



Guida alle operazioni di base

Amazon Redshift



Amazon Redshift: Guida alle operazioni di base

Copyright © 2026 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 dei rispettivi proprietari, che possono o meno essere affiliati, collegati o sponsorizzati da Amazon.

Table of Contents

.....	v
Nozioni di base sui data warehouse serverless	1
Iscrizione a AWS	1
Creazione di un data warehouse con Amazon Redshift serverless	1
Caricamento di dati di esempio	4
Esecuzione di query di esempio	6
Caricamento di dati da Amazon S3	7
Nozioni di base sui data warehouse con provisioning	14
Iscrizione a AWS	16
Definizione delle regole del firewall	17
Fase 1: creazione di un cluster di esempio	17
Fase 2: Configurazione delle regole in ingresso per i client SQL	21
Fase 3: concedere l'accesso a un client SQL ed eseguire query	21
Concessione dell'accesso all'editor di query v2	22
Passaggio 4: caricare i dati da Amazon S3 ad Amazon Redshift	23
Caricamento di dati da Amazon S3 con i comandi SQL	23
Caricamento di dati da Amazon S3 con Query Editor V2	25
Creazione di dati TICKIT nel cluster	25
Fase 5: provare query di esempio utilizzando l'editor di query	26
Fase 6: reimpostazione dell'ambiente	28
Definire e utilizzare un database nel data warehouse	29
Connessione ad Amazon Redshift	30
Creazione di un database	31
Creazione di un utente	32
Creare uno schema	32
Creare una tabella	34
Inserisci le righe di dati in una tabella	35
Seleziona dati da una tabella	35
Caricare i dati	36
Eseguire query sulle tabelle e sulle viste di sistema	36
Visualizza un elenco di nomi di tabella	36
Visualizzazione degli utenti	38
Visualizza le query recenti	38
Determinare l'ID sessione di una query in esecuzione	39

Annullare una query	40
Annulla una query tramite la coda dell'utente con privilegi avanzati	42
Eseguire query sui dati non presenti nel database Amazon Redshift	43
Esecuzione di query nei data lake	43
Esecuzione di query su origini dati remote	44
Accesso ai dati in altri database	44
Addestramento dei modelli di ML con i dati Redshift	45
Scoprire i concetti di Amazon Redshift	46
Risorse di apprendimento aggiuntive	49
Cronologia dei documenti	51

Amazon Redshift non supporterà più la creazione di nuovi Python UDFs a partire dalla Patch 198. Python esistente UDFs continuerà a funzionare fino al 30 giugno 2026. Per ulteriori informazioni, consulta il [post del blog](#).

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.

Nozioni di base sui data warehouse 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. I dati di esempio vengono utilizzati in tutta la documentazione di Amazon Redshift per illustrare le funzionalità. Per iniziare a utilizzare i data warehouse con provisioning Amazon Redshift, consulta [Nozioni di base sui data warehouse con provisioning Amazon Redshift](#).

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

Iscrizione a AWS

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

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

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

Creazione di un data warehouse con Amazon Redshift serverless

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

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

Note

Redshift serverless richiede un VPC Amazon con tre sottoreti in tre zone di disponibilità diverse. Redshift serverless richiede inoltre almeno tre indirizzi IP disponibili. Prima di continuare, assicurati che il VPC Amazon che utilizzi per Redshift serverless abbia tre sottoreti in tre zone di disponibilità diverse e almeno tre indirizzi IP disponibili. Per ulteriori informazioni sulla creazione di sottoreti in un VPC Amazon, consulta [Creare una sottorete](#) nella Guida per l'utente di Amazon VPC. Per ulteriori informazioni sugli indirizzi IP in un Amazon VPC, consulta [Indirizzamento IP per le tue sottoreti VPCs](#).

Per configurare utilizzando le impostazioni predefinite:

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

Scegli Prova la versione di prova gratuita di Redshift serverless.

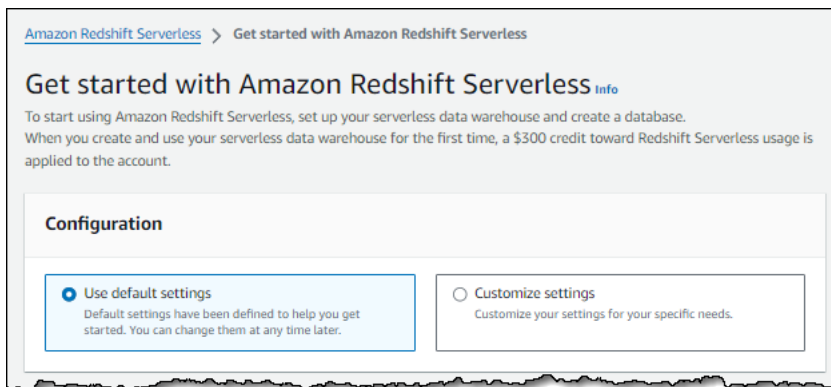
2. In Configuration (Configurazione), scegli Use default settings (Utilizza impostazioni predefinite). Amazon Redshift serverless crea un namespace predefinito a cui è associato un gruppo di lavoro predefinito. Seleziona Save configuration (Salva configurazione).

Note

Un namespace è una raccolta di oggetti di database e utenti. I namespace raggruppano tutte le risorse utilizzate in Amazon Redshift serverless, come schemi, tabelle, utenti, unità di condivisione dati e snapshot.

Un gruppo di lavoro è una raccolta di risorse di calcolo. I gruppi di lavoro includono le risorse di calcolo utilizzate da Amazon Redshift serverless per eseguire attività di calcolo.

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



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

Serverless dashboard [Info](#)

Namespace overview [Info](#)
Namespace data from your account

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

Namespaces / Workgroups [Info](#)

Namespace	Status	Workgroup	Status
default	✔ Available	default	✔ Available

Note

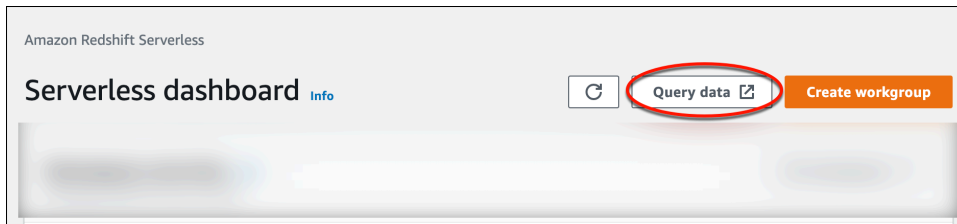
Se Redshift serverless non crea correttamente il gruppo di lavoro, puoi fare quanto segue:

- Risolvi eventuali errori segnalati da Redshift serverless, come la presenza di un numero insufficiente di sottoreti nel VPC Amazon.
- Per eliminare il namespace, scegli default-namespace nella dashboard di Redshift serverless e quindi Operazioni, Elimina spazio dei nomi. L'eliminazione di un namespace richiede diversi minuti.
- Quando apri di nuovo la console Redshift serverless, viene visualizzata la schermata di benvenuto.

