MAHENDRA ENGINEERING COLLEGE
(Autonomous)

Syllabus

Department	Information Technology	Programme Code	2071

VI Semester

Course code	Course Name	Periods/week			Credit	Maximum marks
		L	T	P	C	
15IT14601	DATA ANALYTICS	3	0	0	3	100

Objective(s)	• To familiarize the concepts of big data • To explore the fundamental concepts of big data and analytics • To design applications using Map Reduce Concepts • To analyze its technologies for efficient processing of data using big data frameworks
Outcome(s)	Upon completion of this course, students will be able to • An understanding of different types of quantitative data • An understanding of how to summarize the empirical distribution of data, and create simple visualizations. • A regression analysis, and draw meaningful inference from the results • A basic statistical analysis to solve decisions in complex strategic scenarios

UNIT-I	**INTRODUCTION TO BIG DATA**	**9**

Introduction to Data - Characteristics of Data and Types of digital data: Unstructured- Semi-structured and Structured- Sources of data- Working with unstructured data - Evolution and Definition of big data - Characteristics and Need of big data - Challenges of big data - Data environment versus big data environment.

UNIT-II	**BIG DATA ANALYTICS**	**9**

Overview of Business Intelligence - Data Science and Analytics - Meaning and Characteristics of big data analytics - Need of big data analytics - Classification of analytics - Challenges to big data analytics - Importance of big data analytics - Basic terminologies in big data environment.

UNIT-III	**BIG DATA TECHNOLOGIES AND DATABASES**	**9**

Introduction to NoSQL - Uses - Features and Types – Need – Advantage - Disadvantages and Application of NoSQL - Overview of NewSQL - Comparing SQL - NoSQL and NewSQL - Introduction to MongoDB and its needs - Characteristics of MongoDB - Introduction of apache cassandra and its needs - Characteristics of Cassandra.

UNIT-IV	**HADOOP FOUNDATION FOR ANALYTICS AND MAPREDUCE**	**9**

History –Features - Key advantage and Versions of Hadoop - Essential of Hadoop ecosystems - RDBMS versus Hadoop - Key aspects and Components of Hadoop - Hadoop architectures - Hadoop MapReduce: Introduction to MapReduce - Processing data with Hadoop using MapReduce.

UNIT-V	**YARN FRAMEWORK AND BIG DATA WITH HIVE AND PIG**	**9**

Managing Resources and Applications with YARN - Overview of Hive and its architecture - Hive data types and file format - Hive Query Language (HQL) - Introduction to Pig - Pig latin overview - Data types in Pig and Running Pig - Introduction to Machine Learning - Examples of ML Applications – Supervised Learning – Bayesian Decision Theory

TOTAL PERIODS	**45**

TEXT BOOKS :	
1	Seema Acharya, Subhashini Chellappan – “Big Data and Analytics”, Wiley 2015 Edition.
2	Alex Holmes - “Big Data Black Book”, Dreamtech.
REFERENCES:	
1	Minelli, Chambers, Dhiray- “Big Data Big Analytics”, Wiley.
2	Bart Baesens – “Analytics in a Big Data World”, Wiley.
3	Boris Lublinsky, Kevin T. Smith – “Hadoop Solutions”, Wrox.
4	Chuck Lam – “Hadoop in Action”, Dreamtech
5	Ethem Alpaydin – “Introduction to Machine Learning”, MIT Press 3rd Edition 2015