

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: LY Semester: VIII

Course Code: IUITC801 & 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.

Question No.	Question	Max. Marks	CO	BT Level
Q 1)	Solve any Six questions out of Eight	12		
i)	List the advantages and applications of Big Data Analytics.	2	CO1	Remember
ii)	What is HDFS and what are its features?	2	CO2	Remember
iii)	Explain structured, semi-structured and unstructured data with the help of example .	2	CO2	Understand
iv)	List issues and challenges faced in data stream processing .	2	CO5	Remember
v)	Find Hamming distance between 110011 and 010101, 110001 and 01011 applying the concept of distance measure in Big Data Analytics.	2	CO4	Apply
vi)	Identify the use of market basket model in real life application.	2	CO4	Apply
vii)	Make use of real life example to explain concept of shuffling in Map reduce.	2	CO3	Apply
viii)	Define collaborative filtering.	2	CO6	Remember
Q 2)	Solve any Four questions out of Six .	16		
i)	Compare Traditional and Big Data Approach.	4	CO1	Understand
ii)	List and explain the application of Apache Spark.	4	CO2	Understand
iii)	Apply the updating bucket approach of DGIM algorithm, illustrate it with the help of example.	4	CO4	Apply
iv)	What are combiners? Identify the situation when one use a combiner in Map Reduce job?	4	CO3	Apply
v)	Apply Jaccard distance and cosine distance method to find the distance between the following pairs of set: X= (0,1,2,4,5,3,) and Y= (5,6,7,9,10,8)	4	CO5	Apply

vi)	Define Hub and Authority. Find and analyze the Hub and Authority Score for the following Web.	4	CO6	Analyze
	<pre> graph TD A[A] <--> B[B] A[A] <--> C[C] B[B] <--> C[C] </pre>			
Q 3)	Solve any Two questions out of Three .	16		
i)	Explain the Characteristics, Types and Challenges with respect to Big Data Analytics.	8	CO1	Understand
ii)	List the different recommender system. Explain any one with example in detail	8	CO6	Analyze
iii)	Implement the concept of Map Reduce in Matrix Vector Multiplication .	8	CO3	Apply
Q 4)	Solve any Two questions out of Three .	16		
i)	Suppose data stream consist of integers 4, 8, 5,7,3,6,2,5,1. Let hash function being used is $h(x)=3x+2 \pmod 5$. Show how Flajolet Martin will estimate number of distinct elements in this stream.	8	CO5	Apply
ii)	Explain Hadoop Ecosystem with core components. Explain its physical Architecture. State the limitations of Hadoop.	8	CO2	Understand
iii)	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}. Apply the concept of association mining and describe association rules that have 100% confidence.	8	CO4	Apply