**Lab Exercise- Basics of Setting Up and Using Athena**

Athena is a serverless query service provided by AWS that allows you to analyze data in Amazon S3 using standard SQL. It’s especially useful for analyzing large datasets without the need for managing any infrastructure. Below is a lab exercise that will guide you through the basics of setting up and using Amazon Athena.

**Objective:**

Learn how to set up and use Amazon Athena to query data stored in Amazon S3 using SQL.

**Prerequisites:**

* AWS Account
* Basic knowledge of SQL
* Some data stored in an Amazon S3 bucket (CSV, JSON, Parquet, etc.)

**Step 1: Set Up an S3 Bucket and Upload Data**

**Create an S3 Bucket:**

* Go to the AWS Management Console.
* Navigate to S3 and click on Create bucket.
* Enter a unique bucket name and select a region.
* Click Create bucket.

**Upload Data to the S3 Bucket:**

* Prepare a CSV file (e.g., sample-data.csv) with some sample data. For example:

id,name,age,city

1,John Doe,28,New York

2,Jane Smith,34,Chicago

3,Bob Johnson,45,Los Angeles

* Upload this CSV file to the S3 bucket you created.

**Step 2: Set Up Athena**

**Navigate to Amazon Athena:**

* Go to the AWS Management Console.
* Navigate to Athena under the Analytics section.

**Set Up a Query Result Location:**

* In Athena, click on Settings (top right corner).
* Set the location for query results to your S3 bucket (e.g., s3://your-bucket-name/athena-results/).
* Click Save.

**Step 3: Create a Database and Table in Athena**

**Create a Database:**

In the Athena query editor, run the following SQL command to create a database:

CREATE DATABASE sample\_db;

Select Run Query. The database sample\_db should now be created.

**Create a Table:**

Now, create a table that points to your data in S3:

CREATE EXTERNAL TABLE IF NOT EXISTS sample\_db.sample\_table (

id INT,

name STRING,

age INT,

city STRING

)

ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe'

WITH SERDEPROPERTIES (

'serialization.format' = ','

)

LOCATION 's3://your-bucket-name/';

* Replace your-bucket-name with the actual name of your S3 bucket.
* Click Run Query. The table sample\_table should now be created and linked to your CSV data.

**Step 4: Query Data with Athena**

Basic Select Query:

Run a simple SQL query to retrieve all data from the table:

SELECT \* FROM sample\_db.sample\_table;

Click Run Query to see the results.

**Filtering Data:**

Run a query to filter data, such as finding all individuals from New York:

SELECT \* FROM sample\_db.sample\_table WHERE city = 'New York';

Click Run Query to see the filtered results.

**Aggregation Example:**

Run a query to find the average age of individuals in the table:

SELECT AVG(age) as average\_age FROM sample\_db.sample\_table;

Click Run Query to see the result.

**Step 5: Clean Up**

Drop the Table and Database:

To clean up, you can drop the table and database:

DROP TABLE sample\_db.sample\_table;

DROP DATABASE sample\_db;