

## Practical No.1.

**Aim:** Install, configure and run Hadoop and HDFS and explore HDFS

>Install Virtual Box in your system

<https://download.virtualbox.org/virtualbox/6.1.22/VirtualBox-6.1.22-144080-Win.exe>

>Create a ubuntu VM inside Virtual box using the below ISO

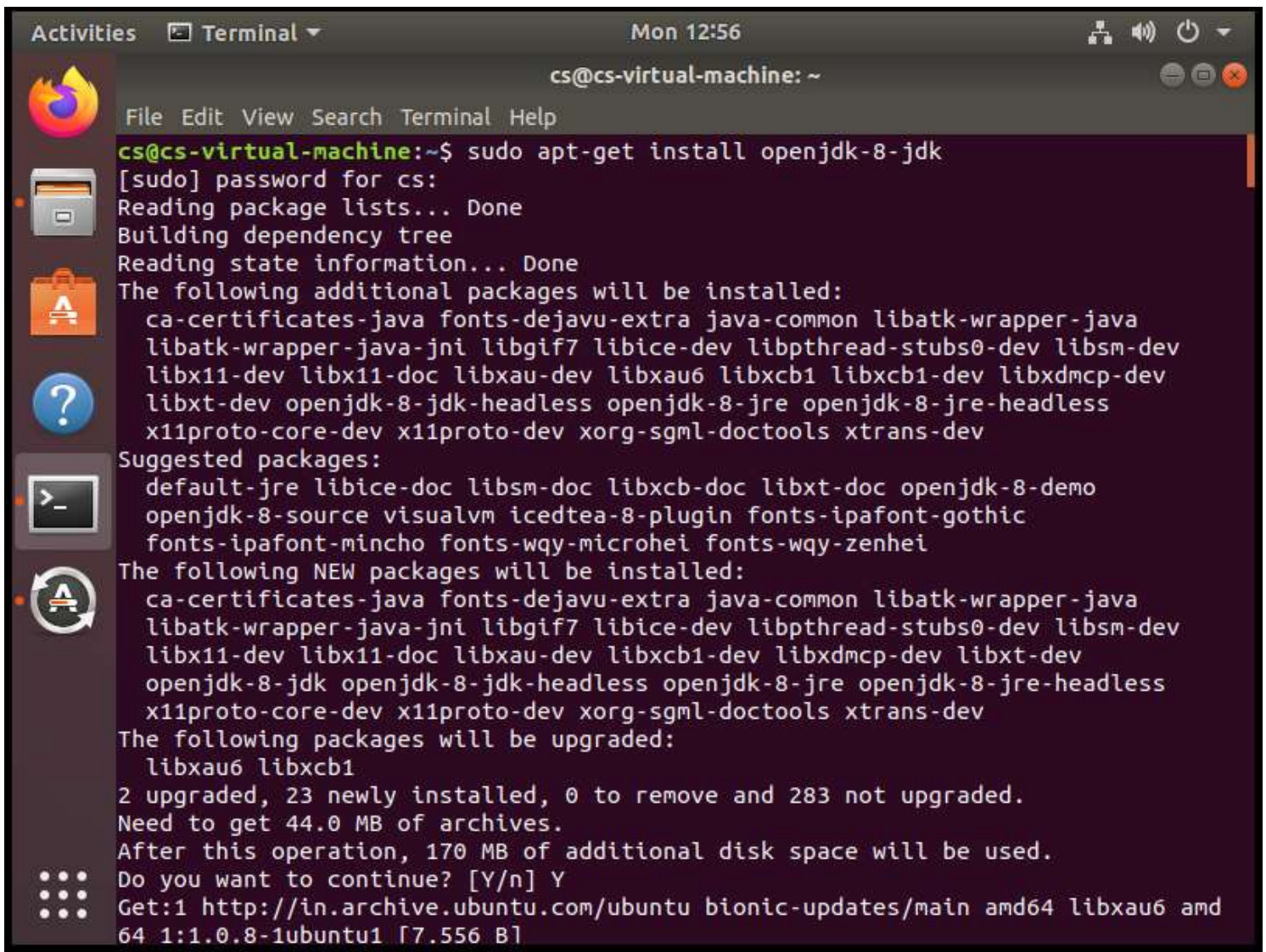
<https://ubuntu.com/download/desktop/thank-you?version=20.04.2.0&architecture=amd64>

>use the following link to check ubuntu installation process

<https://youtu.be/x5MhydijWmc>

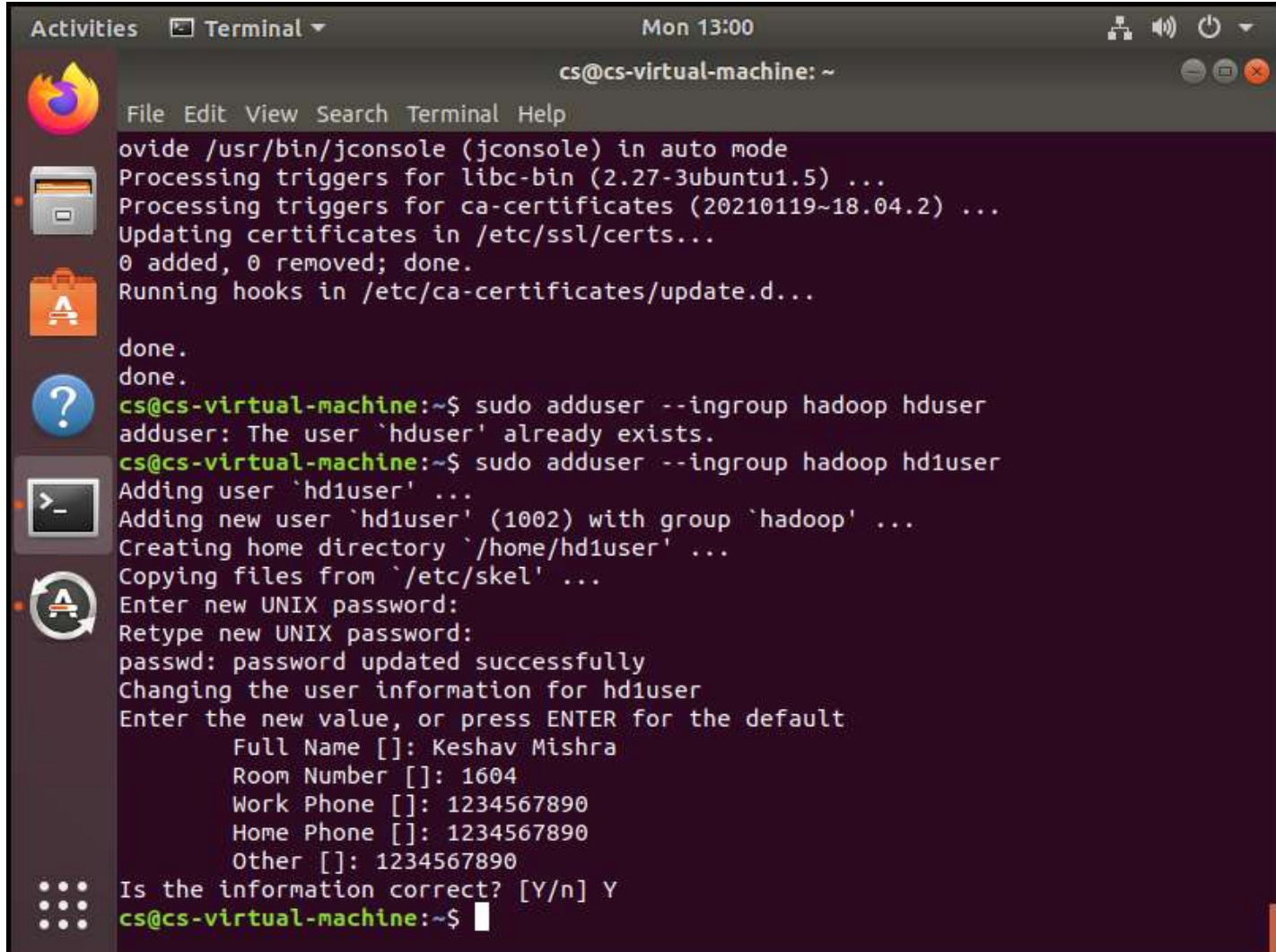
1. First you need to perform some prerequisites

```
sudo apt-get install openjdk-8-jdk
```



