

(3 Hours)

(Total Marks : 80)

Please check whether you have got the right question paper.

- N.B.:**
- 1) **Question No. 1 is compulsory.**
 - 2) Attempt **any THREE** from remaining.
 - 3) Assume suitable **data** if **necessary**.
1.
 - a) Explain the role of combiner in Hadoop with example. (05)
 - b) Explain CAP theorem. How is CAP different from ACID property in databases? (05)
 - c) Explain Cosine distance measure with an example. (05)
 - d) Justify why analysis of big data is important for organizations. (05)
 2.
 - a) Define Sampling. Explain various sampling techniques for Big Data. (10)
 - b) What is Clustering? Explain CURE algorithm for Clustering Big Data. (10)
 3.
 - a) What are the business drivers for NoSQL? Explain in detail any two NoSQL data store with suitable example. (10)
 - b) Explain the PageRank Algorithm with example. (10)
 4.
 - a) Explain Matrix Vector Multiplication using Map Reduce. (10)
 - b) Explain Park Chem Yu (PCY) algorithm and its disadvantages. (10)
 5.
 - a) Explain DGIM Algorithm. Using a stream 10110111010110011100101, and $k = 15$, illustrate how the DGIM algorithm will find the number of 1's. (10)
 - b) Explain the collaborative filtering-based recommendation system in detail. (10)
 6. Write short note on **(Any Two)** : (20)
 - a) Data Stream Management System and its Applications.
 - b) Clique Percolation Method (CPM) for communities detection.
 - c) Hadoop Ecosystem.