Reg. No.										

Manipal Institute of Technology
MANIPAL

(A constituent unit of MAHE, Manipal)

VI SEMESTER B. TECH. (COMPUTER SCIENCE AND ENGINEERING)
MAKEUP EXAMINATIONS, JUNE 2019

SUBJECT: ELECTIVE III - BIG DATA INTEGRATION AND PROCESSING [CRA 4006]
REVISED CREDIT SYSTEM
(18/06/2019)

Time: 3 Hours Max. Marks: 50

Instructions to Candidates:
• Answer **ALL** questions & missing data may be suitably assumed.

1.A. Explain Big Data Management System in comparison with Database Management System. 3M

1.B. What is meant by aggregate queries? List few aggregate functions. 2M

1.C. Consider the employee database 5M
Item (itemid, itemname, unitprice, class, quantityonhand, discount)
Customer(custid, custname, dateofregistration, userid, password)
Customerpurchase (custid, Itemid, quantitypurchase, billno, netprice)

Write the SQL queries for the following.
a. List the minimum unit price
b. List all the classes of item whose average unit price is greater than or equal to 400
c. Display the names of customers who have ordered items but quantitypurchased=0
d.List the items in their decreasing order of quantityonhand and increasing order of discount

2.A. Explain the data model of Aerospike with a neat diagram. 5M

2.B. What is meant by data exchange problem? 2M

2.C. Explain dictionary encoding. Explain why is it used? 3M

3.A. With the help of a detailed example, explain the data integration problem. 5M

3.B. How can aggregation help in business decision making and solve problems related to it? List any 4 aggregational operations to help you extract meaning from large data. 2M

3.C. What are transformations? Explain cross/Cartesian, match/join and filter operations. What is the difference between cross and match operation? 3M

4.A. Explain “split-do-merge” and “big data pipelines” and its significance in Big Data processing. 3M

4.B.	Explain Data Parallelism with a suitable diagram.	2M
4.C.	Explain the difference between a narrow transformation and a wide transformation. Also, list and describe two narrow transformations and give example for two wide transformations.	5M
5.A.	What are two types of operations that help with processing in Spark? Explain the reason for runtime error while applying these operations to an RDD. What is the purpose of "cache" option and what is the issue related with it?	3M
5.B.	Which function can be used to create an integer RDD in Spark? Give an example of its usage.	2M
5.C.	What is the advantage of using a .CSV file in data science? Explain how to export and find the documents in MongoDB.	5M