**NHÓM 10 BIG DATA**

# THÀNH VIÊN:

* Trần Nguyễn Gia Long
* Trần Kim Thiện
* Nguyễn Thanh Danh

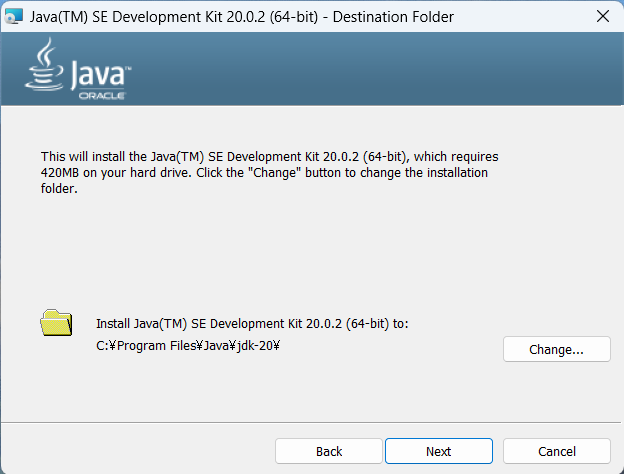
# CÀI ĐẶT HADOOP TRỰC TIẾP TRÊN WINDOWS

## 1. Cài đặt Java Development Kit phiên bản 20.0.2

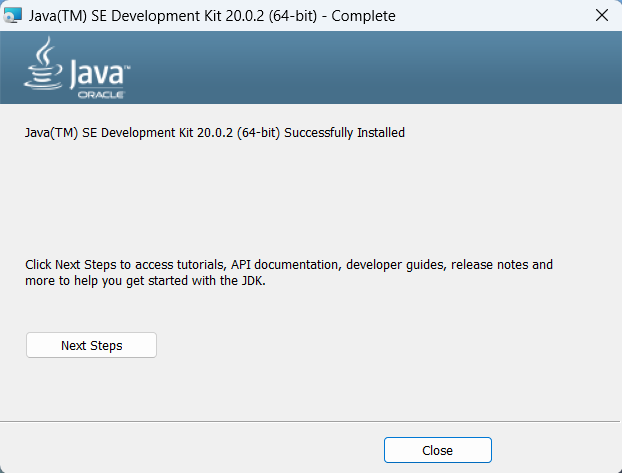
Đầu tiên, tải Java Development Kit 20.0.2 để cài đặt trên máy.



Tiếp tục chọn Next.

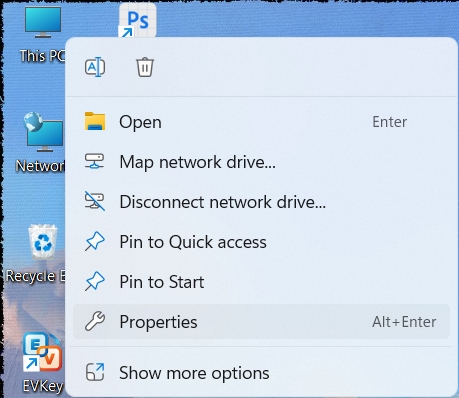


Khi cài đặt hoàn tất sẽ hiện ra thông báo như sau:

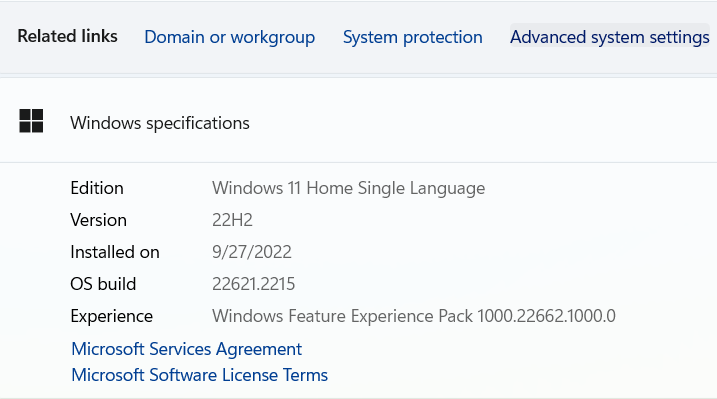


### a. Thiết lập biến môi trường cho Java JDK

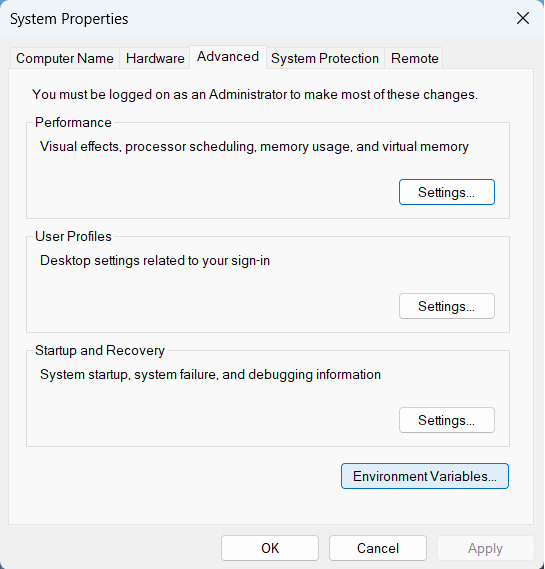
Bấm chuột phải vào This PC ⇒ Properties



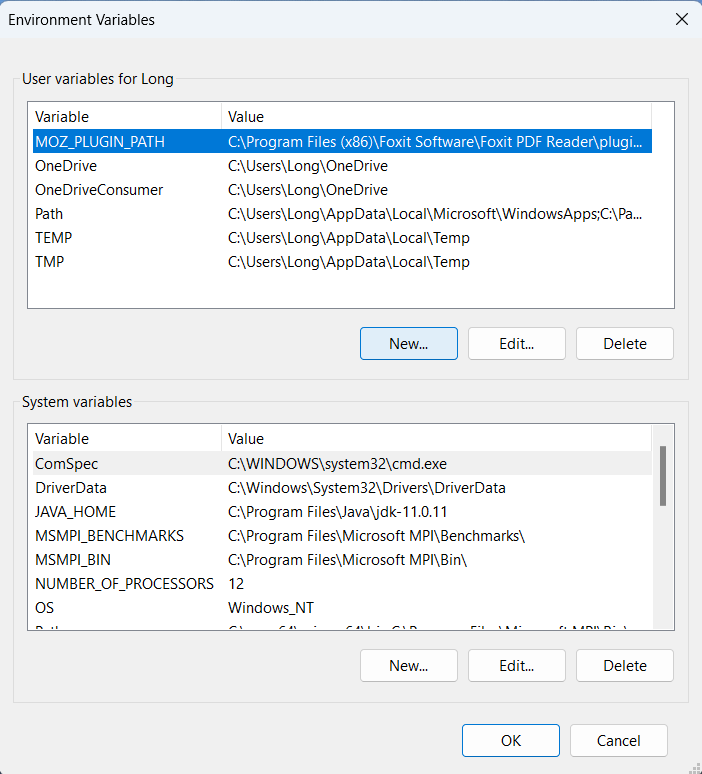
Vào Advanced system setting.



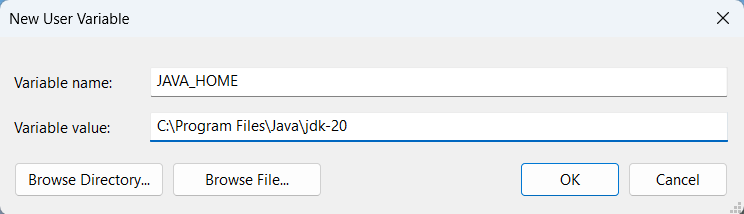
Sau khi vào System Propertise ⇒ Enviroment Variables.

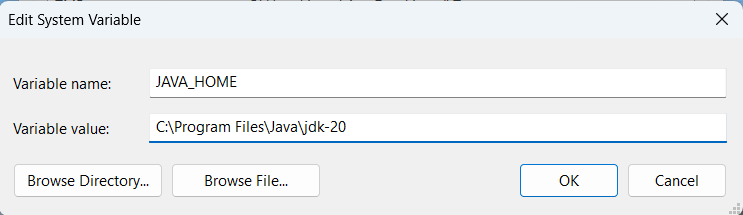


Sẽ xuất hiện màn hình Enviroment Variables:

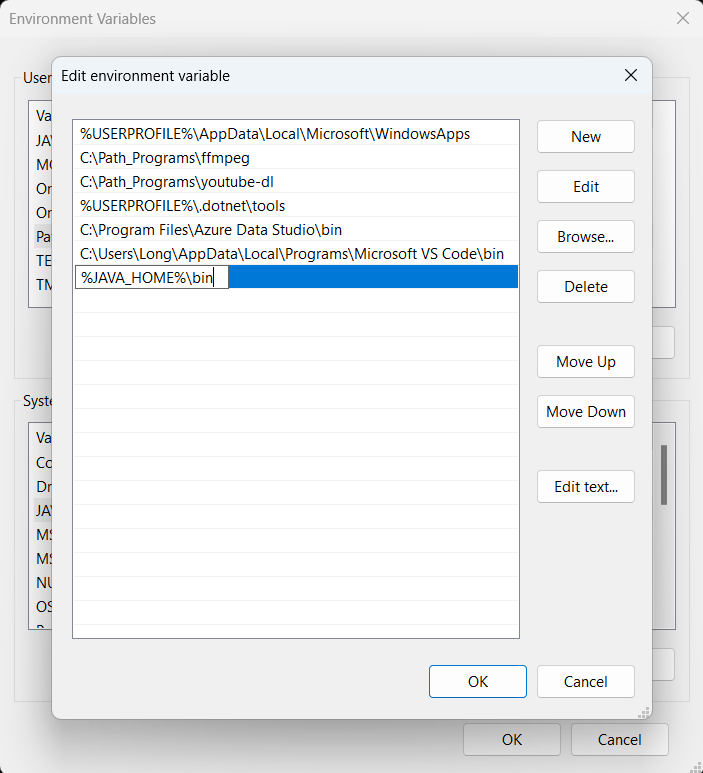


Trong mục **User** và **System Variables** ta cấu hình JAVA\_HOME bằng cách chọn nút *New…*



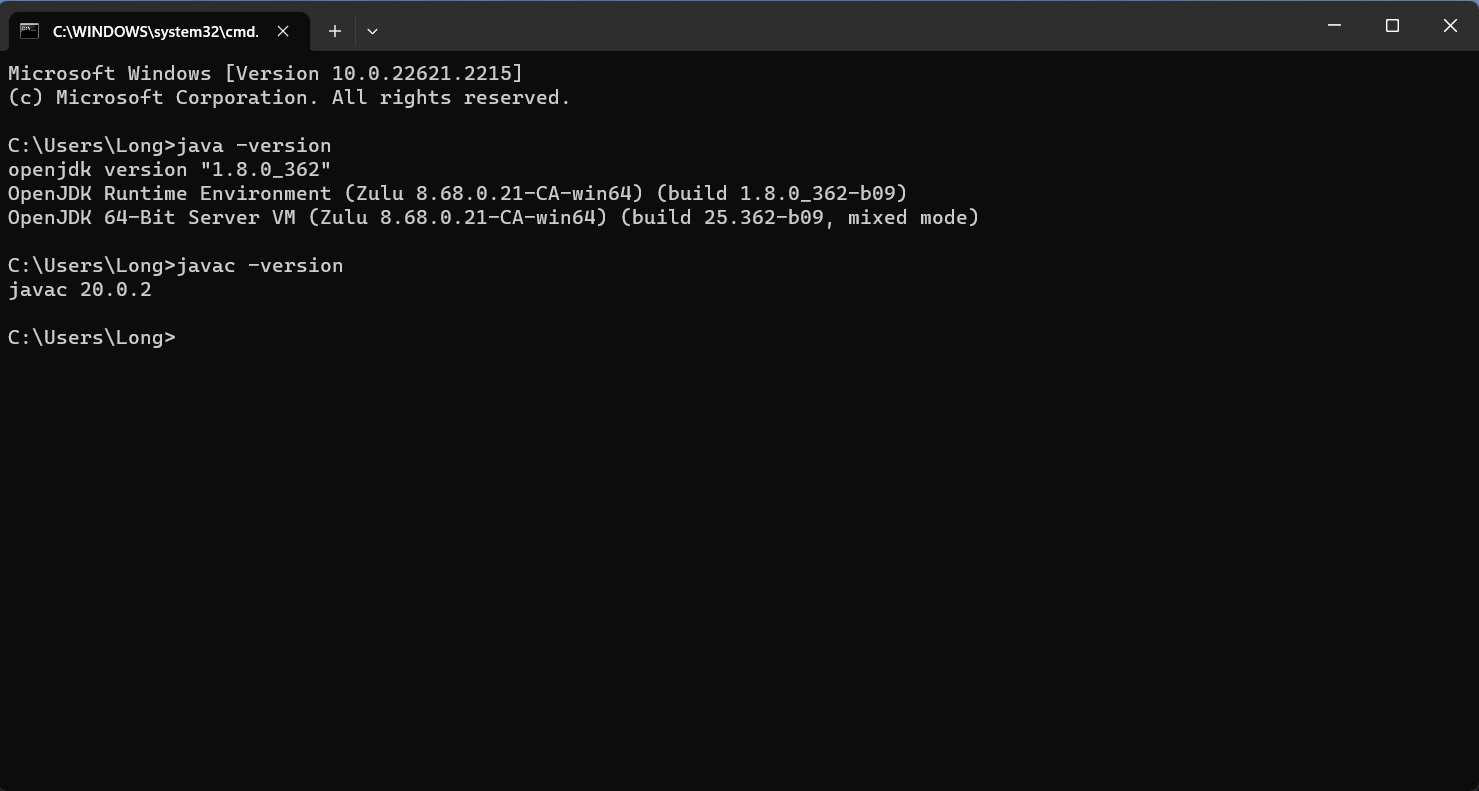


Tiếp tục chỉnh Path cho JDK cho hai mục trên.



Edit thêm lệnh: **%JAVA\_HOME%\bin**

Kiểm tra lại bằng cách dùng *cmd* và bấm lệnh **java -version.**



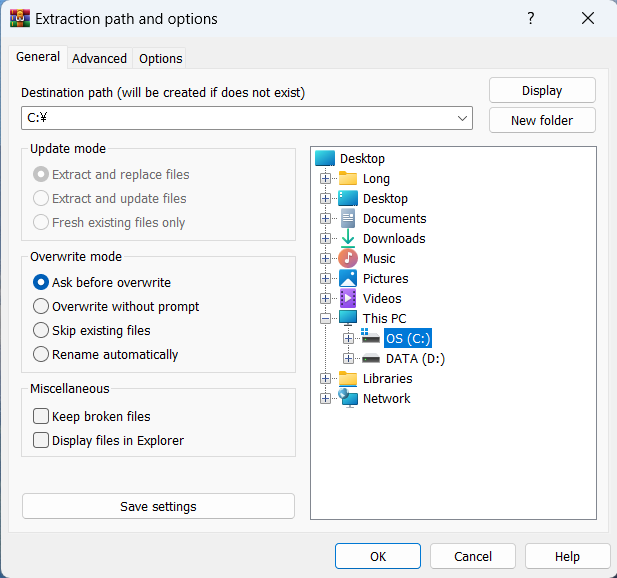
*⇒Phiên bản đã hiện, chứng tỏ java đã cài hoàn thành.*

## 2. Tải và cài đặt Hadoop.

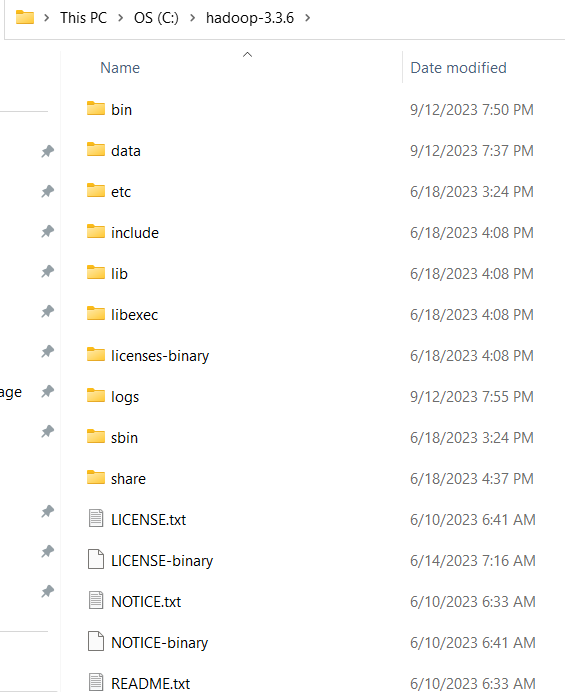
Tải Hadoop về từ trang Apache Hadoop với tên file là *hadoop-3.3.6.tar.gz*.



