

[3 hours]

[80 Mks]

- NB : 1) **Question 1** is compulsory.
 2) Attempt any **three** questions from the **remaining** questions.
 3) **Assume** suitable **data** wherever applicable.

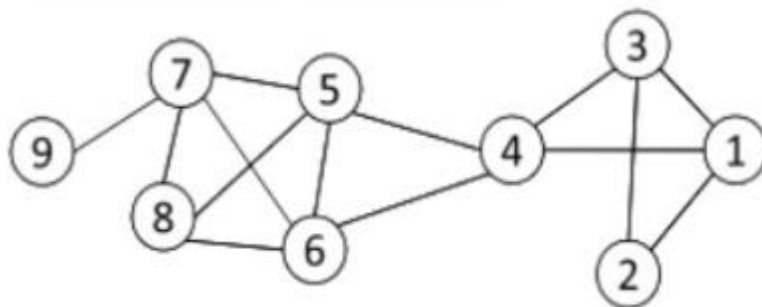
- 1 (a) Write three important characteristics of big data and explain any one with real life example. 5
 (b) Explain Hadoop ecosystem components - Hive and Pig. 5
 (c) Explain MapReduce programming model in detail. 5
 (d) Differentiate between SQL and NoSQL system. 5
- 2 (a) Discuss 1-step Matrix-Matrix Multiplication MapReduce algorithm and apply to the following problem 10

$$\begin{matrix} 2 & 3 & & 3 & 4 \\ 4 & 2 & * & 0 & 2 \end{matrix}$$

 (b) Explain DGIM algorithm for counting ones in a stream with example. 10
- 3 (a) Find the frequent itemset using PCY algorithm with minimum support 3 and $h(i,j)=i*j \text{ mod } 8$

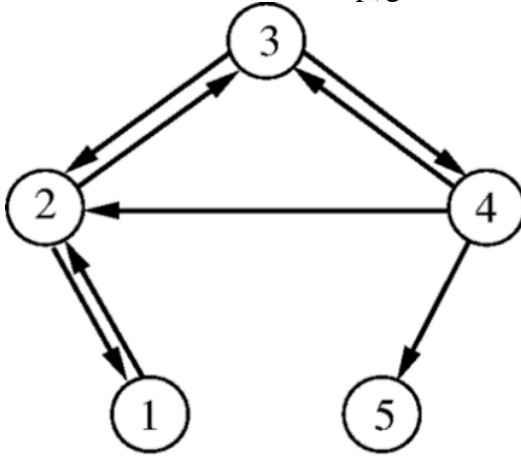
$$\begin{matrix} T1 -- 1,2,3,5 \\ T2 -- 1,2,5 \\ T3 -- 1,3,5 \\ T4 -- 1,2,4 \end{matrix}$$

 (b) Recall all NoSQL design patterns with example. Justify CAP property. 10
- 4 (a) Investigate to find all communities in the graph given below using CPM method. 10



- (b) Comment on usefulness of different types of Recommendation System in real life with example. 10

- 5 (a) Find PAGERANK of each page in the following figure after 3rd iteration. 10



- (b) Illustrate different Relational Algebra operations using MapReduce. 10
- 6 (a) Hadoop advantages and limitations 5
(b) FM algorithm 5
(c) KNN Algorithm 5
(d) CURE algorithm for clustering 5
