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

~~Nov - Dec 2024~~ Jan / Feb 25
(B. Tech.) Program: Computer Engineering Scheme: II
Regular Examination: LY Semester: VII
05-02-25 Course Code: CEC702 and Course Name: Big Data Analytics
Date of Exam: ~~28/11/2024~~ Duration: 02.5 Hours Max. Marks: 60

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

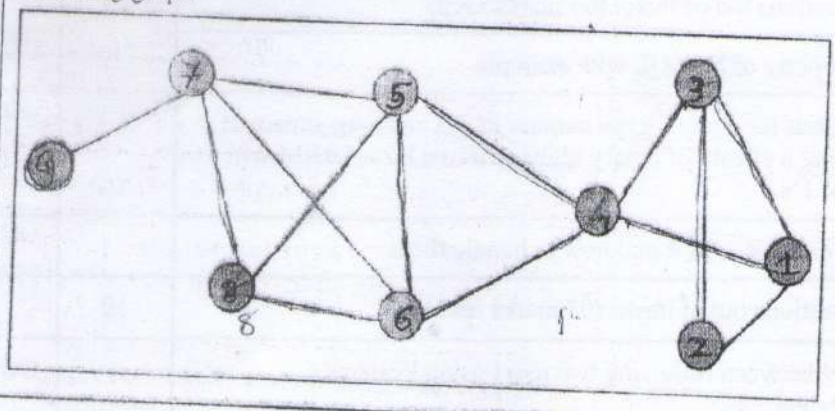
Q. No.	Question	Max. Marks	CO	BT level
Q 1	Solve any two questions out of three: (05 marks each)	10		
a)	Explain BASE property of NOSQL with example.		CO2	U
b)	Give two applications for counting the number of 1's in a long stream of binary values. Using a stream of binary digits, Discuss how, DGIM will find the number of 1's.		CO4	U
c)	Explain different data types in R and how to handle them		CO6	U
Q 2	Solve any two questions out of three: (05 marks each)	10		
a)	Find the similarity between following two users using Pearson's Correlation coefficient.		CO5	

users	Ratings				
	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User1	3	1	2	5	3
User2	_	3	4	4	3

b)	Write a Map Reduce pseudo code to Vector-Matrix multiplication. Demonstrate with an example showing all the steps		CO3	Ap
c)	Explain Bloom filtering for stream data mining.Demonstrate false negative and false positive in bloom filter with example.		CO4	Ap
Q.3	Solve any two questions out of three. (10 marks each)	20		
a)	Explain Hadoop Ecosystem in detail.		CO1	U
b)	Explain NOSQL data architectural patterns with example.		CO2	U

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

~~Nov – Dec 2024~~ Jan / Feb 25
(B. Tech.) Program: Computer Engineering Scheme: II
Regular Examination: LY Semester: VII
Date of Exam: ~~28/11/2024~~ 05-02-25 Course Code: CEC702 and Course Name: Big Data Analytics
Duration: 02.5 Hours Max. Marks: 60

c)	Explain different data visualization techniques in R with examples.		CO6	U
Q.4	Solve any two questions out of three. (10 marks each)	20		
a)	Suppose our stream consists of the integers 3, 1, 4, 1, 5, 9, 2,6, 5. Determine the number of distinct elements if the hash function is: h(x) = 3x+ 7 mod 8.		CO4	Ap
b)	Demonstrate the different communities using CPM method for the following graph 		CO5	Ap
c)	State one step matrix multiplication using mapreduce? Demonstrate same for following two 2x2 metrics M= $\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix}$ N= $\begin{vmatrix} 5 & 6 \\ 7 & 8 \end{vmatrix}$		CO3	

****