Total No. of Printed Pages: 2

PRN No.	

PAPER CODE	U314-243 (ESE)

(AY: 2024-25) December 2024 (ENDSEM) EXAM

T.Y. (SEMESTER - I)

COURSE NAME: BIG DATA ANALYTICS Branch: CSE- AI COURSE CODE: CAUA31203

(T.Y PATTERN 2020)

Time: [1Hr 30 Min] **[Max. Marks: 40]**

(*) Instructions to candidates:

1) **Figures to the right indicate full marks.Use of scientific calculator is allowed**
2) **Use suitable data wherever required**
3) **All questions are compulsory. Solve any two sub question each from Questions 1 and 2**
4) **Solve any one subquestion (2 marks) from Questions 3 ,4 ,5 and 6 and sub question of 4 marks is compulsory from questions 3,4,5,and 6**

Q. No.	Question Description	Max. Marks	CO mapped	BT Level
Q.1	a) Discuss the three main characteristics of Big data, known as 3Vs.	[4]	1	Understand
	b) Explain the role of Distributed Systems in facilitating the processing architecture for Big data.	[4]	1	Understand
	c)Distinguish the processing of structured and unstructured data in the context of big data with examples of each categories.	[4]	1	Understand
Q2	a) Elaborate the working of Hadoop MapReduce framework with a schematic diagram showing all of its components.	[4]	2	Understand
	b) Discuss the working of MapReduce computing as a parallel computing system with a clear structural diagram.	[4]	2	Understand
	c)Write the sample map, reduce function for a word count program, and interpret the working of JobTracker and TaskTracker processes.	[4]	2	Understand
Q3	a) Identify the main features of HDFS and explain their usefulness in a big data context.	[2]	3	Understand
	OR			
	c) Explain why Replication is pursued in HDFS even though it may cause Data redundancy.	[2]	3	Understand
	c)Interpret the workings of HDFS architecture with brief descriptions of the client communications using the	[4]	3	Apply

	name node and datanode.			
Q4	a) Explain the core component of Apache Spark system with its usability in big data processing. OR b) Explain the concept of Resilient Distributed Datasets (RDDs). c)Illustrate the working of cluster computing using Spark framework to perform in-memory computations in detail.	[2] [2] [4]	4 4 4	Understand Understand Apply
Q.5	a) Discuss the role of CPU as a host in CUDA enabled GPU systems. OR b) Express the working of CUDA systems with brief descriptions about flow of processing. c)Explore the utilization of symmetric multiprocessing in GPU core with architectural diagrammatic descriptions.	[2] [2] [4]	5 5 5	Understand Understand Apply
Q.6	a) Describe outliers in Machine Learning and the necessity of its detection. OR b) Discuss the regression analysis in brief for big data computing. c)Interpret the usability of any selected machine learning clustering algorithms to incorporate for Big data processing.	[2] [2] [4]	6 6 6	Understand Understand Apply

********** ALL THE BEST **********