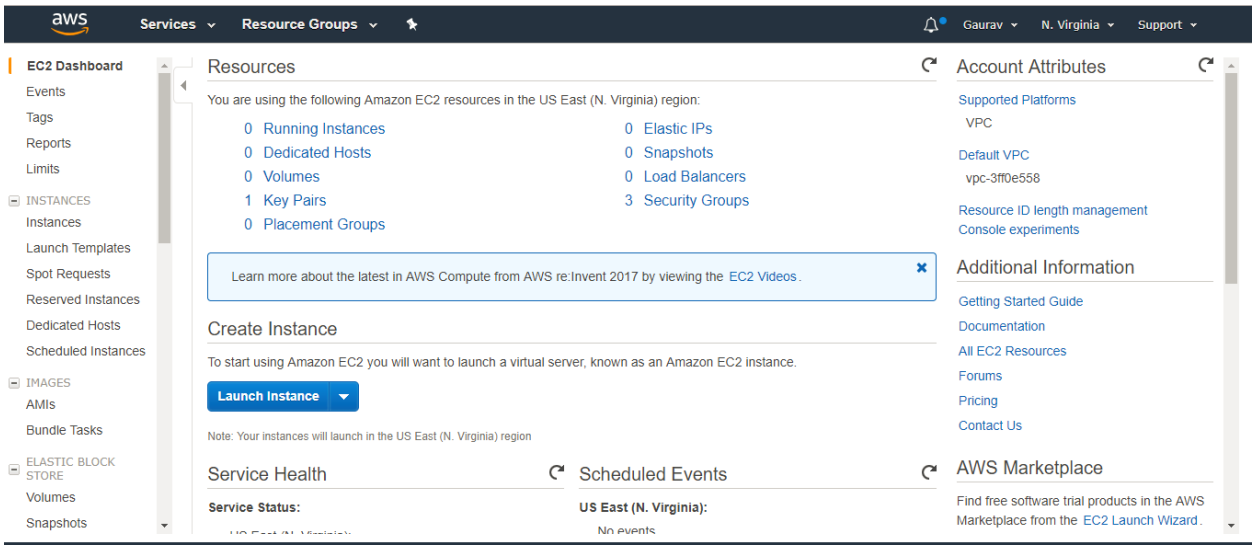
Install CDP Cluster with Cloudera Manager using AMI on AWS EC2:

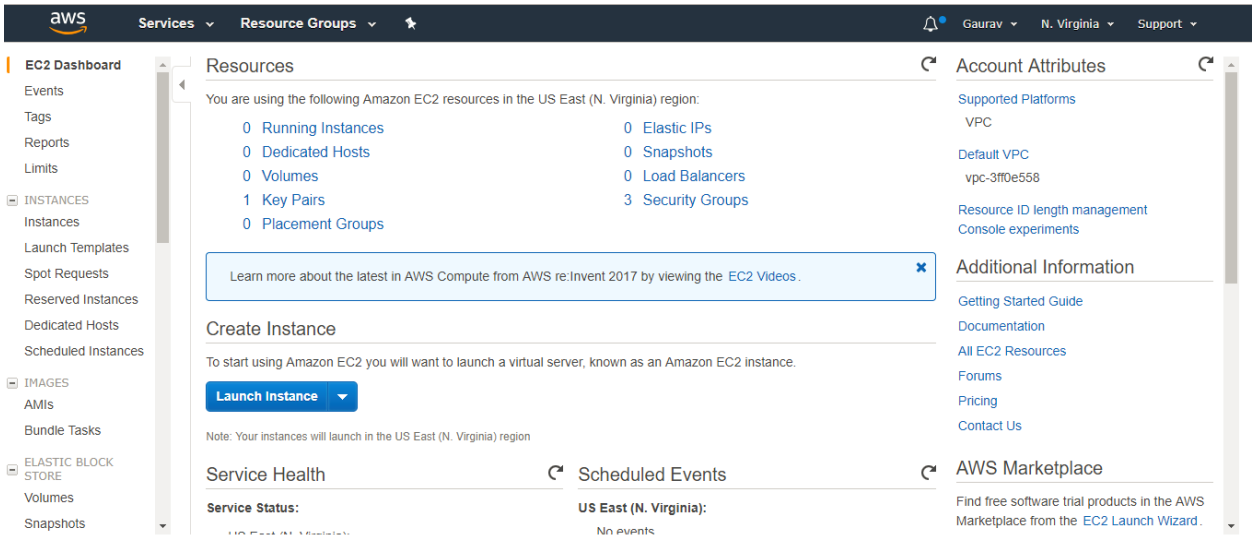
Prerequisites:
Region- **N.Virginia** (You need to select N.Virginia as the region as this is the cheapest region & our AMI is available on N.Virginia)

AMI Name :-**CDP-Cloudera-upGrad**
AMI ID-:ami-0ac405712832c86d0

1. Log in to the AWS console home and click on **'EC2'** under the Compute Section inside the **N.Virginia Region**.



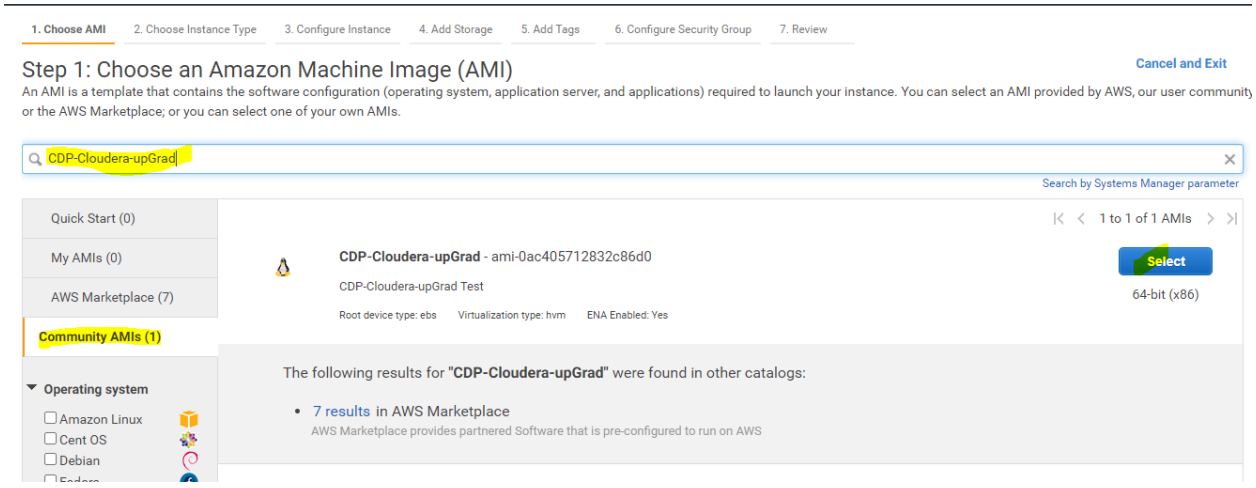
2. Click on '**Launch Instance'**



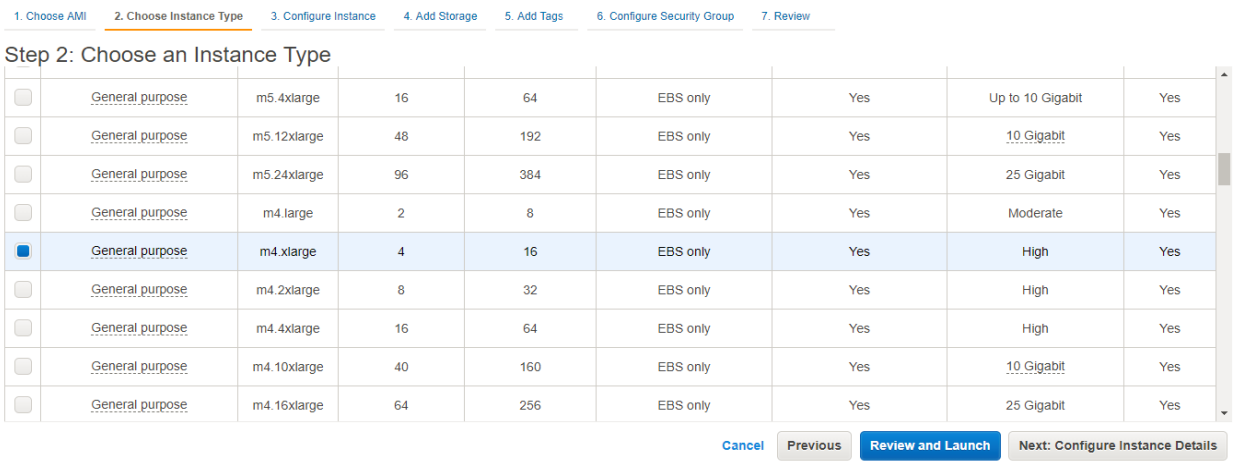
3. Click on '**Community AMI** ' and search our AMI with the name below.

Inside the search, tab enter our AMI Name::**CDP-Cloudera-upGrad** and Select our AMI.

AMI Name: CDP-Cloudera-upGrad



4. Select instance type **'m4.xlarge'** (please select very carefully)
5.click on **'Next Configure Instance details'.**

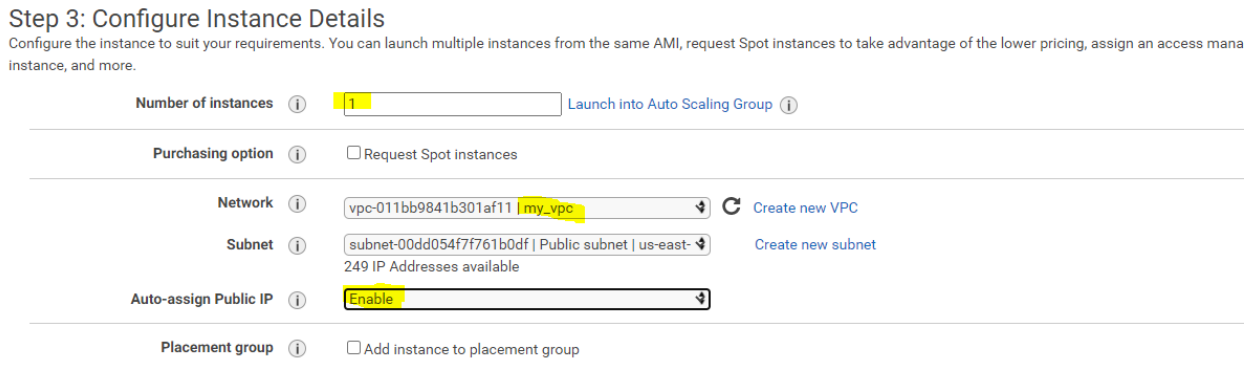


	General purpose	m5.4xlarge	16	64	EBS only	Yes	Up to 10 Gigabit	Yes
	General purpose	m5.12xlarge	48	192	EBS only	Yes	10 Gigabit	Yes
	General purpose	m5.24xlarge	96	384	EBS only	Yes	25 Gigabit	Yes
	General purpose	m4.large	2	8	EBS only	Yes	Moderate	Yes
☑	General purpose	m4.xlarge	4	16	EBS only	Yes	High	Yes
	General purpose	m4.2xlarge	8	32	EBS only	Yes	High	Yes
	General purpose	m4.4xlarge	16	64	EBS only	Yes	High	Yes
	General purpose	m4.10xlarge	40	160	EBS only	Yes	10 Gigabit	Yes
	General purpose	m4.16xlarge	64	256	EBS only	Yes	25 Gigabit	Yes

6. We recommend you to follow the same steps like for CDH 5.15.1

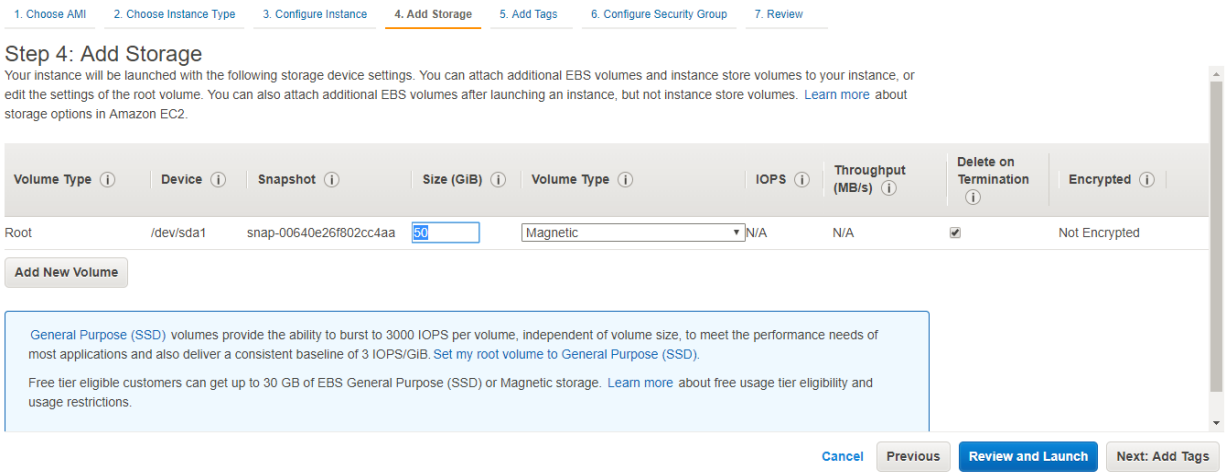
Number of instances- 1

VPC - your vpc name

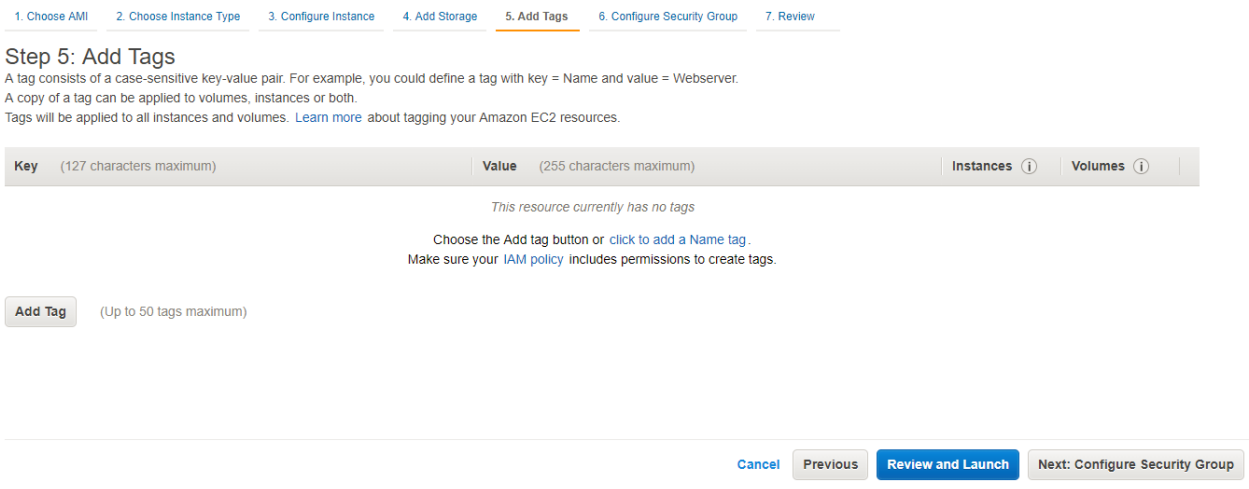


and click on the **'Add Storage'** button (bottom right)**.**

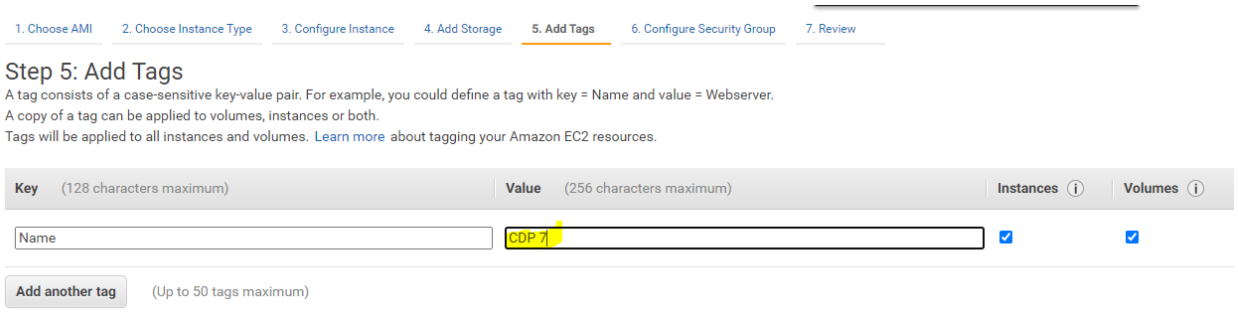
7. We recommend you to add 70GB of storage, and volume type should be ‘Magnetic’ for lower prices. (as pricing is based on the VM usage and configuration). Finally click on the **‘Next: Add Tags’** button.



8. Now click on **‘click to add a Name tag’** as shown in the image below.

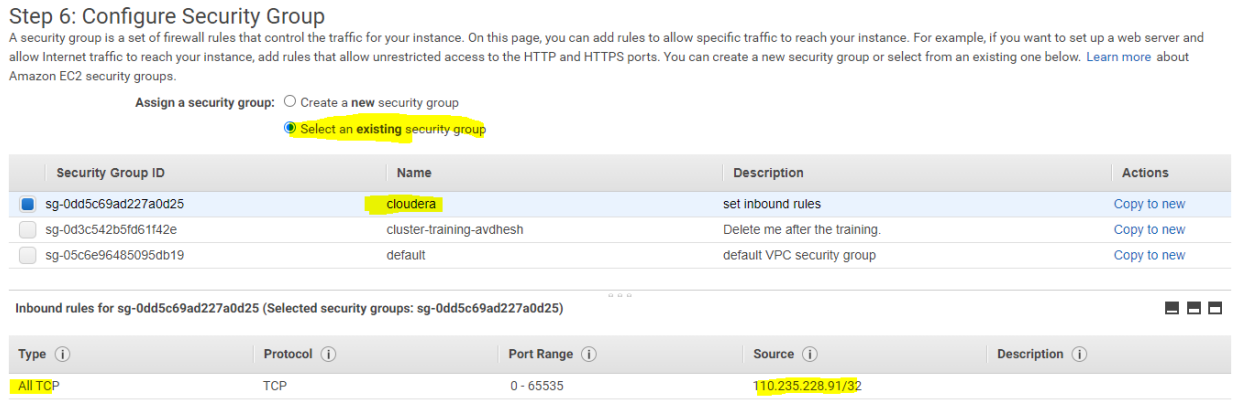


9. Give a name in the cell under **‘Value’**. In our case, we used CDP 7.

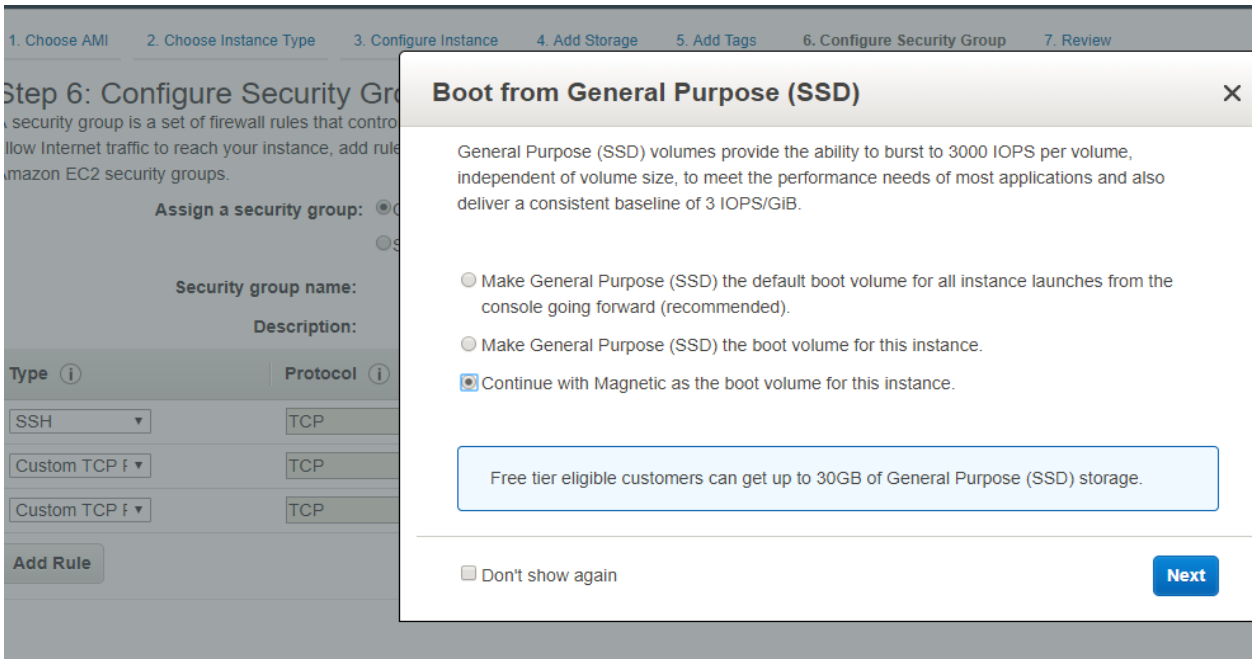


10. Click on **‘Next: Configure Security Group’**

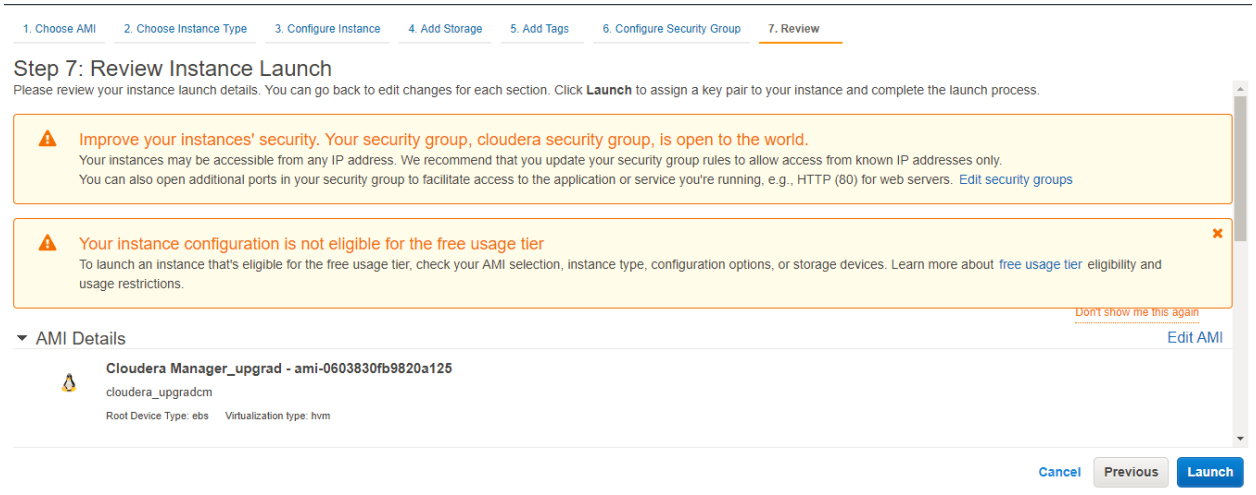
11. Select the **‘Select an existing security group’** and provide any name and description as you wish.



