

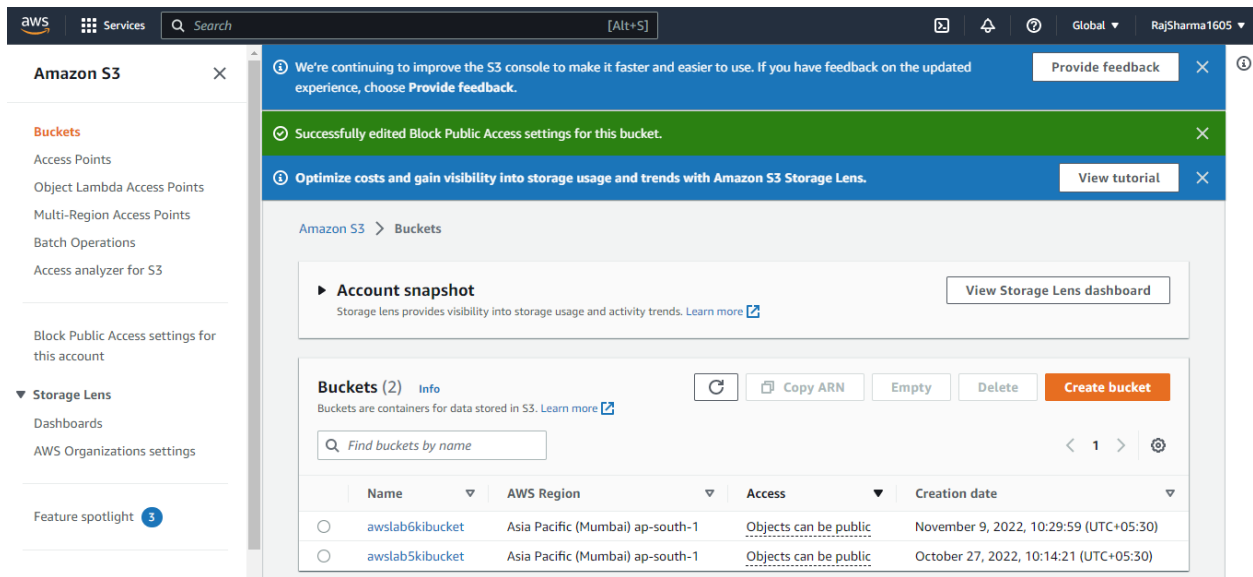
EXP NO : 6

Submitted by : Raj Sharma (RA2011028010055)

Aim: Querying Data in S3 with Amazon Athena

Step 1:

Go to buckets and create two buckets.



Step 2 :

After clicking onto the bucket add files to it.

aws Services Search [Alt+S] Global RajSharma1605

We're continuing to improve the S3 console to make it faster and easier to use. If you have feedback on the updated experience, choose [Provide feedback](#).

Upload succeeded
View details below.

Destination	Succeeded	Failed
s3://awslab6kibucket	2 files, 482.1 KB (100.00%)	0 files, 0 B (0%)

Files and folders Configuration

Files and folders (2 Total, 482.1 KB)

Find by name

Name	Folder	Type	Size	Status	Error
Screenshot (1).png	-	image/png	356.5 KB	Succeeded	-
Screenshot (2).png	-	image/png	125.5 KB	Succeeded	-

Step 3 :
Now go to Amazon athena.

aws Services Search [Alt+S] Mumbai RajSharma1605

Analytics

Amazon Athena

Start querying data instantly.

Amazon Athena is an interactive query service that makes it easy to analyze data in Amazon S3 and other federated data sources using standard SQL.

Begin querying your data

Discover the query editor and start querying right away.

