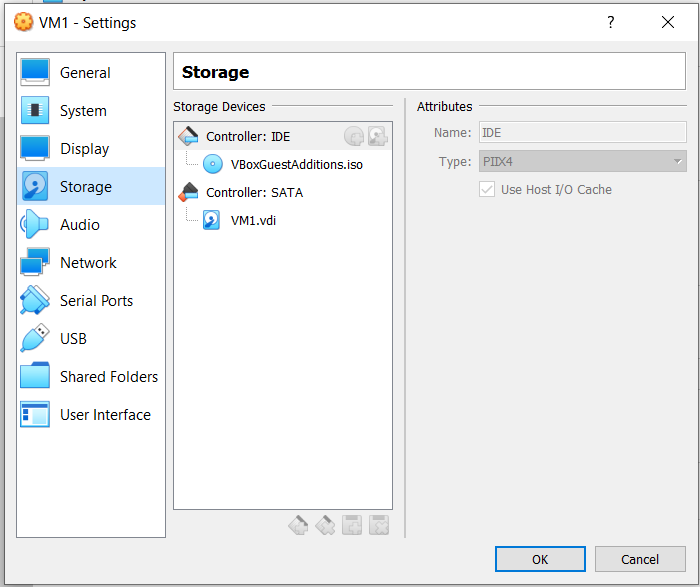
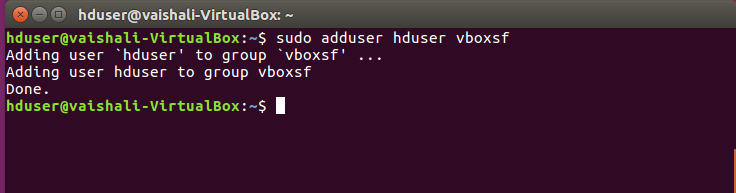
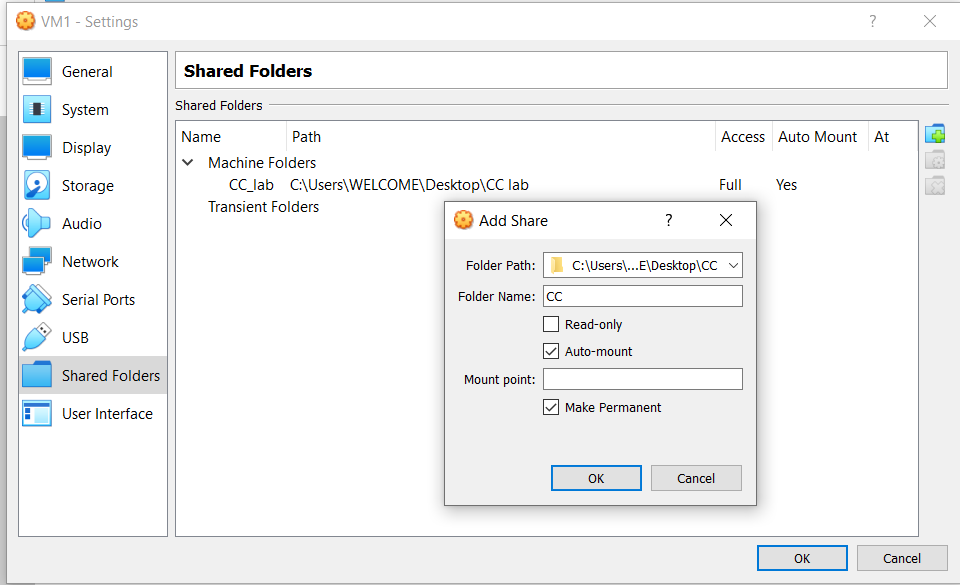
**Ex.No.5 Installation of Single Node Hadoop and Executing Word Count**

