

- N.B:** 1. Question **ONE** is compulsory.  
 2. Attempt any **THREE** out of remaining.  
 3. **Figures to the right** indicate full marks.  
 4. Assume suitable data if **necessary**.

- |             |   |    |
|-------------|---|----|
| <b>Qu-1</b> | a) What are the characteristics of big Data?  | 5  |
|             | b) Explain Hadoop Architectural Model.  | 5  |
|             | c) List the different issues and challenges in data stream query processing.  | 5  |
|             | d) Explain NoSQL data Architecture patterns.  | 5  |
| <b>Qu-2</b> | a) Explain DGIM algorithm for counting ones in a stream with example.   | 10 |
|             | b) Explain Social Network graph clustering algorithm with example.  | 10 |
| <b>Qu-3</b> | a) Explain Model for Recommendation System in detail  | 10 |
|             | b) Explain Matrix - Matrix Multiplication using TWO step MapReduce model.   | 10 |
| <b>Qu-4</b> | a) Explain PageRank algorithm with suitable example.  | 10 |
|             | b) Explain Bloom's filter for stream data mining with example.  | 10 |
| <b>Qu-5</b> | a) Explain PCY algorithm with suitable example.   | 10 |
|             | b) i) Find Jaccard distance $\{1, 2, 3, 4\}$ & $\{2, 3, 5, 7\}$ and $\{a, a, a, b\}$ & $\{a, a, b, b, c\}$<br>ii) Find Hamming Distance between 110011 & 010101 and 11001 & 01011<br>iii) Compute the cosines of the angles between $(3, -1, 2)$ and $(-2, 3, 1)$ . | 10 |
| <b>Qu-6</b> | Write a note on   | 20 |
|             | a) Hadoop Ecosystem   |    |
|             | b) CURE Algorithm   |    |
|             | c) HITS   |    |
|             | d) MapReduce programming model  |    |

**(3 Hours)**

**[Total Marks 80]**

- i. **Q. 1. is Compulsory.**
- ii. **Attempt any three from the remaining.**
- iii. **Assume suitable data.**

- Q. 1** (a) Describe any five characteristics of Big Data. (5)  
 (b) Describe the structure of HDFS in a Hadoop ecosystem using a diagram. (5)  
 (c) Define Social networks and Social Network Mining. (5)  
 (d) Explain Hamming distance measure with an example. (5)
- Q. 2** (a) Describe characteristics of a NoSQL database. (10)  
 (b) Explain concept of Map Reduce using an example. Write Map Reduce pseudocode for "Group By" "aggregation" in a database. (10)
- Q. 3** (a) Why is finding similar items important in Big Data? Illustrate using two example applications. (10)  
 (b) Explain the concept of a Bloom Filter using an example. (10)
- Q. 4** (a) Explain any one algorithm to count number of distinct elements in a Data stream. (10)  
 (b) Draw the diagram showing the structure of the World Wide Web and explain the different parts. (10)
- Q. 5** (a) What are Recommendation Systems? Clearly explain two applications for Recommendation Systems. (10)  
 (b) Explain in detail any one Ranking algorithm used by Search Engines. (10)
- Q. 6** (a) Explain with diagrams the Park Chen Yu (PCY) algorithm for frequent itemset mining. (10)  
 (b) What is a "Community" in a Social Network Graph? Explain any one algorithm for finding communities in a Social Graph. (10)

- N.B:**
1. Question **ONE** is compulsory.
  2. Attempt any **THREE** out of remaining.
  3. **Figures to the right** indicate **full marks**.
  4. Assume suitable data if necessary.

- Qu-1**
- Distinguish between Soft computing and Hard computing.
  - For a fuzzy set,  $A = \left\{ \frac{0.5}{x_1} + \frac{0.4}{x_2} + \frac{0.7}{x_3} + \frac{0.8}{x_4} + \frac{1.1}{x_5} \right\}$  perform Fuzzy complement operation on A.
  - List different operators in Genetic Algorithm. Explain any one in detail.
  - What is artificial neural network? Define characteristics and applications of artificial neural network.
- Qu-2**
- Explain Character Recognition Application with suitable example. Assume necessary parameters such as input-output pairs, learning rate, error threshold etc. Clearly state the impact of learning rate and error threshold on Character Recognition Application.
  - Explain perceptron learning with the help of an example.
- Qu-3**
- State the Classification of GA and explain in detail the concept "Problem solving using GA".
  - Explain in detail Adaptive Neuro-Fuzzy Inference Systems with suitable architecture.
- Qu-4**
- Explain architecture of BAM. How storage and retrieval is performed in BAM.
  - Give weight matrix of Mc Culloch-Pitts neuron model for binary AND function.
- Qu-5**
- With suitable example explain max-min composition and max-product composition.
  - Explain Backpropagation algorithm in detail with the help of flowchart.
- Qu-6**
- Attempt the following
- Radial Basis Function Networks
  - Fuzzy Composition Rules
  - Delta Learning Rule.
  - Defuzzification

\*\*\*\*\*