Homework Problem Set K Submission Form

# Overview

| Your Name | **Nick Videtti** |
| --- | --- |
| Your SU Email | **nvidetti@syr.edu** |

# Instructions

Put your name and SU email at the top. Answer these questions all from the lab. When asked to include screenshots, please follow the screenshot guidelines from the first homework.

Remember as you complete the homework that it is not only about getting it right/correct. We will discuss the answers in class so it’s important to articulate anything you would like to contribute to the discussion in your answer:

* If you feel the question is vague, include any assumptions you've made.
* If you feel the answer requires interpretation or justification, provide it.
* If you do not know the answer to the question, articulate what you tried and how you are stuck.
* Highlight any doubts or questions you would like me to review.

This how you receive credit for answering questions that might not be correct. In addition, you must complete the reflection portion of the homework assignment for full credit. Since most answers will be similar this is an important part of your individual submission.

Complete Part II of this document first, then go back and complete the Reflection in Part I.

# Part I: Reflection

Use this section to reflect on your learning. To achieve the highest grade on the assignment, you must be as descriptive and personal as possible with your reflection.

1. As you completed this assignment, identify what you learned.

**This gave me a better idea of Kafka, KSQL, and the use cases for streams and tables.**

1. What barriers or challenges did you encounter while completing this assignment?

**Some of the questions were a bit vague, but the following question after those usually cleared things up if I misinterpreted it the first time. Also, question 8 seemed to be not something that was covered in class and was mostly copy and paste, although it was pretty interesting to see the Python Kafka consumer in action.**

1. How prepared were you to complete this assignment? What can you do to be better prepared?

**I felt pretty prepared for this assignment. It seemed to be the easiest since maybe the first assignment (AB), which was refreshing as the last assignment of the course and ahead of me starting to work on the Final Project.**

1. Rate your comfort level with this week’s material. Use the rubric provided.

**4 ==> I understand this material and can explain it to others.**  
3 ==> I understand this material.  
2 ==> I somewhat understand the material but sometimes need guidance from others.  
1 ==> I understand very little of this material and need extra help.

# Part II: Questions

**For each question, include a copy of the code required to complete the question along with a screenshot of the code and a screenshot of the output.**

1. Write a Drill query to display only the ATM transactions that ended in an Error status. Show all columns and sort output so the newest errors are first.   
   NOTE: It is strongly suggested you use `backticks` and table aliases when working in Drill.

**SELECT**

**\***

**FROM**

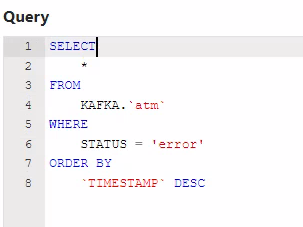
**KAFKA.`atm`**

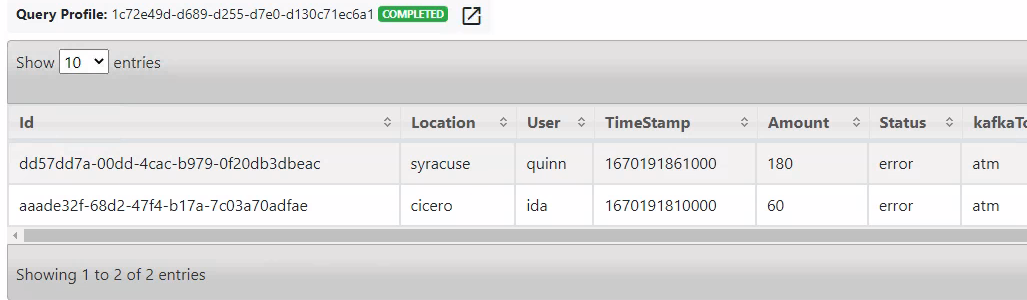
**WHERE**

**STATUS = 'error'**

**ORDER BY**

**`TIMESTAMP` DESC**





1. Write a Drill query to display the total amount withdrawn by user. Do not include error transactions in the totals.

**SELECT**

**UPPER(ATM.USER) ATM\_USER,**

**SUM(ATM.AMOUNT) TOTAL\_AMOUNT\_WITHDRAWN**

**FROM**

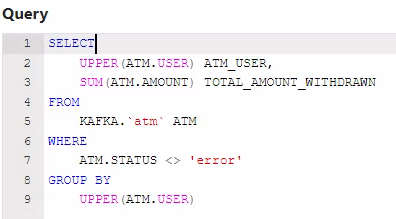
**KAFKA.`atm` ATM**

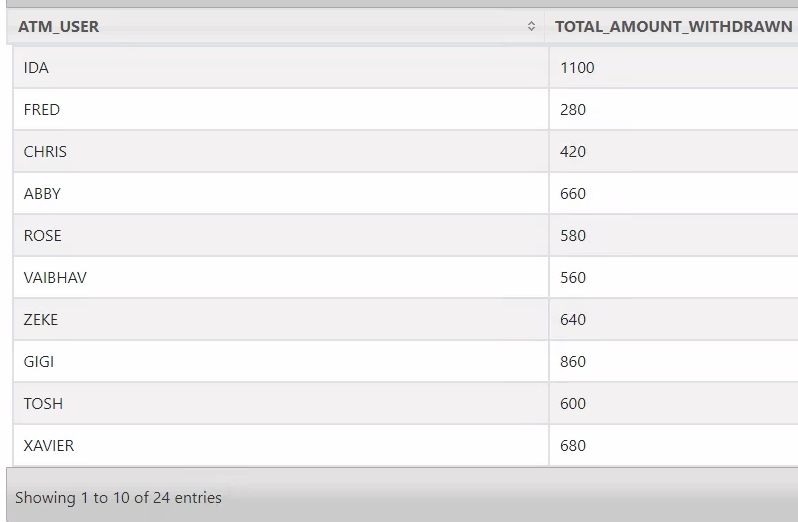
**WHERE**

**ATM.STATUS <> 'error'**

**GROUP BY**

**UPPER(ATM.USER)**

****

****

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.

