



Guida alle operazioni di base

# Amazon Redshift



# Amazon Redshift: Guida alle operazioni di base

Copyright © 2024 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

Amazon Redshift 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 .....	8
Caricamento di dati da Amazon S3 .....	9
Esecuzione di query all'esterno delle origini dati .....	18
Esecuzione di query nei data lake .....	18
Esecuzione di query su origini dati remote .....	19
Accesso ai dati in altri cluster .....	19
Addestramento dei modelli di ML con i dati Redshift .....	20
Cluster con provisioning di Amazon Redshift .....	21
Console dei cluster con provisioning di Amazon Redshift .....	21
Iscrizione a AWS .....	21
Definizione delle regole del firewall .....	22
Connessione ad Amazon Redshift .....	24
Cluster Amazon Redshift e caricamento dei dati .....	24
Utilizzo di un set di dati di esempio .....	25
Utilizzo dei propri dati in Amazon Redshift .....	30
Attività di database comuni .....	44
Attività 1: Creazione di un database .....	45
Attività 2: Creazione di un utente .....	45
Attività 3: Creazione di uno schema .....	46
Attività 4: Creazione di una tabella .....	48
Attività 5: Caricamento di dati di esempio .....	50
Fase 6: Esecuzione di query sulle tabelle di sistema .....	50
Attività 7: Annullamento di una query .....	54
Attività 8: Eliminazione delle risorse .....	57
Panoramica concettuale di Amazon Redshift .....	58
Risorse aggiuntive .....	61
Cronologia dei documenti .....	63
.....	lxv

# Amazon Redshift Serverless

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.

- [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 pagina all'indirizzo <https://portal.aws.amazon.com/billing/signup>.
2. Segui le istruzioni online.

Quando crei un AWS account, viene creato un AWS account utente root. 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).

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/.](https://console.aws.amazon.com/redshiftv2/)

Scegli Prova Amazon Redshift Serverless.

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

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

# Get started with Amazon Redshift Serverless

To start using Amazon Redshift Serverless, set up your serverless data warehouse and create a database.

You will receive \$0 credit towards your Redshift Serverless usage in this account.

## Configuration [Info](#)

### Use default settings

Default settings have been defined to help you get started. You can change them at any time later.

### Customize settings

Customize your settings for your specific needs.

## ▼ How it works



### Using the default settings

Amazon Redshift Serverless creates a default namespace and workgroup. This configuration uses the default settings and becomes active when you associate the default workgroup to the default namespace.



### Customizing the settings

Amazon Redshift Serverless creates a default namespace and workgroup. This configuration becomes active when you associate the default workgroup to the default namespace.

## Namespace [Info](#)

Namespace is a collection of database objects and users. Data properties include database name and password, permissions, and encryption and security.

**⚠** Your data is encrypted by default with an AWS owned key. To choose a different key, choose **Customize settings**.

Namespace name

default

### Database name and password

Database name

dev

Admin user credentials

Created based on IAM credentials

### Permissions

Creazione di un data warehouse con Amazon Redshift serverless

Default IAM role

<arn:aws:iam::694237154968:role/service-role/AmazonRedshift-CommandsAccessRole-20220809T211900> [↗](#)

- 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

## 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.

- 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.

Amazon Redshift Serverless

**Serverless dashboard** [Info](#)

[Query data](#) [Create workgroup](#)



2. Se avvii Query Editor v2 per la prima volta, devi configurare la AWS KMS crittografia prima di procedere. Facoltativamente, è anche possibile specificare l'URI in un bucket S3 per il caricamento dei dati in un momento successivo. Dopo questa operazione, scegli Configura account.

## Configure account

To start using Redshift query editor v2, configure the account.


### AWS KMS encryption

Your data is encrypted by default with a key that AWS owns and manages for you. To choose a different key, customize your encryption settings.

 You can't change the AWS KMS encryption settings after configuring your account. 

Customize encryption settings (advanced)

### S3 bucket - optional

Enter an Amazon S3 path for this account. This path can be used for features such as load data from a local file store. Account users and the S3 bucket must be configured with the proper permissions to secure user's data. [Learn more](#) 

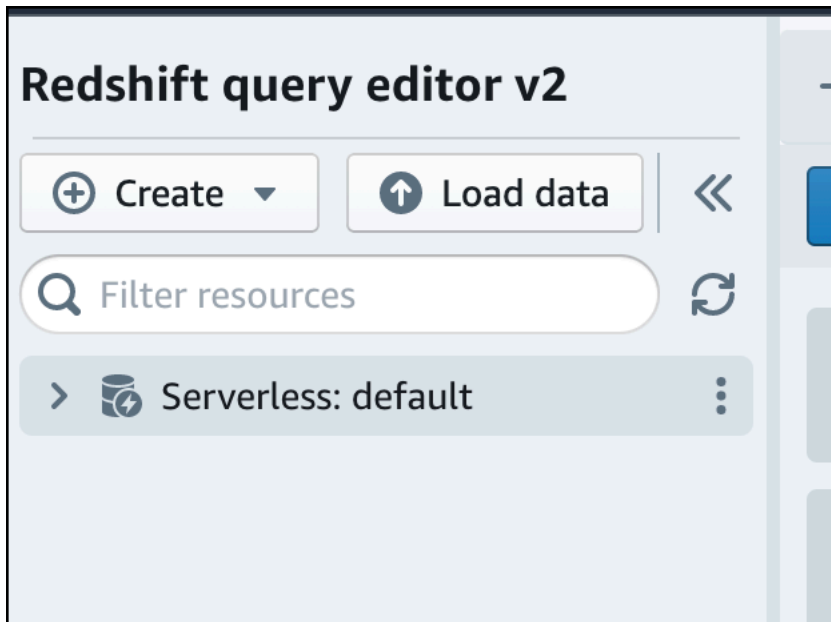
S3 URI

[Cancel](#) [Configure account](#)

Per ulteriori informazioni sulla configurazione dell'editor di query v2, incluse le autorizzazioni necessarie, consulta [Configurazione dell' Account AWS](#) nella Guida alla gestione di Amazon Redshift.

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.

### Connect to default ✕

Authentication [Learn more](#)

Federated user  
The query editor v2 generates a temporary password to connect to the database.

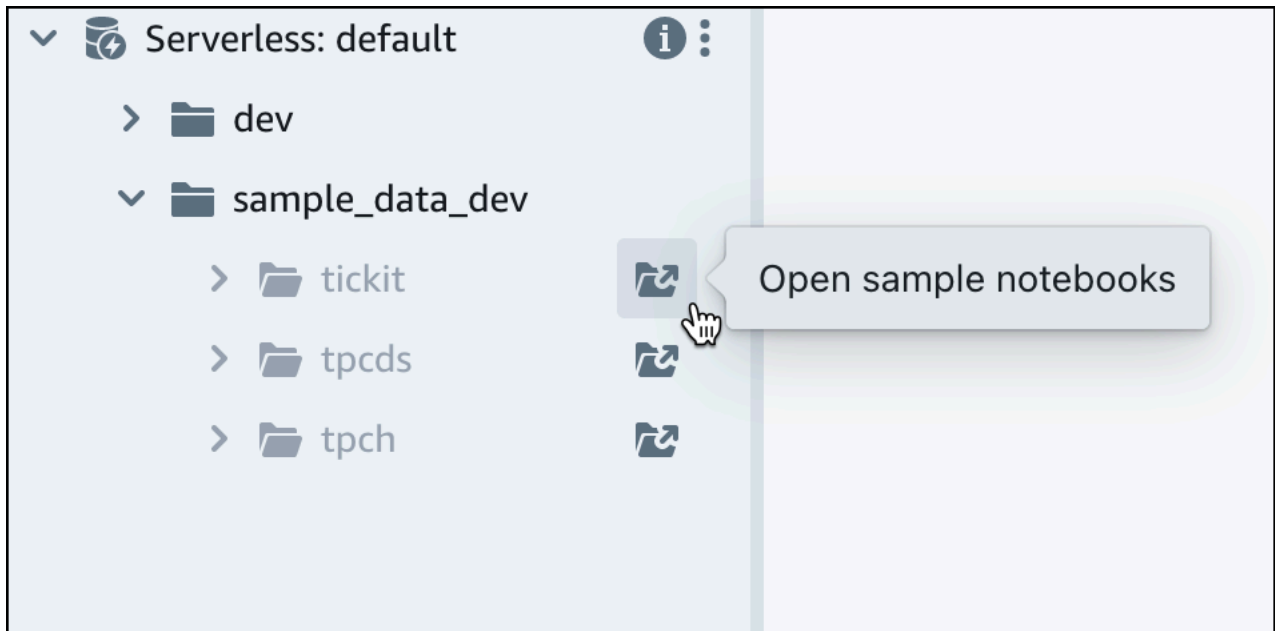
Database user name and password  
Provide a database user and password for the database that you are connecting to. The query editor v2 stores your credentials in AWS Secrets Manager on your behalf.

