Big Data Analytics | Lab 8

**Setup Cluster On Oracle Cloud :- Using Big Data Services**

Raj Panchal  
17CEUBG104

**Setup Oracle Cloud Account :**

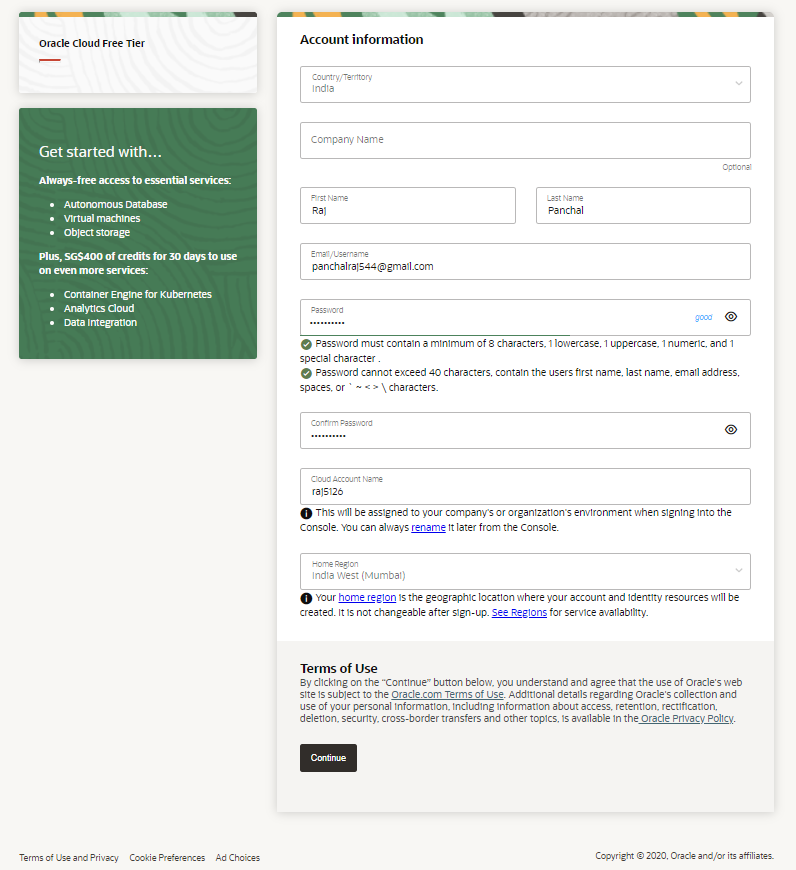
Go to the below link :

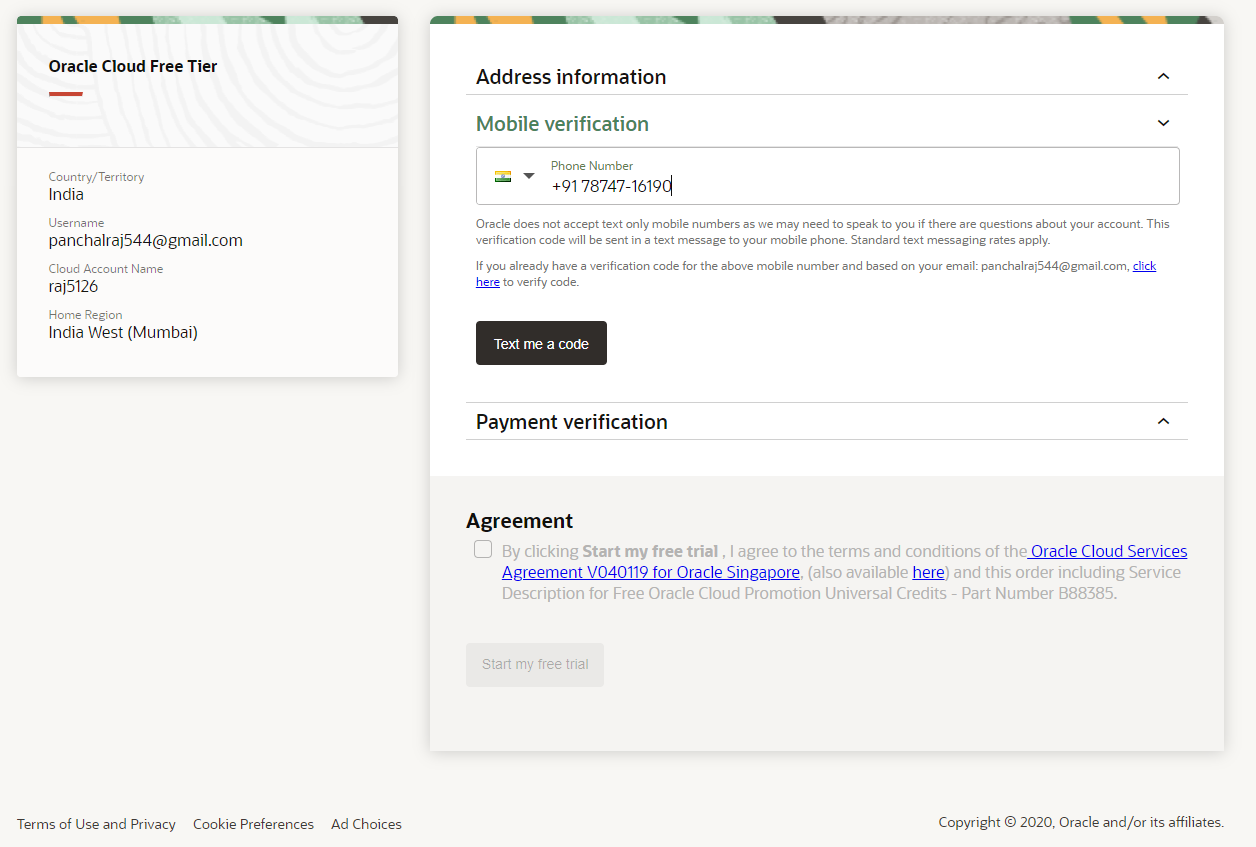
<https://www.oracle.com/cloud/>

Then Click on **Try OCI for Free**.(**OCI** stands for Oracle Cloud Infrastructure.)

Know that oracle offers a Free Tier with no time limits on selection of services like Autonomous Database, Compute, and Storage, as well as **US$300** in free credits to try additional cloud services.

You just need to enter your details and can use it in a few seconds.

