**NHÓM 10 BIG DATA**

# THÀNH VIÊN:

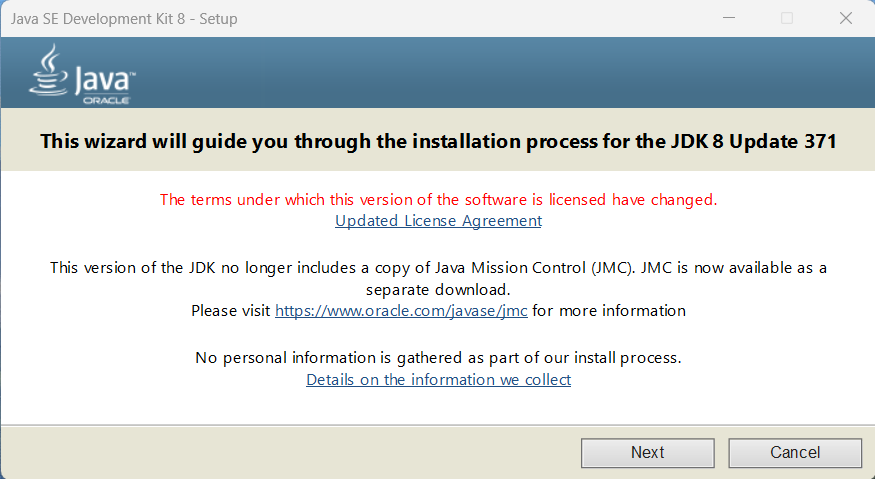
* Trần Nguyễn Gia Long
* Trần Kim Thiện
* Nguyễn Thanh Danh



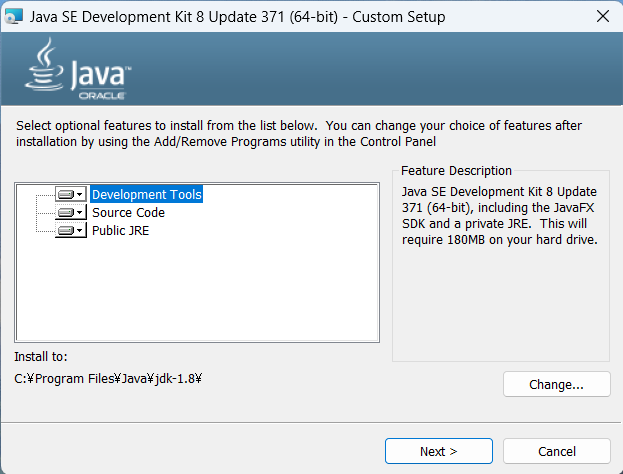
# CÀI ĐẶT HADOOP TRỰC TIẾP TRÊN WINDOWS

## 1. Cài đặt Java Development Kit phiên bản 8

Đầu tiên, tải Java Development Kit 8u371 để cài đặt trên máy.



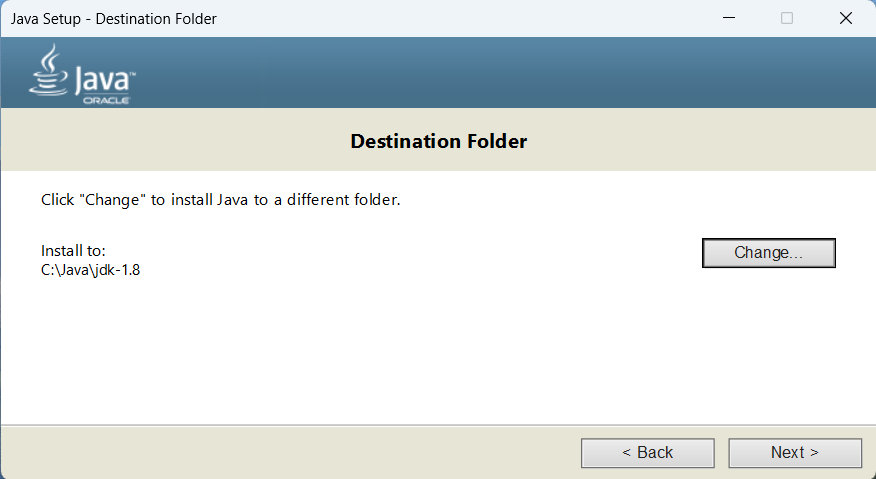
Tiếp tục chọn Next.



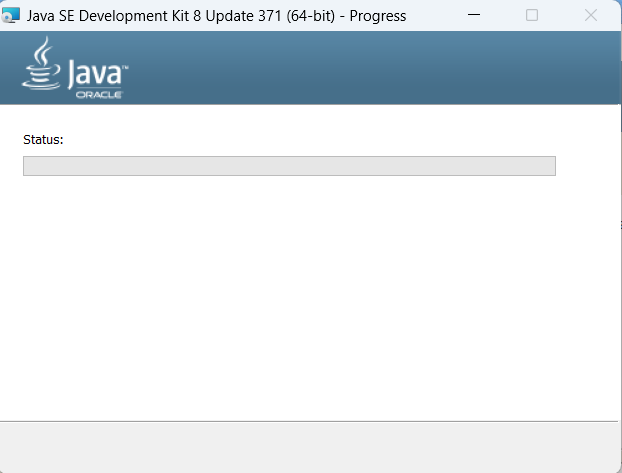
Chỉnh sửa đường dẫn đến thư mục tên ngắn và không khoảng trắng để thuận tiện cho việc tuỳ chỉnh Hadoop sau đó. Ở đây ta sẽ cài đặt ở đường dẫn *C:\Java\jdk-1.8*



Đợi chương trình cài đặt hoàn tất.



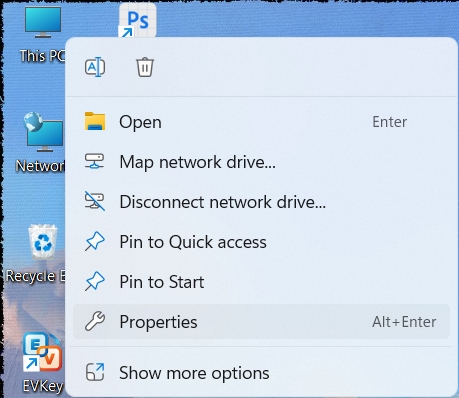
Ta vẫn sẽ điều chỉnh đường dẫn tương tự như trên nếu chương trình yêu cầu.



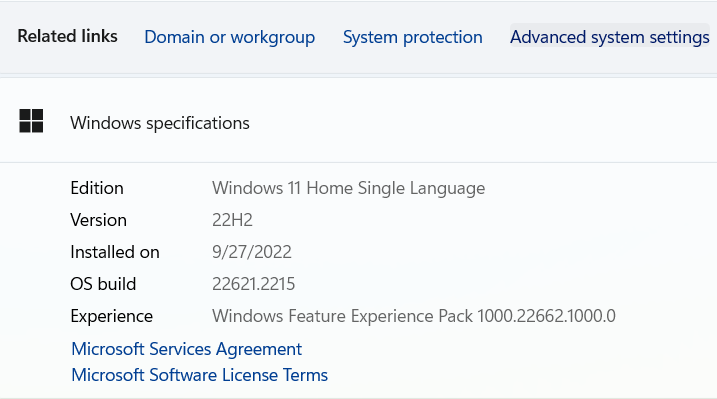
Ta tiếp tục chọn Next và chờ đợi cài đặt.

