

(2 ½ Hours.)

[Total Marks: 75]

- N. B.:** (1) All questions are compulsory.  
 (2) Numbers to the right indicate marks.  
 (3) Make suitable assumptions wherever necessary and state the assumptions made.  
 (4) Answers to the same question must be written together.  
 (5) Mixing of Sub-Questions is not allowed.  
 (6) Draw neat labelled diagrams wherever necessary.

**1. Attempt any three of the following:**

**15**

- Explain the traits of Big Data and give examples where each trait plays a critical role.
- What are the major differences between traditional databases and Big Data platforms?
- Explain the concept of Re-sampling in statistical data analysis.
- Discuss the role of prediction error in building accurate models.
- Explain how neural networks support competitive learning.
- Describe how fuzzy logic is applied in extracting models from data.

**2. Attempt any three of the following:**

**15**

- What is a Map task in MapReduce? Explain with a real-life scenario.
- How do Reduce tasks complement Map tasks in a MapReduce job?
- Explain how union, intersection, and difference operations can be implemented in MapReduce.
- What are the fault tolerance mechanisms in MapReduce during node failures?
- Compare MapReduce with parallel database systems.
- Explain common MapReduce algorithms used in big data processing.

**3. Attempt any three of the following:**

**15**

- Define k-shingles. How are they useful in detecting plagiarism?
- Explain how hashing is applied on shingles for document comparison.
- What is the theory behind locality-sensitive functions?
- Describe the steps involved in finding near-neighbors in large datasets.
- How does the choice of shingle size affect similarity detection?
- Explain how collaborative filtering helps in recommendation systems.

**4. Attempt any three of the following:**

**15**

- What are the key components of a data stream management system?
- Explain how filtering is done in data streams with an example.
- How is moment estimation used in summarizing streaming data?

- d. What are the challenges in counting '1's within a sliding window?
- e. Explain the architecture of Real-Time Analytics Platforms (RTAP).
- f. How can we sample effectively from a high-speed data stream?

**5. Attempt any three of the following:**

**15**

- a. Explain how regression modeling supports decision-making in business intelligence.
  - b. Describe how Principal Component Analysis (PCA) helps in reducing dimensionality in big datasets.
  - c. How can MapReduce be used to implement matrix operations like matrix multiplication?
  - d. Explain the role of Locality-Sensitive Hashing (LSH) in identifying near-duplicate documents.
  - e. What are the key features of Real-Time Analytics Platforms (RTAP) and how do they support stream processing?
  - f. Discuss the application of time series analysis in forecasting trends using big data analytics.
-