CBCS SCHEME

USN | | | | | | | | | | |

21CS71

Seventh Semester B.E./B.Tech. Degree Examination, June/July 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. Describe characteristics and types of Big Data.	10	L2	CO1
	b.	What is Web data? Explain how do you classify data as structured, semistructured, multistructured and unstructured data? Give suitable examples.	10	L2	CO1
		OR			
Q.2	a.	With a neat diagram, explain the function of each of the five layers in big data architecture design.	10	L2	CO1
	b.	Write short notes on Analytics scalability to Big data and Massive Parallel processing platforms.	10	L2	CO1
		Module – 2			
Q.3	a.	With a neat diagram, explain Hadoop main components and ecosystem components.	10	L2	CO2
	b.	Brief out the features of Hadoop and explain HDFS commands with example.	10	L2	CO2
		OR			
Q.4	a.	Explain in detail about Hadoop MapReduce-Frame work with programming model.	10	L2	CO2
	b.	Describe Hadoop YARN and YARN Based Execution Model.	10	L2	CO2
		Module – 3			
Q.5	a.	Discuss NoSQL data architecture pattern with an example.	10	L2	CO3
	b.	Compare and contrast RDBMS and Mongo DB databases.	5	L2	CO3
	c.	What are the different ways of handling Big data problems?	5	L2	CO3
		OR			
Q.6	a.	Discuss the usage of Apache 'S' HBase, MongoDB, Cassandra and Apache 'S' CouchDB.	8	L2	CO3

	b.	With neat diagram, explain the following Shared-Nothing Architecture for Big data tasks : (i) Single server Model (ii) Sharding very large database. (iii) Master Slave distribution Model. (iv) Peer-to-peer distribution Model.	**12**	**L2**	**CO3**
		Module – 4			
Q.7	**a.**	Describe MapReduce execution steps with a neat sketch.	**10**	**L2**	**CO4**
	b.	Define HIVE. List out characteristics and explain Hive architecture in detail.	**10**	**L2**	**CO4**
		OR			
Q.8	**a.**	Demonstrate the pig architecture for scripts data flow and processing.	**10**	**L2**	**CO4**
	b.	Explain with respect to HIVE QL : (i) HIVE QL Data Definition Language (DDL) (ii) HIVE QL Data Manipulation Language (DML)	**10**	**L3**	**CO4**
		Module – 5			
Q.9	**a.**	Explain with an example, Apriori algorithm to evaluate candidate key.	**10**	**L3**	**CO5**
	b.	What are Outliers? Describe the reasons for the presence of outliers in a relationship.	**5**	**L2**	**CO5**
	c.	In Machine Learning explain ANOVA.	**5**	**L2**	**CO5**
		OR			
Q.10	**a.**	Discuss the different phases of text mining process.	**10**	**L2**	**CO5**
	b.	Briefly explain process of Web Usage Mining.	**10**	**L2**	**CO5**

* * * * *