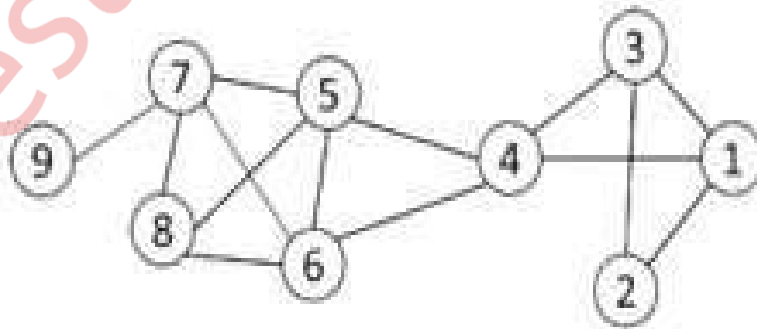15/05/2025 BE IT SEM-VIII C-SCHEME   BDA   QP CODE: 10082370

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 Explain [20]
   a) Big Data and its characteristics
   b) Distance measures for Big Data
   c) Similarity measure can be used in an application to find plagiarism in documents
   d) How failures are handled in MapReduce job

Q2 a) Draw and explain Hadoop ecosystem showing all components [10]
   b) Explain Issues in Data stream query processing [10]

Q3 a) Write a Map reduce pseudo code to multiply two matrices. Illustrate with an example showing all the steps. [10]
   b) Explain Column family store and Graph Store NoSQL architectural pattern with example. [10]

Q4 a) Explain PCY algorithm and its 2 types with neat labeled diagram. [10]
   b) Explain HITS algorithm with example. [10]

Q5 a) Explain Collaborative filtering system. How is it different from content based system. [10]
   b) What is clique percolation method. Write an algorithm on (CPM). [10]
      Also show how the CPM finds clique for the following graph. Explain with steps.

Q6 a) Explain DGIM algorithm. [10]
   b) Explain CURE algorithm. [10]

****

82370