



Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous College Affiliated to University of Mumbai)

MID SEMESTER EXAM

September 2019

Max. Marks : 20
Class : BE Computer
Course Code : CEE71B
Name of the Course: Big Data Analytics

Duration: 60 Minutes
Semester: VII
Branch : Computer Engineering

Instructions:

- (1) All Questions are Compulsory
- (2) Draw neat diagrams

Question No.		Max. Marks	CO	BL
Q 1	Differentiate between HDFS block and InputSplit? Calculate no. of mappers if data size is 1TB and inputsplit size 100MB.	03	CO2	L2
Q 2	NetFlix records which movies each of its customers rented, and also the ratings assigned to those movies by the customers. Movies are rated 1-to-5-stars by customers. Discuss the approach to classify similar customers using Jaccard similarity. Compute the Jaccard bag similarity of the pair of bags: {1,1,1,2}, {1,1,2,2,3}	04	CO3	L4
Q3	Consider a social-networking site that has a relation Friends(User, Friend). This relation has tuples that are pairs (a, b) such that b is a friend of a. The site want to develop statistics about the number of friends members have. Create a sample input file of atleast 8 records. Write a map reduce algorithm to count no. of friends of each person. Show the output of each phase. OR How is val different from var in Scala? Discuss with an example. Write a SCALA program that defines an object .	05	CO2	L3



Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous College Affiliated to University of Mumbai)

	Declare an integer list that has 4 numbers and print all the numbers using for loop.			
Q4	How traditional business approach differ from big data approach ?	04	CO1	L2
Q5	Why is recommendation necessary? Explain content based recommendation with example ? OR What are the drawbacks of traditional clustering algorithms? Explain two pass clustering using CURE algorithm ?	04	CO3	L2

*****END*****