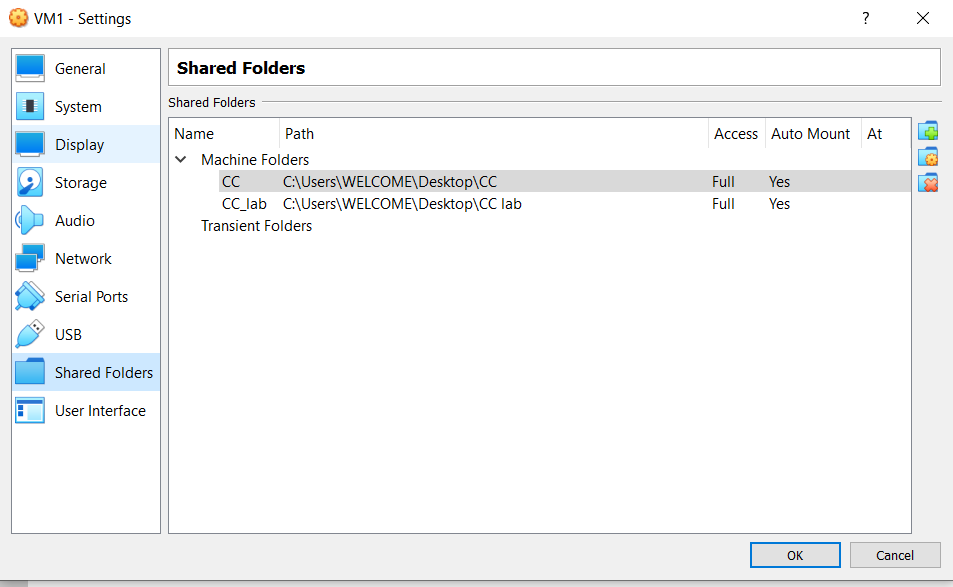
**Program**

1. Adding VBox Guest Additions

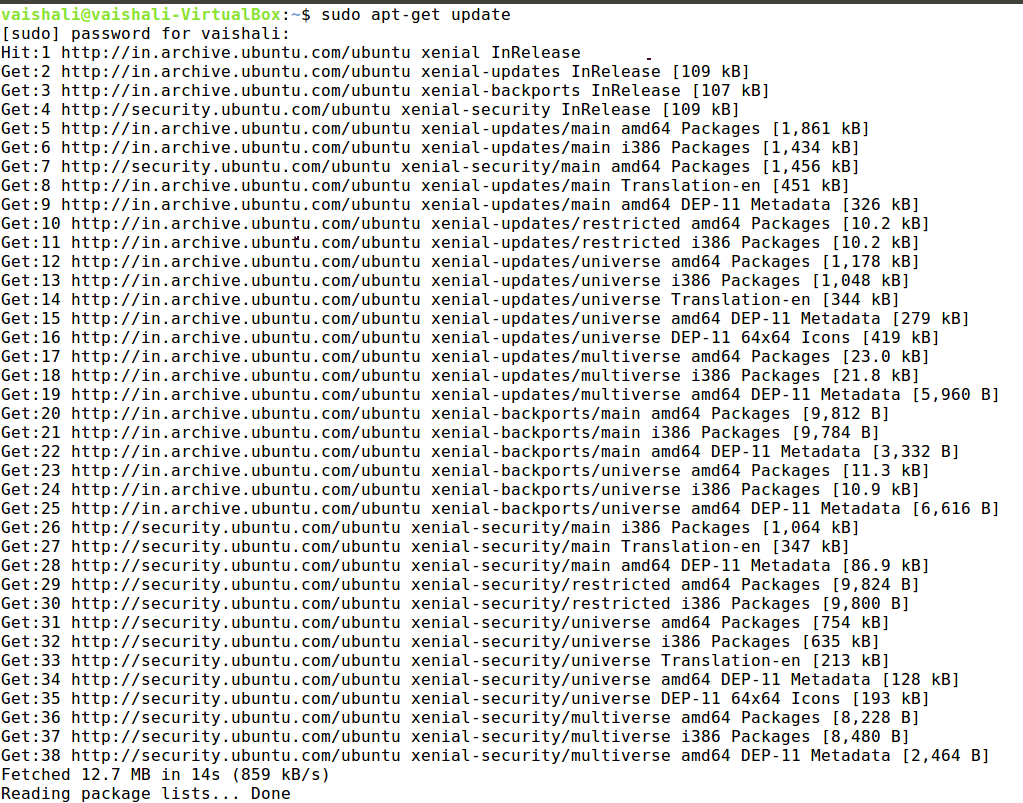




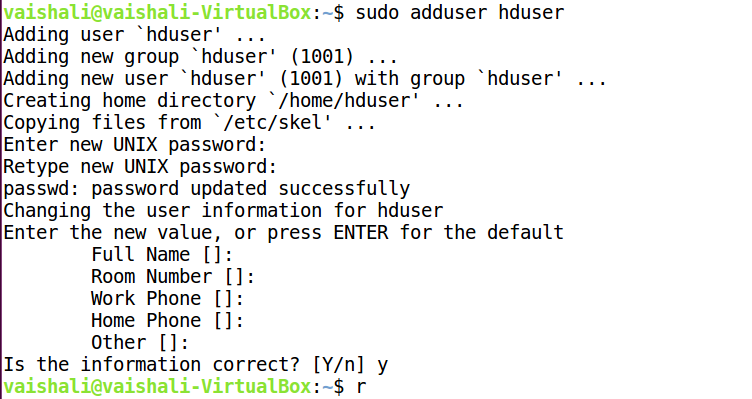




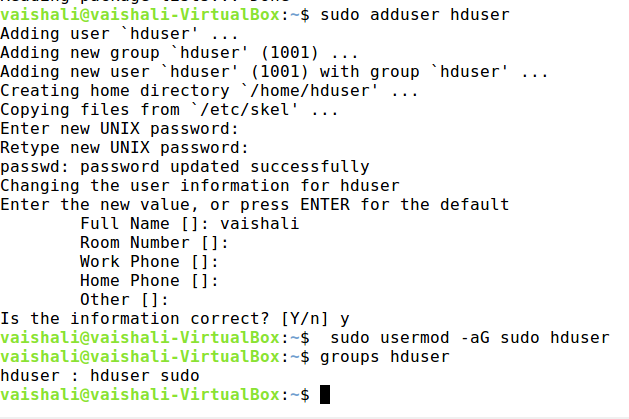
1. Update all the software in Ubuntu



1. Create new user in system using the command **sudo adduser hduser**



4) Add hduser to sudoers list using the command **sudo usermod -aG sudo** **hduser** and verify if the user belongs to Sudo Group using the command **groups hduser.**

