CS/IT 412

Hall Ticket Number:

IV/IV B.Tech (Supplementary) DEGREE EXAMINATION

November, 2019 — **Computer Science & Engineering**

Seventh Semester — **Data Analytics**

Time: Three Hours — **Maximum:** 60 Marks

Answer Question No.1 compulsorily. (1X12 = 12 Marks)

Answer ONE question from each unit. (4X12=48 Marks)

1. Answer all questions (1X12=12 Marks)
   a) Describe the significance of t-test.
   b) Write the R code for two sample t-test
   c) What is Machine Learning?
   d) Define Big Data?
   e) Explain Hadoop ecosystems?
   f) Define Degree of freedom?
   g) Significance of Secondary Name Node in HDFS
   h) Write applications of Map Reduce
   i) What Hadoop eco system contains
   j) Define YARN
   k) Define Hadoop Common
   l) Explain the functionalities of Map Reduce.

UNIT I

2. Explain the characteristics of Big Data 12M

(OR)

3. What is Hypothesis Testing? Explain the following terms with examples
   a)Null Hypothesis b)Alternative Hypothesis
   c) Degrees of Freedom d) P value
   e) How to calculate t test value? f)Type- 1 error & Type-2 error 12M

UNIT II

4. a) What is Null Hypothesis and Alternative Hypothesis with T-Test? 6M
   b) How to calculate t test value? How to calculate t test value? 6M

(OR)

5. a) Write the R code for cluster analysis on iris data set using K-means algorithm iris dataset(Sepal Length, Sepal Width, Petal Length, Petal Width, Species). 6M
   b) Write R code for Hierarchical clustering using single linkage method 6M

UNIT III

6. a) Explain HDFS concepts in detail 6M
   b) Write the R code for cluster analysis on Lung Capacity data set using K-medoids algorithm. Lung Capacity data set (Gender, Height, Smoker, Exercise, Age, Lung Capacity) 6M

(OR)

7. a) Explain HDFS concepts in detail 6M
   b) Explain how YARN runs an application on HDFS? 6M

UNIT IV

8. Explain how HDFS runs a Map Reduce job? 12M

(OR)

9. a) Explain the anatomy of how data read from HDFS 6M
   b) Explain how YARN runs an application on HDFS? 6M