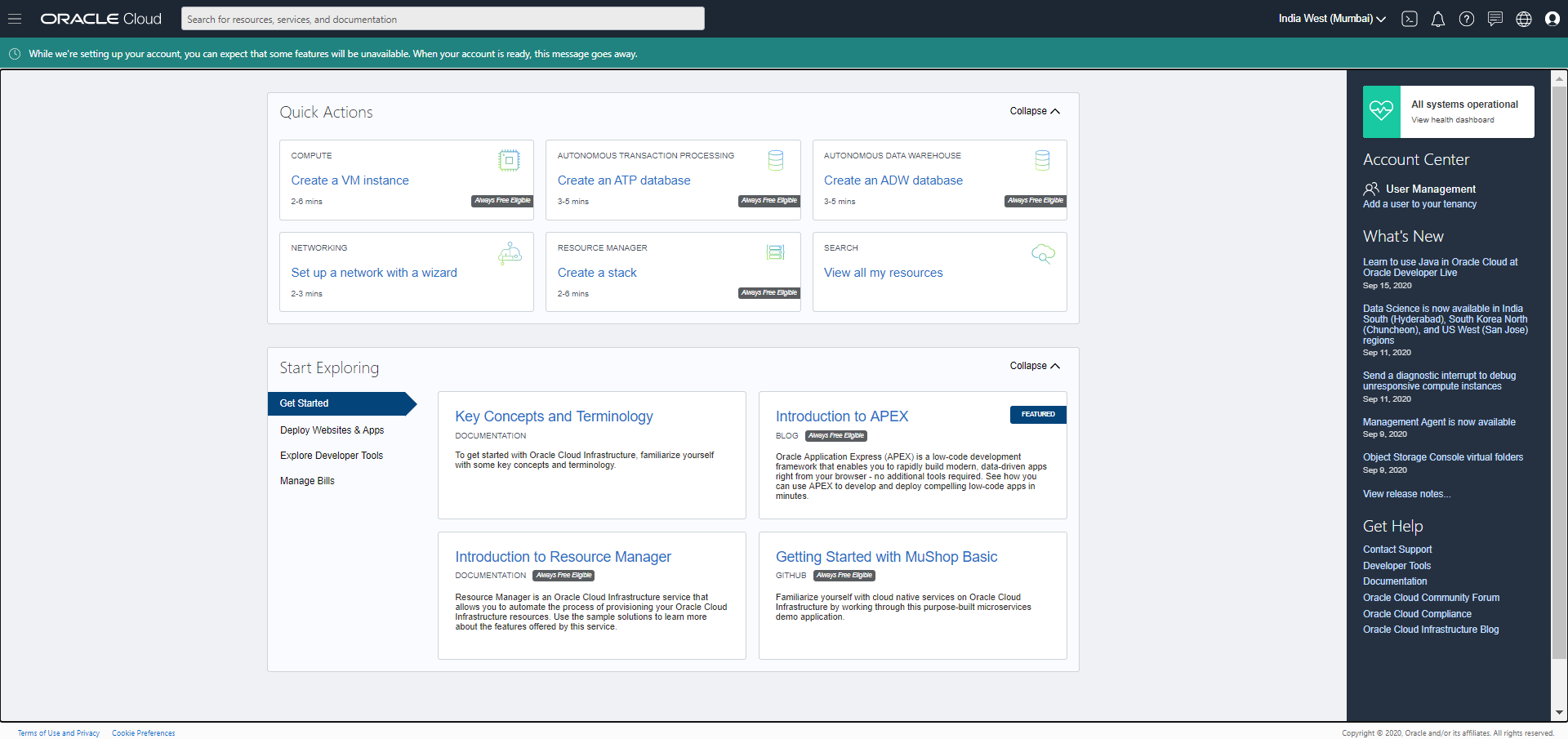




Note : When you first sign up , Oracle will ask you for credit card details but there won't be any charge in free trial. When your trial is over, your account will be limited to Always Free resources until you upgrade your account.

After finishing the process you would be able to OCI screen.Verify that your region is **India West (Mumbai)**.It takes some time setting up your account and available for all features.

Know that tenancy is your cloud name which you provided at time of sign up.(here : **panchalraj544**)

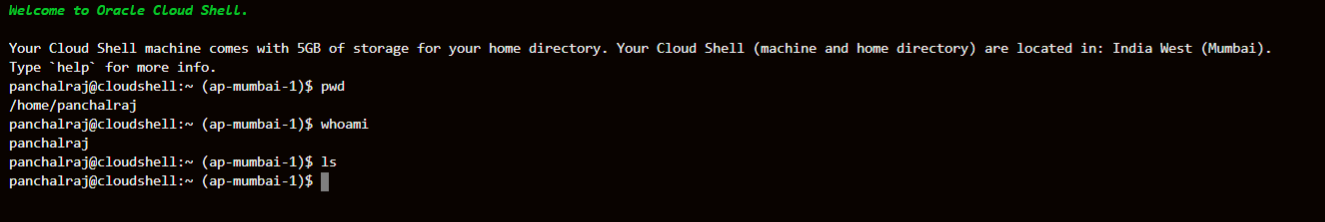


**Interact with Oracle Cloud Shell :**

Click on  from right top to get connected with the cloud shell.

Cloud shell comes with 5 GB of storage for your home directory where you can save your work.

You can also run basic linux commands in Cloud shell.



Click on  explore different menus.

**Create Compartment :**

Compartment will organize the resources used by Big Data Service.

Select from **Identity > Compartments**.

Click **Create Compartment**.

Provide a **NAME** and **DESCRIPTION**, and then click **Create Compartment**.



The created compartment is **Big\_Data\_Analytics\_LAB\_8**.

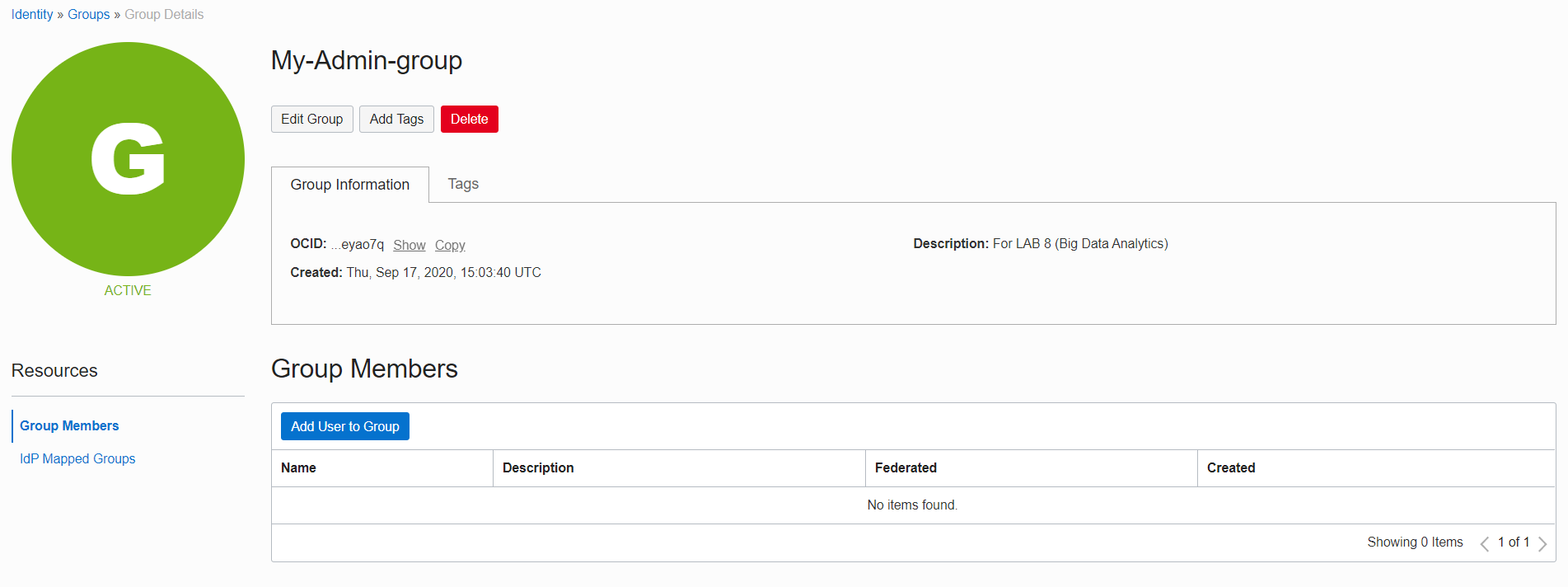
**Create a BDS Group and Add an Admin User:**

You will grant privileges to this group - allowing the group to perform the critical administrative tasks required to manage your cluster lifecycle.

