



# Data Warehousing

---

Yudi Agusta, PhD

Kecerdasan Bisnis, Lecture 03 04 05 06 07

Copyright © Yudi Agusta, PhD 2024



# Definisi Data Warehouse

---

- A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process – Inmon and Hackathorn (1994)
- “A copy of transaction data specifically structured for query and analysis.” Ralph Kimball, *The Data Warehouse Toolkit*, P. 310.
- Sebuah sistem yang terdiri dari beberapa proses dan database yang digunakan untuk menyediakan “infrastruktur data” bagi EIS dan DSS

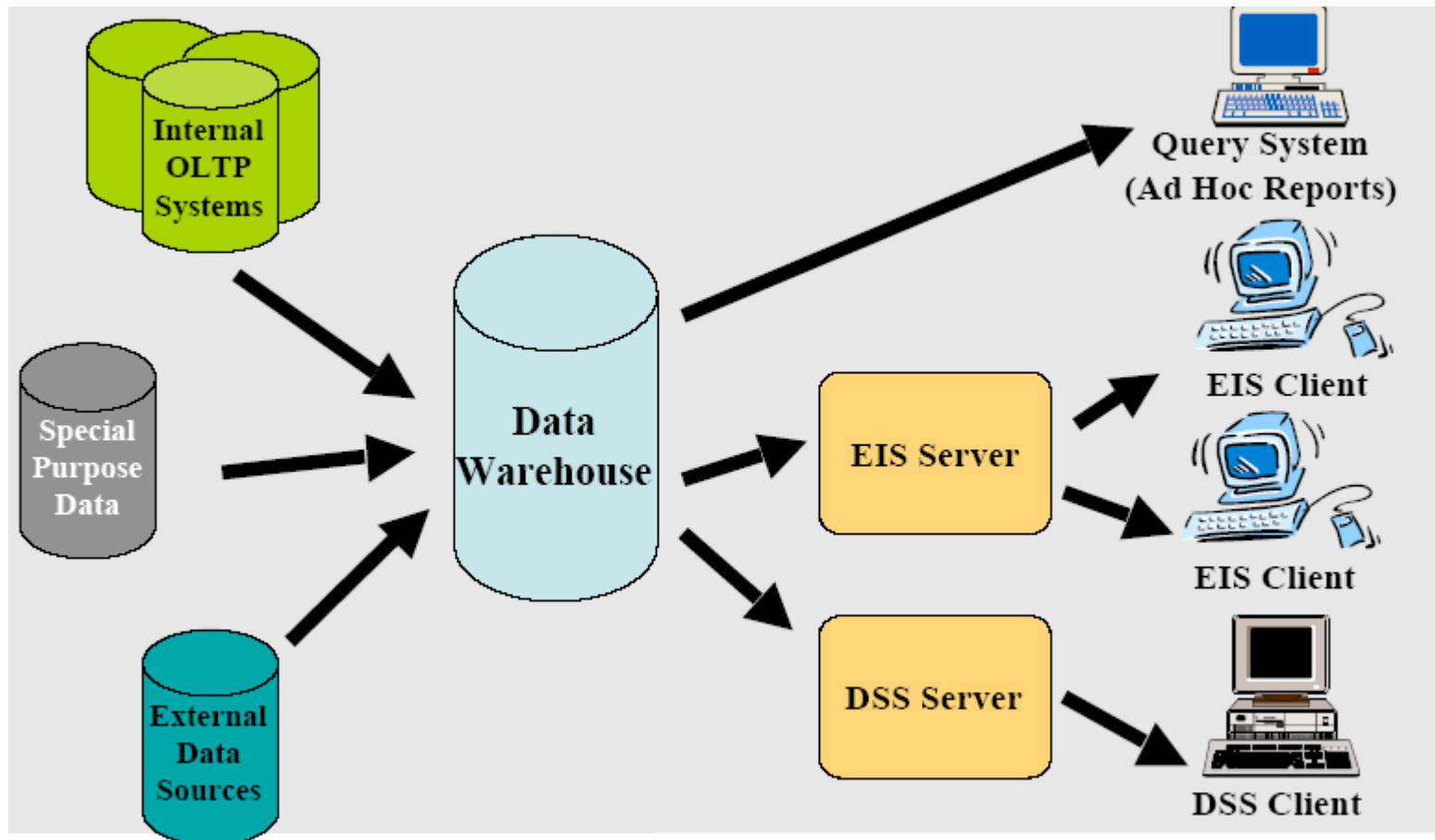


# Proses Data Warehouse

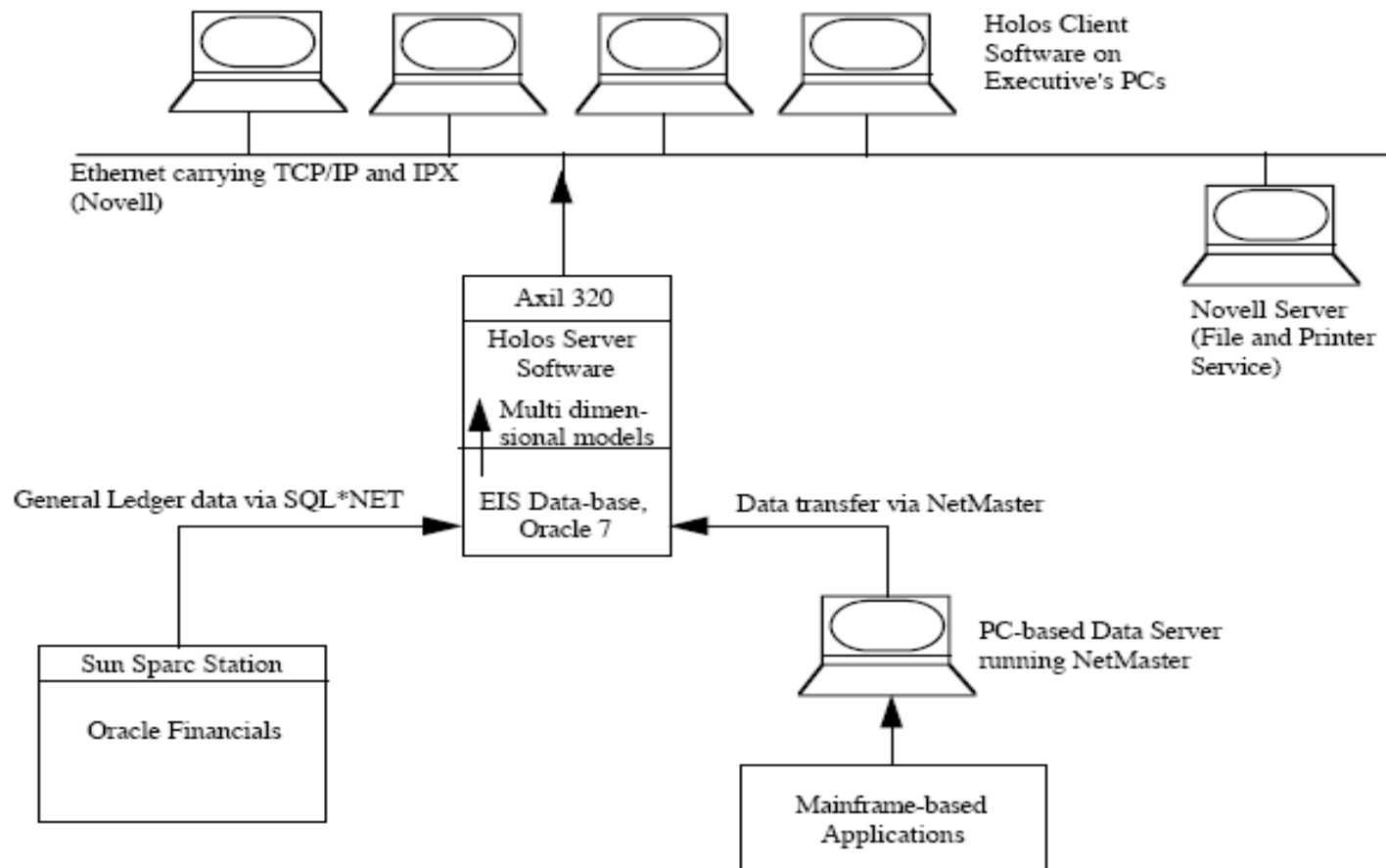
---

- Data warehouse adalah sebuah proses evolusi mencakup pencarian sumber, penyimpanan dan penyediaan data untuk mendukung pengambilan keputusan
- Pengembangan database untuk suatu data warehouse (= sistem komputer) adalah satu komponen dari keseluruhan proses data warehouse

# Arsitektur Data Warehouse Umum



# Sebuah Data Warehouse





# Motivasi Sebuah Data Warehouse

---

- Tekanan terhadap database OLTP untuk proses query terlalu besar
- Data warehouse didisain untuk proses pengambilan data/informasi yang efisien
- Data pada sistem yang berbeda-beda umumnya tidak konsisten, kualitasnya buruk dan disimpan di dalam format yang berbeda
- Mengurangi biaya dalam menyediakan data untuk keperluan pengambilan keputusan
- Mendukung pengambilan keputusan yang lebih cepat dan efisien



# Motivasi Sebuah Data Warehouse

---

- Mendukung untuk memfokuskan diri pada proses bisnis lengkap
- Lebih efisien dalam membaca keperluan konsumen dan pasar
- Mendukung inisiatif baru
- Mendukung untuk bisa tetap kompetitif



