

(3 Hours)

[Total Marks: 80]

Instructions:

1. Question No.1 is compulsory
2. Attempt any Three from the remaining
3. Figures to the right indicate full marks
4. Assume suitable data if necessary.

- Q.1. (a) Compare Big Data Analysis with Traditional Data Mining and warehousing system (5)
 (b) Explain “Shuffle & Sort” phase and “Reducer phase” in Map Reduce. (5)
 (c) Find Jaccard Distance and Cosine Distance between the following pairs of set (5)
 $X=(0,1,2,4,5,3)$ & $Y=(5,6,7,9,10,8)$
 (d) Define Social Networks and Social Network Mining (5)
- Q.2 (a) Explain NoSQL Business drivers and also describe various architecture Patterns (10)
 of NoSQL.
 (b) What is the MapReduce? Explain the role of Combiner with the help of (10)
 an example.
- Q.3. (a) Explain Page Rank algorithm with suitable example (10)
 (b) Elaborate Collaborative Filtering System. How is the system different from a (10)
 content based system.
- Q.4. (a) Explain Park-Chen-Yu algorithm with suitable example. (10)
 (b) Explain the Physical Architecture of Hadoop. State its Limitations. (10)
- Q.5. (a) Describe the Characteristics of Big Data with suitable example. State the types (10)
 of Big data
 (b) Distinguish the following (10)
 (i) Document store & Column family data store.
 (ii) RDBMS & NoSQL database.
- Q.6. Write Short Note on. (any two) (20)
 (a) Hadoop Ecosystem
 (b) Data Stream Management System
 (c) Matrix Multiplication by MapReduce
 (d) Network Traffic Analysis.