Lab01: Set Up a Hadoop 3.2.1 Multi-Node Cluster on Ubuntu (2 Nodes)

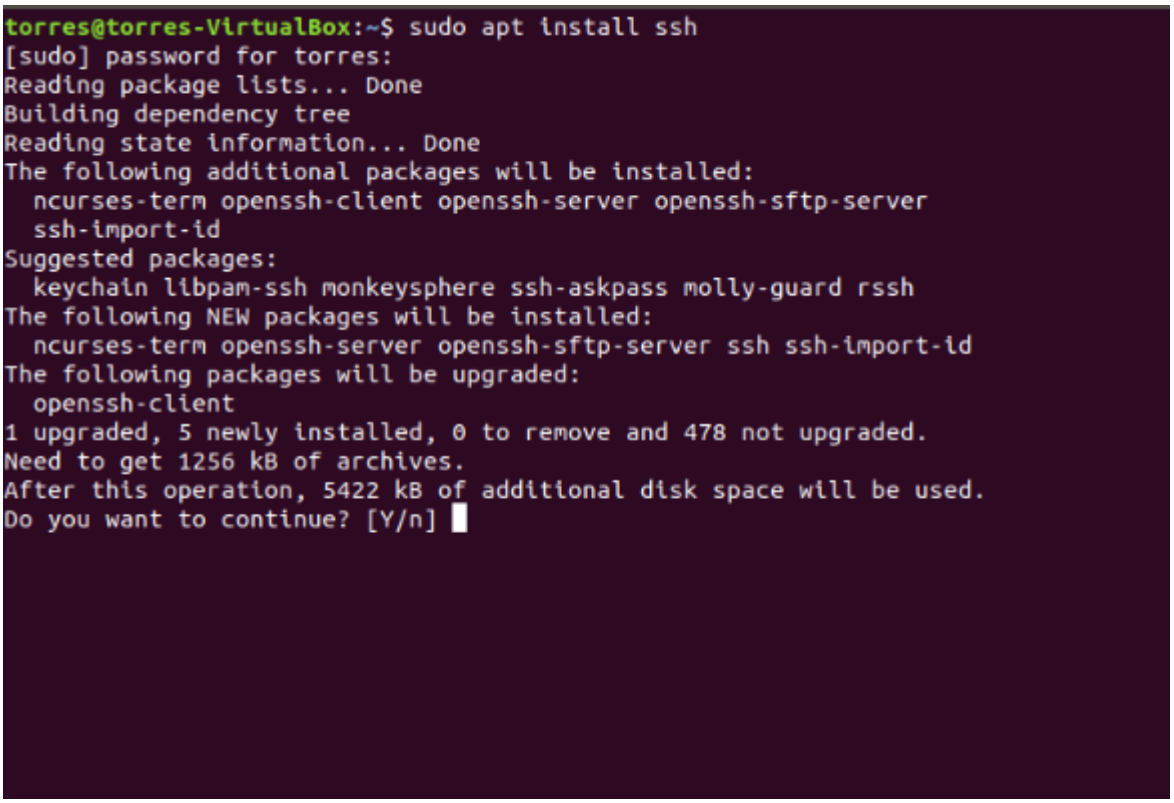
I. Trên toàn bộ máy

1st Step: Cài đặt SSH, PDSH

- Cài SSH

```
sudo apt install ssh
```

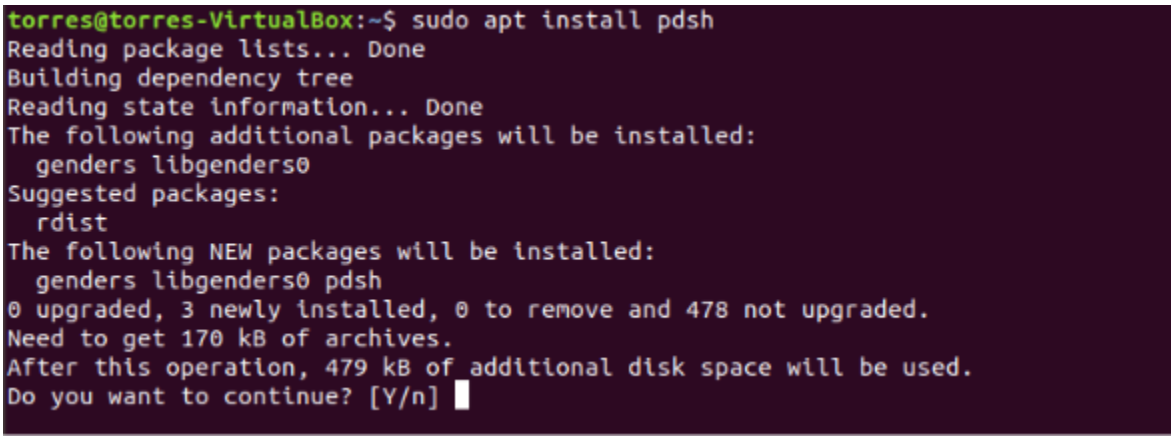
Nếu hỏi phải nhập password. Hãy nhập nó



- Cài PDSH

```
sudo apt install pdsh
```

Như trước, nhấy nhập thông tin khi được yêu cầu

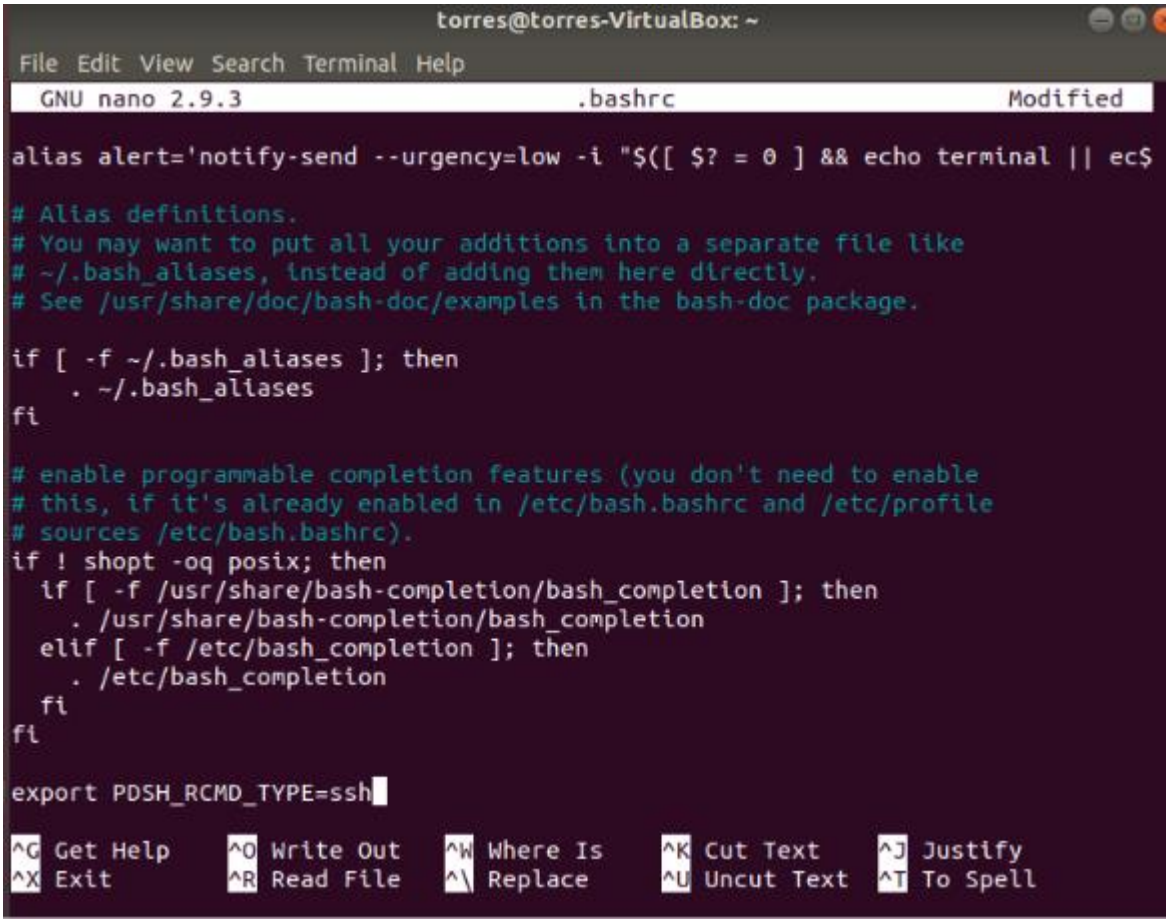


- Mở file .bashrc:

```
Sudo nano .bashrc
```

- Ở cuối file, thêm dòng

```
export PDSH_RCMD_TYPE=ssh
```



- Để lưu file:

`Ctrl X`

`Y`

`Enter`

- Config SSH. Tạo 1 khóa bằng lệnh:

`ssh-keygen -t rsa -P ""`

Bấm Enter mỗi khi cần

```
torres@torres-VirtualBox:~$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/torres/.ssh/id_rsa):
Your identification has been saved in /home/torres/.ssh/id_rsa.
Your public key has been saved in /home/torres/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:4mAZ5iJJkH2jfUJz3ZThFq+41knubh5kvRO7cHyA1os torres@torres-VirtualBox
The key's randomart image is:
+---[RSA 2048]----+
|.o     . o+o     |
|o . = . ..oo     |
| . =o+    o .    |
|...ooo.  o =     |
|o . =o. S B =    |
| . o o . O + *   |
|      . o E B .  |
|       . ..+ +   |
|         ++ .    |
+----[SHA256]-----+
torres@torres-VirtualBox:~$
```

- Copy public key cho authorized_keys với lệnh:

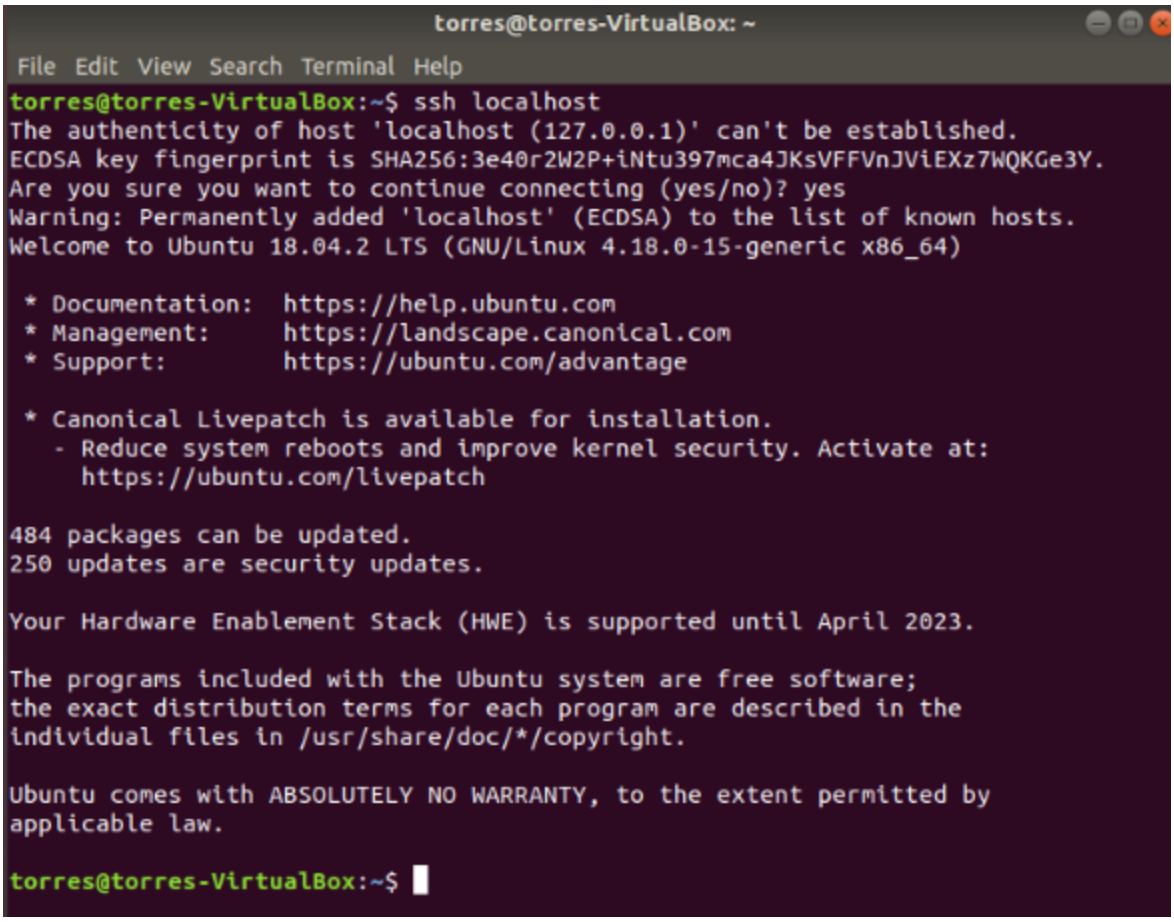
```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

```
torres@torres-VirtualBox:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
torres@torres-VirtualBox:~$
```

- Check SSH bằng việc connect đến localhost

```
ssh localhost
```

Nhập Yes và bấm Enter mỗi khi cần



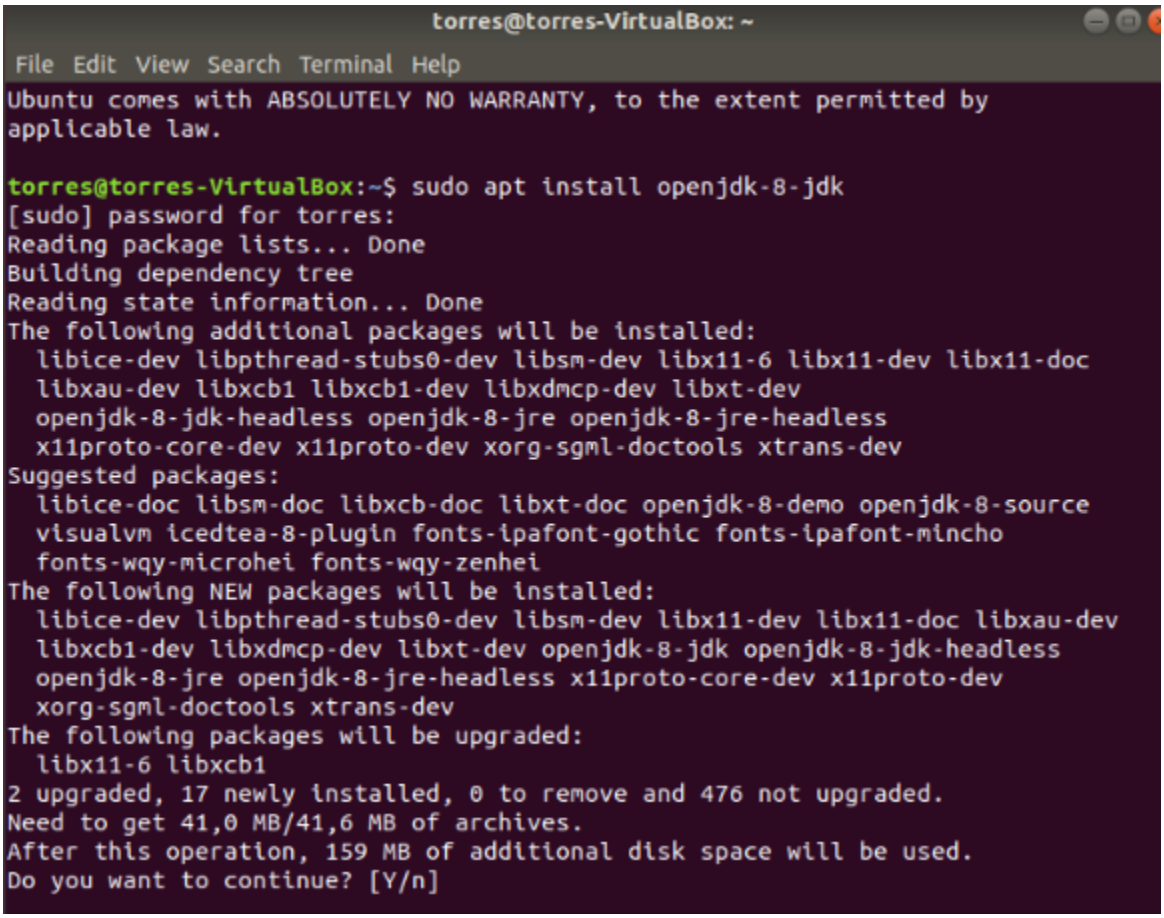
Cài đặt thành công cài SSH, PDSH

2nd Step: Cài đặt Java 8

- Để cài java 8. Nhập lệnh

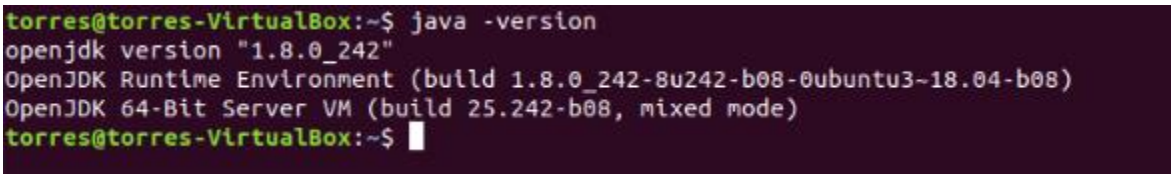
```
sudo apt install openjdk-8-jdk
```

- Như những phần trước, bấm mật khẩu hay trả lời mỗi khi cần



- Check Java version

```
java -version
```



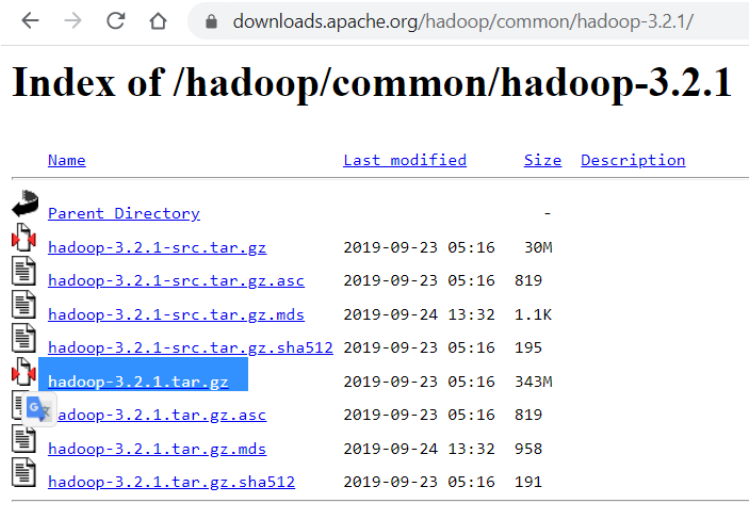
3rd Step: Tải Hadoop

- Tải hadoop bằng lệnh sau:

`sudo wget -P ~`
https://mirrors.sonic.net/apache/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz

```
torres@torres-VirtualBox:~$ sudo wget -P ~ https://mirrors.sonic.net/apache/had
oop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz
--2020-02-02 21:11:42--  https://mirrors.sonic.net/apache/hadoop/common/hadoop-
3.2.1/hadoop-3.2.1.tar.gz
Resolving mirrors.sonic.net (mirrors.sonic.net)... 157.131.0.16, 2001:5a8:601:2
:157:131:0:16
Connecting to mirrors.sonic.net (mirrors.sonic.net)|157.131.0.16|:443... connec
ted.
HTTP request sent, awaiting response... 200 OK
Length: 359196911 (343M) [application/x-gzip]
Saving to: '/home/torres/hadoop-3.2.1.tar.gz'

hadoop-3.2.1.tar.gz    0%[                    ] 744,00K   313KB/s
```

Node: Trong vài trường hợp, link trên sẽ bị chết và bạn không tải được. Lên google gõ hadoop-3.2.1.tar.gz tìm đến link dưới, tải nó về và di chuyển nó từ mục download vào thư mục home



- Giải nén hadoop-3.2.1.tar.gz file:

`tar xzf hadoop-3.2.1.tar.gz`

```
torres@torres-VirtualBox:~$ tar xzf hadoop-3.2.1.tar.gz
```

- Đổi tên hadoop-3.2.1 thành hadoop:

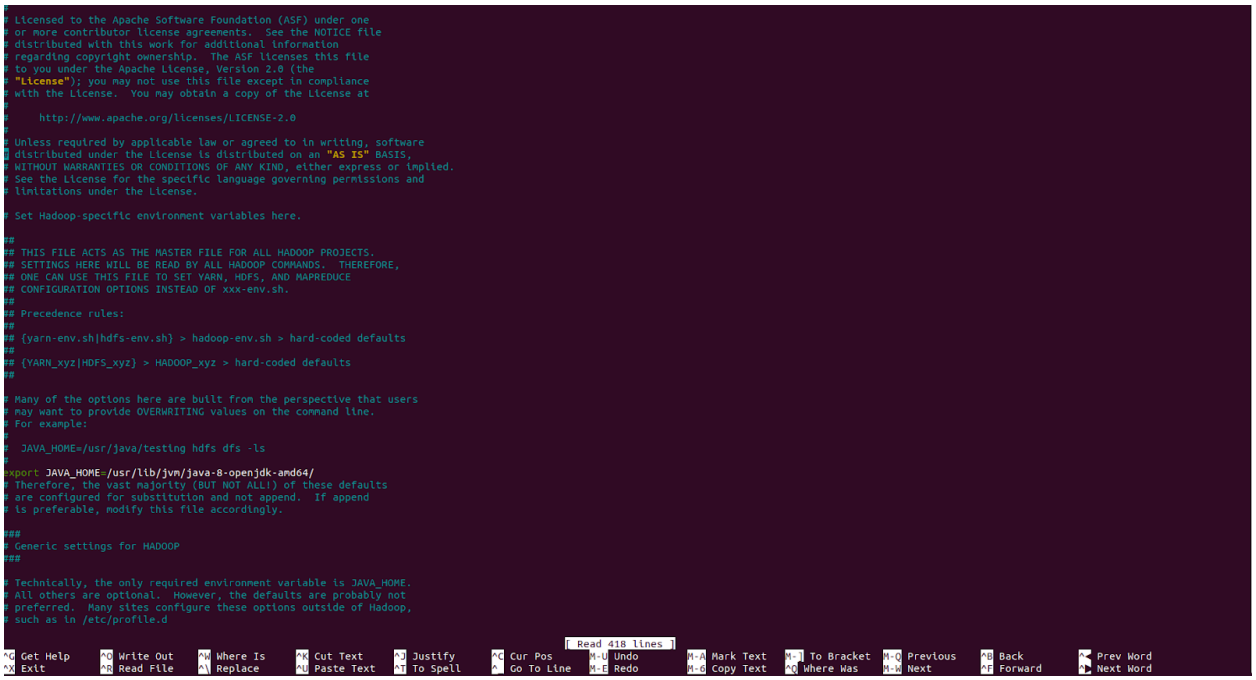
`mv hadoop-3.2.1 hadoop`

- Mở file hadoop-env.sh và sửa JAVA_HOME:

```
Sudo nano ~/hadoop/etc/hadoop/hadoop-env.sh
```

- Thêm dòng này vào:

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



Lưu lại file

- Chuyển hadoop folder đến/usr/local/hadoop:

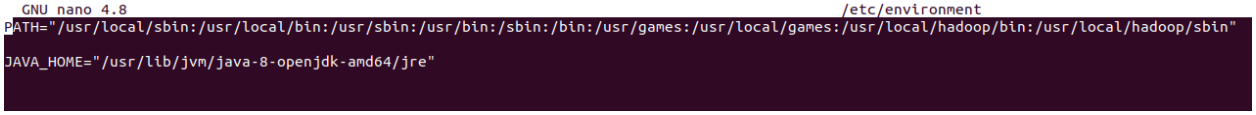
```
sudo mv hadoop /usr/local/hadoop
```

- Mở file environment:

```
sudo nano /etc/environment
```

- Thay thế dòng cũ bằng:

```
PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/usr/local/hadoop/bin:/usr/local/hadoop/sbin"JAVA_HOME="/usr/lib/jvm/java-8-openjdk-amd64/jre"
```



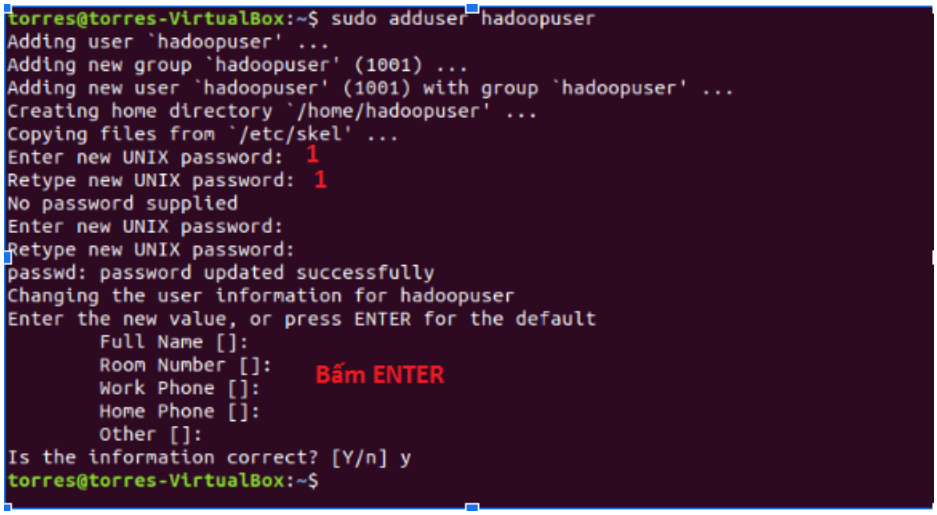
Lưu lại file

4th Step: Tạo 1 user hadoopuser mới trên các máy

- Gõ lệnh:

```
sudo adduser hadoopuser
```

```
//Điền password là 1 cho dễ :D, còn lại cứ bấm Enter
```



- Cấp quyền cho nó, bấm lần lượt các lệnh:

```
sudo usermod -aG hadoopuser hadoopuser
```

```
sudo chown hadoopuser:root -R /usr/local/hadoop/
```

```
sudo chmod g+rwx -R /usr/local/hadoop/
```

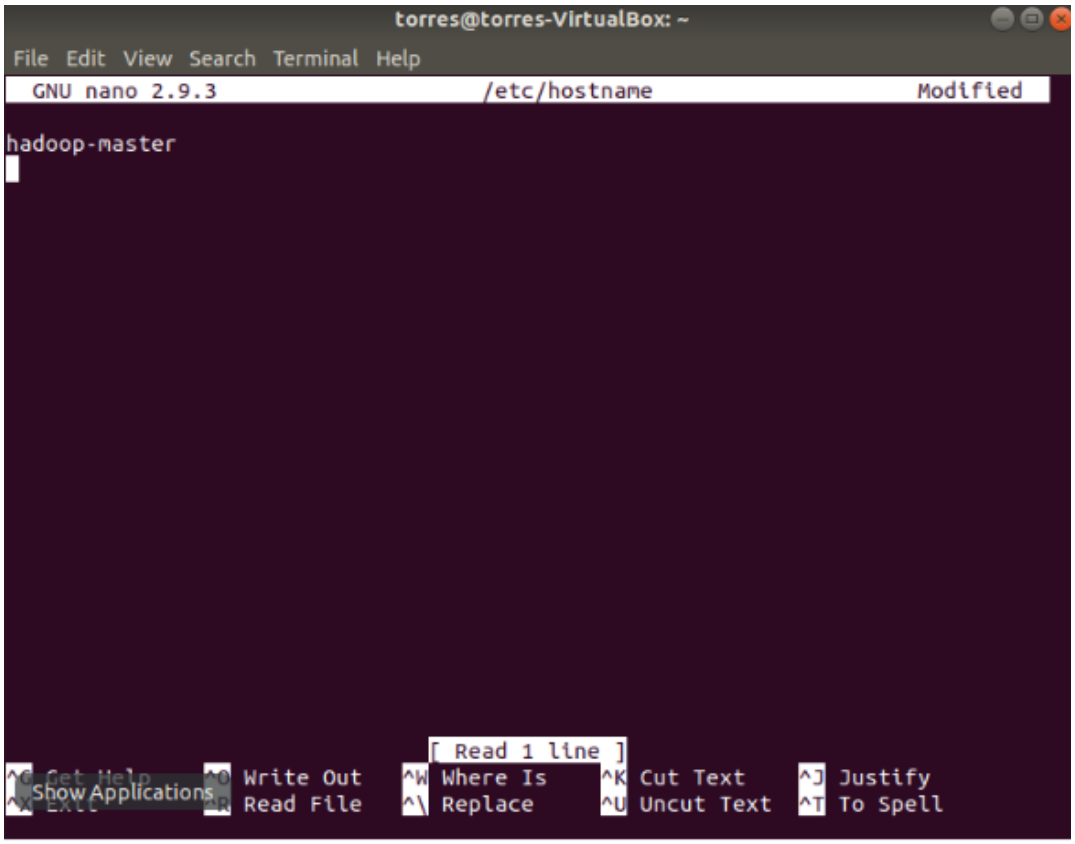
```
sudo adduser hadoopuser sudo
```

```
torres@torres-VirtualBox:~$ sudo usermod -aG hadoopuser hadoopuser
torres@torres-VirtualBox:~$ sudo chown hadoopuser:root -R /usr/local/hadoop
torres@torres-VirtualBox:~$ sudo chmod g+rwx -R /usr/local/hadoop
torres@torres-VirtualBox:~$ sudo adduser hadoopuser sudo
Adding user `hadoopuser' to group `sudo' ...
Adding user hadoopuser to group sudo
Done.
torres@torres-VirtualBox:~$
```

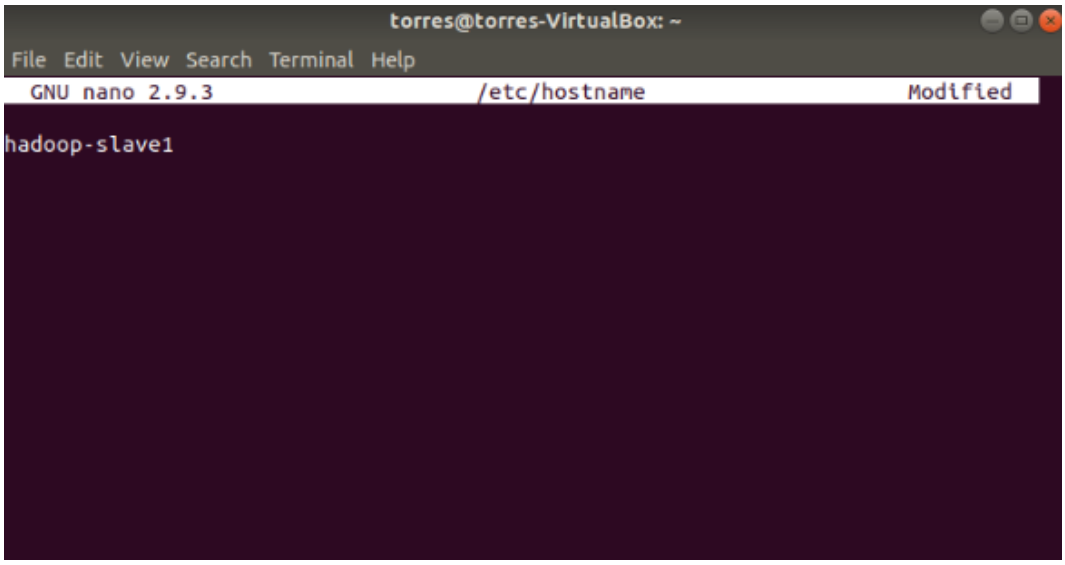
- Thay đổi tên máy:

`sudo nano /etc/hostname`

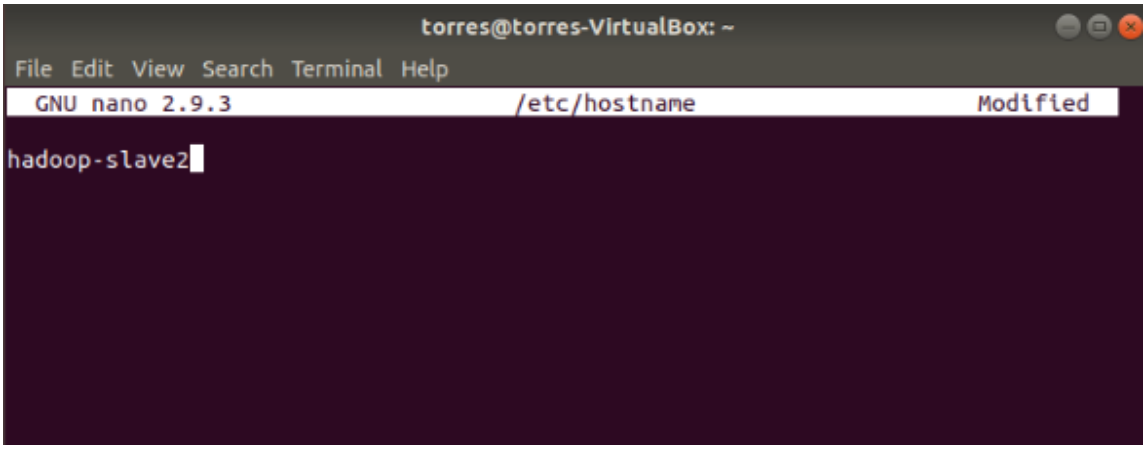
- `Chọn 1 máy là master, sửa file đó thành`



- Chọn 1 máy là slave1, sửa file đó thành



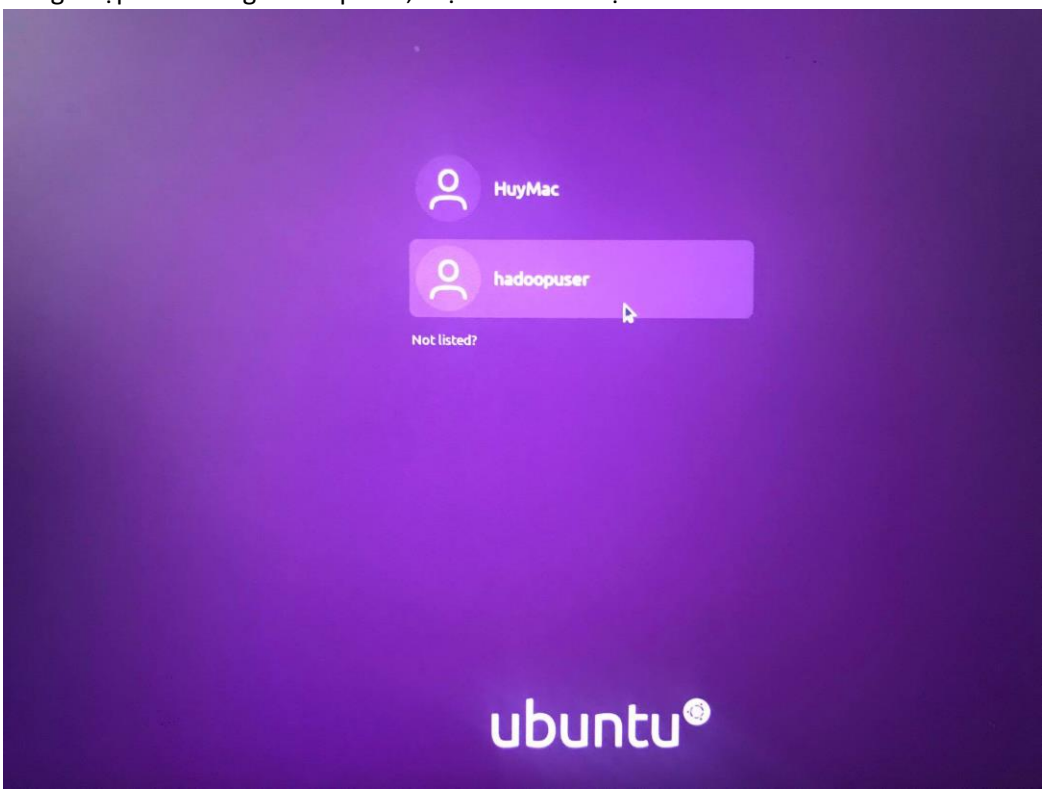
- Chọn 1 máy là slave2, sửa file đó thành



Lưu lại.

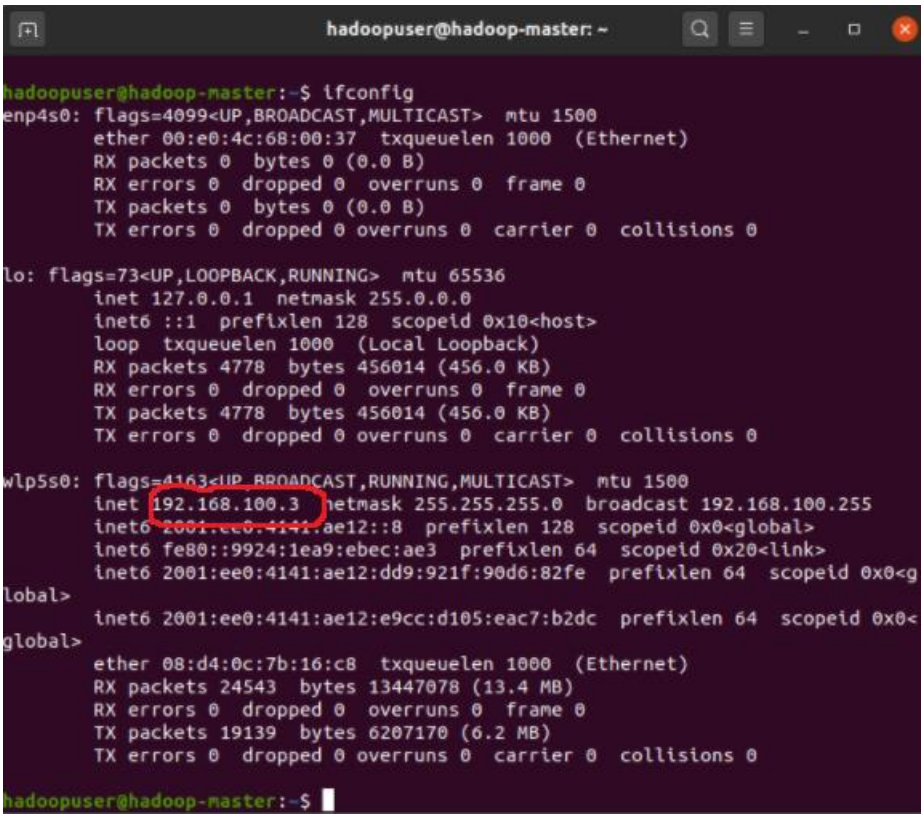
- Gõ lệnh sudo reboot để restart máy.

- Đăng nhập vào thằng hadoopuser, mật khẩu vừa đặt là 1



- Kiểm tra địa chỉ IP từng máy bằng lệnh

```
Ifconfig
```

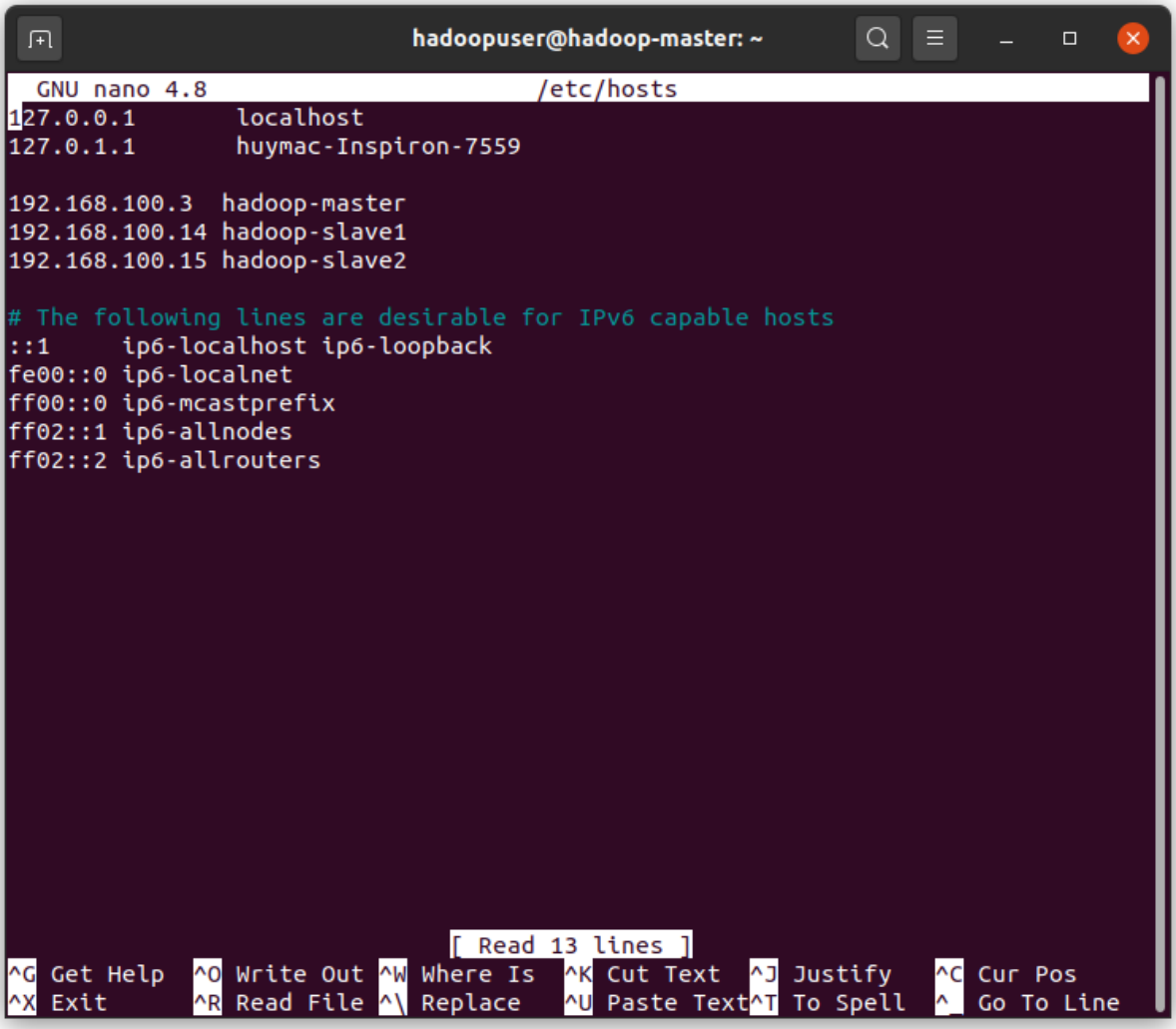


- Kiểm tra địa chỉ IP các máy hadoop-master, hadoop-slave1, hadoop-slave2. TH mình:

192.168.100.3 hadoop-master
192.168.100.14 hadoop-slave1
192.168.100.15 hadoop-slave2

- Mở hosts file trên tất cả máy, thêm 3 dòng trên

```
sudo nano /etc/hosts
```



Lưu lại file, nên thể `sudo reboot` cả 3 máy 1 lần nữa và đăng nhập lại vào hadoopuser

II. Trên máy hadoop-master

1st Step: Dùng SSH để kết nối từ hadoop-master tới hadoop-slave1, hadoop-slave2

- Sinh khóa rsa

```
ssh-keygen -t rsa
```

```
hadoopuser@hadoop-master:~$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hadoopuser/.ssh/id_rsa):
Created directory '/home/hadoopuser/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/hadoopuser/.ssh/id_rsa.
Your public key has been saved in /home/hadoopuser/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:wti4V4CRMVGcbBEIbwd/G0Rnpf/HoT9T1FE65iCZXS4 hadoopuser@hadoop-master
The key's randomart image is:
+---[RSA 2048]----+
|   ..*O+=o o.. . o|
|    .o*=. o = o o |
|     +.+ o = E = o|
|    . * o o o = .o|
|     o + S   . .o |
|      . o     ...o|
|     . .      .. +|
|      .        .+ |
|                .o|
+----[SHA256]-----+
hadoopuser@hadoop-master:~$
```

- Copy khóa tới các máy

`ssh-copy-id hadoopuser@hadoop-master`

