

--	--	--	--	--	--	--	--	--	--



# INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

M.Tech I Semester End Examinations (Supplementary) - May, 2019

Regulation: IARE-R16

## BIG DATA ANALYTICS

Time: 3 Hours

(CSE)

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

- (a) What are the elements of Big Data? Justify how it is both distributed and parallel computed? Explain with an example of exporting all data onto cloud (AWS/Rackspace). [7M]

(b) List out various data generators and describe how to manage data and its architecture for analysis. Assuming a prior probability of 0.3 that a given marketing communication will result in a website visit, how many marketing emails were likely opened if you received 60 website visits? [7M]
- (a) Define Big Data analytics and state what made it powerful? What are the benefits of the Master Data Management (MDM) solution? [7M]

(b) List out the steps that are to be taken reasonably in modelling when a model has to be built for a dataset of 200 patients with 4,000 variables including an indicator of whether or not they had developed cancer in the past year. [7M]

### UNIT – II

- (a) Write tools and list the responsibilities of people who bring needed analytical skills to a project for maximizing the benefits of Big Data analytics for organizations. [7M]

(b) State the reason for the problem and procedure to overcome, when an attempt is done to model the price of the cars at auction and you find that your model has trained well, but subsequently does poorly on new data. [7M]
- (a) Determine the advantages of Big Data analytics strategy which are often defined by the three V's volume, variety and velocity with examples. [7M]

(b) Compare reporting and analysis in the context of Big Data analytics which would help to improve business. [7M]

### UNIT – III

- (a) Explain the installation process of HDFS and explain one of the best ways to assess a model usefulness which is built when marketing department of a company is looking for a way to call customers who are likely to churn and persuade them to stop. [7M]

(b) Justify how MapReduce can best be described as a programming model used to develop Hadoop based applications that can process massive amounts of unstructured data with an example. [7M]

6. (a) Explain the difference between NameNode, Backup Node and Checkpoint NameNode. [7M]  
(b) What is commodity hardware? Explain about the partitioning, shuffle and sort phase. [7M]

#### UNIT – IV

7. (a) Explain about the Hadoop input and output with an example and write a note on data integrity? [7M]  
(b) Sketch daemons architecture in Hadoop Distributed File System? What are the applications of Hadoop and Hadoop YARN? [7M]
8. (a) Discuss Hadoop streaming with RUBY and PYTHON programming language with at least one example. [7M]  
(b) Which of the components retrieve the input splits directly from HDFS to determine the number of map tasks. [7M]

#### UNIT – V

9. (a) In social network analysis, how do you measure the centrality of a node based on the number of links each node has? Explain with an example. Describe the key elements of social media. [7M]  
(b) Compare on different analytics like sentiment analytics, social media analytics and web analytics. [7M]
10. (a) Write the steps and procedure to download tweets of interest using Python or R. [7M]  
(b) List out the reasons why mobile analytics is important? List the different performance metrics to measure the success of mobile application. [7M]