

# Bab 4

# Skema data warehouse

Dedi Gunawan, Ph.D & Dr.Eng. Yusuf Sulistyo N

# OUTLINE

- Pengenalan skema data warehouse
  - Schema (skema)
  - Dimension (dimensi)
  - Measure (jumlah)
  - Fact table(tabel fakta)
  - Multi dimensional data view
- Star schema (skema bintang)
- Snowflake Schema (skema bunga salju)
- Fact constellation schema (skema galaksi)

# Pengenalan skema datawarehouse

# Schema (skema)

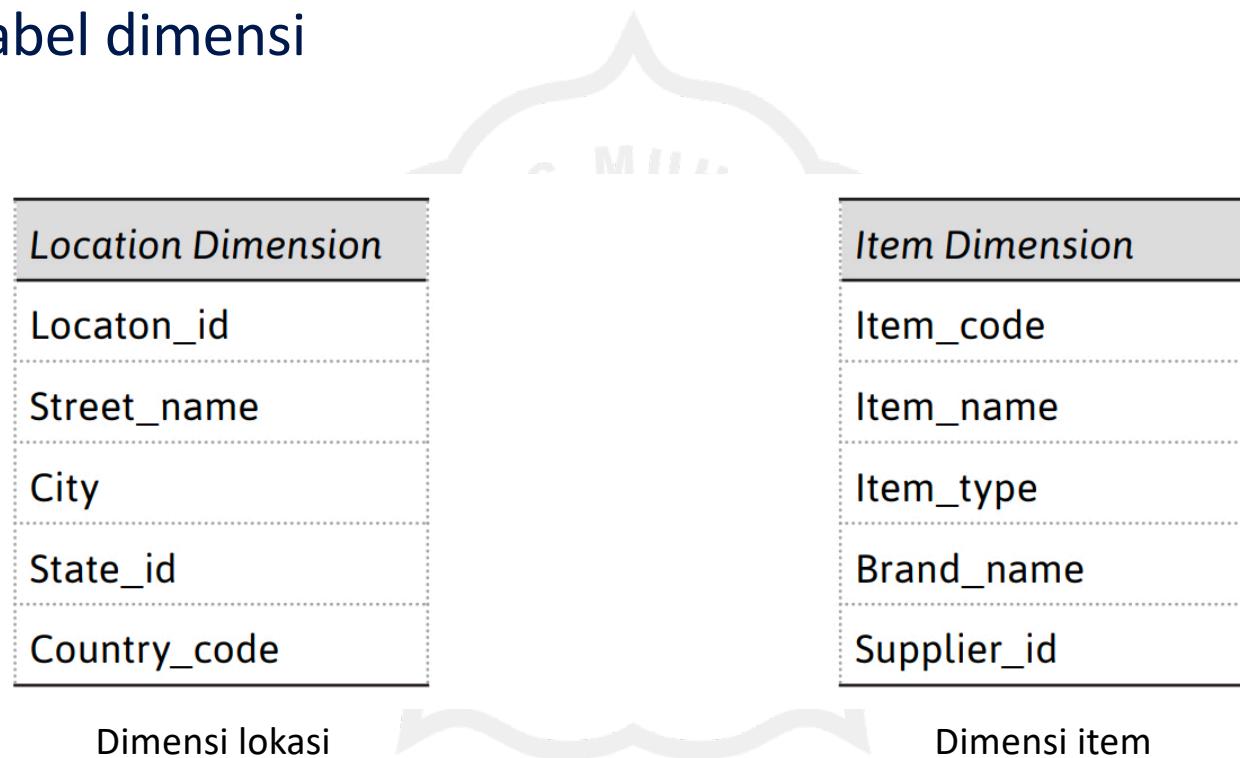
- Skema merupakan deskripsi logis (logical description) dari suatu database.
- Skema merupakan *blueprint* dari keseluruhan database.
- Skema mendeskripsikan bagaimana sekelompok data tersusun dan bagaimana hubungan dan keterkaitan antara satu data dengan data yang lainnya.
- Skema pada datawarehouse terdiri dari nama dan deskripsi suatu record termasuk di dalamnya adalah item-item datayang terkait dengan record tersebut dan agregat dari data.
- Database umumnya menggunakan model relasional, sementara pada data warehouse menggunakan skema.
- Untuk memulai diskusi tentang skema, ada beberapa terminology/istilah yang perlu diketahui diantaranya *dimension*, *Measure*, *Fact table* dan *Multi dimensional data view*.

