

## NIRMA UNIVERSITY

<b>Institute:</b>	Institute of Technology
<b>Name of Programme:</b>	Master of Computer Application (2-Years Programme)
<b>Course Code:</b>	3MCAD351
<b>Course Title:</b>	Big Data Analytics
<b>Course Type:</b>	Departmental Elective
<b>Year of Introduction:</b>	2021-22

### Credit Scheme

L	T	Practical Component				C
		LPW	PW	W	S	
3	0	2	-	-	-	4

### Course Learning Outcomes (CLO):

At the end of the course, students will be able to –

1. explain the significance and challenges of Big Data
2. interpret Big Data using different tools and frameworks
3. utilize Distributed File System with MapReduce programming
4. apply Big Data techniques for useful business applications

### Syllabus:

**Total Teaching hours: 45**

Unit	Syllabus	Teaching hours
Unit-I	<b>Introduction to Data Analytics:</b> Nature of Data, Types of Digital Data, Classification of Digital Data, Structured Data, Semi-Structured Data, Unstructured Data, Characteristics of Data	04
Unit-II	<b>Introduction to Big Data and Big Data Analytics:</b> Introduction to Big Data, Significance of Big Data, Big Data Dimensions, Drivers for Big Data, what is Big Data Analytics, Big Data Analytics - Importance, Issues and Challenges, Applications	05
Unit-III	<b>Hadoop and MapReduce:</b> Introduction to Hadoop, Comparisons of RDBMS and Hadoop, Distributed Computing Challenges, A Brief History of Hadoop, Hadoop Distributed File System, Processing Data with Hadoop, Hadoop YARN, Hadoop Ecosystem, Hadoop in the cloud, Introduction to MapReduce, Algorithms Using MapReduce	12
Unit-IV	<b>NoSQL technologies:</b> Introduction to NoSQL Databases, Types of NoSQL databases, SQL Vs NoSQL, Why NoSQL, Introduction to the Document Database (MongoDB or similar), Data Types and CRUD operations in Document Database, Introduction to the Graph Database (Neo4j or similar), CRUD operations in Graph Database, Relevant Case Studies	12
Unit-V	<b>Introduction to other frameworks:</b> Data Processing Operators in Pig, HiveQL, Querying Data in Hive, Applications on Big Data using Pig and Hive, Fundamentals of HBase and ZooKeeper, Spark Framework and Architecture, Spark essentials and Components.	06
Unit-VI	<b>Mining Data Streams:</b> The Stream Data Model, Sampling Data in a Stream, Filtering Streams, Counting Distinct Elements in a Stream with Case Studies	06

Self-Study: The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Suggested Readings/References:

1. Seema Acharya and Subhashini C, Big Data and Analytics, Wiley India
2. Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press
3. Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman, Big Data for Dummies
4. Tom White, Hadoop: The Definitive Guide, O'reilly Media
5. Dan McCreary, Ann Kelly, Making Sense of NoSQL Manning Publication dreamtech press
6. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill Publishing
7. Bill Franks, Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, John Wiley & sons
8. Glenn J. Myatt, Making Sense of Data, John Wiley & Sons
9. Paul Zikopoulos, Dirk deRoos, Krishnan Parasuraman, Thomas Deutsch, James Giles, David Corrigan, Harness the Power of Big Data The IBM Big Data Platform, Tata McGraw Hill Publications
10. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer

Suggested List of Experiments:

Sr. No.	Title	Hours
1	To study R Programming Language & basic programming constructs. (R Features, Data types, Operators, Loops, Function, Decision making, Vector, Lists, Matrices, Array, Data frame)	04
2	To read data from CSV files in R and analyze it.	04
3	To perform Hadoop installation & introduction to Map Reduce	04
4	To perform word count using Map Reduce	02
5	To analyze real world dataset using Map Reduce	04
6	To perform installation of Document Database (MongoDB or similar) and study basic commands	02
7	To perform CRUD operations on Document Database (MongoDB or similar)	04
8	To perform CRUD operations on Graph Database (Neo4j or similar)	02
9	To study and prepare a report on Big Data Analytics and Cloud	02
10	To perform data querying using Hive	02
-NA-		

Suggested Case List: