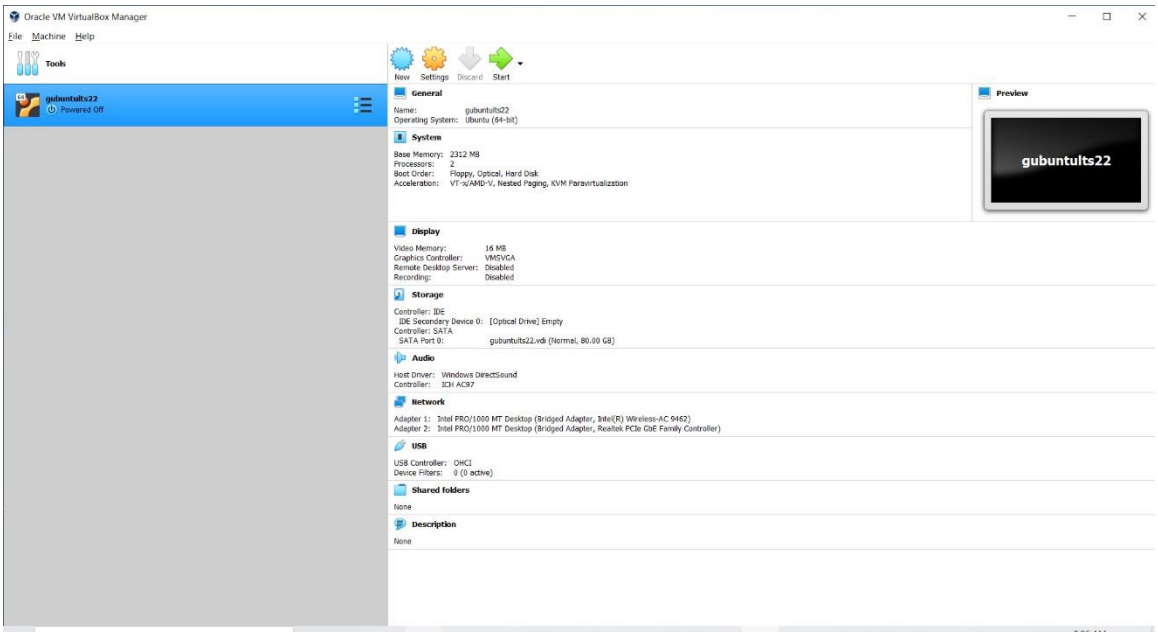
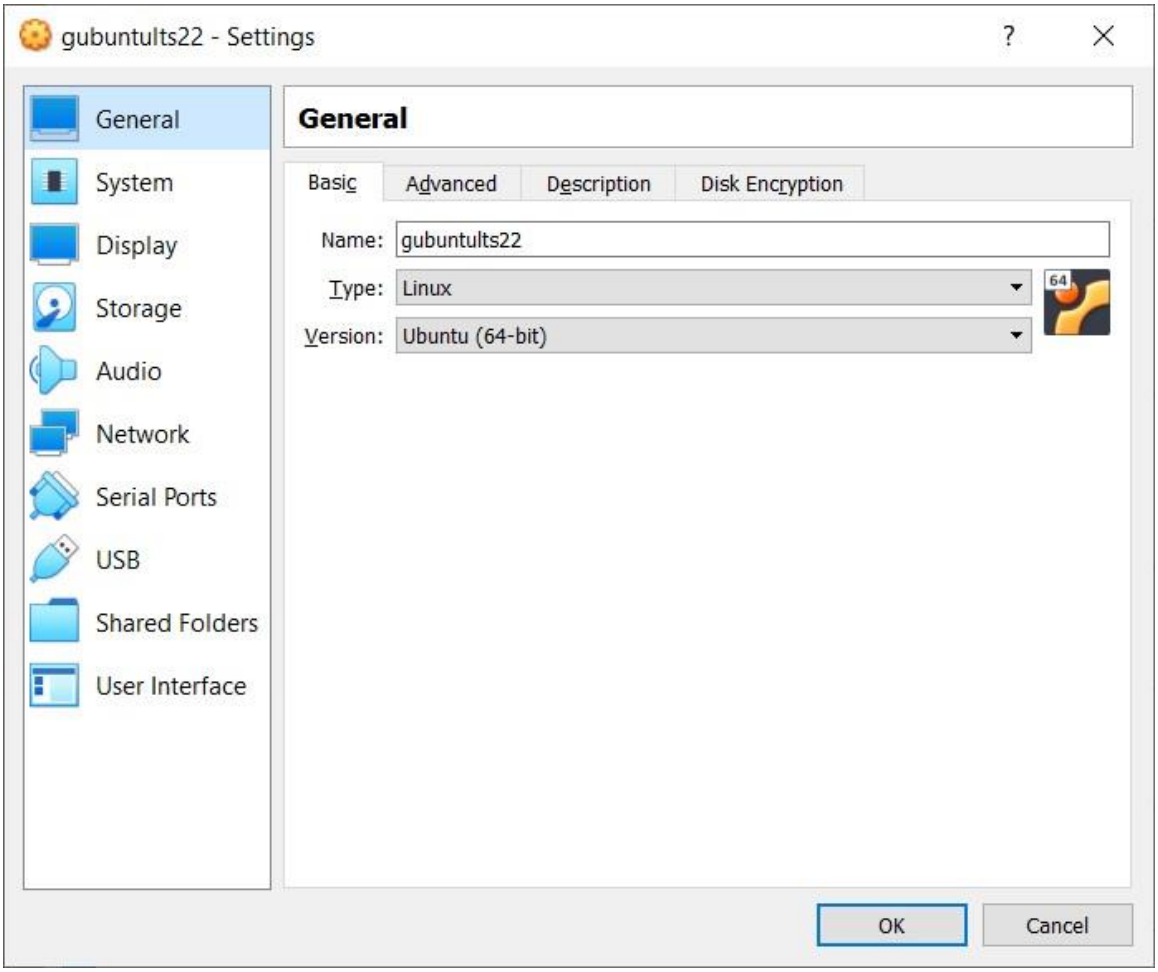
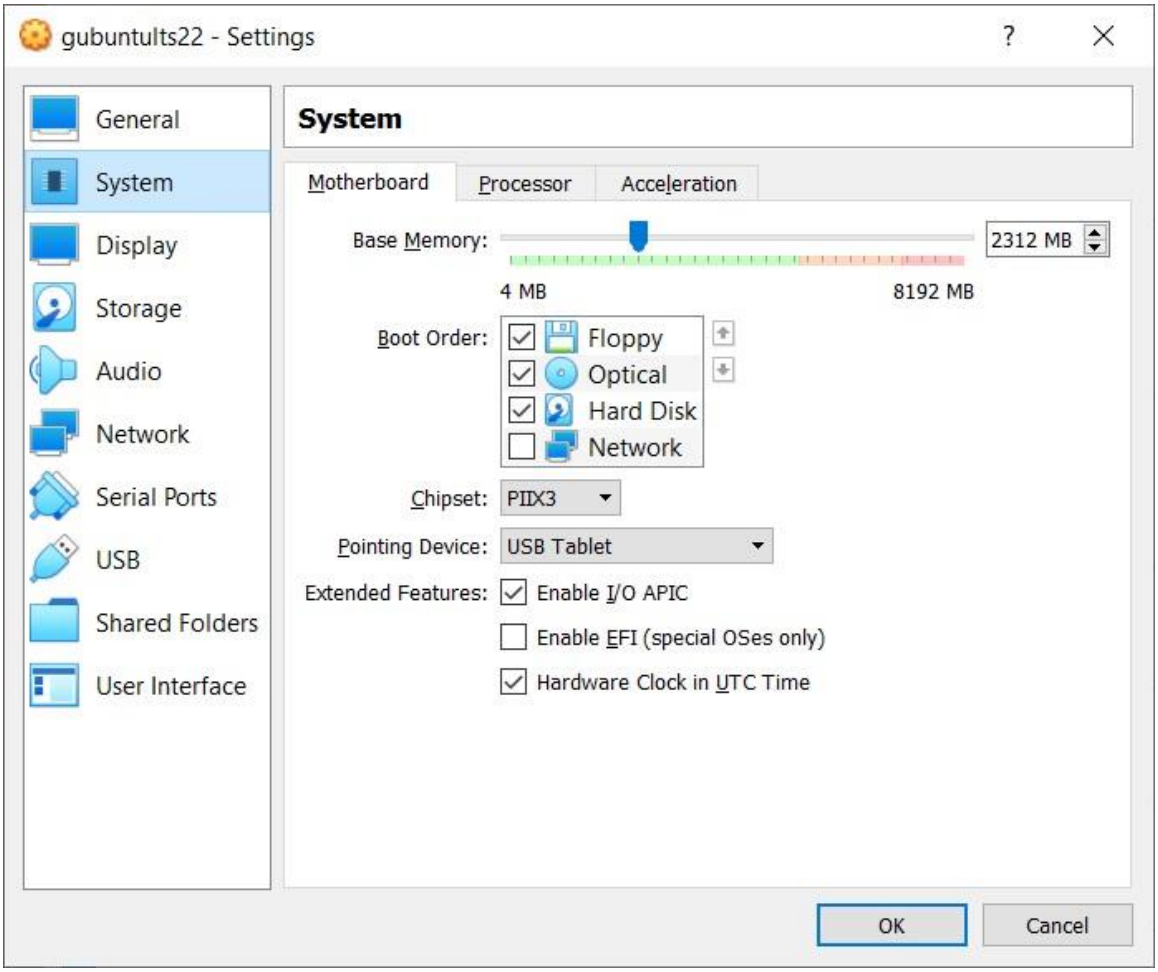
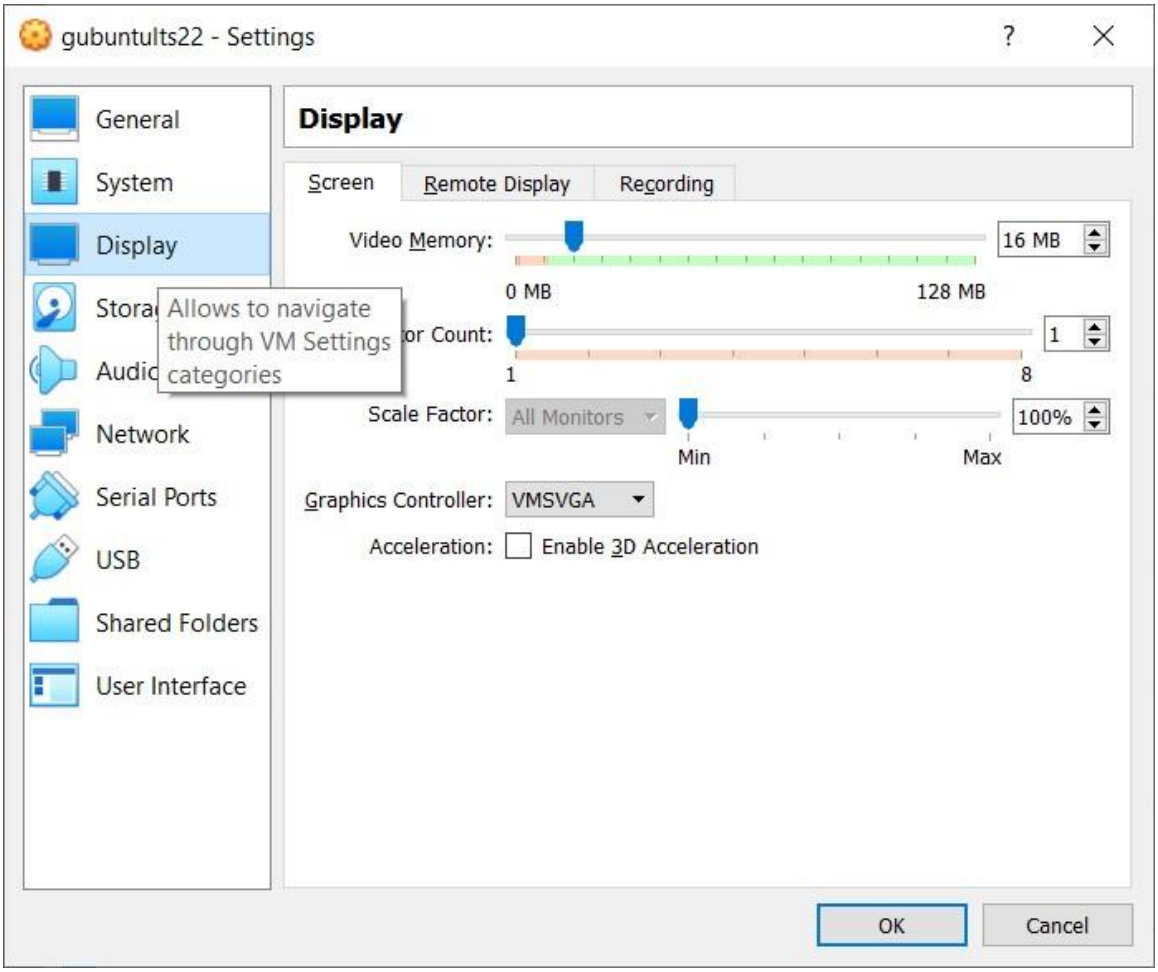
Student :
SFBU- 19599, Manickam Ravisekar - MSCS
Project : MapReduce – Pi

