# Global Terrorism Analysis

Data streaming

Presentado por
**Martín García**
**David Melo**
**Juan Andrés Ruiz**

# Sobre nosotros

Global Terrorism Database

Utilizamos un conjunto de datos que incluye información sobre **atentados terroristas en todo el mundo** desde 1970 hasta 2017 (excepto 1993).

# Sobre nosotros

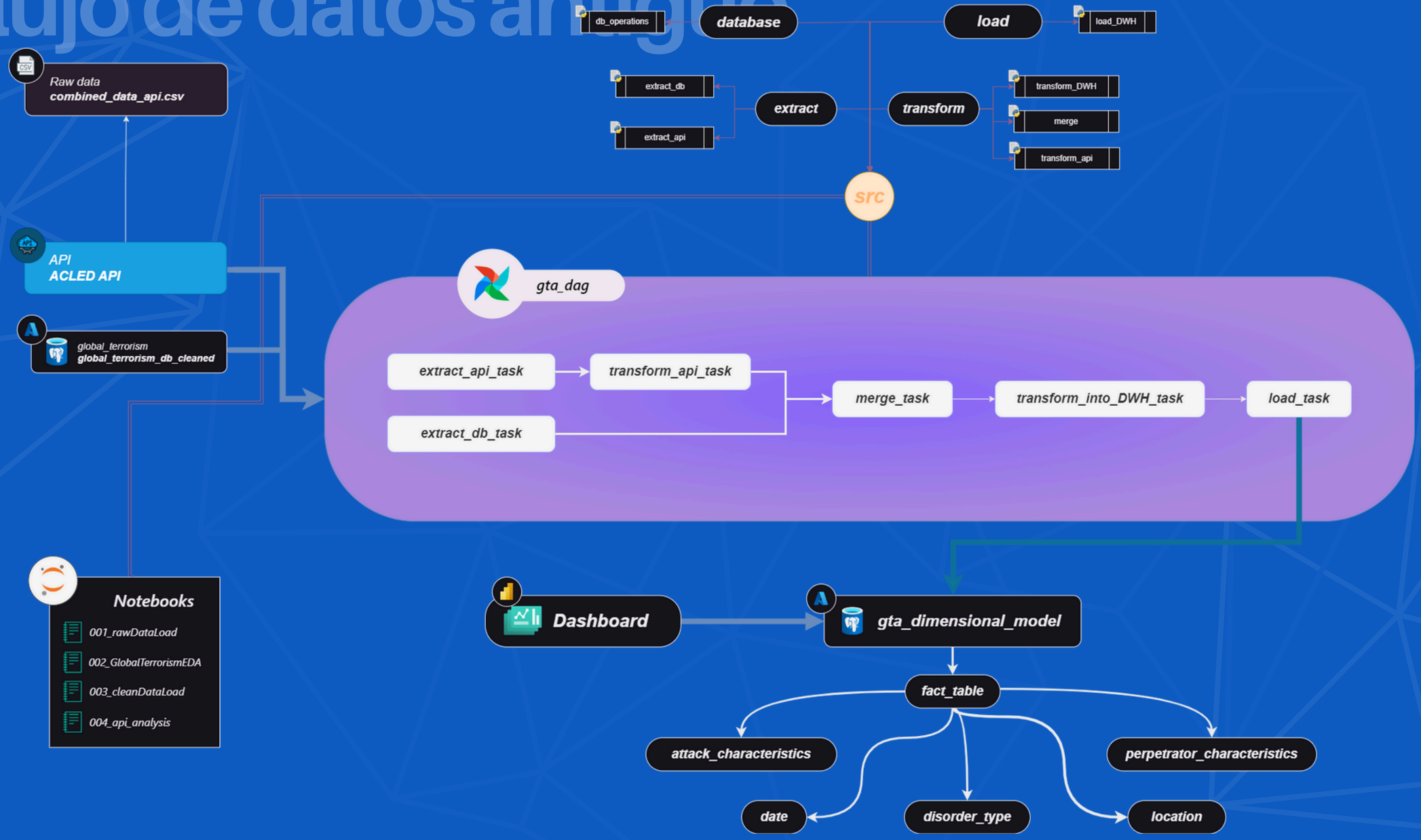
ACLED API

Es la fuente de datos que usamos para realizar un merge entre ambas fuentes de información.

ACLED API

```
def extracting_api_data():
    try:
        logging.info("Starting API data extraction.")

        url = 'https://api.acleddata.com/acled/read.json'
        params = {
            'key': f'{api_key}',
            'email': f'{api_email}',
            'fields': 'event_date|country|disorder_type|actor1',
            'year': '1989|2017',
            'year_where': 'BETWEEN',
            'limit': 5000,
            'page': 1
        }
```

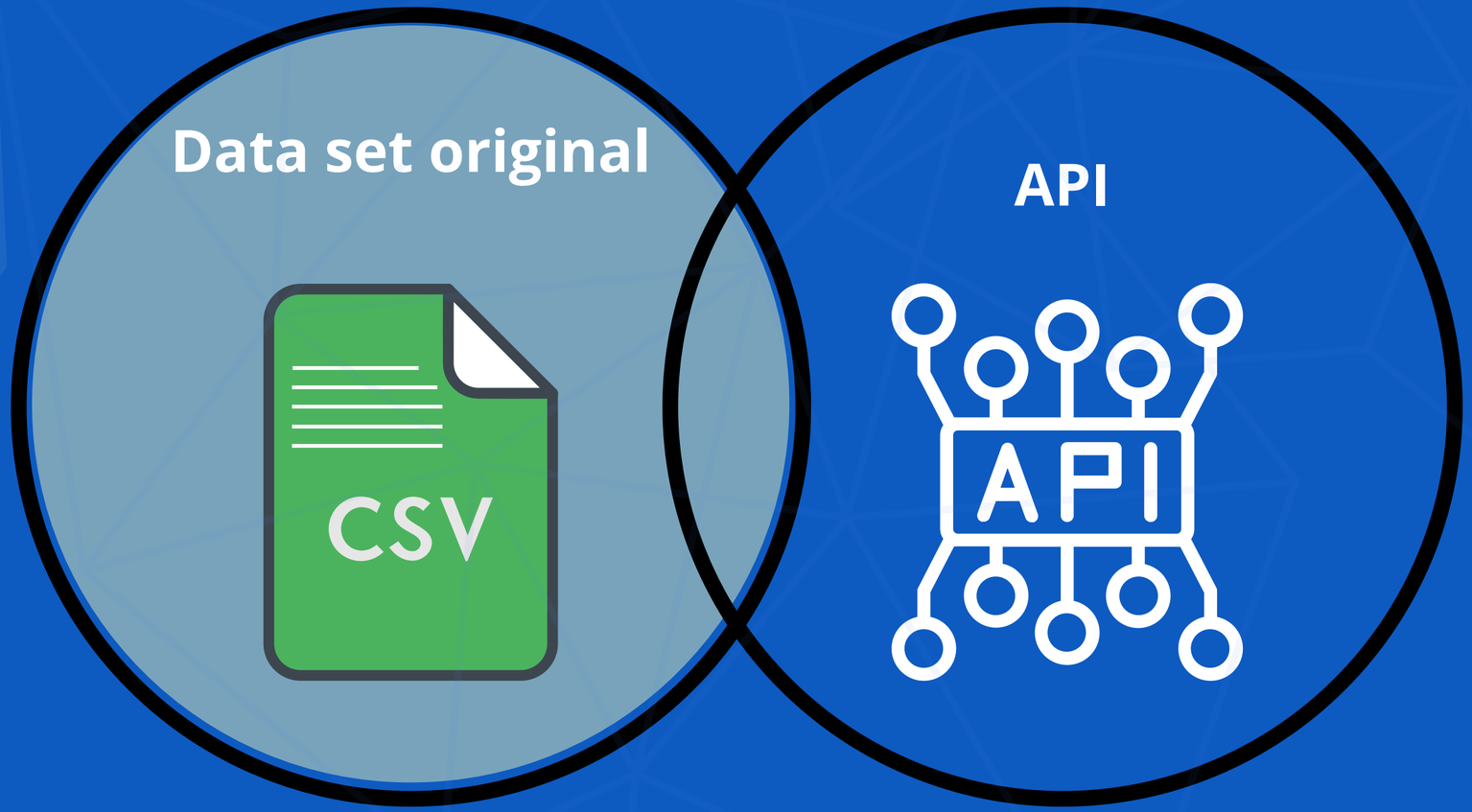
# Flujo de datos antiguo



# Merge

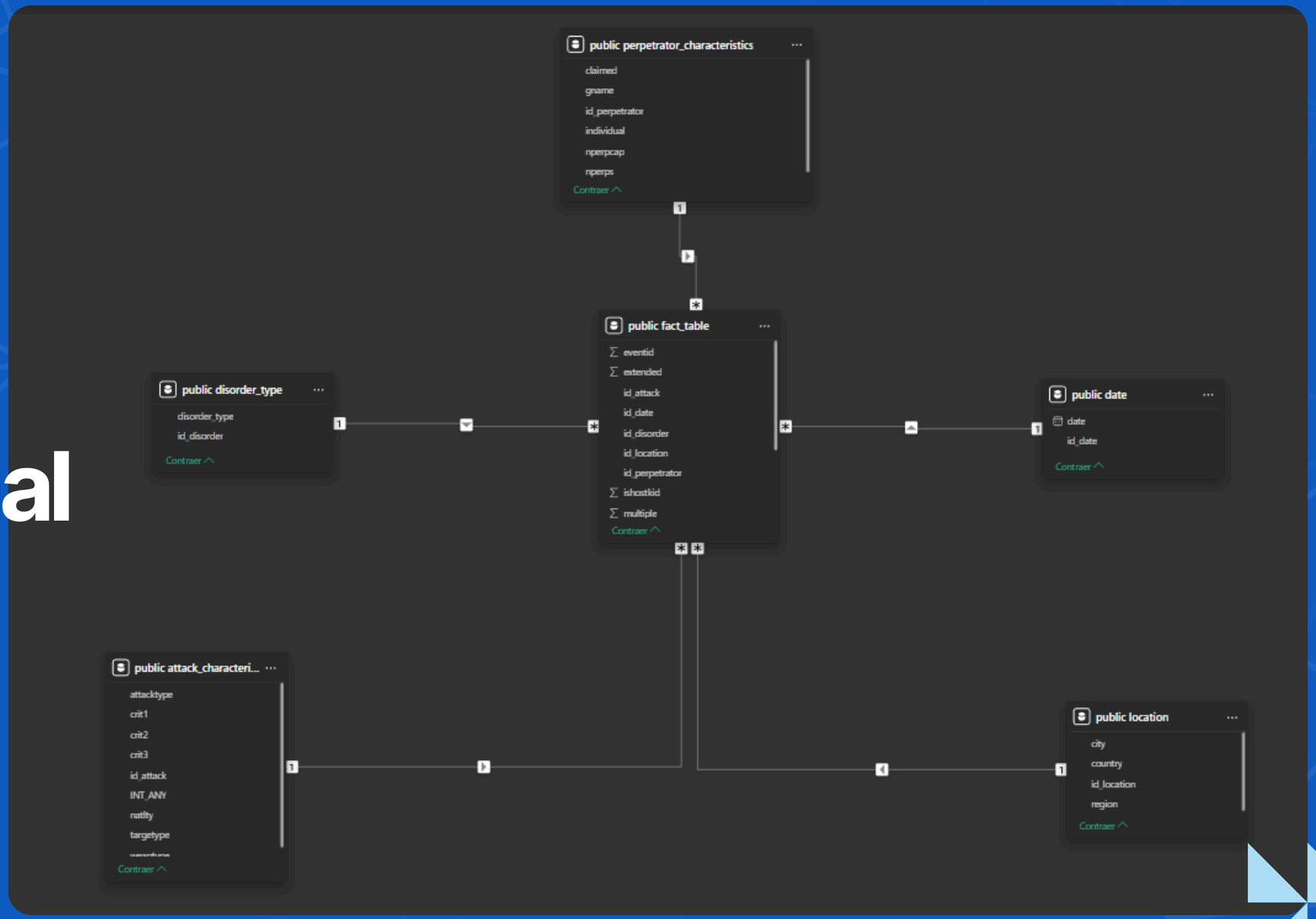
Devuelve todas las filas de la tabla izquierda, incluso si no hay coincidencias en la tabla derecha. Si no hay coincidencia, los valores de la tabla derecha serán NULL.

## *Left join*



# Modelo dimensional

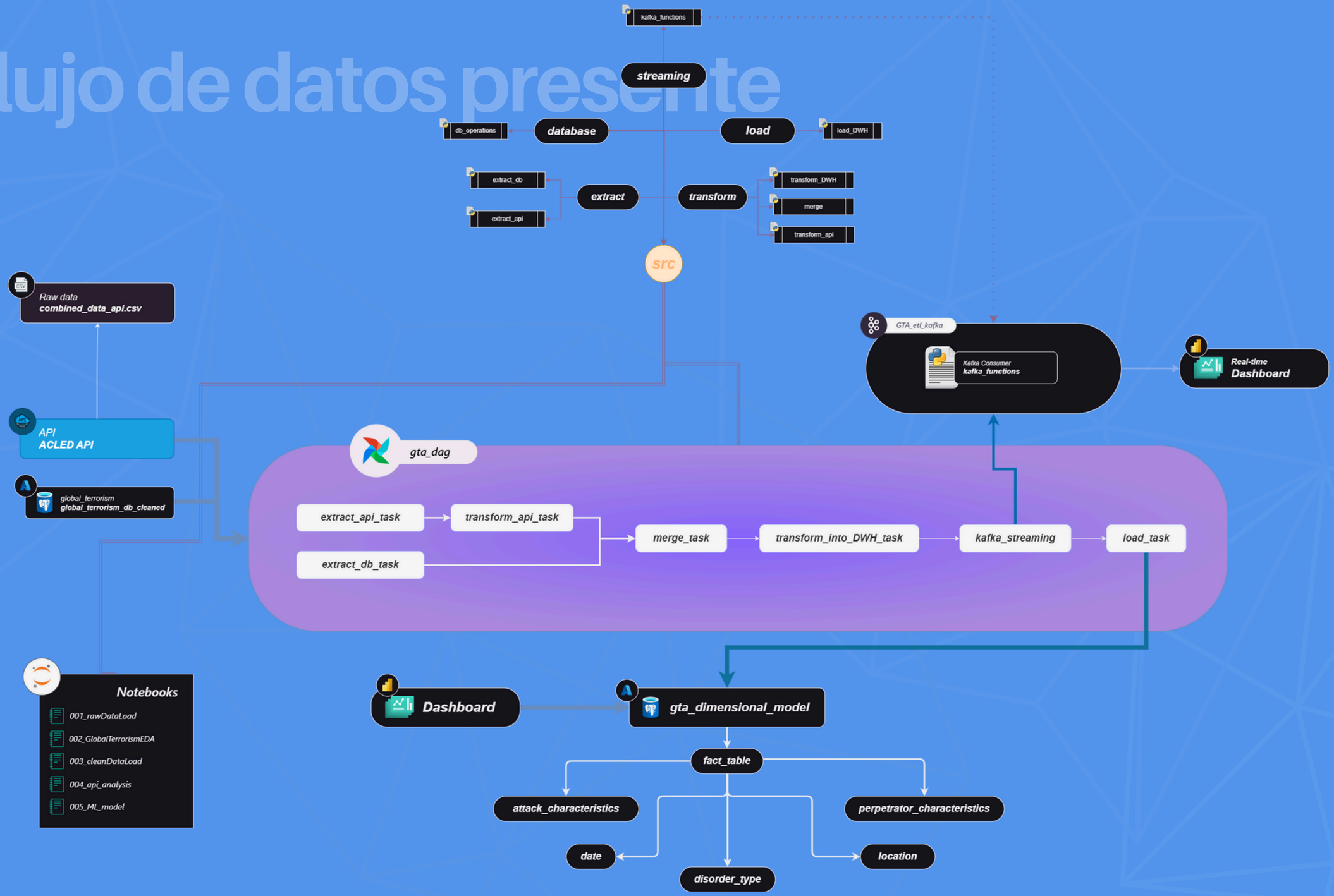
*arquitectura estrella*



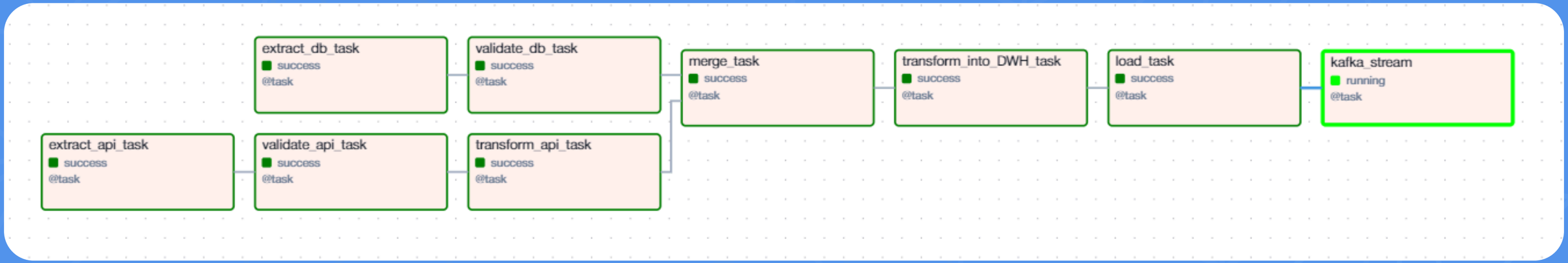
# Dashboard


Año
Todas
Regiones
Todas
Global Terrorism
GTD
GLOBAL TERRORISM DATABASE
Tipo de desorden terrorista
0,68 mil
(0,77%)
87,64 ...
(99,1...)
disor...
Un...
Pol...
Str...
Tasa de muertos y heridos por atentado terrorista
20... (...)
304... (6...)
Tasa de mu...
Suma de ...
Suma de ...
88 mil
Ataques terroristas perpetuados
77 mil
Ataques terroristas exitosos
os terroristas por grupos armados
0,11 ... (0,1...)
0,... (...)
2 ...
(2...)
5,53 mil
(6,31%)
44,97 mil
(51,33%)
gname
Unknown
Taliban
Islamic Sta...
Boko Hara...
Al-Shabaab
Communi...
Maoists
New Peop...
Tehrik-i-Ta...
Revolutio...
Número de atentados terroristas por paises
Recuento de atentados terroristas por mes
10 mil
5 mil
0 mil
Recuento de eventid
enero
febrero
marzo
abril
mayo
junio
julio
agosto
septiembre
octubre
noviembre
diciembre
Mes

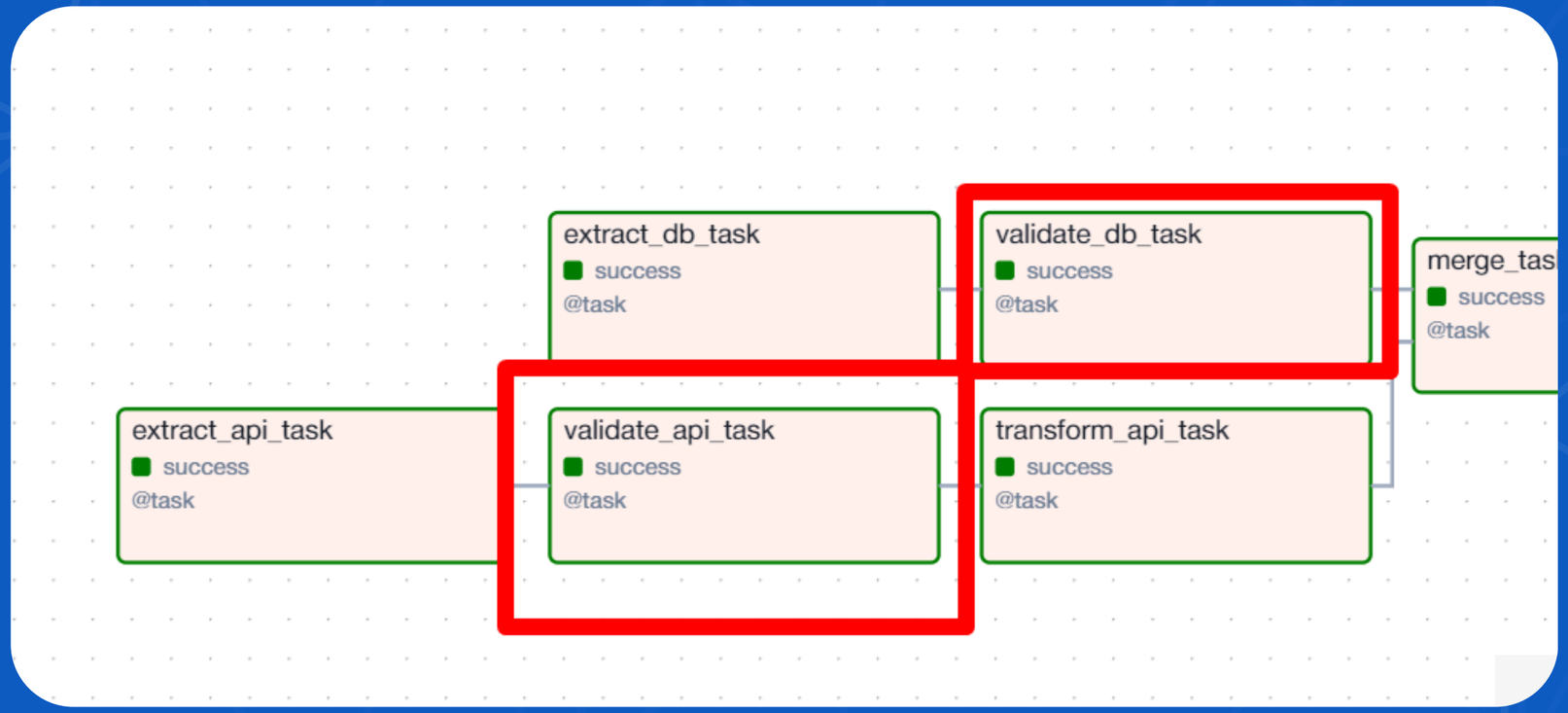
# Flujo de datos presente

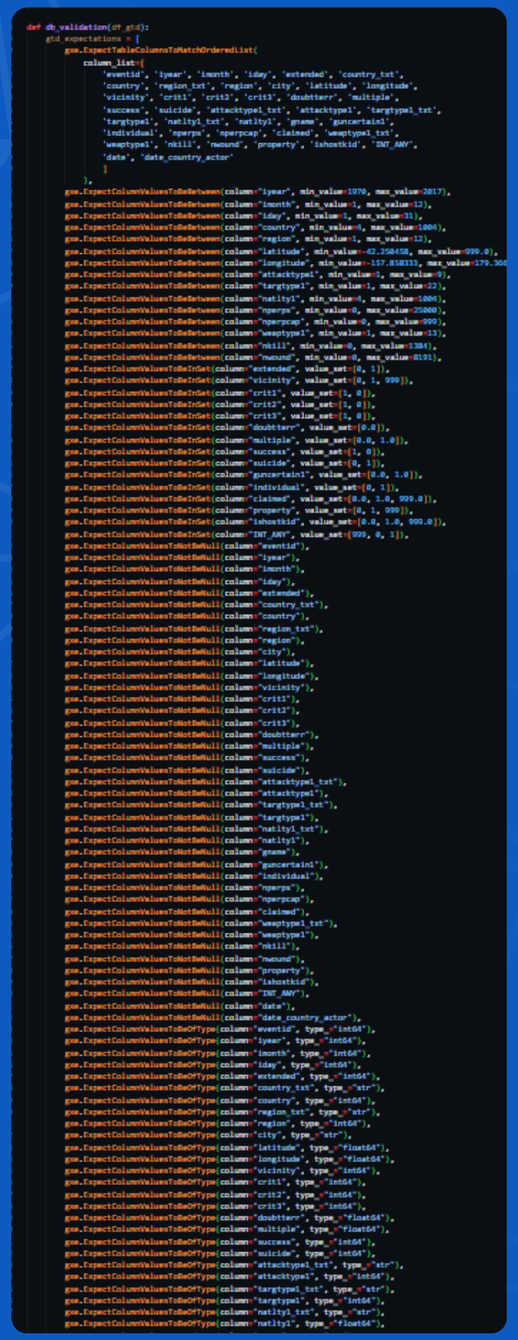


# Flujo de datos presente



# Great Expectations






# *Validaciones para*

- Orden de columnas
- Rangos de valores
- Set de valores
- Nulos
- Tipo de dato

# ML model

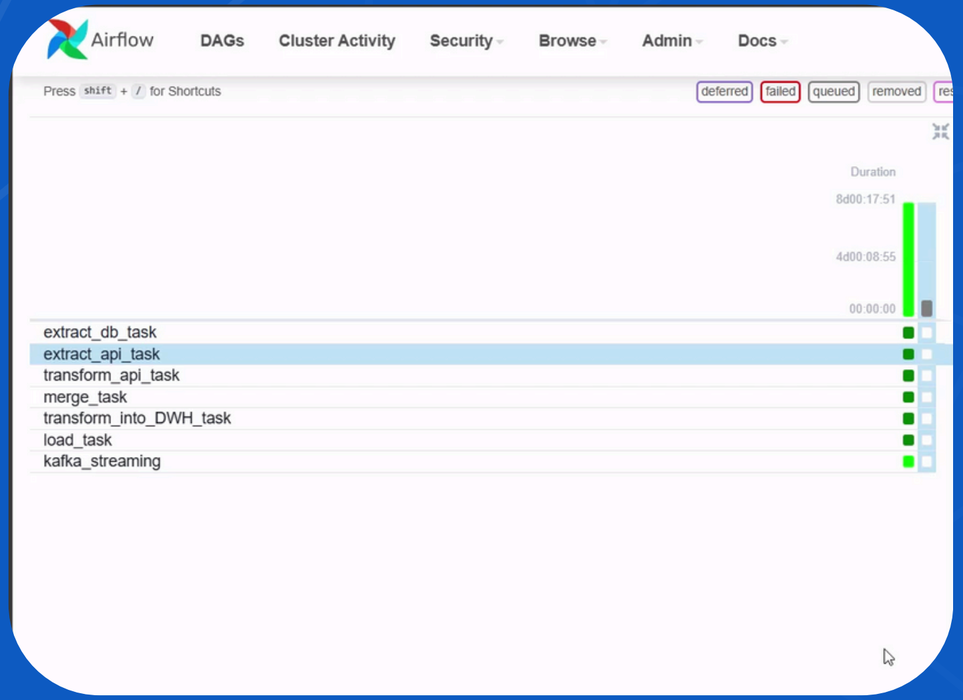
```
Model: Linear Regression
Accuracy: 0.1510
Model: Lasso
Accuracy: 0.1262
Model: Ridge
Accuracy: 0.1510
Model: RidgeCV
Accuracy: 0.1510
Model: LassoCV
Accuracy: 0.1257
Model: LassoLars
Accuracy: 0.1262
Model: LassoLarsCV
Accuracy: 0.1510
Model: LassoLarsIC
Accuracy: 0.1510
Model: XGBRegressor
Accuracy: 0.1571
```

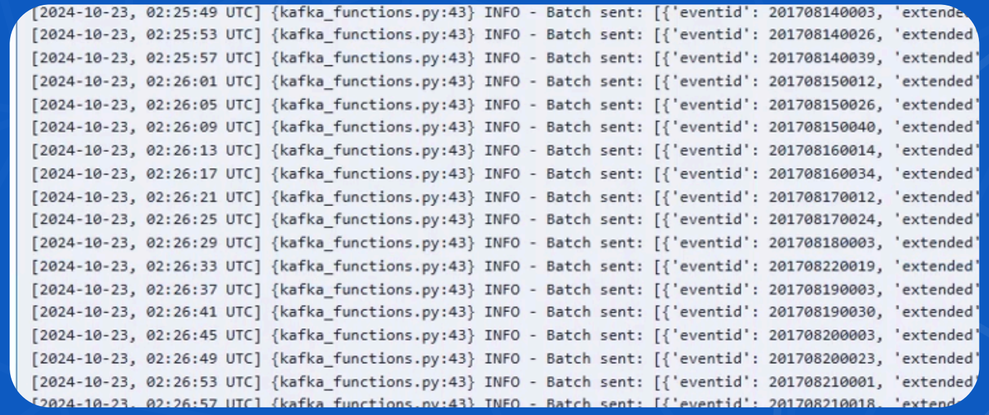
```python
X = df.drop(['eventid', 'id_location', 'id_date',
'id_attack', 'id_perpetrator', 'id_disorder',
'nkill', 'property', 'multiple'], axis=1)

y = df['nkill']
```

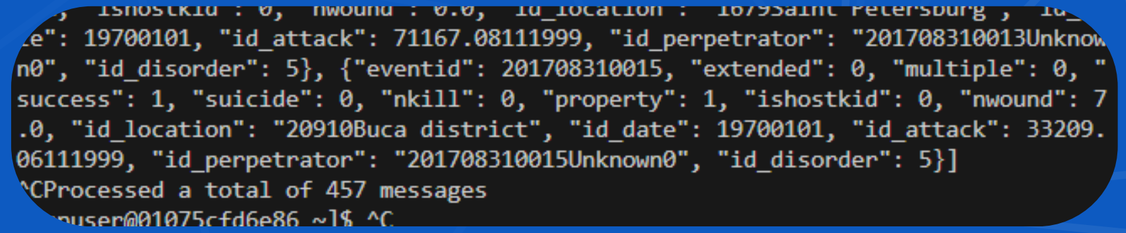
# *Consumer*

kafka_functions.py

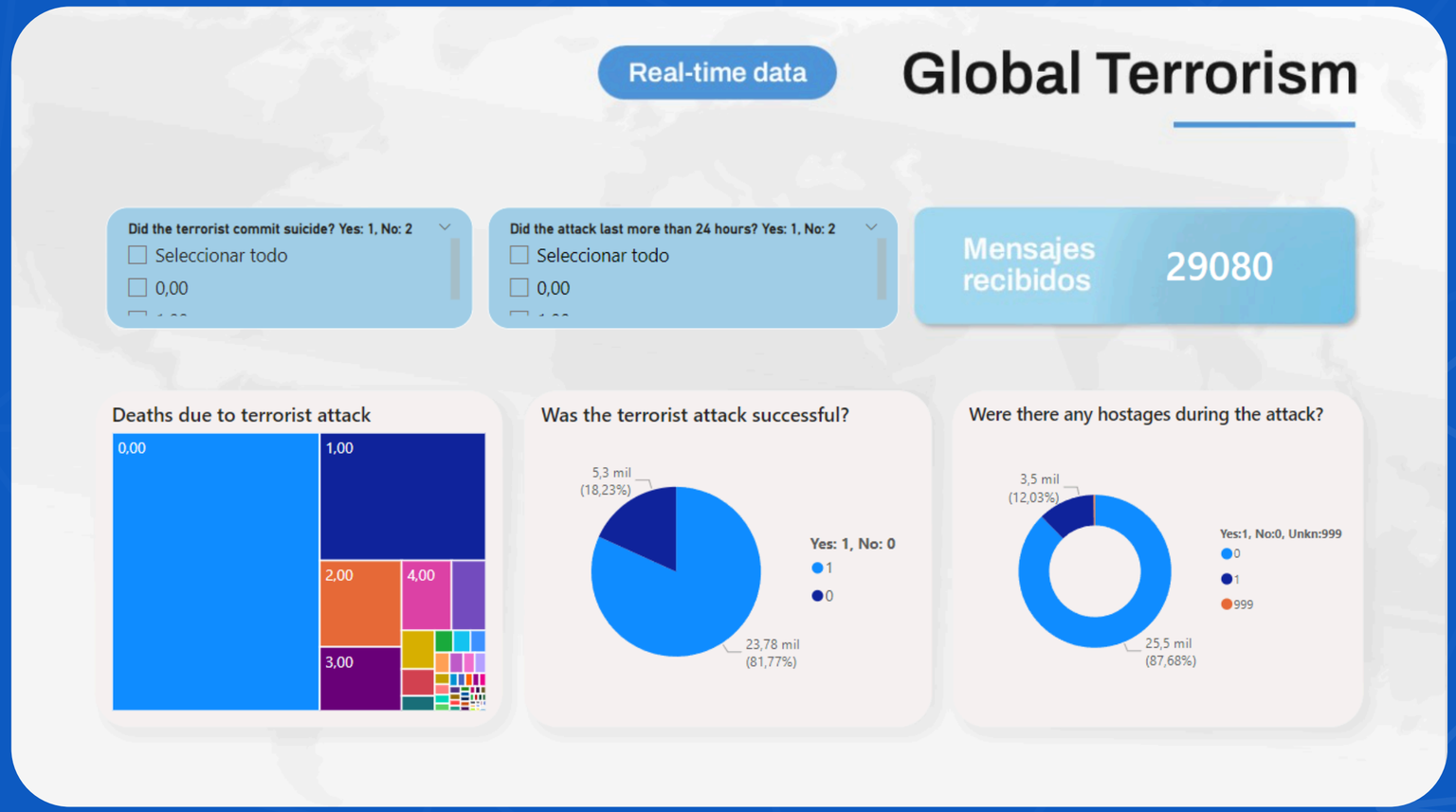
# Kafka



# Real-time dashboard



# Muchas gracias por su atención