Multi Node Cluster Installation

# 1. Virtual Machine Setup

- Change Root Password

|  |
| --- |
| sudo passwd |



- Reboot Server

|  |
| --- |
| reboot |



## Java

### 1.1.1 Download java

|  |
| --- |
| apt update |
| reboot |
| apt install openjdk-8-jdk |



## SSH

### 1.2.1 Download SSH

|  |
| --- |
| apt-get install ssh |
| apt install openssh-server |
| reboot |

A screen shot of a computer

Description automatically generated



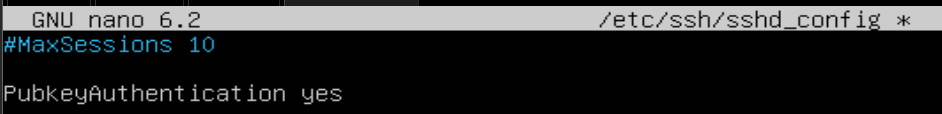


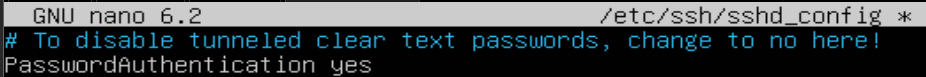
### 1.2.1 Configuration SSH

|  |
| --- |
| nano /etc/ssh/sshd\_config |

- Find # PubkeyAuthentication yes and # PasswordAuthentication yes.

- Uncomment # to PubkeyAuthentication yes andPasswordAuthentication yes ...





- Restart SSH

|  |
| --- |
| service sshd restart |



# 2. Hadoop

## 2.1 Create user hadoopngocphung

|  |
| --- |
| adduser hadoopngocphung |

A screenshot of a computer program

Description automatically generated

## 2.2 Install Hadoop 3.3.6

|  |
| --- |
| su hadoopngocphung |
| wget https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz |
| tar -xzf hadoop-3.3.6.tar.gz |
| mv hadoop-3.3.6 hadoop |
| mv hadoop-3.3.6.tar.gz tools/hadoop/ |
| mv hadoop-3.3.6/ tools/hadoop/ |

A screen shot of a computer

Description automatically generated

- Add the following content to the end of the .bashrc file:

|  |
| --- |
| nano ~/.bashrc |

|  |
| --- |
| # JAVA  export JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64  export PATH=$PATH:$JAVA\_HOME/bin |
| source ~/.bashrc |

## 2.3 Standalone Operation

- By default, Hadoop is configured to run in non-distributed mode as a single Java process. The following example copies all .xml files in the /etc/hadoop directory to the input directory, then finds and displays all results that match the given regular expression.

|  |
| --- |
| mkdir input |

A screen shot of a computer

Description automatically generated

- Copy all the files ending in ".xml" from the "hadoop/etc/hadoop/" directory and place them in the "input" directory.

|  |
| --- |
| cp hadoop/etc/hadoop/\*.xml input |

A screen shot of a computer

Description automatically generated

|  |
| --- |
| tools/hadoop/hadoop/bin/hadoop jar tools/hadoop/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.6.jar grep tools/hadoop/input tools/hadoop/output 'dfs[a-z.]+' |

- Used to run a MapReduce application in Apache Hadoop

A screenshot of a computer program

Description automatically generated

- display the contents of all files in the "output" directory

|  |
| --- |
| cat output/\* |



## 2.4 Create ssh key

- Creatae ssh key

|  |
| --- |
| ssh-keygen -t rsa -P "" |

**A computer screen with white text

Description automatically generated**

|  |
| --- |
| cat /home/hadoopngocphung/.ssh/id\_rsa.pub >> /home/hadoopngocphung/.ssh/authorized\_keys |

- Add the contents of the "id\_rsa.pub" file to the end of the "authorized\_keys" file

|  |
| --- |
| chmod 600 /home/hadoopngocphung/.ssh/authorized\_keys |

- Change the permissions of the "authorized\_keys" file in the .ssh directory of the user "hadoopngocphung"

# 3. Configuring files for Hadoop

## 3.1 File .bashrc

|  |
| --- |
| nano ~/.bashrc |

- Add the following content to the end of the .bashrc file:

|  |
| --- |
| # JAVA  export JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64  export PATH=$PATH:$JAVA\_HOME/bin  # HADOOP  export HADOOP\_HOME=/home/hadoopngocphung/tools//hadoop/hadoop  export PATH=$PATH:$HADOOP\_HOME/bin  export PATH=$PATH:$HADOOP\_HOME/sbin  export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME  export HADOOP\_COMMON\_HOME=$HADOOP\_HOME  export HADOOP\_HDFS\_HOME=$HADOOP\_HOME  export HADOOP\_CONF\_DIR=$HADOOP\_HOME/etc/hadoop  export HADOOP\_YARN\_HOME=$HADOOP\_HOME  export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native  export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_HOME/lib/native" |

A screen shot of a computer program

Description automatically generated

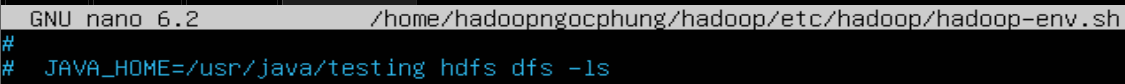
|  |
| --- |
| source ~/.bashrc |

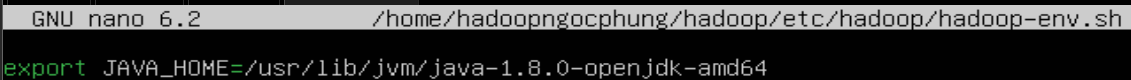
## 3.2 File hadoop-env.sh

|  |
| --- |
| nano ~/tools/hadoop/hadoop/etc/hadoop/hadoop-env.sh |

- find export JAVA\_HOME=... change to java location:

|  |
| --- |
| export JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64 |





## 3.3 File core-site.xml

|  |
| --- |
| nano ~/tools/hadoop/hadoop/etc/hadoop/core-site.xml |

- configure data in file

|  |
| --- |
| <configuration>  <property>  <name>hadoop.tmp.dir </name>  <value>/home/hadoopngocphung/tools/hadoop/tmp</value>  </property>  <property>  <name>fs.defaultFS</name>  <value>hdfs://ngocphung-master:9000</value>  </property>  </configuration> |

A computer screen with yellow text

Description automatically generated

## 3.4 File mapred-site (chỉ cấu hình ở master)

|  |
| --- |
| nano ~/tools/hadoop/hadoop/etc/hadoop/mapred-site.xml |
| <property>  <name>mapreduce.application.classpath</name> <value>$HADOOP\_MAPRED\_HOME/share/hadoop/mapreduce/\*:$HADOOP\_MAPRED\_HOME/share/hadoop/mapreduce/lib/\*</value>  </property>  <property>  <name>mapreduce.jobtracker.address</name>  <value>ngocphung-master:9001</value>  </property>  <property>  <name>mapreduce.framework.name</name>  <value>yarn</value>  </property>  <property>  <name>yarn.app.mapreduce.am.env</name>  <value>HADOOP\_MAPRED\_HOME=/home/hadoopngocphung/tools/hadoop/hadoop</value>  </property>  <property>  <name>mapreduce.map.env</name>  <value>HADOOP\_MAPRED\_HOME=/home/hadoopngocphung/tools/hadoop/hadoop</value>  </property>  <property>  <name>mapreduce.reduce.env</name>  <value>HADOOP\_MAPRED\_HOME=/home/hadoopngocphung/tools/hadoop/hadoop</value>  </property> |

A screenshot of a computer screen

Description automatically generated

## 3.5 File hdfs-site.xml

|  |
| --- |
| nano ~/tools/hadoop/hadoop/etc/hadoop/hdfs-site.xml |

- configure data in file

|  |
| --- |
| <property>  <name>dfs.replication</name>  <value>2</value>  </property>  <property>  <name>dfs.namenode.name.dir</name> <value>/home/hadoopngocphung/tools/hadoop/hadoop/hadoop\_data/hdfs/namenode</value>  </property>  <property>  <name>dfs.datanode.data.dir</name> <value>/home/hadoopngocphung/tools/hadoop/hadoop/hadoop\_data/hdfs/datanode</value>  </property> |

A screenshot of a computer

Description automatically generated

## 3.6 File yarn-site.xml

|  |
| --- |
| nano ~/tools/hadoop/hadoop/etc/hadoop/yarn-site.xml |

- configure data in file

|  |
| --- |
| <property>  <name>yarn.nodemanager.aux-services</name>  <value>mapreduce\_shuffle</value>  </property>  <property>  <name>yarn.nodemanager.env-whitelist</name>  <value>JAVA\_HOME,HADOOP\_COMMON\_HOME,HADOOP\_HDFS\_HOME,HADOOP\_CONF\_DIR,CLASSPATH\_PREPEND\_DISTCACHE,HADOOP\_YARN\_HOME,HADOOP\_MAPRED\_HOME </value>  </property>  <property>  <name>yarn.resourcemanager.scheduler.address</name>  <value>ngocphung-master:9002</value>  </property>  <property>  <name>yarn.resourcemanager.address</name>  <value>ngocphung-master:9003</value>  </property>  <property>  <name>yarn.resourcemanager.webapp.address</name>  <value>ngocphung-master:9004</value>  </property>  <property>  <name>yarn.resourcemanager.resource-tracker.address</name>  <value>ngocphung-master:9005</value>  </property>  <property>  <name>yarn.resourcemanager.admin.address</name>  <value>ngocphung-master:9006</value>  </property> |

A screenshot of a computer program

Description automatically generated

# 4. Configuring network

## 4.1 Configuring for MASTER

- Install Open vSwitch software on the system

|  |
| --- |
| apt-get install openvswitch-switch |

### 4.1.1 Configuring hostname for master

|  |
| --- |
| nano /etc/hostname |

- In this file will appear the default hostname of the machine, delete it and change it to ngocphung-master

