**NHÓM 10 BIG DATA**

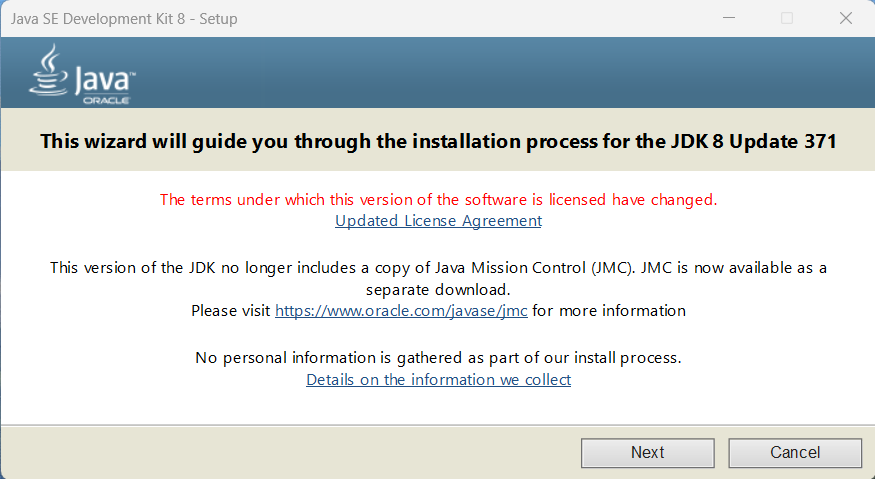
# THÀNH VIÊN:

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

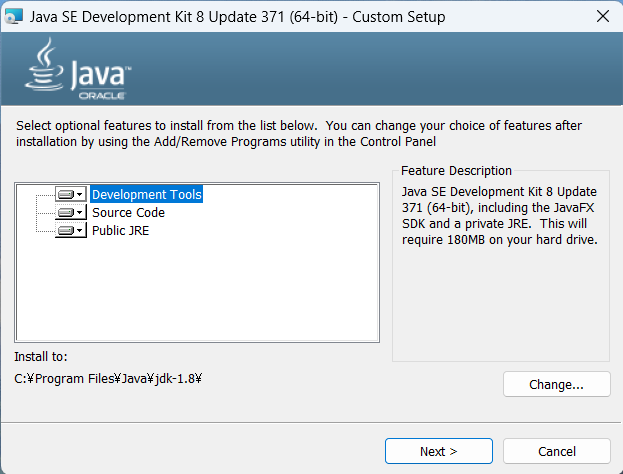
# CÀI ĐẶT SPARK TRỰC TIẾP TRÊN WINDOWS

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

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



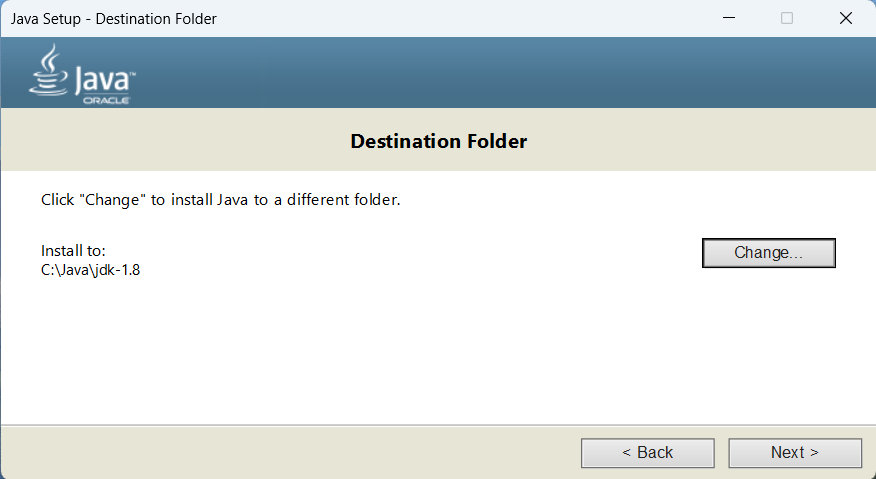
Tiếp tục chọn Next.



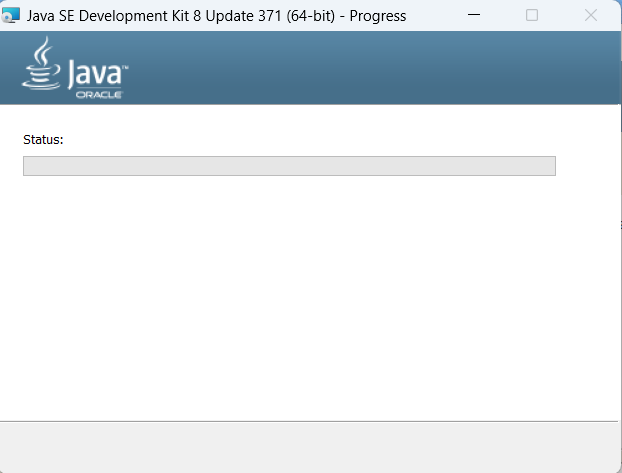
Chỉnh sửa đường dẫn đến thư mục tên ngắn và không khoảng trắng để thuận tiện cho việc tuỳ chỉnh Spark sau đó. Ở đây ta sẽ cài đặt ở đường dẫn *C:\Java\jdk-1.8*



Đợi chương trình cài đặt hoàn tất.



