



Pertemuan #2

Arsitektur Data Warehouse

Praktikum Data Warehouse



Agenda



Arsitektur DW



Komponen DW



Tabel Fakta dan Tabel Dimensi

#1

Arsitektur DW

Apa itu arsitektur data warehouse?

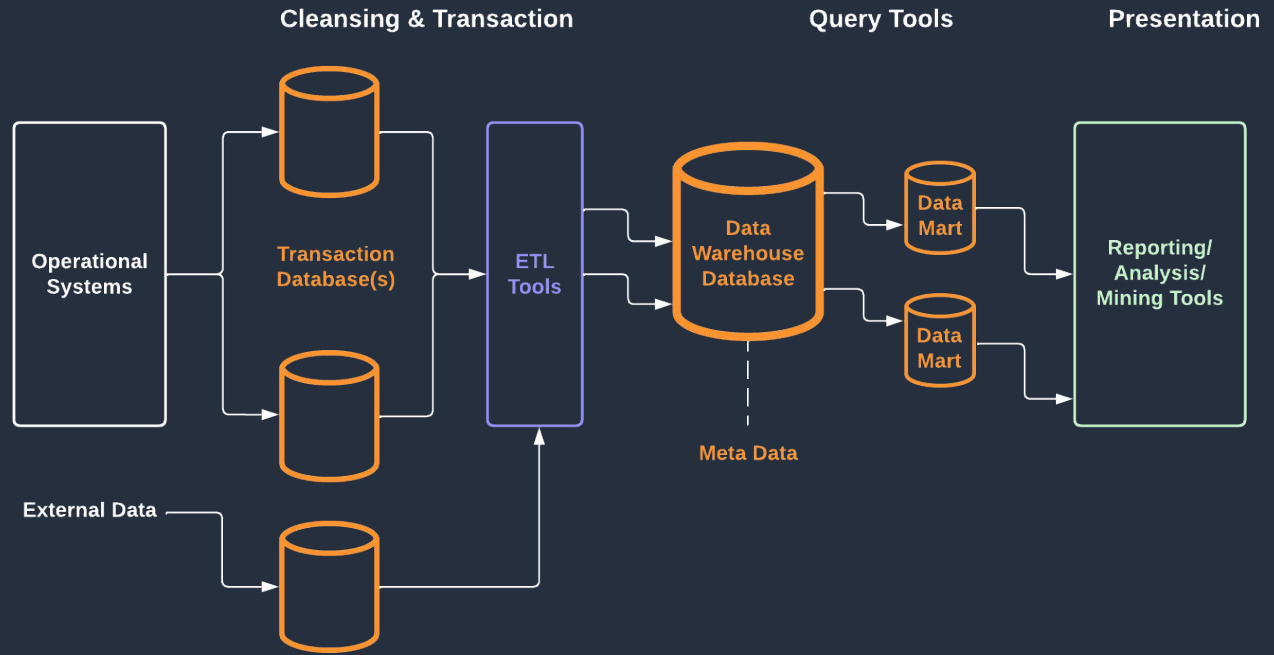
Arsitektur data warehouse merujuk pada kerangka kerja atau susunan komponen yang diperlukan untuk membangun sebuah data warehouse.

→ Pengintegrasian berbagai teknologi dan alat untuk **mengekstrak, mentransformasi, dan memuat data** ke dalam **repositori terpusat** dengan **tujuan analisis dan pelaporan**.

Komponen Arsitektur

Pada umumnya terdiri atas komponen berikut:

- Sumber data operasional, database eksternal, *flat-file*, atau data yang bersumber dari *cloud* lainnya.
- Sourcing, Acquisition, Clean-up and Transformation (ETL) Tools, untuk pemrosesan data ke DW.
- Query Tools, untuk alat interaksi dan manipulasi data dari database.
- Presentation, sebagai alat pelaporan dan visualisasi bagi pengguna akhir.



Komponen Arsitektur

ETL Tools

- Memindahkan data dari berbagai sumber, memprosesnya dengan melakukan transformasi.
- Memuat data ke dalam target database atau sistem.
- Menggabungkan data dari berbagai sumber yang berbeda.
- Menjaga integritas data, memperbaiki kesalahan data, dan meminimalkan duplikasi data.

Query Tools

- Dibagi berdasarkan kategori: Query & Reporting, Apps Development, Data Mining, dan OLAP.
- Melakukan pengambilan data atau querying pada database.
- Memungkinkan pengguna untuk mencari data yang diperlukan untuk tujuan analisis atau pemrosesan data.
- Menggabungkan dan memanipulasi data, melakukan agregasi data, dan membuat laporan.