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

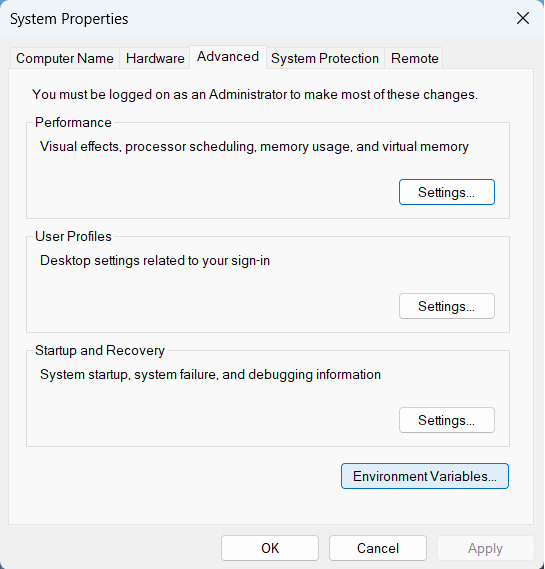
Bấm chuột phải vào This PC ⇒ Properties



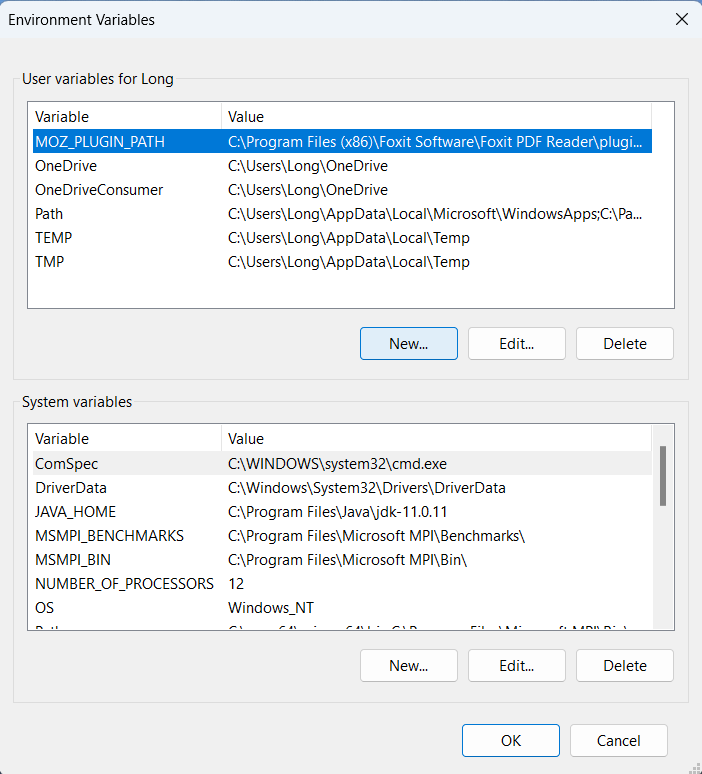
Vào Advanced system setting.



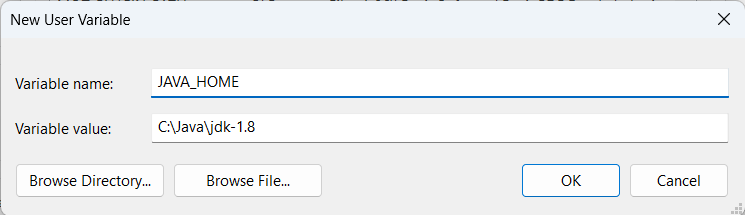
Sau khi vào System Propertise ⇒ Enviroment Variables.

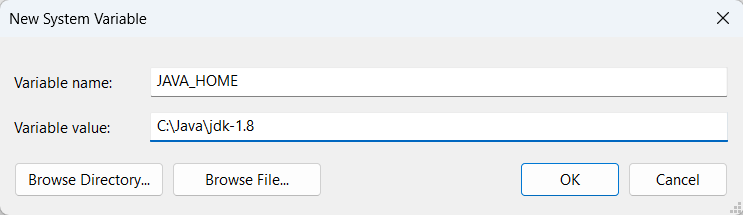


Sẽ xuất hiện màn hình Enviroment Variables:

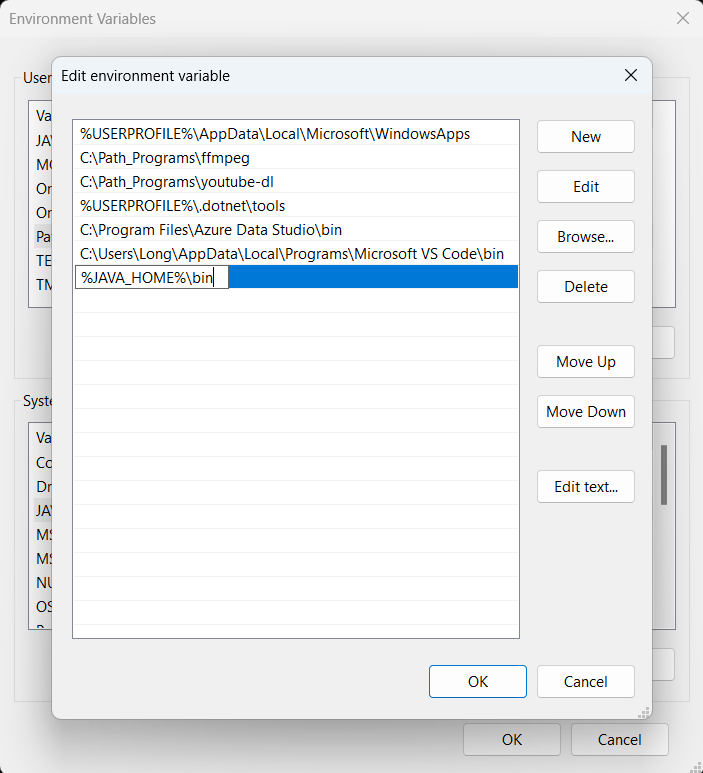


Trong mục **User** và **System Variables** ta cấu hình JAVA\_HOME bằng cách chọn nút *New…*  
Chọn đường dẫn là *C:\Java\jdk-1.8*