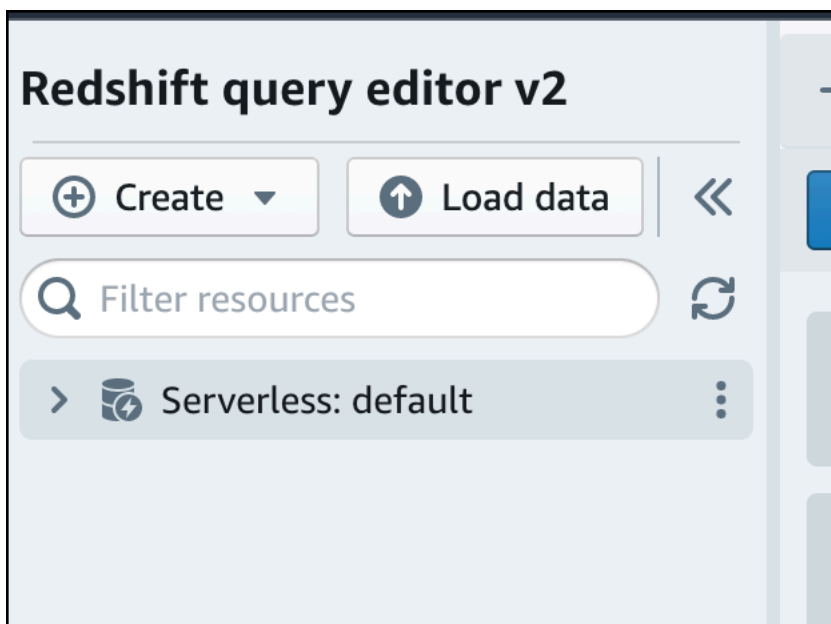
Caricamento di dati di esempio

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

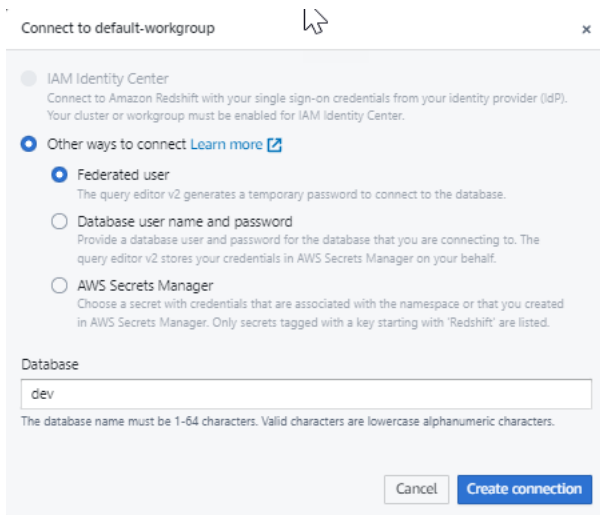
1. Per avviare l'editor di query v2 dalla console Amazon Redshift Serverless, scegli Esegui query sui data. Quando richiami l'editor di query v2 dalla console di Amazon Redshift Serverless, viene visualizzata una nuova scheda del browser con l'editor di query. L'editor di query v2 si connette dal computer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



2. Per questa guida, utilizzerai il tuo account AWS amministratore e quello predefinito AWS KMS key. Per ulteriori informazioni sulla configurazione di Query Editor V2, incluse le autorizzazioni necessarie, consulta [Configurazione dell' Account AWS](#) nella Guida alla gestione di Amazon Redshift. Per informazioni sulla configurazione di Amazon Redshift per l'utilizzo di una chiave gestita dal cliente o per modificare la chiave KMS utilizzata da Amazon Redshift, [consulta AWS KMS Changing the key for a namespace](#).
3. Per connettersi a un gruppo di lavoro, scegliere il nome del gruppo di lavoro nel pannello con struttura ad albero.

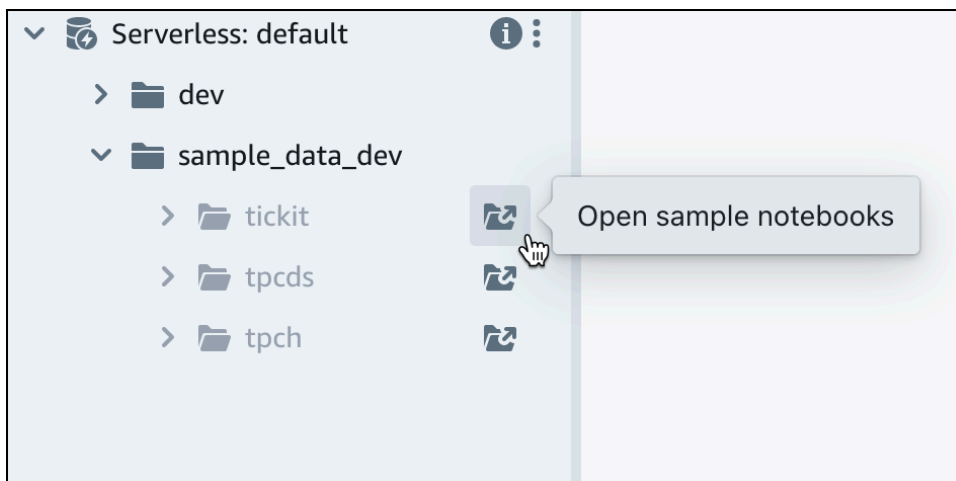


- Quando ti connetti a un nuovo gruppo di lavoro per la prima volta nell'editor di query v2, devi selezionare il tipo di autenticazione da utilizzare per connetterti al gruppo di lavoro. Per questa guida, lascia selezionato Utente federato e scegli Crea connessione.



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

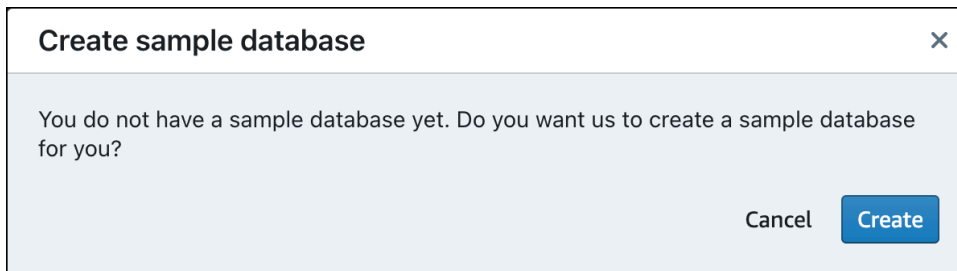
- Nel gruppo di lavoro Amazon Redshift Serverless predefinito, espandi il database `sample_data_dev`. Esistono tre schemi di esempio corrispondenti a tre set di dati di esempio che puoi caricare nel database Amazon Redshift serverless. Scegli il set di dati di esempio che desideri caricare e scegli Apri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



Note

Un notebook SQL è un container per le celle SQL e Markdown. Puoi utilizzare i notebook per organizzare, annotare e condividere più comandi SQL in un singolo documento.

