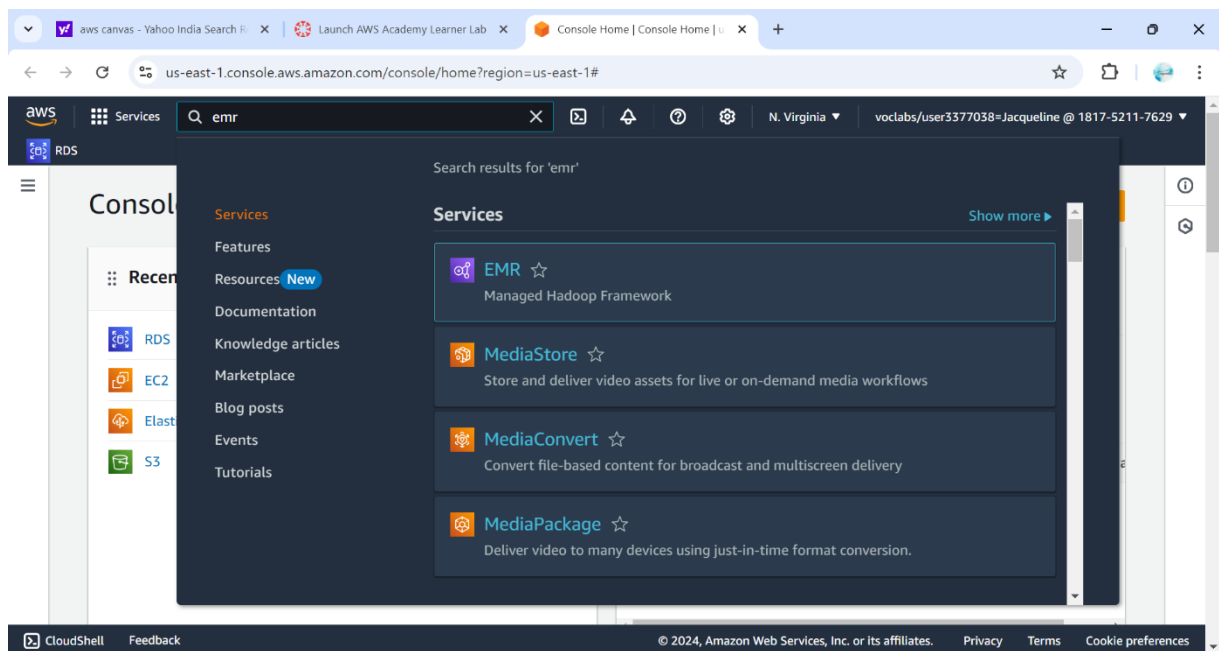


Installation and configuration of cloud Hadoop and demonstrate simple query.