### 4.1.2 Set up static host IP for MASTER

|  |
| --- |
| networkctl |

**A screen shot of a black background

Description automatically generated**

|  |
| --- |
| networkctl status |

**A screenshot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generated

|  |
| --- |
| nano /etc/netplan/00-installer-config.yaml |
| sudo chmod 600 /etc/netplan/00-installer-config.yaml |
| network:  ethernets:  ens33:  dhcp4: false  dhcp6: false  addresses: [192.168.19.1/24]  routes:  - to: default  via: 192.168.19.2  nameservers:  addresses: [192.168.19.1, 8.8.8.8, 8.8.4.4]  version: 2 |

A screenshot of a computer

Description automatically generated

|  |
| --- |
| netplan apply |



- Hệ thống đã được cấu hình theo IP mới, để kiểm tra chạy 1 trong 2 lệnh sau

|  |
| --- |
| networkctl status |

A screenshot of a computer

Description automatically generated

### 4.1.3 Configuring host

|  |
| --- |
| nano /etc/hosts |

- Fill in the content, each slave machine is a line

|  |
| --- |
| 192.168.19.1 ngocphung-master  192.168.19.3 ngocphung-slave1 |

A black and white screen with blue text

Description automatically generated

### 4.1.4 List slave machines (only done on master machine)

|  |
| --- |
| su hadoopngocphung |
| nano ~/tools//hadoop/hadoop/etc/hadoop/workers |

- Fill in the names of slave1 machines****

****

## 4.2 Configuring for SLAVE1

### 4.2.1 Create ngocphung-slave1

- Turn off Master.

- Copy Master, rename to Slave1

A screenshot of a black box

Description automatically generated

- Turn on the slave machine, adjust the static IP and parameters accordingly: hosts, hostname...

### 4.2.2 Set up hostname for slave



|  |
| --- |
| nano /etc/hostname |

- In this file will appear the default hostname of the machine, delete it and change it to ngocphung-master

### 4.2.3 Set up static host IP for SLAVE1

|  |
| --- |
| nano /etc/netplan/00-installer-config.yaml |
| network:  ethernets:  ens33:  dhcp4: false  dhcp6: false  addresses: [192.168.19.3/24]  routes:  - to: default  via: 192.168.19.2  nameservers:  addresses: [192.168.19.1, 8.8.8.8, 8.8.4.4]  version: 2 |

A screenshot of a computer

Description automatically generated

|  |
| --- |
| netplan apply |



- The system has been configured with the new IP, to check run 1 of the following 2 commands

|  |
| --- |
| networkctl status |

A screenshot of a computer

Description automatically generated

## 5. Set up ssh key between nodes

## 5.1 Share key

This operation is only performed on master

- Login with hadoopngocphung



- Create ssh key

|  |
| --- |
| ssh-keygen -t rsa -P "" |

A screenshot of a computer

Description automatically generated

|  |
| --- |
| cat /home/hadoopngocphung/.ssh/id\_rsa.pub >> /home/hadoopngocphung/.ssh/authorized\_keys |



|  |
| --- |
| chmod 600 /home/hadoopngocphung/.ssh/authorized\_keys |



- Share ssh key master - master

|  |
| --- |
| ssh-copy-id -i ~/.ssh/ id\_rsa.pub ngocphung-master |

- Share ssh key master – slave

|  |
| --- |
| ssh-copy-id -i ~/.ssh/id\_rsa.pub ngocphung-slave1 |

A screenshot of a computer program

Description automatically generated

## 5.2 Test connection ssh

- Test connection to master

|  |
| --- |
| ssh hadoopngocphung@ngocphung-master |

A screenshot of a computer

Description automatically generated

|  |
| --- |
| logout |

- Test connection to slave

A screenshot of a computer

Description automatically generated

|  |
| --- |
| ssh hadoopngocphung@ngocphung-slave1 |

# 6. Format namenode

This operation is only performed on the master and only done once.

Update the master configuration information

|  |
| --- |
| hadoop/bin/hdfs namenode -format |



# 7. Create file test

- Go to root directory, create test file

|  |
| --- |
| nano test.sh |

- Contains the following contents

|  |
| --- |
| #!/bin/bash  # test the hadoop cluster by running wordcount  # create input files  mkdir input  echo "Hello World" >input/file1.txt  echo "Hello Hadoop" >input/file2.txt  # create input directory on HDFS  hadoop fs -mkdir -p input1  # put input files to HDFS  hdfs dfs -put ./input/\* input1  # run wordcount  hadoop jar $HADOOP\_HOME/share/hadoop/mapreduce/sources/hadoop-mapreduce-examples-3.3.6-sources.jar org.apache.hadoop.examples.WordCount input1 output1  # print the input files  echo -e "\ninput file1.txt:"  hdfs dfs -cat input1/file1.txt  echo -e "\ninput file2.txt:"  hdfs dfs -cat input1/file2.txt  # print the output of wordcount  echo -e "\nwordcount output:"  hdfs dfs -cat output1/part-r-00000 |

- Authorize and test run

|  |
| --- |
| chmod +x test.sh |
| ./test.sh |

- if wanna run again

|  |
| --- |
| rm -rf input |
| hadoop fs -rm -r input1 |
| hadoop fs -rm -r output1 |

A screenshot of a computer screen

Description automatically generatedA screen shot of a computer

Description automatically generated