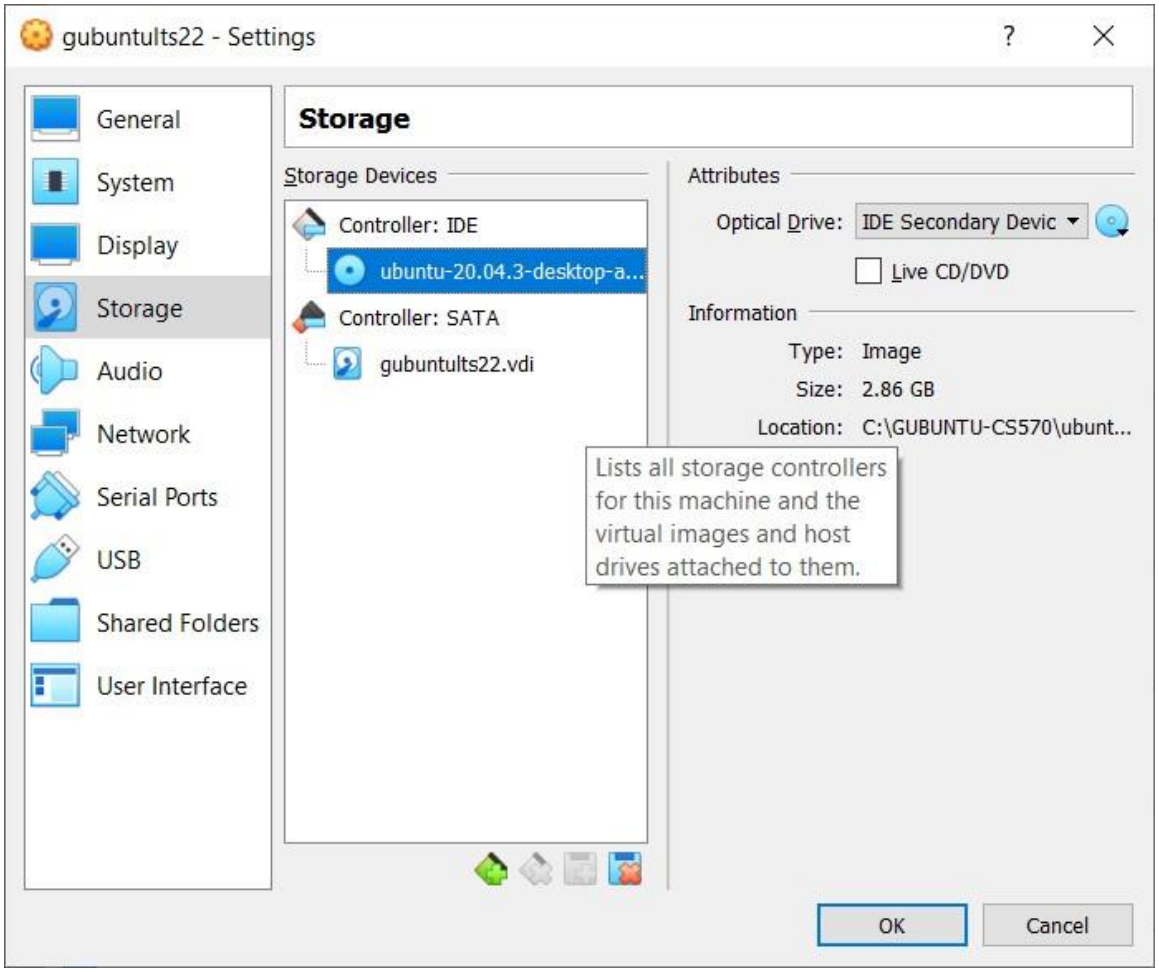
Installation Of Oracle Virtual machine on local host

