aws

시작 안내서

Amazon Redshift

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

Amazon Redshift: 시작 안내서

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

Amazon의 상표 및 트레이드 드레스는 Amazon 외 제품 또는 서비스와 함께, Amazon 브랜드 이미지를 떨어뜨리거나 고객에게 혼동을 일으킬 수 있는 방식으로 사용할 수 없습니다. Amazon이 소유하지 않은 기타 모든 상표는 Amazon과 제휴 관계이거나 관련이 있거나 후원 관계와 관계없이 해당 소유자의 자산입니다.

Table of Contents

..... v
서버리스 데이터 웨어하우스 시작하기 1
AWS에 가입 1
Amazon Redshift Serverless를 사용하여 데이터 웨어하우스 생성 1
샘플 데이터 로드 4
샘플 쿼리 실행 6
Amazon S3에서 데이터 로드 8
프로비저닝된 데이터 웨어하우스 시작하기 16
AWS에 가입 18
방화벽 규칙 확인 19
1단계: 샘플 클러스터 생성 19
2단계: SQL 클라이언트에 대한 인바운드 규칙 구성 22
3단계: SQL 클라이언트에 대한 액세스 권한 부여 및 쿼리 실행 23
쿼리 편집기 v2에 대한 액세스 권한 부여 23
4단계: Amazon S3에서 Amazon Redshift로 데이터 로드 24
SQL 명령을 사용하여 Amazon S3에서 데이터 로드 25
쿼리 에디터 v2를 사용하여 Amazon S3에서 데이터 로드 26
클러스터에 TICKIT 데이터 생성 27
5단계: 쿼리 편집기를 사용하여 예제 쿼리 시도 28
6단계: 환경 재설정 29
데이터 웨어하우스에서 데이터베이스 정의 및 사용 31
Amazon Redshift에 연결 32
데이터베이스 생성 33
사용자 생성 33
스키마를 생성 34
테이블 생성 36
테이블에 데이터 행 삽입 36
테이블에서 데이터 선택 37
데이터 로드 37
시스템 테이블 및 뷰 쿼리 38
테이블 이름 목록 보기 38
사용자 보기 39
최근 쿼리 보기 40
실행 중인 쿼리의 세션 ID 확인 41

쿼리 취소 41
수퍼유저 대기열을 사용해 쿼리 취소 43
Amazon Redshift 데이터베이스에 없는 데이터 쿼리 45
데이터 레이크 쿼리 45
원격 데이터 소스 쿼리 46
다른 데이터베이스의 데이터 액세스 46
Redshift 데이터로 ML 모델 훈련 47
Amazon Redshift 개념 알아보기 48
추가 학습 리소스 51
문서 기록 53

Amazon Redshift는 2025년 11월 1일부터 새 Python UDF 생성을 더 이상 지원하지 않습니다. Python UDF를 사용하려면 이 날짜 이전에 UDF를 생성하세요. 기존 Python UDF는 정상적으로 계속 작동합니다. 자세한 내용은 블로그 게시물을 참조하세요.

Amazon Redshift Serverless 데이터 웨어하우스 시작하기

Amazon Redshift Serverless를 처음 사용하는 경우 Amazon Redshift Serverless 사용을 시작하는 데 도움이 되도록 다음 섹션을 읽는 것이 좋습니다. Amazon Redshift Serverless의 기본 흐름은 서버리스 리소스를 생성하고, Amazon Redshift Serverless에 연결하고, 샘플 데이터를 로드한 다음 데이터에 대해 쿼리를 실행하는 것입니다. 이 안내서에서는 Amazon Redshift Serverless 또는 Amazon S3 버킷에서 샘플 데이터를 로드하도록 선택할 수 있습니다. 샘플 데이터는 Amazon Redshift 설명서 전체에서 기능을 설명하는 데 사용됩니다. Amazon Redshift 프로비저닝 데이터 웨어하우스 사용을 시작하려면 Amazon Redshift 프로비저닝된 데이터 웨어하우스 시작하기 섹션을 참조하세요.

- the section called “AWS에 가입”
- the section called “Amazon Redshift Serverless를 사용하여 데이터 웨어하우스 생성”
- the section called “Amazon S3에서 데이터 로드”

AWS에 가입

아직 계정이 없는 경우 AWS 계정에 가입합니다. 계정이 이미 있는 경우 이 사전 조건 단계를 건너뛰고 기존 계정을 사용할 수 있습니다.

1. https://portal.aws.amazon.com/billing/signup을 엽니다.
2. 온라인 지시 사항을 따릅니다.

   AWS 계정에 가입하면 AWS 계정 루트 사용자가 생성됩니다. 루트 사용자에게는 계정의 모든 AWS 서비스 및 리소스에 액세스할 수 있는 권한이 있습니다. 보안 모범 사례는 관리 사용자에게 관리자 액세스 권한을 할당하고, 루트 사용자만 사용하여 루트 사용자 액세스 권한이 필요한 작업을 수행하는 것입니다.

Amazon Redshift Serverless를 사용하여 데이터 웨어하우스 생성

Amazon Redshift Serverless 콘솔에 처음 로그인하면 서버리스 리소스를 생성하고 구성하는 데 사용할 수 있는 시작하기 환경에 액세스하라는 메시지가 표시됩니다. 이 안내서에서는 Amazon Redshift Serverless의 기본 설정을 사용하여 서버리스 리소스를 생성해보겠습니다.

설정을 더 세밀하게 제어하려면 Customize settings(설정 사용자 지정)를 선택합니다.

> **Note**
>
> Redshift Serverless를 사용하려면 3개의 서로 다른 가용 영역에 3개의 서브넷이 있는 Amazon VPC가 필요합니다. 또한 Redshift Serverless에는 최소 3개의 사용 가능한 IP 주소가 필요합니다. 계속하기 전에 Redshift Serverless에 사용하는 Amazon VPC의 세 가용 영역에 3개의 서브넷이 있고 3개 이상의 사용 가능한 IP 주소가 있는지 확인합니다. Amazon VPC에서 서브넷을 생성하는 방법에 대한 자세한 내용은 Amazon Virtual Private Cloud 사용 설명서에서 서브넷 생성을 참조하세요. Amazon VPC의 IP 주소에 대한 자세한 내용은 VPC 및 서브넷의 IP 주소 지정을 참조하세요.

기본 설정으로 구성하려면:

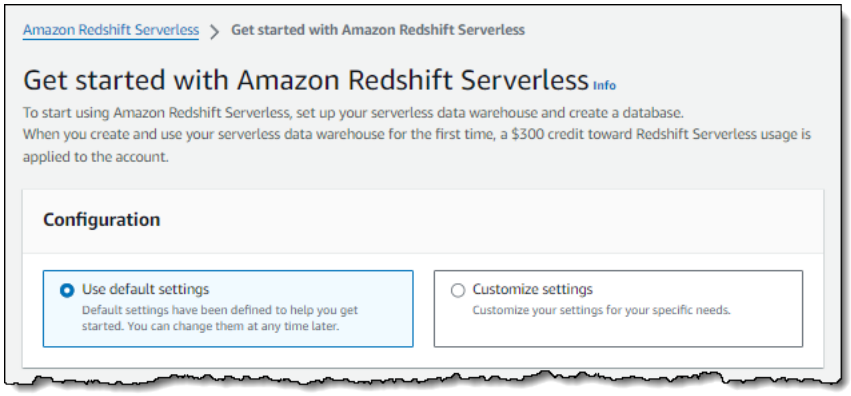
1. AWS Management Console에 로그인한 후 https://console.aws.amazon.com/redshiftv2/에서 Amazon Redshift 콘솔을 엽니다.

   Redshift Serverless 무료 평가판 사용해보기를 선택합니다.

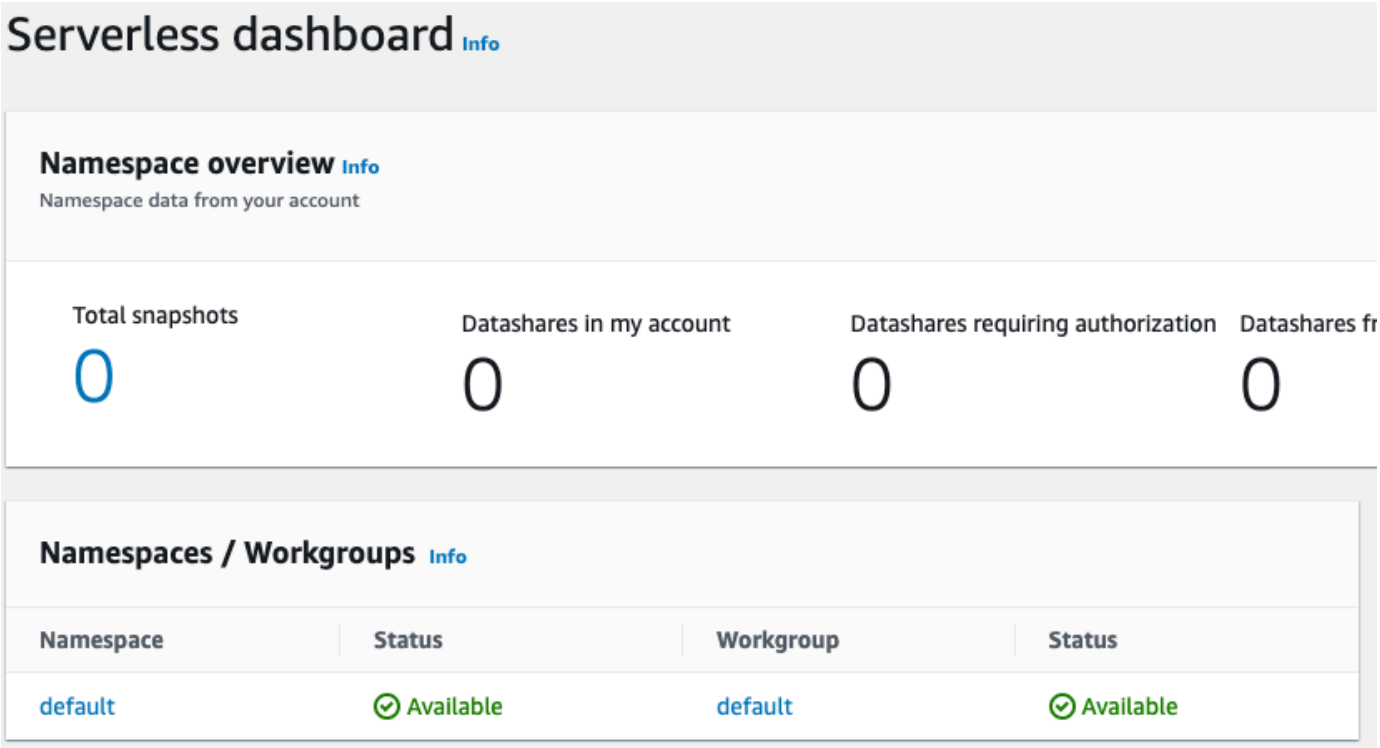
2. Configuration(구성) 아래에서 Use default settings(기본 설정 사용)를 선택합니다. Amazon Redshift Serverless가 이 네임스페이스에 연결된 기본 작업 그룹을 사용하여 기본 네임스페이스를 생성합니다. 구성 저장을 선택합니다.

   > **Note**
   >
   > 네임스페이스는 데이터베이스 객체와 사용자의 모음입니다. 네임스페이스는 스키마, 테이블, 사용자, 데이터 공유 및 스냅샷과 같이 Redshift Serverless에서 사용하는 모든 리소스를 함께 그룹화합니다.
   > 작업 그룹은 컴퓨팅 리소스의 모음입니다. 작업 그룹에는 Redshift Serverless가 컴퓨팅 작업을 실행하는 데 사용하는 컴퓨팅 리소스가 있습니다.

   다음 스크린샷은 Amazon Redshift Serverless의 기본 설정을 보여줍니다.



3. 설정이 완료된 후 계속(Continue)을 선택하여 서버리스 대시보드(Serverless dashboard)로 이동합니다. 서버리스 작업 그룹과 네임스페이스를 사용할 수 있는 것을 확인할 수 있습니다.



> **Note**
>
> Redshift Serverless가 작업 그룹을 성공적으로 생성하지 못한 경우 다음과 같이 해볼 수 있습니다.
>
> - Amazon VPC에 서브넷이 너무 적은 경우 등 Redshift Serverless에서 보고하는 오류를 모두 해결합니다.

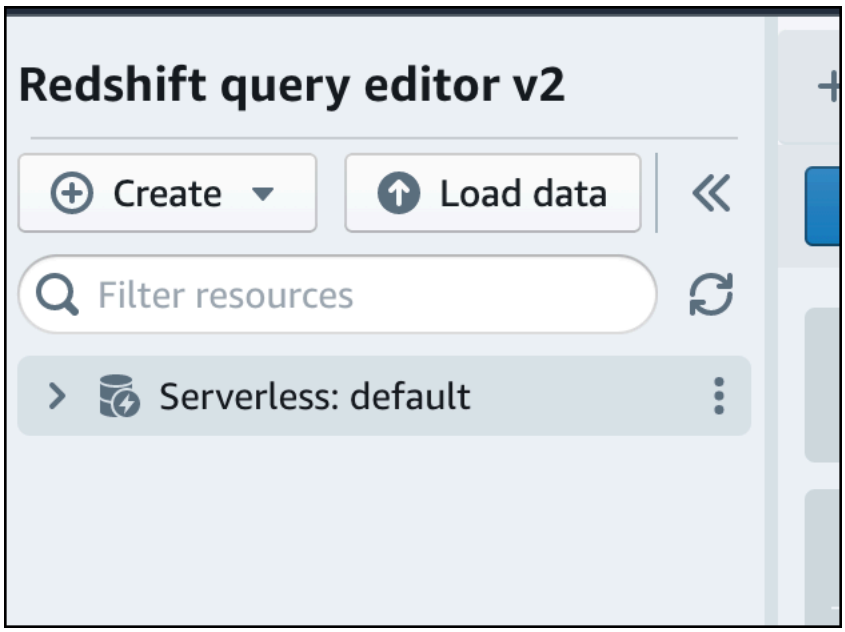
- Redshift Serverless 대시보드에서 기본 네임스페이스를 선택한 다음 작업, 네임스페이스 삭제를 선택하여 네임스페이스를 삭제합니다. 네임스페이스를 삭제하는 데 몇 분 정도 걸립니다.
- Redshift Serverless 콘솔을 다시 열면 시작 화면이 나타납니다.

샘플 데이터 로드

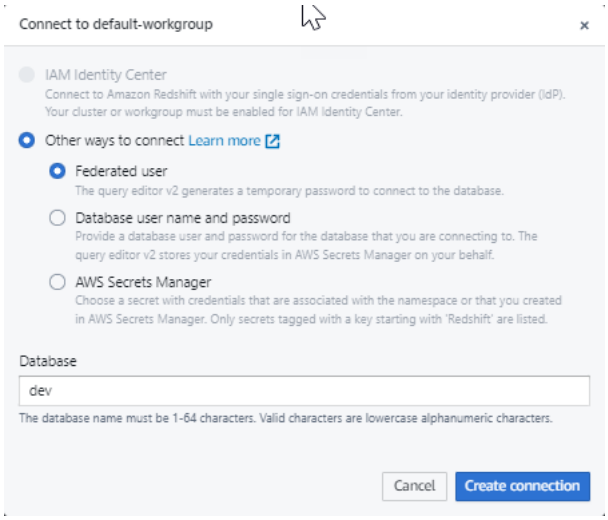
Amazon Redshift Serverless로 데이터 웨어하우스를 설정했으므로 이제 Amazon Redshift 쿼리 에디터 v2를 사용하여 샘플 데이터를 로드할 수 있습니다.

1. Amazon Redshift Serverless 콘솔에서 쿼리 에디터 v2를 실행하려면 데이터 쿼리를 선택합니다. Amazon Redshift Serverless 콘솔에서 쿼리 편집기 v2를 호출하면 쿼리 편집기와 함께 새 브라우저 탭이 열립니다. 쿼리 편집기 v2는 클라이언트 시스템에서 Amazon Redshift Serverless 환경으로 연결합니다.

2. 이 가이드에서는 AWS 관리자 계정과 기본 AWS KMS key를 사용합니다. 필요한 권한을 포함하여 쿼리 에디터 v2를 구성하는 방법에 대한 자세한 내용은 Amazon Redshift 관리 안내서의 AWS 계정 구성을 참조하세요. 고객 관리형 키를 사용하거나 Amazon Redshift에서 사용하는 KMS 키를 변경하도록 Amazon Redshift를 구성하는 방법에 대한 자세한 내용은 네임스페이스의 AWS KMS 키 변경을 참조하세요.
3. 작업 그룹에 연결하려면 트리 보기 패널에서 작업 그룹 이름을 선택합니다.

