Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

A Constituent Institute of MAHE, Manipal

II SEMESTER M.TECH. END SEMESTER EXAMINATIONS, APRIL 2018

SUBJECT: OPEN ELECTIVE - BIGDATA ANALYTICS AND TECHNOLOGIES

[ICT 5281]

REVISED CREDIT SYSTEM

(30/04/2018)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Answer ALL the questions.
- Missing data if any, may be suitably assumed.

1A. Write a program for finding maximum temperature in a year using map-reduce technique. Also explain the logical data flow for the same. 5

1B. What is sharding and replication in NOSQL data bases? Explain. 3

1C. Compare GPFS-FPO and HDFS. 2

2A. State and prove Brewers CAP theorem. 5

2B. Distinguish between clustering and classification. What are the major approaches for clustering? Explain each approach in brief. 3

2C. TABLE Q.2C below, shows the data on sales in a bookstore. The data has two fields namely year and sales in year. Using linear regression model, predict the sales in year 2020. 2

TABLE. Q.2C

Year	Sales
2012	1200
2013	1100
2014	1300
2015	1250
2016	1150

3A. Write mongoDB commands for the following. 5

i. To create a Data base called TEST.
ii. To create a collection called EMPLOYEE
iii. To insert 5 documents (employee id, employee name, department, salary) into the EMPLOYEE collection.
iv. To retrieve documents with salary greater than 50000 and employee id >10.
v. To Update schema by adding new attribute Hobbies for employee id =101.

3B. Explain Pig compilation architecture with a neat diagram. Also write a program in Pig for finding wordcount. 3

3C. Write an AQL module to create a view to fetch all email ids in a document. 2

4A. Build a classification model using Perceptron learning algorithm for following data. 5

X1	1	-1	1	2
X2	2	1	-1	-2
Target class	0	0	1	1

Learning Parameter =0.2, Activation function: Sgn(s)=1 if s>0 else -1.
Initial Weights: w1=1, w2=0.5. bias=0.

4B. List and explain the functions of namenode, secondary namenode and data node in HDFS. 3

4C. In a factory, machines A, B and C manufacture 15%, 25% and 60% of the total production of bolts respectively. The bolts manufactured by the machine A have 4% rate of defects, machine B have 2% rate of defects and machine C have 3% rate of defects. A bolt is drawn at random and is found to be defective. What is the probability that it was produced by B? 2

5A. Consider the input file 'student.csv' consisting of the attributes registration no, name, branch and mark. Write a Stream Processing application which displays total number of students in each branch and maximum mark obtained by a student in each branch except for mechanical branch. 5

5B. What is eviction policy? Explain various eviction policies used in tumbling window. 3

5C. What is a Decision tree? Explain with a suitable example 2