Big Data Analytics SPARK ASSESSMENT

Name: Tshwanelo Mabote

Date: 16 October 2025

Question 1:

a. What are the key features of Apache Spark?

The Key Features of Apache Spark:

General Purpose: It works with multiple programming languages e.g. Python, Java. It is flexible to use in data operating systems.

Fault Tolerance: Apache Sparks uses RDDs to keep track of the usage of data which makes it reliable.

Speed: Spark uses in memory computing and that optimizes the speed of data processing. Fast processing.

Engine: Sparks uses multiple operations like Machine Learning and SQL to process data in batches. Apache Spark engine makes it easier for platform to be used in big data analytics.

b. Explain Spark's Architecture.

Nodes: These are the Physical\ virtual machines where the executers work. Basically, the Server.

Executors: They perform the actual tasks on (Nodes). These are the workers.

Driver program: Contains the spark session. It plans how to execute the job. These are the managers.

Cluster manager: they manage the physical resources e.g. Databricks

Question 2: Refer to Google Collab Notebook

Question 3: Refer to Google Collab Notebook

Question 4: Refer to Google Collab Notebook