

Analisis dan Perancangan Data Warehouse Pada Data Transaksi Supermarket Menggunakan Schema Snowflake

Muhammad Adreansyah¹, Putri Fathia², Gustira³

^{1,2,3}Universitas Potensi Utama, Sumatera Utara, Indonesia

Email: ¹adreansyahlubis@gmail.com, ²putrifathia031@gmail.com, ³hppojob07@gmail.com

Abstract

Dalam supermarket pasti memiliki database yang digunakan untuk mengorganisasikan, menyimpan, dan mengelola data secara efisien disisi lain database berisi sekumpulan data yang terstruktur dari data yang tersimpan secara elektronik untuk menyimpan dan mengelola informasi yang terkait dengan setiap transaksi yang terjadi dalam suatu bisnis atau organisasi. data warehouse menjadi bagian yang mencakup seluruh dari sistem informasi bisnis modern yang memungkinkan pengumpulan, integrasi, dan analisis informasi dari berbagai sumber data secara fungsional. Dalam konteks supermarket, data warehouse berperan dalam mengoptimalkan pengelolaan data transaksi dengan menggunakan metode sebagai proses pengoptimalan database terutama pada data transaksi. Metode yang digunakan adalah schema snowflake untuk merancang struktur hierarkis yang terintegrasi antara data transaksi dan dimensi seperti produk, merk, dan stok dengan memanfaatkan proses yang terdapat pada data warehouse. Proses ETL (Extract, Transform, Load) digunakan untuk mengambil data transaksi dari sumber yang berbeda, mengubahnya menjadi format yang sesuai, dan memuatnya ke dalam data warehouse. Hasilnya adalah data warehouse terstruktur dengan baik yang memungkinkan analisis mendalam terhadap tren penjualan, perilaku pelanggan, dan performa bisnis. Dalam perancangan ini, pendekatan Nine-Step Kimball juga diterapkan untuk memastikan efektivitas data warehouse dengan mengikuti metode yang diterapkan supermarket dapat memastikan bahwa perancangan data warehouse telah memenuhi kebutuhan bisnis dan menghasilkan informasi yang bernilai. Dimana dalam perancangan data warehouse menggunakan schema snowflake, dengan dukungan ETL dan pendekatan Nine-Step Kimball, membantu supermarket mengoptimalkan pengambilan keputusan berdasarkan data transaksi, meningkatkan efisiensi operasional, dan mencapai keunggulan kompetitif dipasar.

Keywords: Database, Data Warehouse, Snowflake, ETL, Nine-Step Kimball

1. PENGANTAR

Dalam era digital yang terus berkembang, data menjadi salah satu aset berharga dalam dunia bisnis. Terutama di industri ritel, supermarket sebagai salah satu bisnis utama menghasilkan volume data yang besar melalui transaksi pembelian konsumen. Banyaknya data yang terdapat didalam database menyulitkan proses management data sehingga untuk menghasilkan data transaksi yang baik untuk kebutuhan proses bisnis dibutuhkan analisa dan perancangan data warehouse untuk mengetahui desain tabel transaksi yang baik dan terstruktur . Data transaksi supermarket mencakup informasi penting seperti data pelanggan, barang yang dibeli, jumlahnya, harga, waktu pembelian, dan metode pembayaran. Pemanfaatan data transaksi ini menjadi krusial dalam mengoptimalkan operasional dan pengambilan keputusan yang efektif bagi supermarket dengan tujuan untuk menghasilkan tabel fakta dari beberapa tabel yang terdapat dalam supermarket sehingga membentuk siklus transaksi yang sesuai dan dirangkum secara keseluruhan melalui tabel fakta sale transaksi melalui pemodelan data menggunakan metode nine-step kimball.

Data warehouse merupakan suatu penyimpanan data dan sistem management yang digunakan untuk mengumpulkan, menyimpan, dan mengintegrasikan data dari berbagai sumber yang berbeda, dengan tujuan untuk analisis bisnis dan pengambilan keputusan yang lebih baik. Data warehouse juga dapat menyimpan berbagai macam informasi baik yang terstruktur ataupun tidak terstruktur[1].

Dalam upaya untuk merancang data warehouse yang efektif untuk mengelola data transaksi dalam lingkungan supermarket, salah satu pendekatan yang dapat digunakan adalah penggunaan schema snowflake. Dalam schema snowflake, dimensi (seperti waktu, produk, pelanggan) dipisahkan menjadi beberapa tabel terkait, sehingga menghasilkan struktur yang lebih kompleks dibandingkan dengan model schema bintang[2]. Schema snowflake memungkinkan pemisahan dimensi menjadi tabel-tabel terkait, sehingga memfasilitasi analisis yang lebih terfokus dan pengelolaan data yang lebih efisien. Dengan memanfaatkan keunikan schema snowflake diharapkan dapat membangun struktur data warehouse yang kompleks dan terorganisir dengan baik. Dalam melakukan perancangan data warehouse dengan schema snowflake terdapat beberapa penelitian yang telah dilakukan sebagai referensi dalam penelitian ini yaitu, Data Warehouse Menggunakan Snowflake Schema Pada Virtual Shop[2], Perancangan Data Warehouse Perpustakaan Perguruan Tinggi Xyz Menggunakan Metode Snowflake Schema[3].

2. METODE PENELITIAN

2.1. Analisis

Pada tahap ini dilakukan 2 analisis yang saling berhubungan yaitu analisis masalah dan analisis kebutuhan data. Analisis masalah dalam proses perancangan datawarehouse dilakukan dengan menggunakan data processing (ETL) untuk mengetahui trend yang muncul dari perkembangan teknologi terhadap penyimpanan basis data, struktur data dan integrasi data. Analisis kebutuhan data yang digunakan dalam penelitian ini yaitu menggunakan data mentah dengan format xls, csv, sql atau dataset lainnya.

2.2. Pengumpulan Data

Pengumpulan data diambil dengan menggunakan metode studi literature untuk menemukan data yang akurat dan berhubungan dengan topik penelitian. Data yang digunakan diperoleh dari karya ilmiah, jurnal, laporan dan literatur yang berhubungan dengan pembahasan penelitian saat ini.

2.3. Pemodelan Data

Pemodelan data menggunakan metode nine-step Kimball memudahkan untuk mendesain data warehouse sehingga data dapat direpresentasikan dalam format yang di tentukan[4]. Pada bagian ini akan dijelaskan 9 tahapan-tahapan yang dilakukan dalam penelitian ini sebagai berikut;

- Choose the Process
Berdasarkan hasil analisis data yang digunakan, proses bisnis yang menjadi fokus penelitian ini adalah pengolahan data transaksi supermarket.

Tabel 1. Choose the Process

Proses Bisnis	Deksripsi	Fungsi yang terlihat
Pengelolaan data transaksi supermarket	Data transaksi supermarket diperoleh dari studi literature berupa data transaksi produk.	Data transaksi, data product, data customers

Pada tahap ini, keputusan yang tepat diambil untuk menentukan representasi apa yang akan dipresentasikan oleh sebuah tabel

fakta. Tingkat detail data yang diperoleh dari model dimensional juga ditentukan dari tahap ini. Oleh karena itu, granularity yang dipilih dalam penelitian ini adalah informasi transaksi penjualan.

Tabel 2. Choose the grain

Grain	Deksripsi	Proses bisnis yang terlibat
informasi transaksi supermarket	Informasi yang diperoleh dari transaksi supermarket berupa tanggal, nama produk, merk, harga, stok, stok terjual, stok tersisa dan total harga.	Data transaksi supermarket

- Identify And Conform the dimensions
Langkah ketiga dalam perancangan data warehouse melibatkan identifikasi dimensi yang berkaitan dengan tabel fakta.

