

### 1. Definición del DAG:

- Utiliza VS Code para crear una instancia de la clase DAG.
- Asigna un identificador único "user\_processing" al DAG.
- Establece la fecha de inicio en el 1 de enero de 2023.
- Programa el DAG para ejecutarse todos los días a medianoche utilizando un ajuste preestablecido de cron.
- Establece `catchup=False` para evitar reponer las ejecuciones perdidas.

### 2. Primera Tarea: Crear Tabla en Postgres

- Crea tu primera tarea con nombre: "create\_table"
- Utiliza una conexión "postgres" creada en la interfaz de usuario de Airflow.
- Ejecuta la siguiente consulta SQL:

```
CREATE TABLE IF NOT EXISTS users (  
nombre TEXT NOT NULL,  
apellido TEXT NOT NULL,  
país TEXT NOT NULL,  
nombre_usuario TEXT NOT NULL,  
contraseña TEXT NOT NULL,  
correo_electronico TEXT NOT NULL  
);
```

### 3. Segunda Tarea: Verificar Disponibilidad de la API

- Crea otra tarea con nombre: "api\_available"
- Crea una conexión HTTP llamada "user\_api" con el host: <https://randomuser.me/>
- Coloca esta tarea después de la tarea "create\_table".

#### 4. Tercera Tarea: Obtener Datos de la API

- Crea una tarea con nombre: "extract\_user"
- Utiliza la conexión "user\_api".
- Realiza una solicitud GET al punto final "api/".
- Filtra la respuesta cargándola con `json.loads`.
- Registra la respuesta en la interfaz de usuario de Airflow.

#### 5. Cuarta Tarea: Procesar Datos Obtenidos por extract\_user

- Crea una tarea con nombre: "process\_user"
- Llama a la función `\_process\_user` que te proporciono en este código. Debes agregarla a tu archivo entre los imports y la definición del DAG.

```
2. def _process_user(ti):
3.     user = ti.xcom_pull(task_ids="extract_user") # fetch data
         pushed by the previous task extract_user
4.     user = user['results'][0]
5.     processed_user = json_normalize({
6.         'firstname': user['name']['first'],
7.         'lastname': user['name']['last'],
8.         'country': user['location']['country'],
9.         'username': user['login']['username'],
10.         'password': user['login']['password'],
11.         'email': user['email'] })
12.     processed_user.to_csv('/tmp/processed_user.csv',
         index=None, header=False)
```

- La función `\_process\_user` obtiene datos de la tarea `extract\_user` usando XCOM.
- Utiliza `json\_normalize` para extraer y formatear los datos para almacenarlos en un archivo CSV.
- Establece la dependencia de "extract\_user" a "process\_user".

## 6. Quinta Tarea: Almacenar Datos Procesados en Postgres

- Nombre de la ultima tarea: "store\_user"
- Llama a la función `\_store\_user`.

```
def _store_user():  
    hook = PostgresHook(postgres_conn_id='postgres')  
    hook.copy_expert(  
        sql="COPY users FROM stdin WITH DELIMITER as ','",  
        filename='/tmp/processed_user.csv'  
    )
```

- Que te proporciono aquí y que debes pegar en tu .py después de los imports y antes de la anterior función
- La función utiliza `PostgresHook` para interactuar con Postgres. Utiliza `copy\_expert` para exportar datos de un archivo CSV a una tabla de Postgres.
- Establece la dependencia de "process\_user" a "store\_user".

## 7. Comprueba que tu flujo se ejecuta correctamente.