IST769 Unit K - Streaming

## Agenda

1. Your Questions
2. Go over Problem Set
3. Unit Coursework Activities

## FRONT MATTER

Start your engines!

PS> **docker-compose down**

PS> **docker-compose up -d jupyter drill zookeeper broker ksqldb-server ksqldb-cli schema-registry connect**

To access KSQL Db Client:PS>  **docker-compose exec ksqldb-cli ksql** [**http://ksqldb-server:8088**](http://ksqldb-server:8088/)

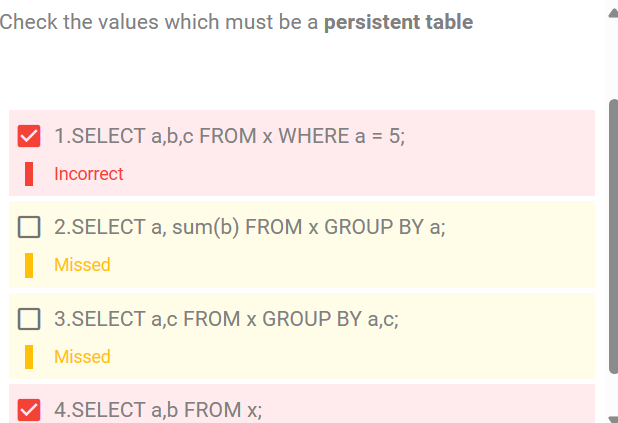
Start the ATM producer example from Jupyter, located at: **/work/examples/Kafka-Producer.ipynb**

**Kafka:**

* What is the storage model? (relational, document, etc…)
  + Streaming
* What are the logical model metaphors? (tables, keys constraints, etc…)
  + Append-Only File (cassandra, Redis)
* How does it scale horizontally? CAP? (after all - its a big data course!)
  + Yes / Partitions on topics
* Purpose of database? OLTP (CRUD) or OLAP/Analytics (CR)?
  + CR
* How do you design for it?
* How to connect to it from:
  + Client / Native Scripting Language  
     Python
  + Drill / SQL
    - Drill
  + Spark / PySpark
    - Spark Streaming

## 1. Your Questions

Ask any questions you have here!

* Does the explain command provide information about a stream?
* 
* Does sink subscribe to a kafka topic?
* Are topics replicated across each broker?
* Can messages sent to a topic be updated?

## 2. Coursework Activities

**Questions:**

1. ATM withdrawals from a specific user
2. Total Amount withdrawn over a specified time period
3. When Total Amount over time period is X display