**Cleveland State University**



***CIS 612: Big Data and Parallel Database Processing Systems***

**Lab Assignment 4**

**Submitted By: Submitted To:**

Shivam Joshi Dr. SS Chung

CSU ID – 2781474

Command which have to follow before initiating creating Hadoop environment and user.

Make sure your system of linux is updated this will help to get all the possible thing installed.

**Pre-requisite thing to do before user creation**

**sudo apt-get update**

**sudo apt-get upgrade**

check your java version and update that as well

**java -version**

Installing JAVA

**sudo apt-get install default-jdk**

again check the make sure it work fine

**java -version**

now its time to Install SSH server which will let us use LOCALHOST

**sudo apt update**

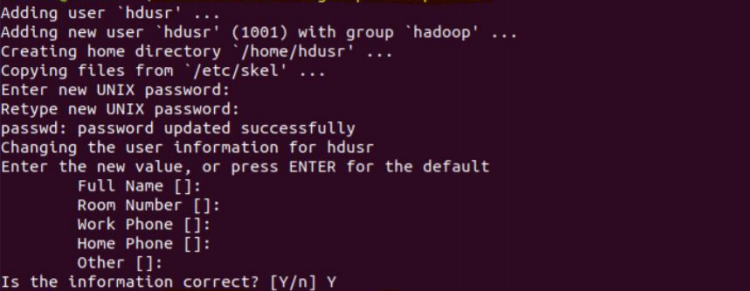
**sudo apt install openshh-server**

**sudo apt-get install ssh**

Now Its time for the USER creation

**sudo addgroup hadoop**

**sudo adduser -ingroup hadoop hdusr (hdusr can be any name according to you)**

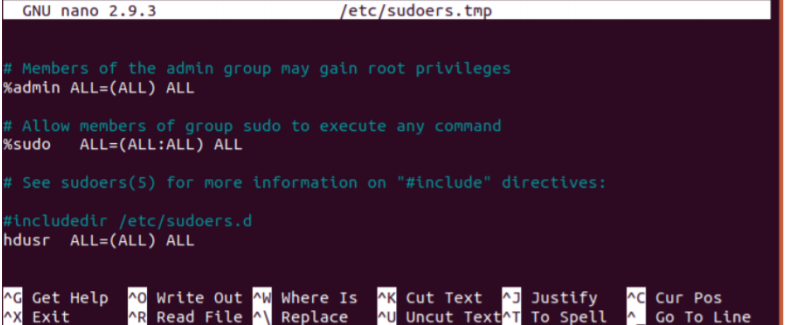
****

**sudo adduser hdusr sudo**

**sudo visudo**

and add this line at the end of the file

**hdusr ALL=(ALL) ALL**



After this change the user to hdusr

**su – hdusr**

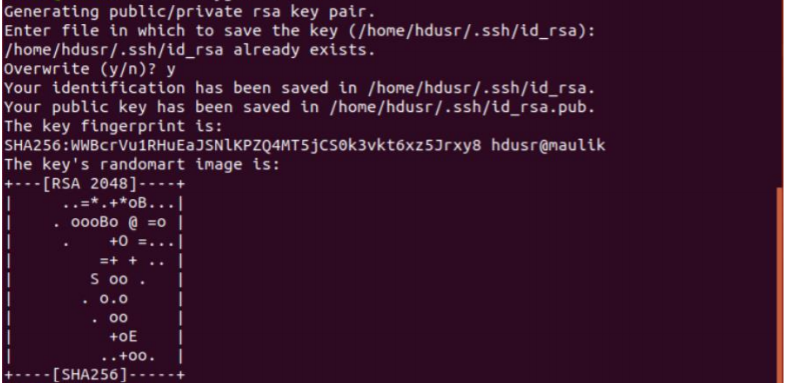
SSH Key Generation for User hdusr

This will create empty password RSA key pair

**ssh-keygen -t rsa -P “”**



Make sure all the double quotes look like this



Enable SSH access with the key generated in the previous step

sudo cat /home/”your username”/.ssh/id\_rsa.pub >> /home/”your username”/.shh/authorized\_keys

NOW REBOOT YOUR SYSTEM

sudo reboot now.