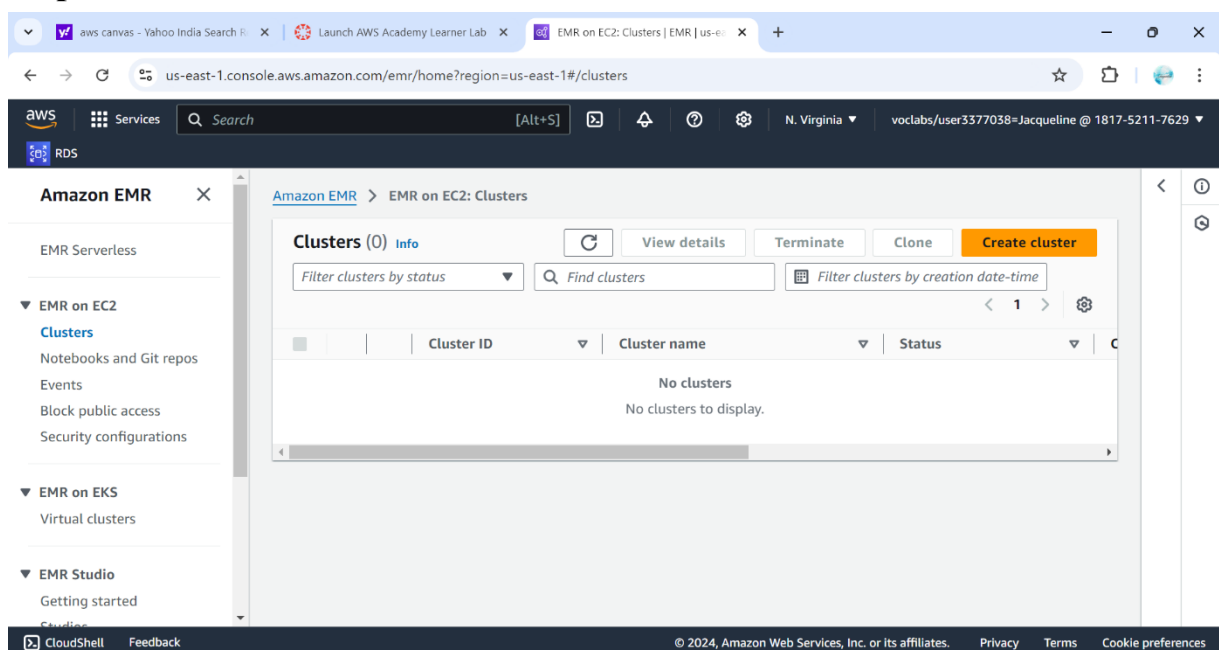
Prepare Screen shots file and also write down the steps.

Make single word or PDF file .

Step 1 : First of all go to the services and in that select EMR to open the EMR console.



Step 2 : Click on Create Cluster.



Step 3 : Fill the cluster name and select Core Hadoop.

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/createCluster

My cluster

Amazon EMR release [Info](#)
A release contains a set of applications which can be installed on your cluster.
emr-7.3.0

Application bundle

Spark Interactive Core Hadoop Flink HBase Presto Trino Custom

☐ AmazonCloudWatchAgent 1.300032.2
☐ HCatalog 3.1.3
☒ Hue 4.11.0
☐ Livy 0.8.0
☒ Pig 0.17.0
☐ Sqoop 1.4.7

☐ Flink 1.18.1
☒ Hadoop 3.3.6
☐ JupyterEnterpriseGateway 2.6.0
☐ Oozie 5.2.1
☐ Presto 0.285
☐ TensorFlow 2.16.1

☐ HBase 2.4.17
☒ Hive 3.1.3
☐ JupyterHub 1.5.0
☐ Phoenix 5.1.3
☐ Spark 3.5.1
☒ Tez 0.10.2

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Step 4 : Configure the following steps.

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/createCluster

☐ AmazonCloudWatchAgent 1.300032.2
☐ HCatalog 3.1.3
☒ Hue 4.11.0
☐ Livy 0.8.0
☒ Pig 0.17.0
☐ Sqoop 1.4.7
☐ Trino 442

☐ Flink 1.18.1
☒ Hadoop 3.3.6
☐ JupyterEnterpriseGateway 2.6.0
☐ Oozie 5.2.1
☐ Presto 0.285
☐ TensorFlow 2.16.1
☐ Zeppelin 0.11.1

☐ HBase 2.4.17
☒ Hive 3.1.3
☐ JupyterHub 1.5.0
☐ Phoenix 5.1.3
☐ Spark 3.5.1
☒ Tez 0.10.2
☐ ZooKeeper 3.9.1

AWS Glue Data Catalogue settings
Use the AWS Glue Data Catalog to provide an external metastore for your application.
☐ Use for Hive table metadata

Operating system options [Info](#)
☒ Amazon Linux release
☐ Customised Amazon Machine Image (AMI)
☒ Automatically apply the latest Amazon Linux updates

Cluster configuration - required [Info](#)
Choose a configuration method for the primary, core and task node groups for your cluster.

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/createCluster

Cluster configuration - required [Info](#)

Choose a configuration method for the primary, core and task node groups for your cluster.