# Tujuan Data Warehouse

---

- Menyediakan akses pada data/informasi perusahaan atau organisasi
- Data di dalam sebuah warehouse bersifat konsisten
  - Satu versi dari kebenaran
- Data warehouse adalah tempat dimana data yang telah dikumpulkan dipublikasikan
- Kualitas data di dalam data warehouse adalah kritikal
  - “Kualitas – ketepatan untuk suatu tujuan”





# Karakteristik Data Warehouse

---

- Berorientasi Subject
- Terintegrasi
- Time Variant
- Not Volatile



## Data Warehouse – Berorientasi Subjek

---

- Data warehouse diorganisasikan oleh “data subyek” yang terkait dengan organisasi
  - Customer, Claim, Shipment, Product
- Ini mungkin bertolak belakang dengan sebagian besar sistem OLTP yang berorientasi proses
  - OLTP = Online Transaction Processing



# Data Warehouse - Terintegrasi

---

- Data di dalam data warehouse distrukturisasi berdasarkan pada model korporasi secara keseluruhan melewati batas fungsional dari kebiasaan yang umumnya ada di dalam suatu perusahaan
- Ini mencakup standar penamaan, sistem pengukuran dan perulangan



# Data Warehouse – Time Variant

---

- Data di dalam data warehouse mempunyai karakter khusus berupa time-series dalam bentuk data historical
- Data terdiri dari suatu seri dari “pemotretan” keadaan yang ditandai dengan waktu dan nilai data pada saat “pemotretan” tersebut
- Ini dapat digunakan untuk melakukan trend analysis dari data tersebut



# Data Warehouse – Not Volatile

---

- Data warehouse tidak diupdate secara terus menerus (seperti inserts, deletes dan changes) seperti dalam sebuah sistem OLTP. Change hanya dilakukan apabila data yang telah dimasukkan dalam data warehouse mengalami salah hitung, konversi, atau ambil
- Data di dalam data warehouse secara periodik di-upload dalam jangka waktu yang sama (misalnya setiap pagi atau setiap tanggal 2)



# Pengembangan Data Warehouse

---

- Mendefinisikan keperluan
- Mendisain data warehouse
  - Database Data Warehouse
  - Extraction, Transformation & Loading Tool
- Pencarian dan pembersihan data
- Mengimplementasikan data warehouse
  - Harus digulirkan dalam bentuk tahapan



# Data Warehousing Sebagai Suatu Proses

---

- **Dinamika fundamental di pusat data warehousing = suatu proses pembelajaran**
  1. Data yang ada di dalam data warehouse akan mengubah pengertian pengguna akhir terhadap pekerjaan mereka. Hal ini berhubungan dengan pencapaian target dan kondisi lingkungan internal dan eksternal
  2. Dengan adanya perubahan pengertian terhadap pekerjaan mereka, keperluan informasi di tingkat operasional juga akan berubah
  3. Dengan berubahnya informasi yang diperlukan di Tingkat operasional, maka disain data warehouse juga akan berubah
  4. Proses akan kembali ke poin 1



# OLTP vs OLAP

---

Yudi Agusta, PhD

Kecerdasan Bisnis, Lecture 04 05 06 07

Copyright © Yudi Agusta, PhD 2024





# OnLine Transaction Processing

---

- On-Line Transaction Processing System
- Transaksi proses (= kegiatan bisnis – penjualan, pemesanan, pembayaran dll)
- Umumnya digunakan di 'bagian depan' bisnis oleh orang-orang yang berinteraksi langsung dengan para pelanggan, supplier, pegawai dll
- Dikarakterisasikan dengan data entry, update dan deletion
- Biasanya terfokus pada satu area bisnis (seperti POS – point-of-sales, inventory, purchasing, human-resources)



# Sistem Informasi = OLTP

---

- Data:
  - Komponen terpenting sebagai penghubung antara mesin (hardware) dan manusia
  - Fakta mentah yang belum di olah
- Informasi:
  - Data yang sudah diolah menjadi bentuk yang berarti bagi penggunaanya untuk tujuan pengambilan keputusan
  - Hasil dari data mentah yang telah diproses untuk memberikan hasil bagi penggunaanya
- Sistem Informasi:
  - Proses untuk mengumpulkan, memproses, menganalisa, dan menyebarkan informasi untuk tujuan tertentu
  - Suatu kombinasi sistematis antara orang, perangkat keras, perangkat lunak, dan database dalam mengumpulkan, mengubah, dan menyebarkan informasi di dalam suatu organisasi



# Data Warehouse - OLAP

---

- Sebuah system yang bertugas mengarsipkan sekaligus melakukan analisis data historis untuk menunjang pengambilan keputusan pada suatu organisasi
- Suatu tempat penyimpanan data yang berasal dari berbagai sumber agar nantinya bisa digunakan untuk pelaporan dan analisis bisnis
- Tempat penyimpanan data pada bisnis yang nantinya digunakan untuk menganalisa permintaan konsumen



# OnLine Analytical Processing

---

- On-Line Analytical Processing
- Umumnya digunakan oleh pembuat kebijakan seperti analis dan manajer
- Data yang ada di dalam sebuah data warehouse umumnya berasal dari bermacam-macam OLTP systems
- Dikarakterisasikan oleh data reading (pelaporan, queries) bukan data entry

