

LAPORAN PERGUDANGAN DATA
***DATA WAREHOUSE* DALAM INDUSTRI *E-COMMERCE* DAN RETAIL**



Disusun oleh:

Marhanny Zahra Nurendra	120450017
Eunike Bunga Santa	121450095
Muhammad Bayu Syuhada	122450007
Johannes Krisjon Silitonga	122450043
Rian Bintang Wijaya	122450094
Bastian Heskia Silaban	122450130

PRODI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
2025

1. Ringkasan Kebutuhan Bisnis

Superstore membutuhkan *data warehouse* untuk mengkonsolidasikan data dari berbagai sumber seperti *Point of Sale* (POS), sistem inventori, dan *platform customer management*. Tujuan utama adalah memperoleh gambaran komprehensif tentang kinerja bisnis lintas kategori produk, segmen pelanggan, dan wilayah geografis untuk mendukung pengambilan keputusan strategis. *Data warehouse* ini akan memungkinkan analisis terkait:

- a. Profitabilitas produk di berbagai kategori dan subkategori
- b. Efektivitas strategi diskon dan promosi
- c. Performa penjualan berdasarkan wilayah geografis
- d. Pola pembelian dan perilaku pelanggan

2. Skema Konseptual Multidimensi

Dalam pengembangan skema konseptual multidimensi ini, kami menggunakan pendekatan gabungan Business/Data Driven sesuai dengan kebutuhan Superstore:

- a. Pendekatan *business driven*
Dari perspektif bisnis, kami mengidentifikasi kebutuhan analisis utama seperti profitabilitas produk berdasarkan kategori, efektivitas strategi diskon, performa penjualan regional, segmentasi pelanggan, dan efisiensi pengiriman.
- b. Pendekatan *data driven*
Dari perspektif data, kami menganalisis ketersediaan data dalam metadata yang mencakup data transaksi penjualan (Sales, Quantity, Discount, Profit), informasi produk (Category, Sub-category), data lokasi geografis (Country, Region, State, City, Postal Code), segmentasi pelanggan (Segment), dan metode pengiriman (Ship Mode).
- c. Hasil penggabungan
Berdasarkan penggabungan kedua pendekatan tersebut, kami mengembangkan skema bintang (star schema) yang optimal dengan satu tabel fakta utama FACT_SALES yang menyimpan data transaksi penjualan dengan berbagai metrik penting, serta lima tabel dimensi yang saling terhubung yaitu DIM_CUSTOMER untuk informasi pelanggan, DIM_PRODUCT untuk detail produk, DIM_LOCATION untuk hierarki lokasi geografis, DIM_TIME untuk analisis temporal, dan DIM_SHIPPING untuk informasi pengiriman.

Struktur multidimensi ini memungkinkan analisis yang komprehensif melalui berbagai perspektif bisnis sambil tetap memaksimalkan penggunaan data yang tersedia.

3. Penjelasan Tiap Komponen

- a. FACT_SALES
 - sales_id: Primary key untuk tabel fakta

- customer_id, product_id, location_id, time_id, shipping_id: Foreign keys ke tabel dimensi
- sales: Nilai penjualan dalam mata uang
- quantity: Jumlah unit produk yang terjual
- discount: Persentase atau nilai diskon yang diberikan
- profit: Keuntungan yang diperoleh

b. DIM_CUSTOMER

- customer_id: Primary key
- customer_name: Nama pelanggan
- segment: Segmentasi pelanggan (Consumer, Corporate, Home Office)
- contact_details: Informasi kontak pelanggan

c. DIM_PRODUCT

- product_id: Primary key
- product_name: Nama produk
- category: Kategori produk (Furniture, Office Supplies, Technology)
- sub_category: Subkategori produk (Chairs, Phones, Storage, dll)

d. DIM_LOCATION

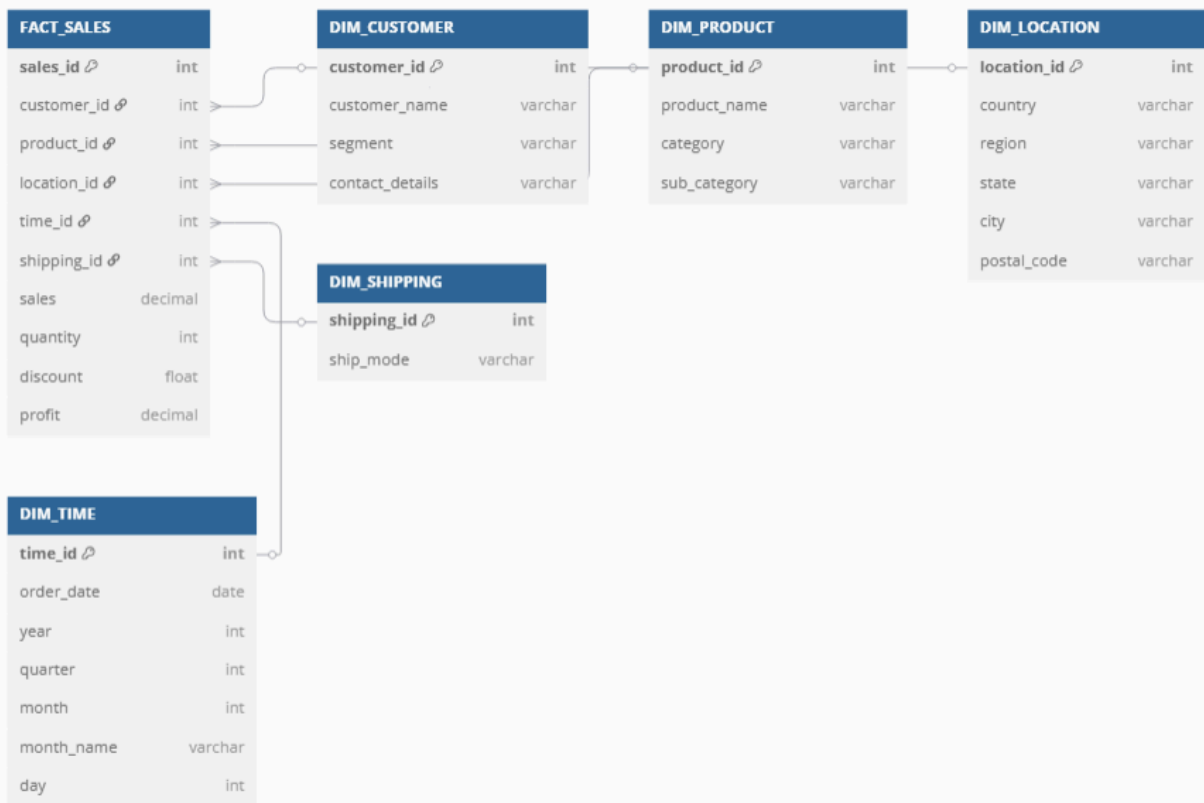
- location_id: Primary key
- country: Negara (sebagian besar United States)
- region: Wilayah geografis (South, West, East, Central)
- state: Provinsi atau negara bagian
- city: Kota
- postal_code: Kode pos

