

NIT2202 Lab 1+2+3 Assessment

Assessment Requirements:

You are given time **until Friday 11:30pm** to complete Lab 1, Lab 2 and Lab 3 on **AWS Academy Data Analytics platform** **independently** without assistance from other people including your Lab Tutor and classmates:

Create a new word document called **Lab_1+2+3_YourName_YourID.docx** where YourName is your real name and YourID is your VU Student ID.

Follow the instructions, capture **FULL** query result screens which must contain six elements: (1) URL (2) User Identifier (3) Resources (4) Query (5) Query results (6)

Taskbar date and time as shown below and label each screen with its corresponding lab number and step or question number underneath.

The screenshot displays the AWS S3 console interface. The address bar shows the URL: `https://s3.console.aws.amazon.com/s3/buckets/mdflurrybucket/object/select?region=us-east-1&prefix=lab1.csv`. The top right corner shows the user identifier: `awsuser @ 0957-8092-8538`. The left sidebar shows the Amazon S3 logo and various navigation options. The main content area displays a SQL query: `SELECT * FROM s3object s LIMIT 5`. Below the query, the query results are shown, indicating that 5 records were returned successfully. The taskbar at the bottom shows the date and time: 4:44 PM, 30/08/2021.

Lab 1 Step 43 query results

Late Submission Penalty

- 1-3 days – 30% each day
- After 3 days – 0 (zero) mark

Deduction

- Missing Cover Sheet or Wrong file name – 10%
- Combined cover sheet file with Lab Assessment file – 10%
- Missing figure numbers/captions – 10%
- Captured screens without 6 elements – up to 50 %
- Missing “Complete All items” – up to 50%

Step 43. Choose Run SQL query

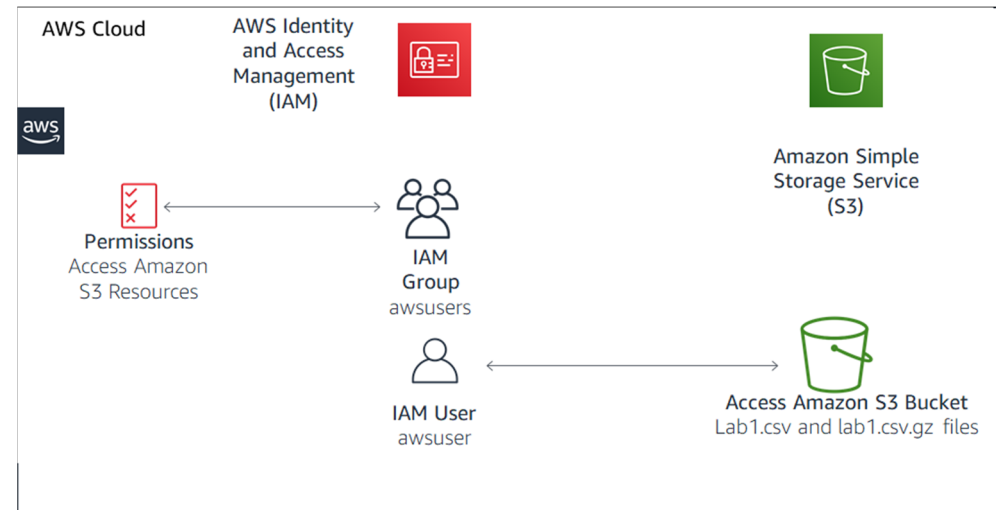
In the Result pane, you should get the total number of records, which is 5.

Step 58. Scroll down the page and choose Run SQL query

You should get results that demonstrate that you can query the compressed file in the same way as a non-compressed file.

Challenge: Run the following SQL query and explain the query and results:

```
SELECT COUNT (s._1) FROM s3Object s WHERE s._1 = '002';
```



Lab 1 Architecture

Lab 01

Step 31. Choose the vertical ellipsis (three dots) next to the Taxidata table, and from the list of options, choose **Preview table**.

In the **Results** window, you will see the first 10 records of the table.

