1. Kafka set up using Docker:

Using JVM Based Apache Kafka Docker Image

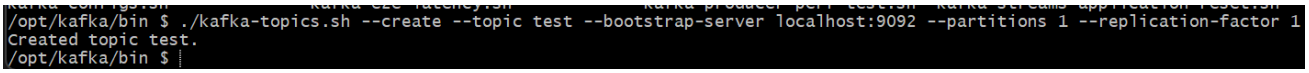
Get the Docker image:

```
$ docker pull apache/kafka:4.1.1
```

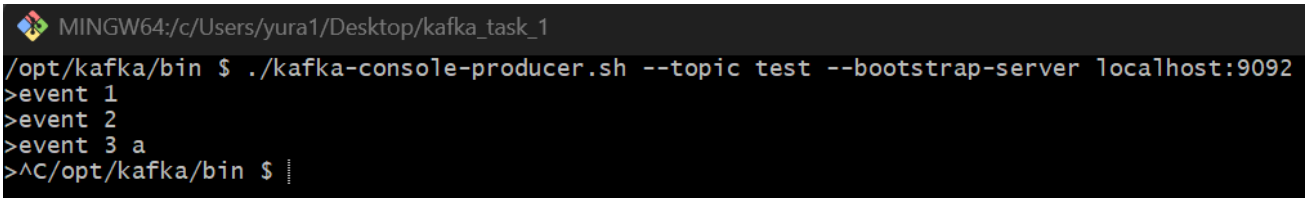
Start the Kafka Docker container:

```
$ docker run -p 9092:9092 apache/kafka:4.1.1
```

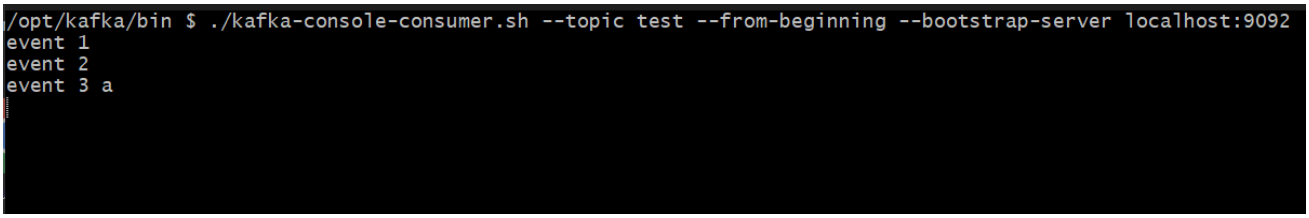
2. Topic creation:



3. Adding a few events:



4. Read events:



Clickstream Data Pipeline using Confluent Platform, ksqlDB, Kafka Connect, Elasticsearch, and Grafana Set up using WSL:

```
[+] Running 11/11
 ✔Network m10_kafkabasics_sql_local-master_default  Created
 ✔Container elasticsearch                           Started
 ✔Container zookeeper                               Started
 ✔Container grafana                                 Started
 ✔Container tools                                   Started
 ✔Container kafka                                   Started
 ✔Container schema-registry                         Started
 ✔Container ksqldb-server                           Started
 ✔Container connect                                 Started
 ✔Container control-center                          Started
 ✔Container ksqldb-cli                              Started
yuraladin@pc:~/m10_kafkabasics_sql_local-master$ docker-compose ps -a
NAME             IMAGE                                             COMMAND                  SERVICE          CREATED          STATUS          PORTS
connect          cnfldemos/cp-server-connect-datagen:0.6.4-7.6.0   "/etc/confluent/dock…"   connect          32 seconds ago   Up 29 seconds   0.0.0.0:8083->8083/tcp, [::]:8083->8083/tcp, 9092/tcp
control-center   confluentinc/cp-enterprise-control-center:7.9.0   "/etc/confluent/dock…"   control-center   32 seconds ago   Up 28 seconds   0.0.0.0:9021->9021/tcp, [::]:9021->9021/tcp
elasticsearch    docker.elastic.co/elasticsearch/elasticsearch:6.3.0   "/usr/local/bin/dock…"   elasticsearch    33 seconds ago   Up 30 seconds   0.0.0.0:9200->9200/tcp, [::]:9200->9200/tcp, 9300/tcp
grafana          grafana/grafana:5.2.4                             "/run.sh"                grafana          33 seconds ago   Up 30 seconds   0.0.0.0:3000->3000/tcp, [::]:3000->3000/tcp
kafka            confluentinc/cp-server:7.9.0                      "/etc/confluent/dock…"   kafka            32 seconds ago   Up 30 seconds   9092/tcp
ksqldb-cli       confluentinc/cp-ksqldb-cli:7.9.0                  "/bin/sh"                ksqldb-cli       32 seconds ago   Up 28 seconds
ksqldb-server    confluentinc/cp-ksqldb-server:7.9.0               "bash -c '\n  echo \"W…"   ksqldb-server    32 seconds ago   Up 29 seconds   0.0.0.0:8088->8088/tcp, [::]:8088->8088/tcp
schema-registry  confluentinc/cp-schema-registry:7.9.0             "/etc/confluent/dock…"   schema-registry  32 seconds ago   Up 29 seconds   8081/tcp
tools            cnfltraining/training-tools:5.4                   "/bin/bash"              tools            33 seconds ago   Up 31 seconds
zookeeper        confluentinc/cp-zookeeper:7.9.0                   "/etc/confluent/dock…"   zookeeper        33 seconds ago   Up 30 seconds   2181/tcp, 2888/tcp, 3888/tcp
yuraladin@pc:~/m10_kafkabasics_sql_local-master$
```

PS: I had to change version of elasticsearch to docker.elastic.co/elasticsearch/elasticsearch:7.17.14 and use connector confluentinc/kafka-connect-elasticsearch:latest to allow it work.

Topics:

```
yuraladin@pc:~/m10_kafkabasics_sql_local-master$ docker-compose exec ksqldb-cli ksql http://ksqldb-server:8088
Debug Options:

                  ===========================================
                  =        _              _ ____  ____      =
                  =      | | _____  __ _| |  _ \| __ )     =
                  =      | |/ / __|/ _` | | | | |  _ \     =
                  =      |   <\__ \ (_| | | |_| | |_) |    =
                  =      |_|\_\___/\__, |_|____/|____/     =
                  =                   |_|                   =
                  =        The Database purpose-built       =
                  =        for stream processing apps       =
                  ===========================================

Copyright 2017-2022 Confluent Inc.

CLI v7.9.0, Server v7.9.0 located at http://ksqldb-server:8088
Server Status: RUNNING

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

ksql> show topics;

 Kafka Topic                 | Partitions | Partition Replicas
------------------------------------------------------------------
 docker-connect-configs      | 1          | 1
 docker-connect-offsets      | 25         | 1
 docker-connect-status       | 5          | 1
 ksql-clusterksql_processing_log | 1          | 1
------------------------------------------------------------------
ksql>
```

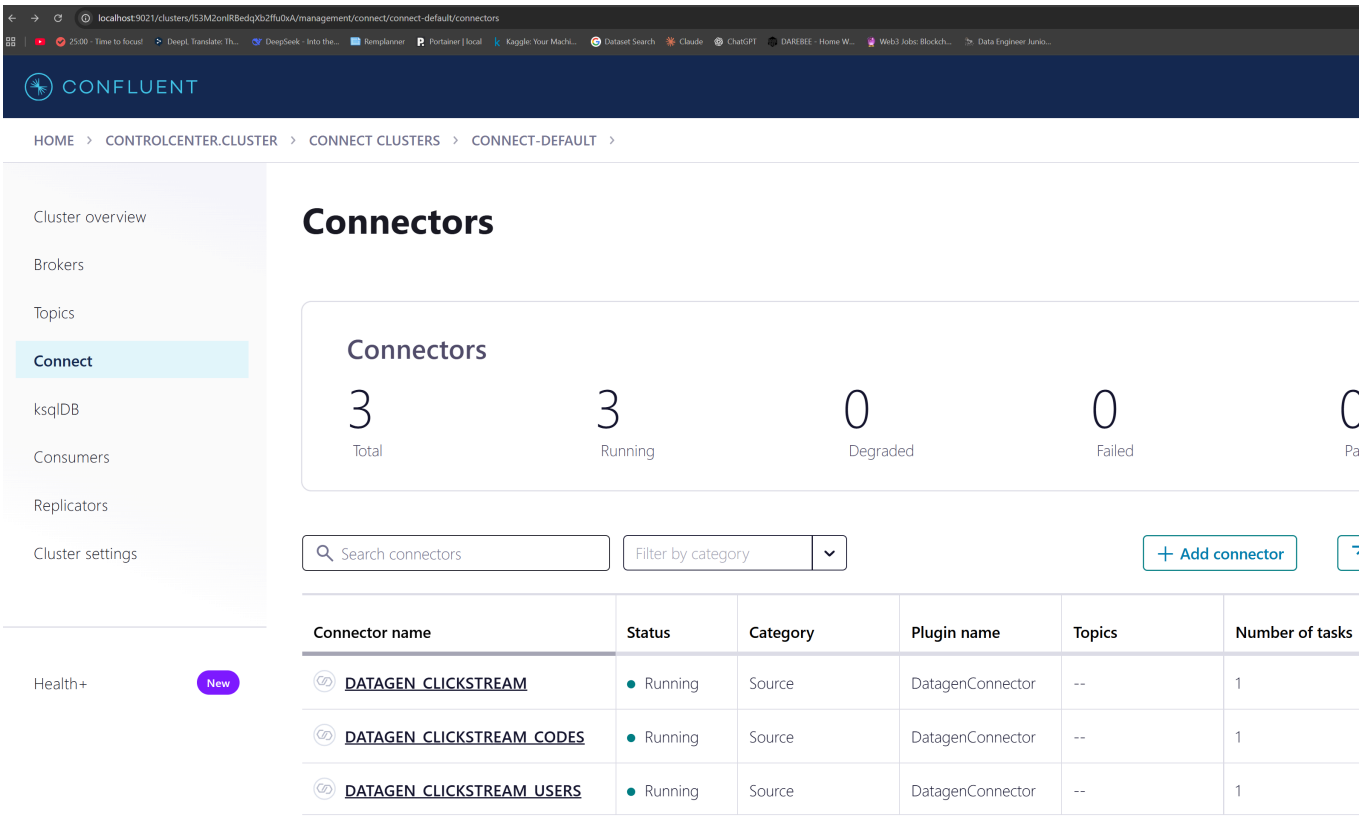
Running sql script:

```
ksql> RUN SCRIPT '/scripts/create-connectors.sql';

CREATE SOURCE CONNECTOR datagen_clickstream_codes WITH ('connector.class'='io.conflu
 Message
--------------------------------------------
 Created connector DATAGEN_CLICKSTREAM_CODES
--------------------------------------------

CREATE SOURCE CONNECTOR datagen_clickstream_users WITH ('connector.class'='io.conflu
 Message
--------------------------------------------
 Created connector DATAGEN_CLICKSTREAM_USERS
--------------------------------------------

CREATE SOURCE CONNECTOR datagen_clickstream WITH ('connector.class'='io.confluent.ka
 Message
----------------------------------------
 Created connector DATAGEN_CLICKSTREAM
----------------------------------------
ksql>
```

