

(3 Hours)

[Total Marks : 80]

- Note : (1) Question No.1 is compulsory.
(2) Attempt any three from the remaining questions.
(3) Assume suitable data.

1. (a) Define the 3 V s of Big Data 05
(b) Find Manhattan distance (L1-norm) and Euclidean distance (L2-norm) for the following points $X_1 = (1, 2, 2)$, $X_2 = (2, 5, 3)$ 05
(c) Explain how are dead ends handled in Page Rank. 05
(d) Give problem in Flajolet-Martin (FM) algorithm to count distinct elements in a stream. 05
2. (a) Explain different NoSQL data architecture patterns. 10
(b) Give Hadoop Ecosystem and briefly explain its components. 10
3. (a) Give 2-step Map Reduce algorithm to multiply two large matrices. 10
(b) i. Give Map Reduce algorithm for Natural Join of two relations. 10
ii. Give Map Reduce algorithm to perform Intersection of two sets.
4. (a) Explain abstract architecture of Data Stream Management system (DSMS). 10
(b) Explain how to compute Page Rank for any web graph. 10
5. (a) Explain Park Chen Yu algorithm for counting frequent item sets. 10
(b) Explain CURE algorithm for large scale clustering. 10
6. (a) Explain with example collaborative and content based filtering in a recommendation system. 10
(b) Write a note on Social Network Graphs. 10