



Lab - 01:

Cài đặt môi trường chạy Hadoop

Lớp: 19_21

Môn: Nhập môn dữ liệu lớn

Sinh viên thực hiện:

Lê Thành Lộc – 19120562

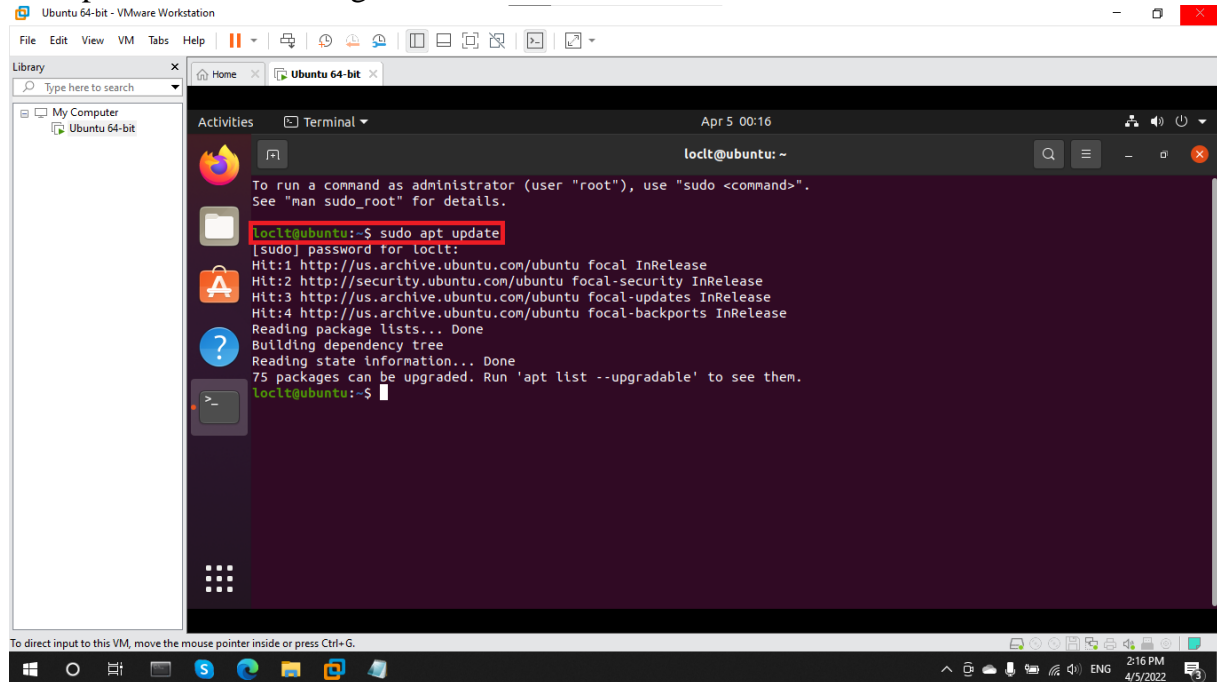
Mục lục

I. Các bước cài đặt và hình ảnh:	3
1. Cập nhật các chỉ mục gói của hệ điều hành Linux	3
2. Cài đặt phiên bản Java 1.8 và kiểm tra xem phiên bản đã được cài đặt chưa	4
3. Cài đặt OpenSSH Server trên Linux	5
4. Tạo SSH Key và SSH certificate với OpenSSH vừa được cài đặt	6
5. Cài đặt hadoop từ link https://downloads.apache.org/hadoop/common/hadoop-3.3.2/hadoop-3.3.2.tar.gz và sau đó tiến hành giải nén bằng lệnh <code>tar xzf</code>	7
7. Thực hiện chỉnh sửa hadoop-env.sh trong thư mục hadoop-3.3.2 được tải về	9
8. Thực hiện chỉnh sửa core-site.xml trong thư mục hadoop-3.3.2 được tải về	10
9. Thực hiện chỉnh sửa hdfs-site.xml trong thư mục hadoop-3.3.2 được tải về	11
10. Thực hiện chỉnh sửa mapred-site.xml trong thư mục hadoop-3.3.2 được tải về	12
11. Thực hiện chỉnh sửa yarn-site.xml trong thư mục hadoop-3.3.2 được tải về	13
II. Chạy thử Hadoop	14
1. Tạo user, chuyển quyền làm việc hiện tại sang tên user vừa tạo và khởi động namenode	14
2. Khởi động Hadoop Service	14
3. Chạy lệnh <code>jps</code> để liệt kê danh sách các JVM (Java HotSpot) đang có quyền truy cập	15
III. Tài liệu tham khảo:	16

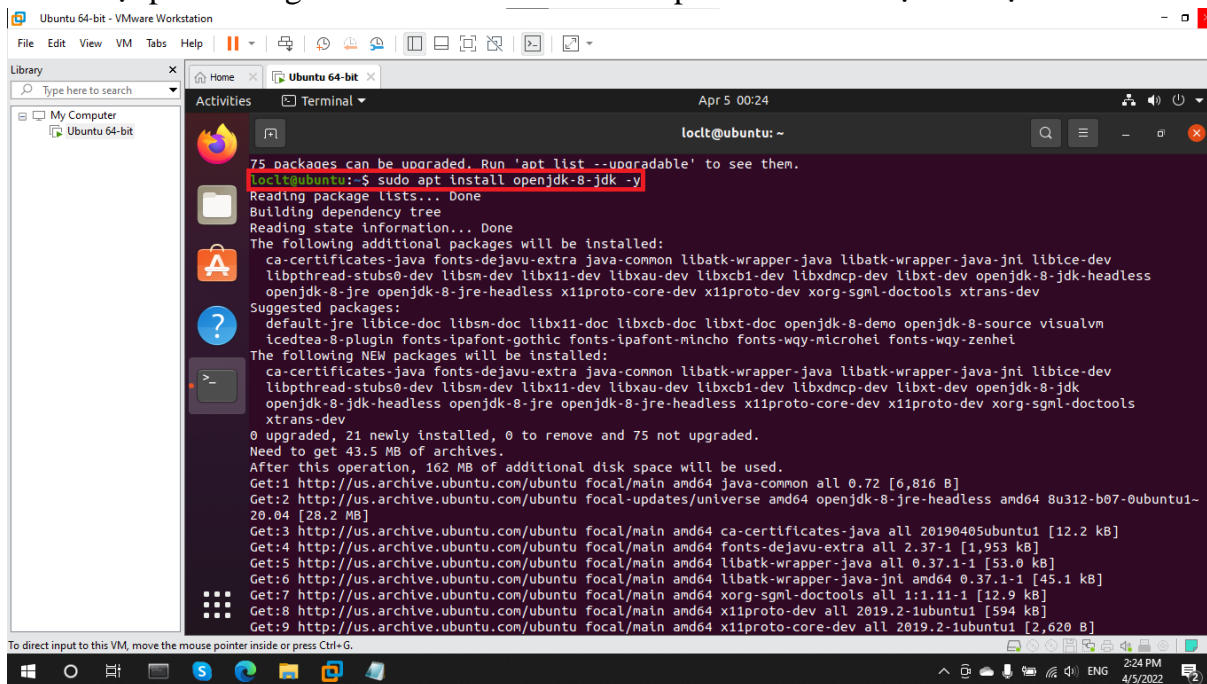
I. Các bước cài đặt và hình ảnh:

- Hệ điều hành được sử dụng: **Ubuntu 20.04**
- Lần lượt thực hiện các bước cài đặt Hadoop như sau:

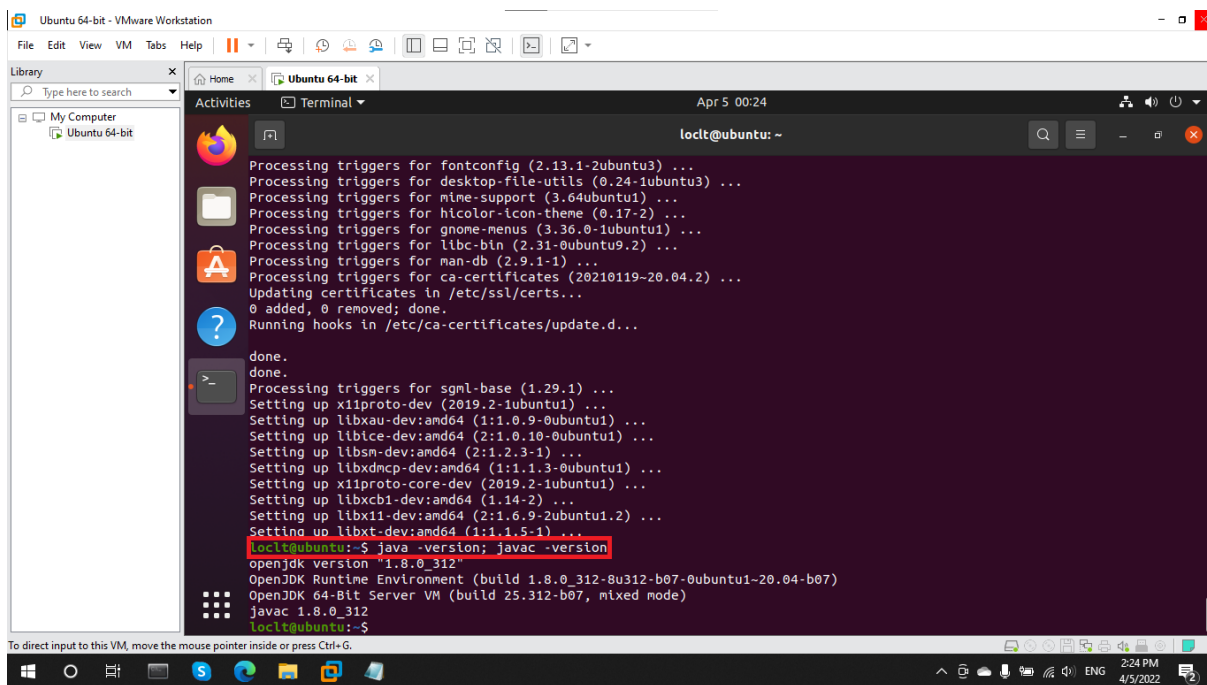
1. Cập nhật các chỉ mục gói của hệ điều hành Linux



