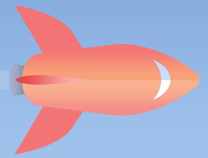


FINAL PROJECT

DIBIMBING -Modeling And Optimization Techniques In Data Warehousing Bootcamp 2023

KELOMPOK 5

Headline

Introduction

Analysis and Visualization

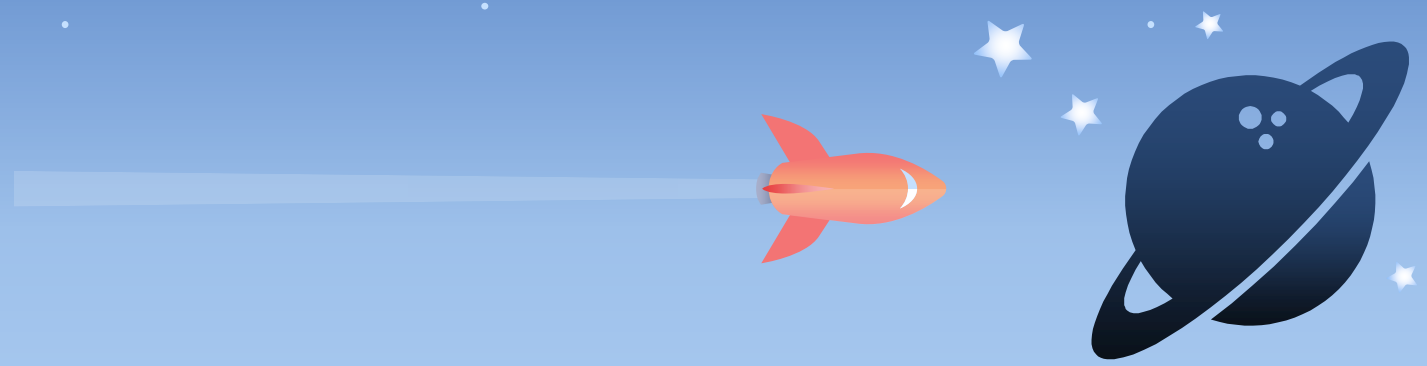
Project Description

Technical improvisation

Introduction

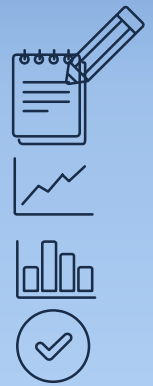
Task

- [ ] ETL/ELT Job Creation using Airflow
- [ ] Data Modeling in Postgres
- [ ] Dashboard Creation with Data Visualization
- [ ] Craft a Presentation Based on Your Work

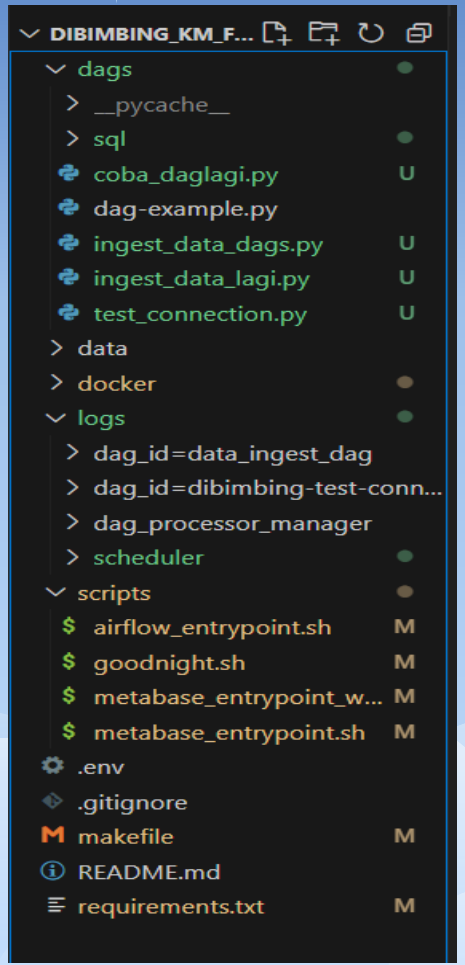


Project Description

ETL/ELT Job Creation using Airflow



Struktur Folder



DAG

```python
from airflow import DAG
from airflow.operators.python_operator import PythonOperator
from airflow.operators.dummy_operator import DummyOperator

from datetime import datetime

# DAG definition
default_args = {
    'owner': 'kel5',
    'depends_on_past': False,
    'start_date': datetime(2023, 11, 30),
    'catchup': False,
}

dag = DAG(
    'ingest_data_dags',
    default_args=default_args,
    description='DAG for ingesting data files into PostgreSQL',
    schedule_interval='@once',
    tags=['ingesting']
)

data_folder_path = 'data/'

def test_conn():
    print("Testing connection...")

    return True
with dag:
# Task to test the connection
test_conn_task = PythonOperator(
    task_id='test_connection',
    python_callable=test_conn,
    dag=dag,
)
```

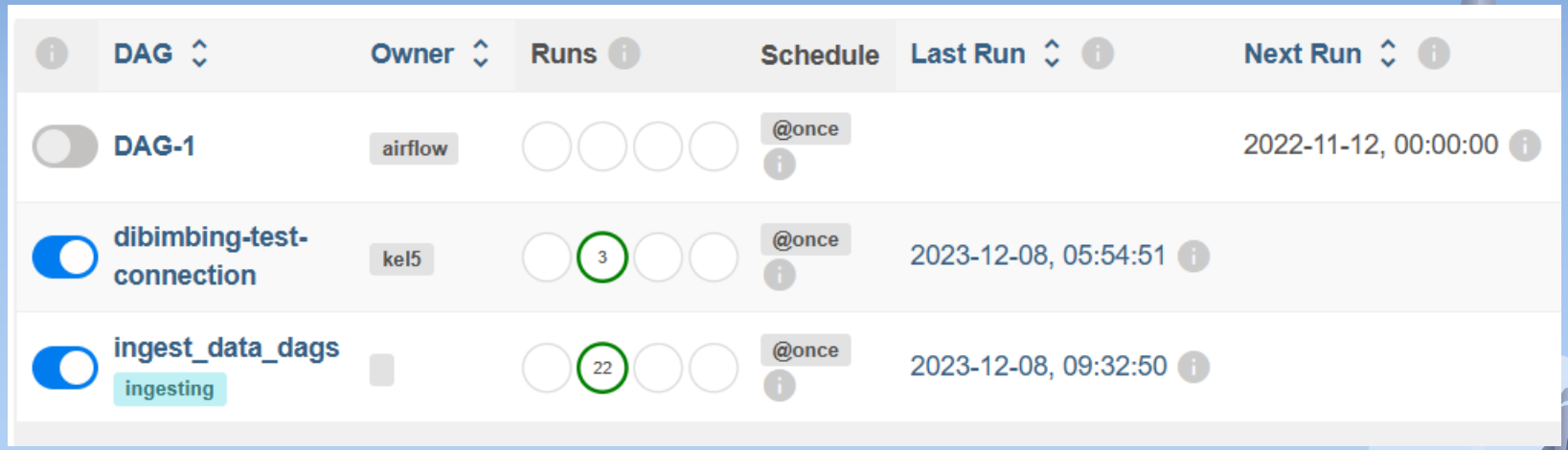
```python
data_folder_path = 'data/'

def test_conn():
    print("Testing connection...")

    return True
with dag:
# Task to test the connection
test_conn_task = PythonOperator(
    task_id='test_connection',
    python_callable=test_conn,
    dag=dag,
)

# CSV task
ingest_csv_task = PythonOperator(
    task_id='ingest_csv',
    python_callable=ingest_csv_files,
    op_kwargs={'folder_path': data_folder_path, 'table_name': 'customers'},
    dag=dag,
)

# JSON tasks
ingest_json_files_login_attempts_task = PythonOperator(
    task_id='ingest_json_login_attempts',
    python_callable=ingest_json_files_login_attempts,
    op_kwargs={'folder_path': data_folder_path, 'table_name': 'login_attempt_history'},
    dag=dag,
)

ingest_json_files_coupons_task = PythonOperator(
    task_id='ingest_json_coupons',
    python_callable=ingest_json_files_coupons,
    op_kwargs={'folder_path': data_folder_path, 'table_name': 'coupons'},
    dag=dag,
)
```

DAG

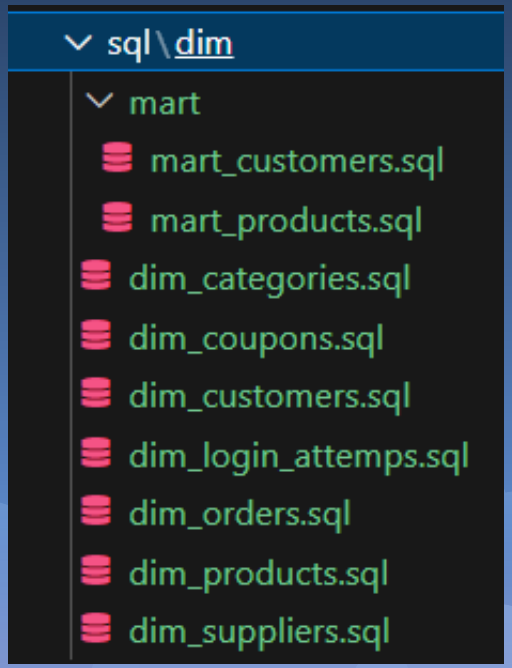
```
# XLS tasks
ingest_xls_files_supplier_task = PythonOperator(
    task_id='ingest_xls_supplier',
    python_callable=ingest_xls_files_supplier,
    op_kwargs={'folder_path': data_folder_path, 'table_name': 'suppliers'},
    dag=dag,
)

ingest_xls_files_product_task = PythonOperator(
    task_id='ingest_xls_product',
    python_callable=ingest_xls_files_product,
    op_kwargs={'folder_path': data_folder_path, 'table_name': 'product'},
    dag=dag,
)

ingest_xls_files_product_category_task = PythonOperator(
    task_id='ingest_xls_product_category',
    python_callable=ingest_xls_files_product_category,
    op_kwargs={'folder_path': data_folder_path, 'table_name': 'product_category'},
    dag=dag,
)

# PARQUET task
ingest_parquet_task = PythonOperator(
    task_id='ingest_parquet',
    python_callable=ingest_parquet_file,
    op_kwargs={'file_path': f'{data_folder_path}/order.parquet', 'table_name': 'orders'},
    dag=dag,
)
```

```
# AVRO task
ingest_avro_task = PythonOperator(
    task_id='ingest_avro',
    python_callable=ingest_avro_file,
    op_kwargs={'file_path': f'{data_folder_path}/order_item.avro', 'table_name': 'order_items'},
    dag=dag,
)

# End of DAG
end = DummyOperator(
    task_id='end',
    trigger_rule='none_failed'
)

# Setting dependencies
test_conn_task >> ingest_csv_task >> end
test_conn_task >> ingest_json_files_login_attempts_task >> end
test_conn_task >> ingest_json_files_coupons_task >> end
test_conn_task >> ingest_xls_files_supplier_task >> end
test_conn_task >> ingest_xls_files_product_task >> end
test_conn_task >> ingest_xls_files_product_category_task >> end
test_conn_task >> ingest_parquet_task >> end
test_conn_task >> ingest_avro_task >> end
```

