

K. J. Somaiya Institute of Engineering and Information Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Subject Code: CSC702

Subject Name: Big Data Analytics

Date:

Date : 06-12-2022

Nov - Dec 2022 (B.Tech) Program: Computer Engineering Examination: LY Semester: VII Course Code: CEC702 Course Name: Big Data Analytics Duration: 2.5 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
Q 1	Solve any six questions out of eight:	12		
i)	What is the difference between structured and unstructured data?	02	CO1	U
ii)	Describe characteristics of a NoSQL databases.	02	CO2	U
iii)	Discuss any four HDFS commands in brief.	02	CO1	U
iv)	Explain Map phase and Reduce phase in MapReduce Framework.	02	CO3	U
v)	List varieties of social network.	02	CO5	U
vi)	What do you mean by data stream?	02	CO4	U
vii)	List basic features of R.	02	CO6	U
viii)	Illustrate Shuffling in MapReduce.	02	CO3	U
Q.2	Solve any four questions out of six.	16		
i)	Explain with example content based filtering.	04	CO5	U
ii)	What are combiners? When should one use combiner in map reduce job? Explain with an example.	04	CO3	U
iii)	Explain any one algorithm to ceunt number of distinct elements in a Data stream.	04	CO4	U
iv)	Explain Hadoop Ecosystem and briefly explain its components.	04	CO1	U
v)	Utilize CAP theorem in Amazon e-commerce application and show its usage.	04	CO2	U
vi)	Write an R script to import "exam.csv" file to a dataframe and perform any three operations on it.	04	CO6	Ap
Q.3	Solve any two questions out of three.	16		

K. J. Somaiya Institute of Engineering and Information Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Subject Code: CSC702

Subject Name: Big Data Analytics

Date: 06-12-2022

i)	With suitable block diagram explain architecture of HDFS. Discuss role of Data node and Name node in HDFS.	08	CO1	U
ii)	Explain concept of Map Reduce using an example. Write Map Reduce pseudocode for "Group By" "aggregation" in a database.	08	CO3	Ap
iii)	Explain different types of charts in R for data visualization in detail.	08	CO6	U
Q.4	Solve any two questions out of three.	16		
i)	Explain the concept of Bloom's Filter in brief. Analyze the algorithm for 8 billion bits, 1 billion members of the set S, and 1 hash function.	08	CO4	An
ii)	What is NoSQL database? List the differences between NoSQL and relational databases. Explain in brief various types of NoSQL databases in practice.	08	CO2	U
iii)	Draw the diagram showing the structure of the World Wide Web and explain the different parts.	08	CO5	U
