



入門指南

# Amazon Redshift



# Amazon Redshift: 入門指南

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon 的商標和商業外觀不得用於任何非 Amazon 的產品或服務，也不能以任何可能造成客戶混淆、任何貶低或使 Amazon 名譽受損的方式使用 Amazon 的商標和商業外觀。所有其他非 Amazon 擁有的商標均為其各自擁有者的財產，這些擁有者可能附屬於 Amazon，或與 Amazon 有合作關係，亦或受到 Amazon 贊助。

# Table of Contents

Amazon Redshift Serverless .....	1
註冊成為 AWS .....	1
使用 Amazon Redshift Serverless 建立資料倉儲 .....	1
載入範例資料 .....	4
執行範例查詢 .....	7
從 Amazon S3 載入資料 .....	9
查詢外部資料來源 .....	16
查詢資料湖 .....	16
查詢遠端資料來源 .....	17
存取其他叢集中的資料 .....	17
使用 Redshift 資料訓練機器學習模型 .....	17
Amazon Redshift 佈建叢集 .....	19
Amazon Redshift 佈建的叢集主控台 .....	19
註冊成為 AWS .....	19
決定防火牆規則 .....	19
連線至 Amazon Redshift .....	21
Amazon Redshift 叢集和資料載入 .....	22
使用範例資料集 .....	22
將您自己的資料引入 Amazon Redshift .....	26
一般資料庫任務 .....	39
任務 1：建立資料庫 .....	39
任務 2：建立使用者 .....	40
任務 3：建立結構描述 .....	40
任務 4：建立資料表 .....	41
任務 5：載入範例資料 .....	43
任務 6：查詢系統資料表 .....	43
任務 7：取消查詢 .....	47
任務 8：清除您的資源 .....	50
Amazon Redshift 概念概觀 .....	51
其他資源 .....	54
文件歷史紀錄 .....	55
.....	lvii

# Amazon Redshift Serverless

若您是 Amazon Redshift Serverless 新手，建議您閱讀以下章節，以協助您開始使用 Amazon Redshift Serverless。Amazon Redshift Serverless 的基本流程是建立無伺服器資源、連接到 Amazon Redshift Serverless、載入範例資料，然後對資料執行查詢。在本指南中，您可以選擇從 Amazon Redshift Serverless 或從 Amazon S3 儲存貯體載入範例資料。

- [the section called “註冊成為 AWS”](#)
- [the section called “使用 Amazon Redshift Serverless 建立資料倉儲”](#)
- [the section called “從 Amazon S3 載入資料”](#)

## 註冊成為 AWS

如果您還沒有 AWS 帳戶，請註冊一個帳戶。如果您已有帳戶，則可略過此事前準備，並使用現有的帳戶。

1. 開啟 <https://portal.aws.amazon.com/billing/signup>。
2. 請遵循線上指示進行。

當您註冊 AWS 帳號時，會建立 AWS 帳號根使用者。root 使用者可以存取帳號中的所有 AWS 服務和資源。作為安全最佳實務，[將管理存取權指派給管理使用者](#)，並且僅使用根使用者來執行[需要根使用者存取權的任務](#)。

## 使用 Amazon Redshift Serverless 建立資料倉儲

第一次登入 Amazon Redshift Serverless 主控台時，系統會提示您存取入門體驗，您可以使用該體驗來建立和管理無伺服器資源。在本指南中，您將使用 Amazon Redshift Serverless 的預設設定來建立無伺服器資源。

若要更精細地控制您的設定，請選擇自訂設定。

若要使用預設設定進行設定：

1. 登入 AWS Management Console 並開啟 Amazon Redshift 主控台，網址為 <https://console.aws.amazon.com/redshiftv2/>。

選擇試用 Amazon Redshift Serverless。

2. 在組態下選擇使用預設設定。Amazon Redshift Serverless 會使用與此命名空間相關聯的預設工作群組建立預設命名空間。選擇 Save configuration (儲存組態)。

下列螢幕擷取畫面顯示 Amazon Redshift Serverless 的預設設定。

# Get started with Amazon Redshift Serverless

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

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

## Configuration [Info](#)

### Use default settings

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

### Customize settings

Customize your settings for your specific needs.

## ▼ How it works



### Using the default settings

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



### Customizing the settings

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

## Namespace [Info](#)

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

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

Namespace name

default

### Database name and password

Database name

dev

Admin user credentials

Created based on IAM credentials

### Permissions

使用 Amazon Redshift Serverless 建立資料倉儲

Default IAM role

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

- 設定完成後，選擇繼續以前往您的 Serverless 儀表板。您可以看到無伺服器工作群組和命名空間可用。

The screenshot shows the 'Serverless dashboard' with a 'Namespace overview' section displaying four metrics: Total snapshots (0), Datashares in my account (0), Datashares requiring authorization (0), and Datashares fr (0). Below this is a table for 'Namespaces / Workgroups' with two columns: Namespace and Status. The table shows 'default' namespace with 'Available' status.

Namespace	Status	Workgroup	Status
default	Available	default	Available

## 載入範例資料

現在您已經使用 Amazon Redshift Serverless 設定資料倉儲，您可以使用 Amazon Redshift 查詢編輯器 v2 載入範例資料。

- 若要從 Amazon Redshift Serverless 主控台啟動查詢編輯器 v2，請選擇查詢資料。當您從 Amazon Redshift Serverless 主控台調用查詢編輯器 v2 時，新的瀏覽器索引標籤會開啟並顯示查詢編輯器。查詢編輯器 v2 會從您的用戶端機器連線到 Amazon Redshift Serverless 環境。

The screenshot shows the 'Serverless dashboard' with a 'Query data' button highlighted by a red circle. The button is located next to a refresh icon and a 'Create workgroup' button.

- 如果您是第一次啟動查詢編輯器 v2，則必須先配置 AWS KMS 加密，然後才能繼續。您亦可選擇將 URI 指定至 S3 儲存貯體，以便稍後載入資料。完成此操作後，選擇設定帳戶。

## Configure account

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


### AWS KMS encryption

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

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

Customize encryption settings (advanced)

### S3 bucket - optional

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

S3 URI

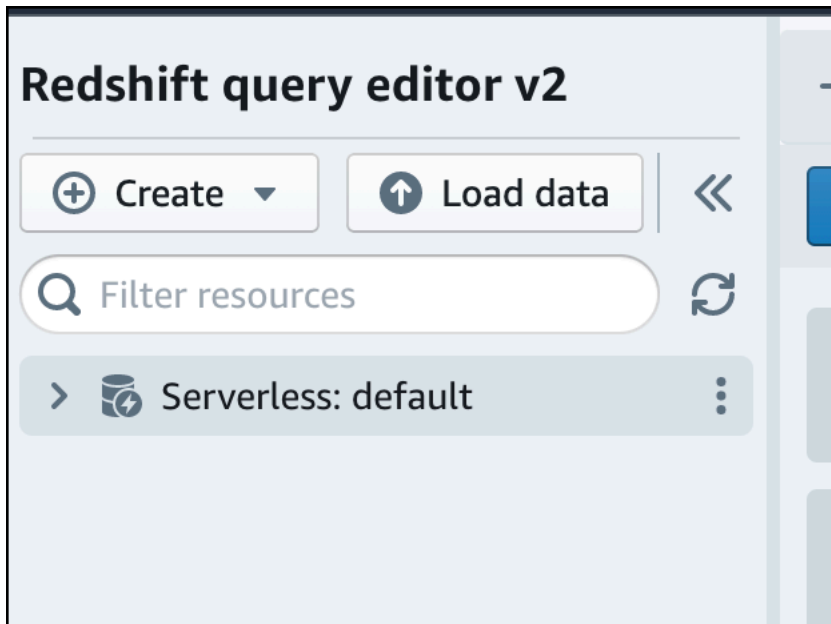
Cancel

Configure account

若要進一步了解如何設定查詢編輯器 v2 (包括需要哪些許可)，請參閱《Amazon Redshift 管理指南》中的[設定您的 AWS 帳戶](#)。

- 若要連接至工作群組，請在樹狀檢視面板中選擇工作群組名稱。





- 第一次在查詢編輯器 v2 中連線至新工作群組時，您必須選取要用於連線至工作群組的驗證類型。在本指南中，請保持聯合身分使用者為選取狀態，然後選擇建立連線。

### Connect to default ✕

Authentication [Learn more](#)

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

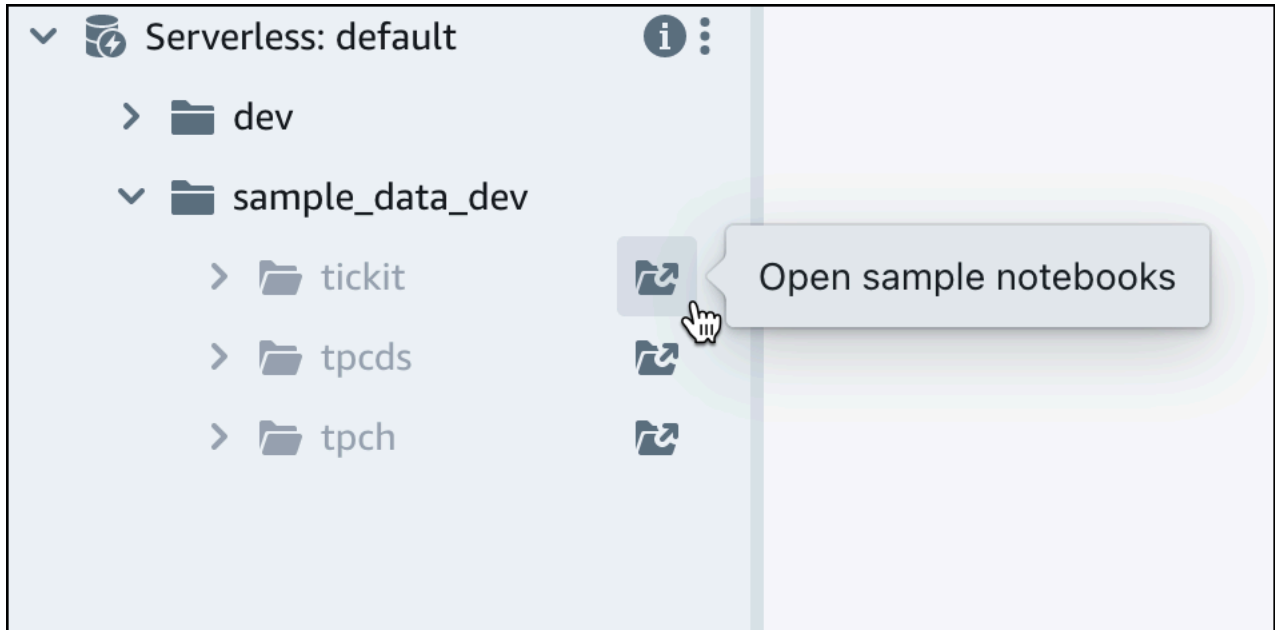
Database user name and password  
Provide a database user and password for the database that you are connecting to. The query editor v2 stores your credentials in AWS Secrets Manager on your behalf.

Database

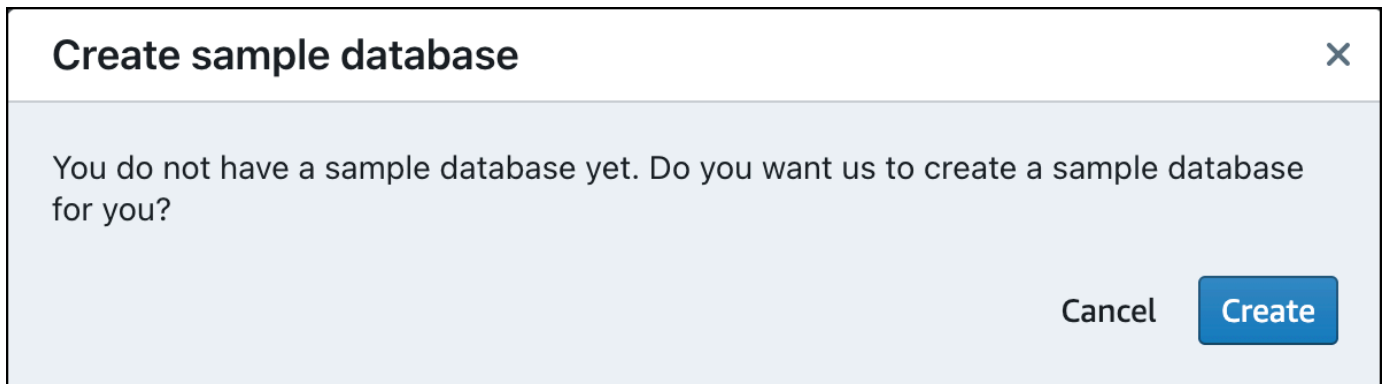
The database name must be 1-64 characters. Valid characters are lowercase alphanumeric characters.

連線之後，您可以選擇從 Amazon Redshift Serverless 或從 Amazon S3 儲存貯體載入範例資料。

5. 在 Amazon Redshift Serverless 預設工作群組下，展開 `sample_data_dev` 資料庫。有三個範例結構描述與三個範例資料集相對應，您可以將這些範例資料集載入 Amazon Redshift Serverless 資料庫。選擇您要載入的範例資料集，然後選擇開啟範例筆記本。



6. 第一次載入資料時，查詢編輯器 v2 會提示您建立範例資料庫。選擇建立。



## 執行範例查詢

設定 Amazon Redshift Serverless 之後，您就可以開始在 Amazon Redshift Serverless 中使用範例資料集。Amazon Redshift Serverless 會自動載入範例資料集 (例如 tickit 資料集)，而且您可以立即查詢資料。

- Amazon Redshift Serverless 完成範例資料的載入後，所有範例查詢都會載入到編輯器中。您可以選擇全部執行，從範例筆記本執行所有查詢。

The screenshot shows a SQL query in the Amazon Redshift console. The query is as follows:

```

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

```

The results are displayed in a table with two columns: eventname and total\_price. The data is as follows:

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

At the bottom right of the console, it shows "Elapsed time: 401 ms" and "Total rows: 9".

您也可以將結果匯出為 JSON 或 CSV 檔案，或在圖表中檢視結果。

The screenshot shows the same SQL query in the Amazon Redshift console. The 'Export' and 'Chart' options are highlighted with a red circle. The 'Chart' option is currently selected, and a chart is visible on the right side of the console. The chart is a line graph with a blue line that starts at the bottom left and goes up to the top right. The Y-axis is labeled 'Click to enter Y axis title' and has a scale from 2 to 8. The X-axis is labeled 'Click to enter X axis title'. The chart area is currently empty, with the text 'Click to enter Plot title' displayed above it.

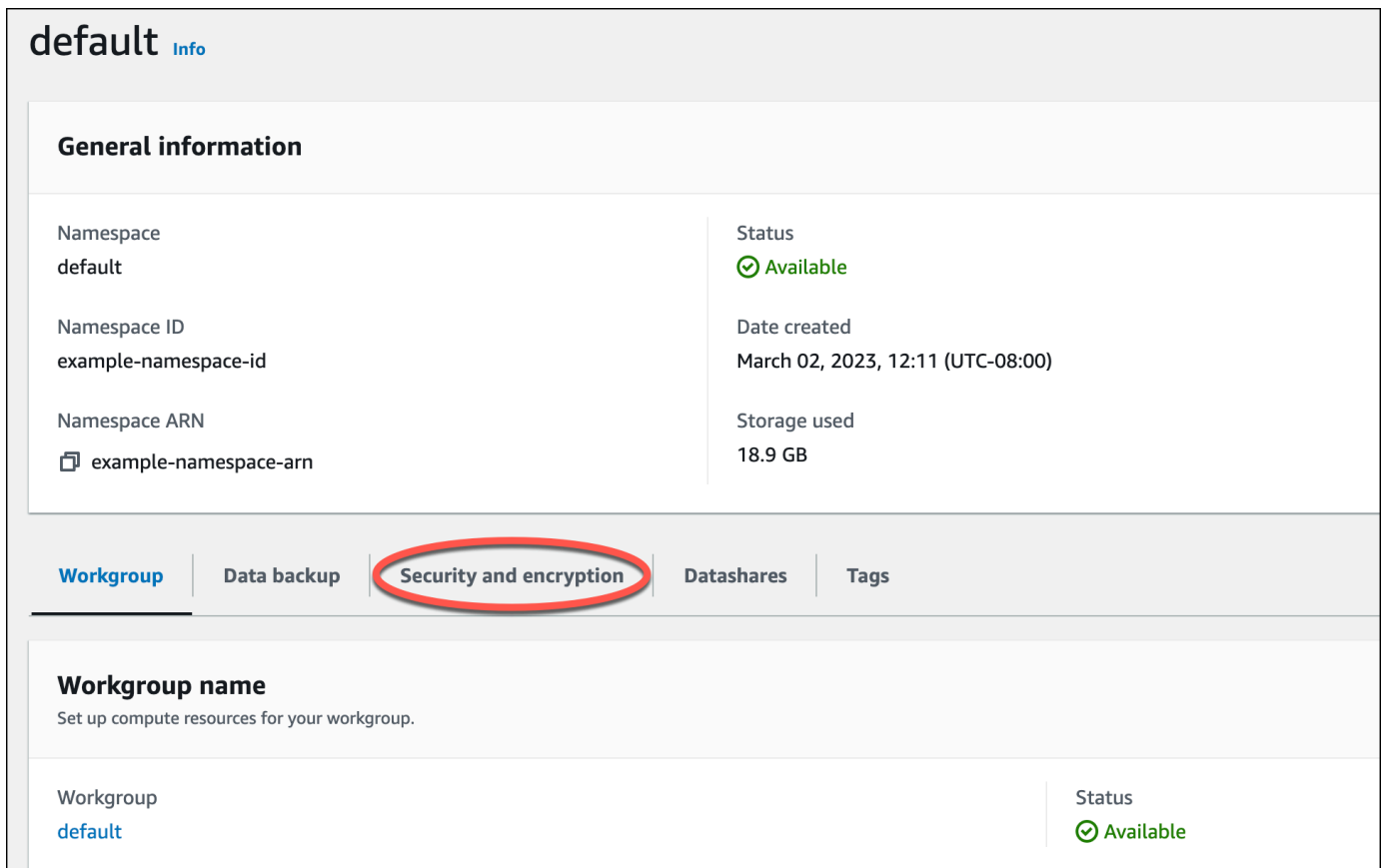
您亦可從 Amazon S3 儲存貯體載入資料。如需進一步了解，請參閱[the section called “從 Amazon S3 載入資料”](#)。

# 從 Amazon S3 載入資料

建立資料倉儲之後，您可以從 Amazon S3 載入資料。

此時，您已經有了一個名為 dev 的資料庫。接下來，您會在資料庫中建立一些資料表、將資料上傳至資料表，然後嘗試執行查詢。為方便起見，在 Amazon S3 儲存貯體中會提供您載入的範例資料。

1. 您必須先建立具有必要權限的 IAM 角色，然後將其附加至無伺服器命名空間，才能從 Amazon S3 載入資料。若要這麼做，請從瀏覽功能表中選擇命名空間組態，然後選擇安全和加密。選擇管理 IAM 角色。



The screenshot displays the configuration page for the 'default' namespace in Amazon Redshift. The page is titled 'default' with an 'Info' link. It features a 'General information' section with the following details:

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

Below the general information, there is a navigation bar with tabs: 'Workgroup', 'Data backup', 'Security and encryption', 'Datashares', and 'Tags'. The 'Security and encryption' tab is circled in red. Underneath, the 'Workgroup name' section is visible, showing the workgroup name as 'default' and its status as 'Available'.

2. 展開管理 IAM 角色功能表，然後選擇建立 IAM 角色。

# Manage IAM roles

## Permissions

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

## Associated IAM roles (1)

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

**Set default** ▼ **Manage IAM roles** ▲

🔍 Search for associated IAM roles or role type

Associate IAM roles  
Create IAM role  
Remove IAM roles

< 1 >

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

3. 選擇您要授與此角色的 S3 儲存貯體存取層級，然後選擇建立 IAM 角色做為預設值。

## 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. 選擇儲存變更。您現在可以從 Amazon S3 載入範例資料。

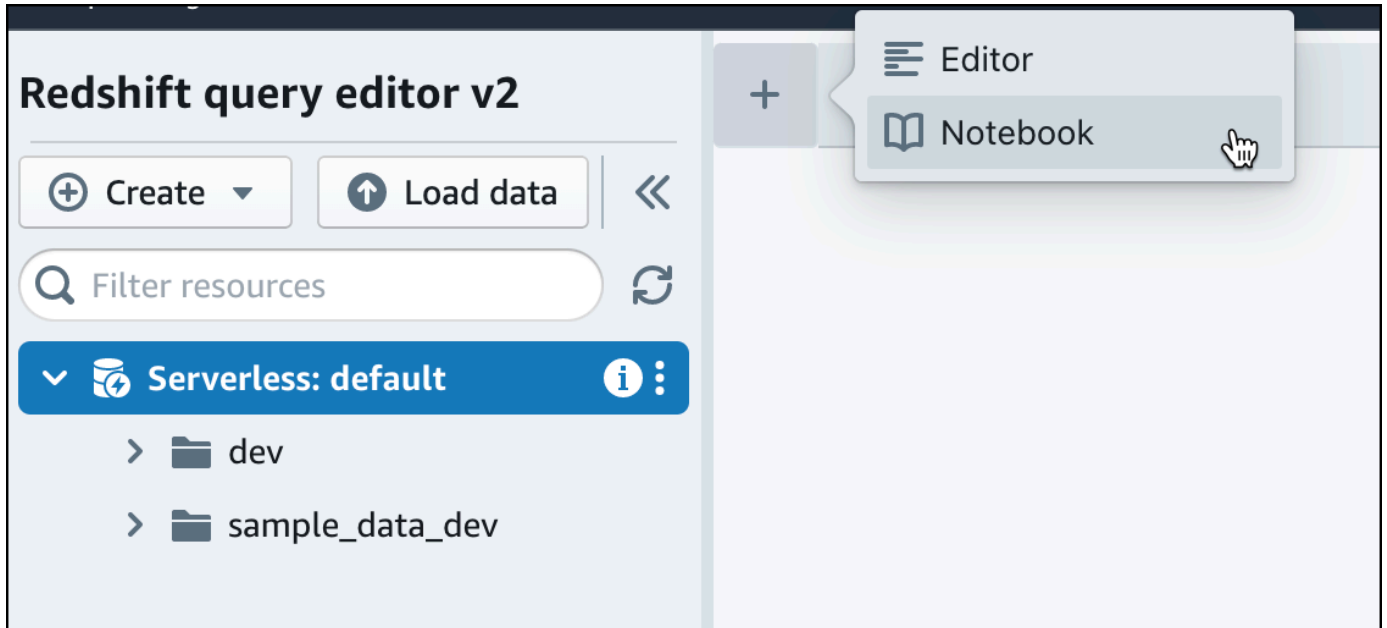
下列步驟會使用公有 Amazon Redshift S3 儲存貯體內的資料，但是您可以使用自己的 S3 儲存貯體和 SQL 命令複寫相同的步驟。

## 從 Amazon S3 載入範例資料

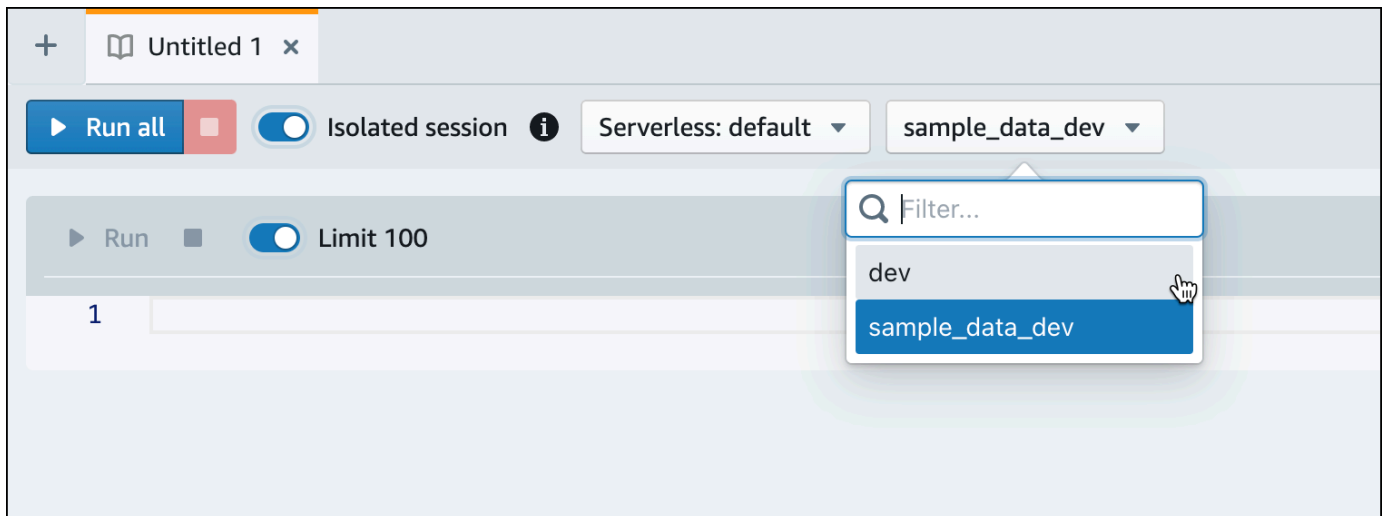
1. 在查詢編輯器 v2 中，選擇



[新增]，然後選擇筆記本以建立新的 SQL 筆記本。



2. 切換到 dev 資料庫。



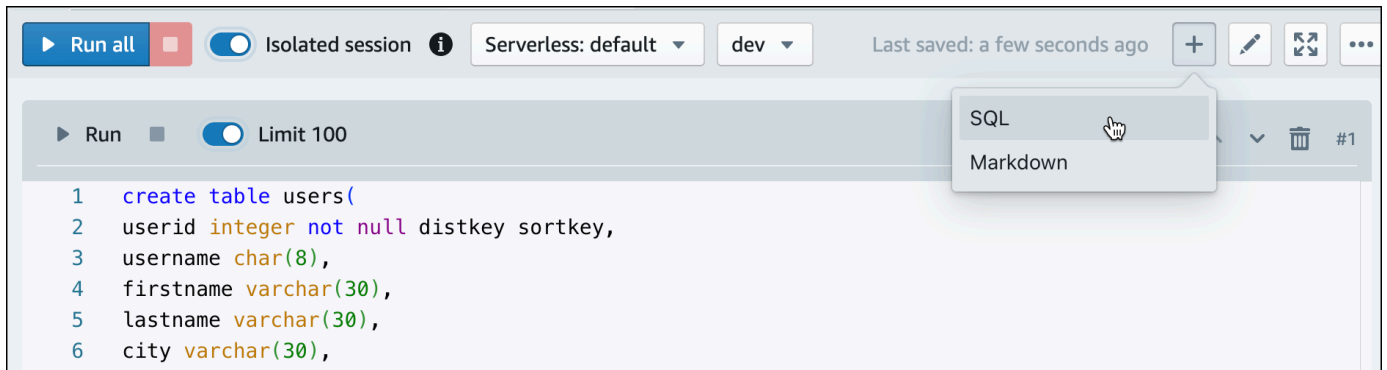
3. 建立資料表。

如果您使用查詢編輯器 v2，請複製並執行以下建立資料表陳述式，以在 dev 資料庫中建立資料表。如需語法的相關資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [CREATE TABLE](#)。

```
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. 在查詢編輯器 v2 中，在筆記本中建立新的 SQL 儲存格。





5. 現在，使用查詢編輯器 v2 中的 COPY 命令，將大型資料集從 Amazon S3 或 Amazon DynamoDB 載入到 Amazon Redshift。如需 COPY 語法的相關資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [COPY](#)。

您可以使用公有 S3 儲存貯體中提供的一些範例資料來執行 COPY 命令。在查詢編輯器 v2 中執行下列 SQL 命令。

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

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

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

6. 載入資料後，請在筆記本中建立另一個 SQL 儲存格，並嘗試一些範例查詢。如需使用 SELECT 命令的相關資訊，請參閱《Amazon Redshift 開發人員指南》中的 [SELECT](#)。若要瞭解範例資料的結構和結構描述，請使用查詢編輯器 v2 進行探索。

