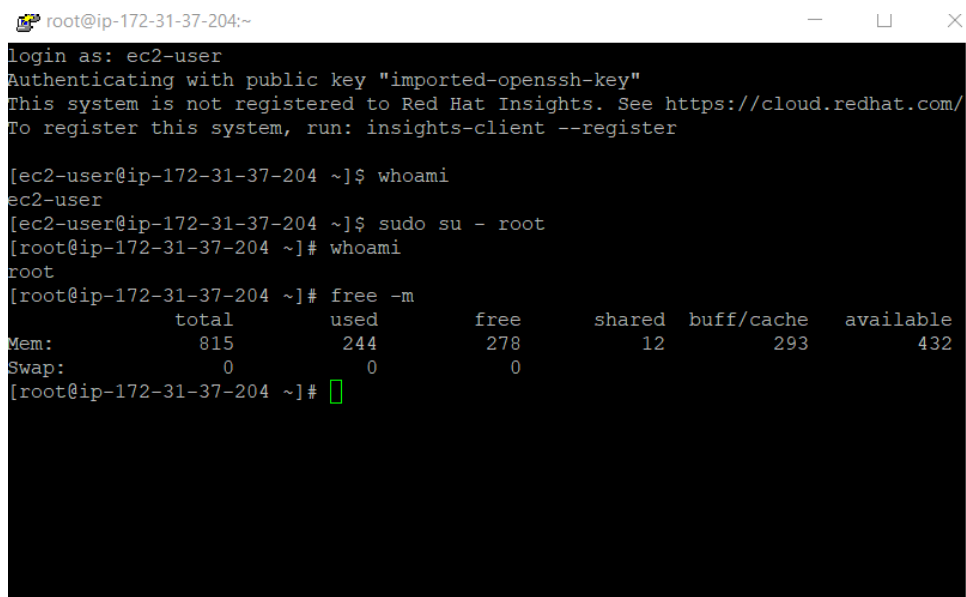


Super Computer with 2+ terabyte storage space:

I want to create a heavy project that includes many fields of computer science, but certainly, my laptop does not have that much storage space. The data that I will need to deploy the project successfully is going to be quite large. Also, my primary goal is that when people check out my project, they don't face any **latency issues**. In previous module, I have created a namenode and datanode using redhat linux operating system. But to increase the efficiency of my system, I need more PCs. So, I decided to launch OS on **AWS cloud**, using the EC2 services. I downloaded multiple instances of redhat linux and made all of them the slave. The storage space at a certain point of time went **beyond 2 terabytes**.

The initial space was as follows:



```
root@ip-172-31-37-204:~  
login as: ec2-user  
Authenticating with public key "imported-openssh-key"  
This system is not registered to Red Hat Insights. See https://cloud.redhat.com/  
To register this system, run: insights-client --register  
  
[ec2-user@ip-172-31-37-204 ~]$ whoami  
ec2-user  
[ec2-user@ip-172-31-37-204 ~]$ sudo su - root  
[root@ip-172-31-37-204 ~]# whoami  
root  
[root@ip-172-31-37-204 ~]# free -m  
              total        used        free      shared  buff/cache   available  
Mem:           815          244          278          12          293          432  
Swap:           0           0           0  
[root@ip-172-31-37-204 ~]#
```

Later I configured the datanode and using the jps command confirmed that it has been started. The steps were the same as shown in the previous module. Same lines were written in the hdfs-site.xml file. The difference was in the core-site.xml file.

The value of the master was:- hdfs://0.0.0.0:9001

It is because, AWS provides two ip, one public and other private. To enable the datanodes to connect to the master this neutral ip was given because, instance in the cloud only knows the private ip. In the core-site.xml file of the slave, the public ip of the master was given. Rest all steps remains same.

```

root@ip-172-31-37-204:/etc/hadoop
-bash: vim: command not found
[root@ip-172-31-37-204 hadoop]# gedit hdfs-site.xml
-bash: gedit: command not found
[root@ip-172-31-37-204 hadoop]# vi hdfs-site.xml
[root@ip-172-31-37-204 hadoop]# vi core-site.xml
[root@ip-172-31-37-204 hadoop]# jps
5217 Jps
[root@ip-172-31-37-204 hadoop]# hadoop-daemon.sh start namenode
starting namenode, logging to /var/log/hadoop/root/hadoop-root-namenode-ip-172-3
[Fatal Error] hdfs-site.xml:12:1: XML document structures must start and end wit
[root@ip-172-31-37-204 hadoop]# jps
5290 Jps
[root@ip-172-31-37-204 hadoop]# vi hdfs-site.xml
[root@ip-172-31-37-204 hadoop]# hadoop-daemon.sh start namenode
starting namenode, logging to /var/log/hadoop/root/hadoop-root-namenode-ip-172-3
[root@ip-172-31-37-204 hadoop]# jps
5368 Jps
[root@ip-172-31-37-204 hadoop]# hadoop-daemon.sh start datanode
starting datanode, logging to /var/log/hadoop/root/hadoop-root-datanode-ip-172-3
1-37-204.ec2.internal.out
[root@ip-172-31-37-204 hadoop]# jps
5459 Jps
5397 DataNode
[root@ip-172-31-37-204 hadoop]#

```

Here we can see that the datanode is connected. Following the similar steps, **100+ datanodes** were created.

```

root@ip-172-31-91-34:/etc/hadoop
Configured Capacity: 2144022429696 (1.95 TB)
Present Capacity: 1879226667008 (1.71 TB)
DFS Remaining: 1879225667584 (1.71 TB)
DFS Used: 999424 (976 KB)
DFS Used%: 0%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0

-----
Datanodes available: 121 (121 total, 0 dead)

Name: 35.154.45.137:50010
Decommission Status : Normal
Configured Capacity: 10724814848 (9.99 GB)
DFS Used: 8192 (8 KB)
Non DFS Used: 1974112256 (1.84 GB)
DFS Remaining: 8750694400 (8.15 GB)
DFS Used%: 0%
DFS Remaining%: 81.59%
Last contact: Tue Sep 22 15:48:37 UTC 2020
:

```

Here, 121 datanodes were connected and the memory was around **1.95 TB**. After connecting more slaves, the storage space went up more.

```
root@ip-172-31-37-204:/etc/hadoop
Configured Capacity: 2793650765824 (2.54 TB)
Present Capacity: 2388343844864 (2.17 TB)
DFS Remaining: 2388342607872 (2.17 TB)
DFS Used: 1236992 (1.18 MB)
DFS Used%: 0%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0

-----
Datanodes available: 149 (274 total, 125 dead)

Name: 13.233.28.233:50010
Decommission Status : Normal
Configured Capacity: 32199651328 (29.99 GB)
DFS Used: 8192 (8 KB)
Non DFS Used: 2060455936 (1.92 GB)
DFS Remaining: 30139187200 (28.07 GB)
DFS Used%: 0%
DFS Remaining%: 93.6%
Last contact: Tue Sep 22 16:51:50 UTC 2020

:
```

In **total 274 slaves were connected**. The present capacity as seen on the first line of the image is **2.17 TB**.

Now, this supercomputer is ready to deal with **BIGDATA** and solve two prominent problems that is volume and velocity.

Now, the master on receiving the data will strip it into 274 block and send each block to its slave's memory. This will increase the efficiency of the system. The system becomes faster and after exceeding the number of slaves to a number, it is the "super computer".

In the next module, we shall be working on this "**Hadoop cluster**", to solve several industrial use-cases.