



# Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058, India

(Autonomous College Affiliated to University of Mumbai)

## Mid Semester Evaluation

Mar 2018

Max. Marks: 30

Class: B.E.

Course Code: CPE8035

Name of the Course: Elective-III Big Data Analytics

Duration: 90 Min

Semester: VIII

Branch: Computer

Instruction:

(1) All questions are compulsory

Q No.		Max. Marks	CO																					
Q.1 ( a )	<p>Find the station wise highest temperature:</p> <table><tr><td>Santa,</td><td>2017,</td><td>38.3</td></tr><tr><td>Colaba,</td><td>2015,</td><td>39.9</td></tr><tr><td>Santa,</td><td>2017,</td><td>40.02</td></tr><tr><td>Colaba,</td><td>2017,</td><td>41.7</td></tr><tr><td>Colaba,</td><td>2015,</td><td>29.3</td></tr><tr><td>Santa,</td><td>2017,</td><td>42.7</td></tr><tr><td>Colaba,</td><td>2015,</td><td>38.1</td></tr></table> <p>Write a program using Mapreduce in java to find the station wise maximum temperature. Clearly state the output at mapper stage and reducer stage</p>	Santa,	2017,	38.3	Colaba,	2015,	39.9	Santa,	2017,	40.02	Colaba,	2017,	41.7	Colaba,	2015,	29.3	Santa,	2017,	42.7	Colaba,	2015,	38.1	4	CO1
Santa,	2017,	38.3																						
Colaba,	2015,	39.9																						
Santa,	2017,	40.02																						
Colaba,	2017,	41.7																						
Colaba,	2015,	29.3																						
Santa,	2017,	42.7																						
Colaba,	2015,	38.1																						
Q.1 ( b )	<p>During check pointing contents of which two files are transferred from one node to another. Also give the name of these two nodes</p> <p>OR</p> <p>Write a command to merge three files a.txt, b.txt, c.txt from hadoop file system and copy it to local file system</p>	3	CO2																					
Q.2 ( a )	Differentiate between Big data and Traditional data. Give 10 parameters.	5	CO1																					
Q.2 ( b )	<p>Write MongoDB query to perform following operations.</p> <p>a) Command to create collection "students".</p> <p>b) Add 2 students with property "name" in collection students.</p> <p>c) Command to return all students.</p> <p>OR</p> <p>Explain Sharding and Replication in NoSQL database.</p>	3	CO3																					
Q.3 ( a )	What is cardinality estimation problem? Explain the algorithm for counting distinct elements in stream	5	CO4																					
Q.3 ( b )	<p>Explain the concept of Bloom Filter with the help of an example</p> <p>OR</p> <p>Explain how to perform Sampling on data in a stream with basic terms. What are the 2 types/methods of Sampling algorithm. Explain any one Sampling algorithm</p>	5	CO4																					
Q.3 ( c )	What is Recommendation System? Explain Collaborative Filtering based recommendation system. How it is different from Content based recommendation system?	5	CO5																					