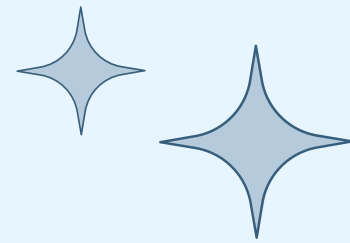


FINAL PROJECT





Anggota Kelompok:

Fajri Yanti

Feni Ayu Putri

Grisella Estefania R

Okta Savira

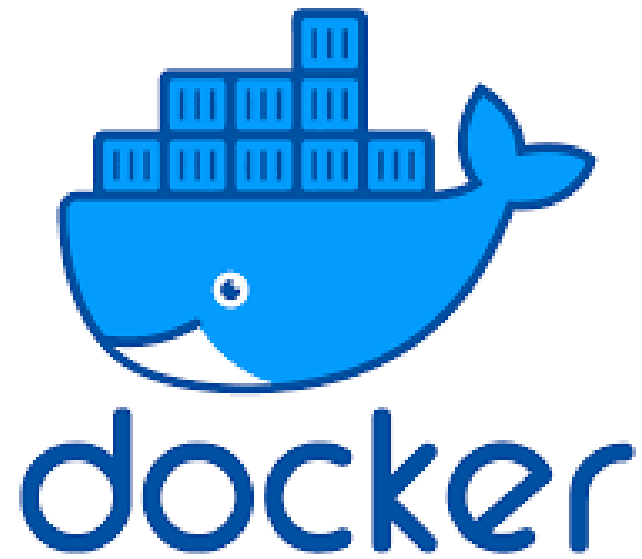


LATAR BELAKANG



Latar belakang project ini adalah meningkatkan pemahaman terhadap data pelanggan dengan membangun data warehouse. Dengan adanya data warehouse data dapat diintegrasikan dan disimpan dari berbagai sumber secara efisien, serta memungkinkan analisis menyeluruh. Kemudian data di visualisasi agar dalam mengidentifikasi tren pelanggan lebih mudah dibaca yang nantinya dapat digunakan untuk meningkatkan layanan pelanggan dan pengambilan keputusan strategis.

TOOLS



TOOLS

Postgresql

- Relational Database Management System (RDBMS): Cocok untuk data yang memiliki hubungan yang kompleks antar entitas, seperti data penjualan, pelanggan, dan produk dalam industri ritel.
- Support untuk Kimball Data Modeling: Mendukung pengimplementasian struktur data berbasis dimensi (dimensional modeling) sesuai prinsip Kimball, memudahkan pemodelan data untuk analisis lebih lanjut.

Apache Airflow

- Workflow Automation: Memungkinkan jadwal dan eksekusi otomatis untuk proses ETL/ELT.
- Skalabilitas dan Pengelolaan Tugas: Memudahkan manajemen tugas-tugas yang kompleks dalam aliran data.

TOOLS

GIT

- Kolaborasi Tim: Membantu tim untuk bekerja bersama, melacak perubahan, dan mengelola kode dengan efisien.
- Riwayat Perubahan: Memungkinkan untuk melihat versi-versi sebelumnya dan melacak evolusi proyek.

DOCKER

- Isolasi Lingkungan: Memastikan konsistensi lingkungan pengembangan, pengujian, dan produksi.
- Portabilitas Aplikasi: Memudahkan deploy aplikasi ke lingkungan apapun tanpa terlalu banyak perubahan.

TOOLS

METABASE

- - Kemudahan Penggunaan: Antarmuka yang ramah pengguna dan intuitif untuk membuat visualisasi data tanpa memerlukan pengetahuan teknis yang mendalam.
- Kustomisasi dan Ekstensibilitas: Memungkinkan penyesuaian visualisasi sesuai kebutuhan bisnis serta integrasi dengan berbagai sumber data.

Setiap tools yang dipilih memiliki keunggulan dalam konteks tertentu yang sesuai dengan tugas yang perlu dijalankan dalam proyek. Kombinasi alat-alat tersebut diharapkan dapat memberikan kerangka kerja yang kuat untuk membangun infrastruktur data yang efektif dan skalabel untuk perusahaan ritel online.

