

AWS Athena S3 EC2 ALB analysis for Accessing Logs

Analyzing ALB Access Logs with Amazon Athena

Step 1: Launch EC2 instance

Start by clicking “Launch Instance” on the EC2 Dashboard

The screenshot shows the AWS Management Console's 'Launch Instance' wizard. The 'Name and tags' section has 'Demo Server A' entered. The 'Application and OS Images (Amazon Machine Image)' section is active, showing a search bar and a grid of AMIs. 'Ubuntu Server 22.04 LTS (HVM), SSD Volume Type' is selected. The 'Summary' panel on the right shows: 1 instance, Canonical Ubuntu 22.04 LTS AMI (ami-0287a05f0ef0e9d9a), t2.micro instance type, new security group, and 1 volume (8 GiB). A 'Free tier' notification is visible. The 'Launch instance' button is highlighted.

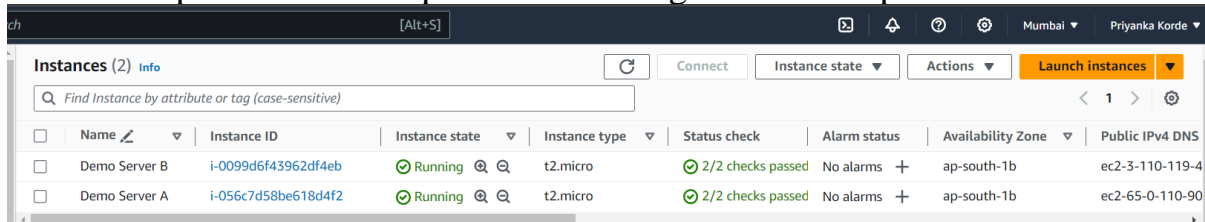
The screenshot shows the 'Configure Instance' step. The 'VPC' is set to 'vpc-048fab17f0190a94e' (default). 'Subnet' is 'No preference'. 'Auto-assign public IP' is 'Enable'. Under 'Firewall (security groups)', 'Create security group' is selected. The 'Security group name' is 'ALB Server A' and the 'Description' is 'ALB Server A'. Under 'Inbound Security Group Rules', 'Security group rule 1 (TCP, 22, 0.0.0.0/0)' is listed. The 'Summary' panel on the right is identical to the previous step. The 'Launch instance' button is highlighted.

The screenshot shows the 'Review Instance' step. It displays the configured 'Inbound Security Group Rules' in detail: Rule 1 (TCP, 22, 0.0.0.0/0) for SSH and Rule 2 (TCP, 80, 0.0.0.0/0) for HTTP. A warning message states: 'Rules with source of 0.0.0.0/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.' The 'Summary' panel on the right is identical to the previous steps. The 'Launch instance' button is highlighted.

Finally, select “**launch instance**” to create the server.

To launch another instance identical to Server-A and name it Server-B:

- Create a new EC2 instance using the same configuration as Server-A.
- During the setup, assign the name “Server-B” to the instance.
- Complete the launch process following the same steps used for Server-A.



	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS
<input type="checkbox"/>	Demo Server B	i-0099d6f43962df4eb	Running	t2.micro	2/2 checks passed	No alarms	ap-south-1b	ec2-3-110-119-4
<input type="checkbox"/>	Demo Server A	i-056c7d58be618d4f2	Running	t2.micro	2/2 checks passed	No alarms	ap-south-1b	ec2-65-0-110-90

Connect the both instance and run commands:

```
sudo apt-get update
sudo apt-get upgrade
sudo apt-get install apache2
cd var/www/html
sudo rm index.html
sudo vi html.html
```

Put simple html code inside index file

Server A: `<html> <h1>Hello, This is Priyanka Korde from Server A !</h1> </html>`

Server B: `<html> <h1>Hello, This is Priyanka Korde from Server B !</h1> </html>`

Open your browser and paste the public IPv4 address of the Server-A

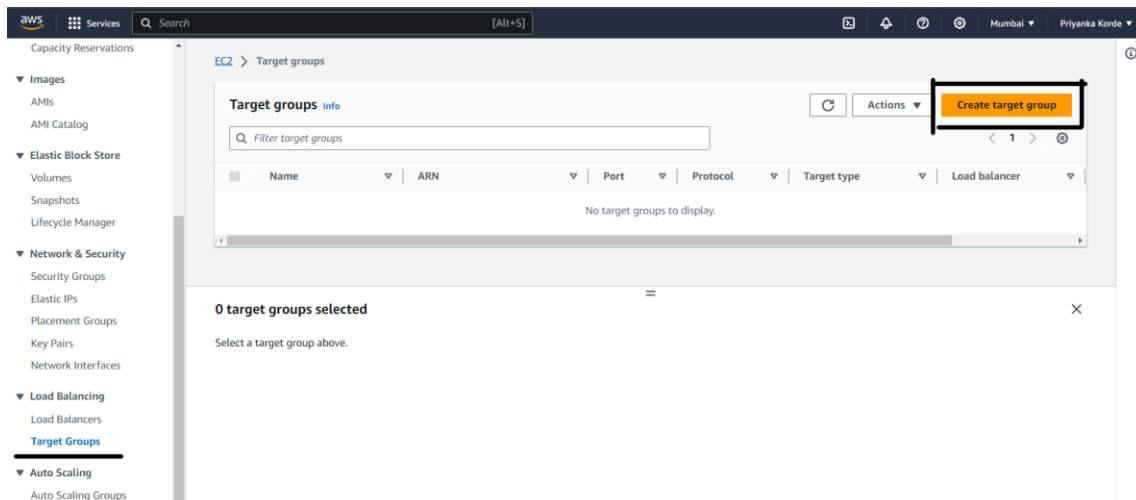


Also verify that Server B is also functioning correctly.

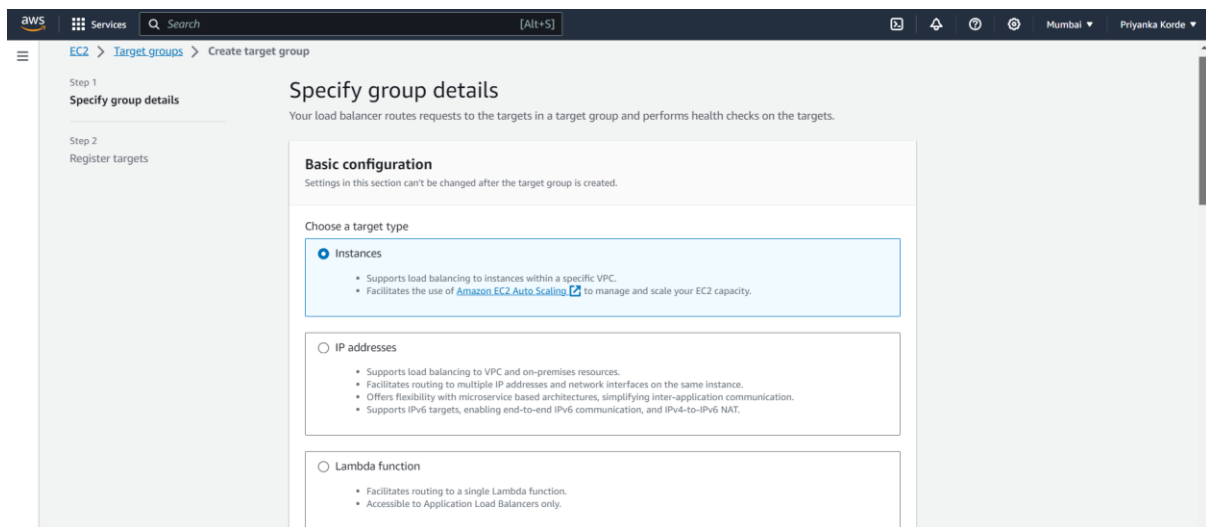


Step 2: Create Target Group

In the EC2 Dashboard, locate the “Load Balancing” section in the left-hand navigation pane and select “Target Groups.”



