RV College of Engineering®
Mysore Road, RV Vidyaniketan Post,
Bengaluru - 560059, Karnataka, India

1RV22AI007

Department of Artificial Intelligence and Machine Learning

Date: 06.11.2025	**Test – 1**	**Max. Marks: 10 + 50**
Semester: VII	**UG**	**Duration: 2 Hrs.**
Course Title: Stream Processing and Analytics		**Course Code: AI372TA**

Common to AIML and CSE (Data Science)

PART A

S. No	Questions	M	BT	CO
1	Identify the most appropriate interaction pattern for the following applications and justify your choice: a)An e-commerce system continuously collects clickstream data. b) A weather data collector sends periodic updates without expecting a reply.	2	3	3
2	List any two needs for streaming data	2	1	1
3	A data collection node loses messages due to a power failure. Which message logging technique (RBML/SBML/HML) would you use to ensure complete recovery, and why?	2	3	3
4	Schematically represent the kafka producer components	2	2	2
5	Summarize any two differences between early rebalance and cooperative rebalance in Kafka.	2	2	2

PART B

S. No	Questions	M	BT	CO
1a.	Consider an IoT-based environmental monitoring network that streams sensor data globally. • Propose an architecture with a neat schematic diagram that balances security, scalability, and fault tolerance. • Justify your choice of interaction pattern and scaling model Explain how checkpointing and logging can jointly ensure reliable recovery across multiple layers and technologies.	10	4	1,3
2a.	Examine the conceptual overlap and differences between soft, near, and non-hard real-time systems and streaming data systems.	5	2	2
2b.	Describe and compare the common interaction patterns with suitable examples.	5	2	3
3a.	A global financial services company processes billions of transactions daily to detect fraudulent activity. The company leverages a real-time fraud detection pipeline using streaming data. The company wants to implement different fault tolerant mechanisms to have secure transactions. What are the potential challenges that occur due to the failure in the stream processing? Discuss how the different stream processing tools handle failures in communication? What happens if a broker crashes?	10	5	3

RV College of Engineering®
Mysore Road, RV Vidyaniketan Post,
Bengaluru - 560059, Karnataka, India

4a.	With a neat diagram elaborate the Kafka data eco system	5	3	3
4b.	Assume a real-time e-commerce platform where every order placed by a customer must be published to a Kafka topic named orders. The orders are later consumed by: • An inventory service (to reduce stock), • A billing service (to generate invoices), and • A notification service (to send order confirmations). This system handles thousands of orders per second. The system needs high throughput and no duplicate orders. Analyse and demonstrate the kafka producer using a program.	5	4	5
5a.	Assume you have a kafka topic named Sales data that contains 6 partitions (P0-P5). A consumer group with two consumers (C1 and C2). Assuming kafka uses RangeAssignor strategy, answer the following questions: (i) Explain how partitions will be assigned between C1 and C2. (ii) What issue can arise if you add a third consumer (C3) to the group? (iii) Give the solution to resolve the issues	6	4	5
5b.	Consider e a Kafka topic named sales_data that contains messages in JSON format. Each message is produced as a UTF-8 encoded JSON string, and there is a need for the Kafka consumer to automatically convert it into a Python/Java dictionary when consuming. Demonstrate the consumer program.	4	4	5