aws

Guía para desarrolladores

AWS Lake Formation

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

AWS Lake Formation: Guía para desarrolladores

Copyright © 2026 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

¿Qué es AWS Lake Formation? 1
Características de la Lake Formation 2
Ingesta y administración de datos 2
Administración de la seguridad 3
Inclusión de los datos en el Catálogo de datos 5
Cómo funciona 6
Flujo de trabajo de administración de permisos de Lake Formation 7
Permisos de metadatos 8
Administración del acceso al almacenamiento 11
Cross-account intercambio de datos en Lake Formation 13
Componentes de Lake Formation 14
Consola de Lake Formation 14
API e interfaz de línea de comandos de Lake Formation 14
Otros servicios de AWS 15
Terminología de Lake Formation 15
Lago de datos 15
Acceso a los datos 15
Modo de acceso híbrido 16
Esquema 16
Flujo de trabajo 16
Data Catalog 16
Datos subyacentes 17
Principal 17
Administrador de lago de datos 17
AWS integraciones de servicios con Lake Formation 18
Recursos adicionales de Lake Formation 19
Blogs 20
Charlas técnicas y seminarios web 20
Arquitectura moderna 20
Recursos de la malla de datos 21
Guías de prácticas recomendadas 21
Introducción a Lake Formation 21
Introducción 22
Completa la inicial AWS tareas de configuración 22

Inscríbase en un Cuenta de AWS ... 22
Configuración AWS Lake Formation ... 22
Configure los recursos de Lake Formation usando CloudFormation plantilla ... 23
Crear un administrador de lago de datos ... 24
Cambio del modelo de permisos predeterminado o uso del modo de acceso híbrido ... 29
Asignar permisos a usuarios de Lake Formation ... 31
Configurar una ubicación de Amazon S3 para el lago de datos ... 32
(Opcional) Configuración de filtrado de datos externo ... 33
(Opcional) Conceda acceso a la clave de cifrado del Catálogo de datos ... 34
(Opcional) Crear un rol de IAM para flujos de trabajo ... 34
Actualización de los permisos de datos AWS Glue al modelo de Lake Formation ... 36
Acerca de los permisos predeterminados ... 37
Enumerar los permisos existentes ... 38
Configurar permisos de Lake Formation ... 41
Conceder a los usuarios permisos de IAM ... 41
Cambiar al modelo de permisos de Lake Formation ... 42
Paso 5: Proteger los nuevos recursos del Catálogo de datos ... 45
Paso 6: Proporcionar a los usuarios una nueva política de IAM ... 46
Paso 7: Limpiar las políticas de IAM existentes ... 47
Configuración de los puntos de conexión de Amazon VPC (AWS PrivateLink) ... 47
Consideraciones para los puntos de conexión VPC de Lake Formation ... 48
Creación del punto de conexión de VPC de la interfaz para Lake Formation ... 48
Creación de una política de punto de conexión de VPC para Lake Formation ... 49
Tutoriales ... 51
Crear un lago de datos a partir de un AWS CloudTrail origen ... 52
Destinatarios previstos ... 53
Requisitos previos ... 54
Paso 1: Crear un usuario de análisis de datos ... 55
Paso 2: Añadir permisos de lectura AWS CloudTrail inicia sesión en el rol de flujo de trabajo ... 56
Paso 3: Crear un bucket de Amazon S3 para el lago de datos ... 56
Paso 4: Registrar una ruta de Amazon S3 ... 57
Paso 5: Conceder permisos de ubicación de datos ... 57
Paso 6: Crear una base de datos en Data Catalog ... 58
Paso 7: Conceder permisos de datos ... 58
Paso 8: Utilizar un esquema para crear un flujo de trabajo. ... 59

Paso 9: Ejecutar el flujo de trabajo 60
Paso 10: Conceder SELECT en las tablas 61
Paso 11: consulte el lago de datos mediante Amazon Athena 61
Creación de un lago de datos a partir de un origen JDBC 62
Destinatarios previstos 63
Requisitos previos 64
Paso 1: Crear un usuario analista de datos 64
Paso 2: Crea una conexión en AWS Glue 65
Paso 3: Crear un bucket de Amazon S3 para el lago de datos 66
Paso 4: Registrar una ruta de Amazon S3 66
Paso 5: Conceder permisos de ubicación de datos 67
Paso 6: Crear una base de datos en el Catálogo de datos 67
Paso 7: Conceder permisos de datos 67
Paso 8: Utilizar un esquema para crear un flujo de trabajo. 68
Paso 9: Ejecutar el flujo de trabajo 70
Paso 10: Conceder SELECCIONAR en las tablas 71
Paso 11: consulte el lago de datos mediante Amazon Athena 71
Paso 12: Consultar los datos del lago de datos mediante Amazon Redshift Spectrum 72
Paso 13: Conceder o revocar los permisos de Lake Formation mediante Amazon Redshift Spectrum 76
Configuración de permisos para formatos de tablas abiertas en Lake Formation 76
Destinatarios previstos 77
Requisitos previos 78
Paso 1: Aprovisionar recursos 79
Paso 2: Configurar los permisos para una tabla de Iceberg 81
Paso 3: Configurar los permisos para una tabla de Hudi 87
Paso 4: Configurar los permisos para una tabla de Hudi 90
Paso 5: Eliminar AWS recursos 92
Gestión de un lago de datos mediante el control de acceso basado en etiquetas 92
Destinatarios previstos 94
Requisitos previos 95
Paso 1: Aprovisionar recursos 95
Paso 2: Registrar la ubicación de sus datos, crear una ontología de etiquetas LF y conceder permisos 97
Paso 3: Crear bases de datos de Lake Formation 101
Paso 4: Conceder permisos de tabla 111

Paso 5: Ejecutar una consulta en Amazon Athena para verificar los permisos 113
Paso 6: Limpiar AWS los recursos 114
Protección de los lagos de datos con control de acceso a nivel de fila 115
Destinatarios previstos 115
Requisitos previos 116
Paso 1: Aprovisionar recursos 117
Paso 2: Consultar sin filtros de datos 118
Paso 3: Configurar los filtros de datos y conceder permisos 120
Paso 4: Consultar con filtros de datos 122
Paso 5: Limpiar AWS los recursos 124
Comparta sus datos de forma segura con Lake Formation 124
Destinatarios previstos 125
Configurar los ajustes de Lake Formation 127
Paso 1: Aprovisione sus recursos mediante plantillas AWS CloudFormation 129
Paso 2: Requisitos previos para compartir entre cuentas de Lake Formation 131
Paso 3: Implementar el uso compartido entre cuentas mediante el método de control de acceso basado en etiquetas 134
Paso 4: Implementar el método de recurso con nombre 140
Paso 5: Limpiar AWS los recursos 144
Compartir los recursos del catálogo de datos con fuentes externas Cuentas de AWS mediante un control de acceso detallado 144
Destinatarios previstos 146
Requisitos previos 146
Paso 1: Proporcionar acceso específico a otra cuenta 147
Paso 2: Proporcionar acceso específico a un usuario de la misma cuenta 149
Permisos de incorporación a Lake Formation 151
Descripción general de los permisos de Lake Formation 153
Métodos para el control de acceso específico 154
Control de acceso a los metadatos 157
Control de acceso a los datos subyacentes 161
Personas de Lake Formation y referencia de permisos IAM 167
AWS Lake Formation personas 167
AWS políticas gestionadas para Lake Formation 169
Permisos sugeridos para las personas 177
Cambiar la configuración predeterminada de su lago de datos 188
Permisos implícitos de Lake Formation 191

Referencia de permisos de Lake Formation 193
Permisos de Lake Formation por tipo de recurso 193
Concesión y revocación de Lake Formation AWS CLI Comandos de la de 196
Permisos de Lake Formation 201
Integración de IAM Identity Center 217
Requisitos previos para la integración de IAM Identity Center con Lake Formation 218
Conexión de Lake Formation con IAM Identity Center 221
Actualización de la integración de IAM Identity Center 227
Eliminación de una conexión de Lake Formation con IAM Identity Center 228
Concesión de permisos a usuarios y grupos 229
Incluir el contexto de usuario de IAM Identity Center en los registros CloudTrail 232
Añadir una ubicación de Amazon S3 a su lago de datos 234
Requisitos de los roles utilizados para registrar ubicaciones 235
Registro de una ubicación de Amazon S3 243
Registro de una ubicación cifrada de Amazon S3 247
Registrar una ubicación de Amazon S3 en otra AWS inscrita 252
Registrar una ubicación de Amazon S3 cifrada en AWS cuentas 254
Dar de baja el registro de una ubicación de Amazon S3 259
Acceso a las ubicaciones de Amazon S3 260
Modo de acceso híbrido 264
Casos de uso comunes del modo de acceso híbrido 266
Cómo funciona el modo de acceso híbrido 268
Configuración del modo de acceso híbrido: escenarios comunes 270
Eliminación de entidades principales y recursos del modo de acceso híbrido 291
Eliminación de entidades principales y recursos del modo de acceso híbrido 292
Recursos adicionales 293
Creación de objetos en el AWS Glue Data Catalog 293
Creación de un catálogo 294
Creación de una base de datos 295
Creación de tablas 296
Creación de vistas del Catálogo de datos 304
Importación de datos mediante flujos de trabajo 353
Esquemas y flujos de trabajo 354
Creación de un flujo de trabajo 355
Ejecución de un flujo de trabajo 359
Inclusión de los datos en el Catálogo de datos 361

Federación de catálogos 363
Características de Catalog Federation 363
Limitaciones 364
Fedérate para Snowflake 364
Federarse para Databricks 373
Bringing Amazon Redshift data into the Data Catalog 381
Ventajas principales 384
Funciones y responsabilidades 384
Requisitos previos 385
Creación de catálogos federados de Amazon Redshift 392
Visualización de objetos del catálogo 401
Actualización de un catálogo federado 403
Acceso a un catálogo federado compartido 405
Eliminación de un catálogo federado 410
Consulta de catálogos federados 411
Recursos adicionales 412
Federación en orígenes de datos externos 412
Flujo de trabajo 413
Requisitos previos 414
Creación de un catálogo federado 418
Visualización de objetos del catálogo 424
Eliminación de un catálogo federado 426
Consulta de catálogos federados 427
Recursos adicionales 428
Integración de Amazon S3 Tables con Data Catalog y Lake Formation 428
Cómo funciona la integración del Catálogo de datos y Lake Formation 430
Requisitos previos 431
Habilitación de la integración de Tablas de Amazon S3 434
Creación de bases de datos y tablas 439
Registrar un bucket de mesa de Amazon S3 en otra AWS cuenta 442
Concesión de permisos 445
Cambiar los controles de acceso para la integración de S3 Tables 448
Habilite Lake Formation con la integración de S3 Tables 449
Cambio de Lake Formation a IAM 451
Creación de un catálogo administrado de Amazon Redshift 454

Administración de permisos para los datos de un recurso compartido de datos de Amazon Redshift ... 458
Requisitos previos ... 460
Configurar permisos en recursos compartidos de datos de Amazon Redshift ... 461
Consultar las bases de datos federadas ... 465
Administración de los permisos de los conjuntos de datos que utilizan metaalmacenes externos ... 466
Flujo de trabajo ... 468
Requisitos previos ... 469
Conexión del Catálogo de datos a un metaalmacén de Hive externo ... 472
Recursos adicionales ... 475
Administrar los permisos de Lake Formation ... 476
Conceder permisos de ubicación de datos ... 476
Concesión de permisos de localización de datos (misma cuenta) ... 477
Concesión de permisos de ubicación de datos (cuenta externa) ... 480
Concesión de permisos sobre una ubicación de datos compartida con su cuenta ... 483
Concesión de permisos de datos ... 484
Permisos de IAM necesarios para conceder permisos de Lake Formation ... 486
Uso del método de recursos con nombre asignado ... 489
Tag-based control de acceso ... 516
Otorgar permisos de lago de datos mediante el método LF-TBAC ... 578
Attribute-based control de acceso ... 586
Ejemplo de escenario de permisos ... 590
Filtrado de datos y seguridad de celda ... 592
Filtros de datos ... 594
Compatibilidad con PartiQL en expresiones de filtro de filas ... 598
Permisos necesarios para consultar tablas con filtrado a nivel de celda ... 601
Administrar filtros de datos ... 601
Consulta de los permisos de bases de datos y tablas ... 617
Revocación de permisos utilizando la consola ... 622
Cross-account intercambio de datos ... 622
Requisitos previos ... 625
Actualización de los ajustes de la versión entre cuentas para compartir datos ... 630
Compartir tablas y bases de datos del catálogo de datos entre Cuentas de AWS o directores de IAM de cuentas externas ... 636
Conceder permisos en una base de datos o tabla compartida con su cuenta ... 642

Conceder permisos de enlace de recursos ... 644
Acceso a los datos subyacentes de una tabla compartida ... 647
Cross-account CloudTrail registro ... 649
Administrar los permisos entre cuentas mediante ambos AWS Glue y Lake Formation ... 654
Visualización de todas las concesiones entre cuentas mediante la operación de GetResourceShares API ... 656
Acceso y visualización de tablas y bases de datos compartidas del Catálogo de datos ... 658
Aceptar una invitación AWS RAM para compartir recursos ... 659
Visualización de tablas y bases de datos compartidas del Catálogo de datos ... 661
Creación de enlaces de recursos ... 663
Cómo funcionan los enlaces de recursos ... 664
Crear un enlace de recursos a una tabla de Catálogo de datos compartida ... 666
Crear un enlace de recursos a una base de datos compartida ... 670
Gestión de enlaces de recursos en AWS Glue API ... 674
Acceso a las tablas entre regiones ... 678
Flujos de trabajo ... 680
Configuración del acceso a las tablas entre regiones ... 684
Seguridad ... 688
Protección de datos ... 688
Cifrado en reposo ... 690
Seguridad de infraestructuras ... 690
Prevención de la sustitución confusa entre servicios ... 691
Registro de eventos de seguridad en AWS Lake Formation ... 692
Integración con Lake Formation ... 693
Uso de la integración de aplicaciones de Lake Formation ... 693
Cómo funciona la integración de aplicaciones de Lake Formation ... 694
Roles y responsabilidades en la integración de aplicaciones de Lake Formation ... 696
Lake Formation flujo de trabajo para las operaciones de API de integración de aplicaciones ... 697
Registro de un motor de consultas de terceros ... 699
Habilitar los permisos para que un motor de consultas de terceros llame a las operaciones de la API de integración de aplicaciones ... 701
Integración de aplicaciones para un acceso completo a la tabla ... 705
Trabajar con otros AWS servicios ... 708
Amazon Athena ... 711
Compatibilidad con formatos de tablas transaccionales ... 713

Recursos adicionales ... 716
Amazon Redshift Spectrum ... 716
Compatibilidad con tipos de tablas transaccionales ... 717
Recursos adicionales ... 718
AWS Glue ... 719
Compatibilidad con tipos de tablas transaccionales ... 720
Recursos adicionales ... 721
Amazon EMR ... 721
Compatibilidad con formatos de tablas transaccionales ... 722
Recursos adicionales ... 723
Quick ... 723
Recursos adicionales ... 724
AWS CloudTrail Lago ... 724
Registro de llamadas a la API de AWS Lake Formation mediante AWS CloudTrail ... 725
Lea acerca de Lake Formation en CloudTrail ... 725
Descripción de los eventos de Lake Formation ... 726
Rastrea el acceso a los datos CloudTrail con Lake Formation ... 728
Prácticas recomendadas, consideraciones y limitaciones de Lake Formation ... 731
Prácticas recomendadas y consideraciones para uso compartido de datos entre cuentas ... 731
Limitaciones de los roles vinculados a servicios ... 734
Limitaciones de acceso a datos entre regiones ... 735
Vistas, consideraciones y limitaciones del catálogo de datos ... 735
Limitaciones de filtrado de datos ... 737
Notas y restricciones para el filtrado a nivel de columna ... 737
Limitaciones de filtrado por celdas ... 739
Consideraciones y limitaciones del modo de acceso híbrido ... 740
Limitaciones a la hora de incorporar los datos del almacén de datos de Amazon Redshift al AWS Glue Data Catalog ... 742
Limitaciones de integración del catálogo de S3 Tables ... 744
Consideraciones y limitaciones del uso compartido de datos del almacén de metadatos de Hive ... 744
Limitaciones de uso compartido de datos de Amazon Redshift ... 746
Limitaciones de la integración de IAM Identity Center ... 748
Prácticas recomendadas y consideraciones sobre el control de acceso basado en etiquetas de Lake Formation ... 749

Consideraciones, limitaciones y regiones compatibles para el control de acceso basado en atributos 752
Solución de problemas de Lake Formation 753
Solución de problemas generales 753
Error: permisos insuficientes para Lake Formation en <ubicación de Amazon S3>. 753
Error: "Permisos de clave de cifrado insuficientes para la API Glue" 754
Mi consulta Amazon Athena o la de Amazon Redshift que usa manifiestos está fallando 754
Error: "Permiso(s) de Lake Formation insuficientes: se requiere crear etiqueta en el catálogo" 754
Error al eliminar administradores de lagos de datos no válidos 754
Error: «La persona que llama no está autorizada temporalmente para realizar esta operación en el recurso...» 754
Solución de problemas de acceso entre cuentas 754
He concedido un permiso entre cuentas Lake Formation pero el destinatario no puede ver el recurso 755
Las entidades principales de la cuenta de destinatario pueden ver el recurso del Catálogo de datos pero no pueden acceder a los datos subyacentes 756
Error: «La asociación falló porque la persona que llamó no estaba autorizada» al aceptar una AWS RAM invitación para compartir recursos 756
Error: “No está autorizado a conceder permisos para el recurso” 757
Error: «Acceso denegado para recuperar la información de AWS la organización» 757
Error: "Organización <organization-ID> no encontrada" 757
Error: "Permisos de Lake Formation insuficientes: combinación ilegal" 757
ConcurrentModificationException en grant/revoke las solicitudes a cuentas externas 757
Error al utilizar Amazon EMR para acceder a datos compartidos entre cuentas 758
Solución de problemas de esquemas y flujos de trabajo 759
<role-ARN>Mi blueprint falló con el mensaje «Usuario: <user-ARN>no está autorizado para realizar: iam: PassRole on resource:» 759
<role-ARN>Mi flujo de trabajo falló con el mensaje «Usuario: <user-ARN>no está autorizado a realizar: iam: PassRole on resource:» 760
Un rastreador en mi flujo de trabajo falló con "El recurso no existe o el solicitante no está autorizado para acceder a los permisos solicitados" 760
Un rastreador de mi flujo de trabajo falló y dijo: «Se produjo un error (AccessDeniedException) al llamar a la CreateTable operación...» 760
Problemas conocidos de AWS Lake Formation 760
Limitación del filtrado de metadatos de las tablas 761

Problema al renombrar una columna excluida 762
Problema con la eliminación de columnas en tablas CSV 762
Las particiones de tabla deben añadirse bajo una ruta común 762
Problema con la creación de una base de datos durante la creación del flujo de trabajo 762
Problema al eliminar un usuario y, a continuación, volver a crearlo 763
Las operaciones de la API del Catálogo de datos no actualizan el valor del parámetro `IsRegisteredWithLakeFormation` 763
Las operaciones de Lake Formation no admiten el registro AWS Glue de esquemas 763
Mensaje de error actualizado 763
Lake FormationAPI de 765
Permisos 766
— operaciones — 766
— data types — 766
Configuración de lagos de datos 767
— operaciones — 767
— data types — 767
Integración de IAM Identity Center 767
— operaciones — 767
— data types — 767
Modo de acceso híbrido 768
— operaciones — 768
— data types — 766
Expedición de credenciales 768
— operaciones — 768
— data types — 769
Etiquetado 769
— operaciones — 769
— data types — 770
API de filtrado de datos 770
— operaciones — 770
— data types — 770
Tipos de datos comunes 771
ErrorDetail 771
Patrones de cadena 771
Regiones admitidas 773
Disponibilidad general 773

AWS GovCloud (US) 773
Optimización de transacciones y almacenamiento 773
Historial de documentos 776
AWSGlosario de 792
..... dccxciii

¿Qué es AWS Lake Formation?

Bienvenido a la Guía para AWS Lake Formation desarrolladores.

AWS Lake Formation le ayuda a gestionar, proteger y compartir datos a nivel mundial de forma centralizada para el análisis y el aprendizaje automático. Con Lake Formation, puede administrar el control de acceso detallado para los datos de su lago de datos en Amazon Simple Storage Service (Amazon S3) y sus metadatos en AWS Glue Data Catalog.

Lake Formation proporciona su propio modelo de permisos que aumenta el modelo de permisos de IAM. El modelo de permisos de Lake Formation permite un acceso específico a los datos almacenados en los lagos de datos, así como a orígenes de datos externos, como los almacenes de datos de Amazon Redshift, las bases de datos de Amazon DynamoDB y los orígenes de datos de terceros, mediante un sencillo mecanismo de concesión o revocación, muy similar al de un sistema de administración de bases de datos relacionales (RDBMS). Los permisos de Lake Formation se aplican mediante controles granulares a nivel de columna, fila y celda en todos los servicios de AWS análisis y aprendizaje automático, incluidos Amazon Athena Amazon Quick, Amazon Redshift Spectrum, Amazon EMR y. AWS Glue

Con el modo de acceso híbrido de Lake Formation para AWS Glue Data Catalog (Data Catalog), puede proteger los datos catalogados y acceder a ellos mediante los permisos de Lake Formation y las políticas de permisos de IAM para Amazon S3 y AWS Glue sus acciones. Con el modo de acceso híbrido, los administradores de datos pueden incorporar los permisos de Lake Formation de forma selectiva e incremental, centrándose en un caso práctico del lago de datos cada vez.

Lake Formation también le permite compartir datos interna y externamente entre varias AWS organizaciones o directamente con los directores de IAM en otra cuenta Cuentas de AWS, lo que proporciona un acceso detallado a los metadatos del catálogo de datos y a los datos subyacentes.

Temas

- Características de la Lake Formation
- AWS Lake Formation: Cómo funciona
- Componentes de Lake Formation
- Terminología de Lake Formation
- AWS integraciones de servicios con Lake Formation
- Recursos adicionales de Lake Formation
- Introducción a Lake Formation

Características de la Lake Formation

Lake Formation le ayuda a descomponer los silos de datos y a combinar diferentes tipos de datos estructurados y no estructurados en un repositorio centralizado. En primer lugar, identifique los almacenes de datos existentes en Amazon S3 o en bases de datos relacionales y NoSQL, y traslade los datos a su lago de datos. A continuación, rastree, catalogue y prepare los datos para su análisis. Después, proporcione a sus usuarios un acceso seguro de autoservicio a los datos a través de los servicios de análisis que elijan.

Puede utilizar la consola de Lake Formation para crear catálogos federados multinivel en el Catálogo de datos y unificar los datos entre los lagos de datos de Amazon S3 y los almacenes de datos de Amazon Redshift. También puede integrar datos de sus bases de datos operativas, por ejemplo Amazon DynamoDB, y de fuentes de datos de terceros BigQuery, como Google o MySQL, entre otras. El Catálogo de datos proporciona un repositorio de metadatos centralizado que facilita la administración y la detección de datos en sistemas dispares.

Para obtener más información, consulte Llevar sus datos al AWS Glue Data Catalog.

Temas

- Ingesta y administración de datos
- Administración de la seguridad
- Inclusión de los datos en el Catálogo de datos

Ingesta y administración de datos

Importe datos de bases de datos que ya estén en AWS

Tras especificar dónde se encuentran sus bases de datos y proporcione sus credenciales de acceso, Lake Formation lee los datos y sus metadatos (esquema) para comprender el contenido de los orígenes de datos. A continuación, importa los datos a su nuevo lago de datos y registra los metadatos en un catálogo central. Con Lake Formation, puede importar datos de bases de datos MySQL, PostgreSQL, SQL Server, MariaDB y Oracle que se ejecuten en Amazon RDS o estén alojadas en Amazon EC2. Son compatibles tanto la carga masiva de datos como la incremental.

Importar datos de otros orígenes externos

Puede usar Lake Formation para mover datos desde bases de datos en las instalaciones conectándose con Java Database Connectivity (JDBC). Identifique sus fuentes de destino y

proporcione las credenciales de acceso en la consola, y Lake Formation leerá y cargará sus datos en el lago de datos. Para importar datos de bases de datos distintas de las enumeradas anteriormente, puede crear trabajos de ETL personalizados con ellas AWS Glue.

Catalogar y etiquetar sus datos

Puede usar AWS Glue rastreadores para leer sus datos en Amazon S3 y extraer el esquema de bases de datos y tablas y almacenar esos datos en un catálogo de datos en el que se puedan realizar búsquedas. A continuación, utilice Lake Formation Control de acceso basado en etiquetas de Lake Formation (TBAC) para administrar los permisos sobre bases de datos, tablas y columnas. Para obtener más información sobre cómo agregar tablas al Catálogo de datos, consulte Creación de objetos en el AWS Glue Data Catalog.

Administración de la seguridad

Defina y gestione los controles de acceso

Lake Formation proporciona un único lugar para administrar los controles de acceso a los datos de su lago de datos. Puede definir políticas de seguridad que restrinjan el acceso a los datos a nivel de base de datos, tabla, columna, fila y celda. Estas políticas se aplican a usuarios y roles de IAM, y a usuarios y grupos cuando se federan a través de un proveedor de identidades externo. Puede utilizar controles detallados para acceder a los datos protegidos por Lake Formation en Amazon Redshift Spectrum, AWS Glue Athena, ETL y Amazon EMR para Apache Spark. Siempre que cree identidades IAM, asegúrese de seguir las mejores prácticas IAM. Para más información, consulte las mejores prácticas de seguridad en la Guía del usuario de IAM.

Modo de acceso híbrido

El modo de acceso híbrido de Lake Formation proporciona la flexibilidad de habilitar selectivamente los permisos de Lake Formation para bases de datos y tablas en su Catálogo de datos. Con el modo de acceso híbrido, ahora tiene una ruta incremental que le permite establecer los permisos de Lake Formation para un conjunto específico de usuarios sin interrumpir las políticas de permisos de otros usuarios o cargas de trabajo existentes. Para obtener más información, consulte Modo de acceso híbrido.

Implantar el registro de auditoría

Lake Formation proporciona registros de auditoría completos CloudTrail para monitorear el acceso y mostrar el cumplimiento de las políticas definidas centralmente. Puede auditar el historial de acceso a los datos en los servicios de análisis y de machine learning que leen los datos de su lago de datos

a través de Lake Formation. Esto le permite ver qué usuarios o roles han intentado acceder a qué datos, con qué servicios y cuándo. Puede acceder a los registros de auditoría de la misma manera que accede a cualquier otro CloudTrail registro mediante la consola CloudTrail APIs and. Para obtener más información sobre CloudTrail los registros, consulteRegistro de llamadas a la API de AWS Lake Formation mediante AWS CloudTrail.

Seguridad de nivel de fila y celda

Lake Formation proporciona filtros de datos que le permiten restringir el acceso a una combinación de columnas y filas. Utilice la seguridad a nivel de filas y celdas para proteger datos confidenciales como la información de identificación personal (PII). Para obtener más información sobre la seguridad a nivel de fila, consulte Filtrado de datos y seguridad de celda en Lake Formation.

Control de acceso basado en etiquetas

Utilice el control de acceso basado en atributos de Lake Formation para administrar cientos o incluso miles de permisos de datos mediante la creación de etiquetas personalizadas denominadas etiquetas LF. Ahora puede definir etiquetas LF y asociarlas a bases de datos, tablas o columnas. A continuación, comparta el acceso controlado a través de los servicios de análisis, de machine learning (ML) y de extracción, transformación y carga (ETL) para su consumo. Las etiquetas LF garantizan que la gobernanza de los datos pueda escalarse fácilmente sustituyendo las definiciones de políticas de miles de recursos por unas pocas etiquetas lógicas. Lake Formation proporciona una búsqueda basada en texto sobre estos metadatos, para que sus usuarios puedan encontrar rápidamente los datos que necesitan analizar.

Control de acceso basado en atributos

Utilice el control de acceso basado en atributos para conceder el acceso a los objetos del Catálogo de datos. El control de acceso basado en atributos (ABAC) es una estrategia de autorización que define los permisos en función de los atributos. AWS llama a estos atributos etiquetas. Puede utilizar ABAC para conceder acceso a las entidades principales de la misma cuenta o de otra cuenta sobre los recursos del Catálogo de datos. Cualquier entidad principal de IAM con claves y valores coincidentes de la etiqueta de IAM o de la etiqueta de sesión tendrá acceso al recurso. Debe tener permisos concesibles sobre los recursos para poder realizar estas concesiones.

Acceso entre cuentas

Las capacidades de administración de permisos de Lake Formation simplifican la protección y la administración de los lagos de datos distribuidos en varias AWS cuentas mediante un enfoque centralizado, lo que proporciona un control de acceso detallado al catálogo de datos y a las

ubicaciones de Amazon S3. Para obtener más información, consulte Cross-account intercambio de datos en Lake Formation.

Inclusión de los datos en el Catálogo de datos

La capacidad de federación le permite crear catálogos federados y establecer permisos sobre conjuntos de datos almacenados en diferentes orígenes de datos como Amazon Redshift sin necesidad de migrar datos o metadatos a Amazon S3 o AWS Glue Data Catalog. Puede utilizar cualquiera de los métodos siguientes para llevar datos y administrar permisos en conjuntos de datos externos en Lake Formation:

Para obtener más información, consulte Inclusión de los datos en el AWS Glue Data Catalog.

- Incorporación de los datos de los almacenes de datos de Amazon Redshift en el AWS Glue Data Catalog: registre un espacio de nombres o un clúster de Amazon Redshift existente en el Catálogo de datos y cree un catálogo federado de varios niveles en el Catálogo de datos.

  Puede acceder a sus datos mediante cualquier motor de consultas compatible con la especificación OpenAPI de catálogo REST de Apache Iceberg, como Amazon EMR sin servidor y Amazon Athena.

  Para obtener más información, consulte Incorporar los datos de Amazon Redshift al AWS Glue Data Catalog.
- Federación en el catálogo de datos desde fuentes de datos externas: conecte el catálogo de datos a fuentes de datos externas mediante AWS Glue conexiones y cree catálogos federados para administrar de forma centralizada los permisos de acceso a los conjuntos de datos mediante Lake Formation. No es necesario migrar los metadatos al Catálogo de datos.

  Para obtener más información, consulte Federación en fuentes de datos externas en el AWS Glue Data Catalog.
- Integración de los cubos de tablas de Amazon S3 con el catálogo de datos: puede publicar y catalogar las tablas de Amazon S3 como objetos del catálogo de datos y registrar el catálogo como una ubicación de datos de Lake Formation desde la consola de Lake Formation o utilizando AWS Glue APIs.

  Para obtener más información, consulte Integración de Amazon S3 Tables con AWS Glue Data Catalog and AWS Lake Formation.
- Creación de catálogos para administrar las tablas de Amazon Redshift en el Catálogo de datos: es posible que no tenga un clúster productor de Amazon Redshift o un recurso compartido de datos

de Amazon Redshift disponibles en la actualidad, pero desee crear y administrar las tablas de Amazon Redshift mediante el Catálogo de datos. Para empezar, puede crear un catálogo AWS Glue gestionado mediante la `glue:CreateCatalog` API o la AWS Lake Formation consola configurando el tipo de catálogo como `Managed` y `Catalog source` como Redshift.

Para obtener más información, consulte Creación de un catálogo gestionado de Amazon Redshift en AWS Glue Data Catalog.

- Integración de Lake Formation con el uso compartido de datos de Amazon Redshift. Utilice Lake Formation para administrar 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.
- Conexión del catálogo de datos a metaalmacenes externos: conéctese AWS Glue Data Catalog a metaalmacenes externos para gestionar los permisos de acceso a los conjuntos de datos de Amazon S3 mediante Lake Formation. No es necesario migrar los metadatos al Catálogo de datos.

  Para obtener más información, consulte Administración de los permisos de los conjuntos de datos que utilizan metaalmacenes externos.
- Integración de Lake Formation con AWS Data Exchange: Lake Formation admite la concesión de licencias de acceso a sus datos mediante AWS Data Exchange. Si está interesado en licenciar sus datos de Lake Formation, consulte Qué es AWS Data Exchange en la Guía del usuario de AWS Data Exchange .

AWS Lake Formation: Cómo funciona

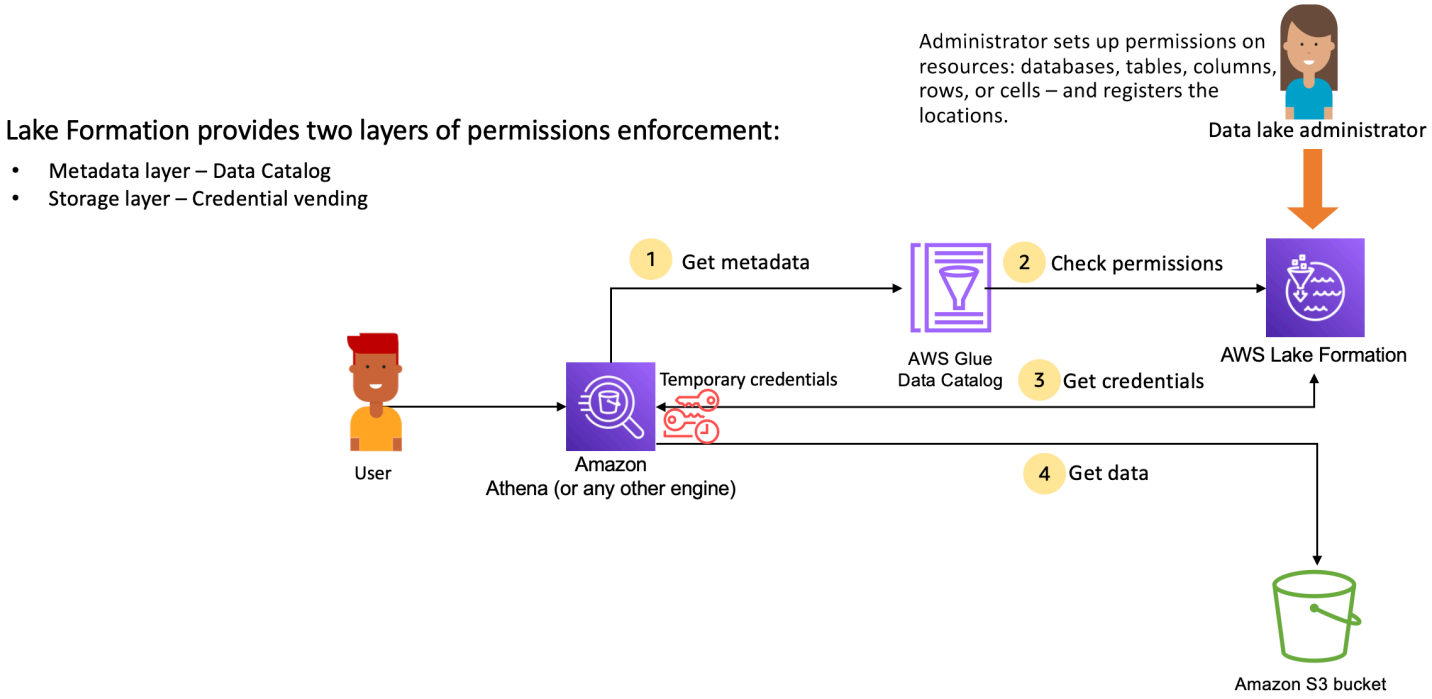
AWS Lake Formation proporciona un modelo de permisos del sistema de administración de bases de datos relacionales (RDBMS) para conceder o revocar el acceso a los recursos del catálogo de datos, como bases de datos, tablas y columnas con datos subyacentes en Amazon S3. Los permisos de Lake Formation, fáciles de gestionar, sustituyen a las complejas políticas de bucket de Amazon S3 y a las correspondientes políticas de IAM.

En Lake Formation, puede implementar permisos en dos niveles:

- Aplicación de permisos a nivel de metadatos en los recursos del Catálogo de datos, como bases de datos y tablas
- Administración de los permisos de acceso al almacenamiento en los datos subyacentes guardados en Amazon S3 en nombre de los motores integrados

Flujo de trabajo de administración de permisos de Lake Formation

Lake Formation se integra con motores de análisis para consultar almacenes de datos de Amazon S3 y objetos de metadatos que están registrados en Lake Formation. El diagrama siguiente ilustra cómo funciona la administración de permisos en Lake Formation.



Pasos generales para la gestión de permisos de Lake Formation

Antes de que Lake Formation pueda proporcionar controles de acceso para los datos de su lago de datos, un administrador del lago de datos o un usuario con permisos administrativos configura las políticas de usuario individuales de las tablas del Catálogo de datos para permitir o denegar el acceso a las tablas del Catálogo de datos utilizando los permisos de Lake Formation.

A continuación, el administrador del lago de datos o un usuario delegado por el administrador concede permisos de Lake Formation a los usuarios en las bases de datos y tablas del Catálogo de datos y registra la ubicación de Amazon S3 de la tabla en Lake Formation.

1. Obtener metadatos: un director (usuario) envía una consulta o un script de ETL a un motor de análisis integrado, como Amazon Athena AWS Glue, Amazon EMR o Amazon Redshift Spectrum. El motor analítico integrado identifica la tabla que se solicita y envía una petición de metadatos al Catálogo de datos.

2. Comprobar permisos. El Catálogo de datos comprueba los permisos del usuario con Lake Formation, y si el usuario está autorizado a acceder a la tabla, devuelve al motor los metadatos que el usuario está autorizado a ver.
3. Obtener credenciales. El Catálogo de datos permite al motor saber si la tabla está administrada por Lake Formation o no. Si los datos subyacentes están registrados en Lake Formation, el motor analítico solicita a Lake Formation que proporcione acceso a los datos mediante la concesión de un acceso temporal.
4. Obtener datos. Si el usuario está autorizado a acceder a la tabla, Lake Formation proporciona acceso temporal al motor analítico integrado. Mediante el acceso temporal, el motor analítico obtiene los datos de Amazon S3 y aplica el filtrado necesario, como el de columnas, filas o celdas. Cuando el motor termina de ejecutar el trabajo, devuelve los resultados al usuario. Este proceso se denomina expedición de credenciales.

   Si la tabla no está administrada por Lake Formation, la segunda llamada del motor analítico se hace directamente a Amazon S3. Para el acceso a los datos se evalúan la política de buckets de Amazon S3 y la política de usuarios de IAM correspondientes.

   Siempre que utilice políticas de IAM, compruebe que sigue las mejores prácticas IAM. Para más información, consulte Prácticas recomendadas de seguridad en IAM en la Guía del usuario de IAM.

Temas

- Permisos de metadatos
- Administración del acceso al almacenamiento
- Cross-account intercambio de datos en Lake Formation

Permisos de metadatos

Lake Formation proporciona autorización y control de acceso para el Catálogo de datos. Cuando un rol de IAM efectúa una llamada a la API del Catálogo de datos desde cualquier sistema, el Catálogo de datos verifica los permisos de datos del usuario y solo devuelve los metadatos a los que el usuario tiene permisos para acceder. Por ejemplo, si un rol de IAM tiene acceso solo a una tabla dentro de una base de datos, y un servicio o un usuario que asume el rol ejecuta la operación `GetTables`, la respuesta contendrá solo esa tabla, con independencia del número de tablas de la base de datos.

Configuración predeterminada: permisos de grupo **`IAMAllowedPrincipal`**

De forma predeterminada, AWS Lake Formation asigna los permisos de todas las bases de datos y tablas a un grupo virtual denominado `IAMAllowedPrincipal`. Este grupo es único y visible solo dentro de Lake Formation. El `IAMAllowedPrincipal` grupo incluye a todos los directores de IAM que tienen acceso a los recursos del catálogo de datos a través de las políticas principales y las políticas de recursos de IAM. AWS Glue Si este permiso existe en una base de datos o tabla, se concederá a todas las entidades principales acceso a la base de datos o tabla.

Si desea proporcionar permisos más específicos en una base de datos o tabla, elimine el permiso `IAMAllowedPrincipal` y Lake Formation aplicará todas las demás políticas asociadas con esa base de datos o tabla. Por ejemplo, si existe una política que permite al Usuario A acceder a la base de datos A con permisos `DESCRIBE`, y el `IAMAllowedPrincipal` existe con todos los permisos, el Usuario A seguirá ejecutando todas las demás acciones, hasta que se revoque el permiso `IAMAllowedPrincipal`.

Además, de forma predeterminada, el grupo `IAMAllowedPrincipal` tiene permisos sobre todas las bases de datos y tablas nuevas cuando se crean. Hay dos configuraciones que controlan este comportamiento. El primero se encuentra en la cuenta y lo habilita para Region-level las bases de datos recién creadas, y el segundo se encuentra en el nivel de la base de datos. Para modificar la configuración predeterminada, consulte Cambio del modelo de permisos predeterminado o uso del modo de acceso híbrido.

Concesión de permisos

Los administradores del lago de datos pueden conceder permisos del Catálogo de datos a las entidades principales para que estas puedan crear y gestionar bases de datos y tablas, y puedan acceder a los datos subyacentes.

Permisos a nivel de bases de datos y tablas

Cuando concede permisos dentro de Lake Formation, el concedente debe especificar la entidad principal a la que conceder permisos, los recursos sobre los que concederlos y las acciones que el concedente debe tener acceso para ejecutar. Para la mayoría de los recursos dentro de Lake Formation, la lista de entidades principales y los recursos para conceder permisos son similares, pero las acciones que puede ejecutar una entidad principal difieren en función del tipo de recurso. Por ejemplo, los permisos `SELECT` están disponibles en las tablas para leerlas, pero los permisos `SELECT` no están permitidos en las bases de datos. El `CREATE_TABLE` permiso está válido en las bases de datos, pero no en las tablas.

Puede conceder AWS Lake Formation permisos mediante dos métodos:

- Método de recurso con nombre. Para elegir los nombres de las bases de datos y las tablas al conceder permisos a los usuarios.
- LF-Tag control de acceso basado (LF-TBAC): los usuarios crean LF-Tags los recursos del catálogo de datos, los asocian a ellos LF-Tags, otorgan `Describe` permisos, los asocian a usuarios individuales y escriben políticas de permisos de LF LF-Tags para diferentes usuarios. Estas LF-Tag-based políticas se aplican a todos los recursos del catálogo de datos asociados a esos LF-Tag valores.

  > **Note**
  >
  > LF-Tags son exclusivos de Lake Formation. Solo son visibles en Lake Formation y no deben confundirse con las etiquetas AWS de recursos.

  LF-TBAC es una función que permite a los usuarios agrupar los recursos en categorías definidas por el usuario LF-Tags y aplicar permisos a esos grupos de recursos. Por lo tanto, es la mejor manera de escalar los permisos a través de un gran número de recursos del Catálogo de datos.

  Para obtener más información, consulte Control de acceso basado en etiquetas de Lake Formation.

Al conceder permisos a una entidad principal, Lake Formation evalúa los permisos como una unión de todas las políticas para ese usuario. Por ejemplo, si tiene dos políticas en una tabla para un principal, donde una política otorga permisos a las columnas col1, col2 y col3 mediante el método de recurso con nombre y la otra política otorga permisos a la misma tabla y principal a col5 y col6 a través LF-Tags, los permisos efectivos serán una unión de los permisos que serían col1, col2, col3, col5 y col6. Esto también incluye filas y filtros de datos.

Permisos de ubicación de datos

Los permisos de ubicación de datos ofrecen a los usuarios no administrativos la posibilidad de crear bases de datos y tablas en ubicaciones específicas de Amazon S3. Si un usuario intenta crear una base de datos o una tabla en una ubicación para la que no tiene permisos, la tarea de creación falla. De este modo, se evita que los usuarios creen tablas en ubicaciones arbitrarias dentro del lago de datos y se controla dónde pueden leer y escribir datos dichos usuarios. Existe un permiso implícito

al crear tablas en la ubicación de Amazon S3 dentro de la base de datos en la que se está creando. Para obtener más información, consulte Conceder permisos de ubicación de datos.

Crear permisos para tablas y bases de datos

De forma predeterminada, los usuarios no administrativos no tienen permisos para crear bases de datos o tablas dentro de una base de datos. La creación de bases de datos se controla a nivel de cuenta mediante los ajustes de Lake Formation, de modo que solo las entidades principales autorizadas pueden crear bases de datos. Para obtener más información, consulte Creación de una base de datos. Para crear una tabla, una entidad principal necesita permiso `CREATE_TABLE` en la base de datos donde se está creando la tabla. Para obtener más información, consulte Creación de tablas.

Permisos implícitos y explícitos

Lake Formation proporciona permisos implícitos en función de la persona y de las acciones que esta lleve a cabo. Por ejemplo, los administradores del lago de datos obtienen automáticamente permisos `DESCRIBE` para todos los recursos del Catálogo de datos, permisos de ubicación de datos para todas las ubicaciones, permisos para crear bases de datos y tablas en todas las ubicaciones, así como y permisos `Grant` y `Revoke` sobre cualquier recurso. Los creadores de bases de datos obtienen automáticamente todos los permisos sobre las bases de datos que crean, y los creadores de tablas obtienen todos los permisos sobre las tablas que crean. Para obtener más información, consulte Permisos implícitos de Lake Formation.

Permisos concedibles

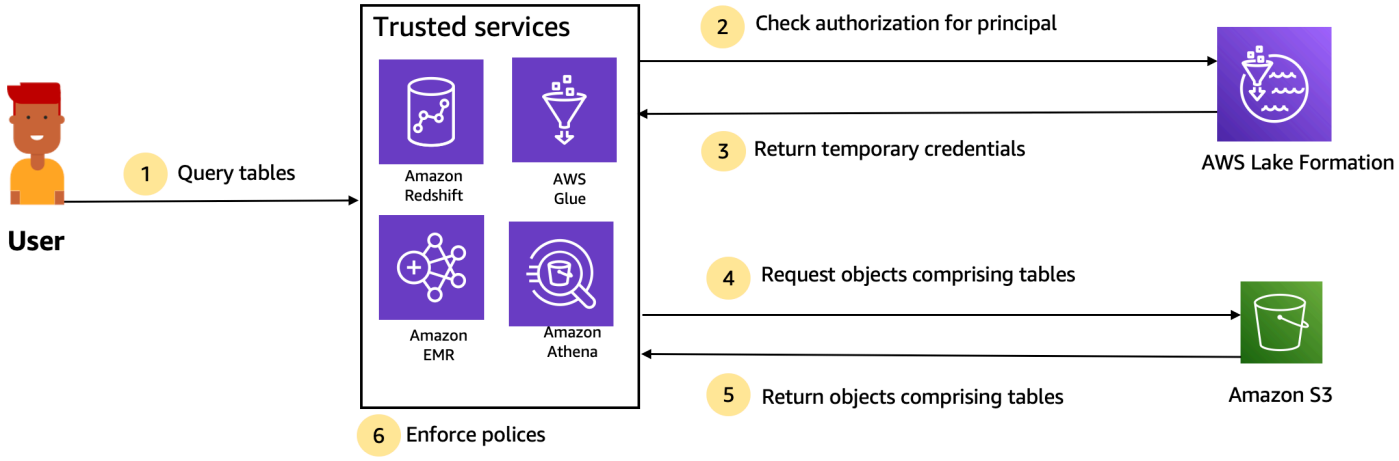
Los administradores del lago de datos tienen la posibilidad de delegar la gestión de los permisos a usuarios no administrativos proporcionándoles permisos concedibles. Cuando a una entidad principal se le proporcionan permisos concedibles sobre un recurso y un conjunto de permisos, esa entidad principal adquiere la capacidad de conceder permisos a otras entidades principales sobre ese recurso.

Administración del acceso al almacenamiento

Lake Formation utiliza la funcionalidad de expendición de credenciales para proporcionar acceso temporal a los datos de Amazon S3. La expedición de credenciales, o expedición de tokens, es un patrón común que proporciona credenciales temporales a usuarios, servicios o alguna otra entidad con el fin de conceder acceso a corto plazo a un recurso.

Lake Formation aprovecha este patrón para proporcionar acceso a corto plazo a servicios de AWS análisis como Athena para acceder a los datos en nombre de la persona que realiza la llamada. Al conceder los permisos, los usuarios no necesitan actualizar sus políticas de bucket de Amazon S3 ni sus políticas de IAM, y no necesitan tener acceso directo a Amazon S3.

El diagrama siguiente muestra cómo Lake Formation proporciona acceso temporal a las ubicaciones registradas:



Trusted services enforce AWS Lake Formation policies (distributed enforcement with fail close).

1. Una entidad principal (usuario) introduce una consulta o solicitud de datos para una tabla a través de un servicio integrado de confianza como Athena, Amazon EMR, Redshift Spectrum o AWS Glue.
2. El servicio integrado comprueba la autorización de Lake Formation para la tabla y las columnas solicitadas y evalúa la autorización. Si el usuario no está autorizado, Lake Formation deniega el acceso a los datos y la consulta falla.
3. En cuanto la autorización tiene éxito y se activa la autorización de almacenamiento para la tabla y el usuario, el servicio integrado recupera las credenciales temporales de Lake Formation para acceder a los datos.
4. El servicio integrado utiliza las credenciales temporales de Lake Formation para solicitar objetos de Amazon S3.
5. Amazon S3 proporciona los objetos de Amazon S3 al servicio integrado. Los objetos de Amazon S3 contienen todos los datos de la tabla.
6. El servicio integrado realiza la aplicación necesaria de las políticas de Lake Formation, como el filtrado a nivel de columna y nivel de fila a nivel de and/or celda. El servicio integrado procesa las consultas y devuelve los resultados al usuario.

Habilitar la aplicación de permisos a nivel de almacenamiento para las tablas del Catálogo de datos

De forma predeterminada, la aplicación a nivel de almacenamiento no está activada para las tablas del Catálogo de datos. Para habilitar la aplicación a nivel de almacenamiento, debe registrar la ubicación en Amazon S3 de sus datos de origen en Lake Formation y proporcionar una función de IAM. Storage-level se habilitarán los permisos para todas las tablas con la misma ruta de ubicación de tabla o el mismo prefijo de la ubicación de Amazon S3.

Cuando un servicio integrado solicita acceso a la ubicación de los datos en nombre de un usuario, el servicio Lake Formation asume este papel y devuelve las credenciales al servicio solicitado con permisos de alcance reducido al recurso para que pueda producirse el acceso a los datos. El rol de IAM registrado debe tener todos los accesos necesarios a la ubicación de Amazon S3, incluidas AWS KMS las claves.

Para obtener más información, consulte Registro de una ubicación de Amazon S3.

Compatible AWS servicios

AWS servicios analíticos como Athena, Redshift Spectrum, Amazon AWS Glue EMR y se Amazon SageMaker AI integran con AWS Lake Formation mediante las Amazon Quick operaciones de la API de venta de credenciales de Lake Formation. Para ver una lista completa de AWS los servicios que se integran con Lake Formation y el nivel de granularidad y los formatos de tabla que admiten, consulteTrabajar con otros AWS servicios.

Cross-account intercambio de datos en Lake Formation

Con Lake Formation, puede compartir los recursos del catálogo de datos (bases de datos y tablas) dentro de una AWS cuenta y entre cuentas en una configuración sencilla mediante el método de recurso designado o LF-Tags. Puede compartir una base de datos completa o seleccionar tablas de una base de datos con cualquiera de los principales de IAM (funciones y usuarios de IAM) de una cuenta, con otras AWS cuentas a nivel de cuenta o directamente con los directores de IAM de otra cuenta.

También puede compartir tablas del Catálogo de datos con filtros de datos para restringir el acceso a los detalles a nivel de fila y a nivel de celda. Lake Formation usa AWS Resource Access Manager (AWS RAM) para facilitar la concesión de permisos entre cuentas. Cuando un recurso se comparte entre dos cuentas, AWS RAM envía invitaciones a la cuenta de destinatario. Cuando un usuario acepta una invitación para AWS RAM compartir, AWS RAM proporciona los permisos necesarios a Lake Formation para que los recursos del catálogo de datos estén disponibles y habilite la aplicación

del nivel de almacenamiento. Para obtener más información, consulte Cross-account intercambio de datos en Lake Formation.

Cuando el administrador del lago de datos de la cuenta del destinatario acepta el AWS RAM recurso compartido, los recursos compartidos están disponibles en la cuenta del destinatario. El administrador del lago de datos concede más permisos de Lake Formation sobre el recurso compartido a entidades principales de IAM adicionales en la cuenta de destinatario, si el administrador tiene permisos `GRANTABLE` sobre el recurso compartido.

Sin embargo, las entidades principales no pueden consultar los recursos compartidos utilizando Athena o Redshift Spectrum sin un enlace de recursos. Un enlace de recursos es una entidad del catálogo de datos y es similar a un Linux-Symlink concepto.

El administrador del lago de datos de la cuenta de destinatario crea un enlace de recurso en el recurso compartido. El administrador concede permisos de `Describe` sobre el enlace de recursos con los permisos requeridos sobre el recurso compartido original a los usuarios adicionales. A continuación, un usuario de la cuenta de destinatario puede utilizar el enlace de recursos para consultar el recurso compartido utilizando Athena y Redshift Spectrum. Para obtener más información sobre los enlaces de recursos, consulte Creación de enlaces de recursos.

Componentes de Lake Formation

AWS Lake Formation se apoya en la interacción de varios componentes para crear y administrar su lago de datos.

Consola de Lake Formation

La consola de Lake Formation se utiliza para definir y gestionar el lago de datos y conceder y revocar los permisos de Lake Formation. Puede usar los esquemas de la consola para descubrir, limpiar, transformar e ingerir datos. También puede habilitar o deshabilitar el acceso a la consola para usuarios individuales de Lake Formation.

API e interfaz de línea de comandos de Lake Formation

Lake Formation proporciona operaciones de API a través de varios SDK específicos de lenguaje y la AWS Command Line Interface (AWS CLI). La API de Lake Formation funciona en combinación con la API de AWS Glue. La API de Lake Formation se centra principalmente en la gestión de los permisos de Lake Formation, mientras que la API de AWS Glue proporciona una API de Catálogo de datos y una infraestructura administrada para definir, programar y ejecutar operaciones de ETL en sus datos.

Para obtener más información acerca de la API de AWS Glue, consulte la Guía para desarrolladores de AWS Glue. Para obtener más información sobre el uso de la AWS CLI, consulte Referencia de comandos de la AWS CLI.

Otros servicios de AWS

Lake Formation utiliza los siguientes servicios:

- AWS Glue para orquestar las tareas y los rastreadores en la transformación de datos mediante las transformaciones de AWS Glue.
- IAM para conceder políticas de permisos a las entidades principales de Lake Formation. El modelo de permisos de Lake Formation amplía el modelo de permisos de IAM para proteger su lago de datos.

Terminología de Lake Formation

A continuación se indican algunos términos importantes que encontrará en esta guía.

Lago de datos

El lago de datos son sus datos persistentes almacenados en Amazon S3 y administrados por Lake Formation mediante un Catálogo de datos. En general, un lago de datos almacena lo siguiente:

- Datos estructurados y no estructurados
- Datos sin procesar y datos transformados

Para que una ruta de Amazon S3 esté dentro de un lago de datos, debe estar registrada en Lake Formation.

Acceso a los datos

Lake Formation proporciona un acceso seguro y específico a los datos a través de un nuevo modelo de concesión/revocación de permisos que aumenta las políticas de AWS Identity and Access Management (IAM).

Los analistas y científicos de datos pueden utilizar toda la cartera de servicios de análisis y machine learning de AWS, como Amazon Athena, para acceder a los datos. Las políticas de seguridad

configuradas de Lake Formation ayudan a garantizar que los usuarios solo puedan acceder a los datos para los que están autorizados.

Modo de acceso híbrido

Gracias al modo de acceso Hybrid, podrá proteger y acceder a los datos catalogados utilizando tanto los permisos de Lake Formation como los de IAM y Amazon S3. El modo de acceso híbrido permite a los administradores de datos incorporar los permisos de Lake Formation de forma selectiva e incremental, centrándose en un caso práctico de lago de datos cada vez.

Esquema

Un esquema es una plantilla de administración de datos que permite incorporar datos fácilmente a un lago de datos. Lake Formation proporciona varios esquemas, cada uno para un tipo de fuente predefinido, como una base de datos relacional o registros de AWS CloudTrail. A partir de un esquema, puede crear un flujo de trabajo. Los flujos de trabajo constan de rastreadores de AWS Glue, trabajos y desencadenadores que se generan para orquestar la carga y actualización de datos. Los esquemas toman como entrada el origen de datos, el destino de estos y la programación para configurar el flujo de trabajo.

Flujo de trabajo

Un flujo de trabajo es un contenedor para un conjunto de trabajos de AWS Glue, rastreadores y desencadenantes relacionados. El flujo de trabajo se crea en Lake Formation y se ejecuta en el servicio AWS Glue. Lake Formation puede seguir el estado de un flujo de trabajo como una entidad única.

Cuando define un flujo de trabajo, selecciona el esquema en el que se basa. A continuación, puede ejecutar flujos de trabajo a petición o según un calendario.

Los flujos de trabajo que cree en Lake Formation son visibles en la consola AWS Glue como un gráfico acíclico dirigido (DAG). Utilizando el DAG, puede seguir el progreso del flujo de trabajo y solucionar problemas.

Data Catalog

El Catálogo de datos es su almacén persistente de metadatos. Se trata de un servicio administrado para almacenar, anotar y compartir metadatos en la nube de AWS del mismo modo que lo haría en un metaalmacén de Apache Hive. Proporciona un repositorio uniforme donde los sistemas dispares

pueden almacenar y encontrar metadatos para rastrear los datos en silos de datos, y luego utilizar esos metadatos para consultar y transformar los datos. Lake Formation utiliza el Catálogo de datos AWS Glue para almacenar metadatos sobre lagos de datos, orígenes de datos, transformaciones y objetivos.

Los metadatos sobre orígenes de datos y objetivos se presentan en forma de bases de datos y tablas. Las tablas almacenan información sobre el esquema, la ubicación, etc. Las bases de datos son colecciones de tablas. Lake Formation proporciona una jerarquía de permisos para controlar el acceso a las bases de datos y tablas del Catálogo de datos.

Cada cuenta AWS dispone de un Catálogo de datos por región de AWS.

Datos subyacentes

Los datos subyacentes se refieren a los datos de origen o datos dentro de los lagos de datos a los que apuntan las tablas del Catálogo de datos.

Principal

Una entidad principal es un usuario o rol de AWS Identity and Access Management (IAM) o un usuario de Active Directory.

Administrador de lago de datos

Un administrador de un lago de datos es una entidad principal que puede conceder a cualquier entidad principal (incluida la propia) permisos sobre cualquier recurso o ubicación de datos del Catálogo de datos. Designe a un administrador del lago de datos como primer usuario del Catálogo de datos. Este usuario puede entonces conceder permisos más específicos de recursos a otras entidades principales.

> **Note**
>
> Los usuarios administrativos de IAM —usuarios con la política administrada `AdministratorAccess` de AWS— no son automáticamente administradores del lago de datos. Por ejemplo, no pueden conceder permisos de Lake Formation sobre objetos del catálogo a menos que se les hayan concedido permisos para hacerlo. Sin embargo, pueden utilizar la consola de Lake Formation o la API para designarse como administradores del lago de datos.

Para obtener información sobre las capacidades de un administrador de lago de datos, consulte Permisos implícitos de Lake Formation. Para obtener información sobre la designación de un usuario como administrador del lago de datos, consulte Crear un administrador de lago de datos.

AWS integraciones de servicios con Lake Formation

Puede utilizar Lake Formation para administrar los permisos de acceso en el nivel de columnas, bases de datos y tablas de los datos almacenados en Amazon S3. Una vez que sus datos estén registrados en Lake Formation, puede utilizar servicios AWS analíticos como AWS Glue Amazon Athena, Amazon Redshift Spectrum o Amazon EMR para consultarlos. Los siguientes AWS servicios se integran con los permisos de Lake Formation AWS Lake Formation y los respetan.

AWS Servicio	Detalles de la integración
AWS Glue	Tema de referencia: Utilizándolo con AWS Lake Formation AWS Glue AWS Glue y Lake Formation comparten el mismo Catálogo de datos. Para las operaciones de la consola (como ver una lista de tablas) y todas las operaciones de la API, los usuarios de AWS Glue solo pueden acceder a las bases de datos y tablas en las que tienen permisos de Lake Formation.
Amazon Athena	Tema de referencia: Uso AWS Lake Formation con Amazon Athena Utilice Lake Formation para conceder o denegar permisos para leer datos en Amazon S3. Cuando los usuarios Amazon Athena seleccion an el catálogo AWS Glue en el editor de consultas, solo pueden consultar las bases de datos, tablas y columnas en las que tienen permisos de Lake Formation. No se admiten las consultas mediante el uso de manifiestos. Actualmente, Lake Formation no admite la administración de permisos en operaciones de escritura como `VACUUM`, `MERGE`, `UPDATE` y `OPTIMIZE` en tablas en formatos de tabla abierta. Además de los directores que se autentican con Athena mediante AWS Identity and Access Management (IAM), Lake Formation apoya

AWS Servicio	Detalles de la integración
	a los usuarios de Athena que se conectan a través del controlador JDBC u ODBC y se autentican a través de SAML. Entre los proveedores SAML compatibles se encuentran Okta y Microsoft Active Directory Federation Service (AD FS).
Amazon Redshift Spectrum	Tema de referencia: Uso AWS Lake Formation con Amazon Redshift Spectrum Cuando los usuarios de Amazon Redshift crean un esquema externo en una base de datos del AWS Glue Data Catalog, solo pueden consultar las tablas y columnas de ese esquema en las que tienen permisos de Lake Formation.
Edición Amazon Quick Enterprise	Referencia: Uso AWS Lake Formation con Quick Cuando un usuario de Amazon Quick Enterprise Edition consulta un conjunto de datos en una ubicación de Amazon S3, el usuario debe tener el `SELECT` permiso de Lake Formation sobre los datos.
Amazon EMR	Referencia: Uso AWS Lake Formation con Amazon EMR Puede integrar los permisos de Lake Formation al crear un clúster de Amazon EMR con un rol de tiempo de ejecución. Una función de tiempo de ejecución es una función de IAM que se asocia a trabajos o consultas de Amazon EMR y, a continuación, Amazon EMR utiliza esta función para acceder a los recursos. AWS

Lake Formation también colabora con AWS Key Management Service (AWS KMS) para permitirle configurar más fácilmente otros servicios integrados para cifrar y descifrar datos en ubicaciones de Amazon Simple Storage Service (Amazon S3).

Recursos adicionales de Lake Formation

Para obtener más información sobre AWS Lake Formation, le recomendamos que continúe profundizando en los conceptos presentados en esta guía mediante los siguientes recursos:

Temas

- Blogs
- Charlas técnicas y seminarios web
- Arquitectura moderna
- Recursos de la malla de datos
- Guías de prácticas recomendadas

Blogs

- Resumen del año 2022 de AWS Lake Formation
- Arquitectura de datos moderna multirregional altamente resiliente
- Uso compartido entre cuentas mediante etiquetas LF para dirigir a las entidades principales de IAM
- Panel de inventario de permisos de Lake Formation
- Malla de datos impulsada por eventos

Charlas técnicas y seminarios web

- re:Invent 2020: Lagos de datos: construya, proteja y comparta fácilmente con AWS Lake Formation
- re:Invent 2022: Creación y funcionamiento de un lago de datos en Amazon S3
- AWS Summit SF 2022: Comprender y conseguir una arquitectura de datos moderna
- AWS Summit ATL 2022: Lagos de datos modernos con AWS Lake Formation, Amazon Redshift y AWS Glue
- AWS Summit ANZ 2022: Lagos de datos, lake houses y malla de datos: ¿qué, por qué y cómo?
- AWS Charlas técnicas en línea: Simplificación de los permisos y la gobernanza del lago de datos

Arquitectura moderna

- Patrones de arquitectura modernos

Recursos de la malla de datos

- Crear una arquitectura de datos moderna y un patrón de malla de datos a escala mediante el control de acceso a AWS Lake Formation basado en etiquetas
- Cómo JPMorgan Chase creó una arquitectura de malla de datos para generar un valor significativo y mejorar su plataforma de datos empresariales
- Crear una malla de datos en AWS

Guías de prácticas recomendadas

- Guías de prácticas recomendadas de AWS Lake Formation

Introducción a Lake Formation

Le recomendamos que lea las siguientes secciones:

- AWS Lake Formation: Cómo funciona. Conozca la terminología esencial y cómo interactúan los distintos componentes.
- Introducción a Lake Formation. Obtenga información sobre los requisitos previos y complete las tareas de configuración importantes.
- AWS Lake Formation tutoriales— Sigue step-by-step los tutoriales para aprender a usar Lake Formation.
- Seguridad en AWS Lake Formation. Comprenda cómo puede ayudar a proteger el acceso a los datos en Lake Formation.

Introducción a Lake Formation

Si no te has registrado AWS o necesitas ayuda para empezar, asegúrate de completar las siguientes tareas.

Temas

- Completa la inicial AWS tareas de configuración
- Configuración AWS Lake Formation
- Actualización AWS Glue de los permisos de datos al AWS Lake Formation modelo
- AWS Lake Formation y puntos finales de VPC de interfaz ()AWS PrivateLink

Completa la inicial AWS tareas de configuración

Para utilizar AWS Lake Formation, primero debe completar las siguientes tareas:

Temas

- Inscríbase en un Cuenta de AWS

Inscríbase en un Cuenta de AWS

Para empezar AWS, necesitas un Cuenta de AWS. Para obtener información sobre cómo crear un Cuenta de AWS, consulte Cómo empezar a utilizar un Cuenta de AWS en la Guía de AWS Account Management referencia.

Configuración AWS Lake Formation

Las siguientes secciones proporcionan información sobre cómo configurar Lake Formation por primera vez. No todos los temas de esta sección son necesarios para empezar a utilizar Lake Formation. Puede usar las instrucciones para configurar el modelo de permisos de Lake Formation para administrar sus AWS Glue Data Catalog objetos y ubicaciones de datos existentes en Amazon Simple Storage Service (Amazon S3).

1. Crear un administrador de lago de datos
2. Cambio del modelo de permisos predeterminado o uso del modo de acceso híbrido

3. the section called “Configurar una ubicación de Amazon S3 para el lago de datos”
4. the section called “Asignar permisos a usuarios de Lake Formation”
5. the section called “Integración de IAM Identity Center”
6. the section called “(Opcional) Configuración de filtrado de datos externo”
7. the section called “(Opcional) Conceda acceso a la clave de cifrado del Catálogo de datos”
8. (Opcional) Crear un rol de IAM para flujos de trabajo

Esta sección muestra cómo configurar los recursos de Lake Formation de dos maneras diferentes:

- Uso de una AWS CloudFormation plantilla
- Uso de la consola de Lake Formation

Para configurar Lake Formation con la AWS consola, vaya aCrear un administrador de lago de datos.

Configure los recursos de Lake Formation usando CloudFormation plantilla

> **Note**
>
> La CloudFormation pila realiza los pasos 1 a 6 de lo anterior, excepto los pasos 2 y 5. Ejecute Cambio del modelo de permisos predeterminado o uso del modo de acceso híbrido y the section called “Integración de IAM Identity Center” manualmente desde la consola de Lake Formation.

1. Inicie sesión en la AWS CloudFormation consola https://console.aws.amazon.com/cloudformation como administrador de IAM en la región EE.UU. Este (Norte de Virginia).
2. Seleccione Lanzar pila.
3. Seleccione Siguiente en la pantalla Crear pila.
4. Introduzca un Nombre de pila.
5. Para DatalakeAdminName y DatalakeAdminPassword, introduzca su nombre de usuario y contraseña para el usuario administrador de Data Lake.
6. Para DatalakeUser1Name y DatalakeUser1Password, introduzca su nombre de usuario y contraseña para el usuario analista de Data Lake.
7. Para DataLakeBucketName, introduzca el nombre del nuevo bucket que se va a crear.

8. Elija Siguiente.
9. En la página siguiente, elija `I acknowledge that CloudFormation might create IAM resources with custom names` y, a continuación, elija Siguiente.
10. Revisa los detalles en la última página y selecciona Acepto que eso AWS CloudFormation podría crear recursos de IAM.
11. Seleccione Crear.

    La creación de la pila puede llevar hasta dos minutos.

Eliminar recursos

Si quieres limpiar los recursos de la CloudFormation pila:

1. De-register el bucket de Amazon S3 que su pila creó y registró como ubicación de lago de datos.
2. Elimine la CloudFormation pila. Esto eliminará todos los recursos creados por la pila.

Crear un administrador de lago de datos

En un principio, los administradores de lagos de datos son los únicos usuarios o funciones AWS Identity and Access Management (IAM) que pueden conceder permisos de Lake Formation sobre las ubicaciones de datos y los recursos del catálogo de datos a cualquier responsable (incluido uno mismo). Para obtener más información acerca de las capacidades de administrador de lago de datos, consulte Permisos implícitos de Lake Formation. De forma predeterminada, con Lake Formation puede crear hasta 30 administradores de lago de datos.

Puede crear un administrador de lagos de datos mediante la consola de Lake Formation o la operación `PutDataLakeSettings` de la API de Lake Formation.

Para crear un administrador de lago de datos, se requieren los permisos siguientes. El usuario `Administrator` tiene estos permisos de forma implícita.

- `lakeformation:PutDataLakeSettings`
- `lakeformation:GetDataLakeSettings`

Si concede a un usuario la política `AWSLakeFormationDataAdmin`, dicho usuario no podrá crear usuarios administradores adicionales de Lake Formation.

Para crear un administrador de lago de datos (consola)

1. Si el usuario que va a ser administrador del lago de datos aún no existe, utilice la consola de IAM para crearlo. De lo contrario, elija un usuario existente para ser administrador del lago de datos.

   > **Note**
   >
   > Le recomendamos que no seleccione un usuario administrativo de IAM (usuario con la política `AdministratorAccess` AWS gestionada) para que sea el administrador del lago de datos.

   Adjunte las siguientes políticas AWS administradas al usuario:

Políticas	¿Obligatoria?	Notas
`AWSLakeFormationDataAdmin`	Obligatorio	Permisos básicos de administrador del lago de datos. Esta política AWS gestionada contiene una denegación explícita de la operación de la API de Lake Formation, `PutDataLakeSetting` que impide a los usuarios crear nuevos administradores de lagos de datos.

Políticas	¿Obligatoria?	Notas
`AWSGlueConsoleFullAccess` , `CloudWatchLogsReadOnlyAccess`	Opcional	Vincule estas políticas si el administrador del lago de datos va a solucionar los flujos de trabajo creados a partir de esquemas de Lake Formation. Estas políticas permiten al administrador del lago de datos consultar la información de solución de problemas en la consola de AWS Glue y en la consola de Amazon CloudWatch Logs . Para obtener más información acerca de los flujos de trabajo, consulte the section called “Importación de datos mediante flujos de trabajo”.
`AWSLakeFormationCrossAccountManager`	Opcional	Vincule esta política para que el administrador del lago de datos pueda conceder y revocar permisos entre cuentas sobre los recursos del Catálogo de datos. Para obtener más información, consulte Cross-account intercambio de datos en Lake Formation.
`AmazonAthenaFullAccess`	Opcional	Adjunte esta política si el administrador del lago de datos va a ejecutar consultas en. Amazon Athena

2. Adjunte la siguiente política en línea, que otorga al administrador del lago de datos permiso para crear el rol vinculado al servicio de Lake Formation. Un posible nombre para la política es `LakeFormationSLR`.

   El rol vinculado al servicio permite al administrador del lago de datos registrar más fácilmente las ubicaciones de Amazon S3 con Lake Formation. Para más información sobre el rol vinculado al servicio Lake Formation, consulte the section called “Cómo utilizar roles vinculados a servicios”.

> **Important**
>
> En todas las políticas siguientes, *<account-id>* sustitúyala por un número de AWS cuenta válido.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "iam:CreateServiceLinkedRole",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "iam:AWSServiceName": "lakeformation.amazonaws.com"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PutRolePolicy"
            ],
            "Resource": "arn:aws:iam::<account-id>:role/aws-service-role/
lakeformation.amazonaws.com/AWSServiceRoleForLakeFormationDataAccess"
        }
    ]
}
```

3. (Opcional) Vincule la siguiente política `PassRole` insertada al usuario. Esta política permite al administrador del lago de datos crear y ejecutar flujos de trabajo. El permiso `iam:PassRole` capacita al flujo de trabajo asumir el rol `LakeFormationWorkflowRole` para crear rastreadores y trabajos, y vincular el rol a los rastreadores y trabajos creados. Un posible nombre para la política es `UserPassRole`.

> **Important**
>
> *<account-id>*Sustitúyalo por un número de AWS cuenta válido.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "PassRolePermissions",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": [
                "arn:aws:iam::111122223333:role/LakeFormationWorkflowRole"
            ]
        }
    ]
}
```

4. (Opcional) Adjunte esta política adicional insertada si su cuenta va a conceder o recibir permisos entre cuentas de Lake Formation. Esta política permite al administrador del lago de datos ver y aceptar AWS Resource Access Manager (AWS RAM) las invitaciones para compartir recursos. Además, para los administradores de lagos de datos de la cuenta AWS Organizations de administración, la política incluye un permiso para habilitar las subvenciones entre cuentas a las organizaciones. Para obtener más información, consulte Cross-account intercambio de datos en Lake Formation.

   Un posible nombre para la política es `RAMAccess`.

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ram:AcceptResourceShareInvitation",
                "ram:RejectResourceShareInvitation",
                "ec2:DescribeAvailabilityZones",
```

```
                "ram:EnableSharingWithAwsOrganization"
            ],
            "Resource": "*"
        }
    ]
}
```

5. Abra la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/ e inicie sesión como usuario con la política `AdministratorAccess` AWS gestionada.
6. Si aparece la ventana de bienvenida a Lake Formation, elija el usuario de IAM que creó o seleccionó en el Paso 1 y, a continuación, elija Comenzar.
7. Si no aparece la ventana de bienvenida a Lake Formation, siga estos pasos para configurar un administrador de Lake Formation.

    a. En el panel de navegación, en Administración, seleccione Roles y tareas administrativas. En la sección de administradores del lago de datos de la página de la consola, seleccione Agregar.

    b. En el cuadro de diálogo Agregar administradores, como tipo de acceso seleccione Administrador del lago de datos.

    c. Para los usuarios y roles de IAM, elija el usuario de IAM que creó o seleccionó en el Paso 1 y, a continuación, seleccione Guardar.

Cambio del modelo de permisos predeterminado o uso del modo de acceso híbrido

Lake Formation comienza con la configuración «Usar solo el control de acceso de IAM» habilitada para que sea compatible con el AWS Glue Data Catalog comportamiento existente. Este ajuste le permite administrar el acceso a sus datos en el lago de datos y sus metadatos mediante políticas de IAM y políticas de bucket de Amazon S3.

Para facilitar la transición de los permisos del lago de datos de un modelo IAM y Amazon S3 a los permisos de Lake Formation, le recomendamos que utilice el modo de acceso híbrido para el Catálogo de datos. Con el modo de acceso híbrido, dispone de una vía incremental en la que puede habilitar los permisos de Lake Formation para un conjunto específico de usuarios sin interrumpir a otros usuarios o cargas de trabajo existentes.

Para obtener más información, consulte Modo de acceso híbrido.

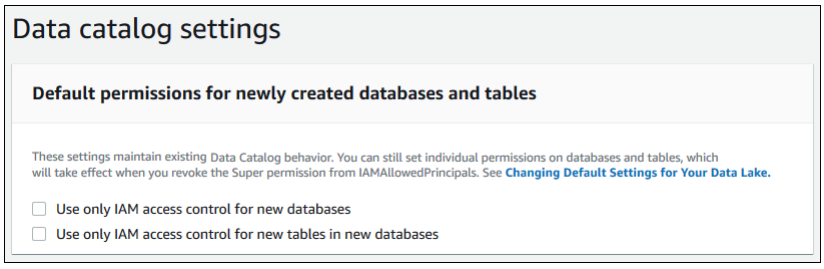
Desactive la configuración predeterminada para mover a todos los usuarios existentes de una tabla a Lake Formation en un solo paso.

> **Important**
>
> Si ya hay bases de datos y tablas de AWS Glue Data Catalog, no siga las instrucciones indicadas en esta sección. En su lugar, siga las instrucciones en the section called "Actualización de los permisos de datos AWS Glue al modelo de Lake Formation".

> **Warning**
>
> Si cuenta con una automatización que crea bases de datos y tablas en el Catálogo de datos, los pasos siguientes podrían provocar un error en los trabajos de automatización y extracción, transformación y carga (ETL) posteriores. Continúe solo después de haber modificado sus procesos existentes o de haber concedido permisos explícitos de Lake Formation a las entidades principales requeridas. Para obtener información sobre los permisos de Lake Formation, consulte the section called "Referencia de permisos de Lake Formation".

Para cambiar la configuración predeterminada del Catálogo de datos

1. Continúe en la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/. Asegúrese de haber iniciado sesión como usuario con la política `AdministratorAccess` AWS administrada.
2. Modifique la configuración del Catálogo de datos:
   a. En el panel de navegación, Administración, seleccione Configuración del Catálogo de datos.
   b. Desmarque ambas casillas y seleccione Guardar.



3. Revoque el permiso de `IAMAllowedPrincipals` a los creadores de bases de datos.

a. En el panel de navegación, en Administración, seleccione Roles y tareas administrativas.

b. En la página de la consola de Roles y tareas administrativas, en la sección Creadores de bases de datos, seleccione el grupo `IAMAllowedPrincipals` y elija Revocar.

   Aparece el cuadro de diálogo para revocar permisos, en el que se muestra que `IAMAllowedPrincipals` tiene el permiso para crear base de datos.

c. Seleccione Revocar.

Asignar permisos a usuarios de Lake Formation

Cree un usuario para tener acceso al lago de datos AWS Lake Formation. Este usuario tiene el privilegio mínimo de permisos para consultar el lago de datos.

Para más información sobre la creación de usuarios o grupos, consulte Identidades IAM en la Guía del usuario de IAM.

Para adjuntar permisos a un usuario no administrador con el fin de acceder a los datos de Lake Formation

1. Abra la consola de IAM en https://console.aws.amazon.com/iam e inicie sesión como usuario con la política `AdministratorAccess` AWS gestionada.
2. Elija Usuarios o Grupos de usuarios.
3. En la lista, seleccione el nombre del usuario o del grupo en el que integrará una política.

   Elija Permisos.
4. Elija Agregar permisos y seleccione Adjuntar políticas directamente. Introduzca `Athena` en el campo de texto Filtrar políticas. En la lista de resultados, marque la casilla de `AmazonAthenaFullAccess`.
5. Pulse el botón Crear política. En la página Crear política, elija la pestaña JSON. Copie la política siguiente y péguela en el editor de políticas.

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
```

```
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess",
                "glue:GetTable",
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetPartitions",
                "lakeformation:GetResourceLFTags",
                "lakeformation:ListLFTags",
                "lakeformation:GetLFTag",
                "lakeformation:SearchTablesByLFTags",
                "lakeformation:SearchDatabasesByLFTags"
            ],
            "Resource": "*"
        }
    ]
}
```

6. Selecciona Siguiente hasta que veas la página de revisión de la política. Escriba un nombre para la política, por ejemplo `DatalakeUserBasic`. Seleccione Crear política y, a continuación, cierre la pestaña Políticas o la ventana del navegador.

Configurar una ubicación de Amazon S3 para el lago de datos

Para utilizar Lake Formation con el fin de administrar y proteger los datos de su lago de datos, primero debe registrar una ubicación de Amazon S3. Al registrar una ubicación, se registran esa ruta de Amazon S3 y todas las carpetas incluidas en esa ruta, lo que permite a Lake Formation aplicar los permisos de nivel de almacenamiento. Cuando el usuario solicita datos de un motor integrado como Amazon Athena, Lake Formation proporciona acceso a los datos en lugar de utilizar los permisos del usuario.

Cuando registras una ubicación, especificas una función de IAM que otorga read/write permisos en esa ubicación. Lake Formation asume esa función al proporcionar credenciales temporales a AWS los servicios integrados que solicitan acceso a los datos de la ubicación registrada de Amazon S3. Puede especificar el rol vinculado al servicio (SLR) de Lake Formation o crear el suyo propio.

Utilice un rol personalizado en las siguientes situaciones:

- Planea publicar las métricas en Amazon CloudWatch Logs. El rol definido por el usuario debe incluir una política para añadir registros a los CloudWatch registros y publicar métricas, además de los permisos de SLR. Para ver un ejemplo de política integrada que otorga los CloudWatch permisos necesarios, consulte. Requisitos de los roles utilizados para registrar ubicaciones
- La ubicación de Amazon S3 está en una cuenta diferente. Para obtener más información, consulte the section called “Registrar una ubicación de Amazon S3 en otra AWS inscrita”.
- La ubicación de Amazon S3 contiene datos cifrados con una Clave administrada de AWS. Para más detalles, consulte Registro de una ubicación cifrada de Amazon S3 y Registrar una ubicación de Amazon S3 cifrada en AWS cuentas.
- Tiene previsto acceder a la ubicación de Amazon S3 mediante Amazon EMR. Para obtener más información sobre los requisitos de los roles, consulte Roles de IAM para Lake Formation en la Guía de administración de Amazon EMR.

El rol que elija debe tener los permisos necesarios, tal y como se describe en Requisitos de los roles utilizados para registrar ubicaciones. Para obtener instrucciones sobre cómo registrar una ubicación de Amazon S3, consulte Añadir una ubicación de Amazon S3 a su lago de datos.

(Opcional) Configuración de filtrado de datos externo

Si tiene intención de analizar y procesar los datos de su lago de datos utilizando motores de consulta de terceros, debe optar por permitir que los motores externos accedan a los datos gestionados por Lake Formation. Si no lo hace, los motores externos no podrán acceder a los datos de las ubicaciones de Amazon S3 que estén registradas en Lake Formation.

Lake Formation es compatible con los permisos a nivel de columna para restringir el acceso a columnas específicas de una tabla. Los servicios analíticos integrados Amazon Athena, como Amazon Redshift Spectrum y Amazon EMR, recuperan los metadatos de las tablas no filtrados del. AWS Glue Data Catalog El filtrado real de las columnas en las respuestas a las consultas es responsabilidad del servicio integrado. Es responsabilidad de los administradores externos gestionar adecuadamente los permisos para evitar el acceso no autorizado a los datos.

Para permitir que motores de terceros accedan a los datos y los filtren (consola)

1. Continúe en la consola de Lake Formation en. https://console.aws.amazon.com/lakeformation/ Compruebe que ha iniciado sesión como una entidad principal con el permiso de IAM en la operación de la API `PutDataLakeSettings` de Lake Formation. El usuario administrador de IAM que ha creado en Inscríbase en un Cuenta de AWS tiene este permiso.

2. En el panel de navegación, Administración, seleccione Configuración de integración de aplicaciones.
3. En la página Configuración de integración de aplicaciones, haga lo siguiente:

    a. Marque la casilla Permitir que motores externos filtren datos en las ubicaciones de Amazon S3 registradas en Lake Formation.

    b. Introduzca los valores de las etiquetas de sesión definidos para motores de terceros.

    c. Para los ID de cuenta de AWS , introduzca los ID de cuenta desde los que los motores de terceros pueden acceder a las ubicaciones registradas en Lake Formation. Pulse Intro después de cada ID de cuenta.

    d. Seleccione Save.

Para permitir que los motores externos accedan a los datos sin validar las etiquetas de sesión, consulte Integración de aplicaciones para un acceso completo a la tabla

(Opcional) Conceda acceso a la clave de cifrado del Catálogo de datos

Si AWS Glue Data Catalog está cifrada, conceda permisos AWS Identity and Access Management (IAM) sobre la AWS KMS clave a los directores que necesiten conceder permisos de Lake Formation en las bases de datos y tablas del catálogo de datos.

Para obtener más información, consulte la Guía para desarrolladores de AWS Key Management Service .

(Opcional) Crear un rol de IAM para flujos de trabajo

Con AWS Lake Formation, puede importar sus datos mediante flujos de trabajo ejecutados por AWS Glue rastreadores. Un flujo de trabajo define el origen de datos y la programación para importar los datos a su lago de datos. Puede definir fácilmente los flujos de trabajo mediante los esquemas o las plantillas proporcionados por Lake Formation.

Cuando creas un flujo de trabajo, debes asignarle una función AWS Identity and Access Management (IAM) que otorgue a Lake Formation los permisos necesarios para ingerir los datos.

En el procedimiento siguiente se presupone que está familiarizado con IAM.

Crear un rol de IAM para flujos de trabajo

1. Abra la consola de IAM en https://console.aws.amazon.com/iam e inicie sesión como usuario con la `AdministratorAccess` AWS política gestionada.
2. En el panel de navegación, seleccione Roles, Crear rol.
3. En la página Crear rol, elija el servicio AWS y, a continuación, Glue. Elija Siguiente.
4. En la página Agregar permisos, busque la política AWSGlueServiceRole gestionada y seleccione la casilla situada junto al nombre de la política en la lista. A continuación, complete el asistente de creación de roles, asignando al rol el nombre `LFWorkflowRole`. Para terminar, seleccione Crear rol.
5. De regreso en la página Roles, busque `LFWorkflowRole` y elija el nombre del rol.
6. En la página Resumen del rol, en la pestaña Permisos, elija Crear política insertada. En la pantalla Crear política, vaya a la pestaña JSON y añada la siguiente política insertada. Un posible nombre para la política es `LakeFormationWorkflow`.

> **Important**
>
> En la siguiente política, *<account-id>* sustitúyala por un Cuenta de AWS número válido.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                 "lakeformation:GetDataAccess",
                 "lakeformation:GrantPermissions"
             ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": ["iam:PassRole"],
            "Resource": [
```

```
                "arn:aws:iam::111122223333:role/LakeFormationWorkflowRole"
            ]
        }
    ]
}
```

A continuación se describen brevemente los permisos de esta política:

- `lakeformation:GetDataAccess` permite que los trabajos creados por el flujo de trabajo escriban en la ubicación de destino.
- `lakeformation:GrantPermissions` permite que el flujo de trabajo conceda el permiso de `SELECT` sobre las tablas de destino.
- `iam:PassRole` permite al servicio asumir el rol de `LakeFormationWorkflowRole` para crear rastreadores y trabajos (instancias de flujos de trabajo) y asociar el rol a los rastreadores y trabajos creados.

7. Compruebe que el rol `LakeFormationWorkflowRole` tenga dos políticas asociadas.
8. Si va a ingerir datos de fuera de la ubicación del lago de datos, añada una política insertada que conceda permisos para leer los datos de origen.

Actualización AWS Glue de los permisos de datos al AWS Lake Formation modelo

AWS Lake Formation los permisos permiten un control de acceso detallado a los datos de su lago de datos. Puede utilizar el modelo de permisos de Lake Formation para gestionar sus AWS Glue Data Catalog objetos y ubicaciones de datos existentes en Amazon Simple Storage Service (Amazon S3).

El modelo de permisos de Lake Formation utiliza permisos generales AWS Identity and Access Management (IAM) para el acceso al servicio de API. Lake Formation usa la funcionalidad Filtrado de datos y seguridad de celda en Lake Formation para restringir el acceso a las tablas en los niveles de columna, fila y celda para los usuarios y sus aplicaciones. En términos comparativos, el modelo AWS Glue concede el acceso a los datos mediante políticas de IAM basadas en identidad y recursos.

Para hacer el cambio, siga los pasos de esta guía.

Para obtener más información, consulte Descripción general de los permisos de Lake Formation.

Acerca de los permisos predeterminados

Para mantener la compatibilidad con versiones anterioresAWS Glue, de forma predeterminada, AWS Lake Formation concede el `Super` permiso al `IAMAllowedPrincipals` grupo sobre todos los recursos del catálogo de AWS Glue datos existentes y concede el `Super` permiso sobre los nuevos recursos del catálogo de datos si está habilitada la configuración de control de acceso para usar solo IAM. Esto hace que el acceso a los recursos del Catálogo de datos y a las ubicaciones de Amazon S3 esté controlado únicamente por las políticas de AWS Identity and Access Management (IAM). El grupo `IAMAllowedPrincipals` incluye a todos los Usuarios y roles de IAM a los que sus políticas de IAM permiten acceder a los objetos de su Catálogo de datos. El permiso `Super` permite a una entidad principal efectuar todas las operaciones compatibles con Lake Formation en la base de datos o tabla sobre la que se concede.

Puede empezar a utilizar Lake Formation para administrar el acceso a sus datos registrando las ubicaciones de los recursos existentes del Catálogo de datos en Lake Formation o utilizando el modo de acceso híbrido. Cuando registre la ubicación de Amazon S3 en modo de acceso híbrido, puede habilitar los permisos de Lake Formation optando por entidades principales para las bases de datos y las tablas bajo esa ubicación.

Para facilitar la transición de los permisos del lago de datos de un modelo IAM y Amazon S3 a los permisos de Lake Formation, le recomendamos que utilice el modo de acceso híbrido para el Catálogo de datos. Con el modo de acceso híbrido, dispone de una vía incremental en la que puede habilitar los permisos de Lake Formation para un conjunto específico de usuarios sin interrumpir a otros usuarios o cargas de trabajo existentes.

Para obtener más información, consulte Modo de acceso híbrido.

Inhabilite la configuración predeterminada del Catálogo de datos para mover todos los usuarios existentes de una tabla a Lake Formation en un solo paso.

Para empezar a utilizar los permisos de Lake Formation con sus bases de datos y tablas existentes del Catálogo de datos AWS Glue, debe hacer lo siguiente:

1. Determine los permisos IAM existentes de sus usuarios para cada base de datos y tabla.
2. Reproduzca estos permisos en Lake Formation.
3. Para cada ubicación de Amazon S3 que contenga datos:
   a. Revoque el permiso `Super` del grupo `IAMAllowedPrincipals` en cada recurso del Catálogo de datos que haga referencia a esa ubicación.

b. Registre la ubicación en Lake Formation.

4. Elimine las políticas de IAM de control de acceso preciso existentes.

> **Important**
>
> Para añadir nuevos usuarios mientras está en el proceso de transición de su Catálogo de datos, debe configurar permisos AWS Glue granulares en IAM como antes. También debe replicar esos permisos en Lake Formation como se describe en esta sección. Si los nuevos usuarios disponen de las políticas de IAM básicas descritas en esta guía, pueden enumerar las bases de datos o tablas que tengan concedido el permiso `Super` a `IAMAllowedPrincipals`. También pueden ver los metadatos de esos recursos.

Siga los pasos de esta sección para actualizar al modelo de permisos Lake Formation.

Temas

- Paso 1: Enumerar los permisos existentes de usuarios y roles
- Paso 2: Configurar permisos equivalentes de Lake Formation
- Paso 3: Conceder a los usuarios permisos de IAM para usar Lake Formation
- Paso 4: Cambiar sus almacenes de datos al modelo de permisos de Lake Formation
- Paso 5: Proteger los nuevos recursos del Catálogo de datos
- Paso 6: Proporcionar a los usuarios una nueva política de IAM para el futuro acceso al lago de datos
- Paso 7: Limpiar las políticas de IAM existentes

Paso 1: Enumerar los permisos existentes de usuarios y roles

Para empezar a utilizar AWS Lake Formation los permisos con las AWS Glue bases de datos y tablas existentes, primero debe determinar los permisos actuales de los usuarios.

> **Important**
>
> Antes de empezar, asegúrese de haber completado las tareas de Introducción.

Temas

- Uso de la operación de la API
- Utilización del Consola de administración de AWS
- Usando AWS CloudTrail

Uso de la operación de la API

Utilice la operación de la ListPoliciesGrantingServiceAccessAPI AWS Identity and Access Management (IAM) para determinar las políticas de IAM asociadas a cada principal (usuario o rol). A partir de las políticas que aparecen en los resultados, puede determinar los permisos IAM que se conceden a la entidad principal. Debe invocar la API para cada entidad principal por separado.

Example

En el siguiente AWS CLI ejemplo, se muestran las políticas asociadas al usuario. `glue_user1`

```
aws iam list-policies-granting-service-access --arn arn:aws:iam::111122223333:user/
glue_user1 --service-namespaces glue
```

El comando devuelve resultados similares al siguiente:

```
{
    "PoliciesGrantingServiceAccess": [
        {
            "ServiceNamespace": "glue",
            "Policies": [
                {
                    "PolicyType": "INLINE",
                    "PolicyName": "GlueUserBasic",
                    "EntityName": "glue_user1",
                    "EntityType": "USER"
                },
                {
                    "PolicyType": "MANAGED",
                    "PolicyArn": "arn:aws:iam::aws:policy/AmazonAthenaFullAccess",
                    "PolicyName": "AmazonAthenaFullAccess"
                }
            ]
        }
```

```
    ],
    "IsTruncated": false
}
```

Utilización del Consola de administración de AWS

También puede ver esta información en la consola AWS Identity and Access Management (IAM), en la pestaña Access Advisor de la página de resumen del usuario o rol:

1. Abra la consola de IAM en https://console.aws.amazon.com/iam/.
2. En el panel de navegación, seleccione Users (Usuarios) o Roles.
3. Elija un nombre de la lista para abrir su página Resumen y elija la pestaña Access Advisor.
4. Inspeccione cada una de las políticas para determinar la combinación de bases de datos, tablas y acciones para las que cada usuario tiene permisos.

   Recuerde inspeccionar los roles además de los usuarios durante este proceso porque sus trabajos de tratamiento de datos podrían estar asumiendo roles para acceder a los datos.

Usando AWS CloudTrail

Otra forma de determinar tus permisos actuales consiste en AWS CloudTrail buscar llamadas a la AWS Glue API en las que el `additionaleventdata` campo de los registros contenga una `insufficientLakeFormationPermissions` entrada. Esta entrada enumera la base de datos y la tabla sobre las que el usuario necesita permisos de Lake Formation para llevar a cabo la misma acción.

Se trata de registros de acceso a datos, por lo que no se garantiza que generen una lista completa de los usuarios y sus permisos. Recomendamos elegir un intervalo de tiempo amplio para capturar la mayoría de los patrones de acceso a los datos de sus usuarios, por ejemplo, varias semanas o meses.

Para obtener más información, consulte Visualización de eventos con el historial de CloudTrail eventos en la Guía del AWS CloudTrail usuario.

A continuación, puede configurar los permisos de Lake Formation para que coincidan con los AWS Glue permisos. Consulte Paso 2: Configurar permisos equivalentes de Lake Formation.

Paso 2: Configurar permisos equivalentes de Lake Formation

Con la información recopilada enPaso 1: Enumerar los permisos existentes de usuarios y roles, conceda AWS Lake Formation permisos que coincidan con los AWS Glue permisos. Utilice cualquiera de los siguientes métodos para las concesiones:

- Utilizar la consola de Lake Formation o la AWS CLI.

  Consulte the section called “Concesión de permisos de datos”.
- Utilice las operaciones `GrantPermissions` o `BatchGrantPermissions` de la API.

  Consulte API de permisos.

Para obtener más información, consulte Descripción general de los permisos de Lake Formation.

Después de configurar los permisos de Lake Formation, continúe con Paso 3: Conceder a los usuarios permisos de IAM para usar Lake Formation.

Paso 3: Conceder a los usuarios permisos de IAM para usar Lake Formation

Para usar el modelo de AWS Lake Formation permisos, los directores deben tener permisos AWS Identity and Access Management (IAM) en Lake Formation. APIs

Cree la siguiente política en IAM y vincúlela a cada usuario que necesite acceder a su lago de datos. Llame a la política `LakeFormationDataAccess`.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "LakeFormationDataAccess",
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess"
            ],
```

```
            "Resource": "*"
        }
    ]
}
```

A continuación, actualice los permisos de Lake Formation una ubicación de datos cada vez. Consulte Paso 4: Cambiar sus almacenes de datos al modelo de permisos de Lake Formation.

Paso 4: Cambiar sus almacenes de datos al modelo de permisos de Lake Formation

Actualice los permisos de Lake Formation una ubicación de datos cada vez. Para ello, repita toda esta sección hasta registrar todas las rutas de Amazon Simple Storage Service (Amazon S3) a las que hace referencia el Catálogo de datos.

Temas

- Verificar los permisos de Lake Formation
- Proteja los recursos del Catálogo de datos existentes
- Active los permisos de Lake Formation para su ubicación de Amazon S3

Verificar los permisos de Lake Formation

Antes de registrar una ubicación, complete una etapa de verificación para asegurarse de que las entidades principales correctas tienen los permisos de Lake Formation necesarios y de que no se conceden permisos de Lake Formation a entidades principales que no deberían tenerlos. Mediante la operación de la API `GetEffectivePermissionsForPath` de Lake Formation, identifique los recursos del Catálogo de datos que hacen referencia a la ubicación de Amazon S3, junto con las entidades principales que tienen permisos sobre dichos recursos.

El siguiente AWS CLI ejemplo devuelve las bases de datos y las tablas del catálogo de datos que hacen referencia al bucket de Amazon S3`products`.

```
aws lakeformation get-effective-permissions-for-path --resource-arn
 arn:aws:s3:::products --profile datalake_admin
```

Observe la opción `profile`. Se recomienda ejecutar el comando como administrador de un lago de datos.

Lo siguiente es un extracto de los resultados obtenidos.

```
{
        "PermissionsWithGrantOption": [
            "SELECT"
        ],
        "Resource": {
            "TableWithColumns": {
                "Name": "inventory_product",
                "ColumnWildcard": {},
                "DatabaseName": "inventory"
            }
        },
        "Permissions": [
            "SELECT"
        ],
        "Principal": {
            "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1",
            "DataLakePrincipalType": "IAM_USER"
        }
 },...
```

> **Important**
>
> Si su Catálogo de datos AWS Glue está cifrado, `GetEffectivePermissionsForPath` devuelve solo las bases de datos y tablas que se crearon o modificaron después de la disponibilidad general de Lake Formation.

Proteja los recursos del Catálogo de datos existentes

A continuación, revoque el permiso `Super` de cada tabla `IAMAllowedPrincipals` y base de datos que haya identificado para la ubicación.

> **Warning**
>
> Si cuenta con una automatización que crea bases de datos y tablas en el Catálogo de datos, los pasos siguientes podrían provocar un error en los trabajos de automatización y extracción, transformación y carga (ETL) posteriores. Continúe solo después de haber

modificado sus procesos existentes o de haber concedido permisos explícitos de Lake Formation a las entidades principales requeridas. Para obtener información sobre los permisos de Lake Formation, consulte the section called “Referencia de permisos de Lake Formation”.

Para revocar **`Super`** de **`IAMAllowedPrincipals`** sobre una tabla

1. Abra la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/. Inicie sesión como administrador del lago de datos.
2. En el panel de navegación, elija Tablas.
3. En la página Tablas, seleccione el botón que hay junto a la tabla que desee.
4. En el menú Acciones, seleccione Restaurar.
5. En el cuadro de diálogo Revocar permisos, en la lista de usuarios y funciones de IAM, desplácese hacia abajo hasta el encabezado Grupo y elija IAMAllowed Principales.
6. En Permisos de tabla, asegúrese de que la opción Super esté seleccionada y, a continuación, elija Revocar.

Para revocar **`Super`** desde **`IAMAllowedPrincipals`** sobre una base de datos

1. Abra la consola en AWS Lake Formation . https://console.aws.amazon.com/lakeformation/ Inicie sesión como administrador del lago de datos.
2. En el panel de navegación, elija Databases (Bases de datos).
3. En la página Bases de datos, seleccione el botón que hay junto a la base de datos que desee.
4. En el menú Actions, seleccione Editar.
5. En la página Editar base de datos, desactive Utilizar solo el control de acceso IAM para las nuevas tablas de esta base de datos y, a continuación, seleccione Guardar.
6. De vuelta en la página Bases de datos, asegúrese de que la base de datos sigue seleccionada y, a continuación, en el menú Acciones, elija Revocar.
7. En el cuadro de diálogo Revocar permisos, en la lista de usuarios y funciones de IAM, desplácese hacia abajo hasta el encabezado Grupo y elija IAMAllowed Principales.
8. En Permisos de base de datos, asegúrese de que la opción Super está seleccionada y, a continuación, elija Revocar.

Active los permisos de Lake Formation para su ubicación de Amazon S3

A continuación, registre la ubicación de Amazon S3 con Lake Formation. Para ello, puede utilizar el proceso descrito en Añadir una ubicación de Amazon S3 a su lago de datos. O bien, utilice la operación de la API `RegisterResource`, descrita en API de expedición de credenciales.

> **Note**
>
> Si se registra una ubicación principal, no es necesario registrar las ubicaciones secundarias.

Tras finalizar estos pasos y comprobar que sus usuarios pueden acceder a sus datos, habrá actualizado con éxito los permisos de Lake Formation. Continúe en el paso siguiente, Paso 5: Proteger los nuevos recursos del Catálogo de datos.

Paso 5: Proteger los nuevos recursos del Catálogo de datos

A continuación, proteja todos los nuevos recursos del Catálogo de datos cambiando la configuración predeterminada del Catálogo de datos. Desactive las opciones para usar únicamente el control de acceso AWS Identity and Access Management (IAM) para las nuevas bases de datos y tablas.

> **Warning**
>
> Si cuenta con una automatización que crea bases de datos y tablas en el Catálogo de datos, los pasos siguientes podrían provocar un error en los trabajos de automatización y extracción, transformación y carga (ETL) posteriores. Continúe solo después de haber modificado sus procesos existentes o de haber concedido permisos explícitos de Lake Formation a las entidades principales requeridas. Para obtener información sobre los permisos de Lake Formation, consulte the section called “Referencia de permisos de Lake Formation”.

Para cambiar la configuración predeterminada del Catálogo de datos

1. Abra la AWS Lake Formation consola en. https://console.aws.amazon.com/lakeformation/ Inicie sesión como usuario administrativo de IAM (el usuario `Administrator` u otro usuario con la política `AdministratorAccess` AWS gestionada).
2. En el panel de navegación, seleccione Configuración.

3. En la página Configuración del Catálogo de datos, desactive ambas casillas de verificación y, a continuación, seleccione Guardar.

El siguiente paso es conceder a los usuarios acceso a bases de datos o tablas adicionales en el futuro. Consulte Paso 6: Proporcionar a los usuarios una nueva política de IAM para el futuro acceso al lago de datos.

Paso 6: Proporcionar a los usuarios una nueva política de IAM para el futuro acceso al lago de datos

Para conceder a sus usuarios acceso a bases de datos o tablas adicionales del catálogo de datos en el futuro, debe proporcionarles la política integral detallada AWS Identity and Access Management (IAM) que se indica a continuación. Llame a la política `GlueFullReadAccess`.

> **Important**
>
> Si adjunta esta política a un usuario antes de revocar `Super` desde `IAMAllowedPrincipals` en cada base de datos y tabla de su Catálogo de datos, ese usuario podrá ver todos los metadatos de cualquier recurso sobre el que se conceda `Super` a `IAMAllowedPrincipals`.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "GlueFullReadAccess",
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess",
                "glue:GetTable",
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetPartitions"
```

```
            ],
            "Resource": "*"
        }
    ]
}
```

> **Note**
>
> Las políticas integradas designadas en este paso y en los anteriores contienen permisos IAM mínimos. Para conocer las políticas sugeridas para los administradores del lago de datos, los analistas de datos y otros personajes, consulte the section called “Personas de Lake Formation y referencia de permisos IAM”.

A continuación, proceda a Paso 7: Limpiar las políticas de IAM existentes.

Paso 7: Limpiar las políticas de IAM existentes

Tras configurar AWS Lake Formation los permisos y crear y adjuntar las políticas generales de control de acceso AWS Identity and Access Management (IAM), complete el siguiente paso final:

- Elimine de los usuarios, grupos y roles las antiguas políticas de IAM de control de acceso específico que replicó en Lake Formation.

Al hacerlo, se asegura de que esas entidades principales ya no tengan acceso directo a los datos de Amazon Simple Storage Service (Amazon S3). A continuación, puede administrar el acceso al lago de datos para esas entidades principales totalmente a través de Lake Formation.

AWS Lake Formation y puntos finales de VPC de interfaz ()AWS PrivateLink

Amazon VPC es un AWS servicio que puede utilizar para lanzar AWS recursos en una red virtual que usted defina. Con una VPC, puede controlar la configuración de la red, como el rango de direcciones IP, las subredes, las tablas de ruteo y las gateways de red.

Si utiliza Amazon Virtual Private Cloud (Amazon VPC) para alojar sus AWS recursos, puede establecer una conexión privada entre su VPC y Lake Formation. Esta conexión se utiliza para que Lake Formation pueda comunicarse con los recursos de su VPC sin pasar por la Internet pública.

Puede establecer una conexión privada entre su VPC y crear un punto final de AWS Lake Formation la VPC de interfaz. Los puntos finales de la interfaz funcionan con una tecnología que le permite acceder de forma privada a Lake Formation APIs sin una puerta de enlace a Internet, un dispositivo NAT, una conexión VPN o una Direct Connect conexión. AWS PrivateLink Las instancias de su VPC no necesitan direcciones IP públicas para comunicarse con Lake Formation. APIs El tráfico entre su VPC y Lake Formation no sale de la red de Amazon.

Cada punto de conexión de la interfaz está representado por una o más interfaces de red elásticas en las subredes.

Para obtener más información, consulte Puntos de conexión de VPC de interfaz (AWS PrivateLink) en la Guía del usuario de Amazon VPC.

Consideraciones para los puntos de conexión VPC de Lake Formation

Antes de configurar un punto de conexión de VPC de interfaz para Lake Formation, asegúrese de revisar las propiedades y limitaciones del punto de conexión de interfaz en la Guía del usuario de Amazon VPC.

Lake Formation es compatible con hacer llamadas a todas sus acciones API desde su VPC. Puede usar Lake Formation con puntos de enlace de VPC en todos los Regiones de AWS que sean compatibles con los puntos de enlace de Lake Formation y Amazon VPC.

Creación del punto de conexión de VPC de la interfaz para Lake Formation

Puede crear un punto final de VPC para el servicio Lake Formation mediante la consola de Amazon VPC o con (). AWS Command Line Interface AWS CLI Para obtener más información, consulte Creación de un punto de conexión de interfaz en la Guía del usuario de Amazon VPC.

Cree un punto de conexión de VPC para Lake Formation utilizando el siguiente nombre de servicio:

- com.amazonaws. *region*.formación lacustre

Si habilita el DNS privado para el punto de conexión, puede hacer solicitudes de API a Lake Formation utilizando su nombre DNS predeterminado para la Región, por ejemplo, `lakeformation.us-east-1.amazonaws.com`.

Para más información, consulte Acceso a un servicio a través de un punto de conexión de interfaz en la Guía del usuario de Amazon VPC.

Creación de una política de punto de conexión de VPC para Lake Formation

Lake Formation es compatible con las políticas de punto de conexión de VPC. Una política de punto final es una política basada en recursos que se adjunta a un punto final de VPC para controlar qué entidades AWS principales pueden usar el punto final para acceder a un servicio. AWS

Puede adjuntar una política a su punto de conexión de VPC que controle el acceso a Lake Formation. La política especifica la siguiente información:

- La entidad principal que puede realizar acciones.
- Las acciones que se pueden realizar.
- Los recursos en los que se pueden llevar a cabo las acciones.

Para más información, consulte Control del acceso a los servicios con puntos de enlace de la VPC en la Guía del usuario de Amazon VPC.

Ejemplo: política de punto de conexión de VPC para acciones de Lake Formation

El ejemplo siguiente de política de punto de conexión de VPC para Lake Formation permite la expedición de credenciales utilizando los permisos de Lake Formation. Puede usar esta política para ejecutar consultas con los permisos de Lake Formation desde un clúster de Amazon Redshift o un Amazon EMR clúster ubicado en una subred privada.

```
{
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "lakeformation:GetDataAccess",
            "Resource": "*",
            "Principal": "*"
        }
    ]
}
```

> **Note**
>
> Si no adjunta una política al crear un punto de conexión, se adjunta una política predeterminada que permite el acceso completo al servicio.

Para obtener más información, consulte estos temas en la documentación de Amazon VPC:

- ¿Qué es Amazon VPC?
- Crear un punto de conexión de interfaz
- Utilizar políticas de punto de conexión de VPC

AWS Lake Formation tutoriales

Los siguientes tutoriales están organizados en tres secciones y proporcionan step-by-step instrucciones sobre cómo crear un lago de datos, ingerir datos, compartir y proteger los lagos de datos mediante AWS Lake Formation:

1. Crear un lago de datos e ingerir datos: aprenda a crear un lago de datos y utilice esquemas para mover, almacenar, catalogar, limpiar y organizar sus datos. También aprenderá a configurar tablas gobernadas. Una tabla gobernada es un nuevo tipo de tabla de Amazon S3 compatible con transacciones atómicas, coherentes, aisladas y duraderas (ACID).

   Antes de comenzar, asegúrese de que ha completado los pasos detallados en Introducción a Lake Formation.

   - Crear un lago de datos a partir de un AWS CloudTrail origen

     Cree y cargue su primer lago de datos utilizando sus propios CloudTrail registros como fuente de datos.

   - Creación de un lago de datos a partir de un origen JDBC en Lake Formation

     Cree un lago de datos utilizando uno de sus almacenes de datos accesibles desde JDBC, como una base de datos relacional, como origen de datos.

2. Proteger los lagos de datos: aprenda a utilizar controles de acceso basados en etiquetas y a nivel de fila para proteger y gestionar eficazmente el acceso a sus lagos de datos.

   - Configuración de permisos para formatos de almacenamiento de tablas abiertas en Lake Formation

     Este tutorial muestra cómo configurar los permisos para los formatos de tablas transaccionales de código abierto (tablas Iceberg de Apache, Hudi de Apache y Linux Foundation Delta Lake) en Lake Formation.

   - Gestión de un lago de datos mediante el control de acceso basado en etiquetas de Lake Formation

     Aprenda a administrar el acceso a los datos dentro de un lago de datos mediante el control de acceso basado en etiquetas en Lake Formation.

   - Protección de los lagos de datos con control de acceso a nivel de fila

Los permisos a nivel de fila de le permiten proporcionar acceso a filas específicas de una tabla en función de las políticas de control y cumplimiento de los datos.

3. Compartir datos: aprenda a compartir sus datos de forma segura entre Cuentas de AWS mediante el control de acceso basado en etiquetas (TBAC) y gestione los permisos específicos en los conjuntos de datos compartidos entre Cuentas de AWS.
   - Compartir un lago de datos utilizando el control de acceso basado en etiquetas de Lake Formation y recursos con nombre

     En este tutorial, aprenderá a compartir sus datos de forma segura entre Cuentas de AWS con Lake Formation.
   - Compartir un lago de datos utilizando el control de acceso específico de Lake Formation

     En este tutorial, aprenderá a compartir rápida y fácilmente conjuntos de datos con Lake Formation al administrar varios Cuentas de AWS con AWS Organizations.

Temas

- Crear un lago de datos a partir de un AWS CloudTrail origen
- Creación de un lago de datos a partir de un origen JDBC en Lake Formation
- Configuración de permisos para formatos de almacenamiento de tablas abiertas en Lake Formation
- Gestión de un lago de datos mediante el control de acceso basado en etiquetas de Lake Formation
- Protección de los lagos de datos con control de acceso a nivel de fila
- Compartir un lago de datos utilizando el control de acceso basado en etiquetas de Lake Formation y recursos con nombre
- Compartir un lago de datos utilizando el control de acceso específico de Lake Formation

Crear un lago de datos a partir de un AWS CloudTrail origen

Este tutorial le guía a través de las acciones que debe realizar en la consola de Lake Formation para crear y cargar su primer lago de datos desde una AWS CloudTrail fuente.

High-level pasos para crear un lago de datos

1. Registre una ruta de Amazon Simple Storage Service (Amazon S3) como lago de datos.

2. Conceda a Lake Formation permisos para escribir en el Catálogo de datos y en las ubicaciones de Amazon S3 del lago de datos.
3. Cree una base de datos para organizar las tablas de metadatos en el Catálogo de datos.
4. Utilice un esquema para crear un flujo de trabajo. Ejecute el flujo de trabajo para incorporar datos de un origen de datos.
5. Configure sus permisos de Lake Formation para permitir que otros administren los datos del Catálogo de datos y del lago de datos.
6. Configure Amazon Athena para consultar los datos que haya importado en su lago de datos de Amazon S3.
7. Para algunos tipos de almacenes de datos, configure Amazon Redshift Spectrum para consultar los datos que importó a su lago de datos de Amazon S3.

Temas

- Destinatarios previstos
- Requisitos previos
- Paso 1: Crear un usuario de análisis de datos
- Paso 2: Añadir permisos de lectura AWS CloudTrail inicia sesión en el rol de flujo de trabajo
- Paso 3: Crear un bucket de Amazon S3 para el lago de datos
- Paso 4: Registrar una ruta de Amazon S3
- Paso 5: Conceder permisos de ubicación de datos
- Paso 6: Crear una base de datos en Data Catalog
- Paso 7: Conceder permisos de datos
- Paso 8: Utilizar un esquema para crear un flujo de trabajo.
- Paso 9: Ejecutar el flujo de trabajo
- Paso 10: Conceder SELECT en las tablas
- Paso 11: consulte el lago de datos mediante Amazon Athena

Destinatarios previstos

En la siguiente tabla se enumeran los roles utilizadas en este tutorial para crear un lago de datos.

Destinatarios previstos

Rol	Description (Descripción)
Administrador de IAM	Tiene la política AWS gestionada:`AdministratorAccess` . Puede crear roles de IAM y buckets de Amazon S3.
Administrador de lagos de datos	Usuario que puede acceder al Catálogo de datos, crear bases de datos y conceder permisos de Lake Formation a otros usuarios. Tiene menos permisos de IAM que el administrador de IAM, pero suficientes para administrar el lago de datos.
Analista de datos	Usuario que puede ejecutar consultas en el lago de datos. Solo tiene permisos suficientes para ejecutar consultas.
Rol de flujo de trabajo	Rol con las políticas de IAM necesarias para ejecutar un flujo de trabajo. Para obtener más información, consulte (Opcional) Crear un rol de IAM para flujos de trabajo.

Requisitos previos

Antes de empezar:

- Asegúrese de haber completado las tareas de Configuración AWS Lake Formation.
- Conozca la ubicación de sus CloudTrail registros.
- Athena requiere que el analista de datos cree un bucket de Amazon S3 para almacenar los resultados de las consultas antes de usar Athena.

Se supone que está familiarizado con el AWS Identity and Access Management (IAM). Para obtener más información acerca de IAM, consulte la Guía del usuario de IAM.

Paso 1: Crear un usuario de análisis de datos

Este usuario tiene el conjunto mínimo de permisos para consultar el lago de datos.

1. Abra la consola de IAM en https://console.aws.amazon.com/iam. Inicie sesión como usuario con la política `AdministratorAccess` AWS gestionada.
2. Cree una tabla con nombre `datalake_user` con la configuración siguiente:
   - Habilite Consola de administración de AWS el acceso.
   - Defina una contraseña y no solicite restablecerla.
   - Adjunte la política `AmazonAthenaFullAccess` AWS gestionada.
   - Vincule la siguiente política insertada. Llame a la política `DatalakeUserBasic`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess",
                "glue:GetTable",
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetPartitions",
                "lakeformation:GetResourceLFTags",
                "lakeformation:ListLFTags",
                "lakeformation:GetLFTag",
                "lakeformation:SearchTablesByLFTags",
                "lakeformation:SearchDatabasesByLFTags"
           ],
            "Resource": "*"
        }
    ]
}
```

Paso 2: Añadir permisos de lectura AWS CloudTrail inicia sesión en el rol de flujo de trabajo

1. Asocie la siguiente política insertada al rol `LakeFormationWorkflowRole`. La política concede permiso para leer tus AWS CloudTrail registros. Llame a la política `DatalakeGetCloudTrail`.

   Para crear la función de `LakeFormationWorkflowRole`, consulte (Opcional) Crear un rol de IAM para flujos de trabajo.

   > **Important**
   >
   > *<your-s3-cloudtrail-bucket>*Sustitúyala por la ubicación de tus CloudTrail datos en Amazon S3.

   JSON

   ```
   {
       "Version":"2012-10-17",
       "Statement": [
           {
               "Effect": "Allow",
               "Action": "s3:GetObject",
               "Resource": ["arn:aws:s3:::<your-s3-cloudtrail-bucket>/*"]
           }
       ]
   }
   ```

2. Compruebe que haya tres políticas adjuntadas al rol.

Paso 3: Crear un bucket de Amazon S3 para el lago de datos

Cree el bucket de Amazon S3 que será la ubicación raíz de su lago de datos.

1. Abra la consola de Amazon S3 en https://console.aws.amazon.com/s3/e inicie sesión como usuario con acceso administrativo.

2. Selecciona Crear depósito y sigue el asistente para crear un depósito con el nombre`<yourName>-datalake-cloudtrail`, donde `<yourName>` aparecen tus iniciales y apellidos. Por ejemplo: `jdoe-datalake-cloudtrail`.

   Para obtener instrucciones detalladas sobre la creación de un bucket de Amazon S3, consulte Crear un bucket.

Paso 4: Registrar una ruta de Amazon S3

Registre una ruta de Amazon S3 como ubicación raíz de su lago de datos.

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/. Inicie sesión como administrador del lago de datos.
2. En el panel de navegación, en Registrar e ingerir, elija las Ubicaciones de los lagos de datos.
3. Seleccione Registrar ubicación y, a continuación, Examinar.
4. Seleccione el bucket `<yourName>-datalake-cloudtrail` que creó anteriormente, acepte el rol de IAM `AWSServiceRoleForLakeFormationDataAccess` predeterminado y, a continuación, elija Registrar ubicación.

   Para obtener más información sobre cómo registrar ubicaciones, consulte Añadir una ubicación de Amazon S3 a su lago de datos.

Paso 5: Conceder permisos de ubicación de datos

Las entidades principales deben tener permisos de ubicación de datos en una ubicación de lago de datos para crear tablas o bases de datos del Catálogo de datos que apunten a esa ubicación. Debe conceder permisos de ubicación de datos al rol de IAM para los flujos de trabajo, de modo que el flujo de trabajo pueda escribir en el destino de la ingesta de datos.

1. En el panel de navegación, bajo Permisos, seleccione Ubicaciones de datos.
2. Elija Conceder y, en el cuadro de diálogo Conceder permisos, seleccione lo siguiente:

   a. En Usuario de IAM y roles, elija `LakeFormationWorkflowRole`.

   b. En Ubicaciones de almacenamiento, elija su bucket de `<yourName>-datalake-cloudtrail`.

3. Elija Conceder.

Para obtener más información sobre permisos de ubicación de datos, consulte Underlying data access control.

Paso 6: Crear una base de datos en Data Catalog

Las tablas de metadatos del Catálogo de datos de Lake Formation se almacenan en una base de datos.

1. En el panel de navegación, Catálogo de datos, elija Tablas.
2. Seleccione Crear base de datos y, en Detalles de la base de datos, introduzca el nombre `lakeformation_cloudtrail`.
3. Deje los demás campos en blanco y elija Crear base de datos.

Paso 7: Conceder permisos de datos

Debe conceder permisos para crear tablas de metadatos en el Catálogo de datos. Como el flujo de trabajo se ejecutará con el rol `LakeFormationWorkflowRole`, debe conceder estos permisos al rol.

1. En la consola de Lake Formation, en el panel de navegación, en Catálogo de datos, seleccione Bases de datos.
2. Elija la base de datos `lakeformation_cloudtrail` y, a continuación, en la lista desplegable Acciones, seleccione Conceder bajo el encabezado Permisos.
3. En el cuadro de diálogo Conceder permisos de datos, seleccione lo siguiente:
    a. En Entidades principales, Usuario de IAM y roles, elija `LakeFormationWorkflowRole`.
    b. En LF-Tags o cataloga los recursos, selecciona los recursos del catálogo de datos con nombre asignado.
    c. En Bases de datos, debería ver que la base de datos `lakeformation_cloudtrail` ya está agregada.
    d. En Permisos de base de datos, seleccione Crear tabla, Modificar y Borrar, y desactive Super si está marcado.
4. Elija Conceder.

Para obtener más información sobre permisos de Lake Formation, consulte Administrar los permisos de Lake Formation.

Paso 8: Utilizar un esquema para crear un flujo de trabajo.

Para leer los CloudTrail registros, entender su estructura y crear las tablas adecuadas en el catálogo de datos, necesitamos configurar un flujo de trabajo que consista en AWS Glue rastreadores, tareas, activadores y flujos de trabajo. Los esquemas de Lake Formation simplifican este proceso.

El flujo de trabajo genera las tareas, los rastreadores y los activadores que descubren e ingieren datos en su lago de datos. Cree un flujo de trabajo basado en uno de los esquemas predefinidos de Lake Formation.

1. En la consola de Lake Formation, en el panel de navegación, elija Esquemas en Ingesta y, a continuación, Usar esquema.
2. En la página Usar un esquema, en Tipo de esquema, elija. AWS CloudTrail
3. En Importar fuente, selecciona una CloudTrail fuente y una fecha de inicio.
4. En Destino de importación, especifique estos parámetros:

Bases de datos de destino	`lakeformation_cloudtrail`
Ubicación de almacenamiento de destino	`s3://<yourName> -datalake-cloudtrail`
Formato de los datos	Parquet

5. Para ver la frecuencia de la importación, seleccione Ejecutar bajo demanda.
6. En Opciones de importación, especifique estos parámetros:

Nombre del flujo de trabajo	`lakeformationcloudtrailtest`
Rol de IAM	`LakeFormationWorkflowRole`
Prefijo de tabla	`cloudtrailtest` **Note** Debe estar en minúscula.

7. Seleccione Crear y espere a que la consola informe de que el flujo de trabajo se ha creado correctamente.

> **Tip**
>
> ¿Ha recibido este mensaje de error?
> `User: arn:aws:iam::<account-id>:user/<datalake_administrator_user> is not authorized to perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/LakeFormationWorkflowRole...`
> Si es así, compruebe que ha sustituido *<account-id>* en la política integrada para el usuario administrador del lago de datos por un número de AWS cuenta válido.

Paso 9: Ejecutar el flujo de trabajo

Como especificó que el flujo de trabajo se ejecuta bajo demanda, debe iniciarlo manualmente.

- En la página Esquemas, seleccione el flujo de trabajo `lakeformationcloudtrailtest` y, en el menú Acciones, elija Iniciar.

  A medida que se ejecuta el flujo de trabajo, puede ver su progreso en la columna Estado de la última ejecución. Pulse el botón de actualización de vez en cuando.

  El estado pasa de EN EJECUCIÓN a Detectando, Importando y FINALIZADO.

  Cuando se complete el flujo de trabajo:

  - El Catálogo de datos tendrá nuevas tablas de metadatos.
  - Sus CloudTrail registros se incorporarán al lago de datos.

  Si se produce un error en el flujo de trabajo, haga lo siguiente:

  a. Seleccione el flujo de trabajo y, en el menú Acciones, elija Ver gráfico.

     El flujo de trabajo se abre en la consola de AWS Glue.

  b. Asegúrese de que se seleccione el flujo de trabajo y elija la pestaña Historial.

  c. En Historial, seleccione la ejecución más reciente y seleccione Ver detalles de la ejecución.

  d. Seleccione un trabajo o un rastreador fallidos en el gráfico dinámico (tiempo de ejecución) y revise el mensaje de error. Los nodos con errores aparecen en rojo o amarillo.

Paso 10: Conceder SELECT en las tablas

Debe conceder el permiso `SELECT` a las nuevas tablas del Catálogo de datos para que el analista de datos pueda consultar los datos a los que apuntan las tablas.

> **Note**
>
> Un flujo de trabajo concede automáticamente el permiso `SELECT` sobre las tablas que crea al usuario que lo ejecutó. Dado que el administrador del lago de datos ejecutó este flujo de trabajo, debe conceder `SELECT` al analista de datos.

1. En la consola de Lake Formation, en el panel de navegación, en Catálogo de datos, seleccione Bases de datos.
2. Elija la base de datos `lakeformation_cloudtrail` y, a continuación, en la lista desplegable Acciones, seleccione Conceder bajo el encabezado Permisos.
3. En el cuadro de diálogo Conceder permisos de datos, seleccione lo siguiente:
   a. En Entidades principales, Usuario de IAM y roles, elija `datalake_user`.
   b. En Recursos de catálogo LF-Tags o recursos, elija Recursos de catálogo de datos con nombre asignado.
   c. En Bases de datos, la base de datos `lakeformation_cloudtrail` ya debería estar seleccionada.
   d. Para Tablas, elija `cloudtrailtest-cloudtrail`.
   e. En Permisos de tabla y columna, elija Seleccionar.
4. Elija Conceder.

El paso siguiente se efectúa como analista de datos.

Paso 11: consulte el lago de datos mediante Amazon Athena

Utilice la Amazon Athena consola para consultar los CloudTrail datos de su lago de datos.

1. Abra la consola de Athena en https://console.aws.amazon.com/athena/e inicie sesión como analista de datos, usuario. `datalake_user`
2. Si es necesario, elija Comenzar para continuar con el editor de consultas de Athena.

3. Para Origen de datos, elija AwsDataCatalog.
4. En Database (Base de datos), elija `lakeformation_cloudtrail`.

   Se rellena la lista de Tablas.
5. En el menú desplegable (3 puntos en horizontal) situado junto a la tabla `cloudtrailtest-cloudtrail`, seleccione Vista previa de la tabla y, a continuación, seleccione Ejecutar.

   La consulta se ejecuta y muestra 10 filas de datos.

   Si no ha utilizado Athena antes, primero debe configurar una ubicación de Amazon S3 en la consola de Athena para almacenar los resultados de las consultas. El `datalake_user` debe disponer de los permisos necesarios para acceder al bucket de Amazon S3 que elija.

> **Note**
>
> Ahora que ha completado el tutorial, conceda permisos de datos y permisos de ubicación de datos a las entidades principales de su organización.

Creación de un lago de datos a partir de un origen JDBC en Lake Formation

Este tutorial le guía por los pasos que debe seguir en la AWS Lake Formation consola para crear y cargar su primer lago de datos desde una fuente JDBC mediante Lake Formation.

Temas

- Destinatarios previstos
- Requisitos previos del tutorial
- Paso 1: Crear un usuario analista de datos
- Paso 2: Crea una conexión en AWS Glue
- Paso 3: Crear un bucket de Amazon S3 para el lago de datos
- Paso 4: Registrar una ruta de Amazon S3
- Paso 5: Conceder permisos de ubicación de datos
- Paso 6: Crear una base de datos en el Catálogo de datos
- Paso 7: Conceder permisos de datos

- Paso 8: Utilizar un esquema para crear un flujo de trabajo.
- Paso 9: Ejecutar el flujo de trabajo
- Paso 10: Conceder SELECCIONAR en las tablas
- Paso 11: consulte el lago de datos mediante Amazon Athena
- Paso 12: Consultar los datos del lago de datos mediante Amazon Redshift Spectrum
- Paso 13: Conceder o revocar los permisos de Lake Formation mediante Amazon Redshift Spectrum

Destinatarios previstos

La siguiente tabla enumera los roles utilizados en este tutorial de JDBC de AWS Lake Formation.

Rol	Description (Descripción)
Administrador de IAM	Un usuario que puede crear usuarios y roles AWS Identity and Access Management (IAM) y buckets de Amazon Simple Storage Service (Amazon S3). Tiene la política `AdministratorAccess` AWS administrada.
Administrador de lago de datos	Usuario que puede acceder al Catálogo de datos, crear bases de datos y conceder permisos de Lake Formation a otros usuarios. Tiene menos permisos de IAM que el administrador de IAM, pero suficientes para administrar el lago de datos.
Analista de datos	Usuario que puede ejecutar consultas en el lago de datos. Solo tiene permisos suficientes para ejecutar consultas.
Rol de flujo de trabajo	Rol con las políticas de IAM necesarias para ejecutar un flujo de trabajo.

Para obtener información sobre los requisitos previos para completar el tutorial, consulte Requisitos previos del tutorial.

Requisitos previos del tutorial

Antes de comenzar eltutorial JDBC de AWS Lake Formation, asegúrese de:

- Completar las tareas de Introducción a Lake Formation.
- Elija el JDBC-accessible banco de datos que desee utilizar para el tutorial.
- Reunir la información necesaria para crear una conexión AWS Glue de tipo JDBC. Este objeto del catálogo de datos incluye la URL del almacén de datos, las credenciales de inicio de sesión y, si el almacén de datos se creó en una Amazon Virtual Private Cloud (Amazon VPC), información de VPC-specific configuración adicional. Para más información, consulte Definición de conexiones en el Catálogo de datos de AWS Glue en la Guía del programador de AWS Glue .

En el tutorial se presupone que está familiarizado con AWS Identity and Access Management (IAM). Para obtener más información acerca de IAM, consulte la Guía del usuario de IAM.

Para empezar, vaya al the section called “Paso 1: Crear un usuario analista de datos”.

Paso 1: Crear un usuario analista de datos

En este paso, debe crear un usuario AWS Identity and Access Management (IAM) para que sea el analista de datos de su lago de datos. AWS Lake Formation

Este usuario tiene el conjunto mínimo de permisos para consultar el lago de datos.

1. Abra la consola de IAM en https://console.aws.amazon.com/iam. Inicie sesión como usuario con la política `AdministratorAccess` AWS gestionada.
2. Cree una tabla con nombre `datalake_user` con la configuración siguiente:
   - Habilite Consola de administración de AWS el acceso.
   - Defina una contraseña y no solicite restablecerla.
   - Adjunte la política `AmazonAthenaFullAccess` AWS gestionada.
   - Vincule la siguiente política insertada. Llame a la política `DatalakeUserBasic`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
```

```
            "Action": [
                "lakeformation:GetDataAccess",
                "glue:GetTable",
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetPartitions",
                "lakeformation:GetResourceLFTags",
                "lakeformation:ListLFTags",
                "lakeformation:GetLFTag",
                "lakeformation:SearchTablesByLFTags",
                "lakeformation:SearchDatabasesByLFTags"
            ],
            "Resource": "*"
        }
    ]
}
```

Paso 2: Crea una conexión en AWS Glue

> **Note**
>
> Omita este paso si ya tiene una conexión en AWS Glue a su origen de datos JDBC.

AWS Lake Formation accede a las fuentes de datos JDBC a través de una conexión. AWS Glue Una conexión es un objeto del Catálogo de datos que contiene toda la información necesaria para conectarse al origen de datos. Puede crear una conexión utilizando la consola de AWS Glue.

Creación de una conexión

1. Abra la AWS Glue consola en https://console.aws.amazon.com/glue/ e inicie sesión como usuario con acceso administrativo.
2. En el panel de navegación, en Data catalog (Catálogo de datos), elija Connections (Conexiones).
3. En la página Conectores, seleccione Crear conexión.
4. En la página Elegir origen de datos, elija JDBC como tipo de conexión. A continuación, elija Siguiente.

5. Continúe con el asistente de conexión y guarde la conexión.

   Para obtener información sobre cómo crear una conexión, consulte Propiedades de la conexión JDBC de AWS Glue en la Guía para desarrolladores de AWS Glue .

Paso 3: Crear un bucket de Amazon S3 para el lago de datos

En este paso, creará el bucket de Amazon Simple Storage Service (Amazon S3) que será la ubicación raíz de su lago de datos.

1. Abra la consola de Amazon S3 en https://console.aws.amazon.com/s3/e inicie sesión como usuario con acceso administrativo.
2. Selecciona Crear depósito y sigue el asistente para crear un depósito con el nombre`<yourName>-datalake-tutorial`, donde `<yourName>` aparecen tu inicial y tu apellido. Por ejemplo: `jdoe-datalake-tutorial`.

   Para obtener instrucciones detalladas sobre la creación de un bucket de Amazon S3, consulte ¿Cómo se crea un bucket de S3? en la Guía del usuario de Amazon Simple Storage Service.

Paso 4: Registrar una ruta de Amazon S3

En este paso, registrará una ruta de Amazon Simple Storage Service (Amazon S3) como ubicación raíz de su lago de datos.

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/. Inicie sesión como administrador del lago de datos.
2. En el panel de navegación, bajo Administración, seleccione Ubicaciones de los lagos de datos.
3. Seleccione Registrar ubicación y, a continuación, Examinar.
4. Seleccione el bucket de `<yourName>-datalake-tutorial` que creó anteriormente, acepte el rol de IAM `AWSServiceRoleForLakeFormationDataAccess` predeterminado y, a continuación, elija Registrar ubicación.

   Para obtener más información sobre cómo registrar ubicaciones, consulte Añadir una ubicación de Amazon S3 a su lago de datos.

Paso 5: Conceder permisos de ubicación de datos

Las entidades principales deben tener permisos de ubicación de datos en una ubicación de lago de datos para crear tablas o bases de datos del Catálogo de datos que apunten a esa ubicación. Debe conceder permisos de ubicación de datos al rol de IAM para los flujos de trabajo, de modo que el flujo de trabajo pueda escribir en el destino de la ingesta de datos.

1. En la consola de Lake Formation, desde panel de navegación, Permisos, seleccione Ubicaciones de datos.
2. Elija Conceder y, en el cuadro de diálogo Conceder permisos:
   a. En Usuario de IAM y roles, seleccione `LakeFormationWorkflowRole`.
   b. En Ubicaciones de almacenamiento, elija su bucket de *<yourName>*`-datalake-tutorial`.
3. Elija Conceder.

Para obtener más información sobre permisos de ubicación de datos, consulte Underlying data access control.

Paso 6: Crear una base de datos en el Catálogo de datos

Las tablas de metadatos del Catálogo de datos de Lake Formation se almacenan en una base de datos.

1. En la consola de Lake Formation, en el panel de navegación, Catálogo de datos, seleccione Bases de datos.
2. Seleccione Crear base de datos y, en Detalles de la base de datos, introduzca el nombre `lakeformation_tutorial`.
3. Deje los demás campos en blanco y elija Crear base de datos.

Paso 7: Conceder permisos de datos

Debe conceder permisos para crear tablas de metadatos en el Catálogo de datos. Como el flujo de trabajo se ejecutará con el rol `LakeFormationWorkflowRole`, debe conceder estos permisos al rol.

1. En la consola de Lake Formation, en el panel de navegación, en Permisos, seleccione Permisos de lago de datos.
2. Seleccione Conceder y, en el cuadro de diálogo Conceder permisos de datos:
   a. En Entidades principales, Usuario de IAM y roles, seleccione `LakeFormationWorkflowRole`.
   b. En LF-Tags o catalogar recursos, elija Recursos de catálogo de datos con nombre.
   c. En Bases de datos, elija la base de datos que creó anteriormente, `lakeformation_tutorial`.
   d. En Permisos de base de datos, seleccione Crear tabla, Modificar y Borrar, y desactive Super si está marcado.
3. Elija Conceder.

Para obtener más información sobre permisos de Lake Formation, consulte Descripción general de los permisos de Lake Formation.

Paso 8: Utilizar un esquema para crear un flujo de trabajo.

El AWS Lake Formation flujo de trabajo genera las AWS Glue tareas, los rastreadores y los activadores que descubren e ingieren datos en su lago de datos. Cree un flujo de trabajo basado en uno de los esquemas predefinidos de Lake Formation.

1. En la consola de Lake Formation, en el panel de navegación, elija Esquemas y, a continuación, Usar esquema.
2. En la página Usar un esquema, en Tipo de esquema, elija Instantánea de base de datos.
3. En Importar origen, para Conexión de base de datos, elija la conexión que acaba de crear, `datalake-tutorial`, o elija una conexión existente para sus orígenes de datos.
4. En Ruta de datos de origen, introduzca la ruta desde la que se van a ingerir los datos en el formulario *<database>*/*<schema>*/*<table>*.

   Puede sustituir el carácter de porcentaje (%) por esquema o tabla. En el caso de las bases de datos que admiten esquemas, introduzca*<database>*/*<schema>*/% para que coincidan con todas las tablas que contiene. *<schema>* *<database>* Oracle Database y MySQL no admiten el esquema en la ruta; en su lugar, introduzca *<database>* /%. En el caso de Oracle Database, *<database>* es el identificador del sistema (SID).

Por ejemplo, si una base de datos Oracle tiene `orcl` como SID, introduzca `orcl/%` para que coincidan con todas las tablas a las que tiene acceso el usuario especificado en la conexión JDCB.

> **Important**
>
> Este campo distingue entre mayúsculas y minúsculas.

5. En Destino de importación, especifique estos parámetros:

Bases de datos de destino	`lakeformation_tutorial`
Ubicación de almacenamiento de destino	`s3://<yourName> -datalake-tutorial`
Formato de los datos	(Elija Parquet o CSV)

6. Para ver la frecuencia de la importación, seleccione Ejecutar bajo demanda.
7. En Opciones de importación, especifique estos parámetros:

Nombre del flujo de trabajo	`lakeformationjdbctest`
Rol de IAM	`LakeFormationWorkflowRole`
Prefijo de tabla	`jdbctest` **Note** Debe estar en minúscula.

8. Seleccione Crear y espere a que la consola informe de que el flujo de trabajo se ha creado correctamente.

> **Tip**
>
> ¿Ha recibido este mensaje de error?

```
User: arn:aws:iam::<account-
id>:user/<datalake_administrator_user> is not authorized to
perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/
LakeFormationWorkflowRole...
```

Si es así, compruebe que ha sustituido *<account-id>* en la política integrada para el usuario administrador del lago de datos por un número de AWS cuenta válido.

Paso 9: Ejecutar el flujo de trabajo

Como especificó que el flujo de trabajo se ejecuta bajo demanda, debe iniciarlo manualmente en. AWS Lake Formation

1. En la consola de Lake Formation, en la página Esquemas, seleccione el flujo de trabajo `lakeformationjdbctest`.
2. Elija Acciones y, a continuación, Pegar.
3. Conforme se ejecuta el flujo de trabajo, vea su progreso en la columna Estado de la última ejecución. Pulse el botón de actualización de vez en cuando.

   El estado pasa de EN EJECUCIÓN a Detectando, Importando y FINALIZADO.

   Cuando finalice el flujo de trabajo:

   - El Catálogo de datos tendrá nuevas tablas de metadatos.
   - Sus datos se incorporan al lago de datos.

   Si el flujo de trabajo falla, haga lo siguiente:

   a. Seleccione el flujo de trabajo. Elija Acciones y, a continuación, elija Ver gráfico.

      El flujo de trabajo se abre en la consola de AWS Glue.

   b. Seleccione un flujo de trabajo y, a continuación, la pestaña Historial.

   c. Seleccione la ejecución más reciente y elija Ver detalles de la ejecución.

   d. Seleccione un trabajo o un rastreador fallidos en el gráfico dinámico (tiempo de ejecución) y revise el mensaje de error. Los nodos con errores aparecen en rojo o amarillo.

Paso 10: Conceder SELECCIONAR en las tablas

Debe conceder el `SELECT` permiso a las nuevas tablas del catálogo de datos AWS Lake Formation para que el analista de datos pueda consultar los datos a los que apuntan las tablas.

> **Note**
>
> Un flujo de trabajo concede automáticamente el permiso `SELECT` sobre las tablas que crea al usuario que lo ejecutó. Como el administrador del lago de datos ejecutó este flujo de trabajo, debe conceder `SELECT` al analista de datos.

1. En la consola de Lake Formation, en el panel de navegación, en Permisos, seleccione Permisos de lago de datos.
2. Seleccione Conceder y, en el cuadro de diálogo Conceder permisos de datos:
   a. En Entidades principales, Usuario de IAM y roles, seleccione `datalake_user`.
   b. En LF-Tags o catalogar recursos, elija Recursos del catálogo de datos con nombre asignado.
   c. En Bases de datos, seleccione `lakeformation_tutorial`.

      Se rellena la lista Tablas.
   d. En Tablas, elija una o más tablas como su origen de datos.
   e. En Permisos de tabla y columna, elija Seleccionar.
3. Elija Conceder.

El paso siguiente se efectúa como analista de datos.

Paso 11: consulte el lago de datos mediante Amazon Athena

Utilice la Amazon Athena consola para consultar los datos de su lago de datos.

1. Abra la consola de Athena en https://console.aws.amazon.com/athena/e inicie sesión como usuario y analista de datos. `datalake_user`
2. Si es necesario, elija Comenzar para continuar con el editor de consultas de Athena.
3. Para Origen de datos, elija AwsDataCatalog.
4. En Database (Base de datos), elija `lakeformation_tutorial`.

Se rellena la lista Tablas.

5. En el menú emergente situado junto a una de las tablas, seleccione Vista previa de la tabla.

   La consulta se ejecuta y muestra 10 filas de datos.

Paso 12: Consultar los datos del lago de datos mediante Amazon Redshift Spectrum

Puede configurar Amazon Redshift Spectrum para consultar los datos que importó a su lago de datos de Amazon Simple Storage Service (Amazon S3). En primer lugar, cree un rol AWS Identity and Access Management (IAM) que se utilice para lanzar el clúster de Amazon Redshift y consultar los datos de Amazon S3. A continuación, conceda a este rol los permisos `Select` en las tablas que desee consultar. Después, conceda permisos al usuario para utilizar el editor de consultas de Amazon Redshift. Por último, cree un clúster de Amazon Redshift y ejecute consultas.

Creará el clúster como administrador y lo consultará como analista de datos.

Para obtener más información sobre Amazon Redshift Spectrum, consulte Uso de Amazon Redshift Spectrum para consultar datos externos en la Guía para desarrolladores de bases de datos de Amazon Redshift.

Para configurar los permisos para ejecutar consultas de Amazon Redshift

1. Abra la consola de IAM en https://console.aws.amazon.com/iam/. Inicie sesión como usuario con la política `AdministratorAccess` AWS administrada.
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. Seleccione la pestaña JSON.
5. Pegue el siguiente documento de política JSON.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
```

```
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess",
                "glue:GetTable",
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetPartitions",
                "lakeformation:GetResourceLFTags",
                "lakeformation:ListLFTags",
                "lakeformation:GetLFTag",
                "lakeformation:SearchTablesByLFTags",
                "lakeformation:SearchDatabasesByLFTags"
            ],
             "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 Revisar política, introduzca el Nombre **`RedshiftLakeFormationPolicy`** para 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.
8. En el panel de navegación de la consola de IAM, seleccione Roles y, a continuación, elija Crear rol.
9. En Select trusted entity (Seleccionar entidad de confianza), elija AWS service (Servicio de).
10. Elija el servicio Amazon Redshift para asumir este rol.
11. Elija el caso de uso Redshift Customizable (Redshift personalizable) para su servicio. A continuación, elija Next: Permissions (Siguiente: permisos).
12. Busque la política de permisos que ha creado, `RedshiftLakeFormationPolicy`, y marque la casilla situada junto al nombre de la política en la lista.
13. Elija Siguiente: Etiquetas.
14. Elija Siguiente: Revisar.
15. En Role name (Nombre del rol), escriba el nombre **`RedshiftLakeFormationRole`**.
16. (Opcional) En Role description (Descripción del rol), escriba una descripción para el nuevo rol.

17. Revise el rol y, a continuación, seleccione Create role (Crear rol).

Para conceder a **`Select`** los permisos de la tabla que se van a consultar en la base de datos de Lake Formation

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/. Inicie sesión como administrador del lago de datos.
2. En el panel de navegación, en Permisos, elija Permisos de lago de datos y seleccione Conceder.
3. Proporcione la información siguiente:
   - Para los usuarios y roles de IAM, elija el rol de IAM que creó, `RedshiftLakeFormationRole`. Cuando se ejecuta el editor de consultas de Amazon Redshift, este utiliza el rol de IAM para dar permiso a los datos.
   - En Database (Base de datos), elija `lakeformation_tutorial`.

     Se rellena la lista de tablas.
   - En Tabla, elija una tabla dentro del origen de datos que desee consultar.
   - Elija el permiso Seleccionar.
4. Elija Conceder.

Para configurar Amazon Redshift Spectrum y ejecutar consultas

1. Abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshift. Inicie sesión como usuario `Administrator`.
2. Elija Create cluster.
3. En la página Crear clúster, introduzca `redshift-lakeformation-demo` el identificador del clúster.
4. Para el tipo de nodo, seleccione dc2.large.
5. Desplácese hacia abajo y, en Configuraciones de base de datos, introduzca o acepte estos parámetros:
   - Nombre de usuario administrador: `awsuser`
   - Contraseña del usuario administrador: (*`Choose a password`*)
6. Amplíe los permisos del clúster y, en la sección Funciones de IAM disponibles, seleccione RedshiftLakeFormationRole. A continuación, seleccione Add IAM role (Añadir rol de IAM).

7. Si debe utilizar un puerto diferente al valor predeterminado de 5439, junto a Configuraciones adicionales, desactive la opción Utilizar valores predeterminados. Amplíe la sección de configuraciones de base de datos e introduzca un nuevo número de puerto de base de datos.
8. Elija Create cluster.

   Se carga la página Clústeres.
9. Espere hasta que el estado del clúster pase a ser Disponible. Seleccione el icono de actualización periódicamente.
10. Conceda permiso al analista de datos para ejecutar consultas en el clúster. Para ello, complete los siguientes pasos.

    a. Abra la consola de IAM en https://console.aws.amazon.com/iam/e inicie sesión como usuario. `Administrator`

    b. En el panel de navegación, seleccione Usuarios y vincule las siguientes políticas administradas al usuario `datalake_user`.

    - `AmazonRedshiftQueryEditor`
    - `AmazonRedshiftReadOnlyAccess`
11. Cierre sesión en la consola de Amazon Redshift y vuelva a iniciarla como usuario `datalake_user`.
12. En la barra de herramientas vertical izquierda, elija el icono EDITOR para abrir el editor de consultas y conectarse al clúster. Si aparece el cuadro de diálogo Conectar a la base de datos, elija el nombre del clúster `redshift-lakeformation-demo` e introduzca el nombre de la base de datos **`dev`**, el nombre de usuario **`awsuser`** y la contraseña que creó. Seleccione Connect to database (Conectar a base de datos).

    > **Note**
    >
    > Si no se le solicitan los parámetros de conexión y ya hay otro clúster seleccionado en el editor de consultas, elija Cambiar conexión para abrir el cuadro de diálogo Conectar a la base de datos.
13. En el cuadro de texto Nueva consulta 1, introduzca y ejecute la siguiente instrucción para asignar la base de datos `lakeformation_tutorial` en Lake Formation al nombre del esquema de Amazon Redshift `redshift_jdbc`:

> **Important**
>
> *<account-id>*Sustitúyalo por un número de AWS cuenta válido y *<region>* por un nombre de AWS región válido (por ejemplo,`us-east-1`).

```
create external schema if not exists redshift_jdbc from DATA CATALOG
 database 'lakeformation_tutorial' iam_role 'arn:aws:iam::<account-id>:role/
RedshiftLakeFormationRole' region '<region>';
```

14. En la lista de esquemas, bajo Seleccionar esquema, elija redshift_jdbc.

    Se rellena la lista de tablas. El editor de consultas muestra solo las tablas en las que se le concedieron permisos para el lago de datos de Lake Formation.

15. En el menú emergente situado junto al nombre de una tabla, seleccione Vista previa de los datos.

    Amazon Redshift devuelve las 10 primeras filas.

    Ahora puede ejecutar consultas en las tablas y columnas para las que tiene permisos.

Paso 13: Conceder o revocar los permisos de Lake Formation mediante Amazon Redshift Spectrum

Amazon Redshift es compatible con la capacidad de conceder y revocar permisos de Lake Formation en bases de datos y tablas mediante instrucciones SQL modificadas. Estas instrucciones son similares a las existentes de Amazon Redshift. Para obtener más información, consulte CONCEDER y REVOCAR en la Guía para desarrolladores de base de datos de Amazon Redshift.

Configuración de permisos para formatos de almacenamiento de tablas abiertas en Lake Formation

AWS Lake Formation admite la administración de permisos de acceso para formatos de tabla abierta (OTF), como Apache Iceberg, Apache Hudi y la fundación Linux Delta Lake. En este tutorial, aprenderá a crear Iceberg, Hudi y Delta Lake con tablas de manifiestos de enlaces simbólicos

durante el AWS Glue Data Catalog uso AWS Glue, a configurar permisos detallados con Lake Formation y a consultar datos con Amazon Athena.

> **Note**
>
> AWS los servicios de análisis no admiten todos los formatos de tablas transaccionales. Para obtener más información, consulte Trabajar con otros AWS servicios. Este tutorial aborda la creación manual de una nueva base de datos y una tabla en el Catálogo de datos utilizando únicamente trabajos de AWS Glue .

Este tutorial incluye una AWS CloudFormation plantilla para una configuración rápida. Puede revisarla y personalizarla para adaptarla a sus necesidades.

Temas

- Destinatarios previstos
- Requisitos previos
- Paso 1: Aprovisionar recursos
- Paso 2: Configurar los permisos para una tabla de Iceberg
- Paso 3: Configurar los permisos para una tabla de Hudi
- Paso 4: Configurar los permisos para una tabla de Hudi
- Paso 5: Eliminar AWS recursos

Destinatarios previstos

Este tutorial está dirigido a administradores de IAM, administradores de lago de datos y analistas empresariales. En la tabla siguiente se enumeran los roles utilizados en este tutorial para crear una tabla gobernada con Lake Formation.

Rol	Description (Descripción)
Administrador de IAM	Usuario que puede crear usuarios de IAM, roles y buckets de Amazon S3. Tiene la política `AdministratorAccess` AWS gestionada.

Rol	Description (Descripción)
Administrador de lago de datos	Usuario que puede acceder al Catálogo de datos, crear bases de datos y conceder permisos de Lake Formation a otros usuarios. Tiene menos permisos de IAM que el administrador de IAM, pero suficientes para administrar el lago de datos.
Analista de negocios	Usuario que puede ejecutar consultas en el lago de datos. Tiene permisos para ejecutar consultas.

Requisitos previos

Antes de empezar este tutorial, debe disponer de una en la Cuenta de AWS que pueda iniciar sesión como usuario con los permisos correctos. Para obtener más información, consulte Inscríbase en un Cuenta de AWS.

En el tutorial, se asume que conoce las políticas y los roles de IAM. Para obtener más información acerca de IAM, consulte la Guía del usuario de IAM.

Debe configurar los siguientes AWS recursos para completar este tutorial:

- Administrador de lago de datos
- Configuración del lago de datos de Lake Formation
- Motor Amazon Athena, versión 3

Para crear un administrador de lago de datos

1. Inicie sesión en la consola de Lake Formation https://console.aws.amazon.com/lakeformation/como usuario administrador. En este tutorial, vamos a crear recursos en la región Este de EE. UU. (Norte de Virginia).
2. Desde la consola de Lake Formation, en Permisos del panel de navegación, seleccione Roles y tareas administrativas.
3. Seleccione Elegir administradores en Administradores del lago de datos.

4. En la ventana emergente para administrar gestores de lagos de datos, en Usuarios y roles de IAM, seleccione Usuario de IAM administrador.
5. Seleccione Save.

Para habilitar la configuración del lago de datos

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/. En el panel de navegación, en Catálogo de datos, seleccione Configuración. Desmarque lo siguiente:
   - Usar solo el control de acceso de IAM para nuevas bases de datos.
   - Usar solo el control de acceso de IAM para nuevas tablas en bases de datos nuevas.
2. En Configuración de la versión entre cuentas, seleccione Versión 3 como versión entre cuentas.
3. Seleccione Save.

Para actualizar el motor Amazon Athena a la versión 3

1. Abra la consola Athena en. https://console.aws.amazon.com/athena/
2. Seleccione el grupo de trabajo y elija el principal.
3. Asegúrese de que el grupo de trabajo tenga una versión mínima de 3. Si no es así, edite el grupo de trabajo, elija Manual en Actualizar el motor de consultas y seleccione la versión 3.
4. Seleccione Guardar cambios.

Paso 1: Aprovisionar recursos

En esta sección se muestra cómo configurar los AWS recursos mediante una CloudFormation plantilla.

Para crear sus recursos utilizando CloudFormation plantilla

1. Inicie sesión en la AWS CloudFormation consola https://console.aws.amazon.com/cloudformationcomo administrador de IAM en la región EE.UU. Este (Virginia del Norte).
2. Seleccione Lanzar pila.
3. Seleccione Siguiente en la pantalla Crear pila.
4. Introduzca un Nombre de pila.
5. Elija Siguiente.

6. En la página siguiente, seleccione Siguiente.
7. Revisa los detalles en la última página y selecciona Reconozco que AWS CloudFormation podría crear recursos de IAM.
8. Seleccione Crear.

   La creación de la pila puede llevar hasta dos minutos.

El lanzamiento de la pila de formación de nube crea los recursos siguientes:

- lf-otf-datalake-123456789012: bucket de Amazon S3 para almacenar datos

> **Note**
>
> El identificador de cuenta adjunto al nombre del bucket de Amazon S3 se sustituye por el identificador de su cuenta.

- lf-otf-tutorial-123456789012: bucket de Amazon S3 para almacenar los resultados de las consultas y los scripts de trabajo AWS Glue
- lficebergdb: base de datos AWS Glue Iceberg
- lfhudidb — Base de datos AWS Glue Hudi
- lfdeltadb — Base de datos Delta AWS Glue
- native-iceberg-create: AWS Glue trabajo que crea una tabla de iceberg en el catálogo de datos
- native-hudi-create: trabajo que crea una tabla Hudi en el catálogo de datos AWS Glue
- native-delta-create: AWS Glue trabajo que crea una tabla Delta en el catálogo de datos
- LF-OTF-GlueServiceRole — Función de IAM que se transfiere para AWS Glue ejecutar las tareas. Este rol incluye las políticas necesarias para acceder a los recursos, como el Catálogo de datos, el bucket de Amazon S3, etc.
- LF-OTF-RegisterRole — Función de IAM para registrar la ubicación de Amazon S3 en Lake Formation. Este rol lleva `LF-Data-Lake-Storage-Policy` asociada.
- lf-consumer-analystuser: usuario de IAM al que consultar los datos usando Athena
- lf-consumer-analystuser-credentials: contraseña del usuario analista de datos almacenada en AWS Secrets Manager

Una vez finalizada la creación de la pila, vaya a la pestaña de resultados y anote los valores de:

- AthenaQueryResultLocation — Ubicación de Amazon S3 para el resultado de la consulta de Athena
- BusinessAnalystUserCredentials — Contraseña para el usuario analista de datos

  Para recuperar el valor de la contraseña:

  1. Elija el valor de `lf-consumer-analystuser-credentials` accediendo a la consola de Secrets Manager.
  2. En la sección Valor secreto, seleccione Recuperar valor secreto.
  3. Anote el valor secreto de la contraseña.

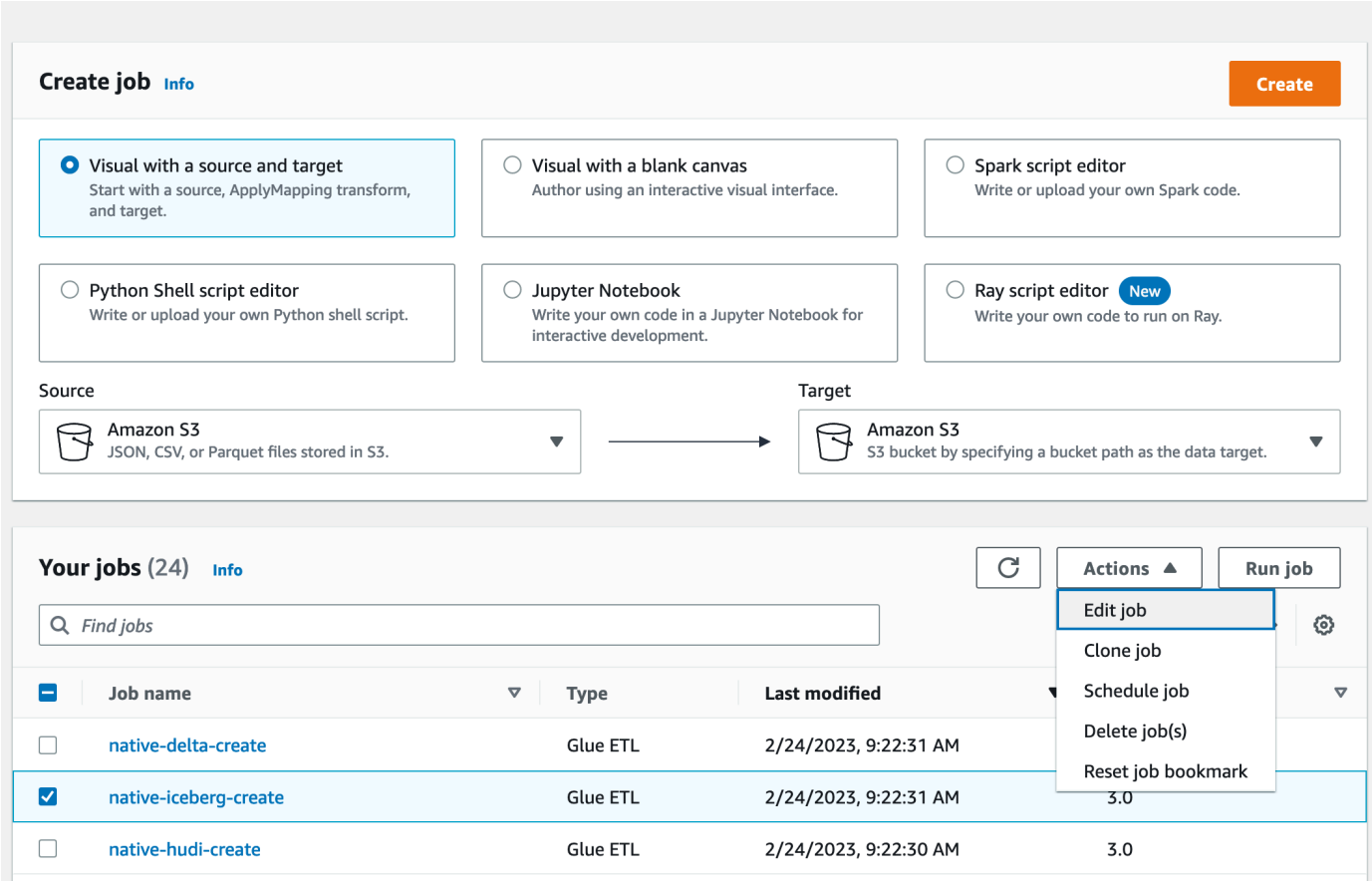
Paso 2: Configurar los permisos para una tabla de Iceberg

En esta sección, aprenderá a crear una tabla de iceberg en AWS Glue Data Catalog, configurar permisos de datos y consultar datos con Amazon Athena. AWS Lake Formation

Creación de una tabla de Iceberg

En este paso, ejecutará un AWS Glue trabajo que creará una tabla de transacciones de Iceberg en el catálogo de datos.

1. Abra la AWS Glue consola https://console.aws.amazon.com/glue/en la región EE.UU. Este (Norte de Virginia) como usuario administrador del lago de datos.
2. En el panel de navegación izquierdo, seleccione trabajos.
3. Seleccione `native-iceberg-create`.



4. En Acciones, seleccione Editar trabajo.
5. En Detalles del trabajo, expanda Propiedades avanzadas y active la casilla situada junto a Usar AWS Glue Data Catalog como metaalmacén de Hive para añadir los metadatos de la tabla en. AWS Glue Data Catalog Esto especifica AWS Glue Data Catalog el metaalmacén de los recursos del catálogo de datos utilizados en el trabajo y permite que los permisos de Lake Formation se apliquen más adelante a los recursos del catálogo.
6. Seleccione Save.
7. Elija Ejecutar. Puede ver el estado del trabajo mientras se ejecuta.

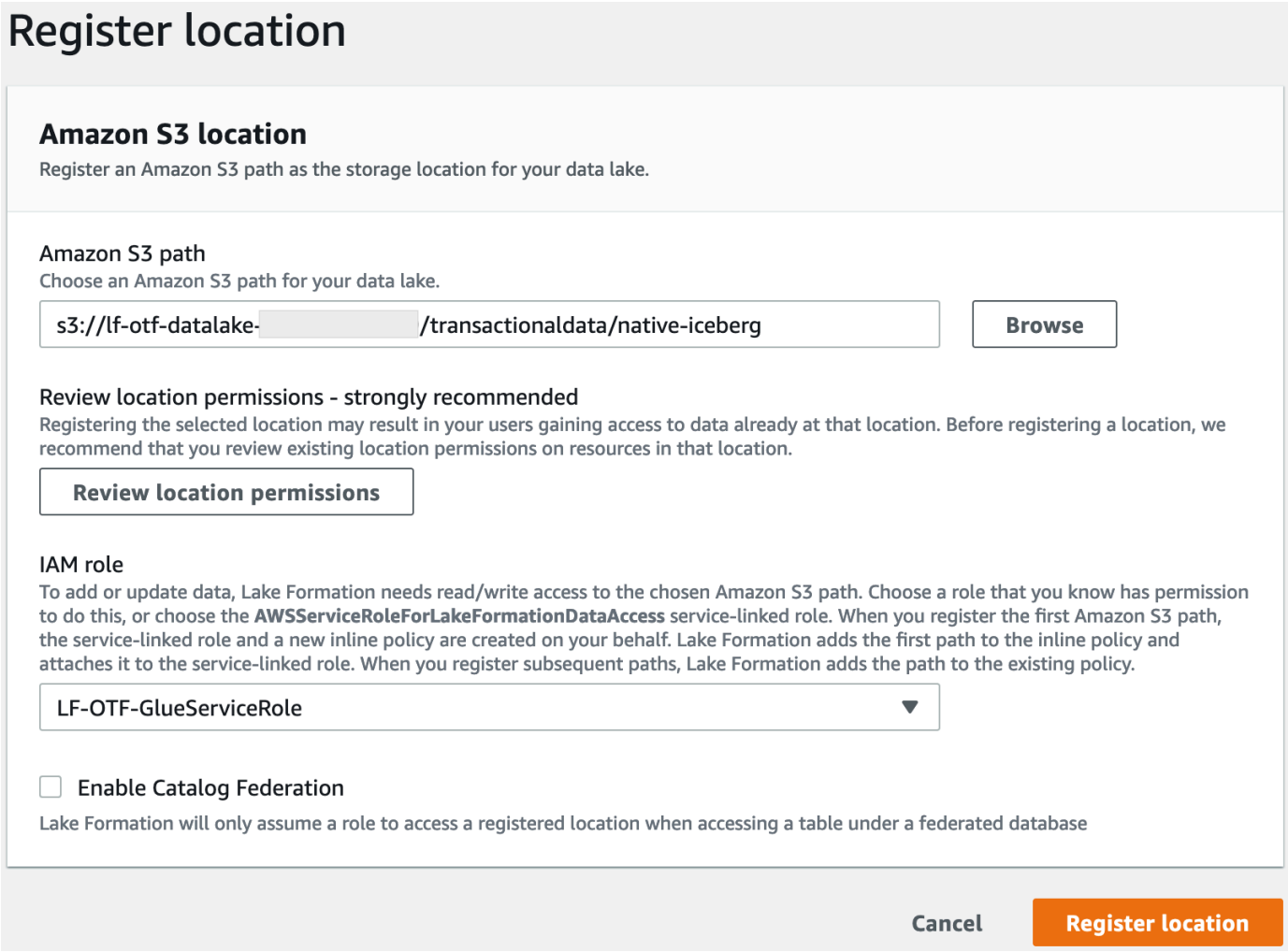
   Para obtener más información sobre los AWS Glue trabajos, consulte Trabajar con trabajos en la AWS Glue consola en la Guía para AWS Glue desarrolladores.

   Este trabajo crea una tabla de Iceberg de nombre `product` en la base de datos `lficebergdb`. Verifique la tabla de productos en la consola de Lake Formation.

Para registrar la ubicación de los datos con Lake Formation

A continuación, registre una ruta de Amazon S3 como ubicación de su lago de datos.

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/como usuario administrador del lago de datos.
2. En el panel de navegación, en Registrar e ingerir, seleccione Ubicación de los datos.
3. Seleccione Registrar ubicación.
4. En la página Registrar ubicación, introduzca:
   - Ruta de Amazon S3: seleccione Examinar y elija `lf-otf-datalake-123456789012`. Elija la flecha de expansión situada junto a la ubicación raíz de Amazon S3 para ir a la `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-iceberg` ubicación.
   - Rol de IAM: seleccione `LF-OTF-RegisterRole` como rol de IAM.
   - Seleccione Registrar ubicación.

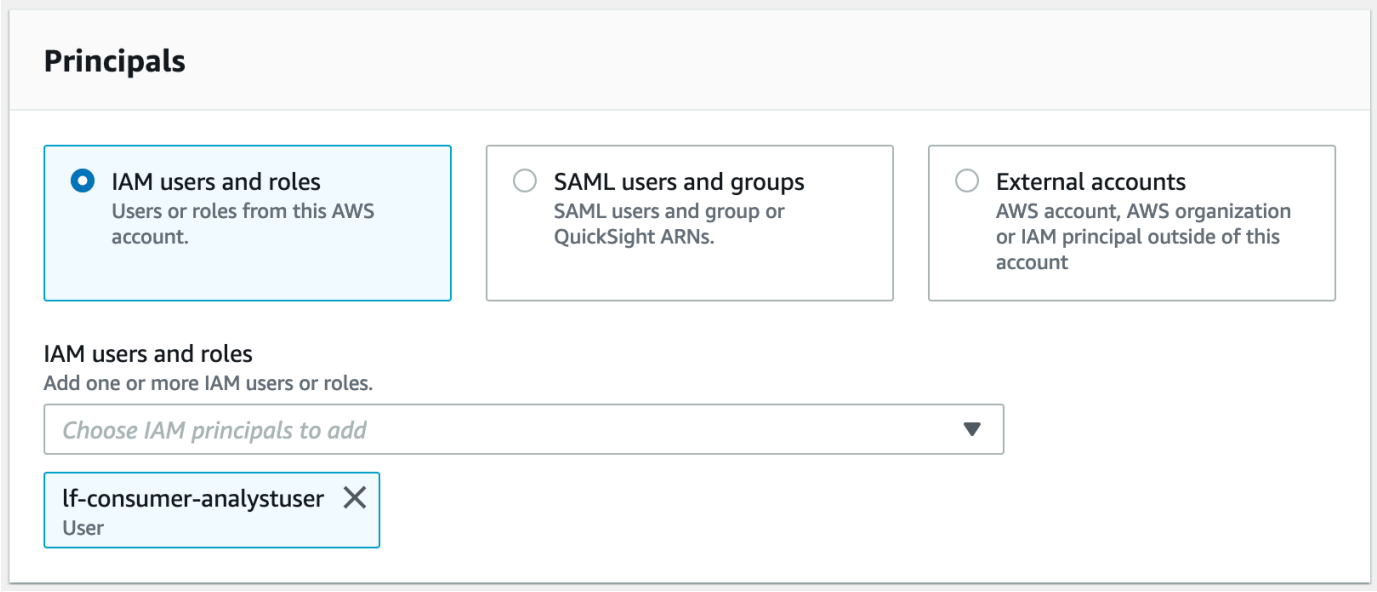


Para obtener más información sobre el registro de una ubicación de datos en Lake Formation, consulte Añadir una ubicación de Amazon S3 a su lago de datos.

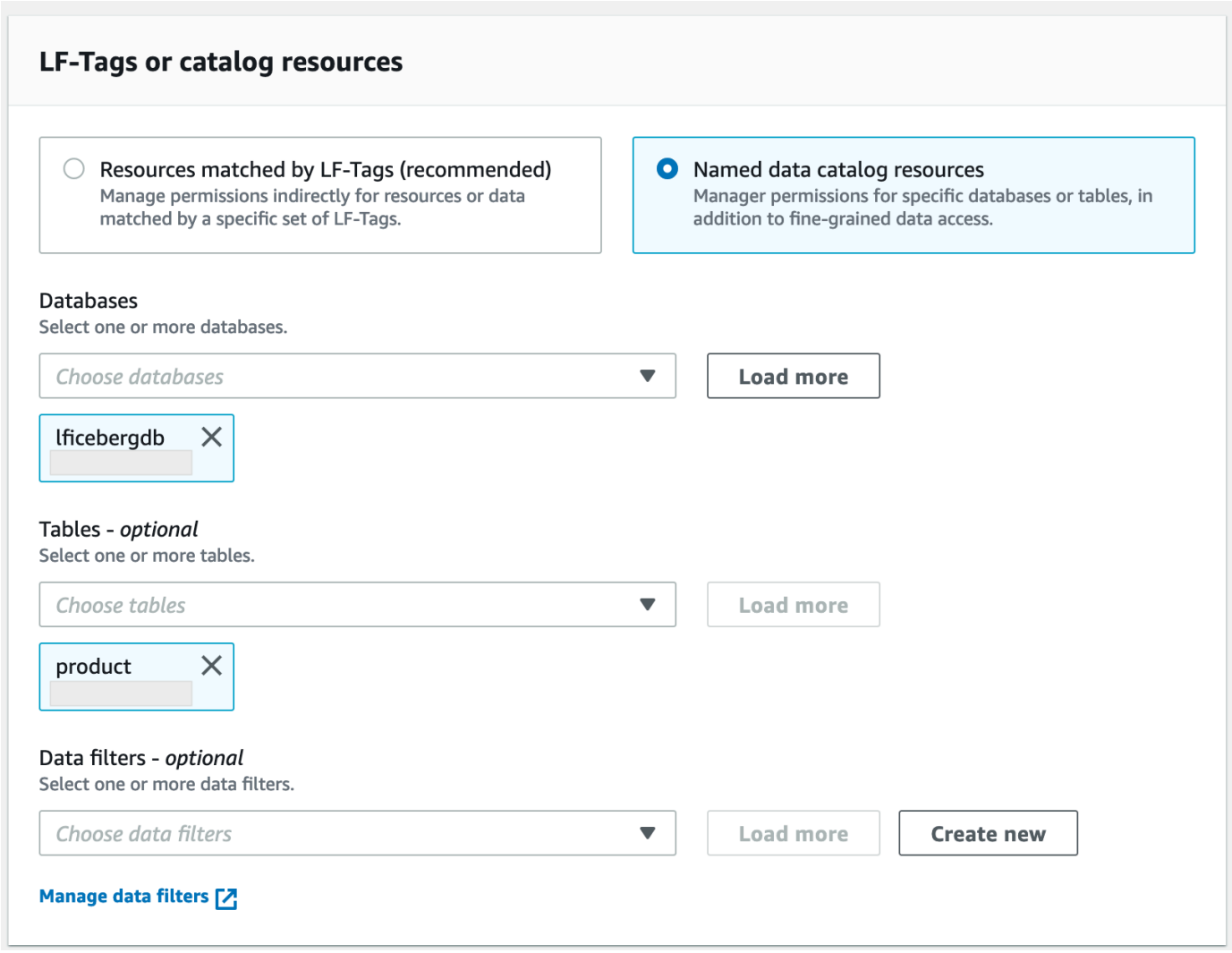
Para conceder permisos a Lake Formation en la tabla de Iceberg

En este paso, concederemos permisos de lago de datos al usuario analista de negocios.

1. En Permisos del lago de datos, seleccione Conceder.
2. En la pantalla Conceder permisos de datos, seleccione Usuarios y roles de IAM.
3. Seleccione `lf-consumer-analystuser` en la lista desplegable.



4. Seleccione un recurso de Catálogo de datos con nombre.
5. En Bases de datos, seleccione `lficebergdb`.
6. En Tablas, seleccione `product`.



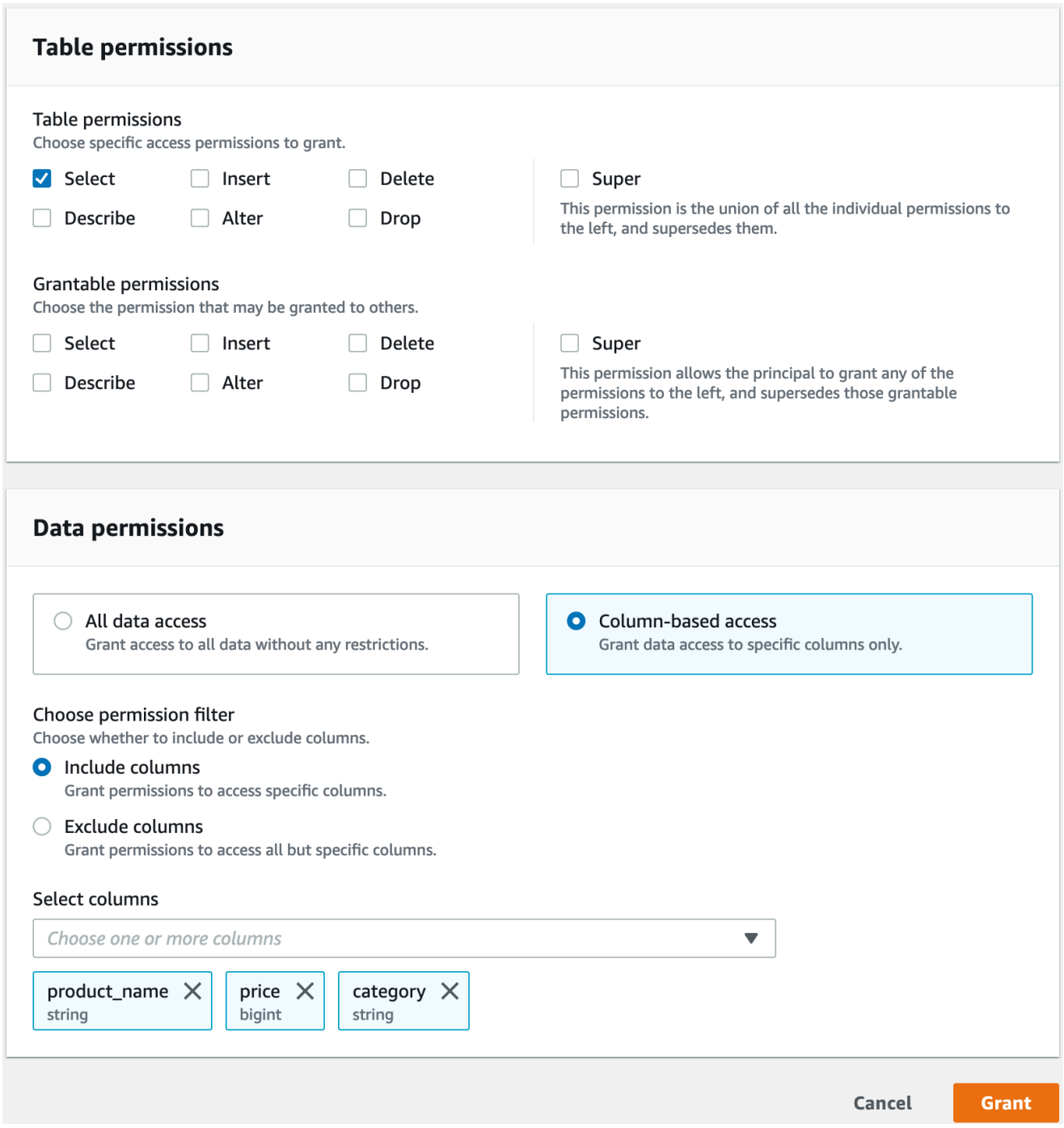
7. A continuación, puede conceder el acceso basado en columnas especificando las columnas.

   a. En Permisos de tabla, marque Seleccionar.

   b. En Permisos de datos, selecciona Column-based Acceso y selecciona Incluir columnas.

   c. Seleccione las columnas `product_name`, `price` y `category`.

   d. Elija Conceder.



Para consultar la tabla de Iceberg con Athena

Ahora puede empezar a consultar la tabla de Iceberg que creó con Athena. Si es la primera vez que ejecuta consultas en Athena, debe configurar una ubicación para los resultados de la consulta. Para obtener más información, consulte Especificación de una ubicación de resultados de consulta.

1. Cierre sesión como usuario administrador del lago de datos e inicie sesión como `lf-consumer-analystuser` en la región EE.UU. Este (Virginia del Norte) con la contraseña indicada anteriormente en el CloudFormation resultado.
2. Abra la consola Athena en https://console.aws.amazon.com/athena/.
3. Elija Configuración y seleccione Administrar.
4. En el cuadro Ubicación del resultado de la consulta, introduzca la ruta al depósito que creó en CloudFormation los resultados. Copie el valor de **`AthenaQueryResultLocation`** (s3://lf-otf-tutorial - 123456789012/athena -results/) y seleccione Guardar.
5. Ejecute la consulta siguiente para obtener una vista previa de los 10 registros almacenados en la tabla de Iceberg:

   ```
   select * from lficebergdb.product limit 10;
   ```

   Para obtener más información sobre la consulta de tablas de Iceberg con Athena, vaya a Consulta de tablas de Iceberg en la Guía del usuario de Amazon Athena.

Paso 3: Configurar los permisos para una tabla de Hudi

En esta sección, aprenderá a crear una tabla de Hudi en AWS Glue Data Catalog, a configurar los permisos de datos y a consultar datos con Amazon Athena. AWS Lake Formation

Para crear una tabla de Hudi

En este paso, ejecutará un AWS Glue trabajo para crear una tabla transaccional de Hudi en el catálogo de datos.

1. Inicie sesión en la AWS Glue consola https://console.aws.amazon.com/glue/en la región EE.UU. Este (Norte de Virginia)

   como usuario administrador del lago de datos.
2. En el panel de navegación izquierdo, seleccione trabajos.
3. Seleccione `native-hudi-create`.
4. En Acciones, seleccione Editar trabajo.
5. En Detalles del trabajo, expanda Propiedades avanzadas y active la casilla situada junto a Usar AWS Glue Data Catalog como metaalmacén de Hive para añadir los metadatos de la tabla en. AWS Glue Data Catalog Esto especifica AWS Glue Data Catalog el metaalmacén de

los recursos del catálogo de datos utilizados en el trabajo y permite que los permisos de Lake Formation se apliquen más adelante a los recursos del catálogo.

6. Seleccione Save.
7. Elija Ejecutar. Puede ver el estado del trabajo mientras se ejecuta.

   Para obtener más información sobre los AWS Glue trabajos, consulte Trabajar con trabajos en la AWS Glue consola en la Guía para AWS Glue desarrolladores.

   Este trabajo crea una tabla de Hudi (vaca) en la base de datos: lfhudidb. Verifique la tabla `product` en la consola de Lake Formation.

Para registrar la ubicación de los datos con Lake Formation

A continuación, registre una ruta de Amazon S3 como ubicación raíz de su lago de datos.

1. Inicie sesión en la consola de Lake Formation https://console.aws.amazon.com/lakeformation/como usuario administrador del lago de datos.
2. En el panel de navegación, en Registrar e ingerir, seleccione Ubicación de los datos.
3. Seleccione Registrar ubicación.
4. En la página Registrar ubicación, introduzca:

   - Ruta de Amazon S3: seleccione Examinar y elija `lf-otf-datalake-123456789012`. Elija la flecha de expansión situada junto a la ubicación raíz de Amazon S3 para ir a la `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-hudi` ubicación.
   - Rol de IAM: seleccione `LF-OTF-RegisterRole` como rol de IAM.
   - Seleccione Registrar ubicación.

Para conceder permisos de lago de datos en la tabla de Hudi

En este paso, concederemos permisos de lago de datos al usuario analista de negocios.

1. En Permisos del lago de datos, seleccione Conceder.
2. En la pantalla Conceder permisos de datos, seleccione Usuarios y roles de IAM.
3. `lf-consumer-analystuser` en el menú desplegable.
4. Seleccione un recurso de Catálogo de datos con nombre.

5. En Bases de datos, seleccione `lfhudidb`.
6. En Tablas, seleccione `product`.
7. A continuación, puede conceder el acceso basado en columnas especificando las columnas.
   a. En Permisos de tabla, marque Seleccionar.
   b. En Permisos de datos, selecciona Column-based Acceso y selecciona Incluir columnas.
   c. Seleccione las columnas `product_name`, `price` y `category`.
   d. Elija Conceder.

Para consultar la tabla de Hudi con Athena

Ahora puede empezar a consultar la tabla de Hudi que creó con Athena. Si es la primera vez que ejecuta consultas en Athena, debe configurar una ubicación para los resultados de la consulta. Para obtener más información, consulte Especificación de una ubicación de resultados de consulta.

1. Cierre sesión como usuario administrador del lago de datos e inicie sesión como `lf-consumer-analystuser` en la región EE.UU. Este (Virginia del Norte) con la contraseña indicada anteriormente en el CloudFormation resultado.
2. Abra la consola Athena en https://console.aws.amazon.com/athena/.
3. Elija Configuración y seleccione Administrar.
4. En el cuadro Ubicación del resultado de la consulta, introduzca la ruta al depósito que creó en CloudFormation los resultados. Copie el valor de **`AthenaQueryResultLocation`** (s3://lf-otf-tutorial - 123456789012/athena -results/) y guárdelo.
5. Ejecute la consulta siguiente para obtener una vista previa de los 10 registros almacenados en la tabla de Hudi:

```
select * from lfhudidb.product limit 10;
```

   Para obtener más información sobre la consulta de tablas de Hudi, vaya a Consulta de tablas de Hudi en la Guía del usuario de Amazon Athena.

Paso 4: Configurar los permisos para una tabla de Hudi

En esta sección, aprenderá a crear una tabla de Delta Lake con un archivo de manifiesto de enlaces simbólicos en el AWS Glue Data Catalog, a configurar los permisos de datos AWS Lake Formation y a consultarlos con Amazon Athena.

Para crear una tabla de Delta Lake.

En este paso, ejecutará un AWS Glue trabajo que creará una tabla transaccional de Delta Lake en el catálogo de datos.

1. Inicie sesión en la AWS Glue consola https://console.aws.amazon.com/glue/en la región EE.UU. Este (Norte de Virginia)

   como usuario administrador del lago de datos.
2. En el panel de navegación izquierdo, seleccione trabajos.
3. Seleccione `native-delta-create`.
4. En Acciones, seleccione Editar trabajo.
5. En Detalles del trabajo, expanda Propiedades avanzadas y active la casilla situada junto a Usar AWS Glue Data Catalog como metaalmacén de Hive para añadir los metadatos de la tabla en. AWS Glue Data Catalog Esto especifica AWS Glue Data Catalog el metaalmacén de los recursos del catálogo de datos utilizados en el trabajo y permite que los permisos de Lake Formation se apliquen más adelante a los recursos del catálogo.
6. Seleccione Save.
7. Seleccione Ejecutar en Acciones.

   Este trabajo crea una tabla de Delta Lake de nombre `product` en la base de datos `lfdeltadb`. Verifique la tabla `product` en la consola de Lake Formation.

Para registrar la ubicación de los datos con Lake Formation

A continuación, registre una ruta de Amazon S3 como ubicación raíz de su lago de datos.

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/el usuario administrador del lago de datos.
2. En el panel de navegación, en Registrar e ingerir, seleccione Ubicación de los datos.
3. Seleccione Registrar ubicación.

4. En la página Registrar ubicación, introduzca:

   - Ruta de Amazon S3: seleccione Examinar y elija `lf-otf-datalake-123456789012`. Elija la flecha de expansión situada junto a la ubicación raíz de Amazon S3 para ir a la `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-delta` ubicación.
   - Rol de IAM: seleccione `LF-OTF-RegisterRole` como rol de IAM.
   - Seleccione Registrar ubicación.

Para conceder permisos de lago de datos en la tabla de Delta Lake

En este paso, concederemos permisos de lago de datos al usuario analista de negocios.

1. En Permisos del lago de datos, seleccione Conceder.
2. En la pantalla Conceder permisos de datos, seleccione Usuarios y roles de IAM.
3. `lf-consumer-analystuser` en el menú desplegable.
4. Seleccione un recurso de Catálogo de datos con nombre.
5. En Bases de datos, seleccione `lfdeltadb`.
6. En Tablas, seleccione `product`.
7. A continuación, puede conceder el acceso basado en columnas especificando las columnas.

   a. En Permisos de tabla, marque Seleccionar.

   b. En Permisos de datos, selecciona Column-based Acceso y selecciona Incluir columnas.

   c. Seleccione las columnas `product_name`, `price` y `category`.

   d. Elija Conceder.

Para consultar la tabla de Delta Lake con Athena

Ahora puede empezar a consultar la tabla de Delta Lake que creó con Athena. Si es la primera vez que ejecuta consultas en Athena, debe configurar una ubicación para los resultados de la consulta. Para obtener más información, consulte Especificación de una ubicación de resultados de consulta.

1. Cierre sesión como usuario administrador del lago de datos e inicie sesión como `BusinessAnalystUser` en la región EE.UU. Este (Virginia del Norte) con la contraseña indicada anteriormente en el CloudFormation resultado.

2. Abra la consola Athena en https://console.aws.amazon.com/athena/.
3. Elija Configuración y seleccione Administrar.
4. En el cuadro Ubicación del resultado de la consulta, introduzca la ruta al depósito que creó en CloudFormation los resultados. Copie el valor de **`AthenaQueryResultLocation`** (s3://lf-otf-tutorial - 123456789012/athena -results/) y guárdelo.
5. Ejecute la consulta siguiente para obtener una vista previa de los 10 registros almacenados en la tabla de Delta Lake:

   ```
   select * from lfdeltadb.product limit 10;
   ```

   Para obtener más información sobre la consulta de tablas de Delta Lake, vaya a Consulta de tablas de Delta Lake en la Guía del usuario de Amazon Athena.

Paso 5: Eliminar AWS recursos

Cómo limpiar los recursos

Para evitar que se le cobren cargos no deseados Cuenta de AWS, elimine los AWS recursos que utilizó para este tutorial.

1. Inicie sesión en la CloudFormation consola https://console.aws.amazon.com/cloudformationcomo administrador de IAM.
2. Elimine la pila de formación de nube. Las tablas que ha creado se eliminan automáticamente con la pila.

Gestión de un lago de datos mediante el control de acceso basado en etiquetas de Lake Formation

Miles de clientes están creando lagos de datos a escala de petabytes. AWS Muchos de estos clientes los utilizan AWS Lake Formation para crear y compartir fácilmente sus lagos de datos en toda la organización. A medida que aumenta el número de tablas y usuarios, los responsables y administradores de datos buscan formas de gestionar fácilmente los permisos en los lagos de datos a escala. El control de acceso basado en etiquetas de Lake Formation (LF-TBAC) resuelve este problema al permitir que los administradores de datos creen etiquetas LF (en función de su clasificación y ontología de datos) que luego se pueden adjuntar a los recursos.

LF-TBAC es una estrategia de autorización que define permisos basados en atributos. En Lake Formation, estos atributos se denominan etiquetas LF. Puede adjuntar etiquetas LF a los recursos del Catálogo de datos y a las entidades principales de Lake Formation. Los administradores de lagos de datos pueden asignar y revocar permisos en los recursos de Lake Formation mediante etiquetas LF. Para obtener más información, consulte Control de acceso basado en etiquetas de Lake Formation.

Este tutorial muestra cómo crear una política de control de acceso basada en etiquetas de Lake Formation utilizando un conjunto de datos AWS público. Además, muestra cómo consultar tablas, bases de datos y columnas que tienen asociadas políticas de acceso basadas en etiquetas de Lake Formation.

Puede utilizar LF-TBAC para los siguientes casos de uso:

- Tiene un gran número de tablas y entidades principales a los que el administrador del lago de datos debe conceder acceso
- Desea clasificar sus datos en función de una ontología y conceder permisos en función de la clasificación
- El administrador del lago de datos desea asignar los permisos de forma dinámica, con acoplamiento flexible

A continuación, se indican los pasos generales para configurar permisos mediante LF-TBAC:

1. El administrador de datos define la ontología de las etiquetas con dos etiquetas LF: `Confidential` y `Sensitive`. Los datos `Confidential=True` tienen controles de acceso más estrictos. Los datos `Sensitive=True` requieren un análisis específico por parte del analista.
2. El administrador de datos asigna diferentes niveles de permisos al ingeniero de datos para crear tablas con diferentes etiquetas LF.
3. El ingeniero de datos crea dos bases de datos: `tag_database` y `col_tag_database`. Todas las tablas `tag_database` están configuradas con `Confidential=True`. Todas las tablas de `col_tag_database` están configuradas con `Confidential=False`. Algunas columnas de la tabla en `col_tag_database` tienen etiquetas `Sensitive=True` para necesidades de análisis específicas.
4. El ingeniero de datos concede permiso de lectura al analista para las tablas con una condición de expresión específica `Confidential=True` y `Confidential=False`, `Sensitive=True`.

5. Con esta configuración, el analista de datos puede centrarse en hacer el análisis con los datos correctos.

Temas

- Destinatarios previstos
- Requisitos previos
- Paso 1: Aprovisionar recursos
- Paso 2: Registrar la ubicación de sus datos, crear una ontología de etiquetas LF y conceder permisos
- Paso 3: Crear bases de datos de Lake Formation
- Paso 4: Conceder permisos de tabla
- Paso 5: Ejecutar una consulta en Amazon Athena para verificar los permisos
- Paso 6: Limpiar AWS los recursos

Destinatarios previstos

Este tutorial está dirigido a administradores de datos, ingenieros de datos y analistas de datos. Cuando se trata de gestionar AWS Glue Data Catalog y administrar los permisos en Lake Formation, los administradores de datos de las cuentas productoras tienen una propiedad funcional en función de las funciones que respaldan y pueden conceder acceso a varios consumidores, organizaciones externas y cuentas.

La tabla siguiente enumera los roles que se utilizan en este tutorial:

Rol	Description (Descripción)
Administrador de datos	El usuario `lf-data-steward` tiene el acceso siguiente: • Acceso de lectura a todos los recursos en el Catálogo de datos • Puede crear etiquetas LF y asociarlas al rol de ingeniero de datos para conceder permisos a otras entidades principales

Rol	Description (Descripción)
Ingeniero de datos	El usuario `lf-data-engineer` tiene el acceso siguiente: • Acceso completo de lectura, escritura y actualización a todos los recursos del Catálogo de datos • Permisos de localización de datos en el lago de datos • Puede asociar etiquetas LF y vincularlas al Catálogo de datos • Puede adjuntar etiquetas LF a los recursos, lo que proporciona acceso a las entidades principales de acuerdo con cualquier política creada por los administradores de datos
Analista de datos	El usuario `lf-data-analyst` tiene el siguiente acceso: • Acceso específico a los recursos compartidos por las políticas de acceso basadas en etiquetas de Lake Formation

Requisitos previos

Antes de empezar este tutorial, debe disponer de una Cuenta de AWS que pueda utilizar para iniciar sesión como usuario administrativo con los permisos correctos. Para obtener más información, consulte Completa la inicial AWS tareas de configuración.

En este tutorial, se supone que está familiarizado con IAM. Para obtener más información acerca de IAM, consulte la Guía del usuario de IAM.

Paso 1: Aprovisionar recursos

Este tutorial incluye una AWS CloudFormation plantilla para una configuración rápida. Puede revisarla y personalizarla para adaptarla a sus necesidades. La plantilla crea tres funciones

diferentes (enumeradas enDestinatarios previstos) para realizar este ejercicio y copia el nyc-taxi-data conjunto de datos en su bucket local de Amazon S3.

- Un bucket de Amazon S3
- Los escenarios apropiados de Lake Formation
- Los recursos adecuados de Amazon EC2
- Tres roles de IAM con credenciales

Crear recursos

1. Inicie sesión en la AWS CloudFormation consola en https://console.aws.amazon.com/cloudformation en la región EE.UU. Este (Norte de Virginia).
2. Seleccione Lanzar pila.
3. Elija Siguiente.
4. En la sección Configuración de usuario, introduzca la contraseña para tres roles: `DataStewardUserPassword`, `DataEngineerUserPassword` y `DataAnalystUserPassword`.
5. Revisa los detalles en la última página y selecciona Acepto que AWS CloudFormation podría crear recursos de IAM.
6. Seleccione Crear.

   La creación de la pila puede tardar hasta cinco minutos.

> **Note**
>
> Después de completar el tutorial, es posible que desee eliminar la pila CloudFormation para evitar seguir incurriendo en cargos. Compruebe que los recursos se hayan eliminado correctamente en el estado de evento de la pila.

Paso 2: Registrar la ubicación de sus datos, crear una ontología de etiquetas LF y conceder permisos

En este paso, el usuario administrador de datos define la ontología de etiquetas con dos etiquetas LF `Confidential` y `Sensitive`; además, ofrece a las entidades principales de IAM específicas adjuntar etiquetas LF recién creadas a los recursos.

Registre la ubicación de los datos y cree una ontología de etiquetas LF.

1. Siga el primer paso como usuario administrador de datos (`lf-data-steward`) para verificar los datos en Amazon S3 y el Catálogo de datos en Lake Formation.

   a. Inicie sesión en la consola de Lake Formation `lf-data-steward` con la contraseña utilizada al implementar la CloudFormation pila. https://console.aws.amazon.com/lakeformation/

   b. En el panel de navegación, en Permisos, elija Roles y tareas administrativas.

   c. En la sección Administradores del lago de datos, elija Agregar.

   d. En la página Agregar administrador, en Usuarios y roles de IAM, elija el usuario `lf-data-steward`.

   e. Seleccione Guardar para añadir `lf-data-steward` como administrador de Lake Formation.

2. A continuación, actualice la configuración del Catálogo de datos para usar el permiso de Lake Formation para controlar los recursos del catálogo en lugar del control de acceso basado en IAM.

   a. En el panel de navegación, Administración, seleccione Configuración del Catálogo de datos.

   b. Desmarque Usar solo el control de acceso de IAM para las nuevas bases de datos.

   c. Desmarque Usar solo el control de acceso de IAM para las nuevas bases de datos.

   d. Haga clic en Guardar.

3. A continuación, registre la ubicación de los datos para el lago de datos.

   a. En el panel de navegación, bajo Administración, seleccione Ubicaciones de los lagos de datos.

   b. Seleccione Registrar ubicación.

   c. En la página Registrar ubicación, en la Ruta de Amazon S3, introduzca `s3://lf-tagbased-demo-`*`Account-ID`*.

d. En rol de IAM, deje el valor predeterminado `AWSServiceRoleForLakeFormationDataAccess` como está.

e. Elija Lake Formation como modo de permiso.

f. Elija Registrar ubicación.

4. A continuación, cree la ontología definiendo una etiqueta LF.

a. En el panel de navegación, en Permisos, elija Etiquetas LF y permisos.

b. Seleccione Agregar etiqueta LF.

c. En Clave, escriba `Confidential`.

d. En Valores, añada `True` y `False`.

e. Elija Añadir etiquetas LF.

f. Repita los pasos para crear la Etiqueta LF `Sensitive` con el valor `True`.

Ha creado todas las etiquetas L necesarias para este ejercicio.

Conceder permisos a los usuarios de IAM

1. A continuación, otorgue a entidades principales específicas de IAM la capacidad de adjuntar etiquetas LF recién creadas a los recursos.

a. En el panel de navegación, en Permisos, elija Etiquetas LF y permisos.

b. En la sección Permisos de etiquetas LF, elija Conceder permisos.

c. En Tipo de permiso, elija Permisos de par clave-valor de etiquetas LF.

d. Seleccione Roles y usuarios de IAM.

e. En Roles y usuarios de IAM, busque y elija el rol `lf-data-engineer`.

f. En la sección Etiquetas LF, agregue la clave `Confidential` con los valores `True` y `False`, y la `key` `Sensitive` con el valor `True`.

g. En Permisos, seleccione Describir y Asociar para Permisos y Permisos concedibles.

h. Elija Conceder.

2. A continuación, conceda permisos `lf-data-engineer` para crear bases de datos en nuestro catálogo de datos y en el bucket de Amazon S3 subyacente creado por AWS CloudFormation.

a. En el panel de navegación, en Administración, seleccione Roles y tareas administrativas.

b. En la sección Creadores de bases de datos, elija Conceder.

c. En Usuarios de AIM y roles, elija el rol `lf-data-engineer`.

d. En Permisos del catálogo, seleccione Crear base de datos.

e. Elija Conceder.

3. A continuación, conceda permisos en el bucket (`s3://lf-tagbased-demo-`*`Account-ID`*) de Amazon S3 al usuario `lf-data-engineer`.

a. En el panel de navegación, bajo Permisos, seleccione Ubicaciones de datos.

b. Elija Conceder.

c. Seleccione Mi cuenta.

d. En Usuarios de AIM y roles, elija el rol `lf-data-engineer`.

e. Para las ubicaciones de almacenamiento, introduzca el depósito de Amazon S3 creado por la CloudFormation plantilla(`s3://lf-tagbased-demo-`*`Account-ID`*).

f. Elija Conceder.

4. A continuación, conceda permisos concedibles `lf-data-engineer` sobre los recursos asociados a la expresión `Confidential=True` de la Etiqueta LF.

a. En el panel de navegación, en Permisos, seleccione Permisos de lago de datos.

b. Elija Conceder.

c. Seleccione Roles y usuarios de IAM.

d. Elija el rol `lf-data-engineer`.

e. En la sección Recursos del catálogo o de la etiqueta LF, seleccione Recursos que coincidan con las etiquetas LF.

f. Seleccione Agregar par clave-valor de etiquetas LF.

g. Añada la clave `Confidential` con los valores `True`.

h. En la sección Permisos de base de datos, seleccione Describir en Permisos de bases de datos y Permisos concedibles.

i. En la sección Permisos de tabla, seleccione Describir, Seleccionar y Modificar, tanto para los Permisos de tabla como para los Permisos concedibles.

j. Elija Conceder.

5. A continuación, otorgue permisos concedibles `lf-data-engineer` sobre los recursos asociados a la expresión `Confidential=False` de la etiqueta LF.

a. En el panel de navegación, en Permisos, seleccione Permisos de lago de datos.

b. Elija Conceder.

c. Seleccione Roles y usuarios de IAM.

d. Elija el rol `lf-data-engineer`.

e. Seleccione Recursos que coincidan con las etiquetas LF.

f. Elija Añadir etiquetas LF.

g. Añada la clave `Confidential` con los valores `False`.

h. En la sección Permisos de base de datos, seleccione Describir en Permisos de bases de datos y Permisos concedibles.

i. En la sección Permisos de tabla y columna, no selecciones nada.

j. Elija Conceder.

6. A continuación, conceda permisos concedibles `lf-data-engineer` sobre los recursos asociados a los pares de clave-valor `Confidential=False` y `Sensitive=True` de la Etiqueta LF.

a. En el panel de navegación, en Permisos, seleccione Permisos de datos.

b. Elija Conceder.

c. Seleccione Roles y usuarios de IAM.

d. Elija el rol `lf-data-engineer`.

e. En la sección Etiquetas LF o recursos del catálogo, seleccione Recursos que coincidan con las etiquetas LF.

f. Seleccione Agregar etiqueta LF.

g. Añada la clave `Confidential` con los valores `False`.

h. Seleccione Agregar par clave-valor de etiquetas LF.

i. Añada la clave `Sensitive` con los valores `True`.

j. En la sección Permisos de base de datos, seleccione Describir en Permisos de bases de datos y Permisos concedibles.

k. En la sección Permisos de tabla, seleccione Describir, Seleccionar y Modificar, tanto para los Permisos de tabla como para los Permisos concedibles.

l. Elija Conceder.

Paso 3: Crear bases de datos de Lake Formation

En este paso, se crean dos bases de datos y se adjuntan etiquetas LF a las bases de datos y columnas específicas con fines de prueba.

Cree sus bases de datos y su tabla para el acceso a nivel de base de datos

1. En primer lugar, cree la base de datos `tag_database`, la tabla `source_data` y adjunte las etiquetas L correspondientes.

   a. En la consola de Lake Formation (https://console.aws.amazon.com/lakeformation/), en Catálogo de datos, elija Bases de datos.

   b. Elija Creación de base de datos.

   c. En Nombre, ingrese `tag_database`.

   d. En Ubicación, introduzca la ubicación de Amazon S3 creada por la CloudFormation plantilla(`s3://lf-tagbased-demo-`*`Account-ID`*`/tag_database/`).

   e. Desmarque Usar solo el control de acceso de IAM para las nuevas tablas de esta base de datos.

   f. Elija Creación de base de datos.

2. A continuación, cree una nueva tabla dentro de `tag_database`.

   a. En la página Bases de datos, seleccione la base de datos `tag_database`.

   b. Seleccione Ver tablas y haga clic en Crear tabla.

   c. En Nombre, escriba `source_data`.

   d. En Base de datos elija la base de datos `tag_database`.

   e. Para el formato de tabla, elija AWS Glue Tabla estándar.

   f. En Los datos se encuentran en, elija Ruta especificada en otra cuenta.

   g. En Incluir ruta, introduzca la ruta que `tag_database` ha creado la CloudFormation plantilla(`s3://lf-tagbased-demo`*`Account-ID`*`/tag_database/`).

   h. En Formato de datos, seleccione CSV.

   i. En Esquema de carga, introduzca la matriz JSON de estructura de columnas siguiente para crear un esquema:

```
[
              {
                   "Name": "vendorid",
```

```
                    "Type": "string"
                },
                {
                    "Name": "lpep_pickup_datetime",
                    "Type": "string"
                },
                {
                    "Name": "lpep_dropoff_datetime",
                    "Type": "string"

                },
                   {
                    "Name": "store_and_fwd_flag",
                    "Type": "string"
                },
                   {
                    "Name": "ratecodeid",
                    "Type": "string"

                },
                   {
                    "Name": "pulocationid",
                    "Type": "string"

                },
                {
                    "Name": "dolocationid",
                    "Type": "string"

                },
                   {
                    "Name": "passenger_count",
                    "Type": "string"

                },
                {
                    "Name": "trip_distance",
                    "Type": "string"

                },
                   {
                    "Name": "fare_amount",
                    "Type": "string"
```

```
                },
                {
                     "Name": "extra",
                     "Type": "string"

                },
                   {
                     "Name": "mta_tax",
                     "Type": "string"

                },
                {
                     "Name": "tip_amount",
                     "Type": "string"

                },
                   {
                     "Name": "tolls_amount",
                     "Type": "string"

                },
                {
                     "Name": "ehail_fee",
                     "Type": "string"

                },
                {
                     "Name": "improvement_surcharge",
                     "Type": "string"

                },
                {
                     "Name": "total_amount",
                     "Type": "string"

                },
                {
                     "Name": "payment_type",
                     "Type": "string"

                }
]
```

j. Seleccione Cargar. Tras cargar el esquema, el esquema de la tabla será similar a la siguiente captura de pantalla:

#	Column Name	Data type
1	vendorid	string
2	lpep_pickup_datetime	string
3	lpep_dropoff_datetime	string
4	store_and_fwd_flag	string
5	ratecodeid	string
6	pulocationid	string
7	dolocationid	string
8	passenger_count	string
9	trip_distance	string
10	fare_amount	string
11	extra	string
12	mta_tax	string
13	tip_amount	string
14	tolls_amount	string
15	ehail_fee	string
16	improvement_surcharge	string
17	total_amount	string
18	payment_type	string

k. Elija Enviar.

3. A continuación, adjunte etiquetas LF de base de datos.

   a. En la página Bases de datos, busque y seleccione `tag_database`.

   b. En el menú Acciones, elija Editar etiquetas LF.

   c. Seleccione Asignar una nueva etiqueta LF.

   d. En Claves asignadas, elija la etiqueta LF `Confidential` que creó anteriormente.

   e. En Valores, elija `True`.

   f. Seleccione Save.

   Esto completa la asignación de la etiqueta LF a la base de datos tag_database.

Cree su base de datos y su tabla para el acceso a nivel de base de datos

Repita los pasos siguientes para crear la base de datos `col_tag_database` y la tabla `source_data_col_lvl`, y adjunte las etiquetas LF columna.

1. En la página Bases de datos, seleccione Crear base de datos.
2. En Nombre, escriba `col_tag_database`.
3. En Ubicación, introduzca la ubicación de Amazon S3 creada por la CloudFormation plantilla(`s3://lf-tagbased-demo-`*`Account-ID`*`/col_tag_database/`).
4. Desmarque Usar solo el control de acceso de IAM para las nuevas tablas de esta base de datos.
5. Elija Creación de base de datos.
6. En la página Bases de datos, seleccione la nueva base de datos (`col_tag_database`).
7. Seleccione Ver tablas y haga clic en Crear tabla.
8. En Nombre, escriba `source_data_col_lvl`.
9. En Base de datos, elija su nueva base de datos (`col_tag_database`).
10. Para el formato de tabla, elija AWS Glue Tabla estándar.
11. En Los datos se encuentran en, elija Ruta especificada en otra cuenta.
12. Introduzca la ruta de Amazon S3 para `col_tag_database` (`s3://lf-tagbased-demo-`*`Account-ID`*`/col_tag_database/`).
13. En Formato de datos, seleccione `CSV`.
14. En `Upload schema`, introduzca el siguiente esquema JSON:

```
[
                {
                     "Name": "vendorid",
                     "Type": "string"


                },
                {
                     "Name": "lpep_pickup_datetime",
                     "Type": "string"


                },
                {
                     "Name": "lpep_dropoff_datetime",
                     "Type": "string"


                },
                   {
                     "Name": "store_and_fwd_flag",
                     "Type": "string"


                },
                   {
                     "Name": "ratecodeid",
                     "Type": "string"


                },
                   {
                     "Name": "pulocationid",
                     "Type": "string"


                },
                {
                     "Name": "dolocationid",
                     "Type": "string"


                },
```

```
                {
                  "Name": "passenger_count",
                  "Type": "string"


              },
              {
                  "Name": "trip_distance",
                  "Type": "string"


              },
                {
                  "Name": "fare_amount",
                  "Type": "string"


              },
              {
                  "Name": "extra",
                  "Type": "string"


              },
                {
                  "Name": "mta_tax",
                  "Type": "string"


              },
              {
                  "Name": "tip_amount",
                  "Type": "string"


              },
                {
                  "Name": "tolls_amount",
                  "Type": "string"


              },
              {
                  "Name": "ehail_fee",
```

```
                    "Type": "string"


               },
               {
                    "Name": "improvement_surcharge",
                    "Type": "string"


               },
               {
                    "Name": "total_amount",
                    "Type": "string"


               },
               {
                    "Name": "payment_type",
                    "Type": "string"


               }
]
```

15. Elija `Upload`. Tras cargar el esquema, el esquema de la tabla será similar a la siguiente captura de pantalla:

#	Column Name	Data type
1	vendorid	string
2	lpep_pickup_datetime	string
3	lpep_dropoff_datetime	string
4	store_and_fwd_flag	string
5	ratecodeid	string
6	pulocationid	string
7	dolocationid	string
8	passenger_count	string
9	trip_distance	string
10	fare_amount	string
11	extra	string
12	mta_tax	string
13	tip_amount	string
14	tolls_amount	string
15	ehail_fee	string
16	improvement_surcharge	string
17	total_amount	string
18	payment_type	string

16. Seleccione Enviar para completar la creación de la tabla.
17. Ahora, asocie la etiqueta LF `Sensitive=True` a las columnas `vendorid` y `fare_amount`.
    a. En la página Tablas, seleccione la tabla que ha creado (`source_data_col_lvl`).
    b. En el menú Acciones, elija Esquema.
    c. Seleccione la columna `vendorid` y elija Editar etiquetas LF.
    d. En Teclas asignadas, seleccione Sensible.
    e. En Valores, elija Verdadero.
    f. Seleccione Save.
18. A continuación, asocie la etiqueta LF `Confidential=False` a `col_tag_database`. Esto es necesario para poder describir la base `lf-data-analyst` de datos cuando se inicia sesión desde. `col_tag_database` Amazon Athena
    a. En la página Bases de datos, busque y seleccione `col_tag_database`.
    b. En el menú Acciones, elija Editar etiquetas LF.
    c. Seleccione Asignar una nueva etiqueta LF.
    d. En Claves asignadas, elija la etiqueta LF `Confidential` que creó anteriormente.
    e. En Valores, elija `False`.
    f. Seleccione Save.

Paso 4: Conceder permisos de tabla

Conceda permisos a los analistas de datos para el uso de las bases de datos `tag_database` y la tabla `col_tag_database` utilizando etiquetas LF `Confidential` y `Sensitive`.

1. Siga estos pasos para conceder permisos al usuario `lf-data-analyst` sobre los objetos asociados a la etiqueta LF `Confidential=True` (base de datos: tag_database) para obtener el permiso `Describe` sobre la base de datos y el permiso `Select` sobre las tablas.
    a. Inicie sesión en la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/as`lf-data-engineer`.
    b. En Permisos, seleccione Permisos de lago de datos.
    c. Elija Conceder.
    d. En Entidades principales, seleccione Roles y usuarios de IAM.

e. En Roles y usuarios de IAM, elija `lf-data-analyst`.

f. En la sección Etiquetas LF o recursos del catálogo, seleccione Recursos que coincidan con las etiquetas LF.

g. Elija Añadir etiquetas LF.

h. En Clave, elija `Confidential`.

i. En Valores, elija `True`.

j. En Permisos de base de datos, seleccione `Describe`.

k. En Permisos de tabla, elija Seleccionar y Describir.

l. Elija Conceder.

2. A continuación, repita los pasos para conceder permisos a los analistas de datos para la expresión `Confidential=False` de las etiquetas LF. Esta Etiqueta LF se utiliza para describir la tabla `col_tag_database` y la tabla `source_data_col_lvl` cuando se inicia sesión como `lf-data-analyst` desde Amazon Athena.

   a. Inicie sesión en la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/as`lf-data-engineer`.

   b. En la página Bases de datos, seleccione la base de datos `col_tag_database`.

   c. Elija Acciones y Conceder.

   d. En Entidades principales, seleccione Roles y usuarios de IAM.

   e. En Roles y usuarios de IAM, elija `lf-data-analyst`.

   f. Seleccione Recursos que coincidan con las etiquetas LF.

   g. Seleccione Agregar etiqueta LF.

   h. En Clave, elija `Confidential`.

   i. En Valores, elija `False`.

   j. En Permisos de base de datos, seleccione `Describe`.

   k. En Permisos de tabla, no seleccione nada.

   l. Elija Conceder.

3. A continuación, repita los pasos para conceder permisos a los analistas de datos para la expresión de etiquetas LF para `Confidential=False` y `Sensitive=True`. Esta Etiqueta LF se utiliza para describir la tabla `col_tag_database` y la tabla `source_data_col_lvl` (nivel de columna) cuando se inicia sesión como `lf-data-analyst` desde Amazon Athena.

a. Inicie sesión en la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/as`lf-data-engineer`.
b. En la página Bases de datos, seleccione la base de datos `col_tag_database`.
c. Elija Acciones y Conceder.
d. En Entidades principales, seleccione Roles y usuarios de IAM.
e. En Roles y usuarios de IAM, elija `lf-data-analyst`.
f. Seleccione Recursos que coincidan con las etiquetas LF.
g. Elija Añadir etiquetas LF.
h. En Clave, elija `Confidential`.
i. En Valores, elija `False`.
j. Elija Añadir etiquetas LF.
k. En Clave, elija `Sensitive`.
l. En Valores, elija `True`.
m. En Permisos de base de datos, seleccione `Describe`.
n. En Permisos de tabla, seleccione `Select` y `Describe`.
o. Elija Conceder.

Paso 5: Ejecutar una consulta en Amazon Athena para verificar los permisos

En este paso, utilice Amazon Athena para ejecutar consultas `SELECT` en las dos tablas (`source_data and source_data_col_lvl`). Utilice la ruta de Amazon S3 como ubicación de los resultados de la consulta (`s3://lf-tagbased-demo-`*`Account-ID`*`/athena-results/`).

1. Inicie sesión en la consola de Athena en https://console.aws.amazon.com/athena/as. `lf-data-analyst`
2. En el editor de consultas de Athena, selecciona `tag_database` en el panel izquierdo.
3. Seleccione el icono de opciones de menú adicionales (tres puntos verticales) situado junto a `source_data` y elija Vista previa de la tabla.
4. Elija Ejecutar consulta.

La consulta tardará unos minutos en ejecutarse. La consulta muestra todas las columnas de la salida porque la etiqueta LF está asociada a nivel de base de datos y la tabla `source_data` heredó automáticamente la `LF-tag` de la base de datos `tag_database`.

5. Ejecute otra consulta con `col_tag_database` y `source_data_col_lvl`.

   La segunda consulta devuelve las dos columnas que estaban etiquetadas como `Non-Confidential` y `Sensitive`.

6. También puede comprobar el comportamiento de la política de acceso basada en etiquetas de Lake Formation en las columnas para las que no tiene concesiones de política. Cuando se selecciona una columna sin etiquetar de la tabla `source_data_col_lvl`, Athena devuelve un error. Por ejemplo, puede ejecutar la siguiente consulta para elegir columnas `geolocationid` sin etiquetar:

   ```
   SELECT geolocationid FROM "col_tag_database"."source_data_col_lvl" limit 10;
   ```

Paso 6: Limpiar AWS los recursos

Para evitar que se le cobren cargos no deseados Cuenta de AWS, puede eliminar los AWS recursos que utilizó para este tutorial.

1. Inicie sesión en la consola de Lake Formation como `lf-data-engineer` y borre las bases de datos `tag_database` y `col_tag_database`.
2. Luego, inicie sesión como `lf-data-steward` y borre todos los Permisos de etiqueta L, los Permisos de datos y los Permisos de ubicación de datos concedidos antes `lf-data-engineer` y `lf-data-analyst`..
3. Inicie sesión en la consola de Amazon S3 como propietario de la cuenta con las credenciales de IAM que utilizó para implementar la CloudFormation pila.
4. Elimine los buckets siguientes:

   - lf-tagbased-demo-accesslogs-*acct-id*
   - lf-tagbased-demo-*acct-id*

5. Inicie sesión en CloudFormation la consola en https://console.aws.amazon.com/cloudformation y elimine la pila que creó. Espera a que el estado de la pila cambie a. `DELETE_COMPLETE`

Protección de los lagos de datos con control de acceso a nivel de fila

AWS Lake Formation Los permisos a nivel de fila le permiten proporcionar acceso a filas específicas de una tabla en función de las políticas de control y cumplimiento de los datos. Si tiene tablas grandes que almacenan miles de millones de registros, necesita una forma de permitir que los diferentes usuarios y equipos accedan solo a los datos que pueden ver. El control de acceso a nivel de fila es una forma sencilla y eficaz de proteger los datos, a la vez que permite a los usuarios acceder a los datos que necesitan para desarrollar su trabajo. Lake Formation proporciona informes centralizados de auditoría y cumplimiento al identificar qué entidades principales accedieron a qué datos, cuándo lo hicieron y a través de qué servicios.

En este tutorial aprenderá cómo funcionan los controles de acceso a nivel de fila en Lake Formation y cómo configurarlos.

Este tutorial incluye una AWS CloudFormation plantilla para configurar rápidamente los recursos necesarios. Puede revisarla y personalizarla para adaptarla a sus necesidades.

Temas

- Destinatarios previstos
- Requisitos previos
- Paso 1: Aprovisionar recursos
- Paso 2: Consultar sin filtros de datos
- Paso 3: Configurar los filtros de datos y conceder permisos
- Paso 4: Consultar con filtros de datos
- Paso 5: Limpiar AWS los recursos

Destinatarios previstos

Este tutorial está dirigido a administradores de datos, ingenieros de datos y analistas de datos. En la siguiente tabla se enumeran los roles y responsabilidades de un propietario y un consumidor de datos.

Rol	Description (Descripción)
Administrador de IAM	Usuario que puede crear usuarios, roles y buckets de Amazon Simple Storage Service (Amazon S3). Tiene la política `AdministratorAccess` AWS gestionada.
Administrador de lagos de datos	Usuario responsable de configurar el lago de datos, crear filtros de datos y conceder permisos a los analistas de datos.
Analista de datos	Usuario que puede ejecutar consultas en el lago de datos. Los analistas de datos que residen en diferentes países (en nuestro caso, EE. UU. y Japón) solo pueden analizar las reseñas de productos de clientes ubicados en su propio país y, por motivos de cumplimiento, no deberían poder ver los datos de clientes ubicados en otros países.

Requisitos previos

Antes de empezar este tutorial, debe disponer de una Cuenta de AWS que pueda utilizar para iniciar sesión como usuario administrativo con los permisos correctos. Para obtener más información, consulte Completa la inicial AWS tareas de configuración.

En este tutorial, se supone que está familiarizado con IAM. Para obtener más información acerca de IAM, consulte la Guía del usuario de IAM.

Cambiar la configuración de Lake Formation

> **Important**
>
> Antes de lanzar la CloudFormation plantilla, desactive la opción Usar solo el control de acceso de IAM para nuevas bases de datos o tablas en Lake Formation siguiendo los pasos que se indican a continuación:

1. Inicie sesión en la consola de Lake Formation https://console.aws.amazon.com/lakeformation/en la región EE.UU. Este (Norte de Virginia) o en la región EE.UU. Oeste (Oregón).
2. En Catálogo de datos, seleccione Configuración.
3. Deseleccione Usar solo el control de acceso de IAM para nuevas bases de datos y Usar solo el control de acceso de IAM para nuevas tablas en las nuevas bases de datos.
4. Seleccione Guardar.

Paso 1: Aprovisionar recursos

Este tutorial incluye una CloudFormation plantilla para una configuración rápida. Puede revisarla y personalizarla para adaptarla a sus necesidades. La CloudFormation plantilla genera los siguientes recursos:

- Usuarios y políticas para:
  - DataLakeAdmin
  - DataAnalystNOSOTROS
  - DataAnalystJP
- Configuración y permisos del lago de datos de Lake Formation
- Una función de Lambda (para recursos CloudFormation personalizados respaldados por Lambda) que se utiliza para copiar archivos de datos de muestra del bucket público de Amazon S3 al bucket de Amazon S3
- Un bucket de Amazon S3 que sirva como nuestro lago de datos
- Una AWS Glue Data Catalog base de datos, una tabla y una partición

Crear recursos

Siga estos pasos para crear sus recursos con la CloudFormation plantilla.

1. Inicie sesión en la CloudFormation consola en https://console.aws.amazon.com/cloudformation en la región EE.UU. Este (Virginia del Norte).
2. Elija Lanzar pila.
3. Seleccione Siguiente en la pantalla Crear pila.
4. Introduzca un Nombre de pila.

5. Para DatalakeAdminUserNamey DatalakeAdminUserPassword, introduzca su nombre de usuario y contraseña de IAM como usuario administrador de Data Lake.
6. Para DataAnalystUsUserNamey DataAnalystUsUserPassword, introduzca el nombre de usuario y la contraseña del nombre de usuario y la contraseña que desee para el usuario analista de datos responsable del mercado estadounidense.
7. Para DataAnalystJpUserNamey DataAnalystJpUserPassword, introduzca el nombre de usuario y la contraseña del nombre de usuario y la contraseña que desee para el usuario analista de datos responsable del mercado japonés.
8. Para DataLakeBucketName, introduzca el nombre de su depósito de datos.
9. Para DatabaseNamey TableNamedéjelo como predeterminado.
10. Elija Siguiente.
11. En la siguiente página, elija Siguiente.
12. Revisa los detalles en la última página y selecciona Acepto que CloudFormation podría crear recursos de IAM.
13. Seleccione Crear.

    La creación de la pila puede tardar un minuto en completarse.

Paso 2: Consultar sin filtros de datos

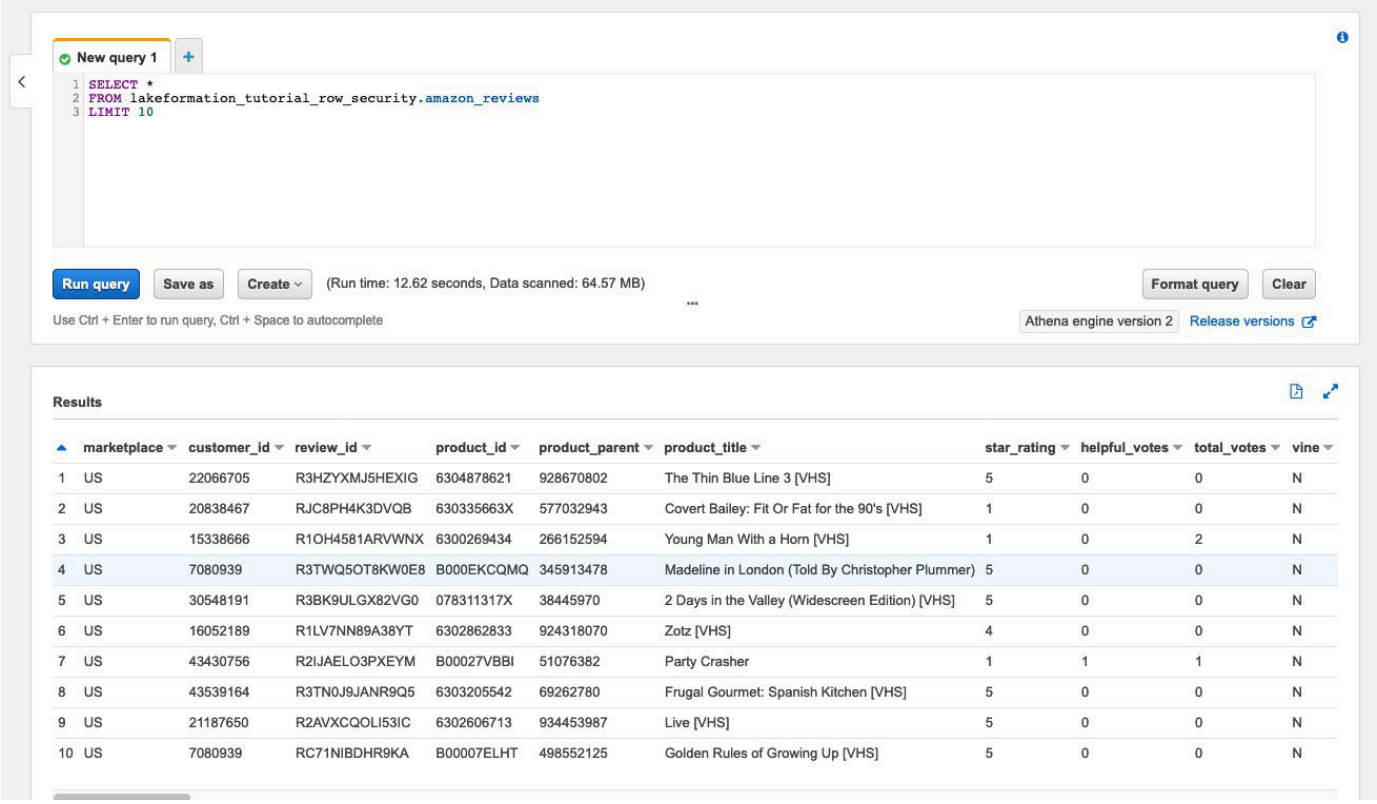
Una vez configurado el entorno, puede consultar la tabla de reseñas de productos. Primero consulte la tabla sin controles de acceso a nivel de fila para asegurarse de que puede ver los datos. Si es la primera vez que ejecuta consultas en Amazon Athena, debe configurar la ubicación de los resultados de la consulta.

Consulte la tabla sin control de acceso a nivel de fila

1. Inicie sesión en la Athena consola https://console.aws.amazon.com/athena/como `DatalakeAdmin` usuario y ejecute la siguiente consulta:

```
SELECT *
FROM lakeformation_tutorial_row_security.amazon_reviews
LIMIT 10
```

La captura de pantalla siguiente muestra el resultado de la consulta. Esta tabla solo tiene una partición, `product_category=Video`, por lo que cada registro es un comentario de revisión de un producto de vídeo.

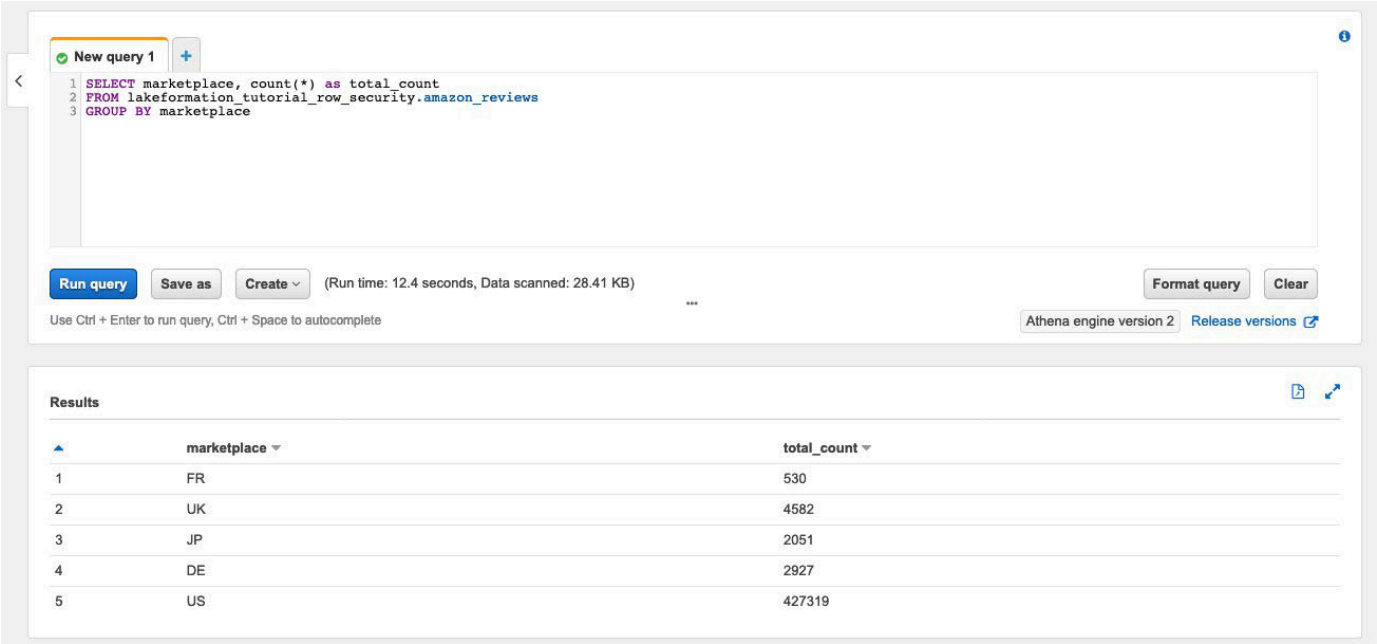


	marketplace	customer_id	review_id	product_id	product_parent	product_title	star_rating	helpful_votes	total_votes	vine
1	US	22066705	R3HZYXMJ5HEXIG	6304878621	928670802	The Thin Blue Line 3 [VHS]	5	0	0	N
2	US	20838467	RJC8PH4K3DVQB	630335663X	577032943	Covert Bailey: Fit Or Fat for the 90's [VHS]	1	0	0	N
3	US	15338666	R1OH4581ARVWNX	6300269434	266152594	Young Man With a Horn [VHS]	1	0	2	N
4	US	7080939	R3TWQ5OT8KW0E8	B000EKCQMQ	345913478	Madeline in London (Told By Christopher Plummer)	5	0	0	N
5	US	30548191	R3BK9ULGX82VG0	078311317X	38445970	2 Days in the Valley (Widescreen Edition) [VHS]	5	0	0	N
6	US	16052189	R1LV7NN89A38YT	6302862833	924318070	Zotz [VHS]	4	0	0	N
7	US	43430756	R2IJAELO3PXEYM	B00027VBBI	51076382	Party Crasher	1	1	1	N
8	US	43539164	R3TN0J9JANR9Q5	6303205542	69262780	Frugal Gourmet: Spanish Kitchen [VHS]	5	0	0	N
9	US	21187650	R2AVXCQOLI53IC	6302606713	934453987	Live [VHS]	5	0	0	N
10	US	7080939	RC71NIBDHR9KA	B00007ELHT	498552125	Golden Rules of Growing Up [VHS]	5	0	0	N

2. A continuación, ejecute una consulta de agregación para recuperar el número total de registros por cada `marketplace`.

```
SELECT marketplace, count(*) as total_count
FROM lakeformation_tutorial_row_security.amazon_reviews
GROUP BY marketplace
```

La captura de pantalla siguiente muestra el resultado de la consulta. La columna `marketplace` tiene cinco valores diferentes. En los pasos siguientes, configurará los filtros basados en filas utilizando la columna `marketplace`.



Paso 3: Configurar los filtros de datos y conceder permisos

En este tutorial se utilizan dos analistas de datos, uno responsable del mercado estadounidense y otro del mercado japonés. Cada analista utiliza Athena para analizar las opiniones de los clientes únicamente para su mercado específico. Cree dos filtros de datos diferentes, uno para el analista responsable del mercado estadounidense y otro para el responsable del mercado japonés. A continuación, conceda a los analistas sus correspondiente permisos.

Cree filtros de datos y conceda permisos

1. Cree un filtro para restringir el acceso a los datos del `marketplace` US.

   a. Inicie sesión en la consola de Lake Formation https://console.aws.amazon.com/lakeformation/en la región EE.UU. Este (Norte de Virginia) como `DatalakeAdmin` usuario.

   b. Elija Filtros de datos.

   c. Elija Crear filtro.

   d. En Nombre del filtro de datos, introduzca `amazon_reviews_US`.

   e. En Base de datos de destino elija la base de datos `lakeformation_tutorial_row_security`.

   f. En Tabla de objetivos, elija la tabla `amazon_reviews`.

   g. En Acceso a nivel de columna, deje el valor predeterminado.

h. En Expresión de filtro de filas, introduzca `marketplace='US'`.

i. Elija Crear filtro.

2. Cree un filtro para restringir el acceso a los datos del `marketplace` japonés.

a. En la página Filtros de datos, seleccione Crear filtro nuevo.

b. En Nombre del filtro de datos, introduzca `amazon_reviews_JP`.

c. En Base de datos de destino elija la base de datos `lakeformation_tutorial_row_security`.

d. En Tabla de destino, elija la tabla `table amazon_reviews`.

e. En Acceso a nivel de columna, deje el valor predeterminado.

f. En Expresión de filtro de filas, introduzca `marketplace='JP'`.

g. Elija Crear filtro.

3. A continuación, conceda permisos a los analistas de datos que utilizan estos filtros de datos. Siga estos pasos para conceder permisos al analista de datos estadounidense (`DataAnalystUS`):

a. En Permisos, elija Permisos de lago de datos.

b. En Permiso de datos, seleccione Conceder.

c. En el Entidades principales, elija Usuarios y roles de IAM y seleccione el rol `DataAnalystUS`.

d. En Etiquetas LF o recursos del catálogo, elija Recursos de Catálogo de datos con nombre.

e. En Database (Base de datos), elija `lakeformation_tutorial_row_security`.

f. Para Tablas (opcional), elija `amazon_reviews`.

g. Para Filtros de datos: opcional, seleccione `amazon_reviews_US`.

h. Para Permisos de filtro de datos, elija Seleccionar.

i. Elija Conceder.

4. Siga estos pasos para conceder permisos al analista de datos japonés (`DataAnalystJP`):

a. En Permisos, elija Permisos de lago de datos.

b. En Permiso de datos, seleccione Conceder.

c. En el Entidades principales, elija Usuarios y roles de IAM y seleccione el rol `DataAnalystJP`.

d. En Etiquetas LF o recursos del catálogo, elija Recursos de Catálogo de datos con nombre.

e. En Database (Base de datos), elija `lakeformation_tutorial_row_security`.

f. Para Tablas (opcional), elija `amazon_reviews`.

g. Para Filtros de datos: opcional, seleccione `amazon_reviews_JP`.

h. Para Permisos de filtro de datos, elija Seleccionar.

i. Elija Conceder.

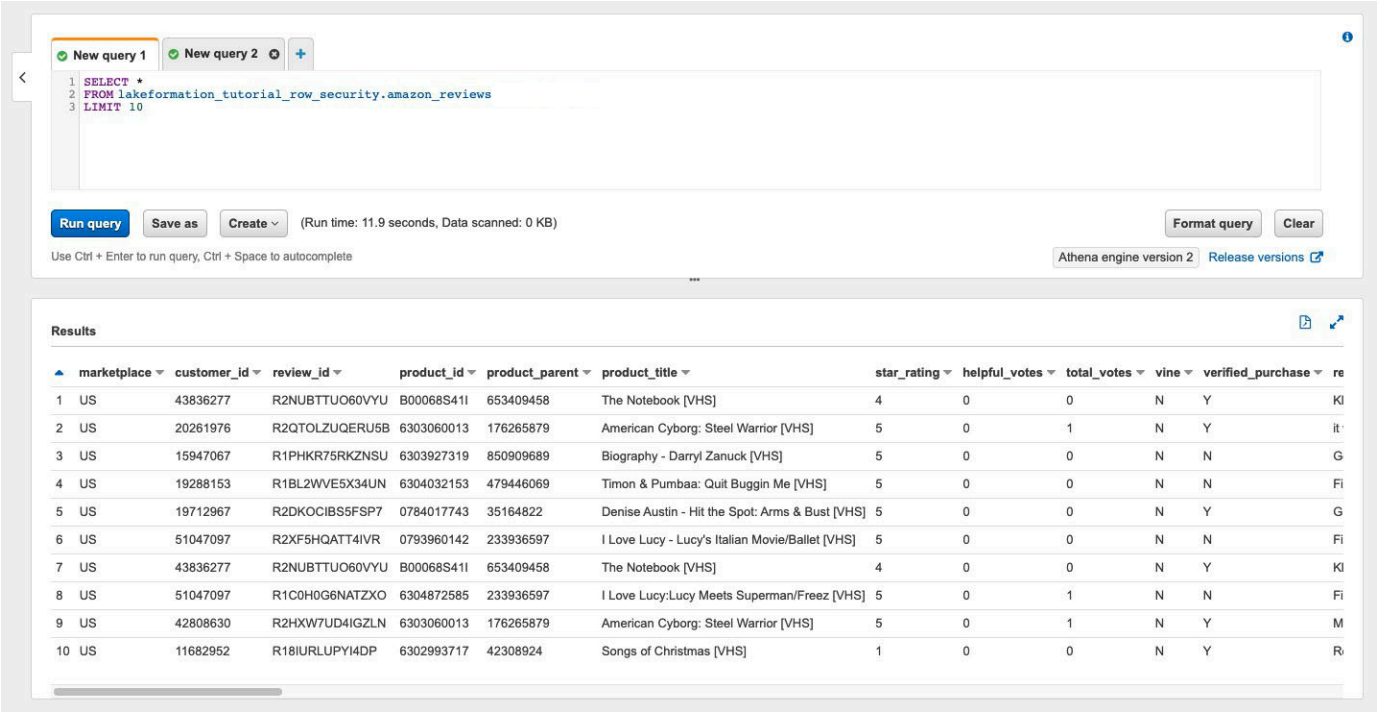
Paso 4: Consultar con filtros de datos

Con los filtros de datos adjuntos a la tabla de reseñas de productos, ejecute algunas consultas y compruebe cómo Lake Formation aplica los permisos.

1. Inicie sesión en la consola de Athena https://console.aws.amazon.com/athena/como usuario. `DataAnalystUS`
2. Ejecute la siguiente consulta para recuperar algunos registros, que se filtran en función de los permisos de nivel de fila que definimos:

```
SELECT *
FROM lakeformation_tutorial_row_security.amazon_reviews
LIMIT 10
```

La captura de pantalla siguiente muestra el resultado de la consulta.



	marketplace	customer_id	review_id	product_id	product_parent	product_title	star_rating	helpful_votes	total_votes	vine	verified_purchase	re
1	US	43836277	R2NUBTTUO60VYU	B00068S41I	653409458	The Notebook [VHS]	4	0	0	N	Y	Kl
2	US	20261976	R2QTOLZUQERU5B	6303060013	176265879	American Cyborg: Steel Warrior [VHS]	5	0	1	N	Y	it
3	US	15947067	R1PHKR75RKZNSU	6303927319	850909689	Biography - Darryl Zanuck [VHS]	5	0	0	N	N	G
4	US	19288153	R1BL2WVE5X34UN	6304032153	479446069	Timon & Pumbaa: Quit Buggin Me [VHS]	5	0	0	N	N	Fi
5	US	19712967	R2DKOCIBS5FSP7	0784017743	35164822	Denise Austin - Hit the Spot: Arms & Bust [VHS]	5	0	0	N	Y	G
6	US	51047097	R2XF5HQATT4IVR	0793960142	233936597	I Love Lucy - Lucy's Italian Movie/Ballet [VHS]	5	0	0	N	N	Fi
7	US	43836277	R2NUBTTUO60VYU	B00068S41I	653409458	The Notebook [VHS]	4	0	0	N	Y	Kl
8	US	51047097	R1C0H0G6NATZXO	6304872585	233936597	I Love Lucy:Lucy Meets Superman/Freez [VHS]	5	0	1	N	N	Fi
9	US	42808630	R2HXW7UD4IGZLN	6303060013	176265879	American Cyborg: Steel Warrior [VHS]	5	0	1	N	Y	M
10	US	11682952	R18IURLUPYI4DP	6302993717	42308924	Songs of Christmas [VHS]	1	0	0	N	Y	R

3. Del mismo modo, ejecute una consulta para contar el número total de registros por mercado.

```
SELECT marketplace , count ( * ) as total_count
FROM lakeformation_tutorial_row_security .amazon_reviews
GROUP BY marketplace
```

El resultado de la consulta solo muestra el US `marketplace` en los resultados. Esto se debe a que el usuario solo puede ver las filas en las que el valor de la columna `marketplace` sea igual a US.

4. Cambie al usuario `DataAnalystJP` y ejecute la misma consulta.

```
SELECT *
FROM lakeformation_tutorial_row_security.amazon_reviews
LIMIT 10
```

El resultado de la consulta solo muestra el `marketplace JP` en los resultados.

5. Ejecute la consulta para contar el número total de registros por `marketplace`.

```
SELECT marketplace, count(*) as total_count
FROM lakeformation_tutorial_row_security.amazon_reviews
GROUP BY marketplace
```

El resultado de la consulta solo muestra la fila que pertenece al `marketplace JP` en los resultados.

Paso 5: Limpiar AWS los recursos

Eliminar recursos

Para evitar que se le cobren cargos no deseados Cuenta de AWS, puede eliminar los AWS recursos que utilizó para este tutorial.

- Elimine la pila de formación de nube.

Compartir un lago de datos utilizando el control de acceso basado en etiquetas de Lake Formation y recursos con nombre

Este tutorial muestra cómo puede configurarse AWS Lake Formation para compartir de forma segura los datos almacenados en un lago de datos con varias empresas, organizaciones o unidades de negocio, sin tener que copiar toda la base de datos. Existen dos opciones para compartir sus bases de datos y tablas con otra persona Cuenta de AWS mediante el control de acceso multicuenta de Lake Formation:

- Control de acceso basado en etiquetas de Lake Formation (recomendado)

  El control de acceso basado en etiquetas de Lake Formation es una estrategia de autorización que define permisos basados en atributos. En Lake Formation, estos atributos se denominan etiquetas LF. Para obtener más detalles, consulte Gestión de un lago de datos mediante el control de acceso basado en etiquetas de Lake Formation.
- Recursos con nombre de Lake Formation

  El método de recursos con nombre de Lake Formation es una estrategia de autorización que define los permisos para los recursos. Los recursos incluyen bases de datos, tablas y columnas. Los administradores del lago de datos pueden asignar y revocar permisos sobre los recursos de Lake Formation. Para obtener más detalles, consulte Cross-account intercambio de datos en Lake Formation.

  Recomendamos utilizar recursos con nombre si el administrador del lago de datos prefiere conceder permisos explícitamente a recursos individuales. Cuando utiliza el método de recurso

designado para conceder permisos de Lake Formation sobre un recurso del catálogo de datos a una cuenta externa, Lake Formation usa AWS Resource Access Manager (AWS RAM) para compartir el recurso.

Temas

- Destinatarios previstos
- Configurar los ajustes del Catálogo de datos de Lake Formation en la cuenta del productor
- Paso 1: Aprovisione sus recursos mediante plantillas AWS CloudFormation
- Paso 2: Requisitos previos para compartir entre cuentas de Lake Formation
- Paso 3: Implementar el uso compartido entre cuentas mediante el método de control de acceso basado en etiquetas
- Paso 4: Implementar el método de recurso con nombre
- Paso 5: Limpiar AWS los recursos

Destinatarios previstos

Este tutorial está dirigido a administradores de datos, ingenieros de datos y analistas de datos. Cuando se trata de compartir tablas del catálogo de datos AWS Glue y administrar permisos en Lake Formation, los administradores de datos de las cuentas productoras tienen una propiedad funcional en función de las funciones que respaldan y pueden conceder acceso a varios consumidores, organizaciones externas y cuentas. La tabla siguiente enumera los roles que se utilizan en este tutorial:

Rol	Description (Descripción)
DataLakeAdminProducer	El usuario de IAM administrador del lago de datos tiene el acceso siguiente: • Acceso completo de lectura, escritura y actualización a todos los recursos del Catálogo de datos • Posibilidad de conceder permisos a los recursos

Rol	Description (Descripción)
	• Puede crear enlaces a recursos para la tabla compartida • Puede adjuntar etiquetas LF a los recursos, lo que proporciona acceso a las entidades principales en función de cualquier política creada por los administradores de datos.
DataLakeAdminConsumer	El usuario de IAM administrador del lago de datos tiene el acceso siguiente: • Acceso completo de lectura, escritura y actualización a todos los recursos del Catálogo de datos • Posibilidad de conceder permisos a los recursos • Puede crear enlaces a recursos para la tabla compartida • Puede adjuntar etiquetas LF a los recursos, lo que proporciona acceso a las entidades principales en función de cualquier política creada por los administradores de datos.
DataAnalyst	El DataAnalyst usuario tiene los siguientes accesos: • Acceso específico a los recursos compartid os por las políticas de acceso basadas en etiquetas de Lake Formation o mediante el método de recursos con nombre

Configurar los ajustes del Catálogo de datos de Lake Formation en la cuenta del productor

Antes de empezar este tutorial, debe disponer de una Cuenta de AWS que pueda utilizar para iniciar sesión como usuario administrativo con los permisos correctos. Para obtener más información, consulte Completa la inicial AWS tareas de configuración.

En este tutorial, se supone que está familiarizado con IAM. Para obtener más información acerca de IAM, consulte la Guía del usuario de IAM.

Configurar los ajustes del Catálogo de datos de Lake Formation en la cuenta del productor

> **Note**
>
> En este tutorial, la cuenta que tiene la tabla de origen se denomina cuenta del productora, y la que necesita acceder a la tabla de origen se denomina cuenta del consumidor.

Lake Formation proporciona su propio modelo de gestión de permisos. Para mantener la compatibilidad con versiones anteriores del modelo de permisos de IAM, el `Super` permiso se concede al grupo `IAMAllowedPrincipals` en todos los AWS Glue Data Catalog recursos existentes de forma predeterminada. Además, la configuración de control de acceso para utilizar solo IAM está activada para los nuevos recursos del Catálogo de datos. Este tutorial utiliza el control de acceso específico utilizando permisos de Lake Formation y políticas de IAM para el control de acceso menos específico. Para obtener más información, consulte Métodos para el control de acceso específico. Por lo tanto, antes de usar una AWS CloudFormation plantilla para una configuración rápida, debe cambiar la configuración del catálogo de datos de Lake Formation en la cuenta del productor.

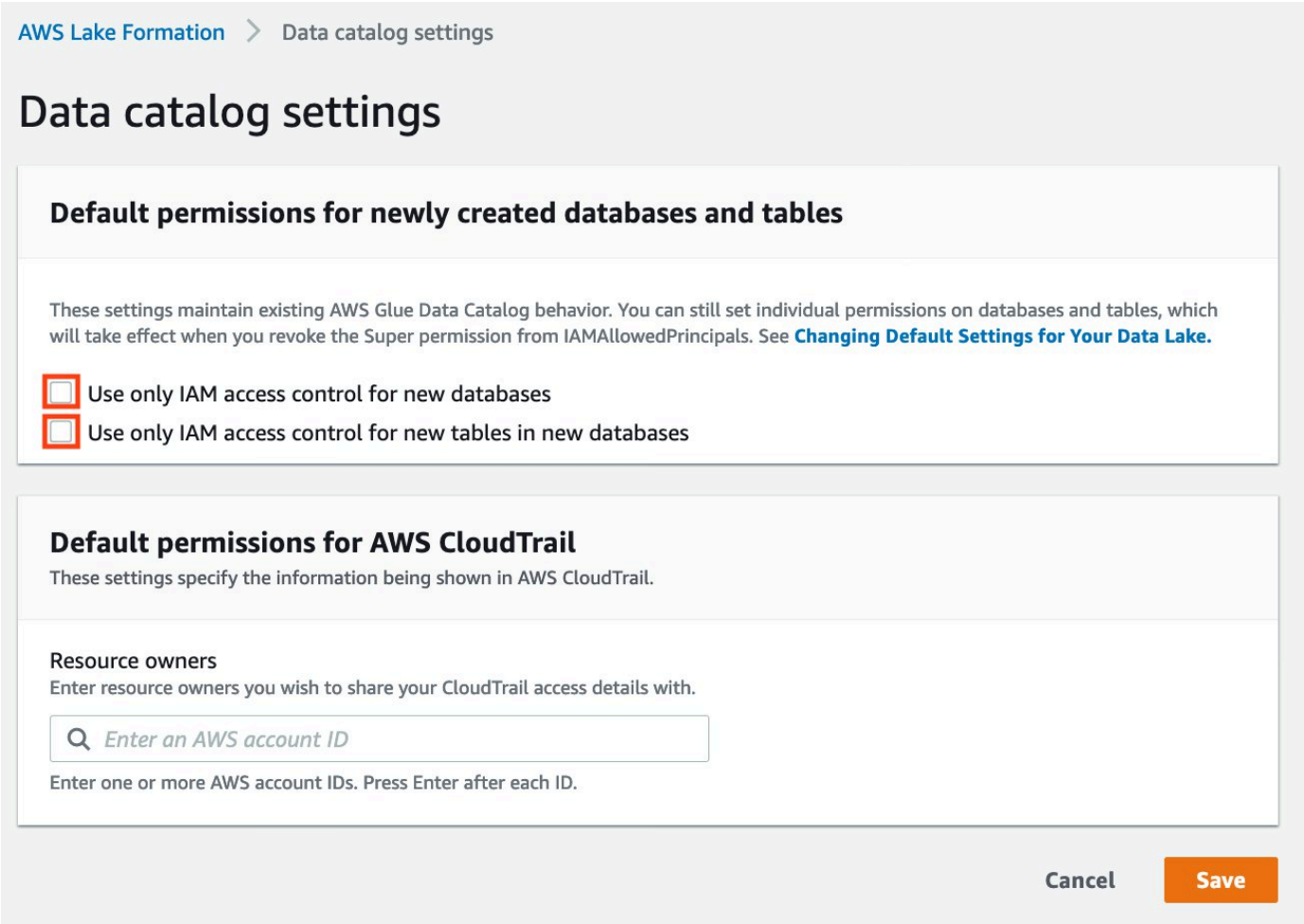
> **Important**
>
> Esta configuración afecta a todas las bases de datos y tablas recién creadas, por lo que se recomienda completar este tutorial en una cuenta que no sea de producción o en una cuenta nueva. Además, si utiliza una cuenta compartida (como la cuenta de desarrollo de su empresa), asegúrese de que no afecta a otros recursos. Si prefiere mantener la configuración de seguridad predeterminada, deberá completar un paso adicional cuando comparta recursos con otras cuentas, en el que revocará el permiso Super predeterminado

de `IAMAllowedPrincipals` en la base de datos o tabla. Trataremos los detalles más adelante en este tutorial.

Para configurar los ajustes del Catálogo de datos de Lake Formation en la cuenta del productor:

1. Inicie sesión Consola de administración de AWS con la cuenta de productor como usuario administrador o como usuario con permiso de la `PutDataLakeSettings` API de Lake Formation.
2. En la consola de Lake Formation, desde panel de navegación, Catálogo de datos, seleccione Configuración.
3. Desmarque Usar solo control de acceso IAM para nuevas bases de datos y Usar solo control de acceso IAM para nuevas tablas en nuevas bases de datos.

   Seleccione Save.



También puede retirar los permisos de `CREATE_DATABASE` para `IAMAllowedPrincipals` en la sección de roles y tareas administrativas, creadores de bases de datos. Solo entonces podrá decidir quién puede crear una nueva base de datos mediante los permisos de Lake Formation.

Paso 1: Aprovisione sus recursos mediante plantillas AWS CloudFormation

La CloudFormation plantilla de la cuenta de productor genera los siguientes recursos:

- Un bucket de Amazon S3 que sirva como lago de datos
- Una función de Lambda (para recursos personalizados respaldados por Lambda CloudFormation). Esta función sirve para copiar archivos de datos de muestra del bucket público de Amazon S3 al bucket de Amazon S3.
- Usuarios y políticas de IAM:. DataLakeAdminProducer
- La configuración y los permisos apropiados de Lake Formation, como:
  - Definición del administrador del lago de datos de Lake Formation en la cuenta del productor
  - Registro de un bucket de Amazon S3 como ubicación del lago de datos de Lake Formation (cuenta de productor)
- Una AWS Glue Data Catalog base de datos, tabla y partición. Como hay dos opciones para compartir recursos Cuentas de AWS, esta plantilla crea dos conjuntos independientes de base de datos y tablas.

La CloudFormation plantilla de la cuenta de consumidor genera los siguientes recursos:

- Usuarios de IAM y políticas:
  - DataLakeAdminConsumer
  - DataAnalyst
- Una AWS Glue Data Catalog base de datos. Esta base de datos sirve para crear enlaces de recursos a recursos compartidos.

Crear sus recursos en la cuenta del productor

1. Inicie sesión en la AWS CloudFormation consola en https://console.aws.amazon.com/cloudformation en la región EE.UU. Este (Norte de Virginia).
2. Seleccione Lanzar pila.

3. Elija Siguiente.
4. En Nombre de la pila, escriba un nombre para la pila, como `stack-producer`.
5. En la sección Configuración de usuario, introduzca la contraseña para `ProducerDatalakeAdminUserName` y `ProducerDatalakeAdminUserPassword`.
6. Para ello DataLakeBucketName, introduce el nombre del depósito de tu lago de datos. El nombre debe ser único de forma global.
7. Para DatabaseNamey TableName, deje los valores predeterminados.
8. Elija Siguiente.
9. En la página siguiente, seleccione Siguiente.
10. Revisa los detalles en la última página y selecciona Sé que AWS CloudFormation podría crear recursos de IAM.
11. Seleccione Crear.

    La creación de la pila puede llevar hasta un minuto.

Crear sus recursos en la cuenta del consumidor

1. Inicie sesión en la AWS CloudFormation consola en https://console.aws.amazon.com/cloudformation en la región EE.UU. Este (Norte de Virginia).
2. Seleccione Lanzar pila.
3. Elija Siguiente.
4. En Nombre de la pila, escriba un nombre para la pila, como `stack-consumer`.
5. En la sección Configuración de usuario, introduzca la contraseña para `ConsumerDatalakeAdminUserName` y `ConsumerDatalakeAdminUserPassword`.
6. Para `DataAnalystUserName` y `DataAnalystUserPassword`, introduzca el nombre de usuario y la contraseña que desee para el usuario de IAM del analista de datos.
7. Para ello DataLakeBucketName, introduce el nombre del depósito de tu lago de datos. El nombre debe ser único de forma global.
8. En DatabaseName, deje los valores predeterminados.
9. Para `AthenaQueryResultS3BucketName`, introduzca el nombre del bucket de Amazon S3 que almacena los resultados de consultas de Amazon Athena. Si no dispone de un bucket de Amazon S3, cree uno nuevo.
10. Elija Siguiente.

11. En la página siguiente, seleccione Siguiente.
12. Revise los detalles en la última página y seleccione Acepto que AWS CloudFormation podría crear recursos de IAM.
13. Seleccione Crear.

    La creación de la pila puede llevar hasta un minuto.

> **Note**
>
> Tras completar el tutorial, elimina la pila CloudFormation para evitar incurrir en cargos. Compruebe que los recursos se hayan eliminado correctamente en el estado de evento de la pila.

Paso 2: Requisitos previos para compartir entre cuentas de Lake Formation

Antes de compartir recursos con Lake Formation, hay requisitos previos tanto para el método de control de acceso basado en etiquetas como para el método de recurso con nombre.

Completar los requisitos previos del control de acceso basado en etiquetas para el intercambio de datos entre cuentas

- Para más información sobre los requisitos para compartir datos entre cuentas, consulte la sección Requisitos previos del capítulo sobre compartir datos entre cuentas.

  Para compartir los recursos del catálogo de datos con la versión 3 o superior de la configuración de la versión Cross Account, el otorgante debe tener los permisos de IAM definidos en la política AWS `AWSLakeFormationCrossAccountManager` gestionada de su cuenta.

  Si utiliza la versión 1 o la versión 2 de la configuración de la versión entre cuentas, antes de poder utilizar el método de control de acceso basado en etiquetas para conceder acceso entre cuentas a los recursos, deberá añadir el siguiente objeto de permisos `JSON` entre cuentas a la política de recursos del Catálogo de datos en la cuenta de productor. De este modo, la cuenta del consumidor permite a la cuenta del consumidor acceder al Catálogo de datos cuando `glue:EvaluatedByLakeFormationTags` es cierto. Además, esta condición se cumple para los recursos sobre los que haya concedido permiso mediante etiquetas de permiso de Lake Formation a la cuenta del consumidor. Esta política es obligatoria para todas las personas Cuenta de AWS a las que se concedan permisos.

La política siguiente debe estar dentro de un elemento `Statement`. En la sección siguiente, analizamos toda la política de IAM.

```
{
    "Effect": "Allow",
    "Action": [
        "glue:*"
    ],
    "Principal": {
        "AWS": [
            "consumer-account-id"
        ]
    },
    "Resource": [
        "arn:aws:glue:region:account-id:table/*",
        "arn:aws:glue:region:account-id:database/*",
        "arn:aws:glue:region:account-id:catalog"
    ],
    "Condition": {
        "Bool": {
            "glue:EvaluatedByLakeFormationTags": true
        }
    }
}
```

Completar los requisitos previos del método de recursos con nombre para compartir entre cuentas

1. Si en su cuenta no existe una política de recursos del Catálogo de datos, las concesiones que efectúe entre cuentas procederán como de costumbre. Sin embargo, si existe una política de recursos del Catálogo de datos, debe añadirle la siguiente instrucción para permitir que sus concesiones entre cuentas tengan éxito si se efectúan con el método de recursos con nombre. Si piensa utilizar solo el método de recursos con nombre, o solo el método de control de acceso basado en etiquetas, puede omitir este paso. En este tutorial, evaluamos ambos métodos, y necesitamos añadir la política siguiente.

   La política siguiente debe estar dentro de un elemento `Statement`. En la sección siguiente, analizamos toda la política de IAM.

```
{
          "Effect": "Allow",
          "Action": [
          "glue:ShareResource"
          ],
          "Principal": {
            "Service":"ram.amazonaws.com"
          },
          "Resource": [
              "arn:aws:glue:region:account-id:table/*/*",
              "arn:aws:glue:region:account-id:database/*",
              "arn:aws:glue:region:account-id:catalog"
          ]
}
```

2. A continuación, añada la política de AWS Glue Data Catalog recursos mediante AWS Command Line Interface (AWS CLI).

   Si concede permisos entre cuentas utilizando tanto el método de control de acceso basado en etiquetas como el método de recursos con nombre, debe establecer el argumento `EnableHybrid` en "true" cuando añada las políticas anteriores. Debido a que esta opción no es actualmente Compatibilidad en la consola, y debe utilizar la API `glue:PutResourcePolicy` y la AWS CLI.

   En primer lugar, cree un documento de políticas (como policy.json) y añada las dos políticas anteriores. *`consumer-account-id`*Sustitúyala por *`account ID`* la que Cuenta de AWS recibe la subvención, *`region`* por la región del catálogo de datos que contiene las bases de datos y tablas sobre las que estás concediendo los permisos y *`account-id`* por la Cuenta de AWS ID del productor.

   Introduzca el siguiente AWS CLI comando. *`glue-resource-policy`*Sustitúyalo por los valores correctos (como file: //policy.json).

   ```
   aws glue put-resource-policy --policy-in-json glue-resource-policy --enable-hybrid
    TRUE
   ```

   Para obtener más información, consulte put-resource-policy.

Paso 3: Implementar el uso compartido entre cuentas mediante el método de control de acceso basado en etiquetas

En esta sección, encontrará los siguientes pasos generales:

1. Definir una etiqueta LF.
2. Asignar la etiqueta LF al recurso de destino.
3. Conceder permisos de etiqueta LF a la cuenta de consumidor.
4. Conceder permisos sobre datos a la cuenta de consumidor.
5. Opcionalmente, revocar los permisos para `IAMAllowedPrincipals` sobre las bases de datos, las tablas y las columnas.
6. Crear un enlace de recursos en la base de datos compartida.
7. Crear una etiqueta LF y asignarla a la base de datos de destino.
8. Conceder permisos de datos de etiqueta LF a la cuenta de consumidor.

Definir una etiqueta LF.

> **Note**
>
> Si ha iniciado sesión en su cuenta de productor, ciérrela antes de completar los pasos siguientes.

1. Inicie sesión en la cuenta del productor como administrador del lago de datos en. https://console.aws.amazon.com/lakeformation/ Utilice el número de cuenta del productor, el nombre de usuario de IAM (el predeterminado es `DatalakeAdminProducer`) y la contraseña que especificó durante la creación de la pila CloudFormation .
2. En la consola de Lake Formation (https://console.aws.amazon.com/lakeformation/), en el panel de navegación, en Permisos, elija Etiquetas LF y Permisos.
3. Seleccione Agregar etiqueta LF.

Asignar la etiqueta LF al recurso de destino.

Asigne la etiqueta LF al recurso de destino y conceda permisos de datos a otra cuenta

Como administrador de un lago de datos, puede adjuntar etiquetas a los recursos. Si tiene previsto utilizar un rol independiente, es posible que tenga que conceder permisos de descripción y vinculación a ese rol.

1. En el panel de navegación, en Catálogo de datos, elija Bases de datos.
2. Seleccione la base de datos (`lakeformation_tutorial_cross_account_database_tbac`) de destino y, en el menú Acciones, elija Editar etiquetas LF.

   En este tutorial, asignará una etiqueta LF a una base de datos, pero también podrá asignar etiquetas LF a tablas y columnas.
3. Seleccione Asignar una nueva etiqueta LF.
4. Añada la clave `Confidentiality` y el valor `public`.
5. Seleccione Save.

Conceder permisos de etiqueta LF a la cuenta de consumidor.

Si aún está en la cuenta del productor, conceda permisos a la cuenta del consumidor para acceder a la etiqueta LF.

1. En el panel de navegación, en Permisos, elija Etiquetas LF y permisos.
2. Elija la pestaña Etiquetas LF y seleccione la clave y los valores de la etiqueta LF que se comparte con la cuenta del consumidor (clave `Confidentiality` y valor `public`).
3. Elija Conceder permisos.
4. En Tipo de permiso, elija Permisos de par clave-valor de etiquetas LF.
5. En Entidades principales, seleccione Cuentas externas.
6. Introduzca el ID Cuenta de AWS de destino.

   Cuentas de AWS dentro de la misma organización aparecen automáticamente. De lo contrario, tendrás que introducir el Cuenta de AWS ID manualmente.
7. En Permisos, seleccione Describir.

   Estos son los permisos concedidos a la cuenta del consumidor. Los permisos concedibles son permisos que la cuenta de consumidor puede conceder a otras entidades principales.
8. Elija Conceder.

En este punto, el administrador del lago de datos del consumidor debería ser capaz de encontrar la etiqueta de la política que se comparte a través de la cuenta del consumidor de la consola Lake Formation, en Permisos, Etiquetas LF y permisos.

Conceder permiso de datos a la cuenta del consumidor

Ahora proporcionaremos acceso a los datos a la cuenta del consumidor especificando una expresión de etiquetas LF y concediendo a la cuenta del consumidor acceso a cualquier tabla o base de datos que coincida con la expresión.

1. En el panel de navegación, en Permisos, Permisos de lago de datos, seleccione Conceder.
2. En el caso de los principales, selecciona Cuentas externas e introduce el Cuenta de AWS ID de destino.
3. Para las etiquetas LF o los recursos del catálogo, elija la clave y los valores de la etiqueta LF que se va a compartir con la cuenta del consumidor (clave `Confidentiality` y valor `public`).
4. En Permisos, bajo Recursos que coinciden con las etiquetas LF (recomendado) elija Agregar etiqueta LF.
5. Seleccione la clave y el valor de la etiqueta que se va a compartir con la cuenta del consumidor (clave `Confidentiality` y valor `public`).
6. Para Permisos de base de datos, seleccione Describir en Permisos de bases de datos para conceder permisos de acceso a nivel de base de datos.
7. El administrador del lago de datos del consumidor debería poder encontrar la etiqueta de política que se comparte a través de la cuenta del consumidor en la consola de Lake Formation https://console.aws.amazon.com/lakeformation/, en Permisos, funciones y tareas administrativas, etiquetas LF.
8. Seleccione Describir en Permisos concedibles para que la cuenta de consumidor pueda conceder permisos a nivel de base de datos a sus usuarios.
9. En Permisos de tabla y columna, elija Seleccionar y Describir en Permisos de tabla.
10. Elija Seleccionar y Describir en Permisos concedibles.
11. Elija Conceder.

Revocar los permisos para **`IAMAllowedPrincipals`** sobre las bases de datos, las tablas y las columnas (opcional).

Al principio de este tutorial, cambió la configuración del Catálogo de datos de Lake Formation. Si omitió esa parte, este paso es obligatorio. Si ha cambiado la configuración del Catálogo de datos de Lake Formation, puede omitir este paso.

En este paso, necesitamos revocar el permiso Super por defecto de `IAMAllowedPrincipals` en la base de datos o tabla. Para obtener más información, consulte Paso 4: Cambiar sus almacenes de datos al modelo de permisos de Lake Formation.

Antes de revocar el permiso para `IAMAllowedPrincipals`, asegúrese de que ha concedido a las entidades principales de IAM existentes el permiso necesario a través de Lake Formation. Esto conlleva tres pasos:

1. Añadir el permiso de IAM al rol o usuario de IAM de destino con la acción `GetDataAccess` de Lake Formation (con la política de IAM).
2. Conceder al rol o usuario de IAM de destino permisos de datos de Lake Formation (alterar, seleccionar, etc.).
3. A continuación, revocar los permisos para `IAMAllowedPrincipals`. De lo contrario, después de revocar los permisos para `IAMAllowedPrincipals`, es posible que las entidades principales de IAM existentes ya no puedan acceder a la base de datos de destino o al Catálogo de datos.

   La revocación del permiso Super para `IAMAllowedPrincipals` es necesaria cuando desea aplicar el modelo de permisos Lake Formation (en lugar del modelo de políticas de IAM) para gestionar el acceso de los usuarios dentro de una misma cuenta o entre varias cuentas utilizando el modelo de permisos de Lake Formation. No tiene que revocar el permiso de `IAMAllowedPrincipals` para otras tablas en las que desee mantener el modelo de política de IAM tradicional.

   En este punto, el administrador del lago de datos de la cuenta del consumidor debería poder encontrar la base de datos y la tabla que se comparten a través de la cuenta del consumidor en la consola de Lake Formation https://console.aws.amazon.com/lakeformation/, en la sección Data Catalog, bases de datos. Si no es así, confirme si lo siguiente está bien configurado:

   1. Están asignados la etiqueta y los valores de política correctos a las bases de datos y tablas de destino.

2. Están asignados el permiso de etiqueta y el permiso de datos correctos a la cuenta del consumidor.
3. Está revocado el permiso super predeterminado de `IAMAllowedPrincipals` en la base de datos o tabla.

Crear un enlace de recursos a la tabla compartida

Cuando un recurso se comparte entre cuentas, y los recursos compartidos no se ubican en el Catálogo de datos de las cuentas consumidoras. Para que estén disponibles, y consultar los datos subyacentes de una tabla compartida utilizando servicios como Athena, necesitamos crear un enlace de recursos a la tabla compartida. Un enlace de recursos es un objeto del Catálogo de datos que es un enlace a una base de datos o tabla local o compartida. Para obtener más información, consulte Creación de enlaces de recursos. Al crear un enlace de recursos, puede:

- Asignar un nombre diferente a una base de datos o tabla que se ajuste a sus políticas de nombres de recursos del Catálogo de datos.
- Utilizar servicios como Athena y Redshift Spectrum para consultar bases de datos o tablas compartidas.

Para crear un enlace de recursos:

1. Si ha iniciado sesión en su cuenta de consumidor, cierre la sesión.
2. Inicie sesión como administrador del lago de datos de la cuenta del consumidor. Utilice el ID de la cuenta de consumidor, el nombre de usuario de IAM (predeterminado DatalakeAdminConsumer) y la contraseña que especificó durante la creación de la CloudFormation pila.
3. En la consola de Lake Formation (https://console.aws.amazon.com/lakeformation/), en el panel de navegación, en Catálogo de datos, bases de datos, elija la base de datos compartida`lakeformation_tutorial_cross_account_database_tbac`.

   Si no ve la base de datos, vuelva a revisar los pasos anteriores para ver si todo está configurado correctamente.
4. Elija Ver las tablas.
5. Elija la tabla compartida `amazon_reviews_table_tbac`.
6. En el menú Acciones, elija Crear recurso.

7. En Nombre del enlace de recursos, introduzca un nombre (para este tutorial, `amazon_reviews_table_tbac_resource_link`).
8. En Base de datos, seleccione la base de datos en la que se creó el enlace al recurso (para esta publicación, la pila CloudFormation n creó la base de datos`lakeformation_tutorial_cross_account_database_consumer`).
9. Seleccione Crear.

   El enlace de recursos aparece en Catálogo de datos, Tablas.

Crear una etiqueta LF y asignarla a la base de datos de destino

Las etiquetas de Lake Formation se encuentran en el mismo Catálogo de datos que los recursos. Esto significa que las etiquetas creadas en la cuenta del productor no están disponibles para su uso al conceder acceso a los enlaces de recursos en la cuenta del consumidor. Es necesario crear un conjunto separado de etiquetas LF en la cuenta del consumidor para utilizar el control de acceso basado en etiquetas LF al compartir los enlaces de recursos en la cuenta del consumidor.

1. Defina la etiqueta LF en la cuenta del consumidor. Para este tutorial, utilizamos la clave `Division` y los valores `sales`, `marketing` y `analyst`.
2. Asigne la clave de etiquetas LF `Division` y el valor `analyst` a la base de datos `lakeformation_tutorial_cross_account_database_consumer`, donde se crea el enlace de recursos.

Conceder permiso de datos de etiquetas LF al consumidor

Como paso final, conceda permiso de datos de etiquetas LF al consumidor.

1. En el panel de navegación, en Permisos, Permisos de lago de datos, seleccione Conceder.
2. En Entidades principales, seleccione los usuarios y roles de IAM y elija el usuario `DataAnalyst`.
3. En las etiquetas LF o los recursos del catálogo, elija los recursos que coincidan con etiquetas LF (recomendado).
4. Elija la clave división y el valor analista.
5. Para los permisos de la base de datos, seleccione Describir en Permisos de la base de datos.
6. En Permisos de tabla y columna, elija Seleccionar y Describir en Permisos de tabla.

7. Elija Conceder.
8. Repita estos pasos para el usuario `DataAnalyst`, donde la clave de la etiqueta LF es `Confidentiality` y el valor es `public`.

   En este punto, el usuario analista de datos de la cuenta del consumidor debería poder encontrar la base de datos y el enlace al recurso y consultar la tabla compartida a través de la consola de Athena en. https://console.aws.amazon.com/athena/ Si no es así, confirme si lo siguiente está bien configurado:

   - Está creado el enlace de recursos para la tabla compartida
   - Ha concedido al usuario acceso a las etiquetas LF compartidas por la cuenta de productor
   - Ha concedido al usuario acceso a las etiquetas LF asociadas al enlace de recursos y a la base de datos en la que se crea el enlace de recursos.
   - Compruebe si ha asignado las etiquetas LF correctas al enlace de recursos y a la base de datos en la que se ha creado el enlace de recursos.

Paso 4: Implementar el método de recurso con nombre

Para utilizar el método de recursos con nombre, le guiaremos a través de los siguientes pasos básicos:

1. De manera opcional, revocar el permiso para `IAMAllowedPrincipals` sobre la base de datos, las tablas y las columnas.
2. Conceder permiso de datos a la cuenta del consumidor.
3. Acepta un recurso compartido de AWS Resource Access Manager.
4. Crear un enlace de recursos para la tabla compartida.
5. Concede permiso de datos para la tabla compartida al consumidor.
6. Concede permiso de datos para el enlace de recursos al consumidor.

Revocar permiso para **`IAMAllowedPrincipals`** sobre las bases de datos, las tablas y las columnas (opcional).

- Al principio de este tutorial, cambiamos la configuración del Catálogo de datos de Lake Formation. Si omitió esa parte, este paso es obligatorio. Para obtener instrucciones, consulte el paso opcional de la sección anterior.

Conceder permiso de datos a la cuenta del consumidor

1. > **Note**
   >
   > Si ha iniciado sesión en la cuenta de productor como otro usuario, cierre primero la sesión.

   Inicie sesión en la consola de Lake Formation https://console.aws.amazon.com/lakeformation/con el administrador del lago de datos de la cuenta del productor con el Cuenta de AWS ID, el nombre de usuario de IAM (el predeterminado es`DatalakeAdminProducer`) y la contraseña especificados durante la creación de la CloudFormation pila.
2. En la página Permisos, en Permisos del lago de datos elija Conceder.
3. En Directores, elija Cuentas externas e introduzca una Cuenta de AWS IDs o AWS más organizaciones. IDs Para más información, consulte: Organizaciones AWS.

   Las organizaciones a las que pertenece la cuenta de productor y que Cuentas de AWS están dentro de la misma organización aparecen automáticamente. De lo contrario, introduzca manualmente el ID de la cuenta o el ID de la organización.
4. Para las etiquetas LF o los recursos del catálogo, elija `Named data catalog resources`.
5. En Bases de datos elija la base de datos `lakeformation_tutorial_cross_account_database_named_resource`.
6. Seleccione Agregar etiqueta LF.
7. En Tablas, seleccione Todas las tablas.
8. En Permisos de columna de tabla, elija Seleccionar y Describir en Permisos de tabla.
9. Elija Seleccionar y Describir en Permisos concedibles.
10. De forma opcional, para Permisos de datos, elija Acceso simple basado en columnas si se requiere una gestión de permisos a nivel de columna.
11. Elija Conceder.

Si no ha revocado el permiso para `IAMAllowedPrincipals`, aparecerá un error de concesión de permisos. En este punto, deberías ver la tabla objetivo que se comparte AWS RAM con la cuenta del consumidor en Permisos y permisos de datos.

Acepta compartir un recurso de AWS RAM

> **Note**
>
> Este paso solo es necesario para el uso compartido Cuenta de AWS basado en una organización, no para el uso compartido basado en una organización.

1. Inicie sesión en la AWS consola https://console.aws.amazon.com/connect/con el administrador del lago de datos de la cuenta de consumidor con el nombre de usuario de IAM (el predeterminado es DatalakeAdminConsumer) y la contraseña especificados durante CloudFormation la creación de la pila.
2. En la AWS RAM consola, en el panel de navegación, en Shared with me, Resource shares, elija el recurso compartido Lake Formation. El estado debe ser Pendiente.
3. Elija Acción y Conceder.
4. Confirme los detalles del recurso y seleccione Aceptar recurso compartido.

   En este punto, el administrador del lago de datos de la cuenta del consumidor debería poder encontrar el recurso compartido en la consola de Lake Formation (https://console.aws.amazon.com/lakeformation/) en Data Catalog, Databases.

Crear un enlace de recursos para la tabla compartida

- Siga las instrucciones de Paso 3: Implementar el uso compartido entre cuentas mediante el método de control de acceso basado en etiquetas (paso 6) para crear un enlace de recursos para una tabla compartida. Asigne un nombre al enlace de recursos `amazon_reviews_table_named_resource_resource_link`. Cree el enlace de recursos en la base de datos `lakeformation_tutorial_cross_account_database_consumer`.

Conceder permiso de datos al consumidor sobre la tabla compartida

Para conceder permiso de datos sobre la tabla compartida al consumidor:

1. En la consola de Lake Formation (https://console.aws.amazon.com/lakeformation/), en Permisos, permisos de Data Lake, seleccione Otorgar.
2. En Entidades principales, seleccione los usuarios y roles de IAM y elija el usuario `DataAnalyst`.

3. Para las etiquetas LF o los recursos del catálogo, elija recursos del Catálogo de datos con nombre.
4. En Bases de datos elija la base de datos `lakeformation_tutorial_cross_account_database_named_resource`. Si no ve la base de datos en la lista desplegable, elija Cargar más.
5. En Tablas, seleccione la tabla `amazon_reviews_table_named_resource`.
6. En Permisos de tabla y columna, elija Seleccionar y Describir en Permisos de tabla.
7. Elija Conceder.

Conceder permiso de datos al consumidor sobre el enlace de recursos

Además de conceder al usuario del lago de datos permiso para acceder a la tabla compartida, también debe concederle permiso para acceder al enlace de recursos.

1. En la consola de Lake Formation (https://console.aws.amazon.com/lakeformation/), en Permisos, Permisos de Data Lake, seleccione Otorgar.
2. En Entidades principales, seleccione los usuarios y roles de IAM y elija el usuario `DataAnalyst`.
3. Para las etiquetas LF o los recursos del catálogo, elija recursos del Catálogo de datos con nombre.
4. En Bases de datos elija la base de datos `lakeformation_tutorial_cross_account_database_consumer`. Si no ve la base de datos en la lista desplegable, elija Cargar más.
5. En Tablas, seleccione la tabla `amazon_reviews_table_named_resource_resource_link`.
6. Para ver los permisos de enlace de recursos, seleccione Describir en Permisos de enlace de recursos.
7. Elija Conceder.

   En este punto, el usuario analista de datos de la cuenta de consumidor debería poder encontrar la base de datos y el enlace de recursos, y consultar la tabla compartida mediante la consola de Athena.

   Si no es así, confirme si lo siguiente está bien configurado:

   - Está creado el enlace de recursos para la tabla compartida

- Ha concedido al usuario acceso a la tabla compartida por la cuenta de productor
- Ha concedido al usuario acceso al enlace de recursos y a la base de datos para la que se ha creado el enlace de recursos

Paso 5: Limpiar AWS los recursos

Para evitar que se le cobren cargos no deseados Cuenta de AWS, puede eliminar los AWS recursos que utilizó para este tutorial.

1. Inicie sesión en la consola de Lake Formation https://console.aws.amazon.com/lakeformation/con la cuenta de productor y elimine o cambie lo siguiente:
   - AWS Resource Access Manager compartir recursos
   - Etiquetas de Lake Formation
   - CloudFormation pila
   - Ajustes de Lake Formation
   - AWS Glue Data Catalog
2. Inicie sesión en la consola de Lake Formation https://console.aws.amazon.com/lakeformation/con la cuenta de consumidor y elimine o cambie lo siguiente:
   - Etiquetas de Lake Formation
   - CloudFormation pila

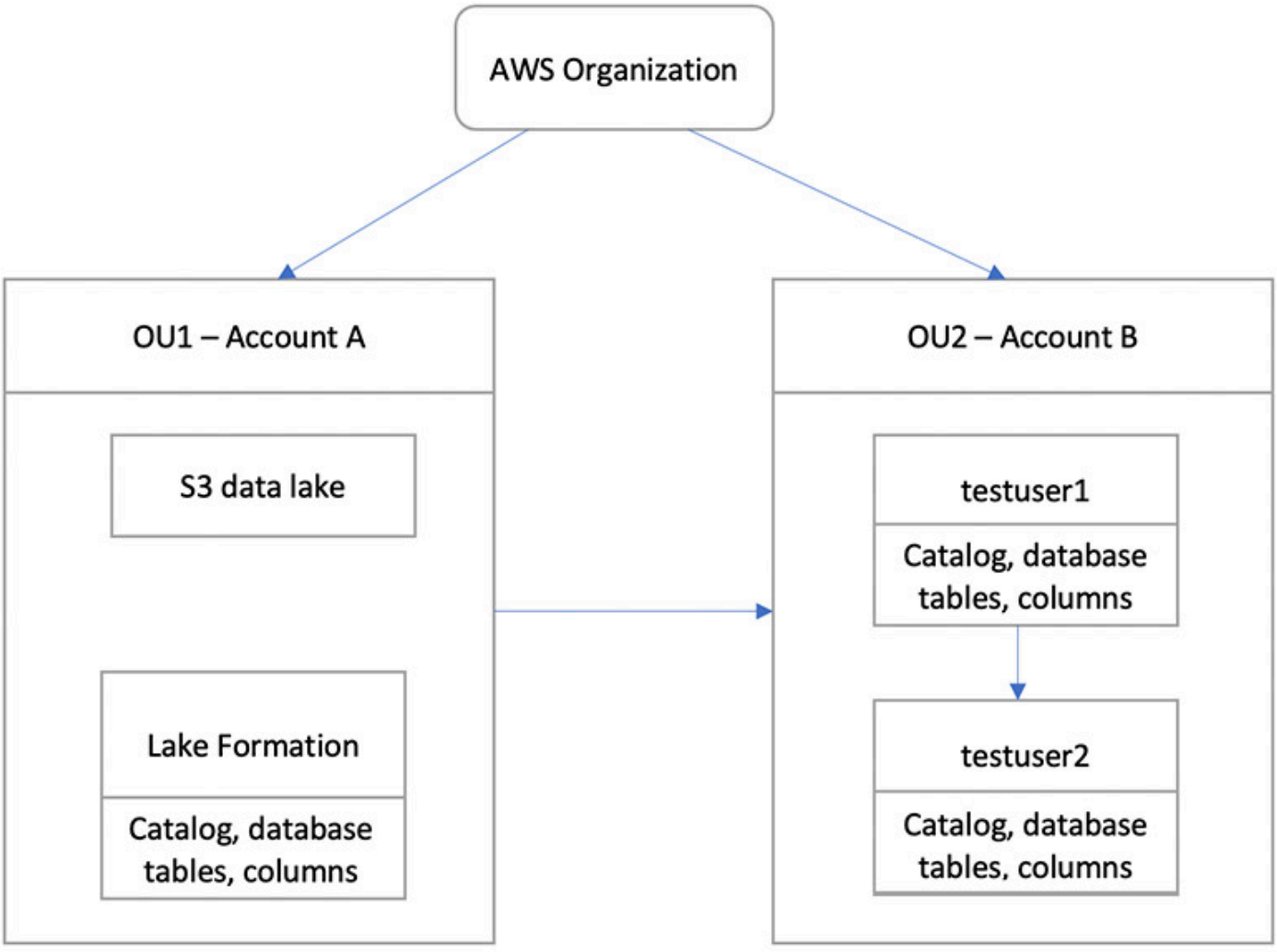
Compartir un lago de datos utilizando el control de acceso específico de Lake Formation

Este tutorial proporciona step-by-step instrucciones sobre cómo puede compartir rápida y fácilmente conjuntos de datos con Lake Formation al administrar varios Cuentas de AWS con AWS Organizations. La definición de permisos específicos sirve para controlar el acceso a datos confidenciales.

Los siguientes procedimientos también muestran cómo un administrador del lago de datos de la Cuenta A puede proporcionar acceso específico para la Cuenta B, y cómo un usuario de la Cuenta B, actuando como administrador de datos, puede conceder acceso específico a la tabla compartida para otros usuarios de su cuenta. Los administradores de datos de cada cuenta pueden delegar

independientemente el acceso a sus propios usuarios, lo que proporciona autonomía a cada equipo o línea de negocio (LOB).

El caso de uso asume que está utilizando AWS Organizations para administrar su Cuentas de AWS. El usuario de la cuenta A en una unidad organizativa (OU1) concede acceso a los usuarios de la cuenta B en OU2. Puede utilizar el mismo enfoque cuando no utilice organizaciones, como cuando solo tiene unas pocas cuentas. El siguiente diagrama ilustra el control de acceso específico de los conjuntos de datos en un lago de datos. El lago de datos está disponible en la Cuenta A. El administrador del lago de datos de la Cuenta A proporciona acceso específico para la Cuenta B. El diagrama también muestra que un usuario de la Cuenta B proporciona acceso a nivel de columna de la tabla del lago de datos de la Cuenta A a otro usuario de la Cuenta B.



Temas

- Destinatarios previstos

- Requisitos previos
- Paso 1: Proporcionar acceso específico a otra cuenta
- Paso 2: Proporcionar acceso específico a un usuario de la misma cuenta

Destinatarios previstos

Este tutorial está dirigido a administradores de datos, ingenieros de datos y analistas de datos. La tabla siguiente enumera los roles que se utilizan en este tutorial:

Rol	Description (Descripción)
Administrador de IAM	Usuario que tiene la política administrada de AWS : `AdministratorAccess` .
Administrador de lago de datos	Usuario que tiene la política AWS `AWSLakeFormationDataAdmin` gestionada: adjunto al rol.
Analista de datos	Usuario que tiene la política AWS gestionada: `AmazonAthenaFullAccess` adjunto.

Requisitos previos

Antes de empezar este tutorial, debe disponer de una Cuenta de AWS que pueda utilizar para iniciar sesión como usuario administrativo con los permisos correctos. Para obtener más información, consulte Completa la inicial AWS tareas de configuración.

En este tutorial, se supone que está familiarizado con IAM. Para obtener más información acerca de IAM, consulte la Guía del usuario de IAM.

Para este tutorial, necesita los siguientes recursos:

- Dos unidades organizativas:
  - OU1 — Contiene la cuenta A
  - OU2 — Contiene la cuenta B
- Una ubicación (bucket) del lago de datos de Amazon S3 en la Cuenta A.

- Un usuario administrador de lagos de datos en la cuenta A. Puede crear un administrador de lagos de datos mediante la consola de Lake Formation (https://console.aws.amazon.com/lakeformation/) o el `PutDataLakeSettings` funcionamiento de la API de Lake Formation.
- Lake Formation configurado en la Cuenta A, y la ubicación del lago de datos de Amazon S3 registrada con Lake Formation en la Cuenta A.
- Dos usuarios en la Cuenta B con las siguientes políticas administradas de IAM:
  - testuser1: tiene adjuntas las políticas AWS `AWSLakeFormationDataAdmin` administradas.
  - testuser2: tiene la política gestionada adjunta. AWS `AmazonAthenaFullAccess`
- Una base de datos testdb en la base de datos Lake Formation para la Cuenta B.

Paso 1: Proporcionar acceso específico a otra cuenta

Conozca cómo un administrador del lago de datos de la Cuenta A proporciona acceso específico a la Cuenta B.

Conceder acceso específico a otra cuenta

1. Inicie sesión Consola de administración de AWS https://console.aws.amazon.com/connect/en la cuenta A como administrador de un lago de datos.
2. Abre la consola de Lake Formation (https://console.aws.amazon.com/lakeformation/) y selecciona Comenzar.
3. en el panel de navegación, elija Bases de datos.
4. Elija Create database (Creación de base de datos).
5. En la sección de detalles de la base de datos, seleccione Base de datos.
6. En Nombre, introduzca un nombre (para este tutorial, utilizamos `sampledb01`).
7. Asegúrese de que no esté marcada la opción para utilizar solo el control de acceso IAM para las nuevas tablas de esta base de datos. Si se deja sin seleccionar, podremos controlar el acceso desde Lake Formation.
8. Elija Creación de base de datos.
9. En la página Bases de datos, elija su base de datos `sampledb01`.
10. En el menú Acciones, elija Conceder.
11. En la sección Conceder permisos, seleccione Cuenta externa.
12. Para el Cuenta de AWS ID o el ID de la AWS organización, introduzca el ID de cuenta de la cuenta B en OU2.

13. Para Tabla, elija la tabla a la que desea que tenga acceso la Cuenta B (para este post, utilizamos la tabla `acc_a_area`). De forma opcional, puede conceder acceso a columnas dentro de la tabla, como hacemos en esta publicación.
14. En Incluir columnas¸ elija las columnas a las que desea que tenga acceso la Cuenta B (para este post, concedemos permisos a tipo, nombre e identificadores).
15. En Columnas, seleccione Incluir columnas.
16. En Permisos de tabla, elija Seleccionar.
17. En Permisos concedibles, elija Seleccionar. Los permisos concedibles son necesarios para que los usuarios administradores de la Cuenta B puedan conceder permisos a otros usuarios de la Cuenta B.
18. Elija Conceder.
19. En el panel de navegación, elija Tablas.
20. Puedes ver una conexión activa en la sección Cuentas de AWS y AWS organizaciones con acceso.

Crear un recurso compartido

Los servicios integrados como Amazon Athena no pueden acceder directamente a las bases de datos o tablas de las cuentas. Por consiguiente, debe crear un enlace de recursos para que Athena pueda acceder a los enlaces de recursos de su cuenta a bases de datos y tablas de otras cuentas. Cree un enlace de recursos a la tabla (`acc_a_area`) para que los usuarios de la Cuenta B puedan consultar sus datos con Athena.

1. Inicia sesión en la AWS consola https://console.aws.amazon.com/connect/en la cuenta B como`testuser1`.
2. En la consola de Lake Formation (https://console.aws.amazon.com/lakeformation/), en el panel de navegación, elija Tablas. Debería ver las tablas a las que la Cuenta A ha proporcionado acceso.
3. Elija la tabla `acc_a_area`.
4. En el menú Acciones, elija Crear enlace de recursos.
5. En Nombre del enlace de recursos, introduzca un nombre (para este tutorial, `acc_a_area_rl`).
6. En Base de datos, elija su base de datos (`testdb`).
7. Seleccione Crear.
8. En el panel de navegación, elija Tablas.

9. Elija la tabla `acc_b_area_rl`.
10. En el menú Acciones, seleccione Ver datos.

    Se le redirige a la consola de Athena, donde debería ver la base de datos y la tabla.

    Ahora puede ejecutar una consulta en la tabla para ver el valor de la columna para la que se proporcionó acceso a testuser1 desde la Cuenta B.

Paso 2: Proporcionar acceso específico a un usuario de la misma cuenta

Esta sección muestra cómo un usuario de la cuenta B (`testuser1`), que actúa como administrador de datos, proporciona acceso específico a otro usuario de la misma cuenta (`testuser2`) al nombre de la columna de la tabla compartida `aac_b_area_rl`.

Conceder acceso específico a un usuario de la misma cuenta

1. Inicie sesión en la AWS consola https://console.aws.amazon.com/connect/en la cuenta B como`testuser1`.
2. En la consola de Lake Formation, en el panel de navegación, seleccione Tablas.

   Puede conceder permisos sobre una tabla a través de su enlace de recursos. Para ello, en la página Tablas, seleccione el enlace del recurso `acc_b_area_rl`, y en el menú Acciones, elija Conceder en destino.
3. En la sección Conceder permisos, seleccione Mi cuenta.
4. En Roles y usuarios de IAM, elija el usuario `testuser2`.
5. En Columna, elija el nombre de la columna.
6. En Permisos de tabla, elija Seleccionar.
7. Elija Conceder.

   Cuando cree un enlace de recursos, solo podrá verlo y acceder a él el usuario. Para permitir que otros usuarios de su cuenta accedan al enlace de recursos, debe conceder permisos sobre el propio enlace de recursos. Debe conceder los permisos DESCRIBIR o ELIMINAR. En la página Tablas, vuelva a seleccionar la tabla y, en el menú Acciones, Conceder.
8. En la sección Conceder permisos, seleccione Mi cuenta.
9. En Roles y usuarios de IAM, elija el usuario `testuser2`.
10. En Permisos de enlace de recursos, seleccione Describir.

11. Elija Conceder.
12. Inicia sesión en la AWS consola en la cuenta B como`testuser2`.

    En la consola de Athena (https://console.aws.amazon.com/athena/), debería ver la base de datos y la tabla. `acc_b_area_rl` Ahora puede ejecutar una consulta en la tabla para consultar el valor de la columna a la que `testuser2` tiene acceso.

Permisos de incorporación a Lake Formation

AWS Lake Formation utiliza el AWS Glue Data Catalog (Catálogo de datos) para almacenar metadatos para los lagos de datos de Amazon S3 y fuentes de datos externas, como Amazon Redshift, en forma de catálogos, bases de datos y tablas. Los metadatos del Catálogo de datos se organizan en una jerarquía de datos de tres niveles que incluye catálogos, bases de datos y tablas. Organiza los datos de diversos orígenes en contenedores lógicos denominados catálogos. Las bases de datos son colecciones de tablas. El Catálogo de datos también contiene enlaces a recursos, que son enlaces a bases de datos y tablas compartidas en cuentas externas, y se utilizan para el acceso entre cuentas a los datos del lago de datos. Cada AWS cuenta tiene un catálogo de datos por región. AWS

Lake Formation proporciona un modelo de permisos de sistema de administración de bases de datos relacionales (RDBMS) para conceder o revocar el acceso a catálogos, bases de datos, tablas y columnas en el Catálogo de datos con datos subyacentes en Amazon S3.

Antes de conocer los detalles del modelo de permisos de Lake Formation, es útil repasar los siguientes antecedentes:

- Los lagos de datos administrados por Lake Formation residen en ubicaciones designadas en Amazon Simple Storage Service (Amazon S3). El Catálogo de datos también contiene objetos de catálogo. Cada catálogo representa datos de orígenes como los almacenes de datos de Amazon Redshift, bases de datos Amazon DynamoDB y orígenes de datos de terceros, como Snowflake, MySQL y más de 30 orígenes de datos externos, que se integran mediante conectores federados.
- Lake Formation mantiene un Catálogo de datos que contiene metadatos sobre los orígenes de datos que se importarán en sus lagos de datos, como los datos de registros y bases de datos relacionales, y sobre los datos de sus lagos de datos en Amazon S3. El Catálogo de datos también contiene metadatos sobre datos de orígenes de datos externos distintos de Amazon S3. Los metadatos se organizan como catálogos, bases de datos y tablas. Las tablas de metadatos contienen el esquema, la ubicación, la partición y otra información sobre los datos que representan. Las bases de datos son colecciones de tablas.
- El Catálogo de datos de Lake Formation es el mismo Catálogo de datos utilizado por AWS Glue. Puede utilizar rastreadores AWS Glue para crear tablas del Catálogo de datos, y trabajos de extracción, transformación y carga (ETL) AWS Glue para poblar los datos subyacentes en sus lagos de datos.

- Los catálogos, bases de datos y tablas del Catálogo de datos se denominan recursos del Catálogo de datos. Las tablas del Catálogo de datos se denominan tablas de metadatos para distinguirlas de las tablas de los orígenes de datos o de los datos tabulares de Amazon S3. Los datos a los que apuntan las tablas de metadatos en Amazon S3 o en los orígenes de datos se denominan datos subyacentes.
- Un principal es un usuario o rol, un usuario o grupo de Amazon Quick, un usuario o grupo que se autentica con Lake Formation a través de un proveedor de SAML o, para el control de acceso entre cuentas, un ID de cuenta, un AWS ID de organización o un ID de unidad organizativa.
- AWS Gluelos rastreadores crean tablas de metadatos, pero también puedes crear tablas de metadatos manualmente con la consola de Lake Formation, la API o AWS Command Line Interface (AWS CLI). Al crear una tabla de metadatos, debe especificar una ubicación. Al crear una base de datos, la ubicación es opcional. Las ubicaciones de las tablas pueden ser ubicaciones de Amazon S3 o ubicaciones de orígenes de datos como una base de datos de Amazon Relational Database Service (Amazon RDS). Las ubicaciones de las bases de datos son siempre ubicaciones de Amazon S3.
- Los servicios que se integran con Lake Formation, como Amazon Athena y Amazon Redshift, pueden acceder al Catálogo de datos para obtener metadatos y comprobar la autorización para ejecutar consultas. Para obtener una lista completa de los servicios integrados, consulte AWS integraciones de servicios con Lake Formation.

Temas

- Descripción general de los permisos de Lake Formation
- Personas de Lake Formation y referencia de permisos IAM
- Cambiar la configuración predeterminada de su lago de datos
- Permisos implícitos de Lake Formation
- Referencia de permisos de Lake Formation
- Integración de IAM Identity Center
- Añadir una ubicación de Amazon S3 a su lago de datos
- Modo de acceso híbrido
- Creación de objetos en el AWS Glue Data Catalog
- Importación de datos mediante flujos de trabajo en Lake Formation

Descripción general de los permisos de Lake Formation

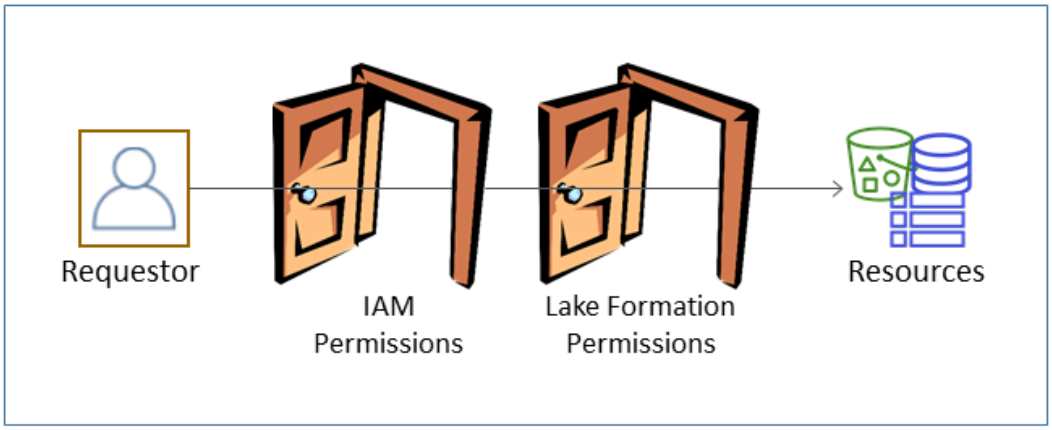
Hay dos tipos principales de permisos en AWS Lake Formation:

- Acceso a los metadatos. Permisos sobre los recursos del Catálogo de datos (Permisos sobre el Catálogo de datos).

  Con estos permisos las entidades principales pueden crear, leer, actualizar y eliminar bases de datos de metadatos y tablas del Catálogo de datos.
- Acceso a los datos subyacentes: permisos en ubicaciones en Amazon Simple Storage Service (Amazon S3) (permisos de acceso a datos y permisos de ubicación de datos).
  - Los permisos del lago de datos permiten a las entidades principales leer y escribir datos en las ubicaciones subyacentes de Amazon S3, datos a los que apuntan los recursos del Catálogo de datos.
  - Los permisos de ubicación de datos permiten a las entidades principales crear y modificar bases de datos y tablas de metadatos que apunten a ubicaciones específicas de Amazon S3.

Para ambas áreas, Lake Formation usa una combinación de permisos de Lake Formation y permisos AWS Identity and Access Management (IAM). El modelo de permisos IAM se compone de políticas de IAM. El modelo de permisos de Lake Formation se implementa como DBMS-style GRANT/REVOKE comandos, como`Grant SELECT on` *`tableName`* `to` *`userName`*.

Cuando una entidad principal efectúa una solicitud para acceder a los recursos del Catálogo de datos o a los datos subyacentes, para que la solicitud tenga éxito, debe pasar las comprobaciones de permisos tanto de IAM como de Lake Formation.



Los permisos de Lake Formation controlan el acceso a los recursos del Catálogo de datos, a las ubicaciones de Amazon S3 y a los datos subyacentes en dichas ubicaciones. Los permisos de IAM controlan el acceso a la Lake Formation y a los recursos y las API de AWS Glue. Así, aunque tenga el permiso Lake Formation para crear una tabla de metadatos en el Catálogo de datos (`CREATE_TABLE`), su operación fallará si no tiene el permiso IAM en la API `glue:CreateTable`. (¿Por qué un permiso `glue:`? porque Lake Formation usa el Catálogo de datos AWS Glue).

> **Note**
>
> Los permisos de Lake Formation se aplican solo en la Región en la que fueron concedidos.

AWS Lake Formation requiere que cada director (usuario o rol) esté autorizado a realizar acciones en los recursos gestionados por la Formación de Lagos. El administrador del lago de datos u otra entidad principal con permisos para conceder permisos de Lake Formation concede a la entidad principal las autorizaciones necesarias.

Cuando concede un permiso de Lake Formation a una entidad principal, puede conceder de forma opcional la capacidad de pasar ese permiso a otra entidad principal.

Puede usar la API de Lake Formation, la AWS Command Line Interface (AWS CLI) o las páginas de permisos de datos y ubicaciones de datos de la consola de Lake Formation para conceder y revocar los permisos de Lake Formation.

Métodos para el control de acceso específico

Con un lago de datos, el objetivo es tener un control de acceso detallado a los datos. En Lake Formation, esto significa un control de acceso específico a los recursos del Catálogo de datos y a las ubicaciones de Amazon S3. Puede lograr un control de acceso específico con uno de los siguientes métodos.

Método	Permisos de Lake Formation	Permisos de IAM	Comentarios
Método 1	Abra	Fine-grained	Este es el método predeterminado por compatibilidad inversa con AWS Glue.

Método	Permisos de Lake Formation	Permisos de IAM	Comentarios
			• Abierto significa que el permiso especial `Super` se concede al grupo `IAMAllowedPrincipals` , donde se crea automáticamente `IAMAllowedPrincipals` e incluye a todos los usuarios y roles de IAM que tienen permitido el acceso a los recursos de su Catálogo de datos por sus políticas de IAM, y el permiso `Super` permite a una entidad principal todas las operaciones de Lake Formation compatibles en la base de datos o tabla en la que se concede. Esto hace que el acceso a los recursos del Catálogo de datos y a las ubicaciones de Amazon S3 esté controlado únicamente por las políticas de IAM. Para obtener más información, consulte Cambiar la configuración predeterminada de su lago de datos y Actualización AWS Glue de los permisos de datos al AWS Lake Formation modelo. • Fine-grainedsignifica que las políticas de IAM controlan todo el acceso a los recursos del catálogo de datos y a los buckets individuales de Amazon S3. En la consola de Lake Formation, este método aparece como Utilizar solo control de acceso IAM.

Método	Permisos de Lake Formation	Permisos de IAM	Comentarios
Método 2	Fine-grained	Coarse-grained	Esta es la opción recomendada. • Fine-grainedacceso significa conceder permisos limitados de Lake Formation a directores individuales sobre los recursos del catálogo de datos, las ubicaciones de Amazon S3 y los datos subyacentes en esas ubicaciones. • Coarse-grainedse traduce en permisos más amplios para las operaciones individuales y para el acceso a las ubicaciones de Amazon S3. Por ejemplo, una política de IAM básica podría incluir `"glue:*"` o `"glue:CreateTables"` en lugar de `"glue:CreateTables"`, dejando que los permisos de Lake Formation controlen si una entidad principal puede o no crear objetos de catálogo. También significa dar a las entidades principales acceso a las API que necesitan para hacer su trabajo, pero bloqueando otras API y recursos. Por ejemplo, puede crear una política de IAM que permita a una entidad principal crear recursos del Catálogo de datos y crear y ejecutar flujos de trabajo, pero que no permita la creación de conexiones AWS Glue o funciones definidas por el usuario. Vea los ejemplos más adelante en esta sección.

Important

Tenga en cuenta lo siguiente:

- De forma predeterminada, Lake Formation tiene activada la configuración de control de acceso Utilizar solo IAM por compatibilidad con el comportamiento existente del Catálogo de datos AWS Glue. Le recomendamos que desactive esta configuración después de pasar a usar los permisos de Lake Formation. Para obtener más información, consulte Cambiar la configuración predeterminada de su lago de datos.
- Los administradores de lagos de datos y los creadores de bases de datos tienen permisos implícitos de Lake Formation que debe comprender. Para obtener más información, consulte Permisos implícitos de Lake Formation.

Control de acceso a los metadatos

Para el control de acceso a los recursos del Catálogo de datos, en el siguiente análisis se asume un control de acceso detallado con los permisos de Lake Formation y un control de acceso más detallado con las políticas de IAM.

Existen dos métodos distintos para la concesión de permisos de Lake Formation sobre los recursos del Catálogo de datos:

- Control de acceso a recursos con nombre. Con este método, el usuario concede permisos sobre bases de datos o tablas específicas especificando nombres de bases de datos o tablas. Las concesiones tienen esta forma:

  Conceda permisos a las entidades principales sobre los recursos [con opción de concesión].

  Con la opción de concesión, puede permitir que el beneficiario conceda los permisos a otras entidades principales.
- Tag-based control de acceso: con este método, se asignan una o más LF-Tags bases de datos, tablas y columnas del catálogo de datos y se conceden permisos sobre una o más de ellas LF-Tags a los directores. Cada uno LF-Tag es un par clave-valor, por ejemplo. `department=sales` Un principal LF-Tags que coincida con el de un recurso del LF-Tags catálogo de datos puede acceder a ese recurso. Este método se recomienda para los lagos de datos con un gran número de bases de datos y tablas. Se explica en detalle en Control de acceso basado en etiquetas de Lake Formation.

Los permisos que una entidad principal tiene sobre un recurso son la unión de los permisos concedidos por ambos métodos.

La siguiente tabla resume los permisos disponibles de Lake Formation en los recursos del Catálogo de datos. Los títulos de las columnas indican el recurso sobre el que se concede el permiso.

Catálogo	Base de datos	Tabla
CREATE_DATABASE	CREATE_TABLE	ALTER
	ALTER	DROP
	DROP	DESCRIBE
	DESCRIBE	SELECT*
		INSERT*
		DELETE*

Por ejemplo, el permiso CREATE_TABLE se concede sobre una base de datos. Esto significa que la entidad principal está autorizada a crear tablas en esa base de datos.

Los permisos con un asterisco (*) se conceden sobre los recursos del Catálogo de datos, pero se aplican a los datos subyacentes. Por ejemplo, el permiso DROP sobre una tabla de metadatos le permite eliminar la tabla del Catálogo de datos. Sin embargo, el permiso DELETE concedido sobre la misma tabla le permite eliminar los datos subyacentes de la tabla en Amazon S3, utilizando, por ejemplo, una instrucción SQL DELETE. Con estos permisos, también puede ver la tabla en la consola de Lake Formation y recuperar información sobre la tabla con la API AWS Glue. Así, SELECT, INSERT y DELETE son tanto permisos del Catálogo de datos como permisos de acceso a los datos.

Al conceder SELECT en una tabla, puede añadir un filtro que incluya o excluya una o más columnas. Esto permite un control de acceso específico sobre las columnas de la tabla de metadatos, limitando las columnas que los usuarios de los servicios integrados pueden ver al ejecutar consultas. Esta capacidad no está disponible solo con políticas de IAM.

También hay un permiso especial denominado Super. El permiso Super permite a una entidad principal efectuar todas las operaciones compatibles con Lake Formation en la base de datos o tabla sobre la que se concede. Este permiso puede coexistir con los demás permisos de Lake

Formation. Por ejemplo, puede conceder `Super`, `SELECT` y `INSERT` sobre una tabla de metadatos. La entidad principal puede efectuar todas las acciones compatibles en la tabla, y cuando revoca `Super`, permanecen los permisos `SELECT` y `INSERT`.

Para obtener más información sobre cada permiso, consulte Referencia de permisos de Lake Formation.

> **Important**
>
> Para consultar una tabla del Catálogo de datos creada por otro usuario, debe tener al menos un permiso de Lake Formation sobre la tabla. Si se le concede al menos un permiso sobre la tabla, también podrá ver la base de datos contenedora de la tabla.

Puede conceder o revocar los permisos del Catálogo de datos utilizando la consola de Lake Formation, la API o la AWS Command Line Interface (AWS CLI). El siguiente es un ejemplo de un AWS CLI comando que otorga al usuario `datalake_user1` permiso para crear tablas en la `retail` base de datos.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

A continuación, se muestra un ejemplo de política de IAM de control de acceso básico que complementa el control de acceso específico con los permisos de Lake Formation. Permite efectuar todas las operaciones sobre cualquier base de metadatos o tabla.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:*Database*",
                "glue:*Table*",
                "glue:*Partition*"
            ],
```

```
            "Resource": "*"
        }
    ]
}
```

El siguiente ejemplo también es básico, pero algo más restrictivo. Permite operaciones de solo lectura en todas las bases de datos de metadatos y tablas del Catálogo de datos de la cuenta y región designadas.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:GetDatabase",
                "glue:GetDatabases"
            ],
            "Resource": "arn:aws:glue:us-east-1:111122223333:*"
        }
    ]
}
```

Compare estas políticas con la siguiente política, que implementa un IAM-based control de acceso detallado. Concede permisos solo sobre un subconjunto de tablas de la base de datos de metadatos de gestión de relaciones con los clientes (CRM) en la cuenta y la región designadas.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
```

```
            "Action": [
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:GetDatabase",
                "glue:GetDatabases"
            ],
            "Resource": [
                "arn:aws:glue:us-east-1:111122223333:catalog",
                "arn:aws:glue:us-east-1:111122223333:database/CRM",
                "arn:aws:glue:us-east-1:111122223333:table/CRM/P*"
            ]
        }
    ]
}
```

Para más ejemplos de políticas de control de acceso básicas, consulte Personas de Lake Formation y referencia de permisos IAM.

Control de acceso a los datos subyacentes

Cuando un AWS servicio integrado solicita acceso a los datos en una ubicación de Amazon S3 cuyo acceso está controlado AWS Lake Formation, Lake Formation proporciona credenciales temporales para acceder a los datos.

Para permitir que Lake Formation controle el acceso a los datos subyacentes en una ubicación de Amazon S3, el usuario registra dicha ubicación con Lake Formation.

Después de registrar una ubicación de Amazon S3, puede comenzar a conceder los siguientes permisos de Lake Formation:

- Permisos de acceso a los datos (`SELECT`, `INSERT`, y `DELETE`)) en las tablas del Catálogo de datos que apuntan a esa ubicación.
- Permisos de ubicación de datos en esa ubicación.

Los permisos de ubicación de datos de Lake Formation controlan la capacidad de crear recursos del Catálogo de datos que apunten a determinadas ubicaciones de Amazon S3. Los permisos de ubicación de datos proporcionan una capa adicional de seguridad a las ubicaciones dentro del lago de datos. Cuando concede el permiso `CREATE_TABLE` o `ALTER` a una entidad principal, también

concede permisos de ubicación de datos para limitar las ubicaciones en las que la entidad principal puede crear o modificar tablas de metadatos.

Las ubicaciones de Amazon S3 son buckets o prefijos bajo un bucket, pero no objetos individuales de Amazon S3.

Puede conceder permisos de ubicación de datos a una entidad principal utilizando la consola de Lake Formation, la API o la AWS CLI. El formato general de una concesión es el siguiente:

```
grant DATA_LOCATION_ACCESS to principal on S3 location [with grant option]
```

Si incluye `with grant option`, el beneficiario puede conceder los permisos a otras entidades principales.

Recuerde que los permisos de Lake Formation siempre funcionan en combinación con los permisos AWS Identity and Access Management (IAM) para un control de acceso detallado. Para read/write los permisos sobre los datos subyacentes de Amazon S3, los permisos de IAM se conceden de la siguiente manera:

Cuando registra una ubicación, especifica una función de IAM que concede read/write permisos en esa ubicación. Lake Formation asume esa función al proporcionar credenciales temporales a los AWS servicios integrados. Un rol normal puede tener adjunta la siguiente política, en la que la ubicación registrada es el bucket `awsexamplebucket`.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket/*"
            ]
        },
```

```
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket"
            ]
        }
    ]
}
```

Lake Formation proporciona un rol vinculado al servicio que puede utilizar durante el registro para crear automáticamente políticas como esta. Para obtener más información, consulte Uso de roles vinculados a servicios para Lake Formation.

Por lo tanto, el registro de una ubicación de Amazon S3 concede los permisos `s3:` IAM necesarios sobre dicha ubicación, donde los permisos vienen especificados por el rol utilizado para registrar la ubicación.

> **Important**
>
> Evite registrar un bucket de Amazon S3 que tenga activada la opción El solicitante paga. Para los buckets registrados en Lake Formation, el rol utilizado para registrar el bucket se considera siempre como el solicitante. Si otra AWS cuenta accede al depósito, se le cobrará al propietario del depósito por el acceso a los datos si el rol pertenece a la misma cuenta que el propietario del depósito.

Para read/write acceder a los datos subyacentes, además de los permisos de Lake Formation, los directores también necesitan el permiso de `lakeformation:GetDataAccess` IAM. Con este permiso, Lake Formation concede la solicitud de credenciales temporales para acceder a los datos.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
```

```
            "Effect": "Allow",
            "Action": "lakeformation:GetDataAccess",
            "Resource": "*"
        }
    ]
}
```

En la política anterior, debe establecer el parámetro Resource en “*” (todos). No se admite la especificación de ningún otro recurso para este permiso. Esta configuración garantiza que Lake Formation pueda administrar el acceso a los datos en todo el entorno del lago de datos de manera eficiente.

> **Note**
>
> Amazon Athena requiere que el usuario cuente con el permiso `lakeformation:GetDataAccess`. Otros servicios integrados requieren que su rol de ejecución subyacente cuente con el permiso `lakeformation:GetDataAccess`.

Este permiso está incluido en las políticas sugeridas en el Personas de Lake Formation y referencia de permisos IAM.

En resumen, para permitir a las entidades principales de Lake Formation leer y escribir datos subyacentes con acceso controlado por los permisos de Lake Formation:

- Registre las ubicaciones de Amazon S3 que contienen los datos con Lake Formation.
- Las entidades principales creadoras de tablas del Catálogo de datos que apunten a ubicaciones de datos subyacentes deben tener permisos de ubicación de datos.
- Las entidades principales que leen y escriben datos subyacentes deben tener permisos de acceso a los datos de Lake Formation en las tablas del Catálogo de datos que apuntan a las ubicaciones de datos subyacentes.
- Las entidades principales que lean y escriban datos subyacentes deben tener el permiso `lakeformation:GetDataAccess` IAM cuando la ubicación de datos subyacente esté registrada en Lake Formation.

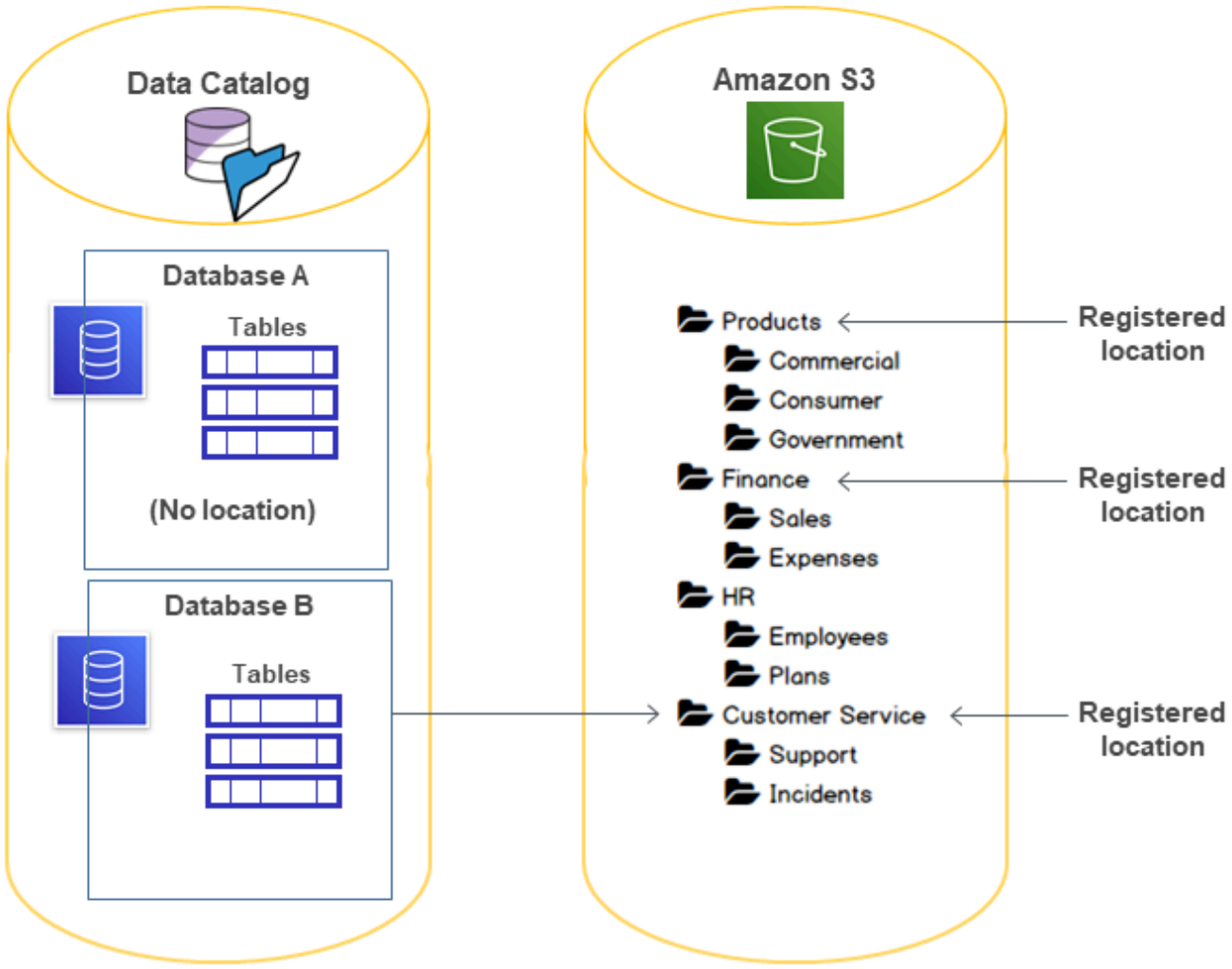
> **Note**
>
> El modelo de permisos de Lake Formation no impide el acceso a las ubicaciones de Amazon S3 a través de la API o la consola de Amazon S3 si tiene acceso a ellas a través de las políticas de IAM o Amazon S3. Puede adjuntar políticas de IAM a las entidades principales para bloquear este acceso.

Más información sobre los permisos de ubicación de datos

Los permisos de ubicación de datos rigen el resultado de las operaciones de creación y actualización en las bases de datos y tablas del Catálogo de datos. Las normas son las siguientes:

- Una entidad principal debe tener permisos explícitos o implícitos de ubicación de datos en una ubicación de Amazon S3 para crear o actualizar una base de datos o tabla que especifique dicha ubicación.
- El permiso explícito `DATA_LOCATION_ACCESS` se concede mediante la consola, la API o. AWS CLI
- Los permisos implícitos se conceden cuando una base de datos tiene una propiedad de ubicación que apunta a una ubicación registrada, la entidad principal tiene el permiso `CREATE_TABLE` en la base de datos y la entidad principal intenta crear una tabla en esa ubicación o en una ubicación secundaria.
- Si a una entidad principal se le conceden permisos de ubicación de datos en una ubicación, la entidad principal tendrá permisos de ubicación de datos en todas las ubicaciones secundarias.
- Un director no necesita permisos de ubicación de datos para realizar read/write operaciones con los datos subyacentes. Basta con tener los o permisos de acceso a los datos `SELECT` o `INSERT`. Los permisos de ubicación de datos se aplican solo a la creación de recursos del Catálogo de datos que apunten a la ubicación.
- Las operaciones de partición (`CreatePartitionBatchCreatePartition`,,`UpdatePartition`) no requieren permisos de ubicación de datos cuando la ubicación de la partición es secundaria de la ubicación de la tabla ya registrada en Lake Formation.

Piense en el escenario que se muestra en el siguiente diagrama.



En este diagrama:

- Los buckets de Amazon S3 `Products`, `Finance` y `Customer Service` están registrados en Lake Formation.
- `Database A` no tiene propiedad de ubicación, y `Database B` tiene una propiedad de ubicación que apunta al bucket `Customer Service`.
- El usuario `datalake_user` tiene `CREATE_TABLE` en ambas bases de datos.
- Se han concedido al usuario `datalake_user` permisos de ubicación de datos solo en el bucket `Products`.

A continuación se muestran los resultados cuando el usuario `datalake_user` intenta crear una tabla de catálogo en una base de datos concreta en una ubicación determinada.

Ubicación en la que **`datalake_user`** intenta crear una tabla

Base de datos y ubicación	Éxito o fracaso	Motivo
Base de datos A en `Finance/Sales`	Fracaso	Sin permiso de ubicación de datos
Base de datos A en `Products`	Correcto	Tiene permiso de ubicación de datos
Base de datos A en `HR/Plans`	Correcto	La ubicación no está registrada
Base de datos B en `Customer Service/Incidents`	Correcto	La base de datos tiene la propiedad de ubicación en `Customer Service`

Para obtener más información, consulte los siguientes temas:

- Añadir una ubicación de Amazon S3 a su lago de datos
- Referencia de permisos de Lake Formation
- Personas de Lake Formation y referencia de permisos IAM

Personas de Lake Formation y referencia de permisos IAM

Esta sección enumera algunas personas sugeridas de Lake Formation y sus permisos AWS Identity and Access Management (IAM) sugeridos. Para obtener información sobre los permisos de Lake Formation, consulte the section called “Referencia de permisos de Lake Formation”.

AWS Lake Formation personas

En la siguiente tabla se muestran las AWS Lake Formation personas sugeridas.

Personas de Lake Formation

Persona	Description (Descripción)
Administrador de IAM (superusuario)	(Obligatorio) Usuario que puede crear usuarios y roles de IAM. Tiene la política `AdministratorAccess` AWS gestionad

Persona	Description (Descripción)
	a. Tiene todos los permisos sobre todos los recursos de Lake Formation. Puede añadir administradores del lago de datos. No puede conceder permisos de Lake Formation si no ha sido designado también administrador del lago de datos.
Administrador del lago de datos	(Obligatorio) Usuario que puede registrar las ubicaciones de Amazon S3, acceder al catálogo de datos, crear bases de datos, crear y ejecutar flujos de trabajo, conceder permisos de Lake Formation a otros usuarios y ver los AWS CloudTrail registros . Tiene menos permisos IAM que el administrador IAM, pero suficientes para administrar el lago de datos. No se pueden añadir otros administradores del lago de datos.
Administrador de solo lectura	(Opcional) Usuario que puede ver entidades principales, recursos del Catálogo de datos, permisos y registros de AWS CloudTrail , sin permisos para las actualizaciones.
Ingeniero de datos	(Opcional) Usuario que puede crear bases de datos, crear y ejecutar rastreadores y flujos de trabajo, y conceder permisos de Lake Formation sobre las tablas del Catálogo de datos que crean los rastreadores y los flujos de trabajo. Le recomenda mos que convierta a todos los ingenieros de datos en creadores de bases de datos. Para obtener más información, consulte Creación de una base de datos.
Analista de datos	(Opcional) Usuario que puede ejecutar consultas en el lago de datos utilizando, por ejemplo, Amazon Athena. Solo tiene permisos suficientes para ejecutar consultas.
Rol de flujo de trabajo	(Obligatorio) Rol que ejecuta un flujo de trabajo en nombre de un usuario. Este rol se especifica cuando se crea un flujo de trabajo a partir de un esquema.

> **Note**
>
> En Lake Formation, los administradores de lagos de datos agregados después de la creación de la base de datos pueden conceder permisos, pero no tienen permisos de acceso a los datos, como SELECT o DESCRIBE, automáticamente. Los administradores que crean bases de datos reciben permisos SUPER en esas bases de datos. Este comportamiento es intencionado: si bien todos los administradores pueden concederse los permisos necesarios, estos permisos no se aplican automáticamente a los recursos preexistentes. Por lo tanto, los administradores deben concederse explícitamente el acceso a las bases de datos que existían antes de que se les asignaran los privilegios de administrador.

AWS políticas gestionadas para Lake Formation

Puede conceder los permisos AWS Identity and Access Management (de IAM) necesarios para trabajar con ellas AWS Lake Formation mediante políticas AWS gestionadas y políticas integradas. Las siguientes políticas AWS gestionadas están disponibles para Lake Formation.

AWS política gestionada: AWSLake FormationDataAdmin

AWSLakeFormationDataAdminla política otorga acceso administrativo AWS Lake Formation y servicios relacionados, como la administración AWS Glue de lagos de datos.

Puede asociar `AWSLakeFormationDataAdmin` a los usuarios, grupos y roles.

Detalles del permiso

- `CloudTrail`— Permite a los directores ver los AWS CloudTrail registros. Esto es necesario para revisar cualquier error en la configuración del lago de datos.
- `Glue`. Permite a las entidades principales ver, crear y actualizar tablas de metadatos y bases de datos en el Catálogo de datos. Esto incluye las operaciones de la API que comienzan con `Get`, `List`, `Create`, `Update`, `Delete` y `Search`. Esto es necesario para administrar los metadatos de las tablas del lago de datos.
- `IAM`. Permite a las entidades principales recuperar información sobre los usuarios y roles de IAM y las políticas asociadas a los roles. Esto es necesario para que el administrador de datos revise y enumere los usuarios y roles de IAM para conceder los permisos de Lake Formation.
- `Lake Formation`. Concede a los administradores de los lagos de datos los permisos necesarios de Lake Formation para administrar los lagos de datos.

- S3. Permite a las entidades principales recuperar información sobre los buckets de Amazon S3 y sus ubicaciones con el fin de establecer la ubicación de datos para los lagos de datos.

```
"Statement": [
        {
            "Sid": "AWSLakeFormationDataAdminAllow",
            "Effect": "Allow",
            "Action": [
                "lakeformation:*",
                "cloudtrail:DescribeTrails",
                "cloudtrail:LookupEvents",
                "glue:CreateCatalog",
  "glue:UpdateCatalog",
                "glue:DeleteCatalog",
  "glue:GetCatalog",
         "glue:GetCatalogs",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:CreateDatabase",
                "glue:UpdateDatabase",
                "glue:DeleteDatabase",
                "glue:GetConnections",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:CreateTable",
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetTableVersions",
                "glue:GetPartitions",
                "glue:GetTables",
                "glue:ListWorkflows",
                "glue:BatchGetWorkflows",
                "glue:DeleteWorkflow",
                "glue:GetWorkflowRuns",
                "glue:StartWorkflowRun",
                "glue:GetWorkflow",
                "s3:ListBucket",
                "s3:GetBucketLocation",
                "s3:ListAllMyBuckets",
                "s3:GetBucketAcl",
                "iam:ListUsers",
                "iam:ListRoles",
```

```
                "iam:GetRole",
                "iam:GetRolePolicy"
            ],
            "Resource": "*"
        },
        {
            "Sid": "AWSLakeFormationDataAdminDeny",
            "Effect": "Deny",
            "Action": [
                "lakeformation:PutDataLakeSettings"
            ],
                "Resource": "*"
        }
    ]
}
```

> **Note**
>
> La política `AWSLakeFormationDataAdmin` no concede todos los permisos necesarios a los administradores del lago de datos. Se necesitan permisos adicionales para crear y ejecutar flujos de trabajo y registrar ubicaciones con el rol vinculado al servicio `AWSServiceRoleForLakeFormationDataAccess`. Para obtener más información, consulte Crear un administrador de lago de datos y Uso de roles vinculados a servicios para Lake Formation.

AWS política gestionada: AWSLake FormationCrossAccountManager

AWSLakeFormationCrossAccountManagerla política proporciona acceso entre cuentas a AWS Glue los recursos a través de Lake Formation y otorga acceso de lectura a otros servicios requeridos, como AWS Organizations y AWS RAM.

Puede asociar `AWSLakeFormationCrossAccountManager` a los usuarios, grupos y roles.

Detalles del permiso

Esta política incluye los siguientes permisos.

- `Glue`. Permite a las entidades principales establecer o eliminar la política de recursos del Catálogo de datos para el control de acceso.

- `Organizations`. Permite a las entidades principales recuperar información sobre cuentas y unidades organizativas (UO) de una organización.
- `ram:CreateResourceShare`. Permite a las entidades principales crear un recurso compartido.
- `ram:UpdateResourceShare`. Permite a las entidades principales modificar algunas propiedades del recurso compartido especificado.
- `ram:DeleteResourceShare`. Permite a las entidades principales eliminar el recurso compartido especificado.
- `ram:AssociateResourceShare`. Permite a las entidades principales añadir la lista especificada de entidades principales y la lista de recursos a un recurso compartido.
- `ram:DisassociateResourceShare`. Permite a las entidades principales eliminar a las entidades principales o recursos especificados de la participación en el recurso compartido especificado.
- `ram:GetResourceShares`. Permite a las entidades principales recuperar detalles sobre los recursos compartidos que le pertenecen o que se comparten.
- `ram:RequestedResourceType`. Permite a las entidades principales recuperar el tipo de recurso (base de datos, tabla o catálogo).
- `AssociateResourceSharePermission`— Permite a los directores añadir o reemplazar el AWS RAM permiso para un tipo de recurso incluido en un recurso compartido. Puede tener exactamente un permiso asociado a cada tipo de recurso en el recurso compartido.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [{
            "Sid": "AllowCreateResourceShare",
            "Effect": "Allow",
            "Action": [
                "ram:CreateResourceShare"
            ],
            "Resource": "*",
            "Condition": {
                "StringLikeIfExists": {
                    "ram:RequestedResourceType": [
                        "glue:Table",
                        "glue:Database",
```

```
                    "glue:Catalog"
                ]
            }
        }
    },
    {
        "Sid": "AllowManageResourceShare",
        "Effect": "Allow",
        "Action": [
            "ram:UpdateResourceShare",
            "ram:DeleteResourceShare",
            "ram:AssociateResourceShare",
            "ram:DisassociateResourceShare",
            "ram:GetResourceShares"
        ],
        "Resource": "*",
        "Condition": {
            "StringLike": {
                "ram:ResourceShareName": [
                    "LakeFormation*"
                ]
            }
        }
    },
    {
        "Sid": "AllowManageResourceSharePermissions",
        "Effect": "Allow",
        "Action": [
            "ram:AssociateResourceSharePermission"
        ],
        "Resource": "*",
        "Condition": {
            "ArnLike": {
                "ram:PermissionArn": [
                    "arn:aws:ram::aws:permission/AWSRAMLFEnabled*"
                ]
            }
        }
    },
    {
        "Sid": "AllowXAcctManagerPermissions",
        "Effect": "Allow",
        "Action": [
            "glue:PutResourcePolicy",
```

```
                "glue:DeleteResourcePolicy",
                "organizations:DescribeOrganization",
                "organizations:DescribeAccount",
                "ram:Get*",
                "ram:List*"
            ],
            "Resource": "*"
        },
        {
            "Sid": "AllowOrganizationsPermissions",
            "Effect": "Allow",
            "Action": [
                "organizations:ListRoots",
                "organizations:ListAccountsForParent",
                "organizations:ListOrganizationalUnitsForParent"
            ],
            "Resource": "*"
        }
    ]
}
```

AWS política gestionada: AWSGlue ConsoleFullAccess

AWSGlueConsoleFullAccessla política otorga acceso total a AWS Glue los recursos cuando una identidad a la que está asociada la política utiliza la Consola de administración de AWS. Si sigue la convención de nomenclatura para los recursos especificados en esta política, los usuarios dispondrán de todas las funciones de la consola. Esta política suele estar asociada a los usuarios de la AWS Glue consola.

Además, AWS Glue Lake Formation asume la función de servicio `AWSGlueServiceRole` para permitir el acceso a servicios relacionados, incluidos Amazon Elastic Compute Cloud (Amazon EC2), Amazon Simple Storage Service (Amazon S3) y Amazon. CloudWatch

AWS managed policy:LakeFormationDataAccessServiceRolePolicy

Esta política está asociada a un rol vinculado a servicio denominado `ServiceRoleForLakeFormationDataAccess` que permite a dicho servicio realizar acciones en recursos previa solicitud por su parte. No puede adjuntar esta política a identidades de IAM.

Esta política permite a los AWS servicios integrados de Lake Formation, como Amazon Athena Amazon Redshift, utilizar la función vinculada al servicio para descubrir los recursos de Amazon S3.

Para obtener más información, consulte, Uso de roles vinculados a servicios para Lake Formation.

Detalles del permiso

Esta política incluye el siguiente permiso.

- `s3:ListAllMyBuckets`: devuelve una lista de todos los buckets pertenecientes al remitente autenticado de la solicitud.

JSON

```
{
 "Version":"2012-10-17",
 "Statement": [
  {
   "Sid": "LakeFormationDataAccessServiceRolePolicy",
   "Effect": "Allow",
   "Action": [
    "s3:ListAllMyBuckets"
   ],
   "Resource": [
    "arn:aws:s3:::*"
   ]
  }
 ]
}
```

Lake Formation actualiza sus políticas AWS gestionadas

Consulta los detalles sobre las actualizaciones de las políticas AWS gestionadas de Lake Formation desde que este servicio comenzó a rastrear estos cambios.

Cambio	Descripción	Fecha
Política `AWSLakeFormationCr`	Lake Formation mejoró la AWSLakeFormationCrossAccountManagerpolítica al	Enero de 2025

Cambio	Descripción	Fecha
`ossAccountManager` actualizada de Lake Formation.	reemplazar el operador de `StringLik e` condición por el `ArnLike` operador que permite a IAM realizar la verificac ión del formato ARN.	
Política `AWSLakeFo rmationDataAdmin` actualizada de Lake Formation.	Lake Formation mejoró la AWSLakeFo rmationDataAdminpolítica al agregar el siguiente AWS Glue Data Catalog CRUD APIs como parte de la función de catálogo múltiple. • pegamento: CreateCatalog • pegamento: UpdateCatalog • pegamento: DeleteCatalog • pegamento: GetCatalog • pegamento: GetCatalogs Este cambio en la política administrada tiene por objeto garantizar que el perfil de administrador de Lake Formation cuente de forma predeterminada con el permiso de IAM para estas nuevas operaciones.	Diciembre de 2024
Política `AWSLakeFo rmationCr ossAccountManager` actualizada de Lake Formation.	Lake Formation mejoró la AWSLakeFo rmationCrossAccountManagerpolítica al agregar elementos Sid a la declaración de política.	Marzo de 2024
Política `AWSLakeFo rmationDataAdmin` actualizada de Lake Formation.	Lake Formation mejoró la AWSLakeFo rmationDataAdminpolítica al agregar un elemento Sid a la declaración de política y eliminar una acción redundante.	Marzo de 2024

Cambio	Descripción	Fecha
Política `LakeFormationDataAccessServiceRolePolicy` actualizada de Lake Formation.	Lake Formation mejoró la LakeFormationDataAccessServiceRolePolicypolítica al agregar un elemento Sid a la declaración de política.	Febrero de 2024
Política `AWSLakeFormationCrossAccountManager` actualizada de Lake Formation.	Lake Formation mejoró la AWSLakeFormationCrossAccountManagerpolítica al agregar un nuevo permiso para permitir el intercambio de datos entre cuentas en modo de acceso híbrido.	Octubre de 2023
Política `AWSLakeFormationCrossAccountManager` actualizada de Lake Formation.	Lake Formation mejoró la AWSLakeFormationCrossAccountManagerpolítica para crear solo un recurso compartid o por cuenta de destinatario cuando el recurso se comparte por primera vez. Todos los recursos compartidos a partir de entonces con la misma cuenta se adjuntan al mismo recurso compartido.	6 de mayo de 2022
Lake Formation inició el seguimiento de los cambios.	Lake Formation comenzó a rastrear los cambios en sus políticas AWS gestionadas.	6 de mayo de 2022

Permisos sugeridos para las personas

Los siguientes son los permisos sugeridos para cada persona. El administrador de IAM no está incluido porque ese usuario tiene todos los permisos sobre todos los recursos.

Temas

- Permisos de administrador del lago de datos
- Permisos de administrador de solo lectura

- Permisos de ingeniero de datos
- Permisos de analista de datos
- Permisos del rol de flujo de trabajo

Permisos de administrador del lago de datos

> **Important**
>
> En las siguientes políticas, *<account-id>* sustitúyalo por un número de AWS cuenta válido y *<workflow_role>* sustitúyelo por el nombre de un rol que tenga permisos para ejecutar un flujo de trabajo, tal y como se define enPermisos del rol de flujo de trabajo.

Tipo de política	Política
AWS políticas gestionadas	• `AWSLakeFormationDataAdmin` • `LakeFormationDataAccessServiceRolePolicy` (política de rol vinculado a servicio) • `AWSGlueConsoleFullAccess` (Opcional) • `CloudWatchLogsReadOnlyAccess` (Optional) • `AWSLakeFormationCrossAccountManager` (Optional) • `AmazonAthenaFullAccess` (opcional) Para obtener información sobre las políticas AWS administradas opcionales, consultethe section called “Crear un administrador de lago de datos”.
Política insertada (para crear el rol vinculado al servicio de Lake Formation)	`{ "Version": "2012-10-17", "Statement": [ { "Effect": "Allow", "Action": "iam:CreateServiceLinkedRole", "Resource": "*",`

Tipo de política	Política

	<pre> "Condition": {
	                "StringEquals": {
	                    "iam:AWSServiceName": "lakeform
	ation.amazonaws.com"
	                }
	            }
	        },
	        {
	            "Effect": "Allow",
	            "Action": [
	                "iam:PutRolePolicy"
	            ],
	            "Resource": "arn:aws:iam:: <account-
	id> :role/aws-service-role/lakeformation.amazonaw
	s.com/AWSServiceRoleForLakeFormationDataAccess"
	        }
	    ]
	}</pre>

(Opcional) Política insertada (política de passrole para el rol de flujo de trabajo). Esto solo es necesario si el administrador del lago de datos crea y ejecuta flujos de trabajo.	<pre>{
	    "Version":"2012-10-17",
	    "Statement": [
	        {
	            "Sid": "PassRolePermissions",
	            "Effect": "Allow",
	            "Action": [
	                "iam:PassRole"
	            ],
	            "Resource": [
	                "arn:aws:iam:: 111122223
	333 :role/<workflow_role> "
	            ]
	        }
	    ]
	}</pre>

Tipo de política	Política

(Opcional) Política en línea (si su cuenta concede o recibe permisos entre cuentas de Lake Formation). Esta política sirve para aceptar o rechazar las invitaciones para compartir AWS RAM recursos y para permitir la concesión de permisos entre cuentas a las organizaciones. `ram:EnableSharingWithAwsOrganization` solo es obligatorio para los administradores del lago de datos de la cuenta de AWS Organizations administración.	{
	"Version":"2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": [
	"ram:AcceptResourceShareInvitation",
	"ram:RejectResourceShareInvitation",
	"ec2:DescribeAvailabilityZones",
	"ram:EnableSharingWithAwsOrganization"
	],
	"Resource": "*"
	}
	]
	}

Permisos de administrador de solo lectura

Tipo de política	Política
Política insertada (básica)	{ "Version":"2012-10-17", "Statement":[{ "Effect":"Allow", "Action":["lakeformation:GetEffectivePermissionsForPath", "lakeformation:ListPermissions", "lakeformation:ListDataCellsFilter", "lakeformation:GetDataCellsFilter", "lakeformation:SearchDatabasesByLFTags",

Tipo de política	Política

```
            "lakeformation:SearchTablesByLFTags"
,
            "lakeformation:GetLFTag",
            "lakeformation:ListLFTags",
            "lakeformation:GetResourceLFTags",
            "lakeformation:ListLakeFormationOpti
ns",
            "cloudtrail:DescribeTrails",
            "cloudtrail:LookupEvents",
            "glue:GetDatabase",
            "glue:GetDatabases",
            "glue:GetConnections",
            "glue:SearchTables",
            "glue:GetTable",
            "glue:GetTableVersions",
            "glue:GetPartitions",
            "glue:GetTables",
            "glue:GetWorkflow",
            "glue:ListWorkflows",
            "glue:BatchGetWorkflows",
            "glue:GetWorkflowRuns",
            "glue:GetWorkflow",
            "s3:ListBucket",
            "s3:GetBucketLocation",
            "s3:ListAllMyBuckets",
            "s3:GetBucketAcl",
            "iam:ListUsers",
            "iam:ListRoles",
            "iam:GetRole",
            "iam:GetRolePolicy"
         ],
         "Resource":"*"
      },
      {
         "Effect":"Deny",
         "Action":[
            "lakeformation:PutDataLakeSettings"
         ],
         "Resource":"*"
      }
   ]
```

Tipo de política	Política
	`}`

Permisos de ingeniero de datos

> **Important**
>
> En las siguientes políticas, *<account-id>* sustitúyalo por un número de AWS cuenta válido y *<workflow_role>* sustitúyalo por el nombre del rol del flujo de trabajo.

Tipo de política	Política
AWS política gestionada	`AWSGlueConsoleFullAccess`
Política insertada (básica)	`{ "Version":"2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "lakeformation:GrantPermissions", "lakeformation:RevokePermissions", "lakeformation:BatchGrantPermissions", "lakeformation:BatchRevokePermissions", "lakeformation:ListPermissions", "lakeformation:AddLFTagsToResource", "lakeformation:RemoveLFTagsFromResource", "lakeformation:GetResourceLFTags", "lakeformation:ListLFTags", "lakeformation:GetLFTag", "lakeformation:SearchTablesByLFTags",`

Tipo de política	Política

	<pre>"lakeformation:SearchDatabasesByLFTa
	gs",
	"lakeformation:GetWorkUnits",
	"lakeformation:GetWorkUnitResults",
	"lakeformation:StartQueryPlanning",
	"lakeformation:GetQueryState",
	"lakeformation:GetQueryStatistics"
	],
	"Resource": "*"
	}
	]
	}</pre>

Política insertada (para las operaciones en tablas gobernadas, incluidas las operaciones dentro de las transacciones)	<pre>{
	"Version":"2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": [
	"lakeformation:StartTransaction",
	"lakeformation:CommitTransaction",
	"lakeformation:CancelTransaction",
	"lakeformation:ExtendTransaction",
	"lakeformation:DescribeTransaction",
	"lakeformation:ListTransactions",
	"lakeformation:GetTableObjects",
	"lakeformation:UpdateTableObjects",
	"lakeformation:DeleteObjectsOnCancel
	"
	],
	"Resource": "*"
	}
	]
	}</pre>

Tipo de política	Política

Política insertada (para el control de acceso a metadatos mediante el método de control de acceso basado en etiquetas de Lake Formation (LF-TBAC))	{
	"Version":"2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": [
	"lakeformation:AddLFTagsToResource",
	"lakeformation:RemoveLFTagsFromResou
	rce",
	"lakeformation:GetResourceLFTags",
	"lakeformation:ListLFTags",
	"lakeformation:GetLFTag",
	"lakeformation:SearchTablesByLFTags"
	,
	"lakeformation:SearchDatabasesByLFTa
	gs"
	],
	"Resource": "*"
	}
	]
	}

Política insertada (política de passrole para el rol de flujo de trabajo)	{
	"Version":"2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: *111122223
	333* :role/*<workflow_role>* "
	]
	}
	]
	}

Permisos de analista de datos

Tipo de política	Política
AWS política gestionada	`AmazonAthenaFullAccess`
Política insertada (básica)	`{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "glue:GetTable", "glue:GetTables", "glue:SearchTables", "glue:GetDatabase", "glue:GetDatabases", "glue:GetPartitions", "lakeformation:GetResourceLFTags", "lakeformation:ListLFTags", "lakeformation:GetLFTag", "lakeformation:SearchTablesByLFTags", "lakeformation:SearchDatabasesByLFTags"], "Resource": "*" }] }`
(Opcional) Política insertada (para las operaciones en tablas gobernadas, incluidas las operaciones dentro de las transacciones)	`{ "Version":"2012-10-17", "Statement": [ { "Effect": "Allow", "Action": [ "lakeformation:StartTransaction", "lakeformation:CommitTransaction", "lakeformation:CancelTransaction", "lakeformation:ExtendTransaction",`

Tipo de política	Política
	"lakeformation:DescribeTransaction", "lakeformation:ListTransactions", "lakeformation:GetTableObjects", "lakeformation:UpdateTableObjects", "lakeformation:DeleteObjectsOnCancel "], "Resource": "*" }] }

Permisos del rol de flujo de trabajo

Este rol cuenta con los permisos necesarios para ejecutar un flujo de trabajo. Se especifica un rol con estos permisos cuando se crea un flujo de trabajo.

> **Important**
>
> En las siguientes políticas, *<region>* sustitúyalo por un identificador de AWS región válido (por ejemplous-east-1), *<account-id>* por un número de AWS cuenta válido, *<workflow_role>* por el nombre del rol del flujo de trabajo y *<your-s3-cloudtrail-bucket>* por la ruta de Amazon S3 a tus AWS CloudTrail registros.

Tipo de política	Política
AWS política gestionada	`AWSGlueServiceRole`
Política insertada (acceso a datos)	{ "Version": "2012-10-17", "Statement": [{ "Sid": "Lakeformation", "Effect": "Allow", "Action": ["lakeformation:GetDataAccess",

Tipo de política	Política

	<pre>"lakeformation:GrantPermissions"
	],
	"Resource": "*"
	}
	]
	}</pre>

Política insertada (política de passrole para el rol de flujo de trabajo)	<pre>{
	"Version":"2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: 111122223333 :role/<workflow_role> "
	]
	}
	]
	}</pre>

Política en línea (para ingerir datos fuera del lago de datos, por ejemplo, AWS CloudTrail registros)	<pre>{
	"Version":"2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": ["s3:GetObject", "s3:ListBucket"],
	"Resource": ["arn:aws:s3::: <your-s3-cloudtrail-bucket> /*"]
	}
	]
	}</pre>

Cambiar la configuración predeterminada de su lago de datos

Para mantener la compatibilidad con versiones anterioresAWS Glue, AWS Lake Formation tiene la siguiente configuración de seguridad inicial:

- El permiso `Super` se concede al grupo `IAMAllowedPrincipals` sobre todos los recursos existentes del Catálogo de datos de AWS Glue.
- La configuración "Usar solo control de acceso de IAM" está activada para los nuevos recursos del Catálogo de datos.

De hecho, esta configuración hace que el acceso a los recursos del catálogo de datos y a las ubicaciones de Amazon S3 se controle únicamente mediante políticas AWS Identity and Access Management (IAM). Los permisos individuales de Lake Formation no están en vigor.

El grupo `IAMAllowedPrincipals` incluye a todos los usuarios y roles de IAM a los que sus políticas de IAM permiten el acceso a los recursos de su Catálogo de datos. El permiso `Super` permite a una entidad principal efectuar todas las operaciones compatibles con Lake Formation en la base de datos o tabla sobre la que se ha concedido.

Para cambiar la configuración de seguridad de modo que el acceso a los recursos del Catálogo de datos (bases de datos y tablas) sea administrado por los permisos de Lake Formation, haga lo siguiente:

1. Cambie la configuración de seguridad predeterminada para los nuevos recursos. Para obtener instrucciones, consulte Cambio del modelo de permisos predeterminado o uso del modo de acceso híbrido.
2. Cambiar la configuración de los recursos existentes del Catálogo de datos. Para obtener instrucciones, consulte Actualización AWS Glue de los permisos de datos al AWS Lake Formation modelo.

Cambiar la configuración de seguridad predeterminada mediante la operación de la API **`PutDataLakeSettings`** de Lake Formation

También puede cambiar la configuración de seguridad predeterminada mediante la operación de la PutDataLakeSettingsAPI Lake Formation. Esta acción toma como argumentos un identificador de catálogo opcional y una DataLakeSettingsestructura.

Para imponer el control de acceso a los metadatos y a los datos subyacentes por parte de Lake Formation en las nuevas bases de datos y tablas, codifique la estructura `DataLakeSettings` de la manera siguiente.

> **Note**
>
> *<AccountID>*Sustitúyalo por un identificador de AWS cuenta válido y *<Username>* por un nombre de usuario de IAM válido. Puede especificar más de un usuario como administrador del lago de datos.

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ],
        "CreateDatabaseDefaultPermissions": [],
        "CreateTableDefaultPermissions": []
    }
}
```

También puede codificar la estructura de la siguiente manera. Omitir el parámetro `CreateDatabaseDefaultPermissions` o `CreateTableDefaultPermissions` equivale a pasar una lista vacía.

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ]
    }
}
```

Esta acción revoca efectivamente todos los permisos de Lake Formation del grupo `IAMAllowedPrincipals` en las nuevas bases de datos y tablas. Cuando cree una base de datos, puede anular esta configuración.

Para imponer el control de acceso a los metadatos y a los datos subyacentes solo mediante IAM en las nuevas bases de datos y tablas, codifique la estructura de `DataLakeSettings` de la manera siguiente.

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ],
        "CreateDatabaseDefaultPermissions": [
            {
                "Principal": {
                    "DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"
                },
                "Permissions": [
                    "ALL"
                ]
            }
        ],
        "CreateTableDefaultPermissions": [
            {
                "Principal": {
                    "DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"
                },
                "Permissions": [
                    "ALL"
                ]
            }
        ]
    }
}
```

Esto concede al grupo `IAMAllowedPrincipals` el permiso `Super` de Lake Formation sobre nuevas bases de datos y tablas. Cuando cree una base de datos, puede anular esta configuración.

> **Note**
>
> En la estructura `DataLakeSettings` anterior, el único valor permitido para `DataLakePrincipalIdentifier` es `IAM_ALLOWED_PRINCIPALS`, y el único valor permitido para `Permissions` es `ALL`.

Permisos implícitos de Lake Formation

AWS Lake Formation concede los siguientes permisos implícitos a los administradores de lagos de datos, a los creadores de bases de datos y a los creadores de tablas.

Administrador de lago de datos

- Tener acceso de `Describe` a todos los recursos del Catálogo de datos excepto a los recursos compartidos desde otra cuenta directamente a otra entidad principal. Este acceso no puede ser revocado por un administrador.
- Disponga de permisos de ubicación de datos en todas las partes del lago de datos.
- Puede conceder o revocar el acceso a cualquier recurso del Catálogo de datos a cualquier entidad principal (incluida la propia). Este acceso no puede ser revocado por un administrador.
- Puede crear bases de datos en el Catálogo de datos.
- Puede conceder el permiso para crear una base de datos a otro usuario.

> **Note**
>
> Los administradores del lago de datos pueden registrar ubicaciones de Amazon S3 solo si disponen de permisos de IAM para hacerlo. Las políticas del administrador del lago de datos sugeridas en esta guía conceden esos permisos. Además, los administradores de los lagos de datos no tienen permisos implícitos para eliminar bases de datos o alter/drop tablas creadas por otros. Sin embargo, pueden concederse a sí mismos permisos para hacerlo.

Para obtener más información sobre los administradores del lago de datos, consulte Crear un administrador de lago de datos.

Creadores de catálogos

- Tienen todos los permisos de catálogo en los catálogos que crean, tienen permisos en las bases de datos y tablas que crean en el catálogo y pueden conceder permisos a otros directores de la misma AWS cuenta para crear bases de datos y tablas en el catálogo. Un creador de catálogos que también tenga la política `AWSLakeFormationCrossAccountManager` AWS administrada puede conceder permisos sobre el catálogo a otras AWS cuentas u organizaciones.

  Los administradores del lago de datos pueden utilizar la consola o la API de Lake Formation para designar a los creadores de catálogos.

  > **Note**
  >
  > Los creadores de catálogos no tienen permisos implícitos en las bases de datos y tablas que otros crean en el catálogo.

Para obtener más información sobre la creación de catálogos, consulte Llevar sus datos al AWS Glue Data Catalog.

Creadores de bases de datos

- Tienen todos los permisos de base de datos en las bases de datos que crean, tienen permisos en las tablas que crean en la base de datos y pueden conceder permisos a otros directores de la misma AWS cuenta para crear tablas en la base de datos. Un creador de bases de datos que también tenga la política `AWSLakeFormationCrossAccountManager` AWS administrada puede conceder permisos sobre la base de datos a otras AWS cuentas u organizaciones.

  Los administradores del lago de datos pueden utilizar la consola de Lake Formation o la API para designar a los creadores de las bases de datos.

  > **Note**
  >
  > Los creadores de bases de datos no tienen permisos implícitos sobre las tablas que otros crean en la base de datos.

Para obtener más información, consulte Creación de una base de datos.

Creadores de tablas

- Tienen todos los permisos sobre las tablas que crean.

- Pueden conceder permisos sobre todas las tablas que creen a las entidades principales de la misma cuenta AWS .
- Puede conceder permisos en todas las tablas que cree a otras AWS cuentas u organizaciones si disponen de la política `AWSLakeFormationCrossAccountManager` AWS gestionada.
- Pueden ver las bases de datos que contienen las tablas que crean.

Referencia de permisos de Lake Formation

Para realizar AWS Lake Formation las operaciones, los directores necesitan tanto los permisos de Lake Formation como los permisos AWS Identity and Access Management (IAM). En general, los permisos IAM se conceden mediante políticas de control de acceso poco específicas, como se describe en the section called “Descripción general de los permisos de Lake Formation”. Puede conceder permisos a Lake Formation mediante la consola, la API o AWS Command Line Interface (AWS CLI).

Para saber cómo conceder o revocar permisos de Lake Formation, consulte the section called “Concesión de permisos de datos” y the section called “Conceder permisos de ubicación de datos”.

> **Note**
>
> Los ejemplos de esta sección muestran cómo conceder permisos a entidades principales en la misma cuenta AWS . Para ver ejemplos de concesiones entre cuentas, consulte the section called “Cross-account intercambio de datos”.

Permisos de Lake Formation por tipo de recurso

A continuación se indican los permisos válidos de Lake Formation disponibles para cada tipo de recurso:

Recurso	Permiso
`Catalog`	`ALL` (`Super`), superusuario
	`ALTER`
	`CREATE_DATABASE`

Recurso	Permiso
	DESCRIBE
	DROP
Database	ALL (Super)
	ALTER
	CREATE_TABLE
	DESCRIBE
	DROP
Table	ALL (Super)
	ALTER
	DELETE
	DESCRIBE
	DROP
	INSERT
	SELECT
View	ALL (Super)
	SELECT
	DESCRIBE
	DROP
Data Catalog	CREATE_DATABASE
Amazon S3 location	DATA_LOCATION_ACCESS

Recurso	Permiso
LF-Tags	DROP
	ALTER
LF-Tag values	ASSOCIATE
	DESCRIBE
	GrantWithLFTagExpression
LF-Tag policy - Database	ALL (Super)
	ALTER
	CREATE_TABLE
	DESCRIBE
	DROP
LF-Tag policy - Table	ALL (Super)
	ALTER
	DESCRIBE
	DELETE
	DROP
	INSERT
	SELECT
Resource link - Database or Table	DESCRIBE
	DROP

Recurso	Permiso
Table with data filters	DESCRIBE
	DROP
	SELECT
Table with column filter	SELECT

Temas

- Concesión y revocación de Lake Formation AWS CLI Comandos de la de
- Permisos de Lake Formation

Concesión y revocación de Lake Formation AWS CLI Comandos de la de

La descripción de cada permiso de esta sección incluye ejemplos de cómo se concede el permiso mediante un AWS CLI comando. Las siguientes son las sinopsis de Lake Formation grant-permissions y revoke-permissions AWS CLI sus comandos.

```
grant-permissions
[--catalog-id <value>]
--principal <value>
--resource <value>
--permissions <value>
[--permissions-with-grant-option <value>]
[--cli-input-json <value>]
[--generate-cli-skeleton <value>]
```

```
revoke-permissions
[--catalog-id <value>]
--principal <value>
--resource <value>
--permissions <value>
[--permissions-with-grant-option <value>]
[--cli-input-json <value>]
[--generate-cli-skeleton <value>]
```

Para obtener descripciones detalladas de estos comandos, consulte grant-permissions y revoke-permissions en la Referencia de comandos de AWS CLI . En esta sección se proporciona información adicional sobre la opción `--principal`.

El valor de la opción `--principal` es uno de los siguientes:

- Nombre de recurso de Amazon (ARN) para un usuario o AWS Identity and Access Management rol (IAM)
- ARN para un usuario o grupo que se autentica a través de un proveedor SAML, como Microsoft Active Directory Federation Service (AD FS)
- ARN de un usuario o grupo de Amazon Quick
- Para los permisos entre cuentas, un identificador de AWS cuenta, un identificador de organización o un identificador de unidad organizativa
- Para un usuario o grupo de IAM Identity Center, ARN de usuario o grupo de IAM Identity Center.

A continuación encontrará la sintaxis y ejemplos para todos los tipos de `--principal`.

La entidad principal es un usuario de IAM

Sintaxis:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/<user-name>
```

Ejemplo:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1
```

La entidad principal es un rol de IAM

Sintaxis:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:role/<role-name>
```

Ejemplo:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:role/workflowrole
```

La entidad principal es un usuario que se autentica a través de un proveedor SAML

Sintaxis:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-
provider/<SAMLproviderName>:user/<user-name>
```

Ejemplos:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
idp1:user/datalake_user1
```

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
AthenaLakeFormationOkta:user/athena-user@example.com
```

La entidad principal es un grupo que se autentica a través de un proveedor SAML

Sintaxis:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-
provider/<SAMLproviderName>:group/<group-name>
```

Ejemplos:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
idp1:group/data-scientists
```

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
AthenaLakeFormationOkta:group/my-group
```

Principal es usuario de Amazon Quick Enterprise Edition

Sintaxis:

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-
id>:user/<namespace>/<user-name>
```

> **Note**
>
> Para *<namespace>*, debe especificar `default`.

Ejemplo:

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-
east-1:111122223333:user/default/bi_user1
```

Principal es un grupo de Amazon Quick Enterprise Edition

Sintaxis:

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-
id>:group/<namespace>/<group-name>
```

> **Note**
>
> Para *<namespace>*, debe especificar `default`.

Ejemplo:

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-
east-1:111122223333:group/default/data_scientists
```

Principal es una AWS cuenta

Sintaxis:

```
--principal DataLakePrincipalIdentifier=<account-id>
```

Ejemplo:

```
--principal DataLakePrincipalIdentifier=111122223333
```

La entidad principal es una organización

Sintaxis:

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-
id>:organization/<organization-id>
```

Ejemplo:

```
--principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/o-
abcdefghijkl
```

La entidad principal es una unidad organizativa

Sintaxis:

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-
id>:ou/<organization-id>/<organizational-unit-id>
```

Ejemplo:

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:ou/o-
abcdefghijkl/ou-ab00-cdefghij
```

Entidad principal es un usuario o grupo de identidades de IAM Identity Center

Example:User

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserID>
```

Example:Group:

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::group/<GroupID>
```

Entidad principal es un grupo de IAM: **`IAMAllowedPrincipals`**

Lake Formation concede el permiso `Super` sobre todas las bases de datos y tablas del Catálogo de datos a un grupo llamado `IAMAllowedPrincipals` de forma predeterminada. Si este permiso de grupo existe en una base de datos o una tabla, todas las entidades principales de la cuenta tendrán acceso al recurso a través de políticas de entidad principal de IAM para AWS Glue. Proporciona compatibilidad retroactiva cuando se empiezan a utilizar permisos de Lake Formation para proteger recursos del Catálogo de datos que anteriormente estaban protegidos mediante políticas de IAM para AWS Glue.

Cuando se usa Lake Formation para administrar los permisos sobre los recursos del Catálogo de datos, primero es necesario revocar el permiso `IAMAllowedPrincipals` de los recursos, o

bien inscribir las entidades principales y los recursos en el modo de acceso híbrido para que los permisos de Lake Formation funcionen.

Ejemplo:

```
--principal DataLakePrincipalIdentifier=IAM_Allowed_Principals
```

Entidad principal es un grupo de IAM: **`ALLIAMPrincipals`**

Al conceder permisos al grupo `ALLIAMPrincipals` en un recurso del Catálogo de datos, todas las entidades principales de la cuenta obtienen acceso al recurso del Catálogo de datos mediante permisos de Lake Formation y permisos de IAM.

Ejemplo:

```
--principal DataLakePrincipalIdentifier=123456789012:IAMPrincipals
```

Permisos de Lake Formation

Esta sección contiene los permisos disponibles de Lake Formation que puede conceder a las entidades principales.

`ALTER`

Permiso	Concedido sobre este recurso	El beneficiario también necesita
`ALTER`	`DATABASE`	`glue:UpdateDatabase`
`ALTER`	`TABLE`	`glue:UpdateTable`
`ALTER`	`LF-Tag`	`lakeformation:UpdateLFTag`

Una entidad principal con este permiso puede alterar los metadatos de una base de datos o tabla del Catálogo de datos. Para las tablas, puede cambiar el esquema de columnas y añadir parámetros de columna. No puede alterar las columnas de los datos subyacentes a los que apunta una tabla de metadatos.

Si la propiedad que se está modificando es una ubicación registrada de Amazon Simple Storage Service (Amazon S3), la entidad principal debe tener permisos de ubicación de datos en la nueva ubicación.

Example

En el siguiente ejemplo, se concede el `ALTER` permiso al usuario de la `datalake_user1` base de datos de `retail` la AWS cuenta 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "ALTER" --resource '{ "Database": {"Name":"retail"}}'
```

Example

El siguiente ejemplo concede `ALTER` al usuario `datalake_user1` en la tabla `inventory` de la base de datos `retail`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "ALTER" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

CREATE_DATABASE

Permiso	Concedido sobre este recurso	El beneficiario también necesita
`CREATE_DATABASE`	Data Catalog	`glue:CreateDatabase`

Una entidad principal con este permiso puede crear una base de metadatos o un enlace de recursos en el Catálogo de datos. La entidad principal también puede crear tablas en la base de datos.

Example

En el siguiente ejemplo, se concede `CREATE_DATABASE` al usuario de la cuenta `datalake_user1` 1111-2222-3333 AWS .

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

Cuando una entidad principal crea una base de datos en el Catálogo de datos, no se conceden permisos a los datos subyacentes. Se conceden los siguientes permisos adicionales para metadatos (junto con la posibilidad de conceder estos permisos a otros):

- `CREATE_TABLE` en la base de datos
- Base de datos de `ALTER`
- Base de datos de `DROP`

Al crear una base de datos, la entidad principal puede especificar de forma opcional una ubicación de Amazon S3. Dependiendo de si la entidad principal tiene permisos de localización de datos, el permiso `CREATE_DATABASE` podría no ser suficiente para crear bases de datos en todos los casos. Es importante tener en cuenta los siguientes puntos.

Caso práctico de creación de una base de datos	Permisos necesarios
La propiedad de ubicación no está especific ada.	`CREATE_DATABASE` es suficiente.
Se especifica la propiedad de ubicación, y la ubicación no está administrada por Lake Formation (no está registrada).	`CREATE_DATABASE` es suficiente.
Se especifica la propiedad de ubicación, y la ubicación es administrada por Lake Formation (está registrada).	Se requiere `CREATE_DATABASE` , además de permisos de localización de datos en la ubicación especificada.

CREATE_TABLE

Permiso	Concedido sobre este recurso	El beneficiario también necesita
CREATE_TABLE	DATABASE	glue:CreateTable

Una entidad principal con este permiso puede crear una tabla de metadatos o un enlace de recursos en el Catálogo de datos dentro de la base de datos especificada.

Example

El siguiente ejemplo concede al usuario datalake_user1 permiso para crear tablas en la retail base de datos de la cuenta 1111-2222-3333. AWS

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Cuando una entidad principal crea una tabla en el Catálogo de datos, se le conceden todos los permisos de Lake Formation sobre la tabla, con la posibilidad de conceder estos permisos sobre otros.

Cross-account Concesiones

Si una cuenta propietaria de una base de datos concede CREATE_TABLE a una cuenta de destinatario, y un usuario de la cuenta de destinatario crea con éxito una tabla en la base de datos de la cuenta propietaria, se aplican las siguientes reglas:

- El usuario y los administradores del lago de datos de la cuenta de destinatario tienen todos los permisos de Lake Formation sobre la tabla. Pueden conceder permisos sobre la tabla a otras entidades principales de su cuenta. No pueden conceder permisos a entidades principales en la cuenta del propietario ni en ninguna otra cuenta.
- Los administradores del lago de datos de la cuenta del propietario pueden conceder permisos sobre la tabla a otras entidades principales de su cuenta.

Permisos de ubicación de datos

Cuando intente crear una tabla que apunte a una ubicación de Amazon S3, dependiendo de si dispone de permisos de ubicación de datos, es posible que el permiso `CREATE_TABLE` no sea suficiente para crear una tabla. Es importante tener en cuenta los tres casos siguientes.

Caso práctico de creación de tablas	Permisos necesarios
Lake Formation no administra la ubicación especificada (no está registrada).	`CREATE_TABLE` es suficiente.
Lake Formation administra la ubicación especificada (está registrada), y la base de datos que la contiene no tiene propiedad de ubicación o tiene una propiedad de ubicación que no es un prefijo de Amazon S3 de la ubicación de la tabla.	Se requiere `CREATE_TABLE` , además de permisos de localización de datos en la ubicación especificada.
Lake Formation administra la ubicación especificada (está registrada), y la base de datos contenedora tiene una propiedad de ubicación que apunta a una ubicación que está registrada y es un prefijo de Amazon S3 de la ubicación de la tabla.	`CREATE_TABLE` es suficiente.

`DATA_LOCATION_ACCESS`

Permiso	Concedido sobre este recurso	El beneficiario también necesita
`DATA_LOCATION_ACCESS`	Ubicación de Amazon S3	(Permisos de Amazon S3 en la ubicación, que deben especificarse en el rol utilizado para registrar la ubicación).

Este es el único permiso de ubicación de datos. Una entidad principal con este permiso puede crear una base de datos o tabla de metadatos que apunte a la ubicación de Amazon S3 especificada. La

ubicación debe estar registrada. Una entidad principal que tiene permisos de localización de datos en una localización también tiene permisos de localización en las localizaciones secundarias.

Example

El siguiente ejemplo concede permisos de ubicación de datos en `s3://products/retail` al usuario `datalake_user1` en la cuenta de AWS 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::products/retail"}}'
```

`DATA_LOCATION_ACCESS` no es necesario para consultar o actualizar los datos subyacentes. Este permiso se aplica únicamente a la creación de recursos del Catálogo de datos.

Para obtener más información sobre permisos de ubicación de datos, consulte Underlying data access control.

DELETE

Permiso	Concedido sobre este recurso	El beneficiario también necesita
`DELETE`	`TABLE`	(No se necesitan permisos de IAM adicionales si la ubicación está registrada).

Una entidad principal con este permiso puede insertar, actualizar y leer datos subyacentes en la ubicación de Amazon S3 especificada por la tabla. La entidad principal también puede ver la tabla en la consola de Lake Formation y recuperar información sobre la tabla con la API AWS Glue.

Example

En el siguiente ejemplo, se concede el `DELETE` permiso al usuario `datalake_user1` de la tabla de la base `retail` de datos de la cuenta `inventory` 1111-2222-3333. AWS

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
```

```
--permissions "DELETE" --resource '{ "Table": {"DatabaseName":"retail",
"Name":"inventory"}}'
```

Este permiso se aplica solo a los datos de Amazon S3 y no a los de otros almacenes de datos como Amazon Relational Database Service (Amazon RDS).

DESCRIBE

Permiso	Concedido sobre este recurso	El beneficiario también necesita
DESCRIBE	Enlace de recurso a tabla Enlace de recurso a base de datos	glue:GetTable glue:GetDatabase
DESCRIBE	DATABASE	glue:GetDatabase
DESCRIBE	TABLE	glue:GetTable
DESCRIBE	LF-Tag	glue:GetTable glue:GetDatabase lakeformation:GetResourceLFTags lakeformation:ListLFTags lakeformation:GetLFTag lakeformation:SearchTablesByLFTags lakeformation:SearchDatabasesByLFTags

Una entidad principal con este permiso puede ver la base de datos, tabla o enlace de recursos especificados. No se concede implícitamente ningún otro permiso del Catálogo de datos ni ningún permiso de acceso a los datos. Las bases de datos y las tablas aparecen en los editores de consultas de los servicios integrados, pero no se puede consultar en ellas salvo que se concedan otros permisos de Lake Formation (por ejemplo, `SELECT`).

