



PART 5

Data Warehouse

VERSI 2025

Data Warehouse

- Data Warehouse (DW) adalah kumpulan data yang dihasilkan untuk mendukung pengambilan keputusan; DW juga merupakan repositori data terkini dan historis yang berpotensi menarik bagi para manajer di seluruh organisasi.
- DW biasanya digunakan untuk menghubungkan dan menganalisis data bisnis dari berbagai sumber. Gudang data adalah inti dari sistem BI yang dibangun untuk analisis dan pelaporan data. Gudang data merupakan perpaduan berbagai teknologi dan komponen yang membantu pemanfaatan data secara strategis.

Karakteristik Data Warehousing

Subject oriented

Data disusun berdasarkan subjek yang rinci, seperti penjualan, produk, atau pelanggan, yang hanya berisi informasi yang relevan untuk mendukung pengambilan keputusan.

Integrated.

Integrasi berkaitan erat dengan orientasi subjek. Gudang data harus menempatkan data dari berbagai sumber ke dalam format yang konsisten.

Time variant (time series)

DW menyimpan data historis. Data tersebut belum tentu memberikan status terkini.

Nonvolatile.

Setelah data dimasukkan ke dalam gudang data, pengguna tidak dapat mengubah atau memperbarui data.

Karakteristik Data Warehousing

Web based.

DW biasanya dirancang untuk menyediakan lingkungan komputasi yang efisien untuk aplikasi berbasis Web.

Relational/ multidimensional.

DW menggunakan struktur relasional atau struktur multidimensi

Client/server.

DW menggunakan arsitektur klien/server untuk menyediakan akses mudah bagi pengguna akhir.

Real time.

DW terbaru menyediakan kemampuan akses dan analisis data secara real-time, atau aktif.

Include metadata.

DW berisi metadata (data tentang data) tentang bagaimana data diorganisasikan dan bagaimana menggunakannya secara efektif.

Type Data Warehouse



**Data Marts
(DMs)**



**Operational
Data Stores
(ODS)**



**Enterprise Data
Warehouses (EDW)**



Data Mart

- Sementara DW menggabungkan basis data di seluruh perusahaan, data mart (DM) biasanya lebih kecil dan berfokus pada subjek atau departemen tertentu.
- DM adalah bagian dari gudang data, biasanya terdiri dari satu area subjek (misalnya, pemasaran, operasional).
- Data mart independen adalah gudang kecil yang dirancang untuk unit bisnis strategis atau departemen, tetapi sumbernya bukan EDW.

Operational Data stores

- Penyimpanan data operasional (ODS) menyediakan berkas informasi pelanggan dalam bentuk yang relatif baru.
- Jenis basis data ini sering digunakan sebagai area persiapan sementara untuk gudang data.
- ODS mirip dengan memori jangka pendek karena hanya menyimpan informasi yang sangat baru.
- ODS mengkonsolidasikan data dari berbagai sistem sumber dan menyediakan tampilan terintegrasi yang hampir seketika (real-time) dari data terkini yang volatil.

Enterprise Data Warehouses (EDW)

- Gudang data perusahaan (EDW) adalah gudang data berskala besar yang digunakan di seluruh perusahaan untuk mendukung pengambilan keputusan.
- Sifat EDW yang berskala besar memungkinkan integrasi data dari berbagai sumber ke dalam format standar untuk aplikasi BI dan pendukung keputusan yang efektif.



Metadata

- Metadata adalah data tentang data.
- Metadata menggambarkan struktur dan beberapa makna tentang data, sehingga berkontribusi pada penggunaan data yang efektif atau tidak efektif.



Data Warehousing Process

Banyak organisasi perlu membuat gudang data—penyimpanan data besar-besaran yang berisi data deret waktu untuk mendukung pengambilan keputusan. Data diimpor dari berbagai sumber daya eksternal dan internal, lalu dibersihkan dan diorganisasikan sesuai dengan kebutuhan organisasi.