```
-- 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;
```

現在您已載入資料並執行了一些範例查詢，您可以探索 Amazon Redshift Serverless 的其他區域。請參閱下列清單，進一步了解如何使用 Amazon Redshift Serverless。

- 您可以從 Amazon S3 儲存貯體載入資料。如需詳細資訊，請參閱[從 Amazon S3 載入資料](#)。
- 您可以使用查詢編輯器 v2，從小於 5 MB 的本機字元分隔檔案載入資料。如需詳細資訊，請參閱[從本機檔案載入資料](#)。
- 您可以使用第三方 SQL 工具使用 JDBC 和 ODBC 驅動程式連線到 Amazon Redshift Serverless。如需詳細資訊，請參閱[連線至 Amazon Redshift Serverless](#)。
- 您也可以使用 Amazon Redshift 資料 API 連接到 Amazon Redshift Serverless。如需詳細資訊，請參閱[使用 Amazon Redshift 資料 API](#)。
- 您可以將 Amazon Redshift Serverless 中的資料與 Redshift ML 搭配使用，透過 CREATE MODEL 命令建立機器學習模型。請參閱[教學課程：建立客戶流失模型](#)，以了解如何建立 Redshift ML 模型。
- 您可以從 Amazon S3 資料湖查詢資料，而無需將任何資料載入 Amazon Redshift Serverless。如需詳細資訊，請參閱[查詢資料湖](#)。

## 查詢 Amazon Redshift 資料庫外部的資料來源

您可以在下文中了解如何開始查詢遠端來源 (包括遠端 Amazon Redshift 叢集) 上的資料的資訊。您也可以在使用 Amazon Redshift 的資訊中了解機器學習 (ML) 模型的資訊。

### 主題

- [查詢您的資料湖](#)
- [查詢遠端資料來源上的資料](#)
- [存取其他 Amazon Redshift 叢集中的資料](#)
- [使用 Amazon Redshift 資料訓練機器學習模型](#)

## 查詢您的資料湖

您可以使用 Amazon Redshift Spectrum 查詢 Amazon S3 檔案中的資料，而無需將資料載入 Amazon Redshift 資料表。Amazon Redshift 提供 SQL 功能，專為儲存在 Amazon Redshift 叢集和 Amazon S3 資料湖中的非常大型資料集的快速線上分析處理 (OLAP) 而設計。您可以查詢多種格式的資料，包括 Parquet、ORC、RCFile、TextFile、SequenceFile、RegexSerde、OpenCSV 和 AVRO。若要在 Amazon S3 中定義檔案的結構，請建立外部結構描述和資料表。然後，您可以使用外部資料目錄，例如 AWS Glue，或您自己的 Apache Hive 中繼存放區。您的任何 Amazon Redshift 叢集都可以立即變更外部資料目錄。

在 AWS Glue 資料目錄中註冊並啟用 AWS Lake Formation 之後，您可以使用 Redshift Spectrum 進行查詢。

Redshift Spectrum 位於獨立於您叢集之外的專屬 Amazon Redshift 伺服器上。Redshift Spectrum 會推送許多 (例如述詞篩選和彙整等) 運算密集的工作到 Redshift Spectrum 層。Redshift Spectrum 還可以智慧擴充以利用大規模並行處理。

您可以在一或多個欄上對外部資料表進行分割，以最佳化資料分割的查詢效能。您可以使用 Amazon Redshift 資料表查詢和連接外部資料表。您可以從多個 Amazon Redshift 叢集存取外部資料表，並從相同 AWS 區域的任何叢集查詢 Amazon S3 資料。當您更新 Amazon S3 資料檔案時，資料可立即從任何 Amazon Redshift 叢集進行查詢。

如需有關 Redshift Spectrum 的詳細資訊，包括如何使用 Redshift Spectrum 和資料湖，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [Amazon Redshift Spectrum 入門](#)。

## 查詢遠端資料來源上的資料

您可以使用聯合查詢將 Amazon RDS 資料庫、Amazon Aurora 資料庫或 Amazon S3 中的資料與 Amazon Redshift 資料庫中的資料聯結在一起。您可以使用 Amazon Redshift 直接查詢操作資料 (無需移動)、套用轉換，以及將資料插入 Redshift 資料表。聯合查詢的某些運算會分散到遠端資料來源。

若要執行聯合查詢，Amazon Redshift 會先與遠端資料來源建立連線。然後，Amazon Redshift 會擷取遠端資料來源中資料表的相關中繼資料，發出查詢，然後擷取結果列。然後，Amazon Redshift 會將結果列分配到 Amazon Redshift 運算節點，以便進一步處理。

如需為聯合查詢設定環境的相關資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的下列其中一個主題：

- [開始使用 PostgreSQL 的聯合查詢](#)
- [開始使用 MySQL 的聯合查詢](#)

## 存取其他 Amazon Redshift 叢集中的資料

使用 Amazon Redshift 資料共用功能，您可以在 Amazon Redshift 叢集或 AWS 帳戶之間以高度安全且更輕鬆的方式共用即時資料，以進行讀取。您可以立即、精細且高效能地跨 Amazon Redshift 叢集存取資料，而無需手動複製或移動資料。在 Amazon Redshift 叢集中更新時，您的使用者可以看到最新且一致的資訊。您可以在不同層次共用資料，例如資料庫、結構描述、資料表、檢視 (包括一般、近期繫結和具體化視觀表)，以及 SQL 使用者定義函數 (UDF)。

Amazon Redshift 資料共用對於下列使用案例特別有用：

- 集中化業務關鍵工作負載 - 使用與多個商業智慧 (BI) 或分析叢集共用資料的中央擷取、轉換和載入 (ETL) 叢集。此方法可為個別工作負載提供讀取工作負載隔離和計費。
- 在環境之間共用資料：在開發、測試和生產環境之間共用資料。您可以透過分享精細層級不同的資料來提高團隊敏捷性。

如需有關資料共用的詳細資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的[開始使用資料共用](#)。

## 使用 Amazon Redshift 資料訓練機器學習模型

使用 Amazon Redshift 機器學習 (Amazon Redshift ML)，您可以透過將資料提供給 Amazon Redshift 來訓練模型。然後，Amazon Redshift ML 會建立可擷取輸入資料中模式的模型。然後，您可以使用這

些模型來產生新輸入資料的預測，而且不會產生額外費用。透過使用 Amazon Redshift ML，您可以使用 SQL 陳述式訓練機器學習模型，並在 SQL 查詢中調用它們以進行預測。您可以透過反覆變更參數並改善訓練資料，繼續提高預測的準確性。

Amazon Redshift ML 可讓 SQL 使用者使用熟悉的 SQL 命令輕鬆建立、訓練和部署機器學習模型。透過使用 Amazon Redshift ML，您可以使用 Amazon Redshift 叢集中的資料，透過 Amazon SageMaker Autopilot 訓練模型，並自動取得最佳模型。然後，您可以在 Amazon Redshift 資料庫中本地化模型並進行預測。

如需有關 Amazon Redshift ML 的詳細資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [開始使用 Amazon Redshift ML](#)。

# Amazon Redshift 佈建叢集

若您是第一次使用 Amazon Redshift，我們建議您閱讀以下章節，以協助您開始使用 Amazon Redshift。

## 主題

- [Amazon Redshift 佈建的叢集主控台](#)
- [連線至 Amazon Redshift 佈建叢集](#)
- [Amazon Redshift 叢集和資料載入](#)
- [一般資料庫任務](#)

## Amazon Redshift 佈建的叢集主控台

在開始設定 Amazon Redshift 叢集之前，請確定您已完成下列事前準備：

- [註冊成為 AWS](#)
- [決定防火牆規則](#)

## 註冊成為 AWS

如果您還沒有 AWS 帳戶，請註冊一個。如果您已有帳戶，則可略過此事前準備，並使用現有的帳戶。

1. 開啟 <https://portal.aws.amazon.com/billing/signup>。
2. 請遵循線上指示進行。

部分註冊程序需接收來電，並在電話鍵盤輸入驗證碼。

當您註冊時 AWS 帳戶，會建立 AWS 帳戶根使用者一個。根使用者有權存取該帳戶中的所有 AWS 服務和資源。作為最佳安全實務，[將管理存取權指派給管理使用者](#)，並且僅使用根使用者來執行[需要根使用者存取權的任務](#)。

## 決定防火牆規則

在本教學課程中，您會於啟動 Amazon Redshift 叢集時指定連接埠。您也會在安全群組中建立傳入規則，以允許經過此連接埠存取您的叢集。

如果您的用戶端電腦在防火牆背後，請確定您知道可以使用的開放連接埠。使用此開放連接埠，您可從 SQL 用戶端工具連接到叢集並執行查詢。如果您不知道開放連接埠，請洽詢對您的網路防火牆規則有所了解的人，以判斷防火牆中開放的連接埠。

雖然 Amazon Redshift 預設使用連接埠 5439，但如果防火牆中未開放此連接埠，則連線無法運作。您無法在建立您的 Amazon Redshift 叢集之後變更其連接埠號碼。因此，請確保您指定了在啟動程序期間可在您環境中運作的開放連接埠。

只有當您將自己的資料引入 Amazon Redshift 時，才會套用此事前準備。如需詳細資訊，請參閱 [將您的資料引入 Amazon Redshift](#)。

登入 Amazon Redshift 主控台之後，您可以建立和管理所有 Amazon Redshift 物件，包括叢集、資料庫和節點。您也可以檢視查詢、執行查詢，以及執行其他資料定義語言 (DDL) 和資料處理語言 (DML) 操作。

如果您是第一次使用 Amazon Redshift，建議您先前往儀表板、叢集和查詢編輯器 v2 頁面開始使用主控台。

若要開始使用 Amazon Redshift 主控台，請觀看以下影片：[開始使用 Amazon Redshift](#)。

您可以在下面找到 Amazon Redshift 主控台之導覽窗格項目的說明：

- Amazon Redshift Serverless – 存取和分析資料，而不必設定、微調和管理 Amazon Redshift 佈建的叢集。
- 佈建的叢集儀表板 - 檢查叢集指標和查詢概觀，瞭解指標資料 (例如 CPU 使用率) 和查詢資訊。使用這些功能可協助您判斷效能資料在指定時間範圍內是否異常。
- 叢集 — 檢視您 AWS 帳戶中的叢集清單、選擇要開始查詢的叢集，或執行叢集相關動作。您也可以從此頁面建立新叢集。
- 查詢編輯器 - 對 Amazon Redshift 叢集上託管的資料庫執行查詢、儲存查詢以供重複使用，或安排它們在將來執行 (僅限查詢編輯器中)。
- 查詢編輯器 v2 – 查詢編輯器 v2 是一種獨立的網頁式 SQL 用戶端應用程式，用來在 Amazon Redshift 資料倉儲上編寫和執行查詢。您可以將查詢結果視覺化為圖表，並透過與團隊中的其他人共用您的查詢來進行協作。
- 查詢和載入 - 取得參考或疑難排解的資訊，例如最近查詢的清單以及每個查詢的 SQL 文字。
- 資料共用 - 身為生產者帳戶管理員，可以授權消費者帳戶存取資料共用，也可以選擇不授權存取。若要使用授權的資料清單，消費者帳戶管理員可以將資料保護與帳戶中的整個 AWS 帳戶或特定叢集命名空間建立關聯。管理員也可以拒絕資料共用。

- 組態 – 從 SQL 用戶端工具透過 Java 資料庫連線 (JDBC) 和開放式資料庫連線 (ODBC) 來連接至 Amazon Redshift 叢集。您也可以設定 Amazon Redshift 託管的虛擬私有雲端 (VPC) 端點。這樣做可在以包含叢集的 Amazon VPC 服務型 VPC 與另一個執行用戶端工具的 VPC 之間提供私有連線。
- 建議 - 取得有關可對 Amazon Redshift 叢集進行變更的具體建議，以排定優化的優先順序。
- AWS Marketplace — 取得與 Amazon Redshift 搭配使用的其他工具或 AWS 服務的相關資訊。
- 警示 - 在叢集指標上建立警示，以檢視效能資料並追蹤指定期間內的指標。
- 事件 - 追蹤事件並取得有關事件發生日期、描述或事件來源等資訊的報告。
- 新功能 - 檢視新的 Amazon Redshift 功能和產品更新。

## 連線至 Amazon Redshift 佈建叢集

若要連接到 Amazon Redshift 叢集，請從叢集頁面展開連接到 Amazon Redshift 叢集，然後執行下列其中一個動作：

- 使用查詢編輯器 v2 在 Amazon Redshift 叢集託管的資料庫上執行查詢。建立叢集後，您可以立即使用查詢編輯器 v2 執行查詢。

如需詳細資訊，請參閱[使用 Amazon Redshift 查詢編輯器 v2 來查詢資料庫](#)。

- 透過複製 JDBC 或 ODBC 驅動程式 URL，使用 JDBC 或 ODBC 驅動程式從您的用戶端工具連接到 Amazon Redshift。

若要使用叢集上的資料，您需要 JDBC 或 ODBC 驅動程式，以便從用戶端電腦或執行個體來連接。將應用程式編寫為使用 JDBC 或 ODBC 資料存取 API 操作，或使用支援 JDBC 或 ODBC 的 SQL 用戶端工具。

如需如何尋找叢集連線字串的相關資訊，請參閱[尋找叢集連線字串](#)。

- 如果您的 SQL 用戶端工具需要驅動程式，您可以下載作業系統特定的驅動程式，以便從用戶端工具連接到 Amazon Redshift。

如需如何為 SQL 用戶端安裝適當驅動程式的相關資訊，請參閱[設定 JDBC 驅動程式 2.0 版的連線](#)。

如需如何設定 ODBC 連線的相關資訊，請參閱[設定 ODBC 連線](#)。



# Amazon Redshift 叢集和資料載入

在本節中，您可以找到兩個教學課程，引導您完成建立 Amazon Redshift 叢集範例的程序。其中一個是使用範例資料集，另一個則是使用您自己的資料集。

## 主題

- [使用範例資料集](#)
- [將您自己的資料引入 Amazon Redshift](#)

## 使用範例資料集

在本教學課程中，您會透過使用範例資料集，逐步完成建立 Amazon Redshift 叢集的程序。Amazon Redshift 會在您建立新叢集時自動載入範例資料集。您可以在建立叢集之後立即查詢資料。

在開始設定 Amazon Redshift 叢集之前，請確定您已完成[註冊成為 AWS](#)和[決定防火牆規則](#)。

在本教學課程中，您將執行下列所示的步驟：



**Step 1:**  
Create cluster



**Step 2:**  
Try sample queries

## 主題

- [步驟 1：建立範例 Amazon Redshift 叢集](#)
- [步驟 2：使用查詢編輯器嘗試範例查詢](#)

### Important

您所建立的範例叢集，會在實際環境中執行。使用本教學課程中設計的範例叢集時，隨需費率為每小時 0.25 USD，直到您刪除此範例叢集為止。如需定價的詳細資訊，請參閱 [Amazon Redshift 定價](#)。如果您有任何問題或進度停滯不前，可以在 [討論論壇](#) 張貼文章來聯繫 Amazon Redshift 團隊。

本教學課程不適用於生產環境，不會深入討論選項。完成此教學中的步驟後，您可以使用 [其他資源](#) 來找到更深入的資訊。此資訊可協助您規劃、部署和維護您的叢集，以及使用您的資料倉儲中的資料。

## 步驟 1：建立範例 Amazon Redshift 叢集

完成事前準備後，就可以根據範例資料集開始建立 Amazon Redshift 叢集。

若要根據範例資料集建立 Amazon Redshift 叢集，請執行下列動作：

1. 登入 AWS Management Console 並開啟 Amazon Redshift 主控台，網址為 <https://console.aws.amazon.com/redshiftv2/>。
2. 若要建立叢集，請執行下列其中一個操作：
  - 在 Amazon Redshift 服務頁面上，選擇建立叢集。建立叢集頁面隨即出現。
  - 在 <https://console.aws.amazon.com/redshiftv2/> 上，選擇已佈建的叢集儀表板，然後選擇建立叢集。
  - 在 <https://console.aws.amazon.com/redshiftv2/> 上，選擇「叢集」，然後選擇「建立叢集」。
3. 在叢集組態區段中，指定叢集識別碼。此識別碼必須是唯一的。識別碼必須介於 1-63 個字元之間，使用有效字元 a-z (僅限小寫字母) 和 - (連字號)。

針對本教學，輸入 **examplecluster**。

4. 如果您的組織符合資格，而且您的叢集是在無法使用 Amazon Redshift 無伺服器的 AWS 區域地方建立，您也許可以在 Amazon Redshift 免費試用計劃下建立叢集。選擇生產或免費試用來回答問題您打算將此叢集用於什麼目的？選擇免費試用時，您可以使用 dc2.large 節點類型建立組態。如需選擇免費試用的相關資訊，請參閱 [Amazon Redshift 免費試用版](#)。如需提供 Amazon Redshift 無伺服器的位 AWS 區域 置清單，請參閱中針對 [Redshift 無伺服器](#) API 列出的端點。Amazon Web Services 一般參考

在您選擇節點類型後，請執行下列其中一個動作：

- 在範例資料中，選擇載入範例資料，將範例資料集載入 Amazon Redshift 叢集。Amazon Redshift 會將範例資料集 Tickit 載入預設 dev 資料庫和 public 結構描述。您可以開始使用查詢編輯器 v2 來查詢資料。
- 若要將自己的資料引入 Amazon Redshift 叢集，請選擇生產。然後，在範例資料中，選擇載入範例資料。如需引入您自己的資料的相關資訊，請參閱[將您自己的資料引入 Amazon Redshift](#)。

Amazon Redshift 會自動將範例資料集載入您的範例 Amazon Redshift 叢集。

5. 在資料庫組態區段中，指定管理員使用者名稱的值。針對管理員密碼，從下列選項中選擇：
  - 產生密碼 - 使用 Amazon Redshift 產生的密碼。

- 手動新增管理員密碼 - 使用您自己的密碼。
- 在中管理管理登入資料 AWS Secrets Manager — Amazon Redshift 會用 AWS Secrets Manager 來產生和管理您的管理員密碼。用 AWS Secrets Manager 於生成和管理密碼的秘密需要支付費用。如需 AWS Secrets Manager 定價的詳細資訊，請參閱 [AWS Secrets Manager 定價](#)。

對於本教學課程，請使用這些值：

- 管理員使用者名稱：輸入 **awsuser**。
- 管理員使用者密碼：輸入密碼的值。

## 6. 選擇建立叢集。

本教學課程使用 Amazon Redshift 查詢編輯器 v2。在 Amazon Redshift 完成建立叢集之後，您可以使用此編輯器立即查詢資料。

您也可以選擇其他支援 JDBC 或 ODBC 驅動程式的 SQL 用戶端工具來處理叢集中的資料。如需詳細資訊，請參閱《Amazon Redshift 管理指南》中的 [使用 SQL 用戶端工具連接到 Amazon Redshift 叢集](#)。

### 步驟 2：使用查詢編輯器嘗試範例查詢

當 Amazon Redshift 建立您的 Amazon Redshift 叢集時，它會自動上傳範例資料集 Tickit。叢集建立可能需要幾分鐘時間才能完成。建立完成之後，叢集狀態會變成 ACTIVE。您可以從範例資料集中檢視範例 Tickit 資料表。

#### 使用查詢編輯器

您可以選擇叢集、dev 資料庫和 public 結構描述，在查詢編輯器 v2 中檢視範例 Tickit 資料表。

建立 Amazon Redshift 叢集之後，在連接到 Amazon Redshift 叢集中，選擇查詢資料。

在查詢編輯器 v2 中，連接到資料庫，然後在樹狀檢視面板中選擇叢集名稱。若出現提示，請輸入連線參數。

當您連接到叢集及其資料庫時，請提供資料庫名稱和使用者名稱。您也可以提供下列其中一種身分驗證方法所需的參數：

#### 資料庫使用者名稱和密碼

使用此方法，也可以為您要連線的資料庫提供密碼。

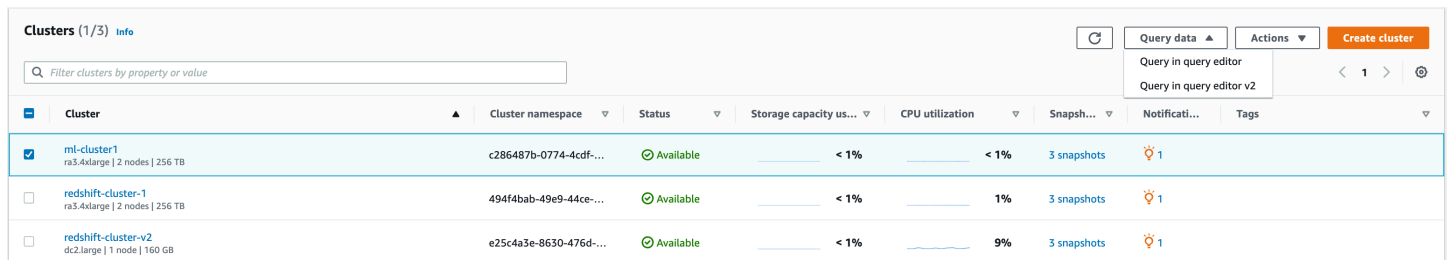
## 臨時憑證

使用此方法，查詢編輯器 v2 會產生一個臨時密碼來連接到資料庫。

當您使用查詢編輯器 v2 選取叢集時，視內容而定，您可以使用內容 (右鍵) 選單來建立、編輯和刪除連線。

根據預設，Amazon Redshift 會建立名為 dev 的預設資料庫，並建立名為 public 的預設結構描述。若要檢視範例資料集的個別資料檔案，請選擇叢集，移至查詢編輯器 v2，然後選擇 dev 資料庫、public 結構描述，以及 Tables。

或者，在導覽窗格中，選擇叢集和您要查詢資料的叢集。然後在查詢資料下，選擇在查詢編輯器中查詢或在查詢編輯器 v2 中查詢，以在指定的查詢編輯器中查詢資料



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

## 嘗試範例查詢

在查詢編輯器之一中嘗試一些範例查詢，如下所示。如需使用 SELECT 命令的相關資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [SELECT](#)。