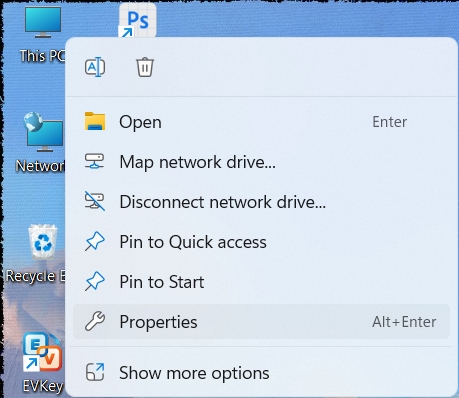
Ta vẫn sẽ điều chỉnh đường dẫn tương tự như trên nếu chương trình yêu cầu.



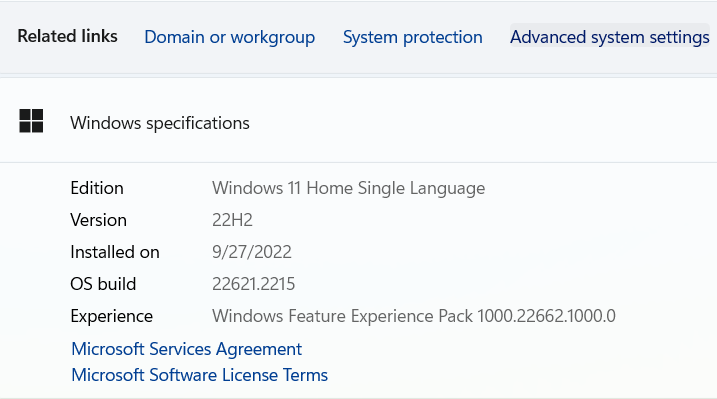
Ta tiếp tục chọn Next và chờ đợi cài đặt.

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

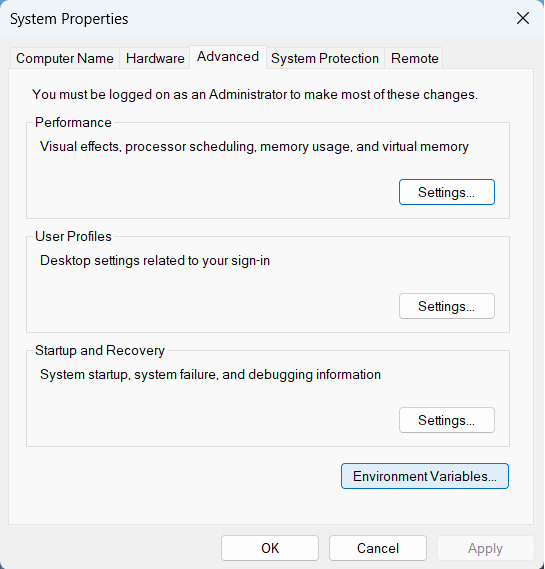
Bấm chuột phải vào This PC ⇒ Properties



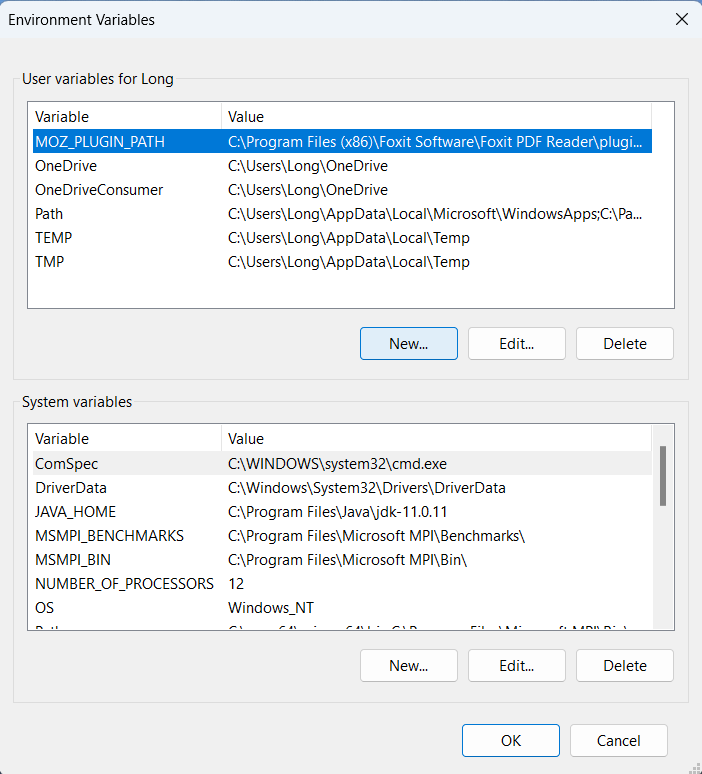
Vào Advanced system setting.



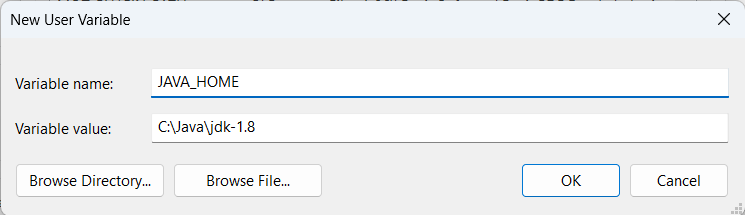
Sau khi vào System Propertise ⇒ Enviroment Variables.

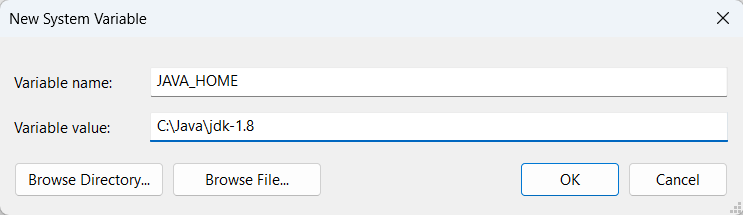


Sẽ xuất hiện màn hình Enviroment Variables:

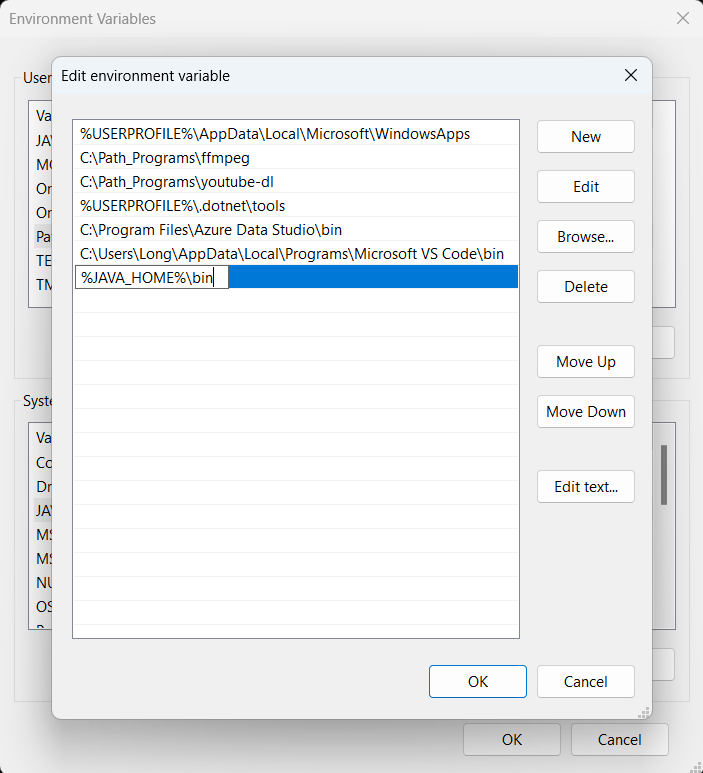


Trong mục **User** và **System Variables** ta cấu hình JAVA\_HOME bằng cách chọn nút *New…*  
Chọn đường dẫn là *C:\Java\jdk-1.8*



