



K J Somaiya Institute of Engineering and Information Technology
An Autonomous Institute Affiliated to University of Mumbai

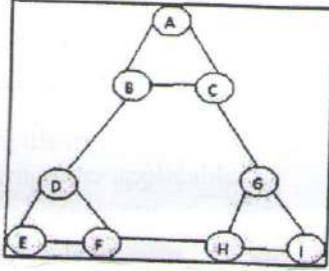
K. J. Somaiya Institute of Engineering and Information Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)
End Semester Exam
Nov – Dec 2021
Program: B. Tech (Computer Engineering)
Examination: LY Semester: VII
Course Code: 1UCEC702 and Course Name: Big Data Analytics
Duration: 03 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 six questions out of eight:	12		
i)	Define Big Data.		CO1	U
ii)	Compare NoSQL and SQL.		CO2	U
iii)	Determine different cases for node failures.		CO3	U
iv)	List the sampling techniques for stream computing.		CO4	U
v)	Calculate cosine distance between {3, -1, 2} and {-2, 3, 1}.		CO5	Ap
vi)	Solve any example of page rank algorithm.		CO6	C
vii)	Determine CURE clustering algorithm drawbacks.		CO4	E
viii)	Explain Collaborative filtering in short.		CO6	U
Q.2	Solve any four questions out of six.	16		
i)	Show why traditional approach is not appropriate for Big Data solutions.		CO1	U
ii)	Explain different No SQL architectural patterns.		CO2	U
iii)	Utilize CAP theorem in Amazon e-commerce application and show its usage.		CO3	Ap
iv)	Explain Flajolet-Martin algorithm in short.		CO4	U
v)	Demonstrate the use of cosine distance in any real world application.		CO5	Ap
vi)	Describe why traditional clustering is not suitable for community detection.		CO6	C
Q.3	Solve any two questions out of three.	16		
i)	Explain distribution models of NoSQL systems.		CO2	U



ii)	Illustrate map-reduce implementation for the following task: a) Reduce side Join of two tables b) Intersection of tables		CO3	An
iii)	Write all the communities for the graph given below by solving it using clique percolation method. 		CO6	C
Q.4	Solve any two questions out of three.	16		
i)	Describe any five characteristics of Big Data.		CO1	U
ii)	Explain the working of DGIM algorithm with an example to count number of ones in a data stream.		CO4	U
iii)	Solve Jaccard distance between i) $\{1,2,3,4\}, \{2,3,5,7\}$ ii) $\{a,a,a,b\}, \{a,a,b,b,c\}$.		CO5	Ap