Steps to install Hadoop In various Modes.

Step 1: First make the linux environment Suitable to install Hadoop by installing various packages.

Sudo apt update

to update the respositories

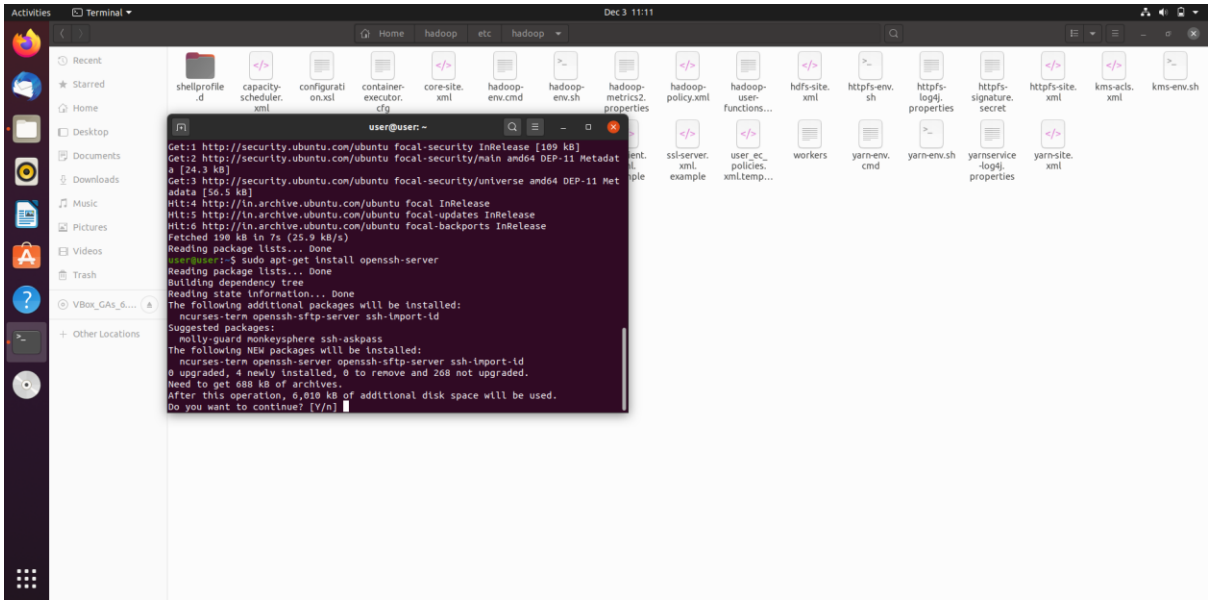
Sudo apt install openjdk-8-jdk openjdk-8-jre

#to install java 1.8 in linux system

Sudo apt-get install openssh-server

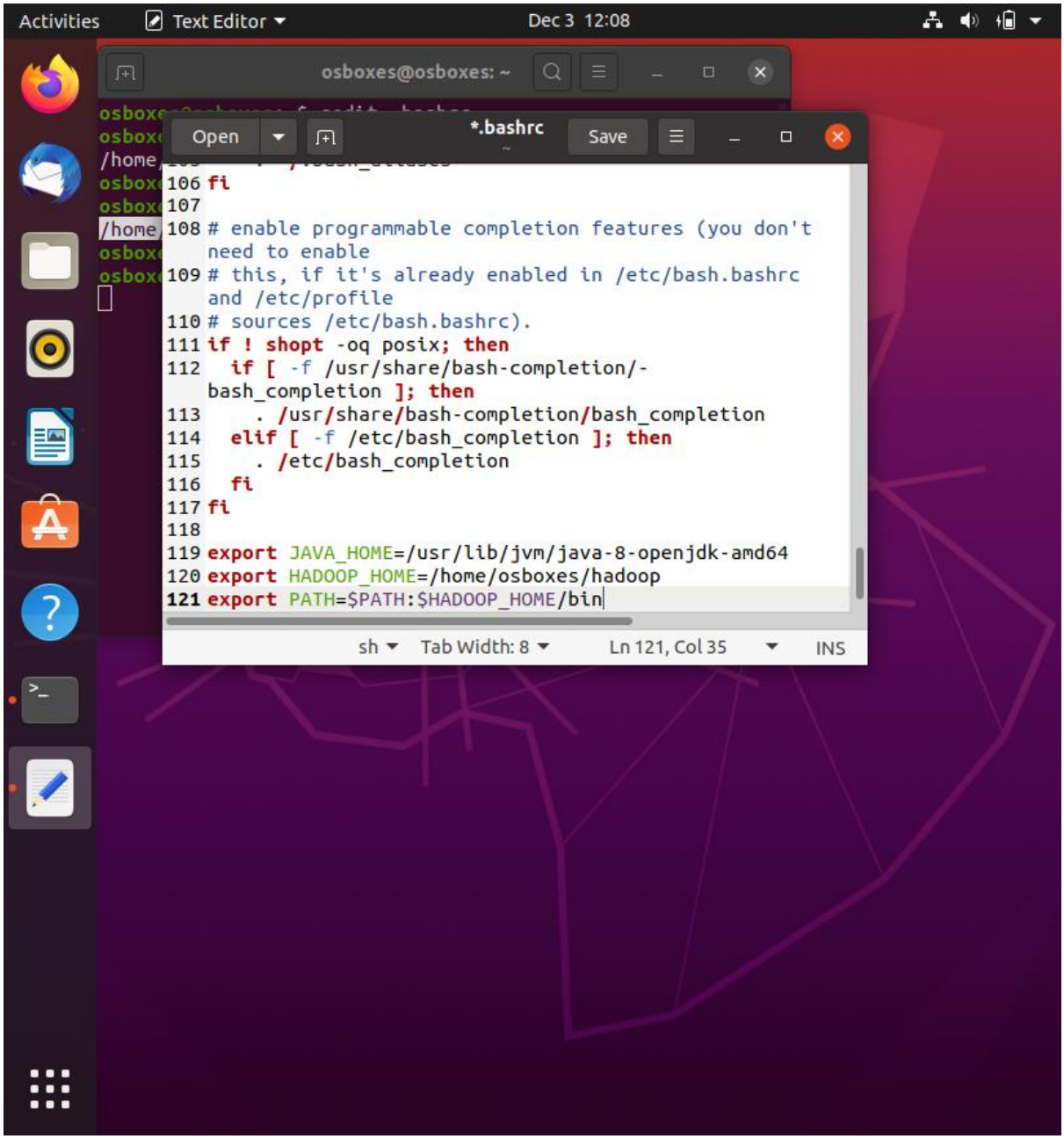
Sudo ufw allow 22

#to install ssh and enable port no.22



Step 2:- Setting JAVA_Home and paths

Goto .bashrc in you home directory with your faviurate text editor and seth those paths



Step 3: Downloading Hadoop

Goto :- https://hadoop.apache.org/release/3.2.1.html

And click on download tar.gz

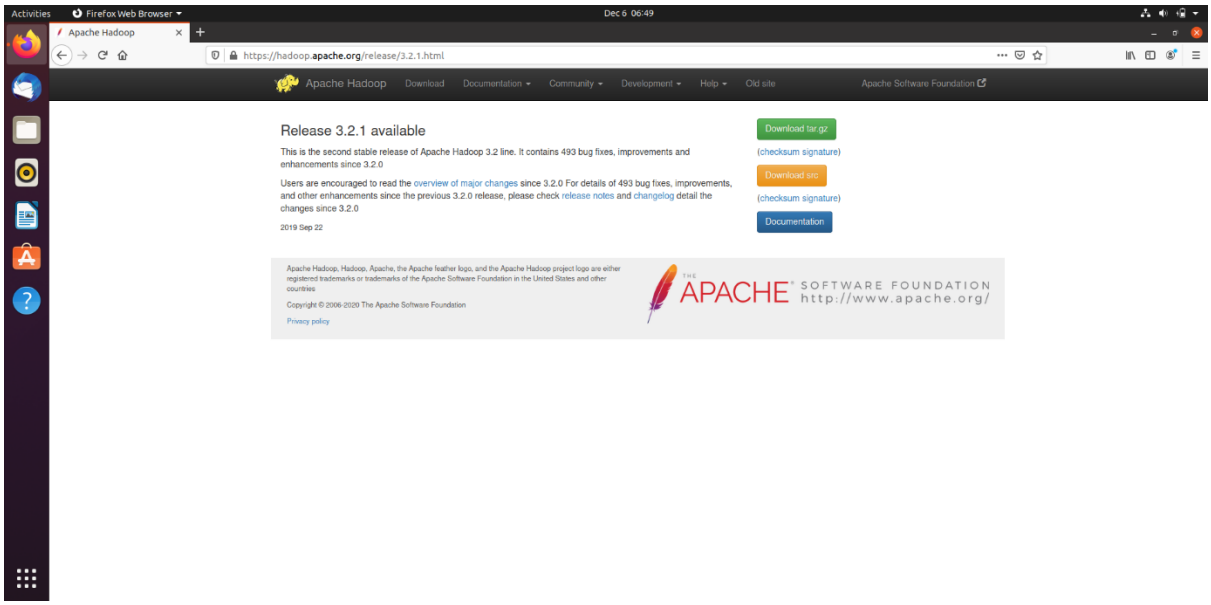
Step 4:- installing/extraction Hadoop in your system

And go to download directory and fire

tar-xvzfhadoop-3.2.1.tar.gz
to extract the file.

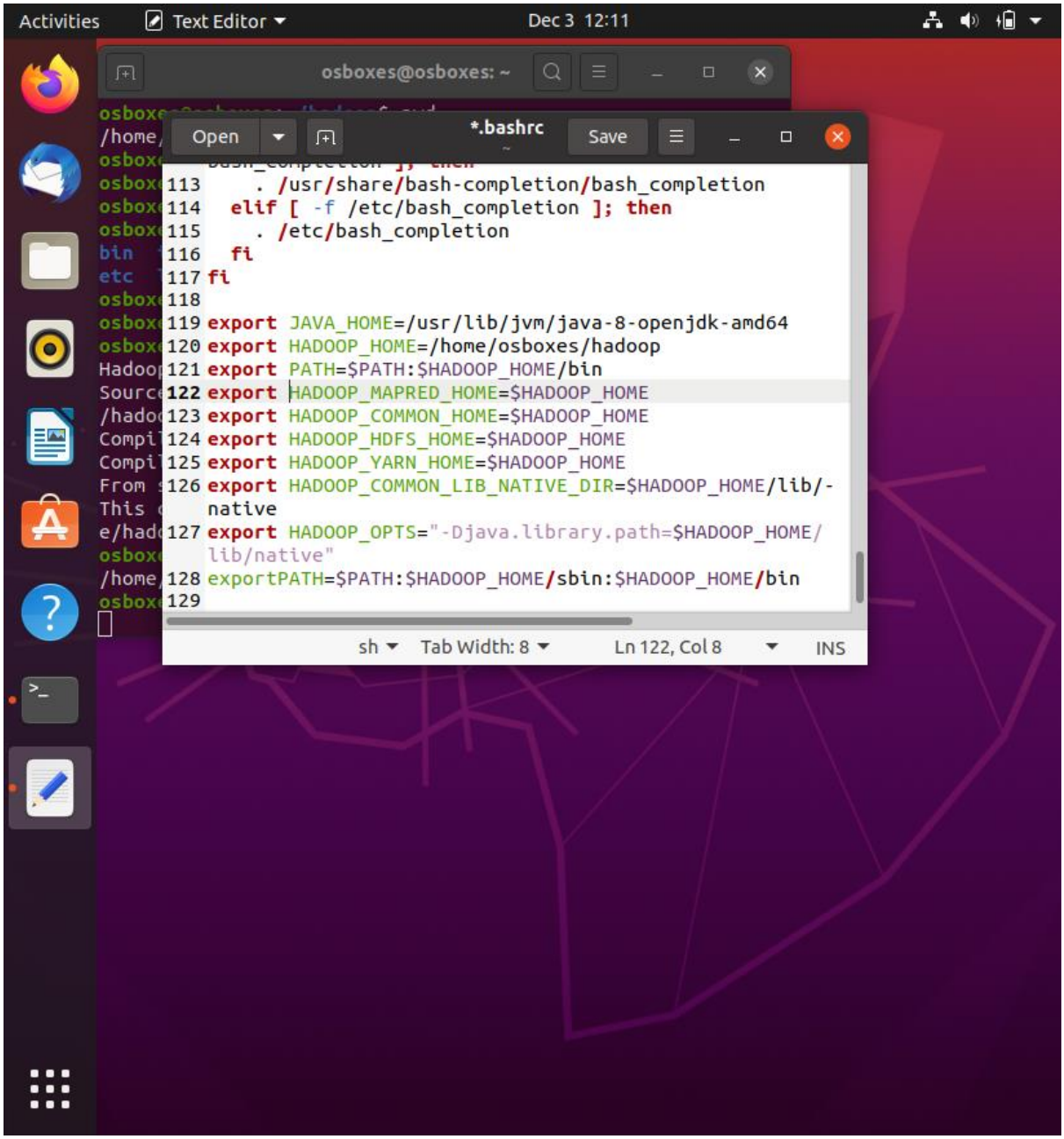
mvhadoop-3.2.1<path-to-directory>
to move Hadoop extracted file to desired directory

ln-shadoop-3.2.1hadoop
to create softlink to access that Hadoop direcotary flawlessly



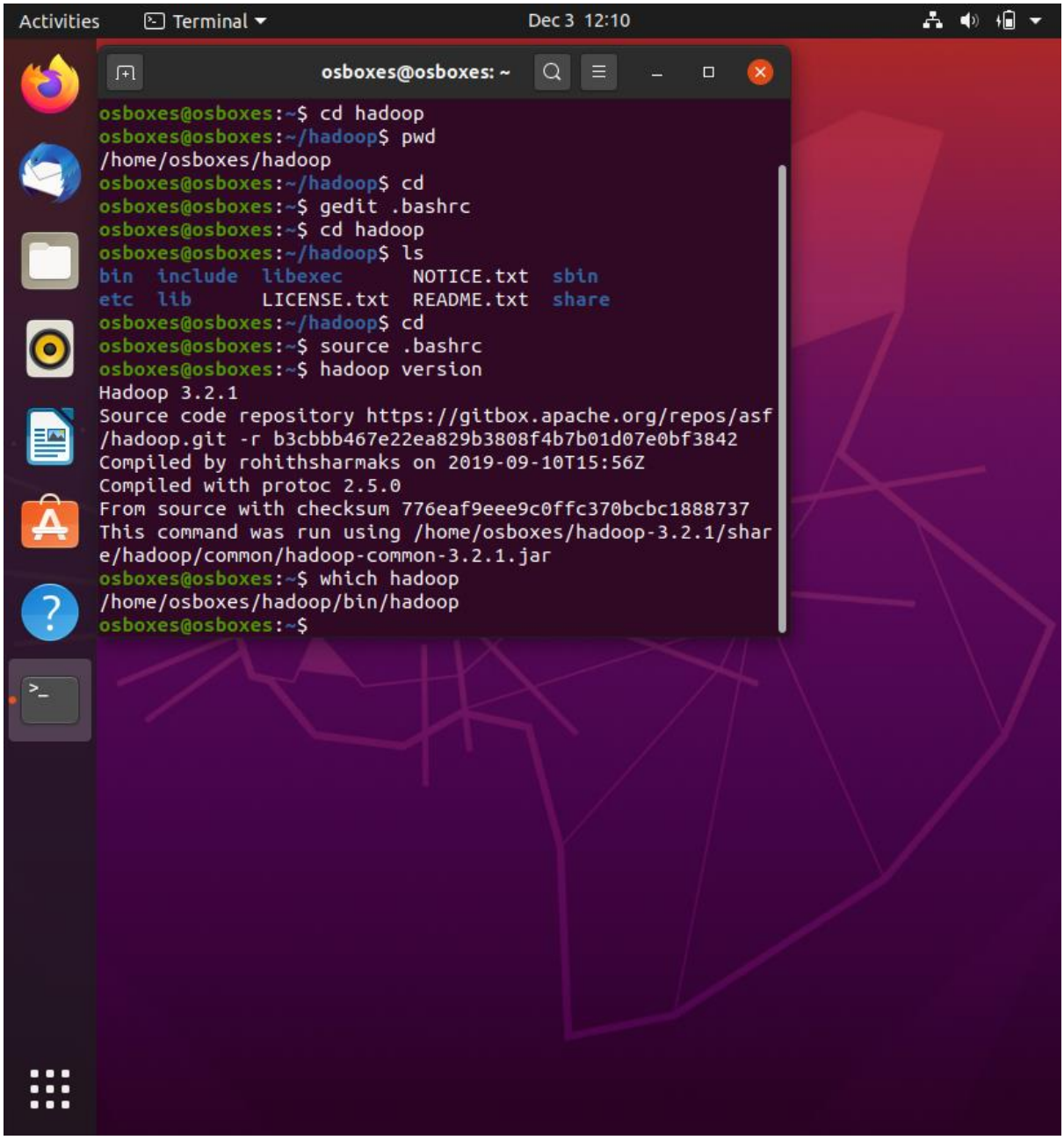
Step 4:- Installing/extraction Hadoop in your linux system

Now set export these paths in your .bashrc file and source it using source .bashrc



Step 5 :- Verify installation.

Run Hadoop version and you will get screen like this.



Congratulations You are done with Standalone mode of Hadoop.

Pseudo Distributed mode

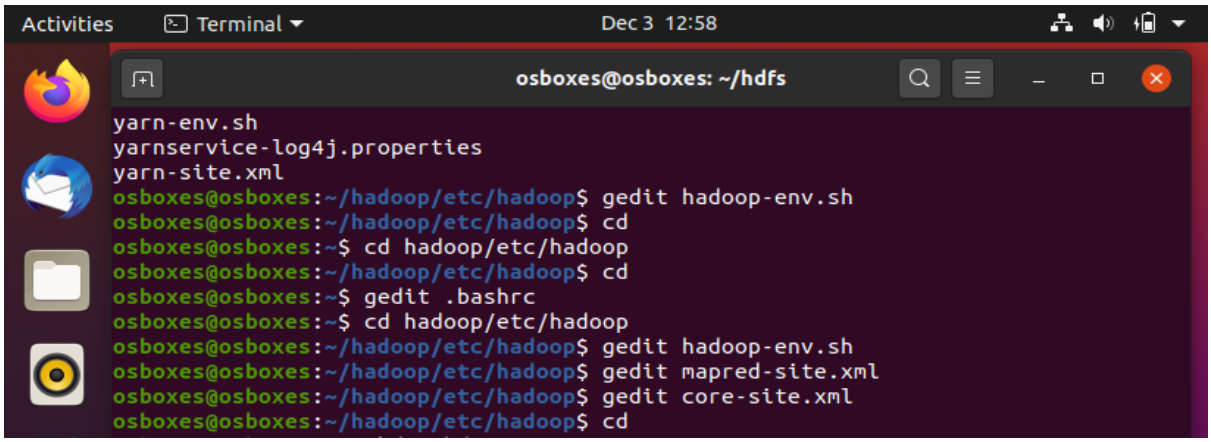
We have to ensure keyless running of all nodes we have to create a public key and assign that to authorized users.

For this just fire.

$ssh-keygen-trsa–P ‘‘’’

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

After this We have to configure some properties in different files of Hadoop

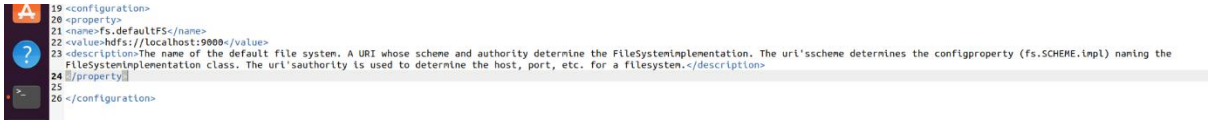


And the configuration for different files are:-

Gedit Hadoop-env.sh

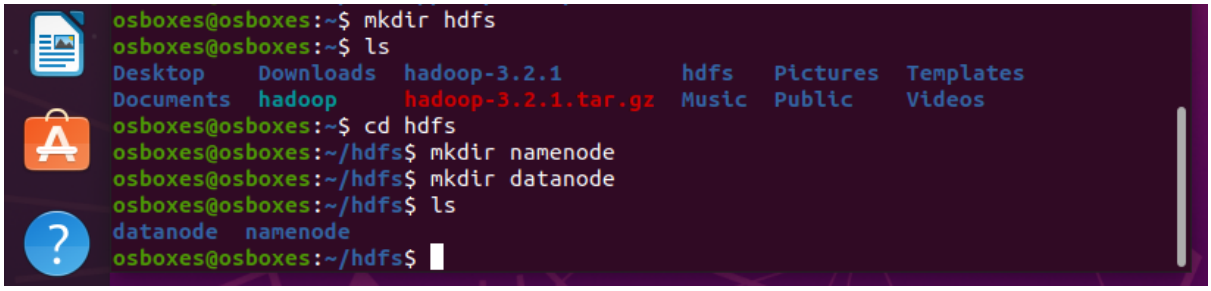
Gedit mapred-site.xml

Gedit core-site.xml



Then go to home home directory using cd command and make dir hdfs with mkdir hdfs command

And in hdfs dir make 2 subdir using mkdir command.



Now go to Hadoop/etc/Hadoop again and configure the last file named as hdfs-site.xml and provide address of those directories.

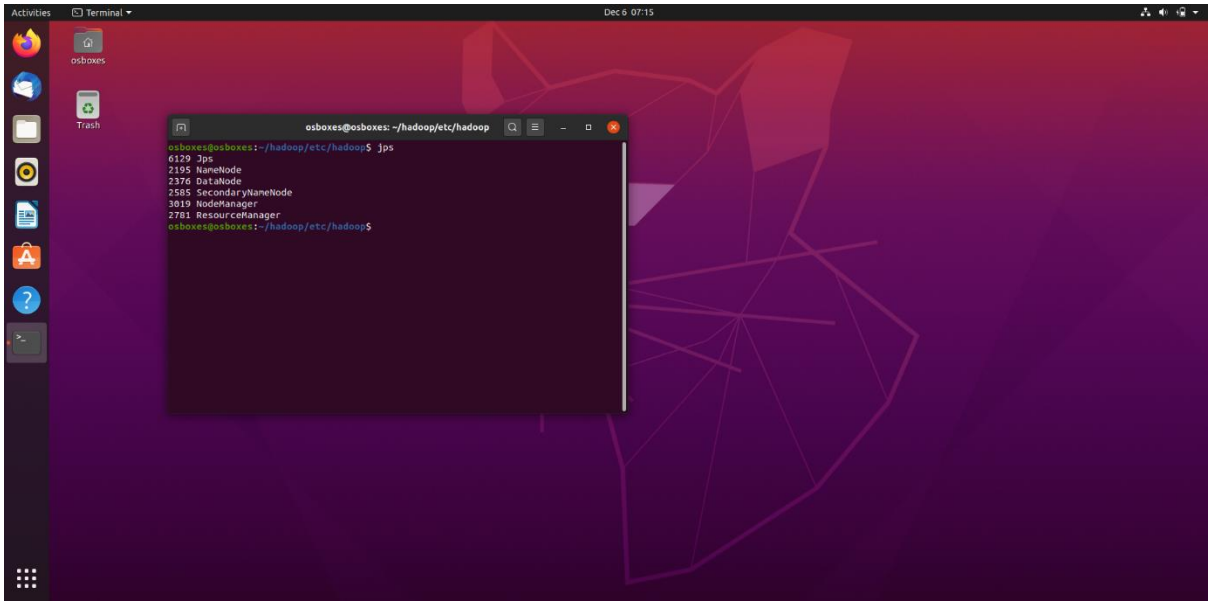
Congratulations Your Are done with pseudo mode installation of Hadoop.

Final Step:- To verify the Hadoop pseudo mode start all the nodes with following commands and your terminal will look like this.

Start-dfs.sh

Star-yarn.sh

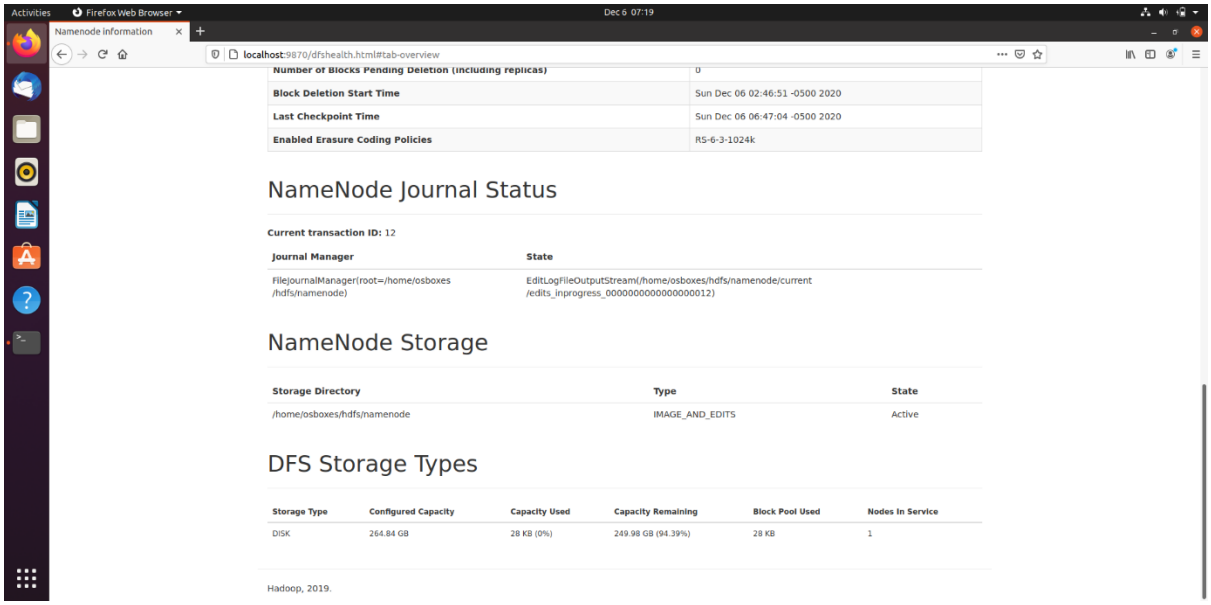
Jps



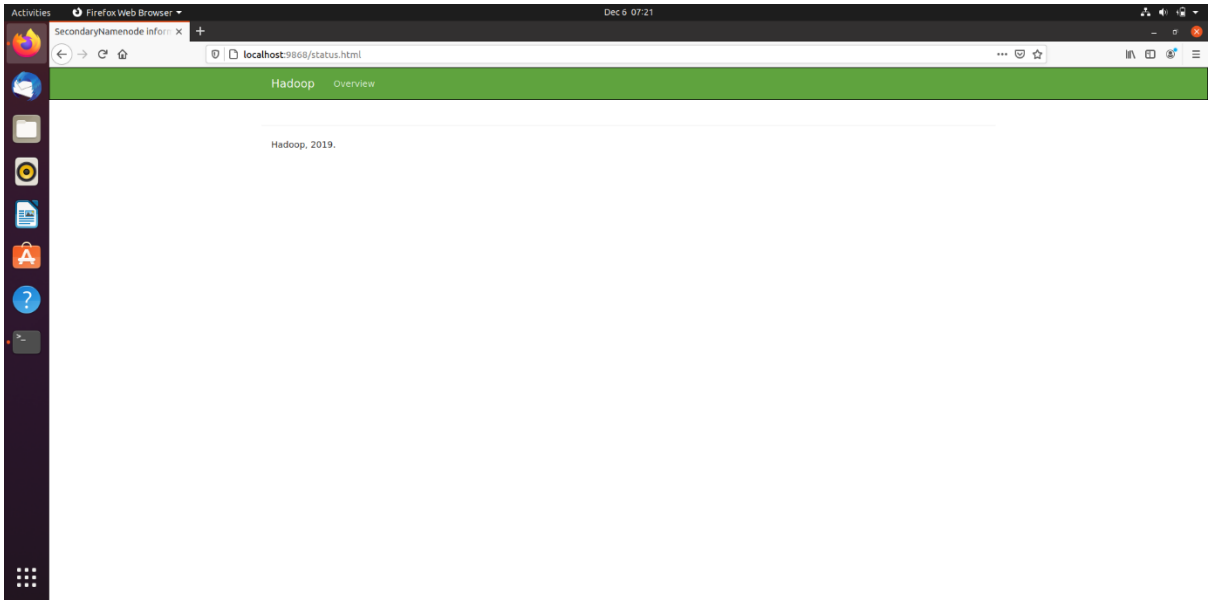
Answer2:

1.web interfaces of Hadoop Namenode

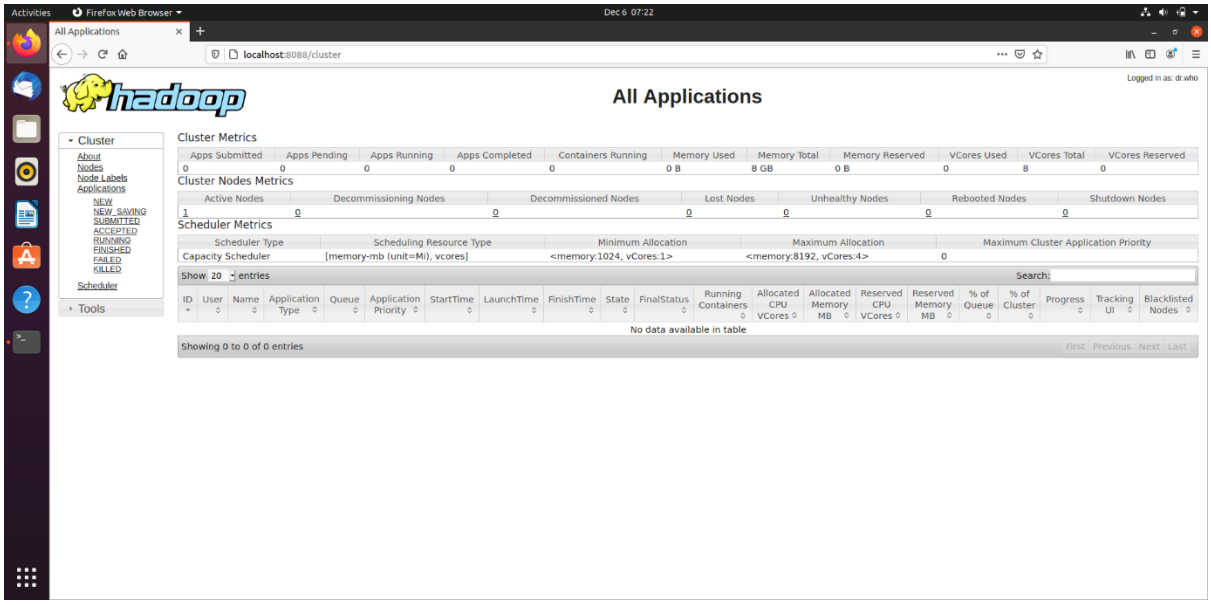
Type **localhost:9870**



2. .web interfaces of Hadoop Secondary Namenode Localhost:9868



3. .web interfaces of Hadoop Resouce manager Localhost:8088



Answer3:- Script to install Hadoop in a single click

!bin/bash

```
#--Please choose appropriate hadoop link for your installation.

bash_file="$HOME/.bashrc"
install_path="/usr/local"

hadoop_url=http://archieve.apache.org/dist/hadoop/common/hadoop-3.2.1/hadoop-
3.2.1.tar.gz
hadoop_file=hadoop-3.2.1.tar.gz
hadoop_folder_name=hadoop-3.2.1
hadoop_xml_folder=$install_path/$hadoop_folder_name/etc/hadoop

user_name="hduser"
group_name="hadoop"

install_flag=0

clear

echo ""; echo ""; echo ""
echo "This script will install Hadoop and Spark to your local directory, modify
your PATH,and add environment variables to your SHELL config file"
read -r -p "Proceed? [y/N] " response
if [[ ! $response =~ ^([yY][eE][sS]|[yY])$ ]]
then
   echo "Aborting..."
   sleep 1
   exit 1
fi

#--Packages currently installed with new versions available are retrieved
#--and upgraded; under no circumstances are currently installed packages
#--removed, nor are packages that are not already installed retrieved
#--and installed.
sudo apt-get update && sudo apt-get upgrade
sudo apt-get -y install software-properties-common
sudo apt-get -y install vim

clear

#--Add a new hadoop group "hadoop"
getent group $group_name 2> /dev/null
if [ ! $? -eq 0 ]; then
   echo "Creating a group 'hadoop' "
   sudo addgroup hadoop
fi

#--Add a new user "hduser"
getent passwd $user_name 2> /dev/null
if [ ! $? -eq 0 ]; then
```

```
    echo "Creating a user 'hduser' "
    sudo adduser --ingroup hadoop hduser
    sudo adduser hduser sudo
fi

#--Check SSH installation/configuration.
ssh -V 2> /dev/null
if [ ! $? -eq 0 ]; then
    echo "Installing SSH"
    sudo apt-get -y install ssh
    sudo apt-get -y install rsync
    ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa
    cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
fi

#--Check JAVA is installed.
javac -version 2> /dev/null
if [ ! $? -eq 0 ]; then
    echo "Installing Java"
    apt-get install -y software-properties-common python-software-properties
    sudo add-apt-repository -y ppa:webupd8team/java
    sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv-keys EEA14886
    sudo apt-get install oracle-java8-installer
    sudo apt-get -y update
fi

#--Check HADOOP is installed.
hadoop version 2> /dev/null
if [ ! $? -eq 0 ]; then
    echo "Installing Hadoop"
    wget $hadoop_url
    sudo tar zxvf $hadoop_file -C $install_path
    sudo chown -R $user_name:$group_name $HADOOP_HOME
    install_flag=1
    sudo rm -rf $hadoop_file
    sleep 1
fi

#--Bashrc and XML file configurations.
_hadoop_config () {

echo "
export JAVA_HOME=/usr/lib/jvm/java-8-oracle
export HADOOP_INSTALL=$install_path/$hadoop_folder_name
export HADOOP_HOME=$install_path/$hadoop_folder_name
export PATH=\$PATH:\$HADOOP_INSTALL/bin:\$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/native"
export HADOOP_CONF_DIR=\$HADOOP_INSTALL/etc/hadoop" >> $bash_file