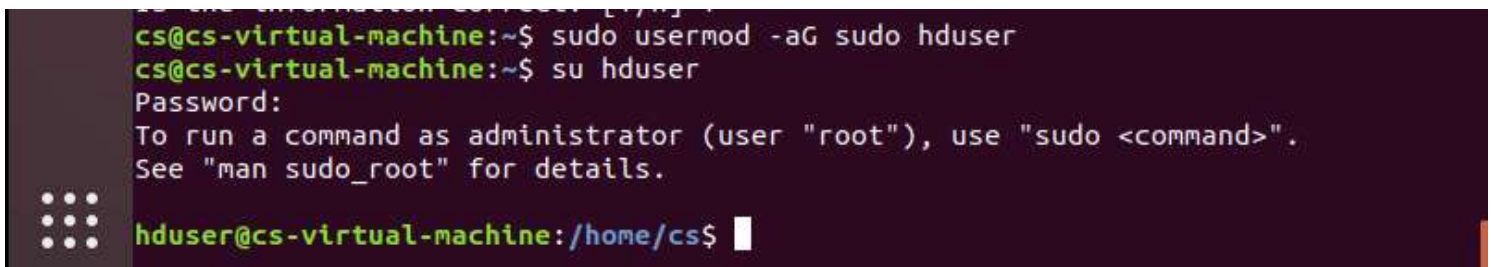
```
Activities Terminal Mon 12:56
cs@cs-virtual-machine: ~
File Edit View Search Terminal Help
cs@cs-virtual-machine:~$ sudo apt-get install openjdk-8-jdk
[sudo] password for cs:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
 ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java
 libatk-wrapper-java-jni libgif7 libice-dev libpthread-stubs0-dev libsm-dev
 libx11-dev libx11-doc libxau-dev libxcb1 libxcb1-dev libxdmcp-dev
 libxt-dev openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless
 x11proto-core-dev x11proto-dev xorg-sgml-doctools xtrans-dev
Suggested packages:
 default-jre libice-doc libsm-doc libxcb-doc libxt-doc openjdk-8-demo
 openjdk-8-source visualvm icedtea-8-plugin fonts-ipafont-gothic
 fonts-ipafont-mincho fonts-wqy-microhei fonts-wqy-zenhei
The following NEW packages will be installed:
 ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java
 libatk-wrapper-java-jni libgif7 libice-dev libpthread-stubs0-dev libsm-dev
 libx11-dev libx11-doc libxau-dev libxcb1-dev libxdmcp-dev libxt-dev
 openjdk-8-jdk openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless
 x11proto-core-dev x11proto-dev xorg-sgml-doctools xtrans-dev
The following packages will be upgraded:
 libxau6 libxcb1
2 upgraded, 23 newly installed, 0 to remove and 283 not upgraded.
Need to get 44.0 MB of archives.
After this operation, 170 MB of additional disk space will be used.
Do you want to continue? [Y/n] Y
Get:1 http://in.archive.ubuntu.com/ubuntu bionic-updates/main amd64 libxau6 amd
64 1:1.0.8-1ubuntu1 [7.556 B]
```

```
sudo addgroup hadoop
sudo adduser --ingroup hadoop hduser
```



```
cs@cs-virtual-machine: ~
File Edit View Search Terminal Help
ovide /usr/bin/jconsole (jconsole) in auto mode
Processing triggers for libc-bin (2.27-3ubuntu1.5) ...
Processing triggers for ca-certificates (20210119~18.04.2) ...
Updating certificates in /etc/ssl/certs...
0 added, 0 removed; done.
Running hooks in /etc/ca-certificates/update.d...
done.
done.
cs@cs-virtual-machine:~$ sudo adduser --ingroup hadoop hduser
adduser: The user `hduser' already exists.
cs@cs-virtual-machine:~$ sudo adduser --ingroup hadoop hd1user
Adding user `hd1user' ...
Adding new user `hd1user' (1002) with group `hadoop' ...
Creating home directory `/home/hd1user' ...
Copying files from `/etc/skel' ...
Enter new UNIX password:
Retype new UNIX password:
passwd: password updated successfully
Changing the user information for hd1user
Enter the new value, or press ENTER for the default
Full Name []: Keshav Mishra
Room Number []: 1604
Work Phone []: 1234567890
Home Phone []: 1234567890
Other []: 1234567890
Is the information correct? [Y/n] Y
cs@cs-virtual-machine:~$
```

```
usermod -aG sudo hduser
su hduser
```



```
cs@cs-virtual-machine:~$ sudo usermod -aG sudo hduser
cs@cs-virtual-machine:~$ su hduser
Password:
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.
hduser@cs-virtual-machine:/home/cs$
```

```
ssh-keygen -t rsa -P ""
cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
```

```
Activities Terminal Mon 13:07
hd1user@cs-virtual-machine: /home/cs
File Edit View Search Terminal Help
hd1user@cs-virtual-machine:/home/cs$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hd1user/.ssh/id_rsa): /home/hduser/.ssh/t
Saving key "/home/hduser/.ssh/t" failed: Permission denied
hd1user@cs-virtual-machine:/home/cs$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hd1user/.ssh/id_rsa): /home/hd1user/.ssh/t
Created directory '/home/hd1user/.ssh'.
Your identification has been saved in /home/hd1user/.ssh/t.
Your public key has been saved in /home/hd1user/.ssh/t.pub.
The key fingerprint is:
SHA256:r3NQdDIXWMLjfnBx7YPFgXv8riyuokqkFfTHEaxXj/Y hd1user@cs-virtual-machine
The key's randomart image is:
+---[RSA 2048]----+
|  . .O. =00.+ . |
|  . . . . * 0.+ = |
|  . .00 0 +000 |
|  . . . . + . .0+ |
|  0 .So . . . .0 |
|  + . . E 0 |
|  . . . . . |
|  . . 0 . . . . |
|  . . . . + .0 . .0 . |
+-----[SHA256]-----+
hd1user@cs-virtual-machine:/home/cs$
```