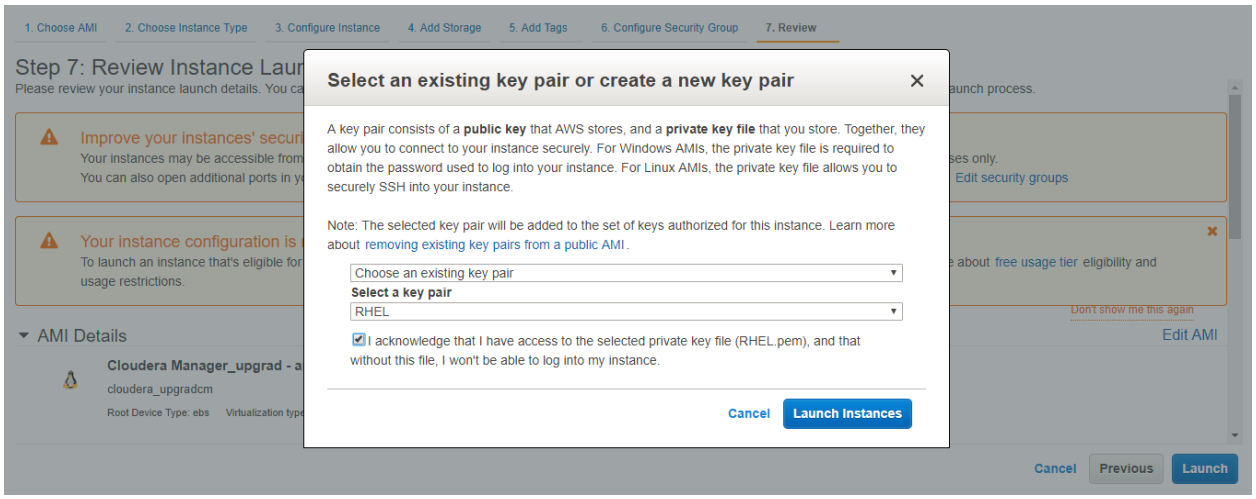
12. Click on **‘Review and Launch’**. The following page will appear as shown in the image below. Click on **‘Continue with Magnetic as the boot volume for this instance’**. Then click **‘Next’**



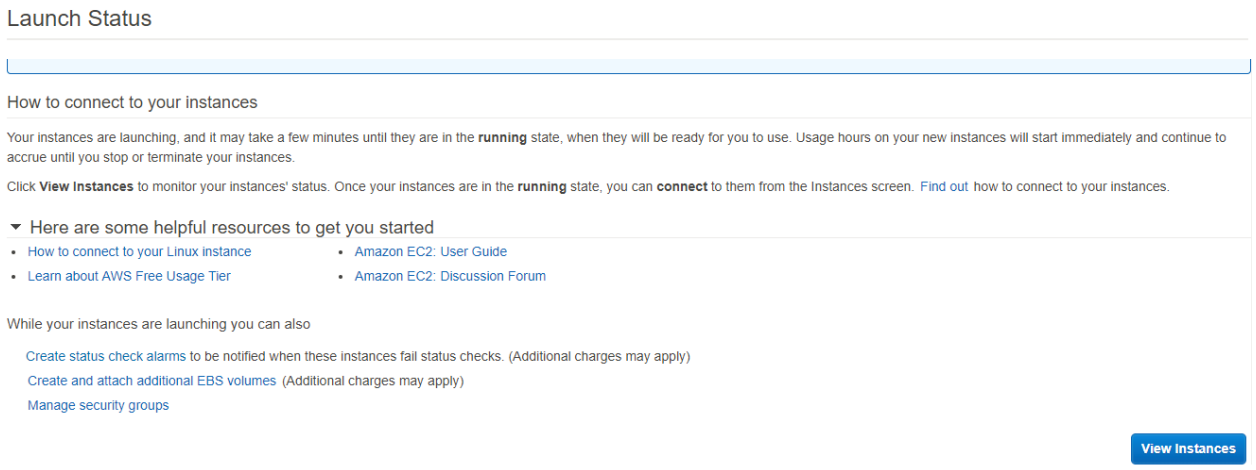
13. Click on the **‘Launch’** button.



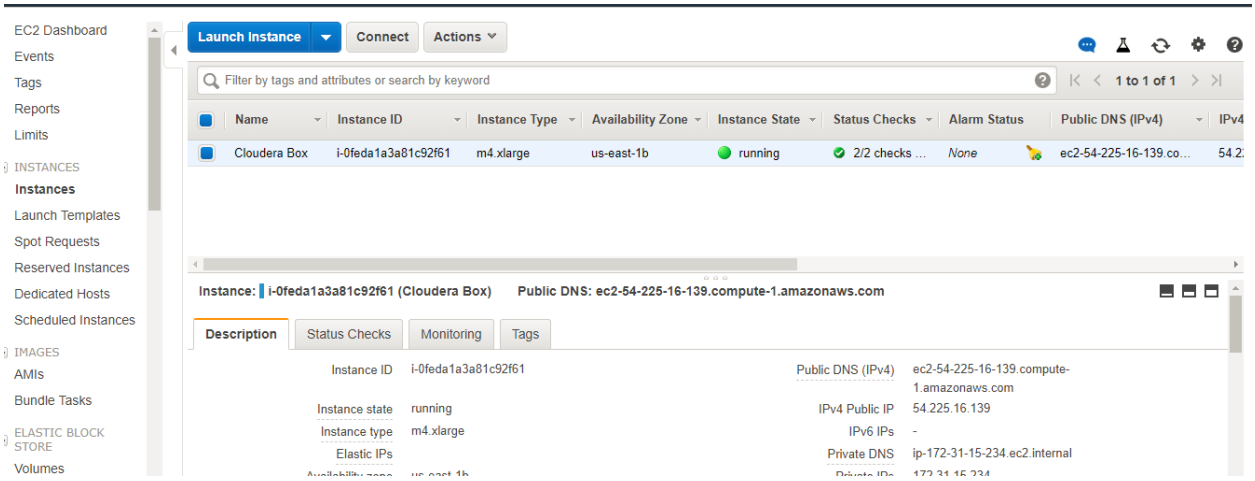
14. Now select **‘Choose an existing key pair’** from the drop-down menu and select your key pair name (in our case, RHEL). Tick the 'I acknowledge….’ and click on **‘Launch Instances’**.



15. Click on **‘View Instances’** and your instances will appear on the screen.



16. Wait until the status check is 2/2 (under **'Status Checks'** column)..

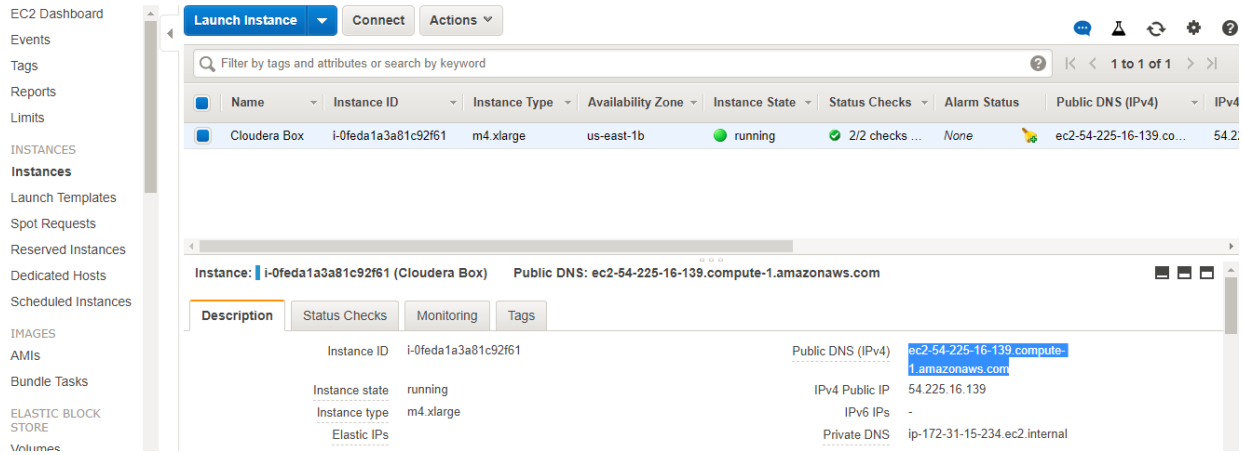


17. Now Access/ log in the instance using putty for window users and install the Hadoop(CDH).

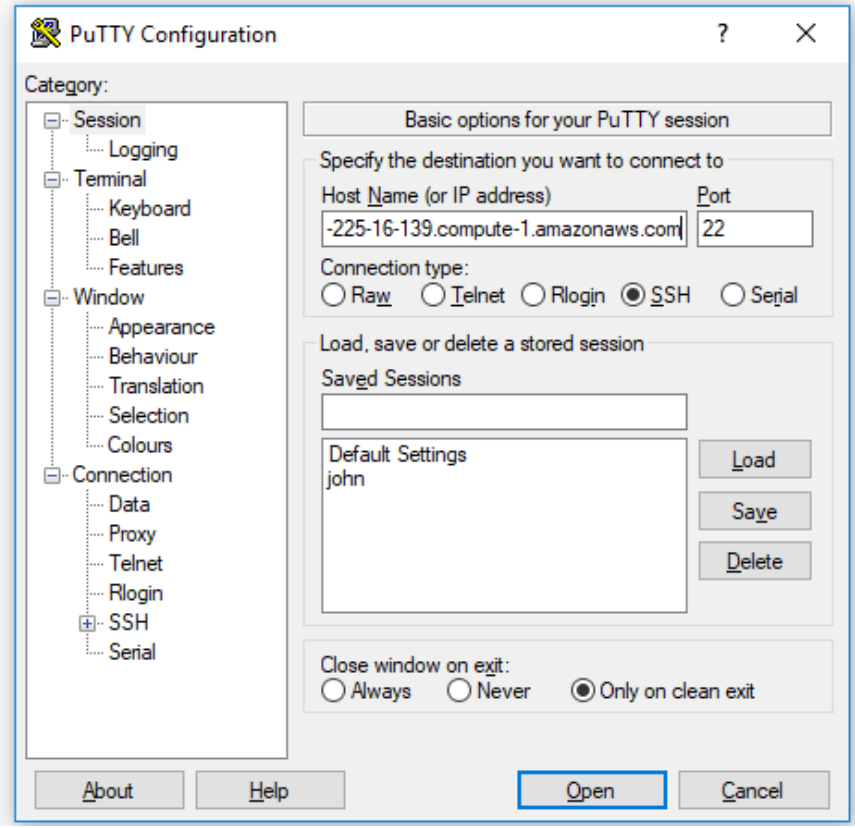
For window users: need .ppk file (convert your RHEL.pem to name.ppk file using PuTTYgen software).

Note:-if you already have a .ppk file please ignore it.

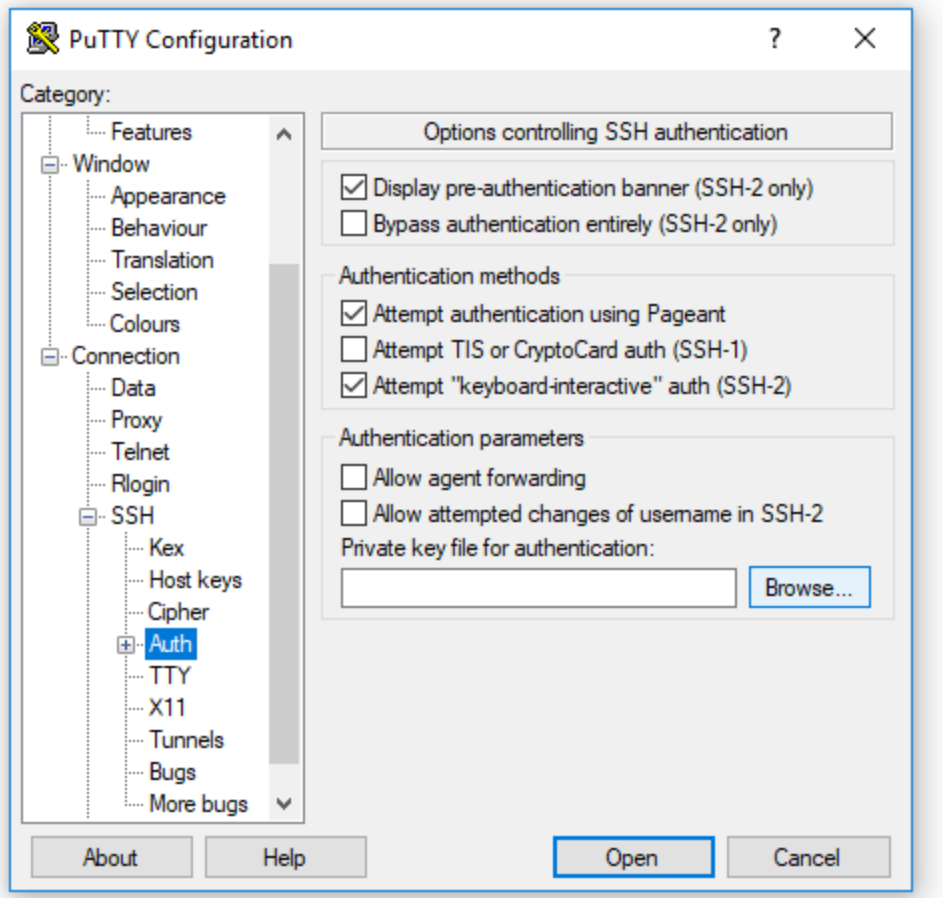
18. Now, open your EC2 dashboard and select your instance. Copy your **'Public DNS (IPv4)'** information as shown in the screenshot.



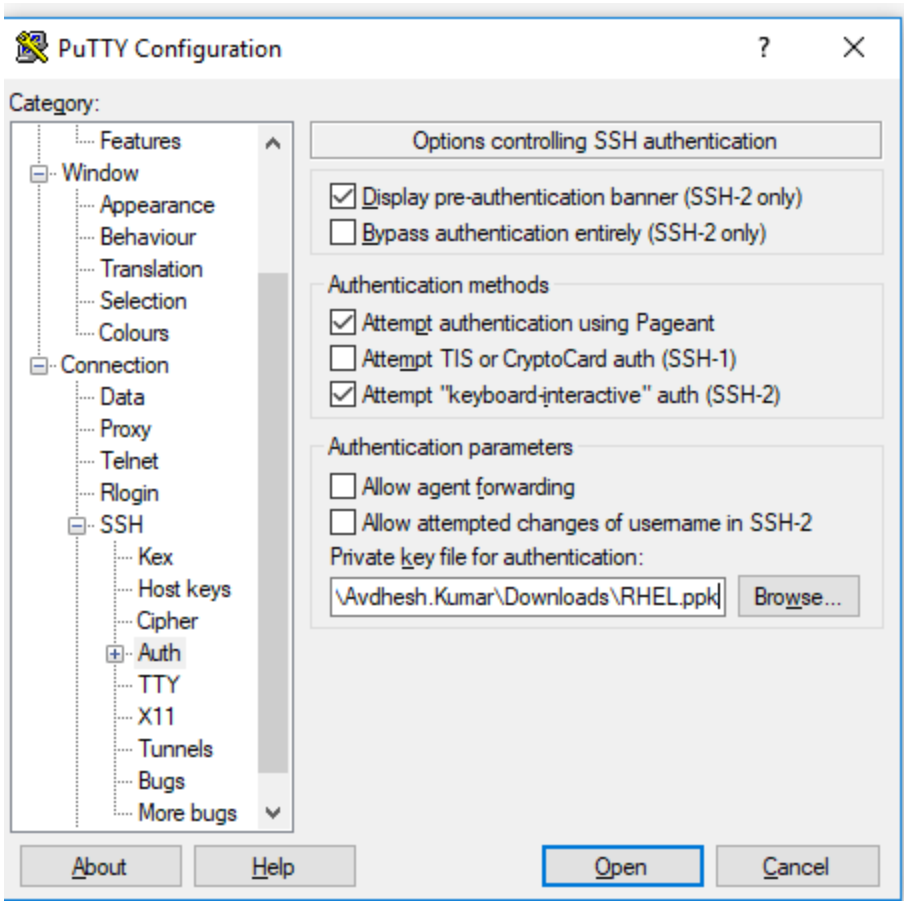
19. Open putty:
Under the **'Host Name'** section, paste the public DNS information of your instance that you just copied.



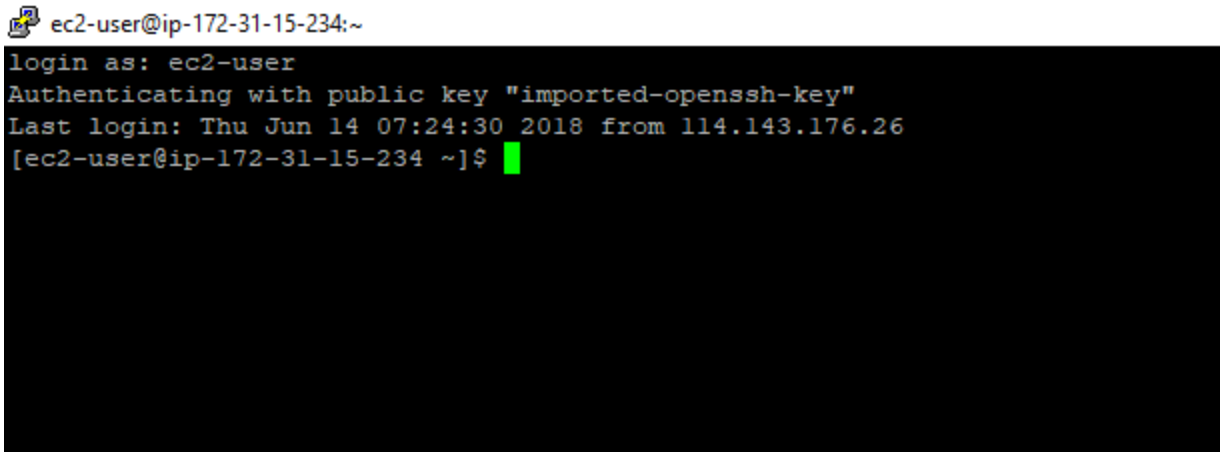
20. On the left-hand side panel, click on **'Connection'**. Then click on **'SSH'** followed by **'Auth'**. In the private key field, click on **'Browse'.**



21. Select the .ppk file(RHEL.ppk) you generated using PuTTYgen and click on ‘Open’. Then, click on open in the PuTTY software.

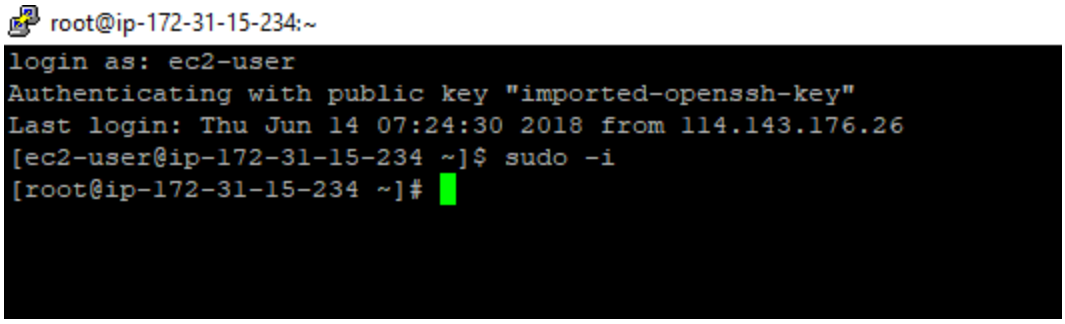


22. Click on **‘Yes’** and login with: **ec2-user**

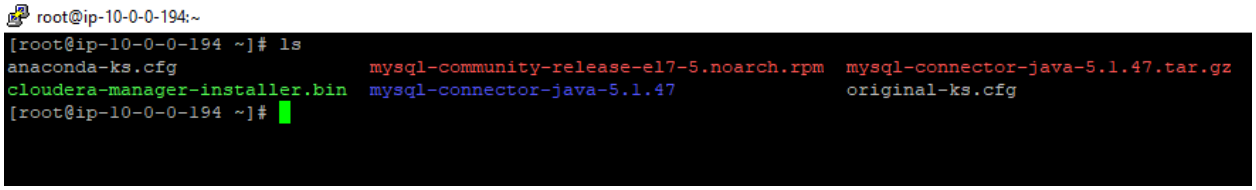


23. Switch to root user using the command

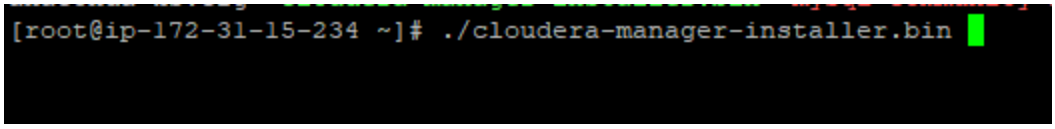
‘sudo -i ’ (enter all commands without quotes unless it is mentioned).



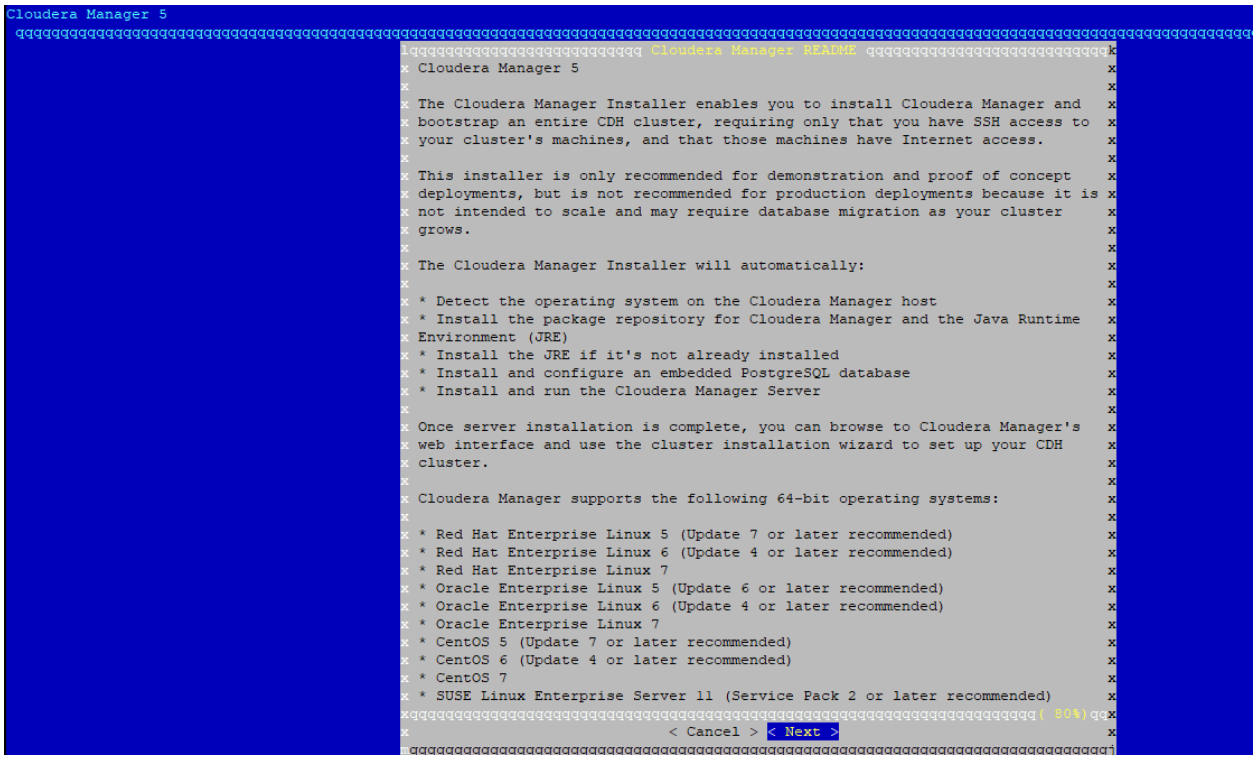
24. Run the command **‘ls’** and verify the ‘**cloudera bin’** is available or not.



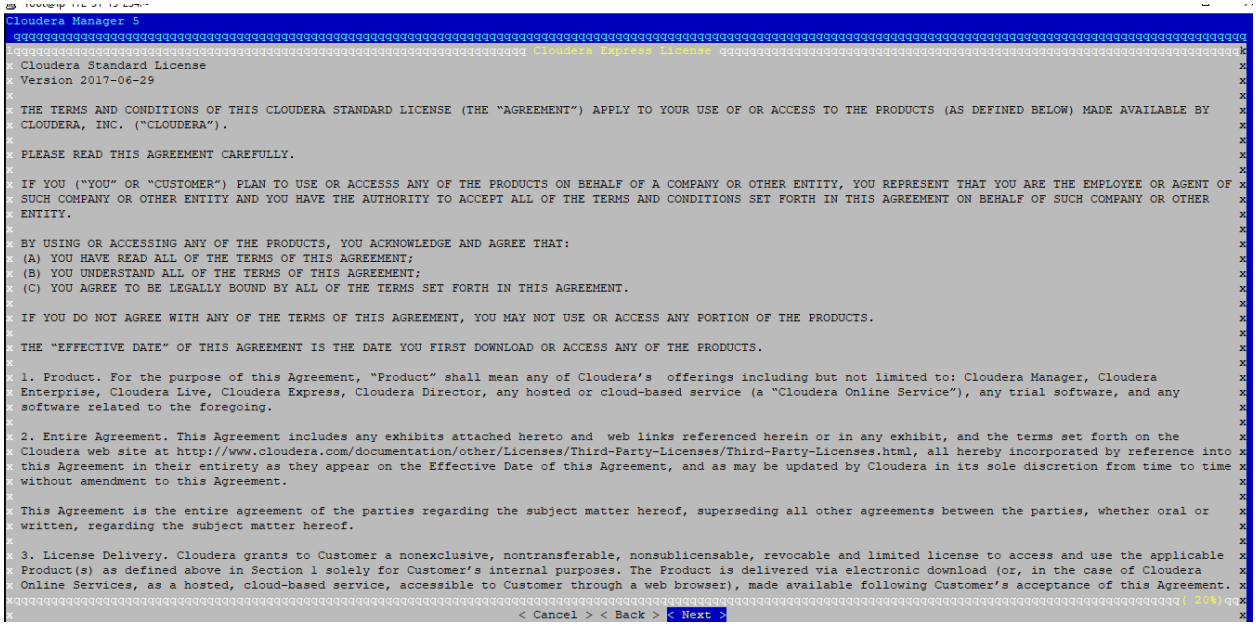
25. Next, run the bin file using the command given below and press the enter
‘ ./cloudera-manager-installer.bin ’



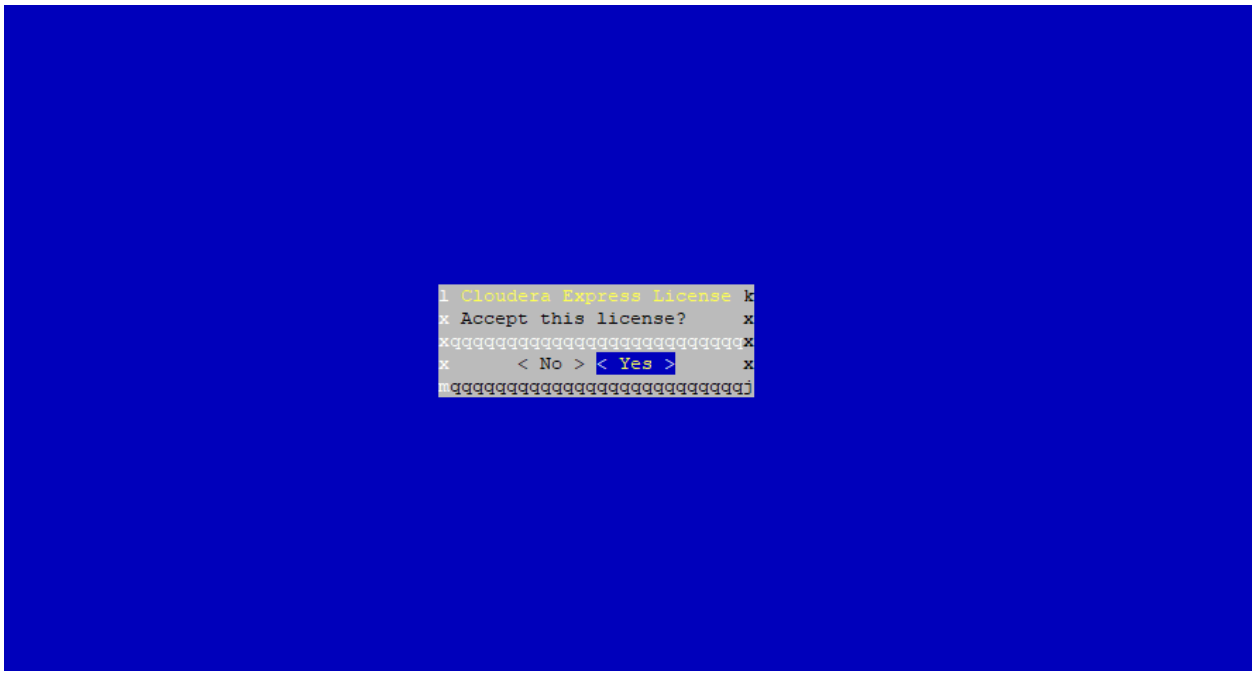
26. Cloudera Manager will start, and the screen will look like the image below. Select **'next'** and press the **enter**.



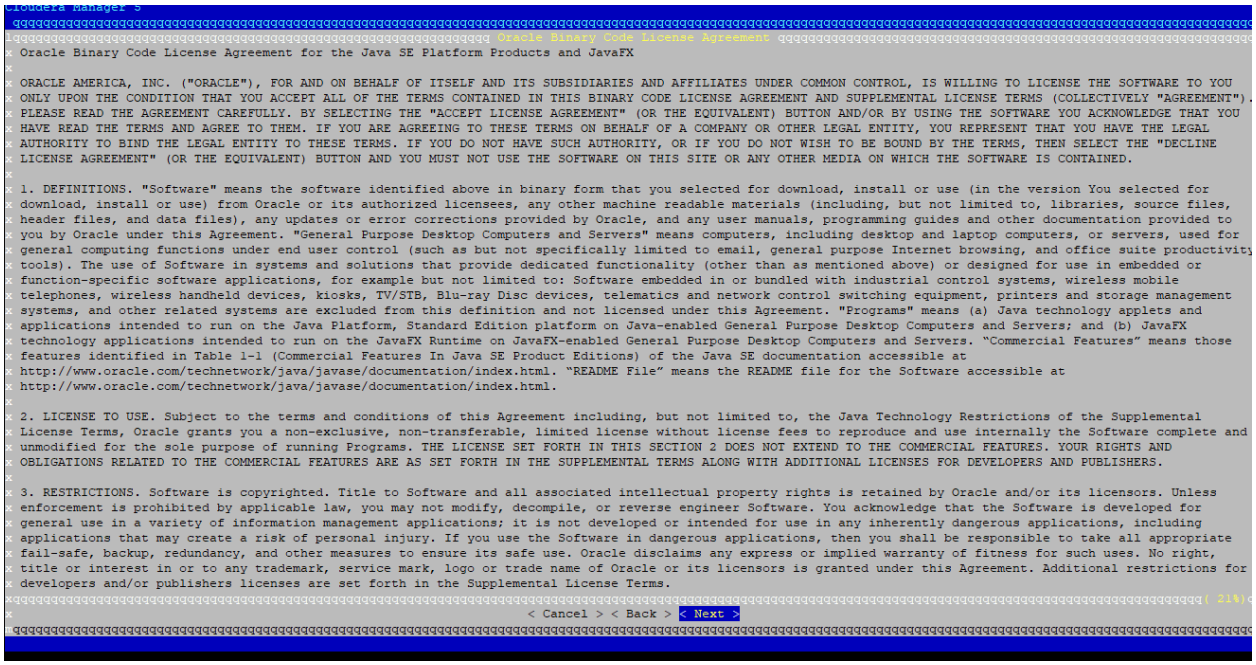
27. Select **next** and press enter :

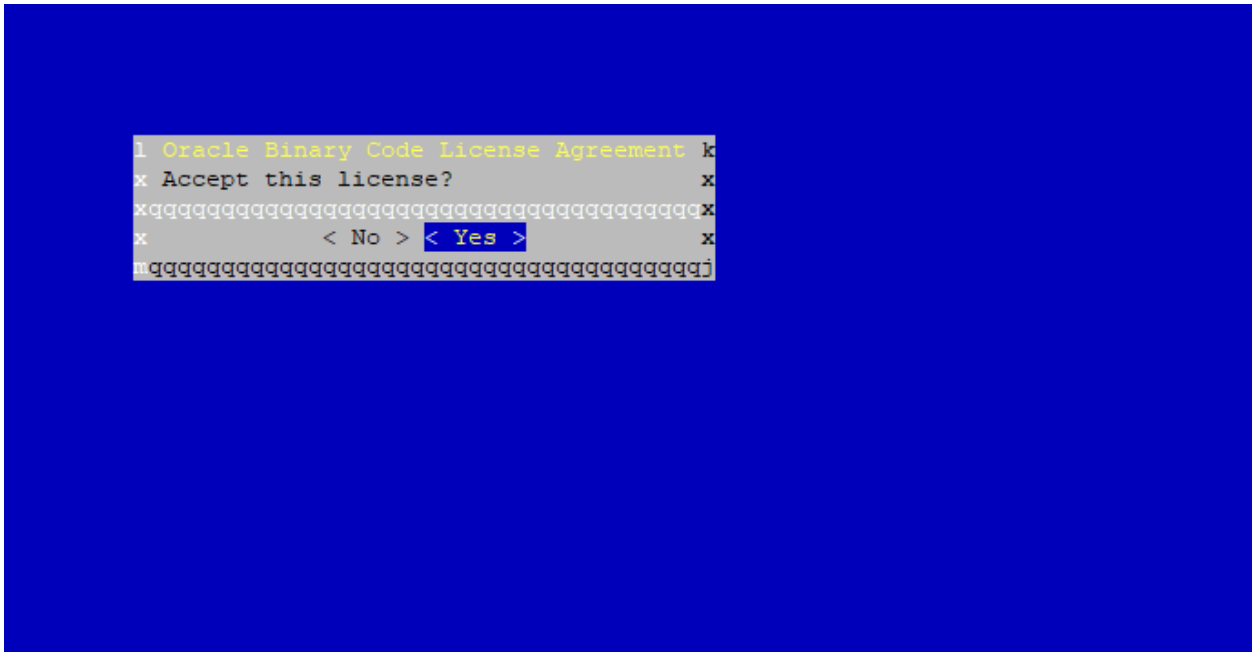


28. Finally, to accept the licence - select **yes**

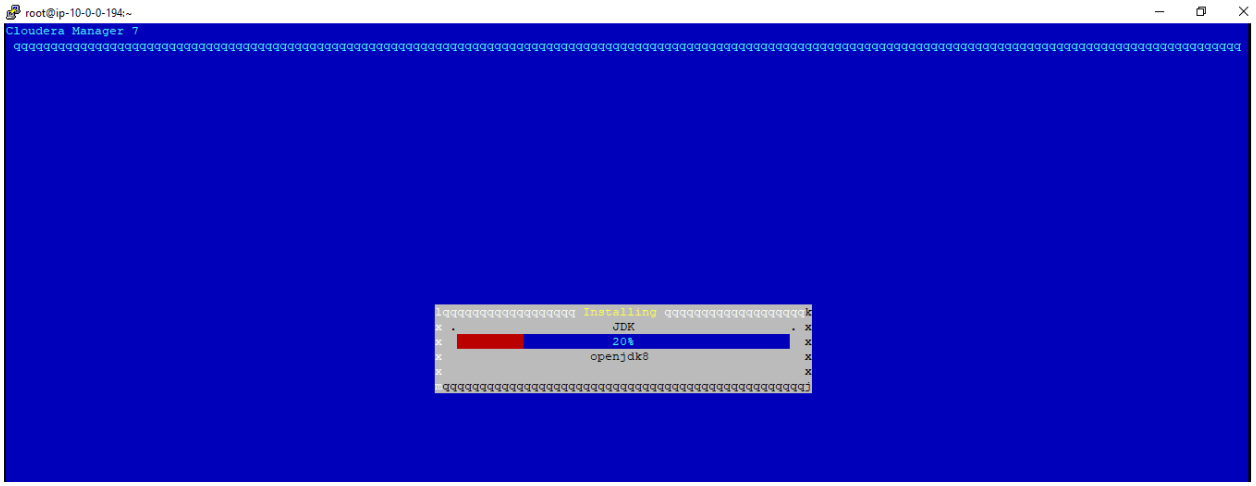


29. Click on **Next** and Accept the Oracle Binary licence

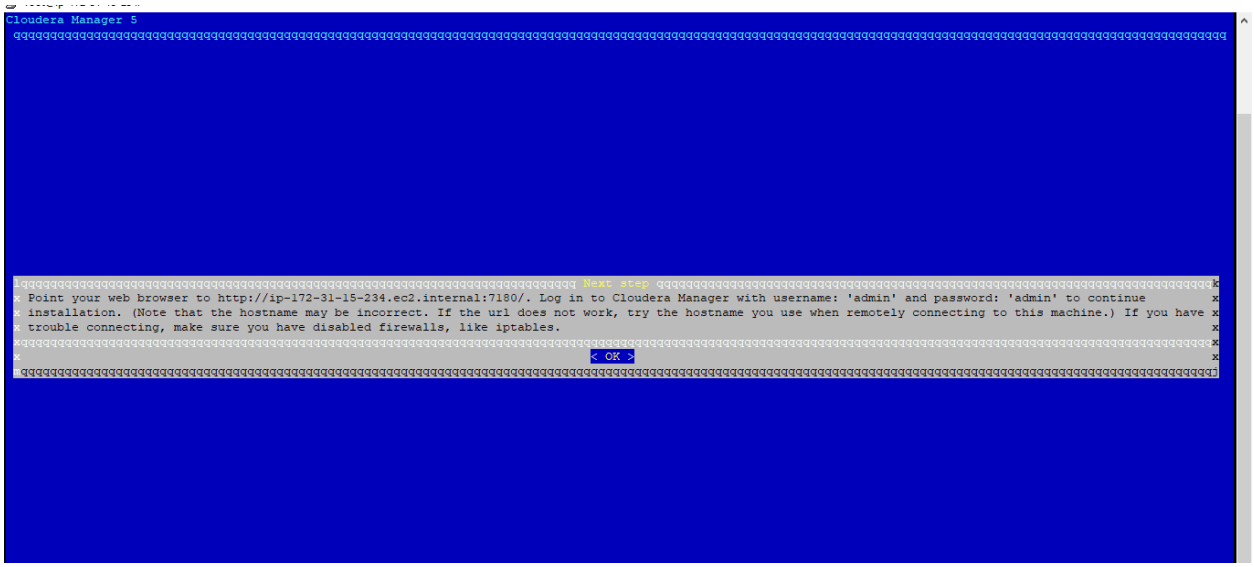




It should take around 10-15 mins

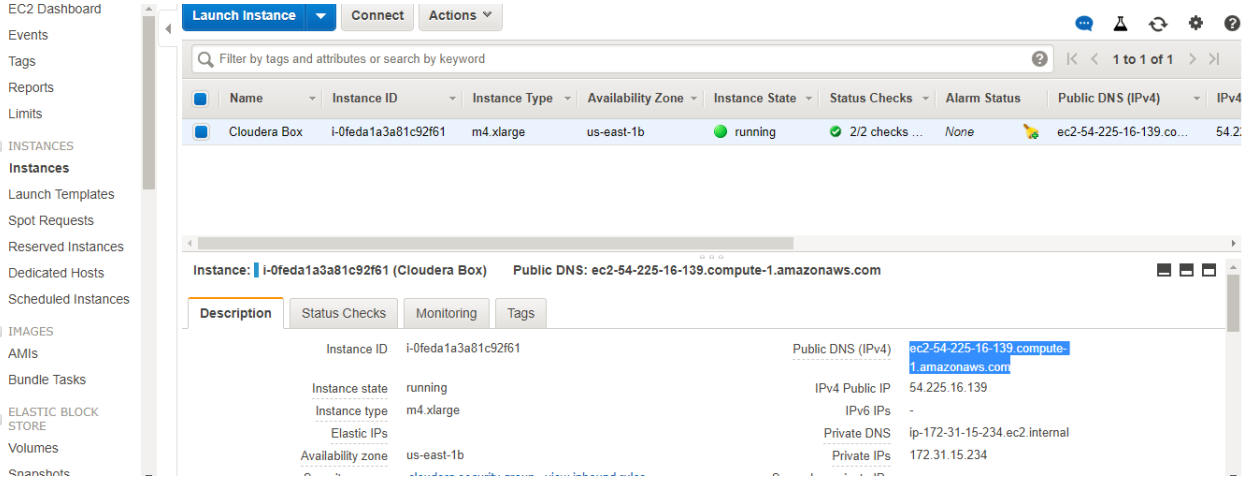


30. Wait until the Cloudera manager is installed and click on **ok**. (as shown in below image)

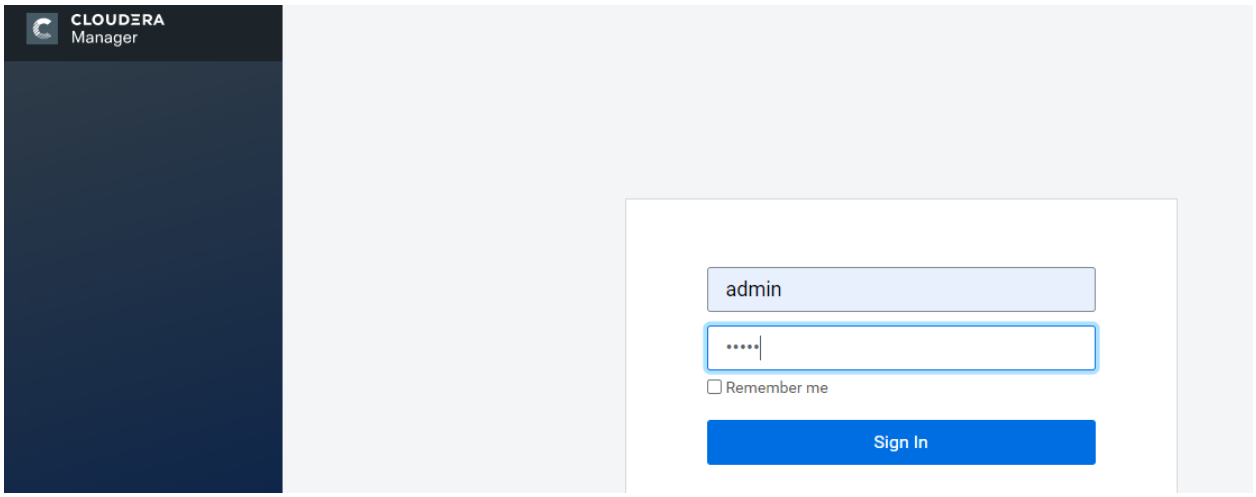


31. Now open a web browser on your local machine; go to the following address **<public-domain-of-ec2>:7180**. This way you will be able to access the Cloudera Manager service that's running on the EC2 instance.

- To get the public domain of ec2, go to the EC2 dashboard, and click on your instance. Then, copy your 'Public DNS'.

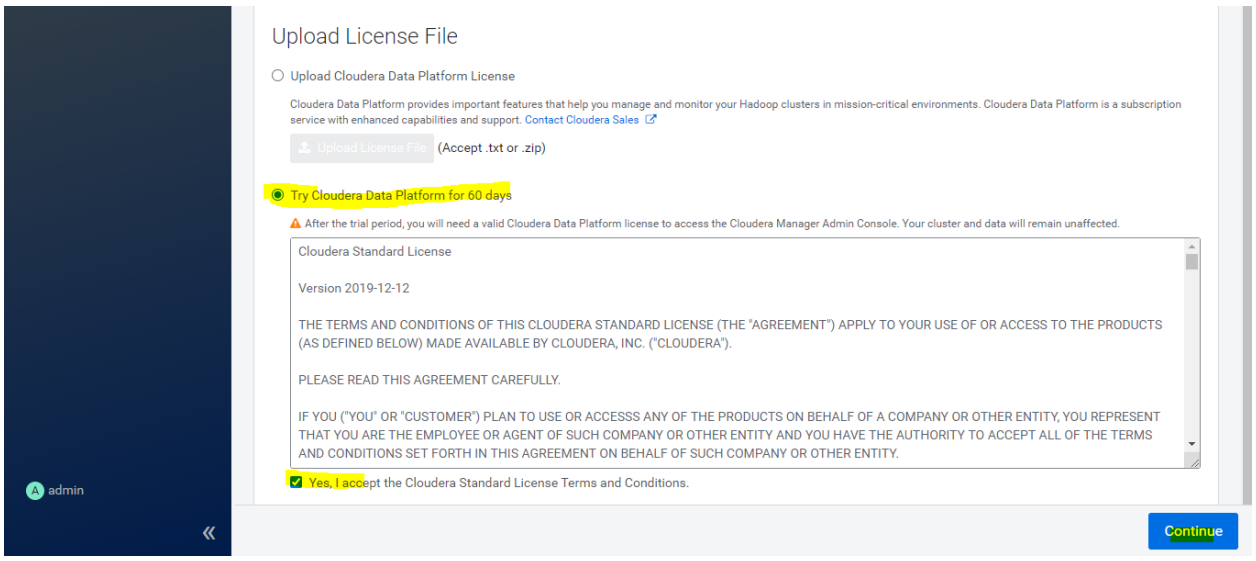


32. Put the copied public IP in place of **<public-domain-of-ec2>** in **<public-domain-of-ec2>:7180**, and open it in your browser. The following page will appear after a few minutes.