Database

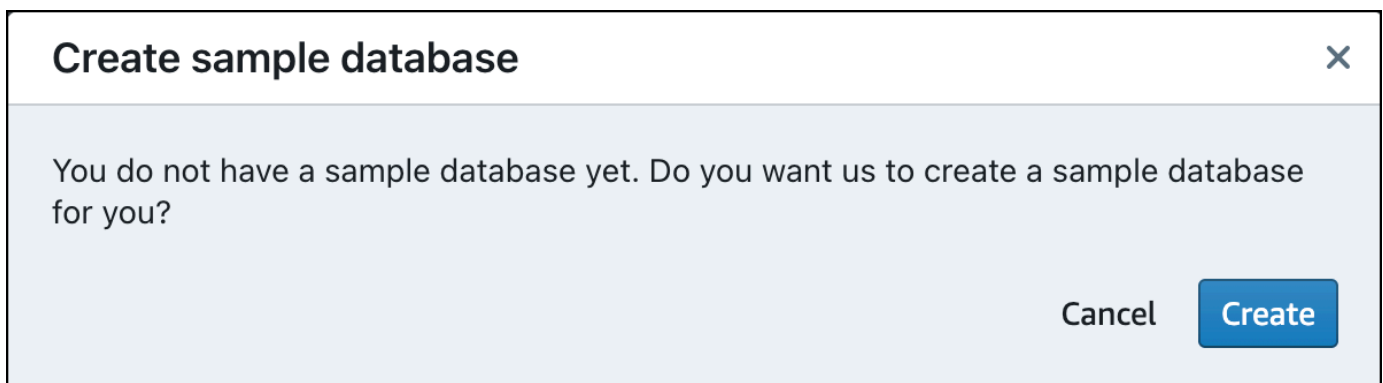
The database name must be 1-64 characters. Valid characters are lowercase alphanumeric characters.

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

5. 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.



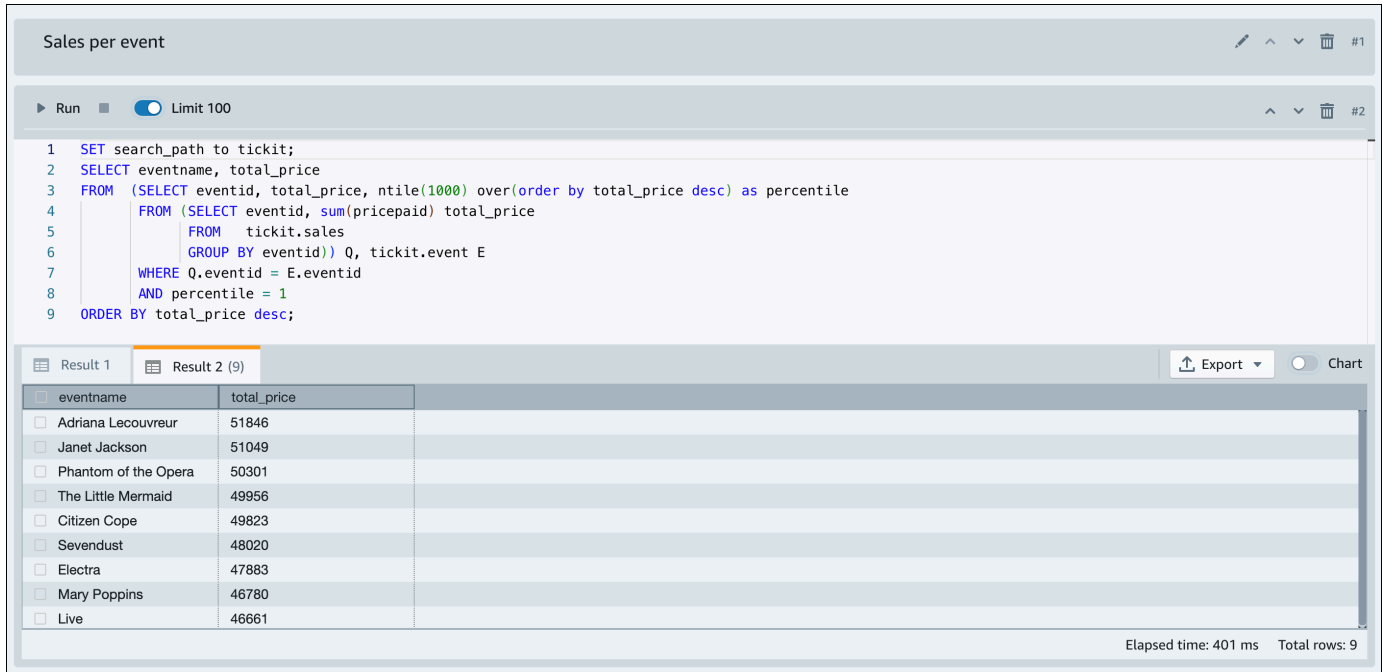
6. Quando si caricano i dati per la prima volta, l'editor di query v2 richiederà di creare un database di esempio. Scegli Crea.



