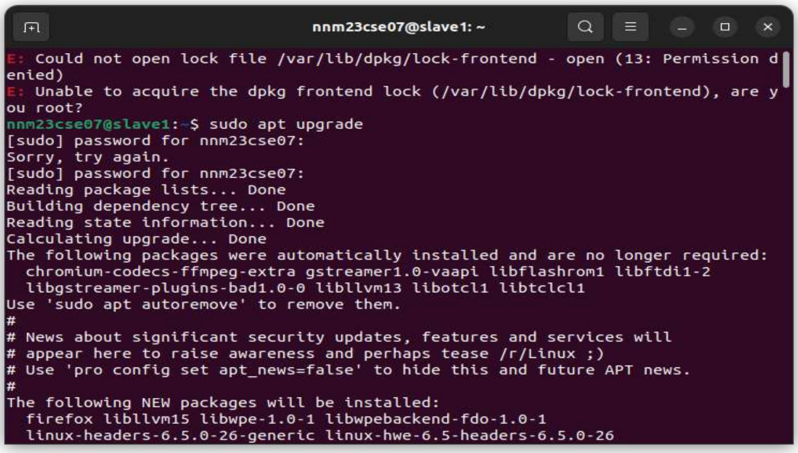
LAB PROGRAM PREREQUISITES: Installation of Hadoop

Link: https://www.youtube.com/watch?v=Slbi-uzPtnw

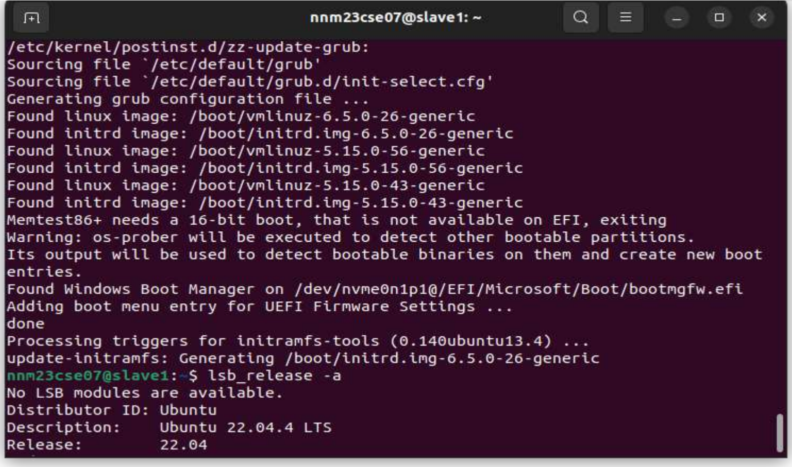
sudo apt upgrade: This is the correct usage to upgrade all upgradable packages on a Ubuntu system before installing Hadoop

The terminal asks for the password for the user, and after the password is entered, the command proceeds.



```
E: Could not open lock file /var/lib/dpkg/lock-frontend - open (13: Permission d
enied)
E: Unable to acquire the dpkg frontend lock (/var/lib/dpkg/lock-frontend), are y
ou root?
nnm23cse07@slave1:~$ sudo apt upgrade
[sudo] password for nnm23cse07:
Sorry, try again.
[sudo] password for nnm23cse07:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
Calculating upgrade... Done
The following packages were automatically installed and are no longer required:
  chromium-codecs-ffmpeg-extra gstreamer1.0-vaapi libflashrom1 libftdi1-2
  libgstreamer-plugins-bad1.0-0 libllvm13 libotcl1 libtclcl1
Use 'sudo apt autoremove' to remove them.
#
# News about significant security updates, features and services will
# appear here to raise awareness and perhaps tease /r/Linux ;)
# Use 'pro config set apt_news=false' to hide this and future APT news.
#
The following NEW packages will be installed:
  firefox libllvm15 libwpe-1.0-1 libwpebackend-fdo-1.0-1
  linux-headers-6.5.0-26-generic linux-hwe-6.5-headers-6.5.0-26
```

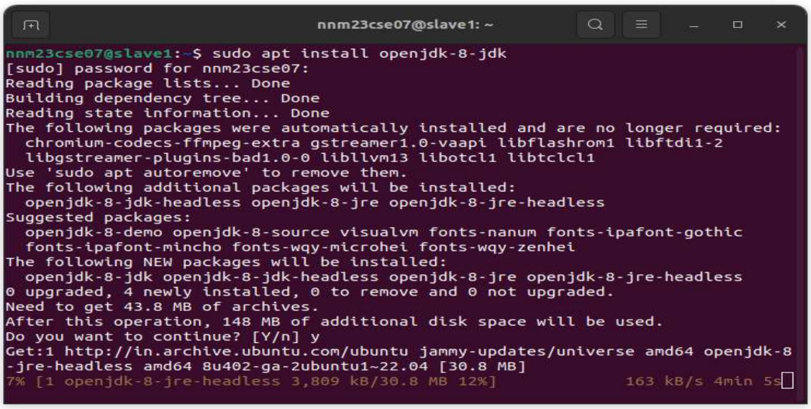
- **Lsb_release -a**: This command lists information about the Linux distribution.
- Output like **Distributor ID: Ubuntu**, **Description: Ubuntu 22.04.4 LTS**, **Release: 22.04**



```
/etc/kernel/postinst.d/zz-update-grub:
Sourcing file `/etc/default/grub'
Sourcing file `/etc/default/grub.d/init-select.cfg'
Generating grub configuration file ...
Found linux image: /boot/vmlinuz-6.5.0-26-generic
Found initrd image: /boot/initrd.img-6.5.0-26-generic
Found linux image: /boot/vmlinuz-5.15.0-56-generic
Found initrd image: /boot/initrd.img-5.15.0-56-generic
Found linux image: /boot/vmlinuz-5.15.0-43-generic
Found initrd image: /boot/initrd.img-5.15.0-43-generic
Memtest86+ needs a 16-bit boot, that is not available on EFI, exiting
Warning: os-prober will be executed to detect other bootable partitions.
Its output will be used to detect bootable binaries on them and create new boot
entries.
Found Windows Boot Manager on /dev/nvme0n1p1@/EFI/Microsoft/Boot/bootmgfw.efi
Adding boot menu entry for UEFI Firmware Settings ...
done
Processing triggers for initramfs-tools (0.140ubuntu13.4) ...
update-initramfs: Generating /boot/initrd.img-6.5.0-26-generic
nnm23cse07@slave1:~$ lsb_release -a
No LSB modules are available.
Distributor ID: Ubuntu
Description:    Ubuntu 22.04.4 LTS
Release:        22.04
```

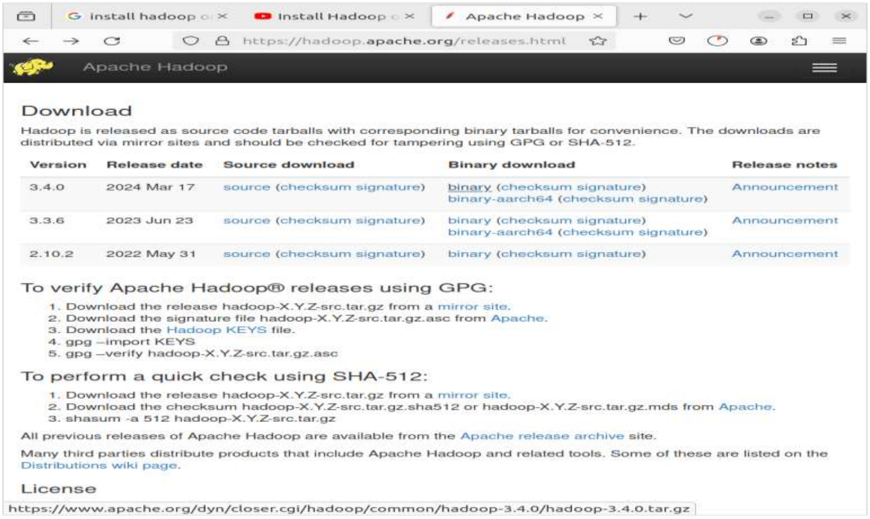
sudo apt install openjdk-8-jdk: The user is attempting to install the OpenJDK 8 JDK (Java Development Kit) using **apt**, the package manager for Ubuntu.

[sudo] password for nnn23cse07:: After entering the **sudo** command, the system prompts the user to enter the password for the **nnn23cse07**



```
nnm23cse07@slave1:~$ sudo apt install openjdk-8-jdk
[sudo] password for nnm23cse07:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following packages were automatically installed and are no longer required:
  chromium-codecs-ffmpeg-extra gstreamer1.0-vaapi libflashrom1 libftdi1-2
  libgstreamer-plugins-bad1.0-0 libllvm13 libotcl1 libtclcl1
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless
Suggested packages:
  openjdk-8-demo openjdk-8-source visualvm fonts-nanum fonts-ipafont-gothic
  fonts-ipafont-mincho fonts-wqy-microhei fonts-wqy-zenhei
The following NEW packages will be installed:
  openjdk-8-jdk openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless
0 upgraded, 4 newly installed, 0 to remove and 0 not upgraded.
Need to get 43.8 MB of archives.
After this operation, 148 MB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://in.archive.ubuntu.com/ubuntu jammy-updates/universe amd64 openjdk-8
-jre-headless amd64 8u402-ga-2ubuntu1~22.04 [30.8 MB]
7% [1 openjdk-8-jre-headless 3,809 kB/30.8 MB 12%]            163 kB/s 4min 5s
```

Download hadoop apache version 3.4.0 from https//hadoop.apache.org/release html



Download

Hadoop is released as source code tarballs with corresponding binary tarballs for convenience. The downloads are distributed via mirror sites and should be checked for tampering using GPG or SHA-512.

Version	Release date	Source download	Binary download	Release notes
3.4.0	2024 Mar 17	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
3.3.6	2023 Jun 23	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
2.10.2	2022 May 31	source (checksum signature)	binary (checksum signature)	Announcement

To verify Apache Hadoop® releases using GPG:

1. Download the release hadoop-X.Y.Z-src.tar.gz from a mirror site.
2. Download the signature file hadoop-X.Y.Z-src.tar.gz.asc from Apache.
3. Download the Hadoop KEYS file.
4. gpg –import KEYS
5. gpg –verify hadoop-X.Y.Z-src.tar.gz.asc

To perform a quick check using SHA-512:

1. Download the release hadoop-X.Y.Z-src.tar.gz from a mirror site.
2. Download the checksum hadoop-X.Y.Z-src.tar.gz.sha512 or hadoop-X.Y.Z-src.tar.gz.mds from Apache.
3. shasum -a 512 hadoop-X.Y.Z-src.tar.gz

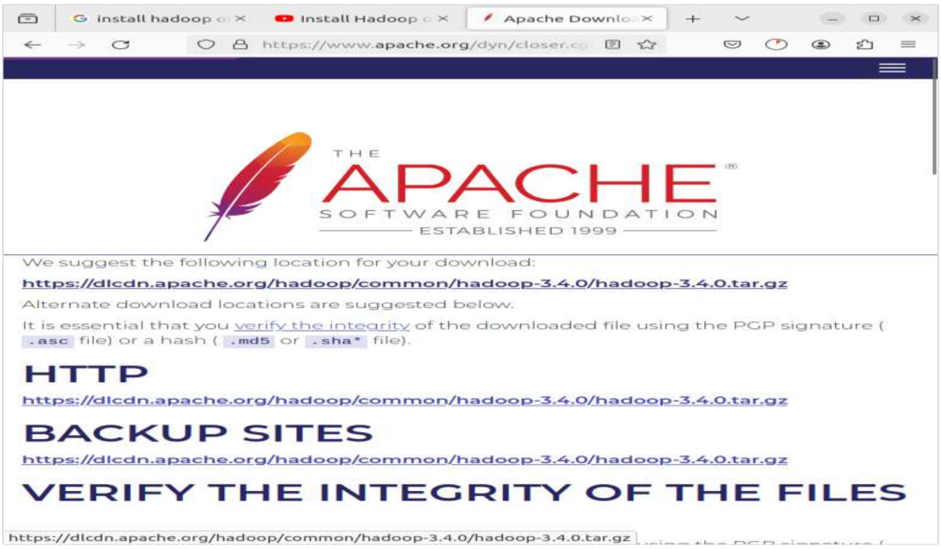
All previous releases of Apache Hadoop are available from the Apache release archive site.

Many third parties distribute products that include Apache Hadoop and related tools. Some of these are listed on the Distributions wiki page.

License

https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz

direct link for downloading a specific version of Apache Hadoop, which is version 3.4.0. The URL provided is https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz



install hadoop o
Install Hadoop o
Apache Downlo
https://www.apache.org/dyn/closer.cg
THE APACHE SOFTWARE FOUNDATION
ESTABLISHED 1999

We suggest the following location for your download:

https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz

Alternate download locations are suggested below.

It is essential that you verify the integrity of the downloaded file using the PGP signature (.asc file) or a hash (.md5 or .sha* file).

HTTP

https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz

BACKUP SITES

https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz

VERIFY THE INTEGRITY OF THE FILES

https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz

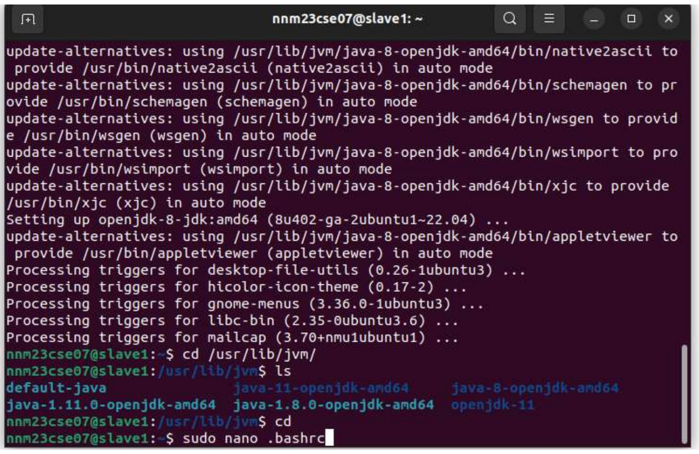
cd /usr/lib/jvm/

- **cd** stands for "change directory". This command changes the current directory to **/usr/lib/jvm/**, which is typically where Java installations are located on a Unix-like operating system.

ls

- **ls** stands for "list". It lists the contents of the current directory. After changing to the **/usr/lib/jvm/** directory

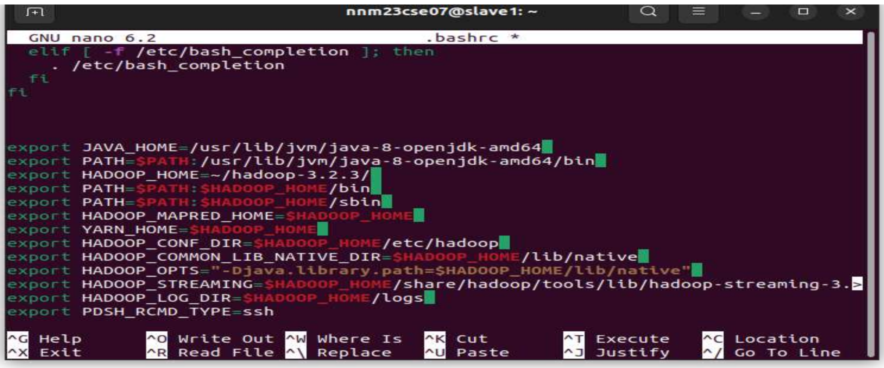
sudo nano .bashrc, is used to edit the **.bashrc** file in the user's home directory using the **nano** text editor



```
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/native2ascii to
 provide /usr/bin/native2ascii (native2ascii) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/schemagen to pr
ovide /usr/bin/schemagen (schemagen) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/wsgen to provid
e /usr/bin/wsgen (wsgen) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/wsimport to pro
vide /usr/bin/wsimport (wsimport) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/xjc to provide
/usr/bin/xjc (xjc) in auto mode
Setting up openjdk-8-jdk:amd64 (8u402-ga-2ubuntu1~22.04) ...
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/appletviewer to
 provide /usr/bin/appletviewer (appletviewer) in auto mode
Processing triggers for desktop-file-utils (0.26-1ubuntu3) ...
Processing triggers for hicolor-icon-theme (0.17-2) ...
Processing triggers for gnome-menus (3.36.0-1ubuntu3) ...
Processing triggers for libc-bin (2.35-0ubuntu3.6) ...
Processing triggers for mailcap (3.70+nmu1ubuntu1) ...
nnm23cse07@slave1:~$ cd /usr/lib/jvm/
nnm23cse07@slave1:/usr/lib/jvm$ ls
default-java               java-11-openjdk-amd64     java-8-openjdk-amd64
java-1.11.0-openjdk-amd64  java-1.8.0-openjdk-amd64  openjdk-11
nnm23cse07@slave1:/usr/lib/jvm$ cd
nnm23cse07@slave1:~$ sudo nano .bashrc
```

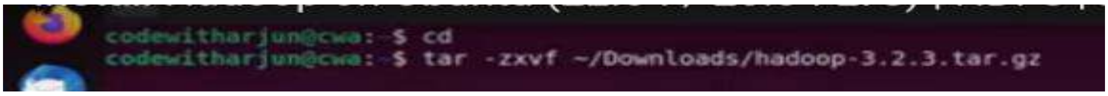
Copy following export commands in **.bashrc** file which opened after this **sudo nano .bashrc command**

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

export PATH=$PATH:/usr/lib/jvm/java-8-openjdk-amd64/bin

export HADOOP_HOME=/hadoop-3.2.3/

export PATH=$PATH:$HADOOP_HOME/bin

export PATH=$PATH:$HADOOP_HOME/sbin

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_CONF_DIR=$HADOOP_HOME/etc/hadoop

export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native

export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

export HADOOP_STREAMING=$HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-3.2.3.jar

export HADOOP_LOG_DIR=$HADOOP_HOME/logs

export PDSH_RCMD_TYPE=ssh
```



```
nnm23cse07@slave1: ~
  GNU nano 6.2                        .bashrc *
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi


export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PATH=$PATH:/usr/lib/jvm/java-8-openjdk-amd64/bin
export HADOOP_HOME=~/hadoop-3.2.3/
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HADOOP_STREAMING=$HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-3.>
export HADOOP_LOG_DIR=$HADOOP_HOME/logs
export PDSH_RCMD_TYPE=ssh

