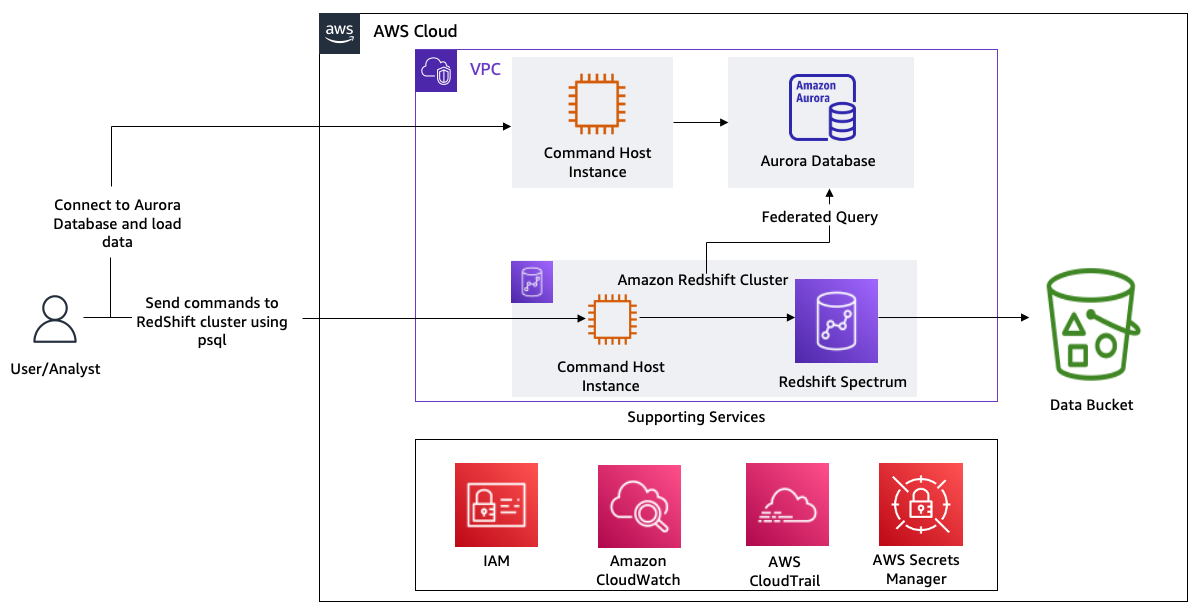
# Lab 2 - Data Analytics Using Amazon Redshift Spectrum

1. Use Amazon Redshift Spectrum to create an external table.
2. Query data stored in Amazon S3.
3. Query Amazon Aurora data from the Amazon Redshift query editor with federated query access.
4. Use the UNLOAD command to save query results to Amazon S3.

UNLOAD Command: This command in Redshift exports the results of your queries to files stored in S3.

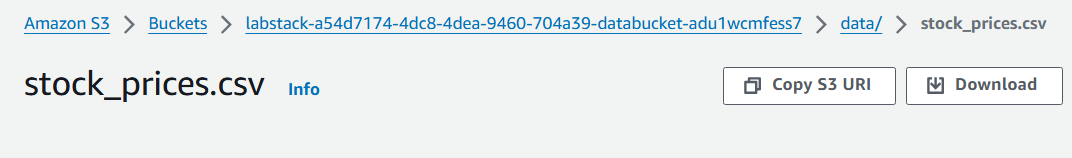
Data API: This is an API that allows you to execute SQL commands and manage your Redshift cluster using code, rather than having to use the Redshift console or other tools manually.

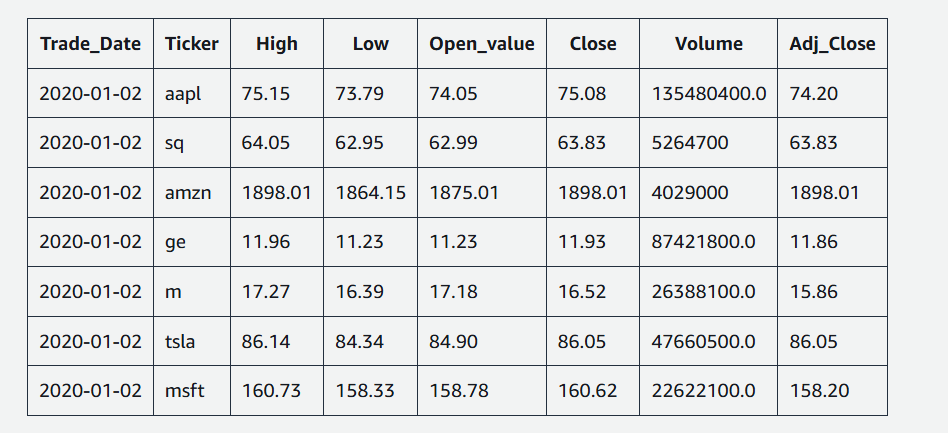
Federated queries allow Amazon Redshift to run SQL queries that pull data from another database, like Amazon Aurora, without moving all the data into Redshift first.



**Task 1: Explore the lab environment**

1.1 Open s3, go tp this location and view this file



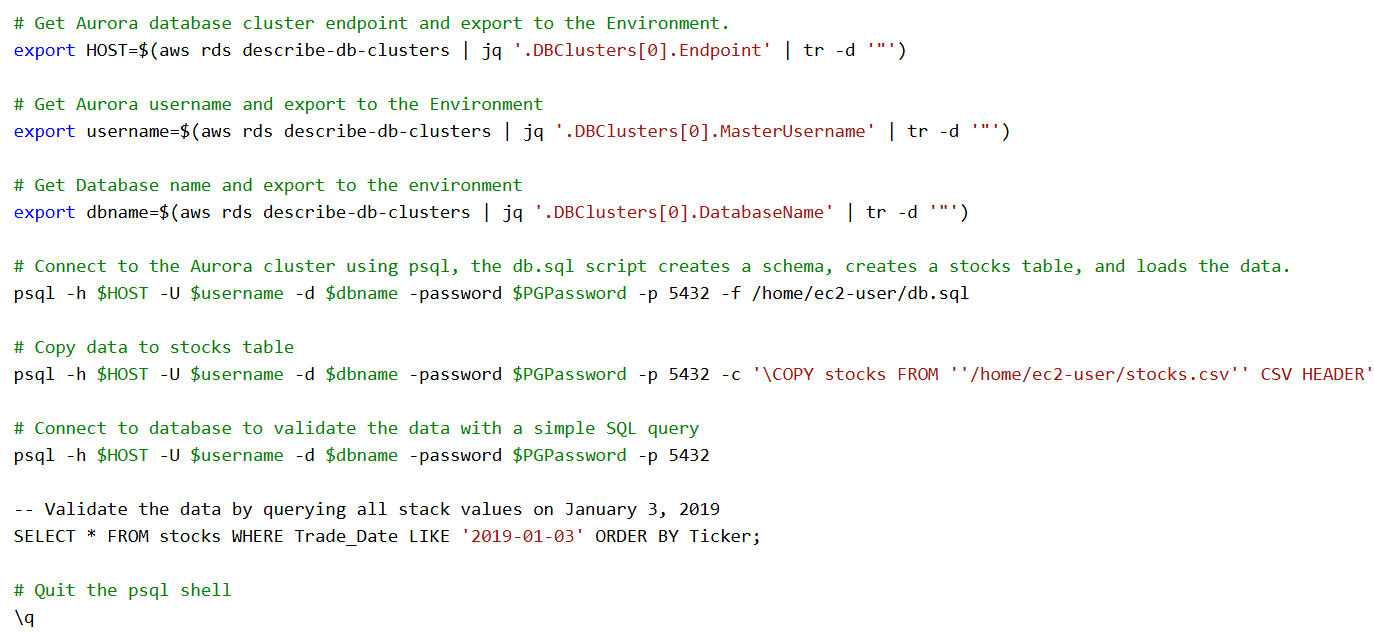


1.2 Load data from S3 to Aurora

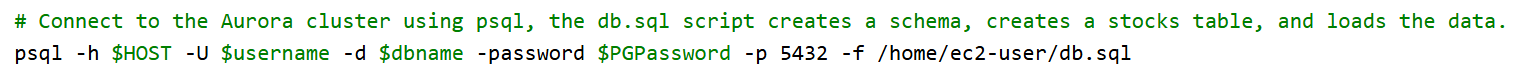
Open link to open CLI

1.3 run this script to do -

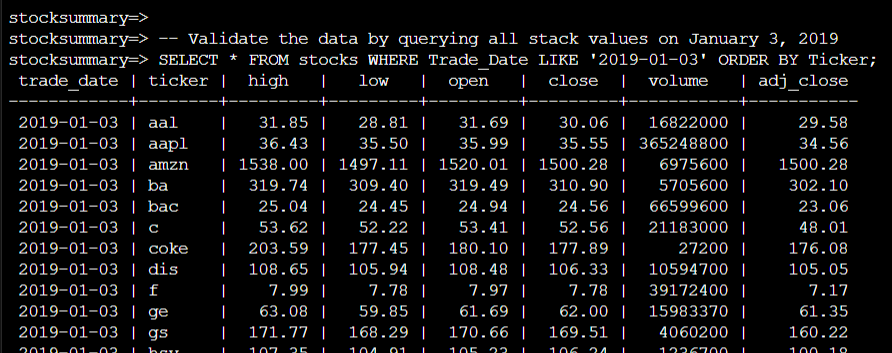
1. Creates a stocksummary schema in the Aurora database
2. Creates a stocks table
3. Copies the stocks CSV data to the stocks table
4. Connects to a database psql console
5. Queries the data from the stocks table
6. Exits the psql shell