Click the “Create target group” button to start the target group creation process , ensure that you have selected the instances as **target type**



- **Name:** Enter a descriptive name for your target group.
- **Protocol:** Choose the protocol (HTTP, HTTPS, etc.) based on your application’s needs.
- **Port:** Specify the port that your instances are listening on.
- **VPC:** Choose the Virtual Private Cloud (VPC) where your instances are located.

Target group name

MyTargetGroup

A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.

Protocol : Port

HTTP 80

1-65535

IP address type

Only targets with the indicated IP address type can be registered to this target group.

☒ IPv4

Each instance has a default network interface (eth0) that is assigned the primary private IPv4 address. The instance's primary private IPv4 address is the one that will be applied to the target.

☐ IPv6

Each instance you register must have an assigned primary IPv6 address. This is configured on the instance's default network interface (eth0). [Learn more](#)

VPC

Select the VPC with the instances that you want to include in the target group. Only VPCs that support the IP address type selected above are available in this list.

-

vpc-048fab17f0190a94e

IPv4: 172.31.0.0/16

Protocol version

☒ HTTP1

Send requests to targets using HTTP/1.1. Supported when the request protocol is HTTP/1.1 or

- **Health Checks:** Define health check settings, including the path as **index.html** health check protocol and then click on next

Health checks

The associated load balancer periodically sends requests, per the settings below, to the registered targets to test their status.

Health check protocol

HTTP

Health check path

Use the default path of "/" to ping the root, or specify a custom path if preferred.

/index.html

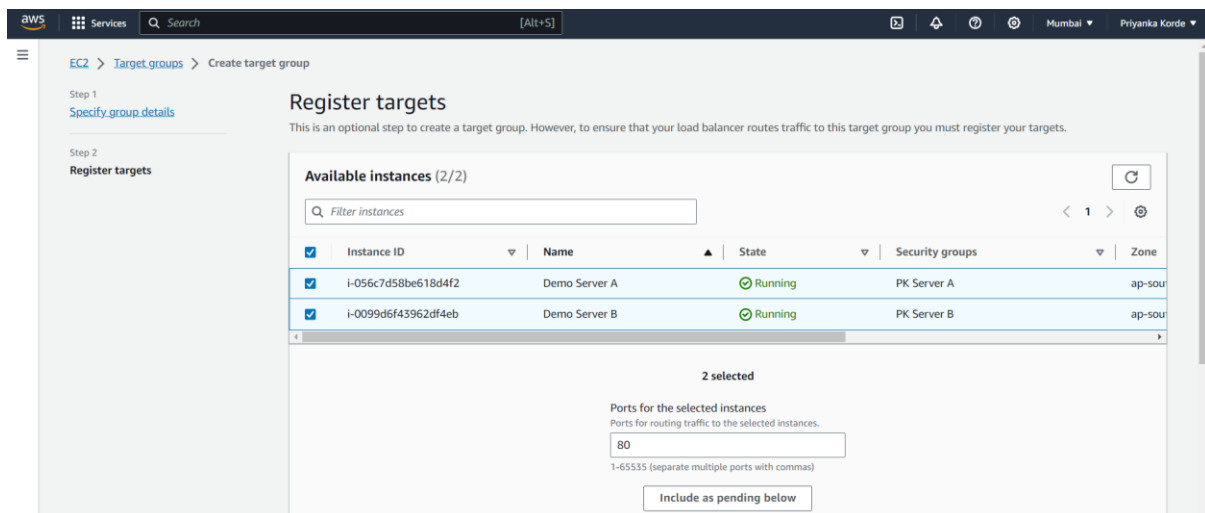
Up to 1024 characters allowed.

► Advanced health check settings

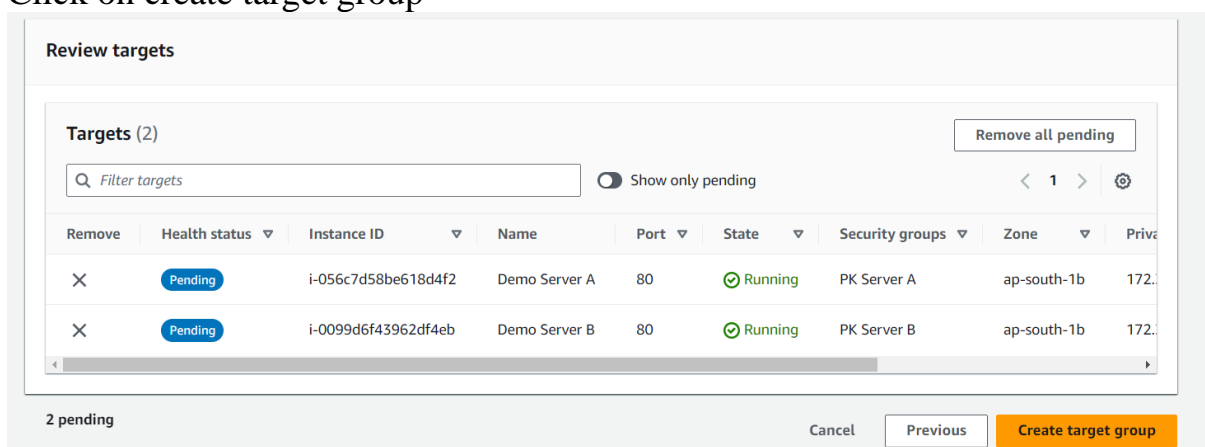
Register Targets

Under the “Targets” section, you’ll need to register instances that should be part of this target group.

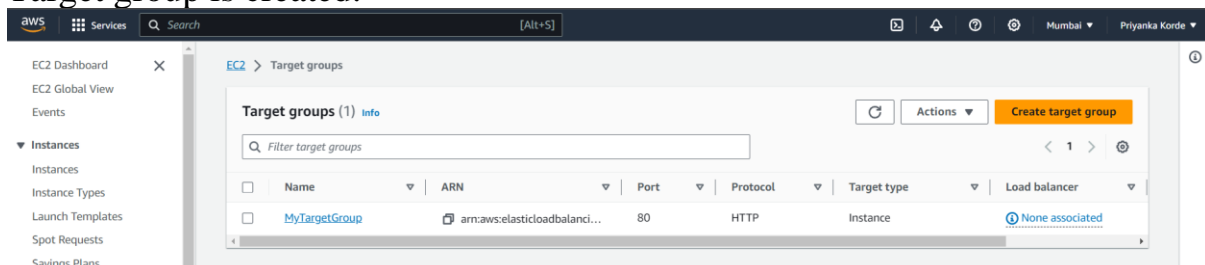
Select the instances by their instance ID or IP address and click on **include as pending below**



Click on create target group

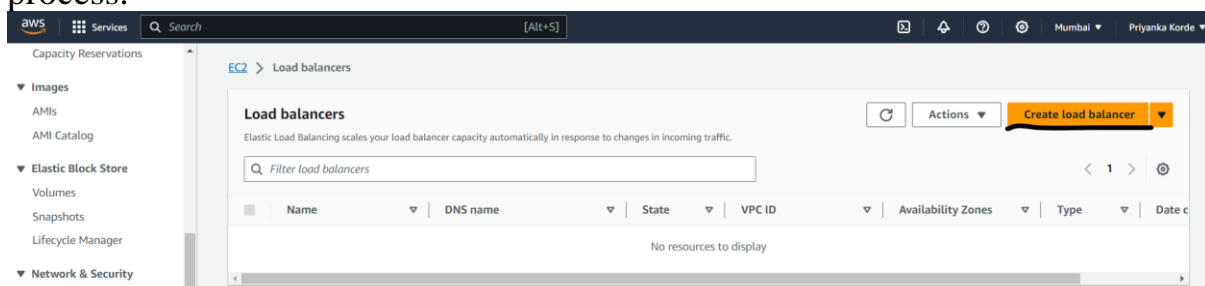


Target group is created.



