



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

無伺服器資料倉儲入門 .....	1
註冊 AWS .....	1
使用 Amazon Redshift Serverless 建立資料倉儲 .....	1
載入範例資料 .....	3
執行範例查詢 .....	6
從 Amazon S3 載入資料 .....	7
佈建資料倉儲入門 .....	15
註冊 AWS .....	17
決定防火牆規則 .....	17
步驟 1：建立範例叢集 .....	18
步驟 2：設定SQL用戶端的傳入規則 .....	20
步驟 3：授予SQL用戶端存取權並執行查詢 .....	21
授予查詢編輯器 v2 的存取權 .....	21
步驟 4：將資料從 Amazon S3 載入到 Amazon Redshift .....	22
使用SQL命令從 Amazon S3 載入資料 .....	22
使用查詢編輯器 v2 從 Amazon S3 載入資料 .....	24
在叢集中建立TICKIT資料 .....	24
步驟 5：使用查詢編輯器嘗試範例查詢 .....	25
步驟 6：重設環境 .....	26
定義和使用資料倉儲中的資料庫 .....	28
連線至 Amazon Redshift .....	28
建立 資料庫 .....	30
建立使用者 .....	30
建立資料架構 .....	31
建立資料表 .....	32
將資料列插入至資料表 .....	33
從資料表選取資料 .....	33
載入資料 .....	34
查詢系統表格和檢視 .....	34
檢視資料表名稱的清單 .....	34
檢視使用者 .....	36
檢視近期查詢 .....	36
確定正在運行的查詢的會話 ID .....	37
取消查詢 .....	37

---

使用超級使用者佇列來取消查詢 .....	39
查詢不在您的 Amazon Redshift 資料庫中的資料 .....	41
查詢資料湖 .....	41
查詢遠端資料來源 .....	42
存取其他資料庫中的資料 .....	42
使用 Redshift 資料訓練機器學習模型 .....	42
了解 Amazon Redshift 概念 .....	44
其他學習資源 .....	47
文件歷史紀錄 .....	48
.....	

# 開始使用 Amazon Redshift Serverless 資料倉儲

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。範例資料會在整個 Amazon Redshift 文件中使用，以示範功能。若要開始使用 Amazon Redshift 佈建的資料倉儲，請參閱 [開始使用 Amazon Redshift 佈建的資料倉儲](#)。

- [the section called “註冊 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 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/註冊>。
2. 請遵循線上指示進行。

當您註冊 AWS 帳戶時，會建立 AWS 帳戶根使用者。根使用者可存取帳戶中的所有 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。

**Note**

Redshift Serverless 需要三個不同可用區域中VPC具有三個子網路的 Amazon。Redshift Serverless 也需要至少 37 個可用的 IP 地址。請確定VPC您用於 Redshift Serverless 的 Amazon 在三個不同的可用區域中有三個子網路，且至少有 37 個可用的 IP 地址，然後再繼續。如需在 Amazon 中建立子網路的詳細資訊VPC，請參閱 Amazon Virtual Private Cloud 使用者指南 中的 [建立子網路](#)。如需 Amazon 中 IP 地址的詳細資訊VPC，請參閱 [VPCs和子網路的 IP 定址](#)。

若要使用預設設定進行設定：

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

選擇嘗試 Redshift Serverless 免費試用。

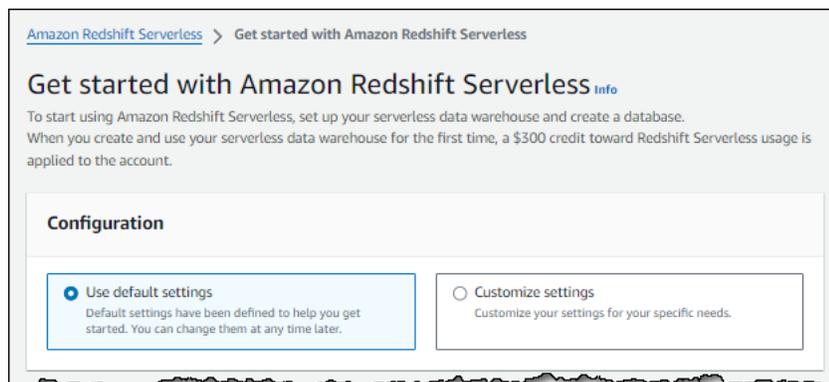
2. 在組態下選擇使用預設設定。Amazon Redshift Serverless 會建立具有與此命名空間關聯之預設工作群組的預設命名空間。選擇 Save configuration (儲存組態)。

**Note**

命名空間是資料庫物件和使用者的集合。命名空間會將您在 Redshift Serverless 中使用的所有資源分組，例如結構描述、資料表、使用者、資料共用和快照。

Workgroup 是運算資源的集合。工作群組包含 Redshift Serverless 用來執行運算任務的運算資源。

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



- 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。

## 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

如果 Redshift Serverless 無法成功建立工作群組，您可以執行下列動作：

- 解決 Redshift Serverless 報告的任何錯誤，例如 Amazon 中的子網路太少 VPC。
- 在 Redshift Serverless 儀表板中選擇預設命名空間，然後選擇動作、刪除命名空間來刪除命名空間。刪除命名空間需要幾分鐘的時間。
- 當您再次開啟 Redshift Serverless 主控台時，會顯示歡迎畫面。

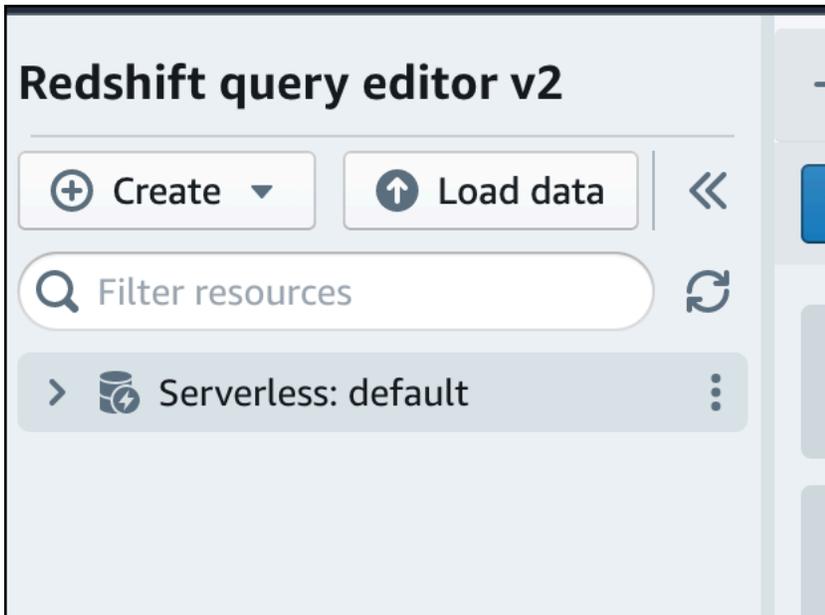
## 載入範例資料

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。

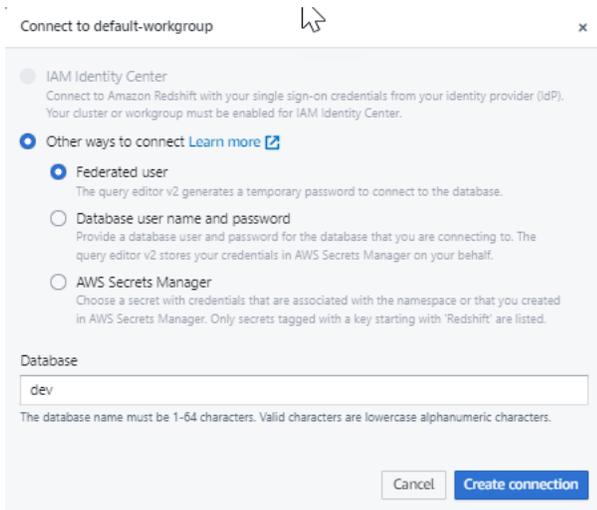
- 若要從 Amazon Redshift Serverless 主控台啟動查詢編輯器 v2，請選擇查詢資料。當您從 Amazon Redshift Serverless 主控台調用查詢編輯器 v2 時，新的瀏覽器索引標籤會開啟並顯示查詢編輯器。查詢編輯器 v2 會從您的用戶端機器連線到 Amazon Redshift Serverless 環境。