^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^/ Go To Line
```

The **tar -zxvf /Downloads/hadoop-3.2.3.tar.gz** command is used to extract the contents of a **tar.gz .** this command extracts all the files from the **hadoop-3.2.3.tar.gz** archive, located in the **/Downloads**



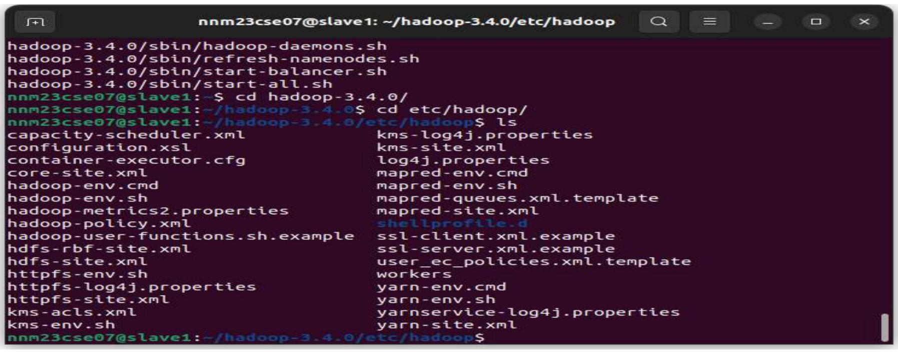
```
codewitharjun@cwa:~$ cd
codewitharjun@cwa:~$ tar -zxvf ~/Downloads/hadoop-3.2.3.tar.gz
```

Once Hadoop is installed, administrators need to configure the system before starting it up and using it.

cd /hadoop-3.4.0/: This command changes the current working directory to the Hadoop installation directory

cd etc/hadoop/: Assuming the user is currently in the **/hadoop-3.4.0** directory, this command changes the directory to **etc/hadoop**

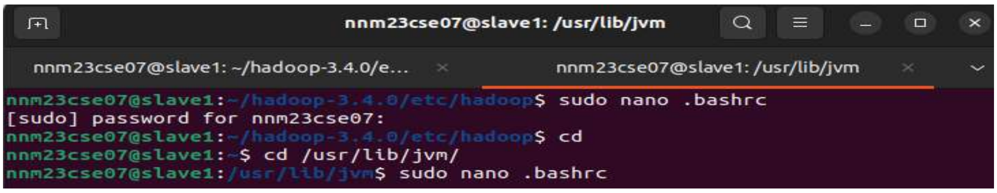
ls: This lists the contents of the current directory, which would be **/hadoop-3.4.0/etc/hadoop**. The output shows various configuration files for Hadoop



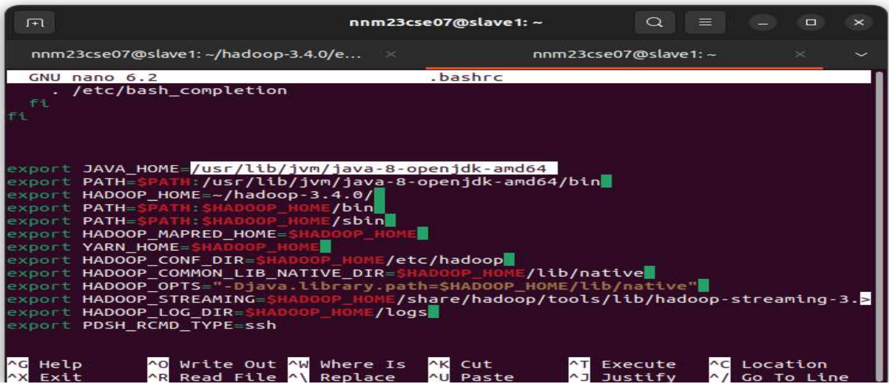
```
nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop
hadoop-3.4.0/sbin/hadoop-daemons.sh
hadoop-3.4.0/sbin/refresh-namenodes.sh
hadoop-3.4.0/sbin/start-balancer.sh
hadoop-3.4.0/sbin/start-all.sh
nnm23cse07@slave1:~$ cd hadoop-3.4.0/
nnm23cse07@slave1:~/hadoop-3.4.0$ cd etc/hadoop/
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ ls
capacity-scheduler.xml                  kms-log4j.properties
configuration.xsl                       kms-site.xml
container-executor.cfg                  log4j.properties
core-site.xml                           mapred-env.cmd
hadoop-env.cmd                          mapred-env.sh
hadoop-env.sh                           mapred-queues.xml.template
hadoop-metrics2.properties              mapred-site.xml
hadoop-policy.xml                       shellprofile.d
hadoop-user-functions.sh.example        ssl-client.xml.example
hdfs-rbf-site.xml                       ssl-server.xml.example
hdfs-site.xml                           user_ec_policies.xml.template
httpfs-env.sh                           workers
httpfs-log4j.properties                 yarn-env.cmd
httpfs-site.xml                         yarn-env.sh
kms-acls.xml                            yarnservice-log4j.properties
kms-env.sh                              yarn-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$
```

sudo nano .bashrc: is used to open the **.bashrc** file in the **nano** text editor with superuser permissions.

cd /usr/lib/jvm/:changes the current working directory to the Java virtual machine installation directory, where different versions of Java can be found.



```
nnm23cse07@slave1: /usr/lib/jvm
nnm23cse07@slave1: ~/hadoop-3.4.0/e...      nnm23cse07@slave1: /usr/lib/jvm
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano .bashrc
[sudo] password for nnm23cse07:
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ cd
nnm23cse07@slave1:~$ cd /usr/lib/jvm/
nnm23cse07@slave1:/usr/lib/jvm$ sudo nano .bashrc
```

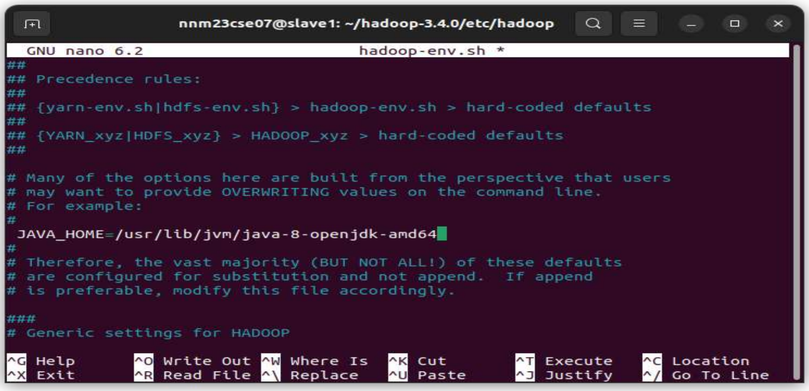


```
  GNU nano 6.2                          .bashrc
    . /etc/bash_completion
  fi
fi


export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PATH=$PATH:/usr/lib/jvm/java-8-openjdk-amd64/bin
export HADOOP_HOME=~/hadoop-3.4.0/
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HADOOP_STREAMING=$HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-3.>
export HADOOP_LOG_DIR=$HADOOP_HOME/logs
export PDSH_RCMD_TYPE=ssh

^G Help       ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location
^X Exit       ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^/ Go To Line
```

sudo nano hadoop-env.h :by using this command following text editor opens

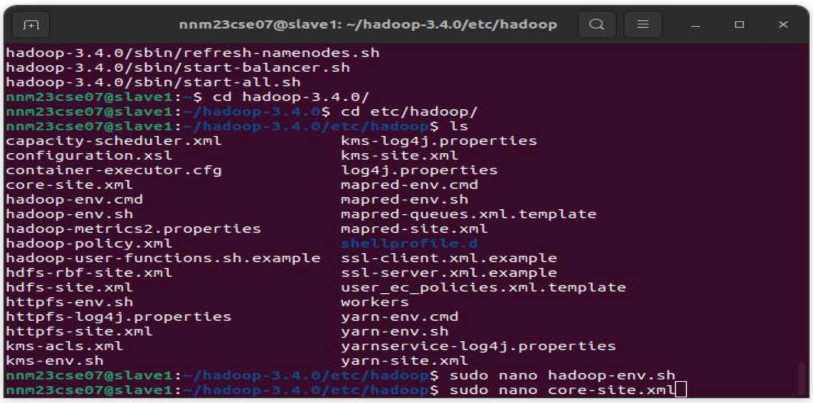
JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64: copy this command as below text editor to set the path for JAVA_HOME

Then press Control o + enter +control x for saving



```
  GNU nano 6.2                    hadoop-env.sh *
##
## Precedence rules:
##
## {yarn-env.sh|hdfs-env.sh} > hadoop-env.sh > hard-coded defaults
##
## {YARN_xyz|HDFS_xyz} > HADOOP_xyz > hard-coded defaults
##

# Many of the options here are built from the perspective that users
# may want to provide OVERWRITING values on the command line.
# For example:
#
 JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
#
# Therefore, the vast majority (BUT NOT ALL!) of these defaults
# are configured for substitution and not append.  If append
# is preferable, modify this file accordingly.

