

## NIRMA UNIVERSITY

<b>Institute:</b>	<b>Institute of Technology, School of Technology</b>
<b>Name of Programme:</b>	<b>BTech CSE</b>
<b>Course Code:</b>	<b>4CS503ME25</b>
<b>Course Title:</b>	<b>Programming for Modern Databases</b>
<b>Course Type:</b>	<b>Department Elective III</b>
<b>Year of Introduction:</b>	<b>2025-26</b>

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. summarise the concept of modern databases (BL2)
2. model modern databases using different tools and frameworks (BL3)
3. apply big data techniques for useful business analytic applications (BL4)
4. design algorithms for mining the data from large volumes. (BL6)

Unit	Contents	Teaching Hours (Total 45)
Unit-I	<b>Introduction to Big Data:</b> Evolution of Modern Databases, Definition of Big Data, Types of Digital Data, Classification of Digital Data, Structured Data, Semi-Structured Data, Unstructured Data, Challenges of Conventional Systems, Big data platforms and data storage, Importance of Big data analytics, Classification of Analytics. Top Challenges Facing Big Data, Terminologies Used in Big Data Environment	08
Unit-II	<b>Hadoop Architecture:</b> Introducing Hadoop and comparisons of RDBMS and Hadoop. Distributed Computing Challenges, Hadoop Overview, Business Value of Hadoop. Hadoop Distributed File System, Processing Data with Hadoop, working with Map Reduce, Hadoop YARN, Hadoop in the Cloud, Applications of Hadoop Ecosystem, Fundamentals of Pig, Hive, HBase, and ZooKeeper, Basic concepts of Apache Spark	10
Unit-III	<b>Processing and Storing Streaming Data:</b> Data Integration and ETL (Extract, Transform, Load) Processes, CAP Theorem - BASE Concept, NoSQL, Types of No SQL databases, Introduction to MongoDB, Data Types in MongoDB, CRUD, Apache Cassandra, Features of Cassandra and CRUD operations, Visualizing Data	09
Unit-IV	<b>Machine Learning for Modern Data:</b> Applying Linear Regression, Clustering, Association rule mining, Decision tree on modern data, Single-Layer Network, Multi-layer Perceptron.	08

Unit-V	<b>Database Security and Optimization:</b> Best Practices for Database security, Authentication and authorization in database systems, Preventing SQL injection and other threats, Indexing and query optimization techniques, Performance tuning and monitoring.	10
--------	---	----

### 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 Franks, *Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*, John Wiley & sons
2. Pete Warden, *Big Data Glossary*, O'Reilly
3. J. Han & M. Kamber, *Data Mining Concepts and Techniques*, Elsevier
4. Glenn J. Myatt, *Making Sense of Data*, John Wiley & Sons
5. Witten & Frank, *Data Mining Practical Machine Learning Tools and Techniques*, Morgan Kaufmann
6. J A Rice, *Mathematical Statistics and Data Analysis*, Wadsworth
7. D.D. Wackerly, W. Mendenhall and R.L. Scheaffer, *Mathematical Statistics with Applications*, Brooks/Cole
8. Seema Acharya and Subhashini C, *Big Data and Analytics*, Wiley

### Suggested List of Experiments:

Sr. No.	Name of Experiments/Exercises	Hours
1	Study and explore various applications using modern databases in different domains. Prepare a study report for any selected application, which includes data sources, analytics visualization algorithms, etc. The domain may be selected from the following: <ul style="list-style-type: none"> <li>• Big Data in Retail</li> <li>• Big Data in Healthcare</li> <li>• Big Data in Education</li> <li>• Big Data in E-commerce</li> <li>• Big Data in Media and Entertainment</li> <li>• Big Data in Finance</li> <li>• Big Data in Travel Industry</li> <li>• Big Data in Telecom</li> <li>• Big Data in Automobile</li> </ul>	04
2	Identify the data sources for big data. Find the technological limitations of conventional data analysis algorithms when performing analytics on big data. Justify your answer with any one of the applications.	02
3	Perform ETL (Extract, Transform, Load) with external systems for selected domains in practical 1.	04
4	Install and configure single node Hadoop cluster. Perform HDFS commands on single node Hadoop Cluster. Design and implement MapReduce program to find phrase frequency from the given dataset.	04

5	Design and implement the MapReduce program to find the sum of given values.	02
6	Design and implement a k-means clustering algorithm using a MapReduce programming model.	04
7	Install and configure Pig. Perform CRUD operations on the given dataset.	02
8	Install and configure MongoDB. Perform CRUD operations on the given dataset.	02
9	Install and configure Cassandra. Perform CRUD operations on the given dataset.	02
10	Install and configure Spark. Analyze the performance of Spark in comparison with Hadoop using a given example scenario.	04