

Lesson 04 Demo 10

Querying S3 Bucket Operations with Athena

Objective: To execute a query within AWS Athena for performing operations on a designated S3 bucket

Tools required: None

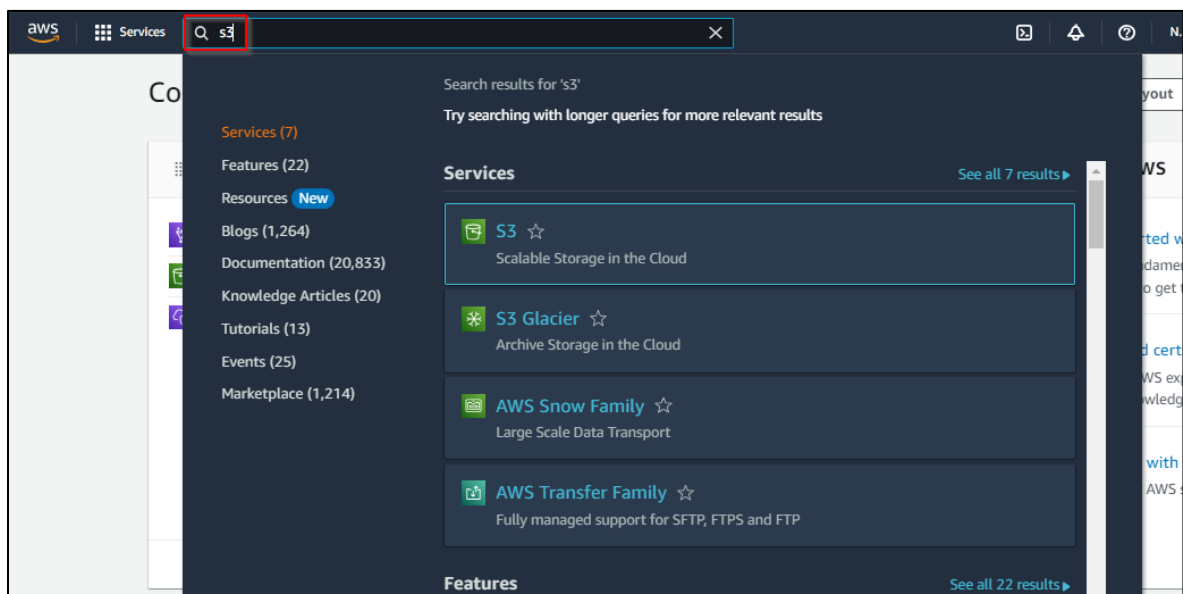
Prerequisites: AWS Lab access with an AWS account created

Steps to be followed:

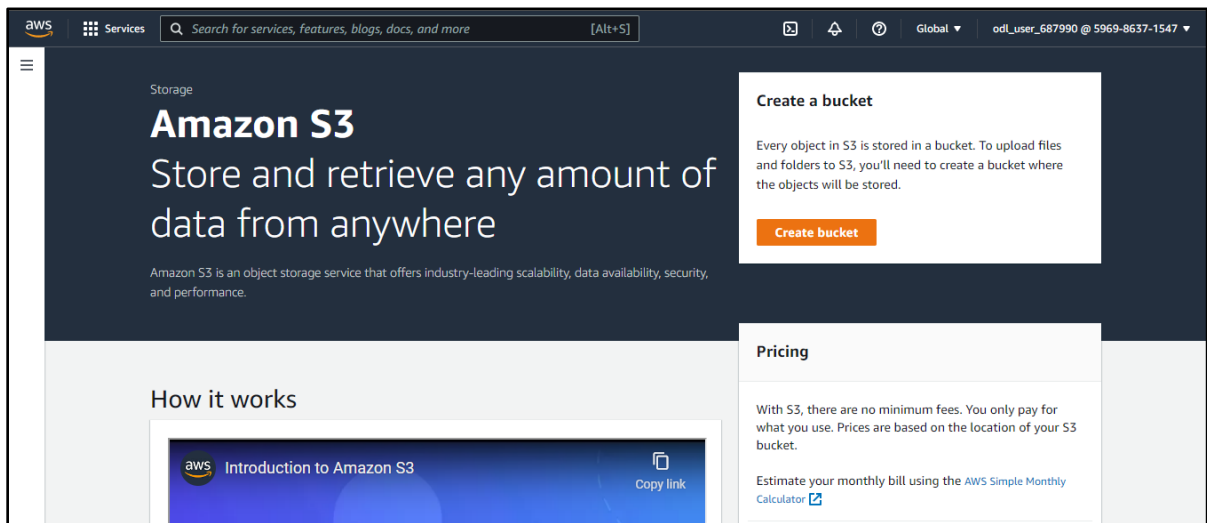
1. Create S3 buckets
2. Execute queries in Athena