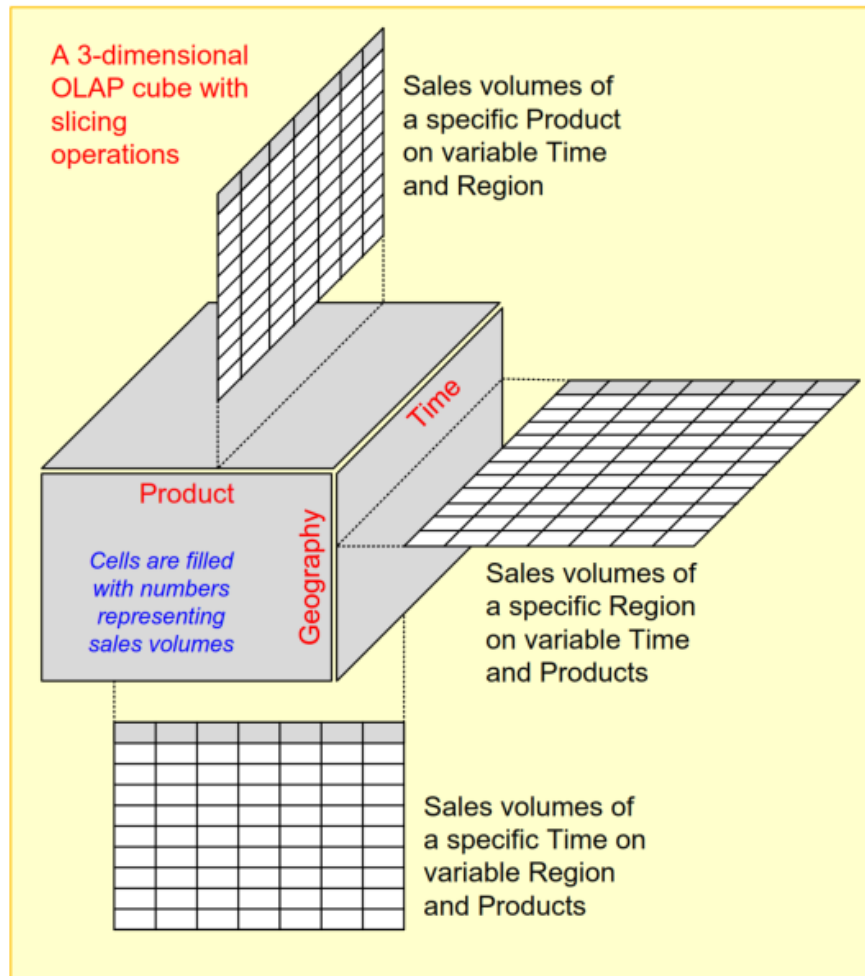


# Fitur OLAP

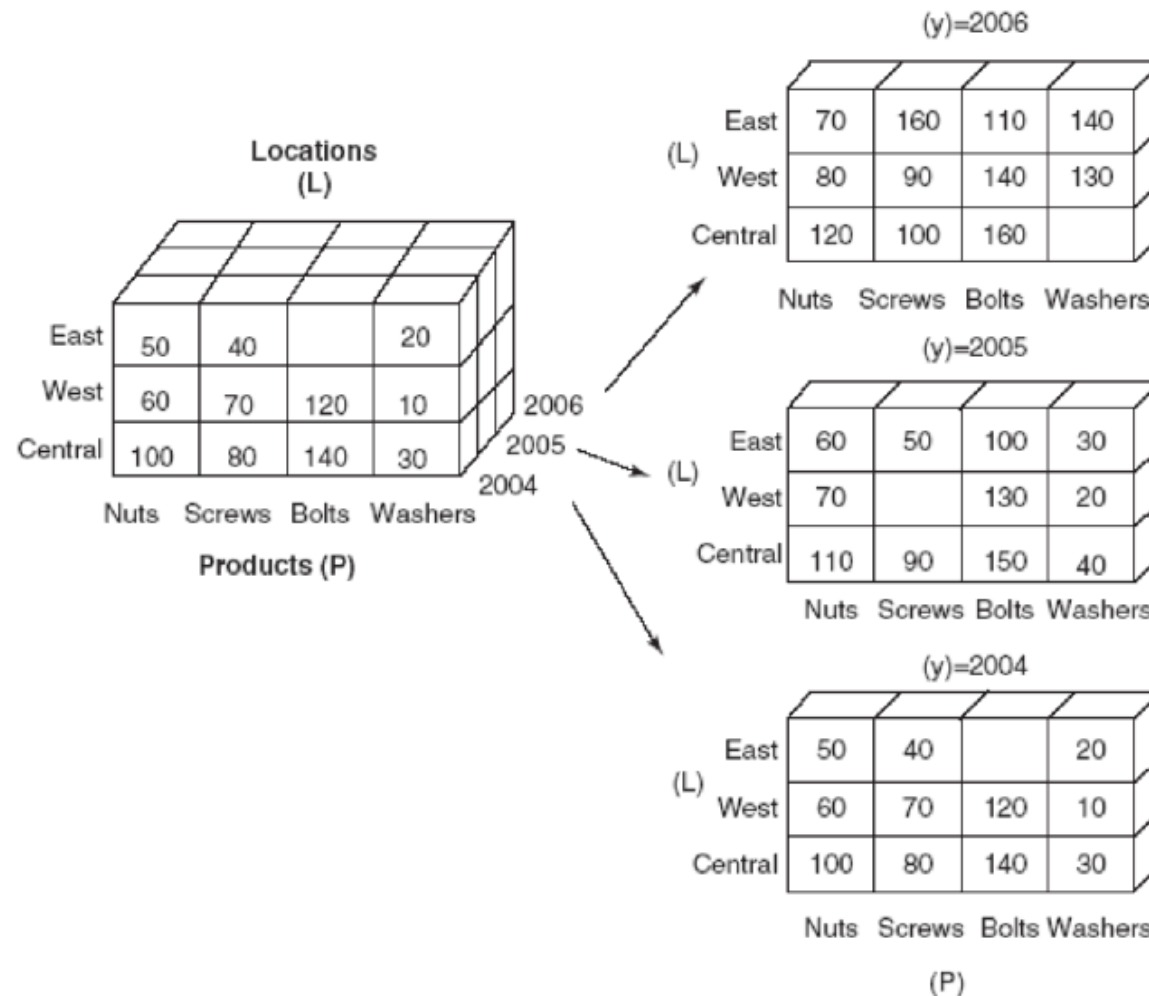
---

- Slice – memotong suatu dimensi pada poin nilai dimensi tertentu
- Dice – memutar sudut pandang melalui dimensi yang berbeda
- Drill Down/Roll Up – fitur untuk melakukan navigasi antar level data antara data paling ringkas (up) hingga data paling detail (down). Dengan drill down, kita dapat melakukan investigasi dari suatu informasi secara detail (contoh: tidak hanya menemukan total penjualan, tetapi juga penjualan berdasarkan region, produk, atau salesperson).

# Slice



# Slice

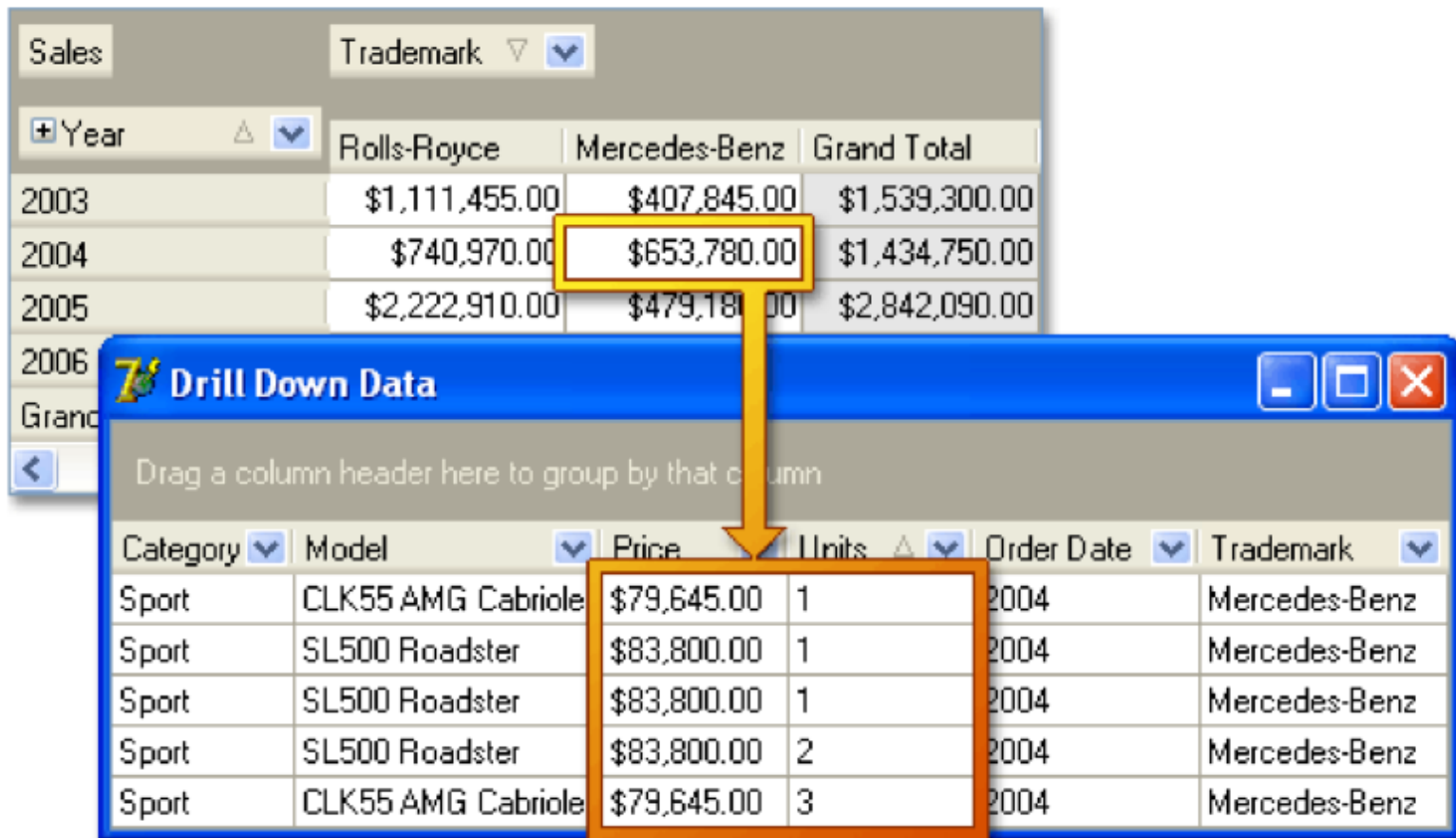


# Dice





# Drill Down – Roll Up



The screenshot illustrates a data drill-down process. The main table shows sales data by year and trademark. A specific data point is highlighted, and a drill-down window provides a detailed view of that data.

**Main Table Data:**

Year	Rolls-Royce	Mercedes-Benz	Grand Total
2003	\$1,111,455.00	\$407,845.00	\$1,539,300.00
2004	\$740,970.00	\$653,780.00	\$1,434,750.00
2005	\$2,222,910.00	\$479,180.00	\$2,842,090.00
2006			
Grand Total			

**Drill Down Data Window:**

Drag a column header here to group by that column

Category	Model	Price	Units	Order Date	Trademark
Sport	CLK55 AMG Cabriolet	\$79,645.00	1	2004	Mercedes-Benz
Sport	SL500 Roadster	\$83,800.00	1	2004	Mercedes-Benz
Sport	SL500 Roadster	\$83,800.00	1	2004	Mercedes-Benz
Sport	SL500 Roadster	\$83,800.00	2	2004	Mercedes-Benz
Sport	CLK55 AMG Cabriolet	\$79,645.00	3	2004	Mercedes-Benz