Lab 02

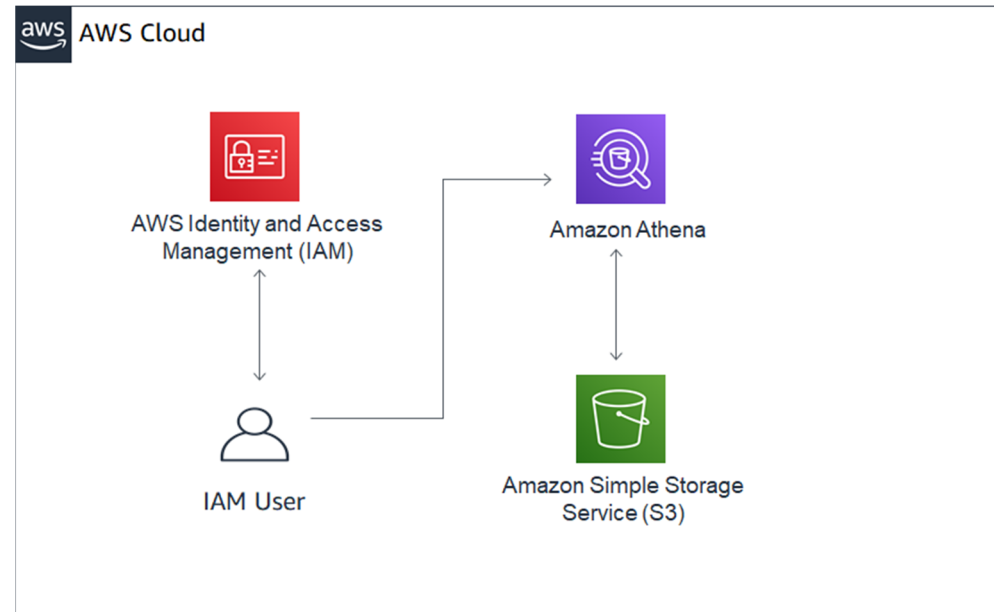
Step 41. Complete all the queries until you create a view to join data from two tables. If you preview the *comparepay* view, you should see the following results:

Vendor 1: cctotal 584502884 cashtotal 250849783

Vendor 2: cctotal 460097126 cashtotal 199181978

Challenge Two

Write a query to compare the average distance for trips that were paid with credit cards and the average distance for trips that were paid with cash in January 2017.



Lab 2 Architecture

Step 56. Choose Run query

Step 62. Choose Run query

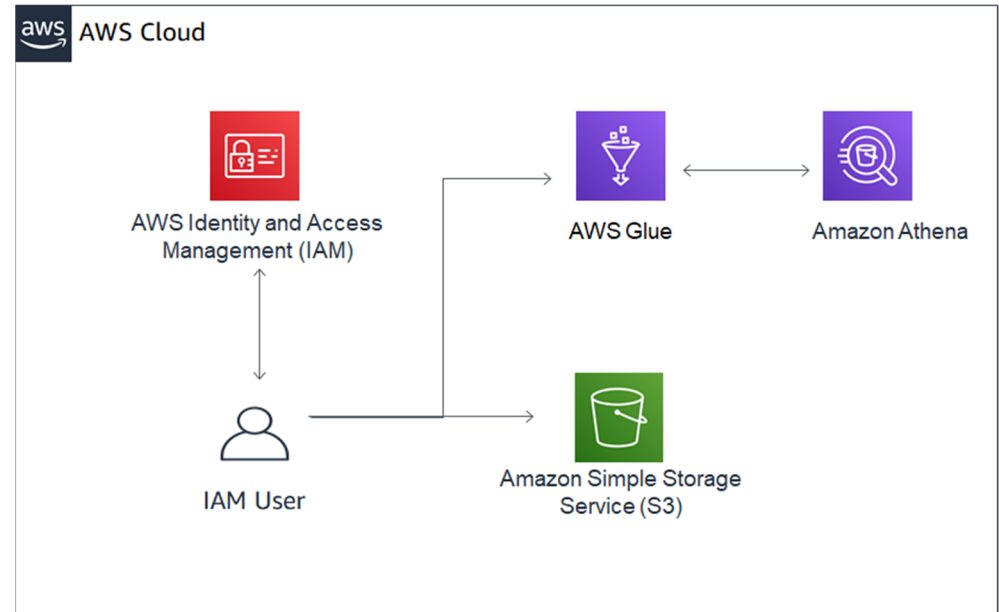
You should see a table of data results from 1950–2018, with the average maximum temperature for each year.

Challenge question

Lab 03

For this challenge, do the following tasks:

- Use AWS Glue to create a table for the weather stations.
- Write a query in Athena to count the number of stations that are not in the US or Canada. The first two characters of the station ID field indicate the country where the station is located. The country codes for the United States is *US* and the country code for Canada is *CA*.



Lab 3 Architecture

Upon completion, check AWS Academy Data Analytics if “Complete All items” are ticked:

▸ Lab 1	Complete All Items	✓
▸ Lab 2	Complete All Items	✓
▸ Lab 3	Complete All Items	✓

There is a deduction for missing “Complete All items”

and then upload the following **two files** via dropbox on VU Collaborate:

- **CoES Assignment Cover Sheet** and
- **Lab_1+2+3_YourName_YourID.docx**