

PERANCANGAN FACT TABLE PENJUALAN MENGGUNAKAN PIPELINE ETL BERBASIS PENTAHO DATA INTEGRATION (PDI) DAN SQLYOG

**Dwi Putri N. Sitanggang¹, Ameiliani Dianira Barahama²,
Irene R. H. T. Tangkawarow³**

^{1,2,3} Jurusan Pendidikan Teknologi Informasi dan Komunikasi, Fakultas Teknik,
Universitas Negeri Manado

e-mail: ¹sitanggangnatalia2@gmail.com, ²ameilianibarahama@unima.ac.id,

³irene.tangkawarow@unima.ac.id

ABSTRAK

Pertumbuhan volume data transaksi dalam dunia bisnis retail menuntut adanya sistem pengolahan data yang cepat, akurat, dan terintegrasi untuk mendukung pengambilan keputusan strategis. Masalah utama yang sering dihadapi adalah tersebarnya data operasional pada berbagai format database yang mengakibatkan keterlambatan dalam penyusunan laporan penjualan. Penelitian ini bertujuan untuk merancang fact table penjualan menggunakan pipeline Extract, Transform, Load (ETL) berbasis Pentaho Data Integration (PDI) dan mengimplementasikannya pada database target menggunakan SQLyog sebagai user interface MySQL. Metodologi yang digunakan mengadopsi pendekatan perancangan multidimensional siklus hidup Kimball (Five-Step Dimensional Design) yang meliputi pemilihan proses bisnis, deklarasi grain, identifikasi dimensi, identifikasi fakta numerik, dan perancangan skema bintang. Hasil penelitian menunjukkan bahwa pipeline ETL yang dibangun berhasil mengotomatisasi pembersihan data, konversi tipe data, penanganan relasi surrogate key, hingga pemuatan data dari sistem OLTP ke dalam star schema secara konsisten. Kesimpulan dari penelitian ini membuktikan bahwa integrasi Pentaho Data Integration dan SQLyog mampu mengoptimalkan efisiensi waktu eksekusi query data warehousing serta meminimalisir redundansi data, sehingga menyediakan fondasi data yang valid bagi keperluan analisis business intelligence.

Kata Kunci: *Data Warehouse, ETL, Fact Table, Pentaho Data Integration, SQLyog.*

ABSTRACT

The growth of transaction data volume in retail businesses demands a fast, accurate, and integrated data processing system to support strategic decision-making. The main problem often faced is the scattered operational data in various database formats, resulting in delays in preparing sales reports. This study aims to design a sales fact table using an Extract, Transform, Load (ETL) pipeline based on Pentaho Data Integration (PDI) and implement it on the target database using SQLyog as the MySQL user interface. The methodology adopted is the Kimball multidimensional lifecycle design approach (Five-Step Dimensional Design) which includes business process selection, grain declaration, dimension identification, numerical fact identification, and star

schema design. The results show that the ETL pipeline successfully automates data cleansing, data type conversion, surrogate key relationship handling, and data loading from the OLTP system to the OLAP star schema consistently. The conclusion proves that the integration of Pentaho Data Integration and SQLyog optimizes the efficiency of data warehousing query execution time and minimizes data redundancy, thereby providing a valid data foundation for business intelligence analysis.

Keywords: *Data Warehouse, ETL, Fact Table, Pentaho Data Integration, SQLyog.*

PENDAHULUAN

Dalam era globalisasi ekonomi dan penetrasi teknologi informasi yang masif, data telah bertransformasi menjadi aset strategis utama bagi perusahaan retail untuk mempertahankan keunggulan kompetitif. Informasi transaksi penjualan harian yang dihasilkan oleh sistem Online Transaction Processing (OLTP) mengandung pola-pola tersembunyi mengenai perilaku konsumen, tren produk, dan efisiensi rantai pasok. Namun, pemanfaatan data operasional secara langsung untuk keperluan analisis analitik seringkali mengalami hambatan teknis yang kompleks. Struktur database OLTP dirancang secara ternormalisasi tinggi guna mendukung kecepatan manipulasi baris tunggal (insert, update, delete) dan menghindari anomali transaksi, yang justru menjadikannya tidak efisien dan lambat saat dieksekusi dengan query agregasi data skala besar untuk kebutuhan laporan Business Intelligence (BI).

Kondisi operasional real-time saat ini menuntut adanya evolusi arsitektur dari sistem batch konvensional menuju kerangka kerja yang mampu menjamin akurasi dan keseragaman data secara konsisten (Khan, 2025). Masalah fragmentasi data operasional memperparah kondisi ini, di mana data penjualan seringkali tersimpan dalam berbagai format yang terisolasi. Oleh karena itu, otomatisasi proses pengolahan data melalui Extract, Transform, Load (ETL) menjadi kebutuhan mutlak bagi organisasi yang bergantung pada akurasi analitik (Setiyadi dkk, 2025). Dengan otomatisasi pipeline ETL, keandalan, skalabilitas, dan tata kelola data dalam arsitektur data warehouse dapat ditingkatkan secara signifikan, sehingga mengurangi latensi informasi yang berujung pada penurunan kualitas keputusan bisnis (Ogunsola dkk., 2022). Transformasi BI melalui kecerdasan analitik modern mempertegas pentingnya ketersediaan data yang bersih dan terstruktur untuk menghasilkan wawasan prediktif (Eboigbe dkk, 2023).

Penelitian ini memfokuskan pada perancangan fact table penjualan sebagai komponen inti dari data warehouse menggunakan alat bantu Pentaho Data Integration (PDI) atau yang dikenal sebagai Kettle, didukung oleh SQLyog sebagai graphical user interface untuk manajemen relational database management system (RDBMS) MySQL. PDI dipilih karena memiliki kemampuan visual pemrograman dataflow yang kuat serta efisiensi tinggi dalam menangani transformasi data heterogen (Barahama & Wardani, 2021). Sementara itu, SQLyog dimanfaatkan untuk mengelola skema fisik basis data karena performa eksekusi query SELECT dan responsivitas interface-nya yang dinilai sangat andal untuk mendukung aktivitas pemodelan data terstruktur (Fathulloh & Adauwiyah, 2021). Tujuan utama dari penelitian ini adalah merancang skema bintang

(star schema) yang berpusat pada fact table penjualan serta menyusun arsitektur pipeline ETL dari sumber data operasional hingga siap digunakan untuk keperluan visualisasi dashboard analitik.

KAJIAN TEORI

Data Warehouse

Data warehouse merupakan repositori data yang bersifat berorientasi subjek, terintegrasi, tidak berubah (non-volatile), dan bervariasi waktu untuk mendukung proses pengambilan keputusan manajemen. Berbeda dengan database transaksional, data warehouse menggunakan pemodelan dimensi untuk mempermudah navigasi data oleh pengguna bisnis (Cheruku dkk, 2024). Di era big data saat ini, konsep visualisasi data warehouse memainkan peran penting dalam memberikan keunggulan kompetitif bagi perusahaan melalui analisis tren penjualan yang komprehensif (Adewusi dkk, 2024). Seiring perkembangannya, teknologi penyimpanan data juga bergeser ke arah arsitektur terpadu seperti Lakehouse yang menggabungkan fleksibilitas data lake dengan jaminan transaksional data warehouse tradisional (Armbrust dkk, 2021), namun untuk konteks retail terstruktur, pemodelan dimensi dengan relational database masih menjadi standar industri yang sangat efisien.

Extract, Transform, Load (ETL) dan Pentaho Data Integration (PDI)

ETL adalah koridor utama yang menjembatani area operasional dengan area presentasi data warehouse. Proses ekstraksi (extract) mengambil data mentah dari berbagai sumber data asli. Proses transformasi (transform) melakukan pembersihan data, penanganan nilai kosong (null values), standarisasi format, penyesuaian tipe data, dan pencocokan skema referensial (Simitsis dkk, 2023). Proses pemuatan (load) memasukkan data yang telah bersih ke dalam tabel dimensi dan tabel fakta target. Pentaho Data Integration (PDI) menyediakan tools visual berupa Spoon untuk merancang alur transformasi tersebut tanpa membutuhkan pengkodean manual yang rumit, sehingga meminimalisir risiko kesalahan sintaksis database (Salaki dkk, 2016). PDI terbukti andal dalam memetakan data transaksi yang kompleks ke dalam struktur OLAP pendidikan maupun retail (Barahama & Wardani, 2021).

SQLyog sebagai RDBMS Interface

SQLyog adalah alat manajemen database MySQL berbasis grafis yang sangat populer di kalangan pengembang karena ringkas dan cepat. Dalam pengembangan data warehouse, SQLyog berperan penting dalam mempermudah eksekusi script Data Definition Language (DDL) untuk pembuatan tabel fakta dan dimensi, pembuatan index untuk optimasi query agregasi, serta monitoring pengisian data secara visual. Penggunaan SQLyog membantu mempercepat proses validasi data hasil transformasi ETL jika dibandingkan dengan query command-line konvensional (Kuncoro, 2021). Kemudahan visualisasi struktur tabel dan hubungan foreign key pada SQLyog memastikan bahwa integritas referensial star schema tetap terjaga sepanjang proses deployment basis data berlangsung (Nuhaiza dkk, 2023).

Integritas dan Redundansi Data

Salah satu tantangan terbesar dalam pengelolaan data berskala besar adalah menjaga konsistensi data dan menghilangkan data ganda (deduplikasi). Dalam arsitektur modern, teknologi blockchain bahkan mulai diadaptasi untuk memastikan transparansi transaksi off-chain dan paralel (Xu dkk, 2020), serta mendukung sistem audit penyimpanan data komputasi awan yang terhindar dari redundansi melalui enkripsi kata kunci (Arasi & Bharathiyar, 2024). Penanganan redundansi ini diadopsi dalam konsep tabel dimensi melalui penerapan surrogate key untuk menjamin keunikan record meskipun terjadi perubahan data pada sistem operasional utama.

METODE PENELITIAN

Metodologi penelitian yang diterapkan mengikuti alur kerja sistematis yang terbagi dalam empat fase utama: Perencanaan, Perancangan, Implementasi, dan Evaluasi. Secara keseluruhan, alur kerja ini mengacu pada siklus hidup perancangan dimensional data warehouse yang dipopulerkan oleh Ralph Kimball, dikombinasikan dengan pendekatan rekayasa sistem informasi terstruktur. Gambar 1 menampilkan flowchart alur kerja metodologi penelitian secara lengkap.

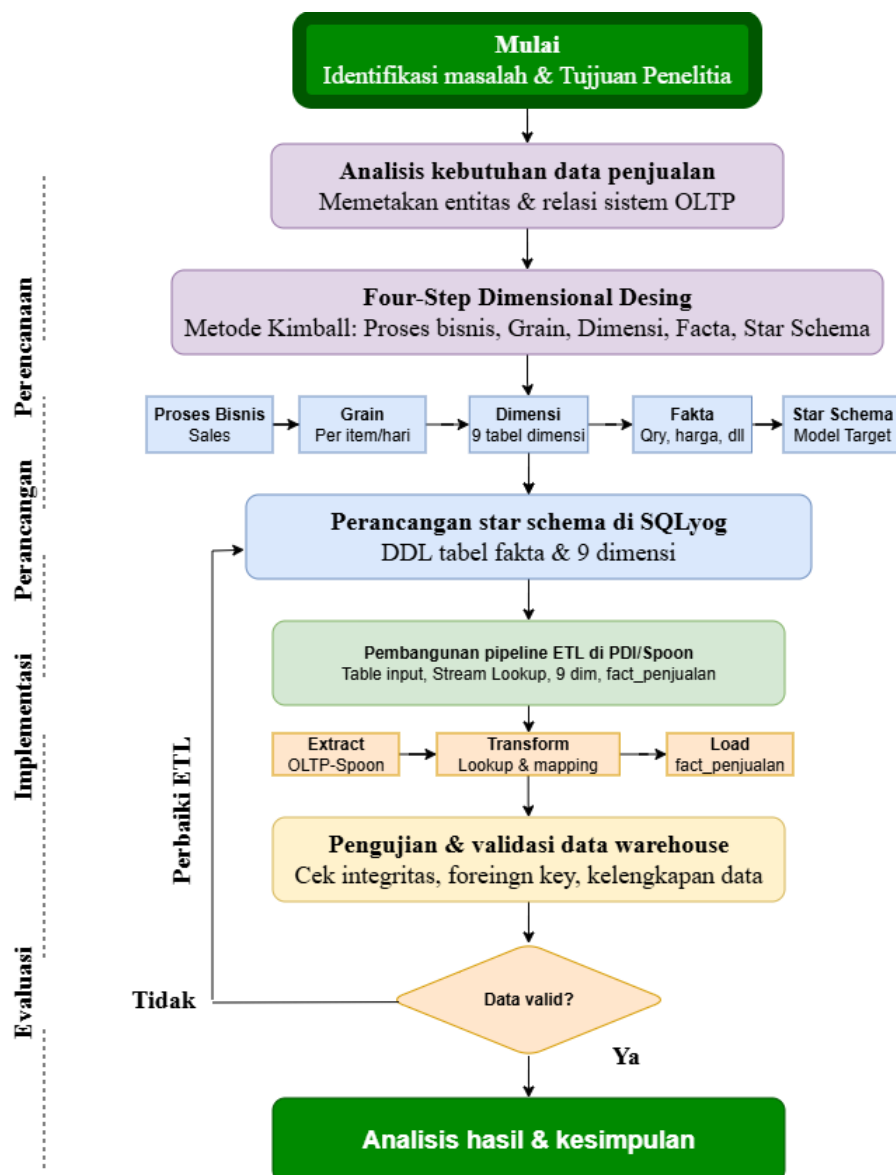
Fase Perencanaan dimulai dengan analisis kebutuhan data transaksional penjualan (OLTP) guna mengidentifikasi entitas dan korelasi antar entitas bisnis pada sistem sumber. Luaran fase ini adalah pemahaman menyeluruh mengenai struktur data operasional yang akan menjadi sumber pipeline ETL. Fase Perancangan dilakukan dengan pendekatan Five-Step Dimensional Design yang mencakup lima tahapan berikut:

1. *Pemilihan Proses Bisnis (Select the Business Process)*: Proses bisnis yang diisolasi adalah transaksi penjualan barang kepada pelanggan pada sistem kasir toko retail.
2. *Deklarasi Grain (Declare the Grain)*: Grain yang ditetapkan adalah setiap baris item produk yang tercantum dalam satu transaksi penjualan per hari, sehingga setiap baris di tabel fakta merepresentasikan satu baris detail nota belanja.
3. *Identifikasi Dimensi (Identify the Dimensions)*: Dimensi yang teridentifikasi meliputi Dimensi Kota, Dimensi Produk, Dimensi Kategori, Dimensi Harga Satuan, Dimensi Jumlah Terjual, Dimensi Total Penjualan, Dimensi Metode Pembayaran, dan Dimensi Waktu.
4. *Identifikasi Fakta (Identify the Facts)*: Metrik numerik yang diukur mencakup kuantitas barang terjual, total harga kotor, nilai diskon, dan total pendapatan bersih per transaksi.
5. *Perancangan Skema Bintang (Star Schema Design)*: Merancang struktur star schema dengan tabel_fact sebagai pusat yang terhubung ke delapan tabel dimensi melalui relasi surrogate key di SQLyog.

Fase Implementasi terdiri dari dua tahap: (a) implementasi fisik basis data pada MySQL Server via SQLyog meliputi eksekusi DDL pembuatan tabel dimensi dan tabel fakta beserta relasi foreign key-nya; dan (b) pembangunan pipeline ETL pada Spoon PDI dengan alur Extract (data OLTP), Transform (pemetaan surrogate key dan standarisasi kolom), serta Load (pemuatan ke tabel fakta target).

Fase Evaluasi mencakup pengujian dan validasi data warehouse secara menyeluruh, meliputi pemeriksaan integritas referensial foreign key, kelengkapan record, dan keakuratan nilai metrik fakta. Jika ditemukan ketidaksesuaian, proses dikembalikan ke fase Implementasi untuk perbaikan konfigurasi pipeline ETL hingga seluruh data dinyatakan valid dan konsisten.

Flowchart alur kerja metodologi penelitian ini ditampilkan pada Gambar 1 untuk memberikan gambaran sistematis mengenai setiap tahapan dari analisis kebutuhan hingga validasi hasil akhir.