2. 在本指南中，您將使用 AWS 管理員帳戶和預設 AWS KMS key。如需設定查詢編輯器 v2 的相關資訊，包括需要哪些許可，請參閱 Amazon Redshift 管理指南 中的 [設定 AWS 帳戶](#) 您的。如需將 Amazon Redshift 設定為使用客戶受管金鑰或變更 Amazon Redshift 使用的 KMS 金鑰的相關資訊，請參閱 [變更命名空間的 AWS KMS 金鑰](#)。
3. 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。

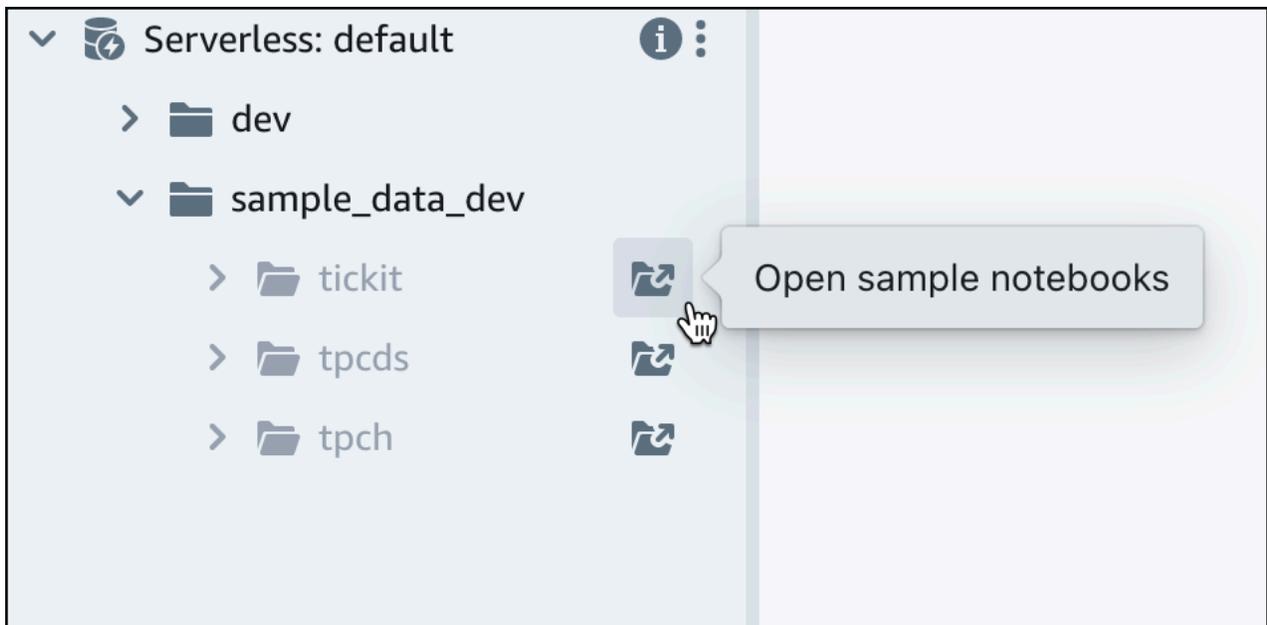


4. 第一次在查詢編輯器 v2 中連線至新工作群組時，您必須選取要用於連線至工作群組的驗證類型。在本指南中，請保持聯合身分使用者為選取狀態，然後選擇建立連線。



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。

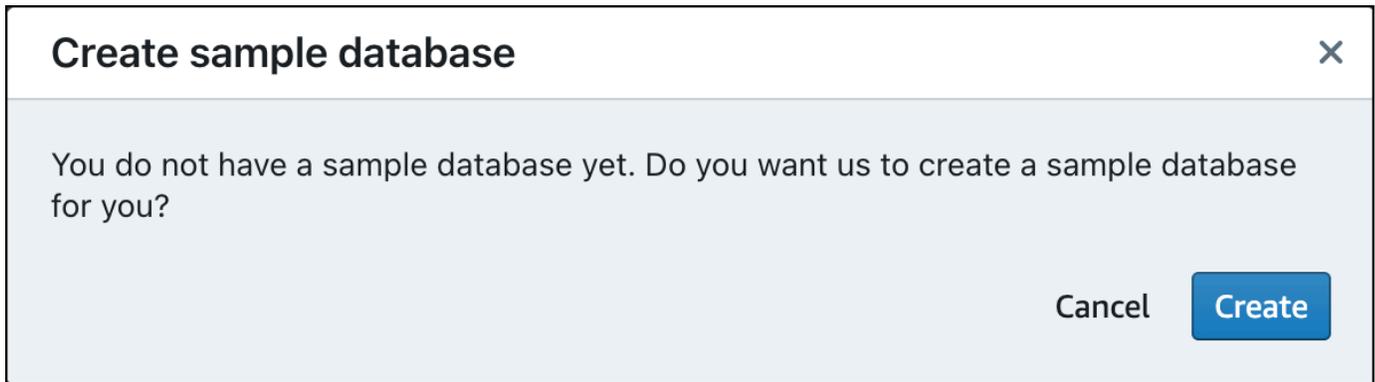
5. 在 Amazon Redshift Serverless 預設工作群組下，展開 `sample_data_dev` 資料庫。有三個範例結構描述與三個範例資料集相對應，您可以將這些範例資料集載入 Amazon Redshift Serverless 資料庫。選擇您要載入的範例資料集，然後選擇開啟範例筆記本。



#### Note

SQL 筆記本是 SQL 和 Markdown 儲存格的容器。您可以使用筆記本在單一文件中組織、註釋和共用多個 SQL 命令。

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 Create (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。

SQL Query:

```

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

```

eventname	total_price
<input type="checkbox"/> Adriana Lecouvreur	51846
<input type="checkbox"/> Janet Jackson	51049
<input type="checkbox"/> Phantom of the Opera	50301
<input type="checkbox"/> The Little Mermaid	49956
<input type="checkbox"/> Citizen Cope	49823
<input type="checkbox"/> Sevendust	48020
<input type="checkbox"/> Electra	47883
<input type="checkbox"/> Mary Poppins	46780
<input type="checkbox"/> Live	46661

Elapsed time: 401 ms 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。

```
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;
```

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許可IAM的角色，並將其連接至無伺服器命名空間，才能從 Amazon S3 載入資料。若要這麼做，請從導覽功能表中選擇命名空間組態，選擇命名空間，然後選擇安全性和加密。然後，選擇管理IAM角色。

## 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. 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 >

<input type="checkbox"/>	<a href="#">IAM roles</a>	Status	Role type
--------------------------	---------------------------	--------	-----------

- 選擇您要授予此角色的 S3 儲存貯體存取層級，然後選擇建立IAM角色作為預設。

## 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.

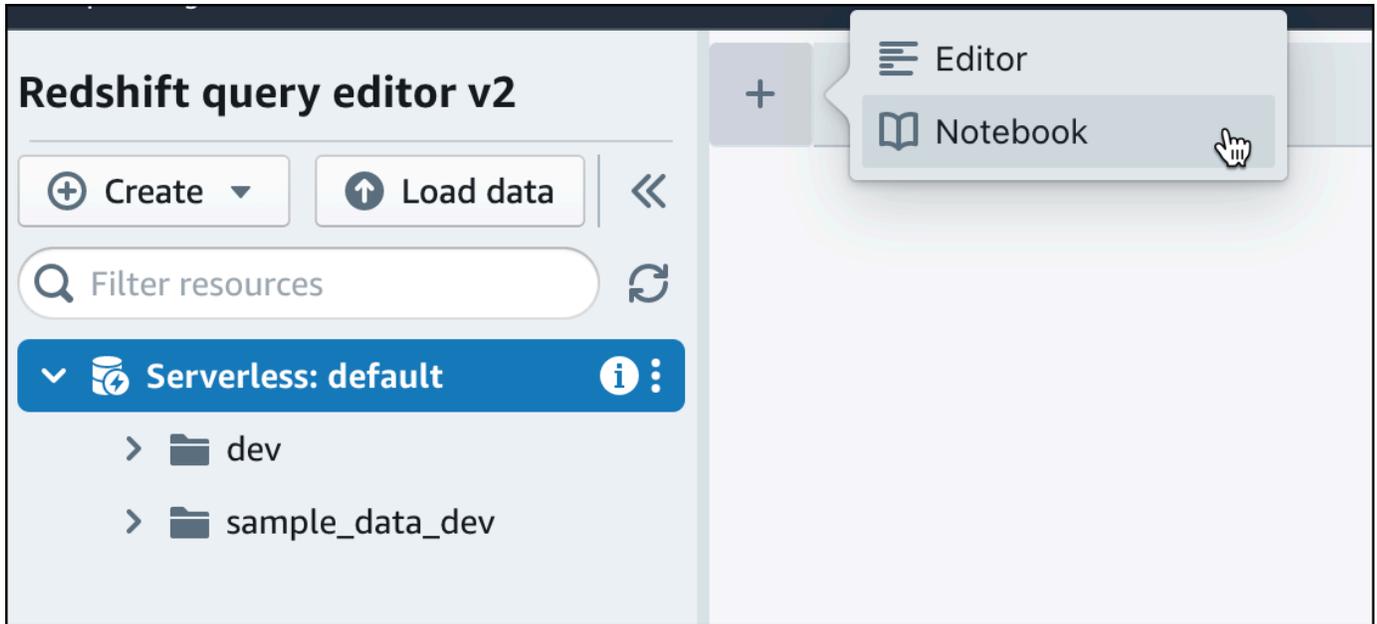
4. 選擇 Save changes (儲存變更)。您現在可以從 Amazon S3 載入範例資料。

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

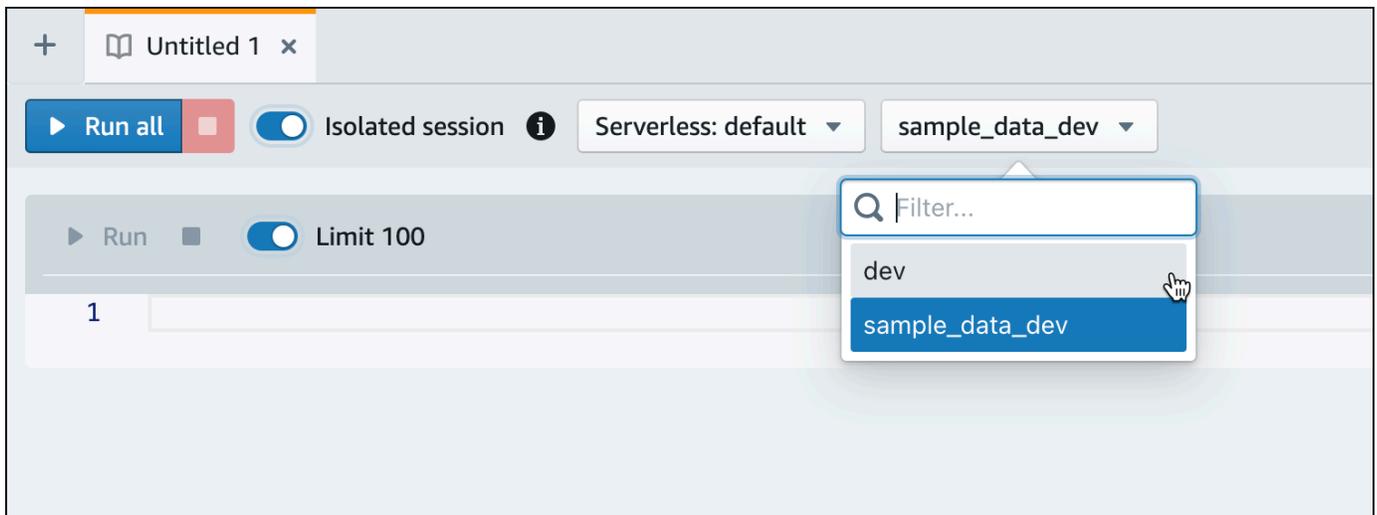
## 從 Amazon S3 載入範例資料

1. 在查詢編輯器 v2 中，選擇 **+** 新增，然後選擇筆記本來建立新的 SQL 筆記本。

新



2. 切換到 dev 資料庫。



3. 建立資料表。

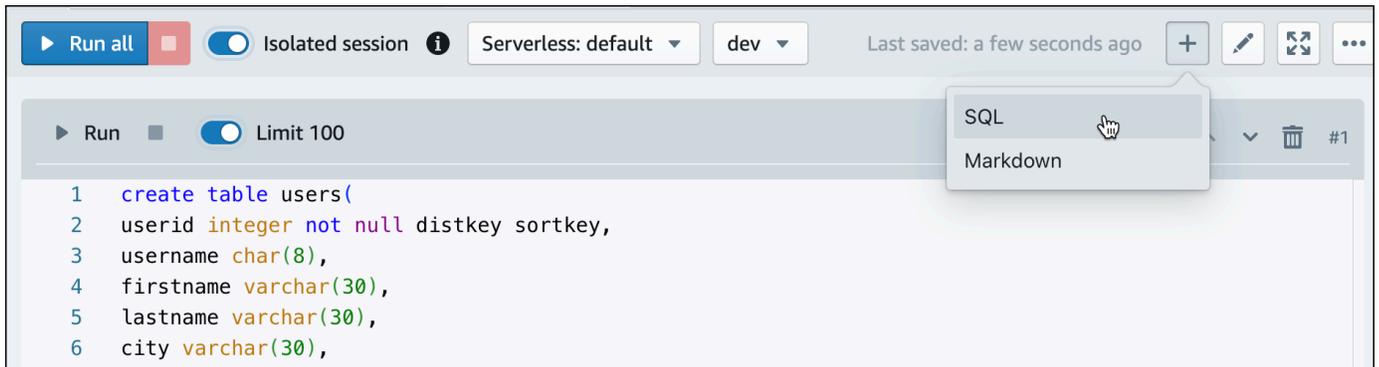
如果您使用查詢編輯器 v2，請複製並執行以下建立資料表陳述式，以在 dev 資料庫中建立資料表。如需語法的詳細資訊，請參閱 Amazon Redshift 資料庫開發人員指南 [CREATE TABLE](#) 中的。

```
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. 現在請在查詢編輯器 v2 中使用 COPY 命令，將大型資料集從 Amazon S3 或 Amazon DynamoDB 載入 Amazon Redshift。如需 COPY 語法的詳細資訊，請參閱 Amazon Redshift 資料庫開發人員指南 [COPY](#) 中的。

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，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 命令的詳細資訊，請參閱 Amazon Redshift 開發人員指南 [SELECT](#) 中的。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 JDBC 和 ODBC 驅動程式，使用第三方 SQL 工具連線至 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 Data API 連線至 Amazon Redshift Serverless。如需詳細資訊，請參閱[使用 Amazon Redshift Data API](#)。
- 您可以使用 Amazon Redshift Serverless 中的資料搭配 Redshift ML，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](#)。

# 開始使用 Amazon Redshift 佈建的資料倉儲

如果您是 Amazon Redshift 的首次使用者，建議您閱讀下列章節，以協助您開始使用佈建叢集。Amazon Redshift 的基本流程是建立佈建的資源、連線至 Amazon Redshift、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 或從 Amazon S3 儲存貯體載入範例資料。範例資料會在整個 Amazon Redshift 文件中使用，以示範功能。

本教學課程示範如何使用 Amazon Redshift 佈建叢集，這些叢集是您管理系統資源 AWS 的資料倉儲物件。您也可以將 Amazon Redshift 與無伺服器工作群組搭配使用，這些工作群組是回應用量自動擴展的資料倉儲物件。若要開始使用 Redshift Serverless，請參閱 [開始使用 Amazon Redshift Serverless 資料倉儲](#)。

建立並登入 Amazon Redshift 佈建主控台後，您可以建立和管理 Amazon Redshift 物件，包括叢集、節點和資料庫。您也可以使用 SQL 用戶端執行查詢、檢視查詢，以及執行其他 SQL 資料定義語言（DDL）和資料操作語言（DML）操作。

## Important

您為此練習佈建的叢集會在即時環境中執行。只要執行，它就會將費用累積到您的 AWS 帳戶。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。本章最後一節說明如何執行此操作。

登入 AWS Management Console 並在 開啟 Amazon Redshift 主控台 <https://console.aws.amazon.com/redshiftv2/>。

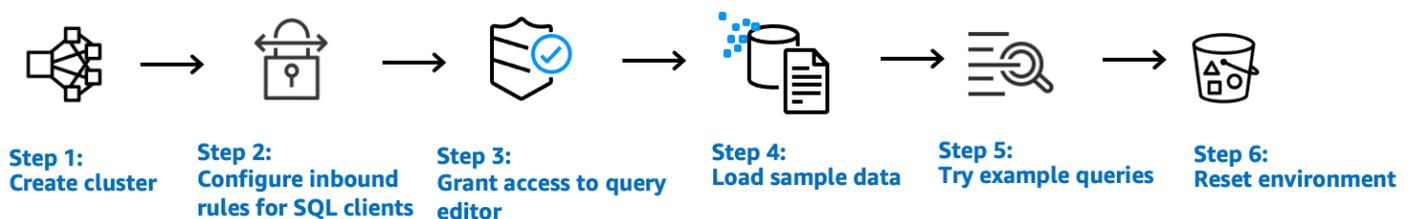
建議您先前往佈建叢集儀表板，開始使用 Amazon Redshift 主控台。

根據您的組態，下列項目會顯示在 Amazon Redshift 佈建主控台的導覽窗格中：

- Redshift Serverless – 無需設定、調整和管理 Amazon Redshift 佈建叢集，即可存取和分析資料。
- 佈建的叢集儀表板 – 檢視 中的叢集清單 AWS 區域、檢查叢集指標和查詢概觀，以取得指標資料（例如 CPU 使用率）和查詢資訊的洞察。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 – 在此 中檢視叢集清單 AWS 區域、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 – 在 Amazon Redshift 叢集上託管的資料庫上執行查詢。我們建議您改用查詢編輯器 v2。
- 查詢編輯器 v2 – Amazon Redshift 查詢編輯器 v2 是獨立的 Web 型 SQL 用戶端應用程式，可在 Amazon Redshift 資料倉儲上撰寫和執行查詢。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，例如最近的查詢清單和每個查詢 SQL 的文字。
- 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共用，消費者帳戶管理員可以將資料共用與帳戶中的整個 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。
- 零ETL整合 – 管理在寫入支援的來源後，讓交易資料可在 Amazon Redshift 中使用的整合。
- IAM Identity Center 連線 – 設定 Amazon Redshift 與 IAM Identity Center 之間的連線。
- 組態 – 透過 Java Database Connectivity ( JDBC ) 和 Open Database Connectivity ( ODBC ) 連線，從 SQL 用戶端工具連線至 Amazon Redshift 叢集。您也可以設定 Amazon Redshift 管理的虛擬私有雲端 ( VPC ) 端點。這樣做會在 VPC 基於的 Amazon VPC 服務之間提供私有連線，其中包含叢集和正在執行用戶端工具 VPC 的另一個叢集。
- AWS 合作夥伴整合 – 與支援的 AWS 合作夥伴建立整合。
- 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 搭配使用之其他工具 AWS 或服務的相關資訊。
- 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。

您將在本教學課程中執行下列步驟。



## 主題

- [註冊 AWS](#)
- [決定防火牆規則](#)
- [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SQL用戶端存取權並執行查詢](#)
- [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](#)

## 註冊 AWS

如果您還沒有 AWS 帳戶，請註冊。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/註冊>。
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，AWS 帳戶根使用者會建立。根使用者有權存取該帳戶中的所有 AWS 服務和資源。作為安全最佳實務，請將管理存取權指派給使用者，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

### Note

本教學假設您的叢集使用預設連接埠 5439，且 Amazon Redshift 查詢編輯器 v2 可用於執行 SQL 命令。它不會進入有關聯網組態的詳細資訊，也不會設定環境中可能需要的 SQL 用戶端。

在某些環境中，您可以在啟動 Amazon Redshift 叢集時指定連接埠。您可以使用此連接埠和叢集的端點 URL 來存取叢集。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，您可以從 SQL 用戶端工具連線至叢集並執行查詢。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。

## 步驟 1：建立範例 Amazon Redshift 叢集

在本教學課程中，您會逐步解說使用資料庫建立 Amazon Redshift 叢集的程序。然後將資料集從 Amazon S3 載入至資料庫中的資料表。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，請確定您已完成任何必要的先決條件，例如 [註冊 AWS](#) 和 [設定防火牆規則](#)。

對於從其他 AWS 資源存取資料的任何操作，您的叢集需要代表您存取資源和資源資料的權限。例如，使用 SQLCOPY 命令從 Amazon Simple Storage Service ( Amazon S3 ) 載入資料。您可以使用 AWS Identity and Access Management ( ) 提供這些許可 IAM。您可以透過您建立並連接至叢集 IAM 的角色來執行此操作。如需憑證和存取許可的詳細資訊，請參閱 Amazon Redshift 資料庫開發人員指南 中的 [憑證和存取許可](#)。

