

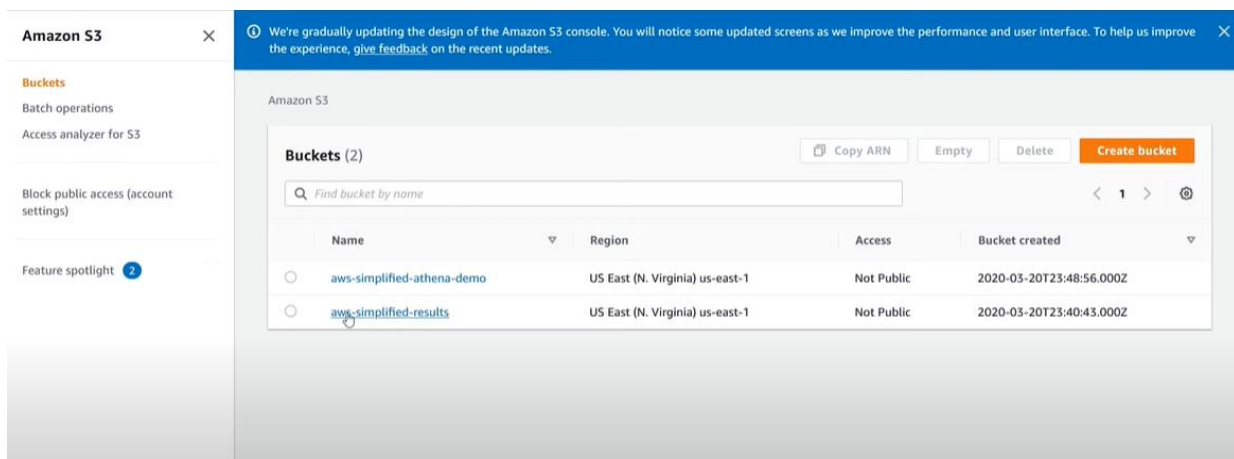
EXPERIMENT 6

Querying Data in S3 with Amazon Athena

Aim: Querying Data in S3 with Amazon Athena

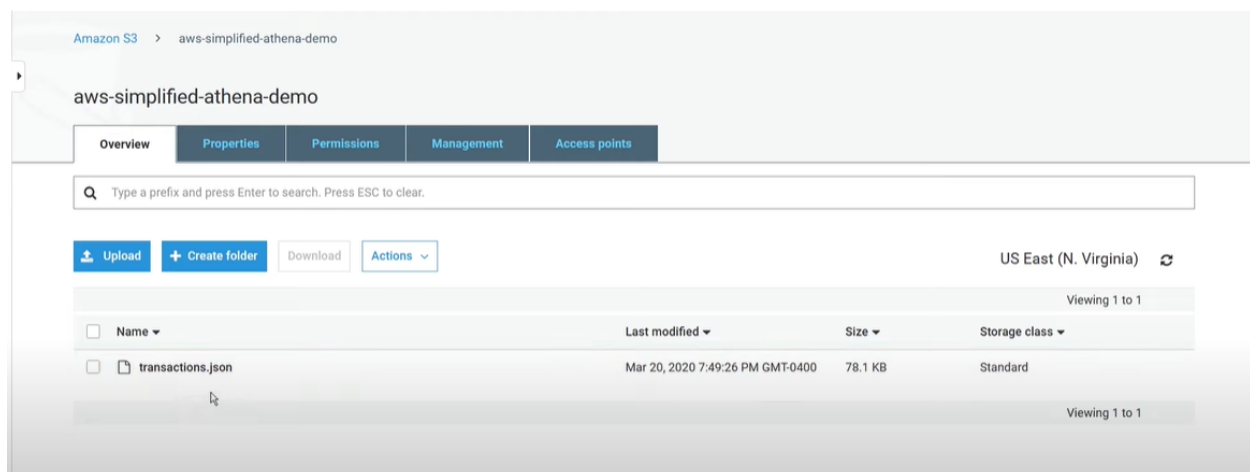
Step 1:

Go to buckets and create two buckets.



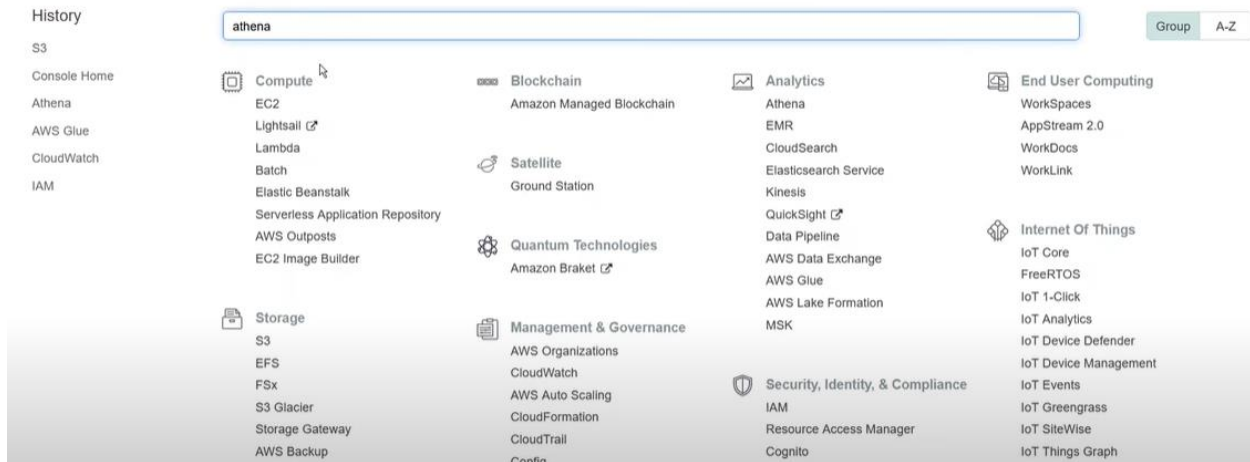
Step 2:

After clicking onto the bucket add files to it.



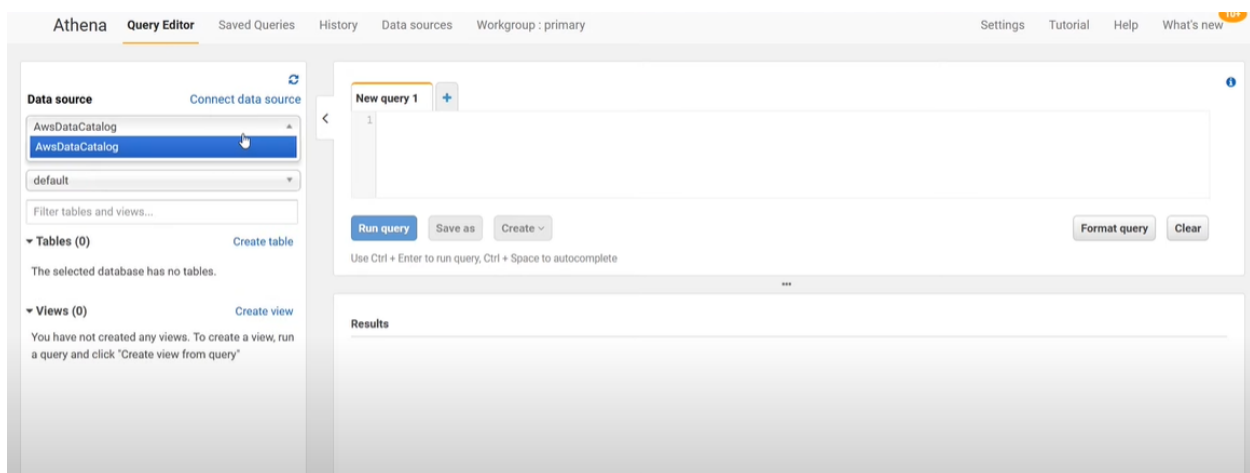
Step 3:

Now go to Amazon Athena.



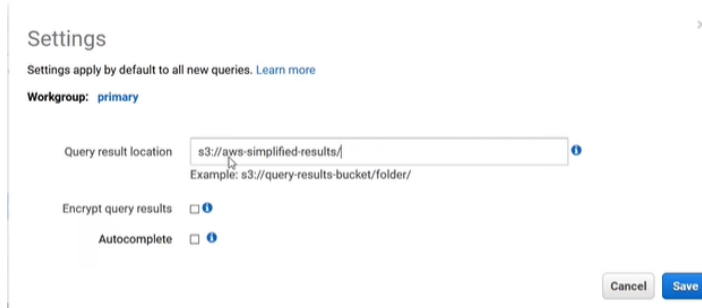
Step 4:

Select Aws Data Catalog in the left side which is present in the data source tab.



Step 5:

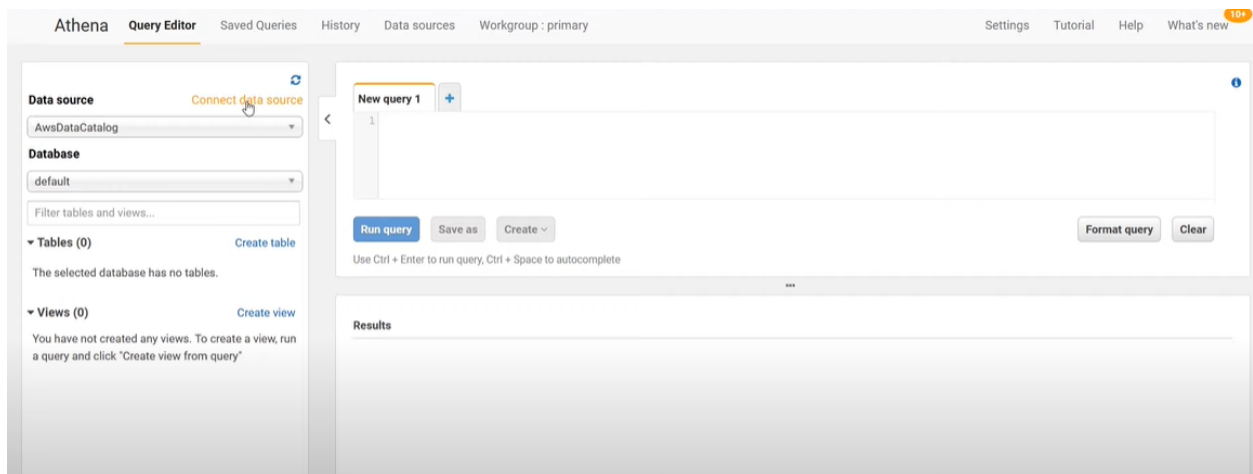
After that go to settings and specify an output path.



The screenshot shows a 'Settings' dialog box with a close button (X) in the top right corner. Below the title, it says 'Settings apply by default to all new queries. [Learn more](#)'. The 'Workgroup' is set to 'primary'. The 'Query result location' is a text input field containing 's3://aws-simplified-results/' with an information icon (i) to its right. Below this field, an example path is shown: 'Example: s3://query-results-bucket/folder/'. There are two checkboxes: 'Encrypt query results' and 'Autocomplete', both of which are currently unchecked and have information icons (i) to their right. At the bottom right, there are 'Cancel' and 'Save' buttons.

Step 6:

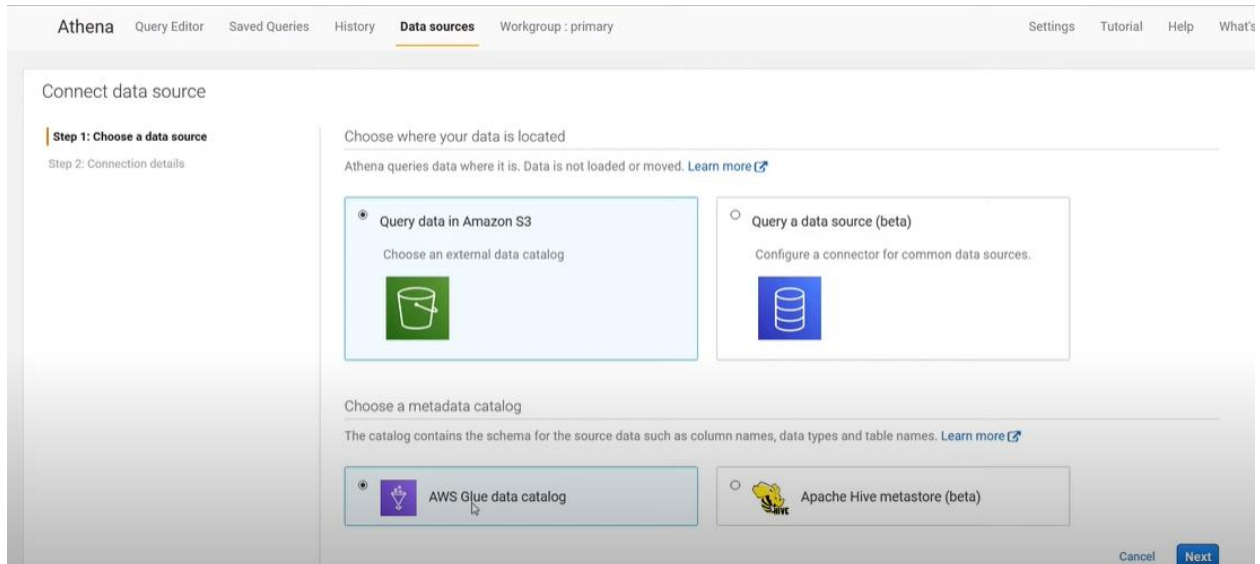
Click on connect data source.



The screenshot shows the Athena Query Editor interface. The top navigation bar includes 'Athena', 'Query Editor' (active), 'Saved Queries', 'History', 'Data sources', and 'Workgroup : primary'. On the right side of the top bar are links for 'Settings', 'Tutorial', 'Help', and 'What's new' with a '100%' zoom indicator. The left sidebar contains a 'Data source' section with a 'Connect data source' link and a dropdown menu showing 'AwsDataCatalog'. Below this is a 'Database' dropdown set to 'default' and a 'Filter tables and views...' input field. The 'Tables (0)' section shows 'The selected database has no tables.' and a 'Create table' link. The 'Views (0)' section shows 'You have not created any views. To create a view, run a query and click "Create view from query"' and a 'Create view' link. The main area is titled 'New query 1' and contains a large text input field for the query. Below the input field are buttons for 'Run query', 'Save as', and 'Create'. At the bottom of the main area are buttons for 'Format query' and 'Clear'. A 'Results' section is visible at the very bottom.

