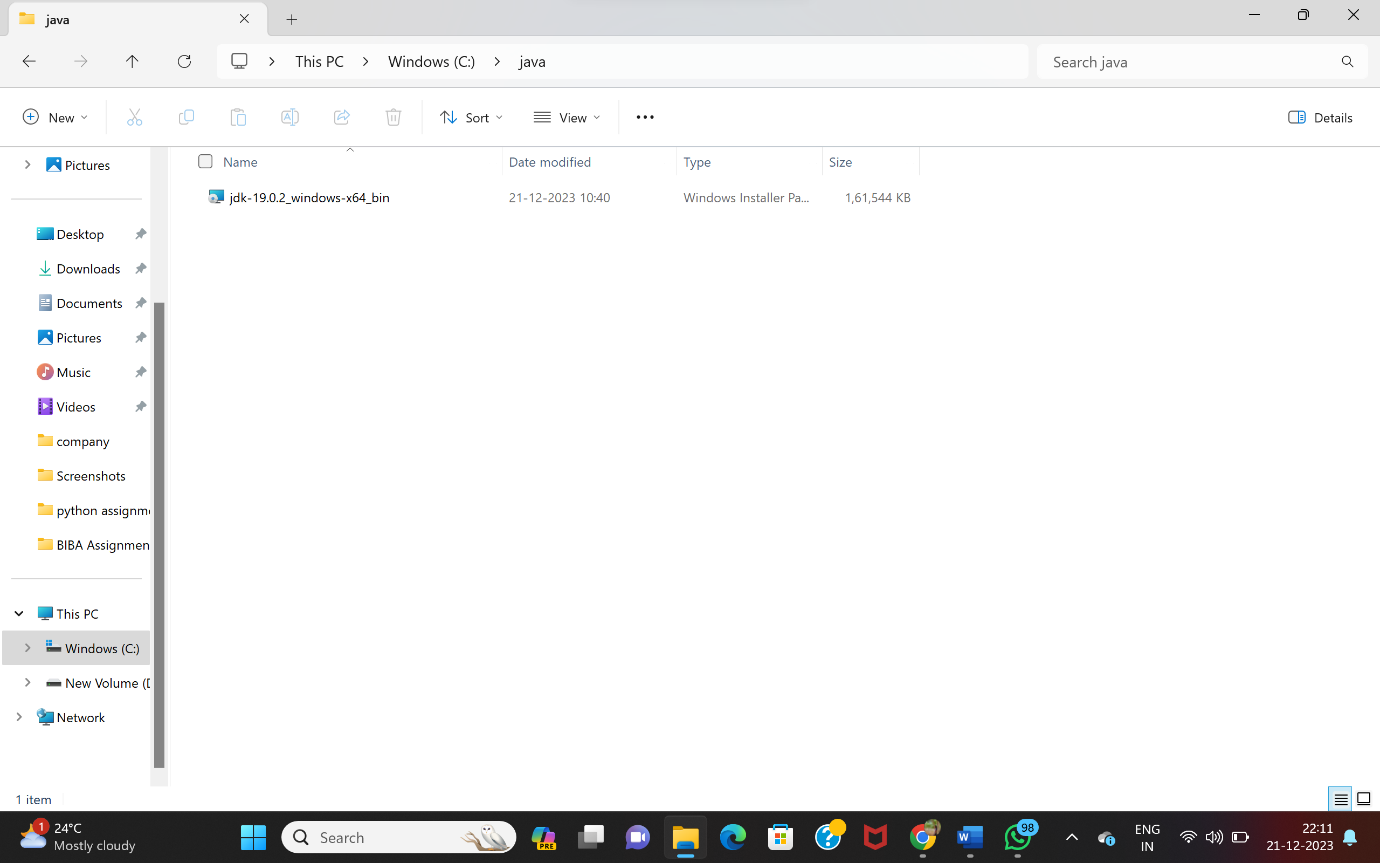
ASSIGNMENT (21-12-2023)