### 建立 Amazon Redshift 叢集

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

#### Important

如果您使用 IAM 使用者憑證，請確定您具有執行叢集操作所需的許可。如需詳細資訊，請參閱 [Amazon Redshift 管理指南](#) 中的 Amazon Redshift 安全。

2. 在 AWS 主控台上，選擇您要建立叢集 AWS 區域的。
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 AWS 區域 支援 RA3 節點類型。如需支援 RA3 節點類型的清單 AWS 區域，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。這樣做會開啟大小計算器，詢問您有關計劃存放在資料倉儲中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。

針對本教學課程，請針對節點類型選擇 `ra3.4xlarge`，針對節點數量選擇 `2`。

如果可用 AZ 組態的選項，請選擇單一 AZ。

- 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。

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。
- 在中管理管理員憑證 AWS Secrets Manager – Amazon Redshift 使用 AWS Secrets Manager 來產生和管理管理員密碼。使用 AWS Secrets Manager 產生和管理密碼的秘密會產生費用。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`。
- 管理使用者密碼：輸入 `Changeit1` 以取得密碼。

6. 針對本教學課程，請建立 IAM 角色，並將其設定為叢集的預設值，如下所述。叢集只能有一個預設 IAM 角色集。

a. 在叢集許可下，針對管理 IAM 角色，選擇建立 IAM 角色。

b. 為 IAM 角色指定 Amazon S3 儲存貯體，以便透過下列其中一種方法存取：

- 選擇無其他 Amazon S3 儲存貯體，以允許建立 IAM 的角色僅存取名為 `redshift` 的 Amazon S3 儲存貯體。
- 選擇任何 Amazon S3 儲存貯體，以允許建立 IAM 的角色存取所有 Amazon S3 儲存貯體。
- 選擇特定 Amazon S3 儲存貯體，為建立 IAM 的角色指定一或多個 Amazon S3 儲存貯體以供存取。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。Amazon Redshift 會自動建立角色，並將該IAM角色設定為叢集的預設值。

由於您從主控台建立IAM角色，因此已連接 AmazonRedshiftAllCommandsFullAccess 政策。這可讓 Amazon Redshift 複製、載入、查詢和分析您IAM帳戶中 Amazon 資源的資料。

如需如何管理叢集預設IAM角色的資訊，請參閱 [Amazon Redshift 管理指南 中的將IAM角色建立為Amazon Redshift 的預設角色](#)。

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，您可能會使用載入範例資料選項建立叢集，並想要開啟增強的 Amazon VPC路由。如果是這樣，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 中設定網路地址轉譯 VPC（NAT）地址，讓叢集存取網際網路。或者，您可以在 中設定 Amazon S3 VPC端點VPC。如需增強型 Amazon VPC路由的詳細資訊，請參閱 Amazon Redshift 管理指南 中的 [增強型 Amazon VPC路由](#)。

8. 選擇建立叢集。在叢集頁面上等待建立Available狀態為 的叢集。

## 步驟 2：設定SQL用戶端的傳入規則

### Note

建議您略過此步驟，並使用 Amazon Redshift 查詢編輯器 v2 存取叢集。

在本教學課程稍後，您可以根據 Amazon VPC服務，從虛擬私有雲端（VPC）內存取叢集。不過，如果您使用來自防火牆外部的SQL用戶端來存取叢集，請務必授予傳入存取。

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。

如需防火牆規則的詳細資訊，請參閱 Amazon EC2使用者指南 中的 [安全群組規則](#)。

- 若要從 Amazon EC2外部用戶端存取，請將傳入規則新增至連接到叢集的安全群組，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，CIDR/IP 為 192.0.2.0/24 可讓該 IP 地址範圍內的用戶端連線至您的叢集。找出適合您環境的正確 CIDR/IP。

## 步驟 3：授予SQL用戶端存取權並執行查詢

若要查詢 Amazon Redshift 叢集託管的資料庫，您有多種SQL用戶端選項。其中包含：

- 連線至叢集，並使用 Amazon Redshift 查詢編輯器 v2 執行查詢。

如果您使用查詢編輯器 v2，則不需要下載和設定SQL用戶端應用程式。您可以從 Amazon Redshift 主控台啟動 Amazon Redshift 查詢編輯器 v2。

- 使用 連線至您的叢集RSQL。如需詳細資訊，請參閱 [Amazon Redshift 管理指南 中的與 Amazon Redshift 連線RSQL](#)。
- 透過 SQL Workbench/J 等SQL用戶端工具連線至叢集。如需詳細資訊，請參閱 Amazon Redshift 管理指南 中的 [使用 SQL Workbench/J 連線至叢集](#)。

本教學課程使用 Amazon Redshift 查詢編輯器 v2 作為在 Amazon Redshift 叢集託管的資料庫上執行查詢的簡單方法。建立叢集後，您可以立即執行查詢。如需使用 Amazon Redshift 查詢編輯器 v2 時考量的詳細資訊，請參閱 Amazon Redshift 管理指南 中的 [使用查詢編輯器 v2 時的考量](#)。

### 授予查詢編輯器 v2 的存取權

管理員第一次為您的 設定查詢編輯器 v2 時 AWS 帳戶，會選擇用來加密查詢編輯器 v2 資源 AWS KMS key 的。Amazon Redshift 查詢編輯器 v2 資源包括已儲存的查詢、筆記本和圖表。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 Amazon Redshift 管理指南 中的 [設定 AWS 帳戶](#)。

若要存取查詢編輯器 v2，您需要獲得許可。管理員可以將 Amazon Redshift 查詢編輯器 v2 的其中一個 AWS 受管政策連接至IAM角色或使用者，以授予許可。這些 AWS 受管政策的撰寫方式具有不同的選項，可控制標記資源如何允許共用查詢。您可以使用IAM主控台 (<https://console.aws.amazon.com/iam/>) 連接IAM政策。如需這些政策的詳細資訊，請參閱 Amazon Redshift 管理指南 中的 [存取查詢編輯器 v2](#)。

您也可以根據所提供受控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主控台政策編輯器來建立自己的政策，請在視覺化編輯器中選擇 SQL Workbench 作為您為其建立政策的服務。查詢編輯器 v2 使用視覺化編輯器和IAM政策模擬器中的服務名稱 AWS SQL Workbench。

如需詳細資訊，請參閱《Amazon Redshift 管理指南》中的[使用查詢編輯器第 2 版](#)。

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

建立叢集後，您可以將 Amazon S3 中的資料載入資料庫資料表。有多種方式可從 Amazon S3 載入資料。

- 您可以使用 SQL 用戶端來執行 SQL CREATE TABLE 命令，在資料庫中建立資料表，然後使用 SQLCOPY命令從 Amazon S3 載入資料。Amazon Redshift 查詢編輯器 v2 是SQL用戶端。
- 您可以使用 Amazon Redshift 查詢編輯器 v2 載入精靈。

本教學課程示範如何使用 Amazon Redshift 查詢編輯器 v2 對CREATE資料表和COPY資料執行SQL命令。從 Amazon Redshift 主控台導覽窗格啟動查詢編輯器 v2。在查詢編輯器 v2 中，建立與以管理員使用者命名dev的examplecluster叢集和資料庫的連線awsuser。在本教學課程中，當您建立連線時，請使用資料庫使用者名稱選擇暫時登入資料。如需使用 Amazon Redshift 查詢編輯器 v2 的詳細資訊，請參閱 [Amazon Redshift 管理指南 中的連線至 Amazon Redshift 資料庫](#)。

### 使用SQL命令從 Amazon S3 載入資料

在查詢編輯器 v2 查詢編輯器窗格中，確認您已連線至examplecluster叢集和dev資料庫。接下來，在資料庫中建立資料表，並將資料載入資料表。在此教學課程中，您載入的資料可在可從許多 存取的 Amazon S3 儲存貯體中使用 AWS 區域。

下列程序會建立資料表，並從公有 Amazon S3 儲存貯體載入資料。

使用 Amazon Redshift 查詢編輯器 v2 複製並執行下列建立資料表陳述式，以在dev資料庫的public結構描述中建立資料表。如需語法的詳細資訊，請參閱 Amazon Redshift 資料庫開發人員指南 [CREATE TABLE](#) 中的 。

使用查詢編輯器 v2 等SQL用戶端建立和載入資料

1. CREATE 對sales資料表執行下列SQL命令。

