IST769 Homework Submission

Name: **Sathish Kumar Rajendiran**
SUID: **666555028**
Email: srajendi@syr.edu
Due Date: **09/14/2021**
Task: **Kafka and KSQL**
Homework #:**10**

Exercise(s):

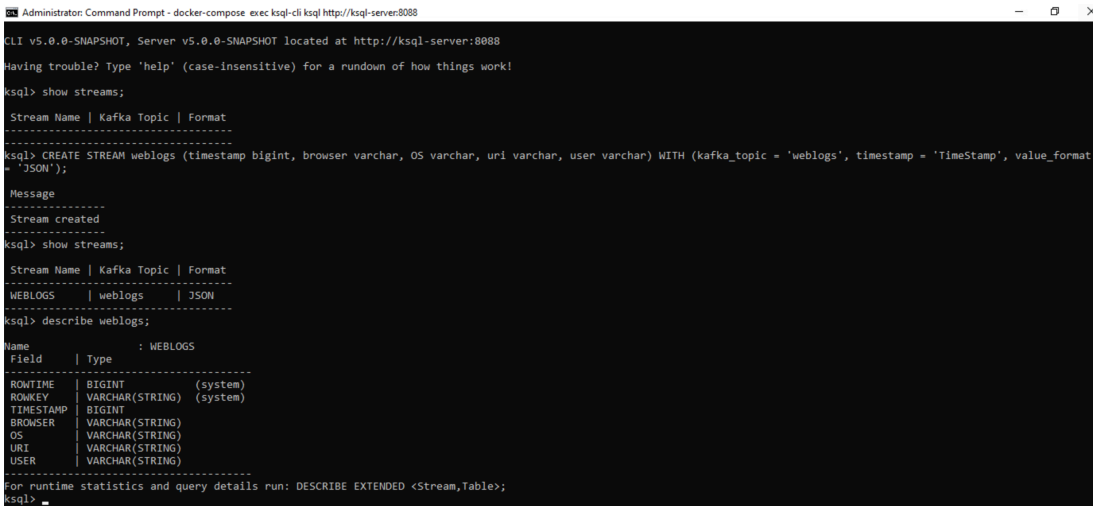
1. Write KSQL to create a stream named **weblogs** from the JSON keys in the **weblogs** Kafka topic. Make sure to set the TIMESTAMP property to the timestamp from the stream.

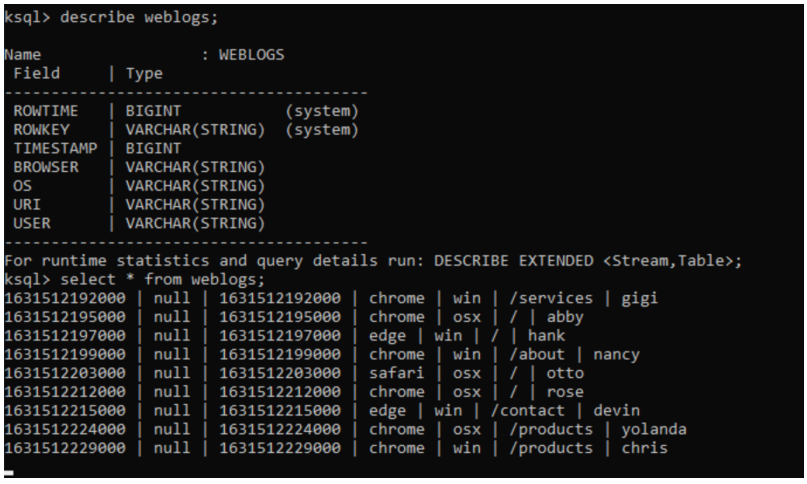
   Solution:

   -- Create Weblogs STREAM

   CREATE STREAM weblogs (timestamp bigint, browser varchar, OS varchar, uri varchar, user varchar) WITH (kafka_topic =’weblogs’, timestamp =’TimeStamp’, value_format =’JSON’);

   Evidence:



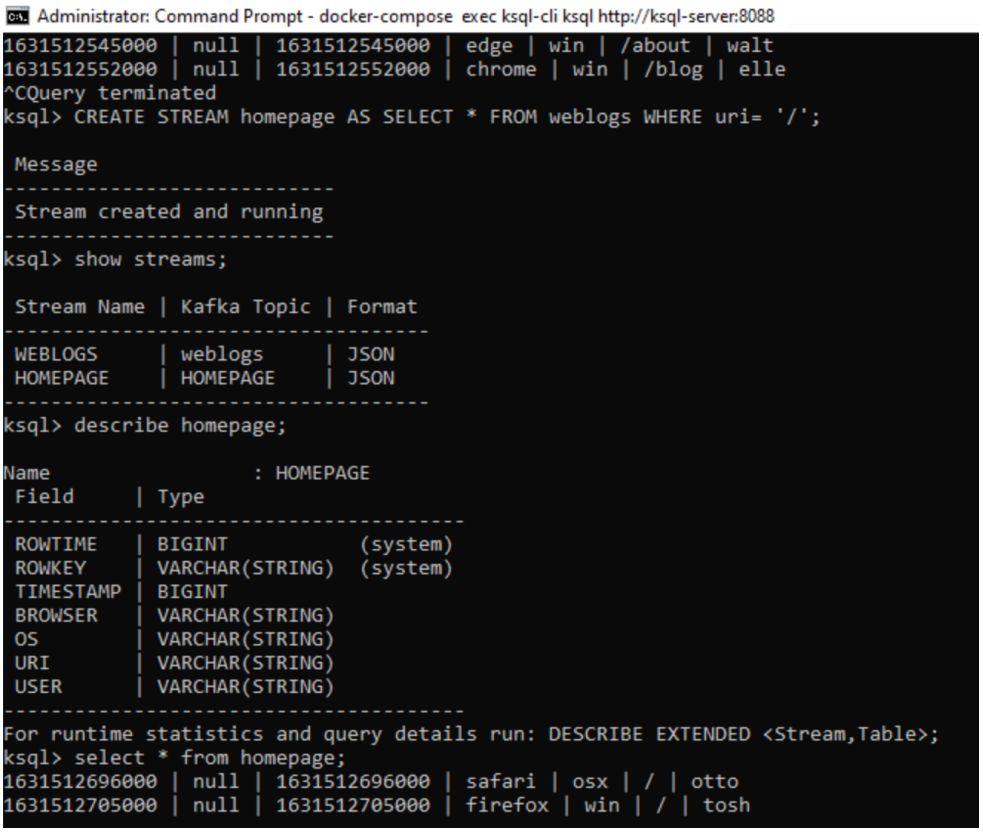


2. Write a KSQL statement create a persistent stream/table called **homepage** which only displays visitors to the root of the website (/). It should display all columns from the **weblogs** stream.

Solution:

```
-- KSQL Statement to create a persistent stream
CREATE STREAM homepage as SELECT * FROM weblogs WHERE uri ='/';
-- View the data
SELECT * FROM homepage;
```

Evidence:



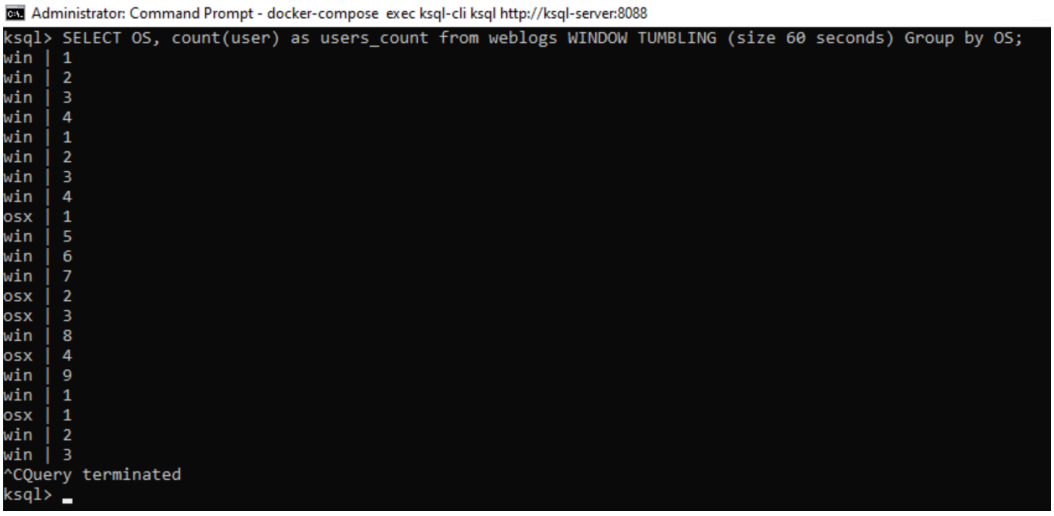
3. Write a KSQL statement to count operating systems users (os) in 60 second windows. After 60 seconds, the counter should reset, and counts should begin again.

Solution:

-- KSQL statement to count the number of operating system users in 60 second window

```
SELECT OS, count(user) as users_count from weblogs WINDOW TUMBLING (size 60 seconds) GROUP BY OS;
```

Evidence:



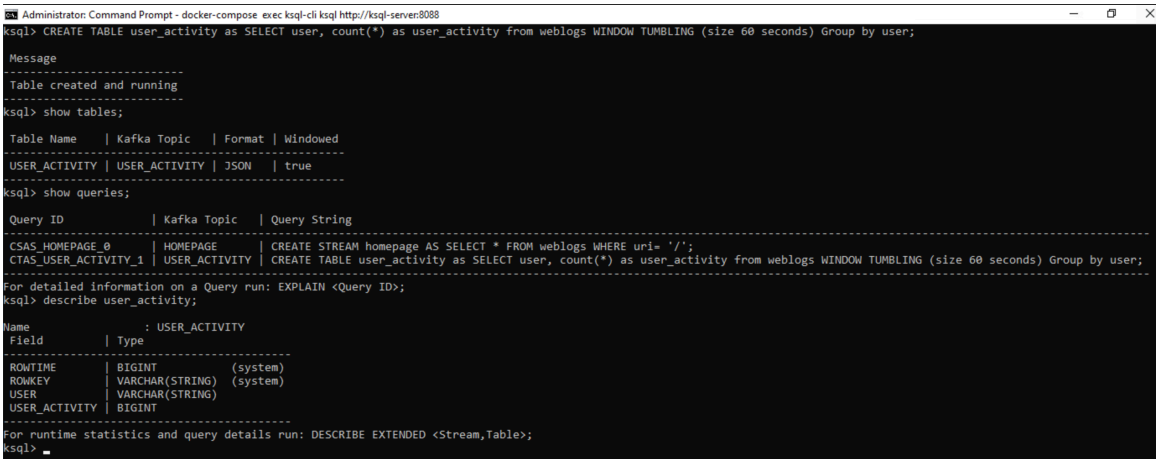
4. Write a KSQL persistent stream/table called **user_activity** which will display a count of user activity on the website within 1-minute sessions.

Solution:

-- create KSQL Persistent stream/table

```
CREATE TABLE user_activity as SELECT user, count (*) as user_activity from weblogs WINDOW TUMBLING (size 60 seconds) GROUP BY user;
```

Evidence:

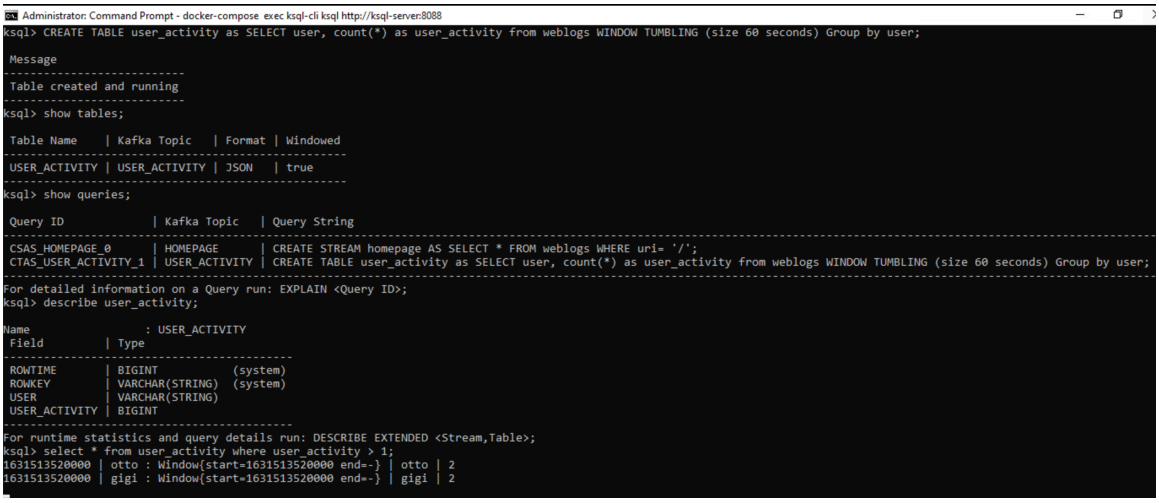


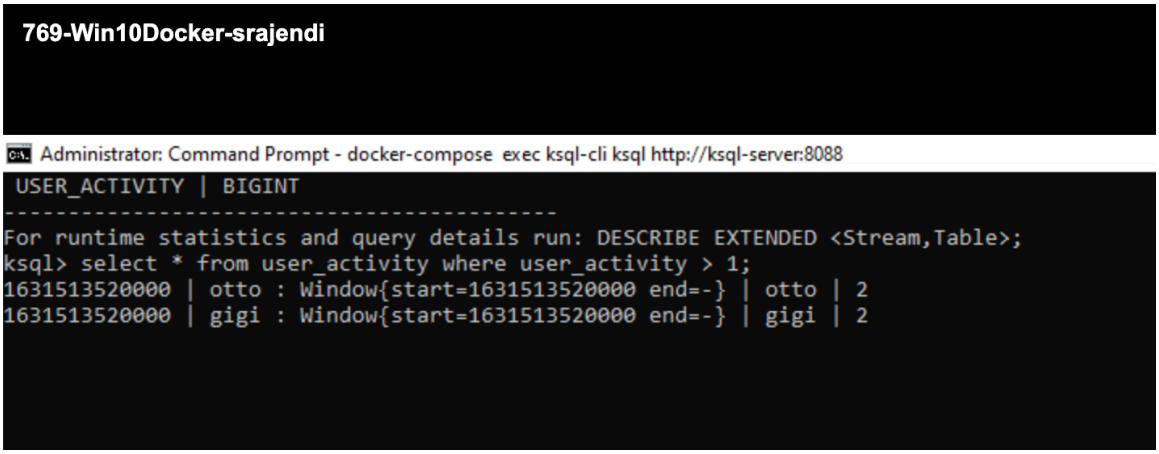
5. Write a KSQL statement to display users who have more than 1 pages of activity in a 1-minute session.

Solution:

```
-- Display results
Select * from user_activity where user_activity > 1;
```

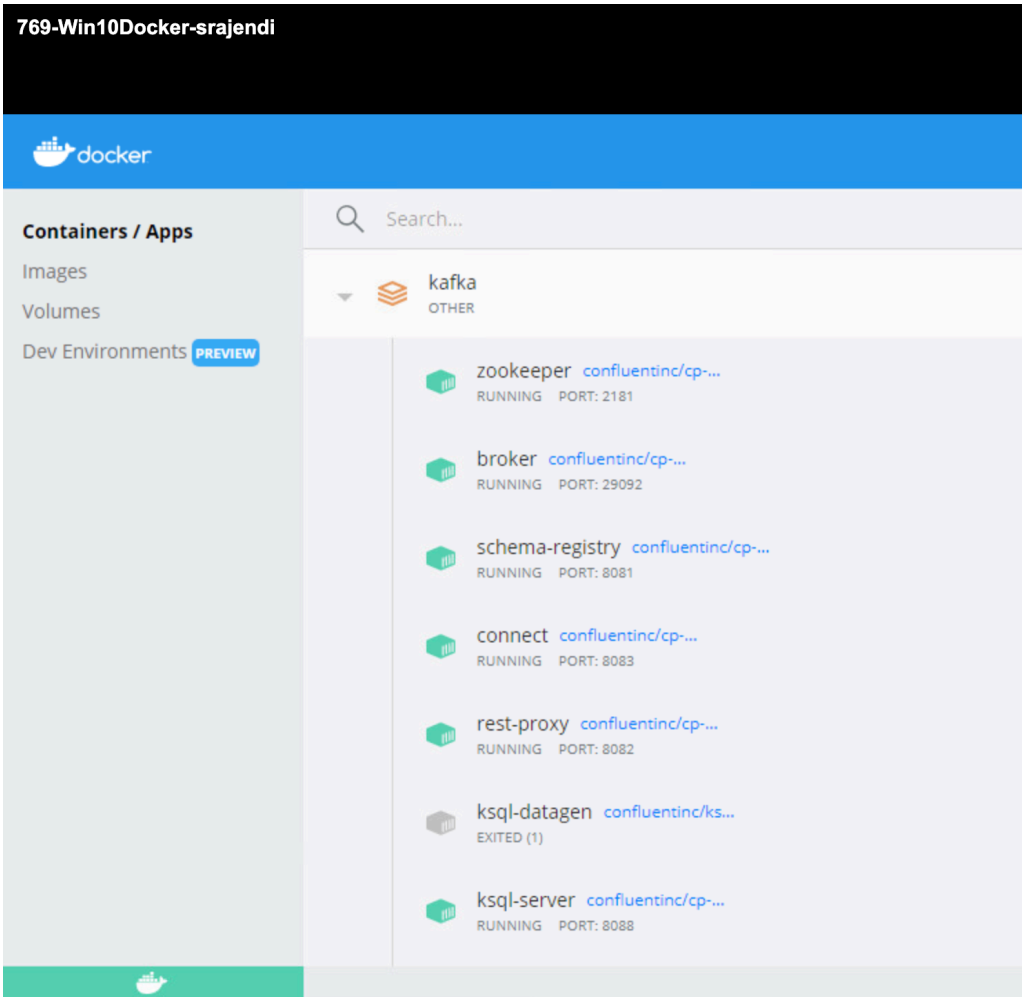
Evidence:





Appendix

```
Administrator: Command Prompt
Microsoft Windows [Version 10.0.19041.1165]
(c) Microsoft Corporation. All rights reserved.

C:\WINDOWS\system32>cd C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka

C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka>docker-compose ps
Name   Command   State   Ports
------------------------------

C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka>docker-compose up -d
Creating network "kafka_default" with the default driver
Creating zookeeper ... done
Creating broker    ... done
Creating schema-registry ... done
Creating rest-proxy      ... done
Creating connect         ... done
Creating ksql-datagen    ... done
Creating ksql-server     ... done
Creating control-center  ... done
Creating ksql-cli        ... done

C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka>docker-compose ps
     Name                   Command              State                                        Ports
-----------------------------------------------------------------------------------------------------------------------------------------------
broker            /etc/confluent/docker/run        Up      0.0.0.0:29092->29092/tcp,:::29092->29092/tcp, 0.0.0.0:9092->9092/tcp,:::9092->9092/tcp
connect           /etc/confluent/docker/run        Up      0.0.0.0:8083->8083/tcp,:::8083->8083/tcp, 9092/tcp
control-center    /etc/confluent/docker/run        Up      0.0.0.0:9021->9021/tcp,:::9021->9021/tcp
ksql-cli          /bin/sh                          Up
ksql-datagen      bash -c echo Waiting for K ...   Up
ksql-server       /etc/confluent/docker/run        Up      0.0.0.0:8088->8088/tcp,:::8088->8088/tcp
rest-proxy        /etc/confluent/docker/run        Up      0.0.0.0:8082->8082/tcp,:::8082->8082/tcp
schema-registry   /etc/confluent/docker/run        Up      0.0.0.0:8081->8081/tcp,:::8081->8081/tcp
zookeeper         /etc/confluent/docker/run        Up      0.0.0.0:2181->2181/tcp,:::2181->2181/tcp, 2888/tcp, 3888/tcp

C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka>
```



Administrator: Command Prompt - docker-compose run ksql-datagen bash

```
ksql-server       /etc/confluent/docker/run        Up      0.0.0.0:8088->8088/tcp,:::8088->8088/tcp
rest-proxy        /etc/confluent/docker/run        Up      0.0.0.0:8082->8082/tcp,:::8082->8082/tcp
schema-registry   /etc/confluent/docker/run        Up      0.0.0.0:8081->8081/tcp,:::8081->8081/tcp
zookeeper         /etc/confluent/docker/run        Up      0.0.0.0:2181->2181/tcp,:::2181->2181/tcp, 2888/tcp, 3888/tcp

C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka>docker-compose run ksql-datagen bash
Creating kafka_ksql-datagen_run ... done
root@ksql-datagen:/#
```

Administrator: Command Prompt - docker-compose run ksql-datagen bash

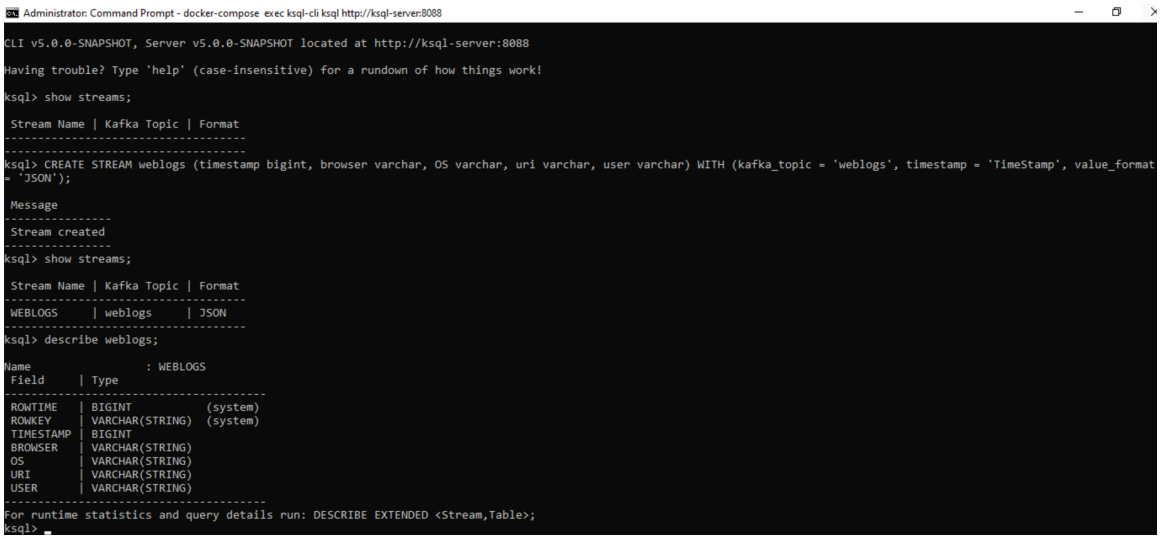
```
C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka>docker-compose run ksql-datagen bash
Creating kafka_ksql-datagen_run ... done
root@ksql-datagen:/# cd /datasets/streaming
root@ksql-datagen:/datasets/streaming# pip install kafka
Collecting kafka
  Downloading https://files.pythonhosted.org/packages/21/71/73286e748ac5045b6a669c2fe44b03ac4c5d3d2af9291c4c6fc76438a9a9/kafka-1.3.5-py2.py3-none-any.whl (207kB)
    100% |████████████████████████████████| 215kB 3.0MB/s
Installing collected packages: kafka
Successfully installed kafka-1.3.5
You are using pip version 8.1.2, however version 21.2.4 is available.
You should consider upgrading via the 'pip install --upgrade pip' command.
root@ksql-datagen:/datasets/streaming# ls
atm-datagen.py  kafka-atm-datagen.py  kafka-atm-stream.sh  kafka-weblogs-datagen.py  kafka-weblogs-stream.sh
root@ksql-datagen:/datasets/streaming#
```

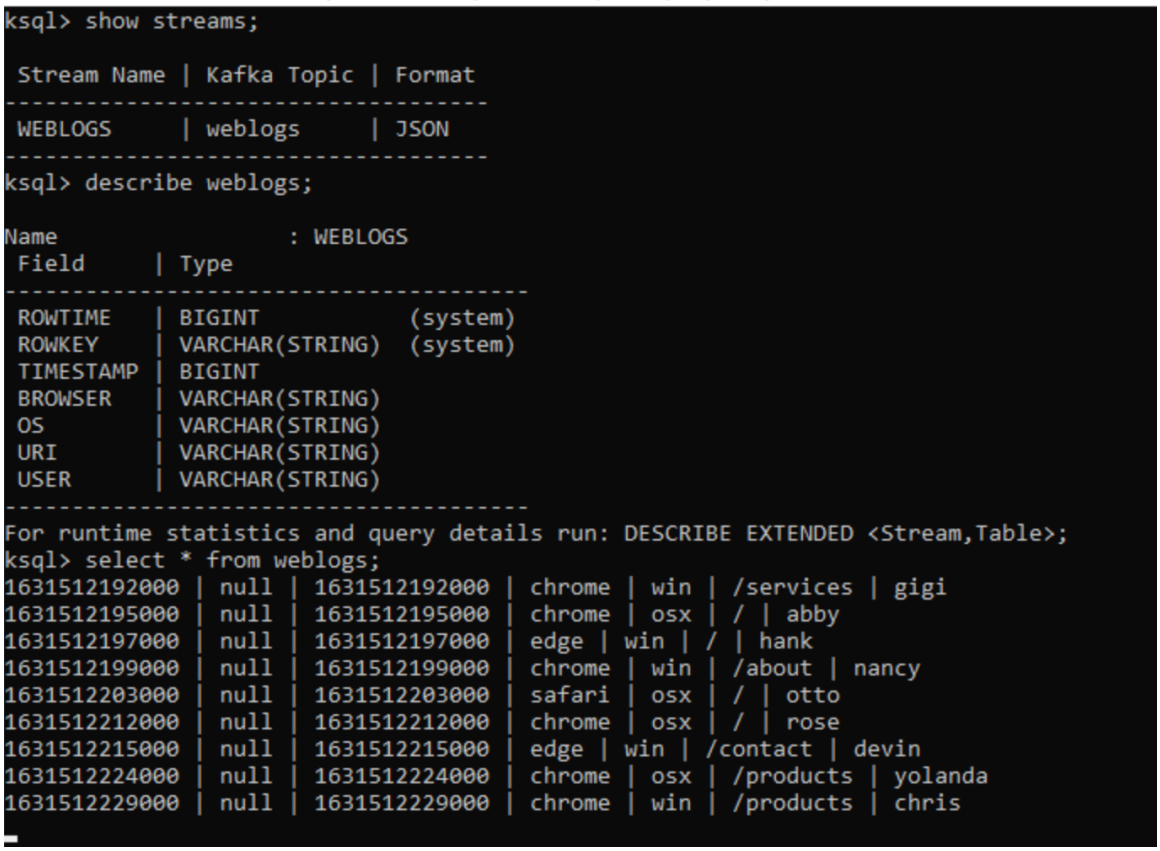
Administrator: Command Prompt - docker-compose run ksql-datagen bash

```
root@ksql-datagen:/datasets/streaming# ls
atm-datagen.py  kafka-atm-datagen.py  kafka-atm-stream.sh  kafka-weblogs-datagen.py  kafka-weblogs-stream.sh
root@ksql-datagen:/datasets/streaming# python kafka-weblogs-datagen.py
{"TimeStamp": 1631511189000, "Browser": "chrome", "OS": "osx", "Uri": "/", "User": "abby"}
{"TimeStamp": 1631511194000, "Browser": "edge", "OS": "win", "Uri": "/services", "User": "hank"}
{"TimeStamp": 1631511197000, "Browser": "firefox", "OS": "win", "Uri": "/services", "User": "tosh"}
{"TimeStamp": 1631511207000, "Browser": "safari", "OS": "osx", "Uri": "/blog", "User": "otto"}
{"TimeStamp": 1631511208000, "Browser": "chrome", "OS": "win", "Uri": "/blog", "User": "chris"}
{"TimeStamp": 1631511217000, "Browser": "firefox", "OS": "win", "Uri": "/services", "User": "surah"}
```

Administrator: Command Prompt - docker-compose exec ksql-cli ksql http://ksql-server:8088

```
Microsoft Windows [Version 10.0.19041.1165]
(c) Microsoft Corporation. All rights reserved.

C:\WINDOWS\system32>cd C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka

C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka>docker-compose exec ksql-cli ksql http://ksql-server:8088

                  ===========================================
                  =        _  __ _____  ____  _             =
                  =       | |/ // ____|/ __ \| |            =
                  =       | ' /| (___ | |  | | |            =
                  =       |  <  \___ \| |  | | |            =
                  =       | . \ ____) | |__| | |____        =
                  =       |_|\_\_____/ \___\_\______|       =
                  =                                         =
                  =  Streaming SQL Engine for Apache Kafka® =
                  ===========================================

Copyright 2017 Confluent Inc.

CLI v5.0.0-SNAPSHOT, Server v5.0.0-SNAPSHOT located at http://ksql-server:8088

Having trouble? Type 'help' (case-insensitive) for a rundown of how things work!

ksql>
```


Administrator: Command Prompt - docker-compose exec ksql-cli ksql http://ksql-server:8088
CLI v5.0.0-SNAPSHOT, Server v5.0.0-SNAPSHOT located at http://ksql-server:8088
Having trouble? Type 'help' (case-insensitive) for a rundown of how things work!
ksql> show streams;
Stream Name | Kafka Topic | Format
ksql> CREATE STREAM weblogs (timestamp bigint, browser varchar, OS varchar, uri varchar, user varchar) WITH (kafka_topic = 'weblogs', timestamp = 'TimeStamp', value_format = 'JSON');
Message
Stream created
ksql> show streams;
Stream Name | Kafka Topic | Format
WEBLOGS | weblogs | JSON
ksql> describe weblogs;
Name : WEBLOGS
Field | Type
ROWTIME | BIGINT (system)
ROWKEY | VARCHAR(STRING) (system)
TIMESTAMP | BIGINT
BROWSER | VARCHAR(STRING)
OS | VARCHAR(STRING)
URI | VARCHAR(STRING)
USER | VARCHAR(STRING)
For runtime statistics and query details run: DESCRIBE EXTENDED <Stream,Table>;
ksql>


Administrator: Command Prompt - docker-compose exec ksql-cli ksql http://ksql-server:8088
ksql> show streams;
Stream Name | Kafka Topic | Format
WEBLOGS | weblogs | JSON
ksql> describe weblogs;
Name : WEBLOGS
Field | Type
ROWTIME | BIGINT (system)
ROWKEY | VARCHAR(STRING) (system)
TIMESTAMP | BIGINT
BROWSER | VARCHAR(STRING)
OS | VARCHAR(STRING)
URI | VARCHAR(STRING)
USER | VARCHAR(STRING)
For runtime statistics and query details run: DESCRIBE EXTENDED <Stream,Table>;
ksql> select * from weblogs;
1631512192000 | null | 1631512192000 | chrome | win | /services | gigi
1631512195000 | null | 1631512195000 | chrome | osx | / | abby
1631512197000 | null | 1631512197000 | edge | win | / | hank
1631512199000 | null | 1631512199000 | chrome | win | /about | nancy
1631512203000 | null | 1631512203000 | safari | osx | / | otto
1631512212000 | null | 1631512212000 | chrome | osx | / | rose
1631512215000 | null | 1631512215000 | edge | win | /contact | devin
1631512224000 | null | 1631512224000 | chrome | osx | /products | yolanda
1631512229000 | null | 1631512229000 | chrome | win | /products | chris

Administrator: Command Prompt - docker-compose exec ksql-cli ksql http://ksql-server:8088

```
1631512545000 | null | 1631512545000 | edge | win | /about | walt
1631512552000 | null | 1631512552000 | chrome | win | /blog | elle
^CQuery terminated
ksql> CREATE STREAM homepage AS SELECT * FROM weblogs WHERE uri= '/';

 Message
----------------------------
 Stream created and running
----------------------------
ksql> show streams;

 Stream Name | Kafka Topic | Format
-----------------------------------
 WEBLOGS     | weblogs     | JSON
 HOMEPAGE    | HOMEPAGE    | JSON
-----------------------------------
ksql> describe homepage;

Name                 : HOMEPAGE
 Field     | Type
---------------------------------------
 ROWTIME   | BIGINT           (system)
 ROWKEY    | VARCHAR(STRING)  (system)
 TIMESTAMP | BIGINT
 BROWSER   | VARCHAR(STRING)
 OS        | VARCHAR(STRING)
 URI       | VARCHAR(STRING)
 USER      | VARCHAR(STRING)
---------------------------------------
For runtime statistics and query details run: DESCRIBE EXTENDED <Stream,Table>;
ksql> select * from homepage;
1631512696000 | null | 1631512696000 | safari | osx | / | otto
1631512705000 | null | 1631512705000 | firefox | win | / | tosh
```

Administrator: Command Prompt - docker-compose exec ksql-cli ksql http://ksql-server:8088

```
---------------------------------------
For runtime statistics and query details run: DESCRIBE EXTENDED <Stream,Table>;
ksql> select * from homepage;
1631512696000 | null | 1631512696000 | safari | osx | / | otto
1631512705000 | null | 1631512705000 | firefox | win | / | tosh
1631512724000 | null | 1631512724000 | edge | win | / | devin
1631512736000 | null | 1631512736000 | firefox | win | / | patty
1631512741000 | null | 1631512741000 | chrome | win | / | nancy
^CQuery terminated
ksql> SELECT OS, count(user) as users_count from weblogs WINDOW TUMBLING (size 60 seconds) Group by OS;
win | 1
win | 2
win | 3
win | 4
win | 1
win | 2
win | 3
```

