

USN

--	--	--	--	--	--	--	--	--	--

21CST701

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

BIG DATA ANALYTICS

(Model Question Paper - II)

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 Structured, Semi-structured and Unstructured data with examples.	(12 Marks)	CO1	L2
1 (b)	What is Data warehouse? Explain the key components in typical data warehouse environment.	(08 Marks)	CO1	L2
OR				
2 (a)	Explain Challenges with Big data.	(10 Marks)	CO1	L2
2 (b)	What is Business Intelligence? How it is different from big data.	(10 Marks)	CO1	L2
3 (a)	Explain in brief the core components of Hadoop ecosystem.	(12 Marks)	CO2	L2
3 (b)	Explain the structure and functionality of the Mapper class in a Java-based MapReduce program. Provide a code snippet as an example.	(8 Marks)	CO3	L3
OR				
4 (a)	Explain the architecture and working of Hadoop YARN. Discuss its main components and how it improves the functionality of Hadoop MapReduce.	(10 Marks)	CO2	L2
4 (b)	Illustrate with example how files are stored in HDFS. Also, justify how HDFS is fault tolerant.	(10 Marks)	CO2	L3
5 (a)	Describe the role of the Mapper in the MapReduce framework. Explain its different phases with example.	(10 Marks)	CO2	L2
5 (b)	Write a MapReduce program to search an employee name in the following data: Input: 001,chp 002,vr 003,pnr 004,prp	(10 Marks)	CO4	L2
OR				

6 (a)	Explain the following terminology of Big Data a. In-Memory Analytics b. In-Database processing c. Symmetric Multit-processor system d. Massively parallel processing e. Shared nothing architecture f. CAP Theorem	(12 Marks)	CO3	L3
6 (b)	Identify and discuss the greatest challenges faced in Big Data analytics.	(8 Marks)	CO2	L2
7 (a)	Discuss the architecture of Hive and explain its key components.	(10 Marks)	CO3	L2
7 (b)	Create a data file for below schemas: Order: CustomerId, ItemId, ItemName, OrderDate, DeliveryDate Customer: CustomerId, CustomerName, Address, City, State, Country i. Create a table for Order and Customer Data. ii. Write a HiveQL to find number of items bought by each customer.	(10 Marks)	CO4	L3
OR				
8 (a)	Define the following key data processing operators in Pig Latin: LOAD, FILTER, FOREACH, GROUP, and JOIN. Explain each operator with an example.	(10 Marks)	CO4	L3
8 (b)	Explain the following relational operators in Pig Latin with an example for each: a) FILTER b)GROUP BY c)JOIN d)UNION e)SPLIT	(10 Marks)	CO4	L3
9 (a)	Explain the apache spark architecture. Explain about the different Cluster Managers in Apache Spark?	(10 Marks)	CO3	L2
9 (b)	Compare Hadoop and Spark.	(5 Marks)	CO2	L2
9 (c)	Compare batch processing and real time processing.	(5 Marks)	CO2	L2
OR				
10 (a)	Explain lazy evaluation, narrow transformation, wide transformation.	(10 Marks)	CO3	L2
10 (b)	Write a Spark program(in python) that reads a file with over 100,000 entries (where each row or line has a <state, mnm_color, count>) and computes and aggregates the counts for each color and state.	(10 Marks)	CO4	L3