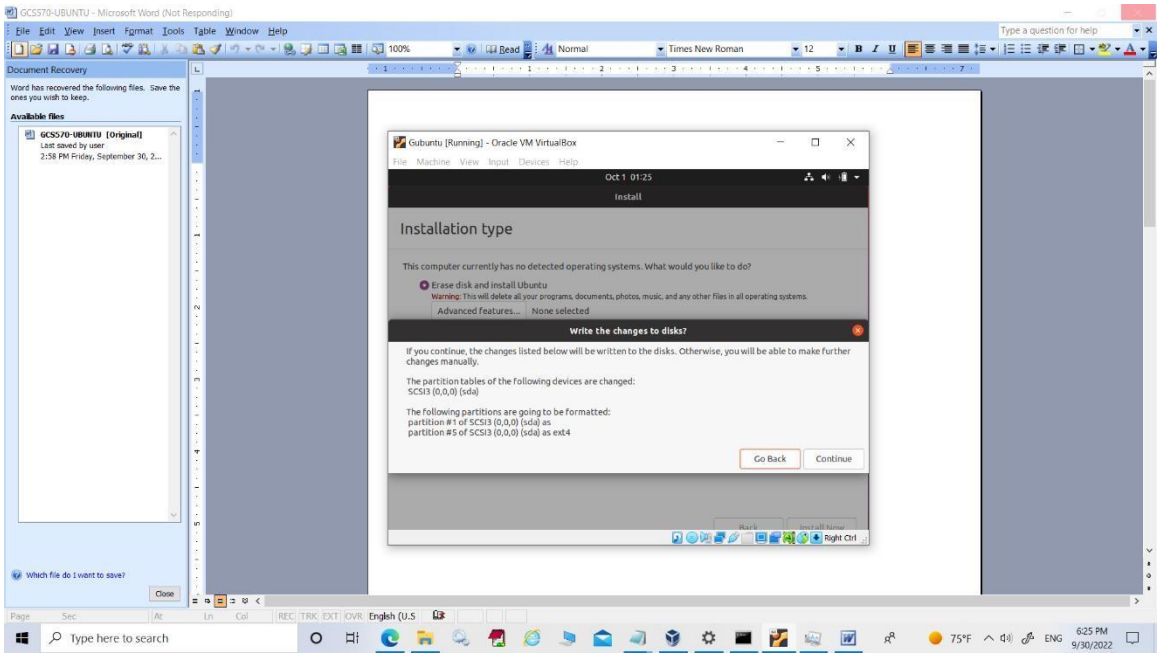


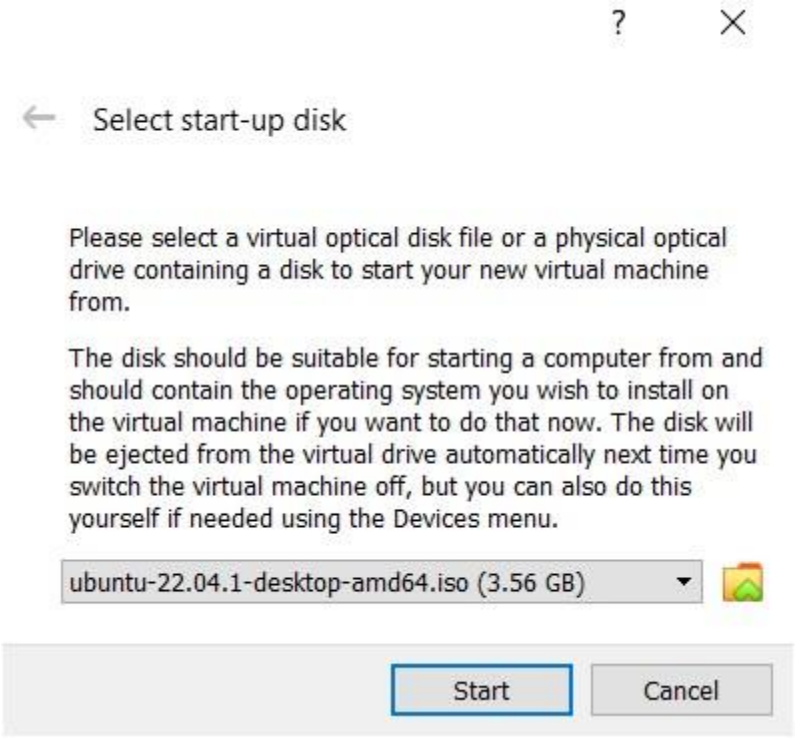

gubuntults22 - Settings
?
General
System
Display
Storage
Audio
Network
Serial Ports
USB
Shared Folders
User Interface
General
Basic
Advanced
Description
Disk Encryption
Name:
gubuntults22
Type:
Linux
Version:
Ubuntu (64-bit)
64
OK
Cancel


gubuntults22 - Settings
General
System
Display
Storage
Audio
Network
Serial Ports
USB
Shared Folders
User Interface
System
Motherboard
Processor
Acceleration
Base Memory:
2312 MB
4 MB
8192 MB
Boot Order:
Floppy
Optical
Hard Disk
Network
Chipset:
PIIX3
Pointing Device:
USB Tablet
Extended Features:
Enable I/O APIC
Enable EFI (special OSes only)
Hardware Clock in UTC Time
OK
Cancel


gubuntults22 - Settings
General
System
Display
Stora
Audic
Network
Serial Ports
USB
Shared Folders
User Interface
Allows to navigate through VM Settings categories
Display
Screen
Remote Display
Recording
Video Memory:
16 MB
0 MB
128 MB
or Count:
1
1
8
Scale Factor:
All Monitors
100%
Min
Max
Graphics Controller:
VMSVGA
Acceleration:
Enable 3D Acceleration
OK
Cancel


gubuntults22 - Settings
General
System
Display
Storage
Audio
Network
Serial Ports
USB
Shared Folders
User Interface
Storage
Storage Devices
Controller: IDE
ubuntu-20.04.3-desktop-a...
Controller: SATA
gubuntults22.vdi
Lists all storage controllers for this machine and the virtual images and host drives attached to them.
Attributes
Optical Drive: IDE Secondary Devic
Live CD/DVD
Information
Type: Image
Size: 2.86 GB
Location: C:\GUBUNTU-CS570\ubunt...
OK
Cancel

