Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY, MANIPAL 576104

(Constituent College of Manipal University)

SECOND SEMESTER M.TECH. DEGREE END SEMESTER EXAMINATION, MAY-2016
SUBJECT: OPEN ELECTIVE-BIG DATA ANALYTICS AND TECHNOLOGIES (ICT-586)
(REVISED CREDIT SYSTEM)

TIME: 3 HOURS **17/05/2016** **MAX. MARKS: 50**

Instructions to candidates

- Answer any **FIVE FULL** questions.
- Missing data, if any, may be suitably assumed.

1A. Consider the input file 'employee.csv' consisting of the attributes emp id, name, dept, salary, designation. Write map reduce code for computing average salary of each department.

1B. Write a Map Reduce logic for performing K-means clustering algorithm.

1C. What is logistic regression? Write the equation of cost function for logistic regression.

(5+3+2)

2A. Explain Hadoop Architecture with neat diagram and explain the concepts Data blocks, Staging and Replication pipelining.

2B. Write an AQL script to fetch all the email ids in a document.

2C. What is the function of Sqoop in Hadoop Ecosystem? Explain. (5+3+2)

3A. Explain different methods used for text analytics with the purpose of

i. Text summarization

ii. Question Answering

3B. Explain different categories of NOSQL and list the properties of NOSQL.

3C. Compare Stream computing and Complex Event Processing.

(5+3+2)

4A. Consider the document structure for book collection in test database as shown in Figure Q.4A. Write mongodb queries for the following

```
{
   _id: ObjectId(7df78ad8902d)
   title: 'NoSQL Overview',
   description: 'No sql database is very fast',
   by_user: 'tutorials point',
   url: 'http://www.tutorialspoint.com',
   tags: ['mongodb', 'database', 'NoSQL'],
   likes: 10
}
```

Figure Q.4A

i To display a list of how many tutorials are written by each user.

ii To display urls of the books with title as "java" or "NOSQL" and user "tutorials point".

iii Mapreduce Mongodb query for displaying the count of books with likes greater than 100.

4B. Compute theta values for predicting the class label using logistic regression gradient descent function for the data given in Table Q.4B.

Table Q.4B

x1	x2	class
1	0	1
0	0	0
1	1	1
0	1	1

4C. When can machine be considered as learning? What are the different types of learning based on the data?

(5+3+2)

5A. Construct the decision tree for the data given in Table Q.5A.

Table Q.5A

id	age	cart_type	Class
1	25	1	B
2	35	0	A
3	30	1	A
4	20	0	B
5	25	1	B

5B. What is tumbling window in Streams Processing language (SPL)? Write different eviction policies used in tumbling window with one example.

5C. Explain components of Hive architecture with a neat diagram.

(5+3+2)

6A. Write SPL code for printing cities with lowest and highest temperature from file 'readings.csv' consisting of date, city, temperature.

6B. Consider 'runs.csv' with contents playerid, name, year, team, runs, opponent_team, Write a Pig script to compute average runs of each player against "team1".

6C. Consider the document structure given in Figure Q.4A and write a java code to display all the documents from mongodb database 'test' and 'collection book'.

(5+3+2)

****