

Devi Ahilya University, Indore, India Institute of Engineering & Technology			IV Year B.E. (Information Technology (Full Time))				
Subject Code & Name	Instructions Hours per Week			Credits			
ITR7E1 Big Data Analytics	L	T	P	L	T	P	Total
	3	1	2	3	1	1	5
Duration of Theory Paper: 3 Hours							

Learning Objectives:

- To increase knowledge of the Big Data landscape.
- Develop a comprehensive knowledge R, Hadoop and Spark for effective data analysis.
- Develop skills in independent managing Big Data projects and related issues.
- Develop ability to carry out research in area of Big Data Analytics.
- .

Prerequisites:

Some programming experience (in any language) is recommended.

COURSE CONTENTS

UNIT- I

Introduction to Big Data : Definition & importance of Big Data - Five Dimensions of Big Data - Volume, Velocity, Variety, Veracity, Valance – Industry examples – Terminologies – Structured Data, Unstructured Data, Semi Structured Data, Streaming Data, Real-Time Data, Meta Data, data at rest, Big Data Applications, Challenges Process of Data Analysis.

UNIT-II

Big Data analytics: Introduction – Concepts - Storing Big Data - Analyzing your data characteristics - Selecting data sources for Analysis - Eliminating Redundant Data - Open Source Technology for Big Data analytics - Predictive Analytics - Crowdsourcing Analytics - Computing Platforms, Limitations, and Emerging Technologies - Consumption of analytics - Modern Analytic Approaches – Ensemble Modeling, Commodity Models, and Text Analysis.

UNIT- III

Introduction to R: R-Environment Setup, Programming with R, Basic Data Types, Vectors, Creating and Naming Vectors, Vector Arithmetic, Introduction to Factors, Factor Levels, Summarizing a Factor, Ordered Factors, Comparing Ordered Factors, Introduction to Data Frame, Subsetting of Data Frames, Extending Data Frames, Sorting Data Frames, Creating a List: Creating a Named List, Accessing List Elements, Reading from CSV files, Writing to CSV file.

UNIT- IV

Introduction to Hadoop: Introduction to learning and knowledge analytics - Big Data From Technology Perspective – Hadoop - Components of Hadoop, Application Development in

Hadoop, The Distributed File System - HDFS, GPFS, Hadoop Cluster Architecture, Batch Processing - Low Latency NoSQL.

UNIT- V

Introduction to Apache Spark: Resilient Distributed Datasets (RDDs), DataFrames & SparkSQL, Spark Operations, Spark Workflow, Python RDD API Examples, Broadcast Variables and Accumulators, Spark's Main Use Cases, Case Studies – Market Basket Analysis using R & Spark, Linear Regression using R & Spark.

Learning Outcomes:

Upon completing the course, students will be able to:

- Apply Knowledge of Big Data to solve real world big data problems.
- Understand the fundamentals of R, Hadoop, Spark programming.
- Work on a real life Project, implementing R, Hadoop, Spark Analytics to create Business Insights.
- Undergo into further research in Big Data.

BOOKS RECOMMENDED:

- [1.] Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Michael Minelli, Michele Chambers, Ambiga Dhira, Wiley India Pvt Ltd, 2013.
- [2.] R for Everyone: Advanced Analytics and Graphics, 1st Ed., Jared P. Lander, Pearson Education, Inc., 2014.
- [3.] Big Data Analytics with R and Hadoop, Vignesh Prajapati, Packt Publishing Ltd, 2013.
- [4.] Big Data Analytics: Turning Big Data into Big Money, Frank J. Ohlhorst, Wiley, 2012.
- [5.] Creating Value with Big Data Analytics: Making Smarter Marketing Decisions, Peter C. Verhoef, Edwin Kooge, Natasha Walk, Taylor & Francis, 2016.
- [6.] K Mark Gardener, Wrox Beginning R: The Statistical Programming Language.
- [7.] Y. anchang Zhao, R and Data Mining: Examples and Case Studies . Elsevier in December 2012.
- [8.] Alex Holmes, Hadoop in practice, Manning Publications, 2012
- [9.] Tom White, Hadoop, The definitive guide, O'Reilly Media, 2010

List of Practical Assignments: (If Applicable)

During the learning of course, students need to do assignments:

- To learn the R Programming language.
 - To explore Rstudio for data Analysis problems.
 - To explore Hadoop for solving the Big data problems.
 - To explore Spark for analysing Big data problems.
-