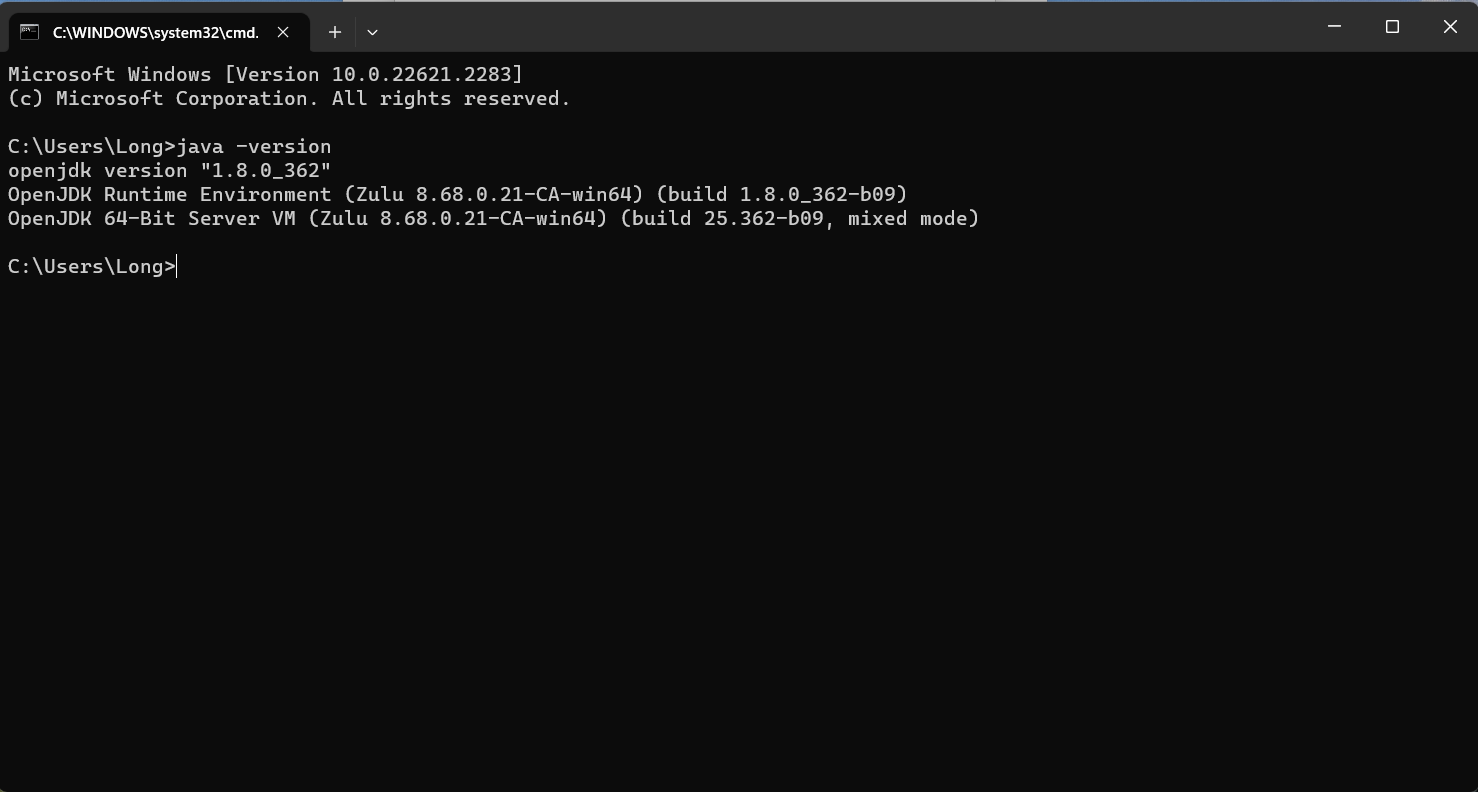


Tiếp tục chỉnh Path cho JDK cho hai mục trên.



Edit thêm lệnh: **%JAVA\_HOME%\bin**

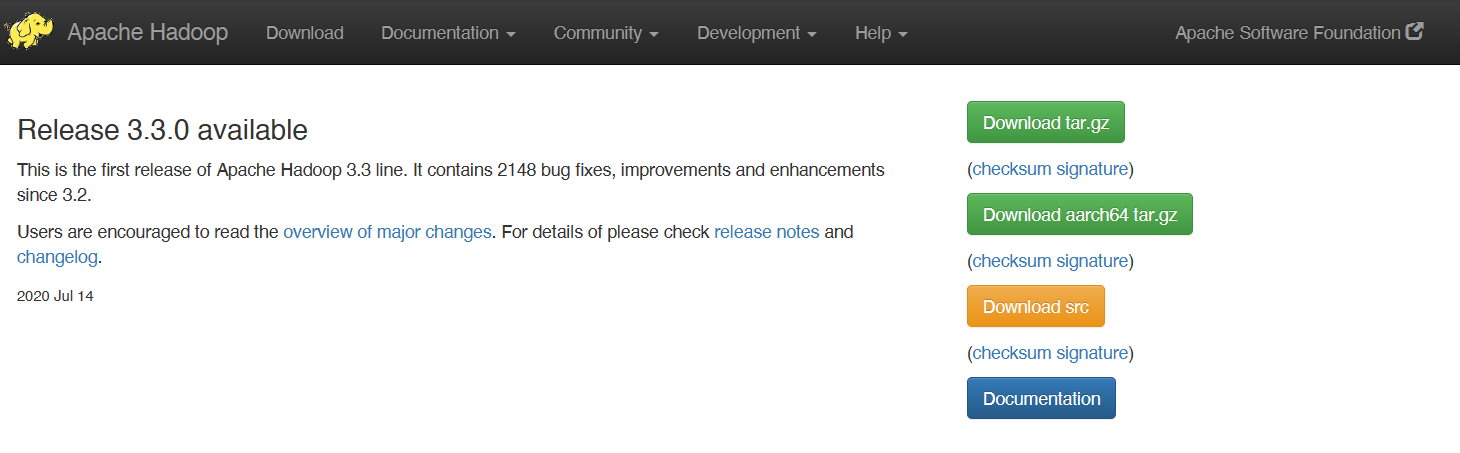
Kiểm tra lại bằng cách dùng *cmd* và bấm lệnh **java -version.**



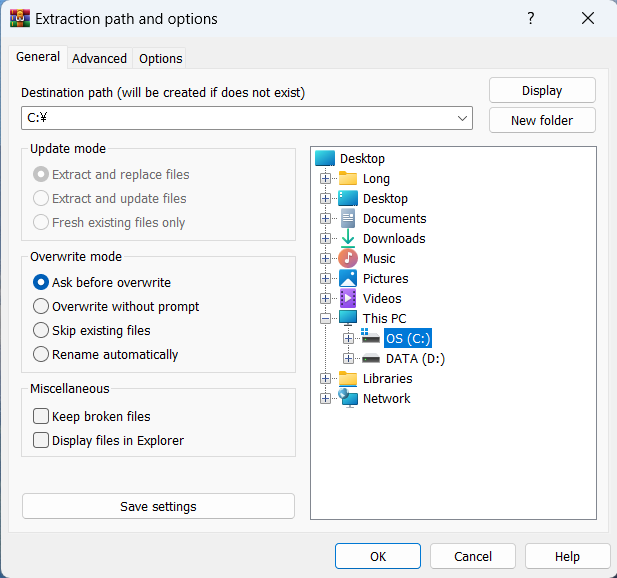
*⇒Phiên bản đã hiện, chứng tỏ java đã cài hoàn tất.*