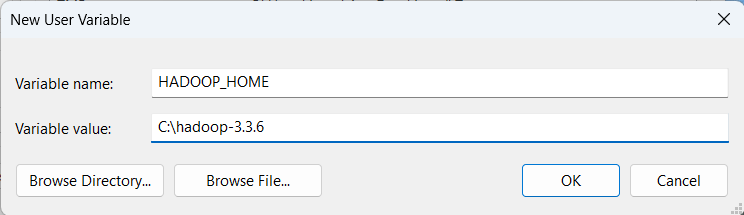
Sau đó giải nén bằng Winrar, và để cho thuận tiện, ta sẽ giải nén vào ổ C.

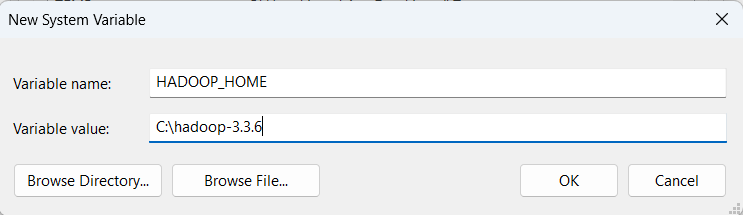


Kết quả:



Tiếp đến là phần cài đặt môi trường cho Hadoop. Tương tự như JDK.

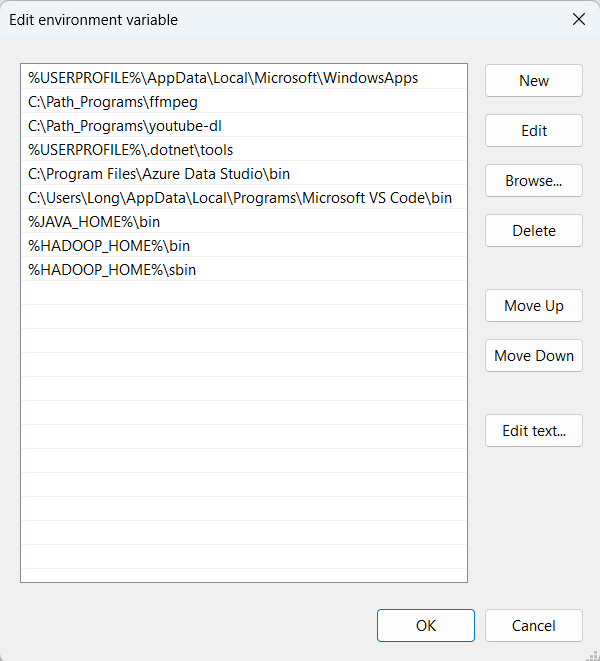




Và bước cuối để cài đặt môi trường Hadoop hoàn tất bằng cách thêm đường dẫn bằng hai biến.

