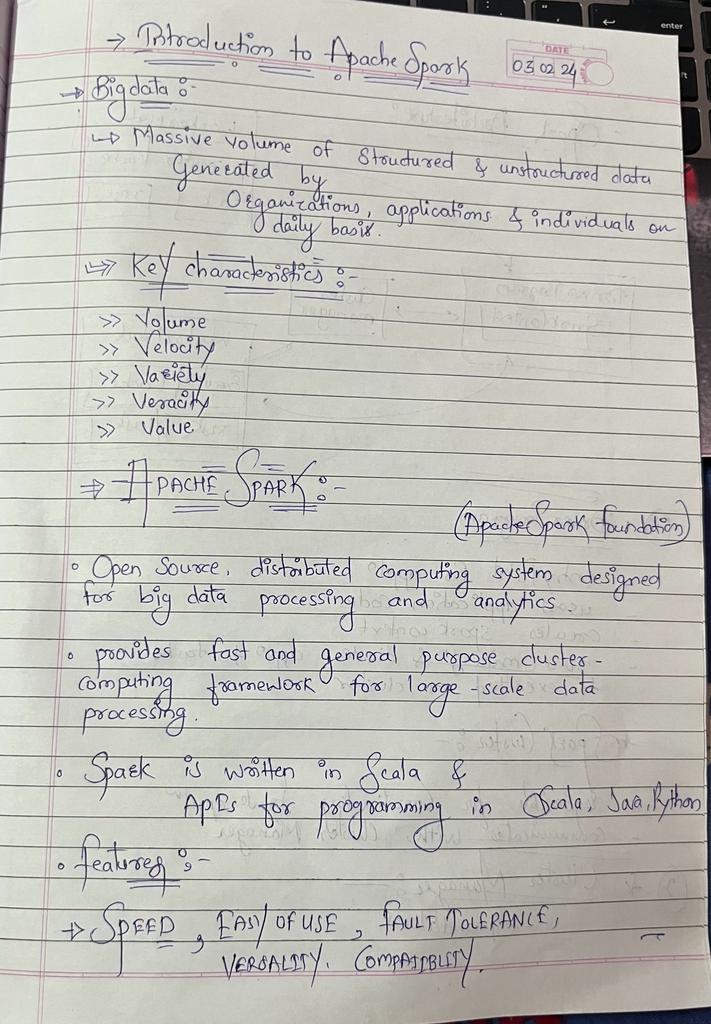
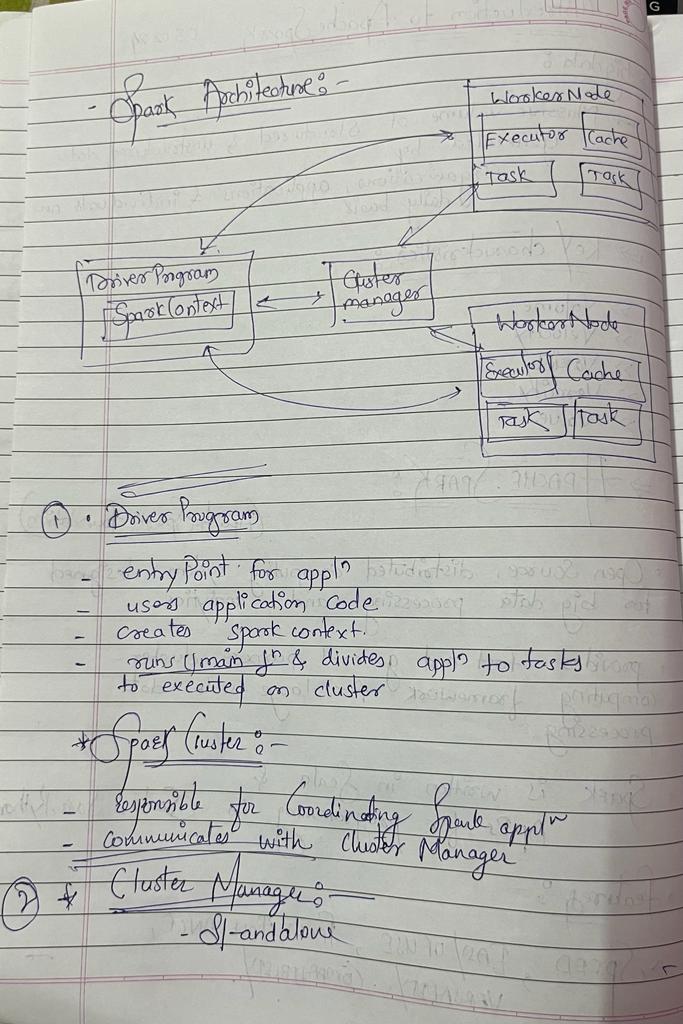
**NAME :- ANIKET SANJAYKUMAR BIYANI**

