

LAPORAN MISI PERTAMA PERGUDANGAN DATA
FASHION RETAIL SALES



Disusun Oleh:

- | | |
|--------------------------------|-----------|
| 1. Sofyan Fauzi Dzaki Arif | 122450116 |
| 2. Sylviani Primaastuti Ananda | 121450042 |
| 3. Rahma Neliyana | 122450036 |
| 4. Renta Siahaan | 122450070 |
| 5. Diana Syafithri | 122450141 |

PROGRAM STUDI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA 2025

A. Ringkasan Kebutuhan dari Misi

1. Kebutuhan Bisnis

Bertujuan dalam proses pengambilan keputusan secara strategis oleh perusahaan retail fashion dengan penyediaan informasi cepat, tepat dan relevan.

- a. Menganalisis perilaku pelanggan untuk mengetahui tren dan preferensi pasar.
- b. Mengidentifikasi item atau produk yang paling banyak di beli.
- c. Mengetahui metode pembayaran yang paling sering digunakan pelanggan.
- d. Memonitoring pendapatan berdasarkan rentang waktu tertentu (harian, mingguan, bulanan, dan tahunan).
- e. Mendukung evaluasi efektivitas strategi penjualan berdasarkan waktu, produk, dan karakteristik pelanggan.

2. Kebutuhan Data

Data yang digunakan bersumber dari berbagai entitas dalam sistem operasional perusahaan dan dikategorikan dalam bentuk fakta dimensi.

- a. Data transaksi (fakta): informasi pembelian, tanggal pembelian, item yang di beli, jumlah pembayaran, rating, review, dan metode pembayaran.
- b. Data pelanggan (dimensi): identitas pelanggan, nama, alamat, serta informasi segmentasi pelanggan.
- c. Data produk (dimensi): informasi terkait produk yang dijual berupa ID produk, nama item, dan kategori produk.
- d. Data waktu (dimensi): informasi waktu yang terstruktur seperti hari, minggu, bulan dan tahun untuk mendukung analisis tren.
- e. Data pembayaran (dimensi): informasi terkait jenis metode pembayaran yang digunakan pelanggan.

3. Kebutuhan Teknis

Dalam kebutuhan teknis, sistem ini dirancang dengan mempertimbangkan efisiensi, skalabilitas, dan kemudahan proses analisis.

- a. Penggunaan skema bintang (star schema): untuk mempercepat proses eksekusi query serta mempermudah proses eksplorasi data.
- b. Proses ETL (extract, transform, load): proses ekstraksi data dengan cara dibersihkan dan ditransformasikan, lalu dimuat ke dalam tabel fakta dimensi.
- c. Integrasi data yang bersih dan konsisten: proses transformasi dengan pembersihan data dari nilai kosong dan standarisasi format data yang digunakan.
- d. Skalabilitas sistem: untuk menangani pertumbuhan data dalam jangka panjang.
- e. Kemudahan akses dan analisis: agar mudah di akses dalam keperluan analisis, visualisasi, dan laporan.

B. Skema Konseptual Multidimensi

1. Tujuan bisnis

Dalam membangun gudang data, tujuan dari gudang data ini adalah untuk membantu perusahaan retail fashion:

1. Menganalisis perilaku pembelian pelanggan
2. Melihat item yang paling banyak dibeli

3. Mengukur performa produk berdasarkan rating review
4. Mengetahui metode pembayaran yang paling sering digunakan
5. Memonitor pendapatan harian, mingguan, bulanan, dll.

2. Desain konseptual awal

- a. Fakta
 - Pembelian (Purchase Fact)
- b. Dimensi
 - Customer (berbasis Reference ID)
 - Produk / Item
 - Tanggal
 - Metode Pembayaran

3. Cek ketersediaan data dan spesifikasi mapping

Field Asli	Tipe Data	Mapping	Tabel Dimensi / Fakta
Customer Reference ID	Integer	Customer_ID	Dim_Customer
Item Purchased	String	Item_Name	Dim_Product
Date Purchase	Date (Float?)	Purchase_Date	Dim_Date
Review Rating	Integer	Review_Score	Fact_Purchase
Payment Method	String	Payment_Method	Dim_Payment_Method
Purchase Amount (USD)	Float	Purchase_Amount	Fact_Purchase