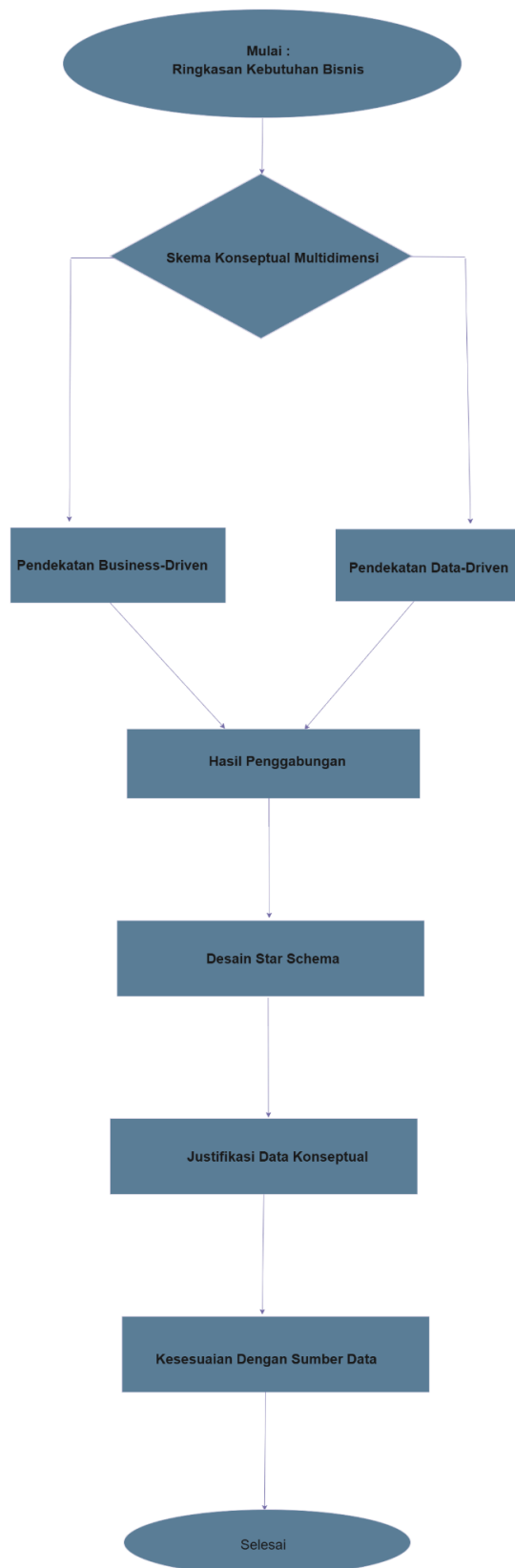
e. DIM_TIME

- time_id: Primary key
- order_date: Tanggal pemesanan lengkap
- year, quarter, month, month_name, day: Granularitas waktu

f. DIM_SHIPPING

- shipping_id: Primary key
- ship_mode: Metode pengiriman (Standard, Second Class, First Class, Same Day)





4. Justifikasi Data Konseptual

Desain konseptual data warehouse untuk proyek Superstore menggunakan pendekatan multidimensional dengan skema bintang (star schema), yang menempatkan tabel fakta penjualan (FACT_SALES) sebagai pusat penyimpanan data numerik yang relevan bagi analisis bisnis. Tabel ini menyimpan metrik utama seperti Sales, Quantity, Discount, dan Profit, yang merupakan ukuran performa transaksi penjualan. Di sekeliling tabel fakta, terdapat beberapa tabel dimensi yang masing-masing mewakili perspektif penting: pelanggan (DIM_CUSTOMER), produk (DIM_PRODUCT), lokasi (DIM_LOCATION), waktu (DIM_TIME), dan pengiriman (DIM_SHIPPING).

Dimensi pelanggan dipilih untuk memungkinkan analisis segmentasi dan perilaku konsumen, sementara dimensi produk berguna dalam mengevaluasi kinerja penjualan berdasarkan kategori dan sub-kategori barang. Dimensi lokasi mendukung analisis berdasarkan wilayah geografis, yang penting untuk logistik dan ekspansi pasar. Dimensi waktu memberikan konteks temporal terhadap performa bisnis dari waktu ke waktu, sedangkan dimensi pengiriman membantu mengevaluasi efisiensi metode pengiriman yang digunakan. Pendekatan skema bintang dipilih karena strukturnya yang sederhana namun optimal untuk kebutuhan query analitik seperti laporan penjualan, analisis diskon, dan tren profitabilitas.

Keterkaitan antara fakta dan dimensi dalam desain ini sangat relevan dengan kebutuhan berbagai stakeholder dalam organisasi: CEO dan Manajer Umum memerlukan data menyeluruh yang mencakup dimensi waktu dan lokasi; Tim Pemasaran fokus pada segmentasi pelanggan melalui dimensi pelanggan; Tim Logistik sangat bergantung pada informasi lokasi untuk optimasi distribusi; sedangkan Tim Data Analyst membutuhkan integrasi waktu, wilayah, dan kategori produk untuk menghasilkan laporan analisis prediktif yang mendalam. Dengan demikian, desain ini memastikan fleksibilitas eksplorasi data dan mendukung pengambilan keputusan strategis berbasis data.

5. Kesesuaian Dengan Sumber Data

Dataset SampleSuperstore.csv menegaskan kesesuaian yang kuat dengan rancangan konseptual yang telah diajukan. Seluruh metrik utama yang dibutuhkan untuk membentuk tabel fakta, seperti Penjualan (Sales), Kuantitas (Quantity), Diskon (Discount), dan Laba (Profit), semuanya tersedia secara langsung dalam data sumber. Selain itu, atribut-atribut esensial untuk pembangunan dimensi-dimensi analisis juga dapat diidentifikasi dengan lengkap; ini mencakup dimensi produk (berdasarkan Product

Name, Category, Sub-Category), dimensi pelanggan (dari Customer Name dan Segment), dimensi lokasi (meliputi City, State, Region, Postal Code), dimensi waktu (menggunakan Order Date), serta dimensi pengiriman (berdasarkan Ship Mode).

Meskipun pengenal unik (ID) seperti `product_id` dan `customer_id` tidak secara eksplisit terdapat dalam dataset awal, identifikasi tersebut dapat diturunkan atau dihasilkan melalui proses ETL (Extract, Transform, Load) yang direncanakan. Dengan demikian, dapat disimpulkan bahwa seluruh komponen arsitektur model yang diusulkan dapat direpresentasikan secara utuh dan efektif memanfaatkan data yang tersedia.