

K. J. Somaiya Institute of Engineering and Information Technology, Sion, Mumbai-22

(Autonomous College Affiliated to University of Mumbai)

End Semester Exam

April - May 2022

(B. Tech) Program: IT

Examination: TY Semester: VI

Course Code: **IUITDLC605I** & Course Name: Big Data Analytics

Duration: 03 Hours

Max. Marks: 60

Instructions:

- (1) All questions are compulsory.
- (2) Draw neat diagrams wherever applicable.
- (3) Assume suitable data, if necessary.

		Max. Marks	CO	BT Level
Qu-1	Solve any Six questions out of Eight	12		
i)	What is Big Data? Give some examples of Big Data.	2	CO1	Remember
ii)	What is NoSQL? What are the NoSQL Business Drivers?	2	CO2	Remember
iii)	What is HDFS and what are its features?	2	CO2	Understand
iv)	What are the issues and challenges faced in data stream processing?	2	CO5	Remember
v)	Find edit distance between the following pair of string: "abcceghi" and "defghsi"	2	CO4	Apply
vi)	Compute the Hamming distance between the following: x= 111111100 and y= 000111111	2	CO4	Apply
vii)	What is a role and effect of PageRank?	2	CO6	Remember
viii)	Make use of real life examples to explain the concept of shuffling in Map reduce.	2	CO3	Apply
Qu-2	Solve any Four questions out of Six .	16		
i)	List different types of data and explain Structured, Semi Structured and Unstructured data by giving example.	4	CO1	Understand
ii)	Explain NoSQL data architecture patterns.	4	CO2	Understand
iii)	Apply the concept of Bloom filter in an example.	4	CO5	Apply

iv)	Apply Map Reduce Algorithm for processing relational data.	4	CO3	Apply
v)	Identify applications of collaborative filtering	4	CO4	Apply
vi)	Compare user-based collaborative filtering and item -based filtering?	4	CO6	Analyze
Qu-3	Solve any Two questions out of Three .	16		
i)	What are 5 V's of Big Data? Explain two examples of big data case studies and indicate which characteristics are satisfied by these cases.	8	CO1	Understand
ii)	Explain collaborative filtering system .Analyze how it is different from content based system	8	CO6	Analyze
iii)	Implement the concept of Map Reduce in Matrix Vector Multiplication .	8	CO3	Apply
Qu-4	Solve any Two questions out of Three .	16		
i)	Suppose data stream consist of integers 1,3,2,1,2,3,4,3,1,2,3,1. Let hash function being used is $h(x) = (6x + 1) \bmod 5$ Show how Flajolet Martin will estimate number of distinct elements in this stream.	8	CO5	Apply
ii)	List the use cases of key-value store, Graph store and document store and explain any one-use case of key-value store in detail.	8	CO2	Understand
iii)	Using suitable example explain PCY algorithm .	8	CO4	Apply