LIBRARY

Pandas

- **Kematangan dan Popularitas:** Pandas adalah salah satu library yang paling populer. Kematangannya telah teruji baik dalam pemrosesan data di lingkungan Data Science.
- **Fleksibilitas dan Fungsionalitas yang Luas:** Pandas memiliki beragam fungsi dan metode yang kuat untuk melakukan manipulasi data, seperti filtering, agregasi, dan transformasi data, membuatnya sangat fleksibel untuk digunakan dalam proyek-proyek data yang beragam.
- **Komunitas yang Besar:** Kehadiran komunitas yang besar menyediakan banyaknya sumber daya yang dapat digunakan untuk mempelajari dan memecahkan masalah yang kompleks.

Polars

- **Performa yang Lebih Cepat:** Polars adalah library yang dirancang dengan fokus pada performa yang lebih tinggi dalam pemrosesan data. Dibangun dengan Rust, Polars menawarkan kecepatan yang lebih tinggi dalam beberapa operasi pemrosesan data dibandingkan dengan Pandas, terutama untuk dataset yang besar.
- **Skalabilitas yang Baik:** Polars menawarkan kemampuan untuk menangani dataset yang besar dengan cepat dan efisien, menjadikannya pilihan yang bagus untuk proyek-proyek dengan kebutuhan pemrosesan data dalam skala besar.

