2. Workflow Orchestration (Apache Airflow)

As part of the data analyst challenge, an **Apache Airflow DAG** was created to automate the data pipeline process.

**Workflow Steps**

Then add a bullet list under it:

* Ingest raw CSV data into PostgreSQL.
* Trigger DBT models to transform the data.
* Validate successful execution of DBT transformations.
* Schedule the DAG to run once every 24 hours (@daily).

An Airflow DAG was created with the following schedule and tasks:

* Runs **daily** (@daily)
* Steps:
  1. **Ingests CSV** data into PostgreSQL.
  2. **Triggers DBT** transformations.
  3. **Validates completion** of DBT models.

**Airflow DAG Code: itsm\_pipeline.py**

from airflow import DAG

from airflow.operators.bash import BashOperator

from airflow.operators.python import PythonOperator

from datetime import datetime, timedelta

default\_args = {

'owner': 'airflow',

'retries': 1,

'retry\_delay': timedelta(minutes=5)

}

def ingest\_data():

import pandas as pd

import psycopg2

df = pd.read\_csv('/path/to/ticket\_dump.csv')

conn = psycopg2.connect(database="itsm", user="postgres", password="your\_password", host="localhost", port="5432")

cursor = conn.cursor()

for \_, row in df.iterrows():

cursor.execute("""

INSERT INTO raw\_tickets (...columns...)

VALUES (...row values...)

""")

conn.commit()

cursor.close()

conn.close()

with DAG(

dag\_id='itsm\_pipeline',

default\_args=default\_args,

start\_date=datetime(2025, 1, 1),

schedule\_interval='@daily',

catchup=False

) as dag:

task\_ingest = PythonOperator(

task\_id='ingest\_csv\_to\_postgres',

python\_callable=ingest\_data

)

task\_dbt\_run = BashOperator(

task\_id='run\_dbt\_models',

bash\_command='cd /path/to/dbt\_project && dbt run'

)

task\_validate = BashOperator(

task\_id='validate\_models',

bash\_command='echo "DBT models completed successfully."'

)

task\_ingest >> task\_dbt\_run >> task\_validate