Step 3 : Create a Load Balancer

Click the “Create Load Balancer” button to start the load balancer creation process.



Select the type of load balancer you want to create:
Choose the appropriate type for your use case and click “Create.”

EC2 > Load balancers > Compare and select load balancer type

Compare and select load balancer type

A complete feature-by-feature comparison along with detailed highlights is also available. [Learn more](#)

Load balancer types

Application Load Balancer

Choose an Application Load Balancer when you need a flexible

Network Load Balancer

Choose a Network Load Balancer when you need ultra-high

Gateway Load Balancer

Choose a Gateway Load Balancer when you need to deploy and

Load balancer name

Name must be unique within your AWS account and can't be changed after the load balancer is created.

A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.

Scheme

Scheme can't be changed after the load balancer is created.

☒ Internet-facing

An internet-facing load balancer routes requests from clients over the internet to targets. Requires a public subnet. [Learn more](#)

☐ Internal

An internal load balancer routes requests from clients to targets using private IP addresses.

IP address type

Select the type of IP addresses that your subnets use.

☒ IPv4

Recommended for internal load balancers.

☐ Dualstack

Includes IPv4 and IPv6 addresses.

Network mapping

The load balancer routes traffic to targets in the selected subnets, and in accordance with your IP address settings.

VPC

Select the virtual private cloud (VPC) for your targets or you can [create a new VPC](#). Only VPCs with an internet gateway are enabled for selection. The selected VPC can't be changed after the load balancer is created. To confirm the VPC for your targets, view your [target groups](#).

-

vpc-048fab17f0190a94e

IPv4: 172.31.0.0/16

+

Select the virtual private cloud (VPC) for your targets or you can [create a new VPC](#). Only VPCs with an internet gateway are enabled for selection. The selected VPC can't be changed after the load balancer is created. To confirm the VPC for your targets, view your [target groups](#).

-

vpc-048fab17f0190a94e

IPv4: 172.31.0.0/16

+

Mappings

Select at least two Availability Zones and one subnet per zone. The load balancer routes traffic to targets in these Availability Zones only. Availability Zones that are not supported by the load balancer or the VPC are not available for selection.

☒ ap-south-1a (aps1-az1)

Subnet

subnet-0376fd1bd78784148

IPv4 address

Assigned by AWS

☒ ap-south-1b (aps1-az3)

Subnet

subnet-04a70e935b6f99900

IPv4 address

Assigned by AWS

☒ ap-south-1c (aps1-az2)

Subnet

subnet-0067bbc64d51d9b16

IPv4 address

Assigned by AWS

Security groups info

A security group is a set of firewall rules that control the traffic to your load balancer. Select an existing security group, or you can [create a new security group](#).

Security groups

Select up to 5 security groups

- PK Server A sg-03db9abb973abe6e6 VPC: vpc-048fab17f0190a94e
- PK Server B sg-031f7dc6356766e9f VPC: vpc-048fab17f0190a94e
- default sg-02dc9b2e3230b6995 VPC: vpc-048fab17f0190a94e

Listeners and routing info

A listener is a process that checks for connection requests using the port and protocol you configure. The rules that you define for a listener determine how the load balancer routes requests to its registered targets.

▼ Listener HTTP:80 Remove

Protocol: HTTP Port: 80 1-65535

Default action: Forward to MyTargetGroup (Target type: Instance, IPv4) HTTP

[Create target group](#)

Listener tags - optional

Review the configuration details to ensure they are correct. Once you're satisfied, click the **“Create”** button.

example, you can have Key = production-webserver, or Key = webserver, and Value = production.

Summary

Review and confirm your configurations. [Estimate cost](#)

Basic configuration Edit

ec2-athena-alb

- Internet-facing
- IPv4

Security groups Edit

- default sg-02dc9b2e3230b6995

Network mapping Edit

VPC vpc-048fab17f0190a94e

- ap-south-1a subnet-0376f41b478784148
- ap-south-1b subnet-04a70e935b6f99900
- ap-south-1c subnet-0067b0c64d51d9b16

Listeners and routing Edit

- HTTP:80 (defaults to PKTargetGroup)

Add-on services Edit

None

Tags Edit

None

Attributes

Certain default attributes will be applied to your load balancer. You can view and edit them after creating the load balancer.

Cancel Create load balancer

You should see a confirmation message indicating that your load balancer has been created successfully.

Your load balancer is now ready to distribute traffic across the registered instances in the target group, providing high availability and scalability for your application. Make sure to update your DNS or application configurations to point to the load balancer's DNS name or IP address.

EC2 > Load balancers

Load balancers (1)

Elastic Load Balancing scales your load balancer capacity automatically in response to changes in incoming traffic.

Filter load balancers

	Name	DNS name	State	VPC ID	Availability Zones	Type	Date c
<input type="checkbox"/>	ec2-athena-alb	ec2-athena-alb-16502599...	Active	vpc-048fab17f0190a94e	3 Availability Zones	application	Octob

Upon refreshing the page, you will receive a response from Server A and B, demonstrating that the load balancer successfully directs traffic to both EC2

instances, ensuring redundancy and optimal load distribution.

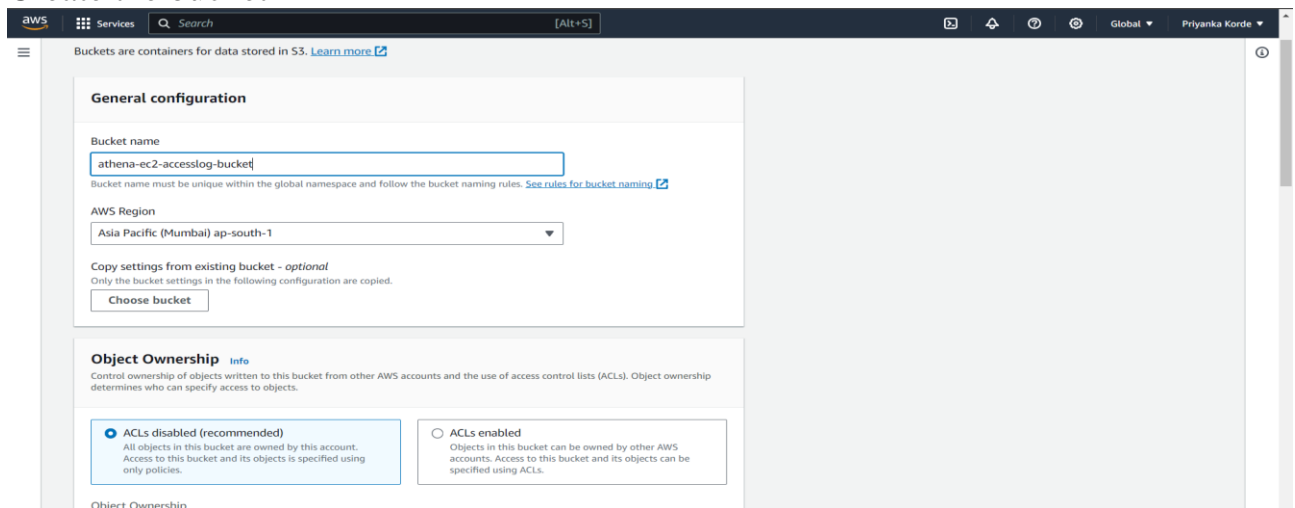


Step 3 : Configure Access Logs

To enabling access logs for your Application Load Balancer (ALB), you'll also need to create an Amazon S3 bucket and set up a policy to allow the ALB to write logs to the bucket.

