



Guida alle operazioni di base

Amazon Redshift



Amazon Redshift: Guida alle operazioni di base

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

I marchi e l'immagine commerciale di Amazon non possono essere utilizzati in relazione a prodotti o servizi che non siano di Amazon, in una qualsiasi modalità che possa causare confusione tra i clienti o in una qualsiasi modalità che denigri o discrediti Amazon. Tutti gli altri marchi non di proprietà di Amazon sono di proprietà delle rispettive aziende, che possono o meno essere associate, collegate o sponsorizzate da Amazon.

Table of Contents

Inizia a usare i data warehouse serverless	1
Iscrizione a AWS	1
Creazione di un data warehouse con Amazon Redshift serverless	1
Caricamento di dati di esempio	4
Esecuzione di query di esempio	6
Caricamento di dati da Amazon S3	8
Inizia a usare i data warehouse con provisioning	16
Iscrizione a AWS	18
Definizione delle regole del firewall	19
Fase 1: creazione di un cluster di esempio	20
Fase 2: Configurazione delle regole in ingresso per i client SQL	23
Passaggio 3: concedere l'accesso a un client SQL ed eseguire le query	24
Concessione dell'accesso all'editor di query v2	24
Passaggio 4: caricare i dati da Amazon S3 ad Amazon Redshift	25
Caricamento di dati da Amazon S3 tramite comandi SQL	25
Caricamento di dati da Amazon S3 utilizzando l'editor di query v2	27
Crea dati TICKIT nel tuo cluster	28
Fase 5: provare query di esempio utilizzando l'editor di query	29
Fase 6: reimpostazione dell'ambiente	30
Definisci e utilizza un database nel tuo data warehouse	32
Connessione ad Amazon Redshift	33
Creazione di un database	34
Creazione di un utente	35
Creare uno schema	35
Creare una tabella	37
Inserisci le righe di dati in una tabella	38
Seleziona dati da una tabella	38
Caricare dati	39
Interroga le tabelle e le viste di sistema	39
Visualizza un elenco di nomi di tabella	39
Visualizzazione degli utenti	41
Visualizza le query recenti	41
Determina l'ID di sessione di una query in esecuzione	42
Annullare un'interrogazione	43

Annulla una query tramite la coda dell'utente con privilegi avanzati	45
Esegui query su dati non presenti nel tuo database Amazon Redshift	47
Esecuzione di query nei data lake	47
Esecuzione di query su origini dati remote	48
Accesso ai dati in altri database	48
Addestramento dei modelli di ML con i dati Redshift	49
Scopri i concetti di Amazon Redshift	50
Risorse di apprendimento aggiuntive	53
Cronologia dei documenti	55
.....	lvii

Inizia a usare i data warehouse Serverless Amazon Redshift

Se utilizzi Amazon Redshift Serverless per la prima volta, consigliamo di iniziare leggendo le seguenti sezioni: Il flusso di base di Amazon Redshift serverless consiste nel creare risorse serverless, connettersi ad Amazon Redshift serverless, caricare dati di esempio e quindi eseguire query sui dati. In questa guida, è possibile scegliere di caricare dati di esempio da Amazon Redshift serverless o da un bucket Amazon S3. I dati di esempio vengono utilizzati in tutta la documentazione di Amazon Redshift per dimostrare le funzionalità. Per iniziare a utilizzare i data warehouse con provisioning di Amazon Redshift, consulta [Inizia a usare i data warehouse con provisioning di Amazon Redshift](#)

- [the section called “Iscrizione a AWS”](#)
- [the section called “Creazione di un data warehouse con Amazon Redshift serverless”](#)
- [the section called “Caricamento di dati da Amazon S3”](#)

Iscrizione a AWS

Se non hai già un AWS account, registrane uno. Se disponi già di un account, puoi ignorare questo prerequisito e utilizzare il tuo account esistente.

1. Apri la <https://portal.aws.amazon.com/billing/registrazione>.
2. Segui le istruzioni online.

Quando si registra un AWS account, viene creato un utente root dell' AWS account. L'utente root ha accesso a tutti i AWS servizi e le risorse dell'account. Come best practice di sicurezza, [assegna l'accesso amministrativo a un utente amministrativo](#) e utilizza solo l'utente root per eseguire [attività che richiedono l'accesso di un utente root](#).

Creazione di un data warehouse con Amazon Redshift serverless

La prima volta che accedi alla console Amazon Redshift serverless, viene richiesto di accedere all'esperienza introduttiva, che puoi utilizzare per creare e gestire risorse serverless. In questa guida, creerai risorse serverless utilizzando le impostazioni predefinite di Amazon Redshift serverless.

Per un controllo più granulare della configurazione, scegli Customize settings (Personalizza impostazioni).

Note

Redshift Serverless richiede un Amazon VPC con tre sottoreti in tre diverse zone di disponibilità. Redshift Serverless richiede inoltre almeno 3 indirizzi IP disponibili. Assicurati che l'Amazon VPC che usi per Redshift Serverless abbia tre sottoreti in tre diverse zone di disponibilità e almeno 3 indirizzi IP disponibili, prima di continuare. Per ulteriori informazioni sulla creazione di sottoreti in [un Amazon VPC](#), consulta [Create a subnet](#) nella Amazon Virtual Private Cloud User Guide. Per ulteriori informazioni sugli indirizzi IP in un Amazon VPC, consulta [Indirizzamento IP per le tue sottoreti VPCs](#).

Per configurare utilizzando le impostazioni predefinite:

1. Accedi AWS Management Console e apri la console Amazon Redshift all'indirizzo. <https://console.aws.amazon.com/redshiftv2/>

Scegli Prova la versione di prova gratuita di Redshift Serverless.

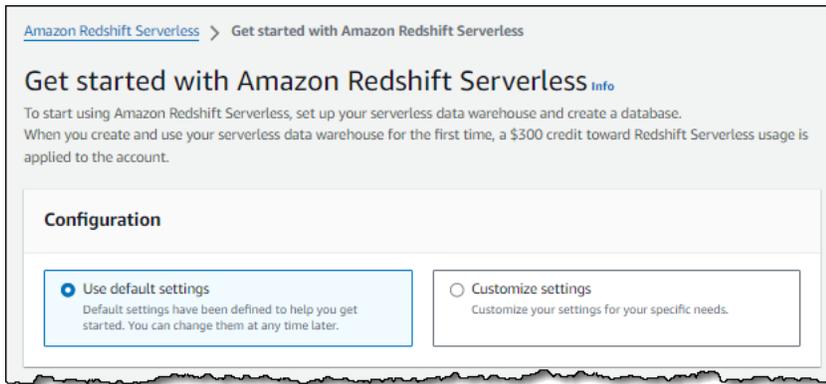
2. In Configuration (Configurazione), scegli Use default settings (Utilizza impostazioni predefinite). Amazon Redshift Serverless crea uno spazio dei nomi predefinito con un gruppo di lavoro predefinito associato a questo spazio dei nomi. Seleziona Save configuration (Salva configurazione).

Note

Un namespace è una raccolta di oggetti e utenti del database. I namespace raggruppano tutte le risorse utilizzate in Redshift Serverless, come schemi, tabelle, utenti, condivisioni di dati e istantanee.

Un gruppo di lavoro è una raccolta di risorse di elaborazione. I gruppi di lavoro ospitano le risorse di elaborazione che Redshift Serverless utilizza per eseguire attività di calcolo.

La schermata seguente mostra le impostazioni predefinite per Amazon Redshift Serverless.



3. Una volta completata la configurazione, scegliere Continue (Continua) per andare a Serverless dashboard (Pannello di controllo serverless). Puoi vedere che il gruppo di lavoro e lo spazio dei nomi serverless sono disponibili.

Serverless dashboard [Info](#)

Namespace overview [Info](#)

Namespace data from your account

Total snapshots	Datashares in my account	Datashares requiring authorization	Datashares fr
0	0	0	0

Namespaces / Workgroups [Info](#)

Namespace	Status	Workgroup	Status
default	Available	default	Available

i Note

Se Redshift Serverless non crea correttamente il gruppo di lavoro, puoi fare quanto segue:

- Risolvi eventuali errori segnalati da Redshift Serverless, come la presenza di un numero insufficiente di sottoreti nel tuo Amazon VPC.

- Elimina lo spazio dei nomi scegliendo lo spazio dei nomi predefinito nella dashboard di Redshift Serverless, quindi scegliendo Azioni, Elimina spazio dei nomi. L'eliminazione di un namespace richiede diversi minuti.
- Quando si apre nuovamente la console Redshift Serverless, viene visualizzata la schermata di benvenuto.

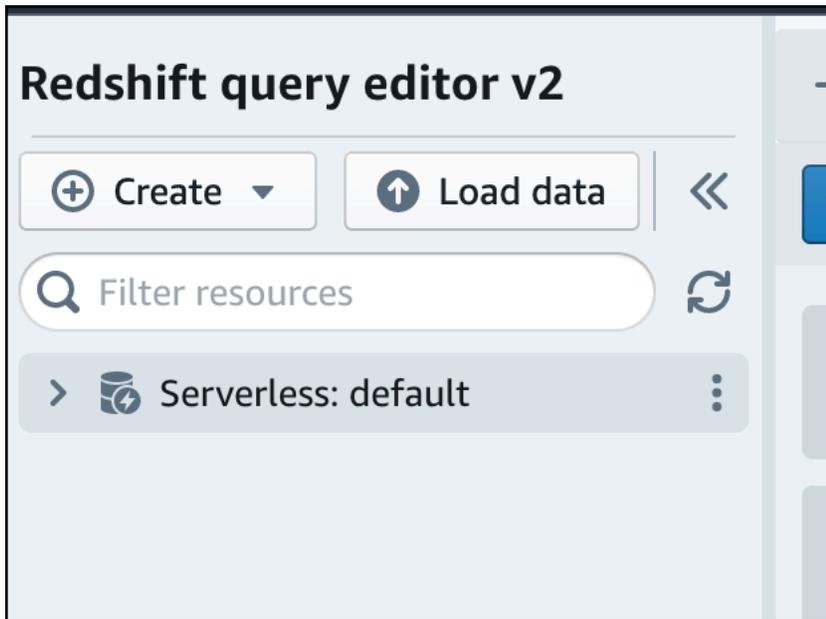
Caricamento di dati di esempio

Ora che è stato configurato il data warehouse con Amazon Redshift Serverless, puoi utilizzare l'editor di query Amazon Redshift v2 per caricare dati di esempio.

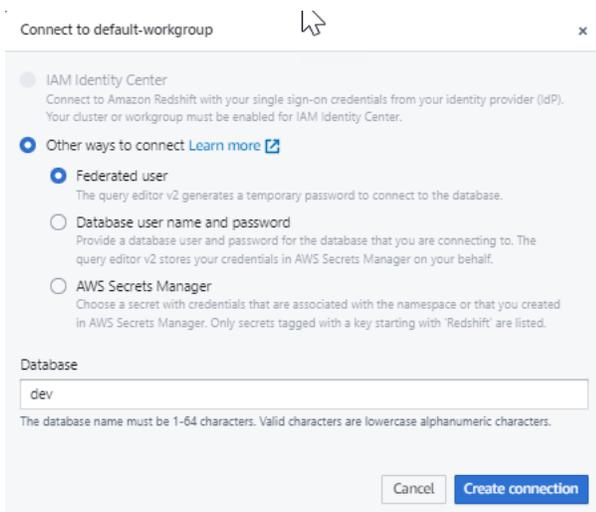
1. Per avviare l'editor di query v2 dalla console Amazon Redshift Serverless, scegli Esegui query sui data. Quando richiami l'editor di query v2 dalla console di Amazon Redshift Serverless, viene visualizzata una nuova scheda del browser con l'editor di query. L'editor di query v2 si connette dal computer client all'ambiente Amazon Redshift Serverless.



2. Per questa guida, utilizzerai il tuo account AWS amministratore e quello predefinito. AWS KMS key Per informazioni sulla configurazione dell'editor di query v2, incluse le autorizzazioni necessarie, consulta [Configuring your](#) nella Account AWS Amazon Redshift Management Guide. Per informazioni sulla configurazione di Amazon Redshift per l'utilizzo di una chiave gestita dal cliente o per modificare la chiave KMS utilizzata da Amazon Redshift, [consulta AWS KMS Changing](#) the key for a namespace.
3. Per connettersi a un gruppo di lavoro, scegliere il nome del gruppo di lavoro nel pannello con struttura ad albero.

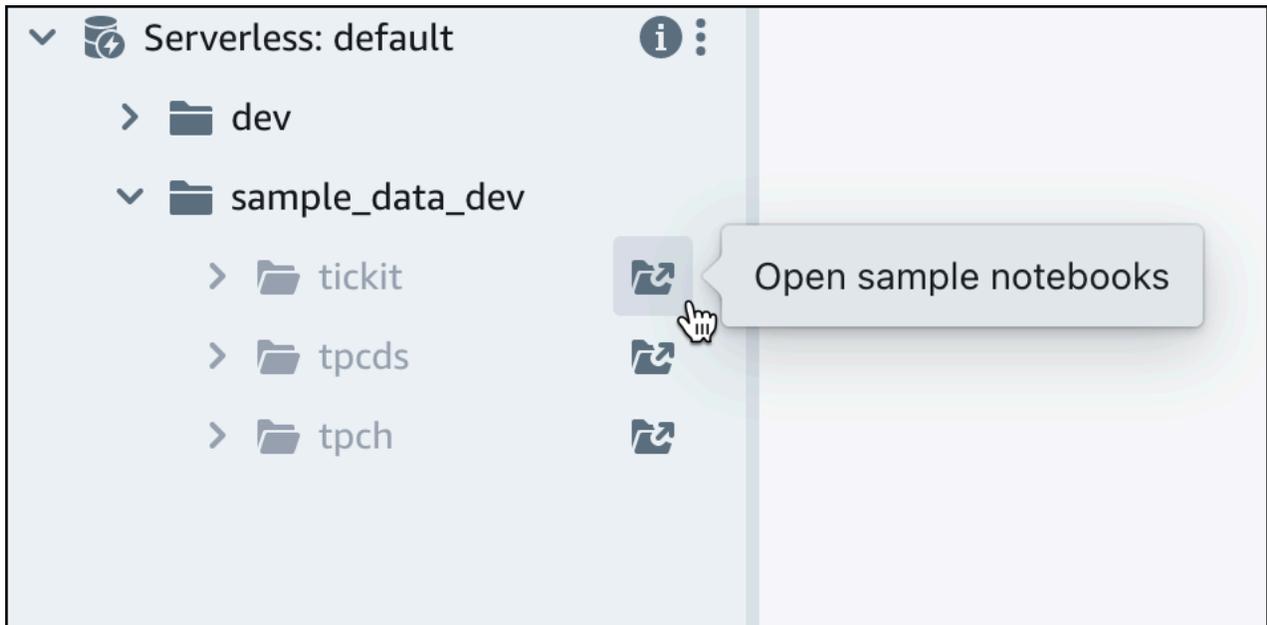


- Quando ti connetti a un nuovo gruppo di lavoro per la prima volta nell'editor di query v2, devi selezionare il tipo di autenticazione da utilizzare per connetterti al gruppo di lavoro. Per questa guida, lascia selezionato Utente federato e scegli Crea connessione.



