

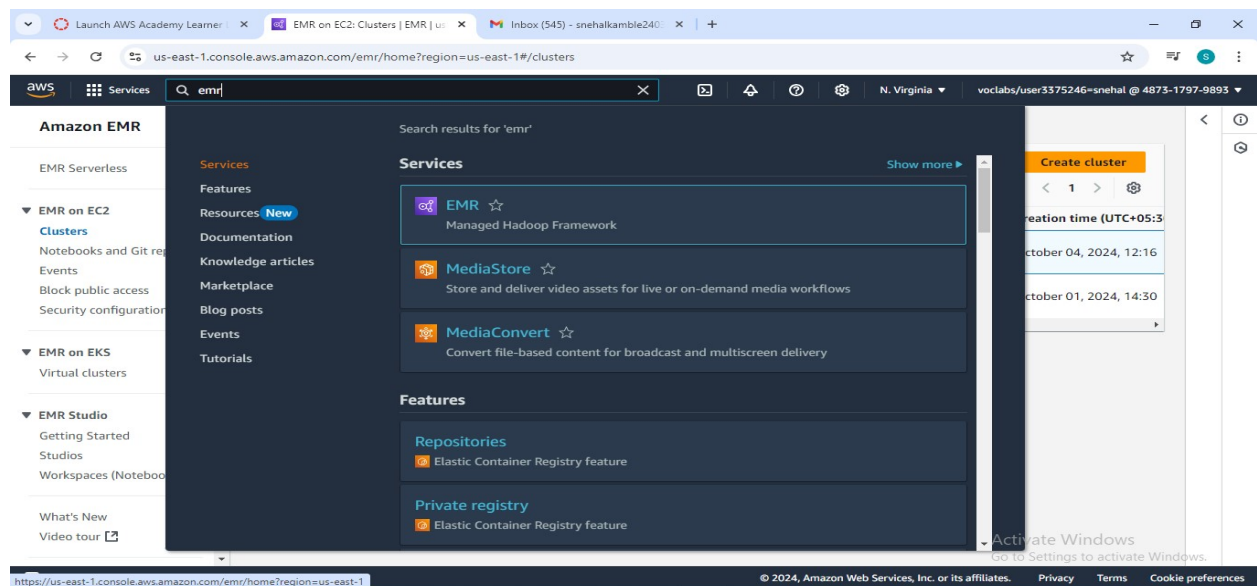
Assignment 10

Name : Srushti Dattatray Pawar

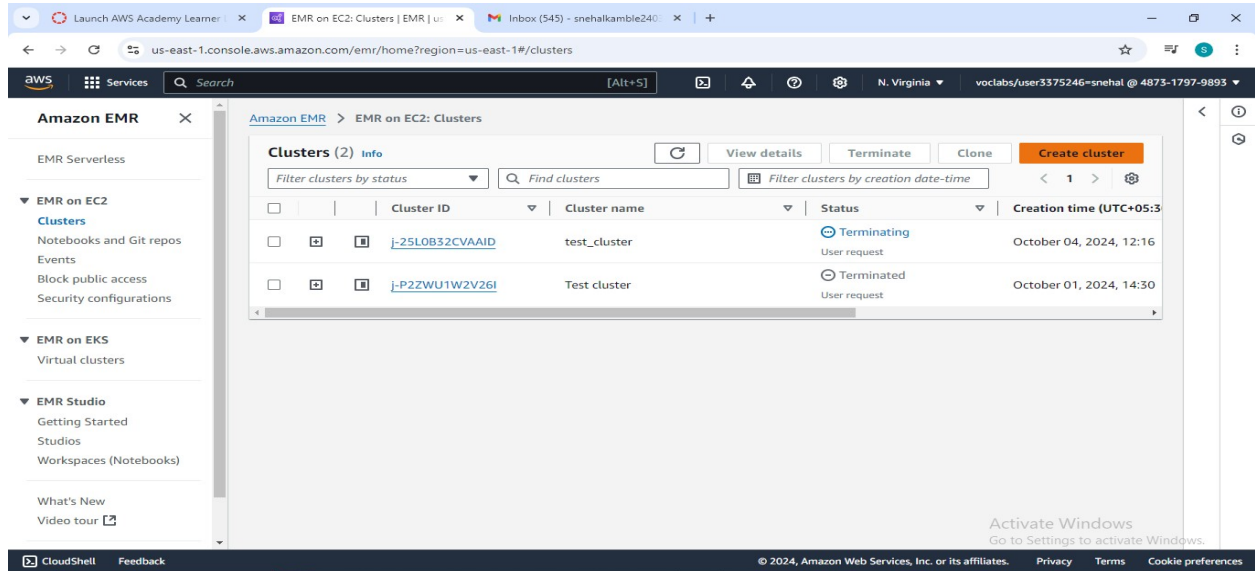
Class: Msc CS Part 2

Q.) Installation and configuration of cloud Hadoop and demonstrate simple query

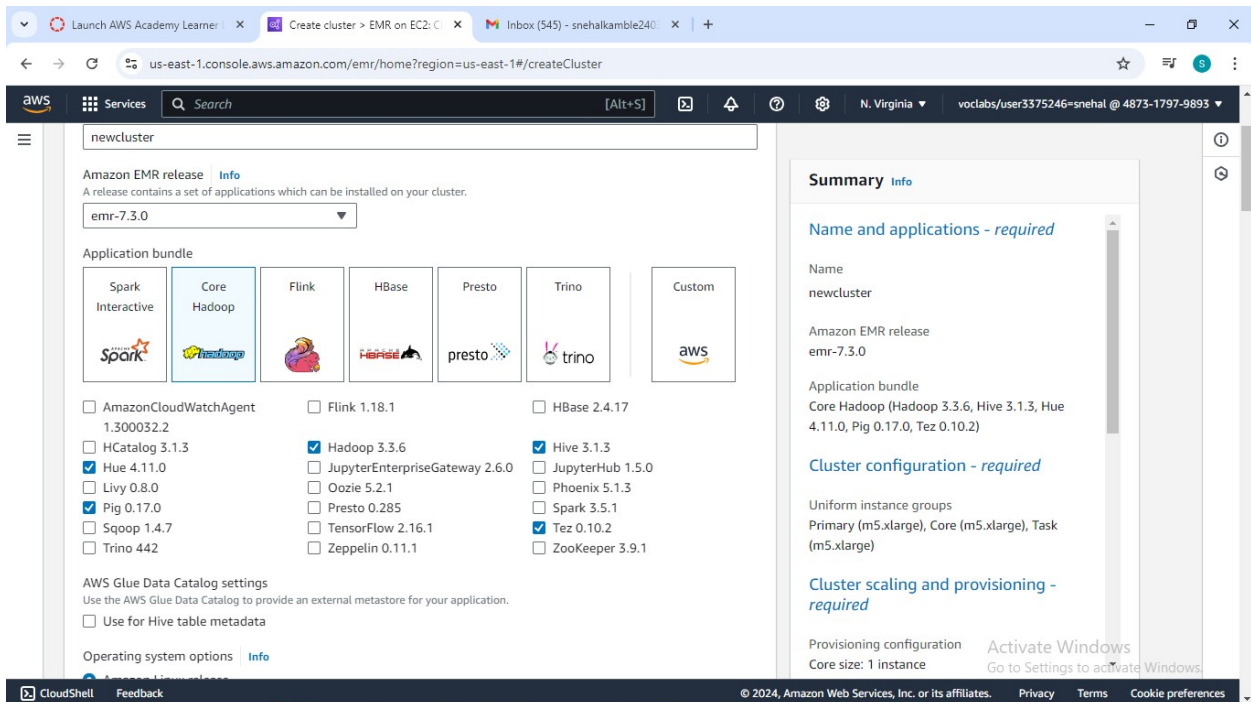
Step 1: Login to your AWS account ,search for EMR under services



Step 2: Click on->Create cluster



Step 3: Give name to cluster and select Core Hadoop application bundle.



Step 4: Select EC2 key pair. Service role->EMR_DefaultRole

Launch AWS Academy Learner

Create cluster > EMR on EC2: C

Inbox (545) - snehalkamble240

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

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user3375246-snehal @ 4873-1797-9893

Cluster configuration - required Info

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

Uniform instance groups

Choose the same EC2 instance type and purchasing option (On-Demand or Spot) for all nodes in your node group. [Learn more](#)

Flexible instance fleets

Choose from the widest variety of provisioning options for the EC2 instances in your cluster. Diversify instance types and purchasing options, and use an allocation strategy. [Learn more](#)

Uniform 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. [Learn more](#)

Node configuration - optional

Core

Choose EC2 instance type

Summary Info

Name and applications - required

Name

newcluster

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

Uniform instance groups

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

Cluster scaling and provisioning - required

Provisioning configuration

Core size: 1 instance

Activate Windows

Go to Settings to activate Windows.

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

Amazon EC2 key pair for SSH to the cluster [Info](#)

Q cluster X Browse Create key pair

You haven't entered an EC2 key. If you're outside a VPN and want to enable SSH or use Hue SQL assistant with this cluster, you must enter an EC2 key.