- Istilah dimension atau dimensi pada datawarehouse merupakan suatu koleksi atau kumpulan informasi mengenai kejadian-kejadian yang bisa diukur.
- Dimensi umumnya merupakan suatu entitas yang digunakan oleh institusi untuk menyimpan datanya.
- Atribut deskriptif diorganisasi atau disusun sebagai kolom-kolom pada tabel dimensi oleh datawarehouse.
- Contoh :
  - attribute dari dimensi siswa bisa berupa nama, nim, umur, tanggal lahir, jenis kelamin.
  - Dimensi alamat meliputi nama jalan, desa, kelurahan, kecamatan, kabupaten dan provinsi

- Tabel dimensi memiliki *primary key* yang bersifat unik, sebagai identifier/penanda suatu record dari dimensi.
- Umumnya dimensi merupakan suatu tabel yang berupa *de-normalized* artinya sangat mungkin terdapat duplikasi data di dalam tabel dimensi tersebut.
- Dalam desain database biasanya setiap tabel harus di normalisasikan untuk menghindari duplikasi data.
- Pada de-normalization biasanya tabel-tabel kecil akan digabungkan menjadi satu sehingga mengurangi penggunaan *join*.
- De-normalization diterapkan untuk kasus-kasus yang lebih mementingkan proses *retrieval* dari pada proses *insert, update* dan *delete*.

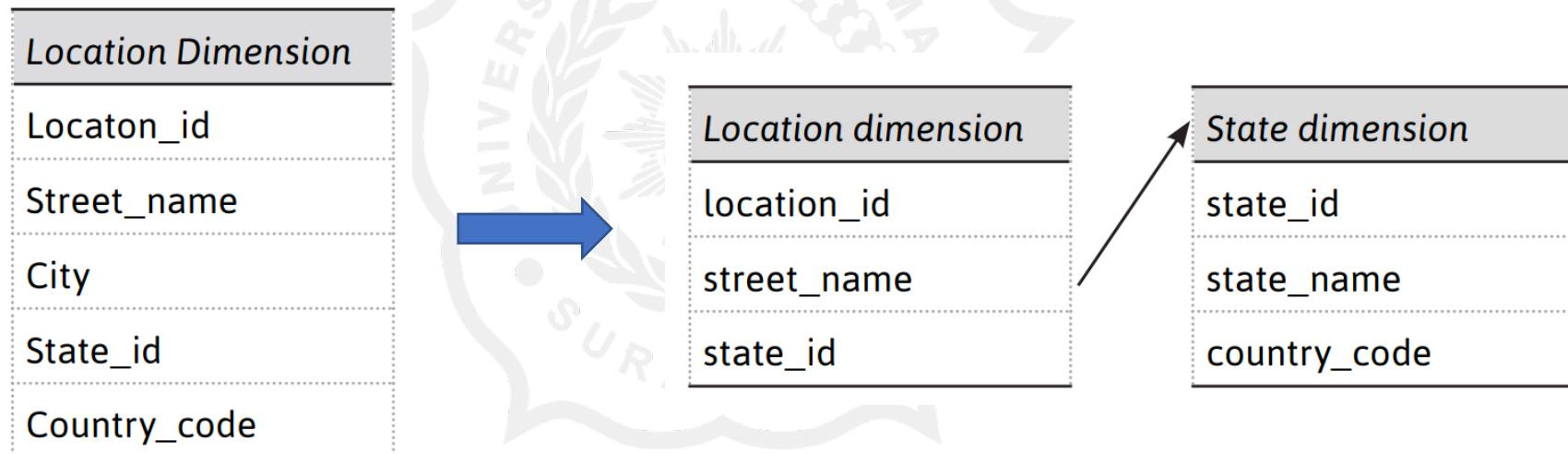
# Dimension (dimensi)

- Contoh tabel dimensi



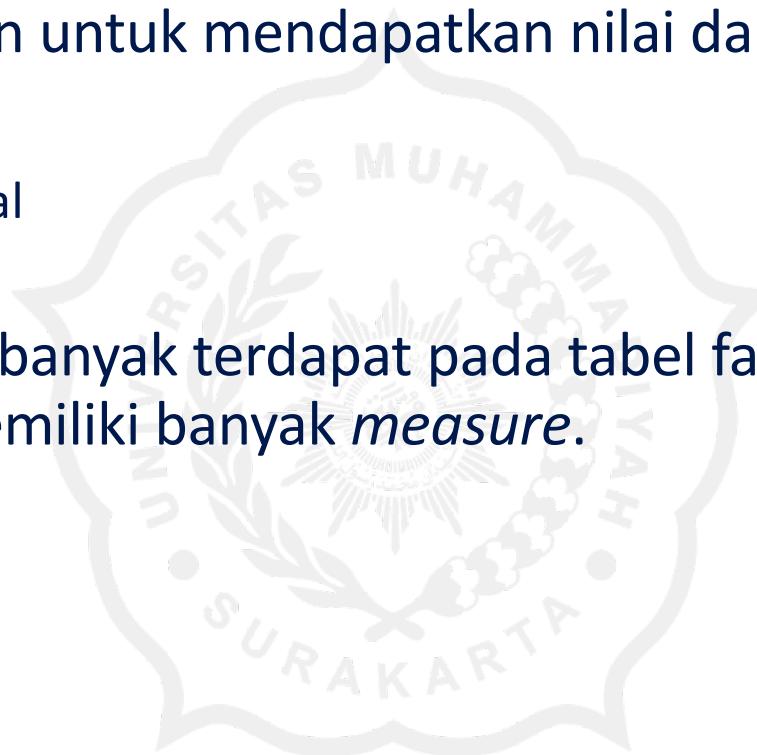
# Dimension (dimensi)

