Cài đặt Hadoop phiên bản 3.2.1

Bước 1: Cài đặt Java

yum install java-1.8.0-openjdk\* -y

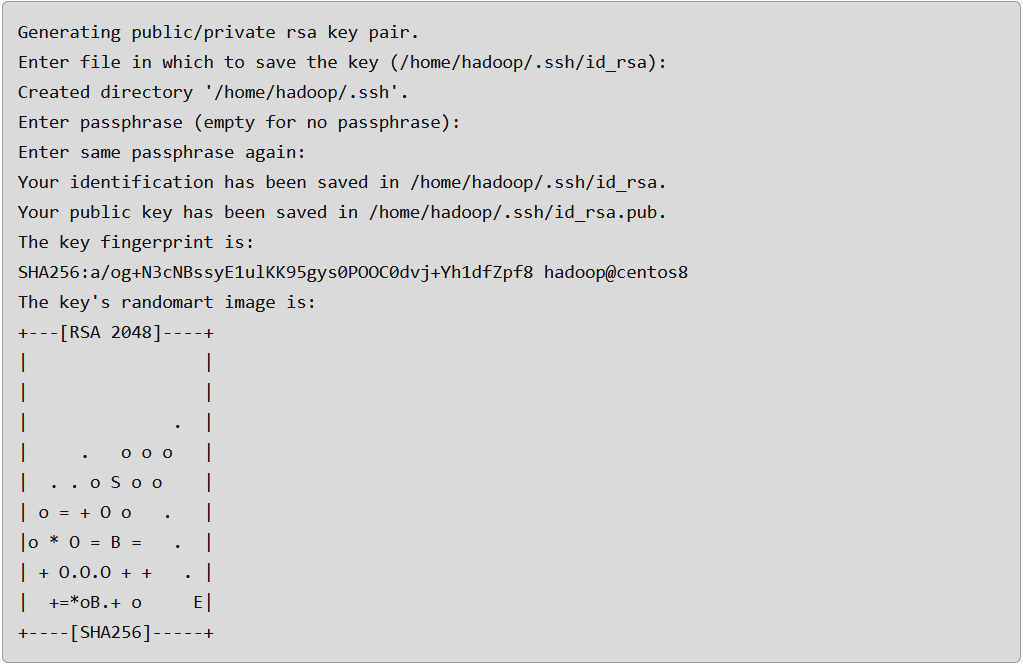
java -version



Bước 2: Cấu hình SSH

ssh-keygen -t rsa

- Chỉ cần nhấn Enter để hoàn thành

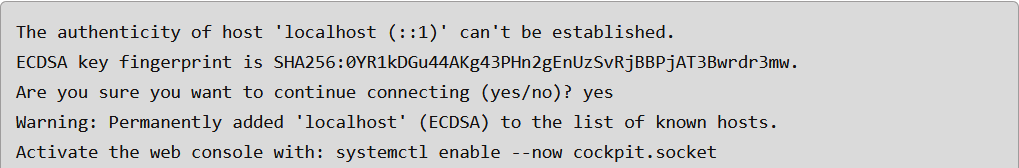


cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

chmod 640 ~/.ssh/authorized\_keys

- Kiểm tra kết nối

ssh localhost



Bước 3: Cài đặt Hadoop bằng đường dẫn

wget <https://downloads.apache.org/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz>

tar -xvzf hadoop-3.2.1.tar.gz

mv hadoop-3.2.1 hadoop

\*Chú ý: Dấu ~ sẽ bằng với đường dẫn /home/tên người dùng trên máy của bạn



- Sửa file ẩn .bashrc

nano ~/.bashrc

- Thêm các dòng này vô ở cuối file

export JAVA\_HOME=

export HADOOP\_HOME=/home/tên người dùng/hadoop

export HADOOP\_INSTALL=$HADOOP\_HOME

export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_HOME=$HADOOP\_HOME

export HADOOP\_HDFS\_HOME=$HADOOP\_HOME

export HADOOP\_YARN\_HOME=$HADOOP\_HOME

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

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

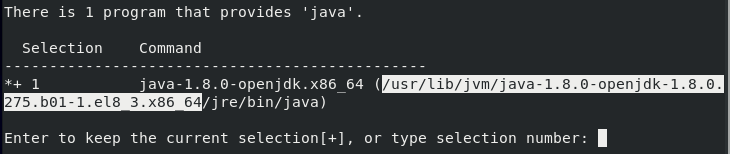
export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_HOME/lib/native"

