BIG DATA ANALYTICS			
Course code:	22IS61	Credits:	03
Prerequisites:	Machine Learning ,Operating System	CIE Marks:	50
Teaching Hours / Week (L : T : P)	2 : 2 : 0	SEE Marks:	50
Total Hours:	40	Exam Hours:	03

Course Objectives:	
1.	Understand the big data platform and its use cases.
2.	Explore the techniques of managing big data using NoSQL, Hadoop.
3.	Use ETL tools for developing business case studies in big data analytics.
4.	Develop the process of map-reduce analytics using Hadoop and related tools.

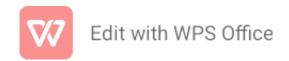| Course Outcomes: At the end of the course, student will be able to:	
CO1	Apply Big Data concepts, tools and applications in engineering and societal problems.
CO2	Apply appropriate solutions for the applications using Hadoop tools.
CO3	Analyze the given data set and identify deep insights from the data set.
CO4	Design and apply appropriate analytics methods based on the nature of the problem, the characteristics of the data, and the desired outcomes.

Mapping of Course outcomes to Program outcomes:														
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3														
CO2	3														
CO3		3													
CO4			2	2	3	2		3	2	2	2	3	3	3	2

Module	Module Contents	Hours
1.	Classification of data, Characteristics, Evolution and definition of Big data, What is Big data, Why Big data, Traditional Business Intelligence Vs Big Data, Typical data warehouse and Hadoop environment. **Big Data Analytics**: What is Big data Analytics, Classification of Analytics, Importance of Big Data Analytics, Technologies used in Big data Environments, Few Top Analytical Tools, NoSQL	08
2.	**Introduction to Hadoop**: Introducing hadoop, why hadoop, why not RDBMS, RDBMS Vs Hadoop, History of Hadoop, Hadoop overview, use case of Hadoop, HDFS (Hadoop Distributed File System), Processing data with Hadoop, Managing resources and applications with Hadoop YARN (Yet Another Resource Negotiator). **Introduction to Map Reduce Programming**: Introduction, Mapper, Reducer, Combiner, Partitioner, Searching, Sorting, Compression.	08
3.	**Introduction to MongoDB:** What is MongoDB, Why MongoDB, Terms used in RDBMS and MongoDB, Data Types in MongoDB, MongoDB Query Language. **Introduction to Cassandra:** Apache cassandra- An Introduction, Features, CQL Datatypes, CQLSH, Keyspaces, CRUD, Collections	08
4.	**Introduction to Hive:** What is Hive, Hive Architecture, Hive data types, Hive file formats, Hive Query Language (HQL), RC File implementation, User Defined Function (UDF). **Introduction to Data Analysis with Spark:** What Is Apache Spark? A Unified Stack, **Programming with RDDs:** RDD Basics, Creating RDDs, RDD Operations, Transformations, Passing Functions to Spark-Python.	08
5.	**Machine Learning with MLlib:** Overview, Machine Learning Basics, Data Types, Algorithms - Feature Extraction, Statistics, Classification and Regression, Clustering, Collaborative Filtering and Recommendation, Dimensionality Reduction, Model Evaluation	08

Text Books:

1. Seema Acharya and Subhashini Chellappan, "Big data and Analytics", Wiley India Publishers, 2nd Edition, 2019. (Ch 1: 1.1, Ch2: 2.1-2.5,2.7,2.9-2.11, Ch3: 3.2,3.5,3.8,3.12, Ch4: 4.1,4.2, Ch 5: 5.1-,5.8, 5.10-5.12, Ch 6: 6.1-6.5, Ch 7: 7.1-7.7, Ch 8: 8.1 - 8.8, Ch 9: 9.1-9.6,9.8)

2. Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia, "Learning Spark: Lightning-Fast Big Data Analysis", O'Reilly, 2015. (Ch 1, Ch 3, Ch 11)

Reference Books:

1. Tom White, Hadoop: The Definitive Guide, ISBN-13: 978- 9352130672, 4th Edition, O"Reilly Media, 2015.
2. Arshdeep Bahga, Vijay Madisetti, Big Data Analytics: A Hands-On Approach, ISBN-13: 978-
0996025577, 1st Edition, VPT Publications, 2018.
3. Eric Sammer, Hadoop Operations: A Guide for Developers and Administrators, ISBN-13: 978-9350239261, 1st edition, O'Reilly Media, 2012.