

IMPLEMENTASI ETL UNTUK MENGHASILKAN FACT TABLE PENJUALAN PADA DATA WAREHOUSE MENGGUNAKAN TOOLS PENTAHO DATA INTEGRATION

Desri Natalis Lasa¹, Ameiliani Dianira Barahama²,
Irene Realyta Halldy Trosi Tangkawarow³

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

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

³irene.tangkawarow@unima.ac.id

ABSTRAK

Pertumbuhan volume data transaksi penjualan mendorong kebutuhan sistem pengelolaan data yang mampu mengintegrasikan berbagai sumber ke dalam satu repositori terpusat untuk analisis bisnis. Penelitian ini bertujuan menerapkan mekanisme Extract, Transform, Load (ETL) serta membangun fact table penjualan dalam lingkungan data warehouse menggunakan Pentaho Data Integration (PDI). Tahapan pengembangan mencakup enam tahapan: (1) pengumpulan data transaksi penjualan, (2) perancangan skema data warehouse berbasis star schema, (3) ekstraksi data dari source database menggunakan komponen Table Input, (4) transformasi data melalui mekanisme Sort Rows, Select Values, dan Stream Lookup untuk membangun delapan tabel dimensi serta fact table, (5) pemuatan data ke dalam dimension table dan fact table menggunakan Table Output, serta (6) pengujian dan validasi konsistensi data antara source database dan data warehouse. Hasil implementasi menunjukkan bahwa proses ETL menggunakan PDI berhasil membangun delapan tabel dimensi, yaitu dimensi produk, kategori, kota, harga satuan, jumlah terjual, total penjualan, metode pembayaran, dan waktu, serta satu fact table penjualan yang mengintegrasikan seluruh dimensi melalui mekanisme Stream Lookup dengan pendekatan star schema. Berdasarkan temuan tersebut, disimpulkan bahwa penerapan ETL dan fact table penjualan menggunakan Pentaho Data Integration terbukti mampu mengoptimalkan tata kelola data sekaligus membangun infrastruktur data warehouse yang andal sebagai landasan pelaporan dan analisis data penjualan.

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

ABSTRACT

The growing volume of sales transaction data drives the need for a data management system capable of integrating multiple sources into a centralized repository for business analysis. This study aims to implement an Extract, Transform, Load (ETL) mechanism and construct a sales fact table within a data warehouse environment using Pentaho Data Integration (PDI). The development process encompasses six phases: (1) sales transaction data collection, (2) data warehouse schema design based on a star schema, (3) data extraction from the source database using the Table Input component, (4) data transformation through the Sort Rows, Select Values, and Stream Lookup mechanism to build eight dimension tables and a fact table, (5) loading data into the dimension table and fact table using the Table Output component, and (6) testing and validation of data consistency between the source database and the data warehouse. The implementation results show that the ETL process using PDI successfully built eight dimension tables, namely product, category, city, unit price, quantity sold, total sales, payment method, and time, as well as one sales fact table that integrates all dimensions through the Stream Lookup mechanism using the star schema approach. Based on these findings, it is concluded that the application of ETL and sales fact table using Pentaho Data Integration is proven to be able to optimize data management while also building a data warehouse infrastructure that is reliable as a basis for reporting and sales data analysis.

(4) data transformation through Sort Rows, Select Values, and Stream Lookup mechanisms to build eight dimension tables and a fact table, (5) data loading into each dimension table and fact table using Table Output, and (6) testing and validation of data consistency between the source database and the data warehouse. Implementation results demonstrate that the ETL process using PDI successfully constructed eight dimension tables, namely product, category, city, unit price, quantity sold, total sales, payment method, and time dimensions, along with one sales fact table integrating all dimensions through the Stream Lookup mechanism with a star schema approach. Based on these findings, it is concluded that the implementation of ETL and a sales fact table using Pentaho Data Integration proves capable of optimizing data governance while establishing a reliable data warehouse infrastructure as the foundation for sales data reporting and analysis.

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

PENDAHULUAN

Perkembangan teknologi informasi yang semakin cepat menuntut organisasi dan perusahaan untuk mengelola data dalam jumlah besar secara lebih efisien guna mendukung proses pengambilan keputusan. Dalam lingkungan bisnis modern, khususnya sektor perbankan, pertumbuhan volume data yang sangat pesat membuat penggunaan basis data tradisional menjadi kurang memadai. Oleh karena itu, diperlukan solusi yang mampu menangani penyimpanan dan analisis data secara lebih optimal. Penerapan data warehouse (DWH) dan data mining kini semakin luas karena mampu menyediakan dukungan analitis terhadap data berskala besar yang sulit diproses menggunakan sistem konvensional (Rohajawati, 2024).

Data warehouse memiliki peran penting sebagai bagian dari infrastruktur data strategis dalam suatu perusahaan. Konsep data warehouse merujuk pada penggabungan data yang berasal dari berbagai sumber ke dalam satu repositori terpusat yang digunakan untuk kebutuhan querying dan reporting. Secara konseptual, data di dalam data warehouse direpresentasikan secara logis melalui tabel dimensi dan tabel fakta (Novandya, 2021). Dalam implementasinya, fact table berfungsi sebagai penyimpan data transaksi, seperti data penjualan, yang kemudian dikaitkan dengan dimension table melalui pendekatan pemodelan dimensional seperti star schema guna menghasilkan analisis yang lebih efektif.