```
-- 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;
```

完成此教學中的步驟後，您可以使用[其他資源](#)來找到更深入的資訊。此資訊可協助您規劃、部署和維護您的叢集，以及使用您的資料倉儲中的資料。

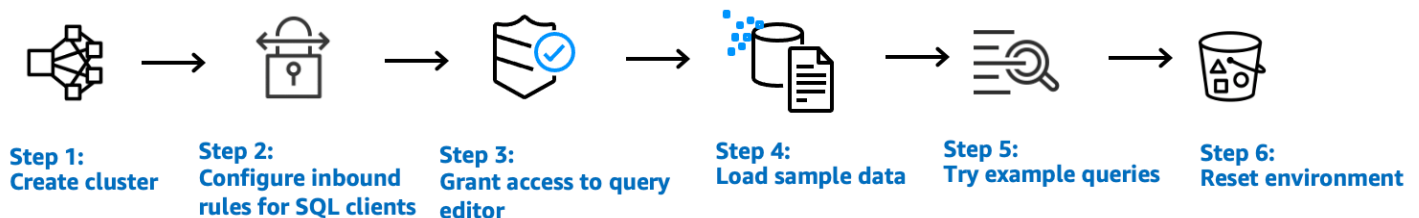
您也可以嘗試將您自己的資料引入 [Amazon Redshift](#) 教學課程，使用自己的資料集建立叢集。

## 將您自己的資料引入 Amazon Redshift

在本教學課程中，您會將自己的資料集引入 Amazon Redshift，逐步完成建立 Amazon Redshift 叢集的程序。您可以使用此範例叢集來評估 Amazon Redshift 服務。

在開始設定 Amazon Redshift 叢集之前，請確定您已完成[註冊成為 AWS](#)和[決定防火牆規則](#)。

在本教學課程中，您將執行下列所示的步驟。



### **⚠ Important**

您所建立的範例叢集，會在實際環境中執行。使用本教學課程中設計的範例叢集時，隨需費率為每小時 0.25 USD，直到您刪除此範例叢集為止。如需定價資訊，請前往 [Amazon Redshift 定價頁面](#)。如果您有任何問題或進度停滯不前，可以在 [討論論壇](#) 張貼文章來聯繫 Amazon Redshift 團隊。

本教學課程不適用於生產環境，不會深入討論選項。完成此教學中的步驟後，您可以使用[其他資源](#)來找到更深入的資訊。此資訊可協助您規劃、部署和維護您的叢集，以及使用您的資料倉儲中的資料。

## 主題

- [步驟 1：建立範例 Amazon Redshift 叢集](#)
- [步驟 2：設定 SQL 用戶端的傳入規則](#)
- [步驟 3：授予其中一個查詢編輯器的存取權並執行查詢](#)
- [步驟 4：將資料從 Amazon S3 載入到 Amazon Redshift](#)
- [步驟 5：使用查詢編輯器嘗試範例查詢](#)
- [步驟 6：重設環境](#)

## 步驟 1：建立範例 Amazon Redshift 叢集

對於從其他 AWS 資源存取資料的任何作業，您的叢集都需要權限，才能代表您存取資源和資源上的資料。範例是使用 COPY 命令從 Amazon Simple Storage Service (Amazon S3) 載入資料。您可以使用 AWS Identity and Access Management (IAM) 提供這些許可。您可以透過連接至叢集的 IAM 角色執行此操作。或者，您可以為具有必要許可的 IAM 使用者提供 AWS 存取金鑰。如需憑證和存取許可的相關資訊，請參閱[憑證和存取許可](#)。

為了最好地保護您的敏感資料並保護您的 AWS 存取登入資料，我們建議您建立 IAM 角色並將其附加到叢集。如需提供存取許可的相關資訊，請參閱[存取其他 AWS 資源的許可](#)。

您即將建立的叢集會實際運作，而非在沙盒中執行。您需要支付叢集的標準 Amazon Redshift 使用費，直到刪除叢集為止。如果您一口氣地完成這裡所述的教學課程，並在完成時刪除叢集，則總費用最少。

### 建立 Amazon Redshift 叢集


1. 登入 AWS Management Console 並開啟 Amazon Redshift 主控台，網址為 <https://console.aws.amazon.com/redshiftv2/>。

#### Important

如果您使用 IAM 使用者憑證，請確定您具備執行叢集操作的必要許可。如需詳細資訊，請參閱《Amazon Redshift 管理指南》中的[控制 IAM 使用者的存取](#)。

2. 在右上角，選擇您 AWS 區域 要建立叢集的位置。
3. 在導覽選單上，選擇叢集，然後選擇建立叢集。建立叢集頁面隨即出現。
4. 在叢集組態區段中，指定叢集識別碼、節點類型和節點的值：

- 叢集識別碼：針對此教學課程輸入 **examplecluster**。此識別碼必須是唯一的。識別碼必須介於 1-63 個字元之間，使用有效字元 a-z (僅限小寫字母) 和 - (連字號)。
- 請從下列方式擇一，調整叢集的大小：

 Note

下列步驟假設支援 RA3 節點類型的 AWS 區域。如需支援 RA3 節點類型的 AWS 區域清單，請參閱《Amazon Redshift 管理指南》中的 [RA3 節點類型概觀](#)。若要深入瞭解每個節點類型和大小的節點規格，請參閱[節點類型詳細資訊](#)。

- 如果您的 AWS 地區支援 RA3 節點類型，請選擇「生產」或「免費試用」來回答這個問題，您打算將此叢集用於什麼目的？
- 如果您不知道要將叢集的大小調整到多大，請選擇協助我選擇。這樣做會啟動調整大小計算器，詢問您有關計劃儲存在資料倉儲中之資料的大小和查詢特性的問題。

如果您知道叢集的所需大小 (也就是節點類型和節點數目)，請選擇我會選擇。然後選擇節點類型和節點數量來調整叢集的大小以進行概念驗證。

- 針對本教學課程，請選擇 dc2.large 做為節點類型，選擇 2 做為節點。
  - 如果您已為叢集選擇了生產，請執行下列其中一個動作：
    - 若要使用 Amazon Redshift 提供的範例資料集，請在範例資料中選擇載入範例資料。Amazon Redshift 會將範例資料集 Tickit 載入預設 dev 資料庫和 public 結構描述。
    - 若要將您自己的資料引入 Amazon Redshift，請繼續進行本教學課程的其餘部分。
5. 在資料庫組態區段中，指定管理員使用者名稱的值。針對管理員密碼，從下列選項中選擇：
- 產生密碼 - 使用 Amazon Redshift 產生的密碼。
  - 手動新增管理員密碼 - 使用您自己的密碼。
  - 在中管理管理登入資料 AWS Secrets Manager — Amazon Redshift 會用 AWS Secrets Manager 來產生和管理您的管理員密碼。用 AWS Secrets Manager 於生成和管理密碼的秘密需要支付費用。如需 AWS Secrets Manager 定價的詳細資訊，請參閱 [AWS Secrets Manager 定價](#)。

對於本教學課程，請使用這些值：

- 管理員使用者名稱：輸入 **awsuser**。

- 管理員使用者密碼：輸入密碼的值。
6. 在本教學課程中，請建立 IAM 角色並將其設定為叢集的預設角色，如下所述。一個叢集只能設定一個預設 IAM 角色。
    - a. 在叢集許可下，針對管理 IAM 角色選擇建立 IAM 角色。
    - b. 透過下列其中一種方法，為 IAM 角色指定要存取的 Amazon S3 儲存貯體：
      - 選擇沒有其他 Amazon S3 儲存貯體，允許建立的 IAM 角色只存取命名為 redshift 的 Amazon S3 儲存貯體。
      - 選擇任何 Amazon S3 儲存貯體，以允許建立的 IAM 角色存取所有 Amazon S3 儲存貯體。
      - 選擇特定的 Amazon S3 儲存貯體，為建立的 IAM 角色指定一或多個要存取的 Amazon S3 儲存貯體。然後從資料表中選擇一或多個 Amazon S3 儲存貯體。
    - c. 選擇建立 IAM 角色作為預設值。Amazon Redshift 會自動建立 IAM 角色，並將其設定為叢集的預設角色。

由於您是從主控台建立 IAM 角色，因此該角色已連接 AmazonRedshiftAllCommandsFullAccess 政策。這可讓 Amazon Redshift 從您的 IAM 帳戶中的 Amazon 資源複製、載入、查詢和分析資料。

如需如何管理叢集的預設 IAM 角色的詳細資訊，請參閱[建立 IAM 角色做為 Amazon Redshift 的預設值](#)。

7. (選用) 在其他組態區段中，關閉使用預設值來修改網路和安全性、資料庫組態、維護、監控和備份設定值。

在某些情況下，您可以使用載入範例資料選項建立叢集，並希望開啟增強型 Amazon VPC 路由。若是這樣，虛擬私有雲端 (VPC) 中的叢集需要存取 Amazon S3 端點，才能載入資料。

若要讓叢集可公開存取，您可以執行下列其中一項作業。您可以在 VPC 中設定網路位址轉譯 (NAT) 位址，以存取網際網路。或者，您可以在 VPC 中設定 Amazon S3 VPC 端點。如需增強型 VPC 路由的相關資訊，請參閱《Amazon Redshift 管理指南》中的[啟用增強型 VPC 路由](#)。

8. 選擇建立叢集。

## 步驟 2：設定 SQL 用戶端的傳入規則

在本教學課程的後面部分，您將根據 Amazon VPC 服務從虛擬私有雲端 (VPC) 內存取叢集。但是，如果您從防火牆的外部使用 SQL 用戶端存取叢集，請務必授予傳入存取權限。



如果您計劃從您的 VPC 內使用 Amazon Redshift 查詢編輯器存取叢集，您可以跳過此步驟。

### 檢查您的防火牆及授予您叢集的傳入存取權限

1. 如果您的叢集需要從防火牆的外部進行存取，請檢查您的防火牆規則。例如，您的用戶端可能是 Amazon Elastic Compute Cloud (Amazon EC2) 執行個體或外部電腦。

如需防火牆規則的相關資訊，請參閱《適用於 Linux 執行個體的 Amazon EC2 使用者指南》中的[安全群組規則](#)。

2. 如要從 Amazon EC2 外部用戶端進行存取，請將輸入規則新增至連接到您叢集的安全群組，以允許傳入流量。您可以在 Amazon EC2 主控台中新增 Amazon EC2 安全群組規則。例如，192.0.2.0/24 的 CIDR/IP 可允許該 IP 地址中的用戶端連接到您的叢集。找出適合您環境的正確 CIDR/IP。

### 步驟 3：授予其中一個查詢編輯器的存取權並執行查詢

若要查詢由 Amazon Redshift 叢集所託管的資料庫，您有兩個選項：

1. Connect 到您的叢集，並 AWS Management Console 使用其中一個查詢編輯器在上執行查詢。

如果您使用其中一個查詢編輯器，則不需要下載和設定 SQL 用戶端應用程式。

2. 透過 SQL 用戶端工具 (例如 SQL Workbench/J) 連接到叢集。如需使用 SQL Workbench/J 的相關資訊，請參閱《Amazon Redshift 管理指南》中的[使用 SQL Workbench/J 連接到叢集](#)。

使用其中一個 Amazon Redshift 查詢編輯器是對 Amazon Redshift 叢集託管的資料庫執行查詢的最簡單方法。建立叢集後，您可以使用 Amazon Redshift 主控台來立即執行查詢。有關使用 Amazon Redshift 查詢編輯器時之注意事項的詳細資訊，請參閱《Amazon Redshift 管理指南》中的[使用查詢編輯器查詢資料庫](#)。

### 授予查詢編輯器 v2 的存取權

系統管理員第一次為您設定查詢編輯器 v2 時 AWS 帳戶，他們會選擇用 AWS KMS key 來加密查詢編輯器 v2 資源的。根據預設，系統會使用 AWS 擁有的金鑰來加密資源。或者，管理員也可以在組態頁面中選擇金鑰的 Amazon Resource Name (ARN)，以使用客戶受管金鑰。設定帳戶後，無法變更 AWS KMS 加密設定。如需詳細資訊，請參閱[設定您的 AWS 帳戶](#)。

若要存取查詢編輯器 v2，您需要許可。管理員可以將下列其中一個 AWS 受管政策附加到 IAM 角色或使用者，以授與許可。這些 AWS 受管理的政策使用不同的選項撰寫，這些選項可控制標記資源允許共用查詢的方式。您可以使用 IAM 主控台 (<https://console.aws.amazon.com/iam/>) 來連接 IAM 政策。

您也可以根據提供的受管政策中允許和拒絕的許可，來建立自己的政策。如果您使用 IAM 主控台政策編輯器建立自己的政策，請選擇 SQL Workbench 作為您在視覺化編輯器中為其建立政策的服務。查詢編輯器 v2 會在視覺化編輯器和 IAM 政策模擬器中使用服務名稱 AWS SQL 工作台。

如需詳細資訊，請參閱[存取查詢編輯器 v2](#)。

## 使用查詢編輯器 v2

若要使用查詢編輯器 v2 查詢資料庫，請參閱[使用查詢編輯器 v2](#)。

## 授予查詢編輯器的存取權

若要使用 Amazon Redshift 查詢編輯器，您需要許可。若要設定存取權，請將 AmazonRedshiftQueryEditor 和 AmazonRedshiftReadOnlyAccess IAM 政策 連接至 IAM 角色，然後將該角色指派給您用於存取叢集的使用者。

## 連接查詢編輯器所需的 IAM 政策

1. 登入 AWS Management Console 並開啟身分與存取權管理主控台，網址為 <https://console.aws.amazon.com/iam/>。
2. 選擇角色。
3. 選擇需要存取查詢編輯器的角色。
4. 在許可索引標籤上，選擇新增許可。
5. 選擇連接政策。
6. 針對政策名稱，選擇 AmazonRedshiftQueryEditor 和 AmazonRedshiftReadOnlyAccess。
7. 選擇新增許可。

若要建立 IAM 角色以允許 Amazon RDS 叢集代表您與其他 AWS 服務進行通訊，請採取下列步驟。本節中使用的值是範例，您可以根據需要選擇值。

## 若要建立 IAM 角色以允許 Amazon Redshift 存取 AWS 服務

1. 開啟 [IAM 主控台](#)。
2. 在導覽窗格中，選擇 Roles (角色)。
3. 選擇 Create Role (建立角色)。
4. 選擇 AWS 服務，然後選擇 Redshift。

5. 在選取您的使用案例下，選擇Redshift – 可自訂，然後選擇下一步：許可。連接許可政策頁面隨即出現。
6. 新增AmazonRedshiftQueryEditor和AmazonRedshiftReadOnlyAccess權限。

選擇下一步：標籤。

7. Add tags (新增標籤) 頁面隨即出現。您可以選擇性地新增標籤。選擇下一步：檢閱。
8. 針對 Role name (角色名稱)，輸入您的角色名稱，例如 **RedshiftQueryEditorAccess**。選擇建立角色。
9. 新角色適用於叢集上使用該角色的所有使用者。若要限制只有特定叢集上的特定使用者或特定區域中的叢集可存取，請編輯角色的信任關係。如需詳細資訊，請參閱《Amazon Redshift 管理指南》中的[限制對 IAM 角色的存取](#)。
10. 將角色與叢集相關聯。您可在建立叢集或新增角色至現有叢集時將 IAM 角色與叢集相關聯。如需詳細資訊，請參閱《Amazon Redshift 管理指南》中的[將 IAM 角色與叢集建立關聯](#)。

#### Note

若要限制對特定資料的存取權，請使用授予所需最低權限的 IAM 角色。

## 使用查詢編輯器

若要使用查詢編輯器來查詢資料庫，請參閱[使用查詢編輯器來查詢資料庫](#)。

## 步驟 4：將資料從 Amazon S3 載入到 Amazon Redshift

使用其中一個 Amazon Redshift 查詢編輯器是將資料載入資料表的最簡單方法。建立叢集之後，您可以使用 Amazon Redshift 主控台將資料從 Amazon S3 載入叢集。

使用查詢編輯器 v2 可簡化使用載入資料精靈時載入資料的程序。如需詳細資訊，請參閱[使用查詢編輯器 v2，將您自己的資料從 Amazon S3 載入到 Amazon Redshift](#)。您也可以使用查詢編輯器 v2 來建立資料表並載入資料。如需詳細資訊，請參閱[使用查詢編輯器從 Amazon S3 載入範例資料](#)。

使用查詢編輯器 v2，將您自己的資料從 Amazon S3 載入到 Amazon Redshift

若要將自己的資料從 Amazon S3 載入到 Amazon Redshift，Amazon Redshift 需要具有所需權限的 IAM 角色，以便從指定的 Amazon S3 儲存貯體載入資料。

首先，連接至資料庫。接下來，在資料庫中建立一些資料表。然後將您自己的資料從 Amazon S3 載入到 Amazon Redshift。如需如何使用查詢編輯器 v2 的相關資訊，請參閱《Amazon Redshift 管理指南》中的[使用查詢編輯器 v2](#)。

在查詢編輯器 v2 載入資料精靈中產生和使用的 COPY 命令，支援 COPY 命令語法可用的所有參數，以便從 Amazon S3 載入資料。如需 COPY 命令及其用於從 Amazon S3 複製載入之選項的詳細資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的[從 Amazon Simple Storage Service 進行 COPY](#)。

### 使用查詢編輯器從 Amazon S3 載入範例資料

目前，您有一個名為 dev 的資料庫且已連接此資料庫。接下來，您會在資料庫中建立一些資料表、將資料上傳至資料表，然後嘗試執行查詢。為方便起見，在 Amazon S3 儲存貯體中會提供您載入的範例資料。

#### Note

如果您是使用 SQL 用戶端工具，請確定 SQL 用戶端已連接至叢集。

完成此步驟後，您可以執行以下操作：

- 嘗試[步驟 5：使用查詢編輯器嘗試範例查詢](#)的範例查詢。
- 在[步驟 6：重設環境](#)重設您的環境。
- 在[其他資源](#)尋找有關 Amazon Redshift 的更多資訊。

#### Note

若要嘗試在查詢編輯器中查詢資料而不載入您自己的資料，請在範例資料中選擇載入範例資料。如果這樣做，Amazon Redshift 會在叢集建立期間自動將其範例資料集載入您的 Amazon Redshift 叢集。

### 從 Amazon S3 載入範例資料

1. 建立資料表。

如果您使用 Amazon Redshift 查詢編輯器，請個別複製並執行以下建立資料表陳述式，以在 dev 資料庫中建立資料表。如需語法的相關資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [CREATE TABLE](#)。

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

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

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

