Semester	3
Course Title	Big Data Analytics
Course Code	MMCA311A
Credits	3
Total Hours of Pedagogy	40
L–T–P-S	3-0-0-0
CIE	50
SEE	50
TOTAL	100
Exam Type	Theory
Exam Hours	3 Hrs

Course Learning Objectives:

1. The analytical process model with their requirements.
2. Descriptive, Predictive and Perspective analytics, Mobile Business Intelligence and big data, Crowdsourcing and Inter-and Trans-Firewall analytics.
3. NoSQL and Comparison between Apache Hadoop and other systems, & Hadoop's evolution.
4. To analyze Hadoop Distributed File System for Big Data.
5. To apply algorithms for handling large datasets and develop Map Reduce applications; setting up and configuring development environment.

Module I

Headings: Big Data Fundamentals

Introduction to Big Data: Classification of digital data: Structured Data, Sources of Structured Data, Ease of working with structured Data, Semi-Structured Data, Sources of Semi-Structured Data, Unstructured Data, Sources of Unstructured Data, Dealing with Unstructured data; Definition of Big Data; Characteristics of Big Data (Volume, Velocity, Variety, Veracity, Value); Why Big Data; Traditional BI versus Big Data;

Big Data Analytics Concepts: Example Applications; Basic Nomenclature; Analytics; Analytical Model Requirements; Analysis Process Model; Job profiles involved;

Data Collection, Sampling, and Preprocessing: Sampling; Types of data elements; Visual Data Explorations & Exploratory Statistical Analysis; Missing values; Outlier Detection and Treatment; Standardizing Data; Categorization;

Text Book	1,2
Chapter	1:1,2 2:1,2
RBT	L2, L3

Module II

Headings: Technologies in Big Data Analytics

Big Data Technology: Hadoop's parallel world – critical components of Hadoop; Old versus New approaches; Data Discovery: Work the way people's minds work; Open-Source Technology for Big Data Analytics; The Cloud and Big data, - Difference between Cloud & Big Data; Predictive Analytics Moves into the Limelight; Software as a Service BI; Mobile business intelligence and big data – Ease of Mobile Application Deployment; Crowdsourcing Analytics, Inter-and Trans-Firewall Analytics; Big Data Analytics combined with Descriptive, Predictive and Prescriptive Analytics; Holistic View of Analytics

Text Book	3
Chapter	3
RBT	L3, L4

Module III

Headings: Modern Data Management and Analytics Systems	
NoSQL – Where it is Used, What is it, Types of NoSQL Databases, Why NoSQL, Advantages of NoSQL, What we miss with NoSQL, Use of NoSQL in industry, NoSQL vendors, SQL versus NoSQL, NewSQL, Comparison of SQL, NoSQL, and NewSQL; **Meet Hadoop:** Data; data storage and analysis; Comparison with other systems: RDBMS, Grid Computing, Volunteer computing; A brief history of Hadoop, Hadoop at Yahoo; Apache Hadoop and Hadoop Ecosystem.	
Text Book	1,4
Chapter	T1 - C4, T4 - C1
RBT	L3, L4

Module IV

Heading: HDFS - Design, Concepts, and Operations **Hadoop Distributed File System:** The design of HDFS, HDFS concepts – Blocks, Name nodes and Datanodes, HDFS Federation, HDFS High-Availability, Basic file system operations, Hadoop file systems, The java interface- Reading data from Hadoop URL, Reading data using the file system API, Writing data, Directories, Querying the file system, deleting data, Data flow - Anatomy of a file read, Anatomy of a file write, Coherency Model, Parallel copying with distcp, Keeping an HDFS cluster balanced, Hadoop Archives	
Text Book	4
Chapter	3
RBT	L3, L4