## 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 displays the Amazon Redshift console interface. At the top, the title "Sales per event" is visible. Below the title, there are controls for running the query, including a "Run" button and a "Limit 100" toggle. The main area contains a SQL query:

```
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, the results are displayed in a table with two columns: "eventname" and "total\_price". The table shows 9 rows of 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 of the console, the execution details are shown: "Elapsed time: 401 ms" and "Total rows: 9".

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

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. Per farlo, scegli Configurazione dello spazio dei nomi dal menu di navigazione, quindi scegli Sicurezza e crittografia. Scegli Gestisci ruoli IAM.

**default** [Info](#)

### 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](#) | [Data backup](#) | **Security and encryption** | [Datashares](#) | [Tags](#)

### 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 ▲

Q Search for associated

Associate IAM roles

Create IAM role

Remove IAM roles

or role type

< 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. Seleziona Salvataggio delle modifiche. È ora possibile caricare i dati di esempio da Amazon S3.

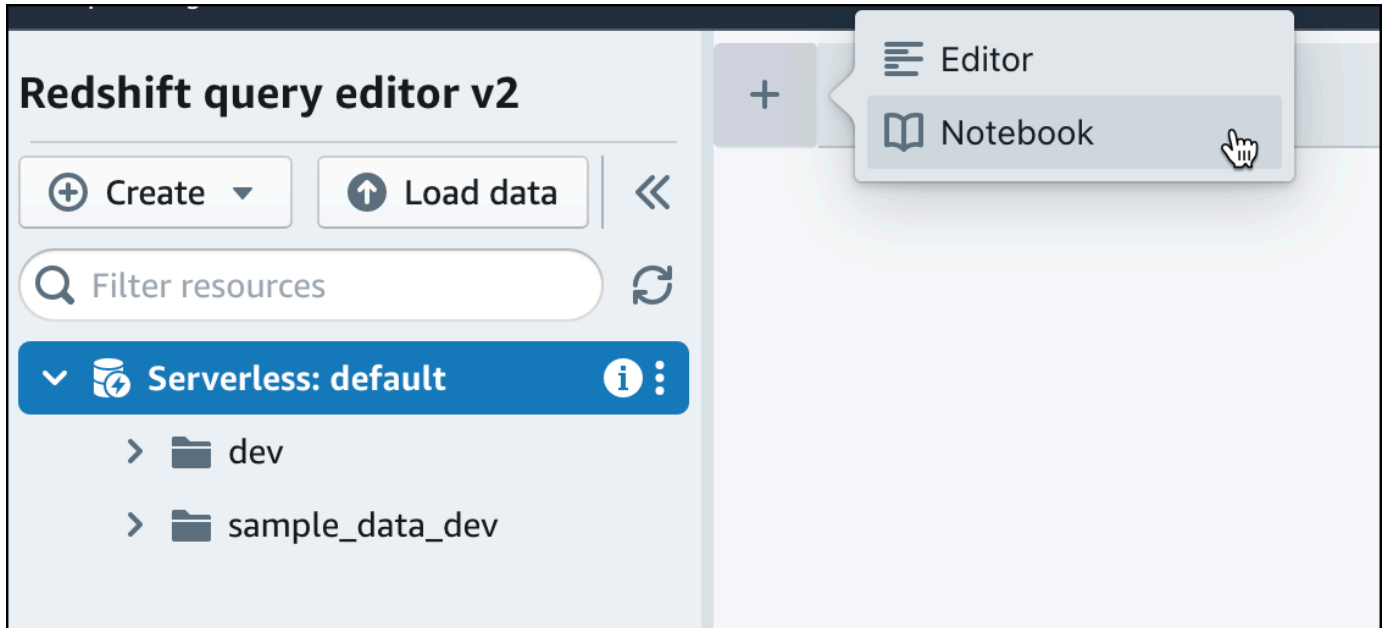
Le fasi seguenti utilizzano i dati all'interno di un bucket S3 pubblico di Amazon Redshift, ma è possibile replicare le stesse fasi utilizzando il proprio bucket S3 e i propri comandi SQL.

## Caricamento di dati di esempio da Amazon S3

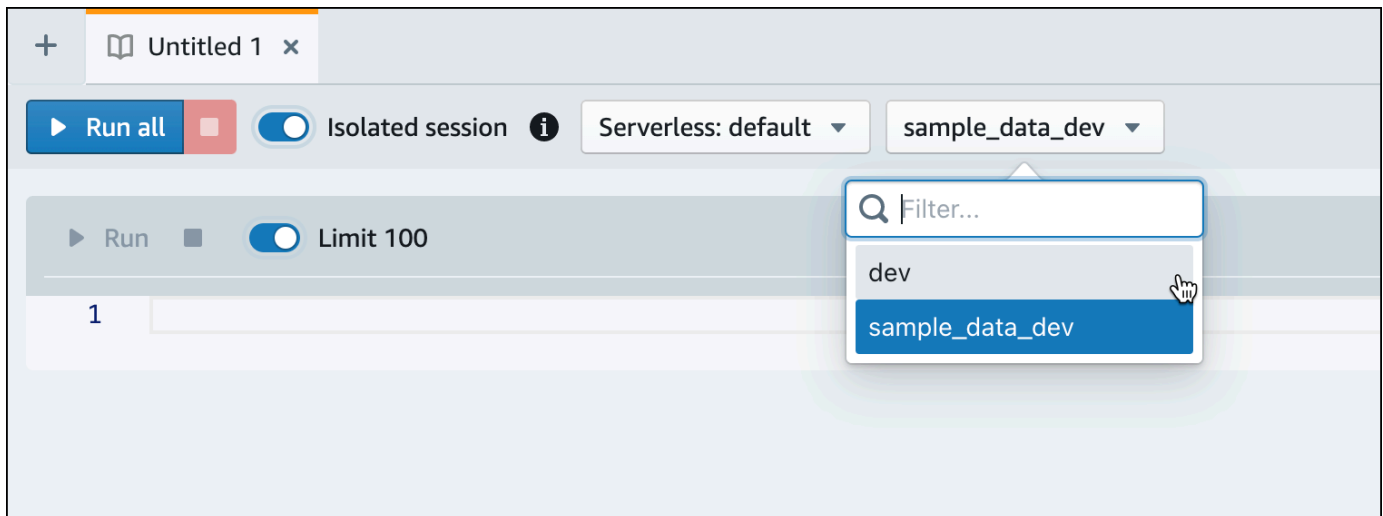
1. Nell'editor di query v2, scegli



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.

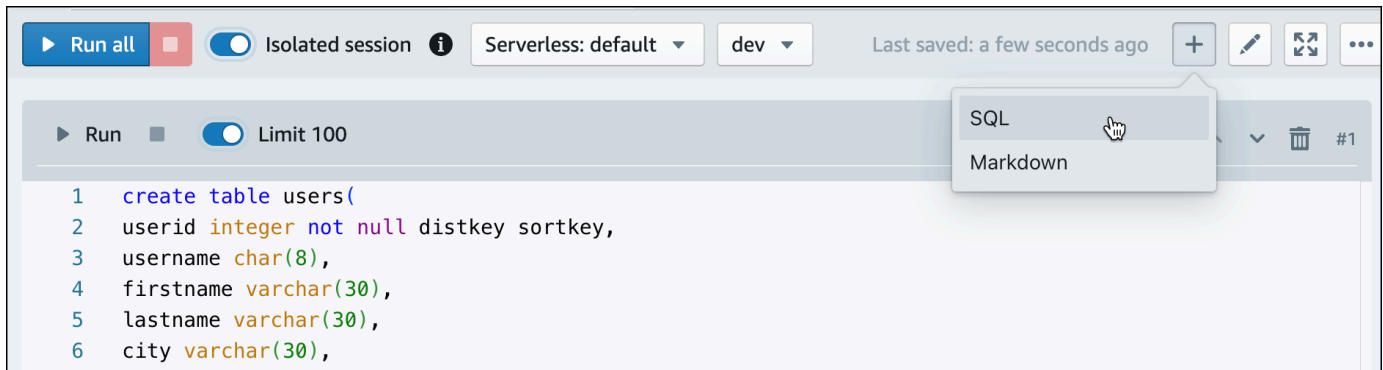


```
create table users(  
userid integer not null distkey sortkey,  
username char(8),  
firstname varchar(30),  
lastname varchar(30),  
city varchar(30),  
state char(2),  
email varchar(100),  
phone char(14),  
likesports boolean,  
liketheatre boolean,  
likeconcerts boolean,  
likejazz boolean,  
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/ticket/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/ticket/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/ticket/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](#).

# Esecuzione di query su origini dati all'esterno del database Amazon Redshift

Di seguito, sono riportate informazioni su come iniziare a eseguire query su dati di origini remote, inclusi i cluster Amazon Redshift remoti. Puoi anche trovare informazioni sui modelli di Machine Learning (ML) di formazione utilizzando Amazon Redshift.

## Argomenti

- [Esecuzione di query nel data lake](#)
- [Esecuzione di query sui dati in origini dati remote](#)
- [Accesso ai dati in altri cluster 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 eseguire query sui dati in molti formati, tra cui Parquet, ORC, RCFile, TextFile, SequenceFile, RegexSerde, OpenCSV e AVRO. È possibile creare schemi e tabelle esterni per definire la struttura dei file in Amazon S3. Quindi, si utilizza un catalogo di dati esterno, ad esempio AWS Glue o il proprio metastore Apache Hive. Le modifiche al tipo di catalogo di dati sono immediatamente disponibili per tutti i cluster Amazon Redshift.

Dopo che i dati sono stati registrati con un catalogo di dati di AWS Glue e dopo averli abilitati con AWS Lake Formation, sarà possibile interrogarli tramite 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. È possibile aggiungere le tabelle esterne da più cluster Amazon Redshift ed eseguire una query sui 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.

## Esecuzione di query sui dati in origini dati remote

È possibile unire i dati da un database Amazon RDS, un database Amazon Aurora o Amazon S3 con i dati nel 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 cluster Amazon Redshift

Utilizzando la condivisione dei dati di Amazon Redshift, puoi condividere i dati in tempo reale in modo sicuro e semplice tra i cluster Amazon Redshift o tra account AWS per scopi di lettura. È possibile avere accesso istantaneo, granulare e a elevate prestazioni ai dati nei cluster Amazon Redshift senza doverli copiare o spostare manualmente. Gli utenti possono visualizzare le informazioni più recenti e coerenti man mano che vengono aggiornate nei cluster Amazon Redshift. È possibile condividere i dati a livelli differenti, come database, schemi, tabelle, viste (incluse viste regolari, tardive e materializzate) e funzioni definite dall'utente (FDU) SQL.

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, consultare [Nozioni di base sulla condivisione dei dati](#) nella Guida per gli sviluppatori di database di Amazon Redshift.

## 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. Con Amazon Redshift ML, è possibile utilizzare i propri dati nei cluster Amazon Redshift per addestrare i modelli con il pilota automatico Amazon SageMaker e ottenere automaticamente il modello migliore. È quindi possibile localizzare i modelli e quindi 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.

# Cluster con provisioning di Amazon Redshift

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

## Argomenti

- [Console dei cluster con provisioning di Amazon Redshift](#)
- [Connessione ai cluster con provisioning di Amazon Redshift](#)
- [Cluster Amazon Redshift e caricamento dei dati](#)
- [Attività di database comuni](#)

## Console dei cluster con provisioning di Amazon Redshift

Prima di iniziare la fase di configurazione di un cluster Amazon Redshift, è necessario soddisfare i seguenti prerequisiti:

- [Iscrizione a AWS](#)
- [Definizione delle regole del firewall](#)

## Iscrizione a AWS

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

1. Apri la pagina all'indirizzo <https://portal.aws.amazon.com/billing/signup>.
2. Segui le istruzioni online.

Nel corso della procedura di registrazione riceverai una telefonata, durante la quale sarà necessario inserire un codice di verifica attraverso la tastiera del telefono.

Quando ti iscrivi a un Account AWS, ne viene creato un Utente root dell'account AWS. 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 amministrativo](#) e utilizza solo l'utente root per eseguire [attività che richiedono l'accesso di un utente root](#).



## Definizione delle regole del firewall

Nell'ambito di questo tutorial, è necessario specificare una porta all'avvio del cluster Amazon Redshift. 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.

Questo prerequisito si applica solo quando in Amazon Redshift si utilizzano i propri dati. Per ulteriori informazioni, consulta [Utilizzo dei propri dati in Amazon Redshift](#).

Dopo aver effettuato l'accesso alla console Amazon Redshift, puoi creare e gestire tutti gli oggetti Amazon Redshift, inclusi cluster, database e nodi. Puoi anche visualizzare le query, eseguire le query ed eseguire altre operazioni DDL (Data Definition Language) e DML (Data Manipulation Language).