- Dùng tổ hợp phím Ctrl+O với Ctrl+X để lưu file

\*Lưu ý: export JAVA\_HOME=tên file của bạn

- Ở terminal dùng tổ hợp Ctrl+Shift+t để mở 1 cửa sổ khác

sudo update-alternatives --config java



- Copy phần bôi đậm Paste vào phần export JAVA\_HOME ở file .bashrc

- Nhập 1 để thoát khỏi chế độ

Bước 4: Cấu hình Hadoop

source ~/.bashrc

nano $HADOOP\_HOME/etc/hadoop/hadoop-env.sh

export JAVA\_HOME=đường dẫn như ở file .bashrc

- Tạo 2 file node

mkdir -p ~/hadoopdata/hdfs/namenode

mkdir -p ~/hadoopdata/hdfs/datanode

- Sửa file core-site.xml

nano $HADOOP\_HOME/etc/hadoop/core-site.xml

- Thêm vào giữa cặp dấu <configuration> </configuration>

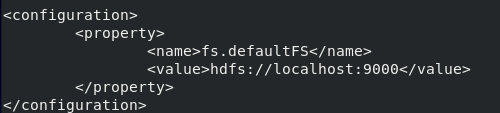
<property>

<name>fs.defaultFS</name>

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

</property>

- Kết quả



- Dùng tổ hợp phím Ctrl+O với Ctrl+X để lưu file

- Sửa file hdfs-site.xml