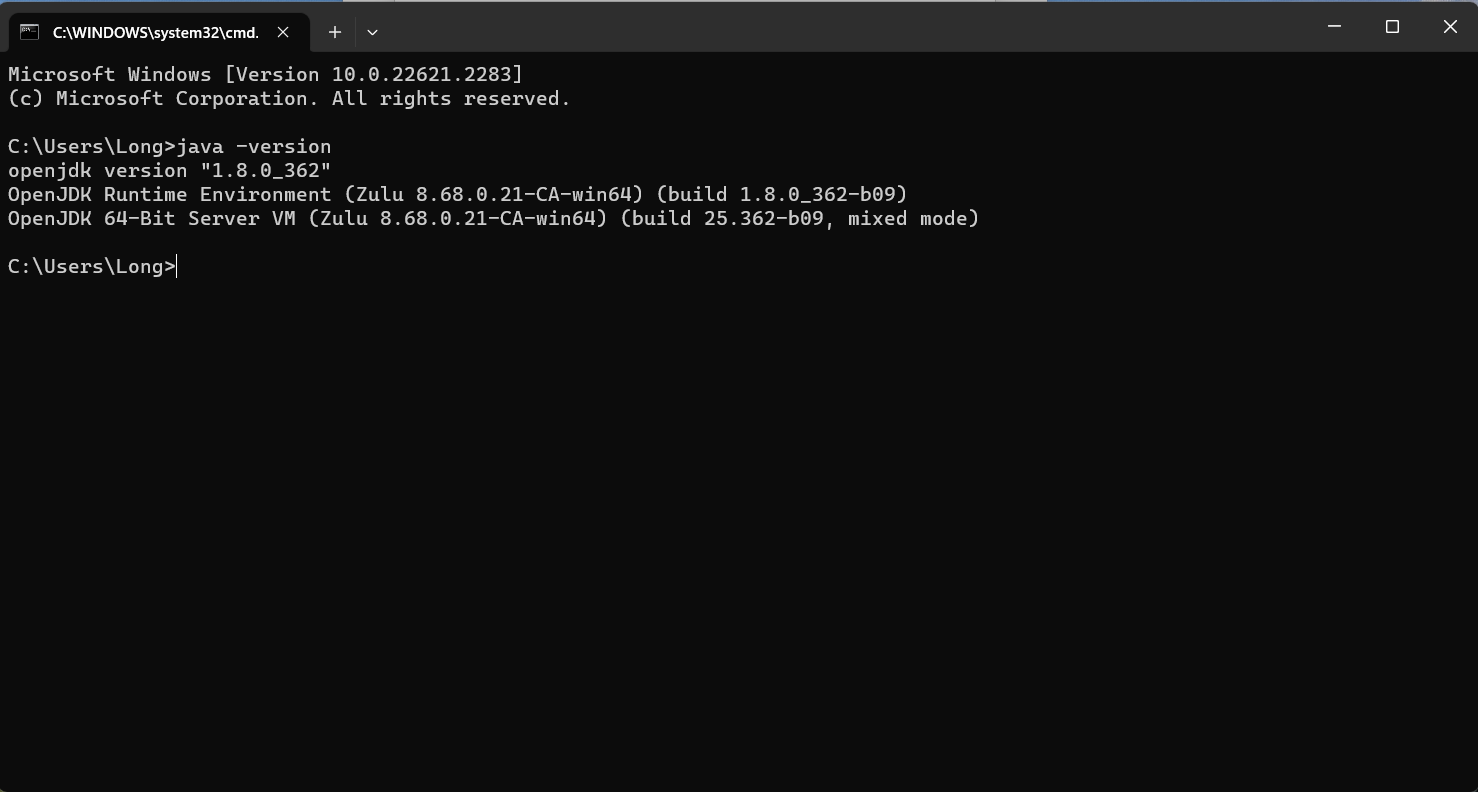


Tiếp tục chỉnh Path cho JDK cho hai mục trên.



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 Python

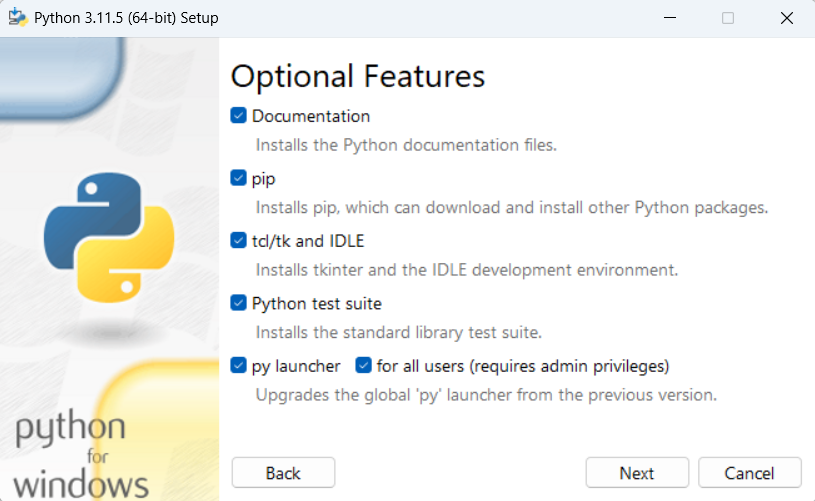
Tiếp theo tải Python phiên bản 3.11.5.



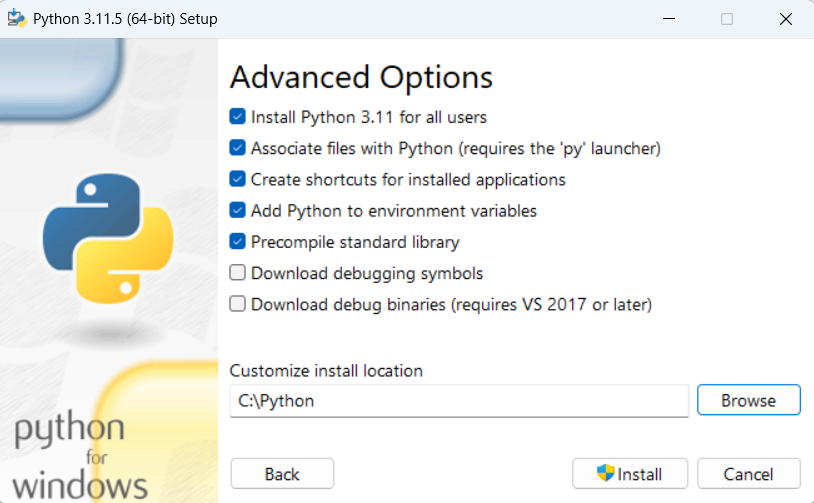
Trong phần cài đặt, tích chọn *Add python.exe to PATH* để thêm Python vào đường dẫn



Tiếp theo là chọn *Customize installation* để tuỳ chỉnh cài đặt.



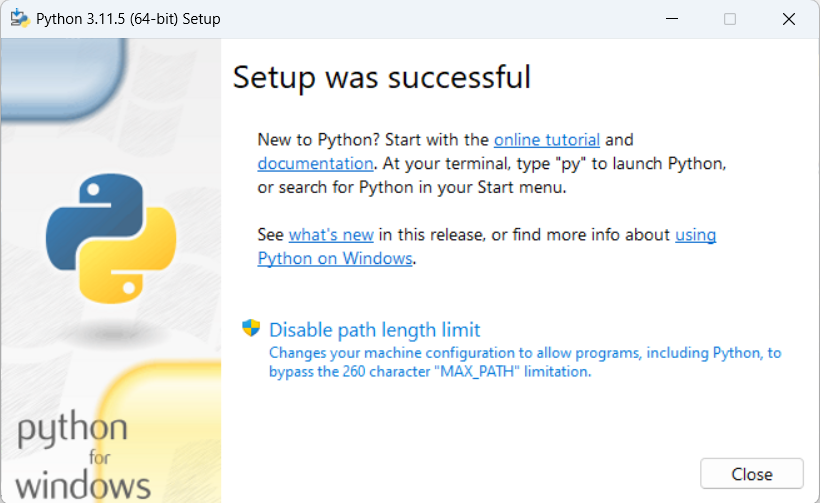
Bấm Next để tuỳ chỉnh tiếp.



Ta tích chọn *Install for all users*.

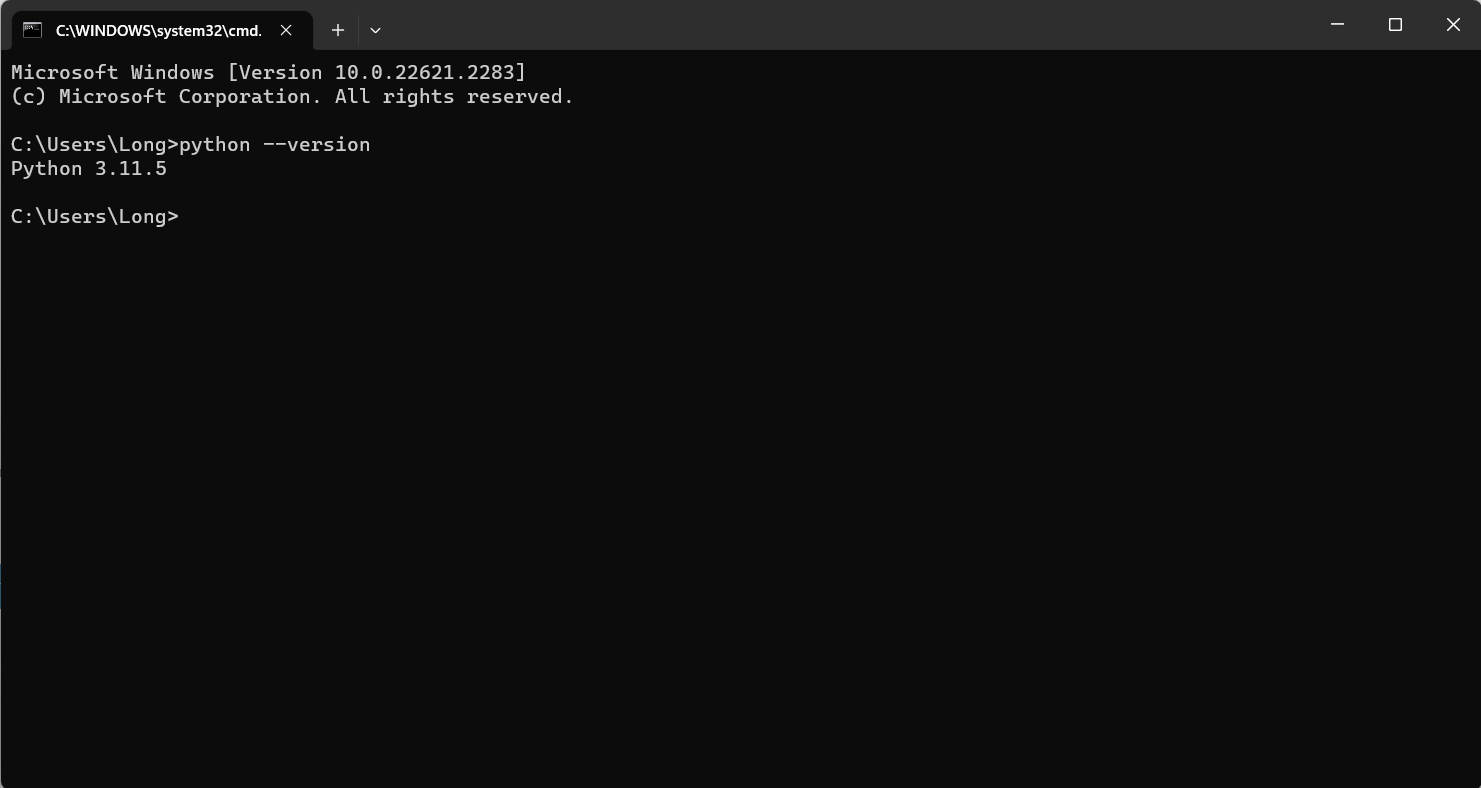
Kế đến là thay đổi đường dẫn. Tạo một thư mục mới, đặt tên là Python vào ổ C và chỉnh đường dẫn cài đặt đến đó.

Sau đó nhấn chọn Cài đặt.



Khi quá trình cài đặt hoàn tất, nhấp tùy chọn *Disable path length limit* ở dưới rồi đóng lại.

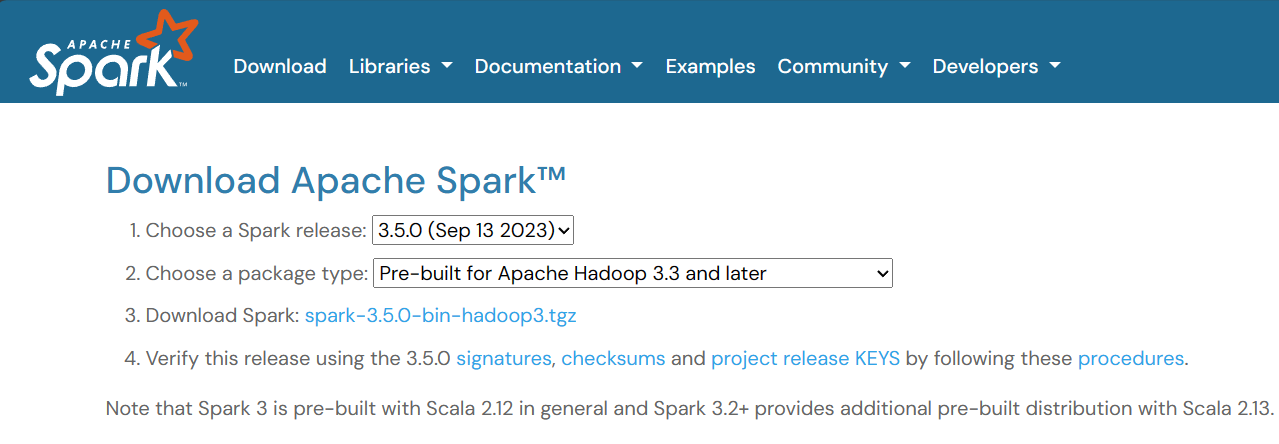
Cuối cùng kiểm tra phiên bản Python bằng Command Prompt.



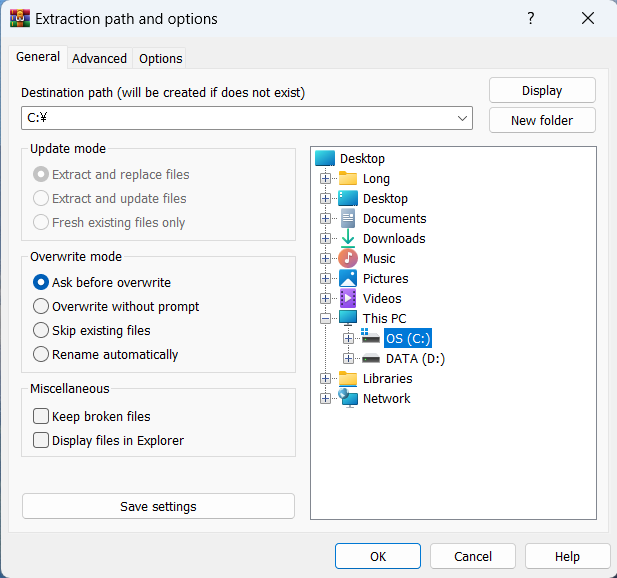
*⇒ Như vậy, Python đã được cài đặt.*

## 3. Tải và cài đặt Spark.

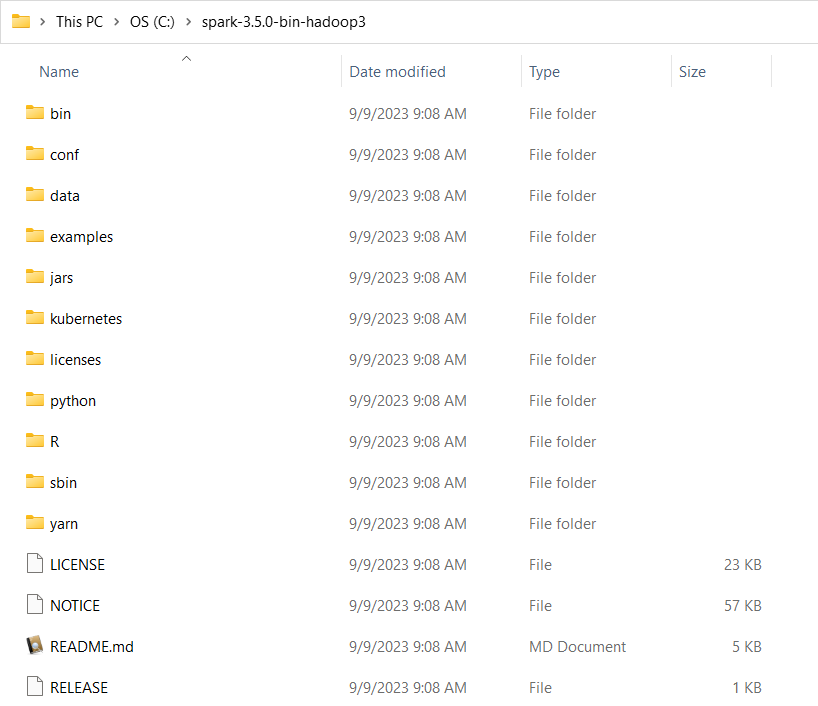
Tải Spark về từ trang Apache Spark với tên file là *spark-3.5.0-bin-hadoop3.tgz*



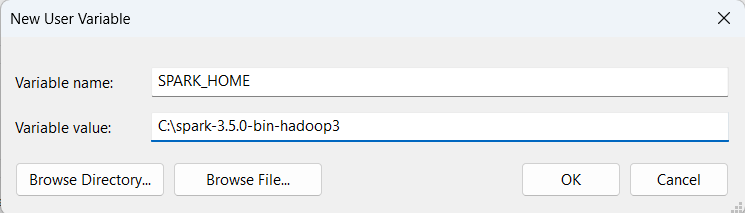
Sau đó giải nén bằng Winrar, và để cho thuận tiện, ta sẽ giải nén vào ổ C.



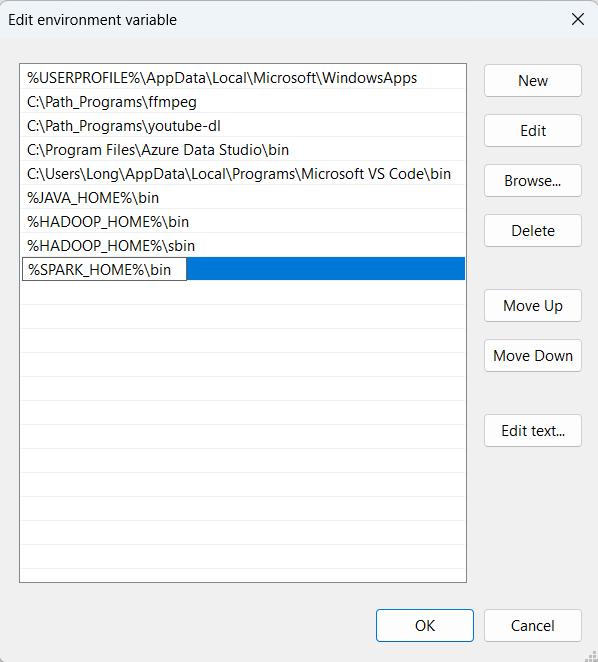
Kết quả:



Tiếp đến là phần cài đặt môi trường cho Spark. Chọn *New* ở phần **User Variables**.



