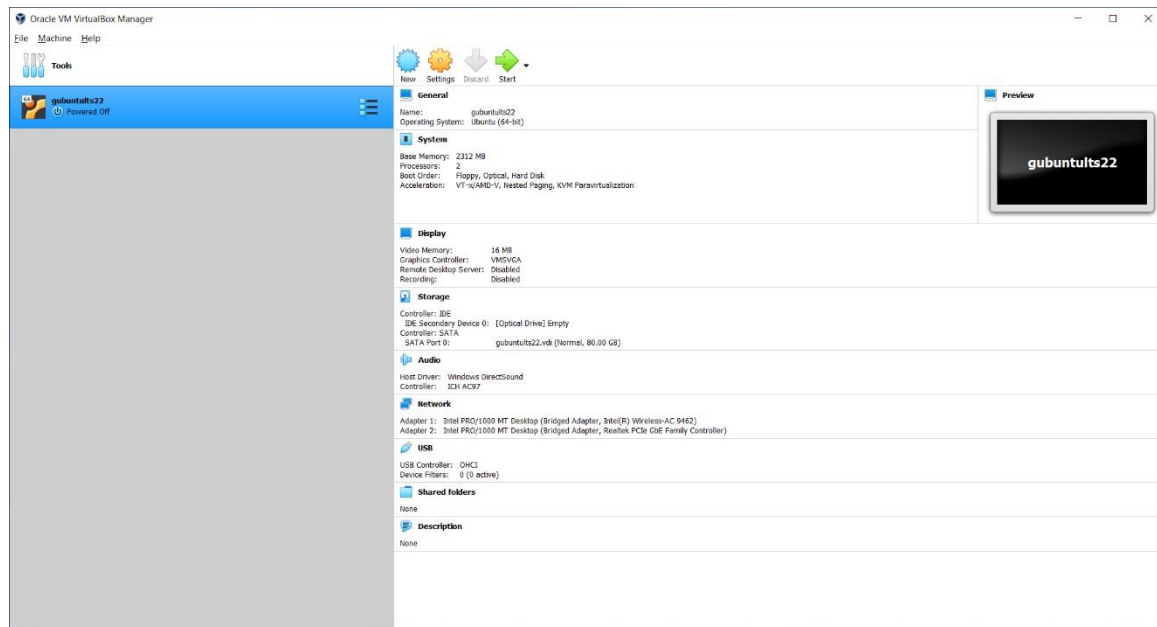
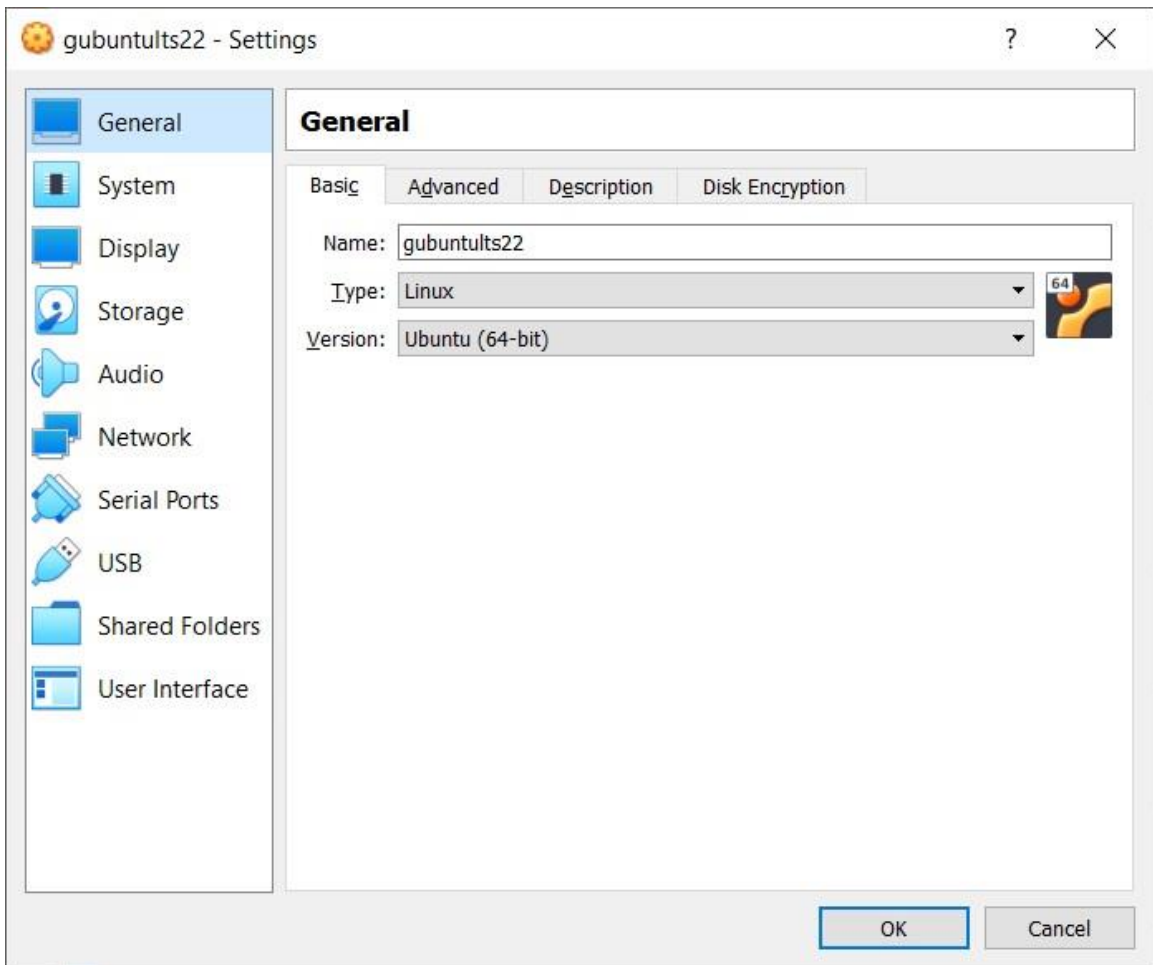
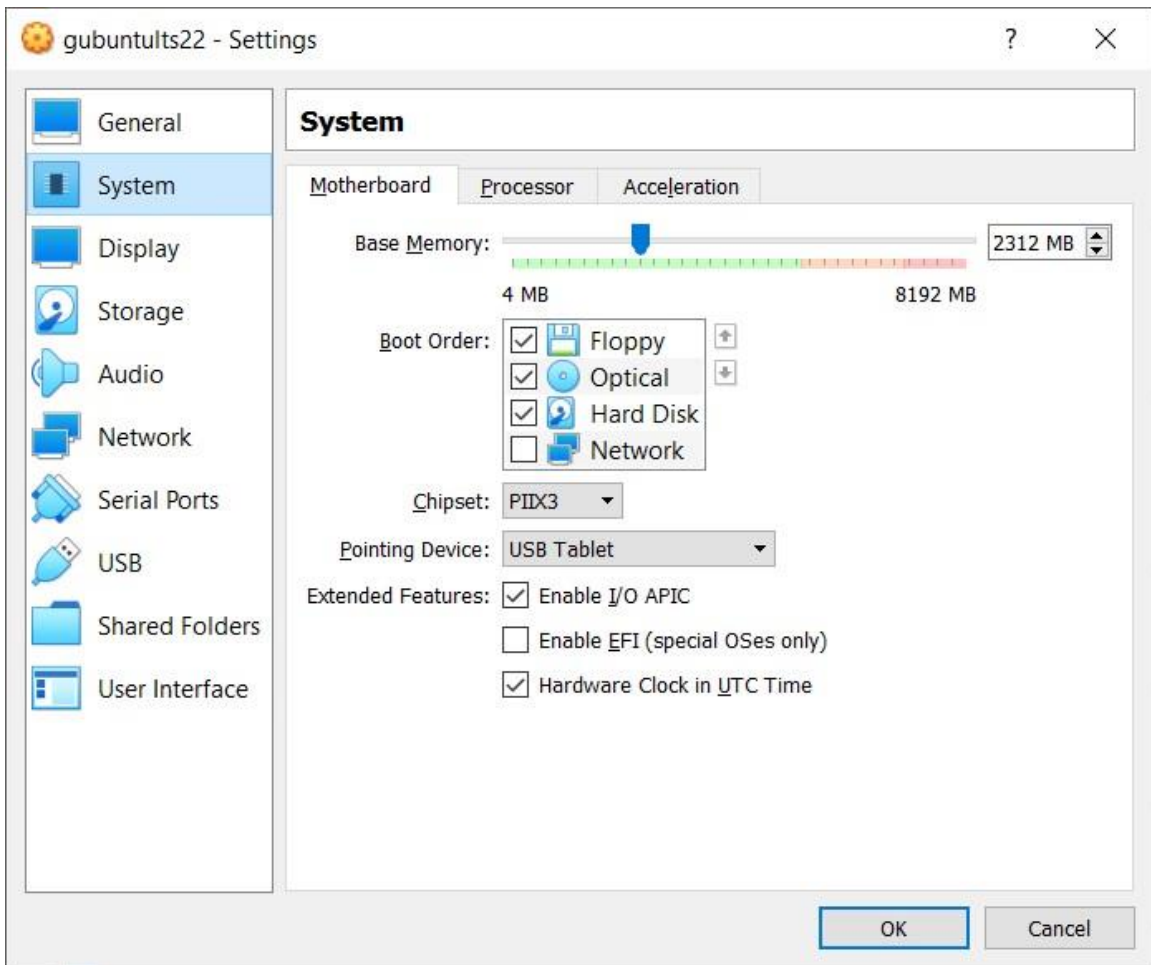