# Mendesign Sebuah Data Warehouse

---

Yudi Agusta, PhD

Kecerdasan Bisnis, Lecture 04 05 06 07

Copyright © Yudi Agusta, PhD 2024

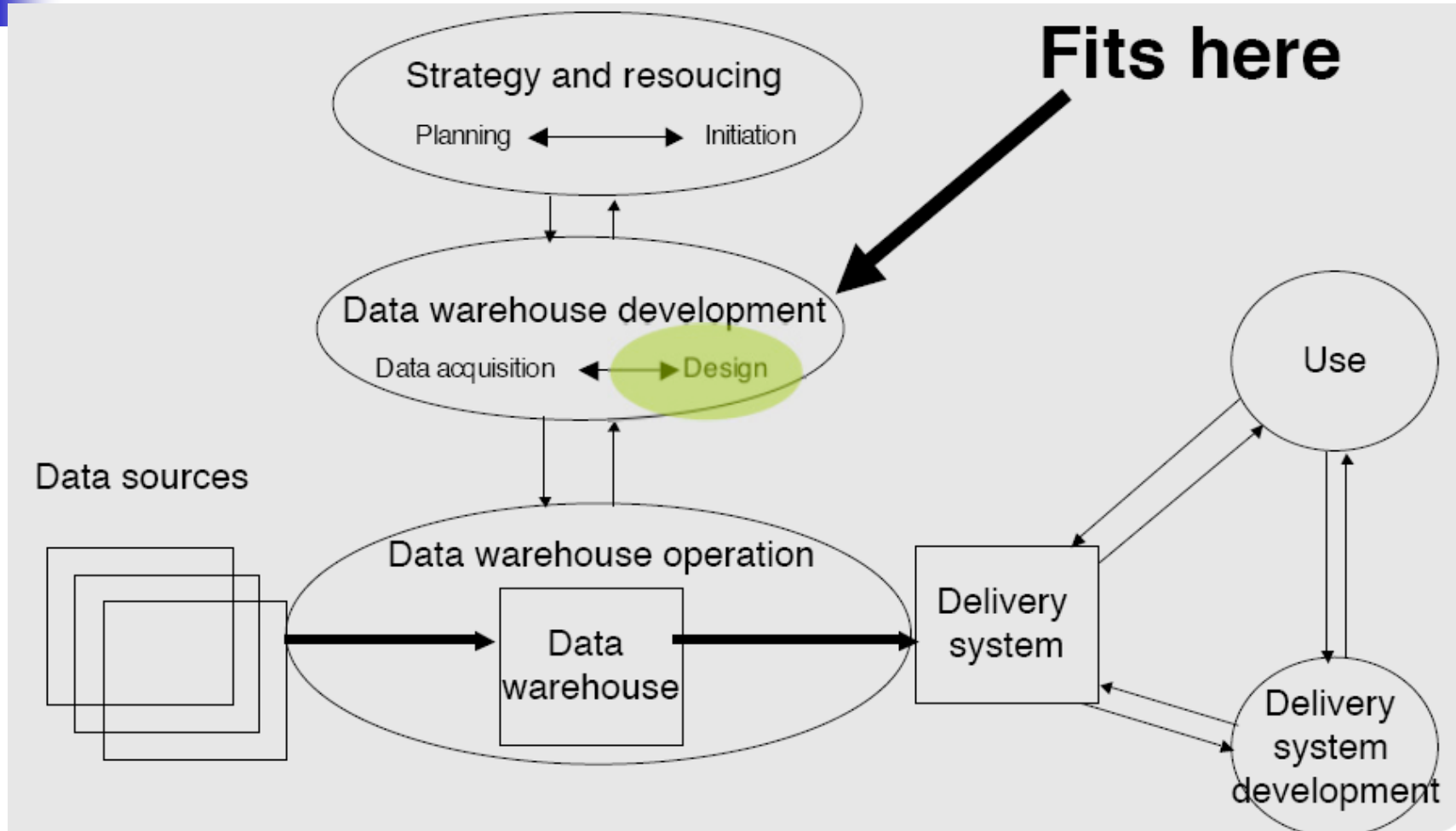


# Mendisain Sebuah Data Warehouse

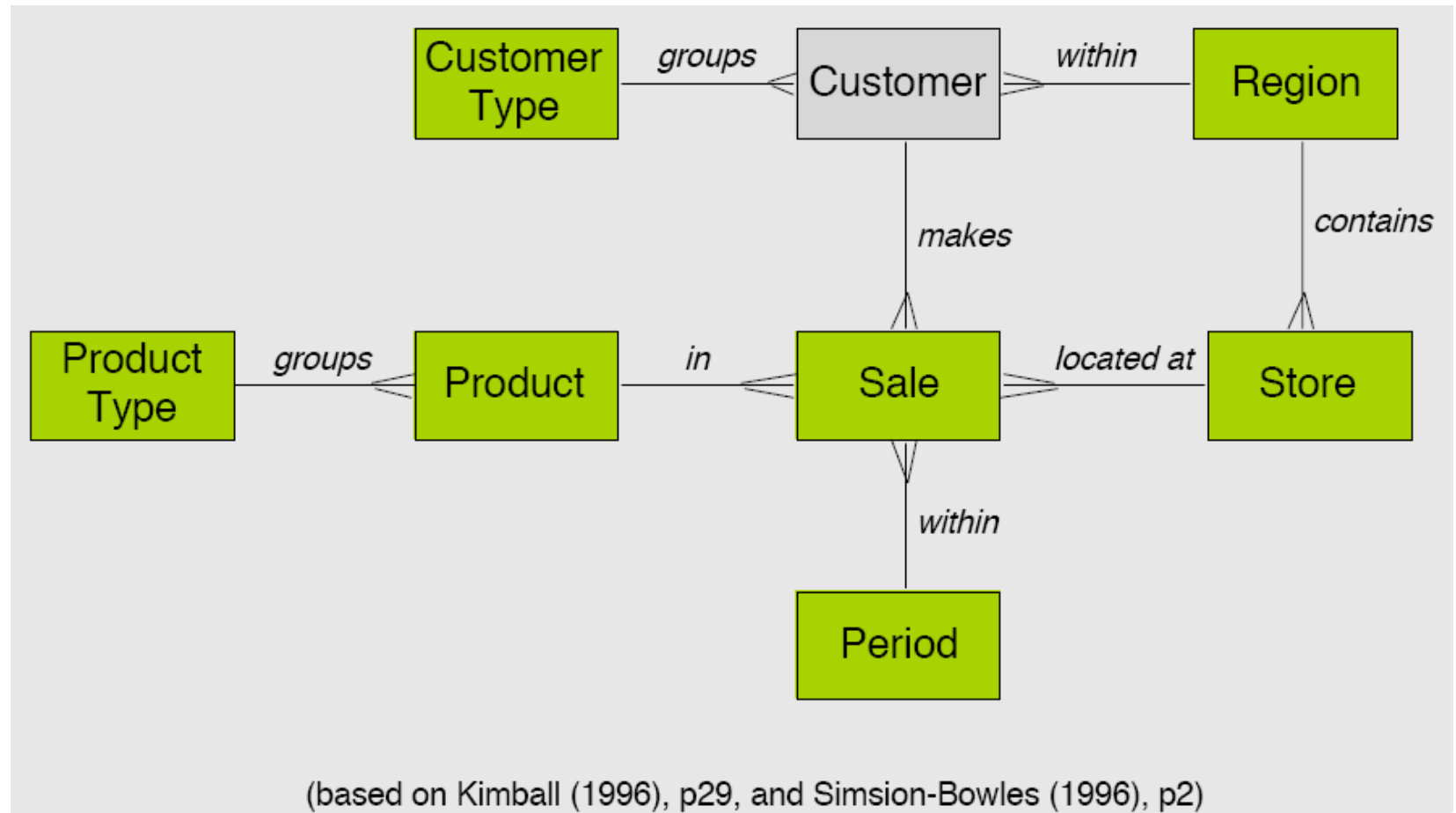
---

- Mendisain database untuk data warehouse adalah kegiatan utama dalam mendisain data warehouse
- Ada dua pendekatan utama dalam perancangan data warehouse
  - Pemodelan dan normalisasi entity relationship diagram (ERD)
  - Pemodelan berdimensi

# Schema Desain Untuk Database



# Schema Entity Relationship Diagram





# Desain Database - ERD Tradisional

---

- Entities and Relationships
- Aturan Normalisasi
  - Umumnya 3NF
  - Menjaga integritas database dengan menghindari anomalies
  - Setiap poin logical direpresentasikan hanya sekali (tidak ada redundancy)
- Pemikiran yang berbeda antara logical dan physical



# Contoh Normalisasi

Sebuah perusahaan manufaktur membuat produk dari beberapa komponen. Setiap produk mempunyai suatu nomor produk yang tersendiri, nama dan waktu perakitan. Semua komponen mempunyai nomor komponen tersendiri, diskripsi, kode supplier dan harga.

Product Code	Name	Time	Component Code	Description	Qty	Supplier	Cost
325	Trolley	0.35	B1378	Wheel	6	S2341	0.22
325	Trolley	0.35	B4532	Chassis	1	S2315	1.23
325	Trolley	0.35	B7649	Handle	1	S8567	0.97
325	Trolley	0.35	B6467	Axle	3	S8567	0.32
563	Cart	0.27	B1378	Wheel	4	S2341	0.22
563	Cart	0.27	B6467	Axle	2	S8567	0.32
563	Cart	0.27	B6468	Frame	1	S3428	0.37



# Database Tradisional (Bukan RDBMS)

---

- Tabel-tabel dalam jumlah yang besar
  - Oracle Financials 1.800, SAP 7 sampai 8.000
- Digunakan secara umum
  - Alamiah dan mudah dimengerti oleh umum
- Penelitian menunjukkan bahwa database tradisional tidak mudah dimengerti oleh profesional IT
  - Khususnya konsep-konsep seperti abstraction, generalisation, sub-types dll





# Permasalahan

---

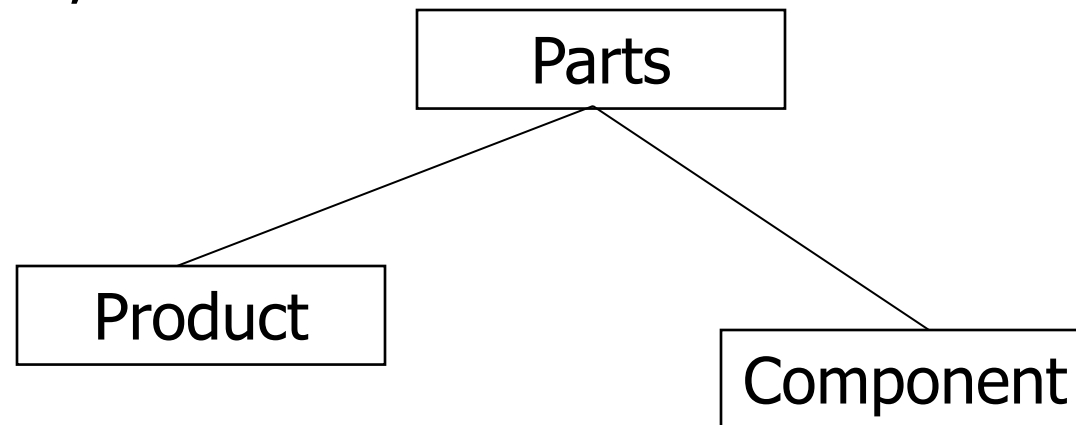
- Sejumlah field yang bukan merupakan atribut kunci sangat tergantung pada primary key
- Ketergantungan berarti bahwa kita hanya bisa mengerjakan satu atribut kalau yang lain sudah dimasukkan
  - ProductCode → Name
  - ProductCode → Time
  - ComponentCode → Deskripsi
  - ComponentCode → Supplier
  - ComponentCode → Cost
  - ComponentCode, Product Code → Quantity



## Database Yang Sudah Dinormalisasikan

---

- Product: ProductCode, Name, Time
- Parts: ProductCode, ComponentCode, Qty
- Component: ComponentCode, Description, Supplier, Cost





# Isi Database Ternormalisasi

*Product*

<b>ProductCode</b>	<b>Name</b>	<b>Time</b>
325	Trolley	0.35
563	Cart	0.27

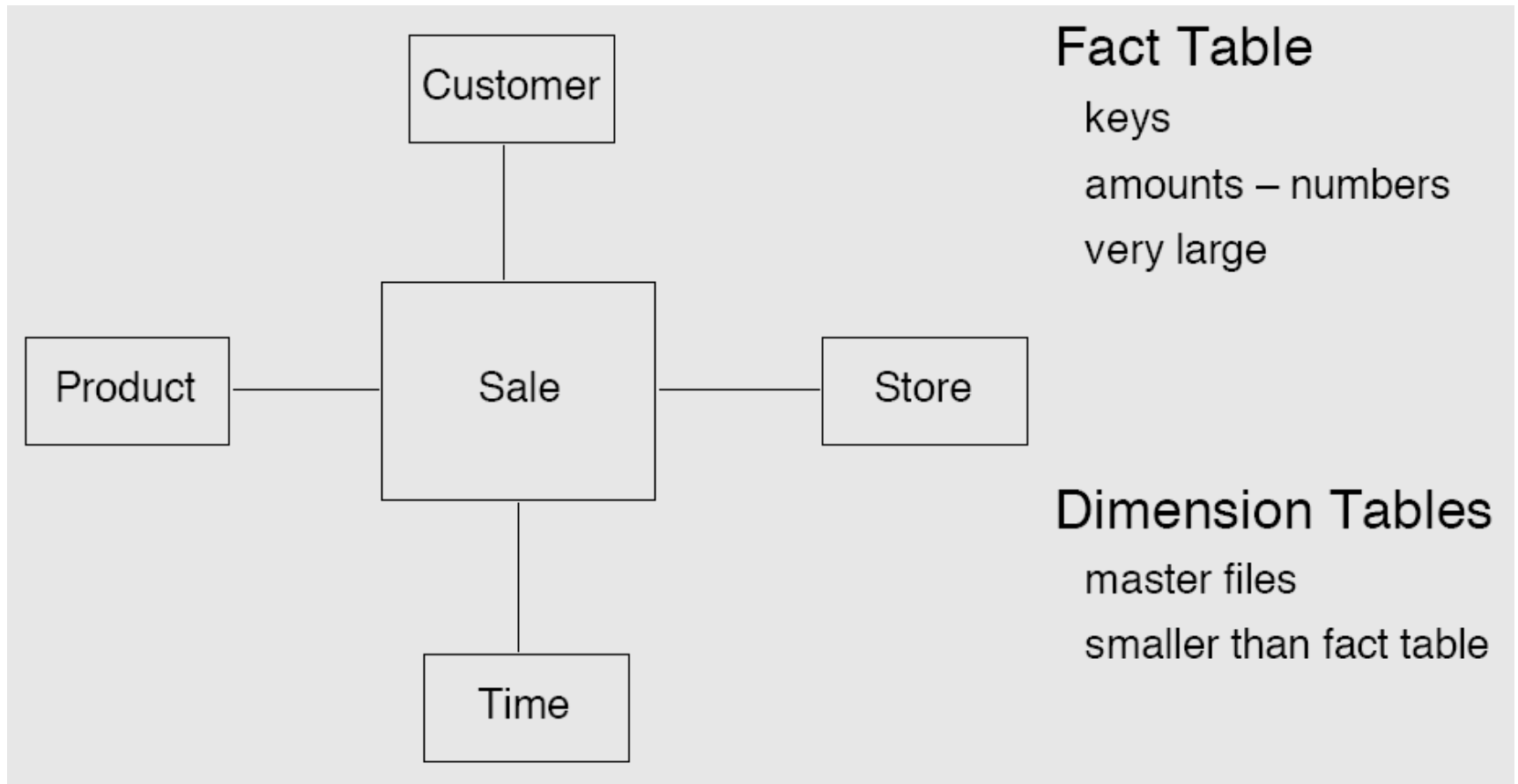
*Parts*

<b>ProductCode</b>	<b>ComponentCode</b>	<b>Qty</b>
325	B1378	6
325	B4532	1
325	B7649	1
325	B6467	3
563	B1378	4
563	B6467	2
563	B6468	1

*Component*

<b>ComponentCode</b>	<b>Description</b>	<b>Supplier</b>	<b>Cost</b>
B1378	Wheel	S2341	0.22
B4532	Chassis	S2315	1.23
B7649	Handle	S8567	0.97
B6467	Axle	S8567	0.32
B6468	Frame	S3428	0.37

# Model Berdimensi





# Tabel Fakta dan Tabel Dimensi

---

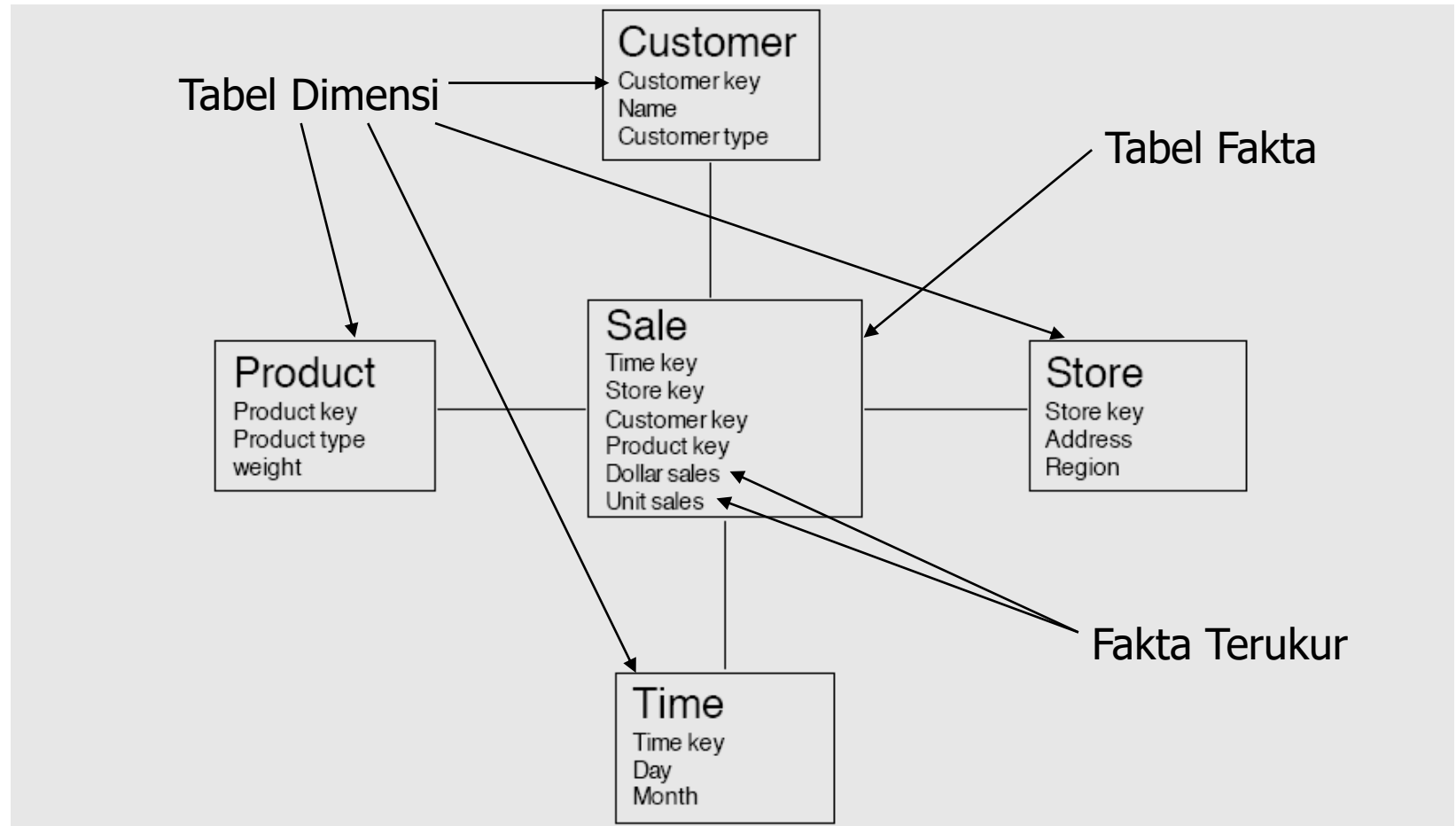
- Tabel Fakta

- Merupakan suatu tabel yang berisikan kode-kode dimensi yang berasal dari tabel dimensi, dan
- Fakta-fakta yang ingin ditampilkan sebagai output dari sistem data warehouse untuk menjawab pertanyaan yang diajukan manajemen