- Dimensi lokasi sebetulnya bisa di-normalisasikan
- Tetapi karena tujuan utamanya adalah proses retrieval maka redundansi pada dimensi lokasi tidak menjadi suatu masalah



# Measure (jumlah/kalkulasi)

- *Measure* digunakan untuk mendapatkan nilai dari suatu tabel dimensi
- Contoh :
  - Jumlah item terjual
  - Jumlah penjualan
- *Measure* biasanya banyak terdapat pada tabel fakta dan umumnya satu tabel fakta bisa memiliki banyak *measure*.

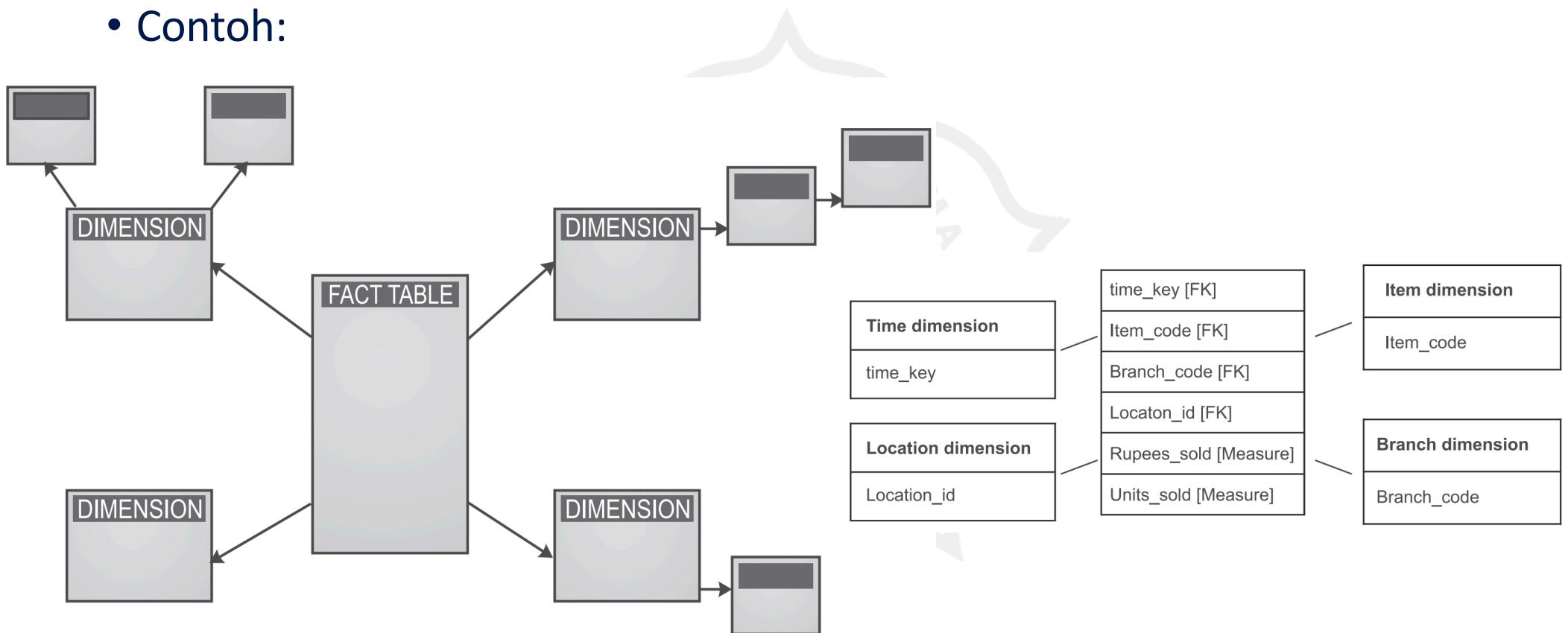


# Fact table (tabel fakta)

- Suatu tabel fakta merupakan sekelompok item-item data yang saling berhubungan.
- Tabel fakta terdiri dari data-data dari tabel dimensi dan measure.
- Tabel fakta hanya memiliki dua jenis kolom yaitu foreign key dan measure
  - Foreign key menghubungkan tabel fakta ke tabel dimensi dan measure berisi nilai dari data
- Tabel fakta biasanya memiliki ukuran yang besar dari pada tabel dimensi

