VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

SLOT: F2 +TF2

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - I
FALL SEMESTER 2025-2026

Programme Name & Branch	**: B.Tech (CSE)**
Course Code and Course Name	**: BCSE402L – BIG DATA ANALYTICS**
Faculty Name(s)	**: Dr. SUGANTHINI. C, Dr. NAGARAJA RAO. A, Dr. PARTHASARATHY**
Class Number(s)	**: VL2025260101959, VL2025260101711, VL2025260101715**
Date of Examination	**: 22.08.2025**
Exam Duration	**: 90 minutes** **Maximum Marks: 50**

General instruction(s):

- Answer All Questions
- M - Max mark; CO – Course Outcome; BL – Blooms Taxonomy Level (1 – Remember, 2 – Understand, 3 – Apply, 4 – Analyse, 5 – Evaluate, 6 – Create)
- Course Outcomes:
  CO1: Recall the characteristics of digital data, data sources, data storage and the applications of big data in different fields.
  CO2: Utilize Hadoop ecosystem tools and Hadoop YARN functions for parallel processing of application tasks.
  CO3: Comprehend the Map Reduce programming model and the Map Reduce Daemon framework.

Q. No	Question	M	CO	BL
1.	A city's transportation department is implementing a new system that uses thousands of sensors on public buses, traffic lights, and public transport stations. These sensors generate real-time data streams on traffic speed, bus locations, and passenger counts. Additionally, the department collects unstructured data from social media feeds about traffic jams and public transit delays. How do the 5 V's of Big Data apply to this scenario? Identify which specific data sources correspond to each V and why?	10	1	2
2.	A ride-sharing company needs to store and analyze two types of data: a. Structured data on completed rides (driver ID, passenger ID, fare, etc.). b. Unstructured data from driver and passenger feedback, including text reviews and star ratings. Recommend appropriate data storage and analysis technologies for each type of data within a Hadoop ecosystem. Explain why certain choices are better than others.	10	1	2
3.	Explain the core components of the Hadoop ecosystem and their respective functions. How do they work together to process big data?	10	2	1
4.	Describe a typical workflow for using the Grunt Shell to analyze a dataset. What are the key commands and steps involved in loading, transforming, and grouping data using Pig Latin?	10	2	3
5.	a. Describe the MapReduce framework. Explain the purpose of the Map and Reduce functions and provide a diagram to show how they work together.	5	3	3
	b. Describe the MapReduce programming model to count the frequency of each word. Explain in detail the map and reduce functions and the role of key-value pairs for the below given sentences. a. The quick brown fox jumps over the lazy dog. b. The dog sat on the mat.	5	3	

****