Se sei un nuovo utente di Amazon Redshift, ti consigliamo di iniziare andando alle pagine Pannello di controllo, Cluster, e editor della query v2 per iniziare a usare la console.

Per informazioni su come iniziare a utilizzare la console Amazon Redshift, guardare il video [Nozioni di base su Amazon Redshift](#).

Di seguito, puoi trovare le descrizioni degli elementi del pannello di navigazione della console Amazon Redshift:

- Amazon Redshift serverless: accede e analizza i dati senza la necessità di configurare, regolare e gestire i cluster con provisioning di Amazon Redshift.
- Provisioned clusters dashboard (Pannello di controllo dei cluster con provisioning): controlla Cluster metrics (Parametri dei cluster) e Query overview (Panoramica delle query) per informazioni dettagliate sui dati dei parametri (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 un elenco di cluster presenti nel tuo AWS account, scegli un cluster per avviare le query o esegui azioni relative al cluster. È inoltre possibile creare un nuovo cluster da questa pagina.
- **Query editor (Editor di query):** esegue query sui database ospitati sul cluster di Amazon Redshift, salva le query per il riutilizzo o pianifica l'esecuzione in un momento successivo (solo nell'editor di query).
- **Query editor v2 (Editor di query v2):** utilizza l'editor di query v2 che è un'applicazione client SQL separata basata su Web che si utilizza per creare ed eseguire query sul data warehouse di 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 AWS account o a specifici namespace del cluster in un account. Un amministratore può anche rifiutare una unità di condivisione dati.
- **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.
- **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.

## Connessione ai cluster con provisioning di Amazon Redshift

Per connettersi ai cluster Amazon Redshift, dalla pagina Cluster, espandi Connect ai cluster Amazon Redshift ed effettua una delle seguenti operazioni:

- Utilizzare l'editor di query v2 per eseguire query sui database ospitati dal cluster Amazon Redshift. Dopo aver creato il cluster, puoi eseguire immediatamente le query con l'editor della 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 utilizzare i dati nel tuo cluster, sono richiesti i driver JDBC o ODBC al fine di stabilire la connettività dal computer client o dall'istanza. 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 lo strumento client SQL richiede un driver, puoi scaricare un driver specifico del sistema operativo per connetterti 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](#).

## Cluster Amazon Redshift e caricamento dei dati

In questa sezione sono riportati due tutorial che illustrano il processo di creazione di un cluster Amazon Redshift di esempio. In uno, si utilizza un set di dati di esempio e nell'altro si utilizza il proprio set di dati.

### Argomenti

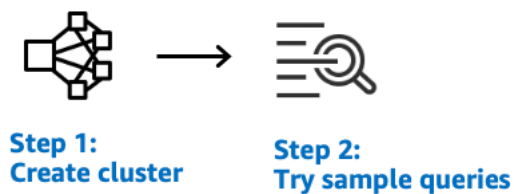
- [Utilizzo di un set di dati di esempio](#)
- [Utilizzo dei propri dati in Amazon Redshift](#)

## Utilizzo di un set di dati di esempio

In questo tutorial verrà illustrato il processo per creare un cluster Amazon Redshift utilizzando un set di dati di esempio. Amazon Redshift carica automaticamente il set di dati di esempio durante la creazione di un nuovo cluster. È possibile eseguire immediatamente una query sui dati dopo la creazione del cluster.

Prima di iniziare la fase di configurazione di un cluster Amazon Redshift, assicurarsi di completare [Iscrizione a AWS](#) e [Definizione delle regole del firewall](#).

In questo tutorial, seguirai le fasi illustrate di seguito.



### Argomenti

- [Fase 1: Creazione di un cluster Amazon Redshift di esempio](#)
- [Fase 2: provare query di esempio utilizzando l'editor di query](#)

#### **⚠ Important**

Il cluster di esempio che crei viene eseguito in un ambiente reale. La tariffa oraria on-demand è pari a 0,25 \$ e consente di utilizzare il cluster campione creato in questo tutorial finché non viene eliminato. Per ulteriori informazioni sui prezzi, consultare [Prezzi di Amazon Redshift](#). Nel caso di domande o problemi, è possibile comunicare con il team di Amazon Redshift inserendo un post nel nostro [forum di discussione](#).

Il presente tutorial non è stato pensato per gli ambienti di produzione, né affronta le diverse opzioni in maniera dettagliata. Dopo aver completato le fasi descritte in questo tutorial, sarà possibile utilizzare la sezione [Risorse aggiuntive](#) per ottenere informazioni più approfondite. Queste informazioni possono aiutare a pianificare, distribuire e gestire i cluster e lavorare con i dati contenuti nel data warehouse.

## Fase 1: Creazione di un cluster Amazon Redshift di esempio

Dopo aver completato i prerequisiti, è possibile iniziare a creare il cluster Amazon Redshift in base a un set di dati di esempio.

Come creare un cluster Amazon Redshift basato su un set di dati di esempio

1. [Accedi AWS Management Console e apri la console Amazon Redshift all'indirizzo https://console.aws.amazon.com/redshiftv2/](https://console.aws.amazon.com/redshiftv2/).
2. Per creare un cluster, effettua una delle operazioni indicate di seguito:
  - Nella pagina del servizio Amazon Redshift, scegliere Crea cluster. Appare la pagina Create cluster (Crea cluster).
  - Su <https://console.aws.amazon.com/redshiftv2/>, scegli la dashboard Provisioned Clusters, quindi scegli Crea cluster.
  - Nel <https://console.aws.amazon.com/redshiftv2/>, scegli Cluster, quindi scegli Crea cluster.
3. Nella sezione Configurazione del cluster, specificare un identificatore del cluster. 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).

Per questo tutorial, inserire **examplecluster**.

4. 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 elencati per l'API [Redshift](#) Serverless nel. Riferimenti generali di Amazon Web Services

Dopo aver scelto il tipo di nodo, effettuare una delle operazioni seguenti:

- In Dati campione, scegliere Carica dati di esempio per caricare il set di dati di esempio nel cluster Amazon Redshift. Amazon Redshift carica il set di dati di esempio Tickit nel database dev e nello schema public predefiniti. È possibile iniziare a utilizzare l'editor di query v2 per eseguire query sui dati.

- Per utilizzare i propri dati nel cluster Amazon Redshift, scegliere Produzione. Quindi in Dati di esempio, scegliere Carica dati di esempio. Per informazioni sull'uso dei propri dati, consultare [Utilizzo dei propri dati in Amazon Redshift](#).

Amazon Redshift carica automaticamente il set di dati di esempio nel cluster Amazon Redshift di esempio.

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 e AWS Secrets Manager utilizza per generare e gestire la password di amministratore. L'utilizzo di 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: specificare un valore per la password.
6. Scegli Create cluster (Crea cluster).

Questo tutorial utilizza l'editor di query Amazon Redshift v2. È possibile utilizzare questo editor per eseguire query sui dati immediatamente, dopo che Amazon Redshift ha terminato la creazione del cluster.

È inoltre possibile scegliere altri strumenti del client SQL che supportano i driver JDBC o ODBC per utilizzare i dati nel cluster. Per ulteriori informazioni, consulta [Connessione a un cluster Amazon Redshift tramite gli strumenti del client SQL](#) nella Guida alla gestione di Amazon Redshift.

## Fase 2: provare query di esempio utilizzando l'editor di query

Quando Amazon Redshift crea il cluster Amazon Redshift, carica automaticamente il set di dati di esempio Tickit. Il completamento della creazione del cluster potrebbe richiedere alcuni minuti. Al termine della creazione, lo stato del cluster diventa ATTIVO. È possibile visualizzare le tabelle Tickit campione dal set di dati campione.

## Utilizzo dell'editor della query

È possibile visualizzare le tabelle Tickit di esempio nell'editor di query v2 scegliendo il cluster, il database dev e lo schema public.

Dopo aver creato il cluster Amazon Redshift, in Connessione ai cluster Amazon Redshift, scegliere Query sui dati.

Dall'editor di query v2, connessi a un database e scegli il nome del cluster nel pannello a vista ad albero. Se richiesto, immettere i parametri di connessione.

Quando ci si connette a un cluster e ai relativi database, viene fornito un Database nome e un Nome utente. È inoltre possibile fornire i parametri necessari per uno dei seguenti metodi di autenticazione:

### Nome utente e password del database

Con questo metodo, fornire anche una Password per il database a cui ti stai connettendo.

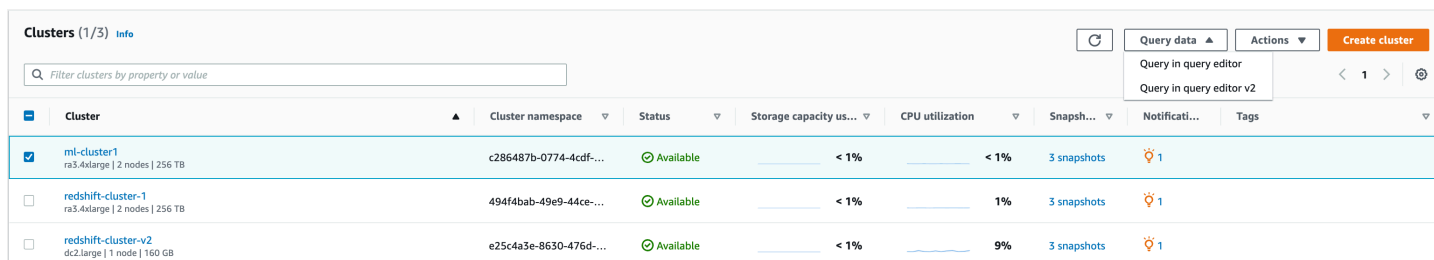
### Credenziali temporanee

Con questo metodo, l'editor di query v2, genera una password temporanea per connettersi al database.

Quando si seleziona un cluster con editor di query v2, a seconda del contesto, è possibile creare, modificare ed eliminare connessioni utilizzando il menu contestuale (clic destro).

Per impostazione predefinita, Amazon Redshift crea un database predefinito denominato dev e uno schema predefinito denominato public. Per visualizzare i singoli file di dati del set di dati di esempio, scegliere un cluster, passare all'editor di query v2 e scegliere il database dev, lo schema public, quindi Tables.

In alternativa, nel pannello di navigazione, scegliere Clusters (Cluster) e il cluster su cui si desidera eseguire query sui dati. In Query dati, scegli Query nell'editor della query o Query nell'editor di query v2 per interrogare i dati nell'editor di query specificato.



Cluster	Cluster namespace	Status	Storage capacity us...	CPU utilization	Snapsh...	Notificati...	Tags
<input checked="" type="checkbox"/> ml-cluster1 ra3.4xlarge   2 nodes   256 TB	c286487b-0774-4cdf-...	Available	< 1%	< 1%	3 snapshots	1	
<input type="checkbox"/> redshift-cluster-1 ra3.4xlarge   2 nodes   256 TB	494f4bab-49e9-44ce-...	Available	< 1%	1%	3 snapshots	1	
<input type="checkbox"/> redshift-cluster-v2 dc2.large   1 node   160 GB	e25c4a3e-8630-476d-...	Available	< 1%	9%	3 snapshots	1	

## Provare le query di esempio

Provare alcune query di esempio in uno degli editor della query, come illustrato di seguito. Per ulteriori informazioni sull'utilizzo del comando SELECT, consultare [SELECT](#) nella Guida per sviluppatori di database di Amazon Redshift.

```
-- 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;
```

Dopo aver completato le fasi descritte in questo tutorial, sarà possibile utilizzare la sezione [Risorse aggiuntive](#) per ottenere informazioni più approfondite. Queste informazioni possono aiutare a pianificare, distribuire e gestire i cluster e lavorare con i dati contenuti nel data warehouse.

È possibile eseguire anche il tutorial [Utilizzo dei propri dati in Amazon Redshift](#) per creare un cluster con il proprio set di dati.

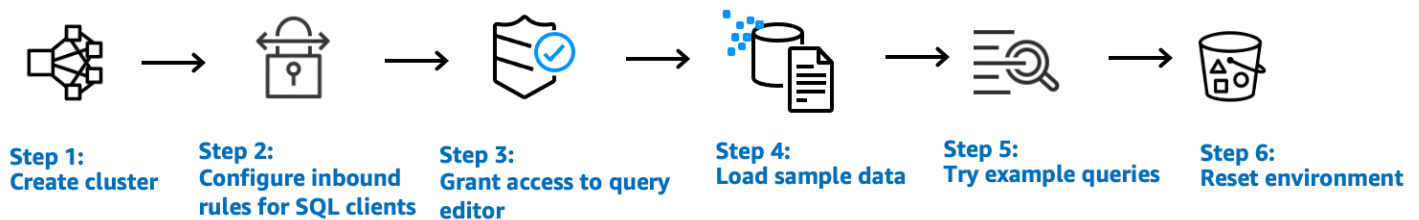


## Utilizzo dei propri dati in Amazon Redshift

In questo tutorial, si esplora il processo per creare un cluster Amazon Redshift utilizzando il proprio set di dati in Amazon Redshift. Puoi usare questo cluster di esempio per valutare il servizio Amazon Redshift.

Prima di iniziare la fase di configurazione di un cluster Amazon Redshift, assicurarsi di completare [Iscrizione a AWS](#) e [Definizione delle regole del firewall](#).

In questo tutorial, seguirai le fasi illustrate di seguito.



### ⚠ Important

Il cluster di esempio che crei viene eseguito in un ambiente reale. La tariffa oraria on-demand è pari a 0,25 \$ e consente di utilizzare il cluster campione creato in questo tutorial finché non viene eliminato. Per informazioni sui prezzi, consultare la [pagina dei prezzi di Amazon Redshift](#). Nel caso di domande o problemi, è possibile comunicare con il team di Amazon Redshift inserendo un post nel nostro [forum di discussione](#).

Il presente tutorial non è stato pensato per gli ambienti di produzione, né affronta le diverse opzioni in maniera dettagliata. Dopo aver completato le fasi descritte in questo tutorial, sarà possibile utilizzare la sezione [Risorse aggiuntive](#) per ottenere informazioni più approfondite. Queste informazioni possono aiutare a pianificare, distribuire e gestire i cluster e lavorare con i dati contenuti nel data warehouse.

### Argomenti

- [Fase 1: Creazione di un cluster Amazon Redshift di esempio](#)
- [Fase 2: Configurazione delle regole in ingresso per i client SQL](#)
- [Fase 3: concessione dell'accesso all'editor di query ed esecuzione delle 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](#)

## Fase 1: Creazione di un cluster Amazon Redshift di esempio

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 conto dell'utente. Un esempio può essere l'utilizzo di un comando COPY per caricare dati da Amazon Simple Storage Service (Amazon S3). Fornisci tali autorizzazioni utilizzando AWS Identity and Access Management (IAM). È possibile eseguire questa operazione tramite un ruolo IAM collegato al cluster. Oppure puoi fornire la chiave di AWS accesso per un utente IAM che dispone delle autorizzazioni necessarie. Per ulteriori informazioni sulle credenziali e le autorizzazioni di accesso, consultare [Credenziali e autorizzazioni di accesso](#).

Per proteggere al meglio i dati sensibili e salvaguardare le credenziali di AWS accesso, consigliamo di creare un ruolo IAM e di collegarlo al cluster. Per ulteriori informazioni sulla concessione delle autorizzazioni di accesso, consultare [Autorizzazioni di accesso ad altre risorse AWS](#).

Il cluster che stai per creare è eseguito in un ambiente live e non in una sandbox. Saranno addebitati i costi standard di utilizzo di Amazon Redshift relativi al cluster finché non viene eliminato. Se il presente tutorial viene completato in un'unica sessione e il cluster eliminato una volta terminate tutte le attività, il costo totale è minimo.

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/](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 [Controllo dell'accesso a utenti IAM](#) nella Guida alla gestione di Amazon Redshift.

2. In alto a destra, 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 una AWS regione che supporti i tipi di nodi RA3. Per un elenco di regioni AWS che supportano i tipi di nodo RA3, consulta l'argomento relativo alla [panoramica dei tipi di nodo RA3](#) nella Guida alla gestione di Amazon Redshift. Per ulteriori informazioni sulle specifiche del nodo per ogni tipo e dimensione di nodo, consulta [Dettagli sul tipo di nodo](#).

- Se la tua AWS regione supporta i tipi di nodi RA3, scegli Produzione o Versione di prova gratuita per rispondere alla domanda Per cosa intendi utilizzare questo cluster?
- Se non si conosce la dimensione del cluster, scegliere Aiutami a scegliere. In questo modo viene avviato un calcolatore di dimensionamento in cui vengono poste domande sulle dimensioni e le caratteristiche delle query dei dati che si prevede di archiviare nel 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 il proof of concept.

- Scegli dc2.large per Node type (Tipo di nodo) e 2 per Nodes (Nodi) per questo tutorial.
- Se hai scelto Produzione per il cluster, procedere in uno dei modi seguenti:
  - 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.
  - Per utilizzare i propri dati in Amazon Redshift, continuare con il resto del tutorial.

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: specificare un valore per 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 ulteriori informazioni su come gestire il ruolo IAM di default per un cluster, consultare [Creazione di un ruolo IAM come default per Amazon Redshift](#).

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, consultare [Abilitazione del routing Amazon VPC avanzato](#) nella Guida alla gestione di Amazon Redshift.

8. Scegli Create cluster (Crea cluster).

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

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.

Se si prevede di accedere al cluster con l'editor di query di Amazon Redshift dall'interno del VPC, è possibile saltare questo passaggio.

