

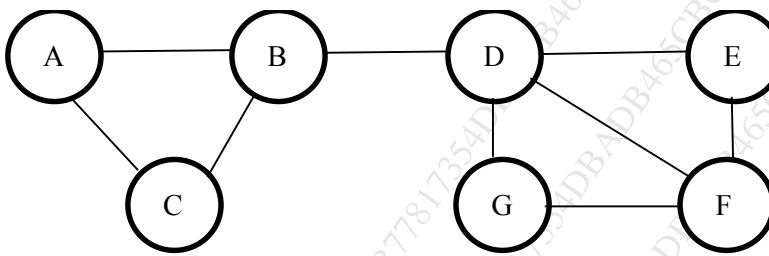
Time: 03 Hours

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) What is function of Map Tasks in the Map Reduce framework? Explain with the help of an example. [5]
- b) Demonstrate how business problems have been successfully solved faster, cheaper and more effectively considering NoSQL Google's MapReduce case study. Also illustrate the business drivers and the findings in it. [5]
- c) Why is HDFS more suited for applications having large datasets and not when there are small files? Elaborate. [5]
- d) Explain the concept of bloom filter with an example [5]
- Q2** a) Name the three ways that resources can be shared between computer systems. Name the architecture used in big data solutions and describe it in detail. [10]
- b) Write a map reduce pseudo code for word count problem. Apply map reduce working on the following document:
- “This is an apple. Apple is red in color”.
- Q3** a) Suppose the stream is 1, 3, 2, 1, 2, 3, 4, 3, 1, 2, 3, 1. Let $h(x) = 6x + 1 \bmod 5$. Show how the Flajolet- Martin algorithm will estimate the number of distinct elements in this stream. [10]
- b) Consider the following data frame given below: [10]
- | subject | class | marks |
|----------------|--------------|--------------|
| 1 | 1 | 56 |
| 2 | 2 | 75 |
| 3 | 1 | 48 |
| 4 | 2 | 69 |
| 5 | 1 | 84 |
| 6 | 2 | 53 |
- i. Create a subset of subject less than 4 by using subset () function and demonstrate the output.
- ii. Create a subset where the subject column is less than 3 and the class equals to 2 by using [] brackets and demonstrate the output.
- Q4** a) What are the Core Hadoop components? Explain in detail. [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) The data analyst of Argon technology Mr. John needs to enter the salaries of 10 employees in R. The salaries of the employees are given in the following table: [10]

Sr. No.	Name of employees	Salaries
1	Vivek	21000
2	Karan	55000
3	James	67000
4	Soham	50000
5	Renu	54000
6	Farah	40000
7	Hetal	30000
8	Mary	70000
9	Ganesh	20000
10	Krish	15000

- i. Which R command will Mr. John use to enter these values demonstrate the output.
- ii. Now Mr. John wants to add the salaries of 5 new employees in the existing table, which command he will use to join datasets with new values in R. Demonstrate the output.

- Q6 a) i. Write the script to sort the values contained in the following vector in ascending order and descending order: (23, 45, 10, 34, 89, 20, 67, 99). Demonstrate the output.
 ii. Name and explain the operators used to form data subsets in R.
- b) How recommendation is done based on properties of product? Elaborate with a suitable example.
