Enrolment No./Seat No_____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–VII EXAMINATION – SUMMER 2025

Subject Code:3170722 **Date:14-05-2025**

Subject Name:Big Data Analytics

Time:02:30 PM TO 05:00 PM **Total Marks:70**

Instructions:

1. **Attempt all questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**
4. **Simple and non-programmable scientific calculators are allowed.**

			MARKS
Q.1	**(a)**	What is Big Data? Explain the five V's of Big Data.	**03**
	(b)	Discuss the key differences between structured, unstructured, and semi-structured data with examples.	**04**
	(c)	Explain Hadoop's architecture in detail with a diagram. How does the NameNode and DataNode work together in Hadoop?	**07**
Q.2	**(a)**	What is a MapReduce function? Explain the purpose of the Map and Reduce phases.	**03**
	(b)	Discuss the concept of Data Locality in Hadoop. Why is it significant?	**04**
	(c)	Write a detailed note on HDFS (Hadoop Distributed File System) and explain its significance in Big Data processing.	**07**
		OR	
	(c)	Discuss how MapReduce handles failures during the execution of jobs.	**07**
Q.3	**(a)**	Define NoSQL databases and explain how they differ from relational databases.	**03**
	(b)	Explain the Document store model in NoSQL databases	**04**
	(c)	Compare and contrast different types of NoSQL databases: Key-Value, Document, Column-family, and Graph databases.	**07**
		OR	
Q.3	**(a)**	Describe how NoSQL databases handle scalability and high availability.	**03**
	(b)	What is Sharding in NoSQL databases? How does it help in handling Big Data?	**04**

	(c)	Describe the architecture of MongoDB. Discuss its data model and how it handles queries and indexing.	**07**
Q.4	**(a)**	Explain the role of Apache Spark in Big Data Analytics. How is it different from Hadoop MapReduce?	**03**
	(b)	Discuss Spark's RDD (Resilient Distributed Dataset) and its features.	**04**
	(c)	Write and explain a simple Spark application for word count using PySpark	**07**

OR

Q.4	**(a)**	What is in-memory processing? Why is it beneficial in Big Data analytics?	**03**
	(b)	Explain Spark's transformation and action operations with examples.	**04**
	(c)	How is Spark used to perform Machine Learning tasks? Discuss Spark MLlib with an example of a classification algorithm.	**07**
Q.5	**(a)**	What is Mahout? Explain its use in scalable machine learning with Big Data.	**03**
	(b)	Explain the concept of data streaming. How does Spark Streaming process real-time data?	**04**
	(c)	Explain how data mining techniques are applied in Big Data Analytics. Mention some of the challenges faced when mining large datasets.	**07**

OR

Q.5	**(a)**	Describe the use of Big Data in retail industries. How can companies benefit from Big Data Analytics in decision-making?	**03**
	(b)	What are the security challenges associated with Big Data?	**04**
	(c)	Explain data warehousing in the context of Big Data. How does Hive enable querying large datasets?	**07**

****