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 client potrebbe essere un'istanza Amazon Elastic Compute Cloud (Amazon EC2) o un computer esterno.

Per ulteriori informazioni sulle regole a livello di firewall, consulta [Regole del gruppo di sicurezza](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

2. Per accedere da un client esterno Amazon EC2, aggiungere una regola di ingresso al gruppo di sicurezza collegato al cluster che consenta il traffico in entrata. Le regole del gruppo di sicurezza Amazon EC2 vengono aggiunte nella console Amazon EC2. Ad esempio, un CIDR/IP di 192.0.2.0/24 consente ai client in tale intervallo di indirizzi IP di connettersi al cluster. Individuare il CIDR/IP corretto per il proprio ambiente.

## Fase 3: concessione dell'accesso all'editor di query ed esecuzione delle query

Per eseguire query sui database ospitati dal cluster Amazon Redshift, sono disponibili due opzioni:

1. Connettiti al tuo cluster ed esegui query AWS Management Console con uno degli editor di query.

Se utilizzi l'editor della query, non devi scaricare e configurare un'applicazione client SQL.

2. Connessione al cluster con uno strumento client SQL, ad esempio SQL Workbench/J. Per ulteriori informazioni sull'uso di SQL Workbench/J, consulta [Connessione al cluster tramite SQL Workbench/J](#) nella Guida alla gestione di Amazon Redshift.

L'uso dell'editor di query di Amazon Redshift è la soluzione più semplice per eseguire query sui database ospitati dal cluster Amazon Redshift. Dopo aver creato il cluster, è possibile eseguire immediatamente le query utilizzando la console Amazon Redshift. Per dettagli sulle considerazioni durante l'utilizzo dell'editor di query di Amazon Redshift, consulta [Esecuzione di query in un database con l'editor della query](#) nella Guida alla gestione di Amazon Redshift.

### Concessione dell'accesso all'editor di query v2

La prima volta che un amministratore configura l'editor di query v2 per l'utente Account AWS, sceglie AWS KMS key quello utilizzato per crittografare le risorse dell'editor di query v2. 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'ARN (Amazon Resource Name) 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, consultare [Configurazione del tuo Account AWS](#)

Per accedere all'editor della query v2, sono necessarie le opportune autorizzazioni. Un amministratore può allegare una delle seguenti politiche AWS gestite 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. Per collegare le policy IAM è possibile utilizzare la console IAM all'indirizzo <https://console.aws.amazon.com/iam/>.

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 ulteriori informazioni, consultare [Accesso all'editor di query v2](#).

Utilizzo dell'editor della query v2

Per utilizzare l'editor di query v2 per interrogare un database, consultare [Utilizzo dell'editor di query v2](#).

Concessione dell'accesso all'editor di query

Per utilizzare l'editor di query Amazon Redshift, è necessaria l'autorizzazione. Per configurare l'accesso, collega le policy IAM `AmazonRedshiftQueryEditor` e `AmazonRedshiftReadOnlyAccess` a un ruolo IAM, quindi assegna il ruolo all'utente che esegue l'accesso al cluster.

Per collegare le policy IAM richieste per l'editor della query

1. [Accedi AWS Management Console e apri la console IAM all'indirizzo https://console.aws.amazon.com/iam/](https://console.aws.amazon.com/iam/).
2. Scegli Ruoli.
3. Scegli il ruolo che deve accedere all'editor di query.
4. Nella scheda Autorizzazioni scegli Aggiungi autorizzazioni.
5. Scegli Collega policy.
6. In Nomi delle policy scegli `AmazonRedshiftQueryEditor` e `AmazonRedshiftReadOnlyAccess`.
7. Scegli Aggiungi autorizzazioni.

Per creare un ruolo IAM che consenta al cluster Amazon Redshift di comunicare con altri servizi AWS per tuo conto, completa la procedura riportata di seguito. I valori utilizzati in questa sezione sono esempi, è possibile scegliere i valori in base alle proprie esigenze.

Creare un ruolo IAM per consentire ad Amazon Redshift di accedere ai servizi AWS

1. Aprire la [console IAM](#).
2. Nel pannello di navigazione, selezionare Ruoli.
3. Selezionare Create role (Crea ruolo).
4. Seleziona Servizio AWS e quindi Redshift.

5. In **Select your use case** (Seleziona il tuo caso d'uso) scegliere **Redshift - Customizable** (Redshift - Personalizzabile) e quindi scegliere **Next: Permissions** (Successivo: Autorizzazioni). Verrà visualizzata la pagina **Attach permissions policy** (Collega policy di autorizzazioni).
6. Aggiungi le **AmazonRedshiftReadOnlyAccess** autorizzazioni **AmazonRedshiftQueryEditore**.  
Scegli **Successivo: Tag**.
7. Viene visualizzata la pagina **Aggiungi tag**. È inoltre possibile aggiungere i tag. Scegli **Prossimo: Rivedi**.
8. Per **Role name** (Nome ruolo), digitare un nome per il ruolo, ad esempio **RedshiftQueryEditorAccess**. Scegliere **Crea ruolo**.
9. Il nuovo ruolo è disponibile per tutti gli utenti nei cluster che usano il ruolo. Per limitare l'accesso solo a utenti specifici in cluster specifici oppure a cluster in regioni specifiche, modificare la relazione di trust per il ruolo. Per ulteriori informazioni, consulta [Limitazione dell'accesso a ruoli IAM](#) nella Guida alla gestione di Amazon Redshift.
10. Associare il ruolo al cluster. È possibile associare un ruolo IAM a un cluster quando si crea il cluster oppure quando si aggiunge il ruolo a un cluster esistente. Per ulteriori informazioni, consulta [Associazione di ruoli IAM ai cluster](#) nella Guida alla gestione di Amazon Redshift.

#### Note

Per limitare l'accesso a dati specifici, utilizza un ruolo IAM che concede il minor numero di privilegi richiesti.

## Utilizzo dell'editor della query

Per utilizzare l'editor della query per eseguire query su un database, consultare [Esecuzione di query in un database utilizzando l'editor della query](#).

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

L'utilizzo di uno degli editor di query di Amazon Redshift è il modo più semplice per caricare i dati nelle tabelle. Dopo aver creato il cluster, puoi caricare i dati da Amazon S3 sul cluster utilizzando la console Amazon Redshift.

L'utilizzo dell'editor di query v2 semplifica il caricamento dei dati quando si utilizza la procedura guidata **Carica dati**. Per ulteriori informazioni, consulta [Caricamento dei dati personali da Amazon S3 ad Amazon Redshift utilizzando l'editor di query v2](#). È anche possibile utilizzare l'editor di query



v2 per creare tabelle e caricare i dati. Per ulteriori informazioni, consulta [Caricare i dati campione da Amazon S3 utilizzando l'editor della query](#).

Caricamento dei dati personali da Amazon S3 ad Amazon Redshift utilizzando l'editor di query v2

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 prima cosa, connettersi a un database. Successivamente, creare alcune tabelle nel database. Caricare quindi i tuoi dati da Amazon S3 ad Amazon Redshift. Per ulteriori informazioni sull'utilizzo dell'editor di query v2, consulta [Utilizzo dell'editor di query v2](#) nella Guida alla gestione di Amazon Redshift.

Il comando COPY generato e utilizzato nell'editor di query v2 della procedura guidata Carica dati supporta tutti i parametri disponibili per la sintassi del comando COPY per caricare i 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.

Caricare i dati campione da Amazon S3 utilizzando l'editor della query

A questo punto sei connesso al database denominato dev. Successivamente, devi creare alcune tabelle nel database, caricare i dati nelle tabelle e provare a eseguire una query. Per maggiore praticità, i dati di esempio da caricare sono disponibili in un bucket Amazon S3.

#### Note

Se utilizzi uno strumento client SQL; verifica che il client SQL sia connesso al cluster.

Dopo aver completato questa fase, è possibile effettuare le seguenti operazioni:

- Provare le query di esempio su [Fase 5: provare query di esempio utilizzando l'editor di query](#).
- Reimpostare l'ambiente su [Fase 6: reimpostazione dell'ambiente](#).
- Per ulteriori informazioni su Amazon Redshift, consultare [Risorse aggiuntive](#).

#### Note

Per provare a eseguire query sui dati nell'editor di query senza caricare i propri dati, scegliere Carica dati di esempio in Dati di esempio. In tal caso, Amazon Redshift carica

automaticamente il proprio set di dati di esempio nel cluster Amazon Redshift durante la creazione del cluster.