Select **Identity > Groups**.

Click **Create Group.**

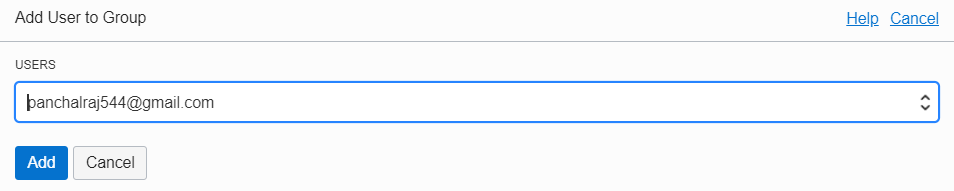
Provide a **NAME** and **DESCRIPTION**, and then click **Create.**



The created Group name is **My-Admin-group.**

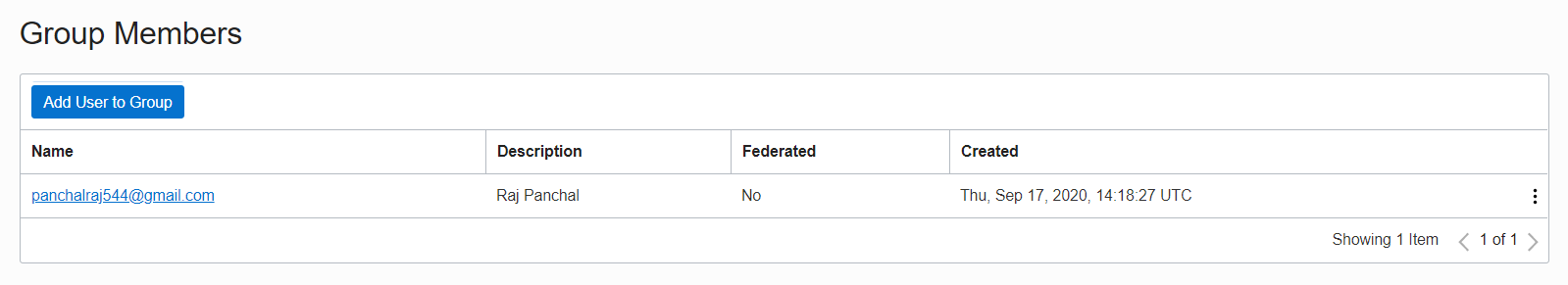
Then click **Add User** **to Group**.

Select **root-user** from list of users and click **Add**.



Here root-user is **panchalraj544@gmail.com**.

You may confirm User within **Group members** list.



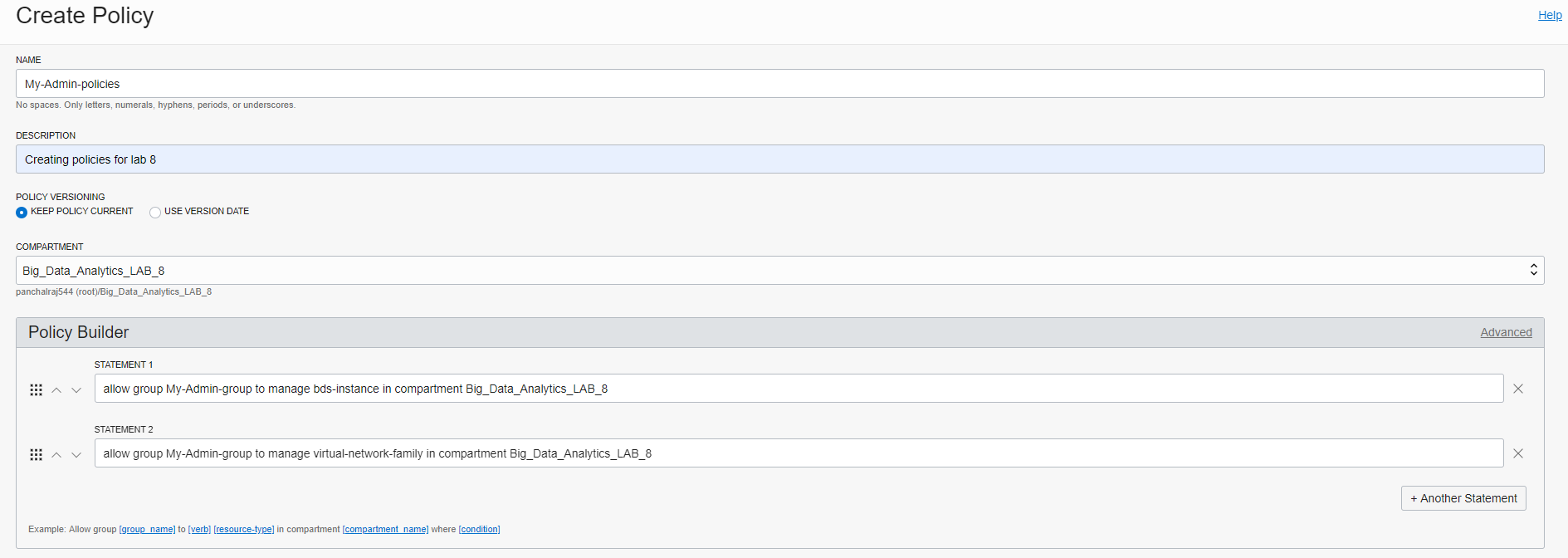
**Create Policies Required to Manage Big Data Service:**

Select **Identity >> Policies**.

Ensure that you are in the previously created compartment.

Click **Create Policy**.

Provide the details as shown below.



The created policy name is **My-Admin-policies**.

Create the following policy in your **root compartment** to allow cluster creation named **My-BDS-policies.**

allow service bdsprod to {VNIC\_READ, VNIC\_ATTACH, VNIC\_DETACH, VNIC\_CREATE, VNIC\_DELETE,VNIC\_ATTACHMENT\_READ, SUBNET\_READ, VCN\_READ, SUBNET\_ATTACH, SUBNET\_DETACH, INSTANCE\_ATTACH\_SECONDARY\_VNIC, INSTANCE\_DETACH\_SECONDARY\_VNIC} in compartment Big\_Data\_Analytics\_LAB\_8

**Create a Virtual Cloud Network:**

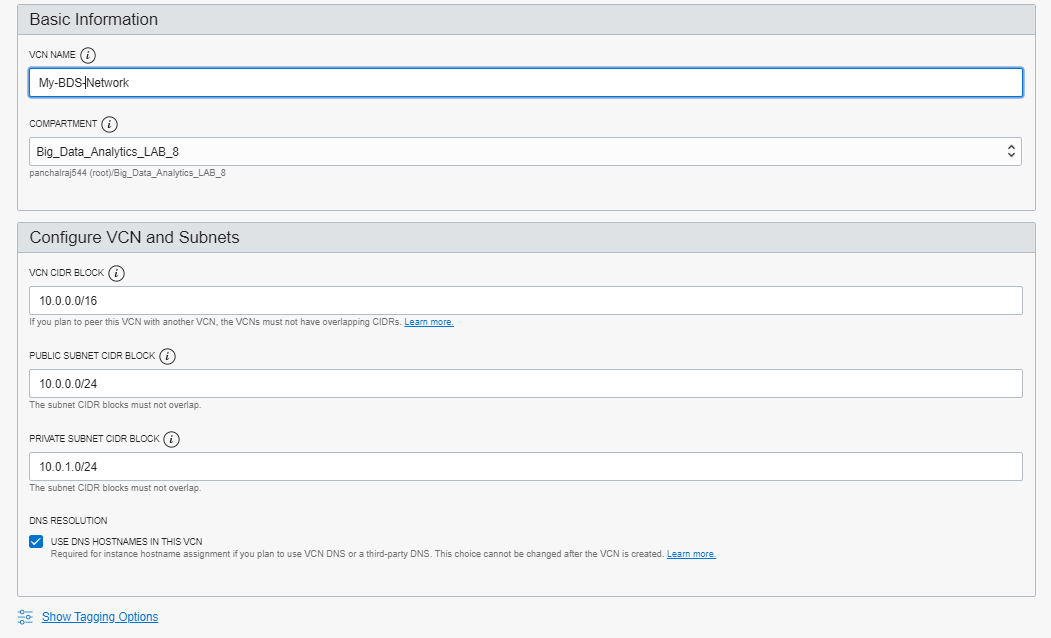
Set up the Virtual Cloud Network that will be used by our Big Data Service.

Select **Networking > Virtual Cloud Networks**.

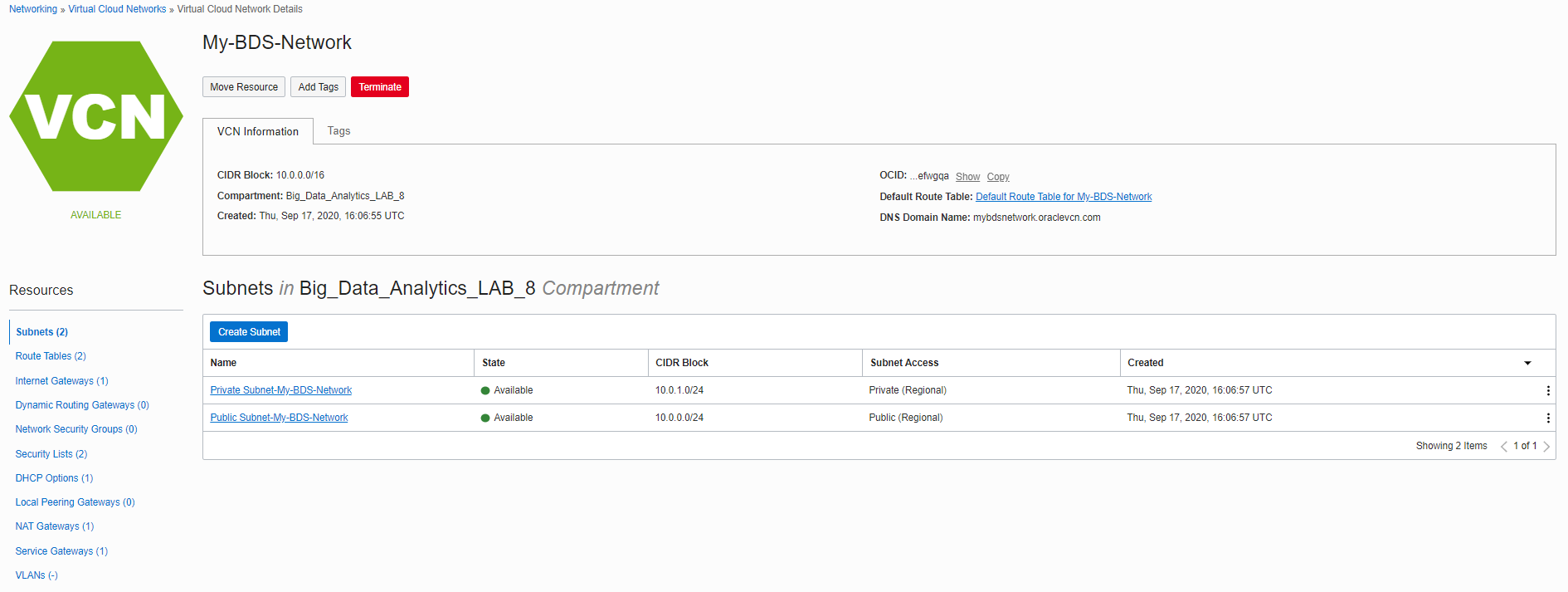
Ensure that you are in the previously created compartment.

Click **Start VCN Wizard**.

Then Select **VCN with Internet Connectivity** and click **Start VCN Wizard**, Add Basic Information as below.



Finally you are able to see your VCN as below.



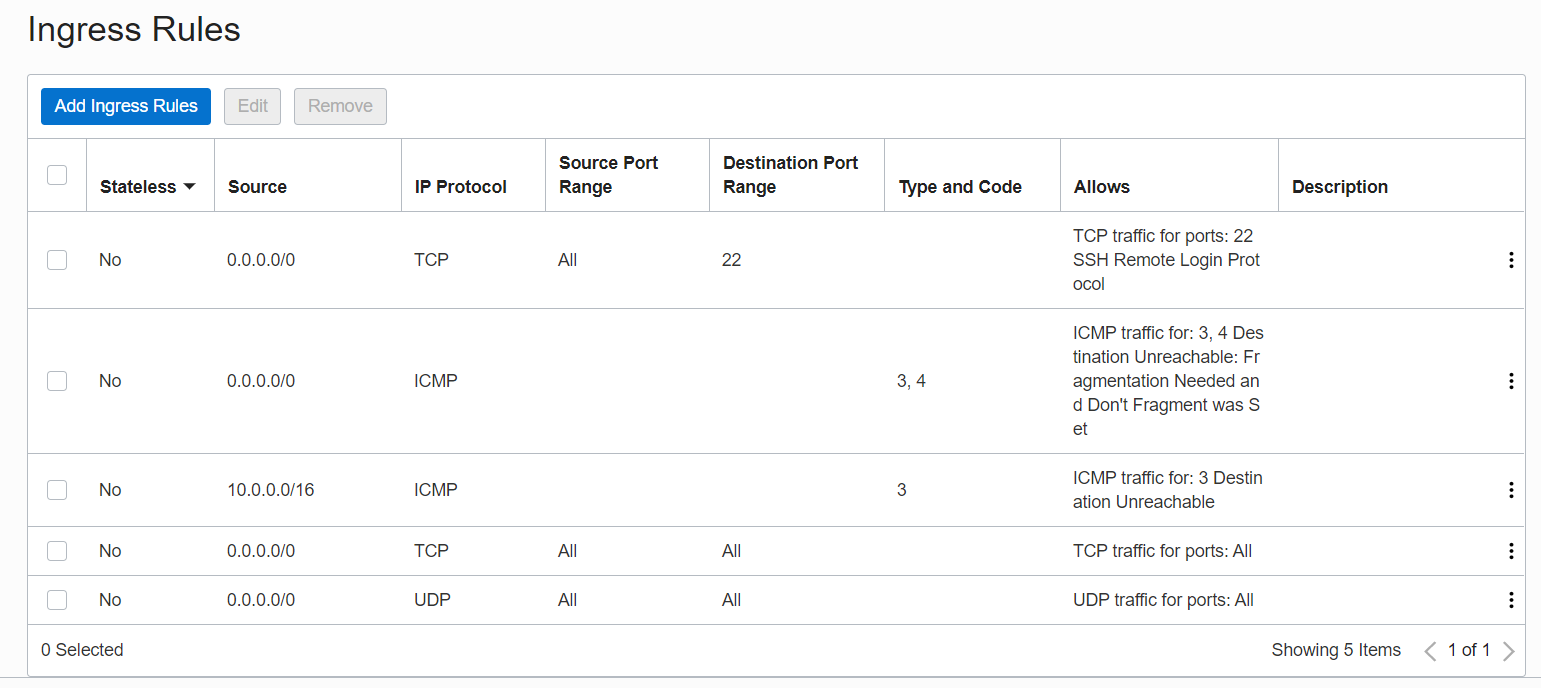
Now we will update security lists next to allow ingress from any host in the VCN.

