



P.E.S. College of Engineering, Mandya - 571 401

(An Autonomous Institution affiliated to VTU, Belagavi)

Seventh Semester, B.E. - Computer Science and Engineering

Semester End Examination; Dec - 2017/Jan - 2018

Big Data Analytics

Time: 3 hrs

Max. Marks: 100

Note: Answer FIVE full questions, selecting ONE full question from each unit.

UNIT - I

- 1 a. One of the sources of Big data is - Channel click information from Set-Top-Box. Suggest a business case analysis of this data helps you to earn money. 10
- b. Big Data Analytics is creating a new market, where customer data from one industry can be collected, categorized, anonymized, and repackages for sale to others-Justify your analysis. 10
- 2 a. A real-time Adaptive Analytics and Decision Engine can help big analysis to track a customer through various stages and engage in a conversation to progress a customer from one stage to the next. List the various stages involved in this process. 10
- b. The data privacy for Big Data is serious business and is causing regulators around the globe to set up a variety of policies and procedures. In view of this, identify important components of data masking algorithm. 10

UNIT - II

- 3 a. List the characteristics of HDFS. 5
- b. Describe the HDFS architecture and HDFS components. 15
- 4 a. MapReduce, programmers need to just design / migrate application into two phases: Map and Reduce. Generate the execution sequence for MapReduce programming. 10
Apply this sequence for searching for specified words in a given text file.
- b. Explain in detail important characteristics or feature of R. 10

UNIT - III

- 5 a. With an example explain running of MapReduce task on RHadoop. 10
- b. Explain with an example File read /write in RHadoop. 10
- 6 a. Identify the various component of the Hadoop streaming MapReduce job. List the Hadoop streaming commands for execution of a MapReduce job. 10
- b. With an example explain MapReduce application. 10

UNIT - IV

- 7 a. Describe the process involved in data analytics project life cycle. 10
- b. Explain ggplot and rCharts data visualization tools. 10

- 8 a. Design a Data Analytical framework to identify the category of a web page belonging to a website. 10
- b. Design a MapReduce problem for calculating the frequency of stock market changes. 10

UNIT - V

- 9 a. Discuss the need for machine learning and explain types of machine-learning algorithms. 10
- b. Explain Linear regression machine learning algorithm with example. 10
- 10 a. Generate the steps involved in the process of Product recommendations in R. 10
- b. List the stages to Implement Clustering algorithm in R and Hadoop. 10

* * *