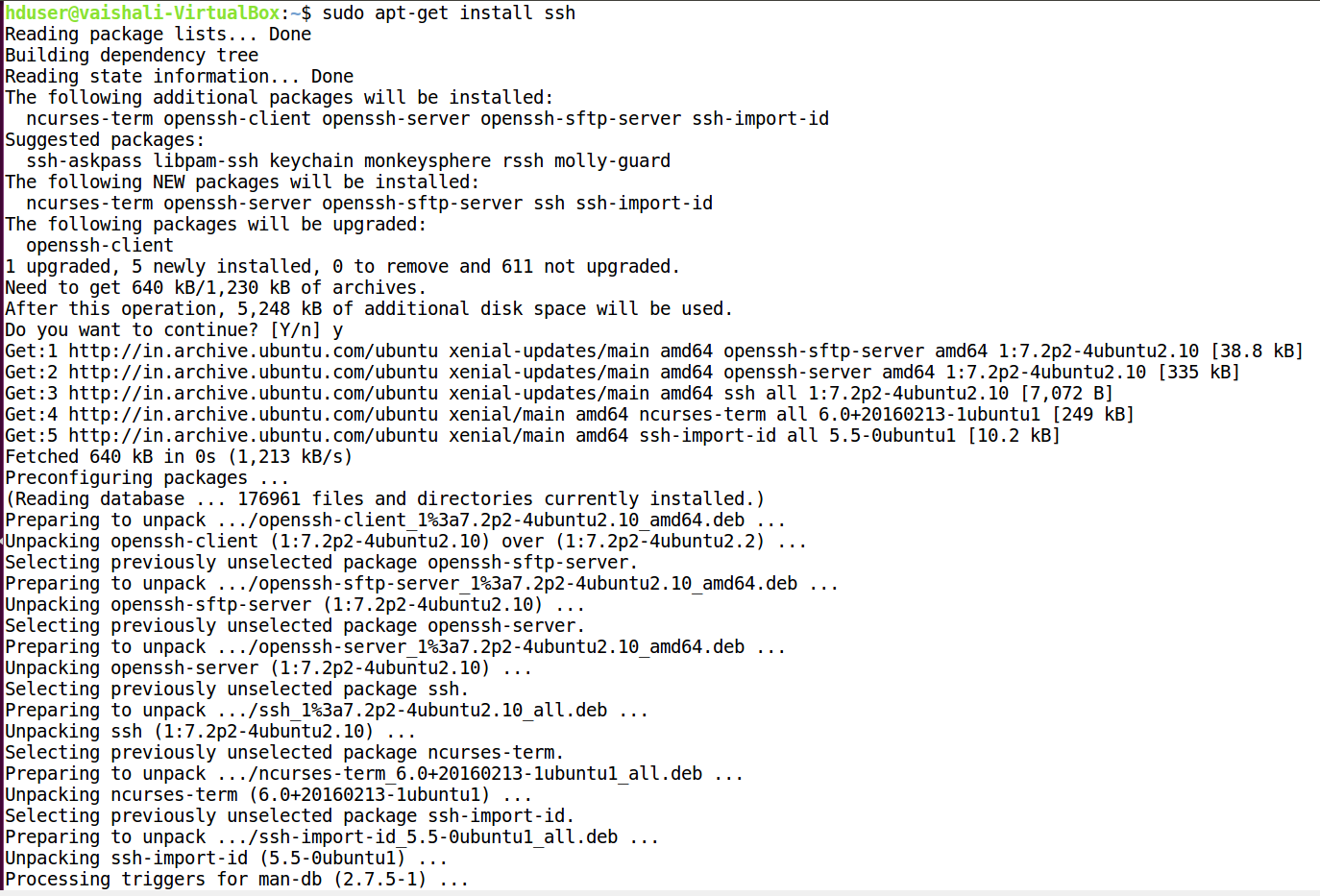


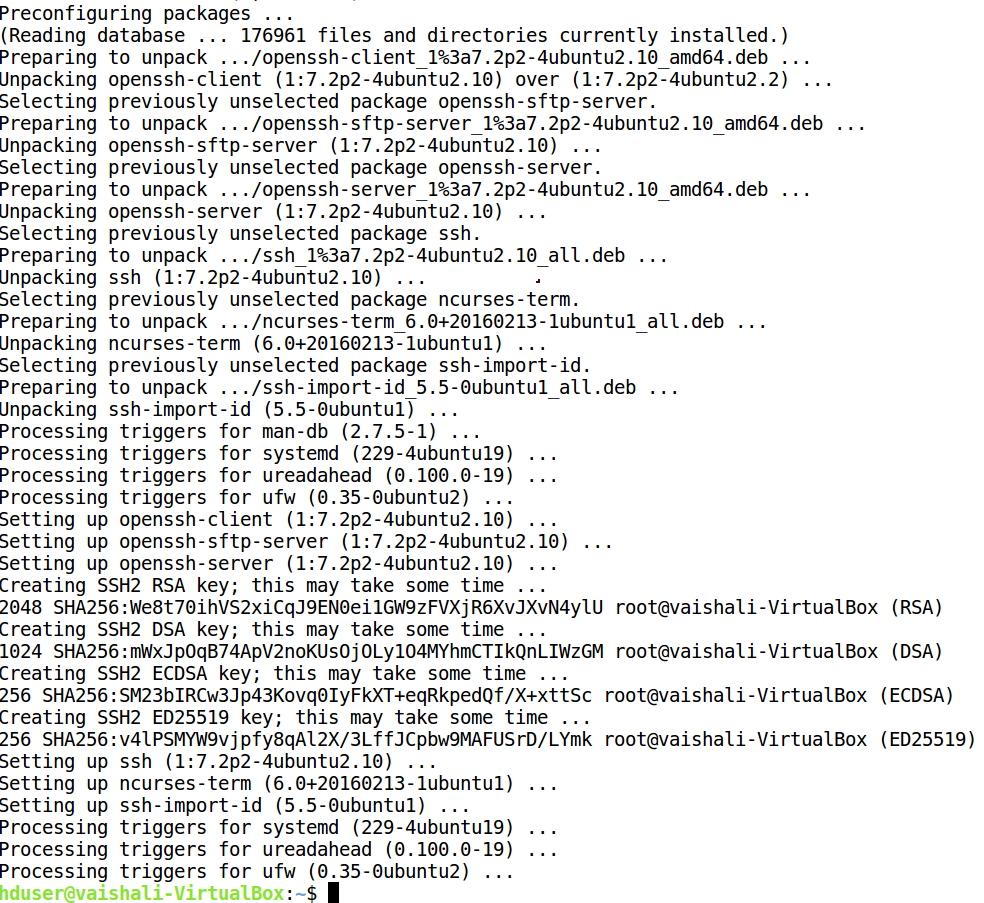
1. Logout and Login as hduser and install Java



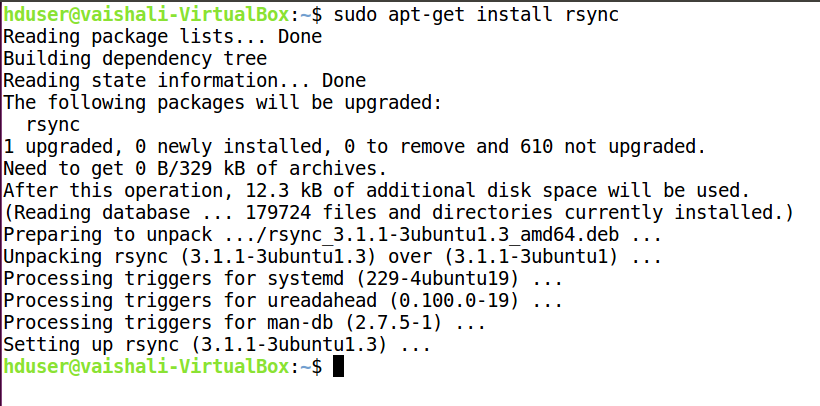
1. Installing Secure Shell

**a) sudo apt-get install ssh**

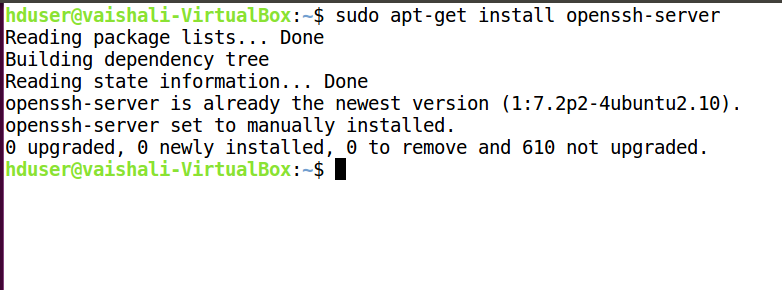




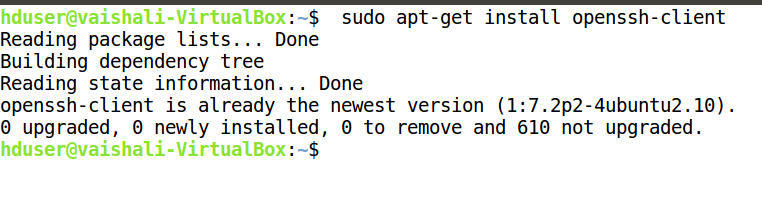
**b) sudo apt-get install rsync**



**c) sudo apt-get install openssh-server**



**d) sudo apt-get install openssh-client**



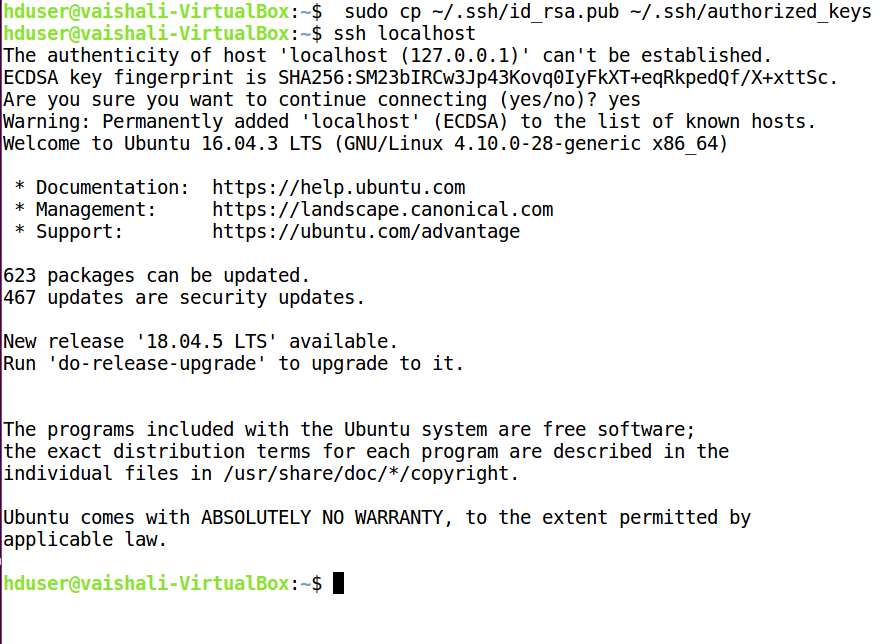
e) Generate key pair and add the public key to the authorized\_keys.

**ssh-keygen -t rsa -P '' -f ~/.ssh/id\_rsa**



**f) sudo cp ~/.ssh/id\_rsa.pub ~/.ssh/authorized\_keys** and verify SSH,

using the command **ssh localhost**



7)Installation of Hadoop 3.1.4

a) Download latest version of Apache Hadoop package

b) Extract and move the hadoop to an installation directory

]$ **sudo tar xfz hadoop-3.1.4.tar.gz**

]$ **sudo mv hadoop-3.1.4 /usr/local/hadoop**

c) Add the JAVA\_HOME to hadoop-env.sh file

]$ **sudo gedit /usr/local/hadoop/etc/hadoop/hadoop-env.sh**

d) Locate java\_home and set **JAVA\_HOME=/usr/local/java**

**e)** Add the following lines to ~/.bashrc ]$ **sudo gedit ~/.bashrc**

#HADOOP VARIABLES START

export HADOOP\_INSTALL=/usr/local/hadoop

export PATH=$PATH:$HADOOP\_INSTALL/bin

export PATH=$PATH:$HADOOP\_INSTALL/sbin

export HADOOP\_MAPRED\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_HOME=$HADOOP\_INSTALL

