

# IMPLEMENTASI PIPELINE ETL/ELT DAN MODEL DIMENSIONAL UNTUK ANALISIS PENJUALAN SHOPEE MENGGUNAKAN POSTGRESQL, DOCKER, DAN APACHE SUPERTSET

Pita Hutabalian<sup>1\*</sup>, Benedict Raditya Pradipta Ginting<sup>2</sup>, Magrozan Qobus Zaidan<sup>3</sup>, Najma Alfisyahrina<sup>4</sup>, Chaerur Rozikin<sup>5</sup>

<sup>1,2,3,4,5</sup>Universitas Singaperbangsa Karawang; Jl. HS.Ronggo Waluyo, Puseurjaya, TelukjambeTimur, Karawang, Jawa Barat 41361;Telp. (0267) 641177

## Keywords:

Shopee; postgresql; docker; apache superset; etl.

## Correspondent Email:

[pitahutabalian@gmail.com](mailto:pitahutabalian@gmail.com)

**Abstrak.** Penelitian ini mengimplementasikan pipeline analitik terintegrasi berbasis PostgreSQL, Docker, dan Apache Superset untuk mengonversi data mentah penjualan toko kosmetik di Shopee menjadi dashboard bisnis interaktif yang diperbarui secara otomatis. Proses Extract, Transform, Load (ETL) dilakukan menggunakan Python dan SQL untuk mengolah data transaksi, produk, diskon, serta pelanggan sebelum dimuat ke dalam data warehouse PostgreSQL yang dirancang menggunakan model Star Schema. Sistem dijalankan dalam lingkungan Docker untuk menjamin konsistensi dan portabilitas antar komponen, sedangkan Apache Superset digunakan untuk menghasilkan visualisasi data secara real-time. Hasil penelitian menunjukkan bahwa sistem ini mampu mengintegrasikan data penjualan secara terpusat, mempercepat proses analisis, serta memperbarui visualisasi secara otomatis tanpa intervensi manual. Pendekatan ini menawarkan solusi yang praktis, ringan, dan efisien bagi pelaku e-commerce dalam menerapkan analitik berbasis data warehouse tanpa memerlukan infrastruktur ETL yang kompleks.



Copyright © JITET (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

**Abstract.** This study implements an integrated analytical pipeline using PostgreSQL, Docker, and Apache Superset to transform raw sales data from a cosmetic store on Shopee into an interactive business dashboard that updates automatically. The Extract, Transform, Load (ETL) process was performed using Python and SQL to process transaction, product, discount, and customer data before loading them into a PostgreSQL-based data warehouse designed with a Star Schema model. The system was deployed within a Docker environment to ensure consistency and portability across components, while Apache Superset was used to generate real-time data visualizations. The results indicate that this system successfully centralized sales data,

*accelerated analysis processes, and enabled automatic visualization updates without manual intervention. This approach provides a practical, lightweight, and efficient solution for e-commerce businesses to implement data warehouse-based analytics without relying on complex ETL infrastructures.*

---

## 1. PENDAHULUAN

Perkembangan e-commerce di Indonesia menunjukkan peningkatan yang pesat dalam beberapa tahun terakhir. Platform seperti Shopee telah menjadi wadah utama bagi pelaku usaha, termasuk toko kosmetik, untuk memperluas jangkauan pasar dan meningkatkan penjualan secara digital. Aktivitas transaksi yang tinggi menghasilkan volume data yang besar dan beragam, mencakup data produk, pelanggan, promosi, serta penjualan. Data tersebut memiliki potensi strategis apabila dikelola dan dianalisis secara terintegrasi melalui sistem data warehouse yang terstruktur dan terpusat.

Namun, sebagian besar pelaku usaha masih menghadapi kendala dalam mengelola data yang tersebar di berbagai format seperti CSV, Excel, maupun API Shopee. Kondisi ini sering menyebabkan munculnya data silo, duplikasi informasi, dan kehilangan nilai penting yang berdampak pada penurunan kualitas analisis. Proses analisis manual pun memerlukan waktu yang lama dan rentan kesalahan, sehingga menghambat pengambilan keputusan berbasis data yang cepat, akurat, dan efisien.

Penelitian terdahulu menyoroti pentingnya penerapan data warehouse dan pipeline data untuk meningkatkan integrasi serta efisiensi analisis. Dinesh dan Devi [4] membuktikan bahwa optimasi proses integrasi berbasis cloud dapat mempercepat pemrosesan data hingga 60%. Khan [7] mengembangkan ETL Maturity Model untuk menilai tingkat kematangan sistem data warehouse, sementara Dhaouadi et al. [3] menegaskan bahwa

pendekatan model dimensional seperti Star Schema mampu meningkatkan performa analisis dibandingkan model relasional klasik. Bahaa et al. [1] menekankan pentingnya penerapan DataOps dalam meningkatkan reliabilitas pipeline data, dan Fan [6] menunjukkan bahwa sistem Enterprise Data Warehouse dapat mengelola data berskala besar dengan latensi rendah.

Dalam konteks e-commerce, penelitian Study on Data Warehousing for E-Commerce Logistics [10] menyoroti pentingnya integrasi data terpusat untuk analisis performa penjualan dan logistik. Sementara itu, The Role of Data Warehousing in the Infrastructure of E-Commerce [12] menjelaskan bahwa data warehouse merupakan komponen kunci untuk memahami perilaku konsumen secara real-time. Namun, hingga saat ini, masih terbatas penelitian yang secara spesifik mengintegrasikan PostgreSQL, Docker, dan Apache Superset dalam satu pipeline analitik yang sederhana, efisien, dan mudah diimplementasikan, khususnya pada analisis penjualan di platform Shopee.

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada implementasi pipeline ETL/ELT dan model dimensional berbasis PostgreSQL yang diintegrasikan dengan Apache Superset melalui Docker container. Superset digunakan untuk mengotomasi proses visualisasi data penjualan secara real-time, menggantikan pipeline ETL kompleks seperti Apache Airflow. Pendekatan ini diharapkan dapat meningkatkan efisiensi analisis data, mempercepat pengambilan keputusan bisnis, serta menjadi solusi ringan

dan terintegrasi bagi pelaku usaha e-commerce dalam menerapkan data-driven decision making.

## 2. TINJAUAN PUSTAKA

### 2.1. Konsep Data Engineering dan Pipeline ETL/ELT

Data engineering merupakan disiplin ilmu yang berfokus pada proses pengumpulan, pengolahan, dan penyajian data agar siap digunakan untuk analisis maupun pengambilan keputusan berbasis data. Salah satu komponennya adalah data pipeline, yaitu alur otomatis yang mengintegrasikan data dari berbagai sumber ke dalam sistem penyimpanan terpusat (data warehouse).

Model Extract, Transform, Load (ETL) terdiri dari tiga tahap utama: ekstraksi data mentah dari sumber, transformasi untuk menyesuaikan format serta meningkatkan kualitas data, dan pemuatan hasilnya ke dalam data warehouse [4]. Sementara itu, pendekatan Extract, Load, Transform (ELT) merupakan pengembangan dari model ETL tradisional, di mana data mentah dimuat terlebih dahulu ke dalam data warehouse sebelum dilakukan transformasi. Pendekatan ELT dianggap lebih efisien pada sistem modern seperti PostgreSQL, BigQuery, atau Snowflake yang mendukung komputasi langsung di sisi basis data [8][11].

Dinesh dan Devi [4] menunjukkan bahwa optimasi proses integrasi berbasis cloud dapat meningkatkan kecepatan pemrosesan data hingga 60%. Khan [7] mengusulkan ETL Maturity Model berbasis Capability Maturity Model Integration (CMMI) untuk menilai kesiapan organisasi dalam mengelola pipeline data, sementara Bahaa et al. [1] menekankan pentingnya konsep DataOps dalam meningkatkan efisiensi, konsistensi, dan reliabilitas proses ETL. Dalam penelitian ini, konsep ETL/ELT diimplementasikan menggunakan Python dan PostgreSQL untuk mengelola data penjualan Shopee secara terstruktur sebelum divisualisasikan melalui

Apache

Superset.

Data engineering merupakan disiplin ilmu yang berfokus pada proses pengumpulan, pengolahan, dan penyajian data agar siap digunakan untuk analisis maupun pengambilan keputusan berbasis data. Salah satu komponennya adalah data pipeline, yaitu alur otomatis yang mengintegrasikan data dari berbagai sumber ke dalam sistem penyimpanan terpusat (data warehouse) [9].

### 2.2. Data Warehouse dan Model Dimensional

Data warehouse (DW) merupakan sistem penyimpanan terpusat yang mengintegrasikan data dari berbagai sumber untuk mendukung analisis historis dan strategis. Secara umum, arsitektur DW terdiri dari tiga lapisan utama: (1) staging area sebagai penampung data mentah, (2) integration layer untuk pembersihan dan transformasi data, serta (3) presentation layer untuk penyajian data dalam bentuk analitik dan visualisasi [3].

Pendekatan dimensional modeling digunakan untuk memudahkan analisis data dengan membagi struktur menjadi dua jenis tabel, yaitu tabel fakta (berisi metrik kuantitatif seperti jumlah penjualan dan pendapatan) dan tabel dimensi (berisi atribut deskriptif seperti produk, pelanggan, dan waktu). Model ini memungkinkan analisis multi-dimensi yang cepat dan mudah dipahami.

