



# **Data Warehouse dan Decision Support System**

**Arif Basofi**

# Referensi

- Data Warehouse, STMIK Global Informatika MDP.
- M. Syukri Mustafa, S.Si., MMSI, Sistem Basis Data II (Data Warehouse), 2008.
- Hanim MA, Data Warehouse 2<sup>nd</sup>

# Memahami Kebutuhan Informasi Strategis

## History dibangunnya DataWarehouse :

- Pada th. 1990 bisnis semakin **komplek**, perusahaan **tersebar secara global** dan **kompetisi** yg berat.
- Usaha **bisnis executif** menjadi **tersebar** dalam **persaingan informasi**.
- Sistem operasional perusahaan menyediakan informasi untuk menjalankan operasi sehari-hari, tetapi executive memerlukan berbagai macam jenis informasi yang dapat mendukung dalam **pengambilan keputusan** yg **strategis**.
- Sistem operasional sangat penting untuk perusahaan, namun **tidak dapat** menyediakan informasi yg strategis. Oleh karena itu, perusahaan **perlu mengarah kepada cara yang baru** dlm memperoleh informasi yg strategis.

# Data Warehouse

## Data Warehouse :

- **Data Warehouse** adalah Pusat repositori informasi yg mampu memberikan database **berorientasi subyek** untuk informasi yg bersifat **historis** yg mendukung **DSS (Decision Suport System)** dan **EIS (Executive Information System)**.
- **Data warehouse** adalah basis data yg menyimpan data sekarang dan masa lalu, yg berasal dari berbagai sistem operasional dan sumber lain (sumber eksternal), yang menjadi perhatian penting bagi manajemen dalam organisasi dan ditujukan **untuk keperluan analisis dan pelaporan manajemen** dalam rangka **pengambilan keputusan**.
- Salinan dari transaksi data yg terstruktur secara spesifik pd **query** dan **analisa** serta **laporan**.
- ∴ **Data warehouse** hanya berisi informasi<sup>2</sup> yg relevan bagi kebutuhan pemakai yg dipakai untuk **pengambilan keputusan**.

# Data Warehouse

## Tujuan :

- Meningkatkan **kualitas** dan **akurasi informasi bisnis** dan mengirimkan informasi ke pemakai dalam bentuk yg dimengerti dan dapat diakses dengan mudah.
- Digunakan untuk **mendukung pengambilan keputusan**, bukan untuk melaksanakan pemrosesan transaksi.
  - Ini mungkin bertolak belakang dgn sebagian besar sistem OLTP yg berorientasi proses.  
**OLTP** = Online Transaction Processing

# Dari Data Warehouse ke Decision Support

- Pembuatan keputusan organisasi memerlukan **view** menyeluruh pada segala aspek perusahaan, shg organisasi membuat data warehouse gabungan yg berisi data yg berasal dari berbagai sumber.
- DBMS juga didesain utk mendukung query OLAP scr efisien dan dioptimalisasi utk mendukung aplikasi decision support.

# Karakteristik Data Warehouse

- Empat karakteristik Data Warehouse:
  - 1) Subject Oriented
  - 2) Integrated
  - 3) Time Variant
  - 4) Non-Volatile

# Karakteristik Data Warehouse...

## 1) Subject Oriented

- Data yg disusun **menurut subyek** berisi hanya informasi yg penting bagi pemrosesan Decision Support (DSS).
- Informasi yg tersimpan dlm database di kelompokkan **berdasarkan subyek tertentu**, misalnya: pelanggan, gudang, pasar, dsb.
- Semua Informasi tsb disimpan dalam suatu sistem *data warehouse*.
- Data-data di setiap **subyek** dirangkum ke dalam **dimensi**, misalnya : periode waktu, produk, wilayah, dsb, shg dapat memberikan nilai sejarah untuk bahan analisa.

# Karakteristik Data Warehouse...

## 1) Subject Oriented...

### Operational Applications

Order  
Processing

Consumer  
Loans

Customer  
Billing

Accounts  
Receivable

Claims  
Processing

Savings  
Accounts

### Data Warehouse Subjects

Sales

Product

Customer

Account

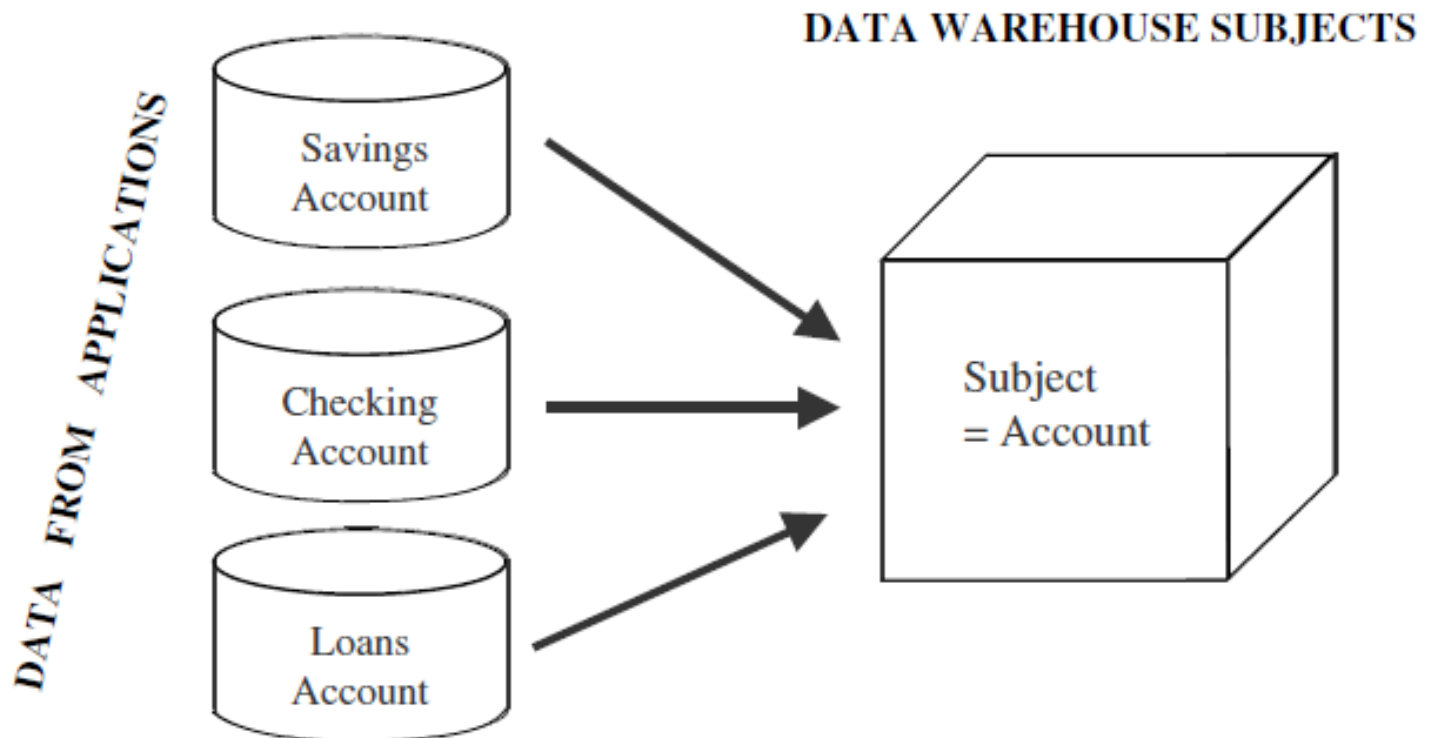
Claims

Policy

# Karakteristik Data Warehouse...

## 2) Integrated

- Data Warehouse dapat menyimpan data<sup>2</sup> yg berasal dari sumber<sup>2</sup> yg terpisah kedalam suatu **format** yg **konsisten** dan saling **terintegrasi** satu dgn lainnya.



# Karakteristik Data Warehouse...

## 3) Time Variant

- Data yg tersimpan dlm rentang 5-10 tahun/lebih, suatu ketika dpt digunakan untuk perbandingan atau perkiraan sbg data yg valid dan akurat, dan data ini tidak dapat diperbaharui.
- Data di dalam data warehouse mempunyai karakter khusus berupa *time-series* dalam bentuk **data historical**.
- Ini dapat digunakan untuk melakukan **trend analysis** dari data tersebut.

# Karakteristik Data Warehouse...

## 3) Time Variant...

- Cara<sup>2</sup> yg digunakan untuk melihat interval waktu dlm mengukur keakuratan suatu data warehouse:
  - a) Paling sederhana : menyajikan data warehouse pada rentang waktu tertentu.
  - b) Menggunakan variasi/perbedaan waktu yg disajikan dalam data warehouse baik implicit maupun explicit; secara explicit dengan unsur waktu dalam hari, minggu, bulan dan waktu tertentu.
  - c) Variasi waktu yg disajikan data warehouse melalui serangkaian *snapshot* yg panjang.

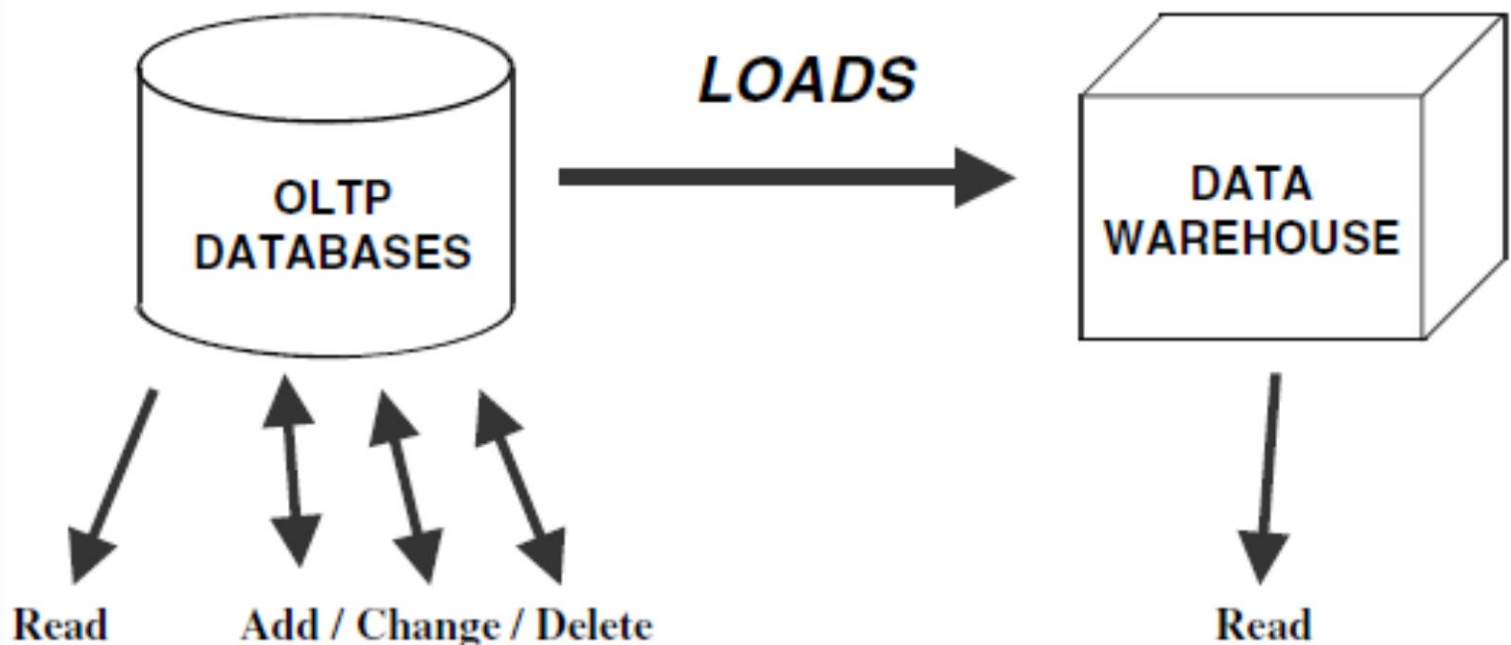
# Karakteristik Data Warehouse...

## 4) Non-Volatile (tidak gampang berubah)

- Data dalam data warehouse tidak diupdate scr real-time/terus menerus (spt inserts, updates, deletes) spt dalam sebuah sistem OLTP, tetapi hanya dpt ditambah dan dilihat.
- Pada data warehouse hanya ada 2 kegiatan memanipulasi data yaitu :
  - loading data (mengambil data), dan
  - akses data (mengakses data warehouse).
- Data di dalam data warehouse scr periodik di *upload* dlm jangka waktu yang sama (misal. setiap sore / setiap tanggal 2)

# Karakteristik Data Warehouse...

## 4) Non-Volatile



# Data Warehouse Vs Database

## Data Warehouse :

- Tidak terikat suatu aplikasi
- Data terpusat
- Historical
- Denormalisasi kecil
- Multiple subject
- Sumber dari semua internal maupun eksternal source
- Fleksibel
- Data oriented
- Umurnya panjang
- Ukuran besar
- Single complex structure

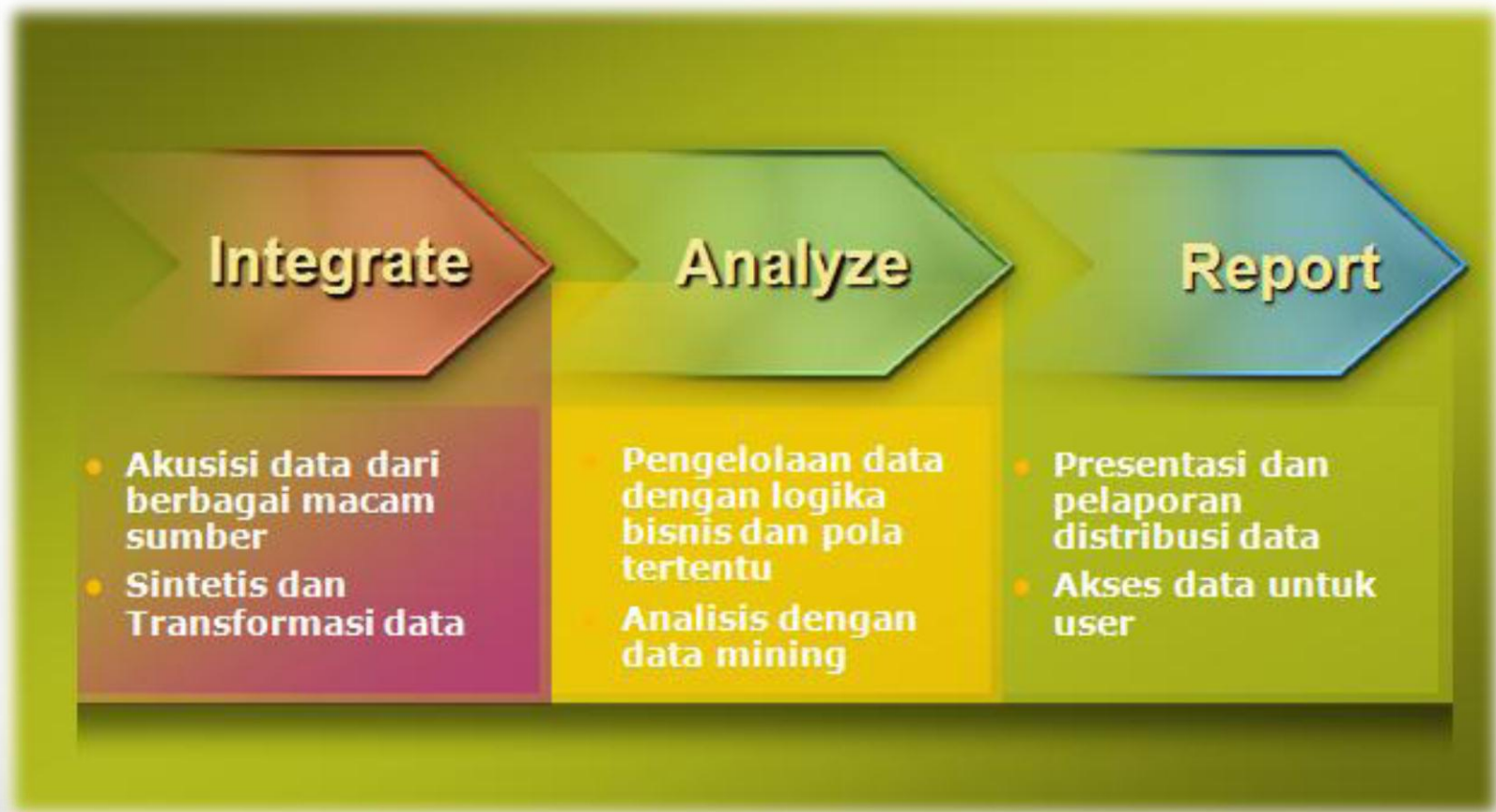
## Database :

- 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

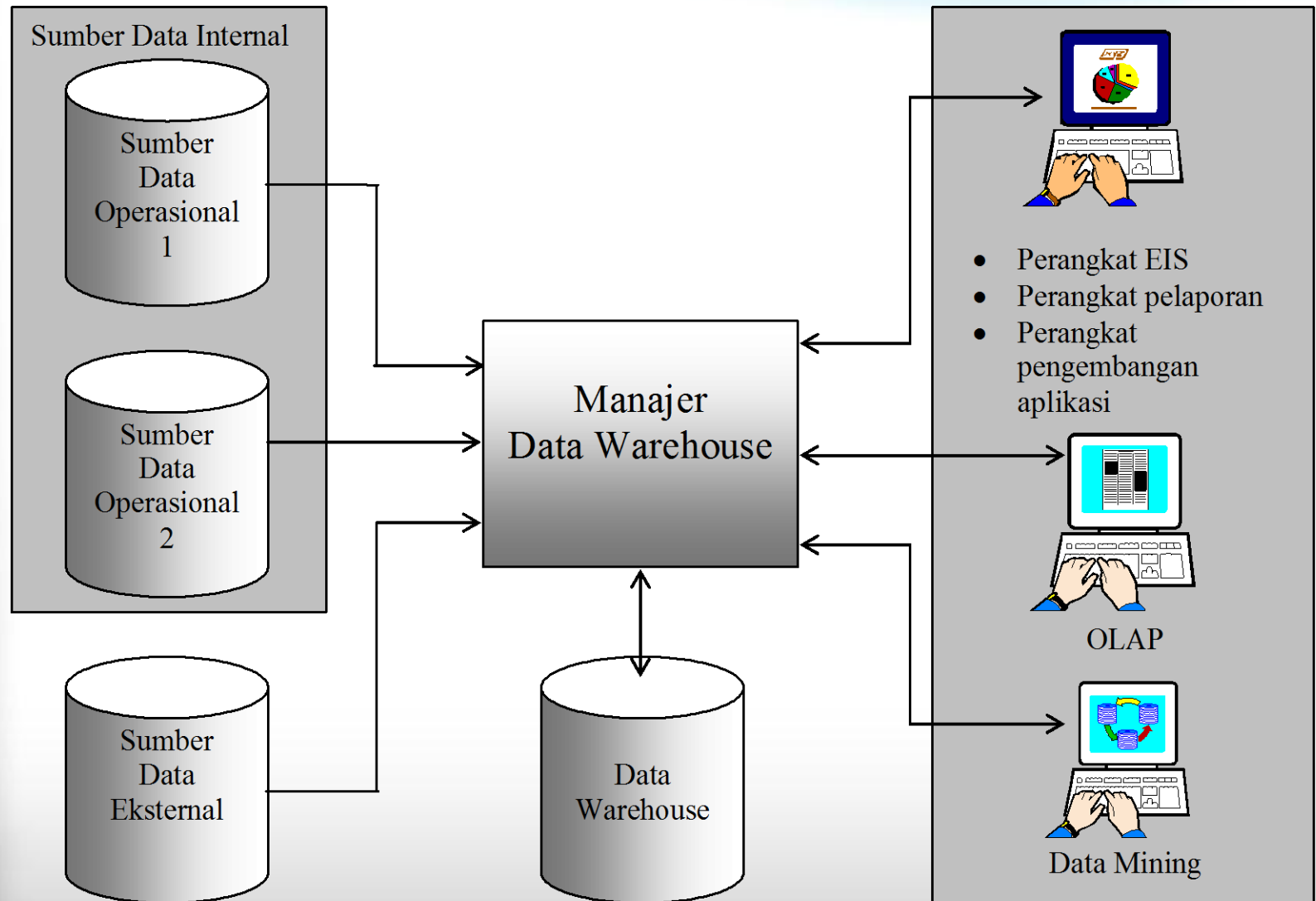
# Sumber Data untuk Data Warehouse

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.
- Berbagai data yang berasal dari sumber digabungkan dan diproses lebih lanjut oleh manajer *data warehouse* dan disimpan dalam basis data tersendiri.
  - Selanjutnya, perangkat lunak seperti OLAP dan *data mining* dapat digunakan oleh pemakai untuk mengakses *data warehouse*.

# Konsep dan Prinsip Data Warehouse



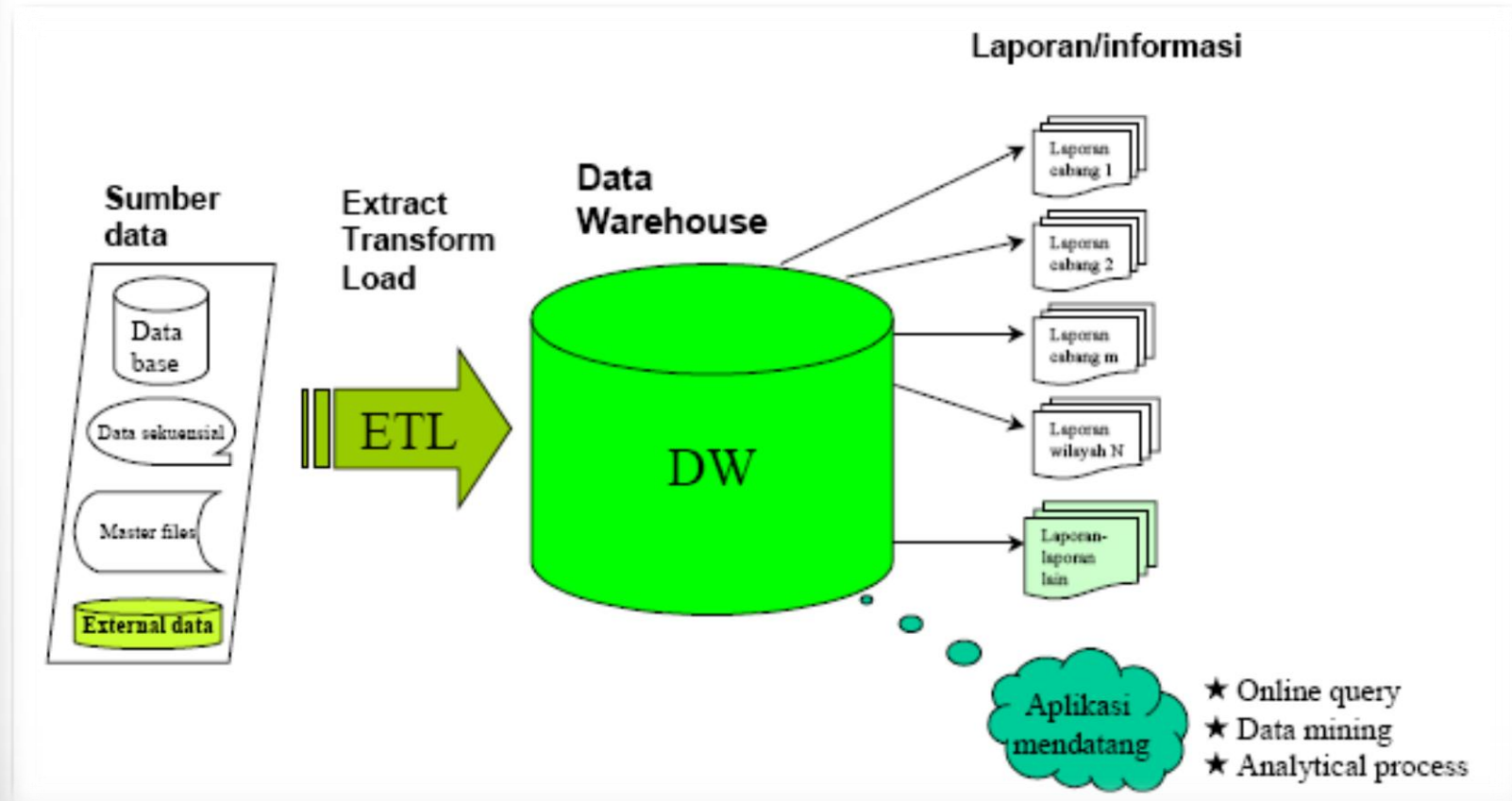
# Konsep dan Prinsip Data Warehouse...



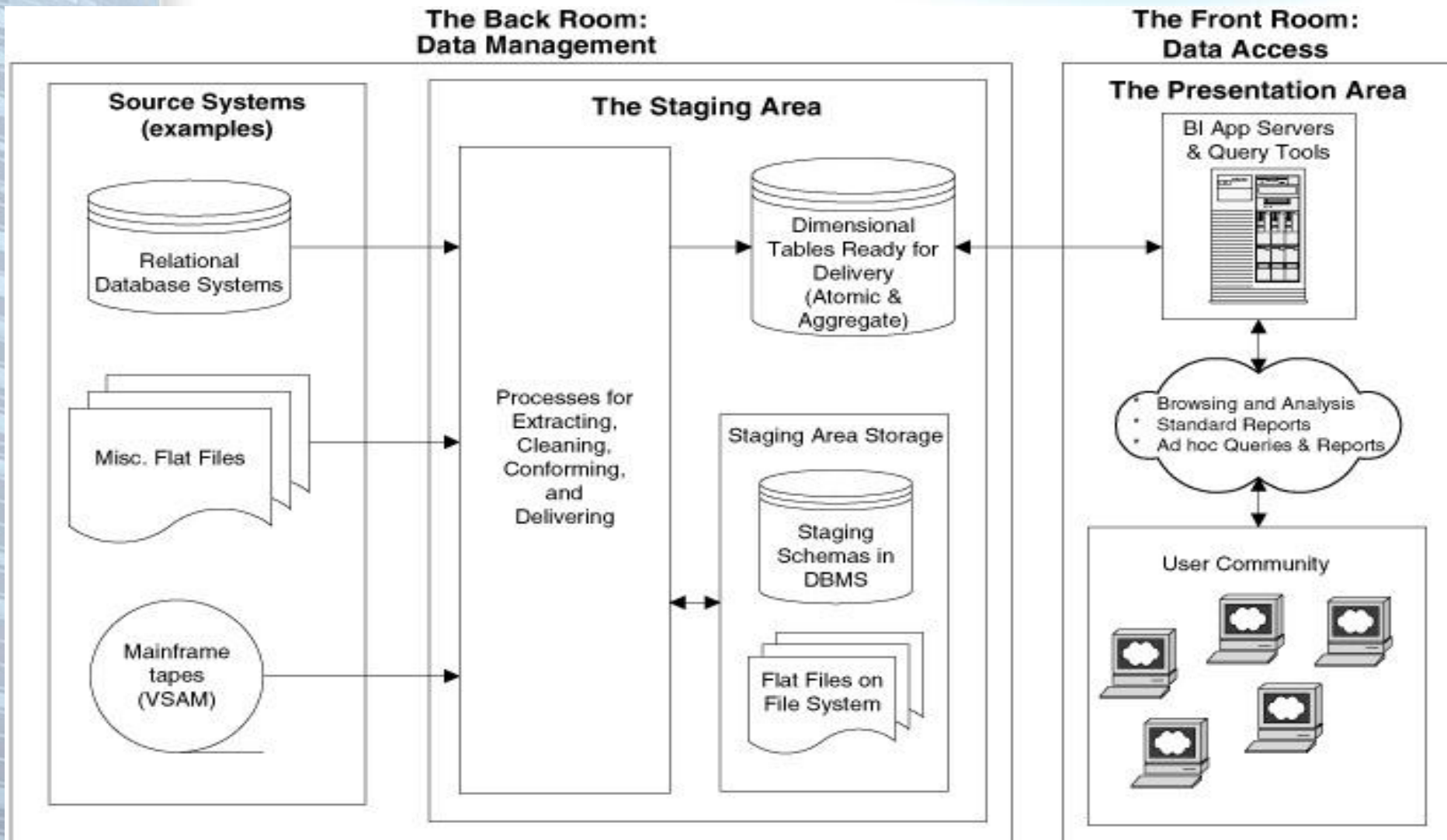
# Langkah Penerapan Data Warehouse



# Arsitektur Data Warehouse



# Arsitektur Data Warehouse (Detail)



# Arsitektur Data Warehouse

Hal yg harus dibuat di dlm perancangan data warehouse:

- **process model**

Tipe apa yang akan dimodelkan?

- **grain**

Apa dasar data dan level atom data yang akan disajikan?

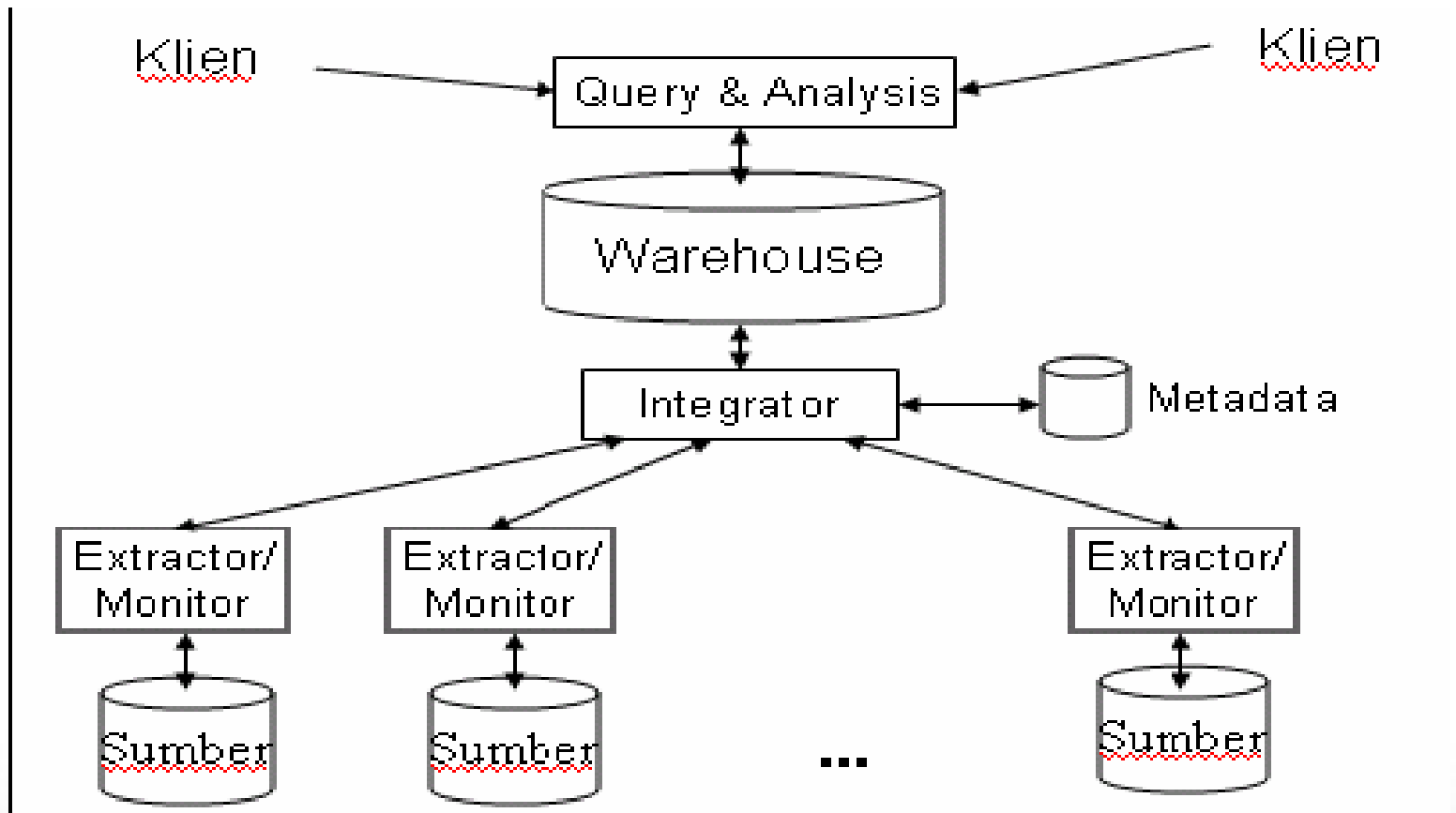
- **dimensi**

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

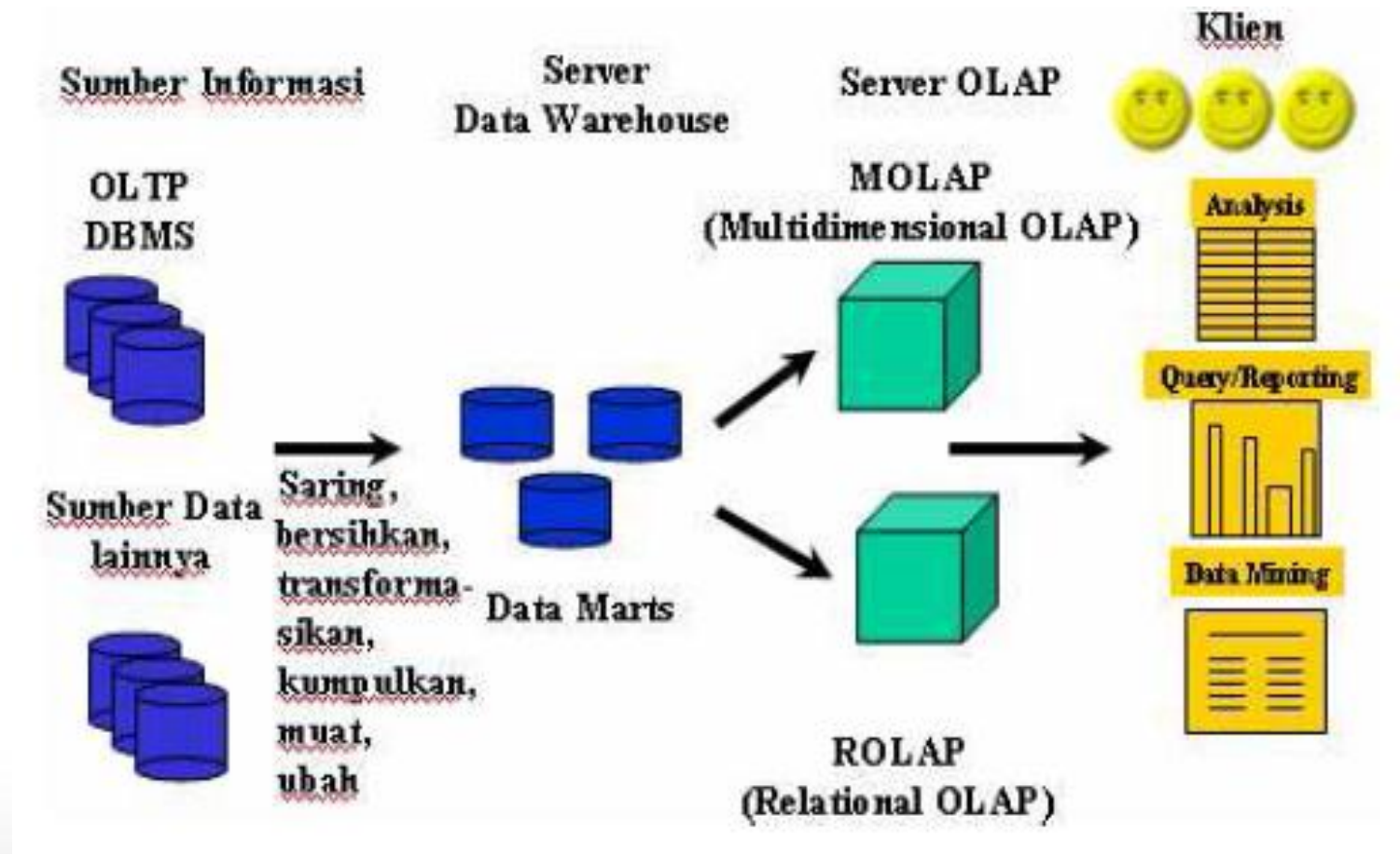
- **ukuran**

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

# Arsitektur Data Warehouse...



# Arsitektur Data Warehouse...



# OLAP (On-Line Analytical Processing)

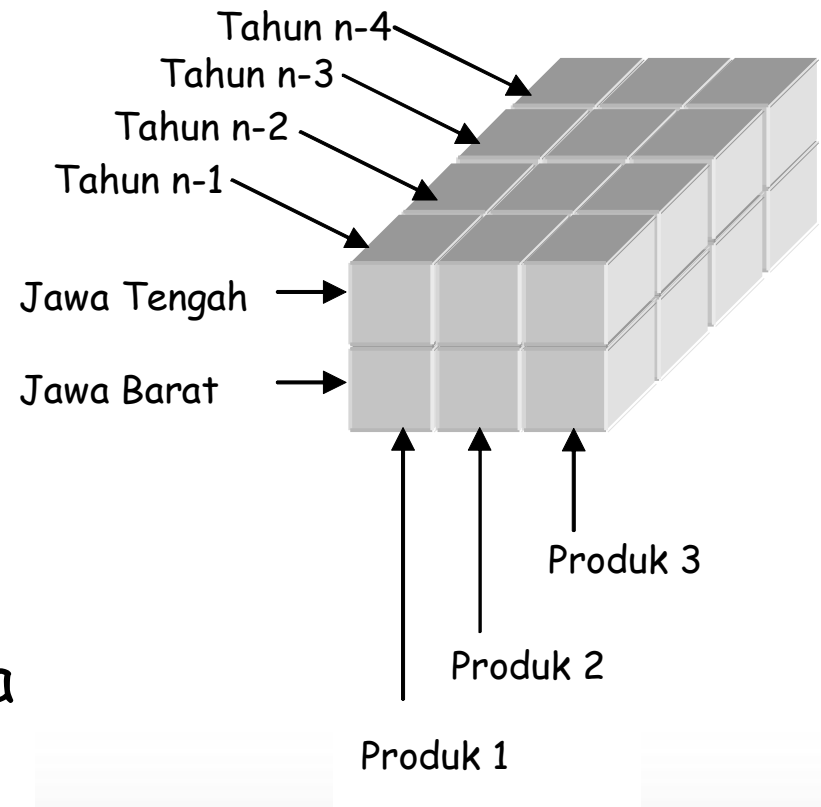
- **OLAP** adalah suatu sistem/teknologi yg dirancang utk **mendukung proses analisis kompleks** dalam rangka mengungkapkan kecenderungan pasar dan faktor-faktor penting dalam bisnis.
- **OLAP** ditandai dgn **kemampuannya menaikkan / menurunkan dimensi data** shg kita dapat menggali data sampai pada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

## Data Warehouse and OLAP as Multidimensional Data Model

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

# Multidimensional Data Model

- **Multidimensional** yang berarti bahwa terdapat **banyak lapisan kolom dan baris** (Ini berbeda dengan tabel pada model relasional yang hanya berdimensi dua).
- 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

## Fact Table (Tabel Fakta):

- Merepresentasikan **proses bisnis**, yaitu model proses bisnis sebagai **bentuk (artifact)** pada data model.
- Mengandung elemen '**pengukuran**' / **metrik** / **fakta** pada bisnis proses.
  - "jumlah penjualan bulanan" pada proses bisnis Penjualan.
  - dll
- Terdapat foreign key untuk tabel dimensi.
- Berisi ribuan kolom.

# Dimension Tables

## Dimension Table (Tabel Dimensi):

- 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...

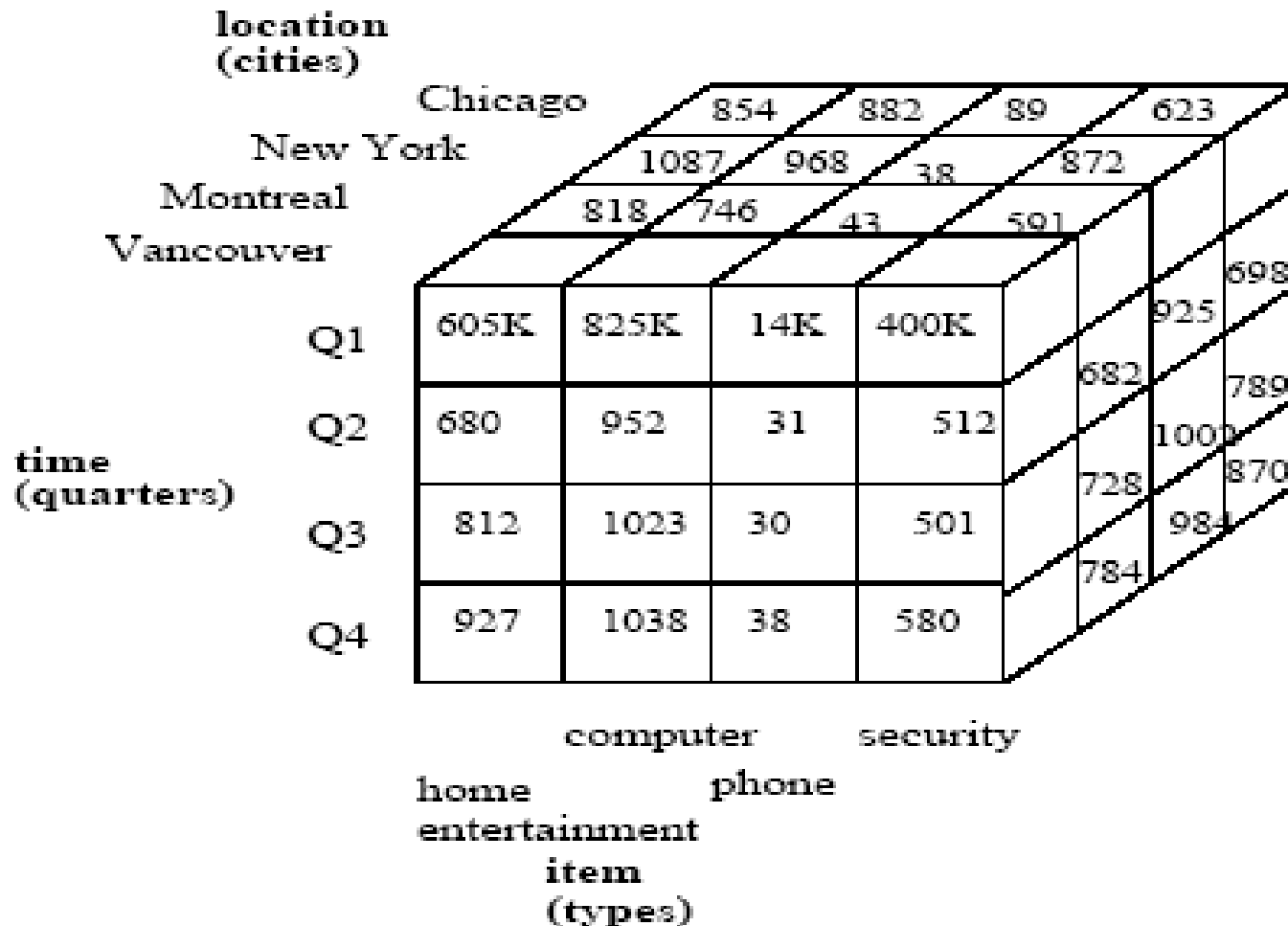
- Atribut pada tabel dimensi merupakan kolom<sup>2</sup> yg ada pada tabel dimensi.
- Pada tabel **dimensi lokasi**, atribut bisa merupakan Kode Pos, kota, kode negara, negara, dll. Secara umum, atribut pada tabel dimensi **digunakan** untuk pelaporan dgn menggunakan batasan query, contoh *where Country='USA'*. Atribut pada tabel dimensi juga mengandung satu / beberapa hierarchical relationships.
- Sebelum mendesain data warehouse, harus ditentukan dulu mengandung elemen apa saja data warehouse yang akan dibuat. Misalkan jika ingin membuat data warehouse yg mengandung "**jumlah penjualan bulanan**" yg melibatkan semua lokasi, waktu, dan produk, maka **dimensinya** adalah :

Lokasi

Waktu

Product

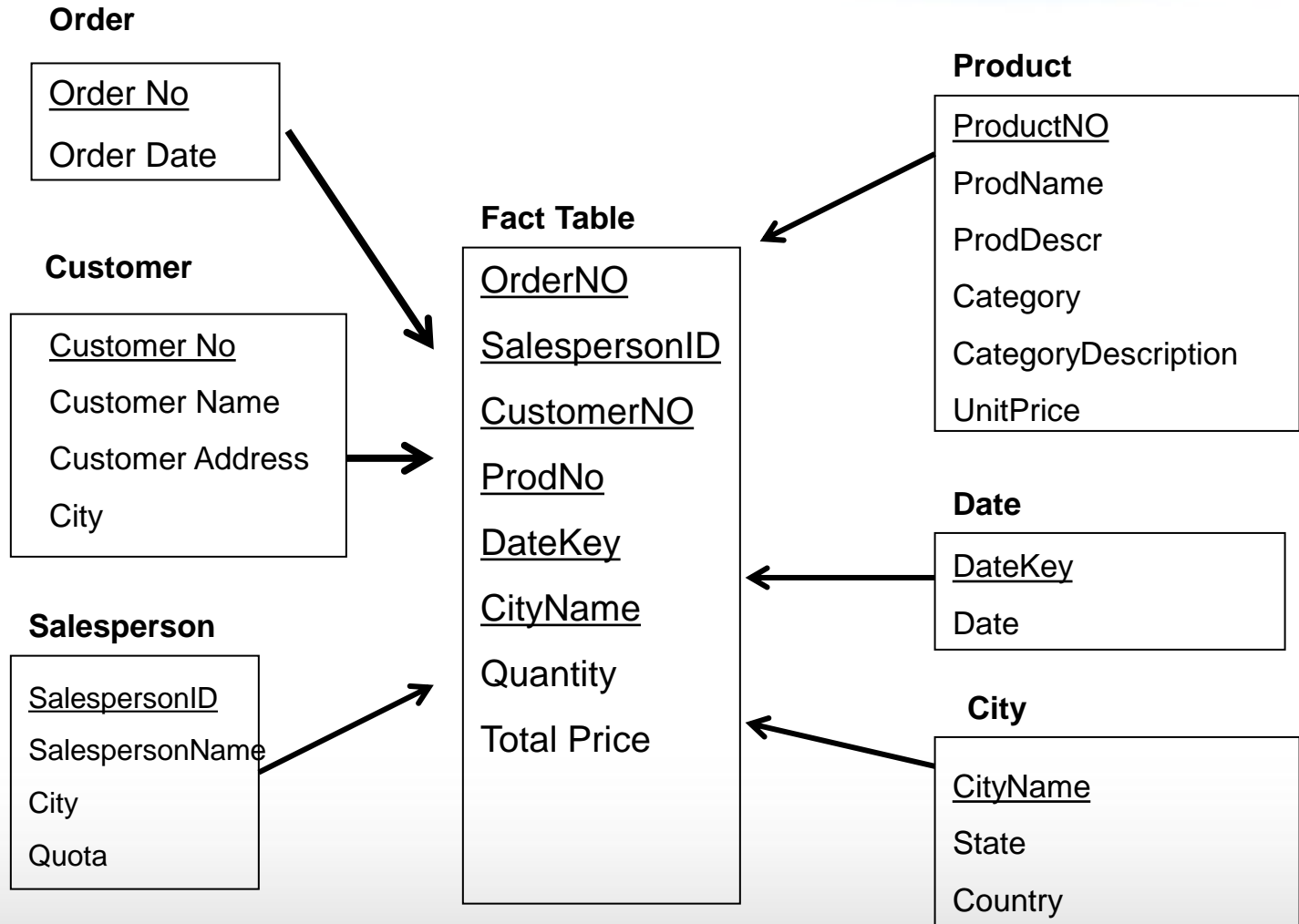
# 3D data cube, according to the dimension time, location, item



# Warehouse Database Schema

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

# Example of a Star Schema



# Example of a Star Schema...

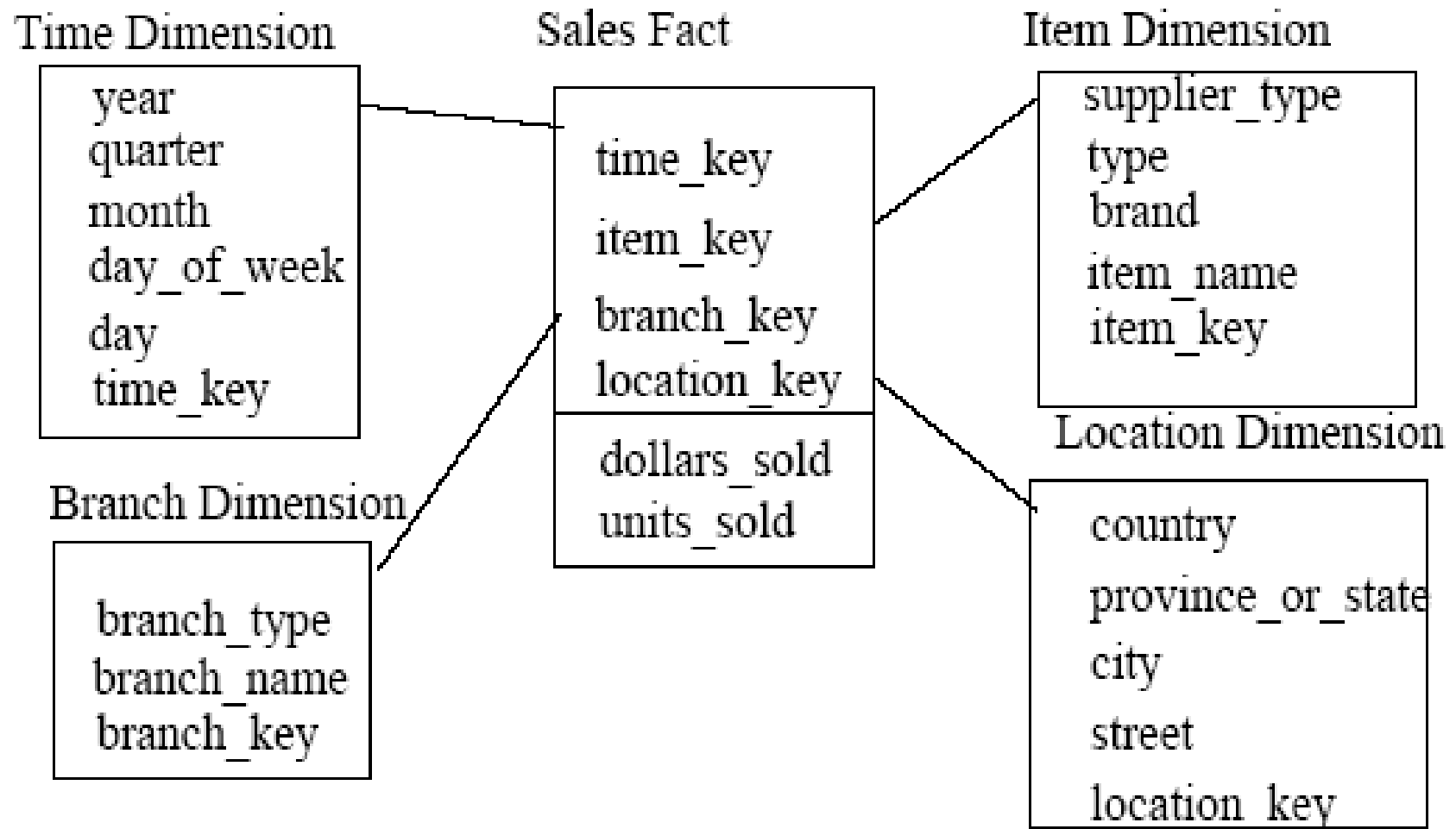
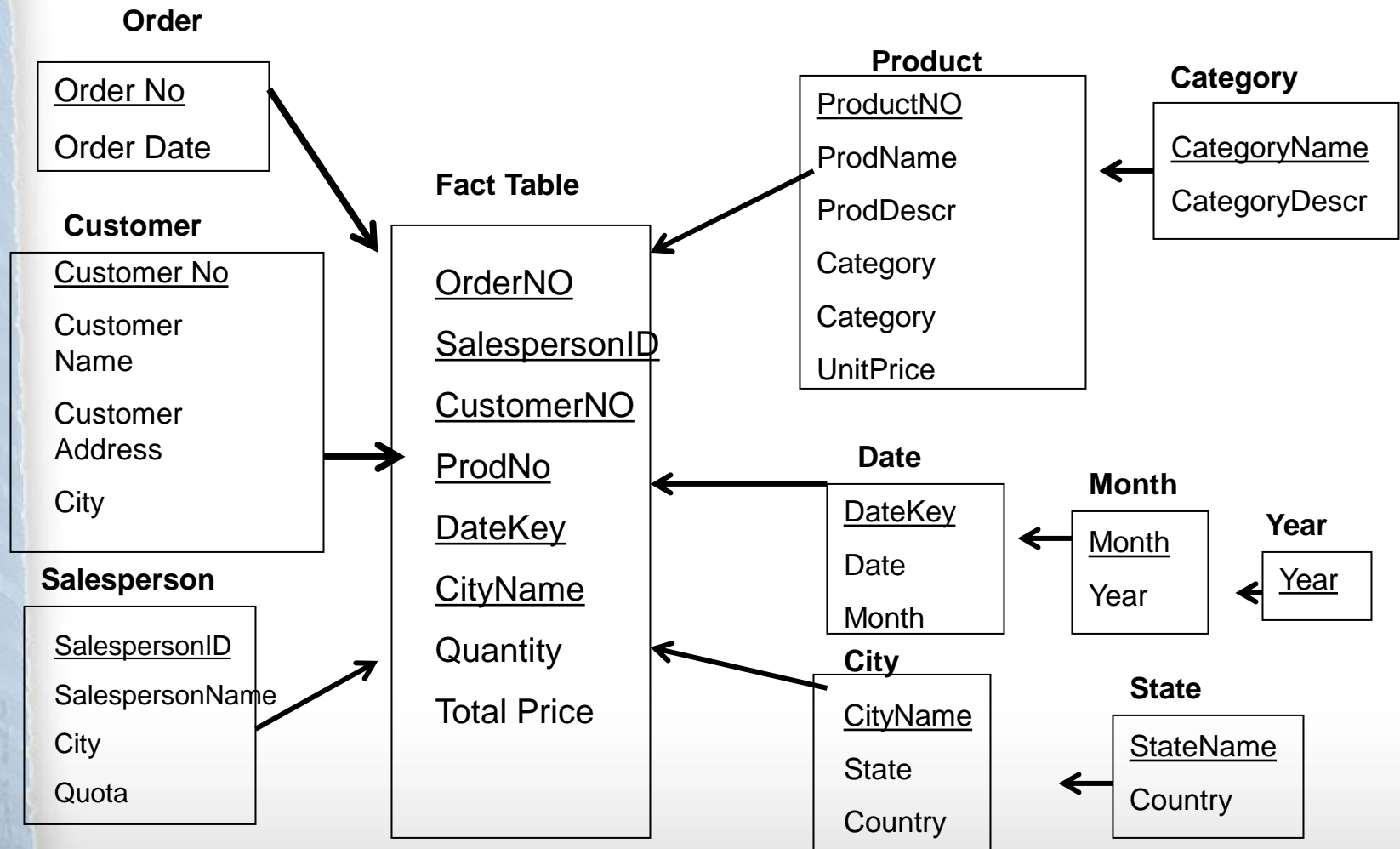


Figure 2.4: Star schema of a data warehouse for sales.

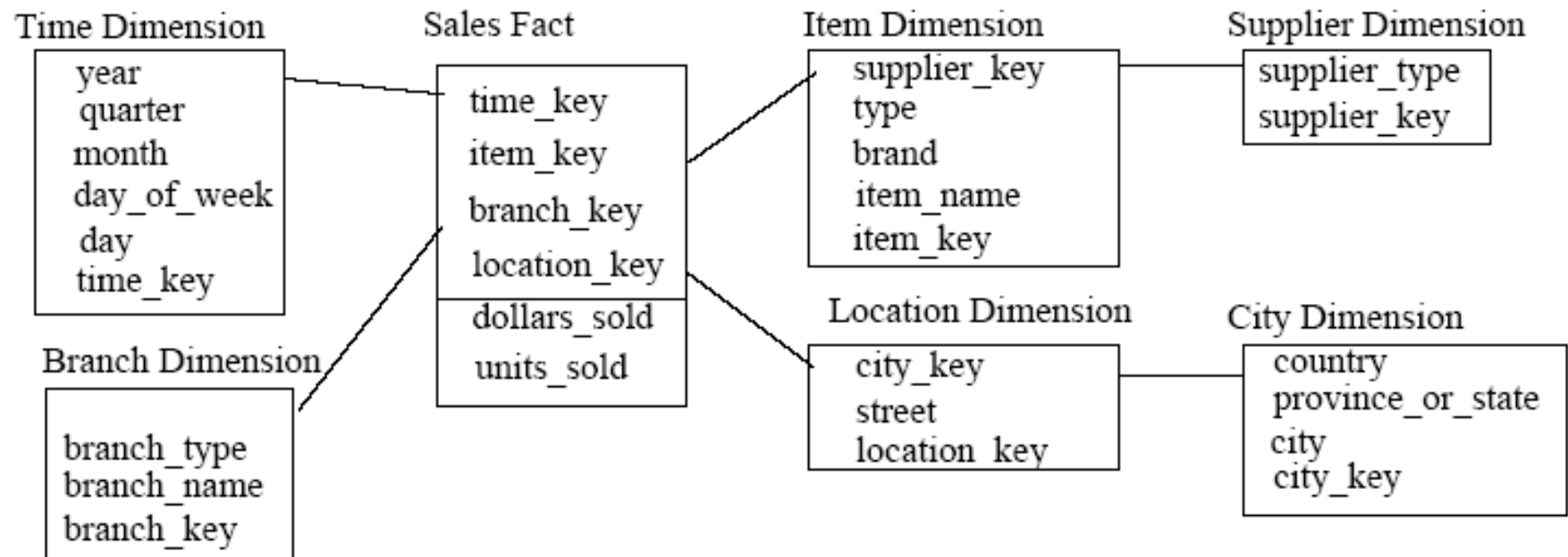
# Star Schema

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

# Example of a Snowflake Schema



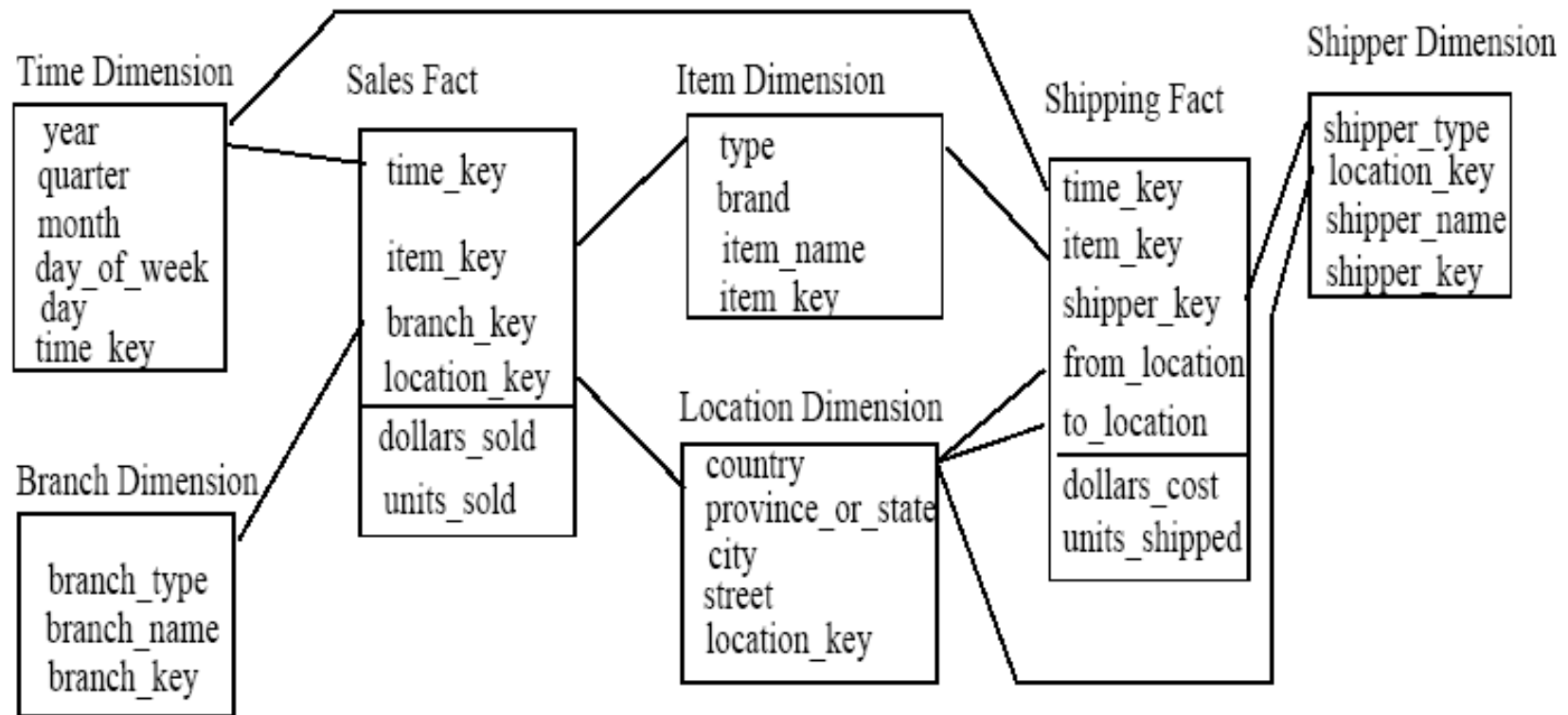
# Snowflake Schema...



# Snowflake Schema

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

# Fact Constellation



# Fact Constellation

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

**S.E.L.E.S.A.I. . .**