Duration: 3hrs **[Max Marks:80]**

N.B. : (1) Question No 1 is Compulsory.
(2) Attempt any three questions out of the remaining five.
(3) All questions carry equal marks.
(4) Assume suitable data, if required and state it clearly.

1 Attempt any FOUR **[20]**

a What are the characteristics of big Data?

b Explain Hadoop Architecture Model.

c Explain NoSQL data Architecture patterns.

d Explain Matrix Vector Multiplication algorithm by MapReduce

e How recommendation is done based on properties of the product?Explain with suitable example

2 **a** What is MapReduce ? Explain How Map and Reduce Work? What is Shuffling in MapReduce? **[10]**

b Why is finding similarity important in BigData? Illustrate using two example application **[10]**

3 **a** Explain Page Rank with Example. Can a Website's Page rank Ever Increase? What are its chances of Decreasing? **[10]**

b For given database D; Minimum support =2 use PCY algorithm to get frequent itemset **[10]**

TID	Items
1	1,2,3
2	2,3,5
3	1,2,3,5
4	2,5

4 **a** Explain the Data Stream Management system with neat diagram **[10]**

b Given a Dim Dataset (1,5,8,10,2} Use the agglomerative clustering algorithm with Euclidean distance to establish hierarchical grouping relationship. Draw the dendrogram. **[10]**

5 a Calculate the Cosine distance measure for given vectors **[10]**

d_1 = **3 2 0 5 0 0 0 2 0 0**

d_2 = **1 0 0 0 0 0 0 1 0 2**

b Explain the concept of Bloom's filter with the helper examples **[10]**

6 a Explain short note on **[20]**

a) Zookeeper

b) CAP theorem

c) Clustering Algorithms

d) Market basket model