Step 1: Create S3 buckets

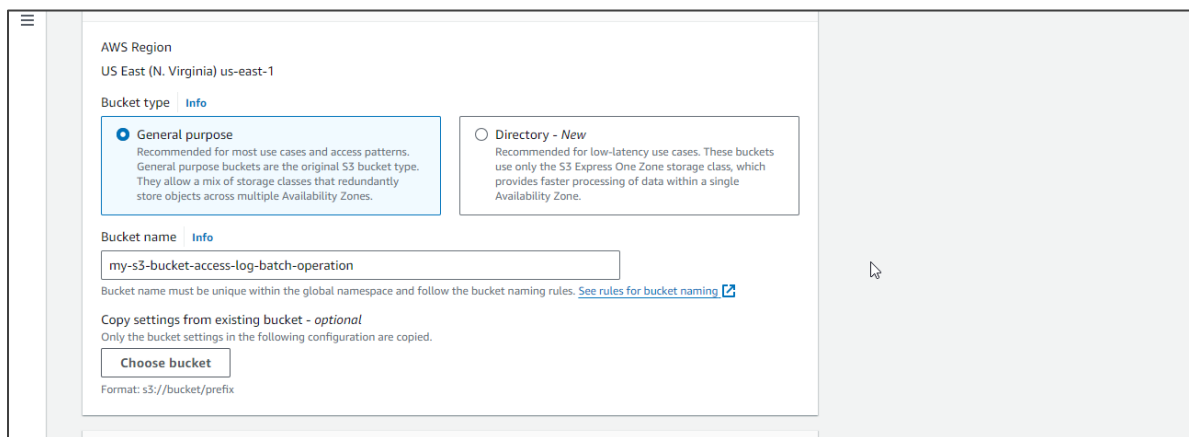
1.1 Open the AWS Management Console and search for **S3 bucket**



1.2 Click on **Create bucket**



1.3 Create two buckets named **my-s3-bucket-access-log-batch-operation** and **s3-bucket-batch-operation-encryption** to aggregate log files from the **s3-bucket-batch-operation-encryption** bucket. Enable **bucket Versioning** for both.



Bucket Versioning

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from application failures. [Learn more](#)

Bucket Versioning

☐ Disable

☒ Enable

Buckets (2) Info

Refresh

Copy ARN

Empty

Delete

Create bucket

Buckets are containers for data stored in S3. [Learn more](#)

Find buckets by name

< 1 > ⚙️

	Name ▲	AWS Region ▼	Access ▼	Creation date ▼
<input type="radio"/>	my-s3-bucket-access-log-batch-operation	US East (N. Virginia) us-east-1	Bucket and objects not public	July 23, 2023, 14:28:00 (UTC+05:30)
<input type="radio"/>	s3-bucket-batch-operation-encryption	US East (N. Virginia) us-east-1	Bucket and objects not public	July 23, 2023, 14:45:27 (UTC+05:30)

1.4 Access the s3-bucket-batch-operation-encryption bucket

Amazon S3 > Buckets

Account snapshot

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

View Storage Lens dashboard

Buckets (2) Info

Refresh

Copy content

Empty

Delete

Create bucket

Buckets are containers for data stored in S3. [Learn more](#)

Find buckets by name

< 1 > ⚙️

	Name ▲	AWS Region ▼	Access ▼	Creation date ▼
<input type="radio"/>	my-s3-bucket-access-log-batch-operation	US East (N. Virginia) us-east-1	Bucket and objects not public	August 4, 2023, 15:55:41 (UTC+05:30)
<input type="radio"/>	s3-bucket-batch-operation-encryption	US East (N. Virginia) us-east-1	Bucket and objects not public	August 4, 2023, 16:02:48 (UTC+05:30)

1.5 Navigate to **Properties**

The screenshot shows the Amazon S3 console interface. The breadcrumb navigation at the top reads "Amazon S3 > Buckets > s3-bucket-batch-operation-encryption". The main heading is "s3-bucket-batch-operation-encryption" with an "Info" link. Below the heading is a tabbed interface with tabs for "Objects", "Properties", "Permissions", "Metrics", "Management", and "Access Points". The "Properties" tab is selected and highlighted with a red box. Under the "Bucket overview" section, there are two rows of information: "AWS Region" with the value "US East (N. Virginia) us-east-1" and "Amazon Resource Name (ARN)" with the value "arn:aws:s3::s3-bucket-batch-operation-encryption". Below this is the "Bucket Versioning" section, which includes a description of versioning and an "Edit" button.

1.6 Scroll down and select **Server access logging**, and then click **Edit** and **Browse S3**

The screenshot shows the "Edit server access logging" page in the Amazon S3 console. The heading is "Edit server access logging" with an "Info" link. Below the heading is the "Server access logging" section, which includes a description and a "Learn more" link. Under this section, there are two radio buttons: "Disable" and "Enable". The "Enable" radio button is selected. Below the radio buttons is a warning box with a red triangle icon and the text "Bucket policy will be updated. When you enable server access logging, the S3 console automatically updates your bucket policy to include access to the S3 log delivery group." Below the warning box is the "Target bucket" section, which contains a text input field with the value "s3://my-s3-bucket-access-log-batch-operation" and a "Browse S3" button highlighted with a red box. Below the input field is the format "Format: s3://bucket/prefix". At the bottom right of the page are "Cancel" and "Save changes" buttons.

1.7 Choose the log-batch file path and click **Choose destination**

Choose destination

S3 Buckets

Buckets that are not in the same Region as your source bucket (US East (N. Virginia) us-east-1) can't be chosen.

Buckets (1/2)

< 1 >

Name	AWS Region	Creation date
<input checked="" type="radio"/> my-s3-bucket-access-log-batch-operation	US East (N. Virginia) us-east-1	August 13, 2024, 11:23:15 (UTC+05:30)
<input type="radio"/> s3-bucket-batch-operation-encryption	US East (N. Virginia) us-east-1	August 13, 2024, 11:23:54 (UTC+05:30)

Cancel

Choose destination

1.8 Click **Save changes**

s3://my-s3-bucket-access-log-batch-operation

Browse S3

Format: s3://<bucket>/<optional-prefix-with-path>

Destination Region

US East (N. Virginia) us-east-1

Destination bucket name

my-s3-bucket-access-log-batch-operation

Destination prefix

-

Log object key format

☒ [DestinationPrefix][YYYY]-[MM]-[DD]-[hh]-[mm]-[ss]-[UniqueString]
 ☐ [DestinationPrefix][SourceAccountid]/[SourceRegion]/[SourceBucket]/[YYYY]/[MM]/[DD]/[YYYY]-[MM]-[DD]-[hh]-[mm]-[ss]-[UniqueString]

To speed up analytics and query applications, use this format.

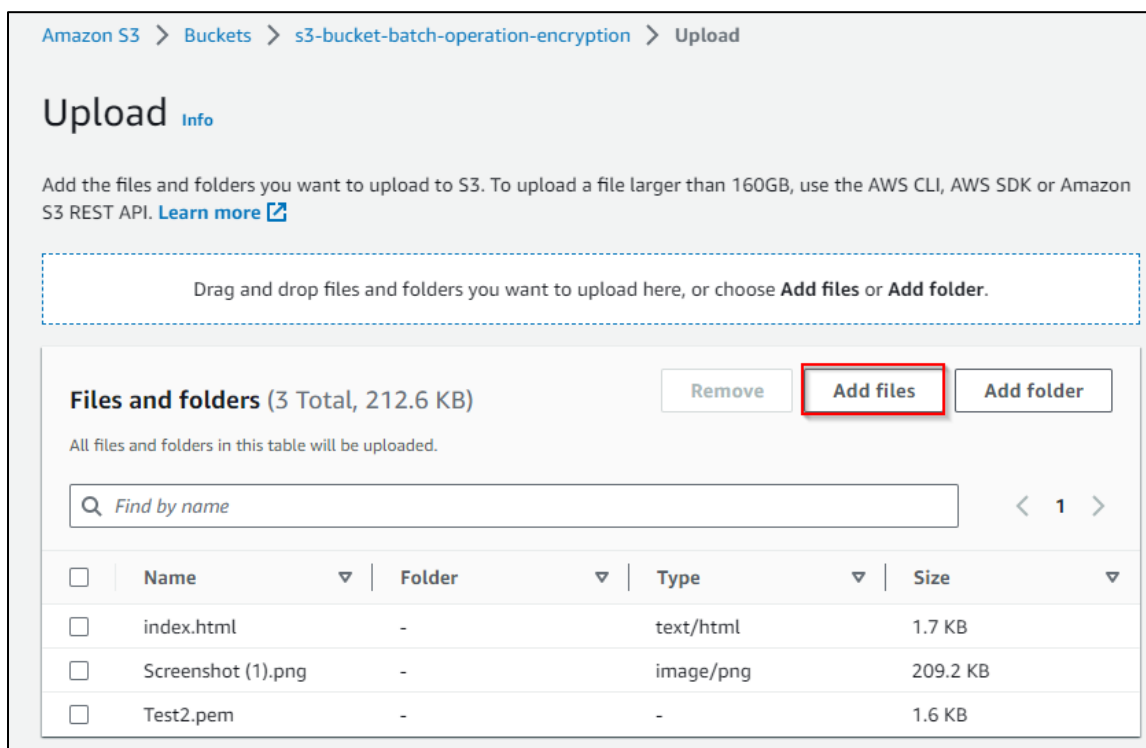
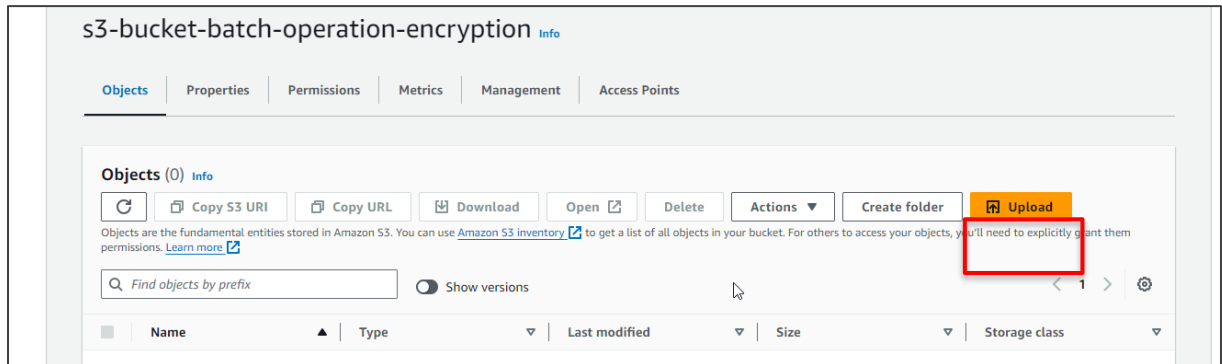
Log object key example

2024-07-01-10-12-56-[UniqueString]

Cancel

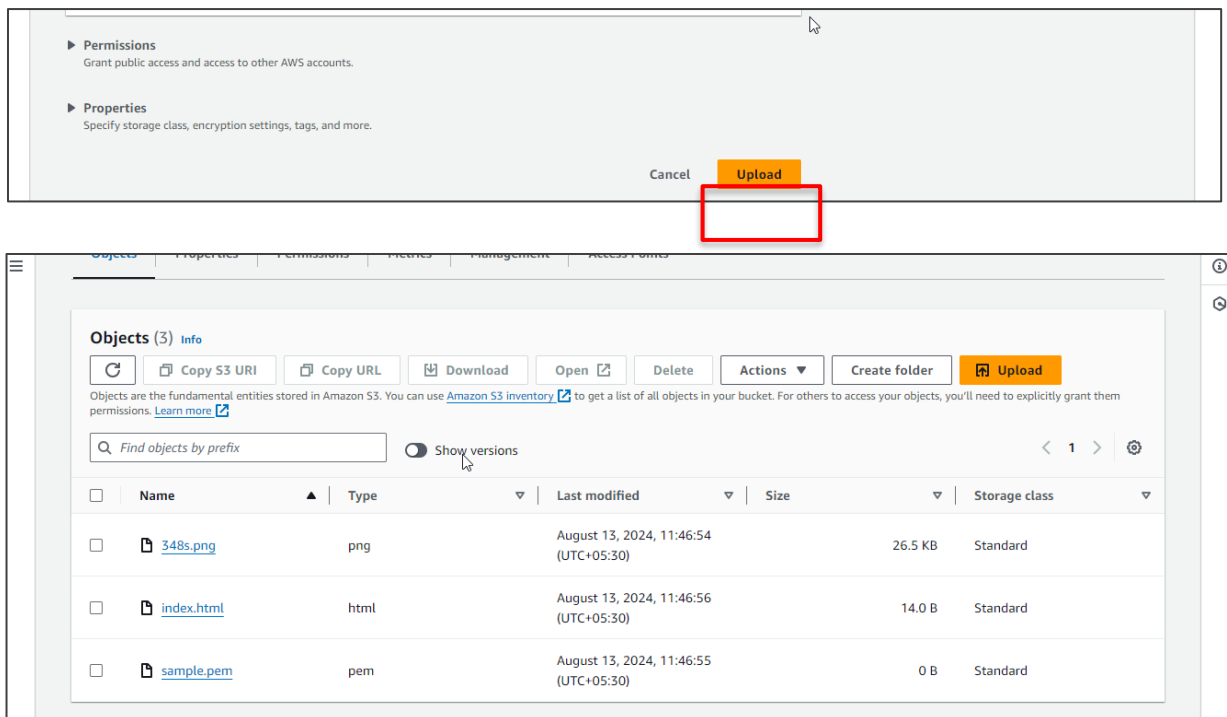
Save changes

1.9 Add HTML, PNG, and PEM files to the bucket using the **Upload** option, and click **Add** files



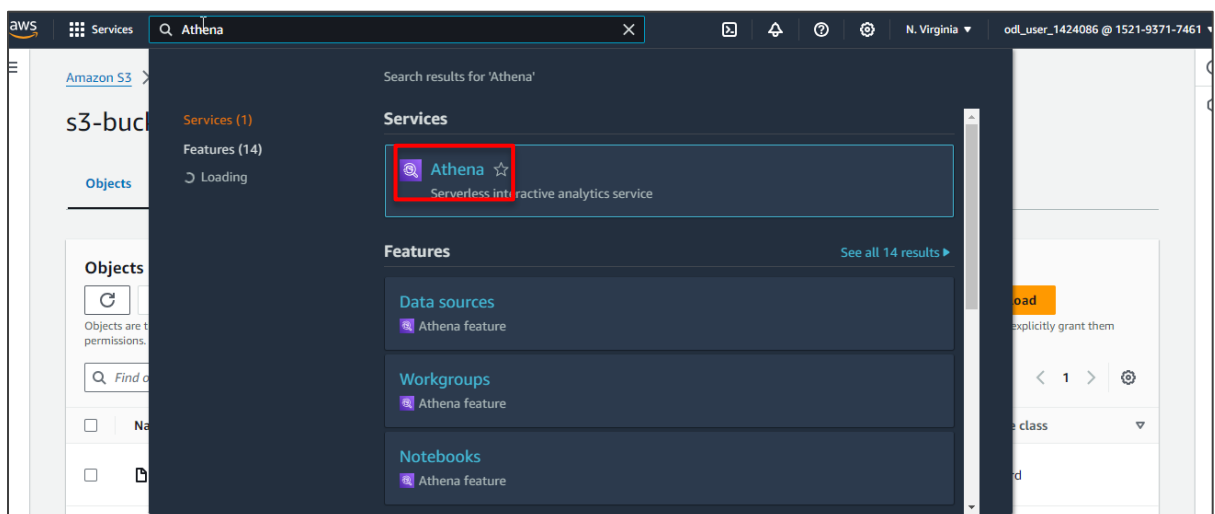
Note: These are sample files that can be created or sourced for demonstration purposes. For example, you can create a simple HTML file, use any image for the PNG file, and generate a PEM file using OpenSSL or any other tool.

