



23/04/2024

Apache AIRFLOW

Realización de dags para tratamiento de datos en postgres importados desde csv

Alumno: PEDRO GARCÍA
(Pedro G - pedrokkdownload@gmail.com)

Módulo: AIRFLOW

Docente: ULISES OJEDA

Estudios: Máster experto en Ciencia y Arquitectura de Datos
Ampliación de ARQUITECTURA DE DATOS

Curso: Edición 59

1. INTRODUCCIÓN	1
2. ENTORNO DEL PROYECTO	2
3. CASO DE USO.....	3
4. COMPONENTES UTILIZADOS	4
Conexiones creadas	4
DAGs creados	5
Dataset creados	5
Dependencias de las tareas y grupos de tareas creados	5
5. EJECUCIÓN DAG_PEDROG_PRACTICA_PARTE_1	6
6. EJECUCIÓN DAG_PEDROG_PRACTICA_PARTE_2	9
Comprobación de los registros cargados en las tablas	11
7. EJECUCIÓN DAG_PEDROG_PRACTICA_PARTE_3	13
8. CONCLUSIONES	17

1. Introducción

Este proyecto corresponde al módulo Apache AIRFLOW del máster de BIGDATA de la empresa Datahack en su edición 59.

Descripción de la situación planteada

El objetivo de la práctica es consolidar los conocimientos mediante un proyecto de Airflow que integre varios de los operadores y componentes estudiados durante las clases.

Se desarrollará la práctica con los siguientes requisitos obligatorios:

- Implementar un workflow con una finalidad abierta.
- Utilizar al menos 2 operadores y 1 sensor.
- Definir y utilizar una conexión.

Se valorará positivamente la interacción con bases de datos, el uso de datasets para el scheduling, buenas prácticas a la hora de definir las tareas y la agrupación de tareas.

Entregable del proyecto

El entregable consta de un archivo zip con los siguientes archivos:

- Memoria explicativa en formato pdf: memoria_pedrog_practica_airflow.pdf
- Archivo docker-compose.yaml para levantar el entorno.
- Directorio dags con el código desarrollado para la definición de los dags.

Software utilizado

Plataforma utilizada:

- Intel con arquitectura de 64 bits

Aplicaciones utilizadas en el desarrollo del proyecto:

- Windows 11 v. 23H2
- Firefox Browser v.123.0
- Docker Desktop v.4.27
- Visual Studio Code v. 1.86
- PowerShell v.7.4.2

2. Entorno del proyecto

docker-compose.yaml

Vamos a utilizar el fichero docker-compose.yaml, con el mismo versionado de las imágenes que las utilizadas en las clases de Airflow. En concreto, los contenedores creados del docker-compose.yaml para la práctica son:

- ☐ Bases de datos: postgres, redis
- ☐ Servicios propios de airflow: webserver, scheduler, worker, triggerer, init

Chequeos desde consola: Comprobamos contenedores y salida de mensajes









CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	PORTS	NAMES
07f8fdb57c50	apache/airflow:2.7.1	"/usr/bin/dumb-init ..."	9 hours ago	Up 9 hours (healthy)	0.0.0.0:8080->8080/tcp	airflow_practica-airflow-webserver-1
100c6a9d581f	apache/airflow:2.7.1	"/usr/bin/dumb-init ..."	9 hours ago	Up 9 hours (healthy)	8080/tcp	airflow_practica-airflow-worker-1
0a8a5d968f1c	apache/airflow:2.7.1	"/usr/bin/dumb-init ..."	9 hours ago	Up 9 hours (healthy)	8080/tcp	airflow_practica-airflow-triggerer-1
c4bb9bfff8ee5	apache/airflow:2.7.1	"/usr/bin/dumb-init ..."	9 hours ago	Up 9 hours (healthy)	8080/tcp	airflow_practica-airflow-scheduler-1
ba45074694aa	apache/airflow:2.7.1	"/bin/bash -c 'funct..."	9 hours ago	Exited (0) 9 hours ago		airflow_practica-airflow-init-1
f9a297dcb953	redis:latest	"docker-entrypoint.s..."	9 hours ago	Up 9 hours (healthy)	6379/tcp	airflow_practica-redis-1
0cc84936f44c	postgres:13	"docker-entrypoint.s..."	9 hours ago	Up 9 hours (healthy)	5432/tcp	airflow_practica-postgres-1

```

airflow-scheduler-1 | [2024-04-22T01:42:16.422+0000] {scheduler_job_runner.py:713} INFO - Setting external_id for <TaskInstance: partel.update_dataset manual__2024-04-22T01:42:11.2998
eued] to 376a009c-9439-49af-b04d-b774f2b4ba38
airflow-scheduler-1 | [2024-04-22T01:42:16.423+0000] {scheduler_job_runner.py:723} INFO - TaskInstance Finished: dag_id=partel, task_id=sensor_csv_disponible, run_id>manual__2024-04-2
99869+00:00, map_index=-1, run_start_date=2024-04-22 01:42:15.642345+00:00, run_end_date=2024-04-22 01:42:16.035025+00:00, run_duration=0.39268, state=success, executor_state=success,
, max_tries=0, job_id=369, pool=default_pool, queue=default, priority_weight=2, operator=HttpSensor, queued_dtm=2024-04-22 01:42:14.095271+00:00, queued_by_job_id=127, pid=2533
airflow-worker-1 | [2024-04-22 01:42:16.591: INFO/ForkPoolWorker-15] Filling up the DagBag from /opt/airflow/dags/partel.py
airflow-worker-1 | [2024-04-22 01:42:17.850: INFO/ForkPoolWorker-15] Running <TaskInstance: partel.update_dataset manual__2024-04-22T01:42:11.299869+00:00 [queued]> on host 180c6a9d581f
airflow-worker-1 | [2024-04-22 01:42:18.363: INFO/ForkPoolWorker-15] Task airflow.providers.celery.executors.celery_executor_utils.execute_command[376a009c-9439-49af-b04d-b774f2b4ba38]
ed in 1.976535009998775: None
airflow-webserver-1 | 172.25.0.1 - - [22/Apr/2024:01:42:18 +0000] "GET /object/grid_data?dag_id=partel&num_runs=25 HTTP/1.1" 200 2905 "http://localhost:8080/dags/partel/grid" "Mozilla
s NT 10.0; Win64; x64; rv:125.0) Gecko/20100101 Firefox/125.0"
airflow-scheduler-1 | [2024-04-22T01:42:19.145+0000] {dagrun.py:653} INFO - Marking run <DagRun partel @ 2024-04-22 01:42:11.299869+00:00: manual__2024-04-22T01:42:11.299869+00:00, st
queued_at: 2024-04-22 01:42:11.328382+00:00, externally triggered: True> successful

```

Chequeos desde docker desktop: Contenedores y Volúmenes (nos interesa el de postgres)

Name	Image	Status	CPU (%)	Port(s)	Last start...
 airflow_practica		Running (6/7)	106.98%		9 hours ago
 airflow-webserver-1 07f8fdb57c50	apache/airflow:2.7.1	Running	0.19%	8080:8080	9 hours ago
 airflow-scheduler-1 c4bb9bfff8ee5	apache/airflow:2.7.1	Running	3.89%		9 hours ago
 airflow-triggerer-1 0a8a5d968f1c	apache/airflow:2.7.1	Running	1.47%		9 hours ago
 airflow-worker-1 100c6a9d581f	apache/airflow:2.7.1	Running	99.14%		9 hours ago
 airflow-init-1 ba45074694aa	apache/airflow:2.7.1	Exited	0%		9 hours ago
 postgres-1 0cc84936f44c	postgres:13	Running	2.15%		9 hours ago
 redis-1 f9a297dcb953	redis:latest	Running	0.14%		9 hours ago

Name	Status ↓	Created
airflow_practica_postgres-db-volume	in use	8 days ago

3. Caso de uso

Se va a desarrollar el código de varios dags que nos permitirán manejar diversos componentes en Airflow para cumplir con los requisitos de la práctica.

Nuestro caso de uso tiene tres partes, cuyas acciones principales detallo a continuación:

Parte 1

- Creación de dos tablas en la base de datos postgres: **tweets_url** y **tweets_local**
- Chequeo de disponibilidad del archivo **tweets_data.csv** en una url de github
- Creación en nuestro host local del archivo **tweets_data_local.csv**
- Generación de un dataset_12

Parte 2

- Lectura del dataset_12
- Descarga temporal desde la url del archivo **tweets_data.csv** para utilización en la siguiente tarea.
- Carga de datos en la tabla **tweets_url** utilizando la descarga realizada en la tarea anterior
- Carga de datos en la tabla **tweets_local** utilizando el fichero local **tweets_data_local.csv** creado en la parte 1
- Generación de un dataset_23

Parte 3

- Lectura del dataset_23
- Ejecución de consultas sobre la tabla **tweets_url** utilizando PostgresOperator
- Ejecución de consultas sobre la tabla **tweets_local** utilizando PythonOperator con función para ejecutar consulta y mostrar resultados

En la url https://github.com/pedrokkdownload/airflow_practica hemos creado un repositorio público para el desarrollo de esta práctica. Se ha colocado el fichero **tweets_data.csv**, que tiene información relacionada con diversas líneas aéreas. La ruta de dicho fichero la hemos utilizado como url para los dags:

https://raw.githubusercontent.com/pedrokkdownload/airflow_practica/main/tweets_data.csv

4. Componentes utilizados

Conexiones creadas










Conexión para trabajar con la base de datos postgres

Edit Connection	
Connection Id *	conexion_postgres_practica
Connection Type *	Postgres <small>Connection Type missing? Make sure you've installed the corresponding Airflow Provider Package.</small>
Description	Conexión a la base de datos postgres para la práctica
Host	postgres
Schema	
Login	airflow
Password	
Port	5432

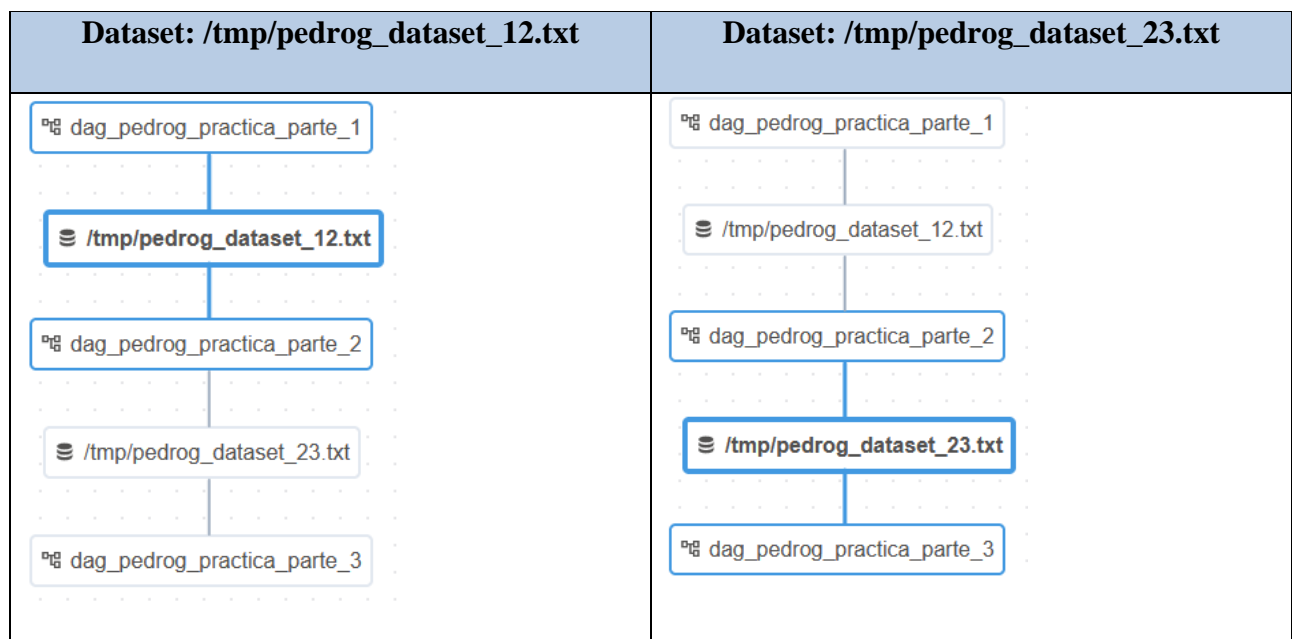
Conexión para trabajar con una url en el host <https://raw.githubusercontent.com/>

Edit Connection	
Connection Id *	conexion_http_practica
Connection Type *	HTTP <small>Connection Type missing? Make sure you've installed the corresponding Airflow Provider Package.</small>
Description	Conexión http a una url de https://raw.githubusercontent.com/ para la práctica
Host	https://raw.githubusercontent.com/
Schema	
Login	
Password	
Port	

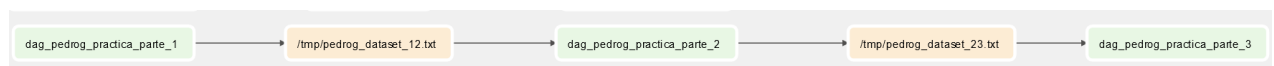
DAGs creados

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks
 dag_pedrog_practica_parte_1 <small>pedrog_practica tweets</small>	airflow		@daily	2024-04-22, 00:00:00	2024-04-23, 00:00:00	
 dag_pedrog_practica_parte_2 <small>pedrog_practica tweets</small>	airflow		Dataset	2024-04-23, 11:51:50	On /tmp/pedrog_dataset_12.txt	
 dag_pedrog_practica_parte_3 <small>pedrog_practica tweets</small>	airflow		Dataset	2024-04-23, 12:23:04	On /tmp/pedrog_dataset_23.txt	

Dataset creados



Dependencias entre los DAGs



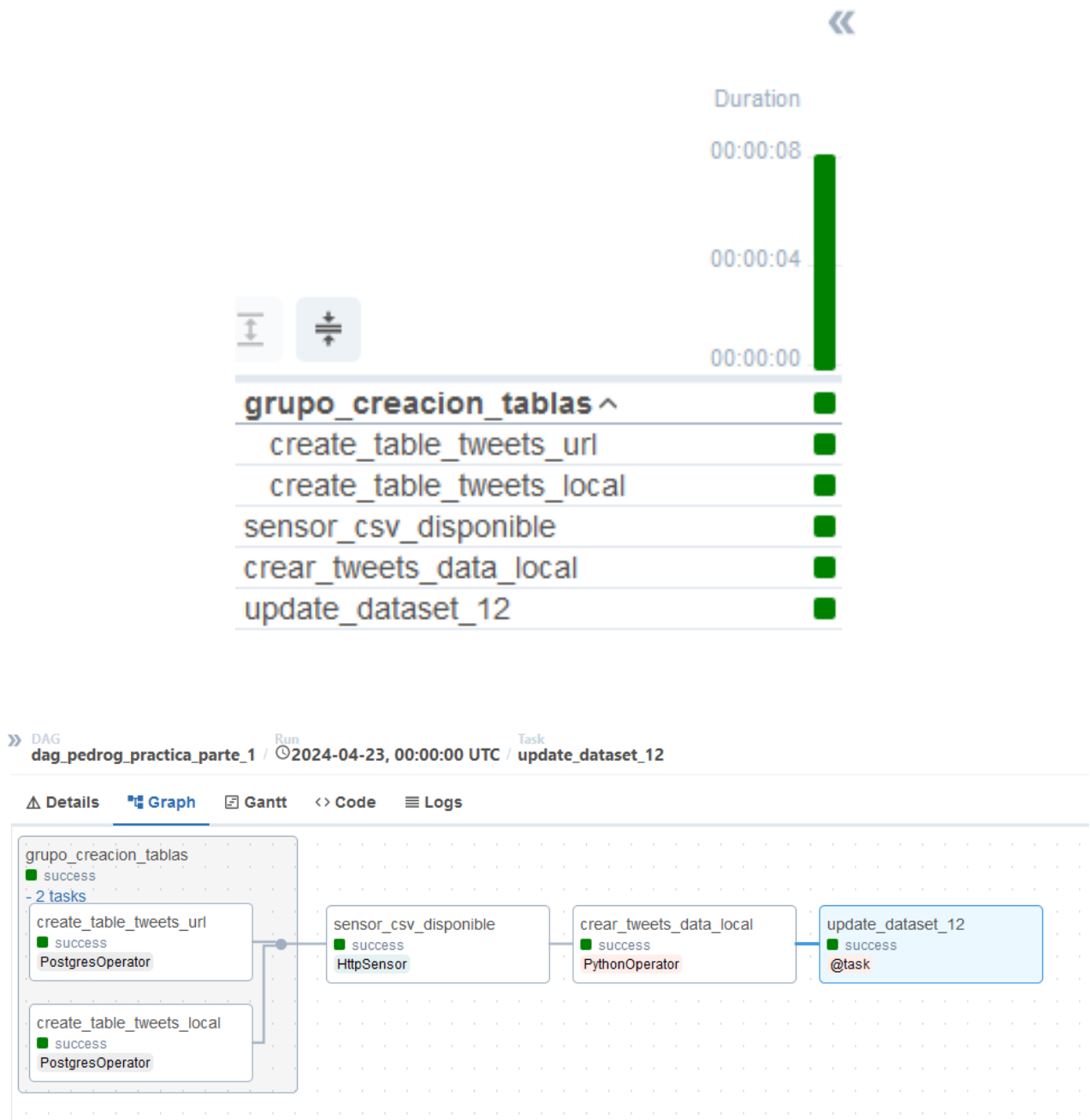
Dependencias de las tareas y grupos de tareas creados

- grupo_creacion_tablas_task >> sensor_csv_disponible >> crear_tweets_data_local_task >> dataset_12_escrito
- dataset_12_leido >> descargar_csv_desde_url_task >> [cargar_tweets_url_task, cargar_tweets_local_task] >> dataset_23_escrito
- dataset_23_leido >> grupo_consultas_tweets_url_task >> grupo_consultas_tweets_local_task

5. Ejecución dag_pedrog_practica_parte_1

Dependencias y orden de las tareas en el DAG

grupo_creacion_tablas_task >> sensor_csv_disponible >> crear_tweets_data_local_task >> dataset_12_escrito



create table tweets url

```

        "user_timezone" text
    );

[2024-04-23, 11:37:01 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 11:37:01 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 11:37:01 UTC] {sql.py:418} INFO - Running statement:
CREATE TABLE IF NOT EXISTS tweets_url(
    "tweet_id" text,
    "airline_sentiment" text,
    "airline_sentiment_confidence" text,
    "negativereason" text NULL,
    "negativereason_confidence" text NULL,
    "airline" text,
    "airline_sentiment_gold" text NULL,
    "name" text,
    "negativereason_gold" text NULL,
    "retweet_count" text,
    "text" text,
    "tweet_coord" text NULL,
    "tweet_created" text,
    "tweet_location" text NULL,
    "user_timezone" text
);
, parameters: None
[2024-04-23, 11:37:02 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_1, task_id=grupo_creacion_tablas.create_table_tweets_url
[2024-04-23, 11:37:02 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 11:37:02 UTC] {taskinstance.py:2776} INFO - 0 downstream tasks scheduled from follow-on schedule check

```

create table tweets local

```

        "user_timezone" text
    );

[2024-04-23, 11:37:01 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 11:37:01 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 11:37:01 UTC] {sql.py:418} INFO - Running statement:
CREATE TABLE IF NOT EXISTS tweets_local(
    "tweet_id" text,
    "airline_sentiment" text,
    "airline_sentiment_confidence" text,
    "negativereason" text NULL,
    "negativereason_confidence" text NULL,
    "airline" text,
    "airline_sentiment_gold" text NULL,
    "name" text,
    "negativereason_gold" text NULL,
    "retweet_count" text,
    "text" text,
    "tweet_coord" text NULL,
    "tweet_created" text,
    "tweet_location" text NULL,
    "user_timezone" text
);
, parameters: None
[2024-04-23, 11:37:01 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_1, task_id=grupo_creacion_tablas.create_table_tweets_local
[2024-04-23, 11:37:02 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 11:37:02 UTC] {taskinstance.py:2776} INFO - 1 downstream tasks scheduled from follow-on schedule check

```

sensor csv disponible

```

[2024-04-23, 10:23:16 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER='***' AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_1' AIRFLOW_CTX_TASK_ID='sensor_csv_disponible'
[2024-04-23, 10:23:16 UTC] {http.py:121} INFO - Poking: pedrokkdownload/***_practica/main/tweets_data.csv
[2024-04-23, 10:23:16 UTC] {base.py:73} INFO - Using connection ID 'conexion_http_practica' for task execution.
[2024-04-23, 10:23:16 UTC] {base.py:287} INFO - Success criteria met. Exiting.
[2024-04-23, 10:23:16 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_1, task_id=sensor_csv_disponible, execution_date=2024-04-22T00:00:00
[2024-04-23, 10:23:16 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 10:23:16 UTC] {taskinstance.py:2776} INFO - 1 downstream tasks scheduled from follow-on schedule check
[2024-04-23, 11:37:03 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_1.sensor_csv_disponible>
[2024-04-23, 11:37:03 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=queueable deps ti=<TaskInstance: dag_pedrog_practica_parte_1.sensor_csv_disponible>
[2024-04-23, 11:37:03 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 11:37:03 UTC] {taskinstance.py:1380} INFO - Executing <Task(HTTPSensor): sensor_csv_disponible> on 2024-04-22 00:00:00+00:00
[2024-04-23, 11:37:03 UTC] {standard_task_runner.py:57} INFO - Started process 4019 to run task
[2024-04-23, 11:37:03 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_1', 'sensor_csv_disponible', 'scheduled__2024-04-22T00:00:00+00:00']
[2024-04-23, 11:37:03 UTC] {standard_task_runner.py:85} INFO - Job 217: Subtask sensor_csv_disponible
[2024-04-23, 11:37:03 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_1.sensor_csv_disponible scheduled__2024-04-22T00:00:00+00:00 [running]> on host ***
[2024-04-23, 11:37:04 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER='***' AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_1' AIRFLOW_CTX_TASK_ID='sensor_csv_disponible'
[2024-04-23, 11:37:04 UTC] {http.py:121} INFO - Poking: pedrokkdownload/***_practica/main/tweets_data.csv
[2024-04-23, 11:37:04 UTC] {base.py:73} INFO - Using connection ID 'conexion_http_practica' for task execution.
[2024-04-23, 11:37:04 UTC] {base.py:287} INFO - Success criteria met. Exiting.
[2024-04-23, 11:37:04 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_1, task_id=sensor_csv_disponible, execution_date=2024-04-22T00:00:00
[2024-04-23, 11:37:04 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 11:37:04 UTC] {taskinstance.py:2776} INFO - 1 downstream tasks scheduled from follow-on schedule check

```

crear tweets data local task

```
[2024-04-23, 09:35:38 UTC] {taskinstance.py:2776} INFO - 1 downstream tasks scheduled from follow-on schedule check
[2024-04-23, 10:23:18 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_1.crear_tweets_data_lo
[2024-04-23, 10:23:18 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_1.crear_tweets_data_lo
[2024-04-23, 10:23:18 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 10:23:18 UTC] {taskinstance.py:1380} INFO - Executing <Task(PythonOperator): crear_tweets_data_local> on 2024-04-22 00:00:00+00:00
[2024-04-23, 10:23:18 UTC] {standard_task_runner.py:57} INFO - Started process 2745 to run task
[2024-04-23, 10:23:18 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_1', 'crear_tweets_data_local', 'scheduled_2024-04-22T00:00:00+00:00']
[2024-04-23, 10:23:18 UTC] {standard_task_runner.py:85} INFO - Job 159: Subtask crear_tweets_data_local
[2024-04-23, 10:23:18 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_1.crear_tweets_data_local scheduled__2024-04-22T00:00:00+00:00 [running]>
[2024-04-23, 10:23:19 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER=*** AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_1' AIRFLOW_CTX_TASK_ID='crear_tweets_data_lo
[2024-04-23, 10:23:19 UTC] {logging_mixin.py:151} INFO - Archivo descargado y guardado en /opt/***/dags/tweets_data_local.csv
[2024-04-23, 10:23:19 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 10:23:19 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_1, task_id=crear_tweets_data_local, execution_date=20240422T000000, sta
[2024-04-23, 10:23:19 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 10:23:19 UTC] {taskinstance.py:2776} INFO - 1 downstream tasks scheduled from follow-on schedule check
[2024-04-23, 11:37:06 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_1.crear_tweets_data_lo
[2024-04-23, 11:37:06 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_1.crear_tweets_data_lo
[2024-04-23, 11:37:06 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 11:37:06 UTC] {taskinstance.py:1380} INFO - Executing <Task(PythonOperator): crear_tweets_data_local> on 2024-04-22 00:00:00+00:00
[2024-04-23, 11:37:06 UTC] {standard_task_runner.py:57} INFO - Started process 4022 to run task
[2024-04-23, 11:37:06 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_1', 'crear_tweets_data_local', 'scheduled_2024-04-22T00:00:00+00:00']
[2024-04-23, 11:37:06 UTC] {standard_task_runner.py:85} INFO - Job 218: Subtask crear_tweets_data_local
[2024-04-23, 11:37:06 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_1.crear_tweets_data_local scheduled__2024-04-22T00:00:00+00:00 [running]>
[2024-04-23, 11:37:06 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER=*** AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_1' AIRFLOW_CTX_TASK_ID='crear_tweets_data_lo
[2024-04-23, 11:37:06 UTC] {logging_mixin.py:151} INFO - Archivo descargado y guardado en /opt/***/dags/tweets_data_local.csv
[2024-04-23, 11:37:06 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 11:37:06 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_1, task_id=crear_tweets_data_local, execution_date=20240422T000000, sta
[2024-04-23, 11:37:06 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 11:37:06 UTC] {taskinstance.py:2776} INFO - 1 downstream tasks scheduled from follow-on schedule check
```

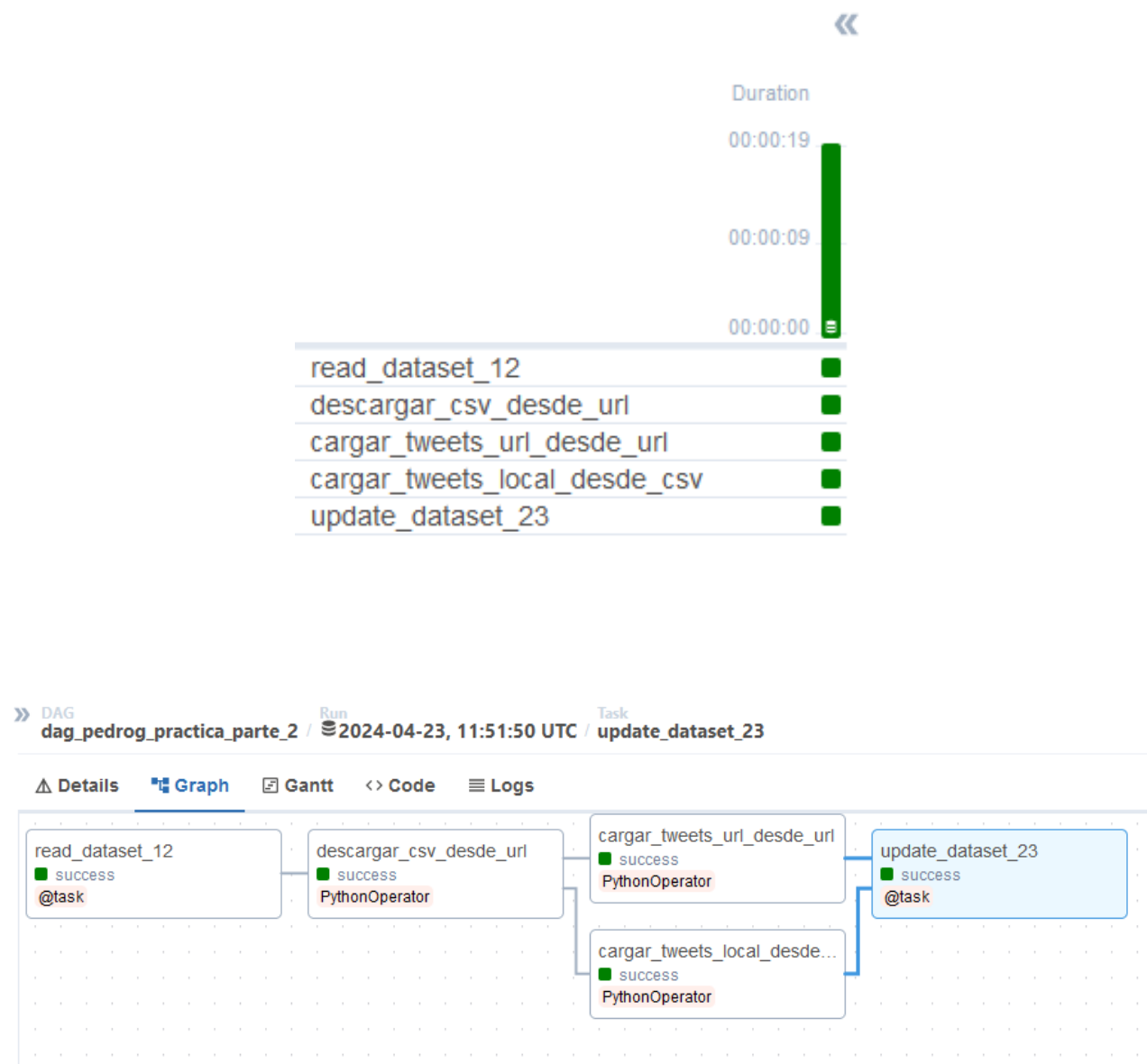
dataset 12 escrito

```
[2024-04-23, 10:23:21 UTC] {standard_task_runner.py:57} INFO - Started process 4724 to run task
[2024-04-23, 10:23:21 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_1', 'update_dataset_12', 'scheduled_2024-04-22T00:00:00+00:00']
[2024-04-23, 10:23:21 UTC] {standard_task_runner.py:85} INFO - Job 161: Subtask update_dataset_12
[2024-04-23, 10:23:21 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_1.update_dataset_12 scheduled__2024-04-22T00:00:00+00:00 [running]> on host
[2024-04-23, 10:23:21 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER=*** AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_1' AIRFLOW_CTX_TASK_ID='update_dataset_12'
[2024-04-23, 10:23:21 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 10:23:21 UTC] {warnings.py:109} WARNING - /home/***/.local/lib/python3.8/site-packages/***/models/baseoperator.py:1473: RemovedInAirflow3Warning: Passing 'execution_da
context["ti"].xcom_push(key=key, value=value, execution_date=execution_date)
[2024-04-23, 10:23:21 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_1, task_id=update_dataset_12, execution_date=20240422T000000, sta
[2024-04-23, 10:23:21 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 10:23:21 UTC] {taskinstance.py:2776} INFO - 0 downstream tasks scheduled from follow-on schedule check
[2024-04-23, 11:37:08 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_1.update_dataset_12 s
[2024-04-23, 11:37:08 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_1.update_dataset_12 sched
[2024-04-23, 11:37:08 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 11:37:08 UTC] {taskinstance.py:1380} INFO - Executing <Task(PythonDecoratedOperator): update_dataset_12> on 2024-04-22 00:00:00+00:00
[2024-04-23, 11:37:08 UTC] {standard_task_runner.py:57} INFO - Started process 4025 to run task
[2024-04-23, 11:37:08 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_1', 'update_dataset_12', 'scheduled_2024-04-22T00:00:00+00:00']
[2024-04-23, 11:37:08 UTC] {standard_task_runner.py:85} INFO - Job 219: Subtask update_dataset_12
[2024-04-23, 11:37:08 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_1.update_dataset_12 scheduled__2024-04-22T00:00:00+00:00 [running]> on host
[2024-04-23, 11:37:08 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER=*** AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_1' AIRFLOW_CTX_TASK_ID='update_dataset_12'
[2024-04-23, 11:37:08 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 11:37:08 UTC] {warnings.py:109} WARNING - /home/***/.local/lib/python3.8/site-packages/***/models/baseoperator.py:1473: RemovedInAirflow3Warning: Passing 'execution_da
context["ti"].xcom_push(key=key, value=value, execution_date=execution_date)
[2024-04-23, 11:37:08 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_1, task_id=update_dataset_12, execution_date=20240422T000000, sta
[2024-04-23, 11:37:08 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 11:37:08 UTC] {taskinstance.py:2776} INFO - 0 downstream tasks scheduled from follow-on schedule check
```

6. Ejecución dag_pedrog_practica_parte_2

Dependencias y orden de las tareas en el DAG

dataset_12_leido >> descargar_csv_desde_url_task >> [cargar_tweets_url_task, cargar_tweets_local_task] >> dataset_23_escrito



dataset 12 leido

```

b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_2/run_id=dataset_triggered_2024-04-23T11:51:50.838452+00:00/task_id=read_dataset_12/attempt=1.log
[2024-04-23, 11:52:43 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_2.read_data
[2024-04-23, 11:52:43 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_2.read_dataset_
[2024-04-23, 11:52:43 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 11:52:44 UTC] {taskinstance.py:1380} INFO - Executing <Task(PythonDecoratedOperator): read_dataset_12> on 2024-04-23 11:51:50.838452+00:00
[2024-04-23, 11:52:44 UTC] {standard_task_runner.py:57} INFO - Started process 4276 to run task
[2024-04-23, 11:52:44 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_2', 'read_dataset_12', 'dataset_triggered_2024
[2024-04-23, 11:52:44 UTC] {standard_task_runner.py:85} INFO - Job 225: Subtask read_dataset_12
[2024-04-23, 11:52:44 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_2.read_dataset_12 dataset_triggered_2024-04-23T11:51:50.838452+0
[2024-04-23, 11:52:44 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER='***' AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_2' AIRFLOW_CTX_TASK
[2024-04-23, 11:52:44 UTC] {logging_mixin.py:151} INFO - Producer update!Producer update!Producer update!Producer update!Producer update!
[2024-04-23, 11:52:44 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 11:52:44 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_2, task_id=read_dataset_12, execution_date=20240423T115
[2024-04-23, 11:52:44 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 11:52:44 UTC] {taskinstance.py:2776} INFO - 1 downstream tasks scheduled from follow-on schedule check

```

descargar csv desde url task

```

b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_2/run_id=dataset_triggered_2024-04-23T11:51:50.838452+00:00/task_id=descargar_csv_desde_url/attempt=1.log
[2024-04-23, 11:52:46 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_2.desca
[2024-04-23, 11:52:46 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_2.descargar
[2024-04-23, 11:52:46 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 11:52:46 UTC] {taskinstance.py:1380} INFO - Executing <Task(PythonOperator): descargar_csv_desde_url> on 2024-04-23 11:51:50.838452+00:00
[2024-04-23, 11:52:46 UTC] {standard_task_runner.py:57} INFO - Started process 4279 to run task
[2024-04-23, 11:52:46 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_2', 'descargar_csv_desde_url', 'dataset_tri
[2024-04-23, 11:52:46 UTC] {standard_task_runner.py:85} INFO - Job 226: Subtask descargar_csv_desde_url
[2024-04-23, 11:52:46 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_2.descargar_csv_desde_url dataset_triggered_2024-04-23T11:51
[2024-04-23, 11:52:46 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER='***' AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_2' AIRFLOW_CTX_
[2024-04-23, 11:52:46 UTC] {python.py:194} INFO - Done. Returned value was: /tmp/tweets_data.csv
[2024-04-23, 11:52:46 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_2, task_id=descargar_csv_desde_url, execution_date=
[2024-04-23, 11:52:46 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 11:52:46 UTC] {taskinstance.py:2776} INFO - 2 downstream tasks scheduled from follow-on schedule check

```

cargar tweets url task

```

[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569591533617307648', 'negative', '1.0', 'Cancelled Flight', '1.0', 'American', '',
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569591393540288512', 'negative', '1.0', 'Cancelled Flight', '1.0', 'American', '',
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569591285150908416', 'positive', '1.0', '', '', 'American', '', 'iambmac', '', '0
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569591136534319105', 'negative', '1.0', 'Bad Flight', '0.6774', 'American', '', '
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569590988395708416', 'positive', '1.0', '', '', 'American', '', 'howiemandel', ''
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569590965880532993', 'negative', '1.0', 'Customer Service Issue', '1.0', 'American
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569590892085915649', 'negative', '1.0', 'Customer Service Issue', '1.0', 'American
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569590191758962688', 'negative', '1.0', 'Late Flight', '0.3358', 'American', '',
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569590013278756865', 'positive', '0.6274', '', '0.0', 'American', '', 'Flora Lola
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569589959088173056', 'negative', '1.0', 'Can't Tell', '1.0', 'American', '', 'you
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569589643487928321', 'positive', '1.0', '', '', 'American', '', 'DrCaseyJRudkin',
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569589460226183168', 'negative', '1.0', 'Late Flight', '1.0', 'American', '', 'ca
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569588816438169600', 'negative', '1.0', 'Cancelled Flight', '1.0', 'American', '',
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569588651925098496', 'positive', '1.0', '', '', 'American', '', 'j1halldc', '', '
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569588591602458624', 'negative', '1.0', 'Customer Service Issue', '1.0', 'American
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569588473050611712', 'positive', '1.0', '', '', 'American', '', 'Laurelinesblog',
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569588464896876545', 'negative', '1.0', 'Bad Flight', '1.0', 'American', '', 'MDDi
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569587813856841728', 'neutral', '0.6759999999999999', '', '0.0', 'American', '',
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569587705937600512', 'negative', '1.0', 'Cancelled Flight', '1.0', 'American', '',
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569587691626622976', 'negative', '0.6684', 'Late Flight', '0.6684', 'American', '',
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569587686496825344', 'positive', '0.3487', '', '0.0', 'American', '', 'KristenRee
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569587371693350008', 'negative', '1.0', 'Customer Service Issue', '1.0', 'American
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569587242672398336', 'neutral', '1.0', '', '', 'American', '', 'sanyabun', '', '0
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569587188687634433', 'negative', '1.0', 'Customer Service Issue', '0.6659', 'Amer
[2024-04-23, 11:52:59 UTC] {logging_mixin.py:151} INFO - Intentando insertar: ('569587140490866689', 'neutral', '0.6771', '', '0.0', 'American', '', 'davidtdtw',
[2024-04-23, 11:52:59 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 11:52:59 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_2, task_id=cargar_tweets_url_desde_url, execut
[2024-04-23, 11:52:59 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 11:52:59 UTC] {taskinstance.py:2776} INFO - 1 downstream tasks scheduled from follow-on schedule check

```

cargar tweets local task

```
b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_2/run_id=dataset_triggered_2024-04-23T11:51:50.838452+00:00/task_id=cargar_tweets_local_desde_csv/attempt=1.log
[2024-04-23, 11:52:48 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_2.cargar_tweets_local_desde_csv>
[2024-04-23, 11:52:48 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_2.cargar_tweets_local_desde_csv>
[2024-04-23, 11:52:48 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 11:52:48 UTC] {taskinstance.py:1380} INFO - Executing <Task(PythonOperator): cargar_tweets_local_desde_csv> on 2024-04-23 11:51:50.838452+00:00
[2024-04-23, 11:52:48 UTC] {standard_task_runner.py:57} INFO - Started process 4285 to run task
[2024-04-23, 11:52:48 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_2', 'cargar_tweets_local_desde_csv', 'dataset_triggered_2024-04-23T11:51:50.838452+00:00']
[2024-04-23, 11:52:48 UTC] {standard_task_runner.py:85} INFO - Job 228: Subtask cargar_tweets_local_desde_csv
[2024-04-23, 11:52:48 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_2.cargar_tweets_local_desde_csv dataset_triggered_2024-04-23T11:51:50.838452+00:00>
[2024-04-23, 11:52:48 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER=*** AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_2' AIRFLOW_CTX_TASK_ID='cargar_tweets_local_desde_csv'
[2024-04-23, 11:52:48 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 11:52:48 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 11:52:48 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_2, task_id=cargar_tweets_local_desde_csv, execution_date=2024-04-23T11:51:50.838452+00:00
[2024-04-23, 11:52:49 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 11:52:49 UTC] {taskinstance.py:2776} INFO - 0 downstream tasks scheduled from follow-on schedule check
```

dataset 23 escrito

```
b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_2/run_id=dataset_triggered_2024-04-23T11:51:50.838452+00:00/task_id=update_dataset_23/attempt=1.log
[2024-04-23, 11:53:01 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_2.update_dataset_23>
[2024-04-23, 11:53:01 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_2.update_dataset_23>
[2024-04-23, 11:53:01 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 11:53:01 UTC] {taskinstance.py:1380} INFO - Executing <Task(PythonDecoratedOperator): update_dataset_23> on 2024-04-23 11:51:50.838452+00:00
[2024-04-23, 11:53:01 UTC] {standard_task_runner.py:57} INFO - Started process 4296 to run task
[2024-04-23, 11:53:01 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_2', 'update_dataset_23', 'dataset_triggered_2024-04-23T11:51:50.838452+00:00']
[2024-04-23, 11:53:01 UTC] {standard_task_runner.py:85} INFO - Job 229: Subtask update_dataset_23
[2024-04-23, 11:53:01 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_2.update_dataset_23 dataset_triggered_2024-04-23T11:51:50.838452+00:00>
[2024-04-23, 11:53:01 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER=*** AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_2' AIRFLOW_CTX_TASK_ID='update_dataset_23'
[2024-04-23, 11:53:01 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 11:53:01 UTC] {warnings.py:109} WARNING - /home/***/.local/lib/python3.8/site-packages/***/models/baseoperator.py:1473: RemovedInAirflow3Warning: Passing 'execution_date' as a keyword argument is deprecated. Use 'context["ti"].xcom_push(key=key, value=value, execution_date=execution_date)' instead.
[2024-04-23, 11:53:01 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_2, task_id=update_dataset_23, execution_date=2024-04-23T11:51:50.838452+00:00
[2024-04-23, 11:53:01 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 11:53:01 UTC] {taskinstance.py:2776} INFO - 0 downstream tasks scheduled from follow-on schedule check
```

Comprobación de los registros cargados en las tablas

Vamos a ver la carga de datos en ambas tablas con una consulta desde consola

```
PS C:\Users\Pedro\Documents\master59\airflow_practica> docker exec -it airflow_practica-postgres-1 bash
root@d7ff66ab6f23:/# psql -U airflow
psql (13.14 (Debian 13.14-1.pgdg120+2))
Type "help" for help.

airflow=# select count(*) from tweets_url;
 count
-----
 14640
(1 row)

airflow=# select count(*) from tweets_local;
 count
-----
 14640
(1 row)
```


Como curiosidad, ambas tablas empiezan la carga de datos a la vez, ya que las tareas están diversificadas y no secuenciales, ejecutándose ambas a la vez, pero la ejecución de la carga en la tabla **tweets_url** finaliza unos 10 segundos más tarde que la ejecución de la carga en la tabla **tweets_local**. La explicación radica en el procedimiento utilizado para la carga de ambas tablas.

Para cargar la tabla **tweets_local** utilizamos el comando COPY de la fuente (archivo local /dags/tweets_data_local.csv) y para cargar la tabla **tweets_url** utilizamos la inserción línea a línea de la fuente (archivo /tmp/tweets_data.csv)

7. Ejecución dag_pedrog_practica_parte_3

Dependencias y orden de las tareas en el DAG

dataset_23_leido >> grupo_consultas_tweets_url_task >> grupo_consultas_tweets_local_task



read_dataset 23

```
b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_3/run_id=manual__2024-04-23T12:23:04.377375+00:00/task_id=read_dataset_23/attempt=1.log
[2024-04-23, 12:23:06 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.read_dataset_23
[2024-04-23, 12:23:06 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.read_dataset_23 manu
[2024-04-23, 12:23:06 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 12:23:06 UTC] {taskinstance.py:1380} INFO - Executing <Task(PythonDecoratedOperator): read_dataset_23> on 2024-04-23 12:23:04.377375+00:00
[2024-04-23, 12:23:06 UTC] {standard_task_runner.py:57} INFO - Started process 4768 to run task
[2024-04-23, 12:23:06 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_3', 'read_dataset_23', 'manual__2024-04-23T12:23:04.37
[2024-04-23, 12:23:06 UTC] {standard_task_runner.py:85} INFO - Job 239: Subtask read_dataset_23
[2024-04-23, 12:23:06 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER=*** AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_3' AIRFLOW_CTX_TASK_ID='re
[2024-04-23, 12:23:06 UTC] {logging_mixin.py:151} INFO - Producer update!Producer update!Producer update!Producer update!Producer update!
[2024-04-23, 12:23:06 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 12:23:06 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_3, task_id=read_dataset_23, execution_date=20240423T122304, st
[2024-04-23, 12:23:06 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 12:23:06 UTC] {taskinstance.py:2776} INFO - 4 downstream tasks scheduled from follow-on schedule check
```

q1 tweets url group by

```
b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_3/run_id=manual__2024-04-23T12:23:04.377375+00:00/task_id=grupo_consultas_tweets_url.q1_tweets_url_
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.g
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1380} INFO - Executing <Task(PostgresOperator): grupo_consultas_tweets_url.q1_tweets_url_group_by> on 2024-04-23 1
[2024-04-23, 12:23:08 UTC] {standard_task_runner.py:57} INFO - Started process 4785 to run task
[2024-04-23, 12:23:08 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_3', 'grupo_consultas_tweets_url.q1_
[2024-04-23, 12:23:08 UTC] {standard_task_runner.py:85} INFO - Job 240: Subtask grupo_consultas_tweets_url.q1_tweets_url_group_by
[2024-04-23, 12:23:08 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_3.grupo_consultas_tweets_url.q1_tweets_url_group_by m
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER=*** AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_3' AIRF
[2024-04-23, 12:23:08 UTC] {sql.py:274} INFO - Executing: SELECT airline, COUNT(*) FROM tweets_url GROUP BY airline;
[2024-04-23, 12:23:08 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 12:23:09 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 12:23:09 UTC] {sql.py:418} INFO - Running statement: SELECT airline, COUNT(*) FROM tweets_url GROUP BY airline;; parameters: None
[2024-04-23, 12:23:09 UTC] {sql.py:427} INFO - Rows affected: 1023
[2024-04-23, 12:23:09 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_3, task_id=grupo_consultas_tweets_url.q1.tw
[2024-04-23, 12:23:09 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 12:23:09 UTC] {taskinstance.py:2776} INFO - 0 downstream tasks scheduled from follow-on schedule check
```

q2 tweets url count

```
b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_3/run_id=manual__2024-04-23T12:23:04.377375+00:00/task_id=grupo_consultas_tweets_url.q2_tweets_url_co
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.gru
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1380} INFO - Executing <Task(PostgresOperator): grupo_consultas_tweets_url.q2_tweets_url_count> on 2024-04-23 12:23:0
[2024-04-23, 12:23:08 UTC] {standard_task_runner.py:57} INFO - Started process 4786 to run task
[2024-04-23, 12:23:08 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_3', 'grupo_consultas_tweets_url.q2.tw
[2024-04-23, 12:23:08 UTC] {standard_task_runner.py:85} INFO - Job 241: Subtask grupo_consultas_tweets_url.q2_tweets_url_count
[2024-04-23, 12:23:08 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_3.grupo_consultas_tweets_url.q2_tweets_url_count manual
[2024-04-23, 12:23:09 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER=*** AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_3' AIRFLOW
[2024-04-23, 12:23:09 UTC] {sql.py:274} INFO - Executing: SELECT COUNT(*) FROM tweets_url;
[2024-04-23, 12:23:09 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 12:23:09 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 12:23:09 UTC] {sql.py:418} INFO - Running statement: SELECT COUNT(*) FROM tweets_url;; parameters: None
[2024-04-23, 12:23:09 UTC] {sql.py:427} INFO - Rows affected: 1
[2024-04-23, 12:23:09 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_3, task_id=grupo_consultas_tweets_url.q2.twee
[2024-04-23, 12:23:09 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 12:23:09 UTC] {taskinstance.py:2776} INFO - 4 downstream tasks scheduled from follow-on schedule check
```


q3 tweets url max

```

b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_3/run_id>manual_2024-04-23T12:23:04.377375+00:00/task_id=grupo_consultas_tweets_url.q3_tweets_url_max
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.grupo
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1380} INFO - Executing <Task(PostgresOperator): grupo_consultas_tweets_url.q3_tweets_url_max> on 2024-04-23 12:23:04.
[2024-04-23, 12:23:08 UTC] {standard_task_runner.py:57} INFO - Started process 4788 to run task
[2024-04-23, 12:23:08 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_3', 'grupo_consultas_tweets_url.q3_tweets
[2024-04-23, 12:23:08 UTC] {standard_task_runner.py:85} INFO - Job 243: Subtask grupo_consultas_tweets_url.q3_tweets_url_max
[2024-04-23, 12:23:08 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_3.grupo_consultas_tweets_url.q3_tweets_url_max manual_2
[2024-04-23, 12:23:09 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER='***' AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_3' AIRFLOW
[2024-04-23, 12:23:09 UTC] {sql.py:274} INFO - Executing: SELECT MAX(airline_sentiment_confidence) FROM tweets_url;
[2024-04-23, 12:23:09 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 12:23:09 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 12:23:09 UTC] {sql.py:418} INFO - Running statement: SELECT MAX(airline_sentiment_confidence) FROM tweets_url; parameters: None
[2024-04-23, 12:23:09 UTC] {sql.py:427} INFO - Rows affected: 1
[2024-04-23, 12:23:09 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_3, task_id=grupo_consultas_tweets_url.q3_tweets
[2024-04-23, 12:23:09 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 12:23:09 UTC] {taskinstance.py:2776} INFO - 0 downstream tasks scheduled from follow-on schedule check

```

q4 tweets url min

```

b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_3/run_id>manual_2024-04-23T12:23:04.377375+00:00/task_id=grupo_consultas_tweets_url.q4_tweets_url_min
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.grupo
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 12:23:08 UTC] {taskinstance.py:1380} INFO - Executing <Task(PostgresOperator): grupo_consultas_tweets_url.q4_tweets_url_min> on 2024-04-23 12:23:04.
[2024-04-23, 12:23:08 UTC] {standard_task_runner.py:57} INFO - Started process 4787 to run task
[2024-04-23, 12:23:08 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_3', 'grupo_consultas_tweets_url.q4_tweets
[2024-04-23, 12:23:08 UTC] {standard_task_runner.py:85} INFO - Job 242: Subtask grupo_consultas_tweets_url.q4_tweets_url_min
[2024-04-23, 12:23:08 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_3.grupo_consultas_tweets_url.q4_tweets_url_min manual_2
[2024-04-23, 12:23:09 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER='***' AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_3' AIRFLOW
[2024-04-23, 12:23:09 UTC] {sql.py:274} INFO - Executing: SELECT MIN(airline_sentiment_confidence) FROM tweets_url;
[2024-04-23, 12:23:09 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 12:23:09 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 12:23:09 UTC] {sql.py:418} INFO - Running statement: SELECT MIN(airline_sentiment_confidence) FROM tweets_url; parameters: None
[2024-04-23, 12:23:09 UTC] {sql.py:427} INFO - Rows affected: 1
[2024-04-23, 12:23:09 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_3, task_id=grupo_consultas_tweets_url.q4_tweets
[2024-04-23, 12:23:09 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 12:23:09 UTC] {taskinstance.py:2776} INFO - 0 downstream tasks scheduled from follow-on schedule check

```

q1 tweets local group by

```

b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_3/run_id>manual_2024-04-23T12:23:04.377375+00:00/task_id=grupo_consultas_tweets_local.q1_tweets_l
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.grupo
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1380} INFO - Executing <Task(PythonOperator): grupo_consultas_tweets_local.q1_tweets_local_group_by> on 2024-04-23 12:23:11.
[2024-04-23, 12:23:11 UTC] {standard_task_runner.py:57} INFO - Started process 4798 to run task
[2024-04-23, 12:23:11 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_3', 'grupo_consultas_tweets_local.
[2024-04-23, 12:23:11 UTC] {standard_task_runner.py:85} INFO - Job 244: Subtask grupo_consultas_tweets_local.q1_tweets_local_group_by
[2024-04-23, 12:23:11 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_3.grupo_consultas_tweets_local.q1_tweets_local_group_by manual_2
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER='***' AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_3' AIRFLOW
[2024-04-23, 12:23:11 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 12:23:11 UTC] {logging_mixin.py:151} INFO - ('Southwest', 2420)
[2024-04-23, 12:23:11 UTC] {logging_mixin.py:151} INFO - ('Virgin America', 504)
[2024-04-23, 12:23:11 UTC] {logging_mixin.py:151} INFO - ('Delta', 2222)
[2024-04-23, 12:23:11 UTC] {logging_mixin.py:151} INFO - ('American', 2759)
[2024-04-23, 12:23:11 UTC] {logging_mixin.py:151} INFO - ('US Airways', 2913)
[2024-04-23, 12:23:11 UTC] {logging_mixin.py:151} INFO - ('United', 3822)
[2024-04-23, 12:23:11 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_3, task_id=grupo_consultas_tweets_local.q1_tweets
[2024-04-23, 12:23:11 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 12:23:11 UTC] {taskinstance.py:2776} INFO - 0 downstream tasks scheduled from follow-on schedule check

```

q2 tweets local count

```
b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_3/run_id=manual__2024-04-23T12:23:04.377375+00:00/task_id=grupo_consultas_tweets_local.q2_tweets_lo
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.g
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.g
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1380} INFO - Executing <Task(PythonOperator): grupo_consultas_tweets_local.q2_tweets_local_count> on 2024-04-23 12
[2024-04-23, 12:23:11 UTC] {standard_task_runner.py:57} INFO - Started process 4797 to run task
[2024-04-23, 12:23:11 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_3', 'grupo_consultas_tweets_local.q
[2024-04-23, 12:23:11 UTC] {standard_task_runner.py:85} INFO - Job 245: Subtask grupo_consultas_tweets_local.q2_tweets_local_count
[2024-04-23, 12:23:11 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_3.grupo_consultas_tweets_local.q2_tweets_local_count
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER='***' AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_3' AIRF
[2024-04-23, 12:23:11 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 12:23:11 UTC] {logging_mixin.py:151} INFO - (14640,)
[2024-04-23, 12:23:11 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_3, task_id=grupo_consultas_tweets_local.q2.
[2024-04-23, 12:23:11 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 12:23:11 UTC] {taskinstance.py:2776} INFO - 0 downstream tasks scheduled from follow-on schedule check
```

q3 tweets local max

```
b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_3/run_id=manual__2024-04-23T12:23:04.377375+00:00/task_id=grupo_consultas_tweets_local.q3_tweets_loc
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.g
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.g
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1380} INFO - Executing <Task(PythonOperator): grupo_consultas_tweets_local.q3_tweets_local_max> on 2024-04-23 12:23
[2024-04-23, 12:23:11 UTC] {standard_task_runner.py:57} INFO - Started process 4800 to run task
[2024-04-23, 12:23:11 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_3', 'grupo_consultas_tweets_local.q3
[2024-04-23, 12:23:11 UTC] {standard_task_runner.py:85} INFO - Job 247: Subtask grupo_consultas_tweets_local.q3_tweets_local_max
[2024-04-23, 12:23:11 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_3.grupo_consultas_tweets_local.q3_tweets_local_max man
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER='***' AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_3' AIRFL
[2024-04-23, 12:23:11 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 12:23:11 UTC] {logging_mixin.py:151} INFO - ('1.0',)
[2024-04-23, 12:23:11 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_3, task_id=grupo_consultas_tweets_local.q3_t
[2024-04-23, 12:23:11 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 12:23:11 UTC] {taskinstance.py:2776} INFO - 0 downstream tasks scheduled from follow-on schedule check
```

q4 tweets local min

```
b567ccabe45f
*** Found local files:
*** * /opt/airflow/logs/dag_id=dag_pedrog_practica_parte_3/run_id=manual__2024-04-23T12:23:04.377375+00:00/task_id=grupo_consultas_tweets_local.q4_tweets_lo
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.g
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1157} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: dag_pedrog_practica_parte_3.g
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1359} INFO - Starting attempt 1 of 1
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1380} INFO - Executing <Task(PythonOperator): grupo_consultas_tweets_local.q4_tweets_local_min> on 2024-04-23 12:23
[2024-04-23, 12:23:11 UTC] {standard_task_runner.py:57} INFO - Started process 4799 to run task
[2024-04-23, 12:23:11 UTC] {standard_task_runner.py:84} INFO - Running: ['***', 'tasks', 'run', 'dag_pedrog_practica_parte_3', 'grupo_consultas_tweets_local.q
[2024-04-23, 12:23:11 UTC] {standard_task_runner.py:85} INFO - Job 246: Subtask grupo_consultas_tweets_local.q4_tweets_local_min
[2024-04-23, 12:23:11 UTC] {task_command.py:415} INFO - Running <TaskInstance: dag_pedrog_practica_parte_3.grupo_consultas_tweets_local.q4_tweets_local_min mar
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1660} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER='***' AIRFLOW_CTX_DAG_ID='dag_pedrog_practica_parte_3' AIRFI
[2024-04-23, 12:23:11 UTC] {base.py:73} INFO - Using connection ID 'conexion_postgres_practica' for task execution.
[2024-04-23, 12:23:11 UTC] {logging_mixin.py:151} INFO - ('0.335',)
[2024-04-23, 12:23:11 UTC] {python.py:194} INFO - Done. Returned value was: None
[2024-04-23, 12:23:11 UTC] {taskinstance.py:1398} INFO - Marking task as SUCCESS. dag_id=dag_pedrog_practica_parte_3, task_id=grupo_consultas_tweets_local.q4_t
[2024-04-23, 12:23:11 UTC] {local_task_job_runner.py:228} INFO - Task exited with return code 0
[2024-04-23, 12:23:11 UTC] {taskinstance.py:2776} INFO - 0 downstream tasks scheduled from follow-on schedule check
```

8. Conclusiones

El seguimiento y avance de este proyecto se ha realizado siguiendo las precisas instrucciones que había para su realización. Los requisitos de un proyecto son la base para que el desarrollo llegue a los objetivos marcados.

A la hora de realizar un proyecto de desarrollo es fundamental llevar a cabo una rigurosa gestión de versiones. No hemos tenido problemas de versiones ya que he utilizado las imágenes empleadas por el docente en los ejemplos realizados en las clases.

La practica se ha dilatado en su desarrollo al haber introducido los diferentes operadores (PythonOperator, PostgresOperator, HttpSensor, etc) en unos grafos que incorporan agrupación de tareas y dataset.

El camino recorrido en la realización de la práctica ha hecho que el aprendizaje de Airflow haya resultado muy satisfactorio.