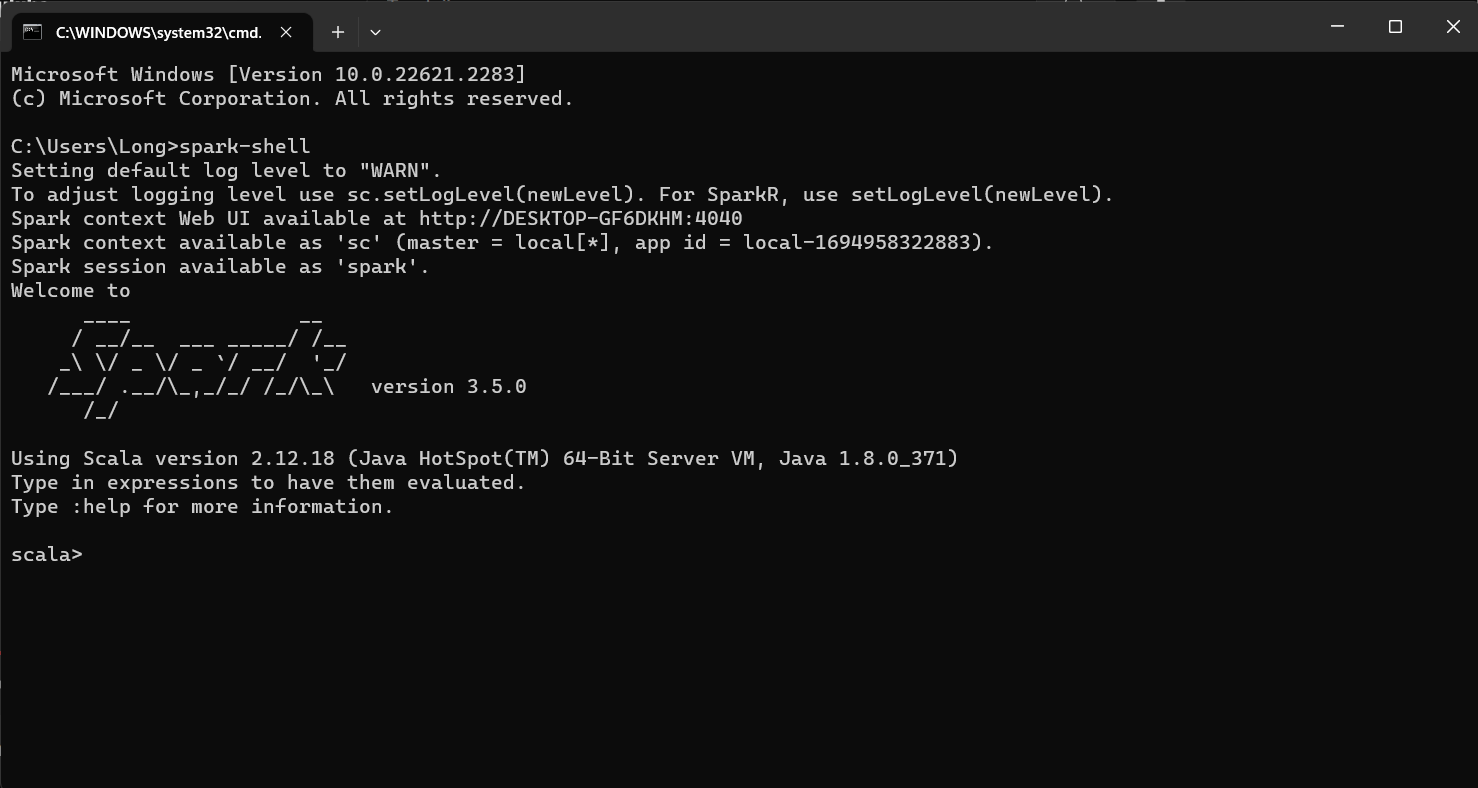
Và bước cuối để cài đặt môi trường Spark hoàn tất bằng cách thêm đường dẫn. Chọn *Edit*, chọn *New* để thêm mới.

**

Chọn OK để xác nhận thay đổi.

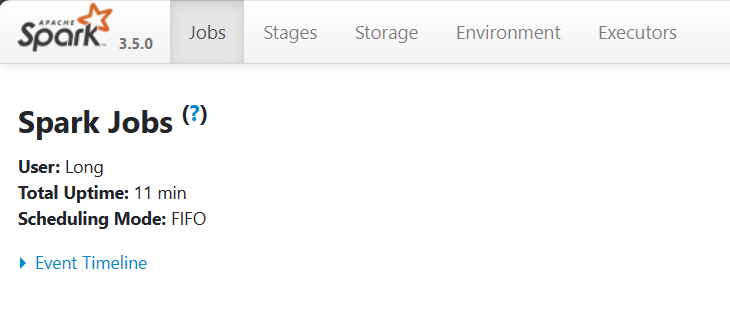
## 3. Khởi chạy Spark

Kiểm tra Spark bằng **Command Prompt** bằng lệnh *spark-shell*.



*⇒Như vậy, Spark đã chạy thành công.*

Web UI của Apache Spark trong <http://localhost:4040/>

****

# TÀI LIỆU THAM KHẢO

<https://phoenixnap.com/kb/install-spark-on-windows-10>