Now Open terminal and login to your hadoop user

su – shiv

SSH setup by connecting to local machine

ssh localhost

Download stable hadoop version and extract it

wget <https://archive.apache.org/dist/hadoop/core/hadoop-2.7.2/hadoop-2.7.2.tar.gz>

tar xzf hadoop-2.7.2.tar.gz

Now CONFIGURE HADOOP ENVIRONMENT VARIABLE (bashrc)

sudo nano .bashrc

#Hadoop Related Options

export HADOOP\_HOME=/home/”your hadoop username”/hadoop-3.2.1

export HADOOP\_INSTALL=$HADOOP\_HOME

export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_HOME=$HADOOP\_HOME

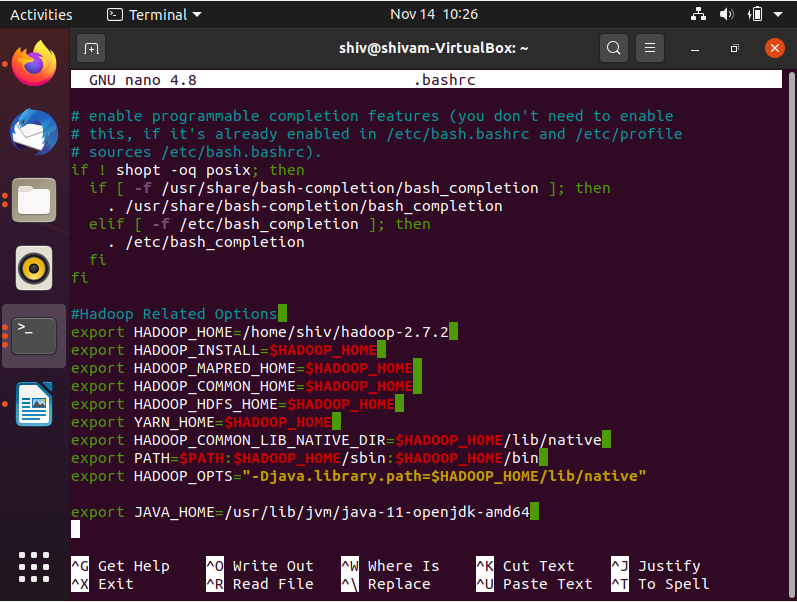
export HADOOP\_HDFS\_HOME=$HADOOP\_HOME

export 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”



shiv@shivam-VirtualBox:~$ tar xzf hadoop-2.7.2.tar.gz



shiv@shivam-VirtualBox:~$ sudo nano .bashrc



shiv@shivam-VirtualBox:~$ which java



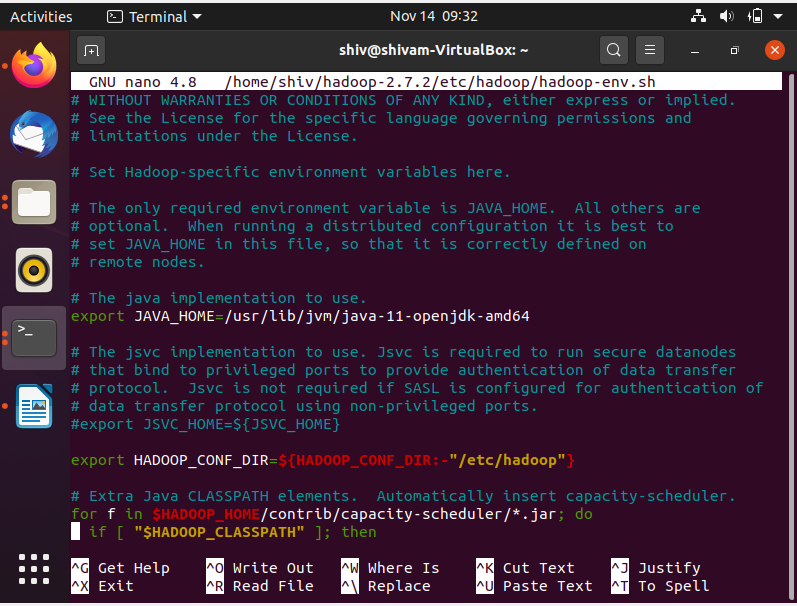
shiv@shivam-VirtualBox:~$ readlink -f /usr/bin/java



