Big Data Analytics

Course Code	20CS4703A	Year	IV	Semester	I
Course Category	PEC	Branch	CSE	Course Type	Theory
Credits	3	L-T-P	3-0-0	Prerequisites	Machine Learning
Continuous Internal Evaluation :	30	Semester End Evaluation:	70	Total Marks:	100

Course Outcomes		
Upon successful completion of the course, the student will be able to		
CO1	Understand the fundamental concepts of big data	L2
CO2	Apply Hadoop Distributed File Systems for storing Big data	L3
CO3	Apply the concepts of Map reduce, Spark and Hive for processing the Big Data	L3
CO4	Apply various analytics mechanisms to design a recommender system.	L3

Contribution of Course Outcomes towards achievement of Program Outcomes & Strength of correlations (3:Substantial, 2: Moderate, 1:Slight)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	2													
CO2	2									1				
CO3	3									1				
CO4						1	1		1	1				2

Syllabus		
Unit No.	**Contents**	**Mapped CO**
I	**Meet Hadoop:** Data, Data Storage and Analysis, Querying All Your data, Beyond Batch, Comparison with Other Systems: Relational database Management Systems, Grid Computing, Volunteer Computing. **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.	**CO1,CO3**
II	**The Hadoop Distributed File system:** The Design of HDFS, HDFS Concepts, The Command Line Interface, Hadoop Filesystems, The Java Interface, Data Flow, Parallel Copying with distcp.	**CO1,CO2**
III	**Hive:** Hive Shell, An Example, Running Hive: Configuring Hive, Hive Services, The Metastore, Comparison with Traditional Databases: Schema on Read Versus Schema on write, Updates, transactions, and Indexes. **HiveQL:** Data Types, Operators and Functions and Tables: managed Tables and External Tables, Partitions and Buckets, Storage Formats, Importing Data, Altering Tables and Dropping Tables.	**CO1,CO3**
IV	**Spark:** Installing Spark, An Example: Spark Applications, Jobs, Stages, and Tasks, A Scala Standalone Application, A Java Example, A Python Example, Resilient Distributed Datasets: Creation, Transformations and Actions, Persistence, Serialization, Shared Variables: Broadcast Variables, Accumulators, Anatomy of a Spark Job: Run:Job Submission, DAG Construction, Task Scheduling, Task Execution	**CO1,CO3**
V	**Use case Study:** Recommendation Systems: Introduction, A Model for Recommendation Systems, Collaborative Filtering System and Content Based Recommendations.	**CO1,CO4**

Learning Resources	
Text Books	
1	Hadoop: The Definitive Guide, Tom While ,Fourth Edition,O'Reilly.
2	Big Data Analytics, RadhaShankarmani, M Vijayalakshmi , Second Edition, 2017, Wiley.
References	
1	Hadoop Essentials: A Quantitative Approach, Henry H. Liu , First Edition, 2012, PerfMath Publishers
2	Big Data and Analytics, Seema Acharya, SubhashiniChellappan, First Edition, 2015, Wiley.
3	Big data analytics with R and Hadoop, VigneshPrajapati,First Edition, 2013, SPD.
4	Spark: The Definitive Guide :Big Data Processing Made simple, Bill Chambers and MateiZaharia , First Edition, 2018 , O'Reilly
e-Resources and other Digital Material	
1	https://nptel.ac.in/courses/106/104/106104189/
2	https://www.coursera.org/specializations/big-data
3	https://www.edx.org/course/big-data-fundamentals