

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

April – May 2023 B. Tech. Program: Computer Engineering Scheme II Examination: LY Semester: VII Course Code:CEC702 Course Name: Big Data Analytics		
Date of Exam: <u>27-05-2023</u>	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 are the characteristics of stream data?	02	CO1	1
ii)	What is Sampling? Explain various sampling techniques for big data.	02	CO4	1
iii)	Explain two applications of Social Network Mining	02	CO5	2
iv)	Define and explain BASE theorem in detail	02	CO3	1
v)	What are the functions of Name Node in Hadoop?	02	CO2	1
vi)	What are the basic features of R?	02	CO6	1
vii)	How Job failure is handled in Hadoop?	02	CO2	2
viii)	Justify Why Big Data is Important for Organizations?	02	CO1	2
Q.2	Solve any four questions out of six.	16		
i)	Explain the characteristics of Hadoop.	04	CO1	1
ii)	Design Map Reduce pseudo code for any two of the following i) Computing selection ii) Projection iii) Intersection	04	CO2	3
iii)	What are the Four ways that NoSQL systems handle big data?	04	CO3	2
iv)	Explain any community detection algorithm for social media networks.	04	CO5	2

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

April – May 2023 B. Tech. Program: Computer Engineering Scheme II Examination: LY Semester: VII Course Code: CEC702 Course Name: Big Data Analytics		
Date of Exam: 27-05-23	Duration: 2.5 Hours	Max. Marks: 60

v)	What are various built in functions in R?	04	CO6	1
vi)	Suppose stream consists of integers 2,1,5,1,5,9,2,3,5,6. Let hash function be $h(x) = 3x \text{ mod } 5$. Determine distinct elements in stream using Flajolet-Martin algorithm.	04	CO4	3
Q.3	Solve any two questions out of three.	16		
i)	Explain the role of each Hadoop ecosystem component. What are the limitations of Hadoop?	08	CO1	2
ii)	Explain DGIM algorithm and solve following problem: Data: 101011000101110110010110110 i) Show initial configuration $N=20$ what is estimate of number of 1's. ii) The following bits enter the window one at a time: 10100111. What is the bucket configuration? Estimate no of 1's in latest $K=14$ bits of window?	08	CO4	3
iii)	Explain the content based recommendation system.	08	CO5	1
Q.4	Solve any two questions out of three.	16		
i)	Explain the steps of Map Reduce execution pipeline with an example	08	CO2	3
ii)	For the following Applications which type of NoSQL data stores would be required explain why? Explain its Architecture? <ul style="list-style-type: none"> ▶ Event Logging ▶ Content management Systems ▶ Blogging platforms 	08	CO3	2
iii)	Explain various data visualization techniques using R.	08	CO6	1