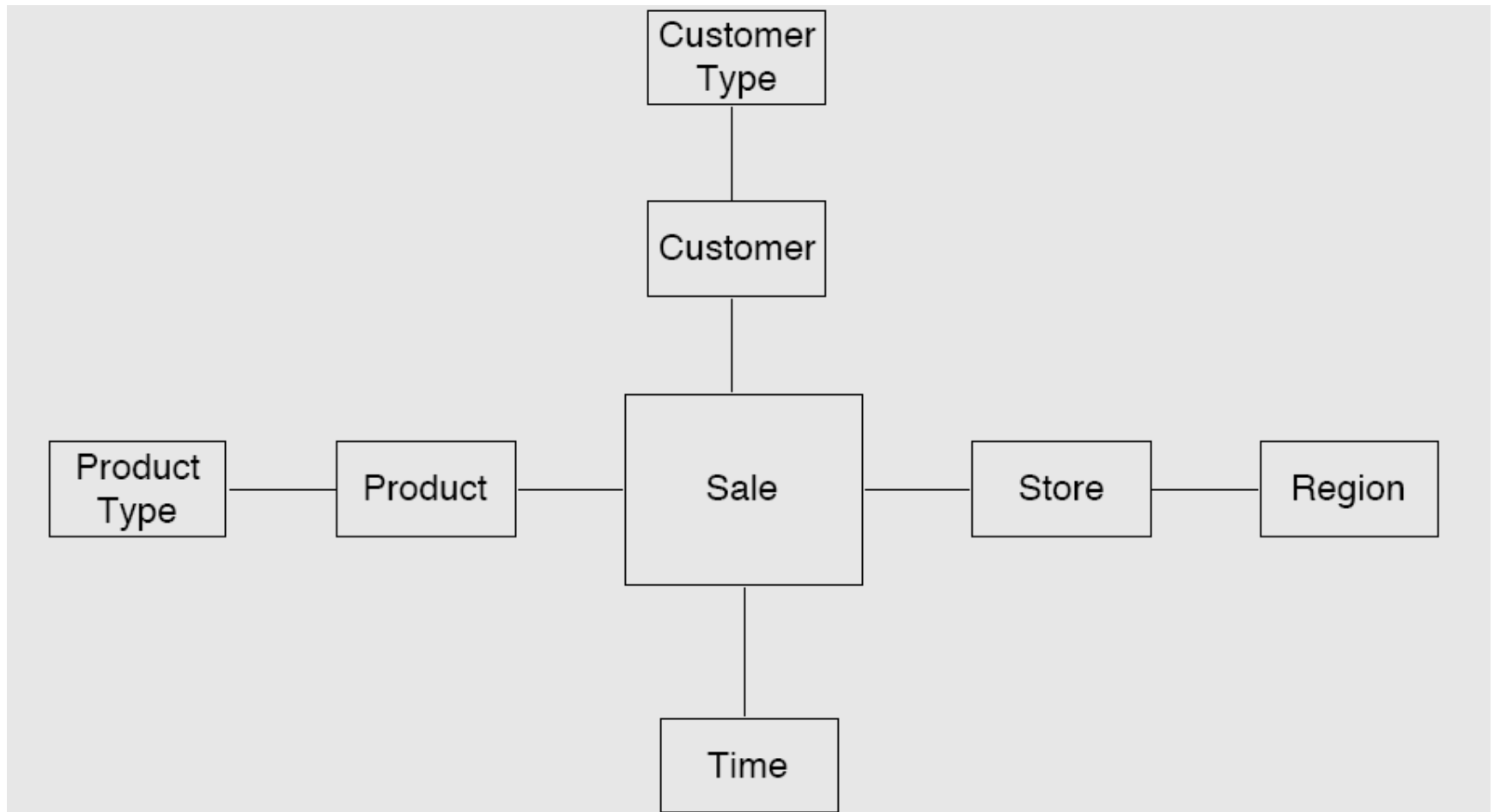
- Tabel Dimensi

- Merupakan tabel-tabel yang berisikan data-data berdimensi, dari dimensi tertinggi ke dimensi terendah, seperti:
  - Dimensi Waktu: Tahun, Bulan, Tanggal, Hari, Libur/Biasa, Akhir Minggu/Hari Biasa, Hari Raya/Hari Biasa
  - Dimensi Tempat: Negara, Provinsi, Kabupaten, Kecamatan, Desa/Kota
- Kode dari tabel dimensi dihubungkan dengan satu atau lebih tabel fakta

# Star Schema (Dengan Atribut)



# Snowflake Schema





# Dimensi

---

- Star schema mungkin (umumnya) mempunyai 10-15 dimensi
- Tampilan yang muncul pada pengguna umumnya mencakup database dengan 6-7 dimensi
- Sistem umum (seperti EIS) mempunyai 20 tampilan yang berbeda-beda dengan 4-5 tabel fakta yang mempunyai dasar penyusunan dan dimensi yang berbeda-beda
- Table dimensi dapat mempunyai banyak dimensi di dalamnya





# Proses Tambahan Dalam Data Warehouse

---

Yudi Agusta, PhD

Kecerdasan Bisnis, Lecture 04 05 06 07

Copyright © Yudi Agusta, PhD 2024



## Dimensi Yang Berubah Secara Perlahan

---

- Karena adanya kebutuhan dukungan data yang berubah-ubah, banyak dimensi (seperti Produk dan Pelanggan) berkembang secara perlahan dalam jangka waktu yang panjang
  - Pelanggan mengubah alamat, preference, dll
  - Tim penjualan mengubah wilayah penjualan dll
- Tiga standar pendekatan yang dilakukan untuk mengakomodasi:
  - Tanpa nilai yang lama
  - Buat record dimensi tambahan
  - Buat field dengan nilai saat ini



# Tipe 1: Timpa

---

- Timpa nilai atribut
  - Contoh: Lokasi retail penjualan yang berubah “Gianyar” menjadi “Klungkung”
- Keuntungan: mudah untuk diimplementasikan
- Kekurangan: tidak ada catatan perubahan



## Tipe 2: Record Baru

---

- Membuat satu tabel dimensi baru untuk setiap versi
  - Perlu untuk menggeneralisasikan kunci dimensi (menambahkan 2 atau 3 digit untuk merekam versi perkembangan tabel dimensi)
- Keuntungan: Secara otomatis memelihara dan membagi catatan
- Kerugian: Lebih kompleks dari proses Timpa



## Tipe 3: Membuat Field dengan Nilai Saat Ini

---

- Buat suatu field yang disebut dengan “Current X”
  - Current Location
  - Old Location
- Berguna bila kita ingin tahu nilai yang lama dan nilai yang baru
  - Contoh: Penyesuaian tim penjualan
- Keuntungan: Sederhana dan Cepat, memungkinkan untuk melakukan perbandingan
- Kerugian: Perubahan yang lainnya yang mengalami perubahan dengan cepat seperti perubahan waktu, pelanggan, dan produk akan sulit diakomodasi



# Memilih Metode

---

- Tipe 1
  - Sangat sederhana. Hanya digunakan bila nilai histori tidak begitu penting
- Tipe 2
  - Secara umum, sangat fleksibel. Kimball mengatakan, ini berguna untuk dimensi sampai ratusan dengan record berjumlah ribuan, khususnya disain dengan DBMS yang bagus
- Tipe 3
  - Tidak sefleksibel Tipe 2, tetapi dapat digunakan untuk dimensi dengan record yang berjumlah jutaan



# Agregat

---


- Sebagian besar data warehouse mempunyai record pada tabel fakta dalam jumlah yang sangat besar (sampai 50 triliun record dan memerlukan media penyimpan sampai 1 – 5 terabytes)
- Agregat (summary sebelum disimpan) adalah cara yang paling efektif untuk meningkatkan performance dari data warehouse
- Agregat record dari tabel fakta yang merepresentasikan summary dari record tabel fakta pada level dasar
- Fakta agregat terletak pada tabel fakta agregat baru atau tabel fakta original
  - Yang mana yang terbaik?