source $bash_file

#--Snappy compression configuration For HADOOP
sudo apt-get install -y libsnappy-dev
sudo cp /usr/lib/x86_64-linux-gnu/lib* $HADOOP_HOME/lib/native/

#--Update JAVA path to hadoop-env.sh
sudo sed -i 's/\${JAVA_HOME}/\/usr\/lib\/jvm\/java-8-oracle/g'
$hadoop_xml_folder/hadoop-env.sh

#--Modify core-site.xml
sudo sed -i -r '/FOOTER/d; s/(<configuration>)//;
                    s/(<\/configuration>)//' $hadoop_xml_folder/core-site.xml
sudo echo "
<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration> "  >> $hadoop_xml_folder/core-site.xml

#--Modify yarn-site.xml
sudo sed -i -r '/FOOTER/d; s/(<configuration>)//;
                    s/(<\/configuration>)//' $hadoop_xml_folder/yarn-site.xml
sudo echo "
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration> " >> $hadoop_xml_folder/yarn-site.xml

#--Modify hdfs-site.xml
sudo sed -i -r '/FOOTER/d; s/(<configuration>)//;
                    s/(<\/configuration>)//' $hadoop_xml_folder/hdfs-site.xml
sudo echo "
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.name.dir</name>
    <value>/hdfs_storage/name</value>
  </property>
  <property>
```

```
    <name>dfs.data.dir</name>
    <value>/hdfs_storage/data</value>
  </property>
</configuration> " >> $hadoop_xml_folder/hdfs-site.xml

#--Modify mapred-site.xml
if [-f $hadoop_xml_folder/mapred-site.xml ]; then
   sudo sed -i -r '/FOOTER/d; s/(<configuration>)//;
               s/(<\/configuration>)//' $hadoop_xml_folder/mapred-site.xml
else
   sudo cp $hadoop_xml_folder/mapred-site.xml.template $hadoop_xml_folder/mapred-
site.xml
   sudo sed -i -r '/FOOTER/d; s/(<configuration>)//;
               s/(<\/configuration>)//' $hadoop_xml_folder/mapred-site.xml
   sudo chown -R $user_name:$group_name $hadoop_xml_folder/mapred-site.xml
fi

sudo echo "
<configuration>
  <property>
    <name>mapred.job.tracker</name>
    <value>localhost:54311</value>
    <description>The host and port that the MapReduce job tracker runs
                 at. If "local", then jobs are run in-process as a single map
                 and reduce task.
    </description>
  </property>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration> " >> $hadoop_xml_folder/mapred-site.xml
}

#--HADOOP configurations:
#-- bashrc file updates
#-- Snappy compression configuration
#-- XML configurations
if [ $install_flag -eq 1 ]; then
   #--Pseudo Hadoop configurations.
   _hadoop_config

   echo "creating required hdfs directories"
   sudo mkdir -p /hdfs_storage/data
   sudo mkdir -p /hdfs_storage/name
   sudo chown -R $user_name:$group_name /hdfs_storage

   echo "creating log directory"
   sudo mkdir -p $HADOOP_HOME/logs
   sudo chown -R $user_name:$group_name $HADOOP_HOME/logs
```

```
    #--Format the HDFS
    hdfs namenode -format

    echo "Hadoop installation and setup is complete."
fi


echo "Installation completed successfully."
```

Copy the above text in a Text file and save the file with .sh extension.
It will automatically Install the Hadoop and make the environment to run it. However sometime you need to make some changes acc. To ur linux platform