```
create table date(  
  dateid smallint not null distkey sortkey,
```

```
caldate date not null,  
day character(3) not null,  
week smallint not null,  
month character(5) not null,  
qtr character(5) not null,  
year smallint not null,  
holiday boolean default('N'));
```

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

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

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

## 2. 使用 COPY 命令從 Amazon S3 載入範例資料。

**Note**

建議您使用 COPY 命令，從 Amazon S3 或 Amazon DynamoDB 將大型資料集載入 Amazon Redshift。如需 COPY 語法的相關資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [COPY](#)。

- a. 下載檔案 [ticketdb.zip](#)，其中包含了個別的範例資料檔案。
- b. 解壓縮個別檔案，並將這些檔案載入您在 AWS 區域 Amazon S3 儲存貯體中的 ticket 資料夾。
- c. 編輯在本教學課程中的 COPY 命令，以便指向 Amazon S3 儲存貯體中的檔案。如需如何使用 Amazon S3 管理檔案的詳細資訊，請參閱《Amazon Simple Storage Service 使用者指南》中的 [建立與設定 S3 儲存貯體](#)。
- d. 為您的叢集提供身分驗證，以代表您存取 Amazon S3 以載入範例資料。您可以透過參考您在先前步驟中建立並設定為叢集預設值的 IAM 角色，以提供身分驗證。

COPY 命令包括 IAM 角色的 Amazon 資源名稱 (ARN) 預留位置、您的儲存貯體名稱 AWS 區域，以及下列範例所示。

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

COPY 命令看起來應該會與以下範例類似。

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

若要載入範例資料，請使用您的值取代下列 COPY 命令中的 *<myBucket>* 和 *<aws-region>*。如果您使用的是 Amazon Redshift 查詢編輯器，請個別執行下列命令。

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

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

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

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

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

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

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

## 步驟 5：使用查詢編輯器嘗試範例查詢

現在，嘗試一些範例查詢，如下所示。如需使用 SELECT 命令的相關資訊，請參閱《Amazon Redshift 開發人員指南》中的 [SELECT](#)。

```
-- 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;
```

您已成功建立 Amazon Redshift 叢集，並使用 Amazon Redshift 查詢編輯器從您自己的資料集查詢資料。

## 步驟 6：重設環境

完成本教學課程後，建議刪除您的範例叢集，以將您的環境重設至先前的狀態。您需要繼續支付 Amazon Redshift 服務的費用，直到刪除叢集為止。

但是，如果您想要嘗試其他 Amazon Redshift 指南中的任務，您可以讓範例叢集繼續執行。

### 刪除叢集

1. 登入 AWS Management Console 並開啟 Amazon Redshift 主控台，網址為 <https://console.aws.amazon.com/redshiftv2/>。
2. 在導覽選單上，選擇叢集來顯示您的叢集清單。
3. 選擇 **examplecluster** 叢集。對於 Actions (動作)，請選擇 Delete (刪除)。Delete cluster (刪除叢集) 頁面隨即出現。
4. 確認要刪除的叢集，然後選擇 Delete cluster (刪除叢集)。

在叢集清單頁面上，叢集的狀態會隨著刪除叢集而更新。

完成本教學課程之後，您可以在[其他資源](#)中找到 Amazon Redshift 的詳細資訊及後續步驟。

## 一般資料庫任務

接下來，您可以尋找一般任務的說明和逐步解說，以開始使用 Amazon Redshift 資料庫。

連接到初始叢集 dev 資料庫之後，即可建立新的資料庫。不論您是選擇使用範例資料集，還是在建立叢集時將自己的資料引入 Amazon Redshift，Amazon Redshift 都會建立 dev 資料庫。

本節中的範例假設如下情況：

- 您已建立 Amazon Redshift 叢集。如需詳細資訊，請參閱 [Amazon Redshift 叢集和資料載入](#)。
- 您已經從 SQL 用戶端工具 (例如 Amazon Redshift 主控台查詢編輯器) 建立到叢集的連線。如需詳細資訊，請參閱 [步驟 3：授予其中一個查詢編輯器的存取權並執行查詢](#)。

### Important

您為此練習所部署的叢集將在實際環境中執行。只要它正在運行，它就會向您的 AWS 帳戶收取費用。如需定價資訊，請參閱 [Amazon Redshift 定價頁面](#)。

為了避免產生不必要的費用，請在使用完後立即刪除您的叢集。練習的最終步驟說明做法。

## 任務 1：建立資料庫

在您確認了叢集已啟動並執行之後，即可建立您自己的第一個資料庫。此資料庫是您實際建立資料表、載入資料，以及執行查詢之處。單一叢集可以代管多個資料庫。例如，您可以在同一個叢集上擁有 SALESDB 資料庫和 ORDERSDB 資料庫。

例如，若要建立名為 **SALESDB** 的資料庫，請在 SQL 用戶端工具中執行下列命令。

```
CREATE DATABASE SALESDB;
```

對於本練習，接受預設值。如需更多命令選項的詳細資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [CREATE DATABASE](#)。

在建立了 SALESDB 資料庫之後，您可以從 SQL 用戶端連接至新的資料庫。使用與您用於目前連線相同的連線參數，但將資料庫名稱變更為 SALESDB。

## 任務 2：建立使用者

依預設，只有您啟動叢集時所建立的管理員使用者才能存取叢集中的初始資料庫。若要授予其他使用者存取權，請建立一個或多個帳戶。資料庫使用者帳戶適用於叢集中的所有資料庫；它們不屬於個別資料庫。

使用 CREATE USER 命令來建立新的使用者。建立新使用者時，您可以指定新使用者的名稱和密碼。我們建議您指定使用者的密碼。它須有 8 至 64 個字元，且必須包含一個大寫字母、一個小寫字母，以及一個數字。

例如，若要建立名為 **GUEST** 的使用者，且密碼為 **ABCd4321**，請執行下列命令。

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

若要以使用者 GUEST 身分連接到 SALESDB 資料庫，請在建立使用者時使用相同的密碼，例如 ABCd4321。

如需更多命令選項的詳細資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [CREATE USER](#)。

## 任務 3：建立結構描述

建立新資料庫之後，即可在目前的資料庫中建立新的結構描述。結構描述是包含具名資料庫物件 (例如資料表、檢視和使用者定義函數 (UDF)) 的命名空間。一個資料庫可以包含一個或多個結構描述，但每個結構描述只能屬於一個資料庫。兩個結構描述可以具有共用相同名稱的不同物件。

您可以在同一個資料庫中建立多個結構描述，以您想要的方式組織資料，或以功能方式將資料分組。例如，您可以建立結構描述來儲存所有暫存資料，以及另一個結構描述來儲存所有報表資料。您也可以建立不同的結構描述，以儲存與相同資料庫中不同業務群組相關的資料。每個結構描述可儲存不同資料庫物件，例如資料表、檢視和使用者定義函數 (UDF)。此外，您可以使用 AUTHORIZATION 子句建立結構描述。這個子句會將擁有權提供給指定的使用者，或設定指定結構描述可使用的磁碟空間上限配額。

Amazon Redshift 會自動為每個新資料庫建立名為 public 的結構描述。如果您在建立資料庫物件時未指定結構描述名稱，則物件會進入 public 結構描述。

若要存取結構描述中的物件，請使用 schema\_name.table\_name 符號限定物件。結構描述的限定名稱包含結構描述名稱和資料表名稱，並以點分隔。例如，您可能有一個 sales 結構定義，其中包含 price 資料表，以及也有 price 資料表的 inventory 結構描述。當您參照 price 資料表時，您必須將其限定為 sales.price 或 inventory.price。

以下範例會為使用者 GUEST 建立名為 **SALES** 的結構描述。

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

如需更多命令選項的詳細資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [CREATE SCHEMA](#)。

若要檢視資料庫中的結構描述清單，請執行以下命令。

