**Assignment Day- 11**

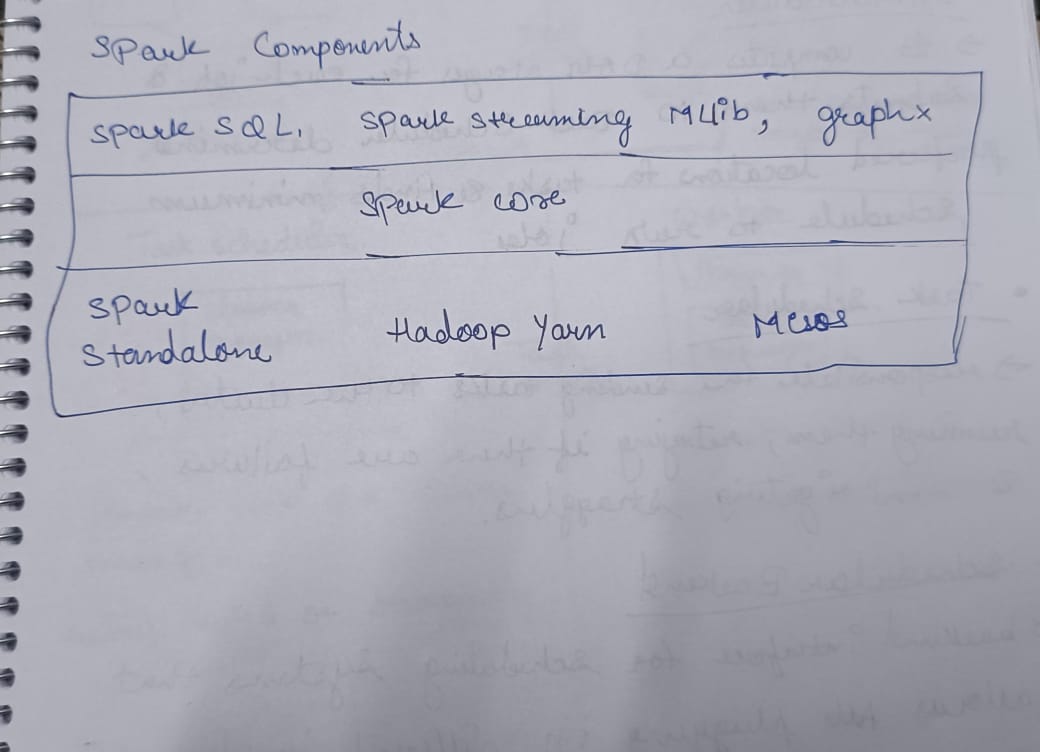
**Name: Kanak Agrawal**

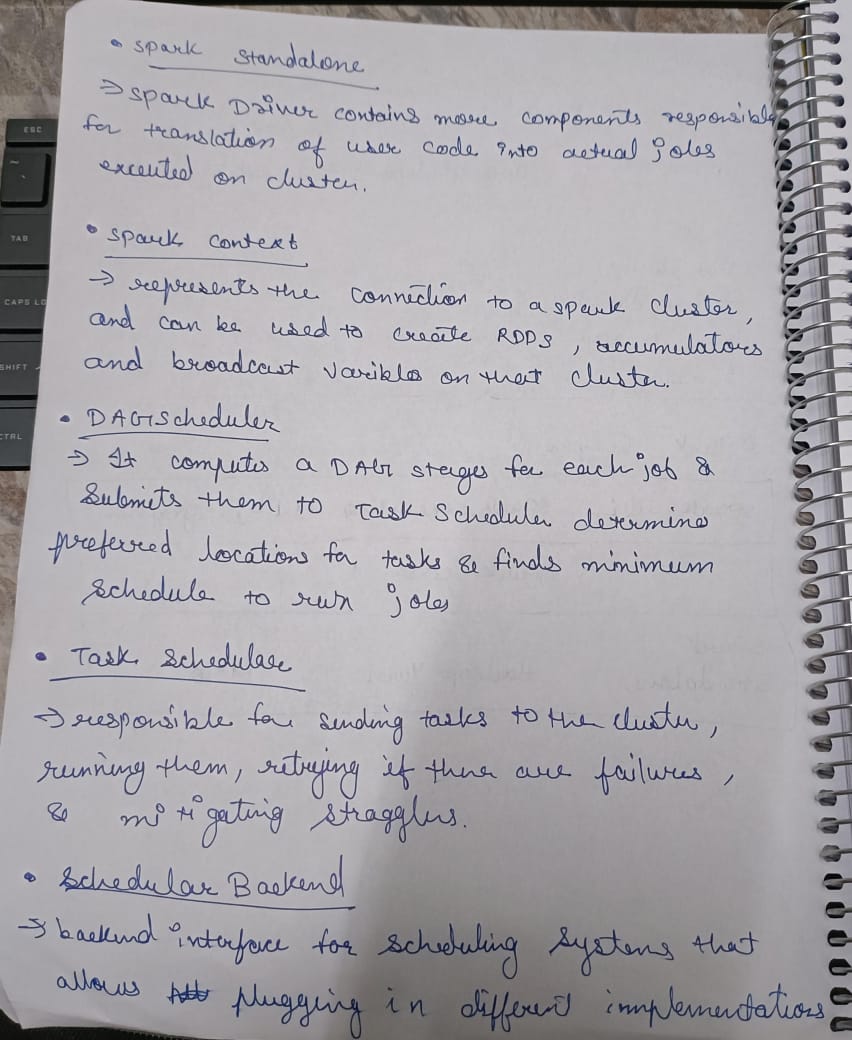
**Email:** [**kanakagrawal2001@gmail.com**](mailto:kanakagrawal2001@gmail.com)

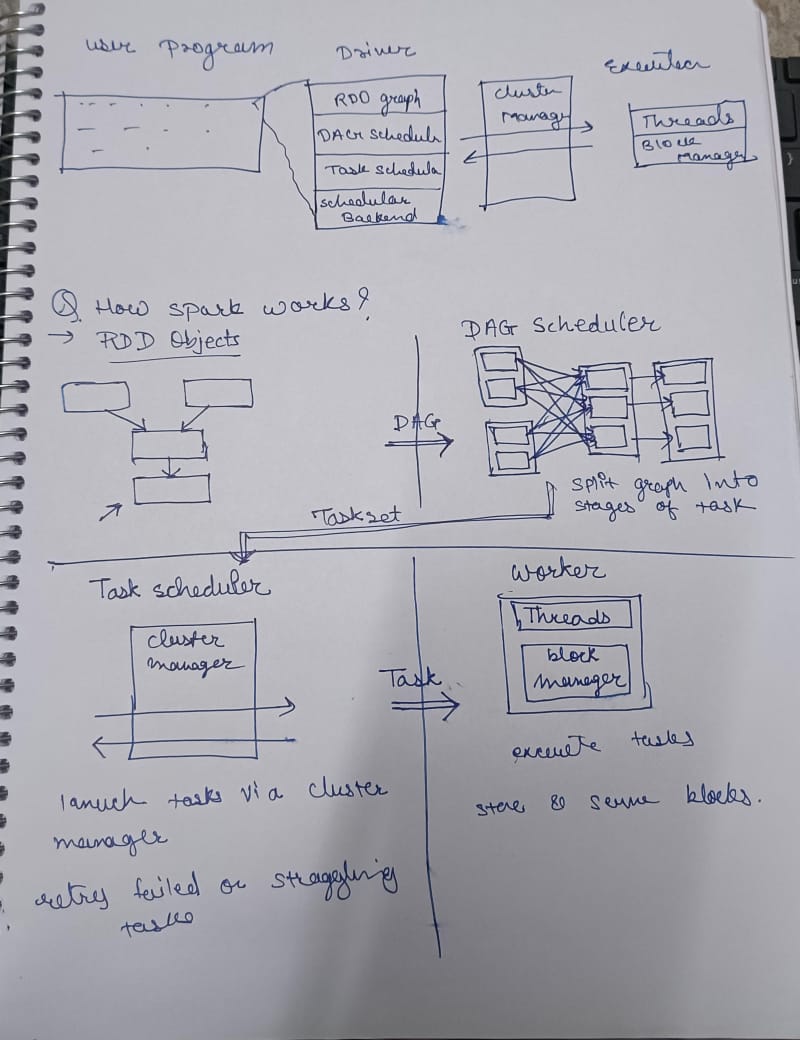
**Date: 03-02-2024**

**Notes on Apache Sparks**

Here is the image of notes





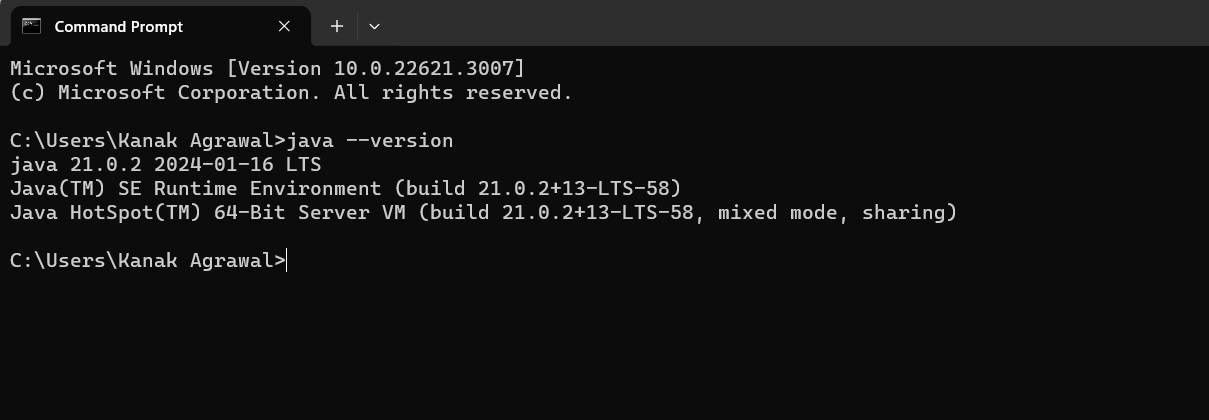


**Installation of Apache Spark and set up**

So firstly, to install pyspark we need to download java and python before downloading Apache spark

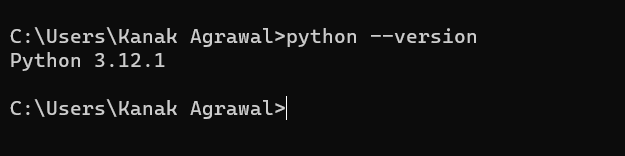
**Install JAVA**

* We need to install java jdk version which is compatible to our system.It is recommnended to download the latest version of jdk.
* After that we need to check in command prompt by typing **java –version** which will display the version of java which we have downloaded.

****

**Install Python**

* After installing java,we need to install python environment which is compatible to our system.It is recommnended to download the latest version of Python.
* After that we need to check in command prompt by typing **python –version** which will display the version of java which we have downloaded.

****

**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.

**Set Environment variable**

* In this firstly we will the add the path of all the three things which are HADOOP, SPARK, JAVA in environmental variable.
* And then we need to open command prompt and then write **spark-shell** which will get downloaded and after that we need to again write in command prompt **pip install pyspark** and it will get downloaded.

