

# Data Warehouse dan Data Mining

Oleh : Asep Jalaludin,S.T.,M.M.

# Data Warehouse

## Definisi :

- o Data Warehouse adalah Pusat repositori informasi yang mampu memberikan database berorientasi subyek untuk informasi yang bersifat historis yang mendukung DSS (Decision Suport System) dan EIS (Executive Information System).
- o Salinan dari transaksi data yang terstruktur secara spesifik pada query dan analisa.
- o Salinan dari transaksi data yang terstruktur spesifik untuk query dan laporan

## Tujuan :

Meningkatkan kualitas dan akurasi informasi bisnis dan mengirimkan informasi ke pemakai dalam bentuk yang dimengerti dan dapat diakses dengan mudah.

# Data Warehouse

- *Data warehouse* adalah basis data yang menyimpan data sekarang dan data masa lalu yang berasal dari berbagai sistem operasional dan sumber yang lain (sumber eksternal) yang menjadi perhatian penting bagi manajemen dalam organisasi dan ditujukan untuk keperluan analisis dan pelaporan manajemen dalam rangka pengambilan keputusan
- *Data warehouse* digunakan untuk mendukung pengambilan keputusan, bukan untuk melaksanakan pemrosesan transaksi
- *Data warehouse* hanya berisi informasi-informasi yang relevan bagi kebutuhan pemakai yang dipakai untuk pengambilan keputusan

# Perbedaan DW dan OLTP

OLTP	Data Warehouse
<ul style="list-style-type: none"><li>● Menangani data saat ini</li><li>● Data bisa saja disimpan pada beberapa platform</li><li>● Data diorganisasikan berdasarkan fungsi atau operasi seperti penjualan, produksi, dan pemrosesan pesanan</li><li>● Pemrosesan bersifat berulang</li><li>● Untuk mendukung keputusan harian (operasional)</li><li>● Melayani banyak pemakai operasional</li><li>● Berorientasi pada transaksi</li></ul>	<ul style="list-style-type: none"><li>● Lebih cenderung menangani data masa lalu</li><li>● Data disimpan dalam satu platform</li><li>● Data diorganisasikan menurut subjek seperti pelanggan atau produk</li><li>● Pemrosesan sewaktu-waktu, tak terstruktur, dan bersifat heuristik</li><li>● Untuk mendukung keputusan yang strategis</li><li>● Untuk mendukung pemakai manajerial yang berjumlah relatif sedikit</li><li>● Berorientasi pada analisis</li></ul>

# Sumber Data untuk DW

1. Data operasional dalam organisasi, misalnya basis data pelanggan dan produk, dan
  2. Sumber eksternal yang diperoleh misalnya melalui Internet, basis data komersial, basis data pemasok atau pelanggan
- o Berbagai data yang berasal dari sumber digabungkan dan diproses lebih lanjut oleh manajer *data warehouse* dan disimpan dalam basis data tersendiri.
  - o Selanjutnya, perangkat lunak seperti OLAP dan *data mining* dapat digunakan oleh pemakai untuk mengakses *data warehouse*

# Empat karakteristik data warehouse

- Subject oriented
- Integrated
- Time variant
- Non-volatile

Yang termasuk karakteristik tambahan data warehouse antara lain :

- Web based.
- Relational/multidimensional.
- Client/server.
- Real time.
- Include metadata.

# Empat karakteristik data warehouse

- Subject oriented
  - Data yang disusun menurut subyek berisi hanya informasi yang penting bagi pemrosesan decision support.
  - Database yang semua informasi yang tersimpan di kelompokkan berdasarkan subyek tertentu misalnya: pelanggan, gudang, pasar, dsb.
  - Semua Informasi tersebut disimpan dalam suatu sistem *data warehouse*.
  - Data-data di setiap subyek dirangkum ke dalam dimensi, misalnya : periode waktu, produk, wilayah, dsb, sehingga dapat memberikan nilai sejarah untuk bahan analisa.

# Empat karakteristik data warehouse

- Integrated
  - Jika data terletak pada berbagai aplikasi yang terpisah dalam suatu lingkungan operasional, encoding data sering tidak seragam sehingga bila data dipindahkan ke data warehouse maka coding akan diasumsikan sama seperti lazimnya.



# Empat karakteristik data warehouse

- Time variant

Data warehouse adalah tempat untuk storing data selama 5 sampai 10 tahun atau lebih, data digunakan untuk perbandingan atau perkiraan dan data ini tidak dapat diperbaharui.