Navigate to the S3 service

Create the bucket



Default encryption Info
Server-side encryption is automatically applied to new objects stored in this bucket.

Encryption type Info

- ☒ Server-side encryption with Amazon S3 managed keys (SSE-S3)
- ☐ Server-side encryption with AWS Key Management Service keys (SSE-KMS)
- ☐ Dual-layer server-side encryption with AWS Key Management Service keys (DSSE-KMS)
Secure your objects with two separate layers of encryption. For details on pricing, see [DSSE-KMS pricing on the Storage tab of the Amazon S3 pricing page](#).

Bucket Key
Using an S3 Bucket Key for SSE-KMS reduces encryption costs by lowering calls to AWS KMS. S3 Bucket Keys aren't supported for DSSE-KMS. [Learn more](#)

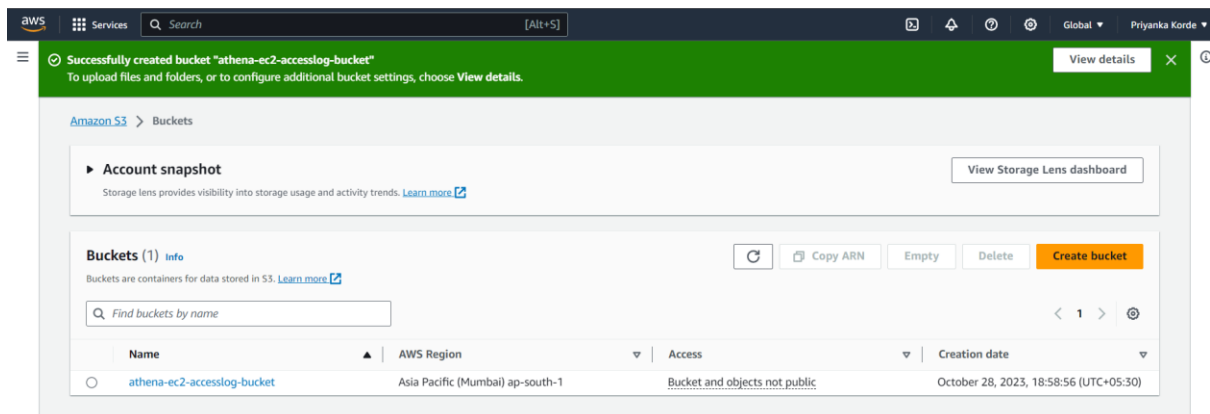
- ☐ Disable
- ☒ Enable

► **Advanced settings**

ⓘ After creating the bucket, you can upload files and folders to the bucket, and configure additional bucket settings.

Cancel **Create bucket**

Review your bucket settings, and if everything looks correct, click the “Create bucket” button.



The bucket creation process has been completed successfully.
Once the bucket is created create 3 folders /prefix /Awslogs/ (youraccountid) for logs.

Step 4: Set Up a Bucket Policy

After creating the S3 bucket, you'll need to set up a bucket policy to allow the ALB to write logs to it:

In the bucket policy editor, you'll need to define a policy that grants the necessary permissions to the ALB.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::account-id:root"
      },
      "Action": "s3:PutObject",
      "Resource": "my-s3-arn"
    }
  ]
}
```

```
}
/
}
```

- Replace account-id with the ID of the AWS account for Elastic Load Balancing for your Region: Asia Pacific (Mumbai) – 718504428378
- Replace *my-s3-arn* with the ARN of the location for your access logs.
arn:aws:s3::bucket-name/prefix/AWSLogs/aws-account-id/*

Bucket policy Edit Delete

The bucket policy, written in JSON, provides access to the objects stored in the bucket. Bucket policies don't apply to objects owned by other accounts. [Learn more](#)

Public access is blocked because Block Public Access settings are turned on for this bucket
To determine which settings are turned on, check your Block Public Access settings for this bucket. [Learn more about using Amazon S3 Block Public Access](#)

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::718504428378:root"
      },
      "Action": "s3:PutObject",
      "Resource": "arn:aws:s3:::athena-ec2-accesslog-bucket/prefix/AWSLogs/413313710573/*"
    }
  ]
}
```

Copy

Step 5: Enabling logs for an Application Load Balancer

Click on the name of your Application Load Balancer (ALB) for which you want to enable access logs.

Inside the ALB details page, navigate to the “Attributes” tab and find the “Access logs” section.

AWS Services Search [Alt+S] Mumbai Priyanka Korde

Capacity Reservations

- ▼ Images
 - AMIs
 - AMI Catalog
- ▼ Elastic Block Store
 - Volumes
 - Snapshots
 - Lifecycle Manager
- ▼ Network & Security
 - Security Groups
 - Elastic IPs
 - Placement Groups
 - Key Pairs
 - Network Interfaces
- ▼ Load Balancing
 - Load Balancers
 - Target Groups
- ▼ Auto Scaling
 - Auto Scaling Groups

Details

Load balancer type Application	Status Active	VPC vpc-048fab17f0190a94e	IP address type IPv4
Scheme Internet-facing	Hosted zone ZP97RAFLXTNZK	Availability Zones subnet-0376fd1bd78784148 ap-south-1a (aps1-az1) subnet-04a70e935b6f99900 ap-south-1b (aps1-az3) subnet-0067bbc64d51d9b16 ap-south-1c (aps1-az2)	Date created October 28, 2023, 18:47 (UTC+05:30)
Load balancer ARN arn:aws:elasticloadbalancing:ap-south-1:413313710573:loadbalancer/app/ec2-athena-alb/b0f59e957a25868c		DNS name Info ec2-athena-alb-1650259997.ap-south-1.elb.amazonaws.com (A Record)	

Listeners and rules Network mapping Security Monitoring Integrations **Attributes** Tags

Attributes Edit

Traffic configuration

Click on “Edit”

Click the “Edit” button to configure access logs.

Enable Access Logs

In the “Access logs” section, do the following:
Navigate to the Monitoring section by scrolling down the page

X-Forwarded-For header
Enables you to append, preserve, or remove the X-Forwarded-For header in the HTTP request before the Application Load Balancer sends the request to the target.

☒ Append
☐ Preserve
☐ Remove

☒ Client port preservation
Indicates whether the X-Forwarded-For header should preserve the source port that the client used to connect to the load balancer.

☒ Preserve host header
Indicates whether the Application Load Balancer should preserve the Host header in the HTTP request and send it to targets without any change.

Monitoring

☒ Access logs
Access logs deliver detailed logs of all requests made to your Elastic Load Balancer. Choose an existing S3 location. If you don't specify a prefix, the access logs are stored in the root of the bucket. Additional charges apply. [Learn more](#)

S3 URI

Format: s3://bucket/prefix/object.

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

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

Specify the S3 bucket where you want to store the access logs. You can choose an existing bucket or create a new one.

Choose an S3 bucket

S3 buckets

Buckets (1/1)

Name	Creation date
<input checked="" type="radio"/> athena-ec2-accesslog-bucket	October 28, 2023, 19:21 (UTC+05:30)

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

Monitoring

☒ Access logs
Access logs deliver detailed logs of all requests made to your Elastic Load Balancer. Choose an existing S3 location. If you don't specify a prefix, the access logs are stored in the root of the bucket. Additional charges apply. [Learn more](#)

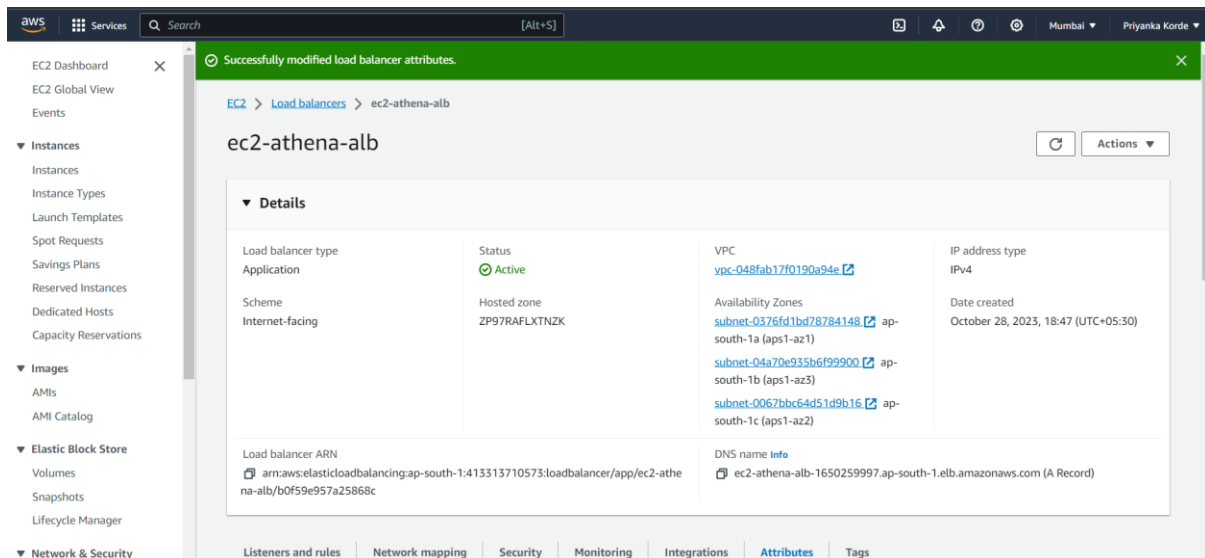
S3 URI

Format: s3://bucket/prefix/object.

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

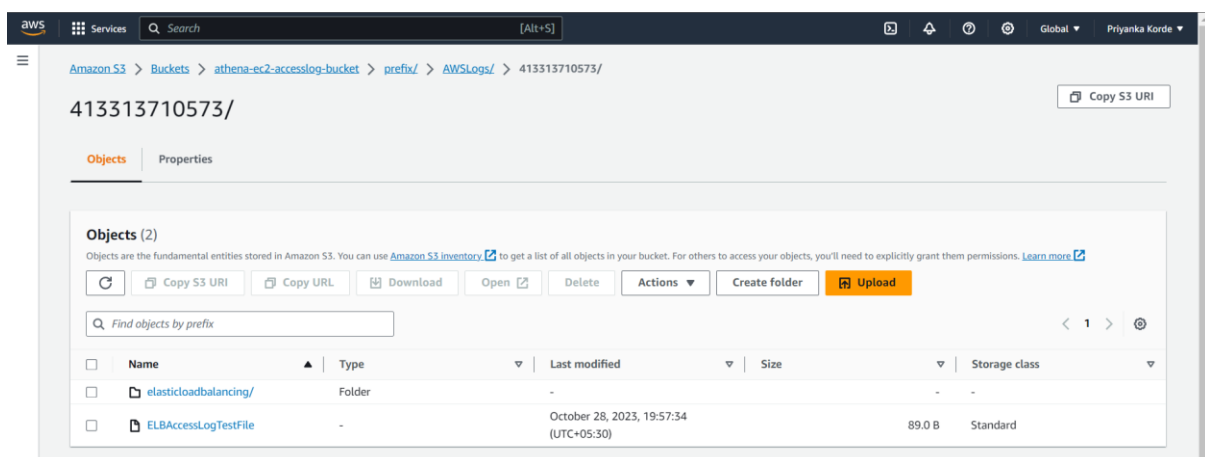
[Cancel](#) [Save changes](#)

Click the “Save” button to save your access log configuration.

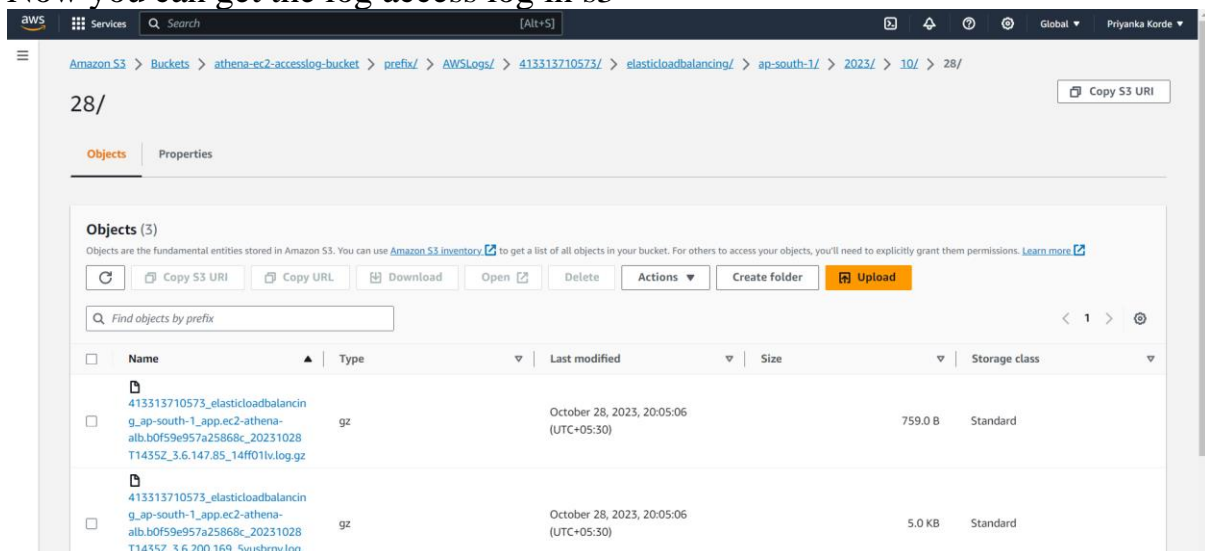


From this point forward, your ALB will start generating access logs and storing them in the specified S3 bucket. You can use these logs for monitoring and analysis of traffic to your load balancer.