1.4 all major commands are stored in



o/p-



**Task 2: Use Redshift Spectrum to analyze data in the Amazon S3 bucket**

2.1 Use URL to open CLI to access psql there

2.2 connect via

cd ~

psql -U dbadmin -h '<INSERT\_REDSHIFT\_CLUSTER\_ENDPOINT>' -d lab -p 5439

2.3 create schema

CREATE EXTERNAL SCHEMA spectrum

FROM DATA CATALOG

DATABASE 'spectrumdb'

IAM\_ROLE 'INSERT\_REDSHIFT\_ROLE'

CREATE EXTERNAL DATABASE IF NOT EXISTS;

2.4 create external table

DROP TABLE IF EXISTS spectrum.stocksummary;

CREATE EXTERNAL TABLE spectrum.stocksummary(

Trade\_Date VARCHAR(15),

Ticker VARCHAR(5),

High DECIMAL(8,2),

Low DECIMAL(8,2),

Open\_value DECIMAL(8,2),

Close DECIMAL(8,2),

Volume DECIMAL(15),

Adj\_Close DECIMAL(8,2)

)

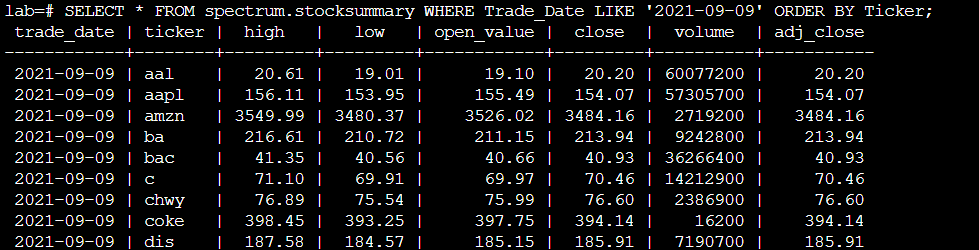
ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION 's3://INSERT\_DATA\_BUCKET\_NAME/data/';

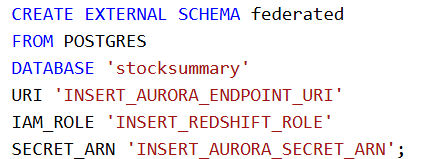
2.5 test query



**Task 3: Analyze data stored in Aurora using federated queries**

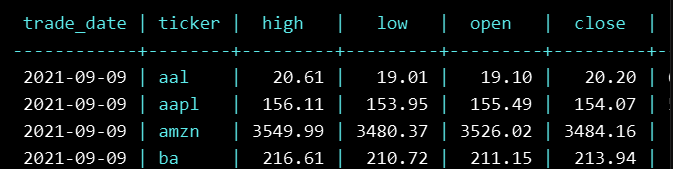
Gain - parallel processing, and data storage

3.1 create schema

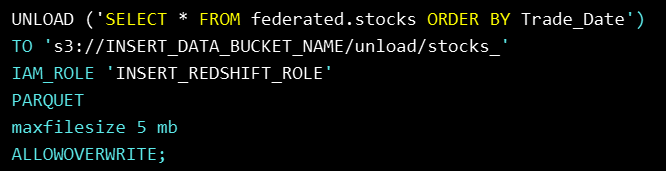


3.2 directly access from tables in aurora DB

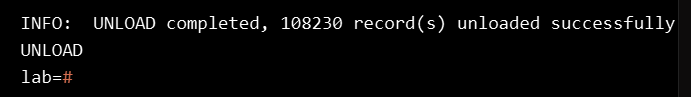
SELECT \* FROM federated.stocks WHERE Trade\_Date LIKE '2021-09-09' ORDER BY Ticker;



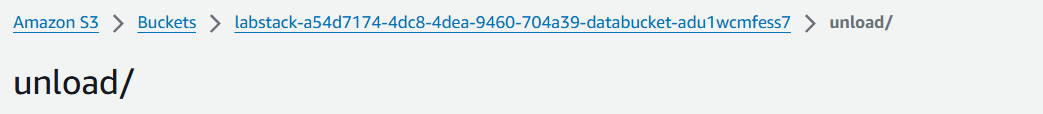
**Task 4: Unload analyzed data to Amazon S3**

4.1 move data raken from aurora to S3

Apache Parquet is the default file format to unload your data from Amazon Redshift

4.2 

4.3 open S3 location to view the transferred data



Challenge - query to access unloaded S3 data in psql

Answer -