☒ **Instance groups**
Choose one instance type per node group

☐ **Instance fleets**
Choose any combination of instance types within each node group

Instance groups

Primary
Choose EC2 instance type

m5.xlarge
4 vCore 16 GiB memory
EBS only storage On-demand price: -
Lowest spot price: -

[Actions](#)

☐ **Use high availability**
Launch highly available, more resilient cluster with three primary nodes on On-Demand Instances. This configuration applies for the lifetime of your cluster. [Find out more](#)

Node configuration - optional

Core
Choose EC2 instance type

m5.xlarge

[Actions](#)

Summary [Info](#)

Name and applications - required

Amazon EMR release
emr-7.3.0

Application bundle
Core Hadoop (Hadoop 3.3.6, Hive 3.1.3, Hue 4.11.0, Pig 0.17.0, Tez 0.10.2)

Cluster configuration - required

Instance groups
Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)

Cluster scaling and provisioning - required

Provisioning configuration
Core size: 1 instance
Task size: 1 instance

Networking - required

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/createCluster

Name

Task - 1

Choose EC2 instance type

m5.xlarge
4 vCore 16 GiB memory
EBS only storage On-demand price: -
Lowest spot price: -

[Actions](#)

Node configuration - optional

[Add task instance group](#)

You can add up to 47 more task instance groups.

EBS root volume

EBS root volume applies to the operating systems and applications that you install on the cluster. [EBS root volume ratio constraints](#)

Size (GiB)	IOPS	Throughput (MiB/s)
15	3000	125

15 - 100 GiB per volume
General purpose SSD (gp3)

3000 - 16000 IOPS per volume.
Choose a maximum ratio of 500:1 between IOPS and volume size.

125 - 1000 MiB/s per volume.
Choose a maximum ratio of 0.25:1 between throughput and IOPS.

Summary [Info](#)

Name and applications - required

Amazon EMR release
emr-7.3.0

Application bundle
Core Hadoop (Hadoop 3.3.6, Hive 3.1.3, Hue 4.11.0, Pig 0.17.0, Tez 0.10.2)

Cluster configuration - required

Instance groups
Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)

Cluster scaling and provisioning - required

Provisioning configuration
Core size: 1 instance
Task size: 1 instance

Networking - required

aws canvas - Yahoo India Search F x Launch AWS Academy Learner Lab x Create cluster > EMR on EC2: Clus x +

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/createCluster

aws

Services

Search

[Alt+S]

RDS

Cluster scaling and provisioning - required info

Choose how Amazon EMR should size your cluster.

Choose an option

Set cluster size manually

Use this option if you know your workload patterns in advance.

Use EMR-managed scaling

Monitor key workload metrics so that EMR can optimise the cluster size and optimise its resource utilisation.

Use custom automatic scaling

To programmatically scale core and task nodes, create custom automatic scaling policies.

Provisioning configuration

Set the size of your core and task instance groups. Amazon EMR attempts to provision this capacity when you launch your cluster.

Name	Instance type	Instance(s) size	Use spot purchasing option
Core	m5.xlarge	1	<input type="checkbox"/>
Task - 1	m5.xlarge	1	<input type="checkbox"/>

Networking - required info

Choose the network settings that determine how you and other entities communicate with your cluster.

Virtual private cloud (VPC)

Info

us-east-1-vpc-013e626860017354c

Browse

Create VPC

Summary info

Name and applications - required

Amazon EMR release
emr-7.3.0

Application bundle
Core Hadoop (Hadoop 3.3.6, Hive 3.1.3, Hue 4.11.0, Pig 0.17.0, Tez 0.10.2)

Cluster configuration - required

Instance groups
Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)

Cluster scaling and provisioning - required

Provisioning configuration
Core size: 1 instance
Task size: 1 instance

Networking - required

CloudShell

Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws canvas - Yahoo India Search F x Launch AWS Academy Learner Lab x Create cluster > EMR on EC2: Clus x +

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/createCluster

aws

Services

Search

[Alt+S]

RDS

EC2 security groups (firewall)

Primary node

EMR-managed security group

EMR will automatically update the selected group.

Create ElasticMapReduce-Primary

Additional security groups - optional

Select up to 4 additional security groups.