Pengelolaan dan pengisian data pada data warehouse dilakukan melalui proses yang dikenal sebagai ETL (Extract, Transform, Load). ETL merupakan metode yang digunakan untuk mengambil data dari berbagai sumber, melakukan penyesuaian atau transformasi sesuai kebutuhan, kemudian memuatnya ke dalam tabel tujuan yang telah disediakan (Andriansyah, 2022). Proses ini memiliki peranan yang sangat penting karena data yang awalnya tidak terstruktur atau beragam formatnya akan diolah menjadi data yang lebih konsisten dan terintegrasi. Dengan demikian, kualitas hasil analisis pada data warehouse dapat ditingkatkan sehingga mendukung pengambilan keputusan perusahaan secara lebih tepat (Prasetia & Kurniawan, 2021).

Salah satu perangkat yang banyak digunakan dalam implementasi ETL adalah Pentaho Data Integration (PDI). Pada lingkungan data warehouse penjualan, Pentaho Data Integration (PDI) digunakan untuk menjalankan proses ekstraksi data dari sumber, melakukan transformasi agar data memiliki format yang seragam dan sesuai dengan kebutuhan analisis, serta memuat data tersebut ke dalam media penyimpanan tujuan (Prasetia & Kurniawan, 2021). Keunggulan pendekatan ini terletak pada kemampuannya menyediakan alur kerja ETL secara visual sehingga proses pengembangan dapat dilakukan dengan lebih mudah tanpa memerlukan pemrograman yang rumit. Selain itu, (Barahama & Wardani, 2021a) juga membuktikan bahwa Pentaho dengan komponen kettle efektif digunakan sebagai alat integrasi data dalam lingkungan pendidikan untuk mendukung pengambilan keputusan berbasis data.

Berbagai penelitian terdahulu telah menunjukkan bahwa penerapan data warehouse berbasis ETL memberikan manfaat yang signifikan dalam berbagai bidang. Penelitian yang dilakukan pada Restoran Khayangan Kuliner membuktikan bahwa implementasi data warehouse menggunakan metode ETL dan OLAP dapat membantu proses analisis data penjualan, mengidentifikasi pola penjualan, serta mendukung penyusunan strategi bisnis dan pengambilan keputusan yang lebih baik (Pratama & Permana, 2023). Selain itu, penelitian mengenai transformasi data warehouse dengan pendekatan single source of truth menggunakan alat ETL IBM InfoSphere DataStage menunjukkan bahwa proses ETL dapat diterapkan dan diuji pada Performance Measurement System (PMS) untuk memantau, mengidentifikasi, dan mengevaluasi kinerja bisnis secara lebih akurat (Rohajawati, 2024). (Barahama & Wardani, 2021b) turut mengembangkan implementasi ETL menggunakan Pentaho Data Integration untuk membangun data warehouse di lingkungan pendidikan, yang membuktikan bahwa alur ETL berbasis Pentaho mampu mengintegrasikan data dari berbagai sumber secara sistematis dan akurat.

Walaupun berbagai penelitian sebelumnya telah memberikan kontribusi yang berarti, kajian yang secara khusus menggabungkan perancangan fact table penjualan dengan implementasi proses ETL menggunakan Pentaho Data Integration dalam satu sistem data warehouse yang terintegrasi masih relatif terbatas. Oleh sebab itu, penelitian ini difokuskan pada perancangan serta implementasi proses ETL dan fact table penjualan pada data warehouse dengan memanfaatkan Pentaho Data Integration (PDI). Aspek kebaruan penelitian ini terletak pada integrasi antara desain fact table berbasis skema dimensional dan otomatisasi proses ETL menggunakan Pentaho Data Integration (PDI) yang diterapkan langsung pada data penjualan. Hasilnya diharapkan dapat menghasilkan data warehouse yang siap digunakan sebagai landasan analisis bisnis secara efisien, akurat, dan terstruktur.

KAJIAN TEORI

Extract, Transform, Load (ETL)

ETL (Extract, Transform, Load) adalah serangkaian proses yang digunakan untuk memindahkan data dari sumber ke dalam data warehouse melalui tiga tahapan utama. Tahap pertama adalah Extract, yaitu proses pengambilan data dari berbagai sumber

heterogen seperti database relasional, file flat, atau sistem operasional. Tahap kedua adalah Transform, yaitu proses pembersihan, validasi, standarisasi, dan konversi data sesuai dengan aturan bisnis yang telah ditetapkan. Tahap ketiga adalah Load, yaitu proses pemuatan data hasil transformasi ke dalam struktur target pada data warehouse (Andriansyah, 2022).

Badiuzzaman dkk (2021) mengembangkan teknik ETL yang efisien dengan fokus pada optimasi performa pemrosesan data di data warehouse skala besar. Hasil penelitian menunjukkan bahwa teknik ETL yang tepat mampu mengurangi waktu pemrosesan secara signifikan serta meningkatkan kualitas data yang dimuat ke dalam data warehouse.

Khan dkk (2024) memberikan tinjauan komprehensif mengenai berbagai teknik, alat, proses, dan evaluasi ETL dalam konteks data warehousing. Penelitian tersebut menyimpulkan bahwa pemilihan teknik dan alat ETL yang tepat merupakan faktor kritis yang menentukan keberhasilan implementasi data warehouse, terutama dalam hal efisiensi waktu pemrosesan, skalabilitas, dan kemampuan penanganan error. Selain itu, (Dinesh & Devi, 2024) mengusulkan pendekatan hybrid optimization pada proses ETL untuk arsitektur data warehouse berbasis cloud yang mampu meningkatkan efisiensi pemrosesan sekaligus mengurangi biaya infrastruktur.

Dalam konteks data multi-sumber, (Pérez dkk, 2024) mengimplementasikan proses ETL untuk membangun dashboard diagnosis dan pengobatan epilepsi genetik dari berbagai sumber data. Penelitian tersebut membuktikan bahwa proses ETL yang dirancang dengan baik mampu mengintegrasikan data dari berbagai sumber yang berbeda format dan struktur menjadi satu kesatuan data yang siap dianalisis. Pendekatan serupa dikembangkan oleh (Gueddoudj dkk, 2023) melalui Os-ETL, yaitu solusi ETL berskala besar berbasis open-source yang mampu mengintegrasikan data heterogen secara efisien dalam lingkungan data warehousing skala enterprise.

Fact Table dan Dimensional Modeling

Fact table adalah komponen utama dalam desain dimensional data warehouse yang berfungsi menyimpan data kuantitatif atau ukuran (measures) dari suatu proses bisnis. Setiap baris pada fact table merepresentasikan satu kejadian bisnis yang dapat diukur, seperti transaksi penjualan, dan dihubungkan ke dimension table melalui foreign key. Penggunaan fact constellation schema sebagai pengembangan dari star schema yang memungkinkan berbagi dimensi antar beberapa fact table, sehingga cocok untuk proses bisnis yang lebih kompleks.

(Pratama & Permana, 2023) mengimplementasikan data warehouse menggunakan metode ETL dan OLAP pada data penjualan restoran dengan merancang fact table penjualan yang dilengkapi dimensi waktu, produk, dan pelanggan. Penelitian tersebut membuktikan bahwa desain fact table yang tepat mampu menghasilkan laporan analitik penjualan yang akurat dan mendukung operasi OLAP seperti drill-down, roll-up, dan slice-and-dice secara efektif. Ghita & Trisminingsih (2021) juga menunjukkan bahwa struktur fact table yang terintegrasi dengan dimensi yang relevan mampu mendukung analisis spasial dan temporal secara bersamaan pada data warehouse komoditas pertanian Indonesia.

ETL Menggunakan Pentaho Data Integration

Pentaho Data Integration (PDI) atau Kettle merupakan alat ETL berbasis open-source yang menyediakan antarmuka grafis drag-and-drop untuk memudahkan perancangan dan eksekusi alur proses ETL tanpa memerlukan pemrograman yang kompleks. Pentaho Data Integration (PDI) terdiri dari dua komponen utama, yaitu Transformation yang mendefinisikan alur pemrosesan data langkah demi langkah, dan Job yang mengatur serta menjadwalkan eksekusi satu atau beberapa Transformation secara terstruktur. Pentaho Data Integration (PDI) mendukung koneksi ke berbagai sumber data termasuk MySQL, PostgreSQL, Oracle, file CSV, dan Microsoft Excel (Prasetya & Kurniawan, 2021).

Prasetya & Kurniawan (2021) mengimplementasikan ETL menggunakan Pentaho pada data warehouse penjualan dan membuktikan bahwa Pentaho Data Integration (PDI) mampu memproses data dari database operasional secara otomatis ke dalam struktur data warehouse dengan tingkat akurasi yang tinggi. Penelitian tersebut menunjukkan bahwa fitur-fitur PDI seperti Table Input, Sort Rows, Merge Join, dan Table Output sangat efektif untuk membangun alur ETL penjualan yang lengkap dan dapat diandalkan. Penggunaan Pentaho Data Integration (PDI) terbukti mempercepat proses pengembangan ETL dibandingkan dengan pemrograman manual.

Lelet dkk (2023) juga memanfaatkan Pentaho dalam pengembangan sistem informasi layanan mahasiswa di Universitas Negeri Manado. Penelitian tersebut membuktikan bahwa Pentaho Data Integration (PDI) dapat diterapkan tidak hanya untuk domain penjualan, tetapi juga untuk berbagai domain sistem informasi lainnya dengan hasil integrasi data yang konsisten dan akurat. Lebih lanjut, (Barahama & Wardani, 2021a) menggunakan Pentaho Data Integration dengan komponen Kettle untuk mendukung penentuan keputusan berbasis data di lingkungan pendidikan, yang semakin memperkuat fleksibilitas dan keandalan Pentaho Data Integration (PDI) dalam berbagai konteks implementasi data warehouse.

Data Warehouse

Data warehouse merupakan sistem basis data yang dirancang khusus untuk keperluan analisis dan pengambilan keputusan bisnis dengan mengintegrasikan data dari berbagai sumber ke dalam satu repositori terpusat yang konsisten. Dhaouadi dkk (2022) mendefinisikan data warehouse sebagai infrastruktur penyimpanan data yang berorientasi pada subjek, terintegrasi, bervariasi terhadap waktu, dan bersifat non-volatile, serta membedakannya dari pendekatan klasik menuju tren modern berbasis cloud dan big data. Karakteristik tersebut menjadikan data warehouse sebagai fondasi utama dalam sistem pendukung keputusan di berbagai sektor bisnis maupun industri.

Rohajawati (2024) menegaskan bahwa transformasi digital mendorong evolusi teknologi data warehouse dari arsitektur konvensional menuju pendekatan yang lebih fleksibel dan skalabel. Penelitian tersebut menunjukkan bahwa efektivitas data warehouse sangat bergantung pada kemampuan sistem dalam mengelola volume data yang besar, kecepatan pemrosesan, serta integrasi dengan berbagai platform modern. Hal ini memperkuat pentingnya implementasi data warehouse yang terancang dengan baik sebagai landasan infrastruktur analitik perusahaan.

Implementasi ETL pada Data Warehouse Penjualan

Implementasi ETL pada data warehouse penjualan bertujuan untuk mengintegrasikan data transaksi dari sistem operasional menjadi struktur dimensional yang siap dianalisis secara multidimensi. Andriansyah (2022) mengimplementasikan proses ETL pada data warehouse laporan harian dan menghasilkan struktur data yang memudahkan pembuatan laporan periodik secara otomatis, sehingga menggantikan proses pelaporan manual yang memakan waktu menjadi proses yang lebih efisien dan akurat.

Tantangan utama dalam implementasi ETL adalah memastikan konsistensi dan kualitas data dari berbagai sumber yang heterogen, khususnya dalam mengelola data dari berbagai cabang atau sistem point-of-sale yang berbeda. Oleh karena itu, validasi dan pembersihan data pada tahap Transform menjadi aspek yang sangat kritis untuk memastikan integritas data yang dimuat ke dalam fact table penjualan. Barahama & Wardani (2021b) menegaskan bahwa penerapan ETL secara menyeluruh menggunakan Pentaho pada lingkungan pendidikan mampu menghasilkan data warehouse yang andal sebagai dasar pelaporan dan analisis data, yang prinsipnya sejalan dengan kebutuhan pada data warehouse penjualan.

METODE PENELITIAN

Penelitian ini merupakan penelitian terapan yang berfokus pada implementasi proses ETL (*Extract, Transform, Load*) dan perancangan *fact table* penjualan pada *data warehouse* menggunakan Pentaho Data Integration (PDI). Pendekatan yang digunakan adalah pendekatan kualitatif-deskriptif dengan studi kasus pada sistem data penjualan.

Alat dan Bahan

Alat yang digunakan dalam penelitian ini meliputi:

1. Pentaho Data Integration (PDI) / Kettle versi 5.0 General Availability Release, digunakan sebagai tools utama dalam perancangan dan pelaksanaan proses ETL (*Extract, Transform, Load*).
2. MySQL, digunakan sebagai database sumber (*source database*) dan database tujuan (*data warehouse*) yang dijalankan melalui layanan XAMPP.
3. SQLyog Community versi 13.3 (64 bit) by Webyog Inc., digunakan sebagai antarmuka grafis (*GUI*) untuk pengelolaan database MySQL, perancangan skema *data warehouse*, dan pengujian query SQL.
4. Laptop/PC dengan spesifikasi minimum RAM 8 GB dan sistem operasi Windows 10 64-bit.

Bahan yang digunakan dalam penelitian ini adalah data transaksi penjualan yang mencakup atribut-atribut sebagai berikut: data produk, kategori produk, kota, harga satuan, jumlah terjual, total transaksi, metode pembayaran, dan waktu transaksi. Data tersebut digunakan sebagai sumber (*source data*) yang akan melalui proses ETL untuk kemudian dimuat ke dalam *fact table* penjualan pada *data warehouse*.

Tahapan Pengembangan Data Warehouse

Penelitian ini dilaksanakan melalui beberapa tahapan sebagai berikut:

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data transaksi penjualan yang mencakup atribut produk, kategori produk, kota, harga satuan, jumlah terjual, total transaksi, metode pembayaran, dan waktu transaksi. Data dikumpulkan dan disimpan dalam database sumber (*source database*) menggunakan MySQL yang dijalankan melalui layanan XAMPP dan dikelola menggunakan SQLyog Community versi 13.3.

2. Perancangan Skema Data Warehouse

Perancangan *data warehouse* menggunakan pendekatan skema bintang (*star schema*) yang terdiri dari satu *fact table* dan beberapa *dimension table*. *Dimension table* yang dirancang meliputi *dim_produk*, *dim_kategori*, *dim_kota*, *dim_hargasatuan*, *dim_jumlahterjual*, *dim_totalpenjualan*, *dim_metodepembayaran*, dan *dim_waktu*. Seluruh *dimension table* tersebut berelasi dengan *fact table* utama yaitu *fact_penjualan*.

3. Proses Extract (Ekstraksi Data)

Tahap ekstraksi dilakukan dengan mengambil data dari *source database* MySQL menggunakan komponen *Table Input* pada Pentaho Data Integration. Data transaksi penjualan diekstrak dari tabel sumber (*tabel_penjualan*) secara keseluruhan untuk kemudian masuk ke tahap transformasi.

4. Proses Transform (Transformasi Data)

Transformasi data dilakukan menggunakan Pentaho Data Integration dengan dua pendekatan berdasarkan jenis tabelnya:

- a. Untuk *dimension table*, proses transformasi dilakukan menggunakan komponen *Table Input* sebagai sumber data, *Add Sequence* (Sk produk) untuk menghasilkan *surrogate key*, *Select Values* untuk memilih dan memetakan atribut yang diperlukan, kemudian *Table Output* untuk memuat data ke dalam *dimension table*. Pendekatan ini diterapkan pada seluruh *dimension table* yang dirancang.
- b. Untuk *fact table*, proses transformasi dilakukan dengan mengambil data dari *tabel_penjualan*, kemudian dilakukan serangkaian *Stream Lookup* secara berurutan untuk memetakan *surrogate key* dari setiap *dimension table* yaitu *dim_produk*, *dim_kategori*, *dim_kota*, *dim_hargasatuan*, *dim_jumlahterjual*, *dim_totalpenjualan*, *dim_metodepembayaran*, dan *dim_waktu*. Setelah seluruh *surrogate key* berhasil dipetakan, komponen *Select Values* digunakan untuk memilih kolom yang akan dimuat ke dalam *fact table*.

5. Proses Load (Pemuatan Data)

Data hasil transformasi dimuat ke dalam masing-masing *dimension table* dan *fact table* penjualan menggunakan komponen *Table Output* pada Pentaho Data Integration. Pemuatan data dilakukan secara berurutan, dimulai dari seluruh *dimension table* terlebih dahulu, kemudian dilanjutkan dengan pemuatan ke *fact table* *fact_penjualan* untuk menjaga integritas referensial data.

6. Pengujian dan Validasi

Pengujian dilakukan untuk memastikan data yang telah dimuat ke dalam *data warehouse* sesuai dengan data sumber. Validasi dilakukan dengan membandingkan

jumlah baris data, nilai total transaksi, jumlah terjual, dan konsistensi *surrogate key* antara *source database* dan *data warehouse* menggunakan query SQL pada SQLyog Community versi 13.3.

HASIL DAN PEMBAHASAN

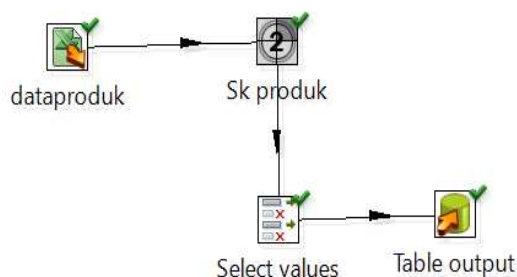
Bagian ini menyajikan hasil implementasi proses ETL (*Extract, Transform, Load*) dan perancangan fact table penjualan pada data warehouse menggunakan Pentaho Data Integration (PDI). Hasil yang diperoleh meliputi desain transformasi ETL untuk setiap tabel dimensi dan struktur fact table penjualan yang berhasil dibangun menggunakan pendekatan *star schema*.

Hasil Implementasi Dimensi

Proses ETL untuk setiap tabel dimensi dirancang menggunakan Pentaho Data Integration dengan alur yang terdiri dari langkah pengambilan data sumber, seleksi dan transformasi kolom, serta pemuatan ke tabel dimensi pada data warehouse. Terdapat delapan tabel dimensi yang berhasil dibangun, yaitu dimensi produk, dimensi kategori, dimensi kota, dimensi harga satuan, dimensi jumlah terjual, dimensi total penjualan, dimensi metode pembayaran, dan dimensi waktu.

Desain Dimensi Produk

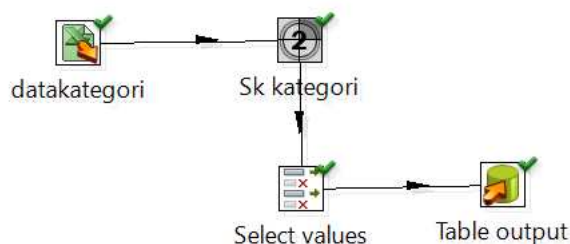
Dimensi produk dibangun untuk menyimpan data master produk yang dijual. Data sumber diambil dari tabel *dataproduk*, kemudian diurutkan menggunakan komponen *Sk produk* (*Sort rows*) berdasarkan kode atau nama produk untuk memastikan keunikan data. Selanjutnya dilakukan seleksi kolom melalui *Select Values* guna memfilter atribut yang relevan, dan data dimuat ke tabel dimensi melalui komponen *Table Output*.



Gambar 1. Dimensi Produk

Desain Dimensi Kategori

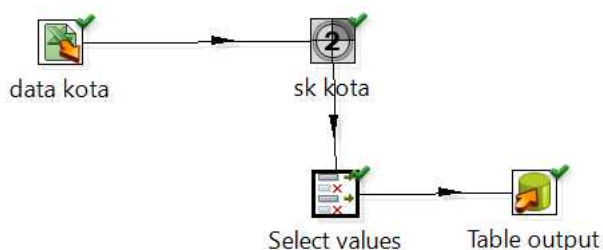
Dimensi kategori berfungsi untuk menyimpan data pengelompokan produk berdasarkan kategorinya. Alur transformasi dimulai dari pembacaan data sumber *datakategori*, kemudian dilakukan pengurutan menggunakan komponen *Sk kategori* (*Sort rows*) berdasarkan kode kategori. Setelah diurutkan, kolom yang dibutuhkan dipilih melalui *Select Values* dan data dimuat ke tabel *dim_kategori* menggunakan *Table Output*.



Gambar 2. Dimensi Kategori

Desain Dimensi Kota

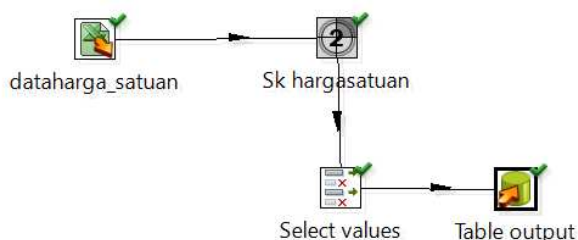
Dimensi kota menyimpan data wilayah atau lokasi penjualan. Data sumber diambil dari tabel *datakota*, selanjutnya diurutkan menggunakan komponen *Sk kota* (*Sort rows*) berdasarkan kode atau nama kota. Proses seleksi atribut dilakukan melalui *Select Values* dan hasil akhirnya dimuat ke tabel *dim_kota* melalui *Table Output*.



Gambar 3. Dimensi Kota

Desain Dimensi Harga Satuan

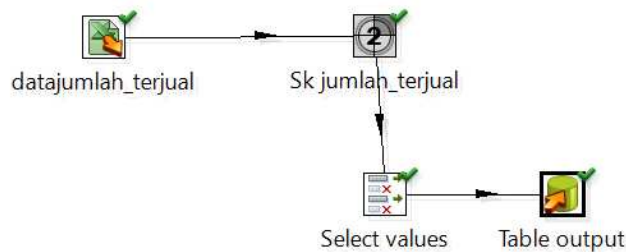
Dimensi harga satuan berfungsi menyimpan informasi harga per unit produk yang dijual. Data sumber diambil dari tabel *datahargasatuan*, kemudian diurutkan menggunakan komponen *Sk hargasatuan* (*Sort rows*) berdasarkan nilai harga atau kode produk. Kolom yang relevan dipilih melalui *Select Values* dan data dimuat ke tabel *dim_hargasatuan* menggunakan *Table Output*.



