SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
FACULTY OF ENGINERING AND TECHNOLOGY
SCHOOL OF COMPUTING

COURSE PLAN

21CSE373T STREAMING ANALYTICS

JANUARY – MAY 2024

Revision History:

Date	Version	Modification done	Modified by	Reviewed by	Authorized by
09-01-2024	1.0	Initial Release	Dr. K.Sornalakshmi	Dr. S. Ganesh Kumar	

Table of Contents

1.0 General Details

Course Code: 21CSE373T

Course Title: Streaming Analytics

Course Time: JANUARY – MAY 2024

Slot: A

Day	Batch			
	Batch 1		Batch 2	
	Hour	Timing	Hour	Timing
Day order 1	1,2	8:00am - 9:40am	6,7	12:30pm - 2:15pm
Day order 2	-	-	-	-
Day order 3	3	9:45am - 10:35am	8	2:20pm - 3:10pm
Day order 4	-	-	-	-
Day order 5	-	-	-	-

Location: University Building, Tech Park

Tutorial Assessment Hour: Batch 1: Day order 3 - 3rd Hour & Batch2: Day order 3 - 8th Hour

2.0 Reference Books

.

1. Garillot F and Mass. G., Stream Processing with Apache Spark, 1st ed., O'Reilly Media, Inc., 2019
2. Narkhede N, Shapira. G, and Palino T., Kafka: The Definitive Guide - Real-Time Data and Stream Processing at Scale, 1st ed., O'Reilly Media, Inc.,2017
3. Ankit Jain, Mastering Apache Storm, 1st ed., Packt Publishing,2017
4. https://docs.mongodb.com/manual/changeStreams/
5. Shakuntala Gupta Edward Navin Sabharwal, "Practical MongoDB Architecting, Developing,and Administering MongoDB"Apress,2016
6. https://aws.amazon.com/dynamodb/features/?pg=dynamodbt&sec=hs

3.0 Prerequisites

Knowledge on any programming language, Operating Systems

4.0 Instructional Objectives

CO-1: Illustrate the concepts and terminologies in stream processing
CO-2: Interpret stream processing applications using Apache Spark Streaming
CO-3: Summarize real-time streaming data pipelines and applications that adapt to the data streams using Kafka
CO-4: Interpret stream processing applications using Apache Storm Streaming
CO-5: Inquire real time data using NoSQL databases & MongoDB

5.0 Overall Assessment Plan

#	Component	Type	Marks
1	Cycle Test - I	Written Test	10
		Lab Test in Apache Spark	10
2	Cycle Test - II	Written Test	10
		Lab Test in Apache Kafka, Apache Storm	10
3	Cycle Test - III	Written Test	5
		Lab Test in MongoDB	5
5	Quiz/Puzzles/Review Questions	Written/Code Demo	10
Total Marks			60

6.0 Tentative Test Schedule

#	Tentative date	Test	Marks	Portion	Duration
1	29-02-2024	Cycle Test - I	7.5	Unit 1 and 2	100 minutes
			5		
2	05-04-2024	Cycle Test - II	7.5	Unit 3 and 4	100 minutes
			5		
3	08-05-2024	Cycle Test - III	5	Unit 5	50 minutes

7.0 Detailed Test Plan

Test	Tentative Date	Type	Marks	Mode
Cycle Test - I	29-02-2024	Written Test	Total: 50 Marks Exam Pattern: Part A MCQ – 10 marks Part B – 5 (Out of 7) * 4 marks – 20 marks Part C – 2 * 10 marks (Either OR choice type)	Physical Exam
	26-02-2024	Programming Test	Total: 10 Marks 2 Apache Spark Commands (Each 5 marks)	Physical Exam using PySpark
Cycle Test - II	05-04-2024	Written Test	Total: 50 Marks Exam Pattern: Part A MCQ – 10 marks Part B – 5 (Out of 7) * 4 marks – 20 marks Part C – 2 * 10 marks (Either OR choice type)	Physical Exam
	10-04-2024	Programming Test	Total: 10 Marks Apache Storm Command (5 marks) Apache Kafka (5 marks)	Physical Exam using Kafka and Storm
Cycle Test - III	08-05-2024	Written Test	Total: 25 Marks Exam Pattern: Part A MCQ – 7 marks Part B – 2 (Out of 3) * 4 marks – 8 marks Part C – 1 * 10 marks (Either OR choice type)	Physical Exam
	25-04-2024	ProgrammingTest	Total: 5 Marks MongoDB Program (5 marks)	PhysicalExam using MongoDB

** Lab tests – A task will be given on the framework based on tutorial sessions. Student can complete the task in the same tutorial session with the help of tutorial commands/web reference and/or framework API reference.

