

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

Nov – Dec 2025		Scheme: IIB	
(B. Tech) Program: Computer Engineering		Regular Examination: LY Semester: VII	
Date of Exam: 02/12/2025		Course Code: CEC702 and Course Name: Big Data Analytics	
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)	In a distributed file system supporting a real-time recommendation engine, how can rack-aware data placement reduce downtime and improve availability during a rack failure?		CO2	U
b)	Write simple pseudo-code for the Map and Reduce functions to compute a natural join between two relations: R(A, B) and S(B, C) and Join attribute is B		CO3	Ap
c)	What is data stream sampling and why do we need it? What happens if sampling is not done properly in data streams?		CO4	Ap
Q 2	Solve any two questions out of three: (05 marks each)	10		
a)	You design a Bloom Filter-based email spam blocking system. Explain how false positives could negatively affect the user experience and propose a method to reduce the FP rate without increasing memory drastically.		CO4	An
b)	Explain cosine similarity and Pearson correlation coefficient in the context of recommender systems.		CO5	U
c)	Describe any five string functions in R and give one suitable example for each.		CO6	U
Q.3	Solve any two questions out of three. (10 marks each)	20		
a)	What are the advantages of Hadoop? Explain Hadoop Architecture and its components with proper diagram.		CO1	U
b)	In which situations would you NOT use a NoSQL database? Why is horizontal scalability easier with NoSQL compared to relational databases?		CO2	Ap
c)	Use MapReduce for matrix – matrix multiplication. Apply MapReduce operation on the following matrices and show the output of each stage.		CO3	Ap

Seat No.:

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

Nov – Dec 2025	Scheme: IIB
(B. Tech) Program: Computer Engineering	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

Q.4	Solve any two questions out of three. (10 marks each)	20																													
a)	Employ the DGIM algorithm. Shown below is a data stream with $N = 24$ and the current bucket configuration. New elements enter the window at the right. Thus, the oldest bit of the window is the left-most bit shown. 1 0 1 0 1 1 0 0 0 1 0 1 1 1 0 1 1 0 0 1 0 1 1 0 (a) What is the largest possible bucket size for $N = 24$? (2M) (b) Show one way of how the above initial stream will be divided into buckets. (2M) (c) What is the estimate of the number of 1's in the latest $k = 14$ bits of this window? (2M) (d) The following bits enter the window, one at a time: 1 0 1 0 1 0 1 1. What is the bucket configuration in the window after this sequence of bits has been processed by DGIM? (4M)		CO4	Ap																											
b)	What is a Community in a Social Network Graph? (2M) Explain steps of Girvan Newman algorithm. (3M) Apply Girvan Newman on sample web graph to detect communities. (5M)		CO5	U																											
c)	Following is the placement data of XYZ college. <table border="1"><thead><tr><th>Year</th><th>Company</th><th>No of Students Placed</th></tr></thead><tbody><tr><td>2022</td><td>Delloite</td><td>10</td></tr><tr><td>2022</td><td>Zeus Learning</td><td>20</td></tr><tr><td>2023</td><td>Delloite</td><td>30</td></tr><tr><td>2023</td><td>Accenture</td><td>25</td></tr><tr><td>2023</td><td>Zeus Learning</td><td>20</td></tr><tr><td>2024</td><td>Delloite</td><td>35</td></tr><tr><td>2025</td><td>LTI Mindtree</td><td>30</td></tr><tr><td>2025</td><td>Accenture</td><td>40</td></tr></tbody></table> Draw following charts using R programming i) Draw a line chart shows Students Placed Over Years for Deloitte Only ii) Draw a pie chart for Placement Share by Company in 2023	Year	Company	No of Students Placed	2022	Delloite	10	2022	Zeus Learning	20	2023	Delloite	30	2023	Accenture	25	2023	Zeus Learning	20	2024	Delloite	35	2025	LTI Mindtree	30	2025	Accenture	40		CO6	An
Year	Company	No of Students Placed																													
2022	Delloite	10																													
2022	Zeus Learning	20																													
2023	Delloite	30																													
2023	Accenture	25																													
2023	Zeus Learning	20																													
2024	Delloite	35																													
2025	LTI Mindtree	30																													
2025	Accenture	40																													