1.10 Click on **Upload**

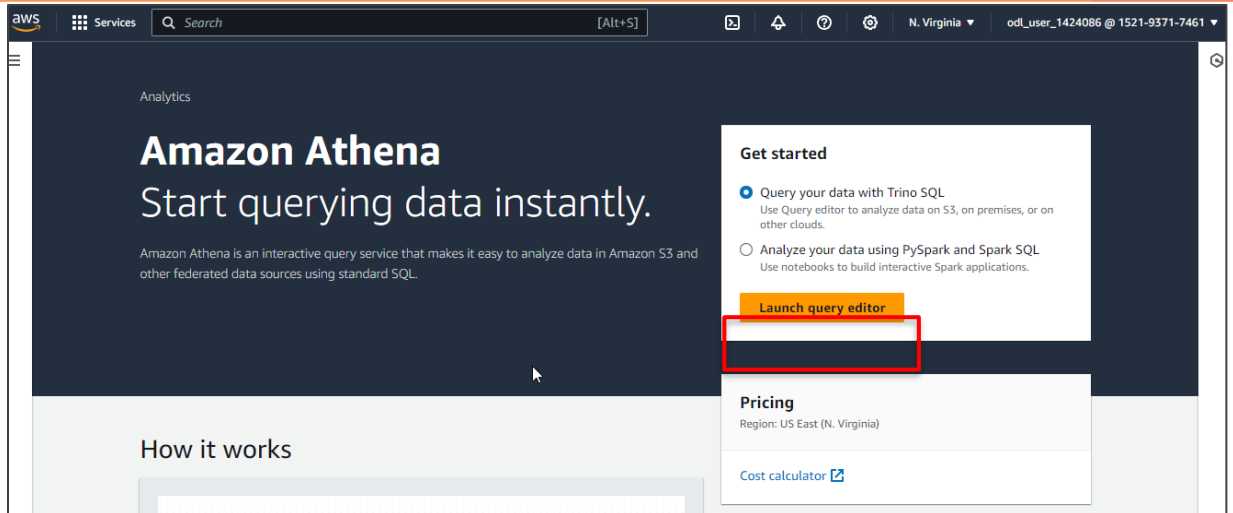


Step 2: Execute queries in Athena

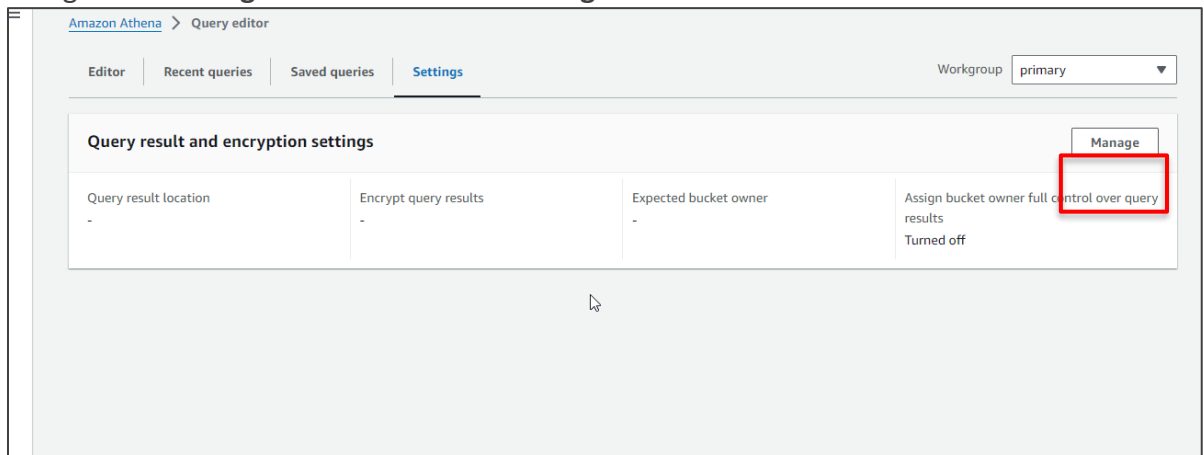
2.1 Navigate to the AWS Management Console and search for **Athena**



2.2 Click on **Launch query editor** to launch the query editor in Athena



2.3 Navigate to **Settings** and then click on **Manage**



2.4 Click on **Browse S3**, select one of the path locations, and then click **Save**

Amazon Athena > Query editor > Manage settings

Manage settings

Query result location and encryption

Location of query result - optional
Enter an S3 prefix in the current region where the query result will be saved as an object.

[View](#) [Browse S3](#)

Expected bucket owner - optional
Specify the AWS account ID that you expect to be the owner of your query results output location bucket.

☐ **Assign bucket owner full control over query results**
Enabling this option grants the owner of the S3 query results bucket full control over the query results. This means that if your query result location is owned by another account, you grant full control over your query results to the other account.

☐ **Encrypt query results**

Choose S3 data set

S3 buckets

Bucket (1/2)

Name	Creation date
<input checked="" type="radio"/> my-s3-bucket-access-log-batch-operation	2023-08-04T15:55:41.000+05:30
<input type="radio"/> s3-bucket-batch-operation-encryption	2023-08-04T16:02:48.000+05:30

[Cancel](#) [Choose](#)

Enter an S3 prefix in the current region where the query result will be saved as an object.