```
hadoopuser@hadoop-master:~$ ssh-copy-id hadoopuser@hadoop-master
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/hadoopuser
/.ssh/id_rsa.pub"
The authenticity of host 'hadoop-master (192.168.205.7)' can't be established.
ECDSA key fingerprint is SHA256:Nsjcx3SrmwVnSNcWxvlYIajjHRdwaET+RGeLkVyoHI4.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter
 out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are promp
ted now it is to install the new keys
hadoopuser@hadoop-master's password:

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'hadoopuser@hadoop-master'"
and check to make sure that only the key(s) you wanted were added.

hadoopuser@hadoop-master:~$
```

`ssh-copy-id hadoopuser@hadoop-slave1`

```
hadoopuser@hadoop-master:~$ ssh-copy-id hadoopuser@hadoop-slave1
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/hadoopuser
/.ssh/id_rsa.pub"
The authenticity of host 'hadoop-slave1 (192.168.205.8)' can't be established.
ECDSA key fingerprint is SHA256:Nsjcx3SrmwVnSNcWxvlYIajjHRdwaET+RGeLkVyoHI4.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter
 out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are promp
ted now it is to install the new keys
hadoopuser@hadoop-slave1's password:

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'hadoopuser@hadoop-slave1'"
and check to make sure that only the key(s) you wanted were added.

hadoopuser@hadoop-master:~$
```

`ssh-copy-id hadoopuser@hadoop-slave2`

```
hadoopuser@hadoop-master:~$ ssh-copy-id hadoopuser@hadoop-slave2
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/hadoopuser
/.ssh/id_rsa.pub"
The authenticity of host 'hadoop-slave2 (192.168.205.9)' can't be established.
ECDSA key fingerprint is SHA256:Nsjcx3SrmwVnSNcWxvlYIajjHRdwaET+RGeLkVyoHI4.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter
 out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are promp
ted now it is to install the new keys
hadoopuser@hadoop-slave2's password:

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'hadoopuser@hadoop-slave2'"
and check to make sure that only the key(s) you wanted were added.

hadoopuser@hadoop-master:~$
```

2nd Step: Config các file

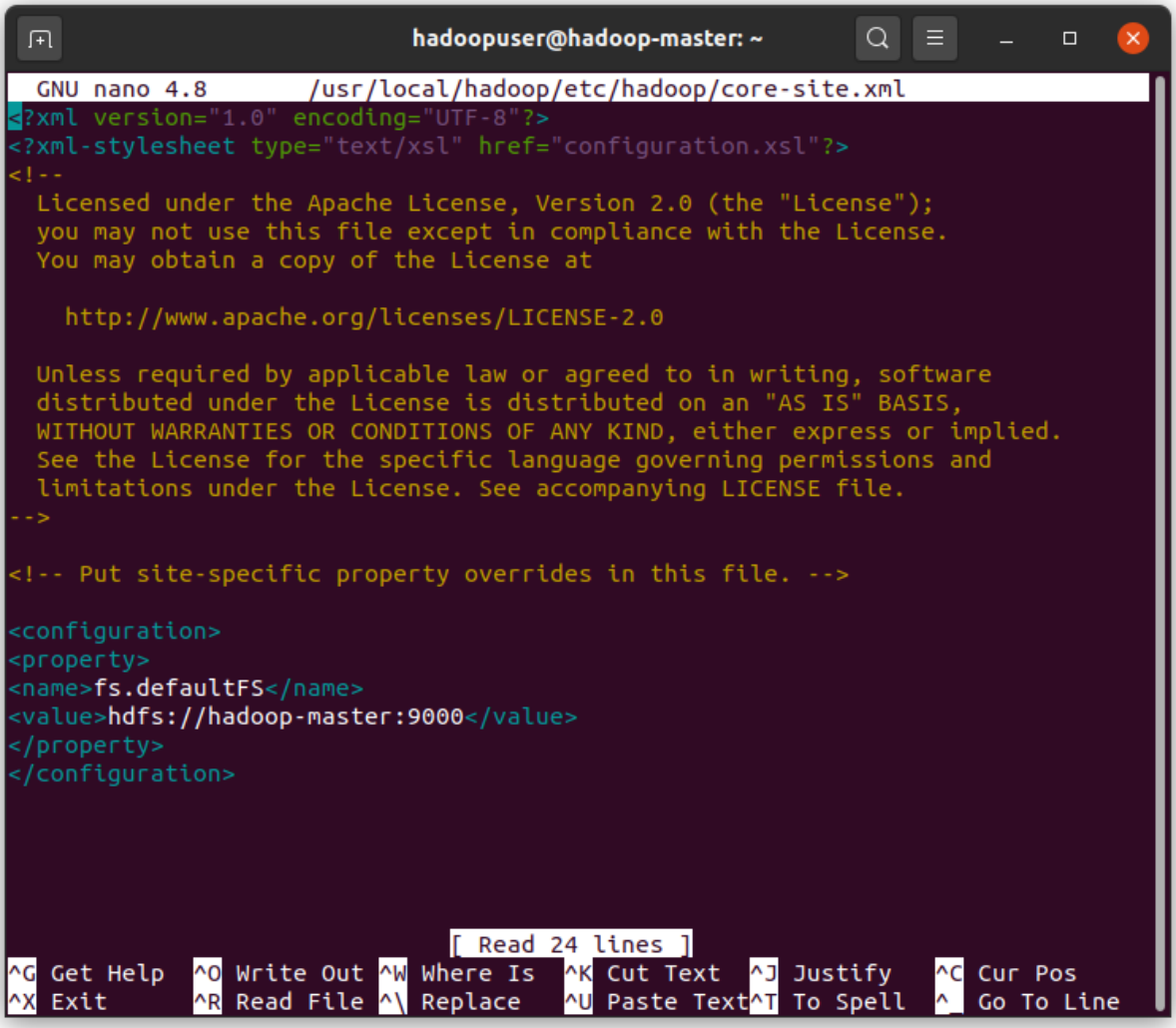
- Sửa file **core-site.xml:**

`sudo nano /usr/local/hadoop/etc/hadoop/core-site.xml`

```
hadoopuser@hadoop-master:~$ sudo nano /usr/local/hadoop/etc/hadoop/core-site.xm
l
[sudo] password for hadoopuser:
```

Thêm các dòng sau:

```
<configuration>
<property>
<name>fs.defaultFS</name>
<value>hdfs://hadoop-master:9000</value>
</property>
</configuration>
```



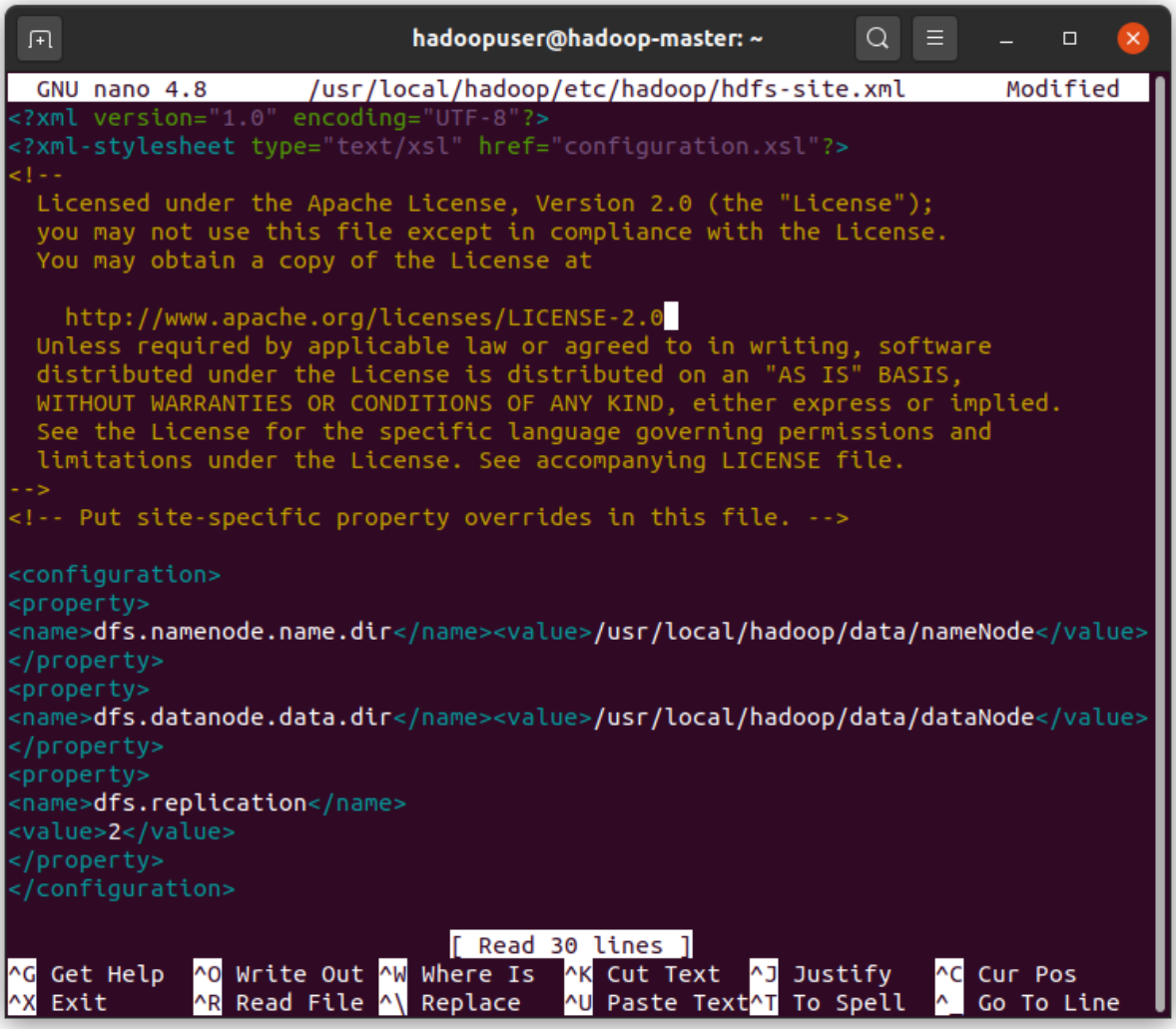
- Sửa file **hdfs-site.xml:**

```
sudo nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml
```

```
hadoopuser@hadoop-master:~$ sudo nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml
```

Thêm các dòng sau:

```
<configuration>

<property>

<name>dfs.namenode.name.dir</name><value>/usr/local/hadoop/data/nameNode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name><value>/usr/local/hadoop/data/dataNode</value>

</property>

<property>

<name>dfs.replication</name>

<value>2</value>

</property>

</configuration>
```



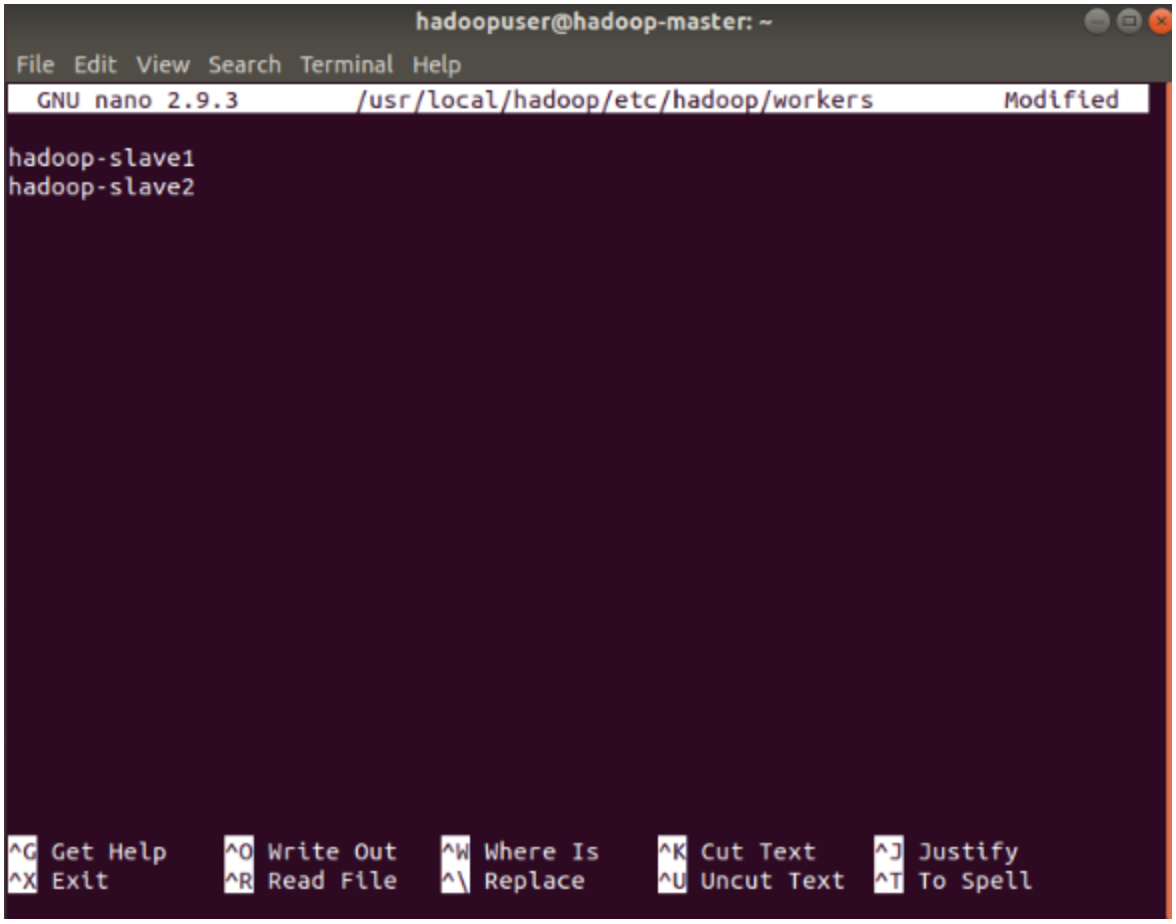
- Sửa file **workers:**

```
sudo nano /usr/local/hadoop/etc/hadoop/workers
```

```
hadoopuser@hadoop-master:~$ sudo nano /usr/local/hadoop/etc/hadoop/workers
```

Thêm các dòng sau:

```
hadoop-slave1
hadoop-slave2
```



3rd Step:

- Gửi các file đã config sang các máy slave

```
scp /usr/local/hadoop/etc/hadoop/* hadoop-slave1:/usr/local/hadoop/etc/hadoop/
```

