**AWS Final Project**

1. Create an S3 bucket with a unique name and upload the CSV file to the S3 bucket (ensure that the file is in UTF-8 format only)

* Open AWS S3 and follow the screen shots.

Graphical user interface, text, website

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

1. Create a crawler to crawl the CSV data and generate a metadata catalog

* Open Aws Glue and follow the screen shots.

Graphical user interface, text

Description automatically generated with medium confidence

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, chat or text message

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

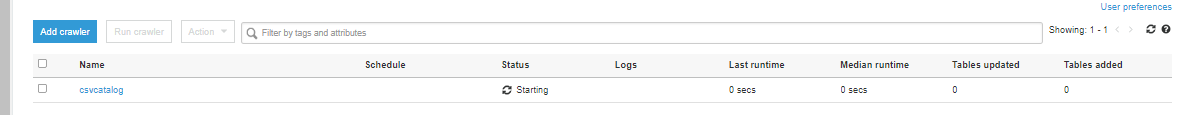
Description automatically generated

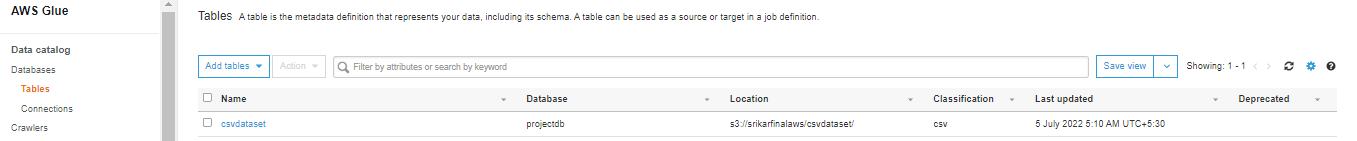
Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated





1. Create a Glue job to transform the data into the Parquet format as CSV is not optimal for data warehouse queries

* Open IAM and create a Role.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

* Open AWS Glue 🡪 jobs and follow the screen shots.

Graphical user interface, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. Add another crawler to crawl the Parquet data files to generate the metadata catalog of the Parquet file in order to query it with Athena

* Follow the 2nd answer to create metadata catalog.

Graphical user interface, text, application

Description automatically generated

* Now open Aws Athena and follow the screen shots.

Graphical user interface, text, application

Description automatically generated

1. Query the data to identify the best-selling item and countries where customers have bought the most-sold item using Athena

A screenshot of a computer

Description automatically generated