[View](#) [Browse S3](#)

You can create and manage lifecycle rules for this bucket
Use Amazon S3 lifecycle rules to store your query results and metadata cost effectively or to delete them after a period of time.
[Learn more](#)

[Lifecycle configuration](#)

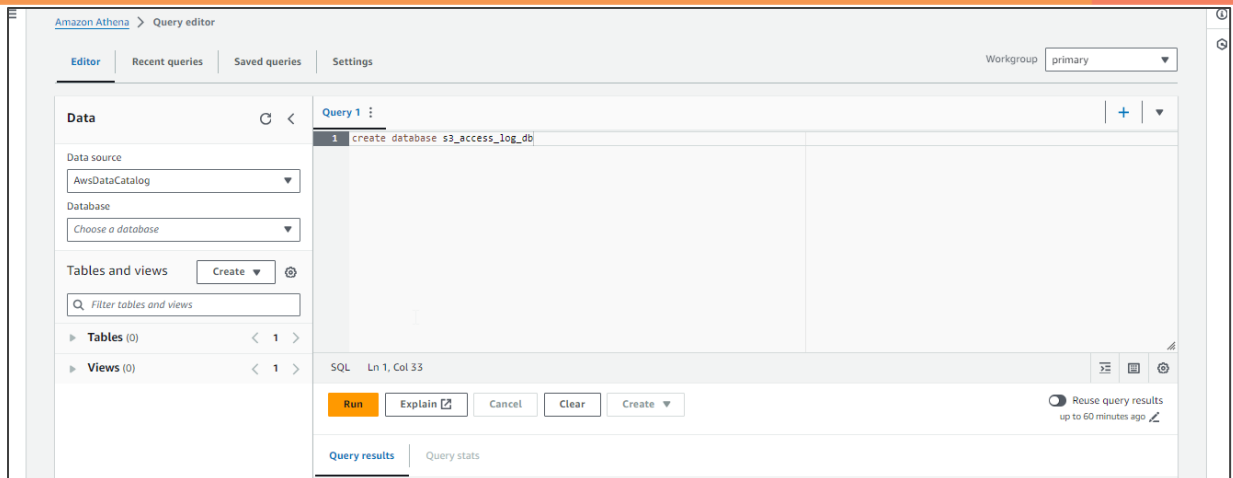
Expected bucket owner - optional
Specify the AWS account ID that you expect to be the owner of your query results output location bucket.

☐ **Assign bucket owner full control over query results**
Enabling this option grants the owner of the S3 query results bucket full control over the query results. This means that if your query result location is owned by another account, you grant full control over your query results to the other account.

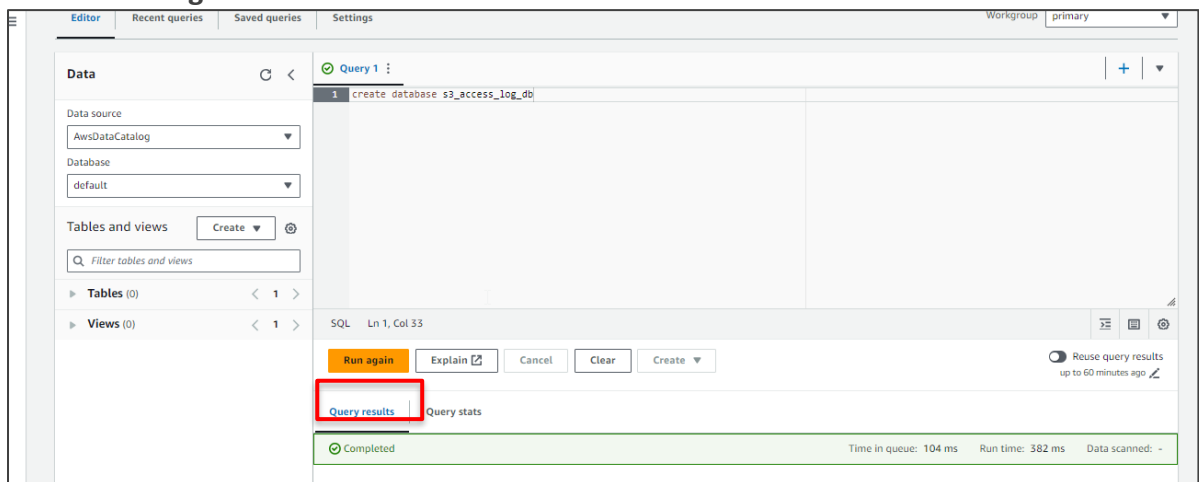
☐ **Encrypt query results**

[Cancel](#) [Save](#)

