



Kampus
Merdeka
INDONESIA JAYA

BUSINESS INTELLIGENCE

Dimensional Modelling

FAST: FOCUS, ADAPTIVE, SINCERE, TRUSTWORTHY

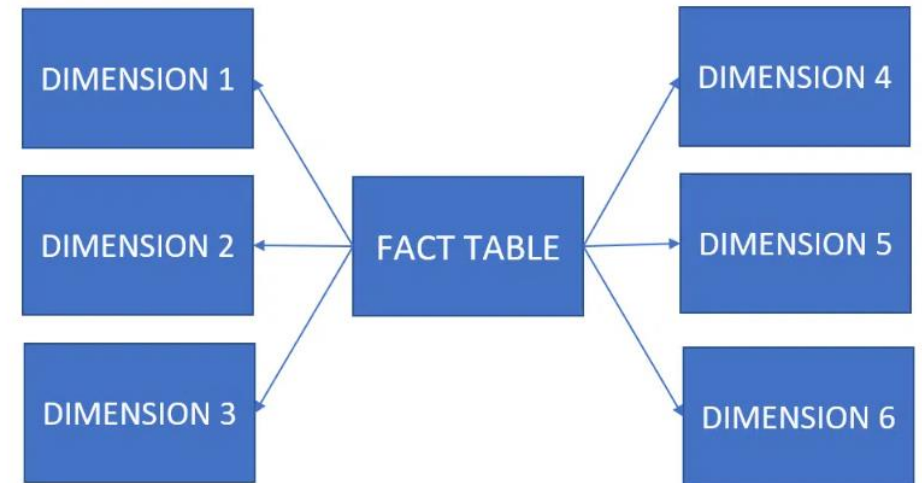
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG

Data Warehouse Modeling

- Data warehouse modeling adalah proses merancang dan mengorganisir model data warehouse.
- Proses ini mencakup perancangan skema database yang tepat sehingga data dapat ditransformasi dan disimpan dengan cara yang bermanfaat bagi end user.
- Teknik yang paling sering digunakan adalah *dimensional modeling*

Dimensional Modeling

- Dimensional modeling adalah teknik pemodelan data yang menyusun data dalam skema yang dapat optimal digunakan untuk proses analisa dan reporting.
- Metode ini melibatkan pengorganisasian data ke dalam fact dan dimension.
- **Fact** digunakan untuk menyimpan data kuantitatif, sedangkan **dimension** untuk mendeskripsikan data.