Dhaouadi et al. [3] menjelaskan bahwa model Star Schema merupakan desain paling populer karena kesederhanaannya dan performa tinggi dalam eksekusi query. Penelitian BCNF – From Normalized to Star Schemas and Back Again [15] juga menunjukkan bahwa Star Schema mampu meningkatkan efisiensi analisis dibandingkan model relasional yang terlalu kompleks. Dalam penelitian ini, model Star Schema diterapkan pada PostgreSQL sebagai struktur utama data warehouse untuk mendukung analisis penjualan kosmetik di Shopee.

### 2.3. Automasi dan Orkestrasi Data

Automasi berperan penting dalam menjaga konsistensi dan keandalan proses pengolahan data. Beberapa framework populer seperti Apache Airflow dan Prefect digunakan untuk mengatur jadwal serta dependensi dalam pipeline ETL secara otomatis [13]. Selain itu, data transformation tool seperti dbt (data build tool) mendukung pengelolaan transformasi SQL modular yang transparan dan terintegrasi dengan version control.

Dalam aspek validasi, Great Expectations digunakan secara luas untuk menjamin kualitas data dengan mendeteksi duplikasi, kesalahan tipe data, dan missing values [5]. Yang et al. [14] mengembangkan metode Internal Data Imputation untuk memperbaiki nilai kosong pada tabel dimensi agar data tetap konsisten.

Pada penelitian ini, proses otomatisasi penuh menggunakan Airflow tidak diterapkan. Sebagai gantinya, sistem mengandalkan Apache Superset sebagai lightweight orchestration layer yang secara otomatis memperbarui visualisasi data setiap kali terjadi perubahan pada database PostgreSQL. Pendekatan ini dinilai lebih sederhana namun tetap mampu menjaga integritas dan keterbaruan data secara real-time di lingkungan Docker.

### 2.4. Penerapan Data Warehouse pada E-Commerce

Penerapan data warehouse dalam bidang e-commerce memiliki peran penting dalam menganalisis performa penjualan, perilaku konsumen, dan efektivitas promosi. Penelitian Study on a Data Warehousing for E-Commerce Logistics [10] menunjukkan bahwa integrasi DW meningkatkan efisiensi analisis logistik dan pelacakan pesanan pelanggan. The Role of Data Warehousing in the Infrastructure of E-Commerce [12] juga menegaskan bahwa DW merupakan komponen utama dalam mendukung pengambilan keputusan bisnis berbasis data.

Theofilou et al. [11] mengembangkan arsitektur scalable data warehouse untuk

pengelolaan data besar di sektor pertanian yang relevan diterapkan pada e-commerce, mengingat kesamaan karakteristik data yang dinamis dan berskala besar. Sementara itu, Ebadifard et al. [5] membuktikan bahwa automasi pipeline ETL mampu meningkatkan performa analitik untuk business intelligence dan machine learning.

Sebagian besar penelitian terdahulu masih berfokus pada pipeline ETL berbasis cloud atau Airflow yang kompleks [1][4][13]. Berbeda dengan itu, penelitian ini menghadirkan pendekatan yang lebih sederhana dan terintegrasi, menggunakan PostgreSQL sebagai data warehouse, Docker sebagai lingkungan terkontainerisasi, dan Apache Superset sebagai platform visualisasi otomatis.

Pendekatan lightweight pipeline ini diharapkan dapat menjadi solusi efisien dan ekonomis bagi pelaku usaha e-commerce, khususnya UMKM, dalam menerapkan analisis data secara real-time dan mendukung data-driven decision making. Sejalan dengan penelitian Harnelia [2] yang menganalisis ulasan produk skincare di e-commerce menggunakan algoritma Support Vector Machine (SVM), penelitian ini juga menyoroti pentingnya pemanfaatan data digital untuk menghasilkan insight bisnis yang relevan bagi sektor ritel dan kecantikan.

## 3. METODE PENELITIAN

### 3.1. Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimen terapan di bidang data engineering, dengan tujuan mengimplementasikan pipeline ETL/ELT yang terintegrasi untuk menganalisis data penjualan toko kosmetik di platform Shopee. Pendekatan ini dipilih karena memungkinkan pengujian langsung terhadap efektivitas integrasi antar komponen sistem mulai dari proses ekstraksi data mentah hingga penyajian visualisasi analitik.

Tahapan penelitian dirancang secara sistematis dalam empat proses utama, yaitu:

1. Ekstraksi data dari berbagai sumber untuk memperoleh dataset mentah;
2. Transformasi data untuk membersihkan, menormalkan, dan menyesuaikan struktur dengan kebutuhan analisis;
3. Pemuatan ke data warehouse berbasis PostgreSQL menggunakan model dimensional (Star Schema); dan
4. Visualisasi interaktif melalui Apache Superset yang dijalankan dalam lingkungan Docker.

Setiap tahapan saling terhubung dalam satu pipeline yang bertujuan meningkatkan efisiensi pengolahan data serta mendukung proses pengambilan keputusan berbasis data (data-driven decision making).