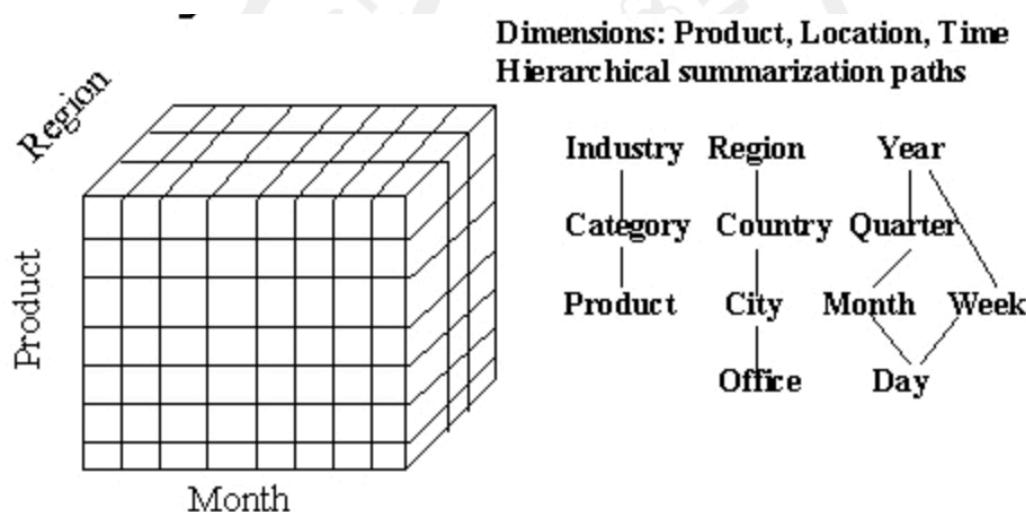
# Hubungan tabel dimensi dengan tabel fakta

- Contoh:



# Multi-dimensional view of data

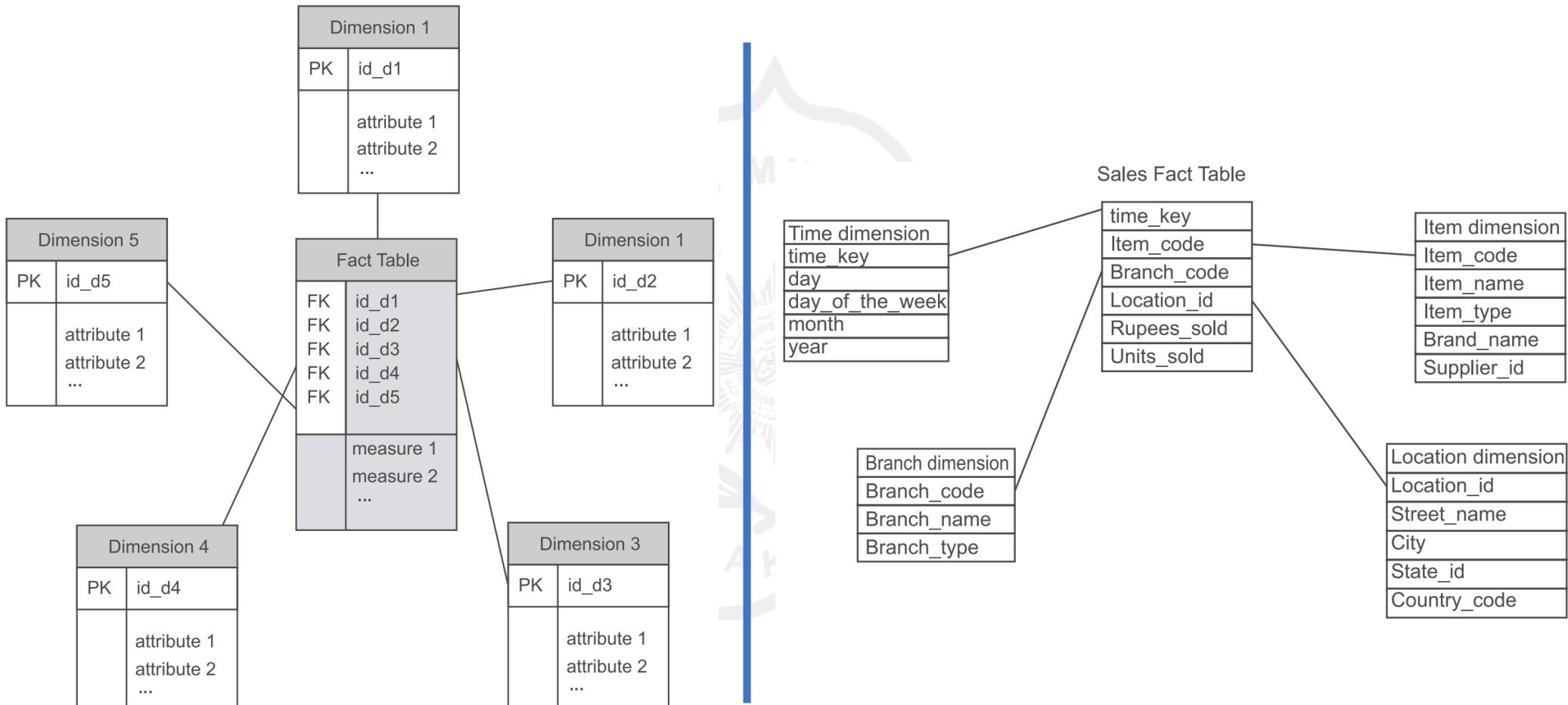
- Data terdiri dari multiple dimensi, dimana dimensi memiliki tingkatan atau hirarki untuk menunjukkan hubungan *parent-child* dari tabel dimensi





