Reg. No. : | | | | | | | | | | | | |

Question Paper Code: U7E03

B.E./B.Tech. DEGREE EXAMINATION, NOV 2024

Seventh Semester

Artificial Intelligence and Data Science

21UAD703-BIG DATA VISUALIZATION

(Regulations 2021)

Duration: Three hours Maximum: 100 Marks

Answer All Questions

PART A - (10 x 2 = 20 Marks)

1.	What is big data?	CO1-U	
2.	Apply the concept of stream processing to monitor real-time social media trends. How would you approach this task?	CO2-App	
3.	What is the role of anomaly detection in data validation?	CO1-U	
4.	Explain the concept of column family in HBase.	CO1-U	
5.	How does a scatter map differ from a traditional scatter chart?	CO1-U	
6.	Why should you avoid using 3D effects in bar charts or pie charts?	CO1-U	
7.	How could you use an incidence matrix to represent a directed graph with three vertices and two edges?	CO2-App	
8.	How can outliers be identified in a scatter chart?	CO1-U	
9.	What is the purpose of visualizing port scan data?	CO1-U	
10.	What role does color coding play in visualizing firewall log data?	CO1-U	

PART – B (5 x 16= 80 Marks)

11.	(a)	Describe the role of Hadoop in the evolution of big data.	CO1-U	(16)
		Or		
	(b)	Describe a common big data use case in the healthcare industry and howit can improve patient outcomes.	CO1-U	(16)
12.	(a)	Consider a situation where you need to store and analyze large sets of unstructured data. Which aggregate data model would you choose for optimizing read performance, and how would you structure the data?	CO2-App	(16)

Or

	(b)	Write a HiveQL query that performs a complex join between three tables containing customer, order, and product information. Explain how you would optimize the query and improve its performance using Hive features like bucketing or partitioning.	CO2-App	(16)
13.	(a)	What are the key principles behind effective data visualization, and how do they influence the design of visualization functions?	CO1-U	(16)
		Or		
	(b)	How do you maintain consistency and standardization across multiple data presentations in an organization?	CO1-U	(16)
14.	(a)	What are scatter maps, and how do they differ from scatter charts in terms of data representation?	CO1-U	(16)
		Or		
	(b)	What are the most effective ways to detect non-linear correlations in big data, and how do they differ from linear correlation analysis?	CO1-U	(16)
15.	(a)	After conducting vulnerability assessment across a company's IT infrastructure, create a visualization that prioritizes vulnerabilities by severity and exploitability. Explain how you would represent the relationships between different vulnerabilities, potential exploits, and affected systems to help security teams prioritize remediation efforts.	CO2-App	(16)
		Or		
	(b)	You are tasked with analyzing network security using port scan data. Design a visualization that highlights suspicious activities across different ports and time periods. How would you represent patterns that could indicate potential threats, and what visual techniques would you use to distinguish between benign and malicious scans?	CO2-App	(16)