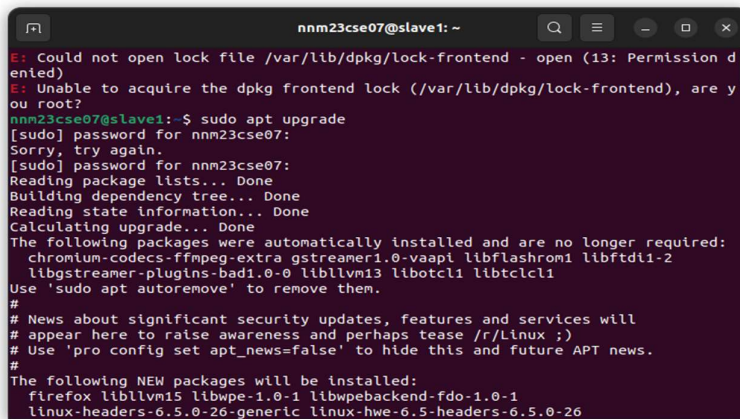


LAB PROGRAM PREREQUISITES: Installation of Hadoop

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

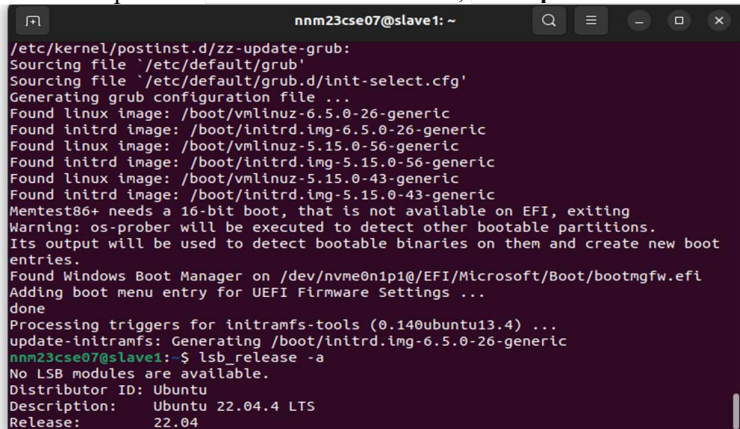
sudo apt upgrade: This is the correct usage to upgrade all upgradable packages on a Ubuntu system before installing Hadoop

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



```
nnm23cse07@slave1: ~  
E: Could not open lock file /var/lib/dpkg/lock-frontent - open (13: Permission denied)  
E: Unable to acquire the dpkg frontend lock (/var/lib/dpkg/lock-frontent), are you 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 libfdt1-2  
  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**



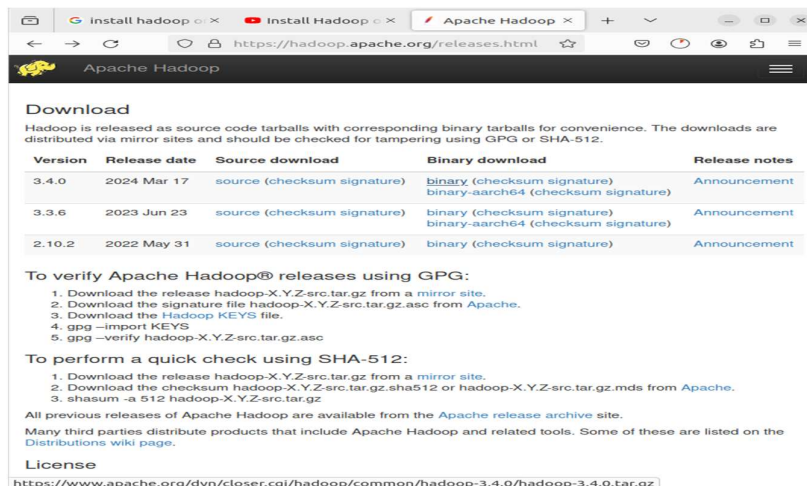
```
nnm23cse07@slave1: ~  
/etc/kernel/postinst.d/zz-update-grub:  
Sourcing file '/etc/default/grub'  
Sourcing file '/etc/default/grub.d/init-select.cfg'  
Generating grub configuration file ...  
Found linux image: /boot/vmlinuz-6.5.0-26-generic  
Found initrd image: /boot/initrd.img-6.5.0-26-generic  
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  
Mentest86+ 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 intransfs-tools (0.140ubuntu13.4) ...  
update-intransfs: 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 nnm23cse07: After entering the **sudo** command, the system prompts the user to enter the password for the **nnm23cse07**

```
nnm23cse07@slave1: ~  
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 libtcl1  
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://d1cdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz>



THE APACHE SOFTWARE FOUNDATION
ESTABLISHED 1999

We suggest the following location for your download:
<https://d1cdn.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://d1cdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz>

BACKUP SITES
<https://d1cdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz>

VERIFY THE INTEGRITY OF THE FILES
<https://d1cdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz>

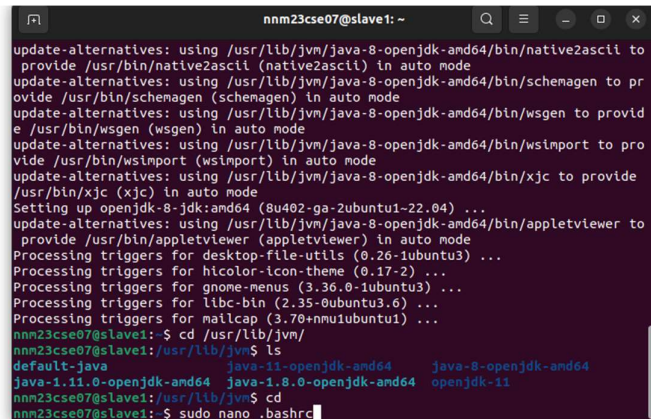
```
cd /usr/lib/jvm/
```

- **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