Gambar 2. Dimensi Harga Satuan

Desain Dimensi Jumlah Terjual

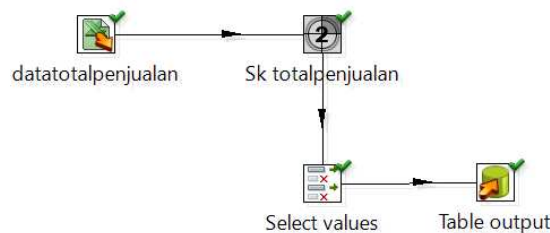
Dimensi jumlah terjual menyimpan data kuantitas produk yang berhasil terjual dalam setiap transaksi. Alur transformasi dimulai dari pembacaan data sumber *datajumlahterjual*, dilanjutkan pengurutan menggunakan komponen *Sk jumlahterjual* (*Sort rows*). Seleksi atribut dilakukan melalui *Select Values* dan data dimuat ke tabel *dim_jumlahterjual* melalui *Table Output*.



Gambar 3. Dimensi Jumlah Terjual

Desain Dimensi Total Penjualan

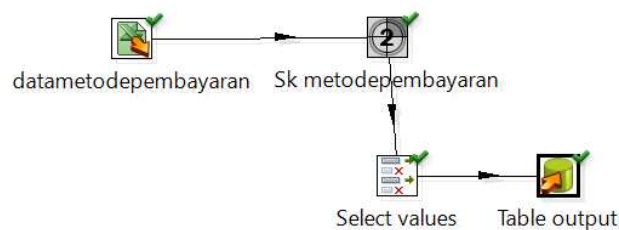
Dimensi total penjualan berfungsi menyimpan data nilai keseluruhan transaksi penjualan. Data sumber diambil dari tabel *datatotalpenjualan*, kemudian diurutkan menggunakan komponen *Sk totalpenjualan* (*Sort rows*) untuk memastikan konsistensi data. Kolom yang dibutuhkan dipilih melalui *Select Values* dan data dimuat ke tabel *dim_totalpenjualan* menggunakan *Table Output*.



Gambar 6. Dimensi Jumlah Total Penjualan

Desain Dimensi Metode Pembayaran

Dimensi metode pembayaran menyimpan informasi jenis pembayaran yang digunakan dalam setiap transaksi, seperti tunai, transfer, atau kartu kredit.

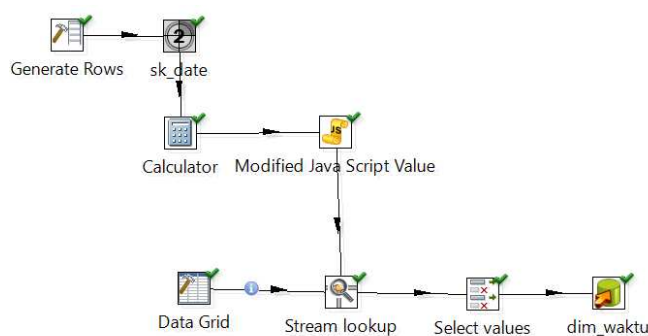


Gambar 7. Dimensi Metode Pembayaran

Data sumber diambil dari tabel *datametodepembayaran*, diurutkan menggunakan komponen *Sk metodepembayaran* (*Sort rows*) berdasarkan kode metode pembayaran. Selanjutnya atribut yang diperlukan dipilih melalui *Select Values* dan dimuat ke tabel *dim_metodepembayaran* melalui *Table Output*.

Desain Dimensi Waktu

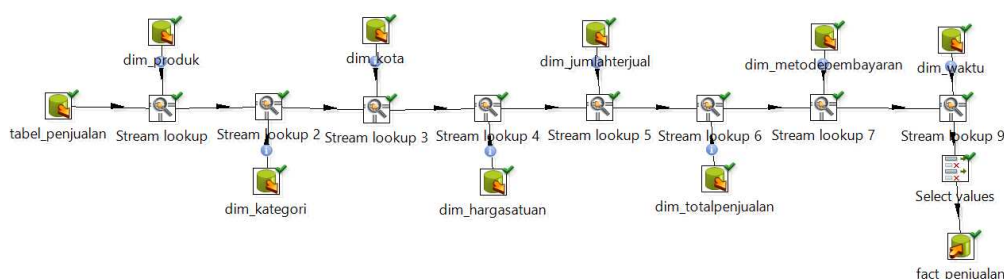
Dimensi waktu menyimpan atribut temporal dari setiap transaksi penjualan seperti tanggal, bulan, kuartal, dan tahun. Data sumber diambil dari tabel *datawaktu*, kemudian diurutkan menggunakan komponen *Sk waktu* (*Sort rows*) berdasarkan urutan tanggal transaksi. Proses seleksi kolom dilakukan melalui *Select Values* dan data dimuat ke tabel *dim_waktu* menggunakan *Table Output*.



Gambar 8. Dimensi Waktu

Hasil Perancangan Fact Table Penjualan

Setelah seluruh tabel dimensi berhasil dibangun, tahap selanjutnya adalah perancangan dan pengisian fact table penjualan. Fact table dirancang menggunakan pendekatan *star schema* yang menghubungkan tabel fakta dengan seluruh tabel dimensi melalui mekanisme *Stream Lookup* pada Pentaho Data Integration.



Gambar 9. Desain Fact Table Penjualan

Gambar 9. menunjukkan alur transformasi ETL untuk pengisian fact table penjualan. Data transaksi penjualan (*tabel penjualan*) digunakan sebagai sumber utama, kemudian dilakukan proses *Stream Lookup* secara berurutan untuk menggabungkan data dari masing-masing dimensi, yaitu dimensi produk, dimensi kategori, dimensi kota,

dimensi harga satuan, dimensi jumlah terjual, dimensi total penjualan, dimensi metode pembayaran, dan dimensi waktu. Setelah semua dimensi berhasil digabungkan, dilakukan seleksi kolom melalui komponen *Select Values* sebelum data akhirnya dimuat ke tabel *fact_penjualan*.

#	Sk produk	Sk kategori	sk kota	Sk hargasatuan	Sk jumlah_terjual	Sk_totalpenjualan	Sk metodepembayaran	sk_waktu	jumlah_transaksi
1	9	2	1	25	1	36	3	20241231	TRX-001
2	15	1	1	6	2	8	4	20241231	TRX-002
3	18	2	2	21	1	32	3	20241231	TRX-003
4	2	1	5	3	3	4	1	20241231	TRX-004
5	7	1	1	10	1	10	3	20241231	TRX-005
6	13	2	2	19	1	30	2	20241231	TRX-006
7	24	1	3	8	2	18	4	20241231	TRX-007
8	24	4	5	14	1	20	3	20241231	TRX-008
9	20	4	5	11	2	23	1	20241231	TRX-009
10	5	1	2	2	4	3	4	20241231	TRX-010
11	10	2	4	22	1	33	2	20241231	TRX-011
12	16	1	2	5	3	11	3	20241231	TRX-012
13	17	2	1	20	1	31	4	20241231	TRX-013
14	3	4	3	15	1	22	3	20241231	TRX-014
15	22	1	3	9	2	19	1	20241231	TRX-015
16	14	2	3	18	1	29	2	20241231	TRX-016
17	1	1	2	5	2	4	4	20241231	TRX-017
18	8	1	5	7	1	1	3	20241231	TRX-018
19	21	3	2	12	1	14	1	20241231	TRX-019
20	23	4	1	17	1	26	3	20241231	TRX-020
21	26	1	1	1	5	3	4	20241231	TRX-021
22	11	2	3	24	1	35	2	20241231	TRX-022
23	19	4	3	16	1	23	3	20241231	TRX-023
24	24	1	4	8	1	6	1	20241231	TRX-024

Gambar 10. Hasil transformasi table_fact

Gambar 10. menunjukkan hasil transformasi fact_table dari dimensi-dimensi yang digunakan sebagai database khusus melalui proses Extract, Transform, Load (ETL) menggunakan tools Pentaho Data Integration-Kettle yang menghasilkan tabel fakta penjualan.

Table 1. Struktur Fact Table Penjualan

Nama Kolom	Tipe Data	Keterangan
sk_produk	INT (FK)	Foreign key yang mengacu pada dimensi produk
sk_kategori	INT (FK)	Foreign key yang mengacu pada dimensi kategori
sk_kota	INT (FK)	Foreign key yang mengacu pada dimensi kota
sk_hargasatuan	INT (FK)	Foreign key yang mengacu pada dimensi harga satuan
sk_jumlah_terjual	INT (FK)	Foreign key yang mengacu pada dimensi jumlah terjual
sk_metodepembayaran	INT (FK)	Foreign key yang mengacu pada dimensi metode pembayaran
sk_totalpenjualan	INT (FK)	Foreign key yang mengacu pada dimensi total penjualan
sk_waktu	BIGINT (FK)	Foreign key yang mengacu pada dimensi waktu
jumlah_transaksi	VARCHAR(11)	Measure yang menyimpan jumlah transaksi penjualan

Tabel 1 Struktur fact table penjualan yang dihasilkan memuat foreign key dari setiap dimensi serta measure berupa data transaksi penjualan yang dapat digunakan untuk keperluan analisis bisnis.



Sk_produk	Sk_kategori	sk_kota	Sk_hargasaduan	Sk_jumlah_terjual	Sk_metodepembayaran	sk_waktu	jumlah_tran
9	2	1	25	1	3	20241231	TRX-001
15	1	1	6	2	4	20241231	TRX-002
18	2	2	21	1	3	20241231	TRX-003
2	1	5	3	3	1	20241231	TRX-004
7	1	1	10	1	3	20241231	TRX-005
13	2	2	19	1	2	20241231	TRX-006
24	1	3	8	2	4	20241231	TRX-007

Gambar 11. Hasil Data pada Tabel fact_penjualan di Database

Gambar 11. menunjukkan hasil data yang telah berhasil dimuat ke dalam tabel fact_penjualan pada database db_penjualan_desri. Data tersebut mencakup seluruh kolom surrogate key yang terhubung dengan tabel dimensi, yaitu Sk_produk, Sk_kategori, Sk_kota, Sk_hargasaduan, Sk_jumlah_terjual, Sk_metodepembayaran, sk_waktu, jumlah_transaksi, dan Sk_totalpenjualan

Hasil implementasi menunjukkan bahwa Pentaho Data Integration mampu menangani proses ETL secara menyeluruh mulai dari pengambilan data sumber hingga pemuatan ke dalam data warehouse. Antarmuka visual berbasis drag-and-drop yang disediakan Pentaho Data Integration (PDI) memudahkan perancangan alur transformasi tanpa memerlukan penulisan skrip secara manual, sehingga proses ETL dapat dilakukan secara lebih efisien dan mudah dipahami. Hal ini sejalan dengan temuan (Prasetya & Kurniawan, 2021) yang menyatakan bahwa Pentaho Data Integration (PDI) mampu memproses data warehouse penjualan secara otomatis dengan tingkat akurasi tinggi.

