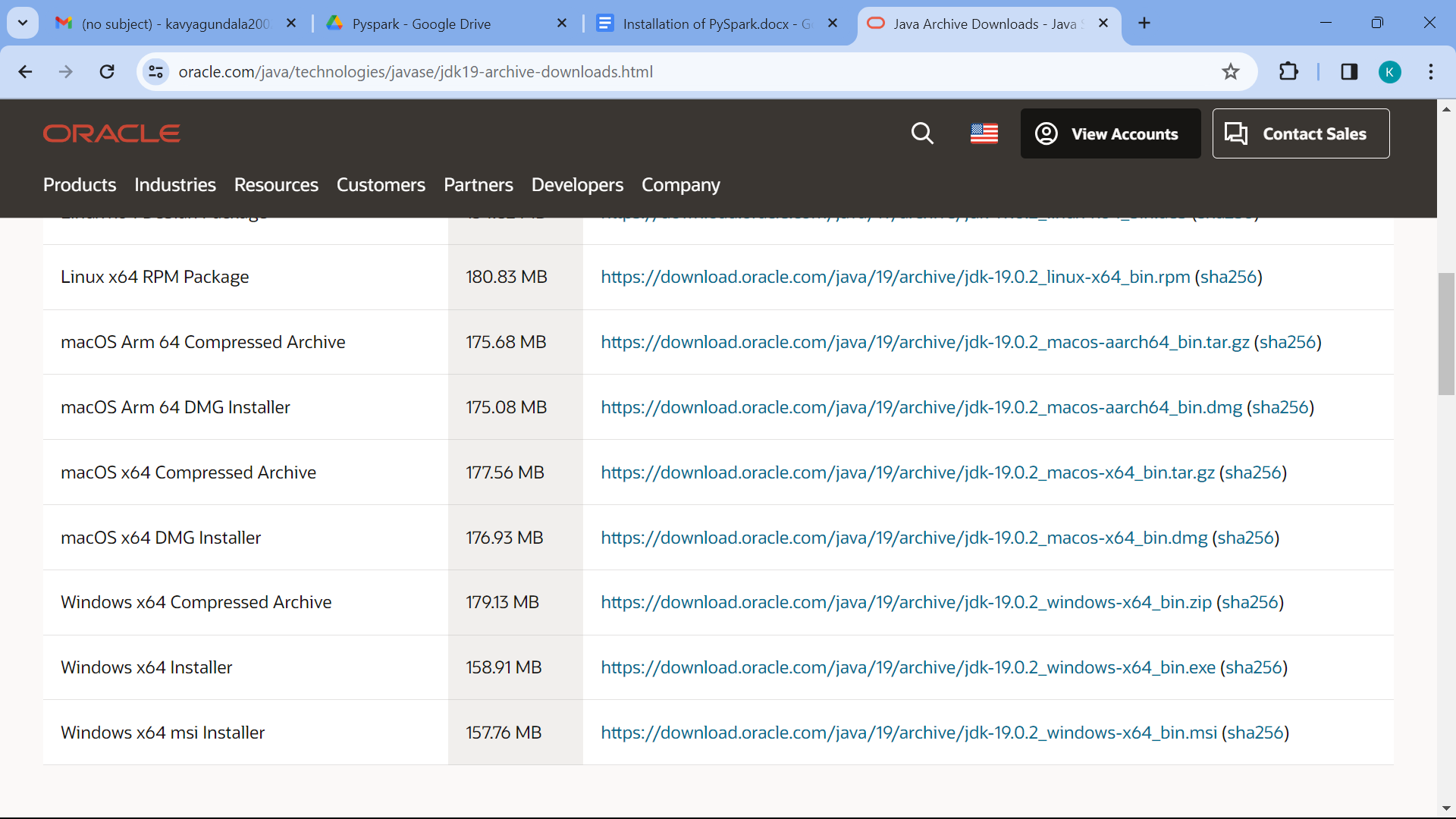
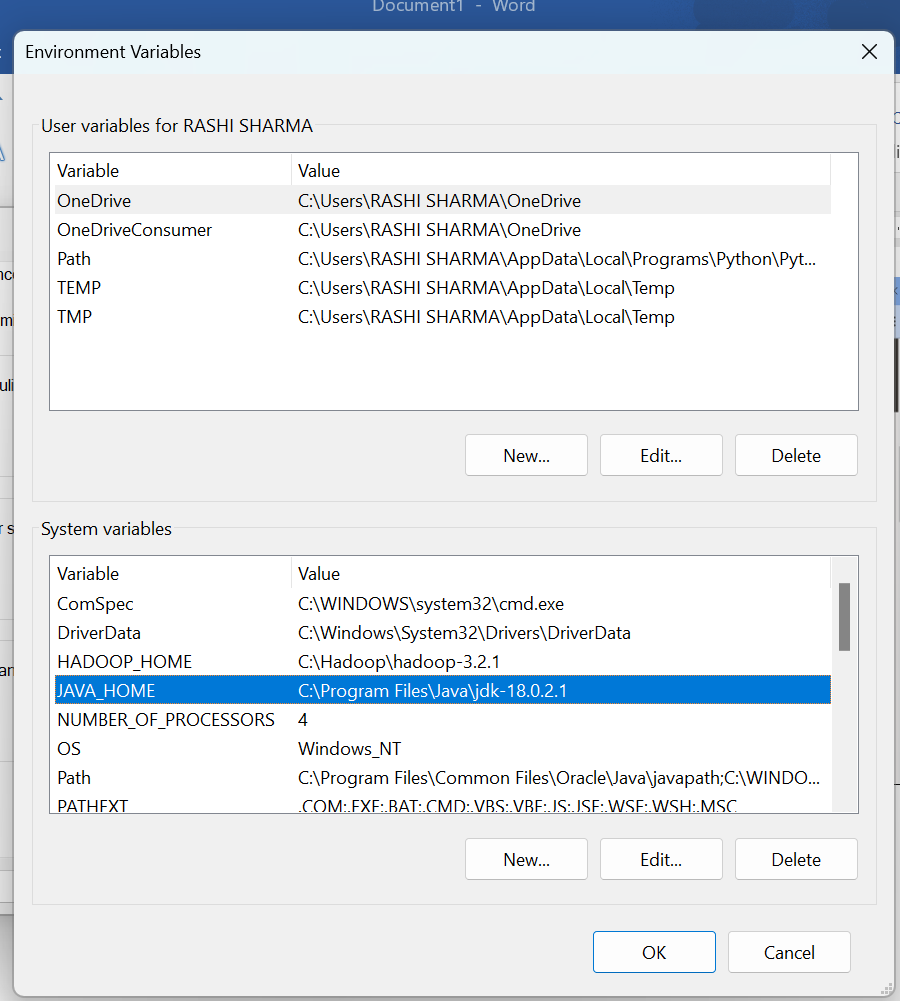
Kavya Gundala

21-12-2023