### 3.2. Sumber dan Jenis Data

Data yang digunakan merupakan data sekunder hasil ekspor dari Shopee Seller Center milik toko kosmetik yang aktif beroperasi. Periode data yang digunakan mencakup Maret 2024 hingga Mei 2025, disimpan dalam format Excel (.xlsx), kemudian dikonversi menjadi CSV untuk proses ekstraksi dan integrasi. Dataset terdiri atas empat kelompok utama:

- Data Produk : memuat informasi kode produk, nama, kategori, harga, stok, dan status ketersediaan.
- Data Penjualan : berisi transaksi harian yang mencakup tanggal, jumlah produk terjual, total pendapatan, serta potongan harga.
- Data Diskon : memuat informasi jenis diskon, nilai potongan, dan periode promosi yang berlangsung.
- Data Voucher : berisi kode voucher, persentase potongan, masa berlaku, dan kategori produk yang berlaku.

Keempat kelompok data tersebut digunakan untuk menganalisis hubungan antara strategi promosi dan performa penjualan, serta mengevaluasi pengaruh diskon terhadap peningkatan volume transaksi. Pendekatan ini

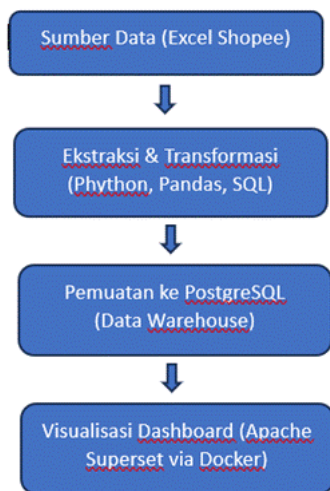
memberikan gambaran menyeluruh terhadap dinamika penjualan toko kosmetik di platform Shopee.

### 3.3. Arsitektur Sistem dan Lingkungan Implementasi

Arsitektur sistem pada penelitian ini terdiri atas tiga komponen utama, yaitu PostgreSQL, Docker, dan Apache Superset, yang saling terintegrasi membentuk pipeline analitik.

1. PostgreSQL (Data Warehouse)  
PostgreSQL berperan sebagai sistem penyimpanan terpusat (centralized storage system) yang menampung hasil ekstraksi dan transformasi data. Struktur database dirancang menggunakan model Star Schema, terdiri atas tabel fakta dan beberapa tabel dimensi untuk mendukung analisis multi-dimensi penjualan.
2. Docker (Platform Kontainerisasi)  
Docker digunakan sebagai container platform untuk menjalankan seluruh layanan (PostgreSQL dan Superset) dalam lingkungan yang terisolasi dan konsisten. Dengan pendekatan containerization, sistem menjadi mudah direplikasi, dideploy, dan dijalankan lintas perangkat tanpa konfigurasi kompleks.
3. Apache Superset (Visualisasi dan Analitik Bisnis)  
Apache Superset berfungsi sebagai platform business intelligence (BI) yang menampilkan hasil analisis data dalam bentuk dashboard interaktif. Superset dihubungkan dengan PostgreSQL melalui koneksi SQLAlchemy, memungkinkan pembaruan visualisasi secara otomatis setiap kali data di warehouse diperbarui.

Arsitektur sistem secara keseluruhan dapat digambarkan sebagai berikut:



Gambar 1. Arsitektur Sistem Pipeline ETL/ELT

Seluruh implementasi dijalankan pada Windows 10 Pro dengan konfigurasi berikut:

- Python 3.10 (library: pandas, sqlalchemy, psycopg2)
- PostgreSQL 14
- Docker Desktop 4.34
- Apache Superset 4.1.2

### 3.4. Tahapan Penelitian

Proses implementasi pipeline dilakukan dalam empat fase utama yang menggambarkan alur kerja ETL/ELT secara lengkap.

#### 3.4.1. Tahap Ekstraksi Data

Pada tahap ini, data diekstrak dari file Excel menggunakan Python (library Pandas). Proses ekstraksi meliputi penggabungan beberapa sumber data menjadi satu dataset utama serta pembersihan awal, seperti:

- Penyeragaman nama kolom,
- Penghapusan baris kosong dan duplikasi, dan
- Konversi format tanggal dan numerik untuk memastikan konsistensi tipe data.

Hasil ekstraksi disimpan dalam format CSV untuk memudahkan proses transformasi berikutnya.

#### 3.4.2. Tahap Transformasi Data

Tahap transformasi dilakukan untuk memastikan kualitas, konsistensi, dan keterpaduan data sebelum dimuat ke dalam data warehouse. Proses ini bertujuan mengubah data mentah menjadi bentuk yang sesuai untuk analisis kuantitatif.

Transformasi meliputi beberapa aktivitas utama, antara lain:

- Normalisasi nama produk dan kategori,
- Penyesuaian satuan harga dan diskon,
- Pembuatan kolom turunan seperti Net Sales, Discount Rate, dan Voucher Impact,
- Pengelompokan data berdasarkan tanggal transaksi dan kategori produk.