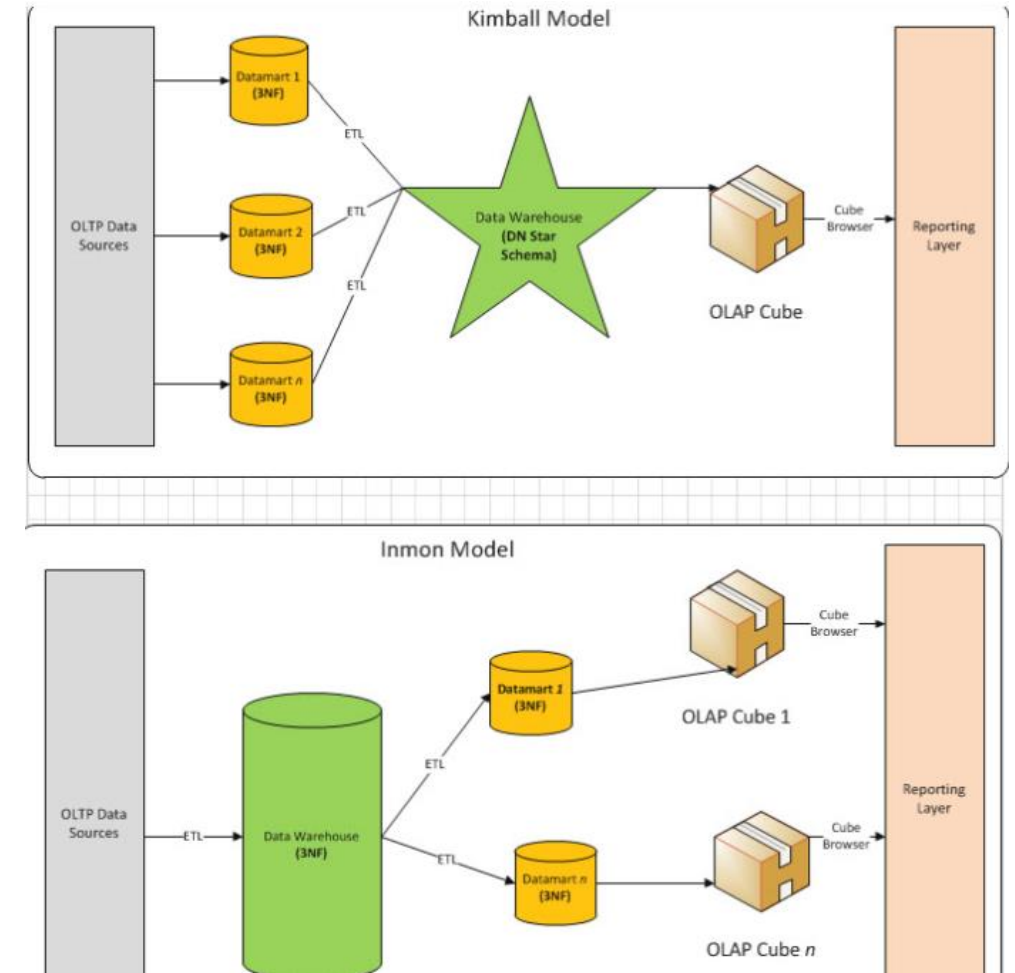
Teknik Dimensional Modeling

Kimball Approach dan **Inmon Approach** adalah dua metodologi paling populer

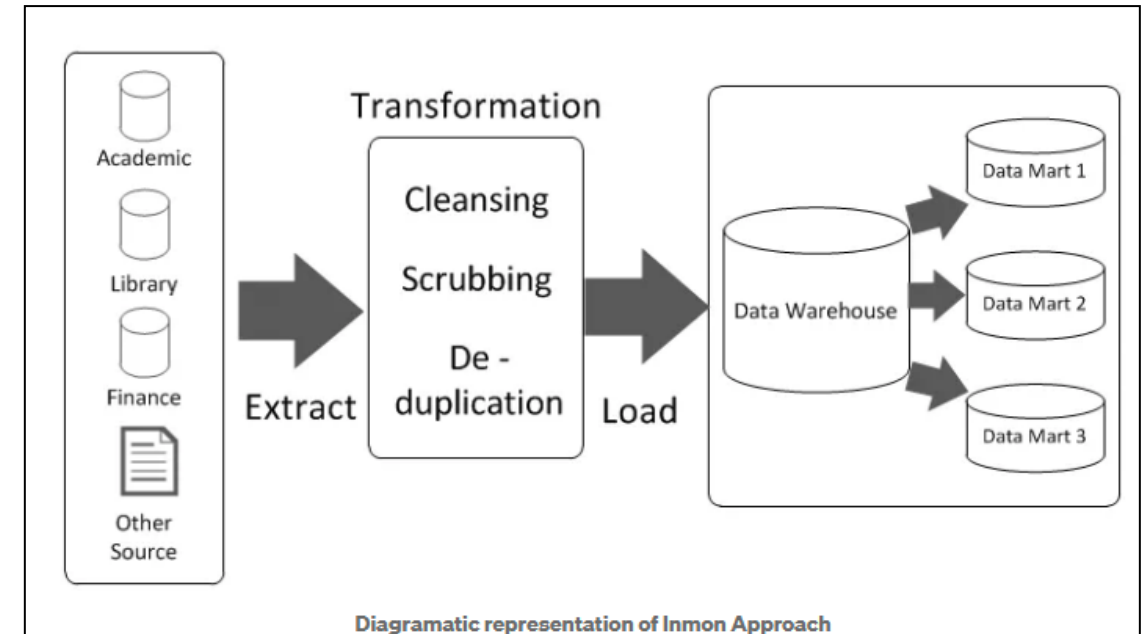
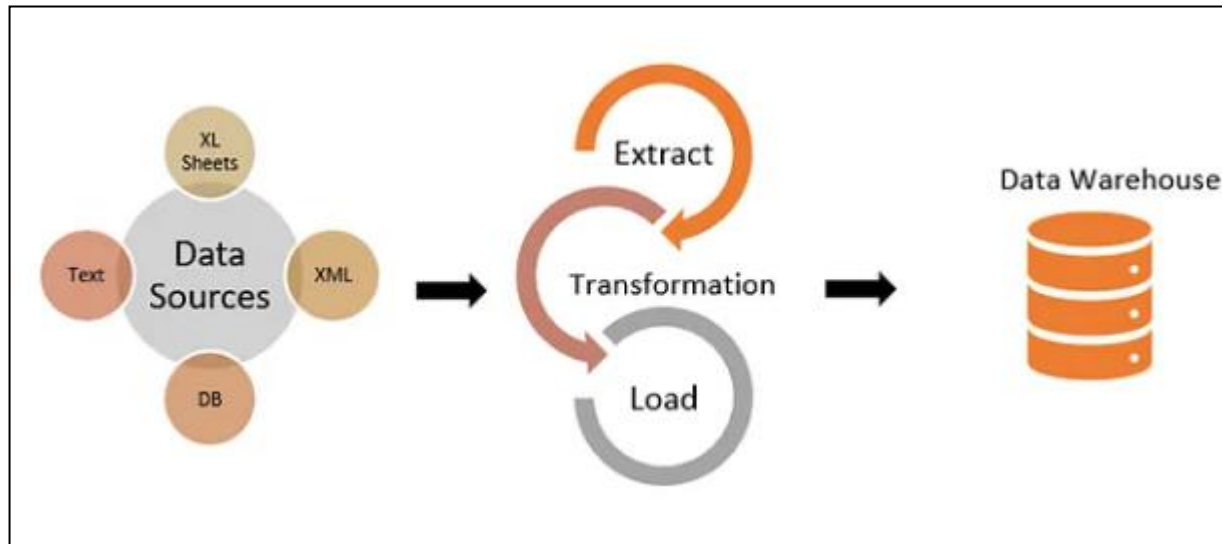
Ralph Kimball



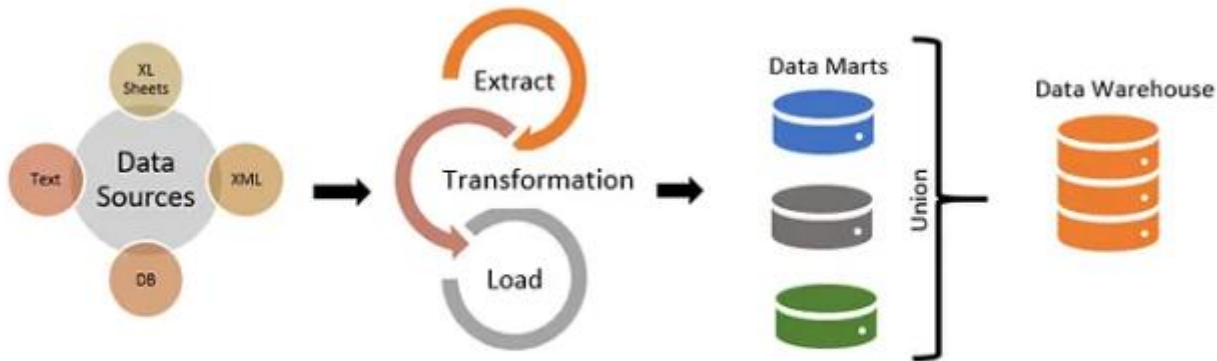
Bill Inmon



Inmon Methodology



Kimball Methodology



Diagrammatic representation of Kimball Approach

4 Step Dimensional Design Process:

- Select the business process
- Declare the grain
- Identify the dimensions
- Identify the facts



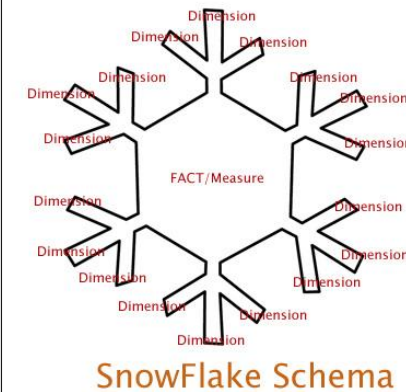
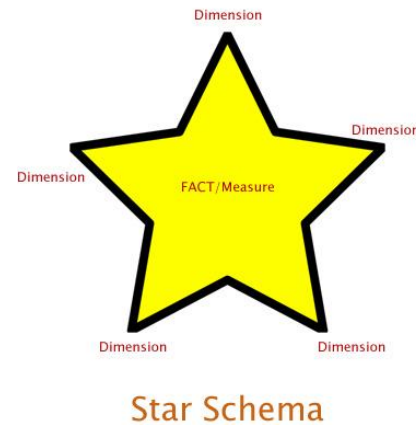
Bagian 3

Skema Data Warehouse

- Perbedaan terpenting antara database OLTP dan data warehouse adalah **SKEMANYA**
- Database OLTP → Dioptimalkan untuk operasi TULIS.
- Data Warehouse → Dioptimalkan untuk operasi BACA
- Ada 3 Skema Data Warehouse yang umum digunakan:
 - Star Schema
 - Snowflake Schema
 - Star Constellation Schema

Star & Snowflake Schema

- **Star Schema:** Tipe skema database di mana suatu fact table berelasi dengan sejumlah dimension table, membentuk pola seperti bintang.
- **Snowflake Schema:** Variasi lebih kompleks dari star schema dengan dimension table yang dinormalisasi sehingga membentuk sejumlah table yang saling terhubung dengan pola seperti salju.