Komponen Arsitektur

Data Mart

- Subset dari data warehouse yang terdiri dari data yang disesuaikan untuk kebutuhan analisis tertentu, seperti untuk departemen atau fungsi bisnis tertentu dalam perusahaan.
- Dibuat dengan cara memilih dan memproses data dari data warehouse yang relevan dan spesifik untuk kebutuhan tertentu.

Data Lake

- Tempat penyimpanan data yang memiliki konsep menyimpan semua jenis data perusahaan dalam bentuk mentah (raw) dan tidak diproses.
- Menyimpan data dalam jumlah besar dan beragam dari berbagai sumber, seperti sistem bisnis, media sosial, dan sensor IoT, dalam satu lokasi yang terpusat.
- Menggunakan teknologi cloud storage untuk menyimpan data dalam skala besar dan murah.

Komponen Arsitektur

Metadata

Informasi atau deskripsi tentang data dan hubungannya.

- Biasanya disimpan dalam sistem manajemen metadata atau *metadata repository*.
- Panduan pemetaan data pada saat transform/ diubah dari lingkup operasional ke data warehouse.

employee_id	first_name	last_name	nin	department_id
44	Simon	Martinez	HH 45 09 73 D	1
45	Thomas	Goldstein	SA 75 35 42 B	2
46	Eugene	Comelsen	NE 22 63 82	2
47	Andrew	Petculescu	XY 29 87 61 A	1
48	Ruth	Stadick	MA 12 89 36 A	15
49	Barry	Scardelis	AT 20 73 18	2
50	Sidney	Hunter	HW 12 94 21 C	6
51	Jeffrey	Evans	LX 13 26 39 B	6
52	Doris	Bemdt	YA 49 88 11 A	3
53	Diane	Eaton	BE 08 74 68 A	1
54	Bonnie	Hall	WW 53 77 68 A	15
55	Taylor	Li	ZE 55 22 80 B	1

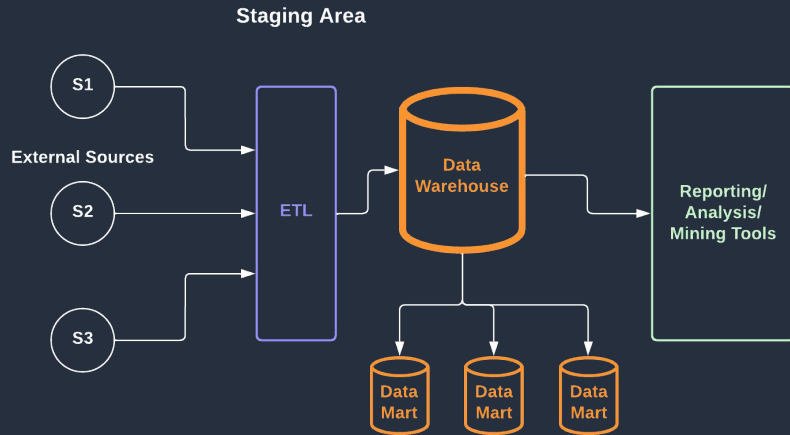
Data

Metadata

Column	Data Type	Description
employee_id	int	Primary key of a table
first_name	nvarchar(50)	Employee first name
last_name	nvarchar(50)	Employee last name
nin	nvarchar(15)	National Identification Number
position	nvarchar(50)	Current position title, e.g. Secretary
department_id	int	Employee department. Ref: Departments
gender	char(1)	M = Male, F = Female, Null = unknown
employment_start_date	date	Start date of employment in organization.
employment_end_date	date	Employment end date. Null if employee still

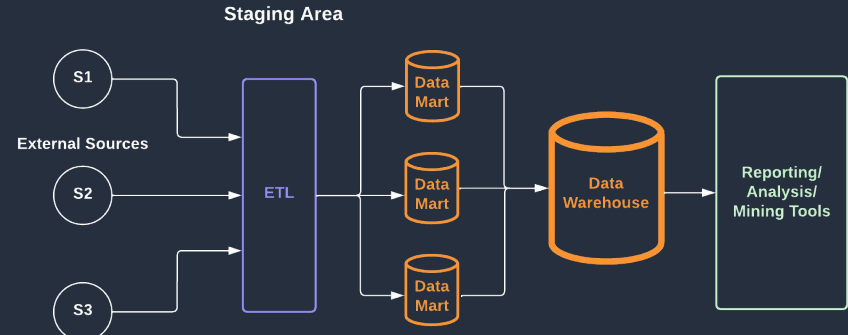
Pendekatan pembangunan DW

Top-down



(Bill Inmon)

