Name: **Enrolment No:**	

UPES

End Semester Examination, May 2024

Course: Stream Processing **Semester: VI**

Program: BT-CSE-Spz-BDATA (H&NH) **Time : 03 hrs.**

Course Code: CSBD3010 **Max. Marks: 100**

Instructions: Answer all the questions in sequential manner.

SECTION A
(5Qx4M=20Marks)

S. No.	Explain the concept of DStreams (Discretized Streams) in Spark Streaming and how they represent streaming data.	**Marks**	**CO**
Q 1	What are the differences between micro-batching and event-driven stream processing?	**4**	**CO1**
Q 2	What is the primary advantage of stream processing over batch processing?	**4**	**CO1**
Q 3	Discuss the benefits of using DataFrames and Datasets in Spark Structured Streaming for expressing streaming computations.	**4**	**CO2**
Q 4	Describe the role of triggers in Spark Structured Streaming and how they control the timing of stream processing.	**4**	**CO2**
Q 5	What are some common challenges faced when scaling distributed log processing systems, and how are these challenges typically addressed?	**4**	**CO3**
	SECTION B **(4Qx10M= 40 Marks)**		
Q 6	What are some common windowing techniques used in stream processing, and how do they affect the processing of streaming data?	**(5+5)**	**CO1**
Q 7	Discuss the types of joins supported for all the possible combinations of static and stream data frame.	**(5+5)**	**CO3**
Q 8	How do stream processing systems handle late arriving data, and what mechanisms are used to ensure accurate processing results?	**(5+5)**	**CO3**
Q 9	Explain the concept of stateful stream processing in Spark Streaming and its applications. OR Distinguish in between stateful and stateless streaming in detail.	**(10)**	**CO4**
	SECTION-C **(2Qx20M=40 Marks)**		
Q 10	Write a Python/Scala program to create a streaming DataFrame from any streaming source. Apply transformations such as filtering or aggregation on the streaming DataFrame, and then write the processed data to multiple sinks, including a file sink and a database sink. Ensure	**20**	**CO2**

	to include error handling mechanisms and discuss the scalability and fault-tolerance aspects of the program.		
Q 11	Discuss the role of various output modes in Structured Streaming. How does each mode affect the output of streaming queries, and when would you choose one mode over another? Explain the concept of triggers in Spark Structured Streaming and how they control the timing of query execution. What are the available trigger types, and how do they impact the behaviour of streaming queries? OR Deliberate any two use cases of stream processing in weather forecasting and stock analytics domain.	**(5+5+5+5)** **(10+10)**	**CO3**