- Star schema merupakan schema yang paling simple.
- Dinamakan star schema karena bentuknya yang mirip bintang dimana tabel fakta berada di tengah-tengah dan tabel dimensi berada pada setiap ujungnya.
- Setiap dimensi pada star schema merepresentasikan satu tabel dimensi yang memiliki berbagai atribut.
- Umumnya, tabel fakta merupakan bentuk normalisasi 3NF pada star schema sementara itu tabel dimensi berbentuk de-normalisasi.

# Star schema



# Karakteristik Star schema

1. Memiliki performa query yang bagus karena tidak memerlukan banyak join.
2. Memiliki struktur yang sederhana dan mudah dipahami.
3. Memerlukan waktu load yang lama karena de-normalisasi pada tabel dimensi, serta menjadikan ukuran data menjadi lebih besar.
4. Paling banyak digunakan dan paling sederhana strukturnya serta dukungan tools yang banyak.

# Kelebihan Star schema

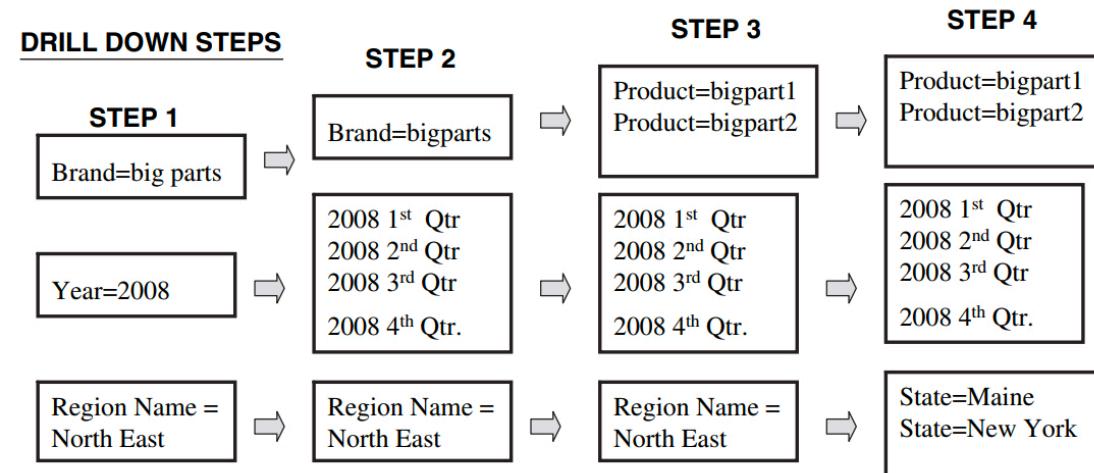
1. Sederhana dan tidak memerlukan banyak join.
2. Star schema memiliki business reporting logic.
3. Memiliki performa query yang tinggi.
4. Memiliki fungsi agregasi yang cepat Karena query yang sederhana.

# Kelemahan Star schema

1. Tidak bisa mengaplikasikan data integrity karena adanya de-normalisasi dan terdapat redundansi data.
2. Proses insert dan update pada star schema bisa menghasilkan anomaly data.
3. Kurang fleksibel untuk data analisis dibandingkan dengan tabel-tabel yang ter-normalisasi
4. Terlalu spesifik untuk kasus tertentu sehingga tidak memungkinkan untuk analisa yang kompleks.

# Menganalisa star schema

- Analisis yang umum dilakukan pada star schema adalah melakukan *drilling down* (penggalian) untuk mendapatkan detail pada level data yang lebih rendah.
- Misalkan divisi marketing melakukan permintaan analisis data sebagai berikut:
  - Tampilkan jumlah total produk dengan merk “**big parts**” yang terjual di region “northeast region” pada tahun “2008”

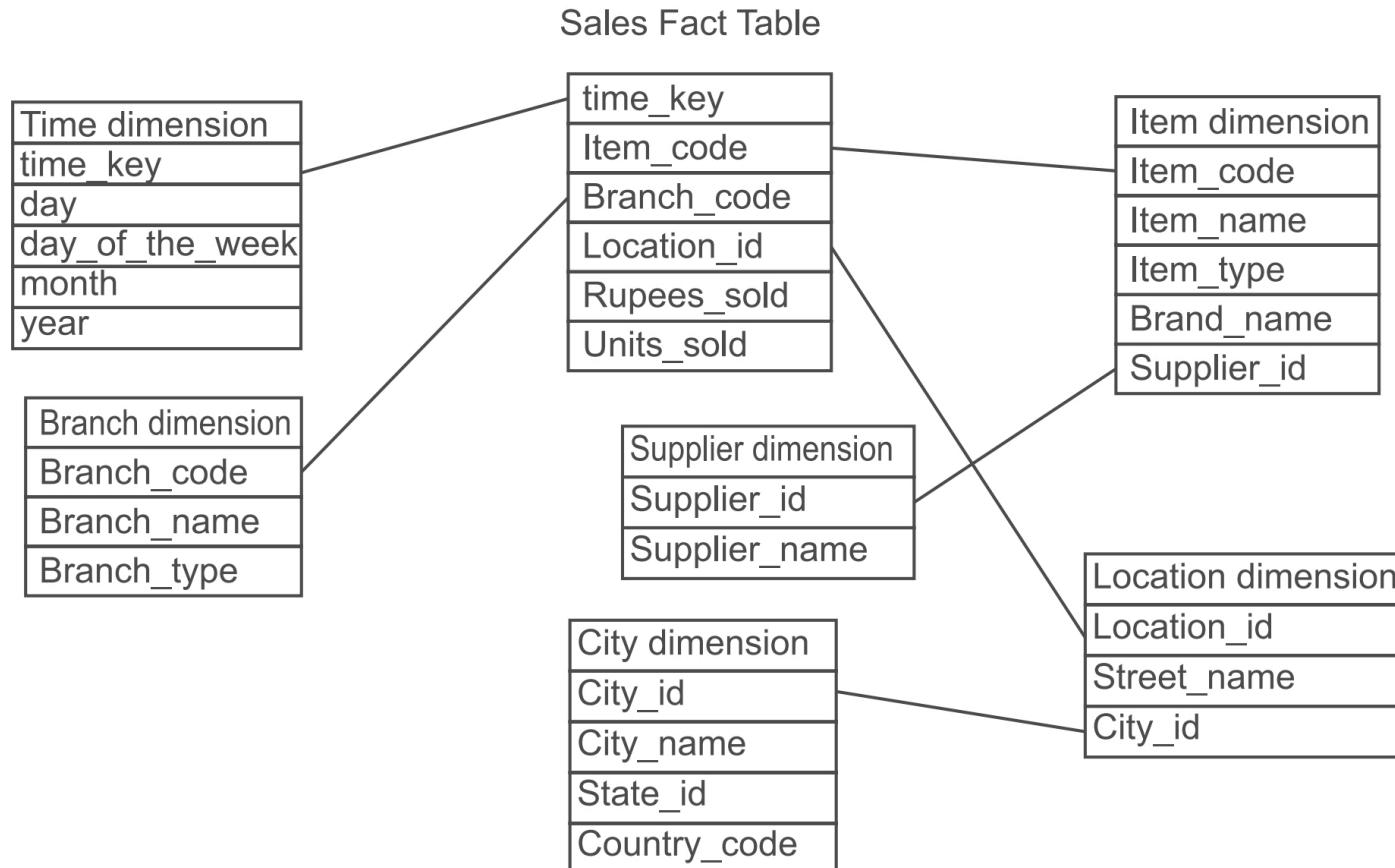


# Snowflake Schema

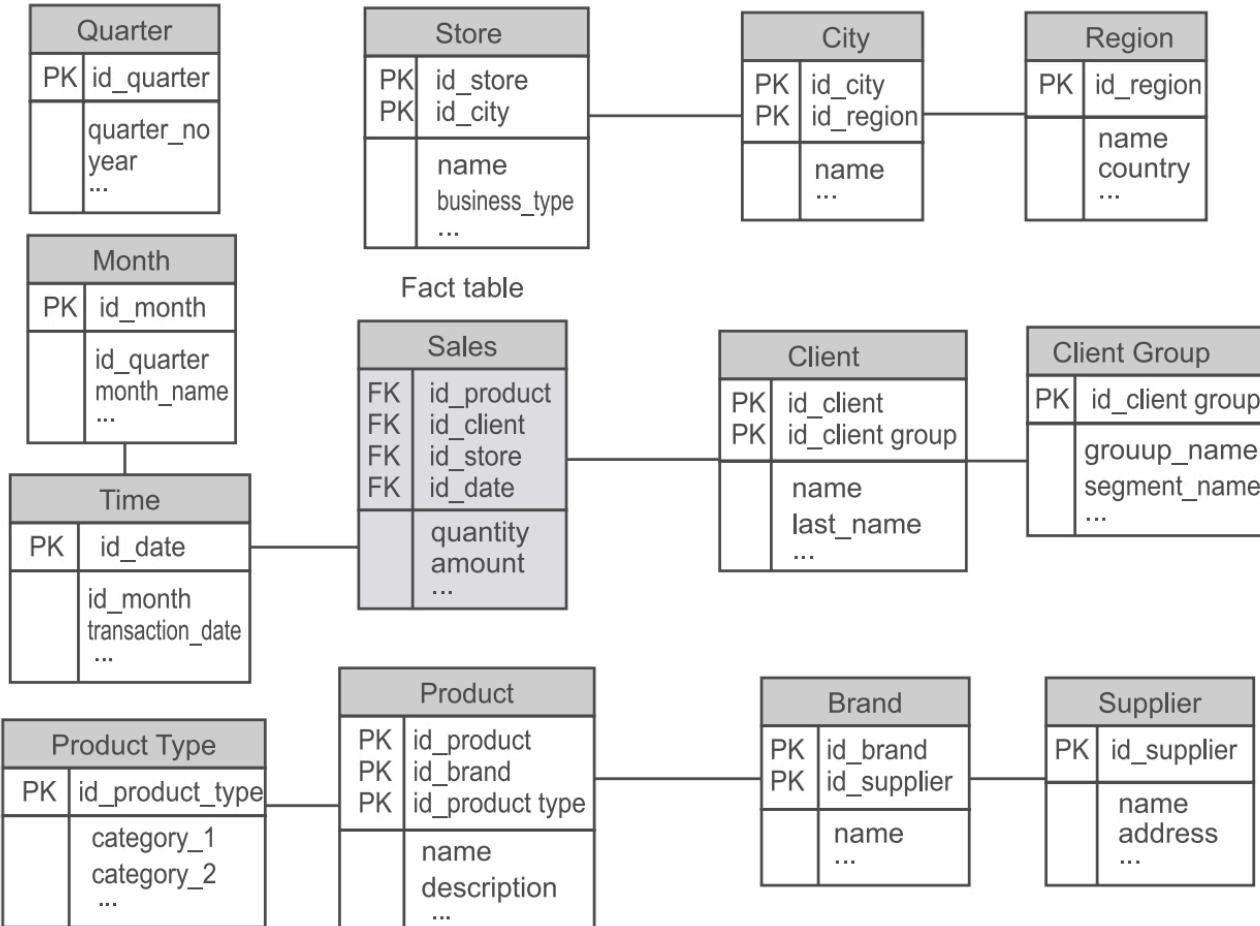
# Snowflake Schema

- Snowflake Schema terdiri dari dimensi tabel yang sudah ter-normalisasi.
- Sehingga sebetulnya Snowflake Schema merupakan perbaikan untuk star scheme dalam hal normalisasi.
- Beberapa tabel perlu di-normalisasikan pada Snowflake Schema yang memisahkan data ke tabel lainnya.
