

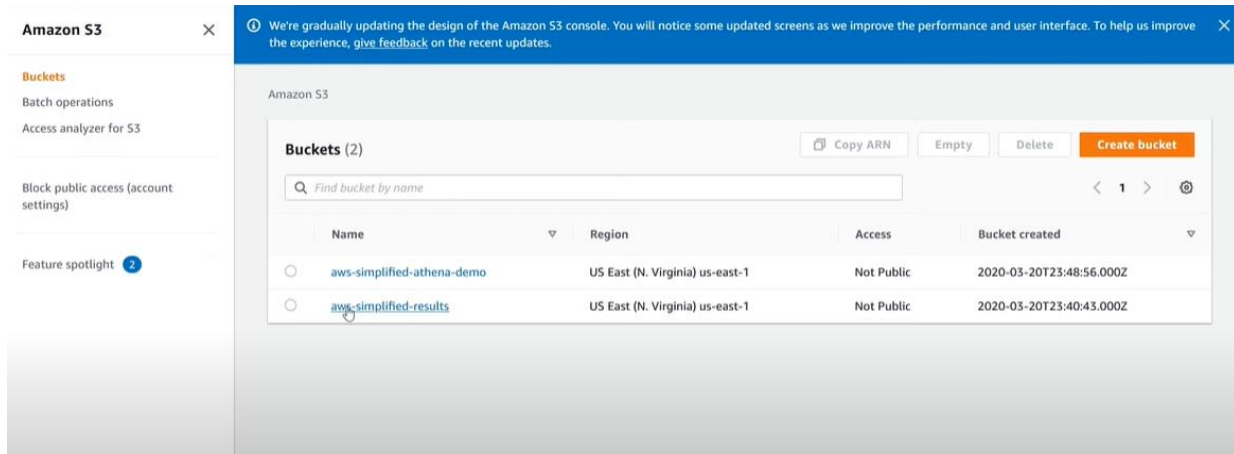
EXP NO : 6

Querying Data in S3 with Amazon Athena

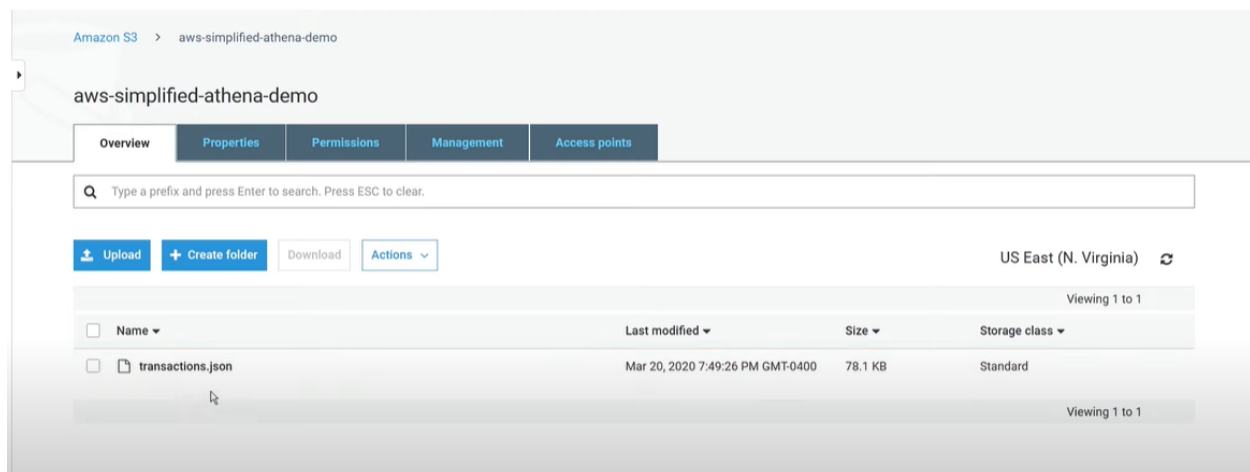
Chaluvadi Jwala Satya Saketh
RA2011028010071