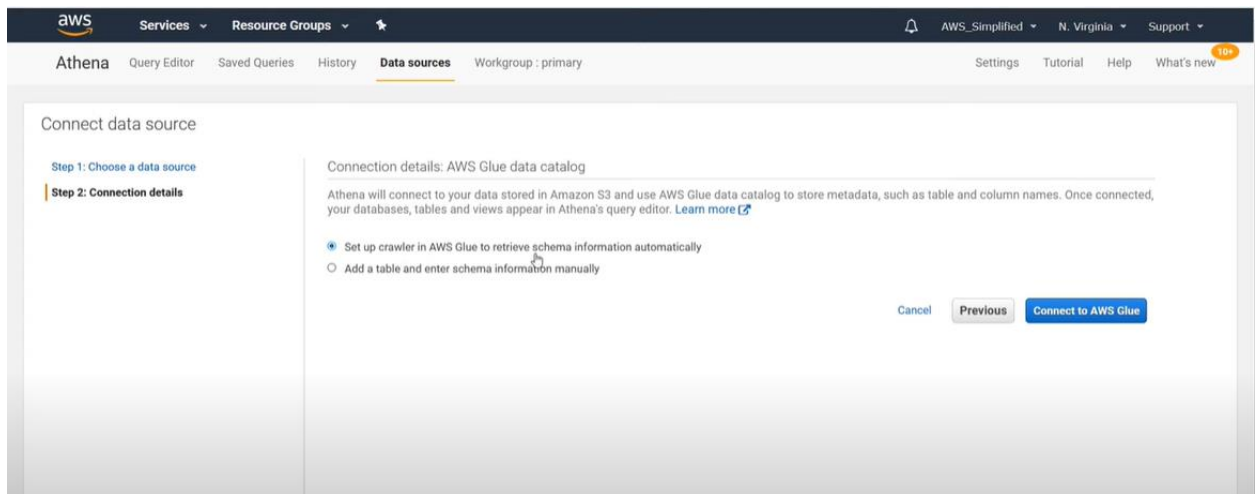
Step 7:

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.

Add crawler

Crawler info

Crawler source type

Data store

IAM Role

Schedule

Output

Review all steps

Add information about your crawler

Crawler name

aws-athena-demo

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

Next

Add crawler

✓ Crawler info

aws-athena-demo

Crawler source type

Data store

IAM Role

Schedule

Output

Review all steps

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

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

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

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

IAM Role

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

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

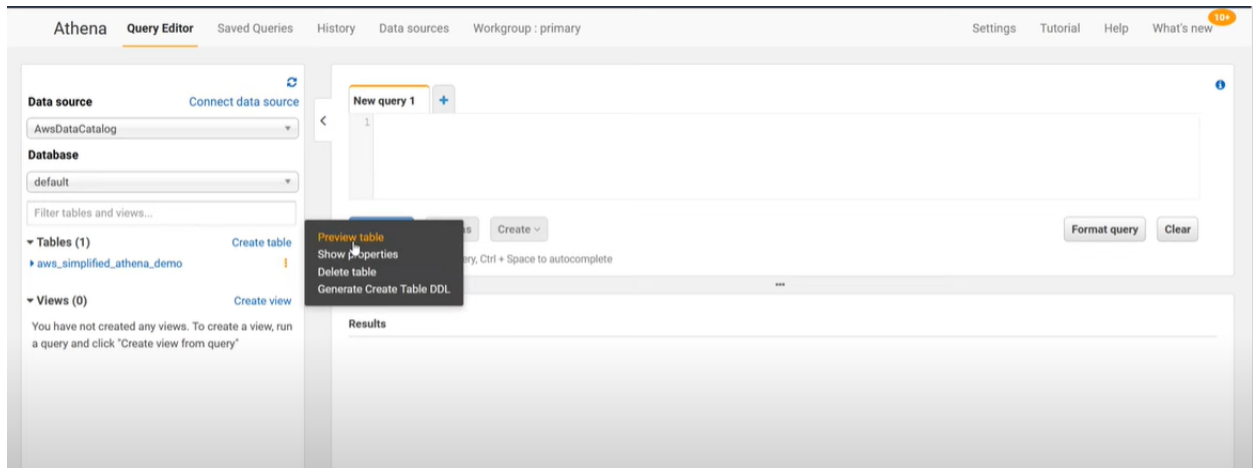
Filter by tags and attributes

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.

