Mid-Semester Test
(EC-2 Regular)

Course No.   : DSECLZG522
Course Title   : Big Data Systems
Nature of Exam   : Closed Book
Weightage   : 30%
Duration   : 2 Hours
Date of Exam   : EN

No. of Pages	= 1
No. of Questions	= 6

Note to Students:
1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made, if any, should be stated clearly at the beginning of your answer.

Q.1. What are the desirable characteristics of a Big Data System?   [03Marks]

Q.2. What is meant by Tunable Consistency in Cassandra NoSQL? In a 3 node Cassandra cluster, a Key space is created with a replication factor of 3. Answer the following questions, in each of the scenarios mentioned below them:
1. Consistency of reads.
2. How many nodes can be lost without data loss.
3. What percentage of data is held in each of the nodes.
        Scenario 1: Write Level - Quorum, Read Level - One
        Scenario 2: Write Level - All, Read Level - One
        Scenario 3: Write Level - One, Read Level - All
        Scenario 4: Write Level - Two, Read Level - One   [06Marks]

Q.3. What is the role of the MapReduce programming paradigm in big data analytics, and how does it implement the divide-and-conquer strategy? What are the distinct types of parallelism used in MapReduce computing paradigm?   [05Marks]

Q.4. What are the basic design principles of Hadoop used for bigdata processing?   [06Marks]

Q.5. Explain the system issues for running a typical parallel program in either a parallel or a distributed manner.   [05Marks]

Q.6. Consider a Big Data Analytics project for "Building recommendation system by analyzing the user ratings and feedback". Illustrate different phases in the life cycle of this big data project.

[05Marks]

****