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

21CST701

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

BIG DATA ANALYTICS

(Model Question Paper - I)

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

Instructions to students

1. Answer FIVE FULL Questions.

Q.No.	Questions	Marks	Course Outcome	RBT Level
1 (a)	Define Big Data. Explain 8 characteristics of Big Data.	**(10 Marks)**	**CO1**	**L2**
1 (b)	Explain with examples the different types of Big Data	**(10 Marks)**	**CO1**	**L2**
	OR			
2 (a)	What is Data warehouse? Explain the key components in typical data warehouse environment.	**(10 Marks)**	**CO1**	**L2**
2 (b)	Differentiate between Data Warehouse and Hadoop.	**(10 Marks)**	**CO1**	**L2**
3 (a)	Explain the hadoop distributed file system architecture with a neat sketch. Illustrate with example how files are stored in HDFS. Also, justify how HDFS is fault tolerant.	**(12 Marks)**	**CO2**	**L2, L3**
3 (b)	Compare Hadoop with traditional SQL databases	**(08 Marks)**	**CO2**	**L2**
	OR			
4 (a)	Write a Java program to implement the word count problem using MapReduce. Explain the Mapper, Reducer, and Driver classes in detail.	**(12 Marks)**	**CO3**	**L3**
4 (b)	Explain the key features and advantages of Hadoop.	**(08 Marks)**	**CO2**	**L2**
5 (a)	Write a MapReuduce program for sorting following data according to name. **Input:** **001,chp** **002,vr** **003,pnr** **004,prp**	**(12 Marks)**	**CO4**	**L3**
5 (b)	Describe the chores of Mapper, Combiner, Partitioner, and Reducer for the word count problem.	**(08 Marks)**	**CO3**	**L3**
	OR			
6 (a)	Define Big Data Analytics. What are the various types of analytics?	**(12 Marks)**	**CO2**	**L2**

6 (b)	Define Data Science and explain its relationship with Big Data analytics.	**(8 Marks)**	**CO2**	**L2**
7 (a)	Explain the concept of SerDe in Hive. Explain the role of SerDe in processing XML data in Hive. Discuss the steps involved in configuring an XML SerDe for a Hive table and how XML data is transformed into a structured format for querying. Additionally, provide an example query on the resulting table and explain how it accesses XML elements.	**(10 Marks)**	**CO4**	**L2, L3**
7 (b)	Describe the different file formats supported by Hive.	**(6 Marks)**	**CO3**	**L2**
7 (c)	Explain the various data types supported by Hive.	**(4 Marks)**	**CO3**	**L2**
	OR			
8 (a)	Describe the anatomy of a Pig program and explain its components.	**(10 Marks)**	**CO3**	**L2**
8 (b)	Write a Pig Latin script to perform the following tasks on a dataset **sales_data (fields: product_id, category, amount, date):** 1. Filter the data for sales in the "Electronics" category. 2. Calculate the total sales amount for each product_id in this category. 3. Sort the results by total amount in descending order. Provide an explanation for each step in your script.	**(10 Marks)**	**CO4**	**L2**
9 (a)	Explain the various components in the apache spark ecosystem.	**(10 Marks)**	**CO3**	**L2**
9 (b)	Explain the different Spark deployment modes.	**(10 Marks)**	**CO3**	**L2**
	OR			
10 (a)	Explain the key features of Apache spark. Explain about the different Cluster Managers in Apache Spark?	**(10 Marks)**	**CO3**	**L2**
10 (b)	Differentiate between Narrow and Wide Transformations in Spark. Provide examples of each.	**(10 Marks)**	**CO3**	**L3**

****