[Explore the query editor](#)

Pricing

Mumbai - Data scanned	\$5 per TB
-----------------------	------------

[Cost calculator](#)

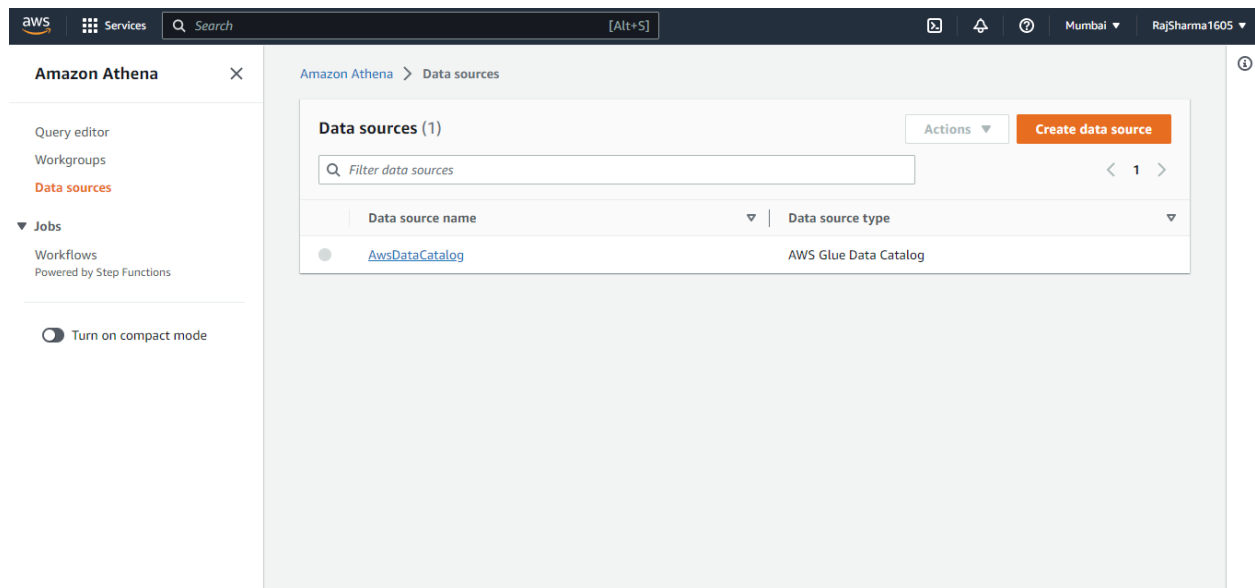
Getting started

Add the data source and identify the table and column names. Then you can begin querying your data right away.