Una volta connesso, puoi scegliere di caricare dati di esempio da Amazon Redshift serverless o da un bucket Amazon S3.

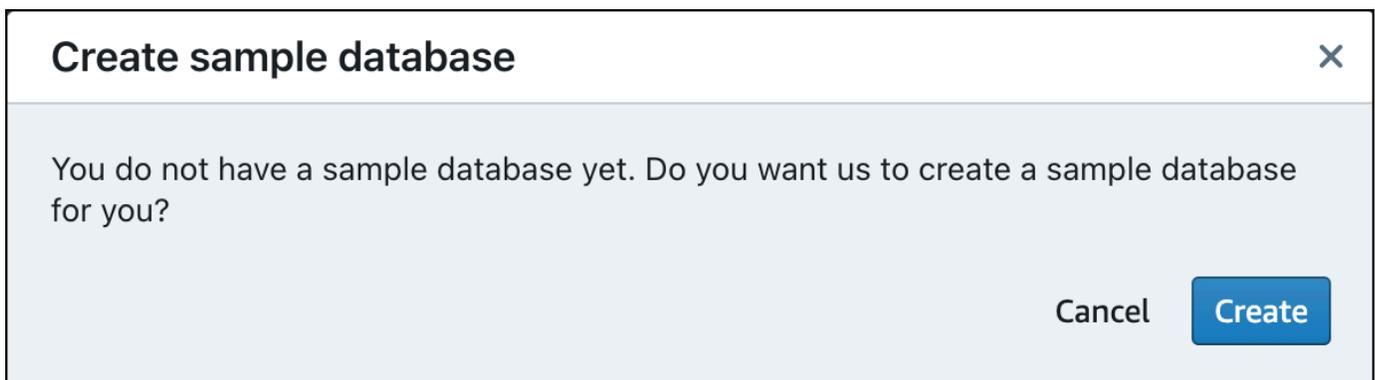
- Nel gruppo di lavoro Amazon Redshift Serverless predefinito, espandi il database `sample_data_dev`. Esistono tre schemi di esempio corrispondenti a tre set di dati di esempio che puoi caricare nel database Amazon Redshift serverless. Scegli il set di dati di esempio che desideri caricare e scegli Apri notebook di esempio.



Note

Un notebook SQL è un contenitore per celle SQL e Markdown. È possibile utilizzare i taccuini per organizzare, annotare e condividere più comandi SQL in un unico documento.

- Quando si caricano i dati per la prima volta, l'editor di query v2 richiederà di creare un database di esempio. Scegli Create (Crea).



Esecuzione di query di esempio

Dopo aver configurato Amazon Redshift serverless, puoi iniziare a utilizzare un set di dati di esempio in Amazon Redshift serverless. Amazon Redshift serverless carica automaticamente il set di dati di esempio, ad esempio il set di dati tickit, ed è possibile eseguire immediatamente le query sui dati.

- Una volta che Amazon Redshift serverless ha terminato il caricamento dei dati di esempio, tutte le query di esempio vengono caricate nell'editor. Puoi scegliere Esegui tutto per eseguire tutte le query dai notebook di esempio.



The screenshot shows the Amazon Redshift console interface. At the top, the title is "Sales per event". Below it, there's a "Run" button and a "Limit 100" toggle. The SQL query is displayed in a code editor:

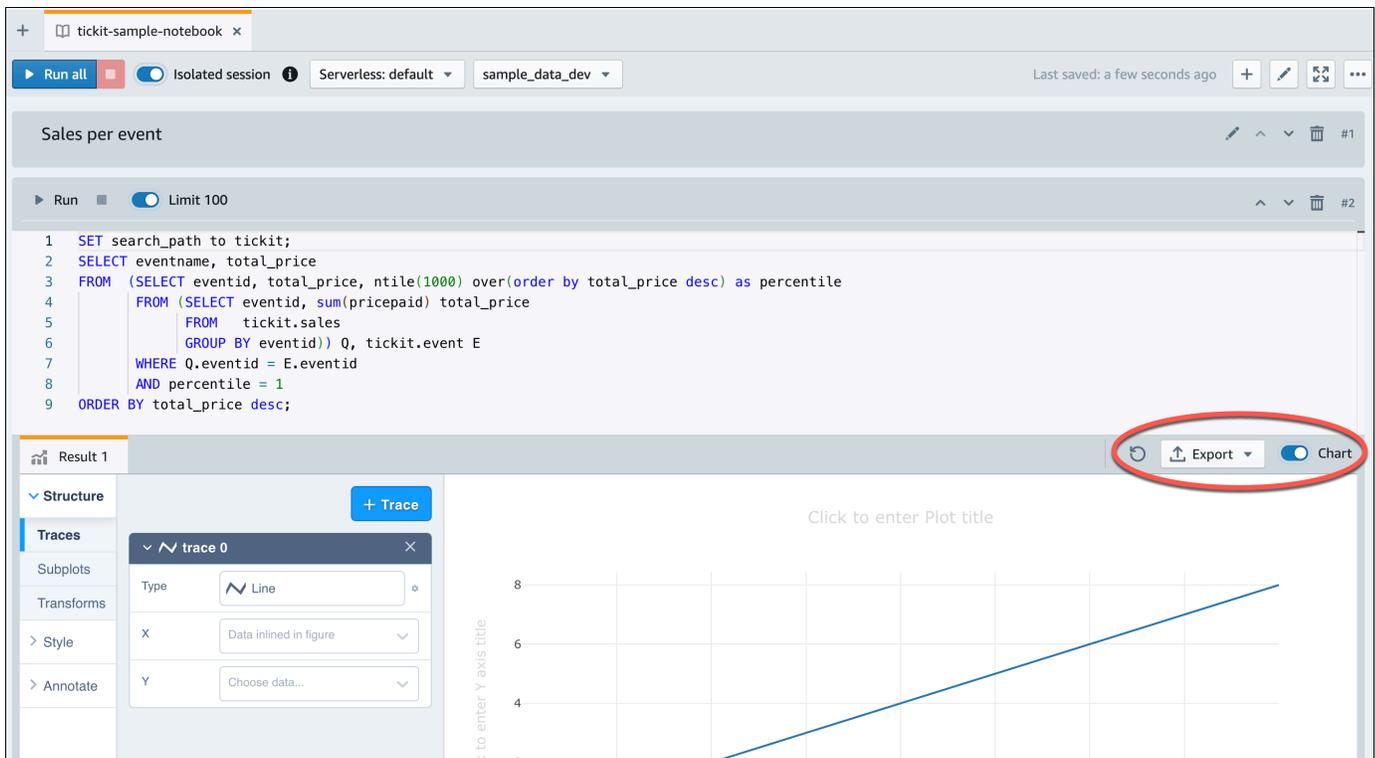
```
1 SET search_path to tickit;
2 SELECT eventname, total_price
3 FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as percentile
4       FROM (SELECT eventid, sum(pricepaid) total_price
5             FROM tickit.sales
6             GROUP BY eventid)) Q, tickit.event E
7 WHERE Q.eventid = E.eventid
8      AND percentile = 1
9 ORDER BY total_price desc;
```

Below the query, there are two tabs: "Result 1" and "Result 2 (9)". The "Result 2 (9)" tab is active, showing a table with two columns: "eventname" and "total_price". The table contains the following data:

eventname	total_price
Adriana Lecouvreur	51846
Janet Jackson	51049
Phantom of the Opera	50301
The Little Mermaid	49956
Citizen Cope	49823
Sevendust	48020
Electra	47883
Mary Poppins	46780
Live	46661

At the bottom right, it says "Elapsed time: 401 ms Total rows: 9".

Puoi anche esportare i risultati come file JSON o CSV o visualizzarli in un grafico.



The screenshot shows the Amazon Redshift console interface. At the top, the title is "Sales per event". Below it, there's a "Run all" button and a "Limit 100" toggle. The SQL query is displayed in a code editor:

```
1 SET search_path to tickit;
2 SELECT eventname, total_price
3 FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as percentile
4       FROM (SELECT eventid, sum(pricepaid) total_price
5             FROM tickit.sales
6             GROUP BY eventid)) Q, tickit.event E
7 WHERE Q.eventid = E.eventid
8      AND percentile = 1
9 ORDER BY total_price desc;
```

Below the query, there are two tabs: "Result 1" and "Result 2 (9)". The "Result 2 (9)" tab is active, showing a table with two columns: "eventname" and "total_price". The table contains the following data:

eventname	total_price
Adriana Lecouvreur	51846
Janet Jackson	51049
Phantom of the Opera	50301
The Little Mermaid	49956
Citizen Cope	49823
Sevendust	48020
Electra	47883
Mary Poppins	46780
Live	46661

At the bottom right, there are two buttons: "Export" and "Chart". The "Export" button is circled in red. The "Chart" button is also circled in red. The "Chart" button is currently disabled.

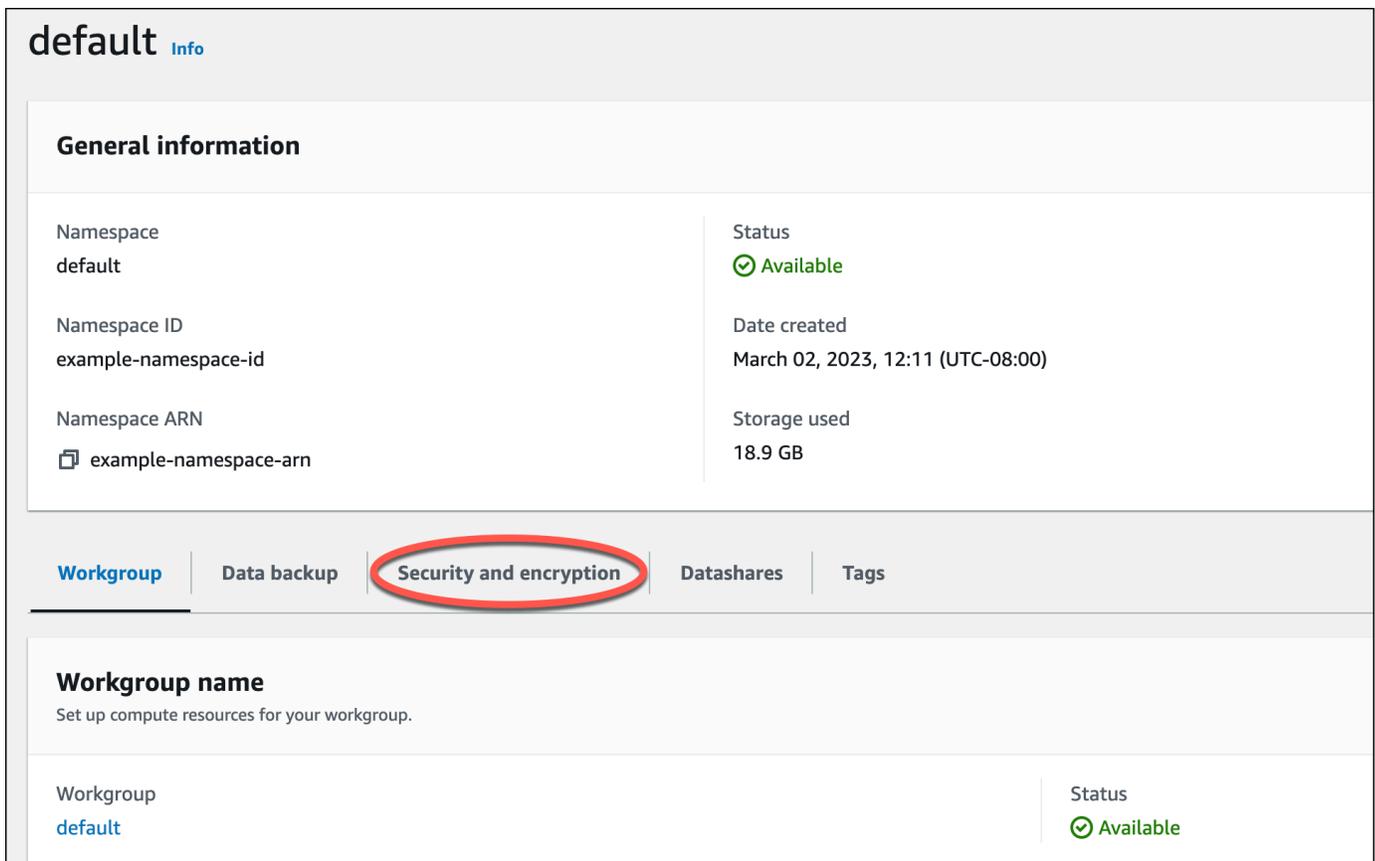
Puoi anche caricare i dati da un bucket Amazon S3. Per ulteriori informazioni, consulta [the section called “Caricamento di dati da Amazon S3”](#).

Caricamento di dati da Amazon S3

Una volta creato il data warehouse, puoi caricare i dati da Amazon S3.

A questo punto, disponi di un database denominato dev. Successivamente, crea alcune tabelle nel database, carica i dati nelle tabelle e prova a eseguire una query. Per maggiore praticità, i dati di esempio da caricare sono disponibili in un bucket Amazon S3.

1. Prima di poter caricare i dati da Amazon S3, è necessario prima creare un ruolo IAM con le autorizzazioni necessarie e collegarlo al proprio spazio dei nomi serverless. A tale scopo, torna alla console Redshift Serverless e scegli la configurazione del namespace. Dal menu di navigazione, scegli il tuo namespace, quindi scegli Sicurezza e crittografia. Scegli Gestisci ruoli IAM.



The screenshot displays the configuration page for a Redshift namespace named 'default'. The 'General information' section shows the namespace is 'Available', created on March 02, 2023, and has used 18.9 GB of storage. Below this, a navigation bar includes tabs for 'Workgroup', 'Data backup', 'Security and encryption', 'Datashares', and 'Tags'. The 'Security and encryption' tab is circled in red. The 'Workgroup name' section below shows the workgroup is 'default' and also 'Available'.

default <small>Info</small>	
General information	
Namespace default	Status ✔ Available
Namespace ID example-namespace-id	Date created March 02, 2023, 12:11 (UTC-08:00)
Namespace ARN example-namespace-arn	Storage used 18.9 GB
Workgroup name Set up compute resources for your workgroup.	
Workgroup default	Status ✔ Available

2. Espandi il menu Gestisci ruoli IAM e scegli Crea ruolo IAM.

Manage IAM roles

Permissions

i Associate an IAM role so that your serverless endpoint can LOAD and UNLOAD data. You can create an IAM role as the default for this configuration that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift Serverless. This policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue. You won't be able to run these SQL commands without an IAM role attached to your namespace.

Associated IAM roles (1)

Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose an IAM role and set it as the default.

Set default ▼ **Manage IAM roles** ▲

🔍 Search for associated IAM roles or role type

Associate IAM roles

Create IAM role

Remove IAM roles

< 1 >

IAM roles [🔗](#) ▼ **Status** ▼ **Role type** ▼

3. Scegli il livello di accesso al bucket S3 che desideri concedere a questo ruolo e scegli Crea ruolo IAM come default.

Create the default IAM role ✕

i Associate an IAM role so that your serverless endpoint can LOAD and UNLOAD data. You can create an IAM role as the default for this configuration that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift Serverless. This policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue. You won't be able to run these SQL commands without an IAM role attached to your namespace.

Specify an S3 bucket for the IAM role to access
To create a new bucket, [visit S3](#)

No additional S3 bucket
Create the IAM role without specifying S3 buckets.

Any S3 bucket
Allow users that have access to your Redshift Serverless data to also access any S3 bucket and its contents in your AWS account.

Specific S3 buckets
Specify one or more S3 buckets that the IAM role being created has permission to access.

4. Scegli **Save changes** (Salva modifiche). È ora possibile caricare i dati di esempio da Amazon S3.

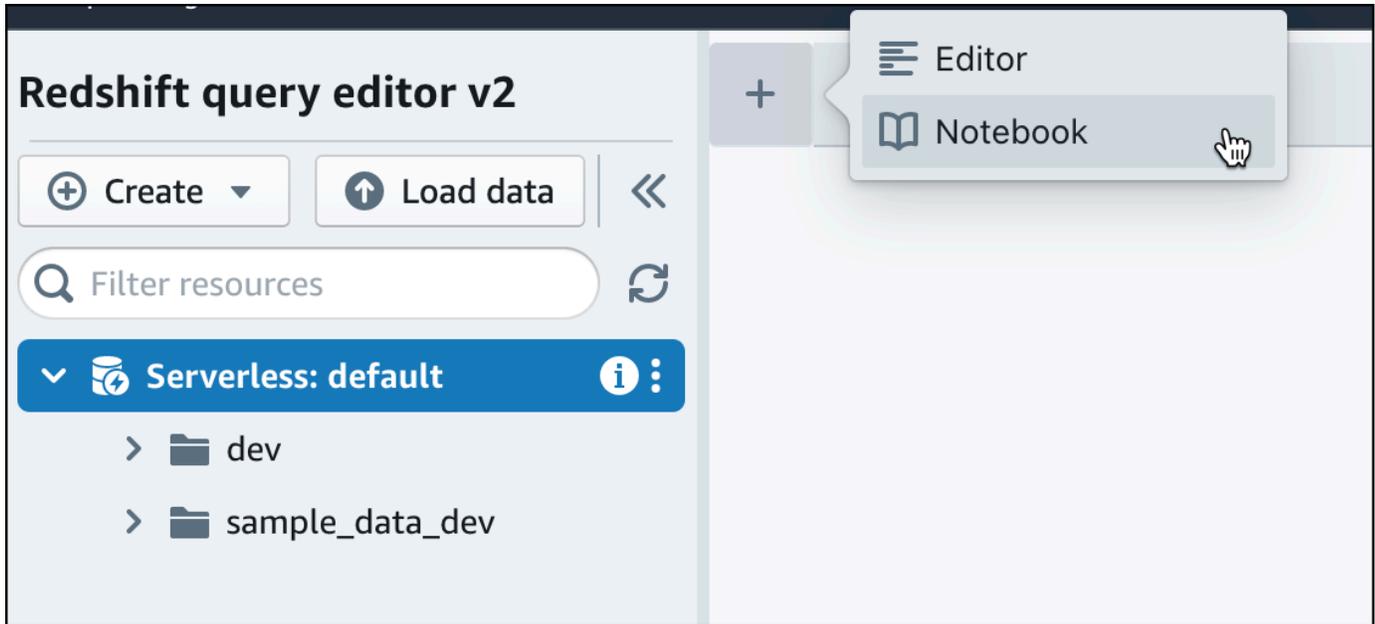
Le fasi seguenti utilizzano i dati all'interno di un bucket S3 pubblico di Amazon Redshift, ma è possibile replicare le stesse fasi utilizzando il proprio bucket S3 e i propri comandi SQL.

Caricamento di dati di esempio da Amazon S3

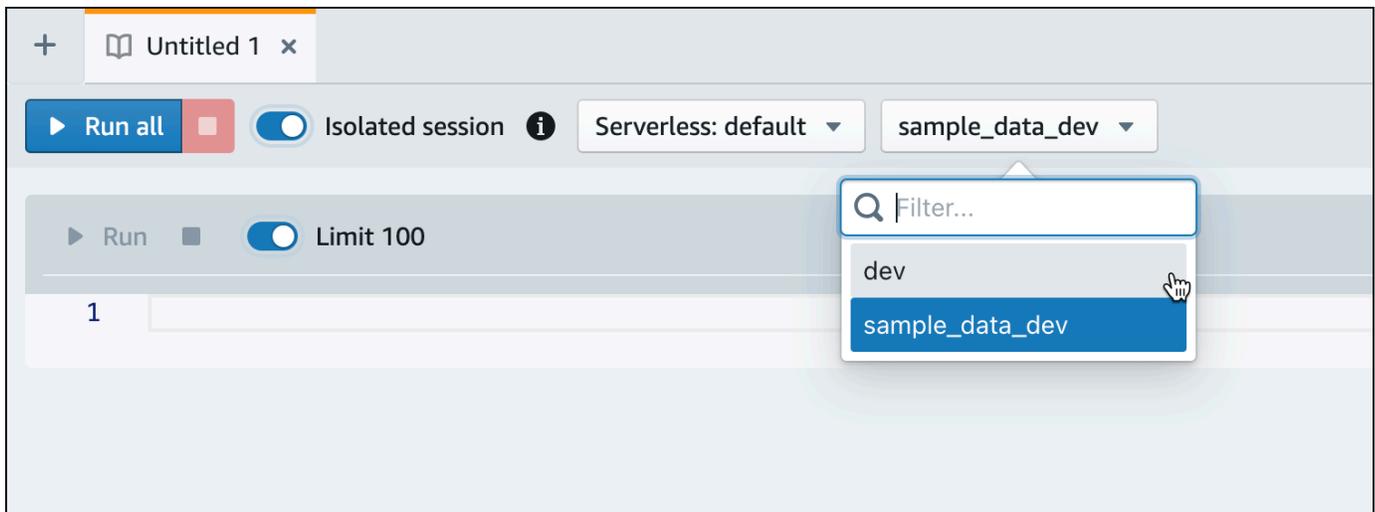
1. Nell'editor di query v2, scegli



Aggiungi, quindi scegli Notebook per creare un nuovo notebook SQL.



2. Passaggio al database dev.



3. Per creare le tabelle.

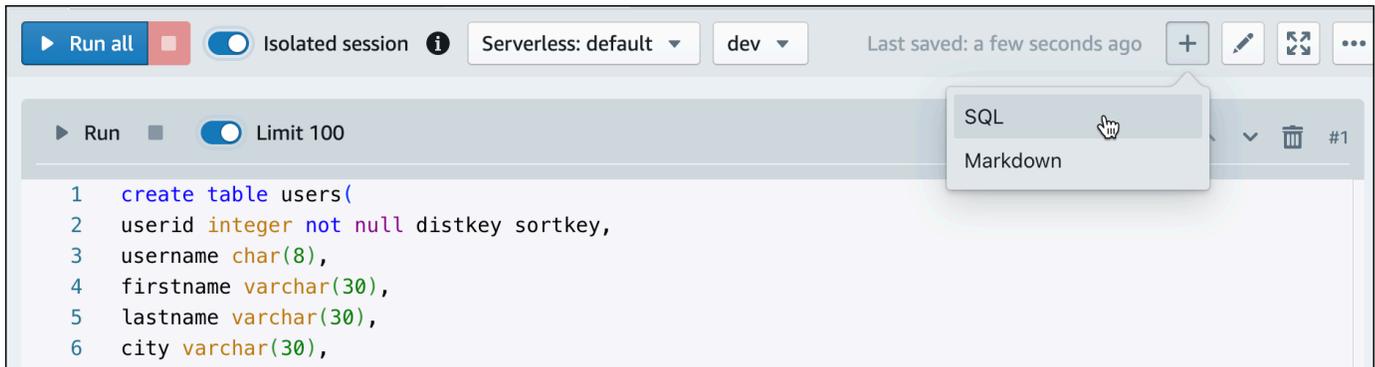
Se utilizzi l'editor di query v2, copia ed esegui le seguenti istruzioni per la creazione delle tabelle nel database dev. Per ulteriori informazioni sulla sintassi, consultare [CREATE TABLE](#) nella Guida per gli sviluppatori di database di Amazon Redshift.

```
create table users(  
userid integer not null distkey sortkey,  
username char(8),  
firstname varchar(30),  
lastname varchar(30),  
city varchar(30),  
state char(2),  
email varchar(100),  
phone char(14),  
likesports boolean,  
liketheatre boolean,  
likeconcerts boolean,  
likejazz boolean,  
likeclassical boolean,  
likeopera boolean,  
likerock boolean,  
likevegas boolean,  
likebroadway boolean,  
likemusicals boolean);
```

```
create table event(  
eventid integer not null distkey,  
venueid smallint not null,  
catid smallint not null,  
dateid smallint not null sortkey,  
eventname varchar(200),  
starttime timestamp);
```

```
create table sales(  
salesid integer not null,  
listid integer not null distkey,  
sellerid integer not null,  
buyerid integer not null,  
eventid integer not null,  
dateid smallint not null sortkey,  
qtysold smallint not null,  
pricepaid decimal(8,2),  
commission decimal(8,2),  
saletime timestamp);
```

4. Nell'editor di query v2, crea una nuova cella SQL nel notebook.



- Utilizza ora il comando COPY nell'editor di query v2 per caricare set di dati di grandi dimensioni da Amazon S3 o da Amazon DynamoDB in Amazon Redshift. Per ulteriori informazioni sulla sintassi di COPY, consultare [COPY](#) nella Guida per gli sviluppatori di database di Amazon Redshift.

Puoi eseguire il comando COPY con alcuni dati di esempio disponibili in un bucket S3 pubblico. Esegui i seguenti comandi SQL nell'editor di query v2.

```
COPY users
FROM 's3://redshift-downloads/tickit/allusers_pipe.txt'
DELIMITER '|'
TIMEFORMAT 'YYYY-MM-DD HH:MI:SS'
IGNOREHEADER 1
REGION 'us-east-1'
IAM_ROLE default;

COPY event
FROM 's3://redshift-downloads/tickit/allevvents_pipe.txt'
DELIMITER '|'
TIMEFORMAT 'YYYY-MM-DD HH:MI:SS'
IGNOREHEADER 1
REGION 'us-east-1'
IAM_ROLE default;

COPY sales
FROM 's3://redshift-downloads/tickit/sales_tab.txt'
DELIMITER '\t'
TIMEFORMAT 'MM/DD/YYYY HH:MI:SS'
IGNOREHEADER 1
REGION 'us-east-1'
IAM_ROLE default;
```

6. Dopo aver caricato i dati, crea un'altra cella SQL nel notebook e prova alcune query di esempio. Per ulteriori informazioni sull'utilizzo del comando SELECT, consultare [SELECT](#) nella Guida per gli sviluppatori di database di Amazon Redshift. Per comprendere la struttura e gli schemi dei dati di esempio, esplora l'utilizzo dell'editor di query v2.

```
-- Find top 10 buyers by quantity.
SELECT firstname, lastname, total_quantity
FROM (SELECT buyerid, sum(qtysold) total_quantity
      FROM sales
      GROUP BY buyerid
      ORDER BY total_quantity desc limit 10) Q, users
WHERE Q.buyerid = userid
ORDER BY Q.total_quantity desc;

-- Find events in the 99.9 percentile in terms of all time gross sales.
SELECT eventname, total_price
FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as
      percentile
      FROM (SELECT eventid, sum(pricepaid) total_price
            FROM sales
            GROUP BY eventid)) Q, event E
WHERE Q.eventid = E.eventid
      AND percentile = 1
ORDER BY total_price desc;
```

Ora che sono stati caricati i dati ed eseguito alcune query di esempio, puoi esplorare altre aree di Amazon Redshift serverless. Consulta l'elenco seguente per ulteriori informazioni su come utilizzare Amazon Redshift serverless.

- Puoi caricare i dati da un bucket Amazon S3. Per ulteriori informazioni, consulta [Caricamento dei dati da Amazon S3](#).
- Puoi utilizzare l'editor di query v2 per caricare dati da un file locale separato da caratteri di dimensioni inferiori a 5 MB. Per ulteriori informazioni, consulta la sezione relativa al [caricamento di dati da un file locale](#).
- Puoi connetterti ad Amazon Redshift serverless con strumenti SQL di terze parti con i driver JDBC e ODBC. Per ulteriori informazioni, consulta [Connessione ad Amazon Redshift serverless](#).
- Puoi anche utilizzare l'Amazon Redshift Data API per connetterti ad Amazon Redshift Serverless. Per ulteriori informazioni, consulta [Uso dell'API dati di Amazon Redshift](#).

- Puoi utilizzare i dati in Amazon Redshift serverless con Redshift ML per creare modelli di machine learning con il comando CREATE MODEL. Consulta il [tutorial: Building customer churn models](#) (Creazione di modelli di abbandono dei clienti) per informazioni su come creare un modello ML di Redshift.
- Puoi eseguire query sui dati da un data lake Amazon S3 senza caricare i dati in Amazon Redshift serverless. Per ulteriori informazioni, consulta [Esecuzione di query in un data lake](#).

Inizia a usare i data warehouse con provisioning di Amazon Redshift

Se sei un utente alle prime armi di Amazon Redshift, ti consigliamo di leggere le seguenti sezioni per iniziare a utilizzare i cluster con provisioning. Il flusso di base di Amazon Redshift consiste nel creare risorse assegnate, connettersi ad Amazon Redshift, caricare dati di esempio e quindi eseguire query sui dati. In questa guida, puoi scegliere di caricare dati di esempio da Amazon Redshift o da un bucket Amazon S3. I dati di esempio vengono utilizzati in tutta la documentazione di Amazon Redshift per dimostrare le funzionalità.

Questo tutorial dimostra come utilizzare i cluster provisioned di Amazon Redshift, che AWS sono oggetti di data warehouse per i quali gestisci le risorse di sistema. Puoi anche utilizzare Amazon Redshift con gruppi di lavoro serverless, che sono oggetti di data warehouse che si ridimensionano automaticamente in base all'utilizzo. Per iniziare a utilizzare Redshift Serverless, consulta [Inizia a usare i data warehouse Serverless Amazon Redshift](#)

Dopo aver creato e effettuato l'accesso alla console di provisioning di Amazon Redshift, puoi creare e gestire oggetti Amazon Redshift, inclusi cluster, nodi e database. Puoi anche eseguire query, visualizzare query ed eseguire altre operazioni SQL Data Definition Language (DDL) e Data Manipulation Language (DML) con un client SQL.

Important

Il cluster fornito per questo esercizio viene eseguito in un ambiente live. Finché è in esecuzione, addebita addebiti a tuo carico. Account AWS Per informazioni sui prezzi, consultare la [pagina dei prezzi di Amazon Redshift](#).

Per evitare costi superflui, è necessario eliminare il cluster al termine dell'esercizio. La sezione finale di questo capitolo spiega come farlo.