Administrator: Command Prompt - docker-compose exec ksql-cli ksql http://ksql-server:8088

```
ksql> SELECT OS, count(user) as users_count from weblogs WINDOW TUMBLING (size 60 seconds) Group by OS;
win | 1
win | 2
win | 3
win | 4
win | 1
win | 2
win | 3
win | 4
osx | 1
win | 5
win | 6
win | 7
osx | 2
osx | 3
win | 8
osx | 4
win | 9
win | 1
osx | 1
win | 2
win | 3
^CQuery terminated
ksql>
```

Administrator: Command Prompt - docker-compose exec ksql-cli ksql http://ksql-server:8088

```
ksql> CREATE TABLE user_activity as SELECT user, count(*) as user_activity from weblogs WINDOW TUMBLING (size 60 seconds) Group by user;

 Message
---------------------------
 Table created and running
---------------------------
ksql> show tables;

 Table Name    | Kafka Topic   | Format | Windowed
---------------------------------------------------
 USER_ACTIVITY | USER_ACTIVITY | JSON   | true
---------------------------------------------------
ksql> show queries;

 Query ID             | Kafka Topic   | Query String
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------
 CSAS_HOMEPAGE_0      | HOMEPAGE      | CREATE STREAM homepage AS SELECT * FROM weblogs WHERE uri= '/';
 CTAS_USER_ACTIVITY_1 | USER_ACTIVITY | CREATE TABLE user_activity as SELECT user, count(*) as user_activity from weblogs WINDOW TUMBLING (size 60 seconds) Group by user;
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------
For detailed information on a Query run: EXPLAIN <Query ID>;
ksql> describe user_activity;

Name                 : USER_ACTIVITY
 Field         | Type
-------------------------------------------
 ROWTIME       | BIGINT           (system)
 ROWKEY        | VARCHAR(STRING)  (system)
 USER          | VARCHAR(STRING)
 USER_ACTIVITY | BIGINT
-------------------------------------------
For runtime statistics and query details run: DESCRIBE EXTENDED <Stream,Table>;
ksql>
```

Administrator: Command Prompt - docker-compose exec ksql-cli ksql http://ksql-server:8088

```
ksql> CREATE TABLE user_activity as SELECT user, count(*) as user_activity from weblogs WINDOW TUMBLING (size 60 seconds) Group by user;

 Message
---------------------------
 Table created and running
---------------------------
ksql> show tables;

 Table Name    | Kafka Topic   | Format | Windowed
---------------------------------------------------
 USER_ACTIVITY | USER_ACTIVITY | JSON   | true
---------------------------------------------------
ksql> show queries;

 Query ID             | Kafka Topic   | Query String
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------
 CSAS_HOMEPAGE_0      | HOMEPAGE      | CREATE STREAM homepage AS SELECT * FROM weblogs WHERE uri= '/';
 CTAS_USER_ACTIVITY_1 | USER_ACTIVITY | CREATE TABLE user_activity as SELECT user, count(*) as user_activity from weblogs WINDOW TUMBLING (size 60 seconds) Group by user;
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------
For detailed information on a Query run: EXPLAIN <Query ID>;
ksql> describe user_activity;

Name                 : USER_ACTIVITY
 Field         | Type
-------------------------------------------
 ROWTIME       | BIGINT           (system)
 ROWKEY        | VARCHAR(STRING)  (system)
 USER          | VARCHAR(STRING)
 USER_ACTIVITY | BIGINT
-------------------------------------------
For runtime statistics and query details run: DESCRIBE EXTENDED <Stream,Table>;
ksql> select * from user_activity where user_activity > 1;
1631513520000 | otto : Window{start=1631513520000 end=-} | otto | 2
1631513520000 | gigi : Window{start=1631513520000 end=-} | gigi | 2
```

769-Win10Docker-srajendi

Administrator: Command Prompt - docker-compose exec ksql-cli ksql http://ksql-server:8088

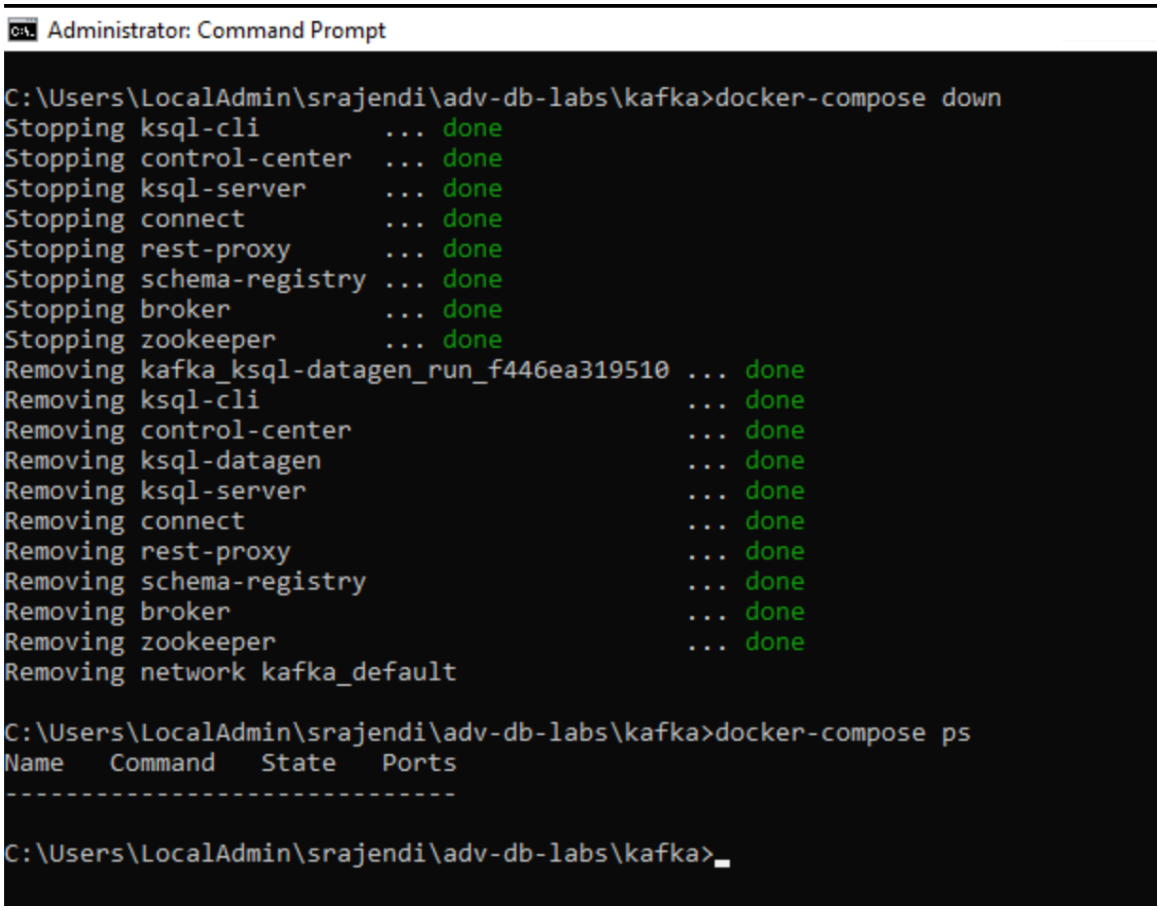
```
 USER_ACTIVITY | BIGINT
----------------------------------------
For runtime statistics and query details run: DESCRIBE EXTENDED <Stream,Table>;
ksql> select * from user_activity where user_activity > 1;
1631513520000 | otto : Window{start=1631513520000 end=-} | otto | 2
1631513520000 | gigi : Window{start=1631513520000 end=-} | gigi | 2
```

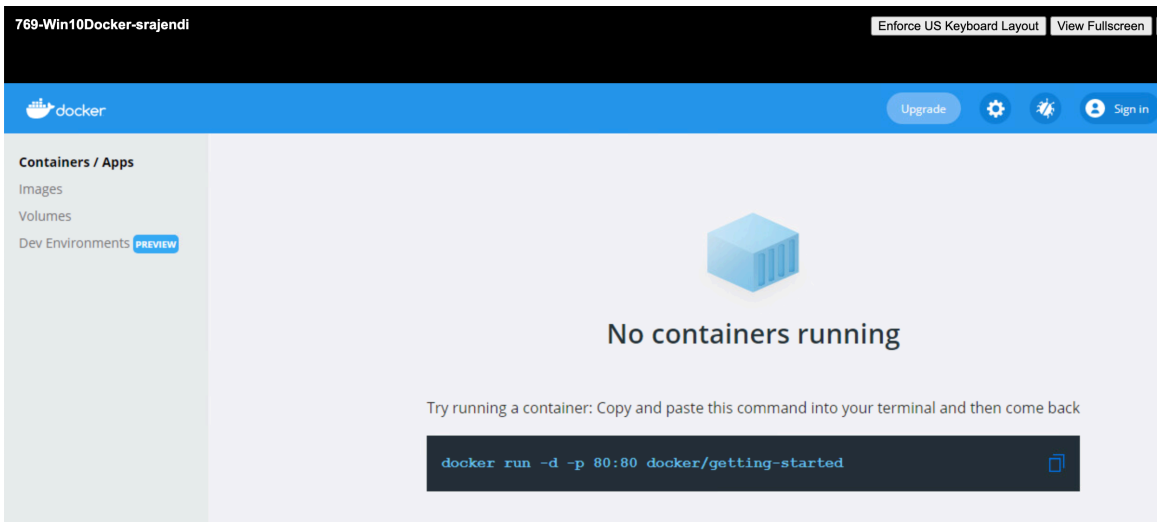
Administrator: Command Prompt

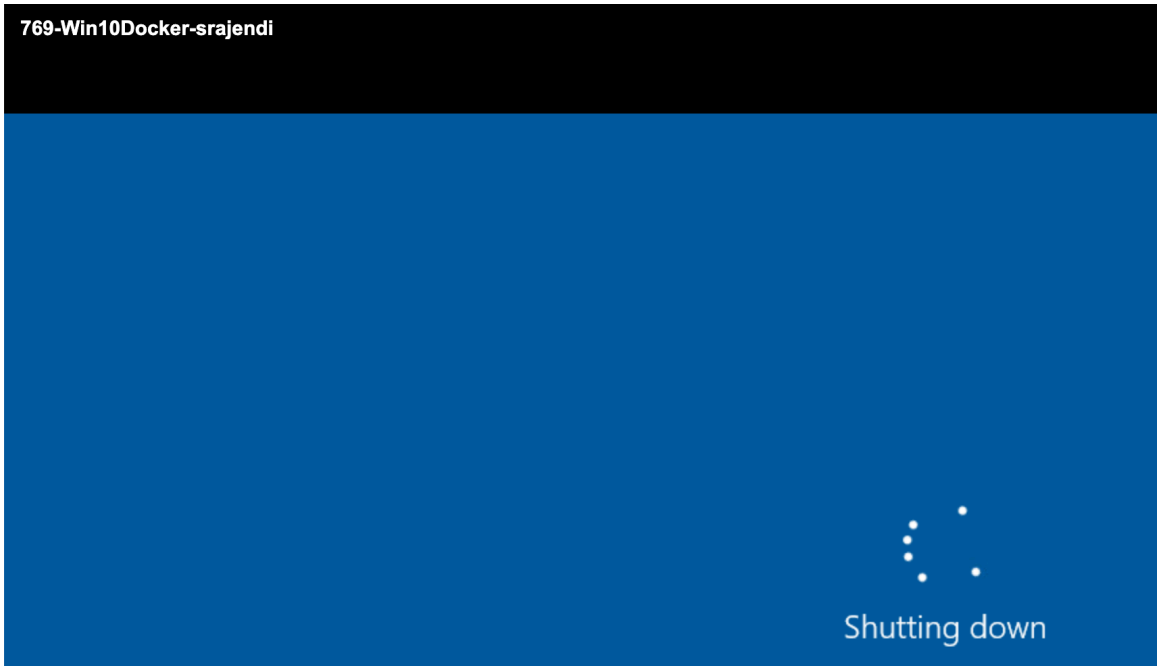
```
{"TimeStamp": 1631517346000, "Browser": "safari", "OS": "osx", "Uri": "/services", "User": "vaibhav"}
{"TimeStamp": 1631517353000, "Browser": "chrome", "OS": "osx", "Uri": "/products", "User": "lisa"}
{"TimeStamp": 1631517361000, "Browser": "edge", "OS": "win", "Uri": "/about", "User": "walt"}
{"TimeStamp": 1631517366000, "Browser": "firefox", "OS": "osx", "Uri": "/blog", "User": "mike"}
{"TimeStamp": 1631517373000, "Browser": "edge", "OS": "win", "Uri": "/about", "User": "walt"}
{"TimeStamp": 1631517381000, "Browser": "edge", "OS": "win", "Uri": "/blog", "User": "hank"}
{"TimeStamp": 1631517390000, "Browser": "chrome", "OS": "win", "Uri": "/blog", "User": "karley"}
{"TimeStamp": 1631517401000, "Browser": "firefox", "OS": "win", "Uri": "/blog", "User": "tosh"}
{"TimeStamp": 1631517406000, "Browser": "firefox", "OS": "win", "Uri": "/", "User": "zeke"}
{"TimeStamp": 1631517415000, "Browser": "safari", "OS": "osx", "Uri": "/about", "User": "vaibhav"}
{"TimeStamp": 1631517417000, "Browser": "chrome", "OS": "win", "Uri": "/", "User": "gigi"}
{"TimeStamp": 1631517419000, "Browser": "chrome", "OS": "win", "Uri": "/blog", "User": "ida"}
{"TimeStamp": 1631517428000, "Browser": "chrome", "OS": "osx", "Uri": "/", "User": "abby"}
{"TimeStamp": 1631517430000, "Browser": "edge", "OS": "win", "Uri": "/", "User": "walt"}
^Croot@ksql-datagen:/datasets/streaming# exit
exit

C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka>
```

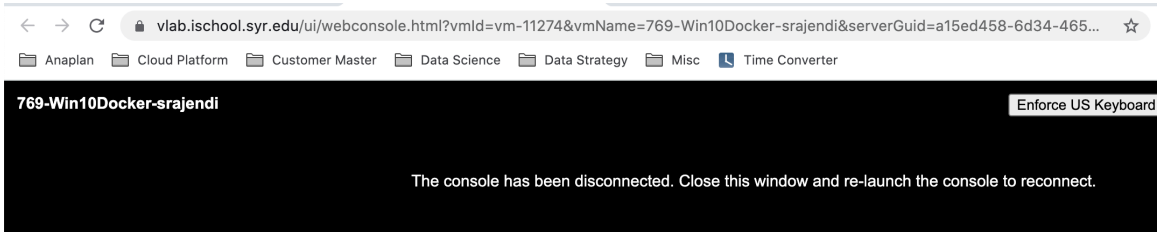
Administrator: Command Prompt

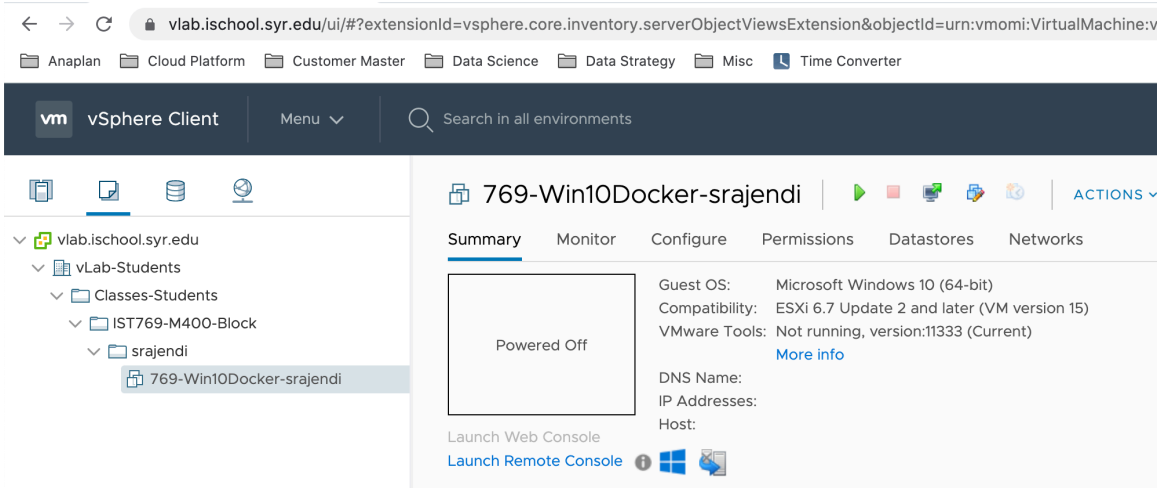
```
1631516940000 | elle : Window{start=1631516940000 end=-} | elle | 2
1631516940000 | elle : Window{start=1631516940000 end=-} | elle | 3
1631516940000 | elle : Window{start=1631516940000 end=-} | elle | 4
1631517000000 | vaibhav : Window{start=1631517000000 end=-} | vaibhav | 2
1631517000000 | vaibhav : Window{start=1631517000000 end=-} | vaibhav | 3
1631517000000 | devin : Window{start=1631517000000 end=-} | devin | 2
1631517000000 | patty : Window{start=1631517000000 end=-} | patty | 2
1631517120000 | lisa : Window{start=1631517120000 end=-} | lisa | 2
1631517180000 | lisa : Window{start=1631517180000 end=-} | lisa | 2
1631517180000 | ida : Window{start=1631517180000 end=-} | ida | 2
1631517240000 | karley : Window{start=1631517240000 end=-} | karley | 2
1631517240000 | karley : Window{start=1631517240000 end=-} | karley | 3
1631517300000 | nancy : Window{start=1631517300000 end=-} | nancy | 2
1631517300000 | nancy : Window{start=1631517300000 end=-} | nancy | 3
1631517300000 | lisa : Window{start=1631517300000 end=-} | lisa | 2
1631517360000 | walt : Window{start=1631517360000 end=-} | walt | 2
^CQuery terminated
ksql> exit
Exiting KSQL.

C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka>
```


Administrator: Command Prompt
C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka>docker-compose down
Stopping ksql-cli ... done
Stopping control-center ... done
Stopping ksql-server ... done
Stopping connect ... done
Stopping rest-proxy ... done
Stopping schema-registry ... done
Stopping broker ... done
Stopping zookeeper ... done
Removing kafka_ksql-datagen_run_f446ea319510 ... done
Removing ksql-cli ... done
Removing control-center ... done
Removing ksql-datagen ... done
Removing ksql-server ... done
Removing connect ... done
Removing rest-proxy ... done
Removing schema-registry ... done
Removing broker ... done
Removing zookeeper ... done
Removing network kafka_default
C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka>docker-compose ps
Name Command State Ports
C:\Users\LocalAdmin\srajendi\adv-db-labs\kafka>


769-Win10Docker-srajendi
Enforce US Keyboard Layout
View Fullscreen
docker
Upgrade
Sign in
Containers / Apps
Images
Volumes
Dev Environments
PREVIEW
No containers running
Try running a container: Copy and paste this command into your terminal and then come back
docker run -d -p 80:80 docker/getting-started


769-Win10Docker-srajendi
Shutting down


vlab.ischool.syr.edu/ui/webconsole.html?vmId=vm-11274&vmName=769-Win10Docker-srajendi&serverGuid=a15ed458-6d34-465...
Anaplan
Cloud Platform
Customer Master
Data Science
Data Strategy
Misc
Time Converter
769-Win10Docker-srajendi
Enforce US Keyboard
The console has been disconnected. Close this window and re-launch the console to reconnect.


vlab.ischool.syr.edu/ui/#?extensionId=vsphere.core.inventory.serverObjectViewsExtension&objectId=urn:vmomi:VirtualMachine:v
Anaplan
Cloud Platform
Customer Master
Data Science
Data Strategy
Misc
Time Converter
vSphere Client
Menu
Search in all environments
vlab.ischool.syr.edu
vLab-Students
Classes-Students
IST769-M400-Block
srajendi
769-Win10Docker-srajendi
769-Win10Docker-srajendi
ACTIONS
Summary
Monitor
Configure
Permissions
Datastores
Networks
Powered Off
Guest OS: Microsoft Windows 10 (64-bit)
Compatibility: ESXi 6.7 Update 2 and later (VM version 15)
VMware Tools: Not running, version:11333 (Current)
More info
DNS Name:
IP Addresses:
Host:
Launch Web Console
Launch Remote Console