Course Code	Course Name	Course Type	Cd	L	T	P	Marks		
							Internal	Final Exam	Total
MCA- 304	Big Data Analytics Using R	PCC	4	4	0	0	40	60	100

Course Outcomes

At the end of the course the student will be able to	
CO1	Understand the fundamental concepts of big data and analytics.
CO2	Explore tools and practices for working with big data.
CO3	Understand the research that requires the integration of large amounts of data.
CO4	Perform hands-on with big data tools using R language.
CO5	Analyze data by utilizing statistical and regression models

Detailed Syllabus
Section-A

UNIT–I: Introduction to Big Data and Hadoop: Introduction to Big Data and Big Data Analytics, Types of Digital Data, Big data characteristics and applications, Understanding Big Data Storage-A General Overview of High-Performance Architecture, HDFS, MapReduce and YARN, Map Reduce Programming Model, History of Hadoop, Analyzing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big Data Strategy, Introduction to Infosphere Big Insights and Big Sheets.

(10Hrs)

UNIT–2: Hadoop File Distributed System (HDFS) & Map Reduce: The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Data Ingest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures, Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.

(10Hrs)

UNIT–3: Introduction to R programming: Introduction and Overview, Installation of R studio, Data Types and Functions, Data Structures: Vectors and Lists, Matrices, Arrays, Data frames, Factors, Data Transformations and Tables, Control Structures.

(10Hrs)

Section-B

UNIT–4: Data Visualization using R : Reading and getting data into R (External Data): Using CSV files, XML files, Web Data, JSON files, Databases, Excel files, Data Cleaning operations like Slicing, filtering, Piping etc., Exploratory data analysis

(10Hrs)

UNIT–5: Statistical & Regression Models and Predictive Analysis : Mean, median, Standard deviation, variance, correlation and covariance Random Forest, Decision Trees, Normal and Binomial distributions, Time Series Analysis, Linear and Multiple Regression, Logistic Regression, Survival Analysis Creating data for analytics through designed experiments, creating data for analytics through active learning, Creating data for analytics through reinforcement learning.

(10Hrs)

Textbooks

S. No.	Name of the Books	Author	Publisher	Edition (Pub. Yr.)
1	Big Data Analytics	Raj Kamal and Preeti Saxena	Mc Graw Hill Education	2nd(2019)
2	Gardener, Beginning R	Mark Gardener	John Wiley & Sons	2nd (2012)

Reference Books

S. No.	Name of the Books	Author	Publisher	Edition (Pub. Yr.)
1	Big Data Analytics with R	Simon Walkowiak	Packt Publishing	(2016)