
Amazon Redshift

시작 안내서



Amazon Redshift: 시작 안내서

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

Amazon의 상표 및 브랜드 디자인은 Amazon 외 제품 또는 서비스와 함께, Amazon 브랜드 이미지를 떨어뜨리거나 고객에게 혼동을 일으킬 수 있는 방식으로 사용할 수 없습니다. Amazon이 소유하지 않은 다른 모든 상표는 각 소유자의 자산으로, 해당 소유자는 Amazon과 제휴를 맺었거나 Amazon과 관련이 있거나 Amazon의 후원을 받는 회사일 수도 있고, 아닐 수도 있습니다.

Table of Contents

Amazon Redshift 시작하기	1
사전 조건	1
AWS에 가입	1
방화벽 규칙 확인	2
클러스터 및 데이터 로드 시작하기	3
샘플 데이터 집합 사용	3
1단계: 샘플 Amazon Redshift 클러스터 생성	3
2단계: 쿼리 편집기를 사용하여 예제 쿼리를 사용해보십시오.	5
Amazon Redshift 에 자체 데이터 가져오기	6
1단계: IAM 역할 생성	6
2단계: 클러스터 생성	7
3단계: SQL 클라이언트에 대한 인바운드 규칙 구성	8
4단계: 쿼리 편집기에 대한 액세스 권한 부여	9
5단계: 샘플 데이터 로드	11
6단계: 쿼리 편집기를 사용하여 예제 쿼리를 사용해보십시오.	14
단계 7: 환경 재설정	15
일반적인 데이터베이스 작업 시작하기	16
작업 1: 데이터베이스 생성	16
작업 2: 사용자를 생성합니다	16
작업 3: 스키마 생성	17
작업 4: 테이블 생성	18
테이블에 데이터 행 삽입	18
테이블에서 데이터 삭제	19
작업 5: 샘플 데이터 로드	19
작업 6: 시스템 테이블 쿼리	19
테이블 이름 목록 보기	20
사용자 보기	21
최근 쿼리 보기	21
실행 중인 쿼리의 프로세스 ID 확인	22
작업 7: 쿼리 취소	22
다른 세션에서 쿼리 취소	24
수퍼유저 대기열을 사용해 쿼리 취소	24
작업 8: 리소스 정리	24
데이터 레이크 쿼리 시작하기	25
원격 데이터 소스 쿼리 시작하기	26
다른 클러스터의 데이터 액세스 시작하기	27
Redshift 데이터로 ML 모델 교육 시작하기	28
추가 리소스	29
문서 기록	30
.....	xxxi

Amazon Redshift 시작하기

을 소개합니다. Amazon Redshift 시작 안내서. Amazon Redshift 는 완벽하게 관리되는 페타바이트급 데이터 웨어하우스 서비스입니다. AWS 클라우드 . Amazon Redshift 데이터 웨어하우스는 컴퓨팅 리소스의 모음입니다. 노드라는 그룹으로 구성되어 있습니다. 클러스터. 각 클러스터는 Amazon Redshift 엔진을 실행하며, 하나 이상의 데이터베이스를 포함합니다.

Amazon Redshift 사용자를 처음 사용할 경우 먼저 다음 단원을 읽는 것이 좋습니다.

- [Amazon Redshift 관리 개요](#)— 이 주제에서는 Amazon Redshift 개요를 확인할 수 있습니다.
- [서비스 하이라이트 및 가격](#)— 이 상품 상세 페이지에서 Amazon Redshift 서비스의 주요 특징 및 가격에 대한 상세 정보를 확인할 수 있습니다.
- [Amazon Redshift 시작하기](#)(이 안내서의) — 이 안내서에서는 Amazon Redshift 사용 자습서를 통해 샘플 클러스터를 생성하고 샘플 데이터로 작업하는 자습서를 확인할 수 있습니다.

이 안내서에서는 다음 중 하나를 안내하는 자습서를 찾을 수 있습니다.

- [Amazon Redshift 클러스터 및 데이터 로드 시작하기 \(p. 3\)](#)
- [일반적인 데이터베이스 작업 시작하기 \(p. 16\)](#)
- [데이터 레이크 쿼리 시작하기 \(p. 25\)](#)
- [원격 데이터 원본의 데이터 쿼리 시작 \(p. 26\)](#)
- [다른 Amazon Redshift 클러스터의 데이터 액세스 시작 \(p. 27\)](#)
- [Amazon Redshift 데이터로 머신 러닝 모델 교육 시작하기 \(p. 28\)](#)

조직에 적합한 경우 Amazon Redshift 무료 평가판 프로그램에 따라 클러스터를 생성할 수 있습니다. 이렇게 하려면 무료 평가판을 클릭하여 dc2.large 노드 유형으로 구성을 만듭니다. 무료 평가판 선택에 대한 자세한 내용은 [Amazon Redshift 무료 평가판](#).

Prerequisites

Amazon Redshift 클러스터 설정을 시작하기 전에 다음 사전 조건을 갖추었는지 확인합니다.

- [AWS에 가입 \(p. 1\)](#)
- [방화벽 규칙 확인 \(p. 2\)](#)

AWS에 가입

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

1. <https://portal.aws.amazon.com/billing/signup>을 엽니다.
2. 온라인 지시 사항을 따릅니다.

등록 절차 중 전화를 받고 전화 키패드를 사용하여 확인 코드를 입력하는 과정이 있습니다.

방화벽 규칙 확인

이 자습서의 일부로 Amazon Redshift 클러스터를 시작할 때 포트를 지정합니다. 또한 보안 그룹에 인바운드 수신 규칙을 생성하여 포트를 통해 클러스터에 대한 액세스를 허용합니다.

클라이언트 컴퓨터가 방화벽을 사용하는 경우 사용할 수 있는 개방된 포트를 알고 있는지 확인합니다. 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 Redshift 는 기본적으로 포트 5439를 사용하지만, 포트가 방화벽에서 차단되면 연결이 해제됩니다. Amazon Redshift 클러스터의 포트 번호는 생성한 후에는 변경할 수 없습니다. 따라서 시작 프로세스 중에 환경에서 작업하는 개방된 포트를 지정해야 합니다.

이 전제 조건은 Amazon Redshift 로 자신의 데이터를 가져올 때만 적용됩니다. 자세한 내용은 [Amazon Redshift 에 자체 데이터 가져오기 \(p. 6\)](#) 단원을 참조하세요.

Amazon Redshift 클러스터 및 데이터 로드 시작하기

이 섹션에서는 사용자가 샘플 Amazon Redshift 클러스터를 만드는 절차를 안내하는 두 가지 자습서를 찾을 수 있습니다. 하나는 샘플 데이터 집합을 사용하고 다른 하나는 자체 데이터 집합을 가져옵니다.

에 대한 사전 조건을 갖추었는지 확인Amazon Redshift 시작 안내서시작하기 전에 단원을 선택합니다. 자세한 내용은 [Prerequisites \(p. 1\)](#) 단원을 참조하십시오.

주제

- [샘플 데이터 집합 사용 \(p. 3\)](#)
- [Amazon Redshift 에 자체 데이터 가져오기 \(p. 6\)](#)

샘플 데이터 집합 사용

이 자습서에서는 샘플 데이터 집합을 사용하여 Amazon Redshift 클러스터를 생성하는 프로세스를 안내합니다. Amazon Redshift 는 새 클러스터를 생성할 때 샘플 데이터 집합을 자동으로 로드하고 새 클러스터를 쿼리 편집기에 연결합니다. 클러스터가 생성되면 즉시 데이터를 쿼리할 수 있습니다.

Amazon Redshift 클러스터 설정을 시작하기 전에,[AWS에 가입 \(p. 1\)](#) 및 [방화벽 규칙 확인 \(p. 2\)](#).

이 자습서에서는 다음 다이어그램에 표시된 단계를 수행합니다.



주제

- [1단계: Amazon Redshift 클러스터 생성 \(p. 3\)](#)
- [2단계: 쿼리 편집기를 사용하여 예제 쿼리를 사용해보십시오. \(p. 5\)](#)

Important

사용자가 생성하는 샘플 클러스터는 실제 환경에서 실행됩니다. 이 자습서에서 설계된 샘플 클러스터 사용 시 주문형 요금은 시간당 \$0.25이며, 샘플 클러스터를 삭제하기 전까지 적용됩니다. 요금에 대한 자세한 내용은 단원을 참조하십시오.[Amazon Redshift 요금](#). 질문이 있거나 문제가 있는 경우 Amazon Redshift 팀에 [토론 포럼](#).

이 자습서는 프로덕션 환경용이 아니므로, 세부 옵션에 대해서는 자세히 설명하지 않습니다. 이 자습서의 단계를 완료한 후, [추가 리소스 \(p. 29\)](#)를 참조하여 자세한 정보를 찾을 수 있습니다. 이 정보는 클러스터를 계획, 배포 및 유지 관리하고, 데이터 웨어하우스에서 데이터를 작업하는 데 도움이 됩니다.

