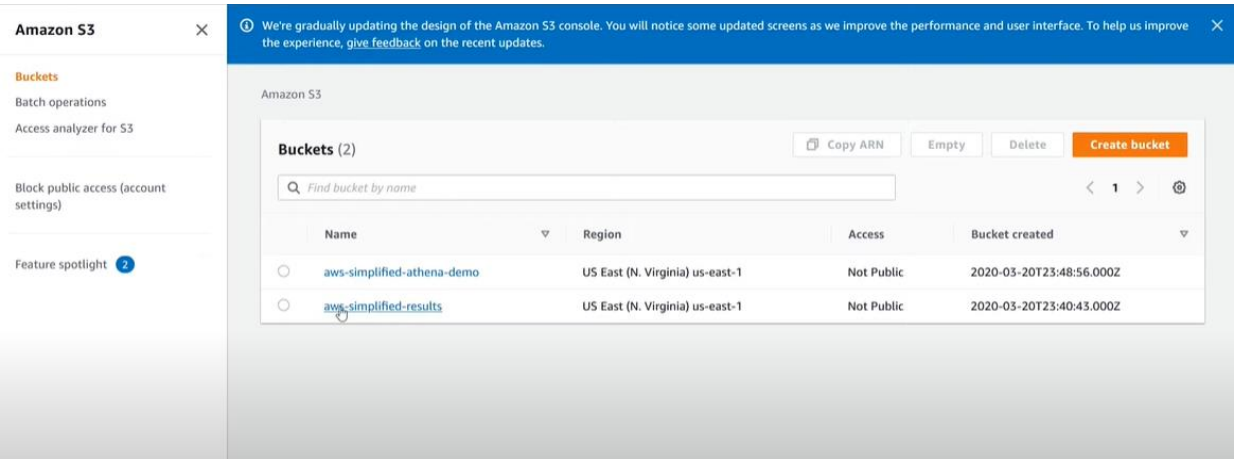
EXPERIMENT 6

Querying Data in S3 with Amazon Athena

Aim: Querying Data in S3 with Amazon Athena

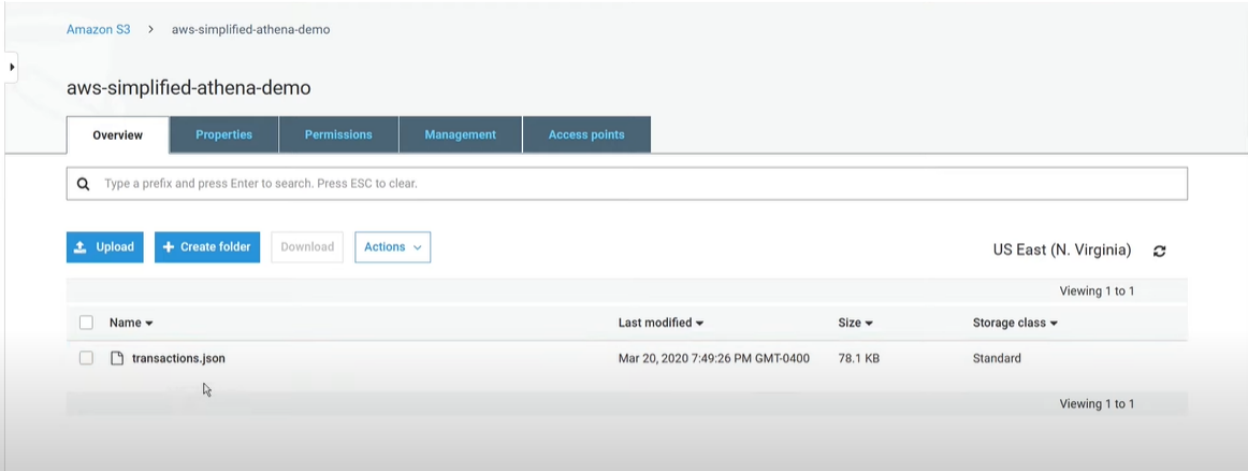
Step 1:

Go to buckets and create two buckets.