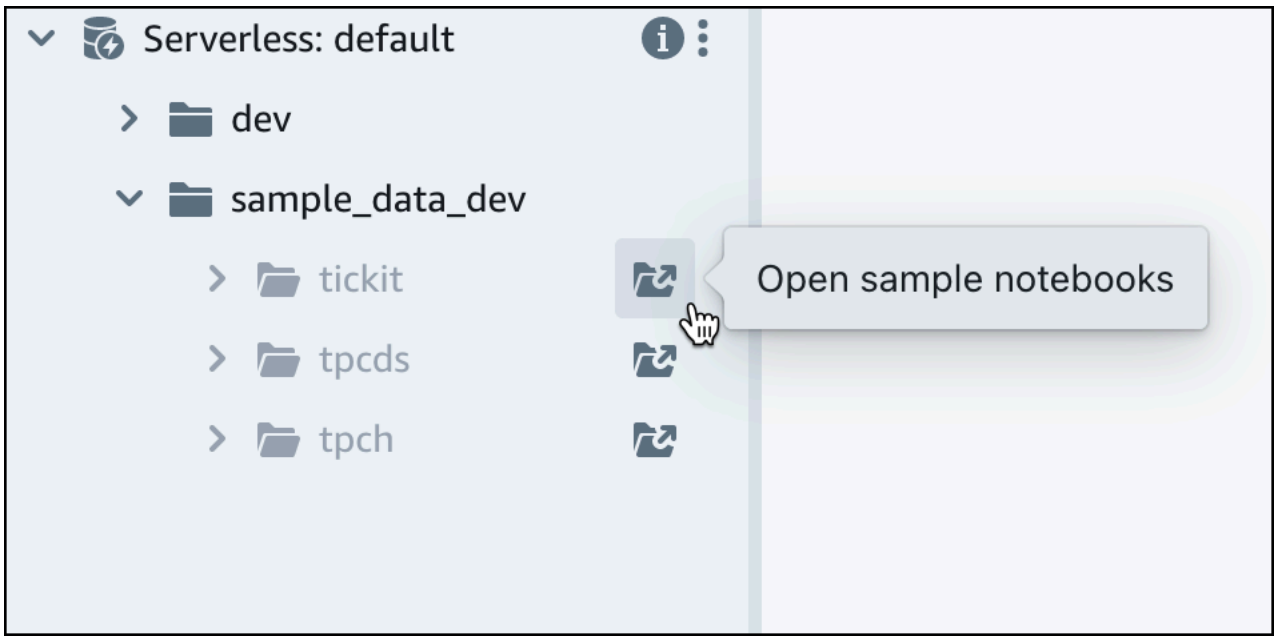


4. 쿼리 에디터 v2에서 새 작업 그룹에 처음 연결하는 경우 작업 그룹에 연결하는 데 사용할 인증 유형을 선택해야 합니다. 이 안내서에서는 페더레이션 사용자를 선택한 상태로 두고 연결 생성을 선택합니다.



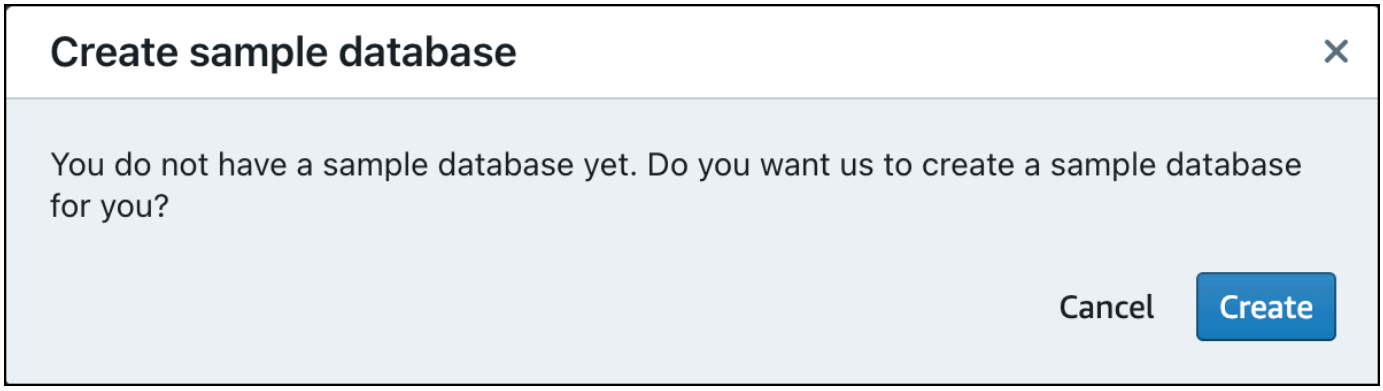
연결되면 Amazon Redshift Serverless 또는 Amazon S3 버킷에서 샘플 데이터를 로드하도록 선택할 수 있습니다.

5. Amazon Redshift Serverless 기본 작업 그룹 아래에서 sample_data_dev 데이터베이스를 펼칩니다. 세 개의 샘플 스키마가 있는데, 이들은 Amazon Redshift Serverless 데이터베이스로 로드할 수 있는 세 개의 샘플 데이터 세트에 해당합니다. 로드하려는 샘플 데이터 세트를 선택하고 샘플 노트북 열기를 선택합니다.



> **Note**
>
> SQL 노트북은 SQL 및 마크다운 셀의 컨테이너입니다. 노트북을 사용하여 하나의 문서에서 여러 SQL 명령을 구성하고 주석을 달고 공유할 수 있습니다.

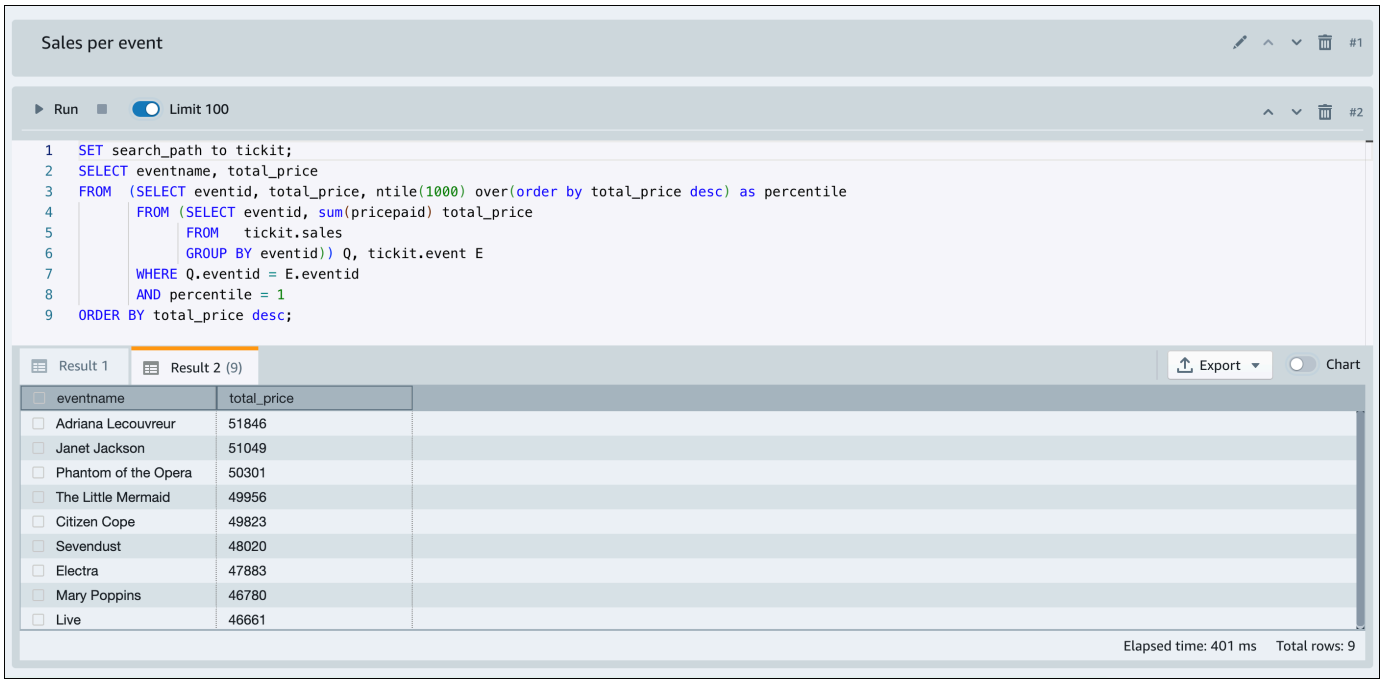
6. 처음으로 데이터를 로드하면 쿼리 에디터 v2에서 샘플 데이터베이스를 만들라는 메시지가 표시됩니다. 생성(Create)을 선택합니다.



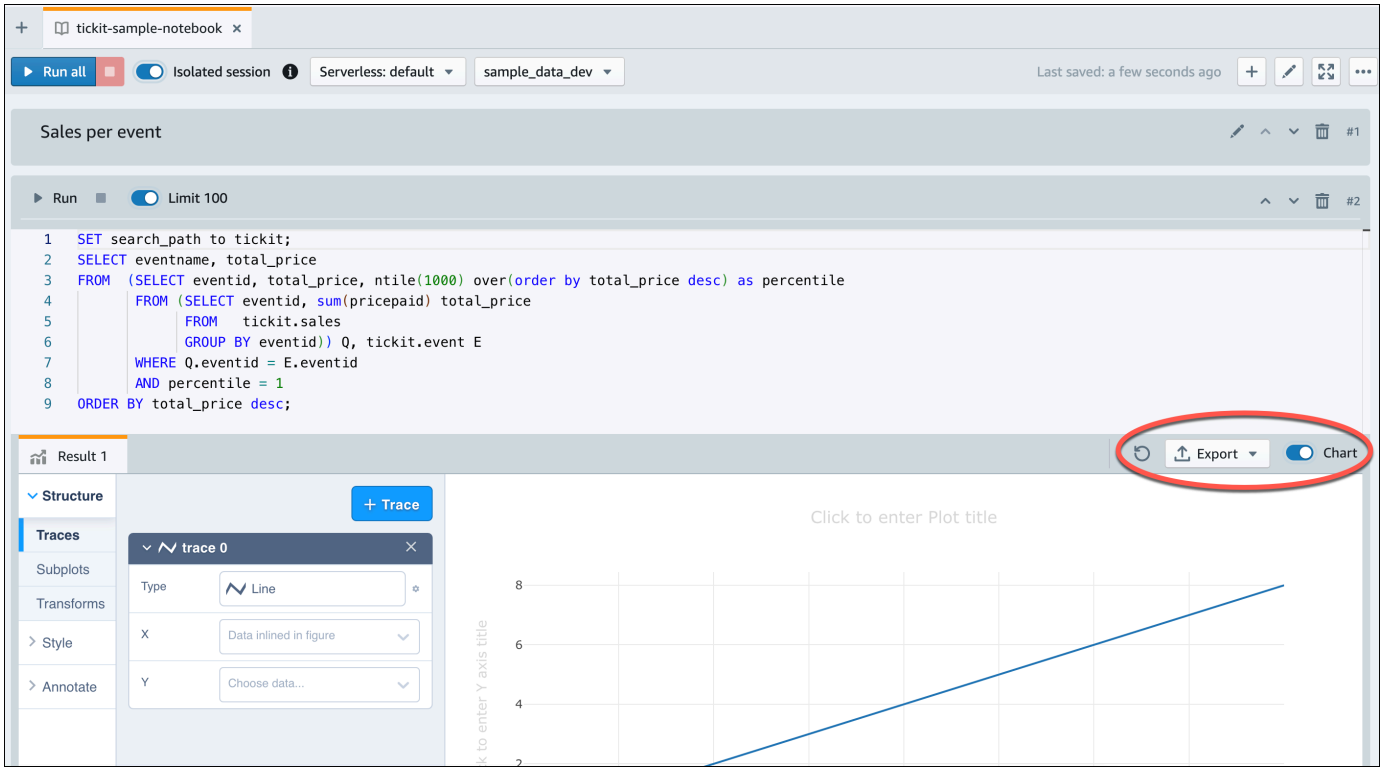
샘플 쿼리 실행

Amazon Redshift Serververless를 설정한 후 Amazon Redshift Serververless의 샘플 데이터 세트를 사용할 수 있습니다. 바로 데이터를 쿼리할 수 있도록 Amazon Redshift Serverless가 tickit 데이터 세트와 같은 샘플 데이터 세트를 자동으로 로드합니다.

- Amazon Redshift Serverless가 샘플 데이터 로드를 완료하면 모든 샘플 쿼리가 편집기에 로드됩니다. 모두 실행을 선택하여 샘플 노트북의 모든 쿼리를 실행할 수 있습니다.



결과를 JSON 또는 CSV 파일로 내보내거나 차트에서 결과를 볼 수도 있습니다.



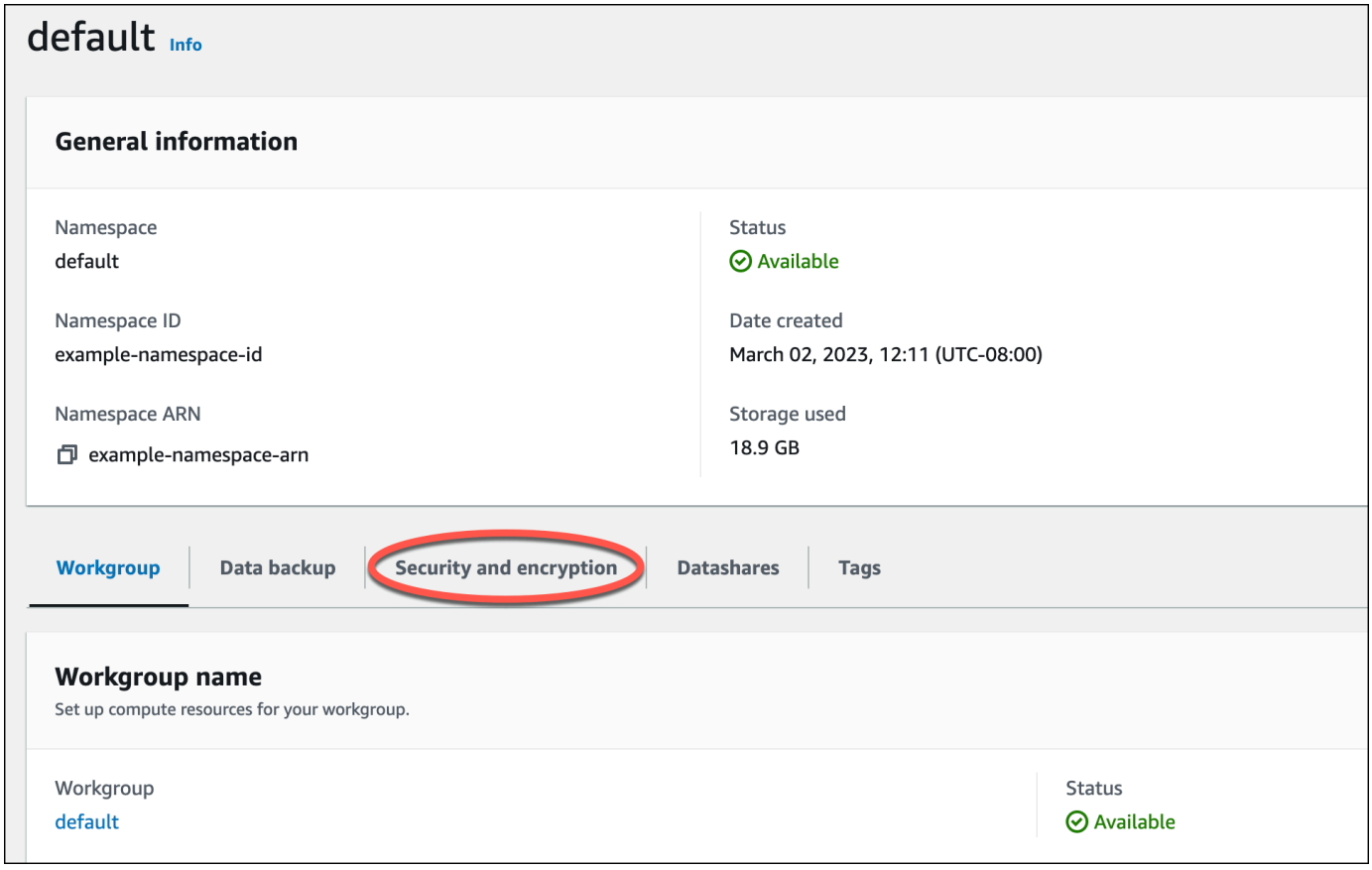
Amazon S3 버킷에서 데이터를 로드할 수도 있습니다. 자세한 내용은 the section called “Amazon S3에서 데이터 로드” 섹션을 참조하세요.