Choose additional security groups

Core and task nodes

EMR-managed security group

EMR will automatically update the selected group.

Create ElasticMapReduce-Core

Additional security groups - optional

Select up to 4 additional security groups.

Choose additional security groups

Steps (0)

Info

Use commands and scripts to tell your cluster where to find and how to process your data. Steps run consecutively unless you enable the Concurrency option.

Filter steps by status

Find steps

< 1 >

⚙

Name	Status	Type	Arguments	Script location
No steps				
You don't have any steps added.				

Add

Summary info

Name and applications - required

Amazon EMR release
emr-7.3.0

Application bundle
Core Hadoop (Hadoop 3.3.6, Hive 3.1.3, Hue 4.11.0, Pig 0.17.0, Tez 0.10.2)

Cluster configuration - required

Instance groups
Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)

Cluster scaling and provisioning - required

Provisioning configuration
Core size: 1 instance
Task size: 1 instance

Networking - required

CloudShell

Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws canvas - Yahoo India Search R x Launch AWS Academy Learner Lab x Create cluster > EMR on EC2: Clus x +

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/createCluster

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user3377038=Jacqueline @ 1817-5211-7629

Bootstrap actions (0) Info

Remove Edit Add

Use bootstrap actions to install software or customise your instance configuration.

Cluster logs Info

Choose where and how to store your log files.

Tags Info

Use tags to search and filter for resources, and track AWS costs associated with your cluster.

Software settings Info

Override the default configurations for specific applications on your cluster.

Security configuration and EC2 key pair Info

Choose a security configuration or create a new one that you can reuse with other clusters.

Security configuration

Select your cluster encryption, authentication, authorisation and instance metadata service settings.

Create a security configuration

Refresh

Browse

Create security configuration

Amazon EC2 key pair for SSH to the cluster Info

Enter a key name or choose Browse to select an Amazon EC2...

Browse

Create key pair

Summary Info

Name and applications - required

Amazon EMR release
emr-7.3.0

Application bundle
Core Hadoop (Hadoop 3.3.6, Hive 3.1.3, Hue 4.11.0, Pig 0.17.0, Tez 0.10.2)

Cluster configuration - required

Instance groups
Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)