1단계: Amazon Redshift 클러스터 생성

사전 조건을 갖추었으면 샘플 데이터 세트를 기반으로 Amazon Redshift 클러스터를 만들 수 있습니다.

샘플 데이터 세트를 기반으로 Amazon Redshift 클러스터를 생성하려면

1. 에 로그인합니다.AWS Management Console에서 Amazon Redshift 콘솔을 엽니다.<https://console.aws.amazon.com/redshift/>.
2. 클러스터를 생성하려면 다음 중 하나를 수행합니다.
 - Amazon Redshift 서비스 페이지에서클러스터 생성. 클러스터 생성 페이지가 표시됩니다.

- 에서 <https://console.aws.amazon.com/redshift/>를 선택하고 대시보드를 선택한 다음 클러스터 생성.
 - 에서 <https://console.aws.amazon.com/redshift/>를 선택하고 클러스터를 선택한 다음 클러스터 생성.
3. 에서 클러스터 구성 섹션에서 클러스터 식별자. 이 식별자는 고유해야 합니다. 식별자는 유효한 문자 a—z (소문자만 해당) 및 - (하이픈)를 사용하여 1-63자여야 합니다.

Enter **examplecluster**이 자습서에 대한 것입니다.

4. 조직에 적합한 경우 Amazon Redshift 무료 평가판 프로그램에 따라 클러스터를 생성할 수 있습니다. 이렇게 하려면 무료 평가판을 클릭하여 dc2.large 노드 유형으로 구성을 만듭니다. 무료 평가판 선택에 대한 자세한 내용은 단원을 참조하십시오. [Amazon Redshift 무료 평가판](#).

다음 예와 같이 선택 내용이 콘솔에 표시됩니다.

나중에 다른 노드 유형을 선택하면 조직에서 더 이상 무료 평가판을 이용할 수 없습니다.

노드 유형을 선택한 후 다음 중 하나를 수행합니다.

- In 샘플 데이터를 선택하고 샘플 데이터 로드를 클릭하여 샘플 데이터 세트를 Amazon Redshift 클러스터로 로드합니다. Amazon Redshift는 샘플 데이터 세트 Tickit을 기본 dev 데이터베이스 및 public스키마입니다. 쿼리 편집기를 사용하여 데이터를 쿼리할 수 있습니다.
- Amazon Redshift 클러스터로 자체 데이터를 가져오려면 프로덕션. 그런 다음 샘플 데이터를 선택하고 샘플 데이터 로드. 사용자 고유의 데이터 가져오기에 대한 자세한 내용은 [Amazon Redshift에 자체 데이터 가져오기 \(p. 6\)](#).

Amazon Redshift는 샘플 데이터 세트를 샘플 Amazon Redshift 클러스터에 자동으로 로드합니다.

5. 에서 데이터베이스 구성 섹션에서 값을 지정합니다. 관리자 이름 및 관리자 암호. 또는 비밀번호 생성을 사용하여 Amazon Redshift에서 생성한 암호를 사용할 수 있습니다.

본 자습서에서는 다음 값을 사용합니다.

- 관리자 이름: **awsuser**를 입력합니다.
 - 관리자 암호: 암호 값을 입력합니다.
6. 클러스터 생성을 선택합니다.

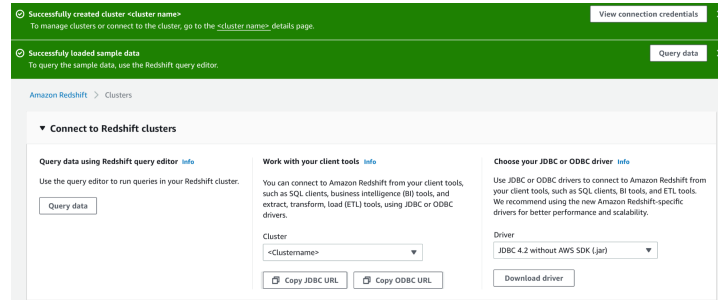
이 자습서에서는 Amazon Redshift 쿼리 편집기를 사용합니다. Amazon Redshift가 클러스터를 생성하는 동안 새 클러스터를 쿼리 편집기에 연결하는 즉시 데이터를 쿼리하는 데 사용할 수 있습니다.

또한 클러스터의 데이터로 작업하기 위해 JDBC 또는 ODBC 드라이버를 지원하는 다른 SQL 클라이언트 도구를 선택할 수도 있습니다. 자세한 내용은 단원을 참조하십시오. [SQL 클라이언트 도구를 사용하여 Amazon Redshift 클러스터에 연결](#)의 Amazon Redshift 클러스터 관리 가이드.

2단계: 쿼리 편집기를 사용하여 예제 쿼리를 사용해보십시오.

Amazon Redshift가 Amazon Redshift 클러스터를 생성할 때 샘플 데이터 세트 Tickit을 자동으로 업로드합니다. 클러스터 생성을 완료하는 데 몇 분 정도 걸릴 수 있습니다. 생성이 완료되면 클러스터 상태가 ACTIVE가 됩니다. 당신은 선택하여 쿼리 편집기에서 샘플 Tickit 테이블을 볼 수 있습니다dev데이터베이스 및public스키마입니다.

Amazon Redshift 클러스터가 생성되면Amazon Redshift 클러스터에 Connect를 선택하고데이터 쿼리.



Amazon Redshift 쿼리 편집기가 나타납니다. Amazon Redshift 는 샘플 데이터 집합과 함께 로드되는 새 클러스터와의 연결을 설정합니다.

기본적으로 Amazon Redshift 는 클러스터 생성 중에 생성된 데이터베이스를데이터베이스 선택라는 기본 스키마라는public. 샘플 데이터 집합의 개별 데이터 파일을 보려면 쿼리 편집기로 이동하여dev데이터베이스 및public스키마입니다.

다음과 같이 쿼리 편집기에서 몇 가지 예제 쿼리를 시도합니다. SELECT 명령 작업에 대한 자세한 내용은 단원을 참조하십시오.선택Amazon Redshift 데이터베이스 개발자 안내서에 나와 있습니다.

```
-- 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;
```


이 자습서의 단계를 완료한 후, [추가 리소스 \(p. 29\)](#)를 참조하여 자세한 정보를 찾을 수 있습니다. 이 정보는 클러스터를 계획, 배포 및 유지 관리하고, 데이터 웨어하우스에서 데이터를 작업하는 데 도움이 됩니다.

또한 다음을 수행할 수도 있습니다. [Amazon Redshift 에 자체 데이터 가져오기 \(p. 6\)](#) 자습서를 사용하여 자체 데이터 집합으로 클러스터를 만들 수 있습니다.

Amazon Redshift 에 자체 데이터 가져오기

이 자습서에서는 자체 데이터 집합을 Amazon Redshift로 가져와 Amazon Redshift 클러스터를 생성하는 프로세스를 안내합니다. 이 샘플 클러스터를 사용하여 Amazon Redshift 서비스를 평가할 수 있습니다.

Amazon Redshift 클러스터 설정을 시작하기 전에, [AWS에 가입 \(p. 1\)](#) 및 [방화벽 규칙 확인 \(p. 2\)](#).

이 자습서에서는 다음 다이어그램에 표시된 단계를 수행합니다.



주제

- 1단계: IAM 역할 생성 (p. 6)
- 2단계: 샘플 Amazon Redshift 클러스터 생성 (p. 7)
- 3단계: SQL 클라이언트에 대한 인바운드 규칙 구성 (p. 8)
- 4단계: 쿼리 편집기에 대한 액세스 권한 부여 (p. 9)
- 5단계: Amazon S3 에서 샘플 데이터 로드 (p. 11)
- 6단계: 쿼리 편집기를 사용하여 예제 쿼리를 사용해보십시오. (p. 14)
- 단계 7: 환경 재설정 (p. 15)

Important

사용자가 생성하는 샘플 클러스터는 실제 환경에서 실행됩니다. 이 자습서에서 설계된 샘플 클러스터 사용 시 주문형 요금은 시간당 \$0.25이며, 샘플 클러스터를 삭제하기 전까지 적용됩니다. 요금에 대한 자세한 내용은 단원을 참조하십시오. [Amazon Redshift 요금 페이지](#). 질문이 있거나 문제가 있는 경우 Amazon Redshift 팀에 [토론 포럼](#).

이 자습서는 프로덕션 환경용이 아니므로, 세부 옵션에 대해서는 자세히 설명하지 않습니다. 이 자습서의 단계를 완료한 후, [추가 리소스 \(p. 29\)](#)를 참조하여 자세한 정보를 찾을 수 있습니다. 이 정보는 클러스터를 계획, 배포 및 유지 관리하고, 데이터 웨어하우스에서 데이터를 작업하는 데 도움이 됩니다.

1단계: IAM 역할 생성

다른 데이터의 데이터에 액세스하는 모든 작업의 경우 AWS 리소스를 사용하려면 클러스터에 사용자를 대신해 리소스와 리소스의 데이터에 액세스할 권한이 필요합니다. 예를 들어 COPY 명령을 사용하여 Amazon Simple Storage Service (Amazon S3) 에서 데이터를 로드합니다. AWS Identity and Access Management(IAM)을 사용하여 그러한 권한을 제공합니다. 이 작업은 클러스터에 연결된 IAM 역할을 통해 수행할 수 있습니다. 또는 제공할 수 있습니다 AWS 필요한 권한이 있는 IAM 사용자에게 액세스 키입니다. 자격 증명 및 액세스 권한에 대한 자세한 내용은 [자격 증명 및 액세스 권한](#)을 참조하십시오.

중요한 데이터를 가장 잘 보호하고 AWS 액세스 자격 증명을 사용하려면 IAM 역할을 만들어 클러스터에 연결하는 것이 가장 좋습니다. 액세스 권한을 제공하는 방법에 대한 자세한 내용은 단원을 참조하십시오. [다른 액세스 권한 AWS 리소스](#).

이 단계에서는 Amazon Redshift S3 버킷에서 데이터를 로드하도록 해주는 IAM 역할을 새로 생성합니다. IAM 역할은 계정에 만들 수 있는, 특정 권한을 지닌 IAM 자격 증명입니다. 다음 단계에서는 해당 역할을 클러스터에 연결합니다.

Amazon Redshift 용 IAM 역할을 생성하려면 다음과 같이 하십시오.

1. AWS Management Console에 로그인하여 <https://console.aws.amazon.com/iam/>에서 IAM 콘솔을 엽니다.
2. 탐색 창에서 역할을 선택합니다.
3. 역할 생성을 선택합니다.
4. 에서AWS서비스그룹에서Redshift.
5. 언더사용 사례를 선택합니다.를 선택하고Redshift - 사용자 지정을 선택한 다음다음: Permissions]]을 선택합니다.
6. Attach permissions policies(권한 정책 연결) 페이지에서 AmazonS3ReadOnlyAccess를 선택합니다. Set permissions boundary(권한 경계 설정)의 기본 설정을 그대로 사용할 수 있습니다. 다음을 선택합니다. Tags]]를 선택합니다.
7. 태그 추가 페이지가 나타납니다. 선택적으로 태그를 추가할 수 있습니다. [다음: 권한(Next: Review)]를 선택합니다.
8. 역할 이름에 역할의 이름을 입력합니다. 이 자습서에서는 **myRedshiftRole**을 입력합니다.
9. 정보를 검토한 후 역할 만들기를 선택합니다.
10. 방금 생성한 역할의 역할 이름을 선택합니다.
11. COPY역할 ARN값을 클립보드에 추가합니다. 이 값은 방금 생성한 역할의 Amazon 리소스 이름 (ARN)입니다. 이 값은 5단계: [Amazon S3에서 샘플 데이터 로드 \(p. 11\)](#)에서 COPY 명령을 사용하여 데이터를 로드할 때 사용합니다.

이제 새로운 역할을 생성했으므로 다음 단계는 이를 클러스터와 연결하는 것입니다. 새 클러스터를 시작할 때 역할을 연결하거나 기존 클러스터에 역할을 연결할 수 있습니다. 다음 단계에서는 역할을 새 클러스터에 연결합니다.

2단계: 샘플 Amazon Redshift 클러스터 생성

사전 조건을 갖추었으므로 이제 Amazon Redshift 클러스터를 시작할 수 있습니다.

생성하려는 클러스터가 활성화됩니다(샌드박스에서 실행되지 않음). 삭제하기 전까지 클러스터에 대해 표준 Amazon Redshift 사용 요금이 청구됩니다. 여기에 설명되어 있는 자습서를 한 번에 끝내고, 실습을 마칠 때 클러스터를 삭제하면 총 청구 비용이 가장 적게 듭니다.

Amazon Redshift 클러스터를 생성하려면 다음과 같이 하십시오.

1. 에 로그인합니다.AWS Management Console에서 Amazon Redshift 콘솔을 엽니다.<https://console.aws.amazon.com/redshift/>.
Important
IAM 사용자 자격 증명을 사용하는 경우 사용자에게 클러스터 작업을 수행하는 데 필요한 권한이 있는지 확인해야 합니다. 자세한 내용은 단원을 참조하십시오.[IAM 사용자에 대한 액세스 제어](#)의Amazon Redshift 클러스터 관리 가이드.
2. 오른쪽 상단에서 을 선택합니다.AWS클러스터를 생성하려는 리전입니다.
3. 탐색 메뉴에서 클러스터를 선택한 다음 클러스터 생성을 선택합니다. 클러스터 생성 페이지가 표시됩니다.
4. 에서클러스터 구성섹션에서 값을 지정합니다.클러스터 식별자,노드 유형, 및노드:
 - 클러스터 식별자: Enter**examplecluster**이 자습서에 대한 것입니다. 이 식별자는 고유해야 합니다. 식별자는 유효한 문자 a—z (소문자만 해당) 및 - (하이픈) 를 사용하여 1-63자여야 합니다.

- 다음 방법 중 하나를 선택하여 클러스터 크기를 조정합니다.

Note

다음 단계에서는 Amazon Redshift 콘솔에 대해 설명합니다. AWSRA3 노드 유형을 지원하는 리전입니다. 의 목록을 보려면 AWSRA3 노드 유형을 지원하는 리전은 단원을 참조하십시오. [RA3 노드 유형 개요](#)의 Amazon Redshift 클러스터 관리 가이드.

- 만약 당신의 AWSRA3 노드 유형을 지원합니다. 프로덕션 또는 무료 평가판 질문에 답하기 위해 클러스터를 어떤 용도로 사용할 계획입니까?

선택할 때 프로덕션에서 다음 중 하나를 수행합니다.

조직에 적합한 경우 Amazon Redshift 무료 평가판 프로그램에 따라 클러스터를 생성할 수 있습니다. 무료 평가판 프로그램을 사용하여 클러스터를 만드는 방법에 대한 자세한 내용은 [샘플 데이터 집합 사용 \(p. 3\)](#). 이렇게 하려면 무료 평가판을 클릭하여 dc2.large 노드 유형으로 구성을 만듭니다. 무료 평가판 선택에 대한 자세한 내용은 단원을 참조하십시오. [Amazon Redshift 무료 평가판](#).

- 클러스터 크기를 잘 모르면 선택할 수 있도록 합니다.. 이렇게 하면 데이터 웨어하우스에 저장할 데이터의 크기 및 쿼리 특성에 대해 질문하는 크기 조정 계산기가 시작됩니다.

클러스터의 필요한 크기 (즉, 노드 유형 및 노드 수) 를 알고 있는 경우 내가 선택할게. 그런 다음 노드 유형의 수노드를 사용하여 개념 증명에 맞게 클러스터의 크기를 조정할 수 있습니다.

- 선택 노드 유형: **dc2.large** 다음으로 바꿉니다. 노드: **2** 이 자습서에 대한 것입니다.
 - 당신이 선택한 경우 프로덕션을 선택한 다음, 다음 중 하나를 수행합니다.
 - Amazon Redshift 가 제공하는 샘플 데이터 집합을 사용하려면 샘플 데이터를 선택하고 샘플 데이터 로드. Amazon Redshift 는 샘플 데이터 세트 Tickit 을 기본 dev 데이터베이스 및 public 스키마입니다.
 - Amazon Redshift 로 데이터를 가져오려면 자습서의 나머지 부분을 계속 진행합니다.
5. 에서 데이터베이스 구성 섹션에서 값을 지정합니다. 데이터베이스 이름 (선택 사항), 데이터베이스 포트 (선택), 관리자 이름, 및 관리자 암호. 또는 비밀번호 생성을 사용하여 Amazon Redshift 에서 생성한 암호를 사용할 수 있습니다.

본 자습서에서는 다음 값을 사용합니다.

- 데이터베이스 이름 (선택 사항): **dev** 를 입력합니다.
 - 데이터베이스 포트 (선택 사항): **5439** 를 입력합니다.
 - 관리자 이름: **awsuser** 를 입력합니다.
 - 관리자 암호: 암호 값을 입력합니다.
6. (선택) 클러스터 사용 권한 섹션에서 다음을 수행합니다. 사용 가능한 IAM 역할 이전에 생성한 IAM 역할을 선택하고 **myRedshiftRole**. 그런 다음 IAM 역할 연결.
- 를 사용하여 클러스터를 생성한 경우 샘플 데이터 로드 옵션을 사용하여 IAM 역할을 클러스터에 연결합니다.
7. (선택) 추가 구성 섹션에서 기본값 사용을 선택하여 수정합니다. 네트워크 및 보안, 데이터베이스 구성, 유지 관리, 모니터링, 및 백업 설정.
8. 클러스터 생성을 선택합니다.

3단계: SQL 클라이언트에 대한 인바운드 규칙 구성

이 자습서의 뒷부분에서 Amazon VPC 서비스를 기반으로 하는 가상 사설 클라우드 (VPC) 내에서 클러스터에 액세스합니다. 그러나 방화벽 외부에서 SQL 클라이언트를 사용하여 클러스터에 액세스하는 경우 인바운드 액세스 권한을 부여해야 합니다.