export HADOOP\_HDFS\_HOME=$HADOOP\_INSTALL

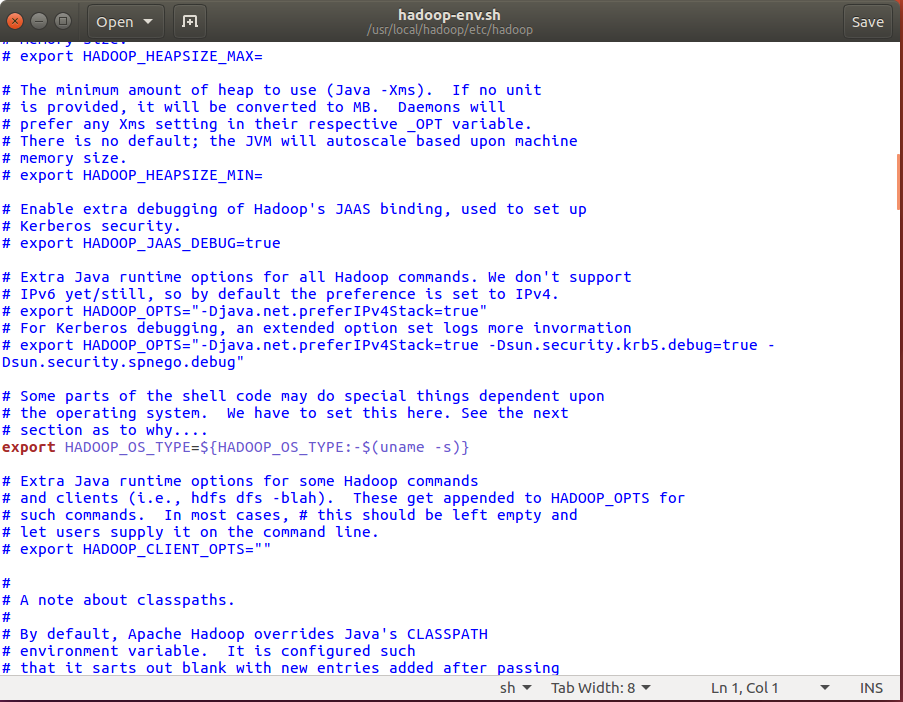
export YARN\_HOME=$HADOOP\_INSTALL

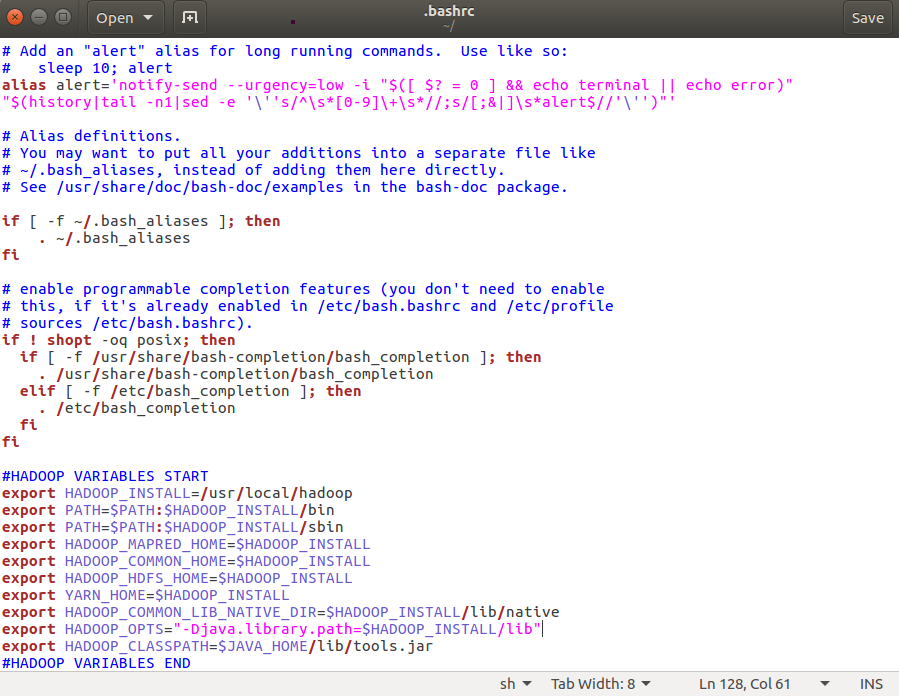
export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_INSTALL/lib/native

export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_INSTALL/lib"

export HADOOP\_CLASSPATH=$JAVA\_HOME/lib/tools.jar

#HADOOP VARIABLES END





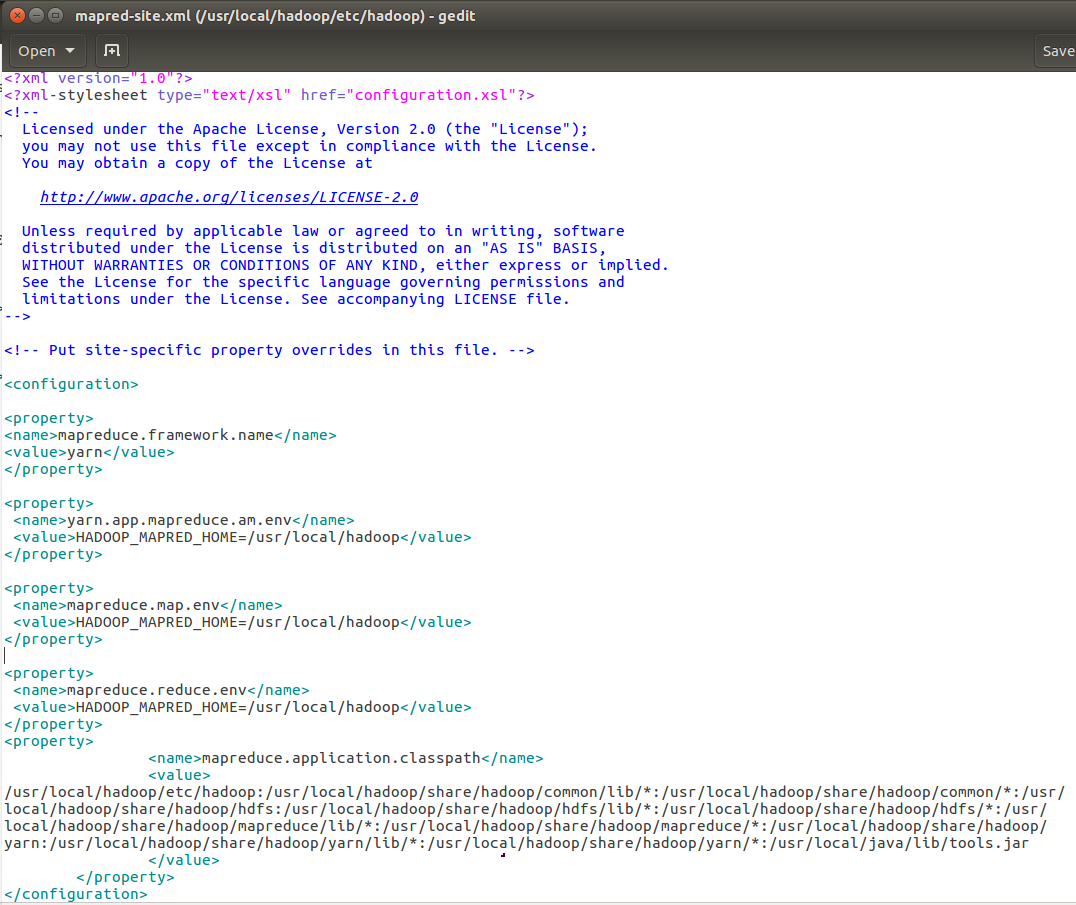
f) Refresh the bashrc file so that our environment variables can be accessed.