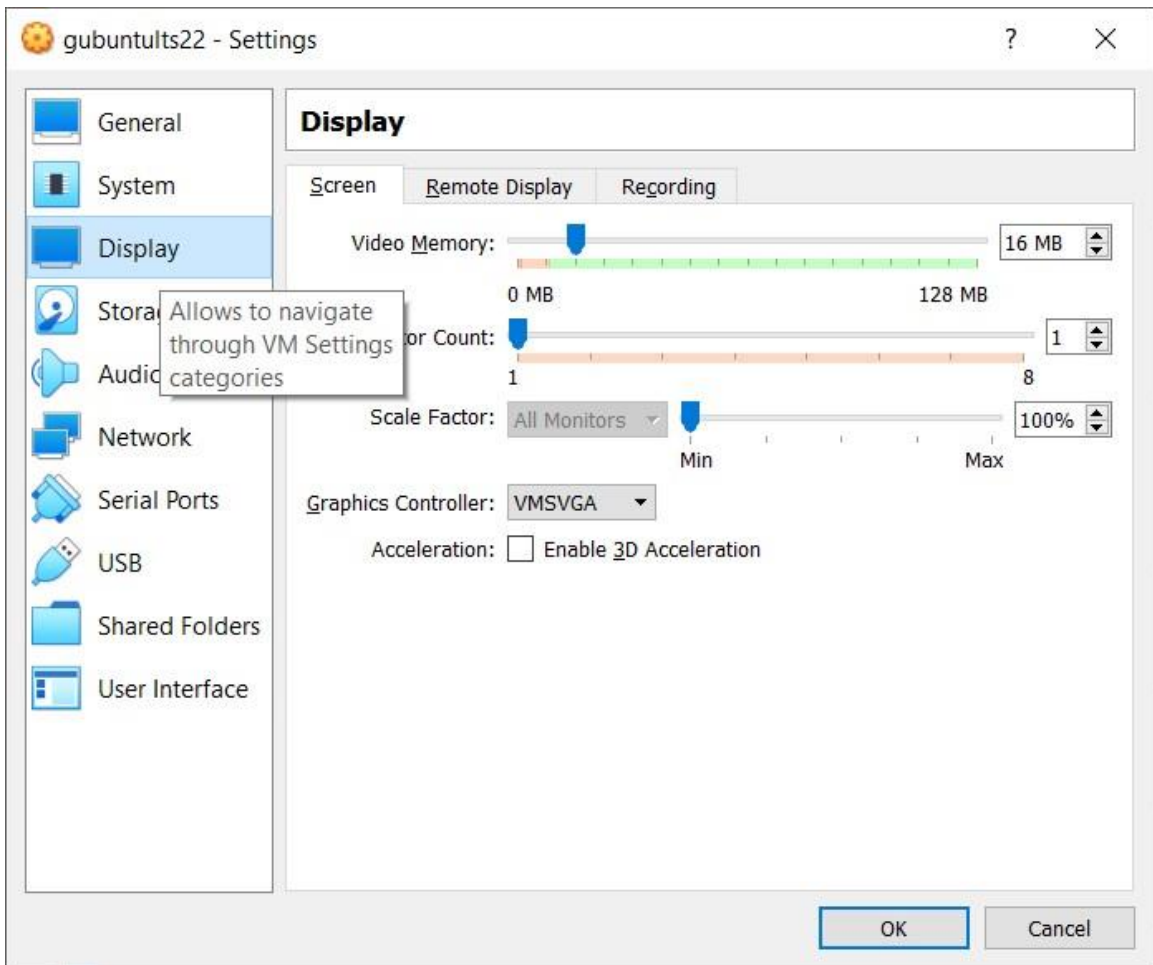
Student :
SFBU- 19599, Manickam Ravisekar - MSCS
Project : MapReduce – Pi