Data Warehouse Process

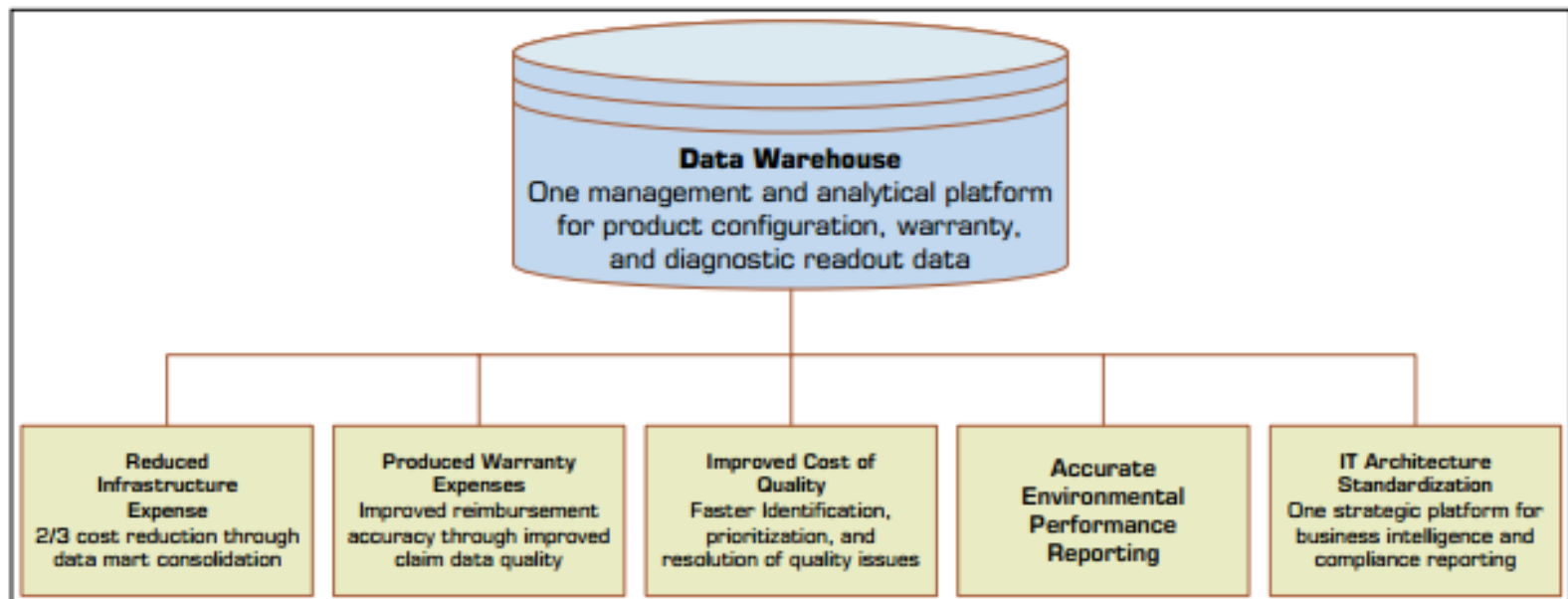


FIGURE 3.3 Data-Driven Decision Making—Business Benefits of the Data Warehouse. Source: Teradata Corp.

Data Warehouse Framework

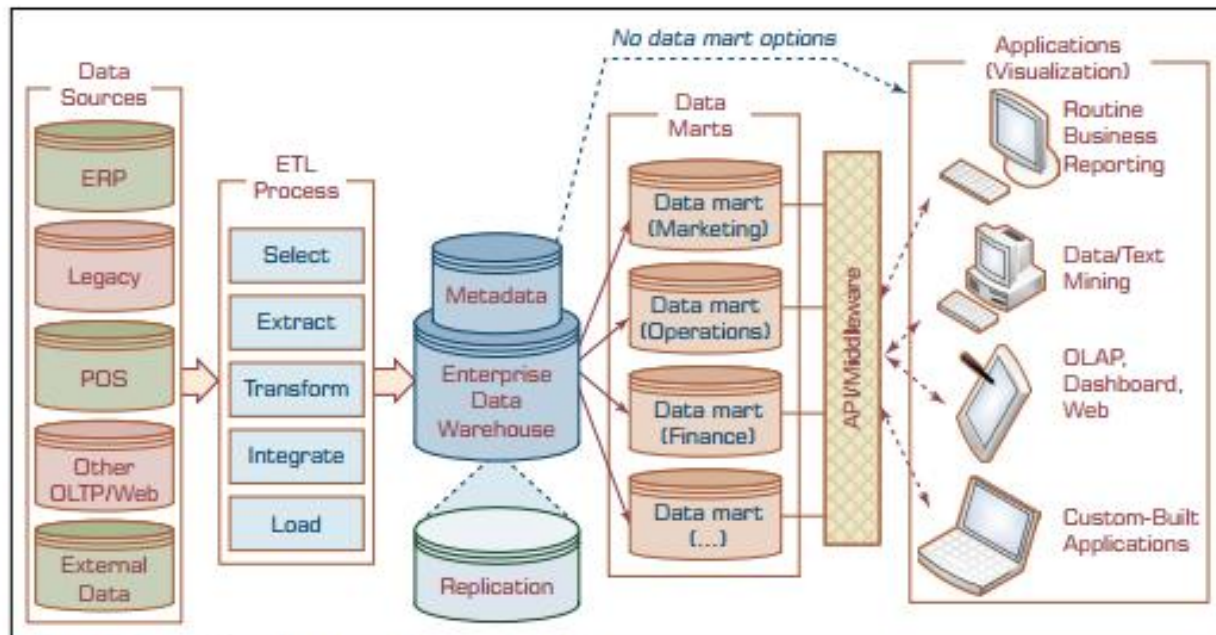


FIGURE 3.4 A Data Warehouse Framework and Views.

Data Warehouse Framework (1)

- **Data sources:** Data bersumber dari beberapa sistem operasional “legacy” yang independen dan mungkin juga dari penyedia data eksternal. Data juga dapat berasal dari sistem OLTP atau system enterprise resource planning (ERP). Data web dalam bentuk log web juga dapat dimasukkan ke gudang data.
- **Data extraction and transformation:** Data diekstraksi dan ditransformasi dengan tepat menggunakan perangkat lunak khusus atau komersial yang disebut ETL

Data Warehouse Framework (2)

- **Data loading:** Data dimuat ke dalam area pementasan, tempat data tersebut ditransformasi dan dibersihkan. Data kemudian siap untuk dimuat ke dalam ke data warehouse dan atau Data Marts.
- **Comprehensive database:** Pada dasarnya, ini adalah EDW untuk mendukung semua analisis keputusan dengan menyediakan informasi ringkasan dan terperinci yang relevan yang berasal dari berbagai sumber.

Data Warehouse Framework (3)

- **Metadata:** Metadata dikelola agar dapat dinilai oleh personel TI dan pengguna. Metadata mencakup program perangkat lunak tentang data dan aturan untuk mengatur ringkasan data yang mudah diindeks dan dicari, terutama dengan perangkat lunak web.
- **Middleware tools:** Middleware memungkinkan akses ke gudang data. Pengguna berpengalaman seperti analis dapat menulis kueri SQL mereka sendiri.

Arsitektur Data Warehousing

- Mengacu pada penggunaan arsitektur n-tier untuk Data Warehousing.
- Hoffer, Prescott, dan McFadden (2007) membedakan arsitektur-arsitektur ini dengan membagi DW menjadi tiga bagian:
 - **Data Warehouse** : berisi data dan perangkat lunak terkait.
 - **Data acquisition (back-end) software**: mengekstrak data dari sistem lama dan sumber eksternal, mengkonsolidasi dan meringkasnya, dan memuatnya ke dalam gudang data.
 - **Client (front-end) software**: memungkinkan pengguna untuk mengakses dan menganalisis data dari gudang

Data Warehousing Architecture (1)

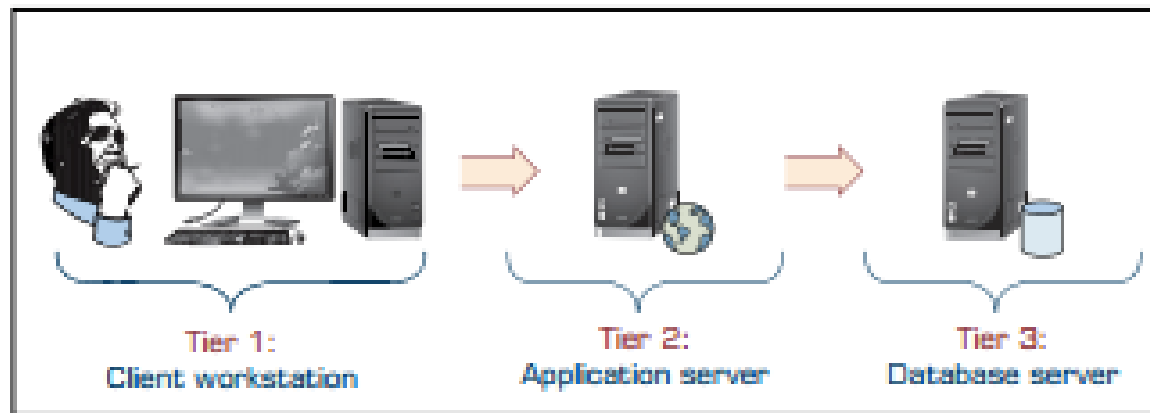


FIGURE 3.5 Architecture of a Three-Tier Data Warehouse.



