# Hadoop on 64 or 32 bit Ubuntu 14.04 (Single node cluster)

# Install Java

Cài đặt môi trường Java do Hadoop framework được viết trên nền Java

Giả sử cài đặt trên hệ điệu hành Ubuntu mới. Chạy tuần tự các lệnh sau:

sudo apt-get update

sudo apt-get install default-jdk

java -version

Nếu hiện thị ra kết quả sau là đúng: java version "1.7.0\_55" OpenJDK Runtime Environment

# Create user

Tạo user để chạy Hadoop

sudo addgroup hadoop //Tạo nhóm

sudo adduser --ingroup hadoop hduser //Tạo user: hduser cho vào nhóm

sudo adduser hduser sudo //Thêm nhóm quyền sudo cho user

Sau khi gõ lệnh này thì nhập các thông tin trên màn hình

# Install SSH

Hadoop dùng SSH để truy cập vào các node.

sudo apt-get install ssh

Kiểm tra sự tồn tại của ssh và sshd

which ssh 🡪 /usr/bin/ssh

$ which sshd 🡪 /usr/sbin/sshd

# Create and setup SSH certificates

Hadoop sử dụng SSH (để truy cập các nút của nó) mà thông thường sẽ yêu cầu người dùng nhập mật khẩu. Tuy nhiên, yêu cầu này có thể được loại bỏ bằng cách tạo và thiết lập chứng chỉ SSH sử dụng các lệnh sau đây. Nếu yêu cầu một tên tập tin chỉ để trống và nhấn phím Enter để tiếp tục.

su hduser

ssh-keygen -t rsa -P ""

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

Chúng ta vừa chuyển sang user: hduser và tạo ra một khóa SSH ủy quyền truy cập user: hduser. Hiện tại Hadoop chưa chạy không ổn định với Ipv6 vì vậy chúng ta sẽ vô hiệu hóa nó. Để làm điều này chúng ta chỉ cần mở file /ect/sysctl.conf và thêm những dòng sau vào cuối file:

net.ipv6.conf.all.disable\_ipv6 = 1  
net.ipv6.conf.default.disable\_ipv6 = 1  
net.ipv6.conf.lo.disable\_ipv6 = 1

Hoặc chạy trực tiếp các lệnh sau từ terminal

echo "net.ipv6.conf.all.disable\_ipv6 = 1" | sudo tee -a /etc/sysctl.conf  
echo "net.ipv6.conf.default.disable\_ipv6 = 1" | sudo tee -a /etc/sysctl.conf  
echo "net.ipv6.conf.lo.disable\_ipv6 = 1" | sudo tee -a /etc/sysctl.conf

Để đảm bảo loại bỏ Ipv6 chúng ta phải thực thi các lệnh sau:

sudo service networking restart

Để chắc ăn thì restart lại máy

sudo shutdown -r now

Kiểm tra lại việc cài đặt certificate SSH cho user hduser

ssh localhost

Nếu không phải nhập password là thành công.

# Install Hadoop

Download bằng trình duyệt với link sau: http://mirrors.sonic.net/apache/hadoop/common/hadoop-2.6.0/hadoop-2.6.0.tar.gz

Download bằng dòng lệnh

wget <http://mirrors.sonic.net/apache/hadoop/common/hadoop-2.6.0/hadoop-2.6.0.tar.gz>

tar xvzf hadoop-2.6.0.tar.gz //Giải nén

Di chuyển sang /usr/local/hadoop và phân quyền cho user: hduser vào thư mục hadoop ( giả sử đang tại thư mục hiện tại )

sudo mv hadoop-2.6.0 /usr/local/hadoop

sudo chown -R hduser:hadoop /usr/local/hadoop //Phân quyền

sudo chmod -R 777 /usr/local/hadoop //Thêm quyền sửa các file

# Setup configuration files

Các file chúng ta sẽ cần phải sửa:

* + ~/.bashrc
  + /usr/local/hadoop/etc/hadoop/hadoop-env.sh
  + /usr/local/hadoop/etc/hadoop/core-site.xml
  + /usr/local/hadoop/etc/hadoop/mapred-site.xml.template
  + /usr/local/hadoop/etc/hadoop/hdfs-site.xml

## File ~/.bashrc

su hduser

sudo gedit ~/.bashrc

Thêm vào cuối của file .bashrc đoạn sau:

#HADOOP VARIABLES START

export JAVA\_HOME=/usr/lib/jvm/java-7-openjdk-amd64

export HADOOP\_INSTALL=/usr/local/hadoop

export PATH=$PATH:$HADOOP\_INSTALL/bin

export PATH=$PATH:$HADOOP\_INSTALL/sbin

export HADOOP\_MAPRED\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_HOME=$HADOOP\_INSTALL

export HADOOP\_HDFS\_HOME=$HADOOP\_INSTALL

export YARN\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_INSTALL/lib/native export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_INSTALL/lib"