Aim: Querying Data in S3 with Amazon Athena

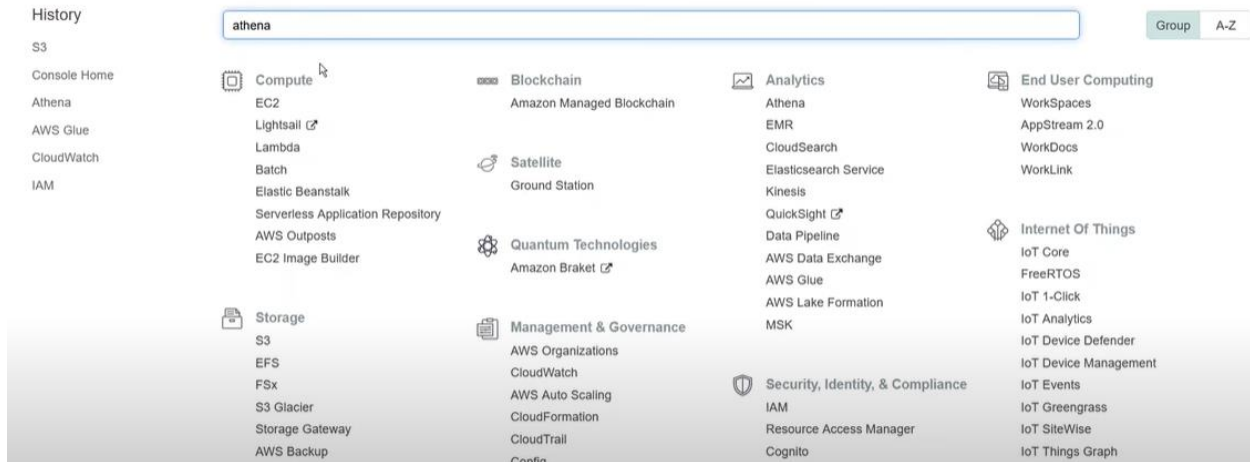
Step 1:
Go to buckets and create two buckets.



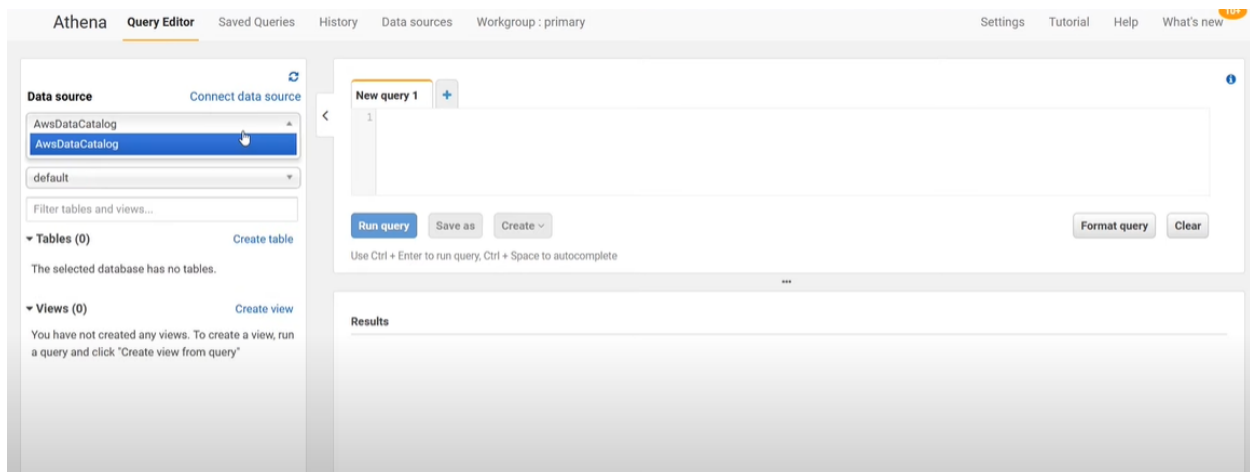
Step 2 :
After clicking onto the bucket add files to it.