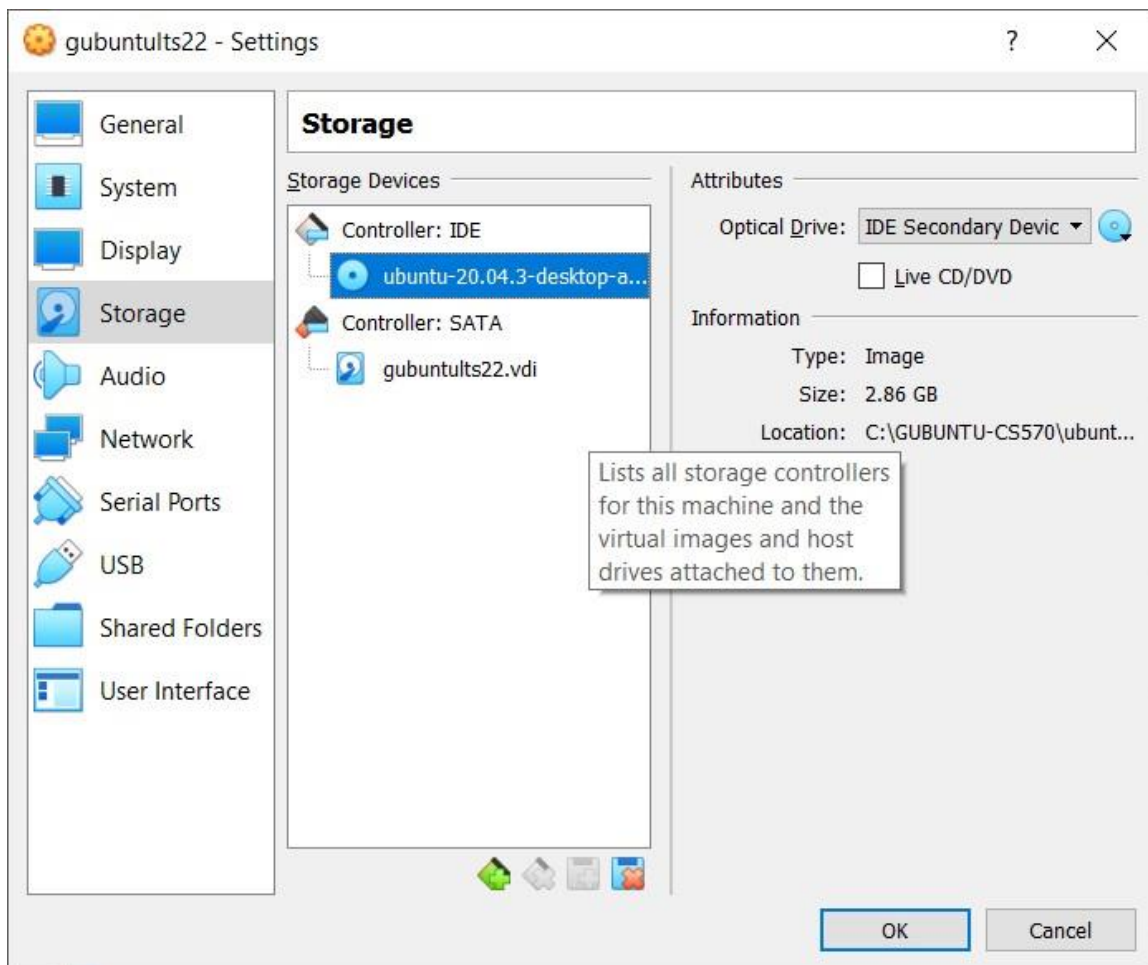
Installation Of Oracle Virtual machine on local host

