Lab Homework 10: Kafka and KSQL

# Learning Objectives

In this exercise, you will practice writing KSQL streams against a Kafka setup:

* Create streams and tables.
* Write queries and persistent queries against streams and tables.
* Use window functions to summarize data over a timeframe.

# Setup

To complete this homework or follow along with the instructor’s class demos from your coursework, you must first complete the steps in this setup section. For this exercise we require the SQL Server instance in the Docker environment. For more details on the how’s and why’s of managing the database environments used in this course, consult Lab 1. Instructions for this assignment:

1. Open the PowerShell prompt on Windows or the Terminal on MacOS.
2. Type:  
   cd adv-db-labs  
   to change the working directory to the repository folder. If you are in the correct spot, your command prompt should have adv-db-labs in it, for instance: PS adv-db-labs>
3. Change into the kafka folder, type:  
   cd kafka  
   if you are in the correct folder, your command prompt should have Kafka in it.
4. Bring up the Kafka environment, type:  
   docker-compose up -d
5. Check to make sure the environment is running, type:  
   docker-compose ps

You should see the Kafka services all **Up**.

1. You are now ready to connect to **ksql-datagen** and start the **Weblogs** data generation script.   
   docker-compose run ksql-datagen bash

From the bash prompt, type these commands to get the **Weblogs** data generation script going:

* 1. **root@ksql-datagen:** cd /datasets/streaming
  2. **root@ksql-datagen/datasets/streaming:** pip install kafka
  3. **root@ksql-datagen/datasets/streaming:** python kafka-weblogs-datagen.py

When the script is running you will see JSON output in your window. This same output is sent to a Kafka topic. Keep this terminal window up and running during the homework.

1. Open another PowerShell / Terminal window and from the kafka folder connect to the ksql instance:  
   docker-compose exec ksql-cli ksql http://ksql-server:8088
2. When you are connected you should be at the ksql> prompt.

# Exercises

Complete each of the following exercises. If you are unsure how to accomplish the task, please consult the coursework videos where there are explanations and demos.

1. Write KSQL to create a stream named **weblogs** from the JSON keys in the **weblogs** Kafka topic. Make sure to set the TIMESTAMP property to the timestamp from the stream.
2. Write a KSQL statement create a persistent stream/table called **homepage** which only displays visitors to the root of the website (/). It should display all columns from the **weblogs** stream.
3. Write a KSQL statement to count operating systems users (os) in 60 second windows. After 60 seconds, the counter should reset, and counts should begin again.
4. Write a KSQL persistent stream/table called **user\_activity** which will display a count of user activity on the website within 1-minute sessions.
5. Write a KSQL statement to display users who have more than 1 pages of activity in a 1-minute session.

# Turning it in

Take your copy and paste each of the solutions to the exercises into the submission template file included with this assignment. Make sure your name and SU email are at the top and turn in your work through the course learning management system.

# Tear-Down

When you are finished with the homework you should stop the environment:

1. From the terminal window where you typed docker-compose up -d type in the following:  
   docker-compose stop