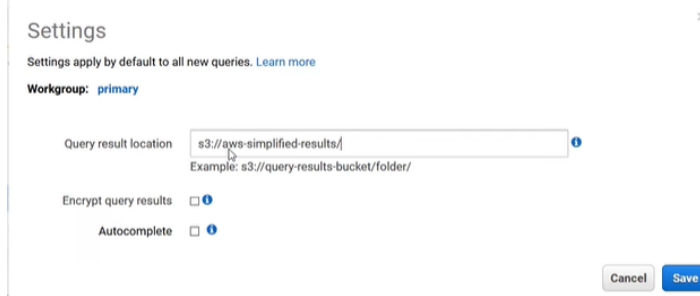
Step 3 :
Now go to Amazon athena.



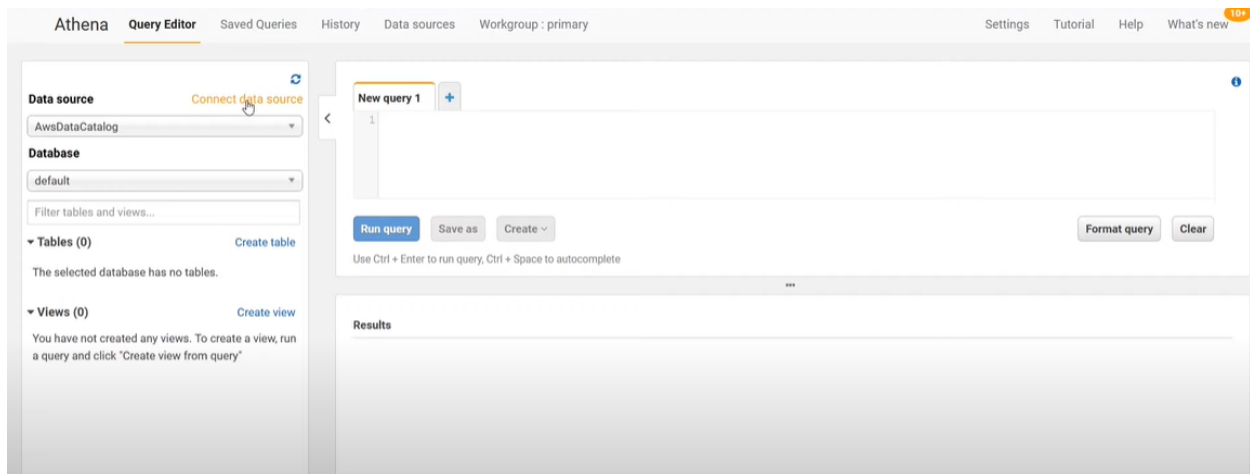
Step 4 :
Select AwsDataCatalog in the left side which is present in the data source tab.



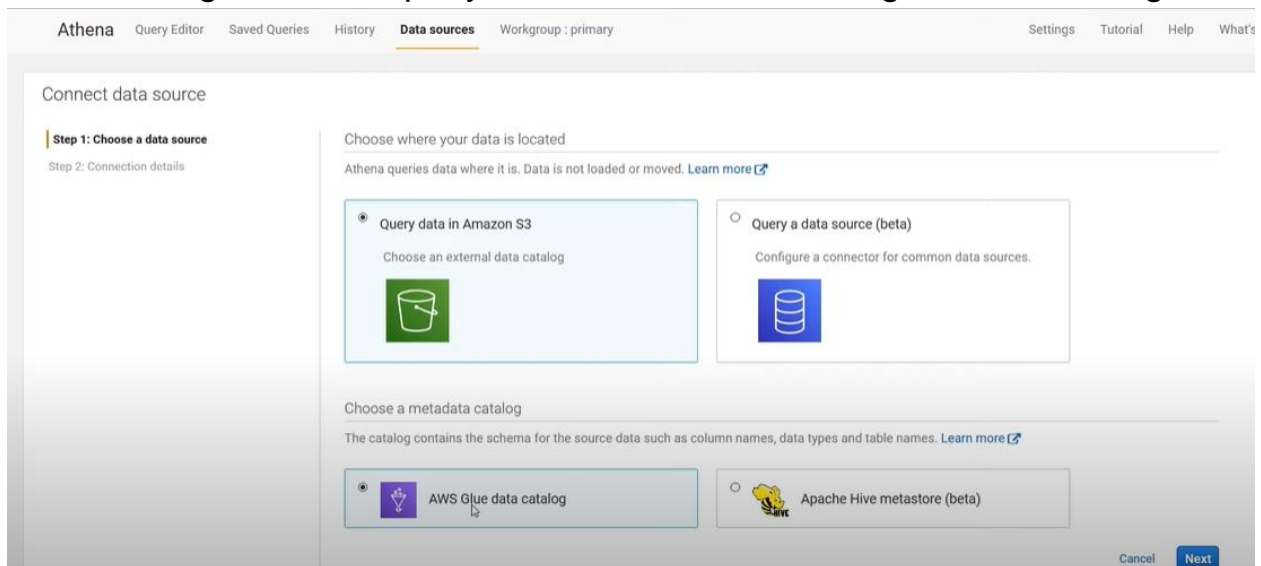
Step 5 :
After that go to settings and specify an output path.