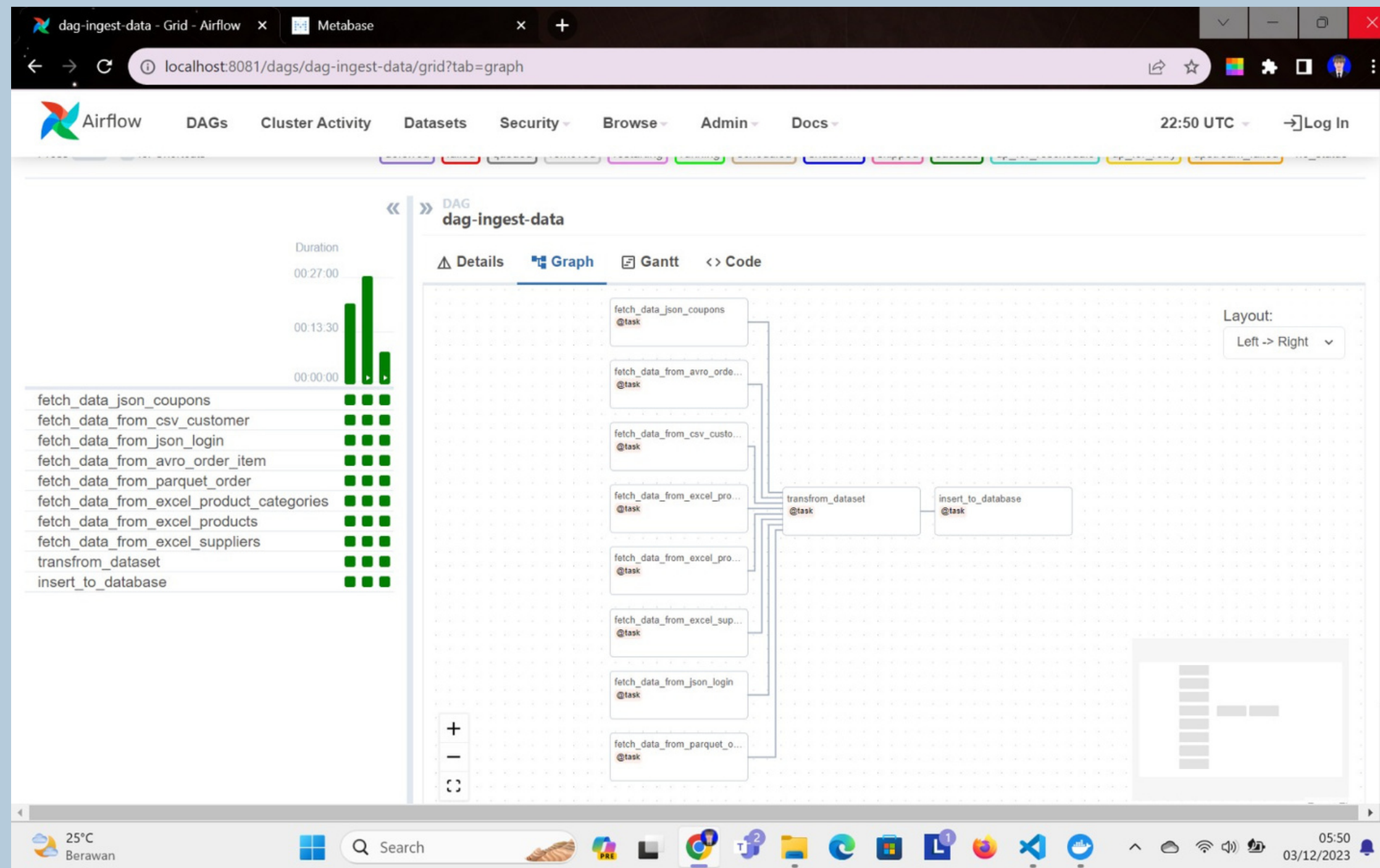
DAG

The screenshot displays the Apache Airflow web interface in a browser window. The address bar shows 'localhost:8081/home'. The top navigation bar includes links for DAGs, Cluster Activity, Datasets, Security, Browse, Admin, and Docs, along with the current time '22:50 UTC' and a 'Log In' button. The main section is titled 'DAGs' and features a filter bar with tabs for 'All' (2), 'Active' (2), 