```
drop table if exists sales;
create table sales(
salesid integer not null,
listid integer not null distkey,
sellerid integer not null,
```

```
buyerid integer not null,  
eventid integer not null,  
dateid smallint not null sortkey,  
qtysold smallint not null,  
pricepaid decimal(8,2),  
commission decimal(8,2),  
saletime timestamp);
```

2. CREATE 對date資料表執行下列SQL命令。

```
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. 使用 COPY命令從 Amazon S3 載入sales資料表。

#### Note

我們建議您使用 COPY命令，從 Amazon S3 將大型資料集載入 Amazon Redshift。如需 COPY語法的詳細資訊，請參閱 Amazon Redshift 資料庫開發人員指南 [COPY](#) 中的。

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。當您在建立default叢集時選擇建立IAM角色作為預設角色時，您可以參考您為叢集建立和設定為的角色來提供身分驗證。 IAM

使用下列SQL命令載入sales資料表。您可以選擇性地從 Amazon S3 下載和檢視[sales資料表的來源資料](#)。

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

```
IAM_ROLE default;
```

4. 使用下列SQL命令載入date資料表。您可以選擇性地從 Amazon S3 下載和檢視[date資料表的來源資料](#)。

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

## 使用查詢編輯器 v2 從 Amazon S3 載入資料

本節說明將您自己的資料載入 Amazon Redshift 叢集。查詢編輯器 v2 可在使用載入資料精靈時簡化載入資料。查詢編輯器 v2 Load 資料精靈中產生和使用的COPY命令支援COPY命令語法可用的許多參數，以從 Amazon S3 載入資料。如需從 Amazon S3 複製負載的COPY命令及其選項的相關資訊，請參閱 Amazon Redshift 資料庫開發人員指南 中的 [COPY Amazon Simple Storage Service](#)。

若要將您自己的資料從 Amazon S3 載入至 Amazon Redshift，Amazon Redshift 需要具有所需權限 IAM 的角色，才能從指定的 Amazon S3 儲存貯體載入資料。

若要將您自己的資料從 Amazon S3 載入至 Amazon Redshift，您可以使用查詢編輯器 v2 載入資料精靈。如需有關如何使用載入資料精靈的資訊，請參閱 Amazon Redshift [管理指南](#) 中的 [從 Amazon S3 載入資料](#)。

## 在叢集中建立TICKIT資料

TICKIT 是範例資料庫，您可以選擇性地載入 Amazon Redshift 叢集，以了解如何在 Amazon Redshift 中查詢資料。您可以建立整組TICKIT資料表，並以下列方式將資料載入叢集：

- 當您在 Amazon Redshift 主控台中建立叢集時，您可以選擇同時載入範例TICKIT資料。在 Amazon Redshift 主控台上，選擇叢集，建立叢集。在範例資料區段中，選取在叢集建立期間，載入範例資料 Amazon Redshift 會自動將其範例資料集載入您的 Amazon Redshift 叢集dev資料庫。
- 若要連線至現有的叢集，請執行下列動作：
  - 在 Amazon Redshift 主控台中，從導覽列中選擇叢集。
  - 從叢集窗格中選擇叢集。
  - 選擇查詢資料，查詢編輯器 v2 中的查詢。

- 在資源清單中展開範例叢集。如果這是您第一次連線到叢集，則會顯示連線至範例叢集。選擇資料庫使用者名稱和密碼。將資料庫保留為 **dev**。awsuser 為使用者名稱和密碼指定 **Changeit1**。
- 選擇建立連線。
- 使用 Amazon Redshift 查詢編輯器 v2，您可以將TICKIT資料載入名為 sample\_data\_dev 的範例資料庫。在資源清單中選擇 sample\_data\_dev 資料庫。在標記節點旁，選擇開啟範例筆記本圖示。確認您要建立範例資料庫。
- Amazon Redshift 查詢編輯器 v2 會建立範例資料庫，以及名為的範例筆記本tickit-sample-notebook。您可以選擇執行全部以執行此筆記本，以查詢範例資料庫中的資料。

若要檢視TICKIT資料的詳細資訊，請參閱 Amazon Redshift 資料庫開發人員指南 中的[範例](#)資料庫。

## 步驟 5：使用查詢編輯器嘗試範例查詢

若要設定並使用 Amazon Redshift 查詢編輯器 v2 查詢資料庫，請參閱Amazon Redshift 管理指南中的[使用查詢編輯器 v2](#)。

現在，嘗試一些範例查詢，如下所示。若要在查詢編輯器 v2 中建立新查詢，請選擇查詢窗格右上角的 + 圖示，然後選擇 SQL。新的查詢頁面隨即出現，您可以在其中複製和貼上下列SQL查詢。

### Note

請務必先在筆記本中執行第一個查詢，這會使用下列SQL命令將search\_path伺服器組態值設定為tickit結構描述：

```
set search_path to tickit;
```

如需使用 SELECT命令的詳細資訊，請參閱 [SELECT](#) Amazon Redshift 資料庫開發人員指南中的。

```
-- 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;
```

## 步驟 6：重設環境

在上述步驟中，您已成功建立 Amazon Redshift 叢集、將資料載入資料表，以及使用 Amazon Redshift 查詢編輯器 v2 等 SQL 用戶端查詢資料。

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。

不過，如果您打算嘗試其他 Amazon Redshift 指南中的任務或 中所述的任務，您可能想要讓範例叢集繼續執行 [執行指令以定義和使用資料倉儲中的資料庫](#)。

### 刪除叢集

1. 登入 AWS Management Console 並在 開啟 Amazon Redshift 主控台 <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)。刪除範例叢集？頁面隨即出現。
4. 確認要刪除的叢集，取消選取建立最終快照設定，然後輸入 **delete** 以確認刪除。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[了解有關 Amazon Redshift 的其他資源](#)中找到 Amazon Redshift 的詳細資訊及後續步驟。

# 執行指令以定義和使用資料倉儲中的資料庫

Redshift 無伺服器資料倉儲和 Amazon Redshift 提供的資料倉儲都包含資料庫。啟動資料倉儲之後，您可以使用SQL指令來管理大部分的資料庫動作。除了少數例外情況之SQL外，所有 Amazon Redshift 資料庫的功能和語法都相同。如需 Amazon Redshift 可用SQL命令的詳細資訊，請參閱亞 Amazon Redshift 資料庫開發人員指南中的[SQL命令](#)。

當您建立資料倉儲時，在大多數情況下，Amazon Redshift 也會建立預設資料dev庫。連線到dev資料庫之後，您可以建立另一個資料庫。

以下各節將逐步介紹使用 Amazon Redshift 資料庫時的常見資料庫任務。這些任務始於創建一個數據庫，如果你繼續最後一個任務，你可以通過刪除數據庫刪除你創建的所有資源。

本節中的範例假設如下情況：

- 您已經建立了一個 Amazon Redshift 資料倉儲。
- 您已經從SQL用戶端工具 (例如 Amazon Redshift 查詢編輯器 v2) 建立到資料倉儲的連線。如需查詢編輯器 v2 的詳細資訊，請參閱亞馬遜 Redshift 管理指南中的[使用 Amazon Redshift 查詢編輯器 v2 查詢資料庫](#)。

## 主題

- [連接到 Amazon Redshift 資料倉儲](#)
- [建立 資料庫](#)
- [建立使用者](#)
- [建立資料架構](#)
- [建立資料表](#)
- [載入資料](#)
- [查詢系統表格和檢視](#)
- [取消查詢](#)

