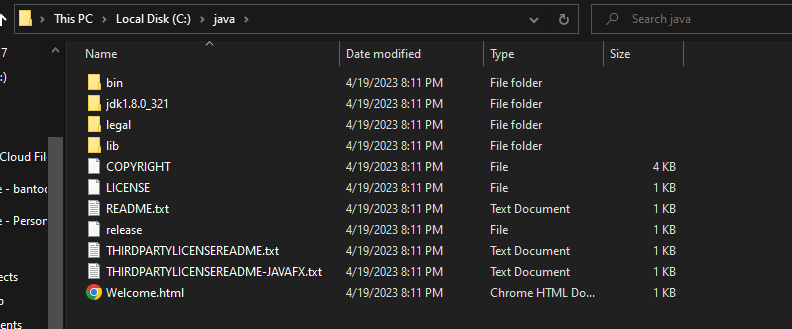
Đỗ Hoàng Thinh - 20133122

HADOOP ECOSYSTEM

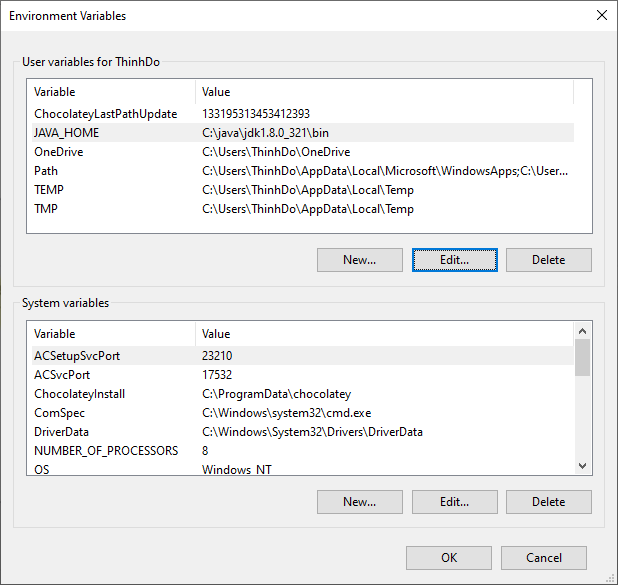
# 0. JAVA

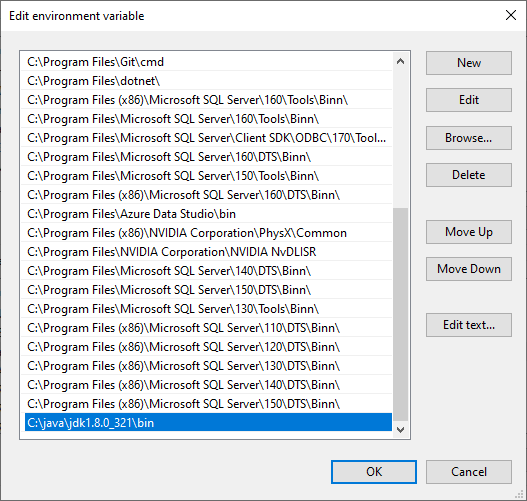
## 0.1. Tải và cài đặt Java

Để Java trong thư mục như sau

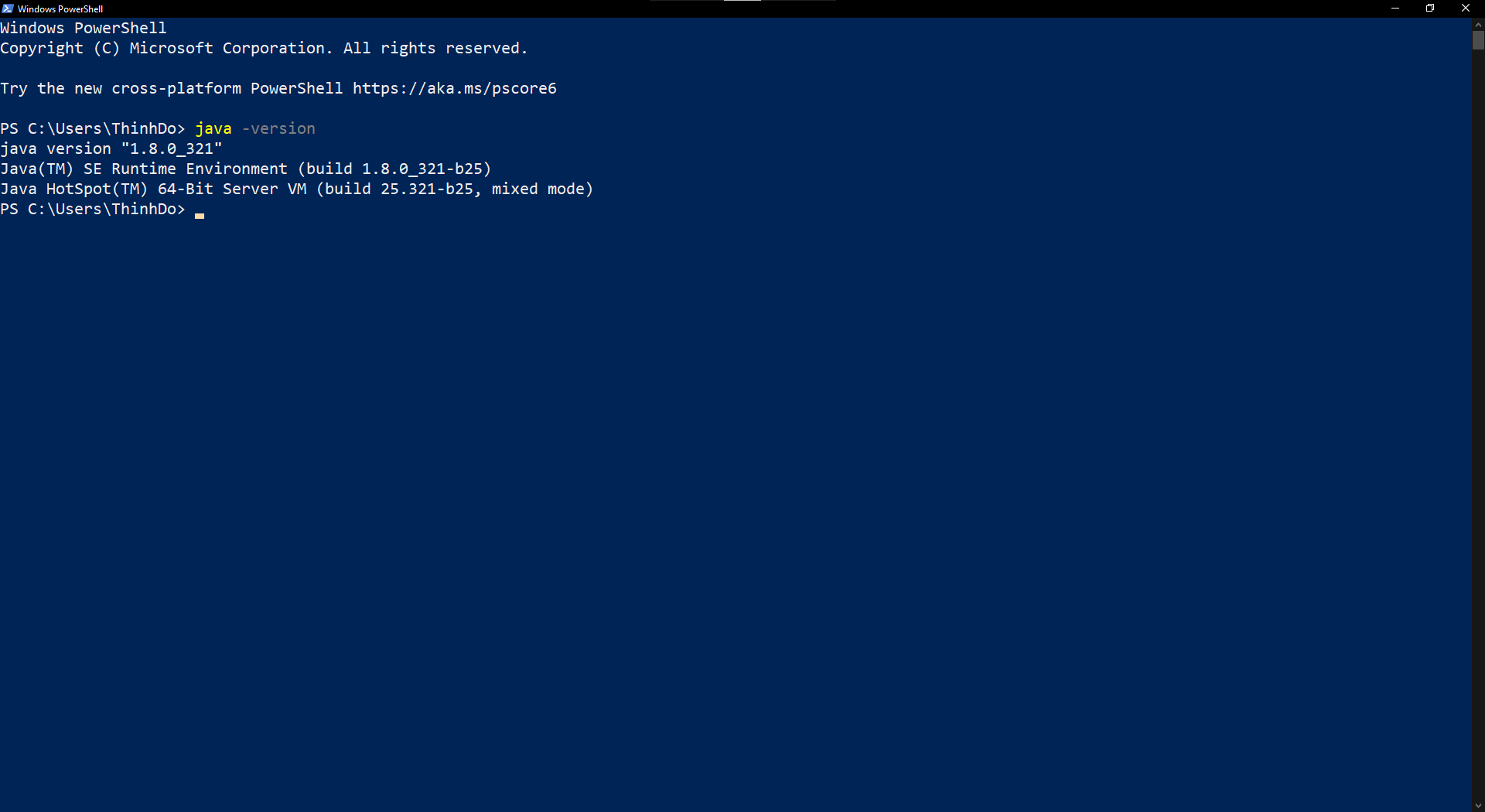


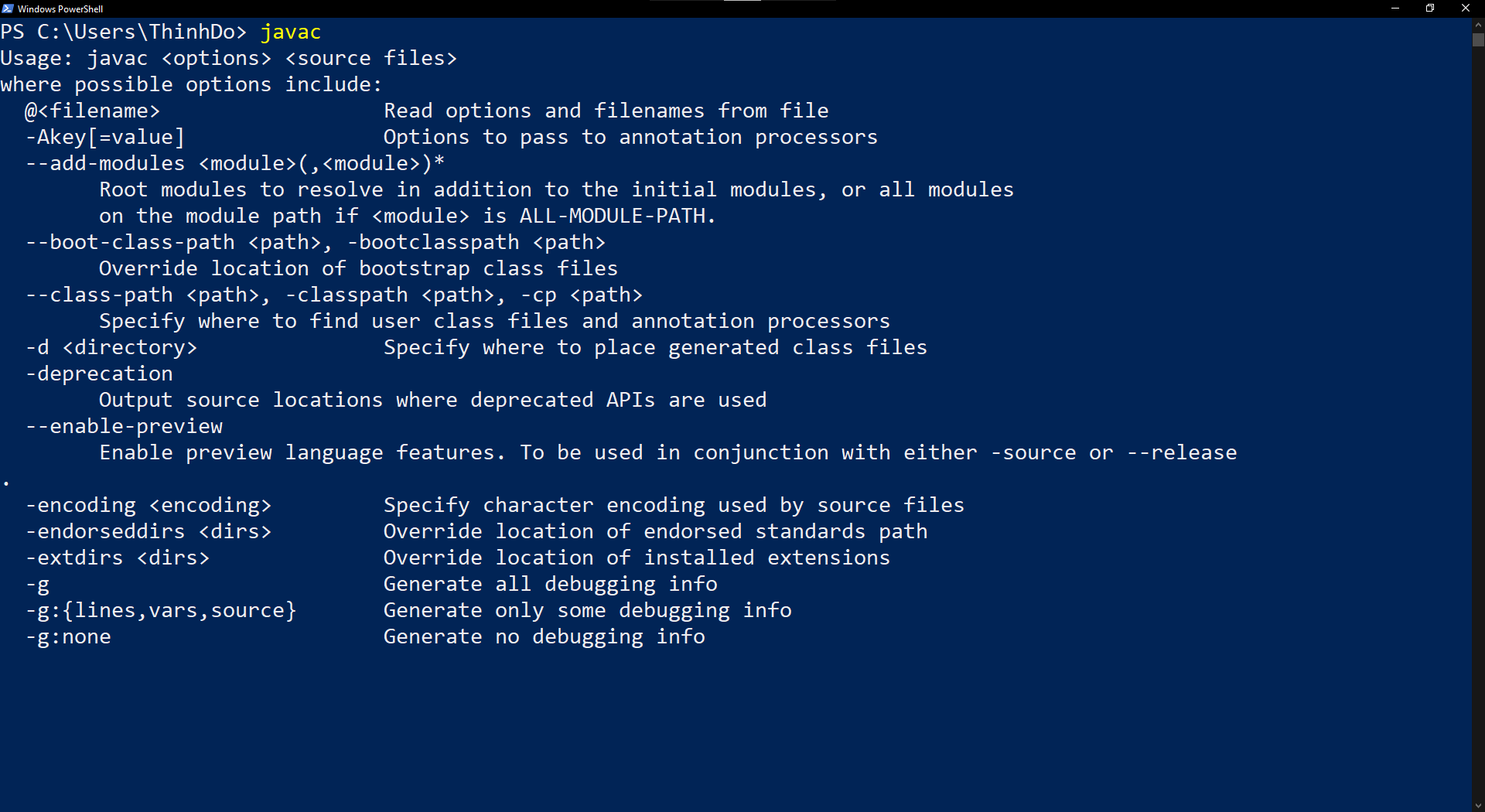
Cài đặt biến môi trường trên Window





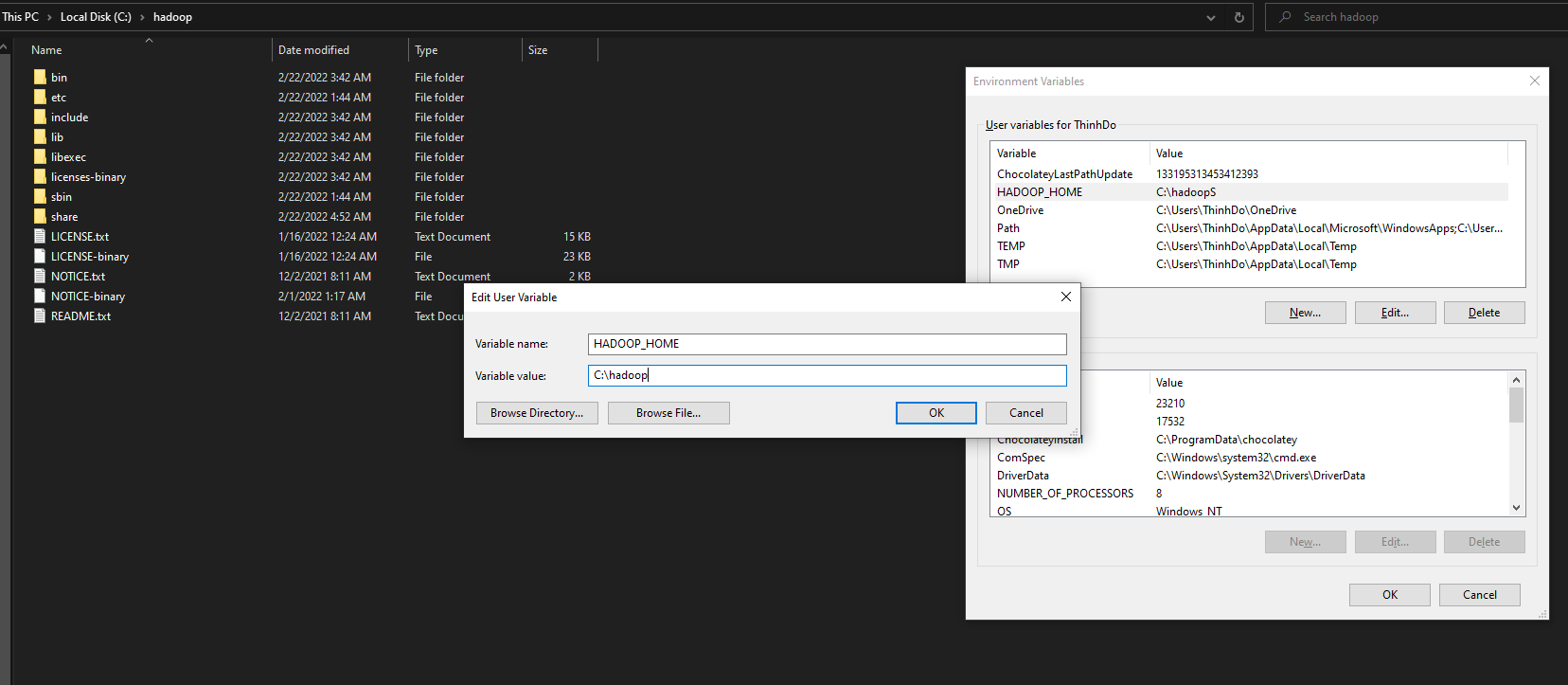
## 0.2. Check Java

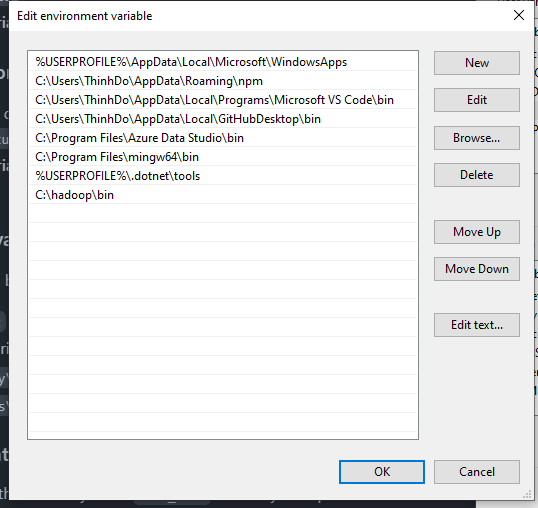