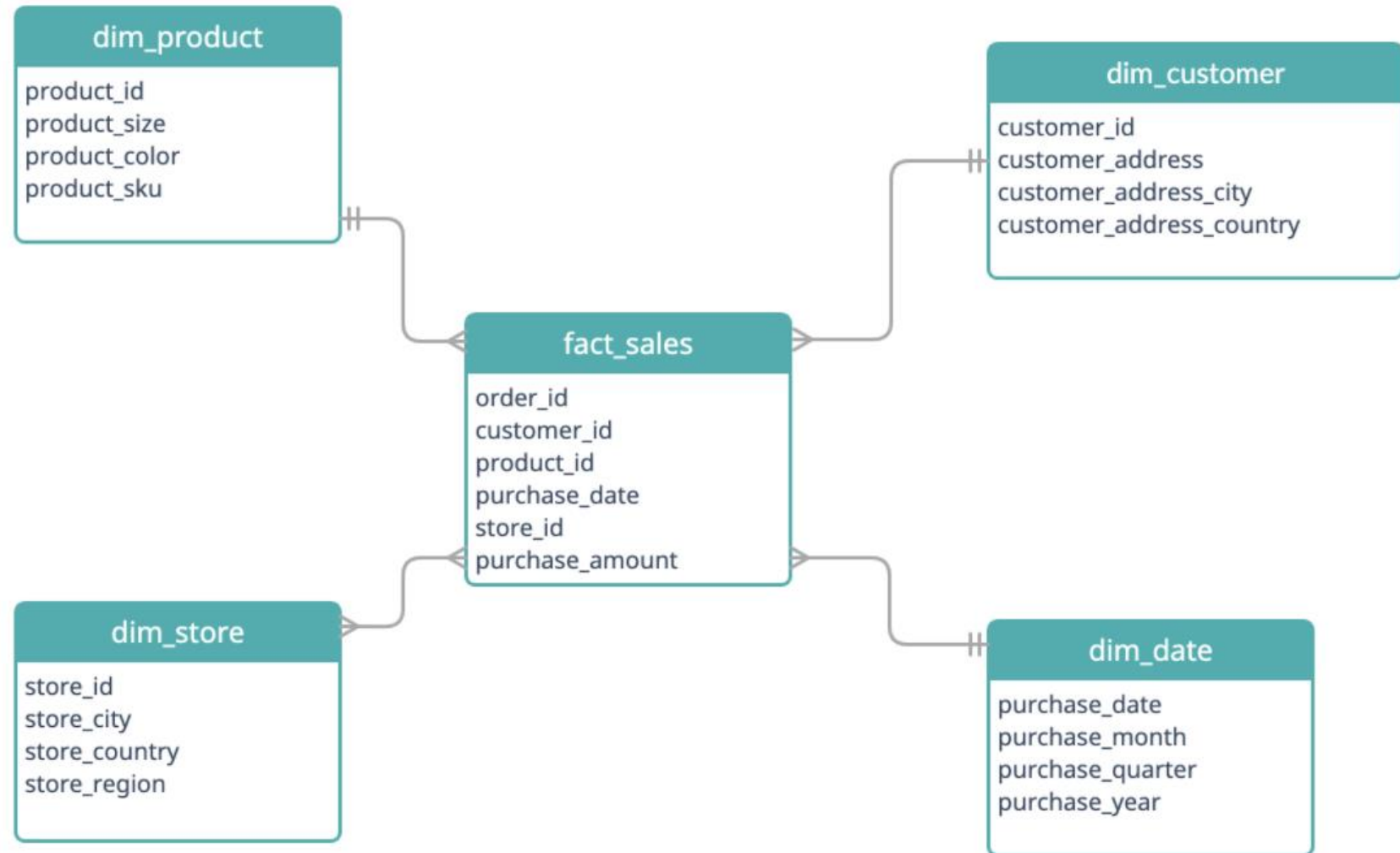
Fact Table

- Menyimpan **data kuantitatif** yang diukur/dianalisa
- Bergantung pada **proses bisnis**, misalnya penjualan, peminjaman, booking, pendaftaran
- Biasanya setiap baris dalam tabel fakta **mewakili sebuah peristiwa atau transaksi** yang diukur/dianalisa
- Terdiri dari key (**primary key, foreign key**) dan **data kuantitatif**

Dimension Table

- Menyimpan informasi tentang **dimensi atau karakteristik/deskripsi** dari data yang diukur
- Dimensi adalah **sudut pandang** yang berbeda dari data, seperti waktu, Lokasi, produk, atau pelanggan
- Terdiri dari **primary key dan atribut deskriptif**
- Berelasi **one-to-many** dengan fact table, di mana satu baris dalam dimension table dapat berhubungan dengan banyak baris dalam fact table).

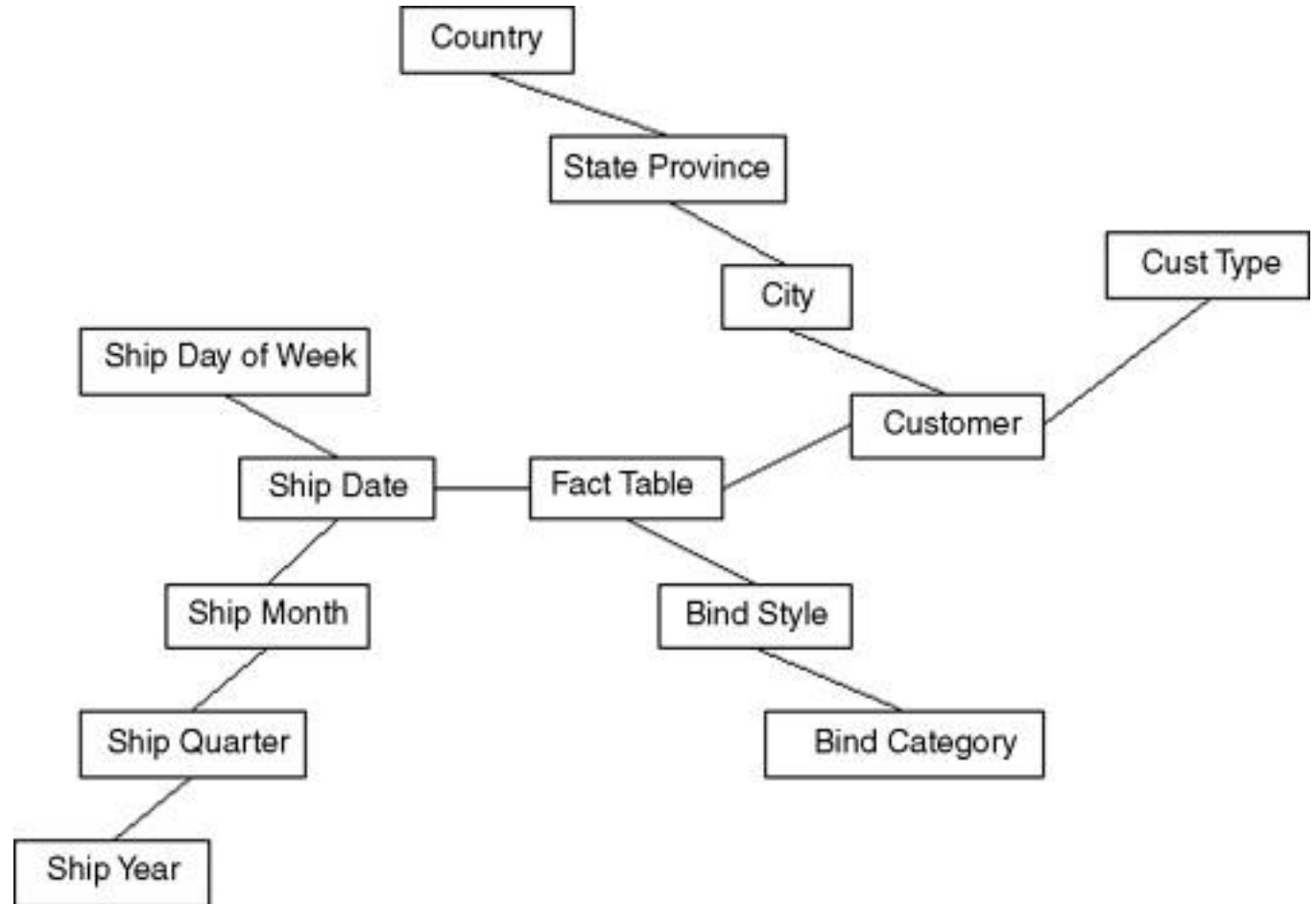
Star Schema



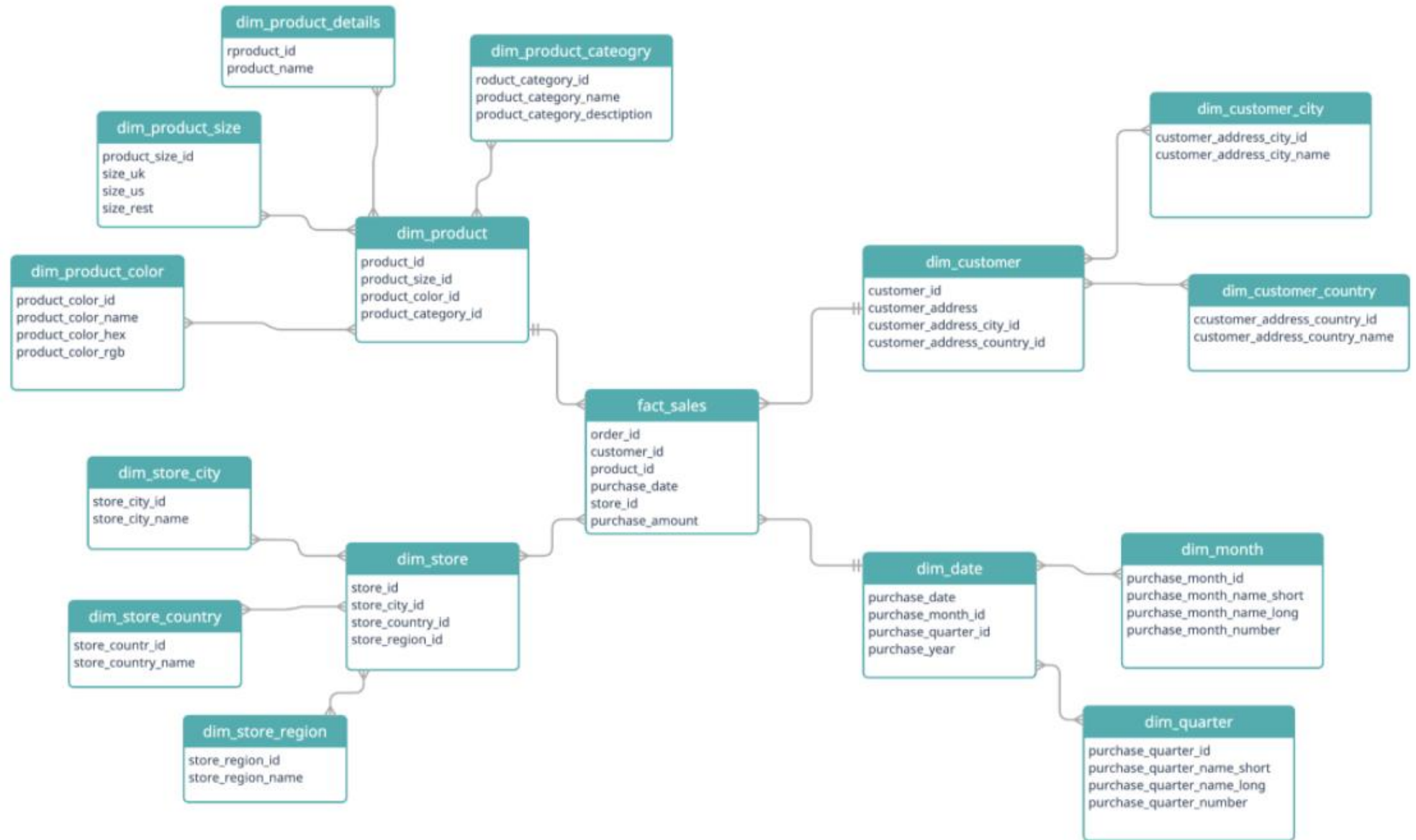
Snowflake Schema

Dalam snowflake schema, dimension table dipecah menjadi "cabang" yang lebih detail atau spesifik → normalized atau denormalized

Dimension table berelasi dengan fact table dan dimension table lainnya



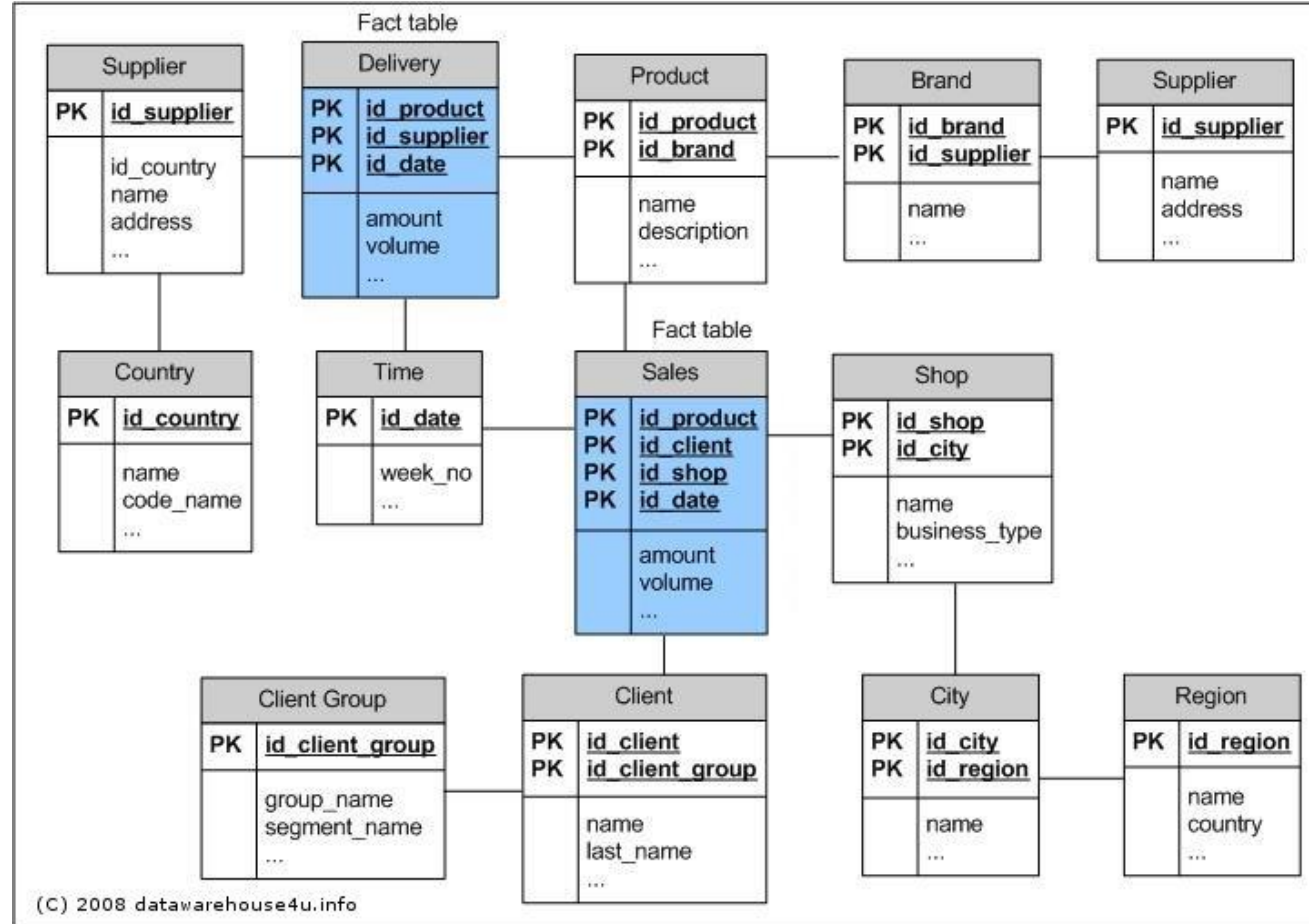
Snowflake Schema



Star Vs Snowflake Schema

Star Schema	Snowflake Schema
Denormalized dimension	Normalized & denormalized dimension
Desain lebih simple	Desain lebih kompleks
Query lebih sederhana	Query lebih kompleks
Eksekusi query lebih cepat	Eksekusi query lebih lama
High	Low
Higher storage	Lower storage
Integritas data lebih buruk	Integritas data lebih baik
Maintanance lebih sulit	Maintenance lebih mudah

Fact Constellation/Galaxy Schema



Conformed Dimension

- Conformed Dimension adalah dimensi yang dipakai bersama oleh beberapa data mart, sehingga memudahkan integrasi dan konsistensi data dalam Data Warehouse.
- Manfaat:
 - Pengurangan duplikasi: Menghindari duplikasi definisi dimensi yang sama sehingga menghemat ruang dan meningkatkan efisiensi
 - Konsistensi data: memastikan bahwa dimensi yang sama di seluruh data marts memiliki arti yang sama
 - Kemudahan integrasi: mempermudah analisis data dari berbagai sumber

Conformed Dimension

Product Description	Open Orders Qty	Inventory Qty	Sales Qty
Baked Well Sourdough	1,201	935	1,042
Fluffy Light Sliced White	1,472	801	922
Fluffy Sliced Whole Wheat	846	513	368

Bus Matrix

- Bus Matrix → matrix yang digunakan dalam desain Data Warehouse untuk merencanakan dan mendokumentasikan hubungan antara fact tables dan dimension tables di seluruh data marts yang berbeda.