**ksql> CREATE STREAM WEBLOGS (**

**>URI VARCHAR,**

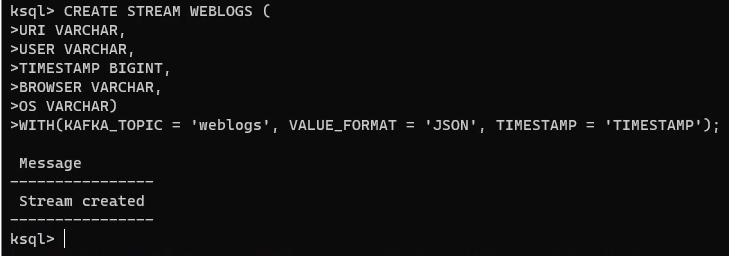
**>USER VARCHAR,**

**>TIMESTAMP BIGINT,**

**>BROWSER VARCHAR,**

**>OS VARCHAR)**

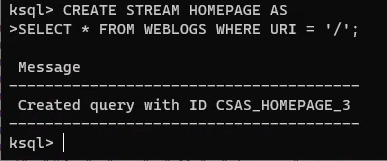
**>WITH(KAFKA\_TOPIC = 'weblogs', VALUE\_FORMAT = 'JSON', TIMESTAMP = 'TIMESTAMP');**

****

1. Write a KSQL statement to create a persistent stream/table called **homepage** that only displays visitors to the root of the website (/). It should display all columns from the weblogs stream.

**ksql> CREATE STREAM HOMEPAGE AS**

**>SELECT \* FROM WEBLOGS WHERE URI = '/';**

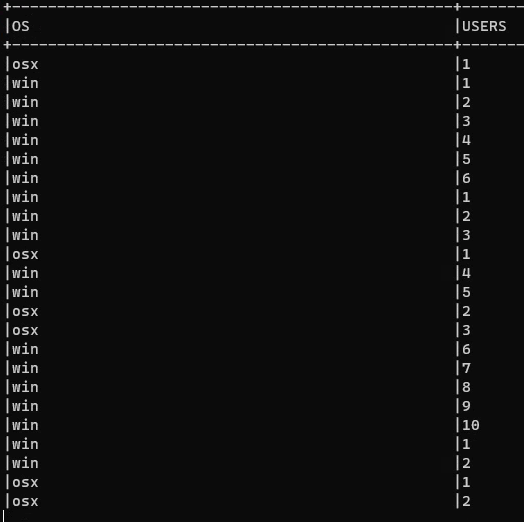


1. Write a KSQL statement to count operating systems users (OS) in 60-second windows. After 60 seconds, the counter should reset and counts should begin again.

**ksql> SELECT OS, COUNT(USER) USERS FROM WEBLOGS >WINDOW TUMBLING (SIZE 60 SECONDS)**

**>GROUP BY OS**

**>EMIT CHANGES;**

****

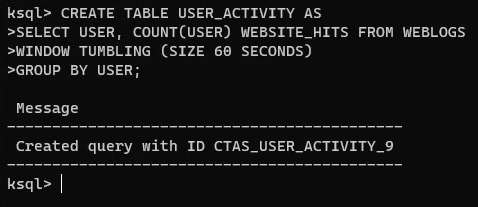
1. Write a KSQL persistent stream/table called **user\_activity** that will display a count of user activity on the website within 1-minute sessions.

**ksql> CREATE TABLE USER\_ACTIVITY AS**

**>SELECT USER, COUNT(USER) WEBSITE\_HITS FROM WEBLOGS**

**>WINDOW TUMBLING (SIZE 60 SECONDS)**

**>GROUP BY USER;**

****

1. Write a KSQL statement to display users who have more than one page of activity in a 1-minute window.

**ksql> SELECT \* FROM USER\_ACTIVITY**

**>WHERE WEBSITE\_HITS > 1**

**>EMIT CHANGES;**

****

1. In Jupyter, write a program to subscribe to the homepage topic generated by the stream/table in Question 4 and display the messages to the console.   
   NOTE: We could easily then write these to Elasticsearch, but we will not do that in this lab.

**import json**

**from confluent\_kafka import Consumer**

**consumer = Consumer({'bootstrap.servers' : 'broker:29092', 'group.id' : '\*'})**

**consumer.subscribe(['HOMEPAGE']) # topic**

**try:**

**while True:**

**msg = consumer.poll(1.0)**

**if msg is None:**

**continue**

**if msg.error():**

**print(f"Consumer error: {msg.error()}")**

**continue**

**raw = msg.value().decode('utf-8')**

**payload = json.loads(raw) # de-serialize JSON in the stream to a Python dictionary!**

**print(f"Received message: {payload}")**

**except KeyboardInterrupt:**

**consumer.close()**

****