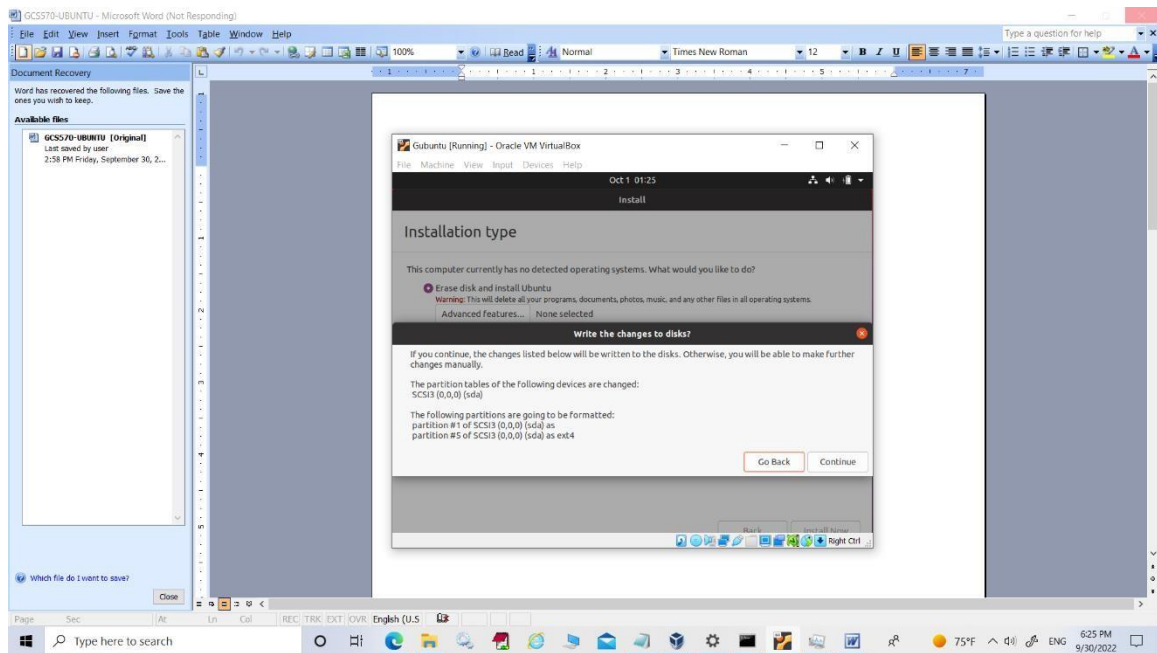













← Select start-up disk

Please select a virtual optical disk file or a physical optical drive containing a disk to start your new virtual machine from.

The disk should be suitable for starting a computer from and should contain the operating system you wish to install on the virtual machine if you want to do that now. The disk will be ejected from the virtual drive automatically next time you switch the virtual machine off, but you can also do this yourself if needed using the Devices menu.

ubuntu-22.04.1-desktop-amd64.iso (3.56 GB) 

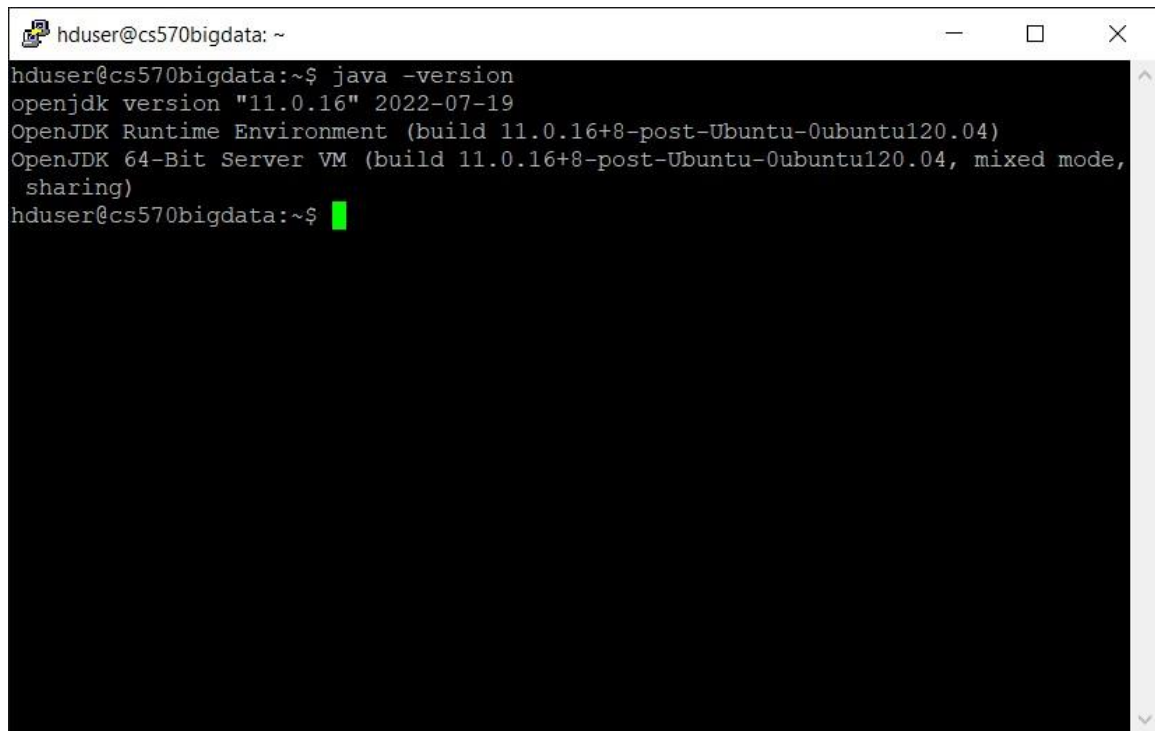
Start

Cancel

After completion of installation Oracle VM , install Java , Hadoop and the script files

Download hadoop
Download Java
Extract hadoop
Java installation steps

```
sudo apt-get install openjdk-11-jre sudo  
apt-get install openjdk-11-jdk java -  
version
```

A terminal window titled 'hduser@cs570bigdata: ~' with standard window controls. The terminal shows the command 'java -version' and its output: 'openjdk version "11.0.16" 2022-07-19', 'OpenJDK Runtime Environment (build 11.0.16+8-post-Ubuntu-0ubuntu120.04)', and 'OpenJDK 64-Bit Server VM (build 11.0.16+8-post-Ubuntu-0ubuntu120.04, mixed mode, sharing)'. The prompt 'hduser@cs570bigdata:~\$' is followed by a green cursor.

```
hduser@cs570bigdata:~$ java -version  
openjdk version "11.0.16" 2022-07-19  
OpenJDK Runtime Environment (build 11.0.16+8-post-Ubuntu-0ubuntu120.04)  
OpenJDK 64-Bit Server VM (build 11.0.16+8-post-Ubuntu-0ubuntu120.04, mixed mode,  
sharing)  
hduser@cs570bigdata:~$
```