## 連接到 Amazon Redshift 資料倉儲

若要 Connect 到 Amazon Redshift 叢集，請從 Amazon Redshift 主控台叢集頁面展開連線到 Amazon Redshift 叢集，然後執行下列其中一個動作：

- 選擇查詢資料以使用查詢編輯器 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 Amazon Redshift 管理[指南中的使用 Amazon Redshift 查詢編輯器 v2 查詢資料庫](#)。

- 在「使用您的用戶端工具」中，選擇您的叢集，然後JDBC使用或驅動程式從用戶端工具透過複製JDBC或ODBC驅動程式連接到 Amazon Redshift。ODBC URL URL從您的用戶端電腦或執行個體使用此功能。編寫應用程式以使用JDBC或ODBC資料存取API作業，或使用支援JDBC或的用戶端工具ODBC。

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，您可以選擇您的JDBC或ODBC驅動程式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詳細資訊，請參閱[設定JDBC驅動程式 2.0 版連線](#)。

如需如何設定ODBC連線的詳細資訊，請參閱[設定ODBC連線](#)。

若要連接到 Redshift 無伺服器資料倉儲，請從 Amazon Redshift 主控台無伺服器儀表板頁面執行下列其中一個動作：

- 使用 Amazon Redshift 查詢編輯器 v2，在 Redshift 無伺服器資料倉儲託管的資料庫上執行查詢。建立資料倉儲之後，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](#)。

- 透過複製或驅動程式，從您的用戶端工具JDBC或ODBC驅動程式 Connect 到 Amazon Redshift。JDBC ODBC URL

若要使用資料倉儲中的資料，您需要JDBC或ODBC驅動程式才能從用戶端電腦或執行個體進行連線。編寫應用程式以使用JDBC或ODBC資料存取API作業，或使用支援JDBC或的用戶端工具ODBC。

如需有關如何尋找連線字串的詳細資訊，請參閱 Amazon Redshift [管理指南中的「連線至無伺服器」](#)。

## 建立 資料庫

確認資料倉儲已啟動並執行之後，您可以建立資料庫。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。數據倉庫可以託管多個數據庫。例如，您可以擁有名為銷售資料的資料庫，以SALESDB及在相同資料倉儲ORDERSDB中命名的訂單資料庫。

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

```
CREATE DATABASE salesdb;
```

### Note

執行命令之後，請務必重新整理資料倉儲中物件的SQL用戶端工具清單，以查看新物件salesdb。

對於本練習，接受預設值。如需更多命令選項的相關資訊，請參閱 Amazon Redshift 資料庫開發人員指南[CREATEDATABASE](#)中的。若要刪除資料庫及其內容，請參閱 Amazon Redshift 資料庫開發人員指南[DROPDATABASE](#)中的。

建立SALESDB資料庫之後，您可以從SQL用戶端連線到新的資料庫。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。

## 建立使用者

依預設，只有您在啟動資料倉儲時建立的管理員使用者才能存取資料倉儲中的預設資料庫。若要授予其他使用者存取權，請建立一個或多個帳戶。資料庫使用者帳戶在資料倉儲中的所有資料庫中都是全域的，而不是每個個別資料庫。

使用CREATEUSER指令建立新使用者。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。

如需其他命令選項的相關資訊，請參閱 Amazon Redshift 資料庫開發人員指南 [CREATEUSER](#) 中的。

## 建立資料架構

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、檢視表和使用使用者定義函數 (UDFs) 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。每個結構描述都可以儲存不同的資料庫物件，例如資料表、檢視表和使用使用者定義函數 (UDFs)。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相關資訊，請參閱 Amazon Redshift 資料庫開發人員指南 [CREATESHEMA](#) 中的。

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。

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}

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此陳述式將結構描述的權限授與使用者。

下列範例會授與使用GUEST者使用SELECT陳述式從SALES結構描述中所有資料表或檢視表選取資料的權限。

```
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與使GUEST用者。

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

## 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。在建立表格時指定欄資訊。

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)
);
```

根據預設，新資料庫物件 (例如資料表) 會在建立資料倉儲期間public建立名為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](#)。

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 v2。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[查詢系統表格和檢視](#)。

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

該 [INSERT](#) 命令將行插入到表中。對於標準大量負載，請使用 [COPY](#) 指令。如需詳細資訊，請參閱 [使用 COPY 命令載入資料](#)。

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');
```

若要將資料插入特定結構定義中的資料表，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

建立資料表並將資料填入資料之後，請使用 SELECT 陳述式來顯示資料表中包含的資料。該 SELECT \* 語句返回一個表中的所有數據的所有列名和行值。使用 SELECT 是驗證最近添加的數據是否正確插入到表中的好方法。

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陳述式查詢資料表的詳細資訊，請參閱[SELECT](#)。

## 載入資料

本指南中的許多範例都使用範TICKIT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。然後，您可以將範例資料上傳到自己的 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命令載入包含存放在 Amazon S3 儲存貯體中的範例資料的表格。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](#)。

## 查詢系統表格和檢視

除了您建立的資料表之外，您的資料倉儲還包含許多系統資料表和檢視表。這些表格和檢視包含有關安裝的資訊，以及系統上執行的各種查詢和程序。您可以查詢這些系統資料表和檢視表，以收集資料庫的相關資訊。如需詳細資訊，請參閱 Amazon Redshift 資料庫開發人員指南中的[系統資料表和檢視參考](#)資料。每個資料表或檢視表的說明會指出表格對所有使用者都可以看見，還是只有超級使用者才能看見。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。

## 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 系統目錄表格。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
```

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
```

```
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

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。

```
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

如需詳細資訊，請參閱 [PG\\_TABLE\\_DEF](#)。

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

## 檢視使用者

您可以查詢 PG\_USER 目錄以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	*****		

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。您可以透過新增至 SELECT 陳述式來篩選查詢，以僅顯示使用者定義的 where usesysid > 1 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	*****		

## 檢視近期查詢

在上一個範例中，使用者識別碼 (使用者識別碼) adminuser 為 100。若要列出最近執行的四個查詢 adminuser，您可以查詢 SYS\_QUERY\_HISTORY 檢視。

您可以使用此檢視來尋找最近執行之查詢的查詢識別碼 (query\_id) 或處理程序識別碼 (session\_id)。您也可以使用此檢視來檢查查詢完成所花費的時間。SYS\_QUERY\_HISTORY 包含查詢字串 (query\_text) 的前 4,000 個字元，可協助您尋找特定查詢。在陳述式中使用 LIMIT 子句來限制結果。

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

**LIMIT 4;**

結果類似如下所示。

```

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);

```

## 確定正在運行的查詢的會話 ID

若要擷取有關查詢的系統資料表資訊，您可能需要指定與該查詢相關聯的工作階段 ID (處理序 ID)。或者，您可能需要為仍在執行的查詢尋找工作階段 ID。例如，如果您需要取消在佈建的叢集上執行時間太長的查詢，則需要工作階段 ID。您可以查詢 STV\_RECENTS 系統資料表以取得執行查詢的工IDs 作階段清單，以及對應的查詢字串。如果您的查詢傳回多個工作階段，您可以查看查詢文字，以判斷您需要的工作階段 ID。

若要判斷執行中查詢的工作階段 ID，請執行下列SELECT陳述式。

```

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

```

## 取消查詢

如果您執行的查詢時間過長或耗用過多資源，請取消查詢。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。下面的查詢從SALES表和USERS表中選擇數據，並通過匹配SELLERID和USERID在WHERE子句中連接兩個表。

