

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

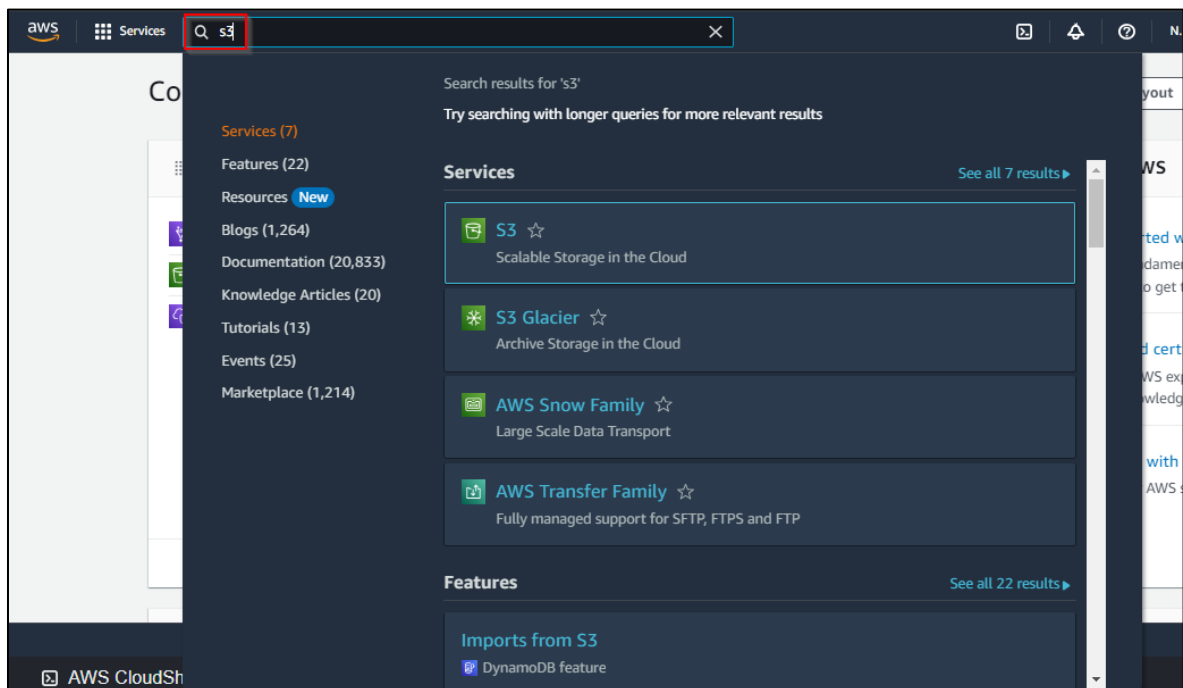
Prerequisites: AWS Lab access with an AWS account created

Steps to be followed:

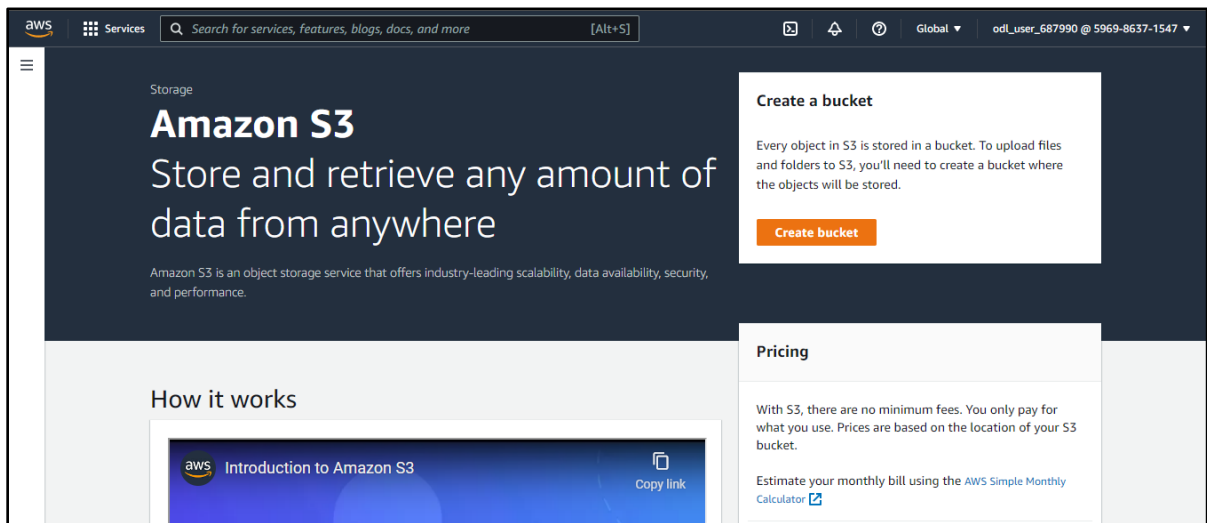
1. Create S3 buckets
2. Execute queries in Athena