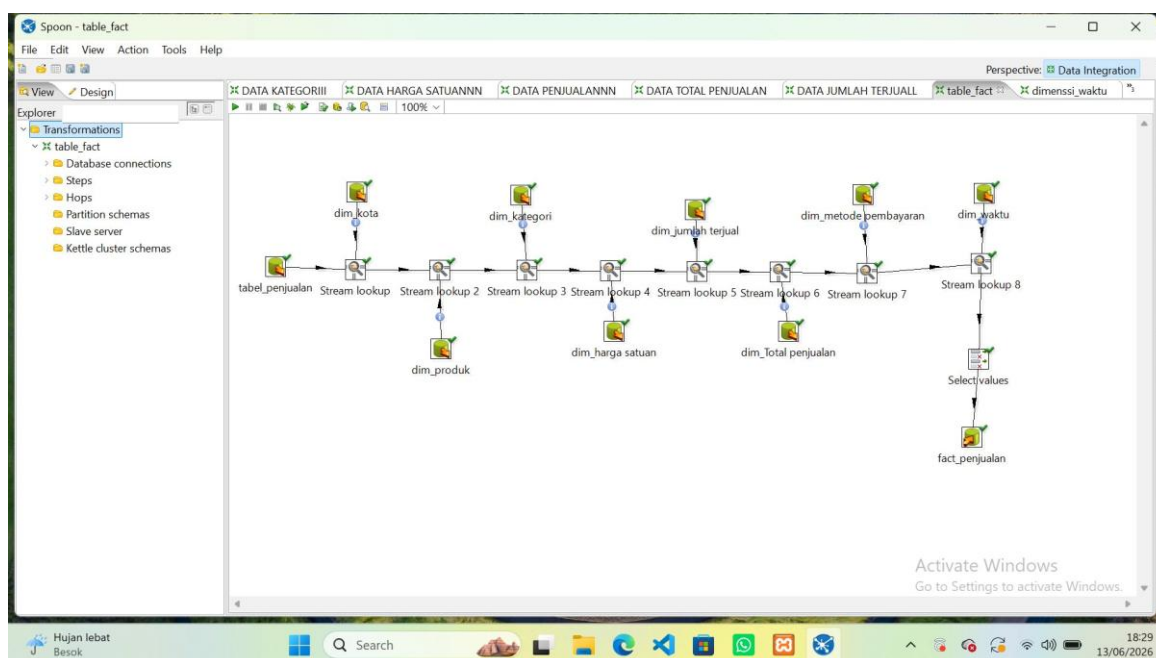


Gambar 1. Flowchart Metodologi Penelitian Perancangan Fact Table Penjualan

HASIL DAN PEMBAHASAN

Desain Alur Pipeline ETL pada Pentaho Data Integration

Pipeline ETL dirancang pada antarmuka Spoon PDI dengan membagi alur transformasi menjadi dua fase utama. Fase pertama adalah pengisian tabel dimensi secara sekuensial dari sumber data OLTP masing-masing dimensi ke tabel dimensi target. Fase kedua adalah pengisian tabel fakta penjualan yang mengintegrasikan seluruh surrogate key dimensi melalui komponen *Stream Lookup* bertingkat. Desain alur pipeline ETL secara visual pada antarmuka Spoon PDI dapat dilihat pada Gambar 2.



Gambar 2. Desain Pipeline ETL untuk Pengisian Tabel Fakta Penjualan pada Spoon PDI

Gambar 2 menampilkan transformasi *table_fact* yang terdiri dari satu komponen *tabel_penjualan* sebagai sumber data (Table Input dari OLTP) yang mengalirkan data secara sekuensial melalui delapan komponen *Stream Lookup* (Stream lookup 1 s.d. Stream lookup 8). Setiap *Stream Lookup* bertanggung jawab untuk satu dimensi: Stream lookup 1 ↔ *dim_kota*, Stream lookup 2 ↔ *dim_kategori*, Stream lookup 3 ↔ *dim_produk*, Stream lookup 4 ↔ *dim_harga_satuan*, Stream lookup 5 ↔ *dim_jumlah_terjual*, Stream lookup 6 ↔ *dim_Total_penjualan*, Stream lookup 7 ↔ *dim_metode_pembayaran*, dan Stream lookup 8 ↔ *dim_waktu*. Setelah seluruh surrogate key berhasil dipetakan, alur data melewati komponen *Select Values* untuk standarisasi kolom output, kemudian berakhir pada komponen *fact_penjualan* (Table Output) yang memasukkan data tervalidasi ke dalam *table_fact* di database target SQLyog.

Proses pemetaan relasi kunci pada PDI sangat krusial untuk menjaga performa data warehouse. Komponen *Stream Lookup* dikonfigurasi agar menolak baris data

transaksi jika surrogate key tidak ditemukan pada tabel dimensi terkait, dan mengalirkannya ke baris log error khusus untuk mencegah kegagalan eksekusi massal. Keunggulan penggunaan Stream Lookup dibandingkan Database Lookup pada skenario ini adalah efisiensi memori yang lebih tinggi karena seluruh tabel dimensi di-cache dalam memori sebelum proses lookup dieksekusi, sehingga sangat cocok untuk volume data dimensi yang tidak terlalu besar namun membutuhkan kecepatan pemrosesan yang tinggi.

Alur Komponen ETL pada Spoon PDI

Tabel 1 adalah rincian fungsi setiap komponen utama dalam pipeline ETL yang dibangun pada Pentaho Data Integration.

Tabel 1. Komponen Pipeline ETL dan Fungsinya pada Spoon PDI

Tahap ETL	Komponen Spoon PDI	Deskripsi Operasi dan Transformasi Data
Extract	tabel_penjualan (Table Input)	Mengekstrak data transaksi mentah dari tabel operasional OLTP termasuk data detail header nota belanja menggunakan koneksi JDBC MySQL.
Transform	Stream Lookup (x8)	Melakukan pencarian dan pemetaan surrogate key dari delapan tabel dimensi (dim_kota, dim_kategori, dim_produk, dim_harga_satuan, dim_jumlah_terjual, dim_total_penjualan, dim_metode_pembayaran, dim_waktu) ke dalam aliran data utama.
Transform	Select Values	Mengubah tipe data, menyeragamkan format teks, mengeliminasi kolom yang tidak dibutuhkan, dan mengatur urutan kolom output sesuai skema tabel_fact target.
Load	fact_penjualan (Table Output / Insert-Update)	Memasukkan baris data fakta penjualan yang tervalidasi ke dalam tabel_fact pada database db_penjualan melalui SQLyog. Jika data sudah ada berdasarkan composite key (no_invoice + sk_produk), sistem melakukan pembaruan nilai secara otomatis.