2. Cài đặt phiên bản **Java 1.8** và kiểm tra xem phiên bản đã được cài đặt chưa

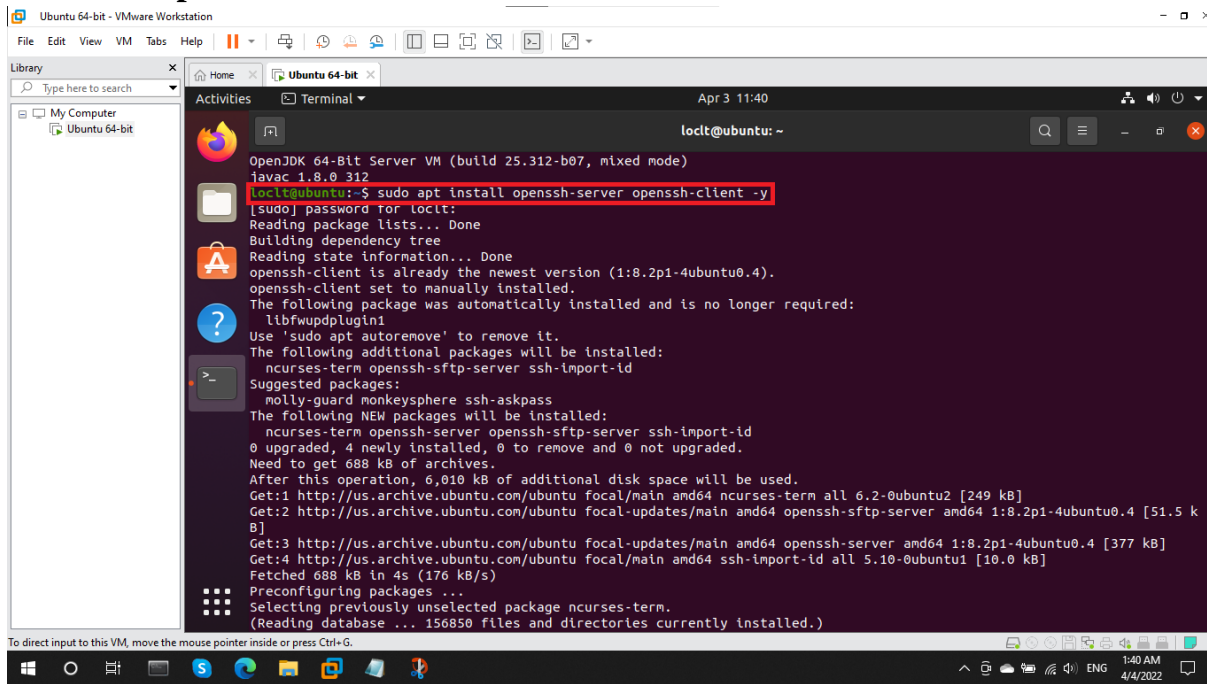


```
locit@ubuntu: ~  
75 packages can be upgraded. Run 'apt list --upgradable' to see them.  
locit@ubuntu:~$ sudo apt install openjdk-8-jdk -y  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
The following additional packages will be installed:  
  ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java libatk-wrapper-java-jni libice-dev  
  libpthread-stubs0-dev libsm-dev libx11-dev libxau-dev 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 libx11-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 libice-dev  
  libpthread-stubs0-dev libsm-dev libx11-dev 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  
0 upgraded, 21 newly installed, 0 to remove and 75 not upgraded.  
Need to get 43.5 MB of archives.  
After this operation, 162 MB of additional disk space will be used.  
Get:1 http://us.archive.ubuntu.com/ubuntu focal/main amd64 java-common all 0.72 [6,816 B]  
Get:2 http://us.archive.ubuntu.com/ubuntu focal-updates/universe amd64 openjdk-8-jre-headless amd64 8u312-b07-0ubuntu1~  
20.04 [28.2 MB]  
Get:3 http://us.archive.ubuntu.com/ubuntu focal/main amd64 ca-certificates-java all 20190405ubuntu1 [12.2 kB]  
Get:4 http://us.archive.ubuntu.com/ubuntu focal/main amd64 fonts-dejavu-extra all 2.37-1 [1,953 kB]  
Get:5 http://us.archive.ubuntu.com/ubuntu focal/main amd64 libatk-wrapper-java all 0.37.1-1 [53.0 kB]  
Get:6 http://us.archive.ubuntu.com/ubuntu focal/main amd64 libatk-wrapper-java-jni amd64 0.37.1-1 [45.1 kB]  
Get:7 http://us.archive.ubuntu.com/ubuntu focal/main amd64 xorg-sgml-doctools all 1:1.11-1 [12.9 kB]  
Get:8 http://us.archive.ubuntu.com/ubuntu focal/main amd64 x11proto-dev all 2019.2-1ubuntu1 [594 kB]  
Get:9 http://us.archive.ubuntu.com/ubuntu focal/main amd64 x11proto-core-dev all 2019.2-1ubuntu1 [2,620 B]  
Processing triggers for fontconfig (2.13.1-2ubuntu3) ...  
Processing triggers for desktop-file-utils (0.24-1ubuntu3) ...  
Processing triggers for mime-support (3.64ubuntu1) ...  
Processing triggers for hicolor-icon-theme (0.17-2) ...  
Processing triggers for gnome-menus (3.36.0-1ubuntu1) ...  
Processing triggers for libc-bin (2.31-0ubuntu9.2) ...  
Processing triggers for man-db (2.9.1-1) ...  
Processing triggers for ca-certificates (20210119~20.04.2) ...  
Updating certificates in /etc/ssl/certs...  
0 added, 0 removed; done.  
Running hooks in /etc/ca-certificates/update.d...  
done.  
done.  
Processing triggers for sgml-base (1.29.1) ...  
Setting up x11proto-dev (2019.2-1ubuntu1) ...  
Setting up libxau-dev:amd64 (1:1.0.9-0ubuntu1) ...  
Setting up libice-dev:amd64 (2:1.0.10-0ubuntu1) ...  
Setting up libsm-dev:amd64 (2:1.2.3-1) ...  
Setting up libxdmcp-dev:amd64 (1:1.1.3-0ubuntu1) ...  
Setting up x11proto-core-dev (2019.2-1ubuntu1) ...  
Setting up libxcb1-dev:amd64 (1.14-2) ...  
Setting up libx11-dev:amd64 (2:1.6.9-2ubuntu1.2) ...  
Setting up libxt-dev:amd64 (1:1.1.5-1) ...  
locit@ubuntu:~$ java -version; javac -version  
openjdk version "1.8.0_312"  
OpenJDK Runtime Environment (build 1.8.0_312-8u312-b07-0ubuntu1~20.04-b07)  
OpenJDK 64-Bit Server VM (build 25.312-b07, mixed mode)  
javac 1.8.0_312  
locit@ubuntu:~$
```



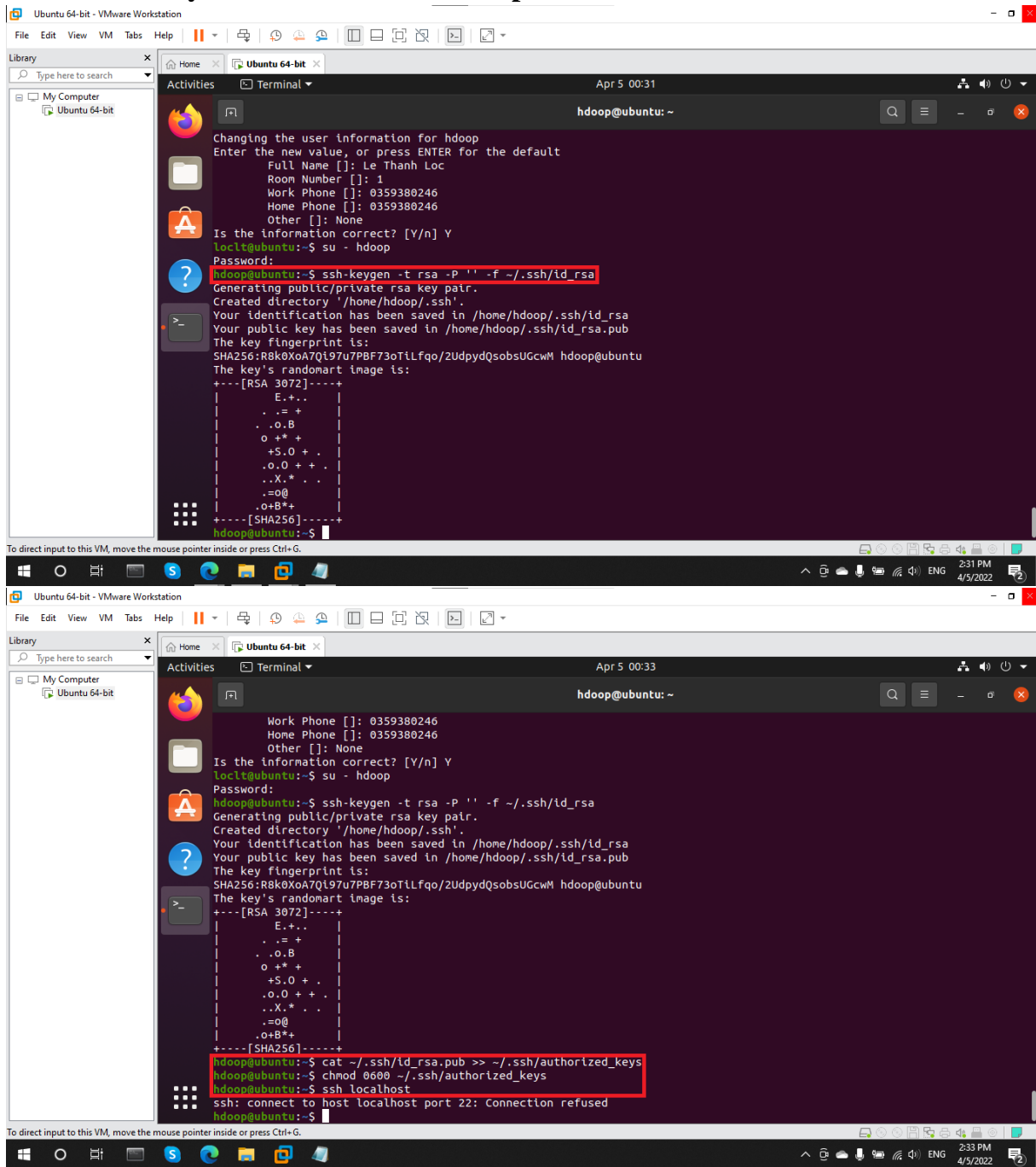
```
locit@ubuntu: ~  
Processing triggers for fontconfig (2.13.1-2ubuntu3) ...  
Processing triggers for desktop-file-utils (0.24-1ubuntu3) ...  
Processing triggers for mime-support (3.64ubuntu1) ...  
Processing triggers for hicolor-icon-theme (0.17-2) ...  
Processing triggers for gnome-menus (3.36.0-1ubuntu1) ...  
Processing triggers for libc-bin (2.31-0ubuntu9.2) ...  
Processing triggers for man-db (2.9.1-1) ...  
Processing triggers for ca-certificates (20210119~20.04.2) ...  
Updating certificates in /etc/ssl/certs...  
0 added, 0 removed; done.  
Running hooks in /etc/ca-certificates/update.d...  
done.  
done.  
Processing triggers for sgml-base (1.29.1) ...  
Setting up x11proto-dev (2019.2-1ubuntu1) ...  
Setting up libxau-dev:amd64 (1:1.0.9-0ubuntu1) ...  
Setting up libice-dev:amd64 (2:1.0.10-0ubuntu1) ...  
Setting up libsm-dev:amd64 (2:1.2.3-1) ...  
Setting up libxdmcp-dev:amd64 (1:1.1.3-0ubuntu1) ...  
Setting up x11proto-core-dev (2019.2-1ubuntu1) ...  
Setting up libxcb1-dev:amd64 (1.14-2) ...  
Setting up libx11-dev:amd64 (2:1.6.9-2ubuntu1.2) ...  
Setting up libxt-dev:amd64 (1:1.1.5-1) ...  
locit@ubuntu:~$ java -version; javac -version  
openjdk version "1.8.0_312"  
OpenJDK Runtime Environment (build 1.8.0_312-8u312-b07-0ubuntu1~20.04-b07)  
OpenJDK 64-Bit Server VM (build 25.312-b07, mixed mode)  
javac 1.8.0_312  
locit@ubuntu:~$
```

3. Cài đặt OpenSSH Server trên Linux



```
OpenJDK 64-Bit Server VM (build 25.312-b07, mixed mode)
javac 1.8.0_312
locit@ubuntu:~$ sudo apt install openssh-server openssh-client -y
[sudo] password for locit:
Reading package lists... Done
Building dependency tree
Reading state information... Done
openssh-client is already the newest version (1:8.2p1-4ubuntu0.4).
openssh-client set to manually installed.
The following package was automatically installed and is no longer required:
  libfwpdpplugin1
Use 'sudo apt autoremove' to remove it.
The following additional packages will be installed:
  ncurses-term openssh-sftp-server ssh-import-id
Suggested packages:
  molly-guard monkeysphere ssh-askpass
The following NEW packages will be installed:
  ncurses-term openssh-server openssh-sftp-server ssh-import-id
0 upgraded, 4 newly installed, 0 to remove and 0 not upgraded.
Need to get 688 kB of archives.
After this operation, 6,010 kB of additional disk space will be used.
Get:1 http://us.archive.ubuntu.com/ubuntu focal/main amd64 ncurses-term all 6.2-0ubuntu2 [249 kB]
Get:2 http://us.archive.ubuntu.com/ubuntu focal-updates/main amd64 openssh-sftp-server amd64 1:8.2p1-4ubuntu0.4 [51.5 kB]
Get:3 http://us.archive.ubuntu.com/ubuntu focal-updates/main amd64 openssh-server amd64 1:8.2p1-4ubuntu0.4 [377 kB]
Get:4 http://us.archive.ubuntu.com/ubuntu focal/main amd64 ssh-import-id all 5.10-0ubuntu1 [10.0 kB]
Fetched 688 kB in 4s (176 kB/s)
Preconfiguring packages ...
Selecting previously unselected package ncurses-term.
(Reading database ... 156850 files and directories currently installed.)
```

4. Tạo SSH Key và SSH certificate với OpenSSH vừa được cài đặt



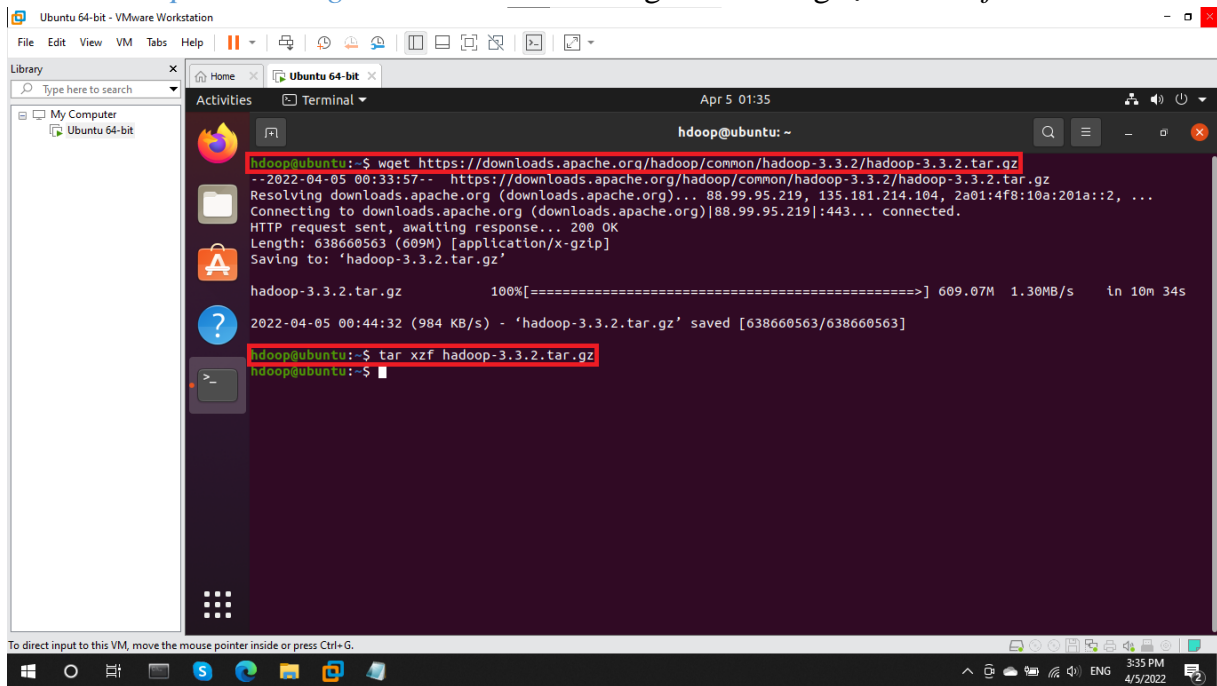
The image shows two screenshots of a terminal window running on an Ubuntu 64-bit VM. The terminal is titled 'hdoop@ubuntu: ~' and shows the following commands and output:

```
Changing the user information for hdoop
Enter the new value, or press ENTER for the default
Full Name []: Le Thanh Loc
Room Number []: 1
Work Phone []: 0359380246
Home Phone []: 0359380246
Other []: None
Is the information correct? [Y/n] Y
loc1t@ubuntu:~$ su - hdoop
Password:
hdoop@ubuntu:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
Created directory '/home/hdoop/.ssh'.
Your identification has been saved in /home/hdoop/.ssh/id_rsa
Your public key has been saved in /home/hdoop/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:R8k0XoA7Q197u7PBF73oTlLfQo/2UdpydQsobsUGcwM hdoop@ubuntu
The key's randomart image is:
+---[RSA 3072]-----+
|      E.+..      |
|      .+=+      |
|      .o.B      |
|      o++      |
|      +S.O +    |
|      .o.O +    |
|      ..X.*     |
|      .=@       |
|      .o+B*+    |
|      +---[SHA256]-----+
hdoop@ubuntu:~$
```