The IP addresses will not be accessible over the internet until you explicitly create a public IP address.

Click **Security List > Default Security List for My-BDS-Network > Add Ingress Rules**.

* Update Source CDR : 0.0.0.0/0
* IP Protocol : TCP
* Destination Port Range : All
* Click **Add Ingress Rules.**

Repeat the same step for **UDP** Protocol.



**Create a Big Data Service Cluster :**

This task is to step through the process of creating a secure, highly available Cloudera Hadoop cluster.It is designed to use all features - but is targeted at small workloads and flexibility.

Goto **Data and AI > Big Data**.

Ensure that you are in the previously created compartment.

Click **Create Cluster** and specify cluster properties.

* Cluster Name : my-bds-cluster
* Cluster Admin Password : <password> (this will be the administrator password for Cloudera Manager)
* Uncheck **SECURE & HIGHLY AVAILABLE(HA).**
* Cluster Version : CDH 6.3-ol7

For Hadoop Nodes :

* Choose Instance Type : Virtual Machine
* Choose Master/Utility Node Shape : VM.Standard2.4
* Block Storage Size per Master/Utility Node (IN GB) : 250
* Choose Worker Node Shape : VM.Standard2.1
* Block Storage Size per Worker Node (IN GB) : 150
* Number of Worker Nodes : 3

For Network Settings :

* CIDR Block : 10.0.0.0/16
* Choose VCN : My-BDS-Network
* Choose Regional Subnet : Public Subnet-My-BDS-Network
* Check **DEPLOY ORACLE-MANAGED SERVICE GATEWAY AND NAT GATEWAY (QUICK START)**

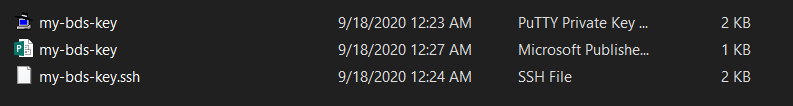
For Additional Options

* Upload your **public key**. The associated private key will be used to make SSH connection to the cluster.

Refer below link for generating public-private key pair in windows 10.

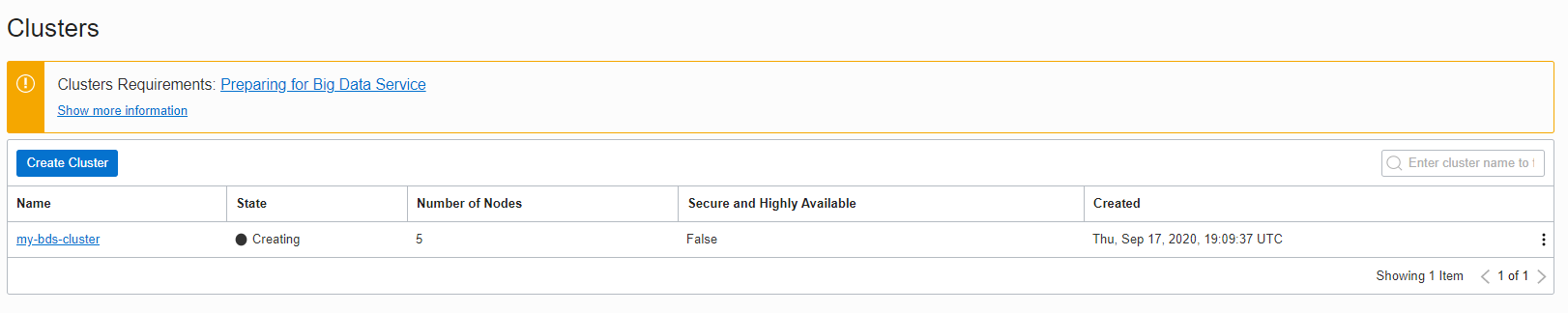
<https://docs.oracle.com/en/cloud/paas/event-hub-cloud/admin-guide/generate-ssh-key-pair-using-puttygen.html>

Here my public key is my-bds-key.pub.



Finally click **Create Cluster**.

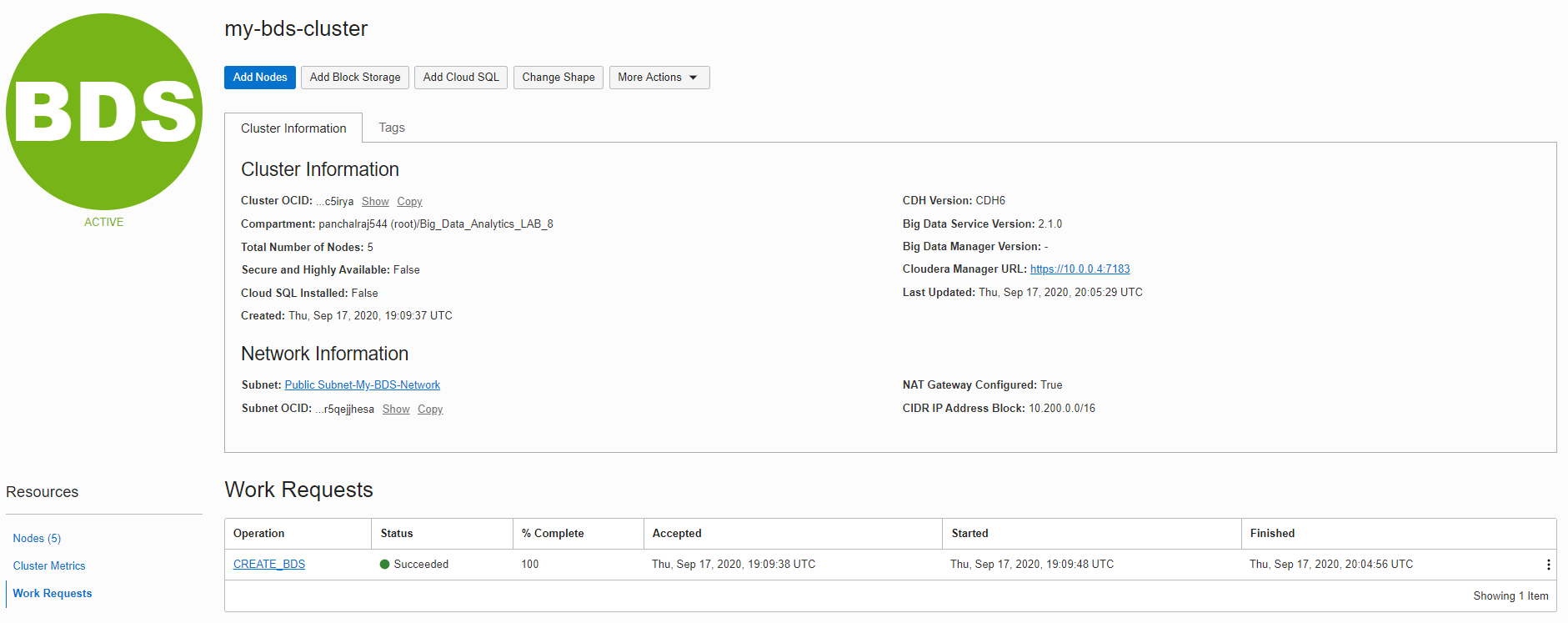
Then you would see the same as above.



