

(3 Hours)

Max. Marks: 80

- N.B.:** (1) Question No. 1 is **compulsory**.
 (2) Attempt any **Three** questions out of remaining **Five** questions.
 (3) **Figures** to the **right** indicate **full** marks.
 (4) Assume suitable data if **necessary**.

Q.1 Answer the following:

[20]

- Explain Different Distance measures for Big Data.
- Explain the Hadoop Architecture with its features.
- Explain CAP Theorem? how it is different from ACID Properties.
- What are the shortcomings of nearest neighbor technique in collaborative filtering method? Suggest some improvements.

Qu-2 a. Write a Map-Reduce Algorithm for Binary search tree. Explain the flow of execution. [10]

Qu-2 b. Suppose a stream consists of the integers 2,1,6,1,5,9,2,3,5. Let the hash functions all be of the form $h(x)=ax+b \text{ mod } 16$ for some a & b . You should treat the result as a 4 bit binary integer. Determine the tail length for each stream element and the resulting estimate of the number of distinct elements if the hash function is :

[10]

- $h(x) = 2x + 3 \text{ mod } 16$
- $h(x) = 4x + 1 \text{ mod } 16$
- $5x \text{ mod } 16$

Qu-3 a. Explain Different types of recommendation system with real time examples. [10]

Qu-3 b. Consider the portion of a Web graph as shown in Figure-1 [10]

[10]

- Compute the hub and authorities scores for all nodes
- Does this graph contain spider traps? Dead ends? If so, which nodes

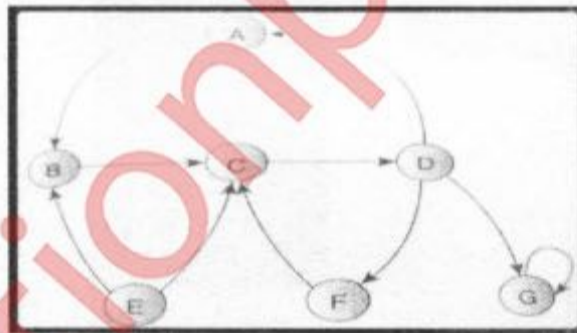


Figure-1 Web graph

Qu-4 a. Write a short note on [10]

- PCY Algorithm
- CURE algorithm

Qu-4 b. Imagine there are 100 baskets, numbered 1, 2, ..., 100 items, similarly numbered. Item I is in basket J if and only if I divides J evenly. For example basket is 24 is the set of items { 1,2,3,4,6,8,12,24}. Describe all the association rules that have 100% confidence. [10]

Qu-5 a. Define Bloom Filter. Explain the concept of Bloom filter Algorithm with example. [10]

Qu-5 b. Explain HITS algorithm with example. [10]

Qu-6 Answer any two of the following [20]

- NoSQL architectural pattern with example.
- Matrix Multiplication by Map Reduce
- List & explain Big data :- 1) Characteristics 2) Types 3) Challenges

----X----