# Complete Data Warehouse

Sales Fact

Time key
Product key
Store key
Promotion key
Dollar sales
Unit sales
Dollar costs
Customer count

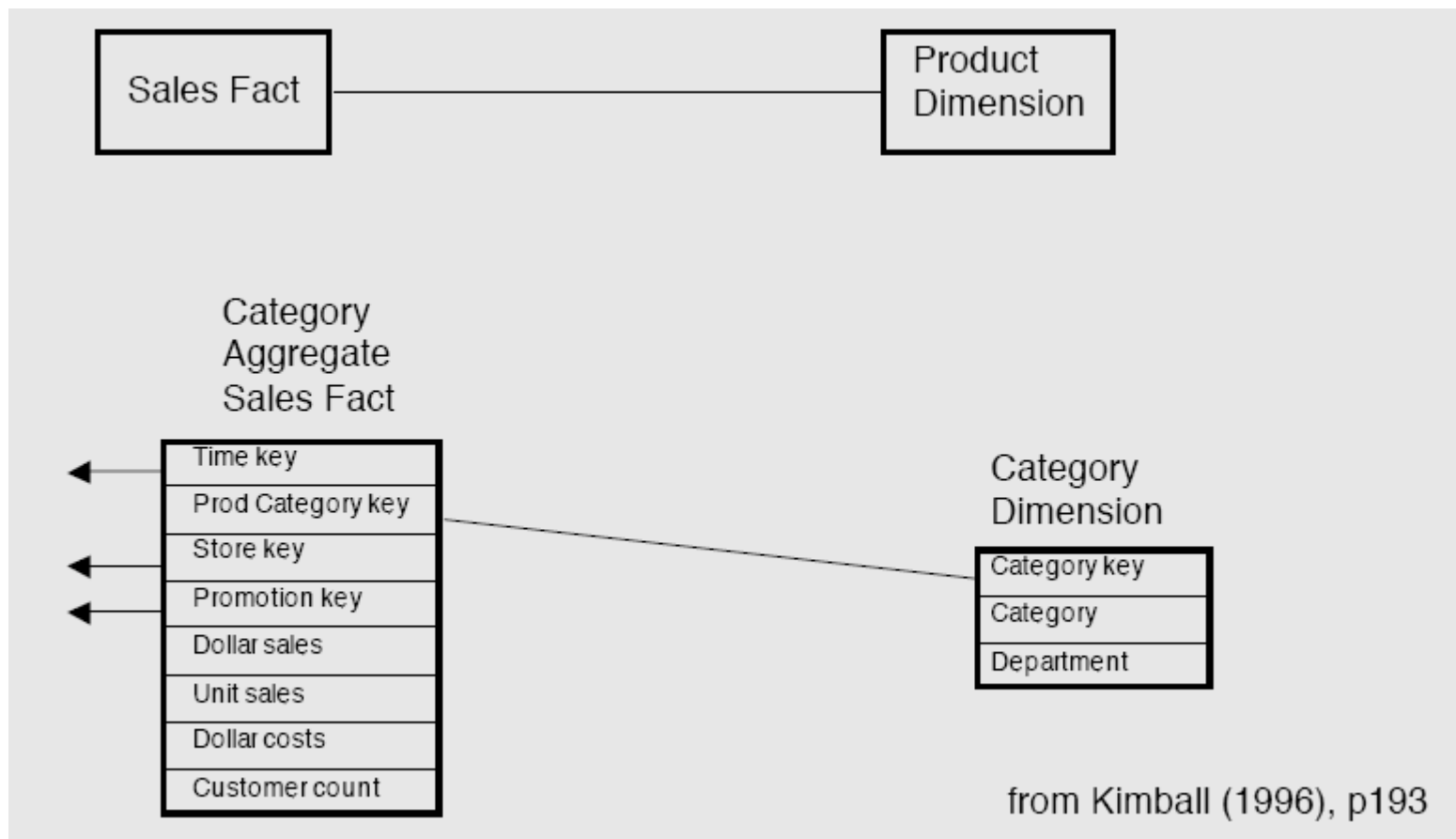
Augmented  
Product  
Dimension

Product key
LEVEL
SKU description
SKU number
package size


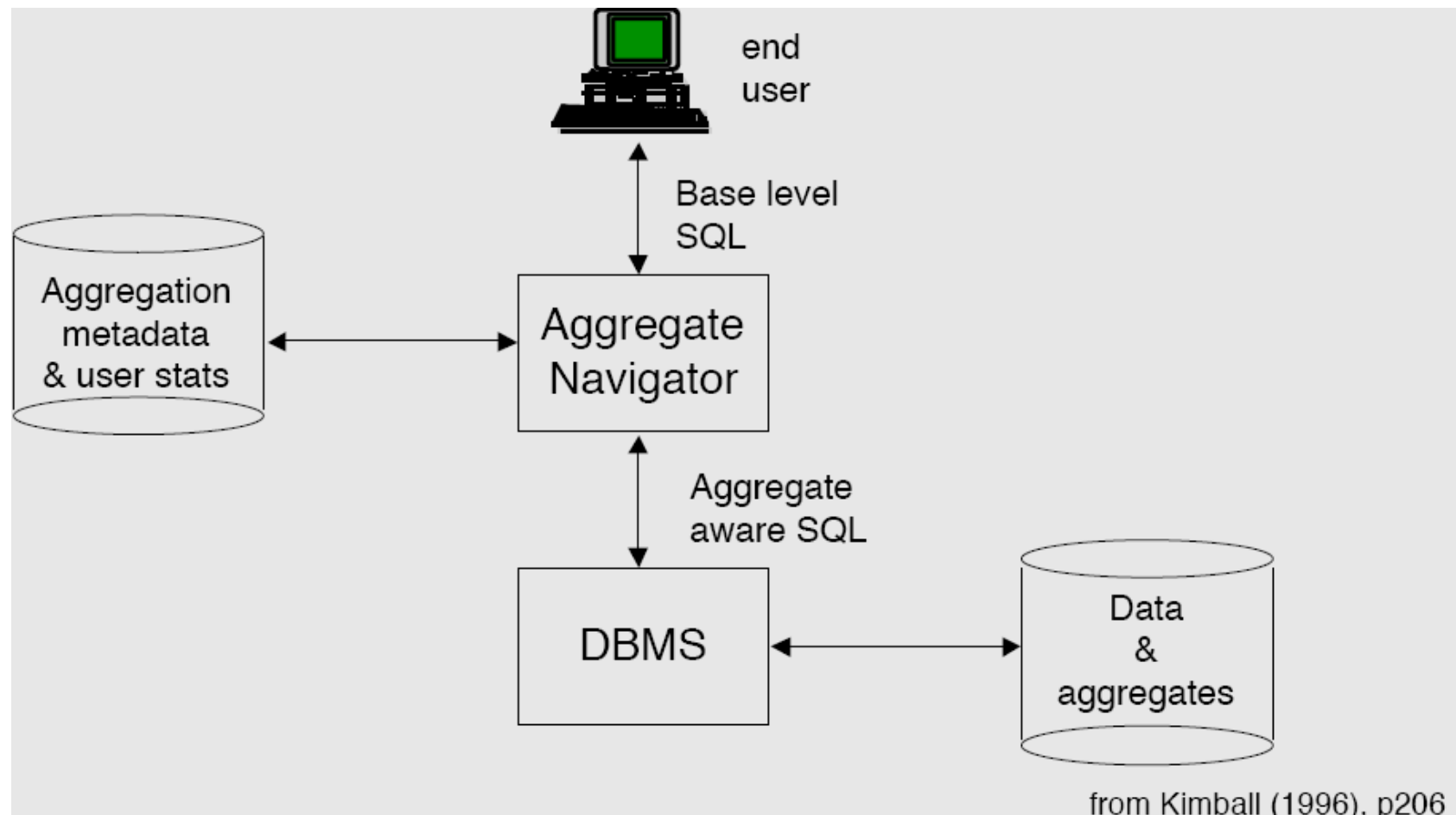
nb. "category dimension records only have values for category and department fields, otherwise null, "NA", or "Total" values



# Tabel Fakta Agregat



# Pemandu Agregat





# Pemandu Agregat

---

- Pemandu agregat secara otomatis mentransformasikan SQL berbasis pengguna ke SQL yang memperhatikan agregat
- Pemandu agregat secara dinamis memilih tabel agregat terbaik untuk digunakan
- Pemandu agregat mengisolasi pengguna dari fakta detail dan memungkinkan DBA untuk melakukan adjustment terhadap agregat



# Data Staging Area

---

- Data Staging merupakan bagian besar dari pengembangan Data Warehouse
- Data Staging merupakan suatu proses untuk mempersiapkan dan memproses data dari berbagai sumber untuk diupload ke dalam data warehouse
- Perencanaan yang matang diperlukan, agar semua input, proses, dan output yang diikutkan sesuai dengan kebutuhan yang ada
- Pengembang perlu memperhatikan proses untuk penyiapan data dimensi dan data fakta dan melakukan test yang mencukupi



# Data Staging Proses

---

- Perencanaan Level Atas
- Penetapan Data Staging Tools
- Menyusun Perencanaan Detail
- Mengisi Tabel Dimensi Sederhana
- Menerapkan Logic Perubahan Dimensi
- Mengisi dimensi yang lainnya
- Menguplod fakta history pada level satuan
- Staging tabel fakta secara bertahap
- Mengisi table aggregate
- Pengoperasian dan Otomatisasi Warehouse



# Data Staging Tools

---

- Tools yang digunakan harus dapat mengakomodasi berbagai data sumber dan dapat mempersiapkan data sesuai kondisi tujuan di data warehouse
- Data Input dapat berupa DBMS, Spreadsheet, ataupun Text
- Pemrosesan data bisa sangat sederhana dalam bentuk copy paste, ataupun yang kompleks dengan berbagai proses perhitungan atau transformasi yang panjang



# Meta Data

---

- Database-nya Database
- Meta Data berfungsi untuk memberikan kemudahan dalam mencari asal data dan untuk keperluan dokumentasi dari data yang ada di dalam tabel-tabel database dari sistem data warehouse
- Meta Data memberikan informasi tentang:
  - Dari mana
  - Proses apa yang dilakukan dalam mendapatkan data yang ada di data warehouse
  - Dan informasi lainnya (masa update, penanggung jawab data, satuan, dll)
- Tabel meta data dapat terdiri dari field:
  - Nama Database, Nama Tabel, Nama Field, Asal Data, Proses Transformasi
- Meta Data dapat juga mencakup informasi terkait pengaturan tampilan seperti tanpa decimal, dengan decimal 2, rata kiri/kanan/tengah, dll