▼ Identity and Access Management (IAM) roles - required [Info](#)

Choose or create a service role and instance profile for the EC2 instances in your cluster.

Amazon EMR service role [Info](#)

The service role is an IAM role that Amazon EMR assumes to provision resources and perform service-level actions with other AWS services.

☒ Choose an existing service role
Select a default service role or a custom role with IAM policies attached so that your cluster can interact with other AWS services.

☐ Create a service role
Let Amazon EMR create a new service role so that you can grant and restrict access to resources in other AWS services.

Service role

EMR_DefaultRole

EC2 instance profile for Amazon EMR

The instance profile assigns a role to every EC2 instance in a cluster. The instance profile must specify a role that can access the resources for your steps and bootstrap actions.

Summary [Info](#)

Name and applications - required

Name
newcluster

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

Uniform instance groups
Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)

Cluster scaling and provisioning - required

Provisioning configuration
Core size: 1 instance

Activate Windows
Go to Settings to activate Windows.

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

☒ Choose an existing service role
Select a default service role or a custom role with IAM policies attached so that your cluster can interact with other AWS services.

☐ Create a service role
Let Amazon EMR create a new service role so that you can grant and restrict access to resources in other AWS services.

Service role

EMR_DefaultRole

EC2 instance profile for Amazon EMR

The instance profile assigns a role to every EC2 instance in a cluster. The instance profile must specify a role that can access the resources for your steps and bootstrap actions.

☒ Choose an existing instance profile
Select a default role or a custom instance profile with IAM policies attached so that your cluster can interact with your resources in Amazon S3.

☐ Create an instance profile
Let Amazon EMR create a new instance profile so that you can specify a custom set of resources for it to access in Amazon S3.

Instance profile

EMR_EC2_DefaultRole

Custom automatic scaling role - optional

When a custom automatic scaling rule triggers, Amazon EMR assumes this role to add and terminate EC2 instances. [Learn more](#)

Custom automatic scaling role

Choose IAM role

Create IAM role

Summary [Info](#)

Name and applications - required

Name
newcluster

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

Uniform instance groups
Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)

Cluster scaling and provisioning - required

Provisioning configuration
Core size: 1 instance

Activate Windows
Go to Settings to activate Windows.

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

Step 5: IN EC2 instance, choose instance profile EMR_EC2_DefaultRole and click->Create
>Create cluster

The screenshot displays the AWS EMR console interface for a cluster. The 'Instances (Hardware)' tab is selected, showing the 'Instance group settings' and 'Instance groups' sections.

Instance group settings

| Cluster scaling option | Core | Task |
|---------------------------|--|--|
| Manually set cluster size | Name and Maximum core nodes in the cluster Core - 1 instances | Name and Maximum task nodes in the cluster Task - 1 1 instances |

