**Name -Monika Gariya**

**Email-** [**monikagariya2023@gmail.com**](mailto:monikagariya2023@gmail.com)

**Data Engineering Batch 1**

**Date – 03-01-2024**

**Topic – Installation of Apache Spark and Setup**

**Today’s Notes**

**Apache PySpark..**

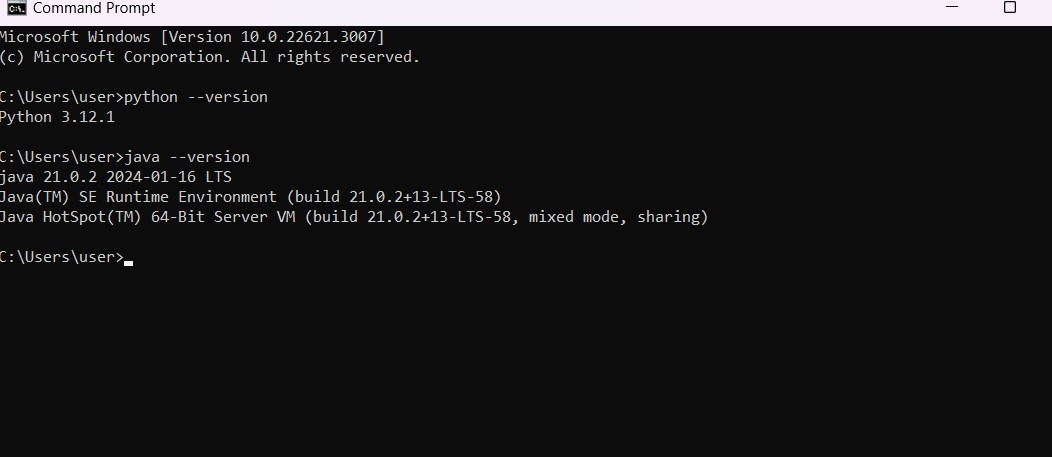
Apache Spark is an open-source distributed computing system that provides a fast and

general-purpose cluster-computing framework for big data processing. PySpark is the Python

API for Apache Spark, allowing you to write Spark applications using Python.

**Python+Spark=PySpark..**

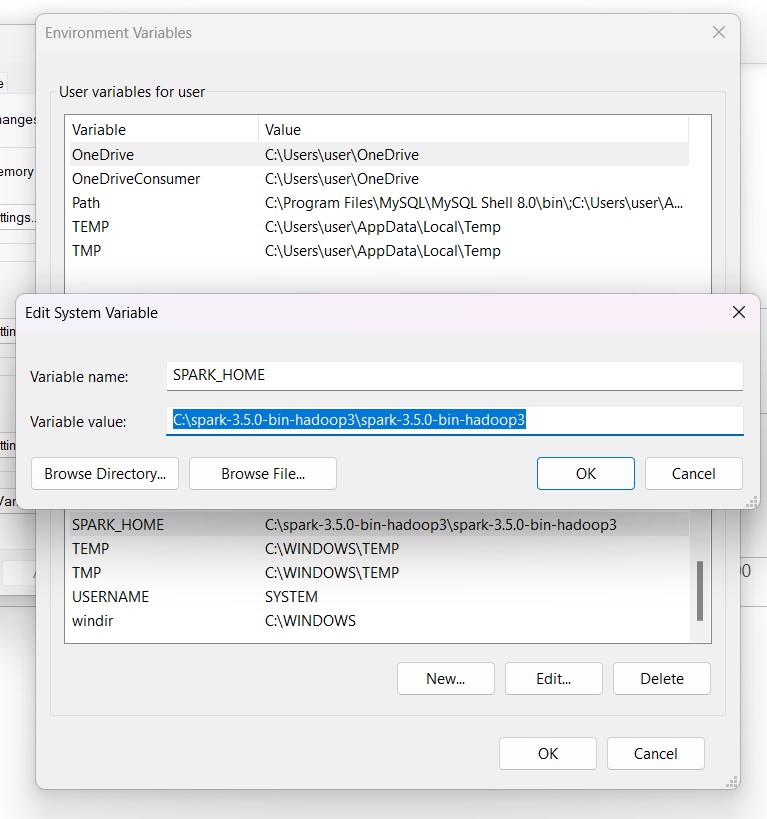
1)Install java and Python on your system

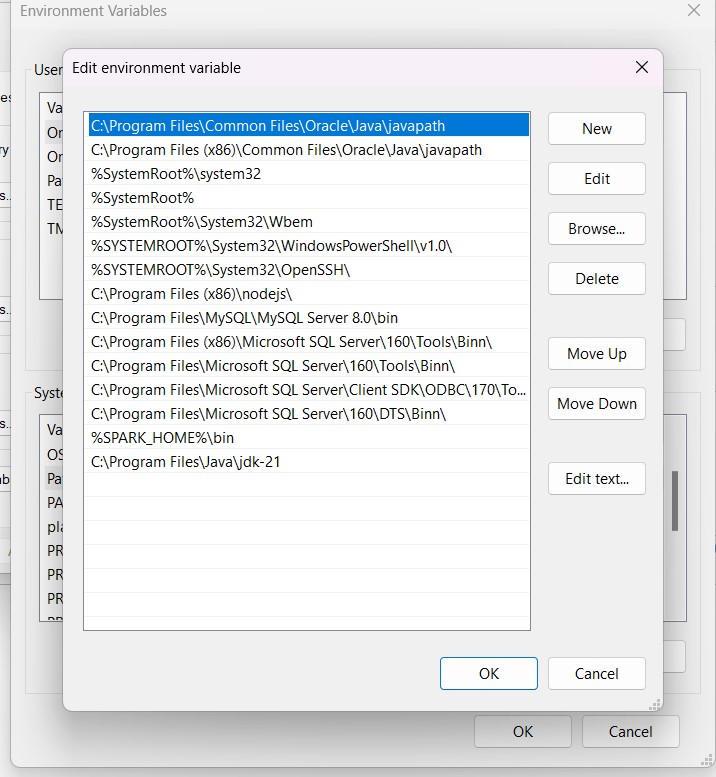


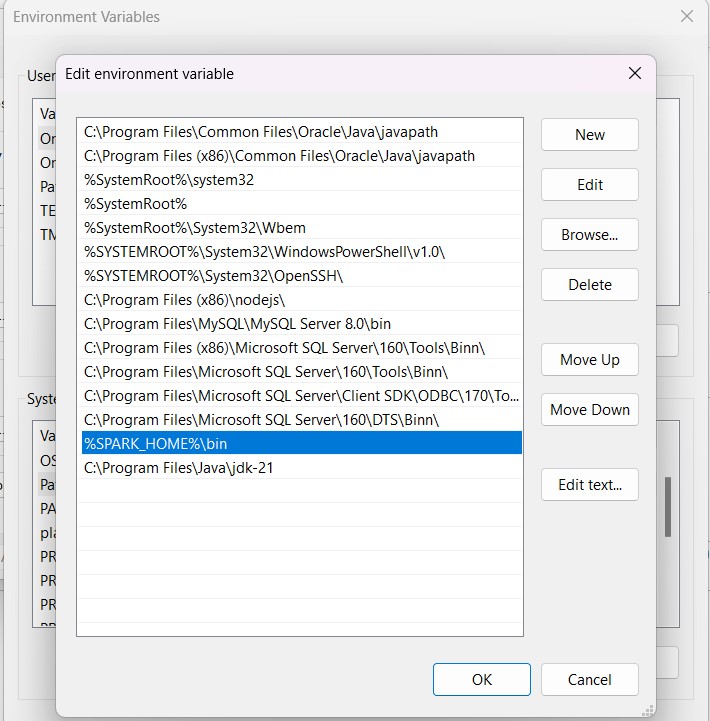
2) **Install Apache Spark**

* Visit the Apache Spark download page.
* Choose the latest version of Spark and download the pre-built package for Hadoop. It will be a tarball (.tgz) file.
* Extract the downloaded tarball to a location on your machine.