```
select * from pg_namespace;
```

輸出應看起來如下列內容。

nsname	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}

如需如何查詢目錄資料表的相關資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [查詢目錄資料表](#)。

使用 GRANT 陳述式將結構描述的許可授予使用者。

下列範例會授予 GUEST 使用者使用 SELECT 陳述式句從 SALES SCHEMA 中的所有資料表或檢視中選擇資料的權限。

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

下列範例會一次將所有可用權限授予 GUEST 使用者。

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

## 任務 4：建立資料表

在您建立新的資料庫之後，請建立資料表來保留您的資料。建立資料表時，您可以為資料表指定任何欄資訊。

例如，若要建立名為 **DEMO** 的資料表，請執行下列命令。

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

您也可以使用 `schema_name.object_name` 符號建立資料表，在 SALES 結構描述中建立資料表。

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

若要檢視和檢查結構描述及其資料表，您可以使用 Amazon Redshift 查詢編輯器。或者，您也可以使用系統檢視來查看結構描述中的資料表清單。如需詳細資訊，請參閱 [任務 6：查詢系統資料表](#)。

根據預設，新的資料庫物件 (例如資料表) 會在叢集建立期間所建立且名為 `public` 的預設結構描述中建立。您可以使用其他結構描述來建立資料庫物件。如需結構描述的相關資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [管理資料庫安全性](#)。

Amazon Redshift 使用 `encoding`、`distkey` 和 `sortkey` 欄進行平行處理。如需設計納入這些元素之資料表的相關資訊，請參閱 [Amazon Redshift 的資料表設計最佳實務](#)。

## 將資料列插入至資料表

在建立資料表之後，可多列資料插入至該資料表。

### Note

[INSERT](#) 命令可將行插入資料表中。如需標準大量載入，請使用 [COPY](#) 命令。如需詳細資訊，請參閱 [使用 COPY 命令載入資料表](#)。

例如，若要將值插入 DEMO 資料表中，請執行下列命令。

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

若要將資料插入結構描述中的資料表，請執行下列命令。

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

## 從資料表選取資料

在您建立資料表並將資料填入其中之後，請使用 `SELECT` 陳述式來顯示資料表中包含的資料。`SELECT *` 陳述式會傳回資料表中的所有資料的所有欄名和列值。使用 `SELECT` 是驗證最近新增的資料是否正確插入到資料表中的好方法。

若要檢視您在 **DEMO** 資料表中輸入的資料，請執行下列命令。

```
SELECT * from DEMO;
```

結果應如下所示。

```
personid | city
-----+-----
         781 | San Jose
         990 | Palo Alto
(2 rows)
```

如需使用 `SELECT` 陳述式來查詢資料表的相關資訊，請參閱 [SELECT](#)。

## 任務 5：載入範例資料

本指南中的大部分範例皆使用 TICKIT 範例資料集。您可以下載檔案 [ticketdb.zip](#)，其中包含了個別的範例資料檔案。

您可以將範例資料載入自己的 Amazon S3 儲存貯體。

若要載入資料庫的範例資料，請先建立資料表。然後使用 `COPY` 命令來載入包含儲存在 Amazon S3 儲存貯體中範例資料的資料表。如需建立資料表和載入範例資料的步驟，請參閱 [步驟 4：將資料從 Amazon S3 載入到 Amazon Redshift](#)。

## 任務 6：查詢系統資料表

除了您建立的資料表外，您的資料庫還包含一些系統資料表。這些系統資料表包含安裝的相關資訊，以及在系統上執行之各種查詢和程序的相關資訊。您可以查詢這些系統資料表來收集資料庫的相關資訊。

**Note**

本文件中每個資料表中的描述會指出是所有使用者都可看見資料表，還是僅超級使用者才能看見資料表。以超級使用者身分登入，查詢僅超級使用者才能看見的資料表。

Amazon Redshift 可讓您存取下列類型的系統資料表：

- [STL 資料表](#)

這些系統資料表是從 Amazon Redshift 日誌檔案產生，以提供系統的歷史記錄。記錄資料表具有 STL 字首。

- [STV 資料表](#)

這些資料表是虛擬系統資料表，包含目前系統資料的快照。快照資料表具有 STV 字首。

- [系統檢視](#)

系統畫面包含在數個 STL 和 STV 系統資料表中找到的資料子集。系統檢視具有 SVV 或 SVL 字首。

- [系統目錄資料表](#)

系統目錄資料表儲存結構描述中繼資料，例如資料表和資料欄的相關資訊。系統目錄資料表有 PG 字首。

若要擷取有關該查詢的系統資料表資訊，您可能需要指定與查詢相關聯的處理程序 ID。如需詳細資訊，請參閱 [判斷執行中查詢的程序 ID](#)。

## 檢視資料表名稱的清單

若要檢視結構描述中所有資料表的清單，您可以查詢 PG\_TABLE\_DEF 系統目錄資料表。您可以先檢查 search\_path 的設定。

```
SHOW search_path;
```

結果應類似以下內容，

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

```
(1 row)
```

下列範例會將 SALES 結構描述新增至搜尋路徑，並顯示 SALES 結構描述中的所有資料表。

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

SHOW search_path;
      search_path
-----
"$user", public, sales
(1 row)

select * from pg_table_def where schemaname = 'sales';
schemaname | tablename | column |          type          | encoding | distkey |
sortkey | notnull
-----+-----+-----+-----+-----+-----+-----
+-----+-----
sales      | demo      | personid | integer                | az64     | f       |
0 | f
sales      | demo      | city      | character varying(255) | lzo      | f       |
0 | f
(2 rows)
```

下列範例顯示目前資料庫中所有結構描述名為 DEMO 之所有資料表的清單。

```
select * from pg_table_def where tablename = 'demo';
schemaname | tablename | column |          type          | encoding | distkey |
sortkey | notnull
-----+-----+-----+-----+-----+-----+-----
+-----+-----
public     | demo      | personid | integer                | az64     | f       |
0 | f
public     | demo      | city      | character varying(255) | lzo      | f       |
0 | f
sales      | demo      | personid | integer                | az64     | f       |
0 | f
sales      | demo      | city      | character varying(255) | lzo      | f       |
0 | f
(4 rows)
```

如需詳細資訊，請參閱 [PG\\_TABLE\\_DEF](#)。

您也可以先選擇要連線的資料庫，使用查詢編輯器 v2 來檢視指定結構描述中的所有資料表。



## 檢視使用者

您可以查詢 PG\_USER 目錄來檢視所有使用者的清單，以及使用者 ID (USESYSID) 和使用者權限。

```
SELECT * FROM pg_user;
  username | usesysid | usecreatedb | usesuper | usecatupd | passwd | valuntil |
  useconfig
-----+-----+-----+-----+-----+-----+-----+
+-----+
 rdsdb    |        1 | true        | true     | true     | ***** | infinity |
 awsuser  |       100 | true        | true     | false    | ***** |          |
 guest    |       104 | true        | false    | false    | ***** |          |
(3 rows)
```

使用者名稱 rdsdb 供 Amazon Redshift 在內部用來執行日常管理和維護任務。您可以將 where usesysid > 1 新增至 select 陳述式來篩選您的查詢，以僅顯示使用者定義的使用者名稱。

```
SELECT * FROM pg_user WHERE usesysid > 1;
  username | usesysid | usecreatedb | usesuper | usecatupd | passwd | valuntil |
  useconfig
-----+-----+-----+-----+-----+-----+-----+
+-----+
 awsuser   |       100 | true        | true     | false    | ***** |          |
 guest     |       104 | true        | false    | false    | ***** |          |
(2 rows)
```

## 檢視近期查詢

在前一個範例中，adminuser 的使用者 ID (USESYSID) 為 100。若要列出五個最近由 adminuser 執行的查詢，您可以查詢 SVL\_QLOG 檢視。

SVL\_QLOG 檢視是來自 STL\_QUERY 資料表的更易用資訊子集。您可以使用此檢視來找出最近執行之查詢的查詢 ID (QUERY) 或處理程序 ID (PID) 您也可以使用此檢視來檢查查詢完成所花費的時間。SVL\_QLOG 包含查詢字串 (SUBSTRING) 的頭 60 個字元，以協助您找出特定查詢。請使用 LIMIT 子句與 SELECT 陳述式搭配，將結果限制為五個資料列。

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

結果類似如下所示。

```

query| pid | elapsed | substring
-----+-----+-----
+-----+-----+-----+-----
892 | 21046 | 55868 | SELECT query, pid, elapsed, substring from svl_qlog WHERE
us
620 | 17635 | 1296265 | SELECT query, pid, elapsed, substring from svl_qlog WHERE us
610 | 17607 | 82555 | SELECT * from DEMO;
596 | 16762 | 226372 | INSERT INTO DEMO VALUES (100);)

```

## 判斷執行中查詢的程序 ID

您可能需要尋找仍在執行之查詢的 PID。例如，如果需要取消正在花費太長時間來執行的查詢，您將需要 PID。您可以查詢 STV\_RECENTS 系統資料表，來取得執行中查詢之處理程序 ID 的清單，以及對應的查詢字串。如果您的查詢傳回多個 PID，則您可以查看查詢文字，來判斷您需要哪個 PID。

若要判斷執行中查詢的 PID，請執行下列 SELECT 陳述式。

```

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

```

## 任務 7：取消查詢

如果您執行的查詢花費太多時間或耗費過多叢集資源，可取消查詢。例如，建立售票員清單，其中包括售票員的名稱和賣出的門票數量。下列查詢從 SALES 資料表和 USERS 資料表中選取資料，並比對 WHERE 子句中的 SELLERID 和 USERID，來聯結這兩個資料表。

```

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

```

結果類似如下所示。

```

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

這是一個複雜查詢。對於本教學課程，您無需擔心此查詢的建構方式。

前一個查詢僅幾秒便完成執行，並傳回 2,102 個資料列。

假設您忘了放入 WHERE 子句。

```

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

```

結果集會包含 SALES 資料表中的所有列乘以 USERS 資料表中的所有列 (49989\*3766)。這稱為笛卡兒聯結，但不建議使用它。結果超過 1 億 8 千 8 百萬個資料列，而且需要很長時間來執行。

若要取消執行中查詢，請使用 CANCEL 命令與的 PID 搭配。

若要尋找處理程序 ID，請啟動一個新工作階段並查詢 STV\_RECENTS 資料表，如前一個步驟所示。以下範例說明如何使結果更具可讀性。若要這麼做，請使用 TRIM 函數修剪結尾空格，並僅顯示查詢字串的前 20 個字元。

```

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

```

結果類似如下所示。

```

pid | btrim | starttime | substring
-----+-----+-----+-----

```

```
610 | adminuser | 2013-03-28 18:39:49.355918 | select sellerid, fir  
(1 row)
```

若要取消 PID 為 610 的查詢，請執行下列命令。

```
CANCEL 610;
```

#### Note

CANCEL 命令不會停止交易。若要停止或還原交易，請使用 ABORT 或 ROLLBACK 命令。若要取消與交易相關聯的查詢，首先取消查詢，然後停止交易。

如果您已取消的查詢與交易相關聯，請使用 ABORT 或 ROLLBACK 命令來取消交易，並捨棄對資料所做的任何變更：

```
ABORT;
```

除非您是以超級使用者身分登入，否則只能取消自己的查詢。超級使用者可以取消所有查詢。

## 取消另一個工作階段的查詢

如果您的查詢工具不支援同時執行查詢，請啟動另一個工作階段來取消查詢。例如，我們在《Amazon Redshift 入門指南》中使用的查詢編輯器不支援多個並行查詢。若要使用查詢編輯器啟動另一個工作階段，請選擇檔案、新視窗，然後使用相同的連線參數進行連接。然後，您可以尋找 PID 並取消查詢。

## 使用超級使用者佇列來取消查詢

如果您的目前工作階段有太多查詢同時執行中，則您可能無法執行 CANCEL 命令，直到另一個查詢完成為止。在此情況下，請使用不同的工作負載管理查詢佇列來執行 CANCEL 命令。

使用工作負載管理，您可執行不同查詢佇列中的查詢，讓您無需等待另一個查詢完成。工作負載管理員會建立個別佇列 (稱為超級使用者佇列)，您可以將其用於故障診斷。若要使用超級使用者佇列，請以超級使用者身分登入，並使用 SET 命令將查詢群組設為 'superuser'。在執行您的命令之後，請使用 RESET 命令來重設查詢群組。

若要使用超級使用者佇列來取消查詢，請執行這些命令。