## Come caricare i dati di esempio da Amazon S3

### 1. Per creare le tabelle.

Se si utilizza l'editor delle query Amazon Redshift, copiare ed eseguire individualmente 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 venue(  
  venueid smallint not null distkey sortkey,  
  venuename varchar(100),  
  venuecity varchar(30),  
  venuestate char(2),  
  venueseats integer);
```

```
create table category(  
  catid smallint not null distkey sortkey,  
  catgroup varchar(10),  
  catname varchar(10),  
  catdesc varchar(50));
```

```
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'));
```

```
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 listing(  
  listid integer not null distkey,  
  sellerid integer not null,  
  eventid integer not null,  
  dateid smallint not null sortkey,  
  numtickets smallint not null,  
  priceperticket decimal(8,2),  
  totalprice decimal(8,2),  
  listtime 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);
```

## 2. Caricare i dati di esempio 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 o Amazon DynamoDB. Per ulteriori informazioni sulla sintassi di COPY, consultare [COPY](#) nella Guida per gli sviluppatori di database di Amazon Redshift.

- a. Scaricare il file [ticketdb.zip](#) contenente i singoli file di dati di esempio.
- b. Decomprimere e caricare i singoli file in una cartella `ticket` del bucket Amazon S3 nella Regione AWS.
- c. Modificare i comandi COPY in questo tutorial in modo che puntino ai file del bucket Amazon S3. Per informazioni su come gestire i file con Amazon S3, consultare [Creazione e configurazione di un bucket S3](#) nella Guida per l'utente Amazon Simple Storage Service.
- d. Per caricare i dati di esempio, è necessario fornire al cluster l'autenticazione per accedere ad Amazon S3 per tuo conto. L'autenticazione sarà fornita referenziando il ruolo IAM creato e impostato come predefinito per il cluster nelle fasi precedenti.

I comandi COPY includono un segnaposto per l'Amazon Resource Name (ARN) per il ruolo IAM, il nome del bucket e un Regione AWS, come illustrato nell'esempio seguente.

```
copy users from 's3://<myBucket>/ticket/allusers_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

Il comando COPY è simile al seguente esempio:

```
copy users from 's3://<myBucket>/ticket/allusers_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

Per caricare i dati di esempio, sostituire `<myBucket>` , e `<aws-region>` nei seguenti comandi COPY con i valori. Se si utilizza l'editor di query di Amazon Redshift, eseguire singolarmente i seguenti comandi.

```
copy users from 's3://<myBucket>/tickit/allusers_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>';
```

```
copy venue from 's3://<myBucket>/tickit/venue_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>';
```

```
copy category from 's3://<myBucket>/tickit/category_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>';
```

```
copy date from 's3://<myBucket>/tickit/date2008_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>';
```

```
copy event from 's3://<myBucket>/tickit/allevvents_pipe.txt'  
iam_role default  
delimiter '|' timeformat 'YYYY-MM-DD HH:MI:SS' region '<aws-region>';
```

```
copy listing from 's3://<myBucket>/tickit/listings_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>';
```

```
copy sales from 's3://<myBucket>/tickit/sales_tab.txt'  
iam_role default  
delimiter '\t' timeformat 'MM/DD/YYYY HH:MI:SS' region '<aws-region>';
```

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

Ora, provare alcune query di esempio, come illustrato di seguito. Per ulteriori informazioni sull'utilizzo del comando SELECT, consultare [SELECT](#) nella Guida per gli sviluppatori di database di Amazon Redshift.

```
-- 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;
```

È stato creato correttamente un cluster Amazon Redshift ed è stata eseguita una query sui dati dal set di dati utilizzando l'editor di query Amazon Redshift.

## Fase 6: reimpostazione dell'ambiente

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, se si desidera provare a eseguire attività descritte in altre guide Amazon Redshift è consigliabile tenere il cluster di esempio in esecuzione.

Per eliminare un cluster

1. [Accedi AWS Management Console e apri la console Amazon Redshift all'indirizzo https://console.aws.amazon.com/redshiftv2/](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). Appare la pagina Create cluster (Crea cluster).
4. Conferma il cluster da eliminare, quindi 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](#).

## Attività di database comuni

Di seguito è riportata una descrizione e una procedura dettagliata per le attività comuni, così da poter iniziare a utilizzare i database Amazon Redshift.

Dopo la connessione al database dev del cluster iniziale, è possibile creare un nuovo database. Indipendentemente dal fatto che si scelga di utilizzare il set di dati di esempio o i propri dati in Amazon Redshift durante la creazione di un cluster, Amazon Redshift crea il database dev.

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

- È stato creato un cluster Amazon Redshift. Per ulteriori informazioni, consulta [Cluster Amazon Redshift e caricamento dei dati](#).
- È stata stabilita una connessione al cluster dallo strumento client di SQL, come l'editor di query della console di Amazon Redshift. Per ulteriori informazioni, consulta [Fase 3: concessione dell'accesso all'editor di query ed esecuzione delle query](#).

### Important

Il cluster distribuito per questo esercizio verrà eseguito in un ambiente reale. Finché è in esecuzione, addebita addebiti sul tuo 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 fase finale dell'esercizio descrive come fare.

## Attività 1: Creazione di un database

Dopo aver verificato che il cluster è attivo e in esecuzione, è possibile creare il primo database. Il database è l'elemento in cui effettivamente saranno create tabelle, saranno caricati i dati e saranno eseguite le query. Un singolo cluster può ospitare più database. Ad esempio, lo stesso cluster può contenere un database SALESDB e un database ORDERSDB.

Ad esempio, per creare un database denominato **SALESDB**, emettere il comando seguente nello strumento client SQL:

```
CREATE DATABASE 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.

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.

## Attività 2: Creazione di un utente

Per impostazione predefinita, solo l'utente amministratore creato all'avvio del cluster ha accesso al database iniziale nel cluster. Per concedere l'accesso ad altri utenti, è necessario creare uno o più account. Gli account utente dei database sono globali per tutti i database in un cluster e non appartengono a singoli 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.

### Attività 3: Creazione di 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 gli oggetti del database specificati, come le tabelle, le viste e le funzioni definite dall'utente (FDU). 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 oggetti di database diversi, come le tabelle, le viste e le funzioni definite dall'utente (FDU). 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.

Nell'esempio seguente viene concesso il privilegio all'utente `GUEST` per selezionare i dati da tutte le tabelle o viste in `SALESCHEMA` utilizzando un'istruzione `SELECT`.

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

Nell'esempio seguente vengono concessi tutti i privilegi disponibili all'utente `GUEST` in una sola volta.

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

## Attività 4: Creazione di una tabella

Dopo aver creato il nuovo database, è necessario creare le tabelle che conterranno i dati. Quando si crea la tabella, è necessario specificare tutte le informazioni sulle colonne.

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

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

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 ed esaminare gli schemi e le relative tabelle, è possibile utilizzare l'editor di query di Amazon Redshift. Oppure è possibile visualizzare l'elenco delle tabelle negli schemi usando le viste di sistema. Per ulteriori informazioni, consulta [Fase 6: Esecuzione di query sulle tabelle di sistema](#).

Per impostazione predefinita, i nuovi oggetti di database, come le tabelle, vengono creati nello schema predefinito denominato `public` creato durante la creazione del cluster. 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.

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 si trova in uno schema, 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](#).

## Attività 5: Caricamento di dati di esempio

Nella maggior parte degli esempi in questa guida viene utilizzato il set di dati di esempio TICKIT. È possibile scaricare il file [ticketdb.zip](#) contenente i singoli file di dati di esempio.

Puoi caricare i dati campione 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](#).

## Fase 6: Esecuzione di query sulle tabelle di sistema

Oltre alle tabelle che hai creato, il database contiene alcune tabelle di sistema. Queste tabelle di sistema contengono informazioni sull'installazione e sui diversi processi e query in esecuzione nel sistema. Puoi eseguire query su queste tabelle di sistema per raccogliere informazioni sul database.

### Note

La descrizione per ogni tabella in questa documentazione indica se una tabella è visibile a tutti gli utenti o solo agli utenti con privilegi avanzati. Per eseguire una query su tabelle visibili solo agli utenti con privilegi avanzati, è necessario accedere come utente con privilegi avanzati.

Amazon Redshift fornisce l'accesso ai tipi seguenti di tabelle di sistema:

- [Tabelle STL](#)

Queste tabelle di sistema vengono generate da file di log di Amazon Redshift per fornire una cronologia del sistema. Le tabelle di registrazione hanno un prefisso STL.

- [Tabelle STV](#)

Queste tabelle sono tabelle di sistema virtuali che contengono snapshot dei dati di sistema correnti. Le tabelle di snapshot hanno un prefisso STV.

- [Viste di sistema](#)

Le visualizzazioni di sistema contengono un sottoinsieme di dati presenti in diverse tabelle di sistema STL e STV. Le visualizzazioni di sistema hanno un prefisso SVV o SVL.

- [Tabelle di catalogo di sistema](#)

Nelle tabelle del catalogo di sistema sono archiviati i metadati dello schema, ad esempio le informazioni su tabelle e colonne. Le tabelle di catalogo di sistema hanno il prefisso PG.

Potresti dover specificare l'ID processo associato a una query per recuperare informazioni delle tabelle di sistema sulla query. Per ulteriori informazioni, consulta [Individua l'ID processo di una query in esecuzione](#).

