KUMARAGURU college of technology
character is life

Register Number:……

M.E/M.TECH/MCA DEGREE EXAMINATIONS: APRIL/MAY 2024

(Regulation 2018)

Second Semester

DATA SCIENCE

P18ITE0012**:** Streaming Analytics

COURSE OUTCOMES

CO1: Explain the need for stream computing

CO2: Comprehend the architecture of stream analytics

CO3: Build data flow management pipelines for streams.

CO4: Process the streaming data

CO5: Deliver the results of streaming analytics

Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. Which of the following is an example of a Large Scale Stream Processing Engine? CO1 [K_2]

   a) Apache Hadoop b) Apache Spark

   c) MongoDB d) MySQL

2. Which of the following is not a characteristic of stream computing? CO1 [K_2]

   a) Processing data in small, continuous streams b) Real-time processing

   c) Ability to process large batches of data at once d) Handling of data in motion

3. Which attribute of streaming analytics architecture refers to the system's ability to process data with minimal delay? CO2 [K_2]

   a) High Availability b) Low Latency

   c) Horizontal Scalability d) Fault Tolerance

4. CO2 [K_2]

List I	List II
A. Apache Zookeeper	i. Ensures the system can continue operating even if certain components fail.
B. Low latency	ii. Commonly used tool for service configuration and management.
C. Horizontal scalability	iii. The system's capacity to process data in parallel by adding more resources.
D. Fault tolerance	iv. Refers to the system's ability to process data with minimal delay.

a) A-i,B-iii,C-ii,D-iv
b) A-ii,B-iii,C-iv,D-i
c) A-iv,B-iii,C-ii,D-i
d) A-iii,B-ii,C-i,D-iv

5. Which data delivery guarantee is provided by Apache Kafka? CO3 [K_2]

a) At Most Once Delivery
b) Best Effort Delivery
c) Exactly Once Delivery
d) At Least Once Delivery

6. What is the difference between an event and a task in data flow management? CO3 [K_2]

a) Events are asynchronous, while tasks are synchronous
b) Events are stateless, while tasks maintain state
c) Events represent actions that have occurred, while tasks represent actions to be executed
d) Events are processed sequentially, while tasks can be executed in parallel

7. Assertion (A): Distributed stream data processing involves coordination, partitioning, and merging of data across multiple nodes. CO4 [K_2]

Reason (R): Coordinating the processing of streaming data ensures synchronization and proper handling of data across distributed systems, while partitioning and merging techniques optimize the distribution and consolidation of data streams.

a) Assertion (A) is correct and Reason (R) is not correct
b) Assertion (A) is not correct but Reason (R) is correct
c) Both Assertion (A) and Reason (R) are correct, and Reason (R) explains Assertion (A).
d) Both Assertion (A) and Reason (R) are not correct

8. What role does coordination play in distributed stream data processing? CO4 [K_2]

a) Synchronizing the processing of data across multiple nodes
b) Managing the distribution of data across multiple nodes
c) Ensuring fault tolerance in the system
d) Optimizing the storage of streaming data

9. Which technique is commonly used for delivering streaming metrics in real-time? CO5 [K_2]

a) Stochastic Optimization
b) Mobile Streaming Apps
c) Batch Processing
d) Static Data Analysis

10. What is the purpose of delivering time series data in streaming analytics? CO5 [K_2]

a) To increase the complexity of data analysis
b) To improve data visualization techniques
c) To provide historical insights into data trends over time
d) To decrease the efficiency of data processing

PART B (10 x 2 = 20 Marks)

11. State the differences between streaming data and static data. CO1 [K_2]
12. Identify the major sources of streaming data. CO1 [K_2]
13. How clock synchronization is achieved in streaming process? CO2 [K_2]
14. State the purpose of Zookeeper. CO2 [K_2]
15. Mention the use of Apache Flume. CO3 [K_2]
16. Identify three options in data delivery and processing. CO3 [K_2]
17. State the purpose of Nimbus. CO4 [K_2]
18. Define NoSQL storage system. CO4 [K_2]
19. Write the techniques used in stochastic optimization methods for streaming. CO5 [K_2]
20. How to deliver time series data? CO5 [K_2]

PART C (6 x 5 = 30 Marks)

21. Summarize the issues in stream data processing. CO1 [K_2]

22. Enumerate real time architecture components and their functionality. CO2 [K_2]

23. Illustrate Apache Flume framework and how it can be used for log management? CO3 [K_3]

24. Elaborate on the properties of distributed streaming data processing system. CO4 [K_2]

25.	Demonstrate key-value store No-SQL database.	CO4	[K_2]
26.	Illustrate any one mobile streaming application with neat sketch.	CO5	[K_3]

Answer any FOUR Questions
PART D (4 x 10 = 40 Marks)

27.	Organize the components of large scale stream processing engine.	CO1	[K_3]
28.	Identify the features of real time architecture.	CO2	[K_3]
29.	How to apply Apache Kafka for real time data flow management? Explain in detail.	CO3	[K_3]
30.	Explain document store processing system MongoDB model.	CO4	[K_3]
31.	Enumerate the ways to apply timed counter for delivering stream data.	CO5	[K_3]

****