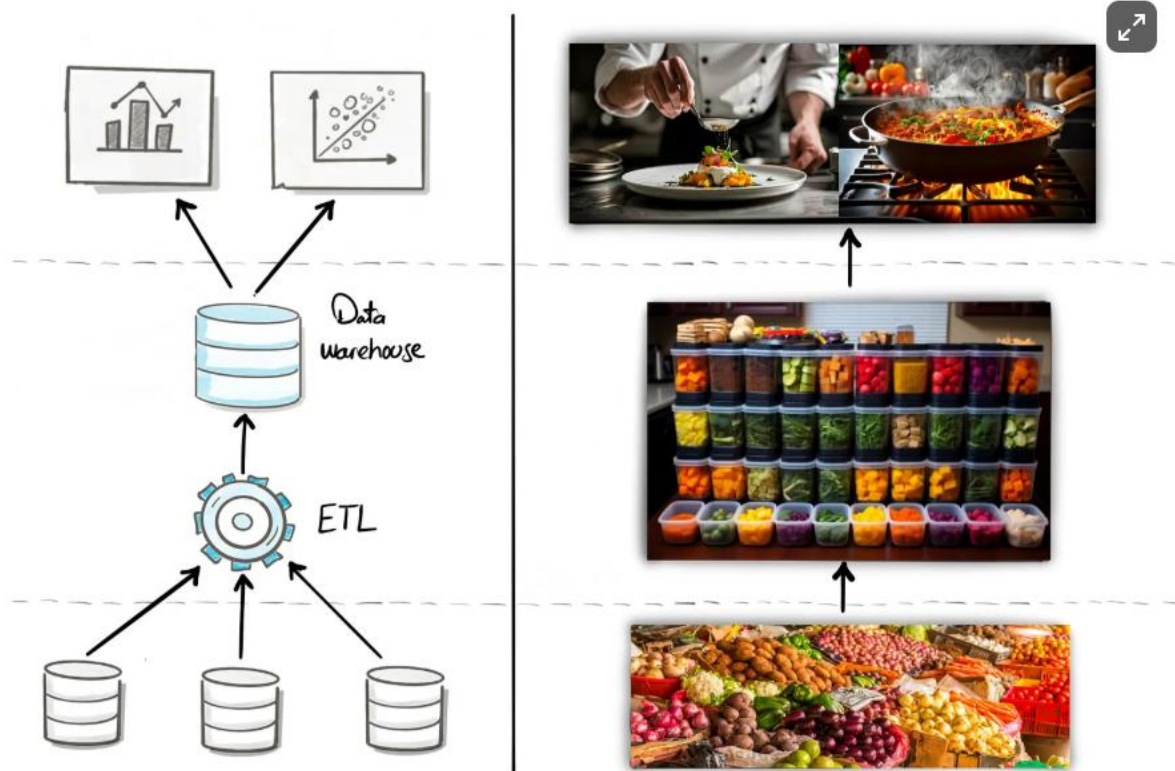


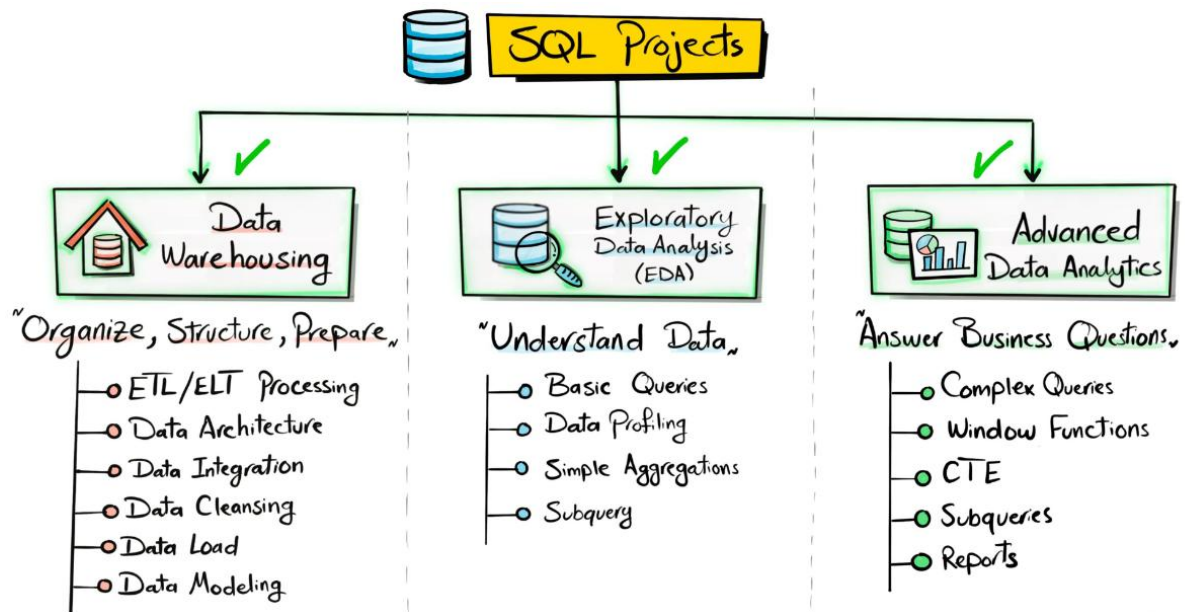
SQL DATA WAREHOUSING FROM SCRATCH

1. What is Data Warehouse ?

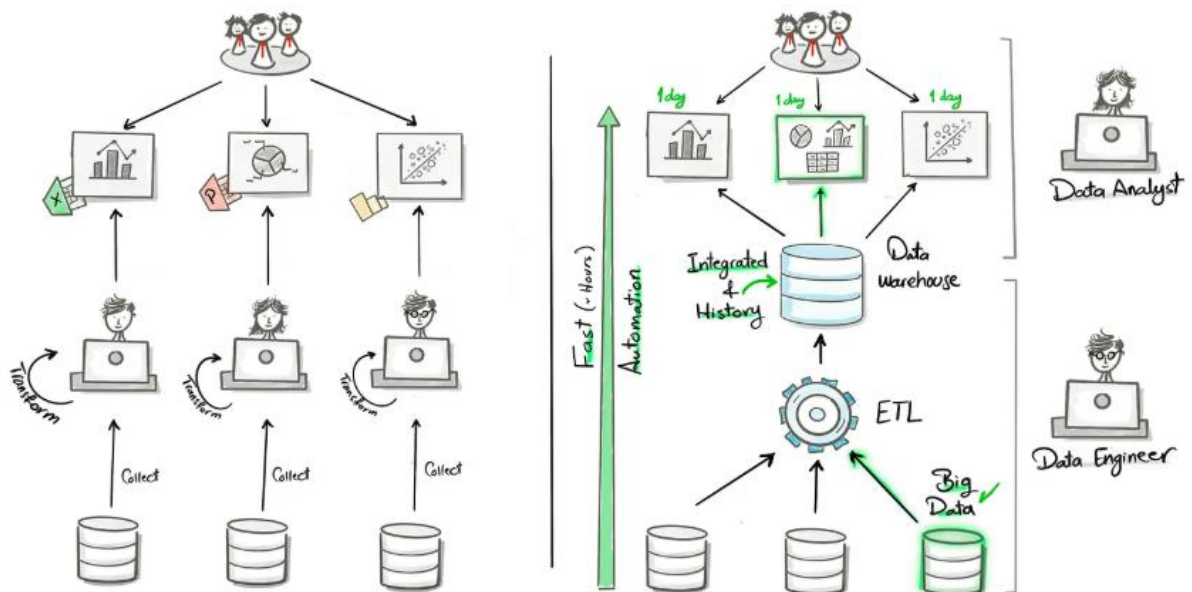


Dalam data warehouse adalah subjek berorientasi pada variasi waktu terintegrasi dan koleksi data non-volatile yang dirancang untuk mendukung manajemen proses pengambilan keputusan