Tabel 3. Identify and Conform the dimensions

Dimension	Deksripsi	Grain
Dim_product	Memiliki atribut idproduct, merk, namaproduct, harga, stock.	Informasi Transaksi Supermarket
Dim_orderproduct	Memiliki atribut idorder, nama order, kuantitas, total harga, waktu_order, metode_pembayaran.	Informasi Transaksi Supermarket
Dim_customers	Memiliki atribut idcustomer, nama, email.	Informasi Transaksi Supermarket
Dim_category	Memiliki atribut idcategory, idproduct, category.	Informasi Transaksi Supermarket

da tahap ini,tabel fakta dipilih dengan mempertimbangkan kemampuannya dalam mengimplementasikan semua tingkat granularitas yang digunakan.Dalam penelitian ini,tabel fakta yang dipilih adalah : fact_sale.

Tabel 4. Choose the Facts

Fakta	Deksripsi	Dimension
Fact_sale	Tabel fakta “fact_sale” memuat informasi mengenai seluruh transaksi penjualan supermarket.	Dim_penjualan,dim_product,dim_orderproduct,dim_customer s,dim_suppliers.

e pre-calculations in the fact table

Pada tahap ini,dilakukan proses perhitungan terhadap tabel fakta “fact_sale” dan hasil perhitungan tersebut disimpan sementara.Beberapa atribut yang dihitung dan disimpan meliputi tanggal transaksi,merk,stock terjual,stock tersisa,harga dan total harga

- Round out the Dimensions Tables

Tabel 5. Round out the Dimensions Tables

Dimension	Atribut	Type (length)	Keterangan
Dim_product	idproduct	int(11)	ID produk
	nama_product	varchar(45)	Nama produk
	merk	varchar(45)	Merk produk
	harga	varchar(45)	Harga asli produk
	stock	varchar(45)	Stok produk tersedia
Dim_orderproduct	idorder	int(11)	ID order
	nama_order	varchar(45)	Nama order
	kuantitas	varchar(45)	Kuantitas
	total_harga	varchar(45)	Total Harga
	waktu_order	varchar(45)	Waktu Order
Dim_customers	metode_pembayaran	varchar(45)	Metode Pembayaran
	idcustomer	int(11)	ID customer
	nama	varchar(45)	Nama customer
Dim_category	email	varchar(45)	Email Customer
	idcategory	int(11)	ID category

idproduct	int(11)	ID product
category	varchar(45)	Category Product

- Decide the Duration of the database and periodicity of update
Pada tahap ini, ditentukan batasan waktu data yang akan diambil dan dipindahkan kedalam tabel fakta. Data yang digunakan dalam desain data warehouse ini meliputi periode selama 1 tahun yang dimulai dari tahun 2021 hingga 2022.
- Track Slowly the changing dimensions
Pada tahap ini terutama dalam tabel dimensi tidak selalu memiliki nilai yang tetap atau statis melainkan bisa memiliki nilai yang dinamis. Untuk mengatasi dan menyelesaikan perubahan yang terjadi dalam dimensi, ada tiga jenis perubahan yang dijelaskan oleh Kimball dan Ross (Kimball and Ross, 2002). Salah satunya perubahan atribut dimensi yang menyebabkan pembuatan catatan dimensi baru dengan surrogate key yang berbeda. Dalam kasus ini atribut harga dan stok akan mengalami perubahan mengikuti perubahan standar harga dan produksi produk di perusahaan dan di pasar.
- Decide the query priorities and the query mode
Pada tahap ini, dilakukan pengolahan ETL (Ekstraksi, Transformasi, dan Pemuatan) agar data dapat diolah, disatukan dan dimuat ke dalam data warehouse dengan format yang konsisten dan siap digunakan untuk analisis bisnis.

2.4. ETL

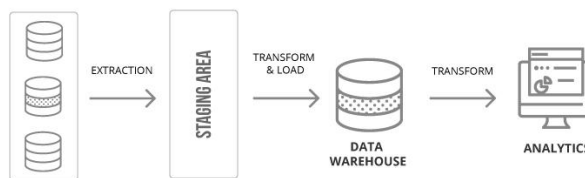
ETL merupakan proses yang sangat penting, karena melalui ETL data dapat disatukan dalam datawarehouse. ETL juga dapat digunakan untuk menggabungkan data dengan sistem yang sudah ada sebelumnya. Proses ETL mencakup ekstraksi data dari sumber, transformasi data kedalam format baru sesuai dengan kebutuhan bisnis dan memuatnya kedalam struktur data target. ETL bertujuan untuk memastikan data yang dikumpulkan disaring, dimanipulasi dan digabungkan dengan baik sehingga dapat disimpan secara efisien dan efektif dalam data warehouse yang nantinya akan digunakan untuk kebutuhan analisis [5]. Berikut adalah penjelasan dari setiap proses ETL;

- a) Ekstraksi Data (Extraction)
Ekstraksi data merupakan proses dimana data ditarik atau diambil dari berbagai sistem operasional menggunakan metode seperti query atau aplikasi ETL (Extract, Transform, Load).
- b) Transformasi Data (Transformation)

Transformasi adalah langkah penting dalam proses pengolahan data di mana data mentah yang telah diekstraksi diproses, disaring, dan diubah sesuai dengan aturan dan format yang berlaku.

c) Pemuatan Data (Load)

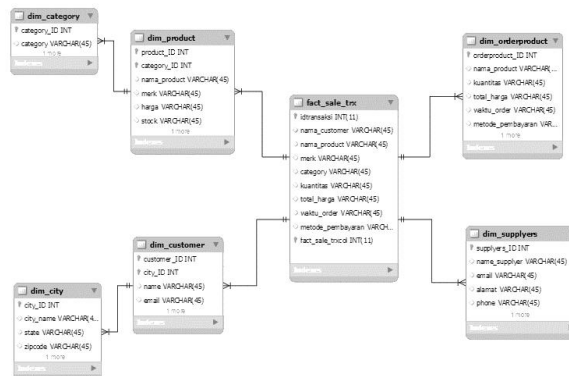
Tahap terakhir dalam proses ETL adalah proses pemuatan data, dimana data hasil transformasi dimuat ke dalam data warehouse. Salah satu cara untuk melakukan proses pemuatan data adalah dengan menjalankan script MYSQL (SQL) secara berkala.



Gambar 1. ETL Process

3. HASIL DAN DISKUSI

Hasil dari perancangan model skema data warehouse menggunakan metode nine-steps Kimball adalah penggunaan skema snowflake. Skema snowflake dipilih karena memiliki struktur yang terorganisir, menghindari redundansi data, dan memiliki keunggulan dalam kompleksitas query dibandingkan dengan skema lain seperti skema star. Skema snowflake memastikan efisiensi dan keakuratan data dalam data warehouse.



