

## BIG DATA

Semester	Subject Code	Category	Lecture Hrs		Theory Hrs		Practical		Credits
			Per week	Per Sem	Per week	Per Sem	Per week	Per Sem	
I	21CPCS1D	ELECTIVE - I	6	90	6	90	0	0	5

### COURSE OBJECTIVE

This course aims at facilitating the student to explore and understand the Big data platform, its architecture and its technology foundations. Work on hadoop platform. Perform mining and analysis on massive data using certain techniques. Also perform analysis through mining techniques.

### COURSE OUTCOME

successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level (K1-K4)
CO1	Understanding the Needs of big data	K2
CO2	Developing map reduce concepts.	K3
CO3	Analyzing the mining of data stream concept.	K3
CO4	Evaluate the relationships or connections between network nodes.	K4
CO5	The Concept of Big Data as Presented Through Social Media.	K4

*Knowledge Level – K1-Remember, K2- Understand, K3-Apply, K4-Analyze*

### MAPPING WITH PROGRAMME OUTCOME

COS	PO1	PO2	PO3	PO4	PO5	PO6
CO1	S	S	S	M	M	M
CO2	S	S	S	S	S	M
CO3	S	M	S	M	S	S
CO4	S	S	M	S	M	S
CO5	S	S	M	M	S	S

*S-Strong, M-Medium and L-Low*

**UNIT I: INTRODUCTION TO BIG DATA****18 Hours**

**Introduction to Big Data**- Big Data Characteristics- Types of Big Data- Traditional Versus Big Data Approach - Technologies Available for Big Data – Hadoop – Introduction - What is Hadoop? - Core Hadoop Components - Hadoop Ecosystem - Physical Architecture - Hadoop Limitations

**Unit II: HADOOP AND MAPREDUCE****18 Hours**

**MapReduce** - MapReduce and The New Software Stack- MapReduce- Algorithms Using MapReduce - Finding Similar Items – Introduction - Nearest Neighbor Search - Applications of Nearest Neighbor Search- Similarity of Documents - Collaborative Filtering as a Similar-Sets Problem - Recommendation Based on User Ratings- Distance Measures.

**UNIT III: MINING DATA STREAMS****18 Hours**

**Mining Data Streams** – Introduction- Data Stream Management Systems- Data Stream Mining - Examples of Data Stream Applications - Stream Queries- Issues in Data Stream Query Processing - Sampling in Data Streams - Filtering Streams - Counting Distinct Elements in a Stream- Querying on Windows – Counting Ones in a Window -Decaying Windows.

**UNIT IV: LINK ANALYSIS****19 Hours**

**Link Analysis** - Introduction- History of Search Engines and Spam –PageRank - Efficient Computation of PageRank - Topic-Sensitive PageRank- Link Spam- Hubs and Authorities Recommendation Systems - Introduction - A Model for Recommendation Systems - Collaborative-Filtering System - Content-Based Recommendations - Mining Social Network Graphs .

**UNIT V: APPLICATIONS OF SOCIAL NETWORK MINING****17 Hours**

**Introduction** - Applications of Social Network Mining - Social Networks as a Graph - Types of Social Networks - Clustering of Social Graphs - Direct Discovery of Communities in a Social Graph - Sim Rank-Counting Triangles in a Social Graph

**Distribution of Marks: Theory 70% and Applications 30%**

## TEXTBOOKS

S. NO	AUTHORS	TITLE	PUBLISHERS	YEAR OF PUBLICATION
1	Radha Shankarmani, M Vijayalakshmi	Big Data Analytics	Wiley Publications, first Edition,	2016

## REFERENCEBOOKS

S. NO	AUTHORS	TITLE	PUBLISHERS	YEAR OF PUBLICATION
1	Seema Acharya, Subhashini Chellappan	Big Data and Analytics	Wiley Publication first edition	Reprint in 2016
2	Hive, Yarn, PIG,	Black Book- Big Data(Covers Hadoop,MapReduce, R, Data visualization)	DT Editorial Services Dream techPress edition	2016

## WEB RESOURCES

1. [https://www.sas.com/en\\_us/insights/big-data/what-is-big-data.html](https://www.sas.com/en_us/insights/big-data/what-is-big-data.html)
2. [https://www.tutorialspoint.com/big\\_data\\_tutorials.htm](https://www.tutorialspoint.com/big_data_tutorials.htm)

## TEACHING METHODOLOGY

- Class room teaching & Group discussions
- Seminars & Smart Class room
- Chart/Assignment & Simulation Model

## SYLLABUS DESIGNERS

- Mrs.G.SANGEETHA LAKSHMI, Assistant professor & HOD, Dept of Computer Science & Applications
- Mrs. B ARULMOZHI , Assistant Professor , Dept of Computer Science & Applications