

BE SEM VII CScheme Summer 2025 CMIN
04/06/2025

Time: 03 Hours

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.

- Q1 a) Mention four characteristics of big data and explain in detail. [5]
 b) Explain Shuffle & Sort phase and Reducer phase in Map Reduce. [5]
 c) Demonstrate how business problems have been successfully solved faster, cheaper and more effectively considering NoSQL Google's Bigtable case study. Also illustrate the business drivers and the findings in it. [5]
 d) List down all six constraints that must be satisfied for representing a stream by buckets using DGIM algorithm with examples. [5]
- Q2 a) The project manager at XYZ Ltd., Ms. Meera, is responsible for maintaining details of all active projects. She has organized the project information in the following table: [10]

Project Id	Project Name	Budget	Status
1	CRM Implementation	120000	In Progress
2	Cloud Infrastructure	180000	Completed
3	Network Upgrade	60000	Not Started
4	E-Commerce Platform	220000	Completed
5	Data Analytics	90000	In Progress

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

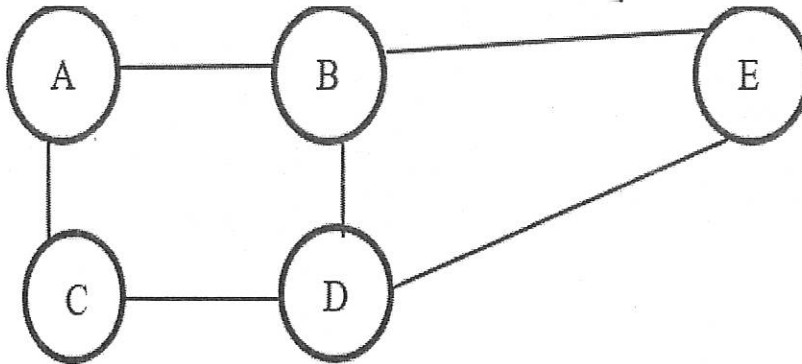
ii) Ms. Meera has recently approved 2 new projects and wants to add their information. The new projects are as follows:

Project Id	Project Name	Budget	Status
6	UX Research	160000	Not Started
7	Cloud Integration	190000	Not Started

Update the Data frame to include the new projects and demonstrate the final output.

- b) Write a short note on variations of NoSQL architectural patterns. [10]

- Q3 a) Suppose the stream is $S = \{10, 12, 8, 15, 6, 9, 14, 7\}$. Let hash functions $h(x) = 5x + 11 \pmod{32}$ for some a and b , treat result as a 5-bit binary integer. Show how the Flajolet- Martin algorithm will estimate the number of distinct elements in this stream. [10]
- b) Explain natural join and grouping and aggregation relational algebraic operation using MapReduce. [10]
- Q4 a) Write a map reduce pseudo code to solve the word count problem. Apply map reduce working on the following document: "Big data is powerful. Big data drives decisions." [10]
- b) With a neat sketch, explain the architecture of the data-stream management system. [10]
- Q5 a) Determine communities for the given social network graph using Girvan- Newman algorithm. [10]



- b) List and discuss various types of data structures in R. [10]
- Q6 a) Describe the components of Hadoop ecosystem with the help of a diagram. [10]
- b) What is recommendation system? How is classification algorithm used in recommendation system? [10]
