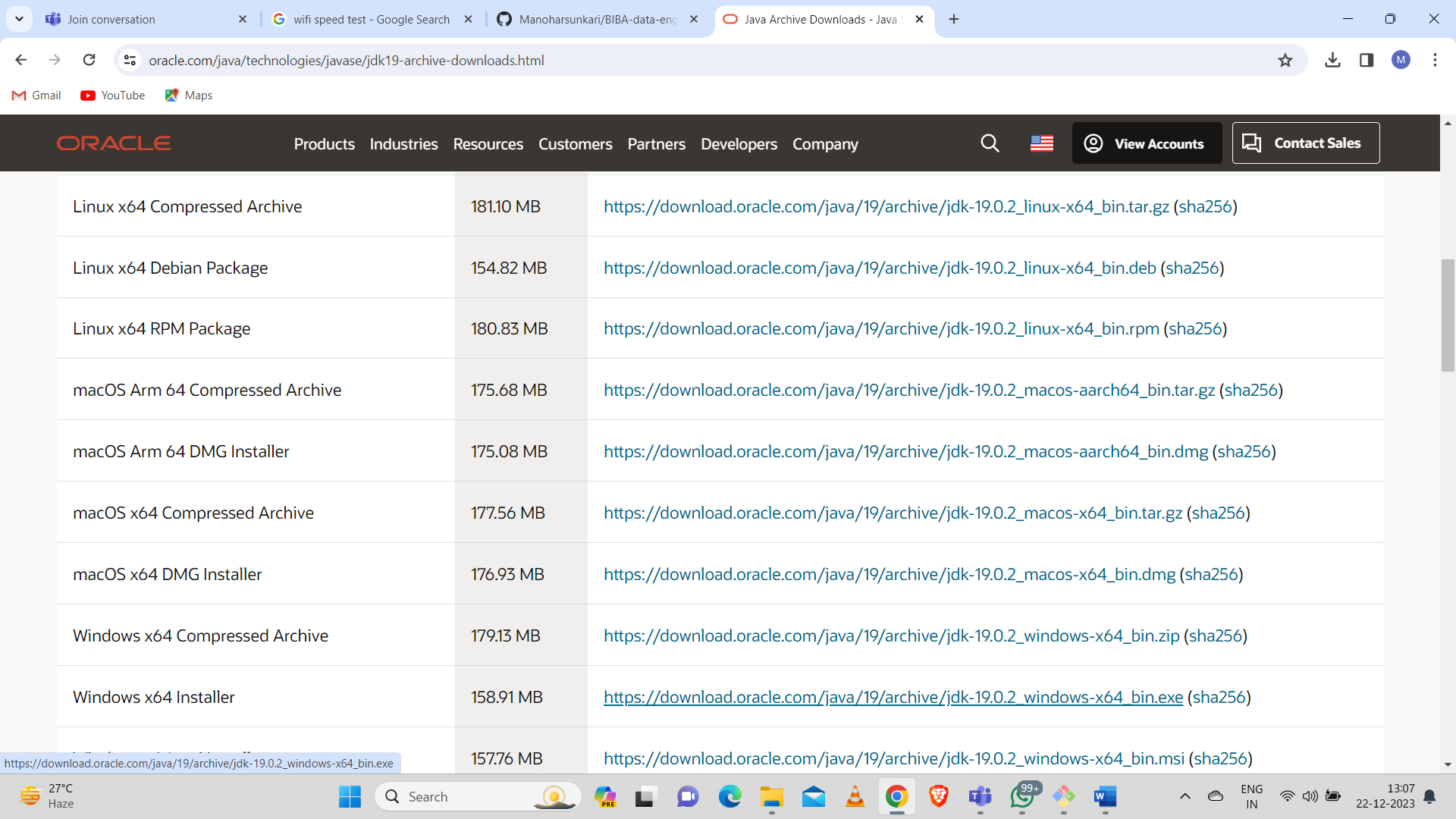
**Sunkari Manohar**

**Assessment – 