

[3 hours]

Marks: 80

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

1 Solve all four. 20

- (a) What is Big Data and give types of big data.
- (b) Elaborate issues of stream processing
- (c) What are the advantages and limitations of Hadoop
- (d) Explain CAP theorem and explain how NoSQL systems guarantees BASE property.

2 (a) Describe the pseudocode for one-step matrix multiplication using mapreduce. 10

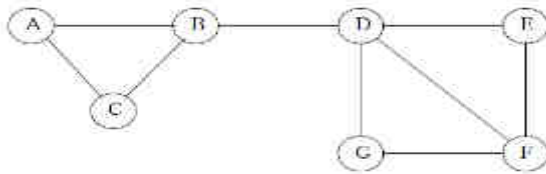
Apply the same to determine the product of matrices M and N:

$$M = \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 4 & 5 & 6 & 7 & 8 \end{matrix} \quad N = \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 4 & 5 & 6 & 7 & 8 \end{matrix}$$

Show output of each stage distinctly.

(b) Show any 5 different relational algebra operations with example. 10

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



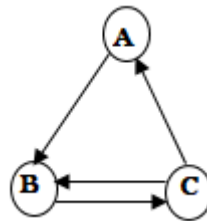
(b) Summarize Bloom's filter with example and its applications. 10

4 (a) Explain collaborative filtering. How it is different from content based filtering? 10

(b) Explain the DGIM algorithm. State the rules used in DGIM that must be followed 10

5 (a) Explain PCY algorithm and its 2 types with neat labeled diagram. 10

(b) Define Hub and Authority. Compute the hub and Authority scores for the web: 10



6 Write short notes on any two: 20

- (i) Cure algorithm.
- (ii) NoSQL data stores with example.
- (iii) Structure of the web

\*\*\*\*\*