

**GUJARAT TECHNOLOGICAL UNIVERSITY****BE- SEMESTER-VII (NEW) EXAMINATION – WINTER 2024****Subject Code:3170722****Date:27-11-2024****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. List out characteristics of Big Data. **03**  
 (b) Distinguish Traditional Vs Big data business approaches. **04**  
 (c) Define HDFS. Explain HDFS Architecture with diagram. **07**
- Q.2** (a) Explain the historical development of Hadoop and its significance in the field of big data processing. **03**  
 (b) Explain following commands of HDFS with syntax and at least one example of each. (i) copyFromLocal (ii) mv (iii) cat **04**  
 (c) Discuss the key components of Hadoop Architecture. **07**
- OR**
- (c) Walk through the process of developing a MapReduce application and explain how MapReduce works in the context of data processing. **07**
- Q.3** (a) Define NoSQL and explain its significance in modern database management systems. **03**  
 (b) Describe the key-value stores data architecture pattern in NoSQL databases and discuss its characteristics. **04**  
 (c) Explain the significance of Spark in modern data analysis, emphasizing its in-memory computing capabilities and its role in enabling interactive data exploration and application development. **07**
- OR**
- Q.3** (a) Explain features and advantages of NoSQL. **03**  
 (b) Explain NoSQL data architecture patterns family document stores in detail. **04**  
 (c) Discuss the fundamental concepts of Spark, such as RDDs, transformations, and actions, and how they contribute to efficient data processing. **07**
- Q.4** (a) Define the concept of data streams and discuss their significance in the context of big data analytics. **03**  
 (b) Explain the concept of graph analytics in the context of big data and discuss its applications and significance. **04**  
 (c) Explain Stream Data Model and Architecture in detail. **07**
- OR**
- Q.4** (a) Describe techniques for sampling data in a stream and discuss methods for filtering streams based on specific criteria. **03**  
 (b) Explain Decaying Window Algorithm **04**  
 (c) Describe application of real-time analytics platform (RTAP) and discuss case studies such as stock market predictions. **07**
- Q.5** (a) Describe important features of Apache Spark. **03**  
 (b) Differentiate between Pig vs Hive. **04**  
 (c) Explain working of Hive with proper steps and diagram. **07**
- OR**
- Q.5** (a) Explain data processing operators in Pig. **03**  
 (b) Explain basic fundamentals of HBase. **04**  
 (c) Define Zookeeper. List the benefits of it. Explain its working in detail. **07**

\*\*\*\*\*