

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:

```
/opt/kafka/bin $ ./kafka-topics.sh --create --topic test --bootstrap-server localhost:9092 --partitions 1 --replication-factor 1  
Created topic test.  
/opt/kafka/bin $
```

3. Adding a few events:

```
MINGW64:/c/Users/yura1/Desktop/kafka_task_1  
/opt/kafka/bin $ ./kafka-console-producer.sh --topic test --bootstrap-server localhost:9092  
>event 1  
>event 2  
>event 3 a  
>^C/opt/kafka/bin $
```

4. Read events:

```
/opt/kafka/bin $ ./kafka-console-consumer.sh --topic test --from-beginning --bootstrap-server localhost:9092  
event 1  
event 2  
event 3 a  
|
```

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 Created
  ✓ Container zookeeper Created
  ✓ Container grafana Created
  ✓ Container tools Created
  ✓ Container kafka Created
  ✓ Container schema-registry Created
  ✓ Container ksqldb-server Created
  ✓ Container connect Created
  ✓ Container control-center Created
  ✓ Container ksqldb-cli Started
yuraladin@pc:~/m10_kafkabasics_sql_local-master$ docker-compose ps -a
NAME           IMAGE          COMMAND                  SERVICE      CREATED        STATUS        PORTS
connect        confluentinc/cp-server-connect-datagen:0.6.4-7.6.0   "/etc/confluent/docker_"    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/docker_"    control-center 32 seconds ago Up 30 seconds  0.0.0.0:9021->9021/tcp, [::]:9021->9021/tcp
elasticsearch   docker.elastic.co/elasticsearch/elasticsearch:6.3.0    "/usr/local/bin/docker_"  elasticsearch 33 seconds ago Up 38 seconds  0.0.0.0:9200->9200/tcp, [::]:9200->9200/tcp, 9306/tcp
grafana        grafana/grafana:5.2.4      "/run.sh"                grafana      33 seconds ago Up 38 seconds  0.0.0.0:3000->3000/tcp, [::]:3000->3000/tcp
kafka          confluentinc/cp-kafka:7.9.0       "/bin/sh"                kafka       32 seconds ago Up 39 seconds  9092/tcp
ksqldb-cli     confluentinc/cp-ksqldb-cli:7.9.0    "/etc/confluent/docker_"  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/docker_"  schema-registry 32 seconds ago Up 29 seconds  8081/tcp
tools          confluentinc/cp-training-tools:5.4.4    "/bin/bash"              tools       33 seconds ago Up 31 seconds
zookeeper      confluentinc/cp-zookeeper:7.9.0     "/etc/confluent/docker_"  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.confluent.connect.datagen.DatagenSourceConnector') AS
Message
-----
Created connector DATAGEN_CLICKSTREAM_CODES
-----

CREATE SOURCE CONNECTOR datagen_clickstream_users WITH ('connector.class'='io.confluent.connect.datagen.DatagenSourceConnector') AS
Message
-----
Created connector DATAGEN_CLICKSTREAM_USERS
-----

CREATE SOURCE CONNECTOR datagen_clickstream WITH ('connector.class'='io.confluent.connect.datagen.DatagenSourceConnector') AS
Message
-----
Created connector DATAGEN_CLICKSTREAM
-----

```

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,"remote_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,"remote_user":"-","time":"11","_time":11,"request":"GET /site/user_status.html HTTP/1.1","status":"407","bytes":2048,"referrer":"-","agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/59.0.3071.115 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,"remote_user":"-","time":"21","_time":21,"request":"GET /site/login.html HTTP/1.1","status":"200","bytes":4006,"referrer":"-","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"}, partition: 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":1457650522282,"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,"first_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:

localhost:9021/clusters/53M2onlRBdqXbfu0xA/management/connect/connect-default/connectors

CONFLUENT

HOME > CONTROLCENTER.CLUSTER > CONNECT CLUSTERS > CONNECT-DEFAULT >

Cluster overview
Brokers
Topics
Connect
ksqldb
Consumers
Replicators
Cluster settings

Connectors

Connectors

Total	Running	Degraded	Failed	Pending
3	3	0	0	0

Search connectors Filter by category Add connector

Connector name	Status	Category	Plugin name	Topics	Number of tasks
DATAGEN_CLICKSTREAM	Running	Source	DatagenConnector	--	1
DATAGEN_CLICKSTREAM_CODES	Running	Source	DatagenConnector	--	1
DATAGEN_CLICKSTREAM_USERS	Running	Source	DatagenConnector	--	1

Load the Streaming Data to ksqldb: Done

localhost:9021/clusters/53M2onlRBdqXbfu0xA/ksqldb1/data-flow

CONFLUENT

HOME > CONTROLCENTER.CLUSTER > KSQLDB

Cluster overview
Brokers
Topics
Connect
ksqldb
Consumers
Replicators
Cluster settings

ksqldb1

Editor Flow Streams Tables Persistent queries Settings

Search

```

graph LR
    KSQL_PROCESSING_LOG["KSQL PROCESSING LOG STREAM"] --> CREATE_TABLE1["CREATE-TABLE"]
    CLICKSTREAM_CODES["CLICKSTREAM_CODES TABLE"] --> CREATE_TABLE1
    CLICKSTREAM_CODES --> CREATE_STREAM1["CREATE-STREAM"]
    CLICKSTREAM_CODES --> CREATE_TABLE2["CREATE-TABLE"]
    CLICKSTREAM_CODES --> CREATE_TABLE3["CREATE-TABLE"]
    CLICKSTREAM_CODES --> CREATE_TABLE4["CREATE-TABLE"]
    CLICKSTREAM_CODES --> CREATE_TABLE5["CREATE-TABLE"]
    CLICKSTREAM_CODES --> CREATE_TABLE6["CREATE-TABLE"]
    CLICKSTREAM_CODES --> CREATE_TABLE7["CREATE-TABLE"]
    CLICKSTREAM_STREAM["CLICKSTREAM STREAM"] --> CREATE_TABLE8["CREATE-TABLE"]
    WEB_USERS_TABLE["WEB USERS TABLE"] --> CREATE_TABLE8
    CLICKSTREAM_STREAM --> CREATE_STREAM2["CREATE-STREAM"]
    WEB_USERS_TABLE --> CREATE_STREAM2
    CREATE_TABLE1 --> EVENTS_PER_MIN_TABLE["EVENTS_PER_MIN TABLE"]
    CREATE_TABLE1 --> ENRICHED_ERROR_CODES_STREAM["ENRICHED_ERROR_CODES STREAM"]
    CREATE_TABLE1 --> ERRORS_PER_MIN_ALERT_TABLE["ERRORS_PER_MIN_ALERT TABLE"]
    CREATE_TABLE1 --> ERRORS_PER_MIN_TABLE["ERRORS_PER_MIN TABLE"]
    CREATE_TABLE1 --> PAGES_PER_MIN_TABLE["PAGES_PER_MIN TABLE"]
    CREATE_TABLE1 --> USER_CLICKSTREAM_STREAM["USER_CLICKSTREAM STREAM"]
    EVENTS_PER_MIN_TABLE --> CHED_ERROR_CODES_COUNT_TABLE["CHED_ERROR_CODES_COUNT TABLE"]
    ENRICHED_ERROR_CODES_STREAM --> CHED_ERROR_CODES_COUNT_TABLE
    ERRORS_PER_MIN_ALERT_TABLE --> CHED_ERROR_CODES_COUNT_TABLE
    ERRORS_PER_MIN_TABLE --> CHED_ERROR_CODES_COUNT_TABLE
    PAGES_PER_MIN_TABLE --> USER_IP_ACTIVITY_TABLE["USER_IP_ACTIVITY TABLE"]
    USER_CLICKSTREAM_STREAM --> CLICK_USER_SESSIONS_TABLE["CLICK_USER_SESSIONS TABLE"]
    USER_CLICKSTREAM_STREAM --> USER_IP_ACTIVITY_TABLE
    
```

New data arriving into stream:

ksqldb1

Editor Flow Streams Tables Persistent queries Settings

```
1 | select * from CLICKSTREAM EMIT CHANGES;
```

● Add query properties
auto.offset.reset = Latest

+Add another field

Running...

Data structure
STREAM

Total messages

Messages/sec

Total message bytes

Filter by keyword

```
{"_TIME":29841,"TIME":"29841","IP":"111.168.57.122","REQUEST":"GET /index.html HTTP/1.1","STATUS":406,"USERID":37,"BYTES":4096,"AGENT":"..."}
```

```
{"_TIME":29831,"TIME":"29831","IP":"233.90.225.227","REQUEST":"GET /images/track.png HTTP/1.1","STATUS":404,"USERID":19,"BYTES":1289,"AG..."}
```

Elasticsearch + Grafana set up:

```
yurraladin@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!

yuraladin@pc:~/m10_kafkabasics_sql_local-master\$

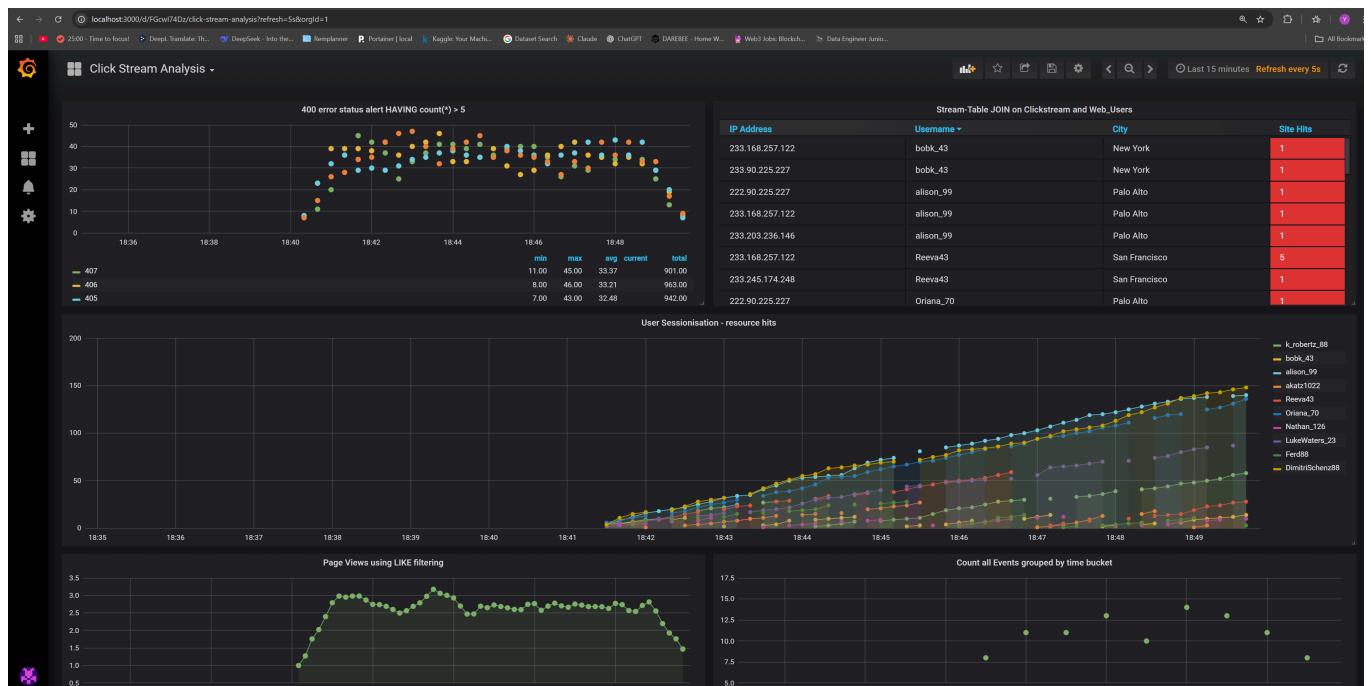
Grafana run:

```

yuraladin@pc:~/m10_kafkabasics_sql_local-master$ docker-compose exec grafana bash /scripts/clickstream-analysis-dashboard.sh
Loading Grafana ClickStream Dashboard

```

Grafana Dashboard:



Connectors:

The screenshot shows the Confluent Control Center interface for managing connectors. On the left, a sidebar menu includes 'Cluster overview', 'Brokers', 'Topics', 'Connect' (which is selected), 'ksqlDB', 'Consumers', 'Replicators', and 'Cluster settings'. The main content area is titled 'Connectors' and displays the following statistics:

Category	Value
Total	10
Running	10
Degraded	0
Failed	0
Paused	0

Below the statistics is a search bar labeled 'Search connectors' and a dropdown menu labeled 'Filter by category'. There are two buttons at the top right: '+ Add connector' and 'Upload connector config file'.

The main table lists the active connectors:

Connector name	Status	Category	Plugin name	Topics	Number of tasks
es_sink_ERRORS_PER_MIN_ALERT	● Running	Sink	ElasticsearchSinkCo...	ERRORS_PER_MIN_...	1
es_sink_ENRICHED_ERROR_CODE...	● Running	Sink	ElasticsearchSinkCo...	ENRICHED_ERROR_...	1
es_sink_EVENTS_PER_MIN	● Running	Sink	ElasticsearchSinkCo...	EVENTS_PER_MIN	1
es_sink_PAGES_PER_MIN	● Running	Sink	ElasticsearchSinkCo...	PAGES_PER_MIN	1
es_sink_USER_IP_ACTIVITY	● Running	Sink	ElasticsearchSinkCo...	USER_IP_ACTIVITY	1
DATAGEN_CLICKSTREAM	● Running	Source	DatagenConnector	--	1
DATAGEN_CLICKSTREAM_CODES	● Running	Source	DatagenConnector	--	1

Sessionize the data

```
yuraladin@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:

