**Bước 1: Kiểm tra điều kiện trước khi cài đặt**

***/\* Cài Openjdk theo câu lệnh sau : \*/***

***sudo apt-get install openjdk-11-jdk -y***

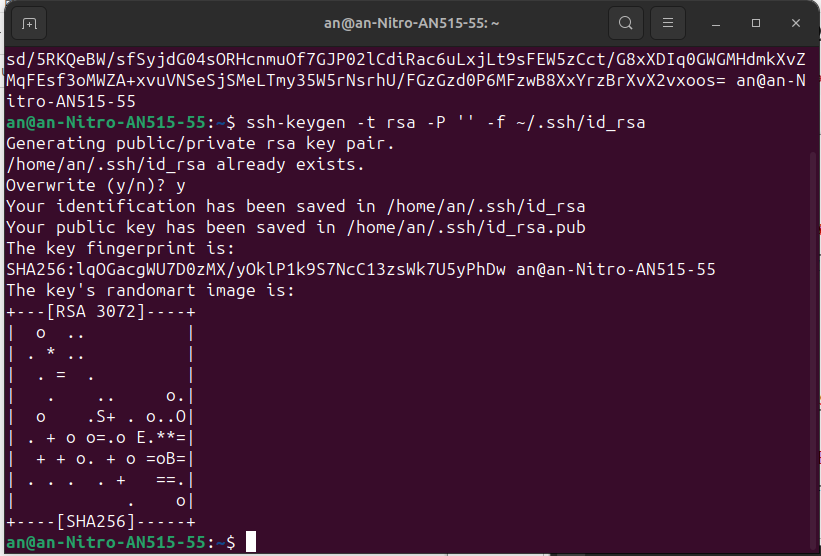
*/\*Cài SSH client và SSH server theo câu lệnh sau :\*/*

***sudo apt-get install -y openssh-server openssh-client***

**Bước 2: Thiết lập Private key cho Hadoop**

*/\* Tạo cặp khóa SSH và xác định vị trí sẽ được lưu trữ :\*/*

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

 Hệ thống tiến hành tạo và lưu cặp khóa

*/\* Sử dụng lệnh cat để lưu public key vào authorized\_keys trong thư mục của SSH:\*/*

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

***/\*****Phân quyền cho người dùng với lệnh****chmod****:\*/*

***chmod 0777 ~/.ssh/authorized\_keys***

*/\*Xác minh mọi thứ được thiết lập chính xác bằng cách ssh đến localhost: \*/*

***ssh localhost***

**Bước 3: Download và Cài đặt Hadoop trên Ubuntu**

*/\*Tải về một phiên bản Hadoop trên trang phân phối chính thức của Hadoop tại :*[*https://hadoop.apache.org/releases.html\*/*](https://hadoop.apache.org/releases.html*/)

