Vidya Vikas Education Trust's
Universal College of Engineering, Kaman Road, Vasai – 401208
Accredited A Grade by NAAC

Department of AIML Engineering
Subject: Big Data Analytics Lab (CSL702)

Experiment No:1

Roll No: 10	Name:Rajeshwar Bura	Div: C	Batch: C1

Code:

Step 1 : Install Java Development Kit

sudo apt update && sudo apt install openjdk-8-jdk

Step 2 : Verify the Java version : java -version

```
sanjay@sanjay-VirtualBox:~$ java --version
openjdk 11.0.20.1 2023-08-24
OpenJDK Runtime Environment (build 11.0.20.1+1-post-Ubuntu-0ubuntu123.04)
OpenJDK 64-Bit Server VM (build 11.0.20.1+1-post-Ubuntu-0ubuntu123.04, mixed mode, sharing)
sanjay@sanjay-VirtualBox:~$
```

Step 3 : Install SSH :

sudo apt install ssh.

Step 4 : Create the hadoop user :

sudo adduser hadoop

```
sanjay@sanjay-VirtualBox:~$ sudo adduser hadoop
Adding user `hadoop' ...
Adding new group `hadoop' (1001) ...
Adding new user `hadoop' (1001) with group `hadoop (1001)' ...
adduser: The home directory `/home/hadoop' already exists.  Not copying from `/etc/skel'.
New password:
BAD PASSWORD: The password is shorter than 8 characters
Retype new password:
passwd: password updated successfully
Changing the user information for hadoop
Enter the new value, or press ENTER for the default
        Full Name []:
        Room Number []:
        Work Phone []:
        Home Phone []:
        Other []:
Is the information correct? [Y/n] y
Adding new user `hadoop' to supplemental / extra groups `users' ...
Adding user `hadoop' to group `users' ...
sanjay@sanjay-VirtualBox:~$ su - hadoop
Password:
hadoop@sanjay-VirtualBox:~$
```

Step 5 : Switch user :

su - hadoop

Step 6 : Configure SSH :

ssh-keygen -t rsa

Department of AIML Engineering
Subject: Big Data Analytics Lab (CSL702)

```
hadoop@sanjay-VirtualBox:~$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hadoop/.ssh/id_rsa):
Created directory '/home/hadoop/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/hadoop/.ssh/id_rsa
Your public key has been saved in /home/hadoop/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:yu8Hsie3mbQ7UifnfH6iam4kFLRRbEb9zVYGUtbaYyg hadoop@sanjay-VirtualBox
The key's randomart image is:
+---[RSA 3072]----+
|     .o+o.  ..+o |
|      .o+ .  o  +|
|      .+   . o.= |
|      .    E..=o.|
|     .  S   ... .|
|     .o.= o      |
|      o*.B       |
|      +o*++ o .  |
|       X@*.+.o   |
+----[SHA256]-----+
hadoop@sanjay-VirtualBox:~$
```

Step 7 : Set permissions :

cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

chmod 640 ~/.ssh/authorized_keys

Step 8 : SSH to the localhost:

ssh localhost

```
hadoop@sanjay-VirtualBox:~/hadoop$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ED25519 key fingerprint is SHA256:2ZsL3K5BKG6h8isZpTufDvB69zWFKS7iFjvsnhWr53I.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'localhost' (ED25519) to the list of known hosts.
Welcome to Ubuntu 23.04 (GNU/Linux 6.2.0-32-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

122 updates can be applied immediately.
To see these additional updates run: apt list --upgradable


The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

hadoop@sanjay-VirtualBox:~$
```

Step 9 : Switch user

su - hadoop

Step 10 : Install Hadoop

wget https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz

Department of AIML Engineering
Subject: Big Data Analytics Lab (CSL702)