Seluruh proses transformasi dilakukan menggunakan SQL Script di PostgreSQL, dengan tujuan meningkatkan efisiensi dan mendukung otomatisasi pada tahap pemuatan.

#### 3.4.3. Tahap Pemuatan Data

Data hasil transformasi dimasukkan ke dalam PostgreSQL menggunakan desain Star Schema, dengan struktur:

- Tabel Fakta (fact\_sales): menyimpan data transaksi, total penjualan, jumlah produk terjual, dan potongan harga.
- Tabel Dimensi (dim\_product, dim\_promotion, dim\_date): menyimpan data deskriptif untuk mendukung analisis multi-dimensi.

Model Star Schema dipilih karena sederhana, mudah dipahami, dan efisien untuk analisis penjualan e-commerce.

#### 3.4.4. Tahap Visualisasi

Tahap terakhir adalah proses visualisasi data dengan menggunakan Apache

Superset yang terhubung langsung ke data warehouse PostgreSQL. Tujuannya adalah menyajikan data dalam bentuk dashboard interaktif untuk mendukung proses pengambilan keputusan.

Dashboard yang dikembangkan menampilkan berbagai komponen analitik, antara lain:

- Tren penjualan harian dan bulanan,
- Kontribusi setiap kategori produk terhadap total penjualan,
- Pengaruh promosi dan diskon terhadap performa produk, serta
- Perbandingan penjualan sebelum dan sesudah periode promosi.

Hasil visualisasi ini memungkinkan pengguna untuk memperoleh insight bisnis secara real-time tanpa perlu melakukan analisis manual yang berulang. Dengan demikian, sistem yang dibangun tidak hanya berfungsi sebagai alat analitik, tetapi juga sebagai pendukung strategis dalam penerapan data-driven decision making.

### 3.5. Validasi dan Evaluasi

Tahap validasi dan evaluasi dilakukan untuk memastikan bahwa sistem yang dikembangkan telah menghasilkan data dan visualisasi yang akurat, konsisten, dan efisien sesuai dengan tujuan penelitian.

Validasi Sistem

Validasi dilakukan untuk memeriksa ketepatan hasil pengolahan data antara sumber asli dan data yang telah dimuat ke data warehouse. Proses validasi mencakup tiga langkah utama, yaitu:

1. Perbandingan hasil agregasi antara data yang tersimpan di PostgreSQL dengan laporan penjualan asli dari Shopee Seller Center. Langkah ini bertujuan memastikan tidak terjadi kehilangan

atau duplikasi data selama proses ekstraksi dan transformasi.

2. Pemeriksaan kesesuaian periode promosi dan diskon pada tabel dimensi dengan catatan kampanye promosi di Shopee, untuk menjamin konsistensi antara data transaksi dan data promosi.
3. Uji coba pembaruan visualisasi di Apache Superset, guna memastikan proses otomatisasi berjalan dengan benar setiap kali data terbaru dimuat ke dalam PostgreSQL.

Hasil validasi menunjukkan bahwa seluruh data penjualan, diskon, dan voucher telah dimuat secara konsisten, tanpa terdapat ketidaksesuaian nilai antara sumber dan hasil akhir di dashboard.

Evaluasi Kinerja Sistem

Evaluasi sistem dilakukan untuk menilai efisiensi dan ketersediaan data (data availability) dalam pipeline yang dibangun. Dua indikator utama digunakan dalam proses evaluasi ini, yaitu:

1. Efisiensi waktu pemrosesan, yakni waktu yang dibutuhkan sistem untuk memproses data dari tahap ekstraksi hingga penyajian visualisasi.
2. Ketersediaan data (data availability), yaitu kemampuan sistem dalam memperbarui dan menampilkan data terbaru secara otomatis tanpa intervensi manual.

Berdasarkan hasil pengujian, pipeline yang dikembangkan mampu memproses sekitar 4.000 baris data dalam waktu kurang dari dua menit, serta menghasilkan dashboard interaktif yang selalu up-to-date dengan data terbaru di PostgreSQL. Hal ini menunjukkan bahwa integrasi antara PostgreSQL, Docker, dan Apache Superset terbukti meningkatkan efisiensi analisis penjualan secara signifikan dibandingkan metode manual.

Temuan tersebut menegaskan bahwa

pendekatan lightweight data warehouse pipeline yang digunakan dalam penelitian ini dapat menjadi solusi efektif dan praktis untuk mendukung penerapan analisis bisnis berbasis data (data-driven business intelligence) di sektor e-commerce.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Implementasi Sistem Pipeline ETL/ELT

Implementasi sistem pipeline pada penelitian ini dirancang menggunakan tiga komponen utama, yaitu PostgreSQL, Docker, dan Apache Superset, yang saling terhubung dalam satu alur proses terintegrasi. Pipeline ini dikembangkan untuk mengotomatisasi proses pengolahan data penjualan toko kosmetik di Shopee, dimulai dari tahap ekstraksi data mentah hingga penyajian visualisasi interaktif berbasis business intelligence.

Proses Extract, Transform, and Load (ETL) dilakukan secara modular menggunakan Python (library Pandas) dan SQL pada PostgreSQL. Data hasil ekspor dari Shopee Seller Center dalam format .xlsx dikonversi menjadi .csv, kemudian diekstraksi ke dalam staging area PostgreSQL melalui koneksi SQLAlchemy.

Tahap transformasi dilakukan untuk memastikan konsistensi dan validitas data sebelum dimuat ke data warehouse. Transformasi yang dilakukan meliputi:

1. Penyeragaman nama kolom dan tipe data (varchar, numeric, date);
2. Penghapusan nilai duplikat dan missing values;
3. Pembuatan kolom turunan seperti net\_sales, discount\_rate, dan voucher\_impact; serta
4. Pengelompokan data berdasarkan kategori produk dan periode penjualan.

Data hasil transformasi kemudian dimuat ke dalam PostgreSQL yang dirancang dengan model dimensional Star Schema, terdiri atas satu fact table (fact\_sales) dan tiga

dimension tables (dim\_product, dim\_promotion, dim\_date).

Implementasi sistem dijalankan menggunakan Docker Compose, yang mengintegrasikan dua kontainer utama, yaitu postgres\_container untuk penyimpanan data warehouse dan superset\_container untuk penyajian visualisasi data. Pendekatan berbasis kontainer ini memungkinkan sistem dijalankan secara konsisten di berbagai lingkungan tanpa konfigurasi manual tambahan, serta mempermudah proses replikasi dan pengujian lintas perangkat.



Gambar 2. Arsitektur Sistem Pipeline ETL/ELT

### 4.2. Desain Model Dimensional

Desain data warehouse pada penelitian ini menggunakan pendekatan Star Schema agar proses analisis lebih cepat, efisien, dan mudah dipahami. Model ini terdiri atas satu tabel fakta dan beberapa tabel dimensi yang terhubung langsung melalui primary key dan foreign key.

Struktur utamanya terdiri atas:

- Tabel Fakta (fact\_sales): menyimpan data kuantitatif seperti total penjualan, jumlah produk terjual, nilai diskon, dan klaim voucher.
- Tabel Dimensi Produk (dim\_product): berisi atribut deskriptif seperti nama produk, kategori, harga, dan stok.
- Tabel Dimensi Promosi (dim\_promotion): memuat jenis promosi, periode diskon, dan nilai potongan.
- Tabel Dimensi Waktu (dim\_date): menyimpan informasi tanggal transaksi, bulan, tahun, dan periode promosi.



Pendekatan ini memungkinkan analisis multi-dimensional (OLAP analysis) seperti drill-down, roll-up, dan slice-dice, sehingga pengguna dapat menelusuri performa penjualan berdasarkan waktu, kategori produk, maupun jenis promosi dengan fleksibel.

#### 4.3. Dashboard Analitik Penjualan di Apache Superset

Dashboard analitik yang dikembangkan menggunakan Apache Superset menampilkan data hasil pengolahan PostgreSQL secara interaktif. Dashboard ini terdiri atas empat komponen utama:

1. Penjualan berdasarkan tipe promosi;
2. Tren penjualan per bulan;
3. Efektivitas penggunaan voucher; dan
4. Perbandingan produk dikunjungi terhadap produk yang dimasukkan ke keranjang.

Data divisualisasikan dalam berbagai bentuk grafik seperti bar chart, line chart, dan comparative plot untuk mempermudah interpretasi performa bisnis secara visual. Koneksi antara PostgreSQL dan Apache Superset dibangun melalui URI SQLAlchemy sehingga setiap kali terjadi pembaruan data di PostgreSQL, visualisasi pada dashboard akan diperbarui secara otomatis tanpa proses manual.



Gambar 3. Dashboard Analitik Penjualan di Apache Superset

#### 4.4. Analisis Hasil Visualisasi

Hasil visualisasi yang diperoleh dari dashboard Superset menunjukkan beberapa pola dan kecenderungan penting dalam aktivitas penjualan toko kosmetik di Shopee.

##### a. Penjualan Berdasarkan Tipe Promosi

Grafik memperlihatkan bahwa promosi diskon reguler memberikan kontribusi terbesar terhadap peningkatan volume transaksi, terutama pada bulan Januari dan Maret 2025, dengan total transaksi mencapai lebih dari 18 per kategori promosi. Sebaliknya, promosi bundling dan kombo hemat menunjukkan performa lebih rendah. Hasil ini sejalan dengan penelitian Study on Data Warehousing for E-Commerce Logistics [10] yang menegaskan pentingnya evaluasi promosi untuk meningkatkan konversi penjualan.