Accedi AWS Management Console e apri la console Amazon Redshift all'indirizzo. <https://console.aws.amazon.com/redshiftv2/>

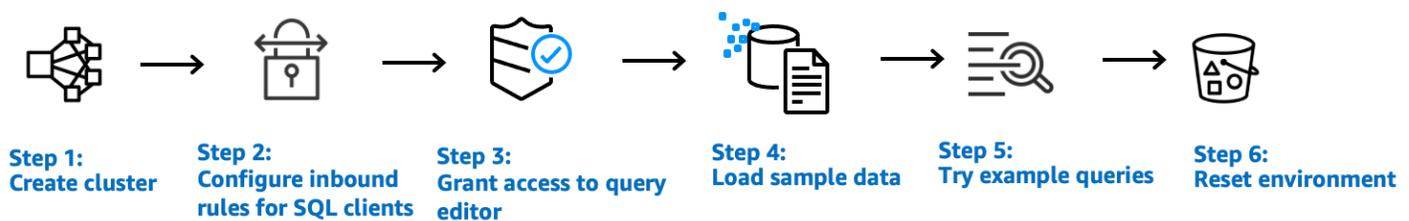
Ti consigliamo di iniziare accedendo alla dashboard dei cluster Provisioned per iniziare a utilizzare la console Amazon Redshift.

A seconda della configurazione, nel pannello di navigazione della console di provisioning di Amazon Redshift vengono visualizzati i seguenti elementi:

- **Redshift Serverless:** accedi e analizza i dati senza la necessità di configurare, ottimizzare e gestire i cluster provisioned di Amazon Redshift.
- **Pannello di controllo dei cluster forniti:** visualizza l'elenco dei cluster presenti Regione AWS, consulta le metriche dei cluster e la panoramica di Query per informazioni dettagliate sui parametri, i dati (come l'utilizzo della CPU) e le informazioni sulle query. L'utilizzo di questi dati può aiutarti a determinare se i dati sulle prestazioni sono anormali in un intervallo di tempo specificato.
- **Cluster:** visualizza l'elenco dei cluster che contiene Regione AWS, scegli un cluster per avviare le query o esegui azioni relative al cluster. È inoltre possibile creare un nuovo cluster da questa pagina.
- **Editor di query:** esegui query su database ospitati nel tuo cluster Amazon Redshift. Ti consigliamo invece di utilizzare l'editor di query v2.
- **Query editor v2** — Amazon Redshift Query Editor v2 è un'applicazione client SQL separata basata sul Web per creare ed eseguire query sul tuo data warehouse Amazon Redshift. Puoi visualizzare i risultati nei grafici e collaborare condividendo le tue domande con altri membri del tuo team.
- **Queries and loads (Query e carichi):** ottiene informazioni per riferimento o risoluzione dei problemi, ad esempio un elenco di query recenti e il testo SQL di ogni query.
- **Unità di condivisione dati** - come amministratore di un account producer puoi autorizzare gli account consumer ad accedere alle unità di condivisione dati o scegliere di non autorizzare alcun accesso. Per utilizzare un datashare autorizzato, un amministratore di account consumer può associare il datashare a un intero o a specifici namespace del cluster in un account. Account AWS Un amministratore può anche rifiutare una unità di condivisione dati.
- **Integrazioni zero-ETL:** gestisci le integrazioni che rendono disponibili i dati transazionali in Amazon Redshift dopo essere stati scritti in fonti supportate.
- **Connessioni IAM Identity Center:** configura la connessione tra Amazon Redshift e IAM Identity Center.
- **Configurations (Configurazioni):** si connette ai cluster di Amazon Redshift dagli strumenti del client SQL tramite le connessioni Java Database Connectivity (JDBC) e Open Database Connectivity (ODBC). Puoi anche configurare un endpoint VPC (Virtual Private Cloud) di Amazon RedShift. In questo modo viene fornita una connessione privata tra un VPC basato sul servizio Amazon VPC che contiene un cluster e un altro VPC che esegue uno strumento client.
- **AWS Integrazione con i partner:** crea l'integrazione con un AWS partner supportato.
- **Advisor - Ricevi consigli specifici sulle modifiche che puoi apportare al cluster Amazon Redshift per dare priorità alle tue ottimizzazioni.**

- **AWS Marketplace:** ottieni informazioni su altri strumenti o AWS servizi compatibili con Amazon Redshift.
- **Allarmi** — Crea allarmi sulle metriche del cluster per visualizzare i dati sulle prestazioni e tenere traccia delle metriche in un periodo di tempo specificato.
- **Eventi** — Tieni traccia degli eventi e ottieni report su informazioni come la data in cui si è verificato l'evento, una descrizione o l'origine dell'evento.
- **Novità:** questa pagina Web elenca le nuove funzionalità di Amazon Redshift e gli aggiornamenti dei prodotti.

In questo tutorial userai la procedura descritta di seguito:



Argomenti

- [Iscrizione a AWS](#)
- [Definizione delle regole del firewall](#)
- [Fase 1: Creazione di un cluster Amazon Redshift di esempio](#)
- [Fase 2: Configurazione delle regole in ingresso per i client SQL](#)
- [Passaggio 3: concedere l'accesso a un client SQL ed eseguire le query](#)
- [Passaggio 4: caricare i dati da Amazon S3 ad Amazon Redshift](#)
- [Fase 5: provare query di esempio utilizzando l'editor di query](#)
- [Fase 6: reimpostazione dell'ambiente](#)

Iscrizione a AWS

Se non ne hai già uno Account AWS, registrane uno. Se disponi già di un account, puoi ignorare questo prerequisito e utilizzare il tuo account esistente.

1. Apri la <https://portal.aws.amazon.com/billing/registrazione>.
2. Segui le istruzioni online.

Parte della procedura di registrazione prevede la ricezione di una telefonata o di un messaggio di testo e l'immissione di un codice di verifica sulla tastiera del telefono.

Quando ti iscrivi a un Account AWS, Utente root dell'account AWS viene creato un. L'utente root dispone dell'accesso a tutte le risorse e tutti i Servizi AWS nell'account. Come best practice di sicurezza, assegna l'accesso amministrativo a un utente e utilizza solo l'utente root per eseguire [attività che richiedono l'accesso di un utente root](#).

Definizione delle regole del firewall

Note

Questo tutorial presuppone che il cluster utilizzi la porta predefinita 5439 e che l'editor di query Amazon Redshift v2 possa essere utilizzato per eseguire comandi SQL. Non fornisce dettagli sulle configurazioni di rete o sulla configurazione di un client SQL che potrebbe essere necessario nel tuo ambiente.

In alcuni ambienti, specifichi una porta quando avvii il cluster Amazon Redshift. Questa porta viene utilizzata insieme all'URL dell'endpoint del cluster per accedere al cluster. È importante inoltre creare una regola in entrata in un gruppo di sicurezza per consentire l'accesso al cluster attraverso la porta.

Se il computer client è protetto da un firewall, è necessario conoscere una porta aperta che è possibile utilizzare. Questa porta aperta consente di stabilire una connessione con il cluster da uno strumento client SQL ed eseguire query. Se non si dispone di questa informazione, chiedere la collaborazione di qualcuno che sia a conoscenza delle regole del firewall di rete al fine di determinare una porta aperta del firewall.

Sebbene Amazon Redshift usi la porta 5439 per impostazione predefinita, se questa porta del firewall non è aperta la connessione non funziona. Non è possibile modificare il numero di porta per il cluster Amazon Redshift dopo la creazione. Pertanto, assicurarsi di specificare una porta aperta che funziona nell'ambiente durante il processo di avvio.

Fase 1: Creazione di un cluster Amazon Redshift di esempio

In questo tutorial, illustrerai il processo per creare un cluster Amazon Redshift con un database. Quindi carichi un set di dati da Amazon S3 nelle tabelle del tuo database. Puoi usare questo cluster di esempio per valutare il servizio Amazon Redshift.

Prima di iniziare a configurare un cluster Amazon Redshift, assicurati di completare tutti i prerequisiti necessari come e. [Iscrizione a AWS](#) [Definizione delle regole del firewall](#)

Per qualsiasi operazione che accede ai dati da un'altra AWS risorsa, il cluster necessita dell'autorizzazione per accedere alla risorsa e ai dati sulla risorsa per tuo conto. Un esempio è l'utilizzo di un comando SQL COPY per caricare dati da Amazon Simple Storage Service (Amazon S3). Fornisci tali autorizzazioni utilizzando AWS Identity and Access Management (IAM). Puoi farlo tramite un ruolo IAM che crei e colleghi al tuo cluster. Per ulteriori informazioni su credenziali e autorizzazioni di accesso, consulta [Credenziali e autorizzazioni di accesso nella Amazon Redshift Database Developer Guide](#).

Come creare un cluster Amazon Redshift

1. Accedi AWS Management Console e apri la console Amazon Redshift all'indirizzo. <https://console.aws.amazon.com/redshiftv2/>

Important

Se si utilizzano le credenziali utente IAM, verificare di disporre delle autorizzazioni necessarie per eseguire le operazioni relative al cluster. Per ulteriori informazioni, consulta [la sezione Sicurezza in Amazon Redshift](#) nella Amazon Redshift Management Guide.

2. Sulla AWS console, scegli Regione AWS dove vuoi creare il cluster.
3. Dal menu di navigazione, scegliere Clusters (Cluster), quindi Create cluster (Crea cluster). Appare la pagina Create cluster (Crea cluster).
4. Nella sezione Configurazione del cluster, specificare i valori per Identificatore del cluster, Tipo di nodo e Nodi:
 - Identificatore del cluster: inserire **examplecluster** per questo tutorial. Questo identificatore deve essere univoco. L'identificatore deve contenere un numero di caratteri compreso tra 1 e 63, dove i caratteri validi sono a-z (solo minuscole) e - (trattino).

- Selezionare uno dei seguenti metodi per dimensionare il cluster:

 Note

Il passaggio seguente presuppone un sistema Regione AWS che supporti i tipi di RA3 nodi. Per un elenco dei tipi Regioni AWS di RA3 nodi di supporto, consulta [Panoramica dei tipi di RA3 nodi](#) nella Amazon Redshift Management Guide. Per ulteriori informazioni sulle specifiche del nodo per ogni tipo e dimensione di nodo, consulta [Dettagli sul tipo di nodo](#).

- Se non si conosce la dimensione del cluster, scegliere Aiutami a scegliere. In questo modo si apre un calcolatore del dimensionamento che ti pone domande sulle dimensioni e sulle caratteristiche di interrogazione dei dati che intendi archiviare nel tuo data warehouse.

Se si conosce la dimensione richiesta del cluster, ovvero il tipo di nodo e il numero di nodi, scegliere Sceglierò. Quindi scegliere il tipo di nodo e il numero di nodi per dimensionare il cluster.

Per questo tutorial, scegli ra3.4xlarge per il tipo di nodo e 2 per il numero di nodi.

Se è disponibile una scelta per la configurazione AZ, scegli Single-AZ.

- Per utilizzare il set di dati di esempio fornito da Amazon Redshift, in Dati di esempio, scegliere Carica dati di esempio. Amazon Redshift carica di default il set di dati di esempio Tickit nel database dev e nello schema public.

5. Nella sezione Configurazione del database specifica i valori per Nome dell'utente amministratore. In Password dell'amministratore seleziona una delle opzioni seguenti:

- Genera una password: utilizza una password generata da Amazon Redshift.
- Aggiungi manualmente una password dell'amministratore: utilizza una tua password.
- Gestisci le credenziali di amministratore in AWS Secrets Manager: Amazon Redshift le AWS Secrets Manager utilizza per generare e gestire la password di amministratore. L'utilizzo AWS Secrets Manager per generare e gestire la password segreta comporta un costo. Per informazioni sui prezzi di AWS Secrets Manager , consulta [Prezzi di AWS Secrets Manager](#).

Per questo tutorial utilizzare i seguenti valori:

- Nome utente amministratore: inserisci **awsuser**.
 - Password utente amministratore: inserisci **Changeit1** la password.
6. Per questo tutorial, create un ruolo IAM e impostatelo come predefinito per il cluster, come descritto di seguito. È possibile che sia disponibile un solo set di ruoli IAM predefinito per un cluster.
- a. In Autorizzazioni cluster, per Gestisci i ruoli IAM, scegli Creazione di ruolo IAM.
 - b. Specificare un bucket Amazon S3 per l'accesso al ruolo IAM con uno dei seguenti metodi:
 - Scegliere Nessun bucket Amazon S3 aggiuntivo per consentire al ruolo IAM creato di accedere solo ai bucket Amazon S3 denominati `redshift`.
 - Scegliere Qualsiasi bucket Amazon S3 per consentire al ruolo IAM creato di accedere a tutti bucket Amazon S3.
 - Scegliere Bucket Amazon S3 specifico per specificare uno o più bucket Amazon S3 a cui il ruolo IAM creato può accedere. Scegli quindi uno o più bucket Amazon S3 dalla tabella.
 - c. Scegliere Creazione di ruolo IAM come predefinito. Amazon Redshift crea e imposta automaticamente il ruolo IAM come predefinito per il cluster.

Poiché hai creato il tuo ruolo IAM dalla console, questo ha allegata la policy `AmazonRedshiftAllCommandsFullAccess`. Ciò consente ad Amazon Redshift di copiare, caricare, analizzare i dati ed eseguire query su di essi dalle risorse Amazon nel tuo account IAM.

Per informazioni su come gestire il ruolo IAM predefinito per un cluster, consulta [Creazione di un ruolo IAM come predefinito per Amazon Redshift](#) nella Amazon Redshift Management Guide.

7. (Facoltativo) Nella sezione Additional configurations (Configurazioni aggiuntive), disattivare Use defaults (Usa valori di default) per modificare le impostazioni di Network and security (Rete e sicurezza), Database configuration (Configurazione del database), Maintenance (Manutenzione), Monitoring (Monitoraggio) e Backup.

In alcuni casi, è possibile creare il cluster con l'opzione Caricare i dati campione e attivare il routing avanzato Amazon VPC. In tal caso, il cluster nel tuo virtual private cloud (VPC) richiede l'accesso all'endpoint Amazon S3 per caricare i dati.

Per rendere il cluster accessibile al pubblico, è possibile eseguire due operazioni. Puoi configurare un indirizzo NAT (Network Address Translation) nel VPC affinché il cluster acceda

a Internet. Oppure puoi configurare un endpoint Amazon S3 VPC nel tuo VPC. Per ulteriori informazioni sul routing avanzato di Amazon VPC, consulta la sezione Enhanced [Amazon VPC routing nella](#) Amazon Redshift Management Guide.

8. Scegli Create cluster (Crea cluster). Attendi che il cluster venga creato con **Available** lo stato nella pagina Clusters.

Fase 2: Configurazione delle regole in ingresso per i client SQL

Note

Ti consigliamo di saltare questo passaggio e accedere al cluster utilizzando Amazon Redshift Query Editor v2.

Più avanti in questo tutorial, si accederà al cluster da un Virtual Private Cloud (VPC) basato sul servizio Amazon VPC. Tuttavia, se per accedere al cluster si utilizza un client SQL esterno al firewall, è necessario concedere l'accesso in entrata.

Per controllare il firewall e concedere l'accesso in ingresso al cluster

1. Se necessario accedere al cluster dall'esterno di un firewall, controlla le regole con cui è configurato. Ad esempio, il tuo cliente potrebbe essere un'istanza Amazon Elastic Compute Cloud (Amazon EC2) o un computer esterno.

Per ulteriori informazioni sulle regole del firewall, consulta le [regole dei gruppi di sicurezza](#) nella Amazon EC2 User Guide.

