

BE SEM VIII (choice based) Summer 2020 INFT

[Time: 3 hours]

[Marks: 80]

1415126

NB: 1) Question 1 is compulsory.

2) Attempt any three questions from the remaining questions.

3) Assume suitable data wherever applicable

- 1 (a) What is Big Data? Explain the role of the Three V's in defining Big Data with examples. 5
 (b) Explain the Core Components of Hadoop with the help of neat diagram. 5
 (c) Explain the concept of Map-Reduce with an example. 5
 (d) Explain what characteristics of social media makes it suitable for big data. 5
- 2 (a) Discuss Matrix-Multiplication MapReduce algorithm and apply to the following problem. 10

$$\begin{array}{cc} 1 & 2 & & 5 & 6 \\ 3 & 4 & * & 7 & 8 \end{array}$$

 (b) Explain architecture of Data Stream Management System with neat diagram. 10
- 3 (a) Discuss Bloom filter process with neat diagram and also explain how bloom filter is useful for big data analytics with an example. 10
 (b) Find the frequent itemset using PCY-algorithm with minimum support 3 and $h(i,j)=i*j \bmod 8$ 10

$$\begin{array}{l} T1 \text{ -- } 1,2,3,5 \\ T2 \text{ -- } 1,2,5 \\ T3 \text{ -- } 1,3,5 \\ T4 \text{ -- } 1,2,4 \end{array}$$
- 4 (a) Discuss Flajolet-Martin algorithm and also show how the Flajolet-Martin algorithm will estimate the number of distinct elements in the stream $S=\{4,2,5,9,1,6,3,7\}$ and hash function is $h(x) = 3x+7 \bmod 32$ 10
 (b) Explain CURE algorithm, clearly stating its advantages over traditional clustering algorithm. 10
- 5 (a) Discuss in detail Structure of Web and also explain Dead ends 10
 (b) Discuss the usefulness of various types of recommendation systems in real-world scenarios with example. 10
- 6 (a) Explain the working of all NOSQL architectural patterns. Identify the applications that can use this pattern. 10
 (b) Explain DGIM algorithm for counting ones in a stream with example. 10
