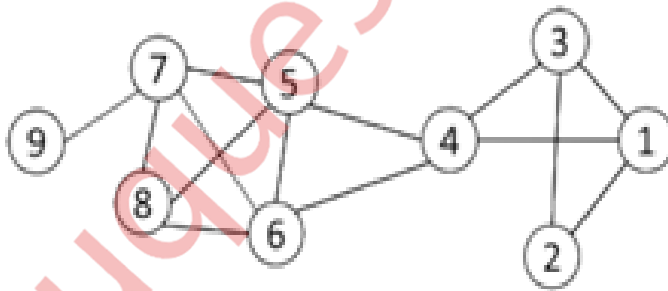**Time: 03 Hours** **Marks: 80**

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 Write short notes on: **[20]**
  a) Big Data and its characteristics
  b) Distance measures for Big Data
  c) The Map and Reduce Tasks
  d) Bloom filter for stream data mining

Q2 a) Explain HDFS architecture. **[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) List the main components of Map reduce execution pipeline. **[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 PageRank algorithm. **[10]**
  b) Explain CURE algorithm. **[10]**

****