Data Warehousing Architecture (2)

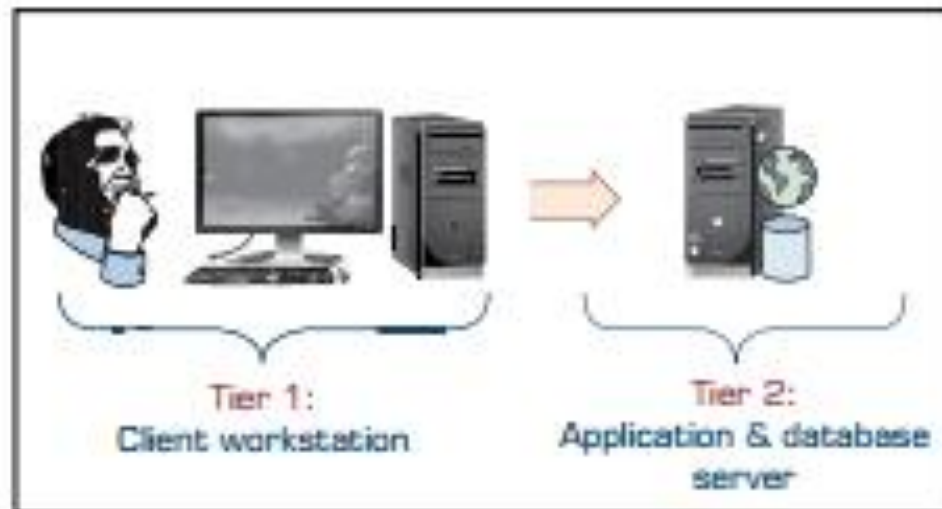


FIGURE 3.6 Architecture of a Two-Tier Data Warehouse.

Data Warehousing Architecture (3)

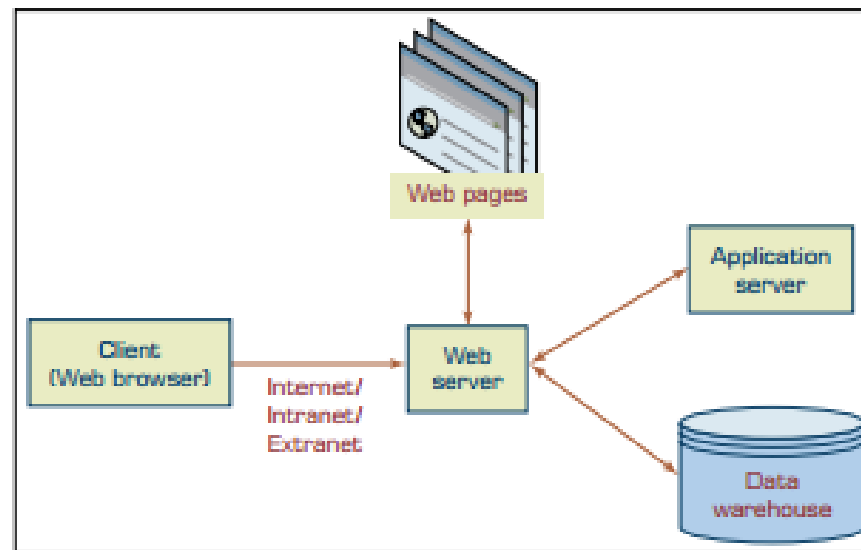


FIGURE 3.7 Architecture of Web-Based Data Warehousing.

Data Warehousing Architecture

Beberapa hal perlu dipertimbangkan saat memutuskan arsitektur mana yang akan digunakan.

- Sistem manajemen basis data (DBMS) mana yang sebaiknya digunakan?
- Apakah pemrosesan paralel dan/atau partisi akan digunakan?
- Apakah alat migrasi data akan digunakan untuk memuat gudang data?
- Alat apa yang akan digunakan untuk mendukung pengambilan dan analisis data?

Akhir bagian Part 5

