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

Jan- Feb: 2026
~~Nov – Dec 2025~~

(B. Tech) Program: Computer Engineering Scheme: IIB
Supplementary ~~Regular~~ Examination: LY Semester: VII
Course Code: CEC702 and Course Name: Big Data Analytics

Date of Exam: ~~02/12/2025~~ 05/02/2026	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)	Compare and contrast ACID vs BASE.		CO2	U
b)	Which algorithm can be used for finding distinct elements in a long data stream? Give two real life applications for the same.		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)	Predict rating for user 2 for Movie 3 using Pearson's Correlation coefficient.		CO5	Ap
b)	Apply suitable Map Reduce algorithm for joining following 2 tables R and S.		CO3	Ap
c)	Insert elements 10 and 7 in Blooms filter of size 5 using following hash functions: h1(x)= x mod 5 and h2(x)=(2 x+ 6) mod 5		CO4	Ap
Q.3	Solve any **two** questions out of three. (10 marks each)	20		
a)	Explain Hadoop Ecosystem in detail.		CO1	U

Q2 a) table:

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

Q2 b) tables:

R	
A	B
Url 1	Url 2
Url 1	Url 4
Url 3	Url 2
Url 4	Url1

S	
B	A
Url 2	Url 5
Url 4	Url 5
Url 2	Url 3
Url 4	Url 3

Jan-Feb: 2026

~~Nov – Dec 2025~~

(B. Tech) Program: Computer Engineering Scheme: IIB

Supplementary ~~Regular~~ Examination: LY Semester: VII
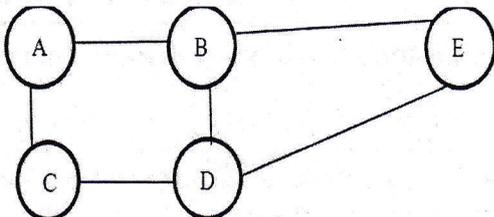
Course Code: CEC702 and Course Name: Big Data Analytics

Date of Exam: ~~02/12/2025~~   Duration: 02.5 Hours   Max. Marks: 60

05/02/2026

b)	List the architectural patterns in NOSQL databases. Discuss the key-value store and document store patterns, focusing on their characteristics, use cases and 2 examples each.		CO2	U
c)	State the "selection", "projection", "union", "intersection" and "difference" relational algebra operations using MapReduce with examples.		CO3	U
Q.4	Solve any **two** questions out of three. (10 marks each)	20		
a)	Suppose our stream consists of the integers 4, 2, 5, 9, 1, 6, 3, 7. Determine the number of distinct elements if the hash function is: $h(x) =. x + 6 \mod 32$. Treat result as 5-bit binary integer.		CO4	Ap
b)	Demonstrate the different communities using Girwan-Newman algorithm for the following graph.		CO5	Ap
c)	Consider above dataframe and using R programming 1) Create a subset of course less than 5 using [] bracket and demonstrate the output. 2) Create a subset where course column is less than 4 or the class equals to 1 by using subset () function and demonstrate the output		CO6	Ap

The table referenced in part c):

Course	id	class	marks
1	11	1	56
2	12	2	75
3	13	1	48
4	14	2	69
5	15	1	84
6	16	2	53

****