```
hadoopuser@hadoop-master:~$ scp /usr/local/hadoop/etc/hadoop/* hadoop-slave1:/u
sr/local/hadoop/etc/hadoop
capacity-scheduler.xml                          100% 8260   293.2KB/s   00:00
configuration.xsl                               100% 1335   791.3KB/s   00:00
container-executor.cfg                          100% 1940     1.1MB/s   00:00
core-site.xml                                   100%  864   776.1KB/s   00:00
hadoop-env.cmd                                  100% 3999     1.9MB/s   00:00
hadoop-env.sh                                   100%   16KB   2.6MB/s   00:00
hadoop-metrics2.properties                      100% 3321     1.4MB/s   00:00
hadoop-policy.xml                               100%   11KB 293.3KB/s   00:00
hadoop-user-functions.sh.example                100% 3414     1.2MB/s   00:00
hdfs-site.xml                                   100% 1051   201.6KB/s   00:00
httpfs-env.sh                                   100% 1484    71.0KB/s   00:00
httpfs-log4j.properties                         100% 1657   975.4KB/s   00:00
httpfs-signature.secret                         100%   21    12.7KB/s   00:00
httpfs-site.xml                                 100%  620   351.4KB/s   00:00
kms-acls.xml                                    100% 3518     1.8MB/s   00:00
kms-env.sh                                      100% 1351   850.0KB/s   00:00
kms-log4j.properties                            100% 1860   990.0KB/s   00:00
kms-site.xml                                    100%  682   429.3KB/s   00:00
log4j.properties                                100%   13KB  11.7MB/s   00:00
mapred-env.cmd                                  100%  951   646.1KB/s   00:00
mapred-env.sh                                   100% 1764     1.2MB/s   00:00
mapred-queues.xml.template                      100% 4113     2.6MB/s   00:00
mapred-site.xml                                 100%  758   817.1KB/s   00:00
/usr/local/hadoop/etc/hadoop/shellprofile.d: not a regular file
ssl-client.xml.example                          100% 2316    57.9KB/s   00:00
```

scp /usr/local/hadoop/etc/hadoop/* hadoop-slave2:/usr/local/hadoop/etc/hadoop/

```
hadoopuser@hadoop-master:~$ scp /usr/local/hadoop/etc/hadoop/* hadoop-slave2:/u
sr/local/hadoop/etc/hadoop/
capacity-scheduler.xml                          100% 8260     1.7MB/s   00:00
configuration.xsl                               100% 1335   128.3KB/s   00:00
container-executor.cfg                          100% 1940     1.9MB/s   00:00
core-site.xml                                   100%  864   737.1KB/s   00:00
hadoop-env.cmd                                  100% 3999     3.2MB/s   00:00
hadoop-env.sh                                   100%   16KB   1.3MB/s   00:00
hadoop-metrics2.properties                      100% 3321     2.9MB/s   00:00
hadoop-policy.xml                               100%   11KB   1.4MB/s   00:00
hadoop-user-functions.sh.example                100% 3414     2.7MB/s   00:00
hdfs-site.xml                                   100% 1051   732.5KB/s   00:00
httpfs-env.sh                                   100% 1484     1.1MB/s   00:00
httpfs-log4j.properties                         100% 1657     1.5MB/s   00:00
httpfs-signature.secret                         100%   21    15.4KB/s   00:00
httpfs-site.xml                                 100%  620   529.2KB/s   00:00
kms-acls.xml                                    100% 3518     2.7MB/s   00:00
kms-env.sh                                      100% 1351   987.0KB/s   00:00
kms-log4j.properties                            100% 1860     1.8MB/s   00:00
kms-site.xml                                    100%  682   302.6KB/s   00:00
log4j.properties                                100%   13KB  11.4MB/s   00:00
mapred-env.cmd                                  100%  951   902.9KB/s   00:00
mapred-env.sh                                   100% 1764     1.6MB/s   00:00
mapred-queues.xml.template                      100% 4113     3.8MB/s   00:00
mapred-site.xml                                 100%  758   799.8KB/s   00:00
/usr/local/hadoop/etc/hadoop/shellprofile.d: not a regular file
ssl-client.xml.example                          100% 2316   228.1KB/s   00:00
```

4th Step: Format HDFS file system

- Format HDFS file system:

```
source /etc/environment
hdfs namenode –format
```

```
hadoopuser@hadoop-master:~$ source /etc/environment
hadoopuser@hadoop-master:~$ hdfs namenode -format
```

5th Step: Start HDFS

- Start HDFS

```
start-dfs.sh
```

```
hadoopuser@hadoop-master:~$ start-dfs.sh
Starting namenodes on [hadoop-master]
Starting datanodes
hadoop-slave1: WARNING: /usr/local/hadoop/logs does not exist. Creating.
hadoop-slave2: WARNING: /usr/local/hadoop/logs does not exist. Creating.
Starting secondary namenodes [hadoop-master]
```

- Trên máy master, kiểm tra:

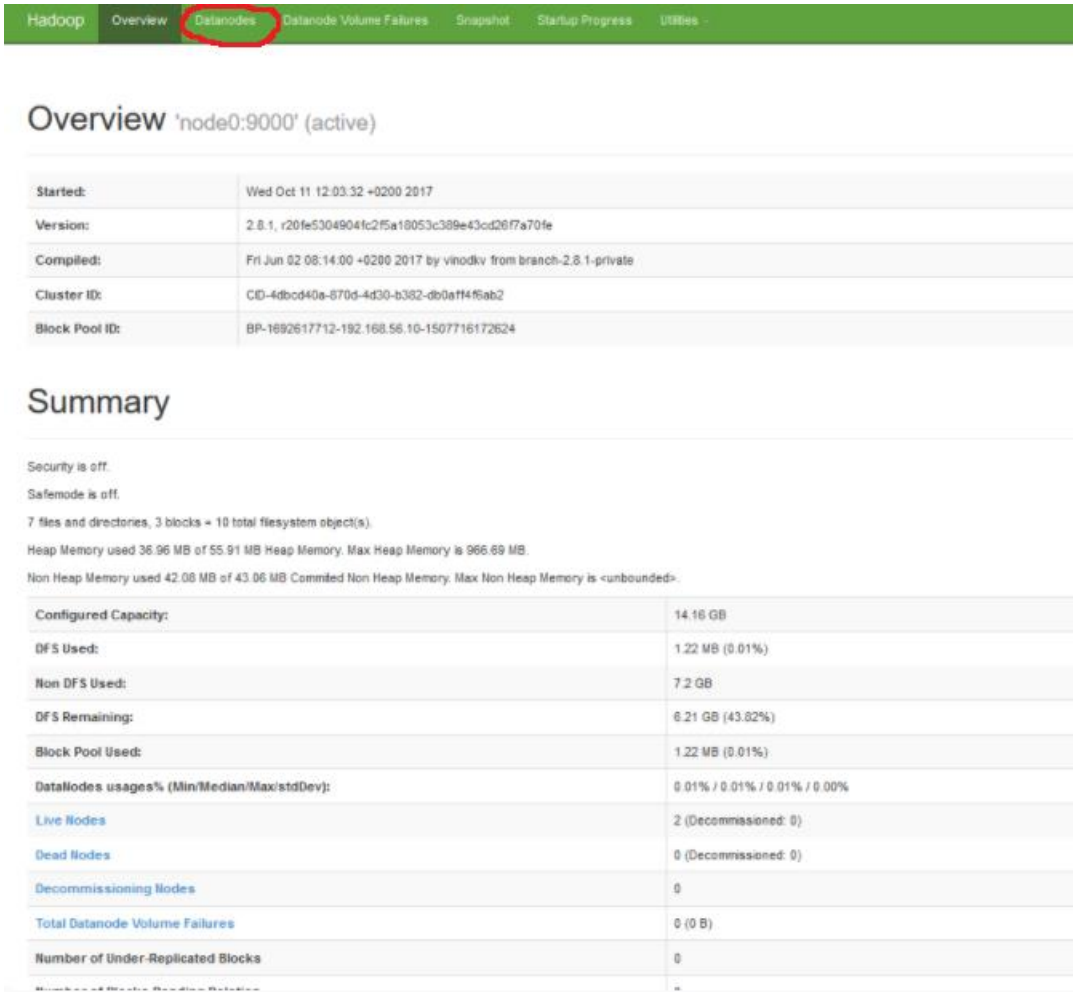
```
hadoopuser@hadoop-master:~$ jps
4138 Jps
3771 NameNode
4014 SecondaryNameNode
hadoopuser@hadoop-master:~$
```

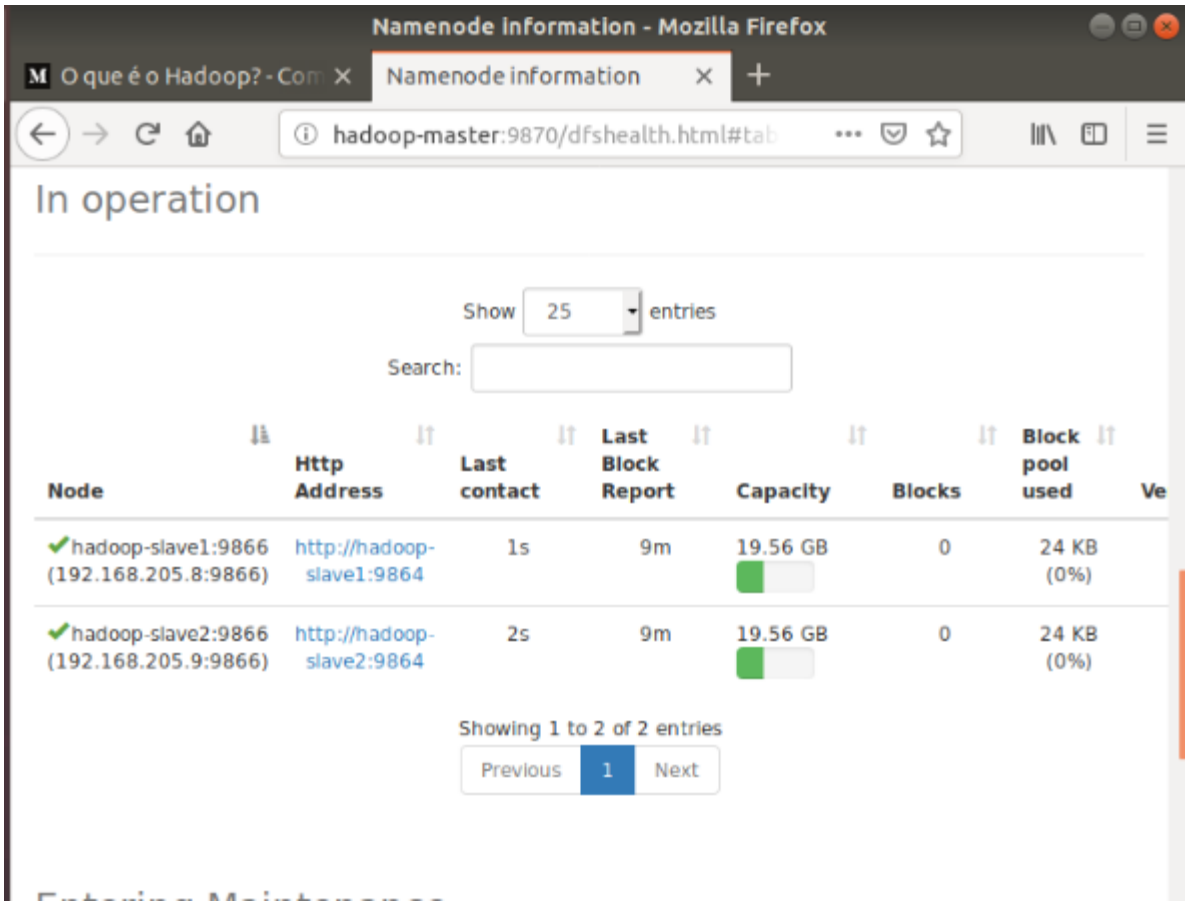
- Trên máy slave1 và slave2

```
hadoopuser@hadoop-slave1:~$ jps
1808 DataNode
2024 Jps
hadoopuser@hadoop-slave1:~$
```

```
hadoopuser@hadoop-slave2:~$ jps
1814 DataNode
2031 Jps
hadoopuser@hadoop-slave2:~$
```

- Check trên web hadoop-master:9870





III. Đẩy dữ liệu lên cụm HDFS

Làm trên máy master

Cách 1: Bằng câu lệnh

- Tạo home directory. Tên /data

```
hdfs dfs -mkdir -p /data
```

- Tạo thư mục books

```
hdfs dfs -mkdir books
```

- Tải 1 số file text trên mạng về máy để đẩy lên

```
cd /home/hadoopuser
```

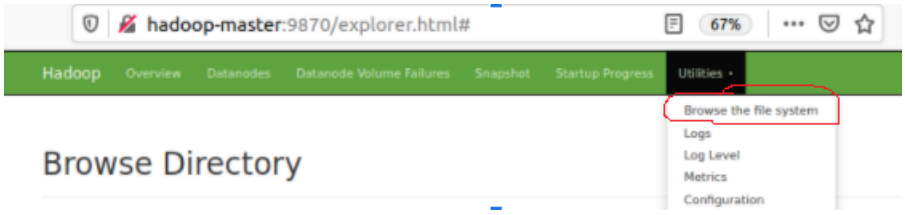
```
wget -O alice.txt
https://www.gutenberg.org/files/11/11-0.txt
```

```
wget -O holmes.txt
https://www.gutenberg.org/files/1661/1661-0.txt

wget -O frankenstein.txt
https://www.gutenberg.org/files/84/84-0.txt
```

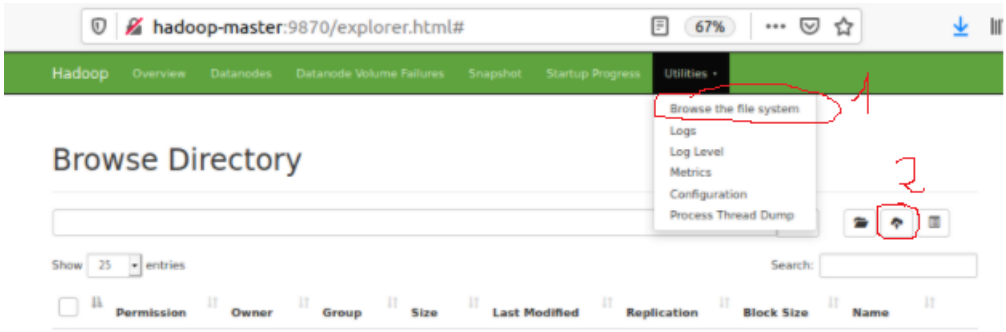
- Up lên cụm HDFS