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 a bucket named **my-s3-bucket-access-log-batch-operation** to aggregate log files from the **s3-bucket-batch-operation-encryption** bucket. Enable **Bucket versioning**.

The screenshot displays the 'General configuration' section of the 'Create bucket' wizard. The 'Bucket name' field is highlighted with a red rectangle and contains the text 's3-bucket-batch-operation-encryption'. Below the field, a note states: 'Bucket name must be unique within the global namespace and follow the bucket naming rules. [See rules for bucket naming](#)'. The 'AWS Region' dropdown menu is set to 'US East (N. Virginia) us-east-1'. At the bottom, there's an optional section titled 'Copy settings from existing bucket - optional' with the subtext 'Only the bucket settings in the following configuration are copied.' and a 'Choose bucket' button.

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 accidental deletions and application failures. [Learn more](#)

Bucket Versioning

☐ Disable

☒ Enable

Buckets (2) Info

↻

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 22, 2022, 14:39:50 (UTC+05:30)
<input type="radio"/>	s3-bucket-batch-operation-encryption	US East (N. Virginia) us-east-1	Bucket and objects not public	July 22, 2022, 14:40:27 (UTC+05:30)

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

Amazon S3 > Buckets

▶ Account snapshot

View Storage Lens dashboard

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

Buckets (2) Info

↻

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**

Amazon S3 > Buckets > s3-bucket-batch-operation-encryption

s3-bucket-batch-operation-encryption [Info](#)

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

Bucket overview

AWS Region	Amazon Resource Name (ARN)	Created
US East (N. Virginia) us-east-1	arn:aws:s3::s3-bucket-batch-operation-encryption	Aug 1, 2023

Bucket Versioning

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every versioning, you can easily recover from both unintended user actions and application failures. [Learn more](#)

[Edit](#)

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

Edit server access logging [Info](#)

Server access logging
Log requests for access to your bucket. [Learn more](#)

Server access logging

☐ Disable

☒ **Enable**

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.

Target bucket

[Browse S3](#)

Format: s3://bucket/prefix

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

1.7 Choose the log-batch file path by clicking **Choose path**

The dialog box titled "Choose destination to upload resources" shows a list of S3 Buckets. The first bucket, "my-s3-bucket-access-log-batch-operation", is selected. The second bucket, "s3-bucket-batch-operation-encryption", is also listed. The "Choose path" button is highlighted in orange.

Name	AWS Region
my-s3-bucket-access-log-batch-operation	US East (N. Virginia) us-east-1
s3-bucket-batch-operation-encryption	US East (N. Virginia) us-east-1

1.8 Click **Save changes**

The "Edit server access logging" page shows the "Server access logging" section. The "Enable" radio button is selected. A warning message states: "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." The "Target bucket" field contains "s3://my-s3-bucket-access-log-batch-operation". The "Save changes" button is highlighted in orange.

Server access logging
Log requests for access to your bucket. [Learn more](#)

Server access logging

☐ Disable

☒ Enable

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.

Target bucket

s3://my-s3-bucket-access-log-batch-operation [Browse S3](#)

Format: s3://bucket/prefix

1.9 Add files (HTML, PNG, and PEM) to the bucket using **Add files**

Amazon S3 > Buckets > s3-bucket-batch-operation-encryption > Upload

Upload [Info](#)

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose **Add files** or **Add folder**.

Files and folders (3 Total, 212.6 KB) Remove **Add files** Add folder

All files and folders in this table will be uploaded.

<input type="checkbox"/>	Name	Folder	Type	Size
<input type="checkbox"/>	index.html	-	text/html	1.7 KB
<input type="checkbox"/>	Screenshot (1).png	-	image/png	209.2 KB
<input type="checkbox"/>	Test2.pem	-	-	1.6 KB

s3-bucket-batch-operation-encryption [Info](#)

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

Objects (3)

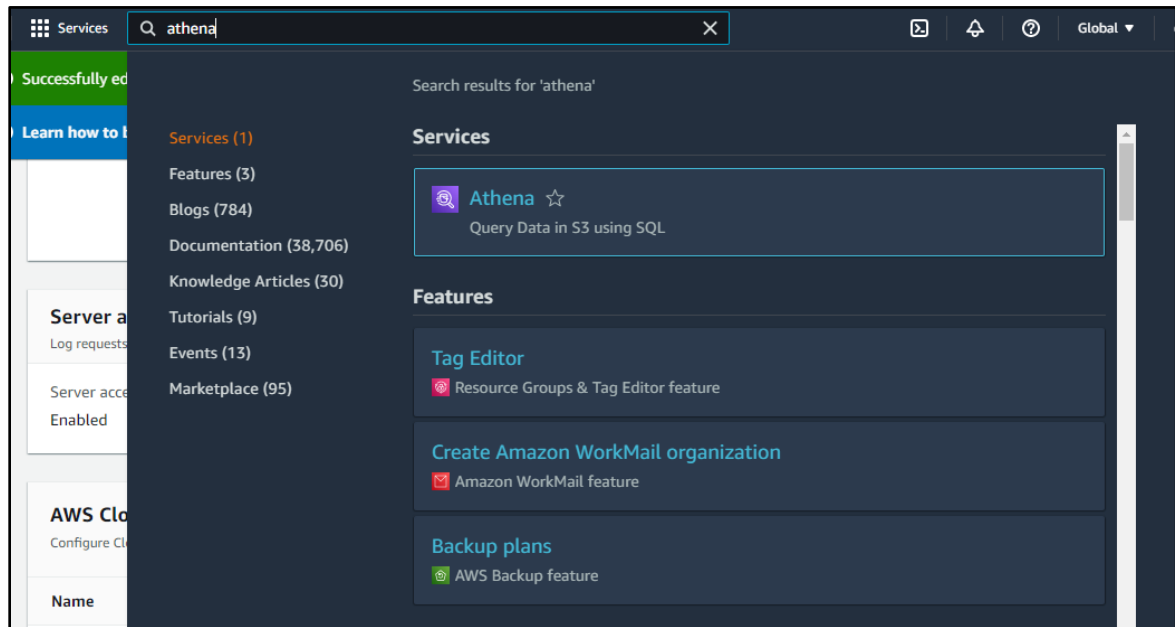
Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 Inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

☐ Show versions < 1 >

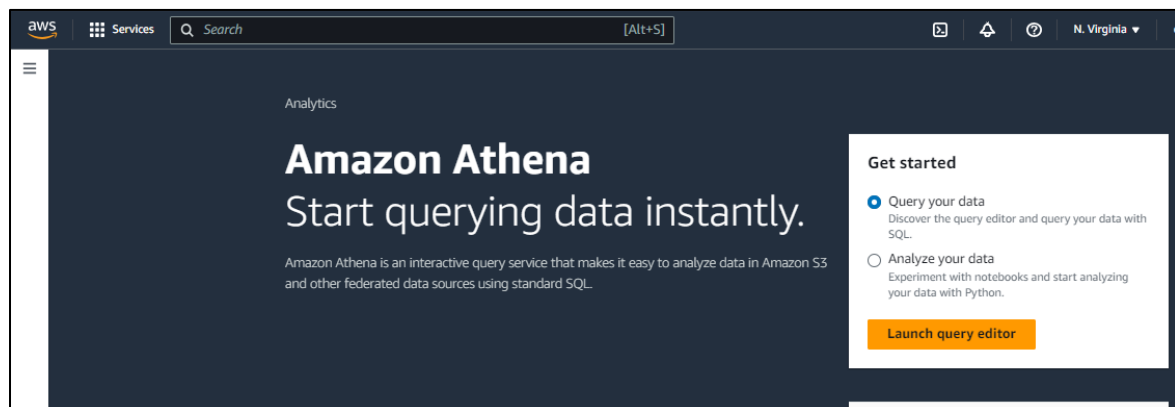
<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	index.html	html	July 22, 2022, 15:01:52 (UTC+05:30)	163.0 B	Standard
<input type="checkbox"/>	key.pem	pem	July 22, 2022, 15:01:54 (UTC+05:30)	1.6 KB	Standard
<input type="checkbox"/>	Resource Group Error.jpg	jpg	July 22, 2022, 15:01:50 (UTC+05:30)	19.7 KB	Standard

Step 2: Execute queries in Athena

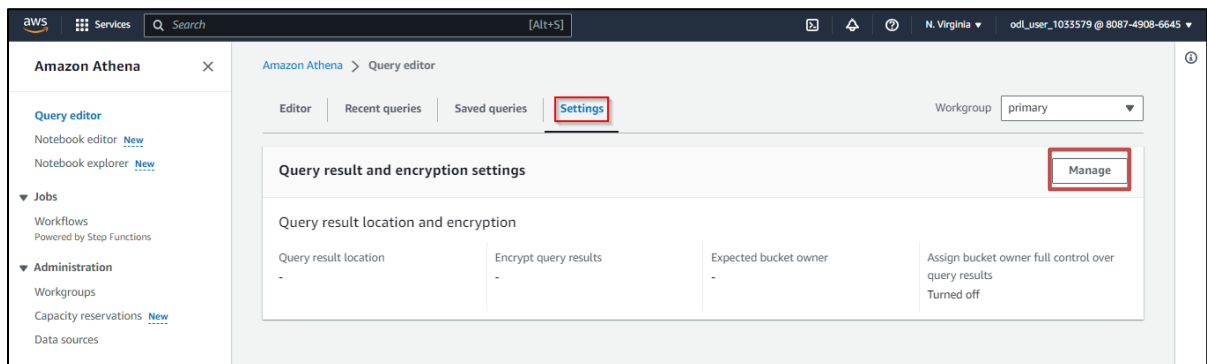
2.1 Navigate to the AWS Management Console and search for **Athena**



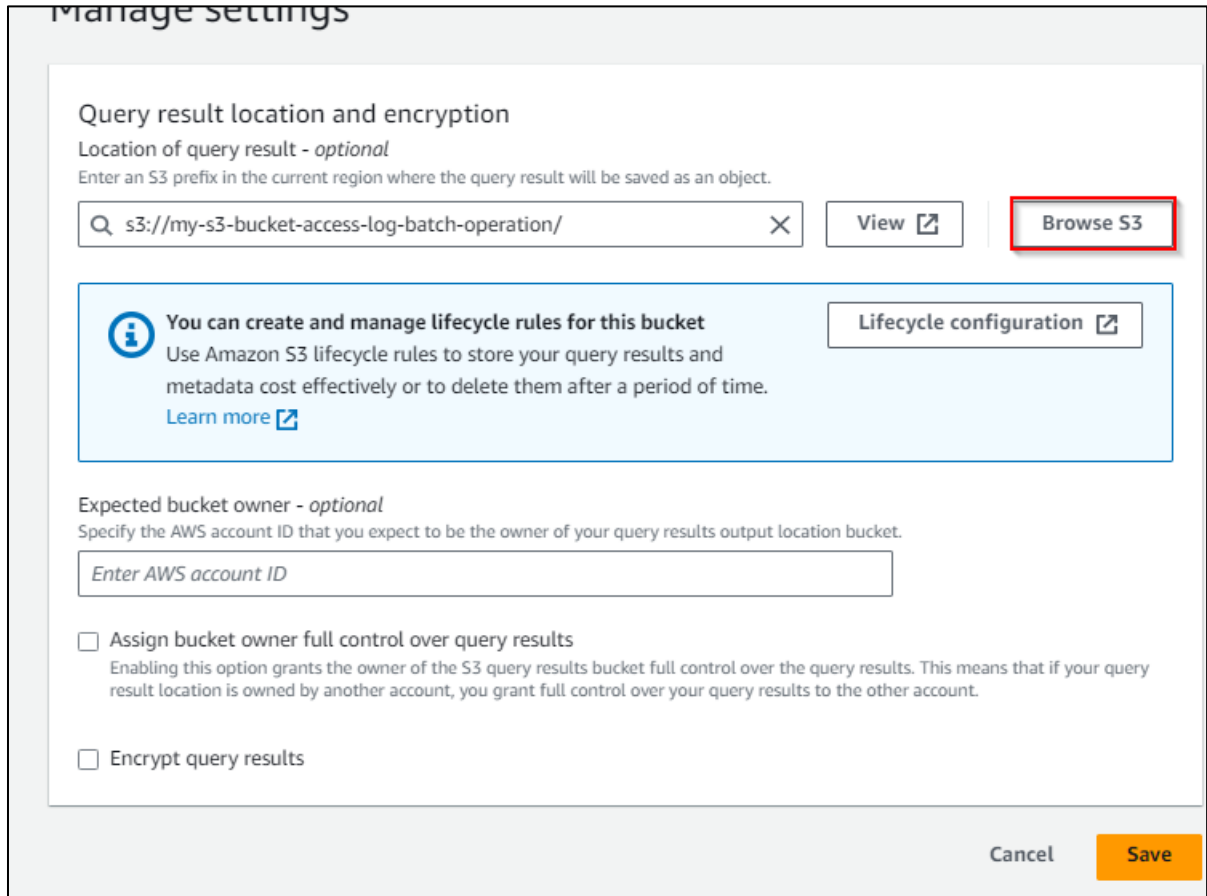
2.2 Launch the query editor in Athena

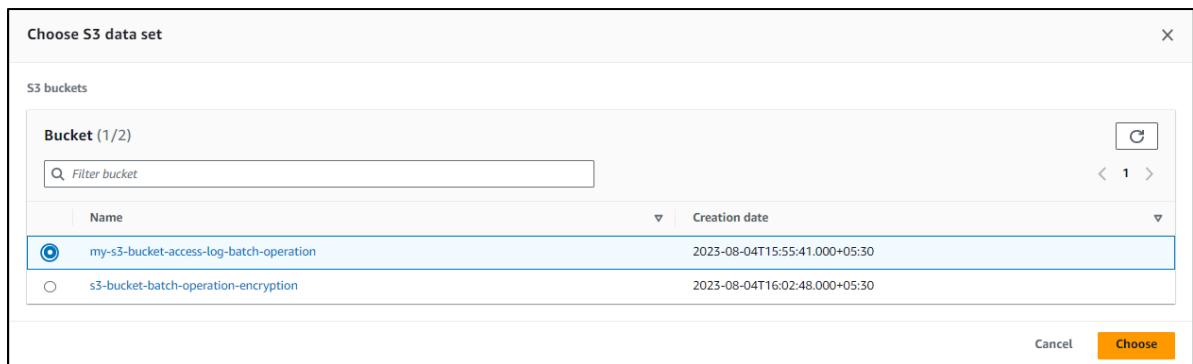


2.3 Click **Settings > Manage**

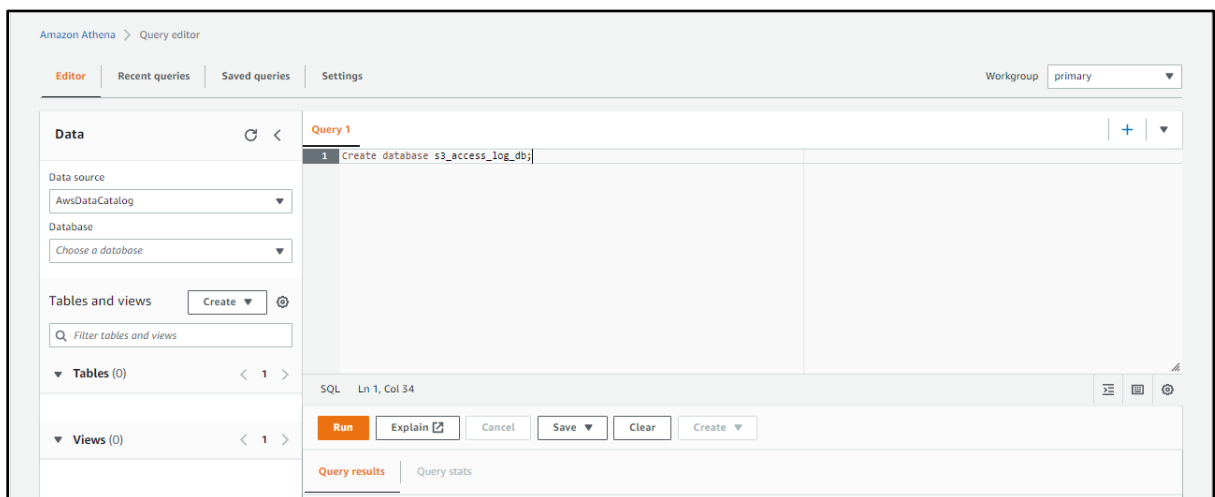


2.4 Click on **Browse S3**, select one of the path locations, and then click **Save**

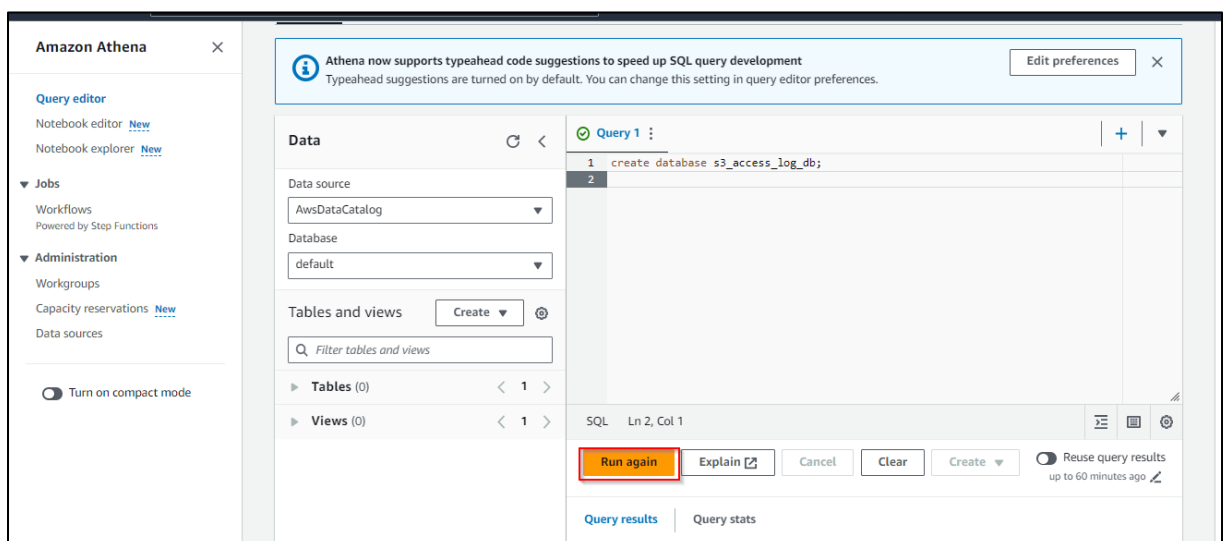




2.5 In the editor, execute the following query:
create database s3_access_log_db;



2.6 Click on **Run again**



```
CREATE EXTERNAL TABLE IF NOT EXISTS s3_access_log_db.mybucket_logs(
  BucketOwner STRING,
  Bucket STRING,
  RequestDateTime STRING,
  RemoteIP STRING,
  Requester STRING,
  RequestID STRING,
  Operation STRING,
  Key STRING,
  RequestURI_operation STRING,
  RequestURI_key STRING,
  RequestURI_httpProtoversion STRING,
  HTTPstatus STRING,
  ErrorCode STRING,
  BytesSent BIGINT,
  ObjectSize BIGINT,
  TotalTime STRING,
  TurnAroundTime STRING,
  Referrer STRING,
  UserAgent STRING,
  VersionId STRING,
  HostId STRING,
  SigV STRING,
  CipherSuite STRING,
  AuthType STRING,
  EndPoint STRING,
  TLSVersion STRING
)
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.RegexSerDe'
WITH SERDEPROPERTIES (
  'serialization.format' = '1', 'input.regex' = '([^\s]*) ([^\s]*) \\[([.]*?)\\] ([^\s]*) ([^\s]*) ([^\s]*) ([^\s]*) ([^\s]*) \\\"([^\s]*) ([^\s]*) (- |([^\s]*)?)\\\" (-|([0-9]*) ([^\s]*) ([^\s]*) ([^\s]*) ([^\s]*) ([^\s]*) ([^\s]*) \\\"([^\s]*) ([^\s]*)\\\" ([^\s]*) (? : ([^\s]*) ([^\s]*) ([^\s]*) ([^\s]*) ([^\s]*) ([^\s]*) ([^\s]*) )? . *$' )
LOCATION 's3://my-s3-bucket-access-log-batch-operation/'
```

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

Amazon Athena

Query editor

Notebook editor [New](#)

Notebook explorer [New](#)

Jobs

Workflows

Powered by Step Functions

Administration

Workgroups

Capacity reservations [New](#)

Data sources

Turn on compact mode

Data

Data source: AwsDataCatalog

Database: **s3_access_log_db**

Tables and views: Create

Filter tables and views

Tables (1)

mybucket_logs

Views (0)

Query 1

```
1 CREATE EXTERNAL TABLE IF NOT EXISTS s3_access_log_db.mybucket_logs(
2   BucketOwner STRING,
3   Bucket STRING,
4   RequestDateTime STRING,
5   RemoteIP STRING,
6   Requester STRING,
7   RequestID STRING,
8   Operation STRING,
9   Key STRING,
10  RequestURI_operation STRING,
