Amazon Web Services

Lab1. Copy RDS Source Data - Prelab

July 2021

Table of Contents

Lab1. Copy RDS Source Data - Prelab

About the lab setup:

RDS Postgres Database is used as a source of ticket sales system for sporting events. It stores transaction information about ticket sales price to selected people and ticket ownership transfer with additional tables for event details. AWS Database Migration Service (DMS) is used for a full data load from the Amazon RDS source to Amazon S3 bucket.

Before the Glue lab starts, you might choose to skip the DMS data migration, instead copy the source data to your S3 bucket directly.

In today's lab, you will copy the data from a centralized S3 bucket to your AWS account, crawl the dataset with AWS Glue crawler for metadata creation and transform the data with AWS Glue to Query data and create a View with Athena and Build a dashboard with Amazon QuickSight.

*****Make sure you are in the same region as where you have deployed your other environments*****

Open AWS CloudShell

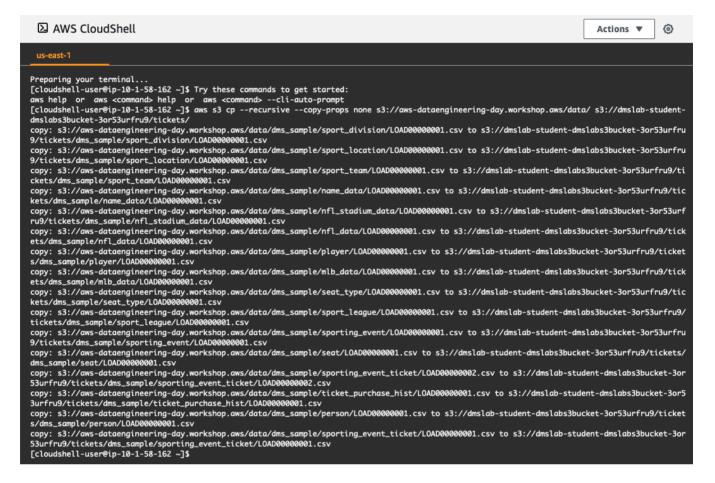Open AWS CloudShell in us-east-1 (N. Virginia) region. It will open a terminal window in the browser. (If there is a pop-up, close it)

Copy Data across from staging Amazon S3 bucket to your S3 bucket

Issue the following command in the terminal, and replace the bucket name with your own one.

```
aws s3 cp --recursive --copy-props none s3://aws-dataengineering-day.workshop.aws/data/
s3://<YourBucketName>/tickets/
```

Lab1. Copy RDS Source Data - Prelab

The data will be copied to your S3 Bucket and you will see the following:

Verify the Data

1. Open the S3 console and view the data that was copied from CloudShell terminal.

2. Your data files in a S3 bucket will look like this:
    BucketName/bucket_folder_name/schema_name/table_name/csv files/

3. In our lab example, this becomes: "/<**BucketName**>/tickets/dms_sample" with a
    separate path for each table_name

Amazon S3 > dmslab-student-dmslabs3bucket-3or53urfru9 > tickets/ > dms_sample/
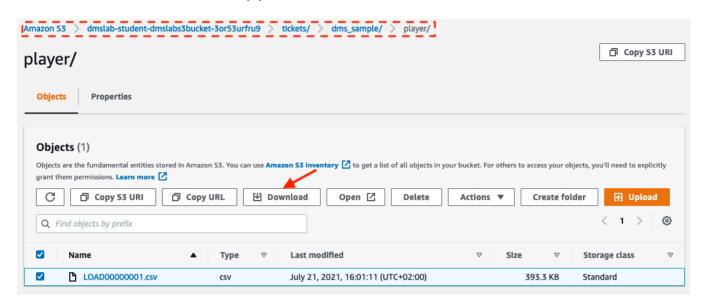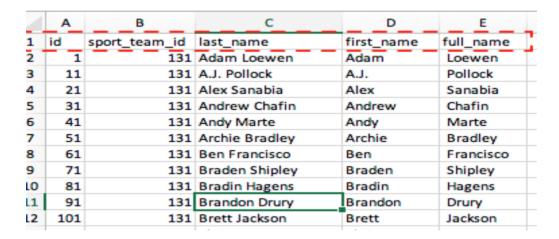
dms_sample/

Copy S3 URI

Objects | Properties

Objects (15)

Objects are the fundamental entities stored in Amazon S3. You can use **Amazon S3 inventory** ☑ to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. **Learn more** ☑

| C | Copy S3 URI | Copy URL | Download | Open ☑ | Delete | Actions ▼ |

| Create folder | ⤓ Upload |

Find objects by prefix

< **1** > ⚙

	Name ▲	Type ▽	Last modified ▽	Size ▽	Storage class ▽
☐	📄 mlb_data/	Folder	-	-	-
☐	📄 name_data/	Folder	-	-	-
☐	📄 nfl_data/	Folder	-	-	-
☐	📄 nfl_stadium_data/	Folder	-	-	-
☐	📄 person/	Folder	-	-	-
☐	📄 player/	Folder	-	-	-
☐	📄 seat_type/	Folder	-	-	-
☐	📄 seat/	Folder	-	-	-
☐	📄 sport_division/	Folder	-	-	-
☐	📄 sport_league/	Folder	-	-	-
☐	📄 sport_location/	Folder	-	-	-
☐	📄 sport_team/	Folder	-	-	-
☐	📄 sporting_event_ticket/	Folder	-	-	-
☐	📄 sporting_event/	Folder	-	-	-
☐	📄 ticket_purchase_hist/	Folder	-	-	-

4. Download one of the files:
    a. Select a table/folder name, tick the check box next to a CSV file name, and choose **Download** option from the **Actions** dropdown list.
    b. Click **Save File**.
    c. Open the file.

Lab1. Copy RDS Source Data - Prelab

Note that column names are included in the file in the first row.

Explore the objects in the S3 directory further.

Next Steps

In the next part of this lab, we will complete the following tasks:
- Extract, Transform and Load Data Lake with AWS Glue

Appendix A: Self-Paced Data Lake Lab

If you If want to re-run the lab by yourself, please follow the lab instruction published in the GitHub:

https://github.com/aws-samples/data-engineering-for-aws-immersion-day