



Manipal Institute of Technology

MANIPAL

(A constituent unit of MAHE, Manipal)

VI SEMESTER B. TECH. (COMPUTER SCIENCE AND ENGINEERING)

REGULAR EXAMINATIONS, April-May 2019

SUBJECT: ELECTIVE III - BIG DATA INTEGRATION AND PROCESSING [CRA 4006]

REVISED CREDIT SYSTEM

(03/05/2019)

Time: 3 Hours

Max. Marks: 50

Instructions to Candidates:

- Answer **ALL** questions & missing data may be suitably assumed.

1.A. Explain aggregation framework used by MongoDB with an example. 3M

1.B. Consider the employee database
Employee(empno, empname, salary, address, dno)
Department(dno, deptname, location)

Write the SQL queries for the following.

- Find all the employees whose name starts with 'a'.
- Find all the employees names who work for the production department.
- Find the average salary of each department.
- Find all departments where total salary is greater than the average of the total salary at all departments. 5M

1.C. What is the advantage of semijoin in distributed computing? Explain the computing steps in semijoin. 2M

2.A. Explain how Aerospike Query Language queries fast, streaming data with a help of neat diagram. 5M

2.B. Explain the concept of Integrated views and schema mapping. 3M

2.C. What are the advantages of Splunk ? 2M

3.A. Explain data fusion. How the voting system and copy detection methods are used to resolve conflicts that may occur because of data fusion? 5M

3.B. In the context of big data pipelines, what is the purpose of:
i. map operation
ii. reduce operation 2M

3.C. Define aggregation. Give any four examples of aggregation operations. 3M

- | | | |
|------|---|----|
| 4.A. | What is Machine Learning? Explain any 2 algorithms implemented in machine learning tools/platforms. | 3M |
| 4.B. | Give any 2 examples of analytical operations that can be performed on Big data to provide insights useful for marketing decision making. | 2M |
| 4.C. | What is Spark SQL? What are its features? Explain how to perform transformation operations using Spark SQL | 5M |
| 5.A. | With the use of a diagram, explain the steps of a Spark pipeline ending with “collect” operation. | 2M |
| 5.B. | Which is the last step/operation in a Spark pipeline? How would you define this operation? Briefly explain 4 common such operations in Spark. | 3M |
| 5.C. | Explain how to export data from MongoDB to CSV file. | 5M |