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

Supplementary Exam Feb-March 2024

Program: B. Tech. Scheme: IIB
Examination: **LY** Semester: **VII**
Course Code: CEC702 Course Name: Big Data Analytics

Date of Exam: 04/3/2024 Duration: 2.5 Hours Max. Marks: 60

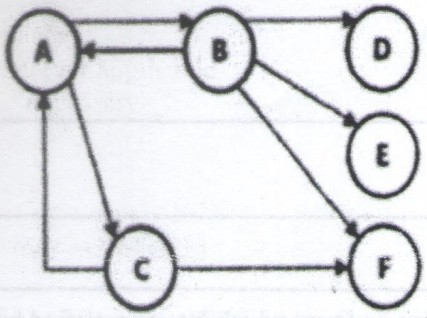
Instructions:
(1)All questions are compulsory.
(2)Draw neat diagrams wherever applicable.
(3)Assume suitable data, if necessary.

		Max. Marks	CO	BT level
Q 1	**Solve any six questions out of eight**	**12**		
i)	Give 2 examples of big data case studies. Indicate which Vs are satisfied by these case studies.	02	CO1	U
ii)	If replication cause data redundancy, then why it is pursued in HDFS?	02	CO2	U
iii)	How CAP is different from ACID properties.	02	CO2	Ap
iv)	Explain relationship between number of map and number of reduce functions required to perform a task	02	CO3	U
v)	What are the typical types of queries that can be executed on streaming data, and how do they differ from standard database queries?	02	CO4	U
vi)	Elaborate how OTT platforms recommendation system works.	02	CO5	E
vii)	Give one example of Direct Discovery of Communities	02	CO5	U
viii)	List different types of data visualization on R	02	CO6	U
Q.2	**Solve any four questions out of six. (M-Marks)**	**16**		
i)	Discuss the main characteristics of Big Data.	04	CO1	U
ii)	List different data architecture patterns in NoSQL? Compare "key value" store and "document" data store patterns with relevant examples.	04	CO2	Ap
iii)	Write Map Reduce Algorithm for following Relational Algebra. 1. Selection 2. Join	04	CO3	U
iv)	Suppose a data stream consists of the integers 3,1,4,1,5,9,2,6,5. Let the hash function being used is h(x)=3x+1 mod 5; Show how the Flajolet-Martin algorithm will estimate the number of distinct element in this stream.	04	CO4	Ap
v)	Define PageRank. Using the webgraph shown below compute the PageRank at every node at the end of the second iteration. Use teleport factor = 0.8.	04	CO5	Ap

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

Supplementary Exam — Feb - March 2024
Program: B. Tech. Scheme: IIB
Examination: **LY** Semester: **VII**
Course Code: CEC702 Course Name: Big Data Analytics

Date of Exam: 04/2/2024 Duration: 2.5 Hours Max. Marks: 60

				
vi)	Write R program to take a string as input from user and check it is palindrome or not.	04	CO6	Ap
Q.3	**Solve any two questions out of three. (M-Marks)**	**16**		
i)	Characterize the components of Hadoop Ecosystem with neat diagram.	08	CO1	U
ii)	Using an example bit stream explain the working of the DGIM algorithm to count the number of 1's (Ones) in a data stream. (3M) Data: 10101100010111010110010110 i) Show initial configuration N=24, what is an estimate of the number of 1's. (2M) ii) The following bits enter the window one at a time:10101011. What is the bucket configuration? Estimate no of 1's in latest K=14 bits of window? (3M)	08	CO4	Ap
iii)	Discuss the role of MapReduce in the efficient computation of PageRank. (4M) How does it address the scalability challenges of processing large web graphs? (4M)	08	CO5	Ap
Q.4	**Solve any two questions out of three. (M-Marks)**	16		
i)	Explain Large Scale File-System Organization.	08	CO2	U
ii)	Write pseudocode for Matrix multiplication using map reduce.	08	CO3	Ap
iii)	Explain a) Reading datasets and Exporting data from R, (3M) b) Manipulating and Processing Data in R, (5M)	08	CO6	U

****