## 2. Tải và cài đặt Hadoop.

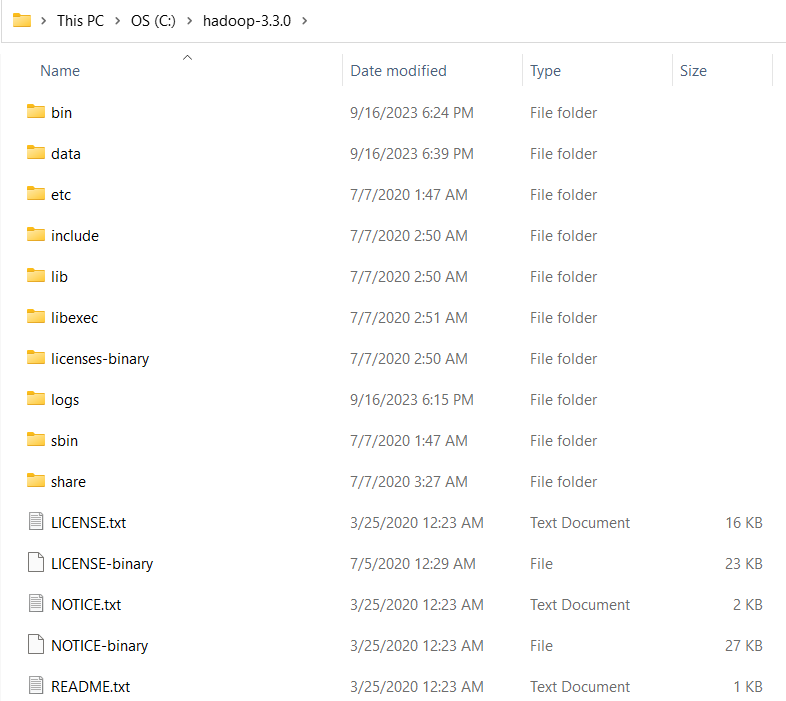
Tải Hadoop về từ trang Apache Hadoop với tên file là *hadoop-3.3.0.tar.gz*.



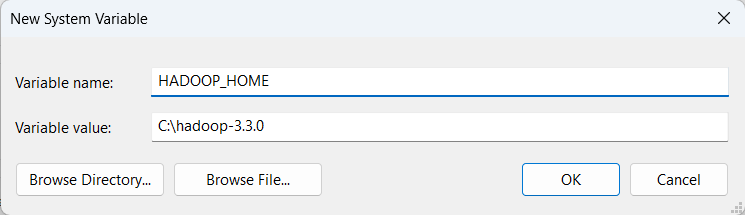
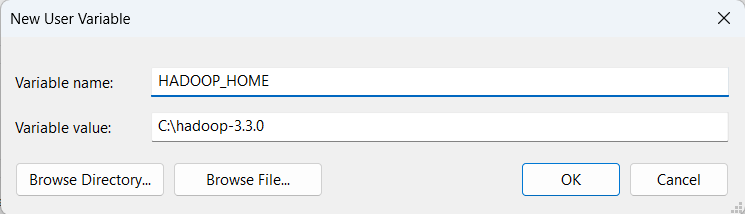
Sau đó giải nén bằng Winrar, và để cho thuận tiện, ta sẽ giải nén vào ổ C.



Kết quả:



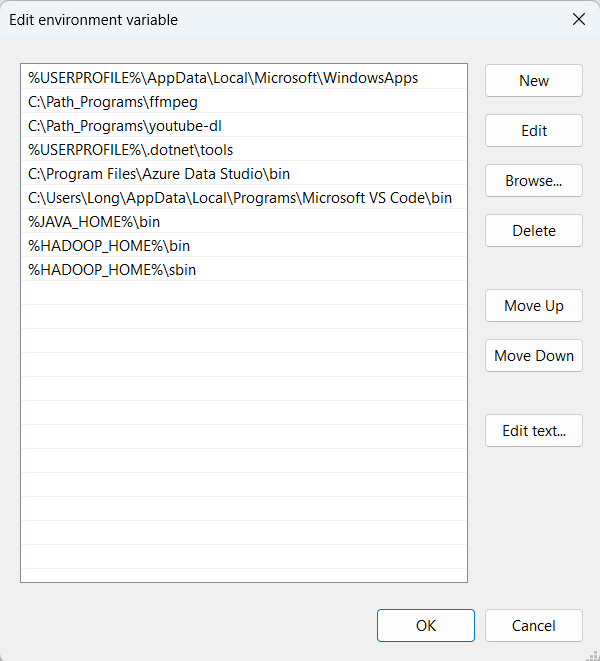
Tiếp đến là phần cài đặt môi trường cho Hadoop. Tương tự như JDK.



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

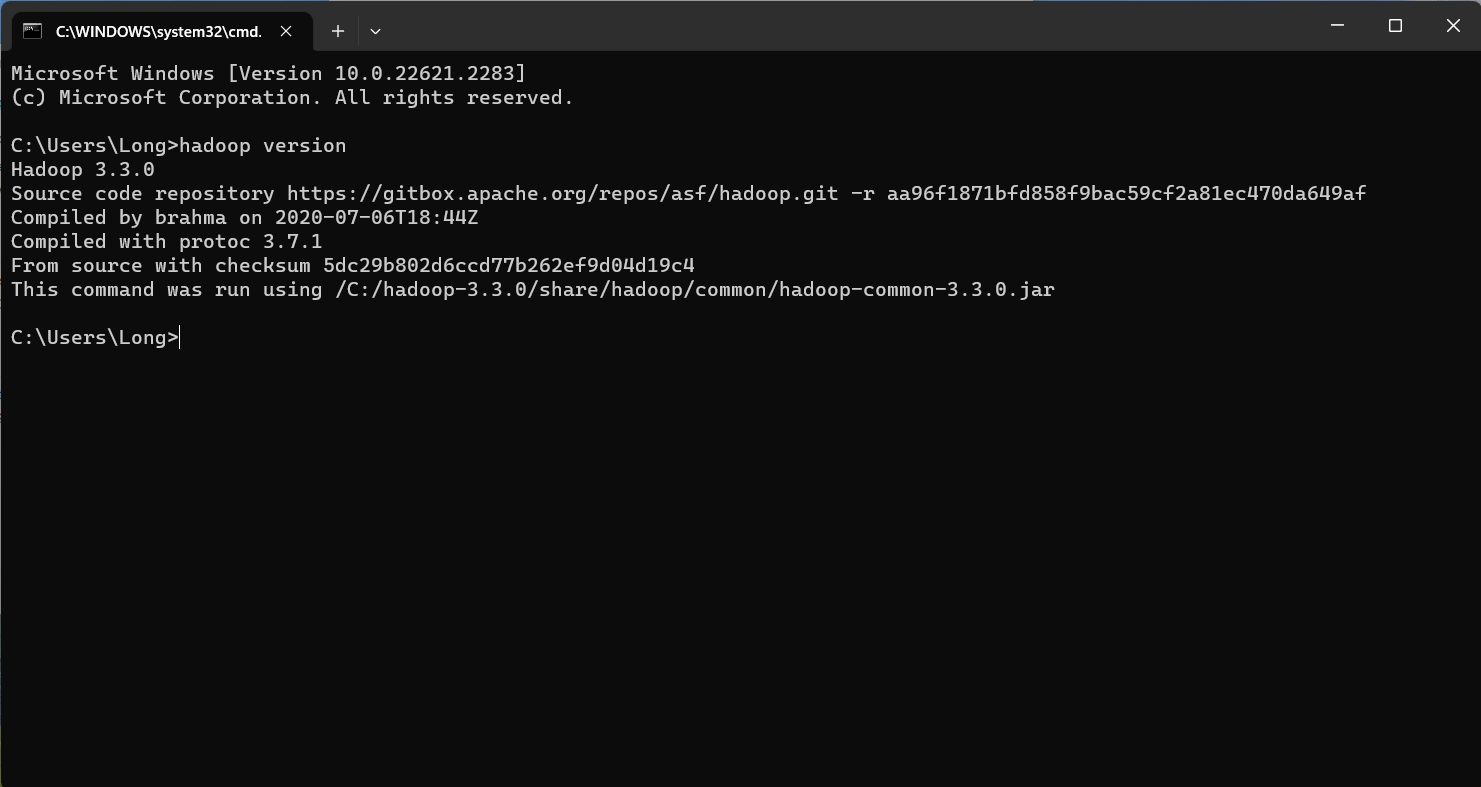
*%HADOOP\_HOME%\bin*

*%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-3.3.0/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.default.name</name>*

*<value>hdfs://0.0.0.0:19000</value>*

*</property>*

*</configuration>*

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

*<configuration>*

*<property>*

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

*<value>yarn</value>*

*</property>*

*<property>*

*<name>mapreduce.application.classpath</name>*

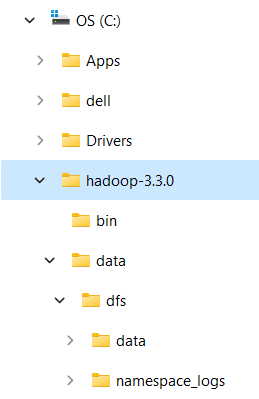
*<value>%HADOOP\_HOME%/share/hadoop/mapreduce/\*,%HADOOP\_HOME%/share/hadoop/mapreduce/lib/\*,%HADOOP\_HOME%/share/hadoop/common/\*,%HADOOP\_HOME%/share/hadoop/common/lib/\*,%HADOOP\_HOME%/share/hadoop/yarn/\*,%HADOOP\_HOME%/share/hadoop/yarn/lib/\*,%HADOOP\_HOME%/share/hadoop/hdfs/\*,%HADOOP\_HOME%/share/hadoop/hdfs/lib/\*</value>*

*</property>*

*</configuration>*

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

* Tạo thư mục “data” trong “C:/Hadoop-3.3.0”
* Tạo thư mục con “datanode” trong “C:/Hadoop-3.3.0/data/dfs/namespace\_logs”
* Tạo thư mục con “namenode” trong “C:/Hadoop-3.3.0/dfs/data”



*<configuration>*

*<property>*

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

*<value>1</value>*

*</property>*

*<property>*

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

*<!-- <value>file:///DIRECTORY 1 HERE</value> -->*

*<value>file:///* *C:/hadoop-3.3.0/data/dfs/namespace\_logs</value>*

*</property>*

*<property>*

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

*<!-- <value>file:///DIRECTORY 2 HERE</value> -->*

*<value>file:///* *C/hadoop-3.3.0/data/dfs/data</value>*

*</property>*

*</configuration>*

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

*<configuration>*

*<property>*

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

*<value>yarn</value>*

*</property>*

*<property>*

*<name>mapreduce.application.classpath</name>*

*<value>%HADOOP\_HOME%/share/hadoop/mapreduce/\*,%HADOOP\_HOME%/share/hadoop/mapreduce/lib/\*,%HADOOP\_HOME%/share/hadoop/common/\*,%HADOOP\_HOME%/share/hadoop/common/lib/\*,%HADOOP\_HOME%/share/hadoop/yarn/\*,%HADOOP\_HOME%/share/hadoop/yarn/lib/\*,%HADOOP\_HOME%/share/hadoop/hdfs/\*,%HADOOP\_HOME%/share/hadoop/hdfs/lib/\*</value>*

*</property>*

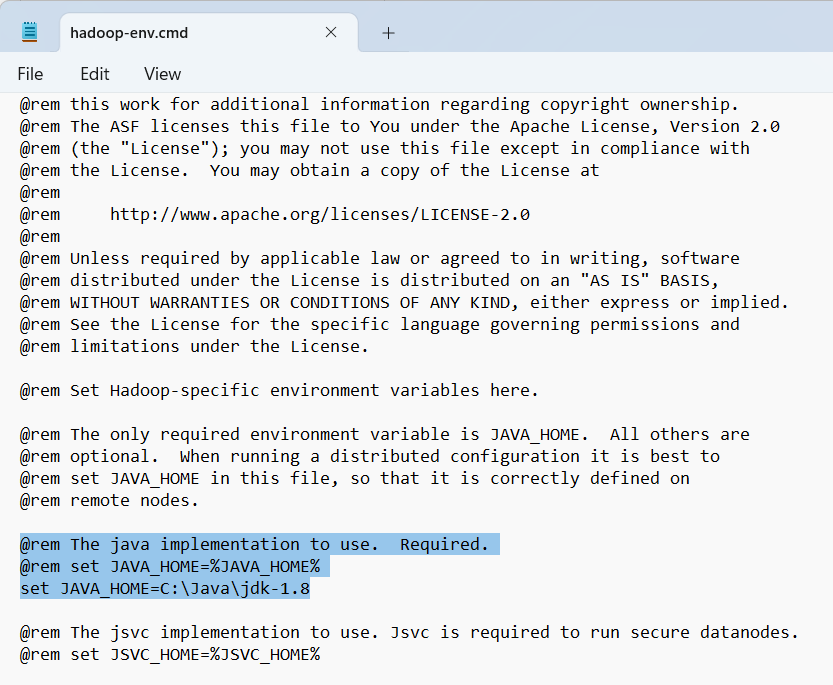
*</configuration>*

Cấu hình **hadoop-env.cmd**:

Ở dòng *set JAVA\_HOME=%JAVA\_HOME%*

Đóng dòng set *JAVA\_HOME=%JAVA\_HOME%* bằng **@rem** và thay *%JAVA\_HOME%* thành đường dẫn cài JDK trong ổ C:

*set JAVA\_HOME= C:\Java\jdk-1.8*

**

## 4. Cập nhật các Hadoop Configurations

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

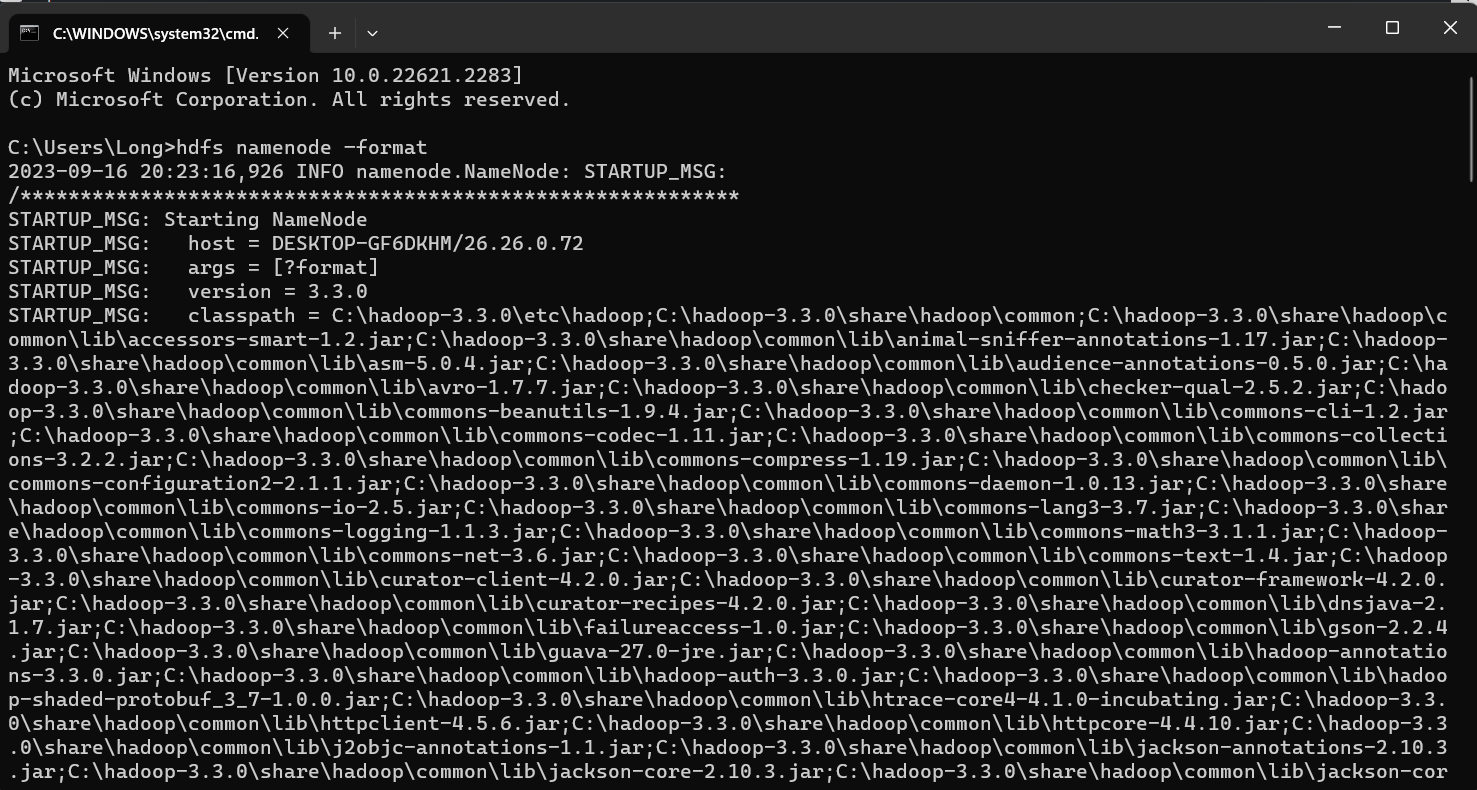
Tải về giải nén ra thấy thư mục bin ở bên trong

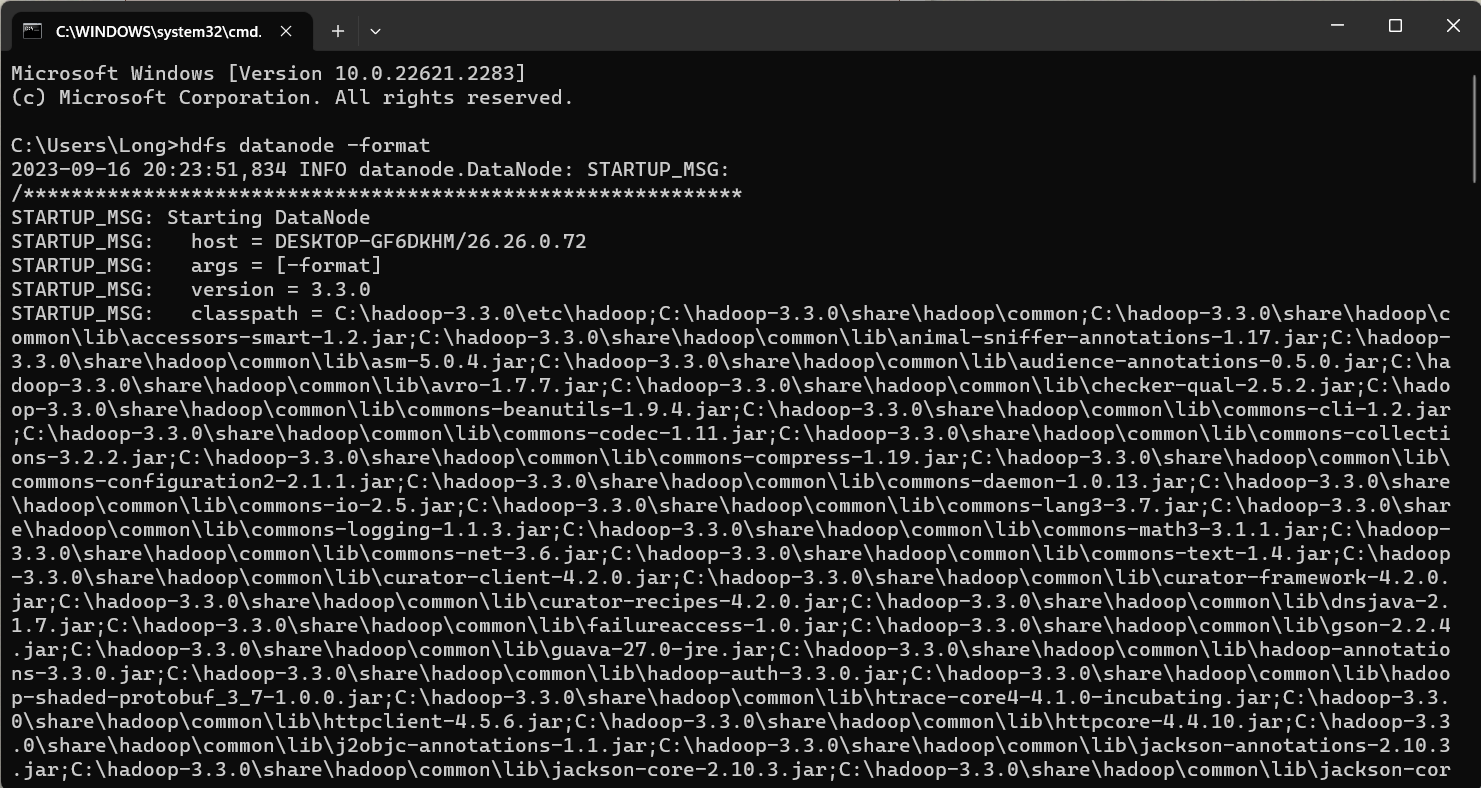
Chép đè thư mục bin này trong thư mục bin của C:\hadoop-3.3.0\bin

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

*hdfs namenode –format*

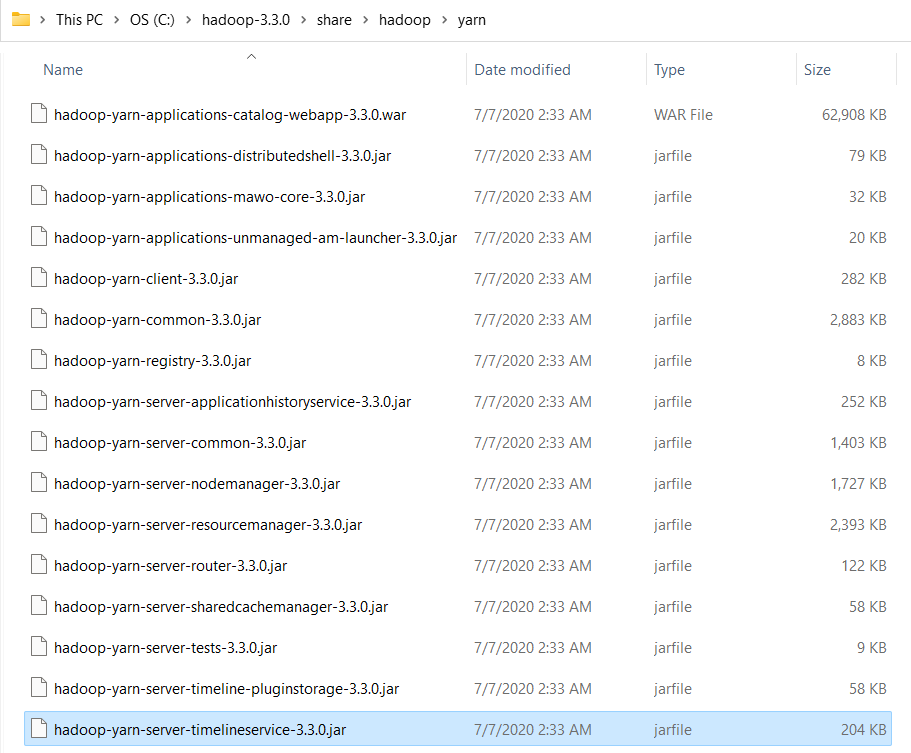
*hdfs datanode -format*





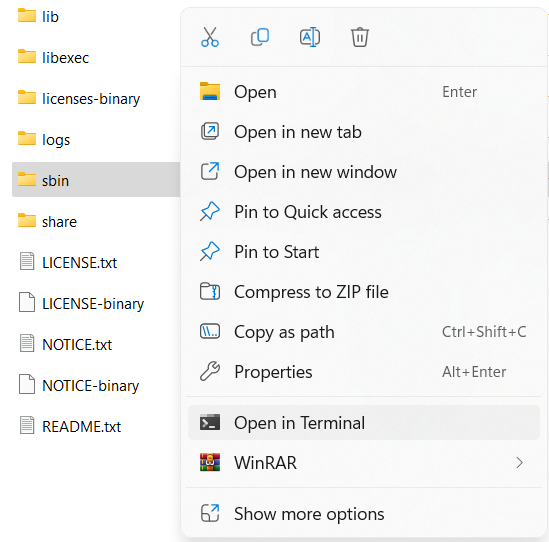
**\*Tiếp theo sao chép file:**

Sao chép file **hadoop-yarn-server-timelineservice-3.3.0.jar** từ đường dẫn *C:/hadoop-3.3.0/share/hadoop/yarn/timelineservice/* vào đường dẫn *C:/hadoop-3.3.0/share/hadoop/yarn/*



## 5. Khởi chạy Hadoop

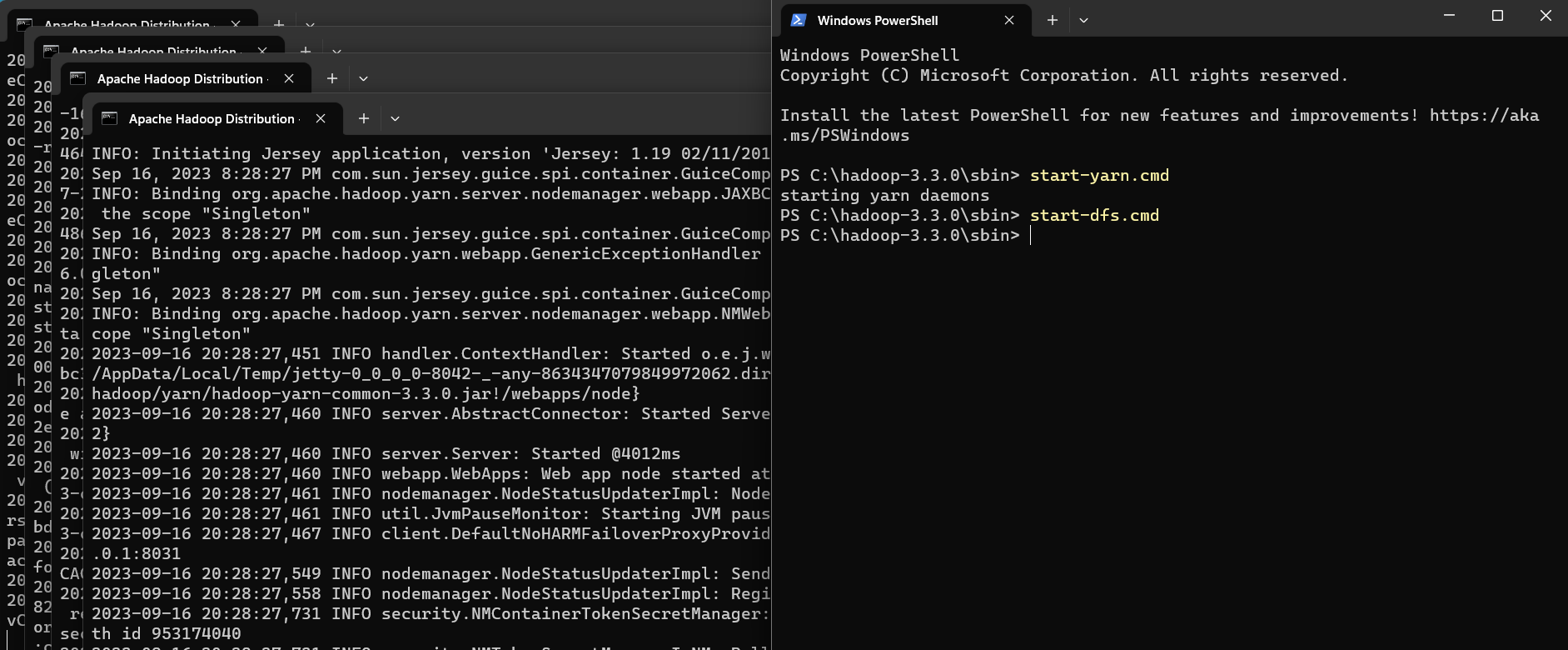
Kiểm tra Hadoop bằng **cmd** trong thư mục C:/hadoop-3.3.0/sbin.