## 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
(1 row)
```

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
(1 row)

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
(2 rows)

```

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

```

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
(4 rows)

```

Per ulteriori informazioni, consultare [PG\\_TABLE\\_DEF](#).

È inoltre possibile utilizzare l'editor di query v2 per visualizzare tutte le tabelle in uno schema specificato scegliendo innanzitutto un database a cui connettersi.

## 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     | ***** |          |

```

```
(3 rows)
```

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    | ***** |          |
(2 rows)
```

## Visualizza le query recenti

Nell'esempio precedente, l'ID utente (`USESYSID`) per `adminuser` è 100. Per elencare le cinque query più recenti eseguite da `adminuser`, è possibile eseguire una query sulla vista `SVL_QLOG`.

La vista `SVL_QLOG` è un sottoinsieme più intuitivo delle informazioni contenute nella tabella `STL_QUERY`. È possibile usare questa vista per trovare l'ID query (`QUERY`) o l'ID processo (`PID`) per una query eseguita di recente. È inoltre possibile utilizzare questa vista per verificare quanto tempo ha richiesto il completamento di una query. `SVL_QLOG` include i primi 60 caratteri della stringa di query (`SUBSTRING`) per semplificare l'individuazione di una query specifica. Usa la clausola `LIMIT` con l'istruzione `SELECT` per limitare i risultati a cinque righe.

```
SELECT query, pid, elapsed, substring from svl_qlog
WHERE userid = 100
ORDER BY starttime desc
LIMIT 4;
```

Il risultato è simile al seguente.

```
query| pid | elapsed | substring
-----+-----+-----+-----
+-----+
892 | 21046 | 55868 | SELECT query, pid, elapsed, substring from svl_qlog WHERE
us
620 | 17635 | 1296265 | SELECT query, pid, elapsed, substring from svl_qlog WHERE us
```



```
610 | 17607 | 82555 | SELECT * from DEMO;
596 | 16762 | 226372 | INSERT INTO DEMO VALUES (100);)
```

## Individua l'ID processo di una query in esecuzione

Potresti dover trovare il PID per una query ancora in fase di esecuzione. Ad esempio, il PID servirà se è necessario annullare una query la cui esecuzione sta durando troppo a lungo. Puoi eseguire una query sulla tabella di sistema STV\_RECENTS per ottenere un elenco degli ID processo per le query in esecuzione, insieme alla stringa di query corrispondente. Se la query restituisce più PID, puoi esaminare il testo della query per determinare il PID necessario.

Per determinare il PID di una query in esecuzione, eseguire l'istruzione SELECT seguente:

```
SELECT pid, user_name, starttime, query
FROM stv_recents
WHERE status='Running';
```

## Attività 7: Annullamento di una query

Se si esegue una query che impiega troppo tempo o utilizza un numero eccessivo di risorse del cluster, è possibile annullarla. 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, usa il comando CANCEL con il PID della query.

Per trovare l'ID processo, avviare una nuova sessione ed eseguire una query sulla tabella STV\_RECENTS, come mostrato nella fase 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.

```
SELECT pid, trim(user_name), starttime, substring(query,1,20)
FROM stv_recents
WHERE status='Running';
```

Il risultato è simile al seguente.

pid	btrim	starttime	substring
610	adminuser	2013-03-28 18:39:49.355918	select sellerid, fir

```
(1 row)
```

Per annullare la query con PID 610, esegui il comando seguente.

```
CANCEL 610;
```

### 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.

## Annulla una query da un'altra sessione

Se lo strumento di query non supporta l'esecuzione simultanea di query, sarà necessario avviare un'altra sessione per annullare la query. Ad esempio, l'editor di query utilizzata in Guida alle operazioni di base di Amazon Redshift non supporta più query simultanee. Per avviare un'altra sessione con l'editor di query, selezionare File, Nuova finestra e connettersi utilizzando gli stessi parametri di connessione. Puoi quindi trovare il PID e annullare la query.

## 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 tramite la coda dell'utente con privilegi avanzati, emettere questi comandi:

```
SET query_group TO 'superuser';
CANCEL 610;
RESET query_group;
```

## Attività 8: Eliminazione delle risorse

Se per completare questo esercizio è stato distribuito un cluster, al termine sarà necessario eliminarlo. L'eliminazione del cluster interrompe l'addebito sull'account AWS .

Per eliminare il cluster, attenersi alla procedura descritta in [Eliminazione di un cluster](#) nella Guida alla gestione di Amazon Redshift.

Se desideri conservare il cluster, conserva anche i dati di esempio come riferimento. Nella maggior parte degli esempi in questa guida sono utilizzate le tabelle create in questo esercizio. Le dimensioni dei dati non hanno un impatto significativo sull'archiviazione disponibile.

Se desideri conservare il cluster ma eliminare i dati di esempio, emettere il comando seguente ed eliminare il database SALESDB:

```
DROP DATABASE SALESDB;
```

Se non è stato creato un database SALESDB o se non si desidera eliminare il database, emettere i comandi seguenti per eliminare solo le tabelle:

```
DROP TABLE DEMO;
DROP TABLE users;
DROP TABLE venue;
DROP TABLE category;
DROP TABLE date;
DROP TABLE event;
DROP TABLE listing;
DROP TABLE sales;
```

# Panoramica concettuale 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. È possibile caricare i dati e iniziare subito a eseguire query nell'editor di query Amazon Redshift v2 o nello strumento di business intelligence (BI) preferito. È possibile usufruire del miglior rapporto prezzo/prestazioni e delle funzionalità SQL familiari in un ambiente facile da usare e senza alcuna 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.
- [Getting started with Amazon Redshift Serverless](#) (Guida introduttiva ad Amazon Redshift Serverless): in questo argomento, vengono fornite ulteriori informazioni su come creare un data warehouse di Amazon Redshift Serverless e iniziare a eseguire query sui dati utilizzando l'editor di query v2.

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

Se la propria organizzazione è idonea e il cluster è stato creato in una Regione AWS in cui Amazon Redshift serverless non è disponibile, potrebbe essere possibile creare un cluster con il programma di prova gratuito di Amazon Redshift. Scegliere Produzione o Versione di prova gratuita per rispondere alla domanda Per cosa si intende utilizzare questo cluster? Se si sceglie Versione di prova gratuita, si crea 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 Regioni AWS in cui Amazon Redshift serverless è disponibile, consultare gli endpoint elencati per l'[API Redshift serverless](#) in 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 le unità di elaborazione Redshift (RPU), i gruppi di sicurezza, i limiti di utilizzo. I gruppi di lavoro dispongono di impostazioni di rete e sicurezza che è possibile configurare utilizzando la console di Amazon Redshift serverless, l'AWS Command Line Interface e le API di Amazon Redshift serverless.

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.

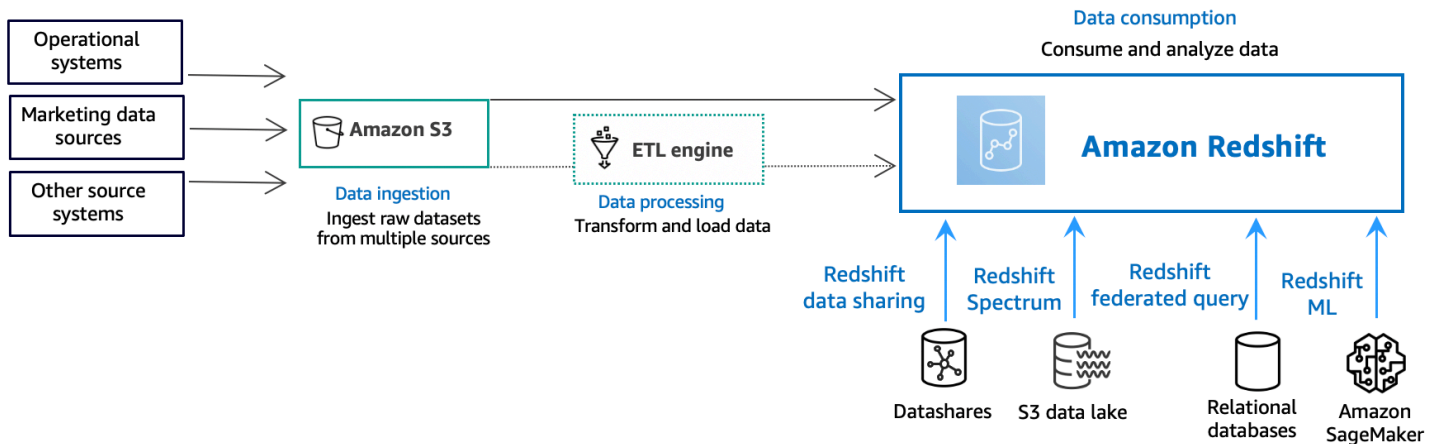
- Database – Un cluster contiene uno o più database.

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 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.
  - [Tabelle e viste di sistema](#): in questo argomento sono descritte le tabelle e le visualizzazioni 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.