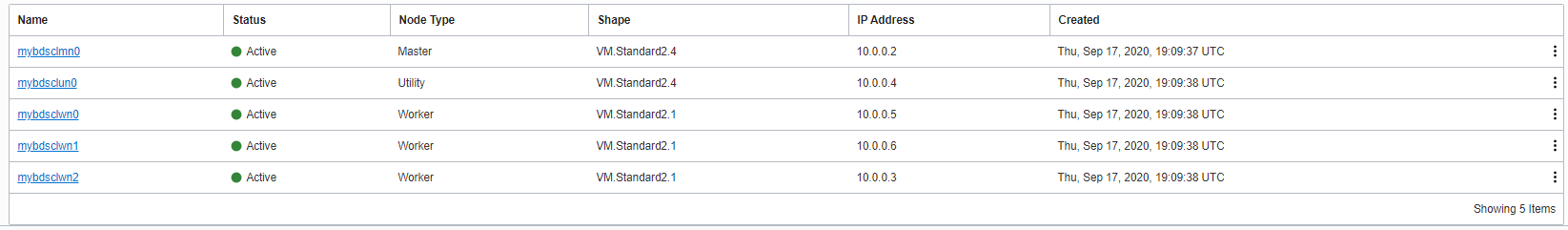
It takes around **50 minutes.**You can monitor its progress at any time.

Select **my-bds-cluster > Work Request**.

After completion of the process you will get like below.



Observe the List of Cluster Nodes:



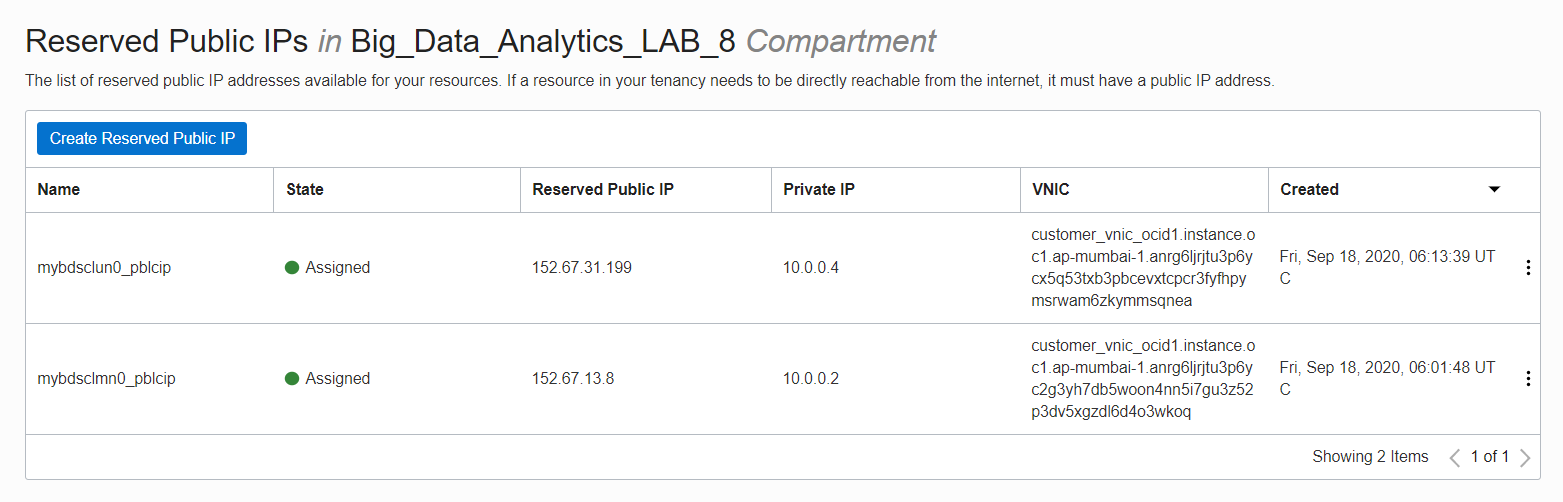
**Map a Private IP Address to a Public IP Address:**

Big Data Service nodes are by default assigned private IP addresses, which aren't accessible from the public internet. One way to make a node accessible from the internet is to map a node's private IP address to a public IP address.

I have followed this link for this task.It is just a 4 Step process and easy to understand.

<https://docs.oracle.com/en/cloud/paas/big-data-service/user/map-private-public-ip-adresses.html>

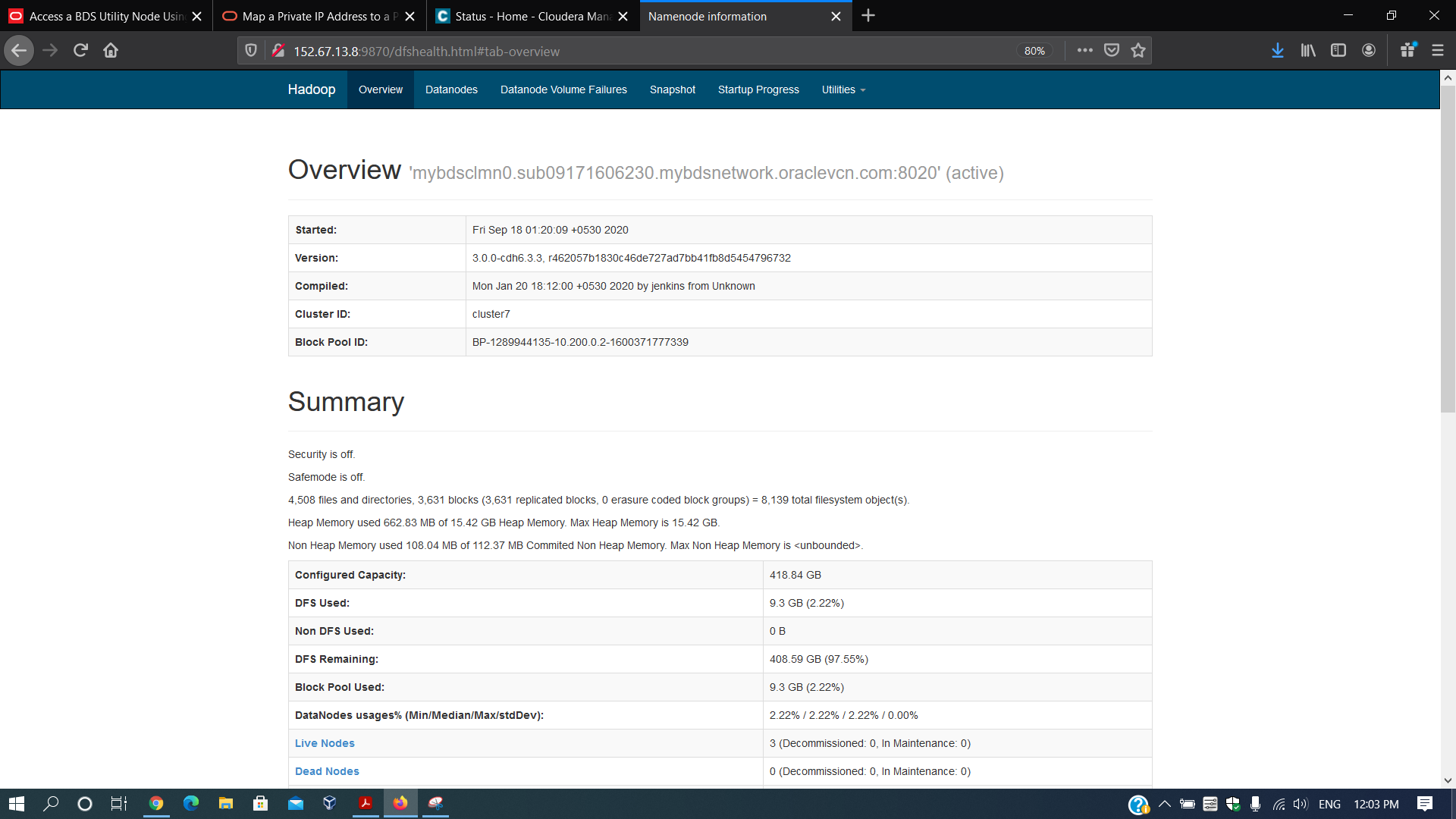
And configure namenode and utility node to public address.