```
hdfs dfs -put alice.txt holmes.txt frankenstein.txt
books
```

- Check trên web



Cách 2: Up bằng web



Lab02: Set Up a MapReduce in Hadoop 3.2.1 Multi-Node Cluster on Ubuntu (2 Nodes)

I. Trên 3 máy master, slave1, slave2

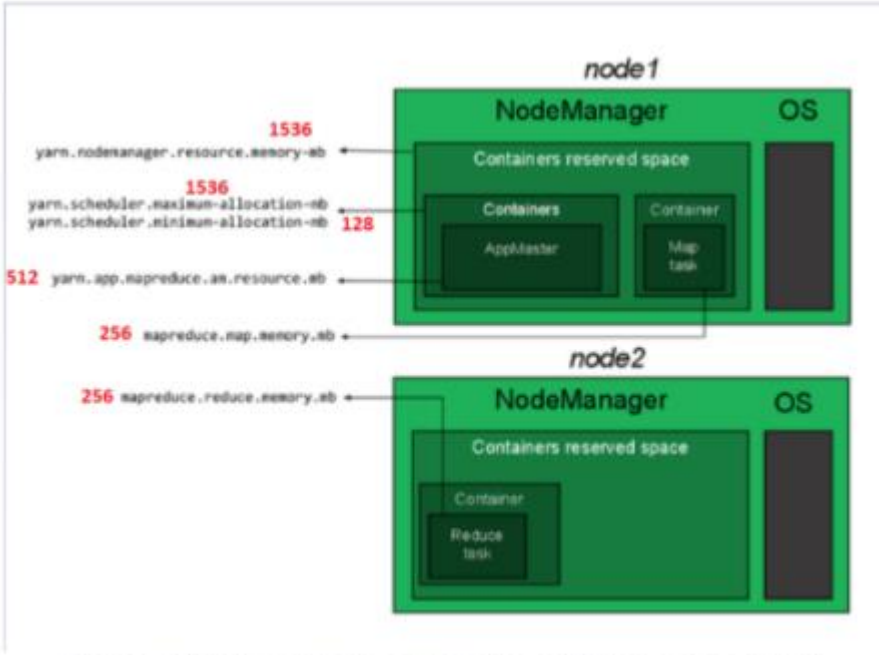


Figure 8: Sample config for 2GB Nodes (will change in next lab)

1st Step: **Configure yarn**

```
export HADOOP_HOME="/usr/local/hadoop"
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
```

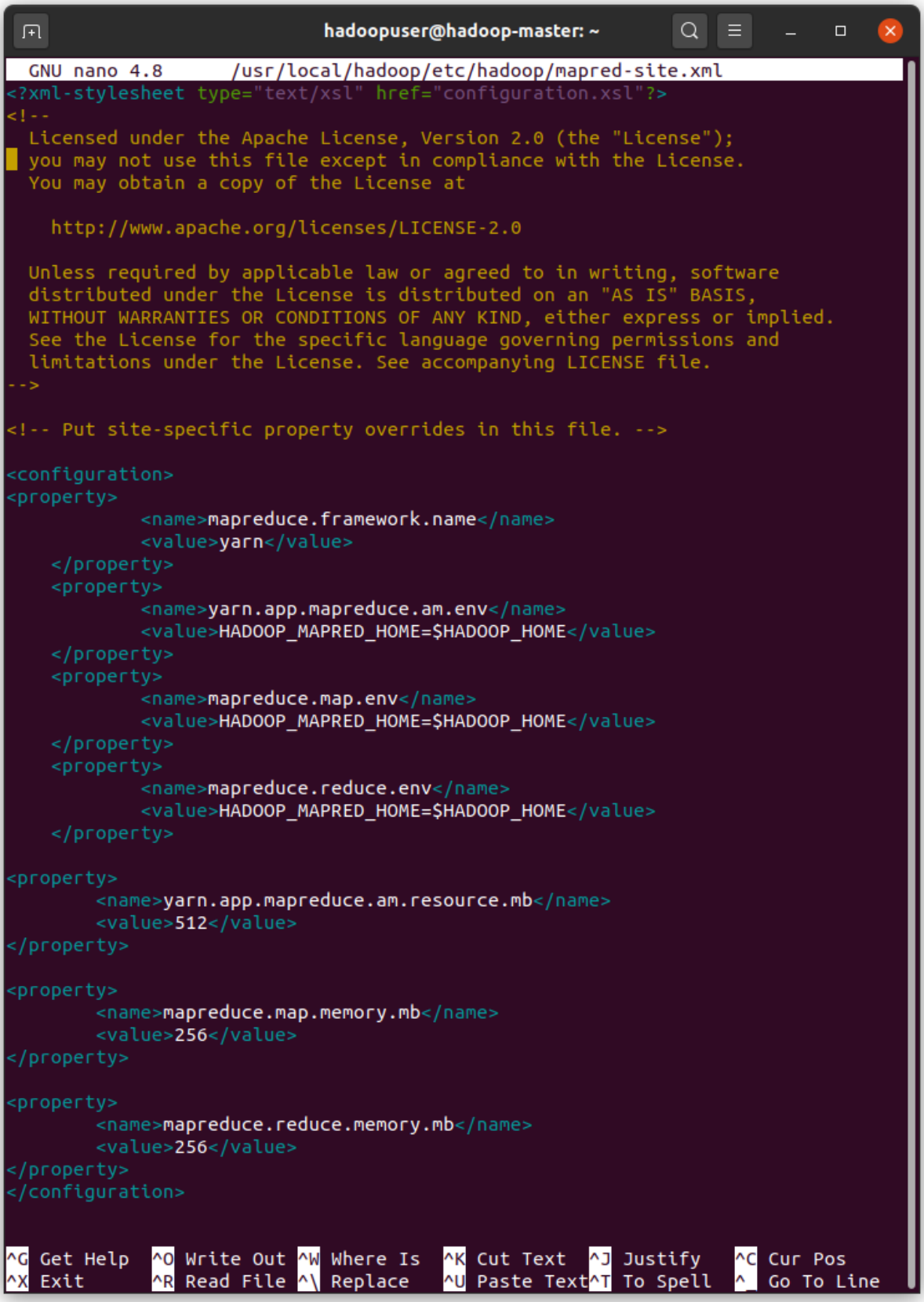
```
hadoopuser@hadoop-master:~$ export HADOOP_HOME="/usr/local/hadoop"
hadoopuser@hadoop-master:~$ export HADOOP_COMMON_HOME=$HADOOP_HOME
hadoopuser@hadoop-master:~$ export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
hadoopuser@hadoop-master:~$ export HADOOP_HDFS_HOME=$HADOOP_HOME
hadoopuser@hadoop-master:~$ export HADOOP_MAPRED_HOME=$HADOOP_HOME
hadoopuser@hadoop-master:~$ export HADOOP_YARN_HOME=$HADOOP_HOME
hadoopuser@hadoop-master:~$
```

2nd Step: Configure mapred-site.xml

```
sudo nano /usr/local/hadoop/etc/hadoop/mapred-site.xml
```

```
<configuration>
<property>
            <name>mapreduce.framework.name</name>
            <value>yarn</value>
  </property>
  <property>
            <name>yarn.app.mapreduce.am.env</name>
  <value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
  </property>
  <property>
            <name>mapreduce.map.env</name>
  <value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
  </property>
  <property>
            <name>mapreduce.reduce.env</name>
  <value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
```

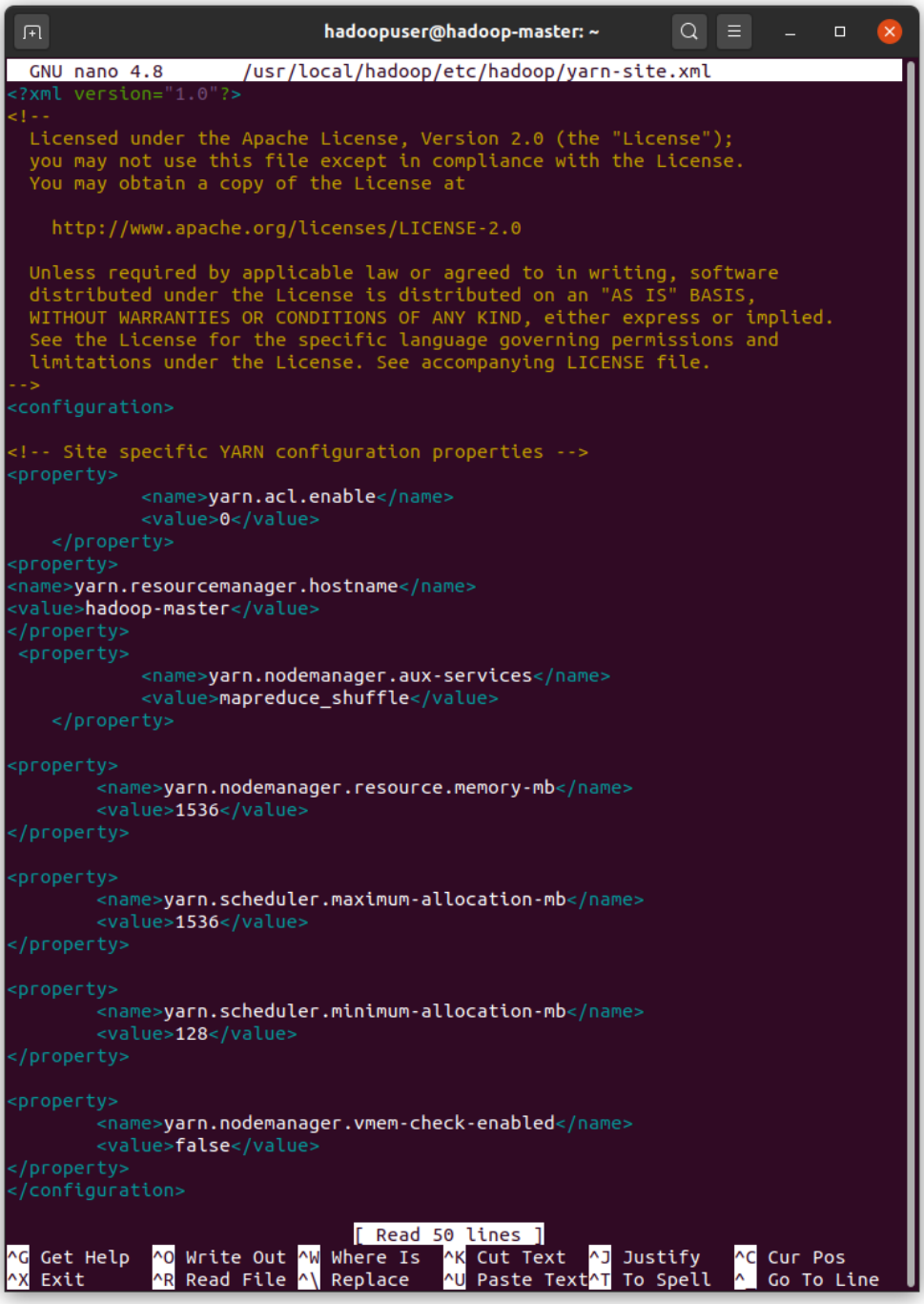
```
  </property>

<property>
        <name>yarn.app.mapreduce.am.resource.mb</name>
        <value>512</value>
</property>

<property>
        <name>mapreduce.map.memory.mb</name>
        <value>256</value>
</property>

<property>
        <name>mapreduce.reduce.memory.mb</name>
        <value>256</value>
</property>
</configuration>
```


hadoopuser@hadoop-master: ~
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/mapred-site.xml
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
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>mapreduce.framework.name</name>
<value>yarn</value>
</property>
<property>
<name>yarn.app.mapreduce.am.env</name>
<value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
</property>
<property>
<name>mapreduce.map.env</name>
<value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
</property>
<property>
<name>mapreduce.reduce.env</name>
<value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
</property>
<property>
<name>yarn.app.mapreduce.am.resource.mb</name>
<value>512</value>
</property>
<property>
<name>mapreduce.map.memory.mb</name>
<value>256</value>
</property>
<property>
<name>mapreduce.reduce.memory.mb</name>
<value>256</value>
</property>
</configuration>
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line

3rd Step: Config yarn-site.xml

```
sudo nano /usr/local/hadoop/etc/hadoop/yarn-site.xml

<configuration>

<!-- Site specific YARN configuration properties -->
<property>
            <name>yarn.acl.enable</name>
            <value>0</value>
  </property>
<property>
<name>yarn.resourcemanager.hostname</name>
<value>hadoop-master</value>
</property>
 <property>
            <name>yarn.nodemanager.aux-services</name>
            <value>mapreduce_shuffle</value>
  </property>

<property>
       <name>yarn.nodemanager.resource.memory-mb</name>
       <value>1536</value>
</property>

<property>
```

```
        <name>yarn.scheduler.maximum-allocation-
mb</name>
        <value>1536</value>
</property>

<property>
        <name>yarn.scheduler.minimum-allocation-
mb</name>
        <value>128</value>
</property>

<property>
        <name>yarn.nodemanager.vmem-check-enabled</name>
        <value>false</value>
</property>
</configuration>
```



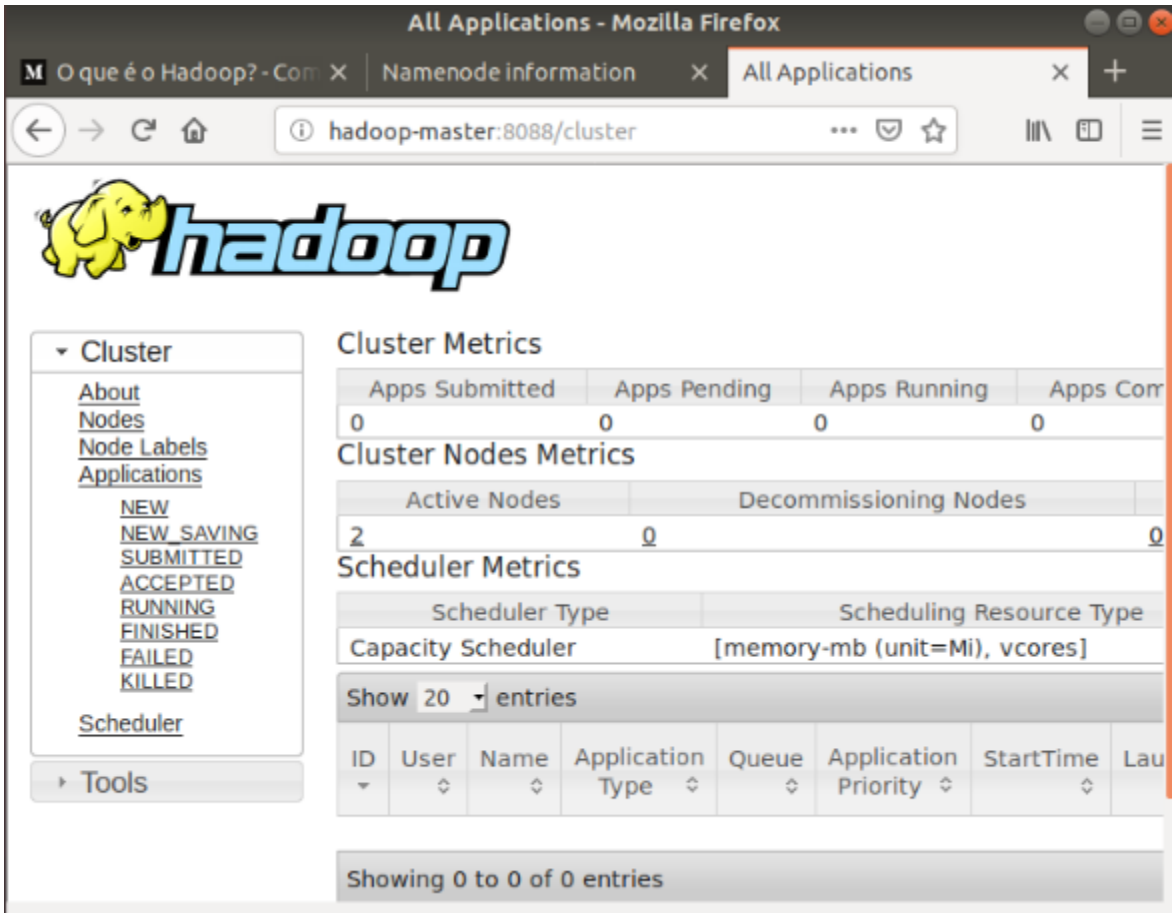
II. Trên máy master

1st Step: Start yarn

`start-yarn.sh`