Subjek berorientasi berarti Gudang selalu berfokus pada area bisnis seperti penjualan pelanggan keuangan dan sebagainya. Terintegrasi karena berjalan dan terintegrasi beberapa sistem sumber biasanya anda membangun data warehouse bukan hanya untuk satu orang sumber tetapi untuk beberapa sumber waktu variasi itu berarti anda dapat menyimpannya data history di dalam data warehouse, non-volatile artinya setelah data masuk ke data warehouse itu tidak dihapus atau dimodifikasi jadi beginilah cara membangunnya.



2. Mengapa kita membutuhkan Data Warehousing?

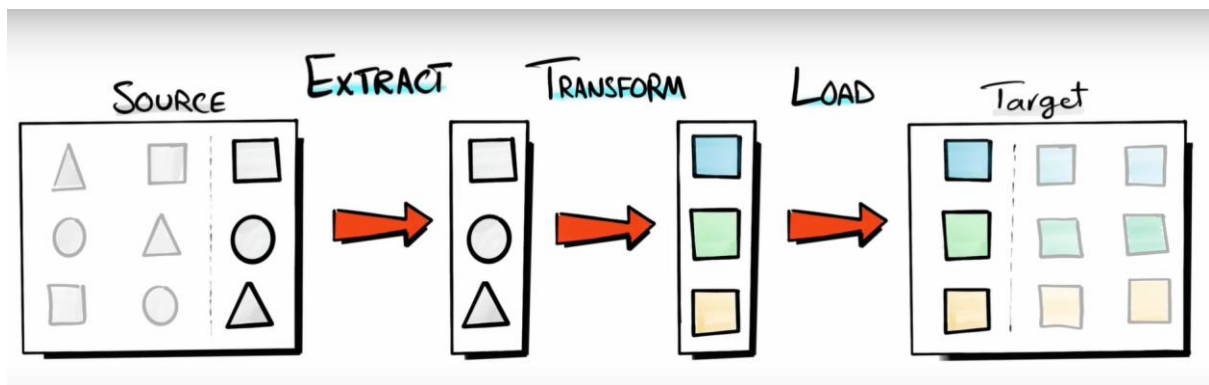


Berdasarkan pengalaman saya, mengelola data tanpa Data Warehouse itu kacau. Setiap tim menarik laporan dari sistem yang berbeda—dan akhirnya menghasilkan angka yang saling bertentangan, laporan yang sudah usang, serta pekerjaan manual yang tiada habisnya.

