Introduction to Amazon Redshift

**SPL-86 - Version 2.1.13**

© 2024 Amazon Web Services, Inc. or its affiliates. All rights reserved. This work may not be reproduced or redistributed, in whole or in part, without prior written permission from Amazon Web Services, Inc. Commercial copying, lending, or selling is prohibited. All trademarks are the property of their owners.

Note: Do not include any personal, identifying, or confidential information into the lab environment. Information entered may be visible to others.

Corrections, feedback, or other questions? Contact us at [*AWS Training and Certification*](https://support.aws.amazon.com/#/contacts/aws-training).

**Introduction**

This lab provides a basic understanding of Amazon Redshift. It demonstrates the basic steps required to get started with Redshift including:

* Creating a Redshift cluster
* Loading data into a Redshift cluster
* Performing queries against the data in cluster

TOPICS COVERED

By the end of this lab you will be able to:

* Launch a Redshift cluster
* Connect an SQL client to the Amazon Redshift cluster
* Load data from an S3 bucket into the Amazon Redshift cluster
* Run queries against data stored in Amazon Redshift

AMAZON REDSHIFT

Amazon Redshift is a fast, fully managed [data warehouse](https://aws.amazon.com/data-warehouse/) that makes it simple and cost-effective to analyze all your data using standard SQL and your existing Business Intelligence (BI) tools.

AMAZON S3

Amazon Simple Storage Service (Amazon S3) makes it simple and practical to collect, store, and analyze data - regardless of format - all at massive scale. S3 is object storage built to store and retrieve any amount of data from anywhere - web sites and mobile apps, corporate applications, and data from IoT sensors.

OTHER AWS SERVICES

During this lab, you may receive error messages when performing actions beyond the steps in this lab guide. These messages will not impact your ability to complete the lab. We recommend you remain within the steps provided by these lab instructions.

PREREQUISITES

Familiarity with relational databases and SQL concepts would be beneficial.

ICON KEY

Various icons are used throughout this lab to call attention to different types of instructions and notes. The following list explains the purpose for each icon:

* **Command:** A command that you must run.
* **Expected output:** A sample output that you can use to verify the output of a command or edited file.
* **Note:** A hint, tip, or important guidance.
* **Caution:** Information of special interest or importance (not so important to cause problems with the equipment or data if you miss it, but it could result in the need to repeat certain steps).
* **Refresh:** A time when you might need to refresh a web browser page or list to show new information.

**Start lab**

1. To launch the lab, at the top of the page, choose **Start lab**.

**Caution:** You must wait for the provisioned AWS services to be ready before you can continue.

1. To open the lab, choose **Open Console**.

You are automatically signed in to the AWS Management Console in a new web browser tab.

**Warning:** Do not change the **Region** unless instructed.

COMMON SIGN-IN ERRORS

**Error: Choosing Start Lab has no effect**

In some cases, certain pop-up or script blocker web browser extensions might prevent the **Start Lab** button from working as intended. If you experience an issue starting the lab:

* Add the lab domain name to your pop-up or script blocker’s allow list or turn it off.
* Refresh the page and try again.

**Task 1: Launch an Amazon Redshift Cluster**

In this task, you will launch an Amazon Redshift cluster. A cluster is a fully managed **data warehouse** that consists of a set of compute nodes. Each cluster runs an Amazon Redshift engine and contains one or more databases.

When you launch a cluster, one of the options you specify is the **node type**. The node type determines the CPU, RAM, storage capacity, and storage drive type for each node. Node types are available in different sizes. Node size and the number of nodes determine the total storage for a cluster.

1. At the top of the AWS Management Console, to the right of **Services** menu, in the search bar, search for

**Amazon Redshift**

 and then choose **Amazon Redshift** from the list.

1. In the left navigation pane, choose **Clusters**.

 You might have to choose the menu  icon to expand the left navigation pane.

1. Choose **Create cluster** to open the Redshift Cluster Creation Wizard.
2. In the **Cluster configuration** section, configure:

* **Cluster identifier:**

lab

* **Node type:**

dc2.large

* **Number of nodes:**

2

1. In the **Database configurations** section, configure:

* **Admin user name:**

master

* For **Admin Password**, select **Manually add the admin password** and configure **Admin user password:**

Redshift123

1. For **Associated IAM roles**, choose **Associate IAM role** , select

**Redshift-Role**

.

1. Choose **Associated IAM Roles** .

 The role grants permission for Amazon Redshift to read data from Amazon S3.

1. In the **Additional configurations** section, deselect  **Use defaults**.
2. Expand  **Network and security**, then configure:

* **Virtual private cloud** *Lab VPC*.
* **VPC security groups:**
  + Deselect **default**.
  + Select **Redshift Security Group**.

1. Expand  **Database configurations**, then configure:

* **Database name**

labdb

1. Scroll to the bottom of the screen, then choose **Create cluster** .

 The cluster will take a few minutes to launch. Please continue with the labs steps. There is no need to wait.

1. Select the name of your Cluster (**lab**) .

 The cluster configuration will be displayed. Spend a few minutes looking at the properties.

1. Wait for the Status of your cluster to display **Available** before continuing to the next task. You should see a success message **lab has been successfully created**. If necessary,  refresh screen.

**Task 2: Use the Redshift Query Editor to Communicate with your Redshift Cluster**

Amazon Redshift can be used via industry-standard SQL. To use Redshift, you require an **SQL Client** that provides a user interface to type SQL. Any SQL client that supports JDBC or ODBC can be used with Redshift.

For this lab, you will use the Amazon Redshift Query editor.

1. In the left navigation pane, choose **Query editor**,then select **Connect to database**then configure:

* **Cluster:**

lab

* **Database name**:

labdb

* **Database user**:

master

1. Choose **Connect** .

**Task 3: Create a Table**

In this task, you will execute SQL commands to create a table in Redshift.

1. Copy this SQL command and paste it into the **Query 1** window, then choose **Run** .

CREATE TABLE users (

userid INTEGER NOT NULL,

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

);

 This command will create a table called **users**. It contains name, address and details about the type of music that the user likes.

**Task 4: Load Sample Data from Amazon S3**

Amazon Redshift can import data from Amazon S3. Various file formats are supported, fixed-length fields, comma-separated values (CSV) and custom delimiters. The data for this lab is pipe-separated (|).

1. Delete the existing query, then paste this SQL command into the **Query 1** window.

COPY users FROM 's3://awssampledbuswest2/tickit/allusers\_pipe.txt'

CREDENTIALS 'aws\_iam\_role=YOUR-ROLE'

DELIMITER '|';

 Before running this command, you will need to insert the ROLE that Redshift will use to access Amazon S3.

1. To the left of the instructions you are currently reading, copy the value for **Role**. It will start with: *arn:aws:iam::*
2. Paste the Role into the query window, replacing the text **YOUR-ROLE**.

 This 2nd line should now look like:

COPY users FROM 's3://awssampledbuswest2/tickit/allusers\_pipe.txt'

CREDENTIALS 'aws\_iam\_role=arn:aws:iam::xxxxxxxxxxxx:role/Redshift-Role'

DELIMITER '|';

1. Choose **Run** .

 The command will take approximately 10 seconds to load **49,990 rows of data**.

**Task 5: Query Data**

Now that you have data in your Redshift database you can query the data using SQL select statements and queries. If you are familiar with SQL, feel free to try additional commands to query the data.

1. Run this query to count the number of rows in the *users* table:

SELECT COUNT(\*) FROM users;

The result shows that there are almost 50,000 rows in the table.

1. Run this query:

SELECT userid, firstname, lastname, city, state

FROM users

WHERE likesports AND NOT likeopera AND state = 'OH'

ORDER BY firstname;

This query displays users in Ohio (OH) who like sports but do not like opera. The list is sorted by their first name.

1. Run this query:

SELECT

city,

COUNT(\*) AS count

FROM users

WHERE likejazz

GROUP BY city

ORDER BY count DESC

LIMIT 10;

This query shows the Top 10 cities where Jazz-loving users live.

CHALLENGE

Try to write a query for these requirements:

* Only display the *firstname* and *lastname*
* of users who like both *Theatre* and *Classical* music
* With a last name is *Smith*

Try to do it yourself before seeing the answer.

If you do not know the answer, [view the answer here](https://us-west-2-tcprod.s3.us-west-2.amazonaws.com/courses/spl-86/v2.1.13.prod-204be596/images/en_us/answer.png).

**Conclusion**

 Congratulations! You have completed the lab. During the lab you successfully:

* Launched a Redshift cluster
* Connected an SQL client to the Amazon Redshift cluster
* Loaded data from an S3 bucket into the Amazon Redshift cluster
* Ran queries against data stored in Amazon Redshift

**End lab**

Follow these steps to close the console and end your lab.

1. Return to the **AWS Management Console**.
2. At the upper-right corner of the page, choose **AWSLabsUser**, and then choose **Sign out**.
3. Choose **End lab** and then confirm that you want to end your lab.

**Additional Resources**

* [Amazon Redshift](http://docs.aws.amazon.com/redshift/latest/mgmt/welcome.html)
* [Amazon Redshift pricing](http://aws.amazon.com/redshift/pricing/)

For more information about AWS Training and Certification, see [*https://aws.amazon.com/training/*](https://aws.amazon.com/training/).

*Your feedback is welcome and appreciated.*  
If you would like to share any feedback, suggestions, or corrections, please provide the details in our [*AWS Training and Certification Contact Form*](https://support.aws.amazon.com/#/contacts/aws-training).