Seat No: _____   Enrollment No: _____

PARUL UNIVERSITY
FACULTY OF IT & COMPUTER SCIENCE
Department of Computer Application
MCA / MSc.IT 2024-25 Mid-Term Examination

Semester: 2

Subject Code: (05201295/05202163)

Subject Name: (Big Data Analytics I/Introduction to Big Data Analytics)

Date: (DD/MM/2025)

Time: (1.30hr)

Total Marks: 40

Instructions:

1. Figures to the right indicate full marks.

2. Make suitable assumptions wherever necessary.

Q.1	Answer the following.	[10]
(a)	1. Name any two industries where Big Data is used. 2. Which social media platform is known for utilizing Big Data extensively? 3. Name the distributed processing framework introduced by Apache for handling Big Data.	[3]
(b)	Objective type/MCQs/True-False/Fill in blanks (7 questions of 1 mark each) 1. Which of the following best defines Big Data? a) A large amount of structured data only b) Data that exceeds the processing capacity of traditional databases c) Small datasets with simple structures d) Only transactional data 2. How does Big Data help in preventing fraudulent activities? a) By reducing the size of datasets b) By using predictive analytics and machine learning to detect anomalies c) By limiting the use of online transactions d) By removing all transactions from the system 3. Which of the following technologies is primarily used for Distributed and Parallel Computing in Big Data? a) SQL Server b) Hadoop c) Excel d) NoSQL 4. What is the primary function of In-Memory Computing in Big Data? a) Storing data in traditional hard drives b) Processing large datasets using RAM for faster computation c) Reducing internet usage d) Improving cloud storage capacity 5. What is the core component of the Hadoop Ecosystem responsible for distributed storage? a) MapReduce b) HDFS (Hadoop Distributed File System) c) Apache Spark d) Kafka	[7]

6. Which of the following tools is used for analyzing data with Hadoop?

a) Microsoft Excel
b) Apache Pig
c) MySQL
d) Google Docs

7. IBM's Big Data strategy includes which of the following platforms for analysis and visualization?

a) Infosphere Big Insights and Big Sheets
b) Google Cloud Platform
c) Microsoft Power BI
d) Tableau

Q.2	Answer the following. (2- or 3-mark questions)	[10]
(a)	1. Differentiate between Hadoop Distributed File System (HDFS) and traditional databases: 2. What are the key components of the Hadoop ecosystem?	[4]
(b)	1. Explain the roles of Name-Node and Data-Node in the Hadoop Distributed File System (HDFS). How do they work together to manage data storage and retrieval? 2. Illustrate the role of Big Data analytics in the retail industry with a real-world case study	[6]
Q.3	Attempt any TWO.	[10]
1	Analyze the role of Hadoop in handling large-scale data. How does its ecosystem support distributed storage and processing compared to traditional databases?	[5]
2	To design a data processing solution using Hadoop for a business that handles large volumes of unstructured data. Which components of the Hadoop ecosystem would you choose, and why?	[5]
3	Enlist and explain in details about main characteristics of Big Data.	[5]
Q.4	Answer the following.	[10]
(a)	Write steps to setting up and installing Hadoop in its two operating modes.	[5]

(b)	Write a Python program that implements basic CRUD operations on a MongoDB database using a menu-driven interface
	Create: Add a new student record with the following details:
	    • Roll Number
	    • Name
	    • Date of Birth
	    • Address
	Read: View the details of all students or search for a student by their roll number.
	Update: Update the details (such as name, date of birth, or address) of an existing student by their roll number.
	Delete: Delete a student record using their roll number.
	Exit: Exit the program.
		[5]

	OR	
(b)	What is the main use of Map Reducer in Hadoop ecosystem? Write a python code to understand the overall programming architecture using Map Reduce API.	[5]