C. Penjelasan Tiap Komponen

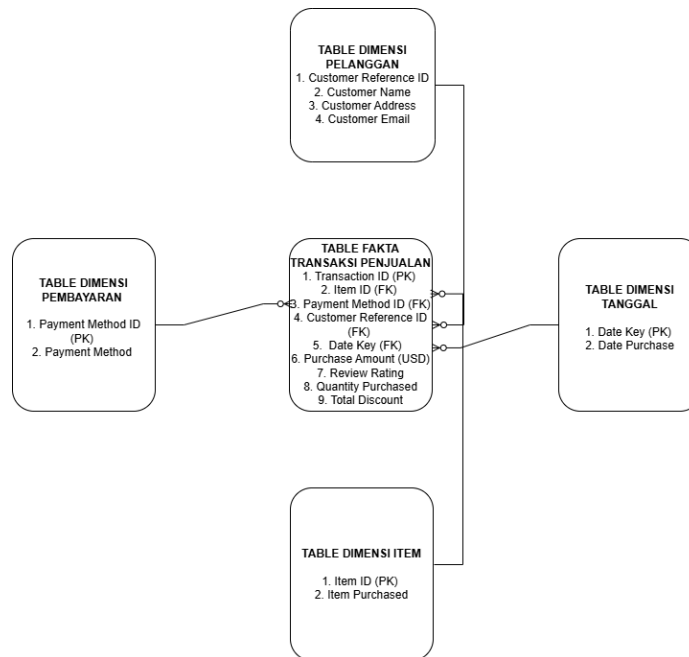
1. Logical Schema

Logical schema merujuk pada struktur tabel yang dirancang untuk mengorganisir dan mengelola data di dalam Data Warehouse. Maka dari itu yang digunakan sebagai table fakta yaitu (Transaksi Penjualan), dan yang bertindak sebagai table dimensi yaitu : Dimensi Tanggal, Dimensi Pelanggan, Dimensi Pembayaran, dan yang terakhir yaitu Dimensi Item.

2. Denormalisasi

Dalam data mentah yang didapatkan bahwasanya table table sudah di pecah menjadi beberapa dimensi dan fakta, maka dari itu perlu diadakan join ataupun penggabungan data menjadi satu yaitu table fakta, karena Denormalisasi sendiri adalah proses menggabungkan beberapa tabel menjadi satu tabel untuk meningkatkan performa query, meskipun dapat menyebabkan redundansi data. Dalam Data Warehouse, tabel fakta biasanya denormalisasi untuk meningkatkan kecepatan eksekusi query, karena penggabungan tabel dimensi dengan tabel fakta sering kali lebih cepat dibandingkan melakukan join pada saat query.

3. Star Schema



Skema ini mendukung analisis data yang efisien karena setiap tabel dimensi dihubungkan langsung dengan tabel fakta tanpa adanya hierarki yang kompleks, yang membuatnya lebih mudah untuk melakukan query analitik

4. ETL Logical Planning

ETL (Extract, Transform, Load) adalah proses penting dalam memindahkan data dari sumber ke dalam Data Warehouse. Tahap-tahap logis dari proses ETL adalah sebagai berikut:

- Extract : data diambil dari kaggle yang dimana didalamnya terdapat informasi informasi mengenai data yang digunakan dan akan di extract ke dalam format csv dan mengambil data data yang penting untuk digunakan.
- Transform : Pembersihan data yang digunakan nantinya agar ketika digunakan untuk data warehouse sudah bersih digunakan seperti nilai Review_Rating yang kosong, Pemetaan Dimensi yang akan digunakan seperti Dimensi Tanggal, Dimensi Pelanggan, Dimensi Pembayaran, dan yang terakhir yaitu Dimensi Item.
- Load : Data yang sudah ditransformasi dimuat ke dalam tabel dimensi dan fakta dalam Data Warehouse. Data transaksi akan dimasukkan ke dalam tabel fakta Transaksi Penjualan, sementara data terkait pelanggan, item, pembayaran, dan tanggal akan dimuat ke dalam tabel dimensi yang sesuai.

D. Justifikasi Desain Konseptual

Dalam pembangunan data warehouse untuk industri fashion retail, desain konseptual menjadi landasan utama untuk menyusun serta menganalisis struktur data yang akan digunakan. Pada konsep fashion retail yang akan dibuat, desain konseptual disusun dengan pendekatan skema bintang, hal ini dikarenakan kemampuannya yang dapat menyederhanakan hubungan antar data, mempercepat query serta dapat memudahkan informasi dari berbagai perspektif bisnis.