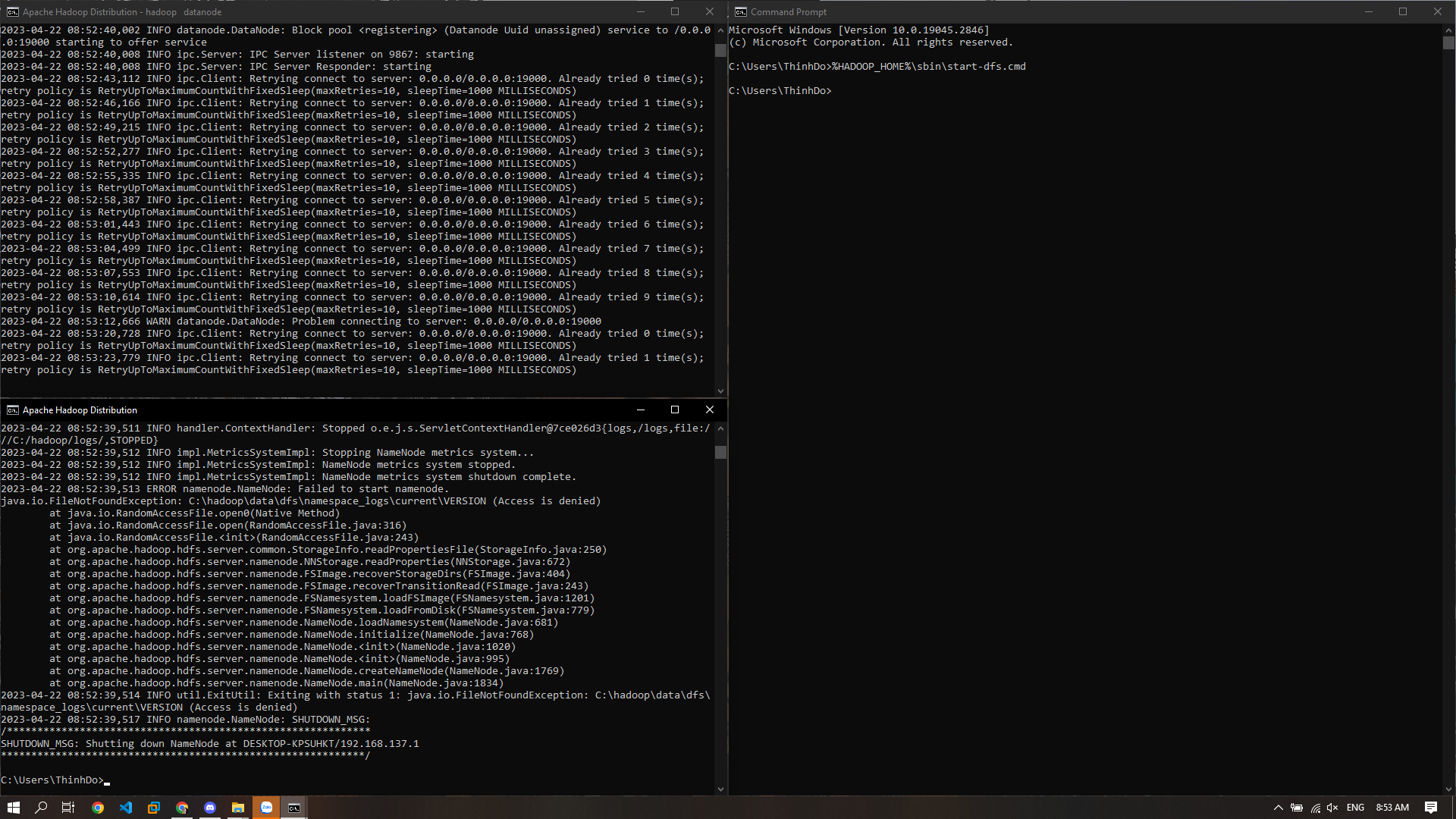
# 1. APACHE HADOOP

## 1.1. Tải và cài đặt hadoop, cài đặt biến môi trường



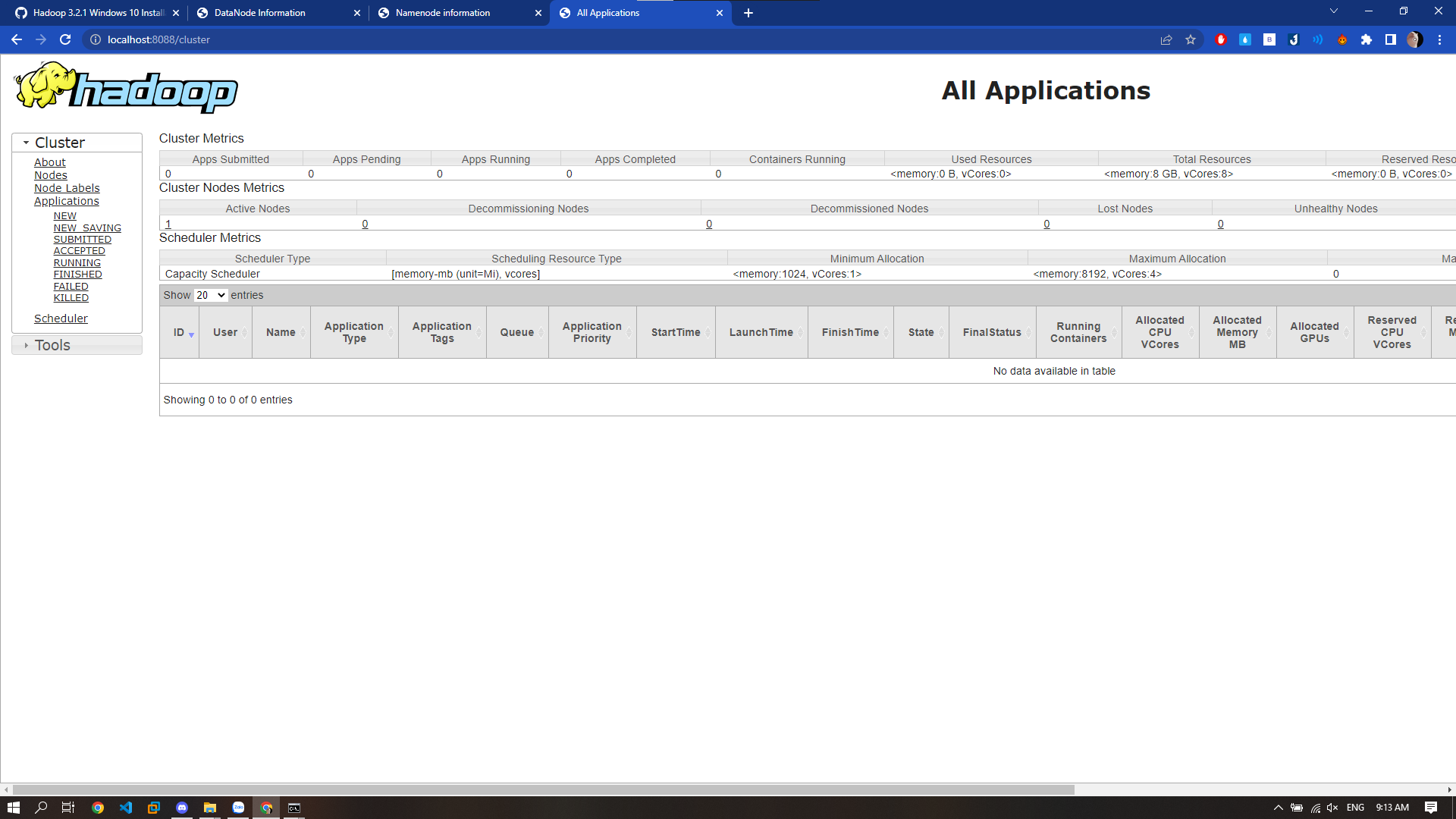


## 1.2. Chạy Hadoop

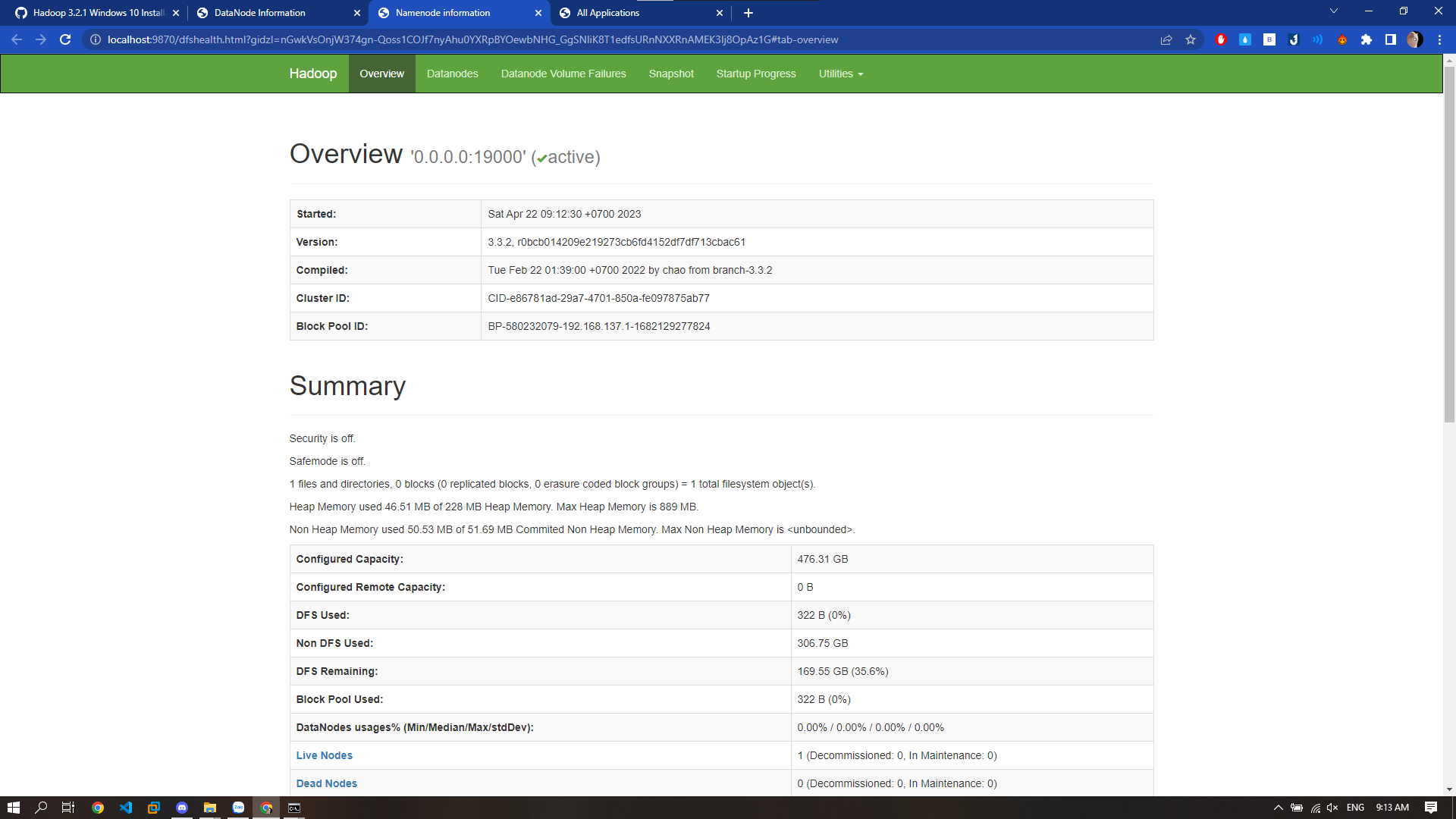


Check cluster:

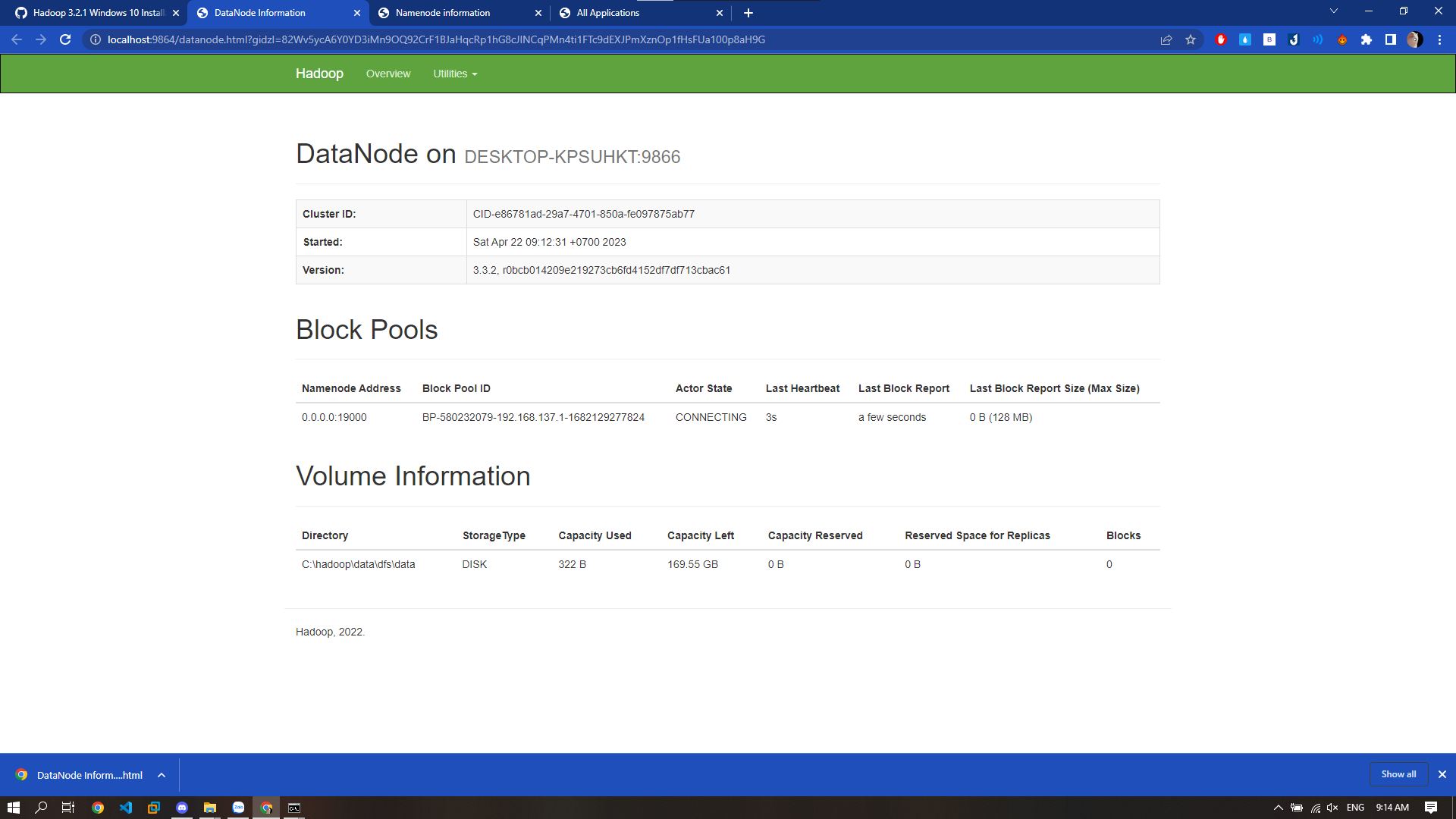
localhost:8088/cluster



Check Hadoop file system

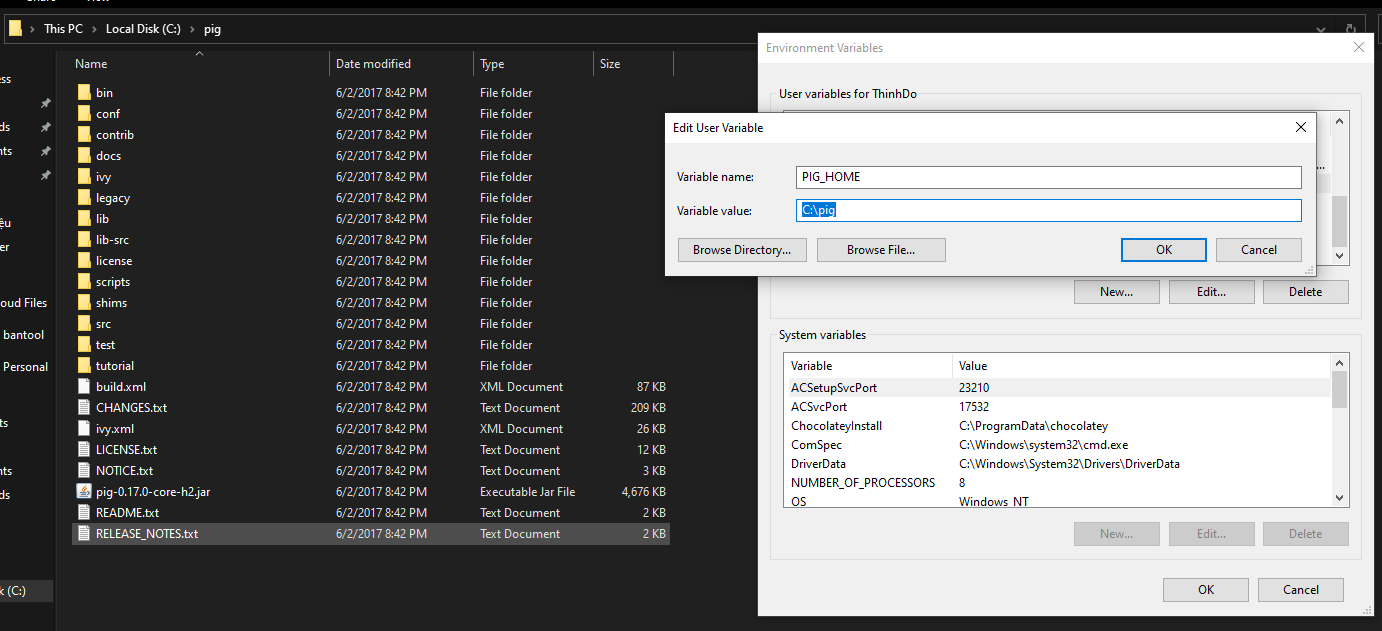
http://localhost:9870/dfshealth.html#tab-overview

Check data nodes

http://localhost:9870/dfshealth.html#tab-datanode

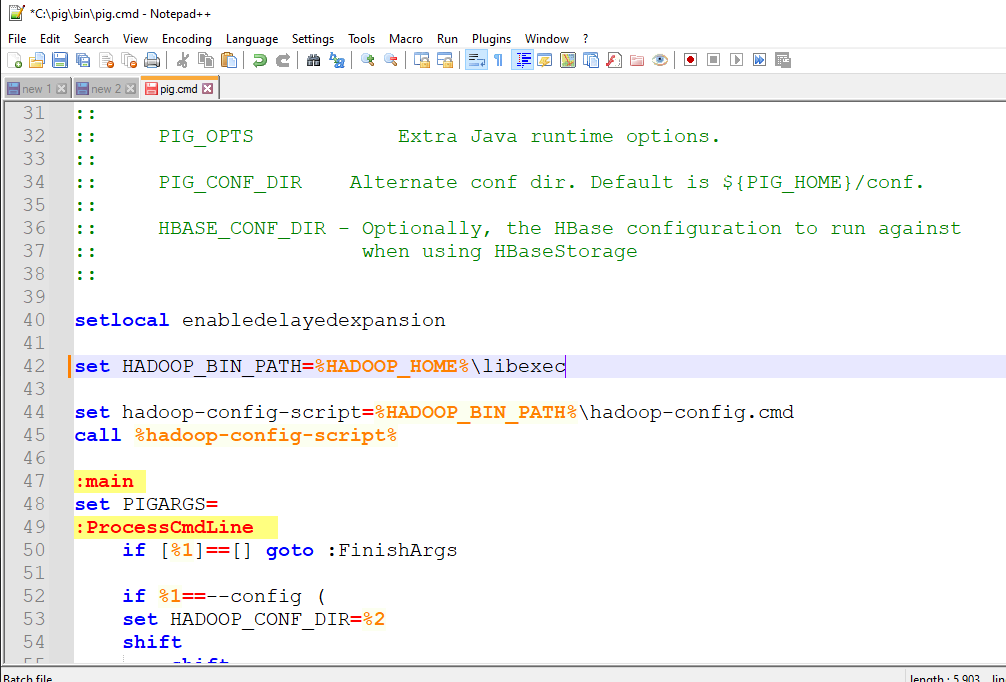
# 2. Apache Pig

## 2.1. Tải, cài đặt và cấu hình cho Apache Pig

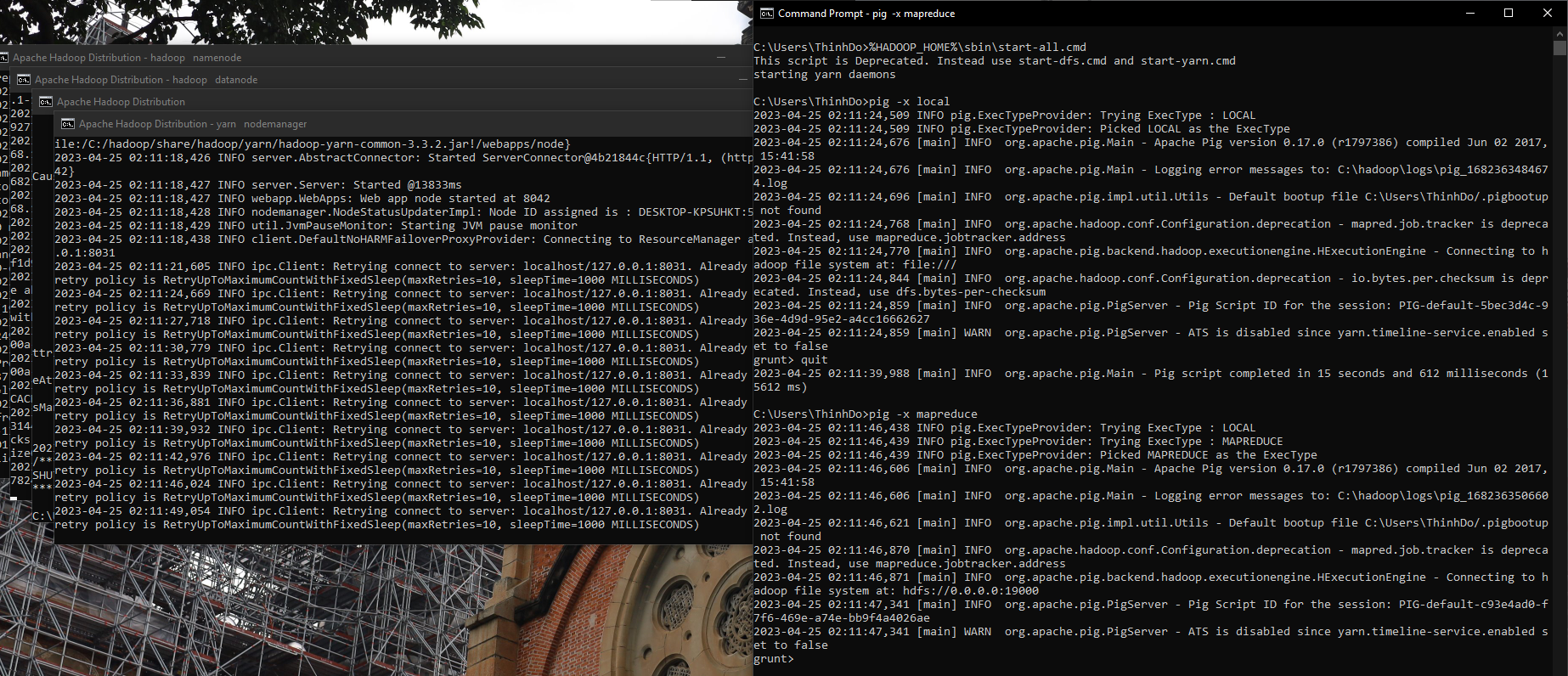


Graphical user interface, text, application, email

Description automatically generated



## 2.2. Chạy Hadoop Pig ở 2 chế độ



## 2.3. Demo HadoopPig

### 2.3.1. Data

Graphical user interface, application, table, Excel

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

### 2.3.2. Run

Text