The second screenshot shows the continuation of the process:

```
Work Phone []: 0359380246
Home Phone []: 0359380246
Other []: None
Is the information correct? [Y/n] Y
loc1t@ubuntu:~$ su - hdoop
Password:
hdoop@ubuntu:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
Created directory '/home/hdoop/.ssh'.
Your identification has been saved in /home/hdoop/.ssh/id_rsa
Your public key has been saved in /home/hdoop/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:R8k0XoA7Q197u7PBF73oTlLfQo/2UdpydQsobsUGcwM hdoop@ubuntu
The key's randomart image is:
+---[RSA 3072]-----+
|      E.+..      |
|      .+=+      |
|      .o.B      |
|      o++      |
|      +S.O +    |
|      .o.O +    |
|      ..X.*     |
|      .=@       |
|      .o+B*+    |
|      +---[SHA256]-----+
hdoop@ubuntu:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
hdoop@ubuntu:~$ chmod 0600 ~/.ssh/authorized_keys
hdoop@ubuntu:~$ ssh localhost
ssh: connect to host localhost port 22: Connection refused
hdoop@ubuntu:~$
```

5. Cài đặt hadoop từ link <https://downloads.apache.org/hadoop/common/hadoop-3.3.2/hadoop-3.3.2.tar.gz> và sau đó tiến hành giải nén bằng lệnh `tar xzf`



```
hadoop@ubuntu: ~  
hadoop@ubuntu:~$ wget https://downloads.apache.org/hadoop/common/hadoop-3.3.2/hadoop-3.3.2.tar.gz  
--2022-04-05 00:33:57-- https://downloads.apache.org/hadoop/common/hadoop-3.3.2/hadoop-3.3.2.tar.gz  
Resolving downloads.apache.org (downloads.apache.org)... 88.99.95.219, 135.181.214.104, 2a01:4f8:10a:201a::2, ...  
Connecting to downloads.apache.org (downloads.apache.org)|88.99.95.219|:443... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 638660563 (609M) [application/x-gzip]  
Saving to: 'hadoop-3.3.2.tar.gz'  
  
hadoop-3.3.2.tar.gz      100%[=====] 609.07M  1.30MB/s   in 10m 34s  
  
2022-04-05 00:44:32 (984 KB/s) - 'hadoop-3.3.2.tar.gz' saved [638660563/638660563]  
  
hadoop@ubuntu:~$ tar xzf hadoop-3.3.2.tar.gz  
hadoop@ubuntu:~$
```

6. Thực hiện chỉnh sửa file **.bashrc**

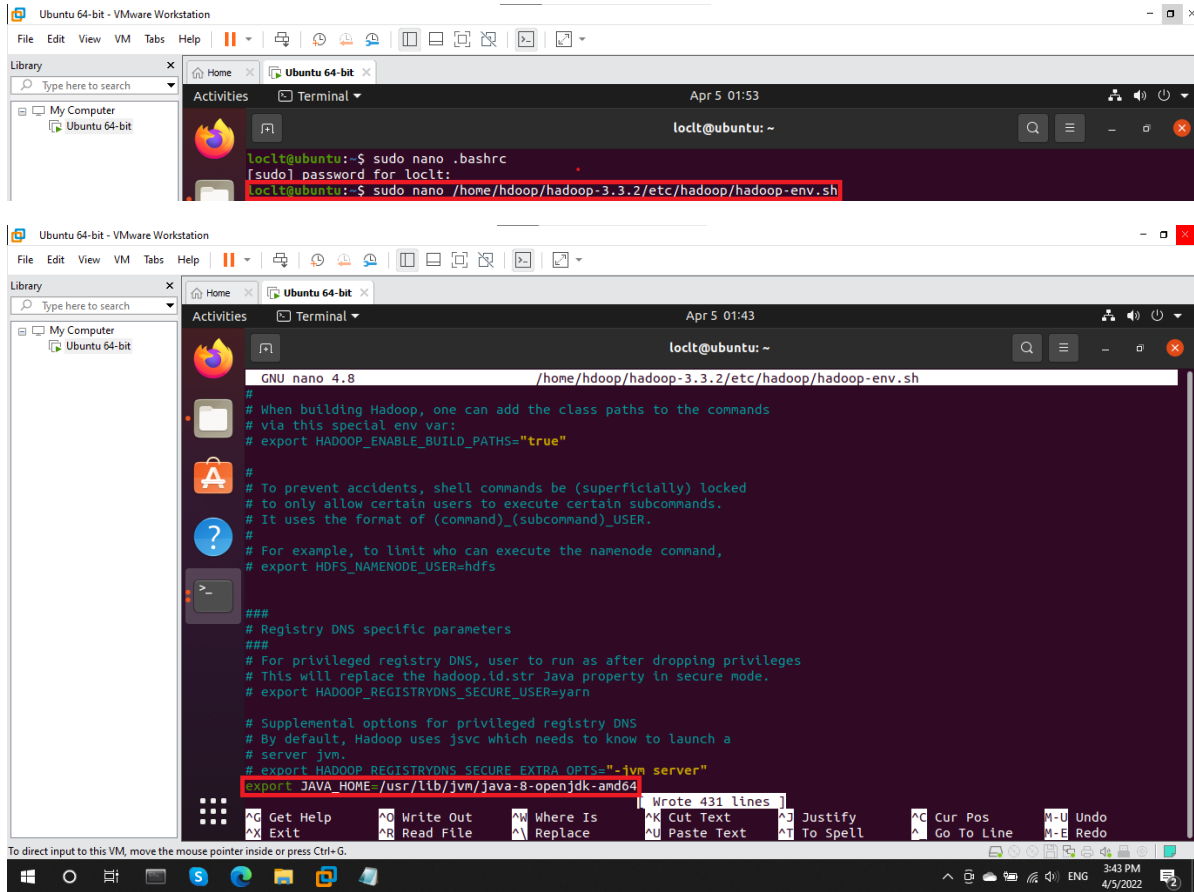
- Nhập lệnh `sudo nano .bashrc` để mở file **.bashrc**.
- Thêm các dòng lệnh `export` như hình vào cuối file.