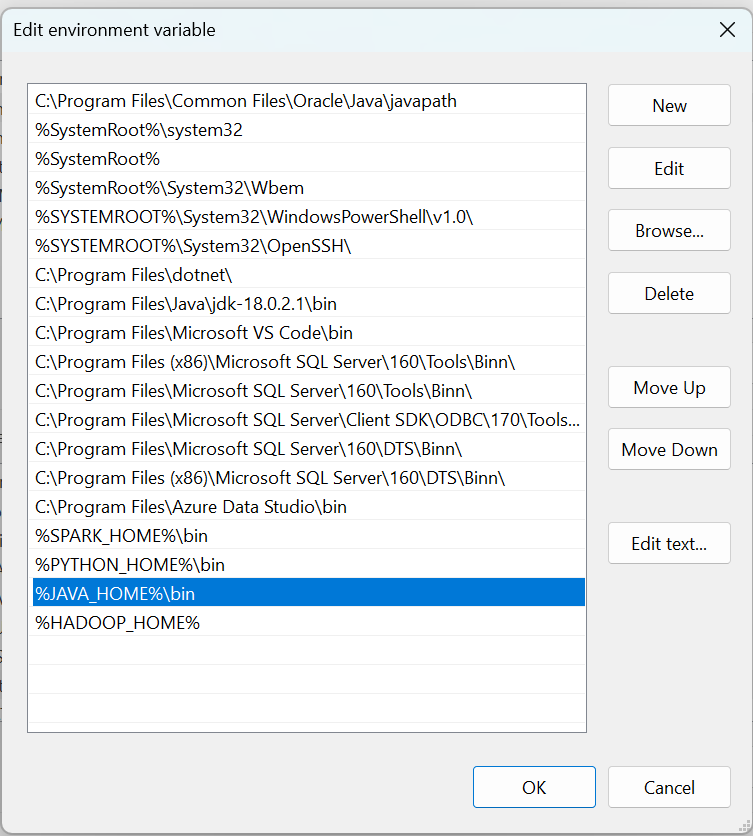
* Installation of PySpark steps
* Installation of JDK-19 (x64 installer)



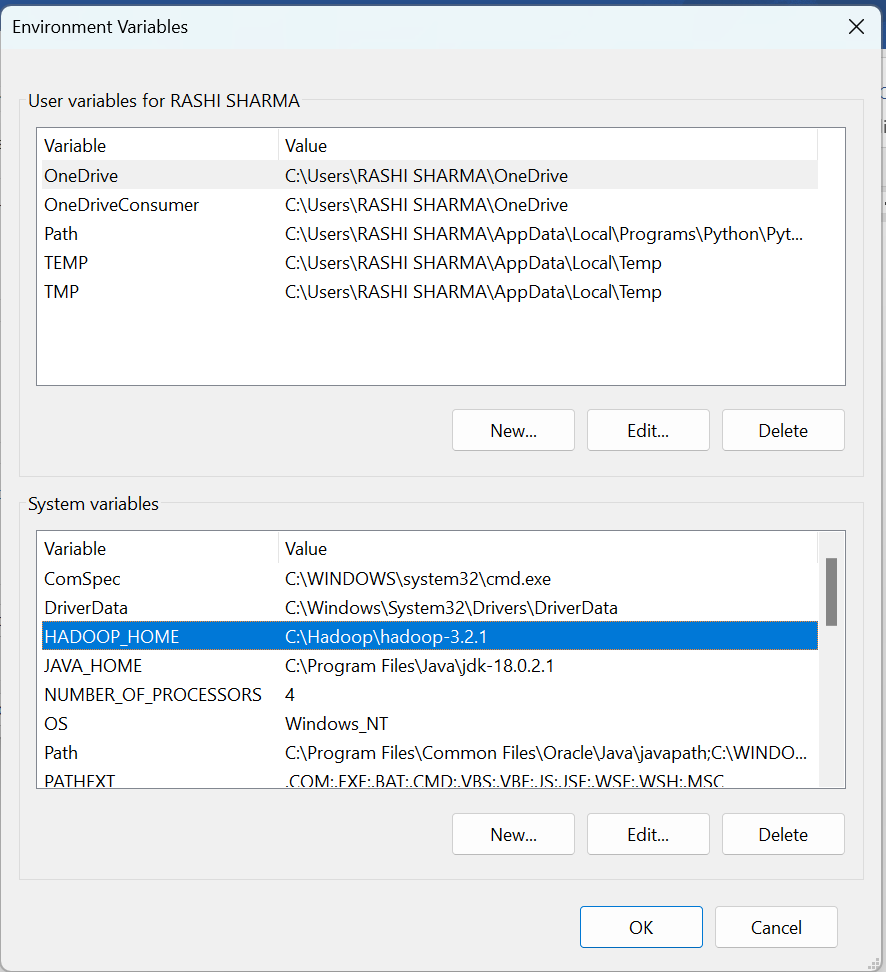
* Download the latest jdk version 19 from the above link. Now we have to set up jdk path in system variables.