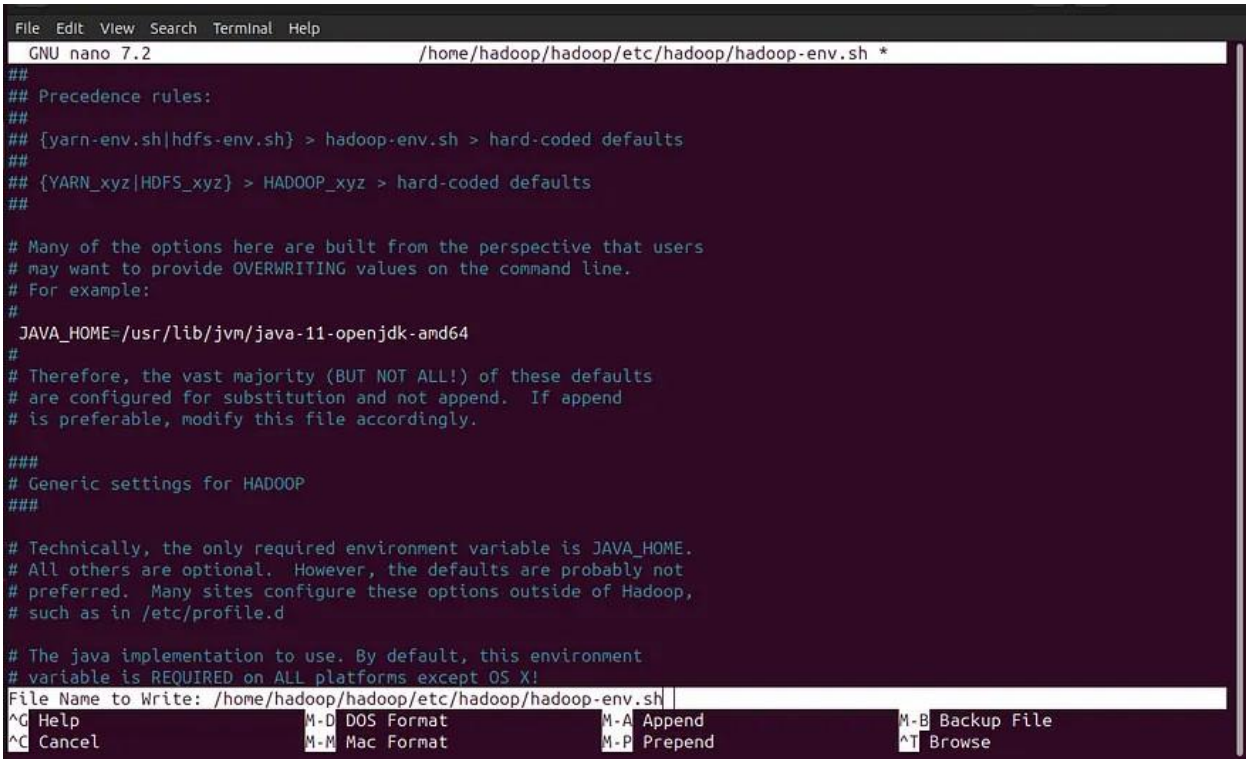
tar -xvzf hadoop-3.3.6.tar.gz
mv hadoop-3.3.6 hadoop

```
hadoop@sanjay-VirtualBox:~$ mv hadoop-3.3.6 hadoop
hadoop@sanjay-VirtualBox:~$ ls
hadoop  hadoop-3.3.6.tar.gz
hadoop@sanjay-VirtualBox:~$
```

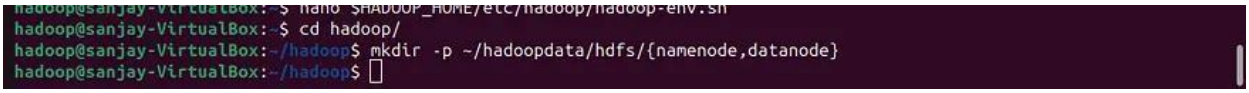
nano ~/.bashrc
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_HOME=/home/hadoop/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

```
File  Edit  View  Search  Terminal  Help
 GNU nano 7.2                                  /home/hadoop/.bashrc *
# Alias definitions.
# You may want to put all your additions into a separate file like
# ~/.bash_aliases, instead of adding them here directly.
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
    . ~/.bash_aliases
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export HADOOP_HOME=/home/hadoop/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

^G Help          ^O Write Out     ^W Where Is      ^K Cut           ^T Execute       ^C Location      M-U Undo
^X Exit          ^R Read File     ^\ Replace       ^U Paste         ^J Justify       ^/ Go To Line    M-E Redo
```

