CO	PO						PSO		
	1	2	3	4	5	6	1	2	3
1.	√		√	√			√	√	
2.	√		√				√	√	
3.	√		√				√	√	
4.	√		√	√		√	√	√	√
5.	√		√	√			√	√	√

BD5151 BIG DATA MINING AND ANALYTICS L T P C
3 0 0 3

OBJECTIVES:

- To understand the computational approaches to Modeling, Feature Extraction.
- To understand the need and application of Map Reduce.
- To understand the various search algorithms applicable to Big Data.
- To analyze and interpret streaming data.
- To learn how to handle large data sets in main memory.
- To learn the various clustering techniques applicable to Big Data.

UNIT I DATA MINING AND LARGE SCALE FILES 9

Introduction to Statistical modeling – Machine Learning – Computational approaches to modeling – Summarization – Feature Extraction – Statistical Limits on Data Mining – Distributed File Systems– Map-reduce – Algorithms using Map Reduce – Efficiency of Cluster Computing Techniques.

UNIT II SIMILAR ITEMS 9

Nearest Neighbor Search – Shingling of Documents – Similarity preserving summaries – Locality sensitive hashing for documents – Distance Measures – Theory of Locality Sensitive Functions – LSH Families – Methods for High Degree of Similarities.

UNIT III MINING DATA STREAMS 9

Stream Data Model – Sampling Data in the Stream – Filtering Streams – Counting Distance Elements in a Stream – Estimating Moments – Counting Ones in Window – Decaying Windows

UNIT IV LINK ANALYSIS AND FREQUENT ITEMSETS 9

Page Rank –Efficient Computation – Topic Sensitive Page Rank – Link Spam – Market Basket Model – Apriori algorithm – Handling Larger Datasets in Main Memory – Limited Pass Algorithm – Counting Frequent Item sets.

UNIT V CLUSTERING 9

Introduction to Clustering Techniques – Hierarchical Clustering –Algorithms – K-Means – CURE – Clustering in Non – Euclidean Spaces – Streams and Parallelism – **Case Study:** Advertising on the Web – Recommendation Systems

TOTAL : 45 PERIODS

OUTCOMES:

Upon completion of the course, the student will be able to

- Design algorithms by employing Map Reduce technique for solving Big Data problems.
- Identify similarities using appropriate measures.
- Point out problems associated with streaming data and handle them.
- Discuss algorithms for link analysis and frequent itemset mining.
- Design solutions for problems in Big Data by suggesting appropriate clustering techniques.

REFERENCES:

1. Jure Leskovec, AnandRajaraman, Jeffrey David Ullman, "Mining of Massive Datasets",Cambridge University Press, Second Edition, 2014.
2. Jiawei Han, MichelineKamber, Jian Pei, "Data Mining Concepts and Techniques", Morgan Kaufman Publications, Third Edition, 2011.
3. Ian H.Witten, Eibe Frank "Data Mining – Practical Machine Learning Tools and Techniques", Morgan Kaufman Publications, Third Edition, 2011.
4. David Hand, HeikkiMannila and Padhraic Smyth, "Principles of Data Mining", MIT Press,2001.

CO	PO						PSO		
	1	2	3	4	5	6	1	2	3
1.	√		√	√			√	√	√
2.			√	√		√	√		
3.	√		√	√			√	√	√
4.	√		√	√			√	√	√
5.	√		√	√			√	√	√

CP5084 PARALLEL ALGORITHMS L T P C
3 0 0 3

OBJECTIVES:

- To learn parallel algorithms development techniques for shared memory and DCM models.
- To study the main classes of fundamental parallel algorithms.
- Learn to design efficient parallel algorithms.
- To study the complexity and correctness models for parallel algorithms.
- To understand parallel solutions for bitwise computation.

UNIT I INTRODUCTION 9

Introduction to Parallel Algorithms – Models of computation – Selection – Merging on EREW and CREW – Median of two sorted sequence – Fast Merging on EREW – Analyzing Parallel Algorithms

UNIT II SORTING & SEARCHING 9

Sorting Networks – Sorting on a Linear Array – Sorting on CRCW, CREW, EREW – Searching a sorted sequence – Searching a random sequence – Bitonic Sort