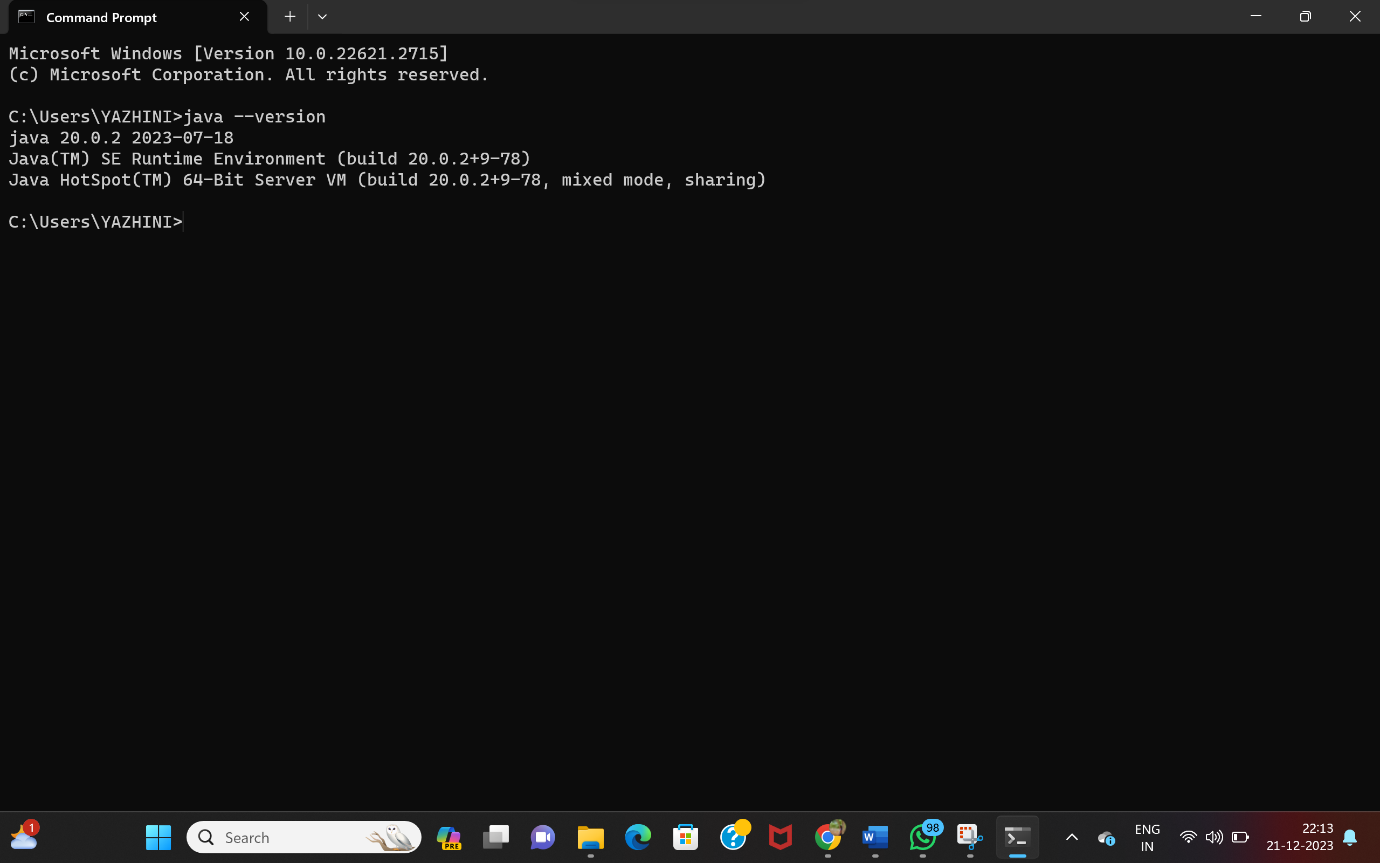
INSTALLATION OF APACHE SPARK

1)Install Java Jdk

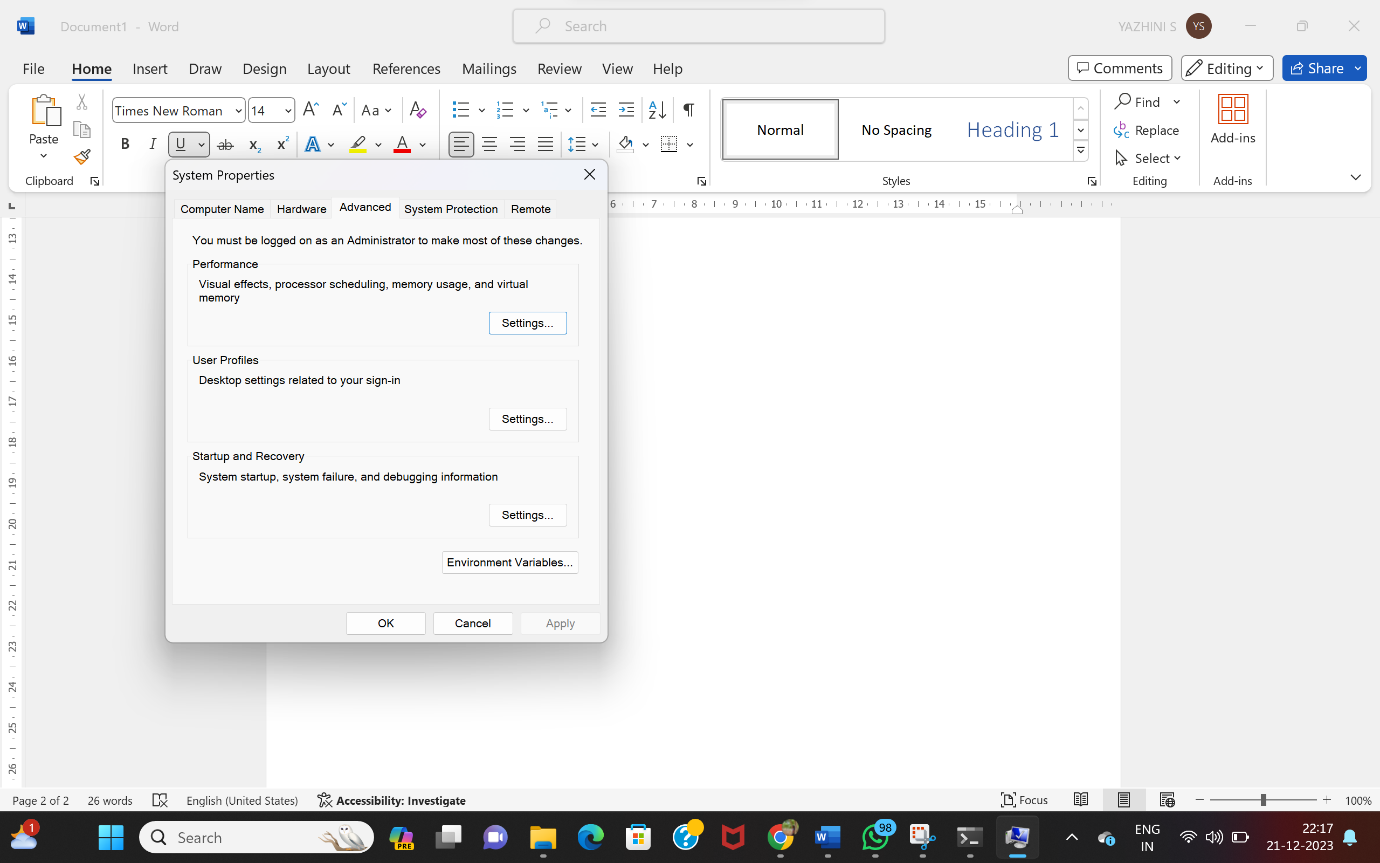


To check its installed

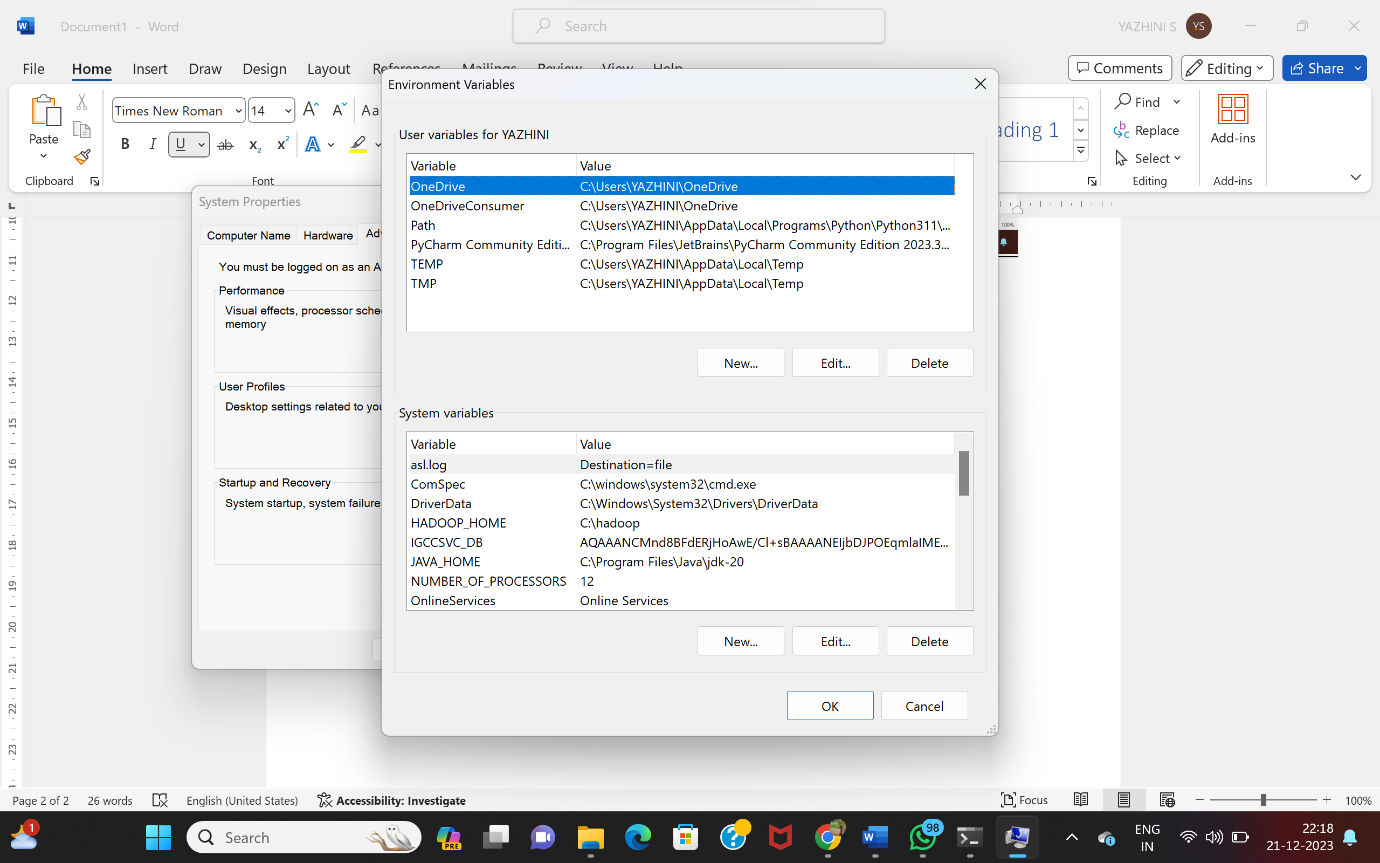
🡪checking the version