su root

1. OpenJDK

apt update

apt install default-jdk

2. SSH

apt-get install ssh

apt install openssh-server

vim /etc/ssh/sshd\_config

bấm chữ i để edit file nè

- Tìm đoạn # PubkeyAuthentication yes. Bỏ dấu # phía trước

- Tìm đoạn # PasswordAuthentication yes. Bỏ dấu # phía trước

xong rùi thì ESC, : rùi nhập wq.

**service sshd restart**

3. Tạo user hadoop

adduser hadoopngoctrang

4. Cài đặt Hadoop 3.3.6

su hadoopngoctrang

cd

cd /home/hadoopngoctrang

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

5. Thiết lập Java\_Home

vim ~/hadoop/etc/hadoop/hadoop-env.sh

bấm chữ i để edit file nè

- Tìm đoạn #export JAVA\_HOME= thêm này zo: /usr/lib/jvm/java-1.11.0-openjdk-amd64

Nhớ bỏ dấu # rùi ESC :wq

\*\*Thiết lập singlenode cluster

6. Standalone Operation

mkdir input

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

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

cat output/\*

7. Pseudo

7.1. Cài ssh key

ssh-keygen -t rsa -P ""

Enter cái nữa

cat /home/hadoopngoctrang/.ssh/id\_rsa.pub >> /home/hadoopngoctrang/.ssh/authorized\_keys

chmod 600 /home/hadoopngoctrang/.ssh/authorized\_keys

7.2. Cấu hình core-site.xml

vim ~/hadoop/etc/hadoop/core-site.xml

bấm chữ i để edit file nè

tìm xún chỗ <configuration> đồ á

sửa thành zị

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

7.3. Cấu hình hdfs-site.xml

vim ~/hadoop/etc/hadoop/hdfs-site.xml

bấm chữ i để edit file nè

tìm xún chỗ <configuration> đồ á

sửa thành zị

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

7.4.

hadoop/bin/hdfs namenode -format

7.5.

hadoop/sbin/start-dfs.sh

jps

~/hadoop/bin/hdfs dfsadmin -report

~/hadoop/bin/hdfs dfs -mkdir /user

~/hadoop/bin/hdfs dfs -mkdir /user/hadoopngoctrang $ ~/hadoop/bin/hdfs dfs -mkdir /user/hadoopngoctrang/input

~/hadoop/bin/hdfs dfs -put hadoop/etc/hadoop/\*.xml /user/hadoopngoctrang/input

~/hadoop/bin/hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.6.jar grep input output ‘dfs[a-z.]+’

~/hadoop/bin/hdfs dfs -get output/ output

cat output/\*

hadoop/sbin/stop-dfs.sh

7.6.

vim ~/.bashrc

bấm chữ i để edit file nè

xong lướt xuống dưới cùng á rùi add mấy này:

export JAVA\_HOME=/usr/lib/jvm/java-1.11.0-openjdk-amd64

export HADOOP\_HOME=/home/hadoopngoctrang/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"

7.7.

vim hadoop/etc/hadoop/mapred-site.xml

bấm chữ i để edit file nè

tìm xún chỗ <configuration> đồ á

sửa thành zị

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

<property>

<name>mapreduce.application.classpath</name>

<value>$HADOOP\_MAPRED\_HOME/share/hadoop/mapreduce/\*:$HADOOP\_MAPRED\_HOME/share/hadoop/mapreduce/lib/\*</value>

</property>

</configuration>

7.8.

vim hadoop/etc/hadoop/yarn-site.xml

bấm chữ i để edit file nè

tìm xún chỗ <configuration> đồ á

sửa thành zị

<configuration>

<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>

</configuration>

7.9

hadoop/sbin/start-dfs.sh

hadoop/sbin/start-yarn.sh

7.10..

rm -rf output

~/hadoop/bin/hdfs dfs -rm -r /user

~/hadoop/bin/hdfs dfs -mkdir /user

~/hadoop/bin/hdfs dfs -mkdir /user/hadoopngoctrang

~/hadoop/bin/hdfs dfs -mkdir /user/hadoopngoctrang/input

~/hadoop/bin/hdfs dfs -put hadoop/etc/hadoop/\*.xml /user/hadoopngoctrang/input

~/hadoop/bin/hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.6.jar grep input output ‘dfs[a-z.]+’

~/hadoop/bin/hdfs dfs -get output/ output

cat output/\*

khi muốn dừng …..

hadoop/sbin/stop-dfs.sh

hadoop/sbin/stop-yarn.sh

eror: connection request

…

1: telnet localhost

2. Hadoop namenode format

(stop-all.sh trước)