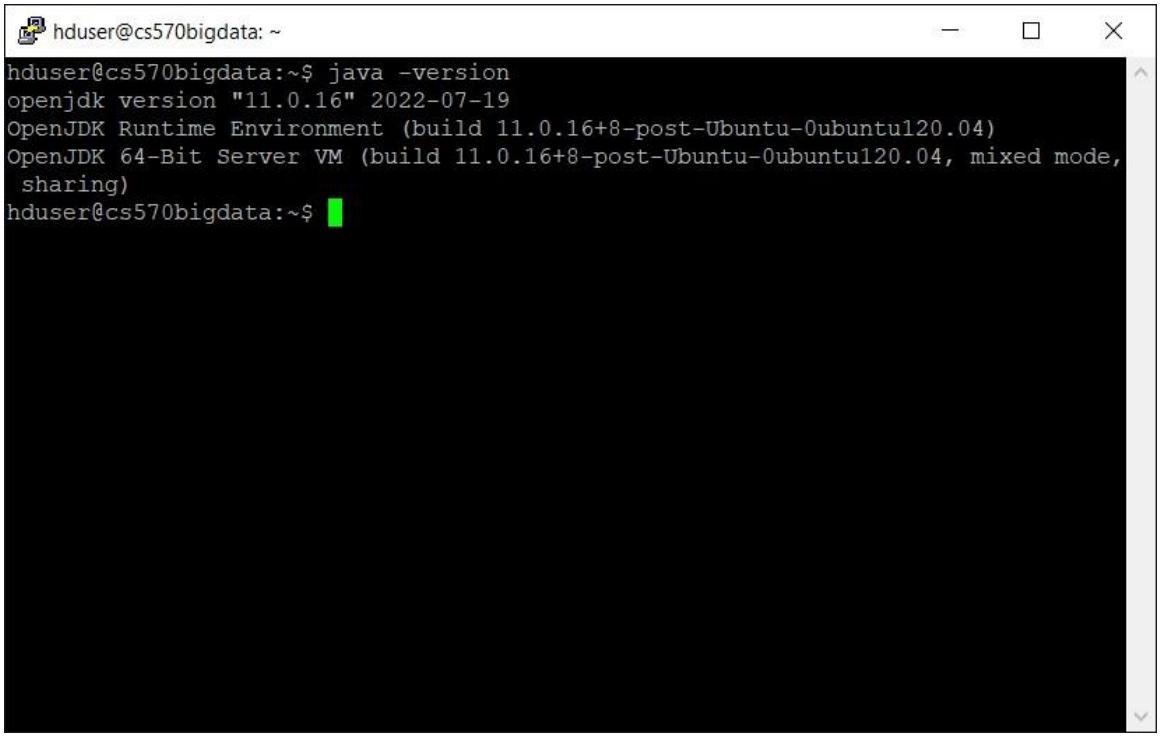




After completion of installation Oracle VM , install Java , Hadoop and the script files

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



Setup hadoop user for Hadoop Installation

sudo addgroup hadoop

sudo adduser --ingroup hadoop hduser

sudo su hduser

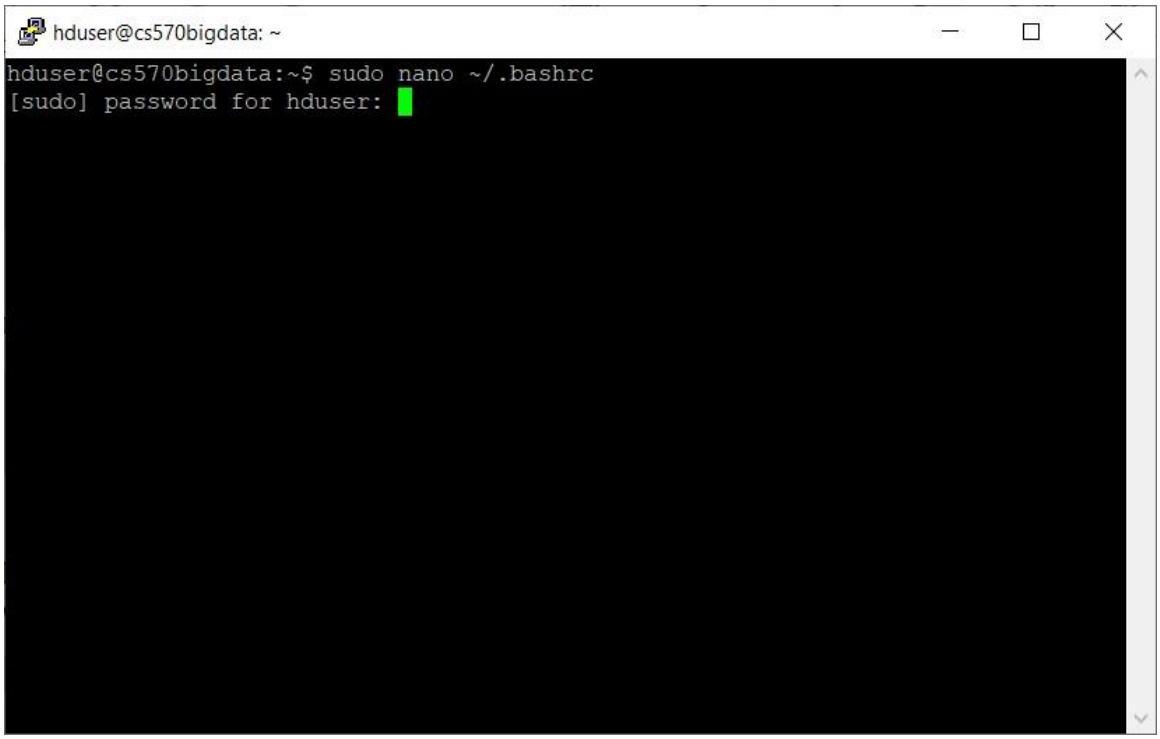
For this I used hadoop version hadoop-2.10.2.tar.gz

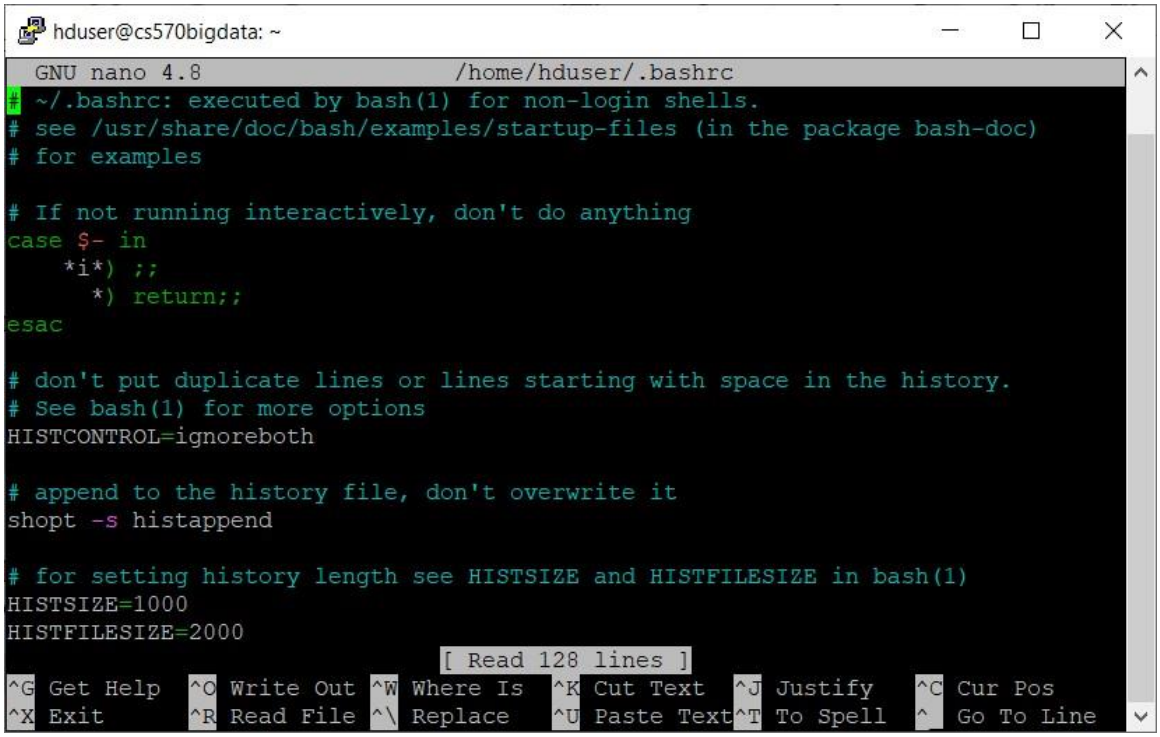
Sudo tar vxzf hadoop-2.10.2.tar.gz –C /usr/local