shiv@shivam-VirtualBox:~$ source ~/.bashrc

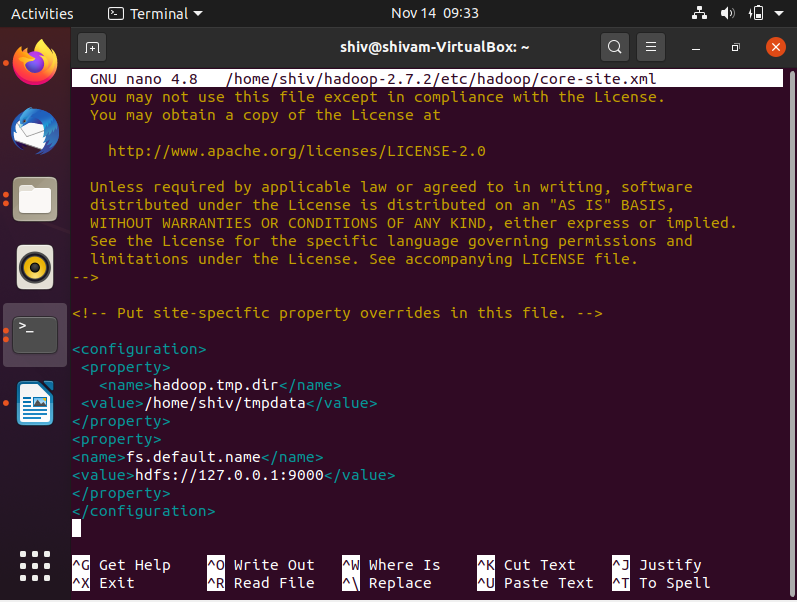
shiv@shivam-VirtualBox:~$ sudo nano $HADOOP\_HOME/etc/hadoop/hadoop-env.sh

export JAVA\_HOME=/usr/lib/jvm/java-11-openjdk-amd64



shiv@shivam-VirtualBox:~$ sudo nano $HADOOP\_HOME/etc/hadoop/hadoop-env.sh

shiv@shivam-VirtualBox:~$ sudo nano $HADOOP\_HOME/etc/hadoop/core-site.xml



<configuration>

<property>

<name>hadoop.tmp.dir</name>

<value>/home/shiv/tmpdata</value> // after home/”you have to enter your hadoop username”