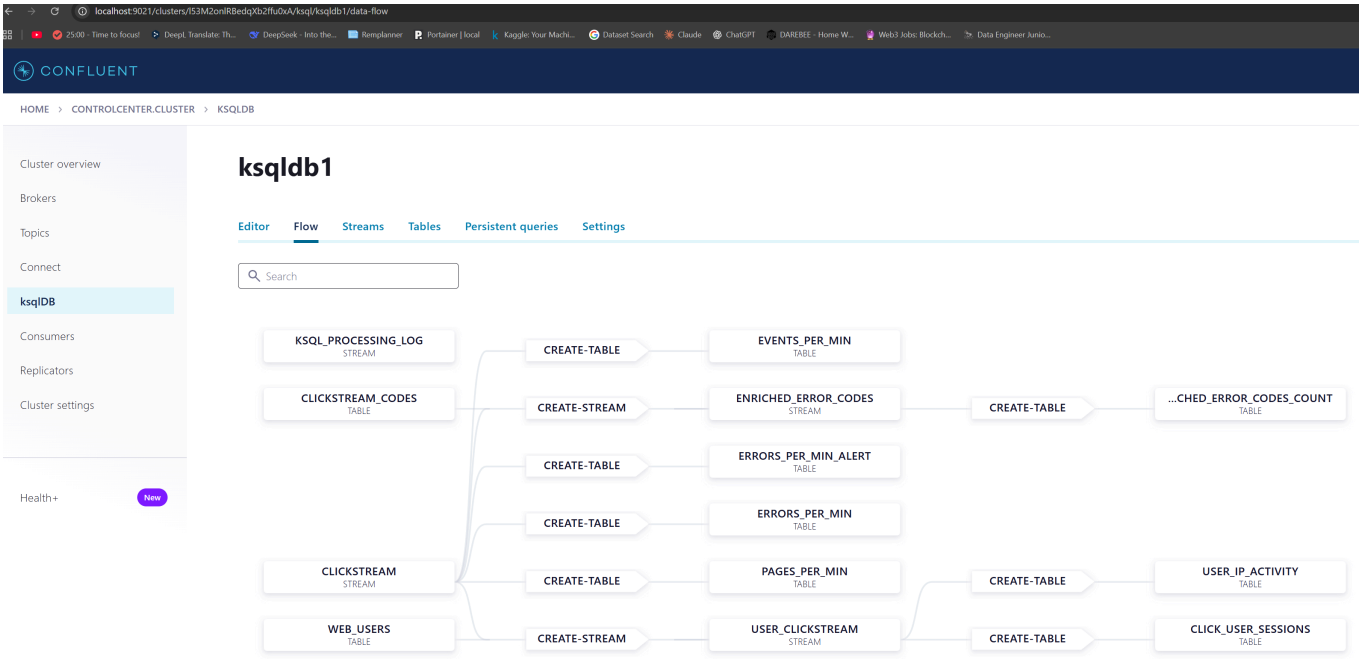
Example of events in clickstream topics:

```
ksql> print clickstream limit 3;
Key format: HOPPING(JSON) or TUMBLING(JSON) or HOPPING(KAFKA_STRING) or TUMBLING(KAFKA_STRING) or KAFKA_STRING
Value format: JSON or KAFKA_STRING
rowtime: 2026/01/03 16:11:42.504 Z, key: [122.173@3328501201639714867/-], value: {"ip":"122.173.165.203","userid":23,"r
mote_user":"-","time":"1","_time":1,"request":"GET /site/login.html HTTP/1.1","status":"200","bytes":"4196","referrer":
-","agent":"Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"}, partition: 0
rowtime: 2026/01/03 16:11:42.806 Z, key: [222.245@3328502296856376376/-], value: {"ip":"222.245.174.248","userid":22,"r
mote_user":"-","time":"11","_time":11,"request":"GET /site/user_status.html HTTP/1.1","status":"407","bytes":"2048","re
errer":"-","agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/59.0.3071.1
5 Safari/537.36"}, partition: 0
rowtime: 2026/01/03 16:11:42.957 Z, key: [222.245@3328502296856375858/-], value: {"ip":"222.245.174.222","userid":19,"r
mote_user":"-","time":"21","_time":21,"request":"GET /site/login.html HTTP/1.1","status":"200","bytes":"4006","referre
:"-","agent":"Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"}, partition: 0
Topic printing ceased
ksql>
ksql> print clickstream_codes limit 3;
Key format: KAFKA_INT
Value format: JSON or KAFKA_STRING
rowtime: 2026/01/03 16:11:42.542 Z, key: 407, value: {"code":407,"definition":"Proxy authentication required"}, partitio
n: 0
rowtime: 2026/01/03 16:11:42.740 Z, key: 404, value: {"code":404,"definition":"Page not found"}, partition: 0
rowtime: 2026/01/03 16:11:42.751 Z, key: 302, value: {"code":302,"definition":"Redirect"}, partition: 0
Topic printing ceased
ksql> print clickstream_users limit 3;
Key format: KAFKA_INT
Value format: JSON or KAFKA_STRING
rowtime: 2026/01/03 16:11:42.403 Z, key: 1, value: {"user_id":1,"username":"AndySims_345324","registered_at":14576505222
82,"first_name":"Ferd","last_name":"Pask","city":"New York","level":"Silver"}, partition: 0
rowtime: 2026/01/03 16:11:42.714 Z, key: 2, value: {"user_id":2,"username":"bobk_43","registered_at":1437809180321,"firs
t_name":"Oriana","last_name":"Rockhill","city":"New York","level":"Silver"}, partition: 0
rowtime: 2026/01/03 16:11:42.740 Z, key: 3, value: {"user_id":3,"username":"LukeWaters_23","registered_at":1444784714328
,"first_name":"Dimitri","last_name":"De Banke","city":"Frankfurt","level":"Gold"}, partition: 0
Topic printing ceased
ksql>
```

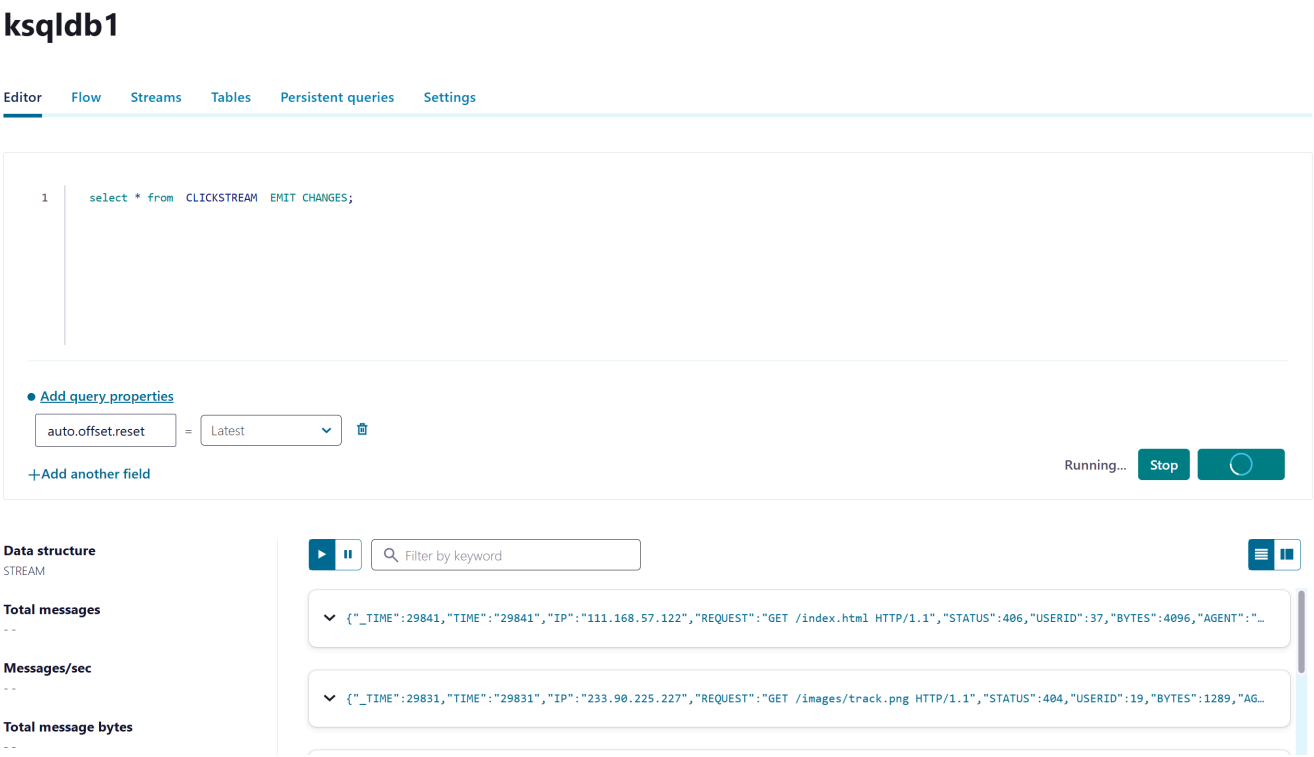
Confluent ui connect:



Load the Streaming Data to ksqlDB: Done



New data arriving into stream:



Elasticsearch + Grafana set up:

```
yura1adin@pc:~/m10_kafkabasics_sql_local-master$ docker-compose exec elasticsearch bash -c /scripts/elastic-dynamic-template.sh

-> Removing kafkaconnect template if it exists already.


-> Loading Elastic Dynamic Template to ensure _TS fields are used for TimeStamp
```

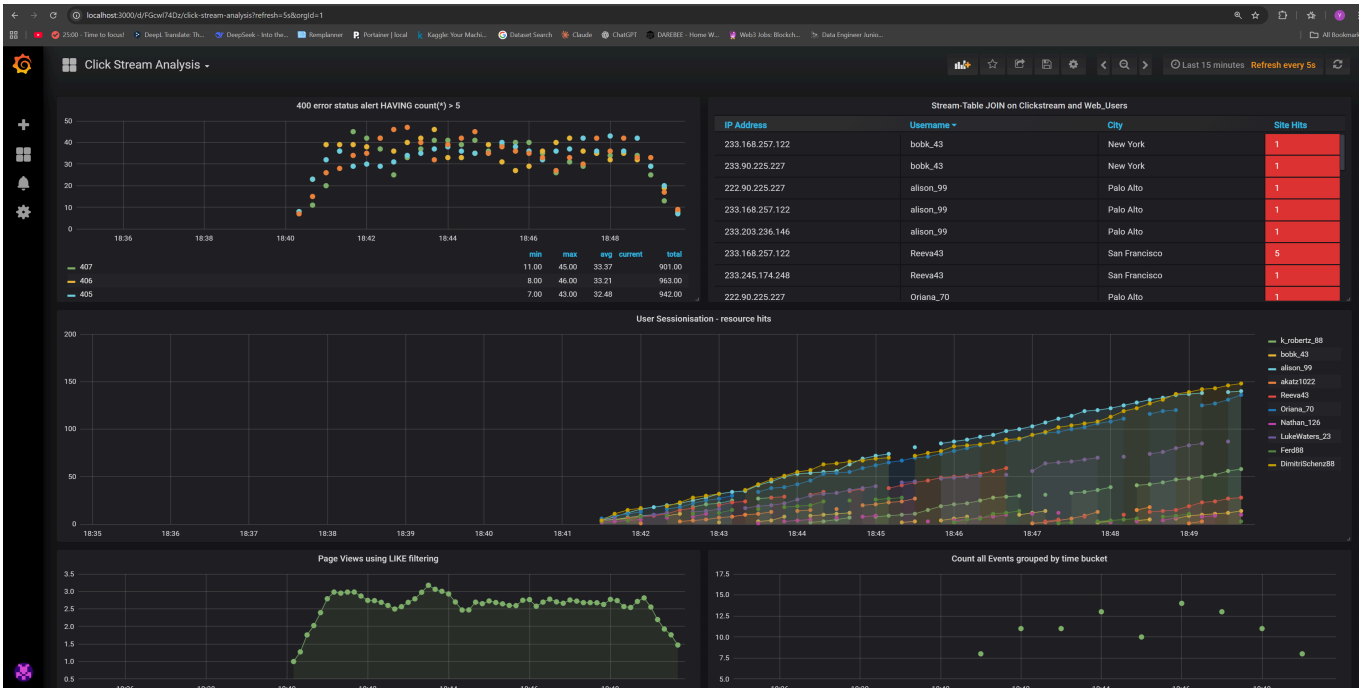
```
docker-compose exec ksqldb-server bash -c '/scripts/ksql-tables-to-grafana.sh'
```

```
        -> Connecting ksqlDB->Elastic->Grafana errors_per_min
        -> Connecting: errors_per_min
                -> Adding Kafka Connect Elastic Source es_sink_ERRORS_PER_MIN
                -> Adding Grafana Source


================================================================
Charting  EVENTS_PER_MIN
        -> Remove any existing Elastic search config
        -> Remove any existing Connect config
        -> Remove any existing Grafana config
        -> Connecting ksqlDB->Elastic->Grafana events_per_min
        -> Connecting: events_per_min
                -> Adding Kafka Connect Elastic Source es_sink_EVENTS_PER_MIN
                -> Adding Grafana Source


================================================================
Charting  PAGES_PER_MIN
        -> Remove any existing Elastic search config
        -> Remove any existing Connect config
        -> Remove any existing Grafana config
        -> Connecting ksqlDB->Elastic->Grafana pages_per_min
        -> Connecting: pages_per_min
                -> Adding Kafka Connect Elastic Source es_sink_PAGES_PER_MIN
                -> Adding Grafana Source


Done!
yura1adin@pc:~/m10_kafkabasics_sql_local-master$
```

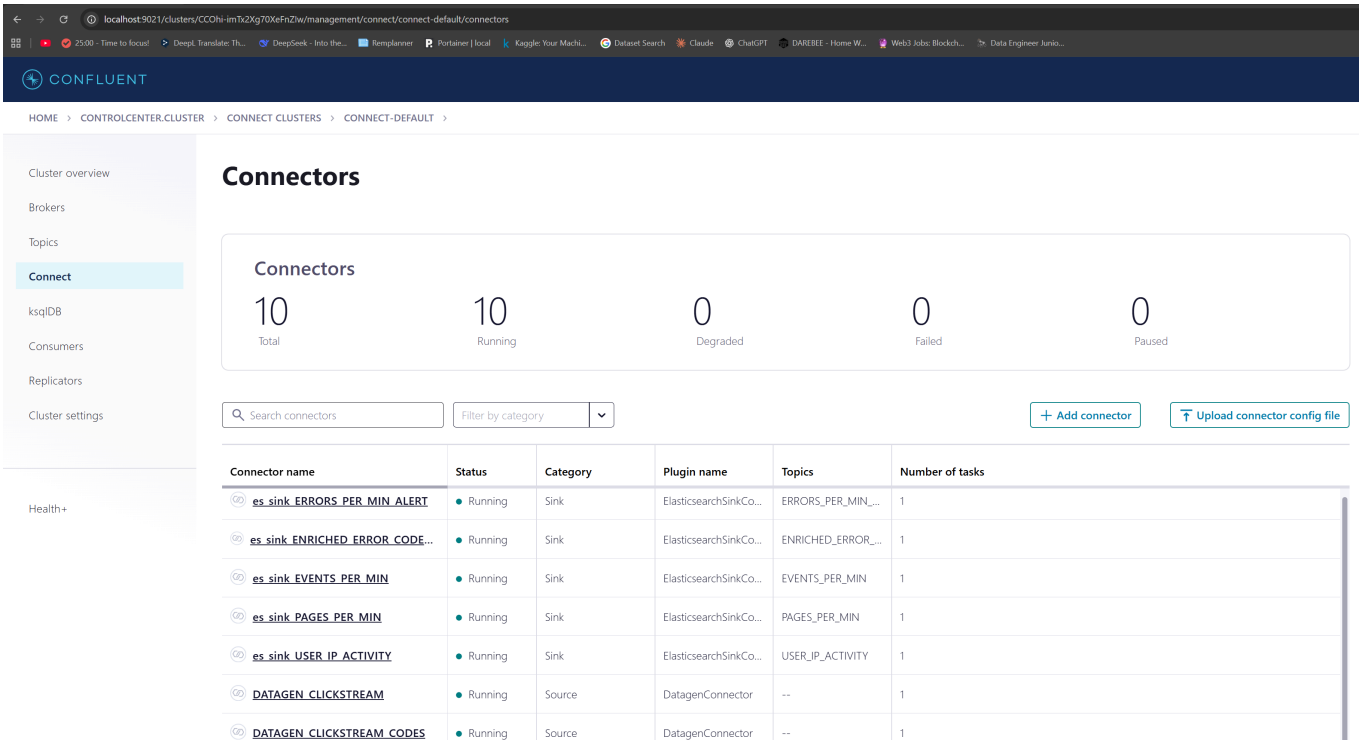
Grafana run:

```
yura1adin@pc:~/m10_kafkabasics_sql_local-master$ docker-compose exec grafana bash /scripts/clickstream-analysis-dashboar
d.sh
Loading Grafana ClickStream Dashboard
```

Grafana Dashboard:



Connectors:



Sessionize the data

```
yura1adin@pc:~/m10_kafkabasics_sql_local-master$ bash sessionize-data.sh
Running for 90 seconds to allow clickstream data to generate
Pausing the clickstream datagen for 35 seconds
Resuming the clickstream datagen for 90 seconds. Then the cycle will repeat
Pausing the clickstream datagen for 35 seconds
Resuming the clickstream datagen for 90 seconds. Then the cycle will repeat
Pausing the clickstream datagen for 35 seconds
Resuming the clickstream datagen for 90 seconds. Then the cycle will repeat
```

Grafana:


Click Stream Analysis
Last 5 minutes
Refresh every 5s
400 error status alert HAVING count(*) > 5
min max avg current total
407 9 40 26 310
406 9 38 26 316
405 15 44 27 327
Stream-Table JOIN on Clickstream and Web_Users
IP Address
Username
City
Site Hits
233.145.28.144 alison_99 Frankfurt 1
233.90.225.227 alison_99 Palo Alto 2
233.145.28.144 Reeva43 San Francisco 1
233.152.245.45 Reeva43 San Francisco 1
233.245.174.248 Nathan_126 San Francisco 1
233.152.245.45 Ferd88 New York 1
222.173.165.103 DimitriSchenz88 Frankfurt 1
233.245.174.248 DimitriSchenz88 Frankfurt 1
User Sessionisation - resource hits
k_robertz_88
bobk_43
alison_99
akatz1022
Reeva43
Oriana_70
Nathan_126
LukeWaters_23
Ferd88
DimitriSchenz88
Page Views using LIKE filtering
avg
Average PAGES 2.109
Count all Events grouped by time bucket
avg current total
Max EVENTS 8.00 40.00
TABLE-STREAM JOIN between ClickStream and Codes Table
Status Code
Definition
Number of events
407 Proxy authentication required 21
406 Not acceptable 27
Press the "Shift + Page Up" on your keyboard to take a screenshot