*/\*Giải nén file nén ra bằng lệnh : \*/*

***tar xvzf hadoop-3.4.0.tar.gz***

**Bước 4: Cấu hình và triển khai Hadoop Single Node (Pseudo-Distributed Mode)**

*/\*****Cấu hình biến môi trường Hadoop ( file .bashrc)***

*Mở file .bashrc bằng trình soạn thảo nano : \*/*

***sudo nano ~/.bashrc***

***/\*****Xác định biến mỗi trường****Hadoop****bằng cách thêm các biến sau vào cuối file\*/*

***#Hadoop Related Options***

***JAVA\_HOME=/usr/lib/jvm/java-11-openjdk-amd64***

***export HADOOP\_HOME=/home/an/app/hadoop-3.4.0***

***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"***

***/\****Áp dụng các thay đổi trên bằng cách thực hiện lệnh sau : \*/

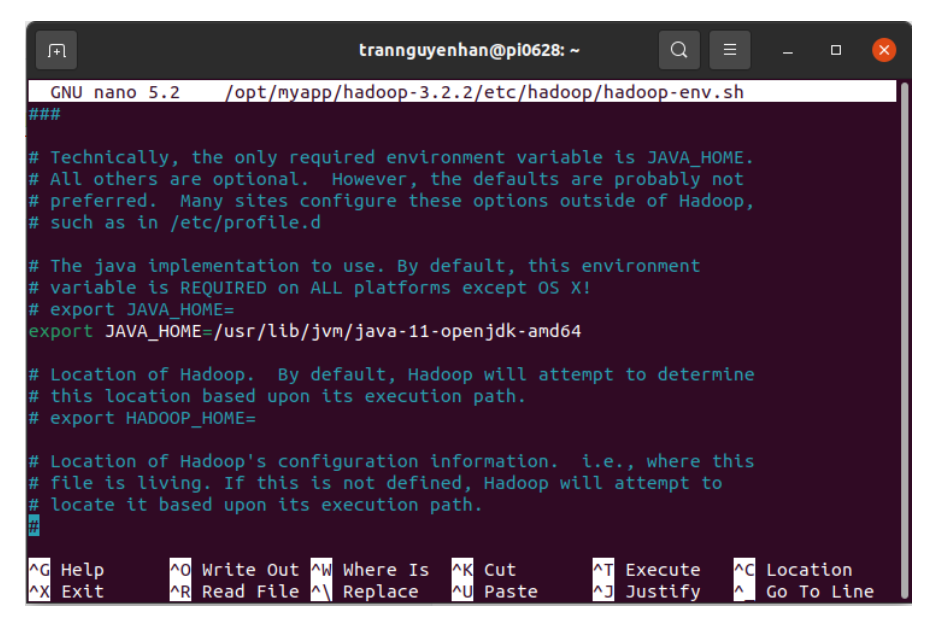
***source ~/.bashrc***

**Bước 5: Chỉnh sửa file hadoop-env.sh**

*/\*Mở file hadoop-env.sh bằng trình soạn thảo nano : \*/*

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

***/\*Tìm tới vị trí như hình bên dưới, bỏ comment ( bỏ dấu #) phần JAVA\_HOME và thêm vào đầy đủ đường dẫn Openjdk :\*/***



Giao diện chỉnh sửa file hadoop-env.sh

*/\* Thêm HADOOP\_SSH\_OPTS: \*/*

**export HADOOP\_SSH\_OPTS="-p 22"**

**Bước 6: Chỉnh sửa file hadoop-env.sh**

*/\*Mở file <hadoop\_directory>/etc/hadoop/core-site.xml bằng trình soạn thảo nano :\*/*

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

*/\*Thêm vào giữa 2 thẻ configuration :\*/*

***<configuration>***

***<property>***

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

***<value>/home/<user>/hadoop-3.2.2/tmpdata</value>***

***</property>***

***<property>***

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

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

***</property>***

***</configuration>***

**Bước 7: Chỉnh sửa file hdfs-site.xml**

*/\** *Mở file core-site.xml bằng trình soạn thảo nano : \*/*

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

*/\*Thêm vào giữa 2 thẻ configuration : \*/*

***<configuration>***

***<property>***

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

***<value>/home/an/hadoop-3.4.0/dfsdata/namenode</value>***

***</property>***

***<property>***

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

***<value>/home/an/app/hadoop-3.4.0/dfsdata/datanode</value>***

***</property>***

***<property>***

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

***<value>2</value>***

***</property>***

*</configuration>*

**Bước 8: Chỉnh sửa file mapred-site.xml**

***/\*Mở file core-site.xml bằng trình soạn thảo nano : \*/***

*sudo nano $HADOOP\_HOME/etc/hadoop/mapred-site.xml*

***/\*Thêm vào giữa 2 thẻ configuration : \*/***

***<configuration>***

***<property>***

***<name>mapreduce.framework.name</name>***

***<value>yarn</value>***

***</property>***

***</configuration>***

**Bước 9: Chỉnh sửa file yarn-site.xml**

*/\** *Mở file core-site.xml bằng trình soạn thảo nano :\*/*

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

*/\*Thêm vào giữa 2 thẻ configuration : \*/*

***<configuration>***

***<property>***

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

***<value>mapreduce\_shuffle</value>***

***</property>***

***<property>***

***<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>***

***<value>org.apache.hadoop.mapred.ShuffleHandler</value>***

***</property>***

***<property>***

***<name>yarn.resourcemanager.hostname</name>***

***<value>127.0.0.1</value>***

***</property>***

***<property>***

***<name>yarn.acl.enable</name>***

***<value>0</value>***

***</property>***

***<property>***

***<name>yarn.nodemanager.env-whitelist</name>***

***<value>JAVA\_HOME,HADOOP\_COMMON\_HOME,HADOOP\_HDFS\_HOME,HADOOP\_CONF\_DIR,CLASSPATH\_PERPEND\_DISTCACHE,HADOOP\_YARN\_HOME,HADOOP\_MAPRED\_HOME</value>***