Bottom-up



(Ralph Kimball)

Pendekatan pembangunan DW

Top-down

Kelebihan:

- Data Mart terbentuk dari Data Warehouse sehingga menyediakan dimensional view yang konsisten.
- Model yang kuat terhadap perubahan bisnis.
- Kemudahan pembuatan Data Mart dari Data Warehouse

Kekurangan:

- Membutuhkan biaya, waktu, dan sumber daya yang tinggi

Bottom-up

Kelebihan:

- Data Mart yang dibuat terlebih dahulu mempercepat proses pembuatan laporan
- Dapat mengakomodasi Data Mart lebih yang kemudian memperluas Data Warehouse
- Membutuhkan biaya, waktu, dan sumber daya yang lebih rendah

Kekurangan:

- Konsistensi lebih rendah dibanding Top-Down

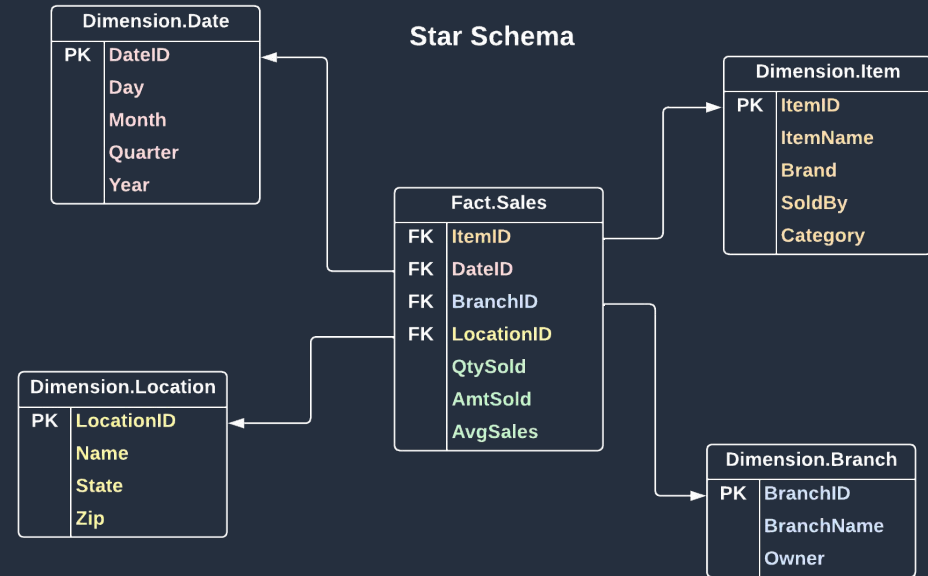
#2

Tabel Fakta dan Tabel Dimensi

Tabel Fakta/ Fact Table

Merupakan tabel utama berisi fakta atau angka terukur dan dapat dihitung.

- Berisi nilai dari sebuah kejadian atau transaksi tertentu misalnya persediaan produk, penjualan, atau pengiriman.
- Memiliki *foreign key* → tabel dimensi.
- Menyimpan nilai numerik dibanding karakter.
- Setiap baris pada tabel fakta merepresentasikan satu nilai/ kuantitas tunggal.
- Berisi data fakta pada tingkat detail atau agregat.