2. Per accedere da un client EC2 esterno Amazon, aggiungi una regola di ingresso al gruppo di sicurezza collegato al cluster che consente il traffico in entrata. Aggiungi le regole EC2 dei gruppi di sicurezza Amazon nella EC2 console Amazon. Ad esempio, uno CIDR/IP di 192.0.2.0/24 consente ai client in quell'intervallo di indirizzi IP di connettersi al cluster. Scoprite quello più adatto al vostro ambiente. CIDR/IP

Passaggio 3: concedere l'accesso a un client SQL ed eseguire le query

Per interrogare i database ospitati dal tuo cluster Amazon Redshift, hai diverse opzioni per i client SQL. Ciò include:

- Connettiti al tuo cluster ed esegui query utilizzando Amazon Redshift Query Editor v2.

Se utilizzi l'editor di query v2, non devi scaricare e configurare un'applicazione client SQL. Avvia Amazon Redshift Query Editor v2 dalla console Amazon Redshift.

- Connect al cluster utilizzando RSQL. Per ulteriori informazioni, consulta [Connecting with Amazon Redshift RSQL](#) nella Amazon Redshift Management Guide.
- Connettiti al tuo cluster tramite uno strumento client SQL, come SQL Workbench/J. Per ulteriori informazioni, consulta [Connettiti al tuo cluster utilizzando SQL Workbench/J nella Amazon Redshift Management Guide](#).

Questo tutorial utilizza Amazon Redshift Query Editor v2 come un modo semplice per eseguire query su database ospitati dal tuo cluster Amazon Redshift. Dopo aver creato il cluster, puoi eseguire immediatamente le query. Per informazioni dettagliate sulle considerazioni relative all'utilizzo dell'editor di query di Amazon Redshift v2, [consulta Considerazioni sull'utilizzo dell'editor di query v2](#) nella Amazon Redshift Management Guide.

Concessione dell'accesso all'editor di query v2

La prima volta che un amministratore configura l'editor di query v2 per te Account AWS, sceglie AWS KMS key quello utilizzato per crittografare le risorse dell'editor di query v2. Le risorse dell'editor di query di Amazon Redshift v2 includono query salvate, notebook e grafici. Per impostazione predefinita, una chiave di proprietà AWS viene utilizzata per crittografare le risorse. In alternativa, un amministratore può utilizzare una chiave gestita dal cliente scegliendo l'Amazon Resource Name (ARN) per la chiave nella pagina di configurazione. Dopo aver configurato un account, le impostazioni di AWS KMS crittografia non possono essere modificate. Per ulteriori informazioni, consulta [Configuring your Account AWS](#) nella Amazon Redshift Management Guide.

Per accedere all'editor della query v2, sono necessarie le opportune autorizzazioni. Un amministratore può collegare una delle policy AWS gestite per Amazon Redshift Query Editor v2 al ruolo o all'utente IAM per concedere le autorizzazioni. Queste politiche AWS gestite sono scritte con diverse opzioni che controllano il modo in cui l'etichettatura delle risorse consente la condivisione

delle query. Puoi utilizzare la console IAM (<https://console.aws.amazon.com/iam/>) per allegare le policy IAM. Per ulteriori informazioni su queste politiche, consulta [Accedere all'editor di query v2](#) nella Amazon Redshift Management Guide.

Puoi anche creare una policy in base alle autorizzazioni consentite e negate nelle policy gestite fornite. Se si utilizza l'editor policy della console IAM per creare le proprie policy, scegliere SQL Workbench come servizio per il quale si crea la policy nell'editor visivo. L'editor di query v2 utilizza il nome di servizio AWS SQL Workbench nell'editor visivo e in IAM Policy Simulator.

Per informazioni, consulta [Utilizzo dell'editor di query v2](#) nella Guida alla gestione di Amazon Redshift.

Passaggio 4: caricare i dati da Amazon S3 ad Amazon Redshift

Dopo aver creato il cluster, puoi caricare dati da Amazon S3 nelle tabelle del database. Esistono diversi modi per caricare dati da Amazon S3.

- È possibile utilizzare un client SQL per eseguire il comando SQL CREATE TABLE per creare una tabella nel database e quindi utilizzare il comando SQL COPY per caricare dati da Amazon S3. L'editor di query Amazon Redshift v2 è un client SQL.
- Puoi utilizzare la procedura guidata di caricamento dell'editor di query di Amazon Redshift v2.

Questo tutorial dimostra come utilizzare l'editor di query Amazon Redshift v2 per eseguire comandi SQL per CREARE tabelle e COPIARE dati. Avvia Query editor v2 dal pannello di navigazione della console Amazon Redshift. All'interno di Query Editor v2, crea una connessione al `examplecluster` cluster e al database denominata `dev` con il nome dell'utente amministratore `awsuser`. Per questo tutorial scegli Credenziali temporanee che utilizzano un nome utente del database quando crei la connessione. Per informazioni dettagliate sull'utilizzo dell'editor di query di Amazon Redshift v2, consulta [Connessione a un database Amazon Redshift nella Amazon Redshift Management Guide](#).

Caricamento di dati da Amazon S3 tramite comandi SQL

Nel riquadro dell'editor di query v2, conferma di essere connesso al `examplecluster` cluster e `dev` al database. Quindi, crea tabelle nel database e carica i dati nelle tabelle. In questo tutorial, i dati che carichi sono disponibili in un bucket Amazon S3 accessibile da molti. Regioni AWS

La procedura seguente crea tabelle e carica dati da un bucket Amazon S3 pubblico.

Usa l'editor di query di Amazon Redshift v2 per copiare ed eseguire la seguente istruzione create table per creare una tabella nello public schema del database. Per ulteriori informazioni sulla sintassi, consultare [CREATE TABLE](#) nella Guida per gli sviluppatori di database di Amazon Redshift.

Per creare e caricare dati utilizzando un client SQL come Query Editor v2

1. Esegui il seguente comando SQL per CREARE la sales tabella.

```
drop table if exists sales;
create table sales(
salesid integer not null,
listid integer not null distkey,
sellerid integer not null,
buyerid integer not null,
eventid integer not null,
dateid smallint not null sortkey,
qtysold smallint not null,
pricepaid decimal(8,2),
commission decimal(8,2),
saletime timestamp);
```

2. Esegui il seguente comando SQL per CREARE la date tabella.

```
drop table if exists date;
create table date(
dateid smallint not null distkey sortkey,
caldate date not null,
day character(3) not null,
week smallint not null,
month character(5) not null,
qtr character(5) not null,
year smallint not null,
holiday boolean default('N'));
```

3. Carica la sales tabella da Amazon S3 utilizzando il comando COPY.

Note

Consigliamo di utilizzare il comando COPY per caricare set di dati di grandi dimensioni in Amazon Redshift da Amazon S3. Per ulteriori informazioni sulla sintassi di COPY, consultare [COPY](#) nella Guida per gli sviluppatori di database di Amazon Redshift.

Per caricare i dati di esempio, è necessario fornire al cluster l'autenticazione per accedere ad Amazon S3 per tuo conto. Fornisci l'autenticazione facendo riferimento al ruolo IAM che hai creato e impostato come cluster quando hai scelto Crea ruolo IAM come predefinito quando hai creato il cluster. default

Carica la sales tabella utilizzando il seguente comando SQL. Facoltativamente, puoi scaricare e visualizzare da Amazon S3 [i dati di origine per sales la](#) tabella. .

```
COPY sales
FROM 's3://redshift-downloads/tickit/sales_tab.txt'
DELIMITER '\t'
TIMEFORMAT 'MM/DD/YYYY HH:MI:SS'
REGION 'us-east-1'
IAM_ROLE default;
```

4. Caricate la date tabella utilizzando il seguente comando SQL. Facoltativamente, puoi scaricare e visualizzare da Amazon S3 [i dati di origine per date la](#) tabella. .

```
COPY date
FROM 's3://redshift-downloads/tickit/date2008_pipe.txt'
DELIMITER '|'
REGION 'us-east-1'
IAM_ROLE default;
```

Caricamento di dati da Amazon S3 utilizzando l'editor di query v2

Questa sezione descrive il caricamento dei tuoi dati in un cluster Amazon Redshift. L'editor di query v2 semplifica il caricamento dei dati quando si utilizza la procedura guidata Load data. Il comando COPY generato e utilizzato nell'editor di query v2 Load data wizard supporta molti dei parametri disponibili nella sintassi del comando COPY per caricare dati da Amazon S3. Per informazioni sul comando COPY e sulle opzioni utilizzate per copiare il caricamento da Amazon S3, consultare

[COPIA da Amazon Simple Storage Service](#) nella Guida per sviluppatori di database Amazon Redshift.

Per caricare i tuoi dati da Amazon S3 ad Amazon Redshift, Amazon Redshift richiede un ruolo IAM con i privilegi necessari per caricare i dati dal bucket Amazon S3 specificato.

Per caricare i tuoi dati da Amazon S3 ad Amazon Redshift, puoi utilizzare la procedura guidata di caricamento dei dati dell'editor di query v2. Per informazioni su come utilizzare la procedura guidata di [caricamento dei dati](#), consulta [Loading data from Amazon S3 nella Amazon Redshift Management Guide](#).

Crea dati TICKIT nel tuo cluster

TICKIT è un database di esempio che puoi caricare facoltativamente nel tuo cluster Amazon Redshift per imparare a interrogare i dati in Amazon Redshift. Puoi creare il set completo di tabelle TICKIT e caricare dati nel tuo cluster nei seguenti modi:

- Quando crei un cluster nella console Amazon Redshift, hai la possibilità di caricare contemporaneamente dati TICKIT di esempio. Sulla console Amazon Redshift, scegli Clusters, Crea cluster. Nella sezione Dati di esempio, seleziona Carica dati di esempio Amazon Redshift carica automaticamente il set di dati di esempio nel dev database del cluster Amazon Redshift durante la creazione del cluster.
- Per connetterti a un cluster esistente, procedi come segue:
 - Nella console Amazon Redshift, scegli Clusters dalla barra di navigazione.
 - Scegli il tuo cluster dal riquadro Cluster.
 - Scegli Query data, Query in Query Editor v2.
 - Espandi examplecluster nell'elenco delle risorse. Se è la prima volta che ti connetti al cluster, viene visualizzato Connect to examplecluster. Scegli nome utente e password del database. Lascia il database così com'è **dev**. **awsuser** Specificare il nome **Changeit1** utente e la password.
 - Scegli Crea connessione.
- Con Amazon Redshift Query Editor v2, puoi caricare i dati TICKIT in un database di esempio denominato sample_data_dev. Scegli il database sample_data_dev nell'elenco delle risorse. Accanto al nodo tickit, scegli l'icona Apri taccuini di esempio. Conferma di voler creare il database di esempio.

- Amazon Redshift Query Editor v2 crea il database di esempio insieme a un notebook di esempio denominato. `tickit-sample-notebook` Puoi scegliere Esegui tutto per eseguire questo notebook per interrogare i dati nel database di esempio.

Per visualizzare i dettagli sui dati TICKIT, consulta Database di [esempio nella Amazon Redshift Database Developer Guide](#).

Fase 5: provare query di esempio utilizzando l'editor di query

Per configurare e utilizzare l'editor di query di Amazon Redshift v2 per interrogare un database, consulta [Working with Query Editor v2](#) nella Amazon Redshift Management Guide.

Ora, provare alcune query di esempio, come illustrato di seguito. Per creare nuove query nell'editor di query v2, scegli l'icona + in alto a destra del riquadro delle query e scegli SQL. Viene visualizzata una nuova pagina di query in cui è possibile copiare e incollare le seguenti query SQL.

Note

Assicurati di eseguire prima la prima query nel notebook, che imposta il valore di configurazione del `search_path` server `tickit` sullo schema utilizzando il seguente comando SQL:

```
set search_path to tickit;
```

Per ulteriori informazioni sull'utilizzo del comando `SELECT`, consulta [SELECT](#) nella Amazon Redshift Database Developer Guide.

```
-- Get definition for the sales table.
SELECT *
FROM pg_table_def
WHERE tablename = 'sales';
```

```
-- Find total sales on a given calendar date.
SELECT sum(qtysold)
FROM sales, date
WHERE sales.dateid = date.dateid
```

```
AND    caldate = '2008-01-05';
```

```
-- Find top 10 buyers by quantity.
SELECT firstname, lastname, total_quantity
FROM    (SELECT buyerid, sum(qtysold) total_quantity
         FROM    sales
         GROUP BY buyerid
         ORDER BY total_quantity desc limit 10) Q, users
WHERE Q.buyerid = userid
ORDER BY Q.total_quantity desc;
```

```
-- Find events in the 99.9 percentile in terms of all time gross sales.
SELECT eventname, total_price
FROM    (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as
         percentile
         FROM (SELECT eventid, sum(pricepaid) total_price
              FROM    sales
              GROUP BY eventid)) Q, event E
WHERE Q.eventid = E.eventid
      AND percentile = 1
ORDER BY total_price desc;
```

Fase 6: reimpostazione dell'ambiente

Nei passaggi precedenti, hai creato con successo un cluster Amazon Redshift, caricato i dati in tabelle e interrogato i dati utilizzando un client SQL come Amazon Redshift Query Editor v2.

Una volta completato il presente tutorial, consigliamo di reimpostare l'ambiente allo stato precedente eliminando il cluster di esempio. I costi del servizio Amazon Redshift continueranno ad essere addebitati fino all'eliminazione del cluster.

Tuttavia, potresti voler mantenere attivo il cluster di esempio se intendi provare attività in altre guide o attività di Amazon Redshift descritte in: [Esegui comandi per definire e utilizzare un database nel tuo data warehouse](#)

Per eliminare un cluster

1. Accedi AWS Management Console e apri la console Amazon Redshift all'indirizzo. <https://console.aws.amazon.com/redshiftv2/>
2. Dal menu di navigazione, scegliere Clusters (Cluster) per visualizzare l'elenco dei cluster.

3. Scegli il cluster `examplecluster`. In Actions (Azioni), scegliere Delete (Elimina). Il cluster di esempi Delete? viene visualizzata la pagina.
4. Conferma il cluster da eliminare, deseleziona l'impostazione Crea istantanea finale, quindi accedi **delete** per confermare l'eliminazione. Scegli Delete cluster (Elimina cluster).

Nella pagina dell'elenco dei cluster, lo stato del cluster viene aggiornato nel corso dell'eliminazione.

Una volta completato il tutorial, sarà possibile trovare ulteriori informazioni su Amazon Redshift e sulle fasi successive in [Risorse aggiuntive per informazioni su Amazon Redshift](#).

Esegui comandi per definire e utilizzare un database nel tuo data warehouse

Sia i data warehouse Serverless Redshift che i data warehouse forniti da Amazon Redshift contengono database. Dopo aver avviato il data warehouse, puoi gestire la maggior parte delle azioni del database utilizzando i comandi SQL. Con poche eccezioni, la funzionalità e la sintassi di SQL sono le stesse per tutti i database Amazon Redshift. Per i dettagli sui comandi SQL disponibili con Amazon Redshift, consulta [i comandi SQL](#) nella Amazon Redshift Database Developer Guide.

Quando crei il tuo data warehouse, nella maggior parte degli scenari, Amazon Redshift crea anche il database predefinito `dev`. Dopo esserti connesso al dev database, puoi creare un altro database.

