

## 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 repositories

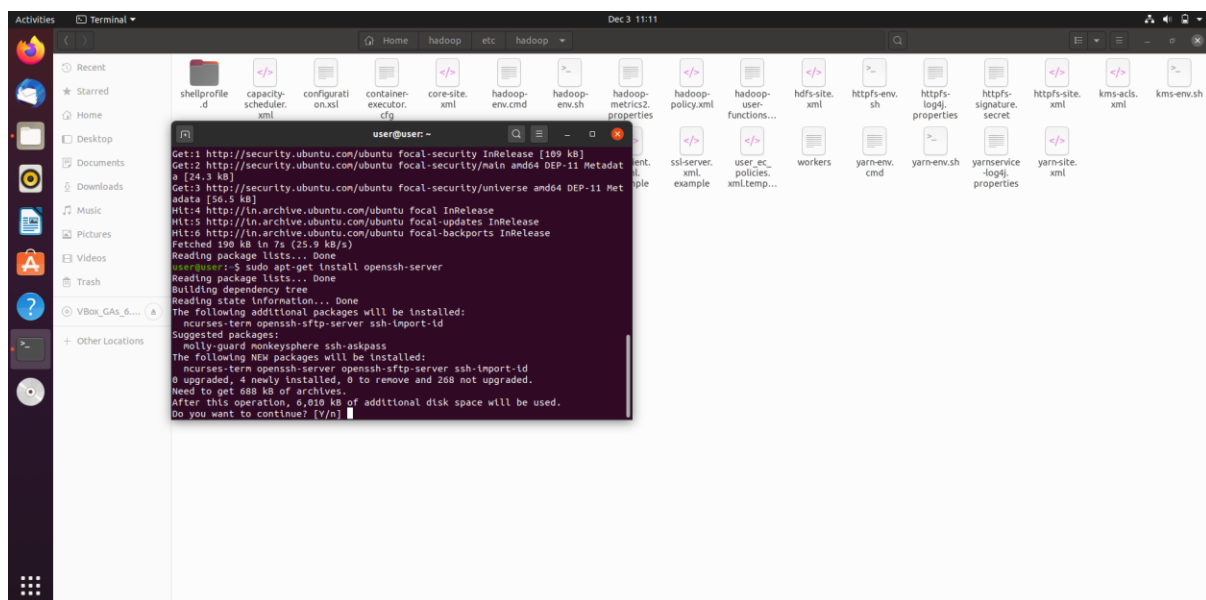
Sudo apt install openjdk-8-jdk openjdk-8-jre

#to install java 1.8 in linux system

Sudo apt-get install openssh-server

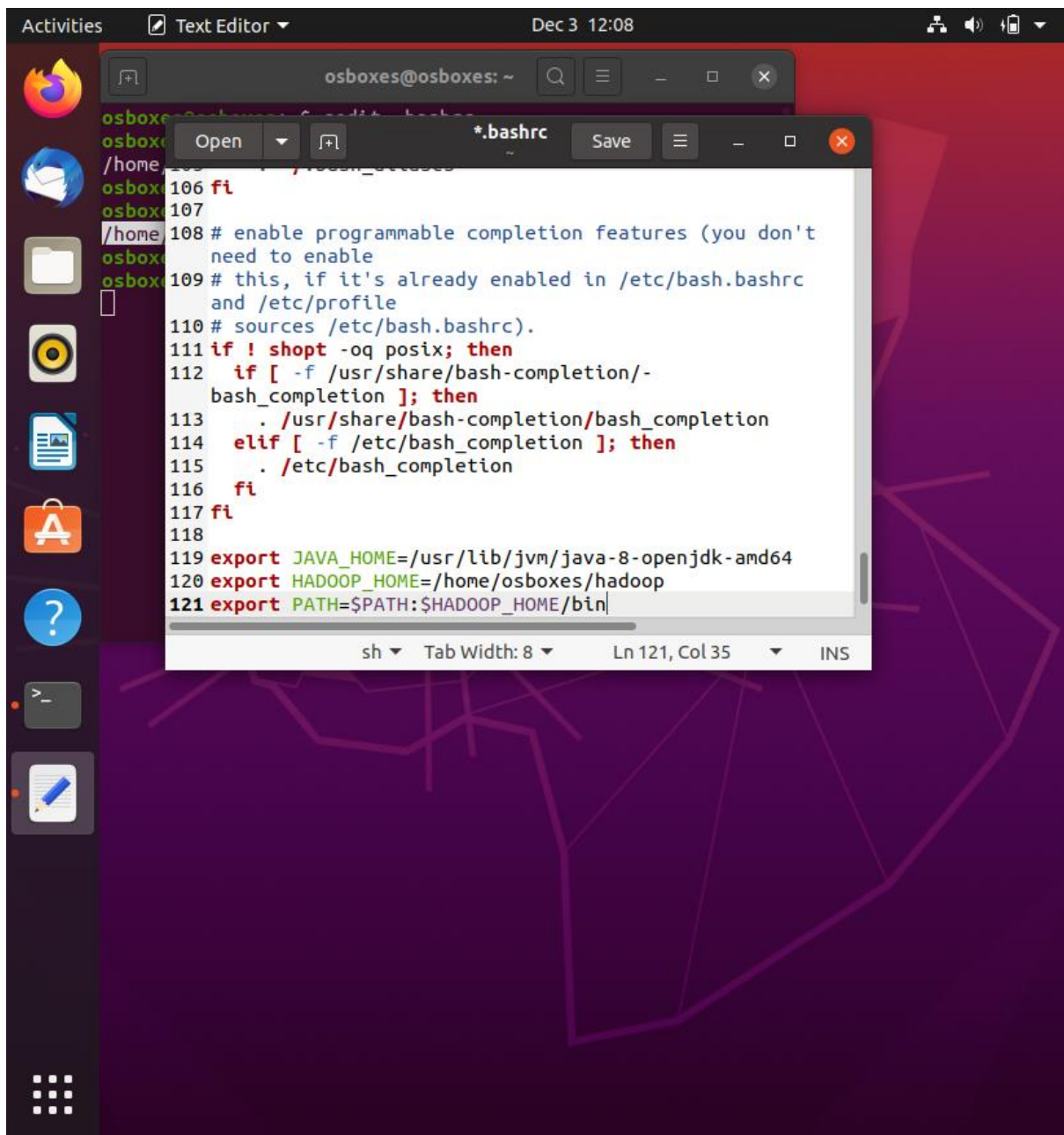
Sudo ufw allow 22

#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-xvzf hadoop-3.2.1.tar.gz
```

# to extract the file.

```
mv hadoop-3.2.1 <path-to-directory>
```

# to move Hadoop extracted file to desired directory

```
ln -s hadoop-3.2.1 hadoop
```

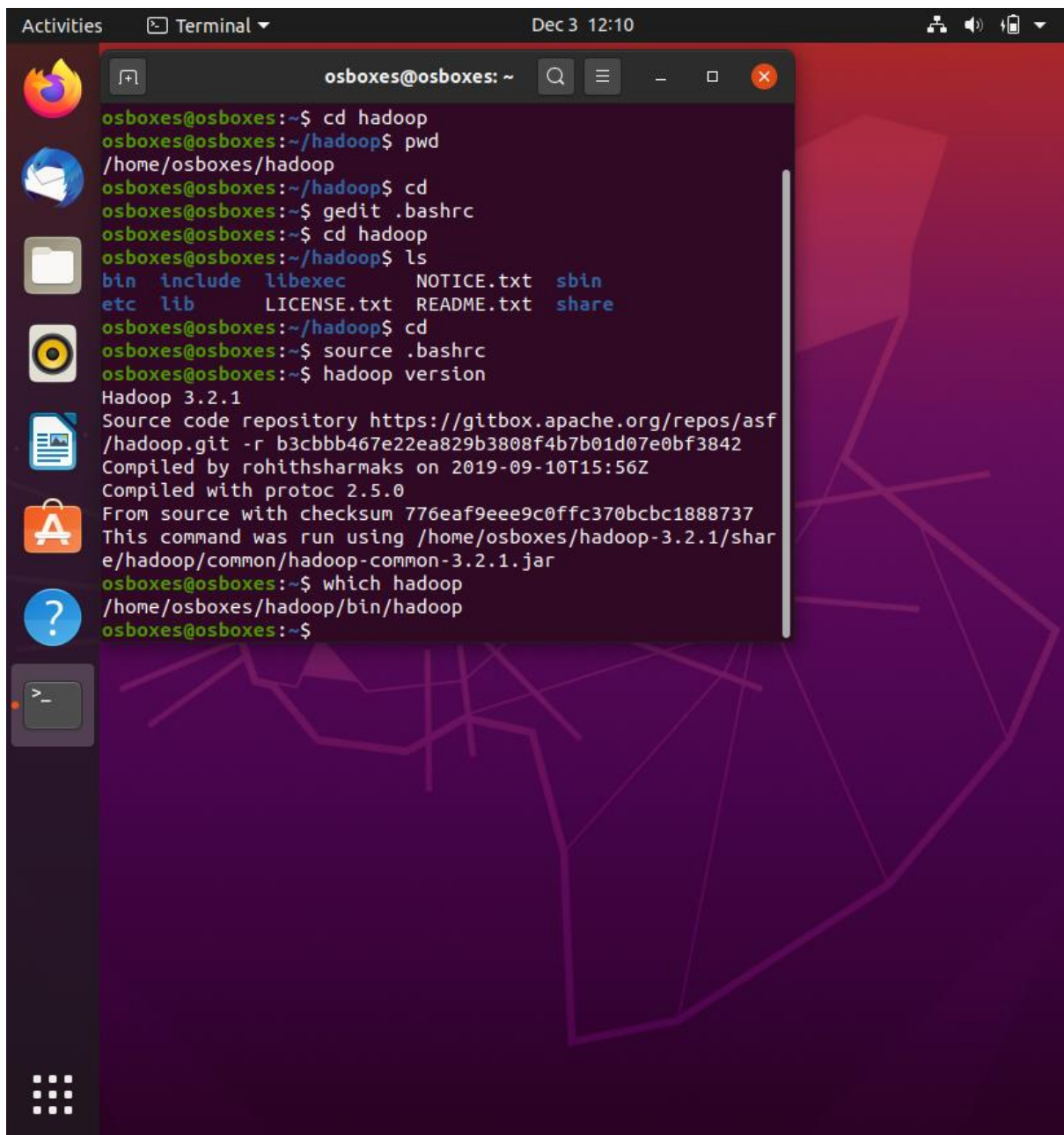
# to create softlink to access that Hadoop directory flawlessly



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

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



A terminal window titled 'osboxes@osboxes: ~' with a search icon, menu icon, and window control buttons. The terminal shows the following commands and output:

```
osboxes@osboxes:~$ cd hadoop
osboxes@osboxes:~/hadoop$ pwd
/home/osboxes/hadoop
osboxes@osboxes:~/hadoop$ cd
osboxes@osboxes:~$ gedit .bashrc
osboxes@osboxes:~$ cd hadoop
osboxes@osboxes:~/hadoop$ ls
bin  include  libexec  NOTICE.txt  sbin
etc  lib      LICENSE.txt  README.txt  share
osboxes@osboxes:~/hadoop$ cd
osboxes@osboxes:~$ source .bashrc
osboxes@osboxes:~$ hadoop version
Hadoop 3.2.1
Source code repository https://gitbox.apache.org/repos/asf
/hadoop.git -r b3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled by rohithsharmaks on 2019-09-10T15:56Z
Compiled with protoc 2.5.0
From source with checksum 776eaf9eee9c0ffc370bcbc1888737
This command was run using /home/osboxes/hadoop-3.2.1/share
hadoop/common/hadoop-common-3.2.1.jar
osboxes@osboxes:~$ which hadoop
/home/osboxes/hadoop/bin/hadoop
osboxes@osboxes:~$
```

Congratulations You are done with Standalone mode of Hadoop.

## Pseudo Distributed mode

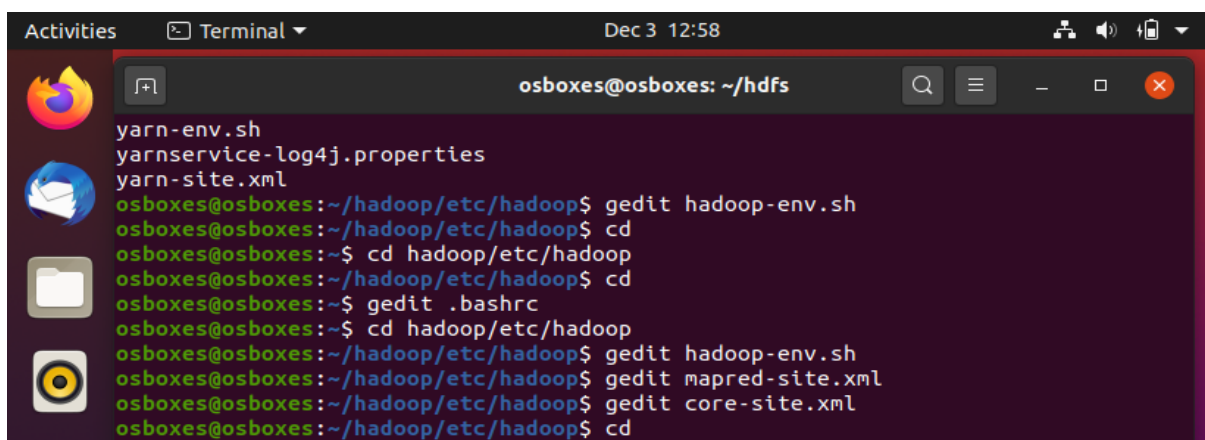
We have to ensure keyless running of all nodes we have to create a public key and assign that to authorized users.

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

A terminal window titled 'osboxes@osboxes: ~/hdfs' showing a series of commands to configure Hadoop. The commands listed are: 'yarn-env.sh', 'yarnservice-log4j.properties', 'yarn-site.xml', 'gedit hadoop-env.sh', 'cd', 'cd hadoop/etc/hadoop', 'gedit .bashrc', 'gedit hadoop-env.sh', 'gedit mapred-site.xml', 'gedit core-site.xml', and 'cd'.

```
osboxes@osboxes: ~/hdfs
yarn-env.sh
yarnservice-log4j.properties
yarn-site.xml
osboxes@osboxes:~/hadoop/etc/hadoop$ gedit hadoop-env.sh
osboxes@osboxes:~/hadoop/etc/hadoop$ cd
osboxes@osboxes:~$ cd hadoop/etc/hadoop
osboxes@osboxes:~/hadoop/etc/hadoop$ cd
osboxes@osboxes:~$ gedit .bashrc
osboxes@osboxes:~$ cd hadoop/etc/hadoop
osboxes@osboxes:~/hadoop/etc/hadoop$ gedit hadoop-env.sh
osboxes@osboxes:~/hadoop/etc/hadoop$ gedit mapred-site.xml
osboxes@osboxes:~/hadoop/etc/hadoop$ gedit core-site.xml
osboxes@osboxes:~/hadoop/etc/hadoop$ cd
```

And the configuration for different files are:-

### Gedit Hadoop-env.sh

A snippet from the Hadoop-env.sh file showing configuration for JAVA\_HOME and export commands.

```
35 # For example:
36 #
37 # JAVA_HOME=/usr/java/testing hdfs dfs -ls
38 export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-and64
39 #
40 # Therefore, the vast majority (BUT NOT ALL!) of these defaults
```

### Gedit mapred-site.xml

A snippet from the mapred-site.xml file showing XML configuration for the mapreduce framework name.

```
18
19 <configuration>
20 <property>
21 <name>mapreduce.framework.name</name>
22 <value>yarn</value>
23 </property>
24 </configuration>
```

### Gedit core-site.xml



```

19 <configuration>
20 <property>
21 <name>fs.defaultFS</name>
22 <value>hdfs://localhost:9000</value>
23 <description>The name of the default file system. A URI whose scheme and authority determine the FileSystemImplementation. The uri's scheme determines the configproperty (fs.SCHEME.impl) naming the
  FileSystemImplementation class. The uri's authority is used to determine the host, port, etc. for a filesystem.</description>
24 </property>
25
26 </configuration>

```

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

And in hdfs dir make 2 subdir using mkdir command.

```

osboxes@osboxes:~$ mkdir hdfs
osboxes@osboxes:~$ ls
Desktop  Downloads  hadoop-3.2.1      hdfs  Pictures  Templates
Documents  hadoop    hadoop-3.2.1.tar.gz  Music  Public    Videos
osboxes@osboxes:~$ cd hdfs
osboxes@osboxes:~/hdfs$ mkdir namenode
osboxes@osboxes:~/hdfs$ mkdir datanode
osboxes@osboxes:~/hdfs$ ls
datanode  namenode
osboxes@osboxes:~/hdfs$

```

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

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3 <!--
4 Licensed under the Apache License, Version 2.0 (the "License");
5 you may not use this file except in compliance with the License.
6 You may obtain a copy of the License at
7
8 http://www.apache.org/licenses/LICENSE-2.0
9
10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the License for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20
21 <property>
22 <name>dfs.replication</name>
23 <value>1</value>
24 <description>Default block replication.
25 The actual number of replications can be specified when the file is created. The default is used if replication is not specified in create time.
26 </description>
27 </property>
28 <property>
29 <name>dfs.namenode.name.dir</name>
30 <value>file://home/osboxes/hdfs/namenode</value>
31 </property>
32 <property>
33 <name>dfs.datanode.name.dir</name>
34 <value>file://home/osboxes/hdfs/datanode</value>
35 </property>
36
37 </configuration>

```

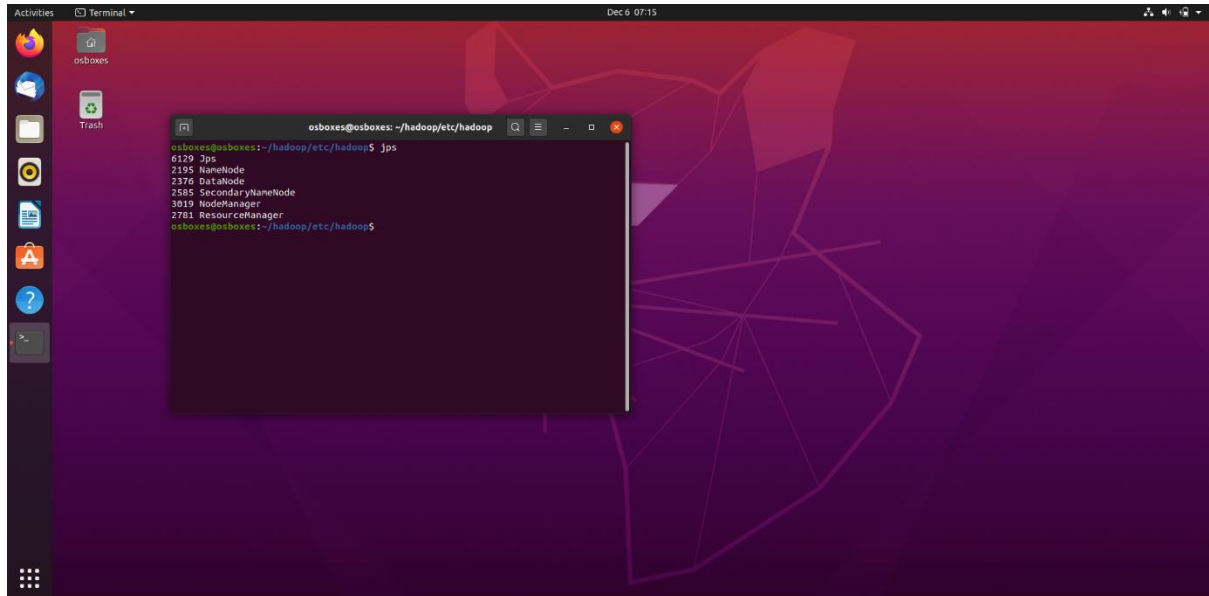
Congratulations You Are done with pseudo mode installation of Hadoop.

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



The screenshot shows a Linux desktop environment with a terminal window open. The terminal window title is 'osboxes@osboxes: ~/hadoop/etc/hadoop'. The command 'jps' has been executed, and the output is as follows:

```
osboxes@osboxes:~/hadoop/etc/hadoop$ jps
6129 Jps
2195 NameNode
2376 DataNode
2585 SecondaryNameNode
3019 NodeManager
2781 ResourceManager
osboxes@osboxes:~/hadoop/etc/hadoop$
```

**Answer2:**

**1.web interfaces of Hadoop Namenode**



Type **localhost:9870**

The screenshot shows the Hadoop NameNode web interface in a Firefox browser. The address bar displays 'localhost:9870/dfshealth.html#tab-overview'. The page contains several sections:

- Number of Blocks Pending Deletion (including replicas):** 0
- Block Deletion Start Time:** Sun Dec 06 02:46:51 -0500 2020
- Last Checkpoint Time:** Sun Dec 06 06:47:04 -0500 2020
- Enabled Erasure Coding Policies:** RS-6-3-1024k
- NameNode Journal Status:**
  - Current transaction ID: 12
  - Journal Manager:** FileJournalManager[root=/home/osboxes/hdfs/namenode]
  - State:** EditLogFileOutputStream[/home/osboxes/hdfs/namenode/current/edits\_inprogress\_00000000000000000012]
- NameNode Storage:**

