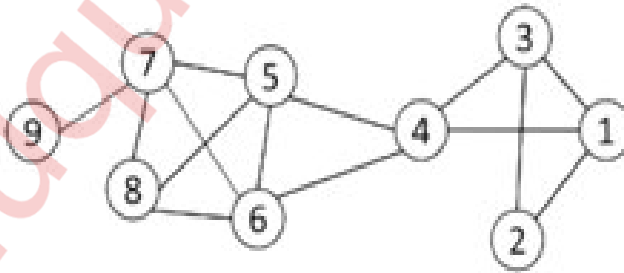4/12/2024 IT SEM-VIII C SCHEME   BDA (DLOC-V) QP CODE: 10065523

Time: 3 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) The Map and Reduce Tasks.
(c) Modified PageRank algorithm.

Q2   a) Draw and explain HDFS architecture showing all components.   **[10]**
b) Explain Column family store and Graph Store NoSQL architectural pattern   **[10]**
with example.

Q3   a) Write a Map reduce pseudo code to multiply two matrices. Illustrate   **[10]**
with an example showing all the steps.
b) Explain Issues in Data stream query processing.   **[10]**

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

Q5   a) Explain Collaborative filtering system. How is it different from content **[10]**
based system.
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 HITS algorithm.   **[10]**
b) Explain CURE algorithm.   **[10]**

_____