# INSTITUTE OF AERONAUTICAL ENGINEERING
## (Autonomous)

M.Tech I Semester End Examinations (Supplementary) - July, 2017

**Regulation: IARE–R16**

**BIG DATA ANALYTICS**

**(Computer Science and Engineering )**

**Time: 3 Hours**                                                                                                              **Max Marks: 70**

**Answer ONE Question from each Unit**
**All Questions Carry Equal Marks**
**All parts of the question must be answered in one place only**

## UNIT – I

1. (a) Discuss various challenges that affect the Big Data phenomenon. Compare and contrast ACID (Atomicity Consistency Isolation Durability) and BASE (Basically Available Soft state Eventual consistency). **[7M]**

   (b) You have taken charge of a model that has been run previously on a small scale every day but now needs to be run on millions of customers daily and given to analysts who will proactively contact customers. List out the tasks and their priority to ensure that your model work on large data as well. **[7M]**

2. (a) Define Bid Data Analytics and state what made it powerful? What are the benefits of the Master Data Management (MDM) solution? **[7M]**

   (b) If an organization retains large data which is not evaluated or analyzed, will not benefit the organization. Explain how Big Data Analytics would benefit the organization in this scenario. **[7M]**

## UNIT – II

3. (a) Describe the drastic changes in the environment of data analysis with Big Data by comparing different analytical tools for analyzing data. **[7M]**

   (b) State the reason for the problem and procedure to overcome, when an attempt is made to modelling a task and it seems that the required level of accuracy is not being achieved. **[7M]**

4. (a) Define the drivers for Big Data - Velocity, Variety and Viscosity with at least two Big Data analytics applications in detail. **[7M]**

   (b) Describe the method by which companies analyze customer data or other types of data in an effort to identify patterns and discover relationships between different data elements. **[7M]**

## UNIT – III

5. (a) Does Hadoop replace existing Data Warehouse? (OR) Hadoop can be an extremely valuable extension to your data warehouse? **[7M]**

   (b) What is MapReduce? How Map Reduce works in Hadoop? What are the components and architecture of MapReduce. **[7M]**

6. (a) What are the most commonly defined input formats in Hadoop? What is a block and block scanner **[7M]**

   (b) Explain what happens if during the PUT operation, HDFS block is assigned a replication factor 1 instead of the default value 3. **[7M]**

## UNIT – IV

7. (a) List out the writable data types used by Hadoop API that have the same features as default java class. **[7M]**

   (b) How to copy a file into HDFS with a different block size to that of existing block size configuration? **[7M]**

8. (a) Can Hadoop be used to create distributed clusters, based on commodity servers, which provide low-cost processing and storage for unstructured data, log files and other forms of Big Data explain with suitable example. **[7M]**

   (b) HDFS stores data using commodity hardware which has higher chances of failures. So, How HDFS ensures the Fault Tolerance capability of the system? **[7M]**

## UNIT – V

9. (a) You are analyzing a cohort of visitors to two websites. You know that the cohort in website A consists of 576 male visitors and 768 female visitors. Assuming all visitors also visit website B, what is the probability that a randomly selected visitor is female? In social network analysis, what is the definition of a clique? **[7M]**

   (b) How does creating a social network marketing plan different from traditional marketing plan? What are the types of results from mobile analytics? **[7M]**

10. (a) Describe the future of sentiment analysis. List out the techniques used in sentiment analysis. **[7M]**

    (b) Write the steps and procedure to perform sentiment analysis on the tweets downloaded from Twitter. **[7M]**

− ○ ○ ◯ ○ ○ −