

Seat No./Enrollment No.: _____

GUJARAT TECHNOLOGICAL UNIVERSITY
BACHELOR OF ENGINEERING – SEMESTER 6 – EXAMINATION – SUMMER 2026

Subject Code: 3164604

Date: 29/05/2026

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.**

- Q.1**
- (a)** Define Big Data and explain any two V's with examples. **03**
 - (b)** Differentiate between traditional data and Big Data approaches. **04**
 - (c)** Describe the transition to Big Data and analyze its impact on business decision making with one Example. **07**

- Q.2**
- (a)** List main components of Hadoop ecosystem. **03**
 - (b)** Explain HDFS architecture with NameNode and DataNode roles. **04**
 - (c)** Illustrate the MapReduce job execution flow and discuss how failures are handled during shuffle and sort phase. **07**

OR

- Q.2**
- (c)** Analyze the role of the JobTracker and TaskTracker in Hadoop's cluster management and discuss how job scheduling and resource allocation are performed during execution? **07**

- Q.3**
- (a)** What is NoSQL? Mention two business drivers. **03**
 - (b)** Explain CAP theorem and its implications for NoSQL design. **04**
 - (c)** Compare key-value, document, and column family stores with examples and select a suitable model for e-commerce analytics- Justify Your choice. **07**

OR

- Q.3**
- (a)** Define stream data and mention two examples of stream sources. **03**
 - (b)** Explain the concept of decaying window in stream processing. **04**
 - (c)** Design a real-time analytics pipeline for sensor data using sampling and filtering techniques. **07**

- Q.4**
- (a)** State any three advantages of Hive over traditional RDBMS. **03**
 - (b)** Describe partitioning and bucketing in Hive with examples. **04**
 - (c)** Develop HiveQL queries to analyze customer transactions and summarize monthly sales using group and aggregate functions. **07**

OR

- Q.4**
- (a)** Write short note on Apache Pig architecture. **03**
 - (b)** Explain difference between Pig Latin and MapReduce approach. **04**

(c) Using Pig and HBase, design a workflow for log analysis to find top 10 URLs accessed per day. **07**

- Q.5**
- (a)** Define RDD in Spark and list its operations. **03**
 - (b)** Discuss advantages of in-memory computing in Spark over Hadoop. **04**
 - (c)** Write a PySpark program to compute word count from text dataset and explain lazy evaluation concept. **07**

OR

- Q.5**
- (a)** List any three Big Data applications in healthcare or finance. **03**
 - (b)** Explain integration of Hadoop and Spark for end-to-end data analysis. **04**
 - (c)** Develop a complete Big Data solution architecture for social-media sentiment analysis including data ingestion, processing, and visualization layers. **07**