Description automatically generated

Text

Description automatically generated

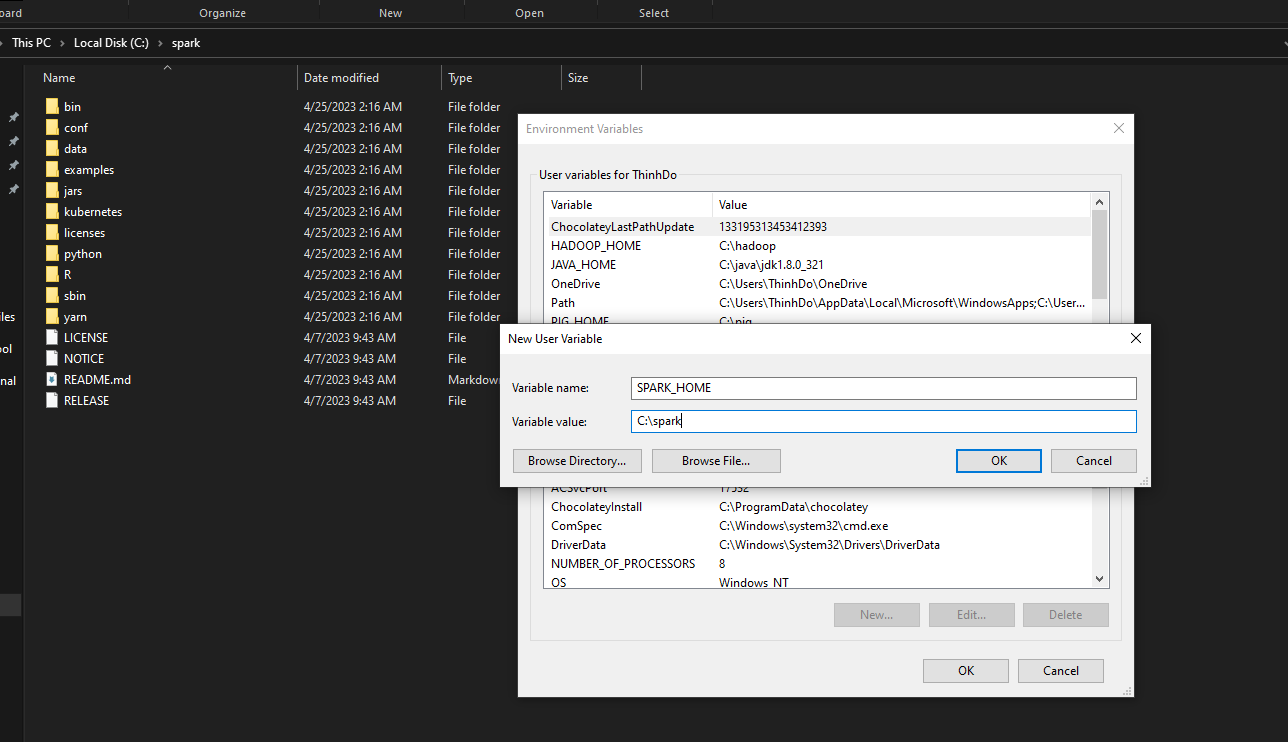
### 2.3.3. Result

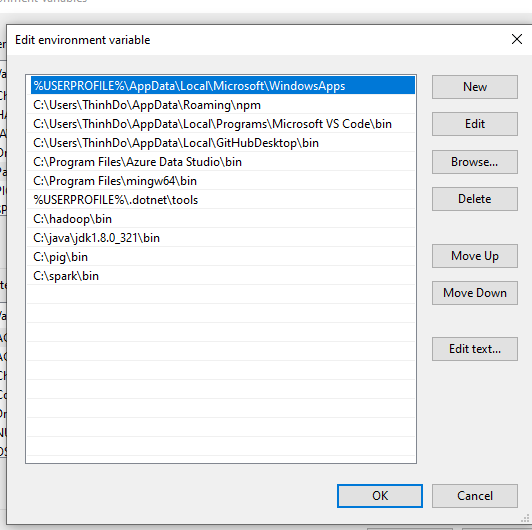
Text

Description automatically generated

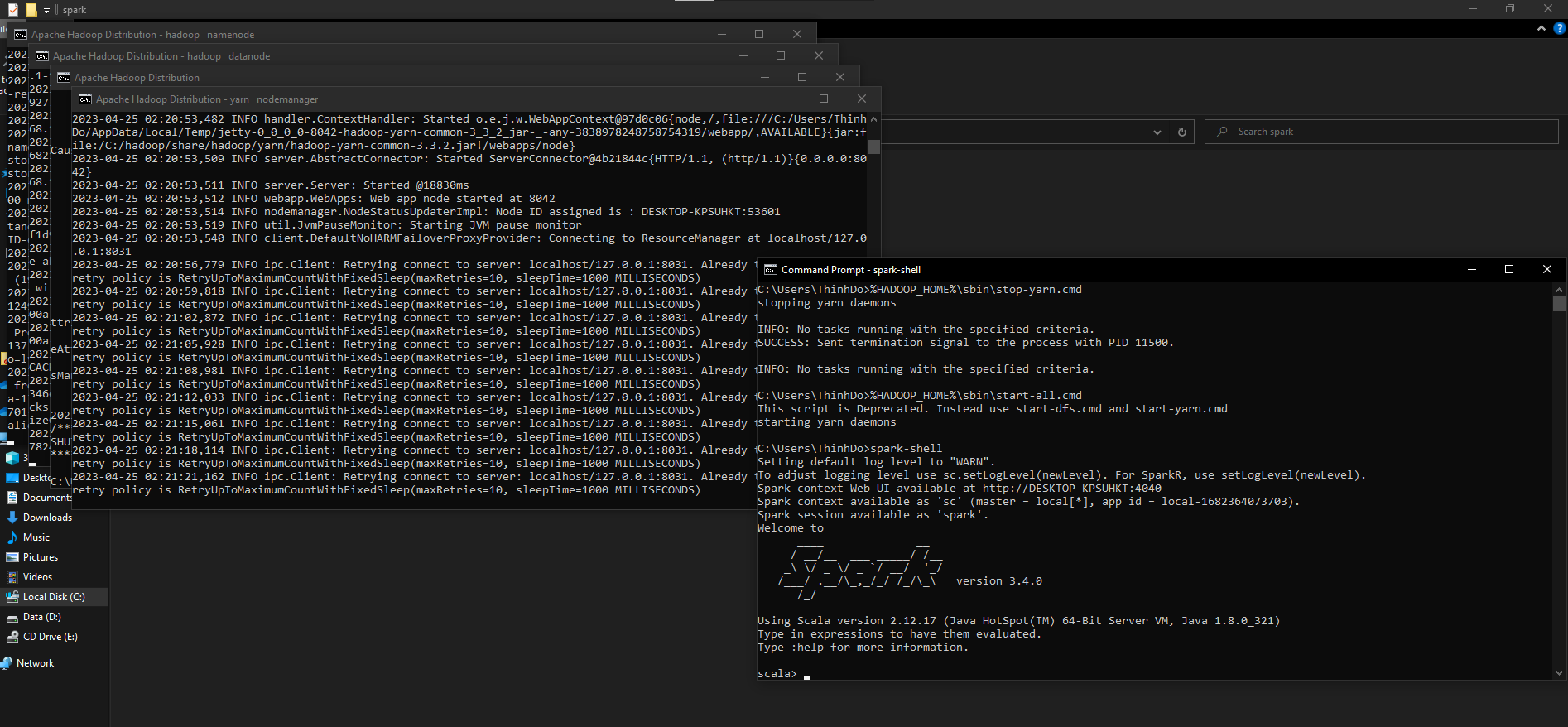
# 3. Apache Spark

## 3.1. Tải và cài đặt Hadoop Spark





## 3.2. Chạy Spark



Shape

Description automatically generated with medium confidence

A screenshot of a computer

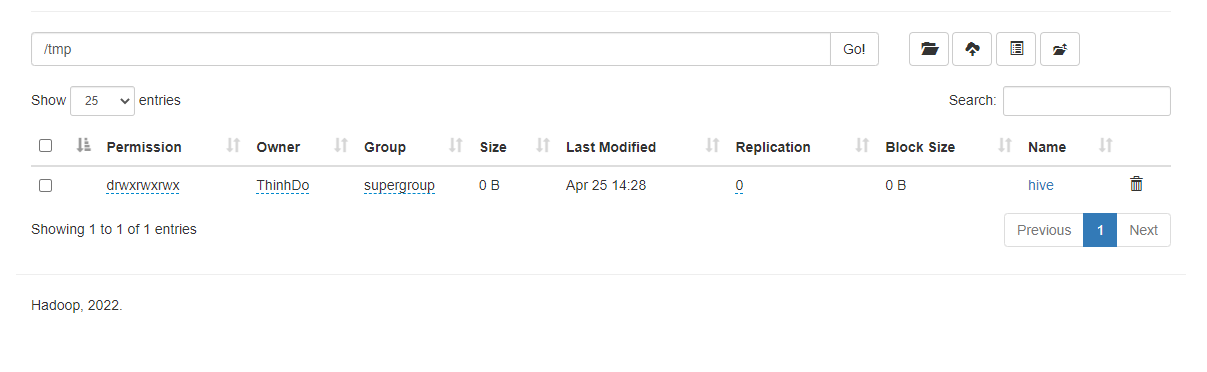
Description automatically generated with medium confidence