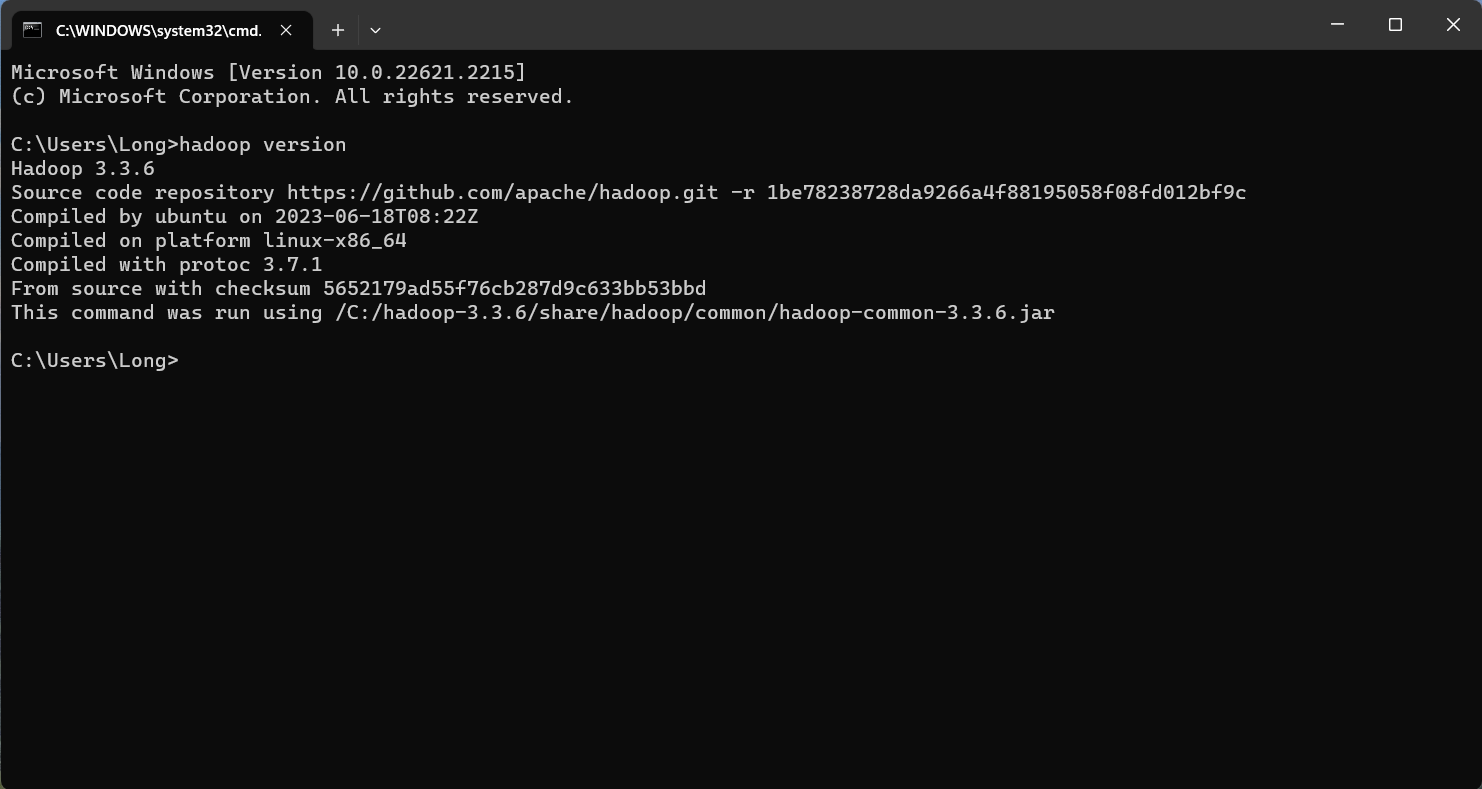
*%HADOOP\_HOME%\bin*

*%HADOOP\_HOME%\sbin*



Nhấn Ok để đóng tất cả các cửa sổ.

Mở cmd để kiểm tra lại bằng lệnh **hadoop version**.



## 3. Cấu hình các tập tin cho Hadoop

Trong thư mục C:/Hadoop-3.3.6/etc/hadoop lần lượt chỉnh sửa các file:

* core-site.xml
* mapred-site.xml
* hdfs-site.xml
* yarn-site.xml
* hadoop-env.cmd

Cấu hình **core-site.xml** như dưới đây:

*<configuration>*

*<property>*

*<name>fs.defaultFS</name>*

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

*</property>*

*</configuration>*

Cấu hình **mapred-site.xml** như dưới đây:

*<configuration>*

*<property>*

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

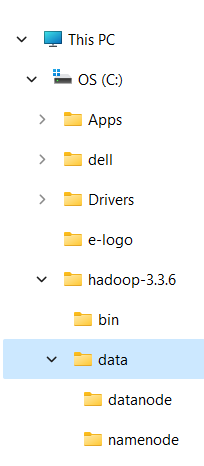
*<value>yarn</value>*

*</property>*

*</configuration>*

Cấu hình **hdfs-site.xml** như dưới đây:

* Tạo thư mục “data” trong “C:/Hadoop-3.3.6”
* Tạo thư mục con “datanode” trong “C:/Hadoop-3.3.6/data”
* Tạo thư mục con “namenode” trong “C:/Hadoop-3.3.6/data”



*<configuration>*

*<property>*

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

*<value>1</value>*

*</property>*

*<property>*

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

*<value>/hadoop-3.3.6/data/namenode</value>*

*</property>*

*<property>*

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

*<value>/hadoop-3.3.6/data/datanode</value>*

*</property>*

*</configuration>*

Cấu hình **yarn-site.xml** như dưới đây:

*<configuration>*

*<property>*

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

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

*</property>*

*<property> <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>*

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

*</property>*

*</configuration>*

Cấu hình **hadoop-env.cmd**:

set JAVA\_HOME=%JAVA\_HOME%

sửa %JAVA\_HOME% thành đường dẫn cài JDK trong ổ C:

set JAVA\_HOME= C:\Progra~1\Java\jdk-20

## 4. Cập nhật các Hadoop Configurations

Tải: https://github.com/s911415/apache-hadoop-3.1.0-winutils

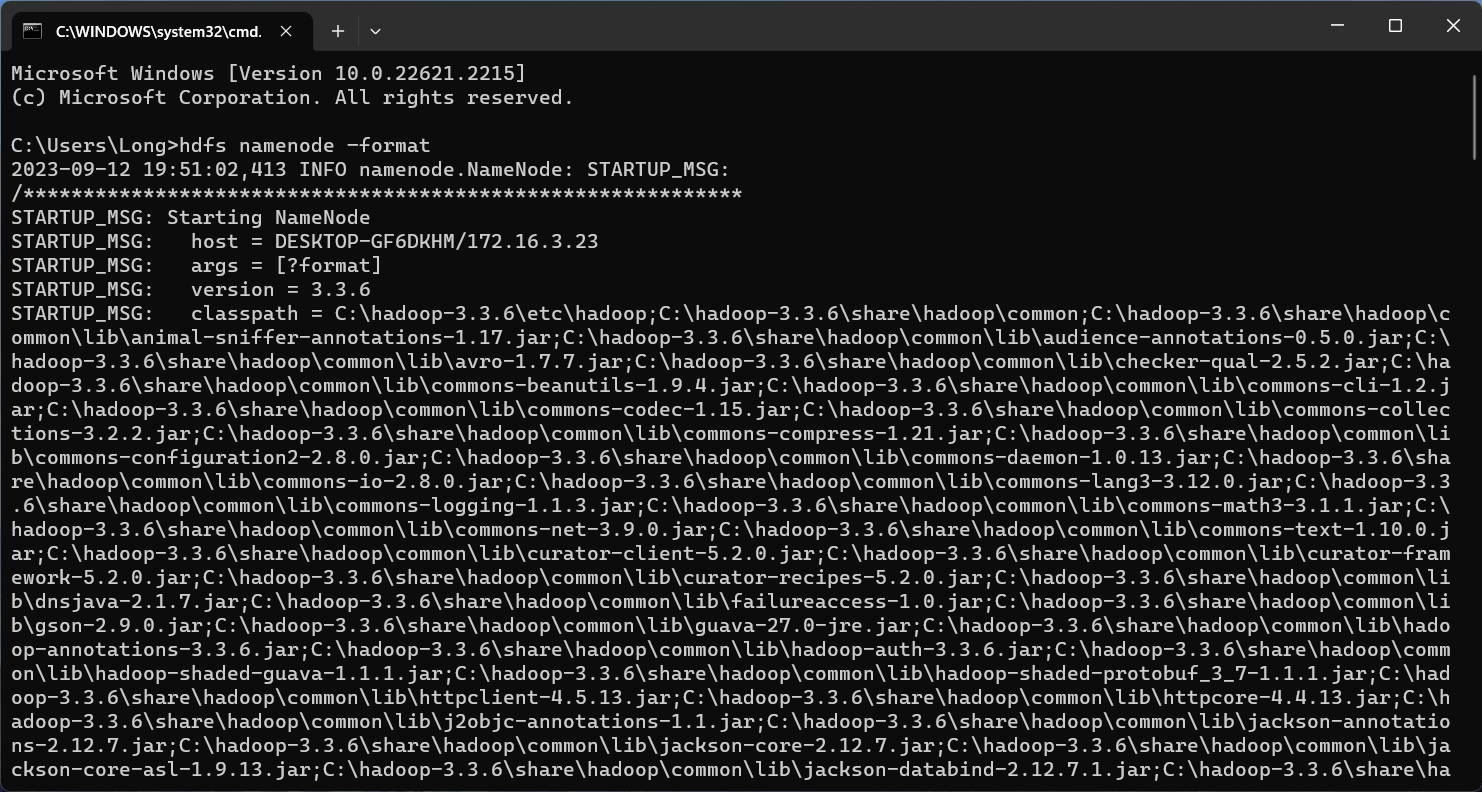
Tải về giải nén ra thấy thư mục bin ở bên trong

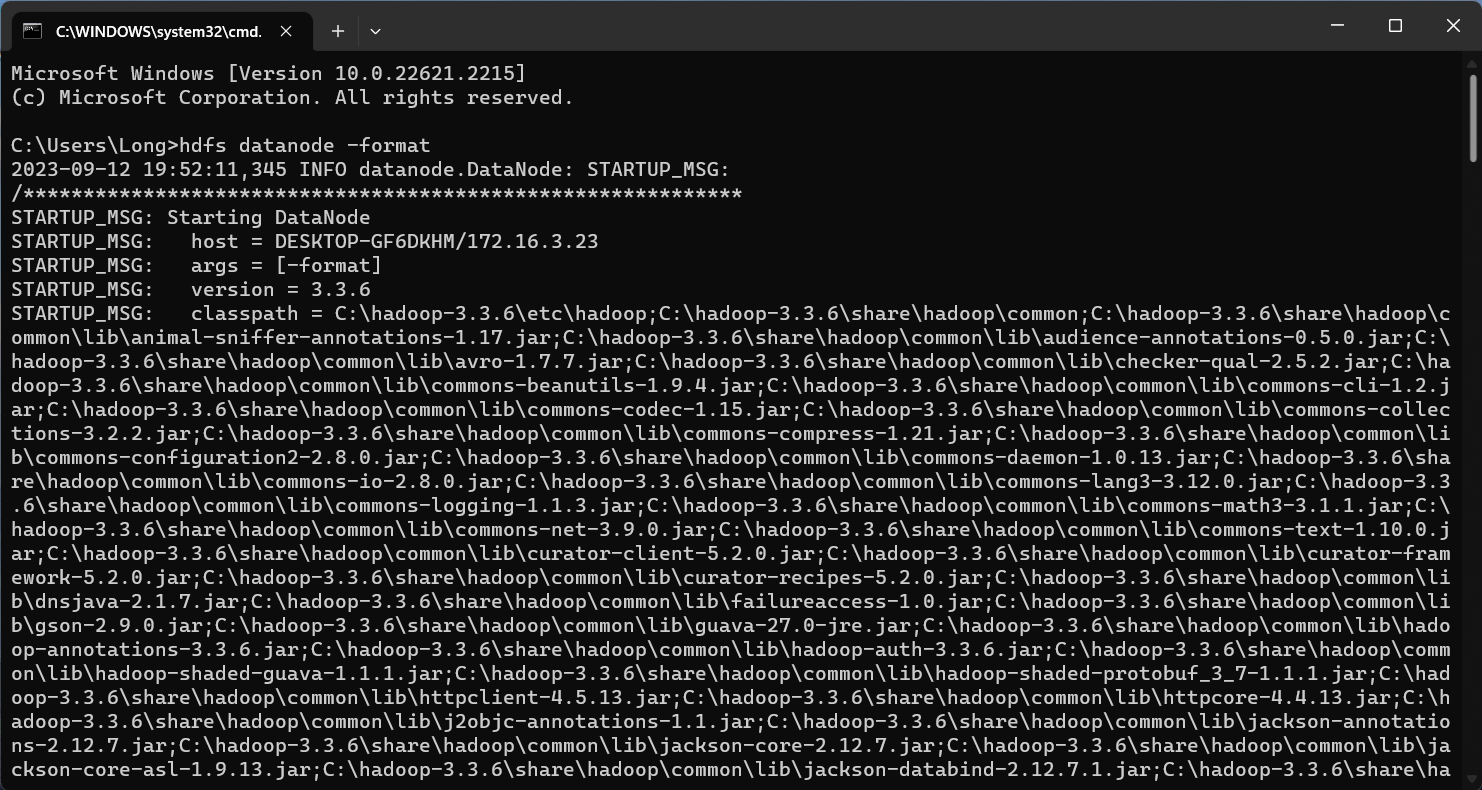
Chép đè thư mục bin này trong thư mục bin của C:\hadoop-3.3.0\bin

Sau đó format lại **namenode** và **datanode** trong cmd.

*hdfs namenode –format*

*hdfs datanode -format*

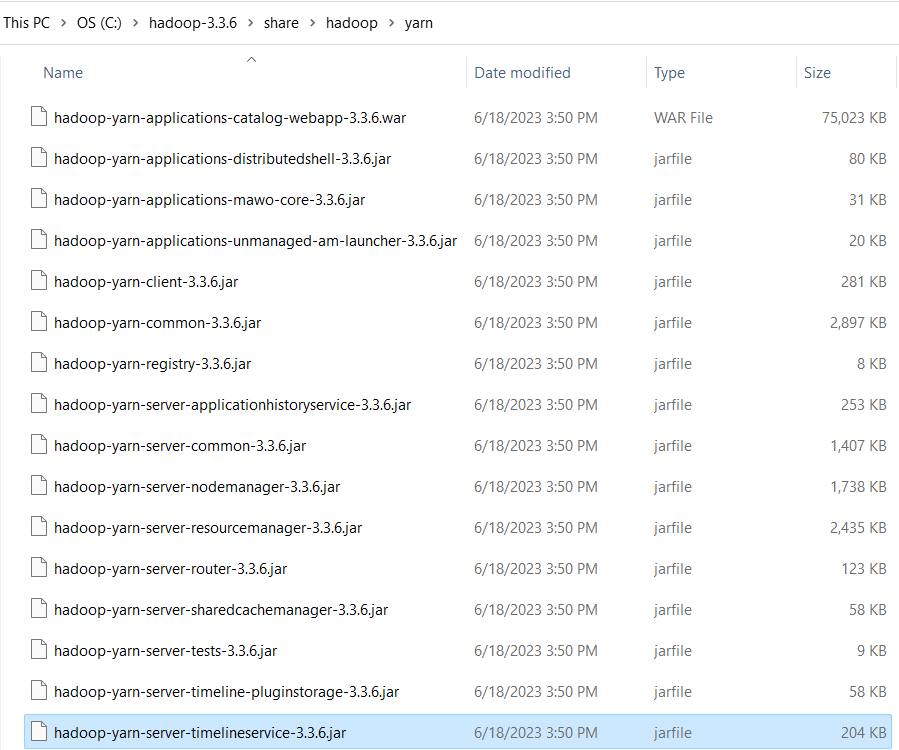




**\*Tiếp theo sao chép file:**

“C:/hadoop-3.3.0/share/hadoop/yarn/timelineservice/hadoop-yarn-server-timelineservice-3.3.0.jar”

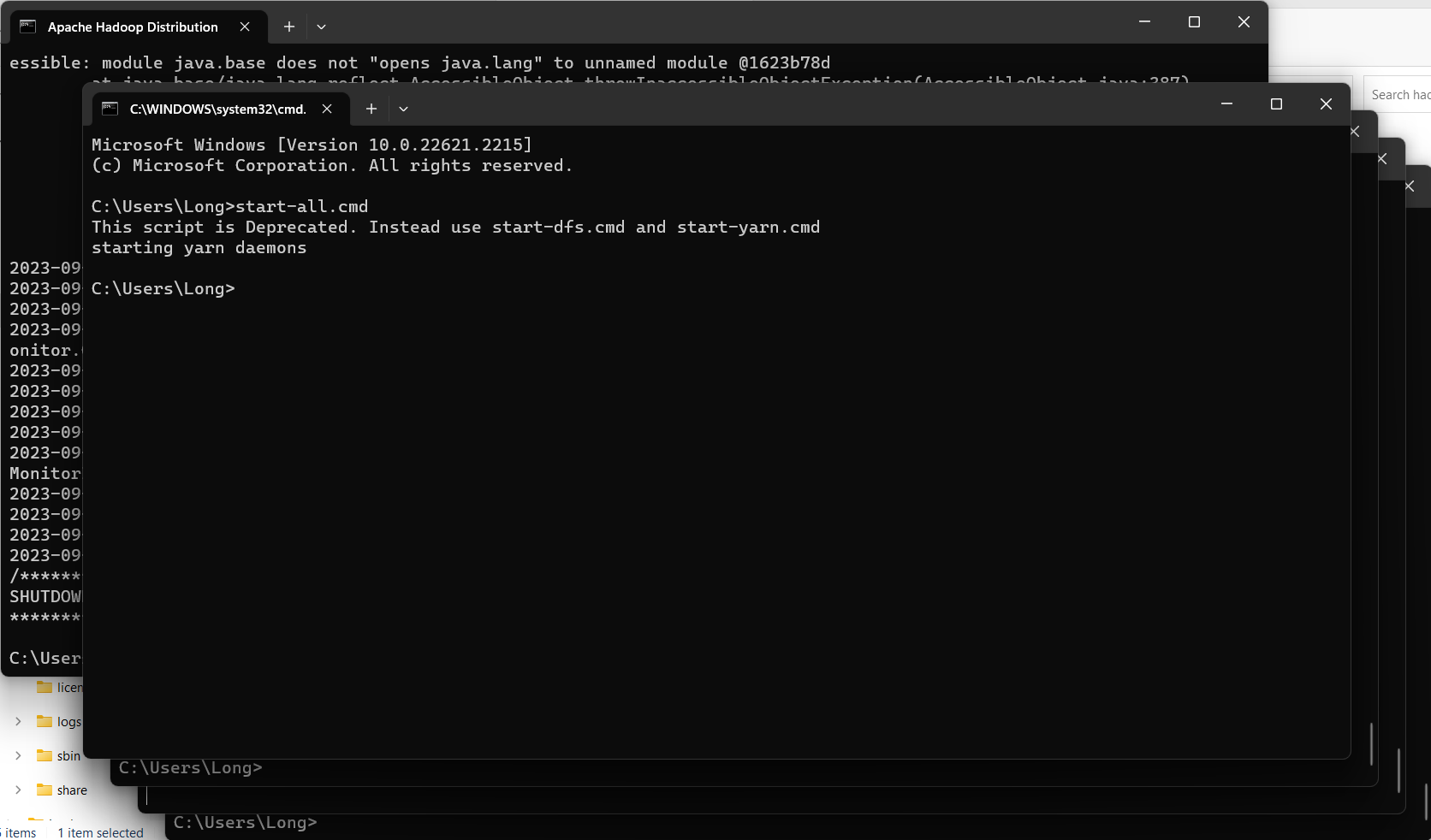
🡪 “C:/hadoop-3.3.0/share/hadoop/yarn/hadoop-yarn-server-timelineservice-3.3.0.jar”



