

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



REPORT LAB 1
CÀI ĐẶT HADOOP TRÊN NỀN TẢNG AWS

MÔN HỌC
NHẬP MÔN DỮ LIỆU LỚN

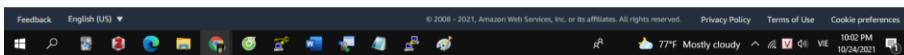
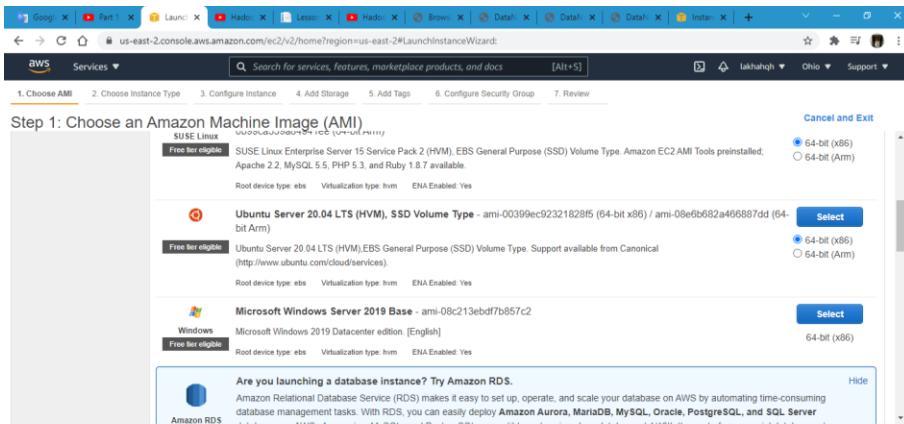
Họ và tên: Trần Ngọc Lan Khanh - MSSV: 18120408
GVTH: ThS. Lê Ngọc Thành

Học kỳ 1 / 2021-2022

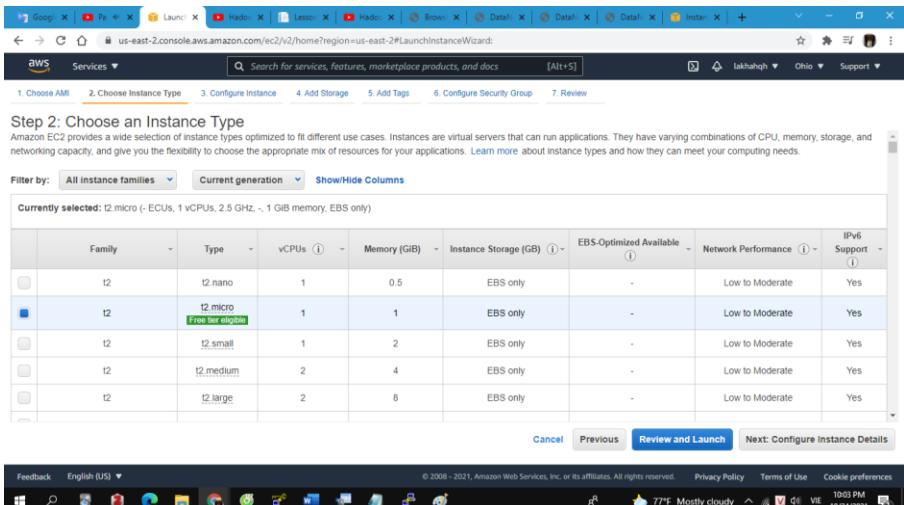
Thành Phố Hồ Chí Minh – Năm 2021

1. Launch Instances:

Launch Ubuntu Server 20.04 LTS



Chọn Instance Type



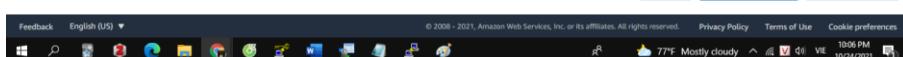
Tạo 4 Instances

Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the same AMI, request Spot Instances to take advantage of the lower pricing, assign an access management role to the instance, and more.

Number of Instances	4	Launch into Auto Scaling Group
<small>Must be a positive integer</small>		
Purchasing option	<input type="checkbox"/> Request Spot Instances	
Network	vpc-802e5beb (default)	<input type="checkbox"/> Create new VPC
Subnet	No preference (default subnet in any Availability Zone)	<input type="checkbox"/> Create new subnet
Auto-assign Public IP	<input type="checkbox"/> Use subnet setting (Enable)	
Placement group	<input type="checkbox"/> Add instance to placement group	
Capacity Reservation	Open	
Domain join directory	<input type="checkbox"/> Create new directory	
IAM role	<input type="checkbox"/> Create new IAM role	

Cancel Previous Review and Launch Next: Add Storage



Step 4: Add Storage

Your instance will be launched with the following storage device settings. You can attach additional EBS volumes and instance store volumes to your instance, or edit the settings of the root volume. You can also attach additional EBS volumes after launching an instance, but not instance store volumes. [Learn more](#) about storage options in Amazon EC2.

Volume Type	Device	Snapshot	Size (GiB)	Volume Type	IOPS	Throughput (MB/s)	Delete on Termination	Encryption
Root	/dev/sda1	snap-04e912a474a57b607	8	General Purpose SSD (gp2)	100 / 3000	N/A	<input checked="" type="checkbox"/>	<input type="checkbox"/> Not Encrypted

Add New Volume

Free tier eligible customers can get up to 30 GB of EBS General Purpose (SSD) or Magnetic storage. [Learn more](#) about free usage tier eligibility and usage restrictions.

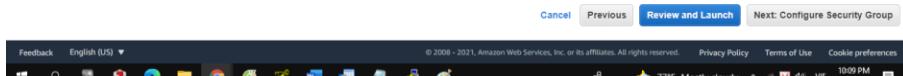


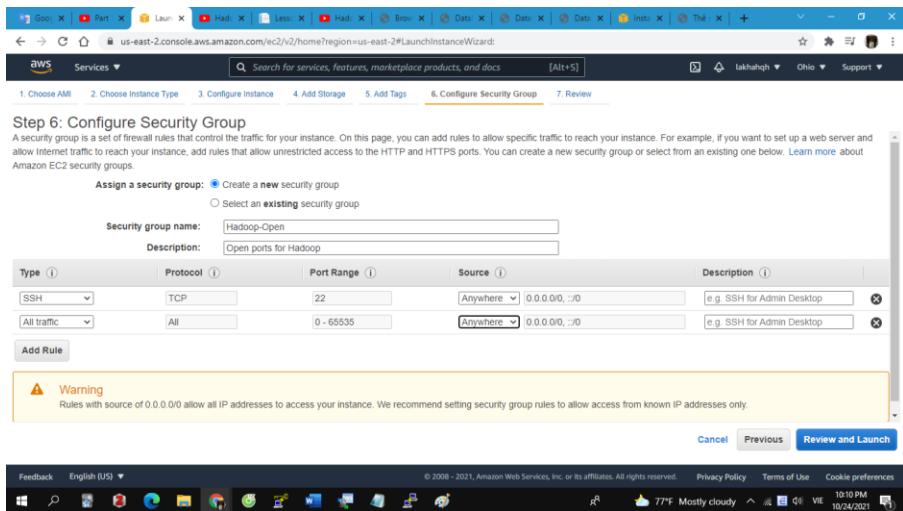
Step 5: Add Tags

A tag consists of a case-sensitive key-value pair. For example, you could define a tag with key = Name and value = Webserver. A copy of a tag can be applied to volumes, instances or both. Tags will be applied to all instances and volumes. [Learn more](#) about tagging your Amazon EC2 resources.

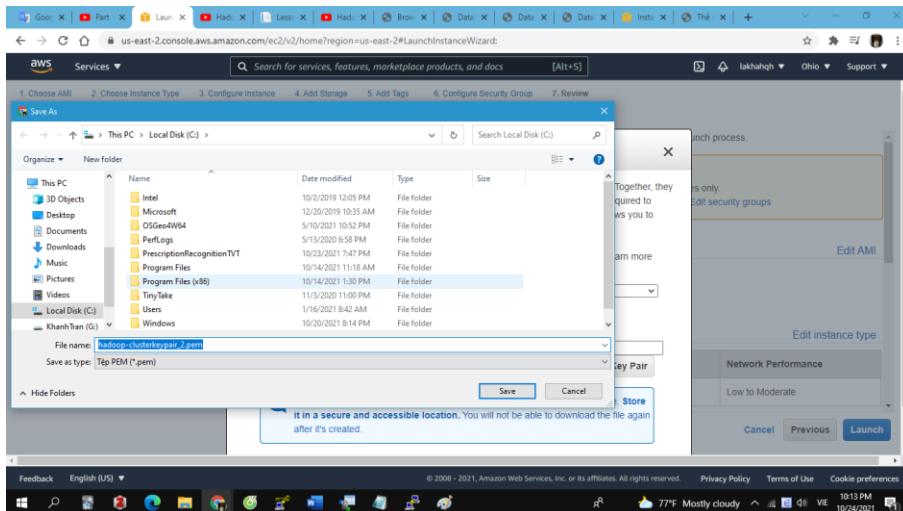
Key	(128 characters maximum)	Value	(256 characters maximum)	Instances	Volumes	Network Interfaces
Name	<input type="text"/> DataNode001					

Add another tag (Up to 50 tags maximum)

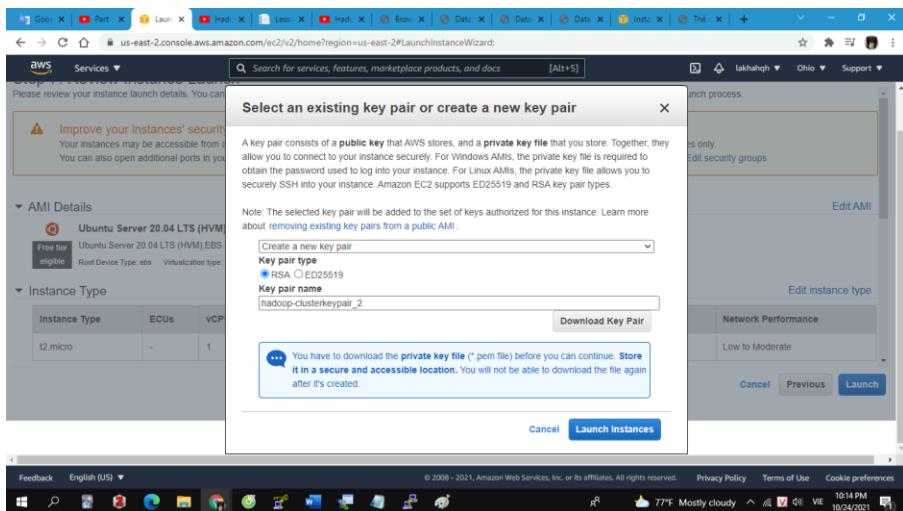




Tạo key pair



Tạo Instances



Instances sau khi tạo

The screenshot shows the AWS EC2 Instances page with four instances listed:

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
NameNode00	i-017c8dc8a5becfa	Running	t2.micro	2/2 checks passed	-	us-east-2a
DataNode003	i-0379742c5bbde5a13	Running	t2.micro	2/2 checks passed	-	us-east-2a
DataNode002	i-0e0d528a4bbcec3aa	Running	t2.micro	2/2 checks passed	-	us-east-2a
DataNode001	i-0bd49c7b95e883e22	Running	t2.micro	2/2 checks passed	-	us-east-2a

Tải putty <https://www.putty.org/>

Tải WinSCP <https://winscp.net/eng/download.php>

2. Ánh xạ các biến môi trường Amazon EC2

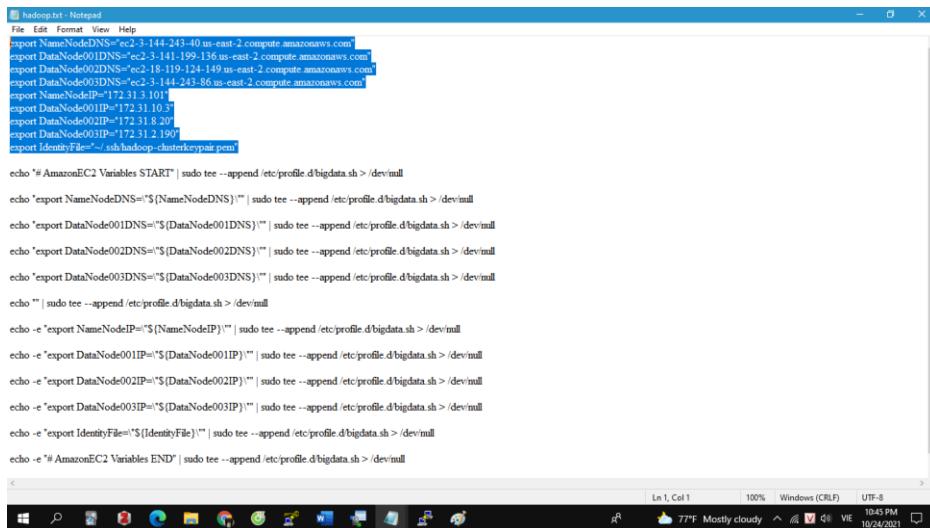
- Trong phần này, ta sẽ thêm các Biến môi trường sử dụng trong các cụm để tham chiếu đến các Node.

Lưu và xem thông tin DNS của các Node

The screenshot shows the AWS EC2 Instances page with the NameNode00 instance selected. The details pane shows:

Instance ID	Public IPv4 address	Private IPv4 addresses
i-017c8dc8a5becfa (NameNode00)	3.144.243.40 open address	172.31.3.101
IPv6 address	Instance state	Public IPv4 DNS
-	Running	ec2-3-144-243-40.us-east-2.compute.amazonaws.com open address
Private IPv4 DNS	Instance type	Elastic IP addresses
ip-172-31-3-101.us-east-2.compute.internal	t2.micro	-

Ta sẽ tiến hành thu tập Public DNS và Private IP của NameNode và các DataNode, lưu vào trong file sau



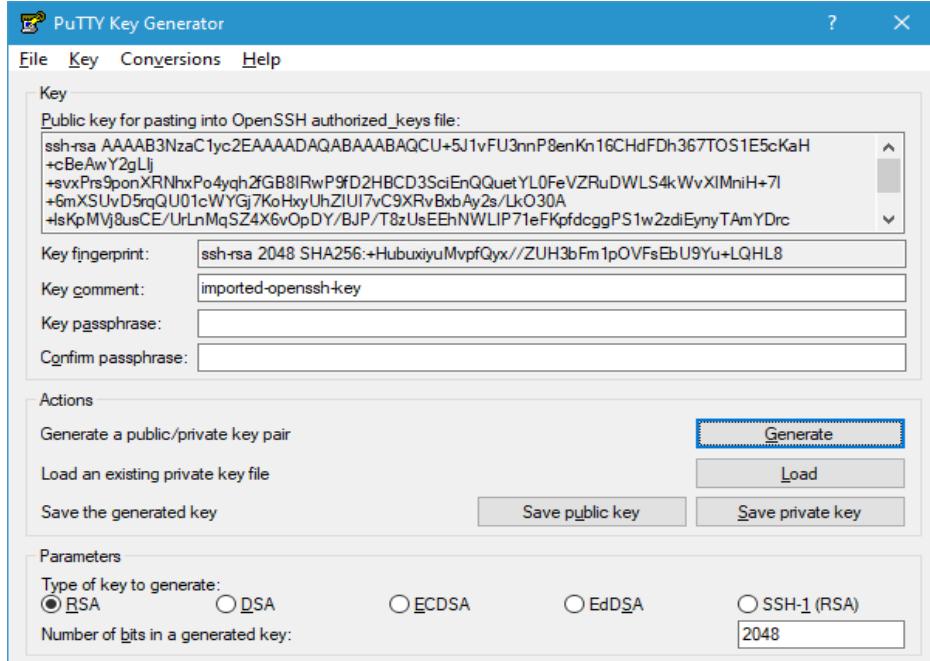
```

hadoop.txt - Notepad
File Edit Format View Help
export NameNodeDNS="ec2-3-144-243-40.us-east-2.compute.amazonaws.com"
export DataNode01DNS="ec2-3-141-199-136.us-east-2.compute.amazonaws.com"
export DataNode02DNS="ec2-3-119-124-149.us-east-2.compute.amazonaws.com"
export DataNode03DNS="ec2-3-1-243-86.us-east-2.compute.amazonaws.com"
export NameNodeIP="172.31.1.101"
export DataNode01IP="172.31.1.107"
export DataNode02IP="172.31.1.207"
export DataNode03IP="172.31.1.197"
export IdentityFile=".ssh/hadoop-clusterkeypair.pem"

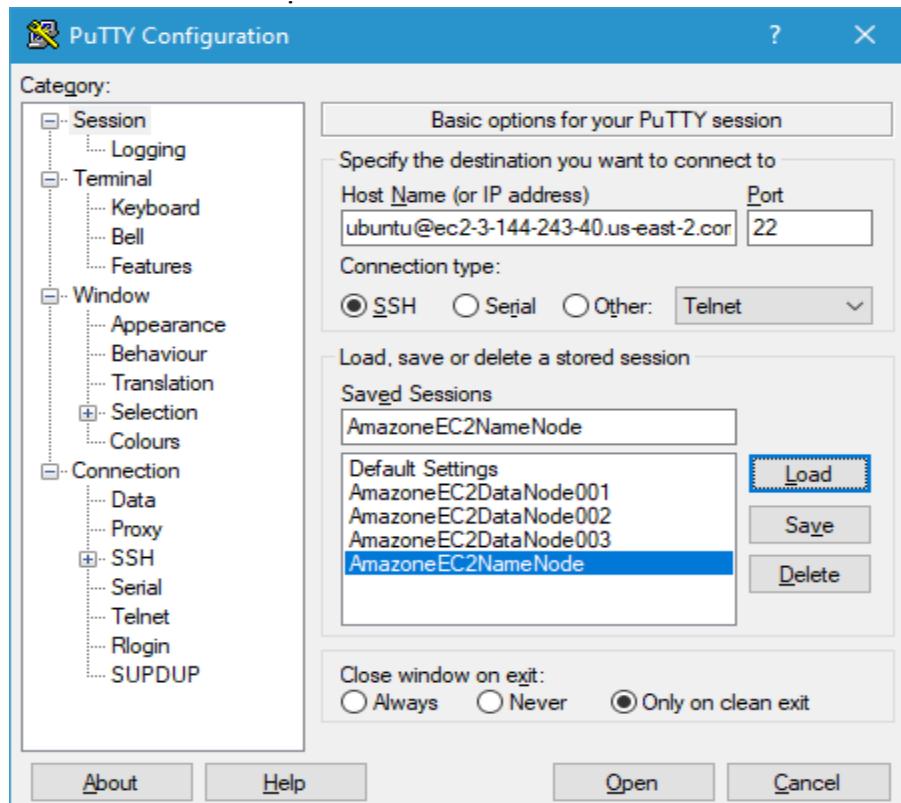
echo "# AmazonEC2 Variables START" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
echo "export NameNodeDNS=\$NameNodeDNS)" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
echo "export DataNode01DNS=\${DataNode01DNS})" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
echo "export DataNode02DNS=\${DataNode02DNS})" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
echo "export DataNode03DNS=\${DataNode03DNS})" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
echo "" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
echo "e 'export NameNodeIP=\$NameNodeIP'" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
echo "e 'export DataNode01IP=\$DataNode01IP'" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
echo "e 'export DataNode02IP=\$DataNode02IP'" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
echo "e 'export DataNode03IP=\$DataNode03IP'" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
echo "e 'export IdentityFile=\$IdentityFile'" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
echo "e '# AmazonEC2 Variables END'" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null

```

Sử dụng PuTTY gen để tạo khóa PPK



Cấu hình PuTTY để tạo kết nối SSH



- Bước tiếp theo ta sẽ tạo biến môi trường chung cho Big Data, các biến này bao gồm
 - Java - JAVA_HOME
 - Hadoop - HADOOP_HOME, Hive - HIVE_HOME, Pig - PIG_HOME

Ta sẽ bắt đầu bằng cách tạo một tệp có tên `bigdata.sh` để sử dụng trong các bản cài đặt sau:

```
sudo touch /etc/profile.d/bigdata.sh
```

```
sudo chmod +x /etc/profile.d/bigdata.sh
```

```
sudo echo -e '#!/bin/bash\n# Environment Variables for Big Data tools\n' | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
```

```
ubuntu@ip-172-31-10-9: ~
```

Ubuntu Pro delivers the most comprehensive open source security and compliance features.

```
https://ubuntu.com/aws/pro
```

11 updates can be applied immediately.
1 of these updates is a standard security update.
To see these additional updates run: apt list --upgradeable

Failed to connect to https://changelogs.ubuntu.com/meta-release-lts. Check your Internet connection or proxy settings

*** System restart required ***
Last login: Fri Oct 15 08:39:20 2021 from 1.53.204.109
ubuntu@ip-172-31-1-10-9: ~ [root]# No such file or directory
-bash: /etc/bin/bashInEnvironment: No such file or directory
ubuntu@ip-172-31-3-101: ~\$ sudo touch /etc/profile.d/bigdata.sh
ubuntu@ip-172-31-3-101: ~\$ sudo chmod +x /etc/profile.d/bigdata.sh
ubuntu@ip-172-31-3-101: ~\$ echo '#!/bin/bashInEnvironment Variables for Big Data tools\n! \$ sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
ubuntu@ip-172-31-3-101: ~\$ []

ubuntu@ip-172-31-2-20: ~

Ubuntu Pro delivers the most comprehensive open source security and compliance features.

```
https://ubuntu.com/aws/pro
```

11 updates can be applied immediately.
1 of these updates is a standard security update.
To see these additional updates run: apt list --upgradeable

Failed to connect to https://changelogs.ubuntu.com/meta-release-lts. Check your Internet connection or proxy settings

*** System restart required ***
Last login: Fri Oct 15 08:04:06 2021 from 1.53.204.109
ubuntu@ip-172-31-2-20: ~ [root]# No such file or directory
-bash: /etc/bin/bashInEnvironment: No such file or directory
ubuntu@ip-172-31-2-20: ~\$ sudo touch /etc/profile.d/bigdata.sh
ubuntu@ip-172-31-2-20: ~\$ sudo chmod +x /etc/profile.d/bigdata.sh
ubuntu@ip-172-31-2-20: ~\$ echo '#!/bin/bashInEnvironment Variables for Big Data tools\n! \$ sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
ubuntu@ip-172-31-2-20: ~\$ []

ubuntu@ip-172-31-2-190: ~

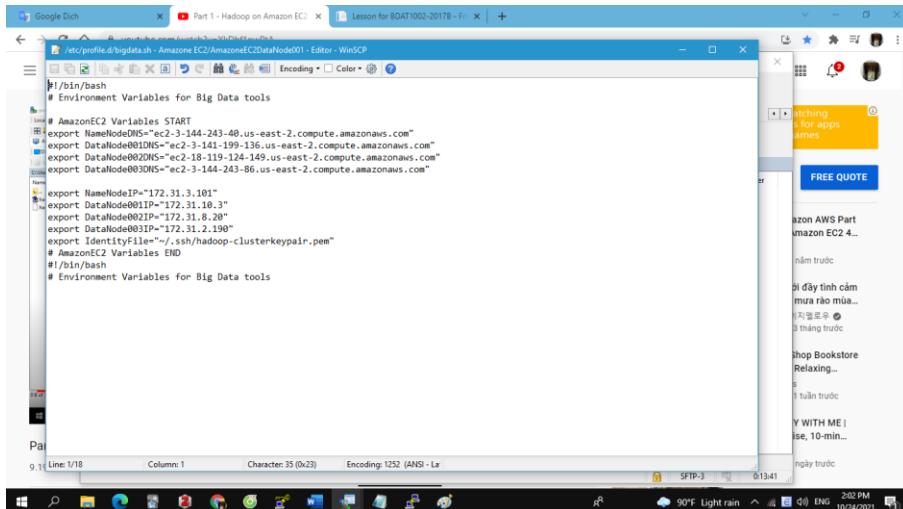
Ubuntu Pro delivers the most comprehensive open source security and compliance features.

```
https://ubuntu.com/aws/pro
```

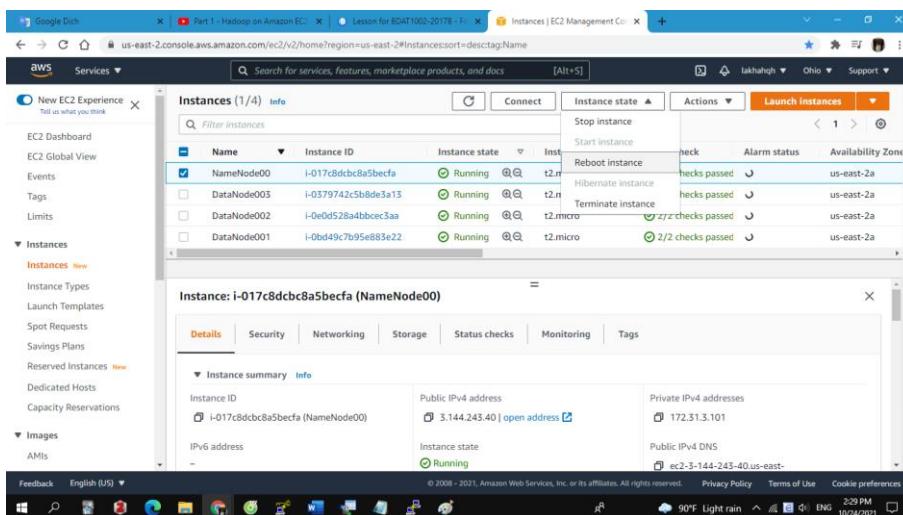
11 updates can be applied immediately.
1 of these updates is a standard security update.
To see these additional updates run: apt list --upgradeable

Failed to connect to https://changelogs.ubuntu.com/meta-release-lts. Check your Internet connection or proxy settings

*** System restart required ***
Last login: Fri Oct 15 08:13:42 2021 from 1.53.204.109
ubuntu@ip-172-31-2-190: ~ [root]# No such file or directory
-bash: /etc/bin/bashInEnvironment: No such file or directory
ubuntu@ip-172-31-2-190: ~\$ sudo touch /etc/profile.d/bigdata.sh
ubuntu@ip-172-31-2-190: ~\$ sudo chmod +x /etc/profile.d/bigdata.sh
ubuntu@ip-172-31-2-190: ~\$ echo '#!/bin/bashInEnvironment Variables for Big Data tools\n! \$ sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
ubuntu@ip-172-31-2-190: ~\$ []



Reboot lại các Instances



Map Nodes in etc/hosts File on Amazon EC2 Instances: quản lý bí danh cho các phiên bản trong cụm publichost=\${NameNodeDNS}

publichost=\${DataNode001DNS}

```
publichost=${DataNode002DNS}
```

```
publichost=${DataNode003DNS}
```

ubuntu@ip-172-31-3-101: ~

```
To see these additional updates run: apt list --upgradable
Failed to connect to https://changelogs.ubuntu.com/meta-release-its. Check your
Internet connection or proxy settings

*** System restart required ***
Last login: Fri Oct 15 08:01:00 2021 from 1.53.204.109
bash: ~!/bin/bash$ Environment: No such file or directory
ubuntu@ip-172-31-3-101:~$ sudo chmod +x /etc/profile.d/bigdata.sh
ubuntu@ip-172-31-3-101:~$ sudo echo '#!/bin/bash\n# Environment Variables for Bi
g Data toolchain\n' | sudo tee -append /etc/profile.d/bigdata.sh > /dev/null
ubuntu@ip-172-31-3-101:~$ sudo chown root /etc/profile.d/bigdata.sh
ubuntu@ip-172-31-3-101:~$ hostname
ip-172-31-3-101
ubuntu@ip-172-31-3-101:~$ publischor="#DataNode001DNS"
ubuntu@ip-172-31-3-101:~$ [ 1
juban@ip-172-31-2-19 ~ - ]+ 100% 0.00s user 0.00s system 0:00.00 total
juban@ip-172-31-2-19:~ - ]+ 100% 0.00s user 0.00s system 0:00.00 total

11 updates can be applied immediately.
1 of these updates is a standard security update.
To see these additional updates run: apt list --upgradable
Failed to connect to https://changelogs.ubuntu.com/meta-release-its. Check your In
ternet connection or proxy settings

*** System restart required ***
Last login: Fri Oct 15 08:01:00 2021 from 1.53.204.109
bash: ~!/bin/bash$ No such file or directory
ubuntu@ip-172-31-3-101:~$ sudo chmod +x /etc/profile.d/bigdata.sh
ubuntu@ip-172-31-3-101:~$ sudo echo '#!/bin/bash\n# Environment Variables for Bi
g Data toolchain\n' | sudo tee -append /etc/profile.d/bigdata.sh > /dev/null
ubuntu@ip-172-31-3-101:~$ sudo chown root /etc/profile.d/bigdata.sh
ubuntu@ip-172-31-3-101:~$ hostname
ip-172-31-3-101
ubuntu@ip-172-31-3-101:~$ publischor="#DataNode001DNS"
ubuntu@ip-172-31-3-101:~$ [ 1
juban@ip-172-31-2-19 ~ - ]+ 100% 0.00s user 0.00s system 0:00.00 total
juban@ip-172-31-2-19:~ - ]+ 100% 0.00s user 0.00s system 0:00.00 total

11 updates can be applied immediately.
1 of these updates is a standard security update.
To see these additional updates run: apt list --upgradable
Failed to connect to https://changelogs.ubuntu.com/meta-release-its. Check your In
ternet connection or proxy settings

*** System restart required ***
Last login: Fri Oct 15 08:01:00 2021 from 1.53.204.109
bash: ~!/bin/bash$ No such file or directory
ubuntu@ip-172-31-3-101:~$ sudo chmod +x /etc/profile.d/bigdata.sh
ubuntu@ip-172-31-3-101:~$ sudo echo '#!/bin/bash\n# Environment Variables for Bi
g Data toolchain\n' | sudo tee -append /etc/profile.d/bigdata.sh > /dev/null
ubuntu@ip-172-31-3-101:~$ sudo chown root /etc/profile.d/bigdata.sh
ubuntu@ip-172-31-3-101:~$ hostname
ubuntu@ip-172-31-3-101:~$ publischor="#DataNode002DNS"
ubuntu@ip-172-31-3-101:~$ [ 1
juban@ip-172-31-2-19 ~ - ]+ 100% 0.00s user 0.00s system 0:00.00 total
juban@ip-172-31-2-19:~ - ]+ 100% 0.00s user 0.00s system 0:00.00 total
```