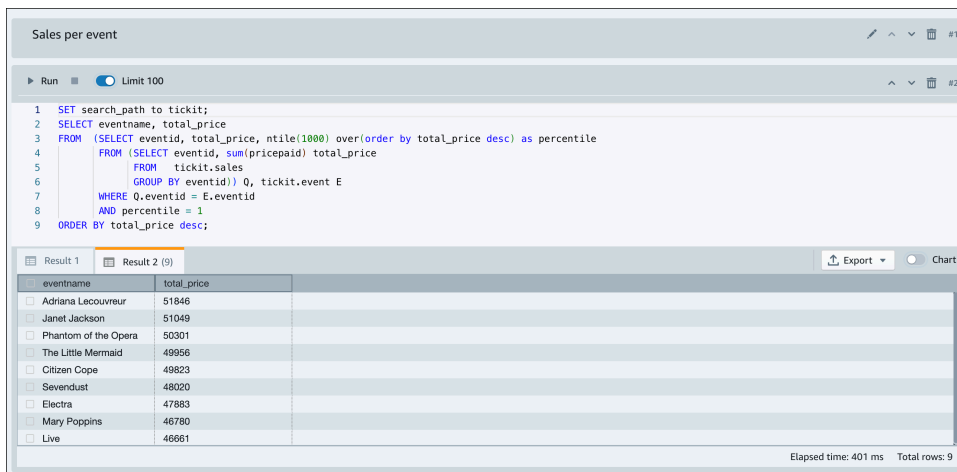
6. Quando si caricano i dati per la prima volta, l'editor di query v2 richiederà di creare un database di esempio. Scegli Create (Crea).



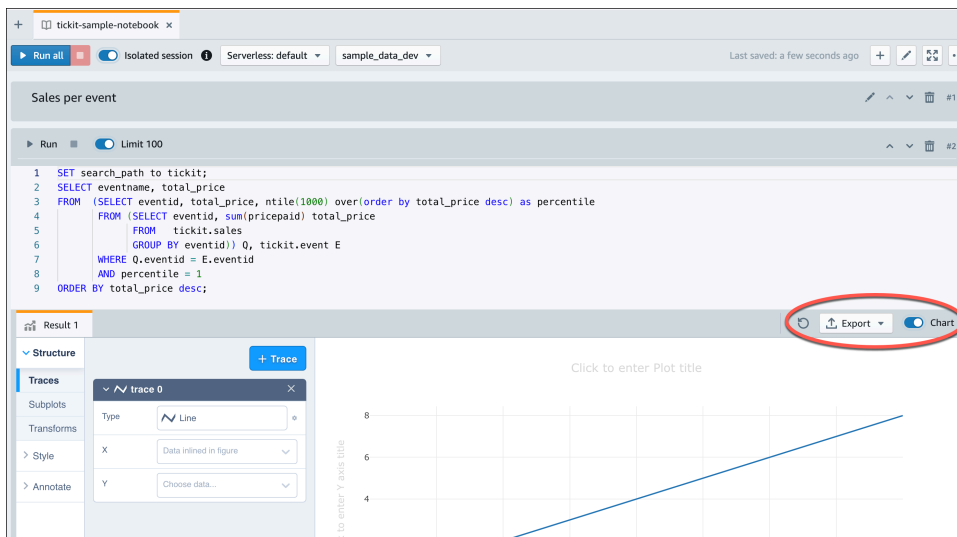
Esecuzione di query di esempio

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

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



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



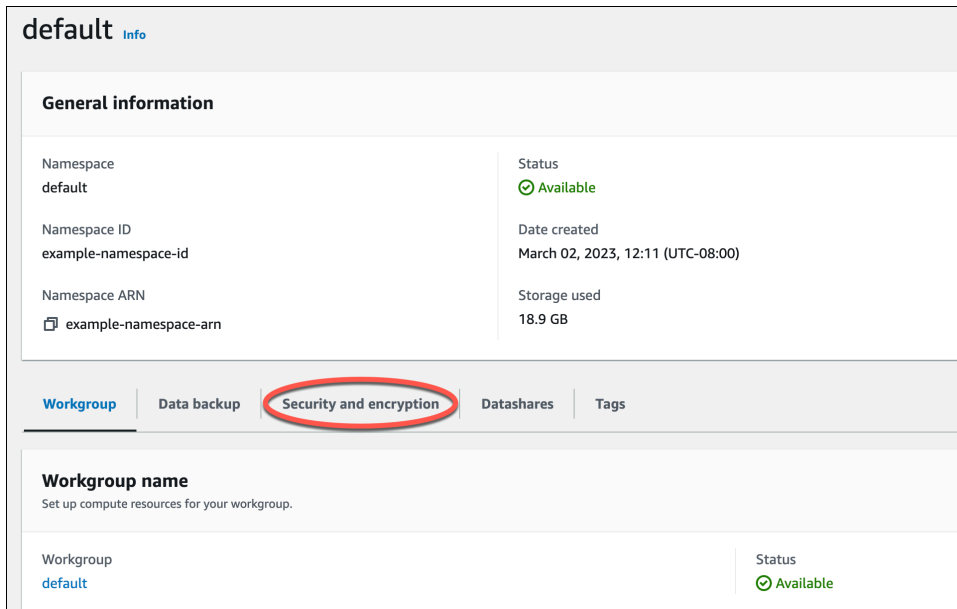
Puoi anche caricare i dati da un bucket Amazon S3. Per ulteriori informazioni, consulta [the section called “Caricamento di dati da Amazon S3”](#).

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. A tale scopo torna alla console Redshift serverless e scegli Configurazione del namespace. Dal menu di navigazione scegli il namespace e quindi 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

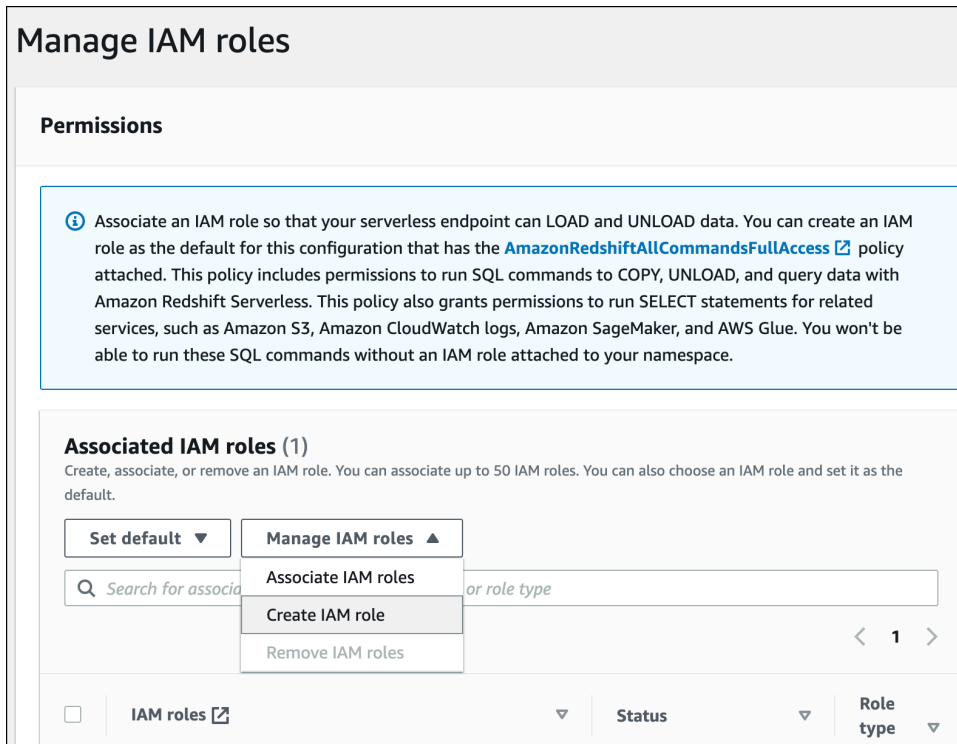
Workgroup | Data backup | **Security and encryption** | Datashares | Tags