Business Process / Event	Time	Customer	Service	Rate Category	Local Svc Provider	Calling Party	Called Party	Long Dist Provider	Internal Organization	Employee	Location	Equipment Type	Supplier	Item Shipped	Account Status
Customer Billing	X	X	X	X	X		X			X				X	
Service Orders	X	X	X		X		X	X	X	X	X			X	
Trouble Reports	X	X	X		X	X	X	X	X	X	X	X	X	X	X
Yellow Page Ads	X	X		X		X		X	X	X				X	
Customer Inquiries	X	X	X	X	X	X	X	X	X	X				X	
Promotions & Communication	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Billing Call Detail	X	X	X	X	X	X	X	X		X	X	X	X	X	X
Network Call Detail	X	X	X	X	X	X	X	X		X	X	X	X	X	X
Customer Inventory	X	X	X	X	X		X	X		X	X	X	X	X	X
Network Inventory	X		X					X	X	X	X	X	X		
Real Estate	X							X	X	X	X				
Labor & Payroll	X							X	X	X					
Computer Charges	X	X	X		X		X	X	X	X	X	X	X		
Purchase Orders	X							X	X	X	X	X	X		
Supplier Deliveries	X							X	X	X	X	X	X		

Bus Matrix

- Dalam arsitektur Data Warehouse, sering ada banyak data marts yang dibuat untuk area bisnis yang berbeda (misalnya penjualan, keuangan, pemasaran).
- Setiap data mart memiliki fakta dan dimensi yang mungkin tumpang tindih atau digunakan bersama.
- Manfaat:
 - Mencatat dan mengelola conformed dimensions dan fakta apa saja yang digunakan oleh tiap data mart.
 - Semacam peta struktur data warehouse untuk komunikasi antara developer dan user

Struktur Bus Matrix

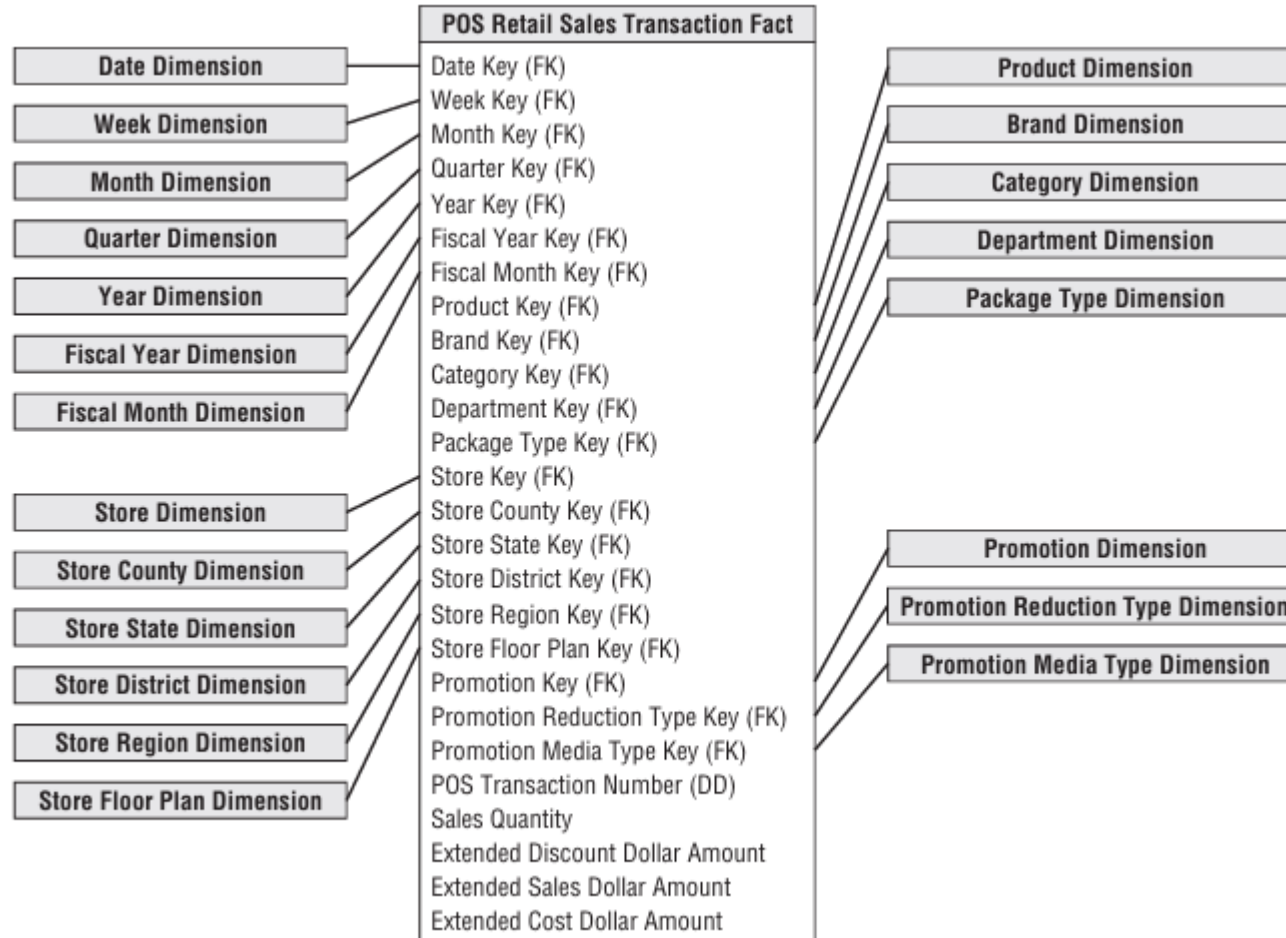
- Baris mewakili **fact tables**
- Kolom mewakili **dimension tables**
- Tanda centang (✓) atau tanda lain menunjukkan hubungan antara fakta dan dimensi