Cluster scaling and provisioning - required

Provisioning configuration
Core size: 1 instance
Task size: 1 instance

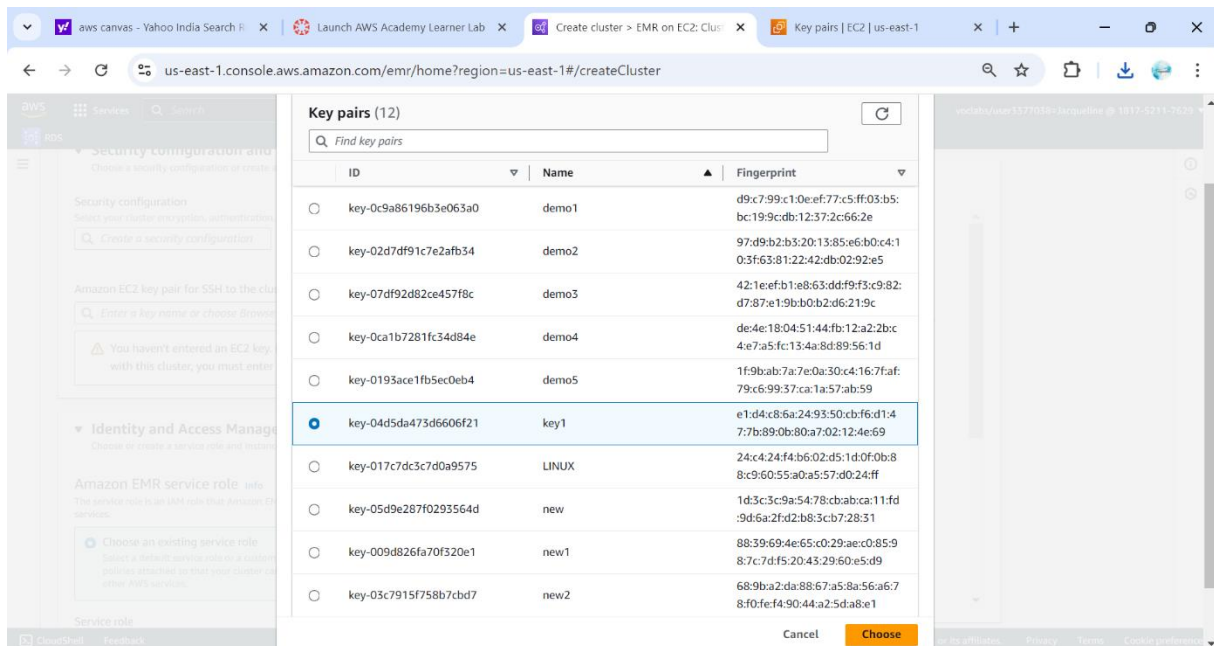
Networking - required

CloudShell

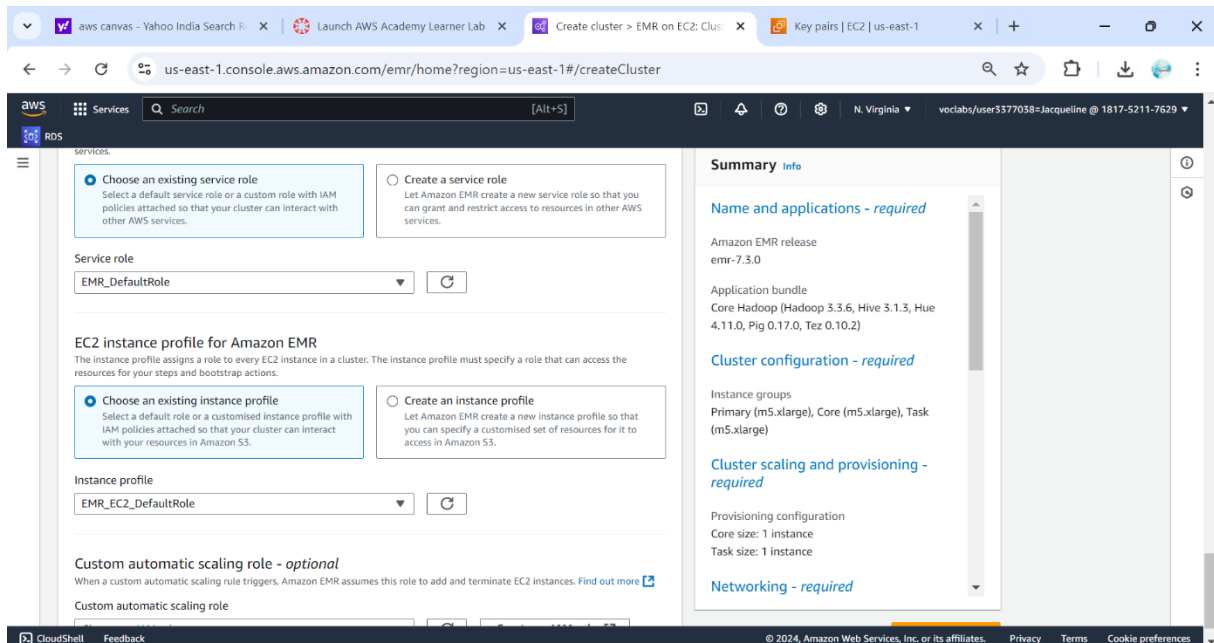
Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Step 5 : Now create a new Key Pair and then choose the created key pair.

