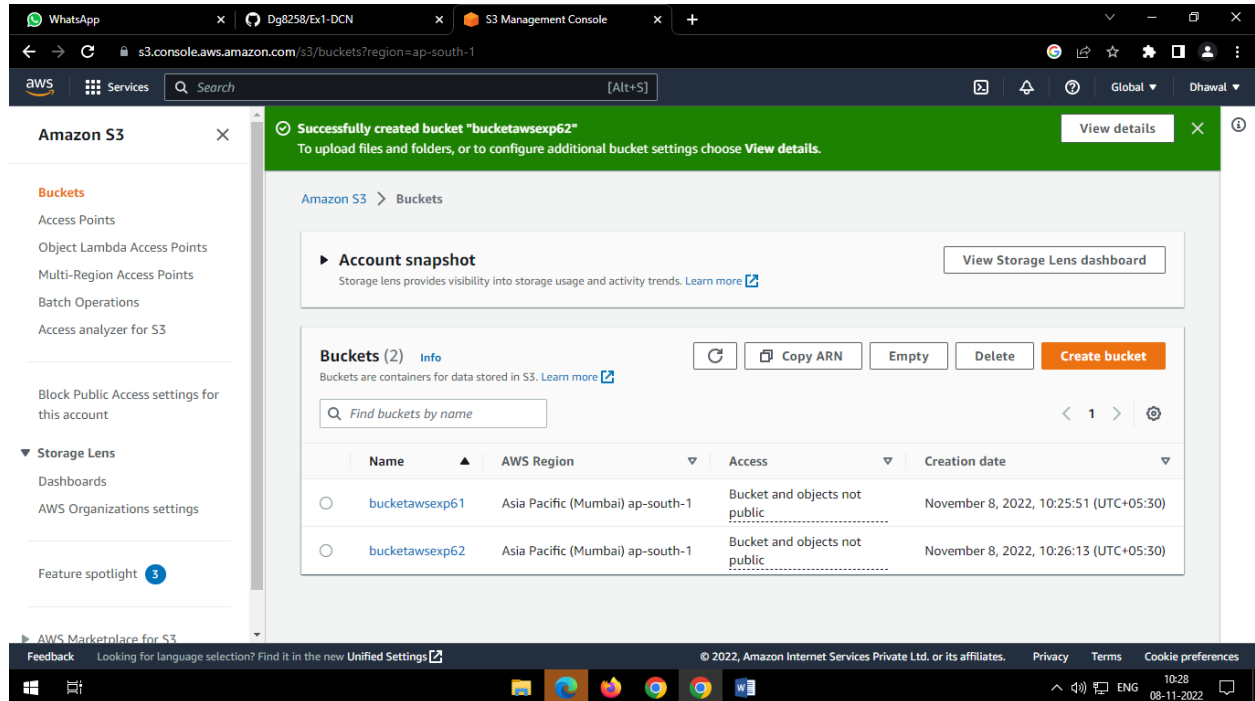


## EXP NO : 6      Querying Data in S3 with Amazon Athena

**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.

The screenshot displays the Amazon S3 console interface. The left sidebar shows the 'Amazon S3' menu with options like Buckets, Access Points, and Storage Lens. The main content area is titled 'bucketawsexp61' and includes tabs for Objects, Properties, Permissions, Metrics, Management, and Access Points. The 'Objects' tab is active, showing a list of objects. Below the tabs, there is a section for 'Objects (1)' with a description and a search bar. A table lists the objects, showing one object: 'DCN Experiment-5.081.pdf' with a size of 434.3 KB and a storage class of Standard. The bottom of the screen shows the Windows taskbar with various application icons and the system clock.