11  RequestURI_key STRING,
12  RequestURI_httpProtoversion STRING,
13  HTTPstatus STRING,
14  ErrorCode STRING,
15  BytesSent BIGINT,
```

SQL Ln 33, Col 1

Run again Explain Cancel Clear Create

Reuse query results up to 60 minutes ago

Query results Query stats

Completed Time in queue: 93 ms Run time: 347 ms Data scanned: -

Amazon Athena

Query editor

Notebook editor [New](#)

Notebook explorer [New](#)

Jobs

Workflows

Powered by Step Functions

Administration

Workgroups

Capacity reservations [New](#)

Data sources

Turn on compact mode

Data

Data source: AwsDataCatalog

Database: s3_access_log_db

Tables and views: Create

Filter tables and views

Tables (1)

mybucket_logs

Views (0)

Query 1

```
1 CREATE EXTERNAL TABLE IF NOT EXISTS s3_access_log_db.mybucket_logs(
2   BucketOwner STRING,
3   Bucket STRING,
4   RequestDateTime STRING,
5   RemoteIP STRING,
6   Requester STRING,
7   RequestID STRING,
8   Operation STRING,
9   Key STRING,
10  RequestURI_operation STRING,
11  RequestURI_key STRING,
12  RequestURI_httpProtoversion STRING,
13  HTTPstatus STRING,
14  ErrorCode STRING,
15  BytesSent BIGINT,
```

