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

May-June 2026
Program: M. Tech. Scheme: II
Regular Examination: FY Semester: II
Course Code: PCEC202 and Course Name: Big Data Analytics
Date of Exam: 27/05/2026 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 process of Knowledge Discovery in Databases (KDD) and its importance.		CO1	U
b)	What are the five Vs of Big Data? Briefly explain each.		CO2	U
c)	Explain rules for bucket formation in DGIM algorithm.(3M) Show the formation for 110001110100.(2M)		CO6	AP
Q 2	Solve any **two** questions out of three: (05 marks each)	10		
a)	When to use MapReduce in Big Data? Justify your answer with example.		CO4	An
b)	Explain techniques and problems with anonymization techniques with example.		CO5	U
c)	Explain DSMS with neat diagram and types of queries.		CO6	U
Q.3	Solve any **two** questions out of three. (10 marks each)	20		
a)	A retail giant like Amazon wants to improve its personalization engine to offer real-time recommendations. Analyze which specific V's of Big Data are drivers here. Explain how the recommendation will work.		CO1	An
b)	Compare Big data analytics with traditional data mining. List pros and cons of both.		CO2	U
c)	How is data moved in and out of Hadoop? Explain the process with examples.		CO3	U
Q.4	Solve any **two** questions out of three. (10 marks each)	20		

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

May-June 2026
Program: M. Tech. Scheme: II
Regular Examination: FY Semester: II
Course Code: PCEC202 and Course Name: Big Data Analytics
Date of Exam: 27/06/2026 Duration: 02.5 Hours Max. Marks: 60

a)	Explain components of Hadoop Ecosystem with neat diagram. Explain working of YARN and PIG.		CO4	U
b)	What is the importance of privacy in data science? Explain privacy-preserving data mining techniques with relevant examples.		CO5	U
c)	What are applications that make use of counting distinct elemments?(2M) Use FM algorithm for stream of the integers 13,11, 14, 11, 15, 19, 12,16, 15. Determine the number of distinct elements if the hash function is: h(x) = 3x+ 7 mod 12. (Use 4 bit binary representation) (6M) Explain combining estimates to decide appropriate value for the same.(2M)		CO6	AP

****