</property>

<property>

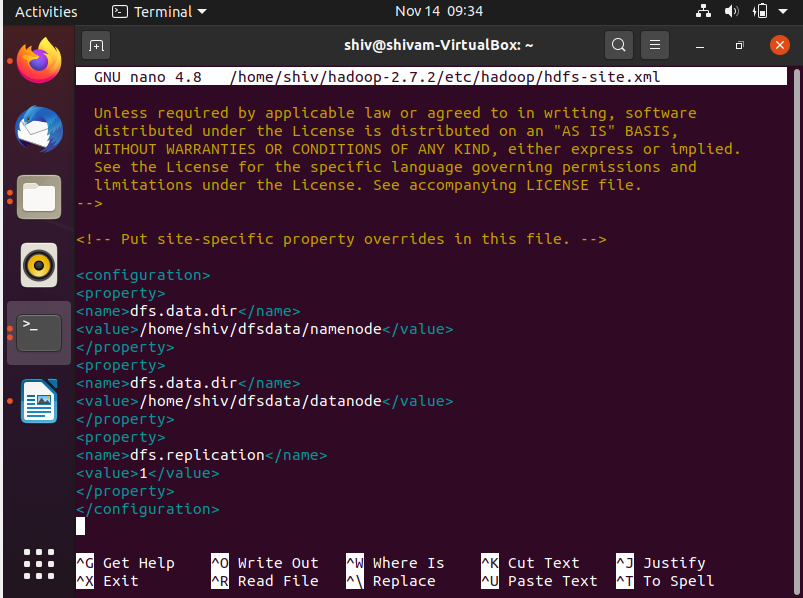
<name>fs.default.name</name>

<value>hdfs://127.0.0.1:9000</value>

</property>

</configuration>

shiv@shivam-VirtualBox:~$ sudo nano $HADOOP\_HOME/etc/hadoop/hdfs-site.xml



<configuration> <property>

<name>dfs.data.dir</name>

<value>/home/shiv/dfsdata/namenode</value> //shiv = your hadoop user name

</property>

<property>

<name>dfs.data.dir</name>

<value>/home/shiv/dfsdata/datanode</value> //shiv = your hadoop user name

</property>

<property>

<name>dfs.replication</name>

<value>1</value>

</property></configuration>

shiv@shivam-VirtualBox:~$ sudo mkdir /home/shiv/dfsdata/namenode

mkdir: cannot create directory ‘/home/shiv/dfsdata/namenode’: No such file or directory

shiv@shivam-VirtualBox:~$ mkdir /home/hdoop/dfsdata/namenode

mkdir: cannot create directory ‘/home/hdoop/dfsdata/namenode’: No such file or directory

shiv@shivam-VirtualBox:~$ sudo mkdir /dfsdata

shiv@shivam-VirtualBox:~$ sudo mkdir /dfsdata/namenode

shiv@shivam-VirtualBox:~$ sudo mkdir /dfsdata/datanode

shiv@shivam-VirtualBox:~$ sudo chmod -777 /home/shiv/dfsdata/namenode

chmod: cannot access '/home/shiv/dfsdata/namenode': No such file or directory

shiv@shivam-VirtualBox:~$ sudo chmod -777 /dfsdata/namenode

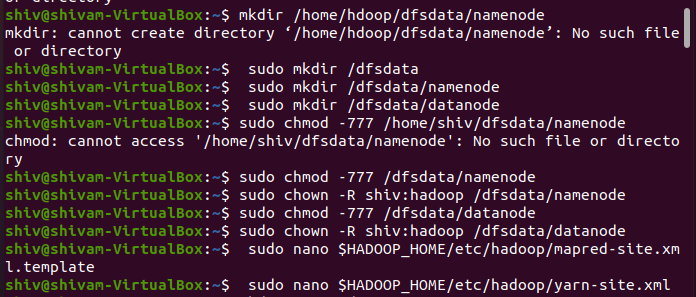
shiv@shivam-VirtualBox:~$ sudo chown -R shiv:hadoop /dfsdata/namenode

shiv@shivam-VirtualBox:~$ sudo chmod -777 /dfsdata/datanode

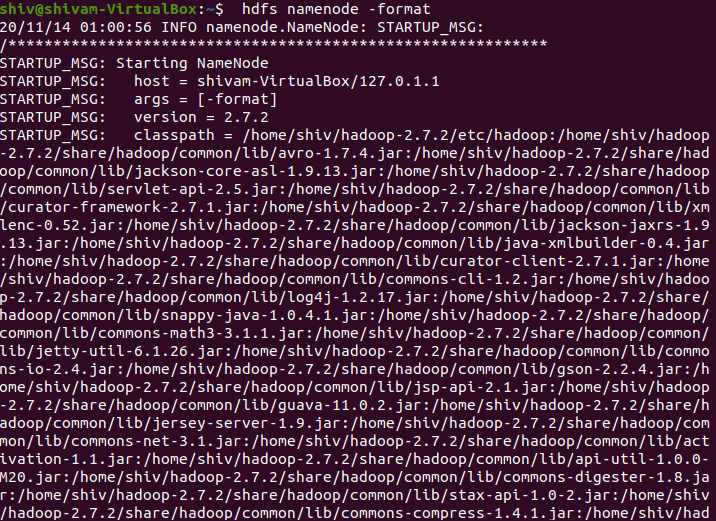
shiv@shivam-VirtualBox:~$ sudo chown -R shiv:hadoop /dfsdata/datanode

