Roll No: [ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ]

BTECH
(SEM V) THEORY EXAMINATION 2023-24
STREAM PROCESSING AND ANALYTICS

TIME: 3 HRS   M.MARKS: 100

Note: 1. Attempt all Sections. If require any missing data; then choose suitably.

SECTION A

1. **Attempt *all* questions in brief.**   2 x 10 = 20

Q no.	Question	Marks	CO
a.	What is scaling?	2	1
b.	What is meant by Windows Aggregation?	2	1
c.	What is meant by Batch Processing?	2	2
d.	What is meant by Data Analysis?	2	2
e.	Explain the Distributed System.	2	3
f.	What is meant by Data Server?	2	3
g.	Define IoT.	2	4
h.	What is meant by Structured Streaming?	2	4
i.	What is meant by Time-based Stream Processing?	2	5
j.	Discuss Spark Streaming Context.	2	5

SECTION B

2. **Attempt any *three* of the following:**   10 x 3 = 30

a.	Differentiate between Stateless and Stateful Processing with suitable examples.	10	1
b.	Write a simple Spark program that prints a sample of the tweets it receives from Twitter every second. What is a Socket?	10	2
c.	Write a script in Spark to read all text files into Spark RDD. Define Spark Components.	10	3
d.	Discuss Structured Streaming Sinks. Define Trigger and Latency.	10	4
e.	Discuss Structured Streaming Vs Spark Streaming.	10	5

SECTION C

3. **Attempt any *one* part of the following:**   10 x 1 = 10

a.	Write short notes on:
	   i. Performance tuning.
	   ii. Time-based streaming process.
		10	5

b.	What are the components of Spark Streaming? What is DStreams?	10	5

4. **Attempt any *one* part of the following:**   10 x 1 = 10

a.	Discuss Sources and Sinks Stream Processing Model.	10	1
b.	What is Apache Spark? What is Stream Processing?	10	1

5. **Attempt any *one* part of the following:**   10 x 1 = 10

a.	Discuss Batch Processing Component in a Streaming Application.	10	2

b.	Write short notes on:
	   i. Latency
	   ii. Throughput
		10	2

Roll No: | | | | | | | | | | | | |

BTECH
(SEM V) THEORY EXAMINATION 2023-24
STREAM PROCESSING AND ANALYTICS

TIME: 3 HRS
M.MARKS: 100

6. **Attempt any *one* part of the following:** 10 x 1 = 10

a.	Discuss Spark's Resilience Model. What is the use of Spark Context. Parallelize() method.	10	3
b.	Write short notes on: i. Cluster Manager. ii. Fault Tolerance in Distributed System.	10	3

7. **Attempt any *one* part of the following:** 10 x 1 = 10

a.	Write short notes on: i. Features of Structured Streaming. ii. Stream Execution.	10	4
b.	How to acquire streaming data from sources? What is a Streaming Dataset?	10	4