```
nnm23cse07@slave1: ~  
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+nmuiubuntu1) ...  
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 *  
ctrl /etc/bash_completion ]; then  
fi . /etc/bash_completion  
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=$HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-3.2.3.jar  
export HADOOP_LOG_DIR=$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 ^_ 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: ~  
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: ~/hadoop-3.4.0$ cd etc/hadoop/  
nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop$ ls  
capacity-scheduler.xml      kms-site.xml  
configuration.xml           kms-log4j.properties  
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... x nnm23cse07@slave1: /usr/lib/jvm x  
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
```



```

nnm23cse07@slave1: ~/hadoop-3.4.0/e...
GNU nano 6.2 .bashrc
/etc/bash_completion
fl
fl

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_CMD_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

```

nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop
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

```

nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop
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.xml           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           shell-profile.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
https-env.sh                workers
https-log4j.properties     yarn-env.cmd
https-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>

```

The screenshot shows a terminal window with the title 'nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop'. The nano editor is open to 'core-site.xml'. The file content is as follows:

```

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>
  <name>hadoop.proxyuser.dataflair.hosts</name> <value>*</value>
  </property>
  <name>hadoop.proxyuser.server.hosts</name> <value>*</value>
  </property>
  <name>hadoop.proxyuser.server.groups</name> <value>*</value>
  </property>
</configuration>

```

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

The screenshot shows a terminal window with the title 'nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop'. The nano editor is open to 'hdfs-site.xml'. The file content is as follows:

```

hadoop-3.4.0/sbin/start-balancer.sh
hadoop-3.4.0/sbin/start-all.sh
nnm23cse07@slave1: ~/hadoop-3.4.0$ cd /etc/hadoop/
nnm23cse07@slave1: /etc/hadoop$ ls
capacity-scheduler.xml      kms-log4j.properties
configuration.xml           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           shell-profile.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
https-env.sh                workers
https-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: /etc/hadoop$ sudo nano hadoop-env.sh
nnm23cse07@slave1: /etc/hadoop$ sudo nano core-site.xml
nnm23cse07@slave1: /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>

```

```

nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop
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>
^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute  ^C Location
^X Exit      ^R Read File ^L Replace   ^U Paste     ^D Justify  ^_ Go To Line

```

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

```

nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop
hadoop-3.4.0/sbin/start-all.sh
nnm23cse07@slave1: ~/hadoop-3.4.0$ cd /etc/hadoop/
nnm23cse07@slave1: /etc/hadoop$ ls
capacity-scheduler.xml      kms-log4j.properties
configuration.xml           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
hdfs-rbf-site.xml           ssl-client.xml.example
hdfs-site.xml               ssl-server.xml.example
https-env.sh                user_ec_policies.xml.template
https-log4j.properties      workers
https-site.xml              yarn-env.cmd
kms-acls.xml                yarn-env.sh
kms-env.sh                  yarnservice-log4j.properties
                             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>

```

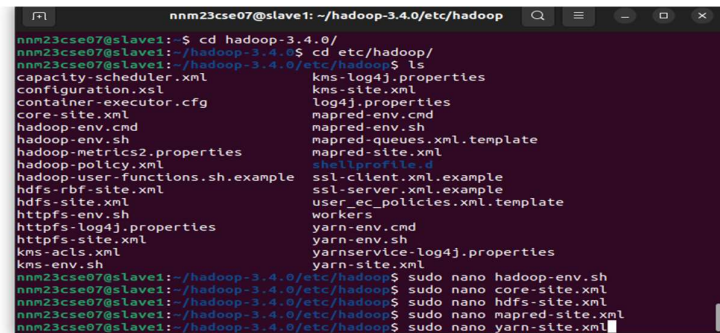
then press control O + Enter +control X

```

nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop
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>
^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute  ^C Location
^X Exit      ^R Read File ^L Replace   ^U Paste     ^D Justify  ^_ Go To Line

```

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

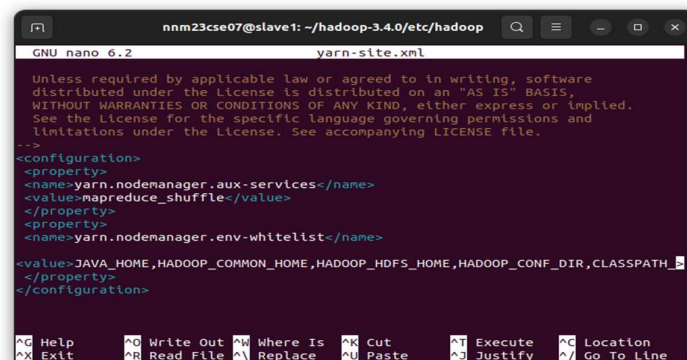


```
nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop
nnm23cse07@slave1:~/hadoop-3.4.0$ cd /etc/hadoop/
nnm23cse07@slave1:~/hadoop-3.4.0/etc/hadoop$ ls
capacity-scheduler.xml      kms-log4j.properties
configuration.xml           kms-site.xml
container-executor.cfg     log4j.properties
core-site.xml              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
https-env.sh               workers
https-log4j.properties     yarn-env.cmd
https-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,CLASSPATH,
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:


```

nnm23cse07@slave1: ~/hadoop-3.4.0/etc/hadoop
hadoop-env.sh          mapred-queues.xml.template
hadoop-metrics2.properties  mapred-site.xml
hadoop-policy.xml       ssh-keygen
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$ ssh localhost
ssh: Could not resolve hostname localhost: 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:Q060GgaA97ZM5UYtbka6JWpXL3zfGolqxoulhsWzss.
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`

```

nnm23cse07@slave1: ~
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:TtpnwKa0W7A7KbFHL0ktbB59rloqSh/Iq0o8C+SuthE nnm23cse07@slave1
The key's randomart image is:
+---[RSA 3072]-----+
|
| E..
|o=..  .
|..BO.  S
|+..  . B .
|.00== + o o
| = %B+o  o
|+B00o..
+---[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

```

nnm23cse07@slave1: ~
6644745-172.16.6.79-1711599895687
2024-03-28 09:54:55,713 INFO common.Storage: Storage directory /tmp/hadoop-nnm23cse07/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_000000000000000000 using no compression
2024-03-28 09:54:55,842 INFO namenode.FSImageFormatProtobuf: Image file /tmp/hadoop-nnm23cse07/dfs/name/current/fsimage.ckpt_000000000000000000 of size 405 bytes 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 collection thread shutdown
2024-03-28 09:54:55,872 INFO namenode.FSNamesystem: Stopping services started for active state
2024-03-28 09:54:55,873 INFO namenode.FSNamesystem: Stopping services started for standby state
2024-03-28 09:54:55,875 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=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

```
nnm23cse07@slave1: ~
ton thread shutdown
2024-03-28 09:54:55,872 INFO namenode.FSNamesystem: Stopping services started for active state
2024-03-28 09:54:55,873 INFO namenode.FSNamesystem: Stopping services started for standby state
2024-03-28 09:54:55,875 INFO namenode.FSImage: FSImageSaver clean checkpoint: txId=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 seconds.
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

```
nnm23cse07@nnamitDelloptilex5070-18-CSL07: ~
SHUTDOWN_MSG: Shutting down NameNode at nnamitDelloptilex5070-18-CSL07/172.16.7.13
/*****
nnm23cse07@nnamitDelloptilex5070-18-CSL07: $ export PDSH_RCMD_TYPE=ssh
nnm23cse07@nnamitDelloptilex5070-18-CSL07: $ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as nnm23cse07 in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [nnamitDelloptilex5070-18-CSL07]
nnamitDelloptilex5070-18-CSL07: Warning: Permanently added 'nnamitdelloptilex5070-18-csl07' (ED25519) to the list of known hosts.
Starting resourcemanager
Starting nodemanagers
nnm23cse07@nnamitDelloptilex5070-18-CSL07: $ jps
10436 ResourceManager
10245 SecondaryNameNode
10054 DataNode
10569 NodeManager
9898 NameNode
10909 Jps
nnm23cse07@nnamitDelloptilex5070-18-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:9800' (vactive)	
Started:	Thu Mar 28 10:21:29 +0530 2024
Version:	3.4.0, rbd0b77f398f626bb7791783192ee7a5dfaee760
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-106673093-172.16.7.13-1711601446433

Summary

LAB PROGRAM BASICS: Creating and managing directory in Hadoop.

hadoop namenode -format