Cd / usr/local

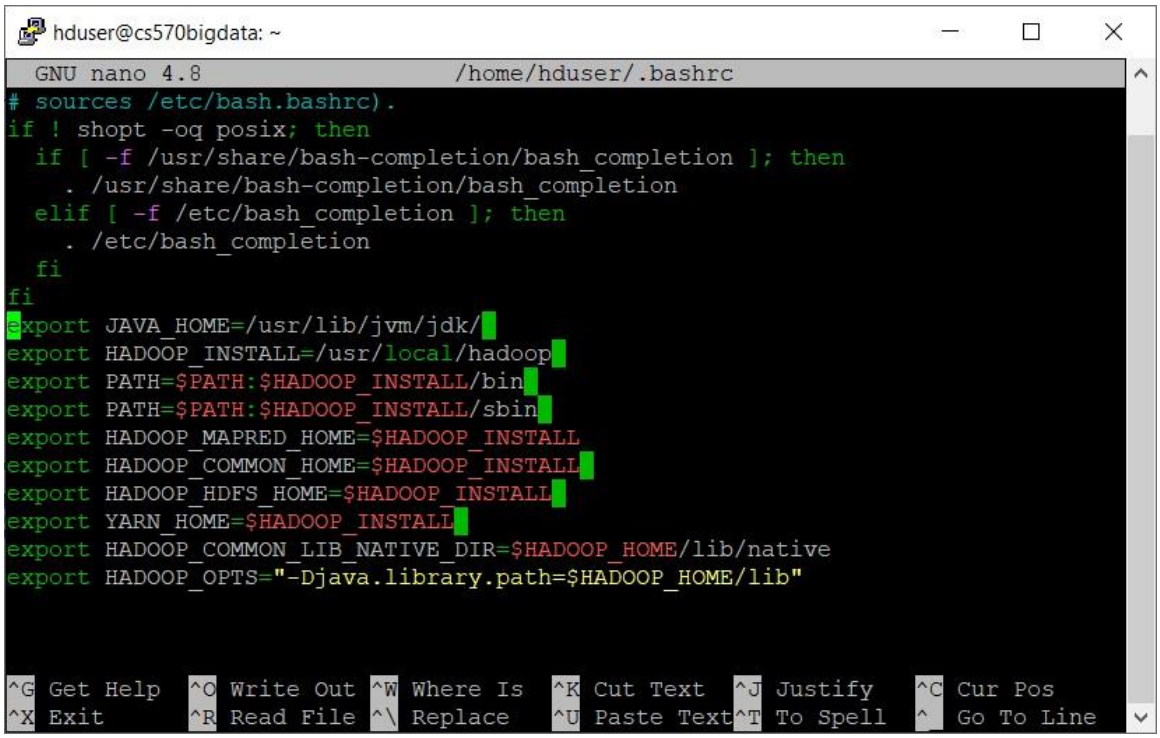
Sudo mv hadoop-2.10.2 hadoop

Sudo chown –R hduser:hadoop hadoop



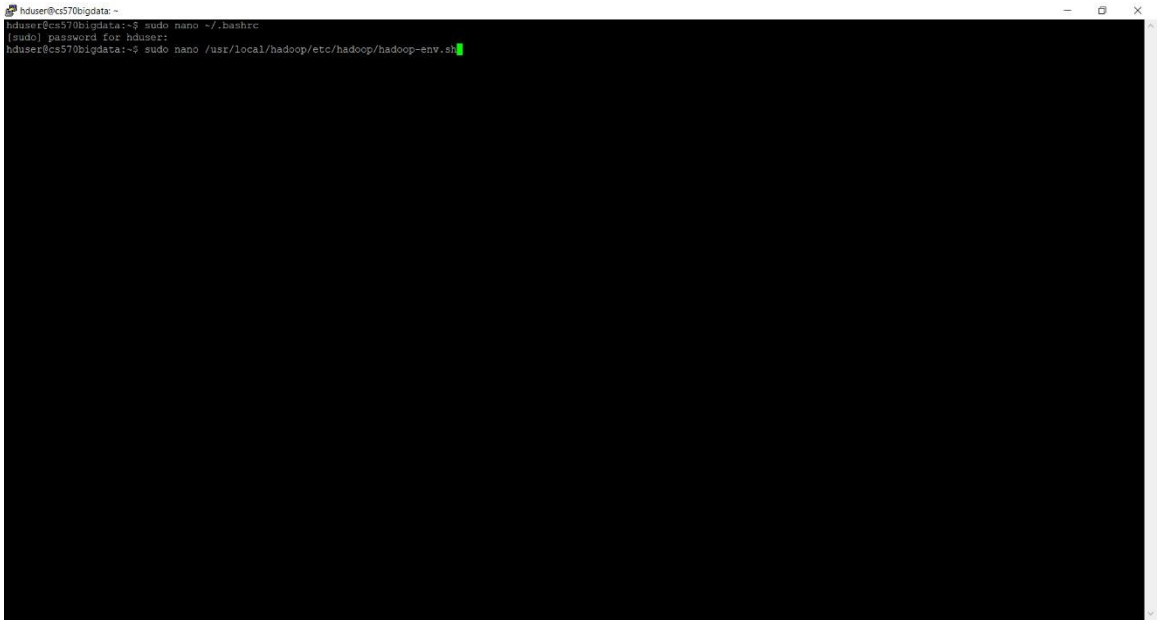


Move to the end and add following lines for hadoop



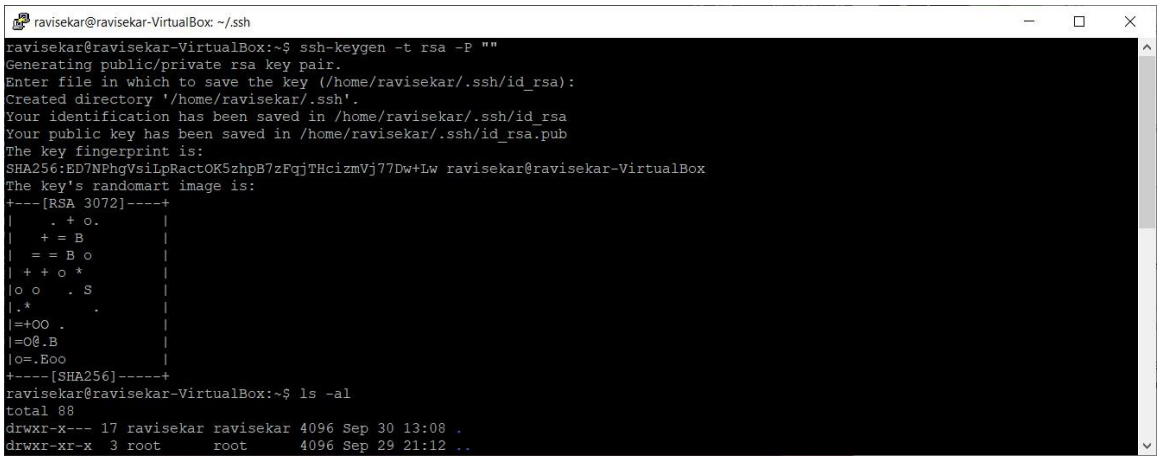
To save use source ~/.bashrc
systemctl reboot –i

