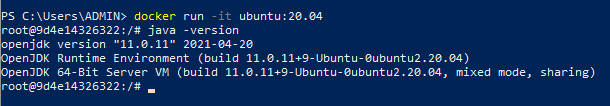
1. **Cài đặt máy ảo Ubuntu**

Sử dụng Ubuntu 20.04 Docker

**#docker pull ubuntu:20.04**

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

**# docker run -it ubuntu:20.04**



Sau đó tạo user bằng

**# adduser minhle**

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**

1. **Cài đặt hadoop**

**2.1 Cài đặt java**

Cài JAVA(JRE)

**# apt install default-jre**

Cài JAVA(JDK)

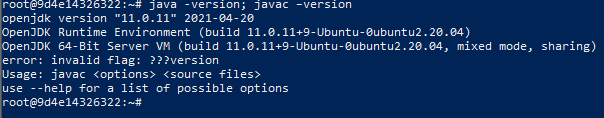
**# apt install default-jdk**

Cài OpenJDK 8:

**#apt install openjdk-8-jdk -y**

Kiểm tra Java đã được cài:

**#java -version; javac –version**



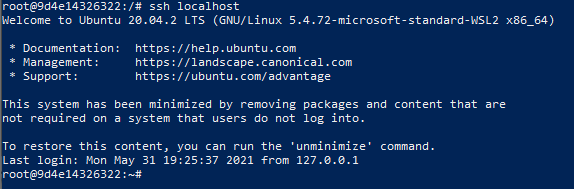
Cài đặt máy chủ và máy khách OpenSSH bằng lệnh sau:

**# apt install openssh-server openssh-client -y**

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

**# apt install openssh-client**

**# apt install openssh-server**



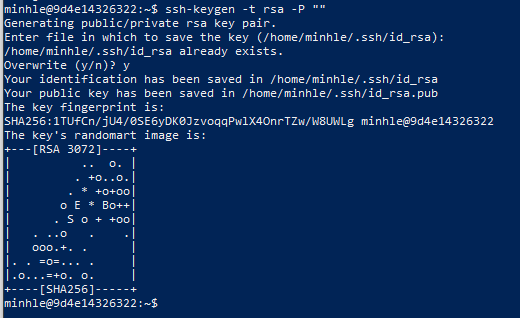
**2.2 Bước vào cài đặt hadoop**

Chuyển đổi người dùng:

**#su – minhle**

Lệnh này sẽ tạo một khóa mới.

**#ssh-keygen -t rsa -P ""**

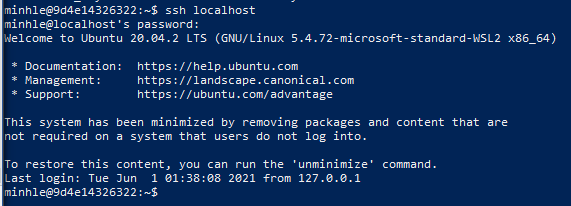


Bật quyền truy cập SSH vào máy cục bộ bằng cách sử dụng khóa này.

**#cat $ HOME / .ssh / id\_rsa.pub >> $ HOME / .ssh / allow\_keys**

Bây giờ hãy kiểm tra thiết lập SSH bằng cách kết nối với localhost với tư cách là người dùng 'hduser'.

**#ssh localhost**



Cài Hadoop:

**#wget** [**https://downloads.apache.org/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz**](https://downloads.apache.org/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz)

**#** **tar xzf hadoop-3.2.1.tar.gz**

Chỉnh sửa file bashrc

**# nano ~/.bashrc**

*#Hadoop Related Options*

*export HADOOP\_HOME=/home/hdoop/hadoop-3.2.1*

*export HADOOP\_INSTALL=$HADOOP\_HOME*

*export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME*

*export HADOOP\_COMMON\_HOME=$HADOOP\_HOME*

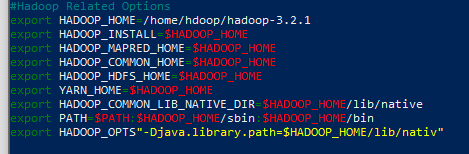
*export HADOOP\_HDFS\_HOME=$HADOOP\_HOME*