##### b. Data Penjualan per Bulan

Tren penjualan bulanan menunjukkan fluktuasi signifikan, dengan puncak penjualan terjadi pada November 2024 sebesar 20–25 juta rupiah, yang diduga dipengaruhi oleh kampanye Shopee 11.11 Big Sale. Penurunan penjualan terjadi pada Februari 2025 akibat keterbatasan stok dan penurunan intensitas promosi. Temuan ini menunjukkan bahwa pipeline mampu mengidentifikasi pola musiman penjualan secara cepat dan akurat, yang dapat dijadikan acuan dalam perencanaan stok dan strategi promosi.

##### c. Efektivitas Penggunaan Voucher

Analisis tingkat penggunaan voucher menunjukkan penurunan bertahap dari Januari 2025 (sekitar 800 klaim) hingga hampir nol pada Mei 2025. Kondisi ini mengindikasikan menurunnya minat pelanggan terhadap voucher atau kurangnya promosi yang relevan. Bagi pelaku usaha, informasi ini penting sebagai dasar untuk memperbarui strategi diskon dan memperkuat promosi berbasis loyalitas pelanggan.

##### d. Perbandingan Produk Dikunjungi vs Dimasukkan ke Keranjang

Grafik perbandingan menunjukkan bahwa rasio add-to-cart rata-rata mencapai 60–70%, namun tidak semua produk yang dikunjungi berakhir pada transaksi pembelian. Hal ini menunjukkan adanya potensi kehilangan pelanggan di tahap checkout. Informasi tersebut berguna untuk meningkatkan kualitas deskripsi produk, strategi harga, dan efektivitas promosi agar tingkat konversi penjualan meningkat.

4.5. Evaluasi Kinerja Pipeline  
Evaluasi kinerja dilakukan untuk menilai efisiensi dan keandalan pipeline yang dibangun. Dua indikator utama yang digunakan adalah waktu pemrosesan (processing time) dan ketersediaan data (data availability).

Pipeline mampu memproses sekitar 4.000 baris data penjualan dalam waktu kurang dari dua menit, mencakup seluruh tahapan extract, transform, load hingga pembaruan visualisasi di Apache Superset. Hasil ini menunjukkan peningkatan efisiensi signifikan dibandingkan proses manual menggunakan Excel yang memerlukan waktu sekitar 15 menit untuk tahapan yang sama.

Integrasi langsung antara PostgreSQL dan Apache Superset melalui Docker memungkinkan pembaruan visualisasi secara otomatis tanpa intervensi pengguna. Hal ini menegaskan bahwa sistem mendukung konsep real-time analytics dengan tingkat ketersediaan data yang tinggi, sehingga hasil analisis dapat digunakan langsung untuk pengambilan keputusan.

Secara keseluruhan, sistem ini dapat dikategorikan sebagai lightweight data analytics pipeline yang efisien dan stabil, terutama bagi pelaku usaha e-commerce berskala kecil hingga menengah. Temuan ini memperkuat hasil penelitian Dhaouadi et al. [3] dan Fan [6] yang menyatakan bahwa model dimensional berbasis data warehouse mampu meningkatkan kecepatan analisis serta reliabilitas data. Dengan demikian, integrasi antara PostgreSQL, Docker, dan Apache Superset dapat dijadikan alternatif solusi praktis dalam implementasi sistem analitik e-

commerce tanpa ketergantungan pada infrastruktur cloud-based ETL yang kompleks dan mahal.

## 5. KESIMPULAN

Berdasarkan hasil penelitian dan implementasi sistem yang telah dilakukan, dapat disimpulkan bahwa:

1. Pipeline analitik terintegrasi berhasil dikembangkan menggunakan PostgreSQL, Docker, dan Apache Superset untuk mengotomatisasi proses pengolahan data penjualan toko kosmetik di Shopee, mulai dari ekstraksi, transformasi, hingga visualisasi interaktif secara real-time.
2. Proses ETL/ELT yang diimplementasikan dengan Python dan SQL mampu menormalkan, membersihkan, serta mengonversi data penjualan, diskon, dan voucher ke dalam struktur data warehouse berbasis model dimensional Star Schema, yang memudahkan analisis multi-dimensi.
3. Evaluasi kinerja sistem menunjukkan peningkatan efisiensi signifikan, dengan waktu pemrosesan  $\pm 4.000$  baris data dalam kurang dari dua menit, dibandingkan proses manual berbasis Excel yang memerlukan waktu sekitar 15 menit.
4. Integrasi antara PostgreSQL dan Apache Superset melalui Docker memungkinkan pembaruan visualisasi data secara otomatis tanpa intervensi manual, sehingga mendukung konsep real-time analytics dan meningkatkan keandalan sistem dalam mendukung pengambilan keputusan bisnis.
5. Dashboard analitik yang dihasilkan memberikan insight strategis terkait tren penjualan, efektivitas promosi, serta perilaku konsumen, yang berguna bagi pelaku usaha e-commerce dalam menerapkan data-driven decision making.

