Enrolment No./Seat No_____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–VI EXAMINATION – SUMMER 2025

Subject Code:3164604 Date:09-06-2025

Subject Name:Big Data Analytics

Time:10:30 AM TO 01:00 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)**	State the difference between traditional data and big data.	**03**
	(b)	Write a short note on predictive analytics.	**04**
	(c)	Explain the "V's" of Big Data in detail with relevant examples.	**07**
Q.2	**(a)**	Write a Map-Reduce code for Word Count.	**03**
	(b)	Explain the steps to set up the Hadoop cluster.	**04**
	(c)	Draw and explain Map-Reduce framework in detail.	**07**
		OR	
	(c)	Draw and discuss HDFS architecture in detail.	**07**
Q.3	**(a)**	Compare and contrast NoSQL and relational databases.	**03**
	(b)	Write CQL queries for the following in Cassandra: 1. Create a keyspace named company 2. Create a table employee with columns: emp_id (PRIMARY KEY), name, dept, salary.	**04**
	(c)	Discuss the architecture and features of Cassandra. How does it manage data distribution and fault tolerance?	**07**
		OR	
Q.3	**(a)**	Differentiate between master-slave and peer-to-peer distribution models.	**03**
	(b)	Write CQL queries for the following in Cassandra: 1. Create a keyspace named university 2. Create a table students with columns: student_id (PRIMARY KEY), name, course, marks	**04**
	(c)	Describe the four ways in which NoSQL systems handle big data problems. Illustrate your answer with suitable examples.	**07**
Q.4	**(a)**	Differentiate between traditional batch processing and stream processing.	**03**
	(b)	Explain the concept of lazy evaluation in Spark with an example.	**04**
	(c)	Describe Flajolet-Martin algorithm with suitable example.	**07**
		OR	
Q.4	**(a)**	Enlist the challenges in mining data streams.	**03**
	(b)	Explain the Spark execution workflow from job submission to task execution.	**04**
	(c)	Explain the concept of counting ones in a window using DGIM algorithm. Illustrate with a bit stream example.	**07**

Q.5	**(a)**	Compare Apache pig with Map Reduce.	**03**
	(b)	Explain the architecture of ZooKeeper.	**04**
	(c)	Explain working of Hive with necessary steps and diagram.	**07**

OR

Q.5	**(a)**	Explain Metastore in Hive.	**03**
	(b)	Explain the data processing operators in PIG.	**04**
	(c)	Discuss the concepts of regions in HBase and storing Big Data with HBase.	**07**

****