Gambar 2. Schema Snowflake

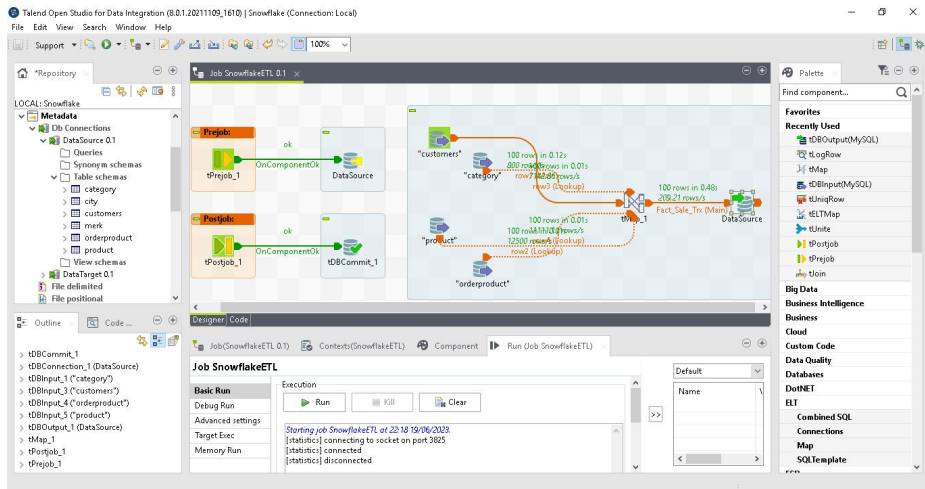
Gambar 2 menunjukkan skema data warehouse yang terdiri dari 7 tabel dimensi dan 1 tabel fakta. Setiap tabel memiliki primary key (PK) yang digunakan untuk mengidentifikasi setiap baris data dengan menggunakan data yang unik. Beberapa tabel juga memiliki foreign key (FK) yang digunakan untuk menjalin hubungan dengan tabel utama, terutama saat melakukan query join. Selain itu, terdapat beberapa atribut yang dijadikan sebagai mandatory yang berarti nilai data pada atribut tersebut tidak boleh kosong atau null.

3.1. Proses ETL

Dalam proses ETL menggunakan platform Talend, data akan diekstrak, ditransformasikan, dan dimuat ke dalam local server penyimpanan data MySQL Workbench untuk gambaran report data dengan cara yang informatif dan interaktif.

1) Extract

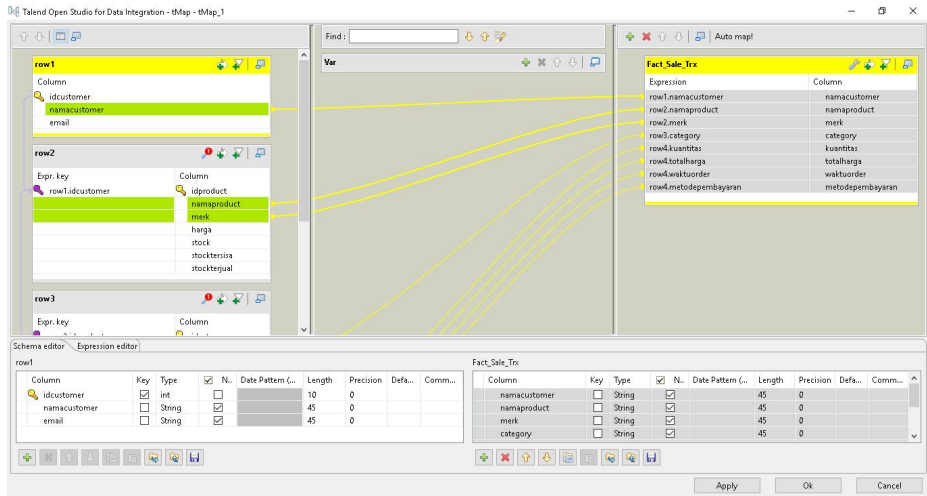
Pada tahap ini sebagai proses awal pengambilan data atau extraction data source dari file .sql untuk dibentuk table pada database sesuai dengan table fakta yang sudah ditentukan.