8.0 Quiz/Puzzles/Review Questions

Total marks – 5. Two activities will be conducted. One for each 2.5 units and score will be calculated for 5 marks.

Test	Tentative Date	Portion
Quiz/Puzzles/Review Questions during theory class	07-03-2024	Units I, II and Unit III (till Topics and Partitions)
	02-05-2024	Remaining topics in Unit III, Unit IV and Unit V

9.0 Detailed Session Plan

#	Topics to be covered	Hours	Ref	Teaching method	Testing method
Unit 1					
1	Introducing Stream Processing, Stream Processing, Examples of Stream Processing, Scaling Up Data Processing, Distributed Stream Processing	1		Lecture	CT1, Quiz 1, Minor Project
2	Stream-Processing Model, Sources and Sinks, Immutable Streams Defined from One Another	1		Lecture	
3	Transformations and Aggregations, Window Aggregations	1		Lecture	
4	Type Conversions, Conditional and loopingstatements	1		Lecture, Demo	
5	Stateless and Stateful Processing, Stateful Streams, An Example: Local Stateful Computation in Scala	1		Lecture, Demo	
6	Stateless or Stateful Streaming, Streaming Architectures, Components of a Data Platform, Architectural Models	1		Lecture, Demo	
7	The Use of a Batch-Processing Component in a Streaming Application	1		Lecture, Demo	
8	Referential Streaming Architectures	1		Lecture, Demo	

9	Streaming Versus Batch Algorithms	1		Lecture, Demo	
Unit 2					
10	Apache Spark as a Stream-Processing Engine	1		Lecture, Demo	CT1, Quiz 1, Minor Project, Hackathon
11	Spark's Distributed Processing Model	1		Lecture, Demo	
12	Spark's Resilience Model	1		Lecture, Demo	
13		1		Lecture, Demo	
14	Introducing Structured Streaming	1		Lecture, Demo	
15		1		Lecture, Demo	
16	The Structured Streaming Programming Model	1		Lecture, Demo	
17		1		Lecture, Demo	
18		1		Lecture, Demo	
Unit 3					
19	Getting Started with Kafka, Kafka, Publish Subscribe messaging model	1		Lecture, Demo	CT2, Quiz 2, Minor Project, Hackathon
20	Kafka Architecture, Messages and Batches, Schemas, Topics and Partitions	1		Lecture, Demo	
21	Producers and consumers, Brokers and Clusters, Multiple sClusters	1		Lecture, Demo	
22	Data Ecosystem, Kafka Producers: Writing messages to Kafka, Kafka Consumers	1		Lecture, Demo	

23	Reading data from Kafka, Stream Processing- Stream Processing Design Patterns	1		Lecture, Demo	
24	Kafka Streams by Examples	1		Lecture, Demo	
25		1		Lecture, Demo	
26	Kafka Streams: Architecture Overview	1		Lecture, Demo	
27		1		Lecture, Demo	
Unit 4					
28	Apache Storm – Introduction	1		Lecture	CT2, Quiz 2, Minor Project, Hackathon
29	Real-Time Processing and Storm Introduction	1		Lecture	
30		1		Lecture	
31	Storm Deployment	1		Lecture, Demo	
32	Topology Development	1		Lecture, Demo	
33	Topology Options	1		Lecture, Demo	
34	Storm Parallelism and Data Partitioning	1		Lecture, Demo	
35	Integration of Storm and Kafka	1		Lecture, Demo	
36		1		Lecture, Demo	
Unit 5					
37	NoSQL Data Bases	1		Lecture	CT3, Quiz 2, Minor Project
38	AWS Cloud Dynamo Database: Amazon DynamoDB features, Serverless	1		Lecture	

39	Introduction to MongoDB, MongoDB Data Model	1		Lecture, Demo	
40	MongoDB Architecture	1		Lecture, Demo	
41	Core Processes, MongoDB Tools	1		Lecture, Demo	
42	Standalone Deployment, Replication	1		Lecture, Demo	
43	Sharding, MongoDB Use Cases	1		Lecture, Demo	
44	Performance Monitoring	1		Lecture, Demo	
45	Social Networking	1		Lecture, Demo	