

**Paper Code:** BCADSE 6.5      **Paper title:** Elective-IV: a. Big Data Analytics      **Teaching Hours –** 5 hrs/week

**Total Teaching Hours:** 60 Hrs.

**Marks:** Th-80+IA-20

**Credits:** 4

### **Unit I**

INTRODUCTION TO BIG DATA- Big Data and its Importance – Four V's of Big Data – Drivers for Big Data – Introduction to Big Data Analytics – Big Data Analytics applications, Architecture Components, Massively Parallel Processing (MPP) Platforms, Unstructured Data Analytics and Reporting, Big Data and Single View of Customer/Product, Data Privacy Protection, Real-Time Adaptive Analytics and Decision Engines. **12 Hrs**

### **Unit II**

INTRODUCTION TO R & HADOOP-Getting Ready to Use R and Hadoop , Installing R ,Installing R Studio, Understanding the features of R language, Installing Hadoop, Understanding Hadoop features ,Learning the HDFS and MapReduce architecture ,Writing HadoopMapReduce Programs, Introducing HadoopMapReduce, Understanding the HadoopMapReduce fundamentals, Writing a HadoopMapReduce example ,Learning the different ways to write HadoopMapReduce in R. **12 Hrs**

### **Unit III**

INTEGRATION OF R & HADOOP-Integrating R and Hadoop ,Introducing RHIPE ,Understanding the architecture of RHIPE Understanding RHIPE samples, Understanding the RHIPE function reference, Introducing R Hadoop ,Understanding the architecture of RHadoop, Understanding RHadoop examples, Understanding the RHadoop function reference. HADOOP STREAMING WITH R Using Hadoop Streaming with R - Introduction, Understanding the basics of Hadoop Streaming, Understanding how to run Hadoop streaming with R, Understanding a MapReduce application, Exploring the Hadoop Streaming R package. **12 Hrs**

### **Unit IV**

DATA ANALYTICS WITH R AND HADOOP -Understanding the data analytics project life cycle – Introduction, Identifying the problem, Designing data requirement ,Preprocessing data ,Performing analytics over data ,Visualizing data, Understanding data analytics problems ,Exploring web pages categorization Case Studies: Computing the frequency of stock market change , Predicting the sale price of blue book for bulldozers. **12 Hrs**

### **Unit V**

UNDERSTANDING BIG DATA ANALYSIS WITH MACHINE LEARNING Introduction to machine learning, Types of machine-learning algorithms ,Supervised machine- learning algorithms, Unsupervised machine learning algorithm, Recommendation algorithms, Steps to generate recommendations in R ,Generating recommendations with R and Hadoop. **12 Hrs**

### **References:**

1. ArvindSathi, “Big Data Analytics: Disruptive Technologies for Changing the Game”, 1st Edition, IBM Corporation, 2012 (Chapter 1,2,3 Unit 1)
2. Big Data Analytics with R and Hadoop, VigneshPrajapati, -Packt Publishing 2013 (Chapters 1,2,3,4,5,6 Unit 2,3,4,5,6)

### **Additional Reading:**

1. Michael Minelli, Michehe Chambers, “Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business”, 1st Edition, AmbigaDhiraj, Wiely CIO Series, 2013. Bill Franks, “Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics”, 1st Edition, Wiley and SAS Business Series, 2012.
2. Tom White, “Hadoop: The Definitive Guide”, 3rd Edition, O'reilly, 2012.