Assignment - 2

Big Data (BCA-DS-311)

Course: BCA DS
Date of Issue: 28.04.2025
Course Unit included: 3rd & 4th
Session: 2024-25

Semester: 2nd
Date of Submission: 05.05.2025
Assignment Number: 2nd
Max. Marks: 20

CO-3:- Applying data modeling techniques to large data sets.

CO-4:- Creating applications for Big Data analytics. Building a complete business data analytic solution.

Q.No.	DESCRIPTION	CO	Revised Bloom's Taxonomy Level	Marks
1	Explain the architecture of Apache Hive and Spark by highlighting their components and workflow. How do the Hive Query Language (HQL) and Spark's in-memory processing differ in managing big data, and what are the scenarios where you would prefer Hive over Spark?	CO-3	RBTL-2	4
2	Discuss the Spark Ecosystem components like Spark SQL, Spark Streaming, MLlib, and GraphX, and compare how data manipulation and analysis are carried out in Spark versus Hive. Illustrate with examples where Spark's speed provides significant advantages over traditional Hive operations.	CO-4	RBTL-2	4

3	What are views and indexes in Hive? How do they help in data management?	CO-3	RBTL-2	4
4	Describe a simple real-life application where Spark can be used for faster processing compared to Hive.	CO-4	RBTL- 3	4
5	Write the steps to create a database and a table in Hive with an example for a student management system.	CO-3	RBTL-3	4