	Course Title: Big Data Analytics	
	Course Code : 21CST701	**No. of Credits: 3: 0: 0 (L-T-P)**	**No. of lecture hours/week : 3**
	Exam Duration : 3 hours	**CIE+ Assignment + SEE = 45+5+50=100**	**Total No. of Contact Hours : 42**

Course Objectives:	**Description**
	1. To know the fundamental concepts of big data and analytics. 2. Apply analytics on Structured and Unstructured Data. 3. Fundamentals of python/R programming languages 4. To explore tools and practices for working with big data

Unit No	Syllabus Content	No of Hours
1	**Types of Digital Data:** Classification of Digital data: Structured Data, Semi Structured Data, Unstructured Data; Introduction to Big Data: Characteristics of Data, Evolution of Big Data, Definition of Big Data, Challenges of Big Data, What is Big Data?, Why Big Data? Traditional Business Intelligence versus Big Data, A Typical Data Warehouse Environment, A Typical Hadoop Environment, and Coexistence of Big Data and Data Warehouse.	9
2	**Hadoop:** Features, Advantages of Hadoop, Versions of Hadoop, Hadoop ecosystem, Hadoop distributions, Hadoop Vs SQL. Introduction to Hadoop: Why Hadoop? RDBMS Vs Hadoop, Distributed computing challenges, History of Hadoop, Hadoop overview, use case of Hadoop, HDFS, Processing data with Hadoop, Managing resources and applications with Hadoop YARN.	9
3	**Introduction to Map Reduce Programming:** Introduction, Mapper, Reducer, Combiner, Partitioner, Searching, sorting, compression. Big Data Analytics: Big Data Analytics. Classification of Analytics, Greatest challenges on Big Data, Big Data Analytics importance, Data Science, Terminologies in Big Data.	8
4	**Introduction to HIVE:** Introduction, HIVE architecture, HIVE data types, HIVE file formats, HIVE query language, RCFile implementation, SerDe, User Defined Functions (UDF) **Introduction to PIG:** Anatomy of PIG, PIG	8

	on Hadoop, PIG philosophy, overview of PIG, Data types in PIG, Running and execution modes of PIG, HDFS commands, Relational operators, Eval function, Complex Data types.	
5	**Introduction to Apache Spark:** Introduction to Apache Spark: The genesis of Spark, Hadoop at Yahoo and Spark early years, What is Apache Spark, Unified Analytics, Apache Spark's Distributed Execution, Spark Application and Spark session, Spark Jobs, Spark stages , Spark tasks, Transformation, Actions and Lazy Evaluation, Narrow and wide transformation, The Spark UI, Your first Standalone application.	8

Course Outcomes	Description	RBT Levels
CO1	Describe Big Data and its importance with its applications.	**L2**
CO2	Differentiate various big data technologies like Hadoop MapReduce, Pig, Hive, Hbase and No-SQL.	**L4**
CO3	Apply tools and techniques to analyze Big Data.	**L4**
CO4	Design a solution for a given problem using suitable Big Data Techniques	**L3**

CO-PO Mapping	PO1	PO2	PO3	PO4	PO5	P06	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	3	3	2					2	2		3
CO2	3	3	3	3	2				2	2		3
CO3	3	3	2	2	2				2	2		3
CO4	2	2	2		2				2	2		3

Strong -3 Medium -2 Weak -1

TEXTBOOKS:

1. Big Data and Analytics by Seema Acharya, SubhasiniChellappan , Wiley Publications, 2015
2. Jules S. Damji, Brooke Wenig, Tathagata Das & Denny Lee Foreword by Matei Zaharia, "Learning Spark: Lightning-Fast Big Data Analysis", O'Reilly, 2020, Second Edition

REFERENCE BOOKS:

1. Illustrateing Big data, Chris Eaton,Dirkderooset al, McGraw Hill, 2012
2. Professional Hadoop Solutions, Boris lublinsky, Kevin T. Smith, Alexey Yakubovich,Wiley, ISBN:9788126551071, 2015

SELF STUDY REFERENCES / WEBLINKS:

1. http://www.bigdatauniversity.com/
2. https://www.coursera.org/specializations/big-data
3. https://www.ibm.com/topics/big-data-analytics

| **COURSE COORDINATOR:** | **Yashaswini H M** |