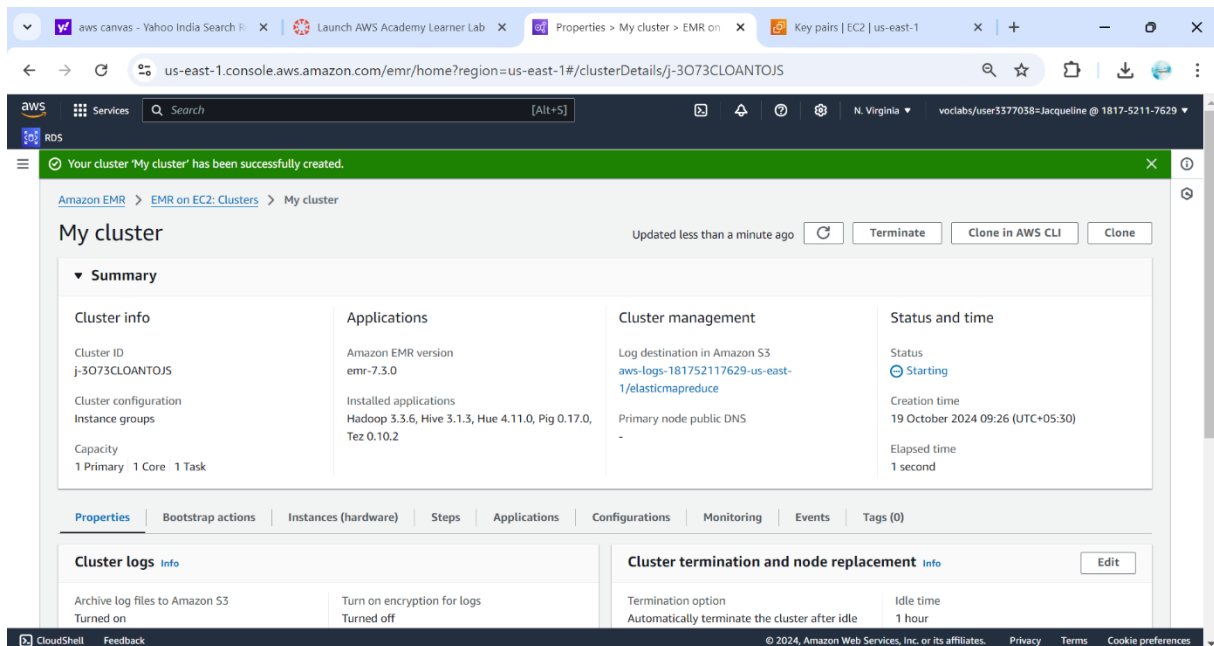


Step 6 : Choose emr_defaultrole and emr_ec2_defaultrole and then click on Create Cluster.

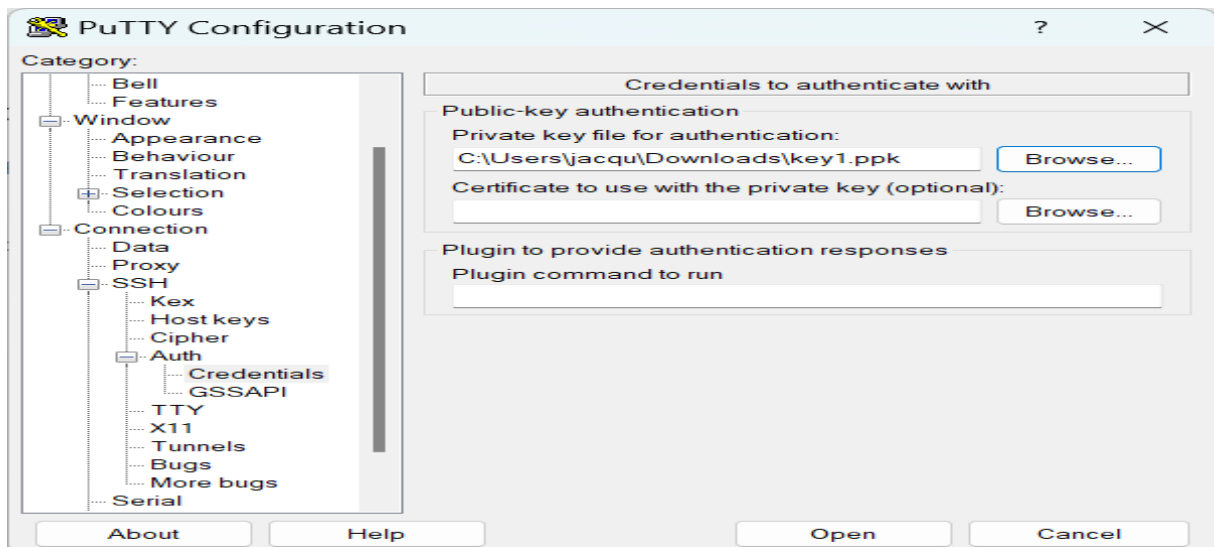


Step :

Step: 7 You see the cluster is successfully created.



Step 8 : Now open the Putty app and load it.



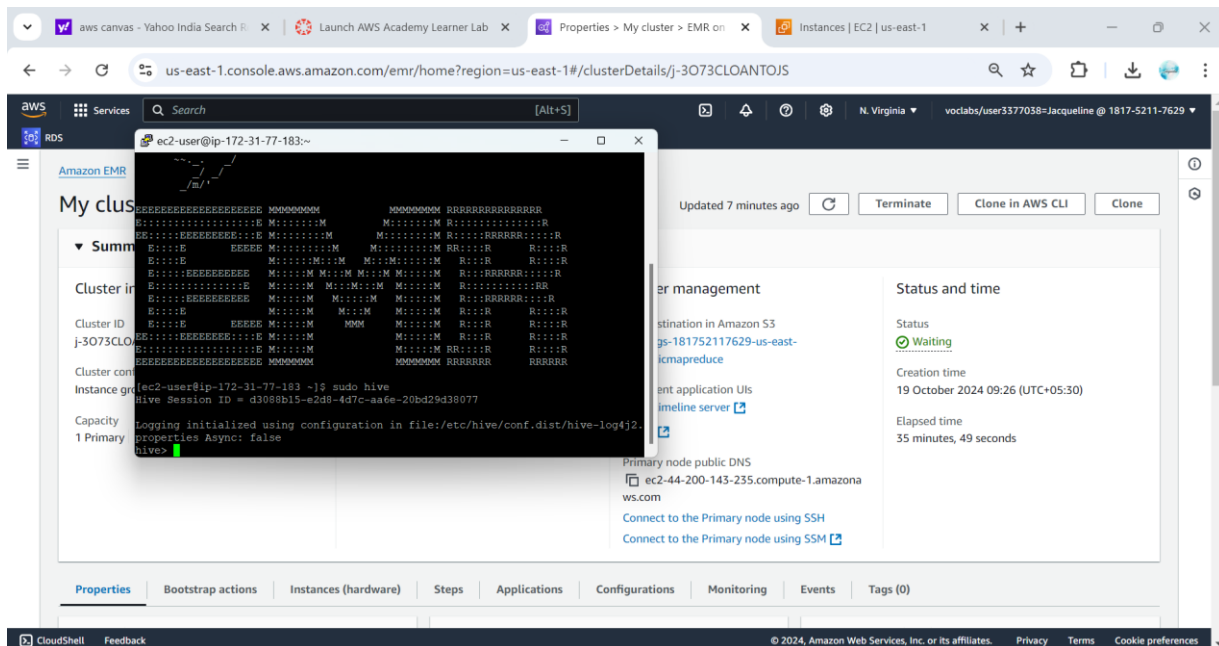
9 Now login as ec2-user.

Step :

