06/06/2026

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) Explain Shuffle & Sort phase and Reducer phase in Map Reduce. [5]

b) Agility is a NoSQL business driver. Justify [5]

c) Explain CAP. How is CAP different from ACID property in databases? [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) Describe the different NoSQL architectural patterns? Explain any two types in detail and also identify any two applications that can use these patterns. [10]

b) Write a map reduce pseudo code to multiply two matrices. Apply map reduce working to perform following matrix multiplication. [10]

$$M = \begin{matrix} 2 & 3 & 4 \\ 1 & 0 & 2 \end{matrix} \quad X \quad V = \begin{matrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{matrix}$$

Q3 a) Suppose the stream is S = {4, 2, 5 ,9, 1, 6, 3, 7}. Let hash functions h(x) = 3x + 7 mod 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) List and explain various functions that allow users to handle data in R workspace with appropriate examples. [10]

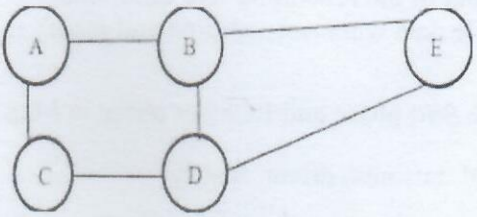
Q4 a) What are the Core Hadoop components? Explain in detail. [10]

b) Explain DGIM algorithm for counting ones in a stream with example. [10]

BE SEM VIII CBSGS (Choice Based) Summer (2026)

CMPN

Q5 a) Determine communities for the given social network graph using Girvan- Newman algorithm. [10]



b) i. Create a Data Frame from the following 4 vectors in R: [10]

stud_id <- c(101:105)

stud_name <- c("Amit", "Neha", "Karan", "Priya", "Rohit")

admission_date <- c("2020-07-15", "2021-02-10", "2020-12-01", "2021-06-25", "2021-09-05")

marks <- c(78, 65, 89, 55, 72)

ii. Display the structure and summary of the above data frame.

iii. Extract only the stud name and marks columns from the data frame.

iv. Extract the student details whose marks are greater than or equal to 70.

Q6 a) i. Write an R script to sort the values contained in the following vector in ascending and descending order. Display both results. [10]

Vector: (85, 42, 67, 29, 93, 58, 74, 36)

ii. List and explain the functions provided by R to combine different sets of data.

b) Describe content-based recommendation system. [10]

****