Amazon S3에서 데이터 로드

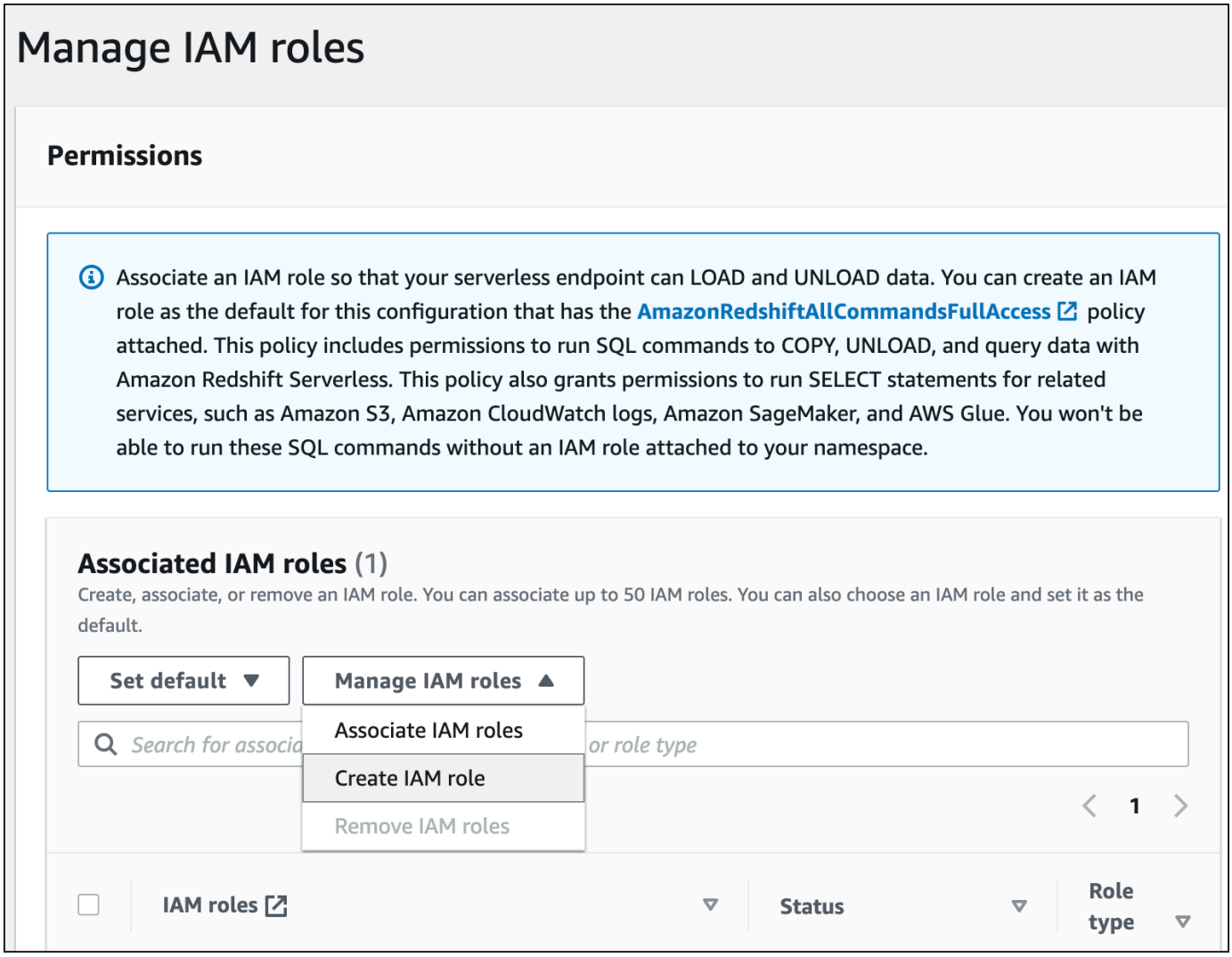
데이터 웨어하우스 생성후 Amazon S3에서 데이터를 로드할 수 있습니다.

이 시점에는 dev라는 이름의 데이터베이스가 있습니다. 다음으로 데이터베이스에 테이블 몇 개를 생성하고, 해당 테이블에 데이터를 업로드하고, 쿼리를 실행해 봅니다. 편의상 Amazon S3 버킷에 있는 샘플 데이터를 로드합니다.

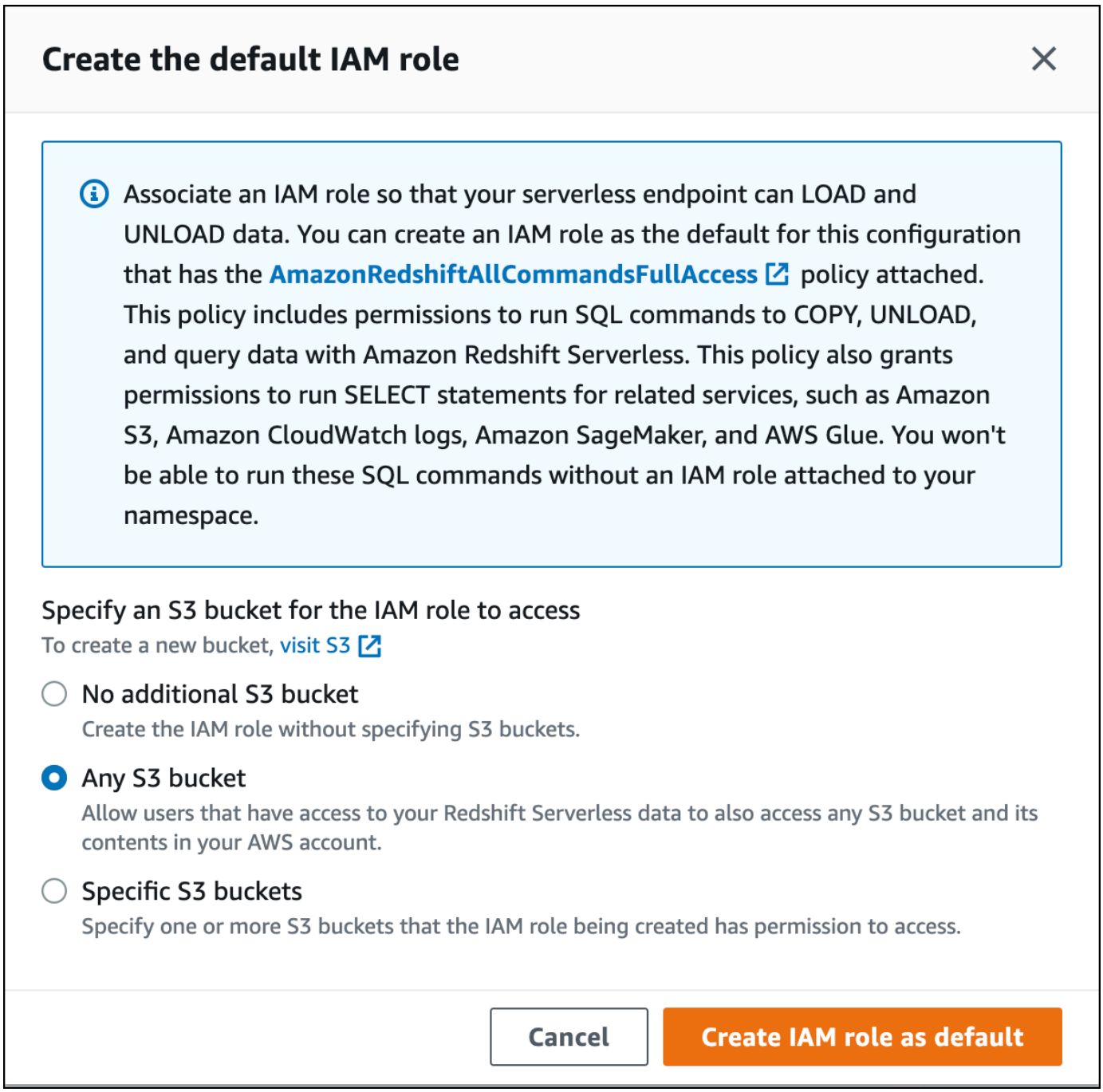
1. Amazon S3에서 데이터를 로드하려면 먼저 필요한 권한을 가진 IAM 역할을 생성하여 서버리스 네임스페이스에 연결해야 합니다. 이렇게 하려면 Redshift Serverless 콘솔로 돌아가 네임스페이스 구성을 선택합니다. 탐색 메뉴에서 네임스페이스를 선택하고 보안 및 암호화를 선택합니다. 그런 다음 IAM 역할 관리를 선택합니다.



2. IAM 역할 관리 메뉴를 확장하고 IAM 역할 생성을 선택합니다.



3. 이 역할에 부여하려는 S3 버킷 액세스 수준을 선택하고 IAM 역할을 기본값으로 생성을 선택합니다.



4. 변경 사항 저장을 선택합니다. 이제 Amazon S3에서 샘플 데이터를 로드할 수 있습니다.

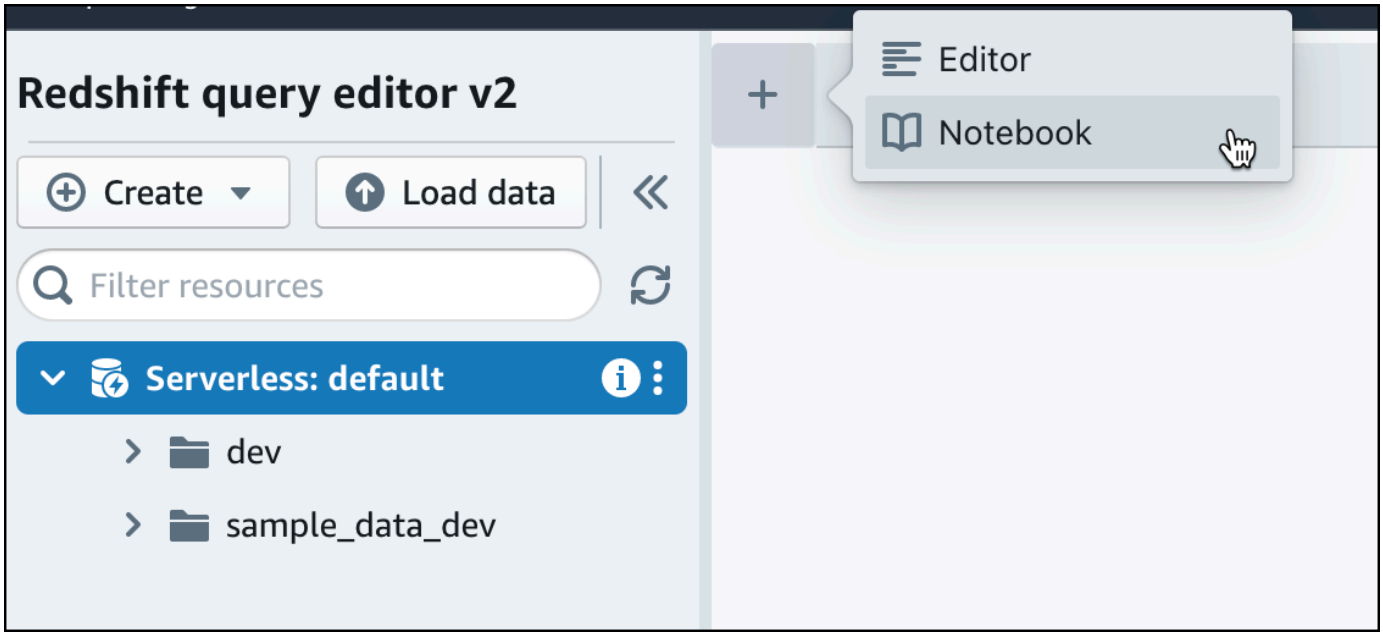
다음 단계는 퍼블릭 Amazon Redshift S3 버킷 내의 데이터를 사용하지만 자체 S3 버킷 및 SQL 명령을 사용하여 동일한 단계를 재현할 수 있습니다.

Amazon S3에서 샘플 데이터 로드

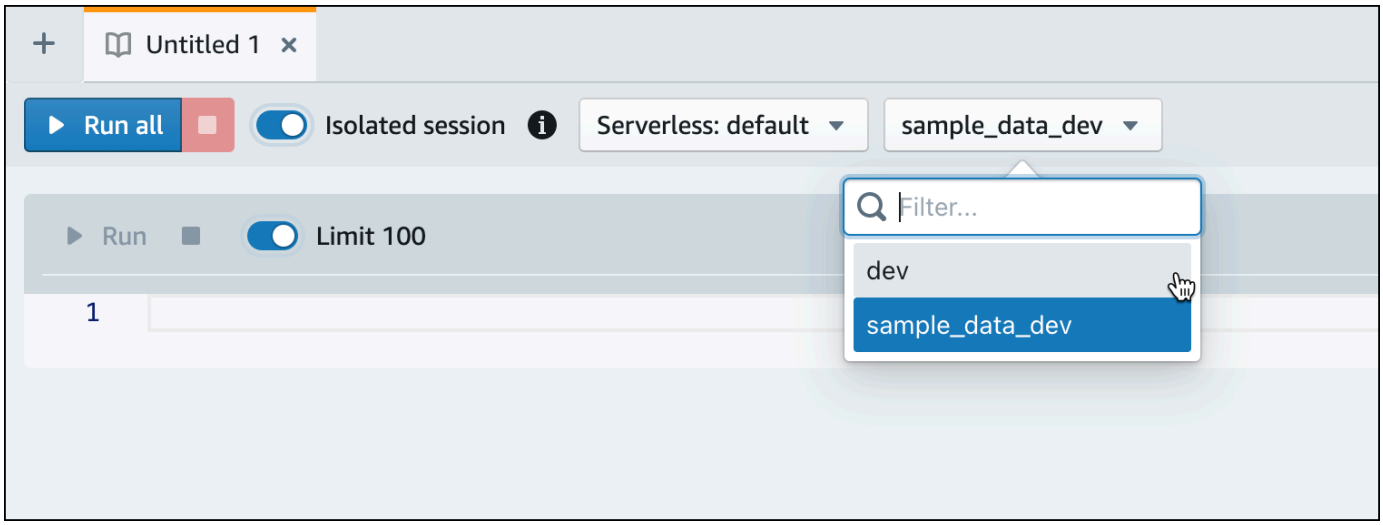
1. 쿼리 에디터 v2에서

   +

   추가를 선택한 다음 노트북을 선택하여 새 SQL 노트북을 생성합니다.



2. dev 데이터베이스로 전환합니다.



3. 테이블을 만듭니다.

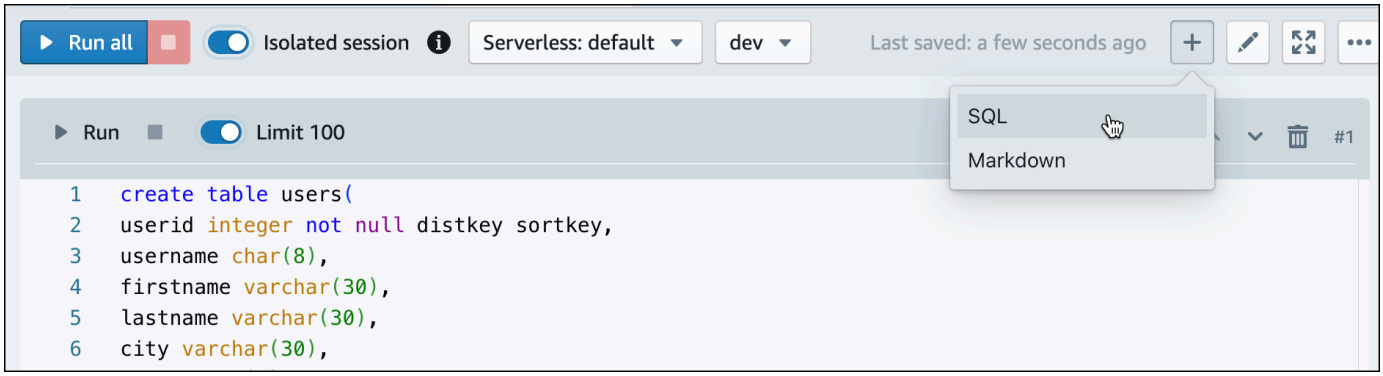
   쿼리 에디터 v2를 사용하는 경우 다음 create table 문을 복사하고 실행하여 dev 데이터베이스에 테이블을 생성합니다. 구문에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 CREATE TABLE 섹션을 참조하세요.

```
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. 쿼리 에디터 v2에서 노트북에 새 SQL 셀을 생성합니다.



5. 이제 쿼리 에디터 v2에서 COPY 명령을 사용하여 Amazon S3 또는 Amazon DynamoDB에서 Amazon Redshift로 대규모 데이터 세트를 로드합니다. COPY 구문에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 COPY 섹션을 참조하세요.

   퍼블릭 S3 버킷에서 사용 가능한 일부 샘플 데이터를 사용하여 COPY 명령을 실행할 수 있습니다. 쿼리 에디터 v2에서 다음 SQL 명령을 실행합니다.

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

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

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

6. 데이터를 로드한 후 노트북에 다른 SQL 셀을 만들고 몇 가지 예제 쿼리를 시도해 보세요. SELECT 명령 작업에 대한 자세한 내용은 Amazon Redshift 개발자 안내서의 SELECT 섹션을 참조하세요. 샘플 데이터의 구조와 스키마를 이해하려면 쿼리 에디터 v2를 사용해 살펴보세요.

```
-- 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;
```

이제 데이터를 로드하고 몇 가지 샘플 쿼리를 실행했으므로 Amazon Redshift Serverless의 다른 영역을 탐색할 수 있습니다. Amazon Redshift Serverless를 사용하는 방법에 대해 자세히 알아보려면 다음 목록을 참조하세요.

- Amazon S3 버킷에서 데이터를 로드할 수 있습니다. 자세한 내용은 Amazon S3에서 데이터 로드를 참조하세요.
- 쿼리 에디터 v2를 사용하여 크기가 5MB 미만인, 문자로 구분된 로컬 파일에서 데이터를 로드할 수 있습니다. 자세한 내용은 로컬 파일에서 데이터 로드를 참조하세요.
- JDBC 및 ODBC 드라이버가 있는 타사 SQL 도구를 사용하여 Amazon Redshift Serverless에 연결할 수 있습니다. 자세한 내용은 Amazon Redshift Serverless에 연결을 참조하세요.
- Amazon Redshift Data API를 사용하여 Amazon Redshift Serverless에 연결할 수도 있습니다. 자세한 내용은 Amazon Redshift 데이터 API 사용을 참조하세요.

- Amazon Redshift Serverless의 데이터를 Redshift ML과 함께 사용하여 CREATE MODEL 명령으로 기계 학습 모델을 생성할 수 있습니다. Redshift ML 모델을 구축하는 방법을 알아보려면 자습서: 고객 이탈 모델 구축을 참조하세요.
- Amazon Redshift Serverless에 데이터를 로드하지 않아도 Amazon S3 데이터 레이크의 데이터를 쿼리할 수 있습니다. 자세한 내용은 데이터 레이크 쿼리를 참조하세요.

Amazon Redshift 프로비저닝된 데이터 웨어하우스 시작하기

Amazon Redshift를 처음 사용하는 경우 프로비저닝된 클러스터 사용을 시작하는 데 도움이 되도록 다음 섹션을 읽는 것이 좋습니다. Amazon Redshift의 기본 흐름은 프로비저닝된 리소스를 생성하고, Amazon Redshift에 연결하고, 샘플 데이터를 로드한 다음, 데이터에 대한 쿼리를 실행하는 것입니다. 이 안내서에서는 Amazon Redshift 또는 Amazon S3 버킷에서 샘플 데이터를 로드하도록 선택할 수 있습니다. 샘플 데이터는 Amazon Redshift 설명서 전체에서 기능을 설명하는 데 사용됩니다.

이 자습서에서는 시스템 리소스를 관리하는 AWS 데이터 웨어하우스 객체인 Amazon Redshift 프로비저닝 클러스터를 사용하는 방법을 보여 줍니다. Amazon Redshift를 서버리스 작업 그룹과 함께 사용할 수도 있습니다. 이는 사용량에 따라 자동으로 규모가 조정되는 데이터 웨어하우스 객체입니다. Redshift 서버리스 사용을 시작하려면 Amazon Redshift Serverless 데이터 웨어하우스 시작하기 단원을 참조하세요.

Amazon Redshift 프로비저닝 콘솔을 생성하고 로그인한 후 클러스터, 노드, 데이터베이스를 포함한 Amazon Redshift 객체를 생성하고 관리할 수 있습니다. 또한 SQL 클라이언트를 사용하여 쿼리를 실행하고, 쿼리를 보고, 기타 SQL 데이터 정의 언어(DDL) 및 데이터 조작 언어(DML) 작업을 수행할 수도 있습니다.

> **Important**
>
> 이 연습에서 프로비저닝하는 클러스터는 실제 환경에서 실행됩니다. 실행되는 동안에는 AWS 계정에 요금이 발생합니다. 요금 정보는 Amazon Redshift 요금 페이지를 참조하세요.
> 불필요한 과금을 피하려면 클러스터를 사용한 후 삭제해야 합니다. 이 장의 마지막 섹션에서는 그 방법을 설명합니다.

AWS Management Console에 로그인한 후 https://console.aws.amazon.com/redshiftv2/에서 Amazon Redshift 콘솔을 엽니다.

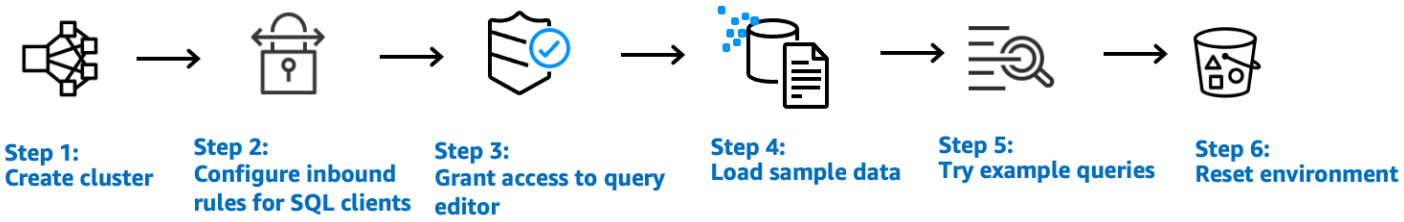
Amazon Redshift 콘솔 사용을 시작하려면 먼저 프로비저닝된 클러스터 대시보드로 이동하는 것이 좋습니다.

구성에 따라 Amazon Redshift 프로비저닝 콘솔의 탐색 창에 다음 항목이 나타납니다.

- Redshift 서버리스 – Amazon Redshift 프로비저닝된 클러스터를 설정, 조정 및 관리할 필요 없이 데이터에 액세스하고 분석합니다.

- 프로비저닝된 클러스터 대시보드 - AWS 리전의 클러스터 목록을 보고, 클러스터 지표 및 쿼리 개요에서 지표 데이터(예: CPU 사용률) 및 쿼리 정보에 대한 인사이트를 확인할 수 있습니다. 이를 통해 지정된 시간 범위 동안 성능 데이터가 비정상인지 판단할 수 있습니다.
- 클러스터 - 이 AWS 리전의 클러스터 목록을 확인하거나, 쿼리를 시작할 클러스터를 선택하거나, 클러스터 관련 작업을 수행합니다. 이 페이지에서 새 클러스터를 생성할 수도 있습니다.
- 쿼리 편집기 – Amazon Redshift 클러스터에 호스팅된 데이터베이스에 대한 쿼리를 실행합니다. 대신 쿼리 에디터 v2를 사용하는 것이 좋습니다.
- 쿼리 에디터 v2 – Amazon Redshift 쿼리 에디터 v2는 Amazon Redshift 데이터 웨어하우스에서 쿼리를 작성하고 실행하는 별도의 웹 기반 SQL 클라이언트 애플리케이션입니다. 결과를 차트로 시각화하고 팀의 다른 사용자와 쿼리를 공유하여 협업할 수 있습니다.
- 쿼리 및 로드(Queries and loads) - 최근 쿼리 목록과 각 쿼리의 SQL 텍스트와 같은 참조 또는 문제 해결을 위한 정보를 제공합니다.
- Datashare – 생산자 계정 관리자로서 소비자 계정에 datashare 액세스 권한을 부여하거나 어떠한 액세스 권한도 부여하지 않도록 선택할 수 있습니다. 권한이 부여된 데이터 공유를 사용하기 위해 소비자 계정 관리자는 데이터 공유를 전체 AWS 계정 또는 계정의 특정 클러스터 네임스페이스와 연결할 수 있습니다. 관리자가 datashare를 거부할 수도 있습니다.
- 제로 ETL 통합 - 지원되는 소스에 작성된 트랜잭션 데이터를 Amazon Redshift에서 사용할 수 있도록 통합을 관리합니다.
- IAM Identity Center 연결 - Amazon Redshift와 IAM Identity Center 간의 연결을 구성합니다.
- 구성(Configurations) – JDBC(Java Database Connectivity) 및 ODBC(Open Database Connectivity) 연결을 통해 SQL 클라이언트 도구에서 Amazon Redshift 클러스터로 연결합니다. Amazon Redshift 관리형 Virtual Private Cloud(VPC) 엔드포인트를 설정할 수도 있습니다. 이렇게 하면 클러스터가 포함된 Amazon VPC 서비스를 기반으로 하는 VPC와 클라이언트 도구를 실행하는 다른 VPC 간에 프라이빗 연결이 제공됩니다.
- AWS 파트너 통합 - 지원되는 AWS 파트너와의 통합을 생성합니다.
- Advisor - 최적화의 우선순위를 지정하기 위해 Amazon Redshift 클러스터에 적용할 수 있는 변경 사항에 대한 구체적인 권장 사항을 제공합니다.
- AWS Marketplace - Amazon Redshift와 함께 작동하는 다른 도구 또는 AWS 서비스에 대한 정보를 제공합니다.
- 경보(Alarms) - 클러스터 지표에 대한 경보를 생성하여 성능 데이터를 보고 지정한 기간 동안 지표를 추적합니다.
- 이벤트(Events) - 이벤트를 추적하고 이벤트가 발생한 날짜, 설명 또는 이벤트 소스와 같은 정보에 대한 보고서를 가져옵니다.
- 새로운 소식(What's new) – 새로운 Amazon Redshift 기능과 제품 업데이트를 봅니다.

이 자습서에서는 다음 단계를 수행합니다.



주제

- AWS에 가입
- 방화벽 규칙 확인
- 1단계: 샘플 Amazon Redshift 클러스터 생성
- 2단계: SQL 클라이언트에 대한 인바운드 규칙 구성
- 3단계: SQL 클라이언트에 대한 액세스 권한 부여 및 쿼리 실행
- 4단계: Amazon S3에서 Amazon Redshift로 데이터 로드
- 5단계: 쿼리 편집기를 사용하여 예제 쿼리 시도
- 6단계: 환경 재설정

AWS에 가입

아직 AWS 계정이 없다면 가입하세요. 계정이 이미 있는 경우 이 사전 조건 단계를 건너뛰고 기존 계정을 사용할 수 있습니다.

1. https://portal.aws.amazon.com/billing/signup을 엽니다.
2. 온라인 지시 사항을 따릅니다.

   등록 절차 중 전화 또는 텍스트 메시지를 받고 전화 키패드로 확인 코드를 입력하는 과정이 있습니다.

   AWS 계정 루트 사용자에 가입하면 AWS 계정 루트 사용자가 만들어집니다. 루트 사용자에게는 계정의 모든 AWS 서비스 및 리소스에 액세스할 권한이 있습니다. 보안 모범 사례는 사용자에게 관리 액세스 권한을 할당하고, 루트 사용자만 사용하여 루트 사용자 액세스 권한이 필요한 작업을 수행하는 것입니다.

방화벽 규칙 확인

> **Note**
>
> 이 자습서에서는 클러스터가 기본 포트 5439를 사용하고 Amazon Redshift 쿼리 에디터 v2를 사용하여 SQL 명령을 실행할 수 있다고 가정합니다. 환경에 필요할 수 있는 네트워킹 구성 또는 SQL 클라이언트 설정에 대해서는 자세히 다루지 않습니다.

일부 환경에서는 Amazon Redshift 클러스터를 시작할 때 포트를 지정합니다. 클러스터의 엔드포인트 URL과 함께 이 포트를 사용하여 클러스터에 액세스합니다. 또한 보안 그룹에 인바운드 수신 규칙을 생성하여 포트를 통해 클러스터에 대한 액세스를 허용합니다.

클라이언트 컴퓨터가 방화벽을 사용하는 경우 사용할 수 있는 개방된 포트를 알아야 합니다. 이 개방된 포트를 사용하여 SQL 클라이언트 도구에서 클러스터에 연결하고 쿼리를 실행할 수 있습니다. 개방된 포트를 모르는 경우 사용자의 네트워크 방화벽 규칙을 알고 있는 사람과 작업하여 방화벽의 개방된 포트를 확인합니다.

Amazon Redshift는 기본적으로 포트 5439를 사용하지만, 포트가 방화벽에서 막혀 있으면 연결이 실패합니다. Amazon Redshift 클러스터의 포트 번호는 생성한 후에는 변경할 수 없습니다. 따라서 시작 프로세스 중에 환경에서 작업하는 개방된 포트를 지정해야 합니다.

1단계: 샘플 Amazon Redshift 클러스터 생성

이 자습서에서는 데이터베이스가 포함된 Amazon Redshift 클러스터를 생성하는 프로세스를 안내합니다. 그런 다음, Amazon S3의 데이터세트를 데이터베이스의 테이블에 로드합니다. 이 샘플 클러스터를 사용하여 Amazon Redshift 서비스를 평가할 수 있습니다.

Amazon Redshift 클러스터 설정을 시작하기 전에 AWS에 가입 및 방화벽 규칙 확인과 같은 필수 사전 요구 사항을 완료했는지 확인합니다.

다른 AWS 리소스의 데이터에 액세스하는 작업의 경우 클러스터에 사용자를 대신해 리소스와 리소스의 데이터에 액세스할 권한이 필요합니다. 예에서는 SQL COPY 명령을 사용하여 Amazon Simple Storage Service(S3)에서 데이터를 로드합니다. AWS Identity and Access Management(IAM)을 사용하여 그러한 권한을 제공합니다. 생성하여 클러스터에 연결한 IAM 역할을 통해 이 작업을 수행할 수 있습니다. 자격 증명 및 액세스 권한에 관한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 자격 증명 및 액세스 권한을 참조하세요.

Amazon Redshift 클러스터를 생성하려면

1. AWS Management Console에 로그인한 후 https://console.aws.amazon.com/redshiftv2/에서 Amazon Redshift 콘솔을 엽니다.

   > **Important**
   >
   > IAM 사용자 자격 증명을 사용하는 경우 사용자에게 클러스터 작업을 수행하는 데 필요한 권한이 있는지 확인해야 합니다. 자세한 내용은 Amazon Redshift 관리 안내서의 Amazon Redshift의 보안을 참조하세요.

2. AWS 콘솔에서 클러스터를 생성하려는 AWS 리전을 선택합니다.
3. 탐색 메뉴에서 클러스터(Clusters)를 선택한 다음 클러스터 생성(Create cluster)을 선택합니다. 클러스터 생성 페이지가 표시됩니다.
4. [클러스터 구성(Cluster configuration)] 섹션에서 [클러스터 식별자(Cluster identifier)], [노드 유형(Node type)] 및 [노드(Nodes)] 값을 지정합니다.

   - [클러스터 식별자(Cluster identifier)]: 이 튜토리얼의 경우 **`examplecluster`**를 입력합니다. 이 식별자는 고유해야 합니다. 식별자는 유효한 문자로 a~z(소문자만) 및 -(하이픈)을 사용하여 1~63자여야 합니다.
   - 다음 방법 중 하나를 선택하여 클러스터 크기를 조정합니다.

     > **Note**
     >
     > 다음 단계에서는 RA3 노드 유형을 지원하는 AWS 리전을 가정합니다. RA3 노드 유형을 지원하는 AWS 리전 목록은 Amazon Redshift 관리 안내서의 RA3 노드 유형 개요를 참조하세요. 각 노드 유형 및 크기의 노드 사양에 대해 자세히 알아보려면 노드 유형 세부 정보를 참조하세요.

     - 클러스터 크기를 얼마로 조정해야 할지 모르는 경우 선택 도움말(Help me choose)을 선택합니다. 이렇게 하면 데이터 웨어하우스에 저장하려는 데이터의 크기 및 쿼리 특성에 대한 질문을 하는 크기 조정 계산기가 열립니다.

       클러스터의 필요한 크기(즉, 노드 유형 및 노드 수)를 알고 있는 경우 직접 선택(I'll choose)을 선택합니다. 그런 다음 노드 유형(Node type)과 노드(Nodes) 수를 선택하여 클러스터 크기를 조정합니다.

이 자습서에서는 노드 유형으로 ra3.4xlarge를 선택하고 노드 수로 2를 선택합니다.

AZ 구성을 선택할 수 있는 경우 단일 AZ를 선택합니다.

- Amazon Redshift가 제공하는 샘플 데이터 집합을 사용하려면 [샘플 데이터(Sample data)]에서 [샘플 데이터 로드(Load sample data)]를 선택합니다. Amazon Redshift는 샘플 데이터 집합 Tickit을 기본 `dev` 데이터베이스와 `public` 스키마로 로드합니다.

5. 데이터베이스 구성 섹션에서 관리자 사용자 이름 값을 지정합니다. 관리자 암호의 경우 다음 옵션 중에서 선택합니다.

   - 암호 생성 - Amazon Redshift에서 생성된 암호를 사용합니다.
   - 관리자 암호 수동 추가 - 내 암호를 사용합니다.
   - AWS Secrets Manager에서 관리자 보안 인증 정보 관리 - Amazon Redshift는 관리자 암호를 생성하고 관리하는 데 AWS Secrets Manager를 사용합니다. AWS Secrets Manager를 사용하여 암호의 보안 암호를 생성하고 관리하면 수수료가 발생합니다. AWS Secrets Manager 요금에 대한 자세한 내용은 AWS Secrets Manager 요금을 참조하세요.

   이 튜토리얼에서는 다음 값을 사용합니다.

   - [관리자 이름(Admin user name)]: **`awsuser`**를 입력합니다.
   - 관리 사용자 암호: 암호에 **`Changeit1`**을 입력합니다.

6. 이 자습서에서는 다음에 설명된 대로 IAM 역할을 생성하고 클러스터의 기본값으로 설정합니다. 클러스터에 대해 설정된 기본 IAM 역할은 하나만 있을 수 있습니다.

   a. 클러스터 권한(Cluster permissions)의 IAM 역할 관리(Manage IAM roles)에서 IAM 역할 생성(Create IAM role)을 선택합니다.

   b. 다음 방법 중 하나로 IAM 역할이 액세스할 Amazon S3 버킷을 지정합니다.

      - 생성된 IAM 역할이 `redshift`로 명명된 Amazon S3 버킷에만 액세스하도록 허용하려면 추가 Amazon S3 버킷 없음(No additional Amazon S3 bucket)을 선택합니다.
      - 생성된 IAM 역할이 모든 Amazon S3 버킷에 액세스하도록 허용하려면 모든 Amazon S3 버킷(Any Amazon S3 bucket)을 선택합니다.
      - 생성된 IAM 역할이 액세스할 Amazon S3 버킷을 하나 이상 지정하려면 특정 Amazon S3 버킷(Specific Amazon S3 buckets)을 선택합니다. 그런 다음 테이블에서 Amazon S3 버킷을 하나 이상 선택합니다.

c. 기본값으로 IAM 역할 생성(Create IAM role as default)을 선택합니다. Amazon Redshift는 IAM 역할을 자동으로 생성하고 클러스터의 기본값으로 설정합니다.

콘솔에서 IAM 역할을 생성했기 때문에 `AmazonRedshiftAllCommandsFullAccess` 정책이 연결되어 있습니다. 이를 통해 Amazon Redshift는 IAM 계정의 Amazon 리소스에서 데이터를 복사, 로드, 쿼리 및 분석할 수 있습니다.

클러스터의 기본 IAM 역할을 관리하는 방법에 대한 자세한 내용은 Amazon Redshift 관리 안내서의 Amazon Redshift의 기본값으로 IAM 역할 생성을 참조하세요.

7. (옵션) Additional configurations(추가 구성) 섹션에서 Use defaults(기본값 사용)를 해제하여 Network and security(네트워크 및 보안), Database configuration(데이터베이스 구성), Maintenance(유지 관리), Monitoring(모니터링) 및 Backup(백업) 설정을 수정합니다.

   경우에 따라 샘플 데이터 로드(Load sample data) 옵션을 사용하여 클러스터를 생성하고 Enhanced Amazon VPC Routing을 설정할 수 있습니다. 이러한 경우 Virtual Private Cloud(VPC)의 클러스터는 데이터를 로드하기 위해 Amazon S3 엔드포인트에 액세스해야 합니다.

   클러스터를 공개적으로 액세스할 수 있도록 하려면 다음 두 가지 중 하나를 수행합니다. 클러스터가 인터넷에 액세스할 수 있도록 VPC에서 네트워크 주소 변환(NAT) 주소를 구성할 수 있습니다. 또는 VPC에서 Amazon S3 VPC 엔드포인트를 구성할 수 있습니다. 향상된 Amazon VPC 라우팅에 관한 자세한 내용은 Amazon Redshift 관리 안내서의 향상된 Amazon VPC 라우팅 단원을 참조하세요.

8. 클러스터 생성을 선택합니다. 클러스터 페이지에서 클러스터가 `Available` 상태로 생성될 때까지 기다립니다.

2단계: SQL 클라이언트에 대한 인바운드 규칙 구성

> **Note**
>
> 이 단계를 건너뛰고 Amazon Redshift 쿼리 에디터 v2를 사용하여 클러스터에 액세스하는 것이 좋습니다.

이 튜토리얼의 뒷부분에서는 Amazon VPC 서비스를 기반으로 하는 Virtual Private Cloud(VPC) 내에서 클러스터에 액세스합니다. 그러나 방화벽 외부에서 SQL 클라이언트를 사용하여 클러스터에 액세스하는 경우 인바운드 액세스 권한을 부여해야 합니다.

방화벽을 확인하고 클러스터에 대한 인바운드 액세스 권한을 부여하려면

1. 방화벽 외부에서 클러스터에 액세스해야 하는 경우 방화벽 규칙을 확인합니다. 예를 들어 클라이언트는 Amazon Elastic Compute Cloud(Amazon EC2) 인스턴스이거나 외부 컴퓨터일 수 있습니다.

   방화벽 규칙에 대한 자세한 내용은 Amazon EC2 사용 설명서의 보안 그룹 규칙을 참조하세요.
2. Amazon EC2 외부 클라이언트에서 액세스하려면 클러스터에 연결된 보안 그룹에 인바운드 트래픽을 허용하는 수신 규칙을 추가합니다. Amazon EC2 콘솔에서 Amazon EC2 보안 그룹 규칙을 추가합니다. 예를 들어 CIDR/IP 192.0.2.0/24를 사용하면 해당 IP 주소 범위의 클라이언트가 사용자의 클러스터에 연결할 수 있습니다. 환경에 맞는 올바른 CIDR/IP를 찾습니다.

3단계: SQL 클라이언트에 대한 액세스 권한 부여 및 쿼리 실행

Amazon Redshift 클러스터에서 호스팅하는 데이터베이스를 쿼리하려는 경우 SQL 클라이언트에 대한 몇 가지 옵션이 있습니다. 다음이 포함됩니다.

- 클러스터에 연결하고 Amazon Redshift 쿼리 에디터 v2를 사용하여 쿼리를 실행합니다.

  쿼리 에디터 v2를 사용하는 경우 SQL 클라이언트 애플리케이션을 다운로드하고 설정할 필요가 없습니다. Amazon Redshift 콘솔에서 Amazon Redshift 쿼리 에디터 v2를 시작합니다.
- RSQL을 사용하여 클러스터에 연결합니다. 자세한 내용은 Amazon Redshift 관리 안내서의 Amazon Redshift RSQL을 사용하여 연결을 참조하세요.
- SQL Workbench/J와 같은 SQL 클라이언트 도구를 통해 클러스터에 연결합니다. 자세한 내용은 Amazon Redshift 관리 안내서의 SQL Workbench/J를 사용하여 클러스터에 연결 단원을 참조하세요.

이 자습서에서는 Amazon Redshift 클러스터에서 호스팅하는 데이터베이스에 대한 쿼리를 실행하는 쉬운 방법으로 Amazon Redshift 쿼리 에디터 v2를 사용합니다. 클러스터를 생성한 후 즉시 쿼리를 실행할 수 있습니다. Amazon Redshift 쿼리 에디터 v2 사용 시 고려 사항에 관한 자세한 내용은 Amazon Redshift 관리 안내서의 쿼리 에디터 v2로 작업할 때 고려할 사항 단원을 참조하세요.

쿼리 편집기 v2에 대한 액세스 권한 부여

관리자가 AWS 계정에 대해 쿼리 편집기 v2를 처음 구성할 때 쿼리 편집기 v2 리소스를 암호화하는 데 사용되는 AWS KMS key를 선택합니다. Amazon Redshift 쿼리 에디터 v2 리소스에는 저장된 쿼리, 노트북, 차트가 포함됩니다. 기본적으로 AWS 소유의 키가 리소스를 암호화하는 데 사용됩니다. 또는

관리자가 구성 페이지에서 키에 대한 Amazon 리소스 이름(ARN)을 선택하여 고객 관리형 키를 사용할 수 있습니다. 계정을 구성한 후에는 AWS KMS 암호화 설정을 변경할 수 없습니다. 자세한 내용은 Amazon Redshift 관리 안내서의 AWS 계정 구성 단원을 참조하세요.

쿼리 편집기 v2에 액세스하려면 권한이 있어야 합니다. 관리자는 Amazon Redshift 쿼리 에디터 v2에 대한 AWS 관리형 정책 중 하나를 IAM 역할 또는 사용자에 연결하여 권한을 부여할 수 있습니다. 이러한 AWS 관리형 정책은 리소스 태깅이 쿼리 공유를 허용하는 방법을 제어하는 다양한 옵션으로 작성됩니다. IAM 콘솔((https://console.aws.amazon.com/iam/))을 사용하여 IAM 사용자 정책을 연결할 수 있습니다. 이러한 정책에 대한 자세한 내용은 Amazon Redshift 관리 가이드의 쿼리 편집기 v2에 액세스를 참조하세요.

제공된 관리형 정책에서 허용 및 거부된 권한을 기반으로 고유한 정책을 생성할 수도 있습니다. IAM 콘솔 정책 편집기를 사용하여 고유한 정책을 생성하는 경우 시각적 편집기에서 정책을 생성할 서비스로 SQL Workbench를 선택합니다. 쿼리 편집기 v2는 시각적 편집기와 IAM 정책 시뮬레이터에서 서비스 이름으로 AWS SQL Workbench를 사용합니다.

자세한 정보는 Amazon Redshift 관리 안내서의 쿼리 편집기 v2 작업을 참조하세요.

4단계: Amazon S3에서 Amazon Redshift로 데이터 로드

클러스터를 생성한 후 Amazon S3의 데이터를 데이터베이스 테이블로 로드할 수 있습니다. Amazon S3에서 데이터를 로드하는 방법에는 여러 가지가 있습니다.

- SQL 클라이언트를 사용해 SQL CREATE TABLE 명령을 실행하여 데이터베이스에 테이블을 생성한 다음, SQL COPY 명령을 사용해 Amazon S3에서 데이터를 로드할 수 있습니다. Amazon Redshift 쿼리 에디터 v2는 SQL 클라이언트입니다.
- Amazon Redshift 쿼리 에디터 v2 로드 마법사를 사용할 수 있습니다.

이 자습서에서는 Amazon Redshift Query Editor V2를 사용해 SQL 명령으로 CREATE 테이블 및 COPY 데이터를 실행하는 방법을 보여 줍니다. Amazon Redshift 콘솔 탐색 창에서 쿼리 에디터 v2를 시작합니다. 쿼리 에디터 v2 내에서 관리자 `awsuser`로 `dev`라는 데이터베이스 및 `examplecluster` 클러스터에 대한 연결을 생성합니다. 이 자습서에서는 연결을 생성할 때 데이터베이스 사용자 이름을 사용하는 임시 자격 증명을 선택합니다. Amazon Redshift 쿼리 에디터 v2 사용에 대한 자세한 내용은 Amazon Redshift 관리 안내서의 Amazon Redshift 데이터베이스에 연결을 참조하세요.

SQL 명령을 사용하여 Amazon S3에서 데이터 로드

쿼리 에디터 v2 쿼리 편집기 창에서 `examplecluster` 클러스터 및 `dev` 데이터베이스에 연결되어 있는지 확인합니다. 다음으로, 데이터베이스에 테이블을 생성하고 테이블에 데이터를 로드합니다. 이 자습서에서는 로드하는 데이터를 여러 AWS 리전에서 액세스할 수 있는 Amazon S3 버킷에서 사용할 수 있습니다.

다음 절차에서는 테이블을 생성하고 퍼블릭 Amazon S3 버킷에서 데이터를 로드합니다.

Amazon Redshift 쿼리 에디터 v2를 사용해 다음 create table 문을 복사하고 실행하여 `dev` 데이터베이스의 `public` 스키마에 테이블을 생성합니다. 구문에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 CREATE TABLE 섹션을 참조하세요.

쿼리 에디터 v2와 같은 SQL 클라이언트를 사용하여 데이터를 생성하고 로드하는 방법

1. 다음 SQL 명령을 실행하여 `sales` 테이블을 생성합니다.

```
  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. 다음 SQL 명령을 실행하여 `date` 테이블을 생성합니다.

```
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. COPY 명령을 사용하여 Amazon S3에서 sales 테이블을 로드합니다.

> **Note**
>
> Amazon S3에서 Amazon Redshift로 대용량 데이터세트를 로드하려면 COPY 명령을 사용하는 것이 좋습니다. COPY 구문에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 COPY 섹션을 참조하세요.

클러스터가 사용자를 대신하여 Amazon S3에 액세스하여 샘플 데이터를 로드할 수 있도록 인증을 제공합니다. 클러스터를 생성할 때 IAM 역할을 기본값으로 생성을 선택한 경우 생성하고 클러스터의 default로 설정한 IAM 역할을 참조함으로써 인증을 제공합니다.

다음 SQL 명령을 사용하여 sales 테이블을 로드합니다. 필요에 따라 Amazon S3에서 sales 테이블의 소스 데이터를 다운로드하고 볼 수 있습니다. .

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

4. 다음 SQL 명령을 사용하여 date 테이블을 로드합니다. 필요에 따라 Amazon S3에서 date 테이블의 소스 데이터를 다운로드하고 볼 수 있습니다. .

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

쿼리 에디터 v2를 사용하여 Amazon S3에서 데이터 로드

이 섹션에서는 자체 데이터를 Amazon Redshift 클러스터에 로드에 관해 설명합니다. 쿼리 에디터 v2는 데이터 로드 마법사를 사용할 때 데이터 로드를 간소화합니다. 쿼리 에디터 v2 데이터 로드 마법사에서

생성되고 사용되는 COPY 명령은 Amazon S3에서 데이터를 로드하기 위해 COPY 명령 구문에 사용할 수 있는 많은 파라미터를 지원합니다. COPY 명령 및 Amazon S3에서 로드를 복사하는 데 사용되는 옵션에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 Amazon S3에서 COPY를 참조하세요.

Amazon S3에서 Amazon Redshift로 데이터를 로드하려면 Amazon Redshift에 지정된 Amazon S3 버킷에서 데이터를 로드하는 데 필요한 권한이 있는 IAM 역할이 필요합니다.

Amazon S3에서 Amazon Redshift로 자체 데이터를 로드하려면 Query Editor V2 로드 데이터 마법사를 사용합니다. 로드 데이터 마법사 사용 방법에 관한 자세한 내용은 Amazon Redshift 관리 가이드의 Amazon S3에서 데이터 로드 섹션을 참조하시기 바랍니다.

클러스터에 TICKIT 데이터 생성

TICKIT은 Amazon Redshift에서 데이터를 쿼리하는 방법을 학습하기 위해 필요에 따라 Amazon Redshift 클러스터에 로드할 수 있는 샘플 데이터베이스입니다. 다음과 같은 방법으로 전체 TICKIT 테이블 집합을 생성하고 클러스터에 데이터를 로드할 수 있습니다.

- Amazon Redshift 콘솔에서 클러스터를 생성할 때 샘플 TICKIT 데이터를 동시에 로드할 수 있습니다. Amazon Redshift 콘솔에서 클러스터, 클러스터 생성을 선택합니다. 샘플 데이터 섹션에서 샘플 데이터 로드를 선택합니다. Amazon Redshift는 클러스터 생성 중에 자동으로 Amazon Redshift 클러스터 `dev` 데이터베이스에 샘플 데이터세트를 로드합니다.
- 기존 클러스터에 연결하려면 다음과 같이 수행합니다.
  - Amazon Redshift 콘솔의 탐색 창에서 클러스터를 선택합니다.
  - 클러스터 창에서 클러스터를 선택합니다.
  - Query Editor V2에서 데이터 쿼리를 선택합니다.
  - 리소스 목록에서 examplecluster를 펼칩니다. 클러스터에 처음 연결하는 경우 examplecluster에 연결이 나타납니다. 데이터베이스 사용자 이름 및 암호를 선택합니다. 데이터베이스를 **`dev`**로 둡니다. 사용자 이름에서 **`awsuser`**, 암호에서 **`Changeit1`**를 지정합니다.
  - 연결 생성을 선택합니다.
- Amazon Redshift 쿼리 에디터 v2를 사용하면 TICKIT 데이터를 sample_data_dev라는 샘플 데이터베이스에 로드할 수 있습니다. 리소스 목록에서 sample_data_dev 데이터베이스를 선택합니다. tickit 노드 옆의 샘플 노트북 열기 아이콘을 선택합니다. 샘플 데이터베이스를 생성할지 확인합니다.
- Amazon Redshift 쿼리 에디터 v2는 tickit-sample-notebook이라는 샘플 노트북과 함께 샘플 데이터베이스를 생성합니다. 모두 실행을 선택하여 이 노트북을 실행하여 샘플 데이터베이스의 데이터를 쿼리합니다.

TICKIT 데이터에 관한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 샘플 데이터베이스를 참조하세요.

5단계: 쿼리 편집기를 사용하여 예제 쿼리 시도

Amazon Redshift 쿼리 에디터 v2를 설정하고 사용하여 데이터베이스를 쿼리하려면 Amazon Redshift 관리 안내서의 쿼리 에디터 v2 작업 단원을 참조하세요.

이제 다음과 같이 몇 가지 예제 쿼리를 시도합니다. Query Editor V2에서 새 쿼리를 생성하려면 쿼리 창의 오른쪽 상단에 있는 + 아이콘을 선택하고 SQL을 선택합니다. 다음 SQL 쿼리를 복사하여 붙여넣을 수 있는 새 쿼리 페이지가 나타납니다.

> **Note**
>
> 먼저 노트북에서 첫 번째 쿼리를 실행해야 합니다. 이는 다음 SQL 명령을 사용하여 `search_path` 서버 구성 값을 `tickit` 스키마로 설정합니다.
>
> ```
> set search_path to tickit;
> ```

SELECT 명령 작업에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 SELECT 단원을 참조하세요.