Por ejemplo, un usuario que tiene `DESCRIBE` en una base de datos puede ver la base de datos y todos los metadatos de la base de datos (descripción, ubicación, etc.). Sin embargo, el usuario no puede averiguar qué tablas contiene la base de datos y no puede eliminar, modificar o crear tablas en la base de datos. Del mismo modo, un usuario que tiene `DESCRIBE` en una tabla puede ver la tabla y los metadatos de la tabla (descripción, esquema, ubicación, etc.), pero no puede soltar, alterar o ejecutar consultas contra la tabla.

A continuación se detallan algunas normas adicionales para `DESCRIBE`:

- Si un usuario tiene otros permisos de Lake Formation sobre una base de datos, tabla o enlace de recursos, `DESCRIBE` se le concede implícitamente.
- Si un usuario tiene `SELECT` solo en un subconjunto de columnas de una tabla (`SELECT` parcial), el usuario estará restringido a ver solo esas columnas.
- No puede conceder `DESCRIBE` a un usuario que tiene selección parcial en una tabla. Por el contrario, no puede especificar listas de inclusión o exclusión de columnas para las tablas sobre las que se concede `DESCRIBE`.

Example

En el siguiente ejemplo, se concede el `DESCRIBE` permiso al usuario `datalake_user1` en el enlace de recursos de la tabla de la base `retail` de datos de la cuenta `inventory-link` 1111-2222-3333. AWS

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DESCRIBE" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory-link"}}'
```

DROP

Permiso	Concedido sobre este recurso	El beneficiario también necesita
DROP	DATABASE	glue:DeleteDatabase
DROP	TABLE	glue:DeleteTable
DROP	LF-Tag	lakeformation:DeleteLFTag
DROP	Enlace de recurso a base de datos Enlace de recurso a tabla	glue:DeleteDatabase glue:DeleteTable

Una entidad principal con este permiso puede borrar una base de datos, tabla o enlace de recurso en el Catálogo de datos. No puede conceder el permiso DROP en una base de datos a una cuenta u organización externa.

> **Warning**
>
> Al eliminar una base de datos se eliminan todas sus tablas.

Example

En el siguiente ejemplo, se concede el DROP permiso al usuario de la base de datos de la cuenta `datalake_user1` 1111-2222-3333`retail`. AWS

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Database": {"Name":"retail"}}'
```

Example

El siguiente ejemplo concede DROP al usuario `datalake_user1` sobre la tabla `inventory` en la base de datos `retail`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Example

El siguiente ejemplo concede DROP al usuario `datalake_user1` sobre el enlace de recursos de tabla `inventory-link` en la base de datos `retail`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail", "Name":"inventory-
link"}}'
```

INSERT

Permiso	Concedido sobre este recurso	El beneficiario también necesita
INSERT	TABLE	(No se necesitan permisos de IAM adicionales si la ubicación está registrada).

Una entidad principal con este permiso puede insertar, actualizar y leer datos subyacentes en la ubicación de Amazon S3 especificada por la tabla. La entidad principal también puede ver la tabla en la consola de Lake Formation y recuperar información sobre la tabla con la API AWS Glue.

Example

El siguiente ejemplo concede el INSERT permiso al usuario `datalake_user1` de la tabla de la base `retail` de datos de la cuenta `inventory` 1111-2222-3333. AWS

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "INSERT" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Este permiso se aplica únicamente a los datos de Amazon S3 y no a los datos de otros almacenes de datos, como Amazon RDS.

SELECT

Permiso	Concedido sobre este recurso	El beneficiario también necesita
SELECT	• TABLE	(No se necesitan permisos de IAM adicionales si la ubicación está registrada).

Una entidad principal con este permiso puede ver una tabla en el Catálogo de datos y consultar los datos subyacentes en Amazon S3 en la ubicación especificada por la tabla. La entidad principal puede ver la tabla en la consola de Lake Formation y recuperar información sobre la tabla con la API AWS Glue. Si se aplicó el filtrado por columnas cuando se concedió este permiso, la entidad principal puede ver los metadatos solo de las columnas incluidas y consultar los datos solo de las columnas incluidas.

> **Note**
>
> Es responsabilidad del servicio de análisis integrado aplicar el filtrado de columnas al procesar una consulta.

Example

En el siguiente ejemplo, se concede el SELECT permiso al usuario `datalake_user1` de la tabla de la base `retail` de datos de la cuenta `inventory` 1111-2222-3333. AWS

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "SELECT" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Este permiso se aplica únicamente a los datos de Amazon S3 y no a los datos de otros almacenes de datos, como Amazon RDS.

Puede filtrar (restringir el acceso a) columnas específicas con una lista de inclusión opcional o una lista de exclusión. Una lista de inclusión especifica las columnas a las que se puede acceder. Una lista de exclusión especifica las columnas a las que no se puede acceder. En ausencia de una lista de inclusión o exclusión, todas las columnas de la tabla son accesibles.

Los resultados de `glue:GetTable` devuelven solo las columnas que la persona que llama tiene permiso para ver. Los servicios integrados como Amazon Athena y Amazon Redshift respetan las listas de inclusión y exclusión de columnas.

Example

El ejemplo siguiente concede `SELECT` al usuario `datalake_user1` sobre la tabla `inventory` utilizando una lista de inclusión.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT"  --resource '{ "TableWithColumns": {"DatabaseName":"retail",
 "Name":"inventory", "ColumnNames": ["prodcode","location","period","withdrawals"]}}'
```

Example

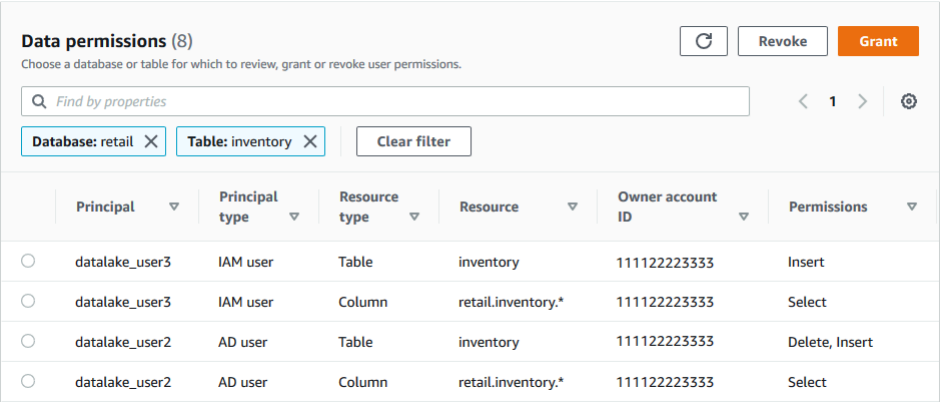
El ejemplo siguiente concede `SELECT` en la tabla `inventory` utilizando una lista de exclusión.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT"  --resource '{ "TableWithColumns": {"DatabaseName":"retail",
 "Name":"inventory", "ColumnWildcard": {"ExcludedColumnNames": ["intkey",
 "prodcode"]}}}'
```

Se aplican las siguientes restricciones al permiso `SELECT`:

- Al conceder `SELECT`, no puede incluir la opción de concesión si se aplica el filtrado por columnas.
- No puede restringir el control de acceso en columnas que son claves de partición.
- A una entidad principal con el permiso `SELECT` sobre un subconjunto de columnas de una tabla no se le puede conceder el permiso `ALTER`, `DROP`, `DELETE` o `INSERT` sobre esa tabla. Del mismo modo, a una entidad principal con el permiso `ALTER`, `DROP`, `DELETE`, o `INSERT` en una tabla no se le puede conceder el permiso `SELECT` con el filtrado de columnas.

El permiso SELECT siempre aparece en la página Permisos de datos de la consola de Lake Formation como una fila separada. En la imagen siguiente se muestra que SELECT se concede a los usuarios datalake_user2 y datalake_user3 en todas las columnas de la tabla inventory.



Super

Permiso	Concedido sobre este recurso	El beneficiario también necesita
Super	DATABASE	glue:*Database*
Super	TABLE	glue:*Table*, glue:*Partition*

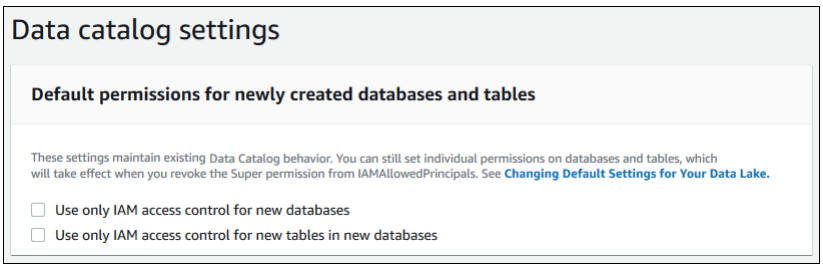
Este permiso permite a una entidad principal efectuar todas las operaciones compatibles con Lake Formation en la base de datos o en la tabla. No puede conceder Super en una base de datos a una cuenta externa.

Este permiso puede coexistir con los demás permisos de Lake Formation. Por ejemplo, puede conceder los permisos Super, SELECT y INSERT sobre una tabla de metadatos. La entidad principal puede entonces efectuar todas las operaciones admitidas en la tabla. Cuando revoca Super, los permisos SELECT y INSERT permanecen, y la entidad principal solo puede efectuar las operaciones de selección e inserción.

En lugar de conceder Super a una entidad principal individual, puede concederla al grupo IAMAllowedPrincipals. El grupo IAMAllowedPrincipals se crea automáticamente e incluye a todos los usuarios y roles de IAM a los que sus políticas de IAM permiten el acceso a los recursos de su Catálogo de datos. Cuando se concede Super a IAMAllowedPrincipals para un recurso

del Catálogo de datos, el acceso al recurso queda efectivamente controlado únicamente por las políticas de IAM.

Puede hacer que se conceda automáticamente el permiso `Super` a `IAMAllowedPrincipals` para los nuevos recursos del catálogo aprovechando las opciones de la página Configuración de la consola de Lake Formation.



- Para conceder `Super` a `IAMAllowedPrincipals` para todas las bases de datos nuevas, seleccione Usar solo control de acceso IAM para bases de datos nuevas.
- Para conceder `Super` a `IAMAllowedPrincipals` para todas las tablas nuevas en bases de datos nuevas, seleccione Usar solo control de acceso IAM para tablas nuevas en bases de datos nuevas.

> **Note**
>
> Esta opción hace que se marque de manera predeterminada la casilla Utilizar sólo el control de acceso IAM para las nuevas tablas de esta base de datos en el cuadro de diálogo Crear base de datos. No hace nada más que eso. Es la casilla de verificación del cuadro de diálogo Crear base de datos que permite la concesión de `Super` a `IAMAllowedPrincipals`.

Estas opciones de la página Configuración están habilitadas de forma predeterminada. Para obtener más información, consulte los siguientes temas:

- the section called “Cambiar la configuración predeterminada de su lago de datos”
- the section called “Actualización de los permisos de datos AWS Glue al modelo de Lake Formation”

SUPER_USUARIO

Permiso	Concedido sobre este recurso	El beneficiario también necesita
`Super user`	`Catalog`	`glue:GetCatalog`

Puede conceder el permiso `Super user` solo a entidades principales específicas sobre los catálogos del Catálogo de datos predeterminado. No puede conceder permisos `Super user` sobre el catálogo predeterminado ni sobre otros tipos de recursos, como bases de datos y tablas, ni a entidades principales de cuentas externas. El permiso `Super user` permite a una entidad principal efectuar todas las operaciones compatibles con Lake Formation en las bases de datos y tablas incluidas en el catálogo concedido.

Con el permiso `Super user`, la entidad principal (beneficiario) puede realizar las siguientes acciones en los recursos (catálogos, bases de datos y tablas) del catálogo:

- Permisos `CREATE_DATABASE` y `DESCRIBE` en el catálogo.
- Permisos `DROP`, `ALTER`, `CREATE_TABLE` y `DESCRIBE` (en efecto, `SUPER`) en todas las bases de datos del catálogo.
- Permisos `DROP`, `ALTER`, `DESCRIBE`, `SELECT`, `INSERT` y `DELETE` (en efecto, `SUPER`) en todas las tablas de todas las bases de datos del catálogo.
- Permisos `All` (en efecto, SUPER) en los catálogos incluidos en el catálogo.
- Permisos concesibles (la posibilidad de conceder estos permisos a otras entidades principales) en todos los catálogos, bases de datos y tablas del catálogo.

Con el permiso `Super user` en un recurso del catálogo, el beneficiario no puede realizar ni delegar las acciones `ALTER` y `DROP` en el catálogo.

ASOCIAR

Permiso	Concedido sobre este recurso	El beneficiario también necesita
`ASSOCIATE`	`LF-Tag`	`glue:GetDatabase`

Permiso	Concedido sobre este recurso	El beneficiario también necesita

		`glue:GetTable`
		
		`lakeformation:AddLFTagsToResource"`
		
		`lakeformation:RemoveLFTagsFromResource"`
		
		`lakeformation:GetResourceLFTags`
		
		`lakeformation:ListLFTags`
		
		`lakeformation:GetLFTag`
		
		`lakeformation:SearchTablesByLFTags`
		
		`lakeformation:SearchDatabasesByLFTags`

Un director con este permiso LF-Tag puede asignarlo LF-Tag a un recurso del catálogo de datos. Al conceder `ASSOCIATE` está otorgando `DESCRIBE` de manera implícita.

Example

En este ejemplo, se concede `datalake_user1` al usuario el `ASSOCIATE` permiso correspondiente LF-Tag a la clave`module`. Concede permisos para ver y asignar todos los valores de esa clave, como indica el asterisco (*).

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Integración de IAM Identity Center

Con AWS IAM Identity Center, puede conectarse a los proveedores de identidad (IdPs) y gestionar de forma centralizada el acceso de los usuarios y grupos a los servicios de AWS análisis. Puede integrar proveedores de identidad como Okta, Ping y Microsoft Entra ID (anteriormente Azure Active Directory) con IAM Identity Center para que los usuarios de su organización accedan a los datos mediante una experiencia de inicio de sesión único. IAM Identity Center también permite conectar otros proveedores de identidad de terceros.

Para obtener más información, consulte los proveedores de identidad compatibles en la Guía del AWS IAM Identity Center usuario.

Puede configurarla AWS Lake Formation como una aplicación habilitada en IAM Identity Center, y los administradores de los lagos de datos pueden conceder permisos detallados a los usuarios y grupos autorizados sobre los recursos. AWS Glue Data Catalog

Los usuarios de la organización pueden iniciar sesión en cualquier aplicación habilitada para Identity Center mediante el proveedor de identidad de su organización y consultar conjuntos de datos aplicando los permisos de Lake Formation. Con esta integración, puede gestionar el acceso a AWS los servicios sin crear varias funciones de IAM.

La propagación confiable de identidades es una AWS IAM Identity Center función que los administradores de Connected Servicios de AWS pueden usar para conceder y auditar el acceso a los datos del servicio. El acceso a estos datos se basa en los atributos del usuario, como las asociaciones de grupo. La configuración de una propagación de identidades fiable requiere la colaboración entre los administradores de Connected Servicios de AWS y los administradores del centro de identidad de IAM. Para obtener más información, consulte Requisitos previos y consideraciones.

Para conocer las limitaciones, consulte Limitaciones de la integración de IAM Identity Center.

Temas

- Requisitos previos para la integración de IAM Identity Center con Lake Formation
- Conexión de Lake Formation con IAM Identity Center
- Actualización de la integración de IAM Identity Center
- Eliminación de una conexión de Lake Formation con IAM Identity Center
- Concesión de permisos a usuarios y grupos
- Incluir el contexto de usuario de IAM Identity Center en los registros CloudTrail

Requisitos previos para la integración de IAM Identity Center con Lake Formation

A continuación, se indican los requisitos previos para integrar IAM Identity Center con Lake Formation.

1. Habilitar IAM Identity Center: la habilitación de IAM Identity Center es un requisito previo para permitir la autenticación y la propagación de identidades.
2. Elegir su origen de identidades: después de habilitar IAM Identity Center, debe tener un proveedor de identidades para administrar los usuarios y grupos. Puede usar el directorio integrado del Identity Center como origen de identidad o usar un IdP externo, como Microsoft Entra ID u Okta.

   Para obtener más información, consulte Administrar su fuente de identidad y conectarse a un proveedor de identidades externo en la Guía del AWS IAM Identity Center usuario.
3. Crear un rol de IAM: el rol que crea la conexión a IAM Identity Center requiere permisos para crear y modificar la configuración de la aplicación en Lake Formation e IAM Identity Center, como se indica en la siguiente política en línea.

   Debe añadir permisos según las prácticas recomendadas de IAM. Los permisos específicos se detallan en los siguientes procedimientos. Para obtener más información, consulte Primeros pasos con IAM Identity Center.

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:CreateLakeFormationIdentityCenterConfiguration",
                "sso:CreateApplication",
                "sso:PutApplicationAssignmentConfiguration",
                "sso:PutApplicationAuthenticationMethod",
                "sso:PutApplicationGrant",
                "sso:PutApplicationAccessScope"
            ],
            "Resource": [
                "*"
            ]
```

```
        }
    ]
}
```

Si compartes recursos del catálogo de datos con organizaciones externas Cuentas de AWS o con organizaciones, debes tener los permisos AWS Resource Access Manager (AWS RAM) para crear recursos compartidos. Para obtener más información sobre los permisos necesarios para compartir recursos, consulte los requisitos previos para compartir Cross-account datos.

Las siguientes políticas en línea contienen los permisos específicos necesarios para ver, actualizar y eliminar las propiedades de la integración de Lake Formation con IAM Identity Center.

- Utilice la siguiente política en línea para permitir que un rol de IAM vea una integración de Lake Formation con IAM Identity Center.

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [

 "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
                "sso:DescribeApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- Utilice la siguiente política en línea para permitir que un rol de IAM actualice una integración de Lake Formation con IAM Identity Center. La política también incluye los permisos opcionales necesarios para compartir recursos con cuentas externas.

  JSON

```
{
```

```
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:UpdateLakeFormationIdentityCenterConfiguration",

 "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
                "sso:DescribeApplication",
                "sso:UpdateApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- Utilice la siguiente política en línea para permitir que un rol de IAM elimine una integración de Lake Formation con IAM Identity Center.

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:DeleteLakeFormationIdentityCenterConfiguration",
                "sso:DeleteApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- Para saber qué permisos de IAM son necesarios para conceder o revocar permisos de lago de datos a usuarios y grupos de IAM Identity Center, consulte Permisos de IAM necesarios para conceder o revocar permisos de Lake Formation.

Descripción de los permisos

- `lakeformation:CreateLakeFormationIdentityCenterConfiguration` – Crea la configuración IdC de Lake Formation.
- `lakeformation:DescribeLakeFormationIdentityCenterConfiguration` – Describe una configuración de IdC existente.
- `lakeformation:DeleteLakeFormationIdentityCenterConfiguration` – Ofrece la posibilidad de eliminar una configuración de IdC de Lake Formation existente.
- `lakeformation:UpdateLakeFormationIdentityCenterConfiguration` – Se usa para cambiar una configuración existente de Lake Formation.
- `sso:CreateApplication` – Se utiliza para crear una aplicación de IAM Identity Center.
- `sso:DeleteApplication`: se utiliza para eliminar una aplicación IAM Identity Center.
- `sso:UpdateApplication`: se utiliza para actualizar una aplicación IAM Identity Center.
- `sso:PutApplicationGrant` – Se utiliza para cambiar la información del emisor de tokens de confianza.
- `sso:PutApplicationAuthenticationMethod` – Otorga acceso de autenticación a Lake Formation.
- `sso:GetApplicationGrant` – Se utiliza para enumerar la información del emisor de tokens de confianza.
- `sso:DeleteApplicationGrant` – Elimina la información del emisor de tokens de confianza.
- `sso:PutApplicationAccessScope` – Añade o actualiza la lista de objetivos autorizados para el ámbito de acceso de una aplicación a IAM Identity Center.
- `sso:PutApplicationAssignmentConfiguration` – Se utiliza para configurar cómo acceden los usuarios a una aplicación.

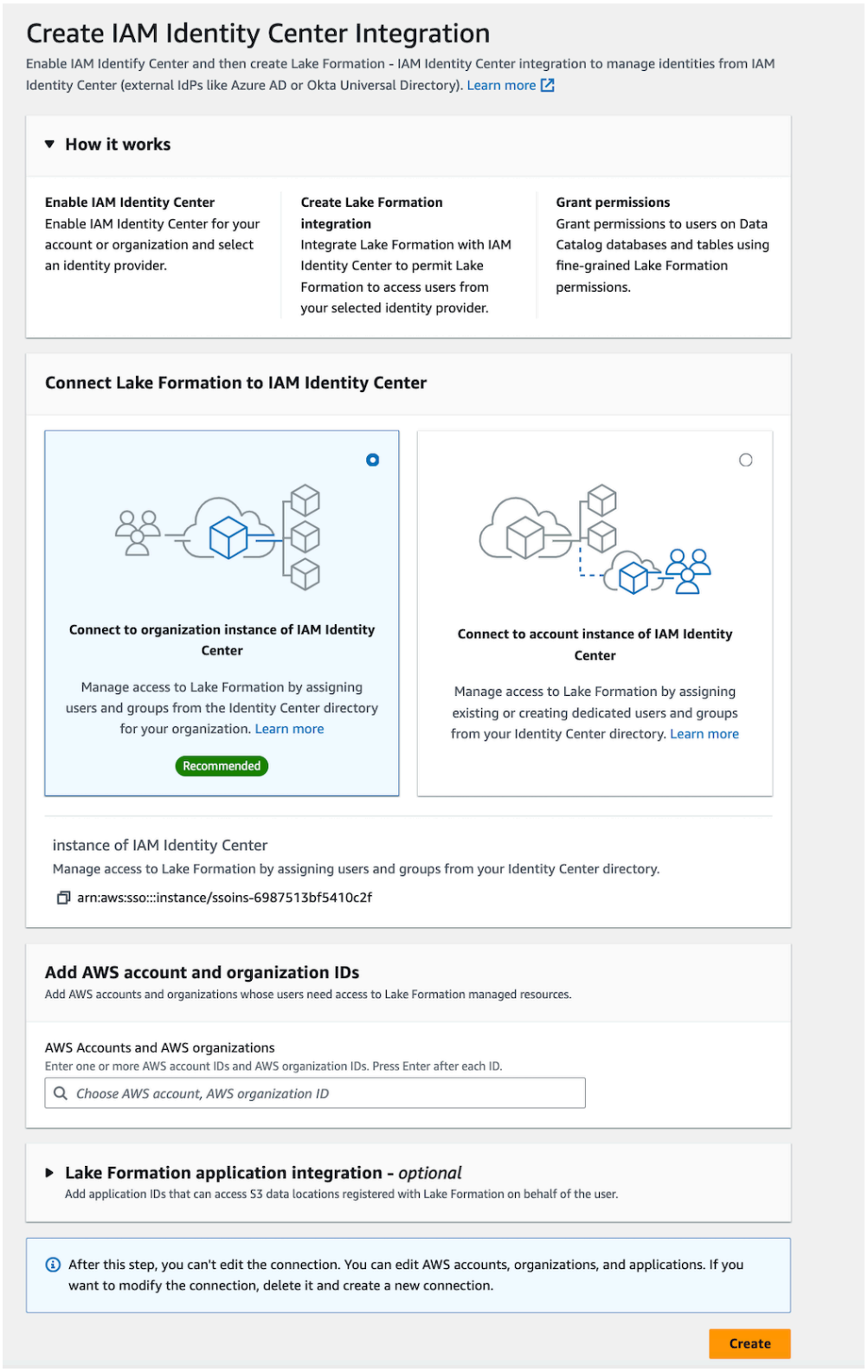
Conexión de Lake Formation con IAM Identity Center

Antes de poder usar IAM Identity Center para administrar identidades y conceder acceso a los recursos del catálogo de datos mediante Lake Formation, debe realizar los siguientes pasos. Puede crear la integración de IAM Identity Center mediante la consola o AWS CLI de Lake Formation.

Consola de administración de AWS

Para conectar Lake Formation con IAM Identity Center

1. Inicie sesión en y abra la Consola de administración de AWS consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación izquierdo, seleccione Integración de IAM Identity Center.



3. (Opcional) Introduzca uno o más Cuenta de AWS identificadores de organización o de unidad and/or organizativa válidos para permitir que las cuentas externas accedan a los recursos del catálogo de datos. Cuando los usuarios o grupos de IAM Identity Center intentan acceder

a los recursos del Catálogo de datos administrados por Lake Formation, Lake Formation asume un rol de IAM para autorizar el acceso a los metadatos. Si el rol de IAM pertenece a una cuenta externa que no tiene una política de AWS Glue recursos ni un AWS RAM recurso compartido, los usuarios y grupos del IAM Identity Center no podrán acceder al recurso aunque tengan permisos de Lake Formation.

Lake Formation usa el servicio AWS Resource Access Manager (AWS RAM) para compartir el recurso con cuentas y organizaciones externas. AWS RAM envía una invitación a la cuenta del beneficiario para aceptar o rechazar el recurso compartido.

Para obtener más información, consulte Aceptar una invitación para compartir recursos de AWS RAM.

> **Note**
>
> Lake Formation permite que los roles de IAM de cuentas externas actúen como roles de operador en nombre de los usuarios y grupos de IAM Identity Center para acceder a los recursos del Catálogo de datos, pero solo se pueden conceder permisos a recursos del Catálogo de datos de la cuenta propietaria. Si intentas conceder permisos a los usuarios y grupos del IAM Identity Center sobre los recursos del catálogo de datos de una cuenta externa, Lake Formation arroja el siguiente error: «no se admiten Cross-account subvenciones para el director».

4. (Opcional) En la pantalla Crear integración de Lake Formation, especifique los ARN de las aplicaciones de terceros que pueden acceder a los datos de las ubicaciones de Amazon S3 registradas en Lake Formation. Lake Formation vende credenciales temporales limitadas en forma de AWS STS tokens a las ubicaciones de Amazon S3 registradas en función de los permisos vigentes, para que las aplicaciones autorizadas puedan acceder a los datos en nombre de los usuarios.
5. (Opcional) En la pantalla Crear una integración con Lake Formation, marque la casilla Amazon Redshift Connect en Trusted Identity Propagation para poder detectar los permisos federados de Amazon Redshift a través de IDC. Lake Formation propaga identidad a posteriores en función de los permisos efectivos, por lo que las aplicaciones autorizadas pueden acceder a los datos en nombre de los usuarios.
6. Seleccione Enviar.

   Una vez que el administrador de Lake Formation finalice los pasos y cree la integración, las propiedades de IAM Identity Center aparecen en la consola de Lake Formation. Al

realizar estas tareas, Lake Formation se convierte en una aplicación habilitada para IAM Identity Center. Las propiedades de la consola incluyen el estado de la integración. Cuando se completa, el estado de la integración muestra `Success`. Este estado indica si la configuración de IAM Identity Center se ha completado.

AWS CLI

- En el siguiente ejemplo se muestra cómo crear la integración de Lake Formation con IAM Identity Center. También puede especificar el `Status` (`ENABLED`, `DISABLED`) de las aplicaciones.

```
aws lakeformation create-lake-formation-identity-center-configuration \
    --catalog-id <123456789012> \
    --instance-arn <arn:aws:sso:::instance/ssoins-112111f12ca1122p> \
    --share-recipients '[{"DataLakePrincipalIdentifier": "<123456789012>"},
                         {"DataLakePrincipalIdentifier": "<555555555555>"}]' \
    --external-filtering '{"AuthorizedTargets": ["<app arn1>", "<app arn2>"],
 "Status": "ENABLED"}'
```

- En el siguiente ejemplo se muestra cómo ver una integración de Lake Formation con IAM Identity Center.

```
aws lakeformation describe-lake-formation-identity-center-configuration
 --catalog-id <123456789012>
```

- El siguiente ejemplo muestra cómo habilitar la autorización. `Redshift:Connect` La autorización se puede HABILITAR o DESHABILITAR.

```
aws lakeformation  create-lake-formation-identity-center-configuration \
--instance-arn <arn:aws:sso:::instance/ssoins-112111f12ca1122p> \
--service-integrations '[{
  "Redshift": [{
    "RedshiftConnect": {
      "Authorization": "ENABLED"
    }
  }]
}]'
```

- Use el `describe-lake-formation-identity-center-configuration` comando para describir la aplicación del centro de identidad de Lake Formation. `Redshift:Connect`la integración de servicios es esencial para la propagación de identidades de IDC entre servicios y clústeres:

```
aws lakeformation describe-lake-formation-identity-center-configuration --catalog-
id <123456789012>
```

Respuesta:

```
{
    "CatalogId": "CATALOG ID",
    "InstanceArn": "INSTANCE ARN",
    "ApplicationArn": "APPLICATION ARN",
    "ShareRecipients": [],
    "ServiceIntegrations": [
        {
            "Redshift": [
                {
                    "RedshiftConnect": {
                        "Authorization": "ENABLED"
                    }
                }
            ]
        }
    ]
}
```

Uso de IAM Identity Center en varios Regiones de AWS

Lake Formation es compatible con IAM Identity Center en múltiples ocasiones. Regiones de AWS Puede ampliar el IAM Identity Center de su región principal Región de AWS a otras regiones para mejorar el rendimiento gracias a la proximidad a los usuarios y la confiabilidad. Cuando se agrega una nueva región al centro de identidad de IAM, puede crear aplicaciones para el centro de identidad de Lake Formation en la nueva región sin replicar las identidades de la región principal. Para obtener más información sobre cómo empezar a usar IAM Identity Center en varias regiones, consulte Multi-Region IAM Identity Center en la guía del usuario del IAM Identity Center.

Actualización de la integración de IAM Identity Center

Tras crear la conexión, puede añadir aplicaciones de terceros para que la integración de IAM Identity Center se integre con Lake Formation y obtener acceso a los datos de Amazon S3 en nombre de los usuarios. También puede eliminar las aplicaciones existentes de la integración de IAM Identity Center. Puede añadir o eliminar aplicaciones mediante la consola de Lake Formation y mediante la AWS CLI operación. UpdateLakeFormationIdentityCenterConfiguration

> **Note**
>
> Tras crear la integración de IAM Identity Center, no podrá actualizar el ARN de la instancia.

Consola de administración de AWS

Para actualizar una conexión existente de IAM Identity Center con Lake Formation

1. Inicie sesión en y abra la Consola de administración de AWS consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación izquierdo, seleccione Integración de IAM Identity Center.
3. Seleccione Añadir en la página Integración de IAM Identity Center.
4. Introduzca uno o más Cuenta de AWS identificadores válidos, identificadores de organizaciones o identificadores de unidades and/or organizativas para permitir que las cuentas externas accedan a los recursos del catálogo de datos.
5. En la pantalla Añadir aplicaciones, introduzca los ID de las aplicaciones de terceros que desee integrar con Lake Formation.
6. Seleccione Añadir.
7. (Opcional) En la página de integración del IAM Identity Center, puede habilitar la propagación de identidades confiable para Amazon Redshift Connect o deshabilitarla. Lake Formation propaga identidad a posteriores en función de los permisos efectivos, por lo que las aplicaciones autorizadas pueden acceder a los datos en nombre de los usuarios.

AWS CLI

Puede añadir o eliminar aplicaciones de terceros para la integración del IAM Identity Center ejecutando el siguiente comando. AWS CLI Cuando se establece el estado de filtrado externo

en ENABLED, IAM Identity Center puede proporcionar administración de identidades para que las aplicaciones de terceros accedan a los datos administrados por Lake Formation. También puede activar o desactivar la integración de IAM Identity Center configurando el estado de la aplicación.

```
aws lakeformation update-lake-formation-identity-center-configuration \
 --external-filtering '{"AuthorizedTargets": ["<app arn1>", "<app arn2>"], "Status":
 "ENABLED"}'\
 --share-recipients '[{"DataLakePrincipalIdentifier": "<444455556666>"}
                     {"DataLakePrincipalIdentifier": "<777788889999>"}]' \
 --application-status ENABLED
```

Si ya tiene una aplicación LF IDC, pero desea agregar la `Redshift:Connect` autorización, puede usar lo siguiente para actualizar su aplicación IDC de Lake Formation. La autorización se puede HABILITAR o DESHABILITAR.

```
aws lakeformation update-lake-formation-identity-center-configuration \
--service-integrations '[{

  "Redshift": [{
    "RedshiftConnect": {
      "Authorization": "ENABLED"
    }
  }]
}]'
```

Eliminación de una conexión de Lake Formation con IAM Identity Center

Si quieres eliminar una integración de IAM Identity Center existente, puedes hacerlo mediante la consola o la DeleteLakeFormationIdentityCenterConfiguration operación de Lake Formation. AWS CLI

Consola de administración de AWS

Para eliminar una conexión existente de IAM Identity Center con Lake Formation

1. Inicie sesión en y abra la Consola de administración de AWS consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación izquierdo, seleccione Integración de IAM Identity Center.
3. Seleccione Eliminar en la página Integración de IAM Identity Center.

4. En la pantalla Confirmar integración, confirme la acción y seleccione Eliminar.

AWS CLI

Para eliminar la integración de IAM Identity Center, ejecuta el siguiente AWS CLI comando.

```
aws lakeformation delete-lake-formation-identity-center-configuration \
    --catalog-id <123456789012>
```

Concesión de permisos a usuarios y grupos

El administrador del lago de datos puede conceder permisos a los usuarios y grupos de IAM Identity Center sobre los recursos del catálogo de datos (bases de datos, tablas y vistas) para permitir un fácil acceso a los datos. Para conceder o revocar los permisos del lago de datos, el otorgante necesita permisos para las siguientes acciones de IAM Identity Center.

- DescribeUser
- DescribeGroup
- DescribeInstance

Puede conceder permisos utilizando la consola de Lake Formation, la API o la AWS CLI.

Para obtener más información sobre cómo conceder permisos, consulte the section called "Concesión de permisos de datos".

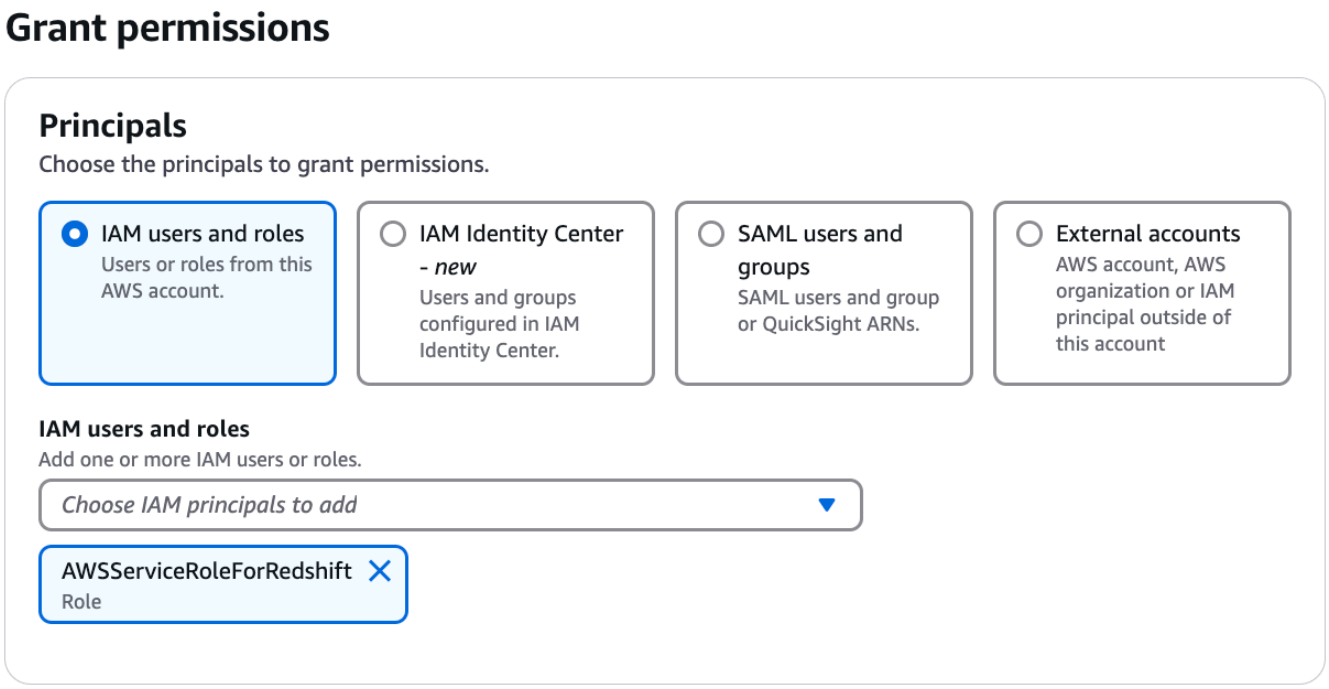
> **Note**
>
> Solo puede conceder permisos a los recursos de su cuenta. Para asignar en cascada los permisos a los usuarios y grupos sobre los recursos compartidos contigo, debes usar AWS RAM los recursos compartidos.

Consola de administración de AWS

Para conceder permisos a usuarios y grupos

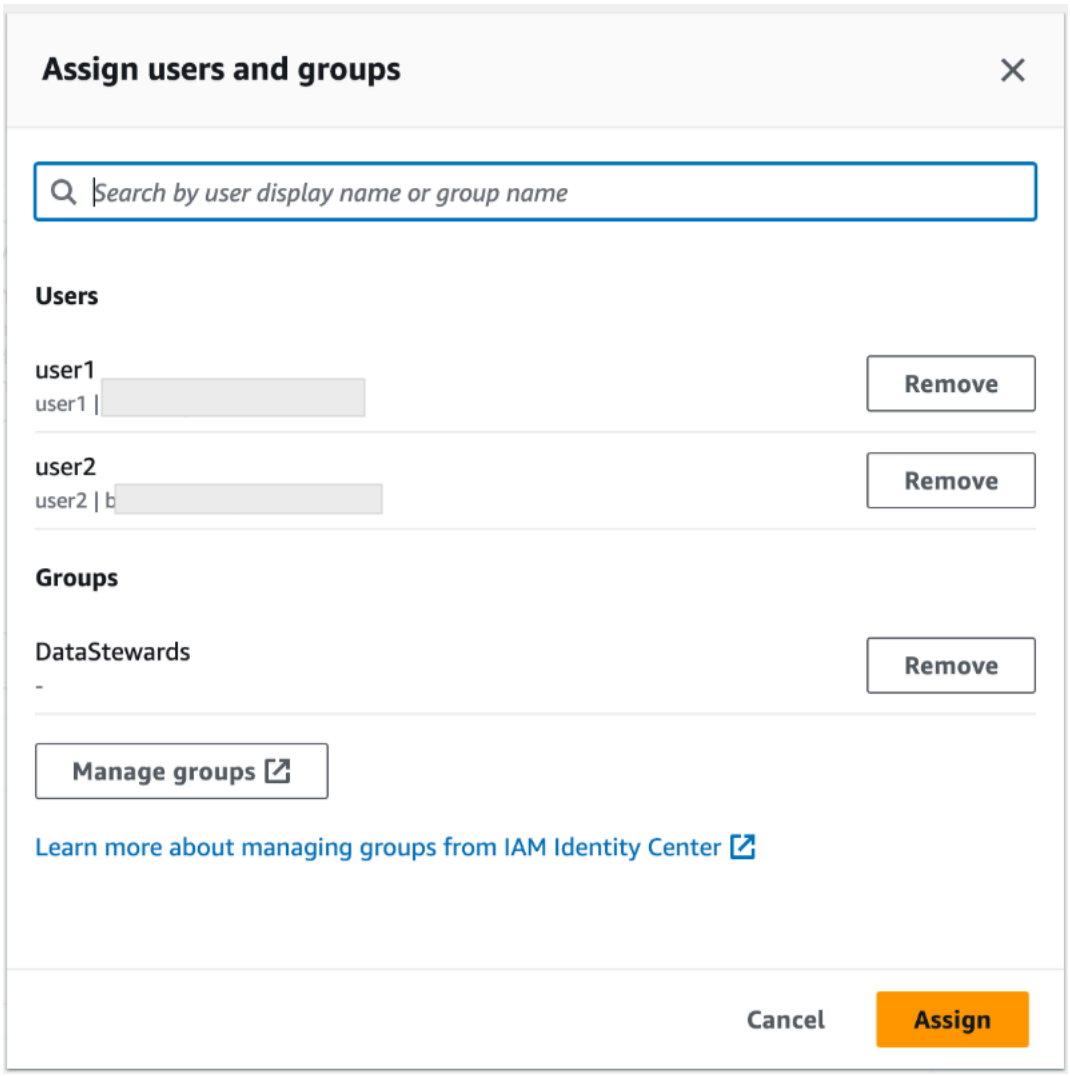
1. Inicie sesión en y abra la Consola de administración de AWS consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

2. Seleccione Permisos de lago de datos bajo Permisos in en la consola de Lake Formation.
3. Seleccione Conceder.
4. En la página Conceder permisos de lagos de datos, elija usuarios y grupos de IAM Identity Center.
5. Seleccione Añadir para elegir los usuarios y grupos a los que va a conceder permisos.



6. En la pantalla Asignar usuarios y grupos, elija los and/or grupos de usuarios a los que desea conceder los permisos.

   Seleccione Asignar.



7. A continuación, elija el método para conceder los permisos.

   Para obtener instrucciones sobre cómo conceder permisos mediante el método de recursos con nombre, consulte Concesión de permisos de datos mediante el método de recurso con nombre.

   Para obtener instrucciones sobre cómo conceder permisos de uso LF-Tags, consulteOtorgar permisos de lago de datos mediante el método LF-TBAC.

8. Elija los recursos del catálogo de datos de los que desea conceder permisos.
9. Elija los permisos del catálogo de datos que desee conceder.
10. Seleccione Conceder.

AWS CLI

En el siguiente ejemplo, se muestra cómo conceder el permiso `SELECT` de usuario de IAM Identity Center en una tabla.

```
aws lakeformation grant-permissions \
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserId> \
--permissions "SELECT" \
--resource '{ "Table": { "DatabaseName": "retail", "TableWildcard": {} } }'
```

Para recuperarlos `UserId` desde el IAM Identity Center, consulte el GetUserId funcionamiento en la referencia de la API del IAM Identity Center.

Incluir el contexto de usuario de IAM Identity Center en los registros CloudTrail

Lake Formation utiliza la funcionalidad de expendición de credenciales para proporcionar acceso temporal a los datos de Amazon S3. De forma predeterminada, cuando un usuario del IAM Identity Center envía una consulta a un servicio de análisis integrado, los CloudTrail registros solo incluyen la función de IAM asumida por el servicio para proporcionar acceso a corto plazo. Si utiliza un rol definido por el usuario para registrar la ubicación de datos de Amazon S3 en Lake Formation, puede optar por incluir el contexto del usuario del IAM Identity Center en los CloudTrail eventos y, a continuación, realizar un seguimiento de los usuarios que acceden a sus recursos.

> **Important**
>
> Para incluir las solicitudes de la API de Amazon S3 a nivel de objeto en el CloudTrail, debe habilitar el registro de CloudTrail eventos para los objetos y los cubos de Amazon S3. Para obtener más información, consulte Habilitar el registro de CloudTrail eventos para los buckets y objetos de Amazon S3 en la guía del usuario de Amazon S3.

Para habilitar la auditoría de la expedición de credenciales en ubicaciones de lagos de datos registradas con roles definidos por el usuario

1. Inicie sesión en la consola de Lake Formation en. https://console.aws.amazon.com/lakeformation/
2. En el panel de navegación de la izquierda, expanda Administración y elija Configuración del Catálogo de datos.
3. En Auditoría mejorada, elija Propagar el contexto proporcionado.
4. Seleccione Save.

También puede habilitar la opción de auditoría mejorada configurando el `Parameters` atributo en la PutDataLakeSettings operación. De forma predeterminada, el valor del parámetro `SET_CONTEXT"` se establece en “verdadero”.

```
{
    "DataLakeSettings": {
        "Parameters": {"SET_CONTEXT": "true"},
    }
}
```

El siguiente es un extracto de un CloudTrail evento con la opción de auditoría mejorada. Este registro incluye tanto el contexto de sesión del usuario de IAM Identity Center como el rol de IAM definido por el usuario que asume Lake Formation para acceder a la ubicación de datos de Amazon S3. Consulte el parámetro on`Behalf0f` en el siguiente extracto.

```
{
         "eventVersion":"1.09",
         "userIdentity":{
            "type":"AssumedRole",
            "principalId":"AROAW7F7MOX4OYE6FLIFN:access-grants-
e653760c-4e8b-44fd-94d9-309e035b75ab",
            "arn":"arn:aws:sts::123456789012:assumed-role/accessGrantsTestRole/access-
grants-e653760c-4e8b-44fd-94d9-309e035b75ab",
            "accountId":"123456789012",
            "accessKeyId":"ASIAW7F7MOX4CQLD4JIZN",
            "sessionContext":{
               "sessionIssuer":{
                  "type":"Role",
                  "principalId":"AROAW7F7MOX4OYE6FLIFN",
```

```
                "arn":"arn:aws:iam::123456789012:role/accessGrantsTestRole",
                "accountId":"123456789012",
                "userName":"accessGrantsTestRole"
             },
             "attributes":{
                "creationDate":"2023-08-09T17:24:02Z",
                "mfaAuthenticated":"false"
             }
          },
          "onBehalfOf":{
              "userId": "<identityStoreUserId>",
              "identityStoreArn": "arn:aws:identitystore::<restOfIdentityStoreArn>"
          }
       },
       "eventTime":"2023-08-09T17:25:43Z",
       "eventSource":"s3.amazonaws.com",
       "eventName":"GetObject",
....
```

Añadir una ubicación de Amazon S3 a su lago de datos

Para añadir una ubicación de datos como almacenamiento en su lago de datos, registre la ubicación (ubicación del lago de datos) con AWS Lake Formation. Luego, puede usar los permisos de Lake Formation para un control de acceso detallado a AWS Glue Data Catalog los objetos que apuntan a esta ubicación y a los datos subyacentes de la ubicación.

Lake Formation también permite registrar una ubicación de datos en modo de acceso híbrido y le proporciona la flexibilidad de habilitar selectivamente los permisos de Lake Formation para bases de datos y tablas en su Catálogo de datos. Con el modo de acceso híbrido, dispone de una ruta incremental que le permite establecer los permisos de Lake Formation para un conjunto específico de usuarios sin interrumpir las políticas de permisos de otros usuarios o cargas de trabajo existentes.

Para más información sobre la configuración del modo de acceso híbrido, consulte Modo de acceso híbrido

Al registrar una ubicación, se registran esa ruta de Amazon S3 y todas las carpetas incluidas en esa ruta.

Por ejemplo, supongamos que tiene una organización de rutas de Amazon S3 como la siguiente:

`/mybucket/accounting/sales/`

Si registra `S3://mybucket/accounting`, la carpeta `sales` también estará registrada y bajo la administración de Lake Formation.

Para obtener más información sobre cómo registrar ubicaciones, consulte Underlying data access control.

> **Note**
>
> Se recomiendan los permisos de Lake Formation para los datos estructurados (organizados en tablas con filas y columnas). Si sus datos contienen datos no estructurados basados en objetos, plantéese la posibilidad de utilizar concesiones de acceso a Amazon S3 para administrar el acceso a los datos.

Temas

- Requisitos de los roles utilizados para registrar ubicaciones
- Registro de una ubicación de Amazon S3
- Registro de una ubicación cifrada de Amazon S3
- Registrar una ubicación de Amazon S3 en otra AWS inscrita
- Registrar una ubicación de Amazon S3 cifrada en AWS cuentas
- Dar de baja el registro de una ubicación de Amazon S3
- Acceso a las ubicaciones de Amazon S3

Requisitos de los roles utilizados para registrar ubicaciones

Debe especificar un rol AWS Identity and Access Management (IAM) al registrar una ubicación de Amazon Simple Storage Service (Amazon S3). AWS Lake Formation asume esa función al acceder a los datos en esa ubicación.

Para registrar una ubicación, puede utilizar uno de los siguientes tipos de rol:

- El rol vinculado al servicio de Lake Formation. Este rol concede los permisos necesarios sobre la ubicación. Utilizar este rol es la forma más sencilla de registrar la ubicación. Para obtener más información, consulte Uso de roles vinculados a servicios para Lake Formation y Limitaciones de los roles vinculados a servicios.

- Un rol definido por el usuario. Utilice un rol definido por el usuario cuando necesite conceder más permisos de los que proporciona el rol vinculado al servicio.

  Debe utilizar un rol definido por el usuario en las circunstancias siguientes:

  - Al registrar una ubicación en otra cuenta.

    Para obtener más información, consulte the section called “Registrar una ubicación de Amazon S3 en otra AWS inscrita” y the section called “Registrar una ubicación de Amazon S3 cifrada en AWS cuentas”.
  - Si utilizó una CMK (`aws/s3`) AWS administrada para cifrar la ubicación de Amazon S3.

    Para obtener más información, consulte Registro de una ubicación cifrada de Amazon S3.
  - Si tiene previsto acceder a la ubicación mediante Amazon EMR.

    Si ya ha registrado una ubicación con el rol vinculado al servicio y desea comenzar a acceder a la ubicación con Amazon EMR, deberá anular el registro de la ubicación y volver a registrarla con un rol definido por el usuario. Para obtener más información, consulte the section called “Dar de baja el registro de una ubicación de Amazon S3”.

Uso de roles vinculados a servicios para Lake Formation

AWS Lake Formation utiliza un rol vinculado a un servicio AWS Identity and Access Management (IAM). Un rol vinculado a un servicio es un tipo único de rol de IAM que está vinculado directamente a Lake Formation. Lake Formation predefine la función vinculada al servicio e incluye todos los permisos que el servicio requiere para llamar a otros AWS servicios en su nombre.

Un rol vinculado a un servicio simplifica la configuración de Lake Formation porque ya no tendrá que agregar manualmente los permisos necesarios. Lake Formation define los permisos de su rol vinculado al servicio y, a menos que se defina lo contrario, solo Lake Formation puede asumir sus roles. Los permisos definidos incluyen las políticas de confianza y de permisos, y que la política de permisos no se puede asociar a ninguna otra entidad de IAM.

Este rol vinculado a servicios confía en los siguientes servicios para asumir el rol:

- `lakeformation.amazonaws.com`

Cuando se utiliza un rol vinculado a un servicio en la cuenta A para registrar una ubicación de Amazon S3 propiedad de la cuenta B, la política de bucket de Amazon S3 (una política basada en

recursos) de la cuenta B debe conceder permisos de acceso al rol vinculado al servicio de la cuenta A.

Para obtener información sobre el uso del rol vinculado a un servicio para registrar una ubicación de datos, consulte Limitaciones de los roles vinculados a servicios.

> **Note**
>
> Las políticas de control de servicios (SCPs) no afectan a las funciones vinculadas al servicio. Para obtener más información, consulte las políticas de control de servicios (SCPs) en la guía del AWS Organizations usuario.

Permisos de rol vinculados al servicio para Lake Formation

Lake Formation utiliza el rol vinculado al servicio denominado `AWSServiceRoleForLakeFormationDataAccess`. Esta función proporciona un conjunto de permisos de Amazon Simple Storage Service (Amazon S3) que permiten al servicio integrado de Lake Formation (por ejemplo) acceder a Amazon Athena las ubicaciones registradas. Al registrar una ubicación de lago de datos, debe proporcionar un rol que tenga los read/write permisos de Amazon S3 necesarios en esa ubicación. En lugar de crear un rol con los permisos de Amazon S3 que se requieren, puede utilizar este rol vinculado con un servicio.

La primera vez que nombre el rol vinculado al servicio como rol con el que registrar una ruta, el rol vinculado al servicio y una nueva política de IAM se crearán en su nombre. Lake Formation añade la ruta a la política insertada y la adjunta al rol vinculado al servicio. Cuando registre rutas posteriores con el rol vinculado al servicio, Lake Formation añade la ruta a la política existente.

Inicie sesión como administrador del lago de datos y registre una ubicación del lago de datos. A continuación, en la consola de IAM, busque el rol `AWSServiceRoleForLakeFormationDataAccess` y vea sus políticas adjuntas.

Por ejemplo, después de registrar la ubicación `s3://my-kinesis-test/logs`, Lake Formation crea la siguiente política en línea y la adjunta a `AWSServiceRoleForLakeFormationDataAccess`.

JSON

```
{
```

```
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3",
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject",
                "s3:AbortMultipartUpload",
                "s3:ListMultipartUploadParts"
            ],
            "Resource": [
                "arn:aws:s3:::my-kinesis-test/logs/*"
            ]
        },
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3ListBucket",
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket",
                "s3:ListBucketMultipartUploads"
            ],
            "Resource": [
                "arn:aws:s3:::my-kinesis-test"
            ]
        }
    ]
}
```

Creación de un rol vinculado a servicio para Lake Formation

No necesita crear manualmente un rol vinculado a servicios. Cuando registra una ubicación de Amazon S3 con Lake Formation en la Consola de administración de AWS, la AWS CLI o la AWS API, Lake Formation crea el rol vinculado al servicio para usted.

> **Important**
>
> Este rol vinculado a servicios puede aparecer en su cuenta si se ha completado una acción en otro servicio que utilice las características compatibles con este rol. Para obtener más información, consulte Un nuevo rol ha aparecido en mi cuenta de IAM.

Si elimina este rol vinculado a servicios y necesita crearlo de nuevo, puede utilizar el mismo proceso para volver a crear el rol en su cuenta. Al registrar una ubicación de Amazon S3 en Lake Formation, Lake Formation crea automáticamente el rol vinculado al servicio.

También puede utilizar la consola de IAM para crear un rol vinculado a servicio con el caso de uso de Lake Formation. En la AWS CLI o en la AWS API, cree una función vinculada al servicio con el nombre del servicio. `lakeformation.amazonaws.com` Para obtener más información, consulte Crear un rol vinculado a un servicio en la Guía del usuario de IAM. Si elimina este rol vinculado al servicio, puede utilizar este mismo proceso para volver a crear el rol.

Edición de un rol vinculado a servicio para Lake Formation

Lake Formation no le permite editar el rol vinculado a servicio `AWSServiceRoleForLakeFormationDataAccess`. Después de crear un rol vinculado al servicio, no podrá cambiar el nombre del rol, ya que varias entidades podrían hacer referencia al rol. Sin embargo, sí puede editar la descripción del rol con IAM. Para obtener más información, consulte Edición de un rol vinculado a servicios en la Guía del usuario de IAM.

Eliminación de un rol vinculado a servicio para Lake Formation

Si ya no necesita usar una característica o servicio que requieran un rol vinculado a un servicio, le recomendamos que elimine dicho rol. Así no tendrá una entidad no utilizada que no se supervise ni mantenga de forma activa. Sin embargo, debe limpiar los recursos de su rol vinculado al servicio antes de eliminarlo manualmente.

> **Note**
>
> Si el servicio de Lake Formation está utilizando el rol cuando intente eliminar los recursos, la eliminación podría producir un error. En tal caso, espere unos minutos e intente de nuevo la operación.

Para eliminar los recursos de Lake Formation utilizados por Lake Formation

- Si ha utilizado el rol vinculado al servicio para registrar las ubicaciones de Amazon S3 en Lake Formation, antes de eliminar el rol vinculado al servicio, debe anular el registro de la ubicación y volver a registrarla con un rol personalizado.

Para eliminar manualmente el rol vinculado a servicios mediante IAM

Utilice la consola de IAM AWS CLI, la o la AWS API para eliminar la función vinculada al `AWSServiceRoleForLakeFormationDataAccess` servicio. Para obtener más información, consulte Eliminación de un rol vinculado a servicios en la Guía del usuario de IAM.

Los siguientes son los requisitos para un rol definido por el usuario:

- Al crear el nuevo rol, en la página Crear rol de la consola de IAM, elija el Servicio de AWS y, a continuación, en Elija un caso de uso, seleccione Lake Formation.

  Si crea el rol utilizando una ruta diferente, asegúrese de que el rol tenga una relación de confianza con `lakeformation.amazonaws.com`. Para obtener más información, consulte Modificación de una política de confianza de rol (consola).
- El rol debe tener una política en línea que conceda read/write permisos a Amazon S3 en la ubicación. La siguiente es una política característica.

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::awsexamplebucket/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::awsexamplebucket"
            ]
        }
    ]
}
```

- Agregue la siguiente política de confianza al rol de IAM para permitir que el servicio de Lake Formation asuma el rol y dispense credenciales temporales a los motores de análisis integrados.

  Para incluir el contexto de usuario del IAM Identity Center en los CloudTrail registros, la política de confianza debe tener el permiso para realizar la `sts:SetContext` acción.

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerAssumeRole1",
            "Effect": "Allow",
            "Principal": {
               "Service": [
                    "lakeformation.amazonaws.com"
                ]
            },
            "Action": [
                "sts:AssumeRole",
                "sts:SetContext"
            ]
        }
    ]
}
```

- El administrador del lago de datos que registra la ubicación debe tener el permiso `iam:PassRole` para el rol.

  La siguiente es una política insertada que concede este permiso. *<account-id>*Sustitúyalo por un número de AWS cuenta válido y *<role-name>* sustitúyalo por el nombre del rol.

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "PassRolePermissions",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
```

```
            ],
            "Resource": [
                "arn:aws:iam::111122223333:role/<role-name>"
            ]
        }
    ]
}
```

- Para permitir que Lake Formation añada CloudWatch registros en Logs y publique métricas, añada la siguiente política en línea.

> **Note**
>
> Escribir en CloudWatch Logs conlleva un cargo.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "Sid1",
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogStream",
                "logs:CreateLogGroup",
                "logs:PutLogEvents"
            ],
            "Resource": [
                 "arn:aws:logs:us-east-1:111122223333:log-group:/aws-
lakeformation-acceleration/*",
                 "arn:aws:logs:us-east-1:111122223333:log-group:/aws-
lakeformation-acceleration/*:log-stream:*"
            ]
        }
    ]
}
```

Registro de una ubicación de Amazon S3

Debe especificar un rol AWS Identity and Access Management (IAM) al registrar una ubicación de Amazon Simple Storage Service (Amazon S3). Lake Formation asume esa función cuando otorga credenciales temporales a los AWS servicios integrados que acceden a los datos en esa ubicación.

> **Important**
>
> Evite registrar un bucket de Amazon S3 que tenga activada la opción El solicitante paga. Para los buckets registrados en Lake Formation, el rol utilizado para registrar el bucket se considera siempre como el solicitante. Si otra AWS cuenta accede al depósito, se le cobrará al propietario del depósito por el acceso a los datos si el rol pertenece a la misma cuenta que el propietario del depósito.

Puede usar la AWS Lake Formation consola, la API de Lake Formation o AWS Command Line Interface (AWS CLI) para registrar una ubicación de Amazon S3.

Antes de empezar

Revise los requisitos del rol utilizado para registrar la ubicación.

Para registrar una ubicación (consola)

> **Important**
>
> Los siguientes procedimientos suponen que la ubicación de Amazon S3 está en la misma AWS cuenta que el catálogo de datos y que los datos de la ubicación no están cifrados. Otras secciones de este capítulo tratan sobre el registro entre cuentas y el registro de ubicaciones cifradas.

1. Abra la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/. Inicie sesión como administrador del lago de datos o como usuario con el permiso de `lakeformation:RegisterResource` IAM.
2. En el panel de navegación, en Administración, seleccione Ubicaciones de los lagos de datos.
3. Elija Registrar ubicación y, a continuación, seleccione Examinar para seleccionar una ruta de Amazon Simple Storage Service (Amazon S3).

4. (Opcional, pero muy recomendable) Seleccione Revisar permisos de ubicación para ver una lista de todos los recursos existentes en la ubicación de Amazon S3 seleccionada y sus permisos.

   El registro de la ubicación seleccionada podría dar lugar a que sus usuarios de Lake Formation accedan a los datos que ya se encuentran en esa ubicación. Revisar esta lista ayuda a garantizar que los datos existentes permanecen seguros.

5. Para el rol de IAM, elija el rol vinculado al servicio `AWSServiceRoleForLakeFormationDataAccess` (valor predeterminado) o un rol de IAM personalizado que cumpla los requisitos de the section called “Requisitos de los roles utilizados para registrar ubicaciones”.

   Puede actualizar una ubicación registrada u otros detalles solo si la registra con un rol de IAM personalizado. Para editar una ubicación registrada con un rol vinculado a un servicio, debe anular el registro de la ubicación y volver a registrarla.

6. Elija la opción Habilitar la federación de catálogos de datos para permitir que Lake Formation asuma un rol y venda credenciales temporales a AWS servicios integrados para acceder a las tablas de bases de datos federadas. Si una ubicación está registrada en Lake Formation y desea utilizar la misma ubicación para una tabla en una base de datos federada, deberá registrar la misma ubicación con la opción Habilitar federación del Catálogo de datos.

7. Seleccione el modo de acceso híbrido para no habilitar los permisos de Lake Formation de forma predeterminada. Cuando registre la ubicación de Amazon S3 en modo de acceso híbrido, puede habilitar los permisos de Lake Formation optando por entidades principales para las bases de datos y las tablas bajo esa ubicación.

   Para más información sobre la configuración del modo de acceso híbrido, consulte Modo de acceso híbrido.

8. Elija Registrar la ruta S3 con la cuenta de propietario y proporcione la cuenta de propietario del bucket de S3 para que Lake Formation pueda verificar el acceso del rol de registro a la ubicación de Amazon S3 y permitir que Lake Formation venda credenciales temporales para los archivos de Amazon S3 en la ubicación registrada.

9. Seleccione Registrar ubicación.

Para registrar una ubicación (AWS CLI)

1. Registro de una nueva ubicación en Lake Formation

   Este ejemplo usa el rol vinculado a un servicio para registrar la ubicación. En su lugar, puede utilizar el argumento `--role-arn` para proporcionar su propio rol.

   *<s3-path>*Sustitúyalo por una ruta de Amazon S3 válida, el número de AWS cuenta por una cuenta válida y *<s3-access-role>* por un rol de IAM que tenga permisos para registrar una ubicación de datos.

   > **Note**
   >
   > No puede editar las propiedades de una ubicación registrada si está registrada con un rol vinculado a un servicio.

   ```
   aws lakeformation register-resource \
    --resource-arn arn:aws:s3:::<s3-path> \
    --use-service-linked-role
   ```

   En el siguiente ejemplo, se utiliza un rol personalizado para registrar la ubicación.

   ```
   aws lakeformation register-resource \
    --resource-arn arn:aws:s3:::<s3-path> \
    --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role>
   ```

2. Para actualizar una ubicación registrada con Lake Formation

   Puede editar una ubicación registrada solo si está registrada con un rol de IAM personalizado. en el caso de una ubicación registrada con un rol vinculado a un servicio, debe anular el registro de la ubicación y volver a registrarla. Para obtener más información, consulte the section called “Dar de baja el registro de una ubicación de Amazon S3”.

   ```
   aws lakeformation update-resource \
    --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role>\
    --resource-arn arn:aws:s3:::<s3-path>
   ```

```
aws lakeformation update-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --use-service-linked-role
```

3. Registrar una ubicación de datos en modo de acceso híbrido con federación

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
 --hybrid-access-enabled
```

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
 --with-federation
```

```
aws lakeformation update-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
 --hybrid-access-enabled
```

4. Registre una ubicación con la cuenta de propietario prevista

   Utilice la `--expected-resource-owner-account` marca para permitir que Lake Formation verifique el acceso del rol de registro a la ubicación de Amazon S3 y venda credenciales temporales para los archivos de Amazon S3 en la ubicación registrada.

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
 --expected-resource-owner-account <123456789012>
```

```
aws lakeformation update-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
 --expected-resource-owner-account <123456789012>
```

Para obtener más información, consulte el funcionamiento de RegisterResourcela API.

> **Note**
>
> Una vez que registre una ubicación de Amazon S3, cualquier AWS Glue tabla que apunte a la ubicación (o a cualquiera de sus ubicaciones secundarias) devolverá el valor del `IsRegisteredWithLakeFormation` parámetro tal y como `true` aparece en la `GetTable` llamada. Existe una limitación conocida por la que las operaciones de la API del Catálogo de datos como `GetTables` y `SearchTables` no actualizan el valor del parámetro `IsRegisteredWithLakeFormation` y devuelven el valor predeterminado, que es falso. Se recomienda utilizar la API`GetTable` para ver el valor correcto del parámetro `IsRegisteredWithLakeFormation`.

Registro de una ubicación cifrada de Amazon S3

Lake Formation se integra con AWS Key Management Service (AWS KMS) para permitirle configurar más fácilmente otros servicios integrados para cifrar y descifrar datos en ubicaciones de Amazon Simple Storage Service (Amazon S3).

Tanto las gestionadas por AWS KMS keys el cliente como Claves administradas por AWS las compatibles. Actualmente, el lado del cliente solo encryption/decryption es compatible con Athena.

Debe especificar un rol AWS Identity and Access Management (de IAM) al registrar una ubicación de Amazon S3. En el caso de las ubicaciones cifradas de Amazon S3, el rol debe tener permiso para cifrar y descifrar datos con el AWS KMS key, o bien la política de claves de KMS debe conceder permisos sobre la clave del rol.

> **Important**
>
> Evite registrar un bucket de Amazon S3 que tenga activada la opción El solicitante paga. Para los buckets registrados en Lake Formation, el rol utilizado para registrar el bucket se considera siempre como el solicitante. Si otra AWS cuenta accede al bucket, se le cobrará al propietario del bucket por el acceso a los datos si el rol pertenece a la misma cuenta que el propietario del bucket.

Lake Formation usa un rol vinculado a servicio para registrar las ubicaciones de los datos. Sin embargo, este rol tiene varias limitaciones. Debido a estas limitaciones, recomendamos crear y

usar un rol de IAM personalizado en su lugar, para mayor flexibilidad y control. El rol personalizado que cree para registrar la ubicación debe cumplir los requisitos especificados en the section called "Requisitos de los roles utilizados para registrar ubicaciones".

> **Important**
>
> Si ha utilizado una Clave administrada de AWS para cifrar la ubicación de Amazon S3, no puede utilizar la función vinculada al servicio Lake Formation. Debe usar un rol personalizado y añadir permisos de IAM a la clave del rol. Los detalles se proporcionan más adelante en esta sección.

Los siguientes procedimientos explican cómo registrar una ubicación de Amazon S3 cifrada con una clave administrada por el cliente o una Clave administrada de AWS.

- Registrar una ubicación cifrada con una clave administrada por el cliente
- Registrar una ubicación cifrada con un Clave administrada de AWS

Antes de empezar

Revise los requisitos del rol utilizado para registrar la ubicación.

Para registrar una ubicación de Amazon S3 cifrada con una clave administrada por el cliente

> **Note**
>
> Si la clave de KMS o la ubicación de Amazon S3 no están en la misma AWS cuenta que el catálogo de datos, siga las instrucciones que se indican en the section called "Registrar una ubicación de Amazon S3 cifrada en AWS cuentas" su lugar.

1. Abra la AWS KMS consola https://console.aws.amazon.com/kmse inicie sesión como usuario administrativo AWS Identity and Access Management (IAM) o como usuario que puede modificar la política de claves de la clave de KMS utilizada para cifrar la ubicación.
2. En el panel de navegación, elija Claves administradas por el cliente y, a continuación, el nombre de la clave de KMS deseada.

3. En la página de detalles de la clave KMS, elija la pestaña Política de claves y, a continuación, siga una de las instrucciones siguientes para añadir su rol personalizado o el rol vinculado al servicio de Lake Formation como usuario de la clave KMS:

   - Si se muestra la vista predeterminada (con las secciones Administradores clave, Eliminación de claves, Usuarios clave y Otras AWS cuentas), en la sección Usuarios clave, agregue su rol personalizado o el rol vinculado al servicio Lake Formation. `AWSServiceRoleForLakeFormationDataAccess`
   - Si se muestra la política de claves (JSON). Edite la política para añadir su rol personalizado o el rol `AWSServiceRoleForLakeFormationDataAccess` vinculado al servicio de Lake Formation al objeto "Permitir el uso de la clave", como se muestra en el siguiente ejemplo.

   > **Note**
   >
   > Si falta ese objeto, agréguelo con los permisos que se muestran en el ejemplo. El ejemplo utiliza el rol vinculado al servicio.

```
        ...
        {
            "Sid": "Allow use of the key",
            "Effect": "Allow",
            "Principal": {
                "AWS": [
                    "arn:aws:iam::111122223333:role/aws-service-role/
lakeformation.amazonaws.com/AWSServiceRoleForLakeFormationDataAccess",
                    "arn:aws:iam::111122223333:user/keyuser"
                ]
            },
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:DescribeKey"
            ],
            "Resource": "*"
        },
        ...
```

4. Abra la AWS Lake Formation consola en. https://console.aws.amazon.com/lakeformation/ Inicie sesión como administrador del lago de datos o como usuario con el permiso de `lakeformation:RegisterResource` IAM.
5. En el panel de navegación, bajo Administración, seleccione Ubicaciones de los lagos de datos.
6. Elija Registrar ubicación y, a continuación, seleccione Examinar para seleccionar una ruta de Amazon Simple Storage Service (Amazon S3).
7. (Opcional, pero muy recomendable) Seleccione Revisar permisos de ubicación para ver una lista de todos los recursos existentes en la ubicación de Amazon S3 seleccionada y sus permisos.

   El registro de la ubicación seleccionada podría dar lugar a que sus usuarios de Lake Formation accedan a los datos que ya se encuentran en esa ubicación. Revisar esta lista ayuda a garantizar que los datos existentes permanecen seguros.
8. Para el rol de IAM, elija el rol vinculado al servicio `AWSServiceRoleForLakeFormationDataAccess` (el predeterminado) o su rol personalizado que cumpla con el the section called “Requisitos de los roles utilizados para registrar ubicaciones”.
9. Seleccione Registrar ubicación.

Para obtener más información sobre el rol vinculado a servicios, consulte Permisos de rol vinculados al servicio para Lake Formation.

Para registrar una ubicación de Amazon S3 cifrada con un Clave administrada de AWS

> **Important**
>
> Si la ubicación de Amazon S3 no está en la misma AWS cuenta que el catálogo de datos, siga las instrucciones que se indican en the section called “Registrar una ubicación de Amazon S3 cifrada en AWS cuentas” su lugar.

1. Cree un rol de IAM que se utilizará para registrar la ubicación. Asegúrese de que cumple los requisitos que figuran en the section called “Requisitos de los roles utilizados para registrar ubicaciones”.
2. Añada la siguiente política insertada al rol. Concede permisos sobre la clave al rol. La especificación `Resource` debe designar el nombre de recurso de Amazon (ARN) del Clave administrada de AWS. Puede obtener el ARN desde la AWS KMS consola. Para obtener el

ARN correcto, asegúrese de iniciar sesión en la AWS KMS consola con la misma AWS cuenta y región Clave administrada de AWS que utilizó para cifrar la ubicación.

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
      ],
      "Resource": "arn:aws:kms:us-
east-1:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"
    }
  ]
}
```

Puede utilizar los alias de clave de KMS en lugar del ID de clave (`arn:aws:kms:region:account-id:key/alias/your-key-alias`).

Para obtener más información, consulte la AWS KMS sección Alias de la Guía para desarrolladores. AWS Key Management Service

3. Abra la AWS Lake Formation consola en. https://console.aws.amazon.com/lakeformation/ Inicie sesión como administrador del lago de datos o como usuario con el permiso de `lakeformation:RegisterResource` IAM.
4. En el panel de navegación, bajo Administración, seleccione Ubicaciones de los lagos de datos.
5. Elija Registrar ubicación y, a continuación, seleccione Examinar para seleccionar una ruta de Amazon S3.
6. (Opcional, pero muy recomendable) Seleccione Revisar permisos de ubicación para ver una lista de todos los recursos existentes en la ubicación de Amazon S3 seleccionada y sus permisos.

El registro de la ubicación seleccionada podría dar lugar a que sus usuarios de Lake Formation accedan a los datos que ya se encuentran en esa ubicación. Revisar esta lista ayuda a garantizar que los datos existentes permanecen seguros.

7. Para Rol de IAM, elija el rol que ha creado en el Paso 1.
8. Seleccione Registrar ubicación.

Registrar una ubicación de Amazon S3 en otra AWS inscrita

AWS Lake Formation le permite registrar ubicaciones AWS de Amazon Simple Storage Service (Amazon S3) en todas las cuentas. Por ejemplo, si AWS Glue Data Catalog está en la cuenta A, un usuario de la cuenta A puede registrar un bucket de Amazon S3 en la cuenta B.

Para registrar un bucket de Amazon S3 en la AWS cuenta B con un rol AWS Identity and Access Management (IAM) en la AWS cuenta A se requieren los siguientes permisos:

- El rol de la cuenta A debe conceder permisos sobre el bucket de la cuenta B.
- La política de bucket de la cuenta B debe conceder permisos de acceso al rol de la cuenta A.

> **Important**
>
> Evite registrar un bucket de Amazon S3 que tenga activada la opción El solicitante paga. Para los buckets registrados en Lake Formation, el rol utilizado para registrar el bucket se considera siempre como el solicitante. Si otra AWS cuenta accede al bucket, se le cobrará al propietario del bucket por el acceso a los datos si el rol pertenece a la misma cuenta que el propietario del bucket.
>
> No puede utilizar el rol vinculado al servicio Lake Formation para registrar una ubicación en otra cuenta. En su lugar, debe utilizar un rol definido por el usuario. El rol debe cumplir los requisitos de the section called “Requisitos de los roles utilizados para registrar ubicaciones”. Para obtener más información sobre el rol vinculado a servicios, consulte Permisos de rol vinculados al servicio para Lake Formation.

Antes de empezar

Revise los requisitos del rol utilizado para registrar la ubicación.

Para registrar una ubicación en otra AWS inscrita

> **Note**
>
> Si la ubicación está cifrada, siga en su lugar las instrucciones de the section called “Registrar una ubicación de Amazon S3 cifrada en AWS cuentas”.

El siguiente procedimiento supone que una entidad principal de la cuenta 1111-2222-3333, que contiene el Catálogo de datos, desea registrar el bucket `awsexamplebucket1` de Amazon S3, que se encuentra en la cuenta 1234-5678-9012.

1. En la cuenta 1111-2222-3333, inicie sesión en la consola de IAM Consola de administración de AWS y ábrala en. https://console.aws.amazon.com/iam/
2. Cree un nuevo rol o consulte uno existente que cumpla con los requisitos de the section called “Requisitos de los roles utilizados para registrar ubicaciones”. Asegúrese de que el rol concede permisos de Amazon S3 a`awsexamplebucket1`.
3. Abra la consola de Amazon S3 en https://console.aws.amazon.com/s3/. Inicie sesión con la cuenta 1234-5678-9012.
4. En la lista Nombre del bucket, seleccione el nombre del bucket, `awsexamplebucket1`.
5. Elija Permisos.
6. En la página Permisos, elija Política de bucket.
7. En el editor de políticas de bucket, pegue la política siguiente. Sustituya *<role-name>* por el nombre de su rol.

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::111122223333:role/<role-name>"
            },
            "Action":"s3:ListBucket",
            "Resource":"arn:aws:s3:::awsexamplebucket1"
```

```
        },
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::111122223333:role/<role-name>"
            },
            "Action": [
                "s3:DeleteObject",
                "s3:GetObject",
                "s3:PutObject"
            ],
            "Resource":"arn:aws:s3:::awsexamplebucket1/*"
        }
    ]
}
```

8. Seleccione Save.
9. Abre la consola en. AWS Lake Formation https://console.aws.amazon.com/lakeformation/ Inicie sesión en la cuenta 1111-2222-3333 como administrador del lago de datos o como usuario con permisos suficientes para registrar ubicaciones.
10. En el panel de navegación, bajo Administración, seleccione Ubicaciones de los lagos de datos.
11. En la página de Ubicaciones de los lagos de datos, seleccione Registrar ubicación.
12. En la página Registrar ubicación, para la ruta de Amazon S3, introduzca el nombre del bucket `s3://awsexamplebucket1`.

> **Note**
>
> Debe escribir el nombre del bucket porque los buckets entre cuentas no aparecen en la lista cuando selecciona Examinar.

13. En Rol de IAM, seleccione su rol.
14. Seleccione Registrar ubicación.

Registrar una ubicación de Amazon S3 cifrada en AWS cuentas

AWS Lake Formation se integra con AWS Key Management Service(AWS KMS) para permitirle configurar más fácilmente otros servicios integrados para cifrar y descifrar datos en las ubicaciones de Amazon Simple Storage Service (Amazon S3).

Ambas son claves administradas por el cliente y Claves administradas por AWS son compatibles. Client-side encryption/decryption no es compatible.

> **Important**
>
> Evite registrar un bucket de Amazon S3 que tenga activada la opción El solicitante paga. Para los buckets registrados en Lake Formation, el rol utilizado para registrar el bucket se considera siempre como el solicitante. Si otra AWS cuenta accede al depósito, se le cobrará al propietario del depósito por el acceso a los datos si el rol pertenece a la misma cuenta que el propietario del depósito.

En esta sección se explica cómo registrar una ubicación de Amazon S3 en las siguientes circunstancias:

- Los datos de la ubicación de Amazon S3 se cifran con una clave KMS creada en AWS KMS.
- La ubicación de Amazon S3 no está en la misma AWS cuenta que AWS Glue Data Catalog.
- La clave KMS está o no en la misma AWS cuenta que el catálogo de datos.

Para registrar un bucket de Amazon S3 AWS KMS cifrado en la AWS cuenta B con un rol AWS Identity and Access Management (IAM) en la AWS cuenta A se requieren los siguientes permisos:

- El rol de la cuenta A debe conceder permisos sobre el bucket de la cuenta B.
- La política de bucket de la cuenta B debe conceder permisos de acceso al rol de la cuenta A.
- Si la clave de KMS está en la cuenta B, la política de claves debe conceder el acceso al rol de la cuenta A y el rol de la cuenta A debe conceder permisos sobre la clave de KMS.

En el siguiente procedimiento, se crea un rol en la AWS cuenta que contiene el catálogo de datos (cuenta A en el análisis anterior). A continuación, utilice este rol para registrar la ubicación. Lake Formation asume este rol al acceder a los datos subyacentes en Amazon S3. El rol asumido tiene los permisos necesarios en la clave de KMS. Como resultado, no tendrá que conceder permisos sobre la clave KMS a las entidades principales que accedan a los datos subyacentes con trabajos ETL o con servicios integrados como Amazon Athena.

> **Important**
>
> No puede utilizar el rol vinculado al servicio Lake Formation para registrar una ubicación en otra cuenta. En su lugar, debe utilizar un rol definido por el usuario. El rol debe cumplir los requisitos de the section called “Requisitos de los roles utilizados para registrar ubicaciones”. Para obtener más información sobre el rol vinculado a servicios, consulte Permisos de rol vinculados al servicio para Lake Formation.

Antes de empezar

Revise los requisitos del rol utilizado para registrar la ubicación.

Para registrar una ubicación cifrada de Amazon S3 en AWS cuentas

1. En la misma AWS cuenta que el catálogo de datos, inicie sesión en la consola de IAM Consola de administración de AWS y ábrala enhttps://console.aws.amazon.com/iam/.
2. Cree un nuevo rol o consulte uno existente que cumpla con los requisitos de the section called “Requisitos de los roles utilizados para registrar ubicaciones”. Asegúrese de que el rol incluye una política que concede permisos de Amazon S3 en la ubicación.
3. Si la clave de KMS no está en la misma cuenta que el Catálogo de datos, añada al rol una política integrada que conceda los permisos necesarios para la clave de KMS. A continuación, se muestra una política de ejemplo. Sustituya Región e ID de cuenta por la región y el número de cuenta de la clave de KMS. *<key-id>*Sustitúyala por el identificador de clave.

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
        "Effect": "Allow",
        "Action": [
            "kms:Encrypt",
            "kms:Decrypt",
            "kms:ReEncrypt*",
            "kms:GenerateDataKey*",
            "kms:DescribeKey"
         ],
```

```
        "Resource": "arn:aws:kms:us-east-1:111122223333:key/<key-id>"
        }
    ]
}
```

4. En la consola de Amazon S3, agregue una política de bucket que conceda los permisos de Amazon S3 necesarios para el rol. A continuación se muestra un ejemplo de política de bucket. Sustituya el ID de AWS cuenta por el número de cuenta del catálogo *<role-name>* de datos, por el nombre de su función y *<bucket-name>* por el nombre del depósito.

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::111122223333:role/<role-name>"
            },
            "Action":"s3:ListBucket",
            "Resource":"arn:aws:s3:::<bucket-name>"
        },
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::111122223333:role/<role-name>"
            },
            "Action": [
                "s3:DeleteObject",
                "s3:GetObject",
                "s3:PutObject"
            ],
            "Resource":"arn:aws:s3:::<bucket-name>/*"
        }
    ]
}
```

5. En AWS KMS, agregue el rol como usuario de la clave KMS.

a. Abra la AWS KMS consola en https://console.aws.amazon.com/kms. A continuación, inicie sesión como usuario administrador o como usuario que puede modificar la política de claves de la clave de KMS utilizada para cifrar la ubicación.

b. En el panel de navegación, seleccione Claves administradas por el cliente y, a continuación, elija el nombre de la clave KMS.

c. En la página de detalles de la clave KMS, en la pestaña Política de claves, si no se muestra la vista JSON de la política de claves, seleccione Cambiar a la vista de políticas.

d. En la sección Política de claves, seleccione Editar y añada el Nombre de recurso de Amazon (ARN) del rol al objeto `Allow use of the key`, como se muestra en el siguiente ejemplo.

> **Note**
>
> Si falta ese objeto, agréguelo con los permisos que se muestran en el ejemplo.

```
...
{
    "Sid": "Allow use of the key",
    "Effect": "Allow",
    "Principal": {
        "AWS": [
            "arn:aws:iam::<catalog-account-id>:role/<role-name>"
        ]
    },
    "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
    ],
    "Resource": "*"
},
...
```

Para más información, consulte Permitir a los usuarios de otras cuentas utilizar una clave KMS en la Guía para desarrolladores de AWS Key Management Service .

6. Abra la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/. Inicie sesión en la cuenta AWS del Catálogo de datos como administrador del lago de datos.
7. En el panel de navegación, bajo Administración, seleccione Ubicaciones de los lagos de datos.
8. Seleccione Registrar ubicación.
9. En la página Registrar ubicación, para la ruta de Amazon S3, introduzca el nombre del bucket como **`s3://<bucket>/<prefix>`**. *`<bucket>`*Sustitúyalo por el nombre del depósito y *`<prefix>`* por el resto de la ruta de la ubicación.

   > **Note**
   >
   > Debe escribir la ruta porque los buckets entre cuentas no aparecen en la lista cuando selecciona Examinar.

10. Para el rol de IAM, elija el rol del paso 2.
11. Seleccione Registrar ubicación.

Dar de baja el registro de una ubicación de Amazon S3

Puede anular el registro de una ubicación de Amazon Simple Storage Service (Amazon S3) si desea que Lake Formation deje de administrarla. Dar de baja una ubicación no afecta a los permisos de ubicación de datos de Lake Formation que se conceden sobre esa ubicación. Puede volver a registrar una ubicación que haya dado de baja y los permisos de ubicación de datos seguirán vigentes. Puede utilizar un rol diferente para volver a registrar la ubicación.

Para dar de baja una ubicación (consola)

1. Abra la AWS Lake Formation consola en. https://console.aws.amazon.com/lakeformation/ Inicie sesión como administrador del lago de datos o como usuario con el permiso de `lakeformation:RegisterResource` IAM.
2. En el panel de navegación, bajo Administración, seleccione Ubicaciones de los lagos de datos.
3. Seleccione una ubicación y, en el menú Acciones, seleccione Editar.
4. Cuando se le indique que confirme, elija Quitar.

Acceso a las ubicaciones de Amazon S3

AWS Lake Formation le permite leer y escribir los archivos de datos subyacentes en Amazon Simple Storage Service (Amazon S3) para las tablas registradas en AWS Glue Data Catalog el (catálogo de datos). Esto le proporciona un conjunto único de permisos tanto para las consultas SQL como para el acceso directo a los archivos mediante las concesiones de la tabla Lake Formation existentes. Lake Formation amplía su funcionalidad de venta de credenciales a las ubicaciones de Amazon S3 registradas en el catálogo de datos. Cuando su aplicación solicita acceso a los archivos de Amazon S3 subyacentes a una tabla del catálogo de datos, Lake Formation evalúa los permisos a nivel de tabla existentes de la persona que llama y, si se autoriza, devuelve las credenciales de Amazon S3 de corta duración y alcance para la ubicación registrada de esa tabla.

Habilite el acceso directo a las ubicaciones de Amazon S3

Para habilitar esta función, debe completar los siguientes pasos.

Requisitos previos

Complete los siguientes requisitos previos antes de activar el indicador de función:

- Especifique la ubicación de la tabla: especifique la ruta a sus datos en Amazon S3 mediante la propiedad LOCATION de los atributos de la tabla. Para obtener más información sobre el formato de ubicación, consulte la ubicación de la tabla en Amazon S3 en la Guía del usuario de Amazon Athena.
- Registre las ubicaciones de Amazon S3: catalogue sus tablas en el catálogo de datos de sus ubicaciones de Amazon S3 y regístrelas en Lake Formation proporcionando la cuenta de propietario del bucket de Amazon S3. Al registrar una ubicación, se registran esa ruta de Amazon S3 y todas las carpetas incluidas en esa ruta. Para obtener más información, consulte Registro de una ubicación de Amazon S3.
- Otorgue permisos de Lake Formation: conceda permisos de Lake Formation `SELECT` (o`SUPER`) sobre tablas a sus científicos de datos y aplicaciones mediante la consola o las API de Lake Formation. Si ya utiliza Lake Formation para acceder a las mesas con servicios como Athena o Amazon EMR, ya está configurado.
- Habilite la integración de las aplicaciones con acceso total a las tablas: habilite la integración de las aplicaciones con acceso total a las tablas para permitir la venta de credenciales en la ubicación de la mesa registrada. Para obtener más información, consulte Integración de aplicaciones para un acceso completo a la tabla.

Habilite el indicador de función

Activa la `fs.s3a.lakeformation.access.grants.enabled` marca en los motores compatibles:

```
fs.s3a.lakeformation.access.grants.enabled = true
```

- Amazon EMR en Amazon EC2: EMR 7.13 o posterior
- Amazon EMR en EKS: EMR 7.13 o posterior
- Amazon EMR Serverless: EMR 7.13 o posterior

Una vez configurado, puede leer o escribir inmediatamente los archivos de datos de EMR mediante las API estándar.

Para obtener más información sobre la integración de Amazon EMR, consulte la venta de credenciales basada en la ruta de Lake Formation en la Guía de administración de Amazon EMR.

Funcionamiento

El acceso a la ubicación de Lake Formation-based Amazon S3 sigue este flujo:

1. Un responsable o un rol solicita acceso a los archivos de datos de Amazon S3 a través de un servicio, como un trabajo de Amazon EMR Spark o una canalización de procesamiento de datos.
2. El complemento AWS SDK integrado con EMR intercepta la solicitud de Amazon S3 y llama a la API Lake Formation. GetTemporaryDataLocationCredentials
3. Lake Formation comprueba si la ruta de Amazon S3 corresponde a una ubicación registrada y a una tabla catalogada, y evalúa los permisos de la persona que llama en la tabla asociada. AWS Glue
4. Si la persona que llama tiene `SELECT SUPER` permisos en la tabla correspondiente con acceso total a la mesa, Lake Formation vende las credenciales temporales relacionadas con esa ubicación de Amazon S3. `SELECT`otorga credenciales de lectura; **`SUPER`** otorga credenciales de lectura y escritura.
5. El complemento utiliza estas credenciales para completar la solicitud de Amazon S3 y Amazon S3 proporciona los datos a la aplicación.
6. Todas las operaciones de venta de credenciales se registran CloudTrail, lo que proporciona un registro de auditoría para el acceso a los datos.

Este proceso es transparente para usted y simplemente utilizará los permisos existentes de Lake Formation con las API estándar basadas en archivos.

```
# Read raw data (Lake Formation-based S3 Location access)
raw_df = spark.read.json("s3://finance-datalake/raw/transactions/dt=2024-03-21/")

# Read governed data (Lake Formation-based S3 Location access)
transactions_df = spark.read.parquet("s3://data-lake/transactions/year=2026/")

# Write processed data (Lake Formation-based S3 Location access)
processed_df.write \
    .mode("append") \
    .partitionBy("transaction_date") \
    .parquet("s3://finance-datalake/processed/transactions/")

print("ETL complete. Records written:", processed_df.count())
```

Requisitos del permiso

Actualmente, la función solo vende las credenciales cuando la persona que llama tiene acceso completo a la tabla, es decir, SELECT en todas las columnas sin filtros de filas o columnas. Si a una tabla se le han aplicado filtros a nivel de fila o de columna, las personas que llaman deben seguir utilizando motores de confianza como Athena, Amazon EMR AWS Glue o Amazon Redshift, que pueden aplicar esos filtros. Esto garantiza que los límites de seguridad se mantengan uniformes.

Cross-account acceso

El acceso simplificado de Lake Formation a las ubicaciones de Amazon S3 funciona con capacidades de uso compartido entre cuentas. Al compartir una tabla con otra AWS cuenta a través de Lake Formation, los destinatarios pueden acceder a los archivos de datos subyacentes de Amazon S3 con sus permisos de Lake Formation, sujetos a los mismos requisitos de acceso completo a la tabla.

La función es compatible con el mecanismo de enlace de recursos de Lake Formation para el acceso entre cuentas. Cuando una cuenta de consumidor tiene permisos en una mesa compartida, Lake Formation vende las credenciales relacionadas con la ubicación de Amazon S3 de la cuenta registrada, lo que permite un acceso fluido a los datos entre cuentas sin necesidad de políticas de bucket de Amazon S3 independientes ni funciones de IAM entre cuentas.

Ubicaciones anidadas de Amazon S3

Cuando varias tablas apuntan a ubicaciones anidadas en el mismo cubo, Lake Formation aplica el siguiente comportamiento:

- Al acceder`s3://bucket`, recibirá los permisos correspondientes a la tabla registrada en el nivel del cubo.
- Al acceder`s3://bucket/folder1`, recibirá los permisos correspondientes a la tabla registrada en esa ruta específica.
- Para acceder a carpetas sin una tabla registrada (por ejemplo,`s3://bucket/folder2`), recibirá los permisos de la ubicación principal registrada más cercana.
- Si hay varias tablas registradas en la misma ubicación, Lake Formation devuelve un error debido a un conflicto de permisos.

Auditoría y conformidad

Se registran todas las operaciones de venta de credenciales CloudTrail, lo que proporciona un registro de auditoría para el acceso a los datos. Cuando Lake Formation vende credenciales a través de la `GetTemporaryDataLocationCredentials` API, CloudTrail registra:

- El principal (usuario o rol)
- Timestamp
- Ubicación de Amazon S3
- AWS GlueTabla asociada

Las llamadas posteriores a la API de Amazon S3 realizadas con esas credenciales también se registran CloudTrail como eventos de datos de Amazon S3, y el contexto las vincula de nuevo a la concesión del permiso de Lake Formation. Esto proporciona a los auditores una visibilidad completa de quién accedió a qué datos, cuándo y a través de qué ruta de permisos, todo ello en un único flujo de CloudTrail registro.

compatible AWS servicios

El acceso simplificado de Lake Formation a las ubicaciones de Amazon S3 funciona con:

- Amazon EMR en Amazon EC2 (EMR 7.13 o posterior)
- Amazon EMR en EKS (EMR 7.13 o posterior)
- Amazon EMR Serverless (EMR 7.13 o posterior)

Complemento de código abierto para servicios de terceros

También puede integrar sus Apache Spark Trino aplicaciones mediante API o mediante un complemento de código abierto proporcionado por AWS. Para obtener más información, consulte aws-lakeformation-accessgrants-plugin-java-v2 en. GitHub

Consideraciones

Tenga en cuenta las siguientes consideraciones al utilizar la venta de credenciales para las ubicaciones de Amazon S3:

- No se admite la venta de credenciales en las ubicaciones de Amazon S3 entre regiones.
- Se admite la venta de credenciales en las ubicaciones de Amazon S3 incluidas como ubicación principal de la tabla.
- El complemento admite Apache Hive los formatos de Delta Lake tabla y Apache Hudi los formatos de tablas. Apache Icebergno es compatible actualmente.
- Actualmente, el complemento no es compatible con el modo Amazon EMR Spark Fine-Grained Access Control (FGAC).

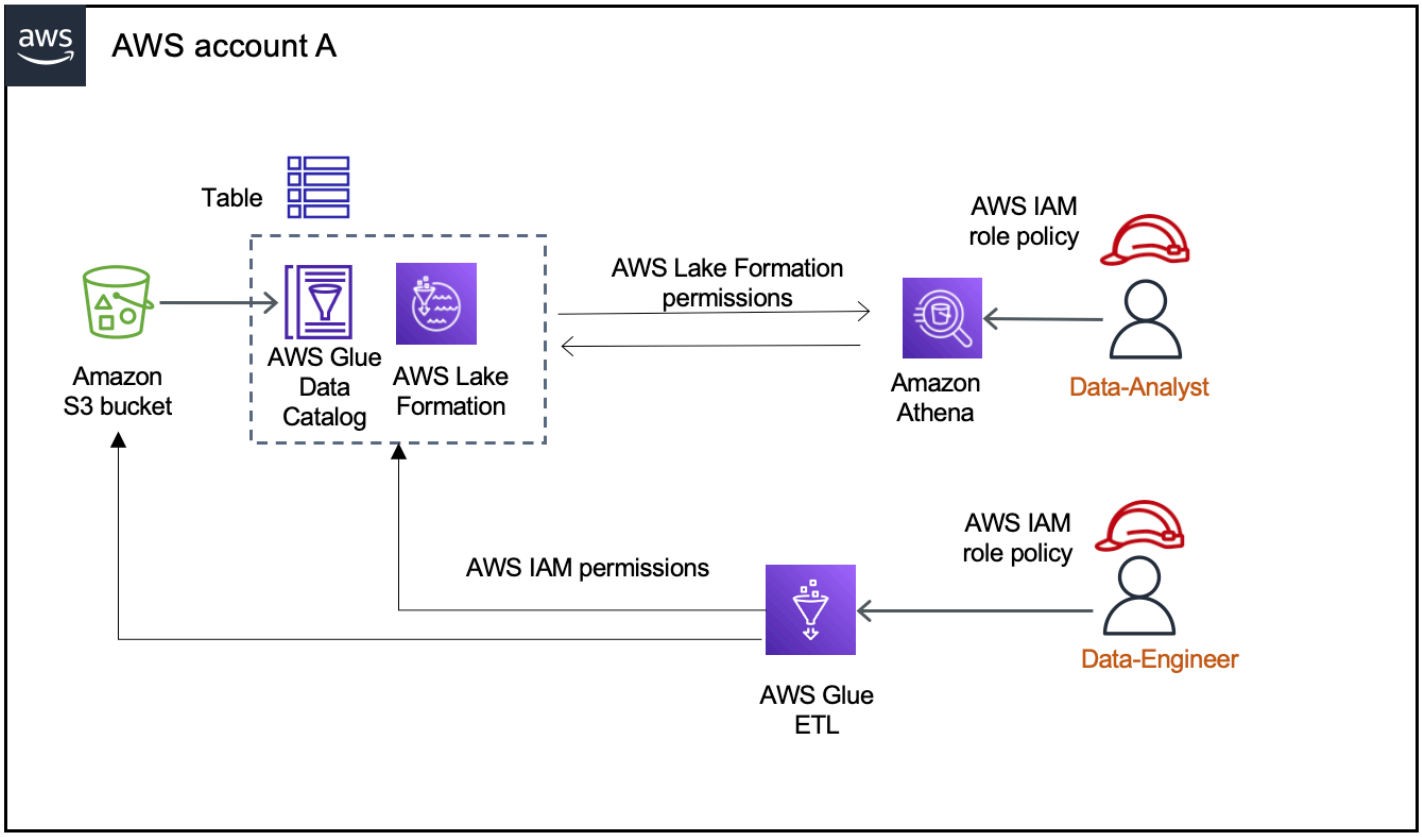
Modo de acceso híbrido

AWS Lake Formation el modo de acceso híbrido admite dos rutas de permiso para los mismos AWS Glue Data Catalog objetos.
En la primera, Lake Formation permite seleccionar entidades principales específicas y concederles permisos de Lake Formation para acceder a catálogos, bases de datos, tablas y vistas mediante inscripción. La segunda vía permite a todos los demás responsables acceder a estos recursos a través de las políticas principales de IAM predeterminadas para Amazon S3 y AWS Glue sus acciones.

Al registrar una ubicación de Amazon S3 en Lake Formation, tiene la opción de aplicar los permisos de Lake Formation para todos los recursos de esta ubicación o utilizar el modo de acceso híbrido. El modo de acceso híbrido solo aplica los permisos `CREATE_TABLE`, `CREATE_PARTITION` y `UPDATE_TABLE` de forma predeterminada. Cuando una ubicación de Amazon S3 se encuentra en el modo híbrido, puede habilitar los permisos de Lake Formation inscribiendo entidades principales para los objetos del Catálogo de datos de esa ubicación. Esto significa que tanto los permisos de Lake Formation como los permisos de IAM pueden controlar el acceso a esos datos. Esto significa que las entidades principales inscritas necesitarán tanto los permisos de Lake Formation como

los permisos de IAM para acceder a los datos, mientras que las entidades principales no inscritas seguirán accediendo a los datos únicamente con los permisos de IAM.

Así, el modo de acceso híbrido proporciona la flexibilidad necesaria para habilitar de forma selectiva Lake Formation para catálogos, bases de datos y tablas de su Catálogo de datos para un conjunto específico de usuarios sin interrumpir el acceso para otros usuarios o cargas de trabajo existentes.



Para ver las consideraciones y limitaciones, consulte [Consideraciones y limitaciones del modo de acceso híbrido](#).

Términos y definiciones

A continuación se detallan las definiciones de los recursos del Catálogo de datos según la configuración de los permisos de acceso:

Recurso de Lake Formation

> Un recurso registrado con Lake Formation. Los usuarios necesitan permisos de Lake Formation para acceder al recurso.

AWS Glue recurso

Un recurso que no está registrado con Lake Formation. Los usuarios solo necesitan permisos de IAM para acceder al recurso porque tiene permisos de grupo `IAMAllowedPrincipals`. Los permisos de Lake Formation no se aplican.

Para obtener más información sobre los permisos de grupo de `IAMAllowedPrincipals`, consulte Permisos de metadatos.

Recursos híbridos

Recursos registrados en modo de acceso híbrido. En función de los usuarios que accedan al recurso, este cambia dinámicamente entre ser un recurso de Lake Formation o un recurso de AWS Glue .

Casos de uso comunes del modo de acceso híbrido

Puede utilizar el modo de acceso híbrido para proporcionar acceso en escenarios de uso compartido de datos de una sola cuenta y entre cuentas:

Escenarios de una sola cuenta

- Convertir un AWS Glue recurso en un recurso híbrido: en este escenario, actualmente no está utilizando Lake Formation, pero desea adoptar los permisos de Lake Formation para los objetos del catálogo de datos. Al registrar la ubicación de Amazon S3 en modo de acceso híbrido, puede conceder permisos de Lake Formation a los usuarios que opten por acceder a bases de datos y tablas específicas que apunten a esa ubicación.
- Convertir un recurso de Lake Formation en un recurso híbrido: actualmente, utiliza los permisos de Lake Formation para controlar el acceso a una base de datos del Catálogo de datos, pero desea proporcionar acceso a nuevas entidades principales mediante los permisos de IAM para Amazon S3 y AWS Glue sin interrumpir los permisos de Lake Formation existentes.

  Al actualizar el registro de una ubicación de datos al modo de acceso híbrido, las nuevas entidades principales pueden acceder a la base de datos del Catálogo de datos que apunta a la ubicación de Amazon S3 mediante las políticas de permisos de IAM sin interrumpir los permisos de Lake Formation de los usuarios existentes.

  Antes de actualizar el registro de ubicación de datos para habilitar el modo de acceso híbrido, primero debe seleccionar las entidades principales que acceden ahora al recurso con permisos de Lake Formation.

Esto tiene como objetivo evitar una posible interrupción del flujo de trabajo actual. También debe conceder permiso de `Super` sobre las tablas de la base de datos al grupo `IAMAllowedPrincipal`.

Cross-account escenarios de intercambio de datos

- Comparta AWS Glue recursos mediante el modo de acceso híbrido: en este escenario, la cuenta del productor tiene tablas en una base de datos que actualmente se comparten con una cuenta de consumidor mediante políticas de permisos de IAM para Amazon S3 y AWS Glue acciones. La ubicación de los datos de la base de datos no está registrada en Lake Formation.

  Antes de registrar la ubicación de los datos en el modo de acceso híbrido, debe actualizar la configuración de la versión entre cuentas a la versión 4. La versión 4 proporciona las nuevas políticas de AWS RAM permisos necesarias para compartir entre cuentas cuando el `IAMAllowedPrincipal` grupo tiene `Super` permiso sobre el recurso. Para los recursos con permisos de grupo `IAMAllowedPrincipal`, puede conceder permisos de Lake Formation a cuentas externas y optar por que utilicen los permisos de Lake Formation. El administrador del lago de datos de la cuenta receptora puede conceder permisos de Lake Formation a entidades principales de la cuenta y optar por ellos para aplicar los permisos de Lake Formation.
- Compartir recursos de Lake Formation utilizando el modo de acceso híbrido: en la actualidad, la cuenta de productor tiene tablas en una base de datos que se comparten con una cuenta de consumidor que aplica los permisos de Lake Formation. La ubicación de los datos de la base de datos no está registrada en Lake Formation.

  En este caso, puede actualizar el registro de ubicación de Amazon S3 al modo de acceso híbrido y compartir los datos de Amazon S3 y los metadatos del Catálogo de datos mediante las políticas de bucket de Amazon S3 y las políticas de recursos del Catálogo de datos con las entidades principales de la cuenta del consumidor. Debe volver a conceder los permisos existentes de Lake Formation y optar por las entidades principales antes de actualizar el registro de la ubicación de Amazon S3. También debe conceder permiso de `Super` sobre las tablas de la base de datos al grupo `IAMAllowedPrincipals`.

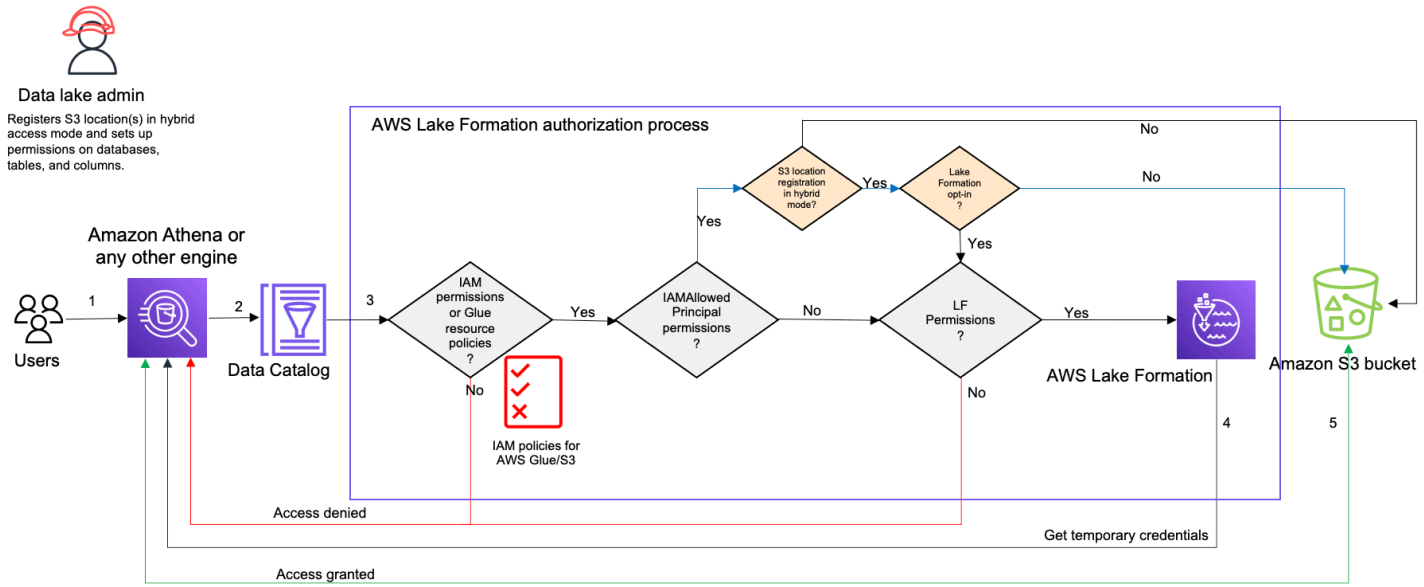
Temas

- Cómo funciona el modo de acceso híbrido
- Configuración del modo de acceso híbrido: escenarios comunes
- Eliminación de entidades principales y recursos del modo de acceso híbrido

- Eliminación de entidades principales y recursos del modo de acceso híbrido
- Recursos adicionales

Cómo funciona el modo de acceso híbrido

El diagrama siguiente muestra cómo funciona la autorización de Lake Formation en el modo de acceso híbrido al consultar los recursos del Catálogo de datos.



Antes de acceder a los datos del lago de datos, un administrador de este o un usuario con permisos administrativos configura políticas de usuario individuales de las tablas del Catálogo de datos para permitir o denegar el acceso a las tablas del Catálogo de datos. Luego, una entidad principal con permisos para la operación de `RegisterResource` registra en Lake Formation la ubicación de Amazon S3 de la tabla en modo de acceso híbrido. Si una ubicación de datos no está registrada con Lake Formation, el administrador concede permisos de Lake Formation a usuarios específicos sobre las bases de datos y tablas del Catálogo de datos y los inscribe para utilizar los permisos de Lake Formation para esas bases de datos y tablas en modo de acceso híbrido.

1. Envía una consulta: un director envía una consulta o un script de ETL mediante un servicio integrado como Amazon Athena, Amazon EMR o AWS Glue Amazon Redshift Spectrum.
2. Solicita datos: el motor analítico integrado identifica la tabla solicitada y envía la solicitud de metadatos al Catálogo de datos (`GetTable`, `GetDatabase`)
3. Comprueba los permisos: el Catálogo de datos verifica los permisos de acceso de la entidad principal que hace la consulta con Lake Formation.

a. Si la tabla no tiene permisos de grupo `IAMAllowedPrincipals` adjuntos, se aplican los permisos de Lake Formation.

b. Si la entidad principal ha optado por utilizar los permisos de Lake Formation en el modo de acceso híbrido y la tabla tiene adjuntos permisos de grupo `IAMAllowedPrincipals`, se aplicarán los permisos de Lake Formation. El motor de consulta aplica los filtros que recibió de Lake Formation y devuelve los datos al usuario.

c. Si la ubicación de la tabla no está registrada en Lake Formation y la entidad principal no ha optado por utilizar los permisos de Lake Formation en el modo de acceso híbrido, el Catálogo de datos comprueba si la tabla tiene permisos de grupo `IAMAllowedPrincipals` adjuntos. Si existe este permiso sobre la tabla, todas las entidades principales de la cuenta obtienen los permisos `Super` o `All` sobre la tabla.

   La venta de credenciales de Lake Formation no está disponible, incluso si se han inscrito, a menos que la ubicación de los datos esté registrada en Lake Formation.

4. Obtener credenciales: el Catálogo de datos comprueba y permite al motor saber si la ubicación de la tabla está registrada en Lake Formation o no. Si los datos subyacentes están registrados en Lake Formation, el motor analítico solicita a Lake Formation credenciales temporales para acceder a los datos del bucket de Amazon S3.

5. Obtener datos: si la entidad principal está autorizada para acceder a los datos de la tabla, Lake Formation proporciona acceso temporal al motor analítico integrado. Mediante el acceso temporal, el motor analítico obtiene los datos de Amazon S3 y aplica el filtrado necesario, como el de columnas, filas o celdas. Cuando el motor termina de ejecutar el trabajo, devuelve los resultados al usuario. Este proceso se denomina “expedición de credenciales”. Para obtener más información, consulte Integración con Lake Formation.

6. Si la ubicación de los datos de la tabla no está registrada en Lake Formation, la segunda llamada desde el motor analítico se hace directamente a Amazon S3. Para el acceso a los datos se evalúan la política de buckets de Amazon S3 y la política de usuarios de IAM correspondientes. Siempre que utilice políticas de IAM, compruebe que sigue las mejores prácticas IAM. Para obtener más información, consulte la sección Prácticas recomendadas de IAM en la Guía del usuario de IAM.

Configuración del modo de acceso híbrido: escenarios comunes

Al igual que con los permisos de Lake Formation, generalmente hay dos tipos de escenarios en los que puede usar el modo de acceso híbrido para administrar el acceso a los datos: proporcionar acceso a los directores dentro de uno Cuenta de AWS y proporcionar acceso a uno externo Cuenta de AWS o principal.

En esta sección se proporcionan instrucciones para configurar el modo de acceso híbrido en los escenarios siguientes:

Administre los permisos en el modo de acceso híbrido desde un solo lugar Cuenta de AWS

- Convertir un AWS Glue recurso en un recurso híbrido— Actualmente proporciona acceso a las tablas de una base de datos a todos los directores de su cuenta mediante permisos de IAM para Amazon S3, AWS Glue pero desea adoptar Lake Formation para administrar los permisos de forma incremental.

- Convertir un recurso de Lake Formation en un recurso híbrido: actualmente utiliza Lake Formation para administrar el acceso a las tablas de una base de datos para todas las entidades principales de su cuenta, pero desea utilizar Lake Formation solo para determinadas entidades principales. Desea proporcionar acceso a los nuevos directores mediante el uso de permisos de IAM para AWS Glue Amazon S3 en la misma base de datos y tablas.

Administre los permisos en modo de acceso híbrido en todas Cuenta de AWS s

- Compartir un AWS Glue recurso mediante el modo de acceso híbrido: actualmente no utiliza Lake Formation para gestionar los permisos de una tabla, pero desea aplicar los permisos de Lake Formation para proporcionar acceso a las entidades principales de otra cuenta.
- Compartir un recurso de Lake Formation mediante el modo de acceso híbrido— Utiliza Lake Formation para administrar el acceso a una tabla, pero desea proporcionar acceso a los directores de otra cuenta mediante permisos de IAM para Amazon S3 AWS Glue y en la misma base de datos y tablas.

Configuración del modo de acceso híbrido: pasos High-level

1. Registre la ubicación de datos de Amazon S3 en Lake Formation seleccionando el modo de acceso híbrido.

2. Las entidades principales deben tener permisos de `DATA_LOCATION` en una ubicación de lago de datos para crear tablas o bases de datos del Catálogo de datos que apunten a esa ubicación.
3. Establezca la Cross-account configuración de la versión en la versión 4.
4. Conceda permisos específicos a roles o usuarios de IAM concretos en bases de datos y tablas. Al mismo tiempo, asegúrese de establecer permisos de `Super` o `All` para el grupo `IAMAllowedPrincipals` de la base de datos y para todas las tablas de la base de datos o para algunas de ellas.
5. Elija las entidades principales y los recursos. Los demás responsables de la cuenta pueden seguir accediendo a las bases de datos y tablas mediante las políticas de permisos de IAM AWS Glue y las acciones de Amazon S3.
6. También puede limpie las políticas de permisos de IAM para Amazon S3 para las entidades principales que hayan optado por usar los permisos de Lake Formation.

Requisitos previos para configurar el modo de acceso híbrido

A continuación se detallan los requisitos previos para configurar el modo de acceso híbrido:

> **Note**
>
> Recomendamos que un administrador de Lake Formation registre la ubicación de Amazon S3 en modo de acceso híbrido y opte por entidades principales y recursos.

1. Otorgue el permiso de ubicación de datos (`DATA_LOCATION_ACCESS`) para crear recursos del Catálogo de datos que apunten a las ubicaciones de Amazon S3. Los permisos de ubicación de datos controlan la capacidad de crear catálogos, bases de datos y tablas del Catálogo de datos que apuntan a ubicaciones concretas de Amazon S3.
2. Para compartir los recursos del catálogo de datos con otra cuenta en modo de acceso híbrido (sin eliminar los permisos de `IAMAllowedPrincipals` grupo del recurso), debe actualizar la configuración de la versión multicuenta a la versión 4 o superior. Para actualizar la versión mediante la consola Lake Formation, elija la versión 4 o la versión 5 en la configuración de la versión entre cuentas en la página de configuración del catálogo de datos.

   También puede usar el `put-data-lake-settings` AWS CLI comando para establecer el `CROSS_ACCOUNT_VERSION` parámetro en la versión 4 o 5:

```
aws lakeformation put-data-lake-settings --region us-east-1 --data-lake-settings
 file://settings
{
"DataLakeAdmins": [
        {
"DataLakePrincipalIdentifier": "arn:aws:iam::<111122223333>:user/<user-name>"
        }
    ],
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "Parameters": {
"CROSS_ACCOUNT_VERSION": "5"
    }
}
```

3.

Para conceder permisos entre cuentas en el modo de acceso híbrido, el otorgante debe tener los permisos de IAM y los servicios necesarios. AWS Glue AWS RAM La política AWS gestionada `AWSLakeFormationCrossAccountManager` concede los permisos necesarios.
Para permitir el intercambio de datos entre cuentas en el modo de acceso híbrido, hemos actualizado la política administrada `AWSLakeFormationCrossAccountManager` añadiendo dos nuevos permisos de IAM:

- RAM: ListResourceSharePermissions
- RAM: AssociateResourceSharePermission

> **Note**
>
> Si no utilizas la política AWS gestionada para el rol de otorgante, añade las políticas anteriores a tus políticas personalizadas.

Ubicación del bucket de Amazon S3 y acceso de los usuarios

Al crear un catálogo, una base de datos o una tabla en AWS Glue Data Catalog, puede especificar la ubicación del depósito de Amazon S3 de los datos subyacentes y registrarlos en Lake Formation. En las tablas siguientes se describe cómo funcionan los permisos para AWS Glue los usuarios

(principales) de Lake Formation en función de la ubicación de datos de Amazon S3 de la tabla o base de datos.

Ubicación de Amazon S3 registrada con Lake Formation

Ubicación de Amazon S3 de una base de datos	AWS Glue usuarios	Usuarios de Lake Formation
Registro en Lake Formation (en modo de acceso híbrido o en modo Lake Formation)	Tenga read/write acceso a la ubicación de datos de Amazon S3 heredando los permisos del IAMAllowedPrincipals grupo (superacceso).	Herede permisos para crear tablas a partir del permiso CREATE TABLE concedido.
Ninguna ubicación de Amazon S3 asociada	Se requiere un permiso DATA LOCATION explícito para ejecutar las instrucciones CREATE TABLE e INSERT TABLE.	Se requiere un permiso DATA LOCATION explícito para ejecutar las instrucciones CREATE TABLE e INSERT TABLE.

IsRegisteredWithLakeFormation propiedad de la tabla

La propiedad `IsRegisteredWithLakeFormation` de una tabla indica si la ubicación de los datos de la tabla está registrada en Lake Formation para el solicitante. Si el modo de permiso de la ubicación está registrado como Lake Formation, la propiedad `IsRegisteredWithLakeFormation` es `true` para todos los usuarios que accedan a la ubicación de datos, ya que se considera que todos los usuarios han optado por participar en esa tabla. Si la ubicación está registrada en modo de acceso híbrido, el valor se establece en `true` solo para los usuarios que hayan optado por esa tabla.

Cómo funciona **`IsRegisteredWithLakeFormation`**

Modo de permisos	Users/Roles	**`IsRegisteredWithLakeFormation`**	Description (Descripción)
Lake Formation	Todos	True	Cuando se registre una ubicación en Lake Formation , la propiedad

Modo de permisos	Users/Roles	**IsRegisteredWithLakeFormation**	Description (Descripción)
			`IsRegisteredWithLakeFormation` se establecerá como verdadera para todos los usuarios. Esto significa que los permisos definidos en Lake Formation se aplican a la ubicación registrada. Lake Formation se encargará de la dispensación de credenciales.

Modo de permisos	Users/Roles	**`IsRegisteredWithLakeFormation`**	Description (Descripción)
Modo de acceso híbrido	Inscrito	True	En el caso de los usuarios que se hayan inscrito para usar Lake Formation para el acceso a los datos y la gobernanza de una tabla, la propiedad `IsRegisteredWithLakeFormation` se establecerá en `true` para dicha tabla. Están sujetos a las políticas de permisos definidas en Lake Formation para la ubicación registrada.

Modo de permisos	Users/Roles	**IsRegisteredWithLakeFormation**	Description (Descripción)
Modo de acceso híbrido	No inscrito	False	En el caso de los usuarios que no se hayan inscrito para utilizar los permisos de Lake Formation, la propiedad `IsRegisteredWithLakeFormation` se establece en `false`. No están sujetos a las políticas de permisos definidas en Lake Formation para la ubicación registrada. En su lugar, los usuarios seguirán las políticas de permisos de Amazon S3.

Convertir un AWS Glue recurso en un recurso híbrido

Siga estos pasos para registrar una ubicación de Amazon S3 en modo de acceso híbrido e incorporar nuevos usuarios de Lake Formation sin interrumpir el acceso a los datos de los usuarios actuales del Catálogo de datos.

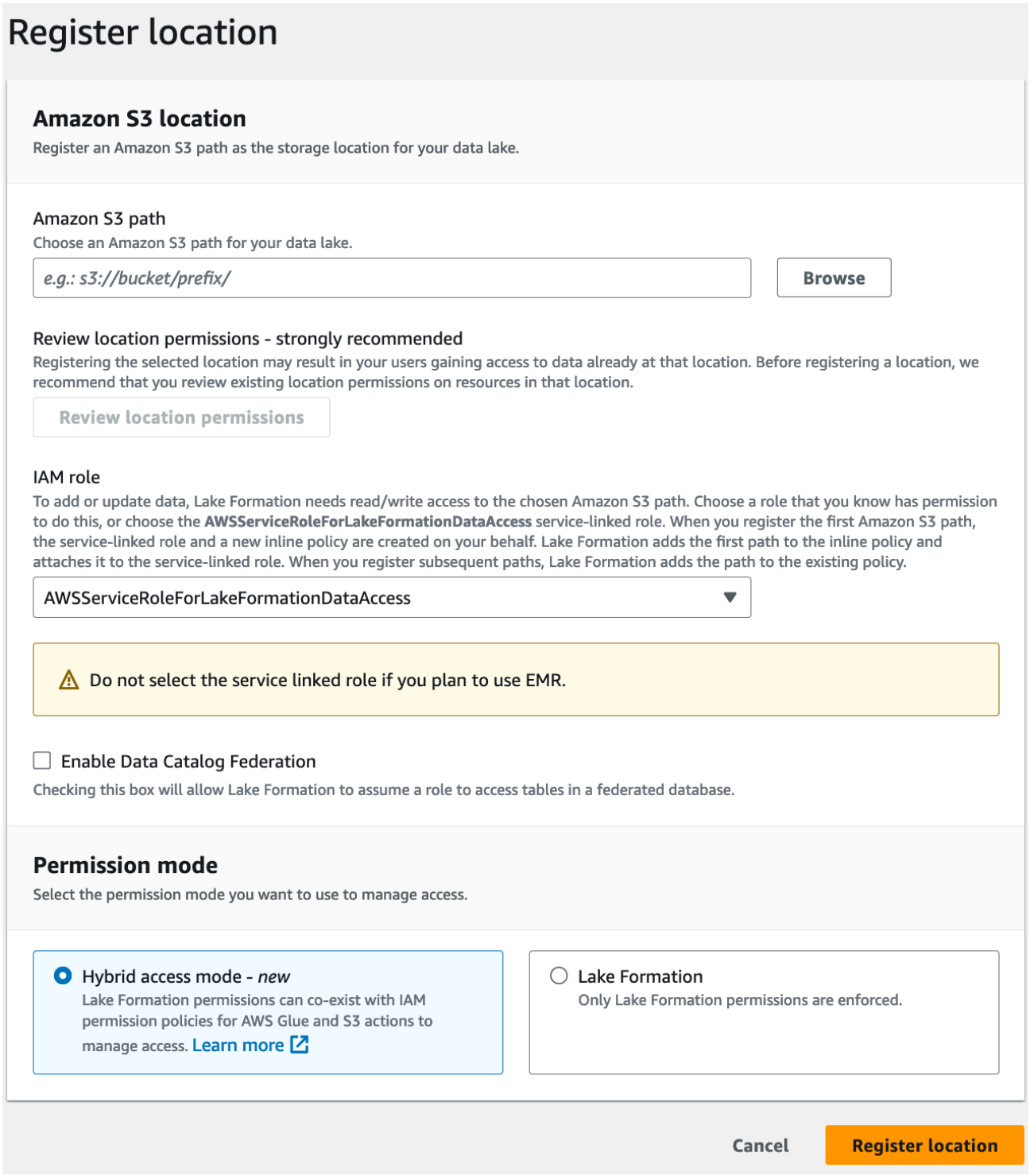
Descripción del escenario: la ubicación de los datos no está registrada en Lake Formation y el acceso de los usuarios a la base de datos y las tablas del Catálogo de datos se determina mediante las políticas de permisos de IAM para acciones de AWS Glue y Amazon S3.
De forma predeterminada, el grupo `IAMAllowedPrincipals` tiene permisos de `Super` en todas las tablas de la base de datos.

Para habilitar el modo de acceso híbrido para una ubicación de datos no registrada en Lake Formation

1. Registre una ubicación de Amazon S3 que habilite el modo de acceso híbrido.

   Console

   1. Inicie sesión en la consola de Lake Formation como administrador del lago de datos.
   2. En Administración del panel de navegación, seleccione las Ubicaciones de los lagos de datos.
   3. Seleccione Registrar ubicación.



4. En la ventana Registrar ubicación, elija la ruta de Amazon S3 que desee registrar en Lake Formation.
5. Para el rol de IAM, elija el rol vinculado al servicio `AWSServiceRoleForLakeFormationDataAccess` (valor predeterminado) o un rol de

IAM personalizado que cumpla los requisitos de Requisitos de los roles utilizados para registrar ubicaciones.

6. Elija el modo de acceso híbrido para aplicar políticas específicas de control de acceso de Lake Formation a las entidades principales y a las bases de datos y tablas del Catálogo de datos que apuntan a la ubicación registrada.

   Seleccione Lake Formation para permitir que Lake Formation autorice las solicitudes de acceso a la ubicación registrada.

7. Seleccione Registrar ubicación.

AWS CLI

El siguiente es un ejemplo para registrar una ubicación de datos con Lake Formation con HybridAccessEnabled:true/false. El valor predeterminado para el parámetro `HybridAccessEnabled` es false. Sustituya la ruta, el nombre del rol y el identificador de AWS cuenta de Amazon S3 por valores válidos.

```
aws lakeformation register-resource --cli-input-json file:file path
json:
    {
        "ResourceArn": "arn:aws:s3:::s3-path",
        "UseServiceLinkedRole": false,
        "RoleArn": "arn:aws:iam::<123456789012>:role/<role-name>",
        "HybridAccessEnabled": true
    }
```

2. Conceda permisos y seleccione entidades principales para utilizar los permisos de Lake Formation para los recursos en modo de acceso híbrido.

Antes de inscribir entidades principales y recursos en el modo de acceso híbrido, verifique que la concesión de los permisos `Super` o `All` al grupo `IAMAllowedPrincipals` existe en las bases de datos y tablas que tienen la ubicación registrada con Lake Formation en el modo de acceso híbrido.

> **Note**
>
> No puede conceder al grupo `IAMAllowedPrincipals` permiso sobre `All tables` dentro de una base de datos. Debe seleccionar cada tabla por separado en el menú desplegable y conceder los permisos. Además, al crear nuevas tablas en la base de

datos, puede optar por usar `Use only IAM access control for new tables in new databases` en la configuración del Catálogo de datos. Esta opción concede el permiso de `Super` al grupo `IAMAllowedPrincipals` automáticamente al crear nuevas tablas en la base de datos.

Console

1. En la consola de Lake Formation, en Catálogo de datos, elija Catálogos, Bases de datos o Tablas.
2. Seleccione un catálogo, base de datos o tabla de la lista y elija Conceder en el menú Acciones.
3. Elija los principales para conceder permisos en la base de datos, las tablas y las columnas mediante el método de recurso con nombre asignado o LF-Tags.

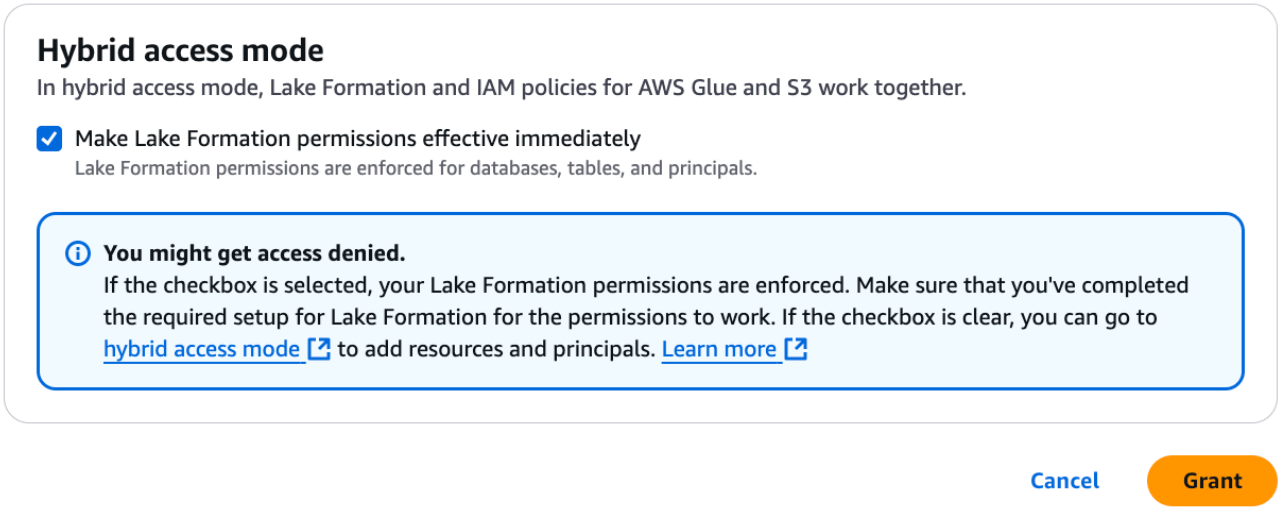
   Como alternativa, elija Permisos de datos, seleccione las entidades principales a las que conceder permisos de la lista y elija Conceder.

   Para obtener más información sobre la concesión de permisos de datos, consulte Concesión de permisos sobre los recursos del Catálogo de datos.

   > **Note**
   >
   > Si concede a una entidad principal el permiso para crear una tablas, también deberá concederle permisos de ubicación de datos (`DATA_LOCATION_ACCESS`). Este permiso no es necesario para actualizar las tablas.
   > Para obtener más información, consulte Conceder permisos de ubicación de datos.

4. Si utiliza el método de recurso con nombre para conceder permisos, la opción de incluir entidades principales y recursos está disponible en la sección inferior de la página de concesión de permisos de datos.

   Seleccione Hacer efectivos inmediatamente los permisos de Lake Formation para habilitar los permisos de Lake Formation para las entidades principales y los recursos.



5. Elija Conceder.

   Al optar por la entidad principal A en la tabla A que apunta a una ubicación de datos, permite que la entidad principal A tenga acceso a la ubicación de esta tabla utilizando los permisos de Lake Formation si la ubicación de datos está registrada en modo híbrido.

AWS CLI

A continuación se muestra un ejemplo para optar por una entidad principal y una tabla en modo de acceso híbrido. Sustituya el nombre del rol, el id de la cuenta de AWS , el nombre de la base de datos y el nombre de la tabla por valores aceptables.

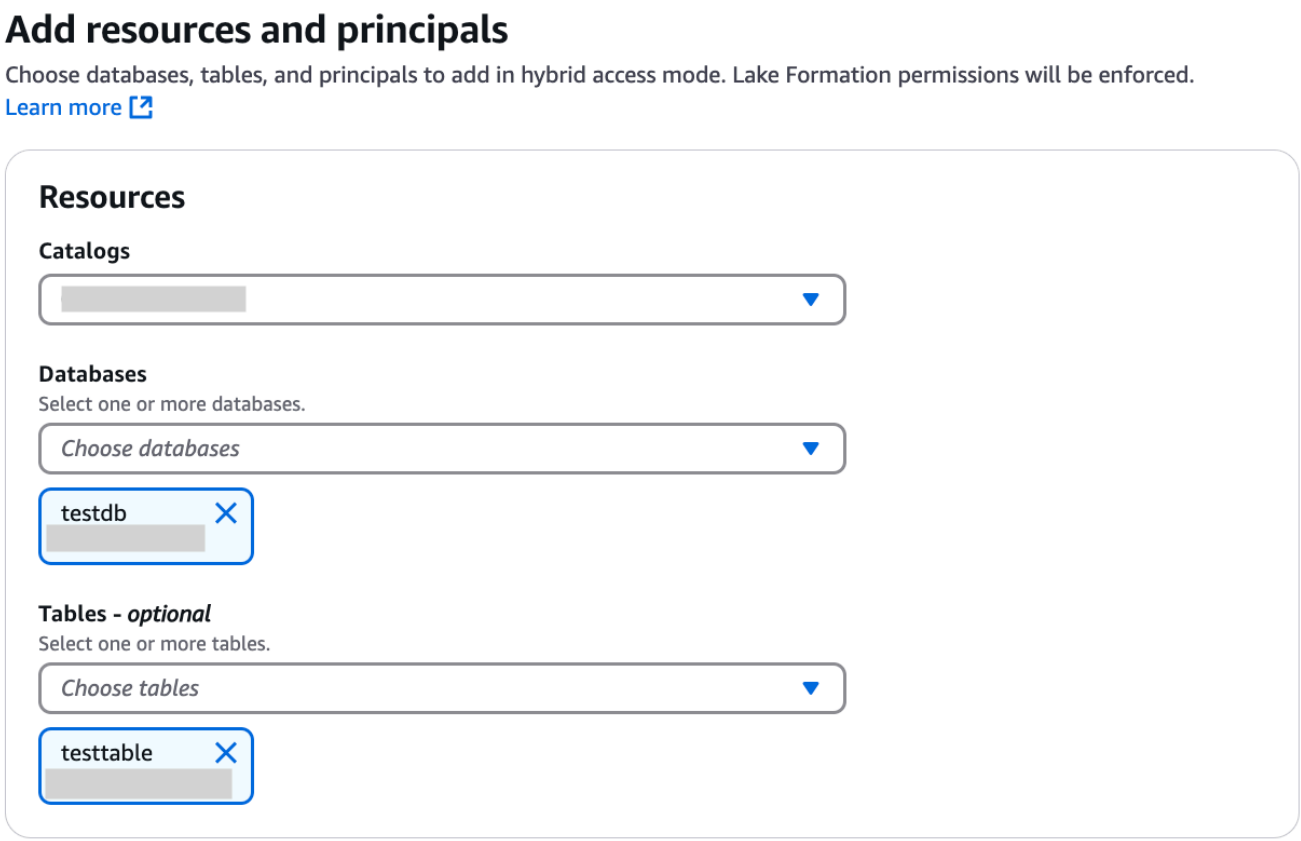
```
aws lakeformation create-lake-formation-opt-in --cli-input-json file://file path
json:
  {
        "Principal": {
            "DataLakePrincipalIdentifier":
 "arn:aws:iam::<123456789012>:role/<hybrid-access-role>"
        },
        "Resource": {
            "Table": {
                "CatalogId": "<123456789012>",
                "DatabaseName": "<hybrid_test>",
                "Name": "<hybrid_test_table>"
            }
        }
    }
```

a. (Optional) Si decides LF-Tags conceder permisos, puedes optar por que los directores usen los permisos de Lake Formation en un paso aparte. Para ello, elija el modo de acceso Híbrido en Permisos de la barra de navegación izquierda.

b. En la sección inferior de la página Modo de acceso híbrido, seleccione Añadir para añadir recursos y entidades principales al modo de acceso híbrido.

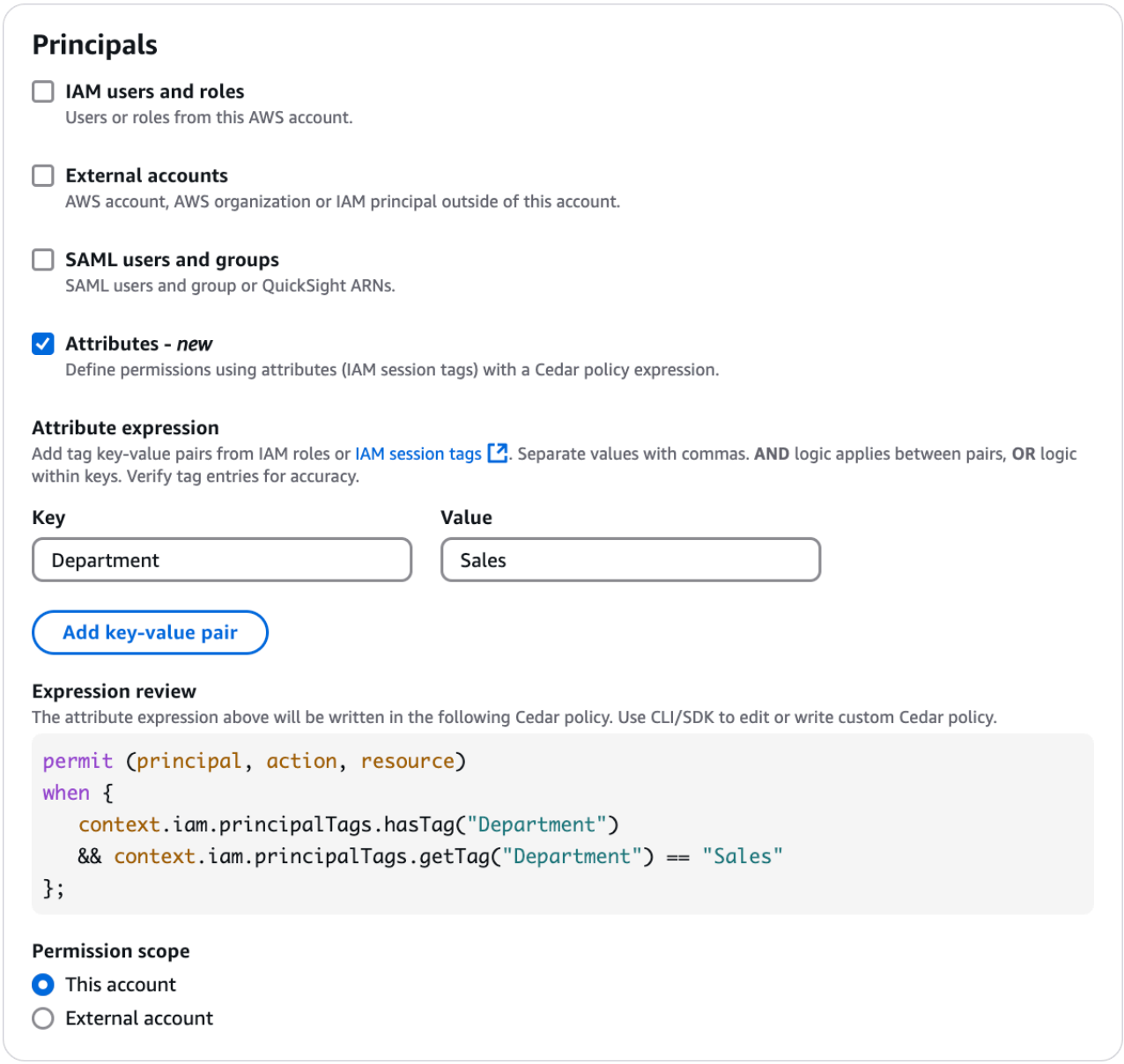
c. En la página Agregar recursos y entidades principales, elija los catálogos, bases de datos y tablas registrados en el modo de acceso híbrido.

   Puede elegir `All tables` bajo una base de datos a la que conceder acceso.



d. Elija la inscripción de entidades principales para utilizar los permisos de Lake Formation en el modo de acceso híbrido.

   - Entidades principales: puede elegir los usuarios y roles de IAM en la misma cuenta o en otra cuenta. También puede elegir usuarios y grupos de SAML.
   - Atributos: seleccione los atributos para conceder permisos en función de los atributos.



- Especifique el par clave-valor para crear una concesión basada en atributos. Revise la expresión de política de Cedar resultante en la consola. Para obtener más información acerca de Cedar, consulte ¿Qué es Cedar? | Referencia GuideLink del lenguaje de la política de Cedar.
- Elija Añadir.

  Se concede acceso a todos los IAM roles/users con atributos coincidentes.

e. Elija Añadir.

Convertir un recurso de Lake Formation en un recurso híbrido

En los casos en los que esté utilizando permisos de Lake Formation para sus bases de datos y tablas del Catálogo de datos, puede editar las propiedades de registro de la ubicación para activar el modo de acceso híbrido. Esto le permite proporcionar a los nuevos directores acceso a los mismos recursos mediante las políticas de permisos y AWS Glue acciones de IAM para Amazon S3 sin interrumpir los permisos existentes de Lake Formation.

Descripción del escenario: en los siguientes pasos se asume que tiene una ubicación de datos registrada en Lake Formation y que ha configurado permisos para entidades principales en bases de datos, tablas o columnas que apuntan a esa ubicación. Si la ubicación se registró con un rol vinculado a un servicio, no podrá actualizar los parámetros de ubicación ni habilitar el modo de acceso híbrido. De forma predeterminada, el grupo `IAMAllowedPrincipals` tiene permisos Super sobre la base de datos y todas sus tablas.

> **Important**
>
> No actualice un registro de ubicación al modo híbrido sin inscribirse en las entidades principales que acceden a los datos de esta ubicación.

Activación del modo de acceso híbrido para una ubicación de datos registrada en Lake Formation

1.
> **Warning**
>
> No recomendamos convertir una ubicación de datos administrada de Lake Formation en un modo de acceso híbrido para evitar interrumpir las políticas de permisos de otros usuarios o cargas de trabajo existentes.

   Opte por las entidades principales existentes que tengan permisos de Lake Formation.

   1. Recopile y revise los permisos que ha concedido a las entidades principales sobre los catálogos, bases de datos y tablas. Para obtener más información, consulte Consulta de los permisos de bases de datos y tablas en Lake Formation.
   2. Elija el modo de acceso Híbrido en Permisos de la barra de navegación izquierda y seleccione Añadir.

3. En la página Agregar entidades principales y recursos, elija los catálogos, bases de datos y tablas de la ubicación de datos de Amazon S3 que desee utilizar en el modo de acceso híbrido. Elija las entidades principales que ya tienen permisos de Lake Formation.
4. Elija Añadir para optar a que las entidades principales utilicen los permisos de Lake Formation en el modo de acceso híbrido.

2. Actualice el bucket/prefix registro de Amazon S3 seleccionando la opción de modo de acceso híbrido.

   Console

   1. Inicie sesión en la consola de Lake Formation como administrador del lago de datos.
   2. En el panel de navegación, vaya a Registrar e ingerir y seleccione las ubicaciones de los lagos de datos.
   3. Seleccione una ubicación y, en el menú Acciones, seleccione Editar.
   4. Elija Modo de acceso híbrido.
   5. Seleccione Save.
   6. En el Catálogo de datos, seleccione la base de datos o tabla y conceda permisos `Super` o `All` al grupo virtual denominado `IAMAllowedPrincipals`.
   7. Compruebe que el acceso de sus usuarios actuales de Lake Formation no se interrumpa al actualizar las propiedades de registro de la ubicación. Inicie sesión en la consola de Athena como entidad principal de Lake Formation y ejecute una consulta de ejemplo en una tabla que apunte a la ubicación actualizada.

      Del mismo modo, compruebe el acceso de AWS Glue los usuarios que utilizan las políticas de permisos de IAM para acceder a la base de datos y a las tablas.

   AWS CLI

   El siguiente es un ejemplo para registrar una ubicación de datos con Lake Formation con HybridAccessEnabled:true/false. El valor predeterminado para el parámetro `HybridAccessEnabled` es false. Sustituya la ruta, el nombre del rol y el identificador de AWS cuenta de Amazon S3 por valores válidos.

   ```
   aws lakeformation update-resource --cli-input-json file://file path
   json:
   {
   ```

```
    "ResourceArn": "arn:aws:s3:::<s3-path>",
    "RoleArn": "arn:aws:iam::<123456789012>:role/<test>",
    "HybridAccessEnabled": true
}
```

Compartir un AWS Glue recurso mediante el modo de acceso híbrido

Comparta datos con otra persona Cuenta de AWS o con un director en otra persona Cuenta de AWS haciendo cumplir los permisos de Lake Formation sin interrumpir el acceso basado en IAM de los usuarios existentes del Catálogo de Datos.

Descripción del escenario: la cuenta del productor tiene una base de datos del catálogo de datos cuyo acceso está controlado mediante las principales políticas y AWS Glue acciones de IAM para Amazon S3. La ubicación de los datos de la base de datos no está registrada en Lake Formation. De forma predeterminada, el grupo `IAMAllowedPrincipals` tiene permisos `Super` sobre la base de datos y todas sus tablas.

Conceder permisos entre cuentas de Lake Formation en modo de acceso híbrido

1. Configuración de una cuenta de productor

    1. Inicie sesión en la consola de Lake Formation con un rol que tenga permiso `lakeformation:PutDataLakeSettings` de IAM.
    2. Vaya a la configuración del Catálogo de datos y seleccione `Version 4` para la configuración de la versión entre cuentas.

        Si utiliza la versión 1 o 2, consulte las instrucciones de Actualización de los ajustes de la versión entre cuentas para compartir datos para actualizar a la versión 3.

        No es necesario hacer cambios en la política de permisos para actualizar de la versión 3 a la 4.
    3. Registre la ubicación en Amazon S3 de la base de datos o tabla que planea compartir en el modo de acceso híbrido.
    4. Compruebe que existe permiso `Super` para el grupo `IAMAllowedPrincipals` en las bases de datos y tablas en las que registró la ubicación de datos en modo de acceso híbrido en el paso anterior.

5. Otorgue permisos de Lake Formation a AWS organizaciones, unidades organizativas (OU) o directamente con un director de IAM en otra cuenta.
6. Si concede los permisos directamente a una entidad principal de IAM, opte por que la entidad principal de la cuenta de consumidor aplique los permisos de Lake Formation en el modo de acceso híbrido activando la opción Hacer efectivos inmediatamente los permisos de Lake Formation.

   Si vas a conceder permisos multicuenta a otra AWS cuenta, al activar la cuenta, los permisos de Lake Formation solo se aplican a los administradores de esa cuenta. El administrador del lago de datos de la cuenta de destinatario tiene que aplicar en cascada los permisos y optar por las entidades principales de la cuenta para hacer cumplir los permisos de Lake Formation para los recursos compartidos que están en modo de acceso híbrido.

   Si eliges la LF-Tags opción Recursos coincidentes para conceder permisos entre cuentas, primero debes completar el paso de concesión de permisos. Puede optar por el modo de acceso híbrido para entidades principales y recursos como un paso separado, seleccionando el modo de acceso híbrido en Permisos en la barra de navegación izquierda de la consola de Lake Formation. Luego, seleccione Agregar para agregar los recursos y las entidades principales a los que desea aplicar los permisos de Lake Formation.

2. Configuración de una cuenta de consumidor

   1. Inicie sesión en la consola de Lake Formation https://console.aws.amazon.com/lakeformation/como administrador de un lago de datos.
   2. Vaya a https://console.aws.amazon.com/ram/homey acepte la invitación para compartir recursos. La pestaña Compartido conmigo de la AWS RAM consola muestra la base de datos y las tablas que se comparten con su cuenta.
   3. Cree un enlace de recursos a la and/or tabla de base de datos compartida en Lake Formation.
   4. Conceda permiso `Describe` en el enlace de recursos y permiso `Grant on target` (para el recurso compartido original) a las entidades principales de IAM de su cuenta (de consumidor).
   5. Conceda a las entidades principales de su cuenta permisos de Lake Formation en la base de datos o tabla compartida. Indique a las entidades principales y a los recursos que apliquen los permisos de Lake Formation en el modo de acceso híbrido activando la opción Hacer efectivos inmediatamente los permisos de Lake Formation.
   6. Pruebe los permisos de Lake Formation de la entidad principal ejecutando consultas Athena de ejemplo. Pruebe el acceso actual de sus AWS Glue usuarios con las políticas principales de IAM para Amazon S3 y AWS Glue sus acciones.

(Opcional) Elimine la política de bucket de Amazon S3 para el acceso a los datos y las políticas principales de IAM para AWS Glue y el acceso a los datos de Amazon S3 para las entidades principales que configuró para usar los permisos de Lake Formation.

Compartir un recurso de Lake Formation mediante el modo de acceso híbrido

Autorice a los nuevos usuarios del Catálogo de datos de una cuenta externa a acceder a las bases de datos y tablas del Catálogo de datos mediante políticas basadas en IAM sin interrumpir los permisos existentes de uso compartido entre cuentas de Lake Formation.

Descripción del escenario. La cuenta del productor tiene una base de datos y tablas administradas por Lake Formation que se comparten con una cuenta externa (de consumidor) a nivel de cuenta o a nivel de entidad principal de IAM. La ubicación de los datos de la base de datos está registrada en Lake Formation. El grupo `IAMAllowedPrincipals` no tiene permisos `Super` en la base de datos ni en sus tablas.

Conceder acceso entre cuentas a nuevos usuarios del Catálogo de datos mediante políticas basadas en IAM sin interrumpir el permiso existente de Lake Formation.

1. Configuración de una cuenta de productor

    1. Inicie sesión en la consola de Lake Formation con un rol que `lakeformation:PutDataLakeSettings`.
    2. Vaya a la configuración del Catálogo de datos y seleccione `Version 4` para la configuración de la versión entre cuentas.

        Si utiliza la versión 1 o 2, consulte las instrucciones de Actualización de los ajustes de la versión entre cuentas para compartir datos para actualizar a la versión 3.

        No es necesario hacer cambios en la política de permisos para actualizar de la versión 3 a la 4.
    3. Recopile y revise los permisos que ha concedido a las entidades principales sobre las bases de datos y las tablas. Para obtener más información, consulte Consulta de los permisos de bases de datos y tablas en Lake Formation.
    4. Vuelva a conceder los permisos entre cuentas existentes de Lake Formation optando por entidades principales y recursos.

> **Note**
>
> Antes de actualizar el registro de una ubicación de datos al modo de acceso híbrido para conceder permisos entre cuentas, debe volver a conceder al menos un recurso compartido de datos entre cuentas por cuenta. Este paso es necesario para actualizar los permisos AWS RAM administrados adjuntos al AWS RAM recurso compartido.
> En julio de 2023, Lake Formation actualizó los permisos AWS RAM gestionados que se utilizan para compartir bases de datos y tablas:
>
> - `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueAllTablesReadWriteForDatabase` (política de uso compartido a nivel de base de datos)
> - `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueTableReadWrite` (política de uso compartido a nivel de tabla)
>
> Las concesiones de permisos multicuenta realizadas antes de julio de 2023 no tienen estos AWS RAM permisos actualizados.
> Si ha concedido permisos para varias cuentas directamente a las entidades principales, tendrá que volver a concederlos individualmente. Si omite este paso, las entidades principales que accedan al recurso compartido podrían obtener un error de combinación ilegal.

5. Vaya a https://console.aws.amazon.com/ram/home.
6. La pestaña Compartidos por mí de la AWS RAM consola muestra los nombres de bases de datos y tablas que has compartido con una cuenta o entidad principal externa.

   Asegúrese de que los permisos adjuntos al recurso compartido tengan el ARN correcto.
7. Comprueba que los recursos del AWS RAM recurso compartido estén en `Associated` estado. Si el estado es `Associating`, espere a que pasen al estado `Associated`. Si el estado pasa a ser `Failed`, deténgase y póngase en contacto con el equipo de servicio de Lake Formation.
8. Elija el modo de acceso Híbrido en Permisos de la barra de navegación izquierda y seleccione Añadir.
9. La página Agregar entidades principales y recursos muestra las bases de datos, and/or las tablas y las entidades principales a las que se puede acceder. Para efectuar las actualizaciones necesarias, añada o quite entidades principales y recursos.

10. Elija las entidades principales con permisos de Lake Formation para la base de datos y las tablas que desea cambiar al modo de acceso híbrido. Elija las bases de datos y tablas.
11. Elija Añadir para que las entidades principales puedan utilizar los permisos de Lake Formation en el modo de acceso híbrido.
12. Conceda permiso `Super` al grupo virtual `IAMAllowedPrincipals` de su base de datos y de las tablas seleccionadas.
13. Edite el registro de Lake Formation de la ubicación de Amazon S3 en modo de acceso híbrido.
14. Otorgue permisos a los AWS Glue usuarios de la cuenta externa (de consumidor) mediante las políticas de permisos de IAM para las AWS Glue acciones de Amazon S3.

2. Configuración de una cuenta de consumidor

    1. Inicie sesión en la consola de Lake Formation https://console.aws.amazon.com/lakeformation/como administrador de un lago de datos.
    2. Vaya a la invitación para compartir recursos https://console.aws.amazon.com/ram/homey acéptela. La pestaña Recursos compartidos conmigo de la AWS RAM página muestra los nombres de las bases de datos y las tablas que se comparten con su cuenta.

       Para AWS RAM compartir, asegúrate de que el permiso adjunto tenga el ARN correcto de la invitación compartida AWS RAM . Comprueba si los recursos del recurso AWS RAM compartido están en `Associated` estado. Si el estado es `Associating`, espere a que pasen al estado `Associated`. Si el estado pasa a ser `Failed`, deténgase y póngase en contacto con el equipo de servicio de Lake Formation.
    3. Cree un enlace de recursos a la and/or tabla de base de datos compartida en Lake Formation.
    4. Conceda permiso `Describe` en el enlace de recursos y permiso `Grant on target` (para el recurso compartido original) a las entidades principales de IAM de su cuenta (de consumidor).
    5. A continuación, configure los permisos de Lake Formation para las entidades principales de su cuenta en la base de datos o tabla compartida.

       En la barra de navegación de la izquierda, en Permisos, elija el modo de acceso Híbrido.
    6. Seleccione Añadir en la sección inferior de la página del modo de acceso híbrido para optar por las entidades principales y la base de datos o tabla compartida de la cuenta de productor.
    7. Conceda permisos a los AWS Glue usuarios de su cuenta mediante las políticas de permisos de IAM para las AWS Glue acciones de Amazon S3.

8. Pruebe los permisos y AWS Glue permisos de Lake Formation de los usuarios ejecutando consultas de ejemplo independientes en la tabla con Athena

   (Opcional) Limpie las políticas de permisos de IAM para Amazon S3 para las entidades principales que se encuentran en el modo de acceso híbrido.

Eliminación de entidades principales y recursos del modo de acceso híbrido

Siga estos pasos para eliminar bases de datos, tablas y entidades principales del modo de acceso híbrido.

Console

1. Inicie sesión en la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En Permisos, seleccione el modo de acceso híbrido.
3. En la página Modo de acceso híbrido, marque la casilla situada junto al nombre de la base de datos o tabla y elija `Remove`.
4. Un mensaje de advertencia le solicitará que confirme la acción. Elija Eliminar .

   Lake Formation ya no exige permisos para esos recursos, y el acceso a este recurso se controlará mediante IAM y AWS Glue permisos. Esto puede provocar que el usuario deje de tener acceso a este recurso si no tiene los permisos de IAM adecuados.

AWS CLI

En el siguiente ejemplo, se muestra cómo eliminar recursos del modo de acceso híbrido.

```
aws lakeformation delete-lake-formation-opt-in --cli-input-json file://file path

json:
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::<123456789012>:role/role name"
    },
    "Resource": {
        "Table": {
            "CatalogId": "<123456789012>",
            "DatabaseName": "<database name>",
```

```
            "Name": "<table name>"
          }
    }
}
```

Eliminación de entidades principales y recursos del modo de acceso híbrido

Siga estos pasos para eliminar bases de datos, tablas y entidades principales del modo de acceso híbrido.

Console

1. Inicie sesión en la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En Permisos, seleccione el modo de acceso híbrido.
3. La página Modo de acceso híbrido muestra los recursos y entidades principales presentes en el modo de acceso híbrido.

AWS CLI

En el ejemplo siguiente, se muestra cómo enumerar todas las entidades principales y recursos que están en el modo de acceso híbrido.

```
aws lakeformation list-lake-formation-opt-ins
```

El siguiente ejemplo muestra cómo enumerar un par específico de entidad principal-recurso.

```
aws lakeformation list-lake-formation-opt-ins --cli-input-json file://file path

json:
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::<account-id>:role/<role name>"
    },
    "Resource": {
```

```
        "Table": {
            "CatalogId": "<account-id>",
            "DatabaseName": "<database name>",
            "Name": "<table name>"
          }
    }
}
```

Recursos adicionales

En la siguiente entrada del blog, le guiaremos a través de las instrucciones para incorporar los permisos de Lake Formation en modo de acceso híbrido para los usuarios seleccionados, mientras que la base de datos ya es accesible para otros usuarios a través de los permisos de IAM y Amazon S3. Revisaremos las instrucciones para configurar el modo de acceso híbrido dentro de una AWS cuenta y entre dos cuentas.

- Presentamos el modo de acceso híbrido AWS Glue Data Catalog para proteger el acceso mediante Lake Formation y las políticas de IAM y Amazon S3.

Creación de objetos en el AWS Glue Data Catalog

AWS Lake Formation utiliza el AWS Glue Data Catalog (catálogo de datos) para almacenar metadatos sobre lagos de datos, fuentes de datos, transformaciones y destinos. Los metadatos son datos sobre los datos subyacentes del conjunto de datos. Cada AWS cuenta tiene un catálogo de datos por AWS región.

Los metadatos del Catálogo de datos se organizan en una jerarquía de datos de tres niveles que incluye catálogos, bases de datos y tablas. Organiza los datos de diversos orígenes en contenedores lógicos denominados catálogos. Cada catálogo representa datos de orígenes como los almacenes de datos de Amazon Redshift, bases de datos Amazon DynamoDB y orígenes de datos de terceros, como Snowflake, MySQL y más de 30 orígenes de datos externos, que se integran mediante conectores federados. También puede crear catálogos nuevos en el Catálogo de datos para almacenar datos en buckets de tablas de S3 o en Redshift Managed Storage (RMS).

Las tablas almacenan información sobre los datos subyacentes, incluida la información sobre esquemas, particiones y ubicaciones de datos. Las bases de datos son colecciones de tablas. El Catálogo de datos también contiene enlaces a recursos, que son enlaces a catálogos, bases de

datos y tablas compartidas en cuentas externas, y se utilizan para el acceso entre cuentas a los datos del lago de datos.

El Catálogo de datos es un objeto de catálogo anidado que contiene catálogos, bases de datos y tablas. Se hace referencia a él mediante el Cuenta de AWS ID y es el catálogo predeterminado de una cuenta y un Región de AWS. El Catálogo de datos utiliza una jerarquía de tres niveles (catalog.database.table) para organizar las tablas.

- Catálogo: el nivel más alto de la jerarquía de metadatos de tres niveles del Catálogo de datos. Puede agregar varios catálogos a un Catálogo de datos mediante la federación.
- Base de datos: el segundo nivel de la jerarquía de metadatos que consta de tablas y vistas. Una base de datos también se denomina esquema en muchos sistemas de datos, como Amazon Redshift y Trino.
- Tabla y vista: el tercer nivel de la jerarquía de datos de tres niveles del Catálogo de datos.

Todas las tablas Iceberg de Amazon S3 se almacenan en el catálogo de datos predeterminado con un ID de catálogo = Cuenta de AWS ID. Puede crear catálogos federados AWS Glue Data Catalog que almacenen definiciones de tablas en Amazon Redshift, Amazon S3 Table Storage u otras fuentes de datos de terceros mediante la federación.

Temas

- Creación de un catálogo
- Creación de una base de datos
- Creación de tablas
- Construcción AWS Glue Data Catalog vistas

Creación de un catálogo

Los catálogos representan el nivel más alto o primer nivel de la jerarquía de metadatos de tres niveles del AWS Glue Data Catalog. Puede utilizar varios métodos para incorporar datos al Catálogo de datos y crear catálogos de varios niveles.

Para obtener más información sobre la creación de catálogos a partir de orígenes de datos externos, consulte Llevar sus datos al AWS Glue Data Catalog.

Para crear una base de datos mediante la consola de Lake Formation, debe iniciar sesión como administrador de lago de datos o creador de catálogo. Un creador de catálogo es una entidad

principal a la que se le ha otorgado el permiso CREATE_CATALOG de Lake Formation. Puede ver una lista de los creadores de catálogos en la página Roles y tareas administrativas de la consola de Lake Formation. Para ver esta lista, debe tener el permiso `lakeformation:ListPermissions` de IAM e iniciar sesión como administrador de lago de datos o como creador de catálogo con la opción de conceder el permiso CREATE_CATALOG.

Creación de una base de datos

Las tablas de metadatos del Catálogo de datos se almacenan en una base de datos. Puede crear tantas bases de datos como necesite, y conceder diferentes permisos de Lake Formation en cada base de datos.

Las bases de datos pueden tener una propiedad de ubicación opcional. Esta ubicación suele estar dentro de una ubicación de Amazon Simple Storage Service (Amazon S3) registrada en Lake Formation. Cuando se especifica una ubicación, las entidades principales no necesitan permisos de ubicación de datos para crear tablas del Catálogo de datos que apunten a ubicaciones dentro de la ubicación de la base de datos. Para obtener más información, consulte Underlying data access control.

Para crear una base de datos mediante la consola de Lake Formation, debe iniciar sesión como administrador del lago de datos o creador de la base de datos. El creador de una base de datos es una entidad principal a la que se le ha otorgado el permiso de CREATE_DATABASE de Lake Formation. Puede ver una lista de los creadores de bases de datos en la página Roles y tareas administrativas de la consola de Lake Formation. Para ver esta lista, debe tener el permiso de `lakeformation:ListPermissions` IAM e iniciar sesión como administrador de un lago de datos o como creador de bases de datos con la opción de conceder el permiso CREATE_DATABASE.

Para crear una base de datos

1. Abra la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/e inicie sesión como administrador de un lago de datos o creador de bases de datos.
2. En el panel de navegación, bajo Catálogo de datos, elija Bases de datos.
3. Elija Creación de base de datos.
4. En el cuadro de diálogo Crear base de datos, introduzca un nombre de base de datos, una ubicación opcional y una descripción opcional.
5. Seleccione de forma opcional Utilizar solo el control de acceso IAM para las nuevas tablas de esta base de datos.

Para obtener más información acerca de esta opción, consulte the section called “Cambiar la configuración predeterminada de su lago de datos”.

6. Elija Creación de base de datos.

Creación de tablas

AWS Lake Formation las tablas de metadatos contienen información sobre los datos del lago de datos, incluida la información del esquema, la información de particiones y la ubicación de los datos. Estas tablas se almacenan en el Catálogo de datos de AWS Glue. Se utilizan para acceder a los datos subyacentes del lago de datos y administrarlos con los permisos de Lake Formation. Las tablas se almacenan dentro de bases de datos en el Catálogo de datos.

Hay varias formas de crear tablas del Catálogo de datos:

- Ejecutar un rastreador en AWS Glue. Consulte Definición de rastreadores en la Guía para desarrolladores de AWS Glue .
- Crear y ejecutar un flujo de trabajo. Consulte the section called “Importación de datos mediante flujos de trabajo”.
- Cree una tabla manualmente utilizando la consola de Lake Formation, la API de AWS Glue o la AWS Command Line Interface (AWS CLI).
- Cree una tabla con Amazon Athena.
- Crea un enlace de recurso a una tabla en una cuenta externa. Consulte the section called “Creación de enlaces de recursos”.

Creación de tablas de Apache Iceberg

AWS Lake Formation admite la creación de tablas Apache Iceberg que utilizan el formato de datos Apache Parquet AWS Glue Data Catalog con datos que residen en Amazon S3. Una tabla en el Catálogo de datos es la definición de metadatos que representa los datos en un almacén de datos. De forma predeterminada, Lake Formation crea tablas Iceberg v2. Para ver la diferencia entre las tablas v1 y v2, consulte Cambios de versión de formato en la documentación de Apache Iceberg.

Apache Iceberg es un formato de tabla abierto para conjuntos de datos analíticos muy grandes. Iceberg permite modificar fácilmente su esquema, o evolución del esquema, de manera que los usuarios pueden añadir, renombrar o eliminar columnas de una tabla de datos sin alterar los datos

subyacentes. Iceberg también ofrece compatibilidad con el control de versiones de datos, que permite a los usuarios hacer un seguimiento de los cambios en los datos a lo largo del tiempo. Esto habilita la característica de viaje en el tiempo, con la que los usuarios pueden acceder a versiones históricas de los datos y consultarlas, así como analizar los cambios en los datos entre actualizaciones y eliminaciones.

Puede usar la consola de Lake Formation o la `CreateTable` operación de la AWS Glue API para crear una tabla Iceberg en el catálogo de datos. Para obtener más información, consulte CreateTable action (Python: create_table).

Cuando cree una tabla de Iceberg en el Catálogo de datos, deberá especificar el formato de la tabla y la ruta del archivo de metadatos en Amazon S3 para poder hacer lecturas y escrituras.

Puede usar Lake Formation para proteger su mesa de iceberg mediante permisos de control de acceso detallados al registrar la ubicación de datos de Amazon S3. AWS Lake Formation En el caso de los datos fuente de Amazon S3 y los metadatos que no estén registrados en Lake Formation, el acceso viene determinado por las políticas de permisos y AWS Glue acciones de IAM para Amazon S3. Para obtener más información, consulte Administrar los permisos de Lake Formation.

> **Note**
>
> El Catálogo de datos no admite la creación de particiones ni la adición de propiedades de tablas de iceberg.

Temas

- Requisitos previos
- Creación de tablas de Iceberg

Requisitos previos

Para crear tablas de Iceberg en el Catálogo de datos y configurar los permisos de acceso a los datos de Lake Formation, debe cumplir los siguientes requisitos:

1. Se requieren permisos para crear tablas de Iceberg sin datos registrados en Lake Formation.

   Además de los permisos necesarios para crear una tabla en el Catálogo de datos, el creador de la tabla requiere los siguientes permisos:

- `s3:PutObject` en el recurso arn:aws:s3:::{bucketName}
- `s3:GetObject` en el recurso arn:aws:s3:::{bucketName}
- `s3:DeleteObject` en el recurso arn:aws:s3:::{bucketName}

2. Se requieren permisos para crear tablas de Iceberg con datos registrados en Lake Formation:

   Para utilizar Lake Formation para administrar y asegurar los datos de su lago de datos, registre su ubicación de Amazon S3 que tiene los datos de las tablas con Lake Formation. Esto es para que Lake Formation pueda vender credenciales a servicios AWS analíticos como Athena, Redshift Spectrum y Amazon EMR para acceder a los datos. Para obtener más información sobre el registro de una ubicación de Amazon S3, consulte Añadir una ubicación de Amazon S3 a su lago de datos.

   Una entidad principal que lee y escribe los datos subyacentes que están registrados en Lake Formation requiere los siguientes permisos:

   - `lakeformation:GetDataAccess`
   - `DATA_LOCATION_ACCESS`

     Una entidad principal que tiene permisos de localización de datos en una localización también tiene permisos de localización en todas las ubicaciones secundarias.

     Para obtener más información sobre permisos de ubicación de datos, consulte Control de acceso a los datos subyacentes.

Para habilitar la compactación, el servicio debe asumir un rol de IAM que tenga permisos para actualizar las tablas del Catálogo de datos. Para obtener más información, consulte Requisitos previos para la optimización de tablas.

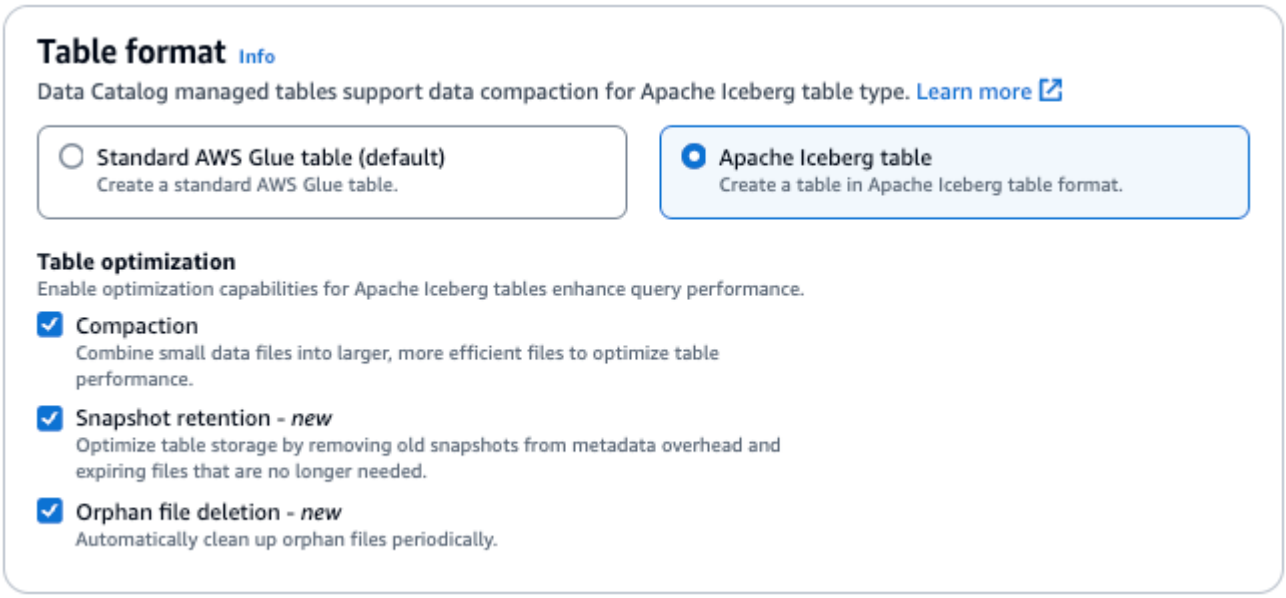
Creación de tablas de Iceberg

Puede crear tablas Iceberg v1 y v2 con la consola Lake Formation o AWS Command Line Interface tal como se documenta en esta página. También puede crear tablas Iceberg utilizando la AWS Glue consola o. Rastreador de AWS Glue Para más información, consulte Catálogo de datos y rastreadores en la Guía para desarrolladores de AWS Glue .

Creación de una tabla de Iceberg

Console

1. Inicie sesión en y abra la Consola de administración de AWS consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En Catálogo de datos, seleccione Tablas y utilice el botón Crear tabla para especificar los siguientes atributos:
   - Nombre de tabla. Escriba un nombre para la tabla. Si utiliza Athena para acceder a las tablas, utilice los consejos para nombres recogidos en la Guía del usuario de Amazon Athena.
   - Base de datos. Elija una base de datos existente o cree una nueva.
   - Descripción. Descripción de la tabla. Puede escribir una descripción para ayudarle a entender el contenido de la tabla.
   - Formato de tabla. Para el formato de la tabla, elija Apache Iceberg.



   - Optimización de tablas
     - Compactación: los archivos de datos se combinan y se reescriben para eliminar los datos obsoletos y consolidar los datos fragmentados en archivos más grandes y eficientes.
     - Retención de instantáneas: las instantáneas son versiones con fecha y hora de una tabla de Iceberg. Las configuraciones de retención de instantáneas permiten a los clientes determinar cuánto tiempo se deben retener las instantáneas y cuántas instantáneas retener. La configuración de un optimizador de retención de instantáneas puede ayudar a administrar la sobrecarga de almacenamiento mediante la eliminación de las instantáneas antiguas e innecesarias y sus correspondientes archivos subyacentes.

- Eliminación de archivos huérfanos: los archivos huérfanos son archivos a los que los metadatos de la tabla de Iceberg ya no hacen referencia. Con el tiempo, estos archivos se pueden acumular, sobre todo después de operaciones como la eliminación de tablas o los errores en los trabajos de ETL. Habilitar la eliminación de archivos huérfanos AWS Glue permite identificar y eliminar periódicamente estos archivos innecesarios, liberando espacio de almacenamiento.

  Para obtener más información, consulte Optimización de las tablas de Iceberg.

- Rol de IAM. Para ejecutar la compactación, el servicio asume un rol de IAM en su nombre. Puede elegir un rol de IAM mediante el menú desplegable. Asegúrese de que el rol tenga los permisos necesarios para habilitar la compactación.

  Para obtener más información sobre los permisos necesarios, consulte Requisitos previos para la optimización de tablas.

- Ubicación. Especifique la ruta a la carpeta de Amazon S3 que almacena la tabla de metadatos. Iceberg necesita un archivo de metadatos y una ubicación en el Catálogo de datos para poder hacer lecturas y escrituras.

- Esquema. Seleccione Agregar columnas para añadir columnas y tipos de datos de las columnas. Tiene la opción de crear una tabla vacía y actualizar el esquema más adelante. El Catálogo de datos admite los tipos de datos de Hive. Para obtener más información, consulte Tipos de datos de Hive.

  Con Iceberg podrá desarrollar el esquema y la partición después de crear la tabla. Puede utilizar consultas de Athena para actualizar el esquema de la tabla y consultas de Spark para actualizar las particiones.

AWS CLI

```
aws glue create-table \
    --database-name iceberg-db \
    --region us-west-2 \
    --open-table-format-input '{
      "IcebergInput": {
           "MetadataOperation": "CREATE",
           "Version": "2"
         }
      }' \
    --table-input '{"Name":"test-iceberg-input-demo",
```

```
            "TableType": "EXTERNAL_TABLE",
            "StorageDescriptor":{
               "Columns":[
                   {"Name":"col1", "Type":"int"},
                   {"Name":"col2", "Type":"int"},
                   {"Name":"col3", "Type":"string"}
                ],
               "Location":"s3://DOC_EXAMPLE_BUCKET_ICEBERG/"
            }
        }'
```

Optimización de las tablas de Iceberg

Lake Formation admite múltiples opciones de optimización de tablas para mejorar la administración y el rendimiento de las tablas Apache Iceberg utilizadas por los motores AWS analíticos y los trabajos de ETL. Estos optimizadores ofrecen un uso eficiente del almacenamiento, un rendimiento mejorado de las consultas y la administración efectiva de los datos. Existen tres tipos de optimizadores de tablas disponibles en Lake Formation:

- Compactación: la compactación de datos compacta archivos de datos pequeños para reducir el uso de almacenamiento y mejorar el rendimiento de lectura. Los archivos de datos se combinan y se reescriben para eliminar los datos obsoletos y consolidar los datos fragmentados en archivos más grandes y eficientes. La compactación se puede configurar para que se ejecute de forma automática o manual según sea necesario.
- Retención de instantáneas: las instantáneas son versiones con fecha y hora de una tabla de Iceberg. Las configuraciones de retención de instantáneas permiten a los clientes determinar cuánto tiempo se deben retener las instantáneas y cuántas instantáneas retener. La configuración de un optimizador de retención de instantáneas puede ayudar a administrar la sobrecarga de almacenamiento mediante la eliminación de las instantáneas antiguas e innecesarias y sus correspondientes archivos subyacentes.
- Eliminación de archivos huérfanos: los archivos huérfanos son archivos a los que los metadatos de la tabla de Iceberg ya no hacen referencia. Con el tiempo, estos archivos se pueden acumular, sobre todo después de operaciones como la eliminación de tablas o los errores en los trabajos de ETL. Habilitar la eliminación de archivos huérfanos AWS Glue permite identificar y eliminar periódicamente estos archivos innecesarios, liberando espacio de almacenamiento.

Puede activar o desactivar los optimizadores de compactación, retención de instantáneas y eliminación de archivos huérfanos para tablas Iceberg individuales del catálogo de datos mediante la AWS Glue consola o las operaciones de la API. AWS CLI AWS Glue

Para obtener más información, consulte Optimización de las tablas Iceberg en la Guía para desarrolladores. AWS Glue

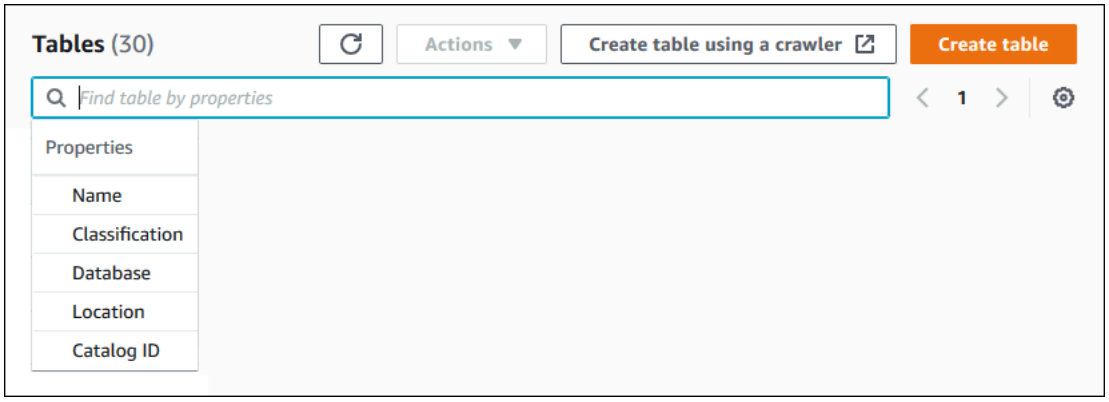
Búsqueda de tablas

Puede usar la AWS Lake Formation consola para buscar tablas del catálogo de datos por nombre, ubicación, base de datos contenedora, etc. Los resultados de la búsqueda muestran solo las tablas en las que tiene permisos de Lake Formation.

Para buscar tablas (consola)

1. Inicie sesión en la consola de Lake Formation Consola de administración de AWS y ábrala en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación, elija Tablas.
3. Coloca el cursor en el campo de búsqueda. El campo tiene el texto marcador de posición Buscar tabla por propiedades.

   Aparece el menú Propiedades, que muestra las distintas propiedades de la tabla por las que buscar.



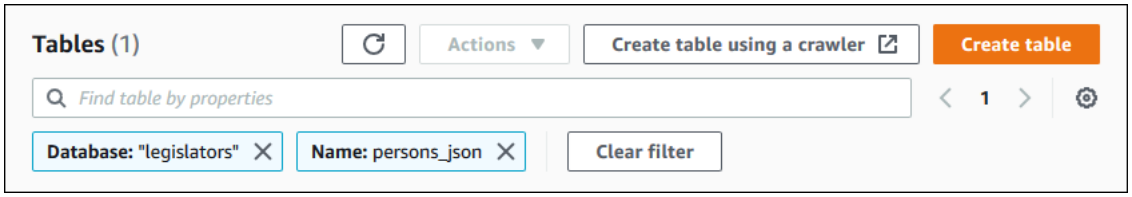
4. Realice una de las siguientes acciones:

   - Buscar por base de datos contenedora.
     1. Seleccione Base de datos en el menú Propiedades y, a continuación, elija una base de datos en el menú Bases de datos que aparece o escriba un nombre de base de datos y pulse Intro.

Se muestran las tablas sobre las que tiene permisos en la base de datos.

2. (Opcional) Para reducir la lista a una sola tabla de la base de datos, vuelva a colocar el cursor en el campo de búsqueda, elija Nombre en el menú Propiedades y elija un nombre de tabla en el menú Tablas que aparece o escriba un nombre de tabla y pulse Intro.

   Aparece la tabla individual y tanto el nombre de la base de datos como el nombre de la tabla aparecen como mosaicos debajo del campo de búsqueda.



   Para ajustar el filtro, cierre cualquiera de los mosaicos o seleccione Borrar filtro.

- Busque por otras propiedades.

  1. Seleccione una propiedad de búsqueda en el menú Propiedades.

     Para buscar por ID de AWS cuenta, seleccione ID de catálogo en el menú Propiedades, introduzca un ID de AWS cuenta válido (por ejemplo, 111122223333) y pulse Entrar.

     Para buscar por ubicación, elija Ubicación en el menú Propiedades y seleccione una en el menú Ubicaciones que aparece. Se devuelven todas las tablas de la ubicación raíz de la ubicación seleccionada (por ejemplo, Amazon S3).

Para buscar tablas, utilice AWS CLI

- En el siguiente ejemplo, se muestra cómo ejecutar una búsqueda parcial. El parámetro `--search-text` permite buscar tablas que contengan el texto especificado en sus metadatos. En este caso, devuelve todas las tablas que tienen la palabra “customer” en el nombre, la descripción u otros campos de metadatos.

```
aws glue search-tables
      --search-text "customer"
      --region Región de AWS
      --max-results 10
      --sort-criteria "FieldName=Name,Sort=ASC"
```

Compartir tablas y bases de datos del catálogo de datos entre AWS Cuentas

Puede compartir los recursos del catálogo de datos (bases de datos y tablas) con AWS cuentas externas concediendo permisos de Lake Formation sobre los recursos a las cuentas externas. A continuación, los usuarios pueden ejecutar consultas y trabajos para unir y consultar tablas en varias cuentas. Con algunas restricciones, cuando comparte un recurso del Catálogo de datos con otra cuenta, las entidades principales de esa cuenta pueden operar con ese recurso como si estuviera en su Catálogo de datos.

No comparte los recursos con directores específicos en AWS cuentas externas, sino que comparte los recursos con una AWS cuenta u organización. Cuando comparte un recurso con una organización AWS , está compartiendo el recurso con todas las cuentas de todos los niveles de esa organización. A continuación, el administrador del lago de datos de cada cuenta externa debe conceder permisos sobre los recursos compartidos a las entidades principales de su cuenta.

Para obtener más información, consulte Cross-account intercambio de datos en Lake Formation y Concesión de permisos sobre los recursos del Catálogo de datos.

Consulte también:

- Acceso y visualización de tablas y bases de datos compartidas del Catálogo de datos
- Requisitos previos

Construcción AWS Glue Data Catalog vistas

En el AWS Glue Data Catalog, una vista es una tabla virtual en la que el contenido se define mediante una consulta SQL que hace referencia a una o más tablas. Puede crear una vista de catálogo de datos que haga referencia a un máximo de 10 tablas mediante editores de SQL para Amazon Athena, Amazon Redshift o Apache Spark mediante EMR Serverless o la versión 5.0. AWS Glue Las tablas de referencia subyacentes de una vista pueden pertenecer a la misma base de datos o a bases de datos diferentes dentro del mismo Cuenta de AWS catálogo de datos.

Puede hacer referencia a AWS Glue tablas estándar y tablas en formatos de tabla abierta (OTF), como Apache Hudi, Linux Foundation Delta Lake y Apache Iceberg, con los datos subyacentes almacenados en las ubicaciones de Amazon S3 registradas en. AWS Lake Formation Además, puede crear vistas a partir de tablas federadas desde recursos compartidos de datos de Amazon Redshift que se compartan con Lake Formation.

Diferenciación de las vistas del Catálogo de datos de otros tipos de vistas

Las vistas del Catálogo de datos difieren de las vistas de Apache Hive, Apache Spark y Amazon Athena. La vista del catálogo de datos es una característica nativa de y es una vista AWS Glue Data Catalog creada por un definidor multidialecto. Puede crear una vista del Catálogo de datos utilizando uno de los servicios de análisis compatibles, como Athena o Amazon Redshift Spectrum, y acceder a la misma vista mediante otros servicios de análisis compatibles. Por otro lado, las vistas Apache Hive, Apache Spark y Athena se crean de forma independiente en cada servicio de análisis, como Athena y Amazon Redshift, y solo están visibles y accesibles dentro de dicho servicio.

¿Qué es una vista de definidor?

Una vista de definidor es una vista SQL que opera en función de los permisos de la entidad principal que la creó. El rol de definidor tiene los permisos necesarios para obtener acceso a las tablas a las que se hace referencia y ejecuta la instrucción SQL que define la vista. El definidor crea la vista y la comparte con otros usuarios mediante AWS Lake Formation un control de acceso detallado.

Cuando un usuario consulta la vista de definidor, el motor de consultas utiliza los permisos del rol de definidor para acceder a las tablas de referencia subyacentes. Este enfoque permite a los usuarios interactuar con la vista sin necesidad de acceder directamente a las tablas de origen, lo que mejora la seguridad y simplifica la administración del acceso a los datos.

Para configurar una vista de definición, la función de IAM del definidor puede estar en la misma AWS cuenta que las tablas base, o en una cuenta diferente, utilizando funciones de definición multicuenta. Para obtener más información sobre los permisos necesarios para el rol de definidor, consulte Requisitos previos para crear vistas.

Un marco para vistas multidialectales

El Catálogo de datos admite la creación de vistas utilizando varios dialectos de lenguaje de consultas estructurado (SQL). SQL es un lenguaje que se utiliza para almacenar y procesar información en una base de datos relacional, y cada motor de AWS análisis utiliza su propia variante de SQL, o dialecto de SQL.

Puede crear una vista del Catálogo de datos en un dialecto SQL mediante uno de los motores de consulta de análisis compatibles. A continuación, puede actualizar la vista utilizando la instrucción `ALTER VIEW` en un dialecto SQL diferente en cualquier otro motor de análisis compatible. Sin embargo, cada dialecto debe hacer referencia al mismo conjunto de tablas, columnas y tipos de datos.