'Paused' (0), 'Running' (0), and 'Failed' (0). There is also a 'Filter DAGs by tag' input and a 'Search DAGs' field. An 'Auto-refresh' toggle is set to 'On'. Below the filter bar is a table listing DAGs:

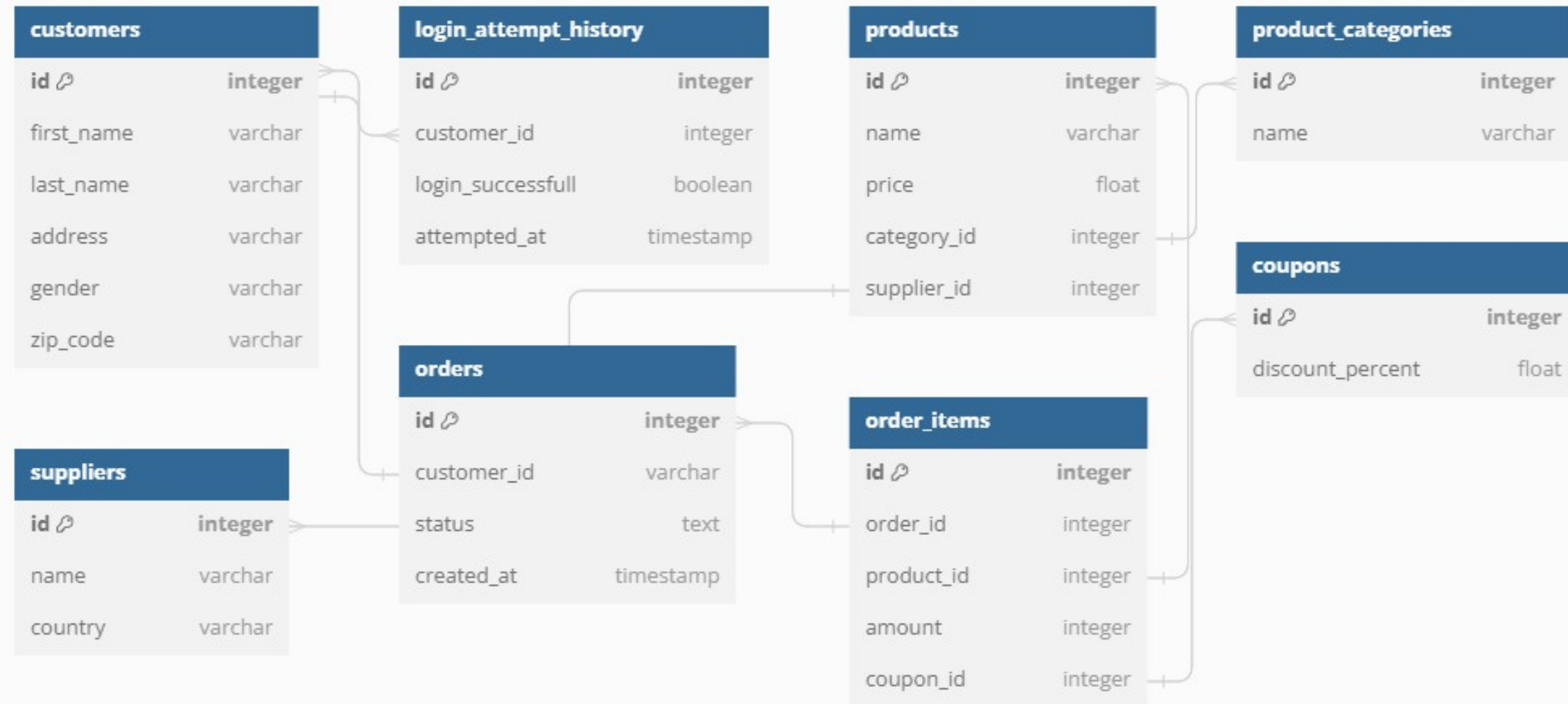
DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions	Links
DAG-1	airflow	2	@once	2023-12-01, 17:05:53		2	[Play] [Stop]	...
dag-ingest-data	Kelompok 9	3	@once	2023-12-01, 16:31:48		10	[Play] [Stop]	...