]$ **source ~/.bashrc**

**g)** Check the Hadoop version

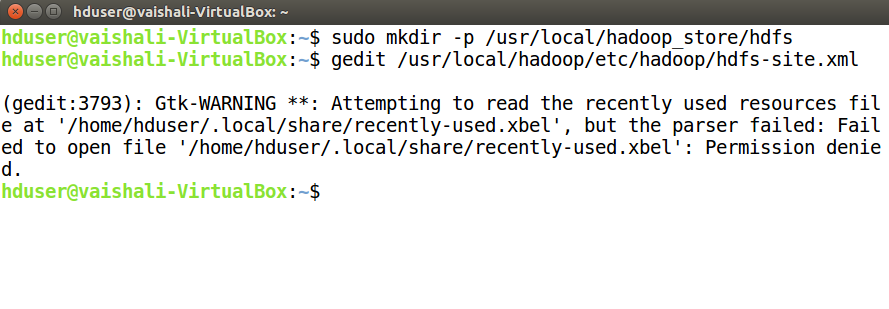


1. a) /usr/local/hadoop/etc/hadoop/mapred-site.xml



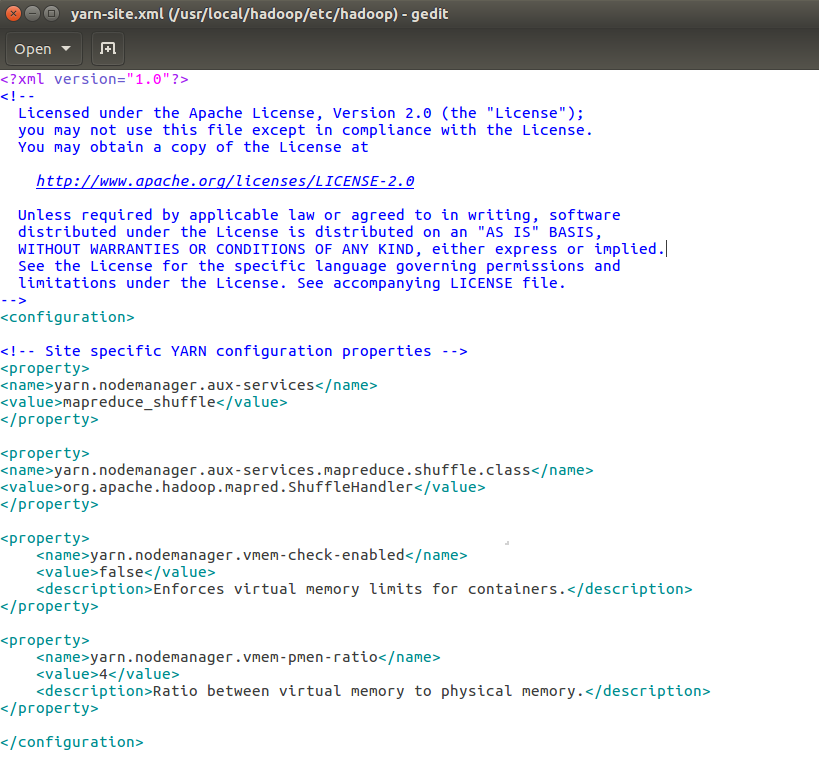
b) Create Namenode and Datanode directories

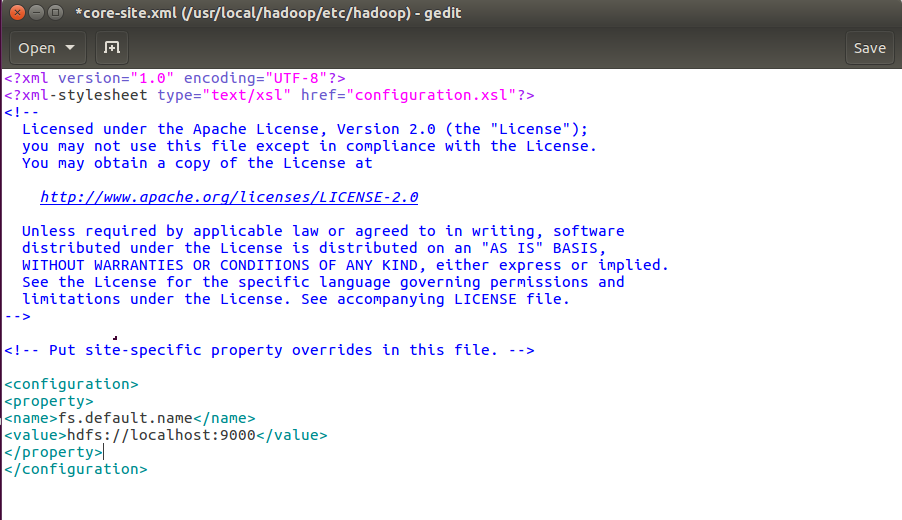
]$ sudo **mkdir -p /usr/local/hadoop\_store/hdfs**



/usr/local/hadoop/etc/hadoop/hdfs-site.xml



c) /usr/local/hadoop/etc/hadoop/yarn-site.xml

d) /usr/local/hadoop/etc/hadoop/core-site.xml

9) Change owner of hadoop\_store

]$ **sudo chown hduser:hduser -R /usr/local/Hadoop**

]$ **sudo chown hduser:hduser -R /usr/local/hadoop\_store**

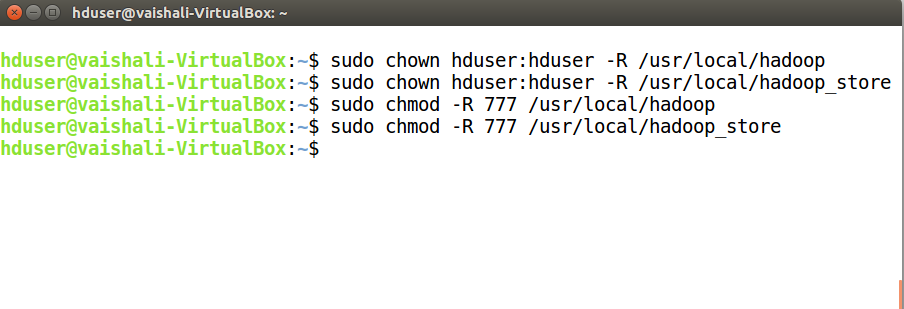
also give the folder the full permission

]$ **sudo chmod -R 777 /usr/local/hadoop**

]$ **sudo chmod -R 777 /usr/local/hadoop\_store**

Format your HDFS, make sure you have logged in as hadoop/hduser user.

