

A

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

Nov – Dec 2024		
(B. Tech.) Program: Computer Engineering Scheme: III		
Regular Examination: LY Semester: VII		
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.
- (4) M indicates marks.

Q. No.	Question	Max. Marks	CO	BT level
Q 1	Solve any two questions out of three: (05 marks each)	10		
a)	A large e-commerce company with 24 X 7 operations is currently handling its data through a traditional RDBMS. They want a better solution to handle increasing clientele and manage the data effectively. Discuss which NoSQL database Model will you suggest and why?	05	CO2	Ap
b)	What are the challenges of querying on large data streams? With respect to data stream querying, give example of a) One time Queries b) Continues Queries c) Predefined Queries d) Ad-hoc Queries	05	CO4	U
c)	Write a script to find the sum, mean, product and standard deviation of the vector in R Programming.	05	CO6	C
Q 2	Solve any two questions out of three: (05 marks each)	10		
a)	Using Flajolet Martin Algorithm count distinct elements in the stream: 2, 3, 7, 1, 4, 5, 6, 7, 4 using $H(x) = (x + 4) \bmod 32$.	05	CO4	Ap
b)	What is deadend in PageRank calculation? How are deadends handled in PageRank calculation?	05	CO5	U
c)	What are combiners? When should one use combiner in map reduce job? Explain with an example.	05	CO3	U
Q.3	Solve any two questions out of three. (10 marks each)	20		
a)	Explain Hadoop Ecosystem and briefly explain its components with neat diagram.	10	CO1	U
b)	Explain concept of Map Reduce using an example. Write Map Reduce pseudocode for "Group By" and "aggregation" in a	10	CO3	Ap

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

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

	database. Explain with an example.																																	
c)	How data visualization is implemented using R programming? Give syntax and describe each with an example.	10	CO6	U																														
Q.4	Solve any two questions out of three. (10 marks each)	20																																
a)	Describe characteristics of a NoSQL databases.(4M) What are the business drivers for NoSQL? (4M) What is the main difference between document store and key-value store (2M)	10	CO2	U																														
b)	Give the name of an algorithm used to count number of one's in a window. (1M) Explain it (5M) with an example considering two new bits arriving (4M) in a stream in detail.	10	CO4	Ap																														
c)	Using the user-movie rating data below, predict User 1's rating for <i>The Dark Knight</i> . Calculate the similarity between User 1 and the other users (User 2 , User 3 , and User 4) and use these similarities to predict the missing rating. <table border="1"><thead><tr><th>User</th><th>Inception</th><th>The Godfather</th><th>Parasite</th><th>Spirited Away</th><th>The Dark Knight</th></tr></thead><tbody><tr><td>User 1</td><td>5</td><td>4</td><td>1</td><td>4</td><td>?</td></tr><tr><td>User 2</td><td>3</td><td>1</td><td>2</td><td>3</td><td>3</td></tr><tr><td>User 3</td><td>4</td><td>3</td><td>4</td><td>3</td><td>5</td></tr><tr><td>User 4</td><td>3</td><td>3</td><td>1</td><td>5</td><td>4</td></tr></tbody></table> Using a user-user collaborative filtering recommendation system, determine which movie(s) would be recommended for User 1. Justify the recommendation by identifying the most similar users to User 1 and explain how the ratings of these similar users contribute to the final recommendation.	User	Inception	The Godfather	Parasite	Spirited Away	The Dark Knight	User 1	5	4	1	4	?	User 2	3	1	2	3	3	User 3	4	3	4	3	5	User 4	3	3	1	5	4	10	CO5	Ap
User	Inception	The Godfather	Parasite	Spirited Away	The Dark Knight																													
User 1	5	4	1	4	?																													
User 2	3	1	2	3	3																													
User 3	4	3	4	3	5																													
User 4	3	3	1	5	4																													
