Name: Rohan Vinayak Chaudhari

Batch: Data Engineering

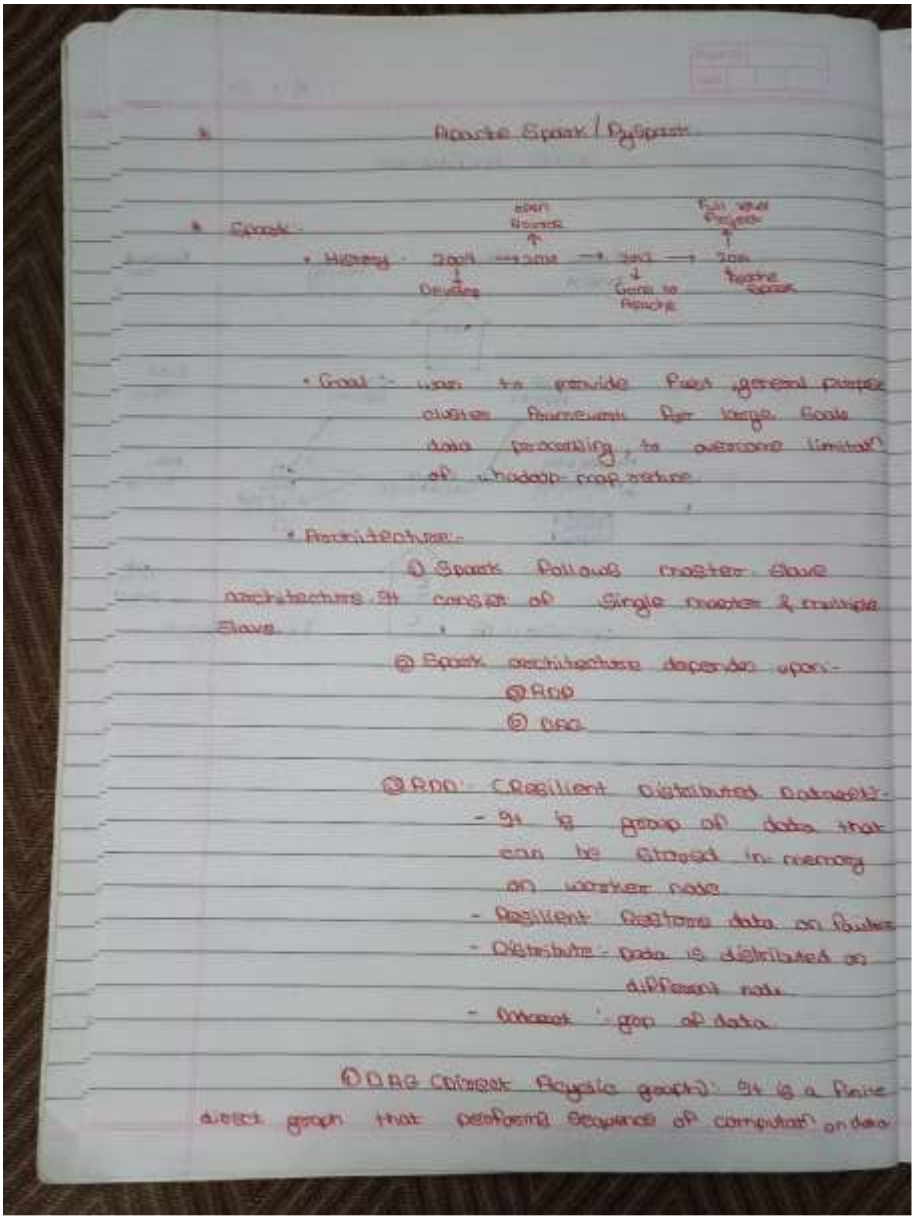
Date:03/02/2024

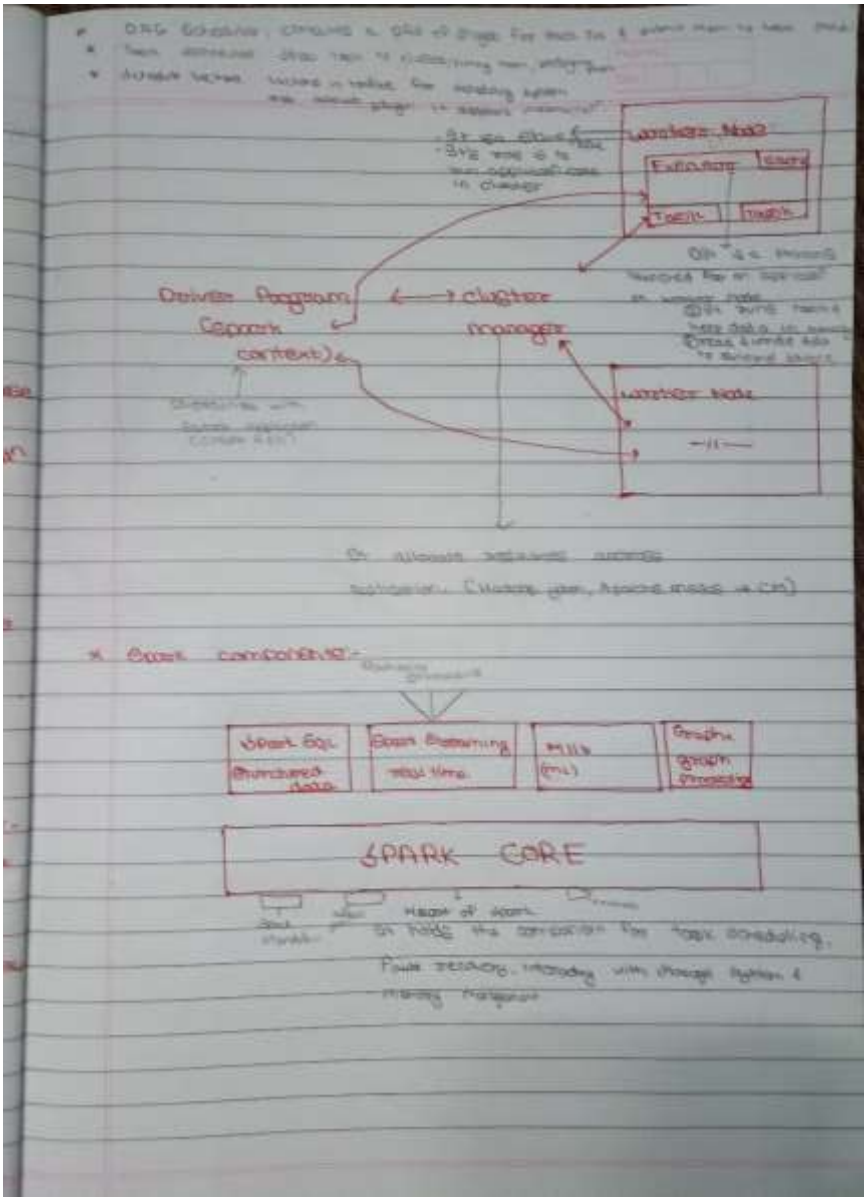
Topic: Pyspark

Solution:

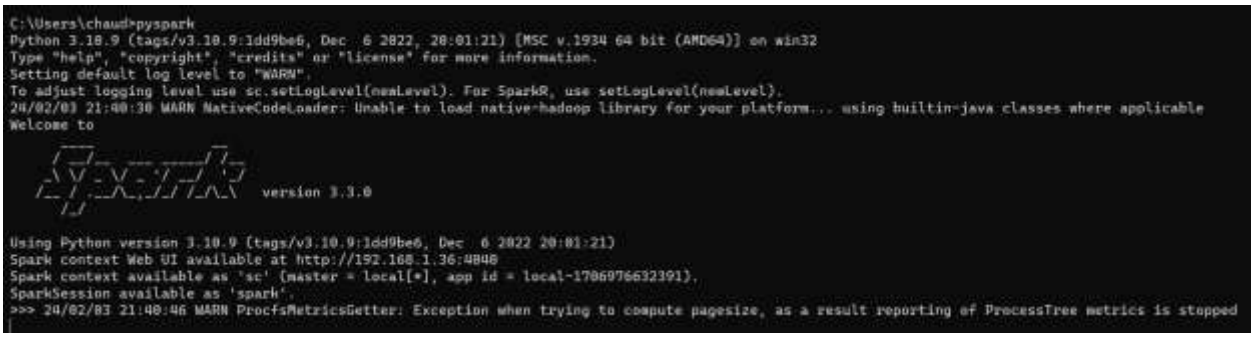
1.Pyspark:

Introduction to pyspark and Basic Architecture:




Worker Node
Executor
cache
Task
Task
Driver Program
SparkContext
Cluster Manager
Worker Node
Spark SQL
Spark Streaming
MLlib
GraphX
SPARK CORE

Pyspark:



Spark shell: