Install JDK & Download and install Hadoop on Ubuntu

root@p930026096:~# aptlupdate && apt-get upgrade -y && apt-get install openjdk-8 -jdk openssh-server openssh-client -y

```
root@p930026096:~# apt update && apt-get upgrade -y && apt-get install openjdk-8
-jdk openssh-server openssh-client -y
Hit:1 http://nova.clouds.archive.ubuntu.com/ubuntu focal InRelease
Hit:2 http://security.ubuntu.com/ubuntu focal-security InRelease
Hit:3 http://nova.clouds.archive.ubuntu.com/ubuntu focal-updates InRelease
Hit:4 http://nova.clouds.archive.ubuntu.com/ubuntu focal-backports InRelease
Reading package lists... Done
Building dependency tree
Reading state information... Done
20 packages can be upgraded. Run 'apt list --upgradable' to see them.
Reading package lists... Done
Building dependency tree
Reading state information... Done
Calculating upgrade... Done
The following packages have been kept back:
  fwupd fwupd-signed gnome-shell-extension-desktop-icons libegl-mesa0
  libfwupd2 libfwupdplugin1 libgbm1 libgl1-mesa-dri libglapi-mesa libglx-mesa0
  libxatracker2 mesa-va-drivers mesa-vdpau-drivers mesa-vulkan-drivers
  ubuntu-advantage-tools
The following packages will be upgraded:
 libpam-modules libpam-modules-bin libpam-runtime libpam0g squashfs-tools
5 upgraded, 0 newly installed, 0 to remove and 15 not upgraded.
```

```
Need to get 512 kB of archives.
After this operation, 0 B of additional disk space will be used.
Get:1 http://nova.clouds.archive.ubuntu.com/ubuntu focal-updates/main amd64 libp
am0g amd64 1.3.1-5ubuntu4.3 [55.4 kB]
Get:2 http://nova.clouds.archive.ubuntu.com/ubuntu focal-updates/main amd64 libp
am-modules-bin amd64 1.3.1-5ubuntu4.3 [41.2 kB]
Get:4 http://nova.clouds.archive.ubuntu.com/ubuntu focal-updates/main amd64 libp
am-runtime all 1.3.1-5ubuntu4.3 [37.3 kB]
Get:5 http://nova.clouds.archive.ubuntu.com/ubuntu focal-updates/main amd64 squa
shfs-tools amd64 1:4.4-1ubuntu0.3 [117 kB]
Fetched 512 kB in 3s (201 kB/s)
Preconfiguring packages ...
(Reading database ... 235630 files and directories currently installed.)
Preparing to unpack .../libpam0g_1.3.1-5ubuntu4.3 amd64.deb ...
Unpacking libpam0g:amd64 (1.3.1-5ubuntu4.3) over (1.3.1-5ubuntu4.2) ...
Setting up libpam0g:amd64 (1.3.1-5ubuntu4.3) ...
(Reading database ... 235630 files and directories currently installed.)
Preparing to unpack .../libpam-modules-bin 1.3.1-5ubuntu4.3 amd64.deb ...
Unpacking libpam-modules-bin (1.3.1-5ubuntu4.3) over (1.3.1-5ubuntu4.2) ...
Setting up libpam-modules-bin (1.3.1-5ubuntu4.3) ...
(Reading database ... 235630 files and directories currently installed.)
Preparing to unpack .../libpam-modules 1.3.1-5ubuntu4.3 amd64.deb ...
```

```
Get:1 http://nova.clouds.archive.ubuntu.com/ubuntu focal-updates/universe amd64
openjdk-8-jre amd64 8u292-b10-0ubuntu1~20.04 [70.0 kB]
Get:2 http://nova.clouds.archive.ubuntu.com/ubuntu focal-updates/universe amd64
openjdk-8-jdk amd64 8u292-b10-0ubuntu1~20.04 [2895 kB]
Fetched 2965 kB in 3s (1095 kB/s)
Selecting previously unselected package openjdk-8-jre:amd64. (Reading database ... 235630 files and directories currently installed.)
Preparing to unpack .../openjdk-8-jre_8u292-b10-0ubuntu1~20.04_amd64.deb ...
Unpacking openjdk-8-jre:amd64 (8u292-b10-0ubuntu1~20.04) ...
Selecting previously unselected package openjdk-8-jdk:amd64.
Preparing to unpack .../openjdk-8-jdk_8u292-b10-0ubuntu1~20.04_amd64.deb ...
Unpacking openj4k-8-jdk:amd64 (8u292-b10-0ubuntu1~20.04) ...
Setting up openjdk-8-jre:amd64 (8u292-b10-0ubuntu1~20.04) ..
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/policytool
to provide /usr/bin/policytool (policytool) in auto mode
Setting up openjdk-8-jdk:amd64 (8u292-b10-0ubuntu1~20.04) ...
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/appletviewer to
provide /usr/bin/appletviewer (appletviewer) in auto mode
Processing triggers for mime-support (3.64ubuntul) ...
Processing triggers for hicolor-icon-theme (0.17-2) ...
Processing triggers for gnome-menus (3.36.0-lubuntul) ...
Processing triggers for libc-bin (2.31-Oubuntu9.2) ...
Processing triggers_for desktop-file-utils (0.24-1ubuntu3) ...
root@p930026096:~#
```

