1. Cài đặt máy ảo Ubuntu

Tải Ubuntu phiên bản 18.04 về máy

C:\Users\ASUS>docker pull ubuntu:18.04

Chạy Ubuntu với thư mục liên kết window có sẵn file cài đặt hadoop-2.9.0.tar.gz

docker run -it -v D:/Hoc/Hiveinstall:/data haolee4:latest

```
PS C:\Users\ASUS> docker run -it -v D:/Hoc/Hiveinstall:/data haolee4:latest
root@25ca3fe7c275:/# cd /data
root@25ca3fe7c275:/data# ls
apache-hive-2.1.0-bin.tar.gz  hadoop-2.9.0  hadoop-2.9.0.tar.gz
root@25ca3fe7c275:/data#
```

Sau đó tạo user bằng

adduser haolee

cài sudo

apt install sudo

cài công cụ editor nano

apt install nano

sau khi cài xong tất cả cập nhật lại cài đặt

apt update

2. cài đặt hadoop

2.1 cài đặt java

sudo apt-get install openjdk-8-jdk

Cài đặt openssh-client và sudo apt openssh-serverđể tạo ssh key

sudo apt install openssh-client

sudo apt install openssh-server

Tạo ssh key

ssh-keygen -t rsa -P ""

cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys

Khởi động SSH và kiểm tra có hoạt động hay không

sudo service ssh restart

ssh localhost

```
haolee@25ca3fe7c275:~$ ssh localhost
Welcome to Ubuntu 18.04.5 LTS (GNU/Linux 4.19.128-microsoft-standard x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage
This system has been minimized by removing packages and content that are
not required on a system that users do not log into.

To restore this content, you can run the 'unminimize' command.
Last login: Wed Dec 23 16:51:21 2020 from ::1
haolee@25ca3fe7c275:~$
```

2.2 Bước vào cài đặt hadoop

Ta vào thư mục data giải nén file hadoop-2.9.0.tar.gz và di chuyển vào thư mục /usr/local

cd /data

tar xvzf hadoop-2.9.0.tar.gz

sudo mv hadoop-2.9.0 /usr/local/

Cấp quyền

sudo chown -R <haolee>:<haolee> /usr/local/hadoop-2.9.0

Chỉnh sửa file bashrc

nano ~/.bashrc

```
fi
#HADOOP VARIABLES START
   export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
   export HADOOP_INSTALL=/usr/local/hadoop-2.9.0
   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 END
#java
export PATH="$PATH:$JAVA_HOME/bin"
#HiVE
export HIVE_HOME=/usr/local/apache-hive-2.1.0-bin
export PATH=$PATH:$HIVE_HOME/bin
```

Lưu lại

source ~/.bashrc

Test Java đã được cài đặt

java –version

```
haolee@25ca3fe7c275:~$ java -version
openjdk version "1.8.0_275"
OpenJDK Runtime Environment (build 1.8.0_275-8u275-b01-0ubuntu1~18.04-b01)
OpenJDK 64-Bit Server VM (build 25.275-b01, mixed mode)
haolee@25ca3fe7c275:~$
```

Chỉnh sửa file hdfs-site.xml

cd /usr/local/hadoop-2.9.0/etc/hadoop

sudo nano hdfs-site.xml

```
<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/usr/local/hadoop-2.9.0/namenode</value>
  </property>
  <property>
   <name>dfs.datanode.data.dir</name>
   <value>file:/usr/local/hadoop-2.9.0/datanode</value>
  </property>
</configuration>
```

Chỉnh sửa file core-site.xml

sudo nano core-site.xml

```
<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
   <name>fs.default.name</name>
   <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

Cuối cùng, định dạng Format Hadoop File system

cd /usr/local/hadoop-2.9.0/bin

hadoop namenode -format

start-all.sh

