Q.P. Code: 25685

(3 Hours) **[Total Marks 80]**

i. **Q. 1. is Compulsory.**
ii. **Attempt any three from the remaining.**
iii. **Assume suitable data.**

Q. 1	**(a)**	Describe any five characteristics of Big Data.	**(5)**
	(b)	Describe the structure of HDFS in a Hadoop ecosystem using a diagram.	**(5)**
	(c)	Define Social networks and Social Network Mining	**(5)**
	(d)	Explain Hamming distance measure with an example.	**(5)**
Q. 2	**(a)**	Describe characteristics of a NoSQL database.	**(10)**
	(b)	Explain concept of Map Reduce using an example. Write Map Reduce pseudocode for "Group By" "aggregation" in a database.	**(10)**
Q. 3	**(a)**	Why is finding similar items important in Big Data? Illustrate using two example applications.	**(10)**
	(b)	Explain the concept of a Bloom Filter using an example.	**(10)**
Q. 4	**(a)**	Explain any one algorithm to count number of distinct elements in a Data stream.	**(10)**
	(b)	Draw the diagram showing the structure of the World Wide Web and explain the different parts.	**(10)**
Q. 5	**(a)**	What are Recommendation Systems? Clearly explain two applications for Recommendation Systems.	**(10)**
	(b)	Explain in detail any one Ranking algorithm used by Search Engines.	**(10)**
Q. 6	**(a)**	Explain with diagrams the Park Chen Yu (PCY) algorithm for frequent itemset mining.	**(10)**
	(b)	What is a "Community" in a Social Network Graph? Explain any one algorithm for finding communities in a Social Graph.	**(10)**