**Lesson 4 Demo 10**

**Query with Athena**

**Objectives:** To run a query in Athena to perform some operations on a specific S3 bucket.

**Prerequisites:** AWS Lab access with an AWS account created

**Steps to be followed:**

1. Login to your AWS lab.
2. Create two S3 buckets.
3. Creating and executing queries in Athena.

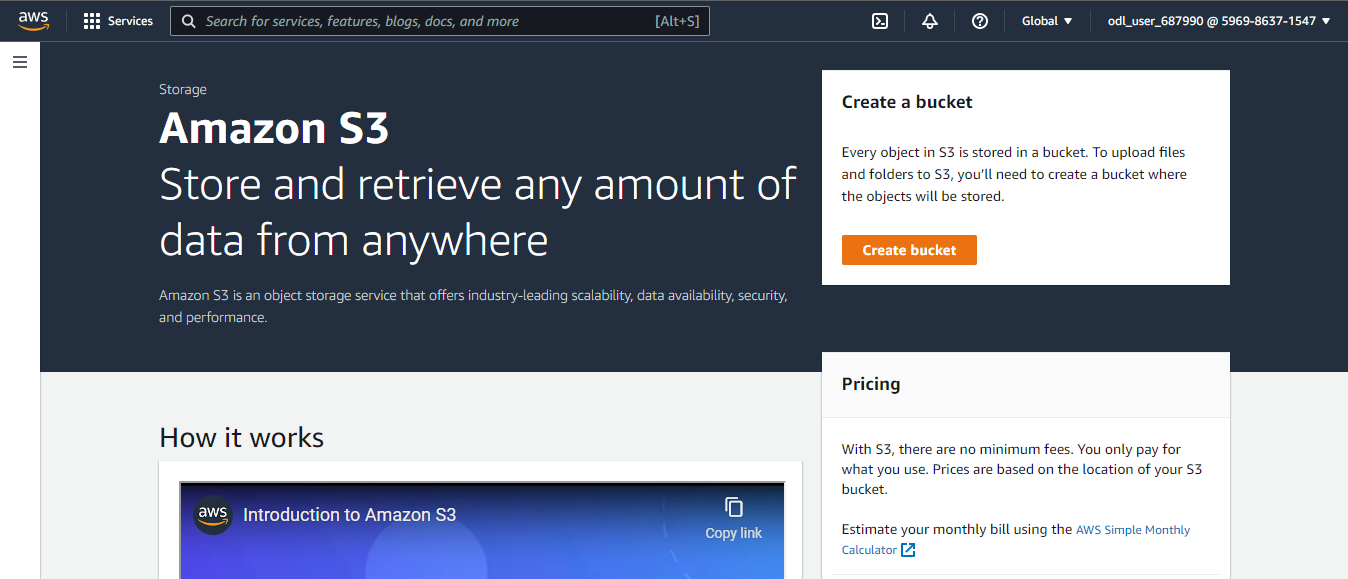
**Step 1: Login to your AWS lab.**

Graphical user interface, text, application