Update the system and install JRE and JDK.

root@p930026096:~# adduser hduser

```
root@p930026096:~# adduser hduser
Adding user `hduser' ...
Adding new group `hduser' (1001) ...
Adding new user `hduser' (1001) with group `hduser' ...
Creating home directory `/home/hduser'
Copying files from `/etc/skel' ...
New password:
Retype new password:
passwd: password updated successfully
Changing the user information for hduser
Enter the new value, or press ENTER for the default
        Full Name []: RoyMa
        Room Number []:
        Work Phone []:
        Home Phone []:
        Other []:
Is the information correct? [Y/n] y
root@p930026096:~#
```

Create a Hadoop user. The username is RoyMa with no room number, work phoe and home phone and others.

root@p930026096:~# su_- hduser

SSH the localhost.

```
hduser@p930026096:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id rsa
```

Generate an SSH key pair and define the storage location.

```
hduser@p930026096:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
hduser@p930026096:~$ chmod 0600 ~/.ssh/authorized_keys
```

"cat" to move public key as authorized_keys to the ssh directory. Set the permissions by "chmod" command.

```
hduser@p930026096:~$ ssh localhost
```

Establish SSH connection to the localhost.

```
hduser@p930026096:-$ wget https://downloads.apache.org/hadoop/common/hadoop-3.2.
2/hadoop-3.2.2.tar.gz
```

Download the Hadoop 3.2.2 mirrors.

hduser@p930026096:~\$ tar xzf hadoop-3.2.2.tar.gz

Extract the zip files.

hduser@p930026096:~\$ nano .bashrc

Edit .bashrc file, adding some environment variables as follow.

```
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
export HADOOP HOME=/home/hduser/hadoop-3.2.2
export HADOOP_INSTALL=$HADOOP HOME
export HADOOP MAPRED HOME=$HADOOP HOME
export HADOOP COMMON HOME=$HADOOP HOME
export HADOOP HDFS HOME=$HADOOP HOME
export YARN HOME=$HADOOP HOME
export HADOOP COMMON LIB NATIVE DIR=$HADOOP HOMET/lib/native
export PATH=$path:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
`G Get Help
                                       ^K Cut Text
                                                    Justify
            ^0 Write Out ^W Where Is
                                                                  ^C Cur Pos
               Read File
                            Replace
                                          Paste Text<sup>T</sup> To Spell
```

hduser@p930026096:~\$ source ~/.bashrc

Apply the changes of .bashrc.

hduser@p930026096:~\$ nano \$HADOOP_HOME/etc/hadoop/hadoop-env.sh Edit Hadoop-env.sh file as follow.

```
## (YARN_xyz|HDFS_xyz} > hADDOP_xyz > hard-coded defaults

## (YARN_xyz|HDFS_xyz} > hADDOP_xyz > hard-coded defaults

## Many of the options here are built from the perspective that users

## Many of the options here are built from the perspective that users

## Many vant to provide OVERORITING values on the command line.

## For example:

## JAVA_HOME=/usr/java/testing hdfs dfs -ls

## Therefore, the vast majority (BUT NOT ALL!) of these defaults

## are configured for substitution and not append. If append

## Generic settings for HADOOP

###

## Generic settings for HADOOP

###

## Technically, the only required environment variable is JAVA_HOME.

## All others are optional. However, the defaults are probably not

## preferred. Nany sites configure these options outside of Hadoop,

## such as in /etc/profile.d

## The java implementation to use. By default, this environment

**variable is ENGUIRED on All_platforms except OS Xi

**export JAVA_HOME=/usr/lib/jwm/java-B-openjdk-andded|

**Location of Hadoop, By default, Hadoop will attempt to determine

## this location based upon its execution path.

## Export JAVA_HOME—

**Location of Madoop's configuration information. i.e., where this

## Location of Madoop's configuration information. i.e., where this

## Location of Madoop's configuration information. i.e., where this

## Location of Madoop's configuration information. i.e., where this

## Location of Madoop's configuration information. i.e., where this

## Location of Madoop's configuration information. i.e., where this

## Location of Madoop's configuration information. i.e., where this

## Location of Madoop's configuration information. i.e., where this

## Location of Madoop's configuration information. i.e., where this

## Location of Madoop's configuration information. i.e., where this

## Location of Madoop's configuration information. i.e., where this

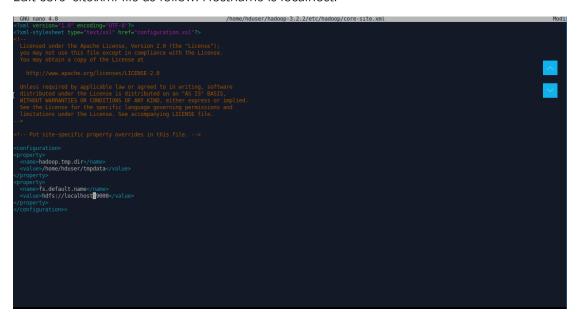
## Location of Madoop's Configuration information. i.e., where this

## Location of Madoop's Configuration information. i.e., where this

## Locatio
```

hduser@p930026096:~\$ nano \$HADOOP_HOME/etc/hadoop/core-site.xml

Edit core-site.xml file as follow. Hostname is localhost.



hduser@p930026096:~\$ nano \$HADOOP_HOME/etc/hadoop/hdfs-site.xml
Edit hdfs-site.xml file as follow.

```
### Anno 1.8

**Chamber 1.8* encoding="UTF-8*">

**Line Line Head and the Line Head and t
```

hduser@p930026096:~\$ nano \$HADOOP_HOME/etc/hadoop/mapred-site.xml

Edit mapred-site.xml as follow.

```
GRU mano 4.8

//mome/hduser/hadoop-3.2.2/etc/hadoop/mapred-site.xml

//mome/hduser/hadoop-3.2.2/etc/hadoop-file.xml

//mome/hduser/hadoop-file.xml

//mome/hduser/hadoop-file.xml
```

iduser@p930026096:~\$ nano \$HADOOP_HOME/etc/hadoop/yarn-site.xml

Edit yarn-site.xml file as follow. Hostname is localhost.

hduser@p930026096:~\$ hdfs namenode -format

Format the namenode at the first time starting Hadoop services.

```
### STARTUP MGG: Starting NameWork | Startup NameWo
```

```
201-10-18 12:04-15 607 1000 conf Configuration: error paraing conf core-site.an

Con ctc.vsts.sc.int.dusspectaclosessed and immoreted data-dure? > code 80 in epilog; expected '<'
at [row.col.ysts.sc.int]; 1821,71*16; // home/durer/ladops.2.co.sc.int.and.

at con.ctc.vsts.sr.StreamScanner.throubsexpectedChartStreamScanner.java:653)
at con.ctc.vsts.sr.StreamScanner.throubsexpectedChartStreamScanner.java:653)
at con.ctc.vsts.sr.StreamScanner.throubsexpectedChartStreamScanner.java:653)
at con.ctc.vsts.sr.StreamScanner.throubsexpectedChartStreamScanner.java:653)
at con.ctc.vsts.sr.StreamScanner.throubsexpectedChartStreamScanner.java:653)
at con.gapache.hadop.conf.ConfigurationsexpectedConfiguration.java:3300)
at con.gapache.hadop.conf.ConfigurationsexpectedConfiguration.java:3300)
at con.gapache.hadop.conf.Configuration.java:3300
at con.gapache.hadop.conf.Configuration.java:3200
at con.gapache.hadop.conf.Configuration.java:3200
at con.gapache.hadop.conf.Configuration.java:2000
at con.ga
```

```
hduser@p930026096:~/hadoop-3.2.2/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [p930026096]
```

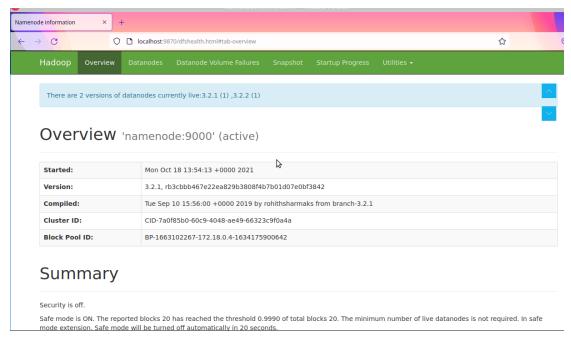
Go to the directory "~/Hadoop-3..2.2.sbin", and start the namenode and datanode.

```
hduser@p930026096:~/hadoop-3.2.2/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
```

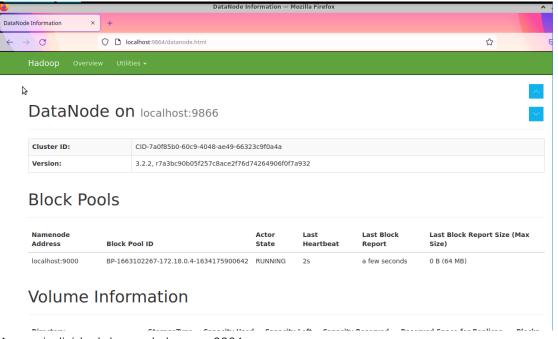
Start the yarn resource and nodemanagers

```
hduser@p930026096:~/hadoop-3.2.2/sbin$ jps
jpjp56896 NodeManager
56193 SecondaryNameNode
57748 Jps
55963 DataNode
56526 ResourceManager
```

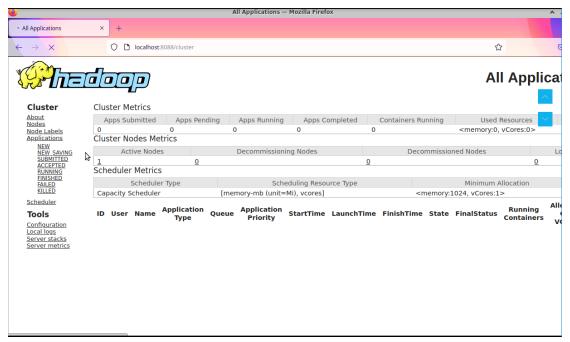
Check if all the daemons are active and running as java processes.



Access the Hadoop namenode UI by port 9870.



Access individual datanode by port 9864.



Access yarn resource manager by port 8088.

Successfully installed Hadoop on Ubuntu!