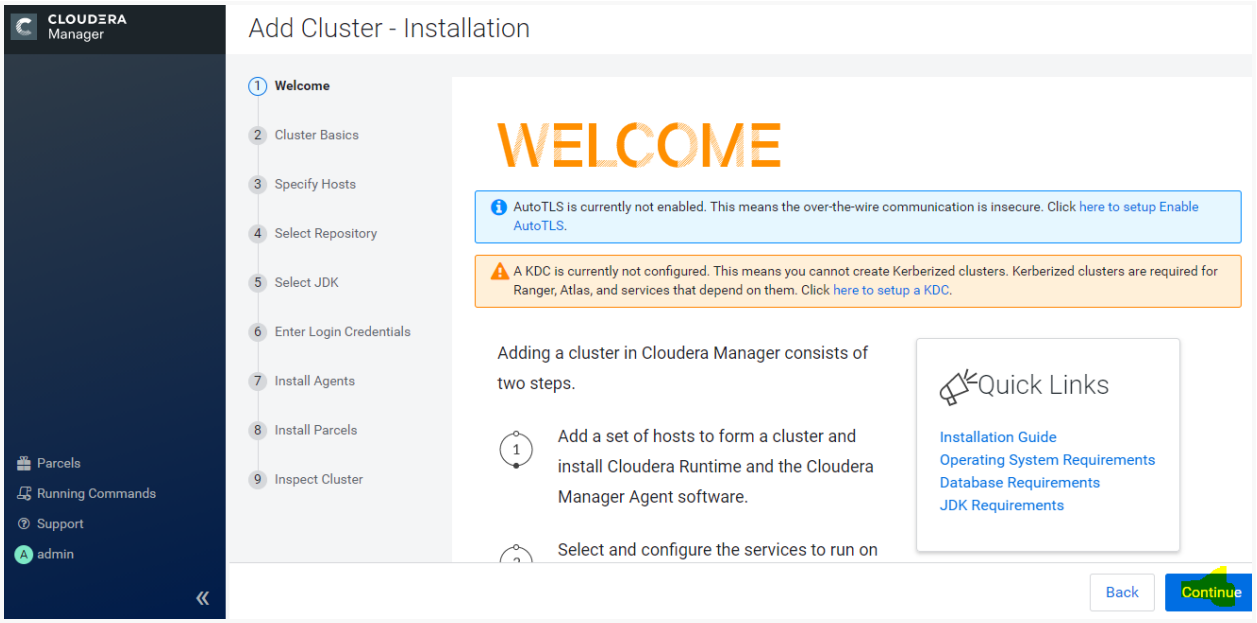


33. Now, log in to Cloudera Manager using **username: admin** and **password: admin**.

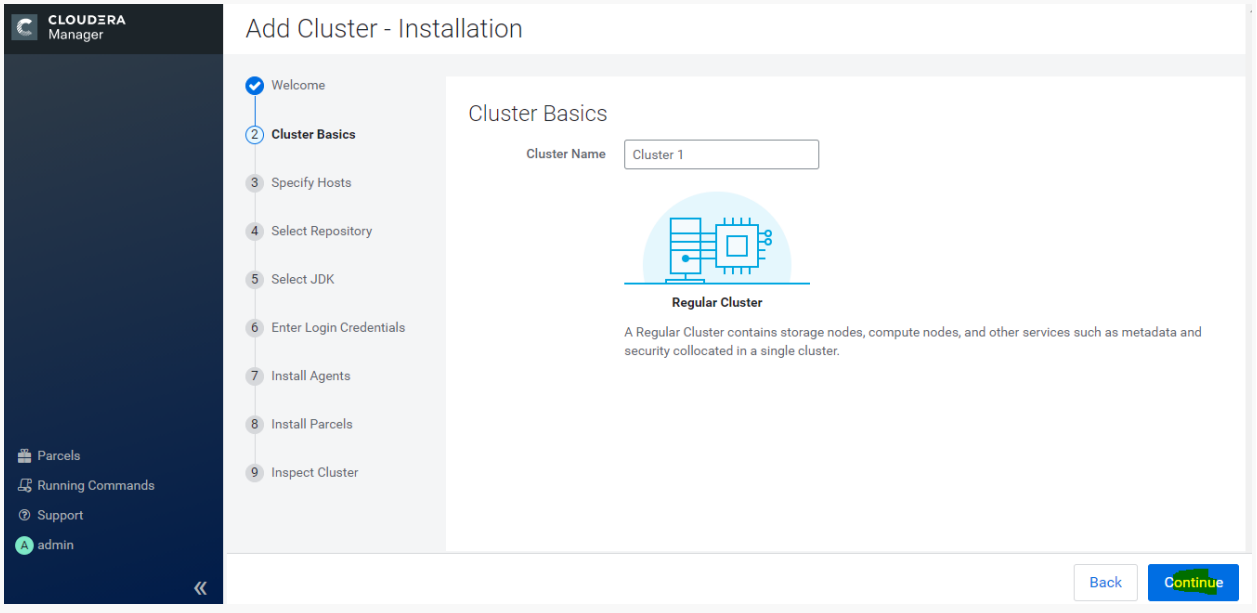
34. Choose Cloudera data platform for 60 days and Accept the terms and conditions, and then, keep clicking on **'Continue'**.



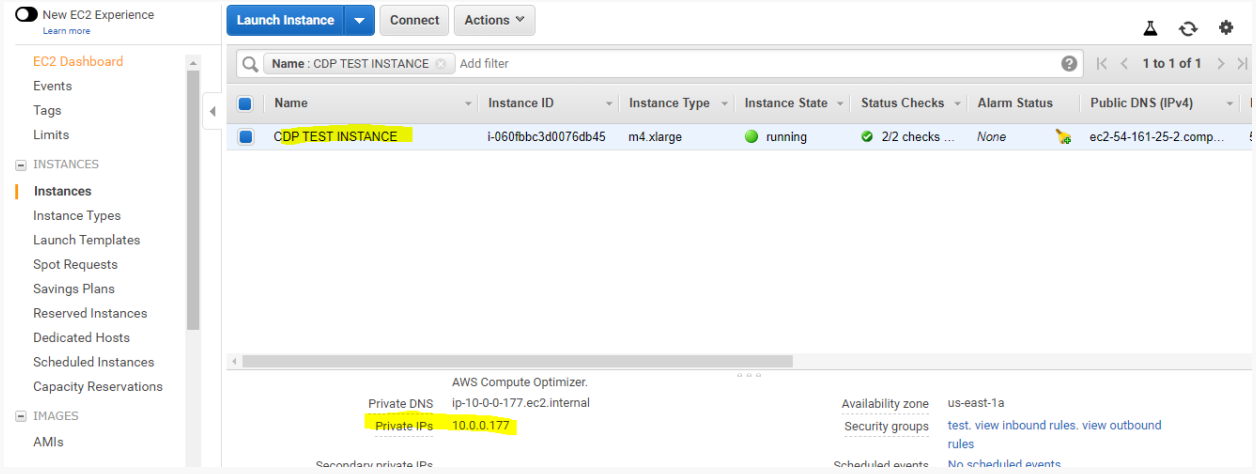
35. Click on **Continue**



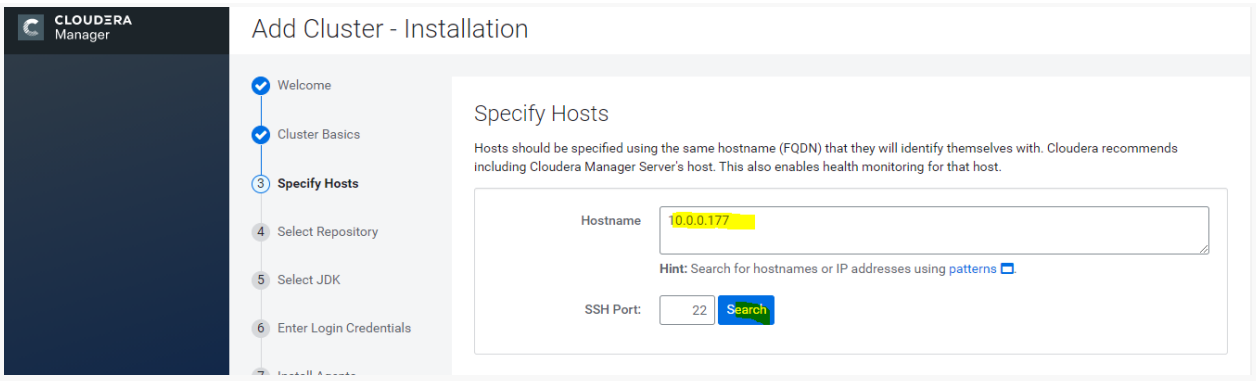
36. Click on **'Continue'**:



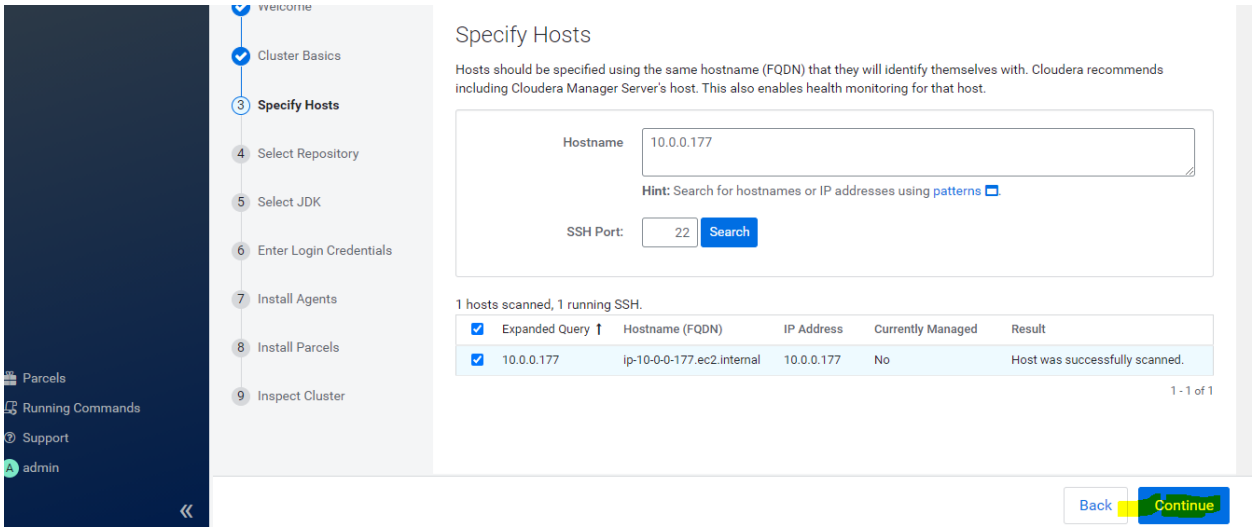
37. When you're asked to search for the instance, give your 'Private IP'.



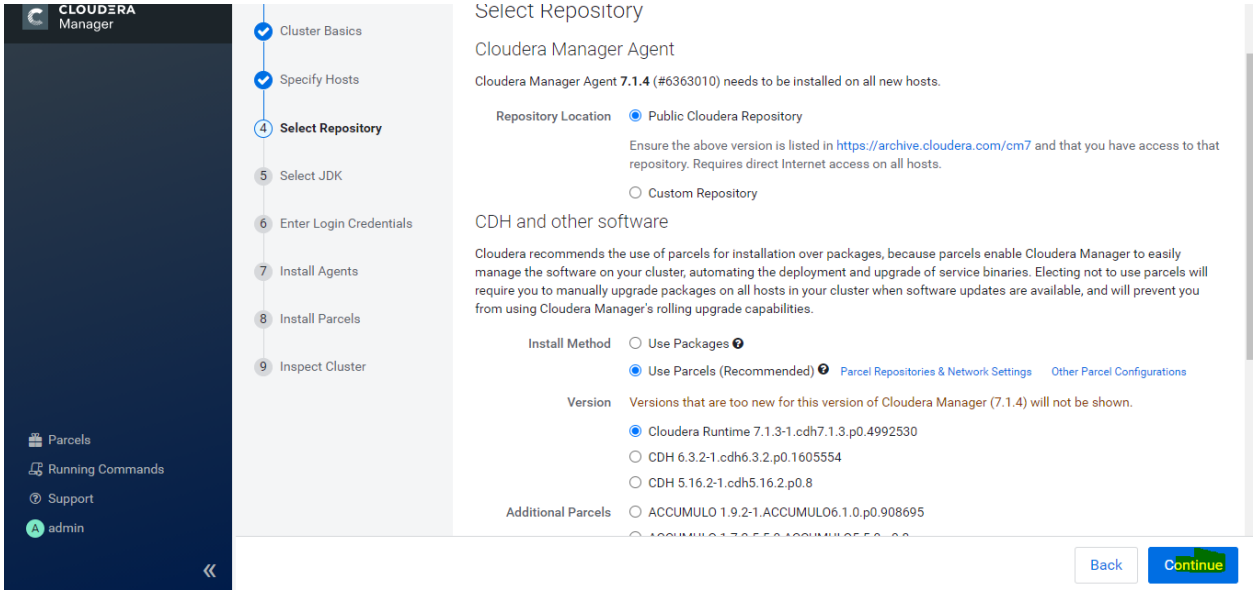
38. After this, the following screen will appear. Type your private IP in the box, and click on **'Search'**



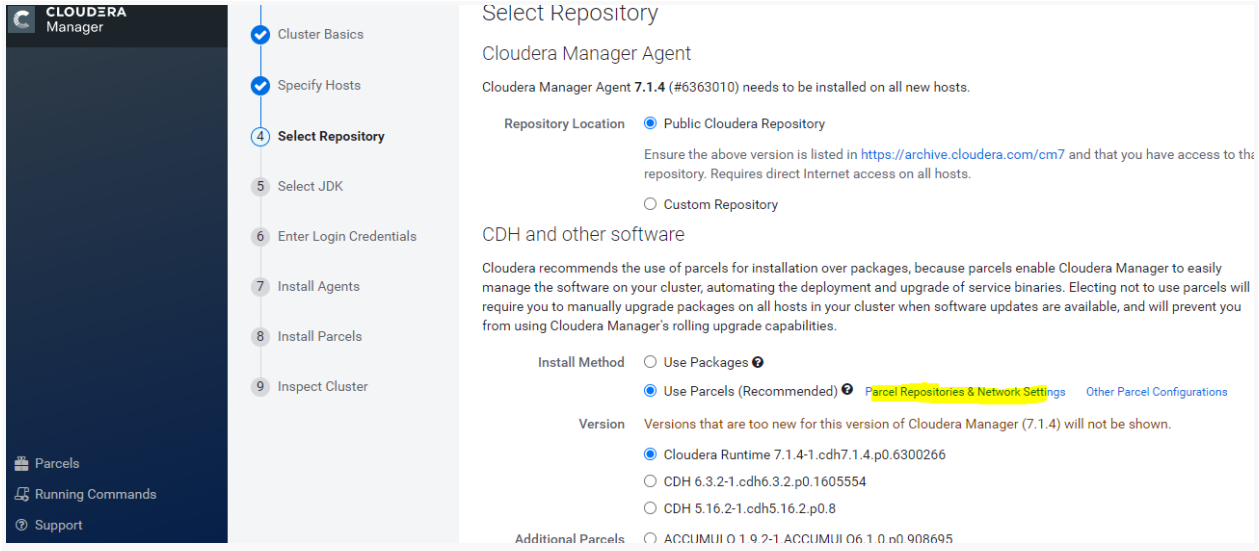
39. Select your private IP from the search results and click on **'Continue'**



40. Click on '**Continue**':



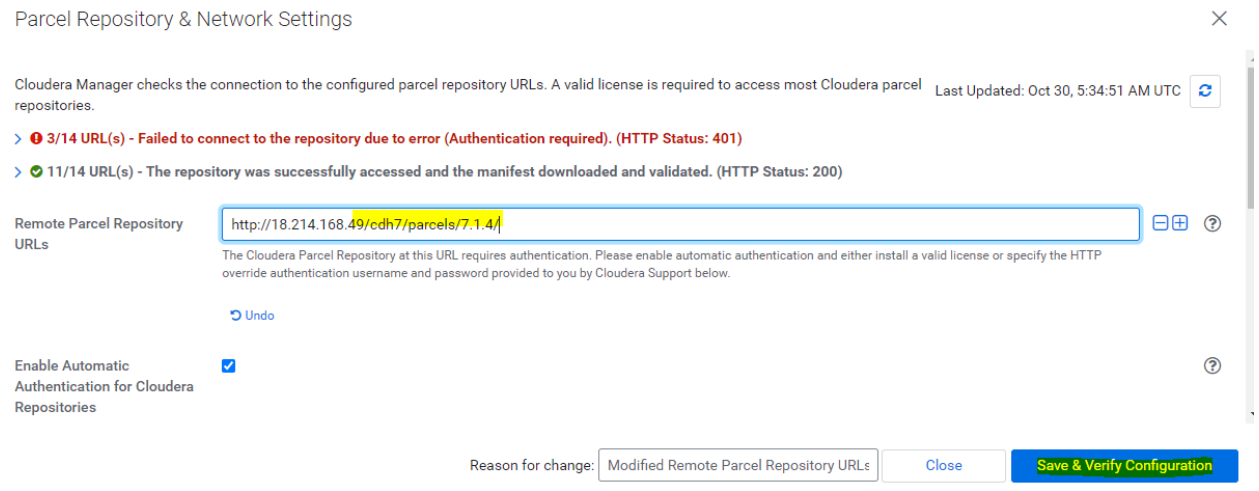
Click on **Parcel Repostories & Network setting**



Remove other parcels using(-)

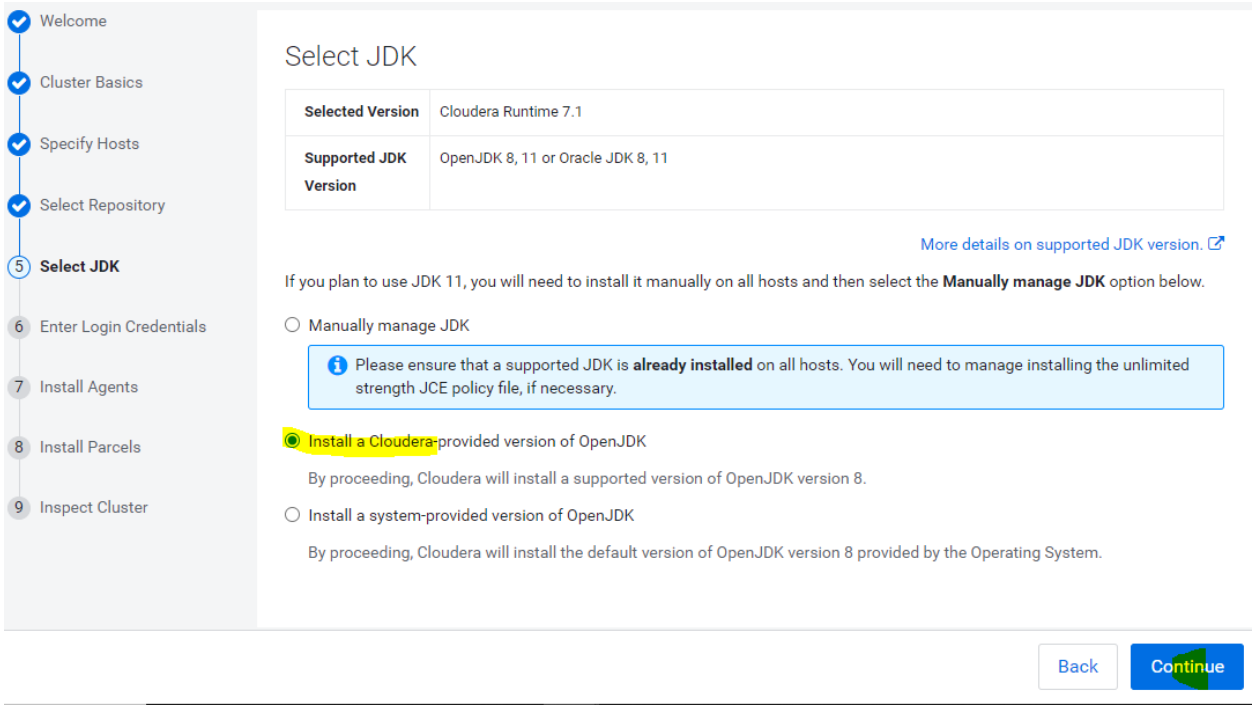
And paste the below url:

http://18.214.168.49/cdh7/parcels/7.1.4/



Click on **Save & Verify Configuration.**

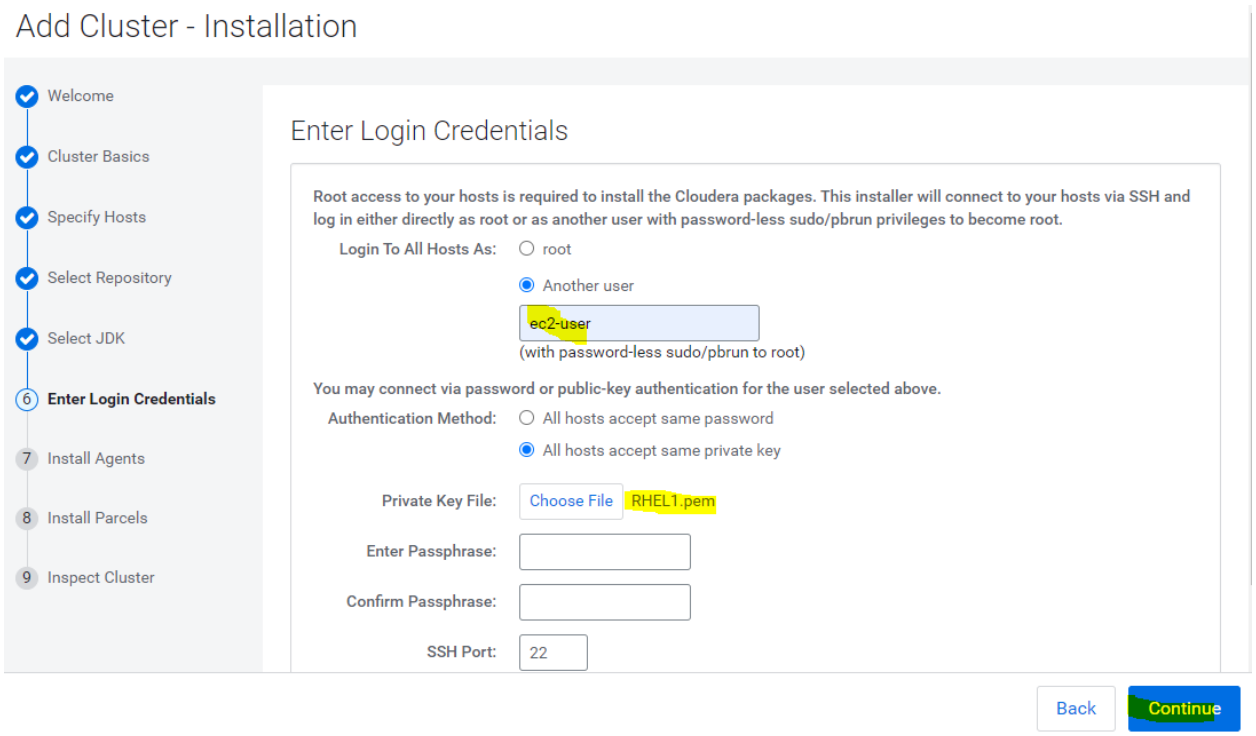
41. Select the Install a Cloudera-provided version of open-jdk and encryption policy then click on **'Continue'.**



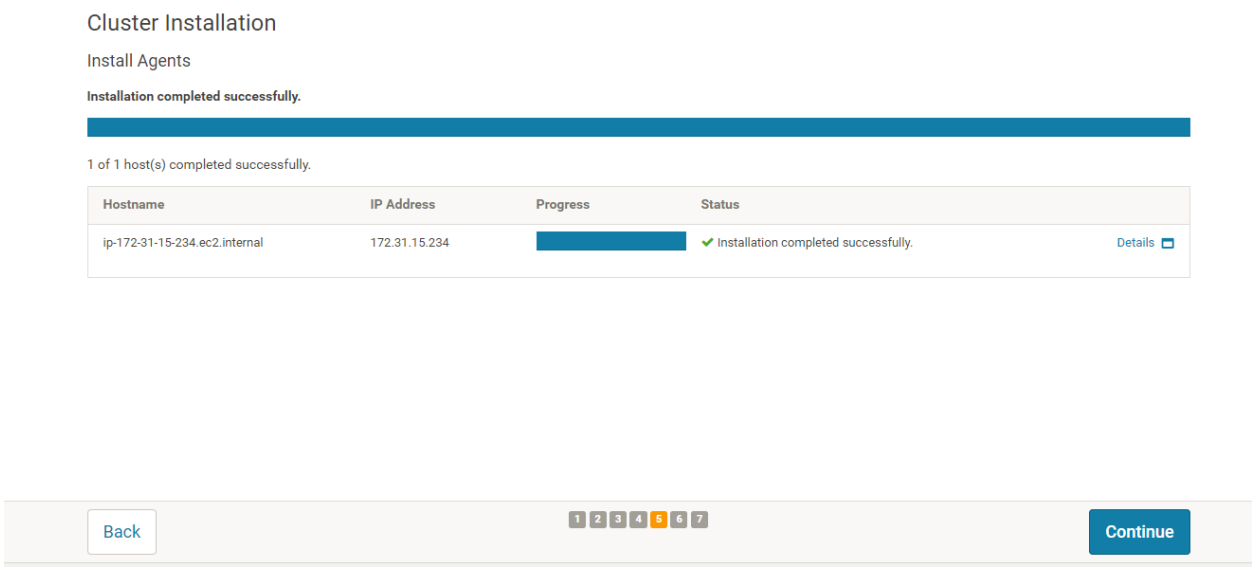
42. Click **'Continue'** again.

43. After this, the following screen will appear. Select **Another User** and type **ec2-user**. Select **'All hosts accept the same private key'** as the authentication method. Browse the **.pem** file in the **Private Key File** option, and click on **'Continue'**.

- Our case: RHEL.pem

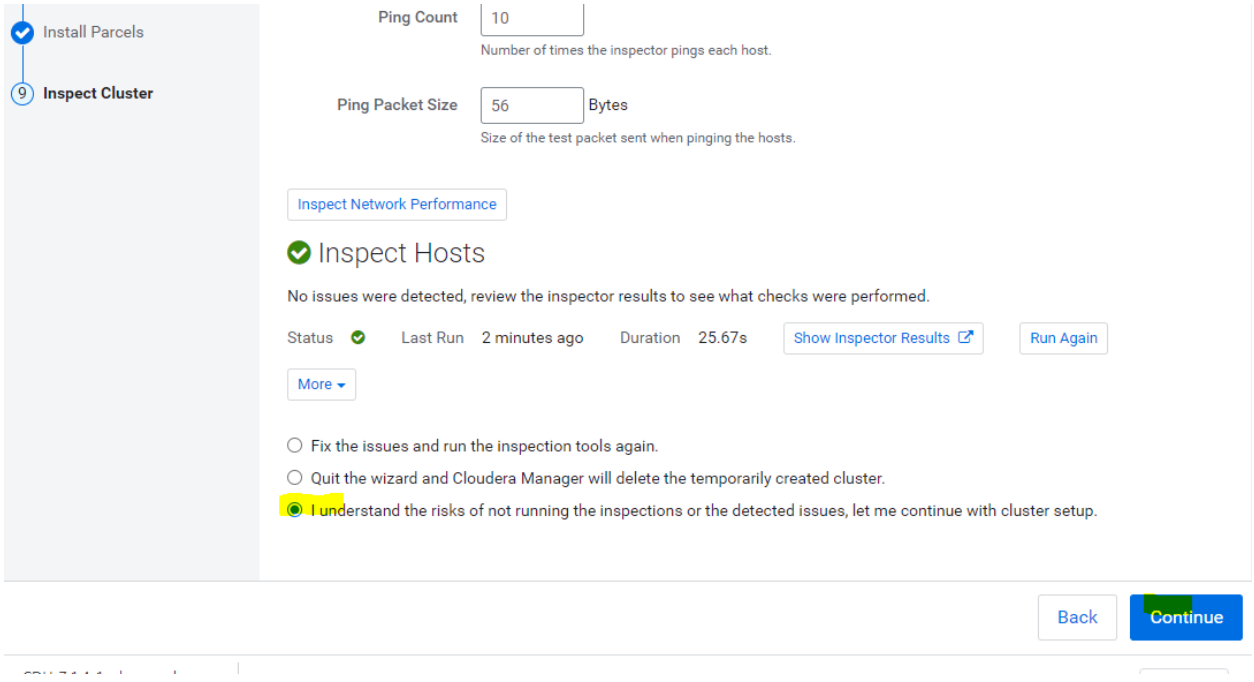


- Click on '**Continue**'

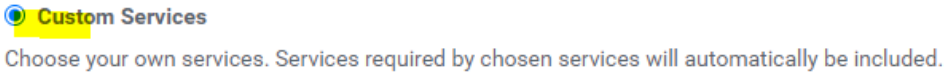


- Click on '**continue**': it takes 30 mins to download and distribute the parcels

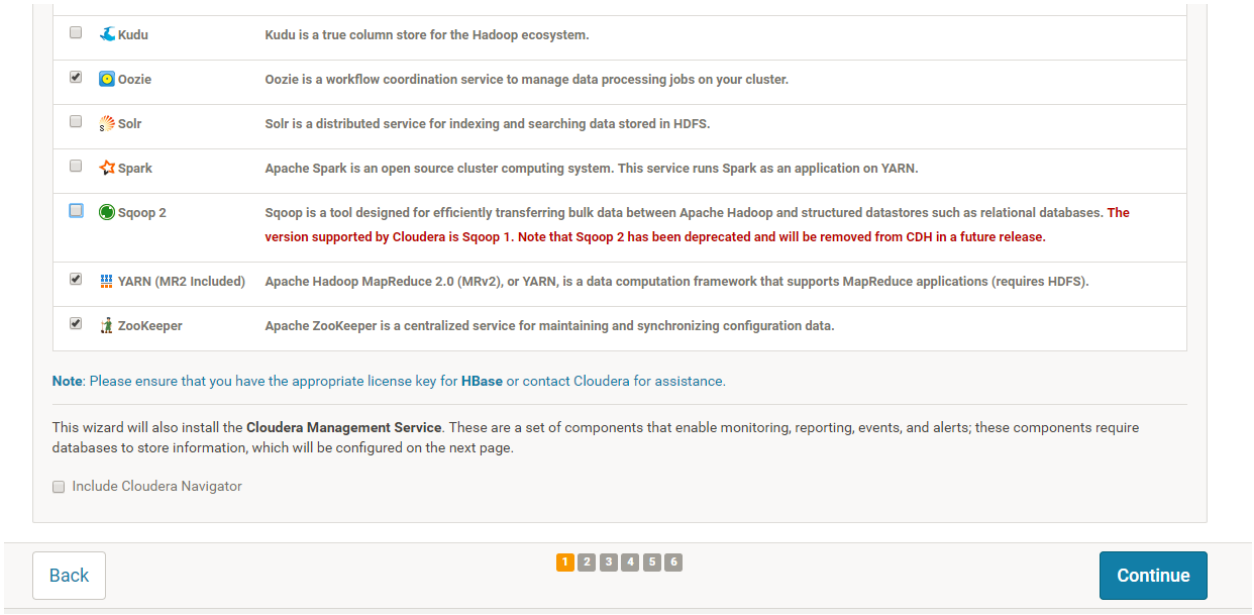
44. Click on '**Continue' after selecting the below radio button as highlighted**.



Select custom services.

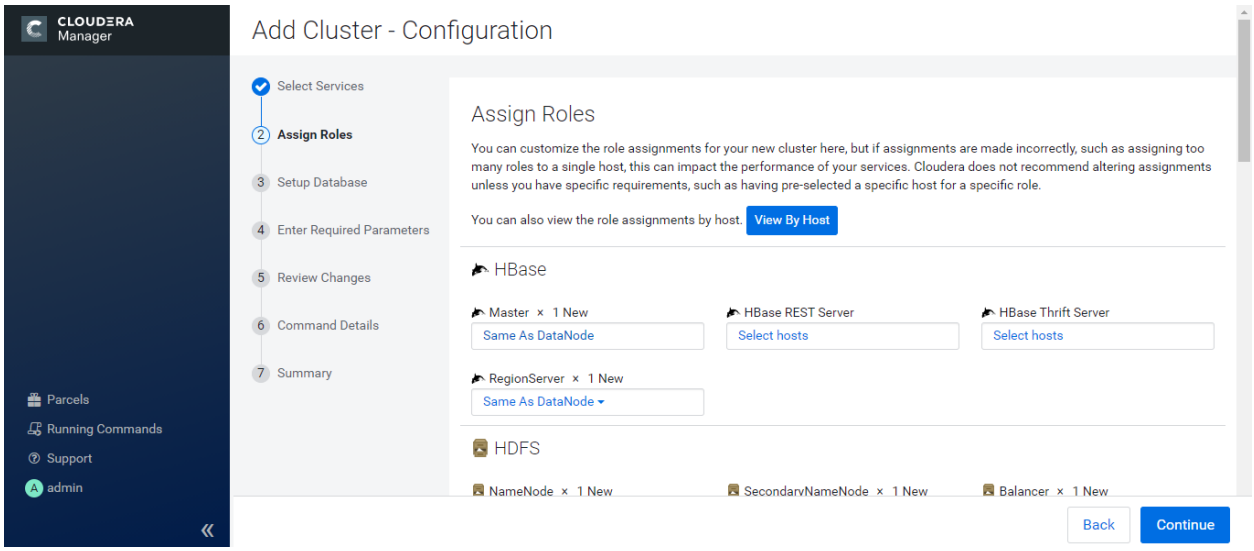


45. After this, the following screen will appear; select **Custom Services and choose HBase, HDFS, Hive,Hive on Tez, Hue,Livy, Phoenix, spark,, Tez, Yarn, ZooKeeper,** and then, click on **Continue**



-

- Click on **‘Continue’**



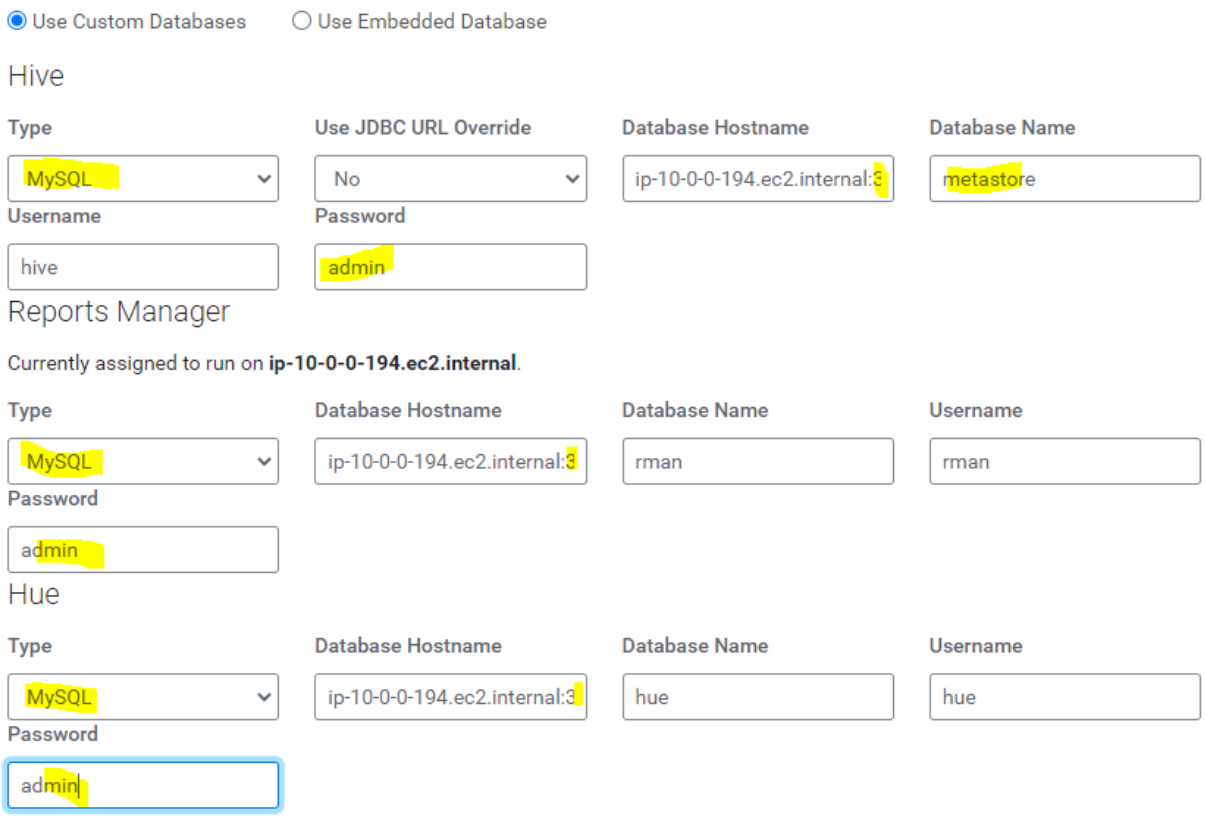
46. Select ‘**use custom data bases’.**

Note :- we already installed MySQL on that AMI with the databases metastore rman hue and oozie.

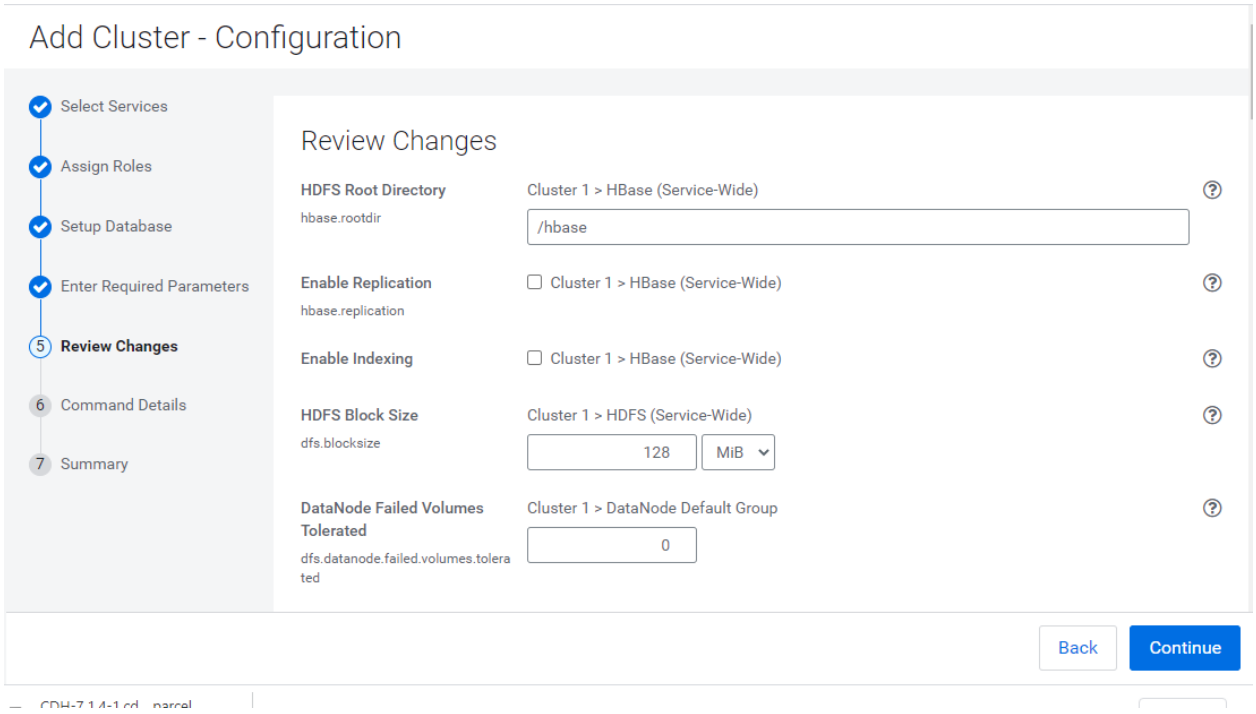
47. Please keep all the usernames and passwords for all the services: Hive, Hue, etc. in a safe place. After paste/ enter these database names and passwords, scroll down, and then, click on **'Test Connections'**.

Change Database Hostname port number is- 3306
Database type - MySQL

For hive service:**-** database name: **metastore** and password is **admin.**
For rman service:- database name: **rman** and password is **admin**
For hue service:- database name: **hue** and password is **admin**
For oozie service :- database name: **oozie** and password is **admin**

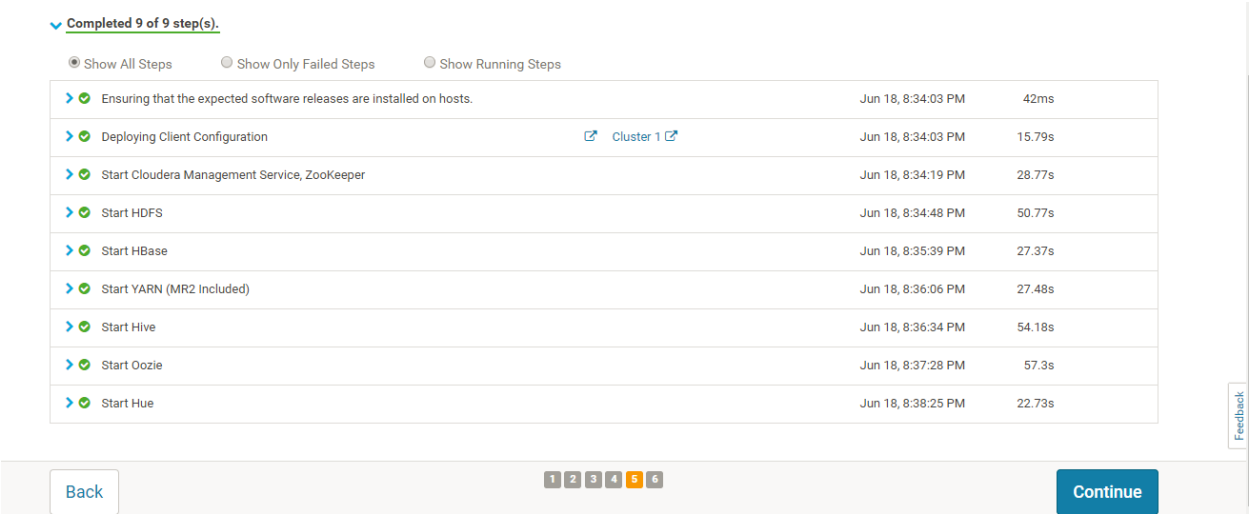


- Click on **Continue**:

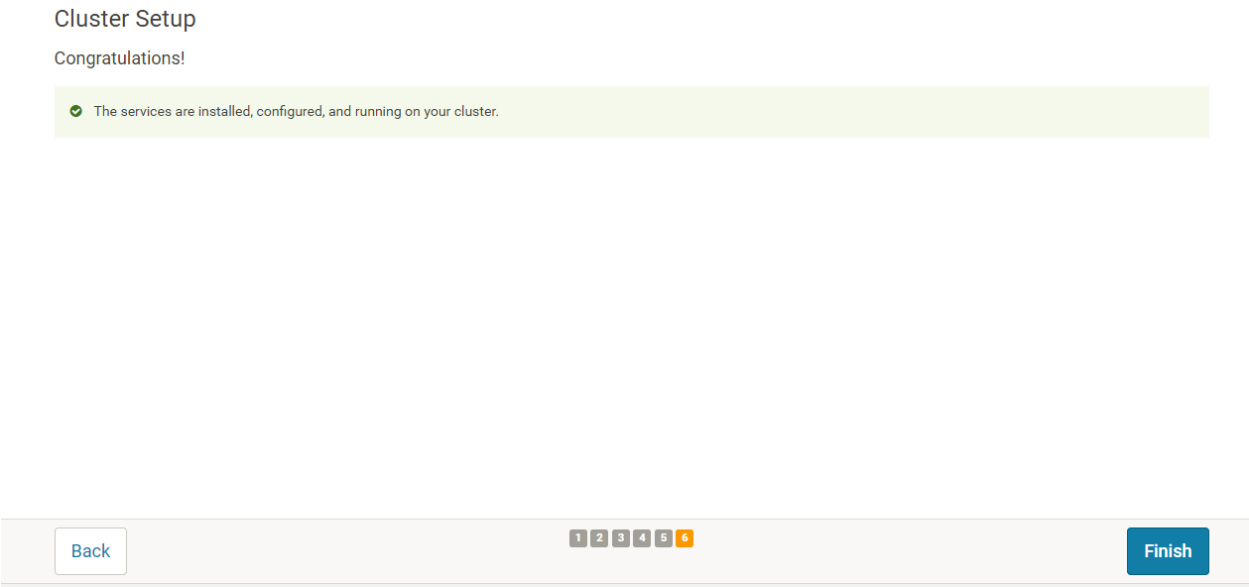


-

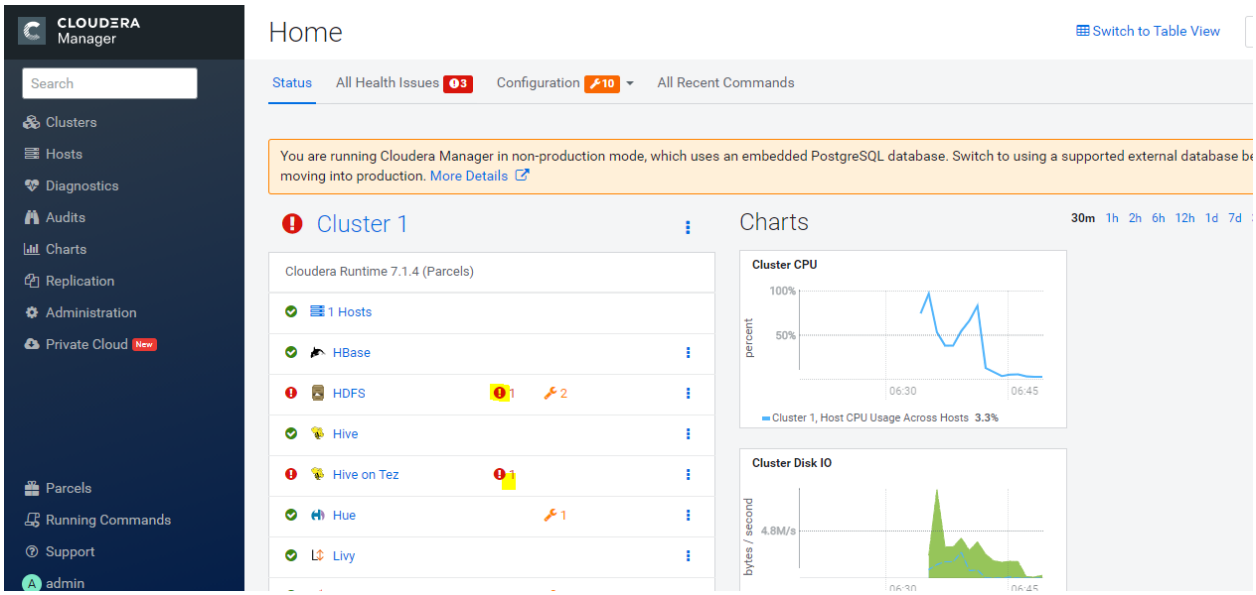
- Deploying all the services takes some time. After this is complete, click on **Continue**



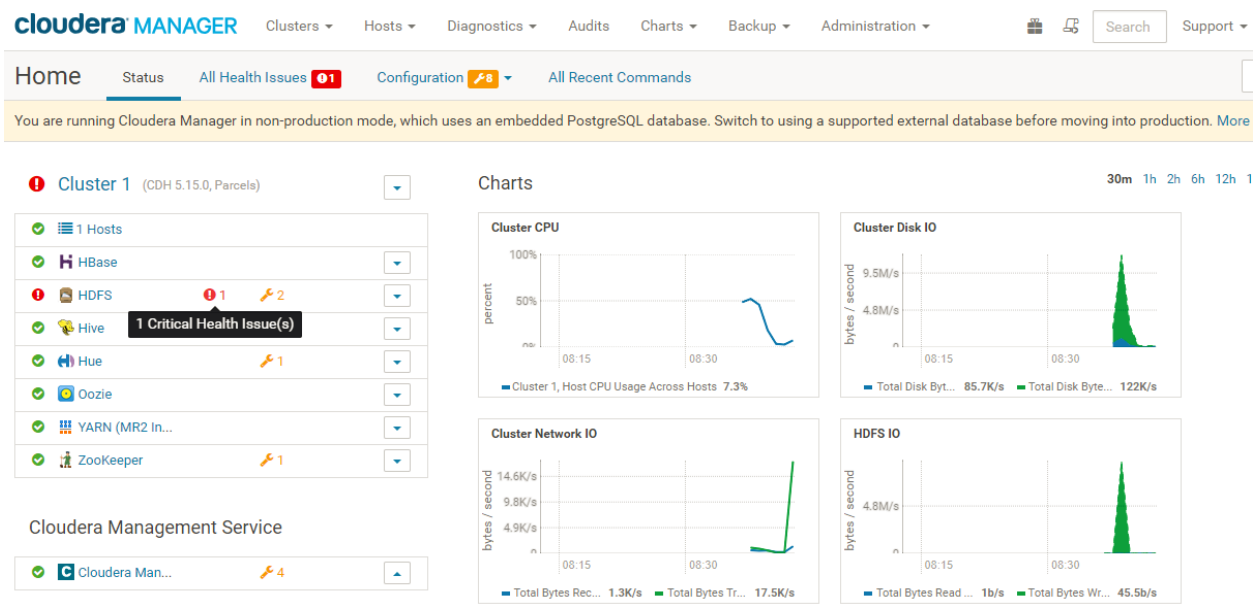
- The following screen will appear on the successful set-up of the cluster. Then, click on **Finish**.



