FACULTY OF COMPUTER SCIENCE

Master of Computer Application (Sem-II)

In Effect from Academic Year 2023-24

Branch Name :	MCA
Program Code :	CS201
Course Name :	Big Data Analytics
Course Code :	3CS2010208T
Pre-requisite Course :	Working knowledge of programming language and database concepts

Course Objectives:

1. To understand basics of Big Data and Big Data Tools (Hadoop, MapReduce)
2. To understand fundamental techniques used for Big data analytics
3. To help a student to perform a variety of "analytics" on different data sets and to arrive at positive conclusions.

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
				Theory		Practical		
Lecture	Tutorial	Practical	Credit	University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	Total
4	-		4	60	40	-	-	100

Course Contents:

Unit No	Topic	Total Hours	Weightage (%)
1	**Overview – Big Data Tools:** **Big Data:** Introduction to Big Data, Big Data characteristics, Types of Big Data, Traditional vs. Big Data approach **Hadoop:**What is Hadoop? Core Hadoop Components, Hadoop Ecosystem (Hbase, Hive, Hcatalog, Pig, Sqoop, Oozie, Mahout, ZooKeeper), Physical Architecture, Hadoop limitations. **NoSQL:** What is NoSQL? NoSQL business drivers, NoSQL case studies (Amazon Dynamo DB, Google's Big Table, MongoDb, Neo4j), NoSQL data architecture patterns (Key-value stores, Graph stores, Column family (Bigtable) stores, Document stores), Variations of NoSQL architectural patterns; Using NoSQL to manage big data **Map Reduce:** MapReduce and New Software stack (Distributed File Systems, Physical Organization of Compute Nodes), The Map Tasks, Grouping by Key, The Reduce Tasks, Combiners, Details of MapReduce Execution, Coping With Node Failures	9	25
2	**Finding Similar items:** Nearest neighbor Search, Applications of Near-Neighbor Search, Similarity of documents (Plagiarism detection, Document clustering, News aggregation), Collaborative Filtering as a Similar-Sets Problem, Recommendations based on user ratings, Distance Measures(Definition of a Distance Measure , Euclidean Distances, Jaccard Distance, Cosine Distance, Edit Distance, Hamming Distance)	9	25
3	**Data Stream Mining:** The Stream Data Model: A Data-Stream-Management System, Examples of Stream Sources, Stream Queries, Issues in Stream Processing. Sampling Data in a Stream, Filtering Streams (The Bloom Filter), Counting Distinct Elements in a Stream, Counting Ones in a Window, Decaying Windows.	6	15

4	**Link Analysis:** PageRank, Efficient Computation of Page Rank (PageRank implementation using MapReduce, Use of Combiners to Consolidate the Result Vector), Topic sensitive Page Rank, link Spam, Hubs and Authorities.	3	10
5	**Recommendation Systems and Social Network Graphs:** **Recommendation Systems:** A Model for Recommendation Systems, Content-Based Recommendations, Collaborative Filtering. **Social Network Graphs:** Application of social network mining, Social network as a graph, Types of social networks, Clustering of social Graphs, Direct discovery of communities in a social Graph, SimRank	9	25

Text Books:

1. Radha Shankarmani, M Vijayalakshmi, Big Data Analytics, 2nd Edition, Wiley
2. Jure Leskovec, AnandRajaraman, Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, Second Edition, 2014.

Reference Books:

1. Seema Acharya, Subhashini Chhellappan, BIG Data and Analytics ,Willey
2. Jiawei Han, Micheline Kamber, Jian Pei, "Data Mining Concepts and Techniques", Morgan Kaufman Publications, Third Edition, 2011.
3. VigneshPrajapati, "Big Data Analytics with R and Haoop", Packet Publishing 2013
4. Big Data Black Book, DreamTech

List of Open Source Software/learning website::

1. http://www.bigdatauniversity.com
2. http://www.mongodb.com
3. http://hadoop.apache.org

Course Learning Outcomes (CLO): On completion of this course, the students will be able to:

CLO	Description	Bloom's Taxonomy Level
CLO1	Understand the importance of Big Data and Big Data Tools for solving real world problems	1 Remembering 2 Understanding
CLO2	Understand the mechanism of fundamental techniques used for the Big Data Analytics.	1 Remembering 2 Understanding
CLO3	Analyze the Big Data framework like Hadoop and NOSQL to efficiently store and process Big Data to generate analytics.	4. Analyze 6. Creating
CLO4	Design of Algorithms to solve Data Intensive Problems using Map Reduce Paradigm	3. Applying 4. Analyze 6. Creating
CLO5	Access and Process Data on Distributed File System	1 Remembering 2 Understanding 3. Applying
CLO6	Understand the implementation of Data Stream Mining, Link Analysis, Recommendation System & Social Network Graphs	1 Remembering 2 Understanding 4. Analyze 5. Evaluate

FACULTY OF COMPUTER SCIENCE

Master of Computer Application (Sem-II)

In Effect from Academic Year 2023-24

Mapping of CLOs with POs & PSOs

Course Learning Outcomes	Program Outcomes (POs)												Program Specific Outcomes(PSOs)	
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CLO1	H	L	L				L				L		H	M
CLO2	H	M	L			L		M		H				M
CLO3		H			M	H		M				M	M	M
CLO4		H	H	H		M		M	M		M	L	H	M
CLO5		M	M	M	H	M	H	M	H				M	M
CLO6	H	M	L			L		M		H	L		M	M

H: High, M: Medium, L: Low