Now check in the s3 bucket you will receive the [ELBAccessLogTestFile](#)

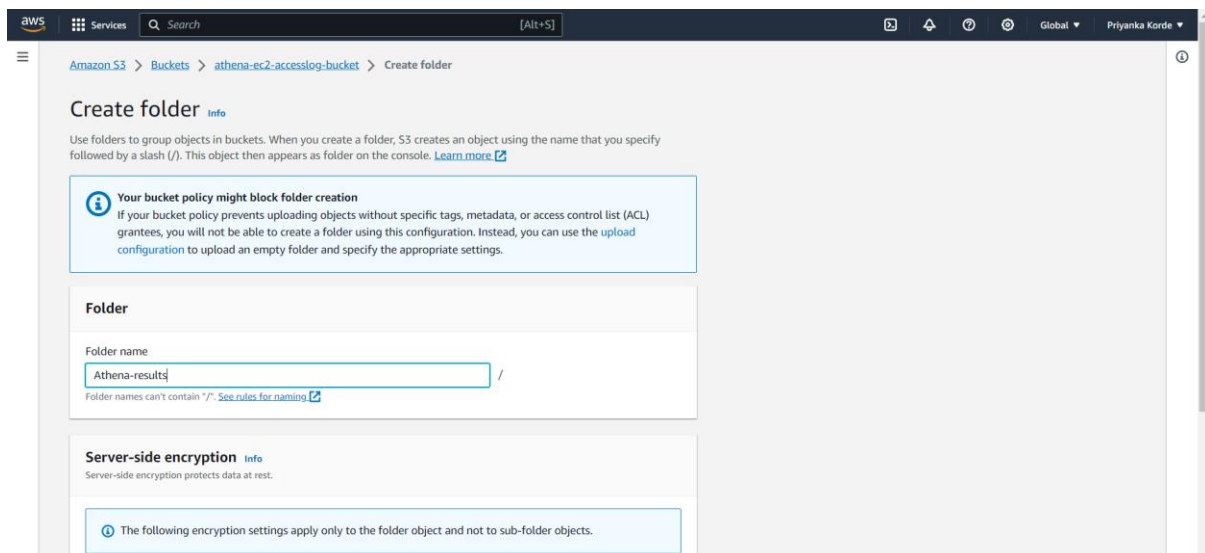


Now you can get the log access log in s3

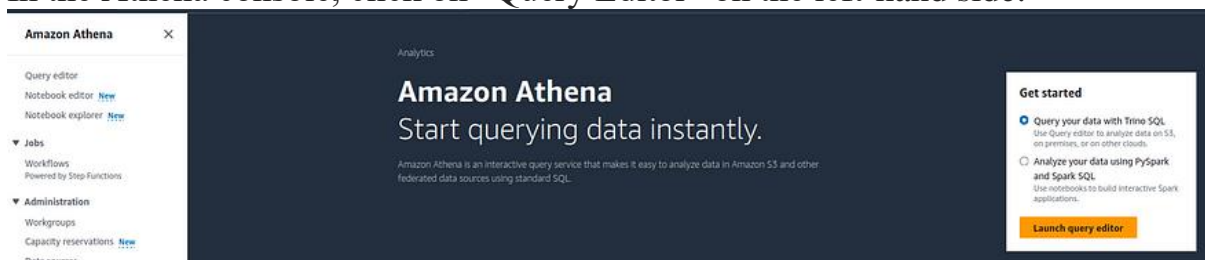


Step 6: Querying AWS Application Load Balancer (ALB) access logs using Amazon Athena:

- Navigate to the Athena service by clicking on “Services” and then selecting “Athena” under the Analytics section.
- Make sure your ALB access logs are being delivered to the S3 bucket, as previously configured.
- Create a folder in the same s3 bucket for the Athena query results.



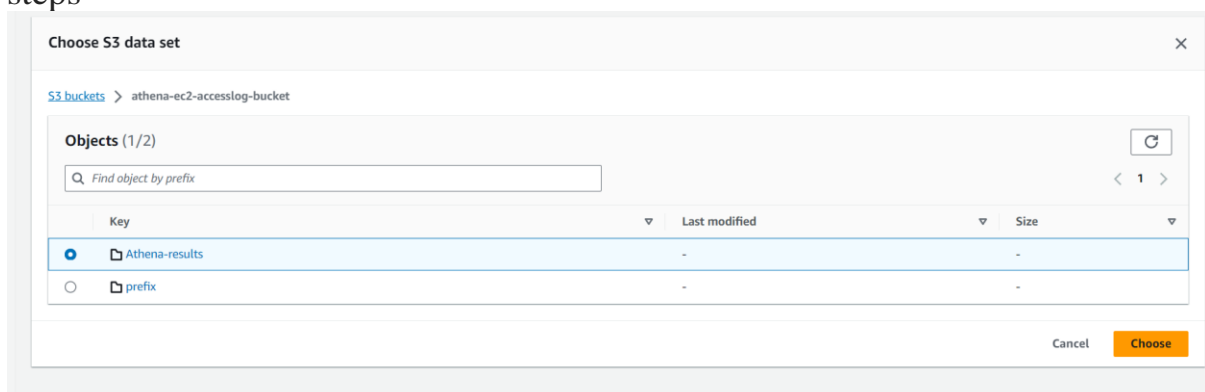
In the Athena console, click on “Query Editor” on the left-hand side.



Click on settings and manage

“Access the settings menu and then choose ‘Manage.’ From there, select the S3 bucket you’ve designated for storing the ALB logs.”

click on **browse** select the bucket and folder and which we have created earlier steps



Click on Choose and Save 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.

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](#)

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**

Amazon Athena

Query editor
Notebook editor [New](#)
Notebook explorer [New](#)

Jobs
Workflows
Powered by Step Functions

Administration
Workgroups
Data sources

Settings successfully updated.

Amazon Athena > Query editor

Editor | Recent queries | Saved queries | **Settings**

Workgroup: primary

Query result and encryption settings

Query result location and encryption

Query result location s3://athena-ec2-accesslog-bucket/Athena-results/	Encrypt query results -	Expected bucket owner -	Assign bucket owner full control over query results Turned off
---	----------------------------	----------------------------	---

Amazon Athena

Query editor
Notebook editor [New](#)
Notebook explorer [New](#)

Jobs
Workflows
Powered by Step Functions

Administration
Workgroups
Data sources

Workgroup query engine upgrade complete
One or more workgroups have been upgraded to Athena engine version 3. To see the workgroups that have been upgraded, use the [Workgroup list](#) page. For information about new features, see the [Athena Engine Version Reference](#).

Amazon Athena > Query editor

Editor | Recent queries | Saved queries | Settings

Workgroup: primary

Athena now supports typeahead code suggestions to speed up SQL query development
Typeahead suggestions are turned on by default. You can change this setting in query editor preferences.

Data

Data source:

Database:

Tables and views:

Tables (0) < 1 >
Views (0) < 1 >

Query 1

```
1 CREATE DATABASE SPriyanka
```

SQL Ln 1, Col 26

☐ Reuse query results

Step 8: Create a Table

After creating the database, you need to create a table that defines the structure of your ALB access logs. You can do this with a CREATE TABLE statement.

```
CREATE EXTERNAL TABLE IF NOT EXISTS alb_logs (
```

```
    type string,
```

```
    time string,
```

```
    elb string,
```

```
    client_ip string,
```

```
    client_port int,
```

```
    target_ip string,
```

```
    target_port int,
```

```
    request_processing_time double,
```

```
    target_processing_time double,
```

```
    response_processing_time double,
```

```
    elb_status_code int,
```

```
    target_status_code string,
```

```
    received_bytes bigint,
```

```
    sent_bytes bigint,
```

```
    request_verb string,
```

```
    request_url string,
```

```
    request_proto string,
```

```
    user_agent string,
```

```
    ssl_cipher string,
```

```
    ssl_protocol string,
```

```
    target_group_arn string,
```

```
    trace_id string,
```

```
    domain_name string,
```

```
    chosen_cert_arn string,
```

```
    matched_rule_priority string,
```

```

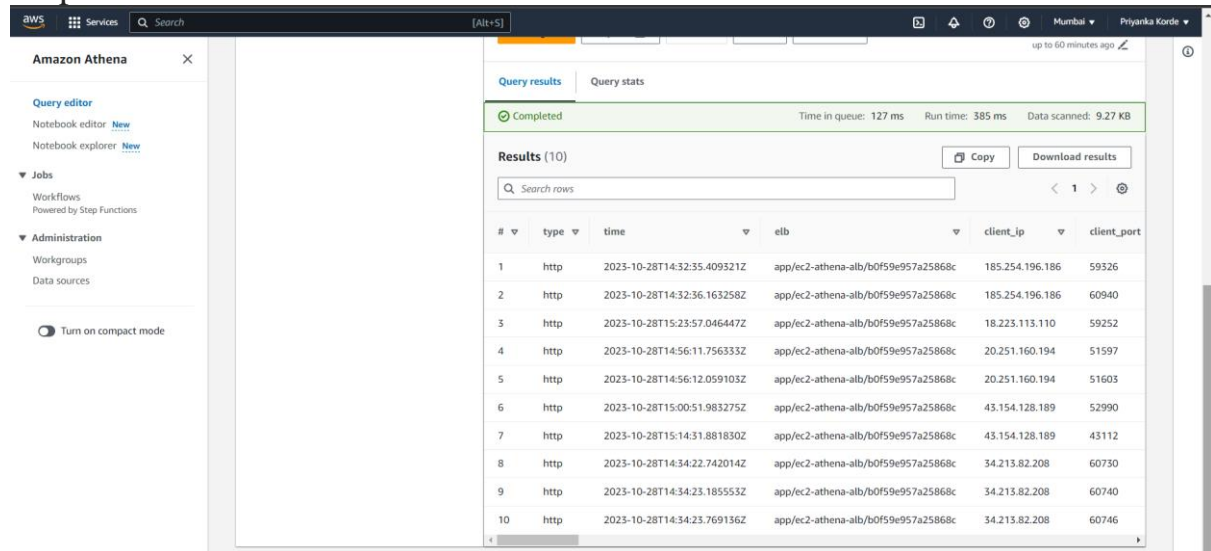
request_creation_time string,
actions_executed string,
redirect_url string,
lambda_error_reason string,
target_port_list string,
target_status_code_list string,
classification string,
classification_reason string
)
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.RegexSerDe'
WITH SERDEPROPERTIES (
'serialization.format' = '1',
'input.regex' =
'^(^ ]*)(^ ]*)(^ ]*)(^ ]*):([0-9]*) (^ ]*)[:-]([0-9]*) ([-.0-9]*) ([-.0-9]*) ([-
.0-9]*) ([-0-9]*) (-[-0-9]*) ([-0-9]*) ([-0-9]*) \"(^ ]*)(.*) (- [^ ]*)\" \"(^\" ]*)\"
([A-Z0-9- _]+) ([A-Za-z0-9- .]*) (^ ]*) \"(^\" ]*)\" \"(^\" ]*)\" \"(^\" ]*)\" ([-.0-9]*)
(^ ]*) \"(^\" ]*)\" \"(^\" ]*)\" \"(^ ]*)\" \"(^[\\s]+?)\" \"(^[\\s]+)\" \"(^ ]*)\" \"(^
]*)\")\"
LOCATION 's3://athena-ec2-accesslog-
bucket/prefix/AWSLogs/413313710573/elasticloadbalancing/ap-south-1/'

```

The screenshot shows the Amazon Athena console interface. On the left, there's a sidebar with navigation options like 'Query editor', 'Jobs', 'Workflows', 'Administration', 'Workgroups', and 'Data sources'. The main area displays a SQL query in a text editor. The query is a Hive-style query with a ROW FORMAT SERDE and a complex REGEX input regex. The query is successful and has completed. The status bar at the bottom shows 'Query successful.' and 'Completed' with a green checkmark. The query execution details show 'Time in queue: 57 ms', 'Run time: 270 ms', and 'Data scanned: -'.

To preview a table in Amazon Athena:

1. Click on the desired table name in the database list.
2. Open the table's menu by clicking the three dots next to the name.
3. Choose "Preview Table" to view a sample of the table's data in a new tab or panel.



query that performs a count of HTTP GET requests grouped by the client IP address:

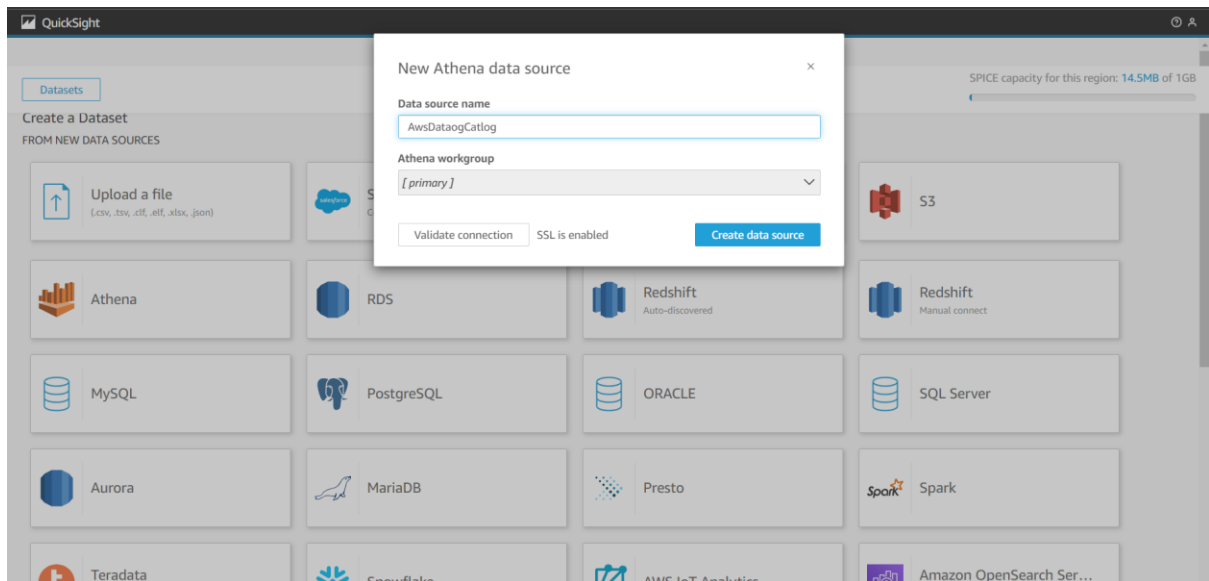
```
SELECT COUNT(request_verb) AS
count,
request_verb,
client_ip
FROM alb_logs
GROUP BY request_verb, client_ip
LIMIT 100;
```

The screenshot displays the Amazon Athena console interface. On the left, the navigation pane includes sections for 'Query editor' (Notebook editor, Notebook explorer), 'Jobs' (Workflows), and 'Administration' (Workgroups, Data sources). The main area is divided into 'Data' (Data source: AwsDataCatalog, Database: default) and 'Tables and views' (Filter tables and views). The 'Query editor' shows a SQL query: `SELECT COUNT(request_verb) AS count, request_verb, client_ip FROM alb_logs GROUP BY request_verb, client_ip LIMIT 100;`. Below the query, the 'Run again' button is highlighted. The 'Query results' tab shows a 'Completed' status with metrics: Time in queue: 241 ms, Run time: 714 ms, Data scanned: 10.42 KB. The 'Results (10)' section displays a table with 10 rows of data.

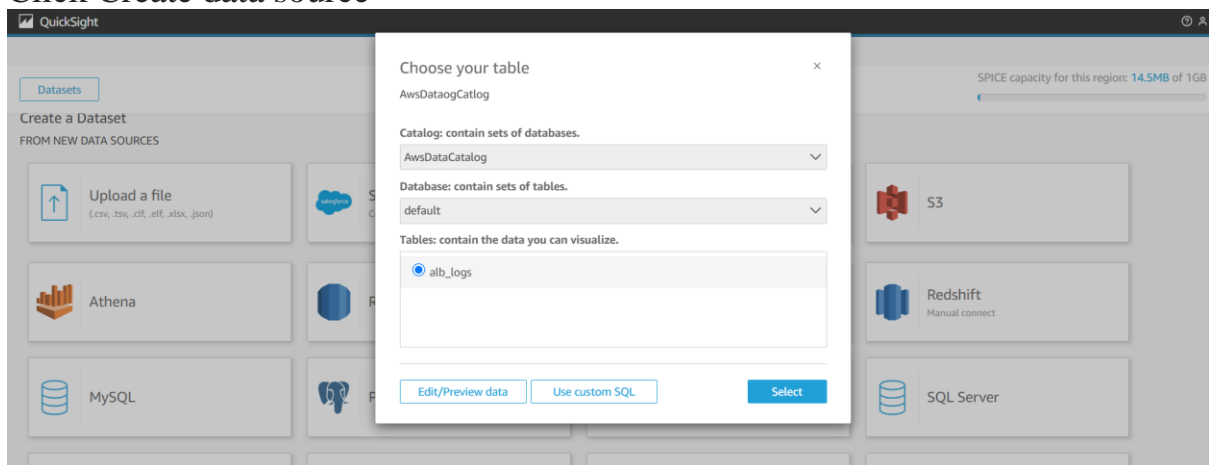
#	count	request_verb	client_ip
1	28	POST	34.213.82.208
2	1	GET	20.251.160.194
3	1	POST	20.251.160.194
4	2	GET	18.223.113.110
5	2	GET	167.248.133.185
6	1	-	167.248.133.185
7	2	GET	185.254.196.186
8	57	GET	34.213.82.208
9	2	HEAD	43.154.128.189
10	2	POST	167.248.133.185

Step 8: Quicksite Analysis

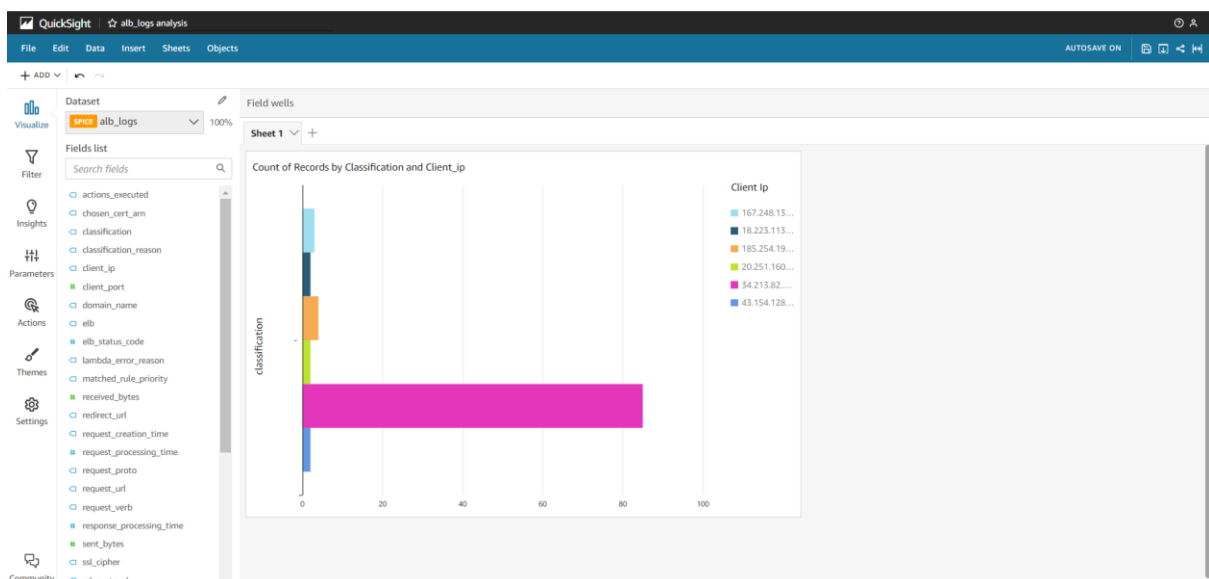
Navigate to New Analysis > New dataset > Athena

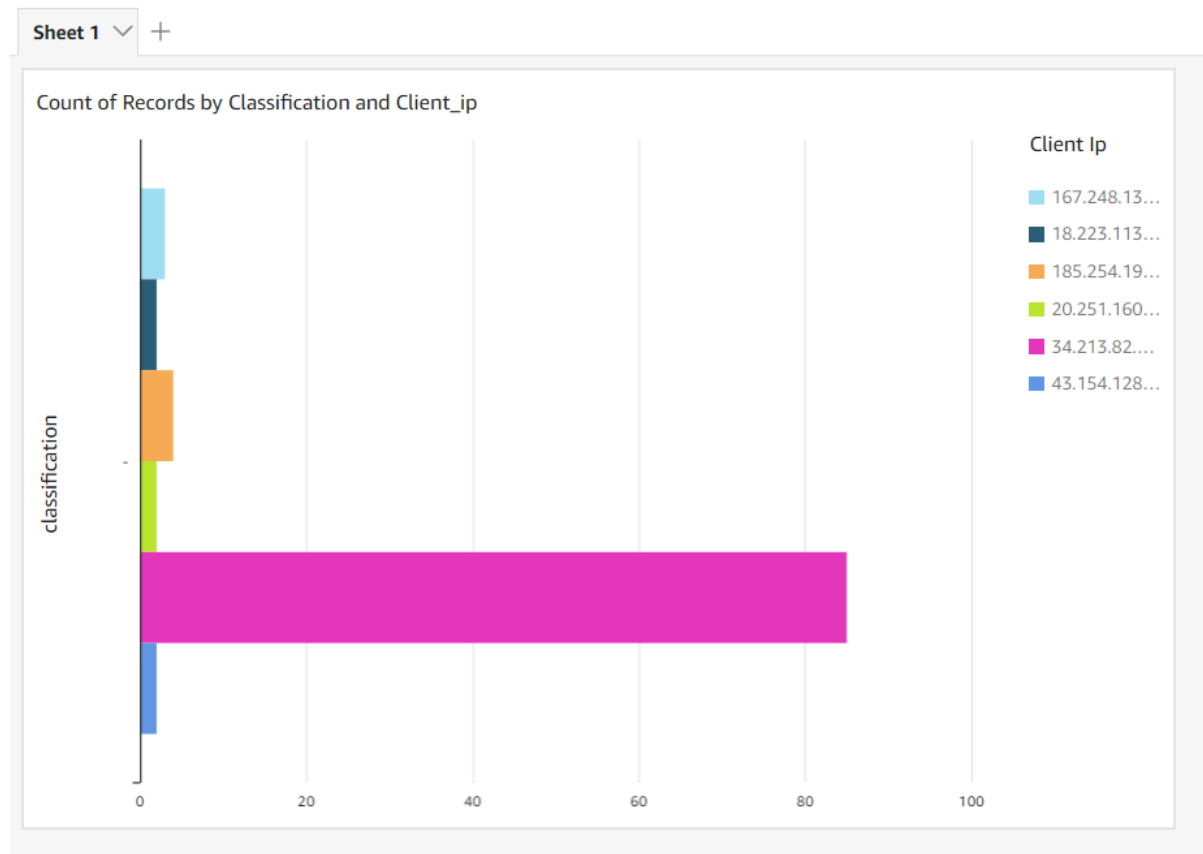


Click Create data source



Select the Y-axis dimension and Group/Color dimension





In This Way you can use the Amazon Quick sight Analysis for Analysing the overall tasks.