Sau đó gõ lệnh theo trình tự để khởi động:

*start-yarn.cmd*

*start-dfs.cmd*

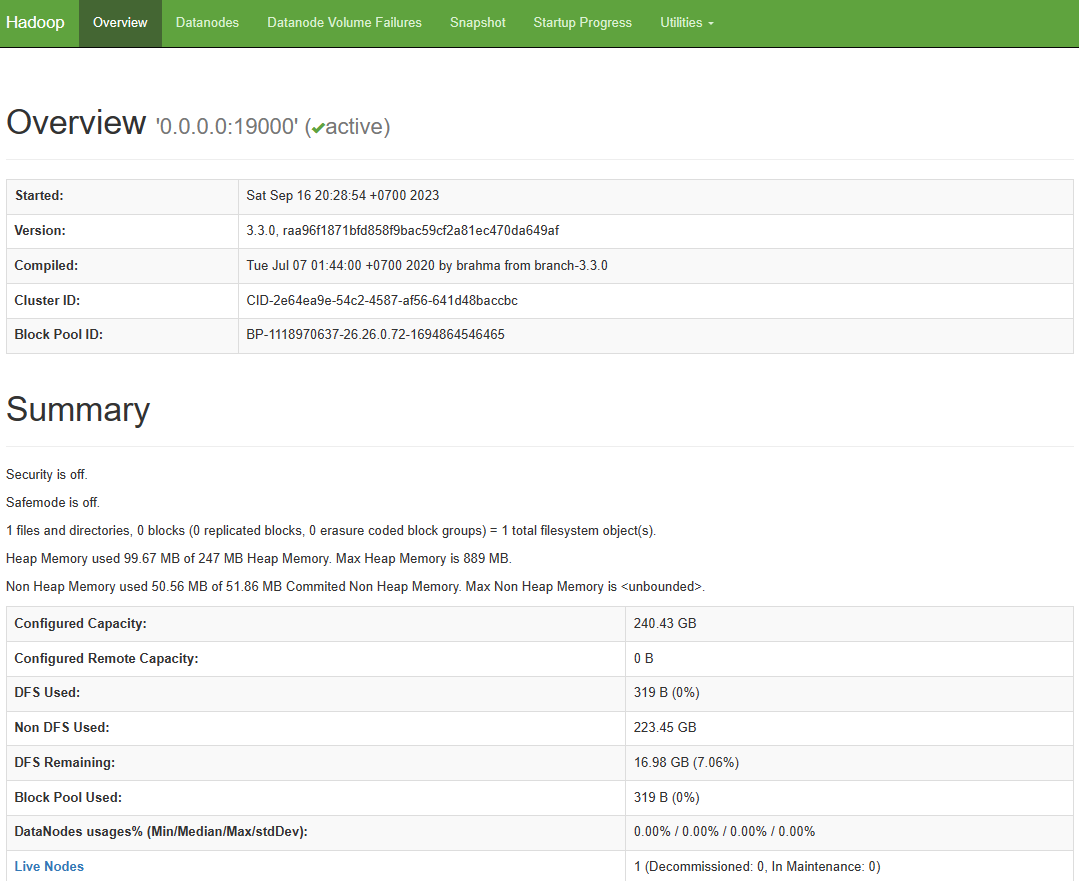
**

Nếu các ứng dụng sau đều có thể chạy bình thường:

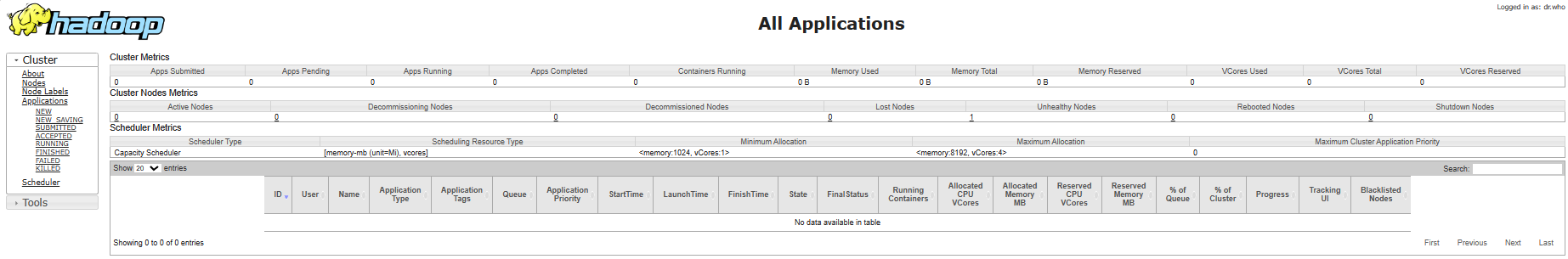
* Hadoop Namenode
* Hadoop datanode
* YARN Resource Manager
* YARN Node Manager

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

HDFS Namenode UI trong <http://localhost:8080>



Trình Quản lí Tài nguyên YARN trong <http://localhost:9870>



Dừng Hadoop lại bằng lệnh *stop-all.cmd*

# TÀI LIỆU THAM KHẢO

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

<https://gist.github.com/vorpal56/5e2b67b6be3a827b85ac82a63a5b3b2e>

<https://muhammadbilalyar.github.io/blogs/How-to-install-Hadoop-on-Window-10/>