```
SET query_group TO 'superuser';
```

```
CANCEL 610;  
RESET query_group;
```

## 任務 8：清除您的資源

如果您部署了叢集來完成本練習，則在完成本練習後請刪除該叢集。刪除叢集會停止向您的 AWS 帳戶產生費用。

若要刪除叢集，請按照《Amazon Redshift 管理指南》中[刪除叢集](#)的步驟進行。

如果想要保留叢集，請保留範例資料以供參考。本指南中的大部分範例皆使用您在本練習中所建立的資料表。資料大小將不會對您的可用儲存空間有任何重大的影響。

如果想要保留叢集，但想要清除範例資料，請執行下列命令來捨棄 SALESDB 資料庫。

```
DROP DATABASE SALESDB;
```

如果您未建立 SALESDB 資料庫，或是不想要捨棄資料庫，請執行下列命令來僅捨棄資料表。

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

# Amazon Redshift 概念概觀

Amazon Redshift Serverless 可讓您無需佈建資料倉儲的一般組態，即可存取和分析資料。系統會自動佈建資源，並有智慧地擴展資料倉儲容量，即使是最嚴苛且無法預測的工作負載，也能為其提供快速的效能。資料倉儲閒置時不會產生費用，因此只需按實際用量支付費用。您可以在 Amazon Redshift 查詢編輯器 v2 或您最愛的商業智慧 (BI) 工具中立即載入資料並開始查詢。在易於使用的零管理環境中，享受最優惠的價格效能和熟悉的 SQL 功能。

如果您是第一次使用 Amazon Redshift，建議您從閱讀下列章節開始：

- [Amazon Redshift Serverless 功能概觀](#) - 在本主題中，您可以找到 Amazon Redshift Serverless 及其主要功能的概觀。
- [服務重點和定價](#) - 在此產品詳細資訊頁面上，您可以找到有關 Amazon Redshift Serverless 重點和定價的詳細資訊。
- [開始使用 Amazon Redshift Serverless](#) - 在本主題中，您可以進一步了解如何建立 Amazon Redshift Serverless 資料倉儲，以及使用查詢編輯器 v2 開始查詢資料。

如果您偏好手動管理 Amazon Redshift 資源，可以針對資料查詢需求建立佈建的叢集。如需詳細資訊，請參閱 [Amazon Redshift 叢集](#)。

如果您的組織符合資格，而且您的叢集是在未提供 Amazon Redshift Serverless 的 AWS 區域中建立，您也許可以在 Amazon Redshift 免費試用方案下建立叢集。選擇生產或免費試用來回答這個問題：您打算將此叢集用於什麼目的？選擇免費試用時，您可以使用 dc2.large 節點類型建立組態。如需有關選擇免費試用的詳細資訊，請參閱 [Amazon Redshift 免費試用版](#)。如需有提供 Amazon Redshift Serverless 的 AWS 區域清單，請參閱 Amazon Web Services 一般參考中針對 [Redshift Serverless API](#) 所列出的端點。

以下是 Amazon Redshift Serverless 的一些關鍵概念。

- 命名空間 – 資料庫物件和使用者的集合。命名空間會將您在 Amazon Redshift Serverless 中使用的資源 (例如結構描述、資料表、使用者、資料庫和快照) 群組在一起。
- 工作群組 - 運算資源的集合。工作群組存放 Amazon Redshift Serverless 用來執行運算任務的運算資源。此類資源的一些範例包括 Redshift 處理單元 (RPU)、安全群組、使用限制。工作群組具有網路和安全設定，您可以使用 Amazon Redshift Serverless 主控台、AWS Command Line Interface 或 Amazon Redshift Serverless API 進行設定。

若要取得有關規劃命名空間和工作群組資源的更多資訊，請參閱 [使用命名空間](#) 和 [使用工作群組](#)。

以下是一些關鍵的 Amazon Redshift 佈建的叢集概念：

- 叢集 – Amazon Redshift 資料倉儲的核心基礎設施元件是叢集。

叢集是由一個或多個運算節點所組成。運算節點會執行已編譯的程式碼。

如果為叢集佈建了兩個或多個運算節點，則會有另外的領導節點協調運算節點。領導節點可處理與應用程式的外部通訊，例如商業智慧工具和查詢編輯器。您的用戶端應用程式只會直接和領導節點互動，外部應用程式不會知道運算節點的存在。

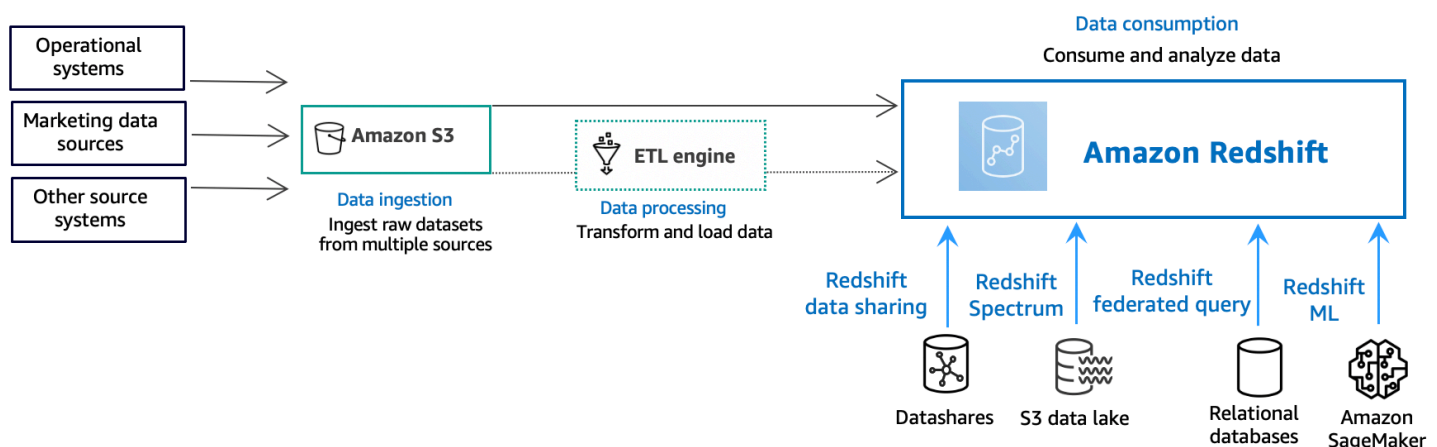
- 資料庫 – 叢集包含一個或多個資料庫。

使用者資料會儲存在運算節點上的一或多個資料庫中。您的 SQL 用戶端會和領導節點進行通訊，然後再由領導節點統籌使用運算節點的查詢執行作業。如需有關運算節點和領導節點的詳細資訊，請參閱[資料倉儲系統架構](#)。在資料庫中，使用者資料會組織成一或多個結構描述。

Amazon Redshift 是關聯式資料庫管理系統 (RDBMS)，因此和其他的 RDBMS 應用程式相容。它提供與典型 RD BMS 相同的功能 (包括線上交易處理 (OLTP) 功能，例如插入和刪除資料)。Amazon Redshift 也針對資料集的高效能批次分析和報告進行了最佳化。

接下來，您可以在 Amazon Redshift 中找到典型資料處理流程的說明，以及流程不同部分的說明。如需 Amazon Redshift 系統架構的詳細資訊，請參閱[資料倉儲系統架構](#)。

下圖說明 Amazon Redshift 中的典型資料處理流程。



Amazon Redshift 資料倉儲是企業級的關聯式資料庫查詢與管理系統。Amazon Redshift 透過多種類型的應用程式來支援用戶端連線，包括商業智慧 (BI)、報告、資料與分析工具。執行分析查詢時，會以多階段操作的方式，擷取、比較和評估大量的資料，來產生最終的結果。

在資料擷取層，不同類型的資料來源會持續將結構化、半結構化或非結構化資料上傳至資料儲存層。此資料儲存區域可做為暫存區，以不同的使用準備狀態儲存資料。儲存的範例可能是 Amazon Simple Storage Service (Amazon S3) 儲存貯體。

在可選資料處理層，來源資料會使用擷取、轉換、載入 (ETL) 或擷取、載入、轉換 (ELT) 管線進行預處理、驗證和轉換。然後使用 ETL 操作對這些原始資料集進行細化。ETL 引擎的一個範例是 AWS Glue。

在資料消耗層，資料會載入您的 Amazon Redshift 叢集，您可以在其中執行分析工作負載。

如需分析工作負載的一些範例，請參閱[查詢外部資料來源](#)。



## 其他資源

如需有關 Amazon Redshift Serverless 的詳細資訊，建議您繼續使用下列 Amazon Redshift 資源，進一步了解本指南中介紹的概念：

- 專題影片：這些影片可協助您了解 Amazon Redshift 功能。
  - 若想概略了解 Amazon Redshift Serverless，請觀看以下影片。[90 秒內解釋 Amazon Redshift Serverless](#)。
  - 若要瞭解如何設定無伺服器資料倉儲並開始查詢資料，請觀看以下影片。[開始使用 Amazon Redshift Serverless](#)。
- [Amazon Redshift 管理指南](#)：本指南建立在本《Amazon Redshift 入門指南》之上。它提供有關建立、管理和監控 Amazon Redshift Serverless 和 Amazon Redshift 佈建叢集的概念和任務的深入資訊。
- [Amazon Redshift 資料庫開發人員指南](#)：本指南也建立在本《Amazon Redshift 入門指南》之上。此指南為資料庫開發人員提供有關設計、建立、查詢及維護資料庫 (組成資料倉儲) 的深入資訊。
  - [SQL 參考](#)：本主題說明 Amazon Redshift 的 SQL 命令和函數參考。
  - [系統資料表和檢視](#)：本主題說明 Amazon Redshift 的系統資料表和檢視。
- Amazon Redshift 教學課程：本主題顯示有關 Amazon Redshift 功能的教學課程。
  - [從 Amazon S3 載入資料](#)：本教學課程說明如何將資料從 Amazon S3 儲存貯體中的資料檔案載入 Amazon Redshift 資料庫資料表。
  - [開始使用資料共用](#)：本節說明如何共用和存取其他 Amazon Redshift 叢集中的資料。
  - [將空間 SQL 函數與 Amazon Redshift 搭配使用](#)：本教學課程示範如何將部分空間 SQL 函數與 Amazon Redshift 搭配使用。
  - [使用 Amazon Redshift Spectrum 查詢巢狀資料](#)：本教學課程說明如何使用 Redshift Spectrum，以使用外部資料表查詢 Parquet、ORC、JSON 和 Ion 檔案格式中的巢狀資料。
  - [設定手動工作負載管理 \(WLM\) 佇列](#)：本教學課程說明如何在 Amazon Redshift 中設定手動工作負載管理 (WLM)。
  - [開始使用 Amazon Redshift ML](#)：本節說明使用者如何使用熟悉的 SQL 命令建立、訓練和部署機器學習模型。
- [新功能](#)：此網頁列出了 Amazon Redshift 的新功能和產品更新。

# 文件歷史紀錄

## Note

如需 Amazon Redshift 中新功能的說明，請參閱[新增功能](#)。

下表說明 Amazon Redshift 入門指南的重要文件變更。

變更	描述	發行日期
文件更新	更新了指南以包含有關開始使用常見資料庫任務、查詢資料湖、查詢遠端來源資料、共用資料，以及使用 Amazon Redshift 資料訓練機器學習模型的新章節。	2021 年 6 月 30 日
新功能	更新了指南以說明新的範例載入程序。	2021 年 6 月 4 日
文件更新	更新了指南以移除原始的 Amazon Redshift 主控台並改善步驟流程。	2020 年 8 月 14 日
新主控台	更新了指南以說明新的 Amazon Redshift 主控台。	2019 年 11 月 11 日
新功能	更新指南來描述快速啟動叢集程序。	2018 年 8 月 10 日
新功能	更新了指南以從 Amazon Redshift 儀表板啟動叢集。	2015 年 7 月 28 日
新功能	更新了指南以使用新的節點類型名稱。	2015 年 6 月 9 日
文件更新	更新了設定 VPC 安全群組的螢幕擷取畫面和程序。	2015 年 4 月 30 日
文件更新	更新了螢幕擷取畫面和程序以符合目前主控台。	2014 年 11 月 12 日
文件更新	已載入中資料從 Amazon S3 資訊移至其本身的區段，並將後續步驟區段移至最終步驟，讓您能更適當地探索文件。	2014 年 5 月 13 日
文件更新	已移除歡迎頁面，並將內容納入主要入門頁面中。	2014 年 3 月 14 日

變更	描述	發行日期
文件更新	這是新版的《Amazon Redshift 入門指南》，描述客戶意見回饋和服務更新。	2014 年 3 月 14 日
新指南	這是《Amazon Redshift 入門指南》的第一版。	2013 年 2 月 14 日

本文為英文版的機器翻譯版本，如內容有任何歧義或不一致之處，概以英文版為準。