

Program Elective-V

Big Data

Course Code	19CS4702A	Year	IV	Semester	I
Course Category	Program Elective-V	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 basic concepts of big data	L2
CO2	Apply the concept of HDFS, Map reduce, Spark for storing and processing of Big data	L3
CO3	Apply Hive for working with Big data and formulate an efficient report	L3
CO4	Apply various analytics mechanisms to design a recommender system.	L3

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, A Brief History of Apache Hadoop. Map Reduce: A Weather Dataset: Data Format, Analyzing the data with Unix Tools, Analyzing the Data with Hadoop: Map and Reduce, Java Map Reduce, Scaling Out: Data Flow, Combiner Functions, Running a Distributed Map Reduce Job.	CO1,CO2
II	The Hadoop Distributed Filesystem: The Design of HDFS, HDFS Concepts, The Command Line Interface, Hadoop File systems, The Java	CO1,CO2

	Interface, Data Flow, Parallel Copying with distcp.	
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, SQL-on Hadoop Alternatives, 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,CO2
V	Use case Study: Recommendation Systems: Introduction, A Model for Recommendation Systems, Collaborative Filtering System and Content Based Recommendations.	CO1,CO4

Text Book
<ol style="list-style-type: none"> 1. Hadoop: The Definitive Guide, Tom White, Fourth Edition, 2015, O'Reilly. 2. Big Data Analytics, RadhaShankarmani, M Vijayalakshmi, Second Edition, 2017, Wiley
References
<ol style="list-style-type: none"> 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
<ol style="list-style-type: none"> 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