

THE KAVERY ENGINEERING COLLEGE*(An Autonomous Institution)***M.E/MCA DEGREE END SEMESTER THEORY EXAMINATIONS, APRIL/MAY 2026***Third Semester**Computer Science and Engineering***POE2404 - Big Data Analytics****(Common to Master of Computer Applications)***Regulations -2024**Duration : 3 Hrs**Maximum : 100 Marks***PART - A (ANSWER ALL QUESTIONS) (Marks 10*2=20)**

<i>Q.No</i>	<i>Questions</i>	<i>BTL</i>	<i>CO</i>	<i>Marks</i>
1.	Show how a Big Data platform can address the Challenges of Conventional Systems for a high-velocity dataset.	L2	1	2
2.	Compare descriptive statistics and inferential statistics in the context of Big Data analytics.	L2	1	2
3.	Show an example of a simple evaluation strategy for comparing two solutions.	L2	2	2
4.	Distinguish between filtering and zooming as interaction techniques with an example.	L2	2	2
5.	Define RTAP.	L1	3	2
6.	Analyze the primary reason why a Stream Data Model is suitable for Stock Market Predictions.	L4	3	2
7.	State the concept of sharding in a database system.	L2	4	2
8.	Compare the primary architectural difference between MapReduce and Hive.	L2	4	2
9.	Define recursion.	L1	5	2
10.	Outline an R function that accepts a number and prints whether it is positive, negative, or zero.	L2	5	2

PART - B (ANSWER ALL QUESTIONS) (Marks 5*16 = 80)

<i>Q.No</i>	<i>Questions</i>	<i>BLT</i>	<i>CO</i>	<i>Marks</i>
11.	a Analyze the statement "Web Data is the Most Popular Big Data" with reference to data analytic professional.	L4	1	16
	Or			
	b Examine the various analysis and reporting tools utilized in Big Data, highlighting their roles, functionalities, and comparative advantages in detail.	L4	1	16
12.	a Apply a visual data analysis dashboard for a retail company to monitor sales performance. Apply different Visualization Techniques to represent various data types, such as sales over time using line chart, product categories using bar chart, and geographic sales distribution (choropleth map).	L3	2	16
	Or			
	b Choose a case study where genetic programming is used to evolve decision rules for classification, and critically evaluate the generated rules for interpretability and effectiveness.	L3	2	16

13.	a	i	Classify assess the architectural components of a Real-time Analytics Platform (RTAP).	L4	3	8
		ii	Examine how the different layers in RTAP work together to provide a seamless analytics experience.	L4	3	8
	Or					
	b	i	Analyze the challenges of "Counting Distinct Elements in a Stream".	L4	3	8
		ii	Compare how the Flajolet-Martin Algorithm addresses the memory limitations of a single machine and justify its use in a real-world scenario like counting unique website visitors.	L4	3	8
14.	a		Explain the different types of NoSQL Databases. Explain the basic data model of each and provide a simple example.	L2	4	16
	Or					
	b		Illustrate the architecture of the Hadoop Distributed File System (HDFS). Explain the function of the NameNode and the DataNode and how they work together to store and retrieve data.	L2	4	16
15.	a	i	Develop an R script to load a dataset containing customer demographics and purchasing behavior. Perform data cleaning operations such as handling missing values, renaming columns, and removing duplicates.	L3	5	8
		ii	Develop an R script to manipulate the cleaned customer dataset using various R data structures. Demonstrate operations such as filtering customers by age group, grouping by region, and summarizing total purchase amounts.	L3	5	8
	Or					
	b	i	Construct an R function that takes a month number (1-12) as input and print the corresponding month name and the number of days.	L3	5	8
		ii	Apply an R function that receives an employee's salary and job role as input. Calculate the updated salary after applying different bonus rates depending on role of Manager, Developer, Analyst, Clerk.	L3	5	8