

PostgreSQL Data Mart with Airflow & dbt

Overview

A self-contained ELT pipeline that ingests raw CSV data into a PostgreSQL data warehouse using Apache Airflow, then models a dimensional data mart with dbt. This project demonstrates data engineering skills in orchestration, transformation, and testing, deployable locally with a single command.

Tech Stack

- **PostgreSQL 13:** Data warehouse and Airflow metastore
- **Redis (latest):** Celery broker for Airflow
- **Apache Airflow 2.5 (CeleryExecutor):** Workflow orchestration
- **dbt 1.10:** Data modeling and testing
- **Docker Compose 3.7:** Single-command local deployment

Workflow

Stage	Tool	Result
Extract & Load	Airflow	Raw ingest DAG loads sales.csv into raw_sales table
Transform	dbt	Builds staging model (stg_raw_sales) and dimensional models (dim_products, dim_customers, fact_sales)
Test	dbt	Runs tests (not null, unique, relationships), all passing

Quick Start

1. Clone the repository:

```
git clone https://github.com/Kentex99/postgres-data-mart.git
cd postgres-data-mart
```

2. Start the stack:

```
docker compose up -d
```

3. Access Airflow UI at <http://localhost:8080> (user: airflow, password: airflow)

4. Trigger the DAG - `ingest_raw_sales_v2` to load raw data. Build the mart with dbt :

Project Structure

5.
 - postgres-data-mart/
 - airflow/ ← DAGs & helpers
 - dbt/sales_dbt/ ← dbt project with models & tests

- data/raw/sales.csv ← sample source file
- docker-compose.yml ← full stack definition
- README.md : project documentation

Extending the Project

- (a) Add another raw CSV to data/raw/
- (b) Declare a new task in the Airflow DAG
- (c) Create a staging model in models/staging/
- (d) Reference it in marts, add tests, and rerun dbt build

Author

Tasie Kenneth – Data Engineering Enthusiast
github.com/Kentex99