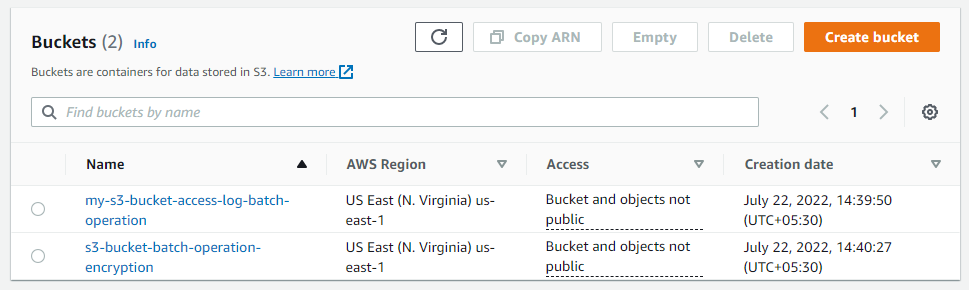
Description automatically generated

**Step 2: Create two S3 buckets.**

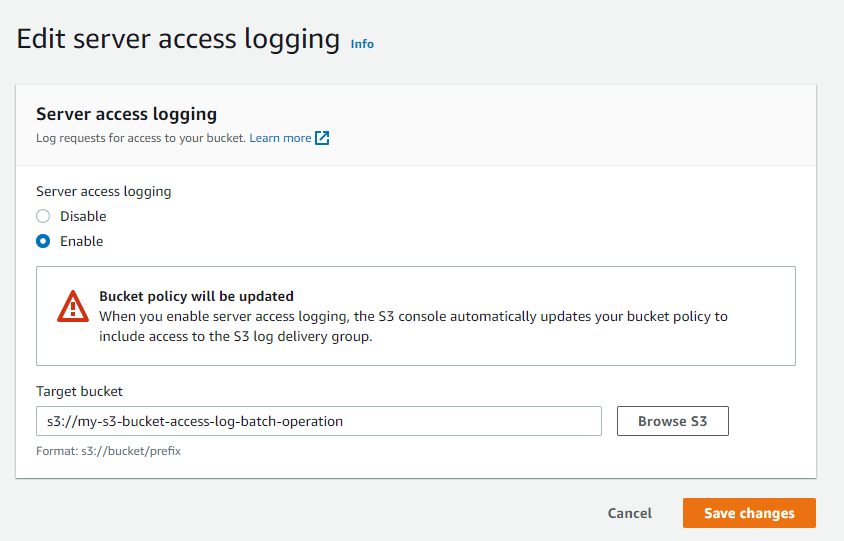
* 1. Select S3 from the AWS search bar.



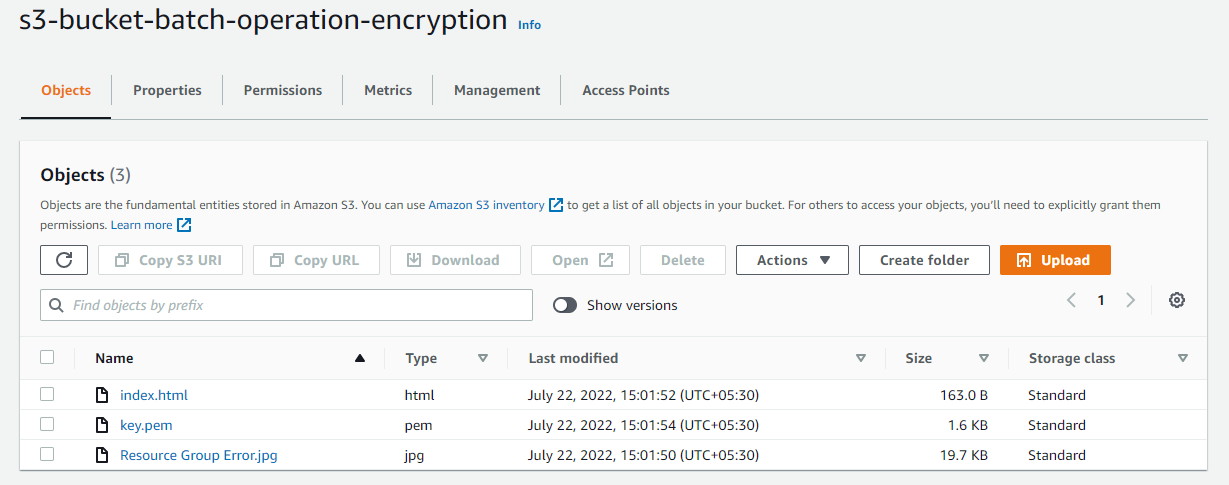
* 1. Create a bucket named **my-s3-bucket-access-log-batch-operation** to add all the log files of the **s3-bucket-batch-operation-encryption** bucket.

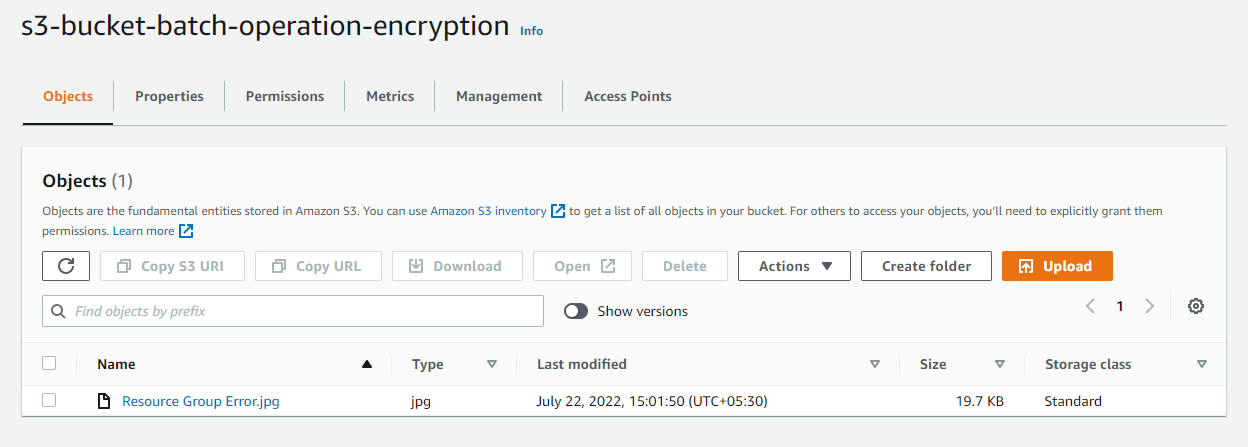


* 1. Add logging to the bucket **s3-bucket-batch-operation-encryption.**
  2. Select the bucket and click **Properties**.
  3. Select **Server access logging** and click the **Edit** option.



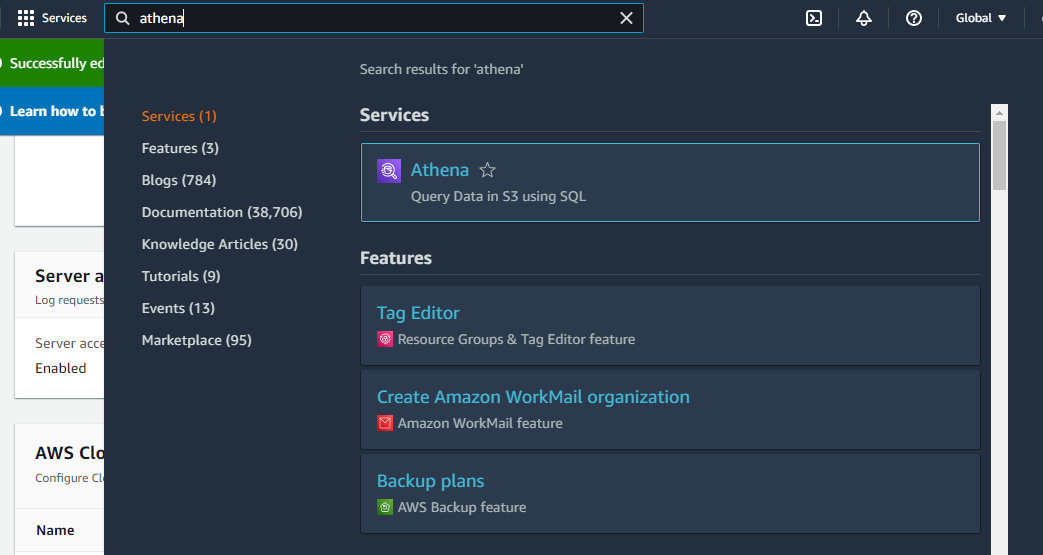
* 1. Click **Save changes**.
  2. Before we perform any operation using Athena, we should have one log file at least. Let’s add files to the monitored bucket and then delete them.





**Step 3: Creating and executing queries in Athena.**

* 1. Go to the search bar in **Services**, and search for Athena



* 1. set up a query in Athena to run on the S3 bucket.

Graphical user interface, website

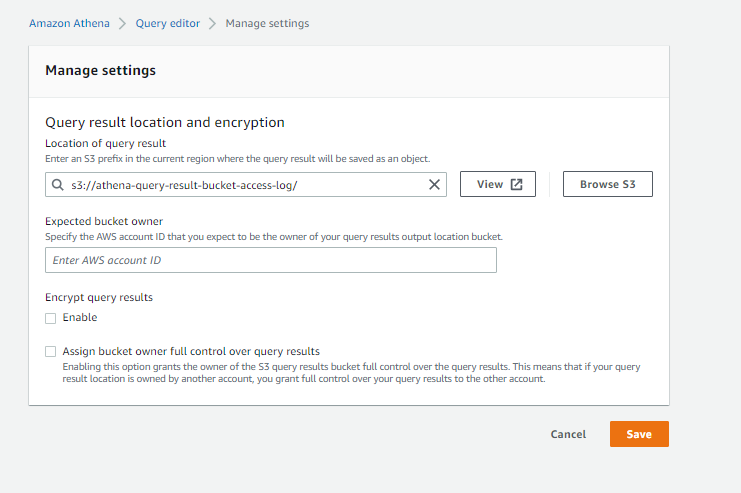
Description automatically generated

* 1. Click on **Settings**

Graphical user interface, application, Teams

Description automatically generated

* 1. Give the query result location in the format.