of java in cmd.



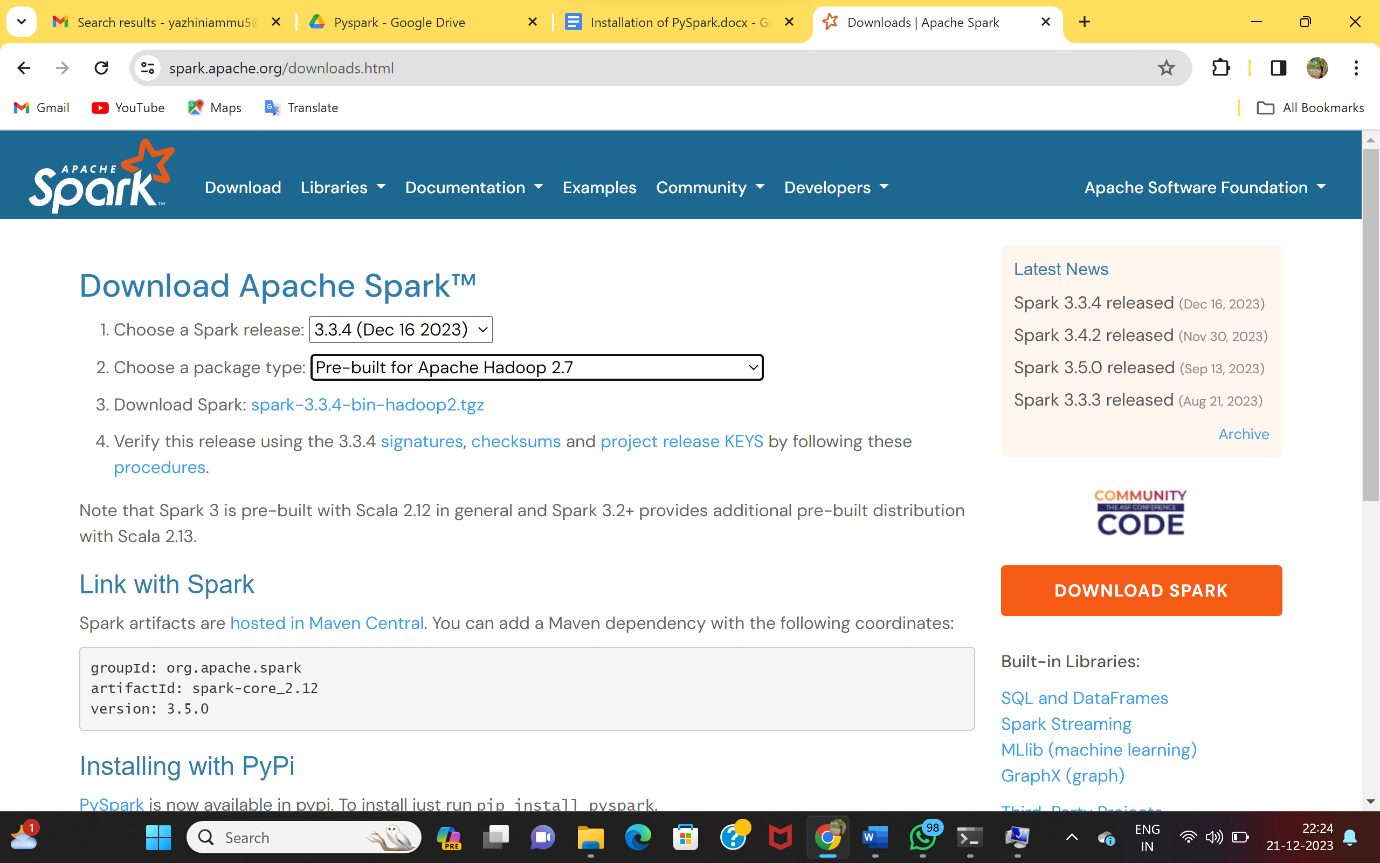
🡪Adding java in System environmental variables



🡪Setting the path of java in the system environmental variables



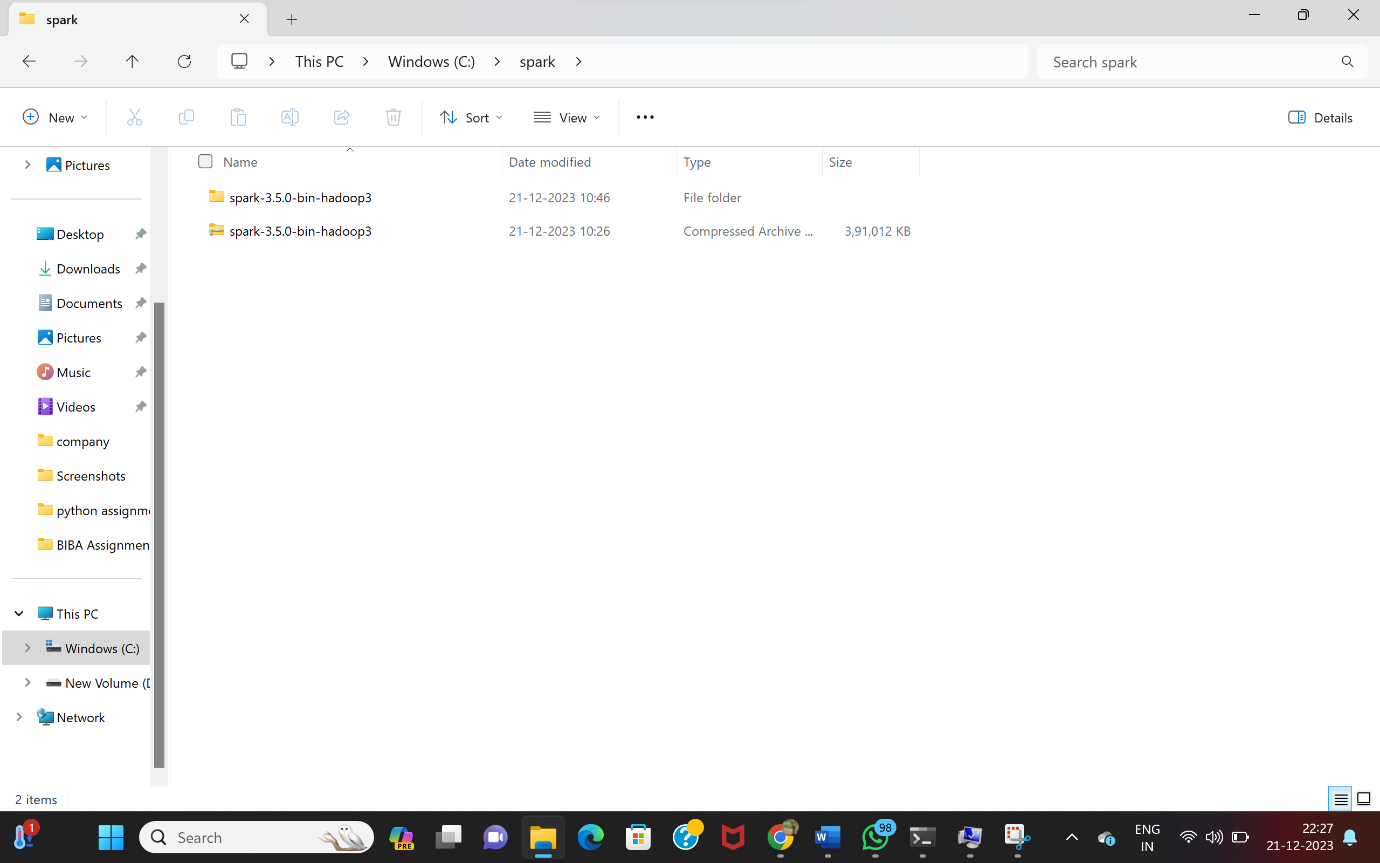
Download Apache spark from Apache software Foundation