]$ **hdfs namenode -format**

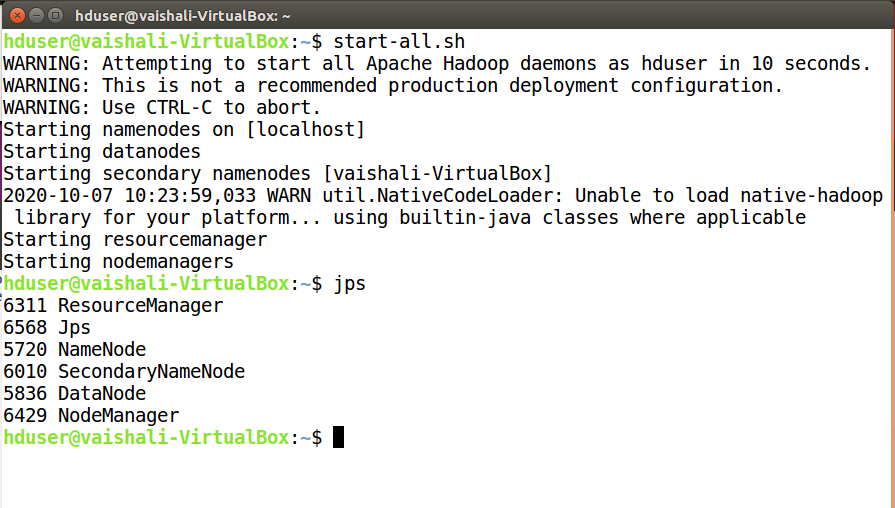


**10)** Start the hadoop dfs and yarn process

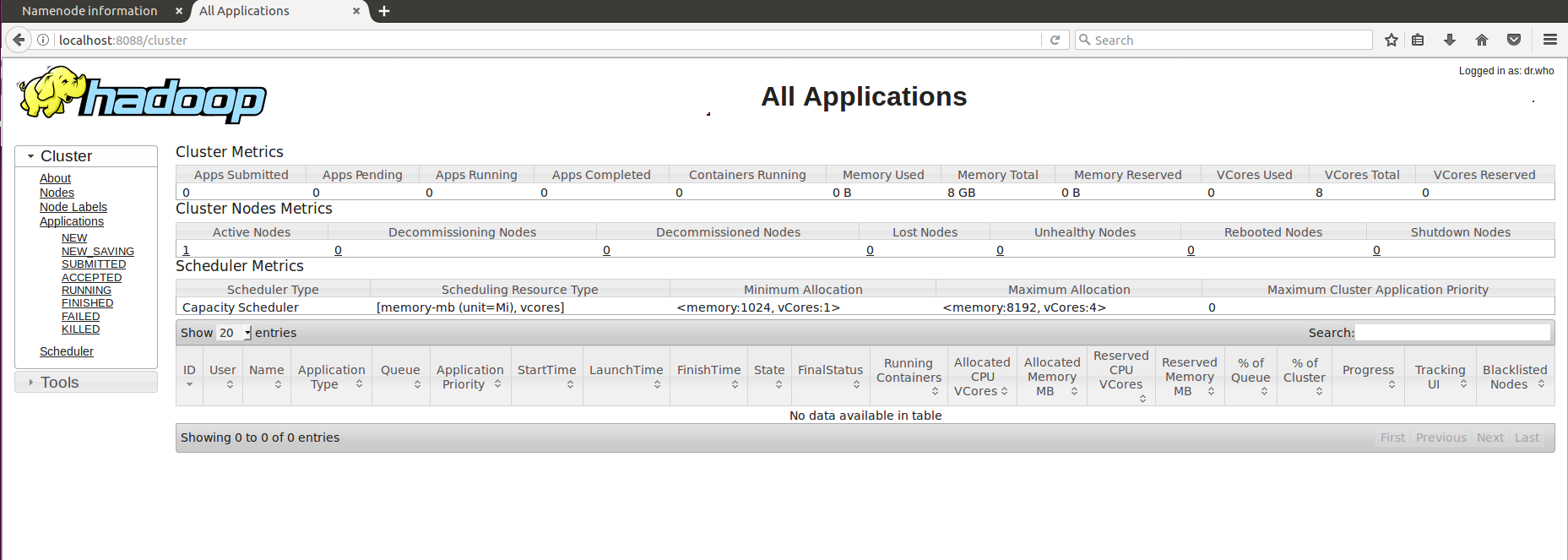
]$ **start-all.sh**

Check the hadoop deamons by running the following command

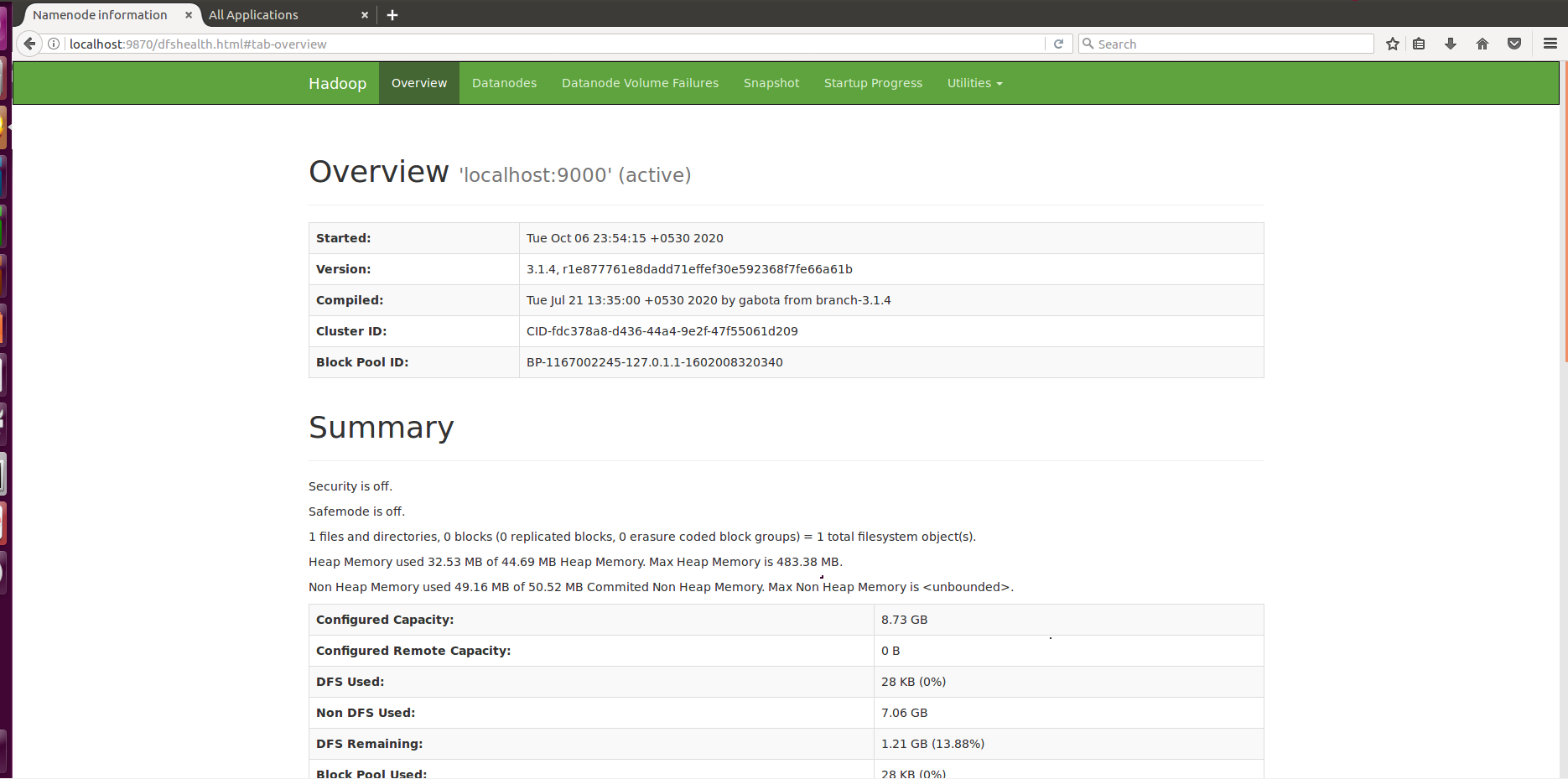
**]$ jps**



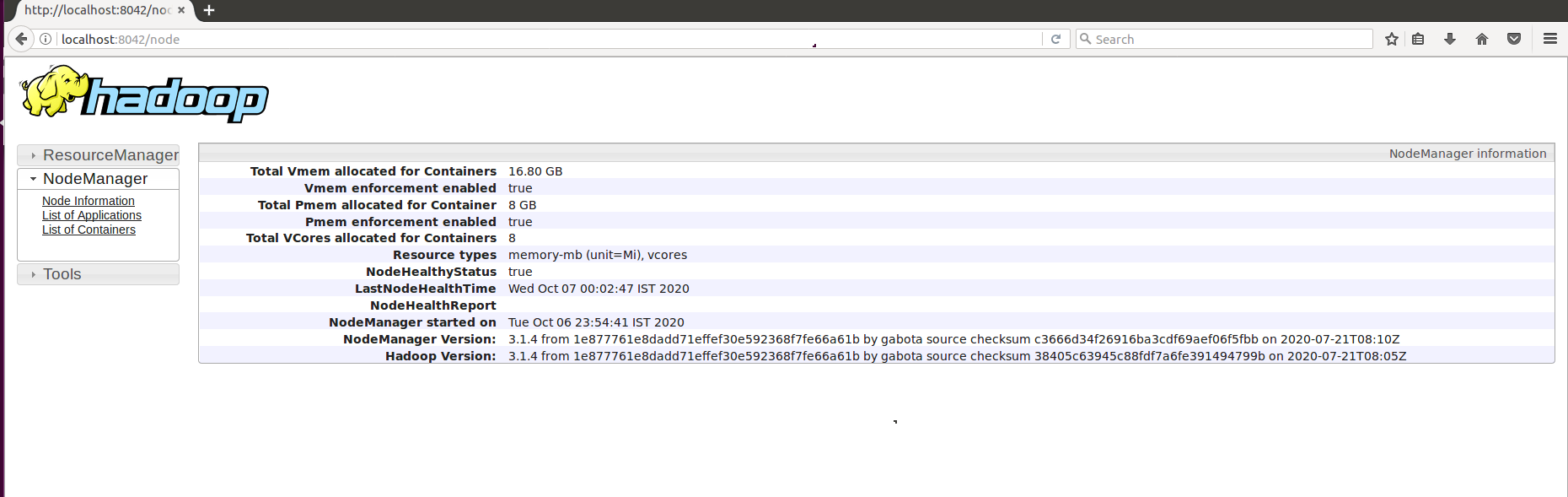
**Resource Manager - localhost:8088**



**NameNode UI