Setup hadoop user for Hadoop Installation

```
sudo addgroup hadoop
```

```
sudo adduser --ingroup hadoop hduser
```

```
sudo su hduser
```

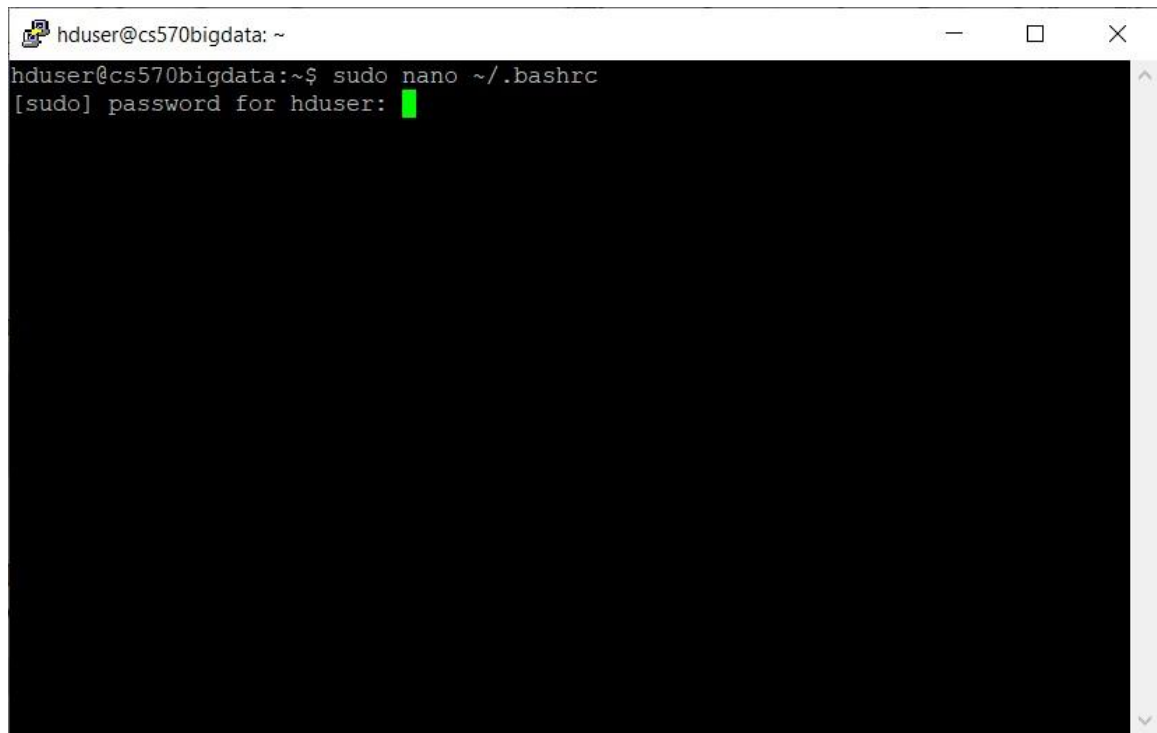
For this I used hadoop version hadoop-2.10.2.tar.gz

```
Sudo tar vxzf hadoop-2.10.2.tar.gz -C /usr/local
```

Cd / usr/local

Sudo mv hadoop-2.10.2 hadoop

Sudo chown -R hduser:hadoop hadoop

A terminal window with a title bar showing 'hduser@cs570bigdata: ~'. The terminal content shows the command 'hduser@cs570bigdata:~\$ sudo nano ~/.bashrc' followed by a prompt '[sudo] password for hduser:' and a green cursor. The terminal window has standard Linux window controls (minimize, maximize, close) in the top right corner.

```
hduser@cs570bigdata: ~
hduser@cs570bigdata:~$ sudo nano ~/.bashrc
[sudo] password for hduser: 
```



```
hduser@cs570bigdata: ~
GNU nano 4.8 /home/hduser/.bashrc
# ~/.bashrc: executed by bash(1) for non-login shells.
# see /usr/share/doc/bash/examples/startup-files (in the package bash-doc)
# for examples

# If not running interactively, don't do anything
case $- in
  *i*) ;;
  *) return;;
esac

# don't put duplicate lines or lines starting with space in the history.
# See bash(1) for more options
HISTCONTROL=ignoreboth

# append to the history file, don't overwrite it
shopt -s histappend

# for setting history length see HISTSIZE and HISTFILESIZE in bash(1)
HISTSIZE=1000
HISTFILESIZE=2000

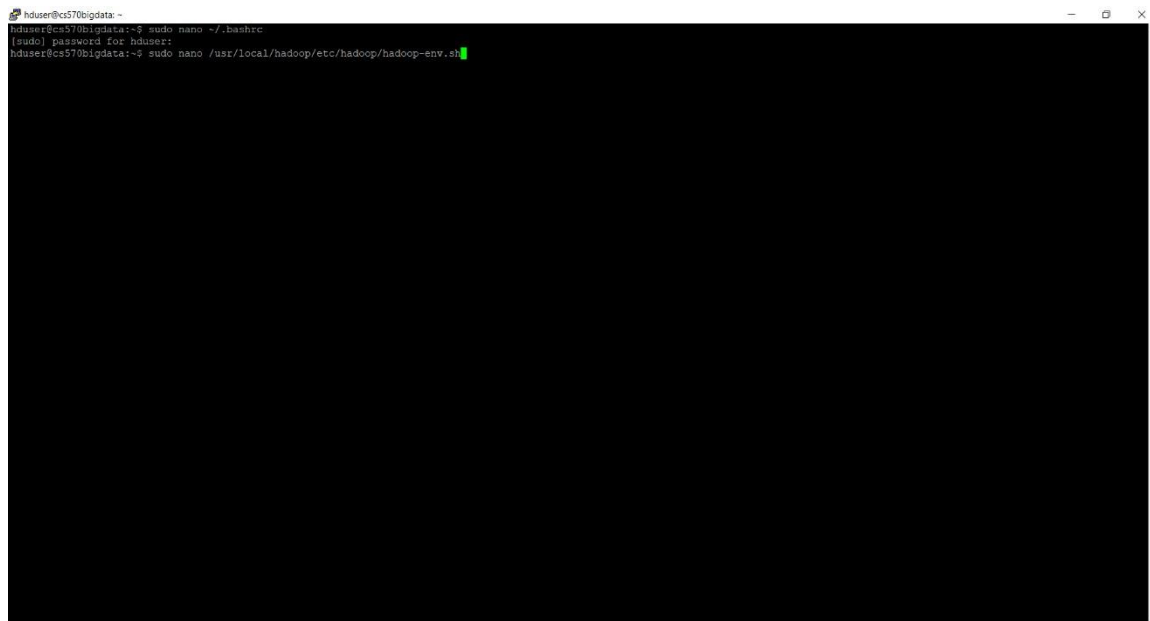
[ Read 128 lines ]
^G Get Help  ^O Write Out  ^W Where Is   ^K Cut Text   ^J Justify    ^C Cur Pos
^X Exit       ^R Read File  ^\ Replace    ^U Paste Text ^T To Spell   ^_ Go To Line
```