SQL Ln 33, Col 1

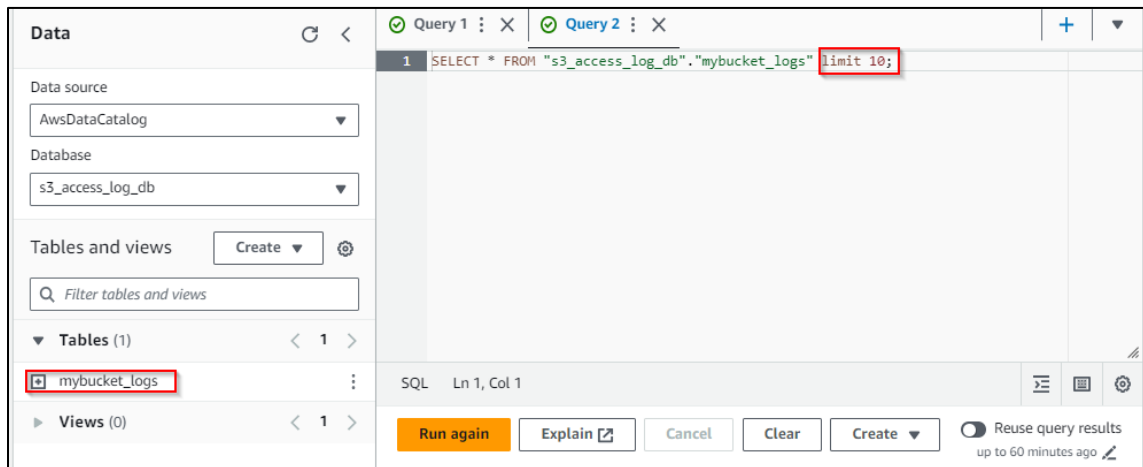
Run again Explain Cancel Clear Create

Reuse query results up to 60 minutes ago

Query results Query stats

Completed Time in queue: 93 ms Run time: 347 ms Data scanned: -

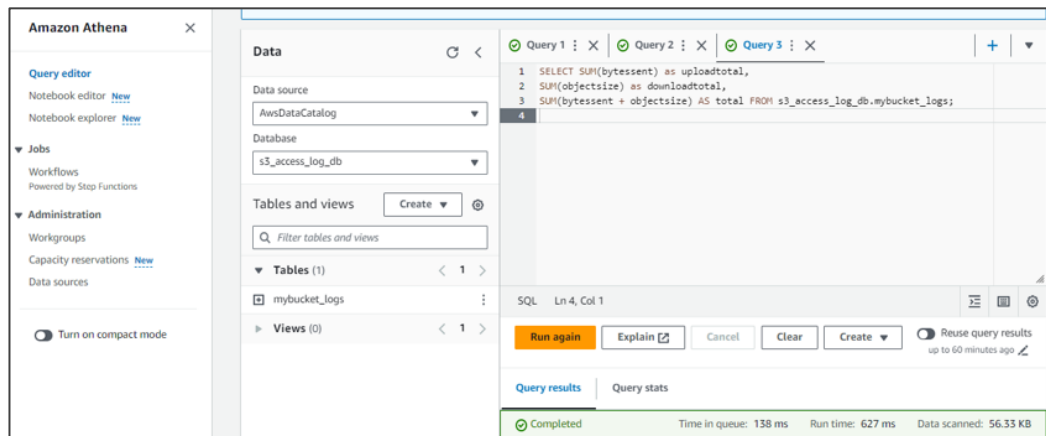
Query successful.



Preview the table: A predefined query with a limit of 10 will be available. Remove the limit to view the complete table data.

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, enabling you to analyze operations on a designated S3 bucket and get insights into data usage patterns.