CSE11202	Unstructured Database	L	T	P	C
Version 1.0		3	0	0	3
Pre-requisite/Exposure	RDMS Concepts, Programming Concepts				
Co-requisite	Logical Ability				

Course Objectives:

1. Understand the Fundamentals of Unstructured Databases: To comprehend the challenges associated with unstructured data and explore various data models suitable for unstructured data storage and manipulation.
2. Develop Skills in Data Handling: To gain hands-on experience in data ingestion, extraction, and querying of unstructured data using modern tools and technologies, with a special focus on MongoDB.
3. Master MongoDB Operations: To understand and utilize MongoDB for creating, managing, and operating on databases, including advanced operations like aggregation, indexing, and replication.
4. Implement Disaster Recovery and Backup Strategies: To develop strategies for backup, recovery, and disaster planning specifically tailored to unstructured databases in various deployment scenarios.

Course Outcomes:

CO1: Distinguish between Different Types of Data Models: Identify and choose appropriate data models for specific use cases involving unstructured data.

CO2: Manage Data Ingestion and Extraction Processes: Effectively handle the ingestion and extraction of unstructured data, and perform data cleaning and pre-processing.

CO3: Utilize MongoDB for Database Operations: Install, configure, and manage MongoDB, and perform CRUD operations efficiently using MongoDB's query language.

CO4: Apply Advanced MongoDB Features: Use advanced MongoDB features such as text search, indexing, MapReduce, and replication to optimize database performance and scalability.

CO5: Design and Implement Backup and Recovery Plans: Create and execute effective backup and disaster recovery plans that ensure data durability and high availability in production environments.

Course Description:

This course offers a comprehensive examination of unstructured databases with a focus on techniques and technologies that handle unstructured data. Students will gain a deep understanding of various data models including document-based, graph-based, key-value, and wide column models. The course covers data ingestion, extraction, and querying of unstructured data, with a particular emphasis on using MongoDB. Advanced features such as MongoDB's indexing, MapReduce, and integration capabilities with other database systems are also explored. Practical skills in MongoDB installation, querying, and database operations are developed, alongside concepts of data backup, recovery, and disaster planning.

Course Content:

Units	**Lecture Hours**
Unit-I: INTRODUCTION TO UNSTRUCTURED DATABASE Overview of unstructured data and its challenges, Introduction to unstructured database, Comparison with structured and semi-structured DBMS, Use cases and benefits of unstructured DBMS, Understanding different data models for unstructured databases, Document-based models (e.g., JSON, XML), Graph-based models, Key-value models, Wide column models, Choosing the appropriate data model for specific use cases.	**7**
Unit-II: DATA INGESTION, EXTRACTION & QUERYING UNSTRUCTURED DATA Techniques for ingesting unstructured data into the database, Extracting structured information from unstructured data, Pre-processing and cleaning unstructured data, Text extraction, parsing, and entity recognition, Handling multimedia and binary data, Introduction to query languages for unstructured databases.	**7**
Unit-III: INTRODUCTION TO MONGODB Overview of MongoDB and its key features, Comparison with traditional relational databases, Understanding NoSQL and document-oriented databases, Installing and setting up MongoDB, MongoDB Data Model, Create Database, Drop Database, Create Collection, Drop Collection, Datatypes, Insert Document, Query Document, Update Document, Delete Document, Projection, Limit Records, Sort Records, Indexing, Aggregation, Replication, Sharding.	**12**

Unit IV: ADVANCED MONGODB: Relationships, Database References, Covered Queries, Analysing Queries, Atomic Operations, Advanced Indexing, Indexing Limitations, ObjectId, MapReduce, Text Search Regular Expression, Studio 3T, GridFS, Capped Collections, Auto-Increment Sequence, MongoDB-Python (Flask) Connection and CRUD Operation	**12**
Unit V: BACKUP, RECOVERY, DISASTER PLANNING AND EMERGING TREDNDS: Backup and recovery procedures for unstructured DBMS, Disaster recovery planning for unstructured databases, MongoDB in Production, Deployment considerations and best practices, Backup and recovery strategies, Advanced features and capabilities of unstructured DBMS, Integration with other database systems and data sources, Handling streaming and real-time data, Exploring cloud-based unstructured DB.	**7**