Le seguenti sezioni illustrano le attività di database più comuni quando si lavora con i database Amazon Redshift. Le attività iniziano con la creazione di un database e se continui fino all'ultima attività puoi eliminare tutte le risorse che crei eliminando il database.

Per gli esempi in questa sezione si assume quanto riportato di seguito:

- Hai creato un data warehouse Amazon Redshift.
- Hai stabilito una connessione al data warehouse dal tuo strumento client SQL, ad esempio l'editor di query Amazon Redshift v2. Per ulteriori informazioni sull'editor di query v2, consulta [Interrogare un database utilizzando l'editor di query Amazon Redshift v2](#) nella Amazon Redshift Management Guide.

Argomenti

- [Connessione ai data warehouse Amazon Redshift](#)
- [Creazione di un database](#)
- [Creazione di un utente](#)
- [Crea uno schema](#)
- [Creare una tabella](#)
- [Caricare i dati](#)
- [Interroga le tabelle e le viste del sistema](#)
- [Annullare una query](#)

Connessione ai data warehouse Amazon Redshift

Per connetterti ai cluster Amazon Redshift, dalla pagina Clusters della console Amazon Redshift, espandi Connect to Amazon Redshift cluster ed esegui una delle seguenti operazioni:

- Scegli Query data per utilizzare l'editor di query v2 per eseguire query su database ospitati dal tuo cluster Amazon Redshift. Dopo aver creato il cluster, puoi eseguire immediatamente le query con l'editor della query v2.

Per ulteriori informazioni, consulta [Interrogare un database utilizzando l'editor di query di Amazon Redshift v2](#) nella Amazon Redshift Management Guide.

- In Lavora con gli strumenti client, scegli il tuo cluster e connettiti ad Amazon Redshift dagli strumenti client utilizzando i driver JDBC o ODBC copiando l'URL del driver JDBC o ODBC. Usa questo URL dal tuo computer o istanza client. Codifica le tue applicazioni affinché usino operazioni API di accesso ai dati JDBC o ODBC e gli strumenti del client SQL che supportano JDBC o ODBC.

Per ulteriori informazioni su come trovare la stringa di connessione del cluster, consultare [Ricerca della stringa di connessione al cluster](#).

- Se il tuo strumento client SQL richiede un driver, puoi scegliere il driver JDBC o ODBC per scaricare un driver specifico del sistema operativo per la connessione ad Amazon Redshift dagli strumenti client.

Per ulteriori informazioni su come installare il driver appropriato per il client SQL, consultare [Configurazione di una connessione del driver JDBC versione 2.0](#).

Per ulteriori informazioni su come configurare una connessione ODBC, consultare [Configurazione di una connessione ODBC](#).

Per connetterti al data warehouse Serverless di Redshift, dalla pagina dashboard Serverless della console di Amazon Redshift, esegui una delle seguenti operazioni:

- Usa l'editor di query Amazon Redshift v2 per eseguire query su database ospitati dal tuo data warehouse Serverless Redshift. Dopo aver creato il tuo data warehouse, puoi eseguire immediatamente le query utilizzando l'editor di query v2.

Per ulteriori informazioni, consultare [Esecuzione di query su un database con l'editor di query v2 di Amazon Redshift](#).

- Connettiti ad Amazon Redshift dagli strumenti client utilizzando i driver JDBC o ODBC copiando l'URL del driver JDBC o ODBC.

Per lavorare con i dati nel data warehouse, sono necessari i driver JDBC o ODBC per la connettività dal computer o dall'istanza client. Codifica le tue applicazioni affinché usino operazioni API di accesso ai dati JDBC o ODBC e gli strumenti del client SQL che supportano JDBC o ODBC.

Per ulteriori informazioni su come trovare la stringa di connessione, consulta [Connecting to Redshift Serverless](#) nella Amazon Redshift Management Guide.

Creazione di un database

Dopo aver verificato che il data warehouse sia attivo e funzionante, puoi creare un database. Il database è l'elemento in cui effettivamente saranno create tabelle, saranno caricati i dati e saranno eseguite le query. Un data warehouse può ospitare più database. Ad esempio, è possibile avere un database per i dati di vendita denominati SALESDB e un database per i dati degli ordini denominati ORDERSDB nello stesso data warehouse.

Per creare un database denominato **SALESDB**, esegui il comando seguente nello strumento client SQL.

```
CREATE DATABASE salesdb;
```

Note

Dopo aver eseguito il comando, assicurati di aggiornare l'elenco degli oggetti dello strumento SQL Client nel tuo data warehouse per vedere il nuovo `salesdb`.

Per questo esercizio, accetteremo le impostazioni predefinite. Per informazioni sulle opzioni di altri comandi, consultare [CREATE DATABASE](#) nella Guida per gli sviluppatori di database di Amazon Redshift. Per eliminare un database e il relativo contenuto, consulta [DROP DATABASE](#) nella Amazon Redshift Database Developer Guide.

Dopo aver creato il database SALESDB, sarà possibile connettersi al nuovo database dal client SQL. Usa gli stessi parametri di connessione usati per la connessione corrente, ma modifica il nome del database in SALESDB.

Creazione di un utente

Per impostazione predefinita, solo l'utente amministratore che hai creato all'avvio del data warehouse ha accesso al database predefinito nel data warehouse. Per concedere l'accesso ad altri utenti, è necessario creare uno o più account. Gli account utente del database sono globali in tutti i database di un data warehouse e non per ogni singolo database.

Utilizzare il comando `CREATE USER` per creare un nuovo utente. Quando crei un nuovo utente, devi specificare il nome del nuovo utente e una password. Si consiglia di specificare una password per l'utente. Deve essere costituita da 8 a 64 caratteri e deve includere almeno una lettera maiuscola, una minuscola e un valore numerico.

Ad esempio, per creare un utente denominato **GUEST** con la password **ABCd4321**, emettere il comando seguente:

```
CREATE USER GUEST PASSWORD 'ABCd4321';
```

Per connettersi al database SALESDB come utente GUEST, utilizzare la stessa password utilizzata al momento della creazione dell'utente, ad esempio ABCd4321.

Per informazioni sulle opzioni di altri comandi, consultare [CREATE USER](#) nella Guida per gli sviluppatori di database di Amazon Redshift.

Crea uno schema

Dopo aver creato un nuovo database, è possibile creare un nuovo schema nel database corrente. Uno schema è uno spazio dei nomi che contiene oggetti di database denominati come tabelle, viste e funzioni definite dall'utente (). UDFs Un database può contenere uno o più schemi e ogni schema appartiene a un solo database. Due schemi possono avere oggetti diversi che condividono lo stesso nome.

È possibile creare più schemi nello stesso database per organizzare i dati nel modo desiderato o per raggruppare i dati in modo funzionale. Ad esempio, è possibile creare uno schema per memorizzare tutti i dati dell'area di gestione temporanea e un altro schema per memorizzare tutte le tabelle di reporting. È inoltre possibile creare schemi diversi per memorizzare i dati relativi a diversi business group che si trovano nello stesso database. Ogni schema può memorizzare diversi oggetti di database, come tabelle, viste e funzioni definite dall'utente (). UDFs Inoltre, è possibile creare schemi

con la clausola `AUTHORIZATION`. Questa clausola conferisce la proprietà a un utente specificato o imposta una quota sulla quantità massima di spazio su disco che lo schema specificato può utilizzare.

Amazon Redshift crea automaticamente uno schema chiamato `public` per ogni nuovo database. Quando non si specifica il nome dello schema durante la creazione di oggetti di database, gli oggetti vanno nello schema `public`.

Per accedere a un oggetto in uno schema, qualificare l'oggetto utilizzando la notazione `schema_name.table_name`. Il nome qualificato dello schema è costituito dal nome dello schema e dal nome della tabella separati da un punto. Ad esempio, è possibile avere uno schema `sales` che ha una tabella `price` e uno schema `inventory` che ha una tabella `price`. Quando si fa riferimento alla tabella `price`, è necessario qualificarla come `sales.price` o `inventory.price`.

Nell'esempio seguente viene creato uno schema denominato **SALES** per l'utente `GUEST`.

```
CREATE SCHEMA SALES AUTHORIZATION GUEST;
```

Per informazioni sulle opzioni di altri comandi, consultare [CREATE TABLE](#) nella Guida per gli sviluppatori di database di Amazon Redshift.

Per visualizzare l'elenco degli schemi nel database, emettere il comando seguente.

```
select * from pg_namespace;
```

L'output visualizzato dovrebbe essere simile al seguente:

nspname	nspowner	nspacl
sales	100	
pg_toast	1	
pg_internal	1	
catalog_history	1	
pg_temp_1	1	
pg_catalog	1	{rdsdb=UC/rdsdb,=U/rdsdb}
public	1	{rdsdb=UC/rdsdb,=U/rdsdb}
information_schema	1	{rdsdb=UC/rdsdb,=U/rdsdb}

Per ulteriori informazioni su come eseguire query sulle tabelle di catalogo, consultare [Query sulle tabelle di catalogo](#) nella Guida per gli sviluppatori di database di Amazon Redshift.

Utilizzare l'istruzione `GRANT` per concedere autorizzazioni agli utenti per gli schemi.

L'esempio seguente concede all'GUESTutente il privilegio di selezionare i dati da tutte le tabelle o viste dello SALES schema utilizzando un'istruzione SELECT.

```
GRANT SELECT ON ALL TABLES IN SCHEMA SALES TO GUEST;
```

L'esempio seguente concede all'utente tutti i privilegi disponibili contemporaneamente. GUEST

```
GRANT ALL ON SCHEMA SALES TO GUEST;
```

Creare una tabella

Dopo aver creato il nuovo database, è necessario creare le tabelle che conterranno i dati. Specificate le informazioni sulla colonna quando create la tabella.

Ad esempio, per creare una tabella denominata **DEMO**, emettere il comando riportato di seguito.

```
CREATE TABLE Demo (  
  PersonID int,  
  City varchar (255)  
);
```

Per impostazione predefinita, i nuovi oggetti di database, come le tabelle, vengono creati nello schema predefinito denominato `public` creato durante la creazione del data warehouse. È possibile utilizzare un altro schema per creare oggetti di database. Per ulteriori informazioni, consultare [Gestione della sicurezza dei database](#) nella Guida per gli sviluppatori di database di Amazon Redshift.

È possibile creare una tabella anche utilizzando la notazione `schema_name.object_name` per creare la tabella nello schema SALES.

```
CREATE TABLE SALES.DEMO (  
  PersonID int,  
  City varchar (255)  
);
```

Per visualizzare e ispezionare gli schemi e le relative tabelle, puoi utilizzare l'editor di query di Amazon Redshift v2. Oppure è possibile visualizzare l'elenco delle tabelle negli schemi usando le viste di sistema. Per ulteriori informazioni, consulta [Interroga le tabelle e le viste del sistema](#).

Le colonne `encoding`, `distkey` e `sortkey` vengono utilizzate da Amazon Redshift per l'elaborazione parallela. Per ulteriori informazioni sulla progettazione di tabelle che includono questi elementi, consultare [Best practice di Amazon Redshift per la progettazione di tabelle](#).

Inserisci le righe di dati in una tabella

Dopo aver creato una tabella, inserire le righe di dati in questa tabella.

Note

Il comando [INSERT](#) inserisce righe in una tabella. Per caricamenti in blocco standard, utilizzare il comando [COPY](#). Per ulteriori informazioni, consultare [Utilizzo del comando COPY per caricare i dati](#).

Ad esempio, per inserire valori nella tabella DEMO, emettere il comando seguente.

```
INSERT INTO DEMO VALUES (781, 'San Jose'), (990, 'Palo Alto');
```

Per inserire dati in una tabella che fa parte di uno schema specifico, esegui il comando seguente.

```
INSERT INTO SALES.DEMO VALUES (781, 'San Jose'), (990, 'Palo Alto');
```

Selezione dati da una tabella

Dopo aver creato una tabella e averla popolata con dati, puoi usare l'istruzione `SELECT` per visualizzare i dati contenuti nella tabella. L'istruzione `SELECT *` restituisce tutti i nomi di colonna e i valori di riga per tutti i dati presenti in una tabella. L'utilizzo di `SELECT` è un buon metodo per verificare che i dati aggiunti di recente siano stati inseriti correttamente nella tabella.

Per visualizzare i dati inseriti nella tabella **DEMO**, emettere il comando seguente:

```
SELECT * from DEMO;
```

Il risultato sarà simile al seguente:

```
personid | city
-----+-----
       781 | San Jose
       990 | Palo Alto
```

```
(2 rows)
```

Per ulteriori informazioni sull'uso dell'istruzione SELECT per eseguire query sulle tabelle, consultare [SELECT](#).

Caricare i dati

Molti degli esempi di questa guida utilizzano il set di dati di esempio di TICKIT. È possibile scaricare il file [ticketdb.zip](#) contenente i singoli file di dati di esempio. Puoi quindi caricare i dati di esempio nel tuo bucket Amazon S3.

Per caricare i dati campione per il database, creare innanzitutto le tabelle. Quindi utilizzare il comando COPY per caricare le tabelle con dati campione archiviati in un bucket Amazon S3. Per la procedura di creazione tabelle e il caricamento di dati di esempio, consultare [Passaggio 4: caricare i dati da Amazon S3 ad Amazon Redshift](#).

Interroga le tabelle e le viste del sistema

Oltre alle tabelle che crei, il tuo data warehouse contiene una serie di tabelle e viste di sistema. Queste tabelle e viste contengono informazioni sull'installazione e sulle varie query e processi in esecuzione sul sistema. È possibile interrogare queste tabelle e viste di sistema per raccogliere informazioni sul database. Per ulteriori informazioni, consulta il [riferimento alle tabelle e alle viste di sistema](#) nella Amazon Redshift Database Developer Guide. La descrizione di ogni tabella o vista indica se una tabella è visibile a tutti gli utenti o solo ai superutenti. Per eseguire una query su tabelle visibili solo agli utenti con privilegi avanzati, è necessario accedere come utente con privilegi avanzati.

Visualizza un elenco di nomi di tabella

Per visualizzare un elenco di tutte le tabelle in uno schema, è possibile eseguire una query sulla tabella del catalogo di sistema PG_TABLE_DEF. È possibile prima esaminare l'impostazione per search_path.

```
SHOW search_path;
```

Il risultato dovrebbe essere simile al seguente,

```
search_path
```

```
-----
$user, public
```

Nell'esempio seguente lo schema SALES viene aggiunto al percorso di ricerca e sono mostrate tutte le tabelle presenti nello schema SALES.

```
set search_path to '$user', 'public', 'sales';
```

```
SHOW search_path;
```

```
search_path
```

```
-----
"$user", public, sales
```

```
select * from pg_table_def where schemaname = 'sales';
```

schemaname	tablename	column	type	encoding	distkey	sortkey	notnull
sales	demo	personid	integer	az64	f		
		0				f	
sales	demo	city	character varying(255)	lzo	f		
		0				f	

L'esempio seguente mostra un elenco di tutte le tabelle denominate DEMO in tutti gli schemi del database corrente.

```
set search_path to '$user', 'public', 'sales';
```

```
select * from pg_table_def where tablename = 'demo';
```

schemaname	tablename	column	type	encoding	distkey	sortkey	notnull
public	demo	personid	integer	az64	f		
		0				f	
public	demo	city	character varying(255)	lzo	f		
		0				f	
sales	demo	personid	integer	az64	f		
		0				f	

```
sales      | demo      | city      | character varying(255) | lzo      | f      |
0 | f
```

Per ulteriori informazioni, consultare [PG_TABLE_DEF](#).