Amazon S3 > Buckets > bucketawsexp61

### bucketawsexp61

Info

Objects | Properties | Permissions | Metrics | Management | Access Points

**Objects (1)**

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

[Refresh](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#)

[Create folder](#) [Upload](#)

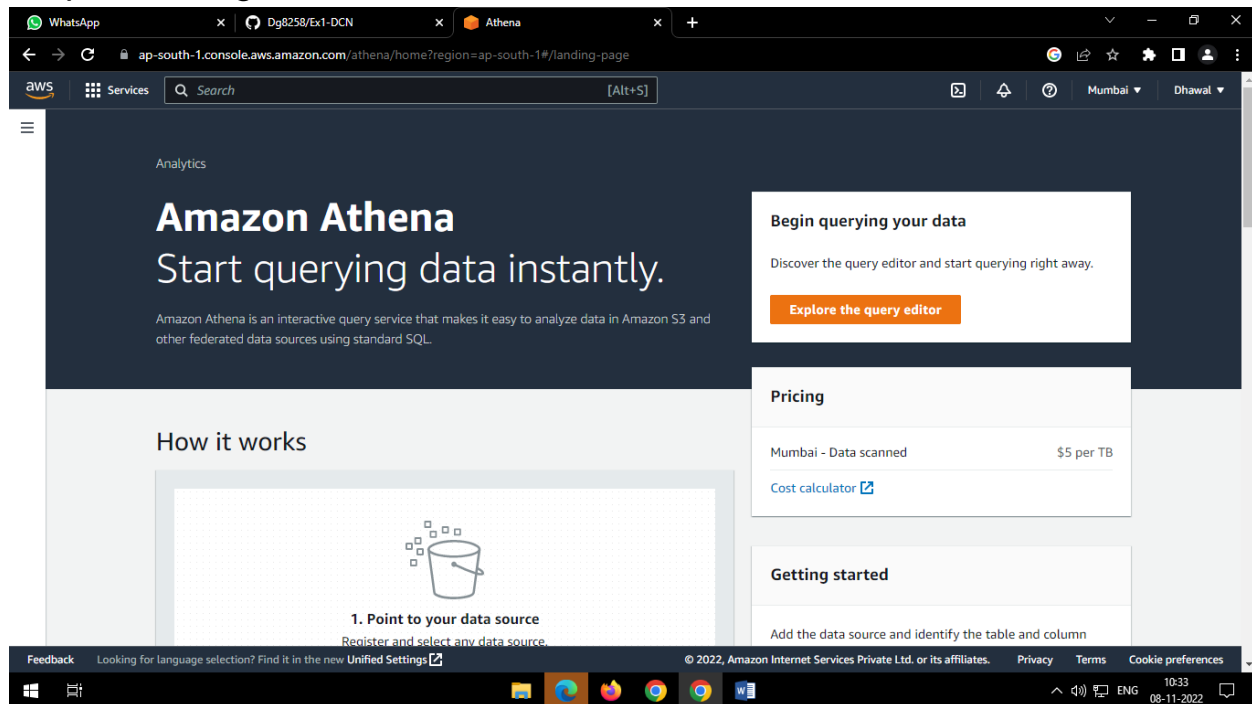
Find objects by prefix

	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	DCN Experiment-5.081.pdf	pdf	November 8, 2022, 10:27:23 (UTC+05:30)	434.3 KB	Standard

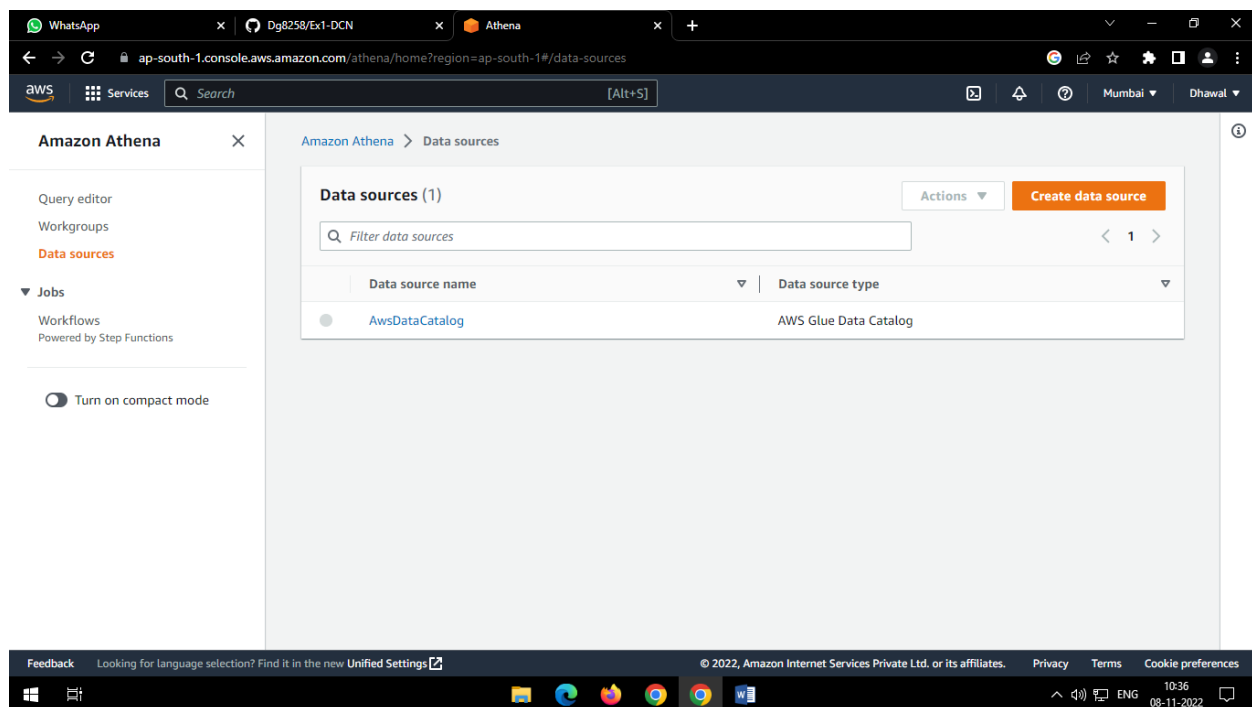
© 2022, Amazon Internet Services Private Ltd. or its affiliates. Privacy Terms Cookie preferences

10:32 08-11-2022

Step 3 : Now go to Amazon athena.

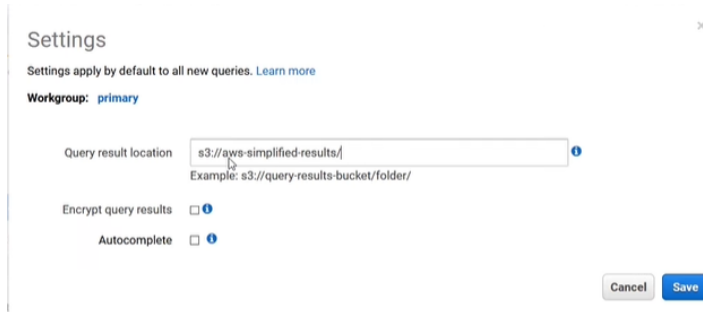


Step 4 :  
Select AwsDataCatalog in the left side which is present in the data source tab.



