

30/12/2023

Duration 3 Hours

Maximum marks 80

- 1) Question 1 is compulsory
- 2) Attempt any three from the remaining questions
- 3) All questions carry equal marks.
- 4) Assume suitable data if necessary.

- Q.1. Attempt any **four** from the following questions
- |   |   |   |
|---|---|---|
| a | Describe any Five characteristics of Big Data Analytics.                      | 5 |
| b | Explain cosine distance measure with an example.                              | 5 |
| c | Explain CAP theorem. How is CAP is different from ACID property in Databases? | 5 |
| d | Explain Hubs and Authorities with neat diagram.                               | 5 |
| e | How Big Data Analytics can be useful in development of Digital India?         | 5 |
- Q.2.a Illustrate with an example the Clique percolation Method (CPM) used in direct discovery of communities in a social graph. 10
- b Illustrate with an example Collaborative filtering system. How it is different from content-based filtering system. 10
- Q.3.a Explain Hadoop Ecosystem with core components? Explain the physical architecture of Hadoop. State its limitations. 10
- b Explain how Hadoop goals are covered in Hadoop Distributed file system. 10
- Q.4.a Explain any one algorithm to count the number of distinct elements in a data Stream. 10
- b Differentiate between RDBMS and a NO-SQL Database. 10
- Q.5 a Explain the concept of Bloom filter using an Example. 10
- b Write the pseudocode for Matrix Vector Multiplication by Map Reduce. Illustrate with an example showing all the steps. 10
- Q.6 a Explain with diagrams Park Chen Yu Algorithm for frequent itemset mining with an example. 10
- b Explain in details any one ranking Algorithm used by search engines with an example. 10

\*\*\*\*\*