**PySpark Assessment 1 : Pyspark Installation** - Abhiram Basa

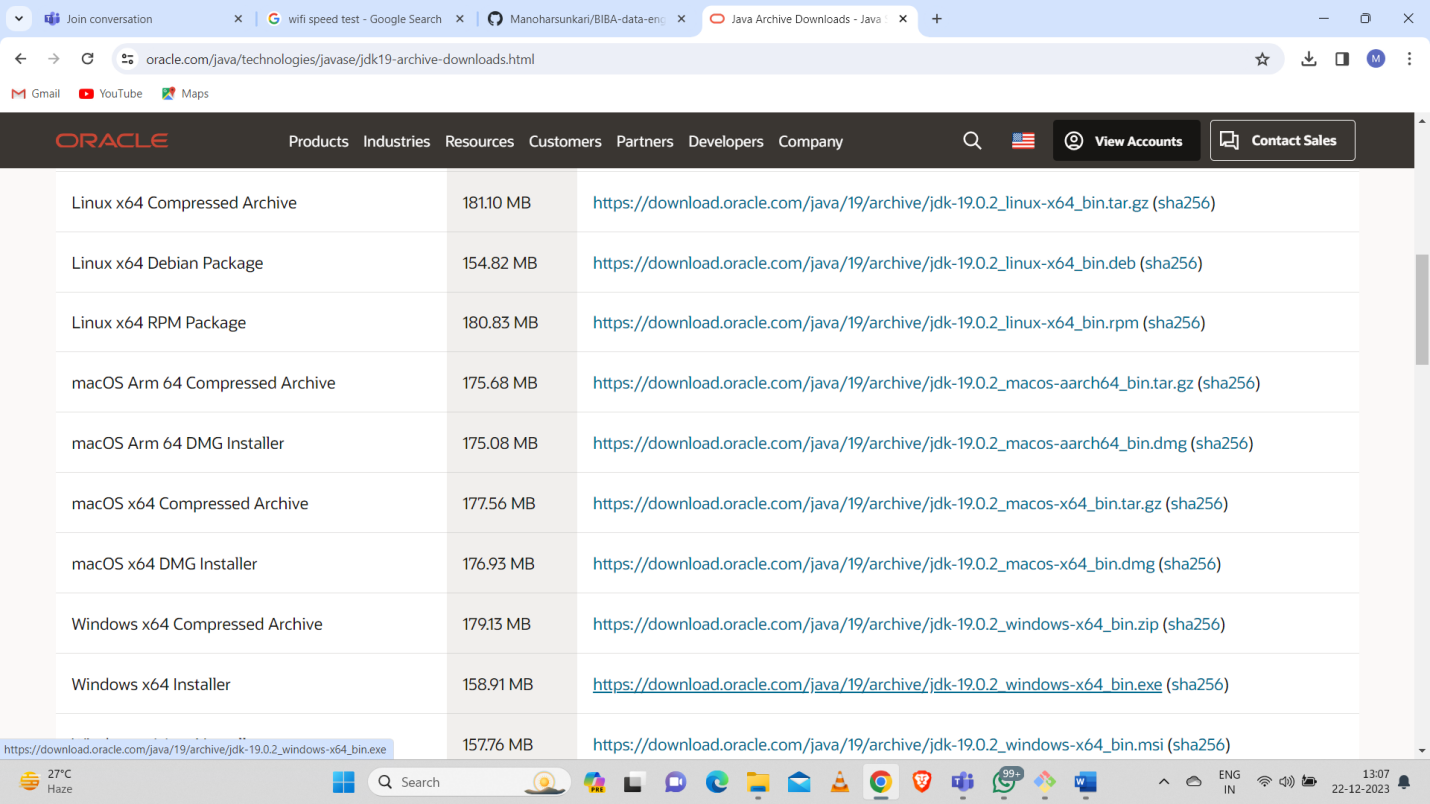
****PySpark**** is an Apache Spark library written in Python.

* PySpark is a Python API which is an analytical processing engine for large-scale powerful distributed data processing
* ****Apache Spark is an open-source**** unified analytics engine used for large-scale data processing.
* PySpark is used to its efficient processing of large datasets.
* Pyspark can be used in pycharm, jupyter notebook and anaconda idle.

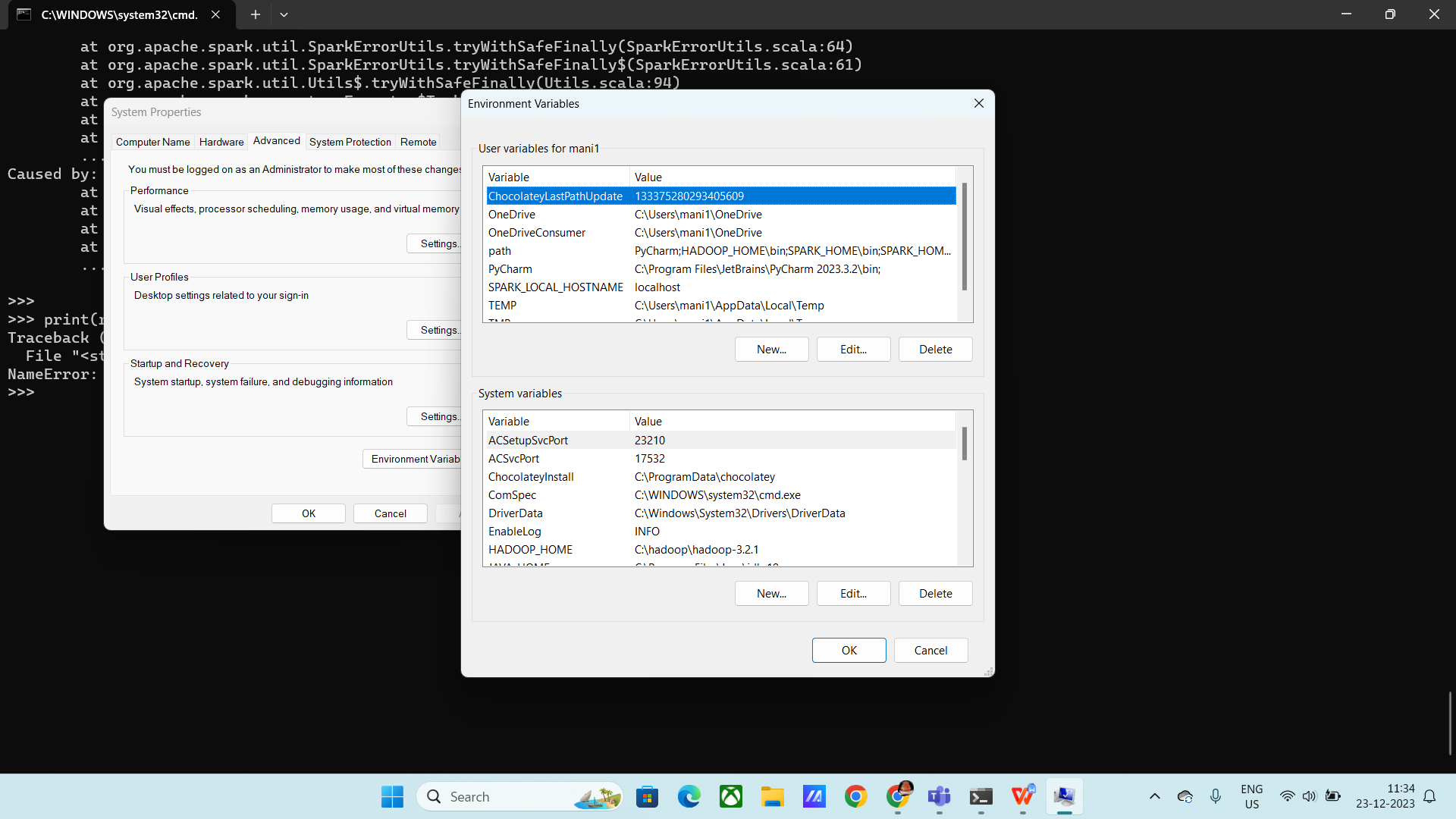
To work with Pyspark we need the following:

* Spark
* Hadoop
* Pyspark module

First installed JDK 19.0.2

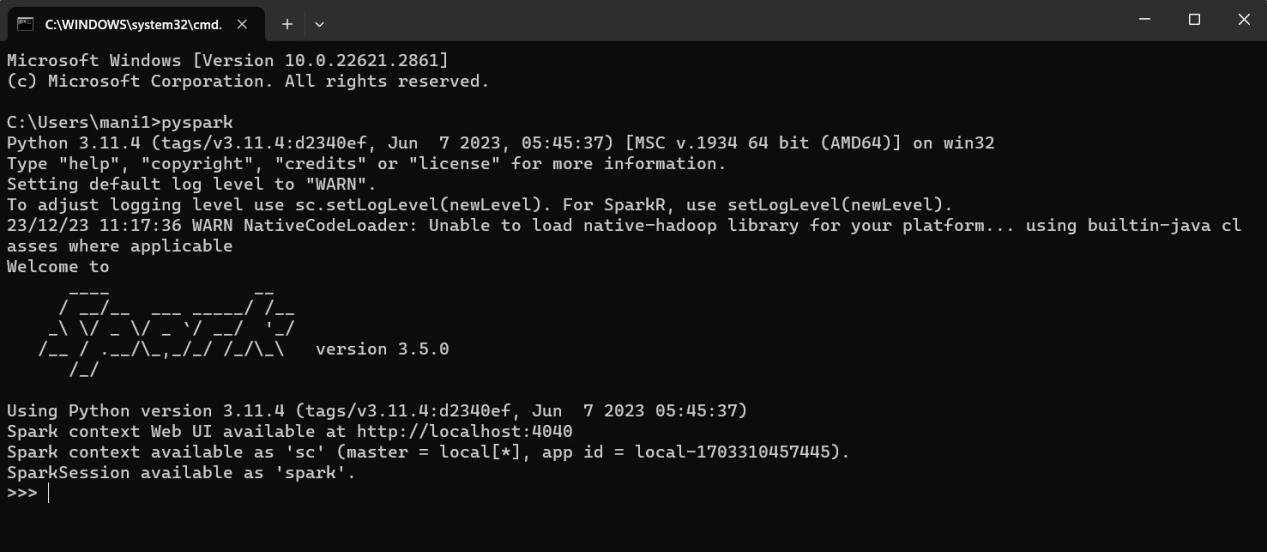


And then edit environment variables for java like JAVA-HOME and add the path of the java in files.



After variable creation in system variables changed path to \bin

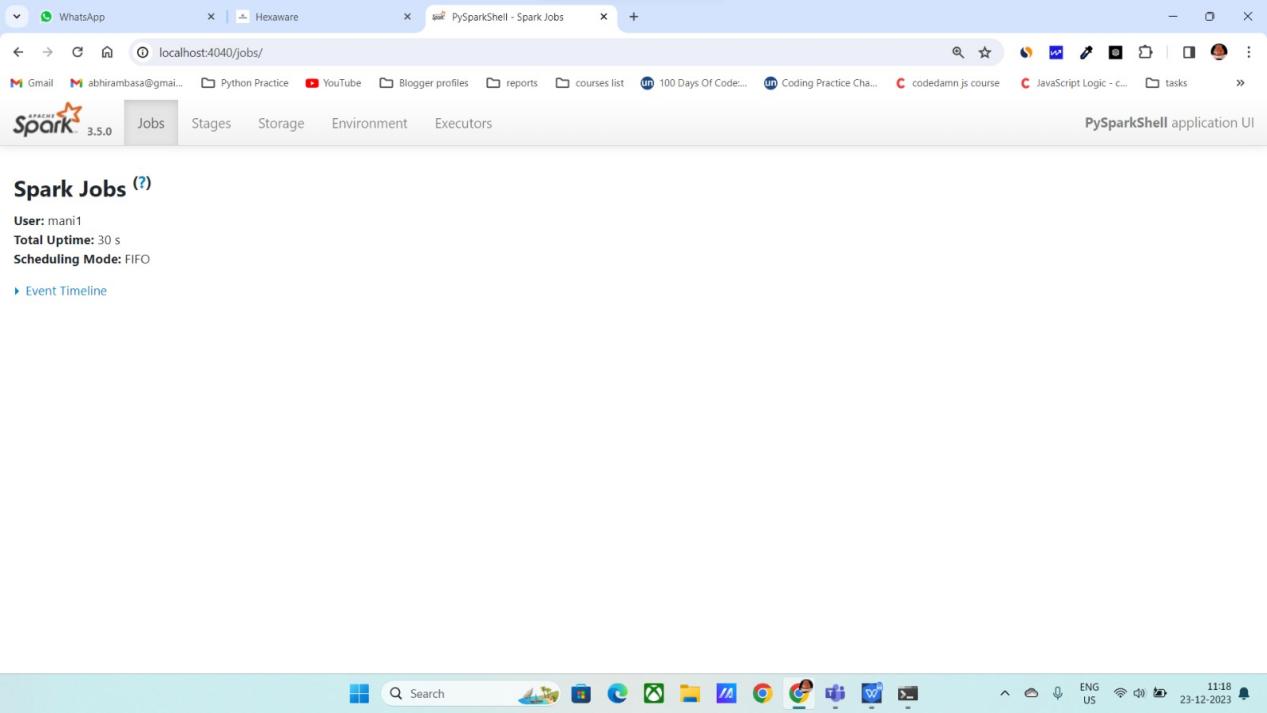
After this installations and editing environmental variables we need to open command prompt and type “pyspark” then it will appear like this:



The above shows pyspark installed successfully..

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

We will get the web ui of pyspark



Now succesfully installed the Pyspark and we can use it to run the programs.