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

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

Associated IAM roles (1)

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

Set default ▼ **Manage IAM roles** ▲

🔍 Search for associated IAM roles or role type

- Associate IAM roles
- Create IAM role**
- Remove IAM roles

< 1 >

<input type="checkbox"/>	IAM roles ?	Status	Role type
--------------------------	-----------------------------	--------	-----------

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

Create the default IAM role ✕

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

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

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

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

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

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

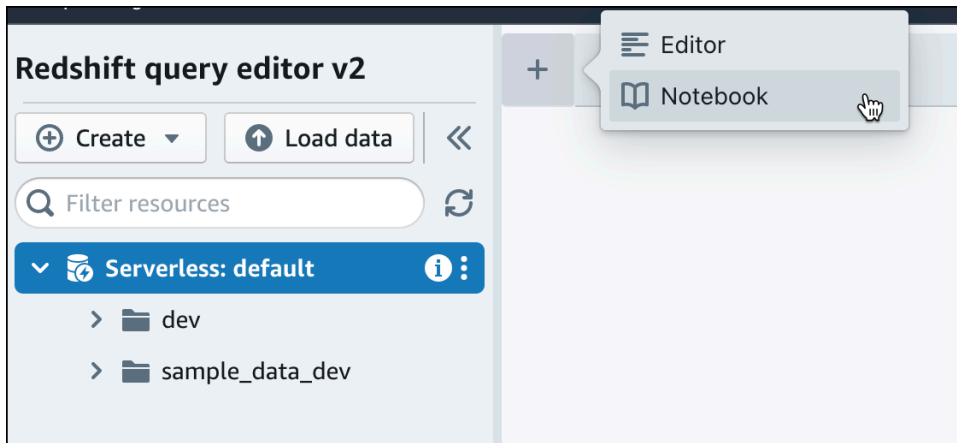
Le fasi seguenti utilizzano i dati all'interno di un bucket S3 pubblico di Amazon Redshift, ma è possibile replicare le stesse fasi utilizzando il proprio bucket S3 e i propri comandi SQL.

Caricamento di dati di esempio da Amazon S3

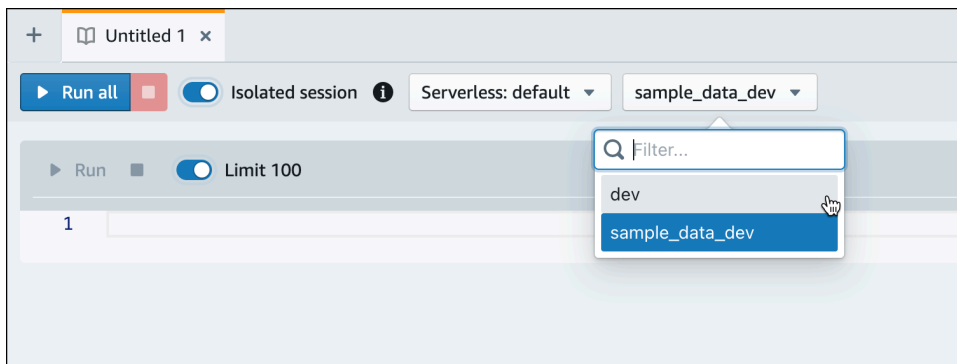
1. Nell'editor di query v2, scegli



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



2. Passaggio al database dev.



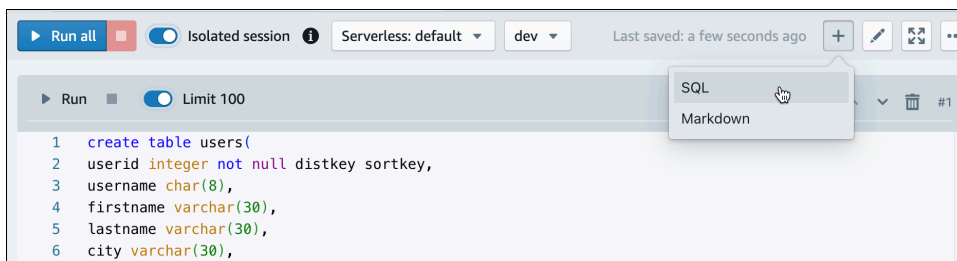
3. Per creare le tabelle.

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

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

```
likeopera boolean,  
likerock boolean,  
likevegas boolean,  
likebroadway boolean,  
likemusicals boolean);  
  
create table event(  
eventid integer not null distkey,  
venueid smallint not null,  
catid smallint not null,  
dateid smallint not null sortkey,  
eventname varchar(200),  
starttime timestamp);  
  
create table sales(  
salesid integer not null,  
listid integer not null distkey,  
sellerid integer not null,  
buyerid integer not null,  
eventid integer not null,  
dateid smallint not null sortkey,  
qtysold smallint not null,  
pricepaid decimal(8,2),  
commission decimal(8,2),  
saletime timestamp);
```

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



5. 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/allevnts_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](#).

Nozioni di base sui data warehouse con provisioning Amazon Redshift

Se utilizzi Amazon Redshift per la prima volta, consigliamo di leggere le seguenti sezioni per iniziare a usare i cluster con provisioning. Il flusso di base di Amazon Redshift consiste nel creare risorse con provisioning, connettersi ad Amazon Redshift, caricare i dati di esempio e quindi eseguire query sui dati. In questa guida puoi scegliere di caricare i dati di esempio da Amazon Redshift o da un bucket Amazon S3. I dati di esempio vengono utilizzati in tutta la documentazione di Amazon Redshift per illustrare le funzionalità.

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

Dopo avere creato la console con provisioning Amazon Redshift e avervi effettuato l'accesso, puoi creare e gestire tutti gli oggetti Amazon Redshift, tra cui cluster, database e nodi. Puoi anche eseguire query e visualizzarle, nonché effettuare altre operazioni DDL e DML SQL con un client SQL.

Important

Il cluster che hai sottoposto a provisioning per questo esercizio viene eseguito in un ambiente reale. Finché è in esecuzione, addebita addebiti al 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 sezione finale di questo capitolo illustra come procedere.

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

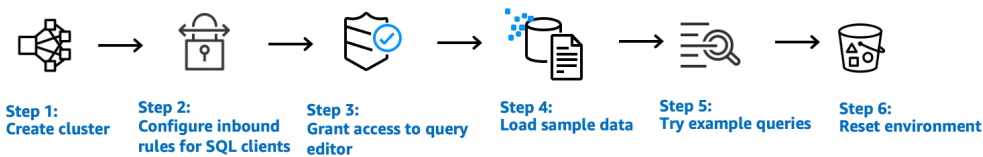
Per iniziare a utilizzare la console Amazon Redshift, consigliamo di procedere innanzitutto con l'accesso al Pannello di controllo dei cluster con provisioning.