2.5 In the editor, execute the query **create database s3_access_log_db;**




2.6 Click on Run again

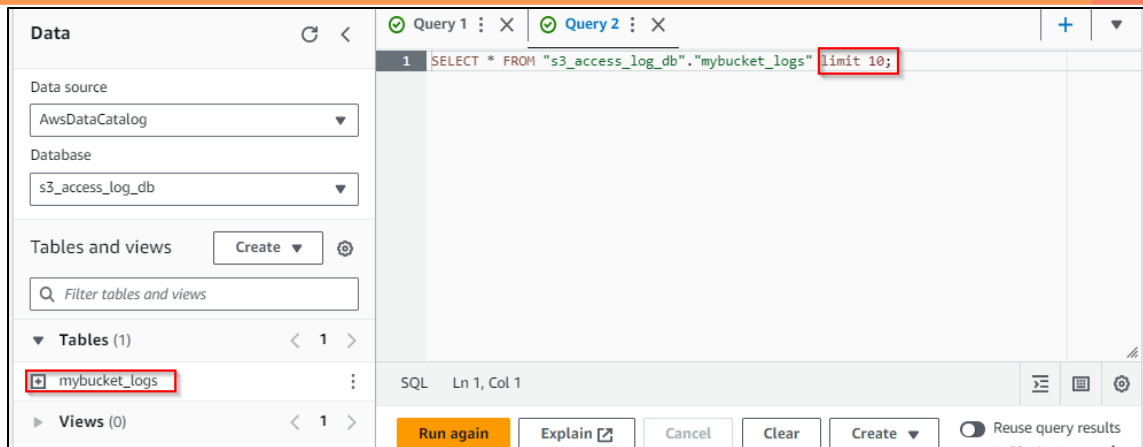


2.7 Execute the given query to create an external table:

CREATE EXTERNAL TABLE IF NOT EXISTS s3_access_log_db.mybucket_logs(

2.8 Execute the code by selecting **s3_access_log_db**, and then click **Run again**

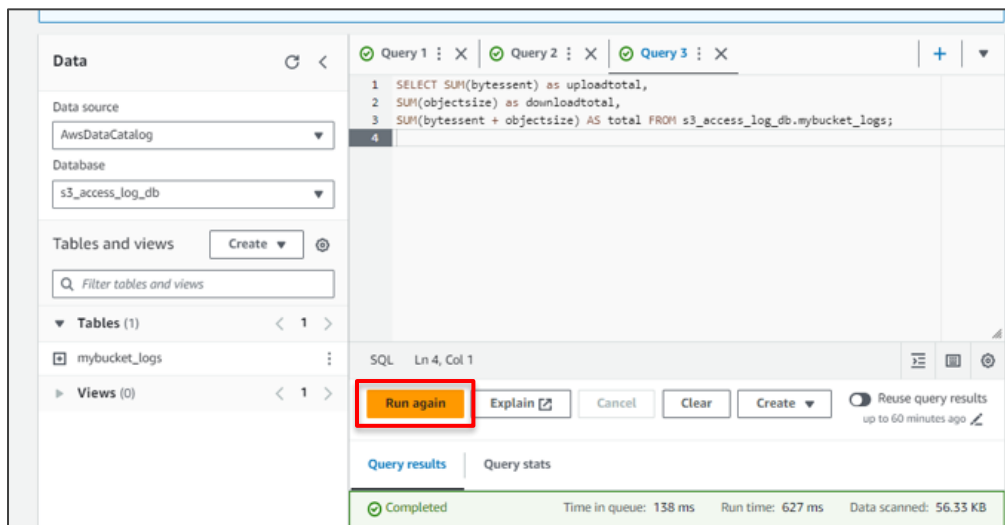
	Query results	Query stats	
	 Completed		Time in queue: 78 ms Run time: 352 ms Data scanned: -
		Query successful.	



Note: A predefined query with a limit of 10 rows will be available.

2.9 Execute the given query to find the exact amount of data uploaded to and downloaded from the monitored bucket:

**SELECT SUM(bytesent) as uploadtotal,
SUM(objectsize) as downloadtotal,
SUM(bytesent + objectsize) AS total FROM s3_access_log_db.mybucket_logs;**



By following these steps, you can now execute queries in AWS Athena, allowing you to analyze operations on a designated S3 bucket and gain insights into data usage patterns.