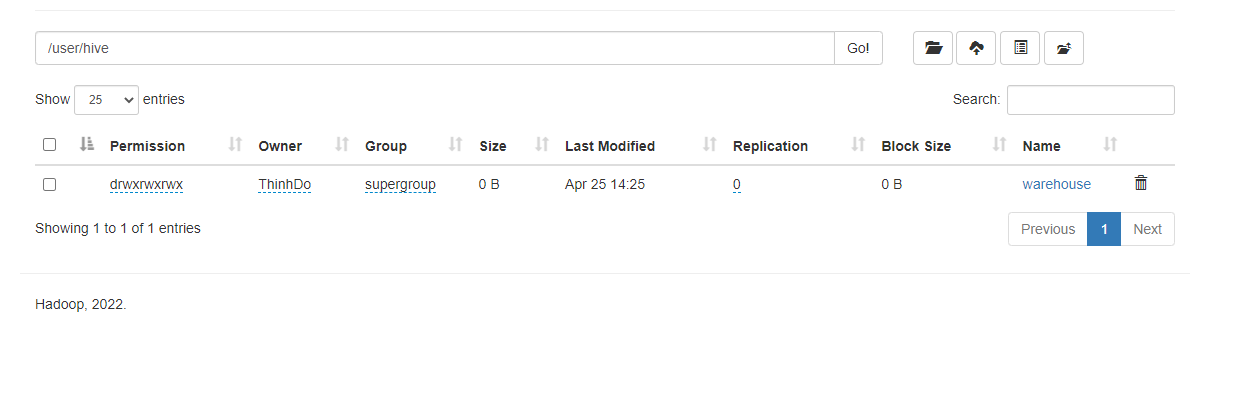
# 4. Apache Hive

## 4.1. Tạo thư mục lưu trữ trên hdfs cho hive

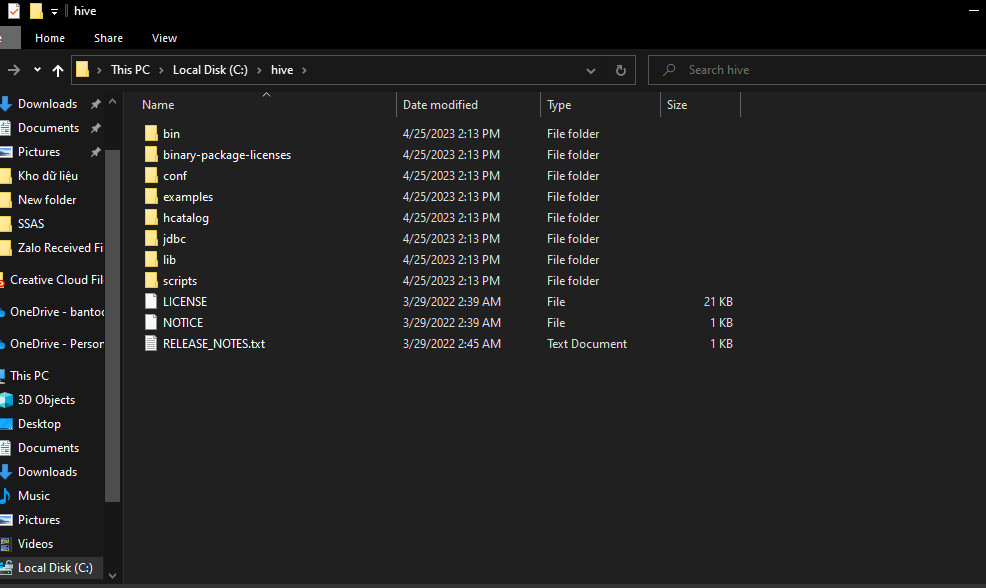
Graphical user interface, text, application, email