VPC 내에서 Amazon Redshift 쿼리 편집기를 사용하여 클러스터에 액세스하려는 경우 이 단계를 건너뛸 수 있습니다.

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) 인스턴스 또는 외부 컴퓨터일 수 있습니다.
2. Amazon EC2 외부 클라이언트에서 액세스하려면 클러스터에 연결된 보안 그룹에 인바운드 트래픽을 허용하는 수신 규칙을 추가하십시오. Amazon EC2 보안 그룹 규칙은 Amazon EC2 콘솔에서 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 찾으십시오.

4단계: 쿼리 편집기에 대한 액세스 권한 부여

Amazon Redshift 클러스터에서 호스팅하는 데이터베이스를 쿼리하려면 다음 두 가지 옵션이 있습니다.

- 클러스터에 연결한 다음 쿼리 편집기를 사용해 AWS Management Console에 대한 쿼리를 실행합니다.

쿼리 편집기를 사용하는 경우 SQL 클라이언트 애플리케이션을 다운로드해 설치할 필요가 없습니다.

- SQL 클라이언트 도구 (예: SQL Workbench/J) 를 통해 클러스터에 Connect.[SQL Workbench/J를 사용하여 클러스터에 연결하기](#)의Amazon Redshift 클러스터 관리 가이드.

Amazon Redshift 쿼리 편집기를 사용하면 Amazon Redshift 클러스터에서 호스팅하는 데이터베이스에 대한 쿼리를 가장 손쉽게 실행할 수 있습니다. 클러스터를 생성한 후 Amazon Redshift 콘솔을 사용하여 즉시 쿼리를 실행할 수 있습니다. Amazon Redshift 쿼리 편집기를 사용할 때 고려해야 할 사항에 대한 자세한 내용은[쿼리 편집기를 사용하여 데이터베이스 쿼리](#)의Amazon Redshift 클러스터 관리 가이드.

주제

- [쿼리 편집기에 대한 액세스 권한 부여](#) (p. 9)
- [쿼리 편집기 사용](#) (p. 10)

쿼리 편집기에 대한 액세스 권한 부여

Amazon Redshift 쿼리 편집기를 사용하려면 권한이 있어야 합니다. 액세스를 설정하려면AmazonRedshiftQueryEditor및AmazonRedshiftReadOnlyAccessIAM 정책을 클러스터에 액세스하는 데 사용하는 IAM 사용자에게 적용합니다.

Amazon Redshift 에 액세스할 IAM 사용자를 이미 생성한 경우AmazonRedshiftQueryEditor및AmazonRedshiftReadOnlyAccess정책을 해당 사용자에게 적용해야 합니다. 아직 IAM 사용자를 생성하지 않은 경우 사용자를 하나 생성하고 이 IAM 사용자에게 정책을 연결합니다.

쿼리 편집기 사용에 필요한 IAM 정책을 연결하려면

1. AWS Management Console에 로그인하여 <https://console.aws.amazon.com/iam/>에서 IAM 콘솔을 엽니다.
2. 사용자를 선택합니다.
3. 쿼리 편집기에 대한 액세스 권한이 필요한 사용자를 선택합니다.
4. [Add permissions]를 선택합니다.
5. 기존 정책 직접 연결(Attach existing policies directly)을 선택합니다.
6. Policy names(정책 이름)로 AmazonRedshiftQueryEditor 및 AmazonRedshiftReadOnlyAccess를 선택합니다.
7. [다음: 권한(Next: Review)]를 선택합니다.
8. [Add permissions]를 선택합니다.

쿼리 편집기 사용

다음 예에서는 쿼리 편집기를 사용하여 아래 작업을 수행합니다.

- SQL 명령을 실행합니다.
- 쿼리 실행 방법에 대한 세부 정보를 봅니다.
- 쿼리를 저장합니다.
- 쿼리 결과 집합을 다운로드합니다.

쿼리 편집기를 사용하려면

1. 에 로그인합니다.AWS Management Console에서 Amazon Redshift 콘솔을 엽니다.<https://console.aws.amazon.com/redshift/>.
2. 탐색 메뉴에서 편집기를 선택한 다음 클러스터의 데이터베이스에 연결합니다.

에서데이터베이스에 연결페이지에서 인증하는 방법에는 두 가지가 있습니다.2013년 5월 22일및AWS 보안 관리자. 본 자습서에서는 을 선택합니다.새 연결 생성및2013년 5월 22일을 선택한 후 다음과 같이 클러스터를 생성할 때 사용한 값을 입력합니다.

- 클러스터: **examplecluster**를 선택합니다.
- 데이터베이스 이름: **dev**를 입력합니다.
- 데이터베이스 사용자: **awsuser**를 입력합니다.

그런 다음연결.

3. 스키마로 퍼블릭을 선택하여 이 스키마를 바탕으로 새 테이블을 생성합니다.
4. 쿼리 편집기 창에서 다음을 입력한 다음 실행을 사용하여 새 테이블을 만듭니다.

```
create table shoes(  
    shoetype varchar (10),  
    color varchar(10));
```

5. 지우기를 선택합니다.
6. 쿼리 편집기 창에서 다음 명령을 입력한 다음 실행을 선택하여 테이블에 행을 추가합니다.

```
insert into shoes values  
( 'loafers', 'brown'),  
( 'sandals', 'black');
```

7. 지우기를 선택합니다.
8. 쿼리 편집기 창에서 다음 명령을 입력한 다음 실행을 선택하여 새 테이블을 쿼리합니다.

```
select * from shoes;
```

쿼리 결과에 결과가 나타납니다.

신발 유형	색상
sandals	black
loafers	brown

9. 실행 세부 정보를 확인하려면 실행을 선택합니다.

10. 쿼리 결과를 파일로 다운로드하려면 내보내기를 선택합니다. 지원되는 파일 형식은 CSV, TXT 및 HTML입니다.

5단계: Amazon S3 에서 샘플 데이터 로드

이 단계에서 우리는 dev라는 데이터베이스를 가지고 있으며 이 데이터베이스에 연결되어 있습니다. 다음으로 데이터베이스에 테이블 몇 개를 생성하고, 해당 테이블에 데이터를 업로드하고, 쿼리를 실행해 봅니다. 편의를 위해, Amazon S3 버킷에서 샘플 데이터를 로드합니다.

Note

SQL 클라이언트 도구를 사용하는 경우 SQL 클라이언트가 클러스터에 연결되어 있는지 확인합니다.

이 단계를 완료한 후 다음을 수행할 수 있습니다.

- 에서 예제 쿼리를 사용해보십시오.[6단계: 쿼리 편집기를 사용하여 예제 쿼리를 사용해보십시오. \(p. 14\).](#)
- 환경 재설정[단계 7: 환경 재설정 \(p. 15\).](#)
- Amazon Redshift 에 대해 자세히 알아보십시오.[추가 리소스 \(p. 29\).](#)

Note

자체 데이터를 로드하지 않고 쿼리 편집기에서 데이터를 쿼리하려면 샘플 데이터 로드 in 샘플 데이터. 이렇게 하면 Amazon Redshift 가 클러스터를 생성하는 동안 자동으로 Amazon Redshift 클러스터로 샘플 데이터 세트를 로드합니다.

Amazon S3 에서 샘플 데이터를 로드하려면

1. 테이블을 만듭니다.

Amazon Redshift 쿼리 편집기를 사용하는 경우, 아래의 테이블 만들기 문을 개별적으로 복사하고 실행하여 dev의 데이터베이스. 구문에 대한 자세한 내용은 단원을 참조하십시오. [테이블 생성](#)의 Amazon Redshift 데이터베이스 개발자 안내서.

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

```
create table venue(  

```

```
venueid smallint not null distkey sortkey,  
venue name varchar(100),  
venue city varchar(30),  
venue state char(2),  
venue seats integer);
```

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

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

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

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

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

2. COPY 명령을 사용하여 Amazon S3 에서 샘플 데이터를 로드합니다.

Note

대용량 데이터 세트는 COPY 명령을 사용하여 Amazon S3 또는 Amazon Amazon DynamoDB 에서 Amazon Redshift로 로드하는 것이 좋습니다. COPY 구문에 대한 자세한 내용은 단원을 참조하십시오. [COPY](#)의 Amazon Redshift 데이터베이스 개발자 안내서.

- 파일 다운로드 [tickitdb.zip](#) 개별 샘플 데이터 파일을 포함합니다.
- 압축을 풀고 개별 파일을 로드합니다. [tickit](#) 폴더에 있는 Amazon S3 버킷에 있는 AWS 리전.
- 이 자습서의 COPY 명령을 편집하여 Amazon S3 버킷의 파일을 가리킵니다. Amazon S3 에서 파일을 관리하는 방법에 대한 자세한 내용은 단원을 참조하십시오. [S3 버킷 생성 및 구성](#)의 Amazon Simple Storage Service 콘솔 사용자 안내서.
- 클러스터에 인증을 제공하여 샘플 데이터를 로드합니다. 역할 기반 인증 또는 키 기반 인증 중 하나를 제공할 수 있습니다. 이 중 역할 기반 인증을 사용하는 것이 좋습니다. 다음을 수행할 수 있습니다. [myRedshiftRole](#)에서 생성한 IAM 역할 [1단계: IAM 역할 생성 \(p. 6\)](#)를 사용하여 Amazon Redshift 가 Amazon S3 버킷에서 데이터를 로드하도록 설정할 수 있습니다. 두 인증 유형에 대한 자세한 내용은 단원을 참조하십시오. [자격 증명](#)의 Amazon Redshift 데이터베이스 개발자 안내서.

이 단계를 수행하려면 이전 단계에서 사용자가 생성하여 클러스터에 연결한 IAM 역할을 참조함으로써 인증을 제공합니다.

Note

Amazon S3 액세스할 수 있는 적절한 권한이 없는 경우 COPY 명령을 실행하면 오류 메시지가 나타납니다. `S3ServiceException: Access Denied`. COPY 명령에 대한 IAM 권한에 대한 자세한 내용은 단원을 참조하십시오. [COPY](#)의 Amazon Redshift 데이터베이스 개발자 안내서.

COPY 명령에는 다음 예와 같이 IAM 역할, 사용자의 버킷 이름, AWS 지역에 대한 Amazon 리소스 이름(ARN)용 자리 표시자가 들어 있습니다.

```
copy users from 's3://<myBucket>/tickit/allusers_pipe.txt'
credentials 'aws_iam_role=<iam-role-arn>'
delimiter '|' region '<aws-region>;'
```

IAM 역할을 사용하여 액세스를 허가하려면, CREDENTIALS 파라미터 문자열의 `<iam-role-arn>`을 [1단계: IAM 역할 생성 \(p. 6\)](#) 단원에서 만든 IAM 역할의 역할 ARN으로 바꾸십시오.

COPY 명령은 다음 예와 유사합니다.

```
copy users from 's3://<myBucket>/tickit/allusers_pipe.txt'
credentials 'aws_iam_role=arn:aws:iam::123456789012:role/myRedshiftRole'
delimiter '|' region '<aws-region>;'
```

샘플 데이터를 로드하려면 `<myBucket>`, `<iam-role-arn>`, 및 `<aws-region>`값을 사용하여 다음 COPY 명령을 실행하십시오. Amazon Redshift 쿼리 편집기를 사용하는 경우 다음 명령을 개별적으로 실행합니다.

```
copy users from 's3://<myBucket>/tickit/allusers_pipe.txt'
credentials 'aws_iam_role=<iam-role-arn>'
delimiter '|' region '<aws-region>;'
```

```
copy venue from 's3://<myBucket>/tickit/venue_pipe.txt'
credentials 'aws_iam_role=<iam-role-arn>'
delimiter '|' region '<aws-region>;'
```

```
copy category from 's3://<myBucket>/tickit/category_pipe.txt'
credentials 'aws_iam_role=<iam-role-arn>'
delimiter '|' region '<aws-region>;'
```

```
copy date from 's3://<myBucket>/tickit/date2008_pipe.txt'
```



```
credentials 'aws_iam_role=<iam-role-arn>'
delimiter '|' region '<aws-region>;
```

```
copy event from 's3://<myBucket>/tickit/allevnts_pipe.txt'
credentials 'aws_iam_role=<iam-role-arn>'
delimiter '|' timeformat 'YYYY-MM-DD HH:MI:SS' region '<aws-region>;
```

```
copy listing from 's3://<myBucket>/tickit/listings_pipe.txt'
credentials 'aws_iam_role=<iam-role-arn>'
delimiter '|' region '<aws-region>;
```

```
copy sales from 's3://<myBucket>/tickit/sales_tab.txt'
credentials 'aws_iam_role=<iam-role-arn>'
delimiter '\t' timeformat 'MM/DD/YYYY HH:MI:SS' region '<aws-region>;
```

6단계: 쿼리 편집기를 사용하여 예제 쿼리를 사용해보십시오.

다음과 같이 이제 몇 가지 예제 쿼리를 시도합니다. SELECT 명령 작업에 대한 자세한 내용은 단원을 참조하십시오. [선택](#)의 Amazon Redshift 개발자 안내서.

```
-- 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;
```

Amazon Redshift 클러스터를 생성하고 Amazon Redshift 쿼리 편집기를 사용하여 자체 데이터 집합에서 데이터를 쿼리했습니다.

단계 7: 환경 재설정

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

그러나 다른 Amazon Redshift 안내서의 작업을 시도해 보려는 경우 샘플 클러스터를 계속 실행하고자 할 수 있습니다.

클러스터를 삭제하려면

1. 에 로그인합니다. AWS Management Console에서 Amazon Redshift 콘솔을 엽니다. <https://console.aws.amazon.com/redshift/>.
2. 탐색 메뉴에서 클러스터를 선택하여 클러스터 목록을 표시합니다.
3. **examplecluster** 클러스터를 선택합니다. [Actions]에 대해 [Delete]를 선택합니다. 클러스터 삭제 페이지가 표시됩니다.
4. 삭제할 클러스터를 확인한 다음 클러스터 삭제를 선택합니다.

클러스터가 삭제되면 클러스터 목록 페이지에서 클러스터 상태가 업데이트됩니다.

이 자습서를 완료한 후 Amazon Redshift 및 다음 단계에 대한 자세한 내용은 [추가 리소스 \(p. 29\)](#).

일반적인 데이터베이스 작업 시작하기

다음에는 Amazon Redshift 데이터베이스를 사용할 수 있도록 일반적인 작업에 대한 설명과 연습을 찾을 수 있습니다.

초기 클러스터에 연결한 후 `dev` 데이터베이스에서 새 데이터베이스를 생성할 수 있습니다. 클러스터를 생성하는 동안 샘플 데이터 세트를 사용할지 아니면 자체 데이터를 Amazon Redshift 로 가져오든 관계없이 Amazon Redshift 는 `dev` 데이터베이스.

이 단원의 예에서는 다음을 가정합니다.

- Amazon Redshift 데이터 웨어하우스 서비스에 가입되어 있어야 합니다. 자세한 내용은 [Prerequisites \(p. 1\)](#) 단원을 참조하십시오.
- Amazon Redshift 클러스터를 생성했습니다. 자세한 내용은 [Amazon Redshift 클러스터 및 데이터 로드 시작하기 \(p. 3\)](#) 단원을 참조하십시오.
- Amazon Redshift 콘솔 쿼리 편집기와 같은 SQL 클라이언트 도구에서 클러스터에 연결을 설정해 두어야 합니다. 자세한 내용은 [4단계: 쿼리 편집기에 대한 액세스 권한 부여 \(p. 9\)](#) 단원을 참조하십시오.

Important

이 연습에서 배포한 클러스터는 실제 환경에서 실행됩니다. 실행 중인 한, 그것은 당신의 AWS 계정. 요금 정보는 [Amazon Redshift 요금 페이지](#). 불필요한 과금을 피하려면 클러스터를 사용한 후 삭제합니다. 이 연습의 마지막 단계에서 그 방법을 설명합니다.

작업 1: 데이터베이스 생성

클러스터가 실행 중인 것을 확인한 후 고유한 첫 번째 데이터베이스를 만들 수 있습니다. 이 데이터베이스에서 실제로 테이블을 만들어 데이터를 업로드하고 쿼리를 실행할 수 있습니다. 하나의 클러스터가 여러 개의 데이터베이스를 호스팅할 수 있습니다. 예를 들어, `SALESDB` 데이터베이스 및 `ORDERSDb` 데이터베이스를 동일한 클러스터에서 사용할 수 있습니다.

예를 들어 이라는 데이터베이스를 만들려면 `SALESDB`에서 SQL 클라이언트 도구에서 다음 명령을 실행합니다.

```
CREATE DATABASE SALESDB;
```

이 연습에서는 기본값을 적용합니다. 더 많은 명령 옵션에 대해서는 [데이터베이스 생성](#)의 Amazon Redshift 데이터베이스 개발자 안내서.

`SALESDB` 데이터베이스를 만든 후 SQL 클라이언트에서 새 데이터베이스에 연결할 수 있습니다. 현재 연결에 사용한 것과 같은 연결 파라미터를 사용하되 데이터베이스 이름을 `SALESDB`으로 변경합니다.

작업 2: 사용자를 생성합니다

기본적으로 클러스터를 시작할 때 만든 관리자 사용자만이 클러스터의 초기 데이터베이스에 액세스할 수 있습니다. 다른 사용자에게 액세스 권한을 부여하려면 하나 이상의 계정을 생성합니다. 데이터베이스 사용자 계정은 개별 데이터베이스가 아닌 클러스터의 모든 데이터베이스에서 전역적입니다.

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

예를 들어 이라는 사용자를 만들려면 **GUEST** 암호를 사용하여 **ABcd4321**에서 다음 명령을 실행합니다.

```
CREATE USER GUEST PASSWORD 'ABcd4321';
```

에 연결하려면 **SALESDB** 데이터베이스를 **GUEST** 사용자의 경우 사용자를 만들 때 동일한 암호 (예: **ABcd4321**).

다른 명령 옵션에 대한 자세한 내용은 [사용자 생성](#)의 Amazon Redshift 데이터베이스 개발자 안내서.

작업 3: 스키마 생성

새 데이터베이스를 만든 후 현재 데이터베이스에서 새 스키마를 만들 수 있습니다. A 스키마 테이블, 뷰, UDF (사용자 정의 함수)와 같은 명명된 데이터베이스 개체를 포함하는 네임스페이스입니다. 데이터베이스는 하나 이상의 스키마를 포함할 수 있으며 각 스키마는 하나의 데이터베이스에만 속합니다. 두 스키마는 동일한 이름을 공유하는 다른 개체를 가질 수 있습니다.

동일한 데이터베이스에 여러 스키마를 만들어 데이터를 원하는 방식으로 구성하거나 데이터를 기능적으로 그룹화할 수 있습니다. 예를 들어 모든 준비 데이터를 저장하는 스키마와 모든 보고 테이블을 저장하는 또 다른 스키마를 만들 수 있습니다. 다른 스키마를 생성하여 동일한 데이터베이스에 있는 서로 다른 업무 그룹과 관련된 데이터를 저장할 수도 있습니다. 각 스키마는 테이블, 뷰, UDF (사용자 정의 함수)와 같은 서로 다른 데이터베이스 개체를 저장할 수 있습니다. 또한 권한 부여 절로 스키마를 만들 수 있습니다. 이 절은 지정된 사용자에게 소유권을 부여하거나 지정된 스키마에서 사용할 수 있는 최대 디스크 공간에 대한 할당량을 설정합니다.

PostgreSQL은 자동으로 **public** 모든 새 데이터베이스. 데이터베이스 개체를 만드는 동안 스키마 이름을 지정하지 않으면 개체는 **public** 스키마를 생성합니다.

스키마의 개체에 액세스하려면 **schema_name.table_name** 표기법을 사용합니다. 스키마의 정규화된 이름은 점으로 구분된 스키마 이름과 테이블 이름으로 구성됩니다. 예를 들어, 당신은 **sales** 스키마가 있는 **price** 테이블과 **inventory** 스키마에는 **price** 테이블을 엽니다. 당신이 참조 할 때 **price** 테이블을 사용하려면 다음과 같이 한정해야 합니다. **sales.price** 또는 **inventory.price**.

다음 예시에서는 이라는 스키마를 생성합니다. **SALES** 해당 사용자에게 대한 **GUEST**.

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

더 많은 명령 옵션에 대해서는 [스키마 생성](#)의 Amazon Redshift 데이터베이스 개발자 안내서.

데이터베이스에서 스키마 목록을 보려면 다음 명령을 실행합니다.

```
select * from pg_namespace;
```

다음과 같이 출력됩니다.

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 [카탈로그 테이블 쿼리](#)의 Amazon Redshift 데이터베이스 개발자 안내서

GRANT 문을 사용하여 사용자에게 스키마에 대한 사용 권한을 부여합니다.

다음 예제에서는 권한을 부여하려면 GUEST를 사용하여 모든 테이블 또는 뷰에서 데이터를 선택할 수 있습니다. SALES SCHEMA SELECT 문을 사용합니다.

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

다음 예제에서는 사용 가능한 모든 권한을 한 번에 GUEST 해당 사용자.

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

작업 4: 테이블 생성

새 데이터베이스를 만든 후 데이터를 보관할 테이블을 만듭니다. 테이블을 생성할 때는 임의의 열 정보를 지정합니다.

다음 예제에서 GUEST 사용자가 SALES DB 새 테이블을 만듭니다.

예를 들어 이라는 테이블을 만들려면 DEMO에서 다음 명령을 실행합니다.

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

또한 사용하여 테이블을 만들 수 있습니다 schema_name.object_name 표기법을 사용하여 테이블을 생성할 수 있습니다. SALES 스키마를 생성합니다.

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

Amazon Redshift 쿼리 편집기를 사용하여 스키마와 해당 테이블을 보고 검사할 수 있습니다. 또는 시스템 뷰를 사용하여 스키마의 테이블 목록을 볼 수 있습니다. 자세한 내용은 [작업 6: 시스템 테이블 쿼리 \(p. 19\)](#) 단원을 참조하십시오.

기본적으로 테이블과 같은 새 데이터베이스 객체는 public 클러스터 생성 중에 생성됩니다. 다른 스키마를 사용하여 데이터베이스 개체를 만들 수 있습니다. 스키마에 대한 자세한 내용은 단원을 참조하십시오. [데이터베이스 보안 관리](#)의 Amazon Redshift 데이터베이스 개발자 안내서.

encoding, distkey, 및 sortkey 열은 Amazon Redshift에서 병렬 처리에 사용됩니다. 이러한 요소를 통합하는 테이블을 디자인하는 방법에 대한 자세한 내용은 단원을 참조하십시오. [Amazon Redshift 모범 사례](#).

테이블에 데이터 행 삽입

테이블을 만든 다음 해당 테이블에 데이터 행을 삽입합니다.

Note

이 **INSERT** 명령은 테이블에 행을 삽입합니다. 표준 대량 로드는 **COPY** 명령을 실행합니다. 자세한 내용은 단원을 참조하십시오. **COPY** 명령을 사용하여 데이터 로드.

예를 들어, 값을 삽입하려면 **DEMO** 테이블에 다음 명령을 실행합니다.

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

테이블에서 데이터 삭제

테이블을 만들고 데이터로 채운 다음 **SELECT** 문을 사용하여 테이블에 포함된 데이터를 표시합니다. **SELECT *** 문은 모든 열 이름과 테이블에 있는 모든 데이터의 행 값을 반환하며, **SELECT** 사용하여 최근에 추가된 데이터가 테이블에 올바르게 삽입되었는지 확인하기에 좋습니다.

에서 입력한 데이터를 보려면 **DEMO** 테이블에 다음 명령을 실행합니다.

```
SELECT * from SALES.DEMO;
```

결과는 다음과 같습니다.

personid	city
781	San Jose
990	Palo Alto

(2 rows)

SELECT 문을 사용하여 테이블을 쿼리하는 자세한 방법은 [선택](#).

작업 5: 샘플 데이터 로드

이 안내서에 나오는 예는 대부분 **TICKIT** 샘플 데이터 집합을 사용합니다. 파일을 다운로드할 수 있습니다. [tickitdb.zip](#) 개별 샘플 데이터 파일이 포함되어 있습니다.

이 자습서의 샘플 데이터는 Amazon S3 버킷에서 제공됩니다. 이러한 버킷은 인증된 모든 AWS 사용자, 그래서 유효한 AWS Amazon S3 작업에 대한 액세스를 허용하는 자격 증명입니다.

데이터베이스를 위한 샘플 데이터를 로드하려면 먼저 테이블을 만든 다음 **COPY** 명령을 사용하여 Amazon S3 버킷에 저장된 샘플 데이터와 함께 테이블을 로드합니다. 테이블을 생성하고 샘플 데이터를 로드하는 단계는 [5단계: Amazon S3에서 샘플 데이터 로드 \(p. 11\)](#).

작업 6: 시스템 테이블 쿼리

데이터베이스에는 생성한 테이블 외에 많은 시스템 테이블이 포함되어 있습니다. 이들 시스템 테이블에는 설치 및 시스템에서 실행 중인 쿼리와 프로세스에 대한 정보가 포함되어 있습니다. 이들 시스템 테이블을 쿼리하여 데이터베이스에 대한 정보를 수집할 수 있습니다.

Note

이 설명서에 나오는 각 테이블에 대한 설명은 테이블이 모든 사용자에게 보이는지 아니면 슈퍼유저에게만 보이는지를 나타냅니다. 슈퍼유저에게만 보이는 테이블을 쿼리하려면 슈퍼유저로 로그인합니다.

Amazon Redshift는 다음 유형의 시스템 테이블에 대한 액세스를 제공합니다.

- **STL 테이블**

이들 시스템 테이블은 Amazon Redshift 로그 파일에서 생성되어 시스템 이력 정보를 제공합니다. 로깅 테이블에는 STL 접두사가 있습니다.

- **STV 테이블**

이 테이블은 현재 시스템 데이터의 스냅샷을 포함하는 가상 시스템 테이블입니다. 스냅샷 테이블에는 STV 접두사가 있습니다.

- **시스템 뷰**

시스템 뷰에는 몇몇 STL 및 STV 시스템 테이블에서 찾을 수 있는 데이터 하위 집합이 포함됩니다. 시스템 뷰에는 SVV 또는 SVL 접두사가 있습니다.

- **시스템 카탈로그 테이블**

시스템 카탈로그 테이블은 테이블과 열에 대한 정보 등 스키마 메타데이터를 저장합니다. 시스템 카탈로그 테이블에는 PG 접두사가 있습니다.

쿼리에 대한 시스템 테이블 정보를 검색하려면 해당 쿼리와 연결된 프로세스 ID를 지정해야 합니다. 자세한 내용은 [실행 중인 쿼리의 프로세스 ID 확인 \(p. 22\)](#) 단원을 참조하십시오.

테이블 이름 목록 보기

스키마의 모든 테이블 목록을 보려면 PG_TABLE_DEF 시스템 카탈로그 테이블을 쿼리하면 됩니다. 먼저 설정을 검사할 수 있습니다.`search_path`.

```
SHOW search_path;
```

결과는 다음과 비슷해야 합니다.

```
search_path
-----
$user, public
(1 row)
```

다음 예제에서는 SALES 스키마를 검색 경로에 추가한 후 SALES 스키마를 생성합니다.

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

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

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

다음 예시에서는 이라는 모든 테이블의 목록을 표시합니다.DEMO현재 데이터베이스의 모든 스키마에서.

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

자세한 내용은 단원을 참조하십시오.[PG_테이블_데프](#).

Amazon Redshift 쿼리 편집기를 사용하여 먼저 연결하려는 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;
 username | usesysid | usecreatedb | usesuper | usecatupd | passwd | valuntil |
 useconfig
+-----+-----+-----+-----+-----+-----+-----+
+-----+
 rdsdb    | 1 | true | true | true | ***** | infinity |
 awsuser  | 100 | true | true | false | ***** | |
 guest    | 104 | true | false | false | ***** | |
(3 rows)
```

사용자 이름 rdsdbAmazon Redshift 에서 일상적 관리 및 유지 작업을 위해 내부적으로 사용합니다. 다음을 추가하여 사용자 정의 사용자 이름만 표시되도록 쿼리를 필터링할 수 있습니다.where usesysid > 1를 SELECT 문에 추가합니다.

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

최근 쿼리 보기

앞의 예에서 사용자 ID (USESYSID)adminuser은 100입니다. 가장 최근 쿼리 5개를 나열하려면adminuser에서 SVL_QLOG 뷰에 대한 쿼리를 실행하여 확인할 수 있습니다.

SVL_QLOG 뷰는 보다 친숙한 STL_QUERY 테이블의 정보 하위 집합입니다. 이 뷰를 사용하여 최근에 실행된 쿼리의 쿼리 ID (QUERY) 또는 프로세스 ID (PID) 를 찾을 수 있습니다. 또한 이 뷰를 사용하여 쿼리를 완료하는 데 걸린 시간을 확인할 수 있습니다. SVL_QLOG에는 특정 쿼리를 찾는 데 도움이 되도록 쿼리 문자열(SUBSTRING)의 첫 60자가 포함되어 있습니다. SELECT 문과 함께 LIMIT 절을 사용하여 결과를 5행으로 제한하십시오.


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

결과는 다음과 같습니다.

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

실행 중인 쿼리의 프로세스 ID 확인

앞의 예에서는 SVL_QLOG 뷰에서 완료된 쿼리의 쿼리 ID와 프로세스 ID (PID) 를 가져오는 방법을 배웠습니다.

아직도 실행 중인 쿼리의 PID를 찾아야 합니다. 예를 들어 실행에 너무 긴 시간이 소요되는 쿼리를 취소해야 하는 경우 PID가 필요합니다. STV_RECENTS 시스템 테이블을 쿼리하여 실행 중인 쿼리의 프로세스 ID 목록과 해당 쿼리 문자열을 가져올 수 있습니다. 쿼리가 여러 개의 PID를 반환하는 경우, 쿼리 텍스트를 보고 필요한 PID를 파악할 수 있습니다.

실행 중인 쿼리의 PID를 확인하려면 다음 SELECT 문을 실행합니다.

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

작업 7: 쿼리 취소

시간이 너무 오래 걸리거나 과도한 클러스터 리소스를 사용하는 쿼리를 실행하는 경우, 해당 쿼리를 취소하십시오. 예를 들어 판매자의 이름과 판매된 티켓 수량이 포함된 티켓 셀러 목록을 생성합니다. 다음 쿼리에서 데이터를 선택 SALES 테이블 및 USERS 테이블에서 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 비슷합니다.

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 집합의 모든 행을 포함 하는SALES테이블의 모든 행을 곱한USERS테이블 (49989*3766) 을 참조하십시오. 이것을 카티전 조인이라고 하는데, 권장하지 않습니다. 결과는 1억 8,800만 개 이상의 행으로 실행하려면 시간이 많이 걸립니다.

실행 중인 쿼리를 취소하려면 쿼리의 PID와 함께 CANCEL 명령을 사용하십시오.

프로세스 ID를 찾으려면 이전 단계에 나온 것처럼 새 세션을 시작하고 STV_RECENTS 테이블을 쿼리합니다. 다음 예에서는 결과를 보다 읽기 가능하게 만드는 방법을 보여 줍니다. 이렇게 하려면 TRIM 함수를 사용하여 후행 공백을 트리밍하고 쿼리 문자열의 첫 20자만 표시합니다.

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

결과는 다음과 같습니다.

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

PID 18764인 쿼리를 취소하려면 다음 명령을 실행합니다.

```
CANCEL 610;
```

Note

취소 명령은 트랜잭션을 중지하지 않습니다. 트랜잭션을 중지하거나 롤백하려면 ABORT 또는 ROLLBACK 명령을 사용합니다. 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.

다른 세션에서 쿼리 취소

동시에 실행되는 쿼리를 쿼리 도구가 지원하지 않는 경우, 쿼리를 취소하려면 다른 세션을 시작합니다. 예를 들어, 우리가 사용하는 쿼리 편집기는 Amazon Redshift 시작하기는 여러 개의 동시 쿼리를 지원하지 않습니다. 쿼리 편집기로 다른 세션을 시작하려면 File, 새 창 동일한 연결 매개 변수를 사용하여 연결합니다. 그러면 PID를 찾아 쿼리를 취소할 수 있습니다.

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 대기열을 사용하려면 수퍼유저로 로그인하고 SET 명령을 사용해 쿼리 그룹을 'superuser'로 설정합니다. 명령을 실행한 후 RESET 명령을 사용하여 쿼리 그룹을 리셋합니다.

수퍼유저 대기열을 사용해 쿼리를 취소하려면 다음 명령을 실행합니다.

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

쿼리 대기열에 대한 자세한 내용은 단원을 참조하십시오. [워크로드 관리 구현](#).

작업 8: 리소스 정리

이 연습을 마치기 위해 클러스터를 배포한 경우 연습을 마치면 클러스터를 삭제합니다. 클러스터를 삭제하면 클러스터가 중지되어 AWS 계정.

클러스터를 삭제하려면 [클러스터 삭제](#)의 Amazon Redshift 클러스터 관리 가이드.

클러스터를 그대로 두려면 참조를 위해 샘플 데이터를 보관하십시오. 이 안내서에 나오는 대부분의 예는 이 연습에서 만드는 테이블을 사용합니다. 데이터 크기는 가용 스토리지에 별 영향을 주지 않습니다.

클러스터를 그대로 두되 샘플 데이터를 정리하려면 다음 명령을 실행해 SALESDB 데이터베이스.

```
DROP DATABASE SALESDB;
```

생성하지 않은 경우 SALESDB 데이터베이스를 삭제하지 않으려면 다음 명령을 실행해 테이블만 삭제합니다.

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

데이터 레이크 쿼리 시작하기

Amazon Redshift Spectrum m을 사용하면 Amazon S3 파일에 있는 데이터를 쿼리할 필요 없이 Amazon Redshift S3 파일을 쿼리할 수 있습니다. 당사는 마루, ORC, RCFile, TextFile, SequenceFile, RegexSerde, OpenCSV, AVRO를 포함하여 다양한 형식의 데이터를 쿼리할 수 있습니다. 외부 스키마와 테이블을 생성하여 Amazon S3에서 파일의 구조를 정의합니다. 그런 다음 다음과 같은 외부 데이터 카탈로그를 사용합니다. AWS Glue 또는 자신의 Apache Hive 메타스토어를 사용할 수 있습니다. Amazon Redshift 클러스터에서 어느 데이터 카탈로그에서나 변경 사항을 즉시 사용할 수 있습니다.

귀하의 데이터가 등록된 후 AWS Glue 데이터 카탈로그 및 AWS Lake Formation, 당사는 Redshift 스펙트럼을 사용하여 쿼리할 수 있습니다.

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

하나 이상의 열에서 외부 테이블을 파티셔닝하여 파티션 제거를 통해 쿼리 성능을 최적화할 수 있습니다. Amazon Redshift 테이블을 사용하여 외부 테이블을 쿼리하고 조인할 수 있습니다. 여러 Amazon Redshift 클러스터에서 외부 테이블에 액세스하여 Amazon S3 데이터를 쿼리할 수 있습니다. AWS 리전. Amazon S3 데이터 파일을 업데이트하면 Amazon Redshift 클러스터의 어느 쿼리에서나 해당 데이터를 즉시 사용할 수 있습니다.

Redshift 스펙트럼 및 데이터 레이크로 작업하는 방법을 포함하여 Redshift 스펙트럼에 대한 자세한 내용은 [Amazon Redshift Spectrum m 시작하기](#) in Amazon Redshift 데이터베이스 개발자 안내서.

원격 데이터 원본의 데이터 쿼리 시작

통합 쿼리를 사용하여 Amazon RDS 데이터베이스, Amazon Aurora 데이터베이스 또는 Amazon S3 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. 페더레이션된 쿼리에 대한 계산 중 일부는 원격 데이터 원본에 배포됩니다.

연동 쿼리를 실행하기 위해 Amazon Redshift는 먼저 원격 데이터 소스에 연결합니다. 그런 다음 Amazon Redshift는 원격 데이터 소스의 테이블에 대한 메타데이터를 검색하고 쿼리를 실행한 다음 결과 행을 검색합니다. 그런 다음 Amazon Redshift는 추가 처리를 위해 결과 행을 Amazon Redshift 컴퓨팅 노드로 배포합니다.

연합 쿼리 사용 환경 설정에 대한 자세한 내용은 Amazon Redshift 데이터베이스 개발자 안내서:

- [PostgreSQL에 연합 쿼리 사용 시작하기](#)
- [MySQL에 연합 쿼리 사용 시작하기](#)

다른 Amazon Redshift 클러스터의 데이터 액세스 시작

Amazon Redshift 데이터 공유를 사용하면 Amazon Redshift 클러스터 또는 AWS 계정을 읽기 위해 사용할 수 있습니다. Amazon Redshift 클러스터 전체에서 데이터를 수동으로 복사하거나 이동할 필요 없이 즉각적이고 세분화된 고성능 데이터에 액세스할 수 있습니다. 사용자는 Amazon Redshift 클러스터에서 업데이트되는 최신 정보와 일관된 정보를 볼 수 있습니다.

Amazon Redshift 데이터 공유는 다음과 같은 사용 사례에 특히 유용합니다.

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

데이터 공유에 대한 자세한 내용은 단원을 참조하십시오. [데이터 공유 시작하기](#)의 Amazon Redshift 데이터베이스 개발자 안내서.

Amazon Redshift 데이터로 머신 러닝 모델 교육 시작하기

Amazon Redshift 기계 학습 (Amazon Redshift ML) 을 사용하면 Amazon Redshift에 데이터를 제공하여 모델을 학습할 수 있습니다. 그런 다음 Amazon Redshift ML은 입력 데이터의 패턴을 캡처하는 모델을 생성합니다. 그런 다음 이러한 모델을 사용하여 추가 비용을 들이지 않고 새 입력 데이터에 대한 예측을 생성할 수 있습니다. Amazon Redshift ML을 사용하면 SQL 문을 사용하여 기계 학습 모델을 학습하고 예측을 위해 SQL 쿼리에서 호출할 수 있습니다. 반복적으로 매개 변수를 변경하고 학습 데이터를 개선하여 예측의 정확도를 계속 향상시킬 수 있습니다.

Amazon Redshift ML을 사용하면 SQL 사용자가 익숙한 SQL 명령을 사용하여 기계 학습 모델을 보다 쉽게 생성, 교육 및 배포할 수 있습니다. Amazon Redshift ML을 사용하면 Amazon Redshift 클러스터의 데이터를 사용하여 Amazon SageMaker에서 모델을 학습할 수 있습니다. 그런 다음 모델을 현지화하고 Amazon Redshift 데이터베이스 내에서 예측을 수행할 수 있습니다.

Amazon Redshift ML에 대한 자세한 내용은 단원을 참조하십시오. [Amazon Redshift ML 시작하기](#)의 Amazon Redshift 데이터베이스 개발자 안내서

추가 리소스

이 자습서를 완료하면 다음 Amazon Redshift 리소스를 사용하여 본 안내서에서 소개한 개념에 대해 더 알아보면 좋습니다.

- [Amazon Redshift 클러스터 관리 가이드](#): 이 가이드는 이를 기반으로 Amazon Redshift 시작하기, 클러스터 생성, 관리 및 모니터링에 대한 개념 및 작업에 대한 깊이 있는 정보를 제공합니다.
- [Amazon Redshift 데이터베이스 개발자 안내서](#): 이 가이드는 이 Amazon Redshift 시작하기. 또한, 사용자의 데이터웨어 하우스를 구성하는 데이터베이스를 설계, 구축, 쿼리하는 방법에 대한 깊이 있는 정보를 데이터베이스 개발자에게 제공합니다.
 - [SQL 참조](#): 이 항목에서는 Amazon Redshift 의 SQL 명령 및 함수 참조에 대해 설명합니다.
 - [시스템 테이블 및 뷰](#): 이 항목에서는 Amazon Redshift 의 시스템 테이블 및 뷰에 대해 설명합니다.
- Amazon Redshift 자습서: 이 항목에서는 Amazon Redshift 기능에 대해 알아보려면 자습서를 보여 줍니다.
 - [Amazon Redshift 와 함께 공간 SQL 함수 사용](#): Amazon Redshift 와 공간 SQL 함수 중 일부를 사용하는 방법을 소개합니다.
 - [Amazon S3 에서 데이터 로드](#): Amazon S3 버킷에 있는 데이터 파일에서 Amazon Redshift 데이터베이스 테이블로 데이터를 로드하는 방법을 설명합니다.
 - [Amazon Redshift Spectrum m을 사용하여 중첩 데이터 쿼리](#): 이 자습서는 Redshift Spectrum을 사용하여 외부 테이블을 사용하여 Parquet, ORC, JSON 및 이온 파일 형식에서 중첩 데이터를 쿼리하는 방법을 설명합니다.
 - [수동 워크로드 관리 \(WLM\) 대기열 구성](#): Amazon Redshift 에서 수동 워크로드 관리 (WLM) 를 구성하는 방법을 설명합니다.
- 동영상 기능: 이 동영상은 Amazon Redshift 기능에 대해 알아보는 데 도움이 됩니다.
 - Amazon Redshift 를 시작하는 방법을 알아보려면 다음 비디오를 시청하십시오. [Amazon Redshift 와 연결](#).
 - Amazon Redshift 데이터 공유의 작동 방식을 알아보려면 다음 비디오를 시청하십시오. [Amazon Redshift 데이터 공유 워크플로](#).
 - Amazon Redshift Machine Learning (ML) 의 작동 방식을 알아보려면 다음 비디오를 시청하십시오. [Amazon Redshift ML](#).
 - Amazon Redshift 콘솔의 쿼리 모니터링 기능을 사용하여 쿼리를 모니터링, 격리 및 최적화하는 방법을 알아보려면 다음 비디오를 시청하십시오. [Amazon Redshift 와 쿼리 모니터링](#).
- [새 소식](#): 이 웹 페이지에는 Amazon Redshift 의 새로운 기능 및 제품 업데이트가 나열되어 있습니다.

문서 기록

다음 표에서는 주요 변경 사항을 설명합니다. Amazon Redshift 시작 안내서.

최신 설명서 업데이트: 2020년 6월 30일

변경 사항	설명	릴리스 날짜
설명서 업데이트	일반적인 데이터베이스 작업 시작, 데이터 레이크 쿼리, 원격 소스에서 데이터 쿼리, 데이터 공유, Amazon Redshift 데이터를 사용한 기계 학습 모델 교육에 대한 새로운 섹션을 포함하도록 가이드가 업데이트되었습니다.	2020년 6월 30일
새로운 기능	새 샘플 로드 절차를 설명하도록 안내서가 업데이트되었습니다.	2020년 6월 4일
설명서 업데이트	원래 Amazon Redshift 콘솔을 제거하고 단계 흐름을 개선하도록 가이드를 업데이트했습니다.	2020년 8월 14일
새로운 콘솔	새 Amazon Redshift 콘솔을 설명하도록 안내서가 업데이트되었습니다.	2019년 11월 11일
새로운 기능	빠른 클러스터 시작 절차를 설명하기 위해 안내서가 업데이트되었습니다.	2018년 8월 10일
새로운 기능	Amazon Redshift 대시보드에서 클러스터를 시작하도록 안내서가 업데이트되었습니다.	2015년 7월 28일
새로운 기능	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 S3 정보의 로드 정보가 자체 단원으로 이동했으며 다음 단계 단원으로 이동되어 찾기 쉬워졌습니다.	2014년 5월 13일
설명서 업데이트	환영합니다 페이지를 없애고 해당 콘텐츠를 기본 시작하기 페이지에 통합하였습니다.	2014년 3월 14일
설명서 업데이트	이것은의 새로운 릴리스입니다 Amazon Redshift 시작 안내서에서 고객 피드백 및 서비스 업데이트가 반영된	2014년 3월 14일
새 설명서	이 설명서는 의 최초 릴리스입니다. Amazon Redshift 시작 안내서.	2013년 14월 2일

기계 번역으로 제공되는 번역입니다. 제공된 번역과 원본 영어의 내용이 상충하는 경우에는 영어 버전이 우선합니다.