NIRMA UNIVERSITY

Institute:	Institute of Technology, School of Technology
Name of Programme:	M Tech CSE, MTech CSE (Cyber Security) / MTech CSE (Data Science)
Course Code:	6CS271ME25
Course Title:	Big Data Systems
Course Type:	Department Elective-I / Core
Year of Introduction:	2025-26

Credit Scheme

L	T	Practical Component				C
		LPW	PW	W	S	
3	0	2	-	-	-	4

Course Learning Outcomes (CLO):

At the end of the course, the students will be able to –

1. outline the significance, challenges, and storage platforms of Big Data in modern computing (BL2)
2. model Big Data applications using Hadoop, NoSQL databases, and distributed computing platforms (BL3)
3. utilize big Data processing frameworks like Hadoop, Pig, and Hive for practical business analytics (BL3)
4. evaluate the performance of Hadoop applications by analyzing job execution, scheduling, and data processing efficiency. (BL5)

Unit	Contents	Teaching Hours (Total 45)
Unit-I	**Introduction to Big Data and Big Data Storage Platforms**: Evolution of Big Data, Types of Big Data, Definition of Big Data, Importance of Big data analytics, Challenges of Conventional Systems, Big data platforms and data storage	06
Unit-II	**Hadoop and HDFS:** Hadoop Ecosystem, Comparisons of RDBMS and Hadoop, Distributed Computing Challenges, Hadoop Overview, Hadoop Ecosystem, Hadoop Environment and setup, Hadoop Configuration and administration, Processing Data with Hadoop, Hadoop YARN, Hadoop in the Cloud	06
Unit-III	**MapReduce:** working with 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	10
Unit-IV	**NoSQL Databases:** Introduction to NoSQL Databases, Types of NoSQL databases, SQL Vs. NoSQL, Why NoSQL, Introduction to the Document Database (MongoDB or similar), Data Types and CRUD operations in Document Database, Introduction to the Graph Database (Neo4j or similar), CRUD operations in Graph Database, Relevant Case Studies	14

Unit-V	**Introduction to other frameworks:** Data Processing Operators in Pig, HiveQL, Querying Data in Hive, Applications on Big Data using Pig and Hive, Fundamentals of HBase and ZooKeeper.	09

Self-Study:
The self-study contents will be declared at the commencement of the semester. Around 10% of the questions will be asked from self-study content.

Suggested Readings/ References:

1. Bill Chambers and Matei Zaharis, Spark: The Definitive Guide: Big Data Processing Made Simple, O'Reilly
2. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer
3. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill
4. Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press
5. Seema Acharya and Subhashini C, Big Data and Analytics, Wiley
6. Dan Mccreary, Ann Kelly, Making Sense of NoSQL Manning Publication dream tech press
7. Paul Zikopoulos, Dirk deRoos, Krishnan Parasuraman, Thomas Deutsch, James Giles, David Corrigan, Harness the Power of Big Data The IBM Big Data Platform, Tata McGraw Hill Publications
8. Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press.

Suggested List of Experiments:

Sr. No.	Name of Experiments/Exercises	Hours
1	Study and explore various applications of big data in different domains. Choose one of them and study in detail. Also, write down the report on different types of digital data generated in selected applications. For example.: • Big Data in Retail • Big Data in Healthcare • Big Data in Education • Big Data in E-commerce • Big Data in Media and Entertainment • Big Data in Finance • Big Data in Travel Industry • Big Data in Telecom	04
2	Set up a single-node Hadoop cluster and apply HDFS commands to the single-node Hadoop Cluster	04
3	a) Design MapReduce algorithms to take a very large file of integers and produce as output: • The largest integer • The average of all the integers. • The same set of integers, but with each integer appearing only once.	06

	The count of the number of distinct integers in the input b) Analyze the impact of different numbers of mappers and reducers on the same definition as practical 3(a). Prepare a conclusive report on the analysis	
4	Set up the MongoDB environment in your system. Import restaurant dataset and perform CRUD operation	04
5	To perform CRUD operations on Graph Database (Neo4j or similar)	04
6	To study and prepare a report on Big Data Analytics and Cloud	04
7	To load an employee dataset into Apache Pig, filter employees earning above ₹50,000 and store the results in HDFS. Then, using Apache Hive, create a table, load the same dataset, and write a query to find the average salary per department, analyzing structured data efficiently.	04