🡪Open a file named “spark” in C folder.

🡪copy the downloaded apache zip file and place in the spark folder.

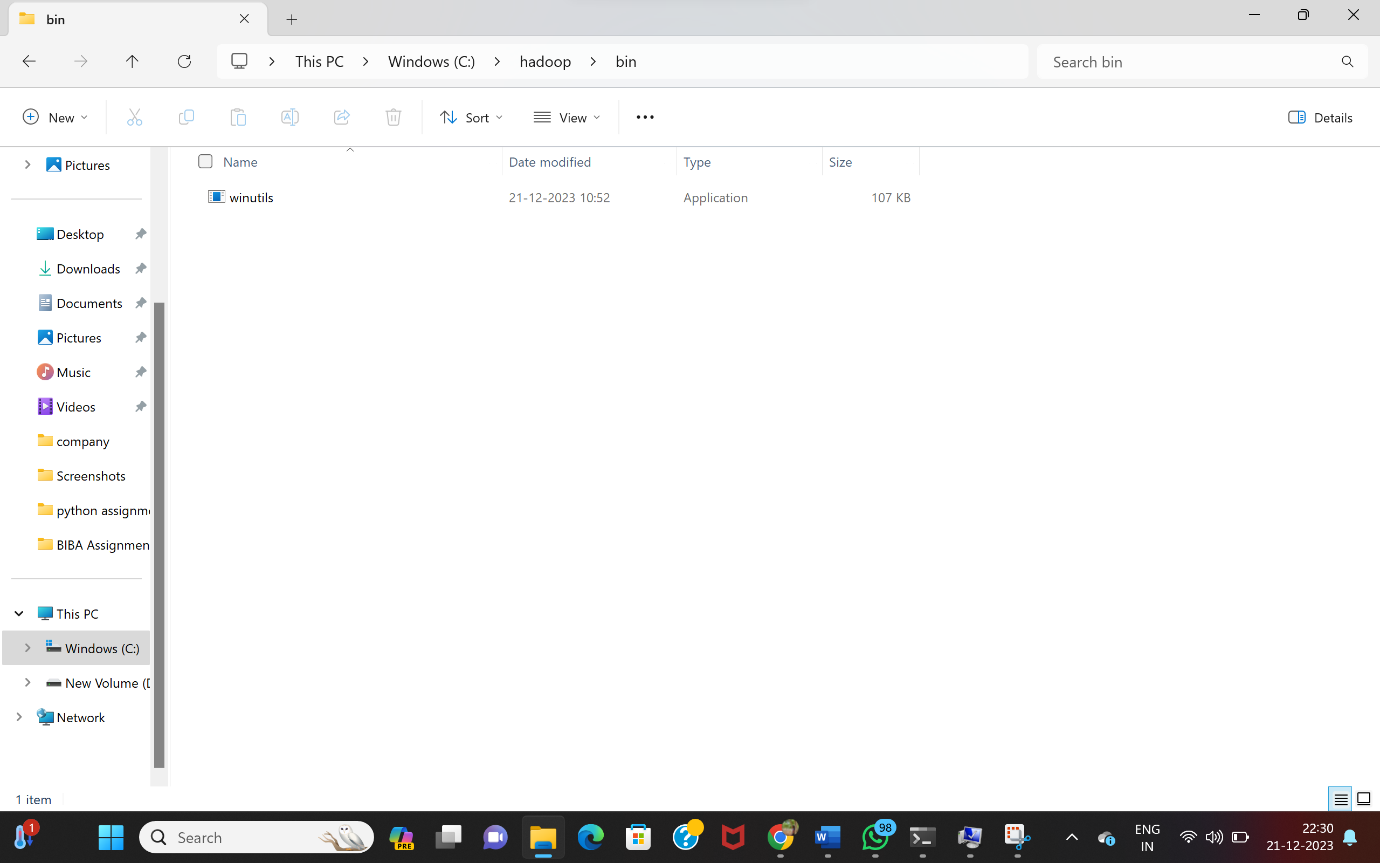
🡪Extract all the files from the zip file and place it in the same folder.