Puede acceder a los distintos dialectos disponibles para la vista mediante la `GetTable` API y la consola. AWS CLI AWS Por lo tanto, la vista del Catálogo de datos está visible y disponible para realizar consultas en los diferentes motores de análisis compatibles.

Al definir un esquema de vista y un objeto de metadatos comunes que puede consultar desde varios motores, las vistas del catálogo de datos le permiten utilizar vistas uniformes de todo el lago de datos.

Para obtener más información sobre cómo se resuelve el esquema para cada dialecto, consulte el enlace a la referencia de la API. Para obtener más información sobre las reglas de coincidencia para los distintos tipos, consulte el enlace a la sección correspondiente del documento de la API.

Integración con permisos de Lake Formation

Se puede utilizar AWS Lake Formation para centralizar la administración de permisos en las AWS Glue Data Catalog vistas de los usuarios. Puede conceder permisos detallados en las vistas del catálogo de datos mediante el método de recurso indicado o LF-Tags compartirlos entre AWS organizaciones y unidades Cuentas de AWS organizativas. También puede compartir las vistas del catálogo de datos y acceder a ellas Regiones de AWS mediante enlaces a recursos. Esto permite a los usuarios proporcionar acceso a los datos sin duplicar el origen de datos y compartiendo las tablas subyacentes.

La declaración `CREATE VIEW` DDL de una vista de catálogo de datos puede hacer referencia a AWS Glue las tablas estándar y a las tablas en formatos de tabla abierta (OTF), como Hudi, Delta Lake e Iceberg, con datos subyacentes almacenados en ubicaciones de Amazon S3 registradas en Lake Formation, así como a las tablas federadas del datashare de Amazon Redshift que se comparten con Lake Formation. Las tablas pueden tener cualquier formato de archivo siempre que el motor utilizado para consultar la vista admita dicho formato. También puede hacer referencia a funciones integradas del motor en el que se ejecuta, aunque es posible que no se permitan otros recursos específicos del motor. Para obtener más información, consulte Vistas, consideraciones y limitaciones del catálogo de datos

Casos de uso

A continuación se detallan los casos de uso importantes de las vistas del Catálogo de datos:

- Crear y administrar permisos en un esquema de vista única. Esto le ayuda a evitar el riesgo de que se existan permisos incoherentes en vistas duplicadas creadas en varios motores.
- Conceda permisos a los usuarios en una vista que haga referencia a varias tablas sin conceder permisos directamente en las tablas de referencia subyacentes.

- Logre el filtrado a nivel de fila en las tablas utilizando LF-Tags (donde la LF-Tags cascada solo llega hasta el nivel de columna) LF-Tags aplicándolo a las vistas y concediendo permisos LF-Tags basados a los usuarios.

compatible AWS servicios de análisis de vistas

Los siguientes servicios AWS de análisis permiten crear vistas del catálogo de datos:

- Amazon Redshift
- Amazon Athena versión 3
- Apache Spark en EMR sin servidor
- Apache Spark en la AWS Glue versión 5.0

Recursos adicionales

Puede obtener más información sobre el Catálogo de datos en esta guía del usuario, así como en los siguientes recursos:

En el siguiente vídeo se muestra cómo crear vistas y consultarlas desde Athena y Amazon Redshift.

Temas

- Requisitos previos para crear vistas
- Creación de vistas del Catálogo de datos mediante instrucciones DDL
- Creación de vistas del catálogo de datos mediante AWS Glue API
- Concesión de permisos del vistas del catálogo de datos
- Vistas materializadas

Requisitos previos para crear vistas

- Para crear vistas en el catálogo de datos, debe registrar las ubicaciones de datos subyacentes de Amazon S3 de las tablas de referencia en Lake Formation. Para obtener más información sobre el registro de datos con Lake Formation, consulteAñadir una ubicación de Amazon S3 a su lago de datos.
- Solo los roles de IAM pueden crear vistas del Catálogo de datos. Otras identidades de IAM no pueden crear vistas del catálogo de datos.

- El rol de IAM que defina la vista debe tener los siguientes permisos:
  - Permiso `SELECT` de Lake Formation con la opción `Grantable` en todas las tablas de referencia, incluidas todas las columnas.
  - Permiso `CREATE_TABLE` de Lake Formation en la base de datos de destino donde se crean las vistas.
  - Una política de confianza para que Lake Formation y AWS Glue sus servicios asuman el cargo.

    JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerAssumeRole1",
            "Effect": "Allow",
            "Principal": {
               "Service": [
                    "glue.amazonaws.com",
                    "lakeformation.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

  - El objetivo: PassRole permiso para AWS Glue y Lake Formation.

    JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerPassRole1",
            "Action": [
                "iam:PassRole"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
```

```
                "StringEquals": {
                    "iam:PassedToService": [
                        "glue.amazonaws.com",
                        "lakeformation.amazonaws.com"
                      ]
                }
            }
        }
    ]
}
```

- AWS Glue y permisos de Lake Formation.

  JSON

```
{
    "Version":"2012-10-17",
                "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "Glue:GetDatabase",
                "Glue:GetDatabases",
                "Glue:CreateTable",
                "Glue:GetTable",
                "Glue:GetTables",
                "Glue:BatchGetPartition",
                "Glue:GetPartitions",
                "Glue:GetPartition",
                "Glue:GetTableVersion",
                "Glue:GetTableVersions",
    "Glue:PassConnection",
                "lakeFormation:GetDataAccess"
            ],
            "Resource": "*"
        }
    ]
}
```

- No puede crear vistas en una base de datos que tenga los permisos `Super` o `ALL` concedidos al grupo `IAMAllowedPrincipals`. Puede revocar el permiso `Super` del grupo `IAMAllowedPrincipals` en una base de datos, ver Paso 4: Cambiar sus almacenes de datos al modelo de permisos de Lake Formation o crear una nueva base de datos con la casilla Usar solo

el control de acceso de IAM para las nuevas tablas de esta base de datos sin marcar en la sección Permisos predeterminados para bases de datos y tablas recién creadas.

Creación de vistas del Catálogo de datos mediante instrucciones DDL

Puede crear AWS Glue Data Catalog vistas mediante editores de SQL para Athena, Amazon Redshift y mediante las API/. AWS Glue AWS CLI

Para crear una vista del Catálogo de datos mediante editores de SQL, elija Athena o Redshift Spectrum y cree la vista utilizando una instrucción de lenguaje de definición de datos (DDL) de tipo `CREATE VIEW`. Tras crear una vista en el dialecto del primer motor, puede utilizar una instrucción DDL de tipo `ALTER VIEW` desde el segundo motor para agregar dialectos adicionales.

Al definir las vistas, es importante que tenga en cuenta lo siguiente:

- Definición de vistas multidialectales: a la hora de definir una vista con varios dialectos, los esquemas de los distintos dialectos deben coincidir. Cada dialecto SQL tendrá una especificación de sintaxis ligeramente distinta. La sintaxis de la consulta que define la vista del Catálogo de datos debe apuntar exactamente a la misma lista de columnas, incluidos los tipos y los nombres, en todos los dialectos. Esta información se almacena en el `StorageDescriptor` de la vista. Los dialectos también deben hacer referencia a los mismos objetos de tabla subyacentes del Catálogo de datos.

  Para agregar otro dialecto a una vista mediante DDL, puede utilizar la instrucción `ALTER VIEW`. Si una instrucción `ALTER VIEW` intenta actualizar la definición de la vista, por ejemplo, modificando el descriptor de almacenamiento o las tablas subyacentes de la vista, la instrucción devuelve un error que indica que la entrada y el descriptor de almacenamiento existente no coinciden. Puede utilizar operaciones de conversión de SQL para asegurarse de que los tipos de columnas de la vista coincidan.
- Actualización de una vista: para actualizar la vista, puede usar la API `UpdateTable`. Si actualiza la vista sin hacer coincidir los descriptores de almacenamiento o las tablas de referencia, puede proporcionar el indicador `FORCE` (consulte la documentación de SQL del motor para conocer la sintaxis). Tras una actualización forzada, la vista incluirá el `StorageDescriptor` forzado y las tablas de referencia. Cualquier otro DDL `ALTER VIEW` debería coincidir con los valores modificados. Una vista que se haya actualizado para incluir dialectos incompatibles pasará a tener el estado “Obsoleto”. El estado de la vista está visible en la consola de Lake Formation y mediante la operación `GetTable`.

- Referencia a un tipo de columna varchar como cadena: no es posible convertir un tipo de columna varchar de Redshift Spectrum en una cadena. Si se crea una vista en Redshift Spectrum con un tipo de columna varchar y un dialecto posterior intenta hacer referencia a dicho campo como cadena, el Catálogo de datos lo tratará como una cadena sin necesidad del indicador `FORCE`.
- Tratamiento de campos de tipos complejos: Amazon Redshift trata todos los tipos complejos como tipos SUPER, mientras que Athena especifica el tipo complejo. Si una vista tiene un campo de tipo SUPER y otro motor hace referencia a esa columna como un tipo complejo concreto, como struct (`<street_address:struct<street_number:int, street_name:string, street_type:string>>`), el Catálogo de datos asume que el campo es de un tipo complejo específico y lo utiliza en el descriptor de almacenamiento, sin necesidad de utilizar el indicador `Force`.

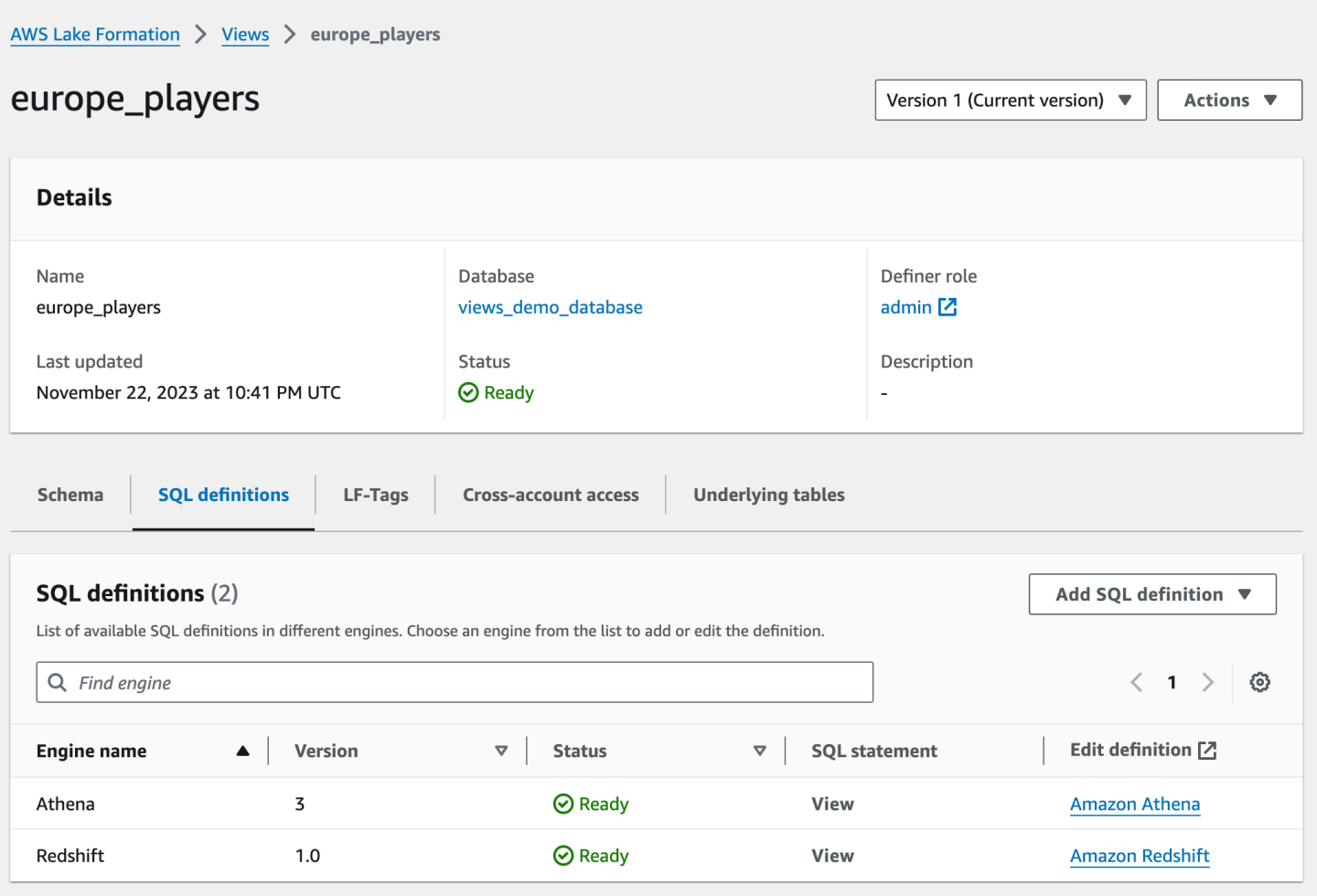
Para obtener más información sobre la sintaxis para crear y administrar vistas del catálogo de datos, consulte:

- Uso de las vistas de AWS Glue Data Catalog de la Guía del usuario de Amazon Athena.
- Sintaxis de vista del Catálogo de datos de Glue en la Guía del usuario de Amazon Athena.
- Creación de vistas en AWS Glue Data Catalog en la Guía para desarrolladores de bases de datos de Amazon Redshift.

  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.

Tras crear una vista del Catálogo de datos, los detalles de la vista están disponibles en la consola de Lake Formation.

1. Seleccione Vistas en el catálogo de datos en la consola de Lake Formation.
2. Aparece una lista de las vistas disponibles en la página de vistas.
3. Seleccione una vista de la lista y la página de detalles mostrará los atributos de la vista.



Esquema

Elija una `Column` fila y seleccione Editar LF-Tags para actualizar los valores de las etiquetas o asignar nuevos valores. LF-Tags

Definiciones de SQL

Puede ver una lista completa de las definiciones de SQL disponibles. Seleccione Añadir definición de SQL y elija un motor de consultas para añadir una definición de SQL. Elija un motor de consulta (Athena o Amazon Redshift) en la columna `Edit definition` para actualizar una definición de SQL.

LF-Tags

Elija Editar LF-Tags para editar los valores de una etiqueta o asignar etiquetas nuevas. Se puede utilizar LF-Tags para conceder permisos sobre las vistas.

Cross-account acceder

Puede ver una lista de Cuentas de AWS las organizaciones y unidades organizativas (OU) con las que ha compartido la vista del catálogo de datos.

Tablas subyacentes

Las tablas subyacentes a las que se hace referencia en la definición de SQL utilizadas para crear la vista se muestran en esta pestaña.

Creación de vistas del catálogo de datos mediante AWS Glue API

Puede usar y AWS Glue CreateTablelas UpdateTableAPI para crear y actualizar vistas en el catálogo de datos. Las operaciones `CreateTable` y `UpdateTable` tienen una nueva estructura `TableInput` para `ViewDefinition`, mientras que las operaciones `SearchTables`, `GetTable`, `GetTables`, `GetTableVersion` y `GetTableVersions` proporcionan la `ViewDefinition` en su sintaxis de salida para las vistas. Además, hay un nuevo campo `Status` en la salida de la API `GetTable`.

Hay dos nuevas AWS Glue conexiones disponibles para validar el dialecto SQL para cada motor de consultas compatible Amazon Athena y Amazon Redshift.

Las API `CreateTable` y `UpdateTable` son asíncronas cuando se utilizan con vistas. Cuando se llama a estas API con varios dialectos SQL, la llamada se valida con cada motor para determinar si el dialecto se puede ejecutar en ese motor y si el esquema resultante de la vista de cada dialecto coincide. El AWS Glue servicio utiliza estas conexiones para realizar llamadas internas a los motores de análisis. Estas llamadas simulan lo que hace el motor para validar si se ejecutó un DDL `CREATE VIEW` o un DDL `ALTER VIEW` de SQL en el motor.

Si el SQL proporcionado es válido y los esquemas de los dialectos de las vistas coinciden, la API de AWS Glue confirma el resultado de forma atómica. La atomicidad permite crear o modificar vistas con varios dialectos sin ningún tiempo de inactividad.

Temas

- Creación AWS Glue conexiones para validar el estado
- Validación del estado de generación de vistas
- Estados y operaciones asíncronos
- Visualización de los escenarios de errores de creación durante las operaciones asincrónicas

Creación AWS Glue conexiones para validar el estado

Para crear o actualizar una AWS Glue Data Catalog vista mediante las `UpdateTable` operaciones `CreateTable` o, debe crear un nuevo tipo de AWS Glue conexión para la validación y proporcionarlo al motor de análisis compatible. Estas conexiones son necesarias para utilizar las vistas del Catálogo de datos con Athena o Amazon Redshift. Puede crear estas conexiones únicamente mediante AWS CLI los AWS SDK o las AWS Glue API. No puedes usar el Consola de administración de AWS para crear la AWS Glue conexión.

> **Note**
>
> Si el rol de definidor de la vista y el rol que realiza la llamada a `CreateTable` o `UpdateTable` no coinciden, ambos necesitarán un permiso `glue:PassConnection` en la declaración de su política de IAM.

Para obtener más información, consulta la documentación de creación de conexión AWS CLI .

AWS CLI comando para crear una conexión

El siguiente es un AWS CLI comando para crear una conexión:

```
aws glue create-connection --region us-east-1
--endpoint-url https://glue.us-east-1.amazonaws.com
--cli-input-json file:///root/path/to/create-connection.json
```

AWS CLI introduzca JSON

Para Amazon Redshift:

```
{
    "CatalogId": "123456789012",
    "ConnectionInput": {
        "ConnectionType": "VIEW_VALIDATION_REDSHIFT",
        "Name": "views-preview-cluster-connection-2",
        "Description": "My first Amazon Redshift validation connection",
        "ConnectionProperties": {
            "DATABASE": "dev",
            "CLUSTER_IDENTIFIER": "glue-data-catalog-views-preview-cluster"
        }
```

```
    }
}
```

Para Amazon Athena:

```
{
    "CatalogId": "123456789012",
    "ConnectionInput": {
        "ConnectionType": "VIEW_VALIDATION_ATHENA",
        "Name": "views-preview-cluster-connection-3",
        "Description": "My first Amazon Athena validation connection",
        "ConnectionProperties": {
            "WORKGROUP_NAME": "workgroup-name"
        }
    }
}
```

Validación del estado de generación de vistas

Al ejecutar las operaciones `CreateTable` o `UpdateTable`, el campo `Status` del resultado de la API `GetTable` muestra los detalles del estado de creación de la vista. Para `create` las solicitudes en las que la tabla aún no existe, AWS Glue crea una tabla vacía durante el proceso asíncrono. Al llamar a `GetTable`, puede incluir un indicador booleano opcional `IncludeStatusDetails`, que muestra información de diagnóstico sobre la solicitud. En caso de error, este indicador muestra un mensaje de error con los estados individuales de cada dialecto.

Los errores durante las operaciones de creación, lectura, actualización y eliminación (CRUD) pueden producirse durante el procesamiento en el servicio AWS Glue/Lake Formation o durante la validación de View SQL en Amazon Redshift o Athena. Cuando se produce un error durante la validación en un motor, el servicio de AWS Glue proporciona el mensaje de error que devuelve el motor.

Campos de estado

Los campos de estado son los siguientes:

- Estado: un estado genérico, independiente de los diferentes tipos de trabajos:
  - QUEUED
  - IN_PROGRESS
  - SUCCESS
  - ERROR

- Acción: indica a qué acción de la tabla se ha llamado; actualmente solo están disponibles las operaciones `CREATE` o `UPDATE`.

  Cuando se trabaja con vistas, es importante distinguir entre operaciones `UPDATE` y `CREATE`. El tipo de operación determina cómo se debe proceder con la consulta de las tablas.

  Una operación `UPDATE` significa que la tabla ya existe en el Catálogo de datos. En este caso, puede seguir consultando la tabla creada anteriormente sin ningún problema. Por otro lado, una operación `CREATE` indica que la tabla no se ha creado correctamente con anterioridad. Si una tabla está marcada como `CREATE`, se producirá un error al intentar consultarla porque la tabla aún no existe en el sistema. Por lo tanto, es imprescindible identificar el tipo de operación (UPDATE o CREATE) antes de intentar consultar una tabla.
- RequestedBy — El ARN del usuario que solicitó el cambio asincrónico.
- UpdatedBy — El ARN del usuario que modificó manualmente por última vez el proceso de cambio asincrónico, por ejemplo, al solicitar una cancelación o modificación.
- Error: este campo solo aparece cuando el estado es FAILED. Este es un mensaje de excepción de nivel principal. Puede haber errores diferentes para cada dialecto.
  - ErrorCode — El tipo de excepción.
  - ErrorMessage — una breve descripción de la excepción.
- RequestTime — una cadena de fecha con formato ISO 8601 que indica la hora en que se inició el cambio.
- UpdateTime — una cadena de fecha con formato ISO 8601 que indica la hora a la que se actualizó el estado por última vez.

Estados y operaciones asíncronos

Al ejecutar una solicitud `glue:CreateTable`, comienza la creación asincrónica de la vista del Catálogo de datos. En las siguientes secciones, este documento describe `Status` la AWS Glue vista que está disponible en una respuesta. `glue:GetTable` Por motivos de brevedad, en esta sección se omite la respuesta completa.

```
{
    "Table": {
        ...
        "Status": {
            ...
            "Action": "CREATE",
```

```
            "State": "QUEUED",
        }
    }
}
```

Los dos atributos anteriores constituyen información de diagnóstico importante que indica el estado de la operación asíncrona, así como las acciones que se pueden realizar en esta vista. A continuación se muestran los posibles valores que pueden adoptar estos atributos.

1. `Status.Action`
   a. CREATE
   b. UPDATE
2. `Status.State`
   a. QUEUED
   b. IN_PROGRESS
   c. SUCCESS
   d. ERROR

También es importante tener en cuenta que algunas actualizaciones de una vista del Catálogo de datos no requieren una operación asíncrona. Por ejemplo, es posible que desee actualizar el atributo de `Description` de la tabla. Como esto no requiere ninguna operación asíncrona, los metadatos de la tabla resultante no tendrán ningún `Status`, y el atributo será `NULL`.

```
{
    "Table": {
        ...,
        "Description": "I changed this attribute!"
    }
}
```

A continuación, en este tema se analiza cómo la información de estado anterior puede afectar a las operaciones que se pueden realizar en una AWS Glue vista.

pegamento: CreateTable

No hay cambios en esta API en comparación con el funcionamiento de `glue:CreateTable` en cualquier tabla de Glue. `CreateTable` acepta llamadas para cualquier nombre de tabla que aún no exista.

pegamento: UpdateTable

Esta operación no se puede realizar en una AWS Glue vista que tenga la siguiente información de estado:

1. Acción == CREATE y Estado == QUEUED
2. Acción == CREATE y Estado == IN_PROGRESS
3. Acción == CREATE y Estado == FAILED
4. Acción == UPDATE y Estado == QUEUED
5. Acción == UPDATE y Estado == IN_PROGRESS

En resumen, puede actualizar una vista del Catálogo de datos solo cuando esta cumpla los siguientes requisitos.

1. Se ha creado con éxito por primera vez.
   a. Acción == CREATE y Estado == SUCCESS
2. Ha alcanzado un estado terminal tras una operación de actualización asíncrona.
   a. Acción == UPDATE y Estado == SUCCESS
   b. Acción == UPDATE y Estado == FAILED
3. Tiene un atributo de estado `NULL` como resultado de una actualización sincrónica.

pegamento: DeleteTable

No hay cambios en esta operación en comparación con el `glue:DeleteTable` funcionamiento de cualquier AWS Glue tabla. Puede eliminar una vista del Catálogo de datos independientemente de su estado.

pegamento: GetTable

No hay cambios en esta operación en comparación con el `glue:GetTable` funcionamiento de cualquier AWS Glue tabla. Sin embargo, no puede consultar una vista del Catálogo de datos desde los motores de análisis hasta que se haya creado correctamente por primera vez. `Action == CREATE and State == SUCCESS`. Tras crear correctamente una vista del Catálogo de datos por primera vez, puede consultar la vista con independencia del estado en que se encuentre.

> **Note**
>
> Toda la información de esta sección se aplica a todas las API de lectura de tablas, como `GetTable`, `GetTables` y `SearchTables`.

Visualización de los escenarios de errores de creación durante las operaciones asincrónicas

Los siguientes ejemplos son representativos de los tipos de errores que pueden resultar de las llamadas a la API de las vistas `CreateTable` o `UpdateTable`. No son exhaustivos, ya que la superficie de error de los errores en las consultas SQL es bastante amplia.

Escenario 1: Error de consulta en Amazon Redshift

La consulta proporcionada para Amazon Redshift incluye un nombre de tabla mal escrito que no se encuentra en el Catálogo de datos durante la validación. El error resultante se muestra en el campo `Status` de la respuesta de `GetTable` de la vista.

Solicitud de `GetTable`:

```
{
    "CatalogId": "123456789012",
    "DatabaseName": "async-view-test-db",
    "TableInput": {
        "Name": "view-athena-redshift-72",
        "Description": "This is an atomic operation",
        "StorageDescriptor": {
            "Columns": [
                { "Name": "col1", "Type": "int" },
                { "Name": "col2", "Type": "string" },
                { "Name": "col3", "Type": "double" }
            ]
        },
        "ViewDefinition": {
            "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
            "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1" ],
            "Representations": [
                {
                    "Dialect": "ATHENA",
                    "DialectVersion": "3",
```

```
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-db\".
\"table_1\"",
                    "ValidationConnection": "athena-connection"
                },
                {
                    "Dialect": "REDSHIFT",
                    "DialectVersion": "1.0",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-
schema\".\"table__1\";",
                    "ValidationConnection": "redshift-connection"
                }
            ]
        }
    }
}
```

Respuesta de `GetTable`:

```
IncludeStatusDetails = FALSE
{
    "Table": {
        "Name": "view-athena-redshift-72",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:39:19-07:00",
        "UpdateTime": "2024-07-11T11:39:19-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:39:19-07:00",
            "UpdateTime": "2024-07-11T11:40:06-07:00",
            "Action": "CREATE",
```

```
            "State": "FAILED"
        }
    }
}

IncludeStatusDetails = TRUE
{
    "Table": {
        "Name": "view-athena-redshift-72",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:39:19-07:00",
        "UpdateTime": "2024-07-11T11:39:19-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:39:19-07:00",
            "UpdateTime": "2024-07-11T11:40:06-07:00",
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "QueryExecutionException",
                "ErrorMessage": "Error received during view SQL validation
 using a connection: [Connection Name: redshift-connection | Query Execution
 Id: ddb711d3-2415-4aa9-b251-6a76ab4f41b1 | Timestamp: Thu Jul 11 18:39:37 UTC
 2024]: Redshift returned error for the statement: ERROR: AwsClientException:
 EntityNotFoundException from glue - Entity Not Found"
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-72",
                    "DatabaseName": "async-view-test-db",
                    "Description": "This is an atomic operation",
```

```
                    "Retention": 0,
                    "StorageDescriptor": {
                        "Columns": [
                            {
                                "Name": "col1",
                                "Type": "int"
                            },
                            {
                                "Name": "col2",
                                "Type": "string"
                            },
                            {
                                "Name": "col3",
                                "Type": "double"
                            }
                        ],
                        "Compressed": false,
                        "NumberOfBuckets": 0,
                        "SortColumns": [],
                        "StoredAsSubDirectories": false
                    },
                    "TableType": "VIRTUAL_VIEW",
                    "IsRegisteredWithLakeFormation": false,
                    "CatalogId": "123456789012",
                    "IsRowFilteringEnabled": false,
                    "VersionId": "-1",
                    "DatabaseId": "<databaseID>",
                    "ViewDefinition": {
                        "IsProtected": true,
                        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
                        "SubObjects": [
                            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
                        ],
                        "Representations": [
                            {
                                "Dialect": "ATHENA",
                                "DialectVersion": "3",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-db\".\"table_1\"",
                                "IsStale": false
                            },
                            {
                                "Dialect": "REDSHIFT",
```

```
                                "DialectVersion": "1.0",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-external-schema\".\"table__1\";",
                                "IsStale": false
                            }
                        ]
                    },
                    "IsMultiDialectView": true
                },
                "ViewValidations": [
                    {
                        "Dialect": "ATHENA",
                        "DialectVersion": "3",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-db
\".\"table_1\"",
                        "UpdateTime": "2024-07-11T11:40:06-07:00",
                        "State": "SUCCESS"
                    },
                    {
                        "Dialect": "REDSHIFT",
                        "DialectVersion": "1.0",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-
external-schema\".\"table__1\";",
                        "UpdateTime": "2024-07-11T11:39:37-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "QueryExecutionException",
                            "ErrorMessage": "Error received during view SQL validation
 using a connection: [Connection Name: redshift-connection | Query Execution Id:
 ddb711d3-2415-4aa9-b251-6a76ab4f41b1 | Timestamp: Thu
 Jul 11 18:39:37 UTC 2024]: Redshift returned error for the statement: ERROR:
 AwsClientException: EntityNotFoundException from glue - Entity Not Found"
                        }
                    }
                ]
            }
        }
    }
}
```

Escenario 2: Conexión a Amazon Redshift no válida

La conexión de Amazon Redshift del siguiente ejemplo tiene un formato incorrecto porque hace referencia a una base de datos de Amazon Redshift que no existe en el punto de conexión proporcionado. cluster/serverless Amazon Redshift no puede validar la vista y el campo `Status` de la respuesta de `GetTable` muestra el error (`"State": "FAILED"` desde Amazon Redshift).

Solicitud de `GetTable`:

```
{
    "CatalogId": "123456789012",
    "DatabaseName": "async-view-test-db",
    "TableInput": {
        "Name": "view-athena-redshift-73",
        "Description": "This is an atomic operation",
        "StorageDescriptor": {
            "Columns": [
                { "Name": "col1", "Type": "int" },
                { "Name": "col2", "Type": "string" },
                { "Name": "col3", "Type": "double" }
            ]
        },
        "ViewDefinition": {
            "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
            "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1" ],
            "Representations": [
                {
                    "Dialect": "ATHENA",
                    "DialectVersion": "3",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-db\".
\"table_1\"",
                    "ValidationConnection": "athena-connection"
                },
                {
                    "Dialect": "REDSHIFT",
                    "DialectVersion": "1.0",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-
schema\".\"table_1\";",
                    "ValidationConnection": "redshift-connection-malformed"
                }
            ]
        }
```

```
    }
}
```

Respuesta de `GetTable`:

```
IncludeStatusDetails = FALSE
{
    "Table": {
        "Name": "view-athena-redshift-73",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:43:27-07:00",
        "UpdateTime": "2024-07-11T11:43:27-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:43:27-07:00",
            "UpdateTime": "2024-07-11T11:43:40-07:00",
            "Action": "CREATE",
            "State": "FAILED"
        }
    }
}

IncludeStatusDetails = TRUE
{
    "Table": {
        "Name": "view-athena-redshift-73",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:43:27-07:00",
        "UpdateTime": "2024-07-11T11:43:27-07:00",
```

```
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:43:27-07:00",
            "UpdateTime": "2024-07-11T11:43:40-07:00",
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "QueryExecutionException",
                "ErrorMessage": "Error received during view SQL validation using a
connection: [Connection Name: redshift-connection-malformed | Query Execution Id:
69bfafd4-3d51-4cb0-9320-7ce5404b1809 | Timestamp: Thu Jul 11 18:43:38 UTC 2024]:
Redshift returned error for the statement: FATAL: database \"devooo\" does not exist"
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-73",
                    "DatabaseName": "async-view-test-db",
                    "Description": "This is an atomic operation",
                    "Retention": 0,
                    "StorageDescriptor": {
                        "Columns": [
                            {
                                "Name": "col1",
                                "Type": "int"
                            },
                            {
                                "Name": "col2",
                                "Type": "string"
                            },
                            {
                                "Name": "col3",
                                "Type": "double"
```

```
                            }
                        ],
                        "Compressed": false,
                        "NumberOfBuckets": 0,
                        "SortColumns": [],
                        "StoredAsSubDirectories": false
                    },
                    "TableType": "VIRTUAL_VIEW",
                    "IsRegisteredWithLakeFormation": false,
                    "CatalogId": "123456789012",
                    "IsRowFilteringEnabled": false,
                    "VersionId": "-1",
                    "DatabaseId": "<databaseID>",
                    "ViewDefinition": {
                        "IsProtected": true,
                        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
                        "SubObjects": [
                            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
                        ],
                        "Representations": [
                            {
                                "Dialect": "ATHENA",
                                "DialectVersion": "3",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-db\".\"table_1\"",
                                "IsStale": false
                            },
                            {
                                "Dialect": "REDSHIFT",
                                "DialectVersion": "1.0",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-external-schema\".\"table_1\";",
                                "IsStale": false
                            }
                        ]
                    },
                    "IsMultiDialectView": true
                },
                "ViewValidations": [
                    {
                        "Dialect": "ATHENA",
                        "DialectVersion": "3",
```

```
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-db
\".\"table_1\"",
                        "UpdateTime": "2024-07-11T11:43:40-07:00",
                        "State": "SUCCESS"
                    },
                    {
                        "Dialect": "REDSHIFT",
                        "DialectVersion": "1.0",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-
external-schema\".\"table_1\";",
                        "UpdateTime": "2024-07-11T11:43:38-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "QueryExecutionException",
                            "ErrorMessage": "Error received during view SQL validation
 using a connection: [Connection Name: redshift-connection-malformed | Query Execution
 Id: 69bfafd4-3d51-4cb0-9320-7ce5404b1809 | Time
stamp: Thu Jul 11 18:43:38 UTC 2024]: Redshift returned error for the statement: FATAL:
 database \"devooo\" does not exist"
                        }
                    }
                ]
            }
        }
    }
}
```

Escenario 3: Error de consulta en Athena

El SQL de Athena no es válido porque la consulta tiene mal escrito el nombre de la base de datos. La validación de consultas de Athena detecta esto y el error resultante aparece en el objeto `Status` de una llamada `GetTable`.

Solicitud de `GetTable`:

```
{
    "CatalogId": "123456789012",
    "DatabaseName": "async-view-test-db",
    "TableInput": {
        "Name": "view-athena-redshift-70",
        "Description": "This is an atomic operation",
```

```
        "StorageDescriptor": {
            "Columns": [
                { "Name": "col1", "Type": "int" },
                { "Name": "col2", "Type": "string" },
                { "Name": "col3", "Type": "double" }
            ]
        },
        "ViewDefinition": {
            "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
            "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1" ],
            "Representations": [
                {
                    "Dialect": "ATHENA",
                    "DialectVersion": "3",
                    "ViewOriginalText": "SELECT * FROM \"gdc--view-playground-db\".
\"table_1\"",
                    "ValidationConnection": "athena-connection"
                },
                {
                    "Dialect": "REDSHIFT",
                    "DialectVersion": "1.0",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-
schema\".\"table_1\";",
                    "ValidationConnection": "redshift-connection"
                }
            ]
        }
    }
}
```

Respuesta de `GetTable`:

```
IncludeStatusDetails = FALSE
{
    "Table": {
        "Name": "view-athena-redshift-70",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:09:53-07:00",
        "UpdateTime": "2024-07-11T11:09:53-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
```

```
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:09:54-07:00",
            "UpdateTime": "2024-07-11T11:10:41-07:00",
            "Action": "CREATE",
            "State": "FAILED",
        }
    }
}

IncludeStatusDetails = TRUE
{
    "Table": {
        "Name": "view-athena-redshift-70",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:09:53-07:00",
        "UpdateTime": "2024-07-11T11:09:53-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:09:54-07:00",
            "UpdateTime": "2024-07-11T11:10:41-07:00",
```

```
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "QueryExecutionException",
                "ErrorMessage": "Error received during view SQL validation using
 a connection: [Connection Name: athena-connection | Query Execution Id: d9bb1e6d-
ce26-4b35-8276-8a199af966aa | Timestamp: Thu Jul 11 18:10:
41 UTC 2024]: Athena validation FAILED: {ErrorCategory: 2,ErrorType: 1301,Retryable:
 false,ErrorMessage: line 1:118: Schema 'gdc--view-playground-db' does not exist}"
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-70",
                    "DatabaseName": "async-view-test-db",
                    "Description": "This is an atomic operation",
                    "Retention": 0,
                    "StorageDescriptor": {
                        "Columns": [
                            {
                                "Name": "col1",
                                "Type": "int"
                            },
                            {
                                "Name": "col2",
                                "Type": "string"
                            },
                            {
                                "Name": "col3",
                                "Type": "double"
                            }
                        ],
                        "Compressed": false,
                        "NumberOfBuckets": 0,
                        "SortColumns": [],
                        "StoredAsSubDirectories": false
                    },
                    "TableType": "VIRTUAL_VIEW",
                    "IsRegisteredWithLakeFormation": false,
                    "CatalogId": "123456789012",
                    "IsRowFilteringEnabled": false,
                    "VersionId": "-1",
                    "DatabaseId": "<databaseID>",
                    "ViewDefinition": {
                        "IsProtected": true,
```

```
                        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
                        "SubObjects": [
                            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
                        ],
                        "Representations": [
                            {
                                "Dialect": "ATHENA",
                                "DialectVersion": "3",
                                "ViewOriginalText": "SELECT * FROM \"gdc--view-
playground-db\".\"table_1\"",
                                "IsStale": false
                            },
                            {
                                "Dialect": "REDSHIFT",
                                "DialectVersion": "1.0",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-external-schema\".\"table_1\";",
                                "IsStale": false
                            }
                        ]
                    },
                    "IsMultiDialectView": true
                },
                "ViewValidations": [
                    {
                        "Dialect": "ATHENA",
                        "DialectVersion": "3",
                        "ViewValidationText": "SELECT * FROM \"gdc--view-playground-db
\".\"table_1\"",
                        "UpdateTime": "2024-07-11T11:10:41-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "QueryExecutionException",
                            "ErrorMessage": "Error received during view SQL validation
 using a connection: [Connection Name: athena-connection | Query Execution Id:
 d9bb1e6d-ce26-4b35-8276-8a199af966aa | Timestamp: Thu J
ul 11 18:10:41 UTC 2024]: Athena validation FAILED: {ErrorCategory: 2,ErrorType:
 1301,Retryable: false,ErrorMessage: line 1:118: Schema 'gdc--view-playground-db' does
 not exist}"
                        }
                    },
                    {
                        "Dialect": "REDSHIFT",
```

```
                        "DialectVersion": "1.0",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-
external-schema\".\"table_1\";",
                        "UpdateTime": "2024-07-11T11:10:41-07:00",
                        "State": "SUCCESS"
                    }
                ]
            }
        }
    }
}
```

Escenario 4: Los descriptores de almacenamiento no coinciden

El SQL proporcionado para el dialecto de Athena selecciona `col1` y `col2`, mientras que el SQL para Redshift selecciona solo `col1`. Esto provoca un error de correspondencia en el descriptor de almacenamiento.

Solicitud de `GetTable`:

```
{
    "CatalogId": "123456789012",
    "DatabaseName": "async-view-test-db",
    "TableInput": {
        "Name": "view-athena-redshift-71",
        "Description": "This is an atomic operation",
        "StorageDescriptor": {
            "Columns": [
                { "Name": "col1", "Type": "int" },
                { "Name": "col2", "Type": "string" },
                { "Name": "col3", "Type": "double" }
            ]
        },
        "ViewDefinition": {
            "Definer": "arn:aws:iam::123456789012:role/GDCViewDdefiner",
            "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1" ],
            "Representations": [
                {
                    "Dialect": "ATHENA",
                    "DialectVersion": "3",
                    "ViewOriginalText": "SELECT col1, col2 FROM \"gdc-view-playground-
db\".\"table_1\"",
```

```
                    "ValidationConnection": "athena-connection"
                },
                {
                    "Dialect": "REDSHIFT",
                    "DialectVersion": "1.0",
                    "ViewOriginalText": "SELECT col1 FROM \"gdc-view-playground-
external-schema\".\"table_1\";",
                    "ValidationConnection": "redshift-connection"
                }
            ]
        }
    }
}
```

Respuesta de `GetTable`:

```
IncludeStatusDetails = FALSE

{
    "Table": {
        "Name": "view-athena-redshift-71",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:22:02-07:00",
        "UpdateTime": "2024-07-11T11:22:02-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:22:02-07:00",
            "UpdateTime": "2024-07-11T11:23:19-07:00",
            "Action": "CREATE",
            "State": "FAILED"
```

```
        }
    }
}

IncludeStatusDetails = TRUE

{
    "Table": {
        "Name": "view-athena-redshift-71",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:22:02-07:00",
        "UpdateTime": "2024-07-11T11:22:02-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:22:02-07:00",
            "UpdateTime": "2024-07-11T11:23:19-07:00",
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "InvalidInputException",
                "ErrorMessage": "Engine and existing storage descriptor mismatch"
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-71",
                    "DatabaseName": "async-view-test-db",
                    "Description": "This is an atomic operation",
                    "Retention": 0,
                    "StorageDescriptor": {
                        "Columns": [
                            {
```

```
                                "Name": "col1",
                                "Type": "int"
                            },
                            {
                                "Name": "col2",
                                "Type": "string"
                            },
                            {
                                "Name": "col3",
                                "Type": "double"
                            }
                        ],
                        "Compressed": false,
                        "NumberOfBuckets": 0,
                        "SortColumns": [],
                        "StoredAsSubDirectories": false
                    },
                    "TableType": "VIRTUAL_VIEW",
                    "IsRegisteredWithLakeFormation": false,
                    "CatalogId": "123456789012",
                    "IsRowFilteringEnabled": false,
                    "VersionId": "-1",
                    "DatabaseId": "<databaseID>",
                    "ViewDefinition": {
                        "IsProtected": true,
                        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
                        "SubObjects": [
                            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
                        ],
                        "Representations": [
                            {
                                "Dialect": "ATHENA",
                                "DialectVersion": "3",
                                "ViewOriginalText": "SELECT col1, col2 FROM \"gdc-view-
playground-db\".\"table_1\"",
                                "IsStale": false
                            },
                            {
                                "Dialect": "REDSHIFT",
                                "DialectVersion": "1.0",
                                "ViewOriginalText": "SELECT col1 FROM \"gdc-view-
playground-external-schema\".\"table_1\";",
                                "IsStale": false
```

```
                            }
                        ]
                    },
                    "IsMultiDialectView": true
                },
                "ViewValidations": [
                    {
                        "Dialect": "ATHENA",
                        "DialectVersion": "3",
                        "ViewValidationText": "SELECT col1, col2 FROM \"gdc-view-
playground-db\".\"table_1\"",
                        "UpdateTime": "2024-07-11T11:23:19-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "InvalidInputException",
                            "ErrorMessage": "Engine and existing storage descriptor
 mismatch"
                        }
                    },
                    {
                        "Dialect": "REDSHIFT",
                        "DialectVersion": "1.0",
                        "ViewValidationText": "SELECT col1 FROM \"gdc-view-playground-
external-schema\".\"table_1\";",
                        "UpdateTime": "2024-07-11T11:22:49-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "InvalidInputException",
                            "ErrorMessage": "Engine and existing storage descriptor
 mismatch"
                        }
                    }
                ]
            }
        }
    }
}
```

Concesión de permisos del vistas del catálogo de datos

Tras crear vistas en el AWS Glue Data Catalog, puede conceder permisos de lago de datos sobre las vistas a los directores de todas las organizaciones y unidades Cuentas de AWS organizativas. Puede conceder permisos mediante el método de recurso LF-Tags indicado o el método de recurso

indicado. Para obtener más información sobre cómo etiquetar recursos, consulte Control de acceso basado en etiquetas de Lake Formation. Para obtener más información sobre cómo conceder permisos directamente sobre vistas, consulte Concesión de permisos para vistas mediante el método de recursos con nombre.

Vistas materializadas

Temas

- Diferenciar las vistas materializadas de otros tipos de vistas
- Casos de uso
- Conceptos clave
- Permisos para vistas materializadas
- Creación y administración de vistas materializadas
- Almacenamiento y acceso a los datos
- Integrating with AWS Lake Formation permisos
- Monitoreo y depuración
- Administrar los trabajos de actualización
- Supervisión y solución de problemas
- Consideraciones y limitaciones

En el catálogo de AWS Glue datos, una vista materializada es una tabla gestionada que almacena el resultado precalculado de una consulta SQL en formato Apache Iceberg. A diferencia de las vistas del catálogo de datos estándar que ejecutan la consulta cada vez que se accede a ellas, las vistas materializadas almacenan físicamente los resultados de la consulta y los actualizan a medida que cambian las tablas de origen subyacentes. Puede crear vistas materializadas con Apache Spark versión 3.5.6 o superior en Amazon Athena, Amazon EMR o. AWS Glue

Las vistas materializadas hacen referencia a las tablas de Apache Iceberg registradas en el catálogo de AWS Glue datos, y los datos precalculados se almacenan como tablas de Apache Iceberg en los buckets de Amazon S3 Tables o en los buckets de uso general de Amazon S3, lo que permite acceder a ellos desde varios motores de consultas, incluidos Amazon Athena, Amazon Redshift y motores de terceros. Iceberg-compatible

Diferenciar las vistas materializadas de otros tipos de vistas

Las vistas materializadas difieren de las vistas del catálogo de AWS Glue datos, las vistas de Apache Spark y las vistas de Amazon Athena en aspectos fundamentales. Si bien las vistas del catálogo de datos son tablas virtuales que ejecutan la definición de la consulta SQL cada vez que se accede a ellas, las vistas materializadas almacenan físicamente los resultados de las consultas precalculadas. Esto elimina los cálculos redundantes y mejora considerablemente el rendimiento de las consultas en el caso de transformaciones complejas a las que se accede con frecuencia.

Las vistas materializadas también difieren de los procesos de transformación de datos tradicionales creados con AWS Glue ETL o con trabajos personalizados de Spark. En lugar de escribir código personalizado para gestionar la detección de cambios, las actualizaciones incrementales y la organización del flujo de trabajo, las vistas materializadas se definen mediante la sintaxis SQL estándar. El catálogo AWS Glue de datos supervisa automáticamente las tablas de origen, detecta los cambios y actualiza las vistas materializadas mediante una infraestructura informática totalmente gestionada.

Casos de uso

Los siguientes son casos de uso importantes de las vistas materializadas:

- Acelere las consultas analíticas complejas: cree vistas materializadas que precalculen las costosas funciones de unión, agregación y ventana. Los motores Spark reescriben automáticamente las consultas posteriores para utilizar los resultados precalculados, lo que reduce la latencia de las consultas y los costes de cálculo.
- Simplifique los procesos de transformación de datos: sustituya los complejos trabajos de ETL que gestionan la detección de cambios, las actualizaciones incrementales y la organización del flujo de trabajo por definiciones simples SQL-based de vistas materializadas. El catálogo de AWS Glue datos gestiona toda la complejidad operativa de forma automática.
- Habilite el análisis de autoservicio con un acceso a los datos gobernado: cree vistas materializadas seleccionadas que transformen los datos sin procesar en conjuntos de datos listos para la empresa. Conceda a los usuarios acceso a vistas materializadas sin exponer las tablas fuente subyacentes, lo que simplifica la gestión de la seguridad y, al mismo tiempo, potencia el análisis de autoservicio.
- Optimice la ingeniería de funciones para el aprendizaje automático: defina vistas materializadas que implementen transformaciones de funciones para los modelos de aprendizaje automático. La capacidad de actualización automática garantiza que los almacenes de funciones se mantengan

actualizados a medida que evolucionan los datos de origen, mientras que la actualización incremental minimiza los costos de procesamiento.

- Implemente un intercambio de datos eficiente: cree vistas materializadas que filtren y transformen los datos para consumidores específicos. Comparta vistas materializadas entre cuentas y regiones utilizando AWS Lake Formation, eliminando la necesidad de duplicar los datos y, al mismo tiempo, manteniendo un gobierno centralizado.

Conceptos clave

Actualización automática

La actualización automática es una capacidad que supervisa continuamente las tablas de origen y actualiza las vistas materializadas de acuerdo con un cronograma que usted defina. Al crear una vista materializada, puede especificar una frecuencia de actualización mediante una programación basada en el tiempo con intervalos de hasta una hora. El catálogo de AWS Glue datos utiliza la infraestructura informática gestionada de Spark para ejecutar las operaciones de actualización en segundo plano, gestionando de forma transparente todos los aspectos de la detección de cambios y las actualizaciones incrementales.

Cuando los datos de origen cambian entre los intervalos de actualización, la vista materializada queda temporalmente obsoleta. Las consultas que acceden directamente a la vista materializada pueden devolver resultados desactualizados hasta que se complete la siguiente actualización programada. En los escenarios que requieren acceso inmediato a los datos más actuales, puede ejecutar una actualización manual mediante el comando `REFRESH MATERIALIZED VIEW` SQL.

Actualización incremental

La actualización incremental es una técnica de optimización que procesa solo los datos que han cambiado en las tablas de origen desde la última actualización, en lugar de volver a calcular toda la vista materializada. El catálogo de AWS Glue datos aprovecha la capa de metadatos de Apache Iceberg para realizar un seguimiento eficiente de los cambios en las tablas de origen y determinar qué partes de la vista materializada requieren actualizaciones.

Este enfoque reduce significativamente los costos de procesamiento y la duración de la actualización en comparación con las operaciones de actualización completa, especialmente en el caso de conjuntos de datos grandes en los que solo un pequeño porcentaje de los datos cambia entre los ciclos de actualización. El mecanismo de actualización incremental funciona automáticamente; no es necesario escribir una lógica personalizada para detectar o procesar los datos modificados.

Reescritura automática de consultas

La reescritura automática de consultas es una capacidad de optimización de consultas disponible en los motores Spark de Amazon Athena, Amazon EMR y. AWS Glue Cuando ejecutas una consulta con tablas base, el optimizador de Spark analiza tu plan de consultas y determina automáticamente si las vistas materializadas disponibles pueden satisfacer la consulta de manera más eficiente. Si existe una vista materializada adecuada, el optimizador reescribe la consulta de forma transparente para utilizar los resultados precalculados en lugar de procesar las tablas base.

Esta optimización se produce sin necesidad de realizar ningún cambio en el código de la aplicación ni en las instrucciones de consulta. El optimizador Spark garantiza que la reescritura automática de consultas solo se aplique cuando la vista materializada sea actual y pueda producir resultados precisos. Si una vista materializada está obsoleta o no cumple completamente con los requisitos de la consulta, el optimizador ejecuta el plan de consultas original utilizando las tablas base y prioriza la corrección por encima del rendimiento.

Función de definición de vistas

Una vista materializada funciona en función de los permisos del rol de IAM que la creó, conocido como rol de definidor de vistas. La función de definición debe tener acceso de lectura a todas las tablas base a las que se hace referencia en la definición de la vista materializada y debe tener permisos de creación de tablas en la base de datos de destino. Cuando el catálogo de AWS Glue datos actualiza una vista materializada, asume la función de definidor para acceder a las tablas de origen y escribir los resultados actualizados.

Este modelo de seguridad permite conceder a los usuarios acceso a las vistas materializadas sin concederles permisos directos sobre las tablas de origen subyacentes. Si la función de definición de vistas pierde el acceso a alguna tabla base, las operaciones de actualización posteriores fallarán hasta que se restablezcan los permisos.

Permisos para vistas materializadas

Para crear y gestionar vistas materializadas, debe configurar AWS Lake Formation los permisos. El rol de IAM que crea la vista materializada (la función de definición) requiere permisos específicos en las tablas de origen y las bases de datos de destino.

Permisos requeridos para el rol definidor

El rol definidor debe contar con los siguientes permisos de Lake Formation:

- En las tablas de origen: permisos SELECT o ALL, sin filtros de filas, columnas ni celdas.

- En la base de datos de destino: permiso CREATE_TABLE
- En el catálogo de AWS Glue datos GetTable y los permisos de la CreateTable API

Cuando crea una vista materializada, el ARN del rol definidor se almacena en la definición de la vista. El catálogo AWS Glue de datos asume esta función al ejecutar operaciones de actualización automática. Si el rol definidor pierde el acceso a las tablas de origen, las operaciones de actualización fallarán hasta que se restablezcan los permisos.

Permisos de IAM para AWS Glue jobs

La AWS Glue función de IAM de su trabajo requiere los siguientes permisos:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetCatalog",
                "glue:GetCatalogs",
                "glue:GetTable",
                "glue:GetTables",
                "glue:CreateTable",
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "cloudwatch:PutMetricData"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject",
                "s3:PutObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket/*"
            ]
        },
```

```
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogGroup",
                "logs:CreateLogStream",
                "logs:PutLogEvents"
            ],
            "Resource": [
                "arn:aws:logs:*:*:*:/aws-glue/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess"
            ],
            "Resource": "*"
        }
    ]
}
```

El rol que utilice para la actualización automática de Materialized View debe tener el PassRole permiso iam: en el rol.

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "iam:PassRole"
      ],
      "Resource": [
```

```
        "arn:aws:iam::111122223333:role/materialized-view-role-name"
      ]
    }
  ]
}
```

Para permitir que Glue actualice automáticamente la vista materializada, el rol también debe contar con la siguiente política de confianza, que habilita al servicio a asumir dicho rol.

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "glue.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

Si la vista materializada se almacena en buckets de tablas de S3, también debe añadir el siguiente permiso al rol.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3tables:PutTableMaintenanceConfiguration"
      ],
      "Resource": "arn:aws:s3tables:*:123456789012:*"
    }
  ]
}
```

Concesión de acceso a vistas materializadas

Para conceder a otros usuarios acceso a consultar una vista materializada, utilice AWS Lake Formation el permiso SELECT en la tabla de vistas materializadas. Los usuarios pueden consultar la vista materializada sin necesidad de tener acceso directo a las tablas de origen subyacentes.

Para obtener información detallada sobre la configuración de los permisos de Lake Formation, consulte Concesión y revocación de permisos en los recursos del catálogo de datos en la Guía para AWS Lake Formation desarrolladores.

Creación y administración de vistas materializadas

Las vistas materializadas se crean mediante la `CREATE MATERIALIZED VIEW` sentencia SQL en los motores Spark. La definición de la vista especifica la consulta SQL que define la lógica de transformación, la base de datos y el nombre de la tabla de destino, y la configuración de actualización opcional. Puede definir transformaciones complejas, incluidas las agregaciones, las uniones de varias tablas, los filtros y las funciones de ventana.

```
CREATE MATERIALIZED VIEW sales_summary
AS
SELECT
    region,
    product_category,
    SUM(sales_amount) as total_sales,
    COUNT(DISTINCT customer_id) as unique_customers
FROM sales_transactions
WHERE transaction_date >= current_date - interval '90' day
GROUP BY region, product_category;
```

Para configurar la actualización automática, incluya el programa de actualización en la definición de la vista:

```
CREATE MATERIALIZED VIEW sales_summary
SCHEDULE REFRESH EVERY 1 HOUR
AS
SELECT region, product_category, SUM(sales_amount) as total_sales
FROM sales_transactions
GROUP BY region, product_category;
```

Puede refrescar manualmente una vista materializada en cualquier momento mediante el `REFRESH MATERIALIZED VIEW` comando:

```
REFRESH MATERIALIZED VIEW sales_summary;
```

Para modificar el programa de actualización de una vista materializada existente, utilice la `ALTER MATERIALIZED VIEW` siguiente instrucción:

```
ALTER MATERIALIZED VIEW sales_summary
ADD SCHEDULE REFRESH EVERY 2 HOURS;
```

Vistas materializadas anidadas

Puede crear vistas materializadas que hagan referencia a otras vistas materializadas como tablas base, lo que permite la transformación de datos en varias etapas. Al crear vistas materializadas anidadas, el catálogo de AWS Glue datos rastrea las dependencias y propaga automáticamente las actualizaciones por la jerarquía de vistas materializadas. Cuando se actualiza una vista materializada base, todas las vistas materializadas posteriores que dependen de ella se actualizan en consecuencia.

Esta capacidad le permite descomponer las transformaciones complejas en etapas lógicas, lo que mejora la capacidad de mantenimiento y permite la actualización selectiva de las capas de transformación en función de sus requisitos de actualización de los datos.

Almacenamiento y acceso a los datos

Las vistas materializadas almacenan los resultados precalculados como tablas de Apache Iceberg en los cubos de S3 Tables o en los cubos S3 de uso general de su cuenta. AWS El catálogo de AWS Glue datos gestiona todos los aspectos del mantenimiento de las tablas Iceberg, incluidas la compactación y la retención de instantáneas, mediante las capacidades de optimización automatizada de S3 Tables.

Como las vistas materializadas se almacenan como tablas de iceberg, puede leerlas directamente desde cualquier Iceberg-compatible motor, incluidos Amazon Athena, Amazon Redshift y plataformas de análisis de terceros. Esta accesibilidad multimotor garantiza que sus datos precalculados permanezcan accesibles en todo su ecosistema de análisis sin duplicar los datos ni convertir el formato.

Integrating with AWS Lake Formation permisos

Puede utilizarla AWS Lake Formation para gestionar permisos detallados en las vistas materializadas. El creador de la vista se convierte automáticamente en el propietario de la vista materializada y puede conceder permisos a otros usuarios o roles utilizando el método AWS Lake Formation de recurso designado o. LF-Tags

Al conceder un `SELECT` permiso a un usuario para acceder a una vista materializada, este puede consultar los resultados precalculados sin necesidad de acceder a las tablas de origen subyacentes. Este modelo de seguridad simplifica la administración del acceso a los datos y permite implementar el principio del privilegio mínimo, proporcionando a los usuarios acceso únicamente a las transformaciones de datos específicas que necesitan.

Puede compartir vistas materializadas entre AWS cuentas, AWS organizaciones y unidades organizativas mediante las funciones AWS Lake Formation de uso compartido entre cuentas. También puede acceder a las vistas materializadas de todas AWS las regiones mediante enlaces a recursos, lo que permite una gobernanza centralizada de los datos con un acceso distribuido a los datos.

Monitoreo y depuración

El catálogo AWS Glue de datos publica en Amazon CloudWatch todas las operaciones de actualización de vistas materializadas y las métricas asociadas. Puede monitorizar la hora de inicio y finalización de la actualización, la duración, el volumen de datos procesados y el estado de la actualización mediante CloudWatch métricas. Cuando las operaciones de actualización fallan, los mensajes de error y la información de diagnóstico se capturan en CloudWatch los registros.

Puede configurar CloudWatch alarmas para recibir notificaciones cuando los trabajos de actualización superen la duración prevista o fallen repetidamente. El catálogo de AWS Glue datos también publica los eventos de cambio para las ejecuciones de actualización correctas y fallidas, lo que le permite integrar las operaciones de visualización materializada en una automatización más amplia del flujo de trabajo.

Para comprobar el estado actual de una vista materializada, utilice el comando `DESCRIBE MATERIALIZED VIEW` SQL, que devuelve los metadatos, incluido el estado de inactividad, la fecha de la última actualización y la configuración del programa de actualización.

Administrar los trabajos de actualización

Iniciar una actualización manual

Active una actualización inmediata fuera del intervalo programado.

Permiso obligatorio: las AWS credenciales utilizadas para realizar la llamada a la API deben tener `glue:GetTable` permiso para la vista materializada.

Para el catálogo de tablas S3:

```
aws glue start-materialized-view-refresh-task-run \
    --catalog-id <ACCOUNT_ID>:s3tablescatalog/<CATALOG_NAME> \
    --database-name <DATABASE_NAME> \
    --table-name <MV_TABLE_NAME>
```

Para el catálogo raíz:

```
aws glue start-materialized-view-refresh-task-run \
    --catalog-id <ACCOUNT_ID> \
    --database-name <DATABASE_NAME> \
    --table-name <MV_TABLE_NAME>
```

Comprobando el estado de actualización

Obtenga el estado de un trabajo de actualización específico:

```
aws glue get-materialized-view-refresh-task-run \
    --catalog-id <CATALOG_ID> \
    --materialized-view-refresh-task-run-id <TASK_RUN_ID>
```

Historial de actualizaciones del listado

Vea todos los trabajos de actualización para obtener una vista materializada:

```
aws glue list-materialized-view-refresh-task-runs \
    --catalog-id <CATALOG_ID> \
    --database-name <DATABASE_NAME> \
    --table-name <MV_TABLE_NAME>
```

> **Note**
>
> Se utiliza <ACCOUNT_ID>:s3tablescatalog/<CATALOG_NAME> para las tablas S3 o <ACCOUNT_ID> para el catálogo raíz.

Detener una actualización en ejecución

Cancelar un trabajo de actualización en curso:

```
aws glue stop-materialized-view-refresh-task-run \
    --catalog-id <CATALOG_ID> \
    --database-name <DATABASE_NAME> \
    --table-name <MV_TABLE_NAME>
```

Supervisión y solución de problemas

Hay tres formas de supervisar los trabajos de actualización de vistas materializadas:

CloudWatch Métricas

Consulte las métricas agregadas de todos sus trabajos de actualización de vistas materializadas en CloudWatch:

Métricas disponibles:

- AWS/Pegue el espacio de nombres con dimensiones:
  - CatalogId: su identificador de catálogo
  - DatabaseName: Base de datos que contiene la vista materializada
  - TableName: nombre de la vista materializada
  - TaskType: Definido en "» MaterializedViewRefresh

Visualización en la consola:

1. Vaya a la CloudWatch consola → Métricas
2. Seleccione el espacio de AWS nombres /Glue
3. Filtrar por dimensiones: CatalogId,,, DatabaseName TableName TaskType
4. Vea las métricas del éxito, el fracaso y la duración del trabajo

Ejemplo de consulta de CloudWatch métricas:

```
{AWS/Glue,CatalogId,DatabaseName,TableName,TaskType} MaterializedViewRefresh
```

Uso de AWS CLI:

```
aws cloudwatch get-metric-statistics \
    --namespace AWS/Glue \
    --metric-name <MetricName> \
    --dimensions Name=CatalogId,Value=<CATALOG_ID> \
                 Name=DatabaseName,Value=<DATABASE_NAME> \
                 Name=TableName,Value=<TABLE_NAME> \
                 Name=TaskType,Value=MaterializedViewRefresh \
    --start-time <START_TIME> \
    --end-time <END_TIME> \
    --period 3600 \
    --statistics Sum \
    --region <REGION>
```

CloudWatch Registros

Vea los registros de ejecución detallados de las ejecuciones individuales de las tareas de actualización:

Grupo de registros: `/aws-glue/materialized-views/<task_run_id>`

Dónde `<task_run_id>` está un UUID (por ejemplo, abc12345-def6-7890-ghij-klmnopqrstuv).

Visualización de registros:

```
# List log streams for a task run
aws logs describe-log-streams \
    --log-group-name /aws-glue/materialized-views/<TASK_RUN_ID> \
    --region <REGION>

# Get log events
aws logs get-log-events \
    --log-group-name /aws-glue/materialized-views/<TASK_RUN_ID> \
    --log-stream-name <LOG_STREAM_NAME> \
    --region <REGION>
```

En CloudWatch la consola:

1. Vaya a CloudWatch → Registrar grupos
2. Busque /aws- -views/ glue/materialized
3. Seleccione el grupo de registros con su ID de ejecución de la tarea
4. Consulta los registros de ejecución detallados, los errores y el resultado de los trabajos de Spark

Notificaciones

Suscríbete a los eventos para recibir notificaciones en tiempo real sobre los cambios en el estado de los trabajos de actualización:

Tipos de eventos disponibles:

- Se inició la tarea de actualización de Glue Materialized View
- La tarea de actualización de Glue Materialized View se realizó correctamente
- Error en la tarea de actualización de Glue Materialized View
- Error al Auto-Refresh invocar la vista materializada de Glue

Creación de una regla:

```
aws events put-rule \
    --name materialized-view-refresh-notifications \
    --event-pattern '{
        "source": ["aws.glue"],
        "detail-type": [
            "Glue Materialized View Refresh Task Started",
            "Glue Materialized View Refresh Task Succeeded",
            "Glue Materialized View Refresh Task Failed",
            "Glue Materialized View Auto-Refresh Invocation Failure"
        ]
    }' \
    --region <REGION>
```

Añadir un objetivo (por ejemplo, un tema de SNS):

```
aws events put-targets \
    --rule materialized-view-refresh-notifications \
    --targets "Id"="1","Arn"="arn:aws:sns:<REGION>:<ACCOUNT_ID>:<TOPIC_NAME>" \
    --region <REGION>
```

Visualización del estado de actualización

Compruebe el estado de sus trabajos de actualización de vistas materializadas mediante la AWS Glue API:

```
aws glue get-materialized-view-refresh-task-run \
    --catalog-id <CATALOG_ID> \
    --materialized-view-refresh-task-run-id <TASK_RUN_ID> \
    --region <REGION>
```

O bien, enumere todas las ejecuciones de actualización recientes:

```
aws glue list-materialized-view-refresh-task-runs \
    --catalog-id <CATALOG_ID> \
    --database-name <DATABASE_NAME> \
    --table-name <MV_TABLE_NAME> \
    --region <REGION>
```

Esto muestra lo siguiente:

- Hora de la última actualización
- Estado de actualización (CORRECTO, FALLIDO, EN EJECUCIÓN, DETENIDO)
- ID de ejecución de la tarea
- Mensajes de error (si falló)

Estados de actualización comunes:

- EN EJECUCIÓN: el trabajo de actualización se está ejecutando actualmente
- CORRECTO: La actualización se completó correctamente
- ERROR: la actualización ha detectado un error
- DETENIDA: la actualización se canceló manualmente

Solución de problemas de actualizaciones fallidas:

Si se produce un error en la actualización, compruebe lo siguiente:

1. Permisos de IAM: asegúrese de que la función de definición tenga acceso a todas las tablas base y a la ubicación de la vista materializada

2. Disponibilidad de la tabla base: compruebe que todas las tablas a las que se hace referencia existan y estén accesibles
3. Validez de la consulta: confirma que la consulta SQL es válida para el dialecto SQL de Spark
4. Límites de recursos: comprueba si has alcanzado los límites de actualización simultánea de tu cuenta

Usa la GetMaterializedViewRefreshTaskRun API para recuperar mensajes de error detallados.

Consideraciones y limitaciones

- Las vistas materializadas solo pueden hacer referencia a las tablas de Apache Iceberg registradas en el catálogo de AWS Glue datos como tablas base.
- La creación de vistas y la reescritura automática de consultas solo están disponibles en los motores de Spark en Apache Spark versión 3.5.6 y versiones posteriores en Amazon Athena AWS Glue , Amazon EMR y (versión 5.1).
- En última instancia, las vistas materializadas son coherentes con las tablas base. Durante la ventana de actualización, las consultas que acceden directamente a la vista materializada pueden devolver datos desactualizados. Para acceder inmediatamente a los datos actuales, ejecute una actualización manual.
- El intervalo mínimo de actualización automática es de una hora. Para los casos de uso que requieran actualizaciones más frecuentes, ejecute las actualizaciones manuales mediante programación mediante el comando. `REFRESH MATERIALIZED VIEW`
- La reescritura de consultas prioriza la corrección por encima del rendimiento. Si una vista materializada está obsoleta o no puede satisfacer los requisitos de consulta con precisión, los motores de Spark ejecutan la consulta original en las tablas base.

Importación de datos mediante flujos de trabajo en Lake Formation

Con AWS Lake Formation, puede importar sus datos mediante flujos de trabajo. Un flujo de trabajo define el origen de datos y la programación para importar los datos a su lago de datos. Es un contenedor para rastreadores AWS Glue, trabajos y desencadenadores que se utilizan para orquestar los procesos de carga y actualización del lago de datos.

Temas

- Esquemas y flujos de trabajo en Lake Formation

- Creación de un flujo de trabajo
- Ejecución de un flujo de trabajo

Esquemas y flujos de trabajo en Lake Formation

Un flujo de trabajo encapsula una actividad compleja de extracción, transformación y carga (ETL) de múltiples tareas. Los flujos de trabajo generan AWS Glue rastreadores, tareas y activadores para organizar la carga y actualización de los datos. Lake Formation ejecuta y rastrea un flujo de trabajo como una única entidad. Puede configurar un flujo de trabajo para que se ejecute bajo demanda o de forma programada.

> **Note**
>
> El escritor de archivos Parquet de Spark no admite caracteres especiales en los nombres de las columnas. Se trata de una limitación técnica del propio escritor, no de un problema de configuración.

Los flujos de trabajo que cree en Lake Formation son visibles en la consola de AWS Glue como un gráfico acíclico dirigido (DAG). Cada nodo del DAG es una tarea, un rastreador o un disparador. Para supervisar el progreso y solucionar problemas, puede hacer un seguimiento del estado de cada nodo del flujo de trabajo.

Cuando se completa un flujo de trabajo de Lake Formation, el usuario que lo ejecutó recibe el permiso `SELECT` de Lake Formation en las tablas del Catálogo de datos que crea el flujo de trabajo.

También puede crear flujos de trabajo en AWS Glue. Sin embargo, como Lake Formation le permite crear un flujo de trabajo a partir de un esquema, la creación de flujos de trabajo es mucho más sencilla y automatizada en Lake Formation. Lake Formation proporciona los siguientes tipos de esquemas:

- Instantánea de la base de datos. Carga o recarga los datos de todas las tablas en el lago de datos desde una fuente JDBC. Puede excluir algunos datos de la fuente en función de un patrón de exclusión.
- Base de datos incremental. Carga solo los datos nuevos en el lago de datos desde una fuente JDBC, en función de los marcadores establecidos anteriormente. El usuario especifica las tablas individuales de la base de datos de origen de JDBC que desee incluir. Para cada tabla, elige las columnas de marcadores y el orden de clasificación de los marcadores para hacer

un seguimiento de los datos que se han cargado previamente. La primera vez que ejecuta un esquema incremental de base de datos sobre un conjunto de tablas, el flujo de trabajo carga todos los datos de las tablas y establece los marcadores para la siguiente ejecución del esquema incremental de base de datos. Por lo tanto, puede utilizar un esquema de base de datos incremental en lugar del esquema de instantánea de base de datos para cargar todos los datos, siempre que especifique cada tabla de los orígenes de datos como parámetro.

- Archivo de registro: carga datos de forma masiva desde fuentes de archivos de registro AWS CloudTrail, incluidos los registros de Elastic Load Balancing y los registros de Application Load Balancer.

Utilice la siguiente tabla como ayuda para decidir si debe utilizar una instantánea de base de datos o un esquema incremental de base de datos.

Utilice la instantánea de la base de datos cuando...	Utilice la base de datos incremental cuando...
• La evolución del esquema es flexible. (Se cambia el nombre de las columnas, se eliminan las columnas anteriores y se añaden nuevas columnas en su lugar). • Se necesita una coherencia total entre el origen y el destino.	• La evolución del esquema es incremental. (Solo hay adición sucesiva de columnas). • Solo se añaden nuevas filas; las filas anteriores no se actualizan.

> **Note**
>
> Los usuarios no pueden editar los esquemas y flujos de trabajo creados por Lake Formation.

Creación de un flujo de trabajo

Antes de empezar, asegúrese de que ha concedido los permisos de datos y de ubicación de datos necesarios al rol `LakeFormationWorkflowRole`. Esto es para que el flujo de trabajo pueda crear tablas de metadatos en el Catálogo de datos y escribir datos en ubicaciones de destino en Amazon S3. Para obtener más información, consulte (Opcional) Crear un rol de IAM para flujos de trabajo y Descripción general de los permisos de Lake Formation.

> **Note**
>
> Lake Formation usa las operaciones `GetTemplateInstance`, `GetTemplateInstances` e `InstantiateTemplate` para crear flujos de trabajo a partir de esquemas. Estas operaciones no son de ámbito público, y solo se utilizan internamente para crear recursos en su nombre. Recibe CloudTrail eventos para crear flujos de trabajo.

Para crear un flujo de trabajo a partir de un esquema

1. Abre la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/. Inicie sesión como administrador del lago de datos o como usuario con permisos de ingeniero de datos. Para obtener más información, consulte Personas de Lake Formation y referencia de permisos IAM.
2. En el panel de navegación, seleccione Esquemas y, a continuación, seleccione Utilizar esquema.
3. En la página Usar un esquema, elija un mosaico para seleccionar el tipo de esquema.
4. En Origen de importación, especifique el origen de datos.

   Si está importando desde un origen JDBC, especifique lo siguiente:

   - Conexión a la base de datos. Elija una conexión de la lista. Cree conexiones adicionales utilizando la consola de AWS Glue. El nombre de usuario JDBC y la contraseña de la conexión determinan los objetos de la base de datos a los que tiene acceso el flujo de trabajo.
   - Ruta de datos de origen: introduzca*<database>*/*<schema>*/*<table>*o*<database>*/*<table>*, según el producto de la base de datos. Oracle Database y MySQL no permiten utilizar un esquema en la ruta. Puede sustituir el carácter de porcentaje (%) por *<schema>* o*<table>*. Por ejemplo, para una base de datos Oracle con un identificador del sistema (SID) de `orcl`, introduzca `orcl/%` para importar todas las tablas a las que tenga acceso el usuario nombrado en la conexión.

     > **Important**
     >
     > Este campo distingue entre mayúsculas y minúsculas. El flujo de trabajo fallará si no coincide entre mayúsculas y minúsculas en alguno de los componentes.

Si especifica una base de datos MySQL, AWS Glue ETL utiliza el controlador JDBC Mysql5 de forma predeterminada, por lo que MySQL8 no es compatible de forma nativa. Puede editar el script del trabajo ETL para utilizar un parámetro `customJdbcDriverS3Path` como el descrito en Valores de tipo de conexión JDBC en la Guía para desarrolladores de AWS Glue para utilizar un controlador JDBC diferente que sea compatible con MySQL8.

Si está importando desde un archivo de registro, asegúrese de que el rol que especifique para el flujo de trabajo (el "rol del flujo de trabajo") tenga los permisos IAM necesarios para acceder a los orígenes de datos. Por ejemplo, para importar AWS CloudTrail registros, el usuario debe tener los `cloudtrail:LookupEvents` permisos `cloudtrail:DescribeTrails` y para ver la lista de CloudTrail registros al crear el flujo de trabajo, y el rol del flujo de trabajo debe tener permisos en la CloudTrail ubicación de Amazon S3.

5. Realice una de las siguientes acciones:

   - Para el tipo de esquema Instantánea de base de datos, identifique de forma opcional un subconjunto de datos a importar especificando uno o varios patrones de exclusión. Estos patrones de exclusión son Unix-style `glob` patrones. Se almacenan como una propiedad de las tablas creadas por el flujo de trabajo.

     Para obtener más información sobre los patrones de exclusión disponibles, consulte Incluir y excluir patrones en la Guía para desarrolladores de AWS Glue .

   - Para el tipo de esquema de base de datos incremental, especifique los siguientes campos. Agregue una fila para cada tabla que desee importar.

     Nombre de la tabla

     Tabla que se va a importar. Debe estar en minúscula.

     Claves de marcadores

     Comma-delimited lista de nombres de columnas que definen las claves de los marcadores. Si está en blanco, la clave principal se utiliza para determinar los datos nuevos. Las mayúsculas y minúsculas de cada columna deben coincidir con las definidas en los orígenes de datos.

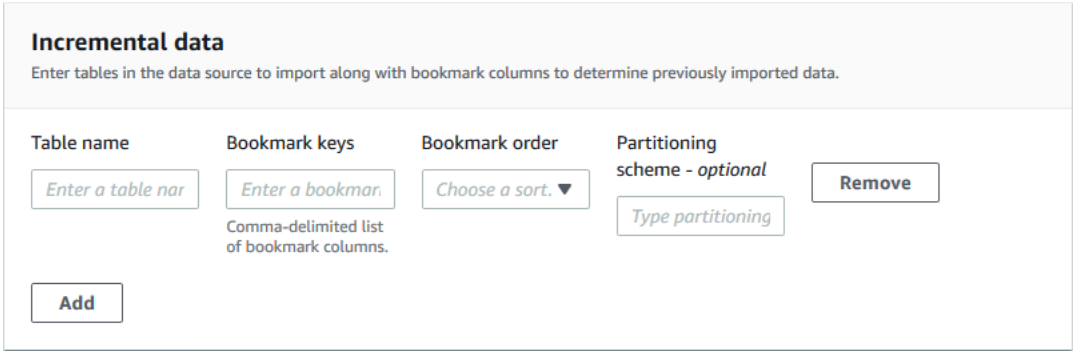
> **Note**
>
> La clave principal se califica como clave marcadora predeterminada solo si es secuencialmente creciente o decreciente (sin huecos). Si desea utilizar la clave principal como clave de marcador y tiene huecos, debe nombrar la columna de clave principal como clave de marcador.

Orden de los marcadores

Si elige Ascendente, las filas con valores superiores a los marcados se identifican como nuevas filas. Si elige Descendente, las filas con valores inferiores a los marcados se identifican como nuevas filas.

Esquema de partición

(Opcional) Lista de columnas de claves de partición, delimitadas por barras (/). Ejemplo: `year/month/day`.



Para más información, consulte Seguimiento de los datos procesados mediante marcadores de trabajo en la Guía para desarrolladores de AWS Glue .

6. En Importar destino, especifique la base de datos de destino, la ubicación de Amazon S3 de destino y el formato de datos.

   Asegúrese de que el rol de flujo de trabajo tenga los permisos de Lake Formation necesarios en la base de datos y en la ubicación de destino de Amazon S3.

   > **Note**
   >
   > Actualmente, los esquemas no admiten el cifrado de datos en el destino.

7. Elija una frecuencia de importación.

   Puede especificar una expresión `cron` con la opción Personalizada.

8. En Opciones de importación:

   a. Introduzca un nombre de flujo de trabajo.

   b. Para el rol, elija el rol `LakeFormationWorkflowRole`, que creó en (Opcional) Crear un rol de IAM para flujos de trabajo.

   c. Si lo desea, especifique un prefijo de tabla. El prefijo se antepone a los nombres de las tablas del Catálogo de datos que crea el flujo de trabajo.

9. Seleccione Crear y espere a que la consola informe de que el flujo de trabajo se ha creado correctamente.

> **Tip**
>
> ¿Ha recibido este mensaje de error?
> `User: arn:aws:iam::<account-id>:user/<username> is not authorized to perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/<rolename>...`
> Si es así, compruebe que lo ha sustituido por *<account-id>* un número de AWS cuenta válido en todas las políticas.

> **Véase también:**
>
> - Esquemas y flujos de trabajo en Lake Formation

Ejecución de un flujo de trabajo

Puede ejecutar un flujo de trabajo desde la consola de Lake Formation, la consola de AWS Glue, la interfaz de línea de comandos de AWS Glue (AWS CLI) o la API.

Para ejecutar un flujo de trabajo (consola de Lake Formation)

1. Abre la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/. Inicie sesión como administrador del lago de datos o como usuario con permisos de ingeniero de

datos. Para obtener más información, consulte Personas de Lake Formation y referencia de permisos IAM.

2. En el panel de navegación, elija Esquemas.
3. En la página Esquemas, seleccione el flujo de trabajo. Después, en el menú Acciones, seleccione Comenzar.
4. Conforme se ejecuta el flujo de trabajo, puede ver su progreso en la columna Estado de la última ejecución. Pulse el botón de actualización de vez en cuando.

   El estado pasa de EN EJECUCIÓN a Detectando, Importando y FINALIZADO.

   Cuando finalice el flujo de trabajo:

   - El Catálogo de datos tendrá nuevas tablas de metadatos.
   - Sus datos se incorporan al lago de datos.

   Si el flujo de trabajo falla, haga lo siguiente:

   a. Seleccione el flujo de trabajo. Elija Acciones y, a continuación, elija Ver gráfico.

      El flujo de trabajo se abre en la consola de AWS Glue.
   b. Asegúrese de que se seleccione el flujo de trabajo y elija la pestaña Historial.
   c. En Historial, seleccione la ejecución más reciente y seleccione Ver detalles de la ejecución.
   d. Seleccione un trabajo o un rastreador fallidos en el gráfico dinámico (tiempo de ejecución) y revise el mensaje de error. Los nodos con errores aparecen en rojo o amarillo.

Véase también:

- Esquemas y flujos de trabajo en Lake Formation

Llevar sus datos al AWS Glue Data Catalog

Puede crear catálogos federados en el AWS Glue Data Catalog (catálogo de datos) y unificar los datos entre los lagos de datos de Amazon S3 y los almacenes de datos de Amazon Redshift. También puede integrar datos de sus bases de datos operativas, por ejemplo Amazon DynamoDB, y de fuentes de datos de terceros, como PostgreSQL BigQuery, Google o MySQL, entre otras. El Catálogo de datos proporciona un repositorio de metadatos centralizado que facilita la administración y la detección de datos en sistemas dispares.

El Catálogo de datos se integra con más de 30 orígenes de datos externos a través de conectores federados. Con esta integración, puede consultar datos de estas fuentes externas sin tener que crear canalizaciones de datos en las que ingerir primero los datos. AWS

Tras catalogar los datos externos, puede utilizarlos para gestionar de forma centralizada AWS Lake Formation los permisos de acceso a los datos en el catálogo de datos. Los administradores de lagos de datos pueden conceder permisos de acceso detallados a otras entidades principales de IAM (usuarios o roles) dentro de la misma cuenta o entre cuentas mediante los métodos de control de acceso basado en etiquetas (etiquetas LF) y recursos con nombre.

Al usar etiquetas LF, los administradores de datos pueden organizar los recursos de forma lógica en función de atributos como el dominio y el nivel de confidencialidad, lo que simplifica la administración de permisos y garantiza controles de acceso consistentes en todos los servicios de análisis y aprendizaje automático, como Athena, Amazon EMR o Redshift Spectrum. AWS Glue

El Catálogo de datos proporciona los siguientes métodos para administrar datos y permisos en conjuntos de datos externos y metaalmacenes externos:

- Integre los depósitos de Amazon S3 Table con Data Catalog: puede publicar y catalogar las tablas de Amazon S3 como objetos del catálogo de datos y registrar el catálogo como una ubicación de datos de Lake Formation desde la consola de Lake Formation o mediante operaciones de AWS Glue API.
- Incorpore los datos de los almacenes de datos de Amazon Redshift al AWS Glue Data Catalog: registre un espacio de nombres o un clúster de Amazon Redshift existente en el catálogo de datos y cree un catálogo federado de varios niveles en el catálogo de datos.

  Puede acceder a sus datos mediante cualquier motor de consultas compatible con la especificación OpenAPI de catálogo REST de Apache Iceberg, como Amazon EMR sin servidor y Amazon Athena.

- Federe Iceberg REST los catálogos remotos al catálogo de datos: federe los Iceberg REST catálogos remotos al catálogo de datos y acceda de forma segura a Iceberg las tablas remotas almacenadas en Amazon S3 mediante motores de análisis. AWS
- Conéctese al catálogo de datos desde fuentes de datos externas: conecte el catálogo de datos a fuentes de datos externas mediante AWS Glue conexiones y cree catálogos federados para administrar de forma centralizada los permisos de acceso a los conjuntos de datos mediante Lake Formation. No es necesario migrar los metadatos al Catálogo de datos.
- Crear catálogos para administrar las tablas de Amazon Redshift en el Catálogo de datos: es posible que no tenga un clúster productor de Amazon Redshift o un recurso compartido de datos de Amazon Redshift disponibles en la actualidad, pero desee crear y administrar las tablas de Amazon Redshift mediante el Catálogo de datos. Para empezar, puede crear un catálogo de AWS Glue administrado mediante la operación de la API `glue:CreateCatalog` o la consola de AWS Lake Formation , configurando el tipo de catálogo como `Managed` y `Catalog source` como Redshift.
- Publicar los recursos compartidos de datos de Amazon Redshift con el Catálogo de datos: publique los recursos compartidos de Amazon Redshift en el Catálogo de datos y utilice Lake Formation para administrar de forma centralizada el acceso a los datos de los recursos compartidos y restringir el acceso de los usuarios.

  Puede consultar los datos mediante Amazon Redshift Spectrum.
- Conectar el Catálogo de datos a metaalmacenes de Hive externos: conecte el Catálogo de datos a metaalmacenes externos para administrar los permisos de acceso en conjuntos de datos en Amazon S3 mediante Lake Formation. No es necesario migrar los metadatos al Catálogo de datos.
- Integre Lake Formation con AWS Data Exchange: Lake Formation admite la concesión de licencias de acceso a sus datos mediante AWS Data Exchange. Si desea conceder licencias para sus datos de Lake Formation, consulte Qué es AWS Data Exchange en la Guía del usuario de AWS Data Exchange .

Temas

- Federación de catálogos a catálogos remotos de Iceberg
- Incorporar los datos de Amazon Redshift al AWS Glue Data Catalog
- Federación en fuentes de datos externas en el AWS Glue Data Catalog
- Integración de Amazon S3 Tables con AWS Glue Data Catalog and AWS Lake Formation
- Cambiar los controles de acceso para la integración de S3 Tables

- Creación de un catálogo gestionado de Amazon Redshift en AWS Glue Data Catalog
- Administración de permisos para los datos de un recurso compartido de datos de Amazon Redshift
- Administración de los permisos de los conjuntos de datos que utilizan metaalmacenes externos

Federación de catálogos a catálogos remotos de Iceberg

La federación de catálogos AWS Glue proporciona acceso directo y seguro a las tablas Iceberg, almacenadas en Amazon S3 y catalogadas en catálogos remotos, mediante AWS motores de análisis. La federación de catálogos sincroniza los metadatos entre el catálogo de datos y los catálogos remotos al acceder a las tablas remotas. Es compatible con una amplia variedad de motores de análisis, incluidos Amazon Redshift, Amazon EMR, Amazon Athena y motores de terceros como Apache Spark AWS Glue, entre otros.

La federación de catálogos se utiliza AWS Glue Data Catalog para comunicarse con sistemas de catálogos remotos para descubrir tablas y con Lake Formation para autorizar el acceso a los datos de las tablas en Amazon S3. Al consultar una tabla federada, Data Catalog descubre la información más reciente de la tabla en el catálogo remoto en el momento de la consulta y obtiene la ubicación de la tabla en Amazon S3, el esquema actual y la información de partición. A continuación, su motor de análisis (Amazon Athena, Amazon Redshift, Amazon EMR) utiliza esta información para acceder a los archivos de datos de Iceberg directamente desde Amazon S3. Lake Formation administra el acceso a las tablas mediante el envío de credenciales específicas a los datos de las tablas almacenados en Amazon S3, lo que permite a los motores aplicar permisos detallados a las tablas federadas.

Características de Catalog Federation

Gobernado con Lake Formation

Los catálogos de iceberg federados de Data Catalog son recursos registrados de Lake Formation, lo que le permite conceder permisos detallados a nivel de fila, columna y celda a las tablas de iceberg en los catálogos federados de Iceberg mediante las subvenciones de Lake Formation. Los catálogos federados de Iceberg y los objetos asociados se pueden compartir de forma segura entre cuentas. AWS Los catálogos de Federated Iceberg también funcionan con el control de Tag-based acceso de Lake Formation, lo que le permite escalar la gobernanza mediante etiquetas.

Configuraciones de red

La federación de catálogos admite conexiones directas a fuentes de catálogo remotas mediante la conectividad HTTPS estándar. También admite la conectividad a través de Amazon VPC si desea mantener el aislamiento de la red y la conectividad mediante el soporte de proxy cuando desea una comunicación segura a través de los firewalls de la organización.

Limitaciones

- Límite de tamaño de los metadatos de la tabla Iceberg: en el caso de los catálogos federados de Iceberg, AWS Glue Data Catalog admite un tamaño máximo de metadatos de 5 MB por llamada a la API REST. AWS Glue Data Catalog rechaza las solicitudes a tablas con metadatos que superen este límite.

Temas

- Fedérate para Snowflake Catálogo Iceberg
- Federate a Databricks Unity Catalog

Fedérate para Snowflake Catálogo Iceberg

AWS Glue Data Catalog se federa para Snowflake usar las credenciales de OAuth2 de un principal de servicio. Snowflake Puede usar esta federación para conectarse a los catálogos y conectarse a Snowflake Horizon ellos. Snowflake Polaris Este mecanismo de autenticación permite a Data Catalog acceder a los metadatos de varios objetos (como catálogos, bases de datos y tablas) de su catálogo, en función de los privilegios asociados al principal de servicio. Para garantizar el acceso a los objetos correctos, es esencial conceder al director de servicio los permisos necesarios para leer los metadatos de estos objetos. Snowflake

Requisitos previos

Antes de crear un catálogo federado en Data Catalog que esté gobernado por Lake Formation, asegúrese de tener los siguientes permisos:

Su principal de IAM (usuario o rol) debe tener los siguientes permisos:

- Permisos de formación de lagos —, `lakeformation:RegisterResource` `lakeformation:DescribeResource`

- AWS Glue permisos —`glue:CreateConnection`,`glue:CreateCatalog`,`glue:GetConnection`, `glue:PassConnection`
- Permisos de Secrets Manager —`secretsmanager:CreateSecret`, `secretsmanager:GetSecretValue`
- Permisos de IAM —`iam:CreateRole`,, `iam:AttachRolePolicy` `iam:PassRole`

Debe ser administrador de un lago de datos de Lake Formation o tener `CREATE_CATALOG` permiso para acceder al catálogo de datos

Cree un catálogo federado

Uso de la consola

1. Inicie sesión en la consola y abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. Elige la AWS región que prefieras en el selector de regiones.
3. En el panel de navegación izquierdo, elija Catálogos.
4. Seleccione Crear catálogo para abrir el flujo de trabajo de creación de catálogo.
5. En el paso Elegir fuente de datos, seleccione una Snowflake de las opciones disponibles.
6. En la etapa Establecer detalles del catálogo, proporciona tres datos: detalles del catálogo, detalles de conexión y detalles de registro.
7. En el contenedor de detalles del catálogo, proporcione un nombre único a su catálogo AWS Glue federado e introduzca el nombre del Snowflake catálogo existente.
8. En el contenedor de detalles de conexiones, puede elegir entre una conexión existente a la que tenga acceso o proporcionar la configuración para crear un conector nuevo.
9. Las nuevas configuraciones de conexión incluyen:
    - Nombre de conexión: nombre único del objeto de AWS Glue conexión.
    - URL de instancia: la URL de punto final de tu Snowflake cuenta actual.
    - Autenticación: especifique la configuración de autenticación que se AWS Glue utiliza para conectarse al servidor de catálogo remoto. AWS Glue admite tanto la autenticación OAuth2 como la personalizada.
    - URL del token: especifique la URL del proveedor de identidad del catálogo remoto.

- ID de cliente de OAuth2: (opcional) especifique el ID de cliente de la credencial de OAuth2 asociada a su catálogo remoto. Si no proporcionas un ID de cliente, la solicitud del token se autentica utilizando únicamente el secreto de cliente almacenado en él. AWS Secrets Manager
- Secreto: almacena y usa el secreto del cliente de OAuth2 usando AWS Secrets Manager o ingresando el valor secreto en el cuadro de texto. Cuando escribes el secreto manualmente en la consola, AWS Glue crea el secreto en tu nombre.
- Ámbito de la URL del token: especifique el ámbito de autenticación de OAuth.
- Filtro de carátulas de catálogo: elija si desea llevar objetos en minúscula o en mayúscula de su catálogo remoto a Data Catalog.

10. Cree un rol de IAM que los directores de servicio de Lake Formation AWS Glue y los directores de servicio de Lake Formation puedan usar para acceder a ubicaciones secretas en Amazon S3 de tablas AWS Secrets Manager remotas de Iceberg, respectivamente. Seleccione el rol de IAM en el menú desplegable de registro. Consulte los pasos 2 y 3 de la siguiente sección de la CLI para obtener más información sobre la política de IAM.
11. Seleccione Probar conexión para comprobar si las propiedades de conexión y el acceso al rol de IAM están configurados correctamente.
12. Selecciona Siguiente para revisar la configuración.
13. Selecciona Crear catálogo en la página de revisión.

Uso de la CLI de

1. Crea un AWS Secrets Manager secreto

   El AWS Glue conector admite dos tipos de autenticación: OAuth2 y personalizada. Cuando utilice la opción OAuth2, utilícela AWS Secrets Manager para almacenar el secreto del cliente del principal de servicio. Snowflake Más adelante, utilizará este secreto al crear la AWS Glue conexión. Para la autenticación personalizada, AWS Secrets Manager utilícela para almacenar y recuperar el token de acceso.

   En el siguiente ejemplo, sustituya*your-snowflake-secret*,*client_secret*, *region* por su propia información.

```
aws secretsmanager create-secret \
--name your-snowflake-secret \
--description "Snowflake secret" \
```

```
--secret-string '{
"USER_MANAGED_CLIENT_APPLICATION_CLIENT_SECRET": "client_secret"
}' \
--region region
```

> **Note**
>
> `USER_MANAGED_CLIENT_APPLICATION_CLIENT_SECRET`es una palabra clave reservada que se AWS Glue utiliza para hacer referencia a un valor secreto del cliente incluido en el secreto. Usa también la misma palabra clave cuando crees el secreto en la consola de Lake Formation.

2. Cree un rol de IAM que dé acceso al objeto de AWS Glue conexión al secreto creado en el paso anterior

   El objeto de AWS Glue conexión requiere acceso al AWS Secrets Manager secreto cuando lo utilizas AWS Secrets Manager para almacenar, recuperar y actualizar tu token secreto de OAuth. El objeto de AWS Glue conexión también requiere acceso para crear, describir y usar las interfaces de red de Amazon VPC cuando utilizas un punto de enlace de Amazon VPC para restringir la conectividad a tu cuenta. Snowflake

   Cree una política de IAM y adjúntela a un rol de IAM. Agregue el principal AWS Glue de servicio a la política de confianza.

   En el siguiente ejemplo *your-secrets-manager-ARNyour-vpc-id*, sustituya y *your-subnet-id1* por su propia información.

   Example Política de IAM

```
{
    "Version": "2012-10-17",
    "Statement": [{
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetSecretValue",
                "secretsmanager:DescribeSecret",
                "secretsmanager:PutSecretValue"
            ],
            "Resource": [
                "your-secrets-manager-ARN"
```

```
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:CreateNetworkInterface",
                "ec2:DeleteNetworkInterface",
                "ec2:DescribeNetworkInterfaces"
            ],
            "Resource": "*",
            "Condition": {
                "ArnEquals": {
                    "ec2:Vpc": "arn:aws:ec2:region:account-id:vpc/your-vpc-id",
                    "ec2:Subnet": ["arn:aws:ec2:region:account-id:subnet/your-
subnet-id1"]
                }
            }
        }
    ]
}
```

Example Política de confianza

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Effect": "Allow",
        "Principal": {
            "Service": "glue.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
    }]
}
```

3. Cree una política de IAM que otorgue a Lake Formation acceso de lectura a la ubicación del catálogo en Amazon S3

   Como propietario de un catálogo federado en Data Catalog, utiliza Lake Formation para conceder a sus equipos de datos acceso detallado a tablas, acceso detallado (a nivel de columnas, filas y celdas) y acceso basado en etiquetas. Lake Formation utiliza una función de IAM que le da acceso a las ubicaciones subyacentes de Amazon S3 de sus tablas Iceberg

remotas. Este acceso permite a Lake Formation vender credenciales de acceso por ámbito específico a los motores de análisis que consultan tablas remotas.

Cree una política de IAM y asígnela a una función de IAM. Agregue al director de servicio de Lake Formation a la política de confianza de roles.

En el siguiente ejemplo, reemplace *amzn-s3-demo-bucketN* y *your-kms-key* por su propia información.

Example Política de IAM

```
{
    "Version": "2012-10-17",
    "Statement": [{
            "Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket1/*",
                "arn:aws:s3:::amzn-s3-demo-bucket2/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket1",
                "arn:aws:s3:::amzn-s3-demo-bucket2"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:Encrypt"
            ],
            "Resource": [
                "your-kms-key"
            ]
        }
```

```
    ]
}
```

Example Política de confianza

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "",
        "Effect": "Allow",
        "Principal": {
            "Service": "lakeformation.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
    }]
}
```

> **Note**
>
> Cuando utiliza la consola de Lake Formation para crear un catálogo federado, la consola utiliza un único rol de IAM con ambas políticas asociadas para completar la configuración.

4. Cree un objeto de conexión AWS Glue

AWS Glue el conector admite los métodos de autenticación OAuth2 y personalizados. El objeto de conexión del catálogo de datos admite una `CATALOG_CASING_FILTER` configuración que permite llevar objetos en minúscula o mayúscula del catálogo remoto a Data Catalog.

El siguiente ejemplo usa la configuración de autenticación OAuth2 para crear una conexión. AWS Glue Modifique el *`highlighted sections`* añadiendo su información.

```
aws glue create-connection \
    --connection-input '{
"Name": "your-glue-connection-to-snowflake-account",
"ConnectionType": "SNOWFLAKEICEBERGRESTCATALOG",
"ConnectionProperties": {
    "INSTANCE_URL": "your-snowflake-account-URL",
    "ROLE_ARN": "your-IAM-role-for-secrets-and-VPC-access",
    "CATALOG_CASING_FILTER": "LOWERCASE_ONLY"
},
```

```
"AuthenticationConfiguration": {
    "AuthenticationType": "OAUTH2",
    "OAuth2Properties": {
        "OAuth2GrantType": "CLIENT_CREDENTIALS",
        "TokenUrl": "your-internal-or-external-token-server-url",
        "OAuth2ClientApplication": {
            "UserManagedClientApplicationClientId": "our-client-id"
        },
        "TokenUrlParametersMap": {
            "scope": "all-apis"
        }
    },
    "SecretArn": "arn:aws:secretsmanager:your-aws-region:your-aws-account-
id:secret:snowflake-secret"
}
}'
```

> **Note**
>
> El campo `OAuth2ClientApplication` es opcional. Si la omites, la solicitud del token se autentica utilizando únicamente el secreto de cliente almacenado en ella AWS Secrets Manager, sin enviar un identificador de cliente.

5. Registre AWS Glue la conexión como un recurso de Lake Formation

   Con el objeto de AWS Glue conexión (creado en el paso 4) y el rol de IAM (creado en el paso 3), ahora puede registrar el objeto de AWS Glue conexión como un recurso administrado por Lake Formation.

   Reemplace el *your-glue-connector-arn* y el *your-IAM-role-ARN-having-LF-access* con su información.

   ```
   aws lakeformation register-resource \
       --resource-arn your-glue-connector-arn \
       --role-arn your-IAM-role-ARN-having-LF-access \
       --with-federation \
       --with-privileged-access
   ```

6. Cree un catálogo federado en Data Catalog

Después de crear un objeto de AWS Glue conexión y registrarlo en Lake Formation, puede crear un catálogo federado en el catálogo de datos:

Proporcione al catálogo federado un nombre único en*your-federated-catalog-name*, haga referencia al catálogo en Snowflake at *catalog-name-in-Snowflake* e introduzca el nombre de la conexión creado anteriormente en. *your-glue-connection-name*

```
aws glue create-catalog \
    --name your-federated-catalog-name \
    --catalog-input '{
    "FederatedCatalog": {
        "Identifier": catalog-name-in-Snowflake",
        "ConnectionName": your-glue-connection-name"
    },
    "CreateTableDefaultPermissions": [],
    "CreateDatabaseDefaultPermissions": []
}'
```

Consideraciones a la hora de integrar con Snowflake

- Cuando depositas recursos (como bases de datos y tablas)Snowflake, Lake Formation no revoca automáticamente los permisos concedidos a ese recurso federado. Para eliminar los permisos de acceso, debe revocar explícitamente los permisos que se otorgaron anteriormente al recurso federado mediante Lake Formation.
- Al montar un catálogo remoto con la `CATALOG_CASING_FILTER='UPPERCASE_ONLY'` configuración, las bases de datos y tablas con identificadores en mayúsculas se federan, pero los objetos con identificadores en minúscula no.
- Puede reutilizar la misma AWS Glue conexión para crear varios catálogos federados. Al eliminar un catálogo, no se eliminará la conexión asociada. Para eliminar una conexión, utilice el `aws glue delete-connection` comando AWS CLI y asegúrese de eliminar primero todos los catálogos asociados.
- No se admiten los espacios de nombres anidados en el Polaris catálogo. Es decir, la federación de catálogos puede acceder a Iceberg tablas remotas que siguen una notación de 3 partes. `catalog.database.table`

Federate a Databricks Unity Catalog

AWS Glue Data Catalog se federa para Databricks usar las credenciales OAuth2 de un director de servicio. Databricks Este mecanismo de autenticación permite acceder AWS Glue Data Catalog a los metadatos de varios objetos (como catálogos, bases de datos y tablas) en Databricks Unity Catalog función de los privilegios asociados al principal del servicio. Para garantizar el acceso a los objetos correctos, es esencial conceder al responsable del servicio los permisos necesarios Databricks para leer los metadatos de estos objetos.

A continuación, la federación de catálogos permite descubrir y consultar Iceberg tablas en su catálogo de Databricks Unity. Para leer las tablas delta, asegúrese de que Iceberg los metadatos estén disponibles para utilizarlas. Uniform Siga el Databricks tutorial y la documentación para crear el servicio principal y los privilegios asociados en su Databricks espacio de trabajo.

Requisitos previos

Antes de crear un catálogo federado en Data Catalog que esté gobernado por Lake Formation, asegúrese de tener los siguientes permisos:

Su director de IAM (usuario o rol) debe tener los siguientes permisos:

- Permisos de Lake Formation —`lakeformation:RegisterResource`, `lakeformation:DescribeResource`
- AWS Glue permisos —`glue:CreateConnection`,`glue:CreateCatalog`,`glue:GetConnection`, `glue:PassConnection`
- Permisos de Secrets Manager —`secretsmanager:CreateSecret`, `secretsmanager:GetSecretValue`
- Permisos de IAM —`iam:CreateRole`,, `iam:AttachRolePolicy` `iam:PassRole`

Debe ser administrador de un lago de datos de Lake Formation o tener `CREATE_CATALOG` permiso para acceder al catálogo de datos.

Cree un catálogo federado

Uso de la consola

1. Inicie sesión en la consola y abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

2. Elija la AWS región que prefiera en el selector de regiones.
3. En el panel de navegación izquierdo, elija Catálogos.
4. Seleccione Crear catálogo para abrir el flujo de trabajo de creación de catálogo.
5. En el paso Elegir fuente de datos, seleccione una Databricks de las opciones disponibles.
6. En el paso Definir los detalles del catálogo, debe proporcionar tres datos: detalles del catálogo, detalles de conexión y detalles de registro.
7. En el contenedor de detalles del catálogo, proporcione un nombre único al catálogo AWS Glue federado e introduzca el nombre del Databricks catálogo existente.
8. En el contenedor de detalles de conexiones, puede elegir entre una conexión existente a la que tenga acceso o proporcionar la configuración necesaria para crear un conector nuevo.
9. Las nuevas configuraciones de conexión incluyen:

    - Nombre de conexión: nombre exclusivo del objeto de AWS Glue conexión.
    - URL del espacio de trabajo: la URL de punto final de tu espacio de Databricks trabajo existente.
    - Autenticación: especifique la configuración de autenticación que se AWS Glue utiliza para conectarse al servidor de catálogo remoto. AWS Glue admite tanto la autenticación OAuth2 como la personalizada.
    - URL del token: especifique la URL del proveedor de identidad del catálogo remoto.
    - ID de cliente de OAuth2: especifique el ID de cliente de la credencial de OAuth2 asociada a su catálogo remoto.
    - Secreto: almacene y utilice el secreto del cliente de OAuth2 utilizando AWS Secrets Manager o introduzca el valor secreto en el cuadro de texto. Al introducir el secreto manualmente en la consola, AWS Glue crea el secreto en tu nombre.
    - Ámbito de la URL del token: especifique el ámbito de autenticación de OAuth.
10. Cree una función de IAM que AWS Glue los directores del servicio de Lake Formation puedan utilizar para acceder a las ubicaciones secretas y Amazon S3 de Iceberg las tablas remotas, respectivamente. Seleccione la función de IAM en el menú desplegable de registro. Consulte los pasos 2 y 3 de la siguiente sección de CLI para obtener detalles sobre la política de IAM.
11. Seleccione Probar conexión para comprobar si las propiedades de conexión y el acceso a los roles de IAM están configurados correctamente. La funcionalidad de conexión de prueba no está disponible cuando se conecta Databricks mediante Amazon VPC.
12. Seleccione Siguiente para revisar la configuración.

13. Selecciona Crear catálogo en la página de revisión.

Uso de la CLI de

1. Crea un AWS Secrets Manager secreto

   El AWS Glue conector admite dos tipos de autenticación: OAuth2 y Custom. Cuando se utiliza la opción OAuth2, se utiliza AWS Secrets Manager para almacenar el secreto del cliente principal del servicio. Databricks Utilizará este secreto más adelante al crear la AWS Glue conexión. Para la autenticación personalizada, AWS Secrets Manager utilícela para almacenar y recuperar el token de acceso.

   En el siguiente ejemplo, sustituya*<databricks-secret>*,*<client_secret>*, *<region>* por su propia información.

   ```
   aws secretsmanager create-secret \
   --name <databricks-secret> \
   --description "Databricks secret" \
   --secret-string '{
   "USER_MANAGED_CLIENT_APPLICATION_CLIENT_SECRET": "<client_secret>"
   }' \
   --region <region>
   ```

   > **Note**
   >
   > `USER_MANAGED_CLIENT_APPLICATION_CLIENT_SECRET`es una palabra clave reservada que se AWS Glue utiliza para hacer referencia a un valor secreto de cliente incluido en el secreto. Usa también la misma palabra clave cuando crees el secreto en la consola de Lake Formation.

2. Cree un rol de IAM que dé acceso al objeto de AWS Glue conexión al secreto creado en el paso anterior

   El objeto de AWS Glue conexión requiere acceso al AWS Secrets Manager secreto cuando se utiliza AWS Secrets Manager para almacenar, recuperar y actualizar el token secreto de OAuth. El objeto de AWS Glue conexión también requiere acceso para crear, describir y utilizar las interfaces de red de Amazon VPC cuando utiliza un punto de enlace de Amazon VPC para restringir la conectividad a su espacio de trabajo. Databricks

Cree una política de IAM y adjúntela a un rol de IAM. Añada el principal AWS Glue de servicio a la política de confianza.

En el siguiente ejemplo *<your-secrets-manager-ARN><your-vpc-id>*, sustituya y *<your-subnet-id1>* por su propia información.

Example Política de IAM

```
{
    "Version": "2012-10-17",
    "Statement": [{
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetSecretValue",
                "secretsmanager:DescribeSecret",
                "secretsmanager:PutSecretValue"
            ],
            "Resource": [
                "<your-secrets-manager-ARN>"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:CreateNetworkInterface",
                "ec2:DeleteNetworkInterface",
                "ec2:DescribeNetworkInterfaces"
            ],
            "Resource": "*",
            "Condition": {
                "ArnEquals": {
                    "ec2:Vpc": "arn:aws:ec2:region:account-id:vpc/<your-vpc-id>",
                    "ec2:Subnet": ["arn:aws:ec2:region:account-id:subnet/<your-
subnet-id1>"]
                }
            }
        }
    ]
}
```

Example Política de confianza

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Effect": "Allow",
        "Principal": {
            "Service": "glue.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
    }]
}
```

3. Cree una política de IAM que dé a Lake Formation acceso de lectura a la ubicación Amazon S3 del catálogo

   Como propietario de un catálogo federado en Data Catalog, utiliza Lake Formation para conceder a sus equipos de datos un acceso detallado a las tablas, un acceso detallado (a nivel de columnas, filas y celdas) y un acceso basado en etiquetas. Lake Formation utiliza una función de IAM que le da acceso a las ubicaciones subyacentes de Amazon S3 de sus tablas Iceberg remotas. Este acceso permite a Lake Formation vender credenciales de acceso limitado a los motores de análisis que consultan tablas remotas.

   Cree una política de IAM y asígnela a una función de IAM. Agregue el director de servicio de Lake Formation a la política de confianza de funciones de IAM.

   En el siguiente ejemplo, reemplace *<your-s3-bucket-N>* y *<your-kms-key>* por su propia información.

   Example Política de IAM

```
{
    "Version": "2012-10-17",
    "Statement": [{
            "Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::<your-s3-bucket-1>/*",
                "arn:aws:s3:::<your-s3-bucket-2>/*"
```

```
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::<your-s3-bucket-1>",
                "arn:aws:s3:::<your-s3-bucket-2>"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:Encrypt"
            ],
            "Resource": [
                "<your-kms-key>"
            ]
        }
    ]
}
```

Example Política de confianza

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "",
        "Effect": "Allow",
        "Principal": {
            "Service": "lakeformation.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
    }]
}
```

> **Note**
>
> Cuando utiliza la consola Lake Formation para crear un catálogo federado, la consola utiliza un único rol de IAM con ambas políticas asociadas para completar la configuración.

4. Cree un objeto de conexión AWS Glue

   Data Catalog admite ConnectionTypeDATABRICKSICEBERGRESTCATALOG: para conectar Data Catalog a. Databricks Este AWS Glue conector admite los métodos de autenticación OAuth2 y Custom.

   El siguiente ejemplo utiliza la configuración de autenticación de OAuth2 para crear una conexión. AWS Glue Modifique el *highlighed sections* añadiendo su información.

```
aws glue create-connection \
    --connection-input '{
"Name": "<your-glue-connection-to-databricks-unity-account>",
"ConnectionType": "DATABRICKSICEBERGRESTCATALOG",
"ConnectionProperties": {
    "INSTANCE_URL": "<your-databricks-workspace-catalog-URL>",
    "ROLE_ARN": "<your-IAM-role-for-secrets-and-VPC-access>"
},
"AuthenticationConfiguration": {
    "AuthenticationType": "OAUTH2",
    "OAuth2Properties": {
        "OAuth2GrantType": "CLIENT_CREDENTIALS",
        "TokenUrl": "<your-internal-or-external-token-server-url>",
        "OAuth2ClientApplication": {
            "UserManagedClientApplicationClientId": "<your-client-id>"
        },
        "TokenUrlParametersMap": {
            "scope": "all-apis"
        }
    },
    "SecretArn": "arn:aws:secretsmanager:<aws-region>:<your-aws-account-
id>:secret:<databricks-secret>"
}
}'
```

5. Registrar AWS Glue la conexión como recurso de Lake Formation

Con el objeto de AWS Glue conexión (creado en el paso 4) y el rol de IAM (creado en el paso 3), ahora puede registrar el objeto de AWS Glue conexión como un recurso gestionado por Lake Formation.

Reemplace el *<your-glue-connector-arn>* y el *<your-IAM-role-ARN-having-LF-access>* con su información.

```
aws lakeformation register-resource \
    --resource-arn <your-glue-connector-arn> \
    --role-arn <your-IAM-role-ARN-having-LF-access> \
    --with-federation \
    --with-privileged-access
```

6. Cree un catálogo federado en Data Catalog

   Tras crear un objeto de AWS Glue conexión y registrarlo en Lake Formation, puede crear un catálogo federado en el catálogo de datos:

   Proporcione al catálogo federado un nombre exclusivo en*<your-federated-catalog-name>*, haga referencia al catálogo en Databricks at *"<catalog-name-in-Databricks>* e introduzca el nombre de la conexión creado anteriormente en. *"<your-glue-connection-name>*

```
aws glue create-catalog \
    --name <your-federated-catalog-name> \
    --catalog-input '{
    "FederatedCatalog": {
        "Identifier": "<catalog-name-in-Databricks>",
        "ConnectionName": "<your-glue-connection-name>"
    },
    "CreateTableDefaultPermissions": [],
    "CreateDatabaseDefaultPermissions": []
}'
```

Consideraciones a la hora de integrarse con Databricks

- Al colocar recursos (como bases de datos y tablas) enDatabricks, Lake Formation no revoca automáticamente los permisos otorgados a ese recurso federado. Para eliminar los permisos de

acceso, debe revocar explícitamente los permisos que se concedieron anteriormente en el recurso federado mediante Lake Formation.

- Puede consultar las tablas de Iceberg almacenadas en Amazon S3 mediante esta integración. Si utiliza cualquier otro formato de tabla o almacenamiento de objetos, puede federar los metadatos de catálogos remotos AWS Glue y enumerar sus bases de datos y tablas, pero las operaciones de consulta `SELECT ColumnFoo from TableBar` fallarán durante la consulta y generarán el error «No se pudo leer la tabla». Apache Iceberg No se admite la ubicación de almacenamiento de objetos.
- Puede reutilizar la misma AWS Glue conexión para crear varios catálogos federados. Al eliminar un catálogo, no se eliminará el objeto de conexión asociado. Para eliminar un objeto de conexión, utilice el `aws glue delete-connection` comando AWS CLI y asegúrese de eliminar primero todos los catálogos asociados.

Incorporar los datos de Amazon Redshift al AWS Glue Data Catalog

Puede gestionar los datos analíticos en los almacenes de datos de Amazon Redshift en AWS Glue Data Catalog el (catálogo de datos) y unificar los lagos de datos de Amazon S3 y los almacenes de datos de Amazon Redshift. Amazon Redshift es un servicio de almacenamiento de datos en la nube totalmente gestionado y a escala de petabytes. AWS Un almacenamiento de datos de Amazon Redshift es una colección de recursos informáticos denominados nodos que se organizan en un grupo llamado clúster. Cada clúster ejecuta un motor Amazon Redshift y contiene una o más bases de datos.

En Amazon Redshift, puede crear clústeres aprovisionados y espacios de nombres sin servidor de Amazon Redshift y registrarlos con el Catálogo de datos. De este modo, puede unificar los datos en el almacenamiento administrado de Amazon Redshift (RMS) y los buckets de Amazon S3, así como acceder a los datos desde motores analíticos compatibles con Apache Iceberg.

Con el registro de los espacios de nombres y los clústeres, puede proporcionar acceso a los datos sin necesidad de copiarlos o moverlos. Para obtener más información sobre el registro de clústeres y espacios de nombres en Amazon Redshift, consulte Registro de clústeres y espacios de nombres de Amazon Redshift en el AWS Glue Data Catalog.

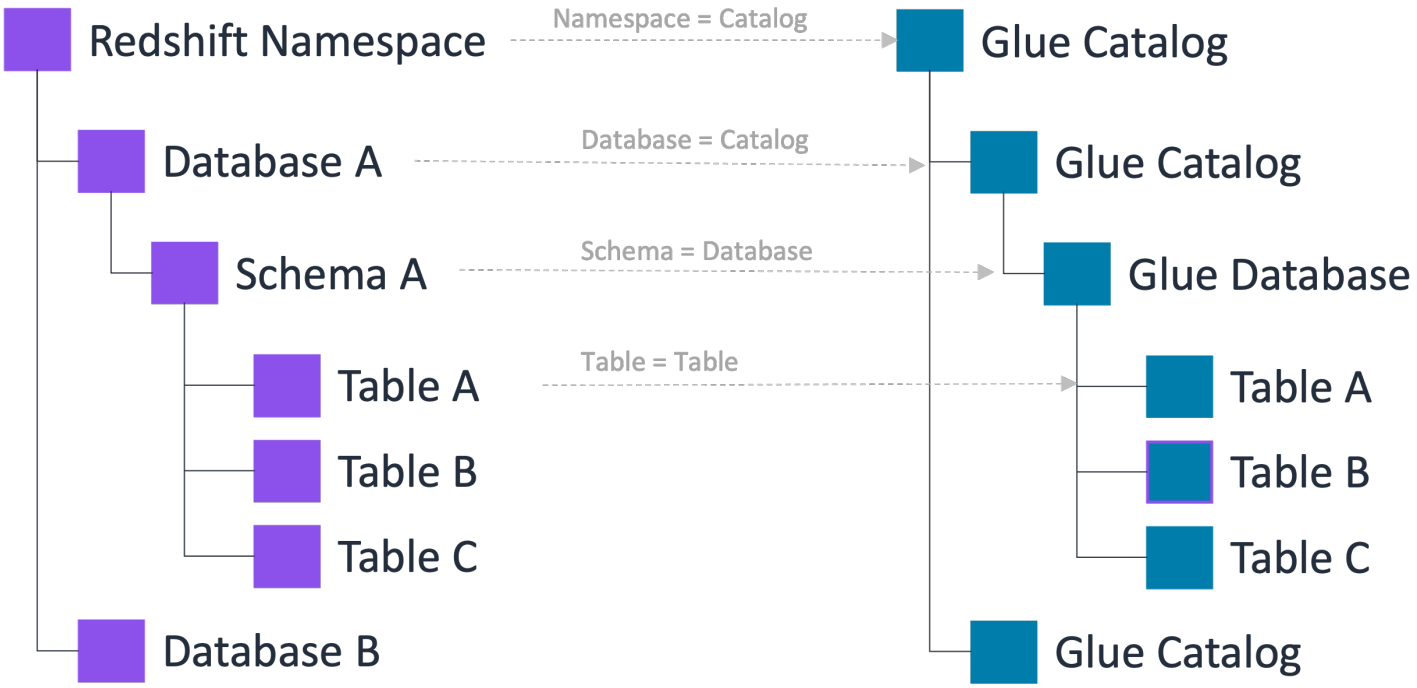
En Amazon Redshift, puede compartir datos a través de recursos compartidos de datos o mediante el registro de espacios de nombres y clústeres con el Catálogo de datos. Con los recursos

compartidos de datos, que funcionan en el nivel de objeto de base de datos individual, debe habilitar el uso compartido de cada tabla o vista. Por el contrario, la publicación de espacios de nombres funciona en el nivel de clúster o espacio de nombres. Al registrar un clúster o un espacio de nombres en el Catálogo de datos, todas las bases de datos y tablas que contiene se comparten automáticamente, sin tener que configurar el uso compartido de objetos individuales.

En el Catálogo de datos, puede crear un catálogo federado para cada espacio de nombres o clúster. Se habla de catálogo federado cuando apunta a una entidad ajena al Catálogo de datos. Las tablas y vistas del espacio de nombres de Amazon Redshift se muestran como tablas individuales en el Catálogo de datos. Puede compartir las bases de datos y tablas en el catálogo federado con entidades principales de IAM y usuarios de SAML seleccionados en la misma cuenta o en otra cuenta con Lake Formation. También puede incluir expresiones de filtro de filas y columnas para restringir el acceso a determinados datos. Para obtener más información, consulte Filtrado de datos y seguridad de celda en Lake Formation.

El Catálogo de datos admite una jerarquía de metadatos de tres niveles que incluye catálogos, bases de datos y tablas (y vistas). Al registrar un espacio de nombres en el Catálogo de datos, la jerarquía de datos de Amazon Redshift se asigna a la jerarquía de tres niveles del Catálogo de datos de la siguiente manera:

- El espacio de nombres de Amazon Redshift pasa a ser un catálogo de varios niveles en el Catálogo de datos.
- La base de datos asociada de Amazon Redshift está registrada como catálogo en el Catálogo de datos.
- El esquema de Amazon Redshift se convierte en una base de datos en el Catálogo de datos.
- La tabla de Amazon Redshift se convierte en una tabla en el Catálogo de datos.



Con esta jerarquía de metadatos de tres niveles, puede acceder a las tablas de Amazon Redshift mediante la notación de tres partes: catalog1/catalog2 «.database.table» del catálogo de datos. Además, los equipos de datos pueden mantener la misma organización que Amazon Redshift utiliza para organizar las tablas dentro de la cuenta del Catálogo de datos.

En Lake Formation, puede administrar de forma segura los datos de Amazon Redshift mediante un control de acceso detallado para los recursos del Catálogo de datos. Con esta integración, puede administrar, proteger y consultar datos analíticos de un único catálogo con un mecanismo de control de acceso común.

Para conocer las limitaciones, consulte Limitaciones a la hora de incorporar los datos del almacén de datos de Amazon Redshift al AWS Glue Data Catalog.

Temas

- Ventajas principales
- Funciones y responsabilidades
- Requisitos previos para administrar los espacios de nombres de Amazon Redshift en AWS Glue Data Catalog
- Creación de catálogos federados de Amazon Redshift
- Visualización de objetos del catálogo

- Actualización de un catálogo federado
- Acceso a un catálogo federado compartido
- Eliminación de un catálogo federado
- Consulta de catálogos federados
- Recursos adicionales

Ventajas principales

El registro de clústeres y espacios de nombres de Amazon Redshift con los lagos de datos de Amazon S3 AWS Glue Data Catalog y los almacenes de datos de Amazon Redshift y unificar los datos entre ellos ofrece las siguientes ventajas:

- Experiencia de consulta uniforme: consulte los datos administrados de Amazon Redshift y los datos en los buckets de Amazon S3 mediante cualquier motor de consultas compatible con Apache Iceberg, como Amazon EMR sin servidor y Amazon Athena, sin tener que mover ni copiar los datos.
- Acceso uniforme a los datos en todos los servicios: no necesita actualizar los nombres de las bases de datos y las tablas de sus canalizaciones de datos cuando accede a las mismas fuentes de datos federadas desde diferentes servicios de AWS análisis, ya que las fuentes de datos están registradas en el catálogo de datos.
- Fine-grained control de acceso: puede aplicar los permisos de Lake Formation para administrar el acceso a las fuentes de datos federadas mediante permisos de control de acceso detallados.

Funciones y responsabilidades

Rol	Responsabilidad
Administrador de clúster productor de Amazon Redshift	Registra el clúster o el espacio de nombres en el Catálogo de datos.
Administrador de lago de datos de Lake Formation	Acepta la invitación al clúster o espacio de nombres, crea catálogos federados y concede acceso a los catálogos federados a otras entidades principales.

Administrador de solo lectura de Lake Formation	Descubre el catálogo federado y consulta las tablas de Amazon Redshift en el catálogo federado.
Rol de transferencia de datos	Amazon Redshift asume en su nombre la transferencia de datos hacia y desde el bucket de Amazon S3.

A continuación, se muestran los pasos generales para proporcionar acceso a los usuarios a un espacio de nombres de Amazon Redshift:

1. En Amazon Redshift, el administrador del clúster productor registra un clúster o un espacio de nombres en el Catálogo de datos.
2. El administrador del lago de datos acepta la invitación al espacio de nombres del administrador del clúster productor de Amazon Redshift y crea un catálogo federado en el Catálogo de datos.

   Tras completar este paso, puede administrar el catálogo de espacios de nombres de Amazon Redshift en el Catálogo de datos.
3. Conceda permisos a los usuarios en catálogos, bases de datos y tablas. Puede compartir todo el catálogo de espacios de nombres o un subconjunto de tablas con usuarios de la misma cuenta o de otra cuenta.

Requisitos previos para administrar los espacios de nombres de Amazon Redshift en AWS Glue Data Catalog

1. Crear un administrador de data lake: cree un rol de IAM que esté autorizado a aceptar la invitación del espacio de nombres y cree los AWS Glue Data Catalog objetos (catálogos, bases de datos tables/views) y conceda permisos de Lake Formation a otros usuarios.

   Para obtener instrucciones paso a paso sobre cómo crear un administrador de lago de datos, consulte Crear un administrador de lago de datos.
2. Actualice los permisos del administrador del lago de datos.

   Además de los permisos de administrador del lago de datos, el administrador del lago de datos necesita los siguientes permisos para aceptar una invitación al espacio de nombres de Amazon

Redshift en Lake Formation, crear o actualizar los recursos del Catálogo de datos y habilitar el acceso al lago de datos:

JSON

```
{
"Version":"2012-10-17",
"Id": "glue-enable-datalake-access",
"Statement": [{
        "Effect": "Allow",
        "Action": [
            "redshift:AssociateDataShareConsumer",
            "redshift:DescribeDataSharesForConsumer",
            "redshift:DescribeDataShares",
            "redshift-serverless:CreateNamespace",
            "redshift-serverless:CreateWorkgroup",
            "redshift-serverless:DeleteNamespace",
            "redshift-serverless:DeleteWorkgroup",
            "ec2:DescribeAccountAttributes",
            "ec2:DescribeSubnets",
            "ec2:DescribeSecurityGroups",
            "ec2:DescribeAvailabilityZones",
            "s3:createBucket",
            "s3:deleteBucket",
            "s3:putBucketPolicy",
            "s3:putEncryptionConfiguration",
            "s3:putLifecycleConfiguration",
            "s3:putBucketVersioning",
            "iam:CreateRole"
        ],
        "Resource": "*"
    },
    {
        "Action": [
            "iam:PassRole"
        ],
        "Effect": "Allow",
        "Resource": "arn:aws:iam::*:role/data transfer role name",
        "Condition": {
            "StringLike": {
                "iam:PassedToService": [
                    "glue.amazonaws.com"
                ]
```

```
                }
            }
        }
    ]
}
```

3. Si el rol de IAM utilizado para crear catálogos federados no es el de administrador del lago de datos, debe conceder al rol el permiso `Create catalog`.

   Cómo crear creadores de catálogos

   a. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   b. En Administración, seleccione Roles y tareas administrativas.

   c. Elija Conceder.

   d. En la pantalla Conceder permisos, seleccione un usuario o rol de IAM.

   e. Seleccione el permiso Crear catálogo.

   f. Opcionalmente, también puede conceder el permiso concesible Crear catálogo. El permiso concesible permite al creador del catálogo conceder el permiso `Create catalog` a otras entidades principales.

   g. Elija Conceder.

   AWS CLI ejemplo de concesión de permisos para crear un catálogo federado.

```
aws lakeformation grant-permissions \
--cli-input-json \
'{
    "Principal": {
     "DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/Admin"
    },
    "Resource": {
        "Catalog": {
        }
    },
    "Permissions": [
        "CREATE_CATALOG",
        "DESCRIBE"
    ]
}'
```

4. Cree un rol de administrador de solo lectura para descubrir los catálogos federados de Amazon Redshift en el Catálogo de datos desde el editor de consultas V2 de Amazon Redshift.

   Para consultar las tablas de Amazon Redshift del catálogo federado desde el editor de consultas V2 de Amazon Redshift, asegúrese de que la política del rol de administrador de solo lectura contenga el ARN del rol vinculado al servicio de Amazon Redshift `AWSServiceRoleForRedshift`.

```
 aws lakeformation put-data-lake-settings
       --region us-east-1 \
       --data-lake-settings \
 '{
   "DataLakeAdmins": [{"DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/
Admin"}],
   "ReadOnlyAdmins":[{"DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/
aws-service-role/redshift.amazonaws.com/AWSServiceRoleForRedshift"}],
   "CreateDatabaseDefaultPermissions":[],
   "CreateTableDefaultPermissions":[],
   "Parameters":{"CROSS_ACCOUNT_VERSION":"4","SET_CONTEXT":"TRUE"}
  }'
```

5. Cree un rol de transferencia de datos que Amazon Redshift pueda asumir en su nombre para transferir datos hacia y desde el bucket de Amazon S3.

   Al habilitar el acceso al lago de datos para motores de consultas compatibles con Apache Iceberg, como Athena, Amazon EMR en Amazon EC2, para acceder a los recursos de Amazon Redshift del Catálogo de datos, debe crear un rol de IAM con los permisos necesarios para realizar la transferencia de datos hacia y desde el bucket de Amazon S3.

   JSON

```
{
    "Version":"2012-10-17",
    "Id": "glue-enable-datalake-access",
    "Statement": [{
      "Sid": "DataTransferRolePolicy",
        "Effect": "Allow",
        "Action": [ "glue:GetCatalog",
                    "glue:GetDatabase",
                    "kms:GenerateDataKey",
                    "kms:Decrypt"],
```

```
        "Resource": "*"
    }
    ]
}
```

6. Añada la siguiente política de confianza a la función de transferencia de datos para AWS Glue que los servicios de Amazon Redshift asuman la función de transferir datos hacia y desde el bucket de Amazon S3.

   JSON

```
      {
    "Version":"2012-10-17",
    "Statement": [{
        "Effect": "Allow",
        "Principal": {
        "Service": [
        "redshift.amazonaws.com",
        "glue.amazonaws.com"
        ]
    },
        "Action": "sts:AssumeRole"
    }]
}
```

7. Añada la siguiente política de claves a la AWS KMS clave si utiliza una clave gestionada por el cliente para cifrar los datos en Amazon cluster/namespace Redshift. Sustituya el número de cuenta por un número de AWS cuenta válido y especifique el nombre de la función de transferencia de datos. De forma predeterminada, los datos del clúster de Amazon Redshift se cifran con una clave de KMS. Lake Formation ofrece una opción para crear la clave de KMS personalizada para el cifrado. Si utiliza una clave administrada por el cliente, debe agregar políticas de claves específicas a la clave.

   Para obtener más información sobre la administración de los permisos de una clave administrada por el cliente, consulte Claves administradas por el cliente.

   JSON

```
{
```

```
    "Version":"2012-10-17",
    "Id": "auto-redshift-3",
    "Statement": [{
            "Sid": "RedshiftAllowAccessPolicy",
            "Effect": "Allow",
            "Principal": {
                "AWS": "*"
            },
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:CreateGrant",
                "kms:DescribeKey"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "kms:CallerAccount": "111122223333",
                    "kms:ViaService": "redshift.us-east-1.amazonaws.com"
                }
            }
        },
        {
            "Sid": "RedshiftServerlessAllowAccessPolicy",
            "Effect": "Allow",
            "Principal": {
                "AWS": "*"
            },
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:CreateGrant",
                "kms:DescribeKey"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "kms:CallerAccount": "111122223333",
                    "kms:ViaService": "redshift-serverless.us-
east-1.amazonaws.com"
```

```
                }
            }
        },
        {
            "Sid": "DirectMetadataAccess",
            "Effect": "Allow",
            "Principal": {
            "AWS": "arn:aws:iam::111122223333:root"
            },
            "Action": [
                "kms:Describe*",
                "kms:Get*",
                "kms:List*",
                "kms:RevokeGrant"
            ],
            "Resource": "*"
        },
        {
            "Sid": "GenerateDataKeyDecryptDataTransferRole",
            "Effect": "Allow",
            "Principal": {
            "AWS": "arn:aws:iam::111122223333:role/data-transfer-role-name"
            },
            "Action": [
                "kms:GenerateDataKey",
                "kms:Decrypt"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                "kms:ViaService": "s3.us-east-1.amazonaws.com"
                }
            }
        }
    ]
}
```

Creación de catálogos federados de Amazon Redshift

En este tema, se describen los pasos que debe seguir para aceptar una invitación a un clúster o espacio de nombres, crear un catálogo federado de varios niveles y conceder permisos a otras entidades principales. Puede completar estas tareas con la consola de Lake Formation, el AWS Command Line Interface (AWS CLI) o APIs/SDKs. Los ejemplos de este tema muestran al productor cluster/namespace, el catálogo de datos y el consumidor de datos en la misma cuenta.

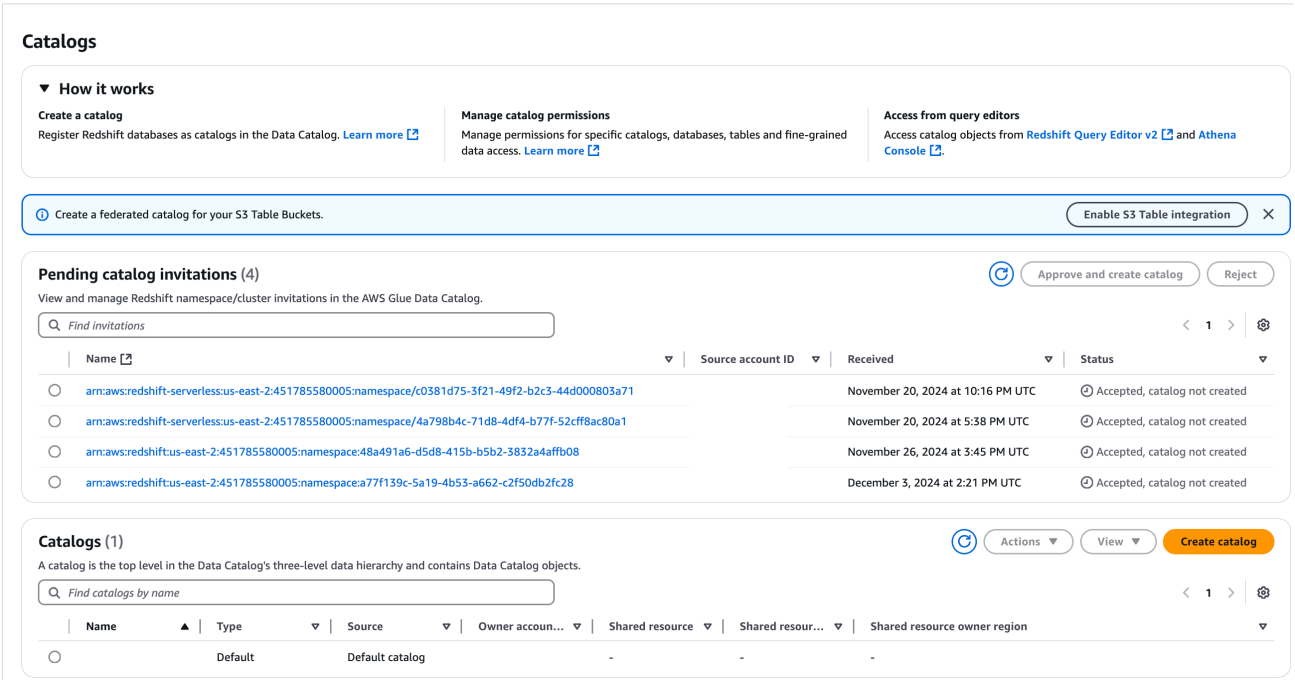
Para obtener más información sobre las capacidades entre cuentas de Lake Formation, consulte Cross-account intercambio de datos en Lake Formation.

Cómo administrar un espacio de nombres de Amazon Redshift en el Catálogo de datos

1. Revise una invitación al espacio de nombres y acéptela.

   Console

   1. Inicie sesión en la consola de Lake Formation como administrador de un lago de datos en https://console.aws.amazon.com/lakeformation/. Vaya a la página Catálogos en Catálogo de datos.
   2. Revise la invitación al espacio de nombres para el que tiene autorización de acceso. La columna Estado indica su estado de participación actual en el espacio de nombres. El estado No aceptada indica que se le ha agregado al espacio de nombres pero aún no ha aceptado la invitación o la ha rechazado.



3. Para responder a una invitación a un espacio de nombres o a un clúster, seleccione el nombre de la invitación y elija Revisar invitación. En Aceptar o rechazar invitación, revise los detalles de la invitación. Seleccione Aceptar para aceptar la invitación o Rechazar para rechazarla. Si rechaza la invitación, no obtendrá acceso al espacio de nombres.

AWS CLI

En los ejemplos siguientes se muestra cómo ver, aceptar y registrar la invitación. Sustituya la Cuenta de AWS identificación por una Cuenta de AWS identificación válida. Sustituya `data-share-arn` por el nombre de recurso de Amazon (ARN) real que hace referencia al espacio de nombres.

1. Consulte una invitación pendiente.

```
aws redshift describe-data-shares \
 --data-share-arn 'arn:aws:redshift:us-
east-1:123456789012:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
ds_internal_namespace' \
```

2. Acepte una invitación.

```
 aws redshift associate-data-share-consumer \
 --data-share-arn 'arn:aws:redshift:us-
east-1:123456789012:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
ds_internal_namespace' \
 --consumer-arn 'arn:aws:glue:us-east-1:123456789012:catalog'
```

3. Registre el clúster o espacio de nombres en la cuenta de Lake Formation. Utilice la operación RegisterResourceAPI para registrar el datashare en Lake Formation. `DataShareArn`es el parámetro de entrada de. `ResourceArn`

> **Note**
>
> Este es un paso obligatorio.

```
aws lakeformation register-resource \
```

```
 --resource-arn 'arn:aws:redshift:us-
east-1:123456789012:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
ds_internal_namespace'
```

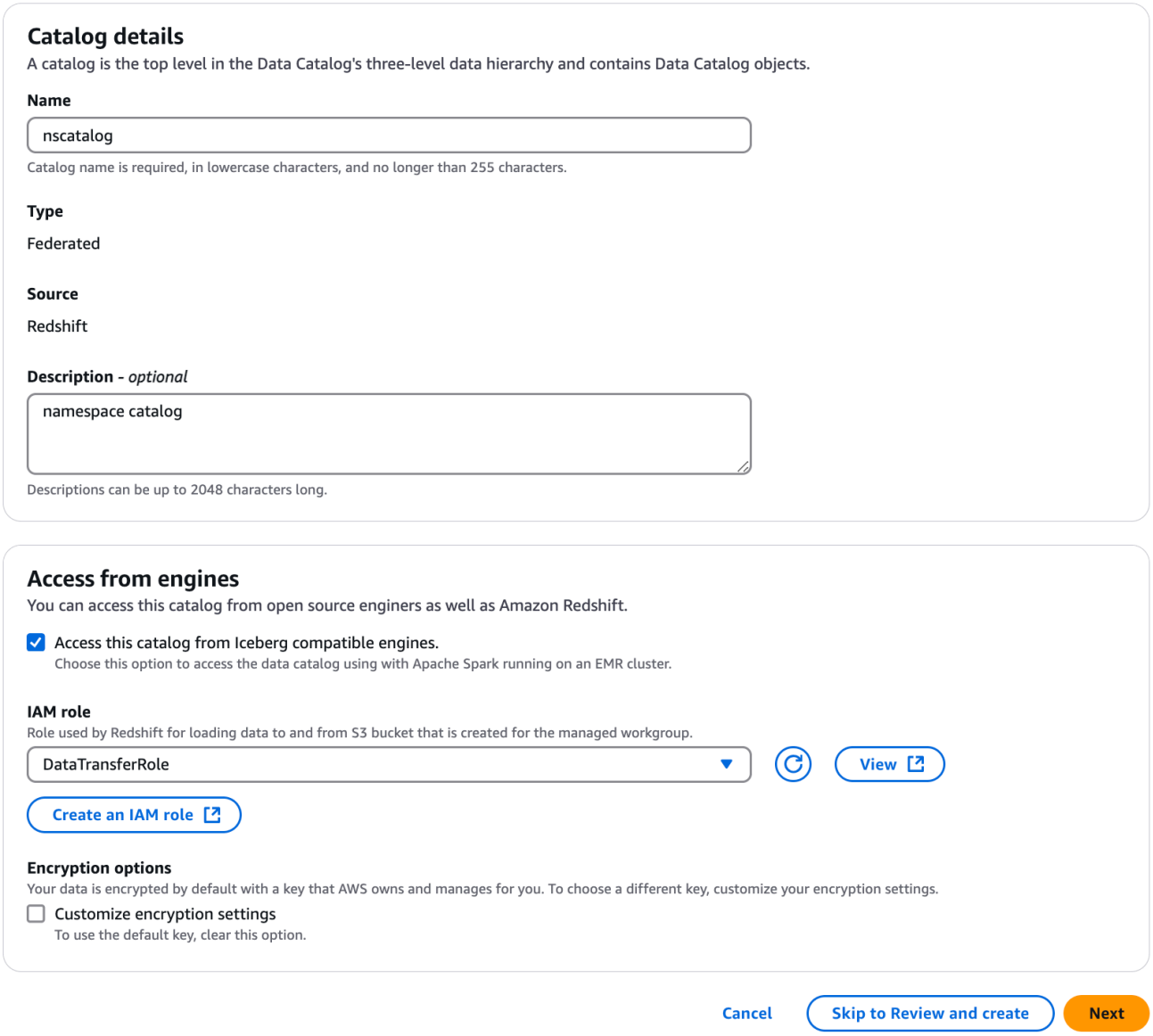
2. Cree un catálogo federado.

   Tras aceptar la invitación, debe crear un catálogo federado en el Catálogo de datos que asigne los objetos del espacio de nombres de Amazon Redshift al Catálogo de datos. Debe ser administrador de lago de datos o un usuario o rol con los permisos necesarios para crear un catálogo.

   Console

   1. Tras aceptar la invitación al espacio de nombres, aparecerá la página Establecer detalles del catálogo.
   2. En la página Establecer detalles del catálogo, escriba un nombre exclusivo para el catálogo. Utilice minúsculas para los nombres de los catálogos. Los nombres de catálogo deben tener 255 caracteres o menos. Este identificador se utiliza para mapear el espacio de nombres internamente en la jerarquía de metadatos (catalogid.dbName.schema.table).
   3. Escriba una descripción para el catálogo. La descripción debe tener 2048 caracteres o menos.
   4. A continuación, elija la casilla de verificación Acceso a este catálogo desde motores compatibles con Iceberg para permitir el acceso a los recursos de Amazon Redshift mediante motores analíticos compatibles con Apache Iceberg, como Athena y Apache Spark en Amazon EMR.

      No necesita habilitar el acceso al lago de datos para acceder a los catálogos federados mediante Amazon Redshift.



5. Para permitir que estos motores de consultas lean y escriban en los espacios de nombres de Amazon Redshift, AWS Glue crea un clúster de Amazon Redshift administrado con los recursos informáticos y de almacenamiento necesarios para realizar operaciones de lectura y escritura sin afectar a las cargas de trabajo del almacén de datos de Amazon Redshift.

   También debe proporcionar un rol de IAM con los permisos necesarios para transferir datos hacia y desde el bucket de Amazon S3.

6. De forma predeterminada, los datos del clúster de Amazon Redshift se cifran mediante una clave AWS gestionada. Lake Formation ofrece una opción para crear la clave de KMS personalizada para el cifrado. Si utiliza una clave administrada por el cliente, debe agregar políticas de claves específicas a la clave.

Elija Personalizar la configuración de cifrado si utiliza una clave gestionada por el cliente para cifrar los datos en Amazon cluster/namespace Redshift. Para usar una clave personalizada, debe agregar una política de claves administradas personalizadas adicionales a la clave de KMS. Para obtener más información, consulte Requisitos previos para administrar los espacios de nombres de Amazon Redshift en AWS Glue Data Catalog.

AWS CLI

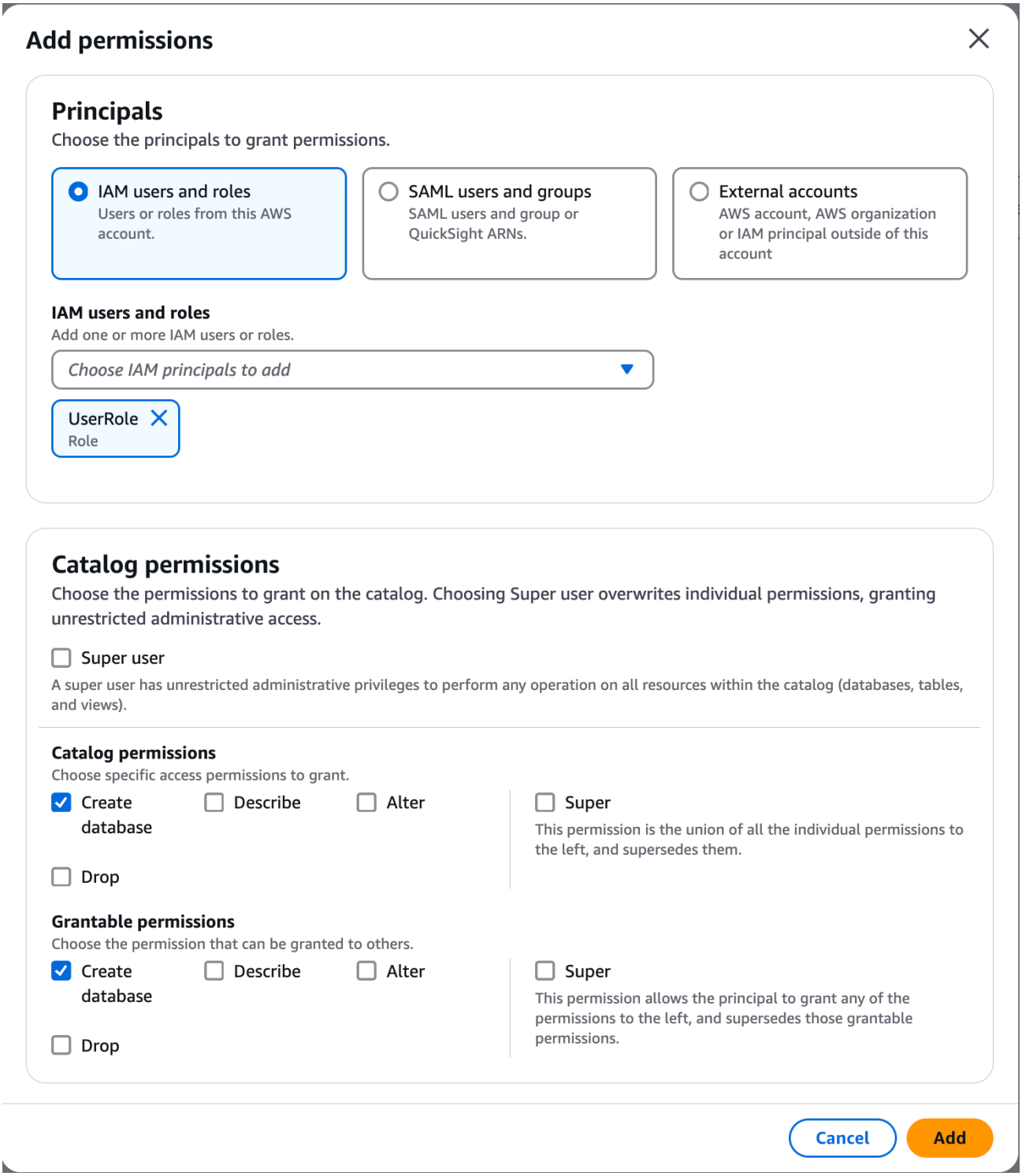
Utilice el siguiente código de ejemplo para crear un catálogo con los datos de Amazon Redshift publicados en el Catálogo de datos mediante la AWS CLI.

```
aws glue create-catalog
--cli-input-json \
'{
    "Name": "nscatalog",
    "CatalogInput": {
        "Description": "Redshift federated catalog",
        "CreateDatabaseDefaultPermissions" : [],
        "CreateTableDefaultPermissions": [],
        "FederatedCatalog": {
            "Identifier": "arn:aws:redshift:us-
east-1:123456789012:datashare:11524d7f-f56d-45fe-83f7-d7bb0a4d6d71/
ds_internal_namespace",
            "ConnectionName": "aws:redshift"
        },
        "CatalogProperties": {
          "DataLakeAccessProperties" : {
            "DataLakeAccess" : true,
            "DataTransferRole" :
 "arn:aws:iam::123456789012:role/DataTransferRole"
         }
       }
    }
}'
```

3. Otorgue permisos a los usuarios de su cuenta o de cuentas externas.

Consola de administración de AWS

1. Seleccione Siguiente para conceder permisos a otros usuarios sobre los catálogos, bases de datos y tablas compartidos.
2. En la pantalla Agregar permisos, elija las entidades principales y los tipos de permisos que desea conceder.



a. En la sección Entidades principales, elija uno de los tipos y, a continuación, especifique las que van a recibir los permisos concedidos.

   - Usuarios y roles de IAM: elija uno o varios usuarios o roles de la lista de usuarios y roles de IAM.

- Usuarios y grupos de SAML: para SAML y Amazon Quick usuarios y grupos, introduzca uno o más nombres de recursos de Amazon (ARN) para los usuarios o grupos federados a través de SAML, o ARN para los usuarios o grupos de Amazon Quick. Pulse Intro después de cada ARN.

  Para obtener información sobre cómo construir los ARN, consulte los comandos grant and revoke. AWS CLI AWS CLI
- Cuentas externas: para AWS AWS la organización o el director de IAM, introduzca uno o más ID de AWS cuenta, ID de organización, ID de unidad organizativa o ARN válidos para el usuario o rol de IAM. Pulse Intro después de cada ID. El ID de una organización consta de una «o-» seguida de 10 a 32 letras minúsculas o dígitos. Un ID de una unidad organizativa comienza por «ou-» seguidos de 4 a 32 letras minúsculas o dígitos (el ID de la raíz que contiene la UO). Esta cadena va seguida de un segundo guion «-» y de 8 a 32 letras minúsculas o dígitos adicionales.

b. En la sección Permisos, seleccione los permisos y los permisos concedibles.

   En Permisos de catálogo, seleccione uno o más permisos para conceder. En Permisos concedibles, selecciona los permisos que el destinatario de la subvención puede conceder a otros directores de su cuenta. AWS Esta opción no es compatible cuando se conceden permisos a una entidad principal de IAM desde una cuenta externa.

   Elija Superusuario para conceder al usuario permisos ilimitados en los recursos (bases de datos, tablas y vistas) del catálogo.

3. Elija Añadir.

AWS CLI

Utilice los siguientes ejemplos para conceder permisos en catálogos, bases de datos y tablas mediante la AWS CLI:

- En el siguiente ejemplo, se muestra cómo conceder permisos en el catálogo federado.

```
aws lakeformation grant-permissions
 --cli-input-cli-json \
   '{
         "Principal": {
              "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/
non-admin"
```

```
        },
        "Resource": {
             "Catalog": {
                   "Id": "123456789012:nscatalog"
              }
          },
          "Permissions": [
                 "DESCRIBE","CREATE_CATALOG"
           ],
          "PermissionsWithGrantOption": [
           ]
   }'
```

- Utilice la siguiente política de ejemplo para conceder permisos sobre una base de datos.

```
aws lakeformation grant-permissions \
  --cli-input-json \
         '{
             "Principal": {

 "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-admin"
             },
             "Resource": {
                 "Database": {
                     "CatalogId": "123456789012:nscatalog/dev",
                     "Name": "public"
                 }
             },
             "Permissions": [
                 "ALL"
             ]
         }'
```

- En el siguiente ejemplo, se muestra cómo conceder permisos en una tabla en la base de datos de Amazon Redshift.

```
aws lakeformation grant-permissions \
  --cli-input-json \
       '{
           "Principal": {

 "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-admin"
           },
```

```
            "Resource": {
                "Table": {
                    "CatalogId": "123456789012:nscatalog2/dev",
                    "DatabaseName": "public",
                    "TableWildcard" : {}
                }
            },
            "Permissions": [
                "ALL"
            ]
        }'
```

4. Seleccione Siguiente para revisar los detalles del catálogo y crear un catálogo federado. El catálogo federado recién creado y los objetos del catálogo aparecen en la página Catálogos.

   Se hace referencia a un catálogo federado de Amazon Redshift con `catalogID = 123456789012:Redshift-federated catalog id`.

Visualización de objetos del catálogo

Tras crear el catálogo federado, puede ver los objetos del catálogo mediante la consola de Lake Formation o la AWS CLI.

Consola de administración de AWS

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. Seleccione Catálogos en Catálogo de datos.
3. Elija un catálogo federado en la lista de la página Catálogos.
4. La página de resumen del catálogo muestra los objetos del catálogo (bases de datos y tablas) sobre los que tiene permisos. La pestaña Permisos muestra las entidades principales de IAM a las que se les han concedido permisos sobre estos objetos.

AWS CLI

- El siguiente AWS CLI ejemplo muestra cómo solicitar el catálogo de nivel superior.

```
aws glue get-catalog \
--catalog-id 123456789012:nscatalog
```

Respuesta

```
{
    "Catalog": {
        "CatalogId": "123456789012:nscatalog",
        "Name": "nscatalog",
        "ResourceArn": "arn:aws:glue:us-east-1:123456789012:catalog/nscatalog",
        "Description": "Redshift published Catalog",
        "CreateTime": "2024-09-05T14:49:16-07:00",
        "FederatedCatalog": {
            "Identifier": "arn:aws:redshift:us-
east-1:123456789012:datashare:b1234589-e823-4a14-ad8e-077085540a50/
ds_internal_namespace",
            "ConnectionName": "aws:redshift"
        },
        "CatalogProperties": {
            "DataLakeAccessProperties": {
                "DataLakeAccess": true,
                "DataTransferRole": "arn:aws:iam::123456789012:role/
DataTransferRole",
                "KmsKey": "AWS_OWNED_KMS_KEY",
                "ManagedWorkgroupName": "123456789012:nscatalog",
                "ManagedWorkgroupStatus": "AVAILABLE",
                "RedshiftDatabaseName": "dev"
            }
        },
        "CatalogIdentifier": "e2309c2c2fb048f1a3069dfdc1c7883e",
        "CreateTableDefaultPermissions": [],
        "CreateDatabaseDefaultPermissions": []
    }
}
```

- En el siguiente ejemplo, se muestra cómo solicitar todos los catálogos de la cuenta.

```
aws glue get-catalogs \
  --recursive
```

- La siguiente solicitud de ejemplo muestra cómo obtener un catálogo en el nivel de base de datos de Amazon Redshift.

```
aws glue get-catlog \
```

```
 --catalog-id 123456789012:namespace catalog name/redshift database name
```

- La siguiente solicitud de ejemplo muestra cómo obtener las bases de datos en el catálogo de nivel de base de datos de Amazon Redshift.

```
aws glue get-databases \
--catalog-id 123456789012:namespace catalog name/redshift database name
```

- La siguiente solicitud de ejemplo muestra cómo obtener una tabla de Amazon Redshift en el catálogo.

```
aws glue get-table \
  --catalog-id 123456789012:parent catalog name/redshift database \
  --database-name redshift schema name \
  --name table name
```

- En el siguiente ejemplo, se muestra cómo obtener todas las tablas de una base de datos de Amazon Redshift.

```
aws glue get-tables \
 --catalog-id 123456789012:namespace catalog name/redshift database name \
 --database-name RS schema name
```

Actualización de un catálogo federado

Puede actualizar un catálogo federado de Amazon Redshift en el catálogo de datos mediante la consola de Lake Formation AWS CLI o la UpdateCatalogoperación de API.

Consola de administración de AWS

Siga estos pasos para actualizar su catálogo federado mediante la consola de Lake Formation.

1. Inicie sesión en y abra la Consola de administración de AWS consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación izquierdo, en Catálogo de datos, elija Catálogos.
3. En la página Catálogos, elija el catálogo federado de Amazon Redshift que desee actualizar.
4. En Acciones, elija Editar.

5. En la pantalla Establecer detalles del catálogo, en la sección Acceso desde motores, elija Acceso a este catálogo desde motores compatibles con Iceberg. Al seleccionar esta opción, se habilitará el acceso al lago de datos para los motores de consulta compatibles con Apache Iceberg.
6. A continuación, cree un nuevo rol de IAM o elija un rol de IAM existente que tenga la política que concede permisos para realizar transferencias de datos hacia y desde el bucket de Amazon S3.

   Para obtener más información sobre los permisos de IAM, consulte Requisitos previos para administrar los espacios de nombres de Amazon Redshift en AWS Glue Data Catalog.
7. De forma predeterminada, los datos del clúster de Amazon Redshift se cifran con una clave administrada mediante una Clave administrada de AWS. Si decide cifrar los datos con una clave administrada por el cliente, cree una clave de KMS o elija una clave existente que tenga los permisos definidos en la sección Requisitos previos para administrar los espacios de nombres de Amazon Redshift en AWS Glue Data Catalog.
8. Seleccione Save.

   Una vez que la operación finaliza correctamente, la página Detalles del catálogo muestra el nombre del grupo de trabajo administrado con el estado Correcto.

AWS CLI

El siguiente es un ejemplo de la entrada de la CLI `update-catalog` con el acceso al lago de datos deshabilitado al establecer el valor del parámetro `DataLakeAacess` en `false`.

```
aws glue update-catalog  --cli-input-json \
'{
    "Name": "nscatalog",
    "CatalogInput": {
        "Description": "Redshift published catalog",
        "CreateDatabaseDefaultPermissions" : [],
        "CreateTableDefaultPermissions": [],
        "FederatedCatalog": {
            "Identifier": "arn:aws:redshift:us-
east-1:123456789012:datashare:11524d7f-f56d-45fe-83f7-d7bb0a4d6d71/
ds_internal_namespace",
            "ConnectionName": "aws:redshift"
        },
        "CatalogProperties": {
```

```
          "DataLakeAccessProperties" : {
            "DataLakeAccess" : false
        }
       }
    }
}'
```

Acceso a un catálogo federado compartido

AWS Lake Formation Las funciones multicuenta permiten a los usuarios compartir de forma segura lagos de datos distribuidos entre varias Cuenta de AWS entidades u AWS organizaciones o directamente con los directores de IAM en otra cuenta, lo que proporciona un acceso detallado a los metadatos y a los datos subyacentes.

Lake Formation utiliza el servicio AWS Resource Access Manager (AWS RAM) para facilitar el intercambio de recursos. Cuando compartes un recurso del catálogo con otra cuenta, AWS RAM envía una invitación a la cuenta del concesionario para que acepte o rechace la concesión del recurso.

Los servicios analíticos integrados, como Amazon Athena y Redshift Spectrum, requieren enlaces a recursos para poder incluir recursos compartidos en las consultas. Los directores deben crear un enlace de recurso en su cuenta AWS Glue Data Catalog a un recurso compartido desde otro. Cuenta de AWS Para obtener más información sobre los enlaces de recursos, consulte Cómo funcionan los enlaces de recursos en Lake Formation.

Un contenedor de enlaces de catálogo es un objeto de catálogo de datos que hace referencia a un catálogo local o federado a nivel de base de datos entre cuentas desde otras cuentas. AWS También puede crear enlaces de bases de datos y enlaces de tablas dentro de un contenedor de enlaces de catálogo. Al crear un enlace a una base de datos o un enlace a una tabla, debe especificar un recurso de destino que resida en el mismo catálogo en el nivel de base de datos de Amazon Redshift de destino (base de datos de Amazon Redshift).

Para crear un contenedor de enlaces a un catálogo, necesita el permiso `CREATE_CATALOG` de Lake Formation o `glue:CreateCatalog`.

Debe tener la versión 4 o superior de la configuración de versiones entre cuentas para compartir bases de datos o tablas del catálogo federado entre sí. Cuenta de AWS

Creación de un contenedor de enlaces de catálogo para un catálogo federado entre cuentas

Puede crear un contenedor de enlaces de catálogo que apunte a un catálogo federado a nivel de base de datos de Redshift en cualquier AWS región mediante la AWS Lake Formation consola, la AWS Glue `CreateCatalog` API o (). AWS Command Line Interface AWS CLI

Cómo crear un contenedor de enlaces de catálogo para un catálogo compartido (consola)

1. Abra la consola en. AWS Lake Formation https://console.aws.amazon.com/lakeformation/ Inicie sesión como una entidad principal que tenga el permiso `CREATE_CATALOG` de Lake Formation.
2. En el panel de navegación, elija Catálogos y, a continuación, elija Crear catálogo.
3. En la página Establecer detalles del catálogo, facilite la siguiente información:

   Name

   Escriba un nombre que cumpla las mismas reglas que el nombre de un catálogo. El nombre puede ser el mismo que el del catálogo compartido de destino.

   Tipo

   Elija el contenedor de enlaces de catálogos como tipo de catálogo.

   origen

   Elija `Redshift`.

   Catálogo de Redshift de destino

   Seleccione un catálogo federado de nivel de base de datos de Redshift o elija un catálogo local (propio) en la lista.

   La lista contiene todos los catálogos compartidos con su cuenta. Observe que el ID de la cuenta del propietario del catálogo aparece con cada catálogo. Si no se muestra un catálogo que sabe que se ha compartido con su cuenta, compruebe lo siguiente:

   - Si no es administrador de un lago de datos, compruebe que el administrador del lago de datos le haya concedido los permisos de Lake Formation sobre el catálogo.
   - Si es administrador de un lago de datos y su cuenta no pertenece a la misma organización de AWS que la cuenta que los concede, asegúrese de haber aceptado la invitación para compartir recursos AWS Resource Access Manager (AWS RAM) del catálogo. Para

obtener más información, consulte Aceptar una invitación para compartir recursos de AWS RAM.

> **Note**
>
> Al crear un contenedor de enlaces de catálogo a través de la consola, es posible que aparezca el menú desplegable Catálogo de Redshift de destino `No matches` al intentar seleccionar un catálogo de Redshift entre cuentas. A pesar de que se muestre, puede especificar manualmente el ARN de destino del catálogo federado de nivel de base de datos de Amazon Redshift (base de datos de Amazon Redshift) en el campo de entrada y el formulario seguirá funcionando correctamente. Por ejemplo: `arn:aws:glue:us-east-1:123456789012:catalog/federated-catalog-redshift/dev`.
>
> Este comportamiento se debe a que la consola solo puede buscar posibles candidatos dentro de la cuenta en la que se ha iniciado sesión actualmente. El menú desplegable está pensado como una característica de autocompletar, pero se pueden especificar los ARN manualmente para el acceso entre cuentas.

4. Para permitir que los motores de consultas de Apache Iceberg lean y escriban en los espacios de nombres de Amazon Redshift AWS Glue , crea un clúster de Amazon Redshift administrado con los recursos informáticos y de almacenamiento necesarios para realizar operaciones de lectura y escritura sin afectar a las cargas de trabajo del almacén de datos de Amazon Redshift. Debe proporcionar un rol de IAM con los permisos necesarios para transferir datos hacia y desde el bucket de Amazon S3.
5. Elija Siguiente.
6. (Opcional) Elija Agregar permisos para conceder permisos a otras entidades principales.

   No obstante, conceder permisos en un contenedor de enlaces de catálogo no otorga permisos en el catálogo (vinculado) de destino. Debe conceder permisos en el catálogo de destino por separado para que el enlace al catálogo esté visible en Athena.
7. A continuación, revise los detalles del contenedor de enlaces de catálogo y elija Crear catálogo.

   A continuación, puede ver el nombre del contenedor de enlaces en la página Catálogos.

   Ahora puede crear enlaces de bases de datos y enlaces de tablas en el contenedor de enlaces de catálogo para permitir el acceso desde los motores de consulta.

Ejemplo de la CLI para la creación de un contenedor de enlaces de catálogo

- En el siguiente ejemplo, el objeto `TargetRedshiftCatalog` especifica el ARN del catálogo federado de nivel de base de datos de Amazon Redshift (base de datos de Amazon Redshift). `DataLakeAccess` debe estar habilitado al crear el contenedor de enlaces de catálogo.

```
aws glue create-catalog \
  --cli-input-json
    '{
        "Name": "linkcontainer",
        "CatalogInput": {
            "TargetRedshiftCatalog": {
               "CatalogArn": "arn:aws:us-east-1:123456789012:catalog/nscatalog/dev"
             },
            "CatalogProperties": {
              "DataLakeAccessProperties" : {
                "DataLakeAccess" : true,
                "DataTransferRole" : "arn:aws:iam::111122223333:role/
DataTransferRole"
             }
           }
        }
    }'
```

Creación de enlaces a recursos en el contenedor de enlaces de catálogo

Puede crear enlaces a recursos para enlaces de bases de datos y tablas dentro de un contenedor de enlaces de catálogo. Al crear enlaces a recursos de base de datos o enlaces a recursos de tabla, debe especificar un recurso de destino que resida en el mismo catálogo de nivel de base de datos de Amazon Redshift de destino (base de datos de Amazon Redshift) que el enlace al que señala el contenedor.

Puede crear un enlace de recurso a una base de datos o tabla de Amazon Redshift compartida mediante la AWS Lake Formation consola, la API o AWS Command Line Interface ()AWS CLI.

- Para obtener instrucciones detalladas, consulte Crear un enlace de recursos a una base de datos de Catálogo de datos compartida.

A continuación, se AWS CLI muestra un ejemplo de cómo crear un enlace a un recurso de base de datos en un contenedor de enlaces de catálogo.

```
aws glue create-database \
  --cli-input-json \
    '{
        "CatalogId": "111122223333:linkcontainer",
        "DatabaseInput": {
            "Name": "dblink",
             "TargetDatabase": {
               "CatalogId": "123456789012:nscatalog/dev",
                "DatabaseName": "schema1"
             }
        }
    }'
```

- Para crear un enlace de recursos de tabla en un contenedor de enlaces de catálogo, primero debe crear una AWS Glue base de datos en el entorno local AWS Glue Data Catalog que contenga el enlace de recursos de tabla.

  Para obtener más información sobre la creación de enlaces de recursos para tablas compartidas, consulte Crear un enlace de recursos a una tabla de Catálogo de datos compartida.

  - Ejemplo de la CLI para la creación de una base de datos que contenga el enlace a los recursos de tabla

    ```
    aws glue create-database \
      --cli-input-json \
          '{
              "CatalogId": "111122223333:linkcontainer",
              "DatabaseInput": {
                  "Name": "db1",
                  "Description": "creating parent database for table link"
              }
          }'
    ```

  - Ejemplo de creación de enlaces de recursos de tabla

    ```
    aws glue create-table \
      --cli-input-json \
    ```

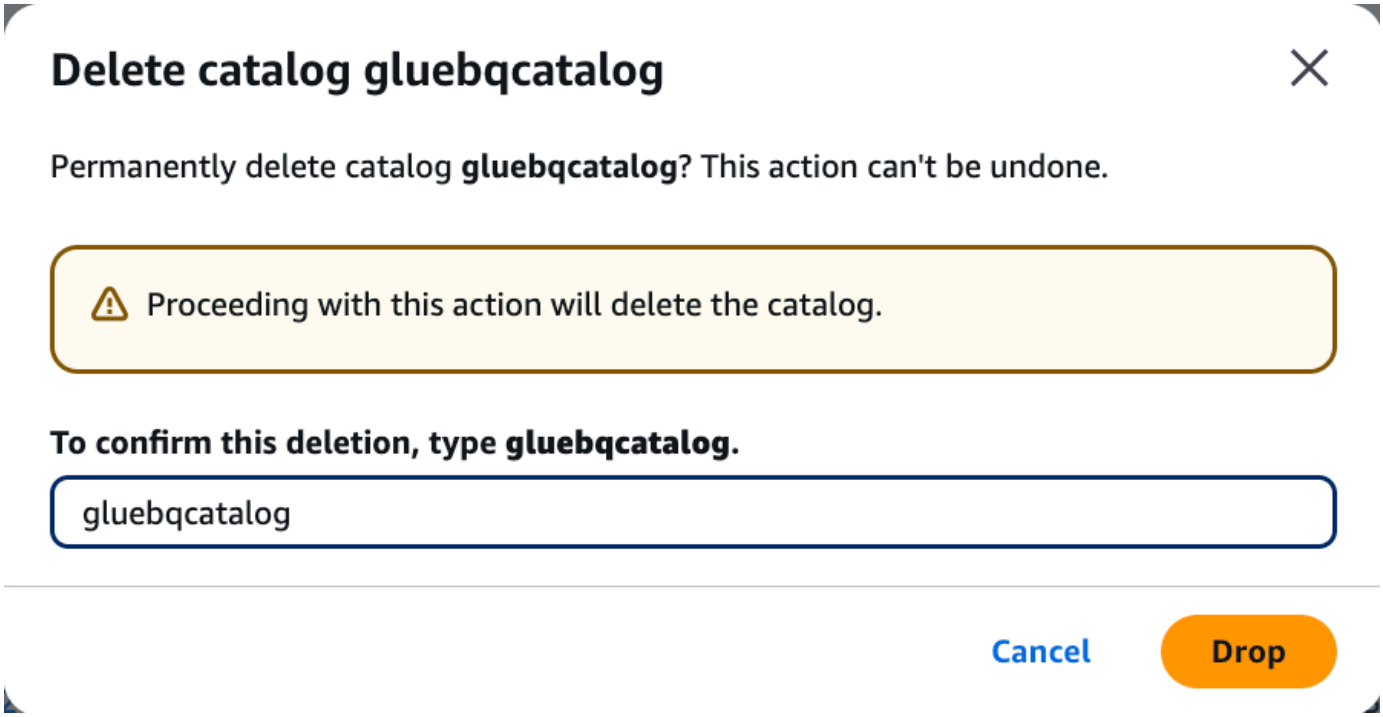
```
    '{
        "CatalogId": "111122223333:linkcontainer",
         "DatabaseName": "db1",
        "TableInput": {
            "Name": "tablelink",
            "TargetTable": {
                "CatalogId": "123456789012:nscatalog/dev",
               "DatabaseName": "schema1",
                "Name": "table1"
             }
        }
    }'
```

Eliminación de un catálogo federado

Puede eliminar los catálogos federados que creó en la `glue:DeleteCatalog` operación o la AWS Glue Data Catalog AWS Lake Formation consola.

Cómo eliminar un catálogo federado (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación, elija Catálogos en Catálogo de datos.
3. Elija el catálogo que desee eliminar de la lista de catálogos.
4. En Acciones, elija Eliminar.
5. Elija Borrar para confirmar y el catálogo federado se eliminará del Catálogo de datos.



Cómo eliminar un catálogo federado (CLI)

- ```
  aws glue delete-catalog
   --catalog-id 123456789012:catalog name
  ```

Consulta de catálogos federados

Tras conceder permisos a otros responsables, estos pueden iniciar sesión y empezar a consultar las tablas de los catálogos federados iniciando sesión en las herramientas de SQL mediante Amazon Redshift, Amazon EMR y ETL. Amazon Athena AWS Glue

Para obtener más información sobre cómo conectarse al punto final de la AWS Glue Data Catalog extensión Rest de Apache Iceberg o a la aplicación Spark independiente, consulte Cómo acceder a esta sección de la Guía para desarrolladores. AWS Glue Data Catalog AWS Glue

Puede usar consultas de lenguaje de definición de datos (DDL) para crear y administrar las tablas de la base de datos mediante Apache Spark en Amazon EMR. Para crear y eliminar tablas en la base de datos de Amazon Redshift, la entidad principal debe tener los permisos `Create table` y `Drop` de Lake Formation.
  ```

Para obtener más información sobre la concesión de permisos del Catálogo de datos, consulte Concesión de permisos sobre los recursos del Catálogo de datos.

Para obtener más información sobre cómo consultar los recursos del catálogo Amazon Athena, consulte Consultas AWS Glue Data Catalog desde la Guía del usuario Amazon Athena de Amazon Athena.

Recursos adicionales

Puede usar Amazon SageMaker Lakehouse para lograr un acceso unificado a los datos tanto en los almacenes de datos como en los lagos de datos. A través de SageMaker Lakehouse, puede utilizar los motores de análisis, aprendizaje automático e inteligencia empresarial preferidos a través de una API REST abierta de Apache Iceberg para garantizar el acceso seguro a los datos con controles de acceso consistentes y detallados.

- SageMaker Taller de Amazon
- Simplifique el acceso a los datos de su empresa con Amazon SageMaker Lakehouse

Federación en fuentes de datos externas en el AWS Glue Data Catalog

Puede conectar el AWS Glue Data Catalog (catálogo de datos) a almacenes de datos como Amazon Redshift o Snowflake, bases de datos en la nube como Amazon RDS Amazon DynamoDB, Oracle y servicios de streaming como Amazon MSK, y sistemas locales como Teradata mediante conexiones. AWS Glue Estas conexiones se almacenan AWS Glue Data Catalog y se registran en él AWS Lake Formation, lo que le permite crear un catálogo federado para cada fuente de datos disponible.

Un catálogo federado es un contenedor de nivel superior que apunta a una base de datos de un sistema de datos externo. Le permite consultar los datos directamente desde el sistema de datos externo sin necesidad de procesos de extracción, transformación y carga (ETL).

Para obtener más información sobre AWS Glue las conexiones, consulte Conexión a los datos en la Guía para AWS Glue desarrolladores.

Los administradores de lagos de datos pueden crear catálogos federados mediante Amazon SageMaker Lakehouse o. Amazon Athena

Luego, los administradores de lagos de datos pueden conceder permisos detallados sobre los objetos del catálogo mediante Lake Formation, controlando el acceso en varios niveles, como el

catálogo, la base de datos, la tabla, la columna, la fila o la celda. Los analistas de datos pueden descubrir y consultar los orígenes de datos catalogados con Athena, y Lake Formation aplica las políticas de acceso definidas. Los analistas pueden unir datos de varios orígenes en una sola consulta sin necesidad de conectarse a cada origen de forma individual.

Temas

- Flujo de trabajo
- Requisitos previos para conectar el Catálogo de datos a orígenes de datos externos
- Crear un catálogo federado mediante un AWS Glue conexión
- Visualización de objetos del catálogo
- Eliminación de un catálogo federado
- Consulta de catálogos federados
- Recursos adicionales

Flujo de trabajo

Un administrador de un lago de datos o un usuario con los permisos necesarios debe completar los siguientes pasos para conectarlo AWS Glue Data Catalog a una fuente de datos externa.

1. Crea una AWS Glue conexión con la fuente de datos. Al registrar la conexión, el rol de IAM utilizado para registrar la conexión debe tener acceso a la función de Lambda y a la ubicación del bucket de desbordamiento de Amazon S3.
2. Registra la conexión con Lake Formation.
3. Crea un catálogo federado en el catálogo de datos mediante una AWS Glue conexión para conectarse a las fuentes de datos disponibles. Las bases de datos, tablas y vistas se catalogan automáticamente en el Catálogo de datos y se registran en Lake Formation.
4. Otorga acceso a catálogos, bases de datos y tablas específicos a los analistas de datos mediante los permisos de Lake Formation. Fine-grained Las políticas de control de acceso se pueden definir en lagos de datos, almacenes y fuentes OLTP mediante Lake Formation, que permite filtros de seguridad a nivel de fila y columna.

   Luego, los analistas de datos pueden acceder a todos los datos a través del catálogo de datos mediante consultas SQL en Athena, sin necesidad de conexiones independientes ni credenciales de orígenes de datos. Los analistas pueden ejecutar consultas SQL federadas que escanean datos de múltiples orígenes y unirlos in situ sin canalizaciones de datos complejas.

Requisitos previos para conectar el Catálogo de datos a orígenes de datos externos

Para conectarse AWS Glue Data Catalog a fuentes de datos externas, registrar la conexión en Lake Formation y configurar los catálogos federados, debe cumplir los siguientes requisitos:

> **Note**
>
> Recomendamos que un administrador del lago de datos de Lake Formation cree las AWS Glue conexiones para conectarse a fuentes de datos externas y cree los catálogos federados.

1. Crear roles de IAM.

   - Cree un rol que tenga los permisos necesarios para implementar los recursos (la función Lambda, el depósito de contenido de Amazon S3, el rol de IAM y la AWS Glue conexión) necesarios para crear una conexión a la fuente de datos externa.
   - Cree un rol que tenga los permisos mínimos necesarios para acceder a las propiedades de la AWS Glue conexión (la función Lambda y el spill bucket de Amazon S3). Este es el rol que incluirá cuando registre la conexión con Lake Formation.

     Para utilizar Lake Formation para gestionar y proteger los datos de su lago de datos, debe registrar la AWS Glue conexión en Lake Formation. Al hacerlo, Lake Formation puede expedir credenciales a Amazon Athena para consultar las orígenes de datos federados.

     El rol debe tener los permisos `Select` o `Describe` en el bucket de Amazon S3 y la función de Lambda.

     - s3: ListBucket
     - s3: GetObject
     - lambda: InvokeFunction

     JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
```

```
    {
      "Effect": "Allow",
      "Action": [
        "s3:*"
      ],
      "Resource": [
        "arn:aws:s3:::amzn-s3-demo-bucket1/object/*",
        "arn:aws:s3:::amzn-s3-demo-bucket1/object"
      ]
    },
    {
      "Sid": "lambdainvoke",
      "Effect": "Allow",
      "Action": "lambda:InvokeFunction",
      "Resource": "arn:aws:lambda:us-east-1:123456789012:function:example-
lambda-function"
    },
    {
      "Sid": "gluepolicy",
      "Effect": "Allow",
      "Action": "glue:*",
      "Resource": "*"
    }
  ]
}
```

- Agregue la siguiente política de confianza al rol de IAM que se utiliza para registrar la conexión:

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": [
                    "lakeformation.amazonaws.com"
              ]
            },
            "Action": "sts:AssumeRole"
        }
```

```
    ]
}
```

- El administrador del lago de datos que registra la ubicación debe tener el permiso `iam:PassRole` para el rol.

  La siguiente es una política insertada que concede este permiso. *<account-id>*Sustitúyalo por un número de AWS cuenta válido y *<role-name>* sustitúyalo por el nombre del rol.

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "PassRolePermissions",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": [
                "arn:aws:iam::111122223333:role/example-role-name>"
            ]
        }
    ]
}
```

- Para crear catálogos federados en el Catálogo de datos, a fin de asegurarse de que el rol de IAM que está utilizando sea el de administrador del lago de datos de Lake Formation, compruebe la configuración del lago de datos (`aws lakeformation get-data-lake-settings`).

  Si no es administrador de un lago de datos, necesita el permiso `CREATE_CATALOG` de Lake Formation para crear un catálogo. En el ejemplo siguiente, se muestra cómo conceder los permisos necesarios para crear catálogos.

```
aws lakeformation grant-permissions \
--cli-input-json \
        '{
            "Principal": {
```

```
             "DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/non-
admin"
            },
            "Resource": {
                "Catalog": {
                }
            },
            "Permissions": [
                "CREATE_CATALOG",
                "DESCRIBE"
            ]
        }'
```

Además del `CREATE_CATALOG` permiso, necesita `DATA_LOCATION_ACCESS` la conexión registrada para crear un catálogo federado. El administrador de un lago de datos tiene permisos de ubicación de datos implícitos y no necesita esta concesión. Para obtener más información, consulte Permisos implícitos de Lake Formation.

En el siguiente ejemplo, se muestra cómo conceder `DATA_LOCATION_ACCESS` en una conexión registrada.

```
aws lakeformation grant-permissions \
  --cli-input-json \
    '{
      "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-admin-
role"
      },
      "Resource": {
        "DataLocation": {
          "CatalogId": "123456789012",
          "ResourceArn": "arn:aws:glue:us-east-1:123456789012:connection/
connection-name"
        }
      },
      "Permissions": ["DATA_LOCATION_ACCESS"]
    }'
```

2. Agregue la siguiente política de claves a la AWS KMS clave si utiliza una clave administrada por el cliente para cifrar los datos de la fuente de datos. Sustituya el número de cuenta por un

número de AWS cuenta válido y especifique el nombre de la función. De forma predeterminada, los datos se cifran mediante una clave de KMS. Lake Formation ofrece una opción para crear la clave de KMS personalizada para el cifrado. Si utiliza una clave administrada por el cliente, debe agregar políticas de claves específicas a la clave.

Para obtener más información sobre la administración de los permisos de una clave administrada por el cliente, consulte Claves administradas por el cliente.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:DescribeKey"
            ],
            "Resource": "arn:aws:kms:us-east-1:123456789012:key/key-1"
        }
    ]
}
```

Crear un catálogo federado mediante un AWS Glue conexión

Para conectarse AWS Glue Data Catalog a fuentes de datos externas, debe usar AWS Glue conexiones que permitan la comunicación con las fuentes de datos externas. Puede crear AWS Glue conexiones mediante la AWS Glue consola, la API Create connection y la consola Amazon SageMaker Lakehouse.

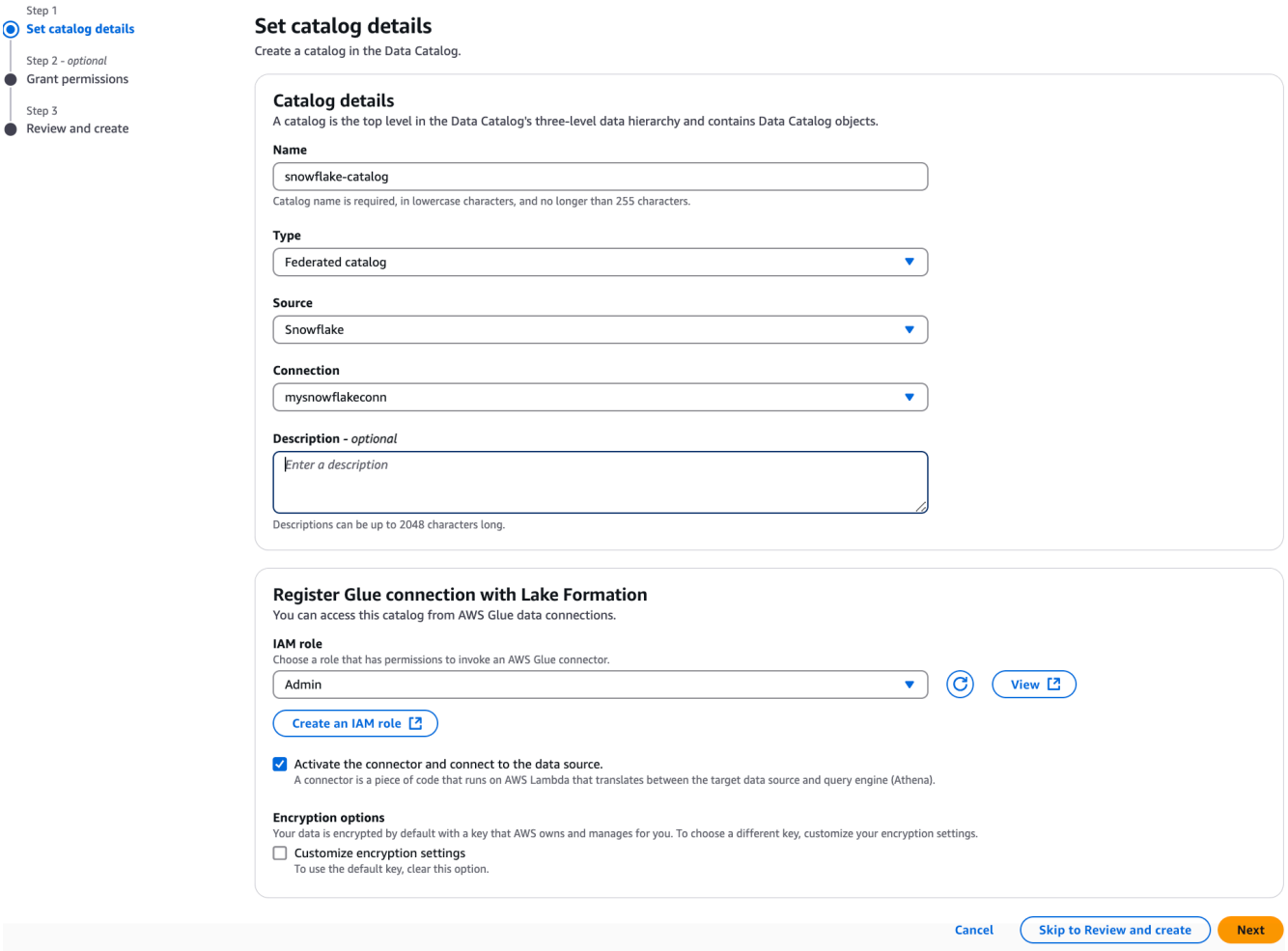
Para obtener instrucciones paso a paso para crear una AWS Glue conexión, consulte Conexión a datos en la Guía para AWS Glue desarrolladores o Creación de conexiones en Amazon SageMaker Lakehouse.

Cuando un usuario ejecuta una consulta en tablas federadas, Lake Formation vende credenciales que invocan una AWS Lambda función especificada en la AWS Glue conexión para recuperar objetos de metadatos de la fuente de datos.

Consola de administración de AWS

Cómo crear un catálogo federado desde un origen de datos externo y configurar los permisos (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación, elija Catálogos en Catálogo de datos.
3. Seleccione la opción Crear catálogo.
4. En la página de detalles Establecer catálogo, facilite la siguiente información:



- Nombre: un nombre exclusivo para el catálogo federado. El nombre no se puede cambiar y debe ir en minúsculas. El nombre puede tener un máximo de 255 caracteres.

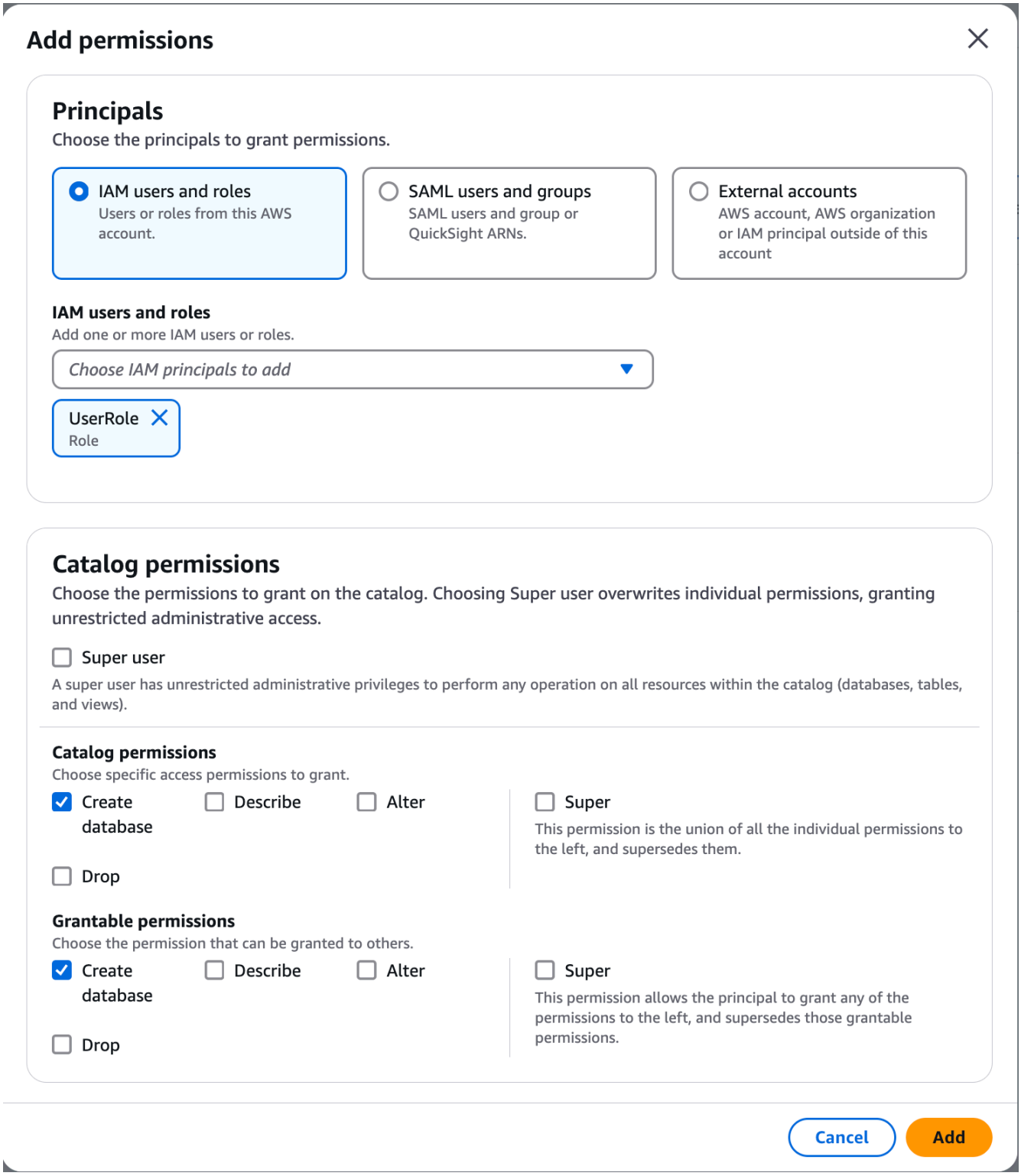
- Tipo: elija catálogo federado como tipo de catálogo.
- Origen: elija un origen de datos en la lista desplegable. Se muestran los orígenes de datos para los que ha creado conexiones. Para obtener más información sobre cómo crear una AWS Glue conexión a una fuente de datos externa, consulte Creación de conexiones para conectores en la Guía para AWS Glue desarrolladores o Creación de conexiones en Amazon SageMaker Lakehouse.
- Conexión: elija una AWS Glue conexión existente a la fuente de datos.
- Descripción: escriba una descripción para el catálogo creado desde el origen de datos.

5. Elija un rol de IAM para que Lake Formation asuma la venta de credenciales para que el motor de consultas acceda a los datos del origen de datos. Esta función debe tener los permisos necesarios para acceder a la AWS Glue conexión e invocar la función Lambda para acceder a los datos de la fuente de datos externa.

   También puede crear un rol nuevo en la consola de IAM.

   Consulte los permisos necesarios en la sección Requisitos previos para conectar el Catálogo de datos a orígenes de datos externos.
6. Seleccione la opción Activar el conector para conectarse al origen de datos para permitir que Athena ejecute consultas federadas.

   Para ver la lista de conectores compatibles, consulte Registro de la conexión en la Guía del usuario de Amazon Athena.
7. Opciones de cifrado: elija la opción Personalizar la configuración de cifrado si desea utilizar una clave personalizada para cifrar el catálogo. Para usar una clave personalizada, debe agregar una política de claves administradas personalizadas adicionales a la clave de KMS.
8. Seleccione Siguiente para conceder permisos a otras entidades principales.
9. En la página Concesión de permisos, elija Agregar permisos.
10. En la pantalla Agregar permisos, elija las entidades principales y los tipos de permisos que desea conceder.



- En la sección Entidades principales, elija uno de los tipos y, a continuación, especifique las que van a recibir los permisos concedidos.

- Usuarios y roles de IAM: elija uno o varios usuarios o roles de la lista de usuarios y roles de IAM.
- Usuarios y grupos de SAML: para SAML y Amazon Quick usuarios y grupos, introduzca uno o más nombres de recursos de Amazon (ARN) para los usuarios o grupos federados a través de SAML, o ARN para los usuarios o grupos de Amazon Quick. Pulse Intro después de cada ARN.

- En la sección Permisos, seleccione los permisos y los permisos concedibles.

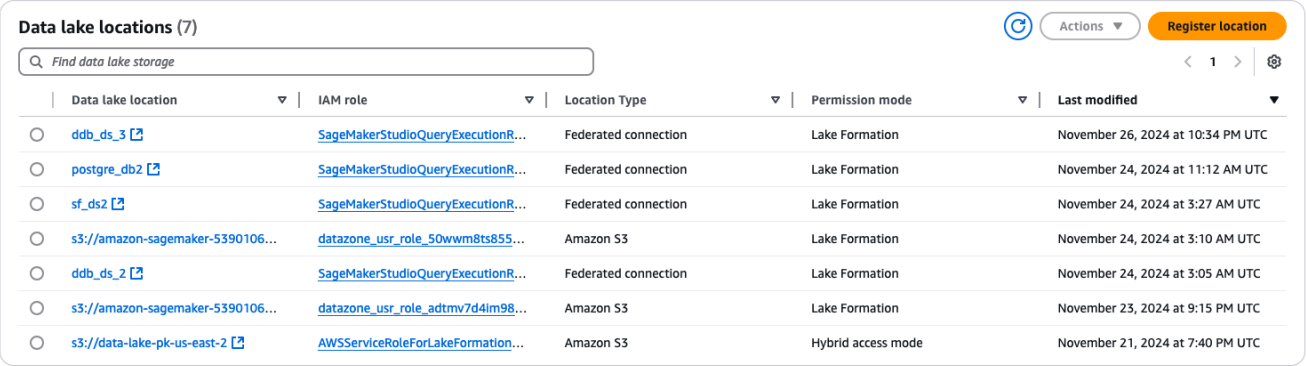
  En Permisos de catálogo, seleccione uno o más permisos para conceder.

  Elija Superusuario para adjudicar permisos administrativos sin restricciones en todos los recursos del catálogo.

  En Permisos concedibles, seleccione los permisos que el destinatario de la subvención puede conceder a otras entidades principales de su cuenta. AWS Esta opción no es compatible cuando se conceden permisos a una entidad principal de IAM desde una cuenta externa.

11. Elija Siguiente para revisar la información y crear el catálogo. La lista Catálogos muestra el nuevo catálogo federado.

    La lista Ubicaciones de datos muestra la conexión federada recién registrada.



AWS CLI

Cómo crear un catálogo federado desde un origen de datos externo y configurar los permisos

1. En el siguiente ejemplo, se muestra cómo crear una conexión. AWS Glue

```
aws glue create-connection
```

```
  --connection-input \
      '{
         "Name": "DynamoDB connection",
         "ConnectionType": "DYNAMODB",
         "Description": "A connection created for DynamoDB",
         "ConnectionProperties": {},
         "AthenaProperties": "spill_prefix": "your_spill_prefix",
         "lambda_function_arn": "Lambda_function_arn",
         "spill_bucket": "Your_Bucket_name",
         "AuthenticationConfiguration": {}
      }'
```

2. El siguiente ejemplo muestra cómo registrar una AWS Glue conexión con Lake Formation.

```
aws lakeformation register-resource
  --cli-input-json \
    {"ResourceArn":"arn:aws:glue:us-east-1:123456789012:connection/
dynamo","RoleArn":"arn:aws:iam::123456789012:role/
AdminTelemetry","WithFederation":true}
```

3. Si no es administrador de un lago de datos, conceda `DATA_LOCATION_ACCESS` la conexión registrada a la entidad principal que crea el catálogo federado. El administrador de un lago de datos tiene permisos de ubicación de datos implícitos y no necesita esta concesión. Para obtener más información, consulte Permisos implícitos de Lake Formation.

   En el siguiente ejemplo, se muestra cómo conceder `DATA_LOCATION_ACCESS` en una conexión registrada.

```
aws lakeformation grant-permissions \
  --cli-input-json \
    '{
      "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-
admin-role"
      },
      "Resource": {
        "DataLocation": {
          "CatalogId": "123456789012",
          "ResourceArn": "arn:aws:glue:us-east-1:123456789012:connection/
connection-name"
        }
```

```
      },
      "Permissions": ["DATA_LOCATION_ACCESS"]
    }'
```

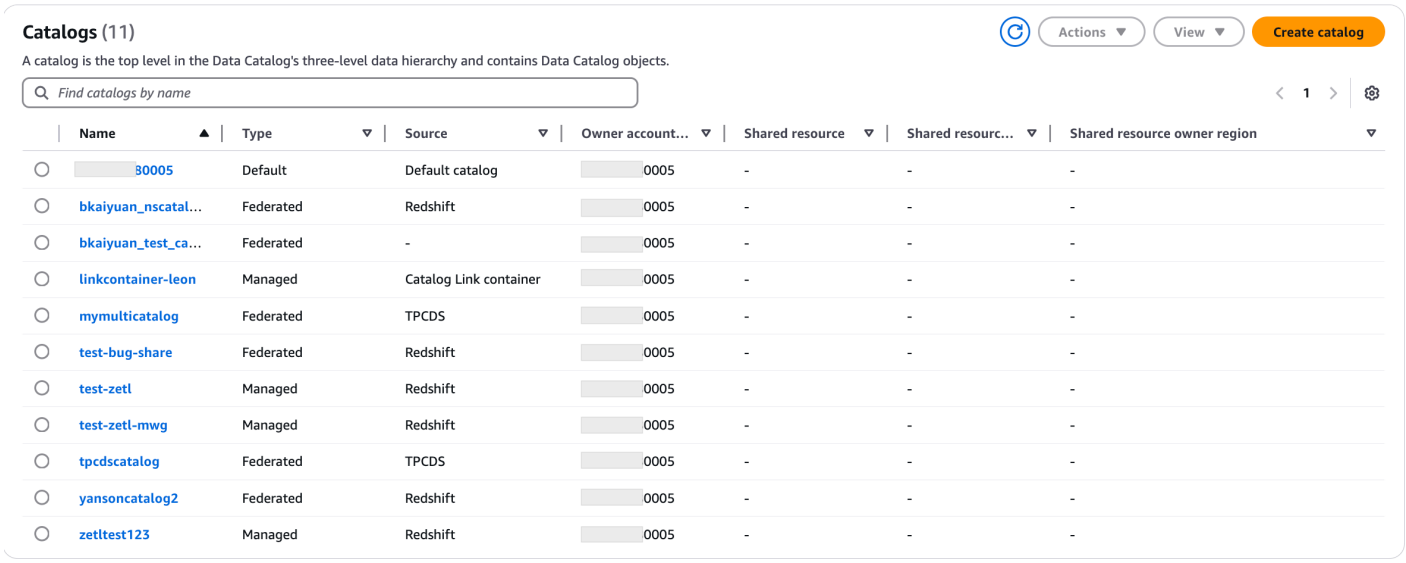
4. En el siguiente ejemplo, se muestra cómo crear un catálogo federado.

```
aws glue create-catalog
 --cli-input-json \
      '{
       "Name":"ddbcatalog",
       "CatalogInput":{"CatalogProperties":{"DataLakeAccessProperties":
{"DataTransferRole":"arn:aws:iam::123456789012:role/role name"}},
       "CreateDatabaseDefaultPermissions":[],
       "CreateTableDefaultPermissions":[],
       "FederatedCatalog":{"ConnectionName":"dynamo","Identifier":"dynamo"}
         }
       }'
```

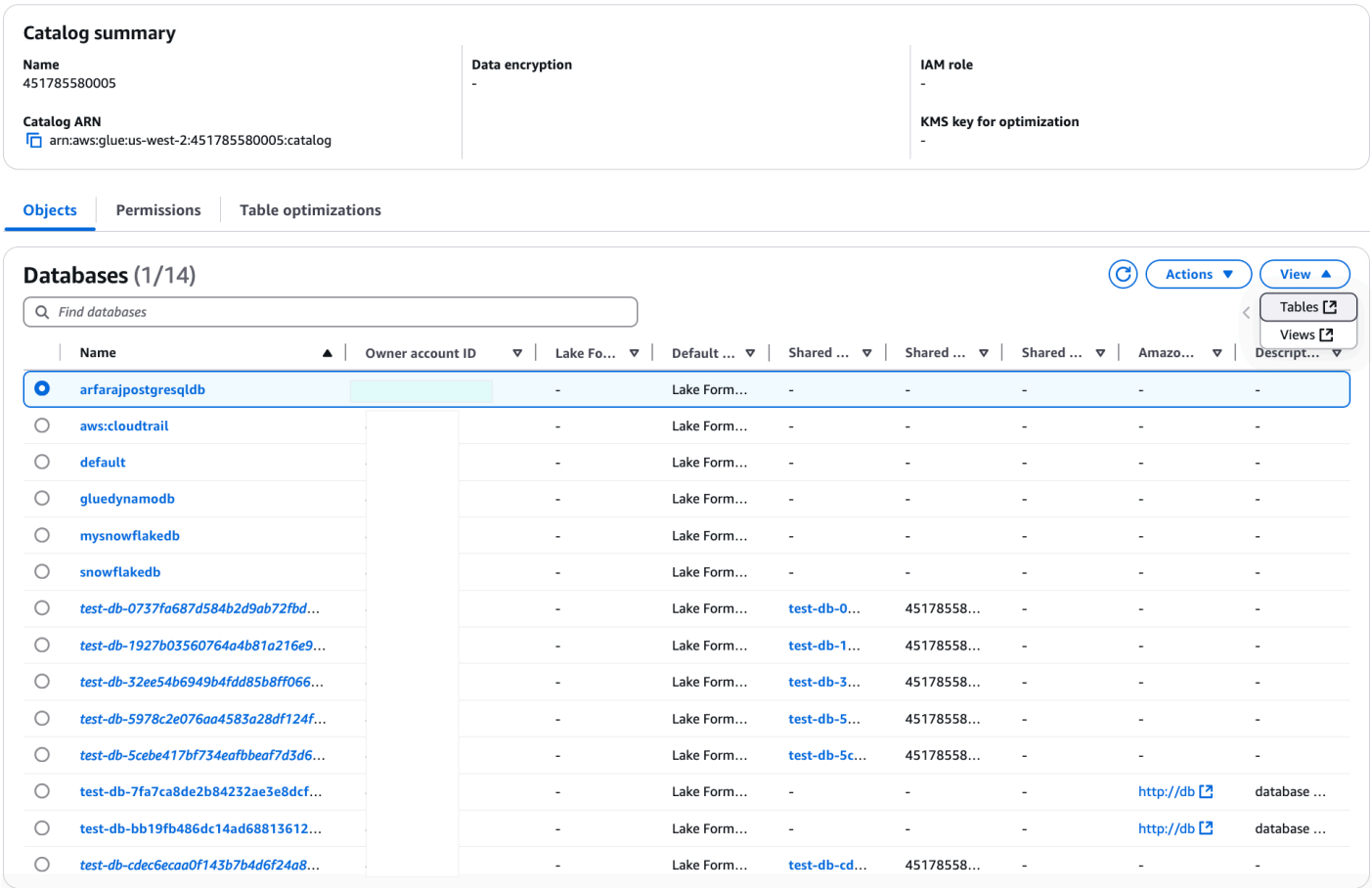
Visualización de objetos del catálogo

Para cada fuente de datos disponible, AWS Glue crea el catálogo correspondiente en AWS Glue Data Catalog. Tras crear un catálogo, puede ver las bases de datos y las tablas del catálogo mediante la consola de Lake Formation o la AWS CLI. Para

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. Seleccione Catálogos en Catálogo de datos. La página de catálogos muestra los catálogos sobre los que tiene permisos.



3. Elija un catálogo de la lista para ver las bases de datos y las tablas que contiene. La lista también contiene las bases de datos de la cuenta y los enlaces a los recursos, que son enlaces a bases de datos y tablas compartidas en cuentas externas y se utilizan para el acceso entre cuentas a los datos del lago de datos.



4. Elija la opción Tablas en Ver para ver y administrar las tablas de la base de datos.

AWS CLI ejemplos de visualización de catálogos y bases de datos

El siguiente ejemplo muestra cómo ver un catálogo mediante AWS CLI

```
aws glue get-catalog \
--catalog-id 123456789012:dynamodbcatalog
```

En el siguiente ejemplo, se muestra cómo solicitar todos los catálogos de la cuenta.

```
aws glue get-catalogs \
 --recursive
```

La siguiente solicitud de ejemplo muestra cómo obtener las bases de datos del catálogo.

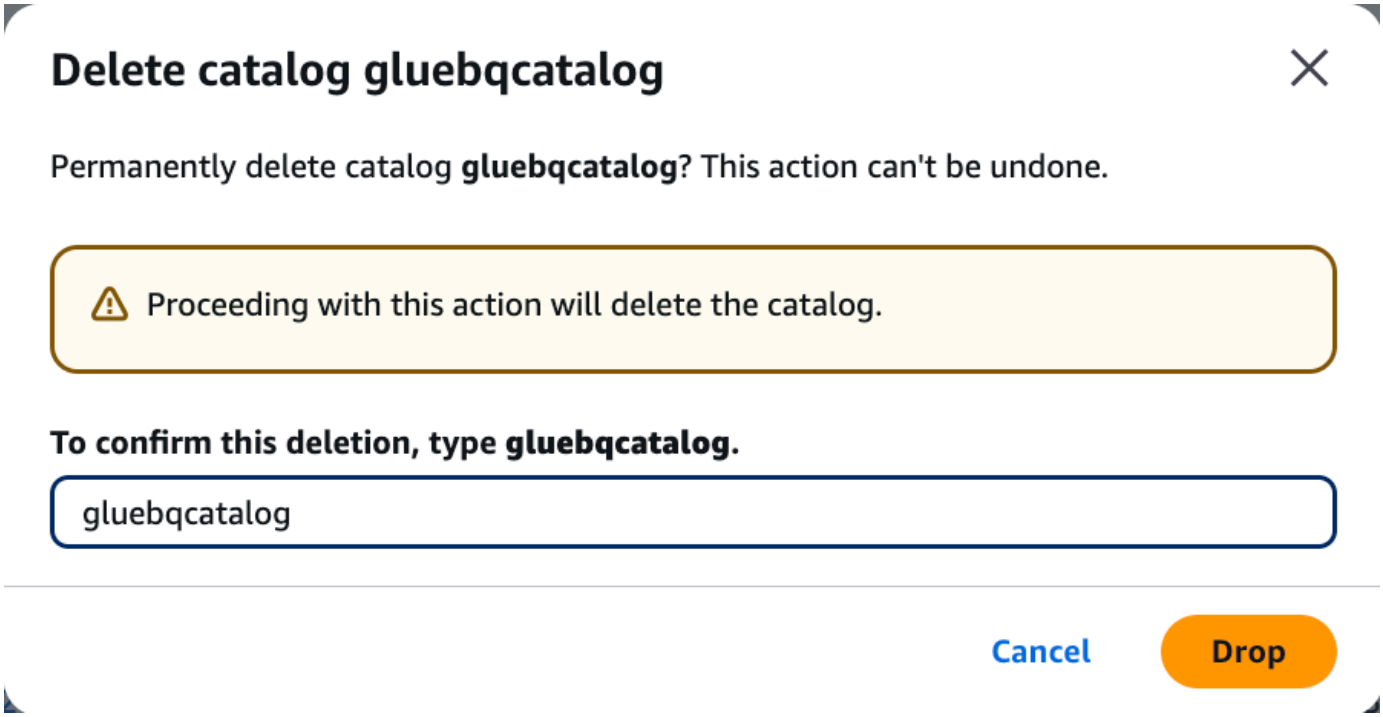
```
aws glue get-database \
--catalog-id 123456789012:dynamodbcatalog
--database-name database name
```

Eliminación de un catálogo federado

Puede eliminar los catálogos federados que creó en la `glue:DeleteCatalog` operación o la AWS Glue Data Catalog AWS Lake Formation consola.

Cómo eliminar un catálogo federado (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación, elija Catálogos en Catálogo de datos.
3. Elija el catálogo que desee eliminar de la lista de catálogos.
4. En Acciones, elija Eliminar.
5. Elija Borrar para confirmar y el catálogo federado se eliminará del Catálogo de datos.



Cómo eliminar un catálogo federado (CLI)

- ```
  aws glue delete-catalog
    --catalog-id 123456789012:catalog name
  ```

Consulta de catálogos federados

Tras conceder permisos a otras entidades principales, pueden iniciar sesión y empezar a consultar las tablas de los catálogos federados mediante Athena.

Para crear y eliminar tablas en la base de datos federada, la entidad principal debe tener los permisos `Create table` y `Drop` de Lake Formation.

Para obtener más información sobre la concesión de permisos del Catálogo de datos, consulte Concesión de permisos sobre los recursos del Catálogo de datos.

Para obtener más información sobre cómo consultar el catálogo de datos Amazon Athena, consulte Consultas AWS Glue Data Catalog desde la Guía del usuario Amazon Athena de Amazon Athena.
  ```

Recursos adicionales

En esta entrada de blog, mostramos cómo los analistas de datos ahora pueden acceder y consultar de forma segura los datos almacenados fuera de los lagos de datos de S3, incluidos los almacenes de datos y las Amazon DynamoDB bases de datos de Amazon Redshift, todo ello a través de una experiencia única y unificada. Los administradores ahora pueden aplicar controles de acceso con distintos niveles de granularidad para garantizar que la información confidencial permanezca protegida y, al mismo tiempo, ampliar el acceso a los datos. Esto permite a las organizaciones acelerar las iniciativas de datos y, al mismo tiempo, mantener la seguridad y la conformidad, lo que permite tomar decisiones más rápidas y basadas en los datos.

- Cataloga y gobierna las consultas federadas de Amazon Athena con Amazon Lakehouse SageMaker

Integración de Amazon S3 Tables con AWS Glue Data Catalog and AWS Lake Formation

Las Tablas de Amazon S3 proporcionan almacenamiento S3 optimizado específicamente para cargas de trabajo de análisis, lo que mejora el rendimiento de las consultas y reduce los costos. Los datos de Tablas de S3 se almacenan en un nuevo tipo de bucket: un bucket de tablas, que almacena tablas como subrecursos. Las tablas de S3 incluyen compatibilidad integrada con el estándar Apache Iceberg, que permite consultar fácilmente datos tabulares en buckets de tablas de Amazon S3 mediante motores de consulta populares como Apache Spark.

Puede integrar Amazon S3 Tables AWS Glue Data Catalog mediante controles de acceso de IAM o con subvenciones de IAM y Lake Formation:

- Control de acceso de IAM: utiliza políticas de IAM para controlar el acceso a las tablas y al catálogo de datos de S3. En este enfoque de control de acceso, se necesitan permisos de IAM tanto en los recursos de S3 Tables como en los objetos del catálogo de datos para acceder a los recursos.
- Control de acceso a Lake Formation: utiliza AWS Lake Formation concesiones además de los permisos de AWS Glue IAM para controlar el acceso a las tablas S3 a través del catálogo de datos. En este modo, los directores requieren permisos de IAM para interactuar con el catálogo de datos, y las concesiones de Lake Formation determinan a qué recursos del catálogo (bases de datos, tablas, columnas, filas) puede acceder el director. Este modo admite tanto el control de acceso amplio (concesiones a nivel de base de datos y tabla) como el control de acceso detallado

(seguridad a nivel de columna y fila). Cuando se configura un rol registrado y se habilita la venta de credenciales, no se requieren permisos de IAM de S3 Tables para el director, ya que Lake Formation vende las credenciales en nombre del principal mediante el rol registrado. El control de acceso de Lake Formation también admite la venta de credenciales para motores de análisis de terceros.

En esta sección se proporciona una guía para configurar la integración en los siguientes escenarios: AWS Lake Formation

- Escenario A: Ha integrado las tablas y el catálogo de datos de S3 mediante los controles de acceso de IAM y ahora piensa AWS Lake Formation utilizarlos. Consulte Cambiar los controles de acceso para la integración de S3 Tables para obtener más información.
- Escenario B: Planea integrar las tablas y el catálogo de datos de S3 utilizando su cuenta AWS Lake Formation y región, y actualmente no los tiene integrados en su cuenta y región. Comience con la Requisitos previos para integrar el catálogo de Tablas de Amazon S3 con el Catálogo de datos y Lake Formation sección y sigaHabilitación de la integración de Tablas de Amazon S3.
- Escenario C: Ha integrado las tablas y el catálogo de datos de S3 mediante IAM AWS Lake Formation y ahora planea usarlo. Consulte Cambiar los controles de acceso para la integración de S3 Tables para obtener más información.

Asegúrese de seguir los pasos que se indican en la sección Integración de las tablas de S3 con los servicios de AWS análisis para disponer de los permisos adecuados para acceder a los recursos de la tabla AWS Glue Data Catalog y a los suyos, y para trabajar con los servicios de AWS análisis.

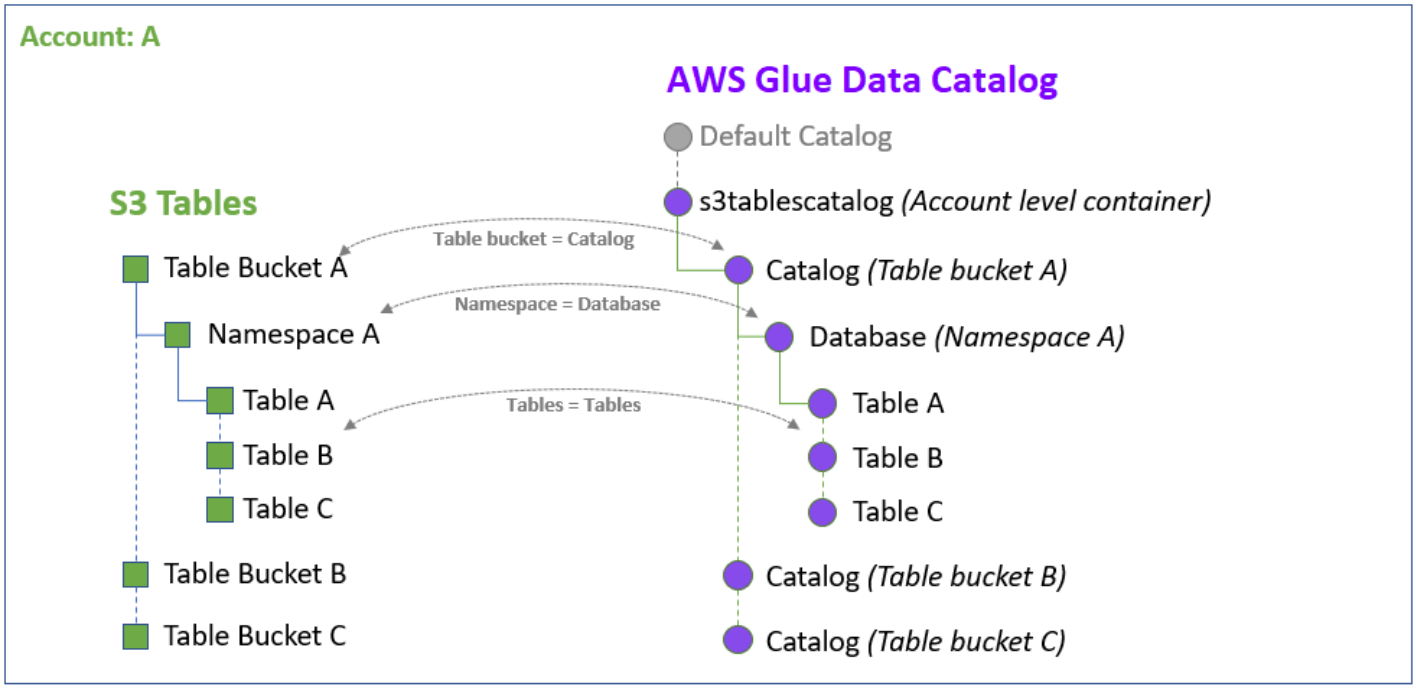
Temas

- Cómo funciona la integración del Catálogo de datos y Lake Formation
- Requisitos previos para integrar el catálogo de Tablas de Amazon S3 con el Catálogo de datos y Lake Formation
- Habilitación de la integración de Tablas de Amazon S3
- Creación de bases de datos y tablas en el catálogo de Tablas de S3
- Registrar un bucket de mesa de Amazon S3 en otro AWS inscrita
- Concesión de permisos

Cómo funciona la integración del Catálogo de datos y Lake Formation

Al integrar el catálogo de tablas de S3 con el Catálogo de datos y Lake Formation, el servicio AWS Glue crea un único catálogo federado denominado `s3tablescatalog` en el Catálogo de datos predeterminado de su cuenta específico de su Región de AWS. La integración mapea todos los recursos del bucket de tablas de Amazon S3 de su Región de AWS cuenta y del catálogo federado de la siguiente manera:

- Los buckets de Tablas de Amazon S3 se convierten en un catálogo de varios niveles en el Catálogo de datos.
- El espacio de nombres de Amazon S3 asociado está registrado como base de datos en el Catálogo de datos.
- Las Tablas de Amazon S3 del bucket de tablas se convierten en tablas del Catálogo de datos.



Tras la integración con Lake Formation, puede crear tablas de Apache Iceberg en el catálogo de cubos de tablas y acceder a ellas mediante motores de AWS análisis integrados, como Amazon Athena Amazon EMR, así como motores de análisis de terceros.

Cuando también habilitas Lake Formation con la integración, se posibilita un control de acceso detallado. AWS Lake Formation Este enfoque de seguridad significa que, además de los permisos

AWS Identity and Access Management (de IAM), debe conceder a su director de IAM con los permisos de Lake Formation en sus tablas antes de poder trabajar con ellos.

Hay dos tipos principales de permisos en AWS Lake Formation:

- Los permisos de acceso a metadatos controlan la capacidad de crear, leer, actualizar y eliminar bases de datos de metadatos y tablas del Catálogo de datos.
- Los permisos de acceso a los datos subyacentes controlan la capacidad de leer y escribir datos en las ubicaciones de Amazon S3 subyacentes a las que apuntan los recursos del Catálogo de datos.

Lake Formation utiliza una combinación de su propio modelo de permisos y el modelo de permisos de IAM para controlar el acceso a los recursos del Catálogo de datos y a los datos subyacentes.

- Para que una solicitud de acceso a los recursos del Catálogo de datos o a los datos subyacentes se atienda correctamente, debe pasar las comprobaciones de permisos tanto de IAM como de Lake Formation.
- Los permisos de IAM controlan el acceso a Lake Formation y a AWS Glue las API y los recursos, mientras que los permisos de Lake Formation controlan el acceso a los recursos del catálogo de datos, a las ubicaciones de Amazon S3 y a los datos subyacentes.

Los permisos de Lake Formation se aplican solo en la región en la que se concedieron, y la entidad principal debe estar autorizada por un administrador de lago de datos o de otra entidad principal con los permisos necesarios para que se le concedan permisos de Lake Formation.

Requisitos previos para integrar el catálogo de Tablas de Amazon S3 con el Catálogo de datos y Lake Formation

Los siguientes son los requisitos previos para habilitar la integración de tablas de Amazon S3 con AWS Glue Data Catalog y AWS Lake Formation.

1. Al habilitar la integración de Tablas de Amazon S3, Lake Formation registra automáticamente la ubicación de las tablas de S3. Para registrar la ubicación del cubo de la mesa en Lake Formation, necesita un IAM role/user con `lakeformation:RegisterResourcelakeformation:RegisterResourceWithPrivilegedAcc` y `lakeformation:CreateCatalog` permisos. Cuando un usuario no administrador con estos permisos registra una ubicación en el catálogo, Lake Formation le concede automáticamente el permiso `DATA_LOCATION_ACCESS` para esa ubicación, lo que permite a la entidad principal

que realiza la llamada llevar a cabo todas las operaciones de Lake Formation compatibles en la ubicación de datos registrada.

2. Al habilitar la integración de tablas de S3, debe elegir un rol de IAM para que Lake Formation provea credenciales que permitan el acceso a los datos. Cree un rol de IAM para el acceso de los datos de Lake Formation a los buckets de tablas de S3. El rol de IAM utilizado para registrar el bucket de tablas con Lake Formation requiere los permisos siguientes:

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "LakeFormationPermissionsForS3ListTableBucket",
            "Effect": "Allow",
            "Action": [
                "s3tables:ListTableBuckets"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3TableBucket",
            "Effect": "Allow",
            "Action": [
                "s3tables:CreateTableBucket",
                "s3tables:GetTableBucket",
                "s3tables:CreateNamespace",
                "s3tables:GetNamespace",
                "s3tables:ListNamespaces",
                "s3tables:DeleteNamespace",
                "s3tables:DeleteTableBucket",
                "s3tables:CreateTable",
                "s3tables:DeleteTable",
                "s3tables:GetTable",
                "s3tables:ListTables",
                "s3tables:RenameTable",
                "s3tables:UpdateTableMetadataLocation",
                "s3tables:GetTableMetadataLocation",
                "s3tables:GetTableData",
```

```
                "s3tables:PutTableData"
            ],
            "Resource": [
                "arn:aws:s3tables:us-east-1:123456789012:bucket/*"
            ]
        }
    ]
}
```

Para obtener más información, consulte Requisitos de los roles utilizados para registrar ubicaciones.

3. Agregue la siguiente política de confianza al rol de IAM para permitir que el servicio de Lake Formation asuma el rol y dispense credenciales temporales a los motores de análisis integrados.

```
{
  "Effect": "Allow",
  "Principal": {
    "Service": "lakeformation.amazonaws.com"
  },
  "Action": [
    "sts:AssumeRole",
    "sts:SetSourceIdentity",
    "sts:SetContext"  # add action to trust relationship when using IAM Identity
 center principals with Lake Formation
  ]
}
```

> **Note**
>
> Se ha actualizado el proceso de integración de los servicios de AWS análisis. Si ha configurado la integración con la versión preliminar, podrá seguir utilizando la integración actual. Sin embargo, el proceso de integración actualizado proporciona mejoras en el rendimiento. Para actualizar la integración:
>
> 1. En primer lugar, elimine su catálogo de tablas de S3 existente en Lake Formation. Para eliminar el catálogo, seleccione el catálogo de `s3tablescatalog` en la lista de catálogos y elija Eliminar en Acciones.
> 2. A continuación, anule el registro de la ubicación de datos de `s3tablescatalog`.

a. En la consola de Lake Formation, en la sección Administraciones, seleccione Ubicaciones de datos.

b. Seleccione una ubicación y, en el menú Acciones, seleccione Quitar.

c. Cuando se le indique que confirme, elija Quitar.

Para obtener instrucciones detalladas sobre cómo anular el registro de una ubicación de datos, consulte la Dar de baja el registro de una ubicación de Amazon S3 sección.

3. A continuación, siga los pasos de integración actualizados de la Habilitación de la integración de Tablas de Amazon S3 sección.

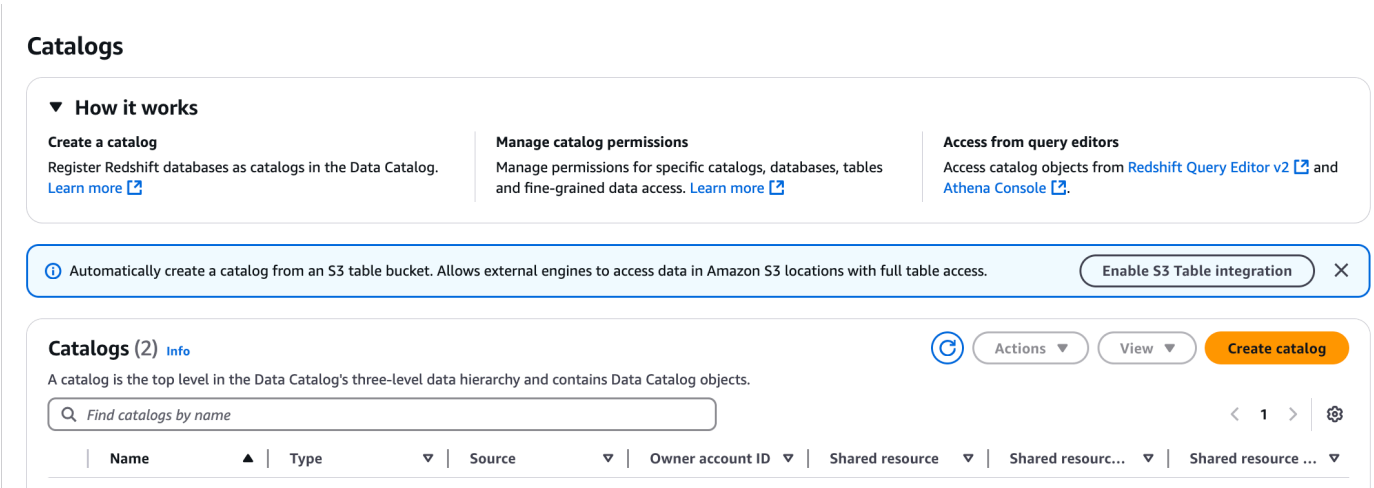
Habilitación de la integración de Tablas de Amazon S3

Puede crear buckets de tablas de Amazon S3 mediante la consola de Amazon S3 e integrarlos con los servicios de análisis de AWS . Para obtener más información, consulte Uso de Tablas de Amazon S3 con servicios de análisis de AWS.

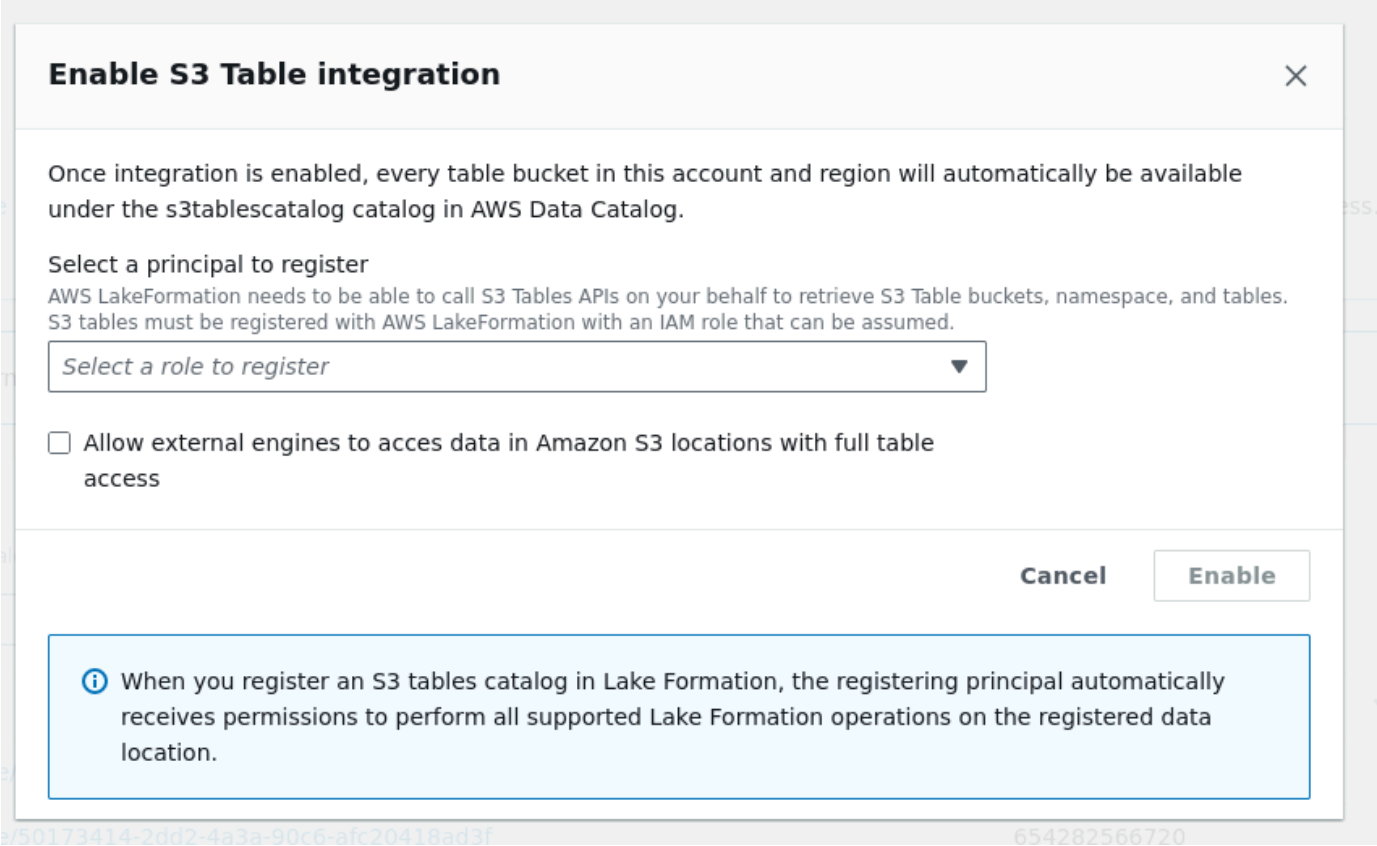
En AWS Lake Formation, puede habilitar la integración de Amazon S3 Tables con la consola Lake Formation AWS Glue Data Catalog y AWS Lake Formation utilizarla o utilizarla AWS CLI.

Cómo integrar Tablas de Amazon S3 con el Catálogo de datos y Lake Formation (consola)

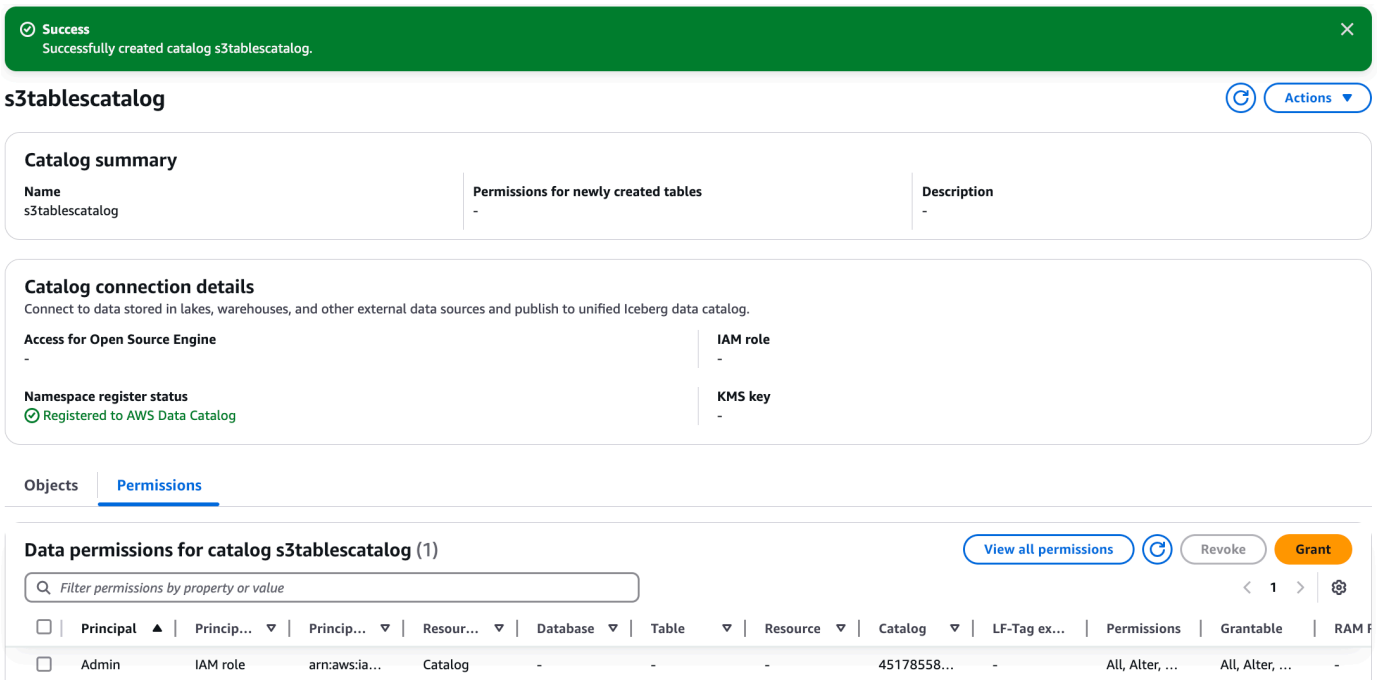
1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación, elija Catálogos en Catálogo de datos.
3. Elija Habilitar la integración de Tablas de S3 en la página Catálogos.



4. Elija un rol de IAM con los permisos necesarios para que Lake Formation asuma la provisión de credenciales a los motores de consultas analíticas. Para conocer los permisos necesarios para el acceso del rol a los datos, consulte step3-permissions en la sección de requisitos previos.



5. Seleccione la opción Permitir que motores externos accedan a datos en las ubicaciones de Amazon S3 con acceso total a las tablas. Al habilitar el acceso completo a las tablas para motores de terceros, Lake Formation devuelve las credenciales al motor externo directamente sin necesidad de validar la etiqueta de sesión de IAM. Esto significa que no puede aplicar controles de acceso detallados de Lake Formation a las tablas a las que se accede.
6. Seleccione Habilitar. El nuevo catálogo de Tablas de S3 se agrega a la lista de catálogos. Al habilitar la integración del catálogo de tablas de S3, el servicio registra la ubicación de los datos del bucket de tablas de S3 en Lake Formation.
7. Elija el catálogo para ver los objetos del catálogo y conceder permisos a otras entidades principales.



Para crear catálogos en varios niveles, consulte la sección Creación de un bucket de tablas en la Guía del usuario de Amazon Simple Storage Service.

Cómo integrar Tablas de Amazon S3 con el Catálogo de datos y Lake Formation (CLI)

Siguiendo la sección de requisitos previos, cree un rol de servicio de IAM que permita a Lake Formation acceder a los recursos de su tabla.

1. Cree un archivo llamado `Role-Trust-Policy.json` que contenga la siguiente política de confianza:

```
{
    "Version": "2012-10-17",
    "Statement": [
      {
        "Sid": "LakeFormationDataAccessPolicy",
        "Effect": "Allow",
        "Principal": {
          "Service": "lakeformation.amazonaws.com"
        },
        "Action": [
            "sts:AssumeRole",
            "sts:SetContext",
            "sts:SetSourceIdentity"
```

```
        ],
        "Condition": {
          "StringEquals": {
            "aws:SourceAccount": "111122223333"
          }
        }
      }
    ]
}
```

2. Cree el rol de servicio de IAM con el siguiente comando:

```
aws iam create-role \
  --role-name S3TablesRoleForLakeFormation \
  --assume-role-policy-document file://Role-Trust-Policy.json
```

3. Cree un archivo llamado `LF-GluePolicy.json` que contenga la siguiente política:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "LakeFormationPermissionsForS3ListTableBucket",
            "Effect": "Allow",
            "Action": [
                "s3tables:ListTableBuckets"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3TableBucket",
            "Effect": "Allow",
            "Action": [
                "s3tables:CreateTableBucket",
                "s3tables:GetTableBucket",
                "s3tables:CreateNamespace",
                "s3tables:GetNamespace",
                "s3tables:ListNamespaces",
                "s3tables:DeleteNamespace",
                "s3tables:DeleteTableBucket",
                "s3tables:CreateTable",
```

```
                "s3tables:DeleteTable",
                "s3tables:GetTable",
                "s3tables:ListTables",
                "s3tables:RenameTable",
                "s3tables:UpdateTableMetadataLocation",
                "s3tables:GetTableMetadataLocation",
                "s3tables:GetTableData",
                "s3tables:PutTableData"
            ],
            "Resource": [
                "arn:aws:s3tables:us-east-1:111122223333:bucket/*"
            ]
        }
    ]
}
```

4. Asocie la política al rol con el siguiente comando:

```
aws iam put-role-policy \
  --role-name S3TablesRoleForLakeFormation \
  --policy-name LakeFormationDataAccessPermissionsForS3TableBucket \
  --policy-document file://LF-GluePolicy.json
```

5. Cree un archivo llamado `input.json` que contenga lo siguiente:

```
{
    "ResourceArn": "arn:aws:s3tables:us-east-1:111122223333:bucket/*",
    "WithFederation": true,
    "RoleArn": "arn:aws:iam::111122223333:role/S3TablesRoleForLakeFormation"
}
```

6. Registre los buckets de tablas con Lake Formation utilizando el siguiente comando:

```
aws lakeformation register-resource \
  --region us-east-1 \
  --with-privileged-access \
  --cli-input-json file://input.json
```

7. Cree un archivo llamado `catalog.json` que contenga el siguiente catálogo:

```
{
   "Name": "s3tablescatalog",
   "CatalogInput": {
```

```
      "FederatedCatalog": {
          "Identifier": "arn:aws:s3tables:us-east-1:111122223333:bucket/*",
          "ConnectionName": "aws:s3tables"
       },
       "CreateDatabaseDefaultPermissions": [],
       "CreateTableDefaultPermissions": [],
       "AllowFullTableExternalDataAccess": "True"
   }
}
```

8. Cree el catálogo `s3tablescatalog` utilizando el siguiente comando. Al crear este catálogo, se rellena AWS Glue Data Catalog con los objetos correspondientes a los grupos de tablas, los espacios de nombres y las tablas.

```
aws glue create-catalog \
  --region us-east-1 \
  --cli-input-json file://catalog.json
```

9. Compruebe que el `s3tablescatalog` catálogo se haya agregado AWS Glue mediante el siguiente comando:

```
aws glue get-catalog --catalog-id s3tablescatalog
```

Creación de bases de datos y tablas en el catálogo de Tablas de S3

Puede crear bases de datos para organizar las tablas de Apache Iceberg y tablas para definir el esquema y la ubicación de los datos en el catálogo de tablas de S3.

Creación de una base de datos (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/e inicie sesión como administrador del lago de datos o creador de bases de datos.
2. En el panel de navegación, elija Bases de datos en Catálogo de datos.
3. Elija Creación de base de datos.
4. En la página Crear base de datos, elija la opción Base de datos y especifique los siguientes detalles:

   - Nombre: un nombre exclusivo para la base de datos.

- Catálogo de datos: elija el catálogo de tablas de S3. La base de datos residirá en este catálogo.
- Descripción: (opcional) agregue una descripción y una ubicación.
- Control de acceso de IAM para tablas nuevas: seleccione de forma opcional utilizar solo el control de acceso de IAM para las nuevas tablas de esta base de datos. Para obtener información sobre esta opción, consulte la sección Cambiar la configuración predeterminada del lago de datos.
- Elija Crear base de datos. Puede ver la base de datos creada en el catálogo de tablas de S3.

Cree una base de datos usando AWS CLI

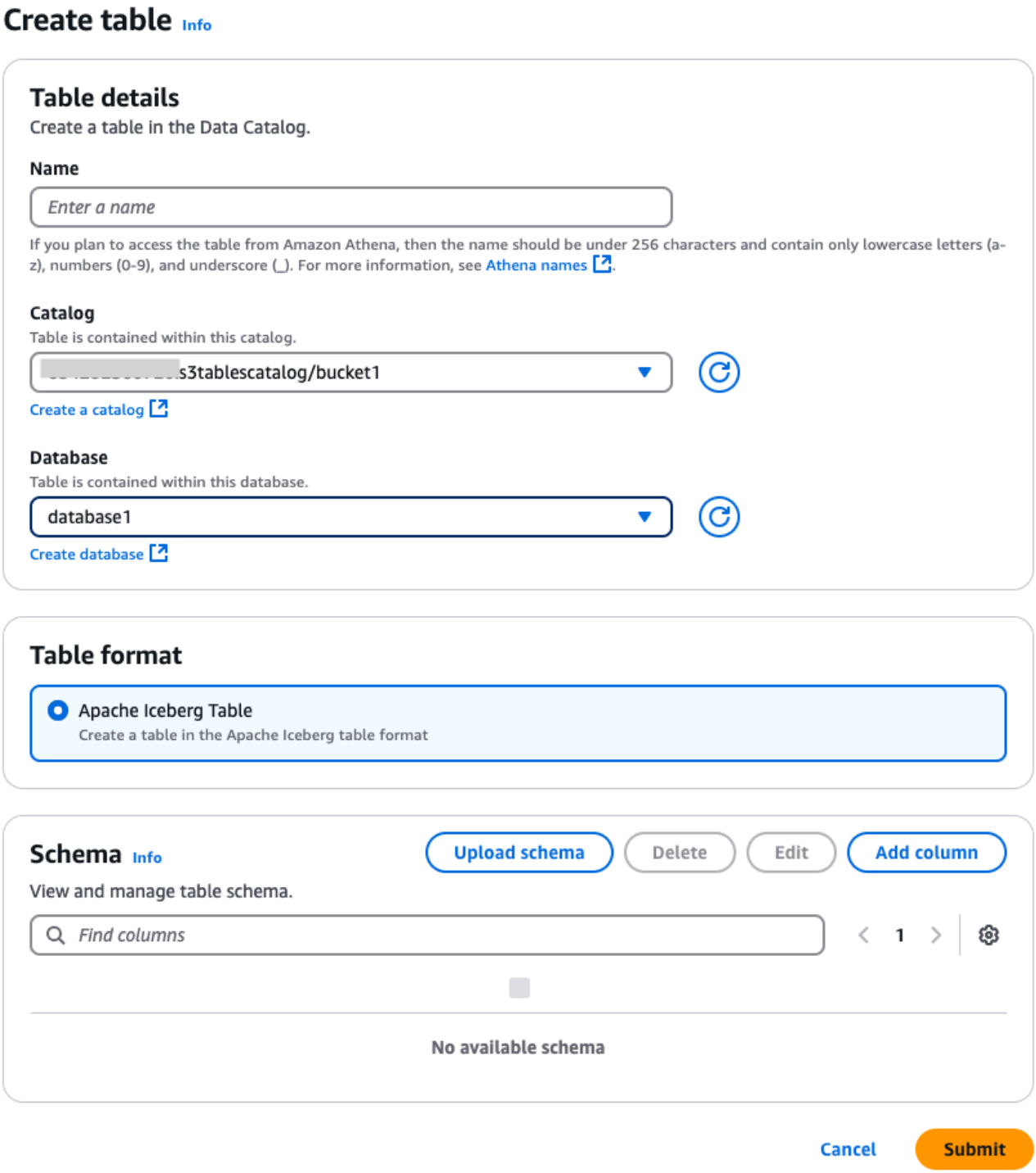
El siguiente comando de la CLI muestra cómo crear una base de datos en el catálogo de tablas de S3.

```
aws glue create-database
--region us-east-1 \
--catalog-id "123456789012:s3tablescatalog/test" \
--database-input \
 '{ "Name": "testglueclidbcreation" }'
```

Crear una tabla (Consola de administración de AWS)

Puede crear tablas de metadatos de Apache Iceberg en el catálogo de tablas de S3 mediante la consola Lake Formation o la AWS Glue `CreateTable` API.

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/e inicie sesión como administrador del lago de datos o como usuario con `CreateTable` permiso.
2. En el panel de navegación, en Catálogo de datos, elija Tablas.
3. Seleccione Create table (Creación de tabla).
4. En la página Crear tabla, especifique los detalles de la tabla.



- Nombre: escriba un nombre único para la tabla.
- Catálogo: elija el catálogo de tablas S3 como catálogo.
- Base de datos: elija la base de datos en el catálogo de tablas de S3.

- Descripción: especifique una descripción para la tabla.
- Esquema: elija Agregar columnas para agregar las columnas y los tipos de datos de las columnas. Tiene la opción de crear una tabla vacía y actualizar el esquema más adelante. Con Iceberg podrá desarrollar el esquema y la partición después de crear la tabla. Puede utilizar consultas de Athena para actualizar el esquema de la tabla y consultas de Spark para actualizar las particiones.

5. Seleccione Enviar.

Cree una tabla (AWS CLI)

```
aws glue create-table \
--database-name "testglueclidbcreation" \
--catalog-id "123456789012:s3tablescatalog/test" \
--region us-east-1 \
--table-input \
'{ "Name": "testtablegluecli", "Parameters": { "format": "ICEBERG" },
 "StorageDescriptor": { "Columns": [ {"Name": "x", "Type": "int", "Parameters":
 {"required": "true"}} ] } }'
```

Registrar un bucket de mesa de Amazon S3 en otro AWS inscrita

Puede registrar grupos de tablas individuales de Amazon S3 (formato ARN`arn:aws:s3tables:us-east-1:`*`account-id`*`:bucket/`*`bucket-name`*:) desde AWS una cuenta y Lake Formation en otra cuenta. Por ejemplo, puede registrar un bucket de tablas desde la cuenta A en Lake Formation de la cuenta B.

Requisitos previos

Antes de iniciar el registro del bucket en cuentas múltiples:

- Cree un bucket de tablas en la cuenta A.
- Cree un rol de IAM en la cuenta B con los permisos adecuados para el registro del bucket.

  Para obtener más información sobre los permisos necesarios para registrar un bucket de tablas en Lake Formation, consulte Requisitos previos para integrar el catálogo de Tablas de Amazon S3 con el Catálogo de datos y Lake Formation.
- Registre el bucket de tablas en la cuenta en la que va a crear el catálogo de tablas de S3.

- Para el acceso entre cuentas, el rol especificado al registrar el bucket de tablas debe ser un rol dentro de la cuenta con los permisos adecuados para acceder al bucket entre cuentas. El rol necesita las acciones de IAM necesarias para las Tablas de S3 a fin de acceder a los recursos del bucket.

Cross-account registro de cubos de mesa

En los siguientes procedimientos, la cuenta A es la cuenta propietaria del recurso y la cuenta B es donde se registrará el bucket de tablas para administrar los permisos de acceso.

1. Inicie sesión Consola de administración de AWS en la cuenta A.

   Abra la consola de Amazon S3 en https://console.aws.amazon.com/s3/.

2. Cree un bucket de tablas. Para obtener más información, consulte Creación de un bucket de tablas en la Guía del usuario de Amazon S3.

3. Registre el bucket de tablas en la cuenta B.

   Usa AWS CLI para registrar el cubo de mesa de la cuenta A con Lake Formation en la cuenta B.

   ```
   aws lakeformation register-resource \
   --resource-arn 'arn:aws:s3tables:us-east-1:account-A-id:bucket/single-bucket-name'
    \
   --role-arn arn:aws:iam::account-B-id:role/role-name \
   --region us-east-1
   ```

   Sustituya el nombre de la cuentaA-id, el nombre de un solo grupo, el nombre de la cuenta B-id y el nombre del rol por sus valores específicos.

4. A continuación, cree un catálogo para el bucket de tablas en la cuenta B.

   Cree un catálogo con la AWS CLI.

   ```
   aws glue create-catalog --region us-east-1 \
   --cli-input-json \
   '{
      "Name": "catalog-name",
      "CatalogInput" : {
         "FederatedCatalog": {
            "Identifier": "arn:aws:s3tables:us-east-1:account A:bucket/single-bucket-
   name",
   ```

```
          "ConnectionName": "aws:s3tables"
      },
      "CreateDatabaseDefaultPermissions": [],
      "CreateTableDefaultPermissions": []
   }
}'
```

Sustituya el nombre del catálogo, la cuenta y el nombre del depósito por sus valores A-id específicos.

En el ejemplo siguiente de la CLI, se muestra cómo ver los detalles del catálogo.

```
aws glue get-catalog \
  --catalog-id account-id:catalog-name \
  --region us-east-1
```

5. A continuación, cree bases de datos y tablas en el catálogo recién creado en la cuenta B.

   Cree una base de datos.

```
aws glue create-database \
  --region us-east-1 \
  --catalog-id "account-B-id:catalog-name" \
  --database-input \
'{
  "Name": "database-name"
}'
```

   Crear una tabla de .

```
aws glue create-table \
  --database-name database-name \
  --catalog-id account-B-id:catalog-name\
  --region us-east-1 \
  --table-input \
  '{
        "Name": "table-name",
        "Parameters": {
            "format": "ICEBERG"
        },
        "StorageDescriptor": {
```

```
            "Columns": [
          {"Name": "x", "Type": "int", "Parameters": {"required": "true"}}
            ]
          }
}'
```

Sustituya el nombre de la base de datos, el nombre de la cuenta, el nombre del catálogo B-id y el nombre de la tabla por sus valores específicos.

6. En el ejemplo siguiente de la CLI, se muestra cómo ver los objetos del catálogo.

   Vea los detalles de la base de datos.

   ```
   aws glue get-database \
     --name database-name \
     --catalog-id account-B-id:catalog-name \
     --region us-east-1
   ```

   Vea los detalles de las tablas.

   ```
   aws glue get-table \
     --name table-name \
     --database-name database-name \
     --catalog-id account-B-id:catalog-name \
     --region us-east-1
   ```

   Sustituya el nombre de la base de datos, el nombre de la cuenta, el nombre del catálogo y el nombre de la B-id tabla por sus valores específicos.

Concesión de permisos

Tras integrar las tablas de S3 AWS Lake Formation, puede conceder permisos sobre el catálogo de tablas de S3 y los objetos del catálogo (depósitos de tablas, bases de datos, tablas) a otros roles y usuarios de IAM de su cuenta. Los permisos de Lake Formation permiten definir controles de acceso en el nivel de tabla, columna y fila para los usuarios de motores de análisis integrados, como Amazon Redshift Spectrum y Athena.

Puede conceder permisos mediante el método de recurso con nombre asignado o el método de control de acceso basado en etiquetas (LF-TBAC) de Lake Formation. Antes de conceder permisos

mediante LF-Tag expresiones LF-Tags y, por tanto, debe definirlas y asignarlas a los objetos del catálogo de datos.

Para obtener más información, consulte Gestión del control LF-Tags de acceso a los metadatos.

Puede compartir bases de datos y tablas con AWS cuentas externas concediendo permisos de Lake Formation a las cuentas externas. A continuación, los usuarios pueden ejecutar consultas y trabajos para unir y consultar tablas en varias cuentas. Cuando comparte un recurso del catálogo con otra cuenta, las entidades principales de esa cuenta pueden operar con ese recurso como si estuviera en su Catálogo de datos.

Al compartir bases de datos y tablas con cuentas externas, el permiso de superusuario no está disponible.

Para obtener instrucciones detalladas sobre la concesión de permisos, consulte la sección Administrar los permisos de Lake Formation.

AWS CLI ejemplo de concesión de permisos en una tabla de Amazon S3

```
aws lakeformation grant-permissions \
--cli-input-json \
'{
    "Principal": {
        "DataLakePrincipalIdentifier":"arn:aws:iam::111122223333:role/DataAnalystRole"
    },
    "Resource": {
        "Table": {
            "CatalogId":"111122223333:s3tablescatalog/amzn-s3-demo-bucket1",
            "DatabaseName":"S3 table bucket namespace <example_namespace>",
            "Name":"S3 table bucket table name <example_table>"
        }
    },
    "Permissions": [
        "SELECT"
    ]
}'
```

Se deben incluir los siguientes parámetros en el comando:

- DataLakePrincipalIdentifier — ARN de usuario, rol o grupo de IAM para conceder permisos

- CatalogId — ID de AWS cuenta de 12 dígitos propietario del catálogo de datos
- DatabaseName — Nombre del espacio de nombres del bucket de la tabla Amazon S3
- Name: nombre de la tabla del bucket de tablas de Amazon S3
- Permissions: permisos que se van a conceder Las opciones incluyen: SELECT, INSERT, DELETE, DESCRIBE, ALTER, DROP, ALLL y SUPER.

Acceso a las tablas de Amazon S3 compartidas

Después de conceder permisos entre cuentas en una base de datos o tabla del catálogo de tablas de S3, para acceder a los recursos, debe crear enlaces de recursos para las bases de datos y tablas compartidas.

1. En la cuenta de destino (la cuenta que recibe los recursos compartidos), cree un enlace a los recursos de la base de datos. Para obtener instrucciones detalladas, consulte Crear un enlace de recursos a una base de datos de Catálogo de datos compartida.

   Ejemplo de la CLI para crear un enlace a los recursos de la base de datos

   ```
   aws glue create-database
   --region us-east-1
   --catalog-id "111122223333"
   --database-input \
   '{
     "Name": "s3table_resourcelink",
     "TargetDatabase": {
       "CatalogId": "011426214932:s3tablescatalog/chmni-s3-table-bucket-011426214932",
       "DatabaseName": "s3_table_ns"
     },
     "CreateTableDefaultPermissions": []
   }'
   ```

2. Conceda permisos entre cuentas para la tabla.

   Ejemplo de la CLI para la concesión de permisos entre cuentas

   ```
   aws lakeformation grant-permissions \
   --region us-east-1 \
   --cli-input-json \
   '{
       "Principal": {
   ```

```
        "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:role/
S3TablesTestExecRole"
    },
    "Resource": {
        "Table": {
            "CatalogId": "011426214932:s3tablescatalog/chmni-s3-table-
bucket-011426214932",
            "DatabaseName": "s3_table_ns",
            "Name": "test_s3_iceberg_table"
        }
    },
    "Permissions": [
        "ALL"
    ]
}'
```

3. Conceda el permiso DESCRIBE de Lake Formation en el enlace de recursos.

   Ejemplo de la CLI para conceder el permiso Describe en el enlace de recursos.

```
aws lakeformation grant-permissions \
    --principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:role/
S3TablesTestExecRole
    --resource Database='{CatalogId=111122223333;, Name=s3table_resourcelink}' \
    --permissions DESCRIBE
```

Cambiar los controles de acceso para la integración de S3 Tables

Una vez que haya integrado Amazon S3 Tables con AWS Glue Data Catalog, puede cambiar la forma en que se controla el acceso a los recursos del catálogo. En esta sección, se explica cómo cambiar el control de acceso en función del modelo de control de acceso actual y deseado. Habilitar Lake Formation le permite usar permisos detallados, como la seguridad a nivel de columna y fila, a través de las subvenciones de Lake Formation, y permite a Lake Formation vender credenciales temporales en nombre de los directores a través de un puesto registrado. Al cambiar el control de acceso AWS Lake Formation a IAM, el control de acceso vuelve a las políticas de IAM estándar, lo que puede resultar adecuado si sus cargas de trabajo no requieren un acceso detallado y si prefiere gestionar los permisos completamente a través de IAM. Ambas rutas de migración implican actualizar los valores predeterminados del catálogo de datos, ajustar los registros de recursos con Lake Formation y coordinar las concesiones de permisos para evitar interrupciones en el acceso durante la transición.

Temas

- Habilite la integración de Lake Formation con S3 Tables con Data Catalog
- Cambie el control de acceso de AWS Lake Formation a IAM

Habilite la integración de Lake Formation con S3 Tables con Data Catalog

En esta sección se describe el flujo de trabajo para migrar el control de acceso de los privilegios de IAM a IAM con AWS Lake Formation concesiones para las tablas de Amazon S3 integradas en. AWS Glue Data Catalog

> **Important**
>
> Al habilitar el control de AWS Lake Formation acceso, se revocarán todos los IAM-based accesos existentes a sus recursos de S3 Tables. Tras completar el paso 1, los usuarios y roles que anteriormente accedían a los datos mediante los permisos de IAM perderán el acceso de forma inmediata. Debe conceder los permisos de Lake Formation en el paso 2 para que los usuarios puedan volver a consultar los datos. Planifique esta migración durante un período de mantenimiento y coordine con su equipo de datos.

Requisitos previos

Para read/write acceder a S3 Tables, además de los permisos de Lake Formation, los directores también necesitan el permiso de `lakeformation:GetDataAccess` IAM. Con este permiso, Lake Formation concede la solicitud de credenciales temporales para acceder a los datos.

Utilización AWS CLI

1. Paso 1: Registrar un depósito en Lake Formation mediante el rol de IAM

   Registre el recurso S3 Tables en Lake Formation.

   > **Note**
   >
   > Si ya tiene un rol, asegúrese de que el acceso híbrido sea falso.

   ```
   aws lakeformation register-resource \
   ```

```
  --resource-arn "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/*" \
  --role-arn "arn:aws:iam::AWSAccountID:role/service-role/LFAccessRole" \
  --with-federation
```

2. Paso 2: Actualizar el AWS Glue catálogo para habilitar el control de acceso a Lake Formation

   Actualice el catálogo con el `CreateDatabaseDefaultPermissions` campo vacío y `CreateTableDefaultPermissions` (establecido en`[]`). Esto garantiza que las bases de datos y las tablas del catálogo se administren mediante las concesiones de Lake Formation en lugar de los permisos IAM-based predeterminados.

```
aws glue update-catalog \
  --catalog-id "s3tablescatalog" \
  --catalog-input '{
    "FederatedCatalog": {
        "Identifier": "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/*",
        "ConnectionName": "aws:s3tables"
    },
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "AllowFullTableExternalDataAccess": "True"
  }'
```

3. Paso 3: Otorgue permisos de Lake Formation a su equipo de datos

   Otorgue permisos a Lake Formation a los directores (roles, usuarios o grupos) que necesiten acceso. Por ejemplo, para conceder acceso de lectura de tabla completa a un rol:

```
aws lakeformation grant-permissions \
  --principal
 DataLakePrincipalIdentifier=arn:aws:iam::AWSAccountID:role/DataTeamRole \
  --resource '{
    "Table": {
        "CatalogId": "AWSAccountID",
        "DatabaseName": "s3tablescatalog/table-bucket-name/namespace",
        "TableWildcard": {}
    }
  }' \
  --permissions "SELECT" "DESCRIBE"
```

   Repita el procedimiento para cada combinación principal y de recursos según sea necesario.

Cambie el control de acceso de AWS Lake Formation a IAM

En esta sección se describe el flujo de trabajo para cambiar el control de acceso de AWS Lake Formation las concesiones a los privilegios de IAM para las tablas de Amazon S3 integradas con. AWS Glue Data Catalog

> **Important**
>
> Si cambias el control de acceso de AWS Lake Formation las concesiones al de IAM, se revocará todo el Formation-based acceso existente de Lake a tus recursos de S3 Tables. Después de completar el paso 2, los usuarios y roles que anteriormente accedieron a los datos a través de las subvenciones de Lake Formation perderán el acceso inmediatamente. Debe conceder el acceso a IAM en el paso 1 antes de actualizar el catálogo. Planifique esta migración durante un período de mantenimiento y coordine con su equipo de datos.

> **Important**
>
> Fine-grained Los controles de acceso, como el acceso a nivel de columnas y los filtros de celdas de datos, junto con los objetos del catálogo de datos, están disponibles solo cuando se utilizan AWS Lake Formation . Antes de proceder a migrar los controles de acceso de IAM AWS Lake Formation a IAM, audite sus subvenciones actuales de Lake Formation `aws lakeformation list-permissions` y determine si las políticas de IAM equivalentes pueden proporcionar el acceso que necesitan sus usuarios. Cualquier director que dependiera de subvenciones detalladas de Lake Formation necesitará un acceso completo a la IAM a nivel de mesa después de migrar el control de acceso.

Requisitos previos

Antes de empezar, asegúrese de lo siguiente:

- Has identificado todas las subvenciones de Lake Formation actualmente en vigor para los recursos que se están migrando. Corre `aws lakeformation list-permissions --resource-type TABLE` a revisarlas.
- Ha preparado políticas de IAM que proporcionan un acceso equivalente a todos los directores afectados.

- El rol de IAM registrado en Lake Formation aún lo tiene `lakeformation:GetDataAccess` (necesario durante el período de transición híbrida).

Utilización AWS CLI

1. Paso 1: Conceder permisos de IAM a los directores

   Adjunte las políticas de IAM a los usuarios o roles a los que necesiten acceso. La política debe incluir tanto AWS Glue los permisos de metadatos como los permisos de datos de S3 Tables.

   > **Note**
   >
   > El siguiente ejemplo de política solo proporciona acceso de lectura.

```
aws iam put-user-policy \
  --user-name GlueIAMAccessUser \
  --policy-name S3TablesIAMAccessPolicy \
  --policy-document '{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "GlueMetadataAccess",
            "Effect": "Allow",
            "Action": [
                "glue:GetCatalog",
                "glue:GetDatabase",
                "glue:GetTable"
            ],
            "Resource": [
                "arn:aws:glue:us-east-1:AWSAccountID:catalog/s3tablescatalog",
                "arn:aws:glue:us-east-1:AWSAccountID:database/
s3tablescatalog/table-bucket-name/namespace",
                "arn:aws:glue:us-east-1:AWSAccountID:table/s3tablescatalog/table-
bucket-name/namespace/*"
            ]
        },
        {
            "Sid": "S3TablesDataAccess",
            "Effect": "Allow",
            "Action": [
```

```
                "s3tables:GetTableBucket",
                "s3tables:GetTable",
                "s3tables:GetTableMetadataLocation",
                "s3tables:GetTableData"
            ],
            "Resource": [
                "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/table-bucket-name",
                "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/table-bucket-name/
table/*"
            ]
        }
    ]
  }'
```

Compruebe que todos los usuarios y roles afectados puedan acceder a las tablas esperadas con sus credenciales de IAM antes de continuar.

2. Paso 2: Actualice el catálogo para restaurar los permisos predeterminados de IAM

   Actualice el catálogo para eso `CreateDatabaseDefaultPermissions` y `CreateTableDefaultPermissions` asígneloALL. `IAM_ALLOWED_PRINCIPALS` `OverwriteChildResourcePermissionsWithDefaultAccept`Configúrelo para que el cambio se propague a todos los recursos secundarios existentes, no solo a los recién creados.

```
aws glue update-catalog \
  --catalog-id "s3tablescatalog" \
  --catalog-input '{
    "FederatedCatalog": {
        "Identifier": "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/*",
        "ConnectionName": "aws:s3tables"
    },
    "CreateDatabaseDefaultPermissions": [{
        "Principal": {"DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"},
        "Permissions": ["ALL"]
    }],
    "CreateTableDefaultPermissions": [{
        "Principal": {"DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"},
        "Permissions": ["ALL"]
    }],
    "OverwriteChildResourcePermissionsWithDefault": "Accept"
  }'
```

3. Paso 3: Anular el registro del recurso de Lake Formation

Una vez que hayas confirmado que todos los accesos funcionan según las políticas de IAM y que los directores no dependen de las subvenciones de Lake Formation, puedes anular el registro del recurso de Lake Formation para completar la migración.

```
aws lakeformation deregister-resource \
  --resource-arn "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/*"
```

> **Note**
>
> Tras anular el registro del recurso, elimínelo de las entidades principales `lakeformation:GetDataAccess` de IAM que ya no lo necesiten.

No es necesario realizar ningún paso`revoke-permissions`.

Creación de un catálogo gestionado de Amazon Redshift en AWS Glue Data Catalog

Es posible que no tenga un clúster productor de Amazon Redshift o un recurso compartido de datos de Amazon Redshift disponibles en la actualidad pero desee crear y administrar tablas de Amazon Redshift mediante el AWS Glue Data Catalog. Para empezar, puede crear un catálogo AWS Glue gestionado mediante la `glue:CreateCatalog` API o la AWS Lake Formation consola configurando el tipo de catálogo como `Managed` y `Catalog source` como Redshift. Este paso hace lo siguiente:

- Crea un catálogo en el Catálogo de datos
- Registra el catálogo como una ubicación de datos de Lake Formation
- Crea un grupo de trabajo administrado de Amazon Redshift sin servidor
- Vincula el grupo de trabajo de Amazon Redshift sin servidor y el Catálogo de datos mediante un objeto de recurso compartido de datos

Cómo crear un catálogo administrado y configurar los permisos (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación, elija Catálogos en Catálogo de datos.
3. Seleccione la opción Crear catálogo.

4. En la página de detalles Establecer catálogo, facilite la siguiente información:

   - Nombre: un nombre exclusivo para el catálogo administrado. El nombre no se puede cambiar y debe ir en minúsculas. El nombre puede tener un máximo de 255 caracteres.
   - Tipo: elija `Managed catalog` como tipo de catálogo.
   - Almacenamiento: elija `Redshift` para el almacenamiento.
   - Descripción: escriba una descripción para el catálogo creado desde el origen de datos.

5. Puede utilizar las aplicaciones de Apache Spark que se ejecutan en Amazon EMR en Amazon EC2 para acceder a las bases de datos de Amazon Redshift en el AWS Glue Data Catalog.

   Para permitir que Apache Spark lea y escriba en el almacenamiento administrado de Amazon Redshift, AWS Glue crea un clúster de Amazon Redshift administrado con los recursos de computación y de almacenamiento necesarios para realizar operaciones de lectura y escritura sin afectar a las cargas de trabajo del almacén de datos de Amazon Redshift. También debe proporcionar un rol de IAM con los permisos necesarios para transferir datos hacia y desde el bucket de Amazon S3. Para conocer los permisos necesarios para el rol de transferencia de datos, consulte el paso 5 de la sección Requisitos previos para administrar los espacios de nombres de Amazon Redshift en AWS Glue Data Catalog.

6. De forma predeterminada, los datos del clúster de Amazon Redshift se cifran mediante una clave AWS gestionada. Lake Formation ofrece una opción para crear la clave de KMS personalizada para el cifrado. Si utiliza una clave administrada por el cliente, debe agregar políticas de claves específicas a la clave.

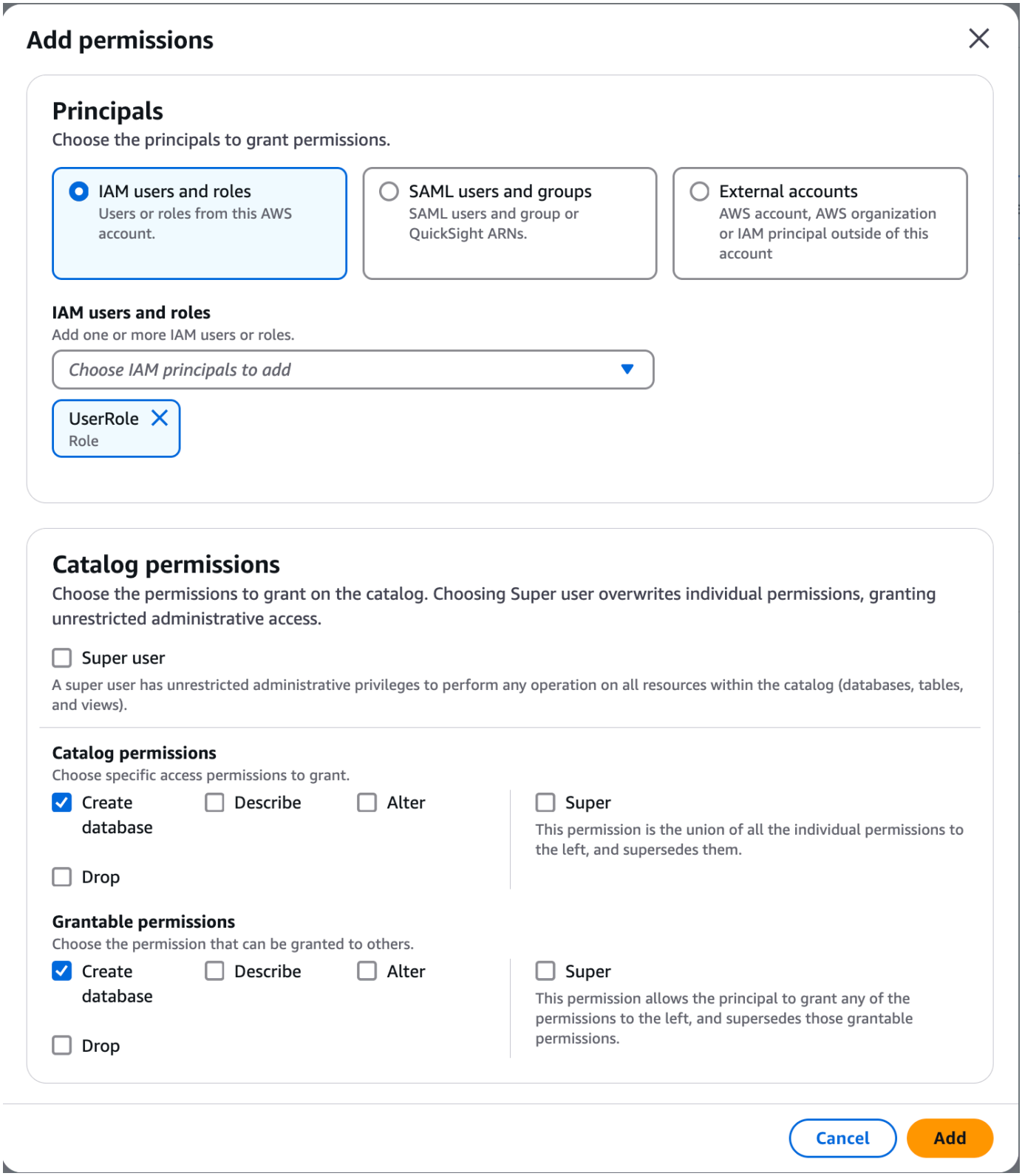
7. Elija Personalización de la configuración de cifrado si utiliza una clave administrada por el cliente para cifrar los datos del almacenamiento administrado de Amazon Redshift. Para usar una clave personalizada, debe agregar una política de claves administradas personalizadas adicionales a la clave de KMS. Para obtener más información, consulte Requisitos previos para administrar los espacios de nombres de Amazon Redshift en AWS Glue Data Catalog.

8. Opciones de cifrado: elija la opción Personalizar la configuración de cifrado si desea utilizar una clave personalizada para cifrar el catálogo. Para usar una clave personalizada, debe agregar una política de claves administradas personalizadas adicionales a la clave de KMS.

9. Seleccione Siguiente para conceder permisos a otras entidades principales.

10. En la página Concesión de permisos, elija Agregar permisos.

11. En la pantalla Agregar permisos, elija las entidades principales y los tipos de permisos que desea conceder.



- En la sección Entidades principales, elija uno de los tipos y, a continuación, especifique las que van a recibir los permisos concedidos.

- Usuarios y roles de IAM: elija uno o varios usuarios o roles de la lista de usuarios y roles de IAM.
- Usuarios y grupos de SAML: para SAML y Amazon Quick usuarios y grupos, introduzca uno o más nombres de recursos de Amazon (ARN) para los usuarios o grupos federados a través de SAML, o ARN para los usuarios o grupos de Amazon Quick. Pulse Intro después de cada ARN.

  Para obtener información sobre cómo construir los ARN, consulte los comandos grant and revoke. AWS CLI AWS CLI

- En la sección Permisos, seleccione los permisos y los permisos concedibles.

  En Permisos de catálogo, seleccione uno o más permisos para conceder.

  Elija Superusuario para adjudicar permisos administrativos sin restricciones en todos los recursos del catálogo.

  En Permisos concedibles, seleccione los permisos que el destinatario de la subvención puede conceder a otros directores de su cuenta. AWS Esta opción no es compatible cuando se conceden permisos a una entidad principal de IAM desde una cuenta externa.

12. Elija Siguiente para revisar la información y crear el catálogo. La lista Catálogos muestra el nuevo catálogo administrado.

Creación de un catálogo federado (CLI)

- En el siguiente ejemplo, se muestra cómo crear un catálogo federado.

```
aws glue create-catalog --cli-input-json file://input.json

{
    "Name": "CatalogName",
    "CatalogInput": {
      "Description": "Redshift published Catalog",
      "CreateDatabaseDefaultPermissions" : [],
      "CreateTableDefaultPermissions": [],
      "CatalogProperties": {
        "DataLakeAccessProperties" : {
          "DataLakeAccess" : "true",
          "DataTransferRole" : "DTR arn",
          "KMSKey": "kms key arn",  // Optional
```

```
          "CatalogType": "aws:redshift"
        }
      }
    }
}
```

Respuesta de Glue get-catalog

```
aws glue get-catalog \
  --catalog-id account-id:catalog-name \
  --region us-east-1

Response:
{
    "Catalog": {
        "Name": "CatalogName",
        "Description": "Glue Catalog for Redshift z-etl use case",
        "CreateDatabaseDefaultPermissions" : [],
        "CreateTableDefaultPermissions": [],
         "CatalogProperties": {
          "DataLakeAccessProperties" : {
            "DataLakeAccess": "true",
            "DataTransferRole": "DTR arn",
            "KMSKey": "kms key arn",
            "ManagedWorkgroupName": "MWG name",
            "ManagedWorkgroupStatus": "MWG status",
            "RedshiftDatabaseName": "RS db name",
            "NamespaceArn": "namespace key arn",
            "CatalogType": "aws:redshift"
         }
       }
    }
```

Administración de permisos para los datos de un recurso compartido de datos de Amazon Redshift

AWS Lake Formation permite gestionar de forma segura los datos de un recurso compartido de datos de Amazon Redshift. Amazon Redshift es un servicio administrado de almacenamiento de datos a escala de petabytes en la nube AWS. Al utilizar la capacidad de compartir datos, Amazon

Redshift le ayuda a compartir datos entre Cuentas de AWS. Para obtener más información sobre el uso compartido de datos, consulte Información general del uso compartido de datos en Amazon Redshift.

En Amazon Redshift, el administrador del clúster productor crea un recurso compartido de datos y lo comparte con el administrador del lago de datos. Para obtener instrucciones paso a paso sobre cómo crear un administrador de lago de datos, consulte Crear un administrador de lago de datos.

Una vez que el usuario (administrador del lago de datos) acepte el recurso compartido de datos, debe crear una base de datos AWS Glue Data Catalog para el recurso compartido de datos específico. Esto es para que pueda controlar el acceso a mediante los permisos de Lake Formation. Lake Formation mapea cada recurso compartido de datos a la base de datos del Catálogo de datos correspondiente. Aparecen como bases de datos federadas en Data Catalog.

Se habla de base de datos federada cuando la base de datos apunta a una entidad ajena al Catálogo de datos. Las tablas y vistas del recurso compartido de datos de Amazon Redshift se muestran como tablas individuales en el Catálogo de datos. Puede compartir la base de datos federada con entidades principales de IAM y usuarios de SAML seleccionados en la misma cuenta o en otra cuenta con Lake Formation. También puede incluir expresiones de filtro de filas y columnas para restringir el acceso a determinados datos. Para obtener más información, consulte Filtrado de datos y seguridad de celda en Lake Formation.

Para proporcionar acceso a los usuarios a un recurso compartido de datos de Amazon Redshift, debe hacer lo siguiente:

1. Actualice la Configuración del Catálogo de datos para habilitar los permisos de Lake Formation.
2. Acepte la invitación para compartir datos del administrador del clúster de productores de Amazon Redshift y registre el recurso compartido de datos en Lake Formation.

   Tras completar este paso, puede administrar el recurso compartido de datos con el Catálogo de datos de Lake Formation.
3. Cree una base de datos federada y defina los permisos en esa base de datos.
4. Conceda permisos a los usuarios para bases de datos y tablas. Puede compartir toda la base de datos o un subconjunto de tablas con usuarios de la misma cuenta o de otra cuenta.

Para conocer las limitaciones, consulte Limitaciones de uso compartido de datos de Amazon Redshift.

Temas

- Requisitos previos para configurar permisos en recursos compartidos de datos de Amazon Redshift
- Configurar permisos en recursos compartidos de datos de Amazon Redshift
- Consultar las bases de datos federadas

Requisitos previos para configurar permisos en recursos compartidos de datos de Amazon Redshift

Actualice la configuración predeterminada del Catálogo de datos

Para habilitar los permisos de Lake Formation para los recursos del Catálogo de datos, le recomendamos que deshabilite la configuración predeterminada del Catálogo de datos en Lake Formation. Para obtener más información, consulte Cambio del modelo de permisos predeterminado o uso del modo de acceso híbrido.

Actualizar permisos

Además de los permisos de administrador de lago de datos (`AWSLakeFormationDataAdmin`), también se requieren los siguientes permisos para aceptar un recurso compartido de datos de Amazon Redshift en Lake Formation:

- `glue:PassConnection on aws:redshift`
- `redshift:AssociateDataShareConsumer`
- `redshift:DescribeDataSharesForConsumer`
- `redshift:DescribeDataShares`

El usuario de IAM administrador del lago de datos tiene implícitos los siguientes permisos.

- data_location_access
- create_database
- lakefomation:registerResource

Configurar permisos en recursos compartidos de datos de Amazon Redshift

En este tema se describen los pasos que debe seguir para aceptar una invitación a compartir datos, crear una base de datos federada y conceder permisos. Puede utilizar la consola de Lake Formation o AWS Command Line Interface (AWS CLI). Los ejemplos de este tema muestran el clúster de productores, el Catálogo de datos y el consumidor de datos en la misma cuenta.

Para obtener más información sobre las capacidades entre cuentas de Lake Formation, consulte Cross-account intercambio de datos en Lake Formation.

Para configurar los permisos de un recurso compartido de datos

1. Revise una invitación al recurso compartido de datos y acéptela.

   Console

   1. Inicie sesión en la consola de Lake Formation como administrador del lago de datos en https://console.aws.amazon.com/lakeformation/. Acceda a la página Intercambio de datos.
   2. Revise los recursos compartidos de datos para los que tiene autorización de acceso. La columna Estado indica su estado de participación actual en el recurso compartido de datos. El estado Pendiente indica que se le ha agregado a un recurso compartido de datos, pero que aún no lo ha aceptado o ha rechazado la invitación.
   3. Para responder a la invitación al recurso compartido de datos, seleccione el nombre del recurso compartido de datos y elija Revisar la invitación. En Aceptar o rechazar recurso compartido de datos, revise los detalles de la invitación. Seleccione Aceptar para aceptar la invitación o Rechazar para rechazarla. Si rechaza la invitación, no obtendrá acceso al recurso compartido de datos.

   AWS CLI

   En los ejemplos siguientes se muestra cómo ver, aceptar y registrar la invitación. Sustituya el ID de Cuenta de AWS por un ID de Cuenta de AWS válido. Sustituya `data-share-arn` por el nombre de recurso de Amazon (ARN) real que hace referencia al recurso compartido de datos.

   1. Consulte una invitación pendiente.

      ```
      aws redshift describe-data-shares \
      ```

```
 --data-share-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds' \
```

2. Acepte un recurso compartido de datos.

```
 aws redshift associate-data-share-consumer \
 --data-share-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds' \
 --consumer-arn 'arn:aws:glue:us-east-1:111122223333:catalog
```

3. Registre el recurso compartido de datos en la cuenta de Lake Formation. Utilice la operación de la API RegisterResource para registrar el recurso compartido de datos en Lake Formation `DataShareArn`es el parámetro de entrada de `ResourceArn`.

> **Note**
>
> Este es un paso obligatorio.

```
aws lakeformation register-resource \
 --resource-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds'
```

2. Cree una base de datos.

   Tras aceptar una invitación para un recurso compartido de datos, debe crear una base de datos que apunte a la base de datos de Amazon Redshift asociada al recurso compartido de datos. Debe ser administrador de un lago de datos para crear una base de datos.

   Console

   1. Seleccione el recurso compartido de datos en el panel Invitaciones y elija Establecer los detalles de la base de datos.
   2. En Establecer los detalles de la base de datos, introduzca un nombre e identificador únicos para el recurso compartido de datos. Este identificador se utiliza para asignar el recurso compartido de datos internamente en la jerarquía de metadatos (dbname.schema.Table).

3. Seleccione Siguiente para conceder permisos a otros usuarios en la base de datos y las tablas compartidas.

AWS CLI

Utilice el siguiente código de ejemplo para crear una base de datos que apunte a la base de datos de Amazon Redshift compartida con Lake Formation mediante AWS CLI.

```
aws glue create-database --cli-input-json \

'{
 "CatalogId": "111122223333",
 "DatabaseInput": {
  "Name": "tahoedb",
  "FederatedDatabase": {
       "Identifier": "arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/federatedds",
       "ConnectionName": "aws:redshift"
   }
 }
 }'
```

3. Concesión de permisos.

Una vez creada la base de datos, puede conceder permisos a otros usuarios de su cuenta o a Cuentas de AWS y organizaciones externas. No podrá conceder permisos de escritura de datos (insertar, eliminar) ni permisos de metadatos (modificar, eliminar, crear) en la base de datos federada asignada a un recurso compartido de datos de Amazon Redshift. Para obtener más información sobre cómo conceder permisos, consulte Administrar los permisos de Lake Formation.

> **Note**
>
> Como administrador del lago de datos, solo puede ver las tablas de las bases de datos federadas. Para realizar cualquier otra acción, debe concederse más permisos sobre dichas tablas.

Console

1. En la pantalla Conceder permisos, seleccione los usuarios a los que va a conceder los permisos.
2. Elija Conceder.

AWS CLI

Utilice los siguientes ejemplos para conceder permisos de bases de datos y tablas mediante AWS CLI:

```
aws lakeformation grant-permissions --input-cli-json file://input.json

{
   "Principal": {
           "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/non-
admin"
   },
   "Resource": {
          "Database": {
                "CatalogId": "111122223333",
                 "Name": "tahoedb"
           }
    },
    "Permissions": [
             "DESCRIBE"
     ],
    "PermissionsWithGrantOption": [

     ]
 }
```

```
aws lakeformation grant-permissions --input-cli-json file://input.json

{
                  "Principal": {
                         "DataLakePrincipalIdentifier":
 "arn:aws:iam::111122223333:user/non-admin"
```

```
                 },
                "Resource": {
                       "Table": {
                            "CatalogId": "111122223333",
                            "DatabaseName": "tahoedb",
                            "Name": "public.customer"
                       }
                },
               "Permissions": [
                      "SELECT"
                ],
               "PermissionsWithGrantOption": [
                       "SELECT"
                 ]
}
```

Consultar las bases de datos federadas

Tras conceder los permisos, los usuarios pueden iniciar sesión y empezar a consultar la base de datos federada mediante Amazon Redshift. Los usuarios ahora pueden usar el nombre de la base de datos local para hacer referencia al recurso compartido de datos de Amazon Redshift en consultas de SQL. En Amazon Redshift, la tabla de clientes del esquema público que se comparte a través del recurso compartido de datos tendrá una tabla correspondiente creada como `public.customer` en el Catálogo de datos.

1. Antes de consultar la base de datos federada mediante Amazon Redshift, el administrador del clúster crea una base de datos a partir de la base de datos del Catálogo de datos mediante el siguiente comando:

   ```
   CREATE DATABASE sharedcustomerdb FROM ARN
    'arn:aws:glue:<region>:111122223333:database/tahoedb' WITH DATA CATALOG SCHEMA
    tahoedb
   ```

2. El administrador del clúster concede permisos de uso en la base de datos.

   ```
   GRANT USAGE ON DATABASE sharedcustomerdb TO IAM:user;
   ```

3. Ahora usted (el usuario federado) puede iniciar sesión en las herramientas de SQL para consultar la tabla.

```
Select * from sharedcustomerdb.public.customer limit 10;
```

Para obtener más información, visite Consulta de AWS Glue Data Catalog en la Guía de administración de Amazon Redshift.

Administración de los permisos de los conjuntos de datos que utilizan metaalmacenes externos

Con la federación de AWS Glue Data Catalog metadatos (federación de catálogos de datos), puede conectar el catálogo de datos a metaalmacenes externos que almacenan los metadatos de sus datos de Amazon S3 y administrar de forma segura los permisos de acceso a los datos mediante AWS Lake Formation. No tiene que migrar los metadatos del metaalmacén externo al Catálogo de datos.

El Catálogo de datos proporciona un repositorio de metadatos centralizado que facilita la administración y la detección de datos en sistemas dispares. Cuando su organización administra los datos del catálogo de datos, puede utilizarlos AWS Lake Formation para controlar el acceso a sus conjuntos de datos en Amazon S3.

> **Note**
>
> Actualmente, solo admitimos la federación de metaalmacenes Hive de Apache (versión 3 y superior).

Para configurar la federación del catálogo de datos, proporcionamos una aplicación AWS Serverless Application Model (AWS SAM) denominada GlueDataCatalogFederation-HiveMetastore. AWS Serverless Application Repository

La implementación de referencia se proporciona GitHub como un proyecto de código abierto en AWS Glue Data Catalog Federation - Hive Metastore.

La AWS SAM aplicación crea e implementa los siguientes recursos necesarios para conectar el catálogo de datos al metaalmacén de Hive:

- Una AWS Lambda función: aloja la implementación del servicio de federación que se comunica entre el catálogo de datos y el metaalmacén de Hive. AWS Glue invoca esta función Lambda para recuperar objetos de metadatos del metabastore de Hive.

- Amazon API Gateway: el punto de conexión del metaalmacén de Hive que actúa como proxy para enrutar todas las invocaciones a la función de Lambda.
- Un rol de IAM: un rol con los permisos necesarios para crear la conexión entre el Catálogo de datos y el metaalmacén de Hive.
- AWS Glue conexión: Amazon API Gateway tipo de AWS Glue conexión que almacena el Amazon API Gateway punto final y una función de IAM para invocarlo.

Al consultar tablas, el AWS Glue servicio realiza una llamada en tiempo de ejecución al metabastore de Hive y recupera los metadatos. La función de Lambda actúa como un traductor entre el metaalmacén de Hive y el Catálogo de datos.

Tras establecer la conexión, para sincronizar los metadatos del metaalmacén de Hive con el Catálogo de datos, debe crear una base de datos federada en el Catálogo de datos utilizando los detalles de conexión del metaalmacén de Hive y asignar esta base de datos a la base de datos de Hive. Una base de datos se denomina base de datos federada cuando apunta a una entidad ajena al Catálogo de datos.

Puede aplicar los permisos de Lake Formation mediante el control de acceso basado en etiquetas y el método de recurso con nombre en la base de datos federada, y compartirlos entre varias unidades organizativas (OU) y varias Cuentas de AWS unidades organizativas (OU). AWS Organizations También puede compartir la base de datos federada directamente con las entidades principales de IAM desde otra cuenta.

Puede definir permisos específicos de columna, fila y celda utilizando los filtros de datos de Lake Formation en tablas de Hive externas. Puede usar Amazon Athena, Amazon Redshift o Amazon EMR para consultar las tablas de Hive externas administradas por Lake Formation.

Para obtener más información sobre el filtrado y el intercambio de datos entre cuentas, consulte:

- Cross-account intercambio de datos en Lake Formation
- Filtrado de datos y seguridad de celda en Lake Formation

Pasos básicos de la federación de metadatos del Catálogo de datos

1. Los usuarios y roles de IAM se crean con los permisos adecuados para implementar la aplicación de AWS SAM y crear bases de datos federadas.

2. Para registrar la ubicación de datos de Amazon S3 en Lake Formation, debe seleccionar la opción `Enable Data Catalog federation` para los conjuntos de datos que utilizan un metaalmacén de Hive externo.
3. Debe configurar los ajustes de la AWS SAM aplicación (nombre de la AWS Glue conexión, URL al metabastore de Hive y parámetros de la función Lambda) e implementar la aplicación. AWS SAM
4. La AWS SAM aplicación despliega los recursos necesarios para conectar el metabastore externo de Hive con el catálogo de datos.
5. Para aplicar los permisos de Lake Formation en la base de datos y las tablas de Hive, crea una base de datos en el Catálogo de datos utilizando los datos de conexión del metaalmacén de Hive y asigna esta base de datos a la base de datos de Hive.
6. Conceda permisos en las bases de datos federadas a las entidades principales de su cuenta o de otra cuenta.

> **Note**
>
> Puede conectar el Catálogo de datos a un metaalmacén de Hive externo, crear bases de datos federadas y ejecutar consultas y scripts de ETL en bases de datos y tablas de Hive sin aplicar los permisos de Lake Formation. En el caso de los datos fuente de Amazon S3 que no estén registrados en Lake Formation, el acceso viene determinado por las políticas de permisos y AWS Glue acciones de IAM para Amazon S3.

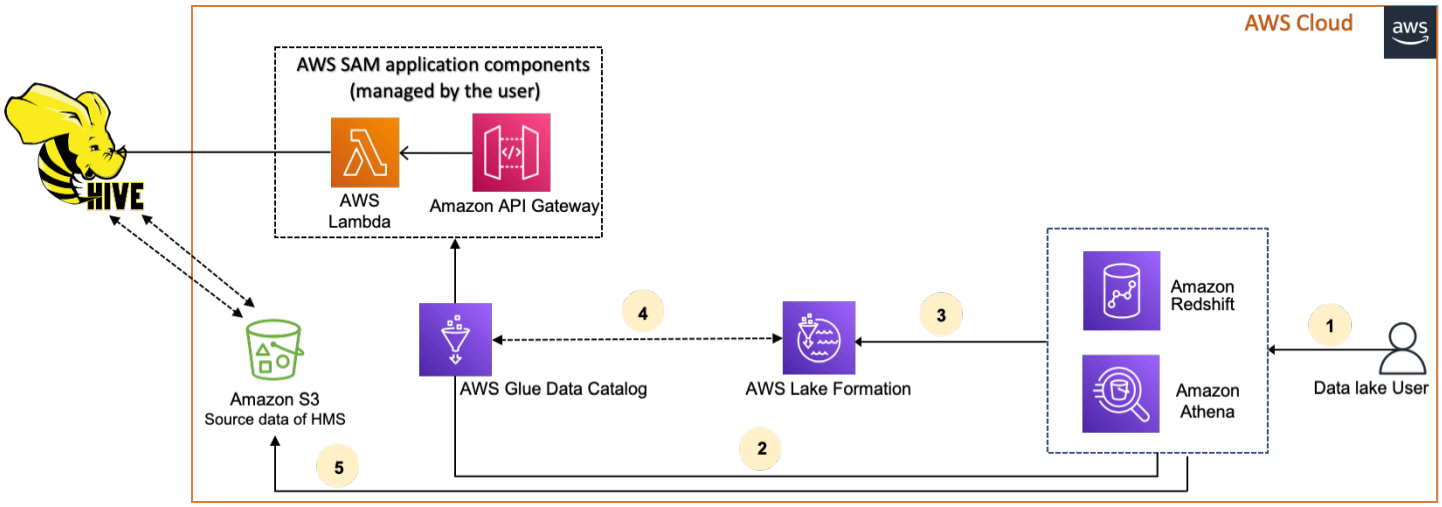
Para conocer las limitaciones, consulte Consideraciones y limitaciones del uso compartido de datos del almacén de metadatos de Hive.

Temas

- Flujo de trabajo
- Requisitos previos para conectar el Catálogo de datos al metaalmacén de Hive
- Conexión del Catálogo de datos a un metaalmacén de Hive externo
- Recursos adicionales

Flujo de trabajo

En el siguiente diagrama se muestra el flujo de trabajo para conectarlo AWS Glue Data Catalog a un metaalmacén de Hive externo.



1. Una entidad principal envía una consulta mediante un servicio integrado como Athena o Redshift Spectrum.
2. El servicio integrado realiza una llamada al catálogo de datos para obtener los metadatos, que a su vez llama al punto final del metaalmacén de Hive disponible en la versión anterior y recibe las respuestas a las solicitudes de metadatos. Amazon API Gateway
3. El servicio integrado envía la solicitud a Lake Formation para verificar la información de la tabla y las credenciales para acceder a la tabla.
4. Lake Formation autoriza la solicitud y suministra credenciales temporales a la aplicación integrada, que permite el acceso a los datos.
5. Con las credenciales temporales recibidas de Lake Formation, el servicio integrado lee los datos de Amazon S3 y comparte los resultados con la entidad principal.

Requisitos previos para conectar el Catálogo de datos al metaalmacén de Hive

Para conectarse AWS Glue Data Catalog a un metabastore externo de Apache Hive y configurar los permisos de acceso a los datos, debe cumplir los siguientes requisitos:

> **Note**
>
> Recomendamos que un administrador de Lake Formation despliegue la AWS SAM aplicación y que solo un usuario privilegiado utilice la conexión del metaalmacén de Hive para crear las bases de datos federadas correspondientes.

1. Crear roles de IAM.

   Para implementar la AWS SAM aplicación

   - Cree un rol que tenga los permisos necesarios para implementar los recursos (función Lambda Amazon API Gateway, rol de IAM y la AWS Glue conexión) necesarios para crear una conexión al metaalmacén de Hive.

   Para crear bases de datos federadas

   Se necesitan los siguientes permisos en los recursos:

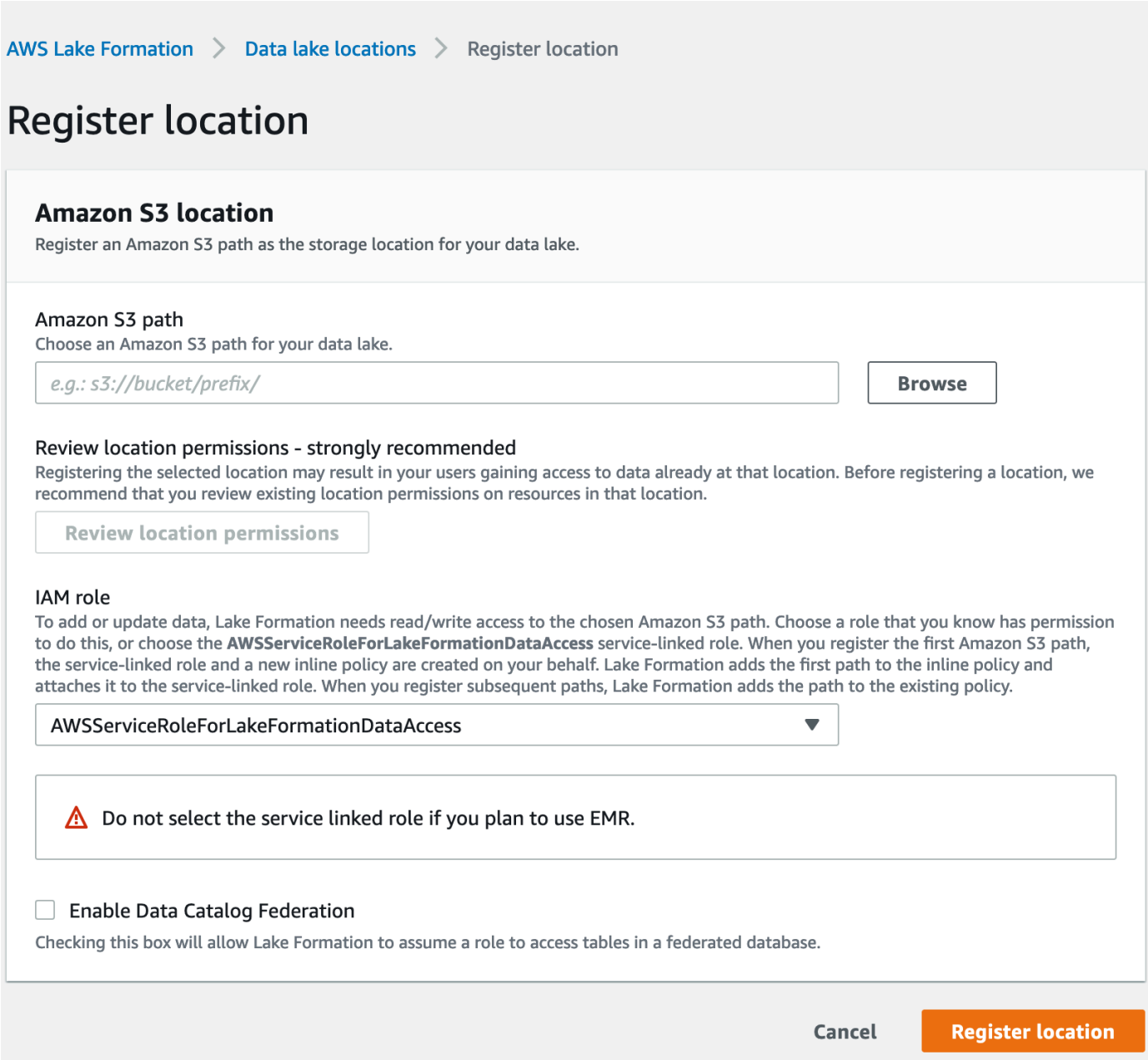
   - `glue:CreateDatabase on resource arn:aws:glue:region:account-id:database/gluedatabasename`
   - `glue:PassConnection on resource arn:aws:glue:region:account-id:connection/hms_connection`

2. Registre la ruta de Amazon S3 en Lake Formation.

   Para utilizar Lake Formation para gestionar y proteger los datos de su lago de datos, debe registrar la ubicación de Amazon S3 que contiene los datos de las tablas en el metaalmacén de Hive con Lake Formation. De este modo, Lake Formation puede vender credenciales a servicios AWS analíticos como Athena, Redshift Spectrum y Amazon EMR.

   Para obtener más información sobre el registro de una ubicación de Amazon S3, consulte Añadir una ubicación de Amazon S3 a su lago de datos.

   Cuando registre la ubicación de Amazon S3, seleccione la casilla de verificación Habilitar la federación del Catálogo de datos para permitir que Lake Formation asuma un rol de acceso a las tablas de una base de datos federada.



Para obtener más información sobre el registro de una ubicación de datos en Lake Formation, consulte Configurar una ubicación de Amazon S3 para el lago de datos.

3. Utilice la versión correcta de Amazon EMR.

   Para utilizar Amazon EMR con las bases de datos federadas del metaalmacén de Hive, debe disponer de Hive versión 3.x o superior y Amazon EMR versión 6.x o superior.

Conexión del Catálogo de datos a un metaalmacén de Hive externo

Para conectarlo AWS Glue Data Catalog a un metaalmacén de Hive, necesitas implementar una AWS SAM aplicación llamada. GlueDataCatalogFederation-HiveMetastore Crea e implementa los recursos necesarios para conectar el metaalmacén de Hive al Catálogo de datos. Puede acceder a la AWS SAM aplicación en. AWS Serverless Application Repository

La AWS SAM aplicación crea la conexión para el metaalmacén de Hive detrás de Amazon API Gateway mediante una función Lambda. La AWS SAM aplicación utiliza un identificador uniforme de recursos (URI) como entrada del usuario y conecta el metabastore externo de Hive al catálogo de datos. Cuando un usuario ejecuta una consulta en las tablas de Hive, el Catálogo de datos llama al punto de conexión de la puerta de enlace de API. El punto de conexión invoca la función de Lambda para recuperar los metadatos de las tablas de Hive.

Para conectar el Catálogo de datos al metaalmacén de Hive y configurar los permisos

1. Implemente el AWS SAM "Hello, World!".
    1. Inicie sesión en Consola de administración de AWS y abra el AWS Serverless Application Repository.
    2. En el panel de navegación, elija Aplicaciones disponibles.
    3. Elija Aplicaciones públicas.
    4. Seleccione la opción Mostrar aplicaciones que crean roles de IAM personalizados o políticas de recursos.
    5. En el cuadro de búsqueda, introduce el nombre GlueDataCatalogFederation-HiveMetastore.
    6. Elija la aplicación de GlueDataCatalogFederation-HiveMetastore.
    7. En Configuración de la aplicación, introduzca la siguiente configuración mínima requerida para la función de Lambda:
        - Nombre de la aplicación: un nombre para AWS SAM la aplicación.
        - GlueConnectionName- Un nombre para la conexión.
        - HiveMetastoreURIs- El URI de tu servidor de metatienda de Hive.
        - LambdaMemory- La cantidad de memoria Lambda en MB de 128 a 10240. El valor predeterminado es 1024.
        - LambdaTimeout- El tiempo de ejecución máximo de la invocación a Lambda en segundos. El valor predeterminado es 30.

- VPCSecurityGroupIdsy VPCSubnetIds: información de la VPC en la que se encuentra el metabastore de Hive.

8. Seleccione Confirmo que esta aplicación puede crear roles de IAM y políticas de recursos personalizados. Para obtener más información, elija el enlace Info.
9. En la sección Configuración de la aplicación, selecciona Implementar. Una vez finalizada la implementación, aparece la función de Lambda en la sección Recursos en la consola de Lambda.

La aplicación se implementa en Lambda. Su nombre va precedido de serverlessrepo- para indicar que la aplicación se implementó desde el AWS Serverless Application Repository. Al seleccionar la aplicación, accederá a la página Recursos, donde se enumeran todos los recursos de la aplicación que se implementaron. Los recursos incluyen la función Lambda que permite la comunicación entre el catálogo de datos y el metabastore de Hive, la AWS Glue conexión y otros recursos necesarios para la federación de bases de datos.

2. Crear una base de datos federada en Data Catalog.

Después de crear una conexión al metaalmacén de Hive, puede crear bases de datos federadas en el Catálogo de datos que apunten a las bases de datos externas del metaalmacén de Hive. Debe crear una base de datos correspondiente en el Catálogo de datos para cada base de datos del metaalmacén de Hive que conecte al Catálogo de datos.

Lake Formation console

1. En la página de Uso compartido de datos, seleccione la pestaña Bases de datos compartidas y, a continuación, Crear base de datos.
2. En Nombre de la conexión, elija el nombre de su conexión al metaalmacén de Hive en el menú desplegable.
3. Introduzca un nombre de base de datos único y el identificador de origen de federación para la base de datos. Este es el nombre que se utiliza en las instrucciones SQL cuando se consultan las tablas. El nombre puede tener un máximo de 255 caracteres y debe ser único en la cuenta.
4. Elija Creación de base de datos.

AWS CLI

```
aws glue create-database \
```

```
'{
 "CatalogId": "<111122223333>",
  "database-input": {
    "Name":"<fed_glue_db>",
    "FederatedDatabase":{
        "Identifier":"<hive_db_on_emr>",
        "ConnectionName":"<hms_connection>"
     }
   }
 }'
```

3. Ver las tablas de la base de datos federada.

   Después de crear la base de datos federada, puede ver la lista de tablas en su metaalmacén de Hive mediante la consola de Lake Formation o la AWS CLI.

   Lake Formation console

   1. Seleccione el nombre de la base de datos en la pestaña Bases de datos compartidas.
   2. En la página Bases de datos, elija Ver tablas.

   AWS CLI

   Los ejemplos siguientes muestran cómo recuperar la definición de la conexión, el nombre de la base de datos y algunas o todas las tablas de la base de datos. Sustituya el ID del catálogo de datos por el Cuenta de AWS ID válido que utilizó para crear la base de datos. Sustituya `hms_connection` por el nombre de la conexión.

   ```
   aws glue get-connection \
   --name <hms_connection>  \
   --catalog-id 111122223333
   ```

   ```
   aws glue get-database \
   --name <fed_glu_db> \
   --catalog-id 111122223333
   ```

   ```
   aws glue get-tables \
   --database-name <fed_glue_db> \
   --catalog-id 111122223333
   ```

```
aws glue get-table \
--database-name <fed_glue_db> \
--name <hive_table_name> \
--catalog-id 111122223333
```

4. Conceder permisos.

   Una vez creada la base de datos, puede conceder permisos a otros usuarios y roles de IAM en su cuenta o a organizaciones Cuentas de AWS y organizaciones externas. No podrá conceder permisos de escritura de datos (insertar, eliminar) ni permisos de metadatos (modificar, eliminar, crear) en las bases de datos federadas. Para obtener más información sobre cómo conceder permisos, consulte Administrar los permisos de Lake Formation.

5. Consultar las bases de datos federadas.

   Tras conceder los permisos, los usuarios pueden iniciar sesión y empezar a consultar la base de datos federada mediante Athena y Amazon Redshift. Los usuarios ahora pueden usar el nombre de la base de datos local para hacer referencia a la base de datos de Hive en las consultas de SQL.

   Ejemplo de sintaxis Amazon Athena de consulta

   Sustituya `fed_glue_db` por el nombre de la base de datos local que creó anteriormente.

   ```
   Select * from fed_glue_db.customers limit 10;
   ```

Recursos adicionales

La siguiente entrada del blog contiene instrucciones detalladas para configurar los permisos de Lake Formation en una base de datos y tablas del metaalmacén de Hive y consultarlos mediante Athena. También ilustramos un caso de uso compartido entre cuentas, en el que un director de Lake Formation en la cuenta de productor A comparte una base de datos y tablas de Hive federadas con la cuenta de consumidor LF-Tag B.

- Consulta tu metaalmacén de Apache Hive con permisos AWS Lake Formation

Administrar los permisos de Lake Formation

Lake Formation proporciona controles de acceso centralizados para los datos de su lago de datos. Puede definir reglas basadas en políticas de seguridad para sus usuarios y aplicaciones por roles en Lake Formation, y la integración con AWS Identity and Access Management autentica a esos usuarios y roles. Una vez definidas las reglas, Lake Formation aplica sus controles de acceso en el nivel de tabla, columna y fila para los usuarios de Amazon Redshift Spectrum y Amazon Athena.

Temas

- Conceder permisos de ubicación de datos
- Concesión de permisos sobre los recursos del Catálogo de datos
- Ejemplo de escenario de permisos
- Filtrado de datos y seguridad de celda en Lake Formation
- Consulta de los permisos de bases de datos y tablas en Lake Formation
- Revocación de permisos mediante la consola de Lake Formation
- Cross-account intercambio de datos en Lake Formation
- Acceso y visualización de tablas y bases de datos compartidas del Catálogo de datos
- Creación de enlaces de recursos
- Acceso a las tablas entre regiones

Conceder permisos de ubicación de datos

Los permisos de ubicación de datos AWS Lake Formation permiten a los directores crear y modificar los recursos del catálogo de datos que apuntan a ubicaciones registradas designadas de Amazon S3. Los permisos de ubicación de datos funcionan con los permisos de datos de Lake Formation para asegurar la información en su lago de datos.

Lake Formation no utiliza el servicio AWS Resource Access Manager (AWS RAM) para conceder permisos de ubicación de datos, por lo que no es necesario aceptar invitaciones para compartir recursos para obtener permisos de ubicación de datos.

Puede conceder permisos de localización de datos utilizando la consola de Lake Formation, la API o la AWS Command Line Interface (AWS CLI).

> **Note**
>
> Para que una concesión tenga éxito, primero debe registrar la ubicación de los datos en Lake Formation.

> **Consulte también:**
>
> - Underlying data access control

Temas

- Concesión de permisos de localización de datos (misma cuenta)
- Concesión de permisos de ubicación de datos (cuenta externa)
- Concesión de permisos sobre una ubicación de datos compartida con su cuenta

Concesión de permisos de localización de datos (misma cuenta)

Siga estos pasos para conceder permisos de ubicación de datos a las entidades principales de su cuenta AWS . Puede conceder permisos utilizando la consola de Lake Formation, la API o la AWS Command Line Interface (AWS CLI).

Consola de administración de AWS

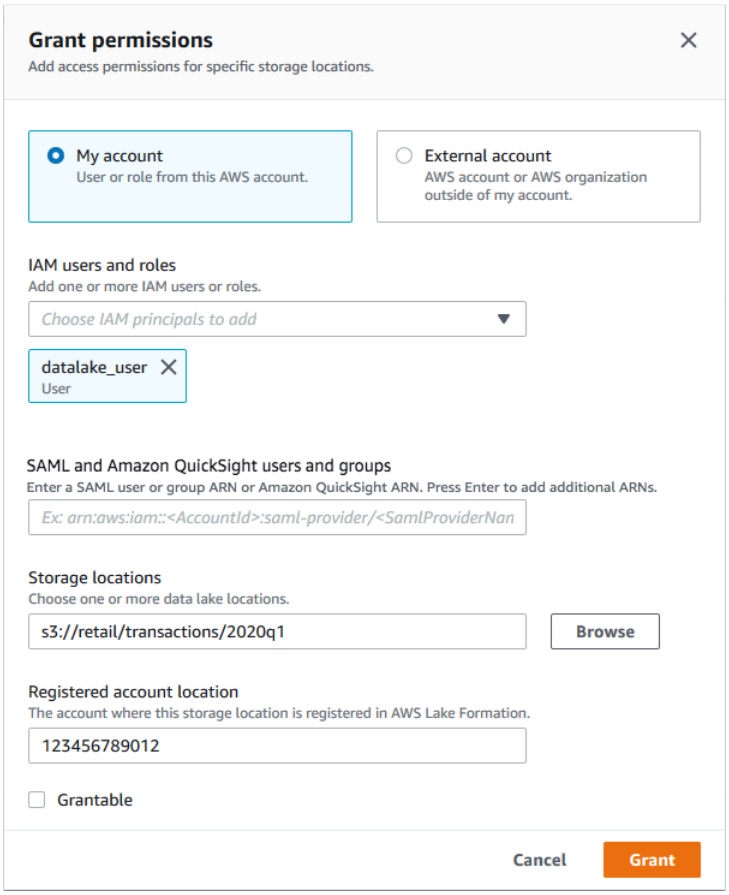
Cómo conceder permisos de ubicación de datos (misma cuenta)

1. Abre la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/. Inicie sesión como administrador del lago de datos o como una entidad principal que dispone de permisos de concesión en la ubicación de datos deseada.
2. En el panel de navegación, bajo Permisos, seleccione Ubicaciones de datos.
3. Elija Conceder.
4. En el cuadro de diálogo Conceder permisos, asegúrese de que el icono Mi cuenta esté seleccionado. A continuación, facilite la siguiente información:
   - Para Usuarios y roles de IAM, elija una o varias entidades principales.

- Para los usuarios y grupos de SAML y Amazon Quick, introduzca uno o más nombres de recursos de Amazon (ARN) para los usuarios o grupos federados a través de SAML o los ARN para los usuarios o grupos de Amazon Quick.

  Introduzca un ARN cada vez y pulse Intro después de cada uno. Para obtener información sobre cómo crear ARN, consulte Concesión y revocación de Lake Formation AWS CLI Comandos de la de.

- Para las ubicaciones de almacenamiento, elija Examinar y busque una ubicación de Amazon Simple Storage Service (Amazon S3). La ubicación debe estar registrada en Lake Formation. Vuelva a seleccionar Examinar para añadir otra ubicación. También puede escribir la ubicación, pero asegúrese de anteponerle `s3://`.
- En Ubicación de la cuenta registrada, introduzca el ID de la AWS cuenta en la que está registrada la ubicación. De forma predeterminada, es su ID de cuenta. En un escenario entre cuentas, los administradores del lago de datos de una cuenta destinataria pueden especificar aquí la cuenta propietaria al conceder el permiso de ubicación de datos a otras entidades principales de la cuenta destinataria.
- (Opcional) Para permitir que las entidades principales seleccionadas concedan permisos de ubicación de datos en la ubicación seleccionada, seleccione Concedible.



5. Elija Conceder.

AWS CLI

Cómo conceder permisos de ubicación de datos (misma cuenta)

- Ejecute un comando `grant-permissions` y conceda `DATA_LOCATION_ACCESS` a la entidad principal, especificando la ruta de Amazon S3 como recurso.

  Example

  En el siguiente ejemplo se conceden permisos de ubicación de datos en `s3://retail` al usuario `datalake_user1`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1
```

```
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail"}}'
```

Example

En el siguiente ejemplo se conceden permisos de ubicación de datos en `s3://retail` al grupo `ALLIAMPrincipals`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "111122223333"}}'
```

Consulte también:

- Referencia de permisos de Lake Formation

Concesión de permisos de ubicación de datos (cuenta externa)

Siga estos pasos para conceder permisos de ubicación de datos a una cuenta u organización externa. AWS

Puede conceder permisos utilizando la consola de Lake Formation, la API o la AWS Command Line Interface (AWS CLI).

Antes de empezar

Asegúrese de que se cumplan todos los requisitos previos de acceso entre cuentas. Para obtener más información, consulte Requisitos previos.

Consola de administración de AWS

Para conceder permisos de ubicación de datos (cuenta externa, consola)

1. Abre la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/. Inicie sesión como administrador del lago de datos.
2. En el panel de navegación, en Permisos, elija Ubicaciones de datos y, a continuación, elija Conceder.

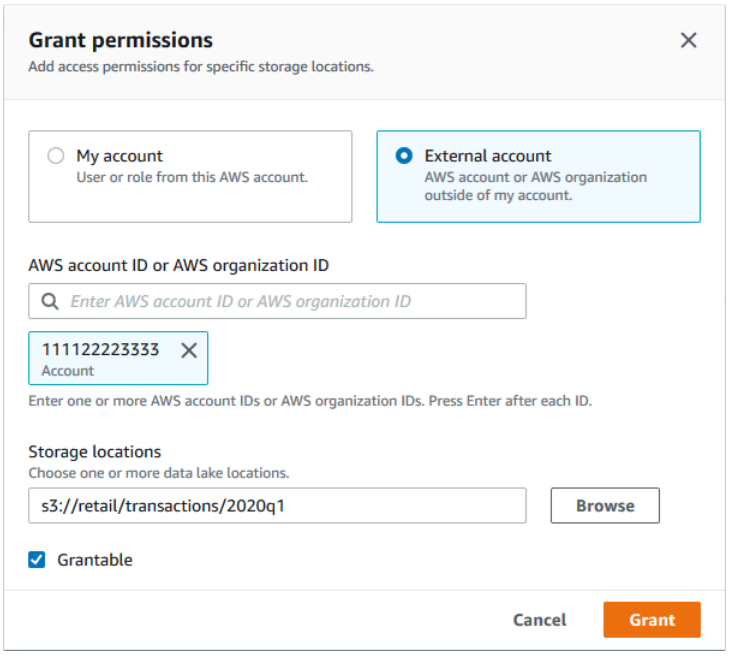
3. En el cuadro de diálogo Conceder permisos, elija el mosaico Cuenta externa.
4. Proporcione la información siguiente:

   - Para el identificador de AWS cuenta o el identificador de AWS organización, introduce números de AWS cuenta, identificadores de organización o identificadores de unidad organizativa válidos.

     Pulse Intro después de cada ID.

     El ID de una organización está formado por en "o-" seguida de 10 a 32 letras minúsculas o dígitos.

     Un ID de unidad organizativa consta de "ou-" seguido de 4 a 32 letras minúsculas o dígitos (el ID de la raíz que contiene la UO). Esta cadena va seguida de un segundo "-" (guión) y de 8 a 32 letras minúsculas o dígitos adicionales.
   - En las ubicaciones de almacenamiento, elija Examinar y busque una ubicación de almacenamiento de Amazon Simple Storage Service (Amazon S3). La ubicación debe estar registrada en Lake Formation.



5. Seleccione Concedible.
6. Elija Conceder.

AWS CLI

Para conceder permisos de ubicación de datos (cuenta externa), AWS CLI)

- Para conceder permisos a una AWS cuenta externa, introduzca un comando similar al siguiente.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333  --permissions
 "DATA_LOCATION_ACCESS" --permissions-with-grant-option
 "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"CatalogId":"123456789012","ResourceArn":"arn:aws:s3::retail/
transactions/2020q1"}}'
```

Este comando concede `DATA_LOCATION_ACCESS` con la opción de concesión a la cuenta 1111-2222-3333 de la ubicación de Amazon S3 `s3://retail/transactions/2020q1`, propiedad de la cuenta 1234-5678-9012.

Para conceder permisos a una organización, introduzca un comando similar al siguiente.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/
o-abcdefghijkl --permissions "DATA_LOCATION_ACCESS" --permissions-
with-grant-option "DATA_LOCATION_ACCESS" --resource '{"DataLocation":
 {"CatalogId":"123456789012","ResourceArn":"arn:aws:s3::retail/
transactions/2020q1"}}'
```

Este comando concede `DATA_LOCATION_ACCESS` con la opción de concesión a organización `s3://retail/transactions/2020q1` de la ubicación de Amazon S3 `o-abcdefghijkl`, propiedad de la cuenta 1234-5678-9012.

Para conceder permisos a una entidad principal en una AWS cuenta externa, introduce un comando similar al siguiente.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3::retail/transactions/2020q1", "CatalogId":
 "123456789012"}}'
```

Este comando concede DATA_LOCATION_ACCESS a una entidad principal en la cuenta 1111-2222-3333 en la ubicación de Amazon S3 s3://retail/transactions/2020q1, propiedad de la cuenta 1234-5678-9012.

Example

En el siguiente ejemplo se conceden permisos de ubicación de datos en s3://retail al grupo ALLIAMPrincipals en una cuenta externa.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "123456789012"}}'
```

Consulte también:

- Referencia de permisos de Lake Formation

Concesión de permisos sobre una ubicación de datos compartida con su cuenta

Después de compartir un recurso del catálogo de datos con su AWS cuenta, como administrador del lago de datos, puede conceder permisos sobre el recurso a otros responsables de su cuenta. Si el permiso ALTER se concede en una tabla compartida y la tabla apunta a una ubicación registrada de Amazon S3, también deberá conceder permisos de ubicación de datos en la ubicación. Del mismo modo, si se concede el permiso CREATE_TABLE o ALTER en una base de datos compartida y la base de datos tiene una propiedad de ubicación que apunta a una ubicación registrada, también deberá conceder permisos de ubicación de datos en la ubicación.

Para conceder permisos de ubicación de datos en una ubicación compartida a una entidad principal de su cuenta, ésta debe haber obtenido el permiso DATA_LOCATION_ACCESS en la ubicación con la opción de concesión. Cuando, a continuación, DATA_LOCATION_ACCESS los conceda a otro principal de su cuenta, debe incluir el ID del catálogo de datos (ID de AWS cuenta) de la cuenta del propietario. La cuenta propietaria es la cuenta que registró la ubicación.

Puede usar la AWS Lake Formation consola, la API o el AWS Command Line Interface ()AWS CLI para conceder permisos de ubicación de datos.

Para conceder permisos en una ubicación de datos compartida con su cuenta (consola)

- Siga los pasos de Concesión de permisos de localización de datos (misma cuenta).

  Para las ubicaciones de almacenamiento, debe escribirlas. En Ubicación de la cuenta registrada, introduce el AWS ID de la cuenta del propietario.

Para conceder permisos en una ubicación de datos compartida con tu cuenta (AWS CLI)

- Introduzca uno de los siguientes comandos para conceder permisos a un usuario o a un rol.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/<user-name>
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"CatalogId":"<owner-account-ID>","ResourceArn":"arn:aws:s3:::<s3-location>"}}'
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:role/<role-name>
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"CatalogId":"<owner-account-ID>","ResourceArn":"arn:aws:s3:::<s3-location>"}}'
```

Concesión de permisos sobre los recursos del Catálogo de datos

Puede conceder permisos de datos a los directores para que AWS Lake Formation puedan crear y administrar los recursos del catálogo de datos y acceder a los datos subyacentes. Puede conceder permisos de lago de datos en catálogos, bases de datos, tablas y vistas. Al conceder permisos en tablas, puede limitar el acceso a columnas o filas específicas de la tabla para un control de acceso aún más específico.

Puede conceder permisos en catálogos, bases de datos, tablas y vistas individuales, o bien con una única operación de concesión, en todas las bases de datos, tablas y vistas de un catálogo o base de datos. Si concede permisos en todas las tablas de una base de datos a entidades principales de IAM, estará concediendo implícitamente el permiso `DESCRIBE` en la base de datos. A continuación, la base de datos aparece en la página Bases de datos de la consola y es devuelta por la operación `GetDatabases` de la API. El mismo principio se aplica en el nivel de catálogo: cuando recibe

permisos para las bases de datos de un catálogo, también obtiene permisos `DESCRIBE` para ese catálogo.

> **Important**
>
> El `DESCRIBE` permiso implícito solo se aplica cuando se conceden permisos a los directores de IAM dentro de la misma cuenta. AWS En el caso de los recursos entre varias cuentas, debe conceder permisos `DESCRIBE` de forma explícita. La concesión automática del permiso `DESCRIBE` no se aplica cuando se usa el control de acceso basado en atributos (ABAC). Al conceder permisos en todas las tablas de una base de datos mediante atributos, Lake Formation no concede permisos `DESCRIBE` implícitos a la base de datos.

Puede conceder permisos mediante el método de recurso con nombre asignado o el método de control de acceso basado en etiquetas (LF-TBAC) de Lake Formation.

Puedes conceder permisos a los directores de las mismas cuentas Cuenta de AWS u organizaciones o a cuentas externas. Cuando los concede a cuentas u organizaciones externas, está compartiendo objetos del Catálogo de datos con esas cuentas u organizaciones. A continuación, las entidades principales de esas cuentas u organizaciones podrán acceder a los objetos del Catálogo de datos de su propiedad y a los datos subyacentes.

> **Note**
>
> Actualmente, el LF-TBAC método permite conceder permisos entre cuentas a directores Cuentas de AWS, organizaciones y unidades organizativas (OU) de IAM.

Al conceder permisos a cuentas u organizaciones externas, debe incluir la opción de concesión. Solo el administrador del lago de datos de la cuenta externa puede acceder a los recursos compartidos hasta que el administrador conceda permisos en los objetos compartidos a otras entidades principales de la cuenta externa.

Puede conceder permisos para el catálogo de datos mediante la AWS Lake Formation consola, la API o el AWS Command Line Interface ().AWS CLI

> **Note**
>
> Al eliminar un objeto del Catálogo de datos, todos los permisos asociados al objeto dejan de ser válidos. Si se vuelve a crear el mismo recurso con el mismo nombre, no se recuperarán los permisos de Lake Formation. Los usuarios deberán volver a configurar los permisos nuevos.

> **Véase también:**
>
> - Compartir tablas y bases de datos del catálogo de datos entre AWS Cuentas
> - Control de acceso a los metadatos
> - Referencia de permisos de Lake Formation

Permisos de IAM necesarios para conceder o revocar permisos de Lake Formation

Todos los responsables, incluido el administrador del lago de datos, necesitan los siguientes permisos AWS Identity and Access Management (IAM) para conceder o revocar los permisos del catálogo de AWS Lake Formation datos o los permisos de ubicación de datos con la API de Lake Formation o la: AWS CLI

- `lakeformation:GrantPermissions`
- `lakeformation:BatchGrantPermissions`
- `lakeformation:RevokePermissions`
- `lakeformation:BatchRevokePermissions`
- `glue:GetTable`, `glue:GetDatabase` o `glue:GetCatalog` para una tabla, base de datos o catálogo a los que esté concediendo permisos con el método de recurso con nombre.

> **Note**
>
> Los administradores del lago de datos tienen permisos implícitos de Lake Formation para conceder y revocar permisos de Lake Formation. Pero aún necesitan los permisos IAM en la concesión de Lake Formation y revocar las operaciones de API.

Los roles de IAM con políticas `AWSLakeFormationDataAdmin` AWS administradas no pueden agregar nuevos administradores de lagos de datos porque esta política contiene una denegación explícita de la operación de la API Lake Formation,`PutDataLakeSetting`.

La siguiente política de IAM se recomienda para las entidades principales que no son administradores del lago de datos y que desean conceder o revocar permisos utilizando la consola de Lake Formation.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:ListPermissions",
                "lakeformation:GrantPermissions",
                "lakeformation:BatchGrantPermissions",
                "lakeformation:RevokePermissions",
                "lakeformation:BatchRevokePermissions",
                "glue:GetCatalogs",
                "glue:GetDatabases",
                "glue:SearchTables",
                "glue:GetTables",
                "glue:GetCatalog",
                "glue:GetDatabase",
                "glue:GetTable",
                "iam:ListUsers",
                "iam:ListRoles",
                "sso-directory:DescribeUser",
                "sso-directory:DescribeGroup",
                "sso:DescribeInstance"
            ],
            "Resource": "*"
        }
    ]
}
```

Todos los `iam:` permisos `glue:` de esta política están disponibles en la política AWS `AWSGlueConsoleFullAccess` gestionada.

Para conceder permisos mediante el control de acceso basado en etiquetas (LF-TBAC) de Lake Formation, los directores necesitan permisos de IAM adicionales. Para obtener más información, consulte Prácticas recomendadas y consideraciones sobre el control de acceso basado en etiquetas de Lake Formation y Personas de Lake Formation y referencia de permisos IAM.

Cross-account permisos

Los usuarios que quieran conceder permisos de Lake Formation para varias cuentas mediante el método de recurso con nombre asignado también deben tener los permisos en la política `AWSLakeFormationCrossAccountManager` AWS gestionada.

Los administradores del lago de datos necesitan esos mismos permisos para conceder permisos entre cuentas, además del permiso AWS Resource Access Manager (AWS RAM) para permitir la concesión de permisos a las organizaciones. Para obtener más información, consulte Permisos de administrador del lago de datos.

El usuario administrativo

Un director con permisos administrativos (por ejemplo, con la política `AdministratorAccess` AWS gestionada) tiene permisos para conceder permisos a Lake Formation y crear administradores de lagos de datos. Para denegar a un usuario o rol el acceso a las operaciones del administrador de Lake Formation, adjunte o añada a su política una instrucción `Deny` para las operaciones de la API del administrador.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Action": [
                "lakeformation:GetDataLakeSettings",
                "lakeformation:PutDataLakeSettings"
            ],
            "Effect": "Deny",
            "Resource": [
                "*"
```

```
            ]
        }
    ]
}
```

> **Important**
>
> Para evitar que los usuarios se añadan como administradores con un script de extracción, transformación y carga (ETL), asegúrese de que todos los usuarios y roles que no sean administradores tengan denegado el acceso a estas operaciones de la API. La política `AWSLakeFormationDataAdmin` AWS gestionada contiene una denegación explícita de la operación de la API de Lake Formation, `PutDataLakeSetting` que impide que los usuarios agreguen nuevos administradores de lagos de datos.

Concesión de permisos de datos mediante el método de recurso con nombre

El método de recurso con nombre del Catálogo de datos es una forma de conceder permisos a los objetos de AWS Glue Data Catalog , como catálogos, bases de datos, tablas, columnas y vistas, utilizando un enfoque centralizado. Le permite definir políticas basadas en recursos que controlan el acceso a recursos específicos del lago de datos.

Al utilizar el método de recurso con nombre para conceder permisos, puede especificar el tipo de recurso y los permisos que desea conceder o revocar para dicho recurso. También puede revocar el permiso más tarde si es necesario, eliminando así los permisos de los recursos asociados.

Puede conceder permisos mediante la AWS Lake Formation consola, las API o el AWS Command Line Interface (AWS CLI).

Temas

- Concesión de permisos de catálogo mediante el método de recurso con nombre
- Concesión de permisos de base de datos mediante el método de recurso con nombre
- Concesión de permisos de tabla mediante el método de recursos con nombre
- Concesión de permisos para vistas mediante el método de recursos con nombre

Concesión de permisos de catálogo mediante el método de recurso con nombre

Los siguientes pasos explican cómo conceder permisos de catálogo utilizando el método de recurso con nombre.

Console

Utilice la página Conceder permisos en la consola de Lake Formation. La página está dividida en las secciones siguientes:

- Tipo de entidad principal: puede conceder permisos a entidades principales específicas o utilizar etiquetas de atributos.
  - Entidades principales – Los usuarios de IAM, roles, usuarios y grupos de IAM Identity Center, usuarios y grupos de SAML, cuentas AWS , organizaciones u unidades organizativas para conceder permisos.

    Entidad principal por atributos: agregue pares de clave-valor de etiquetas desde las etiquetas de sesión de IAM o IAMroles. Las entidades principales con atributos coincidentes reciben acceso al recurso especificado.
  - LF-Tags o recursos del catálogo: los catálogos, las bases de datos, las tablas, las vistas o los enlaces a los recursos para conceder permisos.
  - Permisos. Los permisos de Lake Formation que se conceden.

> **Note**
>
> Para conceder permisos en un enlace a un recurso de base de datos, consulte Conceder permisos de enlace de recursos.

1. Abra la página Conceder permisos.

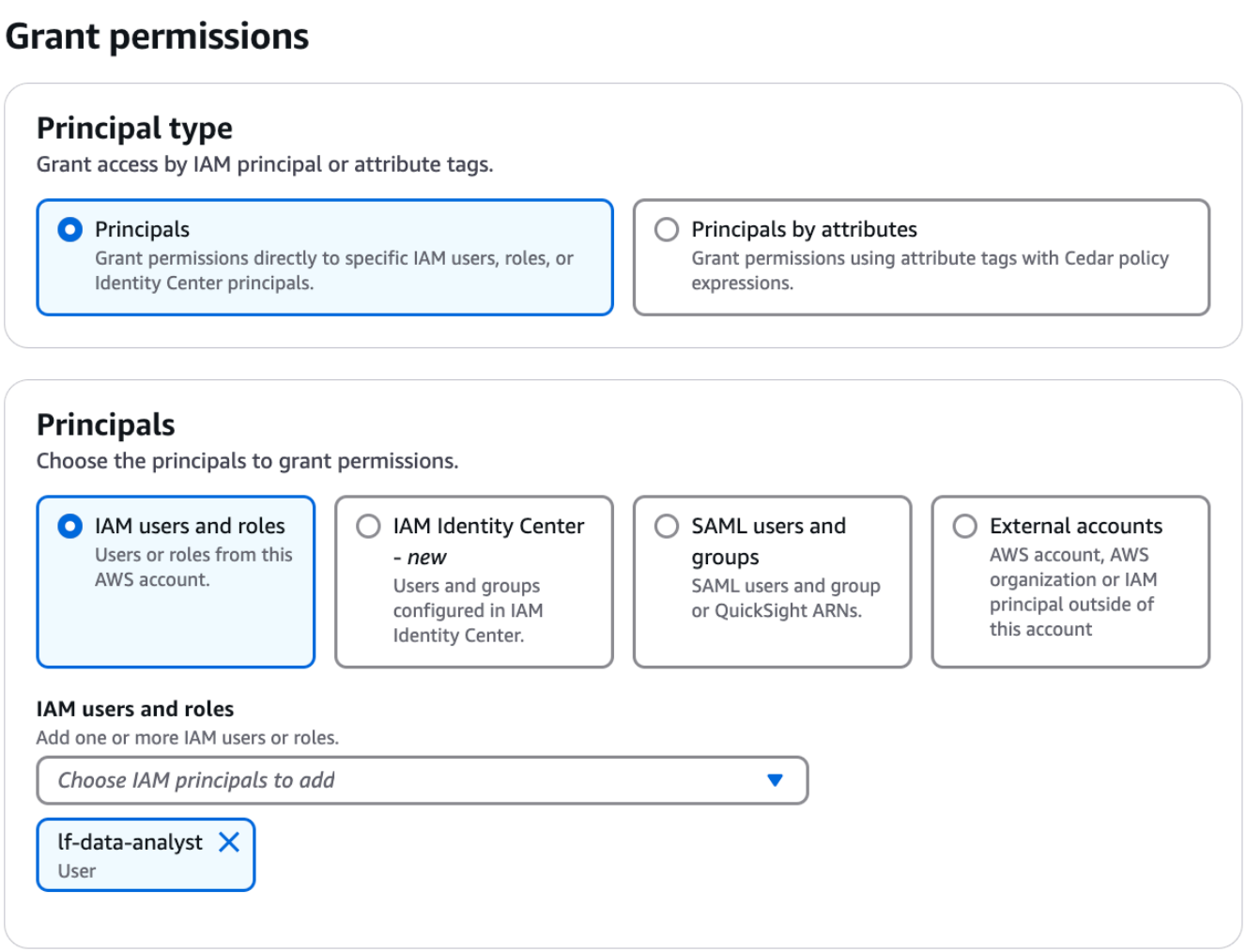
   Abre la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/e inicia sesión como administrador del lago de datos, creador del catálogo o usuario de IAM que tenga permisos de Grantable en el catálogo.

   Realice una de las siguientes acciones:

- En el panel de navegación, en Permisos, seleccione Permisos de datos. A continuación, seleccione Conceder.
- En el panel de navegación, elija Catálogos en Catálogo de datos. A continuación, en la página Catálogos, elija un catálogo y, en el menú Acciones, en Permisos, seleccione Conceder.

> **Note**
>
> Puede conceder permisos en un catálogo a través de su enlace de recursos. Para ello, en la página Catálogos, elija un contenedor de enlace de recursos y, en el menú Acciones, Conceder en el destino. Para obtener más información, consulte Cómo funcionan los enlaces de recursos en Lake Formation.

2. A continuación, en la sección Tipo de entidad principal, elija las entidades principales o especifique los atributos asociados a las entidades principales.



Especifique los directores

Usuarios y roles de IAM

Elija uno o varios usuarios o roles en la lista de usuarios y roles de IAM.

IAM Identity Center

Elija uno o varios usuarios o grupos en la lista de Usuarios y grupos. Seleccione Añadir para añadir más usuarios o grupos.

Usuarios y grupos de SAML

Para los usuarios y grupos de SAML y Quick, introduzca uno o más nombres de recursos de Amazon (ARN) para los usuarios o grupos federados a través de SAML, o ARN para los usuarios o grupos de Amazon Quick. Pulse Intro después de cada ARN.

Para obtener información sobre cómo crear ARN, consulte Concesión y revocación de Lake Formation AWS CLI Comandos de la de.

> **Note**
>
> La integración de Lake Formation con Quick solo es compatible con Quick Enterprise Edition.

Cuentas externas

Para Cuenta de AWS AWS la organización o el director de IAM, introduzca uno o más ID de AWS cuenta, ID de organización, ID de unidad organizativa o ARN válidos para el usuario o rol de IAM. Pulse Intro después de cada ID.

El ID de una organización consta de una «o-» seguida de 10 a 32 letras minúsculas o dígitos.

Un ID de una unidad organizativa comienza por «ou-» seguidos de 4 a 32 letras minúsculas o dígitos (el ID de la raíz que contiene la UO). Esta cadena va seguida de un segundo guion «-» y de 8 a 32 letras minúsculas o dígitos adicionales.

Entidades principales por atributos

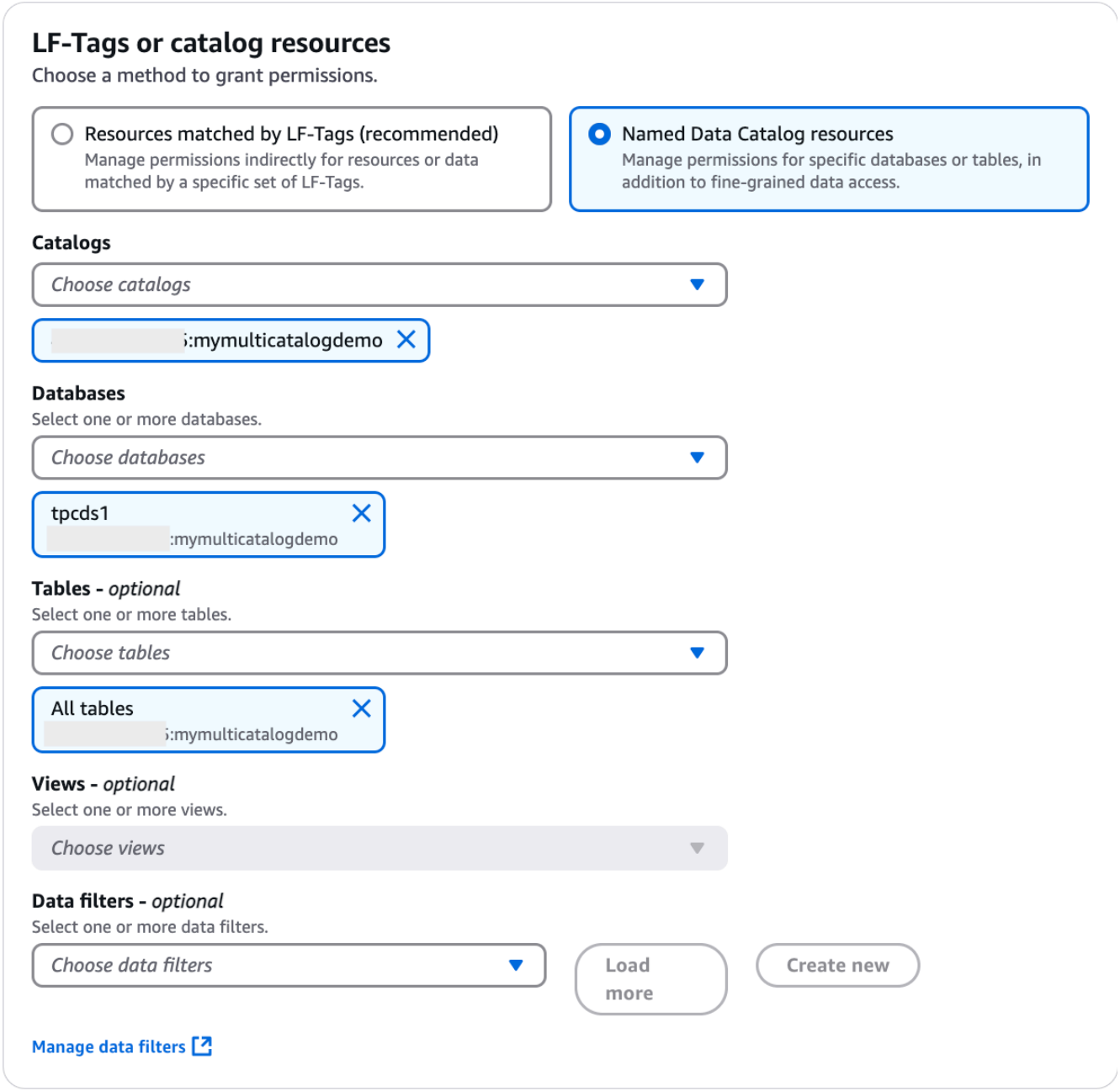
Atributos

Agregue los pares de clave-valor de etiquetas de IAM desde el rol de IAM.

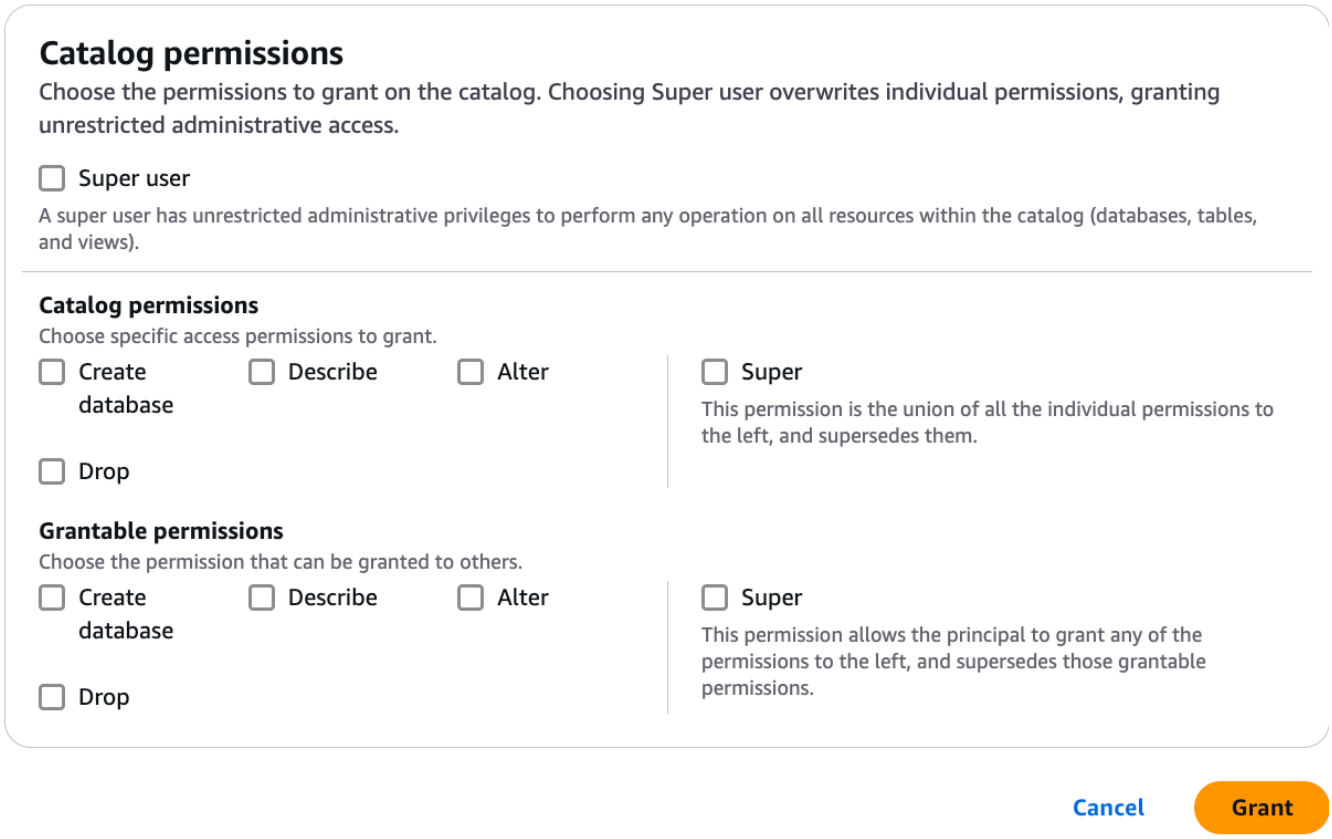
Ámbito de los permisos

Especifique si va a conceder permisos a las entidades principales con atributos coincidentes en la misma cuenta o en otra cuenta diferente.

3. En la sección LF-Tags o recursos del catálogo, elija Recursos del catálogo de datos con nombre asignado.



4. Elija uno o más catálogos de la lista Catálogos. También puede elegir una o más bases de datos, tablas o filtros and/or de datos.
5. En la sección Permisos de catálogo, seleccione los permisos y los permisos concesibles. En Permisos de catálogo, seleccione uno o más permisos para conceder.



Elija Superusuario para conceder privilegios administrativos ilimitados para realizar cualquier operación en todos los recursos del catálogo (bases de datos, tablas y vistas).

> **Note**
>
> Después de conceder `Create database` o `Alter` en un catálogo que tenga una propiedad de ubicación que apunte a una ubicación registrada, asegúrese de conceder también permisos de ubicación de datos en la ubicación a las entidades principales. Para obtener más información, consulte Conceder permisos de ubicación de datos.

6. (Opcional) En Permisos concedibles, seleccione los permisos que el destinatario de la concesión puede conceder a otras entidades principales de su cuenta AWS . Esta opción no es compatible cuando se conceden permisos a una entidad principal de IAM desde una cuenta externa.
7. Elija Conceder.

La página Permisos de datos muestra los detalles de los permisos. Si ha utilizado la opción Entidades principales por atributos para conceder permisos, puede ver los permisos concedidos a `ALLPrincipals` en la lista.

AWS CLI

Para conceder permisos de uso al catálogo AWS CLI, consulte[Creación de catálogos federados de Amazon Redshift](#).

Concesión de permisos de base de datos mediante el método de recurso con nombre

Los siguientes pasos explican cómo conceder permisos de base de datos utilizando el método de recursos con nombre.

Console

Utilice la página Conceder permisos en la consola de Lake Formation. La página está dividida en las secciones siguientes:

- Tipo de entidad principal: la sección Entidades principales incluye los usuarios de IAM, roles, usuarios y grupos de IAM Identity Center, usuarios y grupos de SAML, cuentas de AWS , organizaciones u unidades organizativas para conceder permisos. En la sección Entidades principales por atributos, puede especificar la clave y los valores de los atributos asociados a los roles de IAM.
- LF-Tags o recursos del catálogo: las bases de datos, las tablas, las vistas o los enlaces a los recursos para conceder permisos.
- Permisos. Los permisos de Lake Formation que se conceden.

> **Note**
>
> Para conceder permisos en un enlace a un recurso de base de datos, consulte [Conceder permisos de enlace de recursos](#).

1. Abra la página Conceder permisos.

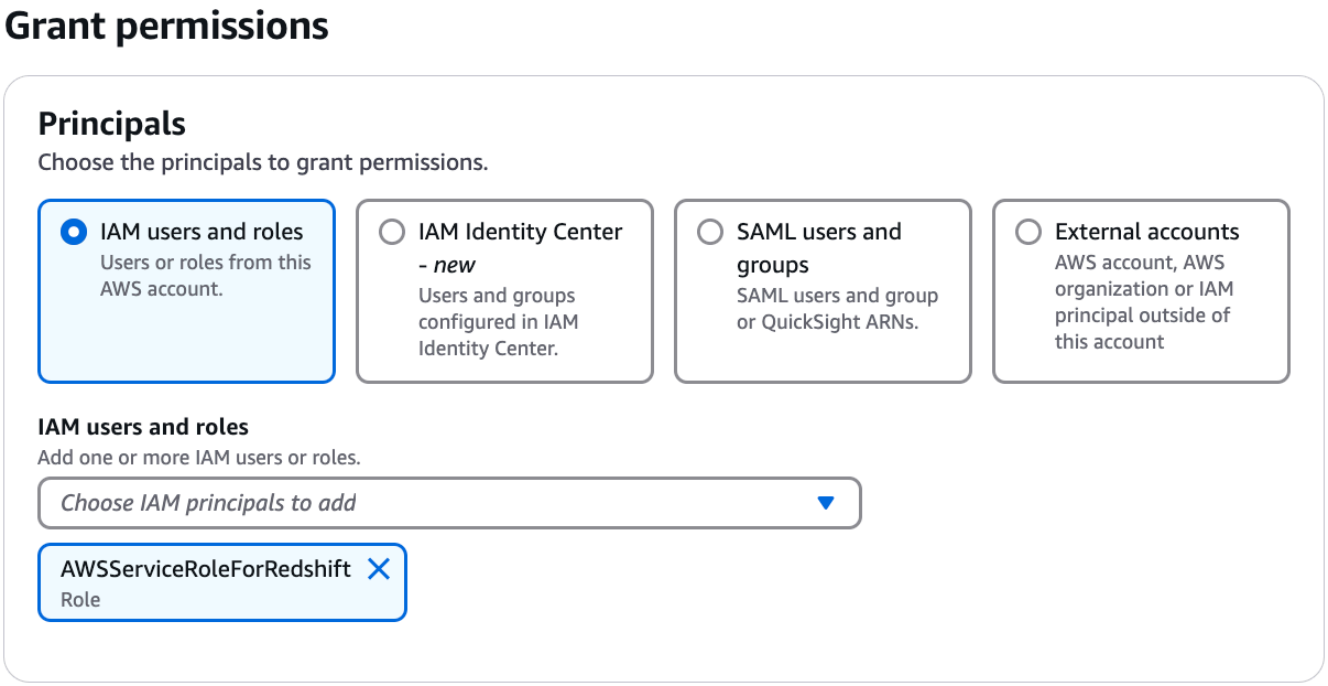
Abre la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/e inicia sesión como administrador del lago de datos, creador de la base de datos o usuario de IAM que tenga permisos de Grantable en la base de datos.

Realice una de las siguientes acciones:

- En el panel de navegación, en Permisos, seleccione Permisos de datos. A continuación, seleccione Conceder.
- En el panel de navegación, elija Bases de datos en Catálogo de datos. A continuación, en la página Bases de datos, elija una y, en el menú Acciones, en Permisos, seleccione Conceder.

> **Note**
>
> Puede conceder permisos en una base de datos desde su enlace de recursos. Para ello, en la página Bases de datos, elija un enlace de recursos y, en el menú Acciones, Conceder en el destino. Para obtener más información, consulte Cómo funcionan los enlaces de recursos en Lake Formation.

2. En la sección Tipo de entidad principal, especifique las entidades principales o conceda permisos a las entidades principales mediante atributos.



Usuarios y roles de IAM

Elija uno o varios usuarios o roles en la lista de usuarios y roles de IAM.

IAM Identity Center

Elija uno o varios usuarios o grupos en la lista de Usuarios y grupos. Seleccione Añadir para añadir más usuarios o grupos.

Usuarios y grupos de SAML

Para los usuarios y grupos de SAML y Quick, introduzca uno o más nombres de recursos de Amazon (ARN) para los usuarios o grupos federados a través de SAML, o ARN para los usuarios o grupos de Amazon Quick. Pulse Intro después de cada ARN.

Para obtener información sobre cómo crear ARN, consulte Concesión y revocación de Lake Formation AWS CLI Comandos de la de.

> **Note**
>
> La integración de Lake Formation con Quick solo es compatible con Quick Enterprise Edition.

Cuentas externas

Para Cuenta de AWS AWS la organización o el director de IAM, introduzca uno o más ID de AWS cuenta, ID de organización, ID de unidad organizativa o ARN válidos para el usuario o rol de IAM. Pulse Intro después de cada ID.

El ID de una organización consta de una «o-» seguida de 10 a 32 letras minúsculas o dígitos.

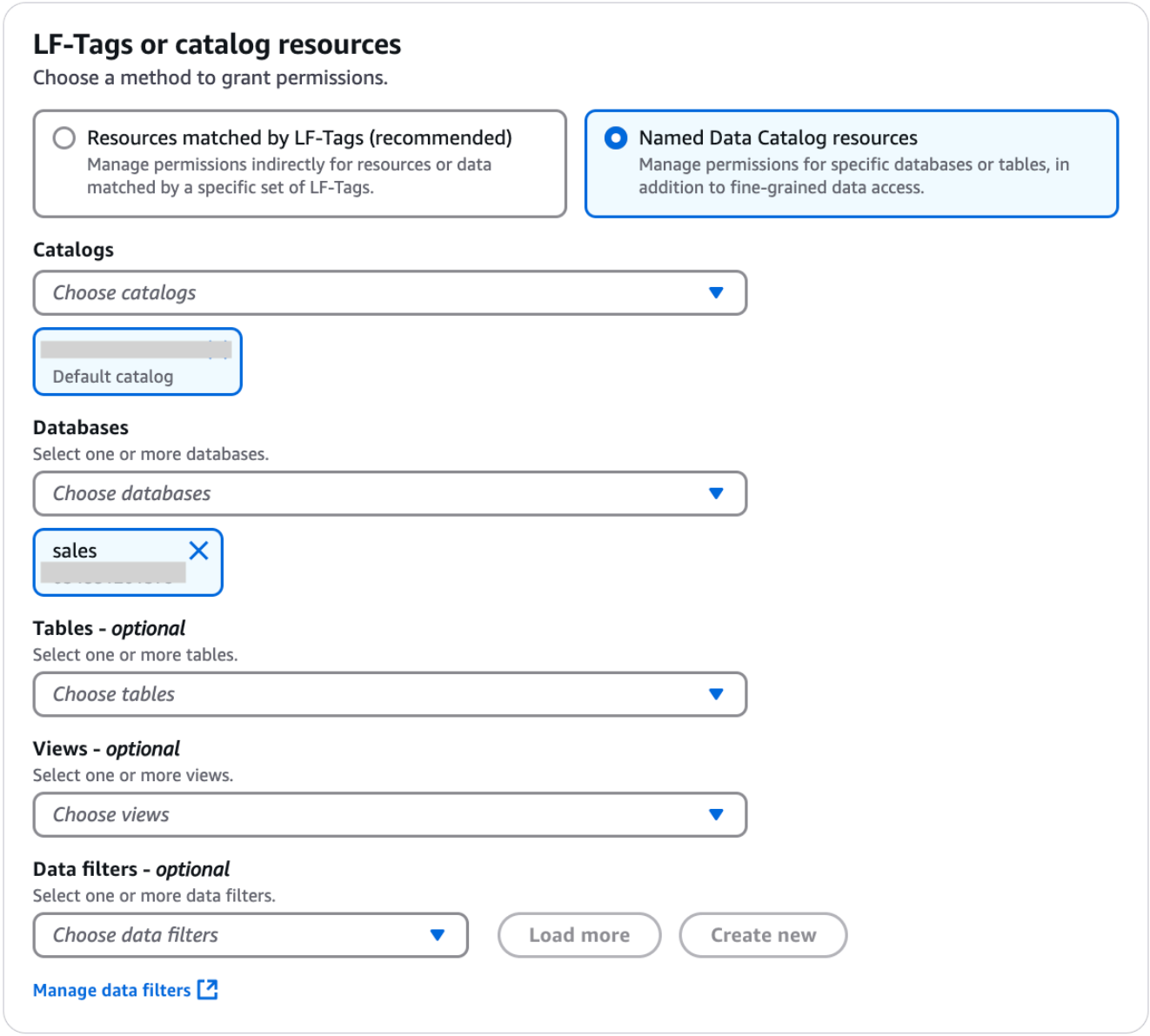
Un ID de una unidad organizativa comienza por «ou-» seguidos de 4 a 32 letras minúsculas o dígitos (el ID de la raíz que contiene la UO). Esta cadena va seguida de un segundo guion «-» y de 8 a 32 letras minúsculas o dígitos adicionales.

Entidades principales por atributos

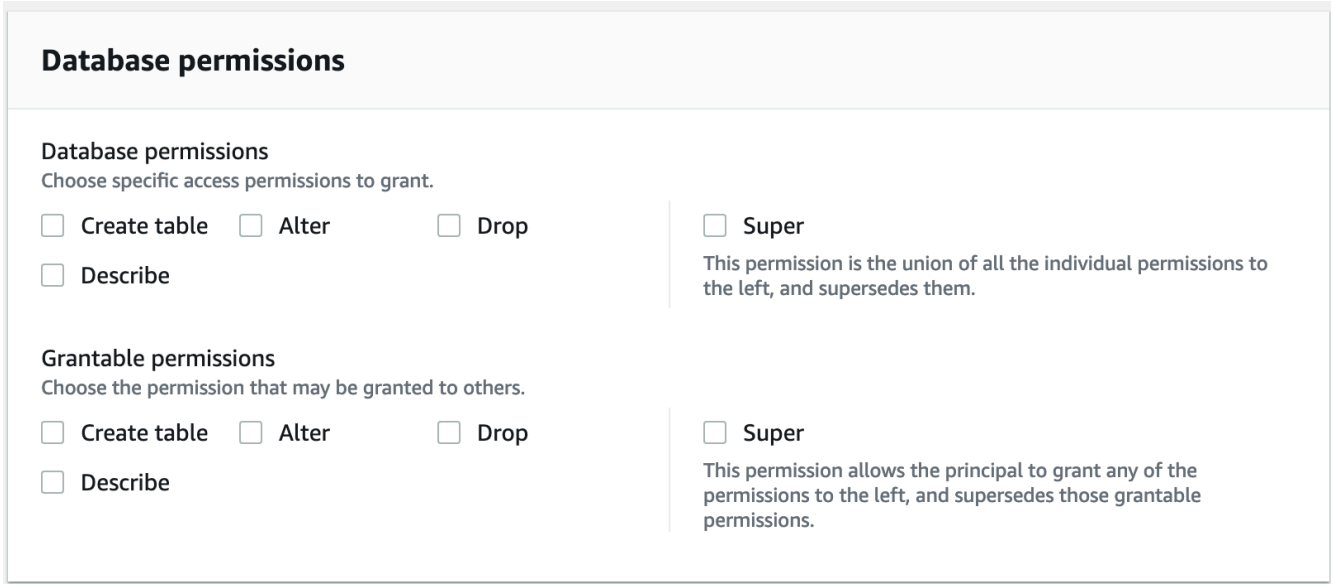
Especifique la clave y el valor o los valores del atributo. Si elige más de un valor, está creando una expresión de atributo con un operador O. Esto significa que si alguno de los valores de las etiquetas de atributo asignados a un rol o usuario de IAM coincide, role/user obtiene permisos de acceso al recurso.

Elija el alcance del permiso especificando si va a conceder permisos a las entidades principales con atributos coincidentes en la misma cuenta o en otra cuenta diferente.

3. En la sección LF-Tags o recursos del catálogo, elija Recursos del catálogo de datos con nombre asignado.



4. Elija una o más bases de datos de la lista Bases de datos. También puede elegir uno o más filtros de and/or datos de tablas.
5. En la sección Permisos, seleccione los permisos y los permisos concedibles. En Permisos de base de datos, seleccione uno o más permisos para conceder.



> **Note**
>
> Después de conceder `Create Table` o `Alter` en una base de datos que tenga una propiedad de ubicación que apunte a una ubicación registrada, asegúrese de conceder también permisos de ubicación de datos en la ubicación a las entidades principales. Para obtener más información, consulte Conceder permisos de ubicación de datos.

6. (Opcional) En Permisos concedibles, seleccione los permisos que el destinatario de la concesión puede conceder a otras entidades principales de su cuenta AWS . Esta opción no es compatible cuando se conceden permisos a una entidad principal de IAM desde una cuenta externa.
7. Elija Conceder.

AWS CLI

Para conceder permisos a la base de datos, utilice el método de recursos con nombre y la AWS Command Line Interface (AWS CLI).

Para conceder permisos a la base de datos mediante el AWS CLI

- Ejecute un comando `grant-permissions` y especifique como recurso una base de datos o el Catálogo de datos, en función del permiso que se conceda.

En los ejemplos siguientes, *<account-id>* sustitúyalo por un identificador de AWS cuenta válido.

Example— Concesión para crear una base de datos

En este ejemplo se concede CREATE_DATABASE al usuario datalake_user1. Dado que el recurso sobre el que se concede este permiso es el Catálogo de datos, la orden especifica una estructura CatalogResource vacía como parámetro resource.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

Example— Concesión para crear tablas en una base de datos designada

En el siguiente ejemplo, se concede CREATE_TABLE sobre la base de datos retail al usuario datalake_user1.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Example— Concesión a una entidad externa AWS cuenta con la opción de concesión

El siguiente ejemplo concede CREATE_TABLE con la opción de concesión de la base de datos retail a la cuenta externa 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "CREATE_TABLE"
 --permissions-with-grant-option "CREATE_TABLE" --resource '{ "Database":
 {"Name":"retail"}}'
```

Example— Concesión a una organización

En el siguiente ejemplo se concede ALTER con la opción de concesión en la base de datos issues a la organización o-abcdefghijkl.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/
```

```
o-abcdefghijkl --permissions "ALTER" --permissions-with-grant-option "ALTER" --
resource '{ "Database": {"Name":"issues"}}'
```

Example- Subvenciona a AlliamPrincipals en la misma cuenta

En el siguiente ejemplo se concede permiso de `CREATE_TABLE` sobre la base de datos `retail` a todas las entidades principales de la misma cuenta. Esta opción permite a todas las entidades principales de la cuenta crear una tabla en la base de datos y crear un enlace de recursos de tabla que permita a los motores de consulta integrados acceder a las bases de datos y tablas compartidas. Esta opción es especialmente útil cuando una entidad principal recibe una concesión entre cuentas y no tiene permiso para crear enlaces de recursos. En este escenario, el administrador del lago de datos puede crear una base de datos de marcadores de posición y conceder permiso `CREATE_TABLE` al grupo `ALLIAMPrincipal`, lo que permitirá a todas las entidades principales de IAM de la cuenta crear enlaces de recursos en la base de datos de marcadores de posición.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals
 --permissions "CREATE_TABLE"  --resource '{ "Database":
 {"Name":"temp","CatalogId":"111122223333"}}'
```

Example- Subvención a AlliamPrincipals en una cuenta externa

En el siguiente ejemplo se concede `CREATE_TABLE` sobre la base de datos `retail` a todas las entidades principales de la misma cuenta. Esta opción permite a cada entidad principal de la cuenta crear una tabla en la base de datos.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals
 --permissions "CREATE_TABLE"  --resource '{ "Database":
 {"Name":"retail","CatalogId":"123456789012"}}'
```

> **Note**
>
> Después de conceder `CREATE_TABLE` o `ALTER` en una base de datos que tenga una propiedad de ubicación que apunte a una ubicación registrada, asegúrese de conceder

también permisos de ubicación de datos en la ubicación a las entidades principales. Para obtener más información, consulte Conceder permisos de ubicación de datos.

Véase también

- Referencia de permisos de Lake Formation
- Conceder permisos en una base de datos o tabla compartida con su cuenta
- Acceso y visualización de tablas y bases de datos compartidas del Catálogo de datos

Concesión de permisos de tabla mediante el método de recursos con nombre

Puede usar la consola de Lake Formation o AWS CLI conceder permisos de Lake Formation en las tablas del catálogo de datos. Puede conceder permisos sobre tablas individuales o, con una única operación de concesión, sobre todas las tablas de una base de datos.

Si concede permisos sobre todas las tablas de una base de datos, estará concediendo implícitamente el permiso `DESCRIBE` sobre la base de datos. A continuación, la base de datos aparece en la página Bases de datos de la consola y es devuelta por la operación `GetDatabases` de la API. Esta concesión automática del permiso `DESCRIBE` no se aplica cuando se usa el control de acceso basado en atributos (ABAC). Al conceder permisos en todas las tablas de una base de datos mediante atributos, Lake Formation no concede permisos `DESCRIBE` implícitos a la base de datos.

Cuando elija `SELECT` como permiso para conceder, tendrá la opción de aplicar un filtro de columnas, de filas o de celdas.

Console

Los siguientes pasos explican cómo conceder permisos de tabla utilizando el método de recursos con nombre y la página Conceder permisos de lago de datos de la consola de Lake Formation. La página está dividida en las estas secciones:

- Tipos de entidad principal: usuarios, roles, cuentas de AWS , organizaciones o unidades organizativas a los que se conceden los permisos. También puede conceder permisos a entidades principales con atributos coincidentes.

- LF-Tags o recursos del catálogo: las bases de datos, las tablas o los enlaces a los recursos para conceder permisos.
- Permisos. Los permisos de Lake Formation que se conceden.

> **Note**
>
> Para conceder permisos sobre un enlace de recursos de tabla, consulte Conceder permisos de enlace de recursos.

1. Abra la página Conceder permisos.

   Abre la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/e inicia sesión como administrador del lago de datos, creador de tablas o usuario al que se le hayan otorgado permisos sobre la mesa con la opción de otorgarlos.

   Realice una de las siguientes acciones:

   - En el panel de navegación, en Permisos, elija Permisos de datos. A continuación, seleccione Conceder.
   - En el panel de navegación, elija Tablas. A continuación, en la página Tablas, elija una y, en el menú Acciones, Permisos, seleccione Conceder.

   > **Note**
   >
   > Puede conceder permisos sobre una tabla a través de su enlace de recursos. Para ello, en la página Tablas, elija un enlace de recursos y, en el menú Acciones, Conceder en el destino. Para obtener más información, consulte Cómo funcionan los enlaces de recursos en Lake Formation.

2. A continuación, en la sección Tipos de entidades principales, especifique entidades principales o entidades principales con atributos coincidentes para conceder permisos.

   Usuarios y roles de IAM

   Elija uno o varios usuarios o roles en la lista de usuarios y roles de IAM.

IAM Identity Center

Elija uno o varios usuarios o grupos en la lista de Usuarios y grupos.

Usuarios y grupos de SAML

Para los usuarios y grupos de SAML y Quick, introduzca uno o más nombres de recursos de Amazon (ARN) para los usuarios o grupos federados a través de SAML, o ARN para los usuarios o grupos de Quick. Pulse Intro después de cada ARN.

Para obtener información sobre cómo crear ARN, consulte Concesión y revocación de Lake Formation AWS CLI Comandos de la de.

> **Note**
>
> La integración de Lake Formation con Quick solo es compatible con Quick Enterprise Edition.

Cuentas externas

Para Cuenta de AWS AWS la organización o el director de IAM, introduzca uno o más Cuenta de AWS ID válidos, ID de organización, ID de unidad organizativa o el ARN del usuario o rol de IAM. Pulse Intro después de cada ID.

El ID de una organización consta de una «o-» seguida de 10 a 32 letras minúsculas o dígitos.

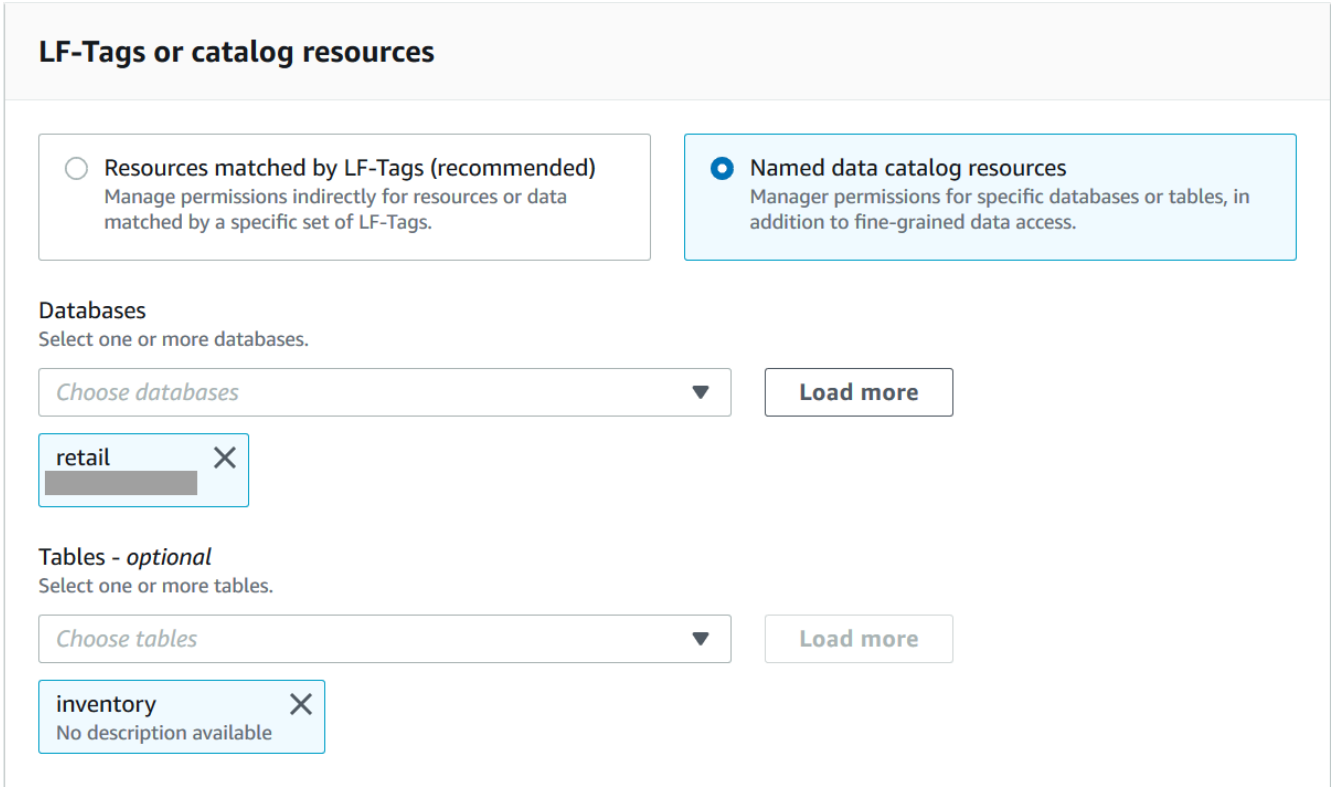
Un ID de una unidad organizativa comienza por «ou-» seguidos de 4 a 32 letras minúsculas o dígitos (el ID de la raíz que contiene la UO). Esta cadena va seguida de un segundo carácter «-» y de 8 a 32 letras minúsculas o dígitos adicionales.

Entidades principales por atributos

Especifique la clave y el valor o los valores del atributo. Si elige más de un valor, está creando una expresión de atributo con un operador O. Esto significa que si alguno de los valores de las etiquetas de atributo asignados a un rol o usuario de IAM coincide, role/user obtiene permisos de acceso al recurso

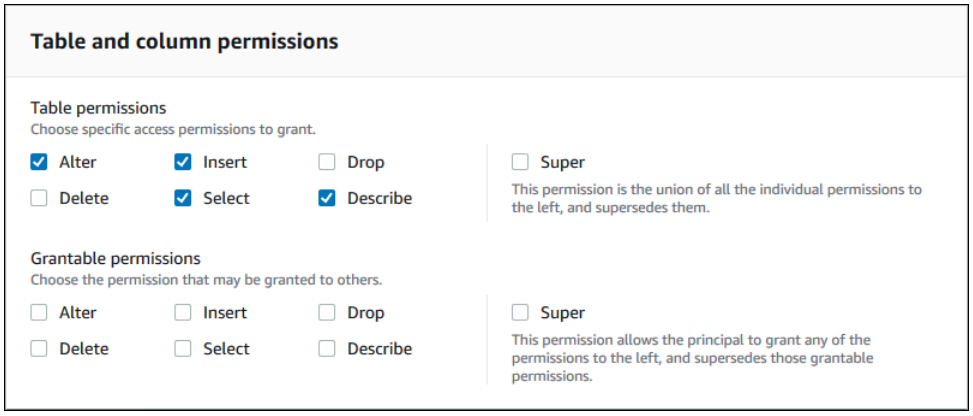
Elija el alcance del permiso especificando si va a conceder permisos a las entidades principales con atributos coincidentes en la misma cuenta o en otra cuenta diferente.

3. En la sección de recursos del catálogo LF-Tags o en la sección de recursos, elija una base de datos. A continuación, seleccione una o más tablas o Todas las tablas.

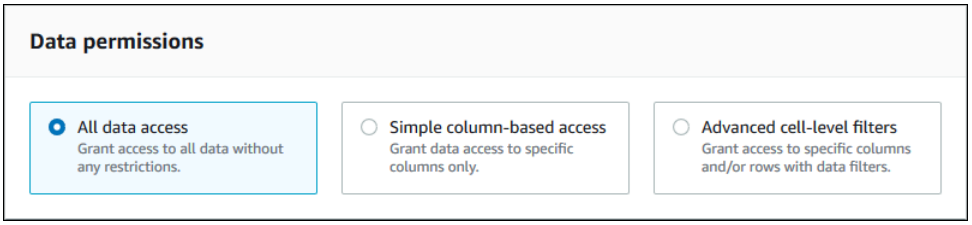


4. Especifique los permisos sin filtrado de datos.

En la sección Permisos, seleccione los permisos de tabla que desee conceder y, opcionalmente, seleccione los permisos que se pueden conceder.



Si concede Seleccionar, la sección Permisos de datos aparece debajo de la sección Permisos de tabla y columna, con la opción Todos los accesos a datos seleccionada de manera predeterminada. Acepte los valores predeterminados.



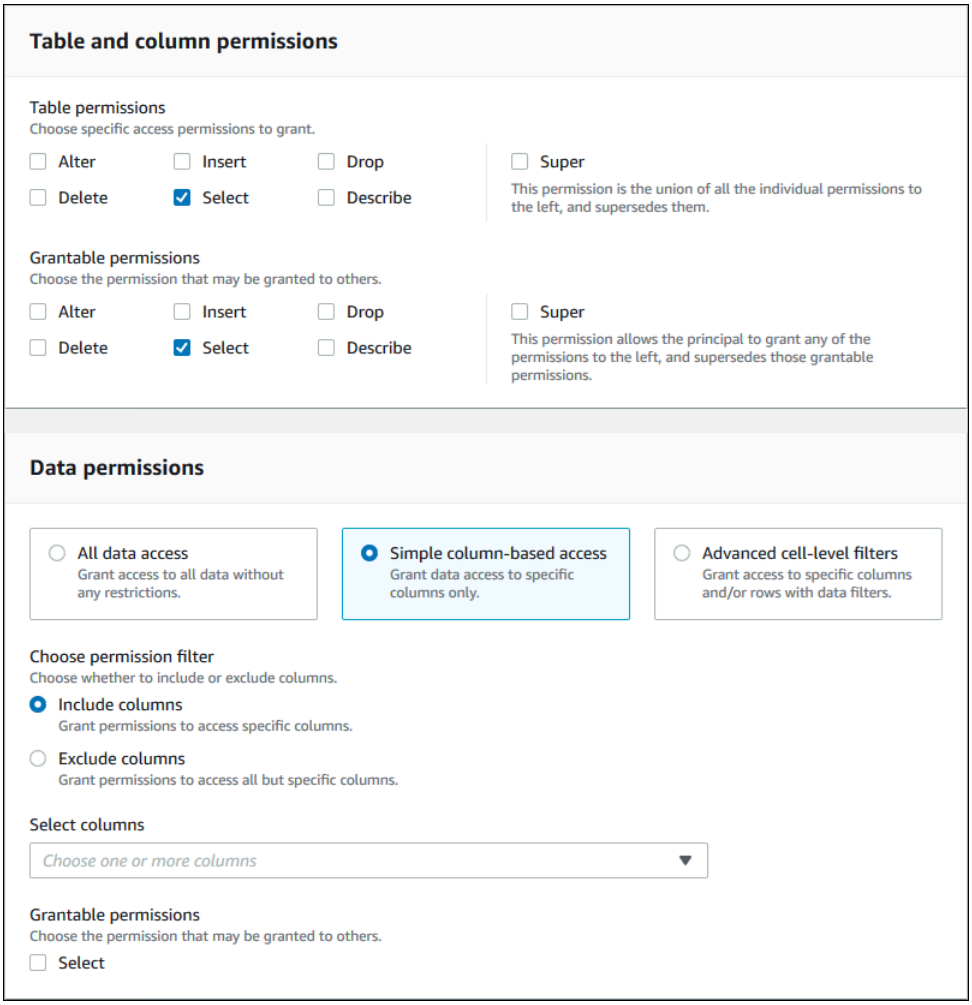
5. Elija Conceder.
6. Especifique el Select permiso con filtrado de datos

   Elija el permiso Seleccionar. No seleccione ningún otro permiso.

   La sección de permisos de datos aparece debajo de la sección de permisos de tabla y columna.
7. Realice una de las siguientes acciones:

   - Aplique solo un filtrado de columnas simple.

     1. Elija Acceso simple basado en columnas.



2. Elija si desea incluir o excluir columnas y, a continuación, elija las que desea incluir o excluir.

   Solo se permite incluir listas cuando se conceden permisos a una AWS cuenta u organización externa.

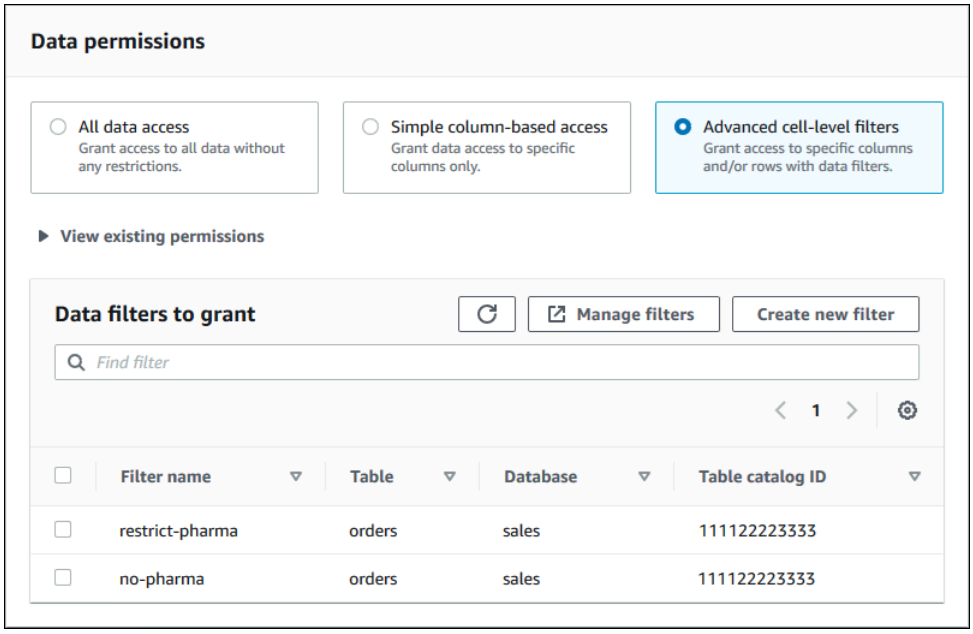
3. (Opcional) En Permisos concedibles, active la opción de concesión para el permiso Seleccionar.

   Si incluye la opción de concesión, el destinatario de esta podrá conceder permisos solo sobre las columnas que le conceda.

> **Note**
>
> También puede aplicar el filtrado por columnas solo creando un filtro de datos que especifique un filtro de columnas y especifique todas las filas como filtro de filas. Sin embargo, esto requiere más pasos.

- Aplicar filtros de columna, fila o celda.

  1. Seleccione Filtros avanzados a nivel de celda.



  2. (Opcional) Amplíe Ver los permisos existentes.
  3. (Opcional) Seleccione Crear filtro nuevo.
  4. (Opcional) Para ver los detalles de los filtros de la lista o para crear filtros nuevos o eliminar los existentes, seleccione Administrar filtros.

     La página Filtros de datos se abre en una nueva ventana del navegador.

     Cuando haya terminado de acceder a la página Filtros de datos, vuelva a la página Conceder permisos y, si es necesario, actualícela para ver los filtros de datos nuevos que haya creado.
  5. Seleccione uno o más filtros de datos para aplicarlos a la concesión.

> **Note**
>
> Si no hay filtros de datos en la lista, significa que no se creó ningún filtro de datos para la tabla seleccionada.

8. Elija Conceder.

AWS CLI

Para conceder permisos de tabla, utilice el método de recursos con nombre y la AWS Command Line Interface (AWS CLI).

Para conceder permisos de tabla mediante el AWS CLI

- Introduzca un comando `grant-permissions` y especifique una tabla como recurso.

Example. Concesión en una sola tabla, sin filtrado

El siguiente ejemplo concede `SELECT` y `ALTER` al usuario `datalake_user1` de la AWS cuenta 1111-2222-3333 de la tabla `inventory` de la base de datos. `retail`

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" "ALTER" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

> **Note**
>
> Si concede el permiso `ALTER` en una tabla cuyos datos subyacentes están en una ubicación registrada, asegúrese de conceder también permisos de ubicación de datos en la ubicación a las entidades principales. Para obtener más información, consulte Conceder permisos de ubicación de datos.

Example– Concesión en todas las tablas con la opción Concesión, sin filtrado

En el siguiente ejemplo, se concede `SELECT` con la opción de concesión en todas las tablas de la base de datos `retail`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" --permissions-with-grant-option "SELECT" --resource '{ "Table":
 { "DatabaseName": "retail", "TableWildcard": {} } }'
```

Example– Concesión con filtrado simple de columnas

El siguiente ejemplo concede SELECT en un subconjunto de columnas de la tabla `persons`. Utiliza un filtrado de columnas simple.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" --resource '{ "TableWithColumns": {"DatabaseName":"hr",
 "Name":"persons", "ColumnNames":["family_name", "given_name", "gender"]}}'
```

Example– Concesión con filtro de datos

En este ejemplo se concede SELECT en la tabla `orders` y se aplica el filtro de datos `restrict-pharma`.

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

A continuación se muestra el contenido del archivo `grant-params.json`.

