[5872]-412

M.E. (Computer)
(Data Science)
BIG DATA ANALYTICS
(2017 Pattern) (Semester - I) (510303)

Time : 3 Hours] *[Max. Marks : 50*
Instructions to the candidates:

1) *Answer Q.1 or Q.2, Q.3 or Q.4, Q.5 or Q.6, Q.7 or Q.8, Q.9 or Q.10, Q.11 or Q.12.*

2) *Figures to the right indicate full marks.*

Q1) List the phases of Data Analytics Life Cycle. Explain the Discovery phase in detail with suitable example. **[9]**

OR

Q2) In real world data, types with missing values for some attributes are a common occurrence. Discuss the various method of preprocessing for handling this problem. **[9]**

Q3) Explain the parallel computing techniques to process & manage Big data produced of high velocity. **[9]**

OR

Q4) Explain the working of Apache Hadoop with HDFS & Map Reduce. **[9]**

Q5) Explain the following with their significance in Hadoop Ecosystem. **[8]**
   a) Apache flume
   b) Pig
   c) Oozie
   d) Zookeeper

OR

Q6) Explain the data analysis of Big data along with the use and description of the Hive & Impola in short. **[8]**

Q7) Explain the Apache Spark architecture along with the various components. Describe the Driver manager component in brief. **[8]**

OR

Q8) Write a short note on the following : **[8]**

a) Spark context

b) Resilent Distributed Datasets (RDD)

Q9) What are the different cluster managers available in Apache spark? Explain each in brief. **[8]**

OR

Q10) Explain in brief the transformation and action in Apache Spark? With suitable example. **[8]**

Q11) List the challenges of Data Visualization. Explain the types of visualization in brief. **[8]**

OR

Q12) List the few data visualization tools and discuss any two applications of data visualization along with the use of suitable example with Python/R. **[8]**

■■■