At the bottom of the table, there is a pagination control showing '1' and a status 'Showing 1-2 of 2 DAGs'. The browser's taskbar at the bottom shows the system clock as '05:50 03/12/2023' and the location as '25°C Berawan'.

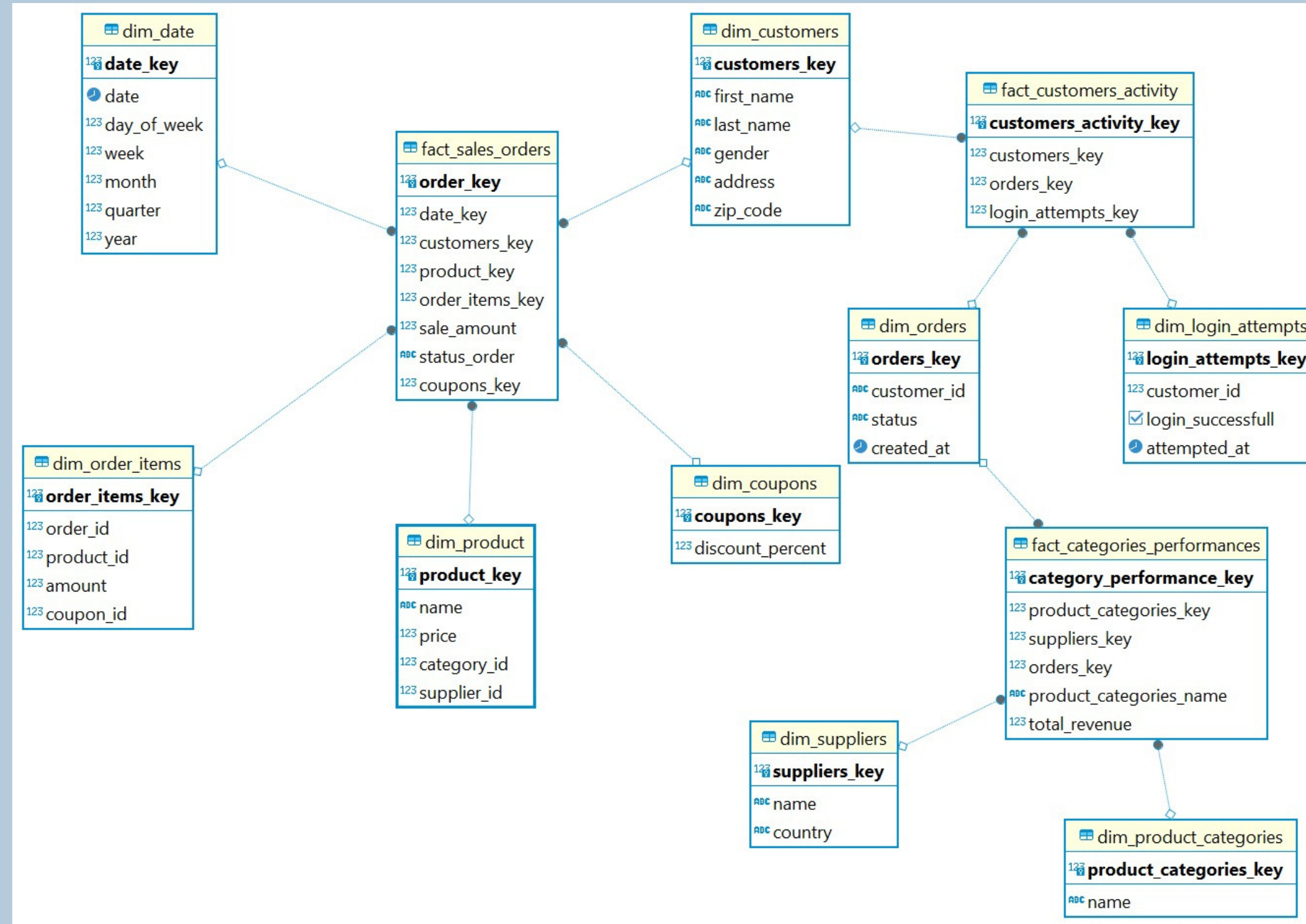
DAG



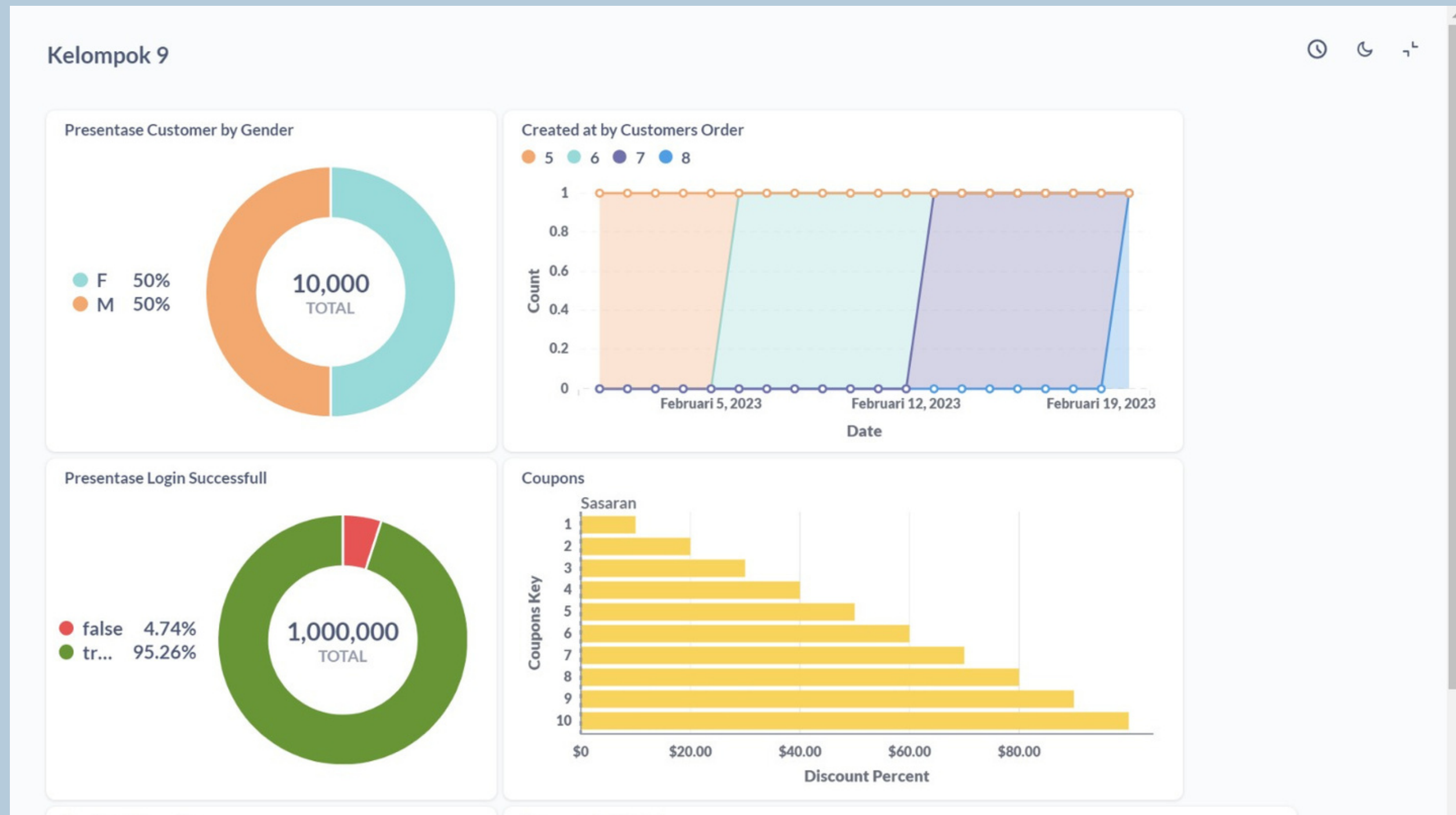
ERD



DATA MODELING ERD



DASHBOARD



DASHBOARD



KESIMPULAN

Proyek ini secara efektif membangun infrastruktur data yang komprehensif untuk perusahaan ritel online. Berikut adalah kesimpulan utama dari proyek ini:

Infrastuktur Data yang Terintegrasi: Dibangun sistem ETL/ELT yang menggunakan Apache Airflow untuk mengelola dan mengolah data dari berbagai format ke dalam data warehouse berbasis PostgreSQL. Penggunaan teknologi seperti Docker dan Git memastikan lingkungan pengembangan yang terisolasi dan manajemen versi yang efisien.

Data Modeling yang Efisien: Melakukan pemodelan data berdasarkan prinsip Kimball, dengan pembentukan tabel-tabel yang informatif dari data yang diberikan serta penggabungan yang kreatif antar tabel. Hal ini memungkinkan analisis mendalam atas data pelanggan, transaksi, dan produk.

KESIMPULAN

Visualisasi Data yang Informatif: Diciptakan dashboard dan visualisasi data menggunakan Metabase yang memberikan gambaran yang jelas mengenai metrik bisnis kunci dan tren yang dapat membantu manajemen dalam mengambil keputusan yang tepat.

Penggunaan Teknologi yang Relevan: Pemilihan alat dan teknologi (seperti Pandas, Polars, Metabase) didasarkan pada kebutuhan spesifik proyek dan memberikan kombinasi yang tepat antara fungsionalitas, kinerja, serta kemudahan penggunaan.

Kesempatan Pengembangan Lanjutan: Proyek ini menciptakan landasan yang kokoh namun juga memberikan ruang untuk pengembangan lanjutan. Terdapat potensi peningkatan melalui penambahan data tambahan, perbaikan performa, dan eksplorasi teknologi baru guna meningkatkan efisiensi dan pemahaman atas data.

Secara keseluruhan, proyek ini berhasil membangun infrastruktur data yang kuat dan siap digunakan untuk mendukung pengambilan keputusan yang lebih baik dan strategis bagi perusahaan ritel online ini.



Terima Kasih



Kelompok 9 - Final Project