21/12/2023**

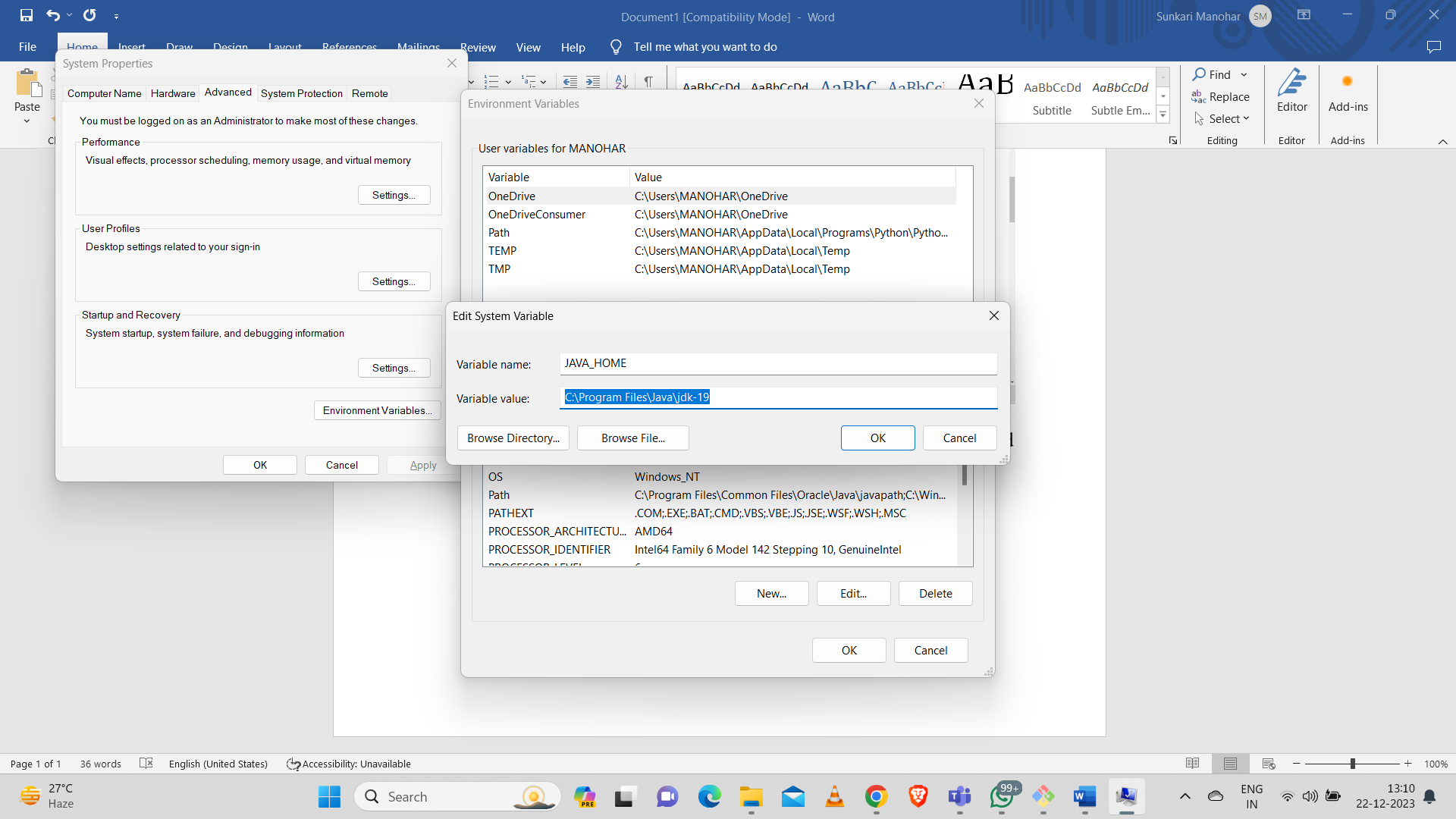
Installation of pyspark:

Pyspark = Python + Apache Spark

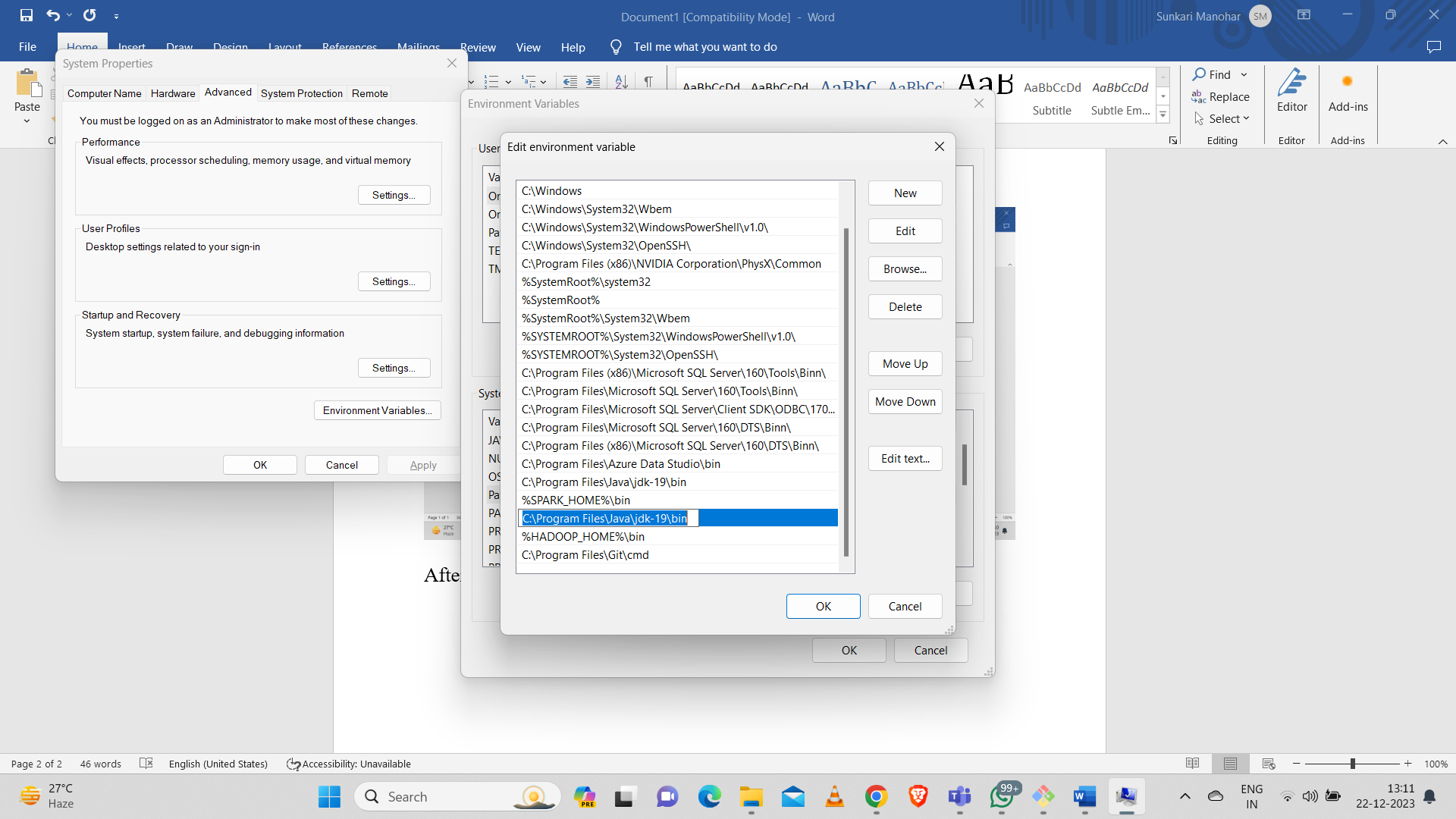
First installed JDK 19.0.2



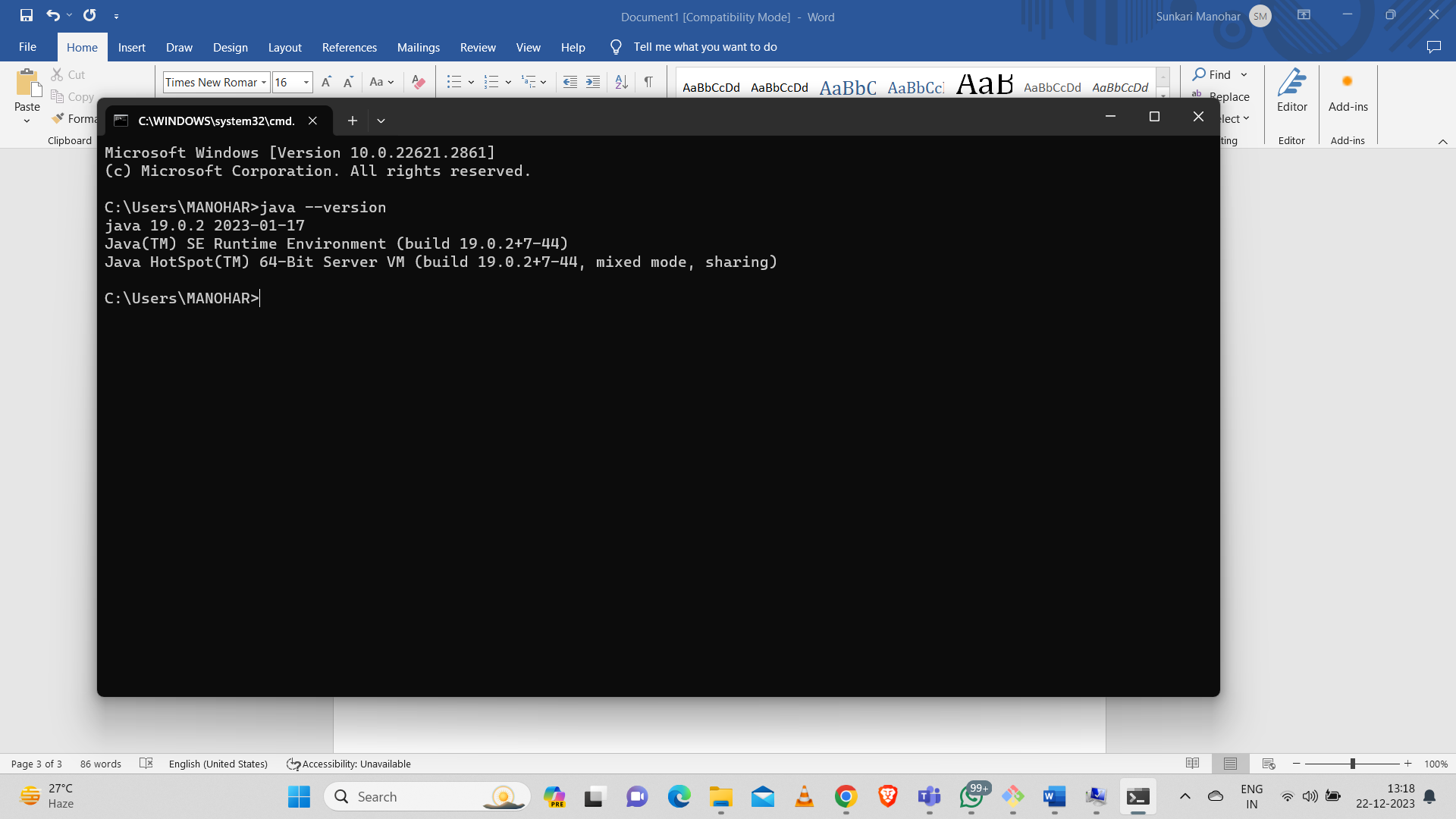
And then edit environment variables for java like JAVA-HOME and add the path of the java in files.



After variable creation in system variables changed path to \bin

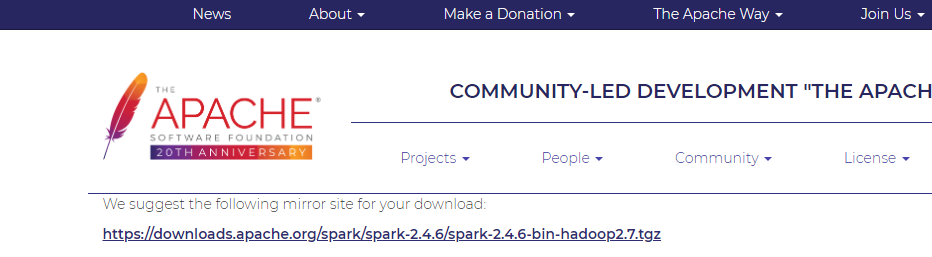


Go to "Command Prompt" and type "java -version" to know the version and know whether it is installed or not.



After this we need to set up python and also edit those environment variables.

