

Enrollment No./Seat No.:

GUJARAT TECHNOLOGICAL UNIVERSITY
Bachelor of Engineering - SEMESTER - VII EXAMINATION - WINTER 2025

Subject Code: 3174807

Date: 24-11-2025

Subject Name: Big Data Engineering and Analytical Methods

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 main challenges faced by conventional data processing systems?	03
(b) Compare structured, semi-structured, and unstructured data with examples.	04
(c) Write a detailed note on different types of Big Data and their applications in real-world scenarios.	07
Q.2 (a) What is a NameNode and DataNode in Hadoop?	03
(b) Explain the architecture of Hadoop and its core components.	04
(c) Describe the working of MapReduce in Hadoop. Explain the entire job execution flow with an example.	07
OR	
(c) Describe Hadoop's use in cloud environments and its benefits for large-scale data processing.	07
Q.3 (a) Write any three differences between SQL and NoSQL databases.	03
(b) Describe the key characteristics of NoSQL databases.	04
(c) Explain in detail the different types of NoSQL databases: key-value, column-family, document, and graph databases, with suitable examples.	07
OR	
(a) List the main types of NoSQL databases.	03
(b) Explain the difference between master-slave and peer-to-peer distribution models.	04
(c) Explain how different NoSQL architectures (document, key-value, column, graph) suit different business use cases.	07
Q.4 (a) What is the use of stream mining in stock market prediction?	03
(b) Explain the concept of counting distinct elements in a stream.	04
(c) Describe the working of a Real-Time Analytics Platform (RTAP) with suitable examples.	07

OR

- (a)** What is a filtering stream? **03**
- (b)** Compare Pig Latin and SQL with suitable examples. **04**
- (c)** Discuss a case study on stock market prediction using data stream mining techniques. **07**

- Q.5**
- (a)** What is Apache Spark? **03**
 - (b)** What is ZooKeeper? Explain its role in maintaining coordination in a Hadoop cluster. **04**
 - (c)** Explain the architecture of Apache Spark in detail with a neat diagram. **07**

OR

- (a)** What is PySpark? **03**
- (b)** Explain the architecture of HBase and the role of HMaster and RegionServer. **04**
- (c)** Explain how to create and manage RDDs, DataFrames, and Datasets in Spark. **07**
