**HADOOP**

**1.** **Triển khai Hadoop single node Ubuntu 20.04.4:**

root

· **Install tools:**

- **openjdk 11:**

apt install openjdk-11-jdk -y

- **openssh-server, openssh-client:**

apt install openssh-server openssh-client -y

· **Thiết lập user:**

- **Tạo cặp khóa SSH và chọn vị trí lưu:**

ssh-keygen -t rsa -P '' -f ~/.ssh/id\_rsa

- **Lưu public key vào authorized\_keys trong thư mục của SSH:**

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

- **Phân quyền cho người dùng:**

chmod 0600 ~/.ssh/authorized\_keys

- **Kiểm tra bằng cách ssh đến localhost:**

ssh localhost

· **Cài đặt Hadoop single node:**

- **Tải file hadoop-3.2.3.zip đã cấu hình về:**

wget -O /opt/hadoop-3.2.3.zip<https://www.dropbox.com/s/5z2k9wrkwj1ara6/hadoop-3.2.3.zip?dl=1>

https://tinyurl.com/twr3ne77

unzip /opt/hadoop-3.2.3.zip -d /opt

- **Cấu hình biến môi trường Hadoop (file .bashrc):**

+ Tải file bashrc.txt:

https://www.dropbox.com/s/rk28mgb85ybe9se/.bashrc?dl=0

[https://tinyurl.com/bde593eu .bashrc](https://tinyurl.com/54dka4hc%20.bashrc)

wget -O bashrc.txt<https://www.dropbox.com/s/quz1usuejvtmjr8/bashrc.txt?dl=1>

https://tinyurl.com/bdzka5eu

+ Copy vào file.bashrc:

cat bashrc.txt > ~/.bashrc

+ Áp dụng thay đổi:

source ~/.bashrc

+ Kiểm tra xem java và hadoop đã được cài đúng trên hệ thống và có thể được truy cập qua terminal không:

java -version

hadoop version

· **Triển khai Hadoop:**

- **Đến thư mục hadoop-3.2.3/sbin và khởi động tất cả daemons:**

pushd $HADOOP\_HOME/sbin && ./start-all.sh

- **Kiểm tra các trình daemon đang chạy:**

jps

**2.** **Chạy file jar trên hadoop:**

root

- **Tải file word-count-hadoop.zip đã cấu hình:**

wget -O word-count-hadoop.zip<https://www.dropbox.com/s/tm3vfnarwrewbie/word-count-hadoop.zip?dl=1>

https://tinyurl.com/2p9cam9x

unzip word-count-hadoop.zip

- **Tạo file jar với Maven.**

- **Chạy file jar trên Hadoop:**

+ Khởi động Hadoop.

+ Đẩy file input lên HDFS:

hdfs dfs -copyFromLocal word-count-hadoop/input.txt /

https://tinyurl.com/3nw6rrmz

+ Kiểm tra file trong HDFS:

hdfs dfs -ls /

+ Chạy file jar:

hadoop jar word-count-hadoop/target/wordcount-V1.jar com.hadoop.mapreduce.WordCount hdfs://localhost:9001/input.txt hdfs://localhost:9001/output

<https://www.dropbox.com/s/ffjcvsn6uhop22q/run_hadoop.sh?dl=0>

https://tinyurl.com/3pvzpsh9

run\_hadoop.sh

wordcount-V1.jar chuyen ra ngoai

- **Kiểm tra file trên HDFS:**

hdfs dfs -ls /

- **Kiểm tra kết quả trong folder input:**

hdfs dfs -cat /output/part-r-00000

class prepare{

# execute 'apt-get update'

exec {'apt-update':

command => '/usr/bin/apt-get update',

}

package { "openjdk-11-jdk": ensure => installed }

package { "openssh-server": ensure => installed }

package { "openssh-client": ensure => installed }

}

class tool{

package { 'unzip': ensure => installed }

}

class execute1{

exec{'create\_key':

path => ['/usr/bin', '/usr/sbin',],

command => '/bin/ssh-keygen -t rsa -P "" -f .ssh/id\_rsa',

}

exec{'save\_publickey':

path => ['/usr/bin', '/usr/sbin',],

command => '/bin/cat .ssh/id\_rsa.pub >> .ssh/authorized\_keys',

}

exec{'privilege':

path => ['/usr/bin', '/usr/sbin',],

command => '/bin/chmod 0600 .ssh/authorized\_keys',

}

}

class download\_hadoop{

exec{'wget1':

path => ['/usr/bin', '/usr/sbin',],

command => '/bin/wget -O /opt/hadoop-3.2.3.zip https://tinyurl.com/twr3ne77',

}

exec{'unzip1':

path => ['/usr/bin', '/usr/sbin',],

command => '/bin/unzip /opt/hadoop-3.2.3.zip -d /opt',

}

}

class execute2{

exec{'rm':

path => ['/usr/bin', '/usr/sbin',],

command => '/bin/rm -r .bashrc',

}

exec{'wget2':

path => ['/usr/bin', '/usr/sbin',],

command => '/bin/wget -O .bashrc https://tinyurl.com/bde593eu',

}

exec{'bash':

path => ['/usr/bin', '/usr/sbin',],

command => '/bin/bash -c "source ~/.bashrc" ',

}

exec{'pushd2':

path => ['/usr/bin', '/usr/sbin',],

cwd => '/opt/hadoop-3.2.3/sbin/',

command => '/bin/bash start-all.sh',

timeout => 800,

}

}

class execute3{

exec{'wget3':

path => ['/usr/bin', '/usr/sbin',],

command => '/bin/wget -O word-count-hadoop.zip https://tinyurl.com/2p9cam9x',

}

exec{'unzip2':

path => ['/usr/bin', '/usr/sbin',],

command => '/bin/unzip word-count-hadoop.zip',

}

exec{'wget5':

path => ['/usr/bin', '/usr/sbin',],

command => 'wget -O pushfile\_hadoop.sh https://tinyurl.com/3nw6rrmz',

}

exec{'hdfs':

path => ['/usr/bin', '/usr/sbin',],

command => 'bash pushfile\_hadoop.sh',

}

exec{'wget4':

path => ['/usr/bin', '/usr/sbin',],

command => 'wget -O run\_hadoop.sh https://tinyurl.com/3pvzpsh9',

}

exec{'jar':

path => ['/usr/bin', '/usr/sbin',],

command => 'bash run\_hadoop.sh',

}

}

if $facts['os']['family'] == 'Debian'{

include prepare

include tool

include execute1

include download\_hadoop

include execute2

include execute3

}

Run file