Gambar 3. Extract Data

2) Transform

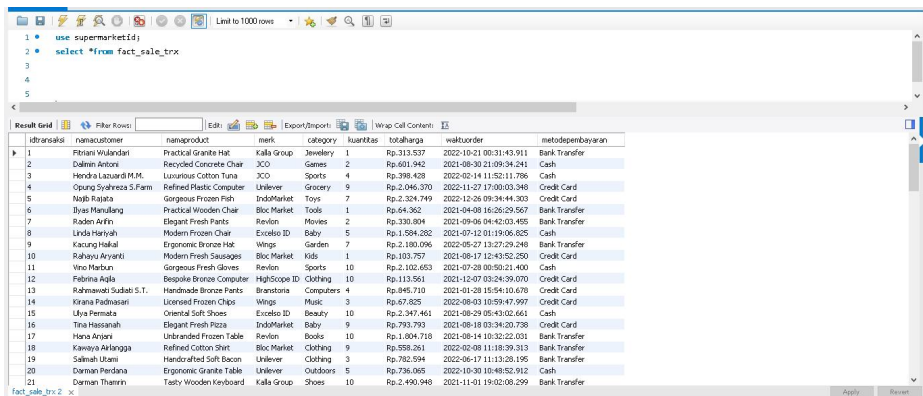
Proses transformasi tabel dalam perancangan data warehouse dilakukan dengan menggunakan komponen Talend tMap yang memungkinkan pengguna untuk melakukan join antar tabel-tabel sesuai dengan model yang telah dirancang.



Gambar 4. Transform Data

3) Load

Tabel-tabel yang telah dibentuk akan disimpan kedalam MySQL Workbench. MySQL Workbench memiliki beberapa cluster yang dirancang khusus untuk membantu para pengembang dalam proses pembuatan tabel. Salah satu cluster tersebut adalah untuk melakukan pemantauan terhadap pembentukan tabel dan query.



Gambar 5. Load Data

4. KESIMPULAN

Secara keseluruhan, penelitian ini membahas hasil dari analisis dan perancangan data warehouse menggunakan schema snowflake serta pemodelan data yang digunakan menggunakan metode pemodelan data nine-steps Kimball. Skema snowflake yang memiliki struktur table dimensi dan sub dimensi untuk mengurangi redundansi data serta yang menjadi fokus pada penelitian ini terbatas pada analisis data terkait dengan transaksi supermarket sehingga kedepannya dapat mempermudah management data yang dibuat terkait proses kegiatan transaksi pada supermarket sehingga dapat digunakan sebagai dasar untuk menyajikan data terintegrasi sebagai data pendukung dalam proses evaluasi dan sebagai dasar untuk proses pengambilan keputusan oleh manajemen terkait data transaksi supermarket yang disertakan Proses ETL (Extract, Transform, Load) untuk merancang data warehouse untuk mempersiapkan data yang berkualitas dan terstruktur agar bisa digunakan dalam analisis kedepannya dan pelaporan. Proses ini membantu memastikan integritas, konsistensi dan akurasi data dalam data warehouse.

REFERENSI

- [1] M. Firdaus, N. Putu, N. Ardiyanti, I. W. Wijaya, K. Sandi, and I. Dewa, "Perancangan dan Implementasi Data Warehouse Penjualan (Studi Kasus : Northwind Sample Database)," vol. 10, no. 1, pp. 175–188, 2021.
- [2] V. Novreza, Y. Munarko, and L. Husniah, "Data Warehouse Menggunakan Snowflake Schema Pada Virtual Shop," *J. Repos.*, vol. 2, no. 1, p. 67, 2020, doi: 10.22219/repositor.v2i1.180.
- [3] A. Dahlan, E. Utami, and E. T. Luthfi, "Perancangan Data Warehouse Perpustakaan Perguruan Tinggi Xyz Menggunakan Metode Snowflake Schema," *Respati*, vol. 8, no. 24, pp. 1–14, 2017, doi: 10.35842/jtir.v8i24.67.
- [4] J. Tunggono, M. R. Faisal, and D. T. Nugrahadi, "Pemanfaatan Data Warehouse Sebagai Sarana Penunjang Penyusunan Borang Akreditasi Standar 3 dan Standar 4," *Kumpul. J. Ilmu Komput.*, vol. 02, no. 01, pp. 96–109, 2015.
- [5] M. P. A. Ariawan, I. M. D. Ardiada, and Y. P. Sudarmojo, "Design of Library Data Warehouse Using OLTP Result of Services Analysis," *Int. J. Eng. Emerg. Technol.*, vol. 3, no. 1, pp. 62–65, 2018.