

USN									
-----	--	--	--	--	--	--	--	--	--

Course Code	22CD61
-------------	--------

Sixth Semester B.E. Degree Examinations, May/June 2026

BIG DATA ANALYTICS

(CSE- DS)

Duration: 3 hrs

Max. Marks: 100

Note: 1. Answer any FIVE full questions choosing ONE full Question from each Module.
 2. Missing data, if any, may be suitably assumed

<u>Q. No</u>	<u>Question</u>	<u>Marks</u>	<u>(RBTl:CO: PI)</u>
<u>Module-1</u>			
1.	a. Discuss the role of Massive Parallel Processing (MPP) platforms in big data analytics.	10	(2 :1: 1.2.1)
	b. Explain the Berkeley data analytics stack and how it simplifies complex analytics tasks.	10	(2 :1: 1.2.1)
(OR)			
2.	a. List and describe different data sources (Internal vs. External) used in modern enterprises.	10	(2 :1: 1.2.1)
	b. Highlight the importance of data storage and analysis strategies for high-velocity data.	10	(2 :1: 1.2.1)
<u>Module-2</u>			
3.	a. What is HDFS safe mode? Explain the conditions under which a name node enters and exits this state.	10	(3:2:1.2.1)
	b. Describe the Apache Sqoop export method with a neat diagram, showing data movement to RDBMS.	10	(2:2:1.2.1)
(OR)			
4.	a. Explain the Hadoop-2 YARN model and its improvements over the traditional MapReduce framework.	10	(3:2:1.2.1)
	b. Demonstrate any six HBase shell commands used for data manipulation and schema definition.	10	(2:2:1.2.1)
<u>Module-3</u>			
5.	a. Explain the CAP theorem and how NoSQL databases choose between consistency, availability, and partition tolerance.	10	(2:3:1.2.1)
	b. Describe the shared-nothing architecture specifically for Sharding very large databases.	10	(2:3:1.2.1)
(OR)			
6.	a. Discuss about the features of Big Table and RC/ORC data stores. Which is preferred for analytical queries?	10	(2:3:1.2.1)

Note: (RBTl - Revised Bloom's Taxonomy Level: CO - Course Outcome: PI- Performance Indicator)

- b. Discuss the usage of MongoDB and Riak in industrial applications with specific use-case examples. 10 (2:3:1.2.1)

Module-4

7. a. Explain the Hive architecture and its characteristics as a data warehousing tool on top of Hadoop. 10 (2:4:2.5.2)
- b. Demonstrate the Pig architecture and how it handles data flow using Pig Latin scripts. 10 (3:4:1.2.1)

(OR)

8. a. Using HiveQL, provide the syntax for DDL operations to add and rename partitions in a table. 10 (2:4:2.5.2)
- b. Discuss about Pig and MapReduce in terms of development speed and execution efficiency. 10 (3:4:1.2.1)

Module-5

9. a. Explain the purpose of web usage analytics and the significance of web graphs in identifying user behavior. 10 (3:5:1.2.1)
- b. Discuss how Cliques are used to discover specific communities within a social network graph. 10 (2:5:1.2.1)

(OR)

10. a. Discuss how to design a predictive analytics framework that integrates both regression and clustering techniques. 10 (3:5:1.2.1)
- b. Write a short note on web content mining and its role in extracting knowledge from unstructured web data. 10 (2:5:1.2.1)

** ** *