


Name:			
Enrolment No:			
<b>UNIVERSITY OF PETROLEUM AND ENERGY STUDIES</b> <b>End Semester Examination, December 2022</b>			
<b>Course: Stream Processing</b> <b>Program: B.Tech-CSE-BD/BAO(Hons.)</b> <b>Course Code: CSBD 4001</b>		<b>Semester: VII</b> <b>Time: 03 hrs.</b> <b>Max. Marks: 100</b>	
<b>Instructions: Explain in short. (60-70 words)</b>			
<b>SECTION A</b> <b>(5Qx4M=20Marks)</b>			
S. No.		Marks	CO
Q 1	Explain the data storage layer in terms of stream processing.	4	CO1
Q 2	Describe any two transformations in Spark with a suitable example of each.	4	CO2
Q 3	Briefly discuss the difference in between stream-stream join.	4	CO3
Q 4	Explain the concept of real time ETL in association with structured streaming.	4	CO3
Q 5	Differentiate in between error and trace log with the help of an example.	4	CO4
<b>SECTION B</b> <b>(4Qx10M= 40 Marks)</b>			
<b>Instruction: Write brief notes. (100-150 words)</b>			
Q 6	Discuss the spark structured model in detail with a suitable example.	10	CO1
Q 7	Clarify the concept of late data handling in streaming with the help of a suitable example.	10	CO1
Q 8	Discourse the concept of check pointing and write ahead logs in terms of stream processing. <p style="text-align: center;">OR</p> Differentiate in between structured and unstructured data in context with data streaming.	10	CO2
Q 9	Discuss the concept of structured streaming API in association with streaming with the help of an example.	10	CO2
<b>SECTION-C</b> <b>(2Qx20M=40 Marks)</b>			
<b>Instruction: Write long answer. (Up to 350 words while explaining)</b> <b>Attempt any part of question no. 10 as there is an option “a” OR “b”.</b>			

**There is no choice for question no.11.**

Q 10	Explain the concept of shuffling in Spark. Discuss its two compression parameters.  OR  Demonstrate Apache Flume architecture and its data flow components in detail.	20	CO4
Q 11	Deliberate the use cases of stream processing in terms of real time stock trades and fraud detection.	20	CO3