Now do the following setting for hduser



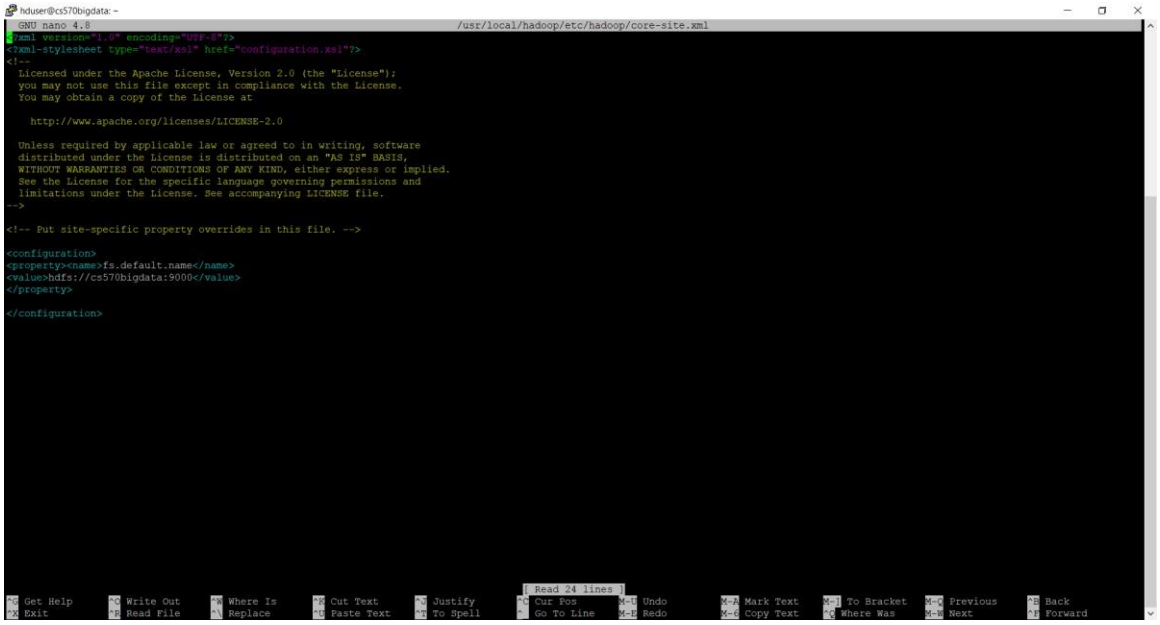
export JAVA_HOME=/usr/lib/jvm/jdk

Ssh generation and creation of authorized keys from public ssh keys

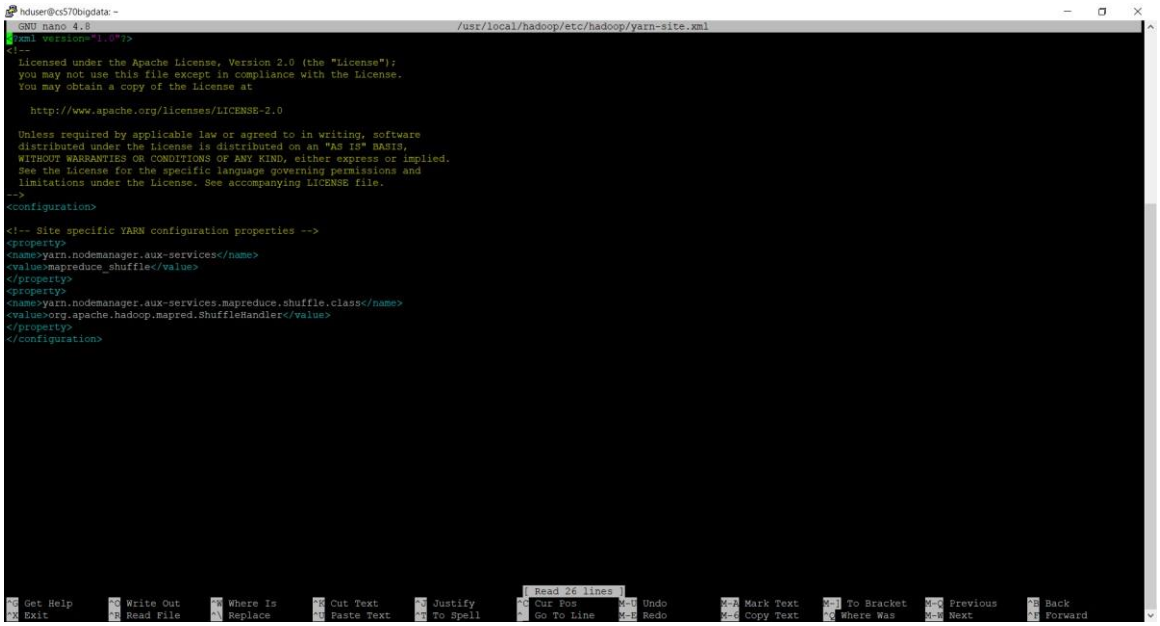


Now we need following files to be set for hadoop :

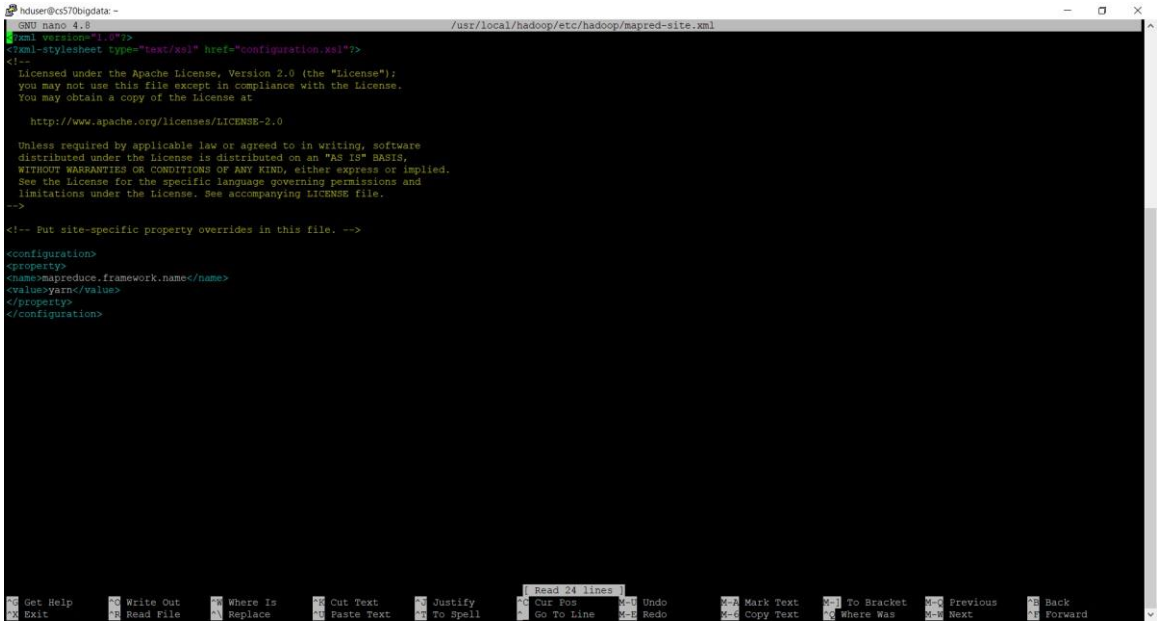
1 . sudo nano /usr/local/hadoop/etc/hadoop/core-site.xml



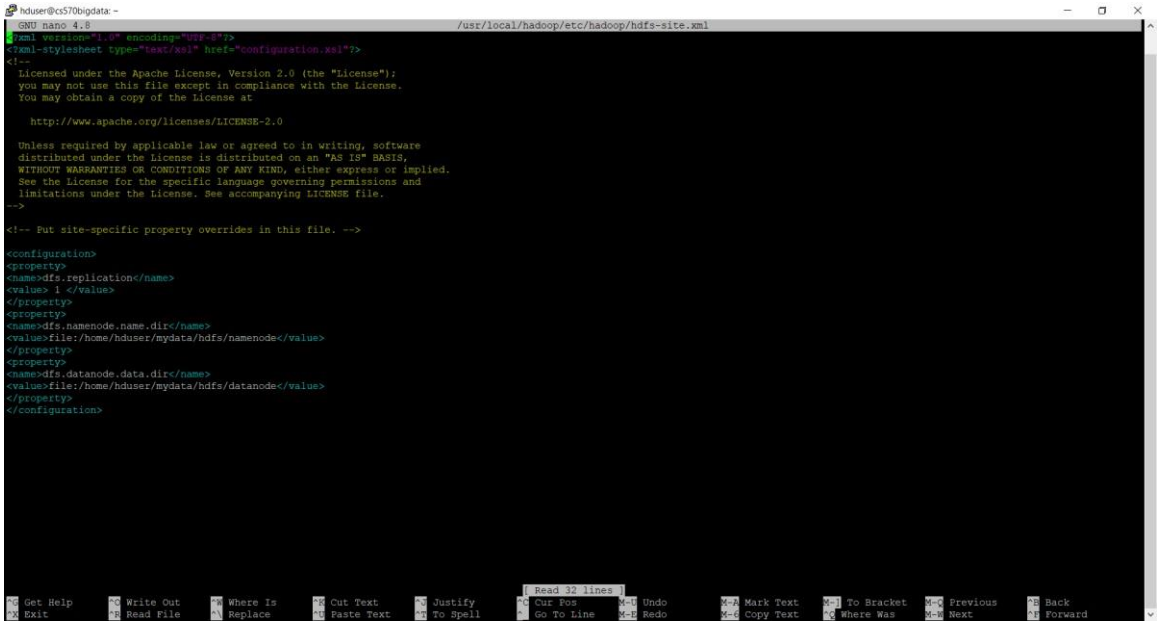
2. sudo nano /usr/local/hadoop/etc/hadoop/yarn-site.xml