```

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。

假設您忘記了輸入WHERE條款。

```

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。

若要取消執行中的查詢，請使用具有查詢工作階段 ID 的CANCEL命令。使用 Amazon Redshift 查詢編輯器 v2，您可以在查詢執行時選擇取消按鈕來取消查詢。

若要尋找工作階段 ID，請啟動新的工作階段並查詢 STV\_RECENTS 資料表，如上一個步驟所示。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。

若要判斷執行中查詢的工作階段 ID，請執行下列SELECT陳述式。

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

結果類似如下所示。

```
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 ...
```

若要取消具有工作階段 ID 的查詢1073791534，請執行下列命令。

```
CANCEL 1073791534;
```

#### Note

命令CANCEL不會停止交易。若要停止或回復交易，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。

如果您取消的查詢與交易相關聯，請使用ABORT或ROLLBACK命令取消交易，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。

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。

如需取消查詢的詳細資訊，請參閱 Amazon Redshift 資料庫開發人員指南[CANCEL](#)中的。

## 使用超級使用者佇列來取消查詢

如果您目前的工作階段有太多同時執行的查詢，您可能無法執行CANCEL命令，直到另一個查詢完成。在這種情況下，請使用不同的工作負載管理查詢佇列執行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，請登入超級使用者，並使用指令將查詢群組設定為「超級使用者」。SET執行命令後，請使用指RESET令重設查詢群組。

若要使用超級使用者佇列取消查詢，請執行這些命令。

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

## 查詢不在您的 Amazon Redshift 資料庫中的資料

接下來，您可以找到有關如何開始查詢遠端來源上的資料的資訊，包括 Amazon S3 資料、遠端資料庫管理員、遠端 Amazon Redshift 資料庫，以及使用 Amazon Redshift 訓練機器學習 (ML) 模型。

### 主題

- [查詢您的資料湖](#)
- [查詢遠端資料庫管理員的資料](#)
- [存取其他 Amazon Redshift 資料庫中的資料](#)
- [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) 而設計的 SQL 功能。您可以查詢多種格式的資料，包括實木地板 ORC、RCFile、TextFile、SequenceFile、RegexSerde、CSV、開啟和 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。

在資料目錄中註冊並啟用 AWS Glue 資料之後 AWS Lake Formation，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 叢集存取外部表格，並從同一 AWS 區域中的任何叢集查詢 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 Redshift Spectrum 的相關資訊，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](#)。

## 查詢遠端資料庫管理員的資料

您可以使用聯合查詢將來自亞馬遜資料RDS庫和 Amazon Aurora 資料庫的資料與 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：

- [開始使用聯合查詢進行 Postgre SQL](#)
- [開始使用聯合查詢至 \[我的\] SQL](#)

## 存取其他 Amazon Redshift 資料庫中的資料

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多 up-to-date 且一致的資訊。您可以在不同層次共用資料，例如資料庫、綱要、表格、視觀表 (包括一般、後期繫結和具體化視觀表)，以及SQL使用者定義函數 (UDFs)。

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 (BIETL) 或分析叢集共用資料的中央擷取、轉換和載入 () 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 [中的管理資料共用任務](#)。

## 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 查詢中呼叫它們以進行預測。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 可讓使用 SQL 者更輕鬆地使用熟悉的 SQL 命令來建立、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 自動 SageMaker 駕駛儀訓練模型，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 Amazon Redshift ML 的相關資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 [開始使用 Amazon Redshift ML](#)。

# 了解 Amazon Redshift 概念

Amazon Redshift Serverless 可讓您在沒有所佈建資料倉儲的所有組態的情況下，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。在零管理環境中 easy-to-use 享受最佳價格效能和熟悉 SQL 的功能。

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 資料倉儲](#) - 在本主題中，您可以進一步了解如何建立 Amazon Redshift Serverless 資料倉儲，以及如何使用查詢編輯器 v2 開始查詢資料。

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

如果您的組織符合資格，且叢集是在無法使用 Amazon Redshift Serverless AWS 區域的中建立，則您可能可以在 Amazon Redshift 免費試用版計畫下建立叢集。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 Amazon Redshift Serverless 可用 AWS 區域 位置的清單，請參閱 [API 中的 Redshift Serverless](#) 列出的 Amazon Redshift 端點 Amazon Web Services 一般參考。

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 Processing Units ( RPU )、安全群組、用量限制。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 進行設定 APIs。

如需取得規劃命名空間和工作群組資源的相關資訊，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。

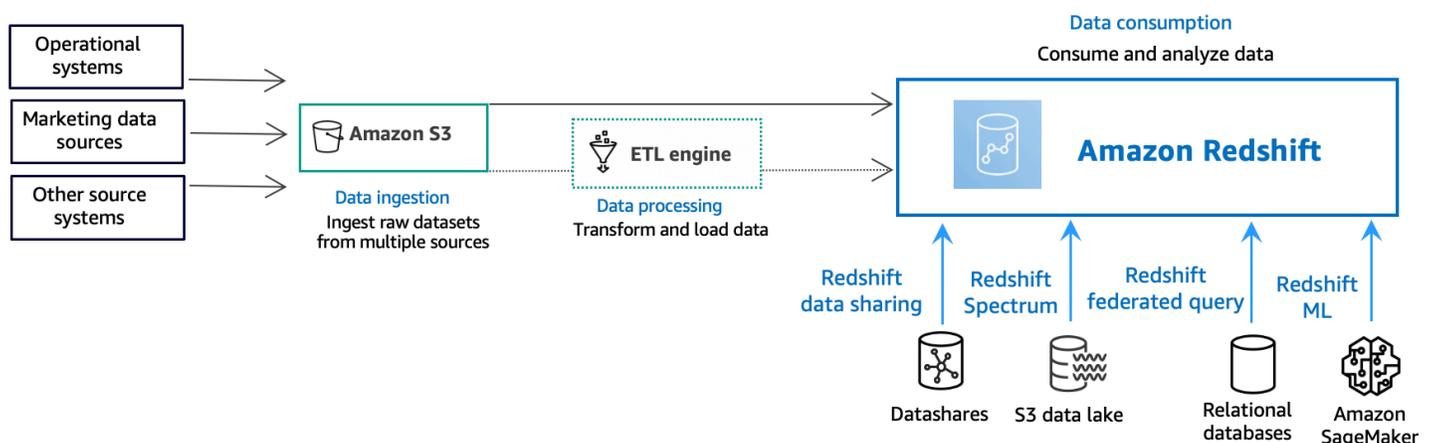
- 資料庫 – 叢集包含一個或多個資料庫。

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用戶端會與領導者節點通訊，進而協調與運算節點執行的查詢。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），與其他RDBMS應用程式相容。它提供與一般相同的功能RDBMS，包括線上交易處理（OLTP）函數，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。

在選用的資料處理層，來源資料會使用擷取、轉換、載入 ( ETL ) 或擷取、載入、轉換 ( ELT ) 管道進行預先處理、驗證和轉換。然後，這些原始資料集會使用 ETL 操作進行精簡。ETL 引擎的範例為 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](#)。

## 了解有關 Amazon Redshift 的其他資源

如需 Amazon Redshift Serverless 的相關資訊，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 參考資料](#)：本主題說明 Amazon Redshift 的 SQL 命令和函數參考。
  - [系統資料表和檢視參考資料](#)：本主題說明 Amazon Redshift 的系統表格和檢視。
- 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 函數與亞馬遜紅移搭配使用](#)：本教學課程示範如何搭配 Amazon Redshift 使用部分空間 SQL 函數。
  - 使用 [Amazon Redshift Spectrum 查詢巢狀資料](#)：本教學課程說明如何使用紅移頻譜，使用外部資料表查詢實木複合地板、ORCJSON、和 Ion 檔案格式的巢狀資料。
  - [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。