Puoi anche utilizzare l'editor di query di Amazon Redshift v2 per visualizzare tutte le tabelle in uno schema specificato scegliendo prima un database a cui connetterti.

Visualizzazione degli utenti

È possibile eseguire una query sul catalogo PG_USER per visualizzare un elenco di tutti gli utenti del database, insieme all'ID utente (USESYSID) e ai privilegi degli utenti.

```
SELECT * FROM pg_user;
```

```
username | usesysid | usecreatedb | usesuper | usecatupd | passwd | valuntil |
useconfig
-----+-----+-----+-----+-----+-----+-----
+-----
rdsdb    |         1 | true        | true     | true     | ***** | infinity |
awsuser  |        100 | true        | true     | false    | ***** |          |
guest    |        104 | true        | false    | false    | ***** |          |
```

Il nome utente `rdsdb` viene utilizzato internamente da Amazon Redshift per eseguire attività di amministrazione e manutenzione di routine. È possibile filtrare la query in modo da visualizzare solo nomi utente definiti dall'utente aggiungendo `where usesysid > 1` all'istruzione `SELECT`.

```
SELECT * FROM pg_user WHERE usesysid > 1;
```

```
username | usesysid | usecreatedb | usesuper | usecatupd | passwd | valuntil |
useconfig
-----+-----+-----+-----+-----+-----+-----
+-----
awsuser  |        100 | true        | true     | false    | ***** |          |
guest    |        104 | true        | false    | false    | ***** |          |
```

Visualizza le query recenti

Nell'esempio precedente, l'ID utente (`user_id`) per `adminuser` è 100. Per elencare le quattro query più recenti eseguite da `adminuser`, è possibile interrogare la vista `SYS_QUERY_HISTORY`.

È possibile utilizzare questa visualizzazione per trovare l'ID della query (`query_id`) o l'ID del processo (`session_id`) per una query eseguita di recente. È inoltre possibile utilizzare questa vista per verificare quanto tempo ha richiesto il completamento di una query. `SYS_QUERY_HISTORY` include i primi 4.000 caratteri della stringa di query (`query_text`) per aiutarti a localizzare una query specifica. Utilizza la clausola `LIMIT` con l'istruzione `SELECT` per limitare i risultati.

```
SELECT query_id, session_id, elapsed_time, query_text
FROM sys_query_history
WHERE user_id = 100
ORDER BY start_time desc
LIMIT 4;
```

Il risultato è simile al seguente.

```
query_id | session_id | elapsed_time | query_text
-----+-----+-----
+-----+-----+-----
892      | 21046     | 55868      | SELECT query, pid, elapsed, substring
from ...
620      | 17635     | 1296265    | SELECT query, pid, elapsed, substring
from ...
610      | 17607     | 82555      | SELECT * from DEMO;
596      | 16762     | 226372     | INSERT INTO DEMO VALUES (100);
```

Determina l'ID di sessione di una query in esecuzione

Per recuperare le informazioni della tabella di sistema su una query, potrebbe essere necessario specificare l'ID di sessione (ID del processo) associato a tale query. In alternativa, potrebbe essere necessario trovare l'ID di sessione per una query ancora in esecuzione. Ad esempio, è necessario l'ID di sessione se è necessario annullare una query la cui esecuzione richiede troppo tempo su un cluster predisposto. È possibile interrogare la tabella di sistema `STV_RECENTS` per ottenere un elenco di sessioni IDs per l'esecuzione delle query, insieme alla stringa di query corrispondente. Se la query restituisce più sessioni, è possibile esaminare il testo della query per determinare l'ID di sessione necessario.

Per determinare l'ID di sessione di una query in esecuzione, esegui la seguente istruzione `SELECT`.

```
SELECT session_id, user_id, start_time, query_text
FROM sys_query_history
```

```
WHERE status='running';
```

Annullare una query

Se esegui una query che richiede troppo tempo o consuma risorse eccessive, annulla la query. Ad esempio, creare un elenco di rivenditori di biglietti che include il nome del rivenditore e la quantità di biglietti venduti. La query seguente seleziona i dati dalle tabelle SALES e USERS e unisce le due tabelle associando SELLERID e USERID nella clausola WHERE.

```
SELECT sellerid, firstname, lastname, sum(qtysold)
FROM sales, users
WHERE sales.sellerid = users.userid
GROUP BY sellerid, firstname, lastname
ORDER BY 4 desc;
```

Il risultato è simile al seguente.

sellerid	firstname	lastname	sum
48950	Nayda	Hood	184
19123	Scott	Simmons	164
20029	Drew	Mcguire	164
36791	Emerson	Delacruz	160
13567	Imani	Adams	156
9697	Dorian	Ray	156
41579	Harrison	Durham	156
15591	Phyllis	Clay	152
3008	Lucas	Stanley	148
44956	Rachel	Villarreal	148

Note

Si tratta di una query complessa. Per questo tutorial, non devi preoccuparti della costruzione di questa query.

La query precedente viene eseguita in alcuni secondi e restituisce 2.102 righe.

Si supponga di aver dimenticato di inserire la clausola WHERE.

```
SELECT sellerid, firstname, lastname, sum(qtysold)
FROM sales, users
GROUP BY sellerid, firstname, lastname
ORDER BY 4 desc;
```

Il set di risultati includerà tutte le righe della tabella SALES moltiplicato per tutte le righe della tabella USERS (49.989 x 3.766). Questa operazione è denominata unione cartesiana ed è sconsigliata. Il risultato corrisponde a oltre 188 milioni di righe e l'esecuzione della query richiede molto tempo.

Per annullare una query in esecuzione, utilizzate il comando CANCEL con l'ID di sessione della query. Con l'editor di query di Amazon Redshift v2 puoi annullare una query selezionando il pulsante Annulla mentre la query è in esecuzione.

Per trovare l'ID di sessione, avvia una nuova sessione e interroga la tabella STV_RECENTS, come mostrato nel passaggio precedente. Nell'esempio seguente viene mostrato come aumentare la leggibilità dei risultati. In questo caso, utilizzare la funzione TRIM per rimuovere gli spazi finali e visualizzare solo i primi 20 caratteri della stringa di query.

Per determinare l'ID di sessione di una query in esecuzione, eseguite la seguente istruzione SELECT.

```
SELECT user_id, session_id, start_time, query_text
FROM sys_query_history
WHERE status='running';
```

Il risultato è simile al seguente.

```
user_id | session_id | start_time | query_text
-----+-----+-----+-----
+-----+-----+-----+-----
100     | 1073791534 | 2024-03-19 22:26:21.205739 | SELECT user_id, session_id,
start_time, query_text FROM ...
```

Per annullare la query con l'ID di sessione 1073791534, esegui il comando seguente.

```
CANCEL 1073791534;
```

Note

Il comando CANCEL non interrompe una transazione. Per l'interruzione o il rollback di una transazione, utilizzare il comando ABORT o ROLLBACK. Per annullare una query associata a una transazione, annullare prima la query e quindi interrompere la transazione.

Se la query annullata è associata a una transazione, utilizzare il comando ABORT o ROLLBACK per annullare la transazione ed eliminare le modifiche apportate ai dati:

```
ABORT;
```

Se non hai effettuato l'accesso come utente con privilegi avanzati, puoi annullare solo le tue query. Un utente con privilegi avanzati può annullare tutte le query.

Se lo strumento di query non supporta l'esecuzione simultanea di query, sarà necessario avviare un'altra sessione per annullare la query.

Per ulteriori informazioni sull'annullamento di una query, consulta [CANCEL](#) nella Amazon Redshift Database Developer Guide.

Annulla una query tramite la coda dell'utente con privilegi avanzati

Se la sessione corrente comprende troppe query in esecuzione simultaneamente, potresti non essere in grado di eseguire il comando CANCEL fino al termine di un'altra query. In questo caso, eseguire il comando CANCEL usando una coda di query di gestione dei carichi di lavoro diversa.

Con la gestione dei carichi di lavoro è possibile eseguire query in code di query diverse in modo da non dover attendere il completamento di un'altra query. Il gestore dei carichi di lavoro crea una coda separata, denominata coda dell'utente con privilegi avanzati, che puoi usare per la risoluzione dei problemi. Per usare la coda dell'utente con privilegi avanzati, è necessario aver effettuato l'accesso come utente con privilegi avanzati e aver impostato il gruppo di query su "superuser" tramite il comando SET. Dopo aver eseguito i comandi, reimposta il gruppo di query usando il comando RESET (REIMPOSTA).

Per annullare una query utilizzando la coda dei superutenti, esegui questi comandi.

```
SET query_group TO 'superuser';  
CANCEL 1073791534;
```

```
RESET query_group;
```

Esegui query su dati non presenti nel tuo database Amazon Redshift

Di seguito, puoi trovare informazioni su come iniziare a interrogare dati su fonti remote, inclusi dati Amazon S3, gestori di database remoti, database Amazon Redshift remoti e modelli di apprendimento automatico (ML) per la formazione con Amazon Redshift.

Argomenti

- [Esecuzione di query nel data lake](#)
- [Interrogazione di dati su gestori di database remoti](#)
- [Accesso ai dati in altri database Amazon Redshift](#)
- [Addestramento dei modelli di machine learning con i dati Amazon Redshift](#)

Esecuzione di query nel data lake

È possibile usare Amazon Redshift Spectrum per eseguire query sui dati nei file Amazon S3 senza doverli caricare nelle tabelle Amazon Redshift. Amazon Redshift fornisce funzionalità SQL progettate per l'elaborazione analitica online veloce (OLAP) di dataset di grandi dimensioni archiviati sia nei cluster Amazon Redshift che nei data lake Amazon S3. È possibile interrogare i dati in molti formati, tra cui Parquet, ORC, RCFFile, TextFile, SequenceFile, RegexSerde, OpenCSV e AVRO. È possibile creare schemi e tabelle esterni per definire la struttura dei file in Amazon S3. Quindi, puoi utilizzare un catalogo di dati esterno, ad esempio il tuo AWS Glue metastore Apache Hive. Le modifiche al tipo di catalogo di dati sono immediatamente disponibili per tutti i cluster Amazon Redshift.

Dopo aver registrato i dati in un AWS Glue Data Catalog e abilitato con AWS Lake Formation, puoi interrogarli utilizzando Redshift Spectrum.

Redshift Spectrum si trova su dei server Amazon Redshift dedicati indipendenti dal cluster. Redshift Spectrum completa numerose attività che richiedono un'importante capacità di calcolo, come l'aggregazione e il filtraggio di predicati, sul livello Redshift Spectrum. Redshift Spectrum inoltre si dimensiona in modo intelligente per trarre vantaggio dall'elaborazione massiva parallela.

È possibile partizionare le tabelle esterne in una o più colonne per ottimizzare le prestazioni delle query tramite l'eliminazione delle partizioni. È possibile eseguire query e join sulle tabelle esterne con le tabelle Amazon Redshift. Puoi accedere a tabelle esterne da più cluster Amazon Redshift e

interrogare i dati di Amazon S3 da qualsiasi cluster nella stessa regione. AWS Quando aggiorni i file di dati Amazon S3, i dati possono essere sottoposti a query immediatamente da qualsiasi cluster Amazon Redshift.

Per ulteriori informazioni su Redshift Spectrum, incluse le modalità di utilizzo di Redshift Spectrum e data lake, consultare [Nozioni di base su Amazon Redshift Spectrum](#) nella Guida per gli sviluppatori di database di Amazon Redshift.

Interrogazione di dati su gestori di database remoti

Puoi unire i dati di un database Amazon RDS e un database Amazon Aurora con i dati del tuo database Amazon Redshift utilizzando una query federata. È possibile utilizzare Amazon Redshift per interrogare direttamente i dati operativi (senza spostarli), applicare trasformazioni e inserire i dati nelle tabelle Redshift. Alcuni dei calcoli per le query federate vengono distribuiti alle origini dati remote.

Per eseguire query federate, Amazon Redshift effettua innanzitutto una connessione all'origine dati remota. Amazon Redshift richiama quindi i metadati relativi alle tabelle nell'origine dati remota, invia query e poi recupera le righe dei risultati. Amazon Redshift distribuisce quindi le righe dei risultati ai nodi di calcolo Amazon Redshift per un'ulteriore elaborazione.

Per informazioni sulla configurazione dell'ambiente per query federate, consultare uno dei seguenti argomenti nella Guida per gli sviluppatori di database di Amazon Redshift:

- [Nozioni di base sull'utilizzo di query federate su PostgreSQL](#)
- [Nozioni di base sull'utilizzo di query federate su MySQL](#)

Accesso ai dati in altri database Amazon Redshift

Utilizzando la condivisione dei dati di Amazon Redshift, puoi condividere dati in tempo reale con elevata sicurezza e maggiore facilità tra cluster AWS o account Amazon Redshift a scopo di lettura. È possibile avere accesso istantaneo, granulare e a elevate prestazioni ai dati nei cluster Amazon Redshift senza doverli copiare o spostare manualmente. I tuoi utenti possono visualizzare le informazioni più consistenti up-to-date e coerenti man mano che vengono aggiornate nei cluster Amazon Redshift. Puoi condividere dati a diversi livelli, ad esempio database, schemi, tabelle, viste (incluse viste regolari, con associazione tardiva e materializzata) e funzioni SQL definite dall'utente (). UDFs

La condivisione dei dati di Amazon Redshift è particolarmente utile per questi casi d'uso:

- Centralizzazione dei carichi di lavoro business-critical: utilizzare un cluster centrale di estrazione, trasformazione e caricamento (ETL) che condivide i dati con più cluster di business intelligence o di analisi. Questo approccio fornisce isolamento del carico di lavoro in lettura e chargeback per singoli carichi di lavoro.
- Condivisione dei dati tra ambienti: condividere i dati tra ambienti di sviluppo, test e produzione. È possibile migliorare l'agilità del team condividendo i dati a diversi livelli di granularità.

Per ulteriori informazioni sulla condivisione dei dati, consulta la sezione [Gestione delle attività di condivisione dei dati](#) nella Amazon Redshift Database Developer Guide.

Addestramento dei modelli di machine learning con i dati Amazon Redshift

Grazie ad Amazon Redshift Machine Learning (Amazon Redshift ML), è possibile addestrare un modello fornendo i dati ad Amazon Redshift. Quindi Amazon Redshift ML crea modelli che acquisiscono i pattern nei dati di input. È possibile quindi utilizzare questi modelli per generare previsioni per i nuovi dati di input senza dover sostenere costi aggiuntivi. Grazie ad Amazon Redshift ML, è possibile addestrare modelli di machine learning utilizzando istruzioni SQL e richiamarli nelle query SQL per la previsione. È possibile continuare a migliorare l'accuratezza delle previsioni modificando iterativamente i parametri e migliorando i dati di addestramento.