```
-- 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;
```

6단계: 환경 재설정

이전 단계에서는 Amazon Redshift 클러스터를 생성하고, 테이블에 데이터를 로드하며, Amazon Redshift 쿼리 에디터 v2와 같은 SQL 클라이언트를 사용하여 데이터를 쿼리했습니다.

이 튜토리얼을 마쳤으면 샘플 클러스터를 삭제하여 환경을 이전 상태로 재설정하는 것이 좋습니다. 클러스터를 삭제할 때까지 Amazon Redshift 서비스에 대해 사용 요금이 계속 부과됩니다.

그러나 다른 Amazon Redshift 안내서의 작업이나 데이터 웨어하우스에서 데이터베이스를 정의하고 사용하는 명령 실행에 설명된 작업을 시도해 보려는 경우 샘플 클러스터를 계속 실행하고자 할 수 있습니다.

클러스터 삭제

1. AWS Management Console에 로그인한 후 https://console.aws.amazon.com/redshiftv2/에서 Amazon Redshift 콘솔을 엽니다.
2. 탐색 메뉴에서 클러스터(Clusters)를 선택하여 클러스터 목록을 표시합니다.
3. `examplecluster` 클러스터를 선택합니다. 작업에 대해 삭제를 선택합니다. examplecluster 삭제 페이지가 나타납니다.
4. 삭제할 클러스터를 확인하고 최종 스냅샷 생성 설정을 선택 취소한 다음 **delete**를 입력하여 삭제를 확인합니다. 클러스터 삭제를 선택합니다.

클러스터가 삭제되면 클러스터 목록 페이지에서 클러스터 상태가 업데이트됩니다.

이 튜토리얼을 마친 후에는 Amazon Redshift에 대해 알아볼 수 있는 추가 리소스에서 Amazon Redshift에 대한 보다 자세한 내용과 다음 단계를 찾아볼 수 있습니다.

데이터 웨어하우스에서 데이터베이스를 정의하고 사용하는 명령 실행

Redshift Serverless 데이터 웨어하우스와 Amazon Redshift 프로비저닝된 데이터 웨어하우스는 모두 데이터베이스를 포함합니다. 데이터 웨어하우스를 시작한 후에는 SQL 명령을 사용하여 대부분의 데이터베이스 작업을 관리할 수 있습니다. 몇 가지 예외는 있지만, SQL의 기능과 구문은 모든 Amazon Redshift 데이터베이스에서 동일합니다. Amazon Redshift에서 사용할 수 있는 SQL 명령에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 SQL 명령을 참조하세요.

데이터 웨어하우스를 생성하면 대부분의 시나리오에서 Amazon Redshift가 기본 dev 데이터베이스도 생성합니다. dev 데이터베이스에 연결한 후 다른 데이터베이스를 생성할 수 있습니다.

다음 섹션에서는 Amazon Redshift 데이터베이스 작업 시 공통 데이터베이스 태스크를 안내합니다. 태스크는 데이터베이스 생성으로 시작되며, 마지막 태스크를 계속 진행하는 경우 데이터베이스를 삭제함으로써 생성한 모든 리소스를 삭제할 수 있습니다.

이 섹션의 예에서는 다음과 같이 가정합니다.

- Amazon Redshift 데이터 웨어하우스를 생성했습니다.
- Amazon Redshift 쿼리 에디터 v2와 같은 SQL 클라이언트 도구에서 데이터 웨어하우스에 대한 연결을 설정했습니다. 쿼리 에디터 v2에 관한 자세한 내용은 Amazon Redshift 관리 안내서의 Amazon Redshift 쿼리 에디터 v2를 사용하여 데이터베이스 쿼리를 참조하세요.

주제

- Amazon Redshift 데이터 웨어하우스에 연결
- 데이터베이스 생성
- 사용자 생성
- 스키마를 생성
- 테이블 생성
- 데이터 로드
- 시스템 테이블 및 뷰 쿼리
- 쿼리 취소

Amazon Redshift 데이터 웨어하우스에 연결

Amazon Redshift 클러스터에 연결하려면 Amazon Redshift 콘솔 클러스터 페이지에서 Amazon Redshift 클러스터에 연결을 확장하고 다음 중 하나를 수행합니다.

- 데이터 쿼리를 선택하여 쿼리 에디터 v2를 사용해 Amazon Redshift 클러스터에서 호스팅하는 데이터베이스에 대한 쿼리를 실행합니다. 클러스터를 생성한 후 바로 쿼리 편집기 v2를 사용해 쿼리를 실행할 수 있습니다.

  자세한 내용은 Amazon Redshift 관리 안내서의 Amazon Redshift 쿼리 에디터 v2를 사용하여 데이터베이스 쿼리를 참조하세요.
- 클라이언트 도구로 작업에서 클러스터를 선택하고 JDBC 또는 ODBC 드라이버 URL을 복사하여 JDBC 또는 ODBC 드라이버를 사용하는 클라이언트 도구에서 Amazon Redshift에 연결합니다. 클라이언트 컴퓨터 또는 인스턴스에서 이 URL을 사용합니다. JDBC 또는 ODBC 데이터 액세스 API 작업을 사용할 수 있는 애플리케이션을 코딩하거나 JDBC 또는 ODBC를 지원하는 SQL 클라이언트 도구를 사용하세요.

  클러스터 연결 문자열을 찾는 방법에 대한 자세한 내용은 클러스터 연결 문자열 찾기를 참조하세요.
- SQL 클라이언트 도구에 드라이버가 필요한 경우 JDBC 또는 ODBC 드라이버 선택 옵션을 사용해 운영 체제별 드라이버를 다운로드하여 클라이언트 도구에서 Amazon Redshift에 연결할 수 있습니다.

  SQL 클라이언트에 적합한 드라이버를 설치하는 방법에 대한 자세한 정보는 JDBC 드라이버 버전 2.0 연결 구성을 참조하세요.

  ODBC 연결을 구성하는 방법에 대한 자세한 내용은 ODBC 연결 구성을 참조하세요.

Redshift 서버리스 데이터 웨어하우스에 연결하려면 Amazon Redshift 콘솔 서버리스 대시보드 페이지에서 다음 중 하나를 수행합니다.

- Amazon Redshift 쿼리 에디터 v2를 사용하여 Redshift 서버리스 데이터 웨어하우스에서 호스팅하는 데이터베이스에 대한 쿼리를 실행합니다. 데이터 웨어하우스를 생성한 후 바로 쿼리 에디터 v2를 사용해 쿼리를 실행할 수 있습니다.

  자세한 내용은 Amazon Redshift 쿼리 편집기 v2를 사용하여 데이터베이스 쿼리을 참조하세요.
- JDBC 또는 ODBC 드라이버 URL을 복사하여 JDBC 또는 ODBC 드라이버로 클라이언트 도구에서 Amazon Redshift에 연결합니다.

데이터 웨어하우스의 데이터를 사용하려면 클라이언트 컴퓨터 또는 인스턴스에서 연결하기 위한 JDBC 또는 ODBC 드라이버가 필요합니다. JDBC 또는 ODBC 데이터 액세스 API 작업을 사용할 수 있는 애플리케이션을 코딩하거나 JDBC 또는 ODBC를 지원하는 SQL 클라이언트 도구를 사용하세요.

연결 문자열을 찾는 방법에 대한 자세한 내용은 Amazon Redshift 관리 안내서의 Redshift 서버리스에 연결을 참조하세요.

데이터베이스 생성

데이터 웨어하우스가 실행 중인 것을 확인한 후 데이터베이스를 생성할 수 있습니다. 이 데이터베이스에서 실제로 테이블을 생성하여 데이터를 업로드하고 쿼리를 실행하게 됩니다. 데이터 웨어하우스는 여러 데이터베이스를 호스팅할 수 있습니다. 예를 들어 동일한 데이터 웨어하우스에 SALESDB라는 판매 데이터용 데이터베이스와 ORDERSDB라는 주문 데이터용 데이터베이스가 있을 수 있습니다.

SALESDB라는 데이터베이스를 생성하려면 SQL 클라이언트 도구에서 다음 명령을 실행합니다.

```
CREATE DATABASE salesdb;
```

> **Note**
>
> 명령을 실행한 후 새 salesdb를 보려면 데이터 웨어하우스에 있는 객체의 SQL 클라이언트 도구 목록을 새로 고쳐야 합니다.

이 연습에서는 기본값을 사용하겠습니다. 더 많은 명령 옵션에 대한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 CREATE DATABASE 섹션을 참조하세요. 데이터베이스와 해당 내용을 삭제하려면 Amazon Redshift 데이터베이스 개발자 안내서의 DROP DATABASE를 참조하세요.

SALESDB 데이터베이스를 생성한 다음 SQL 클라이언트에서 새 데이터베이스에 연결할 수 있습니다. 현재 연결에 사용한 것과 같은 연결 파라미터를 사용하되 데이터베이스 이름을 SALESDB으로 변경합니다.

사용자 생성

기본적으로 데이터 웨어하우스를 시작할 때 생성한 관리자만 데이터 웨어하우스의 기본 데이터베이스에 액세스할 수 있습니다. 다른 사용자에게 액세스 권한을 부여하려면 하나 이상의 계정을 생성합니다.

데이터베이스 사용자 계정은 개별 데이터베이스가 아니라 데이터 웨어하우스의 모든 데이터베이스에서 전역적입니다.

CREATE USER 명령을 사용하여 새 사용자를 생성합니다. 새 사용자를 만들 때 새 사용자의 이름과 암호를 지정합니다. 사용자 암호를 지정하는 것이 좋습니다. 길이가 8~64자여야 하며, 대문자, 소문자, 숫자가 하나 이상 포함되어야 합니다.

예를 들어 이름이 **GUEST**이고 암호가 **ABCd4321**인 사용자를 생성하려면 다음 명령을 실행합니다.

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

GUEST 사용자로 SALESDB 데이터베이스에 연결하려면 사용자를 생성했을 때와 동일한 암호(예:ABCd4321)를 사용합니다.

다른 명령 옵션에 대한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 CREATE USER 섹션을 참조하세요.

스키마를 생성

새 데이터베이스를 생성한 후 현재 데이터베이스에서 새 스키마를 생성할 수 있습니다. 스키마는 테이블, 뷰 및 사용자 정의 함수(UDF)와 같은 명명된 데이터베이스 객체를 포함하는 네임스페이스입니다. 데이터베이스는 하나 또는 여러 개의 스키마를 포함할 수 있으며 각 스키마는 하나의 데이터베이스에만 속합니다. 두 스키마에는 동일한 이름을 공유하는 서로 다른 객체가 있을 수 있습니다.

동일한 데이터베이스에 여러 스키마를 생성하여 원하는 방식으로 데이터를 구성하거나 데이터를 기능적으로 그룹화할 수 있습니다. 예를 들어 모든 스테이징 데이터를 저장하는 스키마와 모든 보고 테이블을 저장하는 다른 스키마를 생성할 수 있습니다. 동일한 데이터베이스에 있는 다른 비즈니스 그룹과 관련된 데이터를 저장하기 위해 다른 스키마를 생성할 수도 있습니다. 각 스키마는 테이블, 뷰 및 사용자 정의 함수(UDF)와 같은 서로 다른 데이터베이스 객체를 저장할 수 있습니다. 또한 AUTHORIZATION 절을 사용하여 스키마를 생성할 수 있습니다. 이 절은 지정된 사용자에게 소유권을 부여하거나 지정된 스키마가 사용할 수 있는 최대 디스크 공간에 대한 할당량을 설정합니다.

Amazon Redshift는 모든 새 데이터베이스에 대해 public이라는 스키마를 자동으로 생성합니다. 데이터베이스 객체를 생성할 때 스키마 이름을 지정하지 않으면 객체가 public 스키마에 들어갑니다.

스키마의 객체에 액세스하려면 schema_name.table_name 표기법을 사용하여 객체를 정규화합니다. 스키마의 정규화된 이름은 점으로 구분된 스키마 이름과 테이블 이름으로 구성됩니다. 예를 들어 price 테이블이 있는 sales 스키마와 price 테이블도 있는 inventory 스키마가 있을 수 있습니다. price 테이블을 참조할 때 반드시 sales.price 또는 inventory.price로 정규화해야 합니다.

다음 예에서는 사용자 GUEST에 대해 **SALES**라는 스키마를 생성합니다.

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

더 많은 명령 옵션에 대한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 CREATE SCHEMA 섹션을 참조하세요.

데이터베이스의 스키마 목록을 보려면 다음 명령을 실행합니다.

```
select * from pg_namespace;
```

출력은 다음과 유사합니다.

```
  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}
```

카탈로그 테이블을 쿼리하는 방법에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 카탈로그 테이블 쿼리를 참조하세요.

GRANT 문을 사용하여 사용자에게 스키마에 대한 권한을 부여합니다.

다음 예에서는 SELECT 문을 사용하여 SALES 스키마의 모든 테이블 또는 뷰에서 데이터를 선택할 수 있는 권한을 GUEST 사용자에게 부여합니다.

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

다음 예에서는 GUEST 사용자에게 사용 가능한 모든 권한을 한 번에 부여합니다.

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

테이블 생성

새 인스턴스를 생성한 후 데이터를 보관할 테이블을 생성합니다. 테이블을 생성할 때 열 정보를 지정합니다.

예를 들어 **DEMO**라는 테이블을 생성하려면 다음 명령을 실행합니다.

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

기본적으로 테이블과 같은 새 데이터베이스 객체는 데이터 웨어하우스 생성 중 생성되는 public이라는 기본 스키마에 생성됩니다. 다른 스키마를 사용하여 데이터베이스 객체를 생성할 수 있습니다. 스키마에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 데이터베이스 보안 관리 섹션을 참조하세요.

schema_name.object_name 표기법으로 테이블을 생성하여 SALES 스키마에서 테이블을 생성할 수도 있습니다.

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

스키마와 해당 테이블을 보고 검사하려면 Amazon Redshift 쿼리 에디터 v2를 사용합니다. 또는 시스템 뷰를 사용하여 스키마의 테이블 목록을 볼 수 있습니다. 자세한 내용은 시스템 테이블 및 뷰 쿼리 단원을 참조하십시오.

Amazon Redshift는 encoding, distkey 및 sortkey 열을 병렬 처리에 사용합니다. 이러한 요소들을 통합하는 테이블의 설계에 대한 자세한 내용은 Amazon Redshift 테이블 설계 모범 사례 섹션을 참조하세요.

테이블에 데이터 행 삽입

테이블을 생성한 후 해당 테이블에 데이터 행을 삽입합니다.

> **Note**
>
> INSERT 명령은 테이블에 행을 삽입합니다. 표준 대량 로드는 COPY 명령을 사용합니다. 자세한 내용은 COPY 명령을 사용하여 데이터 로드 섹션을 참조하세요.

예를 들어 DEMO 테이블에 값을 삽입하려면 다음 명령을 실행합니다.

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

특정 스키마에 있는 테이블에 데이터를 삽입하려면 다음 명령을 실행합니다.

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

테이블에서 데이터 선택

테이블을 만들고 데이터로 채운 다음 SELECT 문을 사용하여 테이블에 포함된 데이터를 표시합니다. SELECT * 문은 테이블의 모든 데이터에 대한 모든 열 이름과 행 값을 반환합니다. SELECT를 사용하면 최근에 추가된 데이터가 테이블에 제대로 삽입되었는지 확인할 수 있습니다.

DEMO 테이블에 입력한 데이터를 보려면 다음 명령을 실행합니다.

