Total No. of Printed Pages: 2

PRN No.	

PAPER CODE	U314-253 (ESE)

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

TY (SEMESTER - I)

COURSE NAME: BIG DATA ANALYTICS **Branch: CSE-AIML** **COURSE CODE: CMUA31203**

(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 five characteristics of Big data popularly known as 5Vs.	[4]	1	Understand
	b) Explain the specific role of Distributed System to structure the processing systems for Big data analytics.	[4]	1	Understand
	c)Distinguish between the structured and unstructured data in the context of big data with examples of each categories.	[4]	1	Understand
Q2	a) Describe the working of map reduce framework with a diagram showing their interactions.	[4]	2	Understand
	b) Draw the Hadoop system's detail architecture and discuss each component's working in detail.	[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. 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 name node and datanode.	[4]	3	Apply

Q4	a) Explain Apache Spark and discuss why it is used in data processing. OR b) Identify the significance of Apache Spark in comparison with Hadoop MapReduce framework. c)Illustrate the working of cluster computing to handle compute-intensive tasks and interpret the usability of the spark framework to manage the overall system in detail.	[2] [2] [4]	4 4 4	Understand Understand Apply
Q.5	a)Classify the usability of the device and host in CUDA-enabled systems. OR b) Express the working of CUDA systems with brief descriptions about flow of processing. c)Explore the workings of GPU computing with a clear architecture diagram showing different components.	[2] [2] [4]	5 5 5	Understand Understand Apply
Q.6	a) Describe Outliers in Machine Learning and Why do We need to Detect Outliers. OR b) Discuss the regression analysis in brief for big data computing. c)Interpret the usability of any selected machine learning classification algorithms to incorporate for Big data processing.	[2] [2] [4]	6 6 6	Understand Understand Apply

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