



Trabajo Grupal

Tarea Unida V: Data Streaming

Identificación

- **Módulo:** Procesamiento de Flujo de Datos Masivos
- **Curso:** Big Data y Business Analytics
- **Integrantes:** Silvia Gabaglio y Stephanie González

1.docker-compose.yml - producer

```
---
version: '3.9'
services:

  redpanda:
    command:
      - redpanda
      - start
      - --smp
      - '1'
      - --reserve-memory
      - 0M
      - --overprovisioned
      - --set
      - redpanda.cluster_id=turning-red
      - --set
      - redpanda.enable_idempotence=true
      - --set
      - redpanda.enable_transactions=true
      - --set
      - redpanda.auto_create_topics_enabled=true
      - --node-id
      - '0'
      - --kafka-addr
      - PLAINTEXT://0.0.0.0:29092,OUTSIDE://0.0.0.0:9092
```



Departamento de Elearning
Facultad Politécnica
Universidad Nacional de Asunción



```
- --advertise-kafka-addr
- PLAINTEXT://redpanda:29092,OUTSIDE://localhost:9092
image: docker.vectorized.io/vectorized/redpanda:v21.11.11
container_name: redpanda
ports:
  - 9092:9092
  - 29092:29092
networks:
  - redpanda_network

ksqldb-server:
  image: confluentinc/cp-ksqldb-server:latest
  hostname: ksqldb-server
  container_name: ksqldb-server
  depends_on:
    - redpanda
  ports:
    - "8088:8088"
  environment:
    KSQL_LISTENERS: "http://0.0.0.0:8088"
    KSQL_BOOTSTRAP_SERVERS: "redpanda:29092"
    KSQL_KSQL_SCHEMA_REGISTRY_URL: "http://schema-registry:8081"
    KSQL_KSQL_LOGGING_PROCESSING_STREAM_AUTO_CREATE: "true"
    KSQL_KSQL_LOGGING_PROCESSING_TOPIC_AUTO_CREATE: "true"
  networks:
    - redpanda_network

ksqldb-cli:
  image: confluentinc/cp-ksqldb-cli:latest
  container_name: ksqldb-cli
  depends_on:
    - redpanda
    - ksqldb-server
  entrypoint: /bin/sh
  tty: true
  networks:
    - redpanda_network

console:
  container_name: redpanda-console
  image: docker.redpanda.com/vectorized/console:v2.2.4
  entrypoint: /bin/sh
  command: -c 'echo "$$CONSOLE_CONFIG_FILE" > /tmp/config.yml;
/app/console'
  environment:
    CONFIG_FILEPATH: /tmp/config.yml
```



Departamento de Elearning
Facultad Politécnica
Universidad Nacional de Asunción



```
CONSOLE_CONFIG_FILE: |
  kafka:
    brokers: ["redpanda:29092"]
    schemaRegistry:
      enabled: true
      urls: ["http://redpanda:8081"]
    redpanda:
      adminApi:
        enabled: true
        urls: ["http://redpanda:9644"]
  networks:
    - redpanda_network
  ports:
    - 8080:8080
  depends_on:
    - redpanda

networks:
  redpanda_network:
    driver: bridge
```

2.Código para levantar redpanda

Se levanta el redpanda en un puerto.

```
PS C:\workspaces\python_proyectos> & c:/workspaces/python_proyectos/env/Scripts/Activate.ps1
(env) PS C:\workspaces\python_proyectos> cd ..
(env) PS C:\workspaces> cd .\panda_proyecto\
(env) PS C:\workspaces\panda_proyecto> ls

Directorio: C:\workspaces\panda_proyecto

Mode                LastWriteTime         Length Name
----                -
-a-----         2/8/2023  1:17 p. m.         2534 docker-compose.yml

(env) PS C:\workspaces\panda_proyecto> docker compose up -d
error during connect: this error may indicate that the docker daemon is not running: Get "http://%2F%2F.%2Fpipe%2Fdocker_engine/
v1.24/containers/json?all=1&filters=%7B%22label%22%3A%7B%22com.docker.compose.config-hash%22%3Atrue%2C%22com.docker.compose.proj
ect%3Dredpanda-quickstart%22%3Atrue%7D%7D": open //./pipe/docker_engine: The system cannot find the file specified.
(env) PS C:\workspaces\panda_proyecto> docker compose up -d
[+] Running 2/2
 ✓ Container redpanda-0          Started          0.7s
 ✓ Container redpanda-console    Started          1.4s
(env) PS C:\workspaces\panda_proyecto> 
```