Tên máy chủ được thêm vào biến môi trường

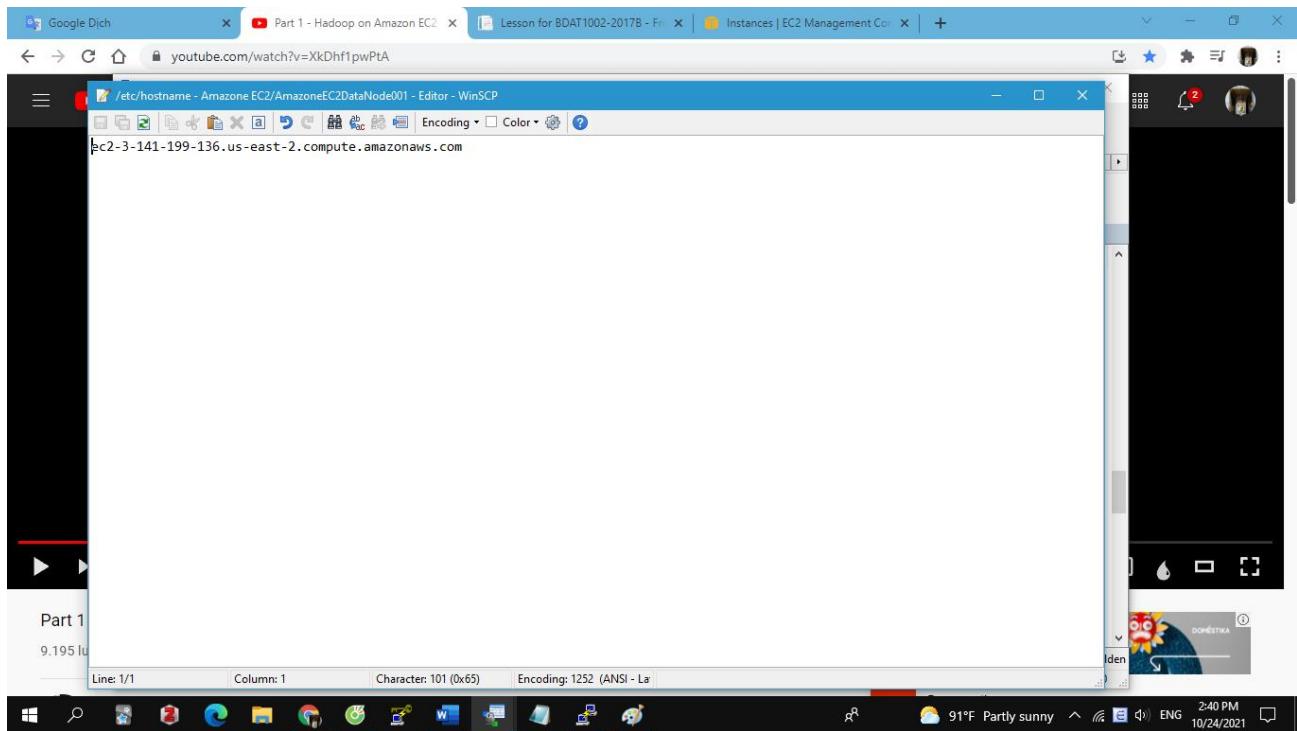
```
sudo hostname ${publichost}
```

```
sudo rm -rf /etc/hostname
```

```
echo -e "${publichost}" | sudo tee --append /etc/hostname > /dev/null
```

```
sudo chown root /etc/hostname
```

Thay đổi hostname sang public DNS



```
sudo rm -rf /etc/hosts
echo -e "127.0.0.1\localhost" | sudo tee --append /etc/hosts > /dev/null
echo -e "127.0.1.1\$publichost" | sudo tee --append /etc/hosts > /dev/null
echo -e "${NameNodeIP}\thadoop-master" | sudo tee --append /etc/hosts > /dev/null
echo -e "${DataNode001IP}\tDataNode001" | sudo tee --append /etc/hosts > /dev/null
echo -e "${DataNode002IP}\tDataNode002" | sudo tee --append /etc/hosts > /dev/null
echo -e "${DataNode003IP}\tDataNode003" | sudo tee --append /etc/hosts > /dev/null
echo -e "\n# The following lines are desirable for IPv6 capable hosts" | sudo tee --append /etc/hosts > /dev/null
echo -e "::1 ip6-localhost ip6-loopback" | sudo tee --append /etc/hosts > /dev/null
echo -e "fe00::0 ip6-localnet" | sudo tee --append /etc/hosts > /dev/null
echo -e "ff00::0 ip6-mcastprefix" | sudo tee --append /etc/hosts > /dev/null
echo -e "ff02::1 ip6-allnodes" | sudo tee --append /etc/hosts > /dev/null
echo -e "ff02::2 ip6-allrouters" | sudo tee --append /etc/hosts > /dev/null
echo -e "ff02::3 ip6-allhosts" | sudo tee --append /etc/hosts > /dev/null
sudo chown root /etc/hosts
```

```

ubuntu@ec2-3-144-243-40:~$ echo -e "#(DataNode002IP)\tDataNode002" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-243-40:~$ echo -e "#(DataNode003IP)\tDataNode003" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-243-40:~$ echo -e "#$ The following lines are desirable for IPv6 capable hosts" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-243-40:~$ echo -e "::1 ip6-localhost ip6-loopback" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-243-40:~$ echo -e "ff00:0:0:0:0:0:0:1 ip6-localnet" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-243-40:~$ echo -e "ff00:0:0:0:0:0:0:1 ip6-mcastprefix" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-243-40:~$ echo -e "ff00:0:0:0:0:0:0:1 ip6-allnodes" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-243-40:~$ echo -e "ff00:0:0:0:0:0:0:1 ip6-allrouters" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-243-40:~$ sudo chown root /etc/hosts
ubuntu@ec2-3-144-243-40:~$ [REDACTED]

ubuntu@ec2-3-144-199-136:~$ echo -e "#(DataNode002IP)\tDataNode002" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-199-136:~$ echo -e "#(DataNode003IP)\tDataNode003" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-199-136:~$ echo -e "#$ The following lines are desirable for IPv6 capable hosts" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-199-136:~$ echo -e "::1 ip6-localhost ip6-loopback" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-199-136:~$ echo -e "ff00:0:0:0:0:0:0:1 ip6-localnet" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-199-136:~$ echo -e "ff00:0:0:0:0:0:0:1 ip6-mcastprefix" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-199-136:~$ echo -e "ff00:0:0:0:0:0:0:1 ip6-allnodes" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-199-136:~$ echo -e "ff00:0:0:0:0:0:0:1 ip6-allrouters" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-3-144-199-136:~$ sudo chown root /etc/hosts
ubuntu@ec2-3-144-199-136:~$ [REDACTED]

ubuntu@ec2-18-110-124-140:~$ echo -e "#(DataNode002IP)\tDataNode002" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-18-110-124-140:~$ echo -e "#(DataNode003IP)\tDataNode003" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-18-110-124-140:~$ echo -e "#$ The following lines are desirable for IPv6 capable hosts" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-18-110-124-140:~$ echo -e "::1 ip6-localhost ip6-loopback" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-18-110-124-140:~$ echo -e "ff00:0:0:0:0:0:0:1 ip6-localnet" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-18-110-124-140:~$ echo -e "ff00:0:0:0:0:0:0:1 ip6-mcastprefix" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-18-110-124-140:~$ echo -e "ff00:0:0:0:0:0:0:1 ip6-allnodes" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-18-110-124-140:~$ echo -e "ff00:0:0:0:0:0:0:1 ip6-allrouters" | sudo tee --append /etc/hosts > /dev/null
ubuntu@ec2-18-110-124-140:~$ sudo chown root /etc/hosts
ubuntu@ec2-18-110-124-140:~$ [REDACTED]

```

```

/etc/hosts - Amazon EC2/AmazonCNameNode - WinSCP
File Home Insert Draw Design Layout References Mailings Review View Zotero Help Foxit Reader PDF Share Comments
Local Mark Files Commands Session Options Remote Help
Editor - WinSCP
[REDACTED]
127.0.0.1 localhost
127.0.1.1 ec2-3-144-243-40.us-east-2.compute.amazonaws.com
172.31.3.101 hadoop-master
172.31.10.3 DataNode001
172.31.8.20 DataNode002
172.31.2.190 DataNode003

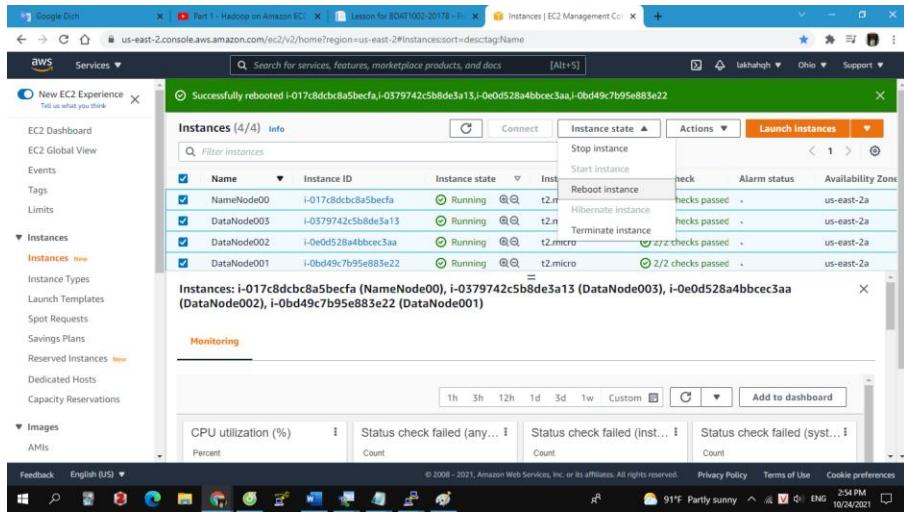
# The following lines are desirable for IPv6 capable hosts
::1 ip6-localhost ip6-loopback
:Fe00:0 ip6-localnet
:FF00:0:0:0:0:0:0:1 ip6-mcastprefix
:FF02::1 ip6-allnodes
:FF02::2 ip6-allrouters
:FF02::3 ip6-allhosts
[REDACTED]

```

127.0.1.1 ec2-3-144-243-40.us-east-2.compute.amazonaws.com: public DNS

172.31.3.101 hadoop-master: NameNode

Reboot



Connect and Install Passwordless SSH: chạy cụm Hadoop trên hệ thống nhiều nút
 sudo rm -rf ~/.ssh/config

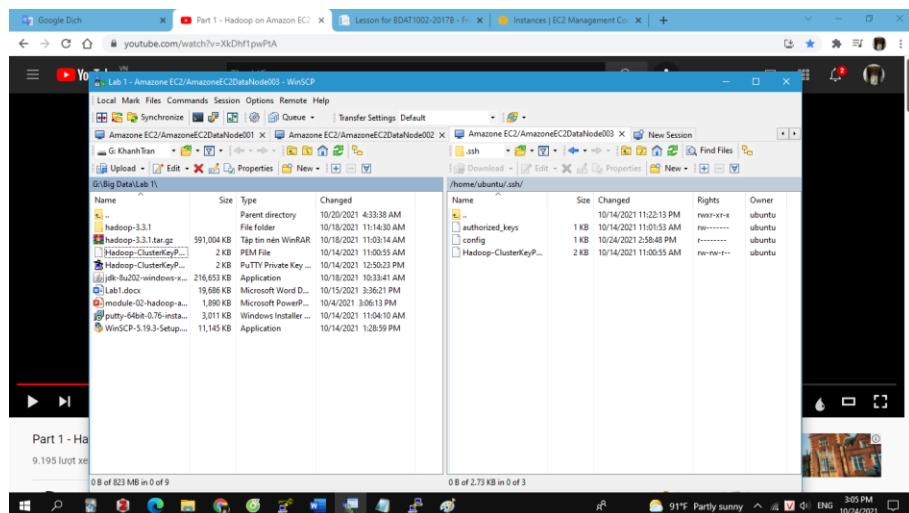
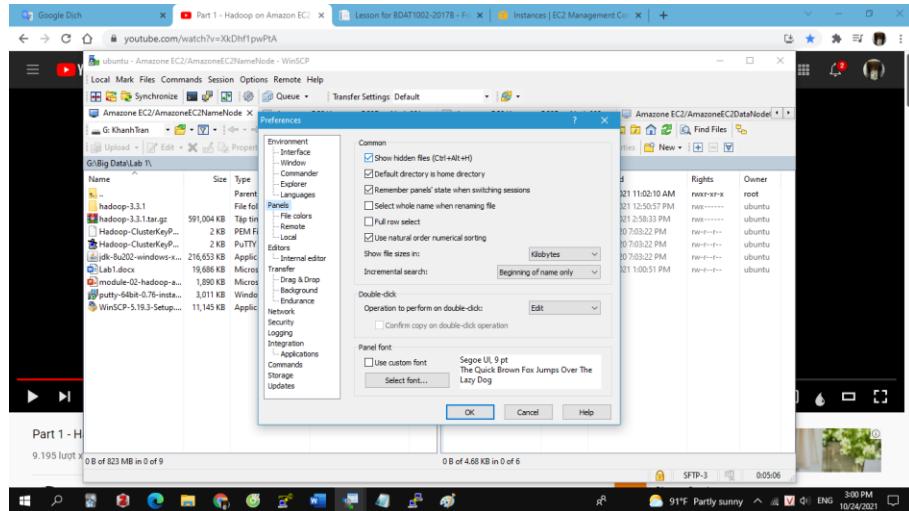
```
echo -e "Host 0.0.0.0" | tee --append ~/.ssh/config > /dev/null
echo -e " HostName ${NameNodeDNS}" | tee --append ~/.ssh/config > /dev/null
echo -e " User ubuntu" | tee --append ~/.ssh/config > /dev/null
echo -e " IdentityFile ${IdentityFile}" | tee --append ~/.ssh/config > /dev/null
echo -e "Host hadoop-master" | tee --append ~/.ssh/config > /dev/null
echo -e " HostName ${NameNodeDNS}" | tee --append ~/.ssh/config > /dev/null
echo -e " User ubuntu" | tee --append ~/.ssh/config > /dev/null
echo -e " IdentityFile ${IdentityFile}" | tee --append ~/.ssh/config > /dev/null
echo -e "Host DataNode001" | tee --append ~/.ssh/config > /dev/null
echo -e " HostName ${DataNode001DNS}" | tee --append ~/.ssh/config > /dev/null
echo -e " User ubuntu" | tee --append ~/.ssh/config > /dev/null
echo -e " IdentityFile ${IdentityFile}" | tee --append ~/.ssh/config > /dev/null
echo -e "Host DataNode002" | tee --append ~/.ssh/config > /dev/null
echo -e " HostName ${DataNode002DNS}" | tee --append ~/.ssh/config > /dev/null
echo -e " User ubuntu" | tee --append ~/.ssh/config > /dev/null
echo -e " IdentityFile ${IdentityFile}" | tee --append ~/.ssh/config > /dev/null
echo -e "Host DataNode003" | tee --append ~/.ssh/config > /dev/null
echo -e " HostName ${DataNode003DNS}" | tee --append ~/.ssh/config > /dev/null
```

```
echo -e " User ubuntu" | tee --append ~/.ssh/config > /dev/null
echo -e " IdentityFile ${IdentityFile}" | tee --append ~/.ssh/config > /dev/null
```

sudo chmod 0400 ~/.ssh/config

sudo chmod 0400 \${IdentityFile}

Chọn show hidden file để hiện file ẩn ssh



sudo chmod 0400 ~/.ssh/config

sudo chmod 0400 ~/.ssh/hadoop-clusterkeypair.pem

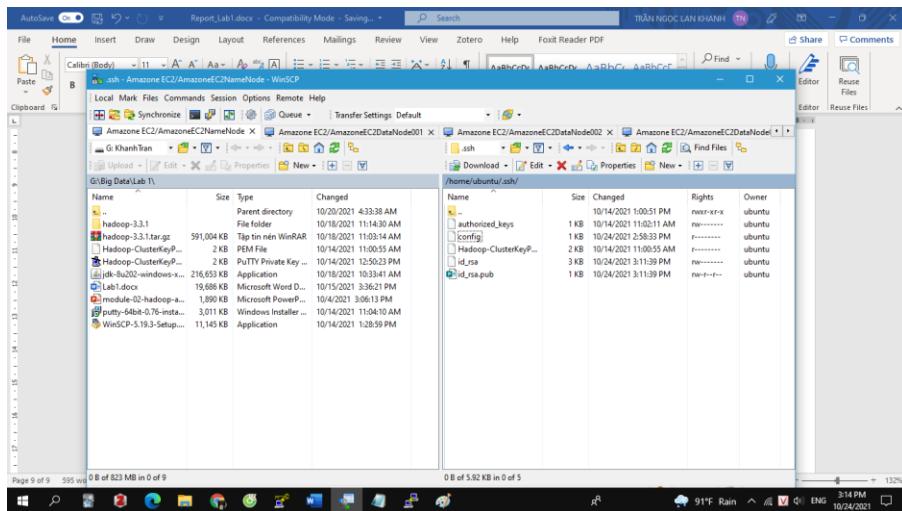
Lab 1

Nhập môn dữ liệu lớn

```
sudo rm -rf ~/.ssh/id_rsa*
```

```
sudo rm -rf ~/.ssh/known_hosts
```

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



sudo chmod 0600 ~/.ssh/id_rsa.pub

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

hosts=0.0.0.0,127.0.0.1,127.0.1.1,hadoop-master,DataNode001,DataNode002,DataNode003

ssh-keyscan -H \${hosts} >> ~/.ssh/known_hosts

```
ubuntu@ec2-3-144-243-40:~$ 
[SHA256]-----+
ubuntu@ec2-3-144-243-40:~$ sudo chmod 0600 ~/.ssh/id_rsa.pub
ubuntu@ec2-3-144-243-40:~$ sudo cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
[SHA256]-----+
cat: /root/.ssh/authorized_keys: No such file or directory
ubuntu@ec2-3-144-243-40:~$ 
ubuntu@ec2-3-144-243-40:~$ 
ubuntu@ec2-3-144-243-40:~$ sudo cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
[SHA256]-----+
ubuntu@ec2-3-144-243-40:~$ hosts=0.0.0.0,127.0.0.1,127.0.1.1,hadoop-master,DataNode001,DataNode002,DataNode003
ubuntu@ec2-3-144-243-40:~$ 
ubuntu@ec2-3-144-243-40:~$ ssh-keyscan -H ${hosts} >> ~/.ssh/known_hosts
# 0.0.0.0:22 SSH-2.0-OpenSSH_8.2p1 Ubuntu-4ubuntu0.2
ubuntu@ec2-3-144-243-40:~$ 
ubuntu@ec2-18-119-124-149:~$ 
[SHA256]-----+
ubuntu@ec2-18-119-124-149:~$ sudo chmod 0600 ~/.ssh/id_rsa.pub
ubuntu@ec2-18-119-124-149:~$ sudo cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
[SHA256]-----+
ubuntu@ec2-18-119-124-149:~$ hosts=0.0.0.0,127.0.0.1,127.0.1.1,hadoop-master,DataNode001,DataNode002,DataNode003
ubuntu@ec2-18-119-124-149:~$ 
ubuntu@ec2-18-119-124-149:~$ ssh-keyscan -H ${hosts} >> ~/.ssh/known_hosts
# 0.0.0.0:22 SSH-2.0-OpenSSH_8.2p1 Ubuntu-4ubuntu0.2
ubuntu@ec2-18-119-124-149:~$ 
```

Chỉ chạy các lệnh sau trên namenode

sudo cat ~/.ssh/id_rsa.pub | ssh -o StrictHostKeyChecking=no DataNode001 'cat >> ~/.ssh/authorized_keys'

```
sudo cat ~/ssh/id_rsa.pub | ssh -o StrictHostKeyChecking=no DataNode002 'cat >> ~/ssh/authorized_keys'
sudo cat ~/ssh/id_rsa.pub | ssh -o StrictHostKeyChecking=no DataNode003 'cat >> ~/ssh/authorized_keys'
```

The screenshot shows two terminal windows side-by-side. Both windows are running on a Windows desktop environment. The top window has the title 'ubuntu@ec2-3-144-243-40 ~' and the bottom window has the title 'ubuntu@ec2-18-119-124-149 ~'. Both windows display the same command being run: 'cat ~/ssh/id_rsa.pub | ssh -o StrictHostKeyChecking=no DataNode002 'cat >> ~/ssh/authorized_keys''. The output of the command is visible in both windows, showing the addition of the public key to the authorized_keys file.

3. Tải Java

```
sudo apt-get -y install default-jdk
```

The screenshot shows three terminal windows side-by-side. All three windows have the title 'ubuntu@ec2-3-144-243-40 ~'. They all display the command 'sudo apt-get -y install default-jdk'. The output of the command is visible in each window, showing the progress of the package download and installation.

```
echo "# JAVA Variables START" | sudo tee --append /etc/profile.d/bigdata.sh >/dev/null
```

```
echo "export JAVA_HOME=/usr/lib/jvm/default-java" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
```

```
echo "PATH=$PATH:$JAVA_HOME/bin" | sudo tee --append /etc/profile.d/bigdata.sh > /dev/null
```

```
echo "# JAVA Variables END" | sudo tee --append /etc/profile.d/bigdata.sh >/dev/null
```