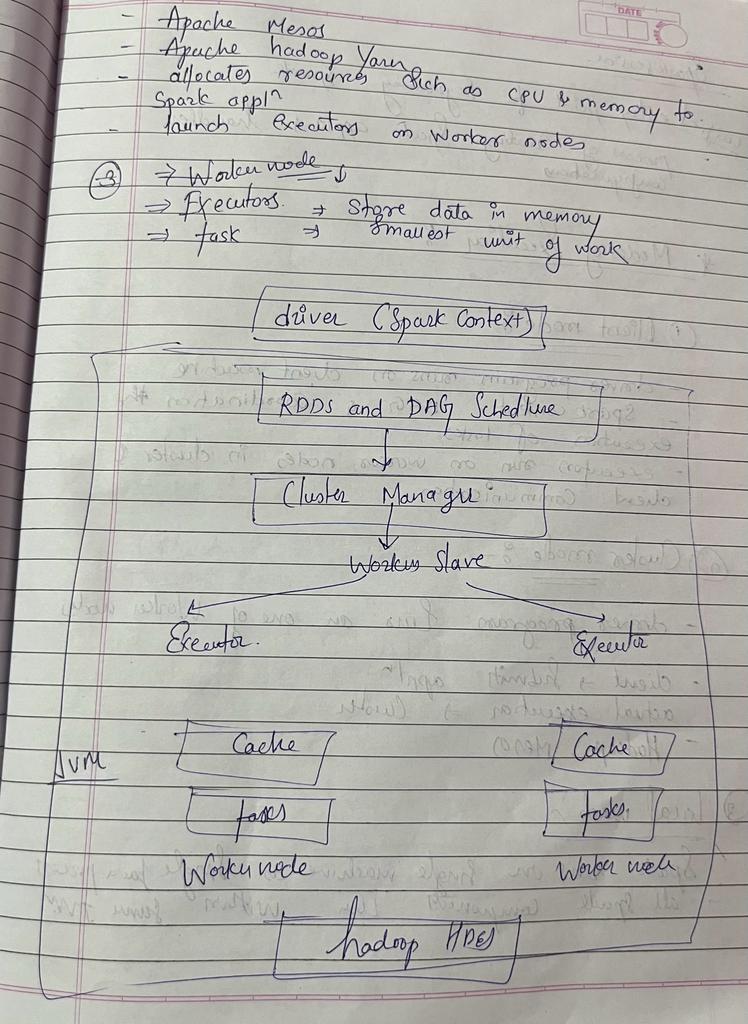
**DATA ENGINEERING BATCH -1**

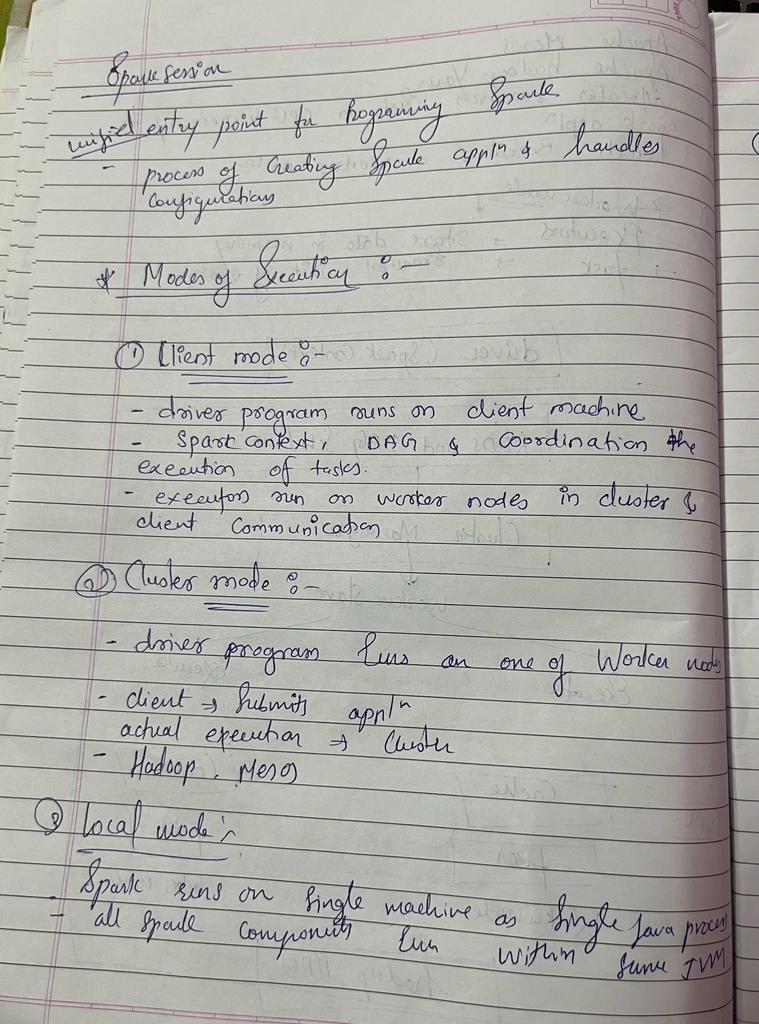
**PYTHON DAY-11 ASSIGNMENT**

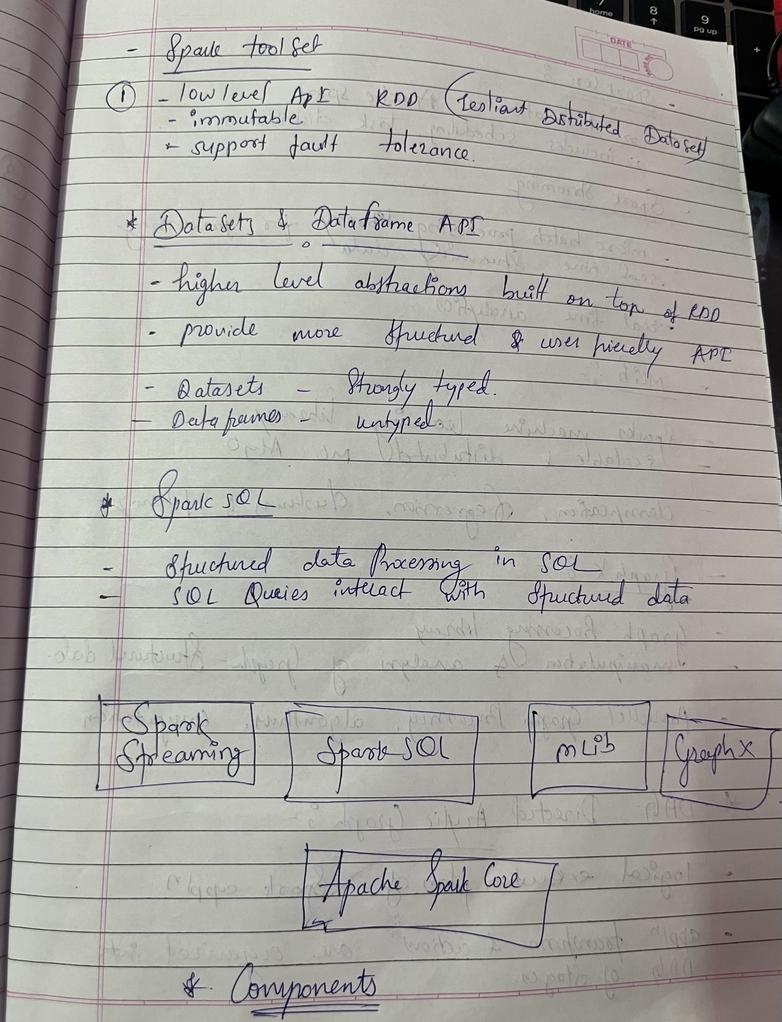
**INTRODUCTION TO APACHE SPARK HANDWRITTEN NOTES**

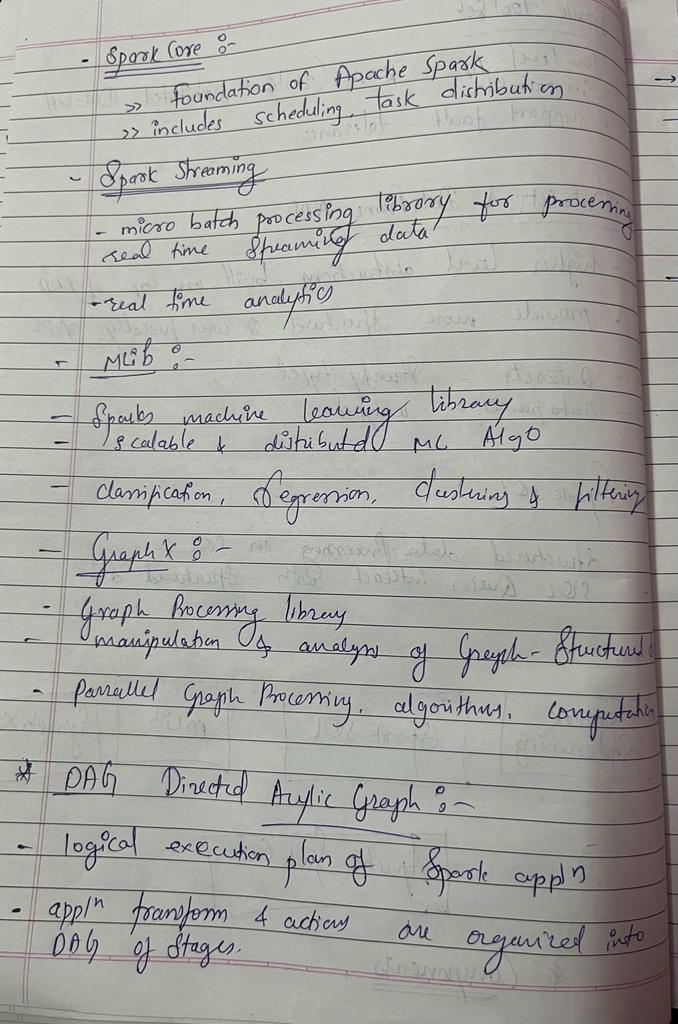
****

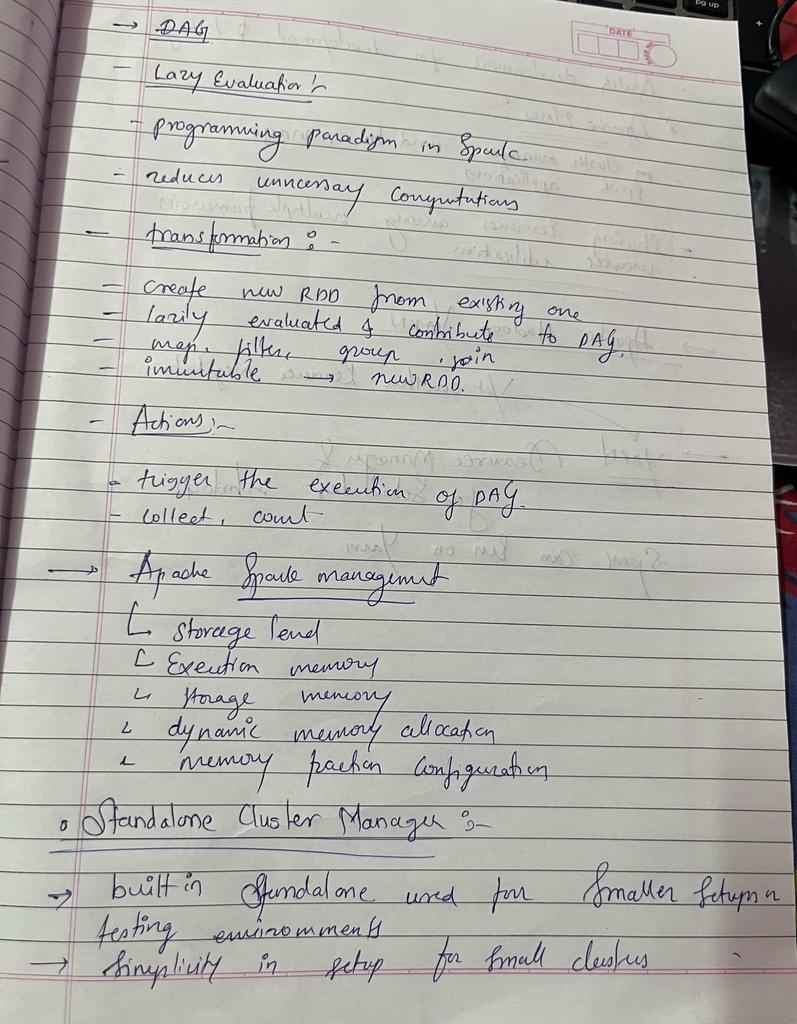


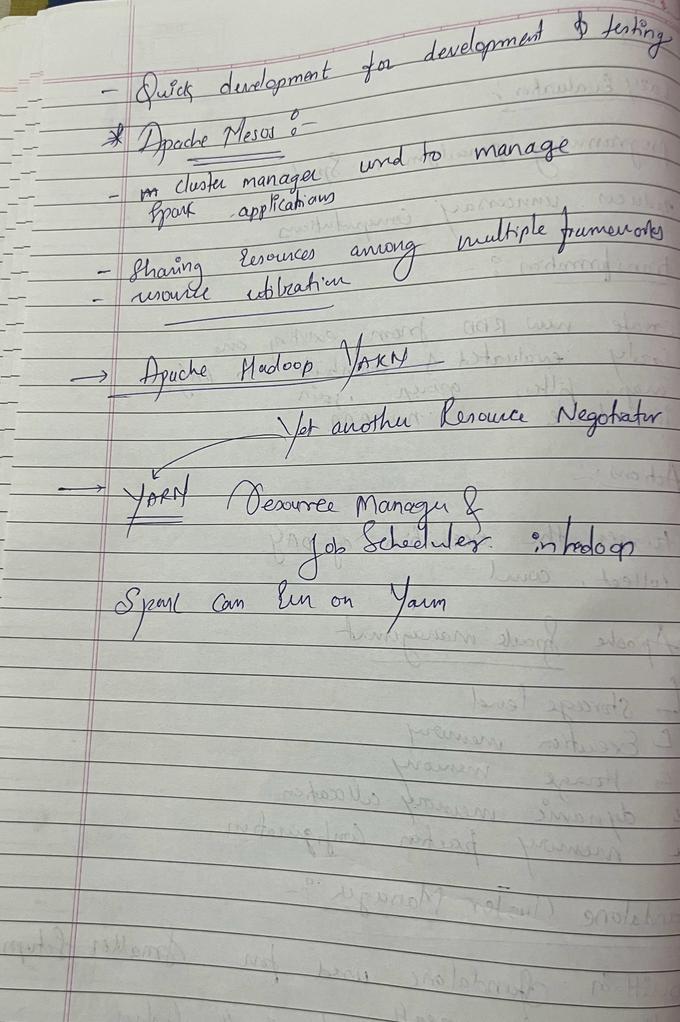










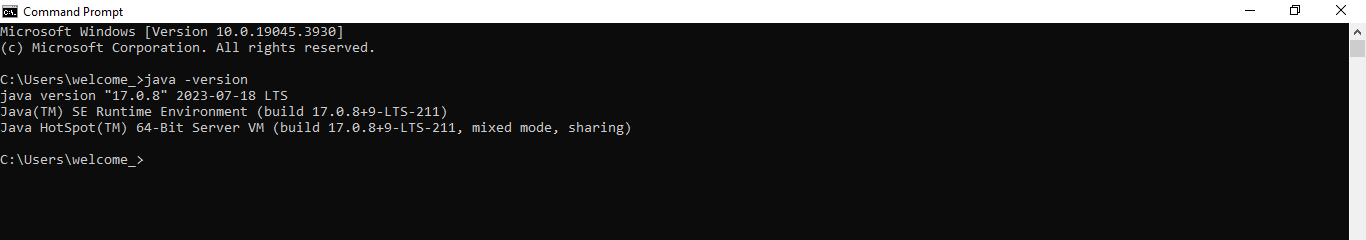


**Pyspark = Python + Apache Spark**

**Installing Required Softwares :**

**Java installation**

As java was already installed on my computer checked as if it is working fine or not



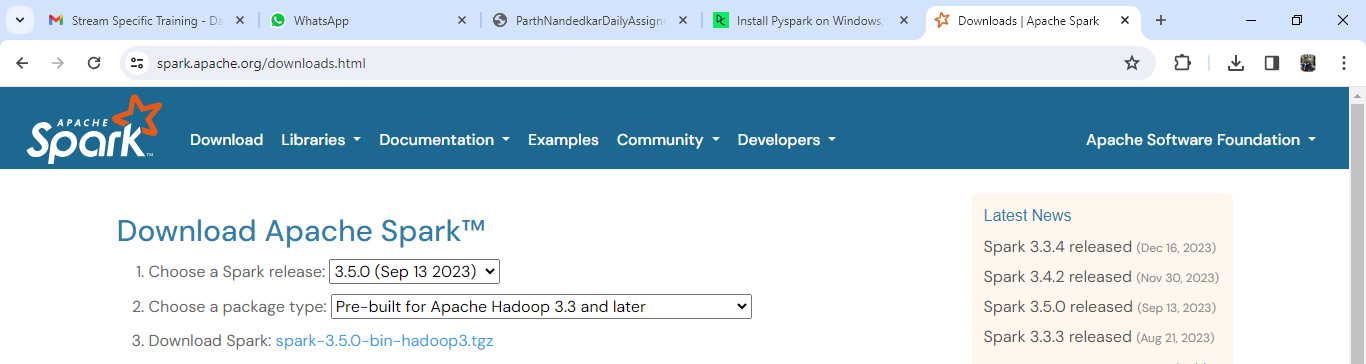
**Python Installation**

Python was also installed in the system also checked version for it

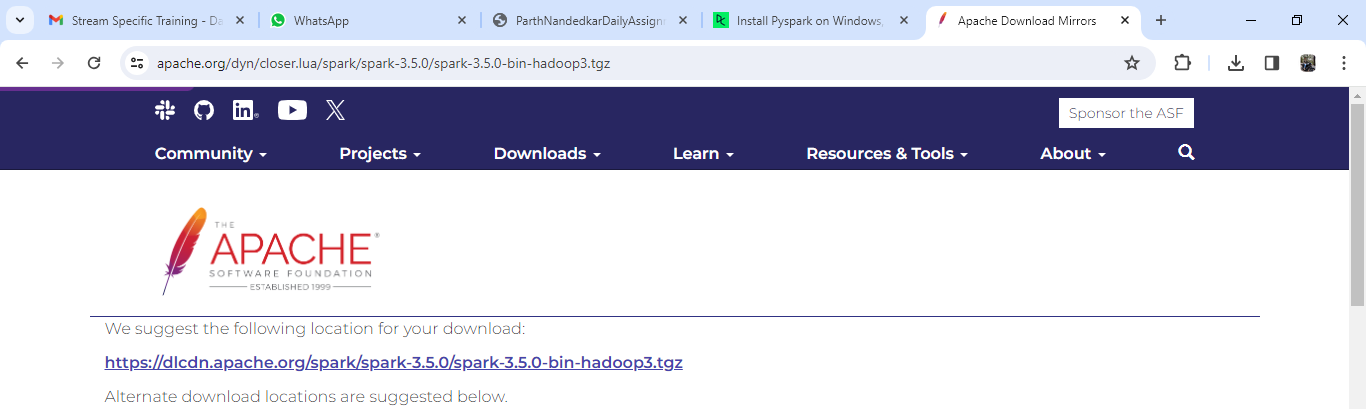


### Installing Pyspark

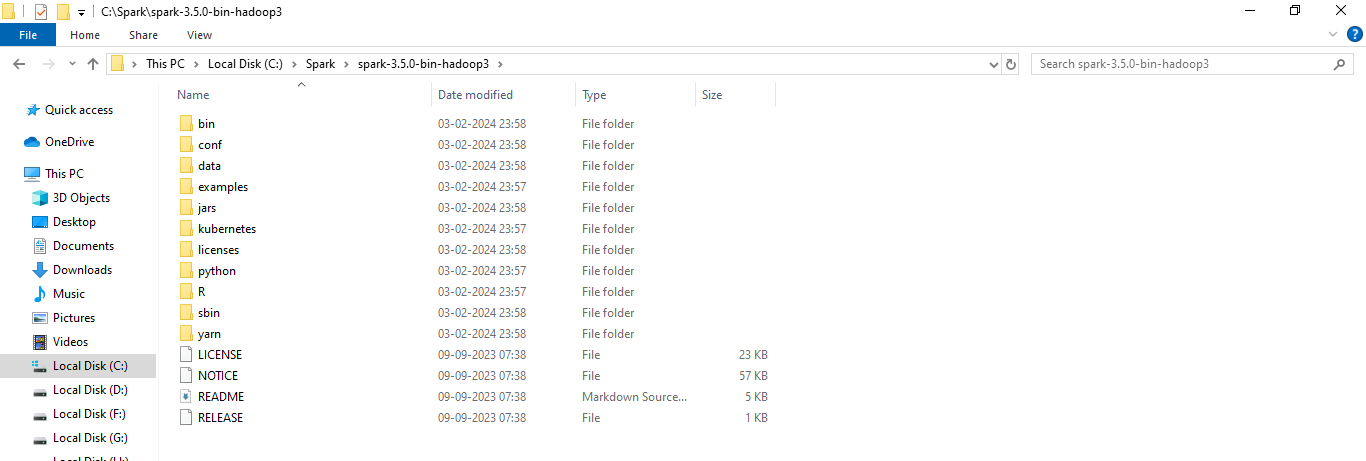
 Head over to the spark homepage



 Select the Spark release and package type as following and download the .tgz file.



**Extract Spark: •**

**Extract the downloaded ZIP file to a location on your machine (e.g., C:\spark\).**

### Download and setup winutils.exe

### Download using these link

### <https://github.com/cdarlint/winutils>

Extract Hadoop: Extract the downloaded Hadoop binary to a location on your machine (e.g., C:\hadoop).