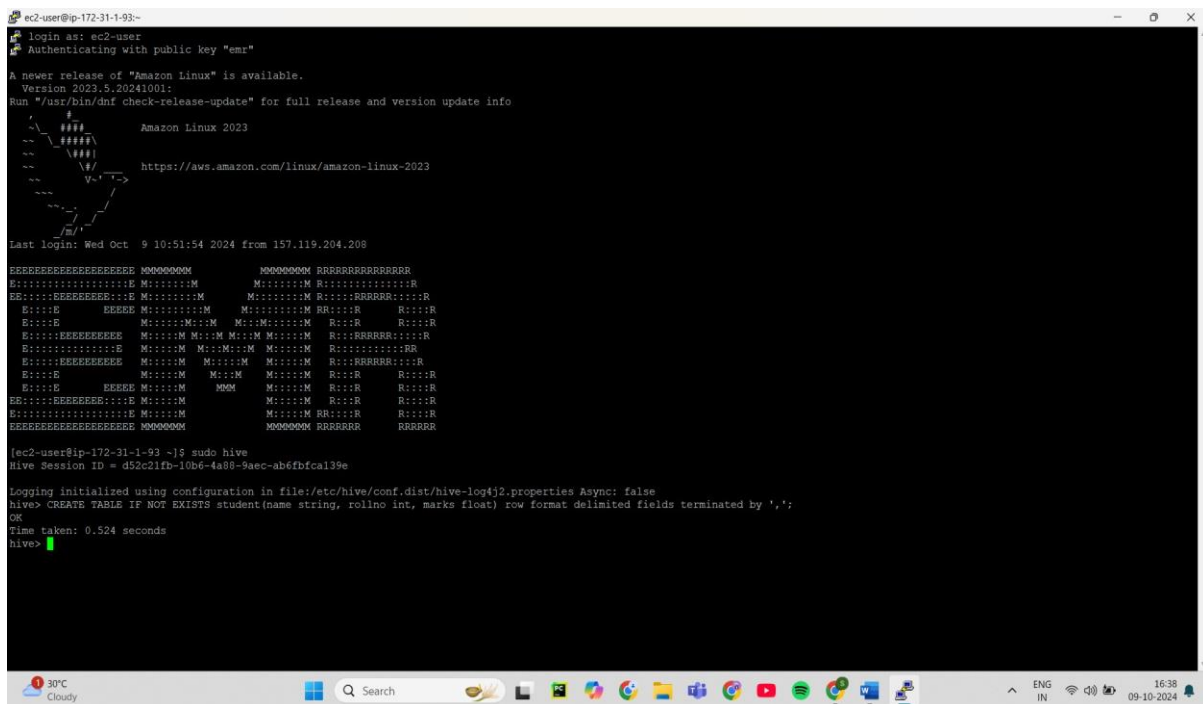
[illegible]

Step

Step10: Using below commands create the database and run the query.



Step11: Enter a command --à sudo hive



aws canvas - Yahoo India SearchLaunch AWS Academy Learner LabProperties > My cluster > EMR onInstances | EC2 | us-east-1

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/clusterDetails/j-3O73CLOANTOJS

Services

Search

[Alt+S]

N. Virginia

voclabs/user3377038=Jacqueline @ 1817-5211-7629

Amazon EMR

EMR Serverless

EMR on EC2

Clusters

Notebooks and Git repos

Events

Block public access

Security configurations

EMR on EKS

Virtual clusters

EMR Studio

Getting started

Studios

Workspaces (notebooks)

What's new

Video tour

Amazon EMR > EMR on EC2: Clusters > My cluster

My cluster

Updated less than a minute ago

Terminate

Summary

Cluster info

Cluster ID
j-3O73CLOANTOJS

Cluster configuration

Instance groups

Capacity

1 Primary 1 Core 1 Task

Applications

Amazon EMR version
emr-7.3.0

Installed applications

Hadoop 3.3.6, Hive 3.1.3, Hue 4.11.0,
Pig 0.17.0, Tez 0.10.2

Cluster management

Log destination in Amazon S3
aws-logs-181752117629-us-east-1/elasticmapreduce

Persistent application UIs
YARN timeline server
Tez UI

Primary node public DNS
ec2-44-200-143-235.compute-1.amazonaws.com

Connect to the Primary node using SSH

Connect to the Primary node using SSM

PropertiesBootstrap actionsInstances (hardware)StepsApplicationsConfigurationsMonitoringEventsTags (0)

CloudShellFeedback

© 2024, Amazon Web Services, Inc. or its affiliates. PrivacyTermsCookie preferences

Account ID: 1817-5211-7629

Federated user: voclabs/user3377038=Jacqueline

Account

Organization

Service Quotas

Billing and Cost Management

Switch role

Sign out