```
{
    "Principal": {"DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"},
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
    "Permissions": ["SELECT"],
    "PermissionsWithGrantOption": ["SELECT"]
}
```

Véase también

- Descripción general de los permisos de Lake Formation
- Filtrado de datos y seguridad de celda en Lake Formation
- Personas de Lake Formation y referencia de permisos IAM
- Conceder permisos de enlace de recursos
- Acceso y visualización de tablas y bases de datos compartidas del Catálogo de datos

Concesión de permisos para vistas mediante el método de recursos con nombre

Los siguientes pasos explican cómo conceder permisos para vistas utilizando el método de recurso con nombre y la página Conceder permisos. La página está dividida en las secciones siguientes:

- Tipos principales: los usuarios de IAM, los roles, los usuarios y grupos del Centro de Identidad de IAM Cuentas de AWS, las organizaciones o las unidades organizativas que conceden permisos. También puede conceder permisos a entidades principales con atributos coincidentes.
- LF-Tags o recursos del catálogo: las bases de datos, las tablas, las vistas o los enlaces a los recursos para conceder permisos.
- Permisos – Los permisos de lagos de datos que se conceden.

Abra el icono Concesión de permisos page

1. Abre la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/e inicia sesión como administrador del lago de datos, creador de la base de datos o usuario de IAM que tenga permisos de Grantable en la base de datos.
2. Realice una de las siguientes acciones:
   - En el panel de navegación, en Permisos, seleccione Permisos de datos. A continuación, seleccione Conceder.
   - En el panel de navegación, en Catálogo de datos, elija Vistas. A continuación, en la página Vistas, elija una y, en el menú Acciones, Permisos, seleccione Conceder.

> **Note**
>
> Puede conceder permisos sobre una vista a través de su enlace de recursos. Para ello, en la página Vistas, elija un enlace de recursos y, en el menú Acciones, elija Conceder en el destino. Para obtener más información, consulte Cómo funcionan los enlaces de recursos en Lake Formation.

Especifique los tipos de entidades principales.

En la sección Tipos de entidades principales, elija Entidades principales o Entidades principales por atributos. Si elige Entidades principales, estarán disponibles las siguientes opciones:

Usuarios y roles de IAM

Elija uno o varios usuarios o roles en la lista de usuarios y roles de IAM.

IAM Identity Center

Elija uno o varios usuarios o grupos en la lista de Usuarios y grupos.

Usuarios y grupos de SAML

Para los usuarios y grupos de SAML y Quick, introduzca uno o más nombres de recursos de Amazon (ARN) para los usuarios o grupos federados a través de SAML, o ARN para los usuarios o grupos de Amazon Quick. Pulse Intro después de cada ARN.

Para obtener información sobre cómo crear ARN, consulte Concesión y revocación de Lake Formation AWS CLI Comandos de la de.

> **Note**
>
> La integración de Lake Formation con Quick solo es compatible con Quick Enterprise Edition.

Cuentas externas

Para Cuenta de AWS AWS la organización o el director de IAM, introduzca uno o más ID de AWS cuenta, ID de organización, ID de unidad organizativa o ARN válidos para el usuario o rol de IAM. Pulse Intro después de cada ID.

El ID de una organización consta de una «o-» seguida de 10 a 32 letras minúsculas o dígitos.

Un ID de una unidad organizativa comienza por «ou-» seguidos de 4 a 32 letras minúsculas o dígitos (el ID de la raíz que contiene la UO). Esta cadena va seguida de un segundo guion «-» y de 8 a 32 letras minúsculas o dígitos adicionales.

> Véase también
>
> - Acceso y visualización de tablas y bases de datos compartidas del Catálogo de datos

Entidades principales por atributos

Especifique la clave y el valor o los valores del atributo. Si elige más de un valor, está creando una expresión de atributo con un operador O. Esto significa que si alguno de los valores de las etiquetas de atributo asignados a un rol o usuario de IAM coincide, role/user obtiene permisos de acceso al recurso

Elija el alcance del permiso especificando si va a conceder permisos a las entidades principales con atributos coincidentes en la misma cuenta o en otra cuenta diferente.

Especifique las vistas

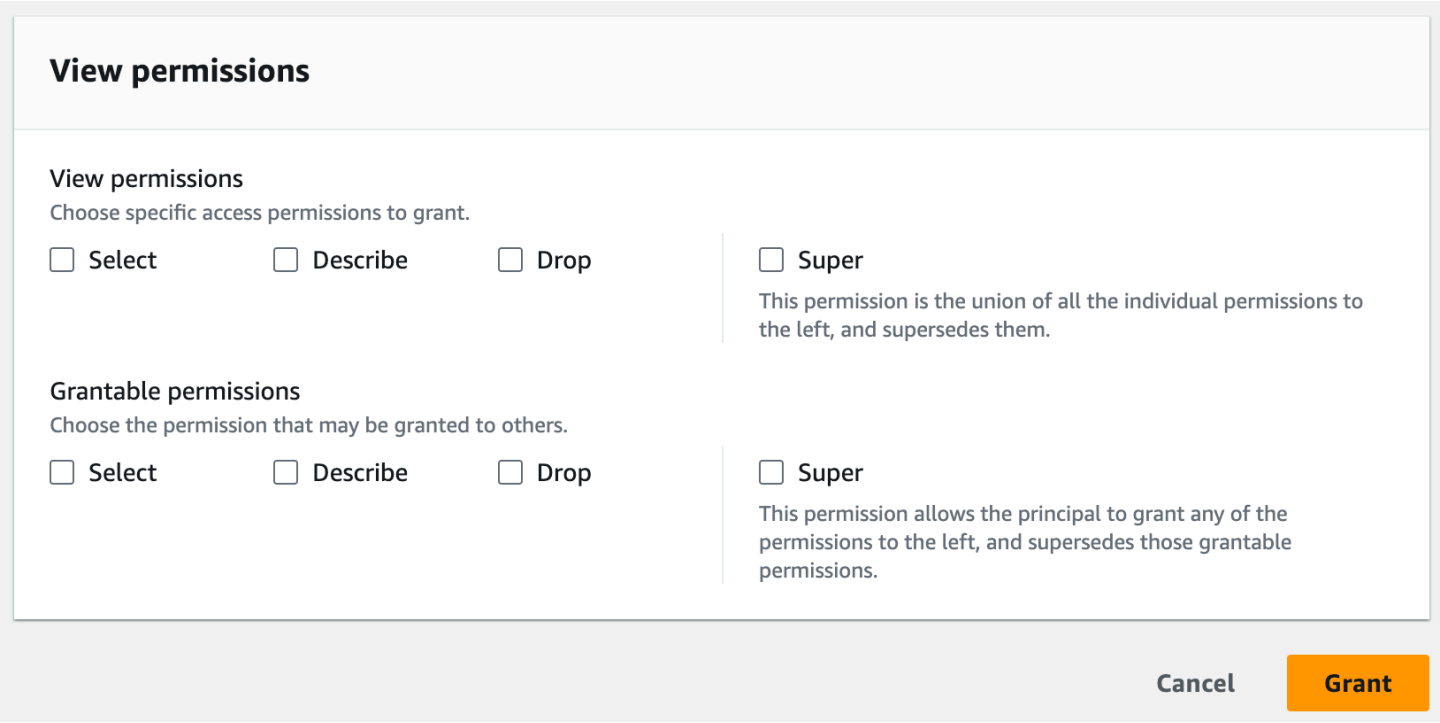
En la sección de recursos del catálogo LF-Tags o en la sección de recursos, elija una o más vistas para conceder permisos.

1. Seleccione un recurso de Catálogo de datos con nombre.
2. Elija una o más vistas de la lista Vistas. También puede elegir uno o más catálogos, bases de datos, tablas y filtros de and/or datos.

   Al conceder permisos de lago de datos a `All tables` dentro de una base de datos, el beneficiario tendrá permisos en todas las tablas y vistas de la base de datos.

Especificar los permisos

En la sección Permisos, seleccione los permisos y los permisos concedibles.



1. En Ver de permisos, seleccione uno o más permisos para conceder.
2. (Opcional) En Permisos concedibles, seleccione los permisos que el destinatario de la concesión puede conceder a otras entidades principales de su cuenta Cuenta de AWS. Esta opción no es compatible cuando se conceden permisos a una entidad principal de IAM desde una cuenta externa.
3. Elija Conceder.

Véase también

- Referencia de permisos de Lake Formation
- Conceder permisos en una base de datos o tabla compartida con su cuenta

Control de acceso basado en etiquetas de Lake Formation

El control de acceso basado en etiquetas (LF-TBAC) de Lake Formation es una estrategia de autorización que define los permisos en función de los atributos. En Lake Formation, estos atributos

se denominan LF-Tags. Puede adjuntarlos LF-Tags a los recursos del catálogo de datos y conceder permisos a los directores de Lake Formation sobre esos recursos utilizando estos LF-Tags recursos. Lake Formation permite realizar operaciones en esos recursos cuando se ha concedido acceso a la entidad principal a un valor de etiqueta que coincide con el valor de la etiqueta del recurso.

LF-TBAC es útil en entornos que crecen rápidamente y ayuda en situaciones en las que la gestión de políticas se vuelve engorrosa.

LF-TBAC es el método recomendado para conceder permisos a Lake Formation cuando hay una gran cantidad de objetos del catálogo de datos, incluidos catálogos federados, bases de datos, tablas y vistas. Lake Formation admite el control de acceso basado en etiquetas para catálogos federados de tablas, almacenes de datos y fuentes de Amazon Redshift datos federadas de Amazon S3 Amazon DynamoDB, como SQL Server y Snowflake.

> **Note**
>
> Las etiquetas de IAM no son lo mismo que. LF-Tags Estas etiquetas no son intercambiables. LF-Tagsse utilizan para conceder permisos a Lake Formation y las etiquetas de IAM se utilizan para definir las políticas de IAM.

Cómo funciona el control de acceso basado en etiquetas de Lake Formation

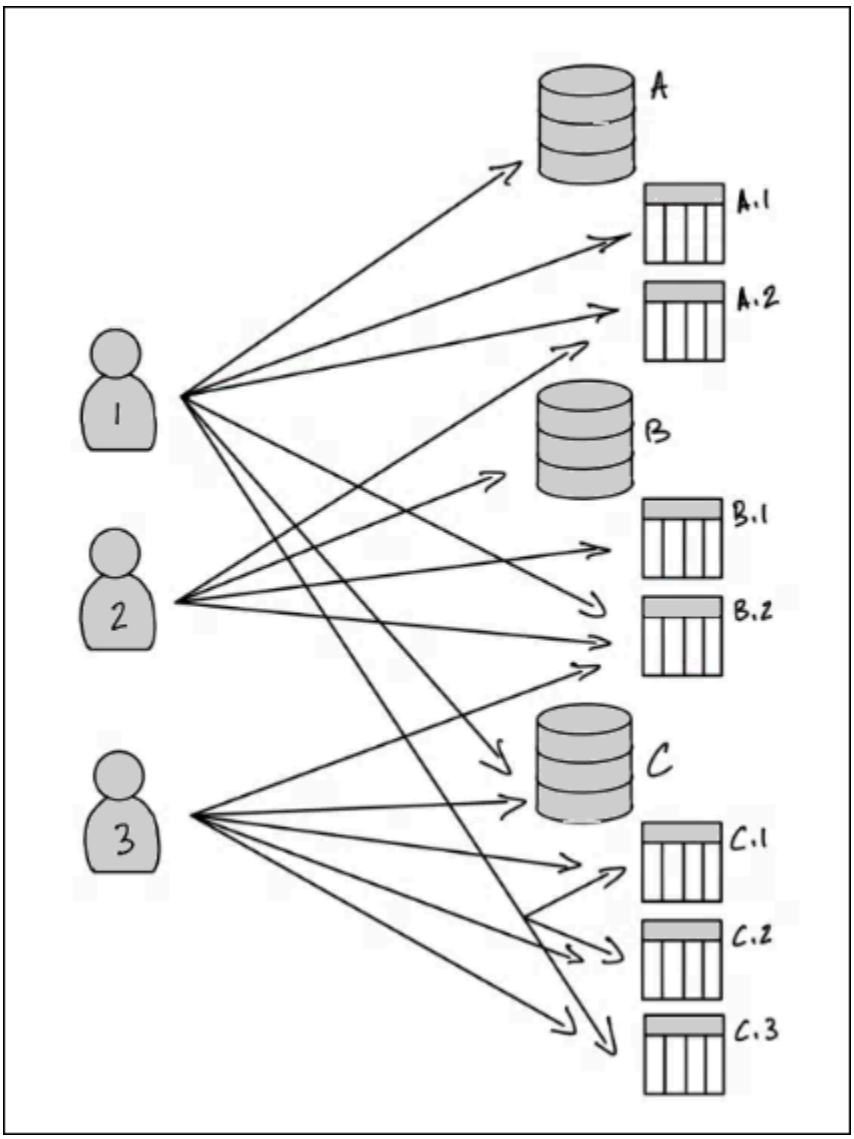
Cada uno LF-Tag es un par clave-valor, como o. `department=sales` `classification=restricted` Una clave puede tener varios valores definidos, como `department=sales,marketing,engineering,finance`.

Para usar el LF-TBAC método, los administradores de lagos de datos y los ingenieros de datos realizan las siguientes tareas.

Tarea	Detalles de la tarea
1. Defina las propiedades y relaciones de LF-Tags.	-
2. Crea a los LF-Tag creadores en Lake Formation.	Añadir creadores LF-Tag
3. Crea el LF-Tag in Lake Formation.	Creando LF-Tags

Tarea	Detalles de la tarea
4. Asigne LF-Tags a los recursos del catálogo de datos.	Asignación LF-Tags a recursos del catálogo de datos
5. Otorgue permisos a otros directores para que los asignen LF-Tags a los recursos, de forma opcional con la opción de conceder.	Administrar los permisos de LF-Tag valor
6. Otorgue LF-Tag expresiones a los directores, de forma opcional con la opción de concesión.	Otorgar permisos de lago de datos mediante el método LF-TBAC
7. (Recomendado) Tras comprobar que los directores tienen acceso a los recursos correctos mediante el LF-TBAC método, revoque los permisos concedidos mediante el método de recurso indicado.	-

Imagine un caso en el que debe conceder permisos a tres entidades principales sobre tres bases de datos y siete tablas.



Para conseguir los permisos indicados en el diagrama anterior utilizando el método de recursos con nombre, tendría que efectuar 17 concesiones, como se indica a continuación (en pseudocódigo).

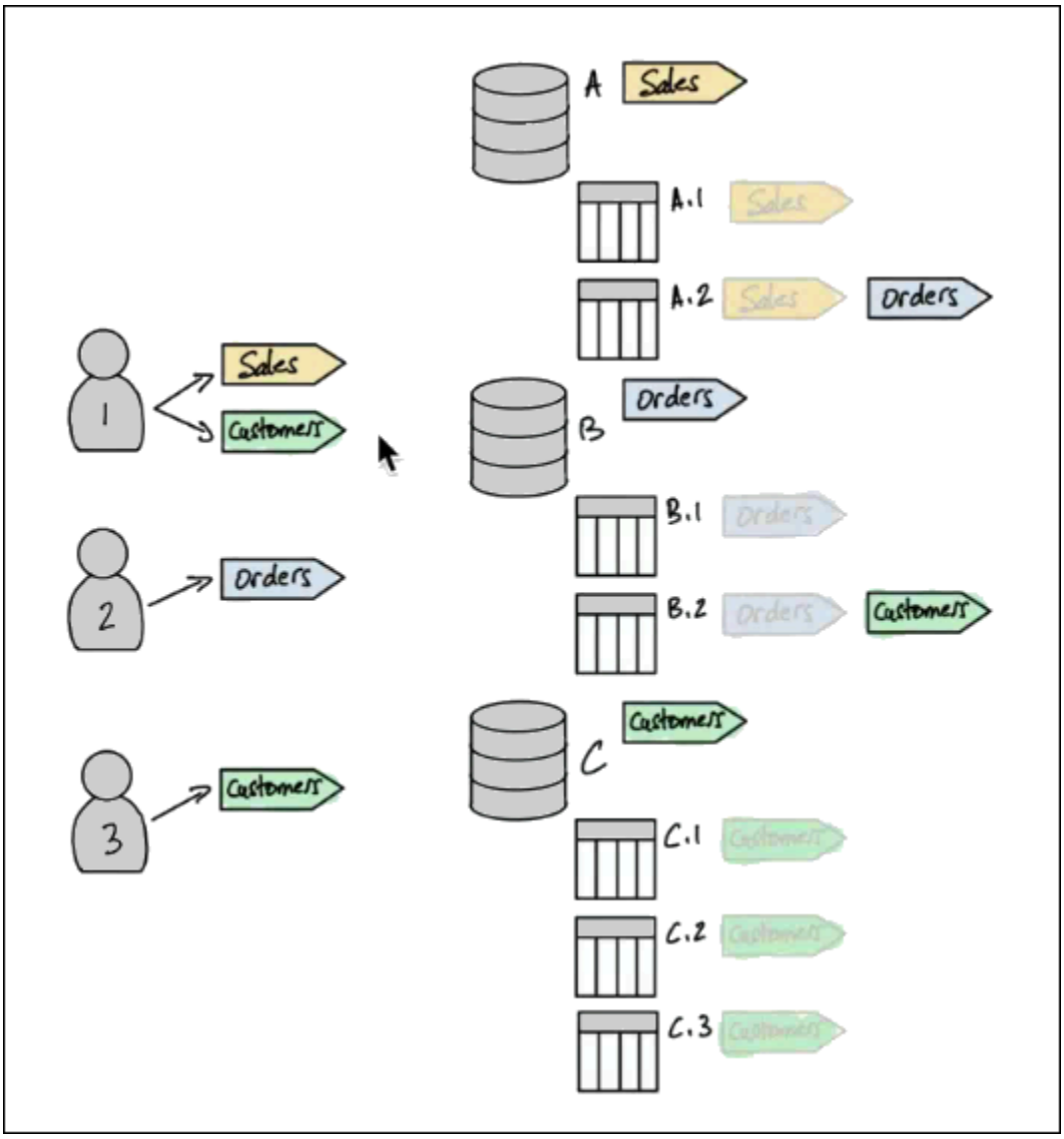
```
GRANT CREATE_TABLE ON Database A TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.1 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table B.2 TO PRINCIPAL 1
...
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 2
GRANT CREATE_TABLE ON Database B TO PRINCIPAL 2
...
GRANT SELECT, INSERT ON Table C.3 TO PRINCIPAL 3
```

Ahora considere cómo otorgaría los permisos mediante LF-TBAC. En el siguiente diagrama se indica que ha asignado LF-Tags a bases de datos y tablas, y que ha concedido permisos LF-Tags a entidades principales.

En este ejemplo, LF-Tags representan áreas del lago de datos que contienen análisis para diferentes módulos de un conjunto de aplicaciones de planificación de recursos empresariales (ERP). Se pueden utilizar para controlar el acceso a los datos analíticos de los distintos módulos. Todos LF-Tags tienen los valores clave `module` y posibles `SalesOrders`, `yCustomers`. Un ejemplo LF-Tag sería el siguiente:

```
module=Sales
```

El diagrama muestra solo los LF-Tag valores.



Asignación de etiquetas a los recursos del Catálogo de datos y herencia

Las tablas se heredan de las bases LF-Tags de datos y las columnas se heredan LF-Tags de las tablas. Los valores heredados pueden anularse. En el diagrama anterior, las atenuadas LF-Tags se heredan.

Debido a la herencia, el administrador del lago de datos solo necesita realizar las cinco LF-Tag asignaciones siguientes a los recursos (en pseudocódigo).

```
ASSIGN TAGS module=Sales TO database A
ASSIGN TAGS module=Orders TO table A.2
ASSIGN TAGS module=Orders TO database B
ASSIGN TAGS module=Customers TO table B.2
ASSIGN TAGS module=Customers TO database C
```

Etiquetar las concesiones a las entidades principales

Tras realizar la asignación LF-Tags a las bases de datos y las tablas, el administrador del lago de datos solo debe conceder cuatro veces a los directores, tal y como se indica LF-Tags a continuación (en pseudocódigo).

```
GRANT TAGS module=Sales TO Principal 1
GRANT TAGS module=Customers TO Principal 1
GRANT TAGS module=Orders TO Principal 2
GRANT TAGS module=Customers TO Principal 3
```

Ahora, un director con la `module=Sales` LF-Tag puede acceder a los recursos del catálogo de datos con la `module=Sales` LF-Tag (por ejemplo, la base de datos A), un director con la `module=Customers` LF-Tag puede acceder a los recursos con la `module=Customers` LF-Tag, etc.

Los comandos de concesión anteriores están incompletos. Esto se debe a que, si bien indican a través LF-Tags de los recursos del catálogo de datos que los directores tienen permisos, no indican exactamente qué permisos de Lake Formation (por ejemplo`SELECT`,`ALTER`) tienen los directores sobre esos recursos. Por lo tanto, los siguientes comandos de pseudocódigo son una representación más precisa de cómo se otorgan los permisos de Lake Formation a los recursos del catálogo de datos. LF-Tags

```
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Sales TO Principal 1
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Sales TO Principal 1
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Customers TO Principal 1
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Customers TO Principal 1
```

```
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Orders TO Principal 2
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Orders TO Principal 2
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Customers TO Principal 3
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Customers TO Principal 3
```

Recopilación. Permisos resultantes sobre los recursos

Teniendo en cuenta los permisos LF-Tags asignados a las bases de datos y tablas del diagrama anterior y los LF-Tags otorgados a los principales del diagrama, la siguiente tabla muestra los permisos de Lake Formation que tienen los principales en las bases de datos y tablas.

Principal	Permisos otorgados mediante LF-Tags
Entidad principal 1	• `CREATE_TABLE` sobre la base de datos A • `SELECT`, `INSERT` sobre la mesa A.1 • `SELECT`, `INSERT` sobre la mesa B.2 • `CREATE_TABLE` sobre la base de datos C • `SELECT`, `INSERT` sobre la mesa C.1 • `SELECT`, `INSERT` sobre la mesa C.2 • `SELECT`, `INSERT` sobre la mesa C.3
Entidad principal 2	• `SELECT`, `INSERT` sobre la mesa A.2 • `CREATE_TABLE` sobre la base de datos B • `SELECT`, `INSERT` sobre la mesa B.1
Entidad principal 3	• `SELECT`, `INSERT` sobre la mesa B.2 • `CREATE_TABLE` sobre la base de datos C • `SELECT`, `INSERT` sobre la mesa C.1 • `SELECT`, `INSERT` sobre la mesa C.2 • `SELECT`, `INSERT` sobre la mesa C.3

Conclusión

En este sencillo ejemplo, el administrador del lago de datos ha podido especificar 17 permisos mediante cinco operaciones de asignación y ocho de concesión. Cuando hay decenas de bases de datos y cientos de tablas, queda clara la ventaja del LF-TBAC método sobre el método de recurso

nombrado. En el caso hipotético de la necesidad de conceder a cada entidad principal acceso a cada recurso, y donde `n(P)` es el número de entidades principales y `n(R)` es el número de recursos:

- Con el método de recurso con nombre, la cantidad de concesiones requeridas es $n(P) \times n(R)$.
- Con el LF-TBAC método, que utiliza una sola LF-Tag, el número total de subvenciones a directores y asignaciones a recursos es $n(P) + n(R)$.

> Véase también
>
> - Gestión del control LF-Tags de acceso a los metadatos
> - Otorgar permisos de lago de datos mediante el método LF-TBAC

Temas

- Gestión del control LF-Tags de acceso a los metadatos
- Gestión de LF-Tag expresiones para el control de acceso a los metadatos
- Administrar los permisos de LF-Tag valor

Gestión del control LF-Tags de acceso a los metadatos

Para utilizar el método de control de acceso basado en etiquetas (LF-TBAC) de Lake Formation para proteger los objetos del catálogo de datos, como catálogos, bases de datos, tablas, vistas y columnas, debe crearlos LF-Tags, asignarlos a los recursos y conceder LF-Tag permisos a los directores.

Antes de poder asignar objetos LF-Tags al catálogo de datos o conceder permisos a los directores, debe definirlos. LF-Tags Solo un administrador de un lago de datos o un director con permisos de LF-Tag creador pueden crear LF-Tags.

LF-Tag creadores

LF-Tag el creador es un director que no es administrador y tiene permisos para crear y administrar LF-Tags. Los administradores de lagos de datos pueden añadir LF-Tag creadores mediante la consola o la CLI de Lake Formation. LF-Tag los creadores tienen permisos implícitos de Lake Formation para actualizar y eliminar LF-Tags, LF-Tags asignar recursos y conceder LF-Tag permisos y permisos de LF-Tag valor a otros directores.

Con las funciones de LF-Tag creador, los administradores de los lagos de datos pueden delegar tareas de administración de etiquetas, como la creación y actualización de claves y valores de las etiquetas, a personas que no sean administradores. Los administradores de los lagos de datos también pueden conceder permisos a LF-Tag los creadores. `Create LF-Tag` A continuación, el LF-Tag creador puede conceder el permiso de creación LF-Tags a otros directores.

Puedes conceder dos tipos de permisos en LF-Tags:

- LF-Tag permisos: `Create LF-Tag`Alter, y`Drop`. Estos permisos son necesarios para crear, actualizar y eliminar LF-Tags.

  Los administradores y LF-Tag creadores de los lagos de datos tienen implícitamente estos permisos LF-Tags cuando los crean y pueden concederlos de forma explícita a los directores para que administren las etiquetas del lago de datos.
- LF-Tag permisos de par clave-valor:`Assign`, y. `Describe` `Grant with LF-Tag expressions` Estos permisos son necesarios para LF-Tags asignarlos a los objetos del catálogo de datos y para conceder permisos sobre los recursos a los directores mediante el control de acceso basado en etiquetas de Lake Formation. LF-Tag los creadores reciben estos permisos de forma implícita al crear. LF-Tags

Tras recibir el `Create LF-Tag` permiso y crearlo correctamente LF-Tags, el LF-Tag creador puede LF-Tags asignar recursos y conceder LF-Tag permisos (`Create LF-Tag`, `AlterDrop`, y) a otras entidades no administrativas para gestionar las etiquetas del lago de datos. Puede administrarlo LF-Tags mediante la consola de Lake Formation, la API o AWS Command Line Interface (AWS CLI).

> **Note**
>
> Los administradores de lagos de datos tienen permisos implícitos de Lake Formation para crear LF-Tags, actualizar y eliminar, LF-Tags asignar recursos y conceder LF-Tag permisos a los directores.

Para conocer las prácticas recomendadas y consideraciones, consulte Prácticas recomendadas y consideraciones sobre el control de acceso basado en etiquetas de Lake Formation

Temas

- Añadir creadores LF-Tag
- Creando LF-Tags

- Actualización LF-Tags
- Eliminando LF-Tags
- Listado LF-Tags
- Asignación LF-Tags a recursos del catálogo de datos
- Visualización LF-Tags asignada a un recurso
- Ver los recursos a los que LF-Tag está asignado a
- Ciclo de vida de un LF-Tag
- Comparación del control de acceso basado en etiquetas de Lake Formation con el control de acceso basado en atributos de IAM

Véase también

- Administrar los permisos de LF-Tag valor
- Otorgar permisos de lago de datos mediante el método LF-TBAC
- Control de acceso basado en etiquetas de Lake Formation

Añadir creadores LF-Tag

De forma predeterminada, los administradores de los lagos de datos pueden crear, actualizar y eliminar LF-Tags, asignar etiquetas a los objetos del catálogo de datos y conceder permisos de etiqueta a los directores. Si desea delegar las operaciones de creación y administración de etiquetas a directores que no sean administradores, el administrador del lago de datos puede crear funciones de LF-Tag creador y conceder `Create LF-Tag` permisos a Lake Formation para las funciones. Con un `Create LF-Tag` permiso concedible, LF-Tag los creadores pueden delegar las tareas de creación y mantenimiento de las etiquetas a otros responsables no administrativos.

Para que los administradores de lagos de datos puedan asignar recursos LF-Tags al catálogo de datos, deben concederse a sí mismos permisos de asociación sobre los LF-Tags que no hayan creado ellos.

> **Note**
>
> Cross-account las concesiones de permisos solo pueden incluir `Associate` permisos `Describe` y permisos. No puede conceder permisos `Create LF-Tag`, `Drop`, `Alter` ni `Grant with LFTag expressions` a las entidades principales de una cuenta diferente.

Temas

- Se requieren permisos de IAM para crear LF-Tags
- Añadir creadores LF-Tag

> **Véase también**
>
> - Administrar los permisos de LF-Tag valor
> - Otorgar permisos de lago de datos mediante el método LF-TBAC
> - Control de acceso basado en etiquetas de Lake Formation

Se requieren permisos de IAM para crear LF-Tags

Debe configurar los permisos para permitir que el director de Lake Formation cree LF-Tags. Agregue la siguiente declaración a la política de permisos del director que debe ser LF-Tag creador.

> **Note**
>
> Si bien los administradores de lagos de datos tienen permisos implícitos de Lake Formation para crear, actualizar y eliminar LF-Tags, asignar LF-Tags recursos y conceder LF-Tags a los directores, los administradores de lagos de datos también necesitan los siguientes permisos de IAM.

Para obtener más información, consulte Personas de Lake Formation y referencia de permisos IAM.

```
{
"Sid": "Transformational",
"Effect": "Allow",
```

```
    "Action": [
        "lakeformation:AddLFTagsToResource",
        "lakeformation:RemoveLFTagsFromResource",
        "lakeformation:GetResourceLFTags",
        "lakeformation:ListLFTags",
        "lakeformation:CreateLFTag",
        "lakeformation:GetLFTag",
        "lakeformation:UpdateLFTag",
        "lakeformation:DeleteLFTag",
        "lakeformation:SearchTablesByLFTags",
        "lakeformation:SearchDatabasesByLFTags"
    ]
}
```

Los directores que asignan LF-Tags recursos y los conceden LF-Tags a los directores deben tener los mismos permisos, excepto los `CreateLFTag` permisos, y. `UpdateLFTag` `DeleteLFTag`

Añadir creadores LF-Tag

Un LF-Tag creador puede crear una LF-Tag, actualizar la clave y los valores de la etiqueta, eliminar etiquetas, asociar etiquetas a los recursos del catálogo de datos y conceder permisos sobre los recursos del catálogo de datos a los directores mediante el LF-TBAC método. El LF-Tag creador también puede conceder estos permisos a los directores.

Puedes crear funciones de LF-Tag creador mediante la AWS Lake Formation consola, la API o AWS Command Line Interface (AWS CLI).

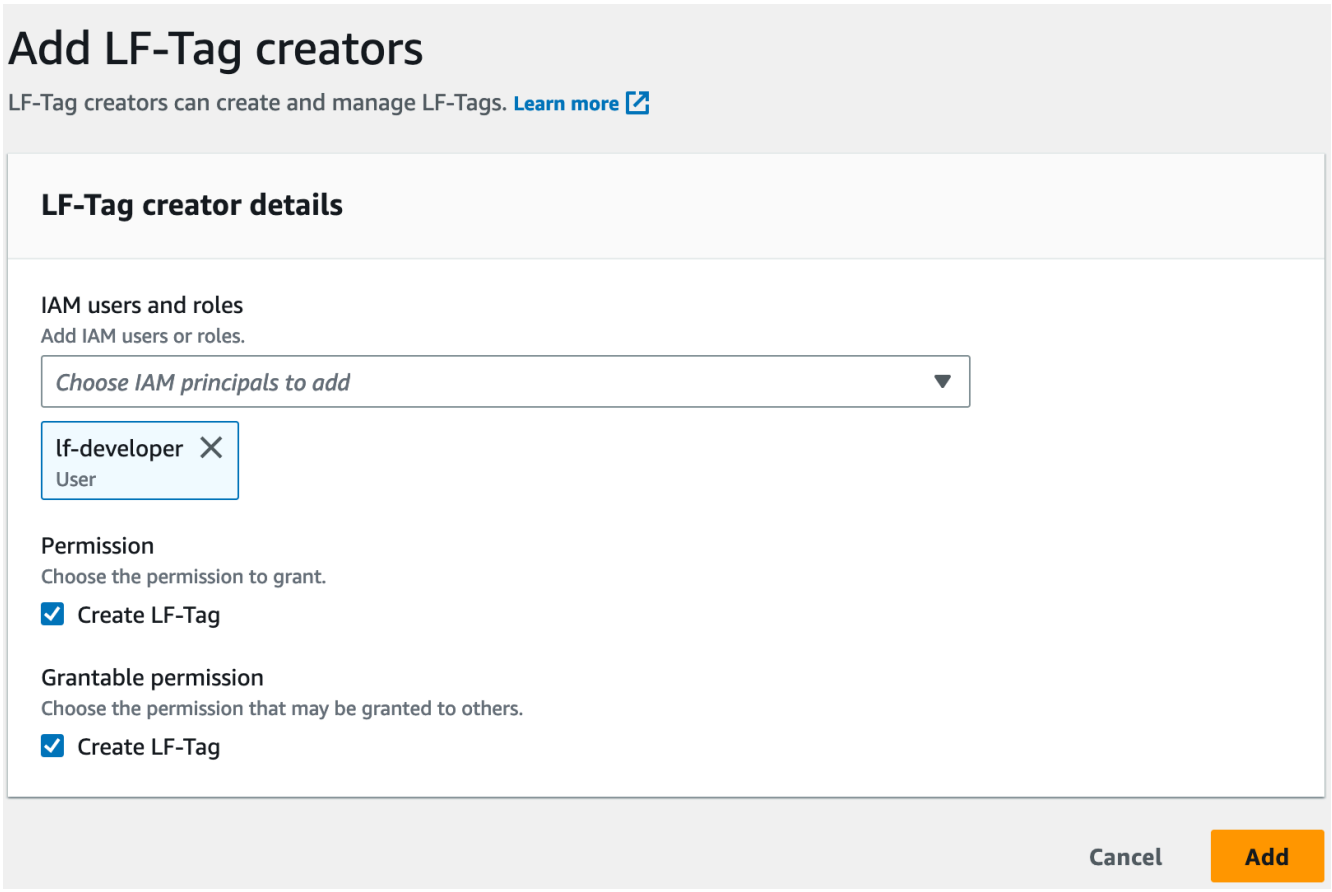
console

Para añadir un LF-Tag creador

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador del lago de datos.
2. En el panel de navegación, en Permisos, elija LF-Tags y permisos.

   En la página LF-Tags y permisos, selecciona la sección de LF-Tag creadores y selecciona Añadir LF-Tag creadores.



3. En la página Añadir LF-Tag creadores, elige un rol o un usuario de IAM que tenga los permisos necesarios para crear. LF-Tags
4. Marque la casilla de verificación del permiso `Create LF-Tag`.
5. (Opcional) Para que las entidades principales seleccionadas puedan conceder permisos `Create LF-Tag` a las entidades principales, seleccione el permiso concedible `Create LF-Tag`.
6. Elija Añadir.

AWS CLI

```
aws lakeformation grant-permissions --cli-input-json file://grantCreate
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:user/tag-manager"
    },
    "Resource": {
        "Catalog": {}
    },
```

```
    "Permissions": [
        "CreateLFTag"
    ],
    "PermissionsWithGrantOption": [
        "CreateLFTag"
    ]
}
```

Los siguientes son los permisos disponibles para un rol de LF-Tag creador:

Permiso	Description (Descripción)
`Drop`	Un director con este permiso LF-Tag puede eliminar a LF-Tag del lago de datos. El director obtiene un `Describe` permiso implícito sobre todos los valores de etiqueta de un LF-Tag recurso.
`Alter`	Un director con este permiso LF-Tag puede añadir o eliminar el valor de una etiqueta de un LF-Tag. El director obtiene un `Alter` permiso implícito sobre todos los valores de etiqueta de a LF-Tag.
`Describe`	Un director con este permiso LF-Tag puede ver los recursos LF-Tag y sus valores cuando los asigna LF-Tags a los recursos o concede permisos a LF-Tags. Puede conceder `Describe` en todos los valores clave o en valores específicos.
`Associate`	Un director con este permiso LF-Tag puede asignarlo LF-Tag a un recurso del catálogo de datos. Al conceder `Associate` está otorgando `Describe` de manera implícita.
`Grant with LF-Tag expression`	Un director con este permiso LF-Tag puede conceder permisos sobre los recursos de un catálogo de datos utilizando la LF-Tag clave y los valores. Conceder `Grant with LF-Tag expression` otorga `Describe` de manera implícita.

Estos permisos se pueden conceder. Una entidad principal que haya recibido estos permisos con la opción de otorgarlos puede otorgarlos a otras entidades principales.

Creando LF-Tags

Todos LF-Tags deben estar definidos en Lake Formation antes de que puedan usarse. A LF-Tag consiste en una clave y uno o más valores posibles para la clave.

Una vez que el administrador del lago de datos haya configurado los permisos de IAM y Lake Formation necesarios para el rol de LF-Tag creador, el director puede crear un LF-Tag. El LF-Tag creador obtiene permiso implícito para actualizar o eliminar cualquier valor de etiqueta LF-Tag y eliminar el. LF-Tag

Puedes crear LF-Tags mediante la AWS Lake Formation consola, la API o el AWS Command Line Interface (AWS CLI).

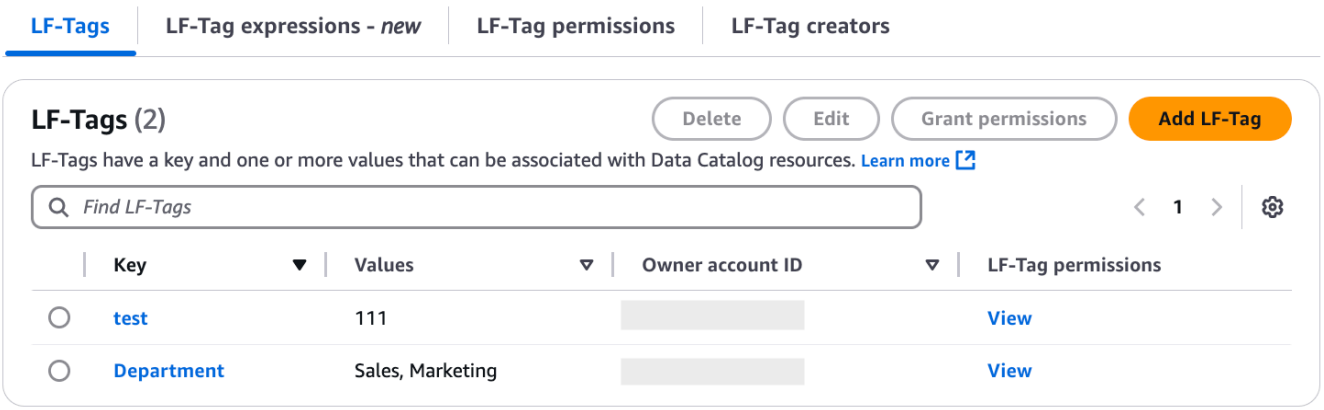
Console

Para crear un LF-Tag

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como director con permisos de LF-Tag creador o como administrador del lago de datos.
2. En el panel de navegación, en Permisos LF-Tags y permisos, elija LF-Tags.

   Se visualizará la página LF-Tags.



3. Elija Añadir LF-Tag.
4. En el cuadro LF-Tag de diálogo Añadir, introduzca una clave y uno o varios valores.

   Cada clave debe tener un valor como mínimo. Para introducir varios valores, introduzca una lista delimitada por comas y, a continuación, pulse Intro, o introduzca un valor cada vez y pulse Añadir después de cada uno. El número máximo de valores permitidos es 1000.

5. Seleccione Agregar etiqueta.

AWS CLI

Para crear un LF-Tag

- Introduzca un comando `create-lf-tag`.

  En el siguiente ejemplo se crea un LF-Tag con clave `module` y valores `Customers` y`Orders`.

```
aws lakeformation create-lf-tag --tag-key module --tag-values Customers Orders
```

Como creador de etiquetas, el director obtiene el `Alter` permiso para ello LF-Tag y puede actualizar o eliminar cualquier valor de etiqueta que contenga LF-Tag. El director LF-Tag creador también puede conceder `Alter` permiso a otro director para que actualice y elimine los valores de las etiquetas que contiene LF-Tag.

Actualización LF-Tags

Para actualizar una `Alter` para la LF-Tag que tiene permiso, agregue o elimine los valores clave permitidos. No puedes cambiar la LF-Tag clave. Para cambiar la clave, elimínala LF-Tag y añade una con la clave requerida. Además del permiso `Alter`, también necesita el permiso IAM `lakeformation:UpdateLFTag` para actualizar los valores.

Al eliminar un LF-Tag valor, no se comprueba la presencia de ese LF-Tag valor en ningún recurso del catálogo de datos. Si el LF-Tag valor eliminado está asociado a un recurso, deja de estar visible para el recurso y las entidades principales a las que se les concedieron permisos en ese par clave-valor ya no tienen los permisos.

Antes de eliminar un LF-Tag valor, si lo desea, puede utilizar el `remove-lf-tags-from-resource`comando de comando para eliminar LF-Tag del catálogo de datos los recursos que tengan el valor que desee eliminar y, a continuación, volver a etiquetar el recurso con los valores que desee conservar.

Solo los administradores del lago de datos, el LF-Tag creador y los directores que tengan `Alter` permisos sobre él LF-Tag pueden actualizar un. LF-Tag

Puede actualizar un LF-Tag mediante la AWS Lake Formation consola, la API o el AWS Command Line Interface (AWS CLI).

Console

Para actualizar una LF-Tag (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador, LF-Tag creador o director de un lago de datos con `Alter` permiso en el LF-Tag.
2. En el panel de navegación, en Permisos LF-Tags y permisos, elija LF-Tags.
3. En la LF-Tagspágina, seleccione un y LF-Tag, a continuación, elija Editar.
4. En el cuadro LF-Tag de diálogo Editar, añada o elimine LF-Tag valores.

   Para añadir varios valores, en el campo Valores introduzca una lista delimitada por comas y, a continuación, pulse Intro, o un valor cada vez y pulse Agregar después de cada uno.
5. Seleccione Save.

AWS CLI

Para actualizar un LF-Tag (AWS CLI)

- Introduzca un comando `update-lf-tag`. Proporcione uno o los dos argumentos siguientes:
  - `--tag-values-to-add`
  - `--tag-values-to-delete`

Example

En el siguiente ejemplo, se reemplaza el valor `vp` `vice-president` por el valor de la LF-Tag clave`level`.

```
aws lakeformation update-lf-tag --tag-key level --tag-values-to-add vice-president
--tag-values-to-delete vp
```

Eliminando LF-Tags

Puede eliminar los LF-Tags que ya no estén en uso. No se comprueba la presencia del recurso LF-Tag en un catálogo de datos. Si el elemento eliminado LF-Tag está asociado a un recurso, deja de estar visible para el recurso y para los principales a los que se les concedieron permisos sobre ese recurso ya LF-Tag no los tienen.

Antes de eliminar un LF-Tag, si lo desea, puede utilizar el `remove-lf-tags-from-resource`comando para eliminarlo LF-Tag de todos los recursos.

Solo los administradores del lago de datos, el LF-Tag creador o un director que tenga `Drop` permiso sobre el LF-Tag pueden eliminar un LF-Tag. Además del `Drop` permiso, el director también necesita el permiso de `lakeformation:DeleteLFTag` IAM para eliminar un LF-Tag.

Puedes eliminar un LF-Tag mediante la AWS Lake Formation consola, la API o el AWS Command Line Interface (AWS CLI).

Console

Para eliminar una LF-Tag (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador del lago de datos.
2. En el panel de navegación, en Permisos LF-Tags y permisos, elija LF-Tags.
3. En la LF-Tagspágina, seleccione un y LF-Tag, a continuación, elija Eliminar.
4. ¿En el entorno Eliminar etiquetas? En el cuadro de diálogo, para confirmar la eliminación, introduzca el valor de la LF-Tag clave en el campo designado y, a continuación, seleccione Eliminar.

AWS CLI

Para eliminar un LF-Tag (AWS CLI)

- Introduzca un comando `delete-lf-tag`. Proporcione la clave del LF-Tag que desea eliminar.

  Example

  En el siguiente ejemplo, se elimina el elemento LF-Tag con la clave`region`.

```
aws lakeformation delete-lf-tag --tag-key region
```

Listado LF-Tags

Puede enumerar aquellos sobre los LF-Tags que tiene `Associate` permisos `Describe` o. Los valores que aparecen con cada LF-Tag clave son los valores sobre los que tiene permisos.

LF-Tag el creador tiene permisos implícitos para ver LF-Tags lo que ha creado.

Los administradores del lago de datos pueden ver todo lo LF-Tags que está definido en la AWS cuenta local y todos aquellos LF-Tags para los `Describe` que se han otorgado `Associate` permisos a la cuenta local desde cuentas externas. El administrador del lago de datos puede ver todos los valores de todos LF-Tags.

Puede publicar listas LF-Tags mediante la AWS Lake Formation consola, la API o AWS Command Line Interface (AWS CLI).

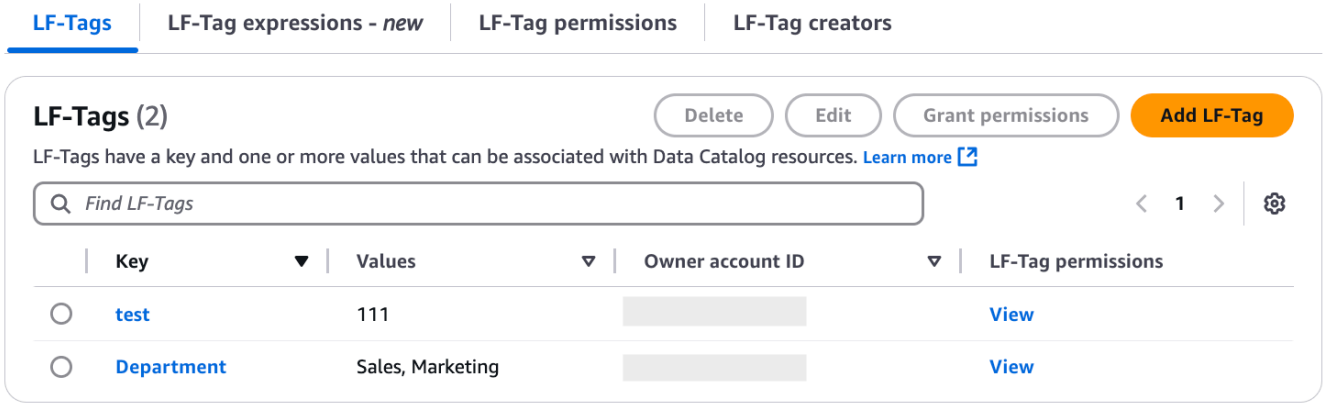
Console

Para enumerar LF-Tags (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como LF-Tag creador, administrador de un lago de datos o como director al que se le hayan concedido permisos LF-Tags y que cuente con el permiso de `lakeformation:ListLFTags` IAM.
2. En el panel de navegación, en Permisos LF-Tags y permisos, selecciona LF-Tags.

   Se visualizará la página LF-Tags.



Comprueba la columna del identificador de la cuenta del propietario para determinar qué cuentas LF-Tags se compartieron con tu cuenta desde una cuenta externa.

AWS CLI

Para enumerar LF-Tags (AWS CLI)

- Ejecute el siguiente comando como administrador de un lago de datos o como director al que se le hayan concedido permisos LF-Tags y que tenga el permiso de `lakeformation:ListLFTags` IAM.

```
aws lakeformation list-lf-tags
```

El resultado es similar al siguiente.

```
{
    "LFTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "level",
            "TagValues": [
                "director",
                "vp",
                "c-level"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "Orders",
                "Sales",
                "Customers"
            ]
        }
    ]
}
```

Para ver también los LF-Tags que se concedieron desde cuentas externas, incluye la opción `--resource-share-type ALL` de comando.

```
aws lakeformation list-lf-tags --resource-share-type ALL
```

El resultado es similar al siguiente. Observe la clave `NextToken`, que indica que la lista continúa.

```
{
    "LFTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "level",
            "TagValues": [
                "director",
                "vp",
                "c-level"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "Orders",
                "Sales",
                "Customers"
            ]
        }
    ],
    "NextToken": "eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ=="
}
```

Repita el comando y añada el `--next-token` argumento para ver las cuentas locales restantes LF-Tags y las LF-Tags que se hayan concedido desde cuentas externas. LF-Tags de cuentas externas siempre están en una página separada.

```
aws lakeformation list-lf-tags --resource-share-type ALL
--next-token eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ==
```

```
{
    "LFTags": [
        {
            "CatalogId": "123456789012",
```

```
            "TagKey": "region",
            "TagValues": [
                "central",
                "south"
            ]
        }
    ]
}
```

API

Puede usar los SDK disponibles para Lake Formation para enumerar las etiquetas que el solicitante tiene permiso para ver.

```
import boto3

client = boto3.client('lakeformation')
...

response = client.list_lf_tags(
    CatalogId='string',
    ResourceShareType='ALL',
    MaxResults=50'
)
```

Este comando devuelve un objeto `dict` con la estructura siguiente:

```
{
    'LFTags': [
        {
            'CatalogId': 'string',
            'TagKey': 'string',
            'TagValues': [
                'string',
            ]
        },
    ],
    'NextToken': 'string'
}
```

Para obtener más información sobre los permisos necesarios, consulte Personas de Lake Formation y referencia de permisos IAM.

Asignación LF-Tags a recursos del catálogo de datos

Puede asignar recursos LF-Tags al catálogo de datos (bases de datos, tablas y columnas) para controlar el acceso a esos recursos. Solo los directores a los que se les concede una coincidencia LF-Tags (y los directores a los que se les concede acceso con el método de recurso indicado) pueden acceder a los recursos.

Si una tabla hereda LF-Tag a de una base de datos o una columna hereda LF-Tag a de una tabla, puede anular el valor heredado asignando un nuevo valor a la clave. LF-Tag

El número máximo LF-Tags que puede asignar a un recurso es 50.

Temas

- Requisitos para administrar las etiquetas asignadas a los recursos
- Asignar LF-Tags a una columna de la tabla
- Asignar LF-Tags a un recurso del catálogo de datos
- Actualización LF-Tags de un recurso
- Eliminar LF-Tag de un recurso

Requisitos para administrar las etiquetas asignadas a los recursos

Para asignar un recurso LF-Tag a un catálogo de datos, debe:

- Tenga el `ASSOCIATE` permiso de Lake Formation en el LF-Tag.
- Tener el permiso de IAM `lakeformation:AddLFTagsToResource`.
- Tener pegamento: GetDatabase permiso en una base de datos de Glue.
- Sea el propietario (creador) del recurso, tener el permiso de Lake Formation `Super` sobre el recurso con la opción `GRANT` o tenga los siguientes permisos con la opción `GRANT`:
  - Para bases de datos de la misma AWS cuenta: `DESCRIBECREATE_TABLE`,`ALTER`, y `DROP`
  - Para bases de datos de una cuenta externa: `DESCRIBE`, `CREATE_TABLE` y `ALTER`
  - Para tablas (y columnas): `DESCRIBE`, `ALTER`, `DROP`, `INSERT`, `SELECT` y `DELETE`

Además, el recurso LF-Tag y el recurso al que se está asignando deben estar en la misma AWS cuenta.

Para eliminar un recurso LF-Tag de un catálogo de datos, debe cumplir estos requisitos y, además, tener el permiso de `lakeformation:RemoveLFTagsFromResource` IAM.

Asignar LF-Tags a una columna de la tabla

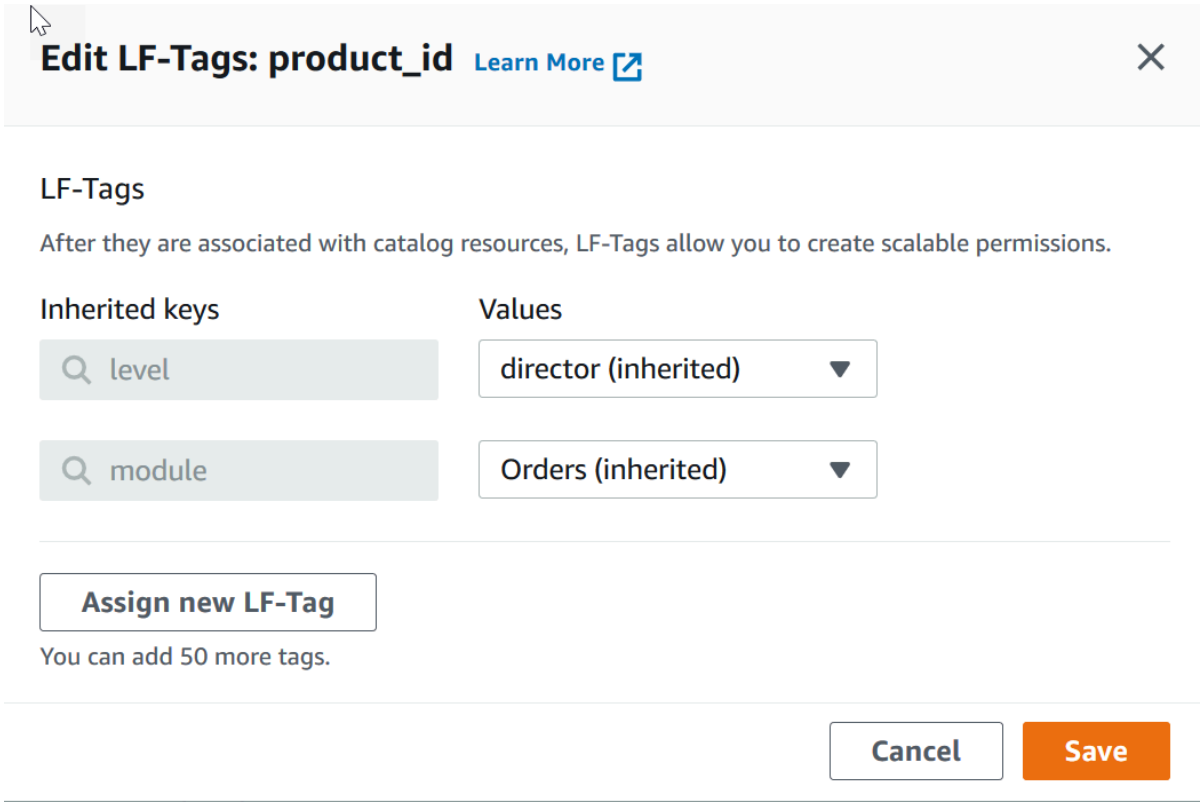
Para asignar LF-Tags a una columna de tabla (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

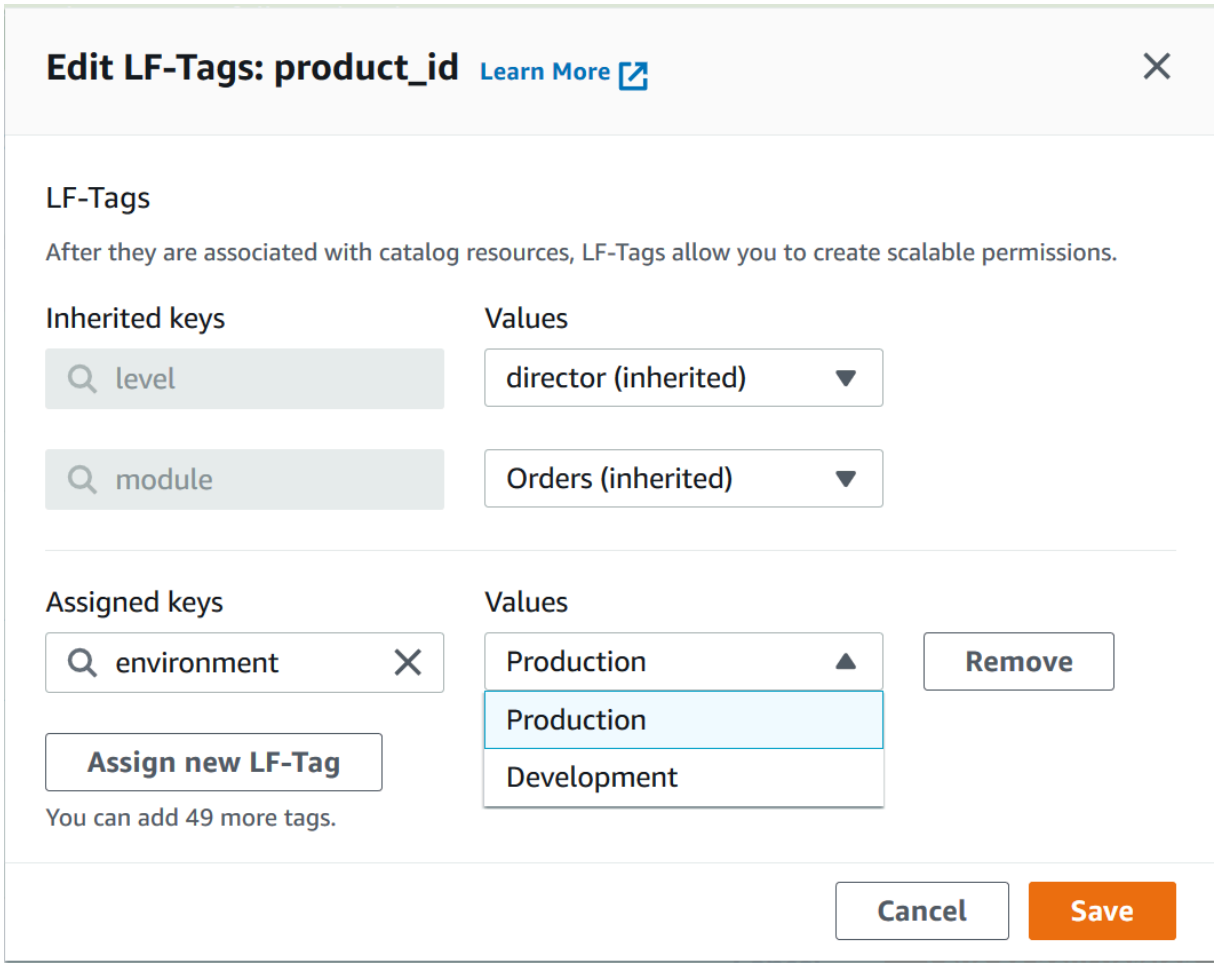
   Inicie sesión como usuario que cumpla con los requisitos enumerados anteriormente.
2. En el panel de navegación, elija Tablas.
3. Elija un nombre de tabla (no el botón de opción situado junto al nombre de tabla).
4. En la página de detalles de la tabla, en la sección Esquema, seleccione Editar esquema.
5. En la página Editar esquema, seleccione una o más columnas y, a continuación, elija Editar LF-Tags.

   > **Note**
   >
   > Si quiere añadir o eliminar columnas y guardar una nueva versión, hágalo en primer lugar. A continuación, edite el LF-Tags.

   Aparece el LF-Tags cuadro de diálogo Editar y muestra los LF-Tags que se hayan heredado de la tabla.



6. (Opcional) En la lista de valores que hay junto a un campo de claves heredadas, seleccione un valor que sustituya al heredado.
7. (Opcional) Selecciona Asignar nuevo LF-Tag. Después, en Teclas asignadas, seleccione una clave y, en Valores, elija un valor para ella.



8. (Opcional) Vuelva a seleccionar Asignar LF-Tag nuevo para añadir otro LF-Tag.
9. Seleccione Save.

Asignar LF-Tags a un recurso del catálogo de datos

Console

Para asignar LF-Tags a una base de datos o tabla del catálogo de datos

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

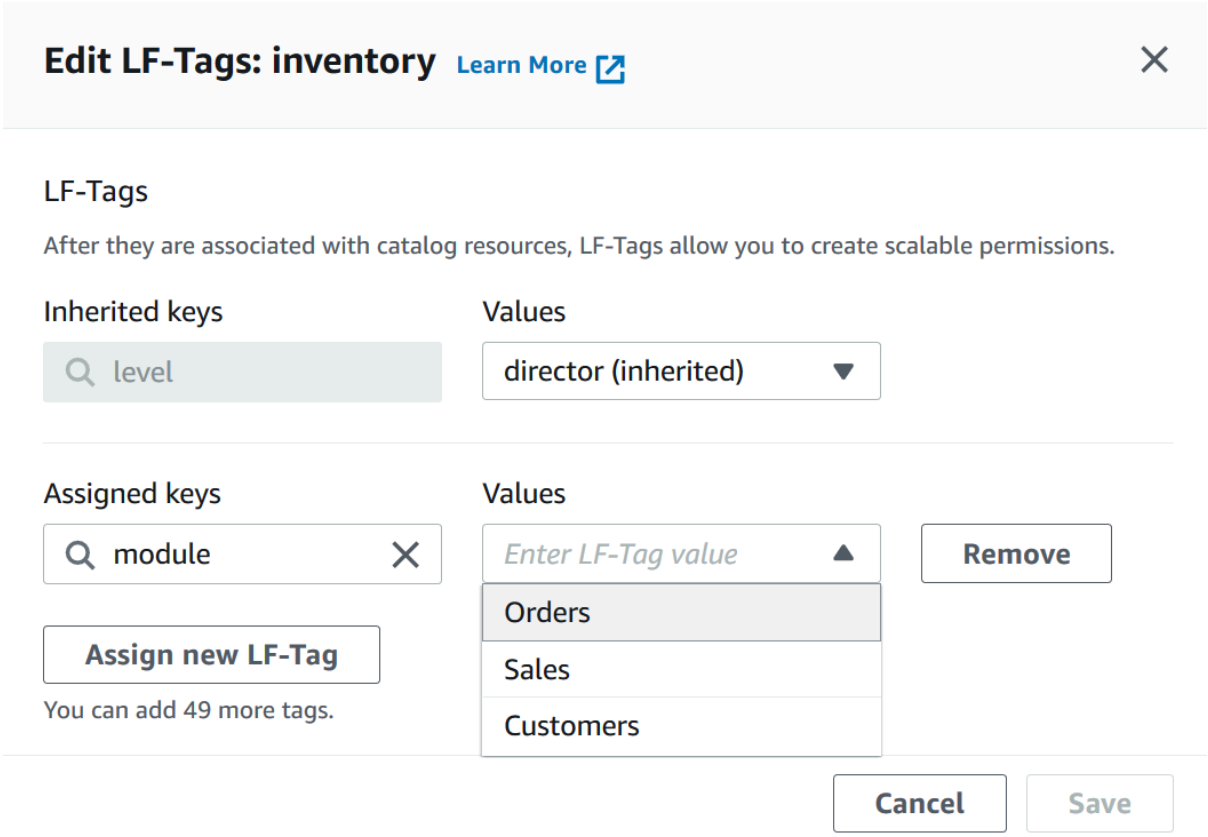
   Inicie sesión como usuario que cumple los requisitos enumerados anteriormente.
2. En el Catálogo de datos del panel de navegación, puede elegir entre:

   - Para asignar LF-Tags a bases de datos, elija Bases de datos.
   - Para asignar LF-Tags a tablas, elija Tablas.

3. Elija una base de datos o una tabla y, en el menú Acciones, elija Editar LF-Tags.

   Aparece el *`resource-name`* cuadro de diálogo Editar LF-Tags:.

   Si una tabla hereda de la base LF-Tags de datos que la contiene, la ventana muestra la heredada LF-Tags. De lo contrario, muestra el texto «No hay ningún elemento heredado LF-Tags asociado al recurso».



4. (Opcional) Si una tabla se ha heredado LF-Tags, en la lista de valores situada junto a un campo de claves heredadas, puede elegir un valor que sustituya al valor heredado.
5. Para asignar uno nuevo LF-Tags, lleve a cabo estos pasos:

   a. Seleccione Asignar nuevo LF-Tag.

   b. En el campo Claves asignadas, elija una LF-Tag clave y, en el campo Valores, elija un valor.

   c. (Opcional) Vuelva a seleccionar Asignar LF-Tag nuevo para asignar uno adicional LF-Tag.

6. Seleccione Save.

AWS CLI

Para asignar LF-Tags a un recurso del catálogo de datos

- Ejecute el comando `add-lf-tags-to-resource`.

  El siguiente ejemplo asigna el LF-Tag `module=orders` a la tabla de la base `orders` de datos`erp`. Utiliza la sintaxis de atajos para el argumento `--lf-tags`. La propiedad `CatalogID` para `--lf-tags` es opcional. Si no se proporciona, se asume el ID de catálogo del recurso (en este caso, la tabla).

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":
 {"DatabaseName":"erp", "Name":"orders"}}' --lf-tags
 CatalogId=111122223333,TagKey=module,TagValues=orders
```

  Si el comando se ejecuta correctamente, verá el siguiente resultado:

```
{
    "Failures": []
}
```

  En el siguiente ejemplo, se asignan dos LF-Tags a la `sales` tabla y se utiliza la sintaxis JSON para el `--lf-tags` argumento.

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":
 {"DatabaseName":"erp", "Name":"sales"}}' --lf-tags '[{"TagKey":
 "module","TagValues": ["sales"]},{"TagKey": "environment","TagValues":
 ["development"]}]'
```

  En el siguiente ejemplo, se asigna el LF-Tag `level=director` a la `total` columna de la tabla. `sales`

```
aws lakeformation add-lf-tags-to-resource --resource '{ "TableWithColumns":
 {"DatabaseName":"erp", "Name":"sales", "ColumnNames":["total"]}}' --lf-tags
 TagKey=level,TagValues=director
```

Actualización LF-Tags de un recurso

Para actualizar un LF-Tag recurso de catálogo de datos (AWS CLI)

- Utilice el comando `add-lf-tags-to-resource`, según se describe en el procedimiento anterior.

  Al agregar un LF-Tag con la misma clave que uno existente LF-Tag, pero con un valor diferente, se actualiza el valor existente.

Eliminar LF-Tag de un recurso

Para eliminar un LF-Tag recurso de un catálogo de datos (AWS CLI)

- Ejecute el comando `remove-lf-tags-from-resource`.

  Si una tabla tiene un LF-Tag valor que anula el valor heredado de la base de datos principal, al quitarlo LF-Tag de la tabla se restaura el valor heredado. Este comportamiento también se aplica a una columna que anula los valores de clave heredados de la tabla.

  En el siguiente ejemplo, se quita el LF-Tag `level=director` elemento de la `total` columna de la `sales` tabla. La propiedad `CatalogID` para `--lf-tags` es opcional. Si no se proporciona, se asume el ID de catálogo del recurso (en este caso, la tabla).

```
aws lakeformation remove-lf-tags-from-resource
--resource ' { "TableWithColumns":
{ "DatabaseName": "erp",  "Name": "sales",  "ColumnNames":[ "total"]}}'
--lf-tags  CatalogId=111122223333,TagKey=level,TagValues=director
```

Visualización LF-Tags asignada a un recurso

Puede ver los LF-Tags que están asignados a un recurso del catálogo de datos. Debe tener el `ASSOCIATE` permiso `DESCRIBE` o permiso en un LF-Tag para verlo.

Console

Para ver los LF-Tags que están asignados a un recurso (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

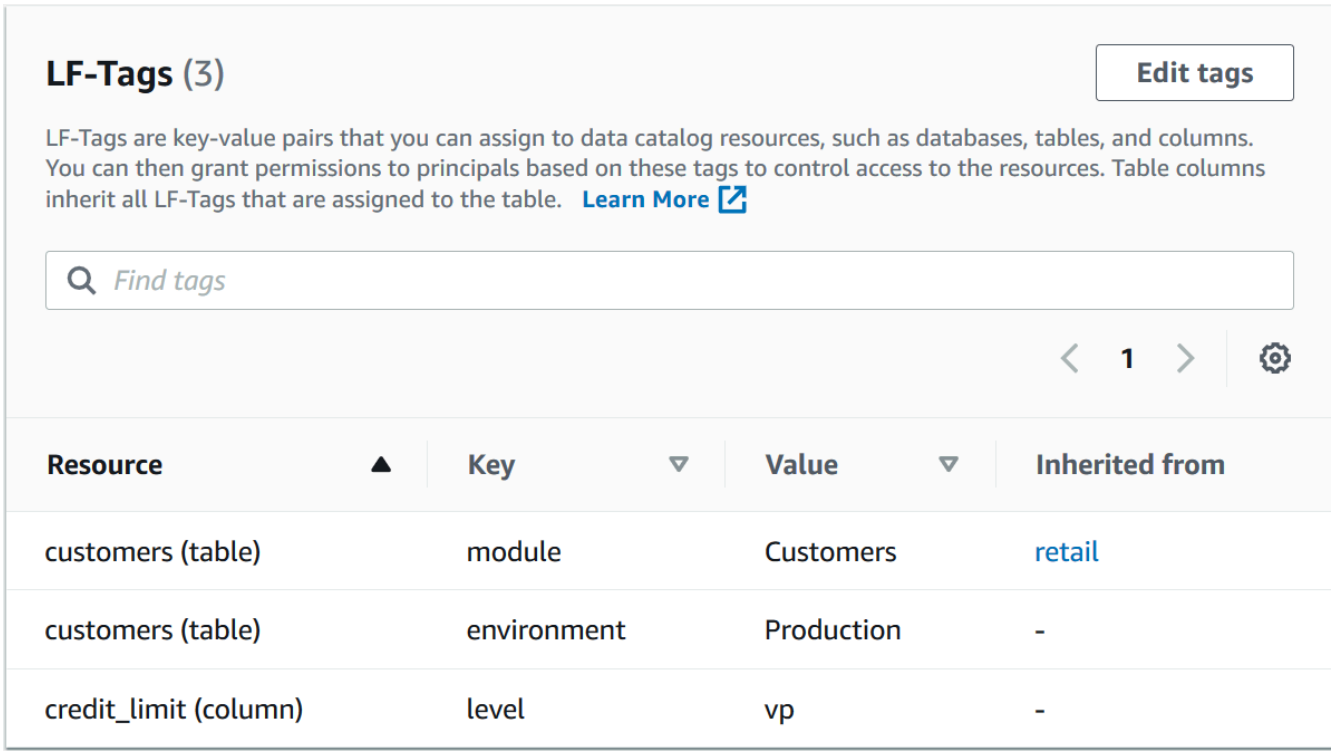
Inicie sesión como administrador del lago de datos, propietario del recurso o usuario al que se le hayan concedido permisos de Lake Formation sobre el recurso.

2. En el Catálogo de datos del panel de navegación, puede elegir entre:

   - Para ver lo LF-Tags asignado a una base de datos, elija Bases de datos.
   - Para ver las LF-Tags asignadas a una tabla, seleccione Tablas.

3. En la página Tablas o Bases de datos, elija el nombre de la base de datos o la tabla. A continuación, en la página de detalles, desplázate hacia abajo hasta la LF-Tagssección.

   La siguiente captura de pantalla muestra lo LF-Tags asignado a una `customers` tabla, que se encuentra en la `retail` base de datos. `module` LF-Tag Se hereda de la base de datos. La `credit_limit` columna tiene el `level=vp` LF-Tag asignado.



AWS CLI

Para ver las LF-Tags que están asignadas a un recurso (AWS CLI)

- Introduzca un comando similar al siguiente.

```
aws lakeformation get-resource-lf-tags --show-assigned-lf-tags --
resource '{ "Table": {"CatalogId":"111122223333", "DatabaseName":"erp",
 "Name":"sales"}}'
```

El comando devuelve el resultado siguiente.

```
{
    "TableTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "sales"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "environment",
            "TagValues": [
                "development"
            ]
        }
    ],
    "ColumnTags": [
        {
            "Name": "total",
            "Tags": [
                {
                    "CatalogId": "111122223333",
                    "TagKey": "level",
                    "TagValues": [
                        "director"
                    ]
                }
            ]
        }
    ]
}
```

Esta salida muestra solo los LF-Tags que se asignan de forma explícita, no los que se heredan. Si desea ver todas las LF-Tags columnas, incluidas las heredadas LF-Tags, omita la `--show-assigned-lf-tags` opción.

Ver los recursos a los que LF-Tag está asignado a

Puede ver todos los recursos del catálogo de datos a los que está asignada una LF-Tag clave en particular. Para ello, necesitará los siguientes permisos de Lake Formation:

- `Describe`o `Associate` en el LF-Tag.
- `Describe` o cualquier otro sobre el recurso.

Además, necesita los siguientes permisos AWS Identity and Access Management (IAM):

- `lakeformation:SearchDatabasesByLFTags`
- `lakeformation:SearchTablesByLFTags`

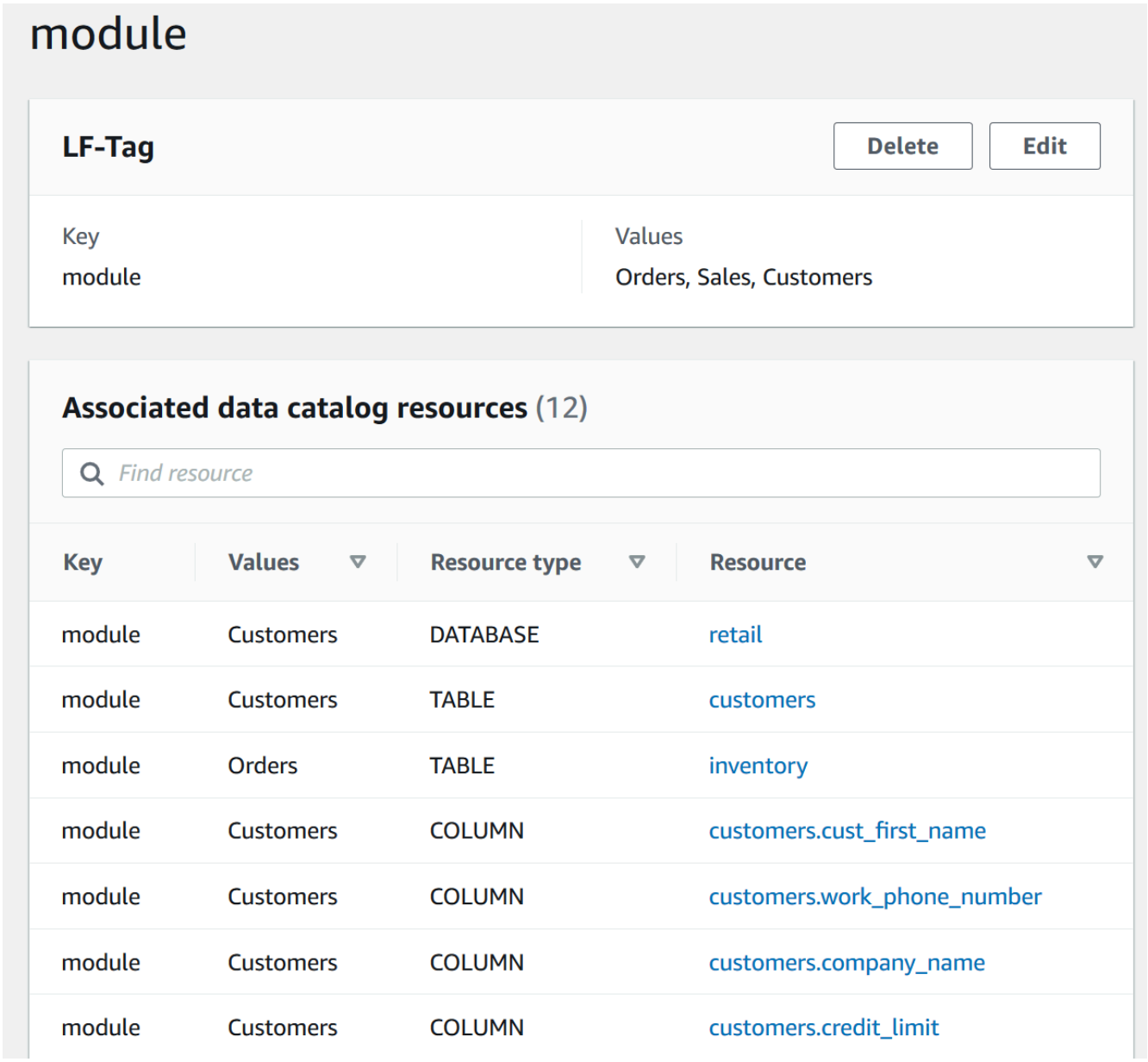
Console

Para ver los recursos a los que LF-Tag está asignado a (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador de un lago de datos o como usuario que cumpla los requisitos enumerados anteriormente.
2. En el panel de navegación, en Permisos LF-Tags y permisos, elija LF-Tags.
3. Elija una LF-Tag clave (no el botón de opción situado junto al nombre de la clave).

   La página de LF-Tag detalles muestra una lista de los recursos a los que se LF-Tag ha asignado.



AWS CLI

Para ver los recursos a los que LF-Tag está asignado a

- Ejecute un comando `search-tables-by-lf-tags` o `search-databases-by-lf-tags`.

Example

En el siguiente ejemplo, se enumeran las tablas y columnas que tienen `level=vp` LF-Tag asignadas las. Para cada tabla y columna de la lista, se muestran todas LF-Tags las asignadas a la tabla o columna, no solo la expresión de búsqueda.

```
aws lakeformation search-tables-by-lf-tags --expression
 TagKey=level,TagValues=vp
```

Para obtener más información sobre los permisos necesarios, consulte Personas de Lake Formation y referencia de permisos IAM.

Ciclo de vida de un LF-Tag

1. El LF-Tag creador Michael crea un LF-Tag`module=Customers`.
2. Michael cede `Associate` el favor LF-Tag al ingeniero de datos Eduardo. Al conceder `Associate` está otorgando `Describe` de manera implícita.
3. Michael concede `Super Custs` a Eduardo con la opción de conceder, para que Eduardo pueda asignarlo LF-Tags a la mesa. Para obtener más información, consulte Asignación LF-Tags a recursos del catálogo de datos.
4. Eduardo lo asigna LF-Tag `module=customers` a la mesa. `Custs`
5. Michael efectúa la siguiente concesión a la ingeniera de datos Sandra (en pseudocódigo).

```
GRANT (SELECT, INSERT ON TABLES) ON TAGS module=customers TO Sandra WITH GRANT OPTION
```

6. Sandra hace la concesión siguiente a la analista de datos Maria.

```
GRANT (SELECT ON TABLES) ON TAGS module=customers TO Maria
```

Ahora María puede ejecutar consultas en la tabla `Custs`.

Véase también

- Control de acceso a los metadatos

Comparación del control de acceso basado en etiquetas de Lake Formation con el control de acceso basado en atributos de IAM

Attribute-based el control de acceso (ABAC) es una estrategia de autorización que define los permisos en función de los atributos. En AWS, estos atributos se denominan etiquetas. Puede adjuntar etiquetas a los recursos de IAM, incluidas las entidades de IAM (usuarios o roles) y a AWS los recursos. Puede crear una única política ABAC o un conjunto pequeño de políticas para sus entidades principales de IAM. Estas políticas ABAC se pueden diseñar para permitir operaciones cuando la etiqueta de la entidad principal coincida con la etiqueta del recurso. ABAC es útil en entornos que crecen con rapidez y ayuda en situaciones en las que la administración de las políticas resulta engorrosa.

Los equipos de seguridad y gobernanza de la nube utilizan IAM para definir las políticas de acceso y los permisos de seguridad para todos los recursos, incluidos los buckets de Amazon S3, las instancias de Amazon EC2 y cualquier recurso al que pueda hacer referencia con un ARN. Las políticas de IAM definen permisos amplios (específicos) para los recursos de su lago de datos, por ejemplo, para permitir o denegar el acceso a nivel de bucket, prefijo o base de datos de Amazon S3. Para obtener más información sobre IAM ABAC, consulte ¿Para qué sirve ABAC? AWS en la Guía del usuario de IAM.

Por ejemplo, puede crear tres roles con la clave de etiqueta `project-access`. Establezca el valor de etiqueta del primer rol en `Dev`, el segundo en `Support` y el tercero en `Marketing`. Asigne etiquetas con el valor adecuado a los recursos. A continuación, puede utilizar una única política que permita el acceso cuando el rol y el recurso estén etiquetados con el mismo valor para `project-access`.

Los equipos de gobierno de datos utilizan Lake Formation para definir permisos detallados para recursos de lagos de datos específicos. LF-Tags se asignan a los recursos del catálogo de datos (bases de datos, tablas y columnas) y se otorgan a los directores. Un director con LF-Tags esa coincidencia con la LF-Tags de un recurso puede acceder a ese recurso. Los permisos de Lake Formation son secundarios a los permisos de IAM. Por ejemplo, si los permisos de IAM no permiten a un usuario acceder a un lago de datos, Lake Formation no concede acceso a ningún recurso de ese lago de datos a ese usuario, incluso si el principal y el recurso coinciden LF-Tags.

El control de acceso basado en etiquetas de Lake Formation (LF-TBAC) funciona con IAM ABAC para proporcionar niveles adicionales de permisos para sus datos y recursos de Lake Formation.

- Los permisos del TBAC de Lake Formation se escalan con innovación. Ya no es necesario que un administrador actualice las políticas existentes para permitir el acceso a nuevos recursos. Por

ejemplo, supongamos que utiliza una estrategia ABAC de IAM con la etiqueta `project-access` para proporcionar acceso a bases de datos específicas de Lake Formation. Al usar LF-TBAC, LF-Tag `Project=SuperApp` se asigna a tablas o columnas específicas, y lo mismo LF-Tag se otorga a un desarrollador para ese proyecto. A través de IAM, el desarrollador puede acceder a la base de datos, y LF-TBAC los permisos le otorgan un mayor acceso a tablas o columnas específicas dentro de las tablas. Si se agrega una tabla nueva al proyecto, el administrador de Lake Formation solo necesita asignar la etiqueta a la nueva tabla para que el desarrollador tenga acceso a la tabla.

- TBAC de Lake Formation requiere menos políticas de IAM. Al utilizar las políticas de IAM para conceder un acceso de alto nivel a los recursos de Lake Formation y TBAC de Lake Formation para gestionar un acceso más preciso a los datos, crea menos políticas de IAM.
- Con el TBAC de Lake Formation, los equipos pueden cambiar y crecer rápidamente. Esto se debe a que los permisos para nuevos recursos se conceden automáticamente de acuerdo con los atributos. Por ejemplo, si un nuevo desarrollador se une al proyecto, es fácil concederle acceso asociando la función de IAM al usuario y, a continuación, asignándole la función necesaria LF-Tags . No es necesario cambiar la política de IAM para respaldar un nuevo proyecto o crear uno nuevo. LF-Tags
- Finer-grained los permisos son posibles utilizando Lake Formation TBAC. Las políticas de IAM otorgan acceso a los recursos de nivel superior, como las bases de datos o las tablas del Catálogo de datos. Con TBAC de Lake Formation, puede conceder acceso a tablas o columnas específicas que contienen valores de datos específicos.

> **Note**
>
> Las etiquetas de IAM no son lo mismo que. LF-Tags Estas etiquetas no son intercambiables. LF-Tags se utilizan para conceder permisos a Lake Formation y las etiquetas de IAM se utilizan para definir las políticas de IAM.

Gestión de LF-Tag expresiones para el control de acceso a los metadatos

LF-Tag las expresiones son expresiones lógicas compuestas por uno o más LF-Tags (pares clave-valor) que se utilizan para conceder permisos sobre AWS Glue Data Catalog los recursos. LF-Tag las expresiones permiten definir reglas que rigen el acceso a los recursos de datos en función de sus etiquetas de metadatos. Puede guardar estas expresiones y reutilizarlas en varias concesiones de permisos, lo que garantiza la coherencia y facilita la administración de los cambios en la ontología de las etiquetas a lo largo del tiempo.

Dentro de una LF-Tag expresión determinada, las claves de las etiquetas se combinan mediante la operación AND, mientras que los valores se combinan mediante la operación OR. Por ejemplo, la expresión de etiqueta `content_type:Sales AND location:US` representa los recursos relacionados con los datos de ventas en EE. UU.

Puede crear hasta 1000 LF-Tag expresiones en una Cuenta de AWS. Estas expresiones proporcionan una forma flexible y escalable de administrar los permisos en función de las etiquetas de metadatos, lo que garantiza que solo los usuarios o las aplicaciones con autorización puedan acceder a recursos de datos específicos en función de las reglas de etiquetas definidas.

LF-Tag las expresiones ofrecen las siguientes ventajas:

- Reutilización: al definir y guardar LF-Tag las expresiones, ya no es necesario replicar manualmente las mismas expresiones al asignar permisos a otros recursos o entidades principales.
- Coherencia: la reutilización de LF-Tag expresiones en varias concesiones de permisos garantiza la coherencia en la forma en que se conceden y administran los permisos.
- Gestión de la ontología de etiquetas: LF-Tag las expresiones ayudan a gestionar los cambios en la ontología de etiquetas a lo largo del tiempo, ya que puede actualizar las expresiones guardadas en lugar de modificar las concesiones de permisos individuales.

Para obtener más información sobre el control de acceso basado en etiquetas, consulte Control de acceso basado en etiquetas de Lake Formation.

LF-Tag creadores de expresiones

LF-Tag el creador de expresiones es un director que tiene permisos para crear y administrar LF-Tag expresiones. Los administradores de lagos de datos pueden añadir creadores de LF-Tag expresiones mediante la consola, la CLI, la API o el SDK de Lake Formation. LF-Tag los creadores de expresiones tienen permisos implícitos de Lake Formation para crear, actualizar y eliminar LF-Tag expresiones, y para conceder permisos de LF-Tag expresión a otros directores.

LF-Tag Los creadores de expresiones que no son administradores de lagos de datos reciben `Grant with LF-Tag expression` permisos implícitos `Alter` y solo para las expresiones que hayan creado. `Drop` `Describe`

Los administradores de lagos de datos también pueden conceder permisos LF-Tag concedibles a los creadores de expresiones. `Create LF-Tag expression` A continuación, el creador de la LF-Tag expresión puede conceder el permiso para crear LF-Tag expresiones a otros directores.

Temas

- Se necesitan permisos de IAM para crear expresiones LF-Tag
- Añada creadores de expresiones LF-Tag
- Crear expresiones LF-Tag
- Actualización de LF-Tag expresiones
- LF-Tag Eliminar expresiones
- Enumerar LF-Tag las expresiones

> **Véase también**
>
> - Administrar los permisos de LF-Tag valor
> - Otorgar permisos de lago de datos mediante el método LF-TBAC
> - Control de acceso basado en etiquetas de Lake Formation

Se necesitan permisos de IAM para crear expresiones LF-Tag

Debe configurar los permisos para permitir que el director de Lake Formation cree LF-Tag expresiones. Agregue la siguiente declaración a la política de permisos del director que debe ser un creador de LF-Tag expresiones.

> **Note**
>
> Si bien los administradores de lagos de datos tienen permisos implícitos de Lake Formation para crear, actualizar LF-Tags y eliminar LF-Tag expresiones, asignar LF-Tags recursos LF-Tags y conceder permisos de LF-Tag expresión a los directores, los administradores de lagos de datos también necesitan los siguientes permisos de IAM.

Para obtener más información, consulte Personas de Lake Formation y referencia de permisos IAM.

```
{
"Sid": "Transformational",
"Effect": "Allow",
    "Action": [
```

```
        "lakeformation:AddLFTagsToResource",
        "lakeformation:RemoveLFTagsFromResource",
        "lakeformation:GetResourceLFTags",
        "lakeformation:ListLFTags",
        "lakeformation:CreateLFTag",
        "lakeformation:GetLFTag",
        "lakeformation:UpdateLFTag",
        "lakeformation:DeleteLFTag",
        "lakeformation:SearchTablesByLFTags",
        "lakeformation:SearchDatabasesByLFTags",
        "lakeformation:CreateLFTagExpression",
        "lakeformation:DeleteLFTagExpression",
        "lakeformation:UpdateLFTagExpression",
        "lakeformation:GetLFTagExpression",
        "lakeformation:ListLFTagExpressions",
        "lakeformation:GrantPermissions",
        "lakeformation:RevokePermissions",
        "lakeformation:BatchGrantPermissions",
        "lakeformation:BatchRevokePermissions"
    ]
}
```

Añada creadores de expresiones LF-Tag

LF-Tag Los creadores de expresiones pueden crear y guardar LF-Tag expresiones reutilizables, actualizar la clave y los valores de las etiquetas, eliminar expresiones y conceder permisos sobre los recursos del catálogo de datos a los directores mediante el LF-TBAC método. El creador de la LF-Tag expresión también puede conceder estos permisos a los directores.

Puede crear funciones de creador de LF-Tag expresiones mediante la AWS Lake Formation consola, la API o AWS Command Line Interface (AWS CLI).

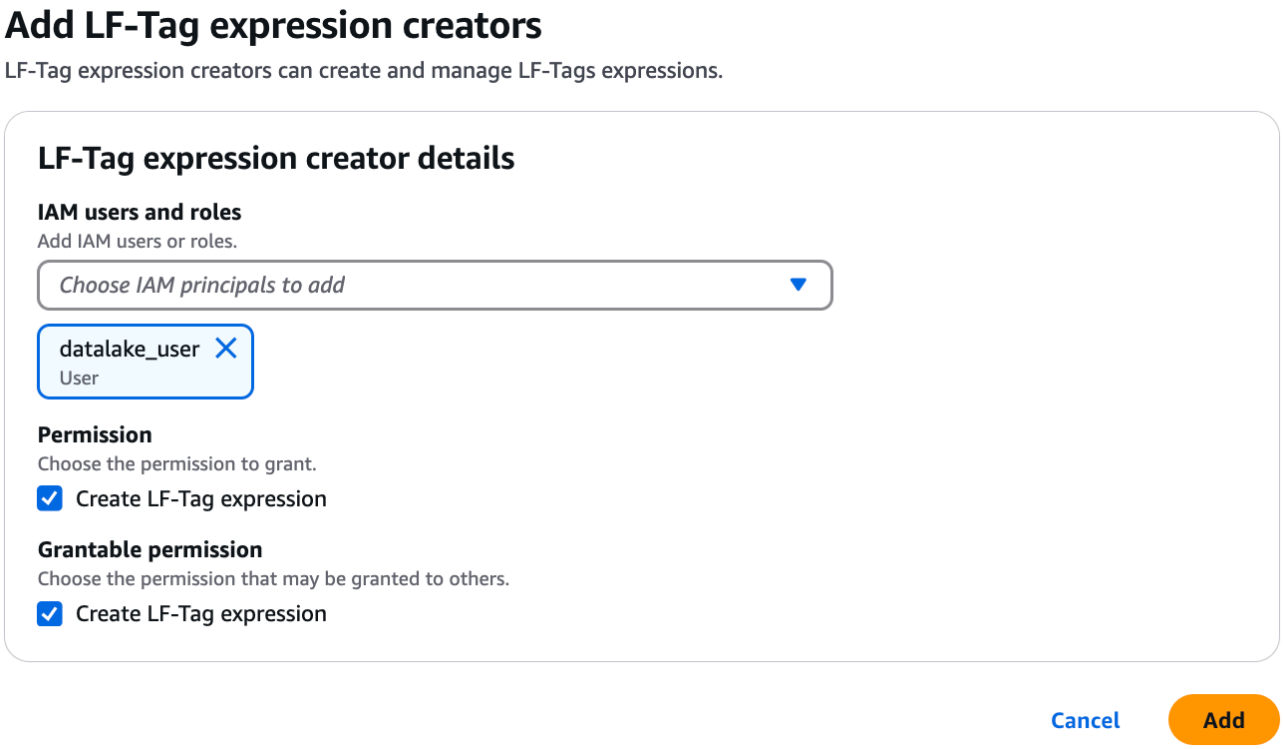
console

Para añadir un creador LF-Tag de expresiones

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador del lago de datos.
2. En el panel de navegación, en Permisos, elija LF-Tags y permisos.
3. Seleccione la pestaña de LF-Tag expresiones.

4. En la sección de creadores de LF-Tag expresiones, selecciona Añadir creadores de LF-Tag expresiones.



5. En la página Añadir creadores de LF-Tag expresiones, elija un rol o usuario de IAM que tenga los permisos necesarios para crear expresiones. LF-Tag
6. Marque la casilla de verificación del permiso `Create LF-Tag expression`.
7. (Opcional) Para que las entidades principales seleccionadas puedan conceder permisos `Create LF-Tag expression` a las entidades principales, seleccione el permiso concedible `Create LF-Tag expression`.
8. Elija Añadir.

AWS CLI

```
aws lakeformation grant-permissions --cli-input-json file://grantCreate
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:user/tag-manager"
    },
    "Resource": {
        "Catalog": {}
```

```
    },
    "Permissions": [
        "CreateLFTagExpression"
    ],
    "PermissionsWithGrantOption": [
        "CreateLFTagExpression"
    ]
}
```

El rol de creador de LF-Tag expresiones tiene la capacidad de crear, actualizar o eliminar LF-Tag expresiones.

Permiso	Description (Descripción)
`Create`	Un director con este permiso puede añadir LF-Tag expresiones en el lago de datos.
`Drop`	Un director con este permiso para una LF-Tag expresión puede eliminar una LF-Tag expresión del lago de datos.
`Alter`	Un director con este permiso para una LF-Tag expresión puede actualiza r el cuerpo de la LF-Tag expresión.
`Describe`	Un director con este permiso para una LF-Tag expresión puede ver el contenido de una LF-Tag expresión.
`Grant with LF-Tag expression`	Este permiso permite al destinatario utilizar la expresión de etiqueta como recurso al conceder permisos de acceso a datos o metadatos . Al conceder `Grant with LF-Tag expression` está otorgando `Describe` de manera implícita.
`Super`	En el caso de las LF-Tag expresiones, el `Super` permiso otorga la capacidad de `Describe` `AlterDrop`,, y concede permisos sobre la expresión de etiqueta a otros principales.

Estos permisos se pueden conceder. Una entidad principal que haya recibido estos permisos con la opción de otorgarlos puede otorgarlos a otras entidades principales.

Crear expresiones LF-Tag

Debe definirlos todos LF-Tags en Lake Formation y asignarlos a los recursos del catálogo de datos antes de que puedan usarse para crear expresiones. Una LF-Tag expresión consta de una clave más y uno o más valores posibles para cada clave.

Una vez que el administrador del lago de datos haya configurado los permisos de IAM y Lake Formation necesarios para la función de creador de LF-Tag expresiones, el director puede crear LF-Tag expresiones reutilizables. El creador de la LF-Tag expresión obtiene permisos implícitos para actualizar el cuerpo de la expresión y eliminarla. LF-Tag

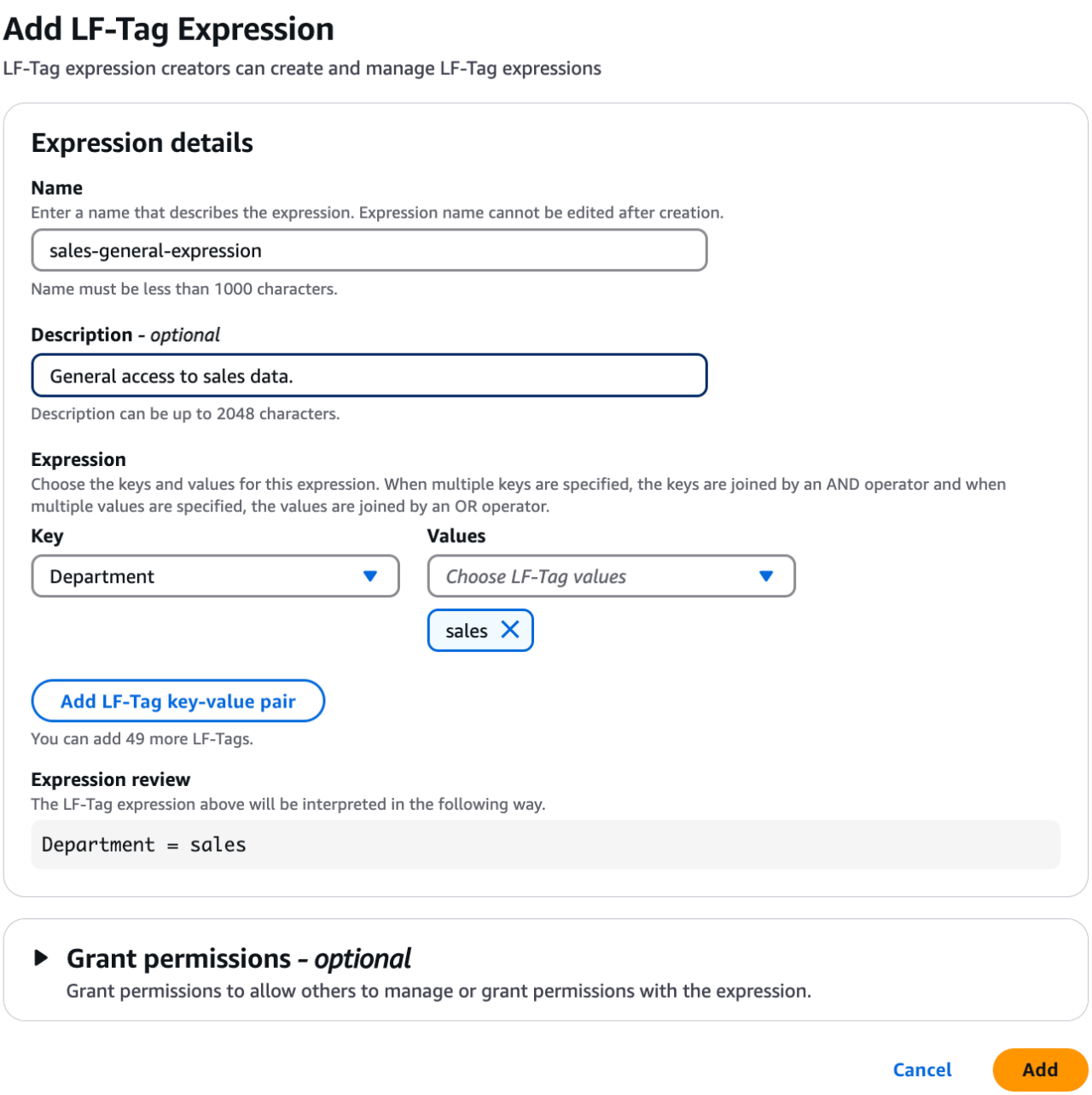
Puede crear LF-Tag expresiones mediante la AWS Lake Formation consola, la API o AWS Command Line Interface (AWS CLI).

Console

Para crear una LF-Tag expresión

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como director con permisos de creador de LF-Tag expresiones o como administrador del lago de datos.
2. En el panel de navegación, en Permisos, elija LF-Tags y permisos.
3. Elija LF-Tag expresiones. Aparece la página Añadir LF-Tag expresiones.



4. Introduzca la información siguiente:

   - Nombre: escriba un nombre exclusivo para la expresión. No puede actualizar el nombre de la expresión.
   - Descripción: opcionalmente, proporcione la descripción de la expresión con sus detalles.
   - Expresión: cree la expresión especificando las claves de etiqueta y sus valores asociados. Puede agregar hasta 50 claves por expresión. Debe tener el permiso `Grant with LF-Tags` de Lake Formation en todas las etiquetas en el cuerpo de la expresión.

Cada clave debe tener un valor como mínimo. Para introducir varios valores, introduzca una lista delimitada por comas y, a continuación, pulse Intro, o introduzca un valor cada vez y pulse Añadir después de cada uno. El número máximo de valores permitidos por clave es 1000.

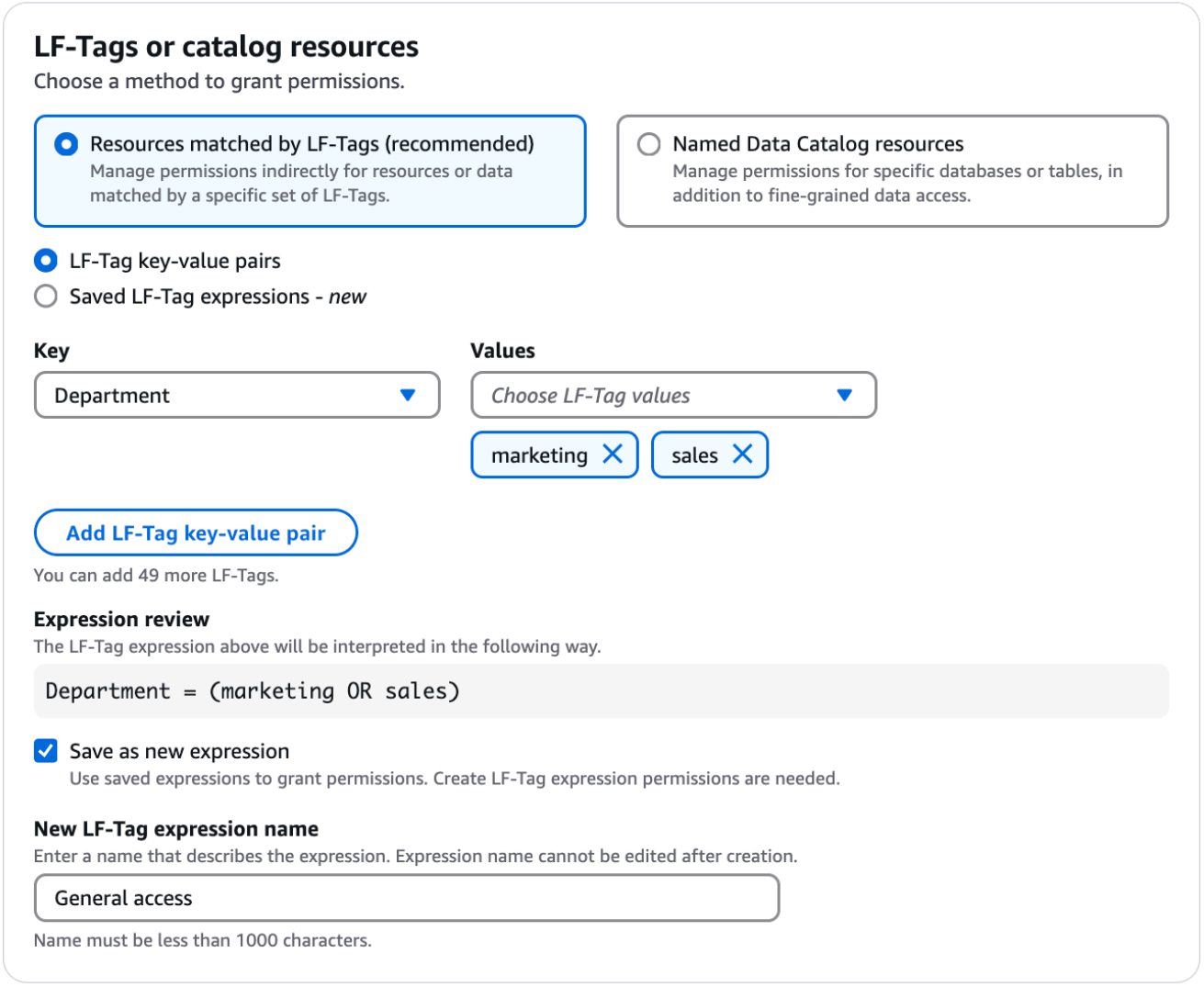
Lake Formation usa la AND/OR lógica para combinar varias claves y valores en una expresión. Dentro de un único par (key : list of values), los valores se combinan mediante el operador lógico OR. Por ejemplo, si el par es (Department : [Sales, Marketing]), significa que la etiqueta coincide si el recurso tiene la etiqueta Departament con el valor Sales O Marketing.

Al especificar varias claves, las une un operador lógico AND. Por lo tanto, si la expresión completa es (Department : [Sales, Marketing]) AND (Location : [US, Canada]), hace coincidir los recursos que tienen la etiqueta Departamento con el valor Ventas O Marketing Y que también tienen la etiqueta Ubicación con el valor EE. UU. O Canadá. El siguiente es otro ejemplo con varias claves y valores:

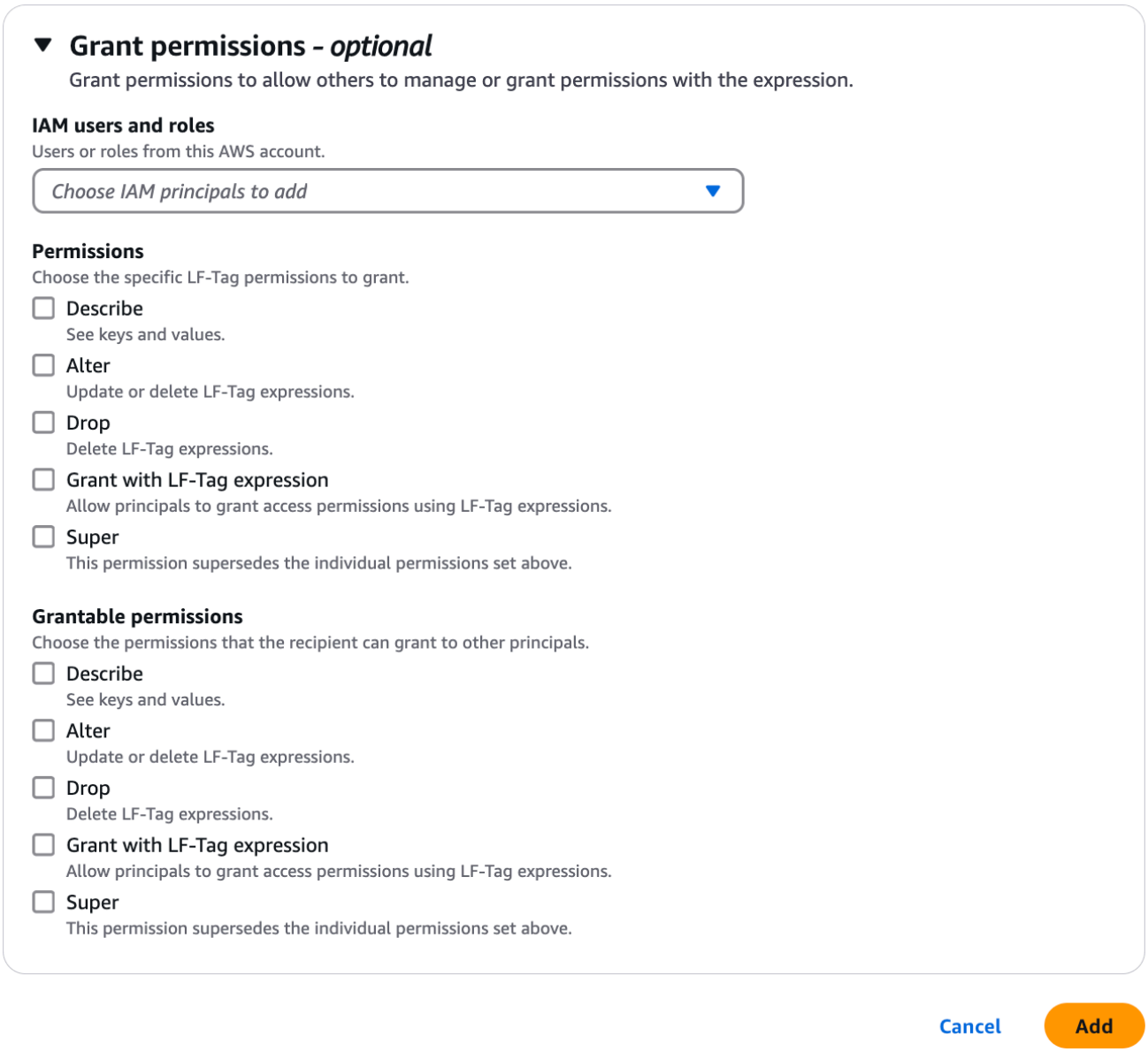
LF-Tag expresión: (ContentType : [Vídeo, audio]) Y (Región: [Europa, Asia]) Y (Departamento: [Ingeniería, ProductManagement]).

Esta expresión coincidiría con los recursos que tengan: - La ContentType etiqueta con el valor Vídeo O Audio Y - La etiqueta Región con el valor Europa O Asia Y - La etiqueta Departamento con el valor Ingeniería ProductManagement O.

También puede guardar una expresión de etiqueta al conceder permisos de lago de datos utilizando LF-Tags. Elija los pares de clave y valor y elija la opción Guardar como nueva expresión. Especifique un nombre que describa la expresión.



5. (Opcional) A continuación users/roles, elija los permisos y los permisos de la expresión que desee concederles en la cuenta. También puede elegir permisos concesibles que permitan a los usuarios conceder estos permisos a otros usuarios de la cuenta. No puede conceder permisos entre cuentas para las expresiones de etiquetas.



6. Elija Añadir .

AWS CLI

Para crear una LF-Tag expresión

- Introduzca un comando `create-lf-tag-expression`.

  En el siguiente ejemplo, se crea una LF-Tag expresión `Department` con la etiqueta con valores `Sales` y `Marketing` Y la etiqueta `Location` con el valorUS.

```
aws lakeformation create-lf-tag-expression \
```

```
-- name "my-tag-expression" \
-- catalog-id "123456789012" \
-- expression '{"Expression":[{"TagKey":"Department","TagValues":
["Sales","Marketing"]},{"TagKey":"Location","TagValues":["US"]}]}'
```

Este comando CLI crea una nueva LF-Tag expresión en AWS Glue Data Catalog. La expresión se puede utilizar para conceder permisos a los recursos del Catálogo de datos, como bases de datos, tablas, vistas o columnas, en función de sus etiquetas asociadas. En este ejemplo, la expresión hará coincidir los recursos que tengan la clave `Department` con los valores `Sales` o `Marketing`, y la clave `Location` con el valor `US`.

Como creador de expresiones de etiquetas, el director obtiene `Alter` permiso para utilizar esta LF-Tag expresión y puede actualizarla o eliminarla. El director creador de la LF-Tag expresión también puede conceder `Alter` permiso a otro director para que actualice y elimine esta expresión.

Actualización de LF-Tag expresiones

Solo los administradores del lago de datos, el creador de la LF-Tag expresión y los directores que tengan `Alter` o tengan `Super` permiso sobre la LF-Tag expresión pueden actualizar una LF-Tag expresión. Además del permiso `Alter`, también necesita el permiso `lakeformation:UpdateLFTagExpression` de IAM y el permiso `Grant with LF-Tag` para todas las claves-valores subyacentes del cuerpo de la nueva expresión para actualizar las expresiones.

Para actualizar una LF-Tag expresión, debe actualizar la descripción, el cuerpo de la expresión y los permisos otorgados a la expresión. No puede cambiar el nombre de la LF-Tag expresión. Para cambiar el nombre, elimine la LF-Tag expresión y añada una con los parámetros necesarios.

Puede actualizar una LF-Tag expresión mediante la AWS Lake Formation consola, la API o AWS Command Line Interface (AWS CLI).

Console

Para actualizar una LF-Tag expresión

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador del lago de datos, LF-Tag creador o director con `Alter` permiso en el LF-Tag.
2. En el panel de navegación, en permisos, selecciona LF-Tags y permisos.

3. Seleccione la pestaña de LF-Tag expresiones.
4. En la sección de LF-Tag expresiones, seleccione una LF-Tag expresión y, a continuación, elija Editar.
5. En el cuadro de diálogo Editar LF-Tag expresión, actualice la descripción y actualice el cuerpo de la expresión añadiendo o quitando claves y valores.

   Para agregar varios valores, en el campo Valores, elija los valores en el menú desplegable.
6. Seleccione Save.

AWS CLI

El comando update-lf-tag-expression de Lake Formation permite actualizar una expresión existente. LF-Tag

```
aws lakeformation update-lf-tag-expression \
-- name expression_name\
-- description new_description \
-- catalog-id catalog_id \
-- expression '{"Expression": [{"TagKey": "tag_key", "TagValues": ["tag_value1",
 "tag_value2", ...]}]}'
```

Esto es lo que significan los parámetros del comando proporcionado:

- name: nombre de la expresión de etiqueta con nombre existente que desea actualizar.
- description: una nueva descripción de la expresión.

  catalog-id: el ID del Catálogo de datos donde reside la expresión de etiqueta con nombre.
- expression: la nueva cadena de expresión de etiqueta con la que desea actualizar la expresión.

LF-Tag Eliminar expresiones

Puede eliminar LF-Tag las expresiones que ya no se utilizan. Si ha concedido permisos a los directores en los recursos del catálogo de datos mediante la LF-Tag expresión, dejarán de tener los permisos.

Solo los administradores del lago de datos, el creador de la LF-Tag expresión o un director con `Drop` permiso sobre la LF-Tag expresión pueden eliminar una LF-Tag expresión. Además del `Drop`

permiso, el director también necesita el permiso de `lakeformation:DeleteLFTagExpression` IAM para eliminar una LF-Tag expresión.

Puede eliminar una LF-Tag expresión mediante la AWS Lake Formation consola, la API o AWS Command Line Interface (AWS CLI).

Console

Para eliminar una LF-Tag expresión (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador del lago de datos, creador de la LF-Tag expresión o responsable que tenga permisos para eliminar la expresión.
2. En el panel de navegación, en Permisos, elija LF-Tags y permisos.
3. Seleccione la pestaña LF-Tag de expresiones.
4. En la sección de LF-Tag expresiones, seleccione una LF-Tag expresión y, a continuación, elija Eliminar.
5. ¿En la opción Eliminar LF-Tag expresión? cuadro de diálogo, para confirmar la eliminación, introduzca el nombre de la LF-Tag expresión en el campo designado y, a continuación, elija Eliminar.

AWS CLI

Para eliminar un LF-Tag (AWS CLI)

- Introduzca un comando `delete-lf-tag-expression`. Proporcione el nombre y el ID de catálogo de la expresión que desea eliminar.

  Example

  En el siguiente ejemplo, se elimina la LF-Tag expresión con el nombre `my-tag-expression` del catálogo de datos con el identificador`123456789012`. El `catalog-id` parámetro es opcional si utiliza la misma cuenta que en su AWS CLI configuración. Tras eliminar una LF-Tag expresión, Lake Formation limpia los registros de permisos asociados a esa expresión. Esto incluye tanto los registros de permisos individuales como los registros de permisos agregados que contienen la expresión eliminada.

  ```
  aws lakeformation delete-lf-tag-expression \
  ```

```
--name "my-tag-expression" \
--catalog-id "123456789012"
```

Enumerar LF-Tag las expresiones

Puede enumerar las LF-Tag expresiones para las que tiene los permisos de descripción. Los administradores de los lagos de datos, los creadores de LF-Tag expresiones y Read-only los administradores pueden ver implícitamente todas las expresiones de etiquetas de sus cuentas.

Puede enumerar LF-Tag las expresiones mediante la AWS Lake Formation consola, la API o AWS Command Line Interface (AWS CLI).

Console

Para enumerar LF-Tag expresiones (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como creador de LF-Tag expresiones, administrador de un lago de datos o como director al que se le hayan concedido permisos sobre LF-Tag las expresiones y que tenga el permiso de `lakeformation:ListLFTagExpressions` IAM.
2. En el panel de navegación, en Permisos LF-Tags y permisos.
3. Seleccione la pestaña de LF-Tag expresiones para ver las expresiones. En esta sección se muestra la información sobre LF-Tag las expresiones existentes, incluido el nombre de la expresión, la propia expresión con enlaces a las etiquetas incluidas y las opciones para crear, editar o eliminar expresiones.

AWS CLI

Para enumerar LF-Tags (AWS CLI)

- Para enumerar LF-Tag expresiones mediante el AWS CLI, puede utilizar el comando list-lf-tag-expressions. La sintaxis de la solicitud es la siguiente:

```
aws lakeformation list-lf-tag-expressions \
-- catalog-id "123456789012" \
-- max-items "100" \
-- next-token "next-token"
```

Donde:

- `catalog-id` es el identificador de AWS cuenta del catálogo de datos para el que desea incluir las expresiones de etiquetas.
- `max-items` especifica el número máximo de expresiones de etiquetas que se van a devolver. Si este parámetro no se utiliza, el valor predeterminado es 100.
- `next-token` es un token de continuación si los resultados se truncaron en una solicitud anterior.

La respuesta incluirá una lista de LF-Tag expresiones y un siguiente token, si corresponde.

Administrar los permisos de LF-Tag valor

Puede conceder los `Alter` permisos a `Drop` los directores LF-Tags para que administren las expresiones de LF-Tag valor. También puede conceder `Describe` `Grant with LF-Tag expressions` permisos a los directores LF-Tags para ver LF-Tags y asignarlos a los recursos del catálogo de datos (bases de datos, tablas y columnas). `Associate` Cuando LF-Tags se asignan a los recursos del catálogo de datos, puede utilizar el método de control de acceso basado en etiquetas (LF-TBAC) de Lake Formation para proteger esos recursos. Para obtener más información, consulte Control de acceso basado en etiquetas de Lake Formation.

Puede conceder estos permisos con la opción de concesión para que otras entidades principales puedan otorgarlos. Los permisos `Grant with LF-Tag expressions`, `Describe` y `Associate` se explican en Añadir creadores LF-Tag.

Puede conceder los `Associate` permisos `Describe` y LF-Tag a una AWS cuenta externa. Un administrador del lago de datos de esa cuenta puede entonces conceder esos permisos a otras entidades principales de la cuenta. Los directores a los que el administrador del lago de datos de la cuenta externa les conceda el `Associate` permiso pueden entonces LF-Tags asignar los recursos del catálogo de datos que haya compartido con su cuenta.

Al hacer concesiones a una cuenta externa, debe incluir la opción de concesión.

Puede conceder permisos LF-Tags mediante la consola de Lake Formation, la API o AWS Command Line Interface (AWS CLI).

Temas

- Listar LF-Tag los permisos mediante la consola
- LF-Tag Otorgar permisos mediante la consola
- Administrar los permisos mediante el LF-Tag AWS CLI

Para obtener más información, consulte Gestión del control LF-Tags de acceso a los metadatos y Control de acceso basado en etiquetas de Lake Formation.

Listar LF-Tag los permisos mediante la consola

Puede utilizar la consola de Lake Formation para ver los permisos concedidos en LF-Tags. Debe ser un LF-Tag creador, un administrador de un lago de datos o tener el `Associate` permiso `Describe` o permiso de un lago de datos LF-Tag para verlo.

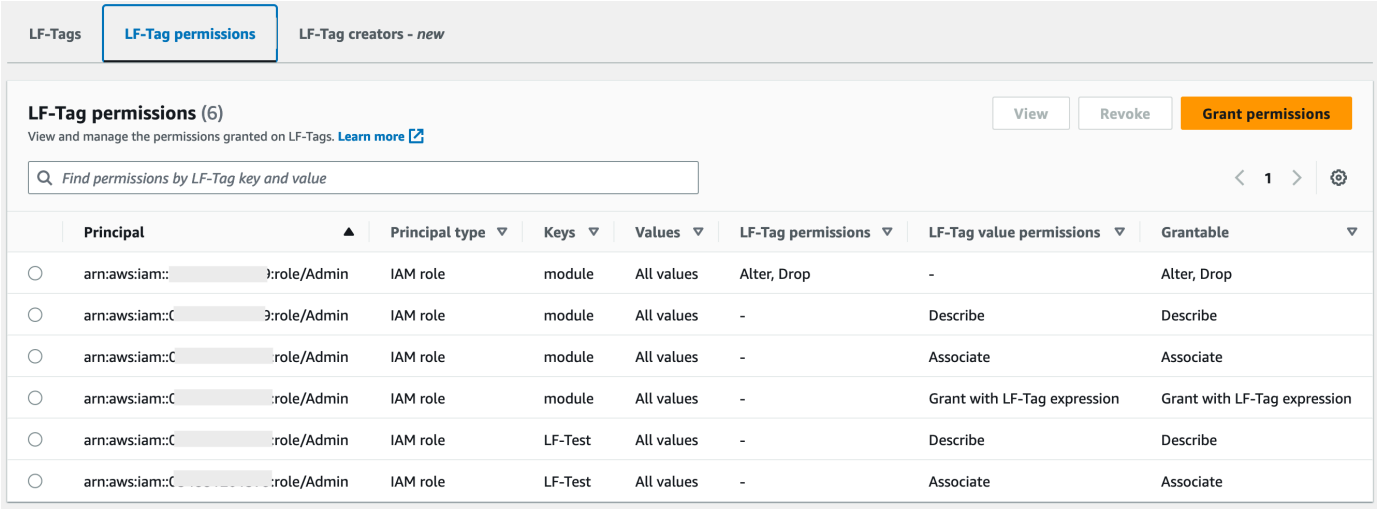
Para enumerar LF-Tag los permisos (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como LF-Tag creador, administrador de un lago de datos o como usuario al que se le LF-Tags hayan otorgado `Describe` los permisos `Drop` `AlterAssociate`,,, o.

2. En el panel de navegación, en Permisos, elija LF-Tags un permiso y elija la sección de LF-Tag permisos.

   La sección de LF-Tag permisos muestra una tabla que contiene el principal, las claves de etiqueta, los valores y los permisos.



LF-Tag Otorgar permisos mediante la consola

En los siguientes pasos se explica cómo conceder permisos LF-Tags mediante la página Otorgar LF-Tag permisos de la consola de Lake Formation. La página está dividida en las estas secciones:

- Tipos de permiso. El tipo de permiso que se va a conceder.
- Entidades principales: los usuarios o roles de IAM, o los usuarios o roles de SAML a los que conceder permisos.
- LF-Tag permisos de par clave-valor: los pares LF-Tag clave-valor a los que se van a conceder permisos.
- LF-Tag permisos: los permisos a los que se van LF-Tags a conceder los permisos.
- LF-Tag permisos de expresión: los permisos LF-Tags para conceder permisos.
- Permisos. Los permisos que se van a conceder.

Abra el icono Otorgar LF-Tag permisos page

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como LF-Tag creador, administrador de un lago de datos o como usuario. LF-Tag Los permisos o pares LF-Tag de valores clave se LF-Tags han concedido con esta opción. `Grant`
2. En el panel de navegación, elija la sección de permisos LF-Tags y LF-Tag permisos.
3. Elija Conceder permisos.

Especificar el tipo de permisos

En la sección Tipo de permisos, seleccione uno.

LF-Tag permisos

Elija los LF-Tag permisos que permiten a los directores actualizar LF-Tag valores o eliminarlos LF-Tags.

LF-Tag permisos de pares clave-valor

Elija los permisos del par LF-Tag clave-valor para permitir a los directores asignar LF-Tags los recursos, vistas LF-Tags y valores del catálogo de datos, y conceder permisos LF-Tags basados en los recursos del catálogo de datos a los directores.

Las opciones disponibles en las siguientes secciones dependen del tipo de permiso.

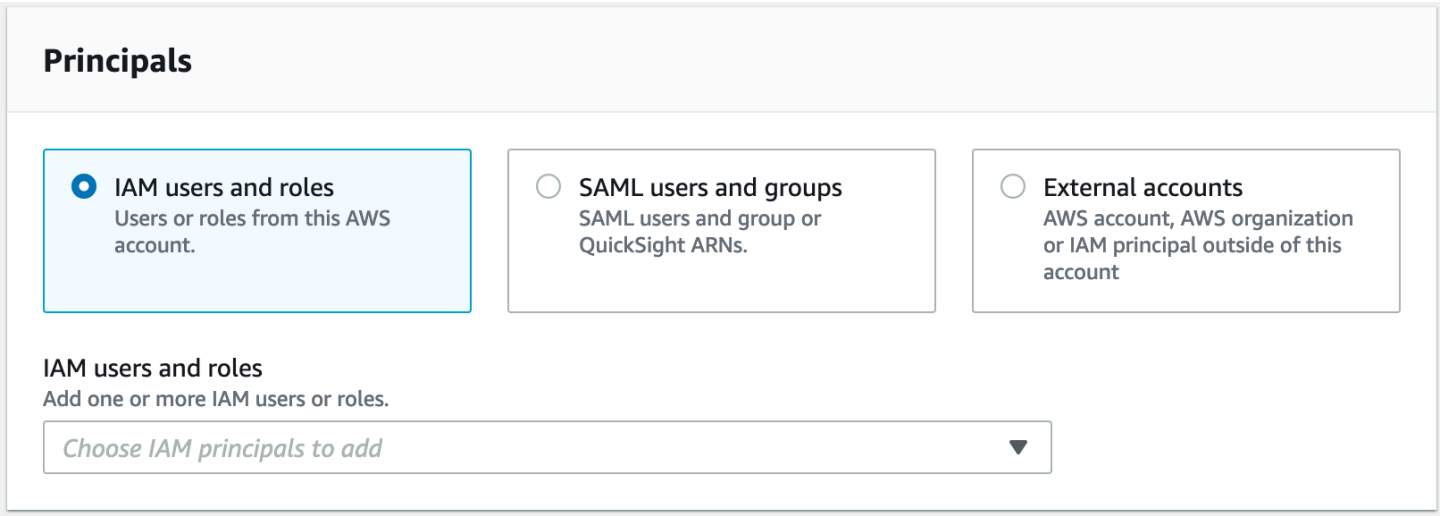
LF-Tag permisos de expresión

Elija los permisos de LF-Tag expresión para permitir a los directores actualizar o eliminar expresiones.

Especificar las entidades principales

> **Note**
>
> No puedes conceder LF-Tag permisos (ni`Drop`) `Alter` a cuentas externas ni a los directores de otra cuenta.

En la sección Entidades principales, elija uno de los tipos y, a continuación, especifique las que van a recibir los permisos concedidos.



Usuarios y roles de IAM

Elija uno o varios usuarios o roles en la lista de usuarios y roles de IAM.

Usuarios y grupos de SAML

Para los usuarios y grupos de SAML y Quick, introduzca uno o más nombres de recursos de Amazon (ARN) para los usuarios o grupos federados a través de SAML, o ARN para los usuarios o grupos de Quick. Pulse Intro después de cada ARN.

Para obtener información sobre cómo crear ARN, consulte Concesión y revocación de Lake Formation AWS CLI Comandos de la de.

> **Note**
>
> La integración de Lake Formation con Quick solo es compatible con Quick Enterprise Edition.

Cuentas externas

Para la AWS cuenta, introduzca uno o más identificadores de AWS cuenta válidos. Pulse Intro después de cada ID.

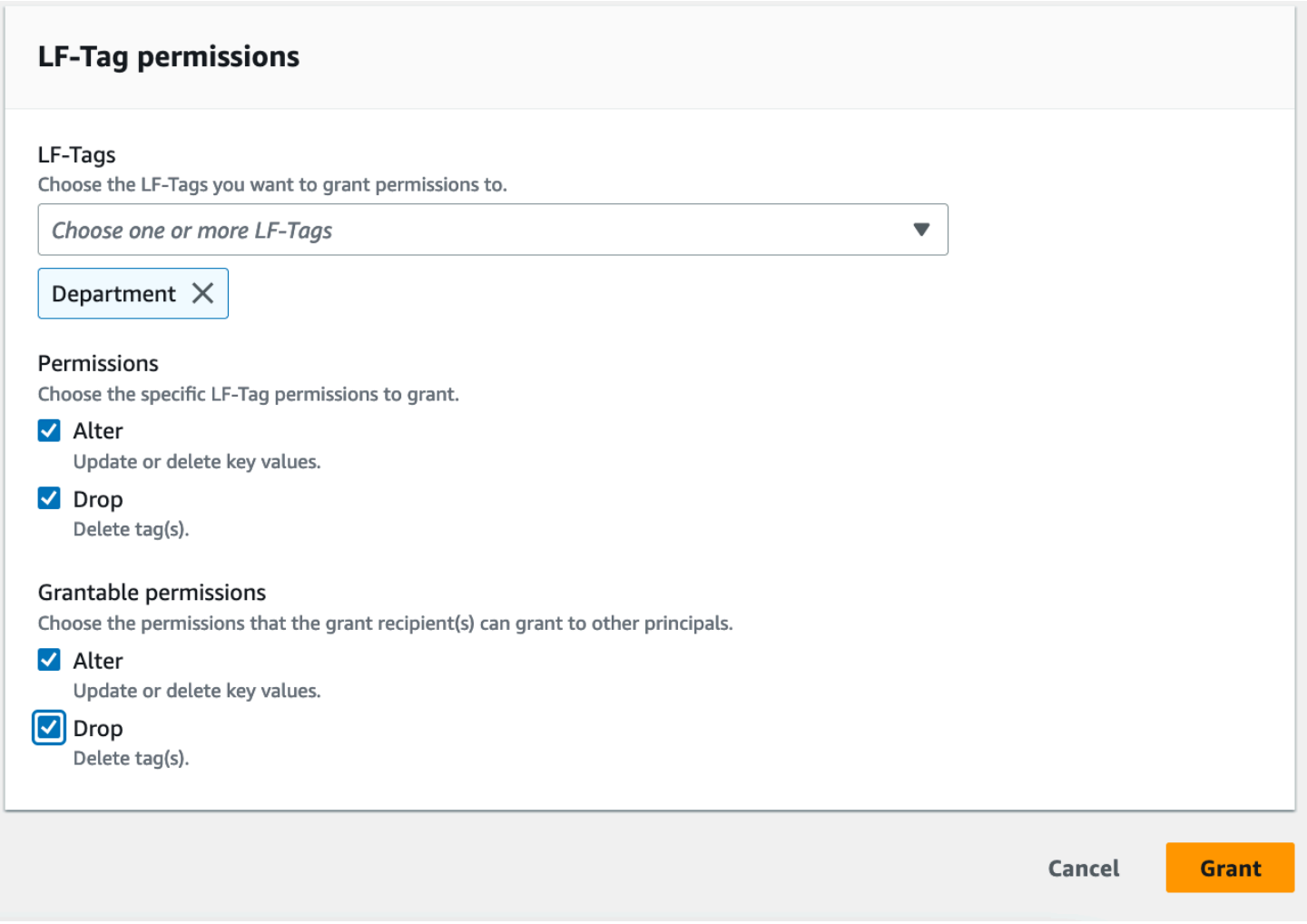
El ID de una organización está formado por una "o-" seguida de 10 a 32 letras minúsculas o dígitos.

Un ID de una unidad organizativa comienza por «ou-» seguidos de 4 a 32 letras minúsculas o dígitos (el ID de la raíz que contiene la UO). Esta cadena va seguida de un segundo guion «-» y de 8 a 32 letras minúsculas o dígitos adicionales.

Para la entidad principal de IAM, escriba el ARN del usuario o el rol de IAM.

Especifique el LF-Tags

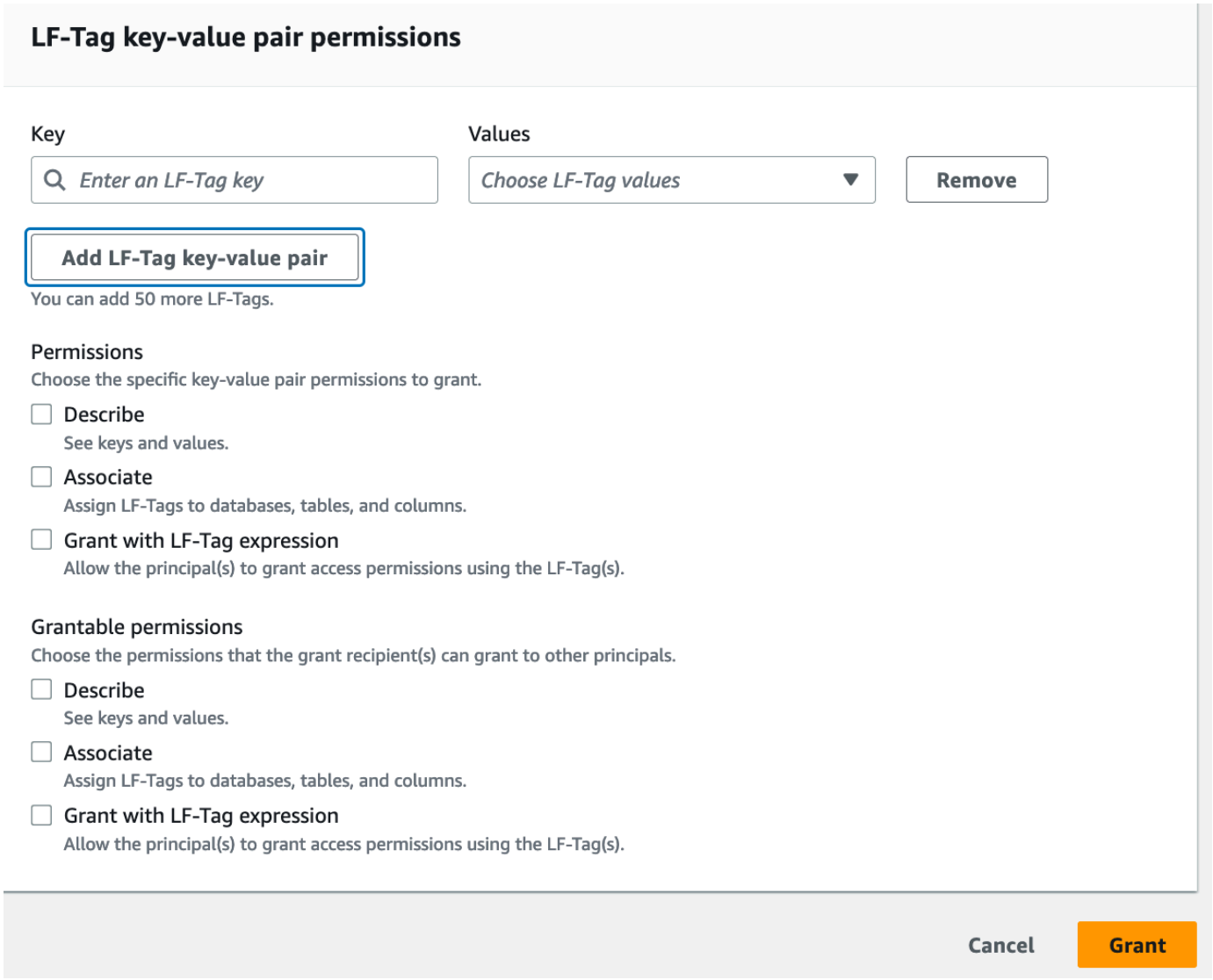
Para conceder permisos LF-Tags, en la sección de LF-Tag permisos, especifique el objeto al LF-Tags que se van a conceder los permisos.



- Elija uno o más LF-Tag mediante el menú desplegable.

Especifique los pares LF-Tag clave-valor

1. Para conceder permisos a los pares LF-Tag clave-valor (primero debe elegir permisos para los pares LF-Tag clave-valor como tipo de permiso), seleccione Añadir par clave-valor para que aparezca la primera fila de campos en los que se especifican la LF-Tag clave y los valores. LF-Tag



2. Sitúe el cursor en el campo Clave y, si lo desea, empiece a escribir para reducir la lista de selección y seleccione una tecla. LF-Tag
3. En la lista Valores, seleccione uno o varios y, a continuación, pulse la tecla Etiqueta, haga clic o pulse fuera del campo para guardar los valores seleccionados.

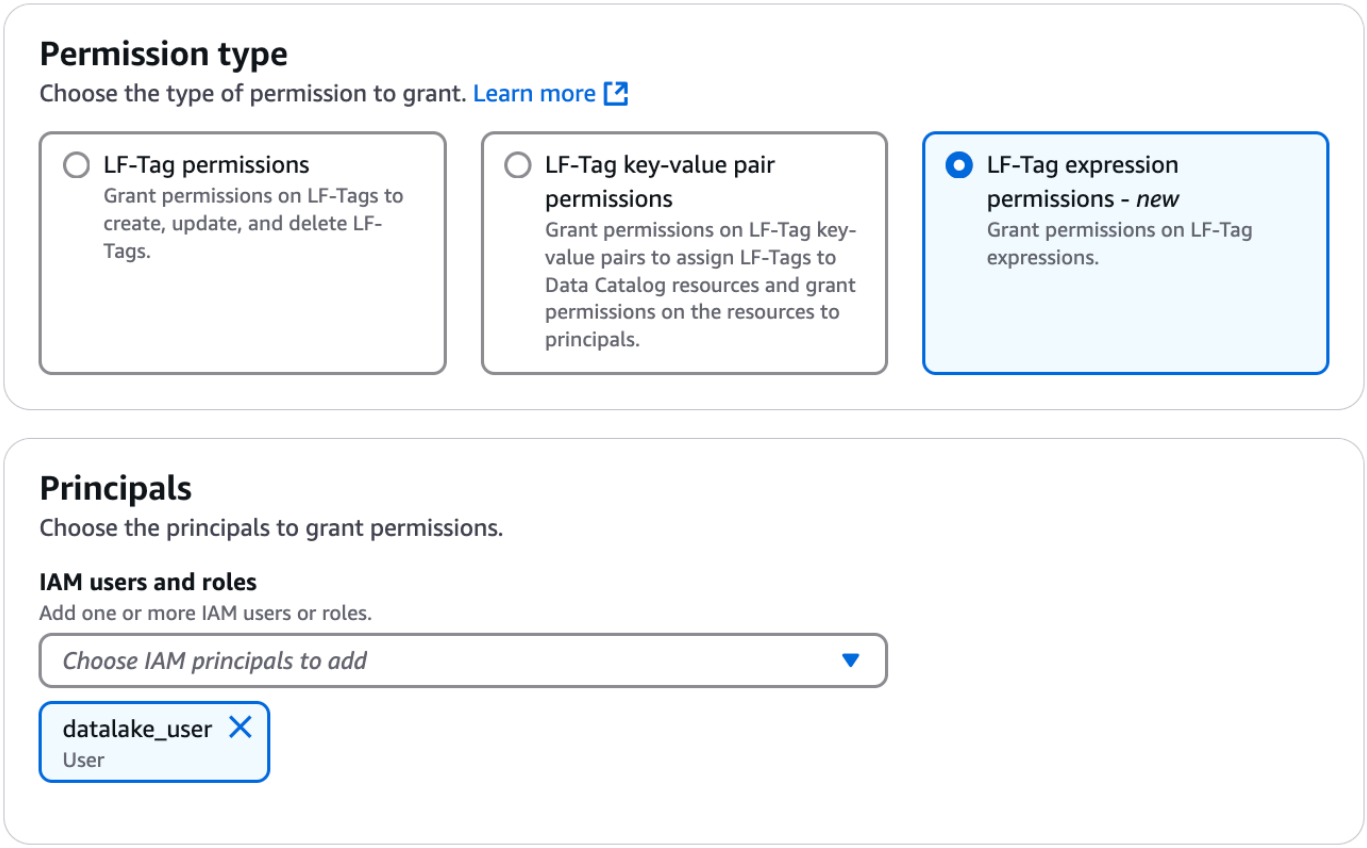
> **Note**
>
> Si una de las filas de la lista Valores tiene el foco, al pulsar Intro se selecciona o anula la selección de la casilla.

Los valores seleccionados aparecen como mosaicos debajo de la lista Valores. Elija la ✖ para eliminar un valor. Seleccione Eliminar para eliminar todo LF-Tag.

4. Para añadir otro LF-Tag, selecciona Añadir de LF-Tag nuevo y repite los dos pasos anteriores.

Especifique las LF-Tag expresiones

1. Para conceder permisos a LF-Tag las expresiones (primero debe elegir elegir permisos de LF-Tag expresión como tipo de permiso).



2. Elija una expresión. LF-Tag
3. Las expresiones seleccionadas aparecen como mosaicos debajo de la lista de LF-Tag expresiones. Elija la ✖ para eliminar una expresión.
4. Para añadir otra LF-Tag expresión, elija otra expresión.

Especificar los permisos

En esta sección se muestran los LF-Tag permisos o los permisos de LF-Tag valor según el tipo de permiso que haya elegido en el paso anterior.

Según el tipo de permiso que haya decidido conceder, seleccione LF-Tag los permisos o los permisos de par LF-Tag clave-valor y los permisos que se pueden conceder.

1. En LF-Tag Permisos, selecciona los permisos que deseas conceder.

   Al conceder Eliminar y Modificar se concede implícitamente Describe.

   Debe conceder los permisos Alterar y Eliminar para todos los valores de las etiquetas.
2. En Permisos LT-Tag con valores clave-valor, seleccione los permisos que desee conceder.

   Al conceder Asociar se concede implícitamente Describir. Elija Otorgar con LF-Tag expresión para permitir que el destinatario de la subvención conceda o revoque los permisos de acceso a los recursos del catálogo de datos mediante el método. LF-TBAC
3. En permisos de LF-Tag expresión, seleccione los permisos que desee conceder.

   Al conceder Eliminar y Modificar se concede implícitamente Describe.

   Al conceder el permiso Super, se conceden todos los permisos disponibles.
4. (Opcional) En Permisos concedibles, selecciona los permisos que el destinatario de la subvención puede conceder a otros directores de su cuenta. AWS
5. Elija Conceder.

Administrar los permisos mediante el LF-Tag AWS CLI

Puede conceder, revocar y enumerar permisos LF-Tags mediante AWS Command Line Interface (AWS CLI).

Para enumerar LF-Tag los permisos (AWS CLI)

- Introduzca un comando `list-permissions`. Debe ser el LF-Tag creador, el administrador de un lago de datos o tener el `Grant with LF-Tag permissions` permiso `Drop` `Alter` `DescribeAssociate`,,,, LF-Tag para verlo.

  El siguiente comando solicita todos LF-Tags aquellos sobre los que tiene permisos.

  ```
  aws lakeformation list-permissions --resource-type LF_TAG
  ```

  El siguiente es un ejemplo de resultado para un administrador de un lago de datos, que ve todo LF-Tags concedido a todos los directores. Non-administrative los usuarios ven que solo LF-Tags se les concede. LF-Tag los permisos concedidos desde una cuenta externa aparecen en una

página de resultados independiente. Para verlos, repita el comando y suministre el `--next-token` argumento con el token devuelto por la ejecución del comando anterior.

```
{
    "PrincipalResourcePermissions": [
        {
            "Principal": {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_admin"
            },
            "Resource": {
                "LFTag": {
                    "CatalogId": "111122223333",
                    "TagKey": "environment",
                    "TagValues": [
                        "*"
                    ]
                }
            },
            "Permissions": [
                "ASSOCIATE"
            ],
            "PermissionsWithGrantOption": [
                "ASSOCIATE"
            ]
        },
        {
            "Principal": {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
            },
            "Resource": {
                "LFTag": {
                    "CatalogId": "111122223333",
                    "TagKey": "module",
                    "TagValues": [
                        "Orders",
                        "Sales"
                    ]
                }
            },
            "Permissions": [
                "DESCRIBE"
```

```
            ],
            "PermissionsWithGrantOption": []
        },
...
    ],
    "NextToken": "eyJzaG91bGRRdWVy...Wlzc2lvbnMiOnRydWV9"
}
```

Puedes enumerar todas las concesiones para una LF-Tag clave específica. El siguiente comando devuelve todos los permisos concedidos en LF-Tag `module`.

```
aws lakeformation list-permissions --resource-type LF_TAG --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

También puede enumerar LF-Tag los valores otorgados a una entidad principal específica para una entidad específica LF-Tag. Al proporcionar el argumento `--principal`, debe indicar el argumento `--resource`. Por lo tanto, el comando solo puede solicitar de manera efectiva los valores otorgados a un principal específico para una LF-Tag clave específica. El siguiente comando muestra cómo hacerlo para el principal `datalake_user1` y la LF-Tag clave`module`.

```
aws lakeformation list-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --resource-type LF_TAG --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

A continuación, se muestra un ejemplo del resultado.

```
{
    "PrincipalResourcePermissions": [
        {
            "Principal": {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
            },
            "Resource": {
                "LFTag": {
                    "CatalogId": "111122223333",
                    "TagKey": "module",
                    "TagValues": [
                        "Orders",
```

```
                        "Sales"
                    ]
                }
            },
            "Permissions": [
                "ASSOCIATE"
            ],
            "PermissionsWithGrantOption": []
        }
    ]
}
```

Para conceder permisos en LF-Tags (AWS CLI)

1. Introduzca un comando similar al siguiente. En este ejemplo, se concede `datalake_user1` al usuario el `Associate` permiso correspondiente LF-Tag a la clave`module`. Concede permisos para ver y asignar todos los valores de esa clave, como indica el asterisco (*).

   ```
   aws lakeformation grant-permissions --principal
    DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
   datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
    {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
   ```

   Al conceder el permiso `Associate`, se otorga el permiso `Describe` de forma implícita.

   En el siguiente ejemplo, se concede `Associate` a la AWS cuenta externa 1234-5678-9012 que contiene la clave`module`, LF-Tag con la opción de concesión. Otorga permisos para ver y asignar solo los valores `sales` y `orders`.

   ```
   aws lakeformation grant-permissions --principal
    DataLakePrincipalIdentifier=123456789012 --permissions "ASSOCIATE"
    --permissions-with-grant-option "ASSOCIATE" --resource '{ "LFTag":
    {"CatalogId":"111122223333","TagKey":"module","TagValues":["sales", "orders"]}}'
   ```

2. Al conceder el permiso `GrantWithLFTagExpression`, se otorga el permiso `Describe` de forma implícita.

   En el siguiente ejemplo, se concede `GrantWithLFTagExpression` a un usuario con la clave`module`, LF-Tag con la opción de concesión. Concede permisos para ver y conceder permisos sobre los recursos del Catálogo de datos utilizando solo los valores `sales` y `orders`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "GrantWithLFTagExpression"
 --permissions-with-grant-option "GrantWithLFTagExpression" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["sales", "orders"]}}'
```

3. En el siguiente ejemplo, se conceden `Drop` permisos a un usuario LF-Tag con la clave`module`, con la opción de concesión. Otorga permisos para eliminar el LF-Tag. Para eliminar una LF-Tag, necesita permisos en todos los valores de esa clave.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "DROP"
 --permissions-with-grant-option "DROP" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

4. En el siguiente ejemplo, se conceden `Alter` permisos al usuario LF-Tag con la clave`module`, con la opción de concesión. Otorga permisos para eliminar el LF-Tag. Para actualizar a LF-Tag, necesita permisos en todos los valores de esa clave.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "ALTER"
 --permissions-with-grant-option "ALTER" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Para revocar los permisos de LF-Tags (AWS CLI)

- Introduzca un comando similar al siguiente. En este ejemplo, se revoca el `Associate` permiso del LF-Tag usuario con la clave`module`. `datalake_user1`

```
aws lakeformation revoke-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Otorgar permisos de lago de datos mediante el método LF-TBAC

Puede conceder los permisos de `ASSOCIATE` Lake Formation `DESCRIBE` y de Lake Formation LF-Tags a los directores para que puedan ver LF-Tags y asignarlos a los recursos del catálogo de datos (bases de datos, tablas, vistas y columnas). Cuando LF-Tags se asignan a los recursos del catálogo

de datos, puede utilizar el método de control de acceso basado en etiquetas (LF-TBAC) de Lake Formation para proteger esos recursos. Para obtener más información, consulte Control de acceso basado en etiquetas de Lake Formation.

Al principio, solo el administrador del lago de datos puede conceder estos permisos. Si el administrador del lago de datos concede estos permisos con la opción de concesión, otras entidades principales podrán otorgarlos. Los permisos `DESCRIBE` y `ASSOCIATE` se explican en Prácticas recomendadas y consideraciones sobre el control de acceso basado en etiquetas de Lake Formation.

Puede conceder los `ASSOCIATE` permisos `DESCRIBE` y LF-Tag a una AWS cuenta externa. Un administrador del lago de datos de esa cuenta puede entonces conceder esos permisos a otras entidades principales de la cuenta. Los directores a los que el administrador del lago de datos de la cuenta externa les conceda el `ASSOCIATE` permiso pueden entonces LF-Tags asignar los recursos del catálogo de datos que haya compartido con su cuenta.

Al hacer concesiones a una cuenta externa, debe incluir la opción de concesión.

Puede conceder permisos LF-Tags mediante la AWS Lake Formation consola, la API o el AWS Command Line Interface (AWS CLI).

> **Note**
>
> Los siguientes pasos no son necesarios para los catálogos de Tablas de S3. Puede utilizarlos LF-Tags para conceder permisos en los catálogos de S3 Tables existentes sin eliminarlos ni volver a crearlos.

Habilitar LF-Tags el soporte para los catálogos federados existentes que utilizan los permisos de Lake Formation

Siga estos pasos si tiene catálogos federados existentes que utilizan permisos de Lake Formation, como Amazon Redshift, Amazon DynamoDB o catálogos que se crearon LF-Tags antes de que estuviera disponible el soporte para los catálogos federados.

1. Eliminar el catálogo existente: llame a la operación de la API `deleteCatalog` para eliminar el catálogo federado existente que usa los permisos de Lake Formation.
2. Cree un nuevo catálogo federado: cree un catálogo nuevo y dirija el nuevo catálogo al existente. namespace/datashare

Use un nombre nuevo para el catálogo: este proceso actualiza los catálogos federados preexistentes para que sean compatibles con la funcionalidad. LF-Tag Si desea utilizar el mismo nombre de catálogo, póngase en contacto con el equipo de AWS soporte para obtener ayuda.

Temas

- Concesión de permisos del catálogo de datos

Véase también

- Administrar los permisos de LF-Tag valor
- Gestión del control LF-Tags de acceso a los metadatos
- Control de acceso basado en etiquetas de Lake Formation

Concesión de permisos del catálogo de datos

Utilice la consola de Lake Formation o conceda permisos AWS CLI a Lake Formation en las bases de datos, tablas, vistas y columnas del catálogo de datos mediante el método de control de acceso basado en etiquetas (LF-TBAC) de Lake Formation.

Console

En los siguientes pasos se explica cómo conceder permisos mediante el método de control de acceso basado en etiquetas (LF-TBAC) de Lake Formation y la página de concesión de permisos de lagos de datos de la consola de Lake Formation. La página está dividida en las secciones siguientes:

- Responsables: los usuarios, las funciones y Cuentas de AWS a los que se deben conceder los permisos.
- LF-Tags o recursos del catálogo: las bases de datos, las tablas o los enlaces a los recursos para conceder permisos.
- Permisos. Los permisos de Lake Formation que se conceden.

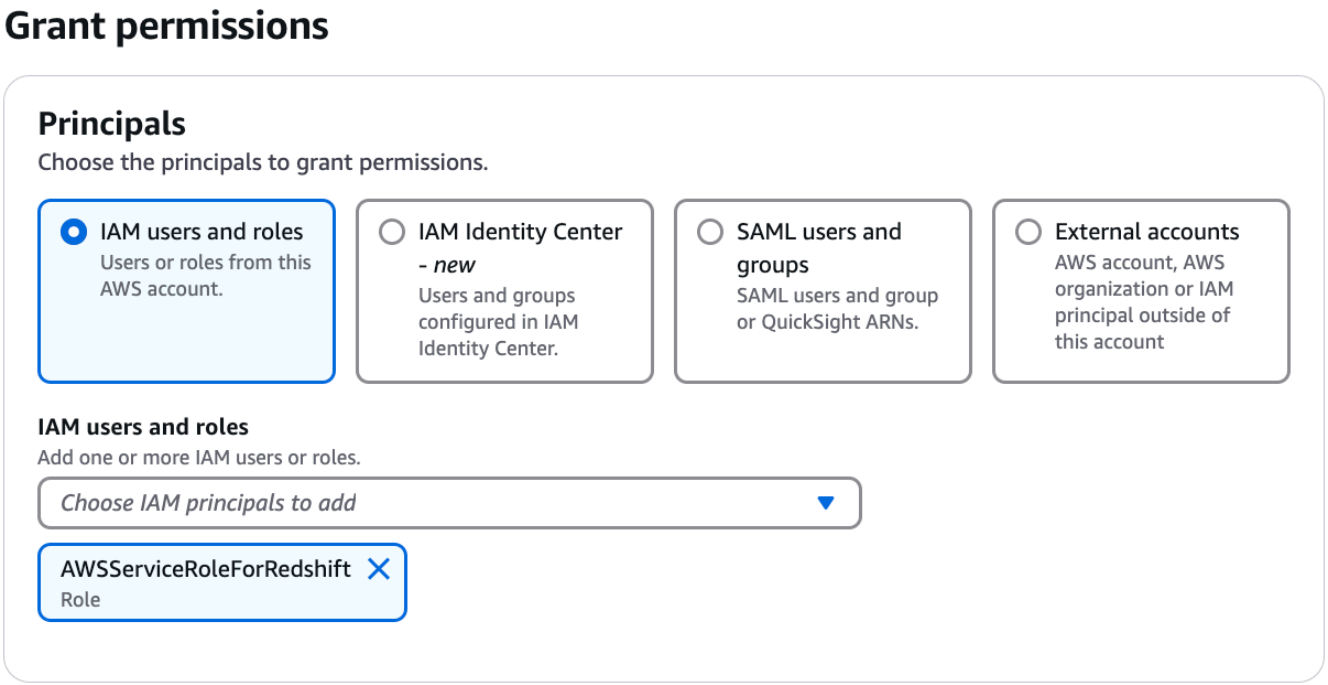
1. Abra la página Conceder permisos de lagos de datos.

   Abra la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/e inicie sesión como administrador de un lago de datos o como usuario al que se le hayan otorgado permisos de Lake Formation sobre los recursos del catálogo de datos LF-TBAC mediante la opción de concesión.

   En el panel de navegación, en Permisos, seleccione Permisos de lago de datos. A continuación, seleccione Conceder.

2. Especifique las entidades principales.

   En la sección Entidades principales, elija uno de los tipos y, a continuación, especifique las que van a recibir los permisos concedidos.



   Usuarios y roles de IAM

   Elija uno o varios usuarios o roles en la lista de usuarios y roles de IAM.

   IAM Identity Center

   Elija uno o varios usuarios o en la lista de Usuarios y grupos.

Usuarios y grupos de SAML

Para los usuarios y grupos de SAML y Quick, introduzca uno o más nombres de recursos de Amazon (ARN) para los usuarios o grupos federados a través de SAML, o ARN para los usuarios o grupos de Quick. Pulse Intro después de cada ARN.

Para obtener información sobre cómo crear ARN, consulte Concesión y revocación de Lake Formation AWS CLI Comandos de la de.

> **Note**
>
> La integración de Lake Formation con Quick solo es compatible con Quick Enterprise Edition.

Cuentas externas

Para Cuentas de AWS AWS la organización o el director de IAM, introduzca uno o más Cuenta de AWS ID, ID de organización, ID de unidad organizativa o ARN válidos para el usuario o rol de IAM. Pulse Intro después de cada ID.

El ID de una organización está formado por una "o-" seguida de 10 a 32 letras minúsculas o dígitos.

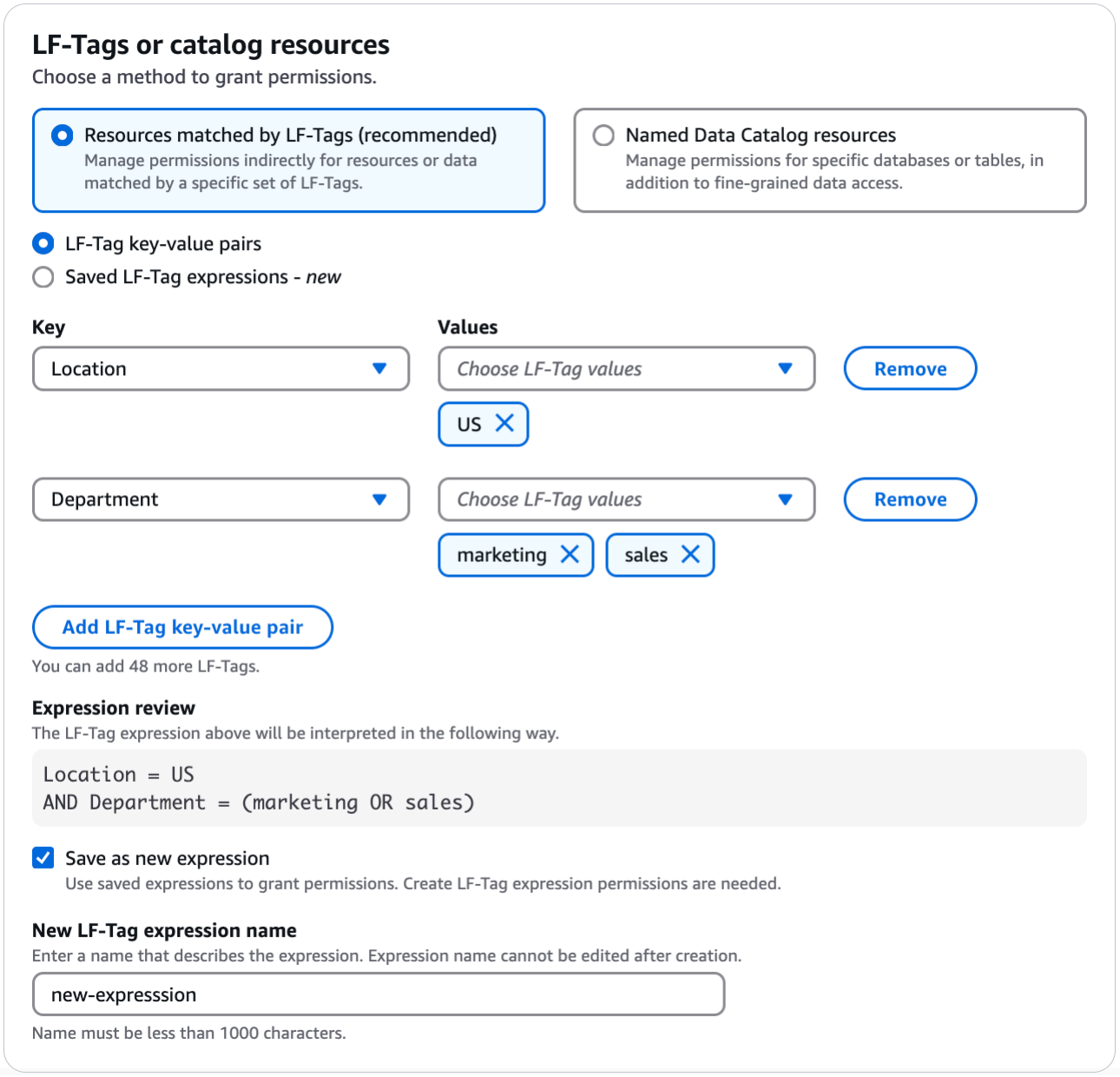
Un ID de una unidad organizativa comienza por «ou-» seguidos de 4 a 32 letras minúsculas o dígitos (el ID de la raíz que contiene la UO). Esta cadena va seguida de un segundo guion «-» y de 8 a 32 letras minúsculas o dígitos adicionales.

3. Especifique el. LF-Tags

Asegúrese de elegir los recursos coincidentes por LF-Tags opción. Elija pares LF-Tag clave-valor o expresiones guardadas LF-Tag .

1. Si elige la opción de pares LF-Tag clave-valor, elija las claves y los valores.

   Si elige más de un valor, está creando una LF-Tag expresión con un `OR` operador. Esto significa que si alguno de los LF-Tag valores coincide con un recurso LF-Tag asignado a un catálogo de datos, se le conceden permisos sobre el recurso.



2. (Opcional) Vuelva a seleccionar Añadir par LF-Tag clave-valor para especificar otro. LF-Tag

   Si especifica más de uno LF-Tag, está creando una LF-Tag expresión con un `AND` operador. Al principal se le conceden permisos sobre un recurso del catálogo de datos solo si al recurso se le ha asignado una coincidencia LF-Tag para cada uno de los LF-Tag elementos de la LF-Tag expresión.

3. Seleccione la opción Guardar como expresión nueva para volver a utilizar la expresión.

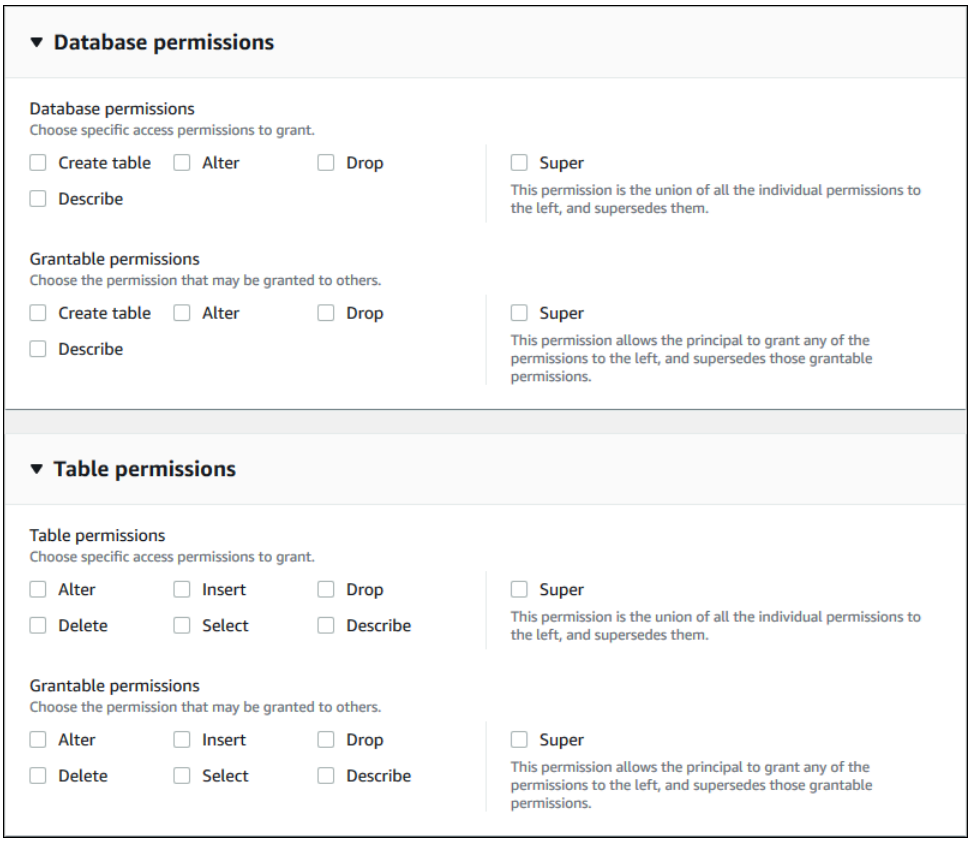
   Necesita `Create LF-Tag expression` para guardar expresiones.

Para obtener más información sobre LF-Tag las expresiones, consulte[Gestión de LF-Tag expresiones para el control de acceso a los metadatos](#).

4. Especifique los permisos.

   Especifique los permisos que se conceden a la entidad principal sobre los recursos coincidentes del catálogo de datos. Los recursos coincidentes son aquellos recursos que se asignaron y LF-Tags que coinciden con una de las LF-Tag expresiones otorgadas al principal.

   Puede especificar los permisos que se van a conceder sobre las bases de datos coincidentes, las tablas coincidentes y las vistas coincidentes.



   En Permisos de base de datos, seleccione los permisos de base de datos que se concederán a la entidad principal en las bases de datos coincidentes.

   En Permisos de tabla, seleccione los que se concederán a la entidad principal en las tablas y vistas coincidentes.

También puede elegir `Select`,`Describe` y permisos de `Drop` en Permisos de tabla y aplicarlos a las vistas.

5. Elija Conceder.

AWS CLI

Puede usar el método AWS Command Line Interface (AWS CLI) y el método de control de acceso basado en etiquetas (LF-TBAC) de Lake Formation para conceder permisos de Lake Formation en las bases de datos, tablas y columnas del catálogo de datos.

Otorgar permisos para lagos de datos mediante el AWS CLI y el LF-TBAC método

- Utilice el comando `grant-permissions`.

  Example

  El siguiente ejemplo otorga la LF-Tag expresión "`module=*`" (todos los valores de la LF-Tag clave`module`) al usuario`datalake_user1`. Ese usuario tendrá el `CREATE_TABLE` permiso para acceder a todas las bases de datos coincidentes, es decir, las bases de datos a las que se les haya asignado la LF-Tag clave`module`, con cualquier valor.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "CREATE_TABLE" --resource '{ "LFTagPolicy":
 {"CatalogId":"111122223333","ResourceType":"DATABASE","Expression":
[{"TagKey":"module","TagValues":["*"]}]}}'
```

  Example

  El siguiente ejemplo concede la LF-Tag expresión "`(level=director) AND (region=west OR region=south)`" al usuario. `datalake_user1` Ese usuario tendrá los permisos `SELECT`, `ALTER`, y `DROP` con la opción de concesión en las tablas que coincidan (tablas a las que se les haya asignado tanto `level=director` como (`region=west` o `region=south`).

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "SELECT" "ALTER" "DROP" --permissions-
with-grant-option "SELECT" "ALTER" "DROP" --resource '{ "LFTagPolicy":
```

```
 {"CatalogId":"111122223333","ResourceType":"TABLE","Expression": [{"TagKey":
 "level","TagValues": ["director"]},{"TagKey": "region","TagValues": ["west",
 "south"]}]}}'
```

Example

El siguiente ejemplo otorga la LF-Tag expresión `module=orders` "» a la AWS cuenta 1234-5678-9012. El administrador del lago de datos de esa cuenta puede entonces conceder la expresión "`module=orders`" a las entidades principales de su cuenta. Esos directores tendrán entonces `CREATE_TABLE` permiso para hacer coincidir las bases de datos propiedad de la cuenta 1111-2222-3333 y compartidas con la cuenta 1234-5678-9012 mediante el método de recurso designado o el método. LF-TBAC

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=123456789012 --permissions "CREATE_TABLE" --
permissions-with-grant-option "CREATE_TABLE" --resource '{ "LFTagPolicy":
 {"CatalogId":"111122223333","ResourceType":"DATABASE","Expression":
[{"TagKey":"module","TagValues":["orders"]}]}}'
```

Attribute-based control de acceso

En AWS Lake Formation, puede conceder acceso a AWS Glue Data Catalog objetos como catálogos, bases de datos, tablas y filtros de datos mediante atributos que son etiquetas de IAM y etiquetas de sesión asociadas a entidades de IAM, como roles y usuarios.

Para obtener más información sobre el uso de etiquetas de sesión, consulte assume-role en la guía del usuario. AWS CLI

Attribute-based el control de acceso (ABAC) es una estrategia de autorización que define los permisos en función de los atributos. AWS llama a estos atributos etiquetas. Puede utilizar ABAC para conceder acceso a las entidades principales de la misma cuenta o de otra cuenta sobre los recursos del Catálogo de datos. Cualquier entidad principal de IAM con claves y valores coincidentes de la etiqueta de IAM o de la etiqueta de sesión tendrá acceso al recurso. Debe tener permisos concesibles sobre los recursos para poder realizar estas concesiones.

ABAC le permite conceder acceso a varios usuarios al mismo tiempo. Cuando los nuevos usuarios se unen a la organización, su acceso a los datos se puede determinar automáticamente en función de sus atributos, como su puesto de trabajo o su departamento, sin necesidad de que los administradores asignen manualmente roles o permisos específicos. Al utilizar atributos en lugar

de roles, ABAC proporciona una forma de administrar el acceso a los datos en diversos sistemas y entornos más ágil y fácil de mantener, lo que, en última instancia, mejora la gobernanza y la conformidad de los datos.

Para obtener más información sobre cómo definir atributos, consulte Definición de permisos en función de los atributos con la autorización de ABAC.

Para obtener información sobre las limitaciones, las consideraciones y AWS las regiones compatibles, consulteConsideraciones, limitaciones y regiones compatibles para el control de acceso basado en atributos.

Temas

- Requisitos previos para conceder permisos mediante atributos
- Concesión de permisos mediante control de acceso basado en atributos

Requisitos previos para conceder permisos mediante atributos

Para conceder permisos mediante el control de acceso basado en atributos (ABAC), debe cumplir los requisitos previos siguientes:

- Actualice la configuración del Catálogo de datos para habilitar los permisos de Lake Formation para objetos del Catálogo de datos. Para obtener más información, consulte la sección Cambio del modelo de permisos predeterminado o uso del modo de acceso híbrido.
- Establezca la configuración de la versión entre cuentas en dos o más.
- Asocie atributos a las entidades de IAM que requieren acceso.
- Solo un administrador del lago de datos o un usuario de IAM con los permisos necesarios pueden conceder acceso a los objetos del Catálogo de datos. Para obtener más información sobre los permisos necesarios, consulte Permisos de IAM.

Concesión de permisos mediante control de acceso basado en atributos

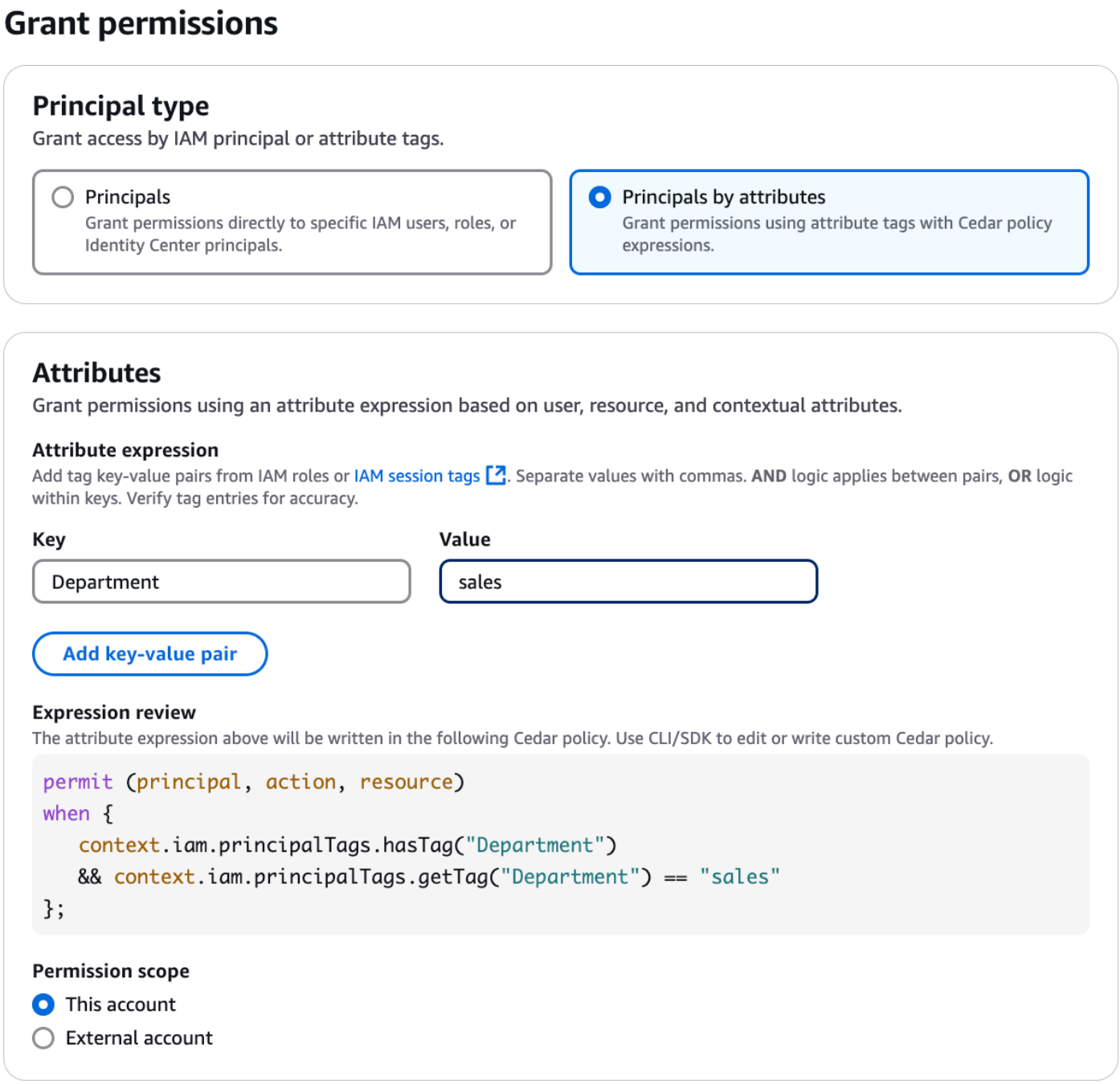
En este tema, se describen los pasos que debe seguir para conceder permisos de acceso basados en atributos a los recursos del Catálogo de datos. Puede utilizar la consola de Lake Formation o la interfaz de línea de AWS comandos (AWS CLI).