http://localhost:9870/**



**YARN UI http://localhost:8042/**

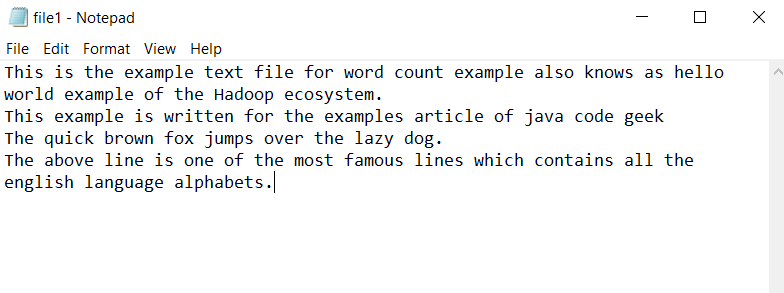


**Execution of wordcount in Hadoop**

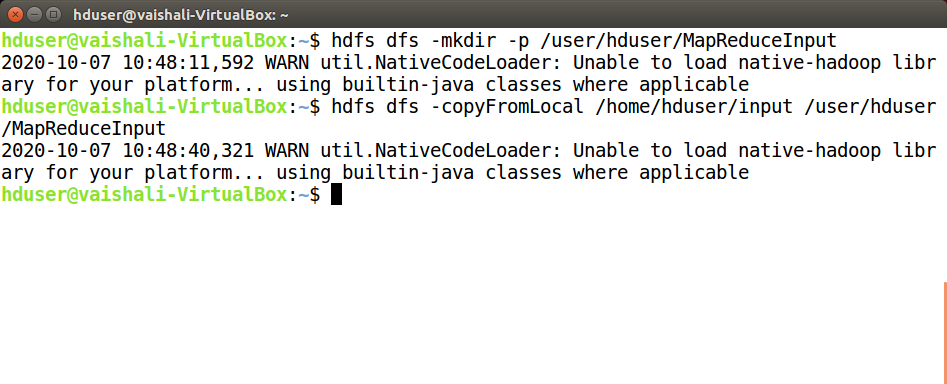
1. Managing files in HDFS.

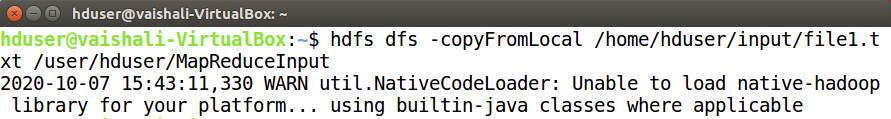
Create an input file with few sentences in it.

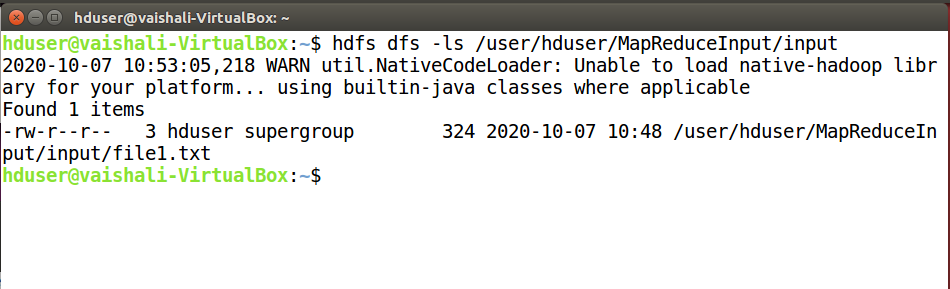
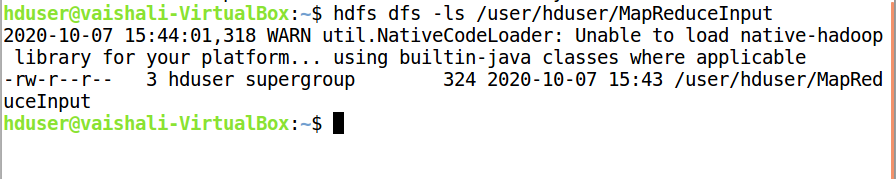
**File1.txt**



