Time: 3 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 a) Distinguish between Name node and Data node. **[5]**

b) List and explain the core business drivers behind the NoSQL movement. **[5]**

c) Mention four characteristics of big data. Elaborate these characteristics with respect to social media websites. **[5]**

d) List and explain the different issues and challenges in data stream query processing. **[5]**

Q2 a) What is a key-value store? What are the benefits of using a key-value store? **[10]**

b) Write a map reduce pseudo code to multiply two matrices. Apply map reduce working to perform following matrix multiplication. **[10]**

$$\begin{matrix}1 & 2\\ 3 & 4\end{matrix} \quad X \quad \begin{matrix}6 & 7\\ 8 & 9\end{matrix}$$

Q3 a) Suppose the stream is S = {2, 1, 6, 1, 5, 9, 2, 3, 5}. Let hash functions h(x) = ax + b mod 16 for some a and b, treat result as a 4-bit binary integer. Show how the Flajolet- Martin algorithm will estimate the number of distinct elements, h(x) = 4x + 1 mod 16. **[10]**

b) Consider the following data frame given below: **[10]**

course	id	class	marks
1	11	1	56
2	12	2	75
3	13	1	48
4	14	2	69
5	15	1	84
6	16	2	53

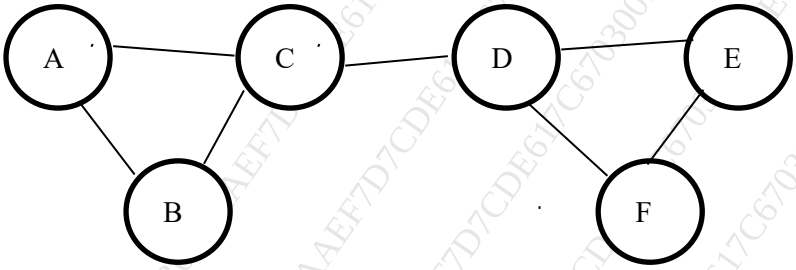
i. Create a subset of course less than 3 by using [ ] brackets and demonstrate the output.

ii. Create a subset where the course column is less than 3 or the class equals to 2 by using subset () function and demonstrate the output.

Q4 a) Explain natural join and grouping and aggregation relational algebraic operation using MapReduce. **[10]**

b) With a neat sketch, explain the architecture of the data-stream management system. **[10]**

Q5 a) Determine communities for the given social network graph using Girvan-Newman algorithm. **[10]**



b) List and discuss various types of data structures in R. **[10]**

Q6 a) i. The following table shows the number of units of different products sold on different days: **[10]**

Product	**Monday**	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
Bread	12	3	5	11	9
Milk	21	27	18	20	15
Cola Cans	10	1	33	6	12
Chocolate bars	6	7	4	13	12
Detergent	5	8	12	20	23

Create five sample numeric vectors from this data.

ii. Name and explain the operators used to form data subsets in R.

b) Define collaborative filtering. Using an example of an e-commerce site like flipkart or amazon describe how it can be used to provide recommendation to users. **[10]**