Kiểm tra Hadoop trong thư mục C:/hadoop-3.3.0/sbin.

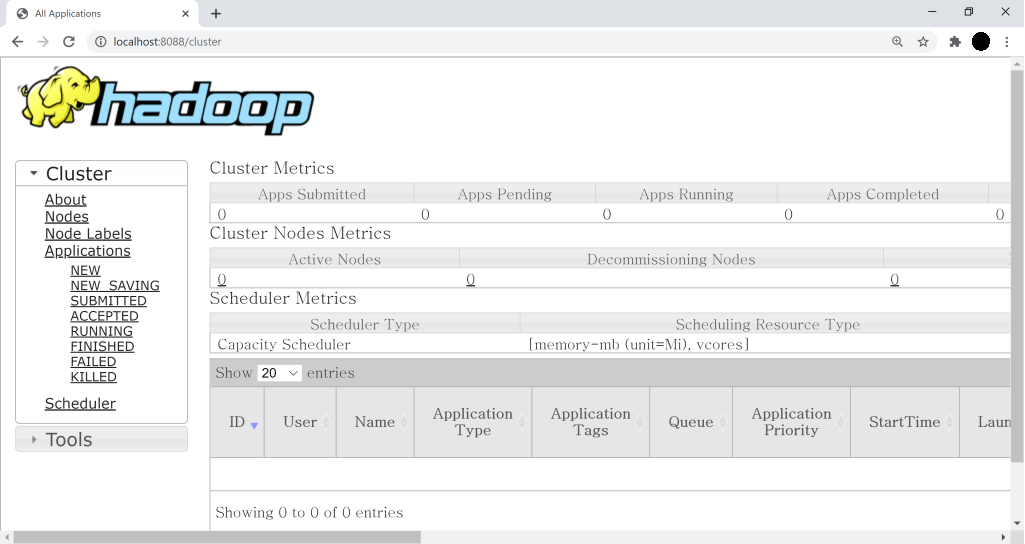
Sau đó gõ lệnh:

*start-all.cmd*

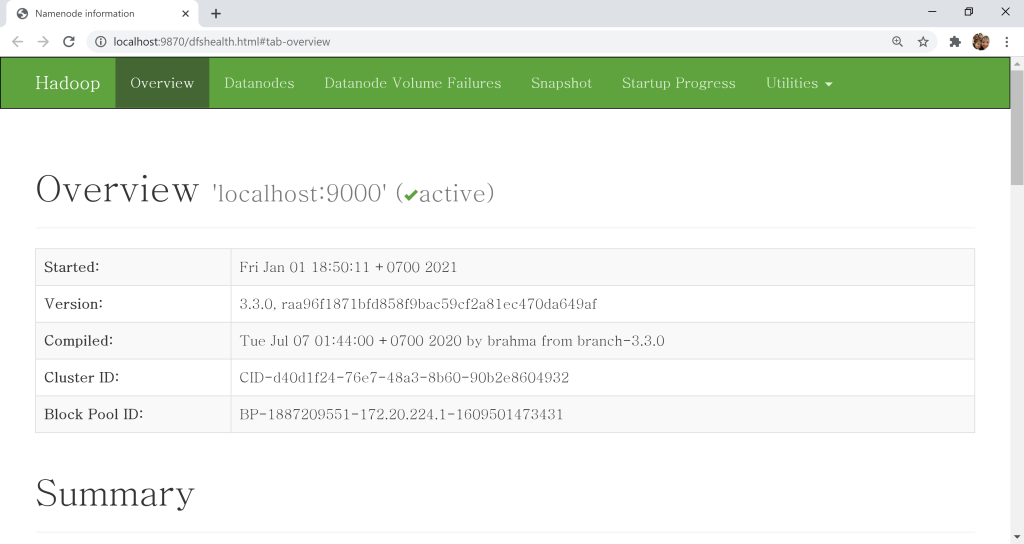


**NHƯ VẬY TA ĐÃ KHỞI ĐỘNG THÀNH CÔNG!**

Vào <http://localhost:8080>



Vào <http://localhost:9870>



# TÀI LIỆU THAM KHẢO

<https://duythanhcse.wordpress.com/2021/01/01/cai-dat-hadoop-tren-windows/>