

CS 6350.002 - Big Data Management and Analytics - S22

Course Homepage

Review Test Submission: Structured Streaming Quiz

Review Test Submission: Structured Streaming Quiz

User	Venkata Kowsik Temididapathi
Course	CS 6350.002 - Big Data Management and Analytics - S22
Test	Structured Streaming Quiz
Started	4/23/22 5:54 PM
Submitted	4/23/22 6:11 PM
Due Date	4/24/22 11:59 PM
Status	Completed
Attempt Score	50 out of 50 points
Time Elapsed	17 minutes
Results Displayed	All Answers, Submitted Answers, Correct Answers

Question 1

10 out of 10 points

Which of the following are features of Structured Streaming?

Hint: See <https://spark.apache.org/docs/latest/structured-streaming-programming-guide.html>

- Selected Answers:
- ☒ You can use Dataframe API
 - ☒ Aggregations, event-time windows, stream-to-batch join can be performed
 - ☒ Built on the Spark SQL Engine
 - ☒ Scalable and Fault-tolerant
- Answers:
- ☒ You can use Dataframe API
 - ☒ Aggregations, event-time windows, stream-to-batch join can be performed
 - ☒ Built on the Spark SQL Engine
 - ☒ Scalable and Fault-tolerant

Question 2

10 out of 10 points

What are the guarantees of Structured Streaming?

Hint: See <https://spark.apache.org/docs/latest/structured-streaming-programming-guide.html>

Selected



Answers: Using continuous processing, it can be process end-to-end data in times as low as 1 millisecond



Fault Tolerance



End-to-end exactly-once processing



Streaming data can be expressed in the same way as batch data

Answers:



Using continuous processing, it can be process end-to-end data in times as low as 1 millisecond



Fault Tolerance



End-to-end exactly-once processing



Streaming data can be expressed in the same way as batch data

Question 3

10 out of 10 points

Which of the following are valid parameters for spark.readstream method?

Selected Answers:



Kafka



Socket



FileSource

Answers:



Kafka



Socket



FileSource

None of the above

RDD

Memory

Question 4

10 out of 10 points

The flow of data in a Structured Streaming application is as follows:

Input Datastream -> Unbounded input table -> Results table -> output

Which of the following are valid output modes?

Selected Answers:



Append



Update



Complete

Answers:



Append

Update and then complete

- ✔ Update
- ✔ Complete

Question 5

10 out of 10 points

By default, how does Structured Streaming know the schema of file based sources?

Hint: See <https://spark.apache.org/docs/latest/structured-streaming-programming-guide.html#schema-inference-and-partition-of-streaming-dataframesdatassets>

Selected Answer: ✔ It has to be specified

- Answers:
- None of the above
 - It can automatically be inferred
 - ✔ It has to be specified
 - Schema is not necessary

Sunday, April 24, 2022 4:10:04 PM CDT

← OK