Concesión de permisos mediante ABAC (Consola de administración de AWS)

1. Abre la consola de Lake Formation en e inicia sesión como administrador del lago de datos, creador del recurso o usuario de IAM que tenga permisos concedibles sobre el recurso. https://console.aws.amazon.com/lakeformation/
2. Realice una de las siguientes acciones:
   - En el panel de navegación, en Permisos, seleccione Permisos de lago de datos. A continuación, seleccione Conceder.
   - En el panel de navegación, elija Catálogos en Catálogo de datos. A continuación, elija un objeto de catálogo (catálogos, bases de datos, tablas y filtros de datos) y, en el menú Acciones, en Permisos, elija Conceder.
3. En la página Concesión de permisos, elija Entidades principales por atributos.
4. Especifique la clave y el valor o los valores del atributo. Si elige más de un valor, está creando una expresión de atributo con un operador `OR`. Esto significa que si alguno de los valores de las etiquetas de atributo asignados a un rol o usuario de IAM coincide, role/user obtiene permisos de acceso al recurso.

   Si elige más de una etiqueta de atributo, está creando una expresión de atributo con un operador `AND`. Al principal se le conceden permisos sobre un recurso del catálogo de datos solo si al IAM role/user se le ha asignado una etiqueta coincidente para cada etiqueta de atributo de la expresión de atributo.

   Revise la expresión de política de Cedar resultante que se muestra en la consola.



5. Elija el alcance de los permisos. Si los beneficiarios pertenecen a una cuenta externa, elija Cuenta externa e introduzca el ID de la AWS cuenta.
6. A continuación, elija la cuenta del Catálogo de datos o cuentas externas. Debe tener los permisos concesibles correspondientes sobre los recursos para completar satisfactoriamente las concesiones de permisos.

7. Especifique qué acciones quiere permitir que realicen las entidades principales (usuarios o roles) que tienen atributos coincidentes. El acceso se concede a las entidades de IAM a las que se les hayan asignado etiquetas y valores que coincidan con al menos una de las expresiones de atributos especificadas. Revise la expresión de política de Cedar resultante en la consola. Para obtener más información acerca de Cedar, consulte ¿Qué es Cedar? | Referencia del lenguaje de políticas de Cedar. GuideLink
8. A continuación, elija los recursos del Catálogo de datos para conceder el acceso. Puede definir estos permisos para varios recursos del Catálogo de datos, incluidos catálogos, bases de datos, tablas y filtros de datos.
9. Elija Conceder.

   Este enfoque le permite controlar el acceso en función de los atributos, lo que garantiza que solo los usuarios o roles con las etiquetas adecuadas puedan realizar acciones específicas en los recursos designados.

Concesión de permisos mediante ABAC (AWS CLI)

El siguiente ejemplo muestra una expresión de atributo que debe cumplirse para recibir todos los permisos disponibles en el recurso. También puede especificar permisos individuales como `Select`, `Describe` o `Drop`. La expresión utiliza la expresión de política de Cedar. Para obtener más información acerca de Cedar, consulte ¿Qué es Cedar? | Referencia GuideLink del lenguaje de políticas de Cedar.

Esta condición comprueba si la entidad principal de IAM tiene una etiqueta `department` y el valor de la etiqueta `department` es igual a `sales`.

```
aws lakeformation grant-permissions
--principal '{"DataLakePrincipalIdentifier": "111122223333:IAMPrincipals"}' \
--resource '{"Database": {"CatalogId": 111122223333, "Name": "abac-db"}}' \
--permissions ALL \
--condition '{"Expression": "context.iam.principalTags.hasTag(\"department\") \
   && context.iam.principalTags.getTag(\"department\") == \"sales\""}'
```

Ejemplo de escenario de permisos

El siguiente escenario ayuda a demostrar cómo se pueden configurar los permisos para proteger el acceso a los datos AWS Lake Formation.

Shirley es administradora de datos. Quiere configurar un lago de datos para su empresa, AnyCompany. Ahora mismo, todos los datos se almacenan en Amazon S3. John es responsable de marketing y necesita acceso de escritura a la información de compra de los clientes (contenida en `s3://customerPurchases`). Un analista de marketing, Diego, se une a John este verano. John necesita poder conceder a Diego acceso para consultar los datos sin involucrar a Shirley.

Mateo, de finanzas, necesita acceso para consultar datos contables (por ejemplo, `s3://transactions`). Quiere consultar los datos de las transacciones en las tablas de una base de datos (`Finance_DB`) utilizadas por el equipo financiero. Su jefe, Arnav, puede darle acceso a la `Finance_DB`. Aunque no debería poder modificar los datos contables, necesita la capacidad de convertir los datos a un formato (esquema) adecuado para la previsión. Estos datos se almacenarán en un bucket separado (`s3://financeForecasts`) que él podrá modificar.

Para resumir:

- Shirley es la administradora del lago de datos.
- John necesita permiso `CREATE_DATABASE` y `CREATE_TABLE` para crear nuevas bases de datos y tablas en el Catálogo de datos.
- John también necesita los permisos `SELECT`, `INSERT` y `DELETE` en las tablas que crea.
- Diego necesita permiso `SELECT` sobre la tabla para ejecutar consultas.

Los empleados de AnyCompany realizan las siguientes acciones para configurar los permisos. Las operaciones de la API mostradas en este escenario muestran una sintaxis simplificada para mayor claridad.

1. Shirley registra en Lake Formation la ruta de Amazon S3 que contiene la información de compra del cliente.

   ```
   RegisterResource(ResourcePath("s3://customerPurchases"), false, Role_ARN )
   ```

2. Shirley concede a John acceso a la ruta de Amazon S3 que contiene la información de compra del cliente.

   ```
   GrantPermissions(John, S3Location("s3://customerPurchases"),
    [DATA_LOCATION_ACCESS]) )
   ```

3. Shirley concede a John permiso para crear bases de datos.

```
GrantPermissions(John, catalog, [CREATE_DATABASE])
```

4. John crea la base de datos `John_DB`. John tiene automáticamente permiso `CREATE_TABLE` sobre esa base de datos porque él la creó.

```
CreateDatabase(John_DB)
```

5. John crea la tabla `John_Table` que apunta a `s3://customerPurchases`. Como él creó la tabla, tiene todos los permisos sobre ella y puede conceder permisos sobre ella.

```
CreateTable(John_DB, John_Table)
```

6. Juan permite a su analista, Diego, acceder a la tabla `John_Table`.

```
 GrantPermissions(Diego, John_Table, [SELECT])
```

7. Juan permite a su analista, Diego, el acceso a `s3://customerPurchases/London/`. Porque Shirley ya registró `s3://customerPurchases`, sus subcarpetas están registradas con Lake Formation.

```
GrantDataLakePrivileges( 123456789012/datalake, Diego, [DATA_LOCATION_ACCESS], [],
S3Location("s3://customerPurchases/London/") )
```

8. John permite a su analista, Diego, crear tablas en la base de datos `John_DB`.

```
GrantDataLakePrivileges( 123456789012/datalake, Diego, John_DB, [CREATE_TABLE],
[] )
```

9. Diego crea una tabla en `John_DB` en `s3://customerPurchases/London/` y obtiene automáticamente permisos `ALTER`, `DROP`, `SELECT`, `INSERT` y `DELETE`.

```
CreateTable( 123456789012/datalake, John_DB, Diego_Table )
```

Filtrado de datos y seguridad de celda en Lake Formation

Al conceder permisos de Lake Formation en una tabla del Catálogo de datos, puede incluir especificaciones de filtrado de datos para restringir el acceso a determinados datos en los resultados de las consultas y en los motores integrados con Lake Formation. Lake Formation utiliza el filtrado

de datos para ofrecer una seguridad a nivel de columna, fila y celda. Puede definir y aplicar filtros de datos en las columnas anidadas si los datos de origen contienen estructuras anidadas.

Con las capacidades de filtrado de datos de Lake Formation, puede implementar los siguientes niveles de seguridad de datos.

Column-level seguridad

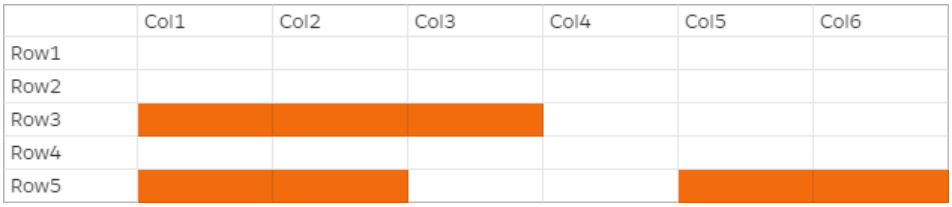
La concesión de permisos en una tabla del catálogo de datos con seguridad a nivel de columna (filtrado de columnas) permite a los usuarios ver solo las columnas específicas y columnas anidadas a las que tienen acceso en la tabla. Considere una tabla `persons` que se utiliza en varias aplicaciones para una gran empresa de comunicaciones multirregional. La concesión de permisos en las tablas del Catálogo de datos con filtrado de columnas puede impedir que los usuarios que no trabajan en el departamento de Recursos Humanos vean información de identificación personal (PII), por ejemplo, un número de seguro social o fecha de nacimiento. También puede definir políticas de seguridad y conceder acceso solo a subestructuras parciales de las columnas anidadas.

Row-level seguridad

La concesión de permisos en una tabla del Catálogo de datos con seguridad a nivel de fila (filtrado de filas) permite a los usuarios ver solo las filas específicas de datos a las que tienen acceso en la tabla. El filtrado se basa en los valores de una o más columnas. Puede incluir estructuras de columnas anidadas al definir las expresiones de filtro de filas. Por ejemplo, si las diferentes oficinas regionales de la empresa de comunicaciones tienen sus propios departamentos de Recursos Humanos, puede limitar los registros de personas que los empleados de RH pueden ver a solo los registros de los empleados de su región.

Cell-level seguridad

Cell-level la seguridad combina el filtrado de filas y el filtrado de columnas para lograr un modelo de permisos muy flexible. Si ve las filas y columnas de una tabla como una cuadrícula, mediante el uso de la seguridad a nivel de celda, puede restringir el acceso a los elementos individuales (celdas) de la cuadrícula en cualquier parte de las dos dimensiones. Es decir, puede restringir el acceso a diferentes columnas en función de la fila. Esto se ilustra en el siguiente diagrama, en el que las columnas restringidas aparecen sombreadas.



Siguiendo con el ejemplo de la tabla de personas, puede crear un filtro de datos a nivel de celda que restrinja el acceso a la columna de direcciones si la fila tiene la columna de país establecida en «Reino Unido», pero que permita el acceso a la columna de direcciones si la fila tiene la columna de país como «EE. UU.».

Los filtros se aplican solo a las operaciones de lectura. Por lo tanto, solo puede conceder el permiso de Lake Formation `SELECT` con filtros.

Cell-level seguridad en columnas anidadas

Lake Formation le permite definir y aplicar filtros de datos con seguridad a nivel de celda en columnas anidadas. Sin embargo, los motores analíticos integrados, como Amazon Athena, Amazon EMR y Amazon Redshift Spectrum, permiten ejecutar consultas en tablas anidadas administradas de Lake Formation con seguridad a nivel de filas y columnas.

Para conocer las limitaciones, consulte Limitaciones de filtrado de datos.

Temas

- Filtros de datos en Lake Formation
- Compatibilidad con PartiQL en expresiones de filtro de filas
- Permisos necesarios para consultar tablas con filtrado a nivel de celda
- Administrar filtros de datos

Filtros de datos en Lake Formation

Puede implementar seguridad a nivel de columna, de fila y de celda mediante el uso de filtros de datos. Selecciona un filtro de datos al conceder el permiso de Lake Formation de `SELECT` sobre las tablas. Si la tabla contiene estructuras de columnas anidadas, puede definir un filtro de datos incluyendo o excluyendo las columnas secundarias y definir expresiones de filtro a nivel de fila en los atributos anidados.

Cada filtro de celdas de datos pertenece a una tabla específica de un Catálogo de datos. Un filtro de datos incluye la siguiente información:

- Nombre del filtro
- Los ID del catálogo de la tabla asociada al filtro
- Nombre de la tabla

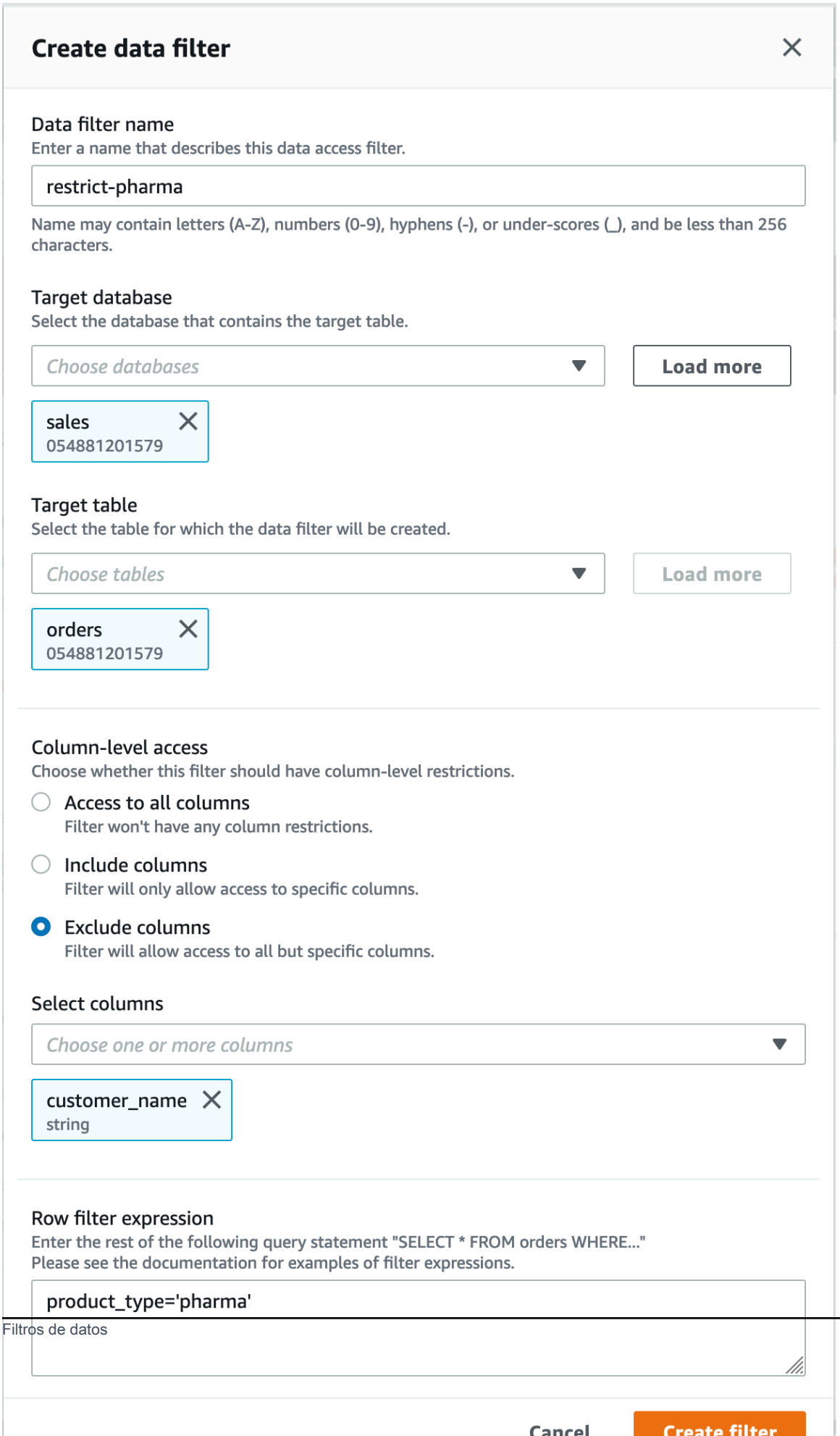
- Nombre de la base de datos que contiene las tablas que se enumerarán.
- Especificación de columna: una lista de columnas y columnas anidadas (con tipos de datos `struct`) para incluir o excluir en los resultados de la consulta.
- Expresión de filtro de filas: una expresión que especifica las filas que se van a incluir en los resultados de la consulta. Con algunas restricciones, la expresión tiene la sintaxis de una cláusula `WHERE` en el lenguaje PartiQL. Para especificar todas las filas, selecciona Acceso a todas las filas en Row-level Acceso en la consola o Uso `AllRowsWildcard` en las llamadas a la API.

  Para obtener más información acerca de lo que se admite en las expresiones de filtro de fila, consulte Compatibilidad con PartiQL en expresiones de filtro de filas.

El nivel de filtrado que recibe depende de cómo rellene el filtro de datos.

- Al especificar el comodín “todas las columnas” y proporcionar una expresión de filtro de filas, solo se establece la seguridad de fila (filtrado de filas).
- Al incluir o excluir columnas específicas y columnas anidadas y especificar «todas las filas» con el comodín de todas las filas, solo establece la seguridad a nivel de columna (filtrado de columnas).
- Cuando se incluyen o excluyen columnas específicas y se proporciona también una expresión de filtro de fila, se establece la seguridad a nivel de celda (filtrado de celdas).

La siguiente captura de pantalla de la consola de Lake Formation muestra un filtro de datos que actúa a nivel de celda. En el caso de las consultas basadas en la tabla `orders`, restringe el acceso a la columna `customer_name` y los resultados de la consulta solo muestran las filas en las que la columna `product_type` contiene «farmacia».


Create data filter
Data filter name
Enter a name that describes this data access filter.
restrict-pharma
Name may contain letters (A-Z), numbers (0-9), hyphens (-), or under-scores (_), and be less than 256 characters.
Target database
Select the database that contains the target table.
Choose databases
Load more
sales
054881201579
Target table
Select the table for which the data filter will be created.
Choose tables
Load more
orders
054881201579
Column-level access
Choose whether this filter should have column-level restrictions.
Access to all columns
Filter won't have any column restrictions.
Include columns
Filter will only allow access to specific columns.
Exclude columns
Filter will allow access to all but specific columns.
Select columns
Choose one or more columns
customer_name
string
Row filter expression
Enter the rest of the following query statement "SELECT * FROM orders WHERE..."
Please see the documentation for examples of filter expressions.
product_type='pharma'
Cancel
Create filter

Tenga en cuenta el uso de comillas simples para encerrar la cadena literal `'pharma'`.

Puede usar la consola de Lake Formation para crear este filtro de datos o proporcionar el siguiente objeto de solicitud a la operación de la API `CreateDataCellsFilter`.

```
{
     "Name": "restrict-pharma",
     "DatabaseName": "sales",
     "TableName": "orders",
     "TableCatalogId": "111122223333",
     "RowFilter": {"FilterExpression": "product_type='pharma'"},
     "ColumnWildcard": {
         "ExcludedColumnNames": ["customer_name"]
     }
}
```

Puede crear todos los filtros de datos que necesite para una tabla. Para ello, necesita un permiso `SELECT` con la opción de concesión que aparece en una tabla. De forma predeterminada, los administradores del lago de datos tienen permiso para crear filtros de datos en todas las tablas de esa cuenta. Normalmente solo se utiliza un subconjunto de los posibles filtros de datos cuando se conceden permisos sobre la tabla a una entidad principal. Por ejemplo, puede crear un segundo filtro de datos para la tabla `orders` que sea un filtro de datos solo con seguridad de filas. En referencia a la captura de pantalla anterior, puede elegir la opción Acceso a todas las columnas e incluir una expresión de filtro de filas de `product_type<>pharma`. El nombre de este filtro de datos podría ser `no-pharma`. Restringe el acceso a todas las filas cuya `product_type` columna esté configurada como «pharma».

El objeto de solicitud de la operación de la API `CreateDataCellsFilter` para este filtro de datos es el siguiente.

```
{
     "Name": "no-pharma",
     "DatabaseName": "sales",
     "TableName": "orders",
     "TableCatalogId": "111122223333",
     "RowFilter": {"FilterExpression": "product_type<>'pharma'"},
     "ColumnNames": ["customer_id", "customer_name", "order_num"
          "product_id", "purchase_date", "product_type",
          "product_manufacturer", "quantity", "price"]
}
```

A continuación, puede conceder SELECT sobre la tabla `restrict-pharma` con el filtro de datos `orders` a un usuario administrativo y SELECT sobre la tabla `orders` con el filtro de datos `no-pharma` a los usuarios no administrativos. En el caso de los usuarios del sector sanitario, concedería SELECT en la tabla `orders` con acceso total a todas las filas y columnas (sin filtro de datos), o quizá con otro filtro de datos que restrinja el acceso a la información sobre precios.

Puede incluir o excluir columnas anidadas al especificar la seguridad a nivel de columna y de fila dentro de un filtro de datos. En el siguiente ejemplo, el acceso al campo `product.offer` se especifica mediante nombres de columna cualificados (entre comillas dobles). Esto es importante en el caso de los campos anidados para evitar que se produzcan errores cuando los nombres de las columnas contienen caracteres especiales y para mantener la compatibilidad con versiones anteriores de las definiciones de seguridad de nivel columnas superior.

```
{
     "Name": "example_dcf",
     "DatabaseName": "example_db",
     "TableName": "example_table",
     "TableCatalogId": "111122223333",
     "RowFilter": { "FilterExpression": "customer.customerName <> 'John'" },
     "ColumnNames": ["customer", "\"product\".\"offer\""]
}
```

Véase también

- Administrar filtros de datos

Compatibilidad con PartiQL en expresiones de filtro de filas

Puede crear expresiones de filtro de filas mediante un subconjunto de tipos de datos PartiQL, operadores y agregaciones. Lake Formation no permite ninguna función PartiQL estándar o definida por el usuario en la expresión del filtro. Puede usar operadores de comparación para comparar columnas con constantes (por ejemplo, `views >= 10000`), pero no puede comparar columnas con otras columnas.

Una expresión de filtro de filas puede ser una expresión simple o una expresión compuesta. La longitud total de la expresión debe ser inferior a 2048 caracteres.

Expresiones simples

Una expresión simple tendrá el siguiente formato: `<column name > <comparison operator ><value >`

- Nombre de la columna

  Puede ser una columna de datos de nivel superior, una columna de partición o una columna anidada presente en el esquema de la tabla y debe pertenecer a los Tipos de datos compatibles indicados a continuación.

- Operador de comparación

  Se admiten los siguientes operadores: `=, >, <, >=, <=, <>,!=, BETWEEN, IN, LIKE, NOT, IS [NOT] NULL`

- Todas las comparaciones de cadenas y coincidencias del patrón `LIKE` distinguen entre mayúsculas y minúsculas. No puede usar el operador IS [NOT] NULL en las columnas de partición.
- Valor de columna

  El valor de la columna debe coincidir con el tipo de datos del nombre de la columna.

Expresión compuesta

Una expresión simple tendrá el formato: `( <simple expression >) <AND/OR >(<simple expression >)`. Las expresiones compuestas se pueden combinar aún más mediante operadores lógicos `AND/OR`.

Tipos de datos compatibles

Los filtros de filas que hacen referencia a una AWS Glue Data Catalog tabla que contiene un tipo de datos no compatible generarán un error. Los siguientes son los tipos de datos admitidos para las columnas y constantes de la tabla, que se asignan a Amazon Redshift los tipos de datos:

- `STRING, CHAR, VARCHAR`
- `INT, LONG, BIGINT, FLOAT, DECIMAL, DOUBLE`
- `BOOLEAN`
- `STRUCT`

Para obtener más información acerca de los tipos de datos en Amazon Redshift, consulte Tipos de datos en la Guía para desarrolladores de bases de datos de Amazon Redshift.

Expresión de filtro de filas

Example

Los siguientes son ejemplos de expresiones de filtro de filas válidas para una tabla con columnas: `country (String), id (Long), year (partition column of type Integer), month (partition column of type Integer)`

- `year > 2010 and country != 'US'`
- `(year > 2010 and country = 'US') or (month < 8 and id > 23)`
- `(country between 'Z' and 'U') and (year = 2018)`
- `(country like '%ited%') and (year > 2000)`

Example

Los siguientes son ejemplos válidos de expresiones de filtro de filas para una tabla con columnas anidadas: `year > 2010 and customer.customerId <> 1`

No se debe hacer referencia a los campos anidados de las columnas de partición al definir expresiones anidadas a nivel de fila.

Las constantes de cadena se deben escribir entre comillas simples.

Palabras clave reservadas

Si la expresión de filtro de filas contiene palabras clave PartiQL, se producirá un error de análisis, ya que los nombres de las columnas pueden entrar en conflicto con las palabras clave. Cuando esto suceda, escape de los nombres de las columnas usando comillas dobles. Algunos ejemplos de palabras clave reservadas son «first», «last», «asc» y «missing». Consulte la especificación PartiQL para obtener una lista de palabras clave reservadas.

Referencia PartiQL

Para obtener más información acerca de PartiQL, consulte https://partiql.org/.

Permisos necesarios para consultar tablas con filtrado a nivel de celda

Los siguientes permisos AWS Identity and Access Management (IAM) son necesarios para ejecutar consultas en tablas con filtrado a nivel de celda.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:StartQueryPlanning",
                "lakeformation:GetQueryState",
                "lakeformation:GetWorkUnits",
                "lakeformation:GetWorkUnitResults"
            ],
            "Resource": "*"
        }
    ]
}
```

Para obtener más información sobre permisos de Lake Formation, consulte Personas de Lake Formation y referencia de permisos IAM.

Administrar filtros de datos

Con el fin de implementar la seguridad a nivel de columna, fila y celda, puede crear y mantener filtros de datos. Cada filtro de datos pertenece a una tabla del Catálogo de datos. Puede crear varios filtros de datos para una tabla y, a continuación, utilizar uno o varios de ellos al conceder permisos sobre la tabla. También puede definir y aplicar filtros de datos en las columnas anidadas que tienen tipos de datos `struct` y que permiten a los usuarios acceder solo a las subestructuras de las columnas anidadas.

Se requiere un permiso `SELECT` con la opción de concesión para crear o ver un filtro de datos. Para que las entidades principales de su cuenta puedan ver y utilizar un filtro de datos, puede conceder el permiso `DESCRIBE` sobre este.

> **Note**
>
> Lake Formation no es compatible con la concesión de permisos `Describe` en un filtro de datos que se comparte desde otra cuenta.

Puede administrar los filtros de datos mediante la AWS Lake Formation consola, la API o AWS Command Line Interface (AWS CLI).

Para obtener más información acerca de los filtros de datos, consulte Filtros de datos en Lake Formation

Creación de un filtro de datos

Puede crear uno o varios filtros de datos para cada tabla del Catálogo de datos.

Para crear un filtro de datos para una tabla del Catálogo de datos (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador del lago de datos, como propietario de la tabla de destino o como entidad principal con permiso de Lake Formation en la tabla de destino.
2. En el panel de navegación, en Catálogo de datos, seleccione Filtros de datos.
3. En la página Filtros de datos, seleccione Crear filtro nuevo.
4. En el cuadro de diálogo Crear filtro de datos, introduzca la información siguiente:

   - Nombre del filtro de datos
   - Base de datos de destino. Indique la base de datos que contiene la tabla.
   - Tabla de destino
   - Column-level acceso: deje esta opción en Acceso a todas las columnas para especificar únicamente el filtrado de filas. Elija Incluir columnas o Excluir columnas para especificar el filtrado de columnas o celdas y, a continuación, especifique las columnas que desea incluir o excluir.

     Columnas anidadas: si aplica el filtro a una tabla que contiene columnas anidadas, puede especificar explícitamente las subestructuras de las columnas de estructura anidada dentro de un filtro de datos.

Al conceder el permiso SELECT a una entidad principal en este filtro, la entidad principal que ejecute la siguiente consulta solo verá los datos de `customer.customerName` y no de `customer.customerId`.

```
SELECT "customer" FROM "example_db"."example_table";
```

Column-level access

Choose whether this filter should have column-level restrictions.

Column-level access

Choose whether this filter should have column-level restrictions.

- ○ Access to all columns
  Filter won't have any column restrictions.
- ◉ Include columns
  Filter will only allow access to specific columns.
- ○ Exclude columns
  Filter will allow access to all but specific columns.

Included columns (4/11)

Choose the columns for column-level access

Find column

1

Name	Type
customer	struct
customerId	string
customerName	string
customerapplication	struct
appId	string
product	struct
offer	struct
listingId	string
prodId	string
type	string
purchaseid	string

Row-level access

Choose whether this filter should have row-level restrictions.

- ○ Access to all rows
- ◉ Filter rows

Row filter expression

Enter the rest of the following query statement SELECT * FROM nested-table WHERE...

Please see the documentation for examples of filter expressions.

```
customer.customerName <> 'John'
```

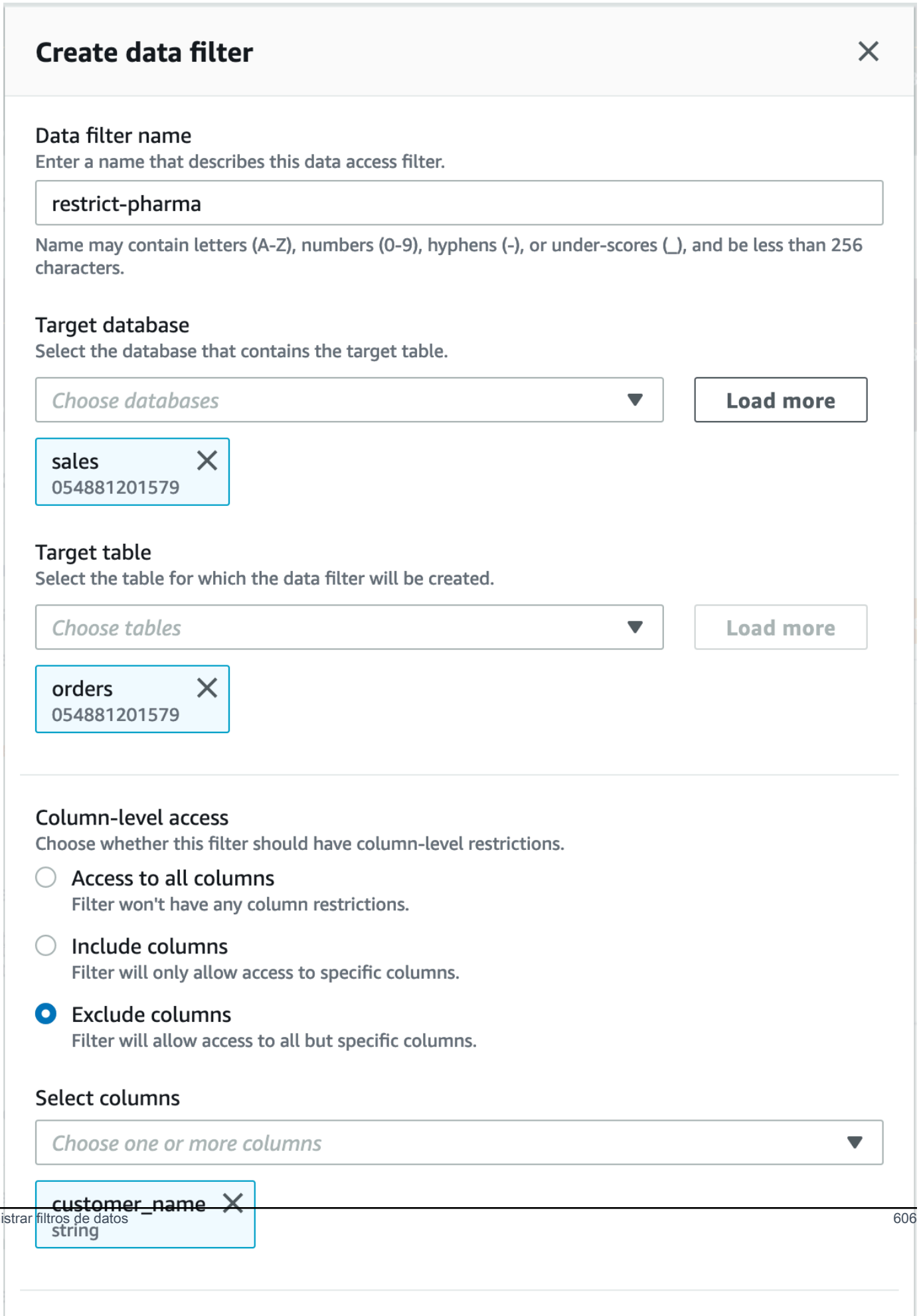
Al conceder permisos a la columna `customer`, la entidad principal recibe el acceso a la columna y a los campos anidados debajo de la columna (`customerName` y `customerID`).

- Expresión de filtro de filas. Introduzca una expresión de filtro para especificar el filtro de filas o celdas. Para conocer los tipos de datos y operadores compatibles, consulte Compatibilidad con PartiQL en expresiones de filtro de filas. Elija Acceso a todas las filas para conceder acceso a todas.

  Puede incluir estructuras de columnas parciales de columnas anidadas en una expresión de filtro de filas para filtrar las filas que contienen un valor concreto.

  Cuando a un director se le conceden permisos para acceder a una tabla con una expresión `Select * from example_nestedtable where customer.customerName <>'John'` de filtro de filas y el Column-levelacceso se establece en Acceso a todas las columnas, los resultados de la consulta solo muestran las filas en las que `customerName <>'John'` el resultado es verdadero.

La siguiente captura de pantalla muestra un filtro de datos que implementa el filtrado de celdas. En las consultas a la tabla `orders`, deniega el acceso a la columna `customer_name` y muestra solo las filas que tienen "pharma" en la columna `product_type`.


Create data filter
Data filter name
Enter a name that describes this data access filter.
restrict-pharma
Name may contain letters (A-Z), numbers (0-9), hyphens (-), or under-scores (_), and be less than 256 characters.
Target database
Select the database that contains the target table.
Choose databases
Load more
sales
054881201579
Target table
Select the table for which the data filter will be created.
Choose tables
Load more
orders
054881201579
Column-level access
Choose whether this filter should have column-level restrictions.
Access to all columns
Filter won't have any column restrictions.
Include columns
Filter will only allow access to specific columns.
Exclude columns
Filter will allow access to all but specific columns.
Select columns
Choose one or more columns
customer_name
string

5. Elija Crear filtro.

Para crear un filtro de datos con políticas de filtrado de celdas en un campo anidado

En esta sección se utiliza el siguiente esquema de ejemplo para mostrar cómo crear un filtro de celdas de datos:

```
[
    { name: "customer", type: "struct<customerId:string,customerName:string>" },
    { name: "customerApplication", type: "struct<appId:string>" },
    { name: "product", type:
 "struct<offer:struct<prodId:string,listingId:string>,type:string>" },
    { name: "purchaseId", type: "string" },
]
```

1. En la página Crear un filtro de datos, escriba un nombre para el filtro.
2. A continuación, utilice el menú desplegable para elegir un nombre de base de datos y un nombre de tabla.
3. En la sección de Column-level acceso, elija Columnas incluidas y seleccione una columna anidada ()`customer.customerName`.
4. En la sección de Row-level acceso, elija la opción Acceso a todas las filas.
5. Elija Crear filtro.

   Al conceder el permiso `SELECT` en este filtro, la entidad principal tiene acceso a todas las filas de la columna `customerName`.
6. A continuación, defina otro filtro de datos para el mismo database/table.
7. En la sección de Column-level acceso, elija Columnas incluidas y seleccione otra columna anidada (`customer.customerid`).
8. En la sección de Row-level acceso, elija Filtrar filas e introduzca una expresión de filtro de filas (`customer.customerid <> 5`).
9. Elija Crear filtro.

   Al conceder el permiso `SELECT` en este filtro, la entidad principal recibe acceso a todas las filas de los campos `customerName` y `customerId`, excepto a la celda en la que el valor es 5 en la columna `customerId`.

Concesión de permisos de filtrado de datos

Puede conceder los permisos `SELECT`, `DESCRIBE` y `DROP` de Lake Formation sobre los filtros de datos a las entidades principales.

Al principio, solo el usuario puede ver los filtros de datos que crea para una tabla. Para que otra entidad principal pueda ver un filtro de datos y conceder permisos del Catálogo de datos con el filtro de datos, debe:

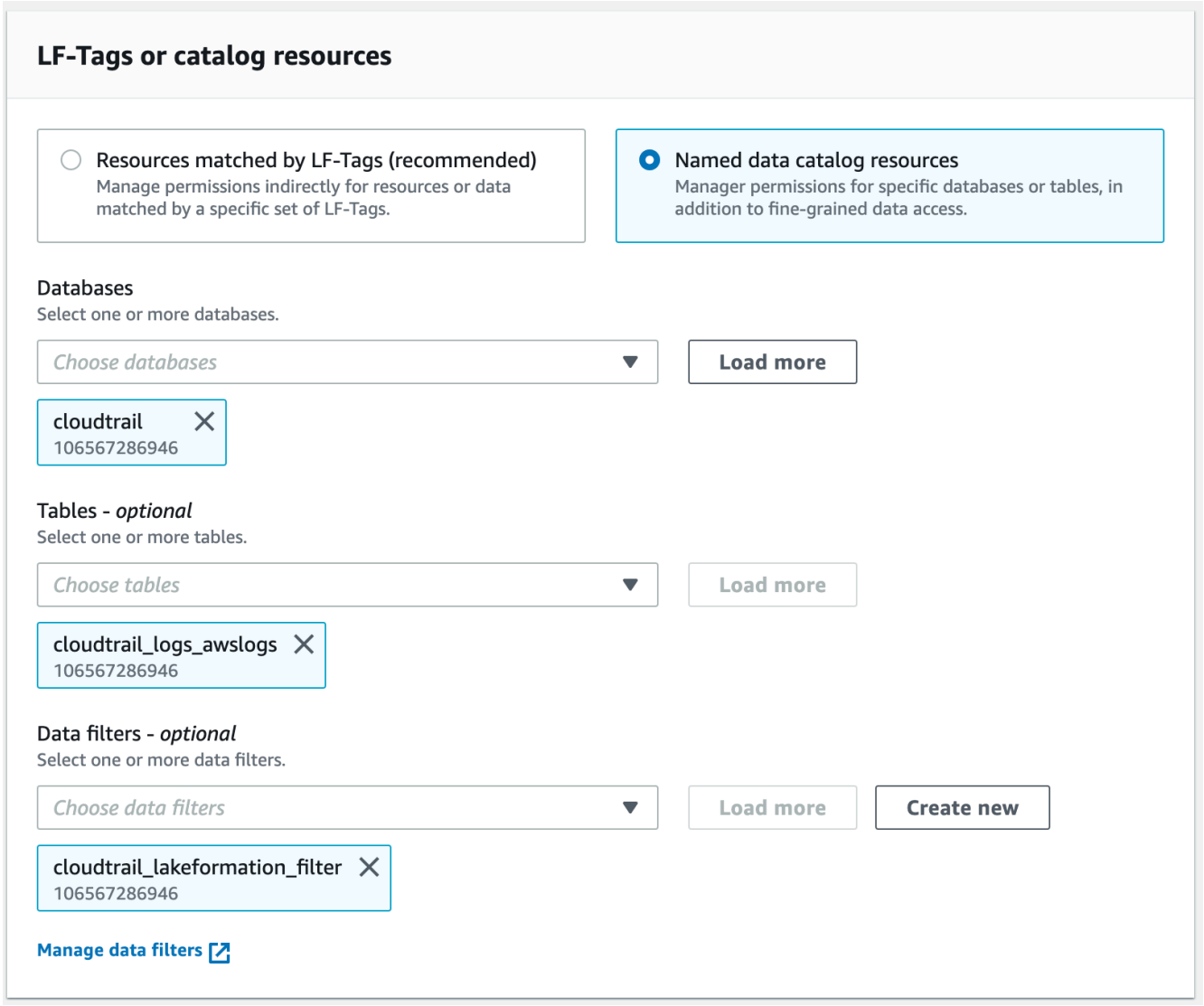
- Conceder `SELECT` sobre una tabla a la entidad principal con la opción de concesión, y aplicar el filtro de datos a la concesión.
- Conceda el permiso `DESCRIBE` o `DROP` sobre el filtro de datos a la entidad principal.

Puede conceder el `SELECT` permiso a una AWS cuenta externa. Un administrador del lago de datos de esa cuenta puede entonces conceder ese permiso a otras entidades principales de la cuenta. Al conceder el permiso a una cuenta externa, debes incluir la opción de concesión para que el administrador de la cuenta externa pueda extender el permiso en cascada a otros usuarios de la his/her cuenta. Al conceder a una entidad principal de su cuenta, la concesión con la opción de concesión es opcional.

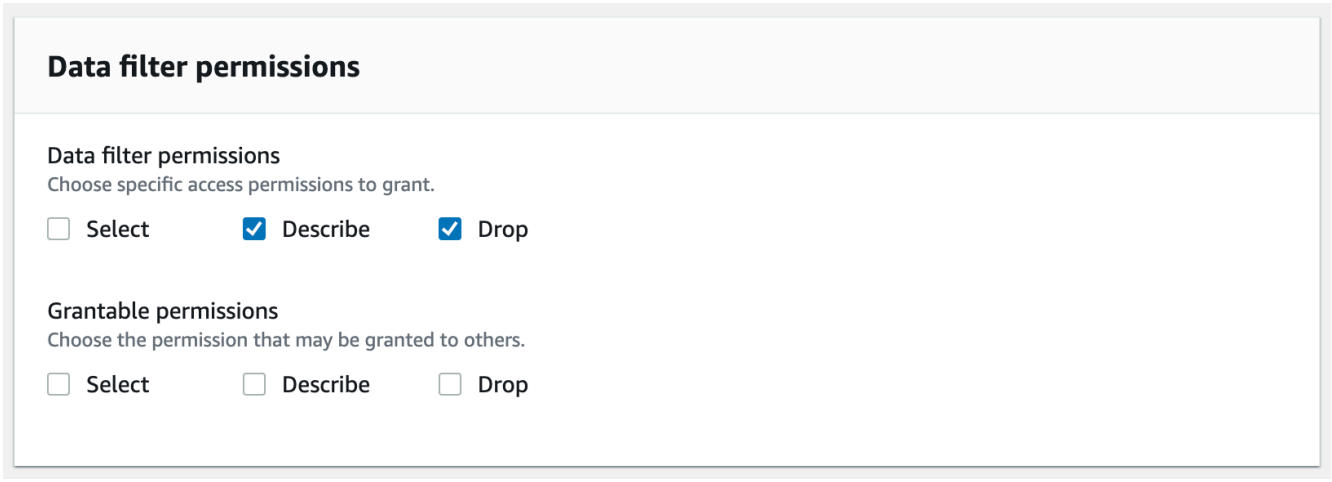
Puedes conceder y revocar permisos sobre los filtros de datos mediante la AWS Lake Formation consola, la API o el AWS Command Line Interface ()AWS CLI.

Console

1. Inicie sesión en la consola de Lake Formation Consola de administración de AWS y ábrala en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación, en Permisos, seleccione Permisos de lago de datos.
3. En la página Permisos, en la sección Permisos de datos, seleccione Conceder.
4. En la página Conceder permisos de datos, seleccione las entidades principales a las que conceder los permisos.
5. En la sección LF-Tags o recursos del catálogo, elija Recursos del catálogo de datos con nombre. A continuación, elija la base de datos, la tabla y el filtro de datos para los que desea conceder permisos.



6. En la sección Permisos del filtro de datos, seleccione los que desea conceder a las entidades principales seleccionadas.



AWS CLI

- Introduzca un comando `grant-permissions`. Especifique `DataCellsFilter` para el argumento de `resource` y `DESCRIBE` o `DROP` para el argumento de `Permissions` y, opcionalmente, para el argumento de `PermissionsWithGrantOption`.

  En el siguiente ejemplo, se concede `DESCRIBE` con la opción de concesión al usuario `datalake_user1` del filtro de datos `restrict-pharma`, que pertenece a la tabla `orders` de la base de datos `sales` de la cuenta AWS 1111-2222-3333.

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

  A continuación se muestra el contenido del archivo `grant-params.json`.

```
{
    "Principal": {"DataLakePrincipalIdentifier":
 "arn:aws:iam::111122223333:user/datalake_user1"},
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
    "Permissions": ["DESCRIBE"],
    "PermissionsWithGrantOption": ["DESCRIBE"]
}
```

Concesión de permisos de datos proporcionados por filtros de datos

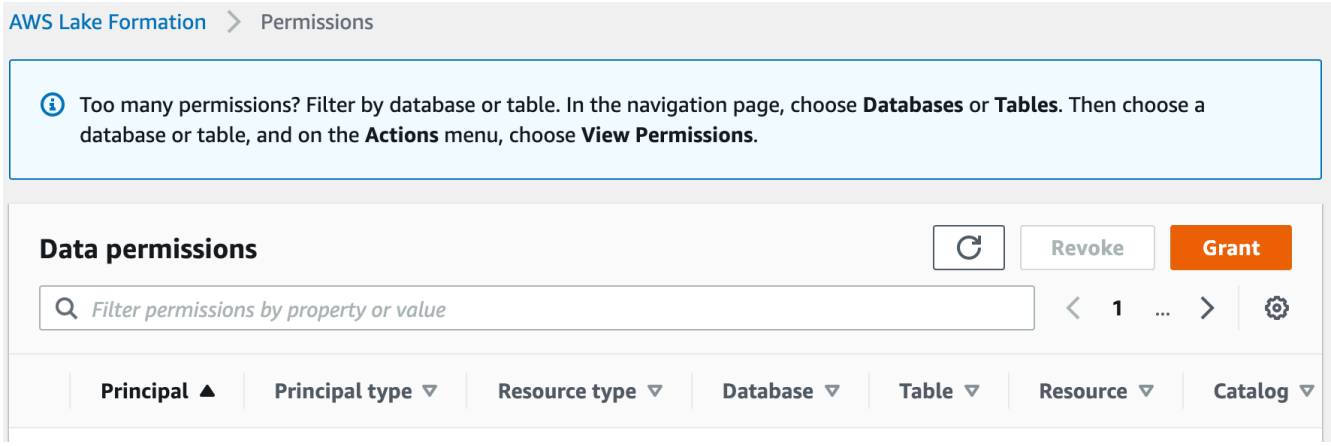
Los filtros de datos representan un subconjunto de datos dentro de una tabla. Para proporcionar acceso a los datos a las entidades principales, es necesario concederles permisos `SELECT`. Con este permiso las entidades principales pueden:

- Ver el nombre real de la tabla en la lista de tablas compartidas con su cuenta.
- Cree filtros de datos en la tabla compartida y conceda permisos a sus usuarios sobre esos filtros de datos.

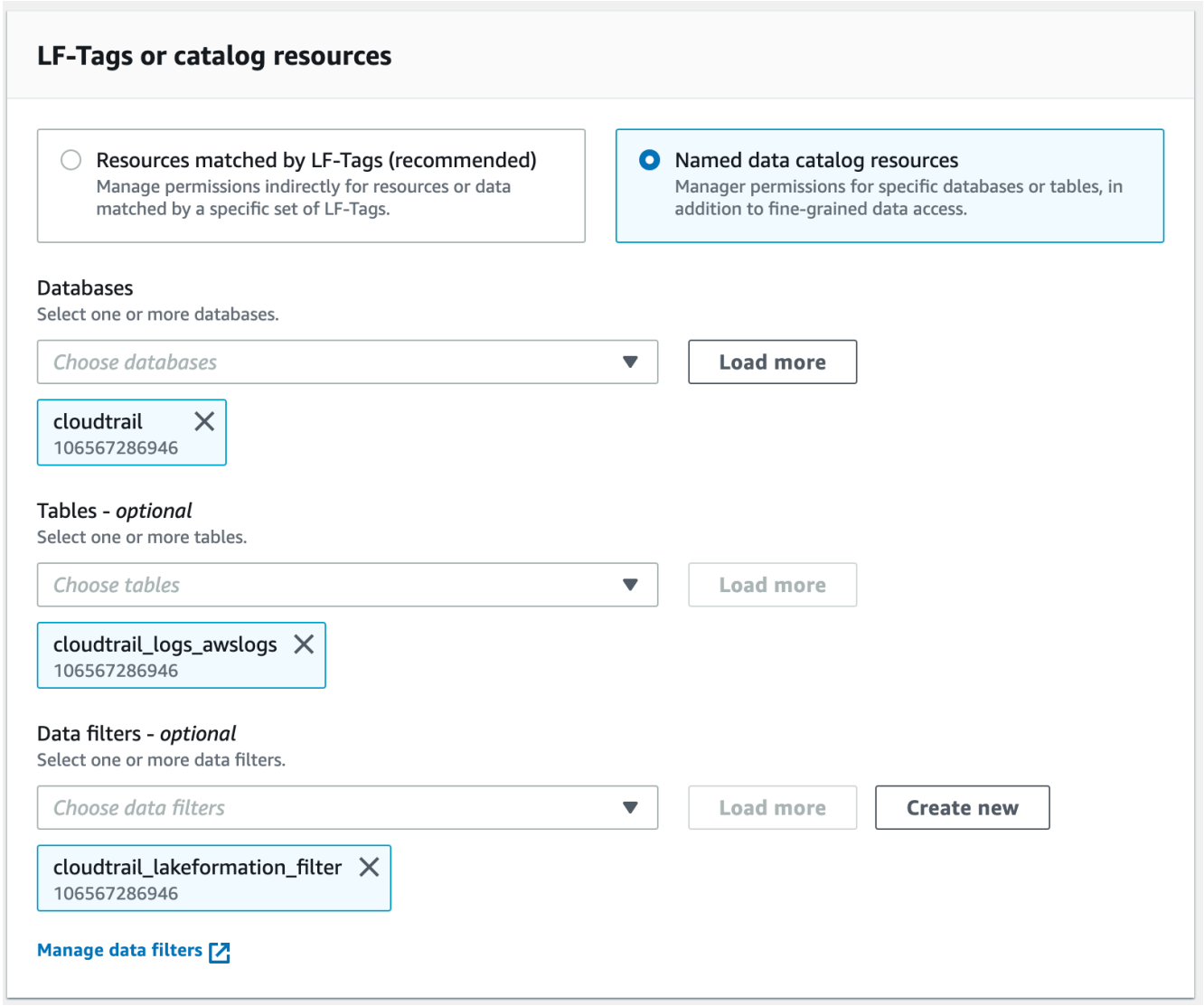
Console

Para conceder permisos SELECT

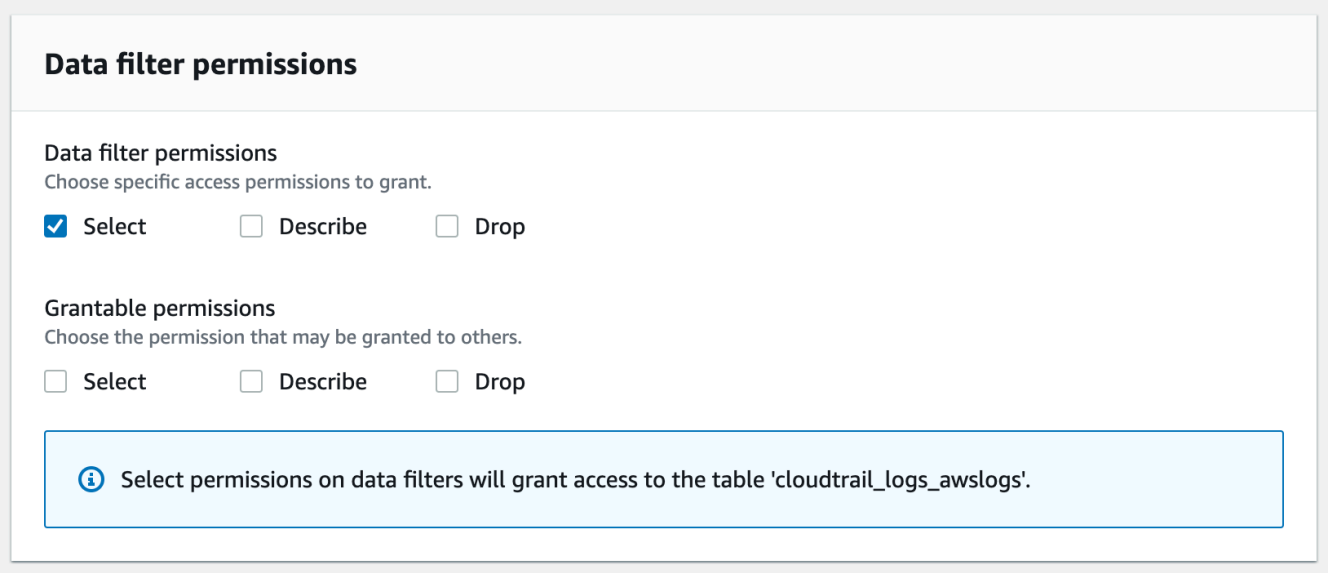
1. Vaya a la página Permisos de la consola de Lake Formation y elija Concesión.



2. Seleccione las entidades principales a las que desea proporcionar acceso y seleccione Recursos del Catálogo de datos con nombre.



3. Para proporcionar acceso a los datos que representa el filtro, elija Seleccionar en Permisos de filtro de datos.



CLI

Introduzca un comando `grant-permissions`. Especifique `DataCellsFilter` el argumento de recurso y `SELECT` para el argumento de permisos.

En el siguiente ejemplo, se concede `SELECT` con la opción de concesión al usuario `datalake_user1` del filtro de datos`restrict-pharma`, que pertenece a la `orders` tabla de la `sales` base de datos en Cuenta de AWS `1111-2222-3333`.

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

A continuación, se muestra el contenido del archivo `grant-params.json`.

```
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
    },
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
```

```
    "Permissions": ["SELECT"]
}
```

Visualización de filtros de datos

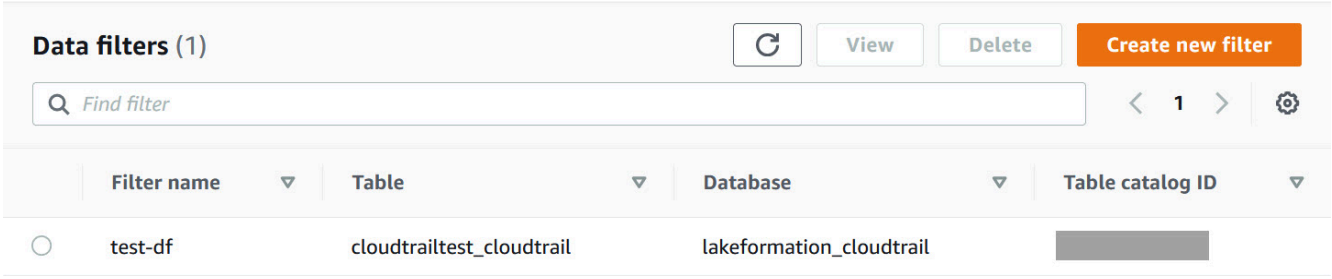
Puede usar la consola de Lake Formation o la API de Lake Formation para ver los filtros de datos. AWS CLI

Para ver los filtros de datos, debe ser administrador de lago de datos o tener los permisos necesarios sobre los filtros de datos.

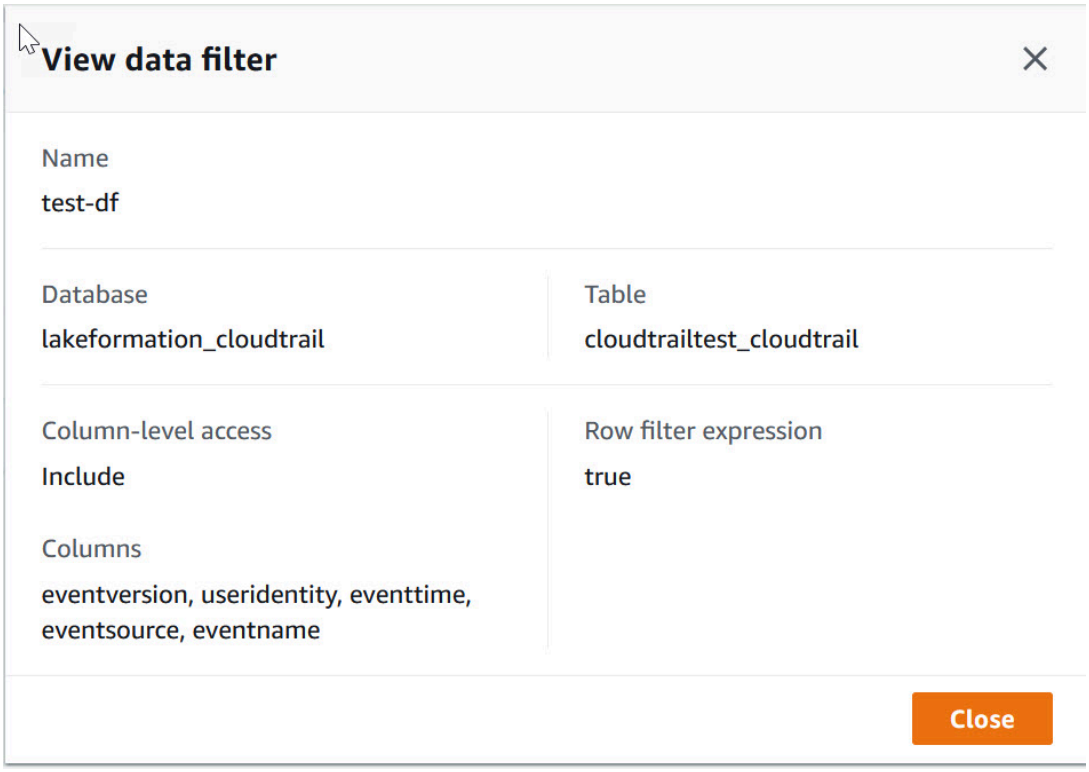
Console

1. Inicie sesión en la consola de Lake Formation Consola de administración de AWS y ábrala en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación, en Catálogo de datos, seleccione Filtros de datos.

   La página muestra los filtros de datos a los que tiene acceso.



3. Para ver los detalles del filtro de datos, elija el filtro de datos y, a continuación, elija Ver. Aparece una nueva ventana con la información detallada del filtro de datos.



AWS CLI

Introduzca un comando `list-data-cells-filter` y especifique un recurso de tabla.

El siguiente ejemplo muestra los filtros de datos de la tabla `cloudtrailtest_cloudtrail`.

```
aws lakeformation list-data-cells-filter --table '{ "CatalogId":"123456789012",
"DatabaseName":"lakeformation_cloudtrail", "Name":"cloudtrailtest_cloudtrail"}'
```

API/SDK

Use la API `ListDataCellsFilter` y especifique un recurso de la tabla.

El ejemplo siguiente utiliza Python para listar los 20 primeros filtros de datos de la tabla `myTable`.

```
response = client.list_data_cells_filter(
    Table = {
        'CatalogId': '111122223333',
        'DatabaseName': 'mydb',
        'Name': 'myTable'
    },
```

```
    MaxResults=20
)
```

Lista de permisos de filtro de datos

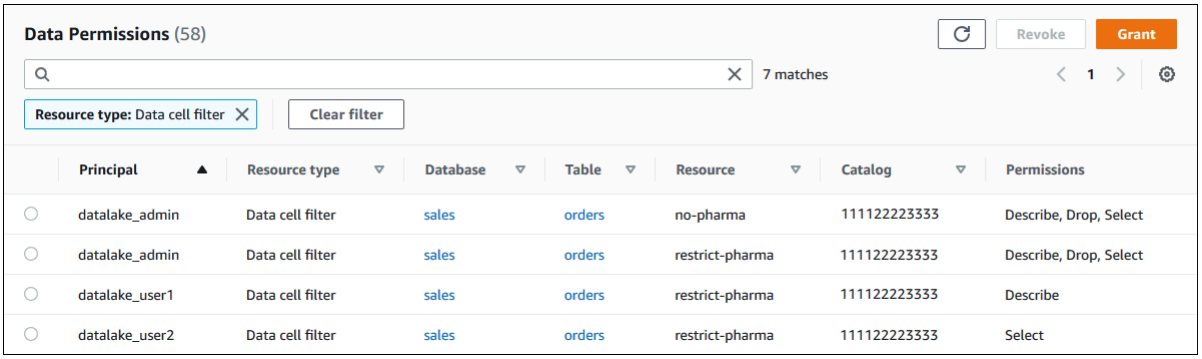
Puede utilizar la consola de Lake Formation para ver los permisos concedidos sobre los filtros de datos.

Para ver los permisos de un filtro de datos, debe ser administrador de lago de datos o tener los permisos necesarios sobre el filtro de datos.

Console

1. Inicie sesión en la consola de Lake Formation Consola de administración de AWS y ábrala en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación, en Permisos, seleccione Permisos de datos.
3. En la página Permisos de datos, haga clic o toque en el campo de búsqueda y, en el menú Propiedades, elija Tipo de recurso.
4. En el menú Tipo de recurso, seleccione Tipo de recurso: filtro de celda de datos.

   Se muestran los filtros de datos para los que tiene permisos. Es posible que tenga que desplazarse horizontalmente para ver las columnas Permisos y Concedibles.



AWS CLI

- Introduzca un comando `list-permissions`. Especifique `DataCellsFilter` para el argumento de `resource` y `DESCRIBE` o `DROP` para el argumento de `Permissions` y, opcionalmente, para el argumento de `PermissionsWithGrantOption`.

El siguiente ejemplo enumera los permisos `DESCRIBE` con la opción de concesión en el filtro de datos `restrict-pharma`. Los resultados se limitan a los permisos concedidos al director `datalake_user1` y a la `orders` tabla de la `sales` base de datos de la AWS cuenta 1111-2222-3333.

```
aws lakeformation list-permissions --cli-input-json file://list-params.json
```

A continuación, se muestra el contenido del archivo `grant-params.json`.

```
{
    "Principal": {"DataLakePrincipalIdentifier":
 "arn:aws:iam::111122223333:user/datalake_user1"},
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
    "Permissions": ["DESCRIBE"],
    "PermissionsWithGrantOption": ["DESCRIBE"]
}
```

Consulta de los permisos de bases de datos y tablas en Lake Formation

Puede consultar los permisos de Lake Formation concedidos sobre una base de datos o tabla del Catálogo de datos. Puede hacerlo mediante la consola de Lake Formation, la API o AWS Command Line Interface (AWS CLI).

Mediante la consola, puede ver los permisos a partir de las páginas Bases de datos o Tablas, o desde la página Permisos de datos.

> **Note**
>
> Si no es administrador de la base de datos ni propietario del recurso, podrá ver los permisos que otras entidades principales tienen sobre el recurso solo si tiene un permiso de Lake Formation sobre el recurso con la opción de concesión.
> Además de los permisos de Lake Formation requeridos, necesita los permisos AWS Identity and Access Management (IAM)`glue:GetDatabases`,`glue:GetDatabase`, `glue:GetTablesglue:GetTable`, `ylakeformation:ListPermissions`.

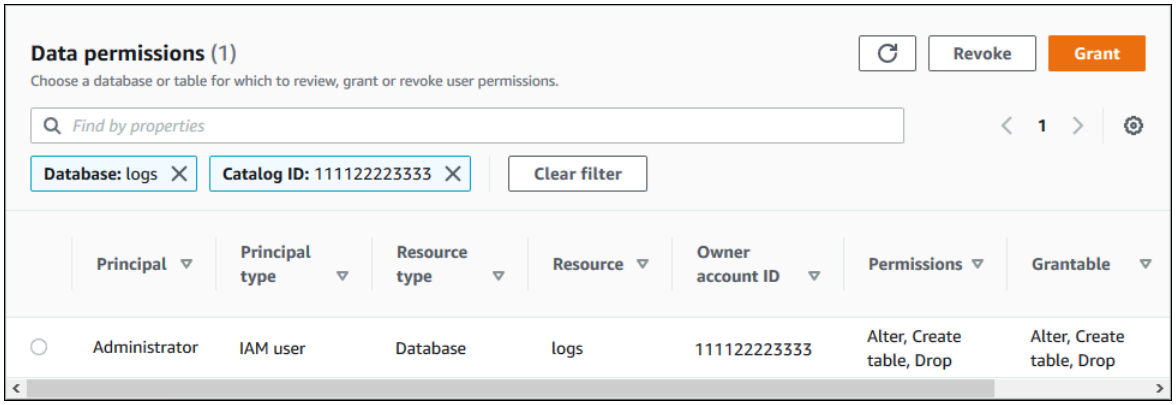
Para ver los permisos sobre una base de datos (consola, empezando por la página Bases de datos)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador del lago de datos, creador de la base de datos o como usuario que tenga algún permiso de Lake Formation sobre la base de datos con la opción de concesión.
2. En el panel de navegación, elija Databases (Bases de datos).
3. Elija una base de datos y, en el menú Acciones, seleccione Ver permisos.

   > **Note**
   >
   > Si elige un enlace de recurso de base de datos, Lake Formation muestra los permisos sobre el enlace de recursos, no sobre la base de datos de destino del enlace de recursos.

   La página Permisos de datos enumera todos los permisos de Lake Formation para la base de datos. El nombre de la base de datos y el ID de catálogo (ID de AWS cuenta) del propietario de la base de datos aparecen como etiquetas debajo del cuadro de búsqueda. Los mosaicos indican que se ha aplicado un filtro para enumerar los permisos únicamente para esa base de datos. Puede ajustar el filtro cerrando un mosaico o eligiendo Borrar filtro.



Para ver los permisos sobre una base de datos (consola, empezando por la página Permisos de datos)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador del lago de datos, creador de la base de datos o como usuario que tenga algún permiso de Lake Formation sobre la base de datos con la opción de concesión.
2. En el panel de navegación, seleccione Permisos de datos.
3. Coloque el cursor en el cuadro de búsqueda y, en el menú de propiedades que aparece, elija Base de datos.
4. En el menú Bases de datos que aparece, elija una.

> **Note**
>
> Si elige un enlace de recurso de base de datos, Lake Formation muestra los permisos sobre el enlace de recursos, no sobre la base de datos de destino del enlace de recursos.

La página Permisos de datos enumera todos los permisos de Lake Formation para la base de datos. El nombre de la base de datos aparece como un mosaico debajo del cuadro de búsqueda. El mosaico indica que se ha aplicado un filtro para enumerar los permisos únicamente para esa base de datos. Puede eliminar el filtro cerrando un mosaico o seleccionando Borrar filtro.

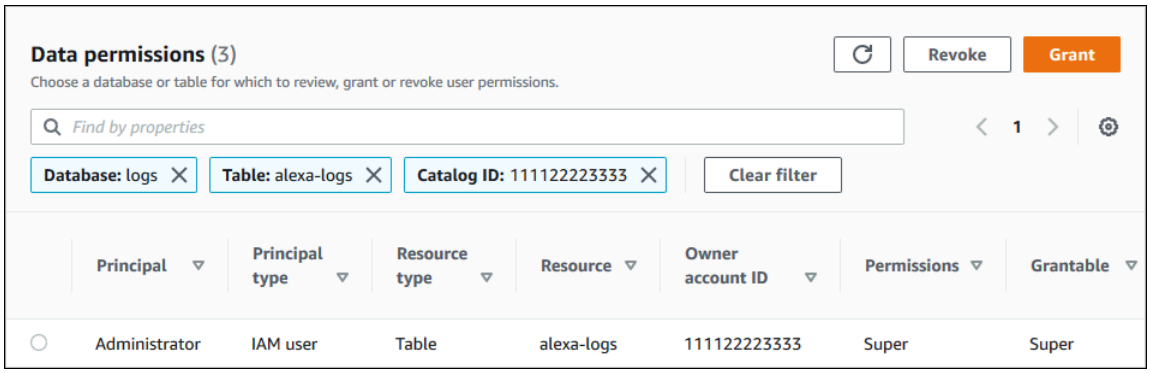
Para ver los permisos sobre una tabla (consola, empezando por la página Tablas)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador del lago de datos, creador de la tabla o como usuario que tenga algún permiso de Lake Formation sobre la tabla con la opción de concesión.
2. En el panel de navegación, elija Tablas.
3. Elija una tabla y, en el menú Acciones, seleccione Ver permisos.

> **Note**
>
> Si elige un enlace de recurso de tabla, Lake Formation muestra los permisos sobre el enlace de recursos, no sobre la tabla de destino del enlace de recursos.

La página Permisos de datos enumera todos los permisos de Lake Formation para la tabla. El nombre de la tabla, el nombre de la base de datos que contiene la tabla y el identificador de catálogo (identificador de AWS cuenta) del propietario de la tabla aparecen como etiquetas en el cuadro de búsqueda. Las etiquetas indican que se ha aplicado un filtro para enumerar los permisos únicamente para esa tabla. Puede ajustar el filtro cerrando una etiqueta o seleccionando Borrar filtro.



Para ver los permisos sobre una tabla (consola, empezando por la página Permisos de datos)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador del lago de datos, creador de la tabla o como usuario que tenga algún permiso de Lake Formation sobre la tabla con la opción de concesión.
2. En el panel de navegación, seleccione Permisos de datos.

3. Coloque el cursor en el cuadro de búsqueda y, en el menú de propiedades que aparece, elija Base de datos.
4. En el menú Bases de datos que aparece, elija una.

   > **Important**
   >
   > Si desea ver los permisos de una tabla que se compartió con su AWS cuenta desde una cuenta externa, debe elegir la base de datos de la cuenta externa que contiene la tabla, no un enlace de recursos a la base de datos.

   La página Permisos de datos enumera todos los permisos de Lake Formation para la base de datos.
5. Coloque el cursor en el cuadro de búsqueda y, en el menú Propiedades que aparece, elija Tabla.
6. En el menú Tablas que aparece, elija una.

   La página Permisos de datos enumera todos los permisos de Lake Formation para la tabla. El nombre de la tabla y el nombre de la base de datos que la contiene aparecen como mosaicos bajo el cuadro de búsqueda. Los mosaicos indican que se ha aplicado un filtro para listar los permisos solo para esa tabla. Puede ajustar el filtro cerrando un mosaico o eligiendo Borrar filtro.

Para ver los permisos en una tabla (AWS CLI)

- Introduzca un comando `list-permissions`.

  En el siguiente ejemplo, se enumeran los permisos de una tabla compartida desde una cuenta externa. La `CatalogId` propiedad es el identificador de AWS cuenta de la cuenta externa y el nombre de la base de datos hace referencia a la base de datos de la cuenta externa que contiene la tabla.

```
aws lakeformation list-permissions  --resource-type TABLE --resource '{ "Table":
 {"DatabaseName":"logs", "Name":"alexa-logs", "CatalogId":"123456789012"}}'
```

Revocación de permisos mediante la consola de Lake Formation

Puede utilizar la consola para revocar todos los tipos de permisos de Lake Formation: de Catálogo de datos, de etiqueta de política, de filtro de datos y de ubicación.

Para revocar los permisos de Lake Formation sobre un recurso (consola)

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como administrador del lago de datos o como usuario al que se le han concedido permisos con la opción de concesión en el recurso.
2. En el panel de navegación, en Permisos, elija Permisos LF-Tags y permisos de Data Lake o Ubicaciones de datos.
3. Seleccione el permiso o la ubicación y, a continuación, elija Revocar.
4. En el cuadro de diálogo que se abre, elija Revocar.

Cross-account intercambio de datos en Lake Formation

Las capacidades multicuenta de Lake Formation permiten a los usuarios compartir de forma segura lagos de datos distribuidos entre varias AWS organizaciones o directamente con los directores de IAM en otra cuenta Cuentas de AWS, lo que proporciona un acceso detallado a los metadatos del catálogo de datos y a los datos subyacentes. Las grandes empresas suelen utilizar varias cuentas Cuentas de AWS, y es posible que muchas de esas cuentas necesiten acceder a un lago de datos gestionado por una sola persona. Cuenta de AWS Los usuarios y los trabajos de AWS Glue extracción, transformación y carga (ETL) pueden consultar y unir tablas en varias cuentas y, aun así, aprovechar las protecciones de datos a nivel de tabla y columna de Lake Formation.

Al conceder permisos de Lake Formation sobre un recurso del catálogo de datos a una cuenta externa o directamente a un director de IAM en otra cuenta, Lake Formation utiliza el servicio AWS Resource Access Manager (AWS RAM) para compartir el recurso. Si la cuenta del beneficiario está en la misma organización que la cuenta del concedente, el recurso compartido estará disponible inmediatamente para el beneficiario. Si la cuenta del concesionario no pertenece a la misma organización, AWS RAM envía una invitación a la cuenta del concesionario para que acepte o rechace la concesión de recursos. Luego, para que el recurso compartido esté disponible, el administrador del lago de datos de la cuenta del concesionario debe usar la AWS RAM consola o aceptar la invitación. AWS CLI

Lake Formation permite compartir los recursos del Catálogo de datos con cuentas externas en modo de acceso híbrido. El modo de acceso híbrido proporciona la flexibilidad de habilitar selectivamente los permisos de Lake Formation para las bases de datos y tablas de su AWS Glue Data Catalog. Con el modo de acceso híbrido, ahora tiene una ruta incremental que le permite establecer los permisos de Lake Formation para un conjunto específico de usuarios sin interrumpir las políticas de permisos de otros usuarios o cargas de trabajo existentes.

Para obtener más información, consulte Modo de acceso híbrido.

Cómo compartir directamente entre cuentas

Las entidades principales autorizadas pueden compartir recursos de forma explícita con una entidad principal de IAM en una cuenta externa. Esta característica es útil cuando el propietario de una cuenta quiere controlar quién puede acceder a los recursos en la cuenta externa. Los permisos que reciba la entidad principal de IAM serán una combinación de concesiones directas y concesiones de cuenta que se transferirán en cascada a las entidades principales. Solo el destinatario de la concesión del permiso puede ver las concesiones directas entre cuentas. La entidad principal que recibe el recurso compartido no puede compartir el recurso con otras entidades principales.

Métodos para compartir los recursos del Catálogo de datos

Con una sola operación de concesión de Lake Formation, puede conceder permisos entre cuentas en los siguientes recursos del Catálogo de datos.

- Una base de datos
- Una tabla individual (con filtrado de columnas opcional)
- Algunas tablas seleccionadas
- Todas las tablas de una base de datos (mediante el comodín Todas las tablas)

Existen dos opciones para compartir sus bases de datos y tablas con otra persona Cuenta de AWS o con los directores de IAM de otra cuenta.

- Control de acceso basado en etiquetas de Lake Formation (LF-TBAC) (recomendado)

  El control de acceso basado en etiquetas de Lake Formation es una estrategia de autorización que define permisos basados en atributos. Puede utilizar el control de acceso basado en etiquetas para compartir los recursos del catálogo de datos (bases de datos, tablas y columnas) con entidades principales, Cuentas de AWS organizaciones y unidades organizativas (OU) de IAM externas. En

Lake Formation, estos atributos se denominan LF-tags. Para más información, consulte Gestión de un lago de datos mediante el control de acceso basado en etiquetas de Lake Formation.

> **Note**
>
> El LF-TBAC método de concesión de permisos del catálogo de datos se utiliza AWS Resource Access Manager para las concesiones entre cuentas.
> Lake Formation ahora admite la concesión de permisos entre cuentas a organizaciones y unidades organizativas mediante el LF-TBAC método.
> Para activar esta prestación, debe actualizar la Configuración de la versión entre cuentas a la Versión 3 o una versión posterior.
> Para obtener más información, consulte Actualización de los ajustes de la versión entre cuentas para compartir datos.

- Recursos con nombre de Lake Formation

  El intercambio de datos entre cuentas de Lake Formation mediante el método de recursos con nombre asignado le permite conceder permisos de Lake Formation con una opción de concesión en las tablas y bases de datos del catálogo de datos a directores Cuentas de AWS, organizaciones o unidades organizativas externas de IAM. La operación de concesión comparte automáticamente esos recursos.

> **Note**
>
> También puedes permitir que el AWS Glue rastreador acceda a un almacén de datos en una cuenta diferente con las credenciales de Lake Formation. Para obtener más información, consulte el Cross-account rastreo en la Guía AWS Glue para desarrolladores.

Los servicios integrados, como Athena y Amazon Redshift Spectrum, requieren enlaces a recursos para poder incluir recursos compartidos en las consultas. Para obtener más información sobre los enlaces de recursos, consulte Cómo funcionan los enlaces de recursos en Lake Formation.

Para ver las consideraciones y limitaciones, consulte Prácticas recomendadas y consideraciones para uso compartido de datos entre cuentas.

Temas

- Requisitos previos

- Actualización de los ajustes de la versión entre cuentas para compartir datos
- Compartir tablas y bases de datos del catálogo de datos entre Cuentas de AWS o directores de IAM de cuentas externas
- Conceder permisos en una base de datos o tabla compartida con su cuenta
- Conceder permisos de enlace de recursos
- Acceso a los datos subyacentes de una tabla compartida
- Cross-account CloudTrail registro
- Administrar los permisos entre cuentas mediante ambos AWS Glue y Lake Formation
- Visualización de todas las concesiones entre cuentas mediante la operación de GetResourceShares API

Temas relacionados

- Descripción general de los permisos de Lake Formation
- Acceso y visualización de tablas y bases de datos compartidas del Catálogo de datos
- Creación de enlaces de recursos
- Solución de problemas de acceso entre cuentas

Requisitos previos

Antes de que su AWS cuenta pueda compartir los recursos del catálogo de datos (catálogos, bases de datos y tablas) con otra cuenta o con las entidades principales de otra cuenta, y antes de que pueda acceder a los recursos compartidos con su cuenta, debe cumplir los siguientes requisitos previos.

Requisitos generales para compartir datos entre cuentas

- Para compartir bases de datos y tablas del Catálogo de datos en modo de acceso híbrido y compartir objetos en los catálogos federados, debe actualizar la Configuración de la versión entre cuentas a la Versión 4.
- Antes de conceder permisos entre cuentas en un recurso del Catálogo de datos, debe revocar todos los permisos de Lake Formation del grupo `IAMAllowedPrincipals` para el recurso. Si la entidad principal que llama tiene permisos entre cuentas para acceder a un

recurso y el permiso `IAMAllowedPrincipals` existe en el recurso, Lake Formation genera `AccessDeniedException`.

Este requisito solo se aplica cuando se registra la ubicación de datos subyacente en el modo de Lake Formation. Si registra la ubicación de datos en modo híbrido, los permisos del grupo `IAMAllowedPrincipals` pueden existir en la base de datos o tabla compartida.

- En las bases de datos que contengan tablas que desee compartir, debe evitar que las nuevas tablas tengan una concesión predeterminada de `Super` a `IAMAllowedPrincipals`. En la consola de Lake Formation, edite la base de datos y desactive Usar solo el control de acceso de IAM para las nuevas tablas de esta base de datos o introduzca el siguiente AWS CLI comando, sustituyéndolo por `database` el nombre de la base de datos. Si la ubicación de datos subyacente está registrada en modo de acceso híbrido, no necesita cambiar esta configuración predeterminada. En el modo de acceso híbrido, Lake Formation le permite aplicar de forma selectiva los permisos de Lake Formation y las políticas de permisos de IAM para Amazon S3 y AWS Glue en el mismo recurso.

```
aws glue update-database --name database --database-input
 '{"Name":"database","CreateTableDefaultPermissions":[]}'
```

- Para conceder permisos entre cuentas, el otorgante debe tener los permisos AWS Identity and Access Management (IAM) necesarios para operar y prestar servicio. AWS Glue AWS RAM La política AWS gestionada `AWSLakeFormationCrossAccountManager` concede los permisos necesarios.

  Los administradores de lagos de datos de las cuentas que reciben recursos compartidos AWS RAM deben tener la siguiente política adicional. Permite al administrador aceptar invitaciones para compartir AWS RAM recursos. También permite al administrador habilitar el intercambio de recursos con las organizaciones.

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ram:AcceptResourceShareInvitation",
```

```
                "ram:RejectResourceShareInvitation",
                "ec2:DescribeAvailabilityZones",
                "ram:EnableSharingWithAwsOrganization"
            ],
            "Resource": "*"
        }
    ]
}
```

- Si quieres compartir los recursos del catálogo de datos con AWS Organizations o las unidades organizativas, debes habilitar el uso compartido con las organizaciones en AWS RAM.

  Para obtener información sobre cómo habilitar el uso compartido con las organizaciones, consulte Habilitar el uso compartido con AWS las organizaciones en la Guía del AWS RAM usuario.

  Debe tener el permiso `ram:EnableSharingWithAwsOrganization` para habilitar el uso compartido con las organizaciones.
- Para compartir recursos directamente con una entidad principal de IAM de otra cuenta, debe actualizar la Configuración de la versión entre cuentas a la Versión 3. Esta configuración está disponible en la página Configuración del Catálogo de datos. Si utiliza la Versión 1, consulte las instrucciones para actualizar la configuración Actualización de los ajustes de la versión entre cuentas para compartir datos.
- No puede compartir los recursos del catálogo de datos cifrados con una clave gestionada por el AWS Glue servicio con otra cuenta. Solo puede compartir los recursos del Catálogo de datos cifrados con la clave de cifrado del cliente, y la cuenta que reciba el recurso compartido debe tener permisos sobre la clave de cifrado del Catálogo de datos para descifrar los objetos.

Cross-account uso de los LF-TBAC requisitos para compartir datos

- Para compartir los recursos del catálogo de datos con AWS Organizations unidades organizativas (OU), debes actualizar la configuración de la versión multicuenta a la versión 3 o superior.
- Para compartir los recursos del Catálogo de datos con la versión 3 de la Configuración de la versión entre cuentas, el concedente debe tener los permisos de IAM definidos en la política administrada AWS `AWSLakeFormationCrossAccountManager` de su cuenta.
- Si utiliza la versión 1 o la versión 2 de la configuración de la versión multicuentas, debe tener una política de recursos del catálogo de datos (`glue:PutResourcePolicy`) que lo habilite LF-TBAC. Para obtener más información, consulte Administrar los permisos entre cuentas mediante ambos AWS Glue y Lake Formation.

- Si actualmente utiliza una política de recursos del catálogo de AWS Glue datos para compartir recursos y desea conceder permisos entre cuentas mediante la versión 3 o superior de la configuración de la versión multicuenta, debe añadir el **`glue:ShareResource`** permiso en la configuración del catálogo de datos mediante la operación de `glue:PutResourcePolicy` API, tal como se muestra en la Administrar los permisos entre cuentas mediante ambos AWS Glue y Lake Formation sección. Esta política no es obligatoria si su cuenta no ha efectuado concesiones entre cuentas utilizando la política de recursos del Catálogo de datos AWS Glue (permiso `glue:PutResourcePolicy` de uso en las versiones 1 y 2) para conceder el acceso entre cuentas.

```
{
      "Effect": "Allow",
      "Action": [
        "glue:ShareResource"
      ],
      "Principal": {"Service": [
        "ram.amazonaws.com"
      ]},
      "Resource": [
        "arn:aws:glue:<region>:<account-id>:table/*/*",
        "arn:aws:glue:<region>:<account-id>:database/*",
        "arn:aws:glue:<region>:<account-id>:catalog"
      ]
    }
```

- Si su cuenta ha realizado acciones entre cuentas mediante la política de recursos del catálogo de AWS Glue datos y actualmente utiliza el método de recurso con nombre asignado o LF-TBAC con la versión 3 de la configuración entre cuentas AWS RAM para compartir recursos, que utiliza para compartir recursos, debe establecer el `EnableHybrid` argumento en `'true'` cuando invoque la operación de `glue:PutResourcePolicy` API. Para obtener más información, consulte Administrar los permisos entre cuentas mediante ambos AWS Glue y Lake Formation.

Se requiere una configuración en cada cuenta que acceda al recurso compartido

- Si comparte recursos con Cuentas de AWS, al menos un usuario de la cuenta de consumidor debe ser administrador de un lago de datos para poder ver los recursos compartidos. Para obtener más información sobre cómo crear un administrador de lago de datos, consulte Crear un administrador de lago de datos.

El administrador del lago de datos puede conceder permisos de Lake Formation sobre los recursos compartidos a otras entidades principales de la cuenta. Otras entidades principales no pueden acceder a los recursos compartidos hasta que el administrador del lago de datos les conceda permisos sobre ellos.

- Los servicios integrados, como Athena y Amazon Redshift Spectrum, requieren enlaces a recursos para poder incluir recursos compartidos en las consultas. Las entidades principales deben crear un enlace de recursos en su Catálogo de datos a un recurso compartido de otra Cuenta de AWS. Para obtener más información sobre los enlaces de recursos, consulte Cómo funcionan los enlaces de recursos en Lake Formation.
- Cuando un recurso se comparte directamente con una entidad principal de IAM, para consultar la tabla utilizando Athena, la entidad principal debe crear un enlace de recursos. Para crear un enlace de recursos, la entidad principal necesita el permiso `CREATE_TABLE` o `CREATE_DATABASE` de Lake Formation y el permiso de `glue:CreateTable` o `glue:CreateDatabase` de IAM.

  Si la cuenta de productor comparte una tabla diferente en la misma base de datos con la misma entidad principal u otra, esta puede consultar la tabla inmediatamente.

> **Note**
>
> Para el administrador del lago de datos y para las entidades principales a las que el administrador del lago de datos haya concedido permisos, los recursos compartidos aparecen en el Catálogo de datos como si fueran recursos locales (de su propiedad). Las tareas de extracción, transformación y carga (ETL) pueden acceder a los datos subyacentes de los recursos compartidos.
> En el caso de los recursos compartidos, las páginas Tablas y Bases de datos de la consola de Lake Formation muestran el ID de cuenta del propietario.
> Cuando se accede a los datos subyacentes de un recurso compartido, los eventos de CloudTrail registro se generan tanto en la cuenta del destinatario del recurso compartido como en la cuenta del propietario del recurso. Los CloudTrail eventos pueden contener el ARN del principal que accedió a los datos, pero solo si la cuenta del destinatario opta por incluir el ARN principal en los registros. Para obtener más información, consulte Cross-account CloudTrail registro.

Actualización de los ajustes de la versión entre cuentas para compartir datos

De vez en cuando, AWS Lake Formation actualiza la configuración del intercambio de datos entre cuentas para distinguir los cambios realizados en el AWS RAM uso y para admitir las actualizaciones realizadas en la función de intercambio de datos entre cuentas. Al hacer esto, Lake Formation crea una nueva versión de la Configuración de la versión entre cuentas.

Principales diferencias en las configuraciones de las versiones entre cuentas

Para obtener más información sobre cómo funciona el intercambio de datos entre cuentas en Configuraciones de la versión entre cuentas diferentes, consulte las secciones siguientes.

> **Note**
>
> Para compartir datos con otra cuenta, el concedente debe haber `AWSLakeFormationCrossAccountManager` administrado los permisos de la política de IAM. Es un requisito previo para todas las versiones.
> Actualizar la Configuración de la versión entre cuentas no afecta a los permisos que el destinatario tiene en los recursos compartidos. Esto se aplica al actualizar de la versión 1 a la versión 2, de la versión 2 a la versión 3 y de la versión 1 a la versión 3. Consulte las consideraciones que se indican a continuación al actualizar las versiones.

Versión 1

Método de recurso con nombre asignado: asigna cada permiso de Lake Formation multicuenta otorgado a un AWS RAM recurso compartido. El usuario (rol de concedente o entidad principal) no requiere permisos adicionales.

LF-TBAC método: Las concesiones de permisos de Cross-account Lake Formation no se utilizan AWS RAM para compartir datos. Debe tener permiso de `glue:PutResourcePolicy`.

Ventajas de actualizar las versiones: versión inicial, no aplicable.

Consideraciones al actualizar las versiones: versión inicial, no aplicable

Versión 2

Método de recurso con nombre asignado: optimiza el número de AWS RAM recursos compartidos al asignar varias concesiones de permisos entre cuentas a un AWS RAM recurso compartido. El usuario no necesita permisos adicionales.

LF-TBAC método: Las concesiones de permisos de Cross-account Lake Formation no se utilizan AWS RAM para compartir datos. Debe tener permiso de `glue:PutResourcePolicy`.

Ventajas de actualizar las versiones: configuración escalable de varias cuentas mediante una utilización óptima de la AWS RAM capacidad.

Consideraciones a la hora de actualizar las versiones: Los usuarios que deseen conceder permisos de Lake Formation para varias cuentas deben disponer de los permisos de la política `AWSLakeFormationCrossAccountManager` AWS gestionada. De lo contrario, debe tener los permisos `ram:AssociateResourceShare` y `ram:DisassociateResourceShare` necesarios para compartir correctamente los recursos con otra cuenta.

Versión 3

Método de recurso con nombre asignado: optimiza la cantidad de AWS RAM recursos compartidos al asignar varias concesiones de permisos entre cuentas a un AWS RAM recurso compartido. El usuario no necesita permisos adicionales.

LF-TBAC método: Lake Formation utiliza AWS RAM subvenciones entre cuentas. El usuario debe añadir la ShareResource declaración glue: al `glue:PutResourcePolicy` permiso. El destinatario debe aceptar las invitaciones para compartir recursos de AWS RAM.

Ventajas de actualizar las versiones: admite las funciones siguientes:

- Permite compartir recursos explícitamente con una entidad principal de IAM en una cuenta externa.

  Para obtener más información, consulte Concesión de permisos sobre los recursos del Catálogo de datos.
- Permite compartir entre cuentas mediante el LF-TBAC método Organizations o unidades organizativas (OU).
- Elimina la sobrecarga que supone mantener AWS Glue políticas adicionales para las subvenciones entre cuentas.

Consideraciones a la hora de actualizar las versiones: Al utilizar un LF-TBAC método para compartir recursos, si el otorgante utiliza una versión anterior a la versión 3 y el destinatario utiliza

la versión 3 o superior, el otorgante recibe el siguiente mensaje de error: «Solicitud de concesión multicuentas no válida». La cuenta de consumidor ha optado por utilizar la versión entre cuentas: v3. Actualice `DataLakeSetting` a `CrossAccountVersion` la versión mínima v3 (servicio: AmazonDataCatalog; código de estado: 400; código de error:)». InvalidInputException Sin embargo, si el otorgante utiliza la versión 3 y el destinatario utiliza la versión 1 o la versión 2, las subvenciones multicuentas utilizadas se tramitan correctamente. LF-Tags

Cross-account las subvenciones concedidas mediante el método de recurso designado son compatibles con distintas versiones. Aunque la cuenta del concedente utilice una versión anterior (versión 1 o 2) y la cuenta del destinatario utilice una versión más reciente (versión 3 o superior), la funcionalidad de acceso entre cuentas funciona correctamente, sin problemas ni errores de compatibilidad.

Para compartir recursos directamente con entidades principales de IAM en otra cuenta, solo el concedente debe utilizar la versión 3.

Cross-account Las concesiones realizadas mediante LF-TBAC el método requieren que los usuarios tengan una política de AWS Glue Data Catalog recursos en la cuenta. Al actualizar a la versión 3, LF-TBAC concede usos AWS RAM. Para permitir que AWS RAM las subvenciones multicuenta se realicen correctamente, debe añadir la `glue:ShareResource` declaración a las políticas de recursos del catálogo de datos existentes, tal y como se muestra en la Administrar los permisos entre cuentas mediante ambos AWS Glue y Lake Formation sección.

Versión 4

El otorgante necesita la versión 4 o superior para compartir los recursos del Catálogo de datos en el modo de acceso híbrido o compartir objetos en un catálogo federado.

Versión 5

La versión 5 de Cross Account mejora el intercambio de recursos entre cuentas, lo que le permite compartir un número ilimitado de tablas con otra cuenta, lo que elimina los límites de asociación de recursos anteriores por tipo de recurso. Para empezar, actualice a la versión 5 para varias cuentas a través de la consola o la API de Lake Formation. Cualquier nueva concesión de permisos entre cuentas utilizará automáticamente patrones comodín en el recurso compartido, en lugar de asociaciones de recursos individuales. Todos los recursos compartidos entre cuentas existentes siguen funcionando y todas las API de Lake Formation existentes siguen siendo compatibles.

Ventajas de la actualización de las versiones: la Cross-account versión 5 mejora el uso compartido entre cuentas, lo que le permite compartir cientos de miles de tablas entre cuentas.

Consideraciones a la hora de actualizar las versiones: las nuevas subvenciones que se obtengan tras la actualización de la versión 5 añadirán patrones de recursos comodín a los recursos compartidos de AWS Resource Manager existentes o crearán nuevos recursos compartidos con patrones comodín. Una vez actualizada a la versión 5, no se admite la degradación a una versión anterior.

Optimizar AWS RAM recursos compartidos

Las nuevas versiones (versión 2 y versiones posteriores) de las subvenciones multicuentas utilizan de manera óptima la AWS RAM capacidad para maximizar el uso multicuenta. Al compartir un recurso con un director externo Cuenta de AWS o de IAM, Lake Formation puede crear un nuevo recurso compartido o asociar el recurso con un recurso compartido existente. Al asociarse a los recursos compartidos existentes, Lake Formation reduce el número de invitaciones a compartir recursos que un consumidor tiene que aceptar. La versión 5 optimiza aún más el uso de la RAM mediante patrones de recursos basados en caracteres comodín en lugar de asociaciones de recursos individuales, lo que reduce considerablemente las asociaciones de recursos por recurso compartido.

Habilitado AWS RAM comparte mediante TBAC o comparte los recursos directamente con los directores

Para compartir recursos directamente con entidades principales de IAM en otra cuenta o para habilitar los recursos compartidos entre cuentas TBAC a organizaciones o unidades organizativas, debe actualizar la Configuración de la versión entre cuentas a la versión 3. Para obtener más información sobre los límites de AWS RAM recursos, consulte. Prácticas recomendadas y consideraciones para uso compartido de datos entre cuentas

Permisos necesarios para actualizar la configuración de las versiones entre cuentas

Si un concedente de permisos entre cuentas tiene permisos de política de IAM gestionados por `AWSLakeFormationCrossAccountManager`, no se requiere ninguna configuración de permisos adicional para el rol o la entidad principal del concedente de permisos entre cuentas. Sin embargo, si el concedente entre cuentas no está utilizando la política administrada, entonces el rol de concedente o entidad principal debe tener concedidos los siguientes permisos de IAM para que la nueva versión de la concesión entre cuentas tenga éxito.

JSON

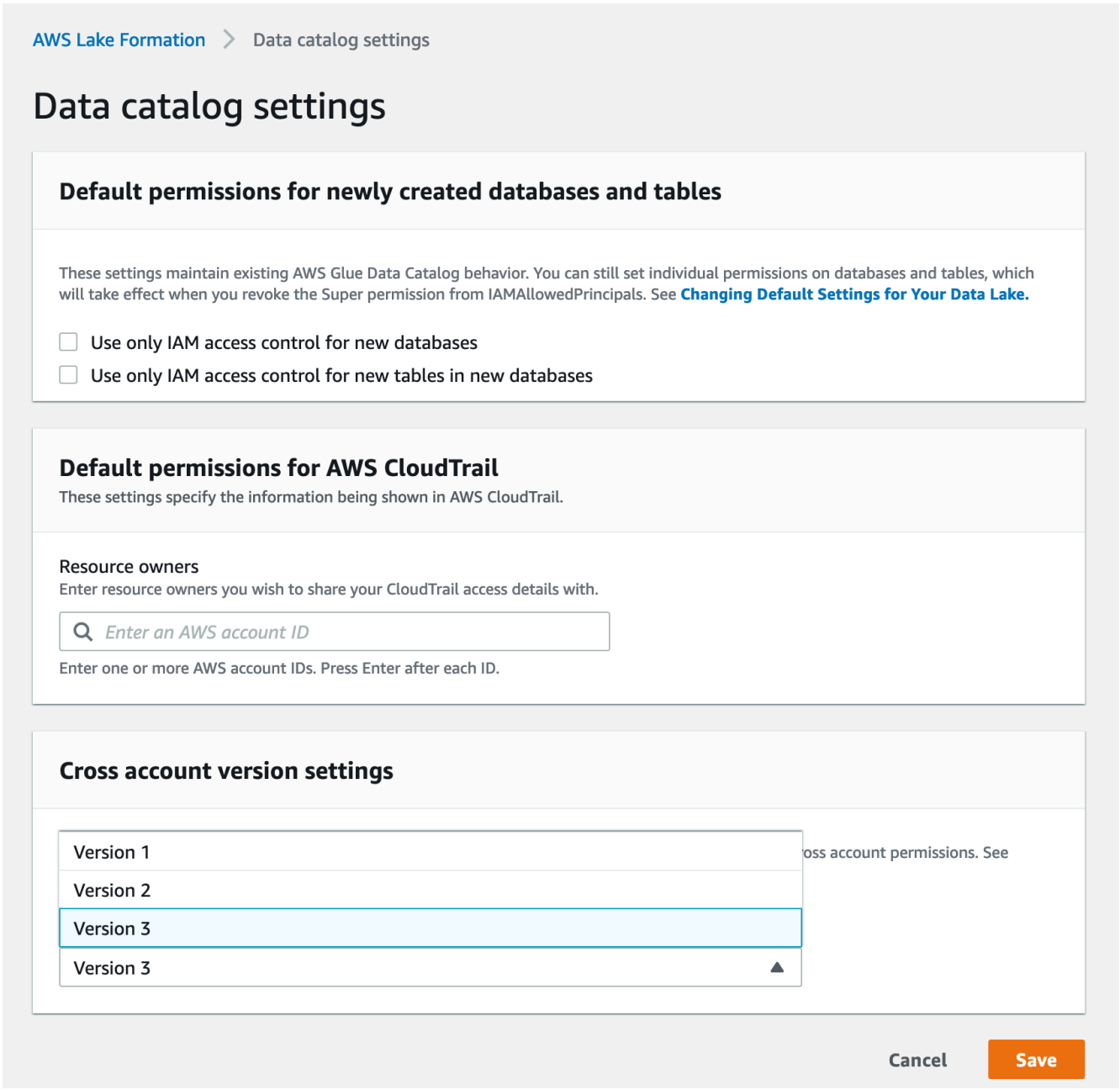
```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Sid": "VisualEditor1",
      "Effect": "Allow",
      "Action": [
         "ram:AssociateResourceShare",
         "ram:DisassociateResourceShare",
         "ram:GetResourceShares"
       ],
     "Resource": "*",
     "Condition": {
       "StringLike": {
         "ram:ResourceShareName": "LakeFormation*"
        }
       }
    }
  ]
}
```

Para habilitar la nueva versión

Siga estos pasos para actualizar la configuración de la versión multicuenta a través de la AWS Lake Formation consola o el AWS CLI.

Console

1. Elija la versión 2, la versión 3, la versión 4 o la versión 5 en la configuración de la versión multicuenta en la página de configuración del catálogo de datos. Si selecciona Versión 1, Lake Formation utilizará el modo de recursos compartidos predeterminado.



2. Seleccione Save.

AWS Command Line Interface (AWS CLI)

Utilice el `put-data-lake-settings` AWS CLI comando para configurar el `CROSS_ACCOUNT_VERSION` parámetro. Los valores aceptados son 1, 2, 3, 4 y 5.

```
aws lakeformation put-data-lake-settings --region us-east-1 --data-lake-settings
 file://settings
{
```

```
    "DataLakeAdmins": [
        {
            "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/test"
        }
    ],
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "Parameters": {
        "CROSS_ACCOUNT_VERSION": "3"
    }
}
```

Important

En cuanto elija la Versión 2 o la Versión 3, todas las nuevas concesiones de recursos con nombre pasarán por el nuevo modo de concesión entre cuentas. Para aprovechar al máximo la AWS RAM capacidad de tus acciones multicuentas existentes, te recomendamos que revoques las concesiones que se concedieron con la versión anterior y las vuelvas a conceder en el nuevo modo.

Compartir tablas y bases de datos del catálogo de datos entre Cuentas de AWS o directores de IAM de cuentas externas

En esta sección se incluyen instrucciones sobre cómo conceder permisos multicuentas sobre los recursos del catálogo de datos a una AWS cuenta externa, una entidad principal de IAM, una AWS organización o una unidad organizativa. La operación de concesión comparte automáticamente esos recursos.

Temas

- Compartir datos mediante el control de acceso basado en etiquetas
- Cross-account compartir datos mediante el método de recurso designado

Compartir datos mediante el control de acceso basado en etiquetas

AWS Lake Formation El control de acceso basado en etiquetas (LF-TBAC) es una estrategia de autorización que define los permisos en función de los atributos. En los siguientes pasos se explica cómo conceder permisos para varias cuentas mediante. LF-Tags

Se requiere una configuración en la cuenta producer/grantor

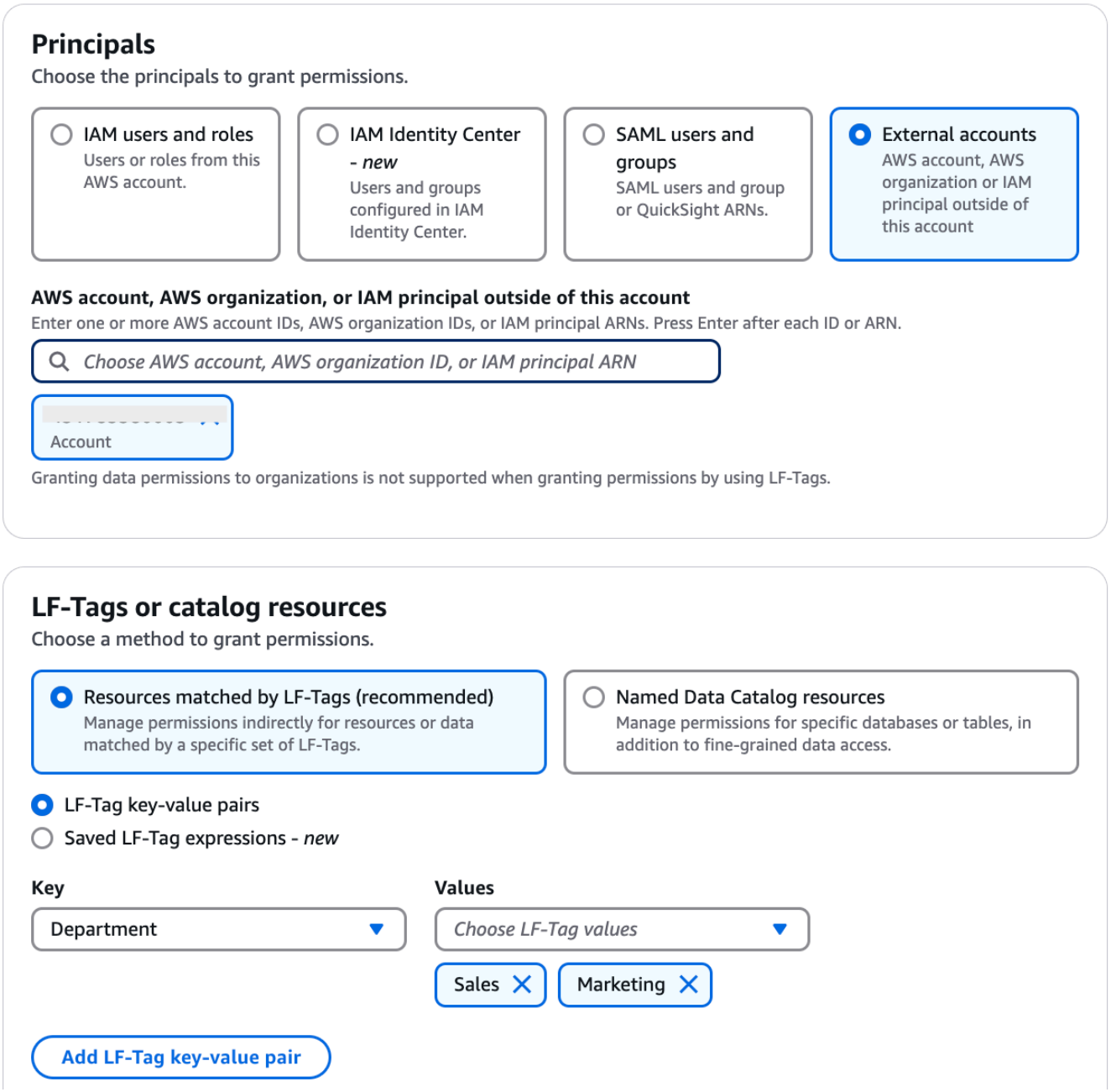
1. Añadir LF-Tags.

    1. Inicie sesión en la consola de Lake Formation como administrador o LF-Tag creador de un lago de datos.
    2. En la barra de navegación izquierda, elija Permisos LF-Tags y permisos.
    3. Elija Añadir LF-Tag.

    Para obtener instrucciones detalladas sobre la creación LF-Tags, consulteCreando LF-Tags.
2. Conceda a Describe and/or Associate permisos pares LF-Tag de valores clave a los directores de IAM de su cuenta o de cuentas externas.

    Al conceder permisos a los pares LF-Tag clave-valor, los directores pueden ver los recursos del catálogo de datos (bases de datos LF-Tags, tablas y columnas) y asignarlos a ellos.
3. A continuación, el administrador del lago de datos o un director de IAM con permiso de asociado pueden asignarlos LF-Tag a bases de datos, tablas o columnas. Para obtener más información, consulte Asignación LF-Tags a recursos del catálogo de datos.
4. A continuación, conceda permisos de datos a cuentas externas mediante LF-Tag expresiones. Esto permite a quien concede o recibe los permisos acceder a los recursos del Catálogo de datos que están etiquetados con las mismas claves y valores.

    1. En el panel de navegación, elija Permisos y Permisos de datos.
    2. Elija Conceder.
    3. En la página Concesión de permisos, para Entidades principales, elija Cuentas externas y especifique el ID de Cuenta de AWS del beneficiario o el rol de IAM de la entidad principal o el nombre de recurso de Amazon (ARN) para la entidad principal (ARN de la entidad principal) si otorga una concesión directa entre cuentas a una entidad principal externa. Debe pulsar Intro después de especificar el ID de cuenta.



4. Para los recursos LF-Tags o para catalogarlos, elija los recursos que coincidan con LF-Tags (recomendado).

   a. Elija la opción pares LF-Tag clave-valor o Expresiones guardadas LF-Tag .

   b. Si elige pares LF-Tag clave-valor, introduzca la clave y los valores asociados al recurso del catálogo de datos LF-Tagque se comparte con la cuenta del concesionario.

      Al concesionario se le conceden permisos sobre los recursos del catálogo de datos a los que se les asignó una coincidencia en la expresión. LF-Tag LF-Tag Si la LF-Tag expresión

especifica varios valores por clave de etiqueta, cualquiera de los valores de la etiqueta puede coincidir.

5. Elija los permisos a nivel de base de datos o de tabla para conceder a los recursos que coincidan con la expresión. LF-Tag

> **Important**
>
> Como el administrador del lago de datos debe conceder permisos sobre los recursos compartidos a las entidades principales de la cuenta beneficiaria, los permisos entre cuentas siempre deben concederse con la opción de concesión.

Para obtener más información, consulte LF-Tag Otorgar permisos mediante la consola.

> **Note**
>
> Las entidades principales que reciban concesiones directas entre cuentas no dispondrán de la opción Permisos concedibles.

La configuración es obligatoria en la cuenta receiving/grantee

1. Inicie sesión en la consola de Lake Formation como administrador del lago de datos de la cuenta consumidora.
2. A continuación, reciba el recurso compartido en la cuenta consumidora.
    1. Abre la AWS RAM consola.
    2. En el panel de navegación, en Compartidos conmigo, elija Recursos compartidos.
    3. Seleccione los recursos compartidos y elija Aceptar el recurso compartido.
3. Al compartir un recurso con otra cuenta, el recurso sigue perteneciendo a la cuenta del productor y no está visible en la consola de Athena. Para que el recurso sea visible en la consola de Athena, debe crear un enlace de recurso que apunte al recurso compartido. Para obtener instrucciones sobre cómo crear un enlace a un recurso, consulte Crear un enlace de recursos a una tabla de Catálogo de datos compartida y Crear un enlace de recursos a una base de datos de Catálogo de datos compartida
    1. Elija Bases de datos o Tablas en el Catálogo de datos.

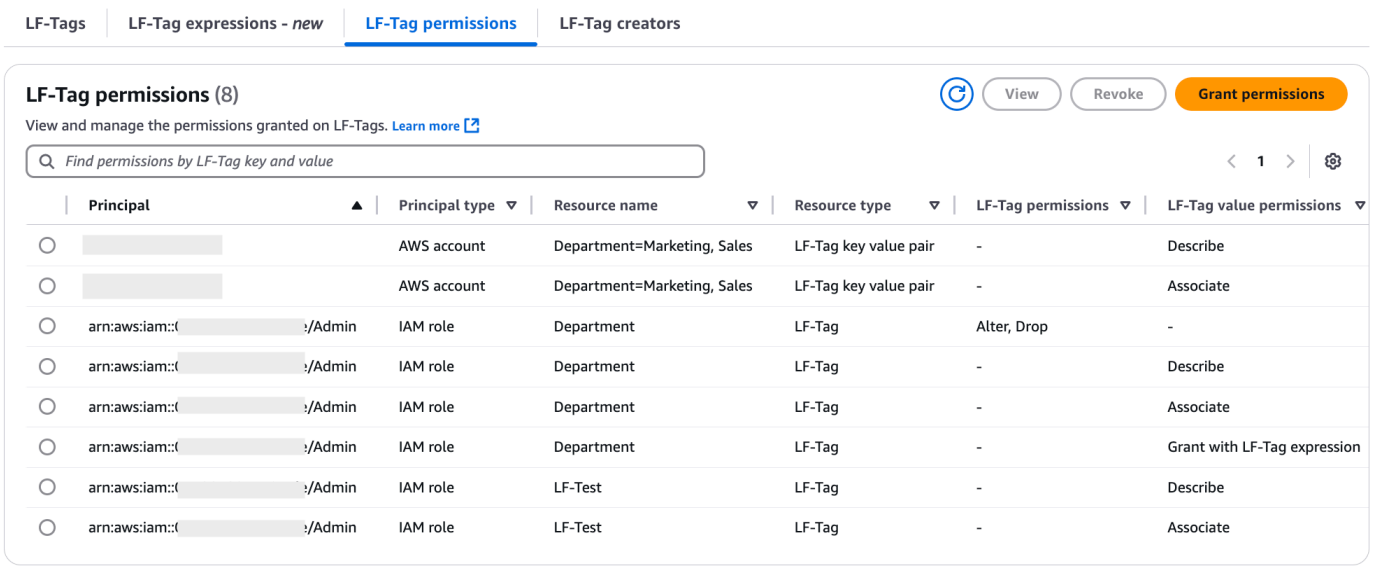
2. En la Databases/Tables página, seleccione Crear, enlace de recursos.
3. Especifique la siguiente información para un enlace de recurso de base de datos:
   - Nombre del enlace de recurso: un nombre exclusivo para el enlace de recurso.
   - Catálogo de destino: el catálogo en el que se crea el enlace de recurso.
   - Región de la base de datos compartida: la región de la base de datos compartida con usted si va a crear el enlace de recurso en una región diferente.
   - Base de datos compartida: elija la base de datos compartida.
   - ID de catálogo de la base de datos compartida: especifique el ID de catálogo de la base de datos compartida.
4. Seleccione Crear. Puede ver el enlace de recurso recién creado en la lista de bases de datos.

Del mismo modo, puede crear un enlace de recurso a una tabla compartida.

4. Ahora conceda el permiso Describir en el enlace de recurso a las entidades principales de IAM con los que va a compartir el recurso.

   1. En la Databases/Tablespágina, seleccione el enlace al recurso y, en el menú Acciones, elija Conceder.
   2. En la sección Conceder permisos, seleccione Usuarios y roles de IAM.
   3. Elija el rol de IAM al que quiera conceder acceso al enlace de recurso.
   4. En la sección de permisos del Enlace de recurso, seleccione Describir.
   5. Elija Conceder.
5. A continuación, concede permisos LF-Tag con valores clave a los principales de la cuenta del consumidor.

   Debería poder encontrar los LF-Tags que se comparten con usted en la cuenta de consumidor de la consola de Lake Formation, en Permisos LF-Tags y permisos. Puede asociar las etiquetas compartidas por el otorgante a los recursos compartidos por la cuenta otorgante, que incluyen: bases de datos, tablas y columnas. Puede conceder permisos adicionales sobre los recursos a otras entidades principales.



1. En el panel de navegación, en Permisos, Permisos de datos, seleccione Conceder.
2. En la página Conceder permisos, seleccione Usuarios y roles de IAM.
3. A continuación, selecciona los usuarios y funciones de IAM de tu cuenta para conceder el acceso a lo compartido databases/tables.
4. A continuación, para buscar recursos LF-Tags o catalogarlos, elija los recursos coincidentes por LF-Tags.
5. A continuación, elige la clave y los valores de lo LF-Tag que se comparte contigo.
6. A continuación, elija los permisos de base de datos y tabla que desea conceder a los usuarios y roles de IAM. También puede elegir permisos concedibles que permitan a los usuarios y roles de IAM conceder permisos a otros. users/roles
7. Elija Conceder.
8. Puede ver las concesiones de permisos en Permisos de datos en la consola de Lake Formation.

Cross-account compartir datos mediante el método de recurso designado

Puedes conceder permisos directamente a los directores de la otra AWS cuenta o a una cuenta externa Cuentas de AWS o AWS Organizations. Otorgar permisos de Lake Formation a organizaciones o unidades organizativas equivale a conceder el permiso a todos los Cuenta de AWS miembros de esa organización o unidad organizativa.

Al conceder permisos a cuentas u organizaciones externas, debe incluir la opción de Permisos concedibles. Solo el administrador del lago de datos de la cuenta externa puede acceder a los recursos compartidos hasta que el administrador conceda permisos sobre los recursos compartidos a otras entidades principales de la cuenta externa.

> **Note**
>
> La opción Permisos otorgables no es compatible cuando se conceden permisos directamente a entidades principales de IAM desde cuentas externas.

Siga las instrucciones de Concesión de permisos de base de datos mediante el método de recurso con nombre para conceder permisos entre cuentas según el método de recurso designado.

Conceder permisos en una base de datos o tabla compartida con su cuenta

Después de compartir con su AWS cuenta un recurso del catálogo de datos que pertenece a otra AWS cuenta, como administrador del lago de datos, puede conceder permisos sobre el recurso compartido a otros responsables de su cuenta. Sin embargo, no puede conceder permisos sobre el recurso a otras cuentas AWS u organizaciones.

Puede usar la AWS Lake Formation consola, la API o AWS Command Line Interface (AWS CLI) para conceder los permisos.

Para conceder permisos en una base de datos compartida (denominado método de recurso, consola)

- Siga las instrucciones en Concesión de permisos de base de datos mediante el método de recurso con nombre. En la lista de bases de datos de los recursos LF-Tags o del catálogo, asegúrese de seleccionar la base de datos en la cuenta externa y no un enlace a un recurso para la base de datos.

  Si no ve la base de datos en la lista, asegúrese de haber aceptado la invitación de la base de datos para compartir del recurso AWS Resource Access Manager (AWS RAM). Para obtener más información, consulte Aceptar una invitación para compartir recursos de AWS RAM.

  Además, para los permisos `CREATE_TABLE` y `ALTER`, siga las instrucciones de Concesión de permisos de localización de datos (misma cuenta) y asegúrese de introducir el ID de la cuenta propietaria en el campo Ubicación de la cuenta registrada.

Para conceder permisos en una tabla compartida (método de recurso con nombre, consola)

- Siga las instrucciones en Concesión de permisos de tabla mediante el método de recursos con nombre. En la lista de bases de datos de los recursos LF-Tags o del catálogo, asegúrese de seleccionar la base de datos en la cuenta externa y no un enlace de recursos para la base de datos.

  Si no ve la tabla en la lista de tablas, compruebe que ha aceptado la invitación para compartir el recurso AWS RAM para la tabla. Para obtener más información, consulte Aceptar una invitación para compartir recursos de AWS RAM.

  Asimismo, para el permiso de `ALTER`, siga las instrucciones de Concesión de permisos de localización de datos (misma cuenta) y asegúrese de introducir el ID de la cuenta propietaria en el campo Ubicación de la cuenta registrada.

Para conceder permisos sobre los recursos compartidos (LF-TBAC método, consola)

- Siga las instrucciones en Concesión de permisos del catálogo de datos. En la sección de recursos LF-Tags o del catálogo, otorgue la LF-Tag expresión exacta que la cuenta externa otorgó a su cuenta o un subconjunto de esa expresión.

  Por ejemplo, si una cuenta externa concedió la LF-Tag expresión `module=customers AND environment=production` a su cuenta con la opción de concesión, como administrador de un lago de datos, puede conceder la misma expresión `module=customers` o `environment=production` a un capital de su cuenta. Puede conceder solo el mismo permiso o un subconjunto de los permisos de Lake Formation (por ejemplo,`SELECT`,`ALTER`, etc.) que se concedieron a los recursos mediante la LF-Tag expresión.

Para conceder permisos en una tabla compartida (método de recurso denominado), AWS CLI)

- Introduzca un comando similar al siguiente. En este ejemplo:
  - El ID AWS de tu cuenta es 1111-2222-3333.
  - La cuenta propietaria de la tabla y que ha concedido el permiso a su cuenta es 1234-5678-9012.
  - El permiso `SELECT` se concede al usuario `datalake_user1` en la tabla compartida `pageviews`. Ese usuario es la entidad principal de su cuenta.

- La tabla `pageviews` se encuentra en la base de datos `analytics`, perteneciente a la cuenta 1234-5678-9012.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "SELECT" --resource '{ "Table": {"CatalogId":"123456789012",
 "DatabaseName":"analytics", "Name":"pageviews"}}'
```

Tenga en cuenta que la cuenta propietaria debe especificarse en la propiedad `CatalogId` del argumento `resource`.

Conceder permisos de enlace de recursos

Sigue estos pasos para conceder AWS Lake Formation permisos en uno o más enlaces de recursos a una entidad principal de tu cuenta. AWS

Después de crear un enlace de recursos, solo podrá verlo y acceder a él por sí mismo. (Esto supone que Control de acceso IAM solo para nuevas tablas en esta base de datos no está activado para la base de datos). Para permitir que otras entidades principales de su cuenta accedan al enlace de recursos, conceda al menos el permiso `DESCRIBE`.

> **Important**
>
> Conceder permisos en un enlace a un recurso no otorga permisos en la tabla o base de datos (vinculada) de destino. Debe conceder los permisos en el destino por separado.

Puede conceder permisos mediante la consola de Lake Formation, la API o AWS Command Line Interface (AWS CLI).

console

Para conceder permisos de enlaces de recursos utilizando la consola de Lake Formation

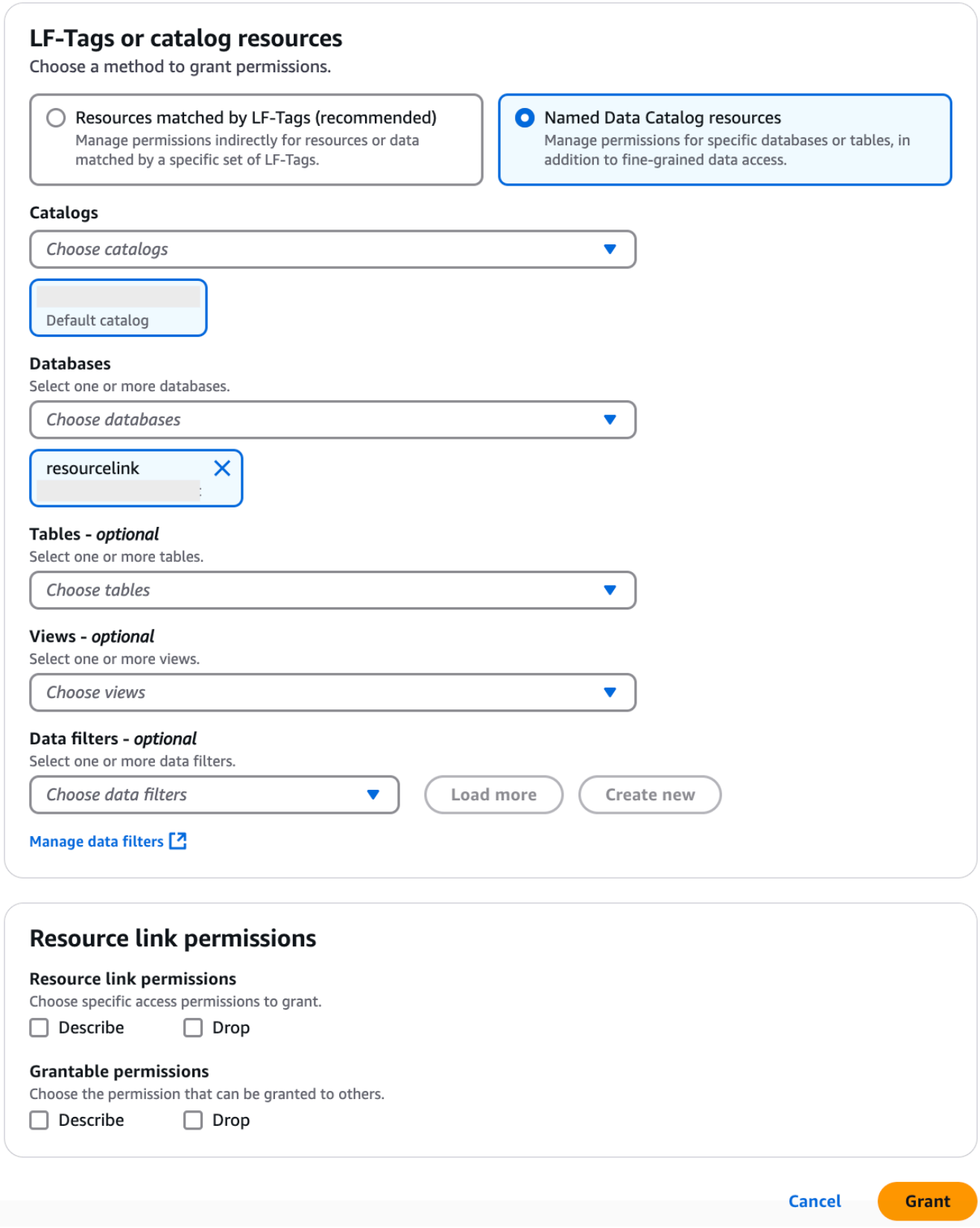
1. Realice una de las siguientes acciones:

   - Para enlaces de recursos de base de datos, siga los pasos en Concesión de permisos de base de datos mediante el método de recurso con nombre para hacer lo siguiente:

1. Seleccione el enlace de recurso en la lista de bases de datos del Catálogo de datos, Bases de datos.
2. Elija Conceder para abrir la página Conceder permisos.
3. Especifique las entidades principales para conceder permisos.
4. Se rellenan los campos Catálogos y Bases de datos.

- Para enlaces de recursos de tabla, siga los pasos de Concesión de permisos de tabla mediante el método de recursos con nombre para hacer lo siguiente:
  1. Seleccione el enlace de recurso en la lista de tablas del Catálogo de datos, Tablas.
  2. Abra la página Conceder permisos.
  3. Especifique las entidades principales.
  4. Se rellenan los campos Catálogos, Bases de datos y Tablas.
  5. Especifique las entidades principales.

2. En Permisos, seleccione los que va a conceder. Si lo desea, seleccione los permisos que pueden concederse.



3. Elija Conceder.

AWS CLI

Para conceder permisos de enlace de recursos mediante AWS CLI

- Ejecute el comando `grant-permissions` y especifique un enlace a un recurso como recurso.

  Example

  En este ejemplo, se concede `DESCRIBE` al usuario de la tabla el `datalake_user1` enlace de recursos de la base `incidents-link` de datos de `issues` la AWS cuenta 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DESCRIBE" --resource '{ "Table": {"DatabaseName":"issues",
 "Name":"incidents-link"}}'
```

Consulte también:

- Creación de enlaces de recursos
- Referencia de permisos de Lake Formation

Acceso a los datos subyacentes de una tabla compartida

Supongamos que la AWS cuenta A comparte una tabla del catálogo de datos con la cuenta B, por ejemplo, concediéndola `SELECT` con la opción de concesión de la tabla a la cuenta B. Para que un principal de la cuenta B pueda leer los datos subyacentes de la tabla compartida, se deben cumplir las siguientes condiciones:

- El administrador del lago de datos de la cuenta B debe aceptar el uso compartido. (Esto no es necesario si las cuentas A y B pertenecen a la misma organización o si la concesión se efectuó con el método de control de acceso basado en etiquetas de Lake Formation).
- El administrador del lago de datos debe volver a conceder a la entidad principal el permiso `SELECT` de Lake Formation que la cuenta A concedió en la tabla compartida.

- La entidad principal debe tener los siguientes permisos IAM sobre la tabla, la base de datos que la contiene y la cuenta A Catálogo de datos.

> **Note**
>
> En la política de IAM siguiente:
>
> - *<account-id-A>*Sustitúyala por el ID AWS de cuenta de la cuenta A.
> - *<region>*Sustitúyalo por una región válida.
> - *<database>*Sustitúyalo por el nombre de la base de datos de la cuenta A que contiene la tabla compartida.
> - *<table>*Sustitúyalo por el nombre de la tabla compartida.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
          "Effect": "Allow",
          "Action": [
            "glue:GetTable",
            "glue:GetTables",
            "glue:GetPartition",
            "glue:GetPartitions",
            "glue:BatchGetPartition",
            "glue:GetDatabase",
            "glue:GetDatabases"
           ],
           "Resource": [
            "arn:aws:glue:us-east-1:111122223333:table/<database>/<table>",
            "arn:aws:glue:us-east-1:111122223333:database/<database>",
            "arn:aws:glue:us-east-1:111122223333:catalog"
           ]
        },
        {
          "Effect": "Allow",
          "Action": [
            "lakeformation:GetDataAccess"
           ],
```

```
            "Resource": [
              "*"
             ]
        }
       ]
}
```

Consulte también:

- Aceptar una invitación para compartir recursos de AWS RAM

Cross-account CloudTrail registro

Lake Formation proporciona un registro de auditoría centralizado de todos los accesos entre cuentas a los datos de su lago de datos. Cuando una AWS cuenta de destinatario accede a los datos de una tabla compartida, Lake Formation copia el CloudTrail evento en los registros de CloudTrail la cuenta propietaria. Los eventos copiados incluyen consultas a los datos por parte de servicios integrados, como Amazon Redshift Spectrum, Amazon Athena y accesos AWS Glue a los datos por parte de los trabajos.

CloudTrail Los eventos de las operaciones entre cuentas en los recursos del catálogo de datos se copian de forma similar.

Como propietario del recurso, si habilita el registro a nivel de objeto en Amazon S3, puede ejecutar consultas que unan los eventos de S3 con CloudTrail los eventos de Lake Formation CloudTrail para determinar las cuentas que han accedido a sus buckets de S3.

Temas

- Incluir las identidades principales en los registros entre cuentas CloudTrail
- Consulta de CloudTrail registros para el acceso entre cuentas de Amazon S3

Incluir las identidades principales en los registros entre cuentas CloudTrail

De forma predeterminada, los CloudTrail eventos entre cuentas que se agregan a los registros del destinatario del recurso compartido y se copian en los registros del propietario del recurso contienen solo el ID AWS principal de la cuenta externa, no el nombre de recurso de Amazon (ARN)

legible para los humanos del principal (ARN principal). Al compartir recursos dentro de límites de confianza, como dentro de la misma organización o equipo, puede optar por incluir el ARN principal en los CloudTrail eventos. A continuación, las cuentas de propietarios de recursos pueden hacer un seguimiento de las entidades principales de las cuentas de destinatarios que acceden a sus recursos en propiedad.

> **Important**
>
> Como destinatario de recursos compartidos, para ver el ARN principal de los eventos en sus propios CloudTrail registros, debe optar por compartir el ARN principal con la cuenta del propietario.
> Si el acceso a los datos se produce a través de un enlace de recursos, se registran dos eventos en la cuenta del destinatario del recurso compartido: uno para el acceso al enlace de recursos y otro para el acceso al recurso de destino. El evento para el acceso al enlace de recursos sí incluye el ARN de la entidad principal. El evento de acceso al recurso de destino no incluye el ARN de la entidad principal sin la activación. El evento de acceso al enlace de recursos no se copia en la cuenta del propietario.

El siguiente es un extracto de un CloudTrail evento multicuenta predeterminado (sin suscripción). La cuenta que efectúa el acceso a los datos es 1111-2222-3333. Este es el registro que se muestra en la cuenta de llamada y en la cuenta del propietario del recurso. Lake Formation rellena los registros de ambas cuentas en el caso entre cuentas.

```
{
    "eventVersion": "1.05",
    "userIdentity": {
        "type": "AWSAccount",
        "principalId": "AROAQGFTBBBGOBWV2EMZA:GlueJobRunnerSession",
        "accountId": "111122223333"
    },
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GetDataAccess",
...
...
    "additionalEventData": {
        "requesterService": "GLUE_JOB",
        "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-G13T0Rmng2"
    },
...
```

```
}
```

Como consumidor de recursos compartidos, cuando opta por incluir el ARN de la entidad principal, el resumen pasa a ser el siguiente. El campo `lakeFormationPrincipal` representa el rol o el usuario final que hace la consulta a través de Amazon Athena, Amazon Redshift Spectrum o trabajos de AWS Glue.

```
{
    "eventVersion": "1.05",
    "userIdentity": {
        "type": "AWSAccount",
        "principalId": "AROAQGFTBBBGOBWV2EMZA:GlueJobRunnerSession",
        "accountId": "111122223333"
    },
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GetDataAccess",
...
...
    "additionalEventData": {
        "requesterService": "GLUE_JOB",
        "lakeFormationPrincipal": "arn:aws:iam::111122223333:role/ETL-Glue-Role",
        "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-G13T0Rmng2"
    },
...
}
```

Para optar por incluir los ARN principales en los registros de varias cuentas CloudTrail

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

   Inicie sesión como usuario de `Administrator` o como usuario con la política de IAM de `Administrator Access`.
2. En el panel de navegación, seleccione Configuración.
3. En la página de configuración del catálogo de datos, en la AWS CloudTrail sección Permisos predeterminados para los propietarios de los recursos, introduzca uno o más identificadores de cuenta del propietario del AWS recurso.

   Pulse Intro después de cada ID de cuenta.
4. Seleccione Save.

Ahora, CloudTrail los eventos entre cuentas almacenados en los registros tanto para el destinatario del recurso compartido como para el propietario del recurso contienen el ARN principal.

Consulta de CloudTrail registros para el acceso entre cuentas de Amazon S3

Como propietario de un recurso compartido, puede consultar CloudTrail los registros de S3 para determinar las cuentas que han accedido a sus buckets de Amazon S3 (siempre que haya habilitado el registro a nivel de objeto en Amazon S3). Esto solo se aplica a las ubicaciones S3 que haya registrado en Lake Formation. Si los consumidores de recursos compartidos optan por incluir los ARN principales en los CloudTrail registros de Lake Formation, puede determinar las funciones o los usuarios que accedieron a los depósitos.

Al ejecutar consultas con Amazon Athena, puede unir los eventos de Lake Formation y CloudTrail los eventos de S3 CloudTrail en la propiedad del nombre de la sesión. Las consultas también pueden filtrar los eventos de Lake Formation en `eventName="GetDataAccess"` y los eventos de S3 en `eventName="Get Object"` o`eventName="Put Object"`.

El siguiente es un extracto de un CloudTrail evento entre cuentas de Lake Formation en el que se accedió a los datos de una ubicación S3 registrada.

```
{
  "eventSource": "lakeformation.amazonaws.com",
  "eventName": "GetDataAccess",
  ..............
  ..............
  "additionalEventData": {
    "requesterService": "GLUE_JOB",
    "lakeFormationPrincipal": "arn:aws:iam::111122223333:role/ETL-Glue-Role",
    "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-B8JSAjo5QA"
   }
}
```

El valor `lakeFormationRoleSessionName` clave,`AWSLF-00-GL-111122223333-B8JSAjo5QA`, se puede unir al nombre de la sesión en la `principalId` clave del evento de S3 CloudTrail . El siguiente es un extracto del CloudTrail evento S3. Muestra la ubicación del nombre de la sesión.

```
{
```

```
   "eventSource": "s3.amazonaws.com",
   "eventName": "Get Object"
   ..............
   ..............
   "principalId": "AROAQSOX5XXUR7D6RMYLR:AWSLF-00-GL-111122223333-B8JSAjo5QA",
   "arn": "arn:aws:sets::111122223333:assumed-role/Deformationally/AWSLF-00-
GL-111122223333-B8JSAjo5QA",
   "session Context": {
     "session Issuer": {
       "type": "Role",
       "principalId": "AROAQSOX5XXUR7D6RMYLR",
       "arn": "arn:aws:iam::111122223333:role/aws-service-role/
lakeformation.amazonaws.com/Deformationally",
       "accountId": "111122223333",
       "user Name": "Deformationally"
     },
   ..............
   ..............
}
```

El nombre de la base de datos se forma de la siguiente manera:

```
AWSLF-<version-number>-<query-engine-code>-<account-id->-<suffix>
```

`version-number`

La versión de este formato, actualmente `00`. Si cambia el formato del nombre de la sesión, la siguiente versión será la `01`.

`query-engine-code`

Indica la entidad que ha accedido a los datos. Los valores actuales son:

`GL`	Trabajo ETL de AWS Glue
`AT`	Athena
`RE`	Amazon Redshift Spectrum

`account-id`

El identificador de la AWS cuenta que solicitó las credenciales de Lake Formation.

`suffix`

Una cadena generada aleatoriamente.

Administrar los permisos entre cuentas mediante ambos AWS Glue y Lake Formation

Es posible otorgar acceso entre cuentas a los recursos del Catálogo de datos y a los datos subyacentes mediante AWS Glue o AWS Lake Formation.

En AWS Glue, se conceden permisos entre cuentas mediante la creación o actualización de una política de recursos del Catálogo de datos. En Lake Formation, concede permisos entre cuentas mediante el modelo de permisos de `GRANT/REVOKE` de Lake Formation y la operación de la API `Grant Permissions`.

> **Tip**
>
> Recomendamos confiar solo en los permisos de Lake Formation para proteger su lago de datos.

Puede ver las subvenciones entre cuentas de Lake Formation mediante la consola Lake Formation o la consola AWS Resource Access Manager (AWS RAM). Sin embargo, esas páginas de la consola no muestran los permisos entre cuentas otorgados por la política de recursos del Catálogo de datos de AWS Glue. Del mismo modo, puede ver las concesiones entre cuentas en la política de recursos del Catálogo de datos mediante la página Configuración de la consola de AWS Glue, pero esa página no muestra los permisos entre cuentas concedidos mediante Lake Formation.

Para garantizar que no pierda ninguna subvención al consultar y gestionar los permisos entre cuentas, Lake Formation y AWS Glue requieren que haga lo siguiente para confirmar que conoce y permite concesiones entre cuentas por Lake Formation y por AWS Glue.

Al conceder permisos para varias cuentas mediante el AWS Glue Política de recursos del catálogo de datos

Si su cuenta (cuenta de cedente o cuenta de productor) no ha realizado concesiones entre cuentas que utilicen AWS RAM para compartir los recursos, puede guardar una política de recursos del Catálogo de datos de la forma habitual en AWS Glue. Sin embargo, si ya se han concedido

concesiones que implican el uso compartido de AWS RAM recursos, debe realizar una de las siguientes acciones para asegurarse de que la política de recursos se guarde correctamente:

- Cuando guarde la política de recursos en la página Configuración de la consola AWS Glue, esta emitirá una alerta indicando que los permisos de la política se sumarán a cualquier permiso concedido mediante la consola Lake Formation. Debe elegir Continuar para guardar la política.
- Al guardar la política de recursos mediante la operación de la API `glue:PutResourcePolicy`, debe establecer el campo `EnableHybrid` en 'TRUE' (tipo = cadena).

  Para actualizar una política de recursos existente, utilice primero la operación de la API `glue:GetResourcePolicy` para recuperar la política actual y, a continuación, modifíquela según sea necesario antes de realizar la llamada a `glue:PutResourcePolicy`.

  > **Note**
  >
  > Al crear políticas AWS Glue de recursos para el acceso entre cuentas, conceda solo los permisos mínimos necesarios para su caso de uso específico.

  Para obtener más información, consulte PutResourcePolicy Action (Python: put_resource_policy) en la Guía para desarrolladores.AWS Glue

Al conceder permisos entre cuentas mediante el método de recursos con nombre de Lake Formation

Si no existe una política de recursos del Catálogo de datos en su cuenta (cuenta de productor), las concesiones entre cuentas de Lake Formation que efectúe procederán como de costumbre. Sin embargo, si existe una política de recursos del Catálogo de datos, debe agregarle la siguiente instrucción para que las concesiones entre cuentas se hagan correctamente si se sigue el método de recurso indicado. *<region>*Sustitúyalo por un nombre de región válido y *<account-id>* por tu ID de cuenta (ID de AWS cuenta de productor).

```
  {
    "Effect": "Allow",
    "Action": [
      "glue:ShareResource"
    ],
    "Principal": {"Service": [
      "ram.amazonaws.com"
    ]},
```

```
    "Resource": [
      "arn:aws:glue:<region>:<account-id>:table/*/*",
      "arn:aws:glue:<region>:<account-id>:database/*",
      "arn:aws:glue:<region>:<account-id>:catalog"
    ]
  }
```

Sin esta declaración adicional, la subvención de Lake Formation es válida, pero queda bloqueada y la cuenta receptora no puede acceder al recurso otorgado. AWS RAM

Important

Si utiliza el método de control de acceso basado en etiquetas (LF-TBAC) de Lake Formation para conceder concesiones entre cuentas, debe tener una política de recursos del catálogo de datos con al menos los permisos especificados en. Requisitos previos

Consulte también:

- Control de acceso a los metadatos(para analizar el método de recursos con nombre asignado frente al método de control de acceso (LF-TBAC) basado en etiquetas de Lake Formation).
- Visualización de tablas y bases de datos compartidas del Catálogo de datos
- Trabajo con la configuración del Catálogo de datos en la consola de AWS Glue en la Guía para desarrolladores de AWS Glue
- Concesión de Cross-Account acceso en la guía para AWS Glue desarrolladores (para ver ejemplos de las políticas de recursos del catálogo de datos)

Visualización de todas las concesiones entre cuentas mediante la operación de GetResourceShares API

Si su empresa concede permisos multicuenta mediante una política de AWS Glue Data Catalog recursos y subvenciones de Lake Formation, la única forma de ver todas las concesiones multicuenta en un solo lugar es mediante la operación de `glue:GetResourceShares` API.

Cuando concedes permisos a Lake Formation en todas las cuentas mediante el método de recurso designado, AWS Resource Access Manager (AWS RAM) crea una política de recursos AWS Identity and Access Management (IAM) y la almacena en tu AWS cuenta. La política otorga los permisos necesarios para acceder al recurso. AWS RAM crea una política de recursos independiente para cada concesión multicuenta. Puede ver todas estas políticas mediante la operación de la API `glue:GetResourceShares`.

> **Note**
>
> Esta operación también devuelve la política de recursos del Catálogo de datos. Sin embargo, si ha activado el cifrado de metadatos en la configuración del catálogo de datos y no tiene permiso para utilizar la AWS KMS clave, la operación no devolverá la política de recursos del catálogo de datos.

Para ver todas las concesiones entre cuentas

- Introduzca el siguiente AWS CLI comando.

```
aws glue get-resource-policies
```

El siguiente es un ejemplo de política de recursos que AWS RAM crea y almacena cuando se conceden permisos sobre una tabla `t` de la base de datos `db1` a la AWS cuenta 1111-2222-3333.

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
         "glue:GetTable",
         "glue:GetTables",
         "glue:GetTableVersion",
         "glue:GetTableVersions",
         "glue:GetPartition",
         "glue:GetPartitions",
         "glue:BatchGetPartition",
```

```
          "glue:SearchTables"
        ],
       "Principal": {"AWS": [
         "111122223333"
       ]},
       "Resource": [
       "arn:aws:glue:us-east-1:111122223333:table/db1/t"
      ]
     }
   ]
}
```

Véase también:

- GetResourceShares Acción (Python: get_resource_policies) en la Guía para desarrolladores AWS Glue

Acceso y visualización de tablas y bases de datos compartidas del Catálogo de datos

Para el administrador del lago de datos y para los directores a los que se les han concedido permisos, los recursos que se comparten con su AWS cuenta aparecen en el catálogo de datos como si fueran recursos de su cuenta. La consola muestra la cuenta propietaria del recurso.

Puede ver los recursos compartidos con su cuenta utilizando la consola de Lake Formation. También puede usar la consola AWS Resource Access Manager (AWS RAM) para ver los recursos que se comparten con su cuenta y los recursos que ha compartido con otras AWS cuentas mediante el método de recurso indicado.

Important

Cuando alguien usa el método de recurso con nombre asignado para conceder permisos entre cuentas en un recurso del catálogo de datos a su cuenta u AWS organización, Lake Formation usa el servicio AWS Resource Access Manager (AWS RAM) para compartir el recurso. Si su cuenta pertenece a la misma AWS organización que la cuenta que la concede, tendrá a su disposición el recurso compartido de forma inmediata.

Sin embargo, si su cuenta no pertenece a la misma organización, AWS RAM envía una invitación a su cuenta para que acepte o rechace el recurso compartido. Luego, para que el recurso compartido esté disponible, el administrador del lago de datos de su cuenta debe usar la AWS RAM consola o la CLI para aceptar la invitación.
La consola de Lake Formation muestra una alerta si hay una invitación para compartir AWS RAM recursos esperando ser aceptada. Solo los usuarios autorizados a ver AWS RAM las invitaciones reciben la alerta.

Consulte también:

- Compartir tablas y bases de datos del catálogo de datos entre AWS Cuentas
- Cross-account intercambio de datos en Lake Formation
- Acceso a los datos subyacentes de una tabla compartida
- Control de acceso a los metadatos(para obtener información sobre el método de recurso designado en comparación con el LF-TBAC método para compartir recursos).

Temas

- Aceptar una invitación para compartir recursos de AWS RAM
- Visualización de tablas y bases de datos compartidas del Catálogo de datos

Aceptar una invitación para compartir recursos de AWS RAM

Si un recurso del catálogo de datos se comparte con su AWS cuenta y esta no pertenece a la misma AWS organización que la cuenta compartida, no tendrá acceso al recurso compartido hasta que acepte una invitación para compartir recursos de AWS Resource Access Manager (AWS RAM). Como administrador de un lago de datos, primero debe consultar AWS RAM las invitaciones pendientes y, a continuación, aceptarlas.

Puede usar la AWS RAM consola, la API o AWS Command Line Interface (AWS CLI) para ver y aceptar las invitaciones.

Para ver y aceptar una invitación para compartir un recurso, haga lo siguiente: AWS RAM (consola)

1. Asegúrese de tener los permisos AWS Identity and Access Management (IAM) necesarios para ver y aceptar las invitaciones a compartir recursos.

   Para obtener información sobre las políticas de IAM sugeridas para los administradores del lago de datos, consulte the section called “Permisos de administrador del lago de datos”.
2. Siga las instrucciones de Aceptar y rechazar invitaciones en la Guía del usuario de AWS RAM .

Para ver y aceptar una invitación para compartir recursos de AWS RAM (AWS CLI)

1. Asegúrese de tener los permisos AWS Identity and Access Management (IAM) necesarios para ver y aceptar las invitaciones a compartir recursos.

   Para obtener información sobre las políticas de IAM sugeridas para los administradores del lago de datos, consulte the section called “Permisos de administrador del lago de datos”.
2. Introduzca el siguiente comando para ver las invitaciones para compartir recursos pendientes.

   ```
   aws ram get-resource-share-invitations
   ```

   El resultado debería ser similar al siguiente.

   ```
   {
       "resourceShareInvitations": [
           {
               "resourceShareInvitationArn": "arn:aws:ram:us-
   east-1:111122223333:resource-share-invitation/a93aa60a-1bd9-46e8-96db-
   a4e72eec1d9f",
               "resourceShareName": "111122223333-123456789012-uswuU",
               "resourceShareArn": "arn:aws:ram:us-east-1:111122223333:resource-
   share/2a4ab5fb-d859-4751-84f7-8760b35fc1fe",
               "senderAccountId": "111122223333",
               "receiverAccountId": "123456789012",
               "invitationTimestamp": 1589576601.79,
               "status": "PENDING"
           }
       ]
   }
   ```

   Observe el estado de `PENDING`.

3. Copie el valor de la clave `resourceShareInvitationArn` en el portapapeles.
4. Pegue el valor en el siguiente comando*<invitation-arn>*, reemplácelo e introduzca el comando.

```
aws ram accept-resource-share-invitation --resource-share-invitation-
arn <invitation-arn>
```

El resultado debería ser similar al siguiente.

```
{
    "resourceShareInvitations": [
        {
            "resourceShareInvitationArn": "arn:aws:ram:us-
east-1:111122223333:resource-share-invitation/a93aa60a-1bd9-46e8-96db-
a4e72eec1d9f",
            "resourceShareName": "111122223333-123456789012-uswuU",
            "resourceShareArn": "arn:aws:ram:us-east-1:111122223333:resource-
share/2a4ab5fb-d859-4751-84f7-8760b35fc1fe",
            "senderAccountId": "111122223333",
            "receiverAccountId": "123456789012",
            "invitationTimestamp": 1589576601.79,
            "status": "ACCEPTED"
        }
    ]
}
```

Observe el estado de `ACCEPTED`.

Visualización de tablas y bases de datos compartidas del Catálogo de datos

Puede ver los recursos que se comparten con su cuenta utilizando la consola de Lake Formation o la CLI de AWS . También puedes usar la consola AWS Resource Access Manager (AWS RAM) o la CLI para ver los recursos que se comparten con tu cuenta y los recursos que has compartido con otras AWS cuentas.

Para ver los recursos compartidos utilizando la consola de Lake Formation

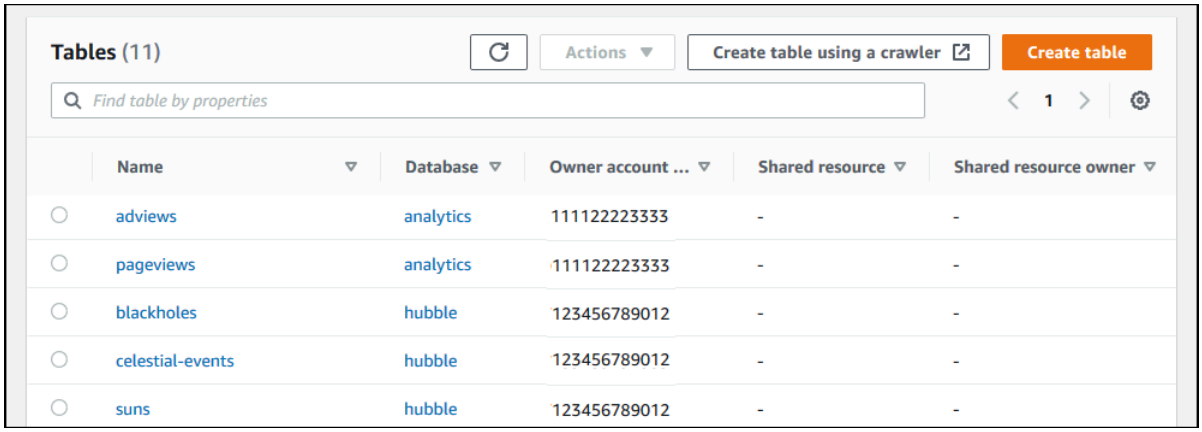
1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.

Inicie sesión como administrador del lago de datos o como usuario al que se han concedido permisos sobre una tabla compartida.

2. Para ver los recursos que se comparten con su AWS cuenta, realice una de las siguientes acciones:

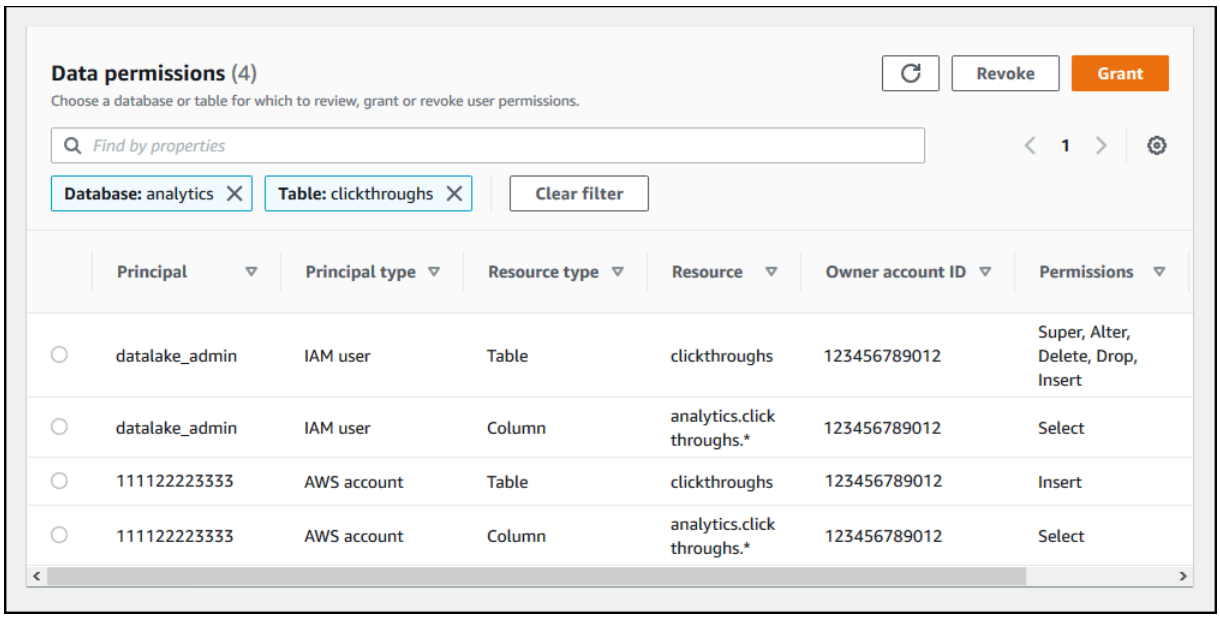
   - Para ver las tablas que se comparten con su cuenta, en el panel de navegación, seleccione Tablas.
   - Para ver las base de datos que se comparten con su cuenta, en el panel de navegación, seleccione Bases de datos.

   La consola muestra una lista de las bases de datos o tablas tanto de su cuenta como compartidas con ella. En el caso de los recursos que se comparten con tu cuenta, la consola muestra el ID de la cuenta AWS del propietario en la columna ID de la cuenta del propietario (la tercera columna de la siguiente captura de pantalla).



3. Para ver los recursos que has compartido con otras AWS cuentas u organizaciones, en el panel de navegación, selecciona Permisos de datos.

   Los recursos que ha compartido aparecen en la página Permisos de datos con el número de cuenta externa en la columna Entidad principal, como se ve en la imagen siguiente.



Para ver los recursos compartidos mediante el AWS RAM consola

1. Asegúrese de tener los permisos AWS Identity and Access Management (de IAM) necesarios para ver los recursos compartidos que utiliza AWS RAM.

   Como mínimo, debe tener el permiso `ram:ListResources`. Este permiso está incluido en la política administrada `AWSLakeFormationCrossAccountManager` de AWS.
2. Inicie sesión en Consola de administración de AWS y abra la AWS RAM consola en https://console.aws.amazon.com/ram/home.
3. Realice una de las siguientes acciones:

   - Para ver los recursos que ha compartido, en el panel de navegación, en Compartidos por mí, seleccione Recursos compartidos.
   - Para ver los recursos que ha compartido, en el panel de navegación, en Compartidos por mí, seleccione Recursos compartidos.

Creación de enlaces de recursos

Los enlaces de recursos son objetos del catálogo de datos que son enlaces a bases de datos y tablas de metadatos, normalmente a bases de datos y tablas compartidas de otras AWS cuentas. Ayudan a permitir el acceso entre cuentas a los datos del lago de datos en todas las regiones. AWS

> **Note**
>
> Lake Formation permite consultar las tablas del catálogo de datos en todas las AWS regiones. Puede acceder a las bases de datos y tablas del catálogo de datos desde cualquier AWS región creando enlaces de recursos en esas regiones que apunten a bases de datos y tablas compartidas en diferentes regiones.

Temas

- Cómo funcionan los enlaces de recursos en Lake Formation
- Crear un enlace de recursos a una tabla de Catálogo de datos compartida
- Crear un enlace de recursos a una base de datos de Catálogo de datos compartida
- Gestión de enlaces de recursos en AWS Glue API

Cómo funcionan los enlaces de recursos en Lake Formation

Un enlace de recursos es un objeto de Catálogo de datos que actúa como enlace con una tabla o base de datos local o compartida. Después de crear un enlace de recursos a una tabla o una base de datos, puede emplear el nombre de dicho enlace donde vaya a utilizar el nombre de la base de datos. `glue:GetTables()` devuelve los enlaces de recursos de tablas y las tablas, ya sean suyas o compartidas, y aparecen como entradas en la página Tablas de la consola de Lake Formation. Los enlaces de recursos a bases de datos actúan de manera similar.

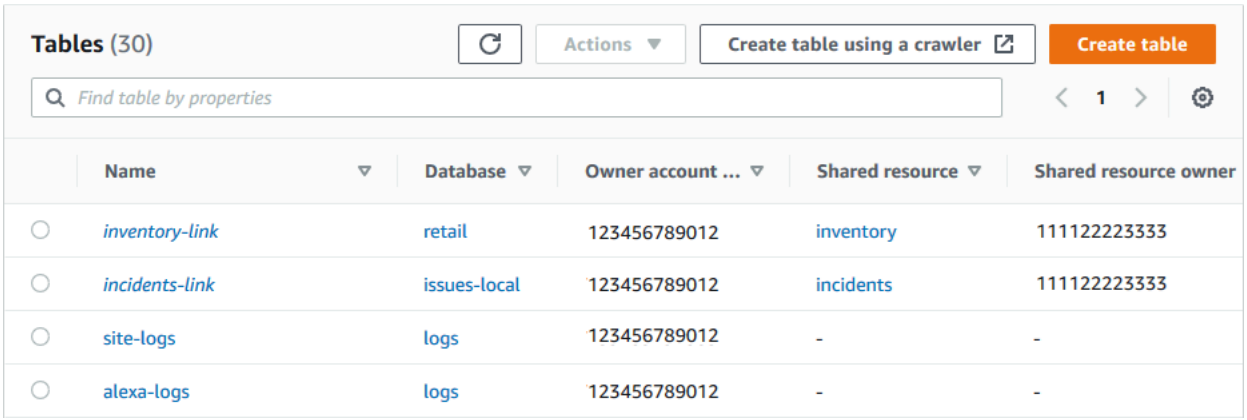
Al crear un enlace de recursos a una base de datos o tabla podrá:

- Asignar un nombre diferente a una base de datos o tabla de su Catálogo de datos. Esto resulta especialmente útil si AWS cuentas diferentes comparten bases de datos o tablas con el mismo nombre, o si varias bases de datos de su cuenta tienen tablas con el mismo nombre.
- Acceda a las bases de datos y tablas del catálogo de datos desde cualquier AWS región creando enlaces de recursos en esas regiones que apunten a la base de datos y las tablas de otra región. Puede ejecutar consultas en cualquier región con estos enlaces de recursos mediante Athena, Amazon EMR y ejecutar trabajos de ETL Spark de AWS Glue , sin copiar los datos de origen ni los metadatos del Glue Data Catalog.
- Utilice AWS servicios integrados como Amazon Athena Amazon Redshift Spectrum para ejecutar consultas que accedan a tablas o bases de datos compartidas. Algunos servicios integrados no

pueden acceder directamente a las bases de datos o tablas entre cuentas. Sin embargo, pueden acceder a los enlaces de recursos presentes en su cuenta a las bases de datos y tablas de otras cuentas.

> **Note**
>
> No es necesario crear un enlace de recursos para hacer referencia a una base de datos o tabla compartida en los scripts de extracción, transformación y carga (ETL) de AWS Glue. Sin embargo, para evitar la ambigüedad cuando varias cuentas de AWS comparten una base de datos o una tabla con el mismo nombre, puede crear y utilizar un enlace de recursos o especificar el identificador del catálogo al invocar las operaciones de ETL.

En el ejemplo siguiente, se muestra la página Tablas de la consola de Lake Formation, que muestra dos enlaces de recursos. Los nombres de los enlaces de recursos siempre se muestran en cursiva. Cada enlace de recursos se muestra junto con el nombre y el propietario del recurso compartido vinculado. En este ejemplo, un administrador de un lago de datos de la AWS cuenta 1111-2222-3333 compartió las tablas `inventory` y `incidents` con la cuenta 1234-5678-9012. A continuación, un usuario de esa cuenta creó enlaces de recursos a esas tablas compartidas.



Las siguientes son notas y restricciones sobre los enlaces de recursos:

- Los enlaces de recursos son necesarios para permitir que los servicios integrados, como Athena y Redshift Spectrum, consulten los datos subyacentes de las tablas compartidas. Las consultas en estos servicios integrados se crean a partir de los nombres de los enlaces de recursos.
- Si asumimos que la configuración para usar solo el control de acceso de IAM para las nuevas tablas de esta base de datos esté desactivada en la base de datos que la contiene, solo la entidad principal que creó un enlace de recursos podrá verla y acceder a ella. Para que otras

entidades principales de su cuenta puedan acceder a un enlace de recursos, conceda el permiso de `DESCRIBE` correspondiente. Para que otros usuarios puedan eliminar un enlace de recursos, concédales el permiso de `DROP`. Los administradores del lago de datos pueden acceder a todos los enlaces de recursos de la cuenta. Para eliminar un enlace de recurso creado por otra entidad principal, el administrador del lago de datos primero debe concederse el permiso de `DROP` para acceder al enlace de recurso. Para obtener más información, consulte Referencia de permisos de Lake Formation.

> **Important**
>
> Conceder permisos en un enlace de recursos no otorga permisos en la tabla o base de datos (vinculada) de destino. Debe conceder los permisos en el destino por separado.

- Para crear un enlace a un recurso, necesita el `CREATE_DATABASE` permiso `CREATE_TABLE` o Lake Formation, así como el permiso `glue:CreateTable` or `glue:CreateDatabase` AWS Identity and Access Management (IAM).
- Puede crear enlaces de recursos a los recursos del catálogo de datos local (de su propiedad), así como a los recursos compartidos con su AWS cuenta.
- Al crear un enlace de recursos, no se comprueba si el recurso compartido de destino existe ni si tiene permisos entre cuentas sobre el recurso. Esto le permite crear el enlace de recurso y el recurso compartido en cualquier orden.
- Si elimina un enlace de recurso, el recurso compartido vinculado no se elimina. Si elimina un recurso compartido, los enlaces de recursos a ese recurso no se eliminan.
- Es posible crear cadenas de enlaces de recursos. Sin embargo, no tiene sentido hacerlo, ya que las API solo siguen el primer enlace de recursos.

> **Véase también:**
>
> - Concesión de permisos sobre los recursos del Catálogo de datos

Crear un enlace de recursos a una tabla de Catálogo de datos compartida

Puedes crear un enlace de recursos a una tabla compartida en cualquier AWS región mediante la AWS Lake Formation consola, la API o AWS Command Line Interface (AWS CLI).

Para crear un enlace de recursos a una tabla compartida (consola)

1. Abre la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/. Inicie sesión como una entidad principal que tenga el permiso de `CREATE_TABLE` de Lake Formation en la base de datos que contiene el enlace de recursos.
2. En el panel de navegación, elija Tablas en Catálogo de datos y, a continuación, elija Crear, Enlace de recursos.
3. En la página Crear enlace de recursos, proporcione la siguiente información:

   Nombre del enlace de recursos

   Introduzca un nombre que cumpla las mismas reglas que el nombre de una tabla. El nombre puede ser el mismo que el de la tabla compartida de destino.

   Base de datos

   La base de datos del catálogo local que contiene el enlace de recursos.

   Región del propietario de la tabla

   Si va a crear el enlace de recursos en una región diferente, seleccione la región de la tabla compartida de destino.

   Tabla compartida

   Seleccione una tabla compartida de la lista o introduzca un nombre de tabla local (propia) o compartida.

   La lista contiene todas las tablas compartidas con su cuenta. Anote la base de datos y el ID de la cuenta del propietario que aparecen en cada tabla. Si no se muestra una tabla que sabe que se ha compartido con su cuenta, verifique lo siguiente:

   - Si no es administrador de un lago de datos, compruebe que el administrador del lago de datos le haya concedido los permisos de Lake Formation sobre la tabla.
   - Si es administrador de un lago de datos y su cuenta no pertenece a la misma organización de AWS que la cuenta que los concede, asegúrese de haber aceptado la invitación para compartir recursos AWS Resource Access Manager (AWS RAM) de la tabla. Para obtener más información, consulte Aceptar una invitación para compartir recursos de AWS RAM.

   Base de datos de la tabla compartida

   Si ha seleccionado una tabla compartida de la lista, este campo se rellena con la base de datos de la tabla compartida en la cuenta externa. De lo contrario, introduzca una base de

datos local (para un enlace de recursos a una tabla local) o la base de datos de la tabla compartida en la cuenta externa.

Propietario de tabla compartida

Si ha seleccionado una tabla compartida de la lista, este campo se rellena con el ID de la cuenta del propietario de la tabla compartida. De lo contrario, introduce tu ID de AWS cuenta (para un enlace de recursos a una tabla local) o el ID de la AWS cuenta que compartió la tabla.

4. Elija Crear para crear el enlace de recursos.

   A continuación, puede ver el nombre del enlace de recursos en la columna Nombre de la página Tablas.

5. (Opcional) Otorgue el permiso de `DESCRIBE` de Lake Formation en el enlace de recursos a las entidades principales que deben poder ver el enlace y acceder a la tabla de destino.

   No obstante, conceder permisos en un enlace a un recurso no otorga permisos en la tabla o base de datos (vinculada) de destino. Debe conceder permisos en la base de datos de destino por separado para que el table/resource enlace esté visible en Athena.

Para crear un enlace de recursos a una tabla compartida en la misma región (AWS CLI)

1. Introduzca un comando similar al siguiente.

```
aws glue create-table --database-name myissues --table-input
 '{"Name":"my_customers","TargetTable":
{"CatalogId":"111122223333","DatabaseName":"issues","Name":"customers"}}'
```

   Este comando crea un enlace de recursos con el nombre `my_customers` a la tabla compartida `customers`, que está en la base de datos `issues` de la cuenta de AWS 1111-2222-3333. El enlace de recursos se almacena en la base de datos local `myissues`.

2. (Opcional) Otorgue el permiso de `DESCRIBE` de Lake Formation en el enlace de recursos a las entidades principales que deben poder ver el enlace y acceder a la tabla de destino.

   No obstante, conceder permisos en un enlace a un recurso no otorga permisos en la tabla (vinculada) de destino. Debe conceder permisos en la base de datos de destino por separado para que el table/resource enlace esté visible en Athena.

Para crear un enlace de recursos a una tabla compartida en una región diferente (AWS CLI)

1. Introduzca un comando similar al siguiente.

```
aws glue create-table --region eu-west-1 --cli-input-json '{
    "CatalogId": "111122223333",
    "DatabaseName": "ireland_db",
    "TableInput": {
        "Name": "rl_useast1salestb_ireland",
        "TargetTable": {
            "CatalogId": "444455556666",
            "DatabaseName": "useast1_salesdb",
            "Region": "us-east-1",
            "Name":"useast1_salestb"
        }
    }
}‘
```

   Este comando crea un enlace de recurso denominado `rl_useast1salestb_ireland` en la región Europa (Irlanda) a la `useast1_salestb` tabla compartida, que se encuentra en la `useast1_salesdb` base de datos de AWS la cuenta 444455556666 de la región EE.UU. Este (Norte de Virginia). El enlace de recursos se almacena en la base de datos local `ireland_db`.

2. Conceda permiso de `DESCRIBE` de Lake Formation a las entidades principales que deben poder ver el enlace y acceder al destino del enlace a través del enlace.

   No obstante, conceder permisos en un enlace a un recurso no otorga permisos en la tabla (vinculada) de destino. Debe conceder permisos en la tabla de destino por separado para que el table/resource enlace esté visible en Athena.

Véase también:

- Cómo funcionan los enlaces de recursos en Lake Formation
- `DESCRIBE`

Crear un enlace de recursos a una base de datos de Catálogo de datos compartida

Puede crear un enlace de recursos a una base de datos compartida mediante la AWS Lake Formation consola, la API o AWS Command Line Interface (AWS CLI).

Para crear un enlace de recursos a una base de datos compartida (consola)

1. Abre la AWS Lake Formation consola en https://console.aws.amazon.com/lakeformation/. Inicie sesión como administrador de un lago de datos o como creador de bases de datos.

   El creador de una base de datos es una entidad principal a la que se le ha otorgado el permiso de `CREATE_DATABASE` de Lake Formation.
2. En el panel de navegación, elija Bases de datos y, a continuación, elija Crear, Enlace de recursos.
3. En la página Crear enlace de recursos, proporcione la siguiente información:

   Nombre del enlace de recursos

   Introduzca un nombre que cumpla las mismas reglas que el nombre de una base de datos. El nombre puede ser el mismo que el de la base de datos compartida de destino.

   Catálogo de destino

   Seleccione el catálogo de destino para el enlace de recursos de la base de datos.

   Región propietaria de la base de datos compartida

   Si va a crear el enlace de recursos en una región diferente, seleccione la región de la base de datos compartida de destino.

   Base de datos compartida

   Elija una base de datos de la lista o introduzca un nombre de base de datos local (propia) o compartida.

   La lista contiene todas las bases de datos compartidas con su cuenta. Observe el ID de la cuenta del propietario que aparece con cada base de datos. Si no ve una base de datos que sabe que se ha compartido con su cuenta, verifique lo siguiente:

   - Si no es administrador de un lago de datos, compruebe que el administrador del lago de datos le haya concedido los permisos de Lake Formation sobre la base de datos.

- Si es administrador de un lago de datos y su cuenta no pertenece a la misma organización AWS que la cuenta concedente, asegúrese de haber aceptado la invitación para compartir recursos AWS Resource Access Manager (AWS RAM) de la base de datos. Para obtener más información, consulte Aceptar una invitación para compartir recursos de AWS RAM.

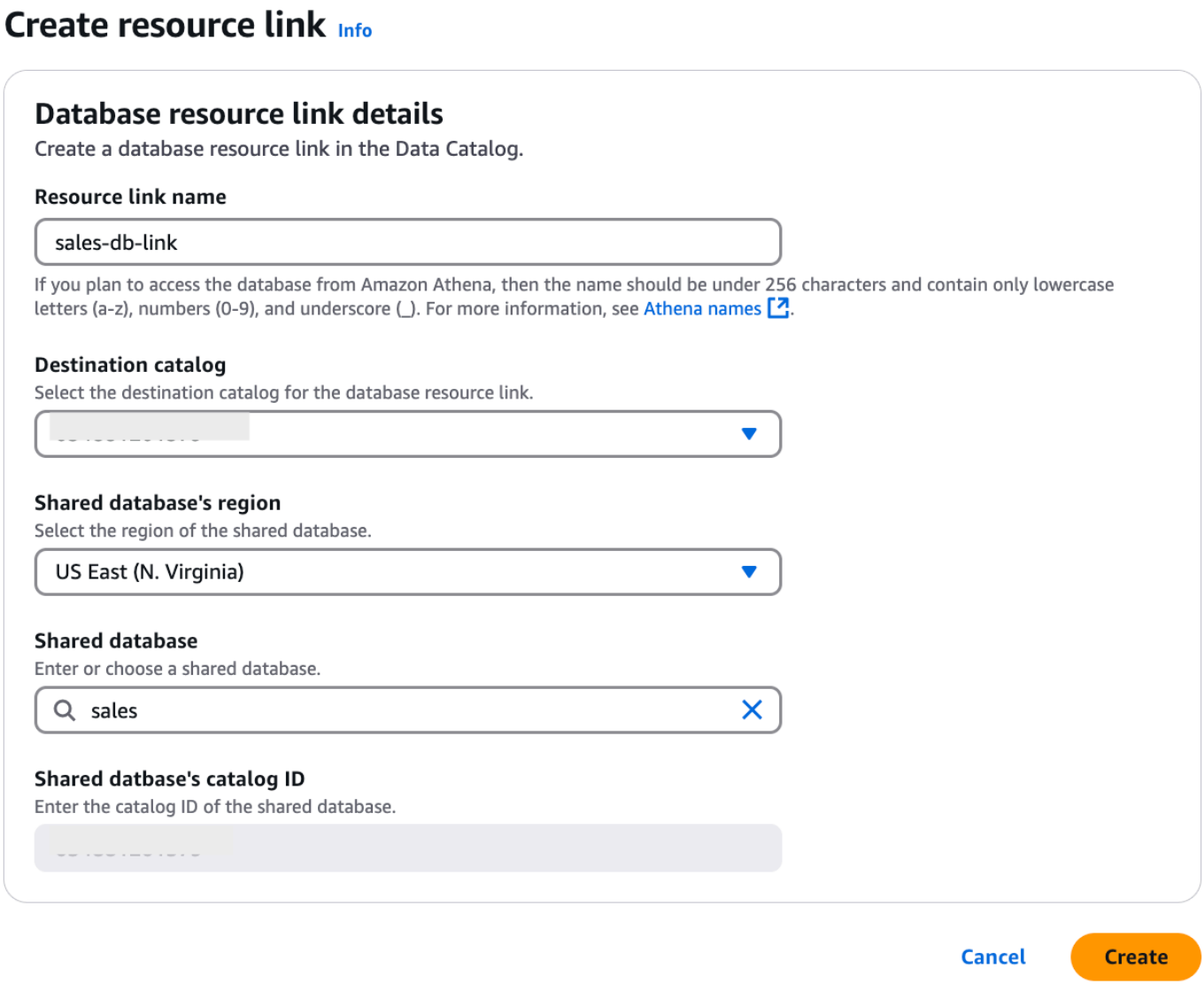
Propietario de la base de datos compartida

Si ha seleccionado una base de datos compartida de la lista, este campo se rellena con el ID de la cuenta del propietario de la base de datos compartida. De lo contrario, introduzca el ID de su AWS cuenta (para un enlace de recursos a una base de datos local) o el ID de la AWS cuenta que compartió la base de datos.

ID de catálogo de la base de datos compartida

Introduzca el ID de catálogo de la base de datos compartida. Al crear un enlace de recursos a una base de datos compartida desde otra cuenta de AWS , debe especificar este ID de catálogo para identificar qué Catálogo de datos de la cuenta contiene la base de datos de origen.

Al seleccionar una base de datos compartida en el menú desplegable, el sistema rellena automáticamente el ID de catálogo de la cuenta propietaria de esa base de datos y que la ha compartido con usted.



4. Elija Crear para crear el enlace de recursos.

   A continuación, puede ver el nombre del enlace de recursos en la columna Nombre de la página Bases de datos.

5. (Opcional) Conceda el permiso de `DESCRIBE` de Lake Formation en el enlace de recursos a las entidades principales de la región de Europa (Irlanda) que deben poder ver el enlace y acceder a la base de datos de destino.

   No obstante, conceder permisos en un enlace a un recurso no otorga permisos en la tabla o base de datos (vinculada) de destino. Debe conceder permisos en la base de datos de destino por separado para que el table/resource enlace esté visible en Athena.

Para crear un enlace de recursos a una base de datos compartida en la misma región (AWS CLI)

1. Introduzca un comando similar al siguiente.

```
aws glue create-database --database-input '{"Name":"myissues","TargetDatabase":
{"CatalogId":"111122223333","DatabaseName":"issues"}}'
```

   Este comando crea un enlace de recursos con el nombre `myissues` de la base de datos compartida`aissues`, que se encuentra en la AWS cuenta 1111-2222-3333.

2. (Opcional) Otorgue el permiso `DESCRIBE` de Lake Formation a las entidades principales del enlace de recursos para que puedan ver el enlace y acceder a la tabla o base de datos de destino.

   No obstante, conceder permisos en un enlace a un recurso no otorga permisos en la tabla o base de datos (vinculada) de destino. Debe conceder permisos en la base de datos de destino por separado para que el table/resource enlace esté visible en Athena.

Para crear un enlace de recursos a una base de datos compartida en una región diferente (AWS CLI)

1. Introduzca un comando similar al siguiente.

```
aws glue create-database --region eu-west-1 --cli-input-json '{
    "CatalogId": "111122223333",
    "DatabaseInput": {
      "Name": "rl_useast1shared_irelanddb",
      "TargetDatabase": {
          "CatalogId": "444455556666",
          "DatabaseName": "useast1shared_db",
          "Region": "us-east-1"
       }
    }
}'
```

   Este comando crea un enlace de recursos denominado `rl_useast1shared_irelanddb` en la AWS cuenta 111122223333 de la región Europa (Irlanda) a la base de datos compartida`useast1shared_db`, que se encuentra en la AWS cuenta 444455556666 de la región EE.UU. Este (Norte de Virginia).

2. Conceda el permiso de `DESCRIBE` de Lake Formation a las entidades principales de la región de Europa (Irlanda) que deben poder ver el enlace y acceder al destino del enlace a través del enlace.

Véase también:

- Cómo funcionan los enlaces de recursos en Lake Formation
- `DESCRIBE`

Gestión de enlaces de recursos en AWS Glue API

En las tablas siguientes se explica cómo las API del Catálogo de datos de AWS Glue gestionan los enlaces a los recursos de bases de datos y tablas. Para todas las operaciones de la API de `Get*`, solo se devuelven las bases de datos y las tablas sobre las que la persona que llama tiene permiso. Además, al acceder a una base de datos o tabla de destino a través de un enlace de recursos, debe tener los permisos AWS Identity and Access Management (IAM) y Lake Formation tanto en el enlace de destino como en el enlace de recursos. El permiso de Lake Formation que se requiere en los enlaces de recursos es `DESCRIBE`. Para obtener más información, consulte `DESCRIBE`.

Operaciones de la API de base de datos

Operación de la API	Gestión de enlaces de recursos
`CreateDatabase`	Si la base de datos es un enlace de recursos, crea el enlace de recursos a la base de datos de destino designada.
`UpdateDatabase`	Si la base de datos es un enlace de recursos, sigue el enlace y actualiza la base de datos de destino. Si el enlace de recursos se debe modificar para enlazarlo a una base de datos diferente, debe eliminarlo y crear uno nuevo.
`DeleteDatabase`	Elimina el enlace de recursos. No elimina la base de datos vinculada (de destino).

Operación de la API	Gestión de enlaces de recursos
`GetDatabase`	Si la persona que llama tiene permisos sobre el destino, sigue el enlace para obtener las propiedades del destino. De lo contrario, devuelve las propiedades del enlace.
`GetDatabases`	Devuelve una lista de bases de datos, incluidos los enlaces de recursos. Para cada enlace de recursos del conjunto de resultado s, la operación sigue el enlace para obtener las propiedades del destino del enlace. Debe especificar `ResourceShareType = ALL` para ver las bases de datos compartidas con su cuenta.

Operaciones de la API de tabla

Operación de la API	Gestión de enlaces de recursos
`CreateTable`	Si la base de datos es un enlace de recursos, sigue el enlace de la base de datos y crea una tabla en la base de datos de destino. Si la tabla es un enlace de recursos, la operación crea el enlace de recursos en la base de datos designada. No es compatible crear un enlace de recursos de tabla a través de un enlace de recursos de base de datos.
`UpdateTable`	Si la tabla o la base de datos designada es un enlace de recursos, actualiza la tabla de destino. Si tanto la tabla como la base de datos son enlaces de recursos, la operación falla.
`DeleteTable`	Si la base de datos designada es un enlace de recursos, sigue el enlace y borra la tabla o el enlace de recursos de tabla en la base de datos de destino. Si la tabla es un enlace de recursos, la operación borra el enlace de recursos de la tabla en la base de datos designada. La eliminación de un enlace de recursos de tabla no elimina la tabla de destino.
`BatchDeleteTable`	Igual que `DeleteTable` .
`GetTable`	Si la base de datos designada es un enlace de recursos, sigue el enlace de la base de datos y devuelve la tabla o el enlace de

Operación de la API	Gestión de enlaces de recursos
	recursos de la tabla de la base de datos de destino. En caso contrario, si la tabla es un enlace de recursos, la operación sigue el enlace y devuelve las propiedades de la tabla de destino.
`GetTables`	Si la base de datos designada es un enlace de recursos, sigue el enlace de la base de datos y devuelve las tablas y los enlaces de recursos de tablas de la base de datos de destino. Si la base de datos de destino es una base de datos compartida de otra AWS cuenta, la operación devuelve solo las tablas compartidas de esa base de datos. No sigue los enlaces de recursos de la tabla en la base de datos de destino. De lo contrario, si la base de datos designada es una base de datos local (propia), la operación devuelve todas las tablas de la base de datos local, y sigue cada enlace de recurso de tabla para devolver las propiedades de la tabla de destino.
`SearchTables`	Devuelve enlaces a recursos de tablas y tablas. No sigue los enlaces para devolver las propiedades de la tabla de destino. Debe especificar `ResourceShareType = ALL` para ver las bases de datos compartidas con su cuenta.
`GetTableVersion`	Igual que `GetTable`.
`GetTableVersions`	Igual que `GetTable`.
`DeleteTableVersion`	Igual que `DeleteTable` .
`BatchDeleteTableVersion`	Igual que `DeleteTable` .

Operaciones de la API en una partición

Operación de la API	Gestión de enlaces de recursos
`CreatePartition`	Si la base de datos designada es un enlace de recursos, sigue el enlace de la base de datos y crea una partición en la tabla designada de la base de datos de destino. Si la tabla es un enlace

Operación de la API	Gestión de enlaces de recursos
	de recursos, la operación sigue el enlace de recursos y crea la partición en la tabla de destino. No es compatible crear una partición mediante un enlace de recursos de tabla y un enlace de recursos de base de datos.
`BatchCreatePartition`	Igual que `CreatePartition` .
`UpdatePartition`	Si la base de datos designada es un enlace de recursos, sigue el enlace de la base de datos y actualiza la partición en la tabla designada de la base de datos de destino. Si la tabla es un enlace de recursos, la operación sigue el enlace de recursos y actualiza la partición en la tabla de destino. No es compatible actualizar una partición mediante un enlace de recursos de tabla y un enlace de recursos de base de datos.
`DeletePartition`	Si la base de datos designada es un enlace de recursos, sigue el enlace de la base de datos y borra la partición en la tabla designada en la base de datos de destino. Si la tabla es un enlace de recursos, la operación sigue el enlace de recursos y borra la partición en la tabla de destino. No es compatible eliminar una partición mediante un enlace de recursos de tabla y un enlace de recursos de base de datos.
`BatchDeletePartition`	Igual que `DeletePartition` .
`GetPartition`	Si la base de datos designada es un enlace de recursos, sigue el enlace de la base de datos y devuelve la información de partición de la tabla designada. En caso contrario, si la tabla es un enlace de recursos, la operación sigue el enlace y devuelve información sobre la partición. Si la tabla y la base de datos son enlaces de recursos, devuelve un conjunto de resultados vacío.

Operación de la API	Gestión de enlaces de recursos
`GetPartitions`	Si la base de datos designada es un enlace de recursos, sigue el enlace de la base de datos y devuelve la información de partición para todas las particiones de la tabla designada. En caso contrario , si la tabla es un enlace de recursos, la operación sigue el enlace y devuelve información sobre la partición. Si la tabla y la base de datos son enlaces de recursos, devuelve un conjunto de resultados vacío.
`BatchGetPartition`	Igual que `GetPartition` .

User-defined funciones, operaciones de API

Operación de la API	Gestión de enlaces de recursos
(Todas las operaciones de la API)	Si la base de datos es un enlace de recursos, sigue el enlace de recursos y efectúa la operación en la base de datos de destino.

Véase también:

- Cómo funcionan los enlaces de recursos en Lake Formation

Acceso a las tablas entre regiones

Lake Formation permite consultar las tablas del catálogo de datos en todas las AWS regiones. Puede acceder a los datos de una región desde otras regiones mediante Amazon Athena, Amazon EMR y AWS Glue ETL creando enlaces de recursos en otras regiones que apunten a las bases de datos y tablas de origen. Con el acceso a las tablas entre regiones, puede acceder a los datos de todas las regiones sin copiar los datos subyacentes o los metadatos en el Catálogo de datos.

Por ejemplo, puede compartir una base de datos o una tabla de una cuenta de productor con una cuenta de consumidor en la Región A. Tras aceptar la invitación a compartir recursos en la Región A, el administrador del lago de datos de la cuenta de consumidor puede crear enlaces de recursos al recurso compartido en la Región B. El administrador de la cuenta de consumidor puede conceder

permisos sobre el recurso compartido a las entidades principales de IAM de esa cuenta en la Región A y conceder permisos de enlace de recursos en la Región B. Con el enlace de recursos, las entidades principales de la cuenta de consumidor pueden consultar los datos compartidos de la Región B.

También puede alojar el origen de datos de Amazon S3 de la región A en una cuenta de productor y registrar la ubicación de los datos en una cuenta central de la región B. Puede crear recursos del Catálogo de datos en la cuenta central, configurar los permisos de Lake Formation y compartir datos con los consumidores de su cuenta o con cuentas externas de la región B. La característica entre regiones permite a los usuarios acceder a estas tablas del Catálogo de datos desde la región C mediante enlaces de recursos.

Con esta característica, puede consultar bases de datos federadas en Apache Hive Metastores entre regiones y también unir tablas de la región local con tablas de otra región al ejecutar consultas.

Lake Formation admite las siguientes características con acceso a tablas entre regiones:

- LF-Tag control de acceso basado
- Fine-grained permisos de control de acceso
- Escribir operaciones en la base de datos o tabla compartida con los permisos adecuados
- Cross-account intercambio de datos a nivel de cuenta y directamente con los directores de IAM

Non-administrative los usuarios con `Create_Table` permisos `Create_Database` y permisos pueden crear enlaces de recursos entre regiones.