The image consists of two screenshots of a VMware Workstation window showing an Ubuntu 64-bit virtual machine. The top screenshot shows the terminal with the command `loclt@ubuntu:~$ sudo nano .bashrc` entered. The bottom screenshot shows the nano editor interface with the `.bashrc` file open. The file content includes standard bash completion settings and a section for Hadoop configuration. The Hadoop configuration section is highlighted with a red box and contains the following lines:

```
#Hadoop Related Options
export HADOOP_HOME=/home/hdoop/hadoop-3.3.2
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH:$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

The bottom screenshot also shows the nano editor's status bar at the bottom, indicating "Wrote 127 lines".

7. Thực hiện chỉnh sửa **hadoop-env.sh** trong thư mục **hadoop-3.3.2** được tải về
- Nhập lệnh `sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/hadoop-env.sh` để mở file.
 - Thêm dòng lệnh **export** vào cuối file như bên dưới và lưu lại.

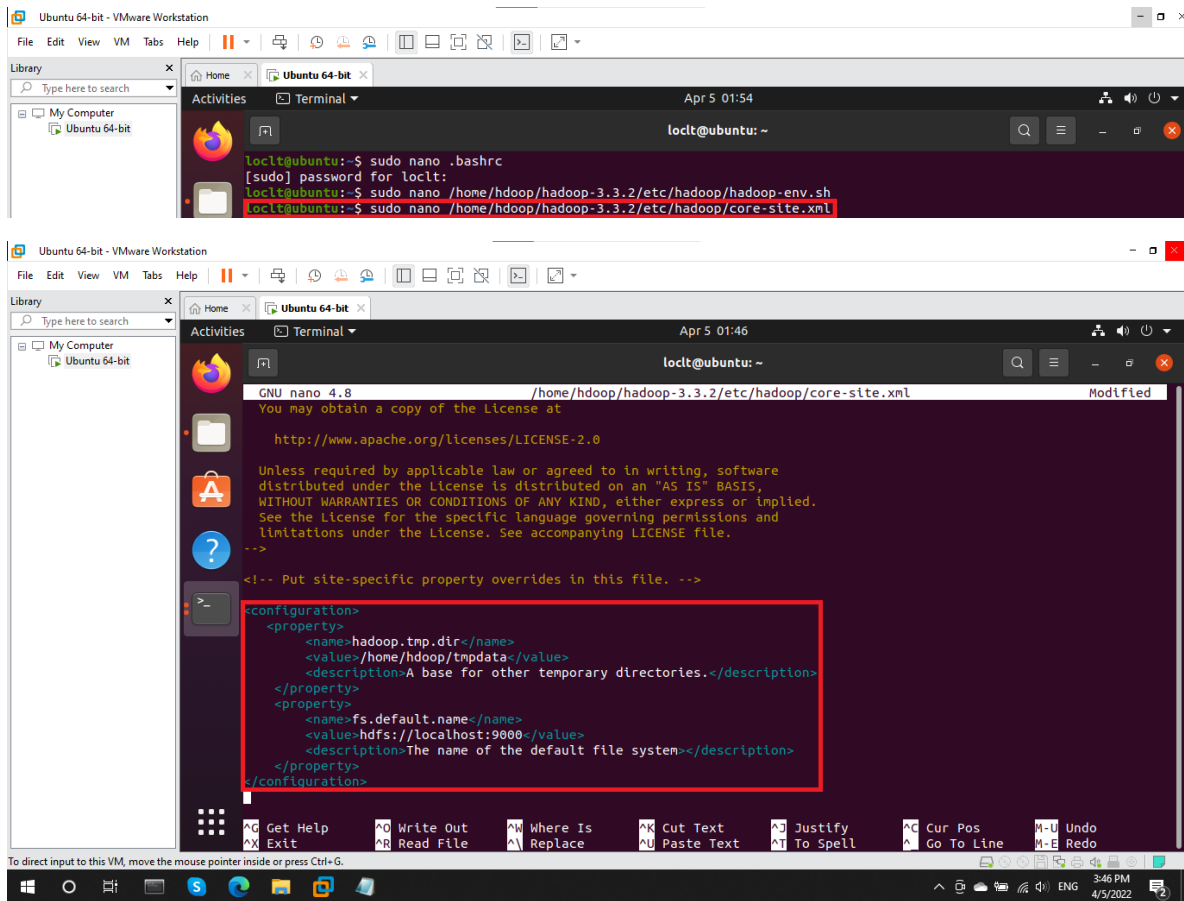


The image consists of two screenshots of a VMware Workstation window showing an Ubuntu 64-bit virtual machine. The top screenshot shows the terminal with the command `sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/hadoop-env.sh` being executed. The bottom screenshot shows the nano editor interface with the file `/home/hdoop/hadoop-3.3.2/etc/hadoop/hadoop-env.sh` open. The file content includes comments about building Hadoop, setting environment variables, and registry DNS parameters. The line `export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64` is highlighted in red. The bottom status bar of the nano editor shows "Wrote 431 lines".

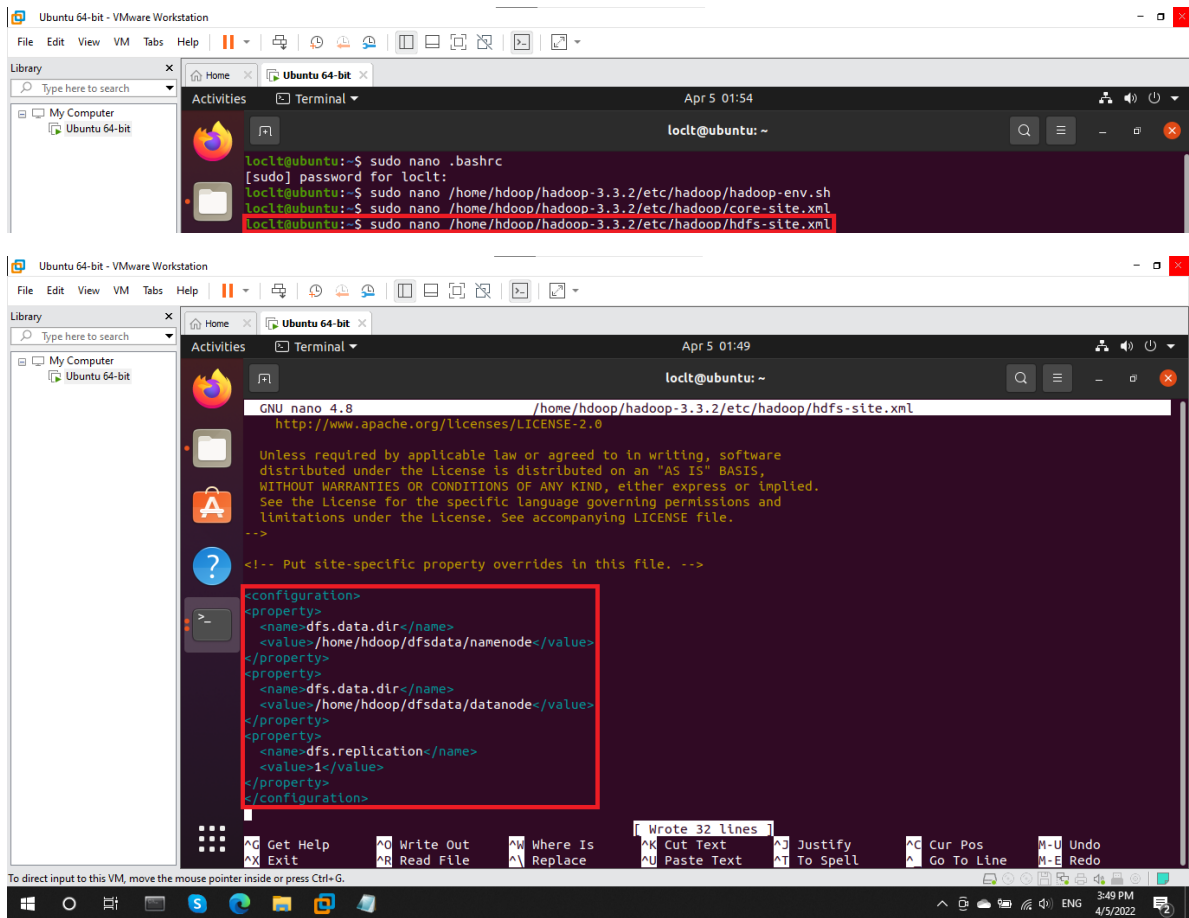
```
loc@ubuntu:~$ sudo nano .bashrc
[sudo] password for loc:
loc@ubuntu:~$ sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/hadoop-env.sh

GNU nano 4.8 /home/hdoop/hadoop-3.3.2/etc/hadoop/hadoop-env.sh
#
# 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
###
# Registry DNS specific parameters
###
# For privileged registry DNS, user to run as after dropping privileges
# This will replace the hadoop.id.str Java property in secure mode.
# export HADOOP_REGISTRYDNS_SECURE_USER=yarn
# Supplemental options for privileged registry DNS
# By default, Hadoop uses jsvc which needs to know to launch a
# server jvm.
# export HADOOP_REGISTRYDNS_SECURE_EXTRA_OPTS="--jvm server"
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

8. Thực hiện chỉnh sửa **core-site.xml** trong thư mục **hadoop-3.3.2** được tải về
- Nhập lệnh `sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/core-site.xml` để mở file.
 - Thêm các dòng lệnh vào giữa các tag **configuration** như hình vẽ bên dưới.



9. Thực hiện chỉnh sửa **hdfs-site.xml** trong thư mục **hadoop-3.3.2** được tải về
- Nhập lệnh `sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/hdfs-site.xml` để mở file.
 - Thêm các dòng lệnh vào giữa các tag **configuration** như hình bên dưới.



The image consists of two screenshots of a terminal window within a VMware Workstation environment, showing the process of editing the `hdfs-site.xml` file.