***</property>***

***</configuration>***

**Bước 10: Format HDFS namenode**

*/\** *Định dạng lại namenode trước khi bắt đầu các dịch vụ đầu tiên : \*/*

***hdfs namenode –format***

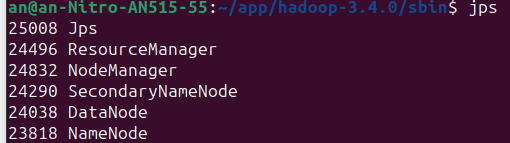
**Bước 11: Start Hadoop Cluster**

*/\* Tại thư mục sbin, thực hiện các lệnh sau để chạy và khởi động Hadoop : \*/*

***./start-all.sh***

*/\*Chạy lệnh jsp để kiểm tra các trình daemon đang chạy : \*/*

***jps***

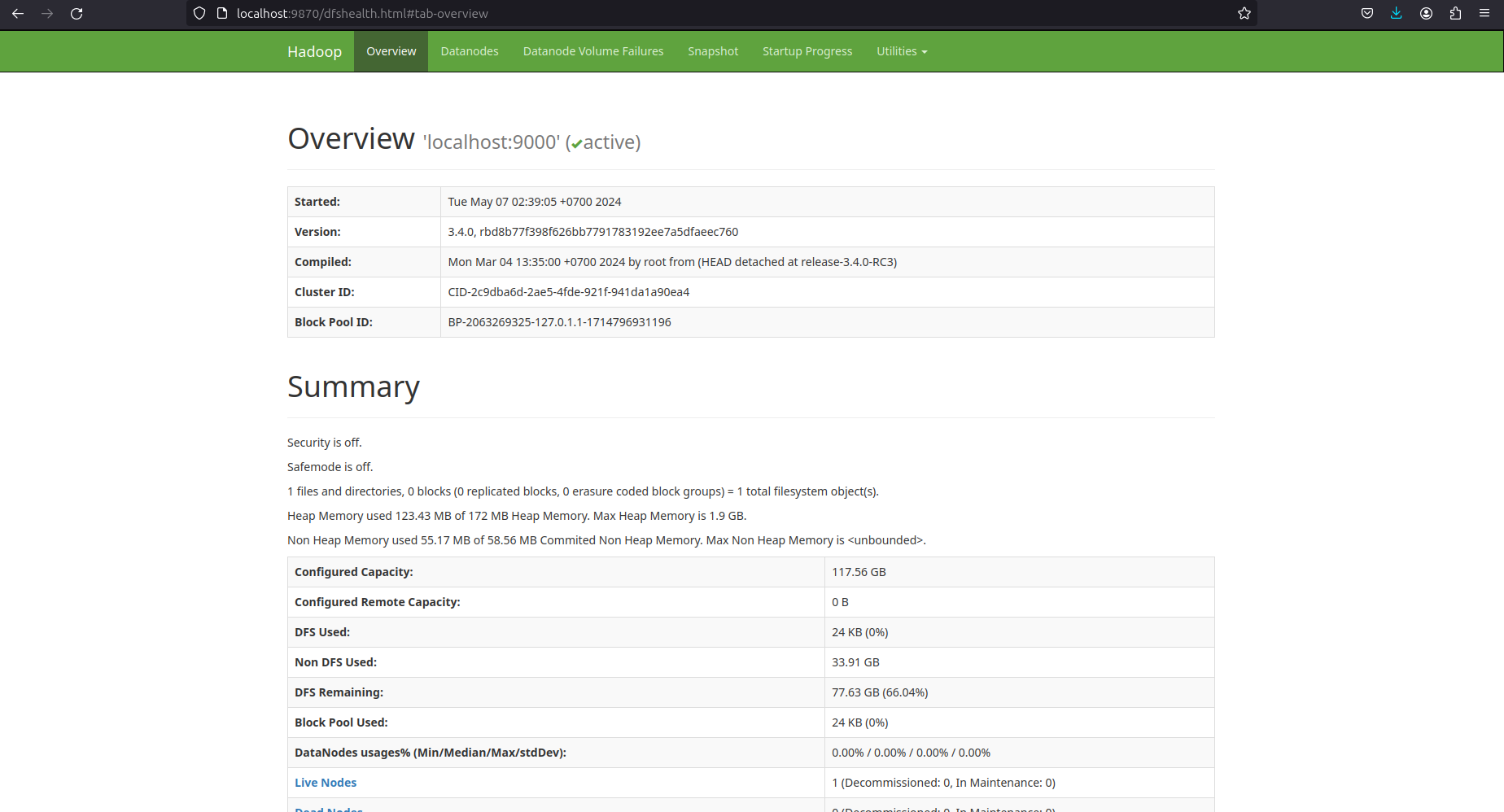


Số trình daemon như sau thì bạn đã cấu hình đúng

***Truy cập Hadoop UI từ trình duyệt***

*/\*Các bạn có thể kiểm tra Hadoop đã được cài đặt thành công hay chưa tại cổng mặc định của namenode là 9870 : \*/*