*export YARN\_HOME=$HADOOP\_HOME*

*export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native*

*export PATH=$PATH:$HADOOP\_HOME/sbin:$HADOOP\_HOME/bin*

*export HADOOP\_OPTS"-Djava.library.path=$HADOOP\_HOME/lib/nativ"*

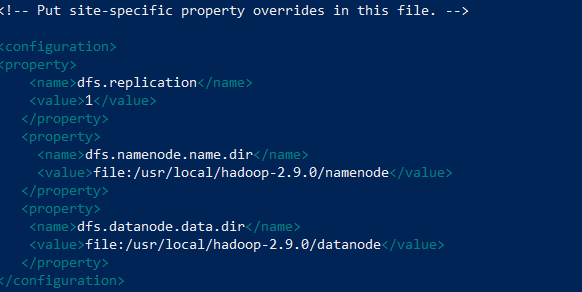


Lưu lại

**# source ~/.bashrc**

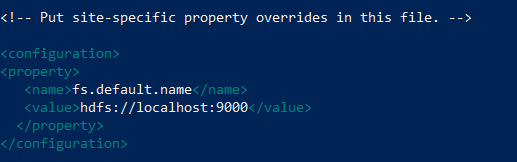
Chỉnh sửa file hdfs-site.xml

**# nano hdfs-site.xml**



Chỉnh sửa file core-site.xml

**# nano core-site.xml**

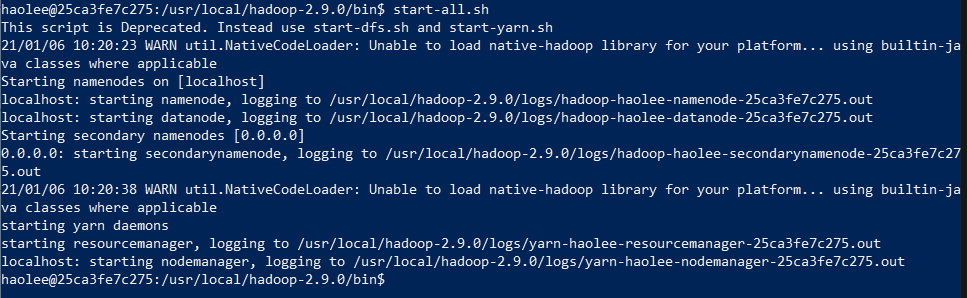


Cuối cùng, định dạng Format Hadoop File system

# cd /usr/local/hadoop-2.9.0/bin

# hadoop namenode -format

# start-all.sh



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

1. **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



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;

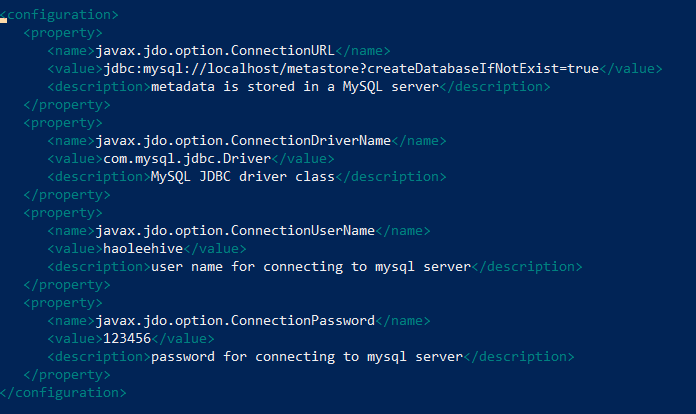
Cần một tài khoản người dùng MySQL để tổ ong sử dụng để truy cập Metastore. Điều rất quan trọng là ngăn tài khoản người dùng này tạo hoặc thay đổi các bảng trong lược đồ cơ sở dữ liệu Metastore.

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

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

mysql> flush privileges;

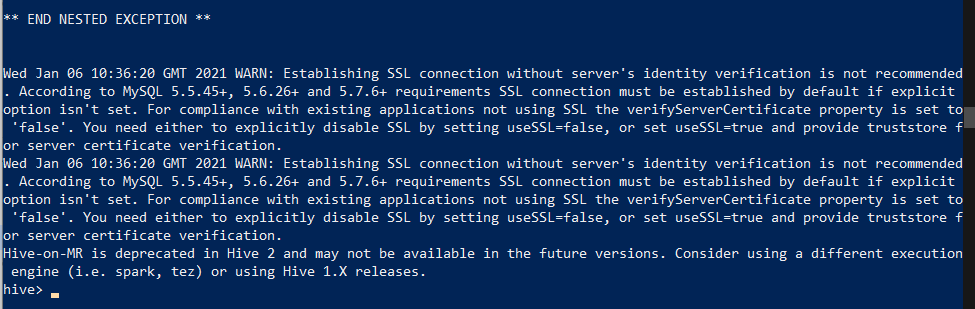
Tạo file Hive-Site.xml (nếu chưa có) trong thư mục $ Hive\_Home / Conf với cấu hình



Cuối cùng, kiểm tra hive có hoạt động

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

# /.hive



Vậy là thành công!