Dec - 24

Cont

Time: 3 Hours **Total Marks: 80**

Note: 1. Question 1 is compulsory
2. Answer any three out of the remaining five questions.
3. Assume any suitable data wherever required and justify the same.

Q.1 a) Explain CAP. How is CAP different from ACID property in databases? **[05]**

b) Secondary Name node is a backup of Name node. Is this statement True or False? Justify your answer. **[05]**

c) List and explain the core business drivers behind the NoSQL movement. **[05]**

d) List down any five constraints that must be satisfied for representing a stream by buckets using DGIM algorithm with examples. **[05]**

Q.2 a) List the architectural patterns in NoSQL databases. Discuss the Key-Value and Document-Oriented patterns, focusing on their characteristics, use cases, and examples. **[10]**

b) Write a map reduce pseudo code for word count problem. Apply map reduce working on the following document: **[10]**

"This is NoSQL. NoSQL handles complex data."

Q.3 a) Explain Map Reduce execution pipeline with suitable example. **[10]**

b) Create a Bloom filter with the following parameters: **[10]**
Size of the bit array m=8
Hash functions:

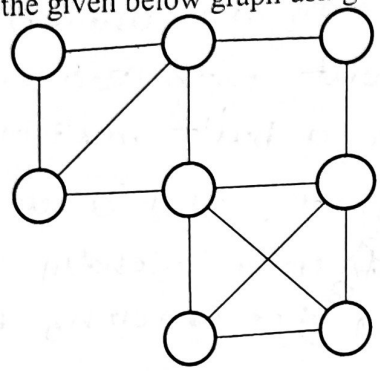
$h1(x)=x \bmod m$
$h2(x)=(2x+1) \bmod m$
$h3(x)=(3x+2) \bmod m$

(i) Insert the following elements into the Bloom filter: 12, 25, 30, 5
(ii) Check if the following elements are present in the Bloom filter: 6, 55
(iii) Discuss the results of your checks, identifying which elements is true positive and which is true negative.

Q.4 a) For the stream of integers: 9, 8, 7, 6, 5, 4, 3, 2. Use the hash function, $h(x)=(2x+1) \bmod 32$ and treat the result as a 5-bit binary integer. Show the steps of the Flajolet-Martin algorithm to estimate the number of distinct elements in this stream. **[10]**

b) Draw a diagram of the typical Hadoop Ecosystem and explain any two components of it. **[10]**

Q.5 a) Write an algorithm for the Clique Percolation Method and discover the communities in the given below graph using Clique Percolation Method with clique k=3. **[10]**



b) i. List and explain the functions provided by R to combine different sets of data. **[10]**

ii. Write the script to sort the values contained in the following vector in ascending order and descending order: (46, 23, 15, 38, 98, 56, 28, 78). Demonstrate the output.

Q.6 a) The project manager at ABC Corp, Mr. Thomas, needs to track information about ongoing projects in the organization. He has the following details about current projects in a table format: **[10]**

ProjectId	ProjectName	Budget
1	Website Redesign	150000
2	Mobile App Launch	100000
3	Data Migration	80000
4	AI Development	200000
5	Cybersecurity Audit	50000

i) Create a Data frame in R for the above project data and display the output.

ii) Show the structure and summary statistics of the Data Frame created.

b) Justify the use of a Content-Based Recommendation System with a specific case study. **[10]**

****