A seconda della configurazione, nel riquadro di navigazione della console con provisioning Amazon Redshift sono visualizzati i seguenti elementi:

- Redshift serverless: accedi ai dati e analizzali senza configurare, ottimizzare e gestire i cluster con provisioning Amazon Redshift.
- Dashboard dei cluster forniti: visualizza l'elenco dei cluster presenti Regione AWS, consulta Cluster Metrics e Query overview per informazioni dettagliate sui parametri, sui dati (come l'utilizzo della CPU) e sulle query. L'utilizzo di questi dati può aiutarti a determinare se i dati sulle prestazioni sono anormali in un intervallo di tempo specificato.
- Cluster: visualizza l'elenco dei cluster che contiene Regione AWS, scegli un cluster per avviare le query o esegui azioni relative al cluster. È inoltre possibile creare un nuovo cluster da questa pagina.
- Editor di query: esegui query sui database ospitati nel cluster Amazon Redshift. Consigliamo invece di utilizzare Query Editor V2.
- Query Editor V2: Amazon Redshift Query Editor V2 è un'applicazione client SQL separata basata sul web che consente di creare ed eseguire query sul data warehouse Amazon Redshift. Puoi visualizzare i risultati nei grafici e collaborare condividendo le tue domande con altri membri del tuo team.
- Queries and loads (Query e carichi): ottiene informazioni per riferimento o risoluzione dei problemi, ad esempio un elenco di query recenti e il testo SQL di ogni query.
- Unità di condivisione dati - come amministratore di un account producer puoi autorizzare gli account consumer ad accedere alle unità di condivisione dati o scegliere di non autorizzare alcun accesso. Per utilizzare un datashare autorizzato, un amministratore di account consumer può associare il datashare a un intero o a specifici namespace del cluster in un account. Account AWS Un amministratore può anche rifiutare una unità di condivisione dati.
- Integrazioni Zero-ETL: gestisci le integrazioni che rendono disponibili i dati transazionali in Amazon Redshift dopo che sono stati scritti in origini supportate.
- Connessioni del Centro identità IAM: configura la connessione tra Amazon Redshift e Centro identità IAM.
- Configurazioni (Configurazioni): si connette ai cluster di Amazon Redshift dagli strumenti del client SQL tramite le connessioni Java Database Connectivity (JDBC) e Open Database Connectivity (ODBC). Puoi anche configurare un endpoint VPC (Virtual Private Cloud) di Amazon RedShift. In questo modo viene fornita una connessione privata tra un VPC basato sul servizio Amazon VPC che contiene un cluster e un altro VPC che esegue uno strumento client.
- AWS Integrazione con i partner: crea l'integrazione con un partner supportato. AWS
- Advisor - Ricevi consigli specifici sulle modifiche che puoi apportare al cluster Amazon Redshift per dare priorità alle tue ottimizzazioni.

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

In questo tutorial userai la procedura descritta di seguito:



Argomenti

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

Iscrizione a AWS

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

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

Nel corso della procedura di registrazione riceverai una telefonata o un messaggio di testo e ti verrà chiesto di inserire un codice di verifica attraverso la tastiera del telefono.

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

Definizione delle regole del firewall

Note

Questo tutorial presuppone che il cluster utilizzi la porta predefinita 5439 e che Amazon Redshift Query Editor V2 possa essere usato per eseguire comandi SQL. Non fornisce dettagli sulle configurazioni di rete o sulla configurazione di un client SQL che potrebbe essere necessario nell'ambiente.

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

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

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

Fase 1: Creazione di un cluster Amazon Redshift di esempio


In questo tutorial viene illustrato come creare un cluster Amazon Redshift con un database. Quindi devi caricare un set di dati da Amazon S3 nelle tabelle del database. Puoi usare questo cluster di esempio per valutare il servizio Amazon Redshift.

Prima di iniziare a configurare un cluster Amazon Redshift, assicurati di completare eventuali operazioni preliminari necessarie come [Iscrizione a AWS](#) e [Definizione delle regole del firewall](#).

Per qualsiasi operazione che accede ai dati da un'altra AWS risorsa, il cluster necessita dell'autorizzazione per accedere alla risorsa e ai dati sulla risorsa per tuo conto. Un esempio consiste nell'uso di un comando SQL COPY per caricare i dati da Amazon Simple Storage Service (Amazon S3). Fornisci tali autorizzazioni utilizzando AWS Identity and Access Management (IAM). A questo scopo utilizza un ruolo IAM che hai creato e collegato al cluster. Per ulteriori informazioni sulle credenziali e sulle autorizzazioni di accesso, consulta [Credenziali e autorizzazioni di accesso](#) nella Guida per sviluppatori di database di Amazon Redshift.


Come creare un cluster Amazon Redshift

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

 Important

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

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

 Note

Il passaggio seguente presuppone un Regione AWS che supporti i tipi di RA3 nodi. Per un elenco dei tipi Regioni AWS di RA3 nodi di supporto, consulta [Panoramica dei tipi di RA3 nodi](#) nella Amazon Redshift Management Guide. Per ulteriori informazioni

sulle specifiche del nodo per ogni tipo e dimensione di nodo, consulta [Dettagli sul tipo di nodo](#).

- Se non si conosce la dimensione del cluster, scegliere **Aiutami a scegliere**. In questo modo viene avviato un calcolatore di dimensionamento in cui vengono poste domande sulle dimensioni e sulle caratteristiche delle query dei dati che prevedi di memorizzare nel data warehouse.

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

Per questo tutorial scegli **ra3.4xlarge** per Tipo di nodo e **2** per Numero di nodi.

Se è disponibile un'opzione per Configurazione AZ, scegli **Single-AZ**.

- Per utilizzare il set di dati di esempio fornito da Amazon Redshift, in **Dati di esempio**, scegliere **Carica dati di esempio**. Amazon Redshift carica di default il set di dati di esempio **Tickit** nel database **dev** e nello schema **public**.
5. Nella sezione **Configurazione del database** specifica i valori per **Nome dell'utente amministratore**. In **Password dell'amministratore** seleziona una delle opzioni seguenti:
- **Genera una password**: utilizza una password generata da Amazon Redshift.
 - **Aggiungi manualmente una password dell'amministratore**: utilizza una tua password.
 - **Gestisci le credenziali di amministratore in Gestione dei segreti AWS**: Amazon Redshift le **Gestione dei segreti AWS** utilizza per generare e gestire la password di amministratore. L'utilizzo **Gestione dei segreti AWS** per generare e gestire la password segreta comporta un costo. Per informazioni sui prezzi di **Gestione dei segreti AWS**, consulta [Prezzi di Gestione dei segreti AWS](#).

Per questo tutorial utilizzare i seguenti valori:

- **Nome utente amministratore**: inserisci **awsuser**.
 - **Password dell'utente amministratore**: inserisci **Changeit1** 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 predefinito per un cluster, consulta [Creazione di un ruolo IAM come predefinito per Amazon Redshift](#) nella Guida alla gestione di Amazon Redshift.

7. (Facoltativo) Nella sezione Configurazioni aggiuntive disattiva Usa impostazioni predefinite per modificare le impostazioni Rete e sicurezza, Configurazione del database, Manutenzione, Monitoraggio e Backup.

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

Per rendere il cluster accessibile al pubblico, è possibile eseguire due operazioni. Puoi configurare un indirizzo NAT (Network Address Translation) nel VPC affinché il cluster acceda a Internet. Oppure puoi configurare un endpoint Amazon S3 VPC nel tuo VPC. Per ulteriori informazioni sul routing avanzato di Amazon VPC, consulta [Routing Amazon VPC avanzato](#) nella Guida alla gestione di Amazon Redshift.

8. Scegli Crea cluster. Attendi che il cluster venga creato con lo stato `Available` nella pagina Cluster.

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

Note

Consigliamo di saltare questa fase e accedere al cluster utilizzando Amazon Redshift Query Editor V2.

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

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

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

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

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, uno CIDR/IP di 192.0.2.0/24 consente ai client in quell'intervallo di indirizzi IP di connettersi al cluster. Scoprite quello più adatto al vostro ambiente. CIDR/IP

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

Per eseguire query sui database ospitati dal cluster Amazon Redshift, sono disponibili diverse opzioni per i client SQL. Ciò include:

- Connettiti al cluster ed esegui query utilizzando Amazon Redshift Query Editor V2.

Se utilizzi Query Editor V2, non devi scaricare e configurare un'applicazione client SQL. Avvia Amazon Redshift Query Editor V2 dalla console Amazon Redshift.

- Connettiti al cluster tramite RSQL. Per ulteriori informazioni, consulta [Connessione con Amazon Redshift RSQL](#) nella Guida alla gestione di Amazon Redshift.

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

In questo tutorial viene utilizzato Amazon Redshift Query Editor V2 come soluzione semplice per eseguire query sui database ospitati dal cluster Amazon Redshift. Dopo avere creato il cluster, puoi eseguire immediatamente query. Per informazioni dettagliate sulle considerazioni durante l'utilizzo di Amazon Redshift Query Editor V2, consulta [Considerazioni sull'utilizzo di Query Editor V2](#) 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. Le risorse di Amazon Redshift Query Editor V2 includono le query salvate, i notebook e i grafici. Per impostazione predefinita, una chiave di proprietà AWS viene utilizzata per crittografare le risorse. In alternativa un amministratore può utilizzare una chiave gestita dal cliente scegliendo il nome della risorsa Amazon (ARN) per la chiave nella pagina di configurazione. Dopo aver configurato un account, le impostazioni di AWS KMS crittografia non possono essere modificate. Per ulteriori informazioni, consulta [Configurazione dell' Account AWS](#) nella Guida alla gestione di Amazon Redshift.

Per accedere all'editor della query v2, sono necessarie le opportune autorizzazioni. Un amministratore può collegare una delle policy gestite da AWS per Amazon Redshift Query Editor V2 al ruolo IAM o all'utente per concedere le autorizzazioni. Queste politiche AWS gestite sono scritte con diverse opzioni che controllano il modo in cui l'etichettatura delle risorse consente la condivisione delle query. Puoi utilizzare la console IAM (<https://console.aws.amazon.com/iam/>) per allegare le policy IAM. Per ulteriori informazioni su queste policy, consulta [Accesso a Query Editor V2](#) nella Guida alla gestione di Amazon Redshift.

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

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


```
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,  
qtsold smallint not null,  
pricepaid decimal(8,2),  
commission decimal(8,2),  
saletime timestamp);
```

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

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

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

Note

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

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

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

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

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

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

Caricamento di dati da Amazon S3 con Query Editor V2

In questa sezione viene illustrato il caricamento di dati in un cluster Amazon Redshift. L'uso di Query Editor V2 semplifica il caricamento di dati quando utilizzi la procedura guidata Carica dati. Il comando COPY generato e usato in Query Editor V2 della procedura guidata Carica dati supporta molti dei 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.

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

Per caricare i dati personali da Amazon S3 in Amazon Redshift, puoi utilizzare la procedura guidata Carica dati di Query Editor V2. Per ulteriori informazioni su come usare la procedura guidata Carica dati, consulta [Caricamento di dati da Amazon S3](#) nella Guida alla gestione di Amazon Redshift.

Creazione di dati TICKIT nel cluster

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

- Quando crei un cluster nella console Amazon Redshift, hai la possibilità di caricare contemporaneamente i dati di esempio TICKIT. Sulla console Amazon Redshift scegli Cluster, Crea cluster. Nella sezione Dati campione seleziona Caricamento dati campione. Amazon Redshift carica il set di dati di esempio nel database dev del cluster Amazon Redshift durante la creazione del cluster.
- Per connetterti a un cluster esistente, esegui le operazioni seguenti:
 - Nella barra di navigazione della console Amazon Redshift scegli Cluster.
 - Scegli il cluster dal riquadro Cluster.
 - Scegli Esegui query sui dati, Esegui query in Query Editor V2.
 - Espandi `examplecluster` nell'elenco delle risorse. Se è la prima volta che ti connetti al cluster, viene visualizzato Connettiti a `examplecluster`. Scegli Nome utente e password del database. Lascia invariato il database **dev**. Specifica **awsuser** per il nome utente e **Changeit1** per la password.
 - Scegli Crea connessione.
- Con Amazon Redshift Query Editor V2 puoi caricare i dati TICKIT in un database di esempio denominato `sample_data_dev`. Scegli il database `sample_data_dev` nell'elenco delle risorse. Accanto al nodo `tickit` scegli l'icona Apri notebook di esempio. Conferma di voler creare il database di esempio.
- Amazon Redshift Query Editor v2 crea il database di esempio insieme a un notebook di esempio denominato `tickit-sample-notebook`. Puoi scegliere Esegui tutto per eseguire questo notebook per effettuare query sui dati nel database di esempio.

Per visualizzare i dettagli dei dati TICKIT, consulta [Database di esempio](#) nella Guida per sviluppatori di database di Amazon Redshift.

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

Per configurare e utilizzare Amazon Redshift Query Editor V2 per eseguire query su un database, consulta [Utilizzo di Query Editor V2](#) nella Guida alla gestione di Amazon Redshift.

Ora, provare alcune query di esempio, come illustrato di seguito. Per creare nuove query in Query Editor V2, scegli l'icona + nell'angolo in alto a destra del riquadro delle query e scegli SQL. Viene visualizzata una nuova pagina di query in cui puoi copiare e incollare le seguenti query SQL.