2. Now you need to disable IPv6. Open the /etc/sysctl.conf file and add the following lines to the end of the file and save it. (One way of opening the file is `sudo nano /etc/sysctl.conf`, after you add the lines you need to press Ctrl+X, Shift Y and Enter)

```
net.ipv6.conf.all.disable_ipv6 = 1
net.ipv6.conf.default.disable_ipv6 = 1
net.ipv6.conf.lo.disable_ipv6 = 1
```

```
hd1user@cs-virtual-machine:/home/cs$ sudo nano /etc/sysctl.conf
[sudo] password for hd1user:
hd1user@cs-virtual-machine:/home/cs$
```

```
Activities Terminal Mon 13:13
hd1user@cs-virtual-machine: /home/cs
File Edit View Search Terminal Help
GNU nano 2.9.3 /etc/sysctl.conf
net.ipv6.conf.all.disable_ipv6=1
net.ipv6.conf.default.disable_ipv6=1
net.ipv6.conf.lo.disable_ipv6=1
[Wrote 3 lines]
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Spell
```

3. Now we download hadoop

```
cd /usr/local
```

```
Wget https://dldn.apache.org/hadoop/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz
```

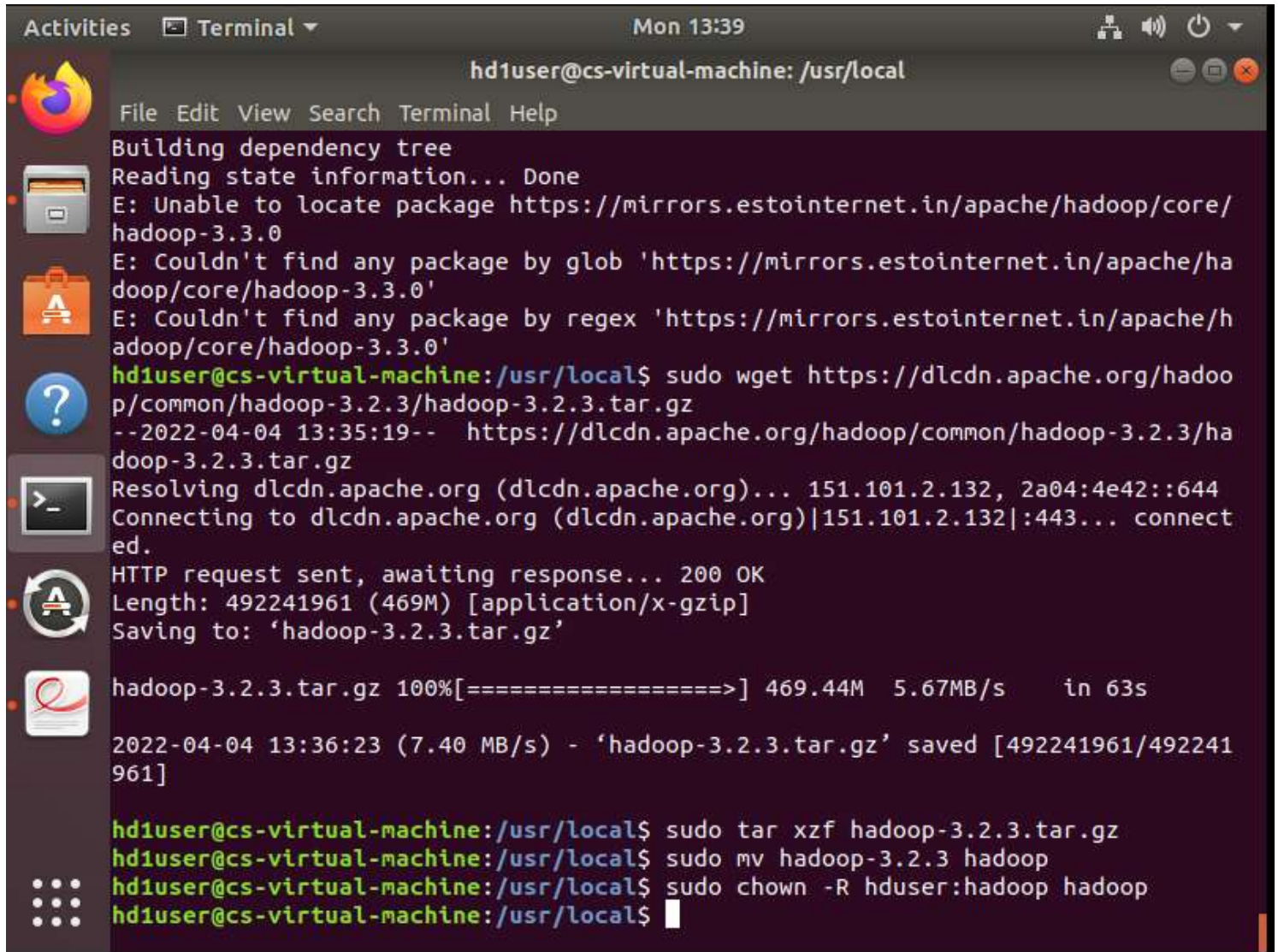
```
hd1user@cs-virtual-machine: /usr/local$ sudo wget https://dldn.apache.org/hadoop
p/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz
--2022-04-04 13:35:19-- https://dldn.apache.org/hadoop/common/hadoop-3.2.3/ha
doo-3.2.3.tar.gz
Resolving dldn.apache.org (dldn.apache.org)... 151.101.2.132, 2a04:4e42::644
Connecting to dldn.apache.org (dldn.apache.org)|151.101.2.132|:443... connect
ed.
HTTP request sent, awaiting response... 200 OK
Length: 492241961 (469M) [application/x-gzip]
Saving to: 'hadoop-3.2.3.tar.gz'

hadoop-3.2.3.tar.gz 100%[=====>] 469.44M 5.67MB/s in 63s

2022-04-04 13:36:23 (7.40 MB/s) - 'hadoop-3.2.3.tar.gz' saved [492241961/492241
961]

hd1user@cs-virtual-machine: /usr/local$
```

```
sudo tar xzf hadoop-3.3.0.tar.gz
sudo mv hadoop-3.3.0 hadoop
sudo chown -R hduser:hadoop hadoop
```



```
Activities Terminal Mon 13:39
hd1user@cs-virtual-machine: /usr/local
File Edit View Search Terminal Help
Building dependency tree
Reading state information... Done
E: Unable to locate package https://mirrors.estointernet.in/apache/hadoop/core/hadoop-3.3.0
E: Couldn't find any package by glob 'https://mirrors.estointernet.in/apache/hadoop/core/hadoop-3.3.0'
E: Couldn't find any package by regex 'https://mirrors.estointernet.in/apache/hadoop/core/hadoop-3.3.0'
hd1user@cs-virtual-machine: /usr/local$ sudo wget https://dlcdn.apache.org/hadoop/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz
--2022-04-04 13:35:19-- https://dlcdn.apache.org/hadoop/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz
Resolving dlcdn.apache.org (dlcdn.apache.org)... 151.101.2.132, 2a04:4e42::644
Connecting to dlcdn.apache.org (dlcdn.apache.org)|151.101.2.132|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 492241961 (469M) [application/x-gzip]
Saving to: 'hadoop-3.2.3.tar.gz'
hadoop-3.2.3.tar.gz 100%[=====>] 469.44M 5.67MB/s in 63s
2022-04-04 13:36:23 (7.40 MB/s) - 'hadoop-3.2.3.tar.gz' saved [492241961/492241961]
hd1user@cs-virtual-machine: /usr/local$ sudo tar xzf hadoop-3.2.3.tar.gz
hd1user@cs-virtual-machine: /usr/local$ sudo mv hadoop-3.2.3 hadoop
hd1user@cs-virtual-machine: /usr/local$ sudo chown -R hduser:hadoop hadoop
hd1user@cs-virtual-machine: /usr/local$
```