###
# Generic settings for HADOOP
^G Help       ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location
^X Exit       ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^/ Go To Line
```

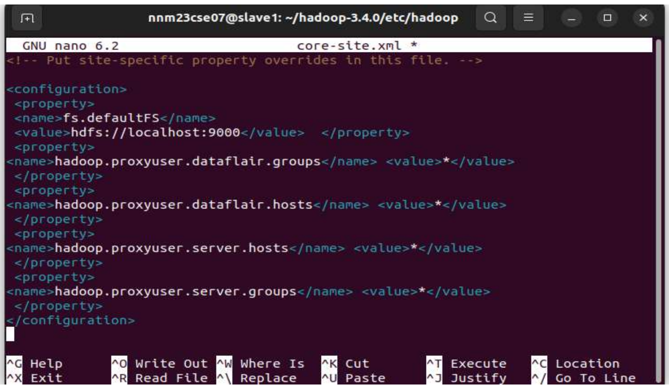
sudo nano /etc/hadoop/core-site.xml :use this command which will open text editor show in below screenshot . **core-site.xml** file is a key configuration file for Hadoop services



```
hadoop-3.4.0/sbin/refresh-namenodes.sh
hadoop-3.4.0/sbin/start-balancer.sh
hadoop-3.4.0/sbin/start-all.sh
nnm23cse07@slave1:~$ cd hadoop-3.4.0/
nnm23cse07@slave1:~/hadoop-3.4.0$ cd etc/hadoop/
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ ls
capacity-scheduler.xml                  kms-log4j.properties
configuration.xsl                       kms-site.xml
container-executor.cfg                  log4j.properties
core-site.xml                           mapred-env.cmd
hadoop-env.cmd                          mapred-env.sh
hadoop-env.sh                           mapred-queues.xml.template
hadoop-metrics2.properties              mapred-site.xml
hadoop-policy.xml                       shellprofile.d
hadoop-user-functions.sh.example        ssl-client.xml.example
hdfs-rbf-site.xml                       ssl-server.xml.example
hdfs-site.xml                           user_ec_policies.xml.template
httpfs-env.sh                           workers
httpfs-log4j.properties                 yarn-env.cmd
httpfs-site.xml                         yarn-env.sh
kms-acls.xml                            yarnservice-log4j.properties
kms-env.sh                              yarn-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano hadoop-env.sh
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano core-site.xml
```

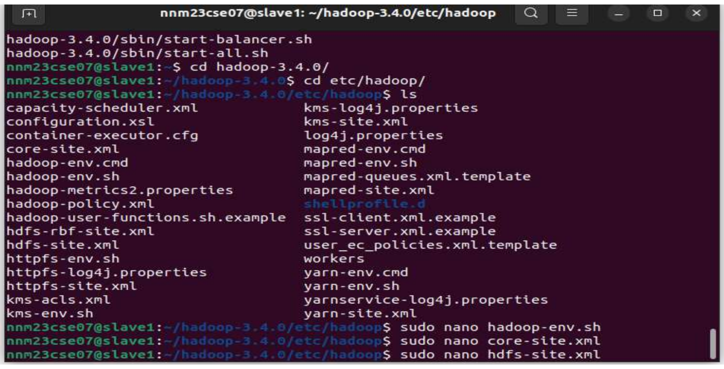
Add below command line in as shown screenshot next to this code

```
<configuration><property>
<name>fs.defaultFS</name>
<value>hdfs://localhost:9000</value></property><property>
<name>hadoop.proxyuser.dataflair.groups</name><value>*</value></property><property>
<name>hadoop.proxyuser.dataflair.hosts</name><value>*</value></property>
<property><name>hadoop.proxyuser.server.hosts</name><value>*</value>
</property><property><name>hadoop.proxyuser.server.groups</name><value>*</value>
</property> </configuration>
```



```
nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop
  GNU nano 6.2                    core-site.xml *
<!-- Put site-specific property overrides in this file. -->

<configuration>
 <property>
 <name>fs.defaultFS</name>
 <value>hdfs://localhost:9000</value>  </property>
 <property>
<name>hadoop.proxyuser.dataflair.groups</name> <value>*</value>
 </property>
 <property>
<name>hadoop.proxyuser.dataflair.hosts</name> <value>*</value>
 </property>
 <property>
<name>hadoop.proxyuser.server.hosts</name> <value>*</value>
 </property>
 <property>
<name>hadoop.proxyuser.server.groups</name> <value>*</value>
 </property>
</configuration>

^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^/ Go To Line
```

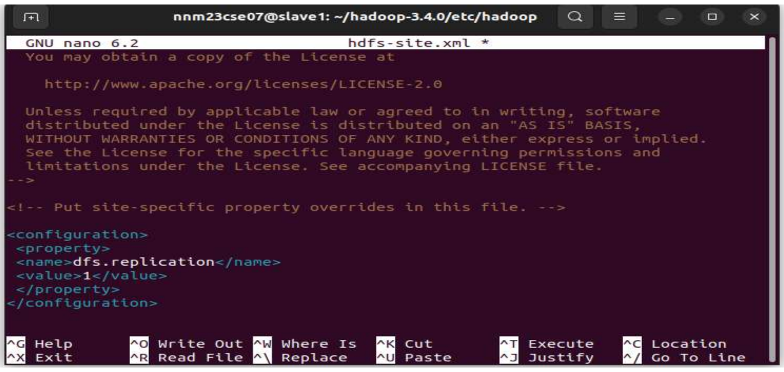
sudo nano /etc/hadoop/hdfs-site.xml   :use following command as in screen shot