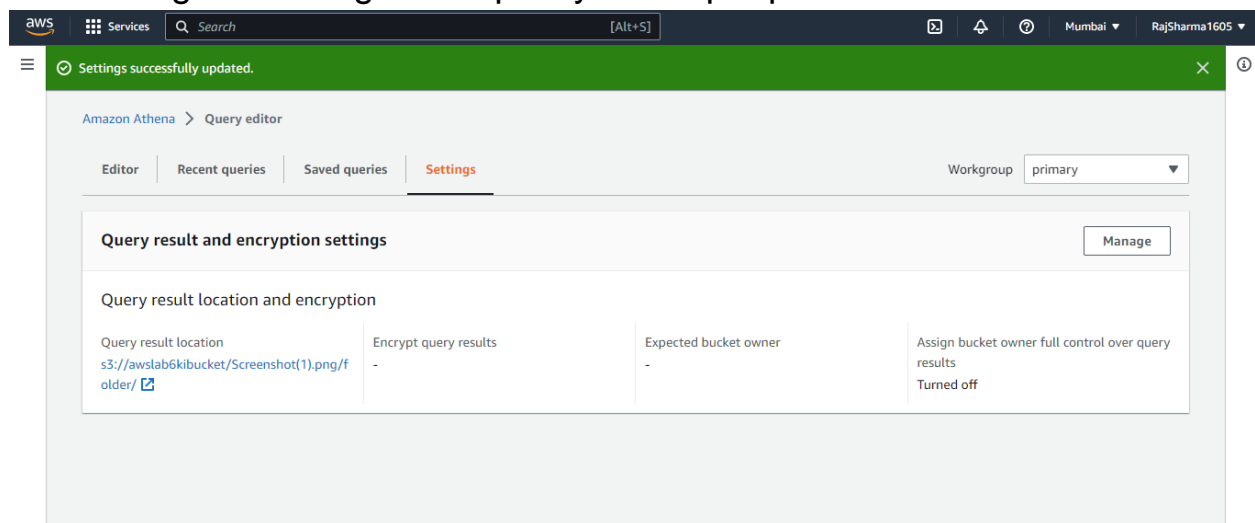
How it works

1. Point to your data source
Register and select any data source.

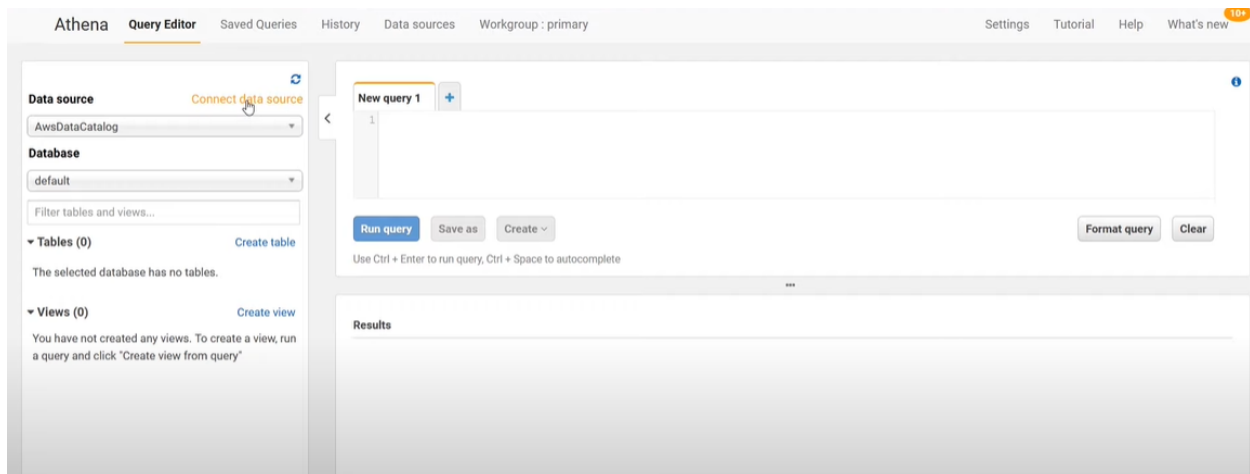
Step 4 :
Select AwsDataCatalog in the left side which is present in the data source tab.



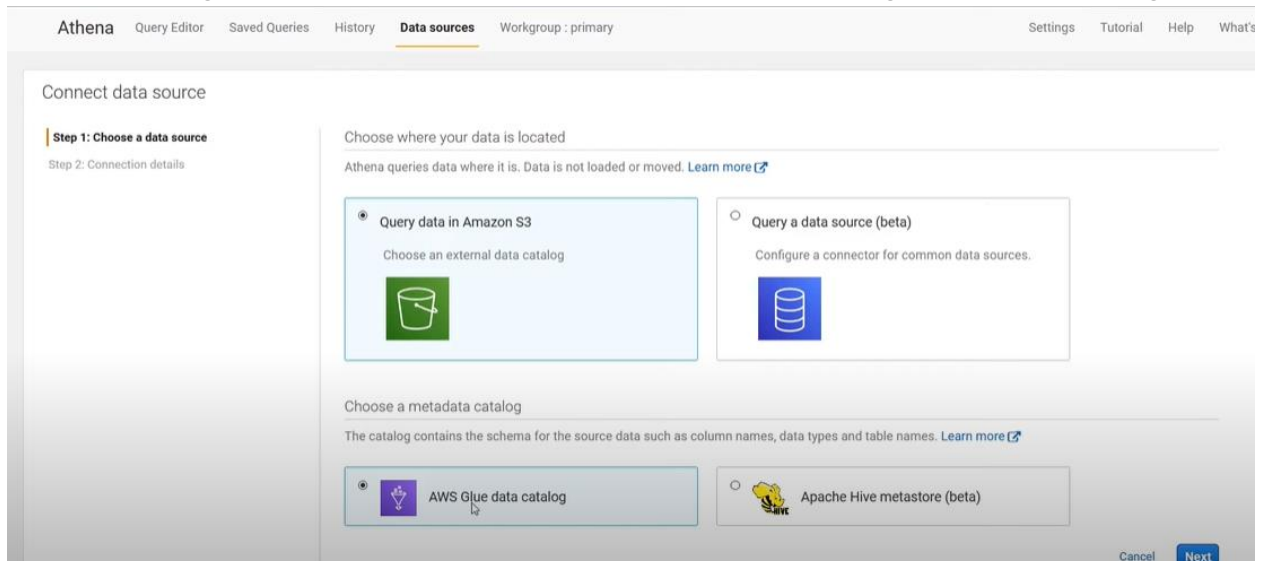
Step 5 :
After that go to settings and specify an output path.