The first screenshot shows the terminal with the following commands and output:

```
loc1t@ubuntu:~$ sudo nano .bashrc
[sudo] password for loc1t:
loc1t@ubuntu:~$ sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/hadoop-env.sh
loc1t@ubuntu:~$ sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/core-site.xml
loc1t@ubuntu:~$ sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/hdfs-site.xml
```

The second screenshot shows the `nano` editor editing the file `/home/hdoop/hadoop-3.3.2/etc/hadoop/hdfs-site.xml`. The file content is as follows:

```
GNU nano 4.8 /home/hdoop/hadoop-3.3.2/etc/hadoop/hdfs-site.xml
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. -->

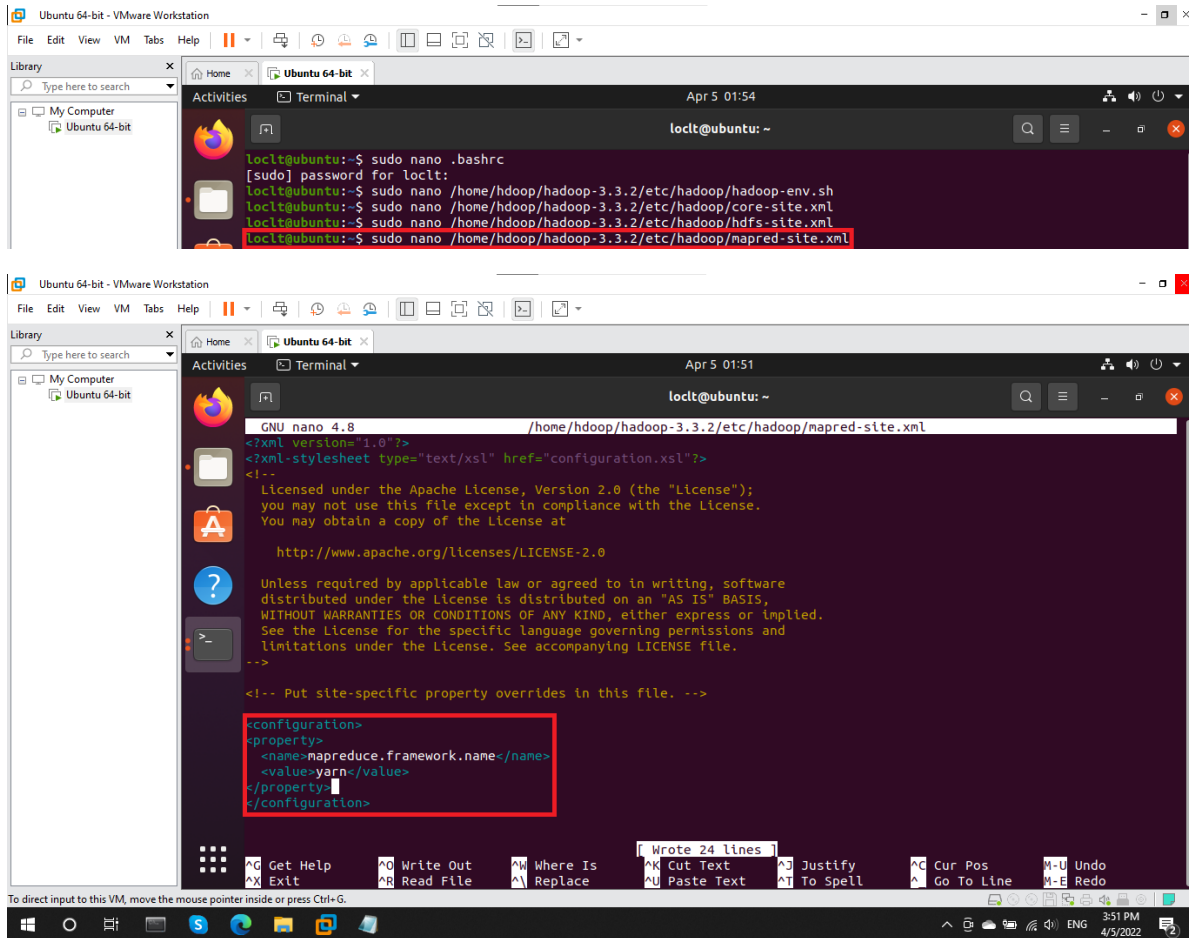
<configuration>
<property>
  <name>dfs.data.dir</name>
  <value>/home/hdoop/dfsdata/namenode</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>/home/hdoop/dfsdata/datanode</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
</configuration>
```

The `<configuration>` tag and its contents are highlighted with a red box. The terminal status bar at the bottom indicates "Wrote 32 lines".

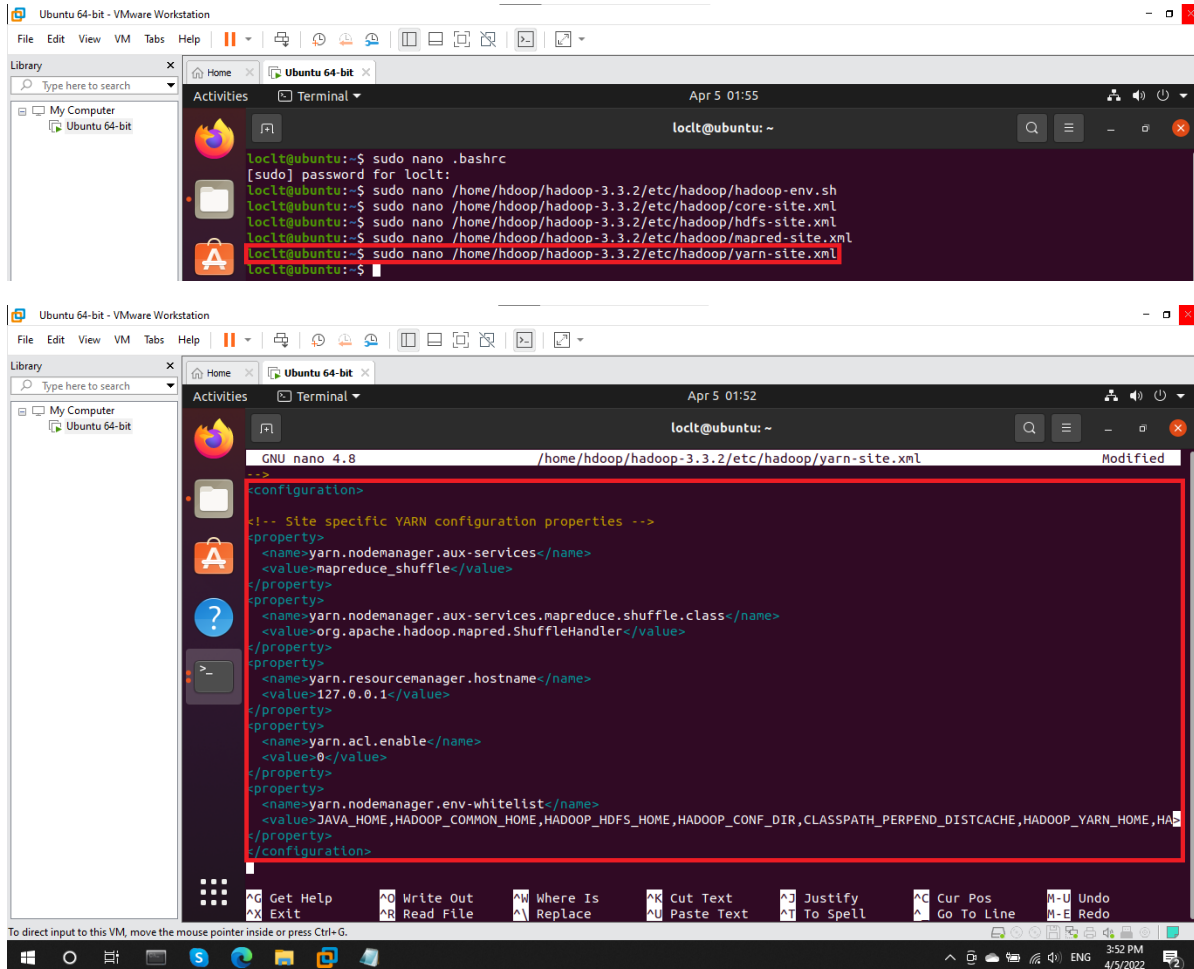
10. Thực hiện chỉnh sửa **mapred-site.xml** trong thư mục **hadoop-3.3.2** được tải về

- Nhập lệnh `sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/mapred-site.xml` để mở file.

- Thêm các dòng lệnh vào giữa các tag **configuration** như hình bên dưới.



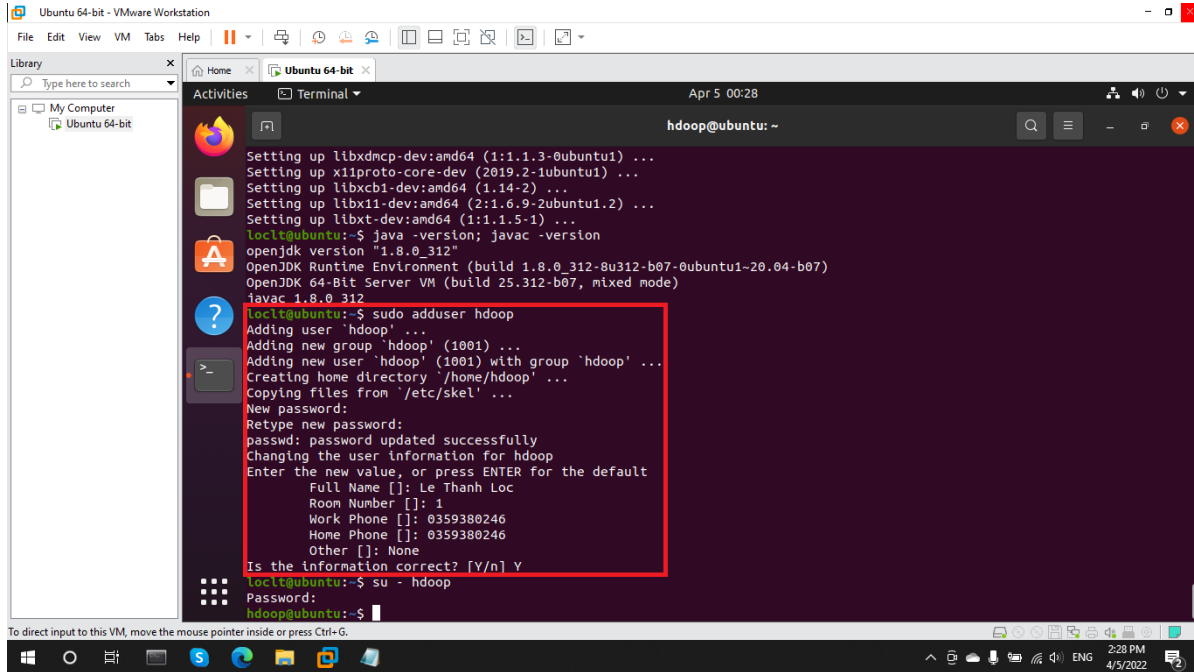
11. Thực hiện chỉnh sửa **yarn-site.xml** trong thư mục **hadoop-3.3.2** được tải về
- Nhập lệnh `sudo nano /home/hadoop/hadoop-3.3.2/etc/hadoop/yarn-site.xml` để mở file.
 - Thêm các dòng lệnh vào giữa các tag **configuration** như hình bên dưới.



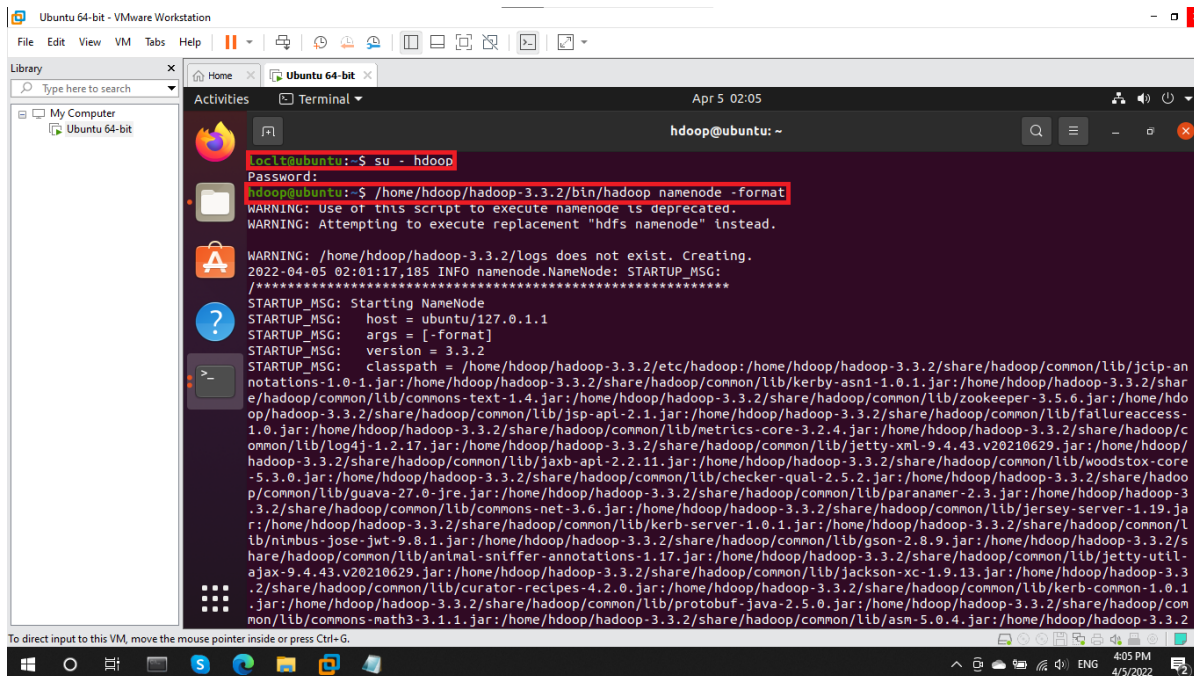
⇒ Hoàn tất các bước cài đặt Hadoop

II. Chạy thử Hadoop

1. Tạo user, chuyển quyền làm việc hiện tại sang tên user vừa tạo và khởi động namenode



```
Setting up libxdmcp-dev:amd64 (1:1.1.3-0ubuntu1) ...
Setting up x11proto-core-dev (2019.2-1ubuntu1) ...
Setting up libxcb1-dev:amd64 (1.14-2) ...
Setting up libx11-dev:amd64 (2:1.6.9-2ubuntu1.2) ...
Setting up libxt-dev:amd64 (1:1.1.5-1) ...
loclt@ubuntu:~$ java -version; javac -version
openjdk version "1.8.0_312"
OpenJDK Runtime Environment (build 1.8.0_312-8u312-b07-0ubuntu1~20.04-b07)
OpenJDK 64-Bit Server VM (build 25.312-b07, mixed mode)
javac 1.8.0_312
loclt@ubuntu:~$ sudo adduser hadoop
Adding user 'hadoop' ...
Adding new group 'hadoop' (1001) ...
Adding new user 'hadoop' (1001) with group 'hadoop' ...
Creating home directory '/home/hadoop' ...
Copying files from '/etc/skel' ...
New password:
Retype new password:
passwd: password updated successfully
Changing the user information for hadoop
Enter the new value, or press ENTER for the default
Full Name []: Le Thanh Loc
Room Number []: 1
Work Phone []: 0359380246
Home Phone []: 0359380246
Other []: None
Is the information correct? [Y/n] Y
loclt@ubuntu:~$ su - hadoop
Password:
hadoop@ubuntu:~$
```

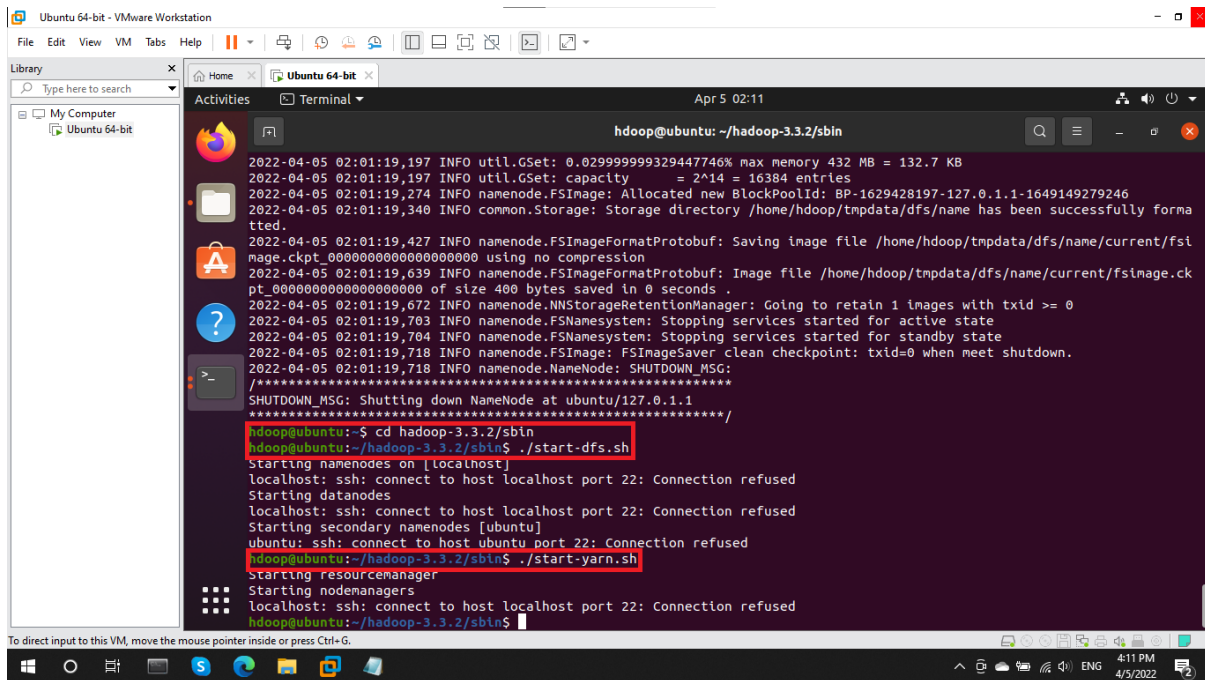


```
loclt@ubuntu:~$ su - hadoop
Password:
hadoop@ubuntu:~$ /home/hadoop/hadoop-3.3.2/bin/hadoop namenode -format
WARNING: Use of this script to execute namenode is deprecated.
WARNING: Attempting to execute replacement "hdfs namenode" instead.
WARNING: /home/hadoop/hadoop-3.3.2/logs does not exist. Creating.
2022-04-05 02:01:17,185 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = ubuntu/127.0.1.1
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.3.2
STARTUP_MSG: classpath = /home/hadoop/hadoop-3.3.2/etc/hadoop:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/jcip-annotations-1.0-1.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/kerby-asn1-1.0.1.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/commons-text-1.4.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/zookeeper-3.5.6.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/jsp-apt-2.1.1.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/failureaccess-1.0.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/metrics-core-3.2.4.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/log4j-1.2.17.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/jetty-xml-9.4.43.v20210629.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/jaxb-apt-2.2.11.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/woodstox-core-5.3.0.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/checker-qual-2.5.2.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/guava-27.0-jre.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/paranamer-2.3.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/commons-net-3.6.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/jersey-server-1.19.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/kerb-server-1.0.1.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/nimbus-jose-jwt-9.0.1.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/gson-2.8.9.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/animal-sniffer-annotations-1.17.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/jetty-util-ajax-9.4.43.v20210629.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/jackson-xc-1.9.13.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/curator-recipes-4.2.0.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/kerb-common-1.0.1.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/protobuf-java-2.5.0.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/commons-math3-3.1.1.jar:/home/hadoop/hadoop-3.3.2/share/hadoop/common/lib/asn-5.0.4.jar:/home/hadoop/hadoop-3.3.2
```

2. Khởi động Hadoop Service

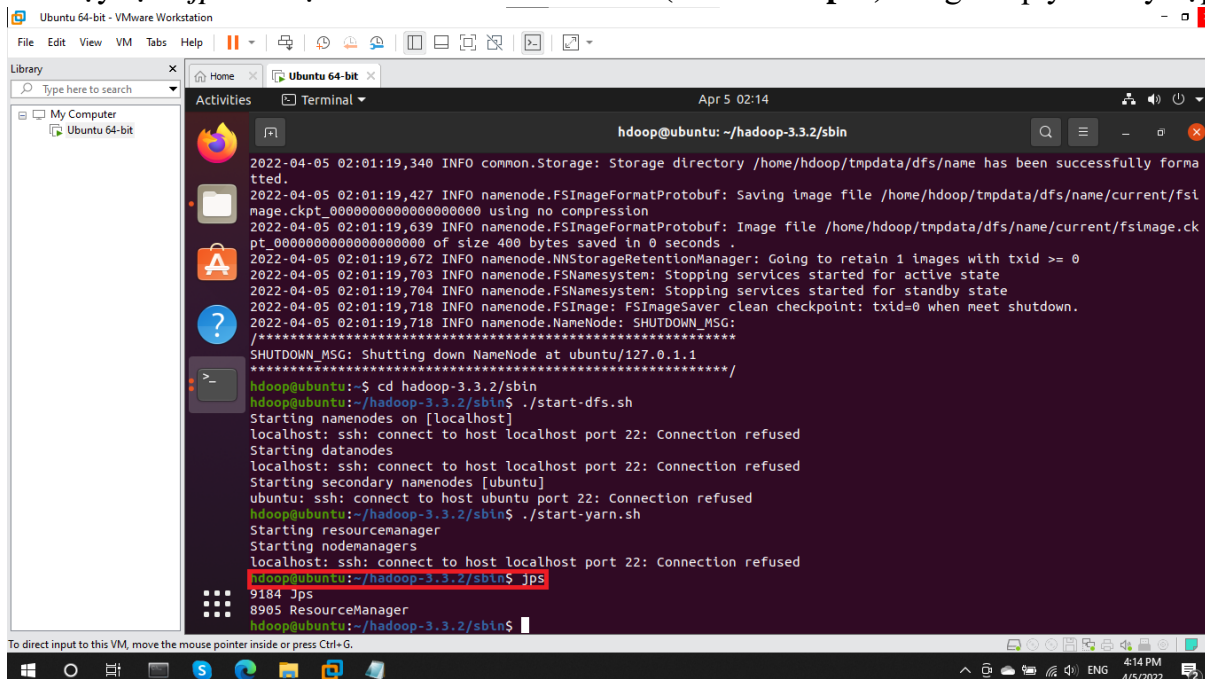
- Chuyển đến thư mục **sbin** trong **hadoop-3.3.2**.

- Chạy 2 lệnh `./start-dfs.sh` và `./start-yarn.sh` để khởi động **NameNode**, **DataNode**, **ResourceManager** và **NodeManagers**



```
hadoop@ubuntu: ~/hadoop-3.3.2/sbin
2022-04-05 02:01:19,197 INFO util.GSet: 0.029999999329447746% max memory 432 MB = 132.7 KB
2022-04-05 02:01:19,197 INFO util.GSet: capacity = 2^14 = 16384 entries
2022-04-05 02:01:19,274 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1629428197-127.0.1.1-1649149279246
2022-04-05 02:01:19,340 INFO common.Storage: Storage directory /home/hadoop/tmpdata/dfs/name has been successfully formatted.
2022-04-05 02:01:19,427 INFO namenode.FSImageFormatProtobuf: Saving image file /home/hadoop/tmpdata/dfs/name/current/fsimage.ckpt_000000000000000000 using no compression
2022-04-05 02:01:19,639 INFO namenode.FSImageFormatProtobuf: Image file /home/hadoop/tmpdata/dfs/name/current/fsimage.ckpt_000000000000000000 of size 400 bytes saved in 0 seconds
2022-04-05 02:01:19,672 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2022-04-05 02:01:19,703 INFO namenode.FSNamesystem: Stopping services started for active state
2022-04-05 02:01:19,704 INFO namenode.FSNamesystem: Stopping services started for standby state
2022-04-05 02:01:19,718 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2022-04-05 02:01:19,718 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at ubuntu/127.0.1.1
*****/
hadoop@ubuntu:~/hadoop-3.3.2/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
localhost: ssh: connect to host localhost port 22: Connection refused
Starting datanodes
localhost: ssh: connect to host localhost port 22: Connection refused
Starting secondary namenodes [ubuntu]
ubuntu: ssh: connect to host ubuntu port 22: Connection refused
hadoop@ubuntu:~/hadoop-3.3.2/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
localhost: ssh: connect to host localhost port 22: Connection refused
hadoop@ubuntu:~/hadoop-3.3.2/sbin$
```

3. Chạy lệnh `jps` để liệt kê danh sách các **JVM (Java HotSpot)** đang có quyền truy cập



```
hadoop@ubuntu: ~/hadoop-3.3.2/sbin
2022-04-05 02:01:19,340 INFO common.Storage: Storage directory /home/hadoop/tmpdata/dfs/name has been successfully formatted.
2022-04-05 02:01:19,427 INFO namenode.FSImageFormatProtobuf: Saving image file /home/hadoop/tmpdata/dfs/name/current/fsimage.ckpt_000000000000000000 using no compression
2022-04-05 02:01:19,639 INFO namenode.FSImageFormatProtobuf: Image file /home/hadoop/tmpdata/dfs/name/current/fsimage.ckpt_000000000000000000 of size 400 bytes saved in 0 seconds
2022-04-05 02:01:19,672 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2022-04-05 02:01:19,703 INFO namenode.FSNamesystem: Stopping services started for active state
2022-04-05 02:01:19,704 INFO namenode.FSNamesystem: Stopping services started for standby state
2022-04-05 02:01:19,718 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2022-04-05 02:01:19,718 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at ubuntu/127.0.1.1
*****/
hadoop@ubuntu:~/hadoop-3.3.2/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
localhost: ssh: connect to host localhost port 22: Connection refused
Starting datanodes
localhost: ssh: connect to host localhost port 22: Connection refused
Starting secondary namenodes [ubuntu]
ubuntu: ssh: connect to host ubuntu port 22: Connection refused
hadoop@ubuntu:~/hadoop-3.3.2/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
localhost: ssh: connect to host localhost port 22: Connection refused
hadoop@ubuntu:~/hadoop-3.3.2/sbin$ jps
9184 jps
8905 ResourceManager
hadoop@ubuntu:~/hadoop-3.3.2/sbin$
```

III. Tài liệu tham khảo:

[1] Powerpoint Hướng dẫn lab 1: Môi trường Hadoop.

[2] Hadoop Installation Guide Step by Step:

<https://www.youtube.com/watch?v=Ih5cuJYYz6Y>

[3] Guide file from video above:

<https://drive.google.com/file/d/17B8ptBee2YCWsqXpTm1UASW5Tezua68O/view>