source ~/.bashrc
JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64



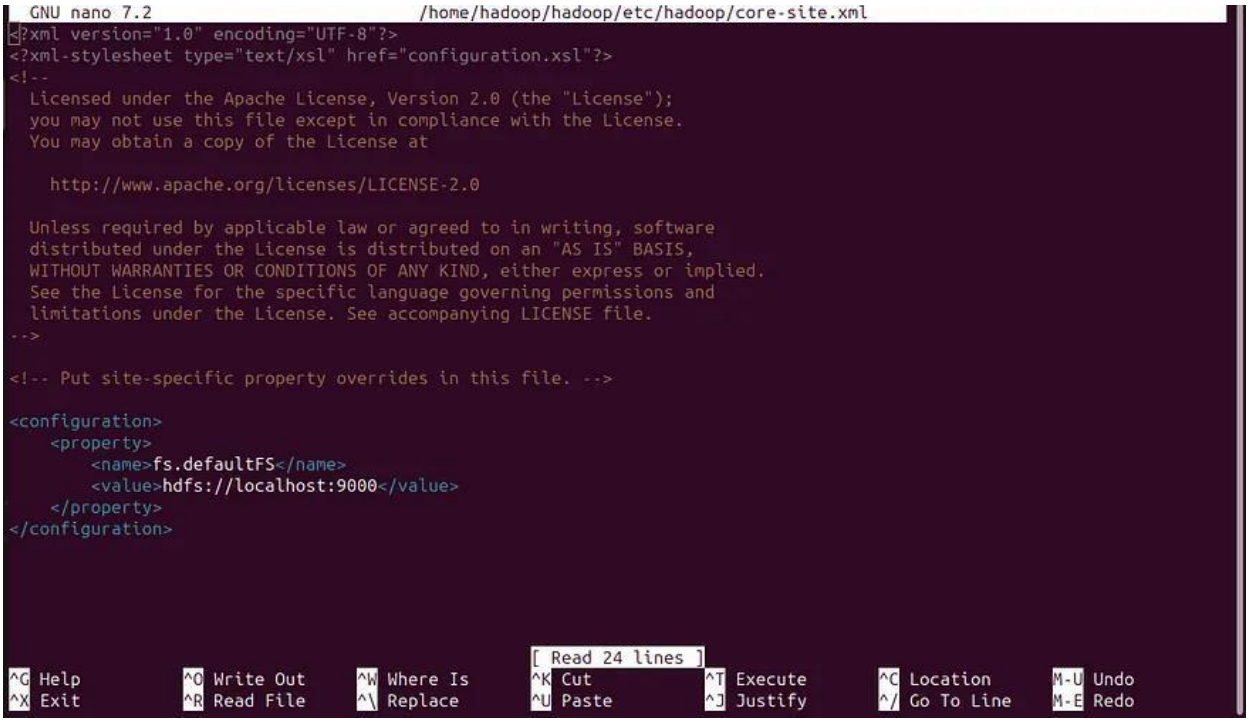
Step 11 : Configuring Hadoop :
cd hadoop/
mkdir -p ~/hadoopdata/hdfs/{namenode,datanode}