# Empat karakteristik data warehouse

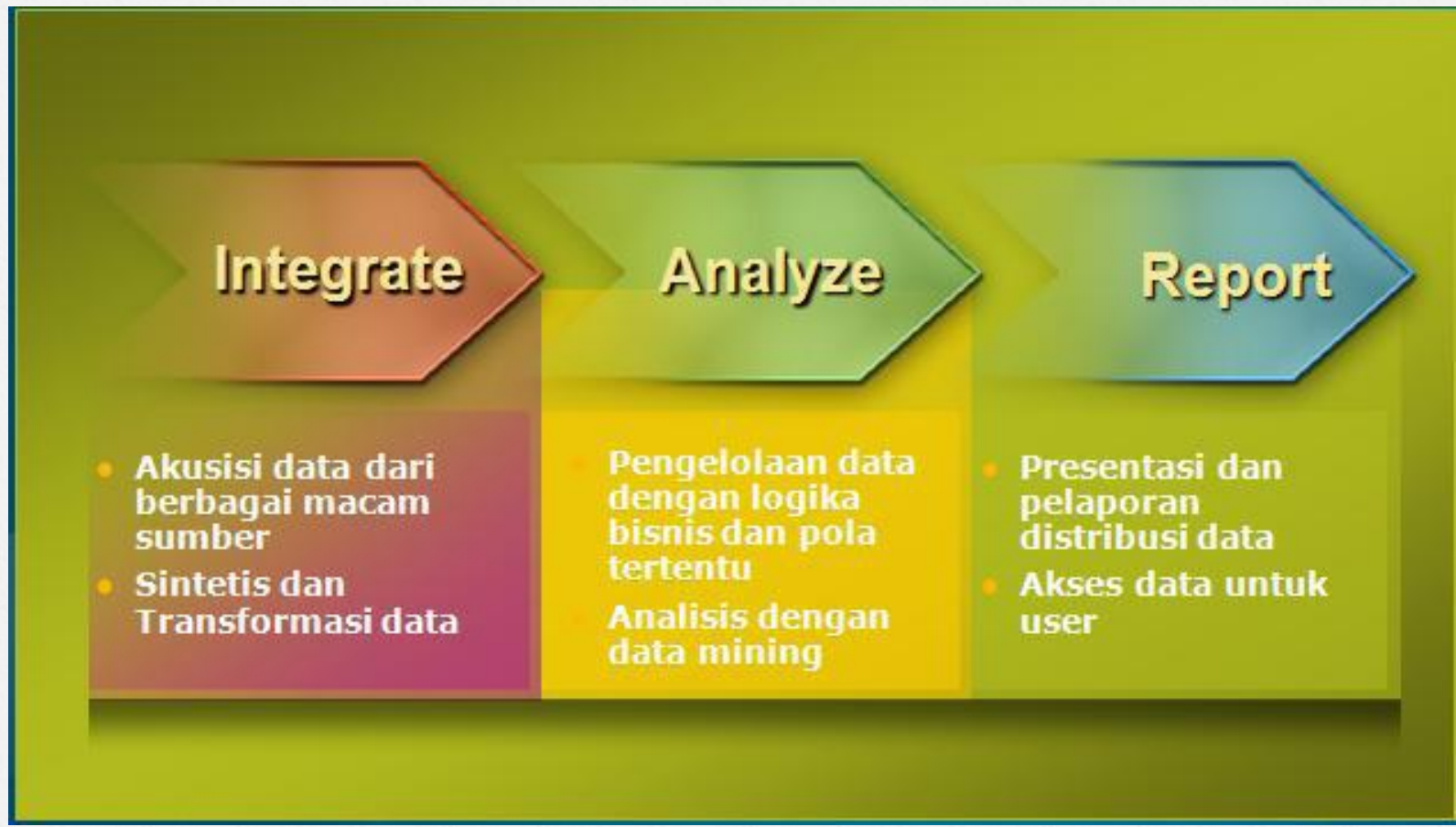
- Non-volatile

***Data tidak dapat diperbaharui atau dirubah tetapi hanya dapat ditambah dan dilihat.***

# Perbedaan Data Warehouse dan Database

- Data Warehouse
  - Tidak terikat suatu aplikasi
  - Data terpusat
  - Historical
  - Denormalisasi kecil
  - Multiple subject
  - Sumber dari dari semua internal maupun eksternal source
  - Fleksibel
  - Data oriented
  - Umurnya panjang
  - Ukuran besar
  - Single complex structure
- Database
  - Aplikasi DSS secara spesifik
  - Tidak terpusat oleh user area
  - Sebagian historical
  - Denormalisasi besar
  - One central subject of concern of user
  - Sumber dari sebagian internal maupun eksternal source
  - Tidak fleksibel, terbatas
  - Project oriented
  - Umurnya pendek
  - Ukuran dari kecil menjadi besar
  - Multi complex structure

# Konsep data warehouse



# Langkah penerapan data warehouse

1. Identifikasi sistem dan unit bisnis yang membutuhkan integrasi data

2. Menetapkan batasan data dan prioritas data yang dibutuhkan untuk putusan organisasional

5. Penerapan Teknologi dan Produk Data warehouse

3. Menampilkan dan memutuskan berbagai alternatif teknologi basis data yang menerapkan konsep data warehouse

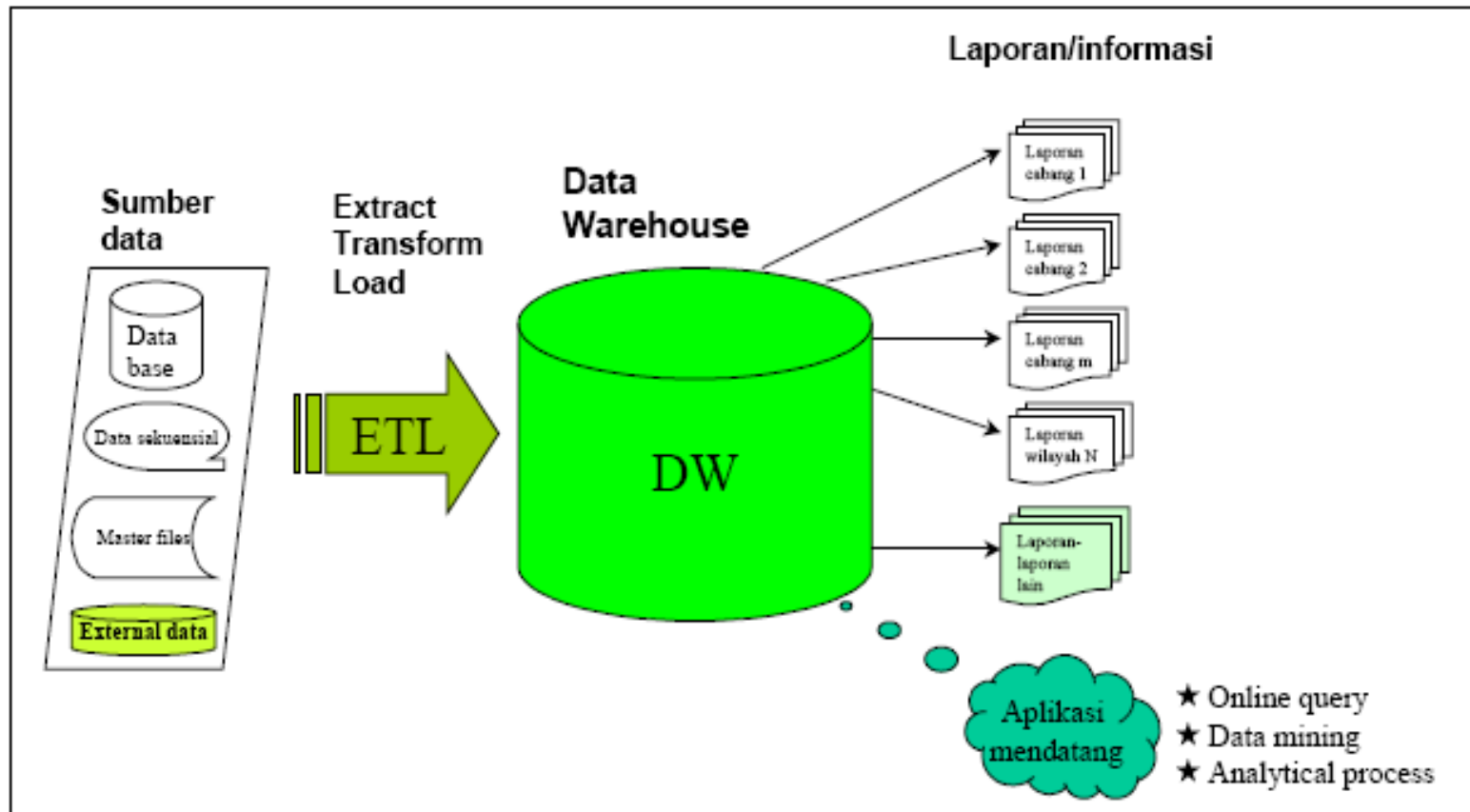
4. Merencanakan pemrosesan dan perancangan standar prosedur operasional implementasi data warehouse

# Proses Data warehouse

Berikut merupakan komponen utama dalam proses data warehouse, sbb :

- Data source.
- Data extraction.
- Data loading.
- Comprehensive database.
- Metadata.
- Middleware tools.

# Proses Data warehouse



# ARSITEKTUR DATA WAREHOUSE

Pilihan berikut harus dibuat didalam perancangan data warehouse

- **process model**

Tipe apa yang akan dimodelkan?

- **grain**

Apa dasar data dan level atom data yang akan disajikan?

- **dimensi**

Dimensi apa yang dipakai untuk masing-masing record tabel fakta?

