Course Code	Course Title	L	T	P	C
BCSE402L	Big Data Analytics	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To introduce the fundamental concepts and importance of big data analytics, emphasizing its relevance in various domains.
2. To equip students with the necessary skills and tools to effectively manage and analyze large-scale data sets, including hands-on experience on relevant technologies.
3. To illustrate the practical application of big data analytics methods and techniques in solving strategic business problems, showcasing real-world examples and case studies.

Course Outcomes

1. Recall the characteristics of digital data, data sources, data storage and the applications of big data in different fields.
2. Utilize Hadoop ecosystem tools and Hadoop YARN functions for parallel processing of application tasks.
3. Comprehend the Map Reduce programming model and the Map Reduce Daemon framework.
4. Apply NoSQL databases for data store management to solve big data problems
5. Analyze and evaluate the use of spark stack components with RDDs, ETL built-in functions for handling big data.

Module:1	Overview of BigData Analytics	5 hours

Introduction - Need of BigData – BigData : Definitions - Characteristics – Evolution - Challenges - Scalability and Parallel processing - Classification of Analytics - Data Storage and Analysis – Use cases of BigData Applications

Module:2	Hadoop for Big data	7 hours

Hadoop and Ecosystem core components – Features, Streaming, pipes Interacting with Hadoop Ecosystem, HDFS: The Design of HDFS- HDFS Concepts - Blocks – Name nodes and Data nodes; Processing Data with Hadoop - Basic File system Operations, Hadoop File systems - Interfaces – I/O - Managing Resources and Applications with Hadoop YARN - Hadoop ecosystem - Hive : Data Types – HQL - Pig : Grunt Shell - Pig Latin data Model & Data Scripts

Module:3	Map Reduce	6 hours

MapReduce Framework - Programming Model – Map Reduce: Map Tasks - Key value pair - Reduce Tasks - Grouping by key – Partitioning – Combiners – Reduce Tasks - MapReduce Execution - Composing Map Reduce for Calculations and Algorithms

Module:4	NoSQL Big Data Management	5 hours

NoSQL Data Store – Data Architecture Patterns – Mongo DB : Data Types - Query Languages - Database commands – Cassandra : CQL Data Types – CRUD – Import and Export - HBase

Module:5	Spark for Big Data Analytics	6 hours

Introduction to Data Analysis with Spark – Functional Programming Basics - Parallel Programming using Resilient Distributed Datasets - Spark SQL – Data Analysis Operations – Spark RDD - Characteristics - Transform and Action Commands - Data Frame Operations – Spark for ETL – Analytics Reporting and Data Visualization		
Module:6	**Data Stream and Real-Time Analytics**	**7 hours**
Data Stream – Concepts & Data Stream Management - Stream Computing Aspects : Sampling , Filtering & Counting Distinct Elements in Streams – Estimating Moments - Frequent Item sets – Handling Larger Datasets for Finding Frequent Item sets – Limited Passes Algorithms		
Module:7	**Graph and Social Network Analytics**	**7 hours**
Graph Model - Representing Graph as Triples – RDF for Graph Databases - SPARQL – Network Organization and Graph Analytics – Social Network Graph Analysis – Topological – Centralities – K-Core – Clustering - Ranking - Counting and Graph matches		
Module:8	**Recent Trends**	**2 hours**
Guest lectures from Industry and, Research and Development Organizations		
	Total Lecture hours:	**45 hours**

Text Book(s)	
1.	Raj Kamal, PreetiSaxena, "Big Data Analytics: Introduction to Hadoop, Spark, and Machine-Learning", 2019, 1st Edition, McGraw Hill.
Reference Books	
1.	Sayan Goswami, Amit Kumar Das, Sourabh Mukherjee, "Big Data Simplified", 2019, 1st Edition, Pearson Education.
2	Subhashini Chellappan, Seema Acharya, "Big Data and Analytics", 2019, 2nd Edition, Wiley.
3.	Tom White,"Hadoop: The Definitive Guide",2009, O'Reilly Media, Inc.
4.	Jure Leskovec, Anand Rajaraman, Jeff Ullman, "Mining of Massive Datasets", 2020, 3rd Edition, Cambridge University Pres.
Mode of Evaluation: CAT, Assignments, Quiz, FAT	

Recommended by Board of Studies	12-05-2023		
Approved by Academic Council	No. 70	Date	24-06-2023