

EXP NO : 6 Querying Data in S3 with Amazon Athena

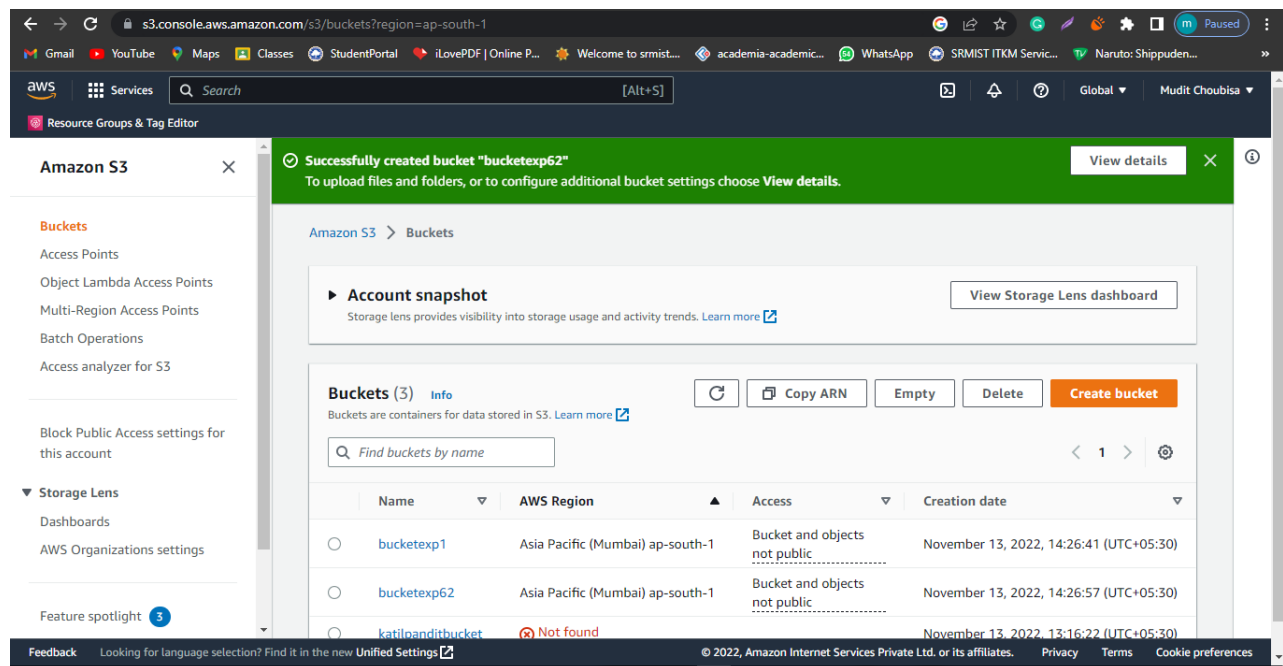
NAME: MUDIT CHOUBISA

REG. NO.:RA2011028010090

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 shows the Amazon S3 console interface for a bucket named 'bucketexp1'. The breadcrumb navigation at the top indicates the path: Amazon S3 > Buckets > bucketexp1. The bucket name 'bucketexp1' is displayed with an 'Info' link. Below this, there are tabs for 'Objects', 'Properties', 'Permissions', 'Metrics', 'Management', and 'Access Points'. The 'Objects' tab is selected, showing a list of objects. A description states: '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](#)'. Action buttons include 'Refresh', 'Copy S3 URI', 'Copy URL', 'Download', 'Open', 'Delete', 'Actions', 'Create folder', and 'Upload'. A search bar prompts 'Find objects by prefix'. The object list table has columns for Name, Type, Last modified, Size, and Storage class. One object is listed: 'DCN Exp 5.pdf' with a type of 'pdf', last modified on 'November 13, 2022, 14:28:14 (UTC+05:30)', a size of '2.3 MB', and a storage class of 'Standard'. The footer contains a 'Feedback' link, a language selection notice, copyright information for 2022, and links for 'Privacy', 'Terms', and 'Cookie preferences'.

Amazon S3 > Buckets > bucketexp1

bucketexp1 [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](#)

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	DCN Exp 5.pdf	pdf	November 13, 2022, 14:28:14 (UTC+05:30)	2.3 MB	Standard

Feedback | Looking for language selection? Find it in the new [Unified Settings](#) | © 2022, Amazon Internet Services Private Ltd. or its affiliates. | [Privacy](#) | [Terms](#) | [Cookie preferences](#)

Step 3 : Now go to Amazon athena.

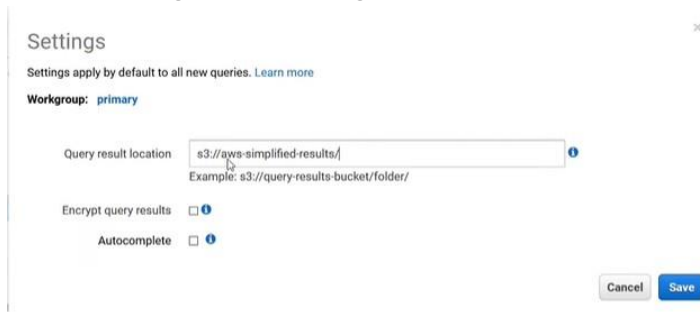
The screenshot shows the Amazon Athena console landing page. The header includes the AWS logo, a search bar, and navigation links. The main content area features the 'Amazon Athena' logo and the text 'Start querying data instantly.' Below this, a description states: 'Amazon Athena is an interactive query service that makes it easy to analyze data in Amazon S3 and other federated data sources using standard SQL.' To the right, there are three sections: 'Begin querying your data' with a button 'Explore the query editor', 'Pricing' showing 'Mumbai - Data scanned' at '\$5 per TB' with a 'Cost calculator' link, and 'Getting started' with instructions to 'Add the data source and identify the table and select the query.' Below the main heading, a section titled 'How it works' shows a diagram of a bucket with a query icon and the text '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.

The screenshot shows the 'Data sources' tab in the Amazon Athena console. The left sidebar contains navigation links: 'Query editor', 'Workgroups', 'Data sources' (highlighted), 'Jobs', 'Workflows', and 'Powered by Step Functions'. The main content area is titled 'Data sources (1)' and includes a search bar 'Filter data sources'. Below the search bar is a table with two columns: 'Data source name' and 'Data source type'. The table contains one entry: 'AwsDataCatalog' with the type 'AWS Glue Data Catalog'. To the right of the table is a 'Create data source' button. The footer includes a feedback link, a language selection notice, and copyright information.

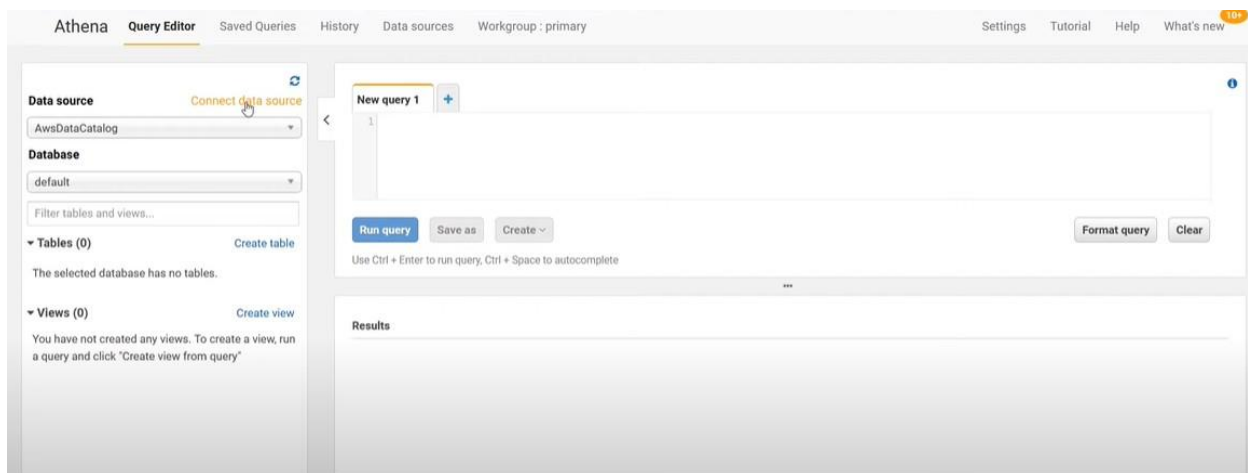
Step 5 :

After that go to settings and specify an output path.



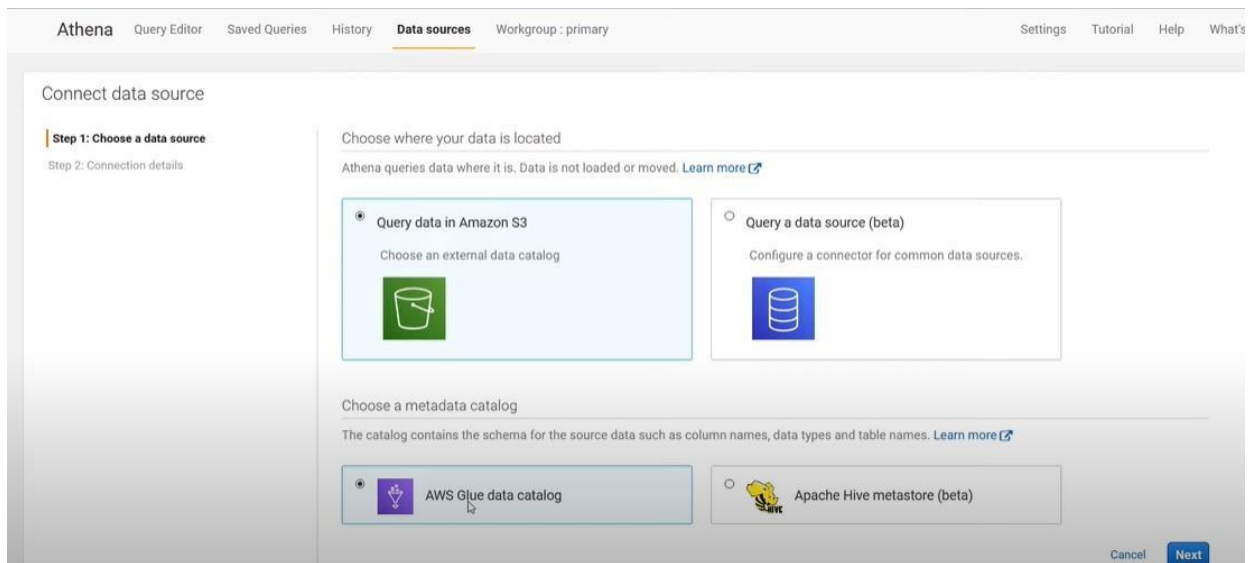
Step 6:

Click on connect data source.



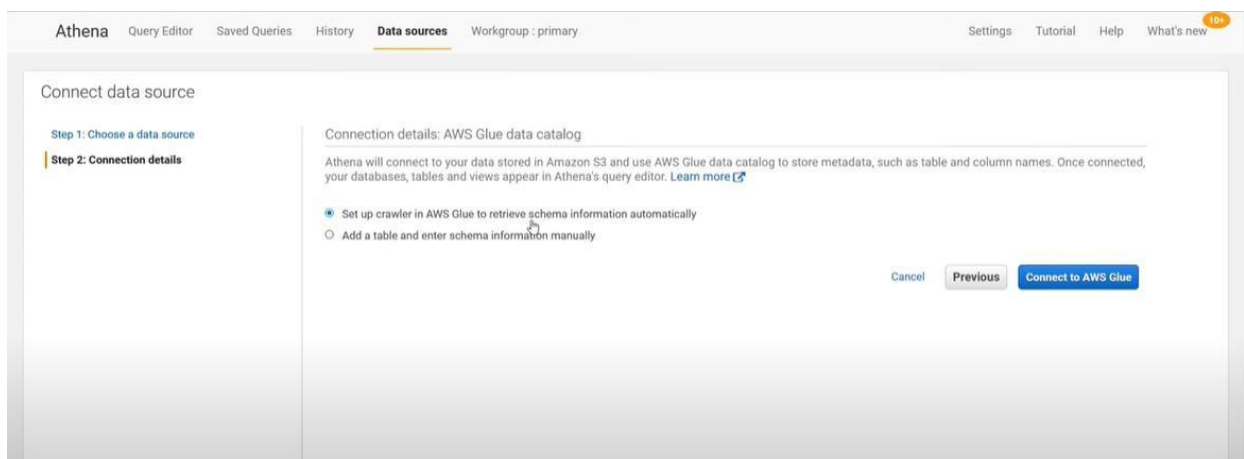
Step 7:

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

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 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

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

aws-athena-demo

Ready

0 secs

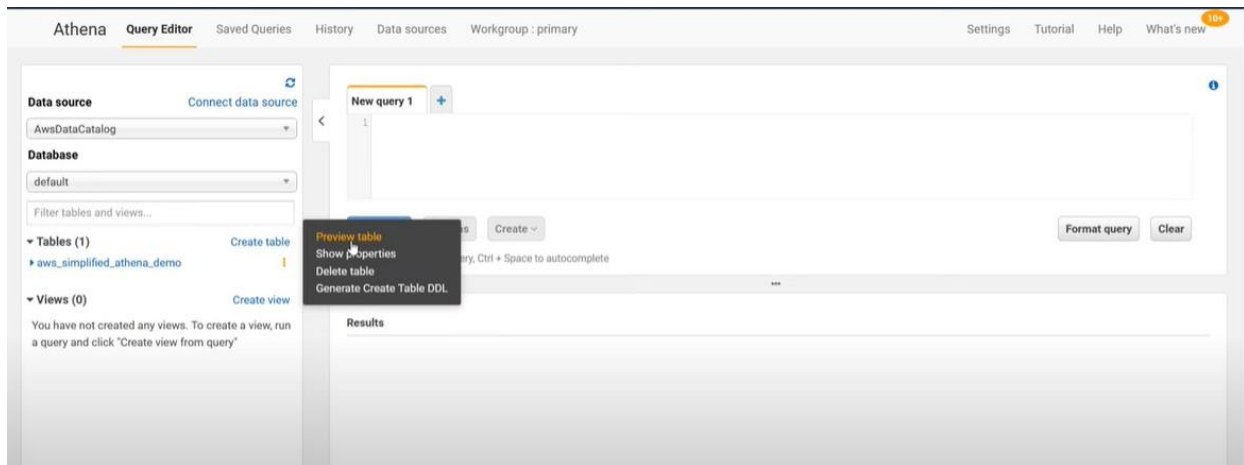
0 secs

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.