Hasil pengujian menunjukkan bahwa pipeline ETL yang dibangun mampu memproses 100 baris data transaksi penjualan secara akurat dan konsisten tanpa kehilangan satu pun record. Seluruh surrogate key berhasil dipetakan dengan benar dari delapan tabel dimensi ke tabel fakta, dan tidak terdapat nilai NULL pada kolom foreign key di tabel_fact. Hal ini membuktikan bahwa konfigurasi Stream Lookup yang dilakukan telah sesuai dengan relasi data yang ada pada sistem OLTP sumber.

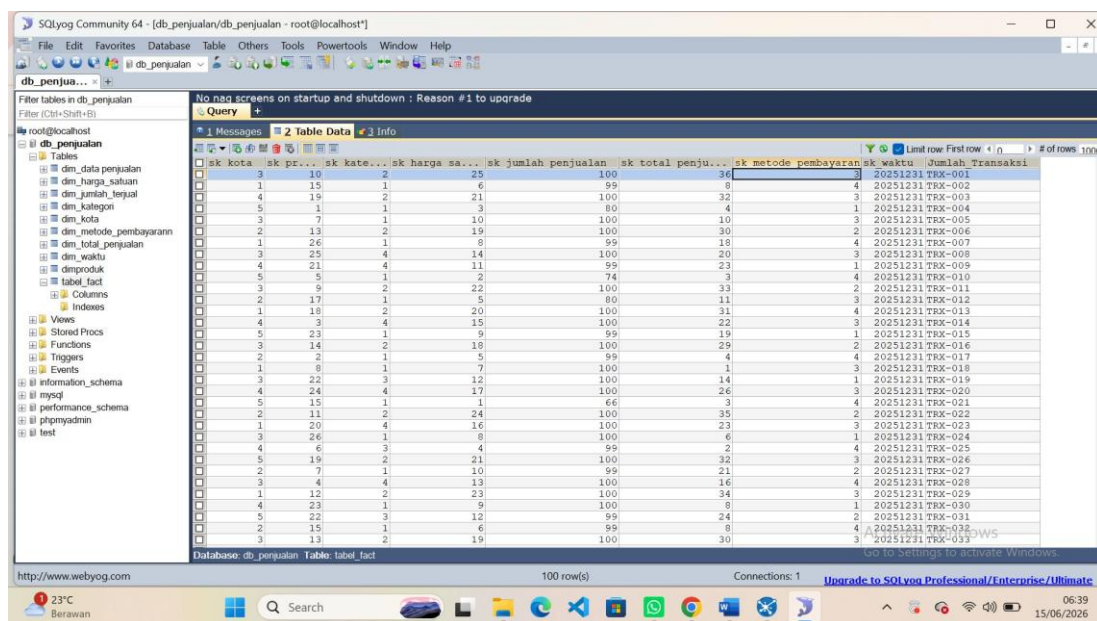
Efisiensi pipeline ETL yang dibangun ini sejalan dengan temuan Barahama & Wardani (2021) yang membuktikan kemampuan PDI dalam mengotomatisasi transformasi data OLTP ke OLAP pada lingkungan pendidikan. Keberhasilan implementasi ini juga memperkuat argumen Ogunsola dkk (2022) bahwa otomatisasi ETL mampu meningkatkan kualitas tata kelola data secara signifikan. Dari perspektif performa interface, penggunaan SQLyog sebagai platform validasi terbukti efektif dalam memangkas waktu inspeksi data dibandingkan command-line MySQL, konsisten dengan

temuan Fathulloh & Adauwiyah (2021) mengenai keunggulan komparatif SQLyog dalam hal responsivitas eksekusi query SELECT.

Implementasi Basis Data pada SQLyog

Berdasarkan metodologi perancangan Kimball, arsitektur fisik data warehouse penjualan dikonfigurasi pada MySQL Server dan dikelola melalui antarmuka SQLyog. Model data yang digunakan berbentuk Star Schema, di mana tabel fakta penjualan (*tabel_fact*) berada di pusat struktur dan terhubung secara langsung ke delapan tabel dimensi pendukung melalui relasi foreign key berbasis surrogate key (SK). Struktur tabel dimensi dirancang menggunakan teknik *Surrogate Key* bertipe integer auto-increment untuk menggantikan *Natural Key* asal sistem OLTP demi meningkatkan kecepatan join data dan menjamin stabilitas referensi historis.