Amazon Redshift ML rende più semplice agli utenti SQL la creazione, l'addestramento e l'implementazione dei modelli di machine learning utilizzando i familiari comandi SQL. Utilizzando Amazon Redshift ML, puoi utilizzare i tuoi dati nei cluster Amazon Redshift per addestrare modelli con SageMaker Amazon AI Autopilot e ottenere automaticamente il modello migliore. Puoi quindi localizzare i modelli e fare previsioni all'interno di un database Amazon Redshift.

Per ulteriori informazioni su Amazon Redshift ML, consultare [Nozioni di base su Amazon Redshift ML](#) nella Guida per gli sviluppatori di database Amazon Redshift.

Scopri i concetti di Amazon Redshift

Amazon Redshift serverless consente di accedere e analizzare i dati senza le configurazioni di un data warehouse con provisioning. Viene eseguito automaticamente il provisioning delle risorse e la capacità del data warehouse viene dimensionata in modo intelligente per fornire prestazioni rapide per carichi di lavoro maggiormente impegnativi e imprevedibili. Quando il data warehouse è inattivo non vengono addebitati costi, si paga solo l'utilizzo. Puoi caricare i dati e iniziare subito a eseguire query nell'editor di query Amazon Redshift v2 o nello strumento di business intelligence (BI) preferito. Goditi il miglior rapporto prezzo/prestazioni e le familiari funzionalità SQL in un easy-to-use ambiente senza amministrazione.

Se Amazon Redshift viene utilizzato per la prima volta, consigliamo di iniziare leggendo le seguenti sezioni:

- [Panoramica delle funzionalità di Amazon Redshift serverless](#): in questo argomento viene riportata una panoramica di Amazon Redshift serverless e le sue funzionalità principali.
- [Punti salienti del servizio e prezzi](#): in questa pagina del prodotto sono disponibili i dettagli sui punti salienti e sui prezzi del servizio Amazon Redshift serverless.
- [Inizia a usare i data warehouse Serverless Amazon Redshift](#). — In questo argomento, puoi saperne di più su come creare un data warehouse Serverless Amazon Redshift e iniziare a interrogare i dati utilizzando l'editor di query v2.

Se preferisci gestire le risorse Amazon Redshift manualmente, puoi creare cluster con provisioning per le proprie esigenze di query sui dati. Per ulteriori informazioni, consultare [Cluster Amazon Redshift](#).

Se la tua organizzazione è idonea e il tuo cluster viene creato in un Regione AWS paese in cui Amazon Redshift Serverless non è disponibile, potresti essere in grado di creare un cluster nell'ambito del programma di prova gratuita di Amazon Redshift. Scegli Produzione o Versione di prova gratuita per rispondere alla domanda Per cosa si intende utilizzare questo cluster? Se scegli Versione di prova gratuita, crei una configurazione con il tipo di nodo dc2.large. Per ulteriori informazioni sulla scelta di una versione prova gratuita, consulta [Prova gratuita di Amazon Redshift](#). [Per un elenco delle aree Regioni AWS in cui è disponibile Amazon Redshift Serverless, consulta gli endpoint Amazon Redshift elencati per l'API Redshift Serverless nel. Riferimenti generali di Amazon Web Services](#)

Di seguito sono riportati alcuni concetti chiave di Amazon Redshift serverless.

- Spazio dei nomi: una raccolta di oggetti di database e utenti. Gli spazi dei nomi raggruppano tutte le risorse utilizzate in Amazon Redshift serverless, come schemi, tabelle, utenti, unità di condivisione dati e snapshot.
- Gruppo di lavoro: una raccolta di risorse di calcolo. I gruppi di lavoro ospitano le risorse di calcolo utilizzate da Amazon Redshift serverless per eseguire attività di calcolo. Alcuni esempi di tali risorse includono Redshift Processing Units (RPU), gruppi di sicurezza, limiti di utilizzo. I gruppi di lavoro dispongono di impostazioni di rete e sicurezza che puoi configurare utilizzando la console Amazon Redshift Serverless, AWS Command Line Interface Amazon Redshift Serverless o Amazon Redshift Serverless. APIs

Per ulteriori informazioni sulla configurazione dello spazio dei nomi e delle risorse del gruppo di lavoro, consulta [Utilizzo dello spazio dei nomi](#) e [Utilizzo dei gruppi di lavoro](#).

Di seguito sono riportati alcuni concetti chiave sui cluster con provisioning di Amazon Redshift:

- Cluster – Il componente centrale dell'infrastruttura di un data warehouse di Amazon Redshift è un cluster.

Un cluster è costituito da uno o più nodi di calcolo. I nodi di calcolo eseguono il codice compilato.

Se viene effettuato il provisioning di un cluster con due o più nodi di calcolo, un ulteriore nodo principale coordina i nodi di calcolo. Il nodo leader gestisce la comunicazione esterna con applicazioni, come strumenti di business intelligence e editor di query. L'applicazione client interagisce direttamente solo con il nodo principale. I nodi di calcolo sono trasparenti alle applicazioni esterne.

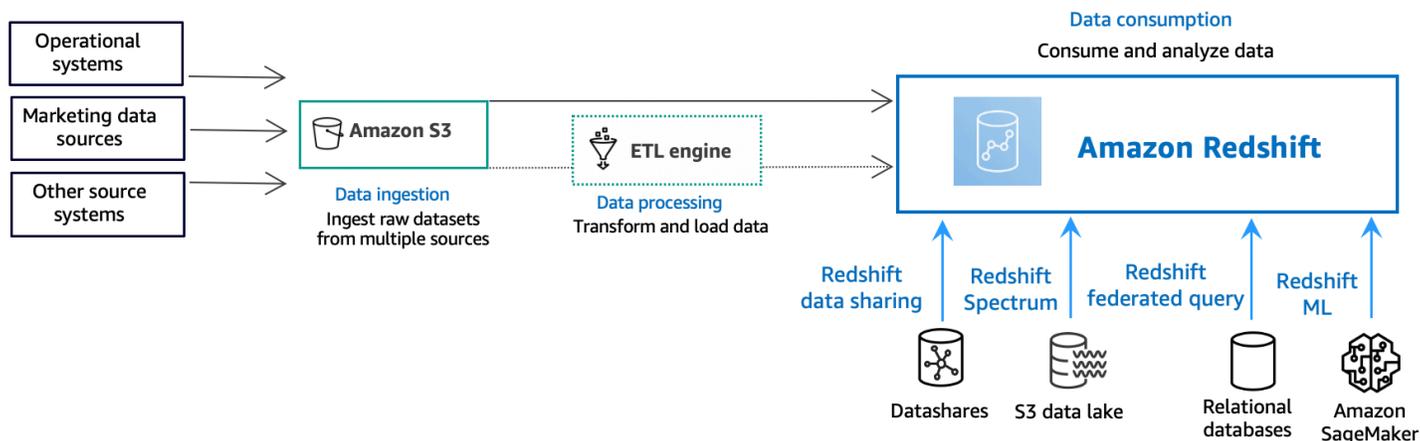
- Database – Un cluster contiene uno o più database.

I dati utente vengono archiviati in uno o più database nei nodi di calcolo. Il client SQL comunica con il nodo principale, che a sua volta coordina l'esecuzione di query con i nodi di calcolo. Per dettagli sui nodi di calcolo e sui nodi principali, consulta [Architettura del sistema di data warehouse](#). All'interno di un database, i dati utente sono organizzati in uno o più schemi.

Amazon Redshift è un sistema di gestione di database relazionali (RDBMS, Relational Database Management System) ed è compatibile con altre applicazioni RDBMS. Offre le stesse funzionalità delle tipiche applicazioni RDBMS, tra cui funzioni di elaborazione di transazioni online (OLTP) come l'inserimento e l'eliminazione di dati. Amazon Redshift è inoltre ottimizzato per l'analisi batch ad alte prestazioni e la creazione di report dei set di dati.

Di seguito, puoi trovare una descrizione del tipico flusso di elaborazione dei dati in Amazon Redshift, insieme alle descrizioni delle diverse parti del flusso. Per ulteriori informazioni sull'architettura di sistema Amazon Redshift, consulta [Architettura del sistema di data warehouse](#).

Il diagramma seguente illustra un tipico flusso di elaborazione dati in Amazon Redshift.



Un data warehouse di Amazon Redshift è un sistema di query e gestione di database relazionali di livello aziendale. Amazon Redshift supporta connessioni client con molti tipi di applicazioni, tra cui strumenti di business intelligence, creazione di report, gestione dei dati e analisi. Attraverso l'esecuzione di query di analisi, puoi recuperare, confrontare e valutare grandi quantità di dati in operazioni in più fasi per produrre un risultato finale.

Presso il livello acquisizione dati, diversi tipi di origini dati caricano continuamente dati strutturati, semistrutturati o non strutturati sul livello di archiviazione dei dati. Questa area di archiviazione dei dati funge da area di gestione temporanea che memorizza i dati in diversi stati di preparazione al consumo. Un esempio di storage potrebbe essere un bucket Amazon Simple Storage Service (Amazon S3).

Al facoltativo livello elaborazione dati, i dati di origine passano attraverso la preelaborazione, la convalida e la trasformazione utilizzando pipeline di estrazione, trasformazione, caricamento (ETL) o di estrazione, caricamento, trasformazione (ELT). Questi set di dati grezzi vengono quindi perfezionati utilizzando le operazioni ETL. Un esempio di motore ETL è AWS Glue.

Presso il livello consumo di dati, i dati vengono caricati nel cluster Amazon Redshift, dove è possibile eseguire carichi di lavoro analitici.

Per alcuni esempi di carichi di lavoro analitici, consulta [Esecuzione di query all'esterno delle origini dati](#).

Risorse aggiuntive per informazioni su Amazon Redshift

Per ulteriori informazioni su Amazon Redshift serverless, consigliamo di continuare con ulteriori informazioni sui concetti introdotti in questa guida utilizzando le seguenti risorse di Amazon Redshift:

- Video in primo piano: questi video ti aiutano a conoscere le funzionalità di Amazon Redshift.
 - Per comprendere Amazon Redshift serverless a un livello elevato, guardare il video seguente. [Amazon Redshift Serverless Explained in 90 Seconds](#) (Amazon Redshift Serverless spiegato in 90 secondi).
 - Per informazioni su come configurare un data warehouse serverless e iniziare a utilizzare i dati, guardare il video seguente. [Getting Started with Amazon Redshift Serverless](#) (Nozioni di base su Amazon Redshift Serverless).
- [Guida alla gestione di Amazon Redshift](#): questa guida si basa sulle Nozioni di base su Amazon Redshift. Offre informazioni dettagliate su concetti e attività relativi a creazione, gestione e monitoraggio di cluster con provisioning di Amazon Redshift Serverless e di Amazon Redshift.
- [Guida per gli sviluppatori di database di Amazon Redshift](#): questa guida si basa sulla Guida alle operazioni di base di Amazon Redshift. Offre informazioni dettagliate per sviluppatori di database relativamente a progettazione, creazione, esecuzione di query e manutenzione dei database presenti all'interno del data warehouse.
 - [Documentazione di riferimento a SQL](#): in questo argomento vengono descritti i comandi SQL e i riferimenti alle funzioni per Amazon Redshift.
 - [Riferimento per tabelle e viste di sistema](#): questo argomento descrive le tabelle e le viste di sistema per Amazon Redshift.
- Tutorial per Amazon Redshift: in questo argomento sono riportati i tutorial sulle funzionalità di Amazon Redshift.
 - [Caricamento di dati da Amazon S3](#): questo tutorial descrive come caricare i dati nelle tabelle di database Amazon Redshift dai file di dati in un bucket Amazon S3.
 - [Guida introduttiva alla condivisione dei dati](#): questa sezione descrive come condividere e accedere ai dati in altri cluster Amazon Redshift.
 - [Utilizzo delle funzioni SQL spaziali con Amazon Redshift](#): questo tutorial mostra come utilizzare alcune delle funzioni SQL spaziali con Amazon Redshift.
 - [Esecuzione di query su dati nidificati con Amazon Redshift Spectrum](#): questo tutorial descrive come utilizzare Redshift Spectrum per eseguire query su dati nidificati nei formati file Parquet, ORC, JSON e Ion utilizzando tabelle esterne.

- [Configurazione di code di gestione del carico di lavoro \(WLM\)](#): in questo tutorial viene descritto come configurare la gestione manuale dei carichi di lavoro (WLM) in Amazon Redshift.
- [Nozioni di base su Amazon Redshift ML](#): questa sezione descrive come gli utenti possono creare, addestrare e implementare modelli di machine learning utilizzando comandi SQL familiari.
- [Novità](#): questa pagina Web elenca le nuove funzionalità di Amazon Redshift e gli aggiornamenti dei prodotti.

Cronologia dei documenti

Note

Per una descrizione delle nuove funzionalità di Amazon Redshift, consulta [What's new](#).

La tabella seguente descrive le importanti modifiche alla documentazione della Amazon Redshift Getting Started Guide.

Modifica	Descrizione	Data di rilascio
Aggiornamento della documentazione	La guida è stata aggiornata in modo da includere nuove sezioni su come iniziare con le attività comuni del database, eseguire query sul data lake, eseguire query sui dati su origini remote, condividere dati e addestrare modelli di machine learning con i dati Amazon Redshift.	30 giugno 2021
Nuova caratteristica	La guida è stata aggiornata per descrivere la nuova procedura di caricamento del campione.	4 giugno 2021
Aggiornamento della documentazione	La guida è stata aggiornata per rimuovere la console Amazon Redshift originale e migliorare il flusso delle fasi.	14 agosto 2020
Nuova console	La guida è stata aggiornata per descrivere la nuova console di Amazon Redshift.	11 novembre 2019
Nuova caratteristica	Guida aggiornata per la descrizione della procedura di avvio rapido del cluster.	10 agosto 2018
Nuova caratteristica	La guida è stata aggiornata per l'avvio dei cluster dal pannello di controllo di Amazon Redshift.	28 luglio 2015
Nuova caratteristica	Guida aggiornata per l'utilizzo dei nuovi nomi dei tipi di nodo.	9 giugno 2015

Modifica	Descrizione	Data di rilascio
Aggiornamento della documentazione	Screenshot e procedura aggiornate per la configurazione dei gruppi di sicurezza VPC.	30 Aprile 2015
Aggiornamento della documentazione	Screenshot e procedure aggiornate in conformità alla console corrente.	12 novembre 2014
Aggiornamento della documentazione	Il caricamento dei dati è stato spostato dalle informazioni su Amazon S3 in una sezione specifica, mentre la sezione sui passaggi successivi è stata spostata nella fase finale per migliorare la reperibilità.	13 maggio 2014
Aggiornamento della documentazione	È stata eliminata la Pagina di benvenuto e il contenuto è stato incorporato nella pagina principale delle Nozioni di base.	14 marzo 2014
Aggiornamento della documentazione	Questa è una nuova versione della Guida alle operazioni di base di Amazon Redshift che tiene conto dei suggerimenti dei clienti e degli aggiornamenti del servizio.	14 marzo 2014
Nuova guida	Questa è la prima versione della Guida alle operazioni di base di Amazon Redshift.	14 febbraio 2013

Le traduzioni sono generate tramite traduzione automatica. In caso di conflitto tra il contenuto di una traduzione e la versione originale in Inglese, quest'ultima prevarrà.