Time: [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   Explain the following: 20

  (a)   Differentiate between Big data and Small data.

  (b)   Explain CAP theorem and explain how NoSQL systems guarantee BASE property.

  (c)   Tell Big data application from visualization perspective.

  (d)   State the decisions we must make in order to design a K-Nearest Neighbor
         algorithm. Explain it for 1NN approach.

2 (a)   Summarize Hadoop Architectural Model with both components in detail. List   10
         advantages and limitations of Hadoop.

  (b)   Apply 1 step Matrix - Matrix Multiplication using MapReduce model and solve the   10
         following example

                    1   3   1   3

                    2   4   *   2   4

3 (a)   For the graph given below use Clique percolation method and find all communities   10
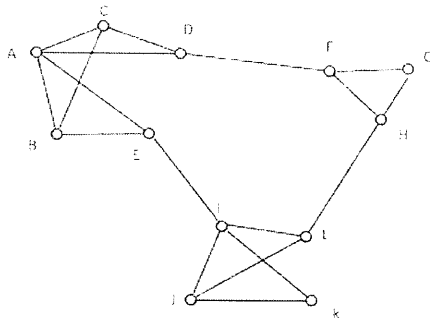
  (b)   Apply HITS algorithm and generate Hub and Authority score after 2 iterations   10
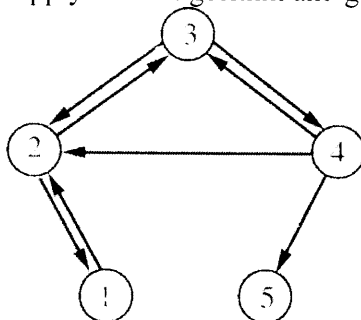
4 (a) What is Recommendation System? Give its different types and an example case study which provides recommendations to users. 10

   (b) Summarize all NoSQL design patterns with example. 10

5 (a) Apply PCY algorithm to find frequent item set with minimum support 40% and hash function h (i, j) = i*j mode 10 10

   T10--- {S, U, N}

   T20— {M, O, N}

   T30--- {T, U, E, S}

   T40— {W, E, D}

   T50— {T, H, U}

   (b) 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 mod 32
   ii)   h(x)= 3x mod 32
   iii)   h(x) = 3x + 2 mod 32

   Show how the Flajolet-Martin algorithm will estimate the number of distinct elements in the stream.

6 (a) Explain Modified PageRank algorithm with example. Discuss problems of page Rank with solution. 10

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

****