Upload input file into HDFS using put command / copyFromLocal command*.*



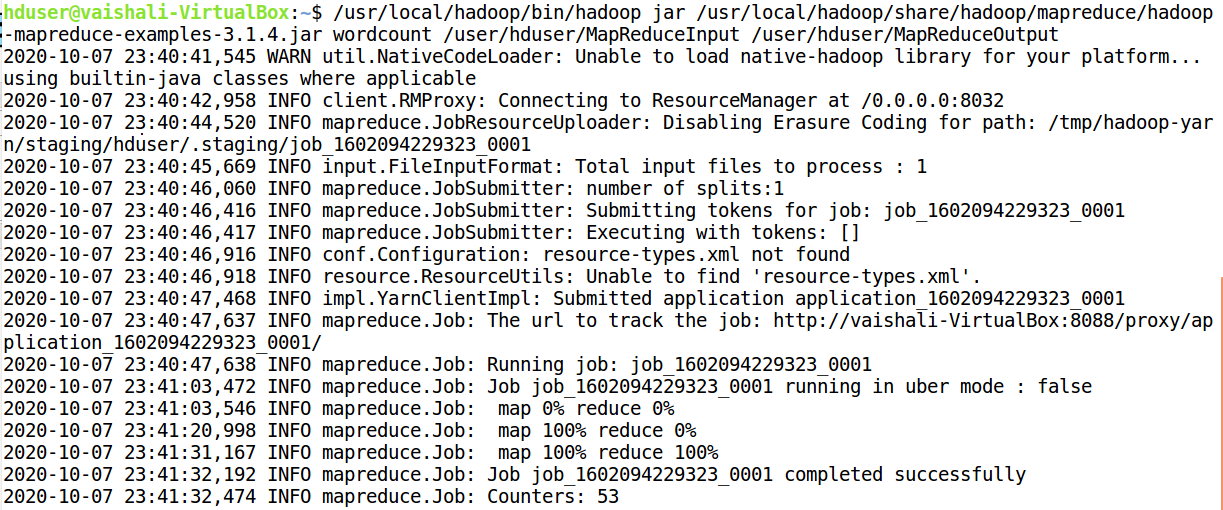


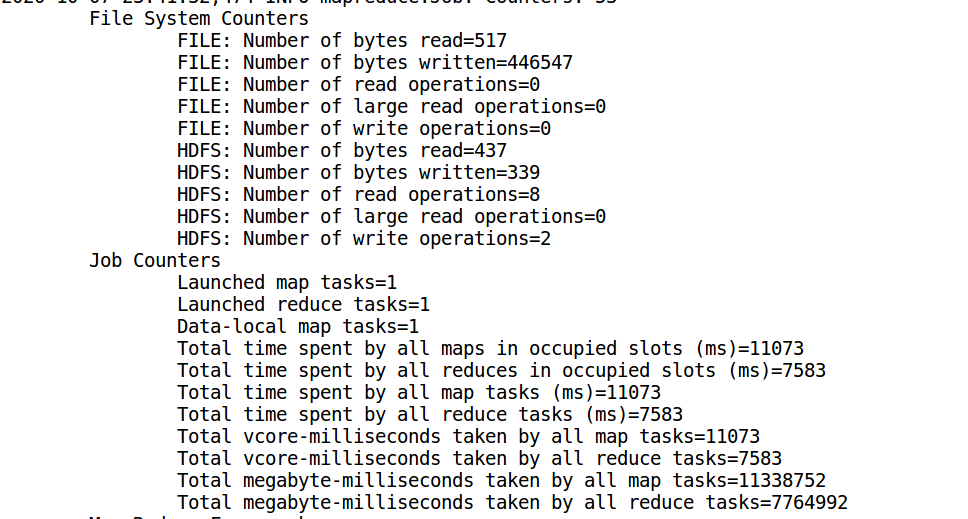


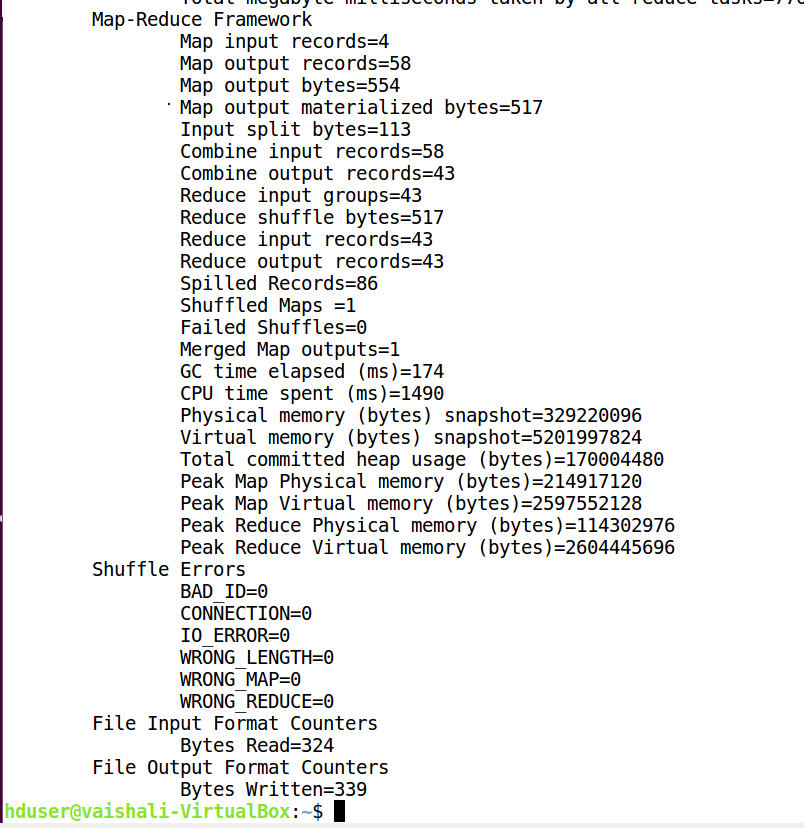
1. Write word count program in Java using Map and Reduce functions. Create

word count program into a jar file using Eclipse. Execute the word count program’s jar file using below command in hduser:

**/usr/local/hadoop/bin/hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.4.jar wordcount /user/hduser/MapReduceInput /user/hduser/MapReduceOutput**



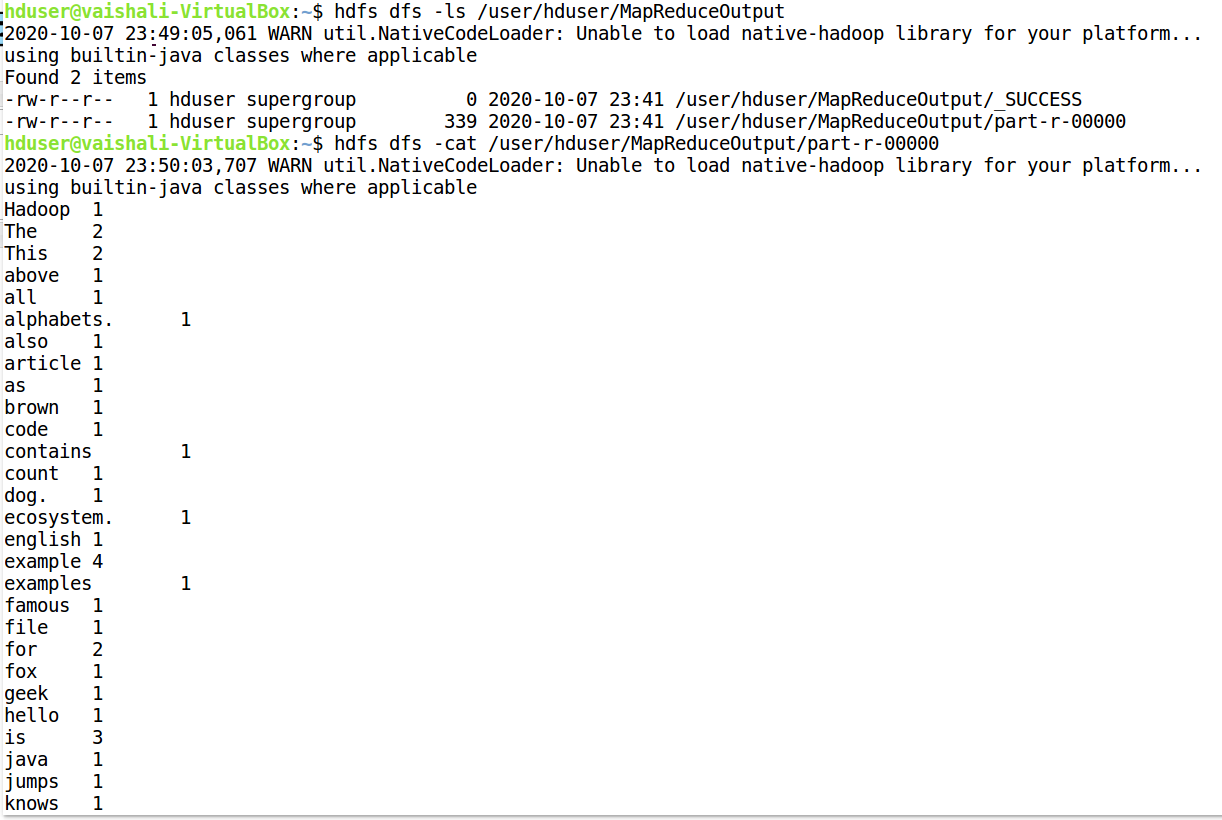




1. **OUTPUT:**

hdfs dfs -ls /user/hduser/MapReduceOutput

hdfs dfs -cat /user/hduser/MapReduceOutput/part-r-00000





1. Finally stop all Hadoop services.