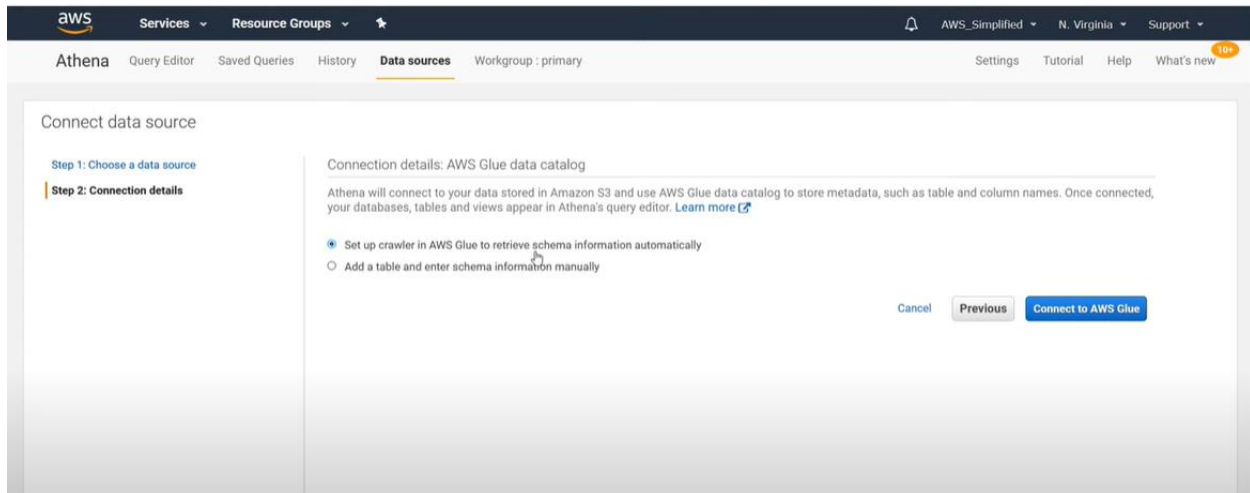
Step 6:
Click on connect data source.



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.



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.

This screenshot shows the 'Add crawler' wizard in the AWS console. The title bar says 'Add crawler'. On the left, there's a sidebar with a list of steps: 'Crawler info' (selected with a green checkmark), 'Crawler source type', 'Data store', 'IAM Role', 'Schedule', 'Output', and 'Review all steps'. The main content area is titled 'Add information about your crawler'. It contains a text input field for 'Crawler name' with the value 'aws-athena-demo'. Below the input field, there's a link that says 'Tags, description, security configuration, and classifiers (optional)'. At the bottom right of the main content area, there is a blue 'Next' button.

This screenshot shows the second step of the 'Add crawler' wizard. The title bar says 'Add crawler'. The sidebar on the left now shows 'Crawler info' as completed and 'Crawler source type' as the current step. The main content area is titled 'Specify crawler source type'. It includes a text block explaining that the user should choose existing catalog tables to specify catalog tables as the crawler source. Below this, there are two radio button options for 'Crawler source type': 'Data stores' (which is selected) and 'Existing catalog tables'. At the bottom right, there are two buttons: 'Back' and '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

demo

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

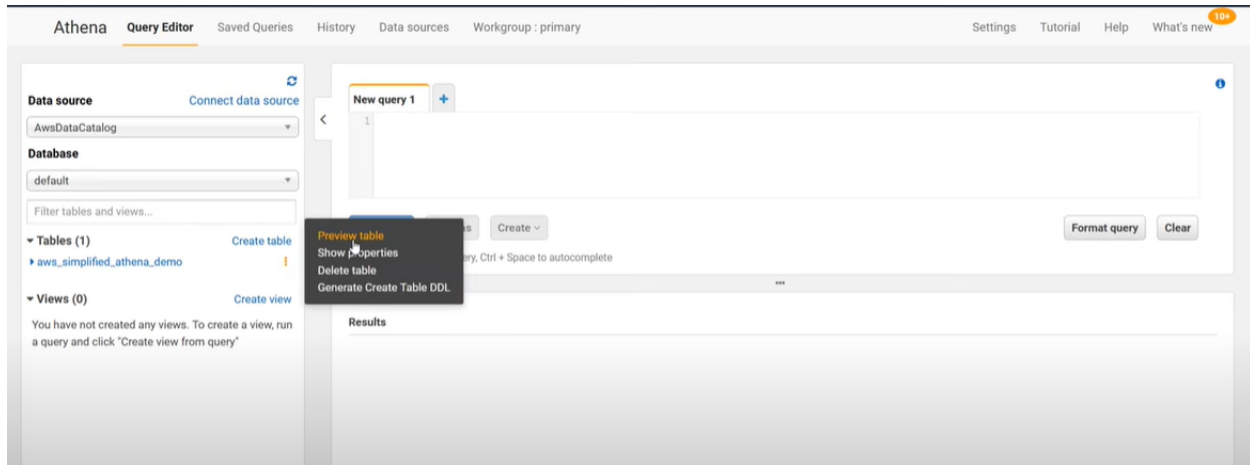
Filter by tags and attributes

Showing: 1 - 1

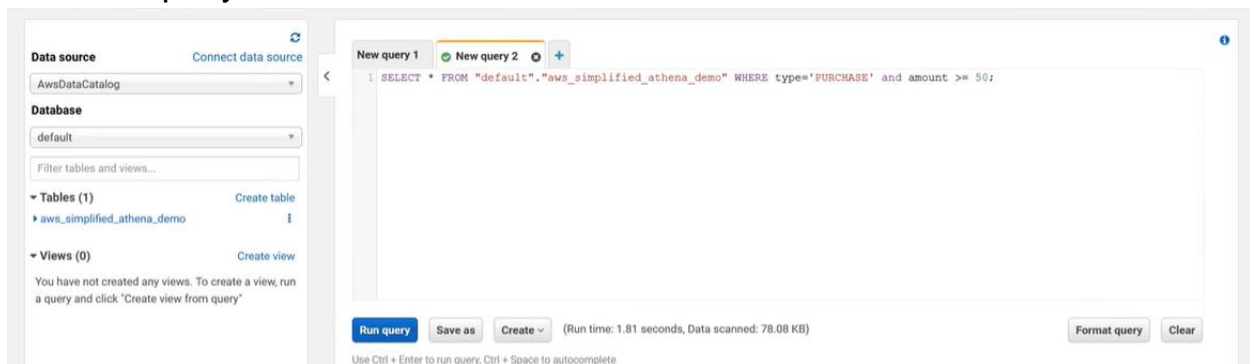
<input checked="" type="checkbox"/>	Name	Schedule	Status	Logs	Last runtime	Median runtime	Tables updated	Tables added
<input checked="" type="checkbox"/>	aws-athena-demo		Ready		0 secs	0 secs	0	0

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.