6. Pendekatan lightweight data analytics pipeline ini terbukti efektif, ekonomis, dan skalabel, sehingga dapat menjadi solusi praktis bagi usaha kecil hingga menengah untuk mengimplementasikan analitik berbasis data warehouse tanpa ketergantungan pada infrastruktur ETL berbasis cloud.
7. Untuk penelitian selanjutnya, disarankan pengembangan sistem dengan penambahan orkestrasi otomatis menggunakan workflow manager (misalnya Apache Airflow atau Prefect), serta integrasi dengan machine learning pipeline untuk memprediksi tren penjualan dan perilaku pelanggan secara lebih komprehensif.

## 6. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Chaerur Rozikin, S.Kom., M.Kom. selaku dosen pembimbing yang telah memberikan arahan, motivasi, dan bimbingan ilmiah selama proses penelitian hingga penyusunan artikel ini. Ucapan terima kasih juga disampaikan atas segala dukungan, waktu, dan kesempatan yang diberikan sehingga penelitian berjudul “Implementasi Pipeline ETL/ELT dan Model Dimensional untuk Analisis Penjualan Shopee Menggunakan PostgreSQL, Docker, dan Apache Superset” dapat terselesaikan dengan baik dan layak untuk dipublikasikan.

## DAFTAR PUSTAKA

- [1] Bahaa, A., et al. (2021). *A systematic literature review for implementing DataOps in the data warehouse lifecycle during the ETL phase*. Journal of Computer Science, 17(8), 1011–1030. <https://doi.org/10.3844/jcssp.2021.1011.1030>
- [2] H. Harnelia, “Analisis Sentimen Review Skincare Skintific Dengan Algoritma Support Vector Machine (SVM),” Jurnal Informatika dan Teknik Elektro Terapan (JITET), vol. 12, no. 2, Apr. 2024.
- [3] Dhaouadi, A., et al. (2022). *Data warehousing process modeling from classical to new paradigms*. Data, 7(8), 113. <https://doi.org/10.3390/data7080113>
- [4] Dinesh, L., & Devi, K. G. (2024). *An efficient hybrid optimization of ETL process in data warehouse of cloud architecture*. Journal of Cloud Computing. <https://doi.org/10.1186/s13677-023-00571-y>
- [5] Ebadifard, N., et al. (2023). *Data extraction, transformation, and loading process automation for algorithmic trading machine learning modelling and performance optimization*. arXiv preprint arXiv:2312.12774.
- [6] Fan, X. (2024). *Enterprise level data warehouse system based on Hive*. Procedia Computer Science, 232, 1172–1178. <https://doi.org/10.1016/j.procs.2024.03.187>
- [7] Khan, M. (2022). *ETL maturity model for data warehouse systems: A CMMI approach*. Computer Standards & Interfaces, 81(3), 103–119. <https://doi.org/10.1016/j.csi.2022.103493>
- [8] Nath, R. P. D., et al. (2020). *High-level ETL for semantic data warehouses*. arXiv preprint arXiv:2006.07180.
- [9] *Optimizing ETL processes for high-volume data in financial applications*. (2023). Journal of Information Systems and Emerging Management, 13(2), 210–218.
- [10] *Study on a data warehousing for e-commerce logistics*. (2023). CEUR Workshop Proceedings, 3680, 1–7. Retrieved from <https://ceur-ws.org/Vol-3680/S3Paper14.pdf>
- [11] Theofilou, A., Gkountaras, N., & Pappa, A. (2025). *Design and implementation of a scalable data warehouse for agricultural big data*. Sustainability, 17(8), 3727. <https://doi.org/10.3390/su17083727>
- [12] *The role of data warehousing in the infrastructure of e-commerce*. (2022). International Journal of Applied Engineering and Management. Retrieved from [https://ijaem.net/issue\\_dcp/The\\_Role\\_of\\_Data\\_Warehousing\\_In\\_the\\_Infrastructure\\_of\\_E-Commerce.pdf](https://ijaem.net/issue_dcp/The_Role_of_Data_Warehousing_In_the_Infrastructure_of_E-Commerce.pdf)

- [13] Uddin, M. K. S., Rahman, M. A., & Islam, M. A. (2024). *A review of implementing AI-powered data warehouse solutions to optimize big data management and utilization*. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.4993596>
- [14] Yang, Y., Abdelhedi, F., Darmont, J., Ravat, F., & Teste, O. (2021). *Internal data imputation in data warehouse dimensions*. arXiv preprint arXiv:2110.01228.
- [15] *BCNF – From normalized to star schemas and back again*. (2023). *ACM SIGMOD Proceedings*, 1–12. <https://doi.org/10.1145/3555041.3589712>