

[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 Solve any four out of following 20

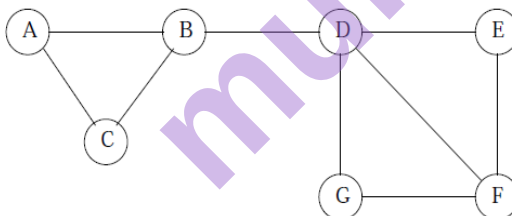
- Explain what characteristics of social media makes it suitable for Big Data.
- Explain the collaborative filtering based recommendation system.
- What are three Vs of Big Data? Give two examples of big data case studies. Indicate which Vs are satisfied by these case studies.
- Explain CAP theorem and explain how NoSQL systems guarantees BASE property.

2 (a) Explain the distributed storage system of Hadoop with the help of a neat diagram.. 10

- (b) Discuss Matrix-Matrix Multiplication. Perform Matrix Multiplication with 1-step Map Reduce method. 10

$$\begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix} * \begin{matrix} 5 & 6 \\ 2 & 3 \end{matrix}$$

3 (a) For the graph given below use Clique percolation and find all communities 10

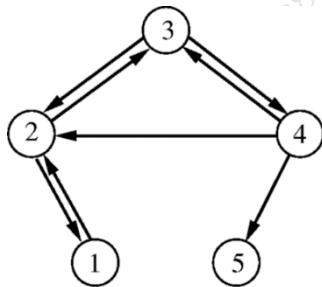


- (b) Give two applications for counting the number of 1's in a long stream of binary values. Using a stream of binary digits, illustrate how DGIM will find the number of 1's. 10

4 (a) Explain Grouping and Aggregation algorithm using MapReduce. Support your answer with a suitable example. 10

- (b) Explain the types of NoSQL data stores and their typical usage. 10

- 5 (a) Explain clearly with diagrams the PCY method of finding frequent itemsets (pairs) in a large data set. 10
- (b) Recall HITS algorithm. Generate Hub and Authority score after 2 iterations for the graph given here. 10



- 6 (a) Suppose a data stream consists of integers 1,3,5,4,6,1,5,9,3,2. Let the hash function used be: 10
- i) $h(x) = x + 1 \bmod 16$
 - ii) $h(x) = 2x + 3 \bmod 16$
 - iii) $h(x) = 3x + 1 \bmod 16$
- Show how the Flajolet-Martin algorithm will estimate the number of distinct elements in the stream.
- (b) Explain CURE algorithm, clearly stating its advantages over traditional clustering algorithm. 10