Note

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

```
set search_path to ticket;
```

Per ulteriori informazioni sull'utilizzo del comando `SELECT`, consulta [SELECT](#) nella Guida per 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;
```

Fase 6: reimpostazione dell'ambiente

Nei passaggi precedenti hai creato un cluster Amazon Redshift, caricato i dati nelle tabelle ed eseguito query sui dati utilizzando un client SQL come Amazon Redshift Query Editor V2.

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

Tuttavia, se desideri provare a eseguire le attività descritte in altre guide di Amazon Redshift o le attività descritte in [Eseguire comandi per definire e utilizzare un database nel data warehouse](#), consigliamo di tenere il cluster di esempio in esecuzione.

Come eliminare un cluster

1. Accedi Console di gestione AWS e apri la console Amazon Redshift all'indirizzo. <https://console.aws.amazon.com/redshiftv2/>
2. Dal menu di navigazione, scegliere Clusters (Cluster) per visualizzare l'elenco dei cluster.
3. Scegli il cluster `examplecluster`. In Actions (Azioni), scegliere Delete (Elimina). Viene visualizzata la pagina Delete `examplecluster`?
4. Conferma il cluster da eliminare, deseleziona l'impostazione Crea snapshot finale, quindi inserisci **delete** per confermare l'eliminazione. Scegli Delete cluster (Elimina cluster).

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

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

Eseguire comandi per definire e utilizzare un database nel data warehouse

Sia i data warehouse Redshift serverless che i data warehouse con provisioning Amazon Redshift contengono database. Dopo avere avviato il data warehouse, puoi gestire la maggior parte delle azioni del database utilizzando i comandi SQL. Salvo qualche eccezione, la funzionalità e la sintassi di SQL sono le stesse per tutti i database Amazon Redshift. Per i dettagli dei comandi SQL disponibili con Amazon Redshift, consulta [Comandi SQL](#) nella Guida per sviluppatori di database di Amazon Redshift.

Quando crei un data warehouse, nella maggior parte degli scenari, Amazon Redshift crea anche il database dev predefinito. Dopo la connessione al database dev puoi creare un nuovo database.

Nelle sezioni seguenti vengono illustrate le attività comuni dei database quando utilizzi i database Amazon Redshift. Le attività iniziano con la creazione di un database e, se continui fino all'ultima attività, puoi eliminare tutte le risorse create rimuovendo il database.

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

- Hai creato un data warehouse Amazon Redshift.
- Hai stabilito una connessione al data warehouse dallo strumento del client SQL, come Amazon Redshift Query Editor V2. Per ulteriori informazioni su Query Editor V2, consulta [Esecuzione di query su un database con Amazon Redshift Query Editor V2](#) nella Guida alla gestione di Amazon Redshift.

Argomenti

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

Connessione ai data warehouse Amazon Redshift

Per connetterti ai cluster Amazon Redshift, nella pagina Cluster della console Amazon Redshift espandi Connettiti ai cluster Amazon Redshift ed effettua una delle seguenti operazioni:

- Scegli Esegui query sui dati per utilizzare Query Editor 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, consulta [Esecuzione di query su un database con Amazon Redshift Query Editor V2](#) nella Guida alla gestione di Amazon Redshift.

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

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

- Se lo strumento del client SQL richiede un driver, puoi scegliere il driver JDBC oppure ODBC per 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, vedere [Configurazione di una connessione del driver JDBC versione 2.2](#).

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

Per connetterti al data warehouse Redshift serverless, dalla pagina Pannello di controllo serverless della console Amazon Redshift, esegui una delle seguenti operazioni:

- Utilizza Amazon Redshift Query Editor V2 per eseguire query sui database ospitati dal data warehouse Redshift serverless. Dopo avere creato il data warehouse, puoi eseguire immediatamente le query utilizzando Query Editor 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 data warehouse, hai bisogno dei driver JDBC oppure ODBC per stabilire la connettività dal computer o dall'istanza client. Codifica le tue applicazioni affinché usino operazioni API di accesso ai dati JDBC o ODBC e gli strumenti del client SQL che supportano JDBC o ODBC.

Per ulteriori informazioni su come trovare la stringa di connessione, consulta [Connessione a Redshift serverless](#) nella Guida alla gestione di Amazon Redshift.

Creazione di un database

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

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

```
CREATE DATABASE salesdb;
```

Note

Dopo avere eseguito il comando, assicurati di aggiornare l'elenco degli oggetti degli strumenti del client SQL nel data warehouse per visualizzare il nuovo salesdb.

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

Creazione di un utente

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

Creare uno schema

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

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

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

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

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

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

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

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

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

```
select * from pg_namespace;
```

L'output visualizzato dovrebbe essere simile al seguente:

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

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

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

Nell'esempio seguente viene concesso il privilegio all'utente GUEST per selezionare i dati da tutte le tabelle o viste nello schema SALES 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;
```

Creare una tabella

Dopo aver creato il nuovo database, è necessario creare le tabelle che conterranno i dati. Quando crei la tabella, specifica 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)  
);
```

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

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

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

Per visualizzare ed esaminare gli schemi e le relative tabelle, puoi utilizzare Amazon Redshift Query Editor V2. Oppure è possibile visualizzare l'elenco delle tabelle negli schemi usando le viste di sistema. Per ulteriori informazioni, consulta [Eseguire query sulle tabelle e sulle viste di sistema](#).

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

Inserisci le righe di dati in una tabella

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

Note

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

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

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

Per inserire dati in una tabella che si trova in uno schema specifico, esegui il comando seguente.

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

Selezione dati da una tabella

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

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

```
SELECT * from DEMO;
```

Il risultato sarà simile al seguente:

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

```
(2 rows)
```

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

Caricare i dati

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. Quindi puoi caricare i dati di esempio nel 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](#).

Eseguire query sulle tabelle e sulle viste di sistema

Oltre alle tabelle che hai creato, il data warehouse contiene numerose tabelle e viste di sistema. Queste tabelle e viste di sistema contengono informazioni sull'installazione e sulle varie query e sui vari processi in esecuzione sul sistema. Puoi eseguire query su queste tabelle e viste di sistema per raccogliere informazioni sul database. Per ulteriori informazioni, consulta [Riferimento di tabelle e viste di sistema](#) nella Guida per sviluppatori di database di Amazon Redshift. La descrizione per ogni tabella o vista 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.

Visualizza un elenco di nomi di tabella

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

```
SHOW search_path;
```

Il risultato dovrebbe essere simile al seguente,

```
search_path
```

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

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

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

```
SHOW search_path;
```

```
search_path
```

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

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

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

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

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

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

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

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

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

Puoi inoltre utilizzare Amazon Redshift Query Editor V2 per visualizzare tutte le tabelle in uno schema specificato scegliendo innanzitutto un database a cui connetterti.

Visualizzazione degli utenti

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

```
SELECT * FROM pg_user;
```

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

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

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

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

Visualizza le query recenti

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

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

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

Il risultato è simile al seguente.

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

Determinare l'ID sessione di una query in esecuzione

Per recuperare le informazioni sulle tabelle di sistema, potresti dover specificare l'ID sessione (ID processo) associato alla query. Potresti dover trovare l'ID sessione per una query ancora in fase di esecuzione. Ad esempio, hai bisogno dell'ID sessione se devi annullare una query la cui esecuzione dura troppo a lungo su un cluster con provisioning. È possibile interrogare la tabella di sistema `STV_RECENTS` per ottenere un elenco di sessioni IDs per l'esecuzione delle query, insieme alla stringa di query corrispondente. Se la query restituisce più ID sessione, puoi esaminare il testo della query per determinare l'ID sessione necessario.

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

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

Annullare una query

Se esegui una query che impiega troppo tempo o utilizza un numero eccessivo di risorse, annullala. 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, utilizza il comando CANCEL con l'ID sessione della query. Con Amazon Redshift Query Editor V2 puoi annullare una query scegliendo il pulsante Annulla mentre la query è in esecuzione.

Per trovare l'ID sessione, avvia una nuova sessione ed esegui query sulla tabella STV_RECENTS, come illustrato 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.

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

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

Il risultato è simile al seguente.

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

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

```
CANCEL 1073791534;
```

Note

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

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

```
ABORT;
```

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

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

Per ulteriori informazioni sull'annullamento di una query, consulta [CANCEL](#) nella Guida per sviluppatori di database di Amazon Redshift.

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

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

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

Per annullare una query utilizzando la coda dell'utente con privilegi avanzati, esegui questi comandi.

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

Eseguire query sui dati non presenti nel database Amazon Redshift

Di seguito sono riportate informazioni su come iniziare a eseguire query sui dati di origini remote, tra cui i dati Amazon S3, i gestori di database remoti, i database Amazon Redshift remoti e i modelli di machine learning (ML) con Amazon Redshift.

Argomenti

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

Esecuzione di query nel data lake

È possibile usare Amazon Redshift Spectrum per eseguire query sui dati nei file Amazon S3 senza doverli caricare nelle tabelle Amazon Redshift. Amazon Redshift fornisce funzionalità SQL progettate per l'elaborazione analitica online veloce (OLAP) di dataset di grandi dimensioni archiviati sia nei cluster Amazon Redshift che nei data lake Amazon S3. È possibile 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 gestori di database remoti

Puoi effettuare il join di dati di un database Amazon RDS e un database Amazon Aurora 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 database 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, consulta [Gestione delle attività di condivisione dati](#) nella Guida per 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 puoi utilizzare i dati nei cluster Amazon Redshift per addestrare i modelli con Amazon SageMaker AI Autopilot e ottenere automaticamente il modello migliore. Puoi quindi localizzare i modelli e fare previsioni all'interno di un database Amazon Redshift.

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

Scoprire i concetti di Amazon Redshift

Amazon Redshift serverless consente di accedere e analizzare i dati senza le configurazioni di un data warehouse con provisioning. Viene eseguito automaticamente il provisioning delle risorse e la capacità del data warehouse viene dimensionata in modo intelligente per fornire prestazioni rapide per carichi di lavoro maggiormente impegnativi e imprevedibili. Quando il data warehouse è inattivo non vengono addebitati costi, si paga solo l'utilizzo. Puoi caricare i dati e iniziare subito a eseguire query nell'editor di query Amazon Redshift v2 o nello strumento di business intelligence (BI) preferito. 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.
- [Nozioni di base sui data warehouse Amazon Redshift serverless](#) in questo argomento puoi scoprire ulteriori informazioni su come creare un data warehouse Amazon Redshift serverless e iniziare a eseguire query sui dati utilizzando Query Editor V2.

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

Se la 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. 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 Regioni AWS in cui Amazon Redshift serverless è disponibile, consulta gli endpoint Amazon Redshift 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.

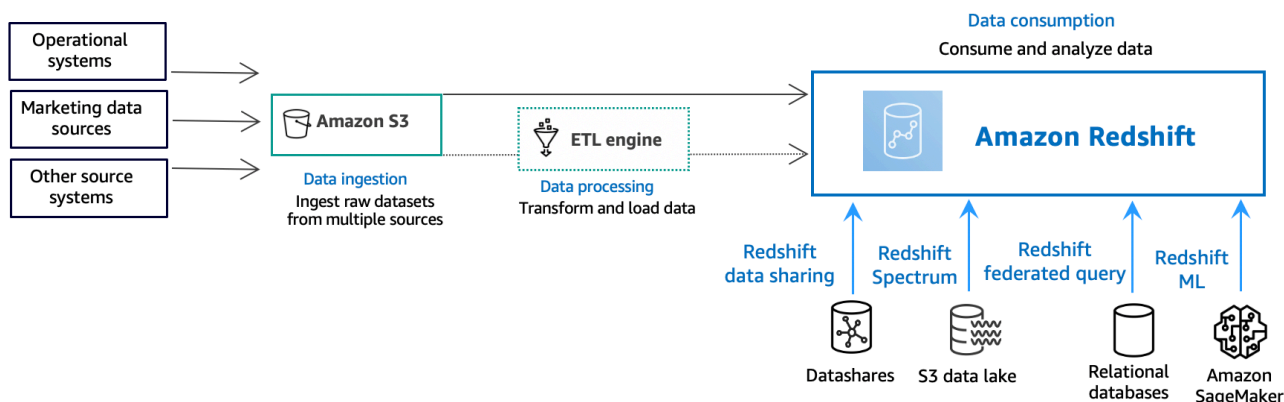
- Database – Un cluster contiene uno o più database.

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.

A livello di importazione dei dati, diversi tipi di origini dati caricano continuamente dati strutturati, semistrutturati o non strutturati sul livello di archiviazione di 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 saperne di più su Amazon Redshift.

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

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

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

Cronologia dei documenti

Note

Per una descrizione delle nuove funzionalità di Amazon Redshift, consulta [Novità](#).

Nella tabella seguente vengono descritte le importanti modifiche apportate alla Guida alle operazioni di base di Amazon Redshift.

Modifica	Description	Data di rilascio
Aggiornamento della documentazione	La guida è stata aggiornata per riflettere le modifiche alle policy gestite da Query Editor V2 e il miglioramento delle autorizzazioni di accesso ai gruppi di lavoro e ai namespace serverless.	21 febbraio 2024
Aggiornamento della documentazione	Gli screenshot e le procedure sono state aggiornate e per riflettere gli ultimi miglioramenti dell'interfaccia della console e i miglioramenti di Query Editor V2.	11 marzo 2023
Nuova funzionalità	La guida è stata aggiornata per includere le procedure e i flussi di lavoro introduttivi di Amazon Redshift serverless. È stata aggiunta una sezione completa sulla creazione e sulla gestione di data warehouse serverless.	12 luglio 2022
Aggiornamento della documentazione	La guida è stata aggiornata per includere Query Editor V2 come interfaccia principale per l'esecuzione di query, sostituendo i riferimenti all'editor di query precedente.	febbraio 2022
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	30 giugno 2021

Modifica	Description	Data di rilascio
	addestrare modelli di machine learning con i dati Amazon Redshift.	
Nuova funzionalità	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 funzionalità	Guida aggiornata per la descrizione della procedura di avvio rapido del cluster.	10 agosto 2018
Nuova funzionalità	La guida è stata aggiornata per l'avvio dei cluster dal pannello di controllo di Amazon Redshift.	28 luglio 2015
Nuova funzionalità	Guida aggiornata per l'utilizzo dei nuovi nomi dei tipi di nodo.	9 giugno 2015
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

Modifica	Description	Data di rilascio
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