nano $HADOOP_HOME/etc/hadoop/core-site.xml

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

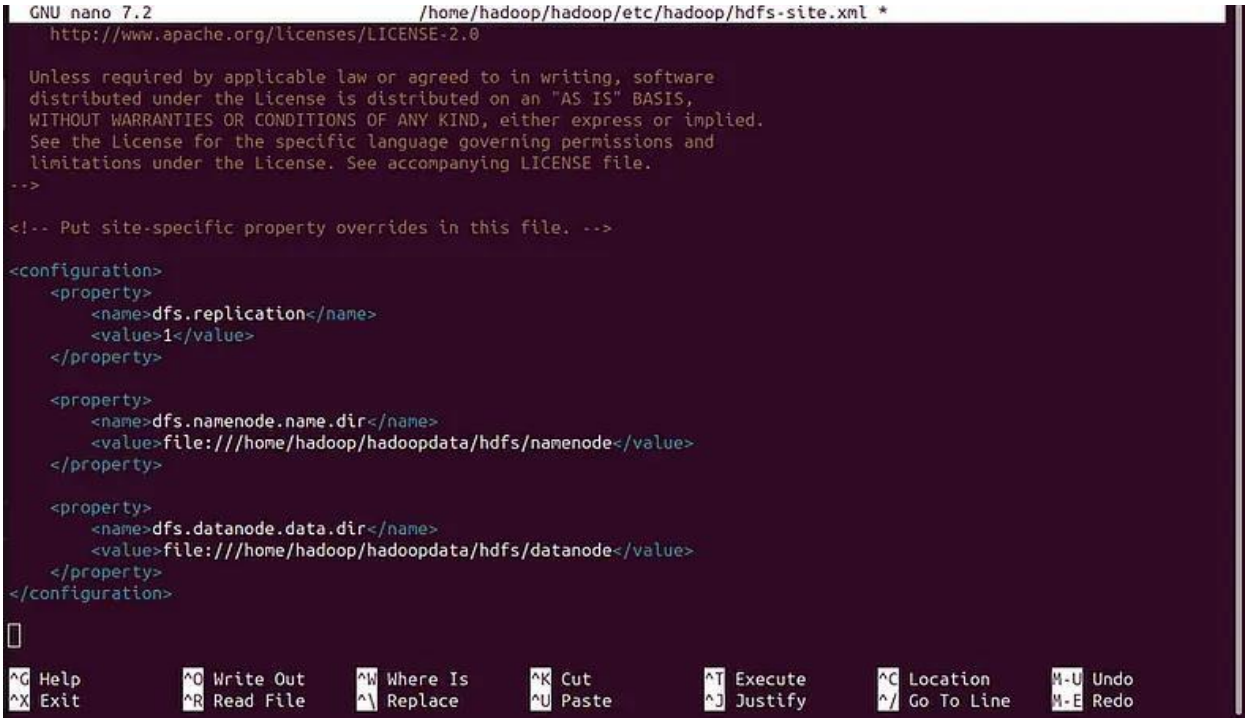
Department of AIML Engineering
Subject: Big Data Analytics Lab (CSL702)



nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml

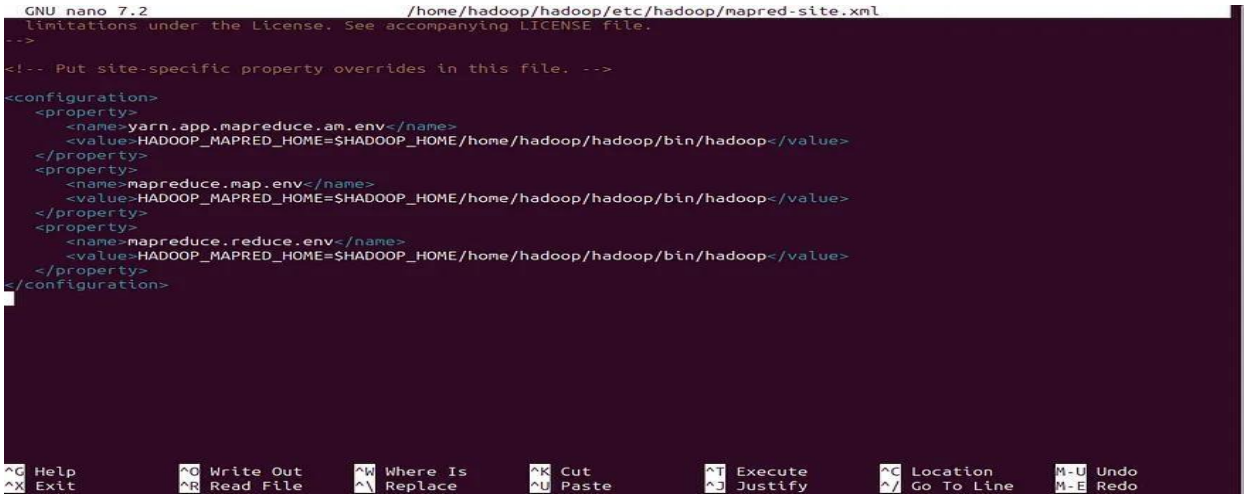
```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>
  </property>
</configuration>
```

