REG. NO												

MANIPAL INSTITUTE OF TECHNOLOGY

(A constituent Institute of MANIPAL UNIVERSITY)

MANIPAL - 576 104, KARNATAKA, INDIA

VII SEMESTER B.TECH. (COMPUTER SCIENCE AND ENGINEERING)
END SEMESTER EXAM - NOV 2017
SUBJECT: BIG DATA ANALYTICS (CSE 4006)
DATE: 28-11-2017

TIME: 3 HOURS   MAX.MARKS: 50

> ### Instructions to Candidates
> - **Note:** Answer all full questions.

1A. Along with an example explain the 'V' in big data which is associated with the retention of data for short term or long term **4M**

1B. Explain the Brewer's CAP theorem **4M**

1C. With what feature of $NoSQL$ the database can be made highly available and fault tolerant in $NoSQL$? Which problem of $RDBMS$ is resolved with the Polyglot persistence of $NoSQL$? **2M**

2A. Write the block diagram of a typical $Hadoop$ environment **3M**

2B. Explain the replication feature in $MongoDB$ along with a diagram **5M**

2C. Write the $MongoDB$ commands to **2M**

> ➢ Export the $JSON$ documents from "Customers" collection in the "test" database into a CSV file "Output.txt" in the D: drive and
> ➢ For finding the documents from the Students collection where the StudName field ends with "a".

3A. Write the commands in $MongoDB$ for the following steps. Write the output of steps b, c and d. **4M**

a. Create a DB '$eatables$', create a collection by the name "$fruits$" and then insert 4 documents into the fruits collection. Use arrays to store the list of fruits in each document.
b. Find those documents from the collection in which the array is having '$banana$' in the first index position.
c. Find documents from the "$food$" collection and display first two elements.

d. Update the element at 0th index position of document with $\_id: 3$ by '*apple*'

3B. Explain the tunable consistency property of Cassandra along with a real-time example. **4M**

3C. ➢ Consider a table student in college database with Name, Age, College and Place as the   **2M**
columns. Export only the columns Name and College to a CSV file in Cassandra.

➢ Consider a collection with _id, StudName, and Grade. Display only the StudName and
Grade from all the documents of the Students collection. The identifier _id should be
suppressed and NOT displayed. Write the syntax in *MongoDB*.

4A. Explain the *Hadoop* high-level architecture with more emphasis on the computational   **4M**
framework

4B. Along with a neat diagram explain different phases of *MapReduce* for processing the   **4M**
following text and write the output of each phase.

Input: *A B R*

C C R

A C B

4C. What is the output of the *MongoDB* syntax: **2M**
$db. Students. find(\{Grade: "VII"\}). pretty(). limit(2);$

What is the equivalent *MongoDB* syntax for the SQL query:
$Select * from Students Where Hobbies in ('Chess', 'Skating');$

5A. Write the program syntax in *RHadoop* starting from token generation till submission of the   **5M**
formatted data into *MapReduce* framework.

5B. Explain the Linear Regression concept along with the programming syntax in R   **5M**