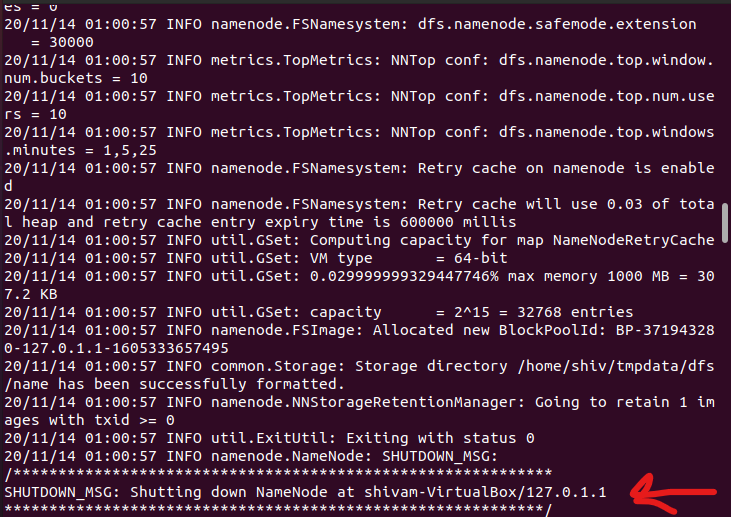
shiv@shivam-VirtualBox:~$ sudo nano $HADOOP\_HOME/etc/hadoop/mapred-site.xml.template

shiv@shivam-VirtualBox:~$ sudo nano $HADOOP\_HOME/etc/hadoop/yarn-site.xml



shiv@shivam-VirtualBox:~$ hdfs namenode -format

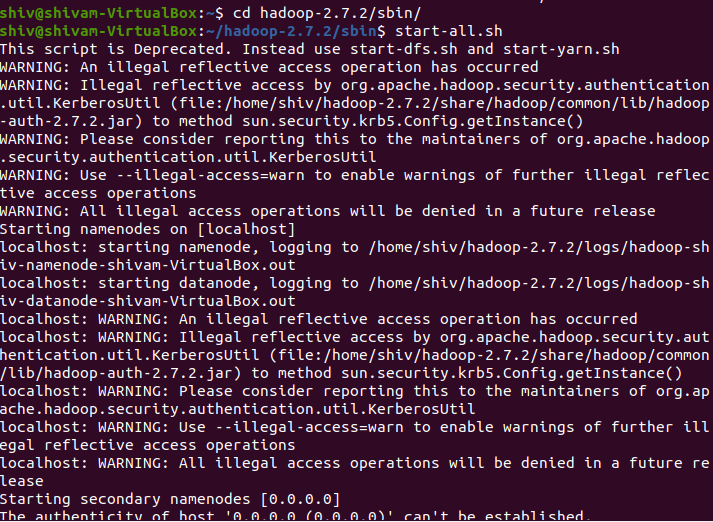




We need to see this msg, which confirm us that NameNode at particular machine is down.

shiv@shivam-VirtualBox:~$ cd hadoop-2.7.2/sbin/

shiv@shivam-VirtualBox:~/hadoop-2.7.2/sbin$ start-all.sh



shiv@shivam-VirtualBox:~/hadoop-2.7.2/sbin$ jps

3714 NodeManager

3202 DataNode

4003 Jps

3046 NameNode

3420 SecondaryNameNode

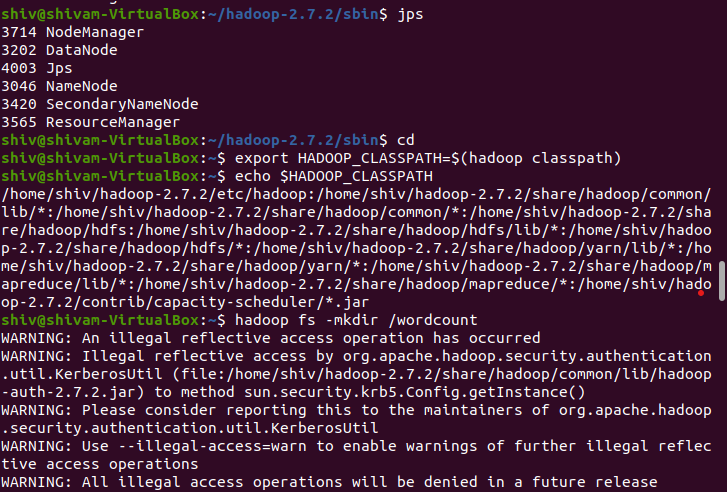
3565 ResourceManager

shiv@shivam-VirtualBox:~/hadoop-2.7.2/sbin$ cd

shiv@shivam-VirtualBox:~$ export HADOOP\_CLASSPATH=$(hadoop classpath)

