

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

May-June 2025-26		
(B. Tech / M. Tech.) Program: B.Tech Scheme: III		
Regular Examination: TY Semester: VI		
Course Code: AIC602 and Course Name: Big Data Analytics		
Date of Exam: 22/05/2026	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)	Explain the 5 V's of Big Data.		CO1	U
b)	Explain the working process of Resilient Distributed Dataset (RDD) in Spark.		CO2	U
c)	Explain six constraints that must be satisfied for representing a stream by buckets using DGIM algorithm with examples.		CO4	U
Q 2	Solve any two questions out of three: (05 marks each)	10		
a)	Explain Parallel SVM for Large Datasets with algorithm steps.		CO5	U
b)	Write a MongoDB query 1. To display all orders along with customer details using collections orders and customers.(2 M) 2. To display only the name and email fields from the customers collection.(2 M) 3. To find the average rating of restaurants in each city. (1 M)		CO2	U
c)	Briefly explain the different types of data used in healthcare and how they help improve patient care.		CO1	U
Q.3	Solve any two questions out of three. (10 marks each).	20		
a)	Write a MapReduce pseudo code for the word count problem. Apply MapReduce on the following document: "Machine learning is useful. Learning machines improve performance."		CO3	AP
b)	Construct a Bloom filter with size of the bit array $m = 8$, Hash functions are 1. $h1(x) = x \bmod m$, 2. $h2(x) = (2x + 1) \bmod m$,		CO4	AP

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

May-June 2025-26 (B. Tech / M. Tech.) Program: B.Tech Scheme: III Regular Examination: TY Semester: VI Course Code: AIC602 and Course Name: Big Data Analytics		
Date of Exam: 22/05/2026	Duration: 02.5 Hours	Max. Marks: 60

	3. $h_3(x) = (3x + 2) \bmod m$. Insert elements 12, 25, 30, and 5. Check whether 6 and 55 are present, and identify the results as false positive or false negative and why.			
c)	Explain the Girvan–Newman algorithm with its steps and determine communities for the given social network graph using Girvan Newman algorithm.		CO6	AP
Q.4	Solve any two questions out of three. (10 marks each)	20		
a)	Find the frequent itemset with minimum support 3 using PCY algorithm using hash function $h(i, j) = i \times j \bmod 8$. $T_1 \rightarrow 1, 3, 5, 7$ $T_2 \rightarrow 1, 3, 5$ $T_3 \rightarrow 1, 5, 7$ $T_4 \rightarrow 1, 2, 7$		CO2	AP
b)	Explain natural join and grouping and aggregation relational algebra operation using MapReduce pseudo code with example.		CO3	U
c)	Apply the Clique Percolation Method (CPM) step by step to identify all communities in the given graph here $K=3$		CO6	AP