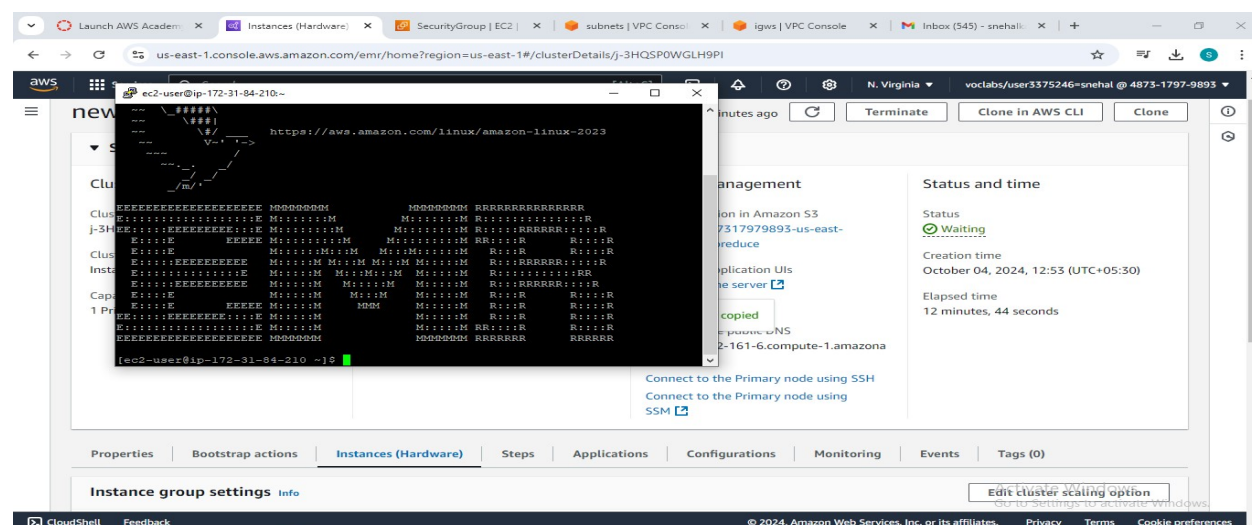
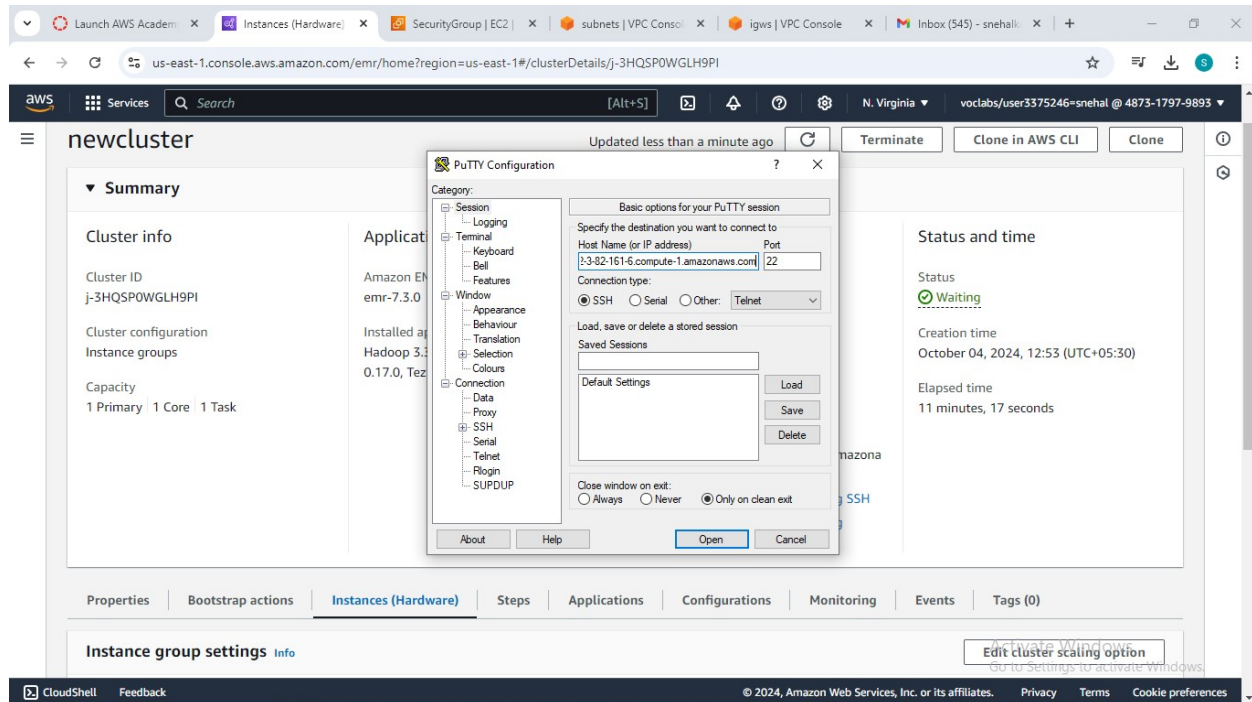
Instance groups (3)

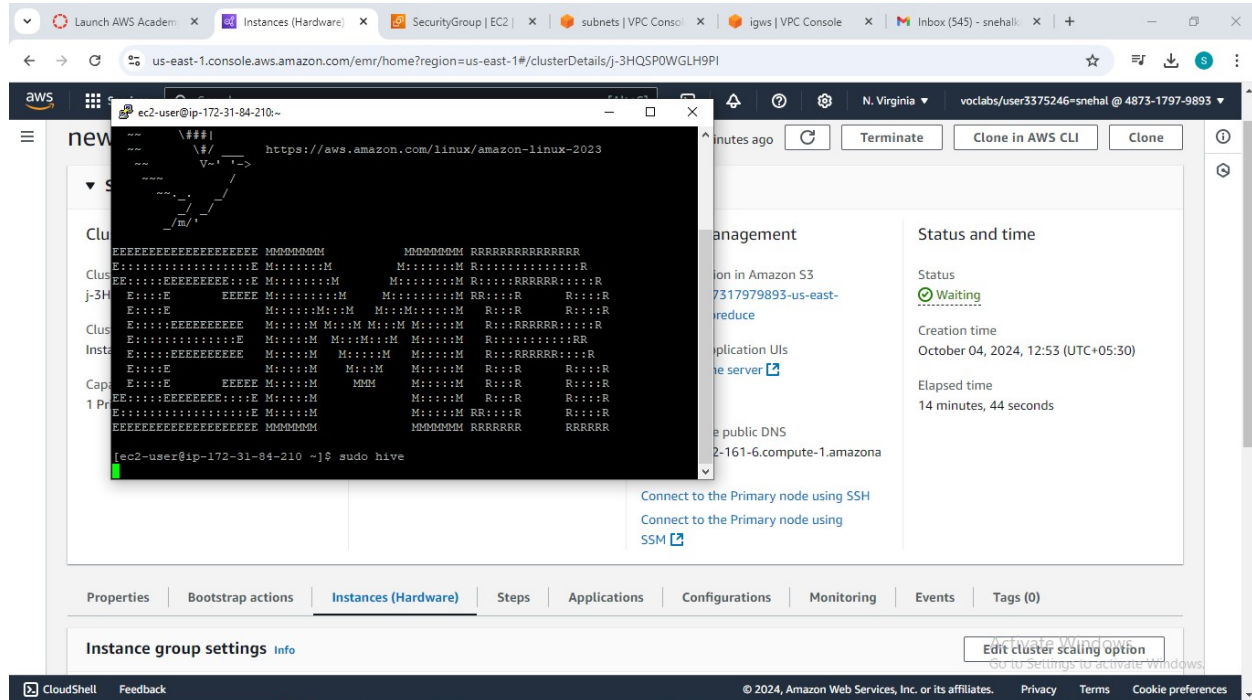
With the instance groups configuration, each node type consists of the same instance type and the same purchasing option for instances: On-Demand or Spot.

| Type and name | ID | Status | Instances | Purchasing option and p... | EBS size (GiB) |
|-----------------|------------------|---------|-----------|----------------------------|----------------|
| Primary | ig-T2PLTL0PZJ7I | Running | 1 | On-Demand | - |
| Core | ig-2JLCX95DPWRZY | Running | 1 | On-Demand | - |
| Task (Task - 1) | ig-28M6ATWGYA8EE | Running | 1 | On-Demand | - |

Step 6: Wait until all instance group's status changed to running

Step 7: Copy the public DNS in puttyit will open a new terminal for ec2 user





Step 8:Type commands:

`sudo hive`

Fire a query for creating table `CREATE TABLE student(`

`id INT,`

`name STRING,`

`rollno STRING UNIQUE DISABLE,`

`class STRING,`

`grade STRING);`

Launch AWS Academy x Instances (Hardware) x SecurityGroup | EC2 x subnets | VPC Console x igws | VPC Console x Inbox (545) - snehal x +

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/clusterDetails/j-3HQSP0WGLH9PI

aws

ec2-user@ip-172-31-84-210:~

```
Query ID = root_20241004074438_81c238d5-e6a2-4d39-9a9e-57c20f4d0866
Total jobs = 1
Launching Job 1 out of 1
Per session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1728027052714_0002)
```

| | VERTICES | MODE | STATUS | TOTAL | COMPLETED | RUNNING | PENDING | FAILED | KILLED |
|-----------|----------|-----------|-----------|-------|-----------|---------|---------|--------|--------|
| Map 1 | | container | SUCCEEDED | 1 | 1 | 0 | 0 | 0 | 0 |
| Reducer 2 | | container | SUCCEEDED | 1 | 1 | 0 | 0 | 0 | 0 |

VERTICES: 02/02 [=====]>>] 100% ELAPSED TIME: 5.82 s

Loading data to table default.student

Cap OK

Time taken: 16.759 seconds

hive> SELECT * FROM student;

OK

| | SNEHAL | 2373020 | MSC | CS | A |
|---|--------|---------|-----|----|---|
| 1 | SNEHAL | 2373020 | MSC | CS | A |

Time taken: 0.284 seconds, Fetched: 1 row(s)

hive>

Connect to the Primary node using SSH

Connect to the Primary node using SSM

Status and time

Status

Waiting

Creation time

October 04, 2024, 12:53 (UTC+05:30)

Elapsed time

21 minutes, 44 seconds

Properties Bootstrap actions Instances (Hardware) Steps Applications Configurations Monitoring Events Tags (0)

Instance group settings Info

Edit cluster scaling option

Go to Settings to activate Windows

CloudShell Feedback

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

And using select * from student; You can view the entry of student