

Course Code: : AI254TA
Semester : VII
Max Marks : 10+50

Date: _____
Time: _____
Duration: _____

27/03/2025
9:00-11:00
120 mins

Stream Processing and Analytics

CIE 2

Note: Answer all the questions

SL. No	Questions	M	BT	CO
PART - A				
1	Give the architecture diagram of Apache Samza	2	2	1
2	Differentiate Apache Storm and samza with reference to message delivery semantics	2	2	2
3	How do you handle fault tolerance using state machine approach	2	2	2
4	Give the difference between stream time and event time. Give an example	2	2	-
5	What is trigger policy and eviction policy?	2	2	2
PART - B				
1	a With a neat diagram, highlight the key differences between traditional DBMS and streaming systems, focusing on query management	6	2	1
	b "Membership type queries usually involve checking if an element belongs to a specific set or group." Justify how bloom filters can be used to solve these queries in real world applications	4	2	3
2	a A leading financial services company, "FinSecure," processes millions of transactions daily across its digital payment platform. The company needed a robust stream processing architecture to handle high transaction volumes in real time to detect fraudulent activities, ensure compliance, and generate operational insights. Build a real-time stream processing system to detect fraudulent transactions within milliseconds using Apache Spark. Support your answer with an architecture diagram and the elaborate the steps in detail	10	3	
3	a Elaborate the different points of failure in a stream processing system with a neat diagram	10	2	
4	a Imagine we are collecting data from a fitness-tracking device such as a Fitbit, and the data is flowing into our streaming system. The head of product marketing has asked us to build a dashboard that shows the average speed for all runners broken down by age groups, such as 12-17, 18-24, 25-34, and so on. The dashboard should be updated every 5 seconds, and the averages should represent data for the last 30 minutes.	10	3	
	(i) Elaborate do we handle this situation using the sliding window technique? Draw the sliding window to support your answer.			



Department of Artificial Intelligence and Machine Learning

		(ii) Explore how various stream processing frameworks facilitate the implementation of sliding window and tumbling window techniques for real-time data processing.		
5	a	Paraphrase the given statement by discussing any three strategies of caching systems. "Caching systems use many different strategies; the ones related to persisting cache entries and keeping the cache fresh."	10	2

M-Marks, BT-Blooms Taxonomy Levels, CO-Course Outcomes

	Particulars	CO1	CO2	CO3	CO4	CO5	L1	L2	L3	L4	L5
Marks Distribution	Max Marks Part A		8	2				10			
	Max Marks Part B	6	20	24				30	20		

Course Outcomes: After completing the course, the students will be able to:-

	Describe the need and the application of real time and stream processing in real world applications.
CO2	Comprehend and apply the various operations like data ingestion, data communication, data analysis, storage for different streaming data applications.
CO3	Investigate and apply streaming concepts using modern tools to solve problems related to social industry.
CO4	Demonstrate a prototype application for streaming data using Kafka as a team / individual.
CO5	Demonstrate solutions for societal and environmental concern problems using modern engineering through writing effective reports.