

**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)**(1X12=12 Marks)*

1. Answer all questions

- Describe the significance of t-test.
- Write the R code for two sample t-test
- What is Machine Learning?
- Define Big Data?
- Explain Hadoop ecosystems?
- Define Degree of freedom?
- Significance of Secondary Name Node in HDFS
- Write applications of Map Reduce
- What Hadoop eco system contains
- Define YARN
- Define Hadoop Common
- 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

- Null Hypothesis
- Alternative Hypothesis
- Degrees of Freedom
- P value
- 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