Storage Directory	Type	State
/home/osboxes/hdfs/namenode	IMAGE_AND_EDITS	Active
- DFS Storage Types:**

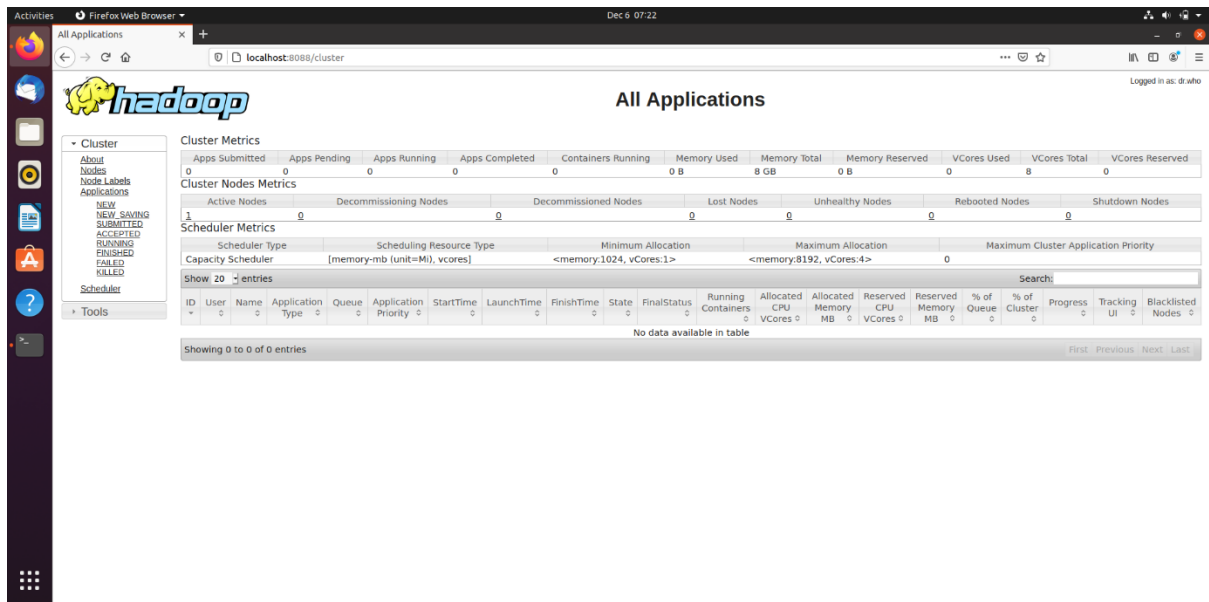
Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	264.84 GB	28 KB (0%)	249.98 GB (94.39%)	28 KB	1

Hadoop, 2019.

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

The screenshot shows the Hadoop Secondary Namenode web interface in a Firefox browser. The address bar displays 'localhost:9868/status.html'. The page has a green header with 'Hadoop' and 'Overview'. Below the header, the text 'Hadoop, 2019.' is visible.

## 3. .web interfaces of Hadoop Resource 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://archive.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'
$shadoop_xml_folder/hadoop-env.sh
```

```
#--Modify core-site.xml
```

```
sudo sed -i -r '/FOOTER/d; s/(<configuration>)/;/
s/(</configuration>)/' $shadoop_xml_folder/core-site.xml
```

```
sudo echo "
```

```
<configuration>
```

```
<property>
```

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

```
<value>hdfs://localhost:9000</value>
```

```
</property>
```

```
</configuration> " >> $shadoop_xml_folder/core-site.xml
```

```
#--Modify yarn-site.xml
```

```
sudo sed -i -r '/FOOTER/d; s/(<configuration>)/;/
s/(</configuration>)/' $shadoop_xml_folder/yarn-site.xml
```

```
sudo echo "
```

```
<configuration>
```

```
<property>
```

```
<name>yarn.nodemanager.aux-services</name>
```

```
<value>mapreduce_shuffle</value>
```

```
</property>
```

```
</configuration> " >> $shadoop_xml_folder/yarn-site.xml
```

```
#--Modify hdfs-site.xml
```

```
sudo sed -i -r '/FOOTER/d; s/(<configuration>)/;/
s/(</configuration>)/' $shadoop_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> " >> $shadoop_xml_folder/hdfs-site.xml

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