

# CBCS SCHEME



--	--	--	--	--	--	--	--	--	--

MMCA311B

**Third Semester MCA Degree Examination, Dec.2025/Jan.2026**  
**Big Data Analytics**

Time: 3 hrs.

Max. Marks: 100

Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.  
 2. M : Marks , L: Bloom's level , C: Course outcomes.

Module - 1				M	L	C
Q.1	a.	Define Big Data and explain its evolution. Discuss the characteristics of Big Data with suitable examples.	10	L2	CO1	
	b.	Differentiate between traditional data processing systems and Big Data systems. Explain why traditional systems are inadequate for handling Big Data.	10	L2	CO1	
OR						
Q.2	a.	Explain the Hadoop ecosystem. Describe the architecture and components of Hadoop, including HDFS, YARN, and MapReduce.	10	L2	CO1	
	b.	Discuss the limitations of Hadoop. Explain the reasons for the shift from Hadoop MapReduce to Apache Spark.	10	L2	CO1	
Module - 2						
Q.3	a.	Describe the design and operations of Hadoop Distributed File System (HDFS). Describe the roles of NameNode, DataNode, and Secondary NameNode.	10	L2	CO2	
	b.	Discuss the Hadoop MapReduce programming model and also explain the job execution flow of a MapReduce program.	10	L2	CO2	
OR						
Q.4	a.	Explain the steps involved in writing MapReduce programs with reference to Word Count and Sorting applications.	10	L2	CO2	
	b.	Explain advanced Hadoop features such as Combiners, Partitioners and Counters. Discuss Hadoop streaming and its integration with Python.	10	L2	CO2	
Module - 3						
Q.5	a.	Explain in-memory big data processing using Apache Spark. Describe the Spark architecture and its main components.	10	L2	CO3	
	b.	Explain Resilient Distributed Datasets (RDDs). Discuss RDD transformations and actions with examples.	10	L2	CO3	

MMCA311B

OR

Q.6	a.	Explain DataFrames and Spark SQL. Describe their role in structured data processing.	10	L2	CO3	
	b.	Explain Spark MLlib and its role in machine learning. Discuss PySpark and the steps involved in running Spark jobs using Python.	10	L2	CO3	
Module - 4						
Q.7	a.	Explain the need for NoSQL databases. Discuss the limitations of traditional RDBMS in handling Big Data.	10	L2	CO4	
	b.	Explain the types of NoSQL databases Key-Value, Document, Column, and Graph with suitable examples.	10	L2	CO4	
OR						
Q.8	a.	Explain HBase architecture. Describe the CRUD operations performed in HBase.	10	L2	CO4	
	b.	Explain Cassandra and MongoDB. Discuss data modelling techniques used for achieving scalability and performance in Big Data systems.	10	L2	CO4	
Module - 5						
Q.9	a.	Explain the role of Big Data tools such as Hive, Pig, Sqoop, Flume, and Oozie in the Hadoop ecosystem.	10	L2	CO5	
	b.	Explain data ingestion in Big Data systems. Describe the working of Flume and Sqoop with suitable examples.	10	L2	CO5	
OR						
Q.10	a.	Discuss real-world applications of Big Data in healthcare, finance, e-commerce, IoT, and social media.	10	L2	CO5	
	b.	Explain real-time analytics using Kafka and Spark Streaming. Discuss the ethical issues and challenges in Big Data analytics.	10	L2	CO5	

CMRIT LIBRARY  
BANGALORE - 560 037

\*\*\*\*\*