nano $HADOOP\_HOME/etc/hadoop/hdfs-site.xml

<property>

                <name>dfs.replication</name>

                <value>1</value>

        </property>

        <property>

                <name>dfs.name.dir</name>

                <value><file:///home/tên> người dùng/hadoopdata/hdfs/namenode</value>

        </property>

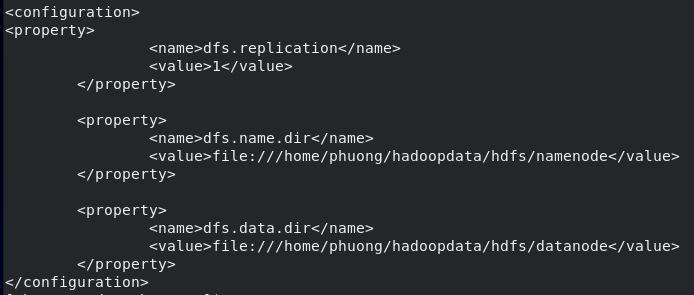
        <property>

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

                <value><file:///home/tên> người dùng/hadoopdata/hdfs/datanode</value>

        </property>

- Kết quả

- Dùng tổ hợp phím Ctrl+O với Ctrl+X để lưu file

- Sửa file yarn-site.xml

nano $HADOOP\_HOME/etc/hadoop/yarn-site.xml

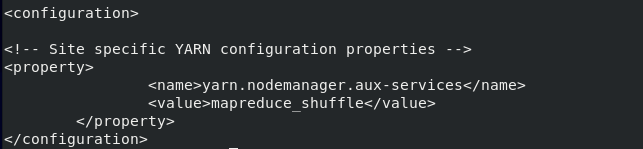
<property>

                <name>yarn.nodemanager.aux-services</name>

                <value>mapreduce\_shuffle</value>

        </property>

- Kết quả