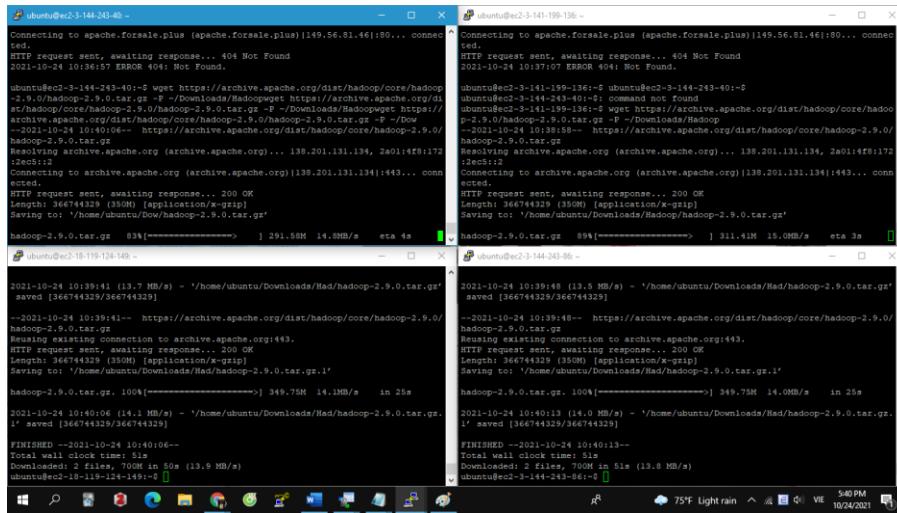
Xác nhận rằng các biến Java đã được thêm vào

```
profile.d - Amazon EC2/AmazonEC2NameNode - WinSCP
Local File Commands Session Options Remote Help
Synchronize Queue Transfer Settings Default
Amazon EC2/AmazonEC2NameNode Amazon EC2/AmazonEC2DataNode001 Amazon EC2/AmazonEC2DataNode002 Amazon EC2/AmazonEC2DataNode003 New Session
K KhanhTran - etcd/profile/bigdata.x - Amazon EC2/AmazonEC2NameNode - Editor - WinSCP
Editing Data (L)
Name: etcd/profile/bigdata.x
Variables for Big Data tools\n
# AmazoneC2 Variables START\n
export NameNodeDNS="ec2-3-144-243-40.us-east-2.compute.amazonaws.com"\n
export DataNode@0IDNS="ec2-3-141-199-136.us-east-2.compute.amazonaws.com"\n
export DataNode@0ZDNS="ec2-18-119-124-149.us-east-2.compute.amazonaws.com"\n
export DataNode@03DNS="ec2-3-144-243-86.us-east-2.compute.amazonaws.com"\n
# AmazoneC2 Variables END\n
# Environment Variables for Big Data tools\n
# JAVA Variables START\n
export JAVA_HOME=/usr/lib/jvm/default-java\n
PATH=$PATH:$JAVA_HOME/bin\n
# JAVA Variables END\n
Line: 1/21 Column: 1 Character: 45 (0x2D) Encoding: 1252 (ANSI - La
0 B of 823 MB in 0 s 715 B of 10.8 KB in 1 of 10
75°F Light rain ENG 5:26 PM
```

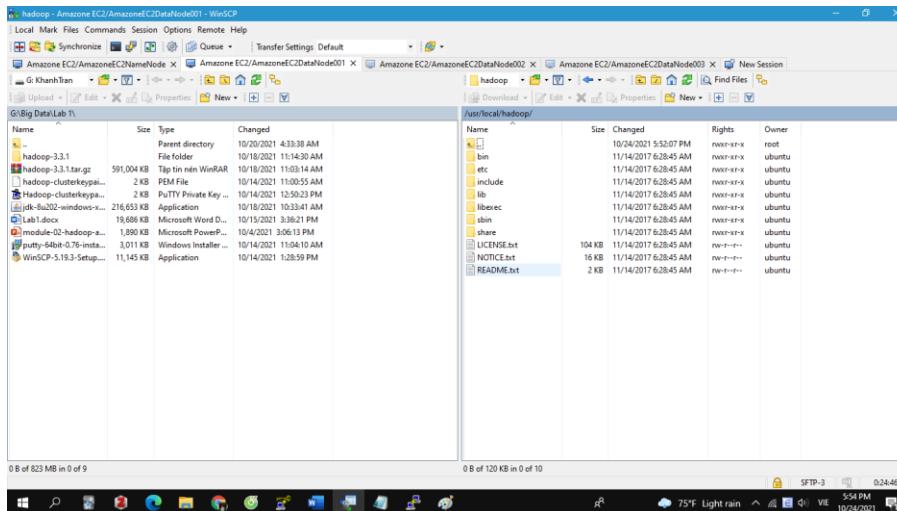
4. Tải Hadoop 2.9.0

Sử dụng đoạn code sau để tải Hadoop về cho từng DataNode và NameNode

```
Wget http://apache.forsale.plus/hadoop/common/hadoop-2.9.0/hadoop-2.9.0.tar.gz  
P~/Downloads/Hadoop
```



Hadoop file được trích xuất tại đây



Đẩy biến môi trường vào cài đặt Hadoop

```

profile.d - Amazon EC2/AmazonECCDataNode01 - WinSCP
Local Mark Files Commands Session Options Remote Help
File Synchronize Queue Transfer Settings Default
Amazon EC2/AmazonECCDataNode01 - /etc/profile.d/bigdata.sh - Amazon EC2/AmazonECCDataNode01 - Editor - WinSCP
G:\KhoaHoc\Lab1
Name
└── hadoop-3.1.1
    ├── hadoop-3.1.1.tar.gz
    ├── hadoop-clusterkeypair.pem
    └── hadoop-clusterkeypair.pem
    Lab1.docx
    module-02-hadoop.sh
    putty-54bf-0.76-intel...
    WinSCP-5.19.3-Setup...
# Environment Variables for Big Data tools

# JAVA Variables START
# JAVA_HOME=/usr/lib/jvm/default-java
PATH=$JAVA_HOME:$JAVA_HOME/bin
# JAVA Variables END
# HADOOP Variables START
export HADOOP_HOME="/usr/local/hadoop"
export HADOOP_CONF_DIR="$HADOOP_HOME/etc/hadoop"
export HADOOP_DATA_HOME="$HOME/hadoop_data/hdfs"
PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
# HADOOP Variables END

# AmazonEC2 Variables
#!/bin/bash

# Environment Variables for Big Data tools

# JAVA Variables START
# JAVA_HOME=/usr/lib/jvm/default-java
PATH=$JAVA_HOME:$JAVA_HOME/bin
# JAVA Variables END
# HADOOP Variables START
export HADOOP_HOME="/usr/local/hadoop"
export HADOOP_CONF_DIR="$HADOOP_HOME/etc/hadoop"
export HADOOP_DATA_HOME="$HOME/hadoop_data/hdfs"
PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
# HADOOP Variables END

Line: 1/28 Column: 1 Character: 35 (0x23) Encoding: 1252 (ANSI - La

```

0 B of 823 MB in 0 of 9 633 B of 10.7 KB in 1 of 10 SFTP-3 0:27:37

Xây dựng các thư mục dữ liệu Hadoop

```

mkdir -p $HADOOP_DATA_HOME/datanode
mkdir -p $HADOOP_DATA_HOME/namenode
mkdir -p $HADOOP_DATA_HOME/tmp

```

```

ubuntu@ec2-3-14-243-40: ~
https://ubuntu.com/aws/pro
79 updates can be applied immediately.
To see these additional updates run: apt list --upgradable
Failed to connect to https://changelogs.ubuntu.com/meta-release-its. Check your Internet connection or proxy settings

Last login: Sun Oct 24 10:29:31 2021 from 1.53.144.9
.bashrc =>#!/bin/bash# Environment: No such file or directory
.bashrc =>#/bin/bash# Environment: No such file or directory
ubuntu@ec2-3-14-243-40:~$ echo $HADOOP_HOME
/usr/local/hadoop
ubuntu@ec2-3-14-243-40:~$ mkdir -p $HADOOP_DATA_HOME/datanode
ubuntu@ec2-3-14-243-40:~$ mkdir -p $HADOOP_DATA_HOME/namenode
ubuntu@ec2-3-14-243-40:~$ mkdir -p $HADOOP_DATA_HOME/tmp
ubuntu@ec2-3-14-243-40:~$ 

ubuntu@ec2-18-119-124-149: ~
https://ubuntu.com/aws/pro
79 updates can be applied immediately.
To see these additional updates run: apt list --upgradable
Failed to connect to https://changelogs.ubuntu.com/meta-release-its. Check your Internet connection or proxy settings

Last login: Sun Oct 24 10:29:31 2021 from 1.53.144.9
.bashrc =>#!/bin/bash# Environment: No such file or directory
.bashrc =>#/bin/bash# Environment: No such file or directory
ubuntu@ec2-18-119-124-149:~$ echo $HADOOP_HOME
/usr/local/hadoop
ubuntu@ec2-18-119-124-149:~$ mkdir -p $HADOOP_DATA_HOME/datanode
ubuntu@ec2-18-119-124-149:~$ mkdir -p $HADOOP_DATA_HOME/namenode
ubuntu@ec2-18-119-124-149:~$ mkdir -p $HADOOP_DATA_HOME/tmp
ubuntu@ec2-18-119-124-149:~$ 

ubuntu@ec2-3-14-243-56: ~
compliance features.
https://ubuntu.com/aws/pro
79 updates can be applied immediately.
To see these additional updates run: apt list --upgradable
Failed to connect to https://changelogs.ubuntu.com/meta-release-its. Check your Internet connection or proxy settings

Last Login: Sun Oct 24 10:29:09 2021 from 1.53.144.9
ubuntu@ec2-3-14-243-56:~$ echo $HADOOP_HOME
/usr/local/hadoop
ubuntu@ec2-3-14-243-56:~$ mkdir -p $HADOOP_DATA_HOME/datanode
ubuntu@ec2-3-14-243-56:~$ mkdir -p $HADOOP_DATA_HOME/namenode
ubuntu@ec2-3-14-243-56:~$ mkdir -p $HADOOP_DATA_HOME/tmp
ubuntu@ec2-3-14-243-56:~$ 

ubuntu@ec2-3-14-243-56: ~

```

```
export JAVA_HOME=/usr/lib/jvm/default-java
```

```
hadoop - Amazon EC2/AmazonEC2NameNode - WinSCP
Local File Commands Session Options Remote Help
Synchronize Queue Transfer Settings: Default
Amazon EC2/AmazonEC2NameNode Amazon EC2/AmazonEC2DataNode001 Amazon EC2/AmazonEC2DataNode002 Amazon EC2/AmazonEC2DataNode003 New Session
File /usr/local/hadoop/etc/hadoop/hadoop-env.sh - Amazon EC2/AmazonEC2NameNode - Editor - WinSCP

# 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.

# Set Hadoop-specific environment variables here.

# The only required environment variable is JAVA_HOME. All others are
# optional. When running a distributed configuration it is best to
# set JAVA_HOME in this file, so that it is correctly defined on
# remote nodes.

# The Java implementation to use.
#export JAVA_HOME=$JAVA_HOME
export JAVA_HOME=/usr/lib/jvm/default-java

# The Jsvc implementation to use. Jsvc is required to run secure datanodes
# that bind to privileged ports to provide authentication of data transfer
# protocol. Jsvc is not required if SASL is configured for authentication of
# data transfer protocol using non-privileged ports.
#export JSVC_HOME=$JSVC_HOME

export HADOOP_CONF_DIR=${HADOOP_CONF_DIR:-"/etc/hadoop"}
```

Thay đổi cấu hình core-site

```
<property>
    <name>hadoop.tmp.dir</name>
    <value>${homefolder}/hadoop_data/hdfs/tmp</value>
    <description>A base for other temporary directories.</description>
</property>

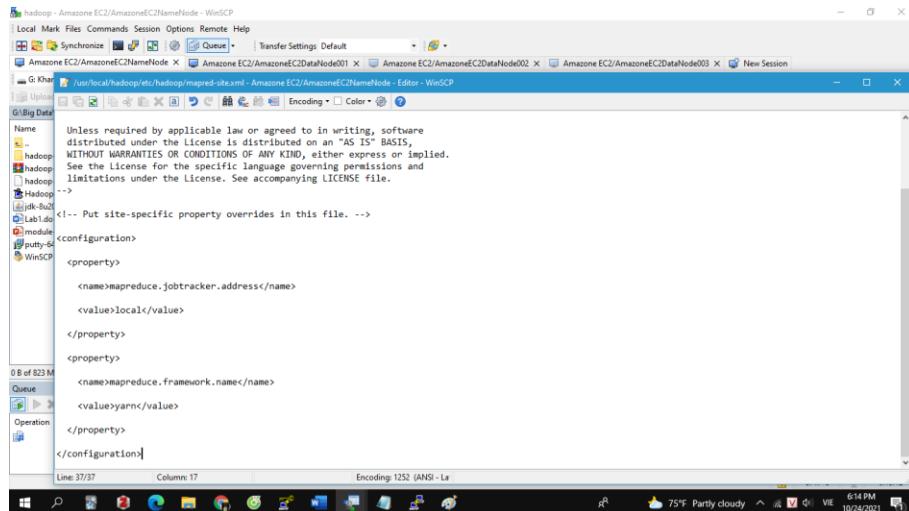
<property>
    <name>fs.defaultFS</name>
    <value>hdfs://$(thisnamenode):9000</value>
    <description>localhost may be replaced with a DNS that points to the NameNode.</description>
</property>

<property>
    <name>dfs.permissions</name>
    <value>false</value>
</property>
</configuration>
```

Thay đổi cấu hình yarn-site.xml

```
hadoop - Amazon EC2/AmazonEC2NameNode - WinSCP
Local Mark File Commands Session Options Remote Help
File Synchronize Queue Transfer Settings Default
Amazon EC2/AmazonEC2NameNode X Amazon EC2/AmazonEC2DataNode001 X Amazon EC2/AmazonEC2DataNode002 X Amazon EC2/AmazonEC2DataNode003 X New Session
G Khanh Tran > Up | Down | Properties New > Find Files
Big Data Lab 1 Properties New > Find Files
Name > /root/local/hadoop/etc/hadoop/yarn-site.xml - Amazon EC2/AmazonEC2NameNode - Editor - WinSCP
Encoding Color
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>mapred.job.tracker</name>
  <value>${thisnamenode}:9001</value>
</property>
<!--<property>
  <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property-->
</configuration>
```

Edit cấu hình cho mapred-site.xml



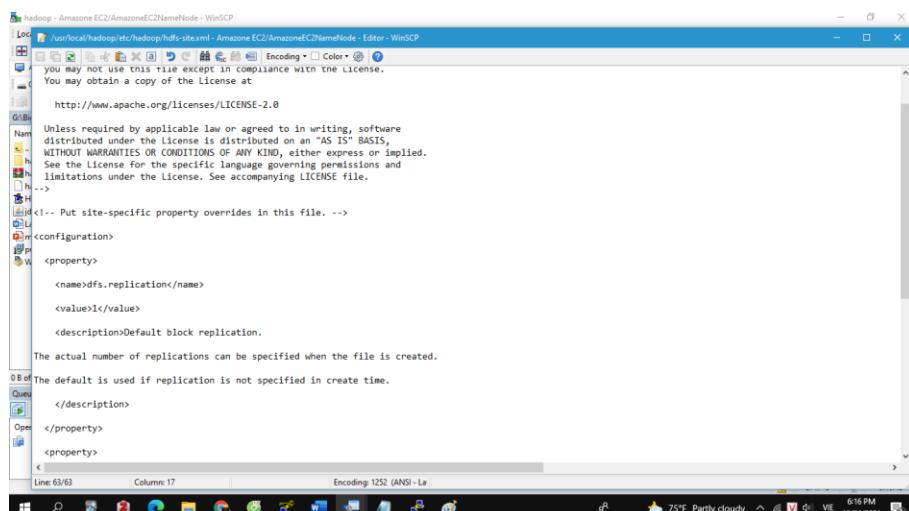
```

hadoop - Amazone EC2/AmazoneEC2NameNode - WinSCP
Local Mark Files Commands Session Options Remote Help
Synchronize Queue Transfer Settings Default
Amazone EC2/AmazoneEC2NameNode X Amazone EC2/AmazoneEC2DataNode01 X Amazone EC2/AmazoneEC2DataNode02 X Amazone EC2/AmazoneEC2DataNode03 X New Session
G:\Big Data
Name
hadoop
hadoop
hadoop
hadoop
-->
<!-- Put site-specific property overrides in this file. -->
<configuration>
<property>
<name>mapreduce.jobtracker.address</name>
<value>local</value>
</property>
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
</configuration>

```

Line: 37/37 Columns: 17 Encoding: 1252 (ANSI - La)

Chỉnh cấu hình hdfs-site.xml



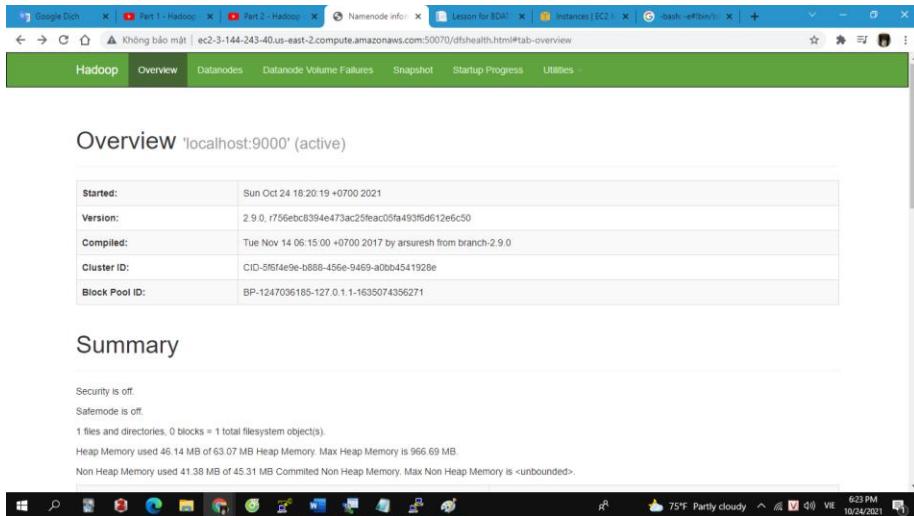
```

hadoop - Amazone EC2/AmazoneEC2NameNode - WinSCP
Loc /usr/local/hadoop/etc/hadoop/hdfs-site.xml - Amazone EC2/AmazoneEC2NameNode - Editor - WinSCP
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>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>
<property>
</configuration>

```

Line: 63/63 Columns: 17 Encoding: 1252 (ANSI - La)

Kiểm tra Namenode



Đoạn code dưới sẽ giúp mọi node sẽ có cùng cấu hình

```
cd $HADOOP_CONF_DIR
```

```
scp hadoop-env.sh core-site.xml hdfs-site.xml mapred-site.xml DataNode001:$HADOOP_CONF_DIR  
scp hadoop-env.sh core-site.xml hdfs-site.xml mapred-site.xml DataNode002:$HADOOP_CONF_DIR  
scp hadoop-env.sh core-site.xml hdfs-site.xml mapred-site.xml
```

DataNode003:\$HADOOP_CONF_DIR

```
ubuntu@ec2-3-144-243-40:~/local/hadoop/etc/hadoop$ 
scp hadoop-env.sh core-site.xml mapred-site.xml DataNode003:$HADOOP_HOME
hadoop-env.sh                                         100%  4710   3.9MB/s  0:00
core-site.xml                                       100% 1599   1.6MB/s  0:00
mapred-site.xml                                     100% 1137   1.0MB/s  0:00
mapred-site.xml                                     100% 956   1.0MB/s  0:00
ubuntu@ec2-3-144-243-40:/usr/local/hadoop/etc/hadoop$ scp hadoop-env.sh core-site.xml
mapred-site.xml                                     100% 956   1.0MB/s  0:00
ubuntu@ec2-3-144-243-40:/usr/local/hadoop/etc/hadoop$ rm $HADOOP_CONF_DIR/core-site.xml
mapred-site.xml                                     100% 956   1.0MB/s  0:00
ubuntu@ec2-3-144-243-40:/usr/local/hadoop/etc/hadoop$ cp $HADOOP_CONF_DIR/core-site.xml
core-site.xml                                       100% 1599   1.6MB/s  0:00
mapred-site.xml                                     100% 1137   1.0MB/s  0:00
mapred-site.xml                                     100% 956   1.0MB/s  0:00
ubuntu@ec2-3-144-243-40:/usr/local/hadoop/etc/hadoop$ rm $HADOOP_CONF_DIR/mapred-site.xml
mapred-site.xml                                     100% 956   1.0MB/s  0:00
ubuntu@ec2-3-144-243-40:/usr/local/hadoop/etc/hadoop$ 
ubuntu@ec2-3-144-243-40:~/local/hadoop$ 

97 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

Failed to connect to https://changelogs.ubuntu.com/meta-release-lts. Check your Internet connection or proxy settings

Last login: Sun Oct 24 10:29:26 2021 from 1.53.144.9
ubuntu@ec2-3-144-199-136:~$ echo $HADOOP_HOME
/usr/local/hadoop
ubuntu@ec2-3-144-199-136:~$ 
ubuntu@ec2-3-144-199-136:~$ mkdir -p $HADOOP_DATA_HOME/datanode
ubuntu@ec2-3-144-199-136:~$ mkdir -p $HADOOP_DATA_HOME/namenode
ubuntu@ec2-3-144-199-136:~$ cp $HADOOP_CONF_DIR/core-site.xml.template $HADOOP_CONF_DIR/datanode-site.xml
ubuntu@ec2-3-144-199-136:~$ cp $HADOOP_CONF_DIR/mapred-site.xml.template $HADOOP_CONF_DIR/mapred-site.xml
ubuntu@ec2-3-144-199-136:~$ 
ubuntu@ec2-3-144-199-136:~$ 

97 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

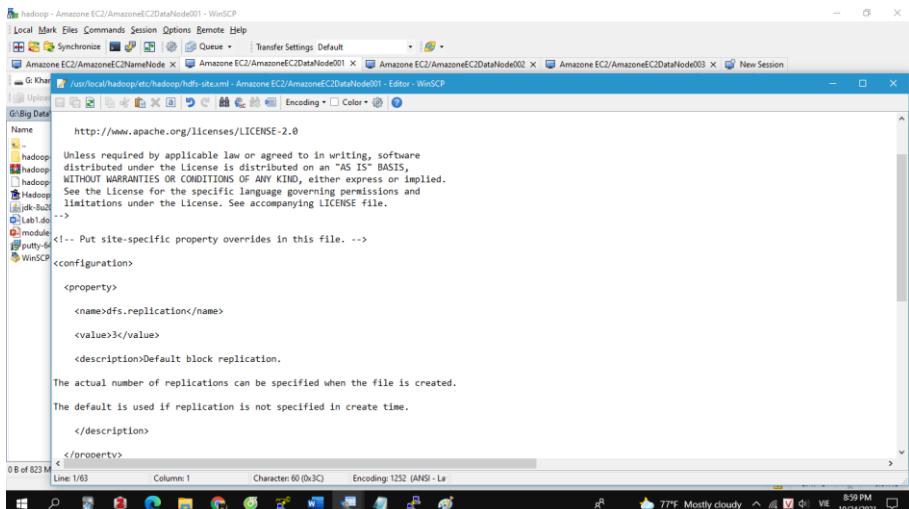
Failed to connect to https://changelogs.ubuntu.com/meta-release-lts. Check your Internet connection or proxy settings

Last login: Sun Oct 24 10:29:05 2021 from 1.53.144.9
ubuntu@ec2-3-144-243-66:~$ echo $HADOOP_HOME
/usr/local/hadoop
ubuntu@ec2-3-144-243-66:~$ 
ubuntu@ec2-3-144-243-66:~$ mkdir -p $HADOOP_DATA_HOME/datanode
ubuntu@ec2-3-144-243-66:~$ mkdir -p $HADOOP_DATA_HOME/namenode
ubuntu@ec2-3-144-243-66:~$ cp $HADOOP_CONF_DIR/core-site.xml.template $HADOOP_CONF_DIR/datanode-site.xml
ubuntu@ec2-3-144-243-66:~$ cp $HADOOP_CONF_DIR/mapred-site.xml.template $HADOOP_CONF_DIR/mapred-site.xml
ubuntu@ec2-3-144-243-66:~$ 
ubuntu@ec2-3-144-243-66:~$ 

975 Party cloudy ▲ 🔍 🌡 VE 6299 PM
10/24/2021
```

Format HDFS cho các cụm còn lại

Cấu hình ba file core-site.xml, hdfs-site.xml, yarn-site.xml giống nhau



Đoạn code sau đây tao master file trong Namenode

```
sudo rm -rf $HADOOP_CONF_DIR/masters
```

```
echo -e "hadoop-master" | sudo tee --append $HADOOP_CONF_DIR/masters > /dev/null
```

```
echo -e "DataNode001" | sudo tee --append $HADOOP_CONF_DIR/masters > /dev/null
```

```

ubuntu@ec2-3-14-243-40:~$ scp /usr/local/hadoop/etc/hadoop/* core-site.xml hdfs-site.xml mapred-site.xml DataNode001:/HADOOP_CONF_DIR
100% 4710 2.4MB/s 00:00
100% 1659 1.6MB/s 00:00
100% 1659 1.6MB/s 00:00
100% 956 853.9KB/s 00:00
ubuntu@ec2-3-14-243-40:~$ scp hadoop-env.sh core-site.xml hdfs-site.xml mapred-site.xml DataNode002:/HADOOP_CONF_DIR
100% 710 1.6MB/s 00:00
100% 1659 1.5MB/s 00:00
100% 1717 1.4MB/s 00:00
100% 956 957.7KB/s 00:00
ubuntu@ec2-3-14-243-40:~$ sudo chmod 755 HADOOP_CONF_DIR/masters
ubuntu@ec2-3-14-243-40:~$ /usr/local/hadoop/etc/hadoop/* > /dev/null
ubuntu@ec2-3-14-243-40:~$ echo -e "DataNode001" | sudo tee --append HADOOP_CONF_DIR/masters > /dev/null
ubuntu@ec2-3-14-243-40:~$ /usr/local/hadoop/etc/hadoop/*

```

Tạo DataNode1 làm Secondary Namenode

Local File Commands Session Options Remote Help

Synchronize Queue Transfer Settings: Default

Amazone EC2/AmazonE2NameNode X Amazone EC2/AmazonE2DataNode001 X Amazone EC2/AmazonE2DataNode002 X Amazone EC2/AmazonE2DataNode003 X New Session

G: K:\ /usr/local/hadoop/etc/hadoop/masters : Amazone EC2/AmazonE2DataNode001 - Editor - WinSCP

Uploading DataNode001

0 B of 823 M Line: 1/2 Column: 1 Character: 104 (0x68) Encoding: 1252 (ANSI - La)

Above: 77°F Mostly cloudy ^ VIE 10/24/2021

Tạo slave file

Local File Commands Session Options Remote Help

Synchronize Queue Transfer Settings: Default

Amazone EC2/AmazonE2NameNode X Amazone EC2/AmazonE2DataNode001 X Amazone EC2/AmazonE2DataNode002 X Amazone EC2/AmazonE2DataNode003 X New Session

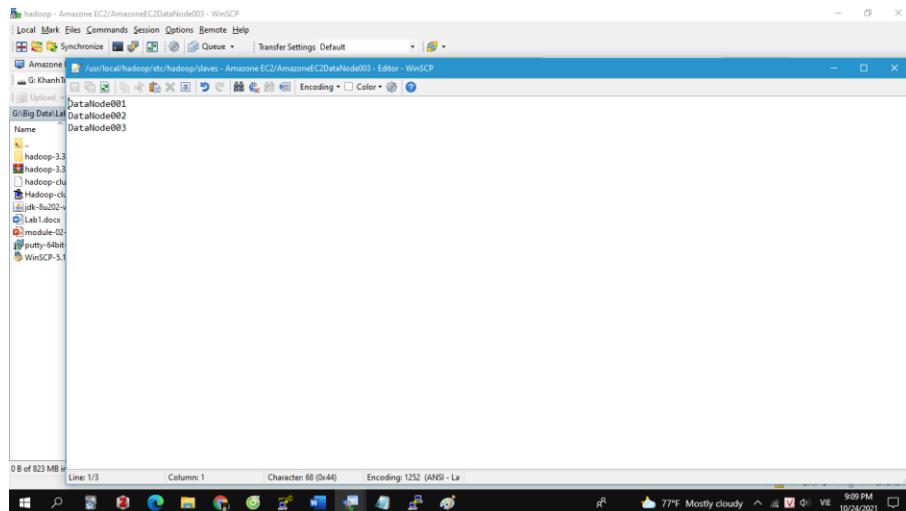
G: K:\ /usr/local/hadoop/etc/hadoop/slaves : Amazone EC2/AmazonE2NameNode - Editor - WinSCP

Uploading DataNode002

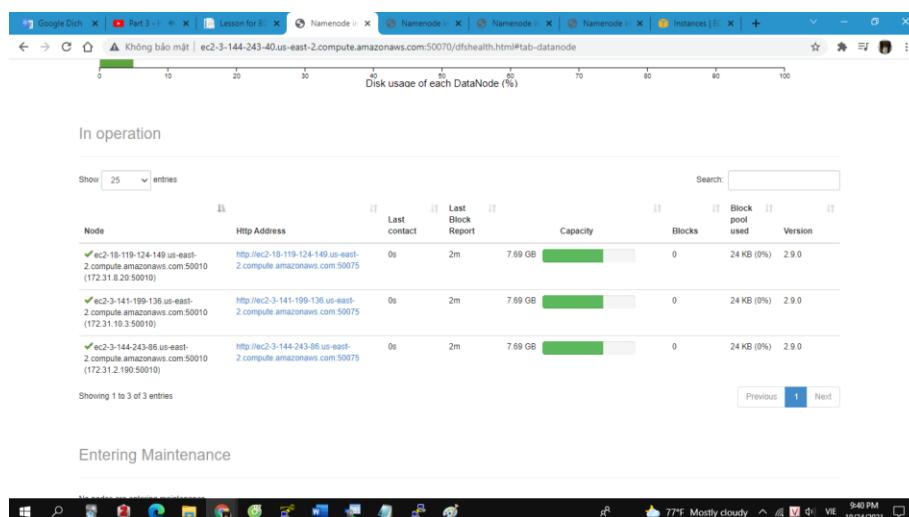
0 B of 823 M Line: 1/3 Column: 1 Character: 68 (0x44) Encoding: 1252 (ANSI - La)

Above: 77°F Mostly cloudy ^ VIE 10/24/2021

Sao chép tệp slaves vào từng Node trong cụm của bạn



Hoàn tất tạo một cụm 4 nodes, với 1 node NameNode và 3 DataNode, trong đó DataNode001 là Secondary NameNode



DataNode on ec2-3-141-199-136.us-east-2.compute.amazonaws.com:50010

Cluster ID:	CID-dd5c2af2-c10d-4e75-89c1-58a168f46169
Version:	2.9.0

Block Pools

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
hadoop-master 9000	BP-153416067-127.0.1.1-1635086227443	RUNNING	0s	5 minutes	0 B (64 MB)

Volume Information

Directory	StorageType	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
/home/ubuntu/hadoop_data/hdfs/datanode/current	DISK	24 KB	2.7 GB	0 B	0 B	0

Hadoop, 2017.

DataNode on ec2-18-119-124-149.us-east-2.compute.amazonaws.com:50010

Cluster ID:	CID-dd5c2af2-c10d-4e75-89c1-58a168f46169
Version:	2.9.0

Block Pools

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
hadoop-master 9000	BP-153416067-127.0.1.1-1635086227443	RUNNING	0s	5 minutes	0 B (64 MB)

Volume Information

Directory	StorageType	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
/home/ubuntu/hadoop_data/hdfs/datanode/current	DISK	24 KB	2.7 GB	0 B	0 B	0

Hadoop, 2017.

DataNode on ec2-3-144-243-86.us-east-2.compute.amazonaws.com:50010

Cluster ID:	CID-dd5c2af2-c10d-4e75-89c1-58a168f46169
Version:	2.9.0

Block Pools

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
hadoop-master 9000	BP-153416067-127.0.1.1-1635086227443	RUNNING	2s	5 minutes	0 B (64 MB)

Volume Information

Directory	StorageType	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
/home/ubuntu/hadoop_data/hdfs/datanode/current	DISK	24 KB	2.7 GB	0 B	0 B	0

Hadoop, 2017.

References:

https://klasserom.azurewebsites.net/Lessons/Binder/1960#CourseStrand_3400

https://www.youtube.com/watch?v=XkDhf1pwPtA&list=PLWsYJ2ygHmWjPsg-6MnQO6WxVWF18OzK_&index=2

<https://us-east-2.console.aws.amazon.com/ec2/v2/home?region=us-east-2#Instances:sort=desc:tag:Name>