- Struktur akhir dari Snowflake Schema adalah tabel fakta berada ditengah-tengah skema setelah semua tabel dimensi di-normalisasikan.

# Snowflake Schema



# Snowflake Schema

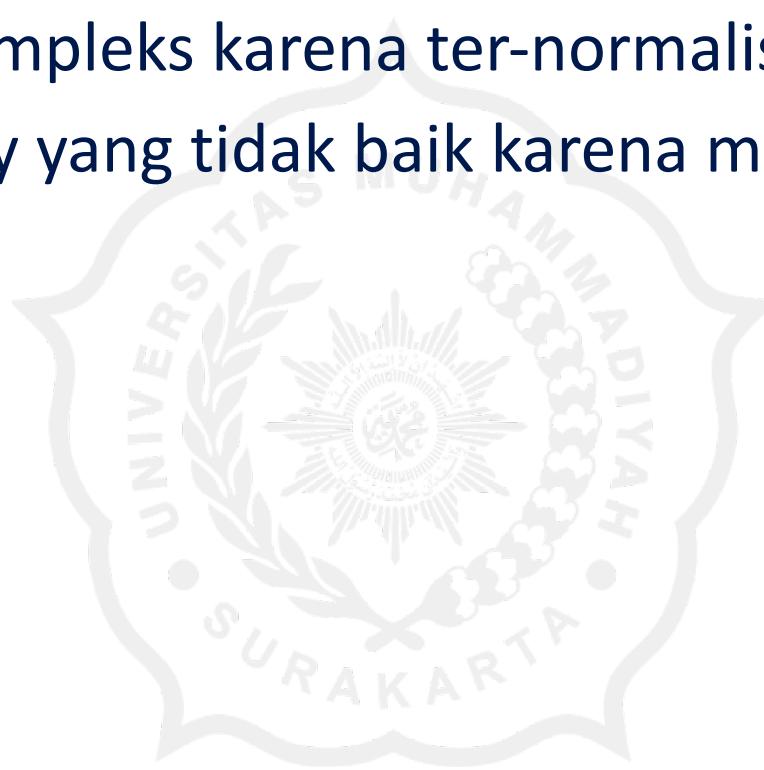


# Keunggulan Snowflake Schema

1. Lebih menghemat storage meskipun memerlukan query yang lebih kompleks.
2. Beberapa tools untuk modeling OLAP memiliki fitur yang sudah dioptimisasi untuk snowflake schema.