4. Now open `$HOME/.bashrc` and add the following lines:

```
export HADOOP_HOME=/usr/local/hadoop
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64
unalias fs &> /dev/null
alias fs="hadoop fs"
unalias hls &> /dev/null
alias hls="fs -ls"
lzohead () {
hadoop fs -cat $1 | lzop -dc | head -1000 | less
}
export PATH=$PATH:$HADOOP_HOME/bin
```

```
hd1user@cs-virtual-machine:/usr/local$ sudo tar xzf hadoop-3.2.3.tar.gz
hd1user@cs-virtual-machine:/usr/local$ sudo mv hadoop-3.2.3 hadoop
hd1user@cs-virtual-machine:/usr/local$ sudo chown -R hduser:hadoop hadoop
hd1user@cs-virtual-machine:/usr/local$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine:/usr/local$ sudo chown -R hd1user:hadoop hadoop
hd1user@cs-virtual-machine:/usr/local$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine:/usr/local$
```

```
hd1user@cs-virtual-machine: /usr/local
File Edit View Search Terminal Help
GNU nano 2.9.3 /home/hd1user/.bashrc
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! 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
  fi
fi
export HADOOP_HOME=/usr/local/hadoop
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64
unalias fs &> /dev/null
alias fs="hadoop fs"
unalias hls &> /dev/null
alias hls="fs -ls"
lzohead () {
hadoop fs -cat $1 | lzop -dc | head -1000 | less
}
export PATH=$PATH:$HADOOP_HOME/bin
```

5. Enter following commands:  
source ~/.bashrc  
cd /usr/local/hadoop/etc/hadoop

```
Activities Terminal Mon 13:47
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop
File Edit View Search Terminal Help
doop/core/hadoop-3.3.0'
E: Couldn't find any package by regex 'https://mirrors.estointernet.in/apache/h
adoop/core/hadoop-3.3.0'
hd1user@cs-virtual-machine:~$ sudo wget https://dlcdn.apache.org/hadoo
p/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz
--2022-04-04 13:35:19-- https://dlcdn.apache.org/hadoop/common/hadoop-3.2.3/ha
doop-3.2.3.tar.gz
Resolving dlcdn.apache.org (dlcdn.apache.org)... 151.101.2.132, 2a04:4e42::644
Connecting to dlcdn.apache.org (dlcdn.apache.org)|151.101.2.132|:443... connect
ed.
HTTP request sent, awaiting response... 200 OK
Length: 492241961 (469M) [application/x-gzip]
Saving to: 'hadoop-3.2.3.tar.gz'

hadoop-3.2.3.tar.gz 100%[=====] 469.44M 5.67MB/s in 63s

2022-04-04 13:36:23 (7.40 MB/s) - 'hadoop-3.2.3.tar.gz' saved [492241961/492241
961]

hd1user@cs-virtual-machine:~$ sudo tar xzf hadoop-3.2.3.tar.gz
hd1user@cs-virtual-machine:~$ sudo mv hadoop-3.2.3 hadoop
hd1user@cs-virtual-machine:~$ sudo chown -R hduser:hadoop hadoop
hd1user@cs-virtual-machine:~$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine:~$ sudo chown -R hd1user:hadoop hadoop
hd1user@cs-virtual-machine:~$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine:~$ source ~/.bashrc
hd1user@cs-virtual-machine:~$ cd /usr/local/hadoop/etc/hadoop
hd1user@cs-virtual-machine:~$
```

6. Add the following line to hadoop-env.sh  
export JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64

```
Activities Terminal Mon 13:55
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop
File Edit View Search Terminal Help
hadoop/core/hadoop-3.3.0'
hd1user@cs-virtual-machine: /usr/local$ sudo wget https://d1cdn.apache.org/hadoop
p/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz
--2022-04-04 13:35:19-- https://d1cdn.apache.org/hadoop/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz
Resolving d1cdn.apache.org (d1cdn.apache.org)... 151.101.2.132, 2a04:4e42::644
Connecting to d1cdn.apache.org (d1cdn.apache.org)|151.101.2.132|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 492241961 (469M) [application/x-gzip]
Saving to: 'hadoop-3.2.3.tar.gz'

hadoop-3.2.3.tar.gz 100%[=====>] 469.44M 5.67MB/s in 63s

2022-04-04 13:36:23 (7.40 MB/s) - 'hadoop-3.2.3.tar.gz' saved [492241961/492241961]

hd1user@cs-virtual-machine: /usr/local$ sudo tar xzf hadoop-3.2.3.tar.gz
hd1user@cs-virtual-machine: /usr/local$ sudo mv hadoop-3.2.3 hadoop
hd1user@cs-virtual-machine: /usr/local$ sudo chown -R hduser:hadoop hadoop
hd1user@cs-virtual-machine: /usr/local$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine: /usr/local$ sudo chown -R hd1user:hadoop hadoop
hd1user@cs-virtual-machine: /usr/local$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine: /usr/local$ source ~/.bashrc
hd1user@cs-virtual-machine: /usr/local$ cd /usr/local/hadoop/etc/hadoop
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop$ sudo nano hadoop-env.s
h
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop$
```



```
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop
GNU nano 2.9.3 hadoop-env.sh
# These options will be appended to the options specified as HADOOP_OPTS
# and therefore may override any similar flags set in HADOOP_OPTS
#
# export HDFS_STORAGECONTAINERMANAGER_OPTS=""
###
# Advanced Users Only!
###
#
# When building Hadoop, one can add the class paths to the commands
# via this special env var:
# export HADOOP_ENABLE_BUILD_PATHS="true"
#
# To prevent accidents, shell commands be (superficially) locked
# to only allow certain users to execute certain subcommands.
# It uses the format of (command)_(subcommand)_USER.
#
# For example, to limit who can execute the namenode command,
# export HDFS_NAMENODE_USER=hdfs
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64
[ Wrote 418 lines ]
^G Get Help    ^O Write Out  ^W Where Is   ^K Cut Text   ^J Justify
^X Exit        ^R Read File  ^\ Replace    ^U Uncut Text ^T To Linter
```