```
hduser@DellOptiPlex5070-20-CSL07: /usr/local/hadoop
24/02/27 10:07:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@DellOptiPlex5070-20-CSL07: /usr/local/hadoop$ sbin/start-yarn.sh
starting yarn daemons
resourceananager running as process 4787. Stop it first.
localhost: nodemanager running as process 5155. Stop it first.
hduser@DellOptiPlex5070-20-CSL07: /usr/local/hadoop$ cd
hduser@DellOptiPlex5070-20-CSL07: /usr/local/hadoop$ sudo vi /usr/local/hadoop/etc/hadoop/yarn-site.xml
hduser@DellOptiPlex5070-20-CSL07: /usr/local/hadoop$ hadoop namenode -format
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

24/02/27 10:09:02 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = DellOptiPlex5070-20-CSL07/172.16.6.79
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 2.7.3
STARTUP_MSG: classpath = /usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/hadoop/common/lib/slf4j-api-1.7.10.jar:/usr/local/hadoop/share/hadoop/common/lib/jersey-core-1.9.jar:/usr/local/hadoop/share/hadoop/common/lib/protobuf-java-2.5.0.jar:/usr/local/hadoop/share/hadoop/common/lib/jetty-util-6.1.26.jar:/usr/local/hadoop/share/hadoop/common/lib/jsr305-3.0.0.jar:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.7.3.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-io-2.4.jar:/usr/local/hadoop/share/hadoop/common/lib/zookeeper-3.4.6.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-httpclient-3.1.jar:/usr/local/hadoop/share/hadoop/common/lib/hadoop-annotations-2.7.3.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-configuration-1.6.jar:/usr/local/hadoop/share/hadoop/common/lib/httpcore-4.2.5.jar:/usr/local/hadoop/share/hadoop/common/lib/stax-api-1.0-2.jar:/usr/local/hadoop/share/hadoop/common/lib/curator-framework-2.7.1.jar:/usr/local/hadoop/share/hadoop/common/lib/jersey-json-1.9.jar:/usr/local/hadoop/share/hadoop/common/lib/api-util-1.0.0-H20.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-beanutils-core-1.8.0.jar:/usr/local/hadoop/share/hadoop/common/lib/guava-11.0.2.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-logging-1.1.3.jar:/usr/local/hadoop/share/hadoop/common/lib/jaxb-impl-2.2.3-1.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-lang-2.6.jar:/usr/local/hadoop/share/hadoop/common/lib/jsp-api-2.1.jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-xxc-1.9.13.jar:/usr/local/hadoop/share/hadoop/common/lib/curator-client-2.7.1.jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-mapper-asl-1.9.13.jar:/u
```

jps Checking Java processes associated with Hadoop

hadoop fs -mkdir /mydir Attempting to create a new directory in HDFS:

```
hduser@DellOptiPlex5070-20-CSL07: /usr/local/hadoop
24/02/27 10:09:04 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at DellOptiPlex5070-20-CSL07/172.16.6.79
*****/
hduser@DellOptiPlex5070-20-CSL07: /usr/local/hadoop$ sbin/start-dfs.sh
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/hadoop-auth-2.7.3.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
24/02/27 10:09:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting namenodes on (localhost)
localhost: namenode running as process 5938. Stop it first.
localhost: datanode running as process 6087. Stop it first.
Starting secondary namenodes [0.0.0.0]
0.0.0.0: secondarynamenode running as process 6277. Stop it first.
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/hadoop-auth-2.7.3.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
24/02/27 10:09:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@DellOptiPlex5070-20-CSL07: /usr/local/hadoop$ sbin/start-yarn.sh
starting yarn daemons
resourceananager running as process 4787. Stop it first.
localhost: nodemanager running as process 5155. Stop it first.
hduser@DellOptiPlex5070-20-CSL07: /usr/local/hadoop$ jps
5938 NameNode
4787 ResourceManager
5155 NodeManager
6277 SecondaryNameNode
6087 DataNode
hduser@DellOptiPlex5070-20-CSL07: /usr/local/hadoop$ hadoop fs-mkdir /mydir
Error: Could not find or load main class fs-mkdir
Caused by: java.lang.ClassNotFoundException: fs-mkdir
hduser@DellOptiPlex5070-20-CSL07: /usr/local/hadoop$ hadoop fs -mkdir /mydir
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/hadoop-auth-2.7.3.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
```

hadoop fs -mkdir /mydir Attempting to create a new directory in HDFS:

hadoop fs -chmod 777 /mydir Changing HDFS directory permissions:

hadoop fs -put inputdata.txt /inputdatasamples Copying a file from the local file system to HDFS:

hadoop fs -ls / Listing contents of a directory in HDFS:

