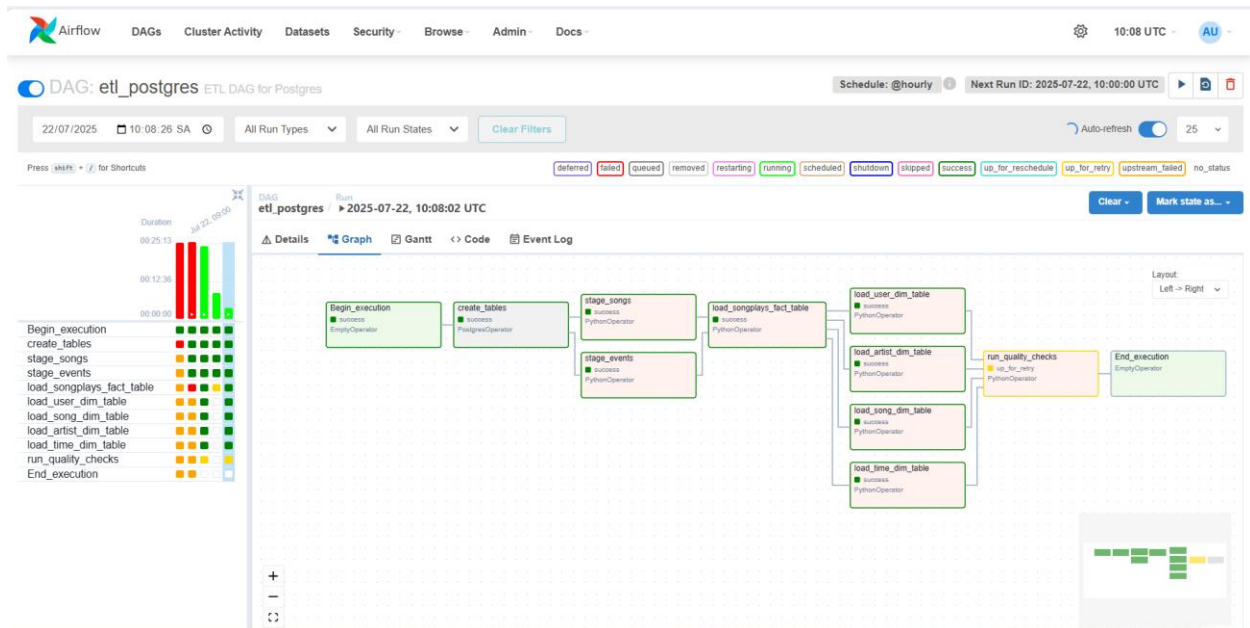


## Hình ảnh các DAG trong airflow



## DAG configuration

```
from dags.sql_queries import SqlQueries
default_args = {
    'owner': 'airflow',
    'depends_on_past': False,
    'start_date': days_ago(1),
    'email_on_failure': False,
    'email_on_retry': False,
    'retries': 5,
    'retry_delay': timedelta(minutes=5),
    'catchup': False,
}
dag = DAG(
    'etl_postgres',
    default_args=default_args,
    description='ETL DAG for Postgres',
    schedule_interval='@hourly',
    catchup=False,
    tags=['example', 'etl']
)
```

## Log của task create\_table

DAG

Run

Task

etl\_postgres ▶ 2025-07-22, 10:18:42 UTC / create\_tables

Clear task Mark state as... Filter DAG by task

Details Graph Gantt Code Event Log Logs XCom Task Duration

All Levels All File Sources Wrap Download See More

```
lastname varchar(256),
length numeric(18,0),
"level" varchar(256),
location varchar(256),
"method" varchar(256),
page varchar(256),
registration numeric(18,0),
sessionid int4,
song varchar(256),
status int4,
ts int8,
useragent varchar(256),
userid int4
);
CREATE TABLE IF NOT EXISTS public.staging_songs (
  num_songs int4,
  artist_id varchar(256),
  artist_name varchar(256),
  artist_latitude numeric(18,0),
  artist_longitude numeric(18,0),
  artist_location varchar(256),
  song_id varchar(256),
  title varchar(256),
  duration numeric(18,0),
  "year" int4
);
CREATE TABLE IF NOT EXISTS public."time" (
  start_time timestamp NOT NULL,
  "hour" int4,
  "day" int4,
  week int4,
```

## Log của stage\_songs :

DAG

Run

Task

etl\_postgres ▶ 2025-07-22, 10:18:42 UTC / stage\_songs

Clear task Mark state as... Filter DAG by task

Details Graph Gantt Code Event Log Logs XCom Task Duration

All Levels All File Sources Wrap Download See More

```
0c6a6447a272
*** Found local files:
***   * /opt/airflow/logs/dag_id=etl_postgres/run_id>manual_2025-07-22T10:18:42.652724+00:00/task_id=stage_songs/attempt=1.log
[2025-07-22, 10:18:47 UTC] {local_task_job_runner.py:123} ▼ Pre task execution logs
[2025-07-22, 10:18:47 UTC] {taskinstance.py:2613} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: etl_postgres.stage_songs manual_2025-07-22T10:18:42.652724+00:00 [queued]>
[2025-07-22, 10:18:47 UTC] {taskinstance.py:2613} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: etl_postgres.stage_songs manual_2025-07-22T10:18:42.652724+00:00 [queued]>
[2025-07-22, 10:18:47 UTC] {taskinstance.py:2866} INFO - Starting attempt 1 of 6
[2025-07-22, 10:18:47 UTC] {taskinstance.py:2889} INFO - Executing <Task(PythonOperator): stage_songs> on 2025-07-22 10:18:42.652724+00:00
[2025-07-22, 10:18:47 UTC] {standard_task_runner.py:104} INFO - Running: ['***', 'tasks', 'run', 'etl_postgres', 'stage_songs', 'manual_2025-07-22T10:18:42.652724+00:00', '--job-id', '59', '--raw', '--subdir', 'DAGS_FQ
[2025-07-22, 10:18:47 UTC] {logging_mixin.py:190} WARNING - /home/***/.local/lib/python3.12/site-packages/***/task/task_runner/standard_task_runner.py:70 DeprecationWarning: This process (pid=1547) is multi-threaded, us
[2025-07-22, 10:18:47 UTC] {standard_task_runner.py:72} INFO - Started process 1548 to run task
[2025-07-22, 10:18:47 UTC] {standard_task_runner.py:105} INFO - Job 59: Subtask stage_songs
[2025-07-22, 10:18:47 UTC] {task_command.py:467} INFO - Running <TaskInstance: etl_postgres.stage_songs manual_2025-07-22T10:18:42.652724+00:00 [running]> on host 0c6a6447a272
[2025-07-22, 10:18:47 UTC] {taskinstance.py:3132} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER=*** AIRFLOW_CTX_DAG_ID='etl_postgres' AIRFLOW_CTX_TASK_ID='stage_songs' AIRFLOW_CTX_EXECUTION_DATE='2025-07-22T10:18:4
[2025-07-22, 10:18:47 UTC] {taskinstance.py:731} ▲▲▲ Log group end
[2025-07-22, 10:18:48 UTC] {etl_dag.py:134} INFO - Staging songs data
[2025-07-22, 10:18:48 UTC] {etl_dag.py:156} INFO - Songs data staged successfully
[2025-07-22, 10:18:48 UTC] {python.py:240} INFO - Done. Returned value was: None
[2025-07-22, 10:18:48 UTC] {taskinstance.py:340} ▶ Post task execution logs
```

## Log của stage\_events :

DAG etl\_postgres / Run 2025-07-22, 10:18:42 UTC / Task stage\_events

Clear task Mark state as... Filter DAG by task

Details Graph Gantt Code Event Log Logs XCom Task Duration

All Levels All File Sources Wrap Download See More

```
0c6a6447a272
*** Found local files:
*** * /opt/airflow/logs/dag_id=etl_postgres/run_id>manual_2025-07-22T10:18:42.652724+00:00/task_id=stage_events/attempt=1.log
[2025-07-22, 10:18:47 UTC] [local_task_job_runner.py:123] Pre task execution logs
[2025-07-22, 10:18:47 UTC] [taskinstance.py:2613] INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: etl_postgres.stage_events manual_2025-07-22T10:18:42.652724+00:00 [queued]>
[2025-07-22, 10:18:47 UTC] [taskinstance.py:2613] INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: etl_postgres.stage_events manual_2025-07-22T10:18:42.652724+00:00 [queued]>
[2025-07-22, 10:18:47 UTC] [taskinstance.py:2866] INFO - Starting attempt 1 of 6
[2025-07-22, 10:18:47 UTC] [taskinstance.py:2889] INFO - Executing <Task(PythonOperator): stage_events> on 2025-07-22 10:18:42.652724+00:00
[2025-07-22, 10:18:47 UTC] [standard_task_runner.py:104] INFO - Running: ['***', 'tasks', 'run', 'etl_postgres', 'stage_events', 'manual_2025-07-22T10:18:42.652724+00:00', '--job-id', '58', '--raw', '--subdir', 'DAGS_F
[2025-07-22, 10:18:47 UTC] [logging_mixin.py:190] WARNING - /home/***/.local/lib/python3.12/site-packages/***/task/task_runner/standard_task_runner.py:70 DeprecationWarning: This process (pid=1546) is multi-threaded, us
[2025-07-22, 10:18:47 UTC] [standard_task_runner.py:105] INFO - Job 58: Subtask stage_events
[2025-07-22, 10:18:47 UTC] [standard_task_runner.py:72] INFO - Started process 1549 to run task
[2025-07-22, 10:18:47 UTC] [task_command.py:467] INFO - Running <TaskInstance: etl_postgres.stage_events manual_2025-07-22T10:18:42.652724+00:00 [running]> on host 0c6a6447a272
[2025-07-22, 10:18:47 UTC] [taskinstance.py:3132] INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER=*** AIRFLOW_CTX_DAG_ID=etl_postgres AIRFLOW_CTX_TASK_ID=stage_events AIRFLOW_CTX_EXECUTION_DATE=2025-07-22T10:18:42.652724+00:00
[2025-07-22, 10:18:47 UTC] [taskinstance.py:731] Log group end
[2025-07-22, 10:18:48 UTC] [etl_dag.py:166] INFO - Staging events data
[2025-07-22, 10:18:48 UTC] [etl_dag.py:191] INFO - loaded event files successfully
[2025-07-22, 10:18:48 UTC] [python.py:240] INFO - Done. Returned value was: None
[2025-07-22, 10:18:48 UTC] [taskinstance.py:340] Post task execution logs
```

## Log của load\_songplays\_fact\_table

DAG etl\_postgres / Run 2025-07-22, 10:18:42 UTC / Task load\_songplays\_fact\_table

Clear task Mark state as... Filter DAG by task

Details Graph Gantt Code Event Log Logs XCom Task Duration

All Levels All File Sources Wrap Download See More

```
0c6a6447a272
*** Found local files:
*** * /opt/airflow/logs/dag_id=etl_postgres/run_id>manual_2025-07-22T10:18:42.652724+00:00/task_id=load_songplays_fact_table/attempt=1.log
[2025-07-22, 10:18:49 UTC] [local_task_job_runner.py:123] Pre task execution logs
[2025-07-22, 10:18:49 UTC] [taskinstance.py:2613] INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: etl_postgres.load_songplays_fact_table manual_2025-07-22T10:18:42.652724+00:00 [queued]>
[2025-07-22, 10:18:49 UTC] [taskinstance.py:2613] INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: etl_postgres.load_songplays_fact_table manual_2025-07-22T10:18:42.652724+00:00 [queued]>
[2025-07-22, 10:18:49 UTC] [taskinstance.py:2866] INFO - Starting attempt 1 of 6
[2025-07-22, 10:18:49 UTC] [taskinstance.py:2889] INFO - Executing <Task(PythonOperator): load_songplays_fact_table> on 2025-07-22 10:18:42.652724+00:00
[2025-07-22, 10:18:49 UTC] [standard_task_runner.py:104] INFO - Running: ['***', 'tasks', 'run', 'etl_postgres', 'load_songplays_fact_table', 'manual_2025-07-22T10:18:42.652724+00:00', '--job-id', '60', '--raw', '--sub
[2025-07-22, 10:18:49 UTC] [standard_task_runner.py:72] INFO - Started process 1553 to run task
[2025-07-22, 10:18:49 UTC] [standard_task_runner.py:105] INFO - Job 60: Subtask load_songplays_fact_table
[2025-07-22, 10:18:49 UTC] [task_command.py:467] INFO - Running <TaskInstance: etl_postgres.load_songplays_fact_table manual_2025-07-22T10:18:42.652724+00:00 [running]> on host 0c6a6447a272
[2025-07-22, 10:18:49 UTC] [taskinstance.py:3132] INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER=*** AIRFLOW_CTX_DAG_ID=etl_postgres AIRFLOW_CTX_TASK_ID=load_songplays_fact_table AIRFLOW_CTX_EXECUTION_DATE=2025-07-22T10:18:42.652724+00:00
[2025-07-22, 10:18:49 UTC] [taskinstance.py:731] Log group end
[2025-07-22, 10:18:49 UTC] [etl_dag.py:49] INFO - Loading songplays fact table
[2025-07-22, 10:18:49 UTC] [etl_dag.py:52] INFO - Songplays fact table loaded successfully
[2025-07-22, 10:18:49 UTC] [python.py:240] INFO - Done. Returned value was: None
[2025-07-22, 10:18:49 UTC] [taskinstance.py:340] Post task execution logs
```

## Log của load\_user\_dim

etl\_postgres / ▶ 2025-07-22, 10:18:42 UTC / load\_user\_dim\_table

Clear task Mark state as... Filter DAG by task

Details Graph Gantt Code Event Log Logs XCom Task Duration

All Levels All File Sources Wrap Download See More

```
0c6a6447a272
*** Found local files:
***   /opt/airflow/logs/dag_id=etl_postgres/run_id>manual_2025-07-22T10:18:42.652724+00:00/task_id=load_user_dim_table/attempt=1.log
[2025-07-22, 10:18:50 UTC] {local_task_job_runner.py:123} ▼ Pre task execution logs
[2025-07-22, 10:18:50 UTC] {taskinstance.py:2613} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: etl_postgres.load_user_dim_table manual__2025-07-22T10:18:42.652724+00:00 [queued]>
[2025-07-22, 10:18:50 UTC] {taskinstance.py:2613} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: etl_postgres.load_user_dim_table manual__2025-07-22T10:18:42.652724+00:00 [queued]>
[2025-07-22, 10:18:50 UTC] {taskinstance.py:2866} INFO - Starting attempt 1 of 6
[2025-07-22, 10:18:50 UTC] {taskinstance.py:2889} INFO - Executing <task(PythonOperator): load_user_dim_table on 2025-07-22 10:18:42.652724+00:00
[2025-07-22, 10:18:50 UTC] {standard_task_runner.py:194} INFO - Running: ['***', 'tasks', 'run', 'etl_postgres', 'load_user_dim_table', 'manual_2025-07-22T10:18:42.652724+00:00', '--job-id', '64', '--raw', '--subdir',
[2025-07-22, 10:18:50 UTC] {logging_mixin.py:190} WARNING - /home/***/.local/lib/python3.12/site-packages/***/task/task_runner/standard_task_runner.py:70 DeprecationWarning: This process (pid=1555) is multi-threaded, us
[2025-07-22, 10:18:50 UTC] {standard_task_runner.py:105} INFO - Job 64: Subtask load_user_dim_table
[2025-07-22, 10:18:50 UTC] {standard_task_runner.py:72} INFO - Started process 1559 to run task
[2025-07-22, 10:18:50 UTC] {task_command.py:467} INFO - Running <TaskInstance: etl_postgres.load_user_dim_table manual_2025-07-22T10:18:42.652724+00:00 [running]> on host 0c6a6447a272
[2025-07-22, 10:18:51 UTC] {taskinstance.py:3132} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER='***' AIRFLOW_CTX_DAG_ID='etl_postgres' AIRFLOW_CTX_TASK_ID='load_user_dim_table' AIRFLOW_CTX_EXECUTION_DATE='2025-07-22
[2025-07-22, 10:18:51 UTC] {taskinstance.py:731} ▲▲▲ Log group end
[2025-07-22, 10:18:51 UTC] {etl_dag.py:63} INFO - Loading user dimension table
[2025-07-22, 10:18:51 UTC] {etl_dag.py:66} INFO - User dimension table loaded successfully
[2025-07-22, 10:18:51 UTC] {python.py:240} INFO - Done. Returned value was: None
[2025-07-22, 10:18:51 UTC] {taskinstance.py:340} ▶ Post task execution logs
```

## Log của các dim khác :

etl\_postgres / ▶ 2025-07-22, 10:18:42 UTC / load\_song\_dim\_table

Clear task Mark state as... Filter DAG by task

Details Graph Gantt Code Event Log Logs XCom Task Duration

All Levels All File Sources Wrap Download See More

```
0c6a6447a272
*** Found local files:
***   /opt/airflow/logs/dag_id=etl_postgres/run_id>manual_2025-07-22T10:18:42.652724+00:00/task_id=load_song_dim_table/attempt=1.log
[2025-07-22, 10:18:50 UTC] {local_task_job_runner.py:123} ▶ Pre task execution logs
[2025-07-22, 10:18:51 UTC] {etl_dag.py:77} INFO - Loading song dimension table
[2025-07-22, 10:18:51 UTC] {etl_dag.py:80} INFO - Song dimension table loaded successfully
[2025-07-22, 10:18:51 UTC] {python.py:240} INFO - Done. Returned value was: None
[2025-07-22, 10:18:51 UTC] {taskinstance.py:340} ▶ Post task execution logs
```

etl\_postgres / ▶ 2025-07-22, 10:18:42 UTC / load\_artist\_dim\_table

Clear task Mark state as... Filter DAG by task

Details Graph Gantt Code Event Log Logs XCom Task Duration

All Levels All File Sources Wrap Download See More

```
0c6a6447a272
*** Found local files:
***   /opt/airflow/logs/dag_id=etl_postgres/run_id>manual_2025-07-22T10:18:42.652724+00:00/task_id=load_artist_dim_table/attempt=1.log
[2025-07-22, 10:18:50 UTC] {local_task_job_runner.py:123} ▶ Pre task execution logs
[2025-07-22, 10:18:51 UTC] {etl_dag.py:91} INFO - Loading artist dimension table
[2025-07-22, 10:18:51 UTC] {etl_dag.py:94} INFO - Artist dimension table loaded successfully
[2025-07-22, 10:18:51 UTC] {python.py:240} INFO - Done. Returned value was: None
[2025-07-22, 10:18:51 UTC] {taskinstance.py:340} ▶ Post task execution logs
```

All LevelsAll File SourcesWrapDownloadSee More

0c6a6447a272  
\*\*\* Found local files:  
\*\*\* /opt/airflow/logs/dag\_id=etl\_postgres/run\_id>manual\_\_2025-07-22T10:18:42.652724+00:00/task\_id=load\_time\_dim\_table/attempt=1.log  
[2025-07-22, 10:18:50 UTC] (local\_task\_job\_runner.py:123) ▶ Pre task execution logs  
[2025-07-22, 10:18:51 UTC] (etl\_dag.py:105) INFO - Loading time dimension table  
[2025-07-22, 10:18:51 UTC] (etl\_dag.py:108) INFO - Time dimension table loaded successfully  
[2025-07-22, 10:18:51 UTC] (python.py:240) INFO - Done. Returned value was: None  
[2025-07-22, 10:18:51 UTC] (taskinstance.py:340) ▶ Post task execution logs