



Sardar Patel Institute of Technology

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

MSE
Oct 2018

Max. Marks: 20
Class: M.Tech (Sem I)
Course Code: CE912

Duration: 60 min.
Time: 10 am-11 am

Name of the Course: Big Data Analytics and Management

Instructions:

- (1) All questions are compulsory
- (2) Draw neat diagrams
- (3) Assume suitable data if necessary

Q.No.		Max Marks	CO
Q.1	What is Hadoop Ecosystem? Explain its components in detail?	05	CO1
Q.2	<p>In order to justify the need of distributed file system, consider following two scenarios A and B and find out total time required (in minutes) to read 1 TB of data.</p> <p>Let there be one machine with 4 I/O channels each channel with capacity of 100 MB/sec</p> <p>Let there be 10 machines each with 4 I/O channels each channel with capacity of 100 MB/sec.</p> <p>Find out the total time needed to read the data in both the scenarios.</p> <p style="text-align: center;">OR</p> <p>a] Name six Hadoop data types</p> <p>b] What is meant by near the data or data locality</p>	5 3+2	CO2
Q.3	<p>What are the drawbacks of traditional clustering algorithm? How CURE clustering algorithm overcomes the problems of traditional clustering?</p> <p style="text-align: center;">OR</p> <p>What is stream computing? Explain a Stream computing algorithm in detail?</p>	05	CO3 CO3
Q.4	Why recommendation is highly in need? What are different key problems in recommendation? How document profile and user profile is created in Content based recommendation?	05	CO3

Suppose there are 6 documents and keywords and their term frequency is as given in a table below calculate doc profile for Doc 3 and Doc 6

Keywords	Important	Machine	Network	Mining	Cooper
Documents					
Doc 1	8	3	6	2	-
Doc 2	7	11	5	-	-
Doc 3	12	8	10	9	7
Doc 4	-	12	-	11	19
Doc 5	14	15	16	-	-
Doc 6	12	-	-	30	15