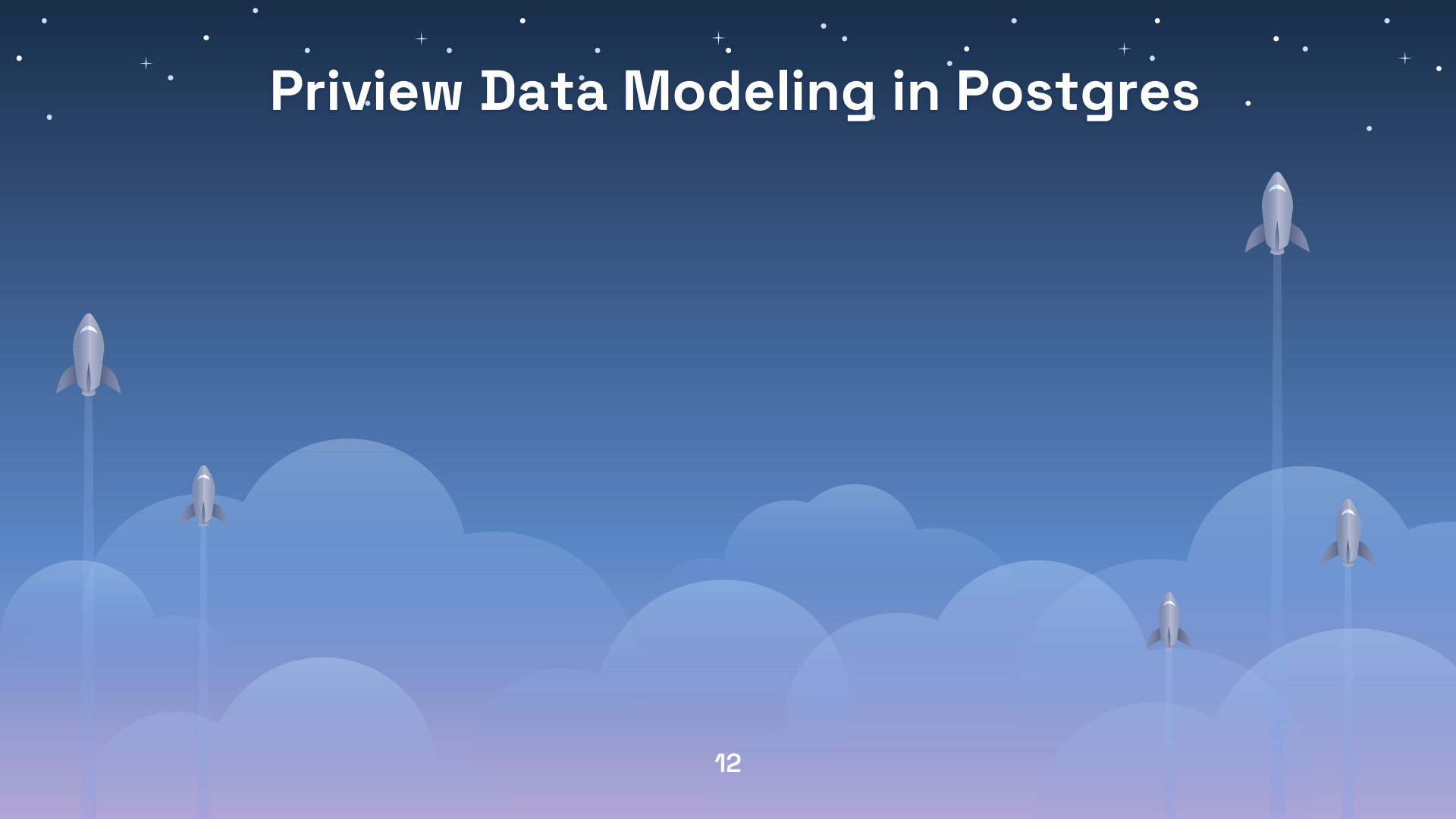
DAG

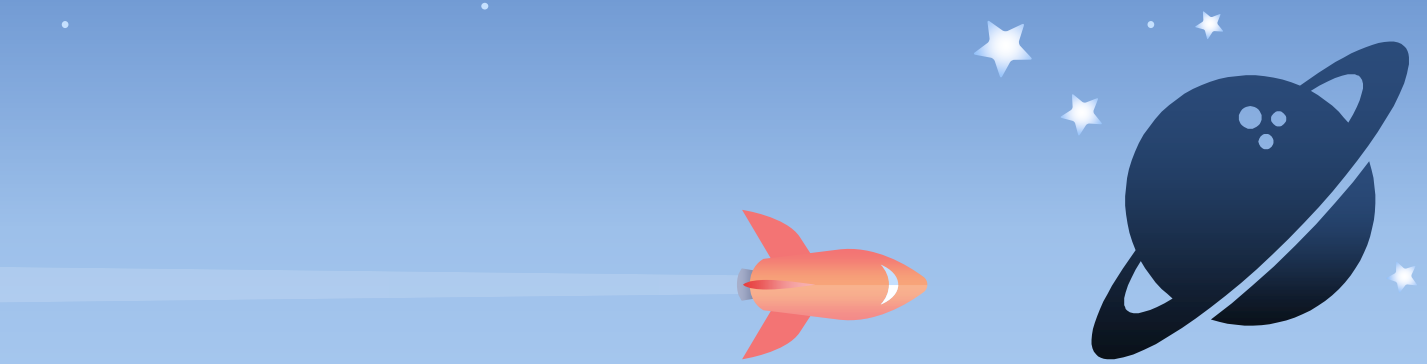

DAG
Owner
Runs
Schedule
Last Run
Next Run
DAG-1
airflow
@once
2022-11-12, 00:00:00
dibimbing-test-connection
kel5
3
@once
2023-12-08, 05:54:51
ingest_data_dags
ingesting
22
@once
2023-12-08, 09:32:50

Data Modeling in Postgres



Priview Data Modeling in Postgres

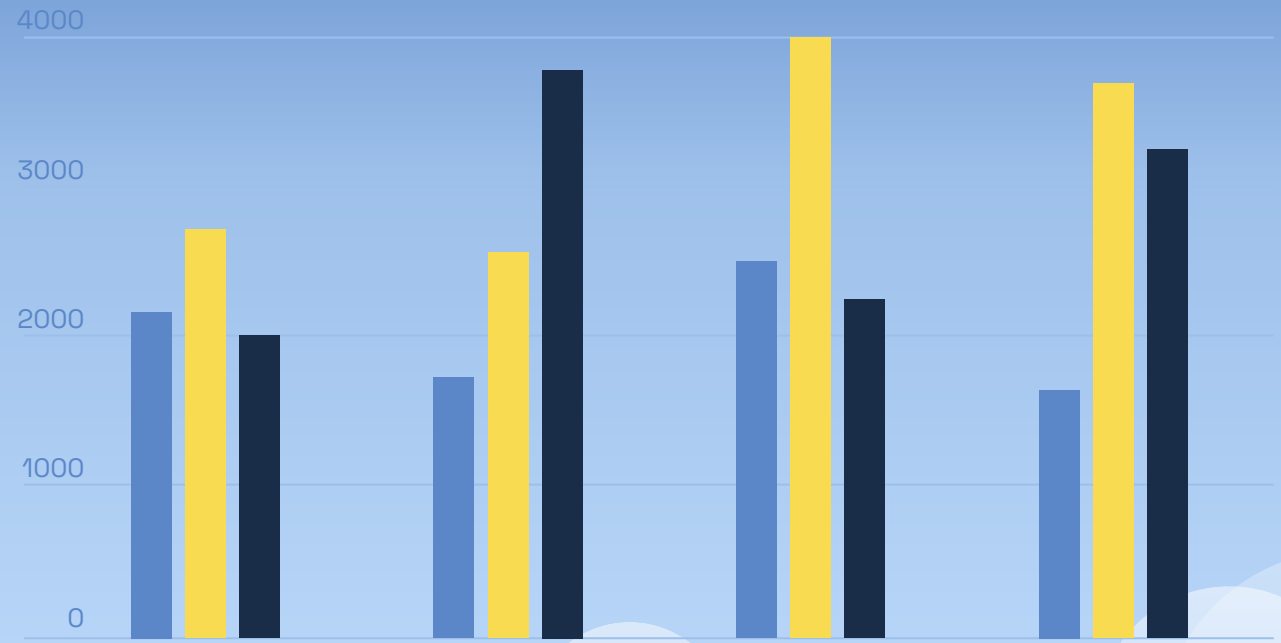




Analysis and Visualization

Dashboard Creation with Data Visualization


Technical improvisation



You can insert graphs from Excel or Google Sheets

Slides Carnival

Free templates for all your presentation needs

For PowerPoint and Google Slides

100% free for personal or commercial use

Ready to use, professional and customizable

Blow your audience away with attractive visuals