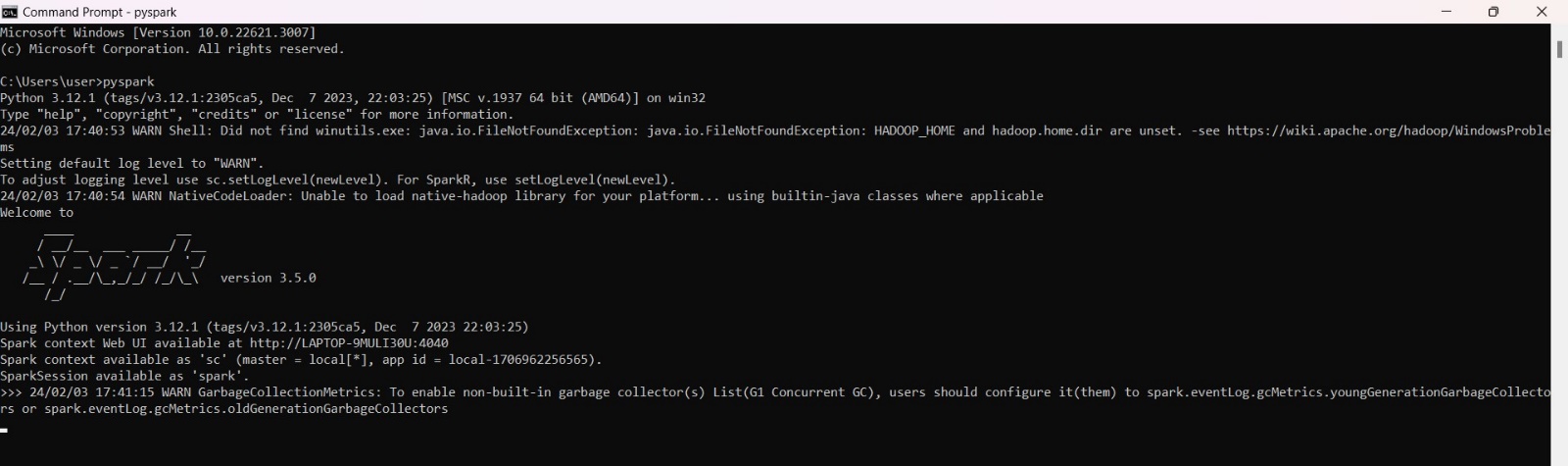
2)Go to the Environmental variable and set the path in the system variable as well as in path and click on ok.



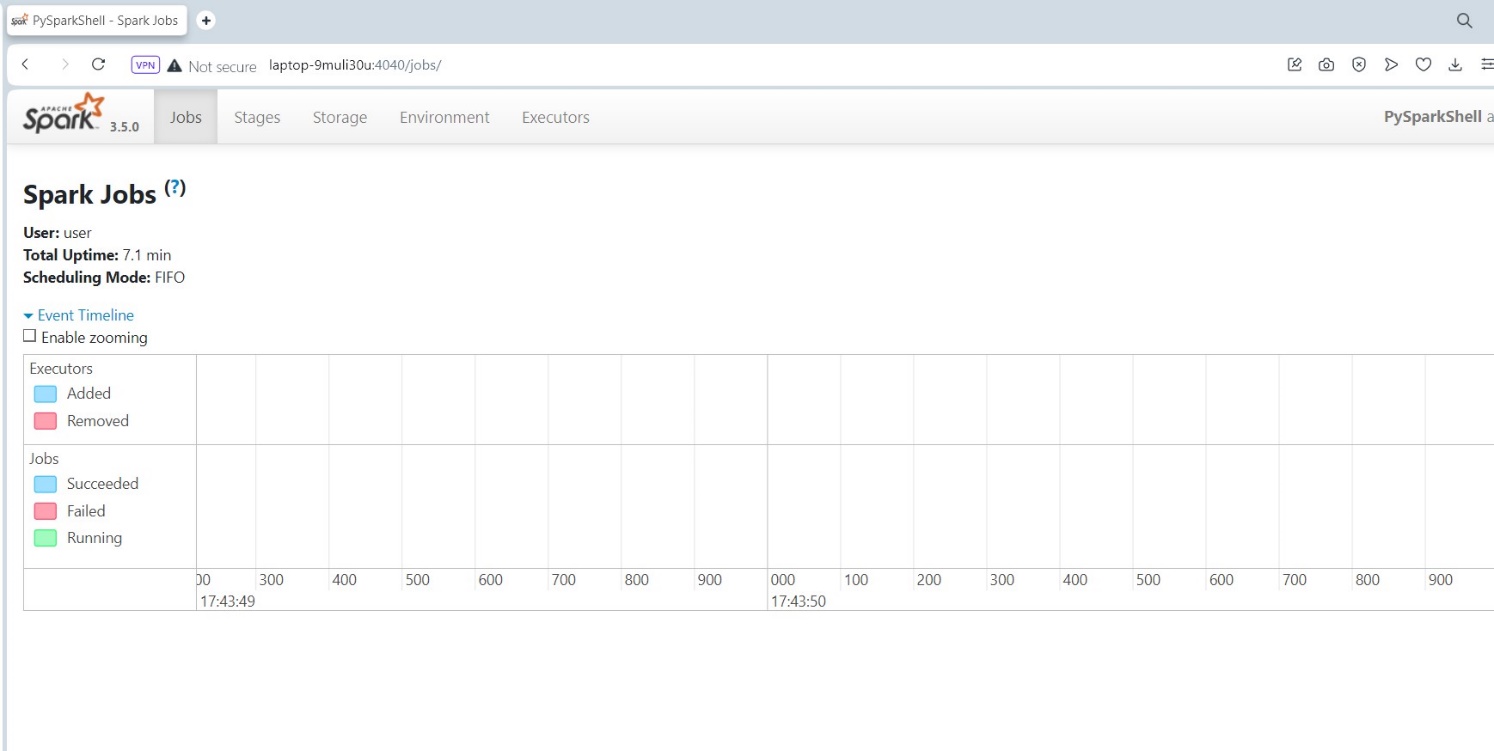




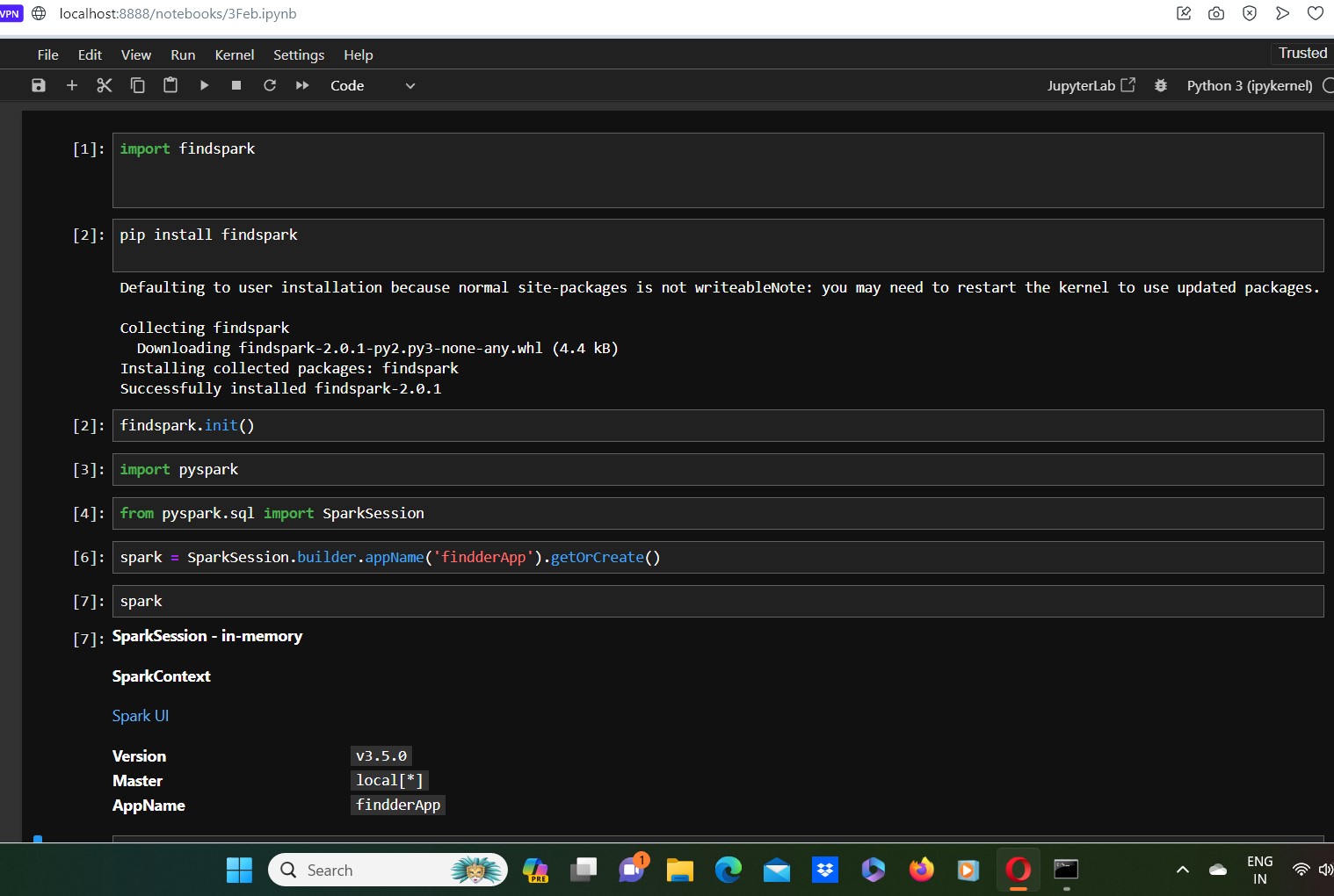
3) Open the command prompt and type “pyspark”.



4copy the link and open it in any browser. And Here we go-



5)Connecting with Jupiter notebook



Today’s notes