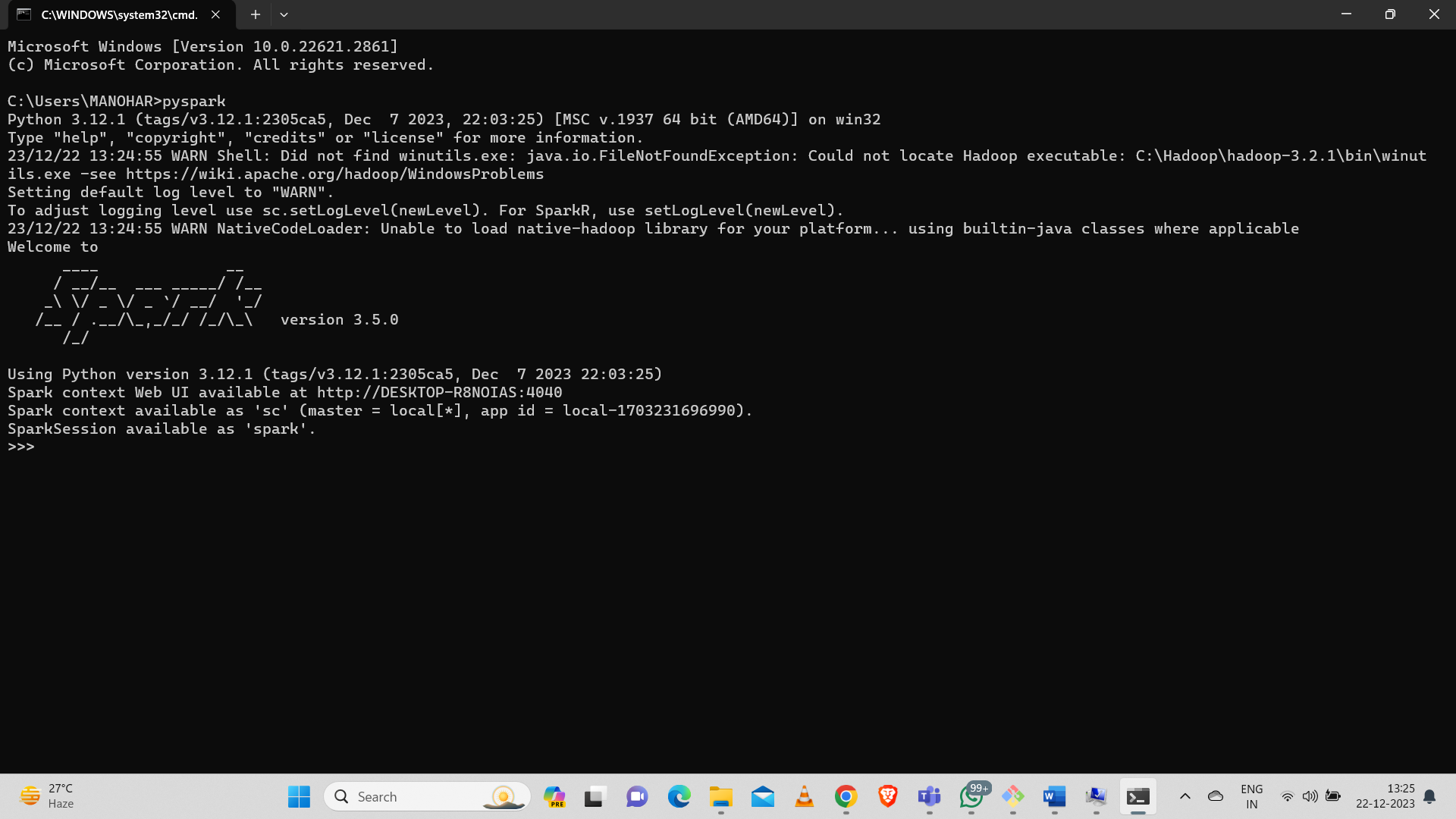
And then installed Spark and Hadoop



After installation of spark and Hadoop created separate folders and extracted the files in the created folders and then edited environmental variables for both spark and Hadoop.

And also installed pyspark using “pip install pyspark” in command prompt.

After this installations and editing environmental variables we need to open command prompt and type “pyspark” then it will appear like this:



The above shows pyspark installed successfully..

To connect to pyspark ui: http://localhost:4040/jobs

We will get the web ui of pyspark