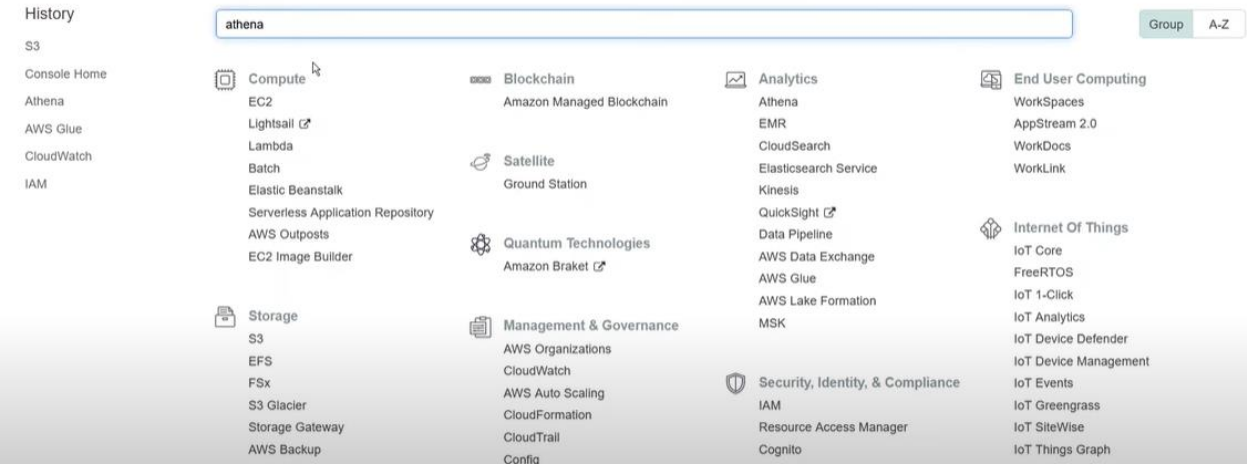
Step 2:

After clicking onto the bucket add files to it.



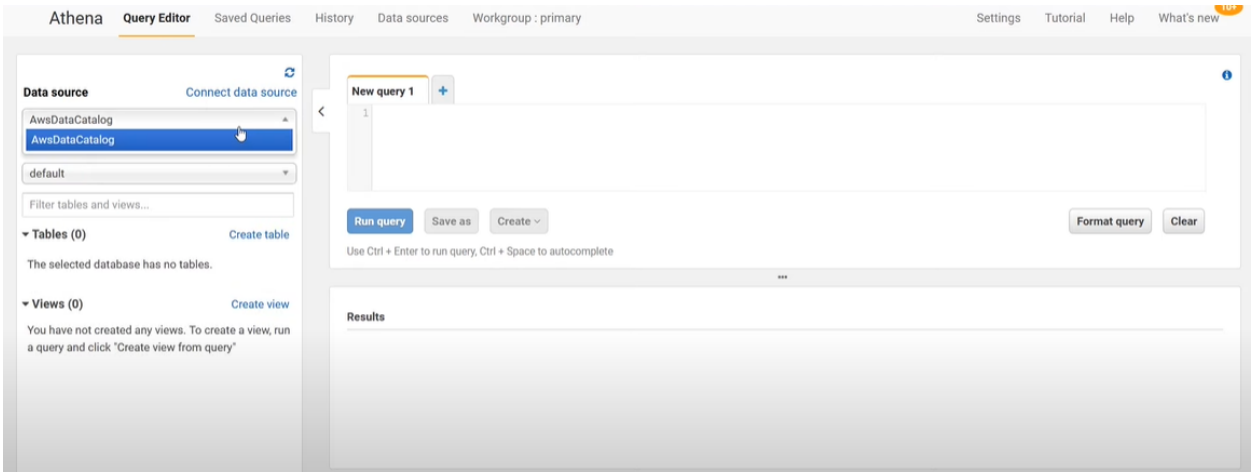
Step 3:

Now go to Amazon Athena.



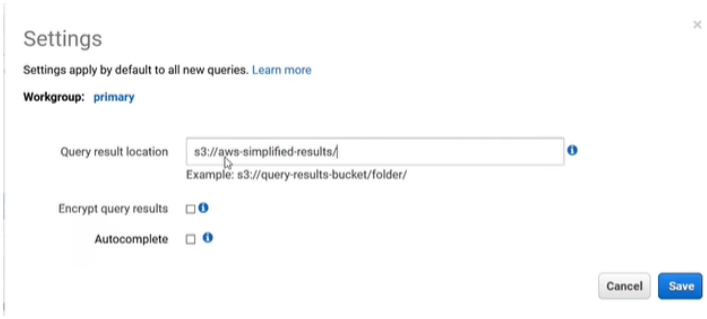
Step 4:

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



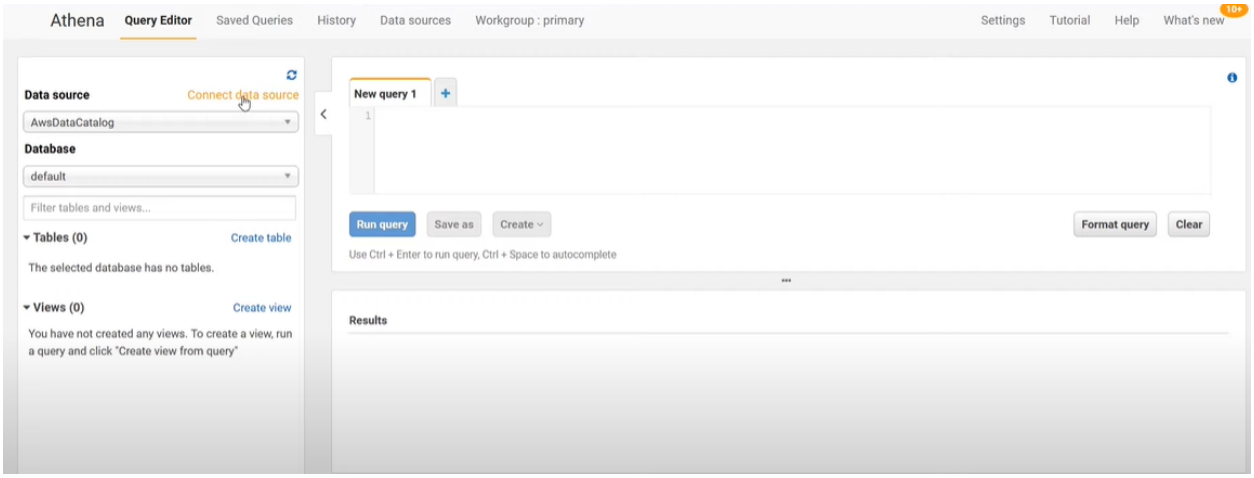
Step 5:

After that go to settings and specify an output path.



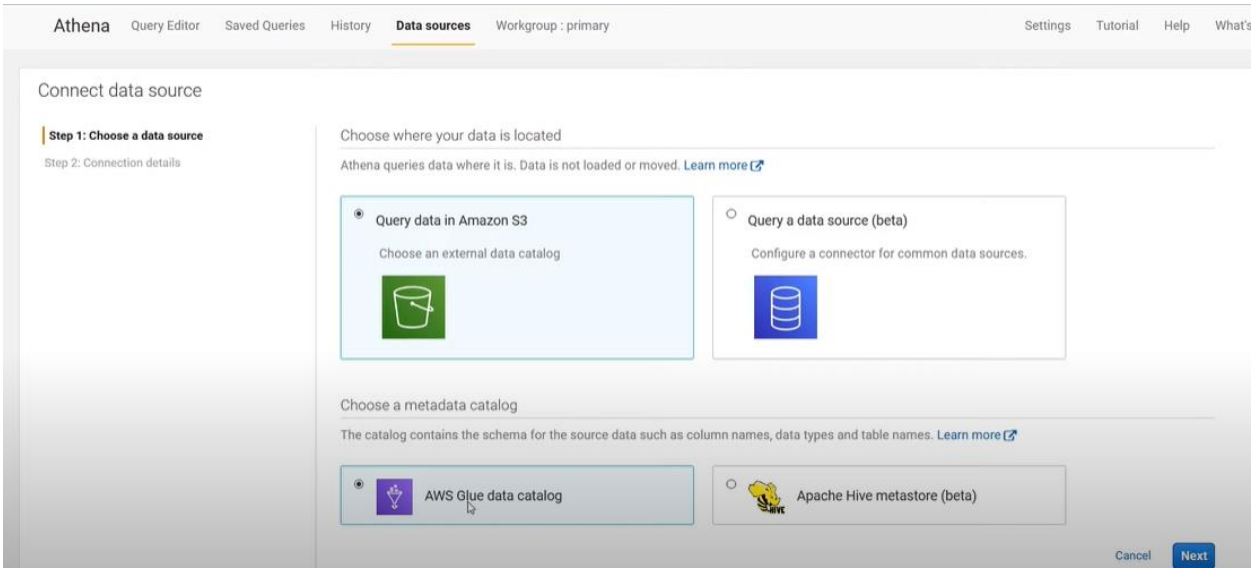
Step 6:

Click on connect data source.



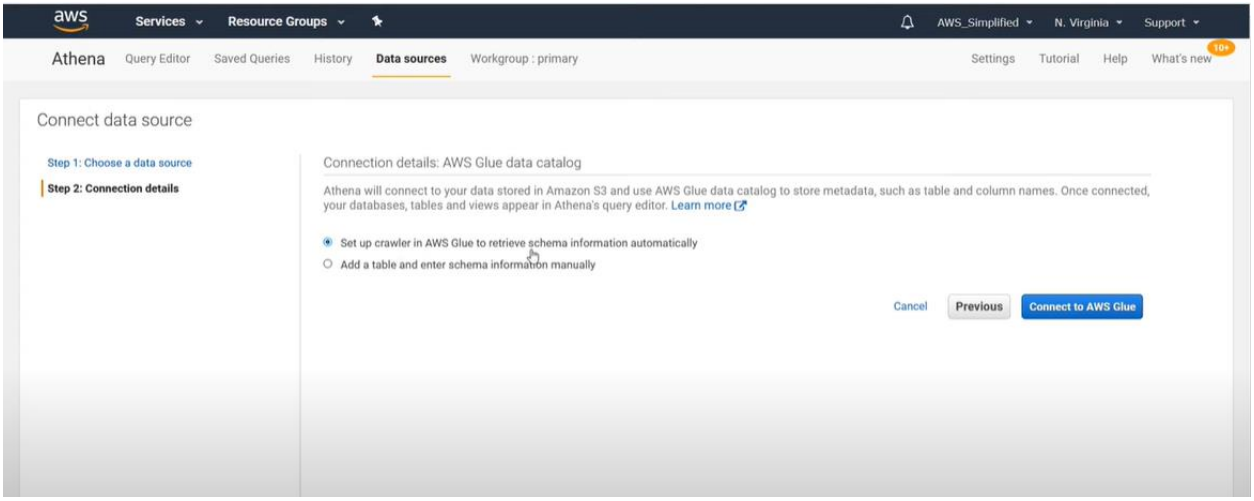
Step 7:

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



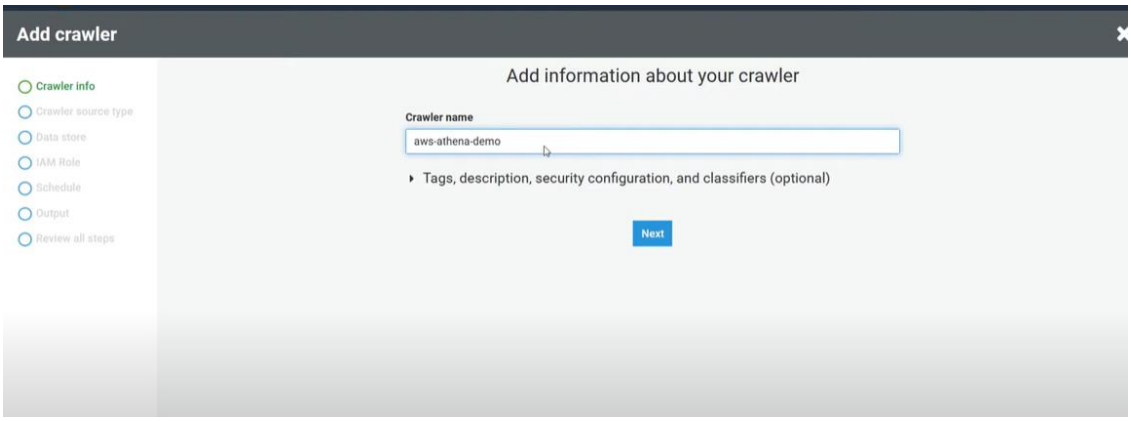
Step 8:

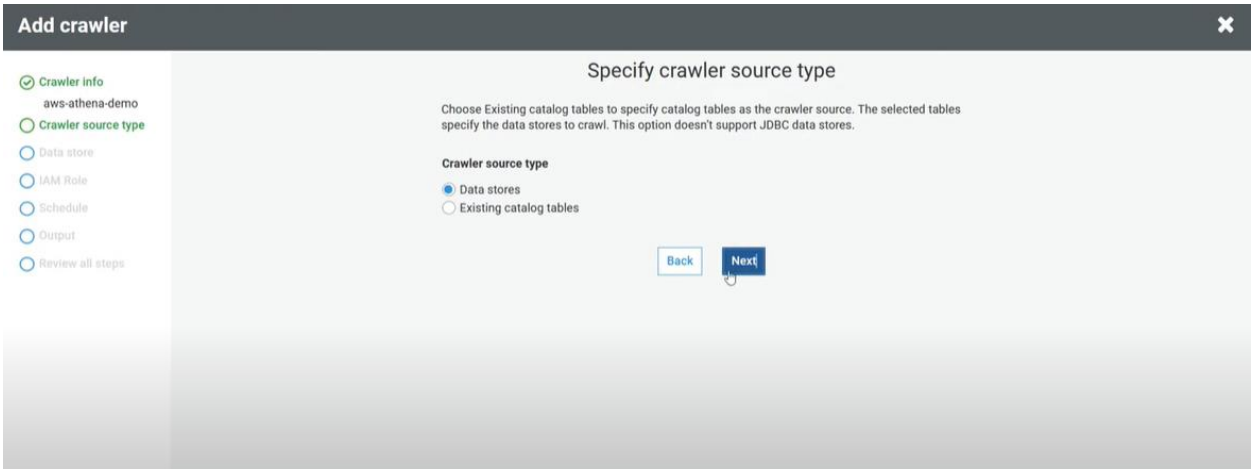
Click on next and select setup a crawler in AWS glue to retrieve schema information automatically.

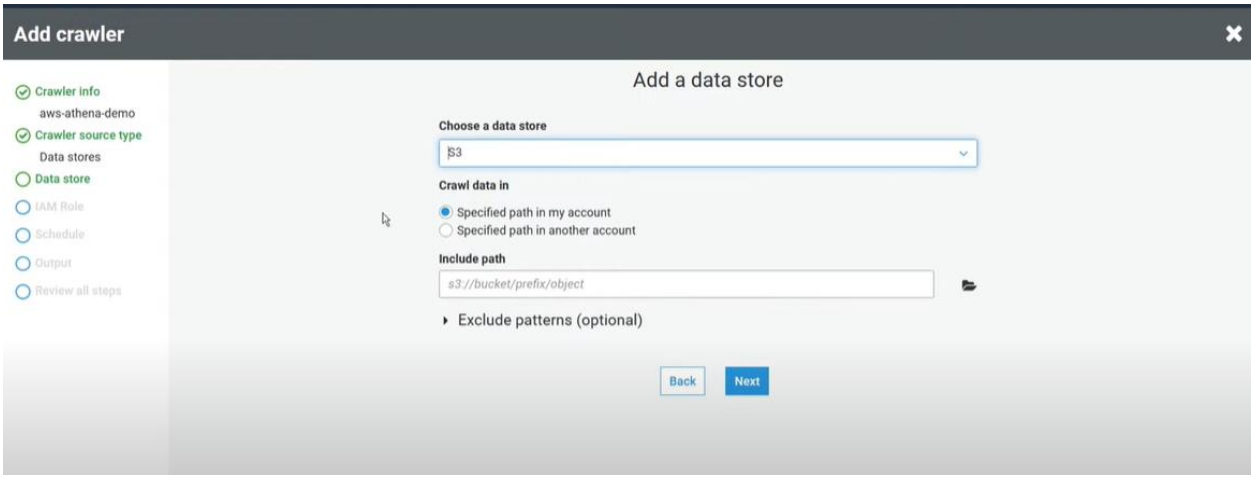


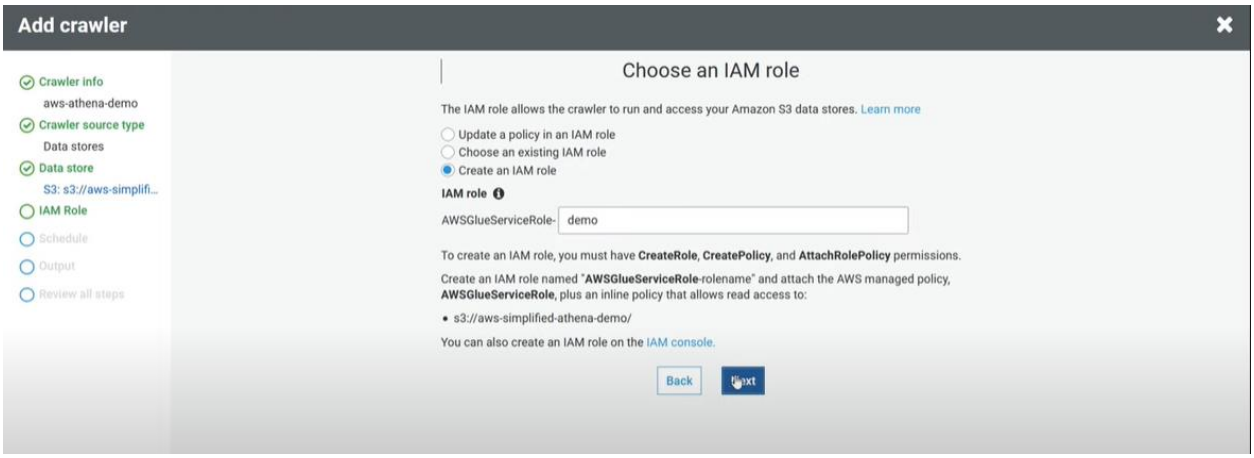
Step 9:

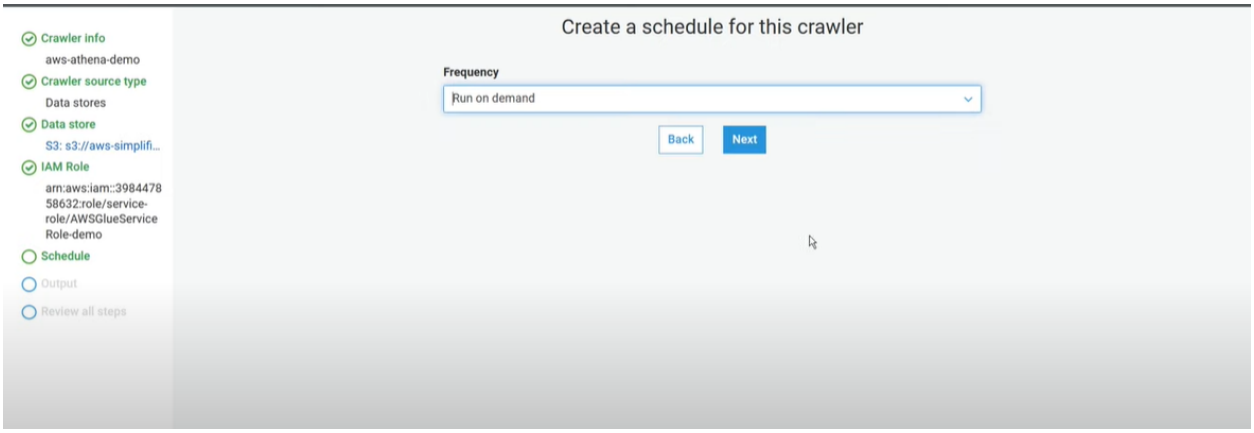
After selecting that it will redirect to a new page and add crawler and follow below steps to add a new crawler after setting up click on finish.

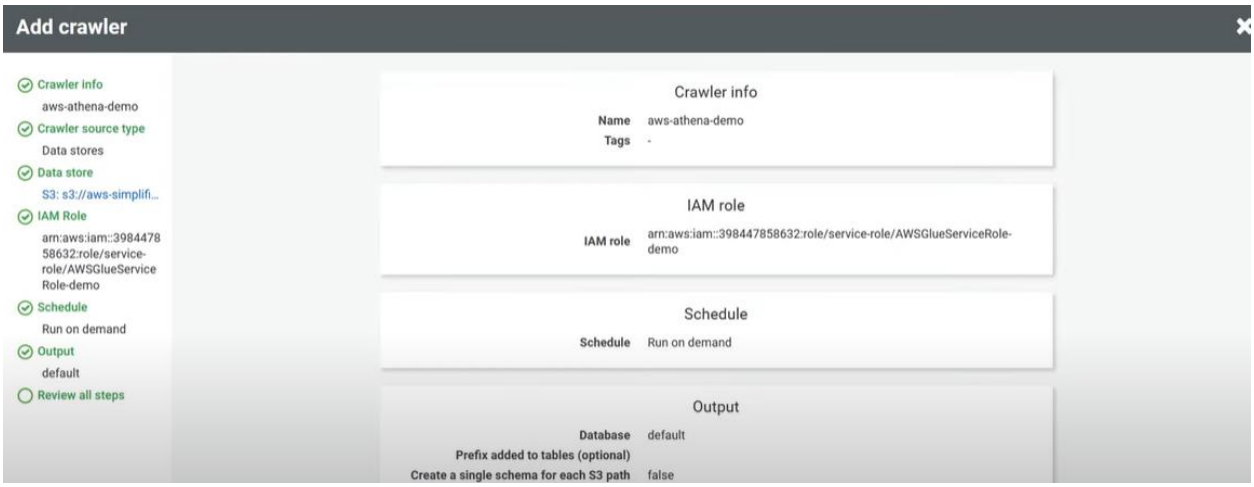





Add crawler
Crawler info
aws-athena-demo
Crawler source type
Data stores
Data store
IAM Role
Schedule
Output
Review all steps
Add a data store
Choose a data store
S3
Crawl data in
Specified path in my account
Specified path in another account
Include path
s3://bucket/prefix/object
Exclude patterns (optional)
Back
Next

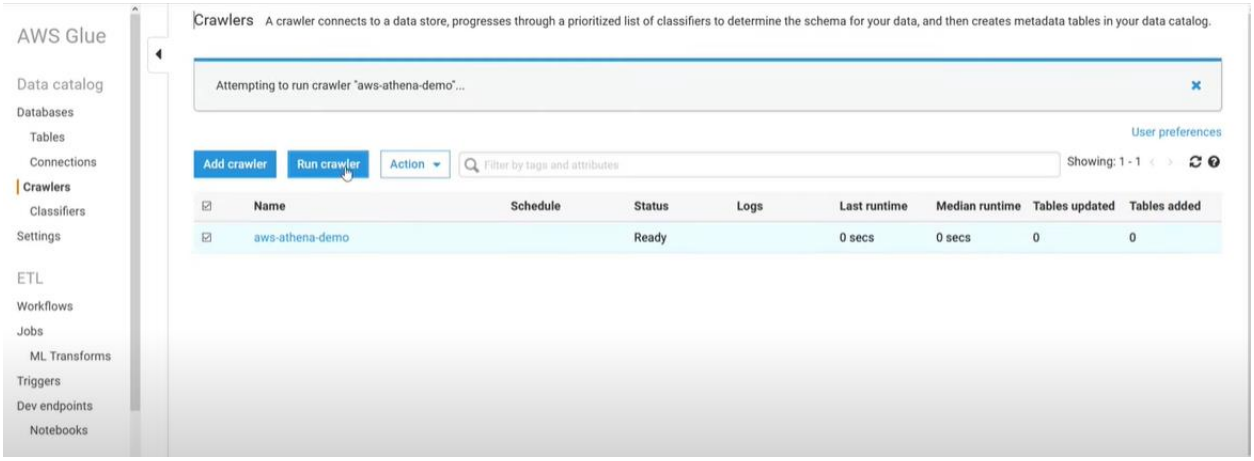

Add crawler
Crawler info
aws-athena-demo
Crawler source type
Data stores
Data store
S3: s3://aws-simplifi...
IAM Role
Schedule
Output
Review all steps
Choose an IAM role
The IAM role allows the crawler to run and access your Amazon S3 data stores. Learn more
Update a policy in an IAM role
Choose an existing 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
arn: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



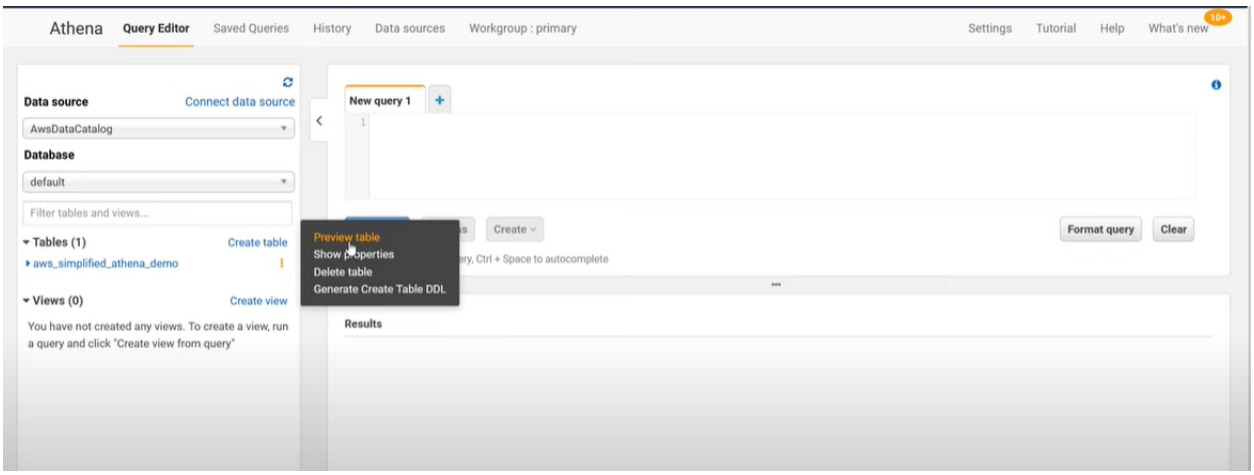
Step 10:

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



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.