

**SREENIVASA INSTITUTE of TECHNOLOGY and MANAGEMENT STUDIES**

**III MCA - I Semester**

<b>T</b>	<b>P</b>	<b>C</b>
<b>4</b>	<b>0</b>	<b>4</b>

**16MCA312**

**BIG DATA ANALYTICS**

**Course Objectives:**

- To explore the fundamental concepts of Big Data
- To Learn Basic concepts of Hadoop
- To Write Hadoop MapReduce Programs for analyzing Big data
- To Explore Hadoop Ecosystem
- To understand the Big data and Hadoop in Real time Scenario

**UNIT – 1: Understanding Big Data**

Datasets, Data Analysis, Data Analytics-Descriptive Analysis, Diagnostics Analytics, Predictive Analytics, Prescriptive Analytics, Big Data Characteristics – volume, velocity, variety, veracity, value, Different Types of Data – Structured Data, Unstructured Data, Semi-Structured Data

**UNIT – II: Hadoop Basics**

Brief history of hadoop, Apache hadoop and the hadoop ecosystem. A weather dataset, analyzing the data with unix tools, analyzing the data with hadoop , Understanding different Hadoop modes, understanding Hadoop Features-Understanding HDFS, Understanding MapReduce, Learning the HDFS and Mapreduce Architecture-Understanding the HDFS architecture, Understanding the MapReduce Architecture,Understanding the HDFS and MapReduce architecture by plot.

**UNIT – III: Writing Hadoop MapReduce Programs**

understanding the basics of MapReduce, Introducing Hadoop MapReduce-Listing Hadoop mapReduce entities,Understanding the Hadoop MapReduce scenario,Understanding the limitations of MapReduce, Writing a Hadoop MapReduce example-Understanding the steps to run a MapReduce job.

**UNIT – IV: Learning Data Analytics**

**Understanding the data analytics project life cycle** -Identifying the problem, Designing data requirement, Preprocessing data, Performing analytics over data, Visualizing data. **Understanding data analytics problems** - Exploring web pages categorization - Identifying the problem,Designing data requirement,Preprocessing data,Performing analytics over data,Visualizing data.

**UNIT – V: Programming with R**

Basic Syntax, Data types, Variables, Operators, Decision Making, Loops, Functions, Vectors, lists, Matrices, Arrays, Data Frames, R Data Interfaces – CSV Files, Excel Files, Database, R charts & graphs , R statistics – Mean, Median, Mode, Linear Regression.

### Course Outcomes:

- Able to know the basics of Big Data.
- Understand Hadoop Basics
- Able to analyze BigData using Hadoop MapReduce Programs
- Explore Hadoop Componenets like Pig, PigLatin, Hive, Hbase, zookeeper
- Know few Real Time Big Data Analysis

### TEXT BOOKS:

1. "Big Data Fundamentals: Concepts, Drivers & Techniques", 1/e, 2016, Thomas Erl, Wajid Khattak, Paul Buhler, Prentice Hall.
2. "Big Data Analytics with R and Hadoop", 1e, 2013, Vignesh Prajapati, Packt Publishing Ltd, UK.

### REFERENCE BOOKS:

1. "The Art of R Programming: A Tour of Statistical Software Design", revised, 2011, [Norman Matloff](#), No Starch Press
2. "Hadoop: The Definitive Guide," 3/e, 2012, Tom White, O'REILLY Publications.
3. "Understanding Big Data: Analytics for Enterprise Class Hadoop and streaming Data" ,2012, Paul Zikopoulos, IBM, Chris Eaton, Paul Zikopoulos, The McGraw-Hill Companies,
4. "Analytics in a Big Data World: The Essential Guide to Data Science and its Applications", 2014, Bart Baesens, Wiley Publications .
5. "Mining of Massive Datasets", 2012, Anand Rajaraman and Jeffrey David Ullman , Cambridge University Press.