🡪Similarly, open another file name Hadoop in the c folder.

🡪Inside Hadoop folder , create another folder named bin.

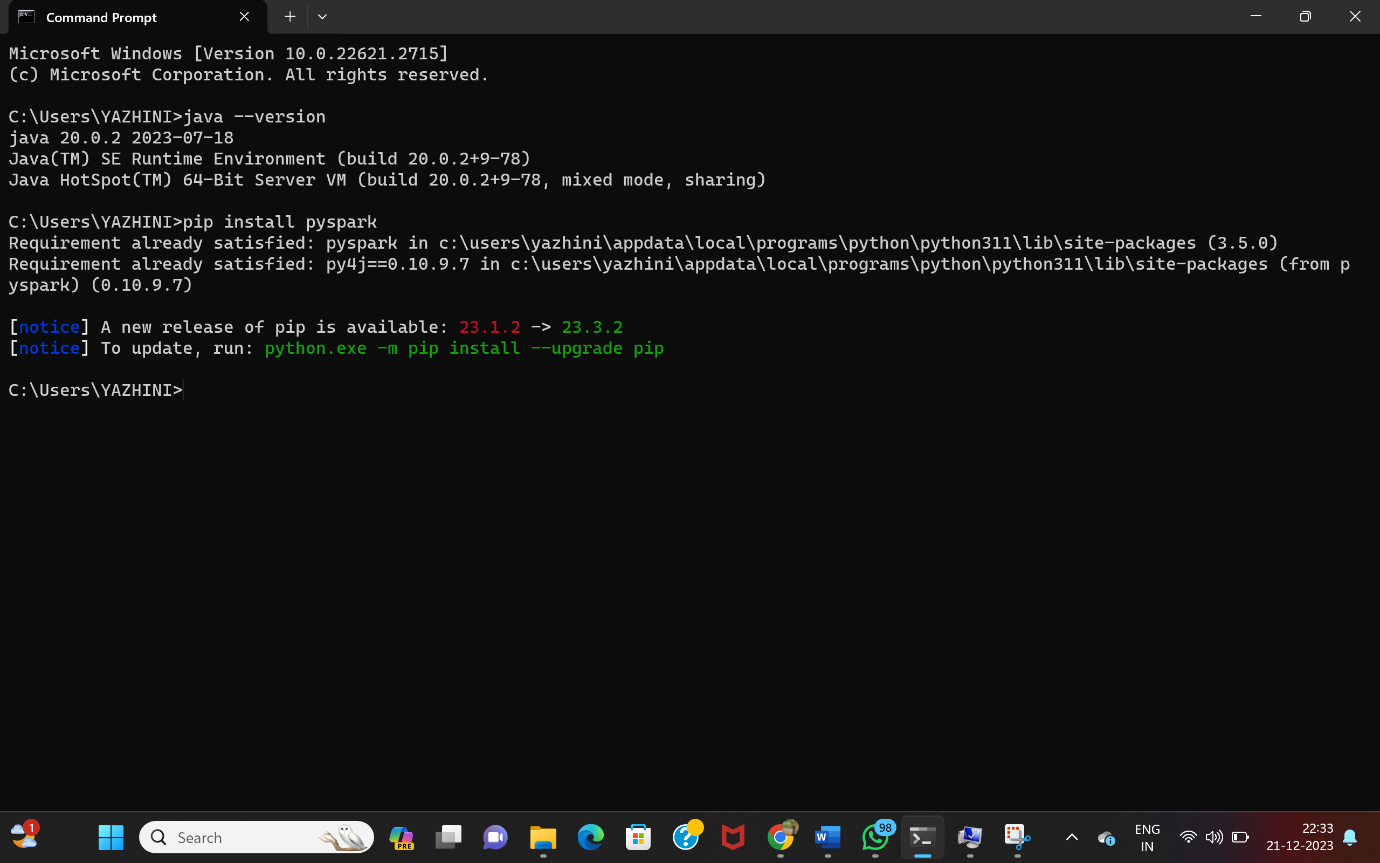
🡪Download winutils.exe from github and place it in the bin folder inside the Hadoop folder.