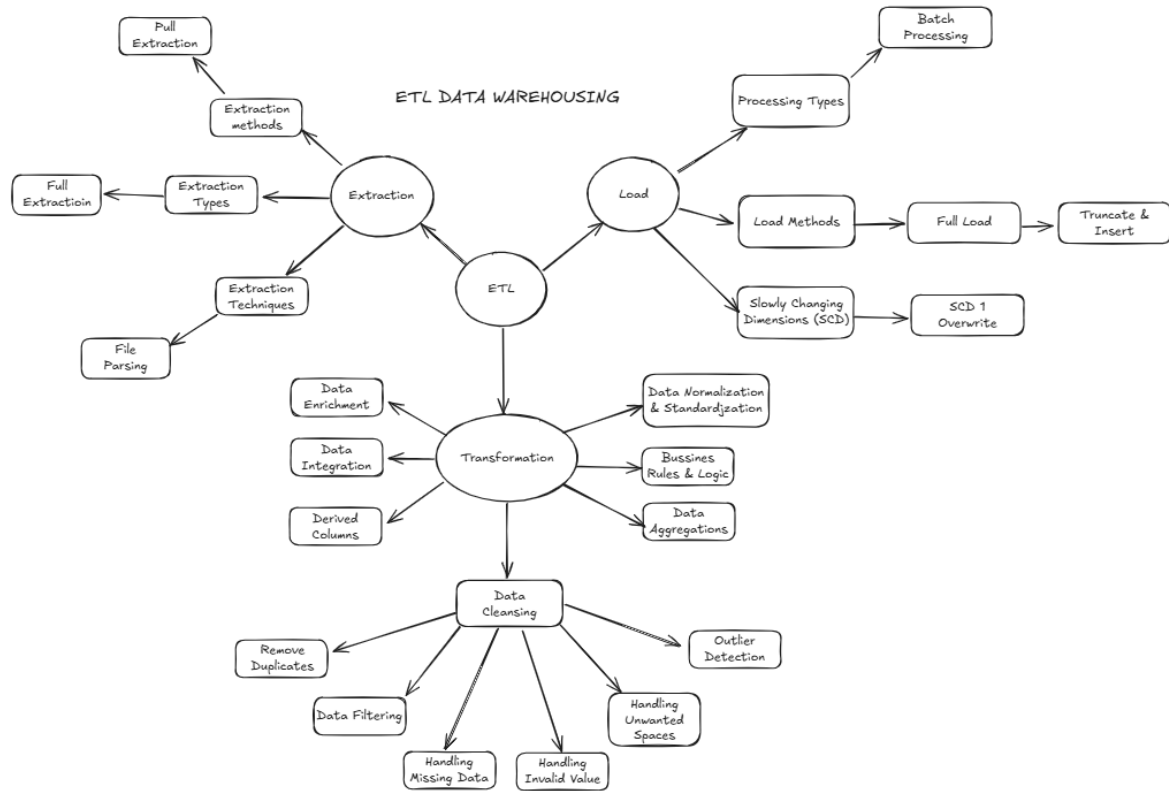
Data Warehouse menyelesaikan masalah ini:

Data Terpusat	: Semua data ada di satu tempat.
Otomatisasi	: Proses ETL menangani pembersihan dan pemuatan data.
Konsistensi	: Satu sumber kebenaran.
Kecepatan	: Laporan siap dalam hitungan jam, bukan minggu.

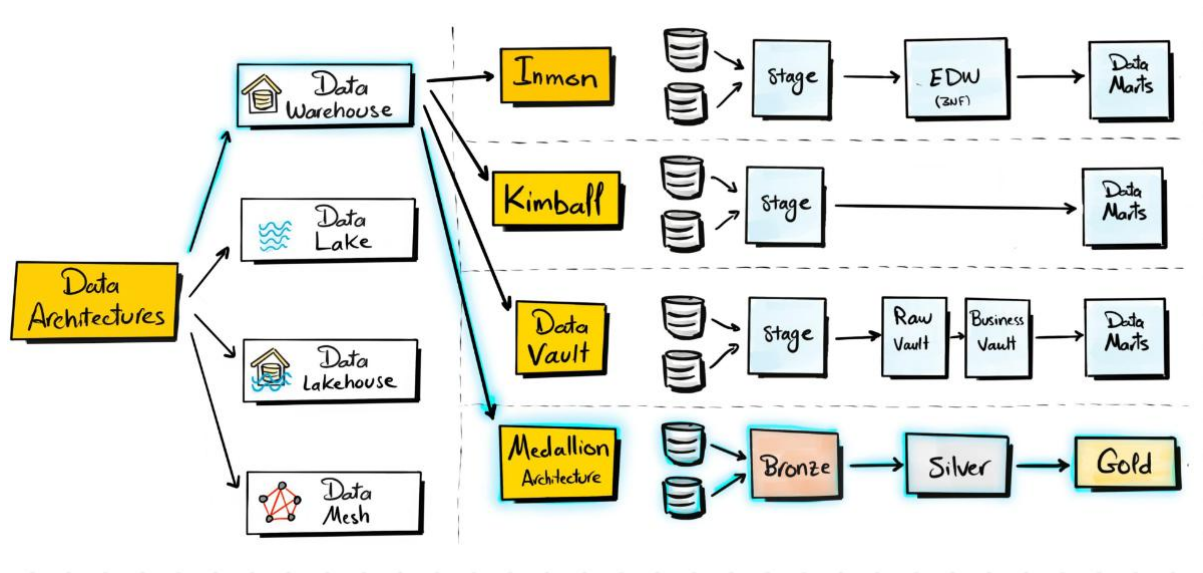
3. Apa itu ETL ?



- **Source**
Sumber data ini bisa berupa *database*, *file* teks, aplikasi, atau sistem lainnya. Perhatikan bahwa data di sumber ini mungkin tidak terstruktur, memiliki format yang berbeda-beda, atau mengandung data yang tidak relevan.
- **Extract**
Proses "Extract" ini berarti mengambil atau menarik data yang dibutuhkan dari berbagai sumber. Pada tahap ini, data mungkin belum diubah atau dibersihkan secara signifikan.
- **Transform**
Proses "Transform" melibatkan pembersihan, standarisasi, penggabungan, pemfilteran, dan konversi data yang diekstrak agar sesuai dengan kebutuhan sistem target.
- **Load**
Proses "Load" adalah tahap akhir di mana data yang telah ditransformasi dimuat ke dalam sistem target, seperti *data warehouse*, *data mart*, atau aplikasi lainnya, untuk keperluan analisis, pelaporan, atau pengambilan keputusan.
- **Target**
Di bagian paling kanan, terdapat sebuah kotak yang diberi label "Target". Kotak ini berisi bentuk-bentuk berwarna yang telah dimuat. Tata letak bentuk di "Target" berbeda dengan di "Source", menunjukkan bahwa data mungkin telah diorganisir atau distrukturkan ulang selama proses ETL. "Target" ini adalah tempat di mana data yang telah diproses siap untuk digunakan.






4. Design Data Architecture





Data Warehouse Project

Project Epics

Aa Name	Q Progress	
 Requirements Analysis		
 Design Data Architecture	0.00% <div></div>	
 Project Initialization	0.00% <div></div>	
+ New page		

Project Tasks

- ▶  Design Data Architecture
- ▶  Project Initialization

5. Analysing Requirements

Sekarang menganalisis kebutuhan itu sangat penting untuk memahami jenis data warehouse yang akan dibangun karena tidak hanya satu standar tentang cara membangunnya dan jika anda menerapkan data warehouse, itulah kita perlu duduk bersama dengan pemegang saham untuk merancang kebutuhan yang dibutuhkan pemegang sahat tersebut baru kita bisa membangun data warehousing.

Project Requirements

Building the Data Warehousing (Data Engineering)

Objective




Mengembangkan gudang data modern menggunakan SQL Server untuk mengkonsolidasikan data penjualan, guna mendukung pelaporan analitik dan pengambilan keputusan yang tepat.

Specifications

- Data Sources: Mengimpor data dari dua sistem sumber (ERP dan CRM) yang disediakan dalam format file CSV.