Move to the end and add following lines for hadoop

```
hduser@cs570bigdata: ~
GNU nano 4.8 /home/hduser/.bashrc
# 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/jdk/
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=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib"
```

To save use source ~/.bashrc
systemctl reboot -i

Now do the following setting for hduser

A terminal window screenshot showing a series of commands being executed. The prompt is 'hduser@cs570bigdata:~'. The first command is 'sudo nano ~/.bashrc', followed by a password prompt '(sudo) password for hduser:'. The second command is 'sudo nano /usr/local/hadoop/etc/hadoop/hadoop-env.sh', followed by a password prompt. The terminal content is as follows:

```
hduser@cs570bigdata:~  
hduser@cs570bigdata:~$ sudo nano ~/.bashrc  
(sudo) password for hduser:  
hduser@cs570bigdata:~$ sudo nano /usr/local/hadoop/etc/hadoop/hadoop-env.sh
```

export JAVA_HOME=/usr/lib/jvm/jdk

Ssh generation and creation of authorized keys from public ssh keys

```
ravisekar@ravisekar-VirtualBox: ~/.ssh
ravisekar@ravisekar-VirtualBox:~$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/ravisekar/.ssh/id_rsa):
Created directory '/home/ravisekar/.ssh'.
Your identification has been saved in /home/ravisekar/.ssh/id_rsa
Your public key has been saved in /home/ravisekar/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:ED7NPhgVsiLpRactOK5zhpB7zFqjTHcizmVj77Dw+Lw ravisekar@ravisekar-VirtualBox
The key's randomart image is:
+---[RSA 3072]-----+
|  . + o. |
| + = B   |
| = = B o  |
| + + o *  |
| o o . S  |
|.*        |
|+=+OO .   |
|=@@.B     |
|o=.Eoo    |
+---[SHA256]-----+
ravisekar@ravisekar-VirtualBox:~$ ls -al
total 88
drwxr-x--- 17 ravisekar ravisekar 4096 Sep 30 13:08 .
drwxr-xr-x  3 root      root      4096 Sep 29 21:12 ..
```

```
ravisekar@ravisekar-VirtualBox: ~/.ssh
ravisekar@ravisekar-VirtualBox:~$ cd .ssh
ravisekar@ravisekar-VirtualBox:~/.ssh$ ls -al
total 16
drwx----- 2 ravisekar ravisekar 4096 Sep 30 13:08 .
drwxr-x--- 17 ravisekar ravisekar 4096 Sep 30 13:08 ..
-rw----- 1 ravisekar ravisekar 2622 Sep 30 13:08 id_rsa
-rw-r--r-- 1 ravisekar ravisekar 584 Sep 30 13:08 id_rsa.pub
ravisekar@ravisekar-VirtualBox:~/.ssh$
```

Now we need following files to be set for hadoop :

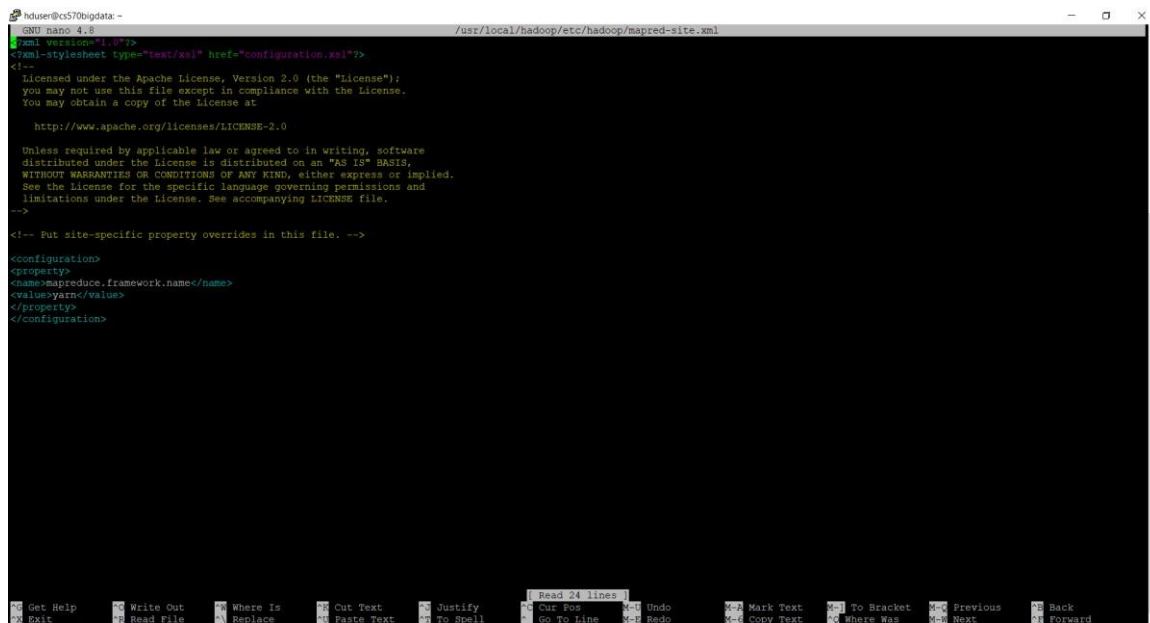
1 . sudo nano /usr/local/hadoop/etc/hadoop/core-site.xml

```
hduser@cs570bigdata: -
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<!-- Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at
http://www.apache.org/licenses/LICENSE-2.0
Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<!-- Put site-specific property overrides in this file. -->
<configuration>
<property><name>fs.default.name</name>
<value>hdfs://cs570bigdata:9000</value>
</property>
</configuration>
```