BUSINESS PROCESSES	Shared Dimensions									
	Date	Customer	Product	Department	Promotion	Reseller	Sales Territory	Employee	Account	Organization
Internet Sales	✓	✓	✓		✓		✓			
Reseller Sales	✓		✓		✓	✓	✓	✓		
Sales Plan	✓		✓				✓			
General Ledger	✓		✓	✓			✓		✓	✓
Inventory	✓		✓							
Customer Surveys	✓	✓								
Customer Service Calls	✓	✓	✓					✓		

Centipede Fact Table

- Centipede fact table → fact table dengan terlalu banyak foreign key yang merujuk pada low cardinality dimension table
- Denormalized dimension table memang membutuhkan space yang lebih tinggi, tetapi yang perlu diperhatikan adalah space untuk Foreign Key pada fact table.
- Bisnis proses secara umum dapat direpresentasikan dengan kurang dari 20 dimension table. Jika lebih dari itu, carilah cara untuk menggabungkan dimensi yang berhubungan menjadi satu dimensi table

Centipede Fact Table

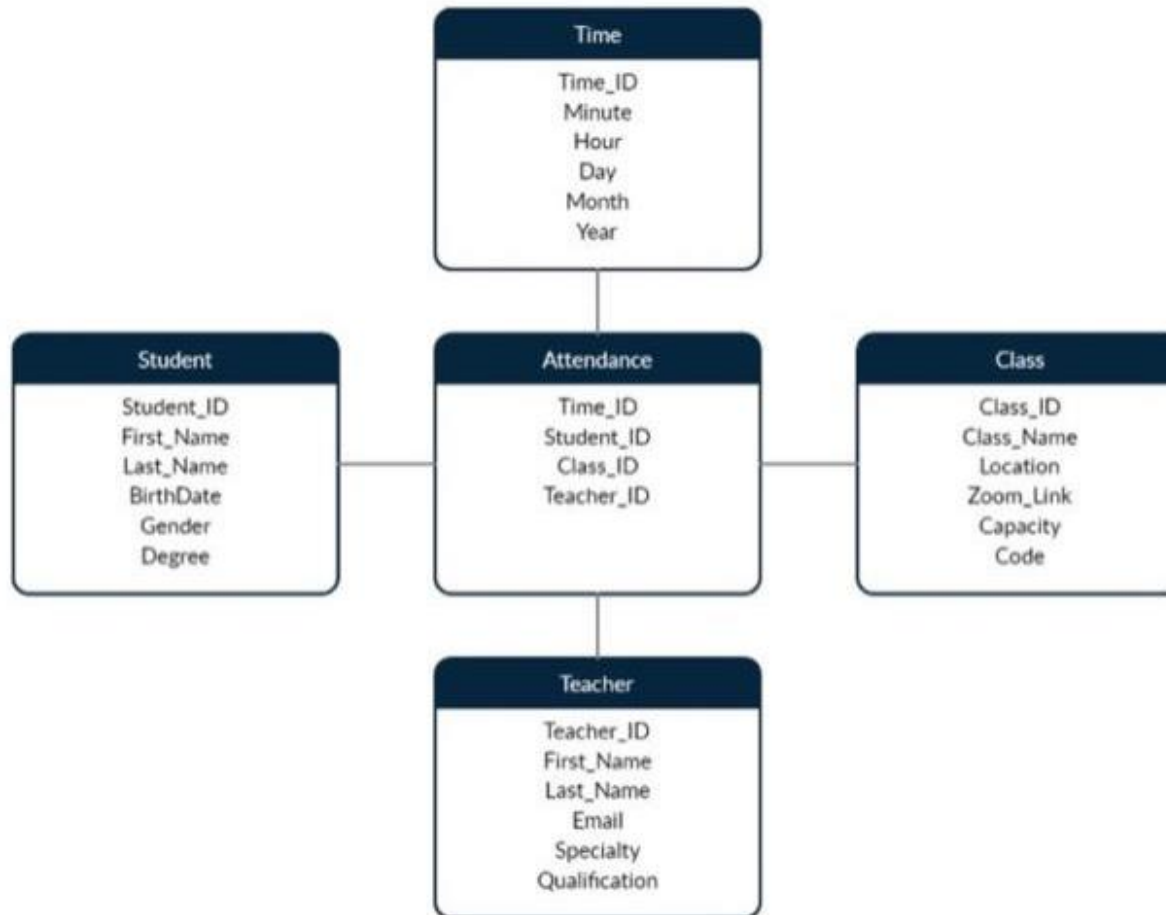




Factless Fact Table

- Factless fact table adalah fact table yang menyimpan relasi antardimensi tanpa ada numerical fact (measures).
- Umumnya digunakan untuk menyimpan data terkait kemunculan event/aktivitas, misalnya kehadiran siswa, kunjungan pelanggan, dst

Factless Fact Table



References

- The Data Warehouse Toolkit – The Definitive Guide to Dimensional Modelling 3rd edition by Ralph Kimball & Margy Ross
- <https://www.learnmsbitutorials.net/ssis-dimension-fact-table-snowflake-star-schema-part1.php>
- <https://builtin.com/articles/star-schema-vs-snowflake-schema>
- <https://www.integrate.io/blog/snowflake-schemas-vs-star-schemas-what-are-they-and-how-are-they-different/>
- <https://medium.com/cloudzone/inmon-vs-kimball-the-great-data-warehousing-debate-78c57f0b5e0e>



Terima kasih