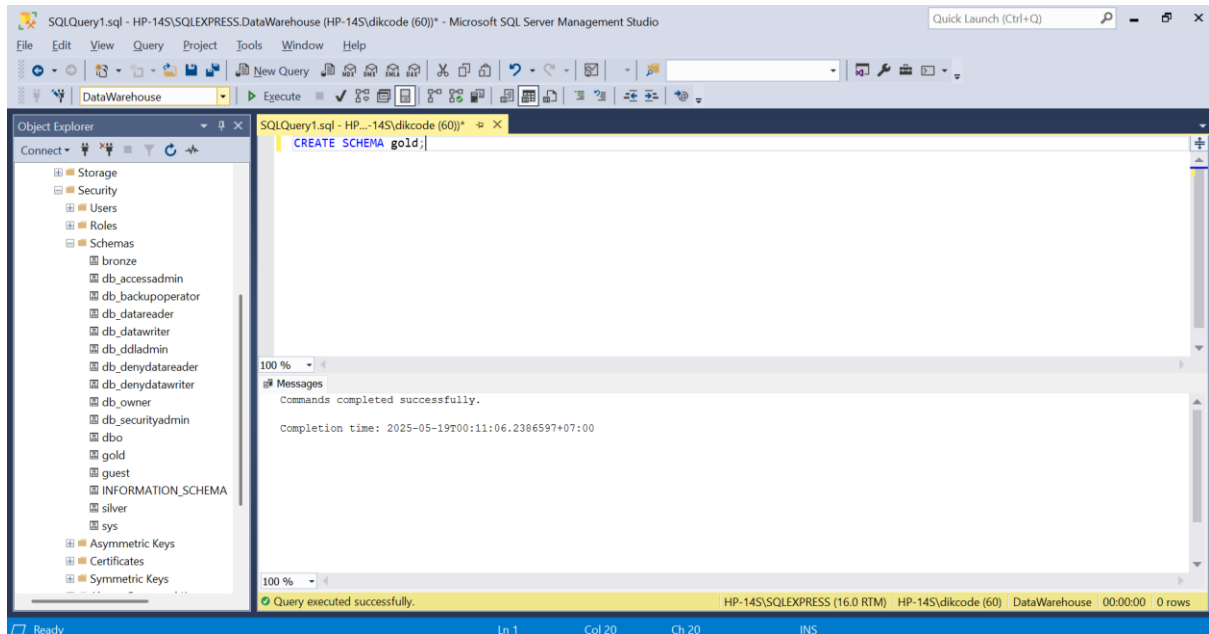
- Data Quality: Membersihkan dan mengatasi masalah kualitas data sebelum dilakukan analisis.
- Integration: Menggabungkan kedua sumber data ke dalam satu model data yang mudah digunakan dan dirancang untuk kueri analitik.
- Scope: Fokus hanya pada dataset terbaru; historisasi data tidak diperlukan.
- Documentation: Menyediakan dokumentasi yang jelas mengenai model data untuk mendukung baik pemangku kepentingan bisnis maupun tim analitik.

6. Design the Layers of Data Warehousing

	 Bronze Layer	 Silver Layer	 Gold Layer
Definition	Raw, unprocessed data as-is from sources	Clean & standardized data	Business-Ready data
Objective	Traceability & Debugging	(Intermediate Layer) Prepare Data for Analysis	Provide data to be consumed for reporting & Analytics
Object Type	Tables	Tables	Views
Load Method	Full Load (Truncate & Insert)	Full Load (Truncate & Insert)	None
Data Transformation	None (as-is)	<ul style="list-style-type: none"> - Data Cleaning - Data Standardization - Data Normalization - Derived Columns - Data Enrichment 	<ul style="list-style-type: none"> - Data Integration - Data Aggregation - Business Logic & Rules
Data Modeling	None (as-is)	None (as-is)	<ul style="list-style-type: none"> - Star Schema - Aggregated Objects - Flat Tables
Target Audience	- Data Engineers	<ul style="list-style-type: none"> - Data Analysts - Data Engineers 	<ul style="list-style-type: none"> - Data Analysts - Business Users

7. What is SQL Server ?

SQL Server adalah sebuah Database Management System (DBMS) yang berfungsi untuk menjalankan dan menyimpan database. Pada dasarnya, database tersebut disimpan di dalam SQL Server. Biasanya, SQL Server diinstal di server lokal (on-premises) atau menggunakan layanan cloud yang menyediakan server SQL.



8. What is SQL SSMS ?

SQL Server Management Studio adalah antarmuka grafis tempat anda dapat pergi dan mulai berinteraksi dengan database tempat anda dapat melihat data menulis query menyelesaikan tugas dan seterusnya.