```
hadoopuser@hadoop-master:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
```

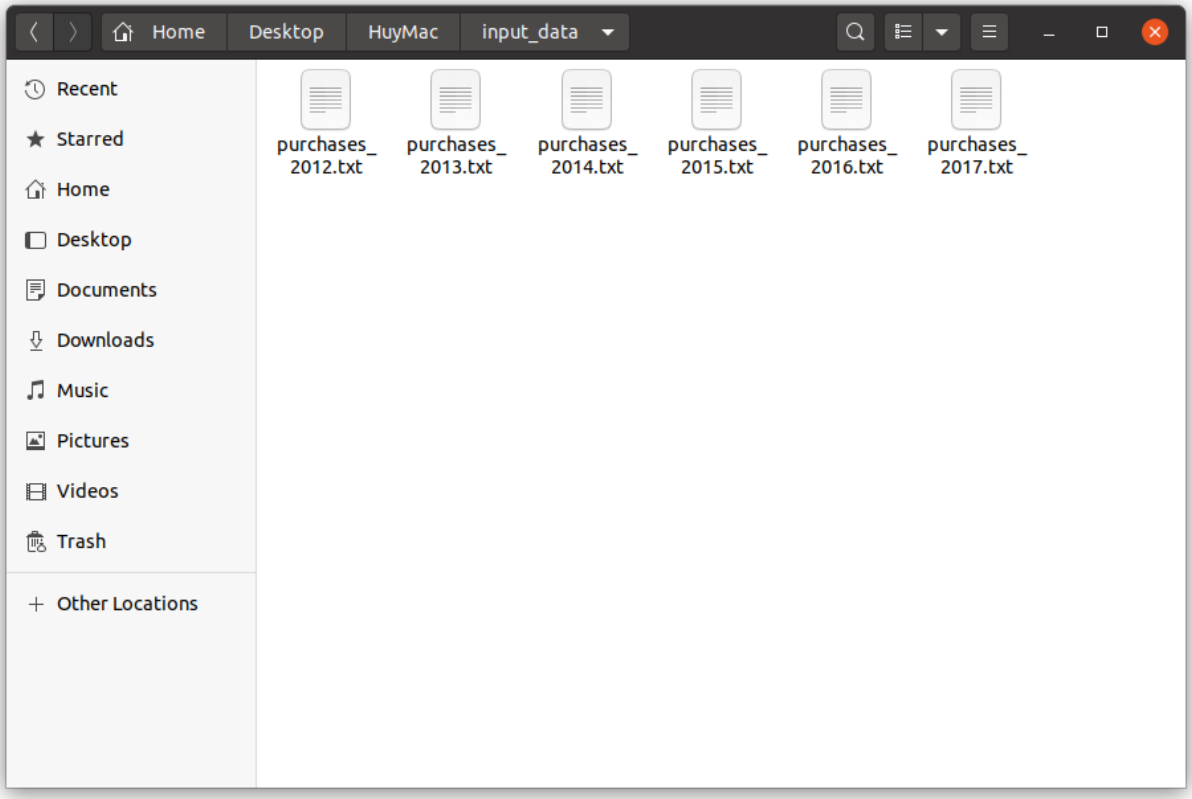
Lên web, truy cập hadoop-master:8088/cluster



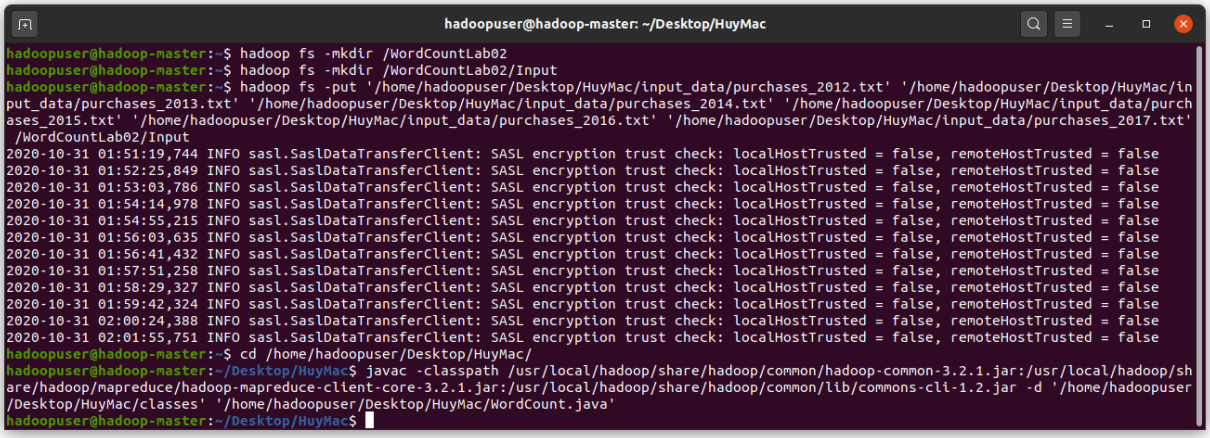
2nd Step: Tạo MapReduce Job

- Chuẩn bị data: Tài trên https://github.com/MacHuy/HDFS-MultiNode/blob/main/HuyMac/input_data/test.txt

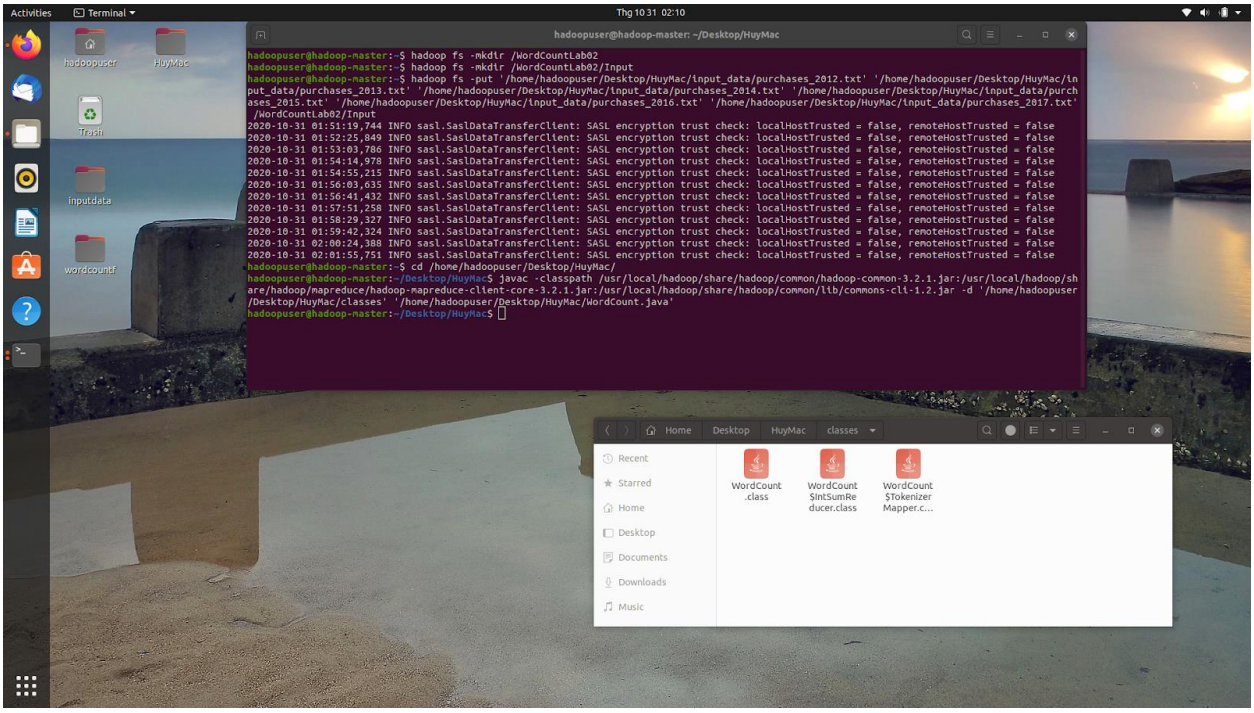
- Hoặc có thể toàn bộ folder HuyMac và move nó để ở Desktop:
  https://github.com/MacHuy/HDFS-MultiNode



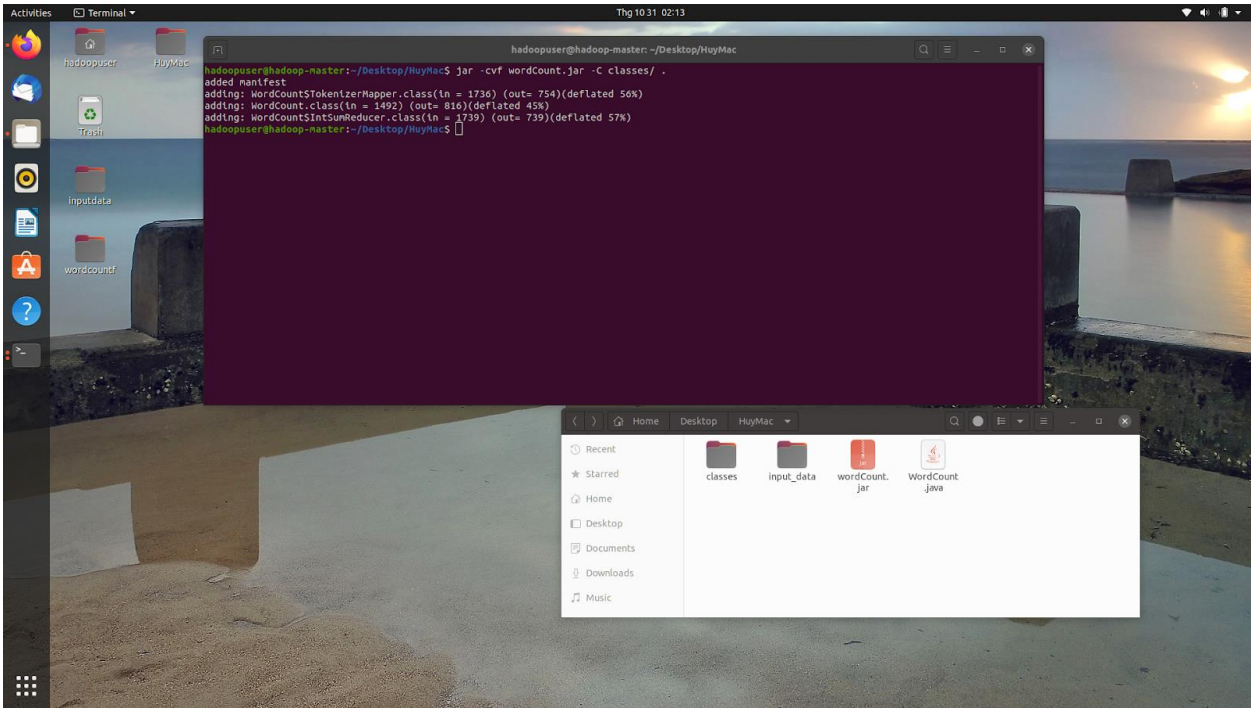
-
- Updata lên HDFS folder /WordCountLab02/Input
  Thực thi file wordcount.java



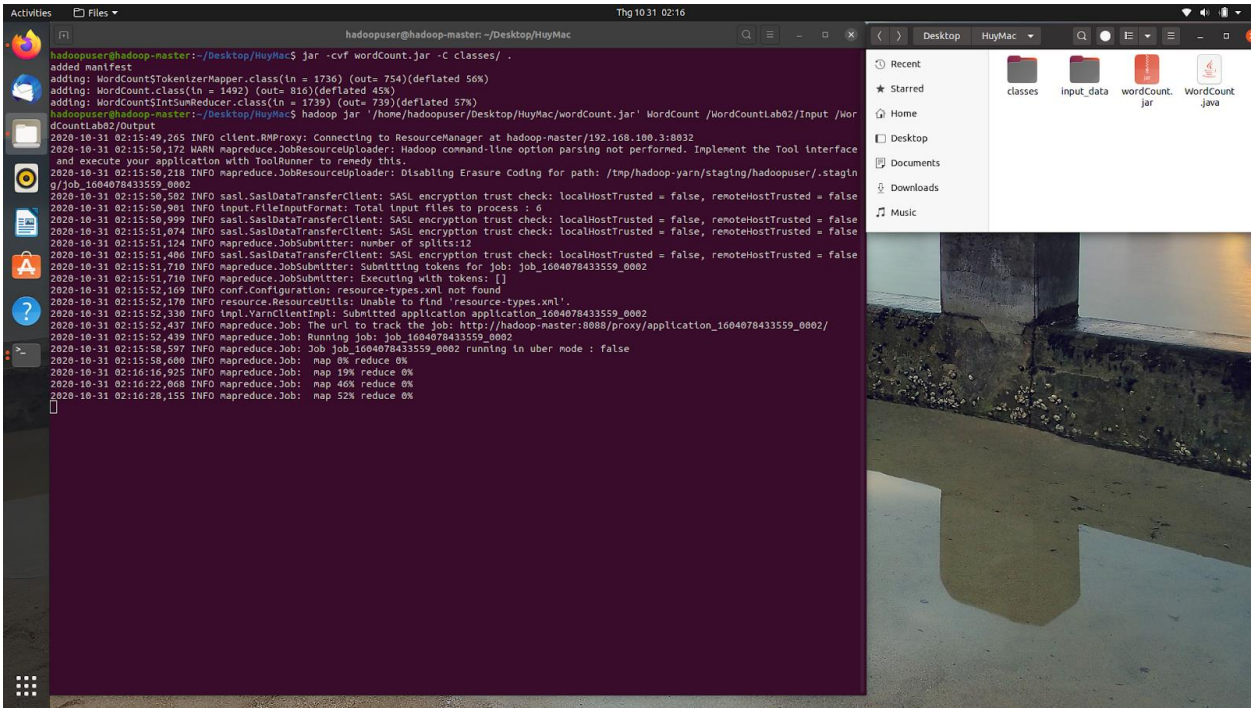