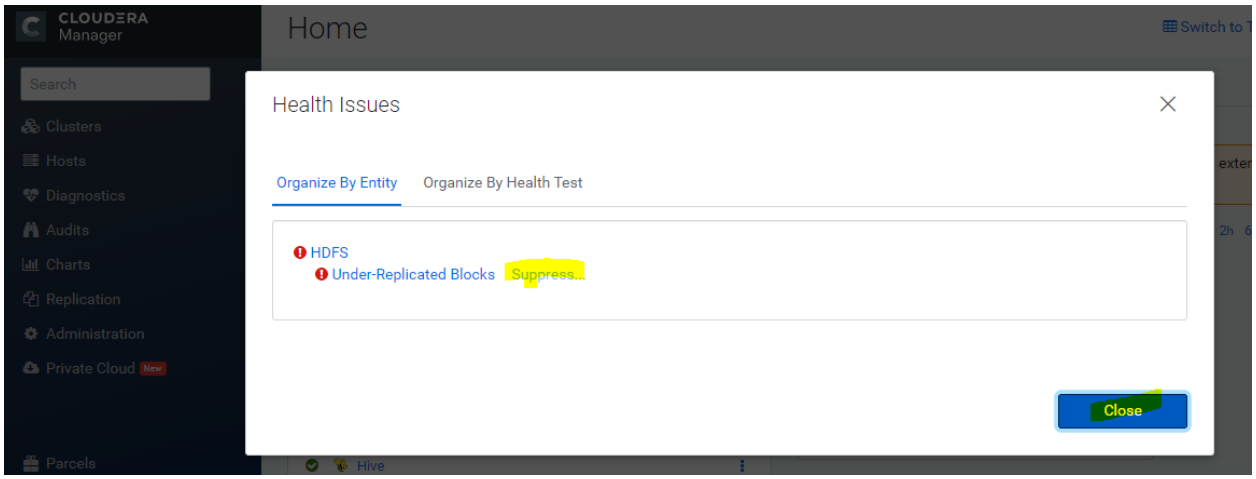
48. After this, the following image will appear. Kindly check whether all the services are green. Please note that it takes some time for all the services to be active



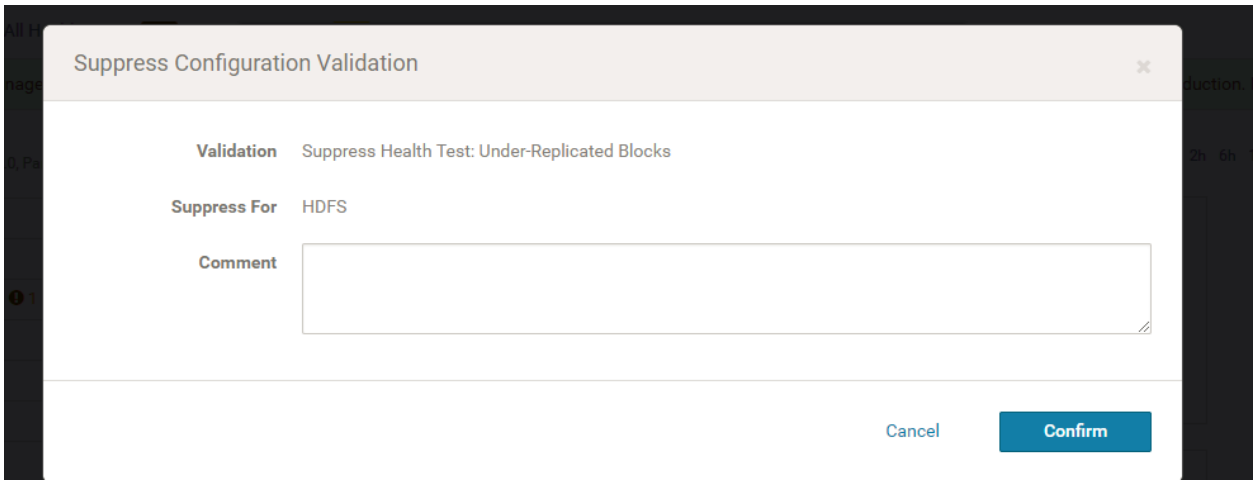
49. Click on **'Critical Health Issue(s)'** (if any).



50. Click on **‘suppress’**:

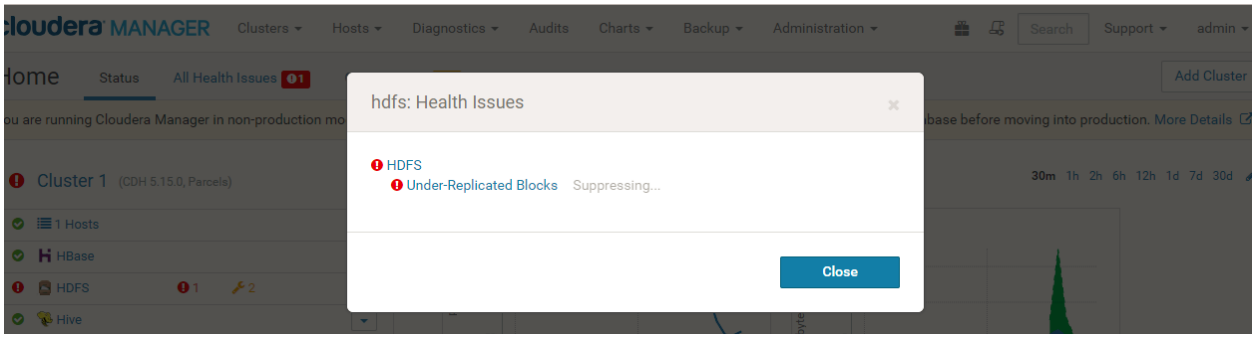


- Click on Confirm:

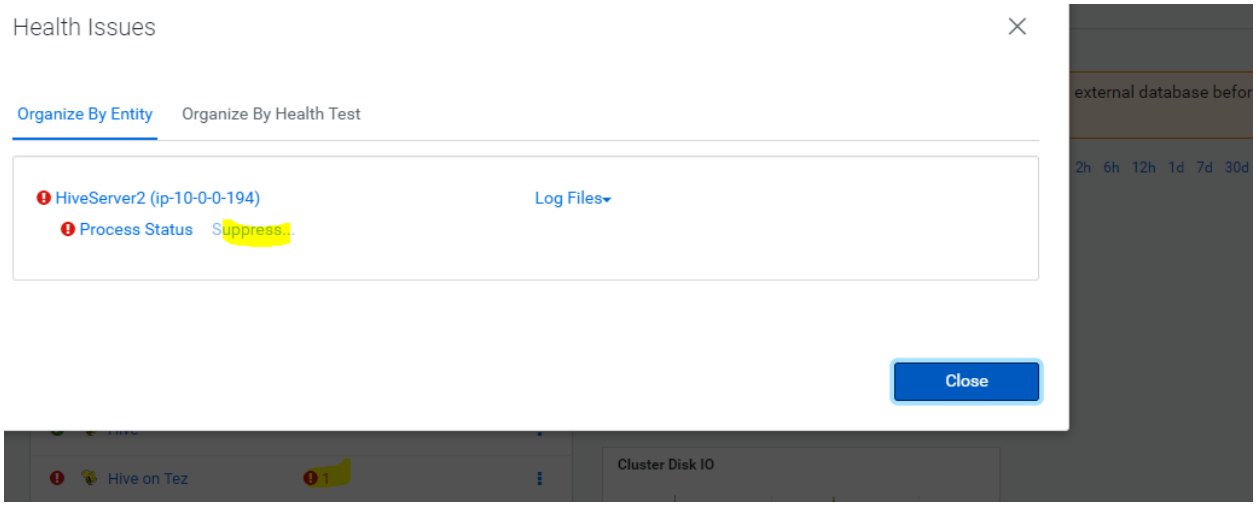


-

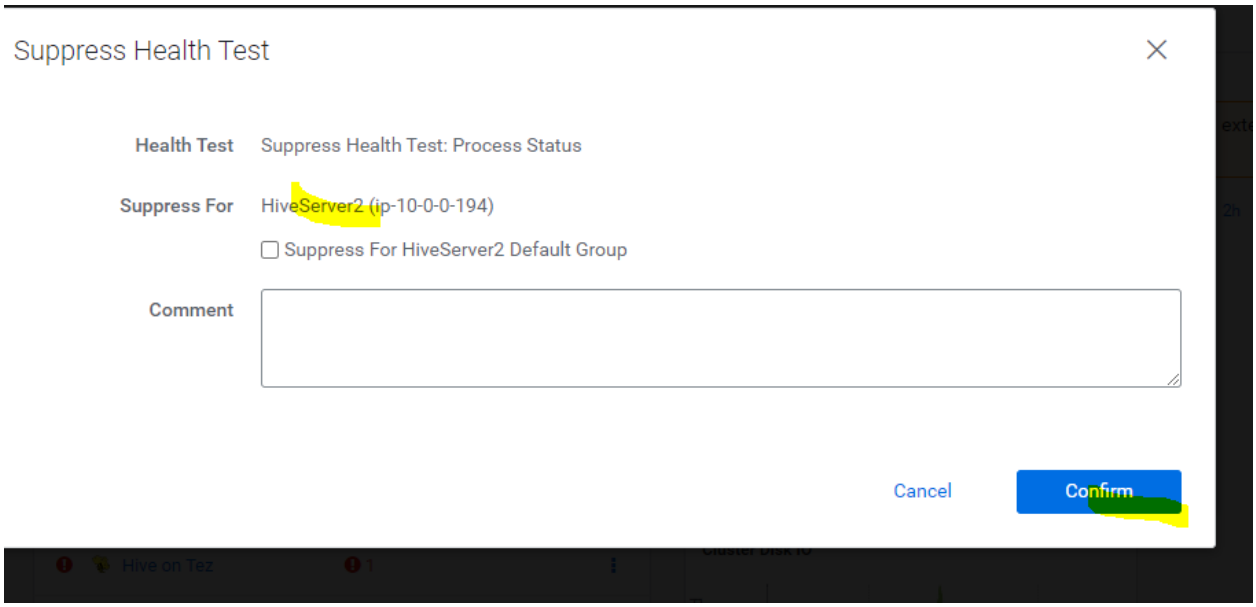
- Click on Close.



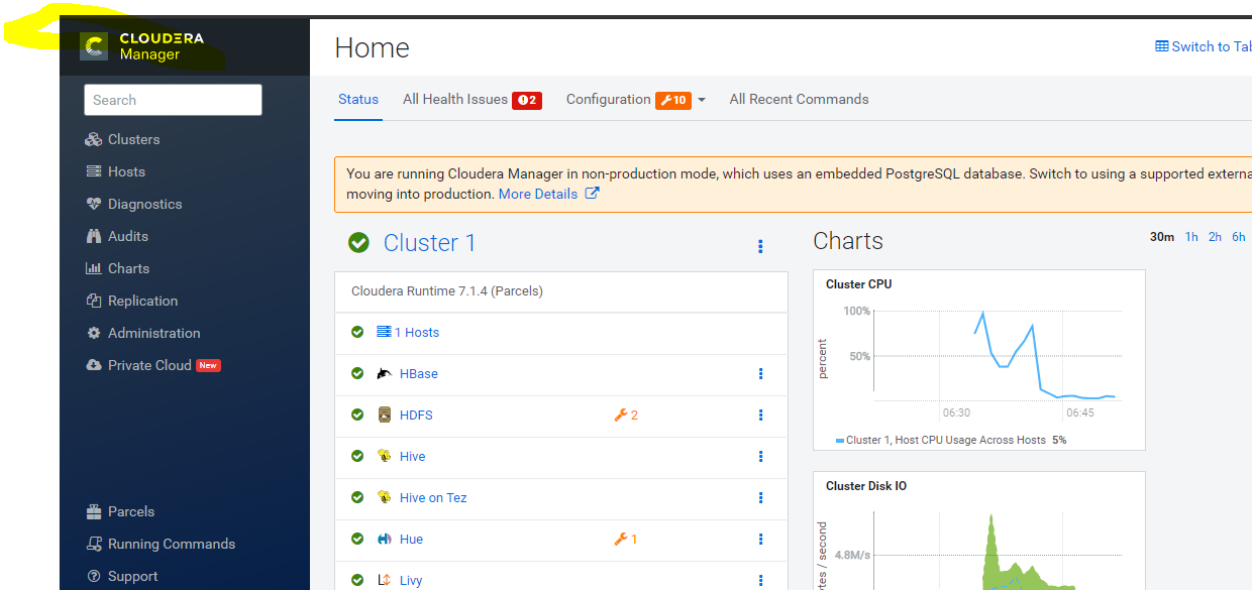
Do the same for the window below



Click on hiveServer2.

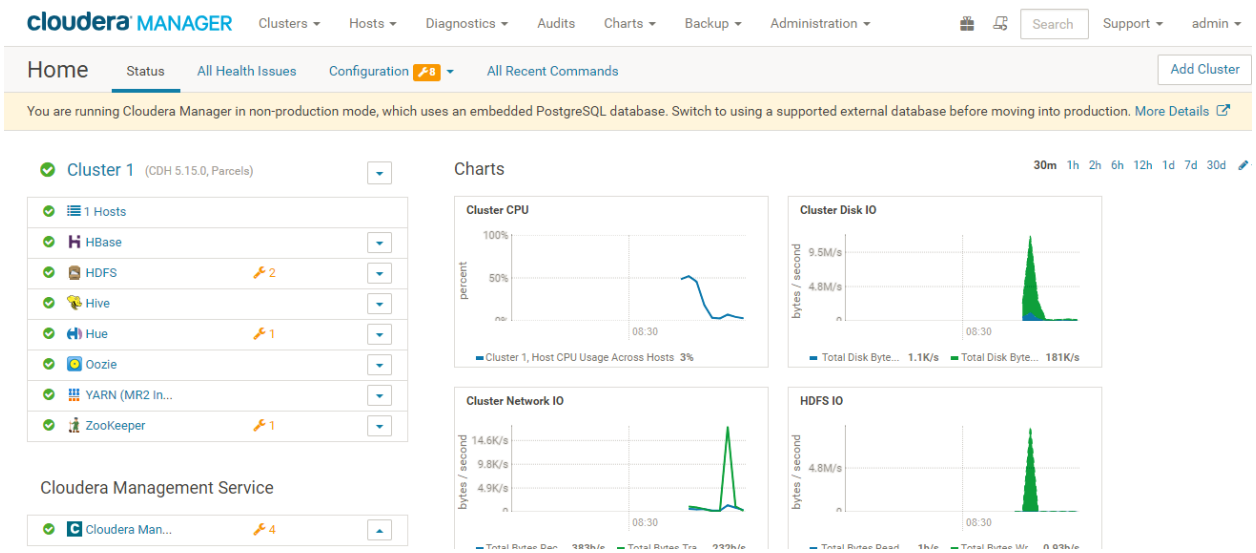


Click on Cloudera Manager.



-

- Now all services are green



51. Set the java path for cloudera-scm-manager or other users:
Again login to ec2 instance using putty and switch to root users;

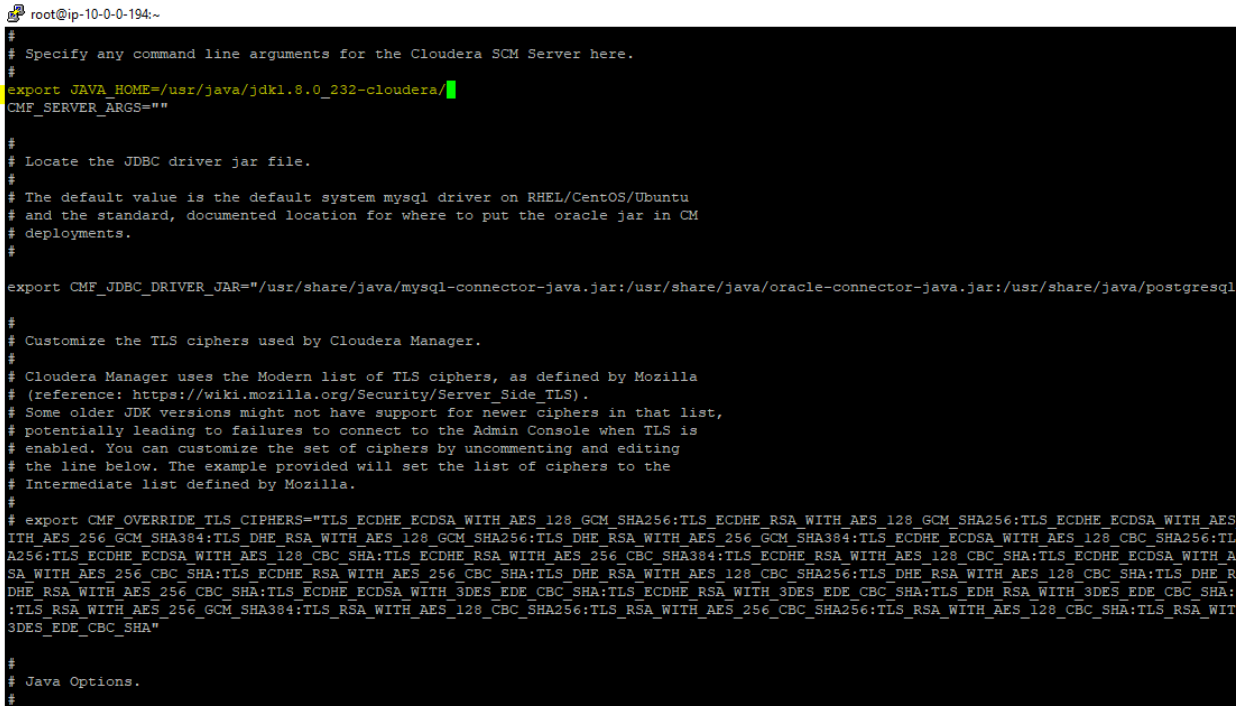
ls /usr/java/jdk1.8.0_232-cloudera/ (enter command without quotes).

```
[ec2-user@ip-172-31-15-234 ~]$ sudo -i
[root@ip-172-31-15-234 ~]# ls /usr/java/jdk1.7.0_67-cloudera/
bin  COPYRIGHT  db  include  jre  lib  LICENSE  man  README.html  release  src.zip  THIRDPARTYLICENSEREADME-JAVAFX.txt  THIRDPARTYLICENSEREADME.txt
[root@ip-172-31-15-234 ~]#
```

vi /etc/default/cloudera-scm-server

Note :- vi takes us to vim text editor. To enter any text we need to enter into insert mode which is achieved by pressing **'i'** and to '**save and quit'** the doc, we need to hit **'esc'** followed by **:wq!** and hit enter.

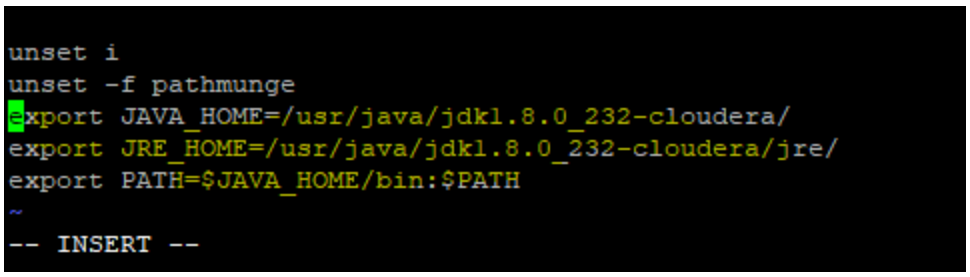
Then, press **' i '** (without quotes) for insert mode, and paste this command as shown:
export JAVA_HOME=/usr/java/jdk1.8.0_232-cloudera/



For others users: open **/etc/profile** and go to the end of the file and insert the variable value.

vi /etc/profile (command)

export JAVA_HOME=/usr/java/jdk1.8.0_232-cloudera/
export JRE_HOME=/usr/java/jdk1.8.0_232-cloudera/jre/
export PATH=$JAVA_HOME/bin:$PATH



Then update the **/etc/profile** using the below commands**:**

source /etc/profile

```
[root@ip-172-31-15-234 ~]# source /etc/profile
[root@ip-172-31-15-234 ~]#
```

Verify the java and JRE version using the below commands as shown in the image below:

java -version

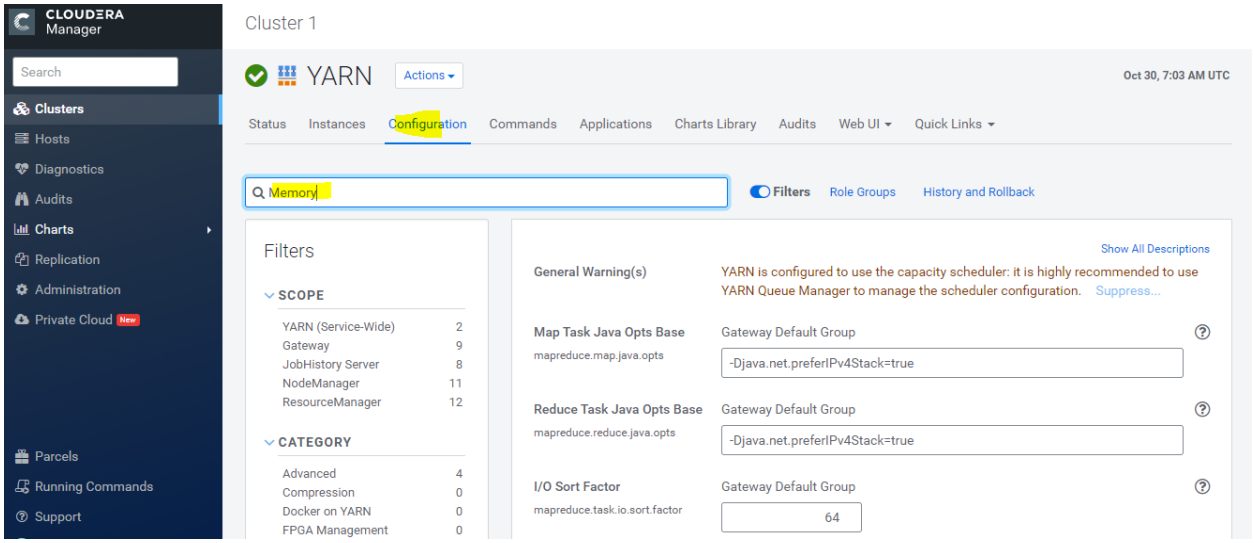
echo $JAVA_HOME

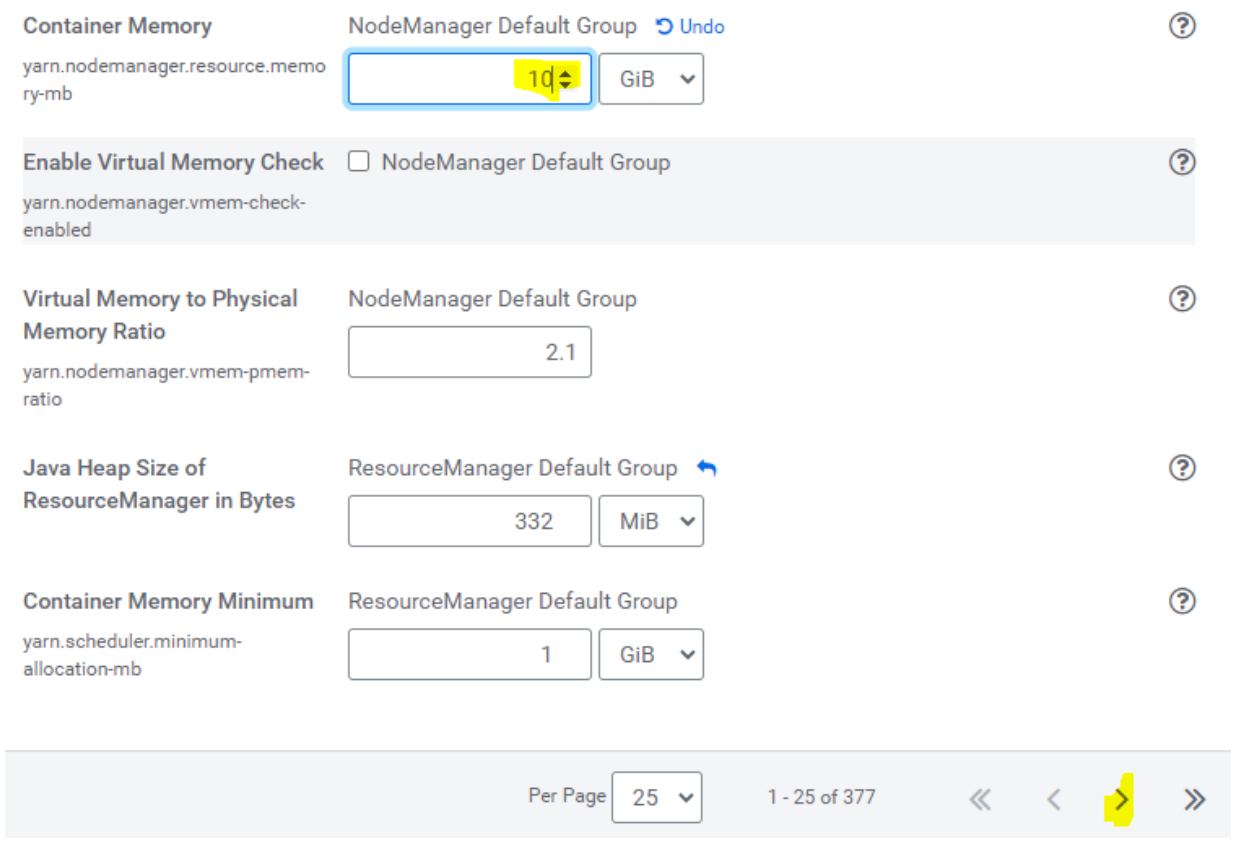
```
[root@ip-10-0-0-194 ~]# java -version
openjdk version "1.8.0_232"
OpenJDK Runtime Environment (AdoptOpenJDK)(build 1.8.0_232-b09)
OpenJDK 64-Bit Server VM (AdoptOpenJDK)(build 25.232-b09, mixed mode)
[root@ip-10-0-0-194 ~]# echo $JAVA_HOME
/usr/java/jdk1.8.0_232-cloudera/
[root@ip-10-0-0-194 ~]# echo $JRE_HOME
/usr/java/jdk1.8.0_232-cloudera/jre/
[root@ip-10-0-0-194 ~]#
```

Configure Yarn tuning:

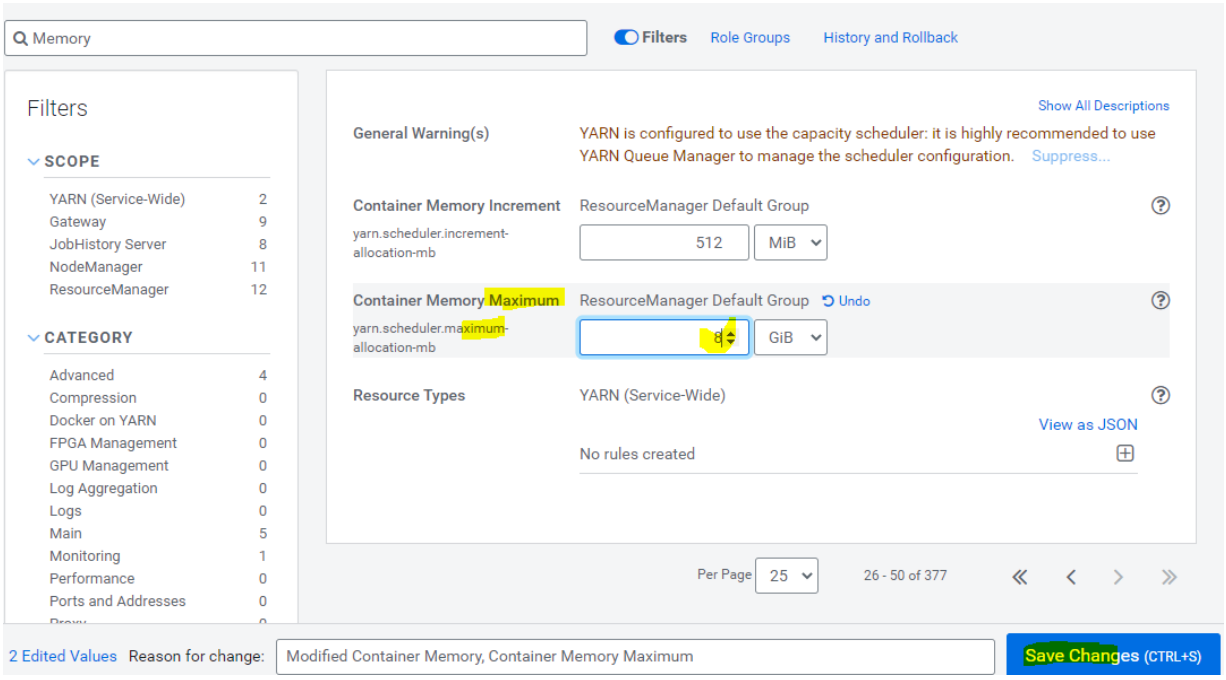
yarn.nodemanager.resource.memory-mb 10GB
yarn.scheduler.maximum-allocation-mb 8GB

Go to yarn services> configuration and search Memory> scroll down





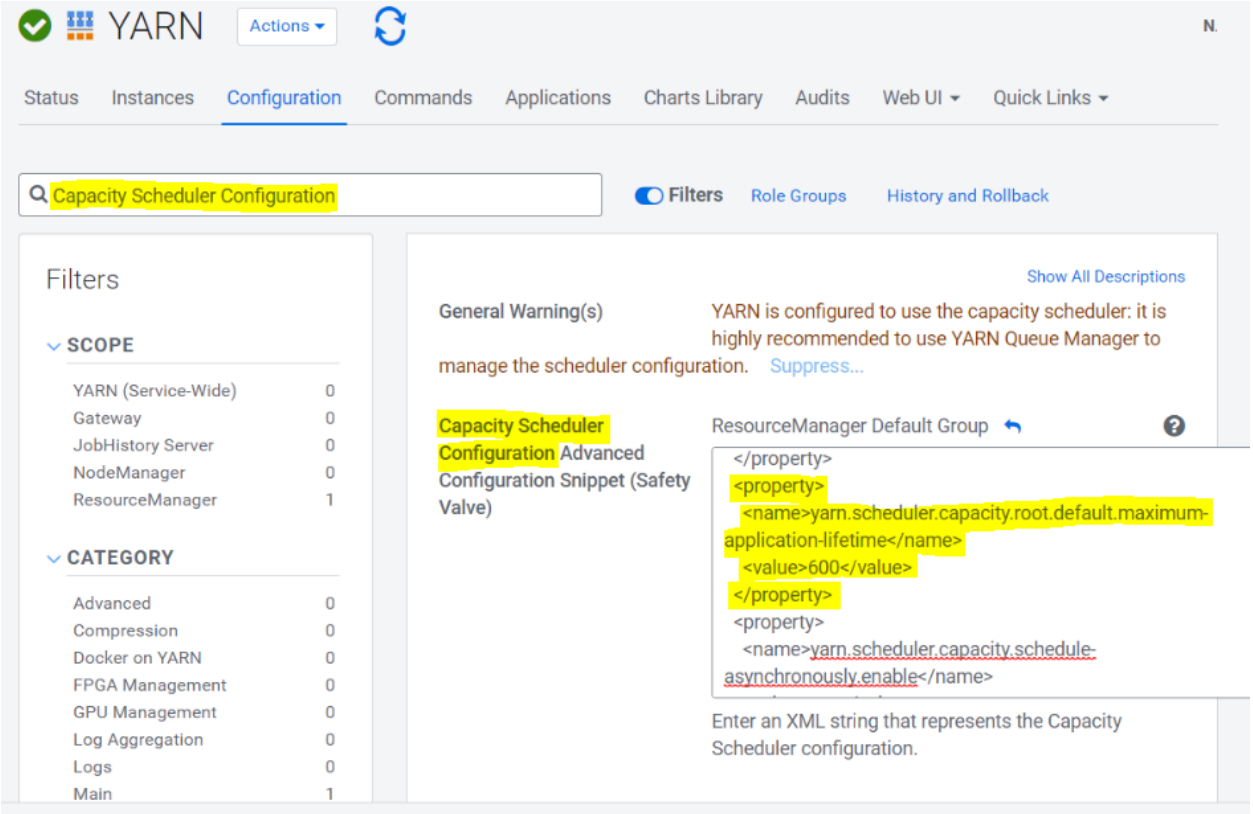
Go to next page



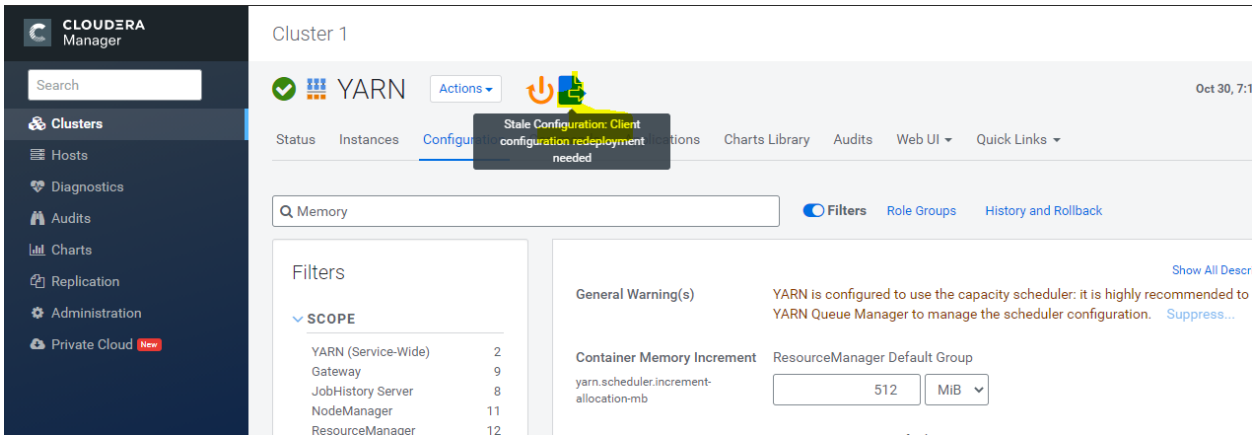
Click on Save Changes.

Next, Go to yarn configuration and search - Capacity Scheduler Configuration and add the following property:

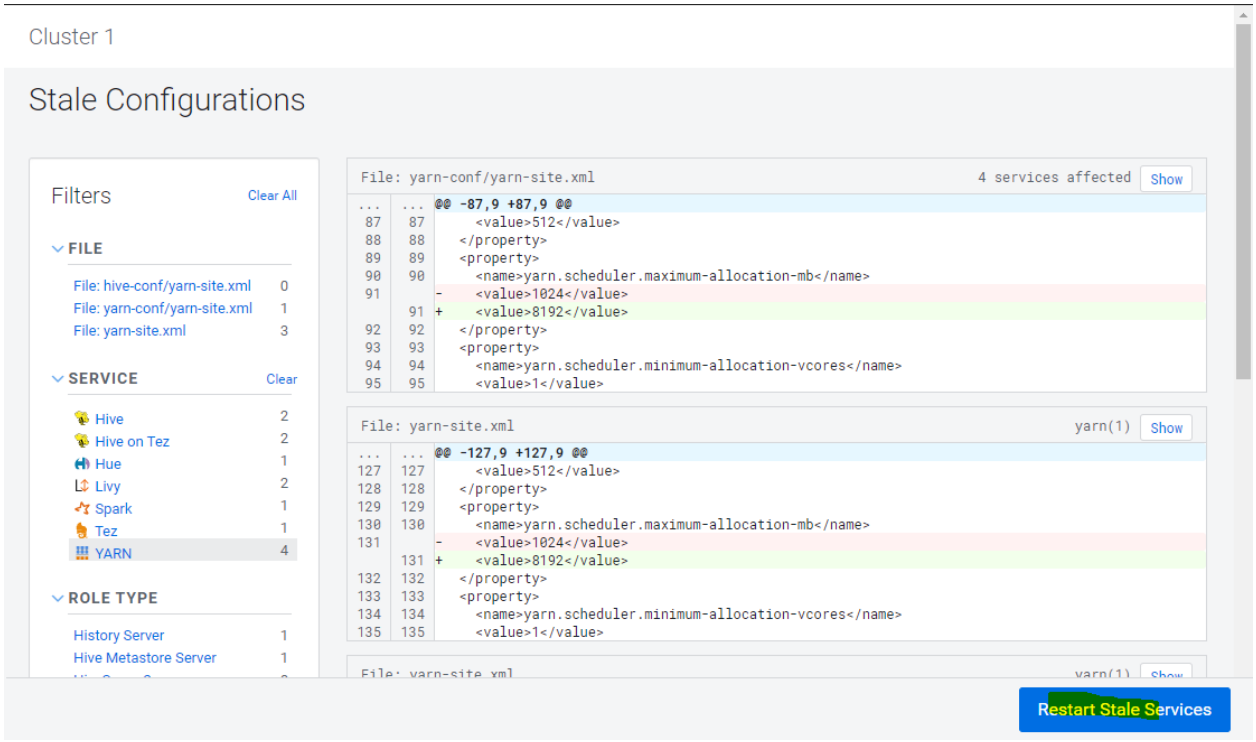
```
<property>
  <name>yarn.scheduler.capacity.root.default.maximum-application-lifetime</name>
  <value>600</value>
</property>
<property>
```

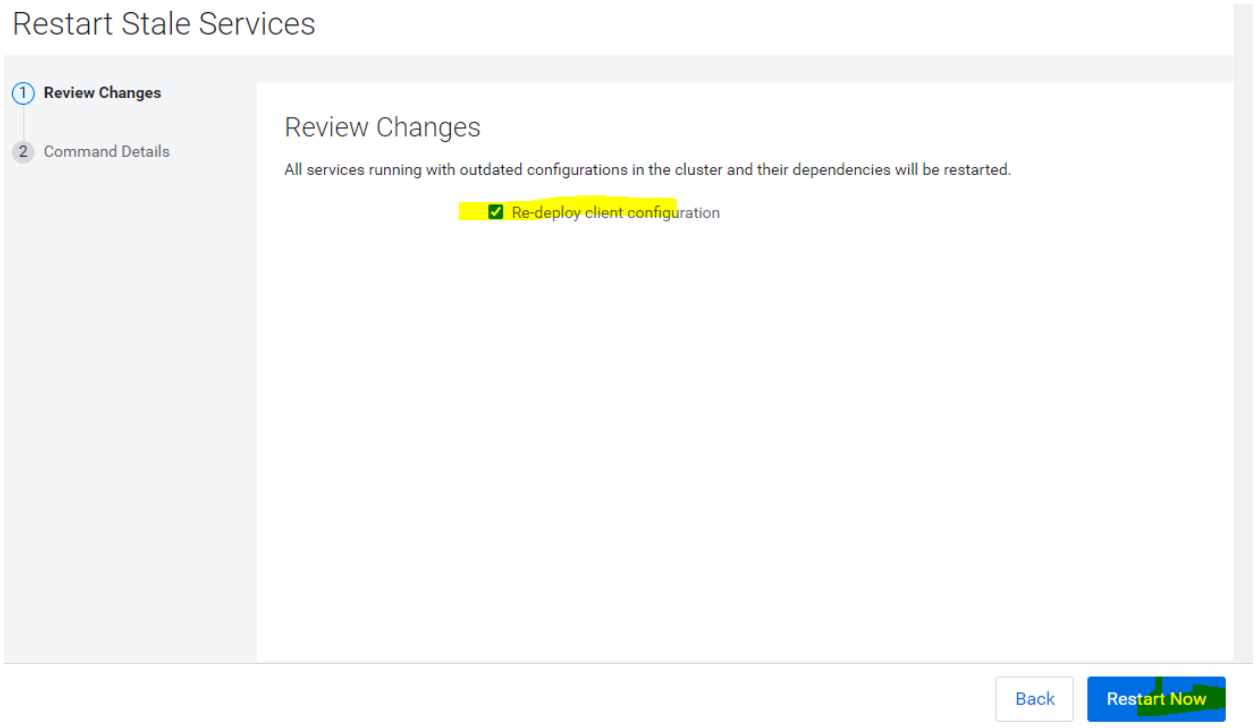


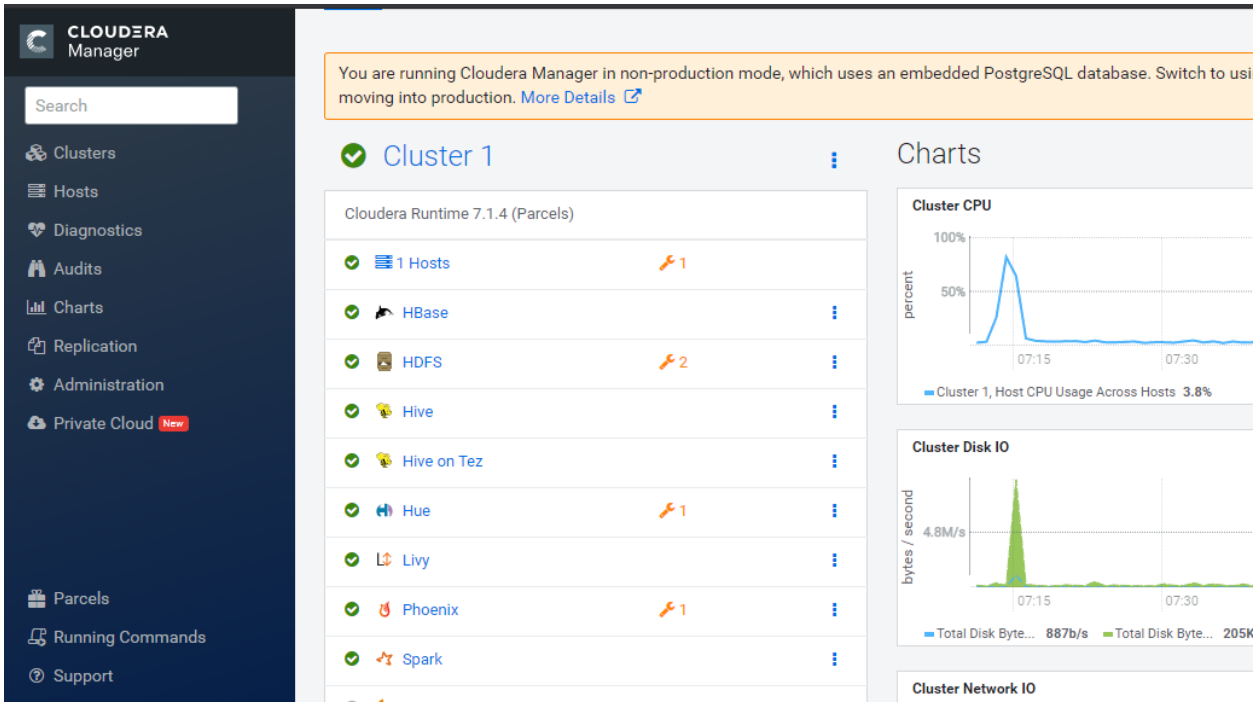
Click on Save Changes. And click on the Stale Configuration button.



