Manipal Institute of Technology
MANIPAL
Constituent Institute of Manipal Academy of Higher Education

VI SEMESTER B. TECH. (COMPUTER SCIENCE AND ENGINEERING)
END SEMESTER EXAMINATIONS, APRIL 2018
SUBJECT: ELECTIVE III - BIG DATA INTEGRATION AND PROCESSING [CRA 4006]
REVISED CREDIT SYSTEM
(26/04/2018)

Time: 3 Hours   Max. Marks: 50

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

1.A.	Explain five requirements for big data systems.	5M
1.B.	Explain any six reasons how BDMS is different from DBMS.	3M
1.C.	Using word count example, explain reduce operation in big data pipelines.	2M
2.A.	With a neat diagram, explain the data model of Aerospike.	5M
2.B.	Explain text search with aggregation in MongoDb using an example.	3M
2.C.	Explain RDD in spark.	2M
3.A.	With a neat diagram, explain lambda architecture.	5M
3.B.	Explain the difference between a narrow transformation and wide transformation using neat diagrams.	3M
3.C.	Explain SPARK MLLib and SparkStreaming.	2M

4.A.   Consider the following relational schema for the SALES database:
        Customer(*CustNo*, CName, City)
        Order(*OrderNo*, OrderDate, CustNo, Amount)
        Order_Item(*OrderNo, ItemNo*, Qty)
        Item(*ItemNo*, ItemName, UnitPrice)
    Write the following queries in SQL:
    i. Find the details of customers who have placed order for the item 1010.
    ii. Find the details of customer who have ordered the item 1010 but not 1011.
    iii. Find the customer numbers of customers who have made total purchase greater than 10000
    iv. Find the name of the customers who made the biggest single order.   5M

4.B.	Explain dictionary encoding with an example.	3M
4.C.	What are correlated queries? Illustrate with an example.	2M
5.A.	Describe the data integration problem in detail with suitable examples.	5M
5.B.	Illustrate Integrated View and Schema Mapping with an example.	3M
5.C.	Draw and label the query architecture of a data integration system.	2M

****