```
haolee@25ca3fe7c275:/usr/local/hadoop-2.9.0/bin$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
21/01/06 10:20:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-ja
va classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop-2.9.0/logs/hadoop-haolee-namenode-25ca3fe7c275.out
localhost: starting datanode, logging to /usr/local/hadoop-2.9.0/logs/hadoop-haolee-datanode-25ca3fe7c275.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop-2.9.0/logs/hadoop-haolee-secondarynamenode-25ca3fe7c27
5.out
21/01/06 10:20:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-ja
va classes where applicable
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop-2.9.0/logs/yarn-haolee-resourcemanager-25ca3fe7c275.out
localhost: starting nodemanager, logging to /usr/local/hadoop-2.9.0/logs/yarn-haolee-nodemanager-25ca3fe7c275.out
haolee@25ca3fe7c275:/usr/local/hadoop-2.9.0/bin$
```

Vậy là namenode đã được tạo, cài đặt hadoop thành công!

3. Cài đặt Hive

Cài đặt mysql

sudo apt-get install mysql-serve

Để cài đặt Hive, ta vào thư mục data giải nén file apache-hive-2.1.0-bin.tar.gz và di chuyển vào thư mục /usr/local

cd /data

sudo tar xvzf apache-hive-2.1.0-bin.tar.gz -C /usr/local

```
haolee@25ca3fe7c275:/usr/local$ ls
apache-hive-2.1.0-bin  bin  etc  games  hadoop-2.9.0  include  lib  man  sbin  share  src
haolee@25ca3fe7c275:/usr/local$
```

Cài đặt MySQL Java Connector

sudo apt-get install libmysql-java

Tạo link mềm để liên kết Hive lib directory hoặc sao chép connector jar to lib folder

#ln–s/usr/share/java/mysql-connector-java.jar
$HIVE_HOME/lib/mysql-connector-java.jar

Tạo lược đồ cơ sở dữ liệu ban đầu bằng cách sử dụng hive-schema-0.14.0.mysql.sql

mysql -u root -p

Enter password:

mysql> CREATE DATABASE metastore;

mysql> USE metastore;

mysql>SOURCE/usr/local/apache-hive-2.1.0 bin/scripts/metastore/upgrade/mysql/ hive-schema-0.14.0.mysql.sql;

Cần một tài khoản người dùng MySQL để tổ ong sử dụng để truy cập Metastore. Điều rất quan trọng là ngăn tài khoản người dùng này tạo hoặc thay đổi các bảng trong lược đồ cơ sở dữ liệu Metastore.

mysql> CREATE USER 'hiveuser'@'%' IDENTIFIED BY 'hivepassword';

mysql> GRANT all on *.* to 'hiveuser'@localhost identified by 'hivepassword';

mysql> flush privileges;

Tạo file Hive-Site.xml (nếu chưa có) trong thư mục $ Hive_Home / Conf với cấu hình

```
<configuration>
   <property>
      <name>javax.jdo.option.ConnectionURL</name>
      <value>jdbc:mysql://localhost/metastore?createDatabaseIfNotExist=true</value>
      <description>metadata is stored in a MySQL server</description>
   </property>
   <property>
      <name>javax.jdo.option.ConnectionDriverName</name>
      <value>com.mysql.jdbc.Driver</value>
      <description>MySQL JDBC driver class</description>
   </property>
   <property>
      <name>javax.jdo.option.ConnectionUserName</name>
      <value>haoleehive</value>
      <description>user name for connecting to mysql server</description>
   </property>
   <property>
      <name>javax.jdo.option.ConnectionPassword</name>
      <value>123456</value>
      <description>password for connecting to mysql server</description>
   </property>
</configuration>
```

Cuối cùng, kiểm tra hive có hoạt động

cd /usr/local/apache-hive-2.1.0-bin/bin

/.hive

```
** END NESTED EXCEPTION **


Wed Jan 06 10:36:20 GMT 2021 WARN: Establishing SSL connection without server's identity verification is not recommended
. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit
option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to
 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore f
or server certificate verification.
Wed Jan 06 10:36:20 GMT 2021 WARN: Establishing SSL connection without server's identity verification is not recommended
. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit
option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to
 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore f
or server certificate verification.
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution
 engine (i.e. spark, tez) or using Hive 1.X releases.
hive>
```

Vậy là thành công!