**INSTALLING PYSPARK**

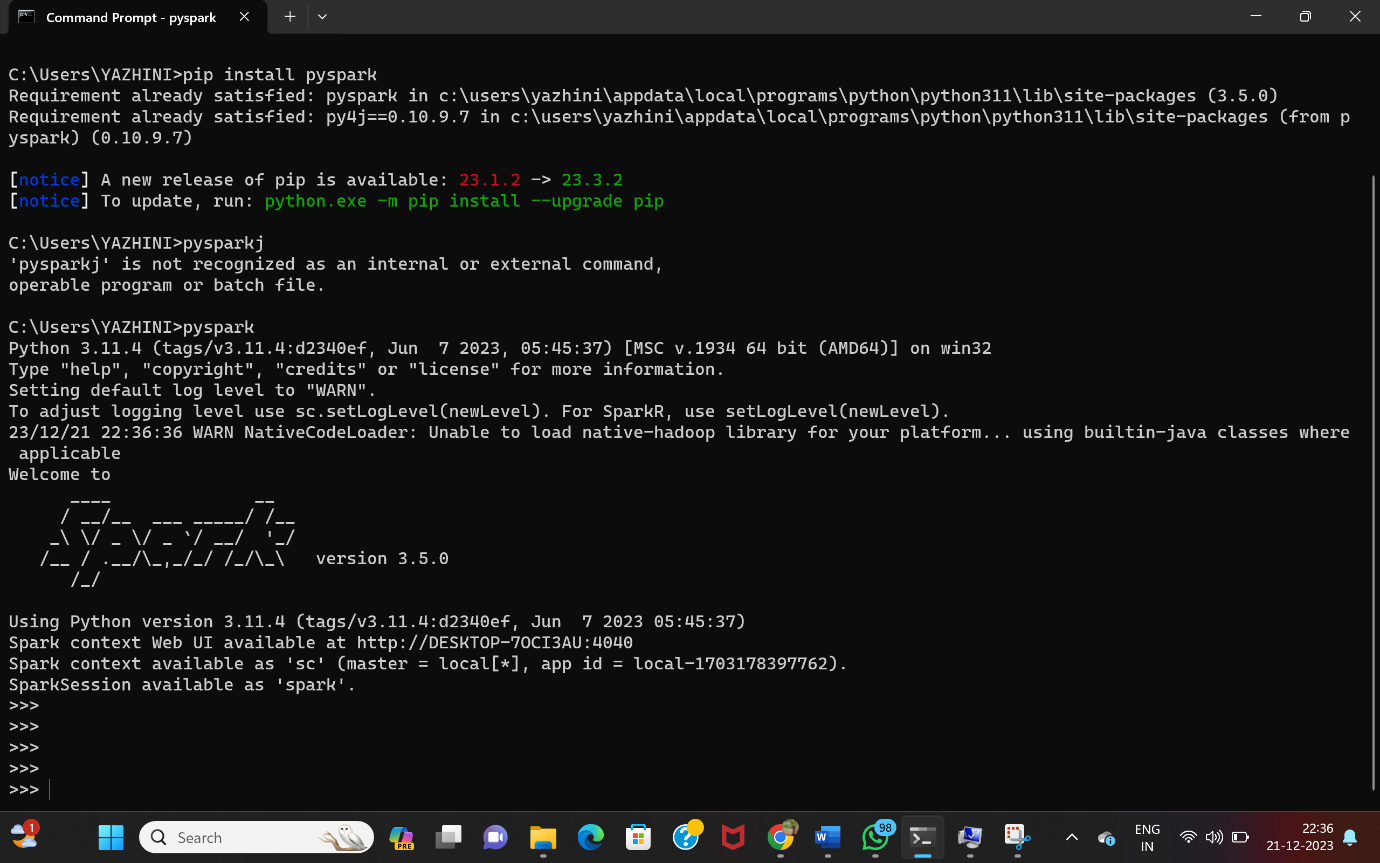
**🡪**Open the command prompt and type “pip install pyspark”

🡪This command will install the pyspark in the machine.



🡪Now, put the path of the spark folder in the system environmental variables same as java.

🡪After that, open command prompt and rum the command “pyspark” or “apache-shell” to see whether the apache spark is installed.



🡪Finally, run “https:\localhost\4040\jobs” in the chrome to see whether the apache and pyspark is working .