2. sudo nano /usr/local/hadoop/etc/hadoop/yarn-site.xml

```
hduser@cs570bigdata: -
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/yarn-site.xml
<?xml version="1.0"?>
<!-- Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at
http://www.apache.org/licenses/LICENSE-2.0
Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<configuration>
<!-- Site specific YARN configuration properties -->
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
<value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
</configuration>
```

3. sudo nano /usr/local/hadoop/etc/hadoop/mapred-site.xml



```
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/mapred-site.xml
<?xml version="1.0"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

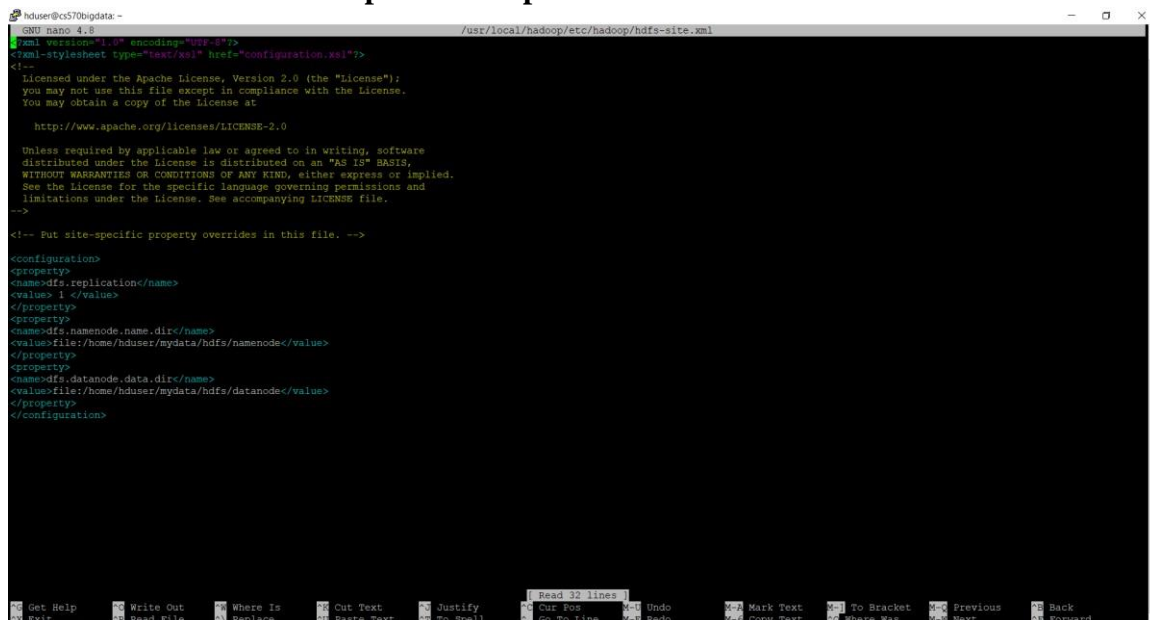
4 For doing hdfs.xml

First complete this task's mkdir

–p mydata/hdfs/namenode mkdir

–p mydata/hdfs/datanode

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



```
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

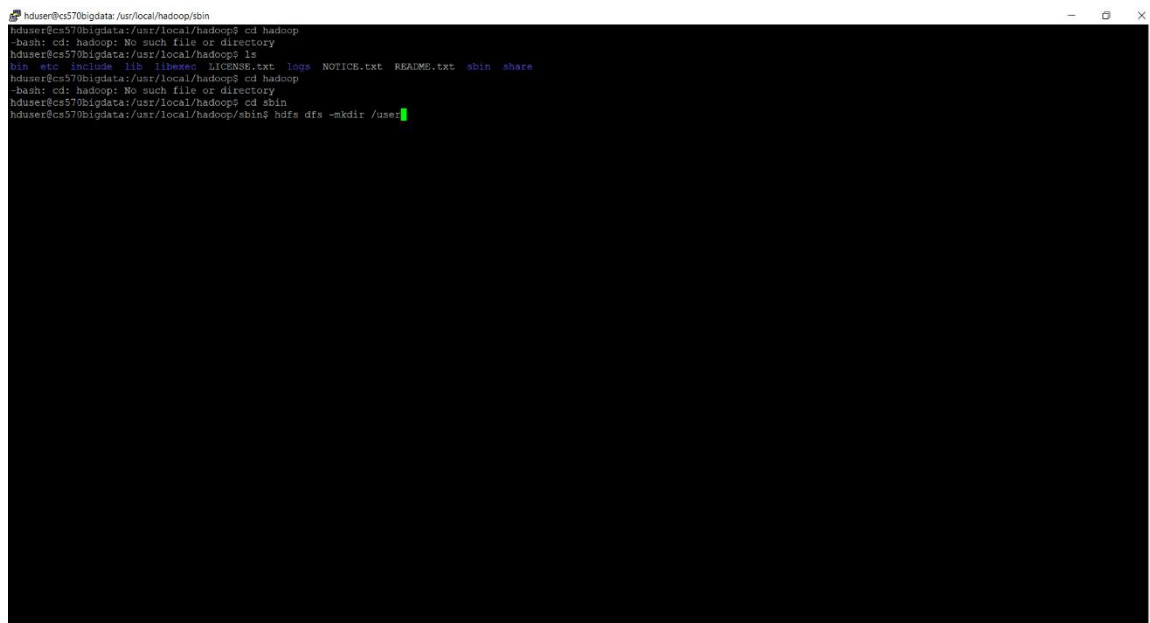
<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>hdfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>hdfs.namenode.name.dir</name>
    <value>file:/home/hduser/mydata/hdfs/namenode</value>
  </property>
  <property>
    <name>hdfs.datanode.data.dir</name>
    <value>file:/home/hduser/mydata/hdfs/datanode</value>
  </property>
</configuration>
```

Create following test data file for hdfs