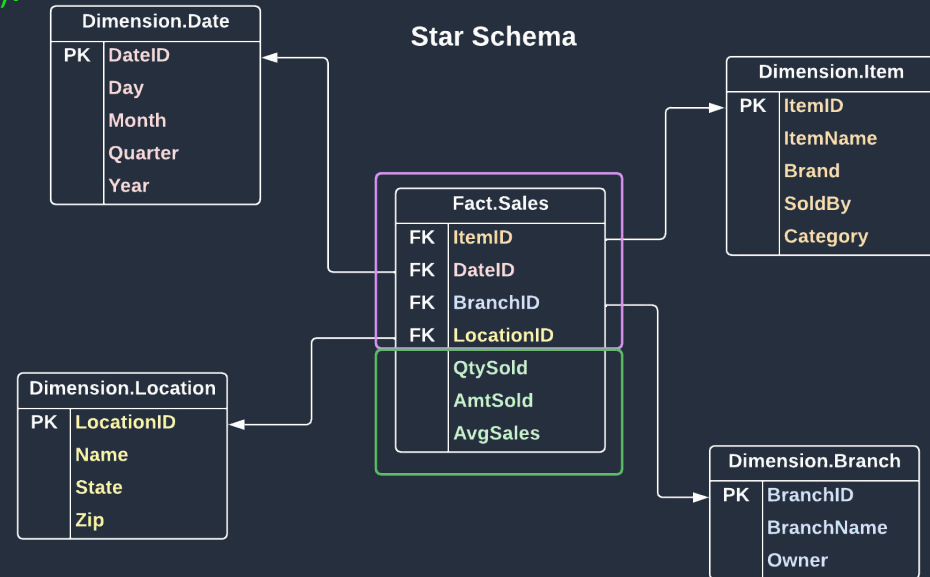


Tabel Fakta/ Fact Table

Tabel fakta biasanya memiliki dua jenis kolom: FOREIGN KEY ke tabel dimensi (kotak ungu) dan MEASURES yang berisi fakta numerik (kotak hijau).

Contoh Measures:

Misalnya kalian punya bisnis keripik, berarti *measure* nya adalah “Jumlah penjualan keripik” atau “Rata-rata penjualan keripik”.



Agregasi Fakta

Individual

FactSalesTransaction
TransactionID
TransactionAmount



Daily Totals

FactSalesDay
Date
TotalSales



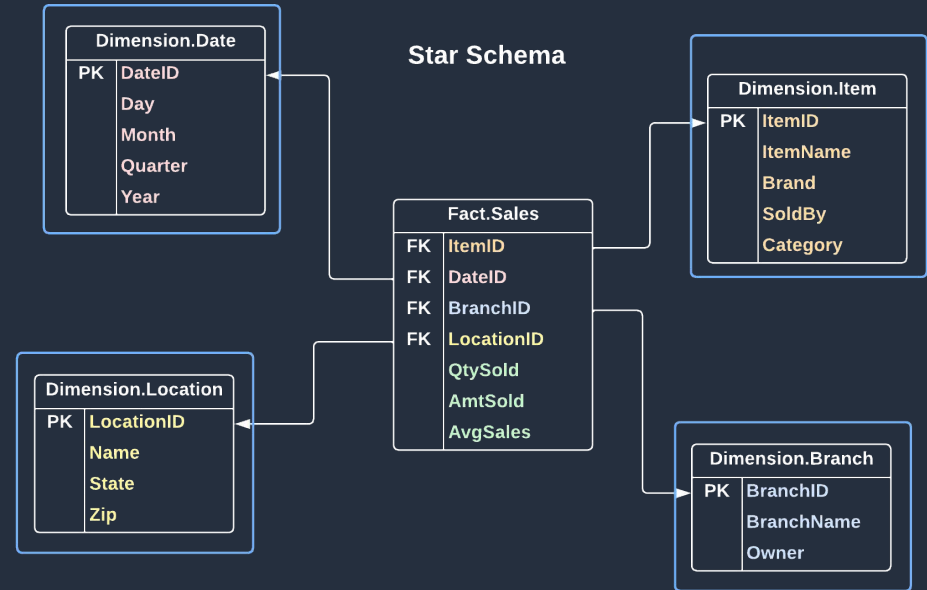
Monthly Totals

FactSalesMonth
Month
TotalSales

Tabel Dimensi/ Dimension Table

Berisi informasi tentang karakteristik dari data yang tersimpan dalam tabel fakta, seperti produk, waktu, dan lokasi.

- Memiliki kolom-kolom yang disebut sebagai atribut dimensi atau dimensi.
- Digunakan untuk menyimpan informasi deskriptif dari fakta.
- Waktu dan lokasi merupakan contoh dimensi yang umum.
- Biasanya menyimpan variasi *formatting* dari dimensi.



Karakteristik Tabel Dimensi

- **Dimension table key** – Primary key dari tabel dimensi mengidentifikasi setiap baris dimensi secara unik.
- **Tabel bersifat lebar** – Sebuah tabel dimensi dapat memiliki banyak atribut atau kolom.
- **Hubungan antar atribut** – Pada skema DW terdapat beberapa dimensi yang saling berhubungan dan ada juga yang tidak.
- **Tidak ternormalisasi** – Untuk efisiensi, sebaiknya query menggunakan atribut yang disalurkan secara langsung dari tabel dimensi ke tabel fakta, tanpa melalui tabel lainnya.

Tugas (tidak dikumpulkan)

Lakukan instalasi SQL Server 2022, kemudian *restore* sampel DW database AdventureWorks (AdventureWorksDW2019.bak) pada laman [berikut](#).

Lakukan eksplorasi terhadap tabel fakta dan dimensi dalam DW AdventureWorks.

Link Video <https://www.youtube.com/watch?v=8IK99x7Y7Mo>

Thank you!

