

THE KAVERY ENGINEERING COLLEGE*(An Autonomous Institution)***B.E/B.TECH DEGREE END SEMESTER THEORY EXAMINATIONS, APRIL-MAY 2025***Eighth Semester**Computer Science and Engineering***CCS368 - Stream Processing***Regulations – 2021***Duration : 3 Hrs****Maximum : 100 Marks****PART – A (ANSWER ALL QUESTIONS) (Marks 10*2=20)**

Q.No	Questions	BLT	CO	Marks
1.	What is data processing, and why is it critical for businesses?	L1	1	2
2.	Name one challenge organizations might face when implementing Data as a Service (DaaS).	L1	1	2
3.	Mention any two primary benefits of Big Data technologies.	L1	2	2
4.	Compare real-time analytics from batch processing.	L1	2	2
5.	How does a document database differ from a relational database?	L3	3	2
6.	What does schema flexibility refer to in NoSQL databases?	L2	3	2
7.	What role do producers play in Apache Kafka?	L2	4	2
8.	How does Kafka handle fault tolerance and data replication?	L3	4	2
9.	What is late data in the context of streaming?	L1	5	2
10.	Differentiate between tumbling windows and sliding windows.	L2	5	2

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

Q.No	Questions	BLT	CO	Marks
11.	a Discuss the concept of data processing and describe in detail each of the primary stages of data processing. How do these stages contribute to transforming raw data into meaningful information.	L2	1	16
	Or			
	b i Compare and Contrast Batch Processing and Stream Processing.	L2	1	8
	ii Describe the main techniques used in data mining, such as clustering, Classification, regression and association rule mining. Explain how each technique works and provide examples of the applications.	L2	1	8
12.	a What is real-time analytics? Analyze how it works and provide two examples of applications that benefit from real- time analytics.	L4	2	16
	Or			
	b Analyze the advantages and limitations of Lambda Architecture and kappa Architecture for processing Big data. Discuss the impact on system design, data processing efficiency, scalability. How do these architectures handle the challenges of data consistency, latency, and fault tolerance? Provide example of real – world applications where each architecture is beneficial preferable.	L4	2	16

13.	a	Explain the differences between the relational model and the document model. How does each model handle data structure and schema flexibility? Provide examples of each type of database and describe a scenario where one might be preferred over the other.	L3	3	16	
		Or				
	b	Describe the basic concepts of graph data models and how they differ from relational data models. Explain how Cypher Query Language is used to query graph databases and provide an example of a simple graph query. Also, briefly describe the purpose of SPARQL in the context of the Semantic Web.	L3	3	16	
14.	a	i	Demonstrate the overall architecture of Apache Kafka, detailing the roles of brokers, producers, and consumers.	L2	4	8
		ii	Compare and Contrast the producer API and Consumer API Kafka.	L2	4	8
		Or				
	b	Describe the Kafka Streams API and its purpose in stream processing. Provide examples of real – world applications that benefit from Kafka Streams and discuss how it simplifies the development of streaming applications.	L2	4	16	
15.	a	Analyze the trade – offs between using tumbling windows and sliding windows in Structured Streaming. Discuss the implications of each window type on data processing performance, accuracy of results, and use cases. Provide examples where one window type must be preferable over the other.	L4	5	16	
		Or				
	b	Discuss the impact of fault tolerance and exactly – once semantics on the design and implementation of a Structured Streaming application. Explain how checkpointing, write – ahead logs, and state management contribute to achieving exactly – once processing guarantees. Provide a detailed example of a streaming application and describe hoe these mechanisms would be implemented to ensure reliability and consistency.	L4	5	16	