DHARMSINH DESAI UNIVERSITY, NADIAD
FACULTY OF TECHNOLOGY
B.TECH. SEMESTER VII [CE]
SUBJECT: (CE720) BIG DATA ANALYTICS

Examination	: Second Sessional	**Seat No**	: _____
Date	: 07/09/2019	**Day**	: Saturday
Time	: 1:45 TO 3:00 PM	**Max. Marks**	: 36

INSTRUCTIONS:
1. Figures to the right indicate maximum marks for that question.
2. The symbols used carry their usual meanings.
3. Assume suitable data, if required & mention them clearly.
4. Draw neat sketches wherever necessary.

Q.1 Do as directed.

(a) A. What is the right number of reducers in MapReduce Application? **[1]**
    B. List out primary phases of a reducer in MapReduce. **[1]**

(b) A Cassandra table with name 'Course' has following columns: id, course_order, course_id, course_owner, title. There is a csv file containing course data located at 'D:/course_data.csv' path. Write Cassandra query to import data from this file into Course table. **[2]**

(c) A. Choose correct option for following syntax in Cassandra: **[2]**
    Primary key((Program,Branch),Student_name, Student_id)
    i.   Primary key with Program as partitioning key gets created.
    ii. Primary key with Student_name and Student_id as partitioning key gets created.
    iii. Primary key with Program and Branch as partitioning key gets created.
    iv. Primary key with Program, Branch, Student_name and Student_id as partitioning key gets created.
    B. Which column is/are clustering column/s in above syntax?

(d) Mention ways to handle one to many entity relation in mongoDB. **[2]**

(e) Why is modeling data important in Big Data Analytics? **[2]**

(f) Describe any two featured data types provided by mongoDB with example usage. **[2]**

Q.2 Attempt *Any Two* from the following questions.

(a) Write Hadoop based map reduce code and explain working with an example to achieve below requirement: **[6]**
    Find mutual friends, given the friend list of each person.

(b) Discuss mongoDB queries to achieve below: **[6]**
    i.   Add extra field with default value to set of records chosen out of many
    ii. Display subset of fields not including -id ordered by two fields- 1st ascending and second descending.
    iii. Find average of a numeric fieldA values within groups of documents having common fieldB value

(c) Write mongoDB based map reduce statements and explain working with an example to achieve below requirement: **[6]**
    Categorize list of people into income tax payers (exemplary three levels based on the tax paid.

Q.3 (a) i. Explain Hinted Handoffs in Cassandra with figure. **[3]**
        ii. Explain two types of replication strategies of Cassandra. **[3]**

    (b) Explain types of Collections in Cassandra with example. **[6]**

OR

Q.3 (a) Answer following questions: **[6]**
        i.   How to use a column which is not part of partitioning key in 'where' clause of Cassandra query?
        ii. Which datatype/column should be used for an attribute which changes only in increments or decrements?
        iii. How to insert a record in Cassandra table which expires after specified time period?
        v.   What is UUID in Cassandra?
        vi. Differentiate UUID and TimeUUID in Cassandra.
        vii. List out write consistency levels in Cassandra.

    (b) Explain various features of Cassandra. **[6]**