Step 6:
Click on connect data source.



Step 7:
After clicking choose a query in amazon s3 and Aws glue data catalog.



Step 8:
Click on next and select setup a crawler in AWS glue to retrieve schema information automatically.

Athena Query Editor Saved Queries History **Data sources** Workgroup : primary Settings Tutorial Help What's new 10+

Connect data source

Step 1: Choose a data source
Step 2: Connection details

Connection details: AWS Glue data catalog

Athena will connect to your data stored in Amazon S3 and use AWS Glue data catalog to store metadata, such as table and column names. Once connected, your databases, tables and views appear in Athena's query editor. [Learn more](#)

☒ Set up crawler in AWS Glue to retrieve schema information automatically
☐ Add a table and enter schema information manually

Cancel Previous **Connect to AWS Glue**

Step 9:

After selecting that it will redirect to a new page and add crawler and follow below steps to add a new crawler after setting up click on finish.

Add crawler

Add information about your crawler

☒ Crawler info
☐ Crawler source type
☐ Data store
☐ IAM Role
☐ Schedule
☐ Output
☐ Review all steps

Crawler name
aws-athena-demo

Tags, description, security configuration, and classifiers (optional)

Next

Add crawler

Specify crawler source type

Choose Existing catalog tables to specify catalog tables as the crawler source. The selected tables specify the data stores to crawl. This option doesn't support JDBC data stores.

Crawler source type
☒ Data stores
☐ Existing catalog tables

Back **Next**

Add crawler

Crawler info

aws-athena-demo

Crawler source type

Data stores

Data store

IAM Role

Schedule

Output

Review all steps

Add a data store

Choose a data store

s3

Crawl data in

☒ Specified path in my account

☐ Specified path in another account

Include path

s3://bucket/prefix/object

Exclude patterns (optional)

Back

Next

Add crawler

Crawler info

aws-athena-demo

Crawler source type

Data stores

Data store

IAM Role

Schedule

Output

Review all steps

Choose an IAM role

The IAM role allows the crawler to run and access your Amazon S3 data stores. [Learn more](#)

☐ Update a policy in an IAM role

☐ Choose an existing IAM role

☒ Create an IAM role

IAM role ⓘ

AWSGlueServiceRole- demo

To create an IAM role, you must have **CreateRole**, **CreatePolicy**, and **AttachRolePolicy** permissions.

Create an IAM role named **"AWSGlueServiceRole-rolename"** and attach the AWS managed policy, **AWSGlueServiceRole**, plus an inline policy that allows read access to:

- s3://aws-simplified-athena-demo/

You can also create an IAM role on the [IAM console](#).

Back

Next

Crawler info

aws-athena-demo

Crawler source type

Data stores

Data store

IAM Role

Schedule

Output

Review all steps

Create a schedule for this crawler

Frequency

Run on demand

Back

Next

Add crawler

Crawler info

aws-athena-demo

Crawler source type

Data stores

Data store

S3: s3://aws-simplifi...

IAM Role

arn:aws:iam::398447858632:role/service-role/AWSGlueServiceRole-demo

Schedule

Run on demand

Output

default

Review all steps

Crawler info

Name

aws-athena-demo

Tags

-

IAM role

arn:aws:iam::398447858632:role/service-role/AWSGlueServiceRole-demo

Schedule

Run on demand

Output

Database

default

Prefix added to tables (optional)

Create a single schema for each S3 path

false

Step 10:

Crawler is successfully created and now click on the crawler and click run crawler.

AWS Glue

Data catalog

Databases

Tables

Connections

Crawlers

Classifiers

Settings

ETL

Workflows

Jobs

ML Transforms

Triggers

Dev endpoints

Notebooks

Crawlers

A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.

Attempting to run crawler "aws-athena-demo"...

Add crawler

Run crawler

Action

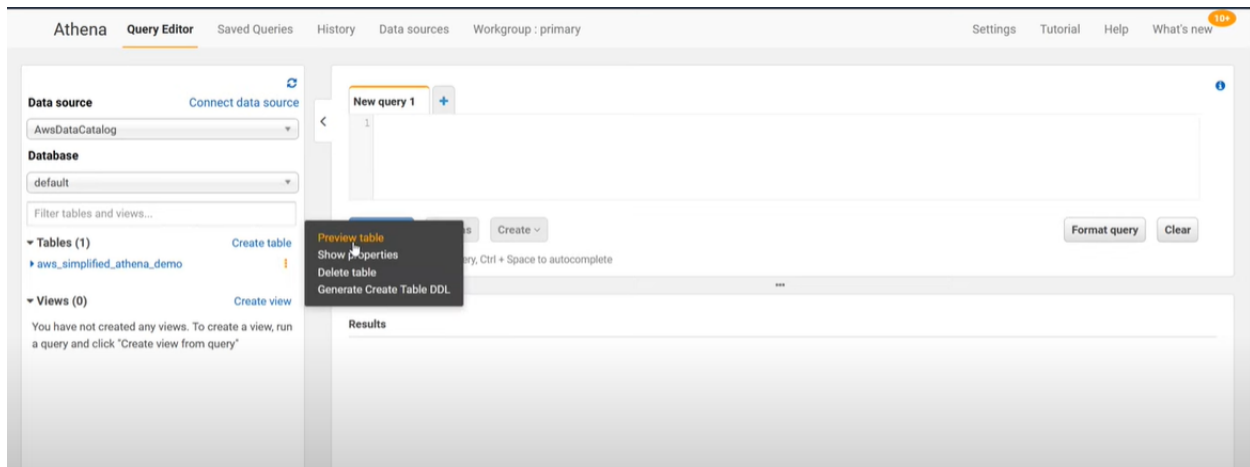
Filter by tags and attributes

Showing: 1 - 1

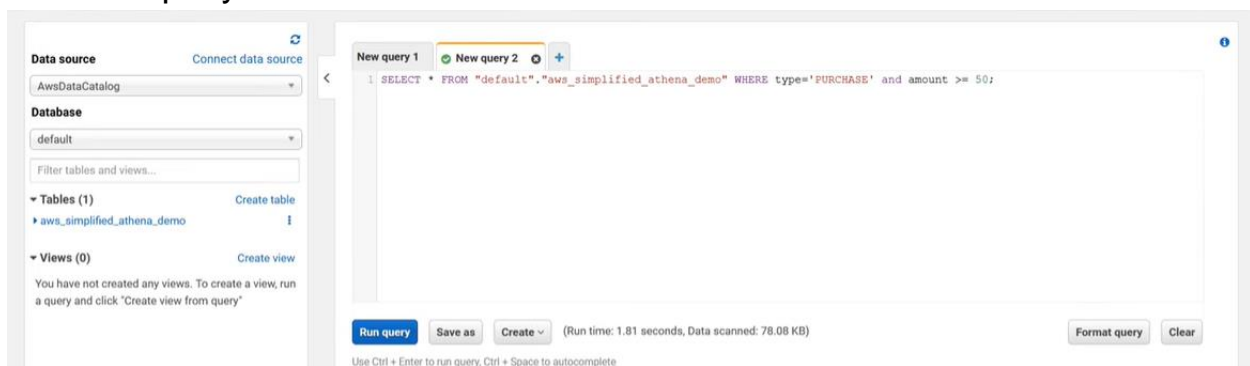
	Name	Schedule	Status	Logs	Last runtime	Median runtime	Tables updated	Tables added
<input checked="" type="checkbox"/>	aws-athena-demo		Ready		0 secs	0 secs	0	0

Step 11:

After running the crawler go back to athena you will see a table created on table column select that and click on preview table.



Step 12:
Now the query can be executed.



Result :

Querying Data in S3 with Amazon Athena is done and output is verified.