Bước 5: Cấu hình Firewall

firewall-cmd --permanent --add-port=9870/tcp

firewall-cmd --permanent --add-port=8088/tcp

firewall-cmd --reload

Bước 6: Khởi động Hadoop

hdfs namenode -format

start-dfs.sh

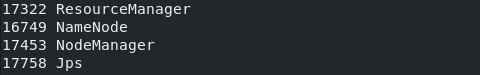


start-yarn.sh



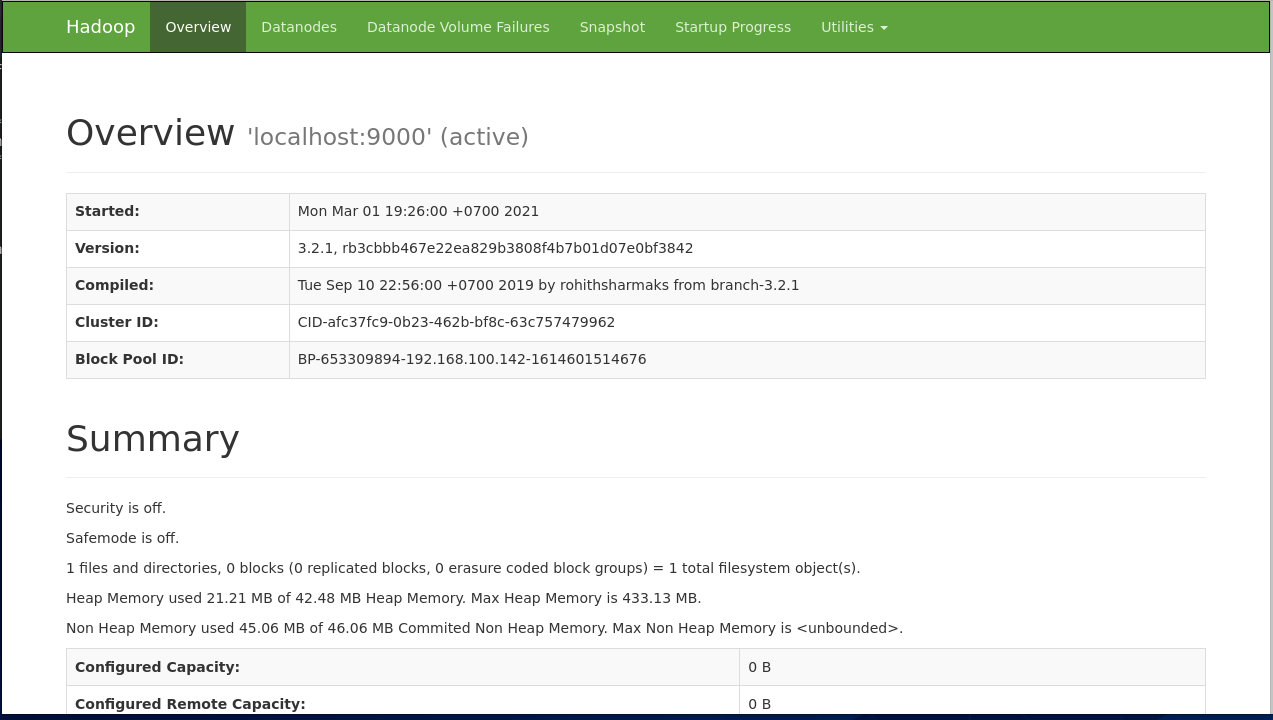
- Kiểm tra trạng thái dịch vụ Hadoop

jps

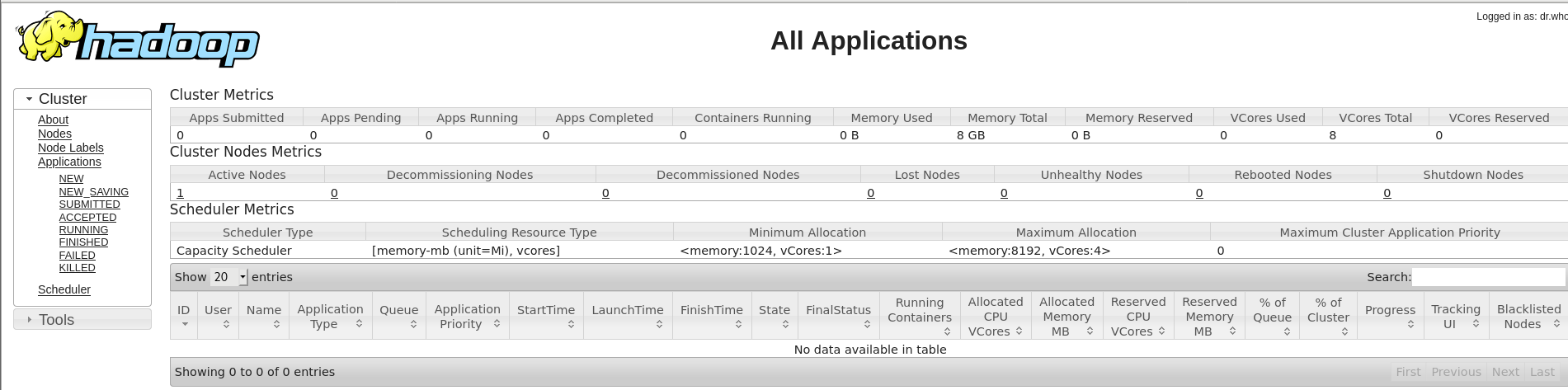


Bước 7: Truy cập Hadoop

Để truy cập Namenode vào trình duyệt web và truy cập <http://localhost:9870>



Để truy cập kiểm tra tài nguyên vào trình duyệt web và truy cập <http://localhost:8088>



Bước 7: Kiểm tra lại Hadoop

- Tạo một số thư mục trong tệp hdfs

hdfs dfs -mkdir /test1

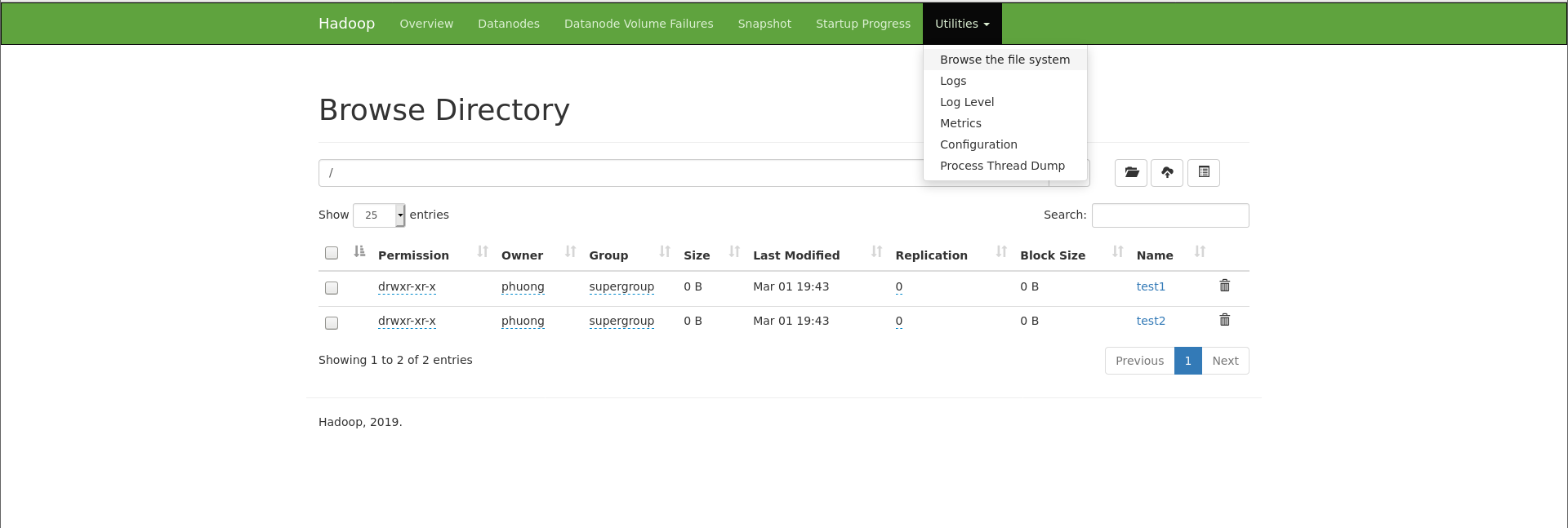
hdfs dfs -mkdir /test2

- Kiểm tra các thư mục trên

hdfs dfs -ls /



- Ta cũng có thể kiểm tra trên web bằng Hadoop Namenode



Bước 8: Dừng Hadoop

stop-dfs.sh

stop-yarn.sh