```
SELECT * from DEMO;
```

결과는 다음과 같아야 합니다.

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

SELECT 문으로 테이블 쿼리에 대한 자세한 내용은 SELECT 섹션을 참조하세요.

데이터 로드

이 안내서에 나오는 많은 예에서는 TICKIT 샘플 데이터세트를 사용합니다. 개별 샘플 데이터 파일이 들어 있는 tickitdb.zip 파일을 다운로드할 수 있습니다. 그런 다음, 샘플 데이터를 자체 Amazon S3 버킷에 업로드할 수 있습니다.

데이터베이스의 샘플 데이터를 로드하려면 먼저 테이블을 생성합니다. 그런 다음 COPY 명령을 사용하여 Amazon S3 버킷에 저장된 샘플 데이터가 있는 테이블을 로드합니다. 테이블을 생성하고 샘플 데이터를 로드하는 단계는 4단계: Amazon S3에서 Amazon Redshift로 데이터 로드 섹션을 참조하세요.

시스템 테이블 및 뷰 쿼리

데이터 웨어하우스에는 생성한 테이블 외에도 많은 시스템 테이블 및 뷰가 포함되어 있습니다. 이러한 테이블 및 뷰에는 설치 및 시스템에서 실행 중인 다양한 쿼리 및 프로세스에 대한 정보가 포함되어 있습니다. 이러한 시스템 테이블 및 뷰를 쿼리하여 데이터베이스에 대한 정보를 수집할 수 있습니다. 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 시스템 테이블 및 뷰 참조를 참조하세요. 각 테이블 또는 뷰에 대한 설명은 테이블이 모든 사용자에게 표시되는지 아니면 슈퍼 사용자에게만 표시되는지를 나타냅니다. 슈퍼 사용자에게만 보이는 테이블을 쿼리하려면 슈퍼 사용자로 로그인합니다.

테이블 이름 목록 보기

스키마의 모든 테이블의 목록을 보려면 PG_TABLE_DEF 시스템 카탈로그 테이블을 쿼리하면 됩니다. 먼저 search_path에 대한 설정을 검토할 수 있습니다.

```
SHOW search_path;
```

결과는 다음과 같습니다.

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

다음 예에서는 검색 경로에 SALES 스키마를 추가하고 SALES 스키마의 모든 테이블을 표시합니다.

```
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
```

다음 예는 현재 데이터베이스의 모든 스키마에서 DEMO라는 모든 테이블의 목록을 보여줍니다.

```
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
```

자세한 내용은 PG_TABLE_DEF 섹션을 참조하세요.

Amazon Redshift 쿼리 에디터 v2를 사용하면 먼저 연결할 데이터베이스를 선택하여 지정된 스키마의 모든 테이블을 볼 수도 있습니다.

사용자 보기

PG_USER 카탈로그를 쿼리하여 모든 사용자의 목록과 사용자 ID(USESYSID) 및 사용자 권한을 볼 수 있습니다.

```
SELECT * FROM pg_user;

  usename   | usesysid | usecreatedb | usesuper | usecatupd |  passwd  | valuntil |
 useconfig
```

```
------------+----------+-------------+----------+-----------+----------+----------
+-----------
 rdsdb      |        1 | true        | true     | true      | ******** | infinity |
 awsuser    |      100 | true        | true     | false     | ******** |          |
 guest      |      104 | true        | false    | false     | ******** |          |
```

사용자 이름 rdsdb는 Amazon Redshift가 일상적 관리 및 유지 태스크를 위해 내부적으로 사용합니다. SELECT 문에 where usesysid > 1을 추가하면 사용자가 정의한 사용자 이름만 표시되도록 쿼리를 필터링할 수 있습니다.

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

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

최근 쿼리 보기

앞의 예에서 adminuser의 사용자 ID(user_id)는 100입니다. adminuser가 실행한 가장 최근의 쿼리 4건을 나열하려면 SYS_QUERY_HISTORY 뷰를 쿼리하면 됩니다.

이 뷰를 사용하여 최근에 실행된 쿼리의 쿼리 ID(query_id) 또는 프로세스 ID(session_id)를 찾을 수 있습니다. 이 뷰를 사용하여 쿼리를 완료하는 데 걸린 시간을 확인할 수도 있습니다. SYS_QUERY_HISTORY에는 특정 쿼리를 찾는 데 도움이 되는 쿼리 문자열(query_text)의 처음 4,000자가 포함되어 있습니다. 결과를 제한하려면 SELECT 문과 함께 LIMIT 절을 사용합니다.

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

결과는 다음과 같습니다.

```
 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);
```

실행 중인 쿼리의 세션 ID 확인

쿼리에 대한 시스템 테이블 정보 검색을 위해 해당 쿼리와 연결된 세션 ID(프로세스 ID)를 지정해야 할 수 있습니다. 또는 아직 실행 중인 쿼리의 세션 ID를 찾아야 할 수도 있습니다. 예를 들어 프로비저닝된 클러스터에서 실행하는 데 너무 오래 걸리는 쿼리를 취소해야 하는 경우 세션 ID가 필요합니다. STV_RECENTS 시스템 테이블을 쿼리하여 실행 중인 쿼리의 세션 ID 목록과 함께 해당 쿼리 문자열을 가져올 수 있습니다. 쿼리가 여러 세션을 반환하는 경우 쿼리 텍스트를 보고 필요한 세션 ID를 파악할 수 있습니다.

실행 중인 쿼리의 세션 ID를 확인하려면 다음 SELECT 문을 실행합니다.

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

쿼리 취소

시간이 너무 오래 걸리거나 리소스를 과도하게 사용하는 쿼리를 실행하는 경우 해당 쿼리를 취소합니다. 예를 들어 티켓 판매자의 이름과 판매된 티켓 수량이 포함된 티켓 판매자 목록을 생성합니다. 다음 쿼리는 SALES 테이블과 USERS 테이블에서 데이터를 선택하고 WHERE 절에서 SELLERID와 USERID를 일치시켜 두 테이블을 조인합니다.

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

결과는 다음과 같습니다.

```
 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**
>
> 이것은 복잡한 쿼리입니다. 이 자습서에서는 이 쿼리가 어떻게 구성되는지에 대해서는 걱정하지 않아도 됩니다.

이전의 쿼리는 몇 초 안에 실행되어 2,102개의 행을 반환합니다.

WHERE 절을 넣는 것을 잊었다고 가정해 보겠습니다.

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

결과 집합에는 SALES 테이블의 모든 행과 USERS 테이블의 모든 행을 곱한 값(49989*3766)이 포함될 것입니다. 이것을 데카르트 조인이라고 하는데, 권장되지 않습니다. 결과는 1억 8,800만 개 이상의 행으로 실행하려면 시간이 많이 걸립니다.

실행 중인 쿼리를 취소하려면 쿼리의 세션 ID와 함께 CANCEL 명령을 사용합니다. Amazon Redshift 쿼리 에디터 v2를 사용하면 쿼리가 실행되는 동안 취소 버튼을 선택하여 쿼리를 취소할 수 있습니다.

세션 ID를 찾으려면 이전 단계에 나온 것처럼 새 세션을 시작하고 STV_RECENTS 테이블을 쿼리합니다. 다음 예에서는 결과를 더 읽기 쉽게 만드는 방법을 보여줍니다. 이렇게 하려면 TRIM 함수를 사용하여 후행 공백을 자르고 쿼리 문자열의 처음 20자만 표시합니다.

실행 중인 쿼리의 세션 ID를 확인하려면 다음 SELECT 문을 실행합니다.

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

결과는 다음과 같습니다.

```
 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  ...
```

세션 ID가 1073791534인 쿼리를 취소하려면 다음 명령을 실행합니다.

```
CANCEL 1073791534;
```

> **Note**
>
> CANCEL 명령은 트랜잭션을 중지하지 않습니다. 트랜잭션을 중지하거나 롤백하려면 ABORT 또는 ROLLBACK 명령을 사용해야 합니다. 트랜잭션에 연결된 쿼리를 취소하려면 먼저 쿼리를 취소한 다음 트랜잭션을 중지합니다.

취소한 쿼리가 트랜잭션과 연결된 경우 ABORT 또는 ROLLBACK 명령을 사용하여 트랜잭션을 취소하고 데이터에 대한 변경 사항을 버립니다.

```
ABORT;
```

수퍼유저로 로그인하지 않은 경우, 본인의 쿼리만 취소할 수 있습니다. 수퍼유저는 모든 쿼리를 취소할 수 있습니다.

동시에 실행되는 쿼리를 쿼리 도구가 지원하지 않는 경우 쿼리를 취소하기 위해 다른 세션을 시작합니다.

쿼리 취소에 관한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 CANCEL 단원을 참조하세요.

수퍼유저 대기열을 사용해 쿼리 취소

현재 세션에 동시에 실행 중인 쿼리가 너무 많다면 다른 쿼리가 끝나기 전까지는 CANCEL 명령을 실행하지 못할 수 있습니다. 이런 경우 다른 워크로드 관리 쿼리 대기열을 사용해 CANCEL 명령을 실행합니다.

워크로드 관리를 사용하면 다른 쿼리 대기열에 있는 쿼리를 실행할 수 있기 때문에 다른 쿼리가 완료되기를 기다릴 필요가 없습니다. 워크로드 관리자는 Superuser 쿼리라고 하는, 문제 해결에 사용할 수 있는 별도의 쿼리를 생성합니다. Superuser 쿼리를 사용하려면 슈퍼 사용자로 로그인하고 SET 명령을 사용해 쿼리 그룹을 'superuser'로 설정해야 합니다. 명령을 실행한 후 RESET 명령을 사용하여 쿼리 그룹을 리셋합니다.

슈퍼 사용자 대기열을 사용해 쿼리를 취소하려면 다음 명령을 실행합니다.

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

Amazon Redshift 데이터베이스에 없는 데이터 쿼리

다음에서 Amazon S3 데이터, 원격 데이터베이스 관리자, 원격 Amazon Redshift 데이터베이스, Amazon Redshift를 사용한 기계 학습(ML) 모델 훈련을 비롯한 원격 소스에서 데이터 쿼리를 시작하는 방법에 대한 정보를 찾을 수 있습니다.

주제

- 데이터 레이크 쿼리
- 원격 데이터베이스 관리자에서 데이터 쿼리
- 다른 Amazon Redshift 데이터베이스의 데이터 액세스
- Amazon Redshift 데이터로 기계 학습 모델 훈련

데이터 레이크 쿼리

Amazon Redshift Spectrum을 사용하여 데이터를 Amazon Redshift 테이블에 로드하지 않고도 Amazon S3 파일에서 데이터를 쿼리할 수 있습니다. Amazon Redshift는 Amazon Redshift 클러스터와 Amazon S3 데이터 레이크 모두에 저장된 초대형 데이터 집합의 빠른 온라인 분석 처리(OLAP)를 위해 설계된 SQL 기능을 제공합니다. Parquet, ORC, RCFile, TextFile, SequenceFile, RegexSerde, OpenCSV 및 AVRO를 비롯한 다양한 형식의 데이터를 쿼리할 수 있습니다. Amazon S3에서 파일 구조를 정의하려면 외부 스키마와 테이블을 생성합니다. 그런 다음 AWS Glue 또는 자체 Apache Hive 메타스토어와 같은 외부 데이터 카탈로그를 사용합니다. 데이터 카탈로그의 변경 사항은 Amazon Redshift 클러스터에서 즉시 사용할 수 있습니다.

데이터가 AWS Glue Data Catalog에 등록되고 AWS Lake Formation에서 사용할 수 있으면 Redshift Spectrum으로 데이터를 쿼리할 수 있습니다.

Redshift Spectrum은 클러스터와 독립적인 전용 Amazon Redshift 서버에 상주합니다. Redshift Spectrum은 조건자 필터링 및 집계 같은 많은 컴퓨팅 집약적 태스크를 Redshift Spectrum 계층에 푸시합니다. 또한 Redshift Spectrum은 대규모 병렬 처리를 활용하도록 지능적으로 확장됩니다.

하나 이상의 열에서 외부 테이블을 분할하여 파티션 제거를 통해 쿼리 성능을 최적화할 수 있습니다. Amazon Redshift 테이블을 사용하여 외부 테이블을 쿼리하고 조인할 수 있습니다. 여러 Amazon Redshift 클러스터에서 외부 테이블에 액세스하고 동일한 AWS 리전의 모든 클러스터에서 Amazon S3 데이터를 쿼리할 수 있습니다. Amazon S3 데이터 파일을 업데이트하면 어느 Amazon Redshift 클러스터에서나 해당 데이터를 즉시 사용할 수 있습니다.

Redshift Spectrum 및 데이터 레이크 작업 방법을 포함하여 Redshift Spectrum에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 Amazon Redshift Spectrum 시작하기를 참조하세요.

원격 데이터베이스 관리자에서 데이터 쿼리

페더레이션 쿼리를 사용하여 Amazon RDS 데이터베이스와 Amazon Aurora 데이터베이스의 데이터를 Amazon Redshift 데이터베이스의 데이터와 조인할 수 있습니다. Amazon Redshift를 사용하여 운영 데이터를 이동하지 않고 직접 쿼리하고, 변환을 적용하고, Redshift 테이블에 데이터를 삽입할 수 있습니다. 연합 쿼리에 대한 계산 중 일부는 원격 데이터 원본에 배포됩니다.

연합 쿼리를 실행하기 위해 Amazon Redshift는 먼저 원격 데이터 원본에 연결합니다. 그런 다음 Amazon Redshift는 원격 데이터 원본의 테이블에 대한 메타데이터를 검색하고 쿼리를 실행한 다음 결과 행을 검색합니다. 그런 다음 Amazon Redshift는 추가 처리를 위해 결과 행을 Amazon Redshift 컴퓨팅 노드에 배포합니다.

연합 쿼리를 위한 환경 설정에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 다음 주제 중 하나를 참조하세요.

- PostgreSQL에 대한 페더레이션 쿼리 사용 시작하기
- MySQL에 대한 연합 쿼리 사용 시작하기

다른 Amazon Redshift 데이터베이스의 데이터 액세스

Amazon Redshift 데이터 공유를 사용하면 Amazon Redshift 클러스터 또는 AWS 계정 간에 읽기용으로 라이브 데이터를 더 안전하고 쉽게 공유할 수 있습니다. 데이터를 수동으로 복사하거나 이동할 필요 없이 Amazon Redshift 클러스터 전반에서 데이터에 대한 즉각적이고 세분화된 고성능 액세스를 사용할 수 있습니다. Amazon Redshift 클러스터에서 업데이트되는 최신 정보와 일관된 정보를 사용자가 볼 수 있습니다. 데이터베이스, 스키마, 테이블, 보기(일반, 후기 바인딩 및 구체화된 보기 포함) 및 SQL 사용자 정의 함수(UDF) 등의 여러 수준에서 데이터를 공유할 수 있습니다.

Amazon Redshift datashare는 다음과 같은 사용 사례에 특히 유용합니다.

- 비즈니스 크리티컬 워크로드 중앙 집중화 – 여러 비즈니스 인텔리전스(BI) 또는 분석 클러스터와 데이터를 공유하는 중앙 추출, 변환 및 로드(ETL) 클러스터를 사용합니다. 이 접근 방식은 개별 워크로드에 대한 읽기 워크로드 격리 및 차지백을 제공합니다.
- 환경 간 datashare – 개발, 테스트 및 프로덕션 환경 간에 데이터를 공유합니다. 다양한 세분화된 수준에서 데이터를 공유하여 팀 민첩성을 향상시킬 수 있습니다.

datashare에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 데이터 공유 작업 관리를 참조하세요.

Amazon Redshift 데이터로 기계 학습 모델 훈련

Amazon Redshift 기계 학습으로 Amazon Redshift에 데이터를 제공하여 모델을 훈련할 수 있습니다. 그런 다음 Amazon Redshift 기계 학습은 입력 데이터의 패턴을 캡처하는 모델을 생성합니다. 이러한 모델을 사용하여 추가 비용 없이 새로운 입력 데이터에 대한 예측을 생성할 수 있습니다. Amazon Redshift 기계 학습을 사용하면 SQL 문으로 기계 학습 모델을 훈련하고 예측을 위해 SQL 쿼리에서 해당 기계 학습 모델을 호출할 수 있습니다. 반복적으로 파라미터를 변경하고 훈련 데이터를 개선하여 예측의 정확도를 계속해서 향상시킬 수 있습니다.

Amazon Redshift 기계 학습을 사용하면 SQL 사용자가 익숙한 SQL 명령으로 기계 학습 모델을 보다 쉽게 생성, 훈련 및 배포할 수 있습니다. Amazon Redshift ML을 사용하면 Amazon Redshift 클러스터의 데이터와 Amazon SageMaker AI Autopilot으로 모델을 훈련하여 자동으로 최상의 모델을 얻을 수 있습니다. 그런 다음 모델을 현지화하고 Amazon Redshift 데이터베이스 내에서 예측을 수행할 수 있습니다.

Amazon Redshift 기계 학습에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서의 Amazon Redshift ML 시작하기 섹션을 참조하세요.

Amazon Redshift 개념 알아보기

Amazon Redshift Serverless를 사용하면 프로비저닝된 데이터 웨어하우스를 구성하지 않아도 데이터를 액세스하고 분석할 수 있습니다. 리소스가 자동으로 프로비저닝하고 데이터 웨어하우스 용량이 지능적으로 크기 조정되어 가장 까다롭고 예측할 수 없는 워크로드에도 빠른 성능을 제공합니다. 데이터 웨어하우스가 유휴 상태일 때는 요금이 발생하지 않으므로 사용량에 대한 요금만 지불합니다. Amazon Redshift 쿼리 에디터 v2 또는 자주 사용하는 비즈니스 인텔리전스(BI) 도구에서 바로 데이터를 로드하고 쿼리를 시작할 수 있습니다. 사용하기 쉽고 관리가 필요 없는 환경에서 최고의 가격 대비 성능과 친숙한 SQL 기능을 활용하세요.

Amazon Redshift를 처음 사용할 경우 먼저 다음 섹션을 읽을 것을 권장합니다.

- Amazon Redshift Serless 기능 개요 – 이 주제에서는 Amazon Redshift Serless의 개요와 주요 기능을 찾아볼 수 있습니다.
- 서비스 하이라이트 및 요금 – 이 제품 세부 정보 페이지에서 Amazon Redshift Serverless의 하이라이트 및 요금에 대한 세부 정보를 찾아볼 수 있습니다.
- Amazon Redshift Serverless 데이터 웨어하우스 시작하기 – 이 주제에서는 Amazon Redshift Serverless 데이터 웨어하우스를 생성하고 쿼리 에디터 v2를 사용하여 데이터 쿼리를 시작하는 방법에 대해 자세히 알아볼 수 있습니다.

Amazon Redshift 리소스를 수동으로 관리하려는 경우 데이터 쿼리 요구 사항에 맞게 프로비저닝된 클러스터를 생성할 수 있습니다. 자세한 내용은 Amazon Redshift 클러스터 섹션을 참조하세요.

조직에 자격이 있으며 Amazon Redshift Serverless를 사용할 수 없는 AWS 리전에 클러스터가 생성되고 있는 경우, Amazon Redshift 무료 평가판 프로그램에서 클러스터를 생성하는 것이 가능할 수 있습니다. 이 클러스터를 어떤 용도로 사용할 계획입니까?라는 질문에 대해 프로덕션 또는 무료 평가판을 선택하여 답합니다. 무료 평가판을 선택하면 dc2.large 노드 유형으로 구성을 생성합니다. 무료 평가판 선택에 대한 자세한 내용은 Amazon Redshift 무료 평가판을 참조하세요. Amazon Redshift Serverless를 사용할 수 있는 AWS 리전 목록은 Amazon Web Services 일반 참조에서 Redshift Serverless API에 대해 나열된 Amazon Redshift 엔드포인트를 참조하세요.

다음은 몇 가지 주요 Amazon Redshift Serverless 개념입니다.

- 네임스페이스 – 데이터베이스 객체와 사용자의 모음입니다. 네임스페이스는 스키마, 테이블, 사용자, 데이터 공유 및 스냅샷과 같이 Amazon Redshift Serverless에서 사용하는 모든 리소스를 함께 그룹화합니다.

- 작업 그룹 – 컴퓨팅 리소스의 모음입니다. 작업 그룹에는 Amazon Redshift Serverless가 컴퓨팅 작업을 실행하는 데 사용하는 컴퓨팅 리소스가 있습니다. 이러한 리소스의 예로는 Redshift 처리 단위(RPU), 보안 그룹, 사용 제한이 있습니다. 작업 그룹에는 Amazon Redshift Serverless 콘솔, AWS Command Line Interface 또는 Amazon Redshift Serverless API를 사용하여 구성할 수 있는 네트워크 및 보안 설정이 있습니다.

네임스페이스 및 작업 그룹 리소스 구성에 대한 자세한 내용은 네임스페이스 작업 및 작업 그룹 사용을 참조하세요.

다음은 Amazon Redshift 프로비저닝 클러스터에 대한 몇 가지 주요 개념입니다.

- 클러스터 – Amazon Redshift 데이터 웨어하우스의 핵심 인프라 구성 요소는 클러스터입니다.

  클러스터는 하나 이상의 컴퓨팅 노드로 구성됩니다. 컴퓨팅 노드는 컴파일된 코드를 실행합니다.

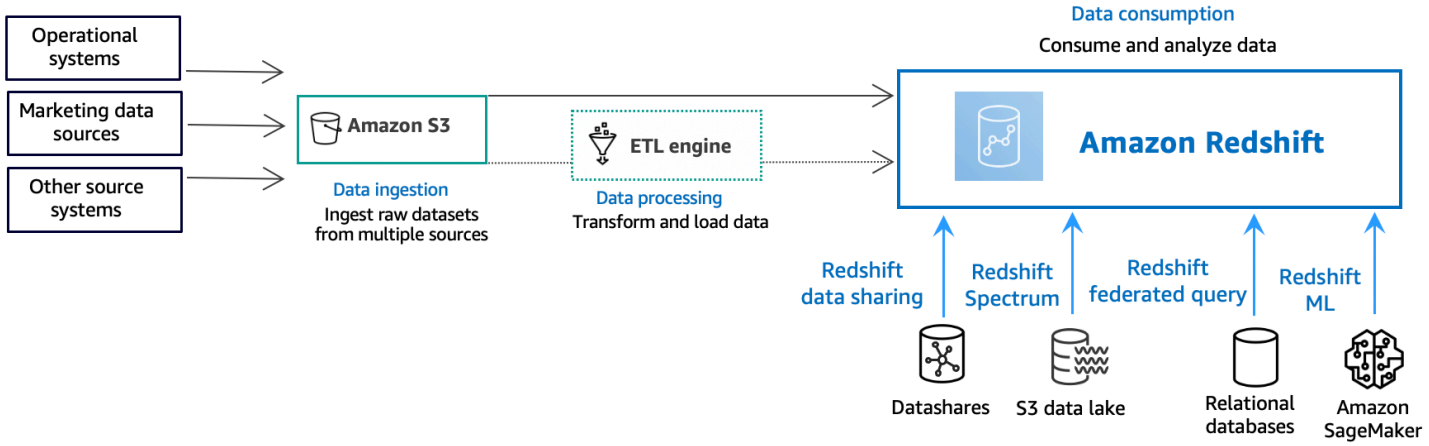
  클러스터에 두 개 이상의 컴퓨팅 노드가 제공된 경우 추가 리더 노드가 컴퓨팅 노드를 조정합니다. 리더 노드는 비즈니스 인텔리전스 도구 및 쿼리 편집기와 같은 애플리케이션과의 외부 통신을 처리합니다. 클라이언트 애플리케이션은 리더 노드와만 직접 상호작용합니다. 컴퓨팅 노드는 외부 애플리케이션에서 인식됩니다.
- 데이터베이스 – 클러스터에는 하나 이상의 데이터베이스가 포함되어 있습니다.

  사용자 데이터는 컴퓨팅 노드에 있는 하나 이상의 데이터베이스에 저장됩니다. SQL 클라이언트는 리더 노드와 통신하고, 리더 노드는 이어서 컴퓨팅 노드와 쿼리 실행을 조정합니다. 컴퓨팅 노드 및 리더 노드에 대한 자세한 내용은 데이터 웨어하우스 시스템 아키텍처를 참조하세요. 데이터베이스 내에서 사용자 데이터는 하나 이상의 스키마로 구성됩니다.

  Amazon Redshift는 관계형 데이터베이스 관리 시스템(RDBMS)이므로 다른 RDBMS 애플리케이션과 호환됩니다. 데이터 삽입 및 삭제와 같은 온라인 트랜잭션 처리(OLTP) 기능을 포함하여 일반적인 RDBMS와 동일한 기능을 제공합니다. 또한 Amazon Redshift는 고성능 배치 분석 및 데이터 집합 보고에 최적화되어 있습니다.

다음은 Amazon Redshift의 일반적인 데이터 처리 흐름에 대한 설명과 흐름의 여러 부분에 대한 설명을 찾을 수 있습니다. Amazon Redshift 시스템 아키텍처에 대한 자세한 내용은 데이터 웨어하우스 시스템 아키텍처를 참조하세요.

다음 다이어그램은 Amazon Redshift의 일반적인 데이터 처리 흐름을 보여줍니다.



Amazon Redshift 데이터웨어 하우스는 엔터프라이즈 급 관계형 데이터베이스 쿼리 및 관리 시스템입니다. Amazon Redshift는 비즈니스 인텔리전스(BI), 보고, 데이터 및 분석 도구를 비롯한 다양한 유형의 애플리케이션과의 클라이언트 연결을 지원합니다. 분석 쿼리를 실행할 때는 다단계 작업을 통해 대용량의 데이터를 가져와서 비교하고 평가하면서 최종 결과를 산출합니다.

데이터 수집 계층에서 다양한 유형의 데이터 원본이 정형, 반정형 또는 비정형 데이터를 데이터 스토리지 계층에 지속적으로 업로드합니다. 이 데이터 스토리지 영역은 다양한 소비 준비 상태로 데이터를 저장하는 준비 영역 역할을 합니다. 스토리지의 예로는 Amazon Simple Storage Service(Amazon S3) 버킷이 있습니다.

선택적 데이터 처리 계층에서 소스 데이터는 추출, 변환, 로드 파이프라인을 사용하여 사전 처리, 검증 및 변환을 거칩니다. 그런 다음 이러한 원시 데이터 집합은 추출, 변환, 로드 작업을 사용하여 구체화됩니다. 추출, 변환, 로드 엔진의 예는 AWS Glue입니다.

데이터 소비 계층에서 데이터는 분석 워크로드를 실행할 수 있는 Amazon Redshift 클러스터로 로드됩니다.

분석 워크로드의 몇 가지 예는 데이터 소스 외부 쿼리를 참조하세요.

Amazon Redshift에 대해 알아볼 수 있는 추가 리소스

Amazon Redshift Serverless에 대해 자세히 알아보려면 다음 Amazon Redshift 리소스를 사용하여 이 안내서에 소개된 개념에 대해 계속해서 자세히 알아보는 것이 좋습니다.

- 기능 동영상: 이 동영상은 Amazon Redshift 기능에 대해 알아보는 데 도움이 됩니다.
  - Amazon Redshift Serverless를 개괄적으로 살펴보려면 다음 동영상을 시청하세요. 90초 만에 설명하는 Amazon Redshift Serverless
  - 서버리스 데이터 웨어하우스를 설정하고 데이터 쿼리를 시작하는 방법을 알아보려면 다음 동영상을 시청하세요. Amazon Redshift Serverless 시작하기
- Amazon Redshift 관리 가이드: 이 가이드는 이 Amazon Redshift 시작 안내서를 기반으로 합니다. 이 안내서는 Amazon Redshift Serverless 및 Amazon Redshift 프로비저닝 클러스터의 생성, 관리 및 모니터링에 대한 개념 및 태스크에 대해 깊이 있게 소개합니다.
- Amazon Redshift 데이터베이스 개발자 안내서: 이 안내서도 Amazon Redshift 시작 안내서를 기반으로 합니다. 이 안내서는 사용자의 데이터 웨어하우스를 구성하는 데이터베이스를 설계, 구축, 쿼리하는 방법에 대한 깊이 있는 정보를 데이터베이스 개발자에게 제공합니다.
  - SQL 참조: 이 주제에서는 Amazon Redshift에 대한 SQL 명령 및 함수 참조에 대해 설명합니다.
  - 시스템 테이블 및 뷰 참조: 이 주제에서는 Amazon Redshift의 시스템 테이블과 뷰에 대해 설명합니다.
- Amazon Redshift 자습서: 이 주제에서는 Amazon Redshift 기능에 대한 자습서를 소개합니다.
  - Amazon S3에서 데이터 로드: 이 튜토리얼에서는 Amazon S3 버킷의 데이터 파일에서 Amazon Redshift 데이터베이스 테이블로 데이터를 로드하는 방법을 설명합니다.
  - 데이터 공유 시작하기: 이 섹션에서는 다른 Amazon Redshift 클러스터에서 데이터를 공유하고 액세스하는 방법을 설명합니다.
  - Amazon Redshift에서 함께 공간 SQL 함수 사용: 이 튜토리얼에서는 Amazon Redshift에서 일부 공간 SQL 함수를 사용하는 방법을 설명합니다.
  - Amazon Redshift Spectrum으로 중첩 데이터 쿼리: 이 튜토리얼에서는 Redshift Spectrum을 사용하여 외부 테이블을 사용하여 Parquet, ORC, JSON 및 Ion 파일 형식의 중첩 데이터를 쿼리하는 방법을 설명합니다.
  - 수동 워크로드 관리(WLM) 대기열 구성: 이 튜토리얼에서는 Amazon Redshift에서 수동 워크로드 관리(WLM)를 구성하는 방법을 설명합니다.
  - Amazon Redshift ML 시작하기: 이 섹션에서는 사용자가 익숙한 SQL 명령으로 기계 학습 모델을 생성, 훈련, 배포하는 방법을 설명합니다.

- 새로운 소식: 이 웹 페이지에는 Amazon Redshift의 새로운 기능과 제품 업데이트가 나와 있습니다.

문서 기록

> **Note**
>
> Amazon Redshift의 새로운 기능에 대한 설명은 새로운 소식을 참조하세요.

다음 테이블에서는 Amazon Redshift 시작 안내서에 대한 중요한 설명서 변경 사항에 대해 설명합니다.

변경 사항	설명	릴리스 날짜
설명서 업데이트	Query Editor V2 관리형 정책 변경 사항과 향상된 서버리스 네임스페이스 및 작업 그룹 액세스 권한을 반영하여 가이드를 업데이트했습니다.	2024년 2월 21일
설명서 업데이트	최신 콘솔 인터페이스의 향상된 기능과 Query Editor V2의 개선 사항을 반영하여 스크린샷 및 절차를 업데이트했습니다.	2023년 3월 11일
새로운 특성	Amazon Redshift Serverless 시작 절차 및 워크플로를 포함하도록 가이드를 업데이트했습니다. 서버리스 데이터 웨어하우스 생성 및 관리에 대한 포괄적인 섹션을 추가했습니다.	2022년 7월 12일
설명서 업데이트	레거시 쿼리 편집기에 대한 언급을 대체하며, Query Editor V2가 기본 쿼리 인터페이스가 된 것을 반영하여 가이드를 업데이트했습니다.	2022년 2월
설명서 업데이트	일반 데이터베이스 태스크 시작하기, 데이터 레이크 쿼리, 원격 소스에서 데이터 쿼리, datashare, Amazon Redshift 데이터로 기계 학습 모델 훈련에 대한 새로운 섹션을 포함하도록 안내서를 업데이트했습니다.	2021년 6월 30일
새로운 특성	새 샘플 로드 절차를 설명하도록 안내서를 업데이트했습니다.	2021년 6월 4일
설명서 업데이트	원래 Amazon Redshift 콘솔을 제거하고 단계 흐름을 개선하도록 안내서를 업데이트했습니다.	2020년 8월 14일

변경 사항	설명	릴리스 날짜
새 콘솔	새 Amazon Redshift 콘솔을 설명하도록 안내서를 업데이트했습니다.	2019년 11월 11일
새로운 특성	빠른 클러스터 시작 절차를 설명하기 위해 안내서가 업데이트되었습니다.	2018년 8월 10일
새로운 특성	Amazon Redshift 대시보드에서 클러스터를 시작하도록 안내서를 업데이트했습니다.	2015년 7월 28일
새로운 특성	새 노드 유형 이름을 사용하도록 안내서가 업데이트되었습니다.	2015년 6월 9일
설명서 업데이트	VPC 보안 그룹 구성에 대한 스크린샷 및 절차가 업데이트되었습니다.	2015년 4월 30일
설명서 업데이트	현재 콘솔에 맞게 스크린샷 및 절차가 업데이트되었습니다.	2014년 11월 12일
설명서 업데이트	Amazon S3에서 데이터 로드 정보가 자체 섹션으로 이동되었으며 다음 단계 섹션이 마지막 단계로 이동되어 찾기 쉬워졌습니다.	2014년 5월 13일
설명서 업데이트	환영합니다 페이지를 없애고 해당 콘텐츠를 기본 시작하기 페이지에 통합하였습니다.	2014년 3월 14일
설명서 업데이트	이 설명서는 고객 피드백 및 서비스 업데이트가 반영된 Amazon Redshift 시작 안내서의 최신 릴리스입니다.	2014년 3월 14일
새 안내서	이 설명서는 Amazon Redshift 시작 안내서의 최초 릴리스입니다.	2013년 14월 2일