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

21CST701

B. E. Degree (Autonomous) Seventh Semester End Examination (SEE), Jan 2025

BIG DATA ANALYTICS

(Model Question Paper - III)

Time: 3 Hours] **[Maximum Marks: 100**

Instructions to students

1. Answer FIVE FULL Questions.

Q. No.	Questions	Marks	Course Outcome	RBT Level
1 (a)	Explain with examples the different types of Big Data	**(10 Marks)**	**CO1**	**L2**
1 (b)	Define Big Data. Explain 8 characteristics of Big Data.	**(10 Marks)**	**CO1**	**L2**
	OR			
2 (a)	What is Hadoop? Explain the key components in typical Hadoop environment.	**(10Marks)**	**CO2**	**L3**
2 (b)	Discuss the importance of big data. Illustrate with a neat diagram the coexistence of big data and data warehouses.	**(10 Marks)**	**CO2**	**L2**
3 (a)	Compare Hadoop with traditional SQL databases	**(10 Marks)**	**CO2**	**L2**
3 (b)	How is the Reducer class implemented in a MapReduce Java program? Explain with a code example.	**(10 Marks)**	**CO4**	**L2**
	OR			
4 (a)	What is Hadoop YARN, and how does it manage resources and applications in a Hadoop cluster?	**(10 Marks)**	**CO2**	**L2**
4 (b)	Illustrate with a word count example the working of map reduce concept.	**(10 Marks)**	**CO4**	**L3**
5 (a)	What is MapReduce. Explain in detail different phases in MapReduce.	**(12 Marks)**	**CO3**	**L2**
5 (b)	Write a user define partitioner class for WordCount problem.	**(08 Marks)**	**CO4**	**L3**
	OR			
6 (a)	Explain CAP Theorem in Big Data Systems	**(10 Marks)**	**CO3**	**L2**
6 (b)	Explain different Big Data Analytics Approaches.	**(6 Marks)**	**CO2**	**L2**
6(c)	Discuss parallel and distributed systems in big data environment.	**(4 Marks)**	**CO2**	**L2**
7 (a)	Explain User Defined Functions (UDFs) in Hive. Describe their purpose. Write a Hive function to convert the values of a field to uppercase.	**(10 Marks)**	**CO4**	**L3**

7 (b)	Explain the RCFile format in Hive and its benefits. Create a table in Hive using the RCFile format and load sample data into it. Write a query to retrieve specific data from this table.	**(10 Marks)**	**CO4**	**L3**
OR				
8 (a)	Describe how GROUP and JOIN operators work in Pig Latin and provide an example where both are used together to calculate the average score of students by course.	**(10 Marks)**	**CO4**	**L3**
8 (b)	Explain the following EVAL Functions with examples: AVG, MAX and COUNT.	**(10 Marks)**	**CO3**	**L2**
9 (a)	Explain the following terms with respect to spark application a. Spark Application b. Spark Session c. Job d. Stage e. Task	**(10 Marks)**	**CO3**	**L2**
9 (b)	Compare Hadoop and Spark. Explain the key features of Apache spark.	**(10 Marks)**	**CO2**	**L2**
OR				
10 (a)	Explain the various components in the apache spark ecosystem.	**(10 Marks)**	**CO3**	**L2**
10 (b)	Explain lazy evaluation, narrow transformation, wide transformation.	**(10 Marks)**	**CO3**	**L2**