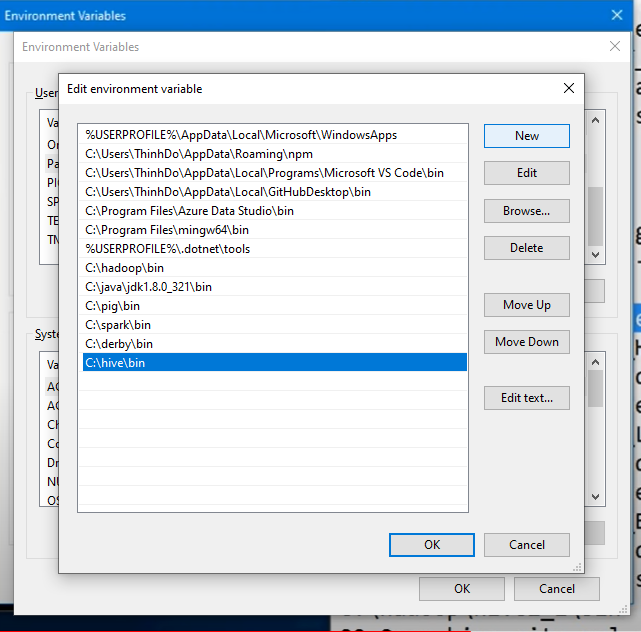
Description automatically generated

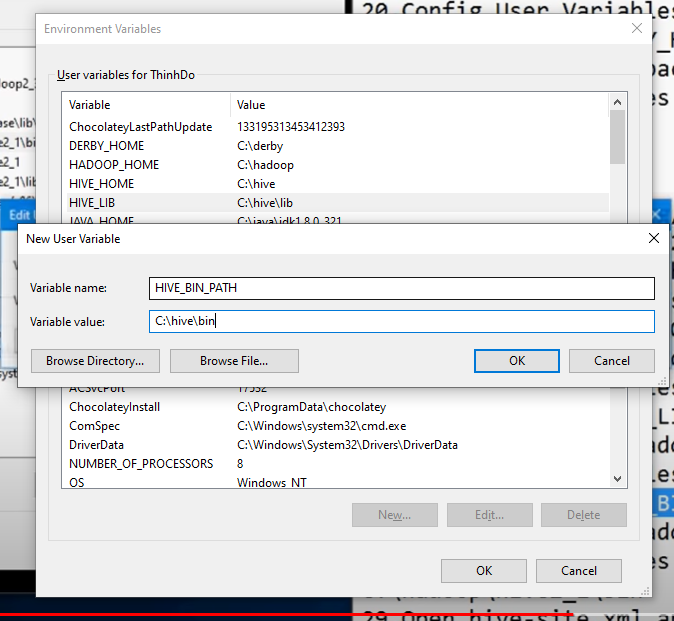


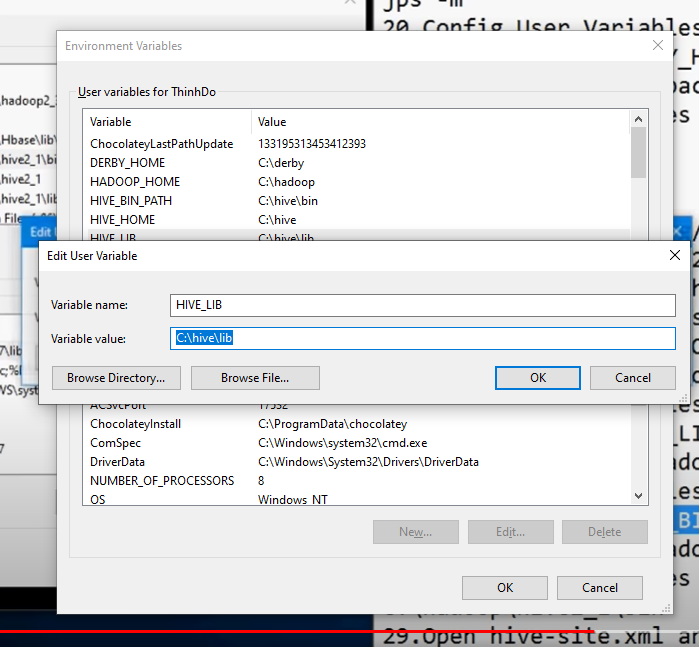


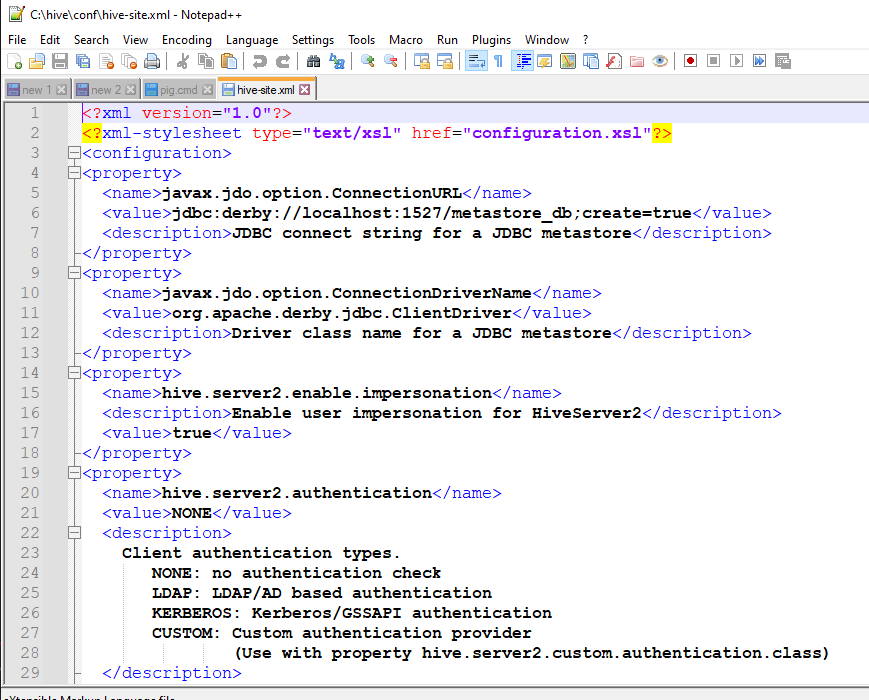
## 4.2. Tải và cài đặt hive











## 4.3. Tải và cài đặt cygwin hổ trợ việc chạy hive

sau khi tải xong cygwin ta tạo một thư mục ‘cygdrive’ ở ổ đĩa C

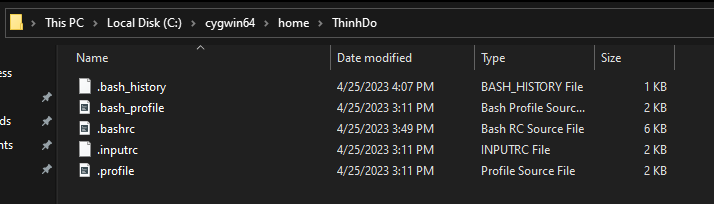
mở cmd lên chạy lệnh sau để cài drive cho cygwin:

mklink /J C:\cygdrive\c C:\

Graphical user interface, text, application

Description automatically generated

## 4.4. Cấu hình file ~/.bashrc



thêm vào các dòng sau ở cuối file

export HADOOP\_HOME='/cygdrive/c/hadoop'

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

export HIVE\_HOME='/cygdrive/c/hive'

export PATH=$PATH:$HIVE\_HOME/bin

export HADOOP\_CLASSPATH=$HADOOP\_CLASSPATH:$HIVE\_HOME/lib/\*.jar

Graphical user interface, text, application

Description automatically generated

vào cygwin chạy câu lệnh sau để cập nhật bashrc: source ~/.bashrc

Text

Description automatically generated with medium confidence

## 4.5. Cài đặt derby

\*coppy hết tất cả thư mục có trong derby/lib cho vào hive/lib\*

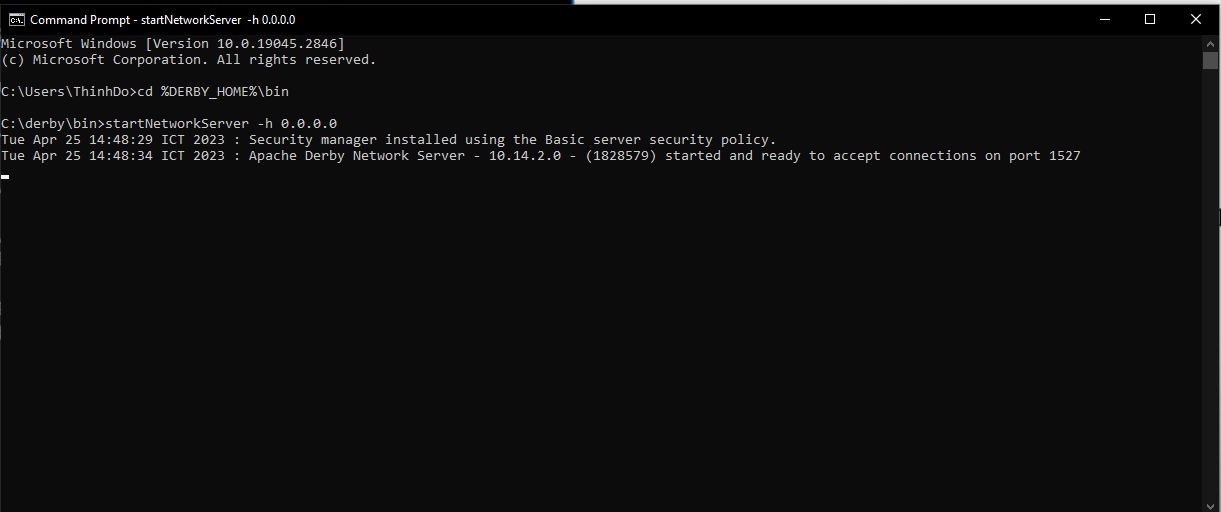
A screenshot of a computer

Description automatically generated with medium confidence

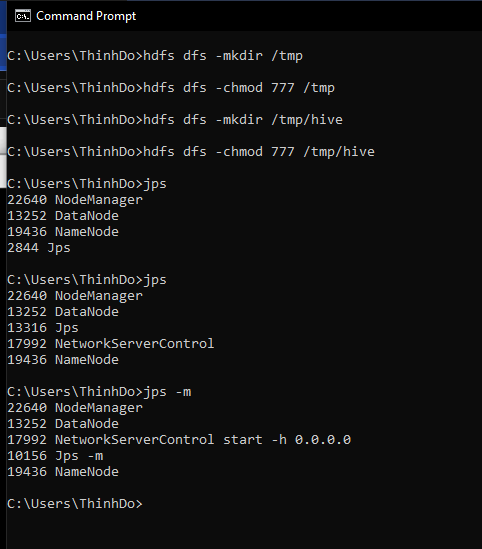
Graphical user interface, text, application, email

Description automatically generated

chạy derby



check jps



## 4.6. Chạy Hive

chọn dbtype cho hive : $schematool -initSchema -dbType derby

chạy hiveserver2: $ hive --service hiveserver2 start

chạy hive (ở một cygwin riêng): $hive