Desain ini pada tabel fakta menggunakan transaksi penjualan yang akan menghubungkan beberapa tabel dimensi yang akan digunakan diantaranya: pelanggan, tanggal, pembayaran, dan item. Pemilihan dimensi ini berdasarkan identifikasi terkait kebutuhan bisnis. Berikut penjelasan justifikasi pemilihan dimensi tersebut:

1. Dimensi pelanggan berperan penting dalam memahami karakteristik konsumen yang melakukan transaksi dapat berupa id pelanggan, nama, alamat dan email sangat dibutuhkan untuk proses segmentasi pelanggan serta analisis pelanggan belanja.
2. Dimensi tanggal untuk memberikan konteks waktu atas setiap transaksi yang terjadi. Informasi waktu dapat berupa jam, hari, minggu, bulan maupun tahun. Dimensi ini berperan untuk manajemen dan mengidentifikasi penjelasan musiman, tren jangka panjang serta memantau efektivitas promosi yang dijalankan pada waktu-waktu tertentu dalam satu tahun.
3. Dimensi pembayaran memberikan informasi tentang aktivitas pelanggan melakukan transaksi, seperti menggunakan uang tunai, kartu debit atau dengan dompet digital dengan variabel berupa id metode pembayaran dan metode pembayaran
4. Dimensi item akan menjelaskan produk jenis produk yang dijual dalam sistem penjualan dengan atribut id item, jenis dan nama produk yang dijual. Keberadaan dimensi sangat penting untuk mengetahui detail dari produk yang paling diminati, serta membantu manajemen inventaris dan pengambilan keputusan untuk restock atau promosi. Dengan adanya dimensi item ini, manajemen produk untuk menyusun strategi produk berdasarkan performa penjualan dan preferensi pasar.

E. Kesesuaian dengan Sumber Data

1. Identifikasi Sumber Data

- a) Sistem Transaksi Penjualan : Menyediakan data pembelian, seperti tanggal pembelian, item yang dibeli, metode pembayaran, jumlah pembayaran, dan rating/review
- b) Sistem Informasi Pelanggan : Berisi informasi identitas pelanggan, seperti nama, alamat, email, segmentasi.
- c) Sistem Informasi Produk : Menyimpan detail produk seperti ID produk, nama item, dan kategori produk.
- d) Sistem Pembayaran : Berisi informasi metode pembayaran yang digunakan pelanggan, seperti tunai, kartu debit, atau e-wallet
- e) Data eksternal (Kaggle) : Menjadi sumber awal yang menyediakan keseluruhan dataset dalam format CSV untuk diekstrak dan diolah

2. Mapping Data

- a) Data transaksi penjualan digunakan sebagai isi dari tabel Fakta_Penjualan (jumlah pembayaran, rating, review)
- b) Tanggal transaksi dipetakan ke dalam Dim_Tanggal untuk mendukung analisis musiman dan tren
- c) Data pelanggan dimasukkan ke dalam Dim_Pelanggan berdasarkan ID dan informasi personal lainnya
- d) Data produk yang dibeli dipetakan ke Dim_Item yang berisi ID, nama item, dan kategori produk

- e) Metode pembayaran pelanggan dipetakan ke dalam Dim_Pembayaran untuk mengetahui preferensi metode transaksi.
3. Ketersediaan Data
- a) Sebagian besar data tersedia dari hasil ekstraksi dataset Kaggle dan sistem operasional penjualan
 - b) Potensi kekurangan data L
 - I. Nilai review bisa saja kosong dan perlu diatasi dalam tahap transformasi
 - II. Tidak semua informasi segmentasi pelanggan mungkin terdefinisi secara eksplisit dalam data mentah.
4. Kualitas Data
- a) Akurasi data penting terutama untuk transaksi dan informasi pembayaran yang menjadi dasar analisis bisnis.
 - b) Kelengkapan dimensi pelanggan, tanggal, produk, dan pembayaran harus dipastikan untuk validitas laporan multidimensi.
 - c) Konsistensi antar sistem dijaga melalui proses ETL dengan transformasi dan validasi, seperti pembersihan nilai kosong dan standarisasi format data.