```
hduser@cs570bigdata: ~/Desktop/inputdata
hduser@cs570bigdata:~/Desktop/inputdata$ cat test.txt
hello world world hello hello
hduser@cs570bigdata:~/Desktop/inputdata$
```

Create directory as mentioned



```
hduser@cs570bigdata: /usr/local/hadoop/sbin
hduser@cs570bigdata: /usr/local/hadoop$ cd hadoop
hduser@cs570bigdata: /usr/local/hadoop$ cd hadoop
hduser@cs570bigdata: /usr/local/hadoop$ ls
bin  etc  include  lib  libexec  LICENSE.txt  logs  NOTICE.txt  README.txt  sbin  share
hduser@cs570bigdata: /usr/local/hadoop$ cd hadoop
hduser@cs570bigdata: /usr/local/hadoop$ cd hadoop
hduser@cs570bigdata: /usr/local/hadoop$ cd sbin
hduser@cs570bigdata: /usr/local/hadoop/sbin$ hdfs dfs -mkdir /user
```

Now copy to file created for input

```
/usr/local/hadoop/bin/hdfs dfs -put '/home/hduser/Desktop/inputdata' /user
```

Next final step for hadoop

```
hdfs namenode -format
```

now everything is set we can start hadoop and run the jar for the word count and get the output as shown in below screen

now start hadoop as shown below by running start-dfs.sh

start-yarn.sh and run jps to check the status of nodes as shown in below screens

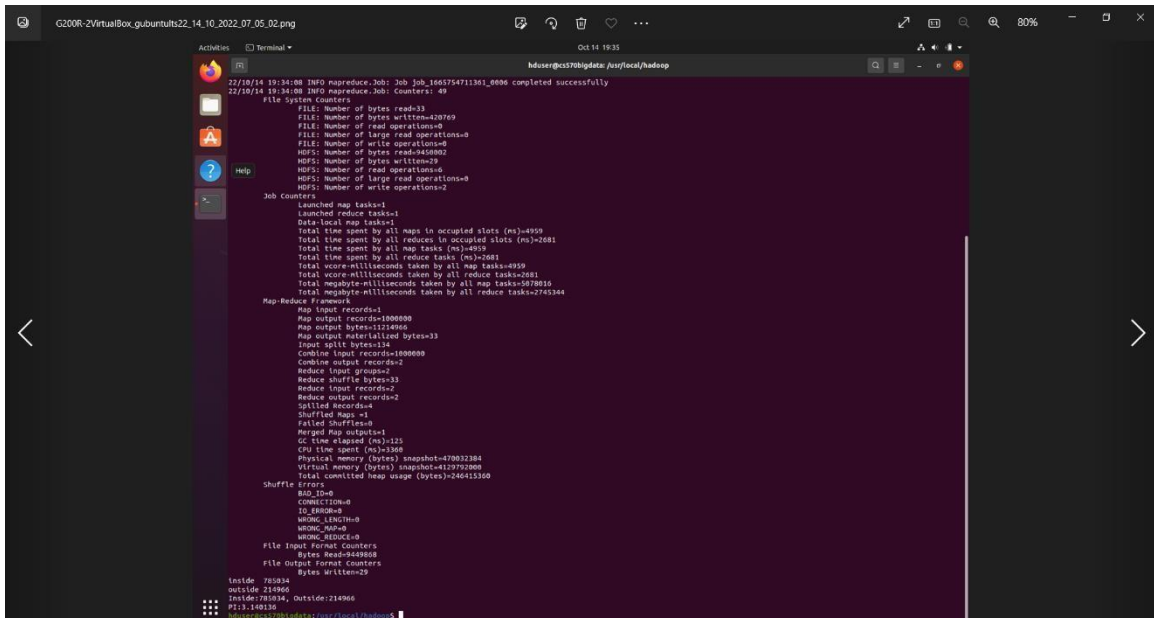
```
Oct 14 18:23
hduser@cs570bigdata:~$ start-dfs.sh
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.10.2.jar)
to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
22/10/14 18:22:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting namenodes on [cs570bigdata]
cs570bigdata: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-cs570bigdata.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-cs570bigdata.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-secondarynamenode-cs570bigdata.out
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.10.2.jar)
to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
22/10/14 18:22:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@cs570bigdata:~$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resourcemanager-cs570bigdata.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-nodemanager-cs570bigdata.out
hduser@cs570bigdata:~$ jps
2672 Jps
2864 DataNode
2245 SecondaryNameNode
2616 ResourceManager
2476 ResourceManager
1917 NameNode
hduser@cs570bigdata:~$
```

Procedure to test mapreduce pi program :

1. Compile the java program(PiCalculation.java)
2. create the jar file for the class generated (picalculation.jar)
3. run the below command for your input data

bin/hadoop jar /home/hduser/gHW/picalculation.jar PiCalculation
/user/hduser/inputdata/PiCalculationInput

**Pi Value Display screen (bottom of 2nd page) output:
One Million and Radius - 200
Random numbers – 1000000 Test
output :**



G200r-output-VirtualBox_gubuntults22_14_10_2022_07_05_56.png

Activities Terminal

Oct 14 19:35

hduser@cs170bigdata: /usr/local/hadoop

```
hduser@cs170bigdata: /usr/local/hadoop$ bin/hadoop dfs -cat /user/hduser/outputwc/part-r-00000
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.10.2.jar)
to method sun.security.krb5.config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
22/10/14 19:35:46 main willNativeCodeLoader: unable to load native-hadoop library for your platform... using builtin-java classes where applicable
inside: 765834
outside: 215866
hduser@cs170bigdata: /usr/local/hadoop$
```