Department of AIML Engineering
Subject: Big Data Analytics Lab (CSL702)



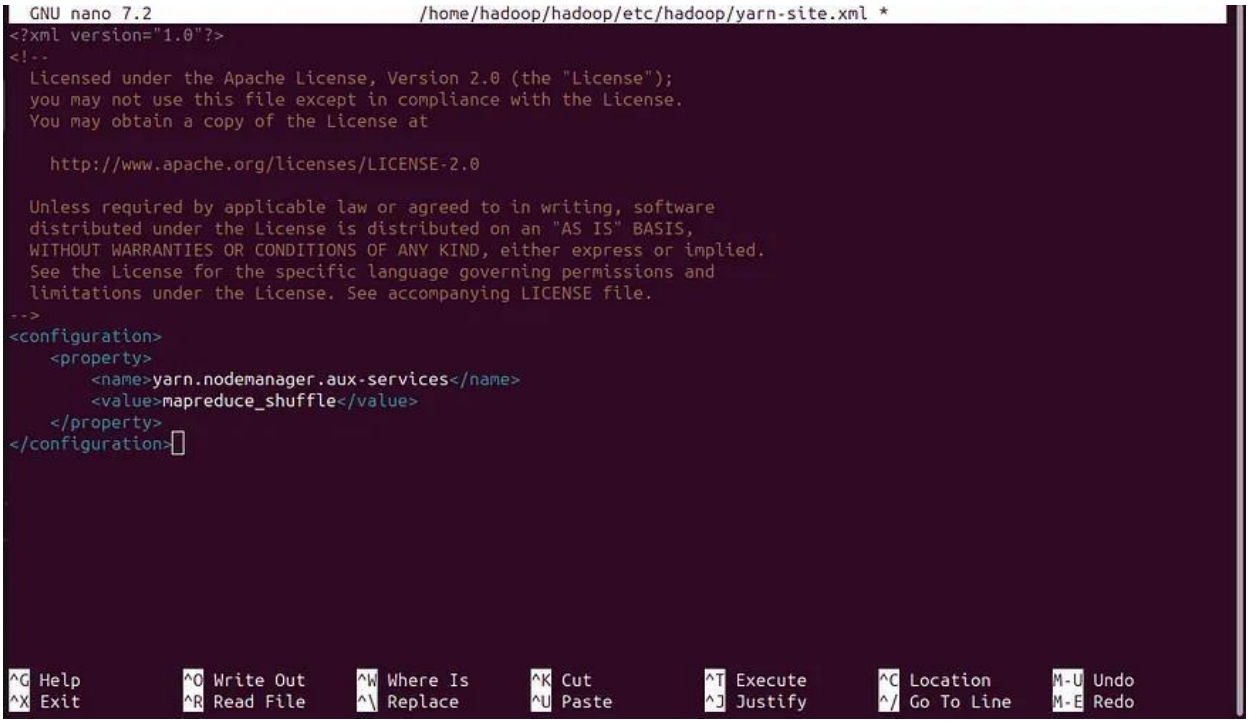
nano $HADOOP_HOME/etc/hadoop/yarn-site.xml

```
<configuration>
   <property>
      <name>yarn.nodemanager.aux-services</name>
      <value>mapreduce_shuffle</value>
   </property>
</configuration>
```



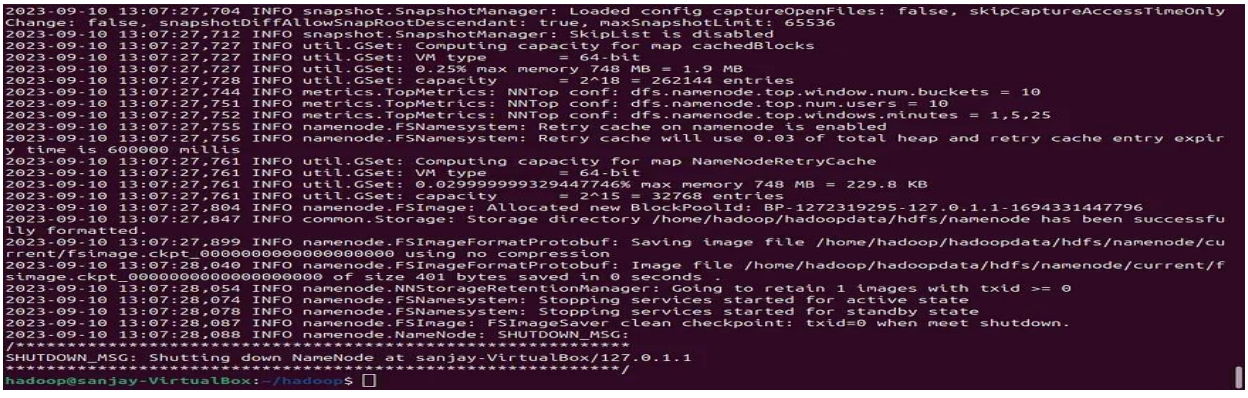
nano $HADOOP_HOME/etc/hadoop/yarn-site.xml

```
<configuration>
   <property>
      <name>yarn.nodemanager.aux-services</name>
      <value>mapreduce_shuffle</value>
   </property>
</configuration>
```



Step 12 : Start Hadoop cluster:

hdfs namenode -format



start-all.sh

```
hadoop@sanjay-VirtualBox:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [sanjay-VirtualBox]
Starting resourcemanager
Starting nodemanagers
hadoop@sanjay-VirtualBox:~$
```

Jps

```
hadoop@sanjay-VirtualBox:~/hadoop$ jps
7235 NodeManager
6677 DataNode
7593 Jps
6554 NameNode
7116 ResourceManager
6893 SecondaryNameNode
hadoop@sanjay-VirtualBox:~/hadoop$
```

Step 13 : Access Hadoop Namenode and Resource Manager :

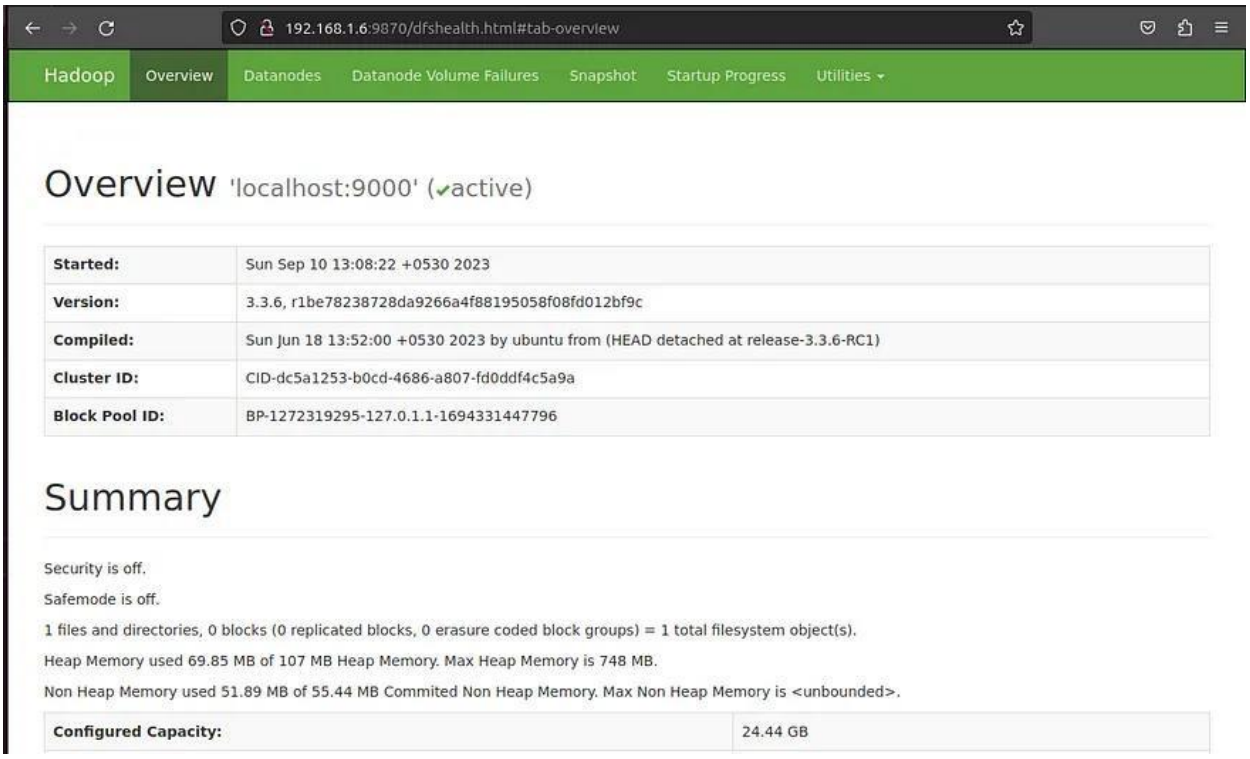
sudo apt install net-tools

ifconfig

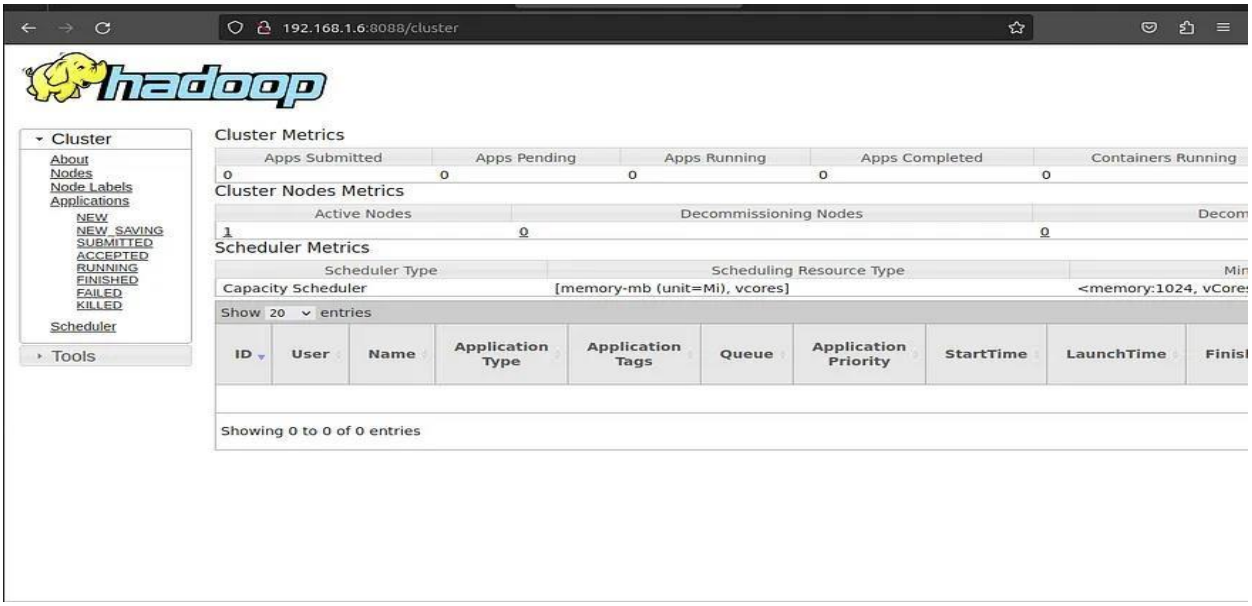
```
hadoop@sanjay-VirtualBox:~/hadoop$ ifconfig
enp0s3: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.1.6  netmask 255.255.255.0  broadcast 192.168.1.255
        inet6 2401:4900:1c28:46c4:f76c:b206:abe3:2d45  prefixlen 64  scopeid 0x0<global>
        inet6 2401:4900:1c28:46c4:ed13:53f4:5c05:50c6  prefixlen 64  scopeid 0x0<global>
        inet6 fe80::112b:300a:9242:51f3  prefixlen 64  scopeid 0x20<link>
        ether 08:00:27:83:31:35  txqueuelen 1000  (Ethernet)
        RX packets 645228  bytes 934388358 (934.3 MB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 93618  bytes 8998032 (8.9 MB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING>  mtu 65536
        inet 127.0.0.1  netmask 255.0.0.0
        inet6 ::1  prefixlen 128  scopeid 0x10<host>
        loop  txqueuelen 1000  (Local Loopback)
        RX packets 3331  bytes 491873 (491.8 KB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 3331  bytes 491873 (491.8 KB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

hadoop@sanjay-VirtualBox:~/hadoop$
```

http://192.168.1.6:9870

Department of AIML Engineering
Subject: Big Data Analytics Lab (CSL702)



http://192.168.1.6:8088



Department of AIML Engineering
Subject: Big Data Analytics Lab (CSL702)

Step 13 :Verify the Hadoop Cluster :

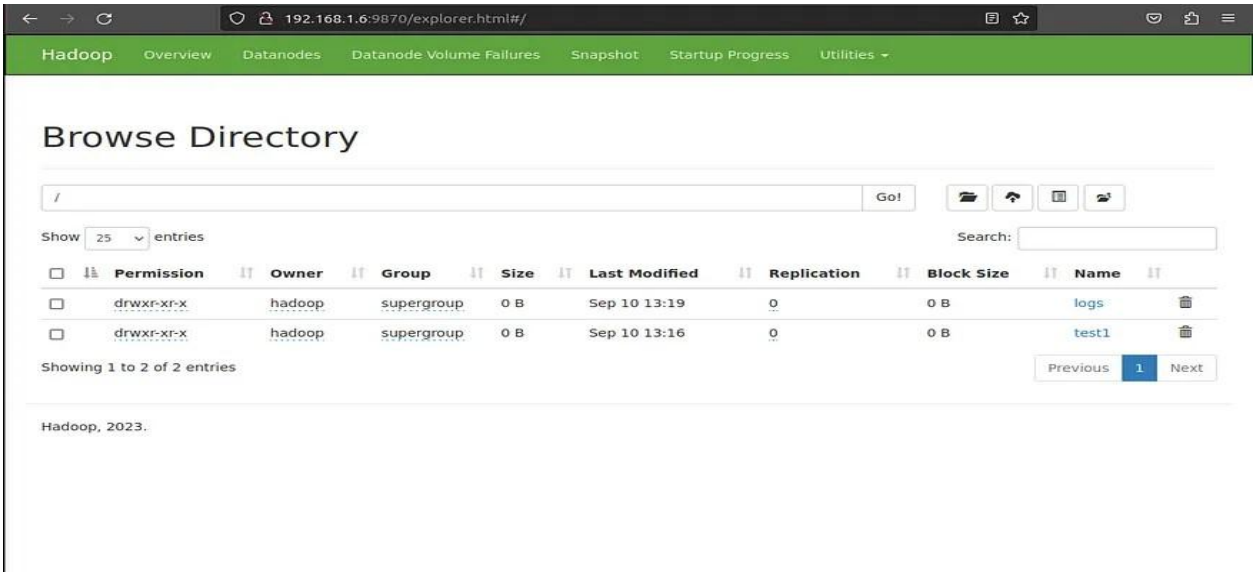
hdfs dfs -mkdir /test1

hdfs dfs -mkdir /logs

hdfs dfs -ls /

```
hadoop@sanjay-VirtualBox:~/hadoop$ hdfs dfs -ls /
Found 2 items
drwxr-xr-x   - hadoop supergroup          0 2023-09-10 13:16 /logs
drwxr-xr-x   - hadoop supergroup          0 2023-09-10 13:16 /test1
hadoop@sanjay-VirtualBox:~/hadoop$
```

hdfs dfs -put /var/log/* /logs/



Step 14 : To stop hadoop services :

stop-all.sh

```
hadoop@sanjay-VirtualBox:~/hadoop$ stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [sanjay-VirtualBox]
Stopping nodemanagers
Stopping resourcemanager
hadoop@sanjay-VirtualBox:~/hadoop$
```