Mallu Sravana Sandhya Assessment-16 22/12/23

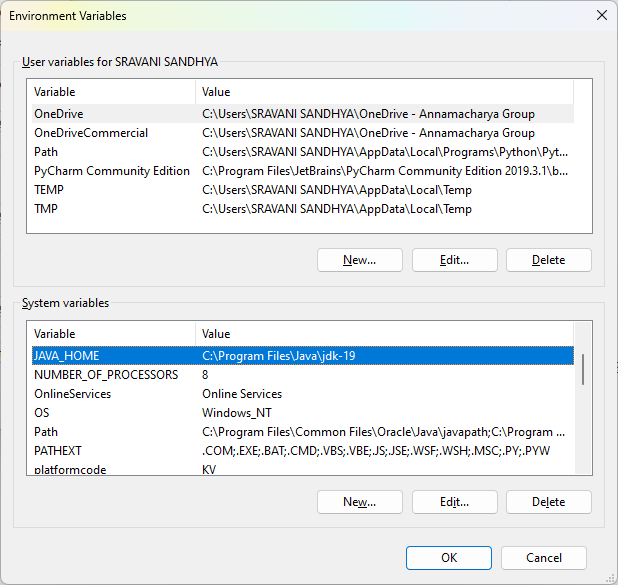
PYSPARK INSTALLATION AND SET UP IN THE LOCAL MACHINE/SYSTEM

To set up pyspark first we should have JDK,python of latetest version,hadoop,spark

First install java version more than 11 in the following website

[Java Archive Downloads - Java SE 19 (oracle.com)](https://www.oracle.com/java/technologies/javase/jdk19-archive-downloads.html)

And install the Java in the system by creating a java folder in local C drive and set the environmental variables in system and add path as shown in the follwing figure



Then install python of latest version and install it in the system the python can be downloaded in the below website

[Download Python | Python.org](https://www.python.org/downloads/)

And set environment variables and path of python as shown in above figure

Similarly download the hadoop from [Apache Download Mirrors](https://www.apache.org/dyn/closer.lua/spark/spark-3.5.0/spark-3.5.0-bin-hadoop3.tgz) website

And create a new folder hadoop in c drive and extract the hadoop files and then set the path in environmental variables

Download winutils.exe from below link and place the file in bin folder present hadoop

[winutils/hadoop-3.2.1/bin/winutils.exe at master · cdarlint/winutils (github.com)](https://github.com/cdarlint/winutils/blob/master/hadoop-3.2.1/bin/winutils.exe)

Same as Hadoop download the spark and set the path in environmental variables

ENVIRONMENTAL VARIBLES:

HADOOP\_HOME:C:\Hadoop\hadoop-3.2.1

JAVA\_HOME:C:\Program Files\Java\jdk-19

PYTHON\_HOME:C:\Users\SRAVANI SANDHYA\AppData\Local\Programs\Python\Python312\python.exe

SPARK\_HOME:C:\Spark\spark-3.5.0-bin-hadoop3

Requried path:

%HADOOP\_HOME%\bin

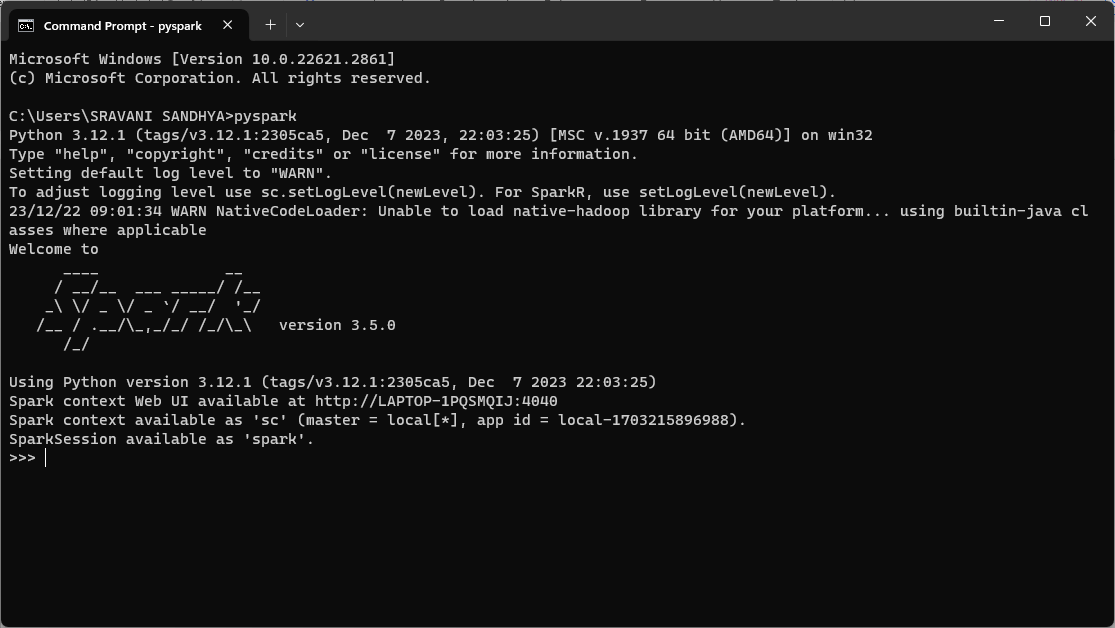
%JAVA\_HOME%\bin

%PYTHON\_HOME%\bin

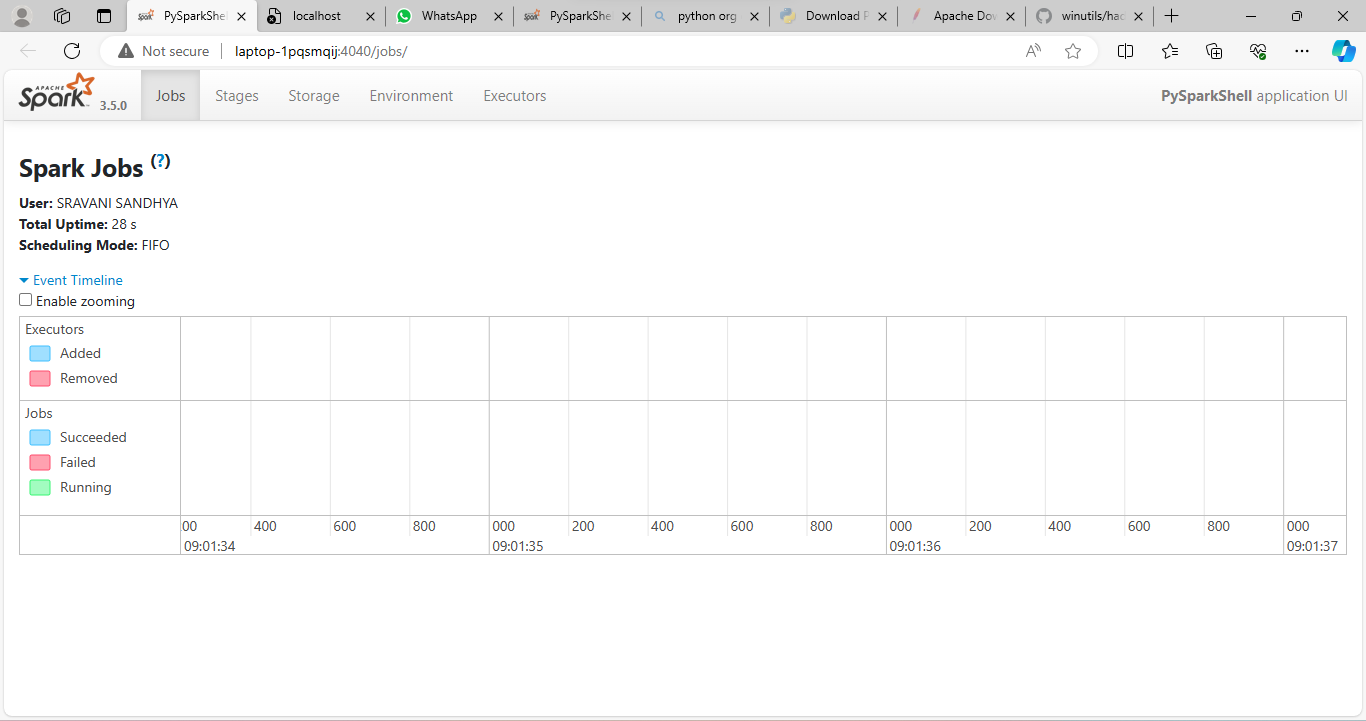
%SPARK\_HOME%\bin

After successful download and installation and setting up of path in environmental variables make sure that enviromental variables and path are as shown above

go to command prompt and type pyspark it will shows like in the picture below



By clicking the web ui we can go to the website



Or by typing http://localhost:4040/jobs/ this link in browser we can get into the website as shown below