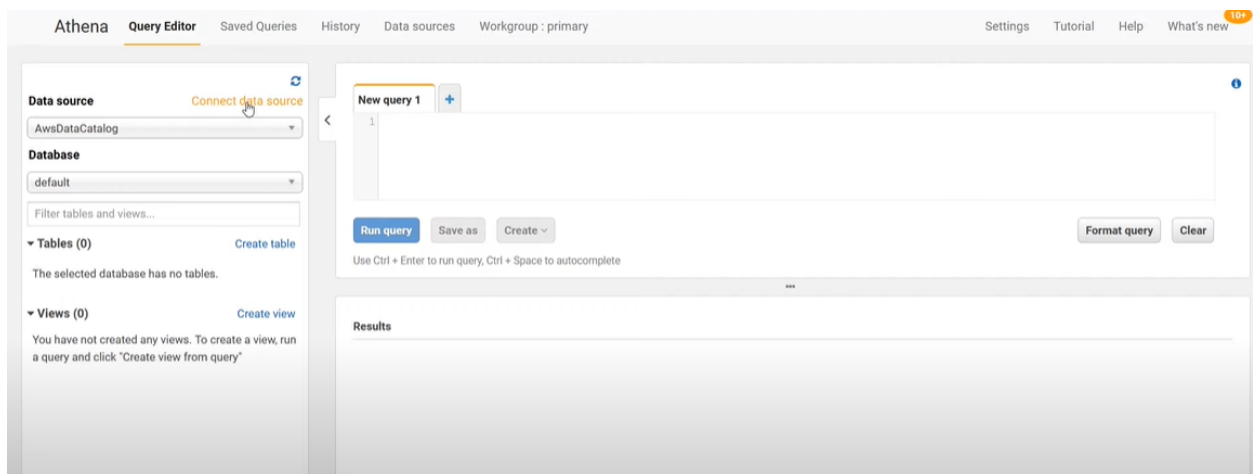
## Step 5 :

After that go to settings and specify an output path.



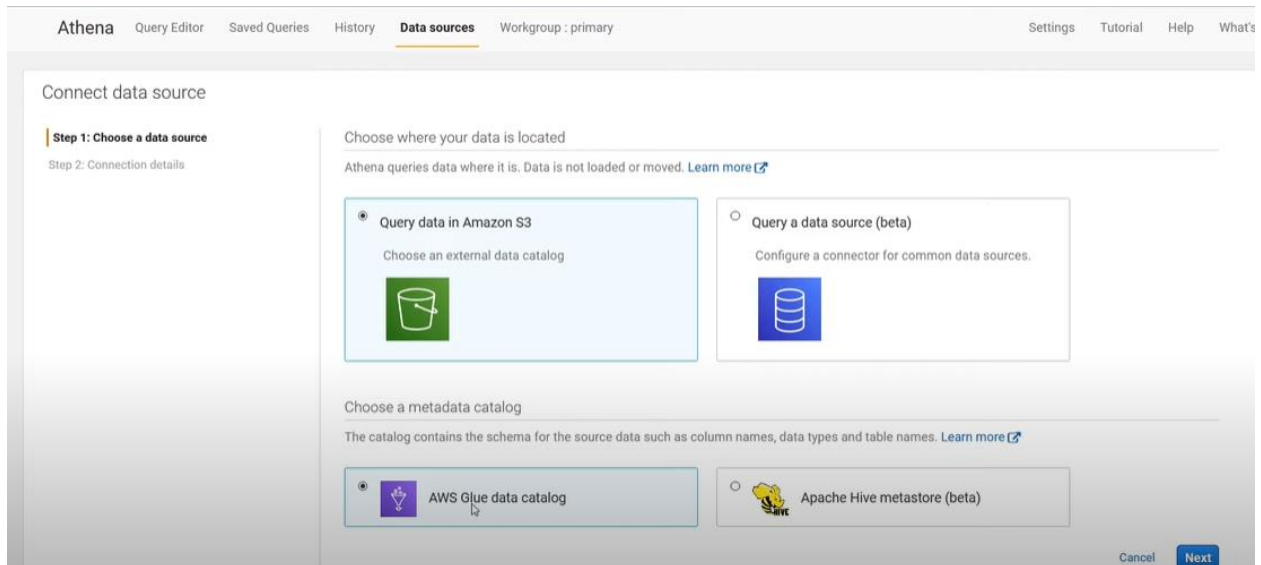
## Step 6:

Click on connect data source.

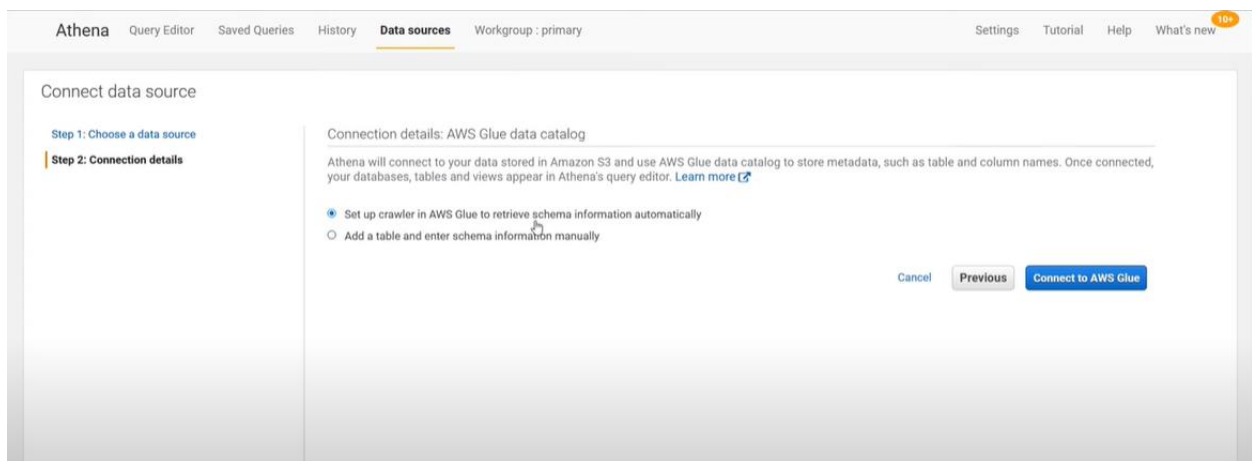


## Step 7:

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.

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

aws-athena-demo

Crawler info

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

aws-athena-demo

Crawler info

Crawler source type

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

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

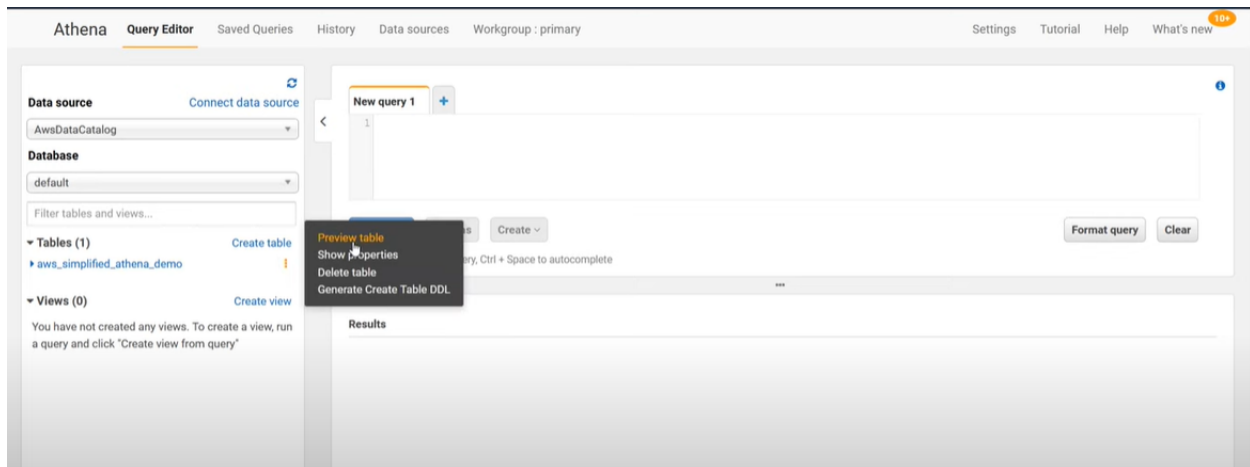
Showing: 1 - 1

	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.