

GUJARAT TECHNOLOGICAL UNIVERSITY**BE - SEMESTER–VI EXAMINATION – SUMMER 2025****Subject Code: 3161607****Date: 28-05-2025****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) What are the key characteristics of Big Data?	03
	(b) What are the challenges of the conventional data management systems when handling Big Data?	04
	(c) Explain the anatomy of a MapReduce job run, including failures and task execution.	07
Q.2	(a) Explain the physical organization of compute nodes in Hadoop.	03
	(b) Explain the process of job scheduling in Hadoop.	04
	(c) Describe the steps involved in setting up a Hadoop cluster with basic commands.	07
	OR	
	(c) Write a Spark program to implement the PageRank algorithm using PySpark or Scala.	07
Q.3	(a) What is NoSQL? List different types of Nosql Databases.	03
	(b) What are the fundamentals of HBase in Big Data processing?	04
	(c) Write a MapReduce program to find the highest and lowest temperatures from a dataset containing daily temperature recordings.	07
	OR	
Q.3	(a) Describe the role of ZooKeeper in Hadoop.	03
	(b) Describe the architecture patterns of NoSQL databases.	04
	(c) Write a Python program using the MongoDB driver (PyMongo) to perform CRUD (Create, Read, Update, Delete) operations on a MongoDB collection storing user profiles.	07
Q.4	(a) What is the difference between master-slave and peer-to-peer NoSQL distribution models?	03
	(b) What is the importance of sampling data in a stream? Provide an example of its application.	04
	(c) Analyze the four ways that NoSQL systems handle Big Data problems.	07
	OR	
Q.4	(a) Mention any two case studies of real-time sentiment analysis using Big Data	03
	(b) How does a shared-nothing architecture help in solving Big Data problems?	04
	(c) Explain the importance of Stream Computing and give examples of its real-time applications, such as stock market predictions.	07
Q.5	(a) What is HiveQL? How does it differ from SQL?	03
	(b) How does PySpark facilitate interactive data processing?	04
	(c) What is Spark's role in Big Data analytics? Discuss its advantages over traditional data processing systems.	07

OR

- Q.5**
- (a)** Define Stream Data Model. **03**
 - (b)** Write a HiveQL query to perform a join operation between two tables: **04**
customers and *orders*, and find all the orders placed by each customer.
 - (c)** Discuss the importance of Graph Analytics Frameworks in Big Data analysis. **07**
Provide examples of real-world applications.
