

Entregable Kafka

Post Work: Data Processing with Kafka end to end

1. El fichero con el que se realizó este entregable es Netflix.csv disponible en <https://www.kaggle.com/datasets> (adjunto) con las siguientes columnas:

- show_id
- title
- director cast
- country
- date_added
- release_year
- rating
- duration
- listed_in
- description
- type
- year_added
- month_added
- season_count

A partir de los datos que se van recopilando en Streaming podríamos realizar recomendaciones de productos en línea basado en las preferencias del usuario.

2. Creación input data

```
Kafka_local > src > main > python > exercise5 > producer.py > ...
1 import time
2 from json import dumps
3 from confluent_kafka import Producer
4 import re
5
6 # Configuración del productor
7 config = {
8     'bootstrap.servers': 'localhost:9092', # Cambia esto con la dirección de tu servidor Kafka
9     'client.id': 'python-producer'
10 }
11
12 # Crear un productor
13 producer = Producer(config)
14
15
```

PROBLEMAS 38 SALIDA CONSOLA DE DEPURACIÓN TERMINAL PUERTOS









"As her father nears the end of his life, filmmaker Kirsten Johnson stages his death in inventive and comical ways to help them both face the inevitable."

s2,TV Show,Blood & Water,,Ama Qamata, Khosi Ngema, Gail Mababane, Thabang Molaba, Dillon Windvogel, Natasha Thahane, Arno Greeff, Xolile Tshabalala, Getmore Sithole, Cindy Mahlangu, Ryle De Morny, Greteli Fincham, Sello Maake Ka-Ncube, Odwa Gwanya, Mekaila Mathys, Sandi Schultz, Duane Williams, Shamilla Miller, Patrick Mofokeng",South Africa,"September 24, 2021",2021,TV-MA,2 Seasons,"International TV Shows, TV Dramas, TV Mysteries","After crossing paths at a party, a Cape Town teen sets out to prove whether a private-school swimming star is her sister who was abducted at birth."

s3,TV Show,Ganglands,Julien Leclercq,"Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabiha Akkari, Sofia Lesaffre, Salim Kechiouche, Noureddine Farihi, Geert Van Rampelberg, Bakary Diombera",,"September 24, 2021",2021,TV-MA,1 Season,"Crime TV Shows, International TV Shows, TV Action & Adventure","To protect his family from a powerful drug lord, skilled thief Mehdi and his expert team of robbers are pulled into a violent and deadly turf war."

s4,TV Show,Jailbirds New Orleans,,,,,"September 24, 2021",2021,TV-MA,1 Season,"Docuseries, Reality TV","Feuds, flirtations and toilet talk go down among the incarcerated women at the Orleans Justice Center in New Orleans on this gritty reality series."

3. Enviar a Topics (movies):

<input type="checkbox"/>	Name	Image	Status	CPU (%)	Port(s)	Last start... ↓
<input type="checkbox"/>	 exercise5		Running (5/5)	8.62%		25 minutes ago
<input type="checkbox"/>	 ksql-cli-1 b6acba5dcfd3	 confluentinc/cp-ksql-cli:latest	Running	0%		25 minutes ago
<input type="checkbox"/>	 control-center f1b08231e994	confluentinc/cp-enterprise-control:	Running	5.08%	9021:9021	25 minutes ago
<input type="checkbox"/>	 ksql-server-1 104674c561f9	 confluentinc/cp-ksql-server:latest	Running	0.95%	8088:8088	25 minutes ago
<input type="checkbox"/>	 kafka-1 f4000541091d	confluentinc/cp-kafka:7.3.3	Running	2.45%	9092:9092	25 minutes ago
<input type="checkbox"/>	 zookeeper-1 23cc0c498e7e	confluentinc/cp-zookeeper:latest	Running	0.14%	2181:2181	25 minutes ago

← → ↻

localhost:9021/clusters/vfefUbITQoOOyHpbsCUH3Q/management/topics

☆

CONFLUENT

HOME > CONTROLCENTER.CLUSTER >

Cluster overview

Brokers

Topics

Connect

ksqlDB

Consumers

Topics

☒ Hide internal topics

Topic name	Partitions
<u>movies</u>	1

movies

Configuration Messages Schema

Message fields

- topic
- partition
- offset
- timestamp
- timestampType
- headers
- key
- value

▶ ||

Jump to offset

+ Produce a new message to this topic

▼ ,

Partition: 0 Offset: 15262 Timestamp: 1705245131359

▼ Movie

Partition: 0 Offset: 15261 Timestamp: 1705245131358

▼ ,

Partition: 0 Offset: 15260 Timestamp: 1705245131358

4. Procesamiento con KSQL:

```
Field | Type
-----|-----
ROWTIME | BIGINT (system)
ROWKEY | VARCHAR(STRING) (system)
PALABRA | VARCHAR(STRING)

Local runtime statistics
-----

(Statistics of the local KSQL server interaction with the Kafka topic movies)
ksql> CREATE TABLE mi_ktable AS
>SELECT palabra,
>count(*)
>FROM movies_stream
>GROUP BY palabra
>EMIT CHANGES;

Message
-----
Table MI_KTABLE created and running. Created by query with query ID: CTAS_MI_KTABLE_2

ksql>
```

MI_KTABLE

[Configuration](#)[Messages](#)[Schema](#)

Message fields

- topic
- partition
- offset
- timestamp
- timestampType
- headers
- key
- value

Filter by keyword

Jump to offset

offset

+ Produce a new message to this topic

International,688

Partition: 0 Offset: 47736 Timestamp: 1705248447889

Comedies,430

Partition: 0 Offset: 47735 Timestamp: 1705248447888

min,1034

Partition: 0 Offset: 47734 Timestamp: 1705248447887

102,29

Partition: 0 Offset: 47733 Timestamp: 1705248447887

5. Resultado en terminal

PROBLEMAS	SALIDA	CONSOLA DE DEPURACIÓN	TERMINAL	PUERTOS
			MA	490
			35	2
			min	994
			Comedies	412
			Dramas	529
			Cast	3
			members	11
			hit	17
			Netflix	1
			show	29
			join	6
			hosts	2
			to	1195
			unpack	2
			drama	15
			of	998
			Season	311
			l	594
			and	1106
			Cristela	1
			Alonzo	2
			offers	5
			tips	5
			for	312
			being	23
			the	1490
			new	103
			kid	11
			at	122
			school	57

PROBLEMAS	SALIDA	CONSOLA DE DEPURACIÓN	TERMINAL	PUERTOS
			season	6
			before	29
			moving	6
			Division	1
			Y	37
			college	22
			football	11
			1445	1
			playing	1
			Keeps	1
			Gabriele	2
			Muccino	3
			Gerard	4
			Butler	2
			Jessica	33
			Biel	5
			Zeta	1
			Jones	56
			Dennis	27
			Quaid	6
			Uma	4
			Thurman	3
			Judy	9
			Greer	5
			Noah	8
			Lomax	2
			James	133
			Tupper	1
			Abella	2
			Wyes	1
			Grant	16
			Goodman	3
			3	109
			2012	34
			PG	350
			13	194
			106	25
			Romantic	192
			Sports	34