- **ukuran**

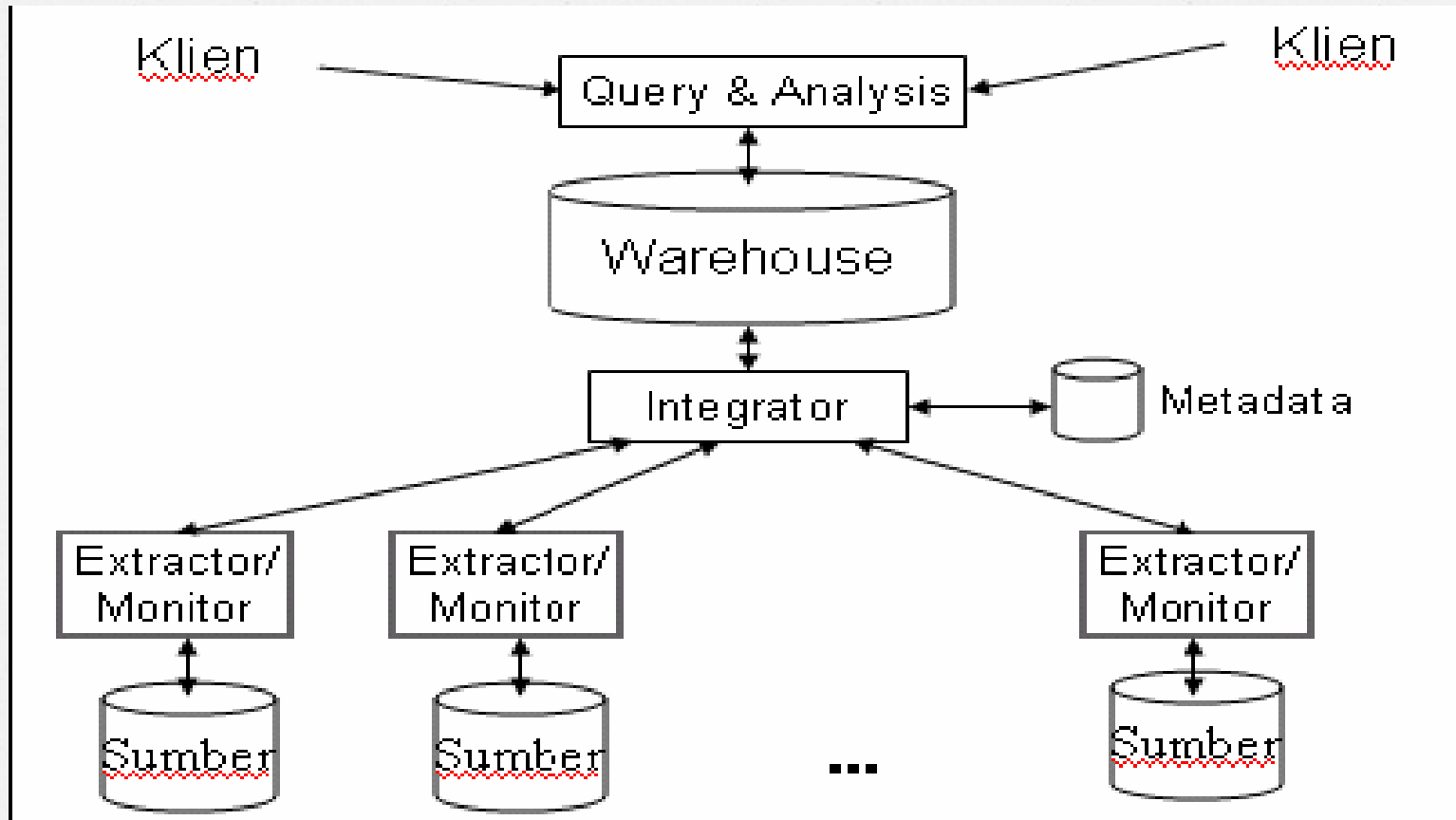
Ukuran apa yang akan mengumpulkan masing-masing record tabel fakta?



Arsitektur data warehouse terbagi dalam tiga bagian yaitu

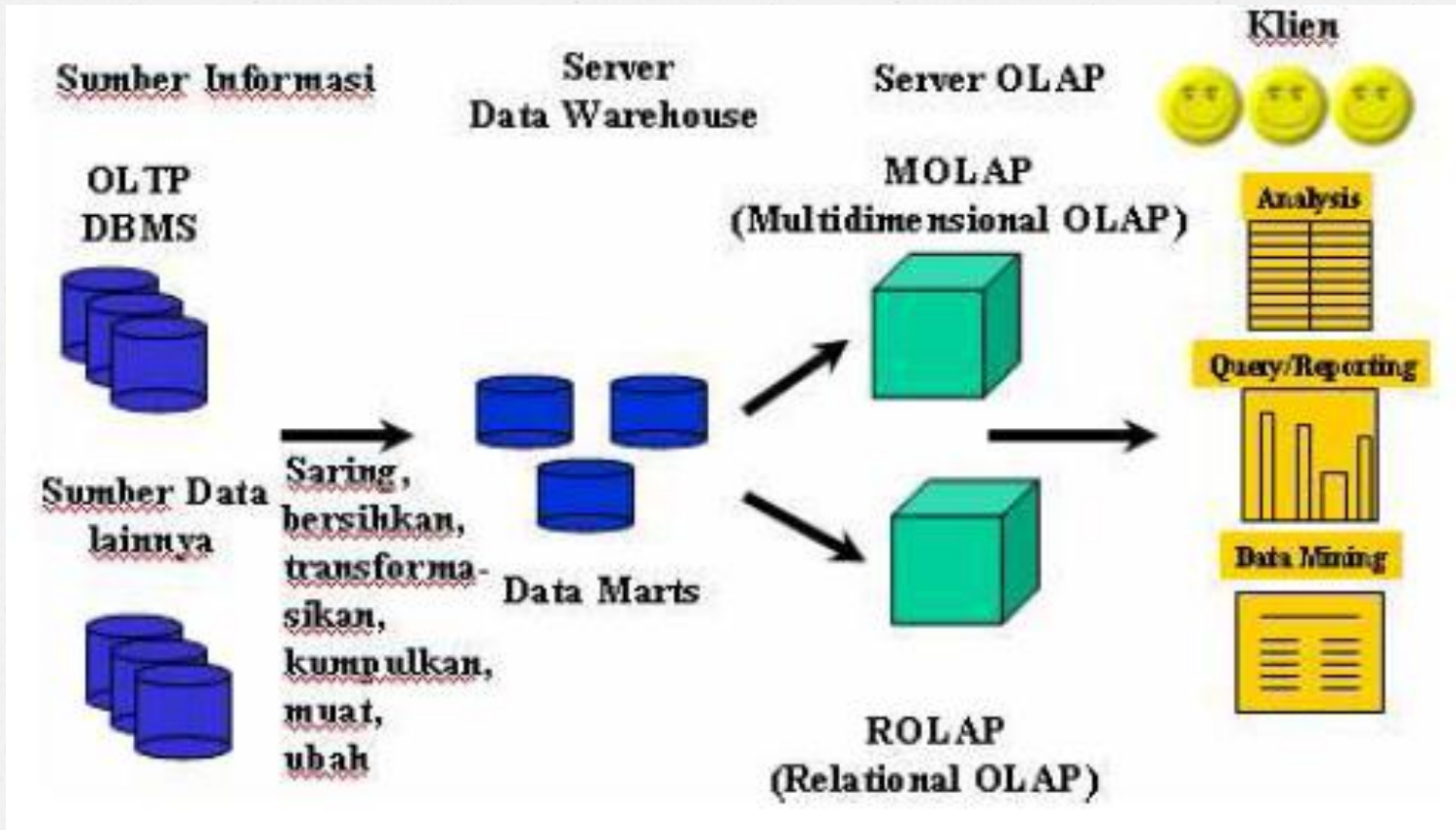
1. Data warehouse itu sendiri, yang terdiri dari data-data dan software yang berasosiasi.
2. Data acquisition (back-end) software, yang mengekstrak data dari system legal dan sumber-sumber eksternal, mengkonsolidasi dan merangkumnya, dan memprosesnya ke dalam data warehouse.
3. Client (front-end) software, yang memungkinkan user mengakses dan menganalisis data dari warehouse.

# ARSITEKTUR DATA WAREHOUSE



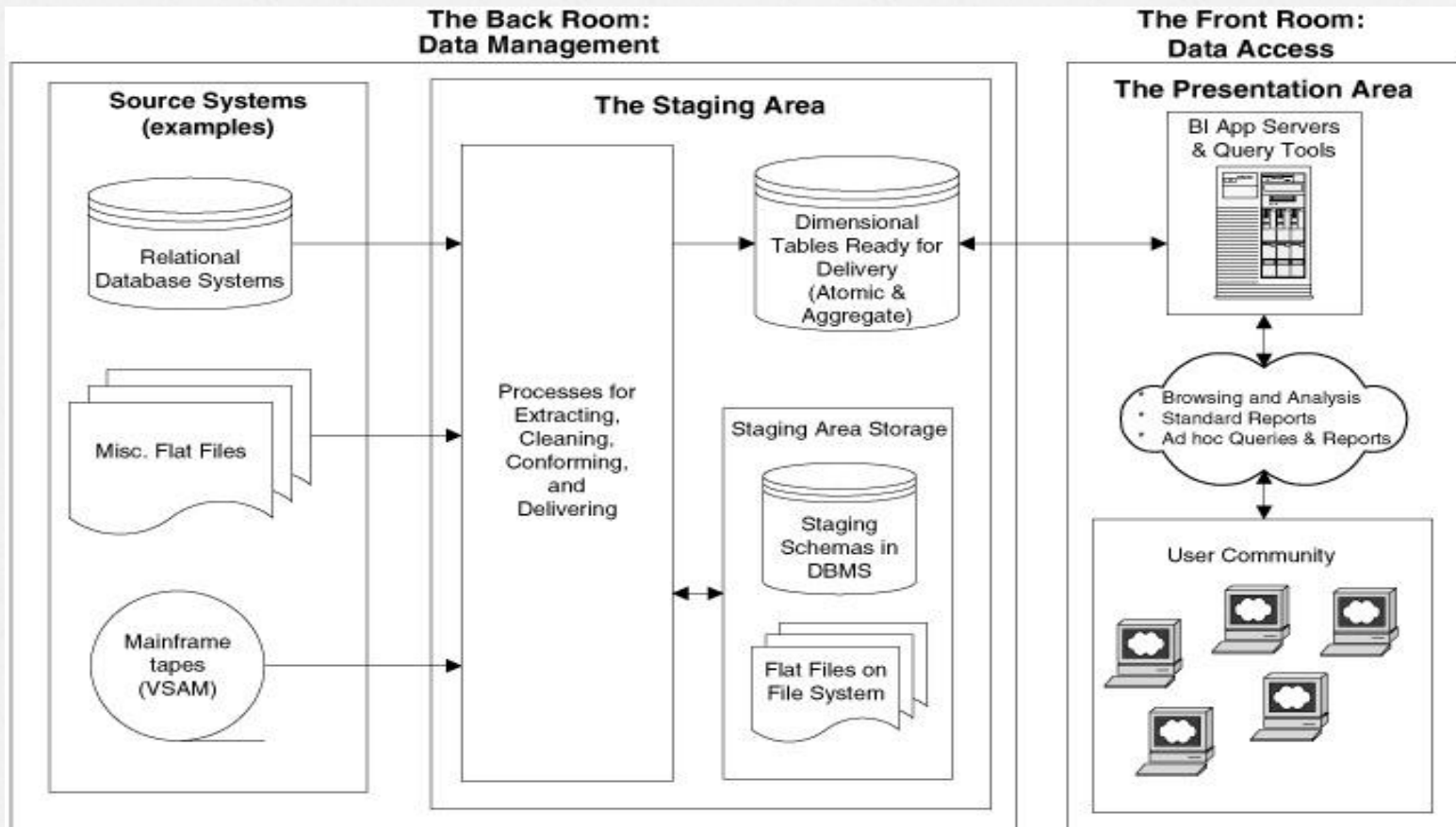
arsitektur dari Data Warehouse

# ARSITEKTUR DATA WAREHOUSE



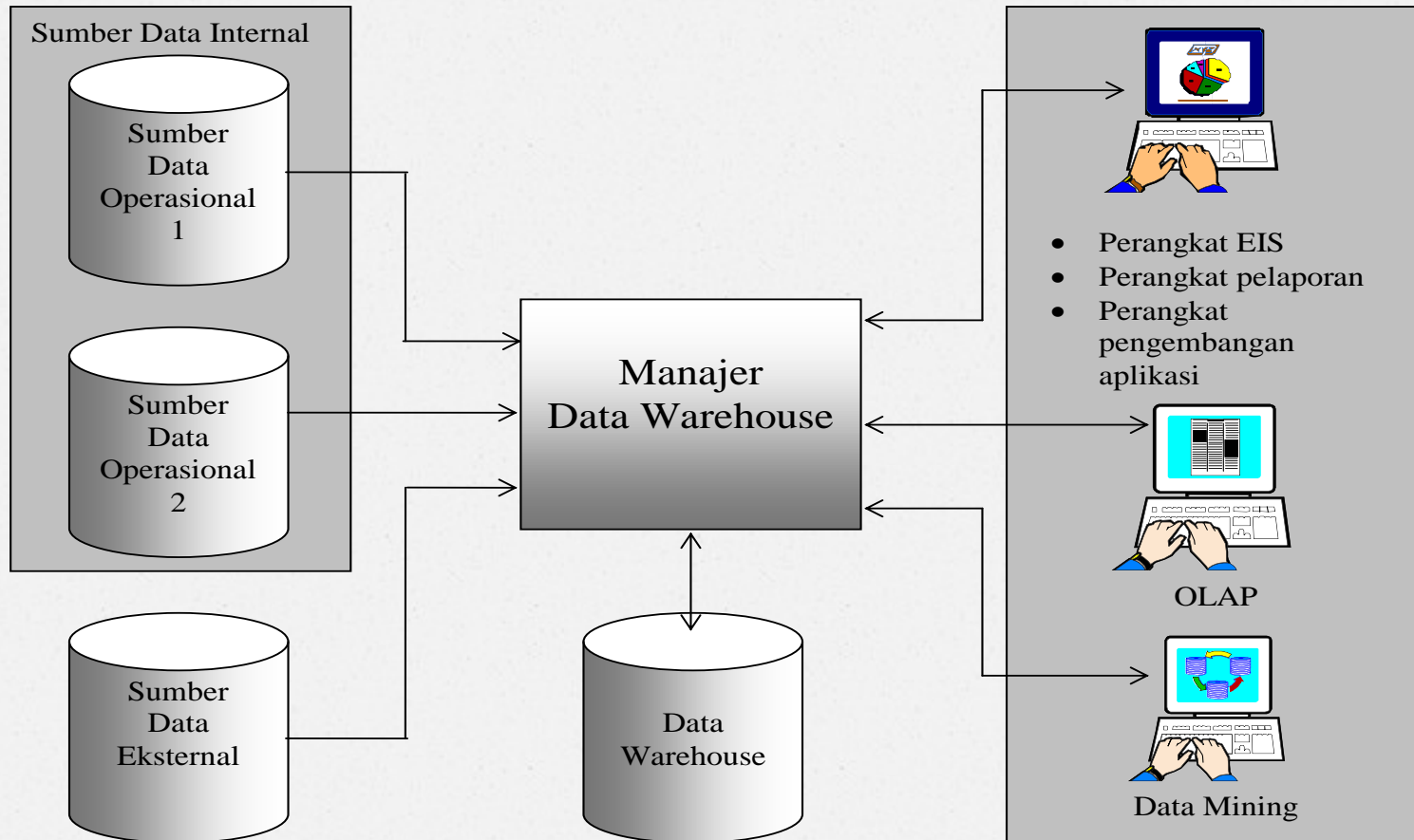
## Arsitektur Data Warehouse

# Data Warehouse Architecture



Asep Jalaludin, S.T., M.M.

# Prinsip Data Warehouse



Asep Jalaludin, S.T., M.M.

# OLAP (On-line analytical processing)

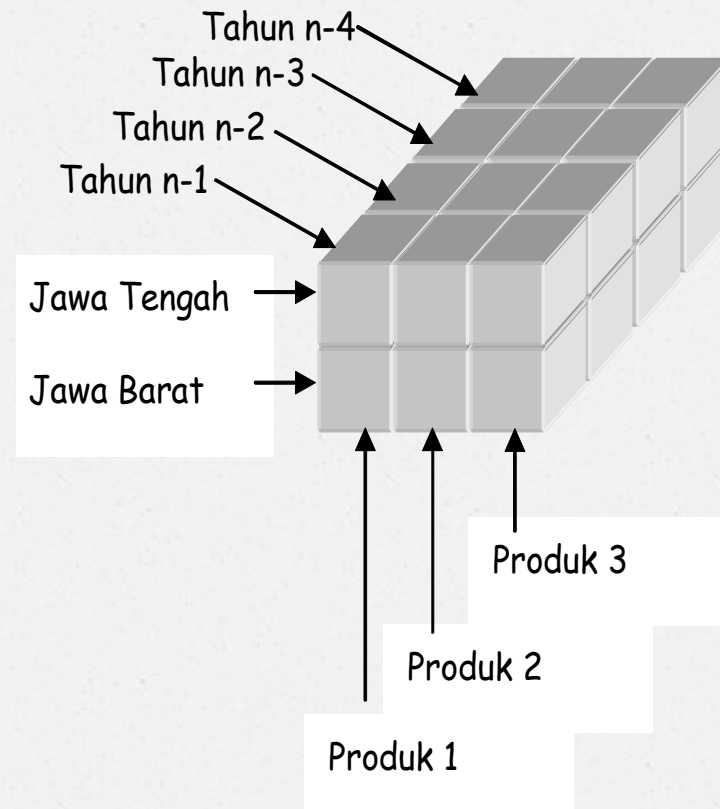
- OLAP adalah suatu sistem atau teknologi yang dirancang untuk mendukung proses analisis kompleks dalam rangka mengungkapkan kecenderungan pasar dan faktor-faktor penting dalam bisnis
- OLAP ditandai dengan kemampuannya menaikkan atau menurunkan dimensi data sehingga kita dapat menggali data sampai pada level yang sangat detail dan memperoleh pandangan yang lebih luas mengenai objek yang sedang kita analisis.
- OLAP secara khusus memfokuskan pada pembuatan data agar dapat diakses pada saat pendefinisian kembali dimensi.
- OLAP dapat digunakan membuat rangkuman dari multidimensi data yang berbeda, rangkuman baru dan mendapatkan respon secara online, dan memberikan view dua dimensi pada data cube multidimensi secara interaktif.

# DW and OLAP as Multidimensional Data Model

- o Data warehouse dan OLAP didasarkan pada multidimensional data model.
- o Model ini merepresentasikan data dalam bentuk data cube, data dimodelkan dan ditampilkan sebagai multiple dimension.
- o Data cube ini didasarkan pada dimensions table dan facts table.

# Multidimensional Data Model

- o Multidimensional yang berarti bahwa terdapat banyak lapisan kolom dan baris (Ini berbeda dengan tabel pada model relasional yang hanya berdimensi dua)
- o Berdasarkan susunan data seperti itu, amatlah mudah untuk memperoleh jawaban atas pertanyaan seperti: “Berapakah jumlah produk 1 terjual di Jawa Tengah pada tahun n-3?”





# Fact Table

- o Merepresentasikan proses bisnis, yaitu model proses bisnis sebagai bentuk (artifact) pada data model.
- o Mengandung elemen ‘pengukuran’ atau metrik atau fakta pada bisnis proses.
  - o “jumlah penjualan bulanan” pada proses bisnis Penjualan.
  - o dll
- o Terdapat foreign key untuk tabel dimensi.
- o Berisi ribuan kolom

# Dimension Tables

- Merepresentasikan *who, what, where, when* and *how* of sebuah pengukuran/artifact.
- Merepresentasikan entities yang real, bukan proses bisnis.
- Memberikan konteks pengukuran (subject)
- Sebagai contoh :

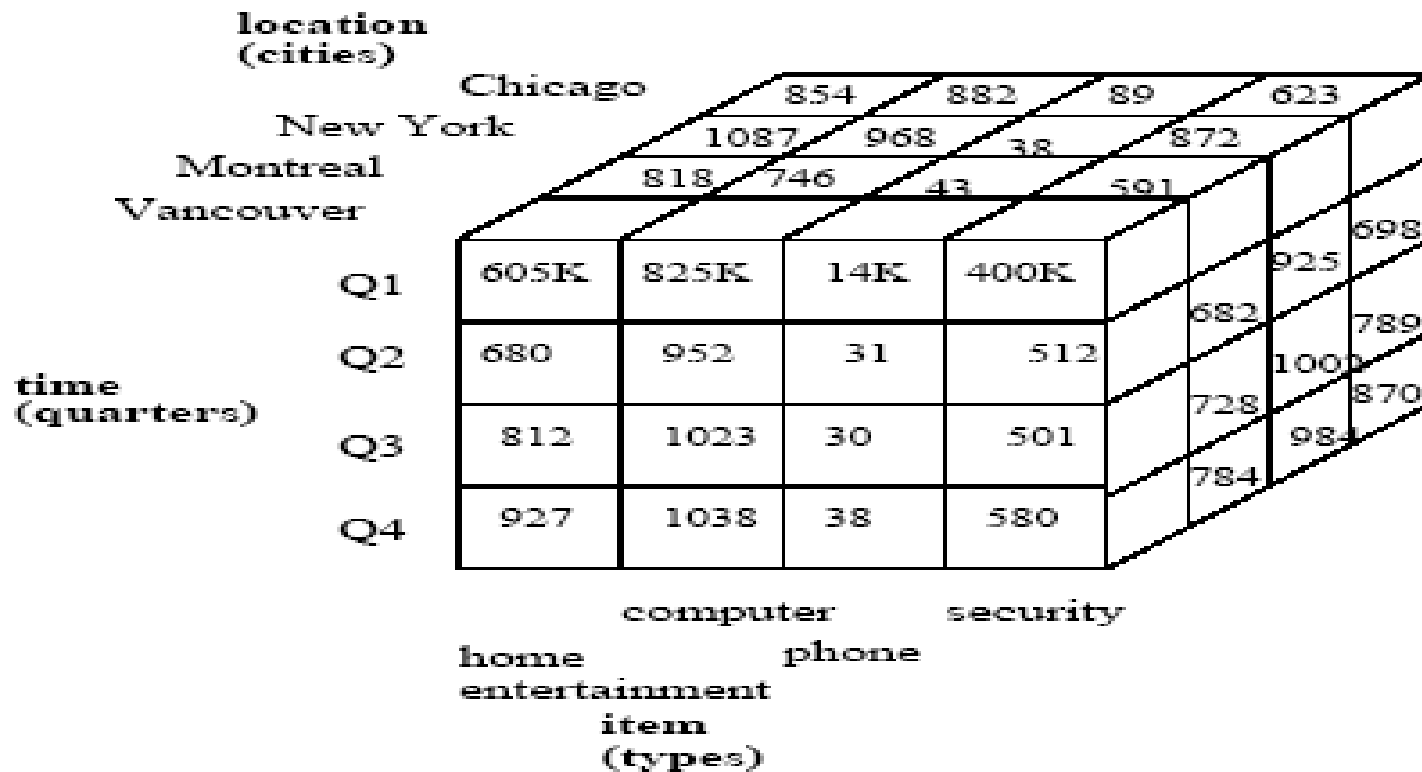
Pada tabel fakta Penjualan. Karakteristik dari pengukuran 'jumlah penjualan bulanan' bisa terdiri dari lokasi (*Where*), waktu (*When*), produk yang terjual (*What*).

# Dimension Tables

- o Atribut pada tabel dimensi merupakan kolom-kolom yang ada pada tabel dimensi.
- o Pada tabel dimensi lokasi, atribut bisa merupakan Kode Pos, kota, kode negara, negara, dll. Secara umum, atribut pada tabel dimensi digunakan untuk pelaporan dengan menggunakan batasan kueri, contoh *where Country='USA'*. Atribut pada tabel dimensi juga mengandung satu atau beberapa hierarchical relationships.
- o Sebelum mendesain data warehouse, harus ditentukan dulu mengandung elemen apa saja data warehouse yang akan dibuat. Misalkan jika ingin membuat data warehouse yang mengandung "jumlah penjualan bulanan" yang melibatkan semua lokasi, waktu, dan produk, maka dimensinya adalah :  
Lokasi  
Waktu Product

o

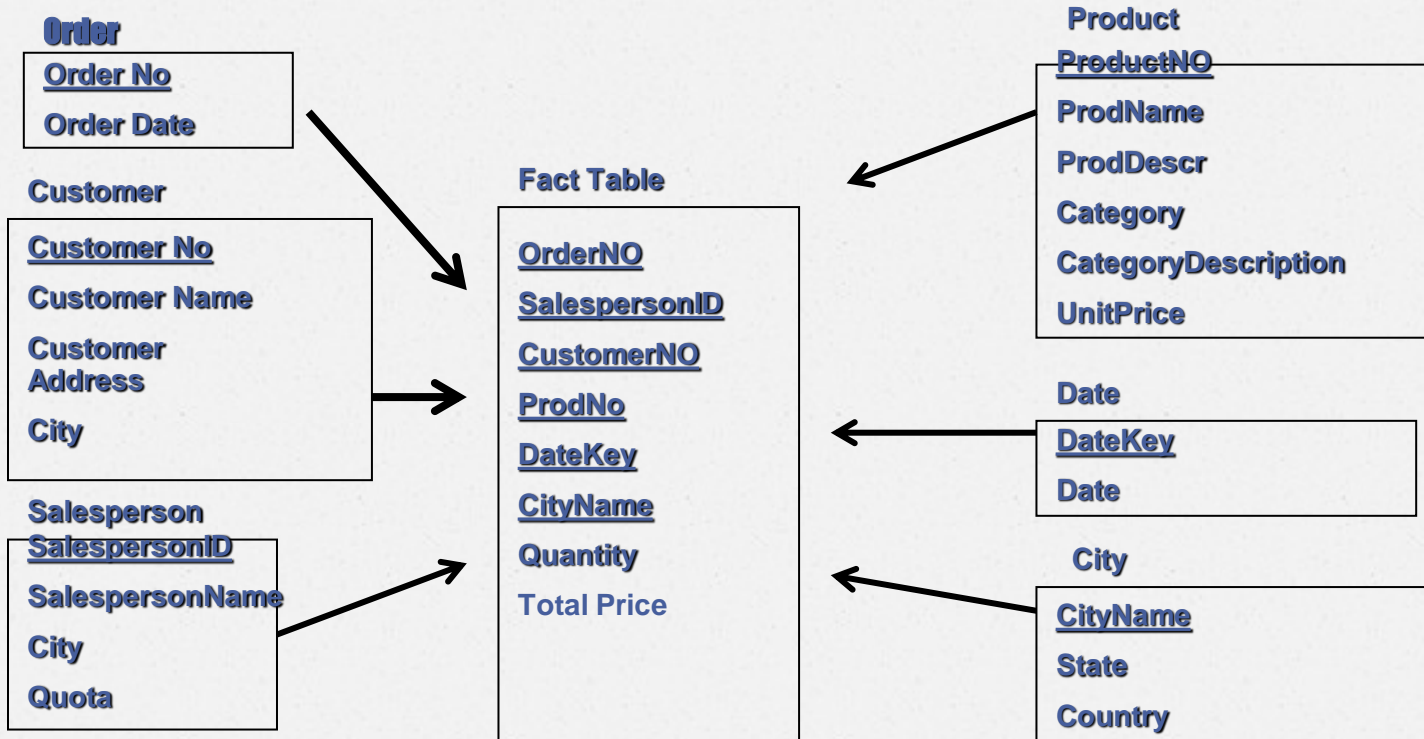
# 3D data cube, according to the dimension time, location, item



# Warehouse Database Schema

- o Bukan ER Diagram
- o Design harus mencerminkan multidimensional view
  - o Star Schema
  - o Snowflake Schema
  - o Fact Constellation Schema

# Example of a Star Schema



### Time Dimension

year  
quarter  
month  
day\_of\_week  
day  
time\_key

### Sales Fact

time\_key  
item\_key  
branch\_key  
location\_key  
dollars\_sold  
units\_sold

### Item Dimension

supplier\_type  
type  
brand  
item\_name  
item\_key

### Branch Dimension

branch\_type  
branch\_name  
branch\_key

### Location Dimension

country  
province\_or\_state  
city  
street  
location\_key

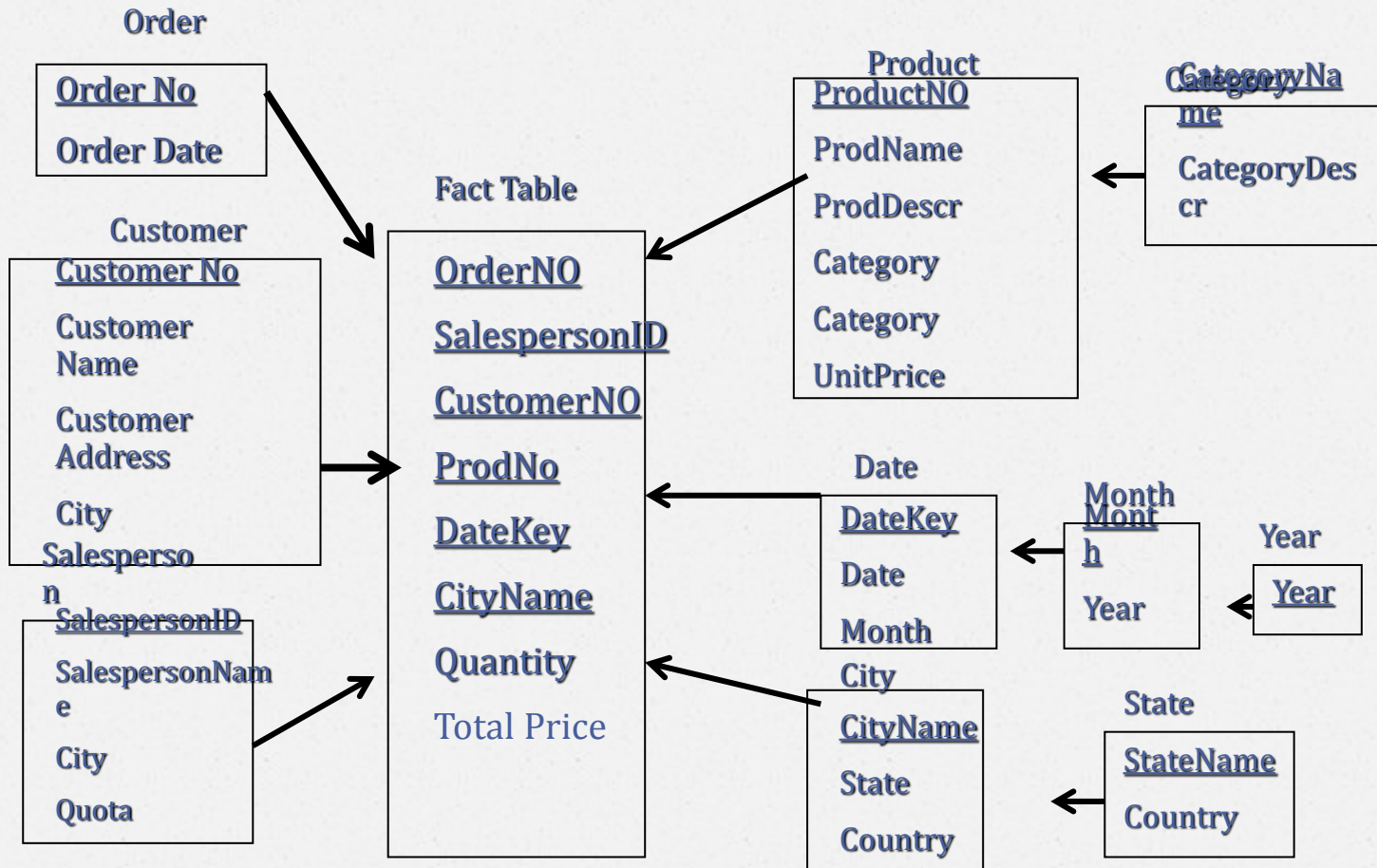
Figure 2.4: Star schema of a data warehouse for sales.

# Star Schema

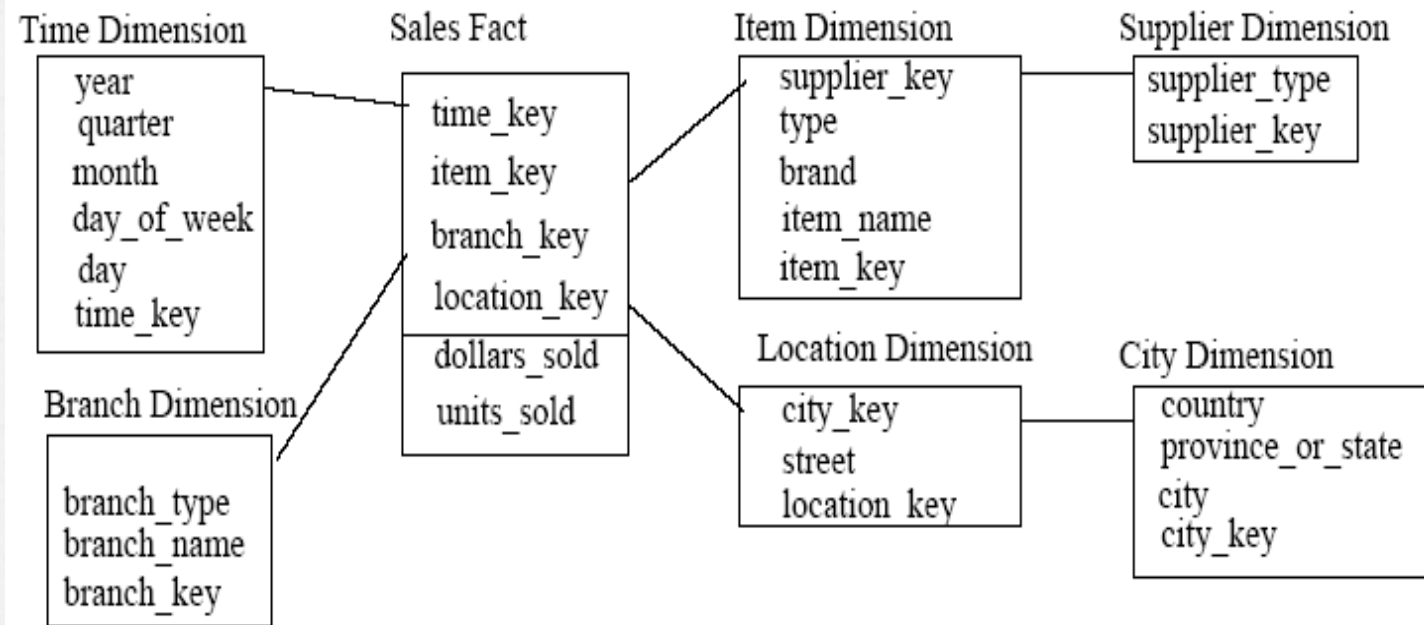
- o Model dimana data warehouse terdiri dari satu tabel pusat yang besar (tabel fakta).
- o Ada satu table untuk tiap dimensi
- o Pada star schema, satu dimensi diwakili oleh satu tabel dan masing-masing tabel diwakili oleh beberapa atribut.



# Example of a Snowflake Schema



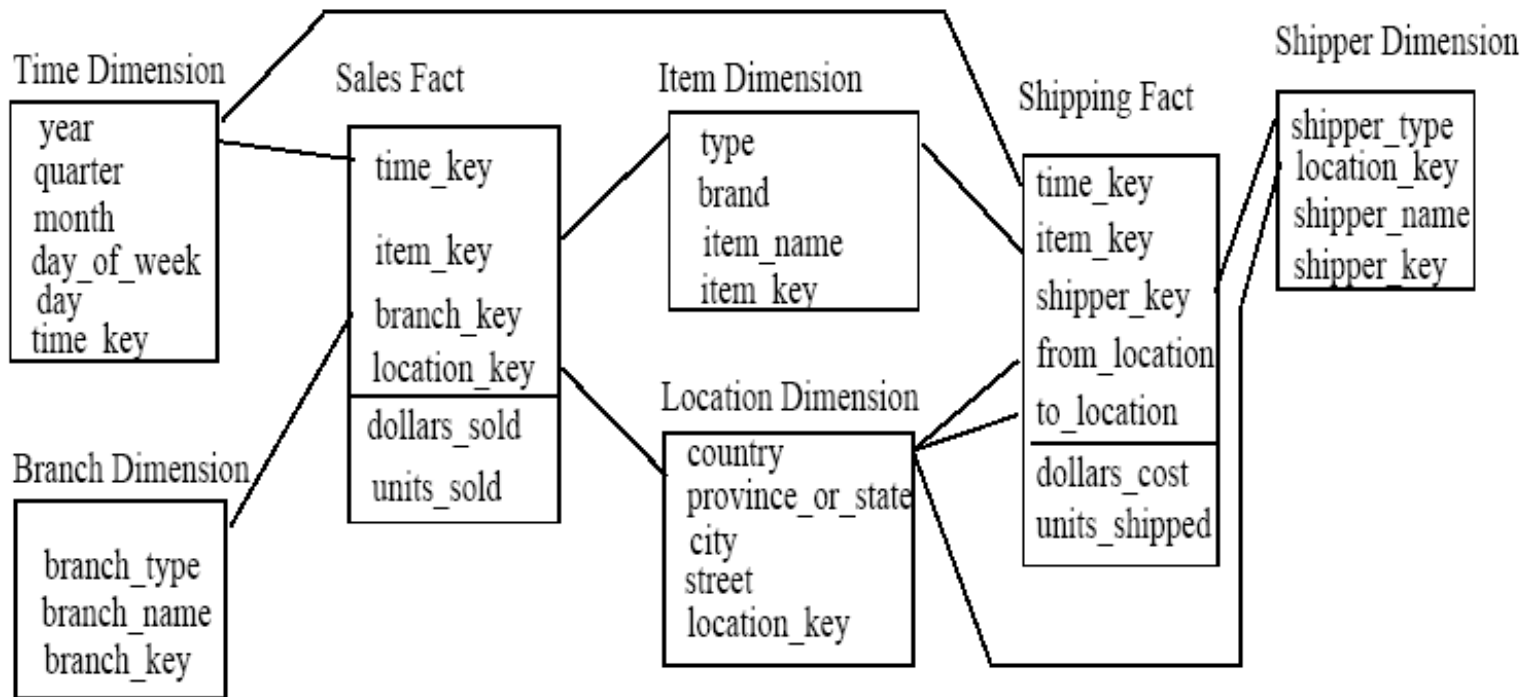
# Snowflake Schema



# Snowflake Schema

- o Ada mekanisme normalisasi tabel dimensi
- o Mudah untuk dimaintain
- o Butuh kapasitas penyimpanan yang relatif lebih kecil
- o Efektifitas browsing/select data berkurang karena harus melibatkan banyak kueri dari berbagai macam tabel

# Fact Constellation



# Fact Constellation

- o Fact Constellation
  - o Ada beberapa tabel fakta yang digunakan bersama-sama (share) beberapa tabel dimensi.
  - o Dapat berupa kumpulan skema star

# Data Warehouse

- *Data warehouse* dapat dibangun sendiri dengan menggunakan perangkat pengembangan aplikasi ataupun dengan menggunakan perangkat lunak khusus yang ditujukan untuk menangani hal ini
- Beberapa contoh perangkat lunak yang digunakan untuk administrasi dan manajemen *data warehouse*:
  - HP Intelligent Warehouse (Hewlett Packard)
  - FlowMark (IBM)
  - SourcePoint (Software AG)

# Petunjuk Membangun DW

- o Menentukan misi dan sasaran bisnis bagi pembentukan *data warehouse*
- o Mengidentifikasi data dari basis data operasional dan sumber lain yang diperlukan bagi *data warehouse*
- o Menentukan item-item data dalam perusahaan dengan melakukan standarisasi penamaan data dan maknanya
- o Merancang basis data untuk *data warehouse*
- o Membangun kebijakan dalam mengarsipkan data lama sehingga ruang penyimpanan tak menjadi terlalu besar dan agar pengambilan keputusan tidak menjadi terlalu lamban.
- o Menarik data produksi (operasional) dan meletakkan ke basis data milik *data warehouse*

# Contoh Software Data Mart

- o SmartMart (IBM)
- o Visual Warehouse (IBM)
- o PowerMart (Informatica)



# OLAP

- *OnLine Analytical Processing*
- Suatu jenis pemrosesan yang memanipulasi dan menganalisa data bervolume besar dari berbagai perspektif (multidimensi). OLAP seringkali disebut analisis data multidimensi.

# OLAP

- Data multidimensi adalah data yang dapat dimodelkan sebagai atribut dimensi dan atribut ukuran
- Contoh atribut dimensi adalah nama barang dan warna barang, sedangkan contoh atribut ukuran adalah jumlah barang

# OLAP : Contoh Data 2 Dimensi

Kota →	Kudus	Magelang	Semarang	...
Triwulan ↓				
1	6.000.000	8.500.000	12.500.000	...
2	4.500.000	3.500.000	14.000.000	...
3	7.600.000	5.500.000	13.700.000	...
4	5.400.000	7.200.000	12.800.00	...

# Kemampuan OLAP

- **Konsolidasi (roll up)** melibatkan pengelompokan data. Sebagai contoh kantor-kantor cabang dapat dikelompokkan menurut kota atau bahkan propinsi. Transaksi penjualan dapat ditinjau menurut tahun, triwulan, bulan, dan sebagainya. Kadangkala istilah *rollup* digunakan untuk menyatakan konsolidasi
- **Drill-down** adalah suatu bentuk yang merupakan kebalikan dari konsolidasi, yang memungkinkan data yang ringkas dijabarkan menjadi data yang lebih detail
- **Slicing and dicing** (atau dikenal dengan istilah *pivoting*) menjabarkan pada kemampuan untuk melihat data dari berbagai sudut pandang

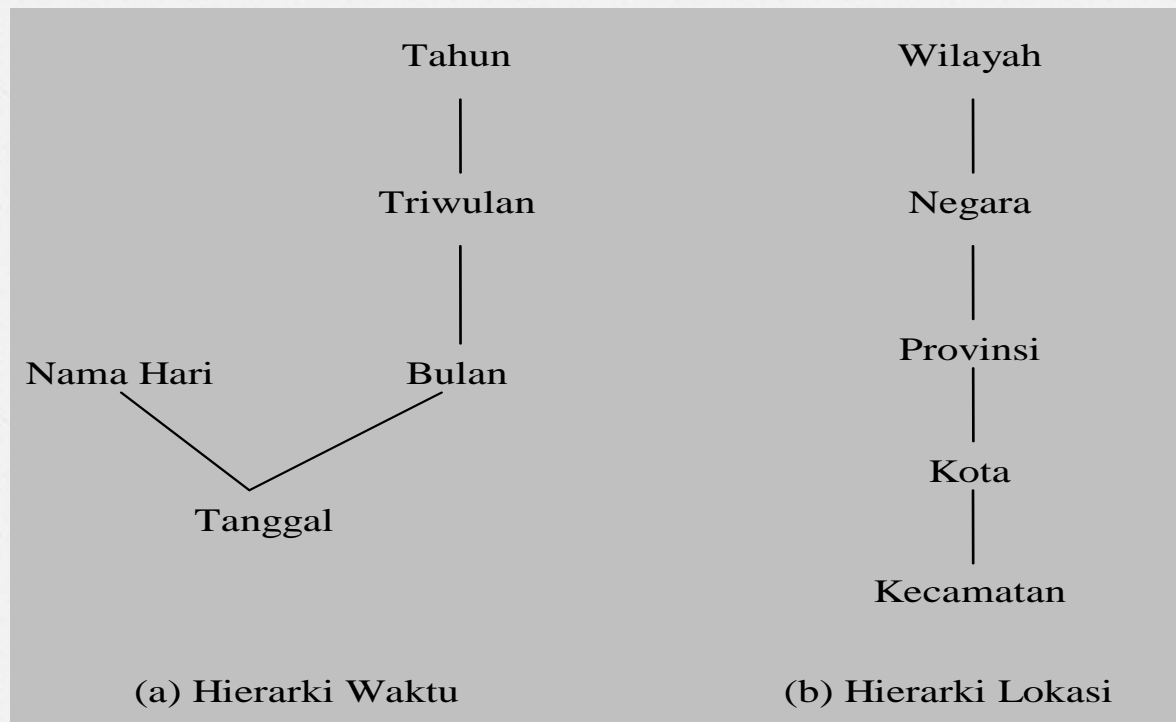
# Contoh Tabel Pivoting

Rasa	Strawberry	Mangga	Nanas	Total
Sirup				
Biasa	3.500.000	1.750.000	500.000	5.750.000
Rendah Kalori	2.300.000	1.500.000	250.000	4.050.000
Total	5.800.000	3.250.000	750.000	9.800.000



Sirup	Rasa	Pendapatan
Biasa	Strawberry	3.500.000
Biasa	Mangga	1.750.000
Biasa	Nanas	500.000
Rendah Kalori	Strawberry	2.300.000
Rendah Kalori	Mangga	1.500.000
Rendah Kalori	Nanas	250.000

# Hierarki Dimensi untuk Drill-down



# PENUTUP

Data warehouse merupakan suatu cara/metode dari suatu database yang berorientasi kepada subjek, non-volatile, time-variance dan terintegrasi yang digunakan untuk mempermudah para pengambil keputusan dalam memecahkan masalah.

Keberadaan data warehouse sangat penting sebagai tools dari DSS, karena data warehouse memang digunakan untuk itu. Dengan adanya data warehouse, diharapkan suatu perusahaan dapat lebih unggul dari kompetitornya dan lebih jeli lagi dalam melihat peluang pasar.

# Daftar Pustaka

- Djoni Darmawikarta, Mengenal Data Warehouse, 2003
- Yudho Giri Sucahyo, Data Mining, 2003
- Yudho Giri Sucahyo, Penerapan Data Mining, 2003
- Jeffrey A. Hoffer, Mary B. Prescott, Fred R. McFadden ; Modern Database Management 8th Edition; 2007