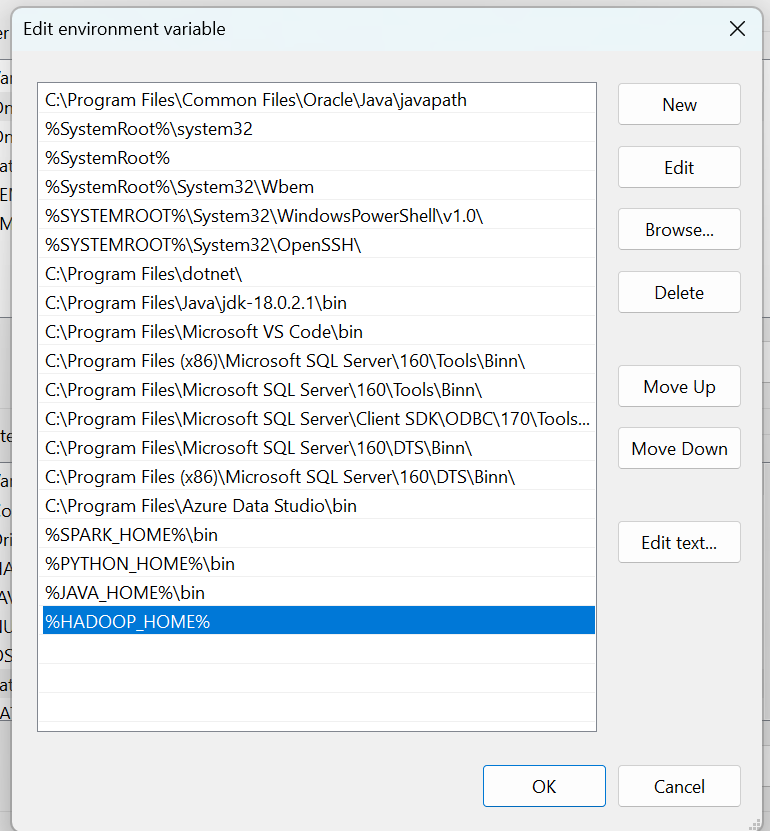
* Now provide java\_home in path.



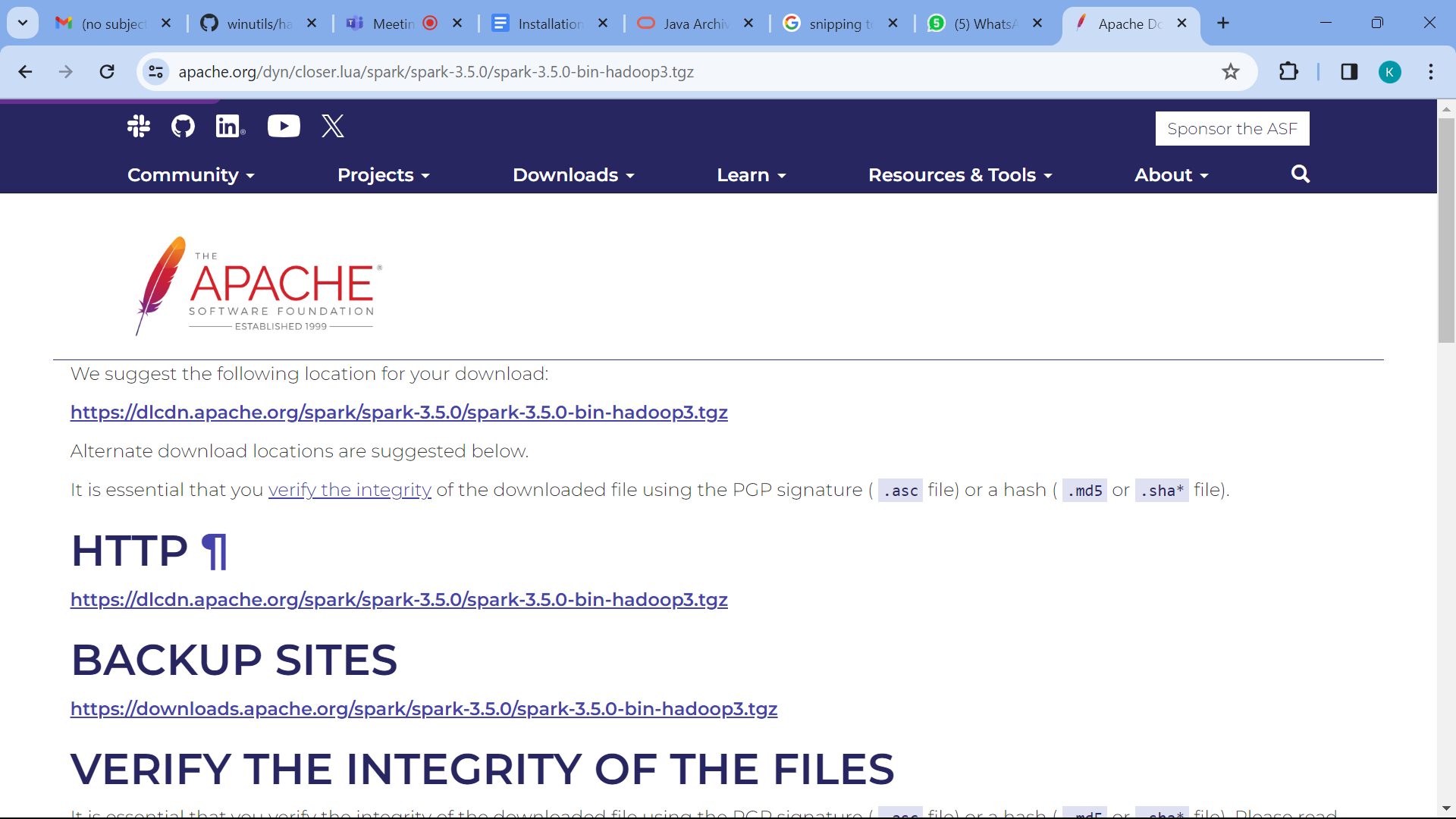
* Installation of Hadoop:
* Install the Hadoop 3.2.1.tar file in your local system. Now create Hadoop folder in c drive. Extract the files from tar file and paste those in Hadoop folder.
* Now we have to set up hadoop path in system variables.



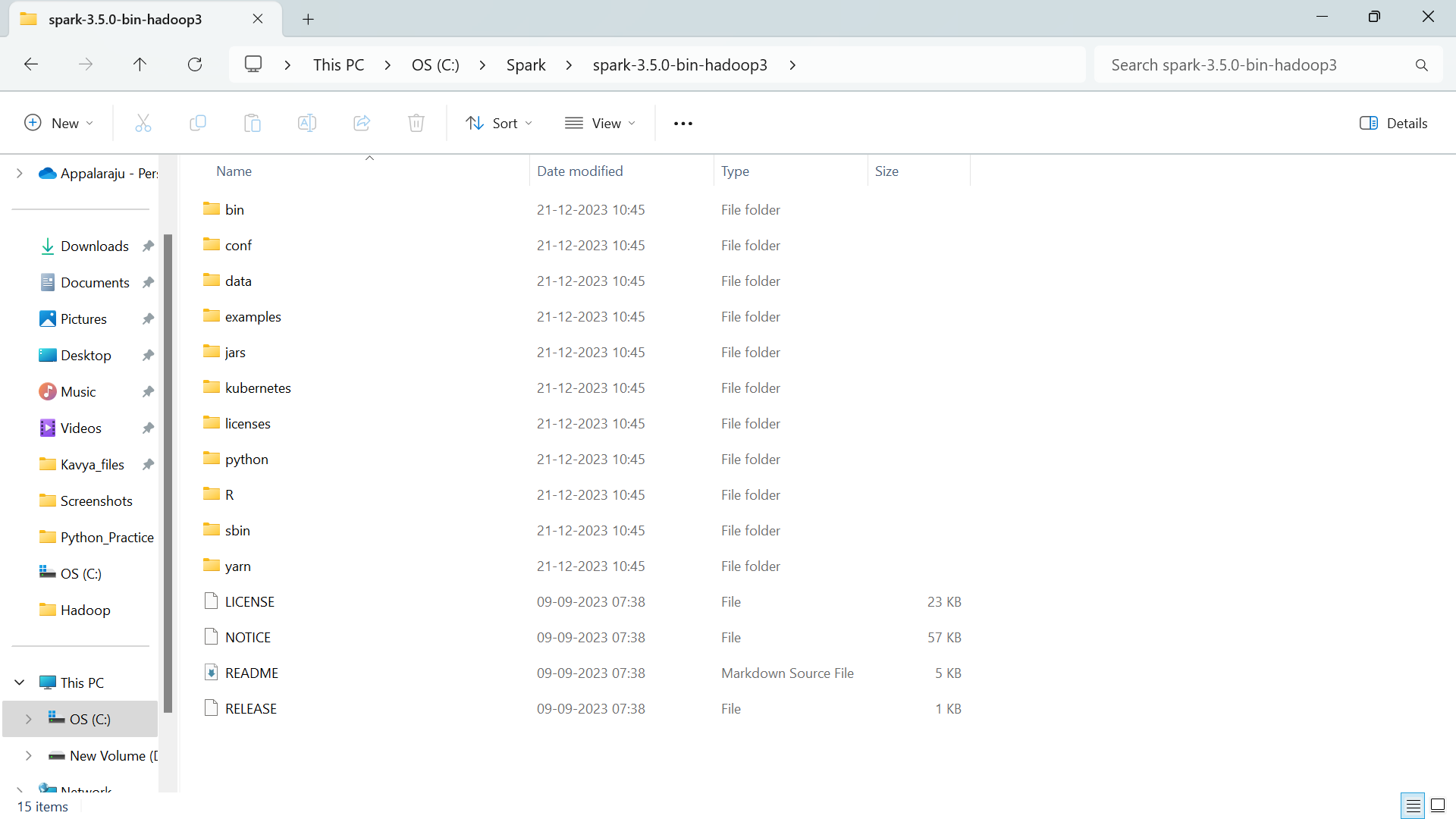
* Now provide hadoop\_home in path.

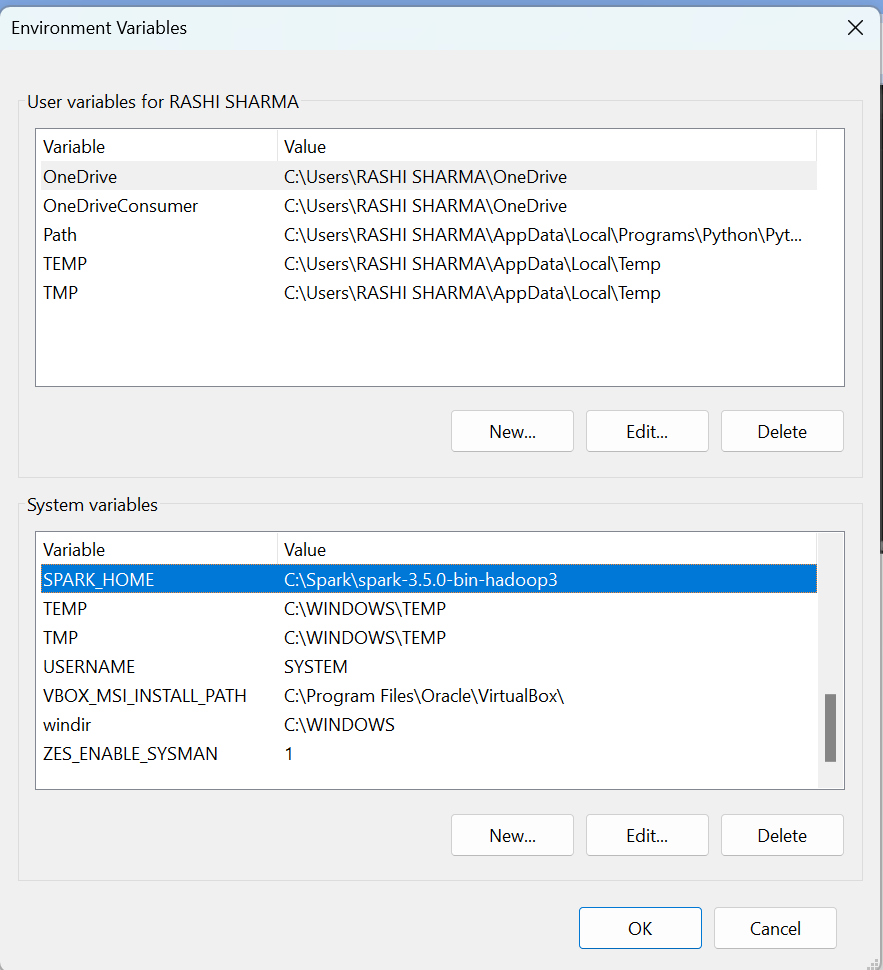


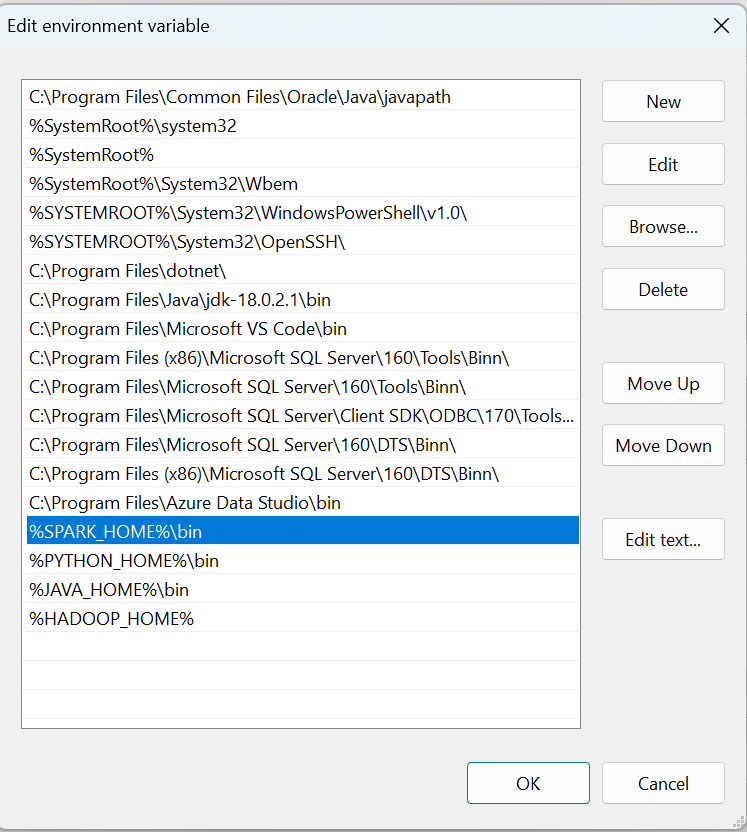
* Installation of PySpark
* We need to install pyspark in command prompt by giving command as “pip install pyspark.
* Install PySpark
* https://www.apache.org/dyn/closer.lua/spark/spark-3.5.0/spark-3.5.0-bin-hadoop3.tgz



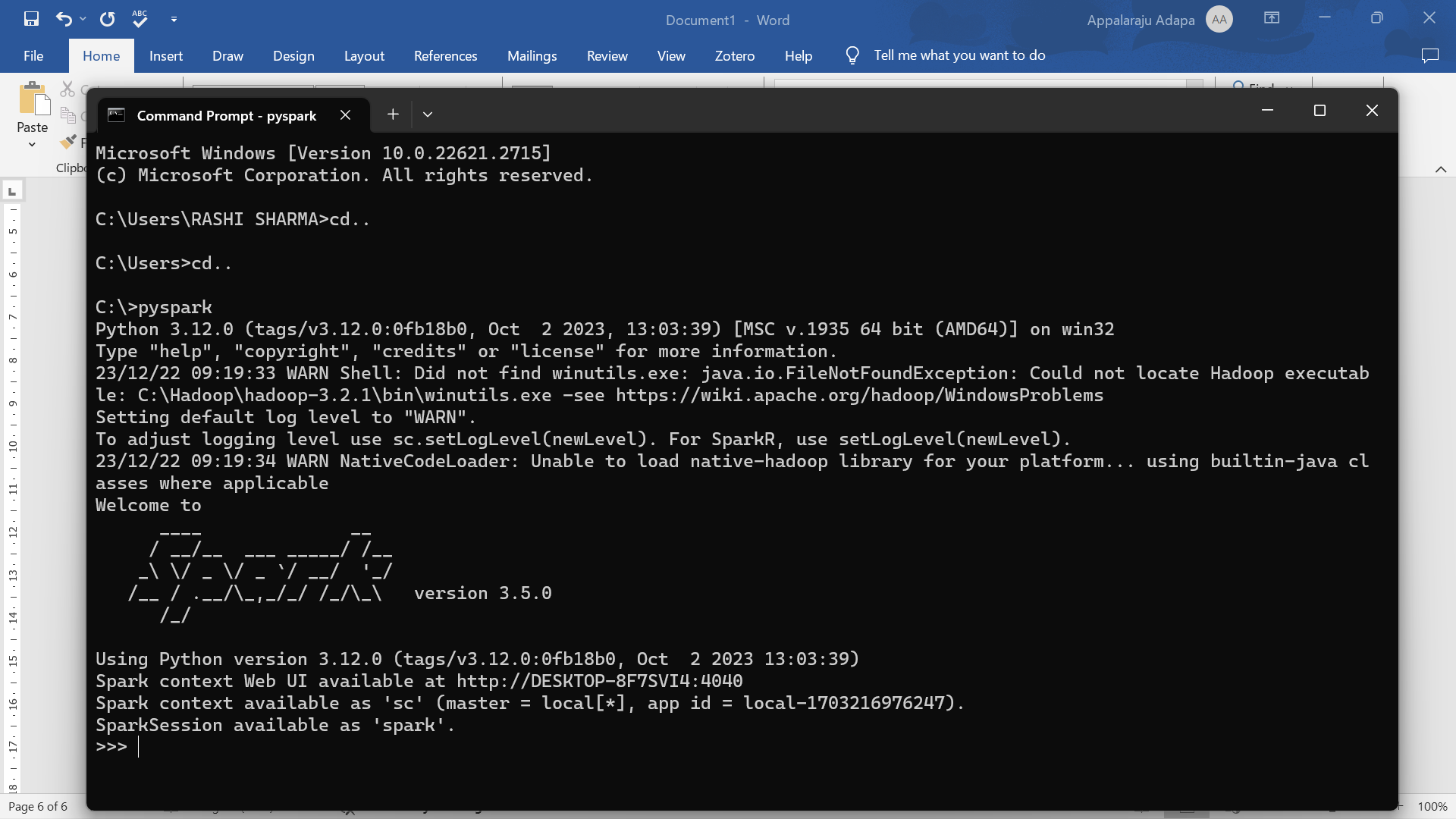
* After downloading the zip file, we need to install the spark file.
* We need to include the path in system variables.







* Now run pyspark in command prompt to check if it is installed or not.



* To connect to pyspark web ui – <http://DESKTOP-8F7SVI4:4040>
* We will get the web ui of pyspark