* 1. Create a **database**

Graphical user interface, text, application

Description automatically generated

* 1. Click on the **Database** tab, and you will find the new database under it.

Graphical user interface, text, application

Description automatically generated

* 1. Execute the query given below:

**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**

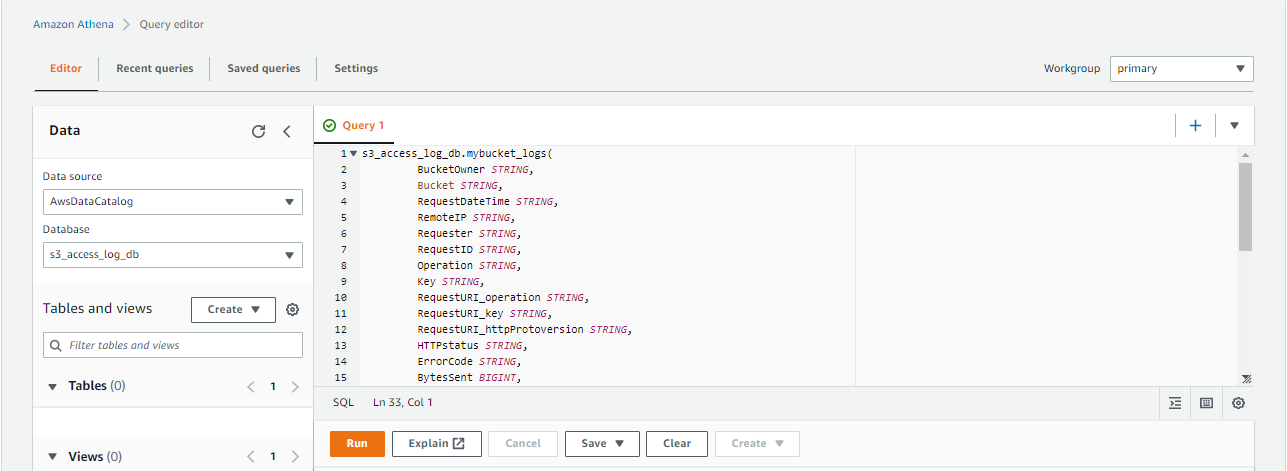
**)**

**ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.RegexSerDe'**

**WITH SERDEPROPERTIES (**

**'serialization.format' = '1', 'input.regex' = '([^ ]\*) ([^ ]\*) \\[(.\*?)\\] ([^ ]\*) ([^ ]\*) ([^ ]\*) ([^ ]\*) ([^ ]\*) \\\"([^ ]\*) ([^ ]\*) (- |[^ ]\*)\\\" (-|[0-9]\*) ([^ ]\*) ([^ ]\*) ([^ ]\*) ([^ ]\*) ([^ ]\*) ([^ ]\*) (\"[^\"]\*\") ([^ ]\*)(?: ([^ ]\*) ([^ ]\*) ([^ ]\*) ([^ ]\*) ([^ ]\*) ([^ ]\*))?.\*$' )**

**LOCATION 's3://my-s3-bucket-access-log-batch-operation/'**



* 1. Execute the query.

Graphical user interface, text, application

Description automatically generated

* 1. When you preview the table, it automatically gives you a predefined query that has a defined limit set to 10. If you remove the limit, you will be able to view the entire table data.

Graphical user interface, text, application

Description automatically generated

* 1. Execute the query given below to find the exact amount of data uploaded to and downloaded from the monitored bucket:

**SELECT SUM(bytessent) as uploadtotal,**

**SUM(objectsize) as downloadtotal,**

**SUM(bytessent + objectsize) AS total FROM s3\_access\_log\_db.mybucket\_logs;**

Graphical user interface, text, application, email

Description automatically generated