> **Note**
>
> Puede crear enlaces de recursos entre regiones en cualquier región y acceder a los datos sin aplicar los permisos de Lake Formation. En el caso de los datos fuente de Amazon S3 que no estén registrados en Lake Formation, el acceso viene determinado por las políticas de permisos y AWS Glue acciones de IAM para Amazon S3.

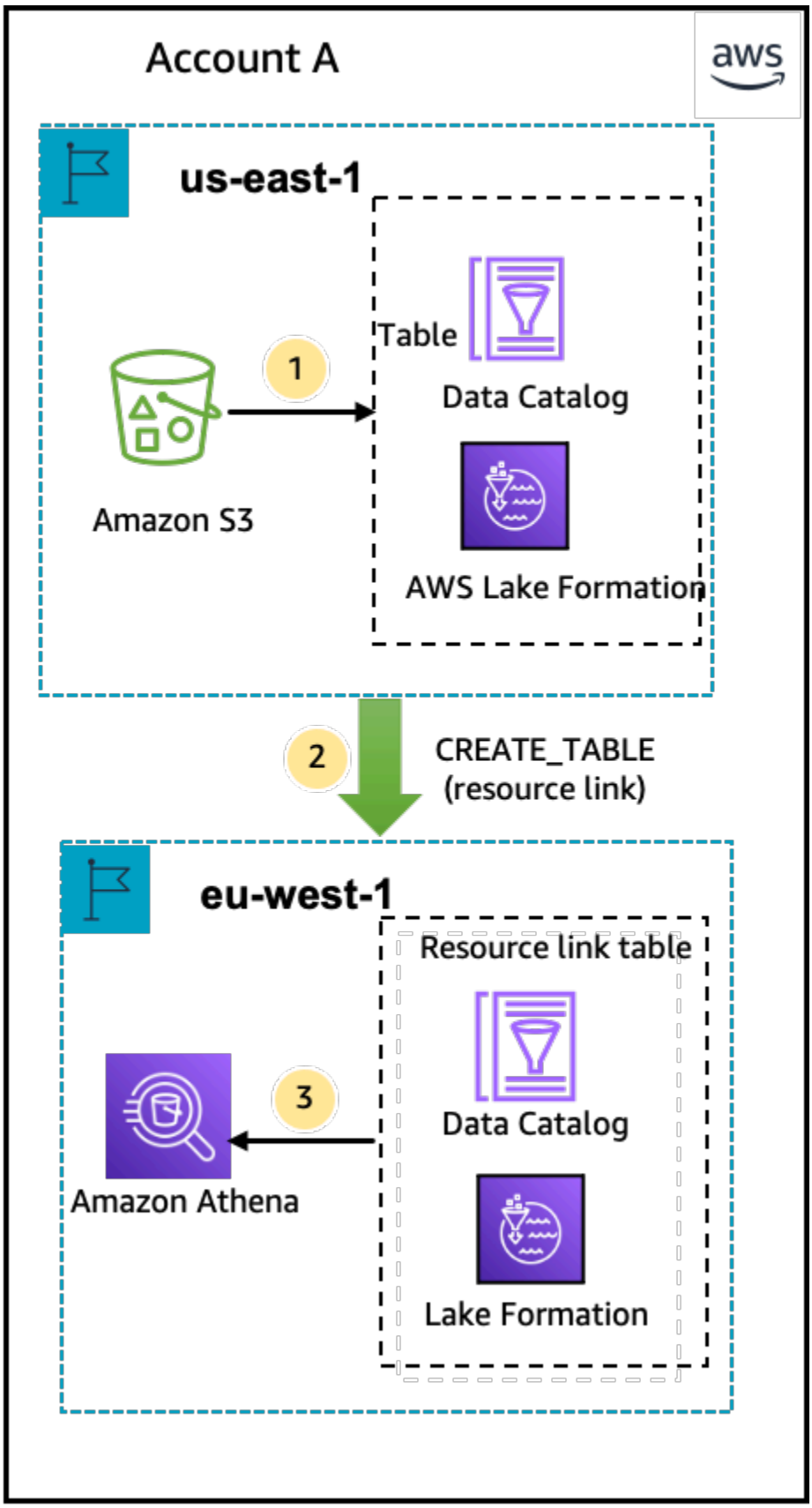
Para conocer las limitaciones, consulte Limitaciones de acceso a datos entre regiones.

Flujos de trabajo

Los siguientes diagramas muestran los flujos de trabajo para acceder a los datos de todas AWS las regiones desde la misma AWS cuenta y desde una cuenta externa.

Flujo de trabajo para acceder a las tablas compartidas dentro de la misma AWS inscrita

En el siguiente diagrama, los datos se comparten con un usuario de la misma AWS cuenta en la región EE.UU. Este (Virginia del Norte) y el usuario consulta los datos compartidos desde la región Europa (Irlanda).


Account A
aws
us-east-1
Table
Data Catalog
1
Amazon S3
AWS Lake Formation
2
CREATE_TABLE
(resource link)
eu-west-1
Resource link table
Data Catalog
3
Amazon Athena
Lake Formation

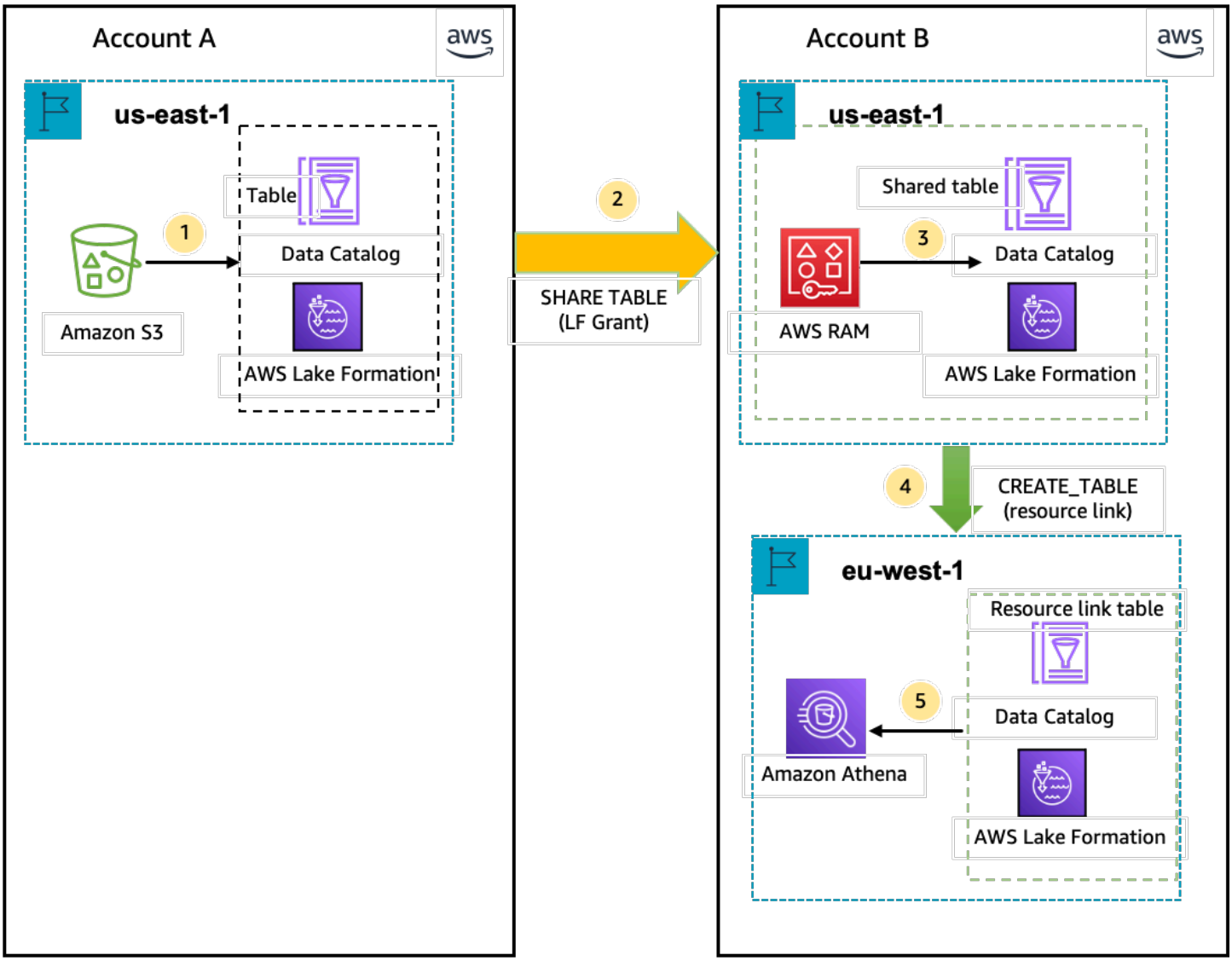
El administrador del lago de datos efectúa las actividades siguientes (pasos 1 y 2):

1. Un administrador de data lake configura una AWS cuenta en las bases de datos y tablas del catálogo de datos y registra una ubicación de datos de Amazon S3 en Lake Formation en la región EE.UU. Este (Norte de Virginia).

   Otorga permiso de `Select` sobre un recurso del Catálogo de datos (tabla de productos en el diagrama) a una entidad principal (usuario) de la misma cuenta.
2. Crea un enlace de recursos en la región Europa (Irlanda) que apunta a la tabla de origen en la región Este de EE. UU. (Norte de Virginia). Otorga permiso de `DESCRIBE` a la entidad principal para utilizar el enlace de recursos de la región de Europa (Irlanda).
3. El usuario consulta la tabla desde la región Europa (Irlanda) mediante Athena.

Flujo de trabajo para acceder a tablas compartidas con un usuario externo AWS inscrita

En el siguiente diagrama, la cuenta del productor (cuenta A) aloja el bucket de Amazon S3, registra la ubicación de los datos y comparte una tabla del Catálogo de datos con una cuenta de consumidor (cuenta B) de la región Este de EE. UU. (Norte de Virginia) y un usuario de la cuenta de consumidor (cuenta B) consulta la tabla de la región de Europa (Irlanda).



1. Un administrador de data lake configura una AWS cuenta (cuenta de productor) con los recursos del catálogo de datos y una ubicación de datos de Amazon S3 registrada en Lake Formation en la región EE.UU. Este (Norte de Virginia).
2. El administrador del lago de datos de la cuenta del productor comparte una tabla del Catálogo de datos con una cuenta de consumidor.
3. El administrador del lago de datos de la cuenta de consumidor acepta la invitación a compartir datos en la región de Este de EE. UU. (Norte de Virginia) y concede el permiso de `Select` para utilizar la tabla compartida con una entidad principal de la misma región.
4. El administrador del lago de datos de la cuenta de consumidor crea un enlace de recursos en la región de Europa (Irlanda) que apunta a la tabla compartida de destino en la región Este de EE.

UU. (Norte de Virginia) y concede al usuario permiso de `DESCRIBE` para acceder al enlace de recursos desde la región de Europa (Irlanda).

5. El usuario consulta la tabla desde la región Europa (Irlanda) mediante Athena.

Configuración del acceso a las tablas entre regiones

Para acceder a los datos de una región diferente, primero debe configurar las bases de datos y las tablas del Catálogo de datos en la región en la que registra su ubicación de datos de Amazon S3. Puede compartir las bases de datos y tablas del Catálogo de datos con las entidades principales de su cuenta o de otra cuenta. A continuación, debe crear administradores de lagos de datos que puedan generar enlaces de recursos que apunten a la ubicación de datos compartidos de destino en las regiones en las que los usuarios consultan los datos.

Para consultar los datos compartidos en la misma cuenta desde una región diferente

En esta sección, la región de la tabla compartida de destino se denomina Región A y los usuarios hacen consultas desde la Región B.

1. Configuración de la cuenta en la Región A (donde se crean y comparten los datos)

   El administrador de un lago de datos debe completar las acciones siguientes:

   a. Registrar una ubicación de datos de Amazon S3.

      Para obtener más información, consulte Añadir una ubicación de Amazon S3 a su lago de datos.

   b. Crear bases de datos y tablas en la cuenta. Esto también lo puede hacer un usuario no administrativo que tenga permisos para crear bases de datos y tablas.

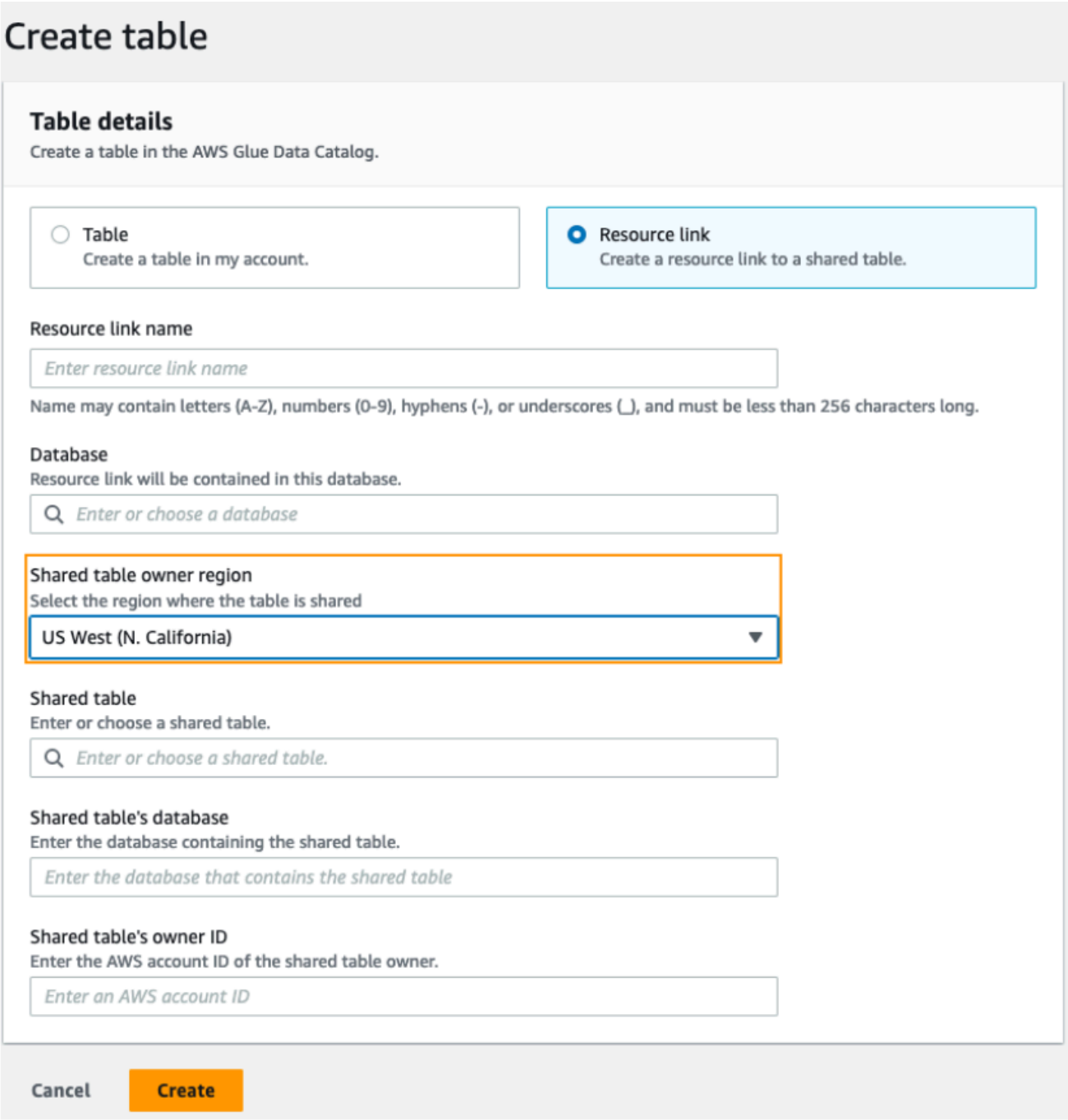
   c. Conceda permisos de datos sobre una tabla a las entidades principales con `Grantable permissions`.

      Para obtener más información, consulte Concesión de permisos sobre los recursos del Catálogo de datos.

2. Configuración de la cuenta en la Región B (donde accede a los datos)

   El administrador del lago de datos debe completar las acciones siguientes:

a. Crear un enlace de recursos en la región B que apunte a la tabla compartida de destino en la región A. Especifique la Región del propietario de la tabla compartida en la pantalla Crear tabla.



Para obtener instrucciones sobre cómo crear enlaces de recursos a bases de datos y tablas, consulte Creación de enlaces de recursos.

b. Conceda permiso de `Describe` a las entidades principales de IAM en el enlace de recursos de la Región B.

   Para obtener más información sobre conceder permisos en enlaces de recursos, consulte Conceder permisos de enlace de recursos.

   Las entidades principales de IAM de la Región B pueden consultar la tabla de destino a través del enlace con Athena.

Para acceder a los datos de varias cuentas de una región diferente

1. Producer/grantor configuración de la cuenta

   El administrador del lago de datos debe completar las acciones siguientes:

   a. Configure la producer/grantor cuenta en la Región A.

   b. Registrar una ubicación de datos de Amazon S3 en la Región A.

   c. Crear bases de datos y tablas. Esto lo puede hacer un usuario no administrativo que tenga permisos para crear bases de datos y tablas.

   d. Otorgue permisos de datos a la consumer/grantee cuenta de una tabla de la región A con`Grantable permissions`.

      Para obtener más información, consulte Compartir tablas y bases de datos del catálogo de datos entre Cuentas de AWS o directores de IAM de cuentas externas.

2. Consumer/grantee configuración de la cuenta

   El administrador del lago de datos debe completar las acciones siguientes:

   a. Acepte la invitación para compartir recursos de AWS RAM la Región A.

   b. Crear un enlace de recursos en la región B que apunte a la tabla compartida. La Región B es donde los usuarios consultarán la tabla.

   c. Conceda permisos de datos en la tabla compartida a las entidades principales de IAM en la Región A.

> **Note**
>
> Debe conceder permisos a la tabla compartida en la misma región en la que se compartió la tabla.

d. Conceda permisos a las entidades principales en el enlace de recursos de la Región B.

   A continuación, las entidades principales de la cuenta de consumidores de la Región B consultan la tabla compartida desde la Región B utilizando Athena.

Seguridad en AWS Lake Formation

La seguridad en la nube AWS es la máxima prioridad. Como AWS cliente, usted se beneficia de una arquitectura de centro de datos y red diseñada para cumplir con los requisitos de las organizaciones más sensibles a la seguridad.

La seguridad es una responsabilidad compartida entre usted AWS y usted. El modelo de responsabilidad compartida la describe como seguridad de la nube y seguridad en la nube:

- Seguridad de la nube: AWS es responsable de proteger la infraestructura que ejecuta AWS los servicios en la AWS nube. AWS también le proporciona servicios que puede utilizar de forma segura. Auditores independientes prueban y verifican periódicamente la eficacia de nuestra seguridad en el marco de los programas de conformidad de AWS. Para obtener más información sobre los programas de cumplimiento que se aplican a AWS Lake Formation, consulte AWS Services in Scope by Compliance Program.
- Seguridad en la nube: su responsabilidad viene determinada por el AWS servicio que utilice. También es responsable de otros factores, incluida la confidencialidad de los datos, los requisitos de la empresa y la legislación y la normativa aplicables.

Esta documentación le ayuda a comprender cómo aplicar el modelo de responsabilidad compartida cuando se utiliza Lake Formation. En los temas siguientes, se le mostrará cómo configurar Lake Formation para satisfacer sus objetivos de seguridad y conformidad. También aprenderá a usar otros AWS servicios que lo ayudan a monitorear y proteger sus recursos de Lake Formation.

Temas

- Protección de datos en Lake Formation
- Seguridad de la infraestructura en AWS Lake Formation
- Prevención de la sustitución confusa entre servicios
- Registro de eventos de seguridad en AWS Lake Formation

Protección de datos en Lake Formation

El modelo de responsabilidad compartida de AWS se aplica a la protección de datos en . Como se describe en este modelo, AWS es responsable de proteger la infraestructura global en la que se ejecutan todos los Nube de AWS. Eres responsable de mantener el control sobre el contenido

alojado en esta infraestructura. También eres responsable de las tareas de administración y configuración de seguridad para los Servicios de AWS que utiliza. Para obtener más información sobre la privacidad de los datos, consulte las Preguntas frecuentes sobre privacidad de datos y los . Para obtener más información sobre la protección de datos en Europa, consulte el Centro del Reglamento General de Protección de Datos (RGPD).

Para fines de protección de datos, le recomendamos que proteja Cuenta de AWS las credenciales y configure usuarios individuales con AWS IAM Identity Center o AWS Identity and Access Management (IAM). De esta manera, solo se otorgan a cada usuario los permisos necesarios para cumplir sus obligaciones laborales. También recomendamos proteger sus datos de la siguiente manera:

- Utiliza la autenticación multifactor (MFA) en cada cuenta.
- Se utiliza SSL/TLS para comunicarse con AWS los recursos. Exigimos TLS 1.2 y recomendamos TLS 1.3.
- Configure la API y el registro de actividad de los usuarios con AWS CloudTrail. Para obtener información sobre el uso de CloudTrail senderos para capturar AWS actividades, consulte Cómo trabajar con CloudTrail senderos en la Guía del AWS CloudTrail usuario.
- Utilice soluciones de AWS cifrado, junto con todos los controles de seguridad predeterminados Servicios de AWS.
- Utiliza servicios de seguridad administrados avanzados, como Amazon Macie, que lo ayuden a detectar y proteger la información confidencial almacenada en Amazon S3.
- Si necesita módulos criptográficos validados por FIPS 140-3 para acceder a AWS través de una interfaz de línea de comandos o una API, utilice un punto final FIPS. Para obtener más información sobre los puntos de conexión de FIPS disponibles, consulte Estándar de procesamiento de la información federal (FIPS) 140-3.

Se recomienda encarecidamente no introducir nunca información confidencial o sensible, como por ejemplo, direcciones de correo electrónico de clientes, en etiquetas o campos de formato libre, tales como el campo Nombre. Esto incluye cuando trabaja o Servicios de AWS utiliza la consola, la API o los SDK. AWS CLI AWS Cualquier dato que introduzca en etiquetas o campos de formato libre utilizados para los nombres se pueden emplear para los registros de facturación o diagnóstico. Si proporciona una URL a un servidor externo, recomendamos encarecidamente que no incluya la información de las credenciales en la URL para validar la solicitud para ese servidor.

Cifrado en reposo

AWS Lake Formation admite el cifrado de datos en las siguientes áreas:

- Datos en el lago de datos de Amazon Simple Storage Service (Amazon S3).

  Lake Formation es compatible con el cifrado de datos con AWS Key Management Service (AWS KMS). Los datos se escriben normalmente en el lago de datos mediante trabajos de extracción, transformación y carga (ETL) de AWS Glue. Para obtener información sobre cómo cifrar los datos escritos por trabajos AWS Glue, consulte Cifrado de datos escritos por rastreadores, trabajos y puntos de conexión de desarrollo en la Guía para desarrolladores de AWS Glue .
- El AWS Glue Data Catalog, que es donde Lake Formation almacena las tablas de metadatos que describen los datos del lago de datos.

  Para obtener más información, consulte Cifrado de su Catálogo de datos en la Guía para desarrolladores de AWS Glue .

Para añadir una ubicación de Amazon S3 como almacenamiento en su lago de datos, debe registrar la ubicación en AWS Lake Formation. A continuación, puede utilizar los permisos de Lake Formation para un control de acceso específico a los objetos AWS Glue Data Catalog que apuntan a esta ubicación y a sus datos subyacentes.

Lake Formation es compatible con el registro de una ubicación de Amazon S3 que contenga datos cifrados. Para obtener más información, consulte Registro de una ubicación cifrada de Amazon S3.

Seguridad de la infraestructura en AWS Lake Formation

Como servicio gestionado, AWS Lake Formation está protegido por los procedimientos de seguridad de red AWS global que se describen en el documento técnico Amazon Web Services: Overview of Security Processes.

Utiliza las llamadas a la API AWS publicadas para acceder a Lake Formation a través de la red. Los clientes deben ser compatibles con la seguridad de la capa de transporte (TLS) 1.0 o una versión posterior. Recomendamos TLS 1.2 o una versión posterior. Los clientes también deben ser compatibles con conjuntos de cifrado con confidencialidad directa total (PFS) tales como Ephemeral Diffie-Hellman (DHE) o Elliptic Curve Ephemeral Diffie-Hellman (ECDHE). La mayoría de los sistemas modernos como Java 7 y posteriores son compatibles con estos modos.

Además, las solicitudes deben estar firmadas mediante un ID de clave de acceso y una clave de acceso secreta que esté asociada a una entidad principal de IAM. También puedes utilizar AWS Security Token Service (AWS STS) para generar credenciales de seguridad temporales para firmar solicitudes.

Prevención de la sustitución confusa entre servicios

El problema de la sustitución confusa es un problema de seguridad en el que una entidad que no tiene permiso para realizar una acción puede obligar a una entidad con más privilegios a realizar la acción. En AWS, la suplantación de identidad entre servicios puede provocar un confuso problema de diputado. La suplantación entre servicios puede producirse cuando un servicio (el servicio que lleva a cabo las llamadas) llama a otro servicio (el servicio al que se llama). El servicio que lleva a cabo las llamadas se puedes manipular para utilizar sus permisos a fin de actuar en función de los recursos de otro cliente de una manera en la que no debe tener permiso para acceder. Para evitarlo, AWS proporciona herramientas que le ayudan a proteger los datos de todos los servicios cuyos directores de servicio tengan acceso a los recursos de su cuenta.

Se recomienda utilizar las claves de contexto de condición global `aws:SourceArn` y `aws:SourceAccount` en las políticas de recursos para limitar los permisos que AWS Lake Formation concede a otro servicio para el recurso. Si se utilizan ambas claves contextuales de condición global, el valor `aws:SourceAccount` y la cuenta del valor `aws:SourceArn` deben utilizar el mismo ID de cuenta cuando se utilicen en la misma declaración de política.

Actualmente, Lake Formation solo admite `aws:SourceArn` en los formatos siguientes:

```
arn:aws:lakeformation:aws-region:account-id:*
```

El ejemplo siguiente muestra cómo se pueden utilizar las claves contextuales de condición global `aws:SourceArn` y `aws:SourceAccount` en Lake Formation para evitar el problema del suplente confuso.

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Sid": "ConfusedDeputyPreventionExamplePolicy",
```

```
      "Effect": "Allow",
      "Principal": {
        "Service": "lakeformation.amazonaws.com"
      },
      "Action": [
        "sts:AssumeRole"
      ],
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "123456789012"
        },
        "ArnEquals": {
          "aws:SourceArn": "arn:aws:lakeformation:aws-region:123456789012:*"
        }
      }
    }
  ]
}
```

Registro de eventos de seguridad en AWS Lake Formation

AWS Lake Formation está integrado con AWS CloudTrail un servicio que proporciona un registro de las acciones realizadas por un usuario, función o AWS servicio en Lake Formation. CloudTrail captura todas las llamadas a la API de Lake Formation como eventos. Las llamadas capturadas incluyen llamadas desde la consola de Lake Formation AWS Command Line Interface, y llamadas en código a las operaciones de la API de Lake Formation.

Para obtener más información acerca del registro de eventos de Lake Formation, consulte Registro de llamadas a la API de AWS Lake Formation mediante AWS CloudTrail.

> **Note**
>
> `GetTableObjects`, `UpdateTableObjects` y `GetWorkUnitResults` son operaciones de planos de datos de gran volumen. Las llamadas a estas APIs plataformas no están registradas actualmente CloudTrail. Para obtener más información sobre las operaciones del plano de datos en CloudTrail, consulte Registrar eventos de datos para senderos en la Guía del AWS CloudTrail usuario.
> Los cambios en Lake Formation para apoyar CloudTrail eventos adicionales se documentarán enHistorial de documentos para AWS Lake Formation.

Integración de servicios de terceros con Lake Formation

La integración con AWS Lake Formation permite a los servicios de terceros acceder de forma segura a los datos de sus lagos de datos basados en Amazon S3. Puede usar Lake Formation como motor de autorización para administrar o hacer cumplir los permisos de su lago de datos con AWS servicios integrados como AWS Glue ETL, Amazon Athena, Amazon EMR y Redshift Spectrum. Lake Formation ofrece dos opciones para integrar servicios:

1. Configuración de integración de la aplicación Lake Formation: Lake Formation puede vender credenciales temporales con alcance reducido en forma de tokens AWS STS a ubicaciones registradas de Amazon S3 en función de los permisos vigentes, de modo que las aplicaciones autorizadas puedan acceder a los datos en nombre de los usuarios.
2. Aplicación centralizada: las operaciones de la API de consulta de Lake Formation recuperan datos de Amazon S3 y filtran los resultados en función de los permisos efectivos. El motor o la aplicación que se integran con la operación de la API de consulta pueden depender de que Lake Formation evalúe los permisos de la identidad que realiza la llamada y filtre los datos de forma segura en función de estos permisos. Third-party los motores de consultas solo ven y funcionan con los datos filtrados.

La provisión de credenciales de Lake Formation no se integra con las consultas de Spark SQL. La venta de credenciales solo funciona con consultas que se ejecutan a través de la biblioteca AWS Glue ETL.

Temas

- Uso de la integración de aplicaciones de Lake Formation

Uso de la integración de aplicaciones de Lake Formation

Lake Formation permite que los servicios de terceros se integren con Lake Formation y obtengan acceso temporal a los datos de Amazon S3 en nombre de sus usuarios mediante el uso GetTemporaryGlueTableCredentialsy GetTemporaryGluePartitionCredentialslas operaciones. Esto permite que los servicios de terceros utilicen la misma función de venta de autorizaciones y credenciales que utilizan el resto de los servicios de AWS análisis. En esta sección se describe cómo utilizar estas operaciones de la API para integrar un motor de consultas de terceros con Lake Formation.

Estas operaciones de la API están deshabilitadas de forma predeterminada. Existen dos opciones para autorizar a Lake Formation a integrar aplicaciones:

- Configurar las etiquetas de sesión de IAM para que se validen cada vez que se invoquen las operaciones de la API de integración de aplicaciones

  Para obtener más información, consulte Habilitar los permisos para que un motor de consultas de terceros llame a las operaciones de la API de integración de aplicaciones.
- Habilitar la opción que permite a los motores externos acceder a los datos en las ubicaciones de Amazon S3 con acceso completo a las tablas

  Esta opción permite que los motores de consultas y las aplicaciones obtengan credenciales sin etiquetas de sesión de IAM si el usuario tiene acceso completo a la tabla. Proporciona beneficios de rendimiento a los motores de consulta y las aplicaciones, además de simplificar el acceso a los datos. Amazon EMR en Amazon EC2 puede aprovechar esta configuración.

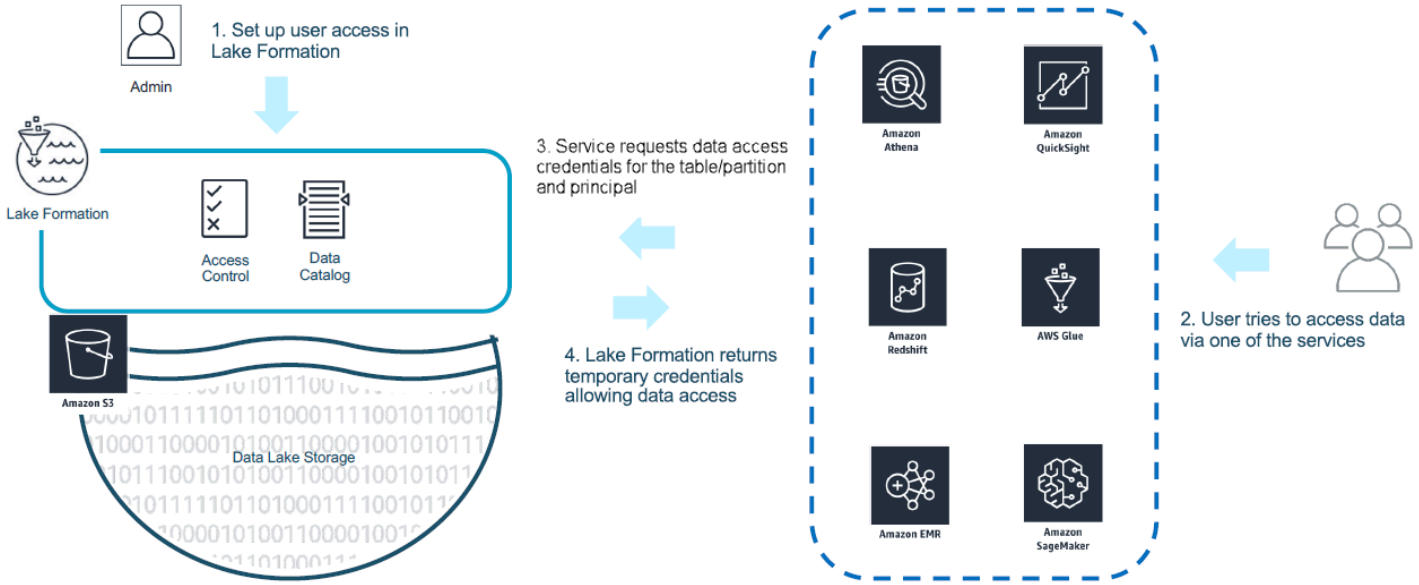
  Para obtener más información, consulte Integración de aplicaciones para un acceso completo a la tabla.

Temas

- Cómo funciona la integración de aplicaciones de Lake Formation
- Roles y responsabilidades en la integración de aplicaciones de Lake Formation
- Lake Formation flujo de trabajo para las operaciones de API de integración de aplicaciones
- Registro de un motor de consultas de terceros
- Habilitar los permisos para que un motor de consultas de terceros llame a las operaciones de la API de integración de aplicaciones
- Integración de aplicaciones para un acceso completo a la tabla

Cómo funciona la integración de aplicaciones de Lake Formation

En esta sección se describe cómo utilizar las operaciones de la API de integración de aplicaciones para integrar una aplicación de terceros (motor de consultas) con Lake Formation.



1. El administrador de Lake Formation efectúa las siguientes actividades:
   - Registra una ubicación de Amazon S3 en Lake Formation proporcionando un rol de IAM (utilizado para suministrar credenciales) que tiene los permisos adecuados para acceder a los datos de la ubicación de Amazon S3
   - Registra una aplicación de terceros para hacer llamadas a las operaciones de la API de suministro de credenciales de Lake Formation. Consulte the section called “Registro de un motor de consultas de terceros”
   - Concede permisos para permitir el acceso a bases de datos y tablas

     Por ejemplo, si desea publicar un conjunto de datos de sesiones de usuario que incluya algunas columnas que contengan información de identificación personal (PII), para restringir el acceso, asigne a estas columnas una LF-TBACetiqueta denominada «clasificación» con el valor «confidencial». A continuación, define un permiso que permita a un analista empresarial acceder a los datos de las sesiones de los usuarios, pero excluye las columnas etiquetadas con classification = sensitive.
2. Una entidad principal (usuario) envía una consulta a un servicio integrado.
3. La aplicación integrada envía la solicitud a Lake Formation solicitando la información de la tabla y las credenciales para acceder a la tabla.
4. Si la entidad principal que efectúa la consulta está autorizada a acceder a la tabla, Lake Formation devuelve las credenciales a la aplicación integrada, lo que permite el acceso a los datos.

> **Note**
>
> Lake Formation no accede a los datos subyacentes al expedir credenciales.

5. El servicio integrado lee los datos de Amazon S3, filtra las columnas según las políticas que ha recibido y devuelve los resultados a la entidad principal.

> **Important**
>
> Las operaciones de la API de suministro de credenciales de Lake Formation permiten una aplicación distribuida con un modelo explícito de denegación en caso de fallo (cierre por error). Esto introduce un modelo de seguridad tripartito entre los clientes, los servicios de terceros y Lake Formation. Se confía en los servicios integrados para hacer cumplir adecuadamente los permisos de Lake Formation (aplicación distribuida).

El servicio integrado es responsable de filtrar los datos leídos de Amazon S3 según las políticas devueltas de Lake Formation antes de que los datos filtrados regresen al usuario. Los servicios integrados siguen un modelo de cierre por error, lo que significa que la consulta debe fallar si no pueden hacer cumplir los permisos de Lake Formation necesarios.

Roles y responsabilidades en la integración de aplicaciones de Lake Formation

A continuación se indican las funciones y las responsabilidades asociadas a las mismas para permitir la integración de aplicaciones de terceros con AWS Lake Formation ellas.

Rol	Responsabilidad
El cliente	• Habilita la configuración de integración de aplicaciones de Lake Formation (consulte the section called “Registro de un motor de consultas de terceros”). • Registra explícitamente los terceros aprobados en Lake Formation (consulte the section called “Registro de un motor de consultas de terceros”).

Rol	Responsabilidad
	• Prueba y valida soluciones de terceros con los permisos de Lake Formation. • Supervisa y audita el uso de terceros de las operaciones de la API de suministro de credenciales de Lake Formation.
El tercero	• Documenta públicamente la capacidad compatible con cada revisión del software y proporciona instrucciones para habilitarla correctam ente. • Anuncia con precisión las capacidades compatibles al llamar a las operaciones de la API de suministro de credenciales de Lake Formation (según la documentación). • Almacena y gestiona de forma segura las credenciales suministradas para evitar la filtración de credenciales y el aumento de privilegios. • Aplica los permisos de acuerdo con las capacidades compatibles y devuelve solo los datos filtrados a los usuarios • No se puede llevar a cabo la consulta si no se pueden aplicar correctamente los permisos necesarios
`AWS Lake Formation`	• Deriva y devuelve correctamente los permisos efectivos para una determinada entidad principal. • Valida las capacidades soportadas por terceros llamada a llamada de la operación de la API. • Devuelve las credenciales de IAM con un alcance reducido solo cuando las capacidades anunciadas del motor coinciden con las definidas en los recursos del catálogo; de lo contrario, devuelve un error.

Lake Formation flujo de trabajo para las operaciones de API de integración de aplicaciones

A continuación se muestra el flujo de trabajo para las operaciones de la API de integración de aplicaciones:

1. Un usuario envía una consulta o solicitud de datos mediante un motor de consultas integrado de terceros. El motor de consultas asume un rol de IAM que representa al usuario o a un grupo de usuarios y recupera las credenciales de confianza para utilizarlas al llamar a las operaciones de la API de integración de aplicaciones.
2. El motor de consultas invoca `GetUnfilteredTableMetadata`. Si se trata de una tabla con particiones, invoca `GetUnfilteredPartitionsMetadata` para recuperar los metadatos y la información sobre políticas del Catálogo de datos.
3. Lake Formation autoriza la solicitud. Si el usuario no tiene los permisos adecuados en la tabla, AccessDeniedExceptionse descarta.
4. Como parte de la solicitud, el motor de consultas envía el filtrado que admite. Hay dos indicadores que se pueden enviar dentro de una matriz: COLUMN_PERMISSIONS y CELL_FILTER_PERMISSION. Si el motor de consultas no admite ninguna de estas funciones y existe una política en la tabla para la función, se lanza una y la consulta no PermissionTypeMismatchExceptionse realiza correctamente. Esto ocurre para evitar la fuga de datos.
5. La respuesta devuelta contiene lo siguiente:
   - El esquema completo de la tabla, que los motores de consultas pueden usar para analizar los datos almacenados.
   - Una lista de columnas autorizadas a las que el usuario tiene acceso. Si la lista de columnas autorizadas está vacía, significa que el usuario tiene permisos de `DESCRIBE`, pero no tiene permisos de `SELECT`, y se produce un error en la consulta.
   - Una marca, `IsRegisteredWithLakeFormation`, que indica si Lake Formation puede vender credenciales a los datos de este recurso. Si el resultado es falso, deben usarse credenciales de clientes para acceder a Amazon S3.
   - Una lista de `CellFilters` si hay alguno que deba aplicarse a las filas de datos. Esta lista contiene columnas y una expresión para evaluar cada fila. Solo debe rellenarse si se envía CELL_FILTER_PERMISSION como parte de la solicitud y hay un filtro de datos en la tabla para el usuario que llama.
6. Una vez recuperados los metadatos, el motor de consultas llama `GetTemporaryGlueTableCredentials` o `GetTemporaryGluePartitionCredentials` solicita AWS credenciales para recuperar los datos de la ubicación de Amazon S3.
7. El motor de consultas lee los objetos relevantes de Amazon S3, filtra los datos según las políticas recibidas en el paso 2 y devuelve los resultados al usuario.

Las operaciones de la API de integración de aplicaciones para Lake Formation incluyen contenido adicional para configurar la integración con motores de consultas de terceros. Puede ver los detalles de la operación en la sección Operaciones de la API de suministro de credenciales.

`QuerySessionContext`Se trata de una estructura a la que los motores de consultas también pueden enviar información Lake Formation para estas operaciones de la API de integración de aplicaciones. Permite Lake Formation almacenar y utilizar contexto adicional para una consulta determinada. A continuación se proporciona un ejemplo de cómo se QuerySessionContextdebe utilizar:

1. El motor de consultas realiza una `GetInternalUnfilteredMetadata` llamada y pasa una estructura QSC que contiene un identificador de consulta único en la solicitud:

```
{
    "QuerySessionContext": {
        "QueryId": "your-unique-identifier-here"
    }
}
```

2. La `GetInternalUnfilteredMetadata` llamada habrá devuelto una `QueryAuthorizationId` cadena en la respuesta. En la siguiente llamada de consulta (y en cualquier otra posterior) que acepte una estructura QSC en la entrada, el motor de consultas pasa la misma estructura QSC que ahora también contiene la información `QueryAuthorizationId` devuelta por. Lake Formation Supongamos que la siguiente llamada es`GetTemporaryGlueTableCredentials`; la solicitud contendrá:

```
{
    "QuerySessionContext": {
        "QueryAuthorizationId": "lf-returned-query-authz-id-here",
        "QueryId": "your-unique-identifier-here"
    },
}
```

Registro de un motor de consultas de terceros

Antes de que un motor de consultas de terceros pueda utilizar las operaciones de la API de integración de aplicaciones, debe habilitar explícitamente los permisos para que el motor de consultas llame a las operaciones de la API en su nombre. Esto se efectúa en unos pocos pasos:

1. Debe especificar AWS las cuentas y las etiquetas de sesión de IAM que requieren permiso para llamar a las operaciones de la API de integración de aplicaciones a través de la AWS Lake Formation consola, la AWS CLI o la API/SDK.
2. Cuando el motor de consultas de terceros asume el rol de ejecución en su cuenta, el motor de consultas debe adjuntar una etiqueta de sesión registrada en Lake Formation que represente al motor de terceros. Lake Formation usa esta marca para validar si la solicitud proviene de un motor aprobado. Para obtener más información acerca de las marcas de sesión, consulte Etiquetas de sesión el la Guía del usuario de IAM.
3. Al configurar un rol de ejecución de un motor de consultas de terceros, debe tener el siguiente conjunto mínimo de permisos en la política de IAM:

   JSON

```
{
  "Version":"2012-10-17",
  "Statement": {"Effect": "Allow",
    "Action": [
      "lakeformation:GetDataAccess",
      "glue:GetTable",
      "glue:GetTables",
      "glue:GetDatabase",
      "glue:GetDatabases",
      "glue:CreateDatabase",
      "glue:GetUserDefinedFunction",
      "glue:GetUserDefinedFunctions",
      "glue:GetPartition",
      "glue:GetPartitions"
    ],
    "Resource": "*"
  }
}
```

4. Configure una política de confianza de roles en el rol de ejecución del motor de consultas para tener un control de acceso preciso sobre qué par clave-valor de etiqueta de sesión puede adjuntarse a este rol. En el ejemplo siguiente, a este rol solo se le permite adjuntar una clave de etiqueta de sesión `"LakeFormationAuthorizedCaller"` y un valor de etiqueta de sesión `"engine1"`, y no se permite ningún otro par de valores clave de etiqueta de sesión.

```
{
    "Sid": "AllowPassSessionTags",
```

```
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::111122223333:role/query-execution-role"
    },
    "Action": "sts:TagSession",
    "Condition": {
    "StringLike": {
        "aws:RequestTag/LakeFormationAuthorizedCaller": "engine1"        }
    }
}
```

Cuando se `LakeFormationAuthorizedCaller` llama a la operación de la STS:AssumeRole API para obtener credenciales para que las utilice el motor de consultas, la etiqueta de sesión debe incluirse en la AssumeRole solicitud. La credencial temporal devuelta se puede usar para efectuar solicitudes a la API de integración de aplicaciones de Lake Formation.

Las operaciones de la API de integración de aplicaciones de Lake Formation requieren que la entidad principal que llama tenga un rol de IAM. El rol de IAM debe incluir una etiqueta de sesión con un valor predeterminado que se haya registrado en Lake Formation. Esta etiqueta permite a Lake Formation verificar si el rol utilizado para llamar a las operaciones de la API de integración de aplicaciones está autorizado a hacerlo.

Habilitar los permisos para que un motor de consultas de terceros llame a las operaciones de la API de integración de aplicaciones

Sigue estos pasos para permitir que un motor de consultas de terceros llame a las operaciones de la API de integración de aplicaciones a través de la AWS Lake Formation consola, el AWS CLI API/SDK quirófano.

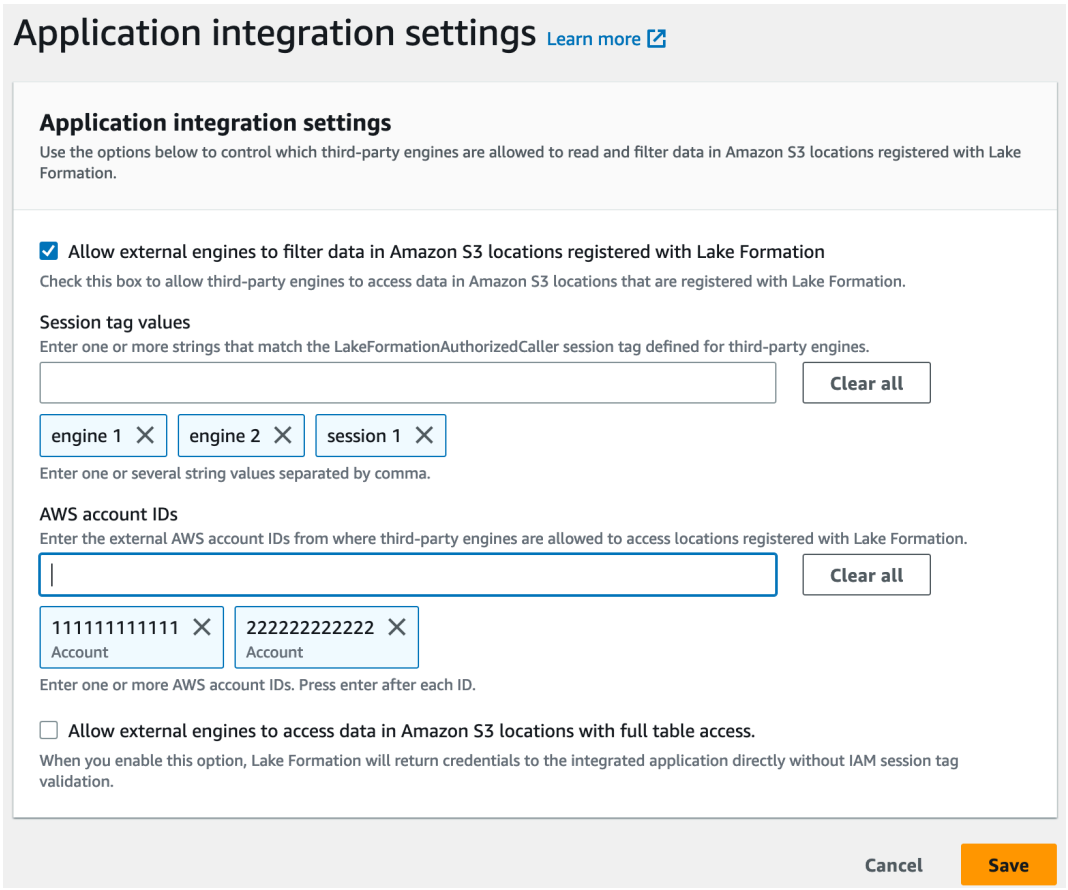
Console

Para registrar tu cuenta para el filtrado de datos externo:

1. Inicie sesión en y abra la Consola de administración de AWS consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación de la izquierda, expanda Administración y, a continuación, elija Configuración de integración de aplicaciones.

3. En la página Configuración de integración de aplicaciones, elija la opción Permitir que motores externos filtren datos en las ubicaciones de Amazon S3 registradas en Lake Formation.
4. Introduzca las etiquetas de sesión que creó para el motor de terceros. Para obtener información sobre las etiquetas de sesión, consulte Pasar etiquetas de sesión en AWS STS en la Guía del AWS Identity and Access Management usuario.
5. Introduzca los ID de cuenta de los usuarios que pueden utilizar el motor de terceros para acceder a la información de metadatos sin filtrar y a las credenciales de acceso a los datos de los recursos de la cuenta actual.

   También puede usar el campo de ID de AWS cuenta para configurar el acceso entre cuentas.



CLI

Utilice el siguiente comando de CLI `put-data-lake-settings` para establecer los parámetros siguientes.

Hay tres campos que se deben configurar cuando se utiliza este AWS CLI comando:

- `allow-external-data-filtering` – (booleano) Indica que un motor de terceros puede acceder a la información de metadatos y a las credenciales de acceso a los datos sin filtrar de los recursos de la cuenta actual.
- `external-data-filtering-allow-list` – (matriz) Una lista de los ID de cuenta que pueden acceder a la información de metadatos sin filtrar y a las credenciales de acceso a los datos de los recursos de la cuenta actual cuando se utiliza un motor de terceros. Si AllowExternalDataFiltering se establece en true, la ExternalDataFilteringAllowList propiedad debe incluir al menos un identificador de cuenta. No se permite una lista vacía.
- `authorized-sessions-tag-value-list` – (matriz) Una lista de valores de etiquetas de sesión autorizados (cadenas). Si se ha adjuntado una credencial de rol de IAM a un par clave-valor autorizado, en caso de que la etiqueta de sesión figure en la lista, la sesión tendrá acceso a la información de metadatos y a las credenciales de acceso a los datos sin filtrar de los recursos de la cuenta configurada. La clave de etiqueta de sesión autorizada se define como `*LakeFormationAuthorizedCaller*`.
- `AllowFullTableExternalDataAccess` - (booleano) Si se debe permitir que un motor de consultas de terceros obtenga credenciales de acceso a los datos sin etiquetas de sesión cuando la persona que llama tiene todos los permisos de acceso a los datos.

Por ejemplo:

```
aws lakeformation put-data-lake-settings --cli-input-json file://
datalakesettings.json

{
  "DataLakeSettings": {
    "DataLakeAdmins": [
      {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111111111111:user/lakeAdmin"
      }
    ],
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "TrustedResourceOwners": [],
    "AllowExternalDataFiltering": true,
    "ExternalDataFilteringAllowList": [
        {"DataLakePrincipalIdentifier": "111111111111"}
        ],
    "AuthorizedSessionTagValueList": ["engine1"],
    "AllowFullTableExternalDataAccess": false
```

```
    }


}
```

API/SDK

Para ello, use la operación de la API `PutDataLakeSetting` para establecer los parámetros siguientes.

Al utilizar esta operación de la API, se deben configurar tres campos:

- `AllowExternalDataFiltering` – (booleano) Indica SI un motor de terceros puede acceder a la información de metadatos y a las credenciales de acceso a los datos sin filtrar de los recursos de la cuenta actual.
- `ExternalDataFilteringAllowList` – (matriz) Una lista de los ID de cuenta que pueden acceder a la información de metadatos sin filtrar y a las credenciales de acceso a los datos de los recursos de la cuenta actual al utilizar un motor de terceros.
- `AuthorizedSessionsTagValueList` – (matriz) Una lista de valores de etiquetas autorizadas (cadenas). Si se ha adjuntado una credencial de rol de IAM a una etiqueta autorizada, la sesión recibirá acceso a la información de metadatos y a las credenciales de acceso a los datos sin filtrar de los recursos de la cuenta configurada. La clave de etiqueta de sesión autorizada se define como `*LakeFormationAuthorizedCaller*`.
- `AllowFullTableExternalDataAccess` - (booleano) Si se debe permitir que un motor de consultas de terceros obtenga credenciales de acceso a los datos sin etiquetas de sesión cuando la persona que llama tiene todos los permisos de acceso a los datos.

Por ejemplo:

```
//Enable session tag on existing data lake settings
public void sessionTagSetUpForExternalFiltering(AWSLakeFormationClient
 lakeformation) {
    GetDataLakeSettingsResult getDataLakeSettingsResult =
 lfClient.getDataLakeSettings(new GetDataLakeSettingsRequest());
    DataLakeSettings dataLakeSettings =
 getDataLakeSettingsResult.getDataLakeSettings();

    //set account level flag to allow external filtering
    dataLakeSettings.setAllowExternalDataFiltering(true);
```

```
    //set account that are allowed to call credential vending or Glue
 GetFilteredMetadata API
    List<DataLakePrincipal> allowlist = new ArrayList<>();
    allowlist.add(new
 DataLakePrincipal().withDataLakePrincipalIdentifier("111111111111"));
    dataLakeSettings.setWhitelistedForExternalDataFiltering(allowlist);

    //set registered session tag values
    List<String> registeredTagValues = new ArrayList<>();
    registeredTagValues.add("engine1");
    dataLakeSettings.setAuthorizedSessionTagValueList(registeredTagValues);

    lakeformation.putDataLakeSettings(new
 PutDataLakeSettingsRequest().withDataLakeSettings(dataLakeSettings));
}
```

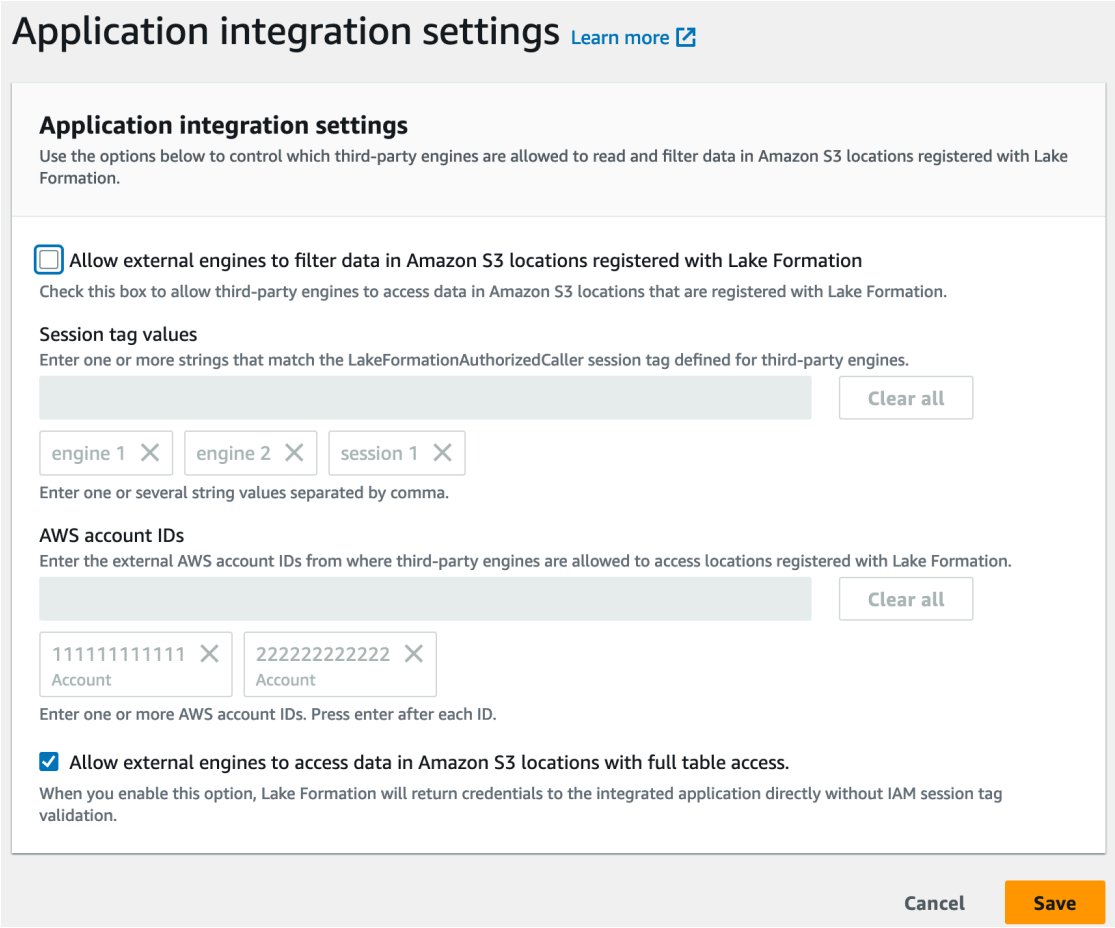
Integración de aplicaciones para un acceso completo a la tabla

Siga estos pasos para permitir que los motores de consultas de terceros accedan a los datos sin la validación de la etiqueta de sesión de IAM:

Console

1. Inicie sesión en la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación de la izquierda, expanda Administración y, a continuación, elija Configuración de integración de aplicaciones.
3. En la página Configuración de integración de aplicaciones, elija la opción Permitir que motores externos accedan a datos en las ubicaciones de Amazon S3 con acceso total a las tablas.

   Al activar esta opción, Lake Formation devuelve las credenciales a la aplicación que efectúa la consulta directamente sin validar la etiqueta de sesión de IAM.



AWS CLI

Utilice el comando CLI `put-data-lake-settings` para establecer el parámetro `AllowFullTableExternalDataAccess`.

```
aws lakeformation put-data-lake-settings –cli-input-json file://put-data-lake-
settings.json –region ap-northeast-1
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111111111111:user/
lakeAdmin"
            }
        ],
        "AllowFullTableExternalDataAccess": true
    }
}
```

Trabajar con otros AWS servicios

AWS servicios como Amazon Athena AWS Glue, Amazon Redshift Spectrum y Amazon EMR se pueden AWS Lake Formation utilizar para acceder de forma segura a los datos de las ubicaciones de Amazon S3 registradas en Lake Formation. Con Lake Formation, puede definir y administrar permisos de control de acceso específicos (FGAC) para sus tablas del AWS Glue Data Catalog. Cada uno de estos AWS servicios es una persona de confianza que llama a Lake Formation, y Lake Formation proporciona acceso a los datos almacenados en Amazon S3 mediante credenciales temporales. Para obtener más información, consulte Cómo funciona la integración de aplicaciones de Lake Formation.

Para aprovechar estas prestaciones, Lake Formation requiere que antes registre la ubicación de Amazon S3 y asigne los permisos adecuados a la entidad principal de IAM para acceder a la tabla, la base de datos y la ubicación de Amazon S3. Para obtener más información, consulte, Administrar los permisos de Lake Formation.

En las siguientes tablas, se enumeran los tipos de permisos de Lake Formation que admiten Amazon Athena, AWS Glue Amazon EMR y Amazon Redshift Spectrum para acceder a datos AWS Glue de tablas estándar y tablas transaccionales (Apache Iceberg, Apache Hudi y Delta Lake de la fundación Linux) con datos almacenados en Amazon S3 y metadatos de tablas en el catálogo de datos.

AWS servicios y tipos de permisos compatibles para tablas y vistas estándar AWS Glue

AWS servicio	Permisos de tabla	Permisos de columna	Permisos de fila y celda
Athena SQL	Acceso de lectura/escritura	Acceso de lectura	Acceso de lectura
Athena para Spark	No admitido	No admitido	No admitido
Redshift Spectrum en un clúster aprovisionado o en Amazon Redshift sin servidor	Acceso de lectura/escritura	Acceso de lectura	Acceso de lectura
Apache Spark en Amazon EMR (EC2)	Acceso de lectura/escritura	Acceso de lectura	Acceso de lectura

AWS servicio	Permisos de tabla	Permisos de columna	Permisos de fila y celda
Apache Hive en Amazon EMR (EC2)	Acceso de lectura/escritura	Acceso de lectura	No compatible
Apache Spark en EMR sin servidor	Acceso de lectura/escritura	Acceso de lectura	Acceso de lectura
Apache Hive en EMR sin servidor	No admitido	No admitido	No admitido
Amazon EMR en EKS	No admitido	No admitido	No admitido
AWS Glue ETL	Acceso de lectura/escritura	AWS Glue La versión 5.0 o superior admite el acceso de lectura.	AWS Glue La versión 5.0 o superior admite el acceso de lectura.

Consideraciones y limitaciones

- Athena Spark no admite la consulta de tablas del Catálogo de datos con permisos de Lake Formation.
- Los usuarios de Athena basados en SAML pueden leer los orígenes de datos protegidos mediante permisos de Lake Formation habilitando la federación basada en SAML 2.0. Los usuarios de SAML pueden insertar datos en tablas de Parquet.
- Apache Spark en EMR sin servidor no admite la consulta de vistas del Catálogo de datos.
- Apache Hive en EMR sin servidor no admite la consulta de tablas con permisos de Lake Formation.
- AWS Glue La versión 5.0 o superior admite controles de acceso detallados en las tablas Iceberg y Hive del catálogo de datos respaldadas por S3. Esta capacidad le permite configurar los controles de acceso a nivel de tabla, fila, columna y celda para las consultas de lectura en sus AWS Glue trabajos de Apache Spark.

  Para obtener más información, consulte Versiones de AWS Glue.

AWS servicios y tipos de permisos compatibles para los formatos de tablas transaccionales

AWS servicio	Iceberg	Hudi	Delta Lake (nativo)	Delta Lake (tablas de enlaces simbólicos)
Athena SQL	Permite leer tablas con permisos de tabla, columna, fila y celda. Las operaciones de escritura requieren acceso completo a la tabla.	Permite leer y crear operaciones en tablas con permisos de tabla, columna, fila y celda. No se admiten las operaciones de escritura.	Athena (versión 3 del motor) admite la lectura de tablas nativas de Delta Lake con permisos de tabla, columna, fila y celda. No se admiten las operaciones de escritura.	Athena (versión 3 del motor) admite la lectura de tablas de enlaces simbólicos de Delta Lake con permisos de tabla, columna, fila y celda. No se admiten las operaciones de escritura.
Redshift Spectrum en un clúster aprovisionado	Permite leer tablas con permisos de tabla, columna, fila y celda. No se admiten las operaciones de escritura.	Permite leer tablas con permisos de tabla, columna, fila y celda. No se admiten las operaciones de escritura.	No compatible	Permite leer de tablas de Delta Lake mediante un manifiesto o de enlaces simbólicos con permisos de tabla, columna, fila y celda. No se admiten las operaciones de escritura.
Apache Spark en Amazon EMR (EC2)	Permite leer tablas con permisos de tabla, columna,	Permite leer tablas con permisos de tabla, columna,	Permite leer tablas con permisos de tabla, columna,	Permite leer tablas con permisos de tabla, columna,

AWS servicio	Iceberg	Hudi	Delta Lake (nativo)	Delta Lake (tablas de enlaces simbólicos)
	fila y celda. Las operaciones de escritura requieren acceso completo a la tabla.	fila y celda. Las operaciones de escritura requieren acceso completo a la tabla.	fila y celda. No se admiten las operaciones de escritura.	fila y celda. Las operaciones de escritura requieren acceso completo a la tabla.
AWS Glue ETL	AWS Glue La versión 5.0 o superior admite la lectura de tablas con permisos de tabla, columna, fila y celda.	Se admite read/write en tablas con permisos a nivel de tabla.	Se admite read/write en tablas con permisos a nivel de tabla.	Se admite read/write en tablas con permisos a nivel de tabla.

Temas

- Uso AWS Lake Formation con Amazon Athena
- Uso AWS Lake Formation con Amazon Redshift Spectrum
- Utilizándolo con AWS Lake Formation AWS Glue
- Uso AWS Lake Formation con Amazon EMR
- Uso AWS Lake Formation con Quick
- AWS Lake Formation Utilizándolo con AWS CloudTrail Lake

Uso AWS Lake Formation con Amazon Athena

Amazon Athena es un servicio de consultas sin servidor que le ayuda a analizar datos estructurados, semiestructurados y no estructurados almacenados en Amazon S3. Puede usar Athena SQL para consultar datos desde formatos de datos CSV, JSON, Parquet y Avro. Athena SQL también admite formatos de tabla como Apache Hive, Apache Hudi y Apache Iceberg. Athena se integra con AWS

Glue Data Catalog para almacenar los metadatos de sus conjuntos de datos en Amazon S3. Athena puede usar Lake Formation para definir y mantener políticas de control de acceso en esos conjuntos de datos.

Estos son algunos casos de uso comunes en los que puede usar Lake Formation con Athena.

- Utilice los permisos de Lake Formation para acceder a los recursos del Catálogo de datos (bases de datos y tablas) desde Athena. Puede usar el método de recurso designado o las etiquetas LF para definir los permisos en la base de datos y las tablas. Para obtener más información, consulte lo siguiente:
  - Concesión de permisos de base de datos mediante el método de recurso con nombre
  - Control de acceso basado en etiquetas de Lake Formation

  > **Note**
  >
  > Los permisos de Lake Formation se aplican solo al utilizar Athena SQL para consultar datos de origen de Amazon S3 y metadatos del Catálogo de datos.
  > Athena Spark no admite la consulta de tablas del Catálogo de datos con permisos de Lake Formation. Los permisos de Lake Formation admiten operaciones de lectura y escritura en bases de datos y tablas.

  > **Note**
  >
  > No puede aplicar filtros de datos cuando usa etiquetas LF para administrar los permisos en los recursos del Catálogo de datos.

- Controle los resultados de las consultas con Filtros de datos en Lake Formation para asegurar las tablas de sus lagos de datos Amazon S3 mediante permisos de columna, fila y celda. Consulte la limitación de la proyección de particiones en la Guía del usuario de Amazon Athena.
- Aplique un control de acceso detallado a los datos disponibles para el usuario de Athena basado en SAML al ejecutar consultas federadas.

  Los controladores JDBC y ODBC de Athena admiten la configuración del acceso federado al origen de datos mediante un proveedor de identidades (IdP) basado en SAML. Utilice Quick integrado con Lake Formation con su rol de IAM actual o con usuarios o grupos de SAML para visualizar los resultados de las consultas de Athena.

> **Note**
>
> Los permisos de Lake Formation para usuarios y grupos de SAML solo se reconocen cuando se utiliza el controlador JDBC u ODBC para enviar consultas a Athena.

Para obtener más información, consulte Uso de Lake Formation y de los controladores JDBC y ODBC de Athena para el acceso federado a Athena.

> **Note**
>
> Actualmente, no se admite la autorización del acceso a las identidades de SAML en Lake Formation en las regiones siguientes:
>
> - Medio Oriente (Baréin): me-south-1
> - Asia-Pacífico (Hong Kong): ap-east-1
> - África (Ciudad del Cabo): af-south-1
> - China (Ningxia): cn-northwest-1
> - Asia-Pacífico (Osaka): ap-northeast-3

- Utilice Cross-account intercambio de datos en Lake Formation para consultar tablas en otra cuenta.

> **Note**
>
> Para obtener más información sobre las limitaciones al utilizar los permisos de Lake Formation para `Views`, consulte Consideraciones y limitaciones.

Compatibilidad con formatos de tablas transaccionales

Con los permisos de Lake Formation puede proteger sus datos transaccionales en sus lagos de datos basados en Amazon S3. La tabla siguiente muestra los formatos de tablas transaccionales compatibles con los permisos de Athena y Lake Formation. Lake Formation impone estos permisos cuando los usuarios de Athena ejecutan sus consultas.

Formato de tabla	Descripción y operaciones permitidas	Permisos de formación de lagos admitidos en Athena
Apache Hudi	Formato utilizado para simplificar el procesamiento incremental de datos y el desarrollo de canalizaciones de datos. Athena admite operaciones de creación y lectura mediante formatos de tablas Apache Hudi en conjuntos de datos de Amazon S3 para los tipos de tablas Hudi Copy on Write (CoW) y Merge On Read (MoR). Athena no admite operaciones de escritura en tablas Hudi. Utilice Athena para consultar conjuntos de datos de Hudi.	Utilice Filtrado de datos y seguridad de celda en Lake Formation para proteger la tabla de Hudi mediante permisos de tabla, columna, fila y celda.
Apache Iceberg	Iceberg administra grandes colecciones de archivos como tablas y admite operaciones de lago de datos analíticos modernos, como las consultas de inserción, actualización y eliminación de registros, y viajes en el tiempo. Para obtener más informaci ón sobre la compatibilidad de Athena con las tablas Iceberg, consulte Uso de tablas Iceberg.	Son compatibles los permisos de tabla, columna, fila y celda. Actualmente, Lake Formation no admite la administración de permisos en operaciones de escritura como `VACUUM`, `MERGE`, `UPDATE` y `OPTIMIZE` en tablas en formatos de tabla abierta.

Formato de tabla	Descripción y operaciones permitidas	Permisos de formación de lagos admitidos en Athena
Linux Foundation Delta Lake	Delta Lake es un proyecto de código abierto que ayuda a implementar arquitecturas de lago de datos modernas basadas habitualmente en Amazon S3 o en Sistema de archivos distribuido de Hadoop (HDFS) Athena admite tablas de Delta Lake creadas mediante una definición de tabla de manifiesto basada en enlaces simbólicos a partir de una tabla AWS Glue Data Catalog de Delta Lake. Para obtener más información, consulte Rastrear las tablas de Delta Lake con rastreadores. AWS Glue Athena (versión 3 del motor) admite la lectura de tablas nativas de Delta Lake. Para obtener más información, consulte Presentamos el soporte nativo de Delta Lake para tablas con AWS Glue rastreadores.	Los permisos de tabla, columna, fila y celda son compatibles con las tablas de enlaces simbólicos y las tablas nativas de Delta Lake.

Recursos adicionales

Publicaciones de blog, vídeos y talleres

- Consultar un conjunto de datos de Apache Hudi en un lago de datos de Amazon S3 con Amazon Athena
- Cree un lago de datos de Apache Iceberg con Amazon Athena, Amazon EMR y AWS Glue
- Insertar, actualizar y eliminar en Amazon S3 con Athena y Apache Iceberg
- Control de acceso basado en etiquetas de LF Taller de Lake Formation sobre las consultas en un lago de datos.

Uso AWS Lake Formation con Amazon Redshift Spectrum

Con Amazon Redshift Spectrum puede consultar y recuperar datos en lagos de datos de Amazon S3 sin tener que cargar datos en nodos de clústeres de Amazon Redshift.

Redshift Spectrum admite dos formas de registrar un catálogo de AWS Glue datos externo habilitado con Lake Formation.

- Uso de un rol de IAM adjunto a un clúster que tenga permiso para el Catálogo de datos

  Para crear un rol de IAM, siga los pasos descritos en el siguiente procedimiento.

  Controlar el acceso al AWS Glue Data Catalog
- Uso de una identidad de IAM federada configurada para gestionar el acceso a recursos AWS Glue Data Catalog externos

  Redshift Spectrum admite la consulta de tablas de Lake Formation mediante identidades de IAM federadas. Las identidades de IAM pueden ser un usuario de IAM o un rol de IAM. Para obtener más información sobre la federación de identidades IAM en Redshift Spectrum, consulte Uso de una identidad federada para administrar el acceso de Amazon Redshift a los recursos locales y a las tablas externas de Redshift Spectrum.

Con la integración de Lake Formation y Redshift Spectrum, puede definir los permisos de control de acceso de fila, columna y celda en las tablas después de registrar sus datos en Lake Formation.

Para obtener más información, consulte Uso de Redshift Spectrum con. AWS Lake Formation

Redshift Spectrum admite lecturas o consultas de `SELECT` en las tablas de esquemas externos administradas por Lake Formation.

Para obtener más información, consulte Creación de esquemas externos para Redshift Spectrum.

Compatibilidad con tipos de tablas transaccionales

La tabla siguiente muestra los formatos de tablas transaccionales compatibles con los permisos de Athena y Lake Formation.

Formatos de tabla compatibles

Formato de tabla	Descripción y operaciones permitidas	Permisos de Lake Formation compatibles con Redshift Spectrum
Apache Hudi	Formato utilizado para simplificar el procesamiento incremental de datos y el desarrollo de canalizaciones de datos. Redshift Spectrum admite operaciones de inserción , eliminación y escritura alterada mediante el formato de tabla Apache Hudi Copy on Write (CoW) en Amazon S3. Para obtener más información, consulte Creación de tablas externas para datos administrados en Apache Hudi.	Utilice Filtrado de datos y seguridad de celda en Lake Formation para proteger las tablas de Hudi mediante permisos de tabla, columna, fila y celda.
Apache Iceberg	Iceberg administra grandes colecciones de archivos como tablas y admite operaciones de lago de datos analíticos modernos, como las consultas	Redshift Spectrum admite tablas Apache Iceberg para efectuar consultas.

Formato de tabla	Descripción y operaciones permitidas	Permisos de Lake Formation compatibles con Redshift Spectrum
	de inserción, actualización y eliminación de registros, y viajes en el tiempo. Para obtener más informaci ón, consulte Uso de tablas de Apache Iceberg con Amazon Redshift.	
Linux Foundation Delta Lake	Delta Lake es un proyecto de código abierto que ayuda a implementar arquitecturas de lago de datos modernos comúnmente construidas sobre Amazon S3 o Sistema de archivos distribuido de Hadoop (HDFS). Redshift Spectrum admite la consulta de tablas Delta Lake. Para obtener más información, consulte Creación de tablas externas para datos administr ados en Delta Lake	Son compatibles los permisos de tabla, columna, fila y celda.

Recursos adicionales

Publicaciones de blog, vídeos y talleres

- Centralice la gobernanza de su lago de datos utilizando, al AWS Lake Formation mismo tiempo, una arquitectura de datos moderna con Amazon Redshift Spectrum
- Utilice Redshift Spectrum para consultar tablas Apache HUDI Copy On Write (CoW) en el lago de datos de Amazon S3

Utilizándolo con AWS Lake Formation AWS Glue

Los ingenieros y DevOps profesionales de datos utilizan AWS Glue Extract, Transform and Load (ETL) con Apache Spark para realizar transformaciones en sus conjuntos de datos en Amazon S3 y cargar los datos transformados en lagos de datos y almacenes de datos para fines de análisis, aprendizaje automático y desarrollo de aplicaciones. Dado que diferentes equipos acceden al mismo conjunto de datos en Amazon S3, es imprescindible conceder y restringir los permisos en función de sus roles.

AWS Lake Formation se basa en AWS Glueél y los servicios interactúan de las siguientes maneras:

- Lake Formation y AWS Glue comparten el mismo Catálogo de datos.
- Las siguientes características de la consola de Lake Formation invocan la consola de AWS Glue:
  - Trabajos: para obtener más información, consulte Agregar trabajos en la Guía para desarrolladores de AWS Glue .
  - Rastreadores: para obtener más información, consulte la sección Catalogación de tablas con un rastreador en la Guía para desarrolladores de AWS Glue .
- Los flujos de trabajo que se generan cuando se utiliza un esquema AWS Gluede Lake Formation son flujos de trabajo de . Puede ver y gestionar estos flujos de trabajo tanto en la consola de Lake Formation como en la consola de AWS Glue.
- Las transformaciones de machine learning se proporcionan con Lake Formation y se basan en las operaciones de la API de AWS Glue. Puede crear y administrar transformaciones de machine learning en la consola de AWS Glue. Para obtener más información, consulte Transformaciones de machine learning en la Guía para desarrolladores de AWS Glue .

Puede utilizar el control de acceso detallado de Lake Formation para gestionar los recursos del Catálogo de datos existentes y las ubicaciones de datos de Amazon S3.

> **Note**
>
> AWS Glue La versión 5.0 o superior admite controles de acceso detallados en las tablas Iceberg y Hive respaldadas por S3. Esta capacidad le permite configurar los controles de acceso a nivel de tabla, fila, columna y celda para las consultas de lectura en sus AWS Glue trabajos de Apache Spark.

Compatibilidad con tipos de tablas transaccionales

Con los permisos de Lake Formation puede proteger sus datos transaccionales en sus lagos de datos basados en Amazon S3. La siguiente tabla muestra los formatos de tablas transaccionales admitidos AWS Glue y los permisos de Lake Formation. Lake Formation hace cumplir estos permisos para AWS Glue las operaciones.

Formatos de tabla compatibles

Formato de tabla	Descripción y operaciones permitidas	Los permisos de Lake Formation son compatibles en AWS Glue
Apache Hudi	Formato de tabla abierta para simplificar el procesamiento incremental de datos y el desarrollo de canalizaciones de datos. Para ver ejemplos, consulte Uso del marco Hudi en AWS Glue.	Los permisos a nivel de tabla están disponibles para las tablas Hudi. Para obtener más informaci ón, consulte la sección sobre Límites.
Apache Iceberg	Formato de tabla abierta que gestiona grandes colecciones de archivos como tablas. Para ver ejemplos, consulte Uso del marco Iceberg en. AWS Glue	AWS Glue La versión 5.0 y las versiones posteriores permiten configurar los controles de acceso a nivel de tabla, fila, columna y celda para las consultas de lectura en las tablas de tareas de Apache Spark para Iceberg. AWS Glue Para obtener más informaci ón, consulte la sección sobre Límites.

Formato de tabla	Descripción y operaciones permitidas	Los permisos de Lake Formation son compatibles en AWS Glue
Linux Foundation Delta Lake	Delta Lake es un proyecto de código abierto que ayuda a implementar arquitecturas de lago de datos modernos comúnmente construid as sobre Amazon S3 o el Sistema de archivos distribui do de Hadoop (HDFS). Para ver ejemplos, consulte Uso del marco Delta Lake en AWS Glue.	Los permisos a nivel de tabla están disponibles para las tablas Delta Lake. Para obtener más informaci ón, consulte la sección sobre Límites.

Recursos adicionales

Publicaciones de blog y repositorios

- Utilice el AWS Glue conector para leer y escribir tablas de Apache Iceberg con transacciones ACID y realizar viajes en el tiempo
- Escritura en tablas de Apache Hudi mediante AWS Glue un conector personalizado
- AWS repositorio de plantillas de Cloudformation y ejemplos de código de pyspark para analizar los datos de streaming mediante AWS Glue Apache Hudi y Amazon S3.

Uso AWS Lake Formation con Amazon EMR

Amazon EMR es una plataforma flexible de clústeres AWS gestionados en la que puede ejecutar cualquier código personalizado en marcos de big data compatibles, como Hadoop Map-Reduce, Spark, Hive, Presto, etc. Las organizaciones también utilizan Amazon EMR para ejecutar aplicaciones de procesamiento de datos por lotes y en streaming en un clúster altamente distribuido. Si utiliza Apache Spark en Amazon EMR, puede ejecutar sus transformaciones de datos y código personalizado en bases de datos y tablas cuyos permisos administre Lake Formation.

Hay tres opciones para implementar Amazon EMR:

- EMR en EC2
- EMR sin servidor
- Amazon EMR en EKS

Para obtener más información, consulte Integrar Amazon EMR con Lake Formation o Uso de EMR Serverless with para un control de acceso detallado AWS Lake Formation

Compatibilidad con formatos de tablas transaccionales

Las versiones 6.15.0 y posteriores de Amazon EMR incluyen compatibilidad con los permisos de control de acceso a nivel de tabla, fila, columna y celda de Lake Formation en los formatos de tabla Apache Hudi, Apache Iceberg y Delta Lake al leer y escribir datos con Spark SQL.

Para conocer las limitaciones, consulte Consideraciones sobre Amazon EMR con Lake Formation.

Formatos de tabla compatibles

Formato de tabla	Descripción y operaciones permitidas	Permisos de Lake Formation admitidos en Amazon EMR
Apache Hudi	Formato de tabla abierta para simplificar el procesamiento incremental de datos y el desarrollo de canalizaciones de datos. Para obtener una lista de las operaciones compatibles, consulte Apache Hudi y Lake Formation.	Amazon EMR es compatibl e con el control de acceso a nivel de tabla, fila, columna y celda con Apache Hudi.
Apache Iceberg	Formato de tabla abierta que gestiona grandes colecciones de archivos como tablas. Para obtener una lista de las operaciones compatibles,	Amazon EMR es compatibl e con el control de acceso a nivel de tabla, fila, columna y celda con Apache Iceberg.

Formato de tabla	Descripción y operaciones permitidas	Permisos de Lake Formation admitidos en Amazon EMR
	consulte Apache Iceberg y Lake Formation.	
Linux Foundation Delta Lake	Delta Lake es un proyecto de código abierto que ayuda a implementar arquitecturas de lago de datos modernos comúnmente construid as sobre Amazon S3 o el Sistema de archivos distribui do de Hadoop (HDFS). Para obtener una lista de las operaciones compatibles, consulte Delta Lake y Lake Formation.	Amazon EMR permite el control de acceso de nivel de tabla, fila, columna y celda con tablas de Delta Lake.

Recursos adicionales

Guía del usuario, publicaciones de blog y talleres

- Integración con Amazon EMR mediante roles de tiempo de ejecución
- Introducción rápida a Apache Hudi, Apache Iceberg y Delta Lake con Amazon EMR en EKS
- Uso de OSS de Delta Lake con EMR sin servidor

Uso AWS Lake Formation con Quick

Quick permite explorar conjuntos de datos gestionados por los permisos de Lake Formation en Amazon S3 mediante Athena.

Tanto los usuarios de las ediciones Standard como Enterprise de Quick se integran con Lake Formation, pero de forma ligeramente diferente.

- Edición empresarial: conceda permisos de control de acceso (FGAC) detallados a usuarios y grupos individuales de Quick para acceder a bases de datos y tablas.
- Edición Standard: permisos a los roles de IAM para acceder a bases de datos y tablas.

> **Note**
>
> De forma predeterminada, Quick usa un rol denominado. `aws-quicksight-service-role-v0` También puede definir funciones personalizadas con los permisos necesarios que permitan acceder rápidamente a Athena.

Para obtener más información, consulte Autorizar conexiones mediante AWS Lake Formation

Recursos adicionales

Publicaciones de blog

- Habilite permisos detallados para los autores de Quick en AWS Lake Formation
- Analice sus datos de forma segura con Quick AWS Lake Formation

AWS Lake Formation Utilizándolo con AWS CloudTrail Lake

AWS CloudTrail Lake permite explorar los almacenes de datos de eventos utilizando Amazon Athena permisos detallados en. AWS Lake Formation

> **Note**
>
> CloudTrail Solo se puede consultar a Lake. Amazon Athena

Para registrar su almacén de datos de eventos de CloudTrail Lake en Lake Formation, consulte Federar un almacén de datos de eventos.

Registro de llamadas a la API de AWS Lake Formation mediante AWS CloudTrail

AWS Lake Formation está integrado con AWS CloudTrail un servicio que proporciona un registro de las acciones realizadas por un usuario, función o AWS servicio en Lake Formation. CloudTrail captura todas las llamadas a la API de Lake Formation como eventos. Las llamadas capturadas incluyen llamadas desde la consola de Lake Formation AWS Command Line Interface, y llamadas de código a las acciones de la API de Lake Formation. Si crea una ruta, puede habilitar la entrega continua de CloudTrail eventos a un bucket de Amazon S3, incluidos los eventos de Lake Formation. Si no configura una ruta, podrá ver los eventos más recientes en la CloudTrail consola, en el historial de eventos. Con la información recopilada por CloudTrail, puede determinar la solicitud que se realizó a Lake Formation, la dirección IP desde la que se realizó la solicitud, quién la hizo, cuándo se realizó y detalles adicionales.

Para obtener más información CloudTrail, consulte la Guía AWS CloudTrail del usuario.

Lea acerca de Lake Formation en CloudTrail

CloudTrail está activado de forma predeterminada al crear una AWS cuenta nueva. Cuando se produce una actividad en Lake Formation, esa actividad se registra como un CloudTrail evento junto con otros eventos de AWS servicio en el historial de eventos. Un evento representa una solicitud específica realizada desde una fuente y contiene información sobre la acción solicitada, la fecha y la hora de la acción y los parámetros de la solicitud. Cada entrada de registro o evento contiene información sobre quién generó la solicitud. La información de identidad del usuario le ayuda a determinar lo siguiente:

- Si la solicitud se realizó con credenciales de usuario root o AWS Identity and Access Management (IAM).
- Si la solicitud se realizó con credenciales de seguridad temporales de un rol o fue un usuario federado.
- Si la solicitud la realizó otro AWS servicio.

Para obtener más información, consulte el elemento userIdentity de CloudTrail .

Puedes ver, buscar y descargar los eventos recientes de tu AWS cuenta. Para obtener más información, consulta Cómo ver eventos con el historial de CloudTrail eventos.

Para tener un registro continuo de los eventos en tu AWS cuenta, incluidos los eventos de Lake Formation, crea un sendero. Un rastro permite CloudTrail entregar archivos de registro a un bucket de Amazon S3. De forma predeterminada, cuando se crea un registro de seguimiento en la consola, este se aplica a todas las regiones de AWS . La ruta registra los eventos de todas las regiones de la AWS partición y envía los archivos de registro al bucket de Amazon S3 que especifique. Además, puede configurar otros AWS servicios, por ejemplo, para analizar más a fondo los datos de eventos recopilados en los CloudTrail registros y actuar en función de ellos. Amazon Athena CloudTrail también puede enviar archivos de registro a Amazon CloudWatch Logs and CloudWatch Events.

Para más información, consulte los siguientes temas:

- Introducción a la creación de registros de seguimiento
- CloudTrail servicios e integraciones compatibles
- Configuración de las notificaciones de Amazon SNS para CloudTrail
- Recibir archivos de CloudTrail registro de varias regiones y recibir archivos de CloudTrail registro de varias cuentas

Descripción de los eventos de Lake Formation

Todas las acciones de la API Lake Formation se registran CloudTrail y se documentan en la Guía para AWS Lake Formation desarrolladores. Por ejemplo, las llamadas a las `RevokePermissions` acciones y `PutDataLakeSettingsGrantPermissions`, generan entradas en los archivos de CloudTrail registro.

En el siguiente ejemplo, se muestra un CloudTrail evento de la `GrantPermissions` acción. La entrada incluye el usuario que concedió el permiso (`datalake_admin`), la entidad principal a la que se concedió el permiso (`datalake_user1`) y el permiso que se concedió (`CREATE_TABLE`). La entrada también muestra que la concesión ha fallado porque la base de datos de destino no estaba especificada en el argumento de `resource`.


```
{
    "eventVersion": "1.08",
    "userIdentity": {
        "type": "IAMUser",
        "principalId": "AIDAZKE67KM3P775X74U2",
        "arn": "arn:aws:iam::111122223333:user/datalake_admin",
        "accountId": "111122223333",
        "accessKeyId": "...",
```

```
        "userName": "datalake_admin"
    },
    "eventTime": "2021-02-06T00:43:21Z",
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GrantPermissions",
    "awsRegion": "us-east-1",
    "sourceIPAddress": "72.21.198.65",
    "userAgent": "aws-cli/1.19.0 Python/3.6.12
 Linux/4.9.230-0.1.ac.223.84.332.metal1.x86_64 botocore/1.20.0",
    "errorCode": "InvalidInputException",
    "errorMessage": "Resource must have one of the have either the catalog, table or
 database field populated.",
    "requestParameters": {
        "principal": {
            "dataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
        },
        "resource": {},
        "permissions": [
            "CREATE_TABLE"
        ]
    },
    "responseElements": null,
    "requestID": "b85e863f-e75d-4fc0-9ff0-97f943f706e7",
    "eventID": "8d2ccef0-55f3-42d3-9ede-3a6faedaa5c1",
    "readOnly": false,
    "eventType": "AwsApiCall",
    "managementEvent": true,
    "eventCategory": "Management",
    "recipientAccountId": "111122223333"
}
```

El siguiente ejemplo muestra una entrada de CloudTrail registro de la `GetDataAccess` acción. Las entidades principales no llaman directamente a esta API. Más bien, `GetDataAccess` se registra cada vez que un AWS servicio principal o integrado solicita credenciales temporales para acceder a los datos en una ubicación de lago de datos registrada en Lake Formation.

```
{
    "eventVersion": "1.05",
    "userIdentity": {
        "type": "AWSAccount",
        "principalId": "AROAQGFTBBBGOBWV2EMZA:GlueJobRunnerSession",
        "accountId": "111122223333"
```

```
    },
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GetDataAccess",
...
...
    "additionalEventData": {
        "requesterService": "GLUE_JOB",
        "lakeFormationPrincipal": "arn:aws:iam::111122223333:role/ETL-Glue-Role",
        "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-G13T0Rmng2"
    },
...
}
```

Rastrea el acceso a los datos CloudTrail con Lake Formation

Lake Formation emite credenciales temporales con alcance limitado para los motores de consulta, como Amazon EMR, AWS Glue y para acceder a las ubicaciones de los datos de las tablas en Amazon S3. Lake Formation utiliza la función de IAM que se especifica al registrar una ubicación de datos en Lake Formation para vender estas credenciales.

Cambio en el comportamiento de CloudTrail registro (febrero de 2026)

Antes de febrero de 2026, cuando un motor de consultas hacía una `s3:GetObject` llamada para obtener datos de consulta, los CloudTrail registros mostraban la función de IAM de registro de Lake Formation como principal.

A partir de febrero de 2026, CloudTrail los registros de `s3:GetObject` las llamadas también incluyen el identificador único de IAM (que comienza porAROA) de la función de IAM que inició la consulta. Este identificador corresponde al rol de IAM que inició la consulta, al trabajo de Amazon EMR AWS Glue o al trabajo de ETL.

Seguimiento del acceso a los datos desde el inicio de la consulta hasta Amazon S3

Cada `s3:GetObject` llamada entrante CloudTrail corresponde a una `lakeformation:GetDataAccess` llamada. Puede usar el identificador de rol de IAM que está presente en ambas entradas de registro para rastrear el rol de IAM de la persona que llama desde el inicio de la consulta hasta el acceso a los datos en Amazon S3.

Para habilitar la información de identidad de origen en los CloudTrail registros de eventos de datos de Amazon S3, siga los pasos que se indican a continuación. Para habilitar los eventos de datos de

Amazon S3 CloudTrail, consulte Habilitar el registro de CloudTrail eventos para buckets y objetos de Amazon S3.

> **Important**
>
> La activación de eventos de datos de Amazon S3 CloudTrail conlleva cargos adicionales.

Para habilitar el seguimiento de la identidad de origen en los registros CloudTrail

1. Utilizando, ejecute el `get-data-lake-settings` comando. En la respuesta, edite el `Parameters` campo que desee añadir`"SET_SOURCE_IDENTITY": "TRUE"`.

   Ejecute el comando `put-data-lake-settings`.

   ```
   "Parameters": {
       "CROSS_ACCOUNT_VERSION": "4",
       "SET_SOURCE_IDENTITY": "TRUE"
   },
   ```

2. En la política de confianza del rol de IAM utilizado para registrar su ubicación de datos de Amazon S3 en Lake Formation, añada`sts:SetSourceIdentity`:

   ```
   {
       "Version": "2012-10-17",
       "Statement": [
           {
               "Effect": "Allow",
               "Principal": {
                   "Service": [
                       "lakeformation.amazonaws.com"
                   ]
               },
               "Action": [
                   "sts:AssumeRole",
                   "sts:SetSourceIdentity"
               ]
           }
       ]
   }
   ```

El siguiente registro muestra la información de identidad de origen en un `s3:GetObject` CloudTrail registro de ejemplo.

```
"sessionContext": {
    "sessionIssuer": {
        "type": "Role",
        "principalId": "AROA2EXAMPLEXYZ3AB6CDKLM",
        "arn": "arn:aws:iam::111122223333:role/LFRegistrationRole",
        "accountId": "111122223333",
        "userName": "LFRegistrationRole"
    },
    "sourceIdentity": "V2.111122223333.AROA3EXAMPLEPQR7ST9UVWXY"
},
"invokedBy": "glue.amazonaws.com"
```

Véase también

- Para obtener información sobre el seguimiento del contexto de usuario del IAM Identity Center en CloudTrail los registros, consulteIntegración de IAM Identity Center.

Supervisión del acceso a los catálogos federados de Amazon Redshift

Los catálogos federados de Amazon Redshift incluidos en los registros AWS Glue Data Catalog no generan eventos en `GetDataAccess` Lake Formation. CloudTrail Para realizar un seguimiento del acceso a los datos de las tablas de Amazon Redshift en el Catálogo de datos, supervise los eventos `GetTable` y `BatchGetTable`, que capturan los patrones de acceso a los metadatos de las tablas federadas de Redshift, incluidos el nombre de la tabla, el nombre de la base de datos y la información del ID del catálogo.

Véase también

- Cross-account CloudTrail registro

Prácticas recomendadas, consideraciones y limitaciones de Lake Formation

Utilice esta sección para encontrar rápidamente las prácticas recomendadas, las consideraciones y las limitaciones de AWS Lake Formation.

Consulte Service Quotas para conocer el número máximo de recursos u operaciones de servicio de su Cuenta de AWS.

Temas

- Prácticas recomendadas y consideraciones para uso compartido de datos entre cuentas
- Limitaciones de los roles vinculados a servicios
- Limitaciones de acceso a datos entre regiones
- Vistas, consideraciones y limitaciones del catálogo de datos
- Limitaciones de filtrado de datos
- Consideraciones y limitaciones del modo de acceso híbrido
- Limitaciones a la hora de incorporar los datos del almacén de datos de Amazon Redshift al AWS Glue Data Catalog
- Limitaciones de integración del catálogo de S3 Tables
- Consideraciones y limitaciones del uso compartido de datos del almacén de metadatos de Hive
- Limitaciones de uso compartido de datos de Amazon Redshift
- Limitaciones de la integración de IAM Identity Center
- Prácticas recomendadas y consideraciones sobre el control de acceso basado en etiquetas de Lake Formation
- Consideraciones, limitaciones y regiones compatibles para el control de acceso basado en atributos

Prácticas recomendadas y consideraciones para uso compartido de datos entre cuentas

Las capacidades multicuenta de Lake Formation permiten a los usuarios compartir de forma segura lagos de datos distribuidos entre varias AWS organizaciones o directamente con los directores de

IAM en otra cuenta Cuentas de AWS, lo que proporciona un acceso detallado a los metadatos del catálogo de datos y a los datos subyacentes.

Tenga en cuenta las siguientes prácticas recomendadas en el uso compartido de datos entre cuentas de Lake Formation:

- No hay límite en cuanto a la cantidad de permisos de Lake Formation que puede conceder a los directores por su propia AWS cuenta. Sin embargo, Lake Formation usa la capacidad AWS Resource Access Manager (AWS RAM) para las concesiones entre cuentas que su cuenta puede realizar con el método de recurso indicado. Para maximizar la AWS RAM capacidad, sigue estas prácticas recomendadas para el método de recurso indicado:
  - Usa el nuevo modo de concesión multicuenta (versión 3 y superior en la configuración de la versión multicuenta) para compartir un recurso con una persona externa Cuenta de AWS. Para obtener más información, consulte Actualización de los ajustes de la versión entre cuentas para compartir datos.
  - Organice las AWS cuentas en organizaciones y conceda permisos a las organizaciones o unidades organizativas. Una concesión a una organización o unidad organizativa cuenta como una concesión.

    La concesión a organizaciones o unidades organizativas también elimina la necesidad de aceptar una AWS Resource Access Manager (AWS RAM) invitación a compartir recursos para obtener la subvención. Para obtener más información, consulte Acceso y visualización de tablas y bases de datos compartidas del Catálogo de datos.
  - En lugar de conceder permisos en muchas tablas individuales de una base de datos, utilice el comodín especial Todas las tablas para conceder permisos en todas las tablas de la base de datos. La concesión en Todas las tablas se considera una única concesión. Para obtener más información, consulte Concesión de permisos sobre los recursos del Catálogo de datos.

  > **Note**
  >
  > Para obtener más información sobre cómo solicitar un límite superior para el número de recursos compartidos AWS RAM, consulte las cuotas de AWS servicio en el Referencia general de AWS.

- Debe crear un enlace de recursos a una base de datos compartida para que dicha base de datos aparezca en los editores de consultas Amazon Athena y en Amazon Redshift Spectrum. Del mismo modo, para poder consultar tablas compartidas con Athena y Redshift Spectrum, debe

crear enlaces de recursos a las tablas. Los enlaces de recursos aparecen entonces en la lista de tablas de los editores de consultas.

En lugar de crear enlaces de recursos para hace consultas en muchas tablas por separado, puede utilizar el comodín Todas las tablas para conceder permisos en todas las tablas de una base de datos. A continuación, cuando cree un enlace de recursos para esa base de datos y lo seleccione en el editor de consultas, tendrá acceso a todas las tablas de esa base de datos para su consulta. Para obtener más información, consulte Creación de enlaces de recursos.

- Cuando comparte recursos directamente con entidades principales de otra cuenta, es posible que la entidad principal de IAM de la cuenta de destinatario no tenga permiso para crear enlaces de recursos para poder consultar las tablas compartidas mediante Athena y Amazon Redshift Spectrum. En lugar de crear un enlace de recursos para cada tabla compartida, el administrador del lago de datos puede crear una base de datos de marcadores de posición y conceder permisos de `CREATE_TABLE` al grupo `ALLIAMPrincipal`. A continuación, todas las entidades principales de IAM de la cuenta de destinatario pueden crear enlaces de recursos en la base de datos del marcador de posición y empezar a consultar las tablas compartidas.

  Consulte el comando CLI de ejemplo para conceder permisos a `ALLIAMPrincipals` en Concesión de permisos de base de datos mediante el método de recurso con nombre.
- Cuando se otorgan permisos entre cuentas directamente a una entidad principal, solo el destinatario puede verlos. El administrador del lago de datos de la cuenta de AWS del destinatario no puede ver estos permisos concedidos directamente.
- Athena y Redshift Spectrum admiten el control de acceso a nivel de columna, pero solo para la inclusión, no para la exclusión. Los trabajos de ETL de AWS Glue no admiten el control de acceso a nivel de columna.
- Cuando se comparte un recurso con su AWS cuenta, solo puede conceder permisos sobre el recurso a los usuarios de su cuenta. No puedes conceder permisos sobre el recurso a otras AWS cuentas, a organizaciones (ni siquiera a la tuya) ni al `IAMAllowedPrincipals` grupo.
- No puede conceder `DROP` ni `Super` sobre una base de datos a una cuenta externa.
- Revoque los permisos entre cuentas antes de eliminar una base de datos o una tabla. De lo contrario, debe eliminar los recursos compartidos huérfanos en AWS Resource Access Manager.

Véase también

- Prácticas recomendadas y consideraciones sobre el control de acceso basado en etiquetas de Lake Formation
- `CREATE_TABLE` en Referencia de permisos de Lake Formation para más normas y limitaciones de acceso entre cuentas.

Limitaciones de los roles vinculados a servicios

Un rol vinculado a un servicio es un tipo especial de rol de IAM al que se vincula directamente. AWS Lake Formation Esta función tiene permisos predefinidos que permiten a Lake Formation realizar acciones en su nombre en todos AWS los servicios.

Las siguientes limitaciones se aplican cuando se utiliza un rol vinculado a un servicio (SLR) para registrar ubicaciones de datos en Lake Formation.

- Las políticas de los roles vinculados a servicios no se pueden modificar una vez creadas.
- Un rol vinculado a un servicio no admite el uso compartido de recursos de catálogo cifrados entre cuentas. Los recursos cifrados requieren permisos AWS KMS clave específicos. Los roles vinculados a servicios tienen permisos predefinidos que no incluyen la posibilidad de trabajar con recursos de catálogo cifrados en todas las cuentas.
- Al registrar varias ubicaciones de Amazon S3, el uso de un rol vinculado al servicio puede hacer que se superen rápidamente los límites de la política de IAM. Esto se debe a que, en el caso de los roles vinculados a un servicio, AWS redacta la política por ti y se va incrementando en un bloque grande que incluye todos tus registros. Puede redactar políticas administradas por el cliente de manera más eficiente, distribuir los permisos entre varias políticas o usar diferentes roles para diferentes regiones.
- Amazon EMR en EC2 no puede acceder a los datos cuando se registran las ubicaciones de datos con roles vinculados a servicios.
- Las operaciones de roles vinculados al servicio eluden tus políticas de control de servicios. AWS
- Cuando se registran ubicaciones de datos con un rol vinculado a un servicio, se actualizan las políticas de IAM con coherencia final. Para obtener más información, consulte la documentación sobre Solución de problemas de IAM en la Guía del usuario de IAM.
- No puede establecer `SET_CONTEXT = TRUE` en la configuración del lago de datos de Lake Formation cuando se utilizan roles vinculados a servicios e IAM Identity Center. La razón es que

los roles vinculados a servicios tienen políticas de confianza inmutables que son incompatibles con la propagación de identidades de confianza que se requiere para la auditoría de `SetContext` con las entidades principales de IAM Identity Center.

Limitaciones de acceso a datos entre regiones

Lake Formation admite las consultas a tablas de catálogo de datos entre Regiones de AWS. Puede acceder a los datos de una región desde otras regiones mediante Amazon Athena Amazon EMR y AWS Glue ETL creando enlaces de recursos en otras regiones que apunten a las bases de datos y tablas de origen. Con el acceso a las tablas entre regiones, puede acceder a los datos de todas las regiones sin copiar los datos subyacentes o los metadatos en el Catálogo de datos.

Se aplican las siguientes limitaciones al acceso a las tablas entre regiones.

- Lake Formation no admite la consulta de tablas del Catálogo de datos de otra región mediante Amazon Redshift Spectrum.
- En la consola de Lake Formation, las vistas de bases de datos y tablas no muestran los nombres de las bases de datos o tablas de la región de origen.
- Para ver la lista de tablas de una base de datos compartida de otra región, primero debe crear un enlace de recursos a la base de datos compartida, luego seleccionar el enlace de recursos y elegir Ver tablas.
- Lake Formation no admite llamadas de enlaces de recursos entre regiones realizadas por usuarios de SAML.
- La característica entre regiones de Lake Formation no implica cargos adicionales por la transferencia de datos.

Vistas, consideraciones y limitaciones del catálogo de datos

Las siguientes consideraciones y limitaciones se aplican a las vistas del catálogo de datos.

- No puede crear una vista del Catálogo de datos desde la consola de Lake Formation. Puede crear vistas mediante el SDK AWS CLI o el SDK.
- Puede crear vistas del Catálogo de datos a partir de 10 tablas. Es un límite invariable. Las tablas de referencia subyacentes de una vista pueden pertenecer a la misma base de datos o a bases de datos diferentes dentro de la misma AWS cuenta.

- Para conocer las consideraciones y limitaciones adicionales específicas de la creación de vistas del Catálogo de datos con Redshift, consulte la sección Vistas, consideraciones y limitaciones del Catálogo de datos de la Guía para desarrolladores de bases de datos de Amazon Redshift. Para Athena, consulte la sección Vistas, consideraciones y limitaciones del Catálogo de datos de la Guía del usuario de Amazon Athena.
- Puede crear vistas del Catálogo de datos en tablas registradas en Lake Formation tanto en el modo de acceso híbrido como en el modo Lake Formation.

  Cuando se utilicen las vistas del Catálogo de datos con el modo de acceso híbrido de Lake Formation, se recomienda asegurarse de que las entidades principales que consuman las vistas tengan habilitados permisos de Lake Formation para las tablas base a las que se hace referencia en la vista sin conceder acceso. Esto garantiza que las tablas base no se revelen a los consumidores a través de los permisos de AWS Glue IAM.
- No hay restricciones en la versión para compartir vistas entre cuentas.
- Las vistas se versionan de modo idéntico que las tablas del Catálogo de datos cuando se usa la instrucción `ALTER VIEW` para un dialecto de vista ya creado. No se puede restaurar una vista anterior porque la versión de la vista cambia con los cambios de datos subyacentes. Al eliminar una versión de vista, esta pasará de forma predeterminada a la siguiente versión más reciente disponible. Al cambiar la versión de la vista, asegúrese de que los datos estén sincronizados con el esquema de la versión de vista seleccionada.
- No se introduce ningún catálogo de APIs datos nuevo. Los existentes `CreateTable` `DeleteTable` y `UpdateTable` `GetTable` APIs se actualizan.
- Amazon Redshift siempre crea vistas con columnas varchar a partir de tablas con cadenas. Debe convertir columnas de cadenas en varchar con una longitud explícita al añadir dialectos de otros motores.
- Al conceder permisos de lago de datos a `All tables` dentro de una base de datos, el beneficiario tendrá permisos en todas las tablas y vistas de la base de datos.
- No se pueden crear vistas:
  - Que hagan referencia a otras vistas.
  - Cuando la tabla de referencia sea un enlace de recurso.
  - Cuando la tabla de referencia esté en otra cuenta.
  - De almacenes de metadatos externos de Hive.
- Las funciones de definición multicuenta no son compatibles con las vistas de Redshift Spectrum Dialect.

- No se admiten los enlaces de recursos para el dialecto de Athena en el editor de consultas de Athena. Para utilizar funciones de definición multicuenta para el dialecto de Atenea, añada la cuenta que aloja las tablas base como fuente de datos en Athena.

Limitaciones de filtrado de datos

Al conceder permisos de Lake Formation en una tabla del Catálogo de datos, puede incluir especificaciones de filtrado de datos para restringir el acceso a determinados datos en los resultados de las consultas y en los motores integrados con Lake Formation. Lake Formation utiliza el filtrado de datos para ofrecer una seguridad a nivel de columna, fila y celda. Puede definir y aplicar filtros de datos en las columnas anidadas si los datos de origen contienen estructuras anidadas.

Notas y restricciones para el filtrado a nivel de columna

Hay tres formas de especificar el filtrado de columnas:

- Mediante el uso de filtros de datos
- Mediante el uso de un filtrado de columnas simple o un filtrado de columnas anidadas.
- TAGsMediante el uso de.

El filtrado de columnas simple solo especifica una lista de columnas para incluir o excluir. Tanto la consola Lake Formation como la API AWS CLI admiten un filtrado de columnas sencillo. Para ver un ejemplo, consulta Grant with Simple Column Filtering.

Las siguientes notas y restricciones corresponden al filtrado de columnas:

- AWS Glue La versión 5.0 o superior admite un control de acceso detallado a través de Lake Formation solo para las tablas Apache Hive y Apache Iceberg.
- Para conceder `SELECT` con la opción de concesión y el filtrado de columnas, debe usar una lista de inclusión, no una lista de exclusiones. Sin la opción de conceder, puede utilizar listas de inclusión o exclusión.
- Para conceder `SELECT` en una tabla con filtrado de columnas, debe haber obtenido la autorización `SELECT` en la tabla con la opción de concesión y sin restricciones de filas. Debe tener acceso a todas las filas.
- Si concede `SELECT` con la opción de concesión y filtrado de columnas a una entidad principal de su cuenta, dicha entidad principal deberá especificar el filtrado de columnas para las mismas

columnas o un subconjunto de las columnas beneficiarias cuando conceda a otra entidad principal. Si concede `SELECT` con la opción de concesión y filtrado de columnas a una cuenta externa, el administrador del lago de datos de la cuenta externa puede conceder `SELECT` en todas las columnas a otra entidad principal de su cuenta. Sin embargo, incluso con `SELECT` en todas las columnas, esa entidad principal tendrá visibilidad solo en las columnas concedidas a la cuenta externa.

- No puede aplicar el filtrado de columnas a las claves de partición.
- A una entidad principal con el permiso `SELECT` sobre un subconjunto de columnas de una tabla no se le puede conceder el permiso `ALTER`, `DROP`, `DELETE` o `INSERT` sobre esa tabla. Para una entidad principal con el permiso `ALTER`, `DROP`, `DELETE` o `INSERT` en una tabla, si concede el permiso `SELECT` con filtrado de columnas, no tiene ningún efecto.

Las siguientes notas y restricciones se aplican al filtrado de columnas anidadas:

- Puede incluir o excluir cinco niveles de campos anidados en un filtro de datos.

  Example

  Col1.Col1_1.Col1_1_1.Col1_1_1_1.Col1_1_1_1_1
- No puede aplicar el filtrado de columnas a los campos anidados dentro de las columnas de partición.
- Si el esquema de la tabla contiene un nombre de columna de nivel superior («cliente»).» dirección») que tiene el mismo patrón de representación de un campo anidado dentro de un filtro de datos (una columna anidada con un nombre de columna de nivel superior `customer` y un nombre de campo anidado `address` se especifica como `"customer"."address"` en un filtro de datos), no puede especificar explícitamente el acceso a la columna de nivel superior o al campo anidado, ya que ambos se representan con el mismo patrón en las listas. inclusion/exclusion Esto es ambiguo y Lake Formation no puede resolverlo si especifica la columna de nivel superior o el campo anidado.
- Si una columna de nivel superior o un campo anidado contiene comillas dobles dentro del nombre, debe incluir una segunda comilla doble cuando especifique el acceso a un campo anidado en la lista de inclusión y exclusión de un filtro de celdas de datos.

  Example

  Ejemplo de nombre de columna anidado con comillas dobles: `a.b.double"quote`

Example

Ejemplo de representación de columna anidada dentro de un filtro de datos: `"a"."b"."double""quote"`

Limitaciones de filtrado por celdas

Tenga en cuenta las siguientes notas y restricciones para el filtrado a nivel de fila y a nivel de celda:

- La seguridad a nivel de celda no se admite en las columnas, vistas y enlaces de recursos anidados.
- Todas las expresiones que se admiten en las columnas de nivel superior también se admiten en las columnas anidadas. Sin embargo, NO se debe hacer referencia a los campos anidados de las columnas de partición al definir expresiones anidadas a nivel de fila.
- La seguridad a nivel de celda está disponible en todas las regiones cuando se utiliza la versión 3 del motor Athena o Amazon Redshift Spectrum. Para otros servicios, la seguridad a nivel de celda solo está disponible en las regiones mencionadas en la Regiones admitidas.
- No se admiten las instrucciones `SELECT INTO`.
- Los tipos de datos `array` y `map` no se admiten en las expresiones de filtro de filas. Solo se admite el tipo `struct`.
- No hay ningún límite para el número de filtros de datos que se pueden definir en una tabla, pero existe un límite de 100 filtros de datos para una sola entidad principal en una tabla.
- Para aplicar un filtro de datos con una expresión de filtro de filas, debe tener `SELECT` con la opción de conceder en todas las columnas de la tabla. Esta restricción no se aplica a los administradores de cuentas externas cuando la concesión se hizo a la cuenta externa.
- Si una entidad principal es miembro de un grupo y tanto la entidad principal como el grupo tienen concedidos permisos sobre un subconjunto de filas, los permisos efectivos de fila de la entidad principal son la unión de los permisos de la entidad principal y los permisos del grupo.
- Los siguientes nombres de columna están restringidos en una tabla para el filtrado a nivel de fila y de celda:
  - ctid
  - oid
  - xmin
  - cmin

  - xmax
  - cmax
  - tableoid
  - insertxid
  - deletexid
  - importoid
  - redcatuniqueid
- Si aplica la expresión de filtro de todas las filas a una tabla de forma simultánea con otras expresiones de filtro con predicados, la expresión de todas las filas prevalecerá sobre todas las demás expresiones de filtro.
- Cuando se conceden permisos en un subconjunto de filas a una AWS cuenta externa y el administrador del lago de datos de la cuenta externa concede esos permisos a un principal de esa cuenta, el predicado de filtro efectivo del principal es la intersección del predicado de la cuenta y cualquier predicado que se haya otorgado directamente al principal.

  Por ejemplo, si la cuenta tiene permisos de fila con el predicado `dept='hr'` y a la entidad principal se le concedió por separado permiso para `country='us'`, la entidad principal solo tiene acceso a las filas con `dept='hr'` y `country='us'`.

Para obtener más información sobre el filtrado de celdas, consulte Filtrado de datos y seguridad de celda en Lake Formation.

Para conocer las consideraciones y limitaciones a la hora de consultar tablas con Amazon Redshift Spectrum con políticas de seguridad de filas, consulte Consideraciones y limitaciones sobre el uso de las políticas de RLS en la Guía para desarrolladores de bases de datos de Amazon Redshift.

Consideraciones y limitaciones del modo de acceso híbrido

El modo de acceso híbrido proporciona la flexibilidad de habilitar selectivamente los permisos de Lake Formation para las bases de datos y tablas de su AWS Glue Data Catalog.
Con el modo de acceso híbrido, ahora tiene una ruta incremental que le permite establecer los permisos de Lake Formation para un conjunto específico de usuarios sin interrumpir las políticas de permisos de otros usuarios o cargas de trabajo existentes.

Las consideraciones y limitaciones indicadas a continuación se aplican al modo de acceso híbrido.

Limitaciones

- Actualizar el registro de ubicaciones de Amazon S3: no es posible editar los parámetros de una ubicación registrada en Lake Formation mediante un rol vinculado a un servicio.
- Opción de incluir al utilizar etiquetas LF: cuando puede conceder permisos de Lake Formation utilizando etiquetas LF, puede incluir entidades principales para aplicar los permisos de Lake Formation como paso consecutivo eligiendo bases de datos y tablas que tengan etiquetas LF adjuntas.
- Acceso en modo de acceso híbrido: el acceso en modo de acceso híbrido en Lake Formation está limitado a los usuarios con permisos de administrador de lago de datos o de administrador de solo lectura.
- Incluir entidades principales: actualmente, solo un rol de administrador del lago de datos puede incluir a las entidades principales en los recursos.
- Incluir todas las tablas de una base de datos: en concesiones entre cuentas, al conceder permisos e incluir todas las tablas de una base de datos, también es necesario incluir la base de datos para que los permisos funcionen.

Consideraciones

- Actualización de la ubicación de Amazon S3 registrada en Lake Formation al modo de acceso híbrido: no recomendamos convertir una ubicación de datos de Amazon S3 que ya esté registrada en Lake Formation a un modo de acceso híbrido, aunque se puede hacer.
- Comportamientos de la API cuando se registra una ubicación de datos en modo de acceso híbrido
  - CreateTable — La ubicación se considera registrada en Lake Formation independientemente del indicador del modo de acceso híbrido y del estado de suscripción. Por lo tanto, el usuario necesita el permiso de ubicación de datos para crear una tabla.
  - CreatePartition/BatchCreatePartition/UpdatePartition(cuando la ubicación de la partición se actualiza para que apunte a la ubicación registrada en el modo híbrido): la ubicación de Amazon S3 se considera registrada en Lake Formation, independientemente del indicador del modo de acceso híbrido y del estado de activación.
  - CreateDatabase/UpdateDatabase (cuando la ubicación de la base de datos se actualiza para que apunte a la ubicación registrada en el modo de acceso híbrido): la ubicación se considera registrada en Lake Formation independientemente del indicador del modo de acceso híbrido y del estado de activación. Así, el usuario necesita el permiso de localización de datos para crear o actualizar una base de datos.

- UpdateTable (cuando la ubicación de una tabla se actualiza para que apunte a la ubicación registrada en el modo de acceso híbrido): la ubicación se considera registrada en Lake Formation independientemente del indicador del modo de acceso híbrido y del estado de activación. Por lo tanto, el usuario necesita permiso de localización de datos para actualizar la tabla. Si la ubicación de la tabla no está actualizada o apunta a una ubicación que no está registrada en Lake Formation, el usuario no necesitará permiso de ubicación de datos para actualizar la tabla.

Limitaciones a la hora de incorporar los datos del almacén de datos de Amazon Redshift al AWS Glue Data Catalog

Puede catalogar y administrar el acceso a los datos analíticos en los almacenes de datos de Amazon Redshift mediante el AWS Glue Data Catalog. Se aplican las siguientes restricciones:

- El uso compartido entre cuentas no es compatible con el nivel de catálogo federado. Sin embargo, puede compartir bases de datos y tablas individuales desde un catálogo federado entre s. Cuenta de AWS
- Debe tener la configuración de versión de Cross Account versión 4 o superior para compartir bases de datos o tablas del catálogo federado entre s. Cuenta de AWS
- El Catálogo de datos solo admite la creación de catálogos de nivel superior.
- Solo puede actualizar la descripción de los catálogos de Redshift Managed Storage (RMS).
- No se admite la configuración de permisos de catálogos federados, así como de las bases de datos y tablas de los catálogos federados, para el grupo `IAMAllowedPrincipals`.

- No se admiten las operaciones de lenguaje de definición de datos (DDL) en el catálogo, incluido el establecimiento de las configuraciones del catálogo, desde motores como Athena, Amazon EMR Spark u otros.
- No se admiten las operaciones de DDL en las tablas de RMS con Athena.
- No se admite la creación de vistas materializadas, ya sea a través de Athena, Apache Spark, AWS Glue Data Catalog the o Amazon Redshift Consumer.
- Athena no admite una experiencia con varios catálogos. No se puede conectar a más de un catálogo específico a la vez. Athena no puede acceder a varios catálogos ni realizar consultas en ellos simultáneamente.

- No se admiten las operaciones de etiquetado y ramificación en tablas de Iceberg mediante Athena y Amazon Redshift.
- No se admite Viaje en el tiempo en las tablas de RMS.
- No se admiten los catálogos de varios niveles con tablas de lago de datos. Todos los datos almacenados en Amazon S3 para su uso con las tablas de lagos de datos deben residir en la configuración predeterminada AWS Glue Data Catalog y no pueden organizarse en catálogos de varios niveles.
- En Amazon Redshift, los recursos compartidos de datos no se agregan al espacio de nombres registrado. Los clústeres y los espacios de nombres son sinónimos y, una vez que publique un clúster en ellos AWS Glue Data Catalog, no podrá añadir nuevos datos.
- Amazon EMR en EC2 no admite la unión de tablas de RMS y tablas de Amazon S3. Solo EMR sin servidor admite esta capacidad.
- No se admiten tablas y esquemas externos.
- Solo se puede acceder a las tablas de RMS desde el punto de conexión de la extensión del catálogo de REST de Iceberg de AWS Glue .
- No se puede acceder a las tablas de Hive desde motores de terceros conectados al catálogo REST de AWS Glue Iceberg.
- Se admitirá el nivel de aislamiento read_committed en las tablas de RMS a través de Spark.
- Los nombres de las bases de datos Redshift no distinguen entre mayúsculas y minúsculas AWS Glue Data Catalog, están restringidos a 128 caracteres y pueden ser alfanuméricos con guiones (-) y guiones bajos (_).
- Los nombres de los catálogos no distinguen entre mayúsculas y minúsculas y están limitados a 50 caracteres, que pueden ser alfanuméricos e incluir guiones (-) y guiones bajos (_).
- Amazon Redshift no admite el uso de los comandos GRANT y REVOKE al estilo de SQL de Lake Formation para administrar los permisos de acceso en las tablas publicadas en el AWS Glue Data Catalog.
- No se aplicarán las políticas de enmascaramiento dinámico de datos y de seguridad de filas asociadas al clúster de Amazon Redshift productor (de origen). En su lugar, se aplicarán los permisos de acceso definidos en Lake Formation a los datos compartidos.
- No se admiten las operaciones de lenguaje de definición de datos (DDL) y lenguaje de manipulación de datos (DML) en enlaces de tabla.
- Si las palabras clave reservadas no se filtran correctamente con caracteres de escape, se producirán errores.

- No se admite el cifrado de datos en escenarios de varios catálogos.

Limitaciones de integración del catálogo de S3 Tables

Amazon S3 Tables se integra con AWS Glue Data Catalog (Data Catalog) y registra el catálogo como una ubicación de datos de Lake Formation. Puede configurar este registro desde la consola de Lake Formation o mediante el servicio APIs.

> **Note**
>
> Si tiene una política basada en recursos de IAM o S3 Tables que restringe los usuarios y las funciones de IAM en función de las etiquetas principales, debe adjuntar las mismas etiquetas principales a la función de IAM que Lake Formation utiliza para acceder a sus datos de S3 (por ejemplo LakeFormationDataAccessRole,) y conceder a esta función los permisos necesarios. Esto es necesario para que la política de control de acceso basada en etiquetas funcione correctamente con la integración de análisis de tablas de S3.

Las siguientes limitaciones se aplican a la integración del catálogo de tablas S3 con Data Catalog y Lake Formation:

- AWS Glue y Lake Formation no admiten nombres de columnas en mayúsculas y minúsculas y convierten todos los nombres de columnas a minúsculas. Debe comprobar que los nombres de las columnas de las tablas sean únicos al convertirse a minúsculas. Use `customer_id` en lugar de `customerId`. El uso de nombres de columna en mayúsculas y minúsculas solo estaba permitido en la versión preliminar.
- La CreateCatalog API no puede crear buckets de tablas en Amazon S3.
- La SearchTables API no puede buscar en las tablas de S3.

Consideraciones y limitaciones del uso compartido de datos del almacén de metadatos de Hive

Con la federación de AWS Glue Data Catalog metadatos (federación de catálogos de datos), puede conectar el catálogo de datos a metaalmacenes externos que almacenan los metadatos de sus datos de Amazon S3 y administrar de forma segura los permisos de acceso a los datos mediante AWS Lake Formation.

Las siguientes consideraciones y limitaciones se aplican a las bases de datos federadas que se crean a partir de bases de datos de Hive:

Consideraciones

- AWS SAM soporte de aplicaciones: usted es responsable de la disponibilidad de los recursos de la aplicación que se AWS SAM despliega (Amazon API Gateway y de la función Lambda). Asegúrese de que la conexión entre el metabastore de Hive AWS Glue Data Catalog y el metabastore de Hive funcione cuando los usuarios ejecuten consultas.
- Requisito de la versión de almacén de metadatos de Hive: puede crear bases de datos federadas con Apache Hive versión 3 o posterior.
- Requisito de base de datos mapeada: cada base de datos de Hive debe estar mapeada a una nueva base de datos en Lake Formation.
- Soporte de federación a nivel de base de datos: puede conectarse al metaalmacén de Hive a nivel de base de datos.
- Permisos en bases de datos federadas: los permisos aplicados a una base de datos federada o a las tablas de una base de datos federada persisten incluso cuando se elimina una tabla de origen o una base de datos. Cuando se vuelve a crear la base de datos o tabla de origen, no es necesario volver a conceder los permisos. Cuando se elimina una tabla federada con permisos de Lake Formation en el origen, los permisos de Lake Formation siguen siendo visibles y puede revocarlos si es necesario.

  Si un usuario elimina una base de datos federada, se pierden todos sus permisos correspondientes. Si se vuelve a crear la misma base de datos con el mismo nombre, no se recuperarán los permisos de Lake Formation. Los usuarios deberán volver a configurar los permisos nuevos.
- IAMAllowedPermisos de grupos principales en bases de datos federadas: según la`DataLakeSettings`, Lake Formation podría establecer permisos para todas las bases de datos y tablas de un grupo virtual denominado`IAMAllowedPrincipal`. `IAMAllowedPrincipal`Se refiere a todos los directores de IAM que tienen acceso a los recursos del catálogo de datos a través de las políticas principales y las políticas de recursos de IAM. AWS Glue Si estos permisos existen en una base de datos o una tabla, todas las entidades principales tienen acceso a la base de datos o tabla.

  Sin embargo, Lake Formation no permite permisos `IAMAllowedPrincipal` en las tablas de bases de datos federadas. Al crear bases de datos federadas, asegúrese de pasar el parámetro `CreateTableDefaultPermissions` como una lista vacía.

Para obtener más información, consulte Cambiar la configuración predeterminada de su lago de datos.

- Unir tablas en las consultas: puede unir las tablas del metaalmacén de Hive con las tablas nativas del Catálogo de datos para ejecutar consultas.

Limitaciones

- Limitación en la sincronización de metadatos entre el metaalmacén de Hive AWS Glue Data Catalog y el metabastore de Hive: después de establecer la conexión con el metabastore de Hive, debe crear una base de datos federada para sincronizar los metadatos del metaalmacén de Hive con el. AWS Glue Data Catalog Las tablas de la base de datos federada se sincronizan en tiempo de ejecución cuando los usuarios ejecutan consultas.
- Limitación al crear tablas nuevas en una base de datos federada: no podrá crear tablas nuevas en bases de datos federadas.
- Limitación de permisos de datos: la compatibilidad con los permisos en las vistas de tabla del metaalmacén de Hive no está disponible.

Limitaciones de uso compartido de datos de Amazon Redshift

AWS Lake Formation le permite gestionar de forma segura los datos de un recurso compartido de Amazon Redshift. Amazon Redshift es un servicio de almacenamiento de datos en la nube totalmente gestionado y a escala de petabytes. AWS Al utilizar la capacidad de compartir datos, Amazon Redshift le ayuda a compartir datos entre Cuentas de AWS. Para obtener más información sobre el uso compartido de datos, consulte Información general del uso compartido de datos en Amazon Redshift.

Las notas y restricciones siguientes se aplican a las bases de datos federadas que se crean a partir de recursos compartidos de datos de Amazon Redshift:

- Requisito de base de datos mapeada: cada recurso compartido de datos de Amazon Redshift debe estar mapeado a una nueva base de datos en Lake Formation. Esto es necesario para mantener nombres de tabla únicos cuando la representación de los objetos de recurso compartido de datos está aplanada en la base de datos del Catálogo de datos.
- Limitación al crear tablas nuevas en una base de datos federada: no podrá crear tablas nuevas en bases de datos federadas.

- Permisos en bases de datos federadas: los permisos aplicados a una base de datos federada o a las tablas de una base de datos federada persisten incluso cuando se elimina una tabla de origen o una base de datos. Cuando se vuelve a crear la base de datos o tabla de origen, no es necesario volver a conceder los permisos. Cuando se elimina una tabla federada con permisos de Lake Formation en el origen, los permisos de Lake Formation siguen siendo visibles y puede revocarlos si es necesario.

  Si un usuario elimina una base de datos federada, se pierden todos sus permisos correspondientes. Si se vuelve a crear la misma base de datos con el mismo nombre, no se recuperarán los permisos de Lake Formation. Los usuarios deberán volver a configurar los permisos nuevos.
- IAMAllowedPermisos de grupos principales en bases de datos federadas: según la`DataLakeSettings`, Lake Formation podría establecer permisos para todas las bases de datos y tablas de un grupo virtual denominado`IAMAllowedPrincipal`. `IAMAllowedPrincipal`Se refiere a todos los directores de IAM que tienen acceso a los recursos del catálogo de datos a través de las políticas principales y las políticas de recursos de IAM. AWS Glue Si estos permisos existen en una base de datos o una tabla, todas las entidades principales tienen acceso a la base de datos o tabla.

  Sin embargo, Lake Formation no permite permisos `IAMAllowedPrincipal` en las tablas de bases de datos federadas. Al crear bases de datos federadas, asegúrese de pasar el parámetro `CreateTableDefaultPermissions` como una lista vacía.

  Para obtener más información, consulte Cambiar la configuración predeterminada de su lago de datos.
- Filtrado de datos: en Lake Formation, puede conceder permisos en una tabla de una base de datos federada con filtrado a nivel de columna y de fila. Sin embargo, no puede combinar el filtrado a nivel de columna y de fila para restringir el acceso con una especificidad a nivel de celda a las tablas de bases de datos federadas.
- Identificador de distinción entre mayúsculas y minúsculas: los objetos de recursos compartidos de datos de Amazon Redshift administrados por Lake Formation solo admitirán nombres de tablas y columnas en minúsculas. No active el identificador de distinción entre mayúsculas y minúsculas en las bases de datos, tablas y columnas de los recurso compartido de datos de Amazon Redshift, si se van a compartir y gestionar mediante Lake Formation.
- Compatibilidad con consultas: puede consultar recursos compartidos de datos de Amazon Redshift administrados por Lake Formation con Amazon Redshift. Athena no admite la consulta de recursos compartidos de datos de Amazon Redshift administrados por Lake Formation.

Para obtener más información acerca de las limitaciones de uso de los recursos compartidos de datos en Amazon Redshift, consulte Limitaciones del uso compartido de datos en la Guía para desarrolladores de bases de datos Amazon Redshift.

Limitaciones de la integración de IAM Identity Center

Con él AWS IAM Identity Center, puede conectarse a los proveedores de identidad (IdPs) y administrar de forma centralizada el acceso de los usuarios y grupos a todos AWS los servicios de análisis. Puede configurarlo AWS Lake Formation como una aplicación habilitada en el Centro de identidades de IAM, y los administradores del lago de datos pueden conceder permisos detallados sobre los recursos a los usuarios y grupos autorizados. AWS Glue Data Catalog

Se aplican las siguientes limitaciones a la integración de Lake Formation con IAM Identity Center:

- No puede asignar usuarios y grupos de IAM Identity Center como administradores de lagos de datos o administradores de solo lectura en Lake Formation.

  Los usuarios y grupos del IAM Identity Center pueden consultar los recursos del catálogo de datos cifrados si utilizan una función de IAM que AWS Glue pueda asumir en su nombre el cifrado y descifrado del catálogo de datos. AWS las claves administradas no admiten la propagación de identidades de forma fiable.
- Los usuarios y grupos de IAM Identity Center solo pueden invocar las operaciones de API que figuran en la política `AWSIAMIdentityCenterAllowListForIdentityContext` proporcionada por el IAM Identity Center.
- Lake Formation permite que los roles de IAM de cuentas externas actúen como roles de operador en nombre de los usuarios y grupos de IAM Identity Center para acceder a los recursos del Catálogo de datos, pero solo se pueden conceder permisos a recursos del Catálogo de datos de la cuenta propietaria. Si intenta conceder permisos a usuarios y grupos de IAM Identity Center sobre recursos del Catálogo de datos de una cuenta externa, Lake Formation devuelve el siguiente error: “Cross-account grants are not supported for the principal”.
- Cuando se utiliza Lake Formation con IAM Identity Center, la configuración de asignaciones de aplicaciones se establece en `false` de forma predeterminada. Si modifica esta configuración directamente a través de la API de IAM Identity Center, debe administrar todas las asignaciones de aplicaciones de forma manual mediante la API. Lake Formation no sincroniza ni administra automáticamente los cambios de asignación realizados fuera de sus flujos de trabajo estándar, lo que puede afectar a los patrones de acceso y los flujos de autorización en su entorno de lago de datos.

Prácticas recomendadas y consideraciones sobre el control de acceso basado en etiquetas de Lake Formation

Puede crear, mantener y asignar etiquetas LF para controlar el acceso a las bases de datos, tablas y columnas del catálogo de datos.

Tenga en cuenta las siguientes prácticas recomendadas al usar el control de acceso basado en etiquetas de Lake Formation:

- Todas las etiquetas LF deben estar predefinidas antes de poder asignarlas a los recursos del Catálogo de datos o concederse a las entidades principales.

  El administrador del lago de datos puede delegar las tareas de administración de etiquetas generando creadores de etiquetas LF con los permisos de IAM necesarios. Los ingenieros y analistas de datos deciden las características y las relaciones de las etiquetas LF. Luego, los creadores de las etiquetas LF crean y mantienen las etiquetas LF en Lake Formation.
- Puede asignar varias etiquetas LF a los recursos del Catálogo de datos. Solo se puede asignar un valor para una clave en particular a un recurso concreto.

  Por ejemplo, puede asignar `module=Orders`, `region=West`, `division=Consumer`, etc. a una base de datos, tabla o columna. No puede asignar `module=Orders,Customers`.
- No puede asignar etiquetas LF a los recursos al crearlos. Solo puede añadir etiquetas LF a los recursos existentes.
- Puede conceder expresiones de etiquetas LF, no solo etiquetas LF individuales, a una entidad principal.

  Una expresión LF-Tag tiene el aspecto siguiente (en pseudocódigo).

  ```
  module=sales AND division=(consumer OR commercial)
  ```

  Una entidad principal a la que se le conceda esta expresión de etiqueta LF solo puede acceder a los recursos del Catálogo de datos (bases de datos, tablas y columnas) que tienen asignado `module=sales` y `division=consumer` o `division=commercial`. Si desea que la entidad principal pueda acceder a los recursos que tienen `module=sales` o `division=commercial`, no incluya ambos en la misma concesión. Haga dos concesiones, una para `module=sales` y otra para `division=commercial`.

La expresión de etiqueta LF más sencilla consiste en una sola etiqueta LF, como `module=sales`.

- Una entidad principal que reciba permisos sobre una etiqueta LF con varios valores puede acceder a los recursos del Catálogo de datos con cualquiera de esos valores. Por ejemplo, si a un usuario se le concede una etiqueta LF con clave= `module` y valores= `orders,customers`, el usuario tiene acceso a los recursos que están asignados o bien a `module=orders` o `module=customers`.
- Debe tener permiso `Grant with LF-Tag expressions` para conceder permisos de datos sobre los recursos del Catálogo de datos mediante el método LF-TBAC. El administrador del lago de datos y el creador de la etiqueta LF reciben este permiso de forma implícita. La entidad principal que tenga el permiso `Grant with LFTag expressions` puede conceder permisos de datos sobre los recursos mediante:
  - el método de recurso con nombre
  - el método LF-TBAC, pero solo usando la misma expresión de etiqueta LF

  Por ejemplo, supongamos que el administrador del lago de datos hace la siguiente concesión (en pseudocódigo).

  ```
  GRANT (SELECT ON TABLES) ON TAGS module=customers, region=west,south TO user1 WITH
   GRANT OPTION
  ```

  En este caso, el `user1` puede conceder `SELECT` en tablas a otras entidades principales mediante el método LF-TBAC, pero solo con la expresión de etiqueta `module=customers, region=west,south` LF completa.
- Si a una entidad principal recibe permisos sobre un recurso con el método LF-TBAC y con el método de recurso con nombre, los permisos que dicha entidad tiene sobre el recurso son la unión de los permisos concedidos por ambos métodos.
- Lake Formation admite conceder `DESCRIBE` y `ASSOCIATE` en etiquetas LF entre cuentas, y conceder permisos sobre los recursos del Catálogo de datos en todas las cuentas mediante el método LF-TBAC. En ambos casos, el principal es un identificador de AWS cuenta.

> **Note**
>
> Lake Formation ahora es compatible con la concesión de permisos entre cuentas a organizaciones y unidades organizativas utilizando el método LF-TBAC. Para utilizar esta prestación, debe actualizar la configuración de la versión entre cuentas a la Versión 3.

Para obtener más información, consulte Cross-account intercambio de datos en Lake Formation.

- Los recursos del Catálogo de datos creados en una cuenta solo se pueden etiquetar con etiquetas LF creadas en la misma cuenta. Las etiquetas LF creadas en una cuenta no se pueden asociar a los recursos compartidos de otra cuenta.
- El uso del control de acceso basado en etiquetas (LF-TBAC) de Lake Formation para conceder acceso entre cuentas a los recursos del catálogo de datos requiere adiciones a la política de recursos del catálogo de datos de su cuenta. AWS Para obtener más información, consulte Requisitos previos.
- Las claves y los valores de las etiquetas LF no pueden superar los 50 caracteres.
- La cantidad máxima de etiquetas LF que puede asignar a un recurso del Catálogo de datos es 50.
- Los siguientes límites son flexibles:
  - El número máximo de etiquetas LF que se pueden crear es de 10 000.
  - El número máximo de valores que se pueden definir para una etiqueta LF es 1000.
- Las etiquetas, las claves y los valores se convierten en minúsculas cuando se almacenan.
- Solo se puede asignar un valor para una etiqueta LF a un recurso concreto.
- Si se conceden varias etiquetas LF a una entidad principal con una sola concesión, la entidad principal solo podrá acceder a los recursos del Catálogo de datos que tengan todas las etiquetas LF.
- Si la evaluación de una expresión de etiqueta LF da como resultado el acceso solo a un subconjunto de columnas de la tabla, pero el permiso Lake Formation que se concede cuando hay una coincidencia es uno de los permisos que requiere acceso completo a las columnas, es decir `Alter`, `Drop`, `Insert` o `Delete`, entonces no se concede ninguno de esos permisos. En su lugar, solo se concede `Describe`. Si el permiso concedido es `All` (`Super`), solo se conceden `Select` y `Describe`.
- No se utilizan caracteres comodín con las etiquetas LF. Para asignar una etiqueta LF a todas las columnas de una tabla, debe asignar la etiqueta LF a la tabla y todas las columnas de la tabla heredarán la etiqueta LF. Para asignar una etiqueta LF a todas las tablas de una base de datos, debe asignar la etiqueta LF a la base de datos y todas las tablas de la base de datos heredarán esa etiqueta LF.
- Puede crear hasta 1000 expresiones de etiquetas LF en una cuenta.
- Puede usar hasta 50 expresiones de etiquetas LF para conceder permisos a una entidad principal en los recursos del Catálogo de datos.

- Al conceder o revocar permisos en una expresión de etiqueta LF integrada, el tamaño de la expresión de etiqueta LF no puede superar los 900 bytes. Para conceder permisos en expresiones de etiquetas LF de mayor tamaño, utilice las expresiones de etiquetas LF guardadas. Para obtener más información, consulte Crear expresiones LF-Tag.
- Para añadir la etiqueta LF a los catálogos federados de Redshift existentes que se crearon antes de la versión de disponibilidad general de la compatibilidad con etiquetas LF para catálogos federados, debe ponerse en contacto con el equipo de soporte para obtener ayuda. AWS

Consideraciones, limitaciones y regiones compatibles para el control de acceso basado en atributos

Las consideraciones y limitaciones indicadas a continuación se aplican al control de acceso basado en atributos (ABAC).

- ABAC no admite la concesión de acceso mediante políticas de etiquetas LF.
- Los permisos que se pueden conceder no están disponibles con ABAC.
- ABAC no admite la concesión de permisos a los usuarios de IAM Identity Center.
- Cuando se utilizan las concesiones de permisos de ABAC en una tabla de Lake Formation, Lake Formation no concede permisos `DESCRIBE` en la base de datos o catálogo principal. Esto difiere de los escenarios ajenos a ABAC, en los que Lake Formation proporciona permisos `DESCRIBE` implícitos a los recursos principales.
- Todas las entidades principales con la clave de etiqueta `AmazonDataZoneProject` siempre se consideran parte de Lake Formation para todos los recursos del Catálogo de datos.
- ABAC solo admite atributos de cadena.

Solución de problemas de Lake Formation

Si tiene problemas al trabajar con AWS Lake Formation, consulte los temas de esta sección.

Temas

- Solución de problemas generales
- Solución de problemas de acceso entre cuentas
- Solución de problemas de esquemas y flujos de trabajo
- Problemas conocidos de AWS Lake Formation
- Mensaje de error actualizado

Solución de problemas generales

Utilice esta información para diagnosticar y solucionar diversos problemas de Lake Formation.

Error: permisos insuficientes para Lake Formation en <ubicación de Amazon S3>.

Se ha intentado crear o modificar un recurso del Catálogo de datos sin permisos de ubicación de datos en la ubicación de Amazon S3 a la que apunta el recurso.

Si una base de datos o tabla del Catálogo de datos apunta a una ubicación de Amazon S3, cuando conceda los permisos Lake Formation `CREATE_TABLE` o `ALTER`, también deberá conceder el permiso `DATA_LOCATION_ACCESS` en la ubicación. Si concede estos permisos a cuentas externas o a organizaciones, debe incluir la opción de concesión.

Una vez concedidos estos permisos a una cuenta externa, el administrador del lago de datos de dicha cuenta deberá conceder los permisos a las entidades principales (usuarios o roles) de la cuenta. Al conceder el `DATA_LOCATION_ACCESS` permiso recibido de otra cuenta, debe especificar el identificador de catálogo (identificador de AWS cuenta) de la cuenta del propietario. La cuenta propietaria es la cuenta que registró la ubicación.

Para obtener más información, consulte Control de acceso a los datos subyacentes y Conceder permisos de ubicación de datos.

Error: "Permisos de clave de cifrado insuficientes para la API Glue"

Se intentó conceder permisos a Lake Formation sin permisos AWS Identity and Access Management (IAM) en la clave de AWS KMS cifrado de un catálogo de datos cifrado.

Mi consulta Amazon Athena o la de Amazon Redshift que usa manifiestos está fallando

Lake Formation no es compatible con las consultas que utilizan manifiestos.

Error: "Permiso(s) de Lake Formation insuficientes: se requiere crear etiqueta en el catálogo"

user/role Debe ser administrador de un lago de datos.

Error al eliminar administradores de lagos de datos no válidos

Debe eliminar todos los administradores de lagos de datos no válidos (funciones de IAM eliminadas que se definen como administradores de lagos de datos) simultáneamente. Si intenta eliminar administradores de lagos de datos no válidos por separado, Lake Formation da un error de entidad principal no válida.

Error: «La persona que llama no está autorizada temporalmente para realizar esta operación en el recurso...»

Se intentó consultar una tabla mediante un enlace a un recurso que puede haber sido redirigido. Para evitar este error, la aplicación debe adoptar el QuerySessionContextprotocolo. Consulte el Lake Formation flujo de trabajo para ver las operaciones de la API de integración de aplicaciones para obtener más información. De lo contrario, espere 24 horas antes de volver a intentar ejecutar la consulta.

Solución de problemas de acceso entre cuentas

Utilice esta información para diagnosticar y solucionar los problemas de acceso entre cuentas.

Temas

- He concedido un permiso entre cuentas Lake Formation pero el destinatario no puede ver el recurso
- Las entidades principales de la cuenta de destinatario pueden ver el recurso del Catálogo de datos pero no pueden acceder a los datos subyacentes
- Error: «La asociación falló porque la persona que llamó no estaba autorizada» al aceptar una AWS RAM invitación para compartir recursos
- Error: “No está autorizado a conceder permisos para el recurso”
- Error: «Acceso denegado para recuperar la información de AWS la organización»
- Error: "Organización <organization-ID> no encontrada"
- Error: "Permisos de Lake Formation insuficientes: combinación ilegal"
- ConcurrentModificationException en grant/revoke las solicitudes a cuentas externas
- Error al utilizar Amazon EMR para acceder a datos compartidos entre cuentas

He concedido un permiso entre cuentas Lake Formation pero el destinatario no puede ver el recurso

- ¿Es el usuario de la cuenta de destinatario administrador del lago de datos? Solo los administradores del lago de datos pueden ver el recurso en el momento del uso compartido.
- ¿Está compartiendo con una cuenta externa a su organización utilizando el método de recursos con nombre? Si es así, el administrador del lago de datos de la cuenta del destinatario debe aceptar una invitación para compartir recursos en AWS Resource Access Manager (AWS RAM).

  Para obtener más información, consulte the section called “Aceptar una invitación AWS RAM para compartir recursos”.
- ¿Está utilizando políticas de recursos a nivel de cuenta (Catálogo de datos) en AWS Glue? En caso afirmativo, si utiliza el método de recursos con nombre, deberá incluir una instrucción especial en la política que autorice a AWS RAM a compartir políticas en su nombre.

  Para obtener más información, consulte the section called “Administrar los permisos entre cuentas mediante ambos AWS Glue y Lake Formation”.
- ¿Tiene los permisos AWS Identity and Access Management (de IAM) necesarios para conceder el acceso entre cuentas?

  Para obtener más información, consulte the section called “Requisitos previos”.

- El recurso sobre el que ha concedido permisos no debe tener ningún permiso de Lake Formation concedido al grupo `IAMAllowedPrincipals`.
- ¿Existe alguna instrucción `deny` sobre el recurso en la política a nivel de cuenta?

Las entidades principales de la cuenta de destinatario pueden ver el recurso del Catálogo de datos pero no pueden acceder a los datos subyacentes

Los directores de la cuenta del destinatario deben tener los permisos AWS Identity and Access Management (IAM) necesarios. Para obtener más información, consulte Acceso a los datos subyacentes de una tabla compartida.

Error: «La asociación falló porque la persona que llamó no estaba autorizada» al aceptar una AWS RAM invitación para compartir recursos

Tras conceder acceso a un recurso a una cuenta diferente, cuando la cuenta receptora intenta aceptar la invitación a compartir el recurso, la acción falla.

```
$ aws ram get-resource-share-associations --association-type PRINCIPAL --resource-
share-arns arn:aws:ram:aws-region:444444444444:resource-share/e1d1f4ba-xxxx-xxxx-xxxx-
xxxxxxxx5d8d
{
    "resourceShareAssociations": [
        {
            "resourceShareArn": "arn:aws:ram:aws-region:444444444444:resource-share/
e1d1f4ba-xxxx-xxxx-xxxx-xxxxxxxx5d8d
",
            "resourceShareName": "LakeFormation-MMCC0XQBH3Y",
            "associatedEntity": "5815803XXXXX",
            "associationType": "PRINCIPAL",
            "status": "FAILED",
            "statusMessage": "Association failed because the caller was not
 authorized.",
            "creationTime": "2021-07-12T02:20:10.267000+00:00",
            "lastUpdatedTime": "2021-07-12T02:20:51.830000+00:00",
            "external": true
        }
    ]
}
```

El error se produce porque el `glue:PutResourcePolicy` es invocado por AWS Glue cuando la cuenta receptora acepta la invitación a compartir recursos. Para resolver el problema, permita que el rol asumido que utilice la producer/grantor cuenta realice la `glue:PutResourcePolicy` acción.

Error: “No está autorizado a conceder permisos para el recurso”

Se ha intentado conceder permisos entre cuentas en una base de datos o tabla que es propiedad de otra cuenta. Cuando una base de datos o una tabla se comparte con su cuenta, como administrador del lago de datos, puede conceder permisos sobre ella solo a los usuarios de su cuenta.

Error: «Acceso denegado para recuperar la información de AWS la organización»

Tu cuenta es una AWS cuenta de administración de Organizations y no tienes los permisos necesarios para recuperar la información de la organización, como las unidades organizativas de la cuenta.

Para obtener más información, consulte Required permissions for cross-account grants.

Error: "Organización <organization-ID> no encontrada"

Se ha intentado compartir un recurso con una organización, pero el uso compartido con organizaciones no está habilitado. Habilite el uso compartido de recursos con organizaciones.

Para obtener más información, consulte Habilitar el uso compartido con AWS Organizaciones en la Guía del AWS RAM usuario.

Error: "Permisos de Lake Formation insuficientes: combinación ilegal"

Un usuario compartió un recurso del Catálogo de datos mientras que los permisos de Lake Formation se concedieron al grupo `IAMAllowedPrincipals` para el recurso. El usuario debe revocar todos los permisos de Lake Formation de `IAMAllowedPrincipals` antes de compartir el recurso.

ConcurrentModificationException en grant/revoke las solicitudes a cuentas externas

Cuando los usuarios realizan varias solicitudes simultáneas de concesión y and/or revocación de permisos para un director en las políticas de etiqueta LF, Lake Formation las arroja.

ConcurrentModificationException Los usuarios deben atrapar la excepción y volver a intentar la grant/revoke solicitud fallida. El uso de versiones por lotes de las operaciones de la `RevokePermissions` API`GrantPermissions`/BatchRevokePermissionsalivia este problema en cierta medida al reducir el número de solicitudes grant/revoke simultáneas. BatchGrantPermissions

Error al utilizar Amazon EMR para acceder a datos compartidos entre cuentas

Al usar Amazon EMR para acceder a los datos compartidos desde otra cuenta, algunas bibliotecas de Spark intentarán llamar a la operación de la API `Glue:GetUserDefinedFunctions`. Como las versiones 1 y 2 de los permisos AWS RAM administrados no admiten esta acción, recibirá el siguiente mensaje de error:

```
"ERROR: User: arn:aws:sts::012345678901:assumed-role/my-
spark-role/i-06ab8c2b59299508a is not authorized to perform:
glue:GetUserDefinedFunctions on resource: arn:exampleCatalogResource
because no resource-based policy allows the glue:GetUserDefinedFunctions
action"
```

Para resolver este error, el administrador del lago de datos que creó el recurso compartido debe actualizar los permisos AWS RAM administrados asociados al recurso compartido. La versión 3 de los permisos administrados de AWS RAM permite a las entidades principales llevar a cabo la acción `glue:GetUserDefinedFunctions`.

Si crea un nuevo recurso compartido, Lake Formation aplica la última versión del permiso AWS RAM gestionado de forma predeterminada y no es necesario que realice ninguna acción. Para habilitar el acceso a los datos entre cuentas para los recursos compartidos existentes, debe actualizar los permisos AWS RAM administrados a la versión 3.

Puede ver los AWS RAM permisos asignados a los recursos compartidos con usted en AWS RAM. Los siguientes permisos se incluyen en la versión 3:

```
Databases
  AWSRAMPermissionGlueDatabaseReadWriteForCatalog
  AWSRAMPermissionGlueDatabaseReadWrite

Tables
  AWSRAMPermissionGlueTableReadWriteForCatalog
  AWSRAMPermissionGlueTableReadWriteForDatabase
```

```
AllTables
  AWSRAMPermissionGlueAllTablesReadWriteForCatalog
  AWSRAMPermissionGlueAllTablesReadWriteForDatabase
```

Para actualizar la versión de los permisos AWS RAM administrados de los recursos compartidos existentes

Usted (administrador del lago de datos) puede actualizar los permisos AWS RAM administrados a una versión más reciente siguiendo las instrucciones de la Guía del AWS RAM usuario o puede revocar todos los permisos existentes para el tipo de recurso y volver a concederlos. Si revoca los permisos, AWS RAM elimina el AWS RAM recurso compartido asociado al tipo de recurso. Al volver a conceder los permisos, AWS RAM crea nuevos recursos compartidos adjuntando la última versión de los permisos gestionados. AWS RAM

Solución de problemas de esquemas y flujos de trabajo

Utilice esta información para diagnosticar y solucionar problemas de esquemas y flujos de trabajo.

Temas

- <role-ARN>Mi blueprint falló con el mensaje «Usuario: <user-ARN>no está autorizado para realizar: iam: PassRole on resource:»
- <role-ARN>Mi flujo de trabajo falló con el mensaje «Usuario: <user-ARN>no está autorizado a realizar: iam: PassRole on resource:»
- Un rastreador en mi flujo de trabajo falló con "El recurso no existe o el solicitante no está autorizado para acceder a los permisos solicitados"
- Un rastreador de mi flujo de trabajo falló y dijo: «Se produjo un error (AccessDeniedException) al llamar a la CreateTable operación...»

<role-ARN>Mi blueprint falló con el mensaje «Usuario: <user-ARN>no está autorizado para realizar: iam: PassRole on resource:»

Se ha intentado crear un esquema por parte de un usuario que no tiene permisos suficientes para pasar el rol elegido.

Actualice la política de IAM del usuario para poder pasar el rol, o pídale que elija un rol diferente con los permisos necesarios para pasar el rol.

Para obtener más información, consulte the section called “Personas de Lake Formation y referencia de permisos IAM”.

<role-ARN>Mi flujo de trabajo falló con el mensaje «Usuario: <user-ARN>no está autorizado a realizar: iam: PassRole on resource:»

El rol que especificó para el flujo de trabajo no tenía una política en línea que permitiera que el rol se pasara a sí mismo.

Para obtener más información, consulte the section called “(Opcional) Crear un rol de IAM para flujos de trabajo”.

Un rastreador en mi flujo de trabajo falló con "El recurso no existe o el solicitante no está autorizado para acceder a los permisos solicitados"

Una posible causa es que el rol pasado no tenía permisos suficientes para crear una tabla en la base de datos de destino. Concede al rol el permiso `CREATE_TABLE` sobre la base de datos.

Un rastreador de mi flujo de trabajo falló y dijo: «Se produjo un error (AccessDeniedException) al llamar a la CreateTable operación...»

Una posible causa es que el rol de flujo de trabajo no tuviera permisos de ubicación de datos en el almacén de destino. Conceda permisos de ubicación de datos al rol.

Para obtener más información, consulte the section called “`DATA_LOCATION_ACCESS`”.

Problemas conocidos de AWS Lake Formation

Revise estos problemas conocidos para ver si AWS Lake Formation.

Temas

- Limitación del filtrado de metadatos de las tablas
- Problema al renombrar una columna excluida
- Problema con la eliminación de columnas en tablas CSV
- Las particiones de tabla deben añadirse bajo una ruta común
- Problema con la creación de una base de datos durante la creación del flujo de trabajo
- Problema al eliminar un usuario y, a continuación, volver a crearlo

- Las operaciones de la API del Catálogo de datos no actualizan el valor del parámetro IsRegisteredWithLakeFormation
- Las operaciones de Lake Formation no admiten el registro AWS Glue de esquemas

Limitación del filtrado de metadatos de las tablas

AWS Lake Formation los permisos a nivel de columna se pueden usar para restringir el acceso a columnas específicas de una tabla. Cuando un usuario recupera metadatos sobre la tabla utilizando la consola o una API como `glue:GetTable`, la lista de columnas del objeto de tabla contiene solo los campos a los que tiene acceso. Es importante comprender las limitaciones de este filtrado de metadatos.

Aunque Lake Formation pone a disposición de los servicios integrados metadatos sobre los permisos de las columnas, el filtrado real de las columnas en las respuestas a las consultas es responsabilidad del servicio integrado. Los clientes de Lake Formation compatibles con el filtrado a nivel de columna, incluidos Amazon Athena, Amazon Redshift Spectrum y Amazon EMR, filtran los datos en función de los permisos de columna registrados en Lake Formation. Los usuarios no podrán leer datos a los que no deberían tener acceso. Actualmente, AWS Glue ETL no es compatible con el filtrado de columnas.

> **Note**
>
> Los clústeres de EMR no se administran completamente mediante AWS. Por lo tanto, es responsabilidad de los administradores de los EMR asegurar adecuadamente los clústeres para evitar el acceso no autorizado a los datos.

Algunas aplicaciones o formatos pueden almacenar metadatos adicionales, incluidos los nombres y tipos de las columnas, en el mapa de `Parameters` como propiedades de la tabla. Estas propiedades se devuelven sin modificar y son accesibles por cualquier usuario con permiso de `SELECT` sobre cualquier columna.

Por ejemplo, el Avro SerDe almacena una representación en JSON del esquema de la tabla en una propiedad de la tabla denominada`avro.schema.literal`, que está disponible para todos los usuarios con acceso a la tabla. Le recomendamos que evite almacenar información sensible en las propiedades de las tablas y que sea consciente de que los usuarios pueden conocer el esquema completo de las tablas con formato Avro. Esta limitación es específica de los metadatos sobre una tabla.

AWS Lake Formation elimina cualquier propiedad de la tabla, empezando por `spark.sql.sources.schema` cuando responde a una solicitud `glue:GetTable` o similar, si la persona que llama no tiene `SELECT` permisos en todas las columnas de la tabla. Esto impide a los usuarios acceder a metadatos adicionales sobre tablas creadas con Apache Spark. Cuando se ejecutan en Amazon EMR, las aplicaciones Apache Spark siguen pudiendo leer estas tablas, pero es posible que no se apliquen ciertas optimizaciones y que no se admitan los nombres de columnas que distinguen entre mayúsculas y minúsculas. Si el usuario tiene acceso a todas las columnas de la tabla, Lake Formation devuelve la tabla sin modificar con todas las propiedades de la tabla.

Problema al renombrar una columna excluida

Si utiliza permisos a nivel de columna para excluir una columna y, a continuación, cambia el nombre de la columna, esta dejará de estar excluida de las consultas, como `SELECT *`.

Problema con la eliminación de columnas en tablas CSV

Si crea una tabla del Catálogo de datos con el formato CSV y luego elimina una columna del esquema, las consultas podrían devolver datos erróneos y es posible que no se respeten los permisos a nivel de columna.

Solución alternativa: cree una tabla nueva en su lugar.

Las particiones de tabla deben añadirse bajo una ruta común

Lake Formation espera que todas las particiones de una tabla estén bajo una ruta común que se establece en el campo de ubicación de la tabla. Cuando utilice el rastreador para añadir particiones a un catálogo, esto funcionará sin problemas. Pero si añade particiones manualmente, y estas particiones no están bajo la ubicación establecida en la tabla principal, el acceso a los datos no funciona.

Problema con la creación de una base de datos durante la creación del flujo de trabajo

Al crear un flujo de trabajo a partir de un esquema utilizando la consola de Lake Formation, puede crear la base de datos de destino si no existe. Al hacerlo, el usuario que ha iniciado sesión obtiene el permiso de `CREATE_TABLE` sobre la base de datos creada. Sin embargo, el rastreador que genera el flujo de trabajo asume el papel de este cuando intenta crear una tabla. Esto produce un error porque el rol no tiene el permiso `CREATE_TABLE` en la base de datos.

Solución: si crea la base de datos a través de la consola durante la configuración del flujo de trabajo, antes de ejecutar el flujo de trabajo, debe dar al rol asociado al flujo de trabajo el permiso `CREATE_TABLE` sobre la base de datos que acaba de crear.

Problema al eliminar un usuario y, a continuación, volver a crearlo

El siguiente escenario da como resultado permisos erróneos de Lake Formation devueltos por `lakeformation:ListPermissions`:

1. Cree un usuario y conceda permisos a Lake Formation.
2. Elimine el usuario.
3. Vuelva a crear el usuario con el mismo nombre.

`ListPermissions` devuelve dos entradas, una para el usuario anterior y otra para el usuario nuevo. Si intenta revocar los permisos concedidos al usuario anterior, se revocan los permisos del nuevo usuario.

Las operaciones de la API del Catálogo de datos no actualizan el valor del parámetro **`IsRegisteredWithLakeFormation`**

Existe una limitación conocida por la que las operaciones de la API del Catálogo de datos como `GetTables` y `SearchTables` no actualizan el valor del parámetro `IsRegisteredWithLakeFormation` y devuelven el valor predeterminado, que es falso. Se recomienda utilizar la API`GetTable` para ver el valor correcto del parámetro `IsRegisteredWithLakeFormation`.

Las operaciones de Lake Formation no admiten el registro AWS Glue de esquemas

Las operaciones de Lake Formation no admiten AWS Glue tablas que contengan un `SchemaReference` en el `StorageDescriptor` para utilizarlas en el Registro de esquemas.

Mensaje de error actualizado

AWS Lake Formation ha actualizado las excepciones específicas del recurso al mensaje de `EntityNotFound` error general para las siguientes operaciones de API a fin de cumplir con los objetivos de seguridad y cumplimiento.

- RevokePermissions
- GrantPermissions
- GetResourceLFTags
- GetTable
- GetDatabase

AWS Lake FormationAPI de

> **Note**
>
> Ya está disponible la referencia actualizada la API de servicio de AWS Lake Formation.

Contenido

- API de permisos
  - Operaciones
  - Data Types
- Configuración del lago de datos APIs
  - Operaciones
  - Data Types
- Integración del centro de identidad de IAM APIs
  - Operaciones
  - Data Types
- API de modo de acceso híbrido
  - Operaciones
  - Data Types
- API de expedición de credenciales
  - Operaciones
  - Data Types
- API de etiquetado
  - Operaciones
  - Data Types
- API de filtrado de datos
  - Operaciones
  - Tipos de datos
- Tipos de datos comunes
  - Estructura ErrorDetail

- Patrones de cadena

API de permisos

En la sección API de permisos se describen las operaciones y los tipos de datos necesarios para conceder y revocar permisos en AWS Lake Formation. Consulte la Guía de referencia de la API de Lake Formation para ver todas las operaciones y tipos de datos de la API de AWS Lake Formation.

Operaciones

- GrantPermissions
- RevokePermissions
- BatchGrantPermissions
- BatchRevokePermissions
- GetEffectivePermissionsForPath
- ListPermissions
- GetDataLakePrincipal

Data Types

- Resource
- DatabaseResource
- TableResource
- TableWithColumnsResource
- DataCellsFilterResourcee
- DataLocationResource
- DataLakePrincipal
- PrincipalPermissions
- PrincipalResourcePermissions
- DetailsMap
- ColumnWildcard
- BatchPermissionsRequestEntry

- BatchPermissionsFailureEntry

Configuración del lago de datos APIs

En esta sección figuran las operaciones de la API de configuración del lago de datos y los tipos de datos para gestionar los administradores del lago de datos.

Operaciones

- GetDataLakeSettings
- PutDataLakeSettings

Data Types

- DataLakeSettings

Integración del centro de identidad de IAM APIs

Esta contiene las operaciones para crear y administrar la integración de Lake Formation con IAM Identity Center.

Operaciones

- CreateLakeFormationIdentityCenterConfiguration
- DeleteLakeFormationIdentityCenterConfiguration
- DescribeLakeFormationIdentityCenterConfiguration
- UpdateLakeFormationIdentityCenterConfiguration

Data Types

- ExternalFilteringConfiguration

API de modo de acceso híbrido

En la sección API de modo de acceso híbrido se describen las operaciones y los tipos de datos necesarios para configurar el modo de acceso híbrido en AWS Lake Formation. Consulte la Guía de referencia de la API de Lake Formation para ver todas las operaciones y tipos de datos de la API de AWS Lake Formation.

Operaciones

- CreateLakeFormationOptIn
- DeleteLakeFormationOptIn
- ListLakeFormationOptIns

Data Types

- Resource
- DatabaseResource
- TableResource
- Resource Info
- LakeFormationOptInsInfo
- DataLocationResource

API de expedición de credenciales

La sección API de expedición de credenciales describe las operaciones y los tipos de datos relacionados con el uso del servicio de AWS Lake Formation para la expedición de credenciales y para registrar y administrar un recurso del lago de datos.

Operaciones

- RegisterResource
- DeregisterResource
- ListResources
- GetUnfilteredTableMetadata

- GetUnfilteredPartitionsMetadata
- GetTemporaryGluePartitionCredentials
- GetTemporaryGlueTableCredentials
- UpdateResource

Data Types

- FilterCondition
- RowFilter
- ResourceInfo

API de etiquetado

La sección API de etiquetado describe las operaciones y los tipos de datos relacionados con una estrategia de autorización que define un modelo de permisos sobre los atributos o las etiquetas de pares clave-valor.

Operaciones

- GetLFTagExpression
- ListLFTagExpressions
- DeleteLFTagExpression
- UpdateLFTagExpression
- CreateLFTagExpression
- AddLFTagsToResource
- RemoveLFTagsFromResource
- GetResourceLFTags
- ListLFTags
- CreateLFTag
- GetLFTag
- UpdateLFTag

- DeleteLFTag
- SearchTablesByLFTags
- SearchDatabasesByLFTags

Data Types

- LFTagKeyResource
- LFTagPolicyResource
- TaggedTable
- TaggedDatabase
- LFTag
- LFTagPair
- LFTagError
- ColumnLFTag

API de filtrado de datos

En la sección API de filtro de datos se describe cómo administrar los filtros de celdas de datos en AWS Lake Formation.

Operaciones

- CreateDataCellsFilter
- DeleteDataCellsFilter
- ListDataCellsFilter
- GetDataCellsFilter
- UpdateDataCellsFilter

Tipos de datos

- DataCellsFilter
- RowFilter

Tipos de datos comunes

Los tipos de datos comunes describen diversos tipos de datos comunes en AWS Lake Formation.

Estructura ErrorDetail

Contiene detalles sobre un error.

Campos

- `ErrorCode` – cadena UTF-8, con 1 byte de largo como mínimo y 255 bytes de largo como máximo, que coincide con el Single-line string pattern.

  El código asociado a este error.
- `ErrorMessage` – cadena de descripción de un máximo de 2048 bytes de largo, que coincide con el URI address multi-line string pattern.

  Mensaje que describe el error.

Patrones de cadena

La API utiliza las siguientes expresiones regulares para definir qué es un contenido válido para diversos miembros y parámetros de cadena:

- Patrón de cadena de línea única – `"[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\t]*"`
- Patrón de cadena de varias líneas de la dirección URI – `"[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*"`
- Patrón de cadena personalizado n.º 3: “`^\w+\.\w+\.\w+$`”
- Patrón de cadena personalizado n.º 4: “`^\w+\.\w+$`”
- Patrón de cadena personalizado n.º 5: “`arn:aws:iam::[0-9]*:role/.*`”
- Patrón de cadena personalizado n.º 6: “`arn:aws:iam::[0-9]*:user/.*`”
- Patrón de cadena personalizado n.º 7: “`arn:aws:iam::[0-9]*:group/.*`”
- Patrón de cadena personalizado n.º #8 – `"arn:aws:iam::[0-9]*:saml-provider/.*"`
- Patrón de cadena personalizado n.º #9 – `"^([\p{L}\p{Z}\p{N}_.:\/=+\-@%]*)$"`
- Patrón de cadena personalizado n.º #10 – `"^([\p{L}\p{Z}\p{N}_.:\*\/=+\-@%]*)$"`

- Patrón de cadena personalizado n.º #11 – "[\p{L}\p{N}\p{P}]*"

Regiones admitidas

Esta sección contiene información sobre el soporte Regiones de AWS y la funcionalidad de Lake Formation.

Disponibilidad general

Para conocer los servicios Regiones de AWS compatibles AWS Lake Formation, consulte la lista de AWS servicios disponibles por región.

Para obtener una lista de los puntos de conexión de Lake Formation de cada región y las Service Quotas de Lake Formation, consulte puntos de conexión y cuotas de AWS Lake Formation.

AWS GovCloud (US)

Para obtener una descripción general de las diferencias entre AWS GovCloud (US) la región y el estándar Regiones de AWS, consulte En qué AWS Lake Formation se diferencia AWS GovCloud (US).

Optimización de transacciones y almacenamiento

Las funciones de tablas gobernadas, soporte de transacciones y optimización del almacenamiento de Lake Formation están disponibles en las siguientes ubicaciones: Regiones de AWS

Nombre de la región	Parámetro de la región	Punto de conexión
Este de EE. UU. (Norte de Virginia)	us-east-1	lakeformation.us-east-1.amazonaws.com lakeformation-fips.us-east-1.amazonaws.com
Este de EE. UU. (Ohio)	us-east-2	lakeformation.us-east-2.amazonaws.com lakeformation-fips.us-east-2.amazonaws.com

Nombre de la región	Parámetro de la región	Punto de conexión
Oeste de EE. UU. (Oregón)	us-west-2	lakeformation.us-west-2.amazonaws.com lakeformation-fips.us-west-2.amazonaws.com
Asia-Pacífico (Mumbai)	ap-south-1	lakeformation.ap-south-1.amazonaws.com
Asia-Pacífico (Seúl)	ap-northeast-2	lakeformation.ap-northeast-2.amazonaws.com
Asia-Pacífico (Singapur)	ap-southeast-1	lakeformation.ap-southeast-1.amazonaws.com
Asia-Pacífico (Sídney)	ap-southeast-2	lakeformation.ap-southeast-2.amazonaws.com
Asia-Pacífico (Tokio)	ap-northeast-1	lakeformation.ap-northeast-1.amazonaws.com
Europa (Fráncfort)	eu-central-1	lakeformation.eu-central-1.amazonaws.com
Europa (Irlanda)	eu-west-1	lakeformation.eu-west-1.amazonaws.com
Europa (Londres)	eu-west-2	lakeformation.eu-west-2.amazonaws.com
Europa (Estocolmo)	eu-north-1	lakeformation.eu-north-1.amazonaws.com

Nombre de la región	Parámetro de la región	Punto de conexión
Canadá (centro)	ca-central-1	lakeformation.ca-central-1.amazonaws.com
América del Sur (São Paulo)	sa-east-1	lakeformation.sa-east-1.amazonaws.com

Historial de documentos para AWS Lake Formation

En la siguiente tabla se describen los cambios importantes en la documentación de AWS Lake Formation.

Cambio	Descripción	Fecha
Documentación actualizada para el control de acceso basado en atributos	Lake Formation permite conceder permisos a las entidades principales con atributos coincidentes en los recursos del Catálogo de datos. Para obtener más información, consulte Control de acceso basado en atributos.	17 de abril de 2025
Integración de Amazon S3 Tables con AWS Lake Formation y AWS Glue Data Catalog	Ahora puede integrar y catalogar las tablas S3 como AWS Glue Data Catalog objetos y registrar el catálogo como una ubicación de datos de Lake Formation. Para obtener más información, consulte Creación de un catálogo de Tablas de Amazon S3 en el AWS Glue Data Catalog.	13 de marzo de 2025
`AWSLakeFormationCrossAccountManager` Política actualizada de Lake Formation.	Lake Formation mejoró la AWSLakeFormationCrossAccountManagerpolítica al reemplazar el operador de `StringLike` condición por el `ArnLike` operador	25 de enero de 2025

	que permite a IAM realizar la verificación del formato ARN.	
Actualización sobre un cambio en la política	Documentó el cambio en la AWSLakeFormationDataAdminpolítica.	3 de diciembre de 2024
Actualizaciones de varios catálogos	AWS Glue Data Catalog le permite crear catálogos federados y unificar los datos de los lagos de datos de Amazon S3 y los almacenes de datos de Amazon Redshift, así como integrar los datos de sus bases de datos operativas, como Amazon DynamoDB, y de fuentes de datos de terceros, como Snowflake, MySQL, etc. Para obtener más información, consulte Inclusión de los datos en el AWS Glue Data Catalog.	3 de diciembre de 2024
Actualización de la documentación para las expresiones de etiquetas LF	Puede guardar las expresiones de etiquetas LF y reutilizarlas para conceder permisos en los recursos del Catálogo de datos. Para obtener más información, consulte Administración de expresiones de etiquetas LF para el control de acceso a los metadatos.	7 de noviembre de 2024

Actualizada la documentación relativa a las vistas del Catálogo de datos	Puede crear vistas AWS Glue Data Catalog utilizando AWS Glue APIs además de utilizar DDLs Amazon Athena y Amazon Redshift. Para obtener más información, consulte Creación de vistas del Catálogo de datos.	7 de agosto de 2024
Agregada documentación sobre la dispensación de credenciales auditables	Lake Formation le permite incluir el contexto del usuario del IAM Identity Center en los CloudTrail eventos y, a continuación, realizar un seguimiento de los usuarios que acceden a sus recursos. Para obtener más información, consulte Incluir el contexto de usuario del IAM Identity Center en CloudTrail los registros.	14 de julio de 2024
Actualización sobre un cambio en la política	Documentó el cambio (añadió una declaración IDs y eliminó los permisos redundantes) en las políticas AWSLakeFormationCrossAccountManager y AWSLake Formation DataAdmin.	14 de marzo de 2024
Actualización de la configuración de Lake Formation	Se han actualizado los pasos descritos en la sección Configuración de AWS Lake Formation.	7 de febrero de 2024

Actualización sobre un cambio en la política	Se han añadido nuevos permisos a la política insertada en un rol vinculado a un servicio. Para obtener más información, consulte Uso de roles vinculados a servicios para Lake Formation.	7 de febrero de 2024
Actualización sobre un cambio en la política	Documentó el cambio en la LakeFormationDataAccessServiceRolePolicypolítica.	2 de febrero de 2024
Limitaciones de Lake Formation consolidadas	Se ha creado una sección unificada de las limitaciones y consideraciones de Lake Formation. Para obtener más información, consulte Limitaciones de Lake Formation.	15 de diciembre de 2023
Se ha añadido documentación para la compactación de Iceberg	Para mejorar el rendimiento de lectura de los servicios de AWS análisis, como Athena y Amazon EMR, y los trabajos de AWS Glue ETL, AWS Glue Data Catalog proporciona compactación gestionada (un proceso que compacta objetos pequeños de Amazon S3 para convertirlos en objetos más grandes) para las tablas Iceberg del catálogo de datos. Para obtener más información, consulte Optimización de las tablas de Iceberg.	25 de noviembre de 2023

Se ha añadido documentación para la integración de IAM Identity Center	Las integraciones de IAM Identity Center permiten a los usuarios y grupos acceder a los recursos del catálogo de datos haciendo cumplir los permisos de Lake Formation. Para obtener más información consulte Integración de IAM Identity Center.	25 de noviembre de 2023
Se ha añadido documentación para las vistas del catálogo de datos	Puede crear vistas en las AWS Glue Data Catalog que se haga referencia a un máximo de 10 tablas mediante los editores de SQL Amazon Athena o Amazon Redshift. Para obtener más información, consulte Creación de vistas.	25 de noviembre de 2023
Actualizado el cambio de política	Documentó el cambio en la AWSLakeFormationCrossAccountManagerpolítica.	25 de octubre de 2023

Agregada documentación para el modo de acceso híbrido	El modo de acceso híbrido proporciona la flexibilidad de habilitar selectivamente los permisos de Lake Formation para las bases de datos y tablas de su AWS Glue Data Catalog. Con el modo de acceso híbrido, ahora tiene una ruta incremental que le permite establecer los permisos de Lake Formation para un conjunto específico de usuarios sin interrumpir las políticas de permisos de otros usuarios o cargas de trabajo existentes. Para obtener más información, consulte Modo de acceso híbrido.	26 de septiembre de 2023
Agregada documentación para crear tablas de Apache Iceberg	Ahora puede crear tablas Apache Iceberg que utilicen el formato de datos Apache Parquet AWS Glue Data Catalog con datos que residan en Amazon S3. Para obtener más información, consulte Creación de tablas Iceberg.	16 de agosto de 2023

Agregada documentación para el acceso a datos entre regiones	Lake Formation permite consultar las tablas del catálogo de datos en todas las AWS regiones. Puede acceder a los datos de una región desde otras regiones mediante Athena, Amazon EMR y ejecutar AWS Glue ETL creando enlaces de recursos en otras regiones que apunten a las bases de datos y tablas de origen. Puede conectar el Catálogo de datos a metaalmacenes externos que almacenan los metadatos de sus datos de Amazon S3 y gestionar de forma segura los permisos de acceso a los datos mediante AWS Lake Formation. Para obtener más información, consulte Acceso a tablas entre regiones.	30 de junio de 2023
Contenido reorganizado	Reorganizados los capítulos de la guía para adaptarlos a la experiencia del usuario de Lake Formation.	15 de mayo de 2023

Agregada documentación para la federación de HMS	Puede conectar el Catálogo de datos a metaalmacenes externos que almacenan los metadatos de sus datos de Amazon S3 y gestionar de forma segura los permisos de acceso a los datos mediante AWS Lake Formation. Para obtener más información, consulte Administración de los permisos de los conjuntos de datos que utilizan metaalmacenes externos.	15 de abril de 2023
Agregada documentación sobre el uso compartido de datos de Amazon Redshift	Ahora puede gestionar de forma segura los datos de un recurso compartido de datos de Amazon Redshift con los permisos de Lake Formation . Lake Formation admite la concesión de licencias de acceso a sus datos mediante AWS Data Exchange. Para obtener más información, consulte Intercambio de datos en AWS Lake Formation.	30 de noviembre de 2022
Compatibilidad para compartir datos entre cuentas directamente con las entidades principales	Agregada información sobre cómo compartir datos directamente con las entidades principales de IAM en otra cuenta. Para obtener más información, consulte Uso compartido de datos entre cuentas en AWS Lake Formation.	10 de noviembre de 2022

Support para AWS RAM habilitar el intercambio de datos mediante TBAC	Se agregó información sobre el método LF-TBAC para conceder permisos al catálogo de datos, que se utiliza para las concesiones entre cuentas. AWS Resource Access Manager	10 de noviembre de 2022
Agregada una sección sobre cómo trabajar con otros servicios	Se agregó información sobre cómo AWS servicios como Athena, AWS Glue Redshift Spectrum y Amazon EMR pueden utilizar Lake Formation para acceder de forma segura a los datos de las ubicacion es de Amazon S3 registrad as en Lake Formation. Para obtener más información, consulte Trabajar con otros AWS servicios.	10 de noviembre de 2022
???	Agregada información sobre cómo solucionar un error al utilizar Amazon EMR para acceder a los datos entre cuentas. Para obtener más información, consulte Error al utilizar Amazon EMR para acceder a datos compartidos entre cuentas.	7 de noviembre de 2022

Actualizaciones del uso compartido de recursos entre cuentas	Agregada una descripción sobre cómo funcionan los recursos compartidos entre cuentas en Lake Formation . Documentó el cambio en la política. AWSLakeFormationCrossAccountManager	6 de mayo de 2022
Nuevos tutoriales	Agregados nuevos tutoriales para crear tablas gobernadas, proteger los lagos de datos y compartir lagos de datos. Para obtener más detalles, consulte la sección Comenzar.	20 de abril de 2022
Nueva página de inicio de Lake Formation	Se actualizó la página de inicio de Lake Formation para incluir enlaces a tutoriales que proporcionan step-by-step instrucciones sobre cómo construir un lago de datos, ingerir datos, compartir y proteger lagos de datos con Lake Formation.	20 de abril de 2022

Compatibilidad para el suministro de credenciales	Agregada información sobre el suministro de credencia les, compatible con Lake Formation para permitir que los servicios de terceros se integren con Lake Formation mediante las operaciones de la API de suministro de credenciales. Para obtener más información, consulte Cómo funciona el suministr o de credenciales en Lake Formation.	28 de febrero de 2022
Compatibilidad para tablas gobernadas y filtrado de datos avanzado	Agregada información acerca de tablas gobernada s por Lake Formation, que admiten transacciones ACID, compactación automática de datos y consultas de viaje en el tiempo. Agregada informaci ón sobre la creación de filtros de datos para admitir la seguridad a nivel de columna, fila y celda. Para obtener más información, consulte Tablas gobernadas en Lake Formation y Filtrado de datos y seguridad a nivel de celda en Lake Formation.	30 de noviembre de 2021

Compatibilidad con puntos de conexión de interfaz de VPC	Se agregó información sobre la creación de un punto final de interfaz de nube privada virtual (VPC) para Lake Formation, de modo que la comunicación entre su VPC y Lake Formation se lleve a cabo de forma completa y segura dentro de la red. AWS Para obtener más informaci ón, consulte Uso de Lake Formation con puntos de conexión de VPC.	11 de octubre de 2021
Compatibilidad con las políticas de punto de conexión de VPC	Agregada información acerca de la compatibilidad con las políticas de punto de conexión de nube privada virtual (VPC) en Lake Formation. Para obtener más informaci ón, consulte Uso de Lake Formation con puntos de conexión de VPC.	11 de octubre de 2021
Compatibilidad con control de acceso basado en etiquetas.	El control de acceso basado en etiquetas de Lake Formation proporciona una forma nueva y más escalable de administrar el acceso a los recursos del Catálogo de datos y a los datos subyacent es mediante etiquetas LF. Para obtener más informaci ón, consulte Control de acceso basado en etiquetas de Lake Formation.	7 de mayo de 2021

Nuevo requisito de suscripción para el filtrado de datos en Amazon EMR.	Agregada información sobre el requisito de optar por permitir que Amazon EMR filtre los datos gestionados por Lake Formation. Para obtener más información, consulte Permitir el filtrado de datos en Amazon EMR.	9 de octubre de 2020
Compatibilidad para conceder permisos entre cuentas completos en las bases de datos del Catálogo de datos	Agregada información sobre la concesión de permisos completos de Lake Formation en bases de datos del Catálogo de datos entre cuentas de AWS , incluido `CREATE_TABLE` . Para obtener más información, consulte Compartir bases de datos del Catálogo de datos.	1 de octubre de 2020
Support para que Amazon Athena los usuarios se autentiquen mediante SAML.	Agregada información sobre la compatibilidad para los usuarios de Athena que se conectan a través del controlador JDBC u ODBC y se autentican a través de proveedores de identidad SAML, como Okta y Microsoft Active Directory Federation Service (AD FS). Para más información, consulte Integraciones de servicio AWS con Lake Formation.	30 de septiembre de 2020

Compatibilidad para acceso entre cuentas con un Catálogo de datos cifrado	Agregada información sobre la concesión de permisos entre cuentas cuando el Catálogo de datos está cifrado. Para obtener más información, consulte Requisitos previos para el acceso entre cuentas.	30 de julio de 2020
Compatibilidad para el acceso entre cuentas al lago de datos	Se agregó información sobre la concesión de AWS Lake Formation permisos en las bases de datos y tablas del catálogo de datos a AWS cuentas y organizaciones externas, y sobre el acceso a los objetos del catálogo de datos compartidos desde cuentas externas. Para obtener más información, consulte Acceso entre cuentas.	7 de julio de 2020
Integración con Amazon Quick	Se agregó información sobre cómo conceder permisos de Lake Formation a los usuarios de Quick Enterprise Edition para que puedan acceder a los conjuntos de datos que residen en las ubicaciones registradas de Amazon S3. Para obtener más información, consulte Concesión de permisos de ubicación de datos.	29 de junio de 2020

Actualizaciones de los capítulos de configuración y de introducción	Reorganizados y mejorados los capítulos de configuración e introducción. Se actualizaron los permisos recomendados AWS Identity and Access Management (IAM) para el administrador del lago de datos.	27 de febrero de 2020
Support para AWS Key Management Service	Se agregó información sobre cómo el soporte de Lake Formation para AWS Key Management Service (AWS KMS) simplifica la configuración de servicios integrados para leer y escribir datos cifrados en ubicaciones registradas de Amazon Simple Storage Service (Amazon S3). Se agregó información sobre cómo registrar las ubicaciones de Amazon S3 cifradas con AWS KMS keys. Para obtener más información, consulte the section called “Añadir una ubicación de Amazon S3 a su lago de datos”.	27 de febrero de 2020
Actualizaciones de los esquemas y las políticas de IAM de los administradores de lagos de datos	Aclarados los parámetros de entrada para los esquemas de bases de datos incrementales. Actualizadas las políticas de IAM necesarias para el administrador de un lago de datos.	20 de diciembre de 2019

Reescritura y actualización de los capítulos de seguridad	Mejorados los capítulos de seguridad y actualización.	29 de octubre de 2019
El superpermiso reemplaza a todos los permisos	Actualizados los capítulos de seguridad y actualización para reflejar la sustitución del permiso `All` por `Super`.	10 de octubre de 2019
Adiciones, correcciones y aclaraciones	Efectuadas adiciones, correcciones y aclaraciones de acuerdo con los comentarios. Revisado el capítulo de seguridad. Actualizados los capítulos de seguridad y actualización para reflejar la sustitución del grupo `Everyone` por `IAMAllowedPrincipals` .	11 de septiembre de 2019
Nueva guía	Esta es la versión inicial de la Guía para desarrolladores de AWS Lake Formation .	8 de agosto de 2019

AWSGlosario de

Para ver la terminología más reciente de AWS, consulte el Glosario de AWS en la Referencia de Glosario de AWS.

Las traducciones son generadas a través de traducción automática. En caso de conflicto entre la traducción y la version original de inglés, prevalecerá la version en inglés.