**Using Environment Variables**

**Set path for**

**JAVA\_HOME - C:\Program Files\Java\jdk-17**

**SPARK\_HOME C:\spark\extractedfile\_name**

**PYTHON**

**HADOOP\_HOME c:hadoop\bin**

### C:\Users\welcome_\Pictures\Screenshots\Screenshot (839).png

**Configure System Environment Variables**

**Edit System Path:**

**Add the following entries to the system PATH environment variable:**

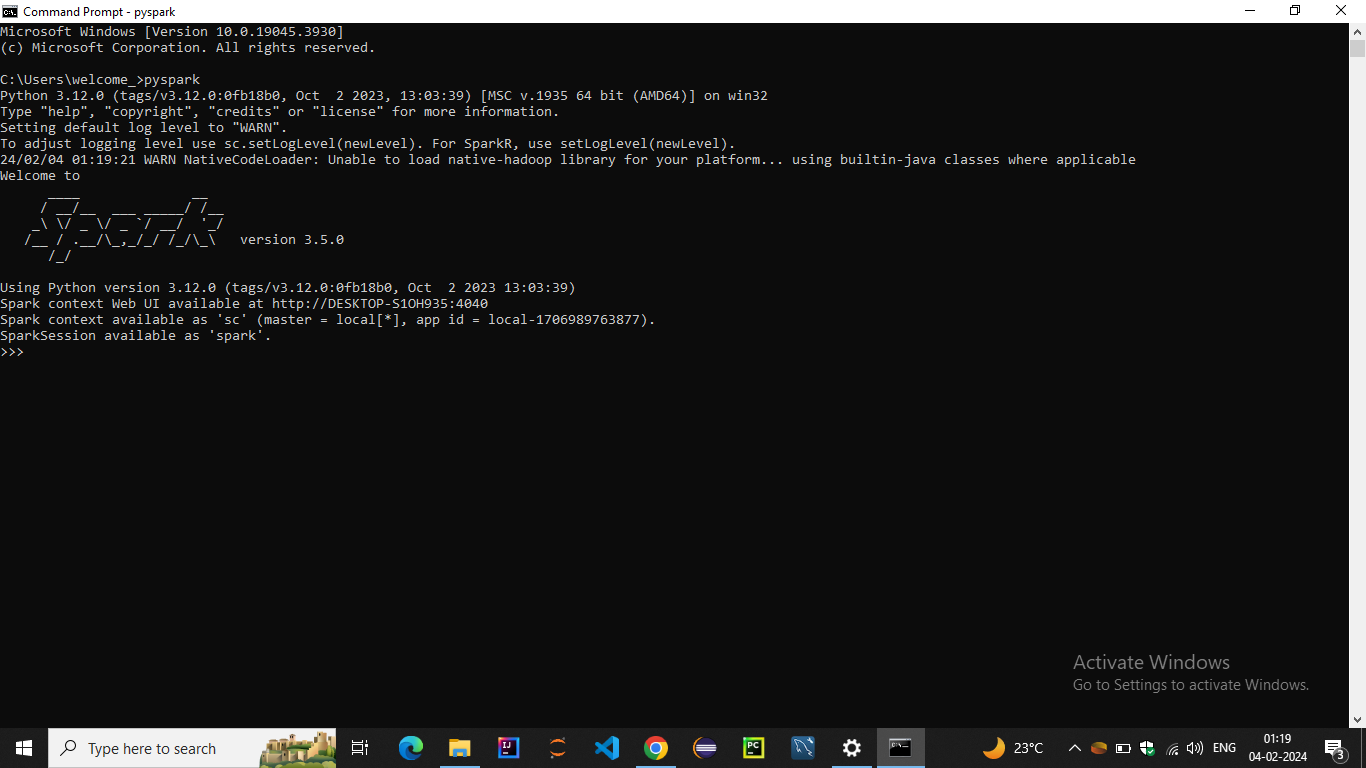
**%JAVA\_HOME%\bin**

**%SPARK\_HOME%\bin**

**Run on cmd or powershell**

**Pyspark**

**OUTPUT**

****