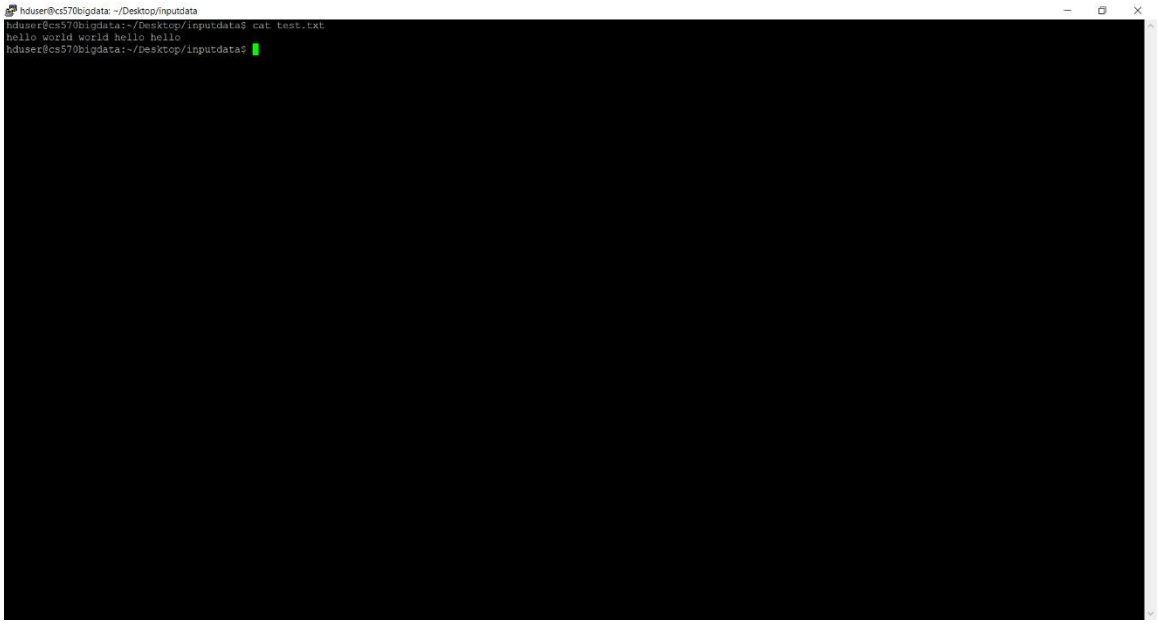
3. sudo nano /usr/local/hadoop/etc/hadoop/mapred-site.xml



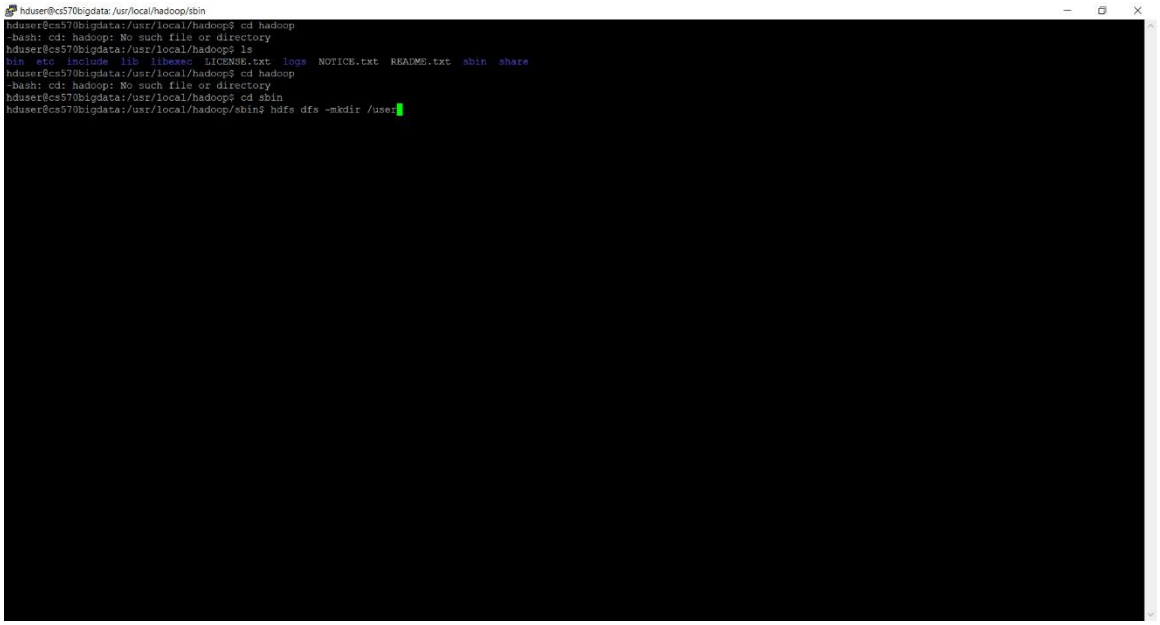
4 For doing hdfs.xml
First complete this task's mkdir
–p mydata/hdfs/namenode mkdir
–p mydata/hdfs/datanode
sudo nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml



Create following test data file for hdfs



Create directory as mentioned



Now copy to file created for input

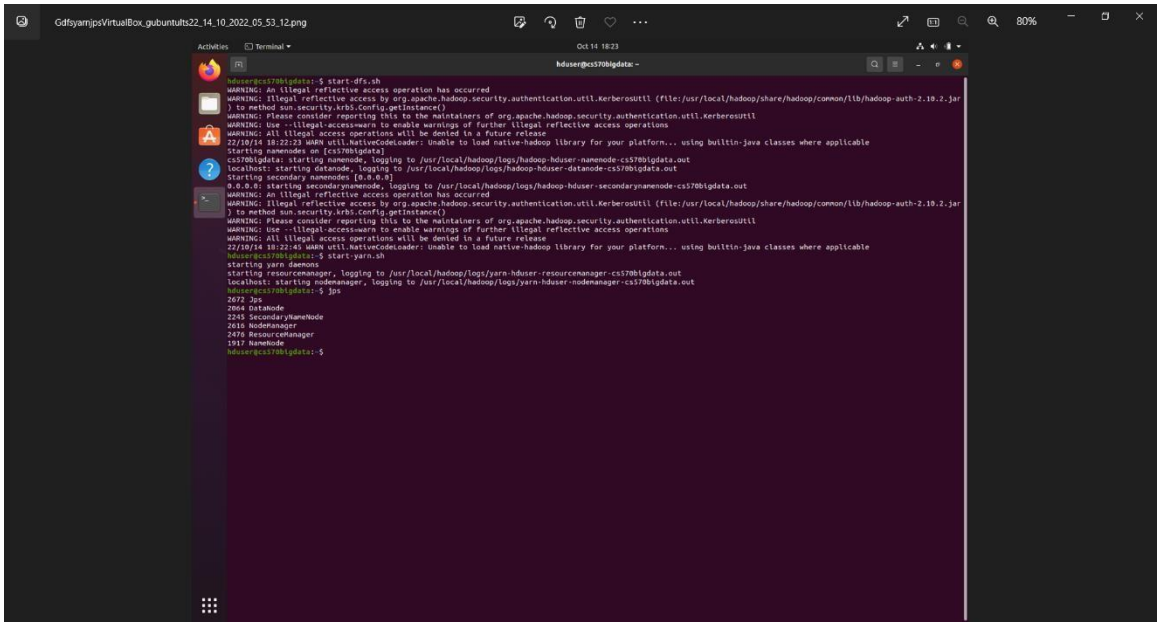
/usr/local/hadoop/bin/hdfs dfs -put '/home/hduser/Desktop/inputdata' /user

