



Amazon Web Services

Lab1. Copy RDS Source Data - Prelab
July 2021

Table of Contents

About the lab setup:	3
Open AWS CloudShell	3
Copy Data across from staging Amazon S3 bucket to your S3 bucket.....	3
Verify the Data.....	4
Next Steps	6
Appendix A: Self-Paced Data Lake Lab.....	6

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:

```

AWS CloudShell
Actions

us-east-1

Preparing your terminal...
[cloudshell-user@ip-10-1-58-162 ~]$ Try these commands to get started:
aws help or aws <command> help or aws <command> --cli-auto-prompt
[cloudshell-user@ip-10-1-58-162 ~]$ aws s3 cp --recursive --copy-props none s3://aws-dataengineering-day.workshop.aws/data/ s3://dmslab-student-
dmslabs3bucket-3or53urfru9/tickets/
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/sport_division/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urfru
9/tickets/dms_sample/sport_division/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/sport_location/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urfru
9/tickets/dms_sample/sport_location/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/sport_team/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urfru9/ti
ckets/dms_sample/sport_team/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/name_data/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urfru9/tic
kets/dms_sample/name_data/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/nfl_stadium_data/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urf
ru9/tickets/dms_sample/nfl_stadium_data/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/nfl_data/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urfru9/tick
ets/dms_sample/nfl_data/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/player/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urfru9/ticket
s/dms_sample/player/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/mlb_data/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urfru9/tick
ets/dms_sample/mlb_data/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/seat_type/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urfru9/tic
kets/dms_sample/seat_type/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/sport_league/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urfru9/
tickets/dms_sample/sport_league/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/sporting_event/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urfru
9/tickets/dms_sample/sporting_event/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/seat/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urfru9/tickets/
dms_sample/seat/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/sporting_event_ticket/LOAD000000002.csv to s3://dmslab-student-dmslabs3bucket-3or
53urfru9/tickets/dms_sample/sporting_event_ticket/LOAD000000002.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/ticket_purchase_hist/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or5
3urfru9/tickets/dms_sample/ticket_purchase_hist/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/person/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or53urfru9/ticket
s/dms_sample/person/LOAD000000001.csv
copy: s3://aws-dataengineering-day.workshop.aws/data/dms_sample/sporting_event_ticket/LOAD000000001.csv to s3://dmslab-student-dmslabs3bucket-3or
53urfru9/tickets/dms_sample/sporting_event_ticket/LOAD000000001.csv
[cloudshell-user@ip-10-1-58-162 ~]$
```

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

Lab1. Copy RDS Source Data - Prelab

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](#)

Refresh Copy S3 URI Copy URL Download Open Delete Actions

Create folder Upload

< 1 > Settings

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	mlb_data/	Folder	-	-	-
<input type="checkbox"/>	name_data/	Folder	-	-	-
<input type="checkbox"/>	nfl_data/	Folder	-	-	-
<input type="checkbox"/>	nfl_stadium_data/	Folder	-	-	-
<input type="checkbox"/>	person/	Folder	-	-	-
<input type="checkbox"/>	player/	Folder	-	-	-
<input type="checkbox"/>	seat_type/	Folder	-	-	-
<input type="checkbox"/>	seat/	Folder	-	-	-
<input type="checkbox"/>	sport_division/	Folder	-	-	-
<input type="checkbox"/>	sport_league/	Folder	-	-	-
<input type="checkbox"/>	sport_location/	Folder	-	-	-
<input type="checkbox"/>	sport_team/	Folder	-	-	-
<input type="checkbox"/>	sporting_event_ticket/	Folder	-	-	-
<input type="checkbox"/>	sporting_event/	Folder	-	-	-
<input type="checkbox"/>	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

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

player/ Copy S3 URI

Objects Properties

Objects (1)

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](#)

Refresh Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

Find objects by prefix

<input checked="" type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input checked="" type="checkbox"/>	LOAD00000001.csv	csv	July 21, 2021, 16:01:11 (UTC+02:00)	393.3 KB	Standard

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

	A	B	C	D	E
1	id	sport_team_id	last_name	first_name	full_name
2	1	131	Adam Loewen	Adam	Loewen
3	11	131	A.J. Pollock	A.J.	Pollock
4	21	131	Alex Sanabia	Alex	Sanabia
5	31	131	Andrew Chafin	Andrew	Chafin
6	41	131	Andy Marte	Andy	Marte
7	51	131	Archie Bradley	Archie	Bradley
8	61	131	Ben Francisco	Ben	Francisco
9	71	131	Braden Shipley	Braden	Shipley
10	81	131	Bradin Hagens	Bradin	Hagens
11	91	131	Brandon Drury	Brandon	Drury
12	101	131	Brett Jackson	Brett	Jackson

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 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>