2023-CSE5BDC-T5-W - Big Data Management on the Cloud

Big Data Management on the Cloud Week 5. Quizzes or Notes

**Topic 1: Spark Streaming**

Which of the following are challenges faced when supporting event time query semantics? (Choose one answer.)

We may need to keep state information for past windows of time for a long time

Which of the following are common between batch query processing using dataframes and data stream processing using dataframes? (Choose two answers.)

You can use the same dataframes query operations like, join, filter, etc. for both types of query processing.

Both can use the code optimisations offered by the Tungsten engine.

## Topic 2: Spark GraphFrames

Which of the following ways of querying Big Data uses the Spark DataFrames API? (Choose three answers.)

GraphFrames

Spark Streaming

SparkSQL

Which of the following data are required when building a directed GraphFrame? (Choose two answers.)

-A column specifying the id of the vertices of the graph

-A column specifying the destination vertices for the graph edges