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



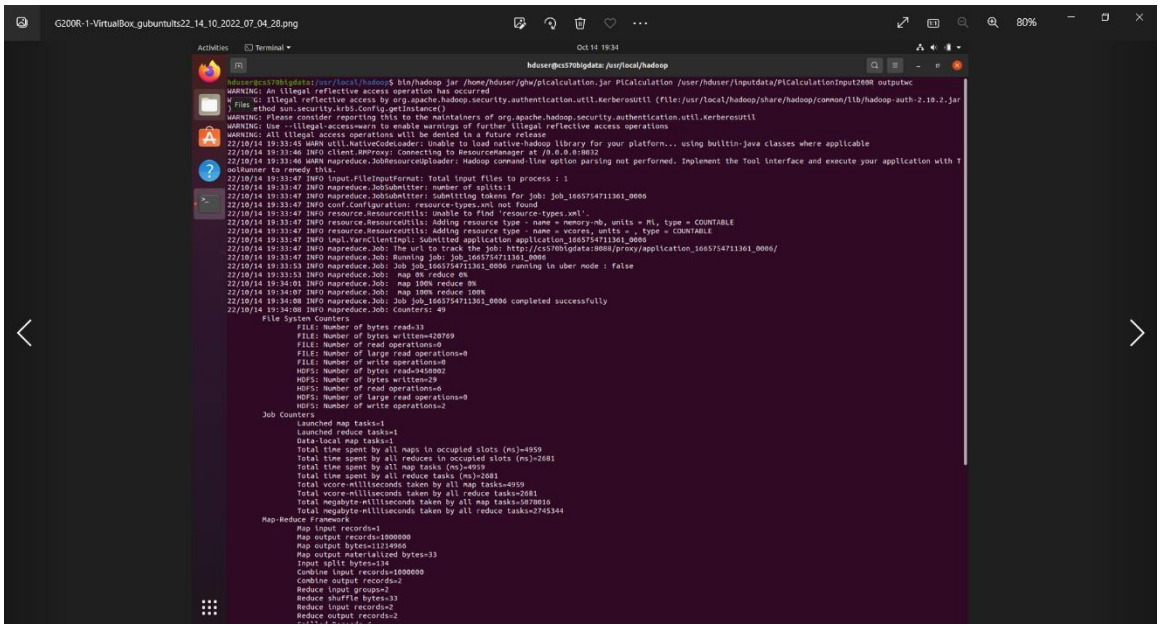
Procedure to test mapreduce pi program :

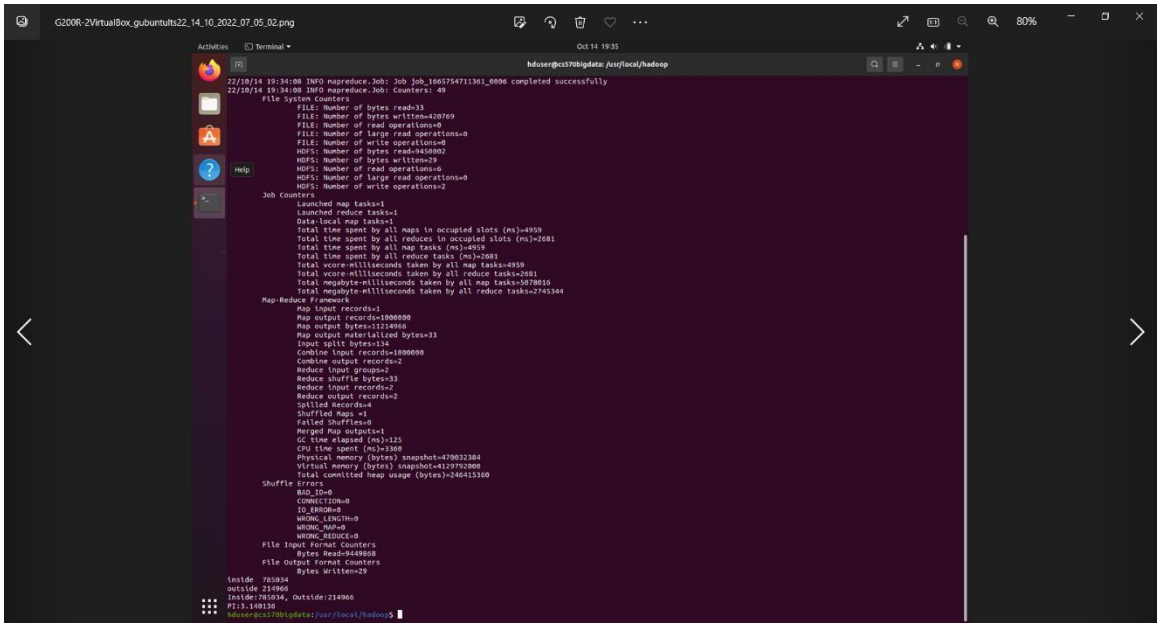
1. Compile the java program(PiCalculation.java)
2. create the jar file for the class generated (picalculation.jar)
3. run the below command for your input data

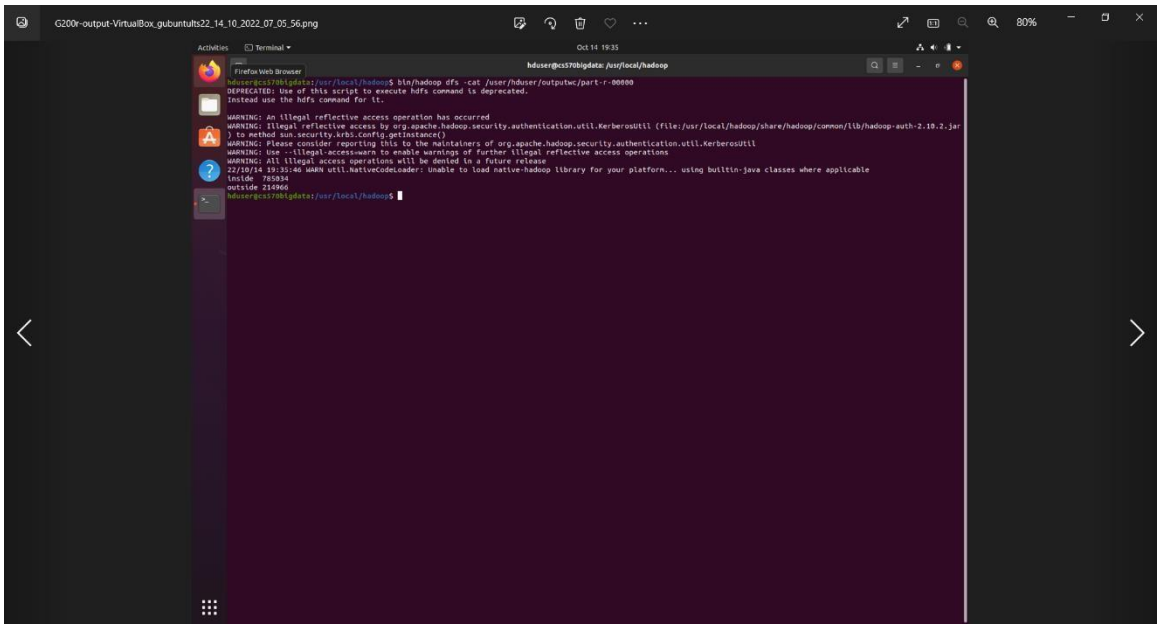
bin/hadoop jar /home/hduser/ghw/picalculation.jar PiCalculation /user/hduser/inputdata/PiCalculationInput

Tests for Random Numbers :

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
80%
Activities
Terminal
Oct 14 19:35
hduser@cs570bigdata: /usr/local/hadoop
Firefox Web Browser
hduser@cs570bigdata:/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 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
inside 785034
outside 214966
hduser@cs570bigdata:/usr/local/hadoop$