Click on Restart Stale Services





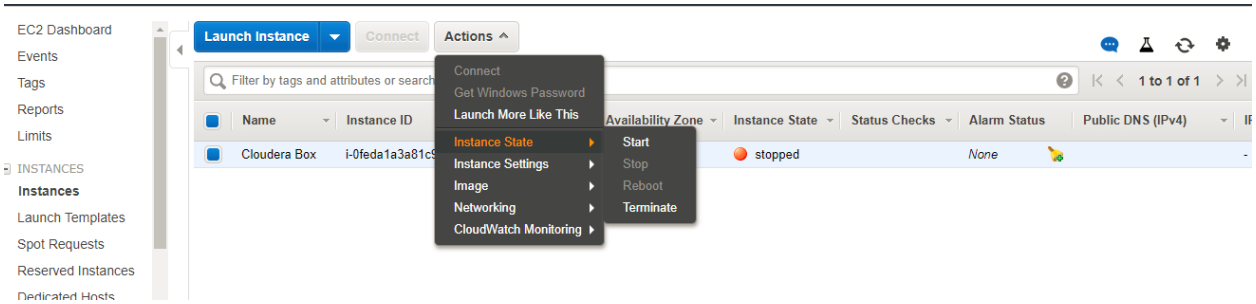


Note: Please stop the instance if you are not using the ec2 dashboard.

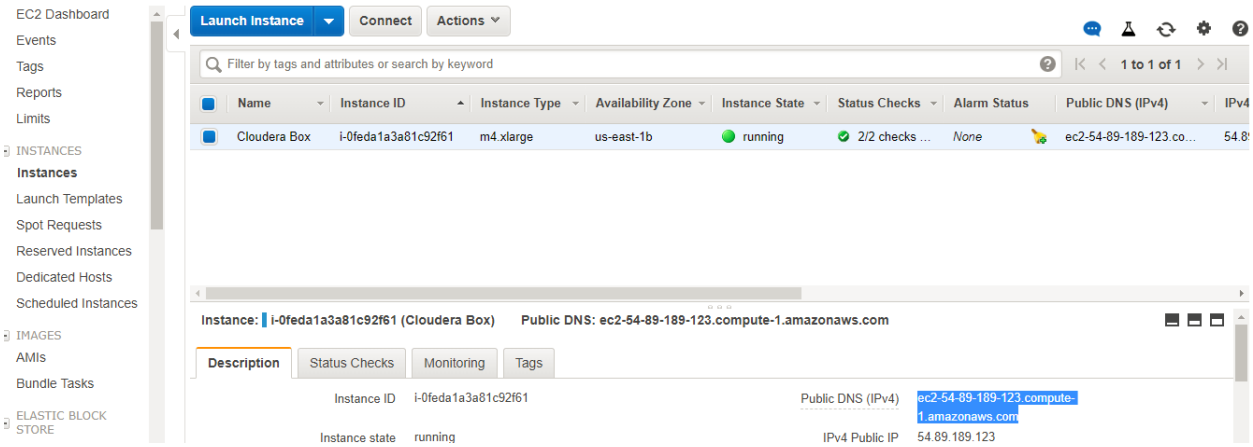
If your instance is running then the cost is Rs.19/hours

After I stopped my instance and restarted it, my Cloudera management services and other services stopped into my browser. The following screen appeared.

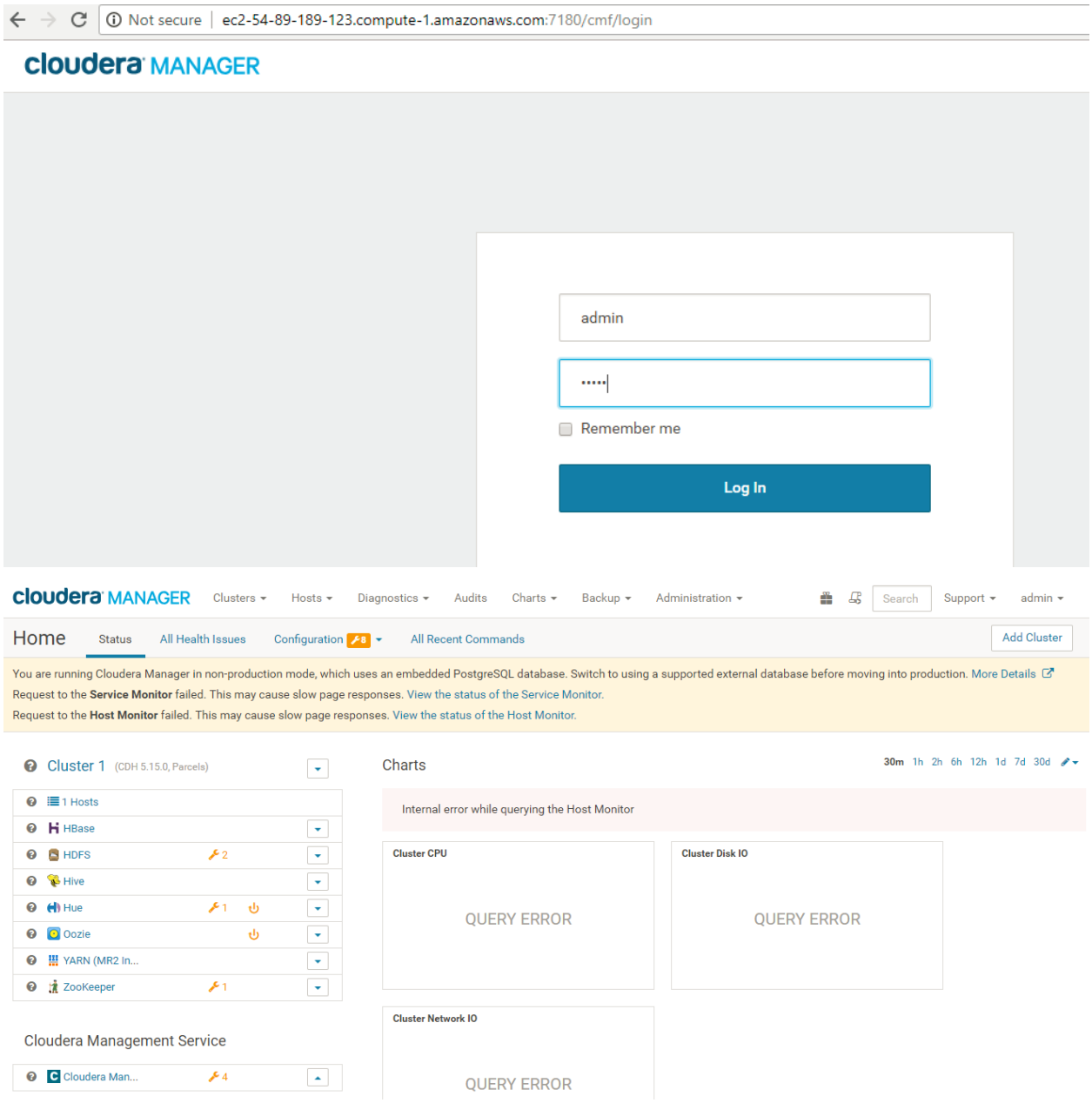
A. Start the instance from the EC2 dashboard and wait until the status check is 2/2.



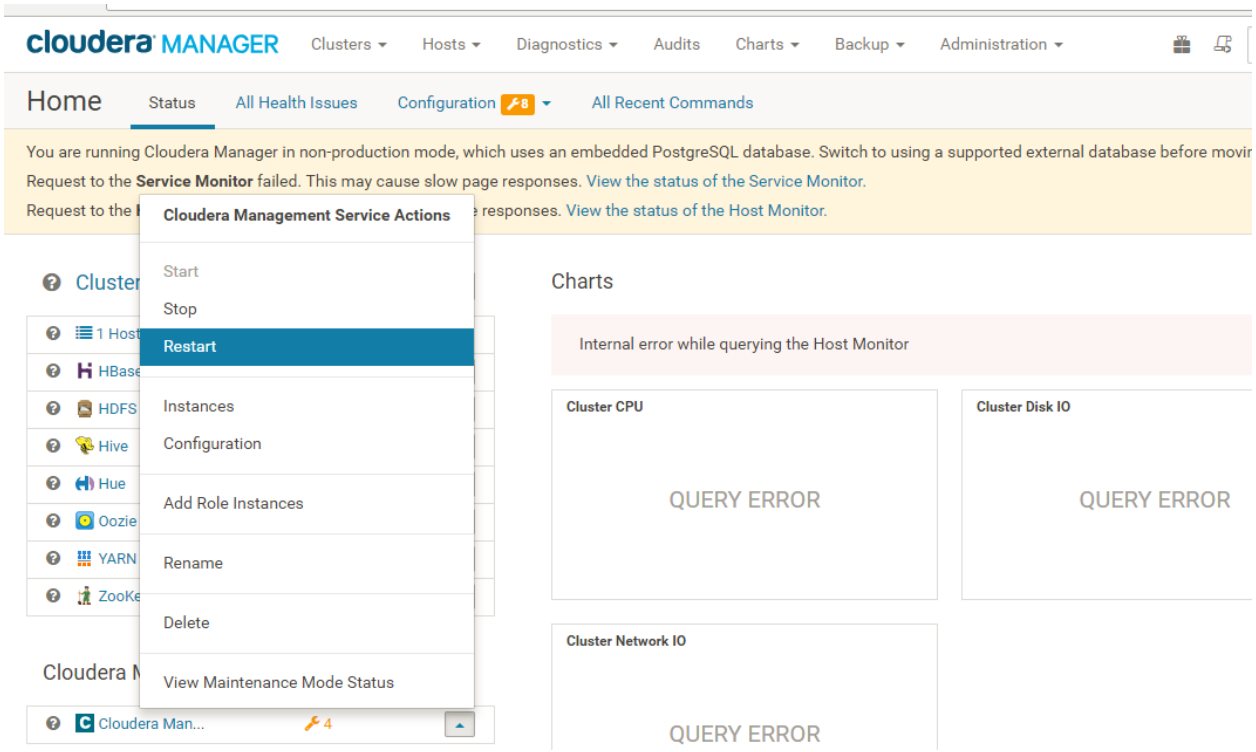
B. Copy the public IP and paste into the browser



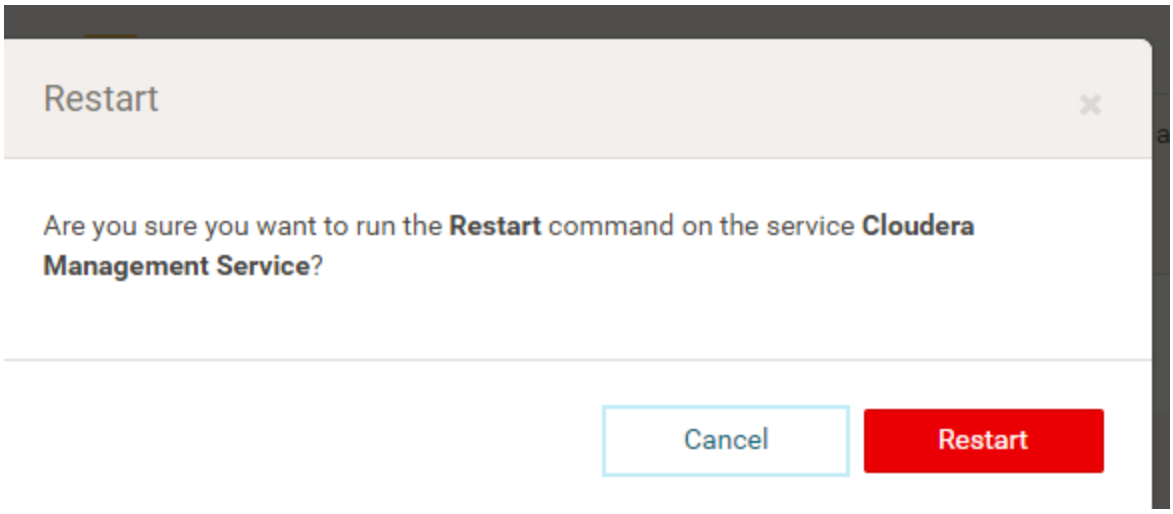
http://publicip:7180



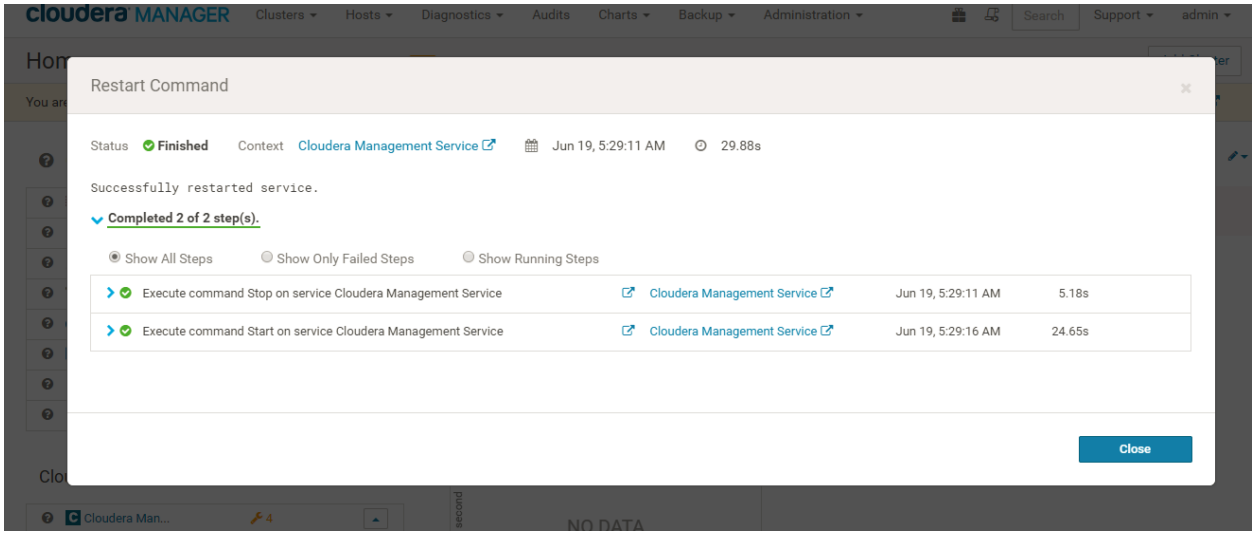
To solve this problem, go to the page above, scroll down, and click on **'Cloudera Management Service';** then, click on **'Restart'**



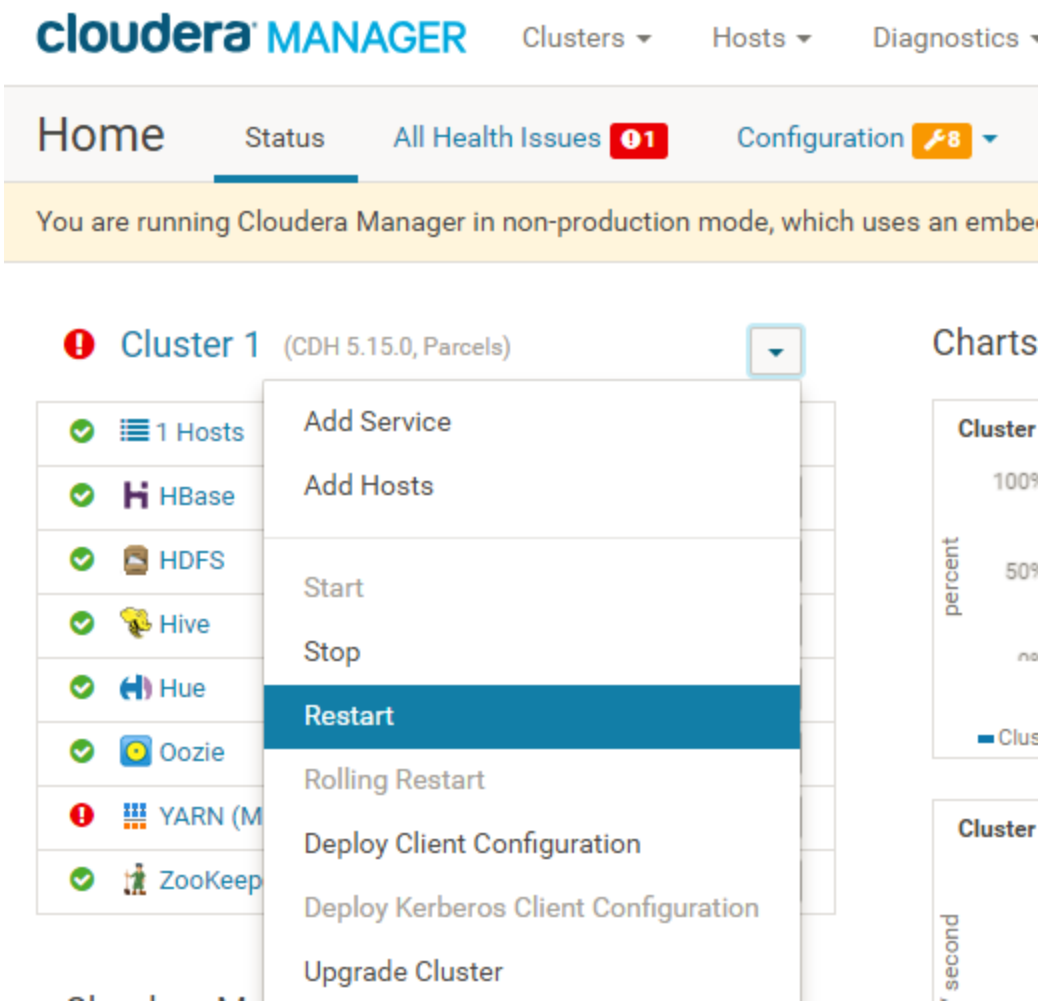
Click on **'Restart':**



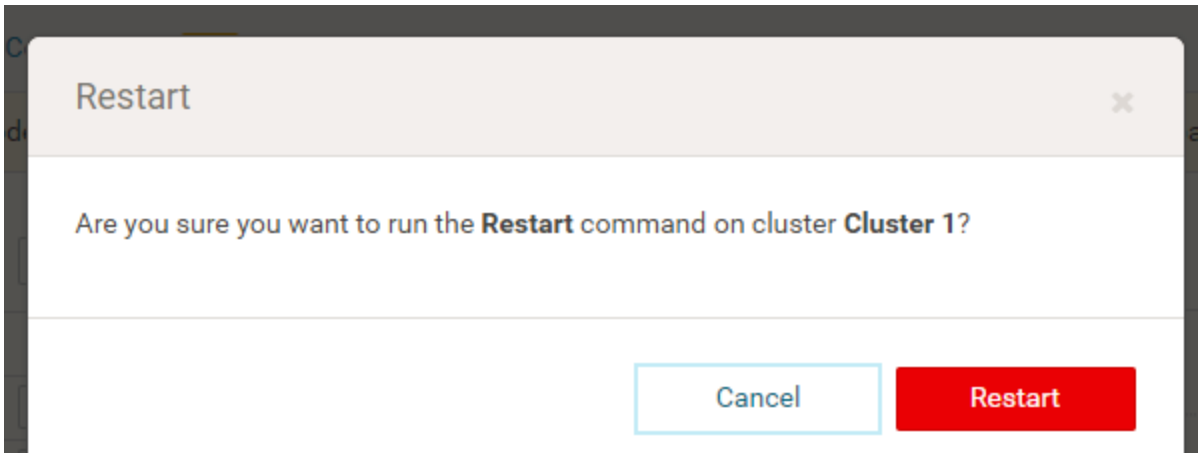
Click on **‘Close’:**



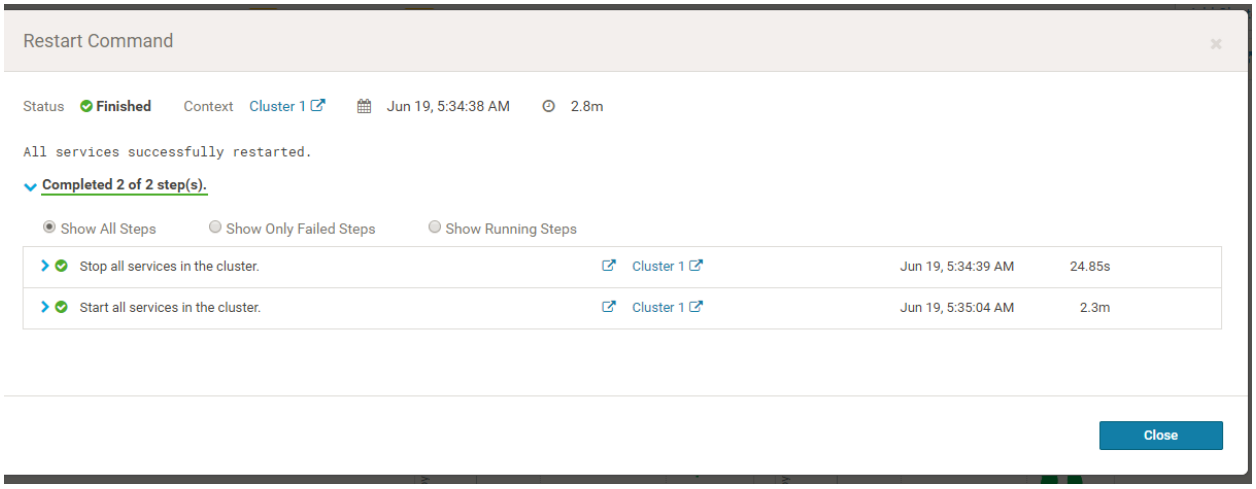
Now **Restart** the cluster1 services:



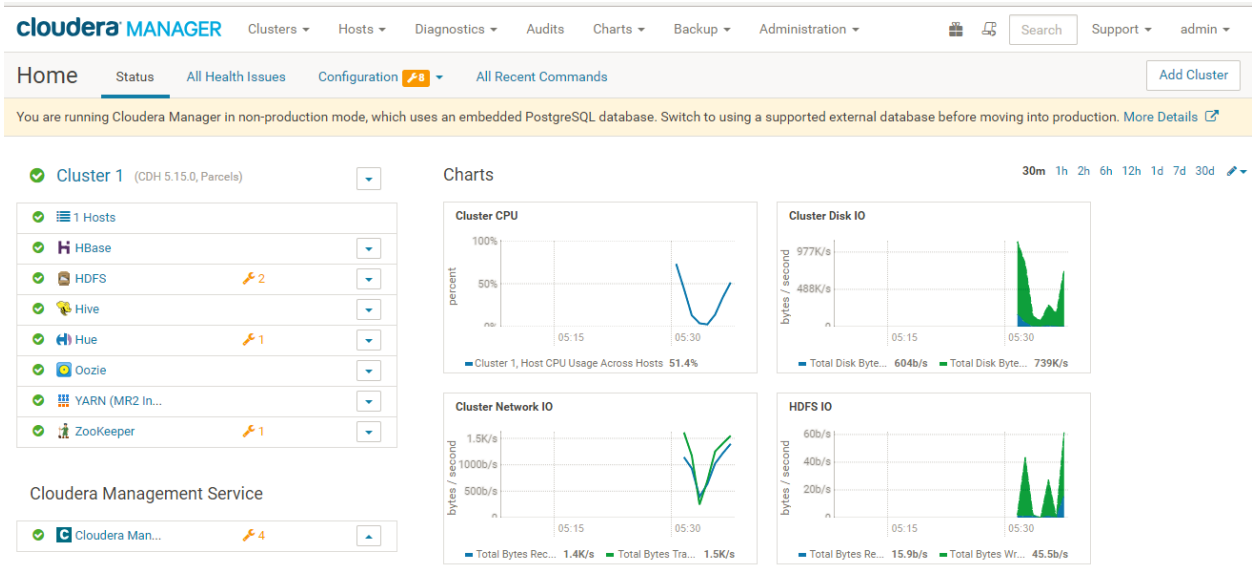
Click on **Restart:**



Click on **'Close'.**



All services in green.



Note: Once you have finished the module, terminate this instance. Even if the instance is stopped some charges are still applicable.