3. Instalación del KSQLdb-cli y KSQL-server

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL

For more help on how to use Docker, head to https://docs.docker.com/go/guides/

● PS C:\workspaces\panda_proyecto> docker compose up -d
[+] Running 47/4
✓ console 7 layers [██████████] 0B/0B Pulled 194.7s
✓ ksqldb-server 12 layers [██████████] 0B/0B Pulled 215.4s
✓ redpanda 7 layers [██████████] 0B/0B Pulled 181.5s
✓ ksqldb-cli 17 layers [██████████] 0B/0B Pulled 134.6s
[+] Running 5/5
✓ Network panda_proyecto_redpanda_network Created 0.3s
✓ Container redpanda Started 4.9s
✓ Container ksqldb-server Started 5.8s
✓ Container redpanda-console Started 5.9s
✓ Container ksqldb-cli Started 7.3s
○ PS C:\workspaces\panda_proyecto> 
```

4.consumer

```
1  import json
2  from kafka import KafkaConsumer
3
4  consumer = KafkaConsumer(
5      bootstrap_servers=["localhost:9092"],
6      group_id="demo-group",
7      auto_offset_reset="earliest",
8      enable_auto_commit=False,
9      consumer_timeout_ms=1000,
10     value_deserializer=lambda m: json.loads(m.decode('ascii'))
11 )
12
13 consumer.subscribe("CONTAR_SYBOLOS")
14
15 try:
16     for message in consumer:
17         topic_info = f"topic: {message.partition}|{message.offset}"
18         message_info = f"key: {message.key}, {message.value}"
19         print(f"{topic_info}, {message_info}")
20         print('hola')
21 except Exception as e:
22     print(f"Error occurred while consuming messages: {e}")
23 finally:
24     consumer.close()
```



Departamento de Elearning
Facultad Politécnica
Universidad Nacional de Asunción



5. Las 4 consultas utilizadas en sql

Promedio de precio ponderado

```
CREATE TABLE promedio AS select s,avg(p) from finnhub_trades_stream group by s  
EMIT CHANGES;
```

Consulta promedio de precio ponderado

```
Select * from promedio
```

```
+-----+-----+
|S|KSQL_COL_0|
+-----+-----+
|AAPL|174.10195555555552|
|AMZN|132.7125|
|BINANCE:BTCUSD|25686.68375|
Query terminated
ksql> select * from promedio;
+-----+-----+
|S|KSQL_COL_0|
+-----+-----+
|AAPL|174.12587450980382|
|AMZN|132.742263636362|
|BINANCE:BTCUSD|25733.20504761903|
Query terminated
ksql> CREATE TABLE promedio AS select s,avg(p) from finnhub_trades_stream group by s EMIT CHANGES;[]
```

Cantidad de transacciones por símbolo

```
CREATE TABLE cantidad_transacciones AS SELECT s,count(s) AS  
cantidad_transaccion_simbolo FROM finnhub_trades_stream GROUP BY s EMIT  
CHANGES;
```

```
select * from cantidad_transacciones;
```

```
>CREATE TABLE cantidad_transacciones AS SELECT s,count(s) AS cantidad_transaccion_simbolo FROM finnhub_trades_stream GROUP BY s EMIT CHANGES;

Message
-----
Created query with ID CTAS_CANTIDAD_TRANSACCIONES_18
-----
ksql> select * from cantidad_transacciones;
+-----+-----+
|S|CANTIDAD_TRANSACCION_SIMBOLO|
+-----+-----+
|AAPL|1|
|BINANCE:BTCUSD|3|
Query terminated
ksql> []
```



Departamento de Elearning
Facultad Politécnica
Universidad Nacional de Asunción



Precio Máximo

```
CREATE TABLE precio_maximo AS SELECT s,MAX(p) AS MAX_PRICE FROM  
finnhub_trades_stream GROUP BY s EMIT CHANGES;
```

```
select * from precio_maximo;
```

```
ksql> CREATE TABLE precio_maximo AS SELECT s,MAX(p) AS MAX_PRICE FROM finnhub_trades_stream GROUP BY s EMIT CHANGES;  
  
Message  
-----  
Created query with ID CTAS_PRECIO_MAXIMO_14  
-----  
  
ksql> select * from precio_maximo;  
+-----+-----+  
|S|MAX_PRICE|  
+-----+-----+  
|AAPL|174.44|  
|AMZN|133.075|  
Query terminated  
ksql> 
```

Precio Mínimo

```
CREATE TABLE precio_minimo AS SELECT s,MIN(p) AS MIN_PRICE FROM  
finnhub_trades_stream GROUP BY s EMIT CHANGES;
```

```
select * from precio_minimo;
```

```
ksql> select * from precio_minimo;  
+-----+-----+  
|S|MIN_PRICE|  
+-----+-----+  
|AAPL|174.34|  
|AMZN|132.9|  
|BINANCE:BTCUSDT|25982.8|  
Query terminated  
ksql> 
```