- Xuất hiện 3 file class trong folder /classes



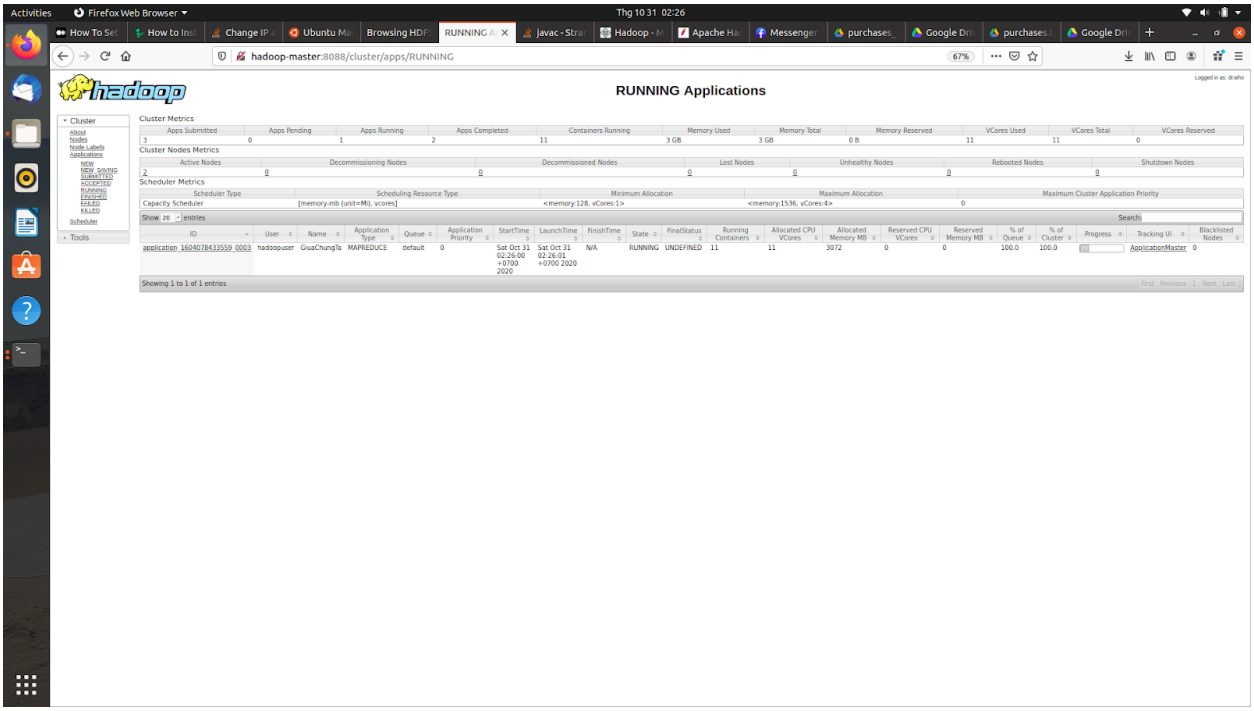
- Tạo file jar



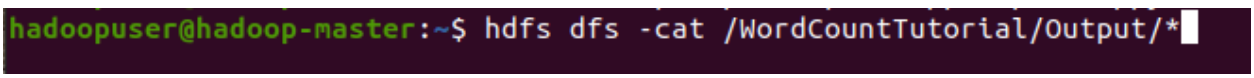
- Sử dụng file jar thực thi chạy job mapreduce

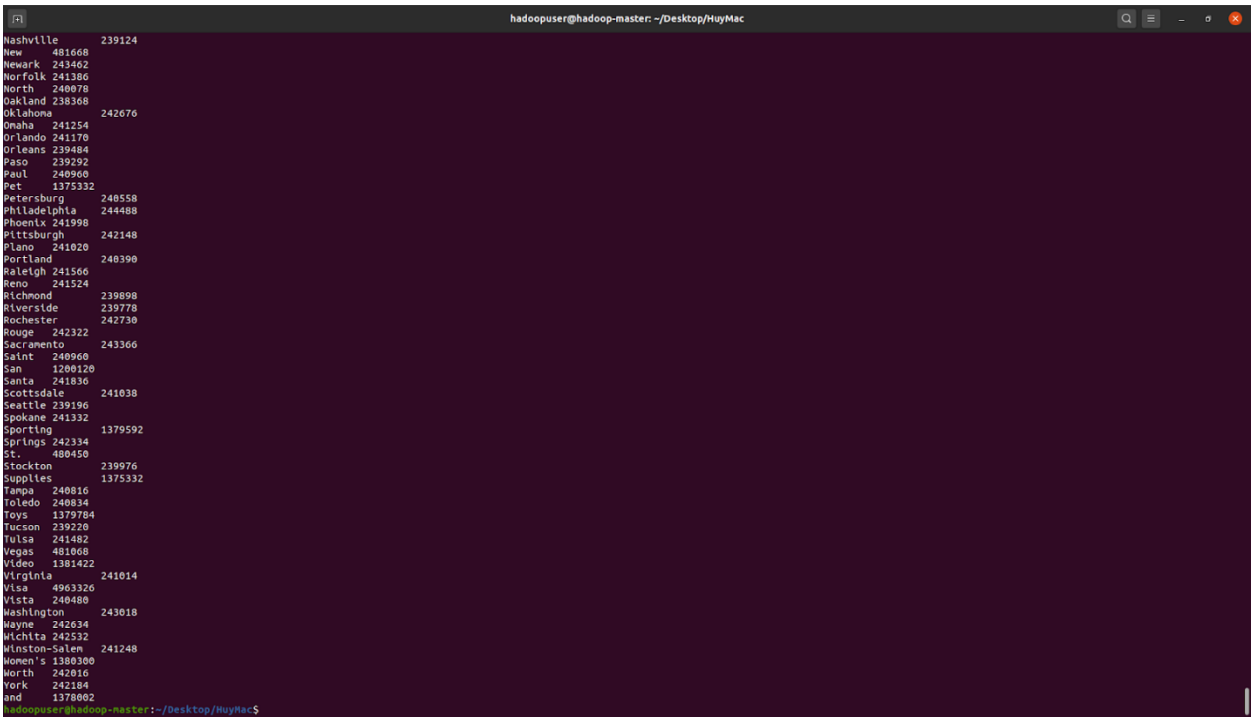


- Lên web, check job đang chạy

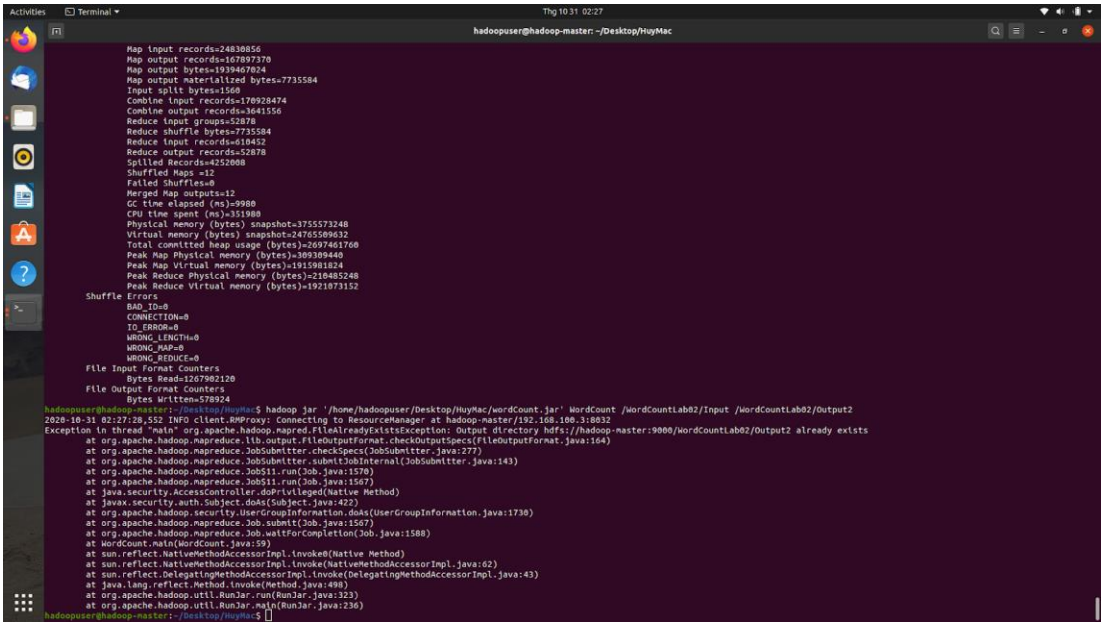


- Xem kết quả

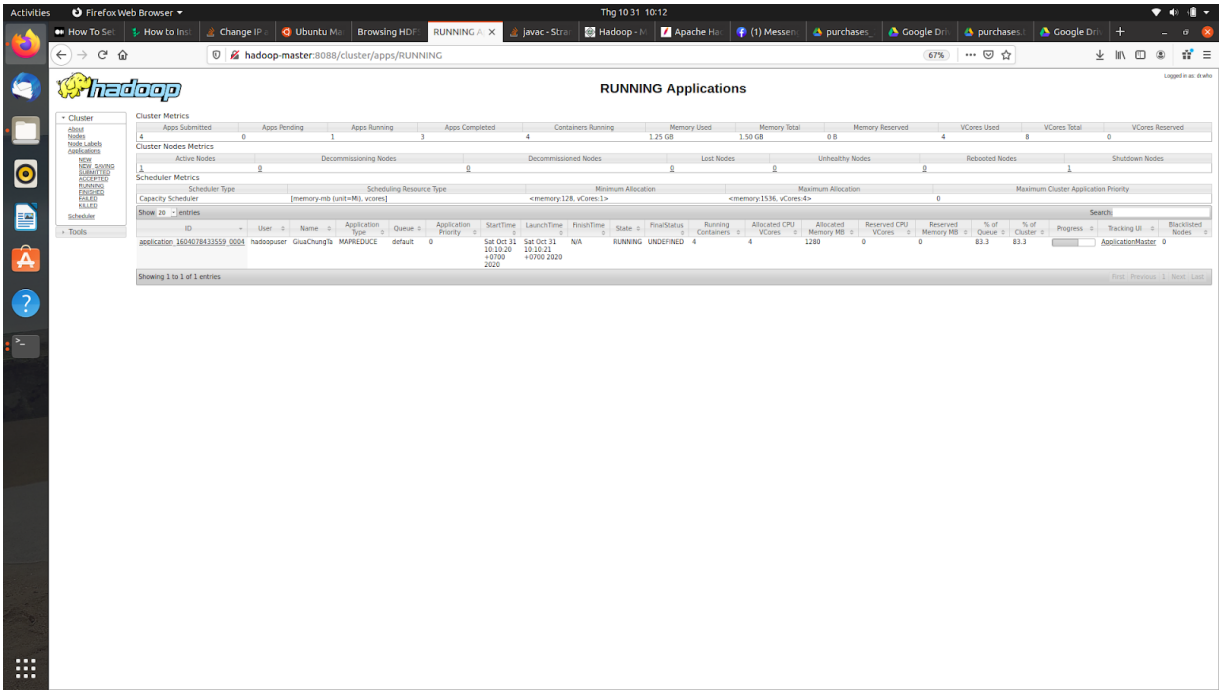




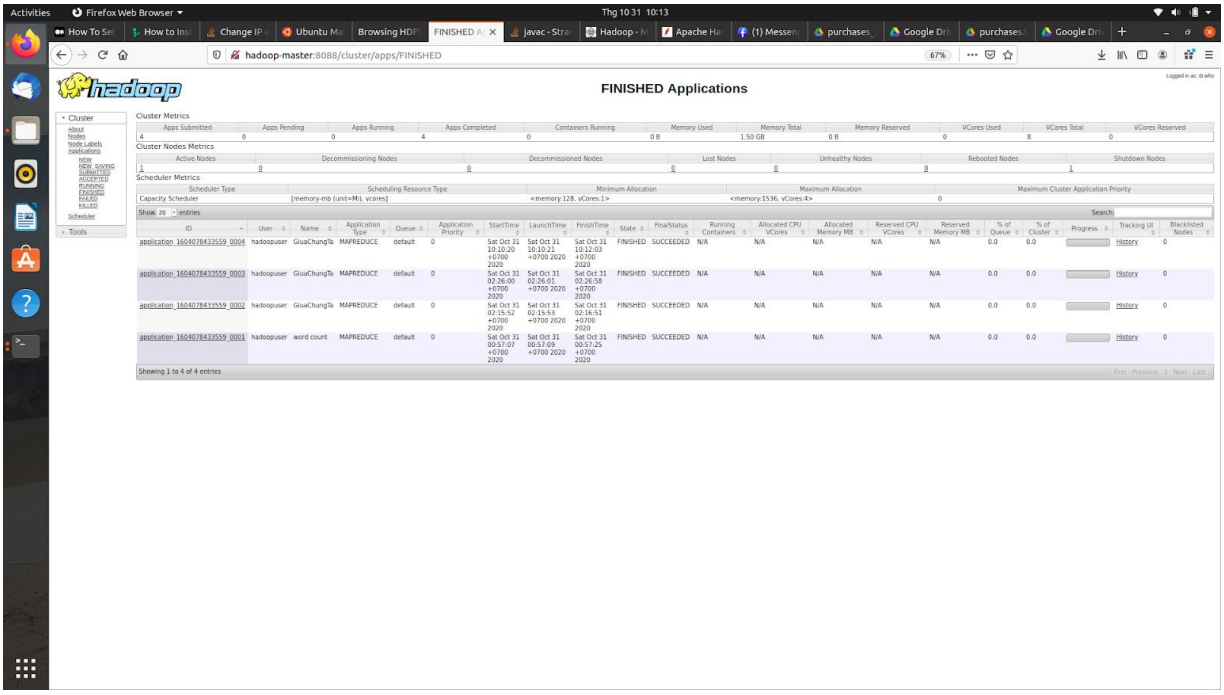
TH chạy lại job, nhớ đổi sang file output mới, folder output cũ đã có, HDFS ko cho phép overwrite



Trường hợp chỉ có 1 datanode slave1



- Check các job finished



Toàn bộ về Hadoop 3.2.1 Multi-Node Cluster and Mapreduce Job:

https://github.com/MacHuy/HDFS-MultiNode (pdf file)