shiv@shivam-VirtualBox:~$ echo $HADOOP\_CLASSPATH

shiv@shivam-VirtualBox:~$ hadoop fs -mkdir /wordcount



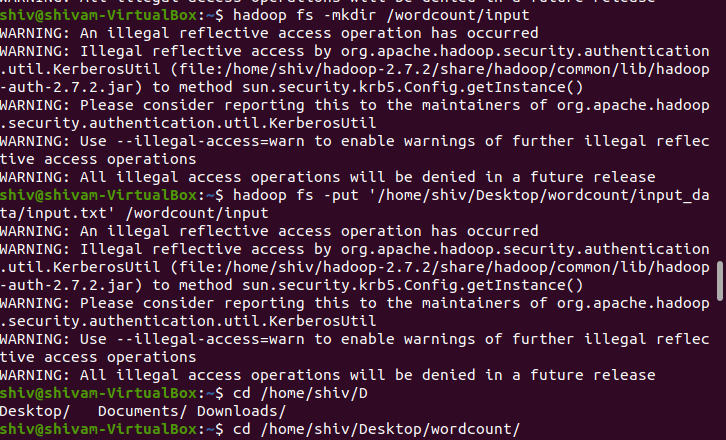
shiv@shivam-VirtualBox:~$ hadoop fs -mkdir /wordcount/input

shiv@shivam-VirtualBox:~$ hadoop fs -put '/home/shiv/Desktop/wordcount/input\_data/input.txt' /wordcount/input

shiv@shivam-VirtualBox:~$ cd /home/shiv/D

Desktop/ Documents/ Downloads/

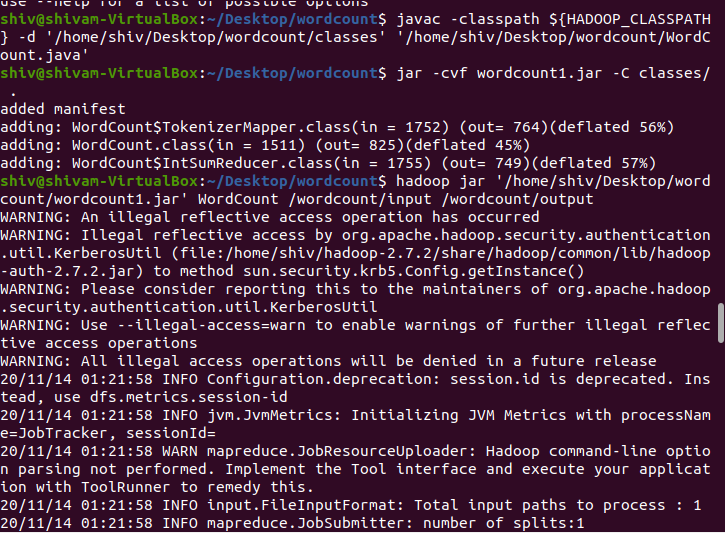
shiv@shivam-VirtualBox:~$ cd /home/shiv/Desktop/wordcount/



shiv@shivam-VirtualBox:~/Desktop/wordcount$ javac -classpath ${HADOOP\_CLASSPATH} -d '/home/shiv/Desktop/wordcount/classes' '/home/shiv/Desktop/wordcount/WordCount.java'

shiv@shivam-VirtualBox:~/Desktop/wordcount$ jar -cvf wordcount1.jar -C classes/ .

shiv@shivam-VirtualBox:~/Desktop/wordcount$ hadoop jar '/home/shiv/Desktop/wordcount/wordcount1.jar' WordCount /wordcount/input /wordcount/output



shiv@shivam-VirtualBox:~/Desktop/wordcount$ hadoop dfs -cat /wordcount/output/\*