```
nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop
hadoop-3.4.0/sbin/start-balancer.sh
hadoop-3.4.0/sbin/start-all.sh
nnm23cse07@slave1:~$ cd hadoop-3.4.0/
nnm23cse07@slave1:~/hadoop-3.4.0$ cd etc/hadoop/
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ ls
capacity-scheduler.xml                kms-log4j.properties
configuration.xsl                     kms-site.xml
container-executor.cfg                log4j.properties
core-site.xml                         mapred-env.cmd
hadoop-env.cmd                        mapred-env.sh
hadoop-env.sh                         mapred-queues.xml.template
hadoop-metrics2.properties            mapred-site.xml
hadoop-policy.xml                     shellprofile.d
hadoop-user-functions.sh.example      ssl-client.xml.example
hdfs-rbf-site.xml                     ssl-server.xml.example
hdfs-site.xml                         user_ec_policies.xml.template
httpfs-env.sh                         workers
httpfs-log4j.properties               yarn-env.cmd
httpfs-site.xml                       yarn-env.sh
kms-acls.xml                          yarnservice-log4j.properties
kms-env.sh                            yarn-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano hadoop-env.sh
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano core-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano hdfs-site.xml
```

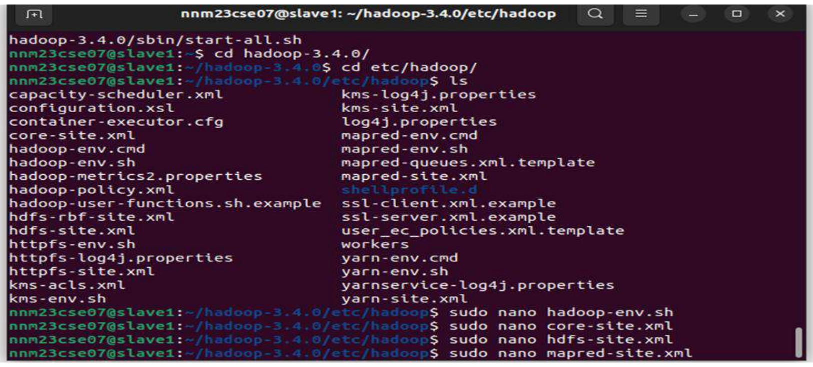
Add this command as shown in screenshoot

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
</configuration>
```



```
  GNU nano 6.2                          hdfs-site.xml *
  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>dfs.replication</name>
 <value>1</value>
 </property>
</configuration>
```

sudo nano /etc/hadoop/mapred-site.xml : this command line used as shown below



```
hadoop-3.4.0/sbin/start-all.sh
nnm23cse07@slave1:~$ cd hadoop-3.4.0/
nnm23cse07@slave1:~/hadoop-3.4.0$ cd etc/hadoop/
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ ls
capacity-scheduler.xml                  kms-log4j.properties
configuration.xsl                       kms-site.xml
container-executor.cfg                  log4j.properties
core-site.xml                           mapred-env.cmd
hadoop-env.cmd                          mapred-env.sh
hadoop-env.sh                           mapred-queues.xml.template
hadoop-metrics2.properties              mapred-site.xml
hadoop-policy.xml                       shellprofile.d
hadoop-user-functions.sh.example        ssl-client.xml.example
hdfs-rbf-site.xml                       ssl-server.xml.example
hdfs-site.xml                           user_ec_policies.xml.template
httpfs-env.sh                           workers
httpfs-log4j.properties                 yarn-env.cmd
httpfs-site.xml                         yarn-env.sh
kms-acls.xml                            yarnservice-log4j.properties
kms-env.sh                              yarn-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano hadoop-env.sh
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano core-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano hdfs-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano mapred-site.xml
```

write following command line in text editor shown below shown screenshot

<configuration>

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

</property><property>

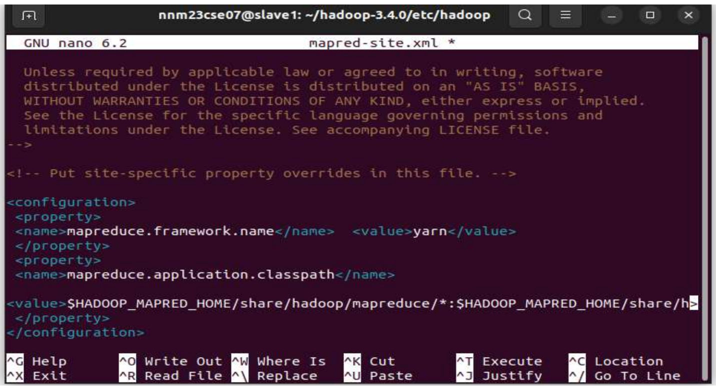
<name>mapreduce.application.classpath</name>

<value>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/lib/*</value>

</property>

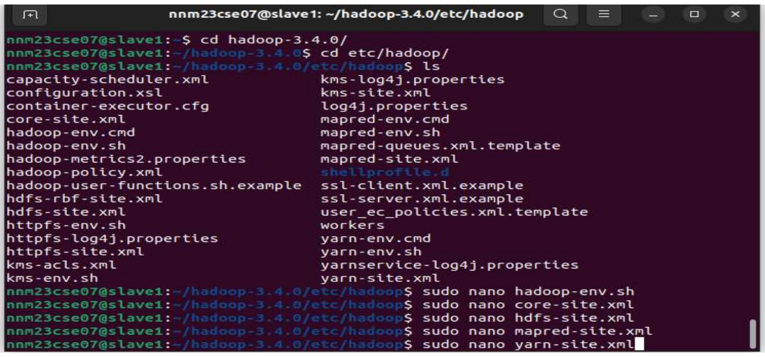
</configuration>

then press control O + Enter +control X



```
  GNU nano 6.2                          mapred-site.xml *

  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>
 <property>
 <name>mapreduce.application.classpath</name>

<value>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPRED_HOME/share/h>
 </property>
</configuration>
```

sudo nano /etc/hadoop/yarn-site.xml :Enter this command as below screenshot

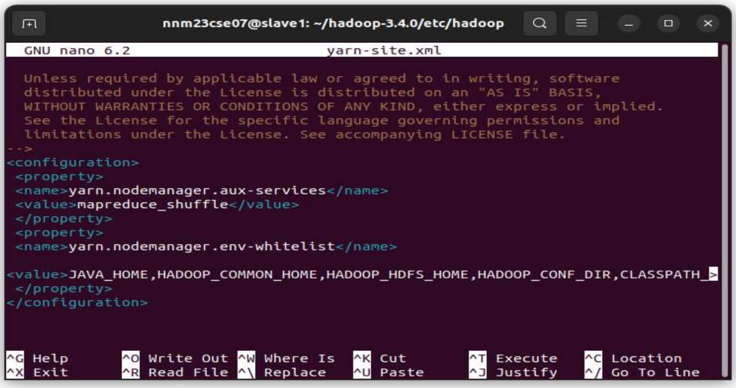


```
nnm23cse07@slave1:~$ cd hadoop-3.4.0/
nnm23cse07@slave1:~/hadoop-3.4.0$ cd etc/hadoop/
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ ls
capacity-scheduler.xml                  kms-log4j.properties
configuration.xsl                       kms-site.xml
container-executor.cfg                  log4j.properties
core-site.xml                           mapred-env.cmd
hadoop-env.cmd                          mapred-env.sh
hadoop-env.sh                           mapred-queues.xml.template
hadoop-metrics2.properties              mapred-site.xml
hadoop-policy.xml                       shellprofile.d
hadoop-user-functions.sh.example        ssl-client.xml.example
hdfs-rbf-site.xml                       ssl-server.xml.example
hdfs-site.xml                           user_ec_policies.xml.template
httpfs-env.sh                           workers
httpfs-log4j.properties                 yarn-env.cmd
httpfs-site.xml                         yarn-env.sh
kms-acls.xml                            yarnservice-log4j.properties
kms-env.sh                              yarn-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano hadoop-env.sh
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano core-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano hdfs-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano mapred-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano yarn-site.xml
```