Penggunaan komponen Stream Lookup dalam pembangunan fact table terbukti efektif untuk menggabungkan data dari delapan tabel dimensi secara berurutan dalam satu alur transformasi. Pendekatan star schema yang diterapkan menghasilkan struktur data warehouse yang sederhana namun optimal untuk proses query analitik, karena tabel fakta hanya perlu berelasi langsung dengan tabel dimensi tanpa hierarki yang kompleks.

Hasil penelitian ini juga diperkuat oleh temuan (Barahama & Wardani, 2021a) yang membuktikan bahwa Pentaho Data Integration dengan komponen Kettle mampu menjalankan alur ETL secara terstruktur dalam berbagai konteks implementasi data warehouse, termasuk di lingkungan pendidikan. Pendekatan serupa yang diterapkan dalam penelitian ini mengonfirmasi bahwa mekanisme ETL berbasis Pentaho Data Integration (PDI) bersifat adaptif dan dapat diandalkan untuk berbagai domain data.

Secara keseluruhan, data warehouse penjualan yang dibangun telah memenuhi kebutuhan integrasi data dari berbagai sumber ke dalam satu repositori terpusat yang siap digunakan sebagai dasar pengambilan keputusan bisnis berbasis data (data-driven decision making).

KESIMPULAN

Penelitian ini berhasil mengimplementasikan proses ETL dan fact table penjualan pada data warehouse menggunakan Pentaho Data Integration. Proses ETL yang dirancang mampu mengekstrak, mentransformasi, dan memuat data ke dalam delapan tabel dimensi, yaitu dimensi produk, kategori, kota, harga satuan, jumlah terjual, total penjualan, metode pembayaran, dan waktu, serta satu fact table penjualan yang mengintegrasikan seluruh dimensi tersebut melalui mekanisme *Stream Lookup*. Pendekatan *star schema* yang diterapkan menghasilkan struktur data warehouse yang efisien dan optimal untuk mendukung analisis data penjualan. Implementasi ini membuktikan bahwa Pentaho Data Integration merupakan tools yang efektif dalam membangun data warehouse karena kemudahan penggunaannya dan kemampuannya menangani integrasi data dari berbagai sumber secara terstruktur. Untuk penelitian selanjutnya, disarankan agar data warehouse yang telah dibangun diintegrasikan dengan tools visualisasi seperti Tableau atau Power BI guna menghasilkan dashboard analitik yang lebih interaktif dalam mendukung pengambilan keputusan bisnis.

DAFTAR PUSTAKA

- Andriansyah, D. (2022). Implementasi Extract-Transform-Load (ETL) Data Warehouse Laporan Harian Pool. *Jurnal Teknik Informatika STMIK Antar Bangsa*, 8(2), 45-49.
- Badiuzzaman Biplob, M., & Mokammel Haque, M. (2021, December). Development of an efficient etl technique for data warehouses. In *Proceedings of the International Conference on Big Data, IoT, and Machine Learning: BIM 2021* (pp. 243-255). Singapore: Springer Singapore.
- 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.
- Dhaouadi, A., Bousselmi, K., Gammoudi, M. M., Monnet, S., & Hammoudi, S. (2022). Data warehousing process modeling from classical approaches to new trends: Main features and comparisons. *Data*, 7(8), 113.
- Dinesh, L., & Devi, K. G. (2024). An efficient hybrid optimization of ETL process in data warehouse of cloud architecture. *Journal of Cloud Computing*, 13(1), 12.
- Ghita, F., & Trisminingsih, R. (2021). Pengujian Data Warehouse SOLAP untuk Komoditas Pertanian Indonesia. *Jurnal Ilmu Komputer dan Agri-Informatika*, 8(1), 42-56.

- Gueddoudj, E. Y., Chikh, A., & Attia, A. (2023). Os-ETL: A High-Efficiency, Open-Scala Solution for Integrating Heterogeneous Data in Large-Scale Data Warehousing. *Ingénierie des Systèmes d'Information*, 28(3).
- Khan, B., Jan, S., Khan, W., & Chughtai, M. I. (2024). An Overview of ETL Techniques, Tools, Processes and Evaluations in Data Warehousing. *Journal on Big Data*, 6.
- Lelet, M., Kaparang, D. R., & Pratasik, S. (2023). Pengembangan Sistem Informasi Layanan Mahasiswa Fakultas Teknik Universitas Negeri Manado. *EduTIK: Jurnal Pendidikan Teknologi Informasi Dan Komunikasi*, 3(1), 138-148.
- Novandya, A. (2021). Data Warehouse: Fact Constellation Schema Dengan Pemodelan Data Dimensional Powell Pada Wedding Organizer. *Reputasi: Jurnal Rekayasa Perangkat Lunak*, 2(2), 59-65.
- Pérez Garriga, A., Honrath, P., Wolking, S., Coldewey, B., Bozkir, S. A., May, P., ... & Lipprandt, M. (2024, August). Multi-Source Data ETL (Extract, Transform, Load) for a Genetic Epilepsy Diagnosis and Treatment Dashboard. In *Medical Informatics Europe*.
- Praselia, I. P. W., & Kurniawan, I. N. H. (2021). Data Warehouse Implementasi ETL (Extract, Transform, Load) pada Data warehouse Penjualan Menggunakan Tools Pentaho. *TIERS Information Technology Journal*, 2(1).
- Pratama, I. P. A. E., & Permana, I. K. W. H. (2023). Analisa Pola Penjualan Berbasis Data Warehouse Menggunakan Metode ETL dan OLAP (Studi Kasus: Restoran Khayangan Kuliner). *TEMATIK*, 10(1), 124-130.
- Rohajawati, S. (2024). Effectiveness of Digital Transformation in Data Warehouse Technology. *The Indonesian Journal of Computer Science*, 13(1).