#HADOOP VARIABLES

## /usr/local/hadoop/etc/hadoop/hadoop-env.sh

Sửa JAVA\_HOME trong file hadoop-env.sh như sau:

export JAVA\_HOME=/usr/lib/jvm/java-7-openjdk-amd64

## /usr/local/hadoop/etc/hadoop/core-site.xml

Cài đặt lại thuộc tính trong file cấu hình. Tạo và phân quyền cho user hduser vào thư mục vừa tạo

su hduser

sudo mkdir -p /app/hadoop/tmp

sudo chown hduser:hadoop /app/hadoop/

Copy nội dung sau vào file core-site.xml (ghi đè vào thẻ configuration)

<configuration>

<property>

<name>**hadoop.tmp.dir**</name>

<value>**/app/hadoop/tmp**</value>

<description>**A base for other temporary directories.**</description>

</property>

<property>

<name>**fs.default.name**</name>

<value>**hdfs://localhost:54310**</value>

<description>**The name of the default file system... .**</description>

</property>

</configuration>

## /usr/local/hadoop/etc/hadoop/mapred-site.xml

Trong thư mục /usr/local/hadoop/etc/hadoop không có file mapred-site.xml chúng ta sẽ tạo ra file này từ file file mapred-site.xml.template

cp /usr/local/hadoop/etc/hadoop/mapred-site.xml.template /usr/local/hadoop/etc/hadoop/mapred-site.xml

Copy nội dung sau vào file mapred-site.xml (ghi đè vào thẻ configuration)

<configuration>

<property>

<name>**mapred.job.tracker**</name>

<value>**localhost:54311**</value>

<description>**The host and port that the MapReduce job tracker runs at...** </description>

</property>

</configuration>

## /usr/local/hadoop/etc/hadoop/hdfs-site.xml

Tương tự ta cũng phải tạo thư mục và phân quyền cho user hdusser

su hduser

sudo mkdir -p /usr/local/hadoop\_store/hdfs/namenode

sudo mkdir -p /usr/local/hadoop\_store/hdfs/datanode

sudo chown -R hduser:hadoop /usr/local/hadoop\_store

Copy nội dung sau vào file hdfs-site.xml (ghi đè vào thẻ configuration)

<configuration>

<property>

<name>**dfs.replication**</name>

<value>**1**</value>

<description>**Default block replication...** </description>

</property>

<property>

<name>**dfs.namenode.name.dir**</name>

<value>**file:/usr/local/hadoop\_store/hdfs/namenode**</value>

</property>

<property>

<name>**dfs.datanode.data.dir**</name>

<value>**file:/usr/local/hadoop\_store/hdfs/datanode**</value>

</property>

</configuration>

# Format the new Hadoop filesystem

Để bắt đầu chạy được hệ thông chúng ta file định dạng lại file.

su hduser

hadoop namenode –format

# Starting Hadoop

start-all.sh //Starting hadoop

jps //Kiểm tra các service được chạy

6129 Jps

5484 NameNode

5811 SecondaryNameNode

5969 ResourceManager

6094 NodeManager

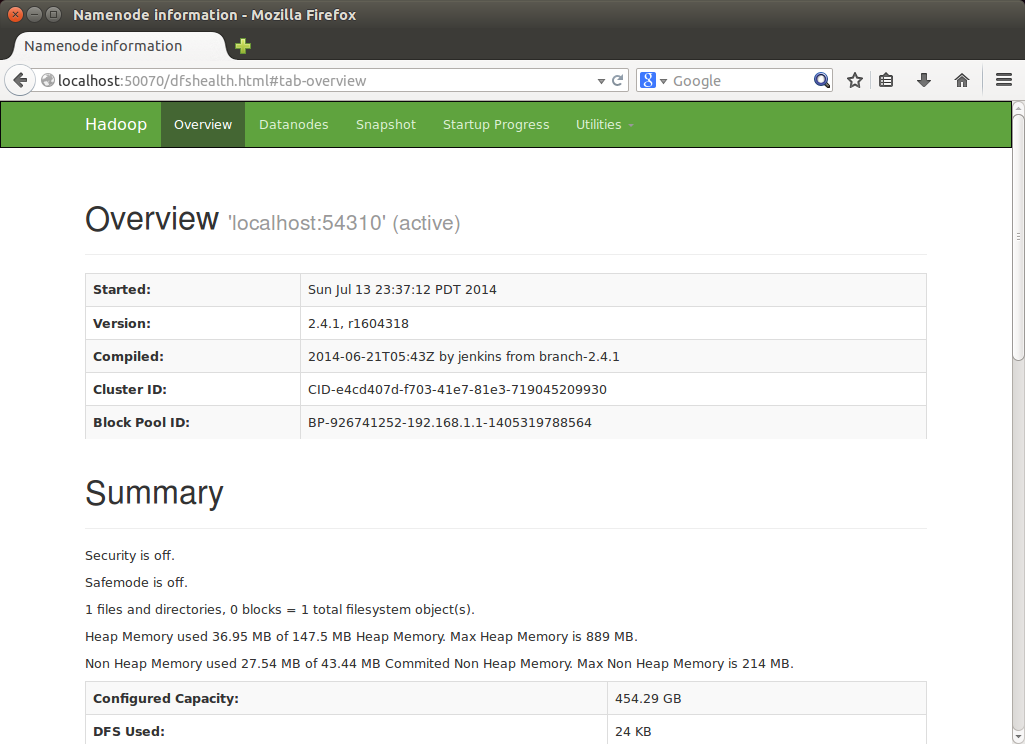
5610 DataNode

# Storping Hadoop

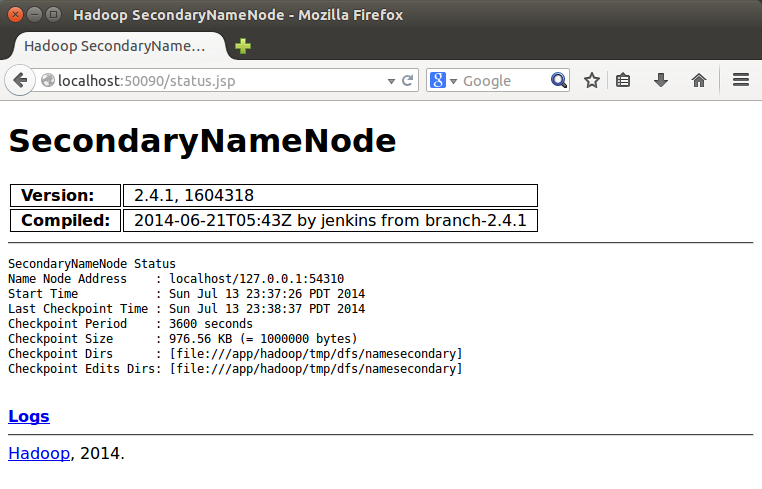
stop-all.sh

# Hadoop website interfaces

http://localhost:50070/ - web UI of the NameNode daemon

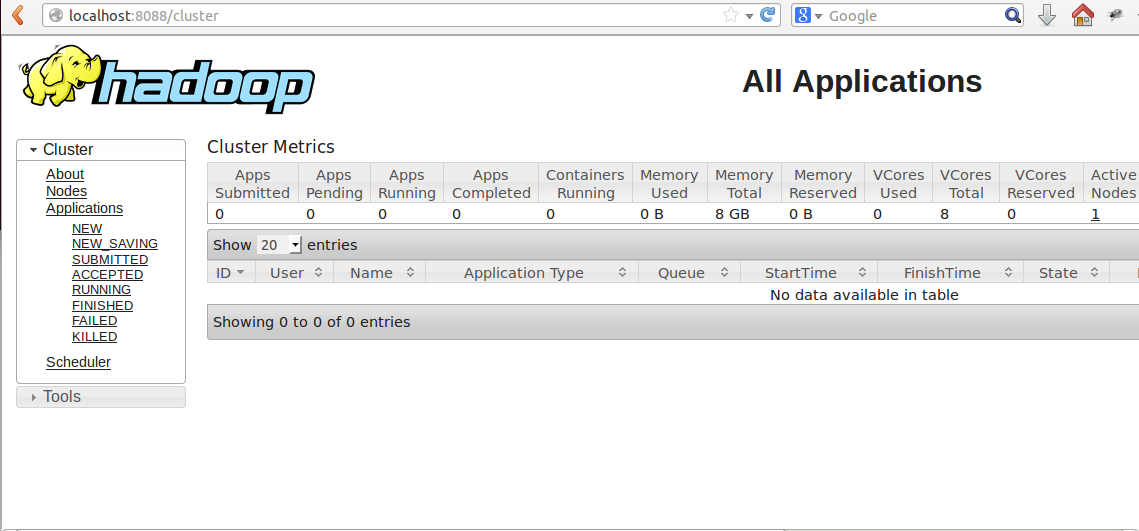


SecondaryNameNode



DataNode

  
<http://localhost:8088> – ResourceManager



Chiennb edit from http://www.bogotobogo.com/Hadoop/BigData\_hadoop\_Install\_on\_ubuntu\_single\_node\_cluster.php