Name of the Programme : **M.Sc. in Data Science**
Course Code : **CSD-526**
Title of the Course : **Big Data Analytics**
Number of Credits : **4(4L)**
Contact Hours : **60 hours (60L-0T-0P)**
Effective from AY : **2023-24**

Pre-requisites for the Course	Programming Language, SQL queries, and exposure to Linux Environment.	
Course Objectives:	The course objective is to equip students with a comprehensive understanding of Big Data platforms, with a specific focus on Apache Hadoop and its ecosystem.	
Content: Theory	**UNIT I: INTRODUCTION TO BIG DATA AND HADOOP** Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, Analysing Data with Unix tools, Analysing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big Data Strategy, Introduction to Infosphere BigInsights and Big Sheets.	**15 hours**
	UNIT II: HDFS(Hadoop Distributed File System) The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Data Ingest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures.	**15 hours**
	UNIT III: Map Reduce Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle, and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.	**15 hours**
	Unit IV: Hadoop Eco System Pig: Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators. Hive: Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data, and User-Defined Functions. Hbase: HBasics, Concepts, Clients, Example, Hbase Versus RDBMS. Big SQL: Introduction	**15 hours**
Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Self-study/Flipped classroom	
References/Readings	1. Franks, B. (2012). Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics. John Wiley & Sons. 2. Liebowitz, J. (Ed.). (2013). Big data and business analytics. CRC press. 3. Warden, P. (2011). Big data glossary. O'Reilly Media, Inc..	

Course Outcomes	Upon completion of the course, learners will be able to: 1. Develop an understanding of the principles, concepts, and technologies underlying big data analytics. 2. Acquire skills in processing and transforming large datasets using distributed computing frameworks like Apache Spark, enabling parallel and scalable data processing. 3. Apply machine learning algorithms to big data 4. Analyze case studies and real-world applications of big data analytic

(Back to Index)