Homework Problem Set K Submission Form

# Overview

|  |  |
| --- | --- |
| Your Name | Sana Khan |
| Your SU Email | [Skhan53@syr.edu](mailto:Skhan53@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.
2. What barriers or challenges did you encounter while completing this assignment?
3. How prepared were you to complete this assignment? What can you do to be better prepared?
4. 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

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**

A screenshot of a computer

Description automatically generated with medium confidence

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 Write a Drill query to display the total amount withdrawn by user.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

**SELECT a.`User`, SUM(Amount) AS TotalAmountWithdran**

**FROM kafka.`atm` as a WHERE status = ‘ok’**

**GROUP BY a.`UserA screenshot of a computer

Description automatically generated with medium confidence**

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.

**CREATE STREAM weblogs (Uri varchar, User varchar, TimeStamp bigint, Browser varchar, OS varchar)**

**>WITH (KAFKA\_TOPIC='weblogs', VALUE\_FORMAT='JSON', TIMESTAMP='TimeStamp');**  
A screen shot of a computer

Description automatically generated with medium confidence

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 = '/';**

**show queries;**

|  |
| --- |
|  |

A screenshot of a computer

Description automatically generated with medium confidence

|  |
| --- |
|  |
|  |
|  |

**Show streams;**

A screenshot of a computer

Description automatically generated with medium confidence

**Describe weblogs;**

A black screen with white text

Description automatically generated with low confidence

**SELECT \* FROM homepage EMIT CHANGES;**

A screen shot of a computer

Description automatically generated with medium confidence

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.

**SELECT Os, COUNT(\*) FROM weblogs WINDOW TUMBLING (SIZE 60 SECONDS) GROUP BY Os EMIT CHANGES;**

A picture containing text, screenshot

Description automatically generated

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.

**CREATE TABLE user\_activity AS SELECT User, COUNT(\*) FROM weblogs WINDOW TUMBLING (SIZE 60 SECONDS) GROUP BY User EMIT CHANGES;**

**A picture containing text, screenshot

Description automatically generated**

1. Write a KSQL statement to display users who have more than one page of activity in a 1-minute window.   
   A picture containing text, screenshot

   Description automatically generated
2. 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.   
   A picture containing text, screenshot, line, receipt

   Description automatically generated

NOTE: We could easily then write these to Elasticsearch, but we will not do that in this lab.