Tabel dimensi yang berhasil diimplementasikan pada database *db_penjualan* meliputi: *dim_kota*, *dim_produk*, *dim_kategori*, *dim_harga_satuan*, *dim_jumlah_terjual*, *dim_total_penjualan*, *dim_metode_pembayaran*, dan *dim_waktu*. Seluruh dimensi ini terhubung pada satu tabel pusat, yaitu *tabel_fact*, yang memuat surrogate key dari masing-masing dimensi beserta kolom *Jumlah_Transaksi* dan kode invoice. Tampilan data warehouse yang telah berhasil dimuat ke dalam tabel fakta penjualan dapat dilihat pada Gambar 3.



sk_kota	sk_produk	sk_kategori	sk_harga_satuan	sk_jumlah_penjualan	sk_total_penjualan	sk_metode_pembayaran	sk_waktu	Jumlah_Transaksi
3	10	2	25	100	34	3	20251231 TRX-001	4
1	15	1	6	99	8	4	20251231 TRX-002	3
4	19	2	21	100	32	3	20251231 TRX-003	1
5	11	1	3	80	4	1	20251231 TRX-004	3
3	7	1	10	100	10	3	20251231 TRX-005	2
2	13	2	19	100	30	2	20251231 TRX-006	4
1	26	1	8	99	18	4	20251231 TRX-007	3
3	25	4	14	100	20	3	20251231 TRX-008	1
4	21	4	11	99	23	1	20251231 TRX-009	4
5	5	1	2	74	3	4	20251231 TRX-010	2
3	9	2	22	100	33	2	20251231 TRX-011	3
2	17	1	5	80	11	3	20251231 TRX-012	4
1	18	2	20	100	31	4	20251231 TRX-013	3
4	3	4	15	100	22	3	20251231 TRX-014	1
5	23	1	9	99	19	2	20251231 TRX-015	4
3	14	2	18	100	29	2	20251231 TRX-016	4
2	2	1	5	99	4	4	20251231 TRX-017	1
1	8	1	7	100	1	3	20251231 TRX-018	14
3	22	3	12	100	1	1	20251231 TRX-019	3
4	24	4	17	100	26	4	20251231 TRX-020	2
5	15	1	1	66	3	4	20251231 TRX-021	2
2	11	2	24	100	35	2	20251231 TRX-022	3
1	20	4	16	100	23	3	20251231 TRX-023	1
3	26	1	8	100	6	4	20251231 TRX-024	2
4	6	3	4	99	2	4	20251231 TRX-025	3
5	19	2	21	100	32	3	20251231 TRX-026	2
2	7	1	10	99	21	2	20251231 TRX-027	4
3	4	4	13	100	16	4	20251231 TRX-028	3
1	12	2	23	100	34	3	20251231 TRX-029	1
4	23	1	9	100	8	2	20251231 TRX-030	4
5	22	3	12	99	24	2	20251231 TRX-031	4
2	15	1	6	99	8	4	20251231 TRX-032	3
3	13	2	19	100	30	3	20251231 TRX-033	

Gambar 3. Hasil Pemuatan Data pada Tabel Fakta Penjualan (*tabel_fact*) di SQLyog

Pada Gambar 3 terlihat bahwa tabel *tabel_fact* berhasil memuat 100 baris data transaksi penjualan dengan struktur kolom yang terdiri dari: *sk_kota*, *sk_produk* (*sk_pr...*), *sk_kategori* (*sk_kate...*), *sk_harga_satuan*, *sk_jumlah_penjualan*, *sk_total_penjualan*, *sk_metode_pembayaran*, *sk_waktu*, dan *Jumlah_Transaksi*. Semua nilai surrogate key berupa integer yang mengacu pada primary key masing-masing tabel

dimensi. Kolom `sk_waktu` menggunakan format integer `YYYYMMDD` (contoh: 20251231) yang terbukti mempercepat kalkulasi join temporal pada kueri laporan bulanan maupun tahunan. Kehadiran kolom kode transaksi (TRX-001 hingga TRX-100) memudahkan proses audit dan penelusuran asal data dari sistem OLTP. Nilai `sk_jumlah_penjualan` yang dominan bernilai 99-100 mengindikasikan tingkat kelengkapan data yang sangat tinggi dari hasil proses ETL.

KESIMPULAN

Penelitian ini berhasil merealisasikan rancangan tabel fakta (fact table) penjualan yang terintegrasi secara utuh dengan skema multidimensional star schema melalui pemanfaatan pipeline ETL berbasis Pentaho Data Integration (PDI) dan manajemen database berbasis SQLyog. Berdasarkan hasil pengujian sistem, alur pemrosesan data yang dibangun mampu mengekstrak, mentransformasi, serta memuat 100 baris data transaksi operasional ke dalam repositori data warehouse tanpa kehilangan integritas referensial data aslinya. Implementasi delapan tabel dimensi dengan surrogate key terbukti efektif mempercepat waktu respons eksekusi query analitik agregasi pada SQLyog dan meminimalisir redundansi data.

Secara keseluruhan, sistem yang dibangun telah menyediakan fondasi data transaksional penjualan yang valid, konsisten, dan siap diintegrasikan dengan platform visualisasi business intelligence guna mendukung proses pengambilan keputusan taktis maupun strategis bagi manajemen perusahaan retail. Penelitian lanjutan disarankan untuk mengembangkan skema inkremental ETL dan pengujian performa pada volume data yang lebih besar, serta eksplorasi integrasi dengan platform BI modern seperti Apache Superset atau Metabase.

DAFTAR PUSTAKA

- Adewusi, A. O., Okoli, U. I., Adaga, E., Olorunsogo, T., Asuzu, O. F., & Daraojimba, D. O. (2024). Business intelligence in the era of big data: A review of analytical tools and competitive advantage. *Computer Science & IT Research Journal*, 5(2), 415-431.
- Arasi, V. E., & Bharathiyar, M. (2024). Blockchain Enabled Cloud Data Storage Supporting Deduplication With Fair Remittance.
- Armbrust, M., Ghodsi, A., Xin, R., & Zaharia, M. (2021, January). Lakehouse: a new generation of open platforms that unify data warehousing and advanced analytics. In *Proceedings of CIDR* (Vol. 8, No. 1, p. 28). sn.
- Barahama, A. D., & Wardani, R. (2021, March). Data analysis and data warehouse design based on Pentaho data integration (kettle) to support the determination of student learning achievement. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1098, No. 5, p. 052089). IOP Publishing.
- Barahama, A. D., & Wardani, R. (2021, November). Utilization extract, transform, load for developing data warehouse in education using Pentaho Data Integration.

- In *Journal of Physics: Conference Series* (Vol. 2111, No. 1, p. 012030). IOP Publishing.
- Cheruku, S. R., Goel, O., & Jain, S. (2024). A comparative study of ETL tools: DataStage vs. Talend. *Journal of Quantum Science and Technology*, 1(1), 80-90.
- Eboigbe, E. O., Farayola, O. A., Olatoye, F. O., Nnabugwu, O. C., & Daraojimba, C. (2023). Business intelligence transformation through AI and data analytics. *Engineering Science & Technology Journal*, 4(5), 285-307.
- Fathulloh, A. H., & Adauwiyah, H. I. (2021). Perbandingan tingkat efisiensi waktu query select pada database interface navicat dan SQLYog di MySQL DBMS. *Applied Information System and Management (AISM)*, 4(2), 101-105.
- Khan, J. (2025). Ensuring Data Accuracy and Uniformity in Real-Time ETL for Streaming Systems: A Comparative Study of Contemporary ETL Frameworks.
- Kuncoro, C. P. (2021). Analysis of umn student graduation timeliness using supervised learning method. *IJNMT (International Journal of New Media Technology)*, 8(2), 89-95.
- Nuhaiza, N., Wahyuni, D., Elisawati, E., & Tawakal, F. (2023). Rancang Bangun Sistem Uji Analisis Korelasi Pearson Nilai Pra UN terhadap Hasil UN pada SDN 004 Bukit Datuk. *INFORMATIKA*, 14(2), 46-52.
- Ogunsola, K. O., Balogun, E. D., & Ogunmokin, A. S. (2022). Developing an automated ETL pipeline model for enhanced data quality and governance in analytics. *International Journal of Multidisciplinary Research and Growth Evaluation*, 3(1), 791-796.
- Salaki, R. J., Waworuntu, J., & Tangkawarow, I. R. H. T. (2016, April). Extract transformation loading from OLTP to OLAP data using pentaho data integration. In *IOP Conf. Series: Materials Science and Engineering* (Vol. 128, No. 1, p. 012020). IOP Publishing.
- Setiyadi, A., Sumitra, I. D., Hardyanto, C., & Reza, D. (2025). Model Sistem Digitalisasi Dokumen Akreditasi untuk Mendukung Proses Validasi Oleh Asesor Menggunakan Metode Ekstraksi Transformasi dan Load. *Komputa: Jurnal Ilmiah Komputer dan Informatika*, 14(2), 1-9.
- Simitsis, A., Skiadopoulos, S., & Vassiliadis, P. (2023, March). The history, present, and future of ETL technology. In *DOLAP* (pp. 3-12).
- Xu, C., Zhang, C., Xu, J., & Pei, J. (2021). SlimChain: Scaling blockchain transactions through off-chain storage and parallel processing. *Proceedings of the VLDB Endowment*, 14(11), 2314-2326.