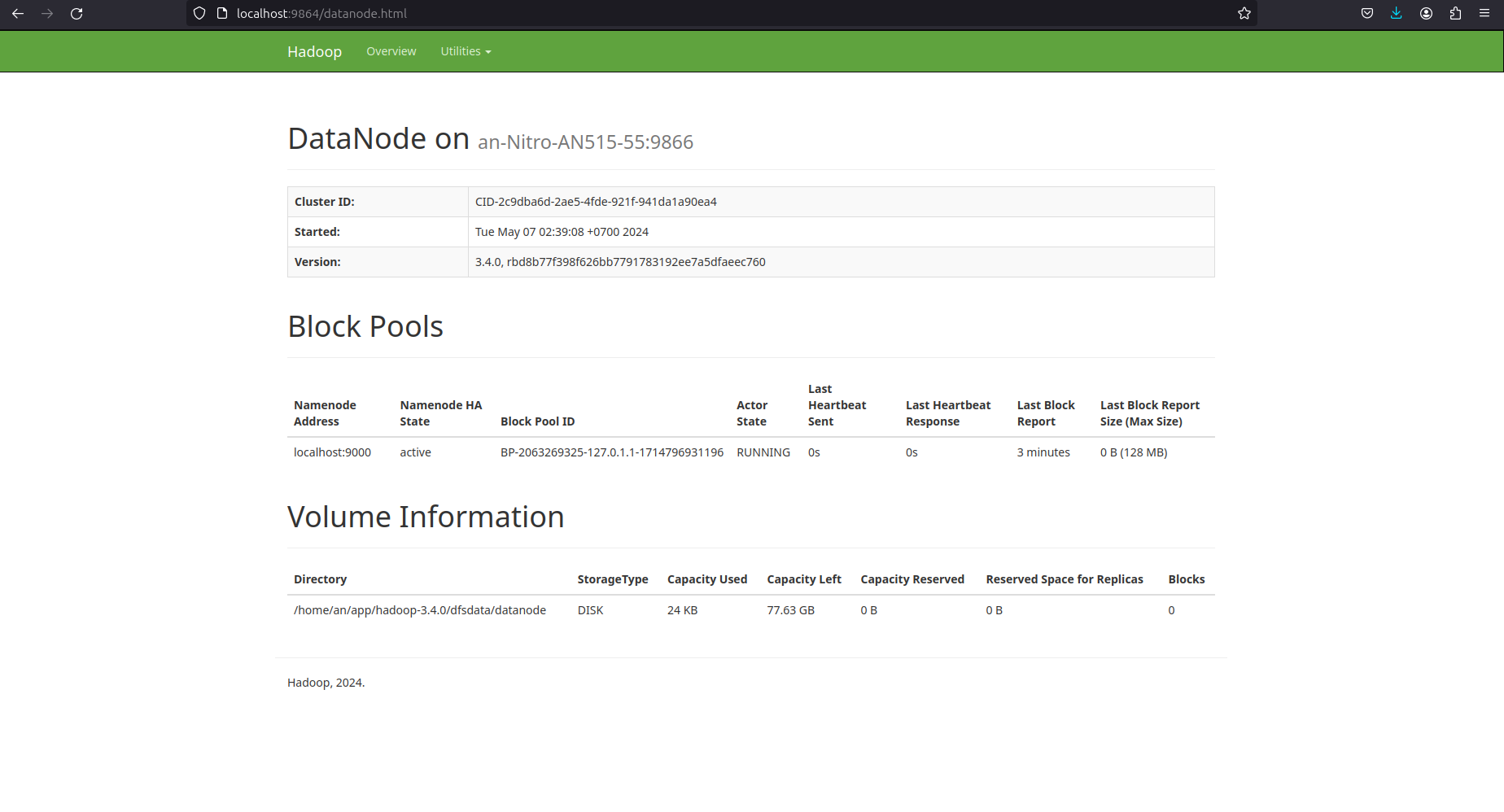
***localhost:9870***

Truy cập Hadoop UI từ trình duyệt

***Kiểm tra datanode tại cổng mặc định 9864:***

*/\*Kiểm tra datanode tại cổng mặc định 9864 : \*/*

***localhost:9864***

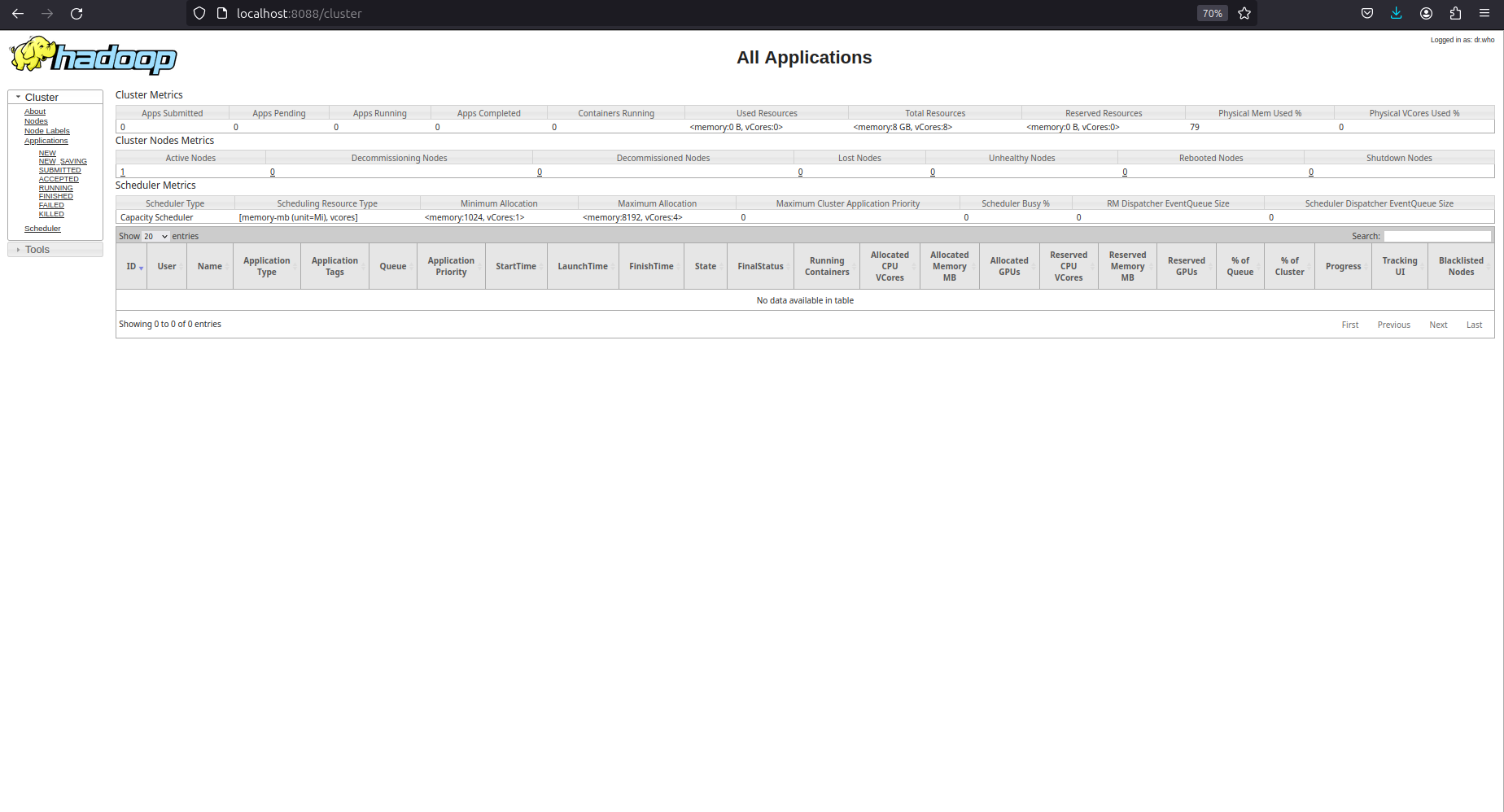


datanode tại cổng mặc định 9864

***Kiểm tra Yarn resource manager tại cổng 8088:***

***/\*Kiểm tra Yarn resource manager tại cổng 8088 : \*/***

***locahost:8088***

Yarn resource manager tại cổng 8088