# Kelemahan Snowflake Schema

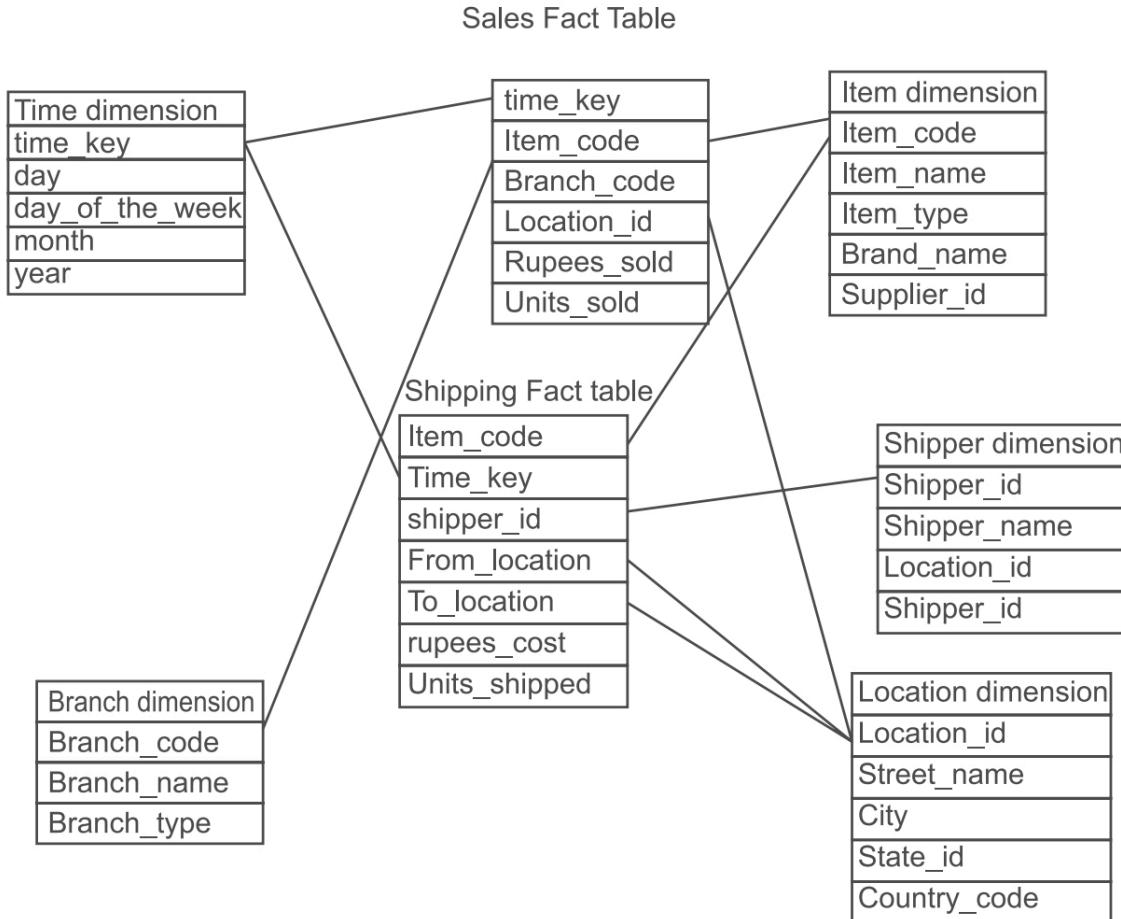
1. Query lebih kompleks karena ter-normalisasi
2. Performa query yang tidak baik karena melibatkan banyak join





- Yang membedakan Galaxy Schema dengan schema yang lainnya adalah pada schema ini terdapat beberapa tabel fakta
- Galaxy schema merupakan digunakan untuk menangani online analytical processing (OLAP) dimana beberapa tabel fakta melakukan sharing tabel-tabel dimensi
- Galaxy schema merupakan perbaikan dari skema star

# Galaxy Schema



- Kelebihan dari galaxy scheme adalah kemampuannya untuk menyediakan support pada user karena memiliki tabel fakta yang bisa digunakan untuk berbagai keperluan
- Sedangkan kekurangan dari schema ini adalah desain yang rumit dikarenakan schema ini mempertimbangkan berbagai jenis agregasi.

Salamat Gamsahabnida Gratias Ago Vos Dankie dat U Sas efcharistó Thank You Merci Asante Teşekkürler Grazie Köszönöm Tack Obrigado Hvala Arigato Ngiyabonga Sağ Olun Gràcies Xièxiè Nín Dankon Danke Dziekujemy Spasibo Va Multumesc Go raibh maith agat Gracias

Dioilch yn fawr bekka bér Khokhobchai Than Dank U Kasih