Basavarajeswari Group of Institutions

2022 SCHEME

BALLARI INSTITUTE OF TECHNOLOGY & MANAGEMENT

(Autonomous Institute under Visvesvaraya Technological University, Belagavi)

USN										

Course Code | 2 | 2 | C | D | 6 | 1 |

Sixth Semester B.E. Degree Examinations, June/July 2025

BIG DATA ANALYTICS

(CSE- Data Science)

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

Q. No		*Question*	*Marks*	*(RBTL:CO: PI)*
		Module-1		
1.	**a.**	Explain the concept of big data and how it differs from traditional data.	**07**	**(2 :1: 1.2.1)**
	b.	Classify different sources of big data with examples.	**07**	**(2 :1: 1.2.1)**
	c.	Compare structured, semi-structured, and unstructured data types.	**06**	**(2 :1: 1.2.1)**
		(OR)		
2.	**a.**	List the features of grid computing. How does it differ from cluster and cloud computing.	**06**	**(2 :1: 1.2.1)**
	b.	Explain the role and functionality of each of the five layers in the design of big data architecture.	**08**	**(2 :1: 1.2.1)**
	c.	Explain what big query is in google cloud. Add a clear diagram to show how it works in the cloud.	**06**	**(2 :1: 1.2.1)**
		Module-2		
3.	**a.**	Illustrate the components of the Hadoop ecosystem.	**06**	**(3:2:1.2.1)**
	b.	Summarize the working of the Hadoop Distributed File System (HDFS).	**08**	**(2:2:1.2.1)**
	c.	Interpret the steps of MapReduce in Hadoop data processing.	**06**	**(2:2:1.2.1)**
		(OR)		
4.	**a.**	Explain the role of Hadoop YARN in resource management with neat sketch.	**07**	**(2:2:1.2.1)**
	b.	Compare the features of Hive and HBase in the Hadoop ecosystem.	**06**	**(2:2:1.2.1)**
	c.	Demonstrate the use of HDFS commands with suitable examples.	**07**	**(2:2:1.2.1)**
		Module-3		
5.	**a.**	Explain the concept of NoSQL databases and how they differ from traditional relational databases.	**06**	**(2:3:1.2.1)**
	b.	Illustrate document datastore with example.	**07**	**(2:3:1.2.1)**
	c.	Explain four aggregated data models (Schemaless) of NOSQL.	**07**	**(2:3:1.2.1)**
		(OR)		
6.	**a.**	Make use of MongoDB query language and database commands to list the functions of each with example.	**07**	**(3:3:1.2.1)**

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

	b.	List features of MongoDB and outline the replication and AutoSharding.	**06**	**(2:3:1.2.1)**
	c.	Make use of Casandra Query Language commands to list the function of each with example.	**07**	**(3:3:1.2.1)**

Module-4

7.	**a.**	Illustrate dataflow sequence and wokflow steps between Hive and Hadoop with neat sketch.	**07**	**(2:4:2.5.2)**
	b.	Show MapReduce process diagrammatically to depict a client submitting a job, the workflow and creating the outputs.	**07**	**(3:4:1.2.1)**
	c.	Explain components of Hive architecture with neat sketch.	**06**	**(2:4:1.2.1)**

(OR)

8.	**a.**	Demonstrate the use of the LOAD, STORE, and DUMP commands in data processing workflow.	**06**	**(2:4:1.2.1)**
	b.	Illustrate how do you create partitions and buckets in a Hive database.	**07**	**(2:4:2.5.2)**
	c.	Summarize all the Pig data types and provide an example for each.	**07**	**(2:4:1.2.1)**

Module-5

9.	**a.**	Describe the regression analysis to predict the value of the dependent variable in case of linear regression.	**10**	**(3:5:1.2.1)**
	b.	Discuss three phases for web phase mining.	**10**	**(2:5:1.2.1)**

(OR)

10.	**a.**	Illustrate the phases of the text mining process.	**10**	**(2:5:2.5.2)**
	b.	Make use of equations for computing PageRank using the in-degree as conferring authority and the relative authority of the parents over linked children	**10**	**(2:5:2.5.2)**

** ** **

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