7. Run the following commands:  
sudo mkdir -p /app/hadoop/tmp  
sudo chown hduser:hadoop /app/hadoop/tmp

```
Activities Terminal Mon 13:57
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop
File Edit View Search Terminal Help
doop-3.2.3.tar.gz
Resolving dlcdn.apache.org (dlcdn.apache.org)... 151.101.2.132, 2a04:4e42::644
Connecting to dlcdn.apache.org (dlcdn.apache.org)|151.101.2.132|:443... connect
ed.
HTTP request sent, awaiting response... 200 OK
Length: 492241961 (469M) [application/x-gzip]
Saving to: 'hadoop-3.2.3.tar.gz'

hadoop-3.2.3.tar.gz 100%[=====] 469.44M 5.67MB/s in 63s

2022-04-04 13:36:23 (7.40 MB/s) - 'hadoop-3.2.3.tar.gz' saved [492241961/492241
961]

hd1user@cs-virtual-machine:/usr/local$ sudo tar xzf hadoop-3.2.3.tar.gz
hd1user@cs-virtual-machine:/usr/local$ sudo mv hadoop-3.2.3 hadoop
hd1user@cs-virtual-machine:/usr/local$ sudo chown -R hduser:hadoop hadoop
hd1user@cs-virtual-machine:/usr/local$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine:/usr/local$ sudo chown -R hd1user:hadoop hadoop
hd1user@cs-virtual-machine:/usr/local$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine:/usr/local$ source ~/.bashrc
hd1user@cs-virtual-machine:/usr/local$ cd /usr/local/hadoop/etc/hadoop
hd1user@cs-virtual-machine:/usr/local/hadoop/etc/hadoop$ sudo nano hadoop-env.s
h
hd1user@cs-virtual-machine:/usr/local/hadoop/etc/hadoop$ sudo mkdir -p /app/had
oop/tmp
hd1user@cs-virtual-machine:/usr/local/hadoop/etc/hadoop$ sudo chown hd1user:had
oop /app/hadoop/tmp
hd1user@cs-virtual-machine:/usr/local/hadoop/etc/hadoop$
```

8. Make the following changes in core-site.xml (this file is present in /usr/local/hadoop/etc/hadoop)

Add the following between <configuration> and </configuration>

```
<property>
  <name>hadoop.tmp.dir</name>
  <value>/app/hadoop/tmp</value>
  <description>A base for other temporary directories.</description>
</property>
```

```
<property>
  <name>fs.default.name</name>
  <value>hdfs://localhost:54310</value>
  <description>The name of the default file system. A URI whose
  scheme and authority determine the FileSystem implementation. The
  uri's scheme determines the config property (fs.SCHEME.impl) naming
  the FileSystem implementation class. The uri's authority is used to
```

determine the host, port, etc. for a filesystem.</description>  
</property>



The image shows a terminal window titled "Terminal" with a system clock of "Mon 14:02". The user is logged in as "hd1user" at "cs-virtual-machine" in the directory "/usr/local/hadoop/etc/hadoop". The terminal output shows the following sequence of commands and their results:

```
2022-04-04 13:36:23 (7.40 MB/s) - 'hadoop-3.2.3.tar.gz' saved [492241961/492241961]
hd1user@cs-virtual-machine:~/usr/local$ sudo tar xzf hadoop-3.2.3.tar.gz
hd1user@cs-virtual-machine:~/usr/local$ sudo mv hadoop-3.2.3 hadoop
hd1user@cs-virtual-machine:~/usr/local$ sudo chown -R hduser:hadoop hadoop
hd1user@cs-virtual-machine:~/usr/local$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine:~/usr/local$ sudo chown -R hd1user:hadoop hadoop
hd1user@cs-virtual-machine:~/usr/local$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine:~/usr/local$ source ~/.bashrc
hd1user@cs-virtual-machine:~/usr/local$ cd /usr/local/hadoop/etc/hadoop
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano hadoop-env.sh
h
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo mkdir -p /app/hadoop/tmp
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo chown hd1user:hadoop /app/hadoop/tmp
```

```
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop
GNU nano 2.9.3 core-site.xml
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>hadoop.tmp.dir</name>
<value>/app/hadoop/tmp</value>
<description>A base for other temporary directories.</description>
</property>
<property>
<name>fs.default.name</name>
<value>hdfs://localhost:54310</value>
<description>The name of the default file system. A URI whose
scheme and authority determine the FileSystem implementation. The
uri's scheme determines the config property (fs.SCHEME.impl) naming
the FileSystem implementation class. The uri's authority is used to
determine the host, port, etc. for a filesystem.</description>
</property>
</configuration>
[ Wrote 34 lines ]
^G Get Help      ^O Write Out    ^W Where Is     ^K Cut Text     ^J Justify
^X Exit          ^R Read File   ^\ Replace      ^U Uncut Text  ^T To Spell
```

9. In the file `mapred-site.xml` Add the following between `<configuration>` and `</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>
```

```
Activities Terminal Mon 14:05
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop
File Edit View Search Terminal Help
hd1user@cs-virtual-machine:~/usr/local$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine:~/usr/local$ sudo chown -R hd1user:hadoop hadoop
hd1user@cs-virtual-machine:~/usr/local$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine:~/usr/local$ source ~/.bashrc
hd1user@cs-virtual-machine:~/usr/local$ cd /usr/local/hadoop/etc/hadoop
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano hadoop-env.s
h
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo mkdir -p /app/had
oop/tmp
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo chown hd1user:had
oop /app/hadoop/tmp
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano core-site.xml
Use "fg" to return to nano.
[1]+  Stopped                  sudo nano core-site.xml
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano core-site.xml
Use "fg" to return to nano.
[2]+  Stopped                  sudo nano core-site.xml
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano mapred-site.
xml
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano core-site.xml
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano mapred-site.
xml
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$
```

```
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop
GNU nano 2.9.3 mapred-site.xml Modified
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>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>
</configuration>
```

^G Get Help   ^O Write Out   ^W Where Is   ^K Cut Text   ^J Justify  
^X Exit   ^R Read File   ^\ Replace   ^U Uncut Text   ^T To Spell

10. In the file hdfs-site.xml Add the following between <configuration> and </configuration>

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

```
<description>Default block replication.
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.
</description>
</property>
```

```
Activities Terminal Mon 14:08
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop

File Edit View Search Terminal Help
hd1user@cs-virtual-machine:~/usr/local$ sudo nano $HOME/.bashrc
hd1user@cs-virtual-machine:~/usr/local$ source ~/.bashrc
hd1user@cs-virtual-machine:~/usr/local$ cd /usr/local/hadoop/etc/hadoop
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano hadoop-env.sh
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo mkdir -p /app/hadoop/tmp
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo chown hd1user:hadoop /app/hadoop/tmp
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano core-site.xml
Use "fg" to return to nano.
[1]+ Stopped sudo nano core-site.xml
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano core-site.xml
Use "fg" to return to nano.
[2]+ Stopped sudo nano core-site.xml
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano mapred-site.xml
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano core-site.xml
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano mapred-site.xml
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$ sudo nano hdfs-site.xml
hd1user@cs-virtual-machine:~/usr/local/hadoop/etc/hadoop$
```

```
Activities Terminal Mon 14:08
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop

File Edit View Search Terminal Help
GNU nano 2.9.3 hdfs-site.xml

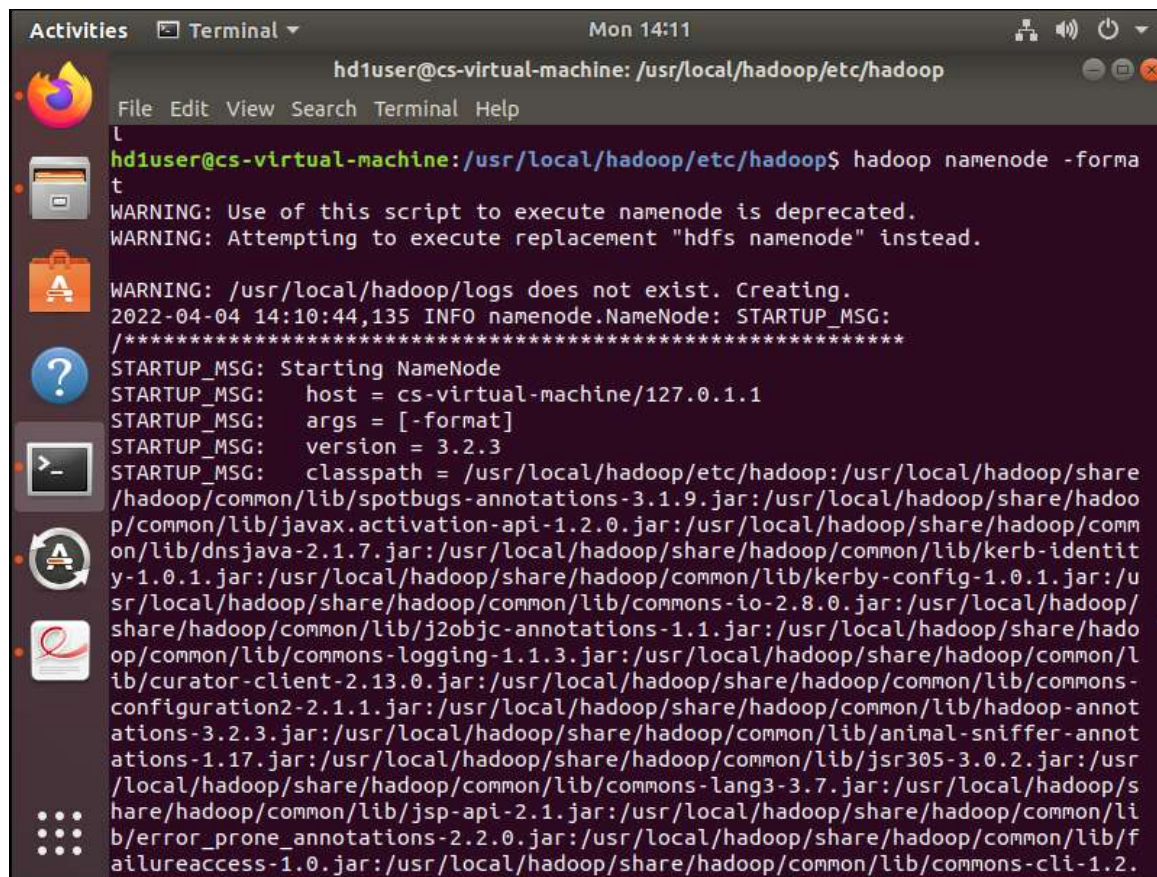
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
<description>Default block replication.
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.
</description>
</property>
</configuration>

[ Wrote 28 lines ]
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Spell
```

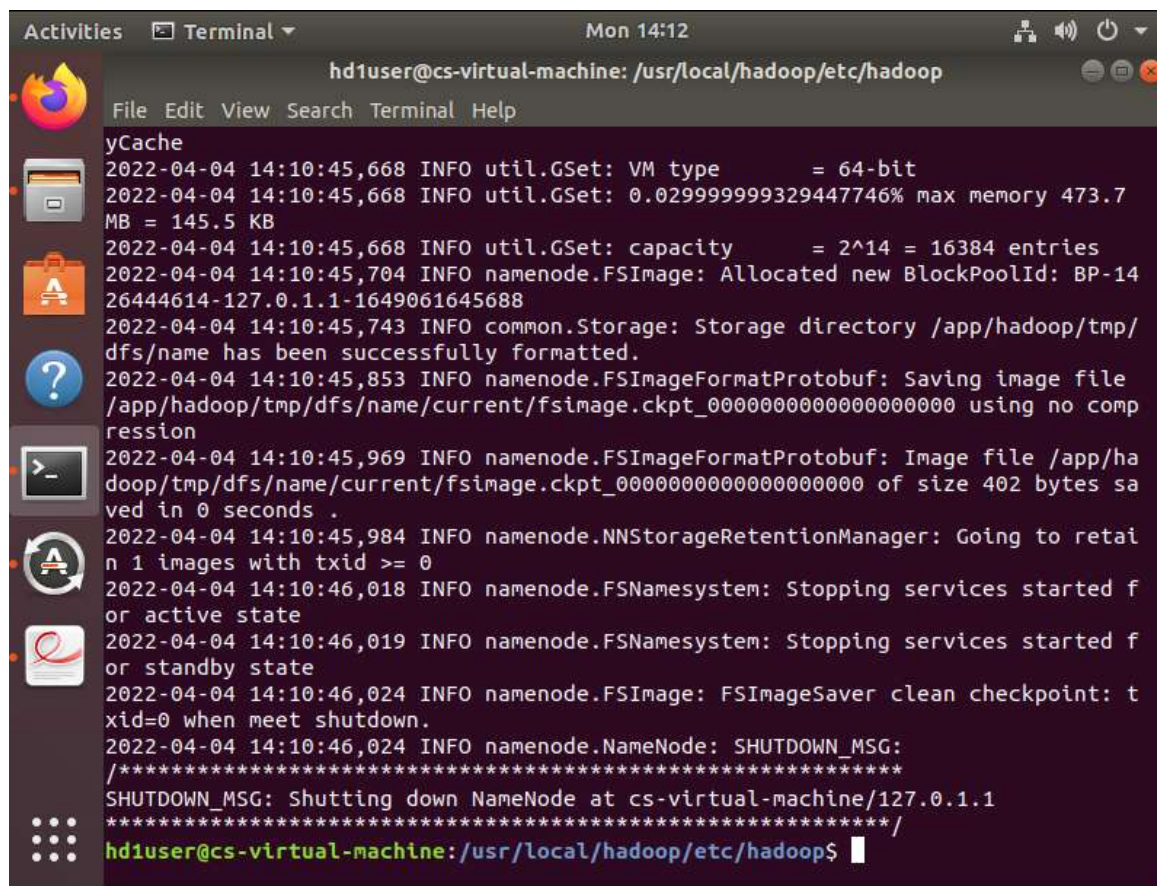
11. Finally we format namenode by the following commands:

hadoop namenode -format



```
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop$ hadoop namenode -format
WARNING: Use of this script to execute namenode is deprecated.
WARNING: Attempting to execute replacement "hdfs namenode" instead.

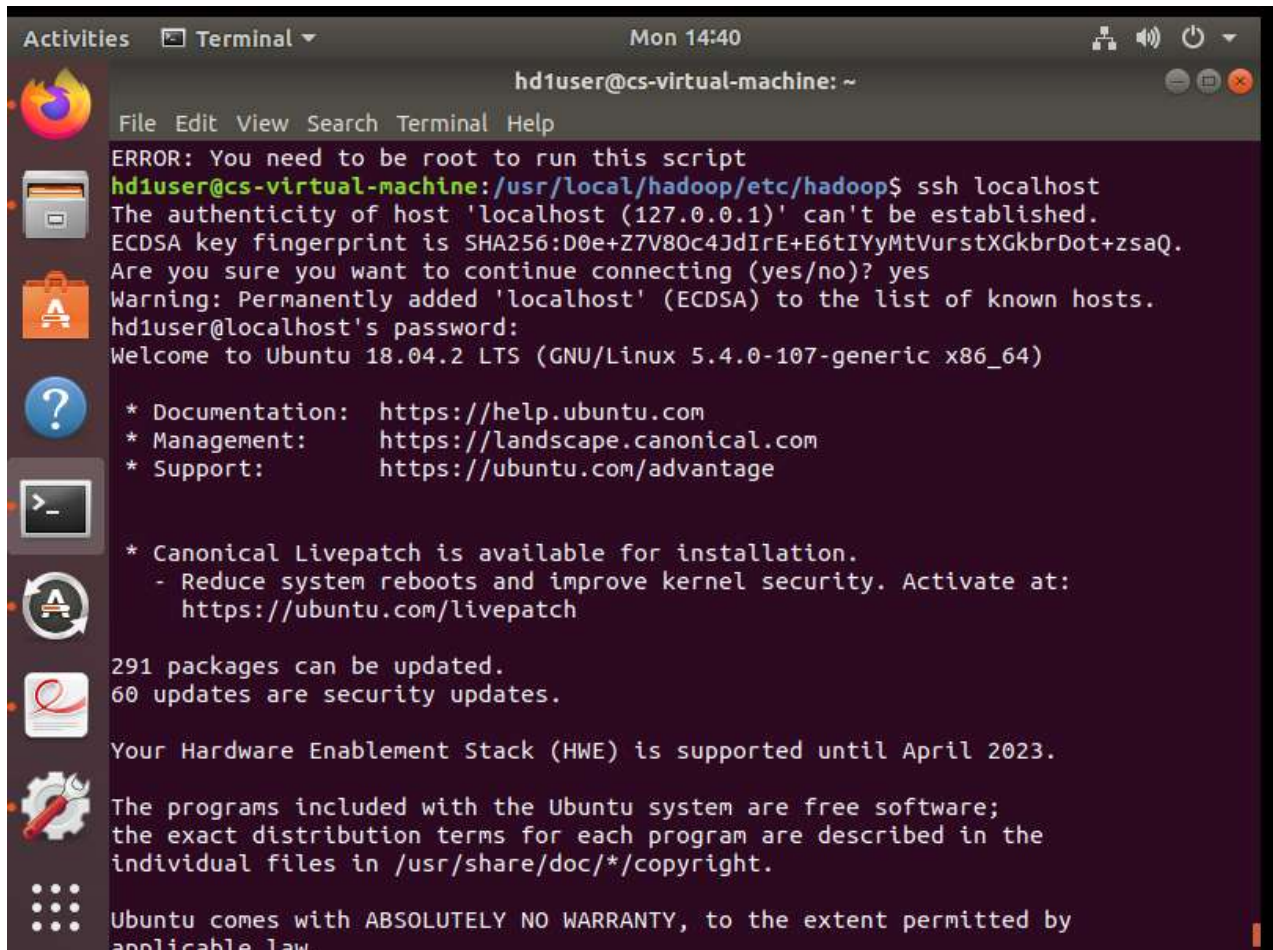
WARNING: /usr/local/hadoop/logs does not exist. Creating.
2022-04-04 14:10:44,135 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = cs-virtual-machine/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.2.3
STARTUP_MSG:   classpath = /usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share
/hadoop/common/lib/spotbugs-annotations-3.1.9.jar:/usr/local/hadoop/share/hadoo
p/common/lib/javax.activation-api-1.2.0.jar:/usr/local/hadoop/share/hadoop/comm
on/lib/dnsjava-2.1.7.jar:/usr/local/hadoop/share/hadoop/common/lib/kerb-identit
y-1.0.1.jar:/usr/local/hadoop/share/hadoop/common/lib/kerby-config-1.0.1.jar:/u
sr/local/hadoop/share/hadoop/common/lib/commons-io-2.8.0.jar:/usr/local/hadoop/
share/hadoop/common/lib/j2objc-annotations-1.1.jar:/usr/local/hadoop/share/hado
op/common/lib/commons-logging-1.1.3.jar:/usr/local/hadoop/share/hadoop/common/l
ib/curator-client-2.13.0.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-
configuration2-2.1.1.jar:/usr/local/hadoop/share/hadoop/common/lib/hadoop-annot
ations-3.2.3.jar:/usr/local/hadoop/share/hadoop/common/lib/animal-sniffer-annot
ations-1.17.jar:/usr/local/hadoop/share/hadoop/common/lib/jsr305-3.0.2.jar:/usr
/local/hadoop/share/hadoop/common/lib/commons-lang3-3.7.jar:/usr/local/hadoop/s
hare/hadoop/common/lib/jsp-api-2.1.jar:/usr/local/hadoop/share/hadoop/common/li
b/error_prone_annotations-2.2.0.jar:/usr/local/hadoop/share/hadoop/common/lib/f
ailureaccess-1.0.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-cli-1.2.
```



```
yCache
2022-04-04 14:10:45,668 INFO util.GSet: VM type      = 64-bit
2022-04-04 14:10:45,668 INFO util.GSet: 0.029999999329447746% max memory 473.7
MB = 145.5 KB
2022-04-04 14:10:45,668 INFO util.GSet: capacity    = 2^14 = 16384 entries
2022-04-04 14:10:45,704 INFO namenode.FSImage: Allocated new BlockPoolId: BP-14
26444614-127.0.1.1-1649061645688
2022-04-04 14:10:45,743 INFO common.Storage: Storage directory /app/hadoop/tmp/
dfs/name has been successfully formatted.
2022-04-04 14:10:45,853 INFO namenode.FSImageFormatProtobuf: Saving image file
/app/hadoop/tmp/dfs/name/current/fsimage.ckpt_000000000000000000 using no comp
ression
2022-04-04 14:10:45,969 INFO namenode.FSImageFormatProtobuf: Image file /app/ha
doo/tmp/dfs/name/current/fsimage.ckpt_000000000000000000 of size 402 bytes sa
ved in 0 seconds .
2022-04-04 14:10:45,984 INFO namenode.NNStorageRetentionManager: Going to retai
n 1 images with txid >= 0
2022-04-04 14:10:46,018 INFO namenode.FSNamesystem: Stopping services started f
or active state
2022-04-04 14:10:46,019 INFO namenode.FSNamesystem: Stopping services started f
or standby state
2022-04-04 14:10:46,024 INFO namenode.FSImage: FSImageSaver clean checkpoint: t
xid=0 when meet shutdown.
2022-04-04 14:10:46,024 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at cs-virtual-machine/127.0.1.1
*****/
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop$
```



Now to start hadoop, we need to run command: ssh localhost



```
hd1user@cs-virtual-machine: ~
ERROR: You need to be root to run this script
hd1user@cs-virtual-machine: /usr/local/hadoop/etc/hadoop$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:D0e+Z7V80c4JdIrE+E6tIYyMtVurstXGkbrDot+zsaQ.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
hd1user@localhost's password:
Welcome to Ubuntu 18.04.2 LTS (GNU/Linux 5.4.0-107-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

 * Canonical Livepatch is available for installation.
   - Reduce system reboots and improve kernel security. Activate at:
     https://ubuntu.com/livepatch

291 packages can be updated.
60 updates are security updates.

Your Hardware Enablement Stack (HWE) is supported until April 2023.

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

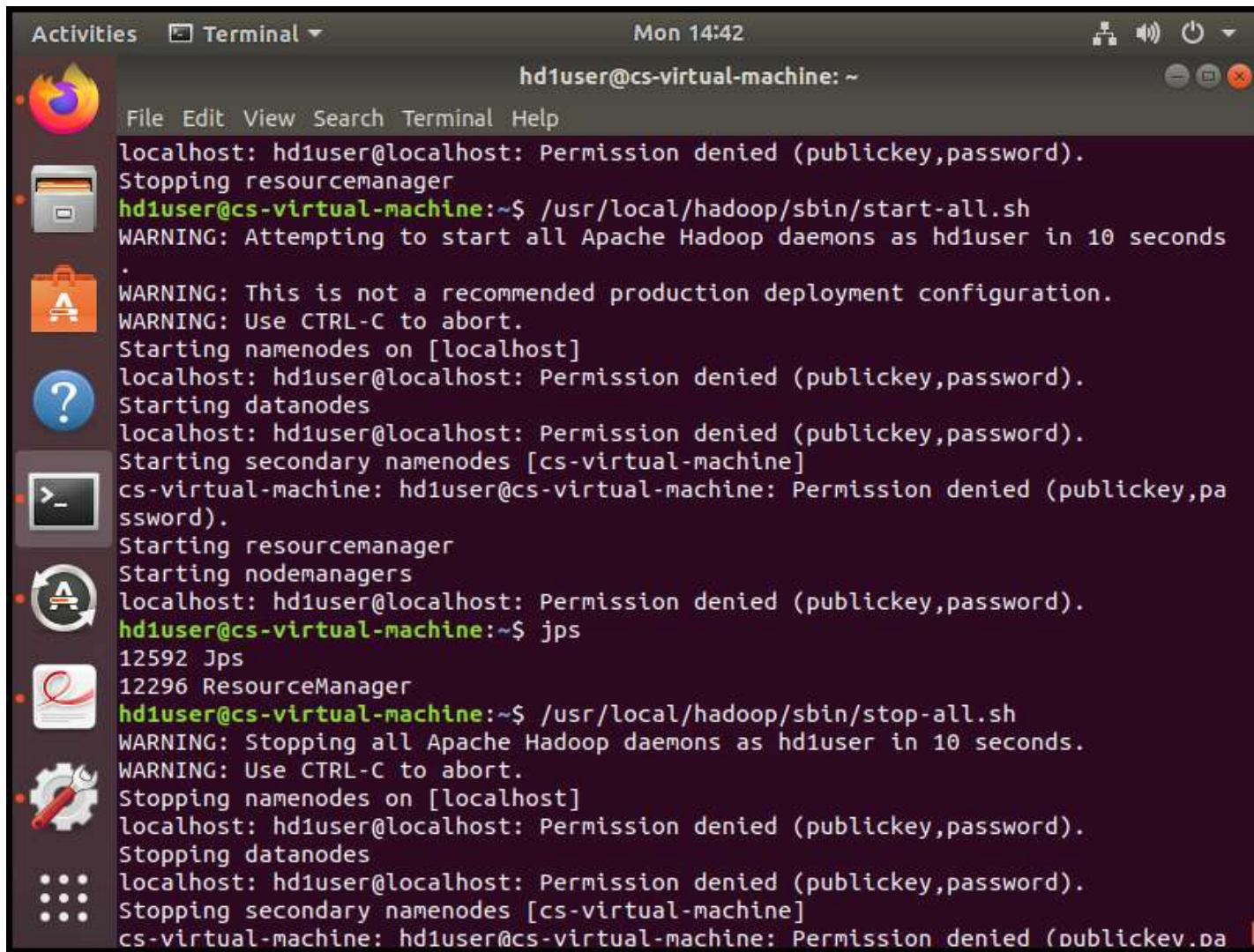
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.
```

Now to start hadoop, we need to run command:  
ssh localhost

This command is to start hadoop services:  
/usr/local/hadoop/sbin/start-all.sh

This command is to check that all hadoop services are running (6 services should appear): jps

This command is to stop hadoop services:  
/usr/local/hadoop/sbin/stop-all.sh



The image shows a terminal window titled "Terminal" with the user "hd1user@cs-virtual-machine". The terminal output shows the execution of `/usr/local/hadoop/sbin/start-all.sh` and `/usr/local/hadoop/sbin/stop-all.sh`. Both commands result in "Permission denied (publickey,password)" errors on the local host and the virtual machine. The terminal also shows the execution of `jps` and `jps` commands, which output the following processes: 12592 Jps and 12296 ResourceManager.

```
hd1user@cs-virtual-machine: ~
File Edit View Search Terminal Help
localhost: hd1user@localhost: Permission denied (publickey,password).
Stopping resourcemanager
hd1user@cs-virtual-machine:~$ /usr/local/hadoop/sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hd1user in 10 seconds
.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
localhost: hd1user@localhost: Permission denied (publickey,password).
Starting datanodes
localhost: hd1user@localhost: Permission denied (publickey,password).
Starting secondary namenodes [cs-virtual-machine]
cs-virtual-machine: hd1user@cs-virtual-machine: Permission denied (publickey,pa
ssword).
Starting resourcemanager
Starting nodemanagers
localhost: hd1user@localhost: Permission denied (publickey,password).
hd1user@cs-virtual-machine:~$ jps
12592 Jps
12296 ResourceManager
hd1user@cs-virtual-machine:~$ /usr/local/hadoop/sbin/stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as hd1user in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
localhost: hd1user@localhost: Permission denied (publickey,password).
Stopping datanodes
localhost: hd1user@localhost: Permission denied (publickey,password).
Stopping secondary namenodes [cs-virtual-machine]
cs-virtual-machine: hd1user@cs-virtual-machine: Permission denied (publickev.da
```

\*\*\*Incase if ssh fails

[https://linuxhint.com/fix\\_connection\\_refused\\_ubuntu/](https://linuxhint.com/fix_connection_refused_ubuntu/)

```
$ sudo apt list --installed | grep openssh-server
```

```
$ sudo apt install openssh-server
```

```
$ sudo apt list --installed | grep openssh-server
```

```
$ sudo service ssh status
```

```
$ sudo service ssh start
```

```
$ sudo service ssh restart
```

\*\*Incase if all services not starting for hadoop

<https://gautambangalore.medium.com/resolve-permission-issue-among-datanodes-with-namenode-to-establish-secure-shell-ssh-without-a-1e76e2863111>

```
$ ssh-keygen -t rsa -P „" -f ~/.ssh/id_rsa
```

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

```
$ chmod 0600 ~/.ssh/authorized_keys
```

```
sudo service network-manager restart
```

```
sudo service ssh restart
```