Enter this below command line as shown below screen shot

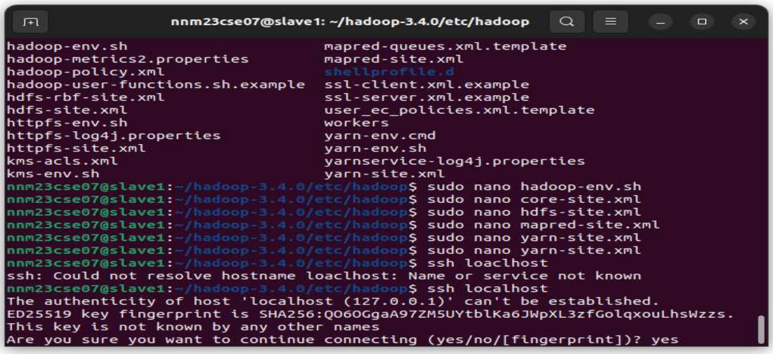
```
<configuration>
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.env-whitelist</name>
<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSP
ATH_PREP                    END_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
</property>
</configuration>
```

Then press Control O + Enter + Control x



```
  GNU nano 6.2                          yarn-site.xml

  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>
 <property>
 <name>yarn.nodemanager.aux-services</name>
 <value>mapreduce_shuffle</value>
 </property>
 <property>
 <name>yarn.nodemanager.env-whitelist</name>

<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_>
 </property>
</configuration>
```

|enter the command **ssh localhost** which initiates an SSH (Secure Shell) session to the local machine:

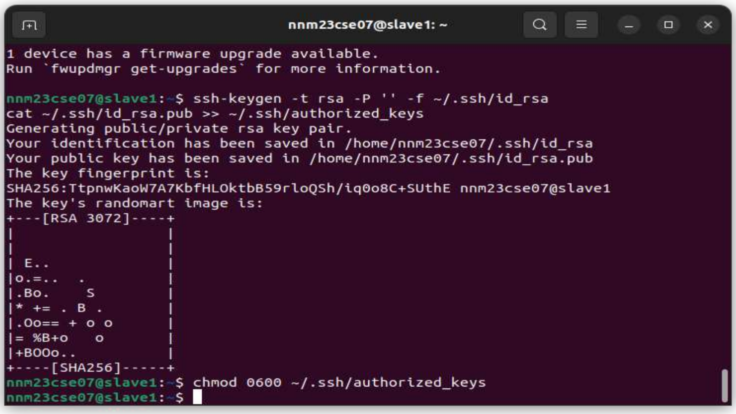


```
hadoop-env.sh                        mapred-queues.xml.template
hadoop-metrics2.properties           mapred-site.xml
hadoop-policy.xml                    shellprofile.d
hadoop-user-functions.sh.example     ssl-client.xml.example
hdfs-rbf-site.xml                    ssl-server.xml.example
hdfs-site.xml                        user_ec_policies.xml.template
httpfs-env.sh                        workers
httpfs-log4j.properties              yarn-env.cmd
httpfs-site.xml                      yarn-env.sh
kms-acls.xml                         yarnservice-log4j.properties
kms-env.sh                           yarn-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano hadoop-env.sh
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano core-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano hdfs-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano mapred-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano yarn-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ sudo nano yarn-site.xml
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ ssh loaclhost
ssh: Could not resolve hostname loaclhost: Name or service not known
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ED25519 key fingerprint is SHA256:QO6OGgaA97ZM5UYtblKa6JWpXL3zfGolqxouLhsWzzs.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
```

Enter below command as in the screen shot

ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa

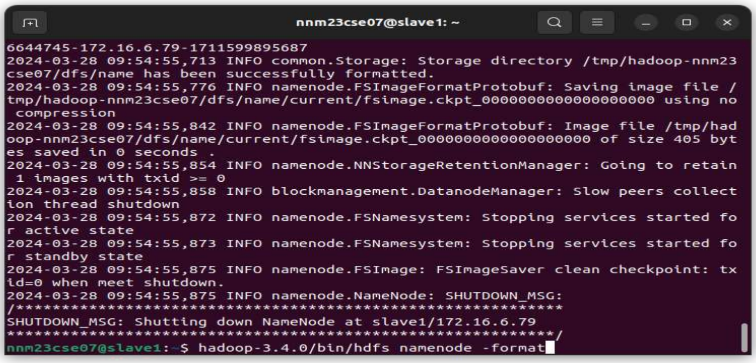
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys



```
1 device has a firmware upgrade available.
Run `fwupdmgr get-upgrades` for more information.

nnm23cse07@slave1:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
Generating public/private rsa key pair.
Your identification has been saved in /home/nnm23cse07/.ssh/id_rsa
Your public key has been saved in /home/nnm23cse07/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:TtpnwKaoW7A7KbfHLOktbB59rloQSh/iq0o8C+SUthE nnm23cse07@slave1
The key's randomart image is:
+---[RSA 3072]----+
|                 |
|                 |
| E..             |
|o.=..  .         |
|.Bo.    S        |
|* += . B .       |
|.Oo== + o o      |
|= %B+o   o       |
|+BOOo..          |
+----[SHA256]-----+
nnm23cse07@slave1:~$ chmod 0600 ~/.ssh/authorized_keys
nnm23cse07@slave1:~$ 
```

chmod 0600 ~/.ssh/authorized_keys command is used to set the permissions of the **authorized_keys** file to be more secure

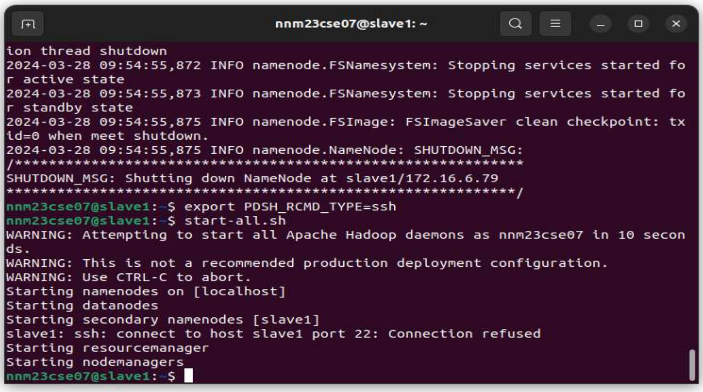
hadoop-3.2.3/bin/hdfs namenode –format is used in the context of setting up Apache Hadoop, a framework for distributed storage and processing of large data sets



```
6644745-172.16.6.79-1711599895687
2024-03-28 09:54:55,713 INFO common.Storage: Storage directory /tmp/hadoop-nnm23
cse07/dfs/name has been successfully formatted.
2024-03-28 09:54:55,776 INFO namenode.FSImageFormatProtobuf: Saving image file /
tmp/hadoop-nnm23cse07/dfs/name/current/fsimage.ckpt_0000000000000000000 using no
 compression
2024-03-28 09:54:55,842 INFO namenode.FSImageFormatProtobuf: Image file /tmp/had
oop-nnm23cse07/dfs/name/current/fsimage.ckpt_0000000000000000000 of size 405 byt
es saved in 0 seconds .
2024-03-28 09:54:55,854 INFO namenode.NNStorageRetentionManager: Going to retain
 1 images with txid >= 0
2024-03-28 09:54:55,858 INFO blockmanagement.DatanodeManager: Slow peers collect
ion thread shutdown
2024-03-28 09:54:55,872 INFO namenode.FSNamesystem: Stopping services started fo
r active state
2024-03-28 09:54:55,873 INFO namenode.FSNamesystem: Stopping services started fo
r standby state
2024-03-28 09:54:55,875 INFO namenode.FSImage: FSImageSaver clean checkpoint: tx
id=0 when meet shutdown.
2024-03-28 09:54:55,875 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at slave1/172.16.6.79
************************************************************/
nnm23cse07@slave1:~$ hadoop-3.4.0/bin/hdfs namenode -format
```

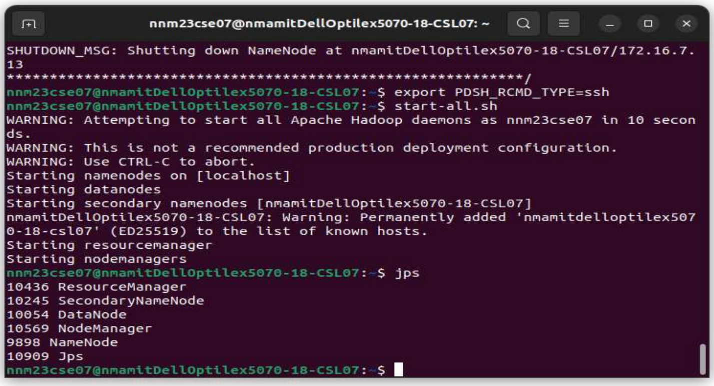
export PDSH_RCMD_TYPE=ssh

start-all.sh   Start NameNode daemon and DataNode daemon



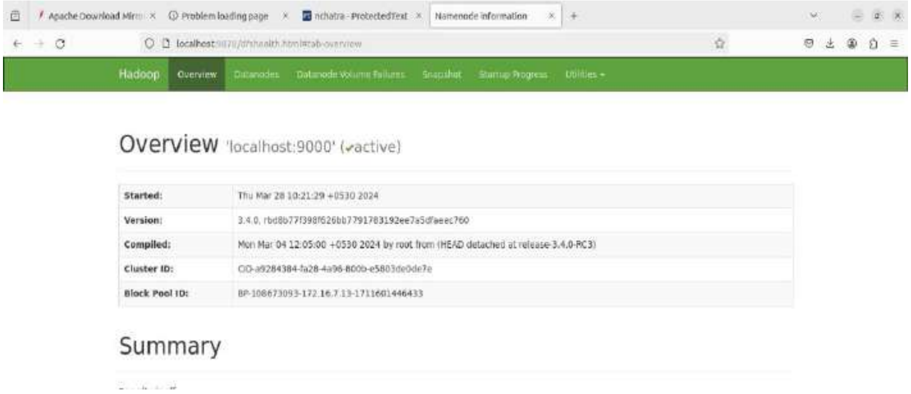
```
ion thread shutdown
2024-03-28 09:54:55,872 INFO namenode.FSNamesystem: Stopping services started fo
r active state
2024-03-28 09:54:55,873 INFO namenode.FSNamesystem: Stopping services started fo
r standby state
2024-03-28 09:54:55,875 INFO namenode.FSImage: FSImageSaver clean checkpoint: tx
id=0 when meet shutdown.
2024-03-28 09:54:55,875 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at slave1/172.16.6.79
************************************************************/
nnm23cse07@slave1:~$ export PDSH_RCMD_TYPE=ssh
nnm23cse07@slave1:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as nnm23cse07 in 10 secon
ds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [slave1]
slave1: ssh: connect to host slave1 port 22: Connection refused
Starting resourcemanager
Starting nodemanagers
nnm23cse07@slave1:~$
```

jps   : It is used to list the instrumented HotSpot Java Virtual Machines (JVMs) on a target system



```
SHUTDOWN_MSG: Shutting down NameNode at nmamitDellOptilex5070-18-CSL07/172.16.7.
13
************************************************************/
nnm23cse07@nmamitDellOptilex5070-18-CSL07:~$ export PDSH_RCMD_TYPE=ssh
nnm23cse07@nmamitDellOptilex5070-18-CSL07:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as nnm23cse07 in 10 secon
ds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [nmamitDellOptilex5070-18-CSL07]
nmamitDellOptilex5070-18-CSL07: Warning: Permanently added 'nmamitdelloptilex507
0-18-csl07' (ED25519) to the list of known hosts.
Starting resourcemanager
Starting nodemanagers
nnm23cse07@nmamitDellOptilex5070-18-CSL07:~$ jps
10436 ResourceManager
10245 SecondaryNameNode
10054 DataNode
10569 NodeManager
9898 NameNode
10909 Jps
nnm23cse07@nmamitDellOptilex5070-18-CSL07:~$
```

localhost:9870 refers to a network address that is used to access a service running on the local machine using a web browser or other network client



Overview 'localhost:9000' (✓active)

Started:	Thu Mar 28 10:21:29 +0530 2024
Version:	3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaeec760
Compiled:	Mon Mar 04 12:05:00 +0530 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID:	CID-a9284384-fa28-4a96-800b-e5803de0de7e
Block Pool ID:	BP-108673093-172.16.7.13-1711601446433

Summary