Module V

Heading: MapReduce Framework - Data Analysis, Application Development, and Execution in Hadoop **Map Reduce:** A weather dataset – data format; Analyzing the data with Unix Tools; Analyzing the Data with Hadoop - Map and Reduce, - Java MapReduce; Scaling out, Data Flow, - Combiner Functions; Hadoop streaming – Ruby, Python; Hadoop Pipes Compiling and Running; **Developing a Map Reduce Application:** The Configuration API - Combining Resources, - Variable Expansion; Writing a Unit Test - Mapper, - Reducer; Running Locally on Test Data - Running a Job in a Local Job Runner, Testing the Driver; Running on a Cluster - Packaging, - Launching a Job, - The MapReduce Web UI, Retrieving the Results, - Debugging a Job, - Hadoop Logs, - Remote Debugging.	
Text Book	4
Chapter	2, 5
RBT	L3, L4

Course Outcomes (Course Skill Set):
At the end of the course, the student will be able to:

Sl. No.	Course Outcomes	PO	RBT
1	Demonstrate the classification of digital data, analyze Big Data characteristics and analytics concepts, and apply preprocessing techniques to prepare data for effective analysis and decision-making.	PO1, PO 2	L2, L3
2	Analyze Big Data technologies and Apply integrated analytics approaches to create effective business intelligence solutions.	PO1, PO2	L3, L4
3	Develop skills to apply NoSQL databases and Hadoop ecosystem components for big data solutions.	PO1, PO2, PO3	L3, L4
4	Analyze the design, concepts, and data flow of HDFS to evaluate its role	PO1, PO2,	L3, L4

	in reliable and scalable big data storage.	PO3	
5	Develop and execute MapReduce applications on Hadoop for large-scale data analysis and processing.	PO1, PO2, PO3	L3, L4

Suggested Learning Resources:

Text Books:

Sl. No.	Name of the author	Title of the Book	Name of the publisher	Edition and Year
1	Seema Acharya & Subhasini Chellappan	"Big Data Analytics"	Wiley	1st Edition, 2015
2	Bart Baesens	" Analytics in a Big Data World : The Essential Guide to Data Science and its Applications"	Wiley	2014
3	Michael Minelli, Michehe Chambers, Ambiga Dhiraj	"Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses"	Wiley CIO Series	1st Edition, 2013
4	Tom E White	"Hadoop: The Definitive Guide"	O'reilly	3rd Edition, 2012

Reference Books:

Sl. No.	Name of the author	Title of the Book	Name of the publisher	Edition and Year
1	Boris lublinsky, Kevin t. Smith, & Alexey Yakubovich	"Professional Hadoop Solutions"	Wiley, ISBN:9788126551071	2015
2	Chris Eaton, Dirk deroos et al	"Understanding Big data "	McGraw Hill	2012
3	Vignesh Prajapati	"Big Data Analytics with R and Haoop"	Packet Publishing	2013
4	Tom Plunkett, Brian Macdonald et al	"Oracle Big Data Handbook"	Oracle Press, Technologies, McGraw Hill Education	2014

Web links and Video Lectures (e-Resources):

1. https://www.slideteam.net/big-data-analytics-powerpoint-presentation-slide.html
2. https://www.slideshare.net/slideshow/big-data-ppt-31616290/31616290
3. https://www.youtube.com/watch?v=iANBytZ26MI
4. https://www.youtube.com/watch?v=OP8BsGnqi9c
5. https://www.youtube.com/watch?v=0buKQHokLK8
6. https://www.youtube.com/watch?v=xh4gy1lbL2k

Activity-Based Learning (Suggested Activities in Class)/Practical-Based Learning

1. NoSQL commands
2. Java Interface for Reading data using the file system API - HDFS.
3. Java Interface for Writing data using the file system API - HDFS.
4. Java Programs on MapReduce-Map function, Reduce Function, MapReduce Application

CO-PO Mapping Table

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	3	2					
CO2	3	3						
CO3	3	3	3					
CO4	3	3	3					
CO5	3	3	3					