

## Ejercicio 7.

### Creación de Tabla en Hive

1.

```
CREATE EXTERNAL TABLE airport_trips (  
  tpep_pickup_datetime date,  
  airport_fee string,  
  pyment_type int,  
  tolls_amount float,  
  total_amount float  
)  
row format delimited fields terminated by ';' location '/user/hive/warehouse'
```

2. Esquema de la tabla airport\_trips

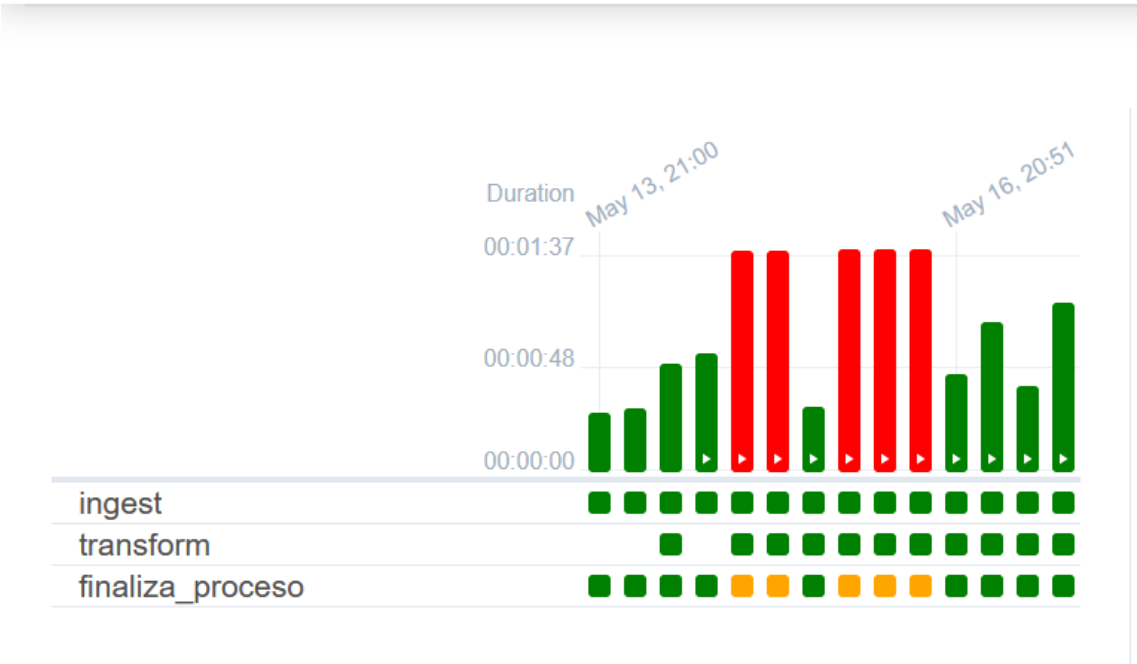
col_name	data_type	comment	
# col_name	data_type	comment	
	[NULL]	[NULL]	
tpep_pickup_datetime	date		
airport_fee	string		
pyment_type	int		
tolls_amount	float		
total_amount	float		
	[NULL]	[NULL]	
# Detailed Table Information	[NULL]	[NULL]	

3. Scripts para automatizar descarga y transformación (Python)

```
hadoop@0126ee263ae9:~/scripts$ ls  
'Scripts_Jupyter_Notebook' derby.log ingest_7.sh spark-warehouse transformation.py  
Transform_airport.py ingest.sh pyspark_jupyter.sh start-services.sh  
hadoop@0126ee263ae9:~/scripts$
```

4. Airflow (Automatización de pipeline)

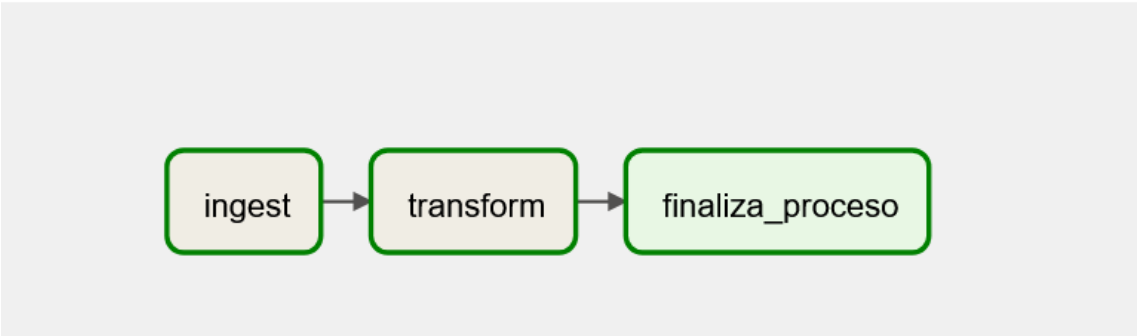
Vista Grid:



Vista de Resumen:

DAG	
Ingesta_transformacion	
DAG Details	
DAG Runs Summary	
Total Runs Displayed	14
<div></div> Total success	9
<div></div> Total failed	5
First Run Start	2025-05-16, 19:57:07 -03
Last Run Start	2025-05-17, 10:16:25 -03
Max Run Duration	00:01:37
Mean Run Duration	00:01:04
Min Run Duration	00:00:26
DAG Summary	

Vista de Grafo:



5. Tabla de datos en Hive:

tpep_pickup_datetime	airport_fee	pyment_type	tolls_amount	total_amount	
2020-12-31	[NULL]	2	0	11,8	
2020-12-31	[NULL]	2	0	4,3	
2020-12-31	[NULL]	2	0	17,3	
2020-12-31	[NULL]	2	0	21,8	
2020-12-31	[NULL]	2	6,12	33,92	
2020-12-31	[NULL]	2	0	8,3	
2020-12-31	[NULL]	2	0	12,3	
2020-12-31	[NULL]	2	0	9,3	