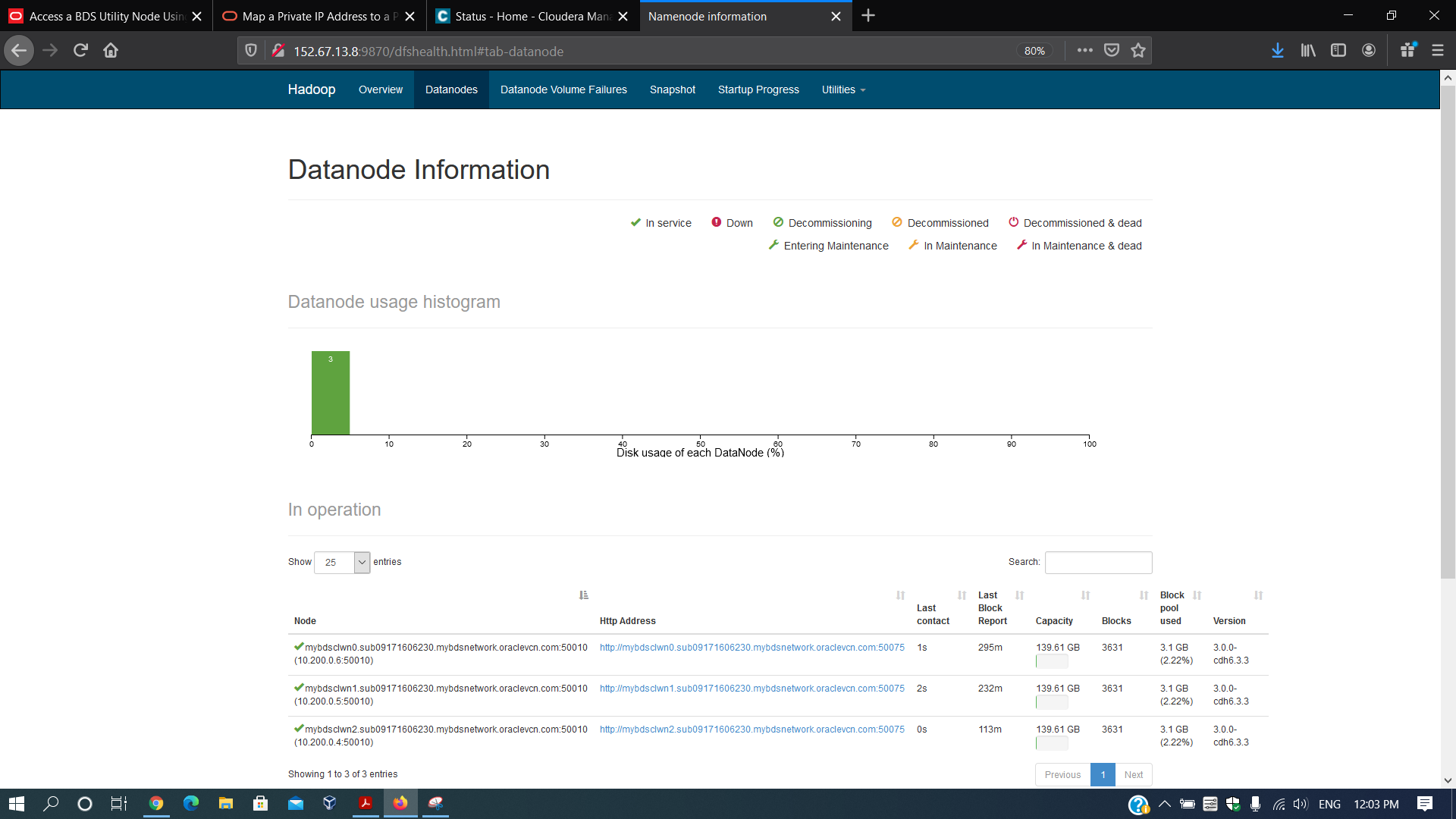


Now I can access HDFS GUI via namenode and Cloudera manager via utility node.

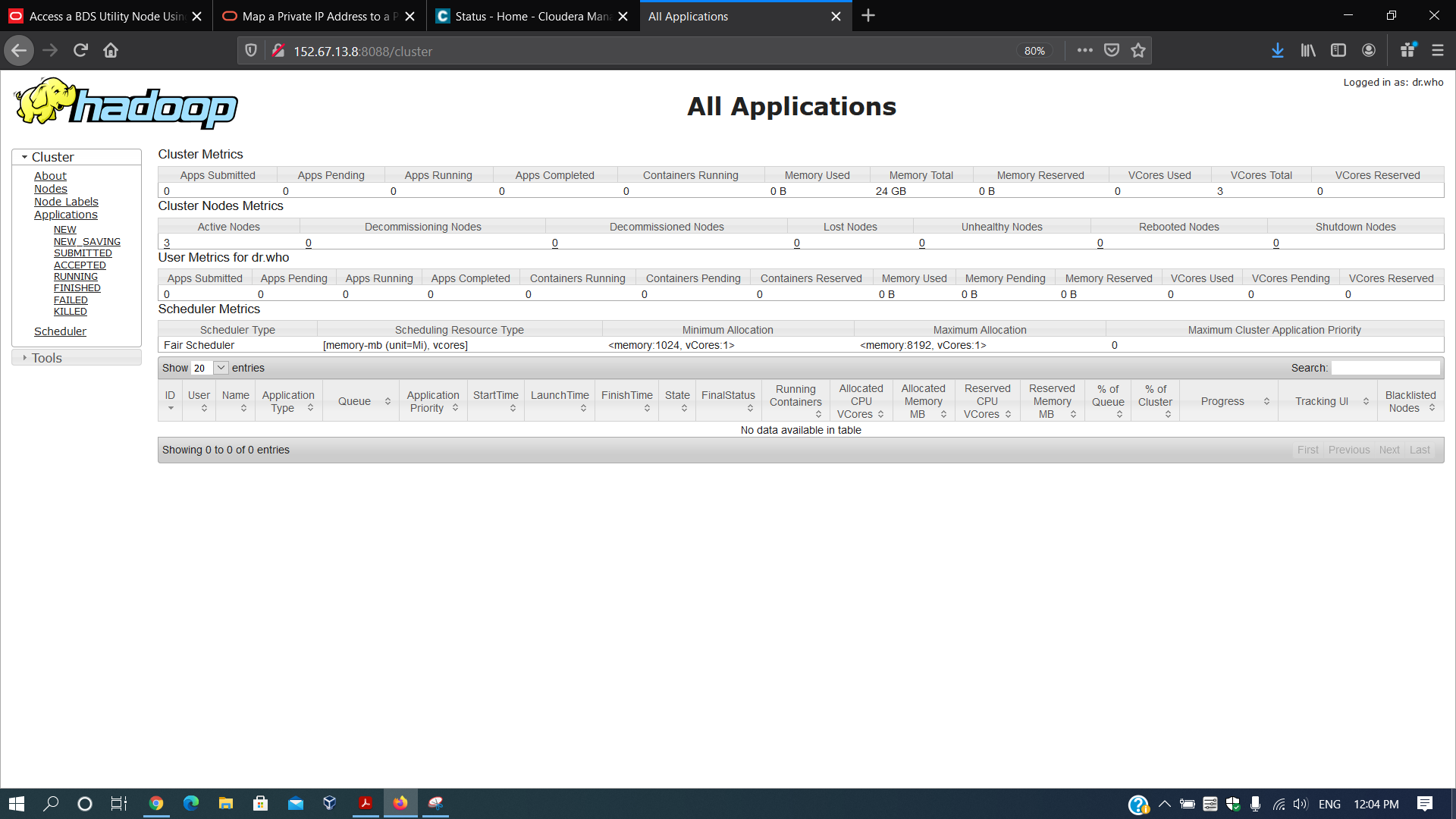
**Verify Cluster’s presence of GUI:**

NameNode Information at : <http://152.67.13.8:9870/dfshealth.html#tab-overview>





Cluster information at : <http://152.67.13.8:8088/cluster>

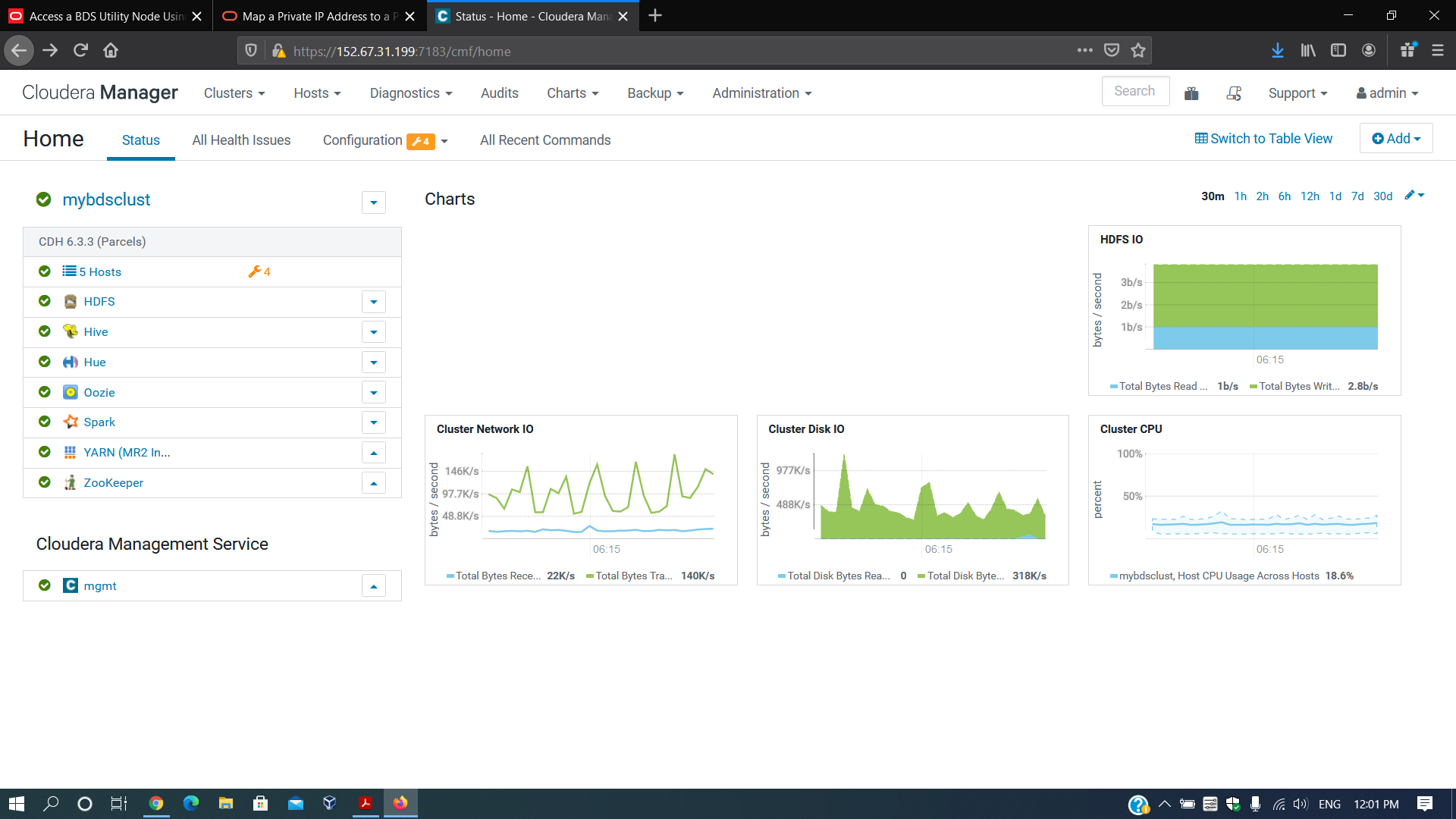


Cloudera Manager at : <https://152.67.31.199:7183/cmf/home>

It will ask for Username and Password.

**Username** : admin

**Password** : <password> (what you provided during creation of cluster.)



**Access main node cluster using SSH from workstation :**

Download **PuTTY** on your windows 10 from the link below.

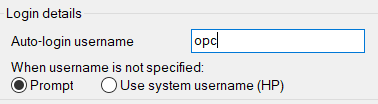
* <https://www.putty.org/>

Open PuTTY.

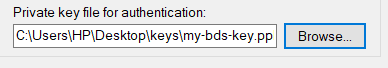
Enter your **Name node public ip** in the Host Name field.



Goto **Connection > Data** and Set below configuration.



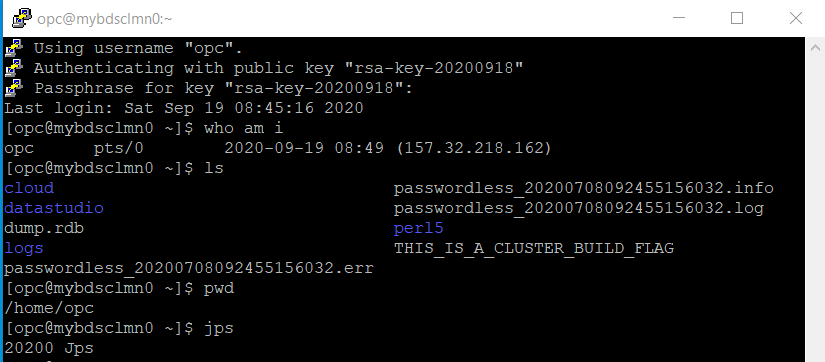
Goto **SSH > Auth** and select your **private key** file associated with the public key that you have setup during cluster creation.



Then click **Open.**

You will get some messages the first time. Click **Yes** and then enter **Passphrase**.

Now you can run a sample command as below.



Document Link : <https://docs.google.com/document/d/1_behwbXqP9fD80jDdzKFxGfAvc6AZlAOkjjXBZFK-qE/edit?usp=sharing>