# Studi Kasus Design Data Warehouse Usaha Retail

---

Yudi Agusta, PhD

Kecerdasan Bisnis, Lecture 04 05 06 07

Copyright © Yudi Agusta, PhD 2024





# Studi Kasus Usaha Retail

---

- Latar Belakang
  - Perusahaan grocery besar dengan perkiraan 500 outlet
  - Setiap outlet mempunyai sekitar 60000 produk dalam tampilannya
  - Teknologi yang digunakan
    - SKU – Stock Keeping Unit
    - UPC – Universal Product Code



# Studi Kasus Usaha Retail

---

- Problem Statement:
  - Perlu untuk memaksimalkan keuntungan dan tetap menjaga stok agar tetap ada
  - Perlu informasi untuk mendukung pengambilan keputusan dalam hal penetapan harga dan promosi. Tipe promosi yang dilaksanakan:
    - Discount harga sementara
    - Reklame surat kabar
    - Tampilan lemari dan lorong
    - Kupon



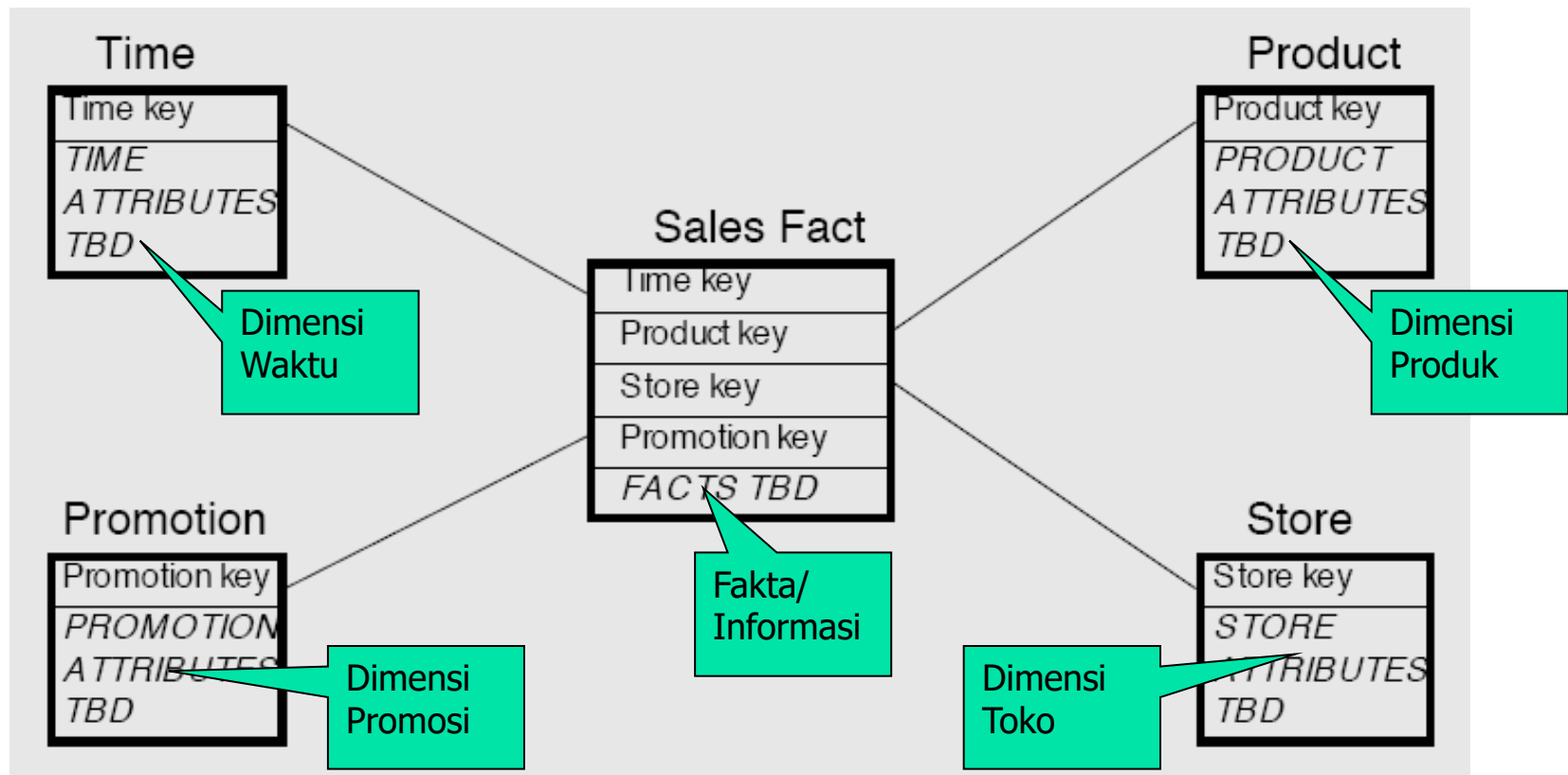
# Usaha Retail

---

- Proses yang perlu dianalisa
  - Memilih Proses Bisnis
    - Pergerakan barang harian yang mencakup transaksi penjualan dan pembelian barang
  - Memilih fakta/informasi untuk menjawab permasalahan
    - **Unit Terjual** by Toko by Promosi by Hari
  - Memilih variable dimensi
    - Waktu, Produk, Toko dan Promosi

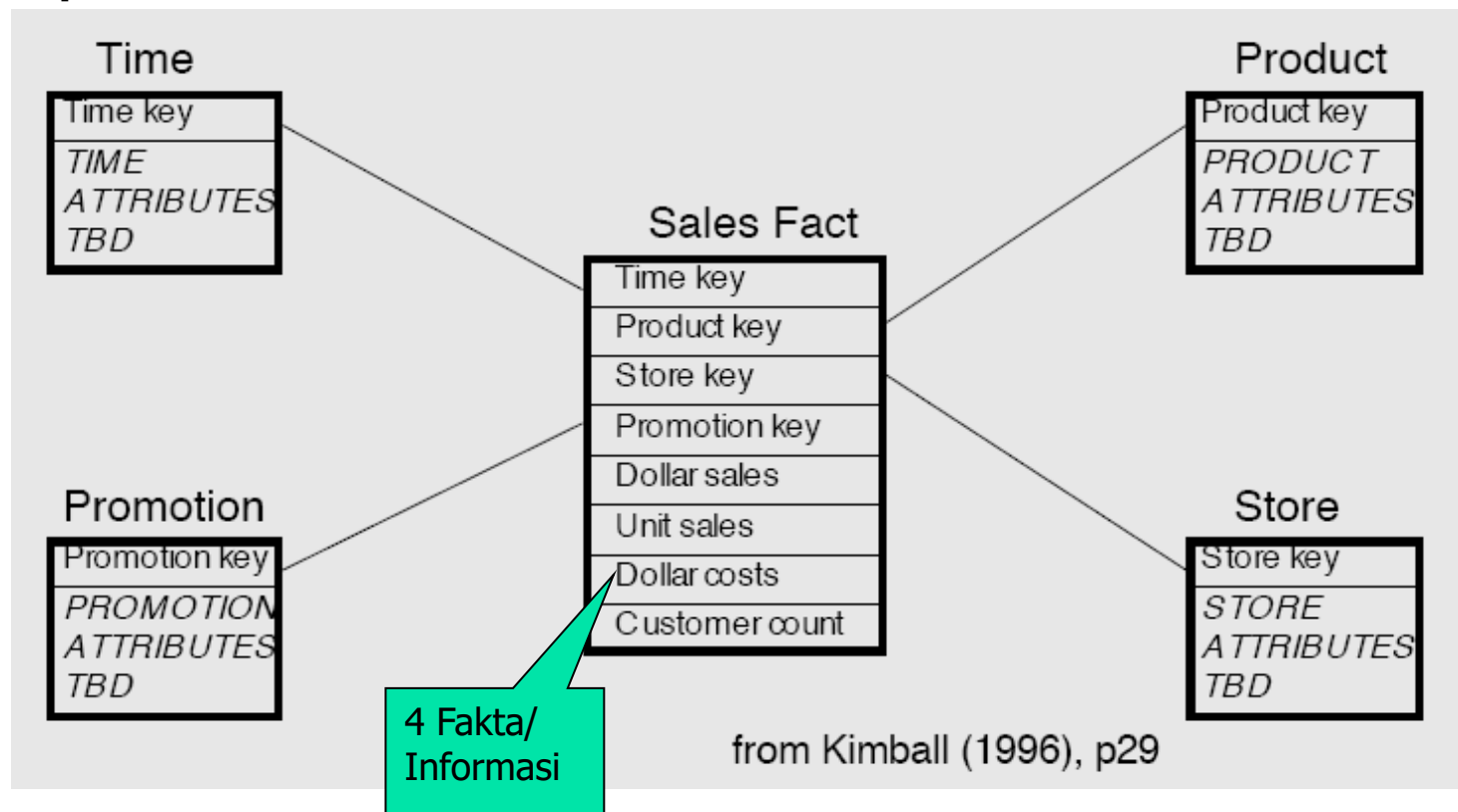
# Disain Data Warehouse Usaha Retail

- Fakta/Informasi Untuk Menjawab Permasalahan dan Varibel Dimensi



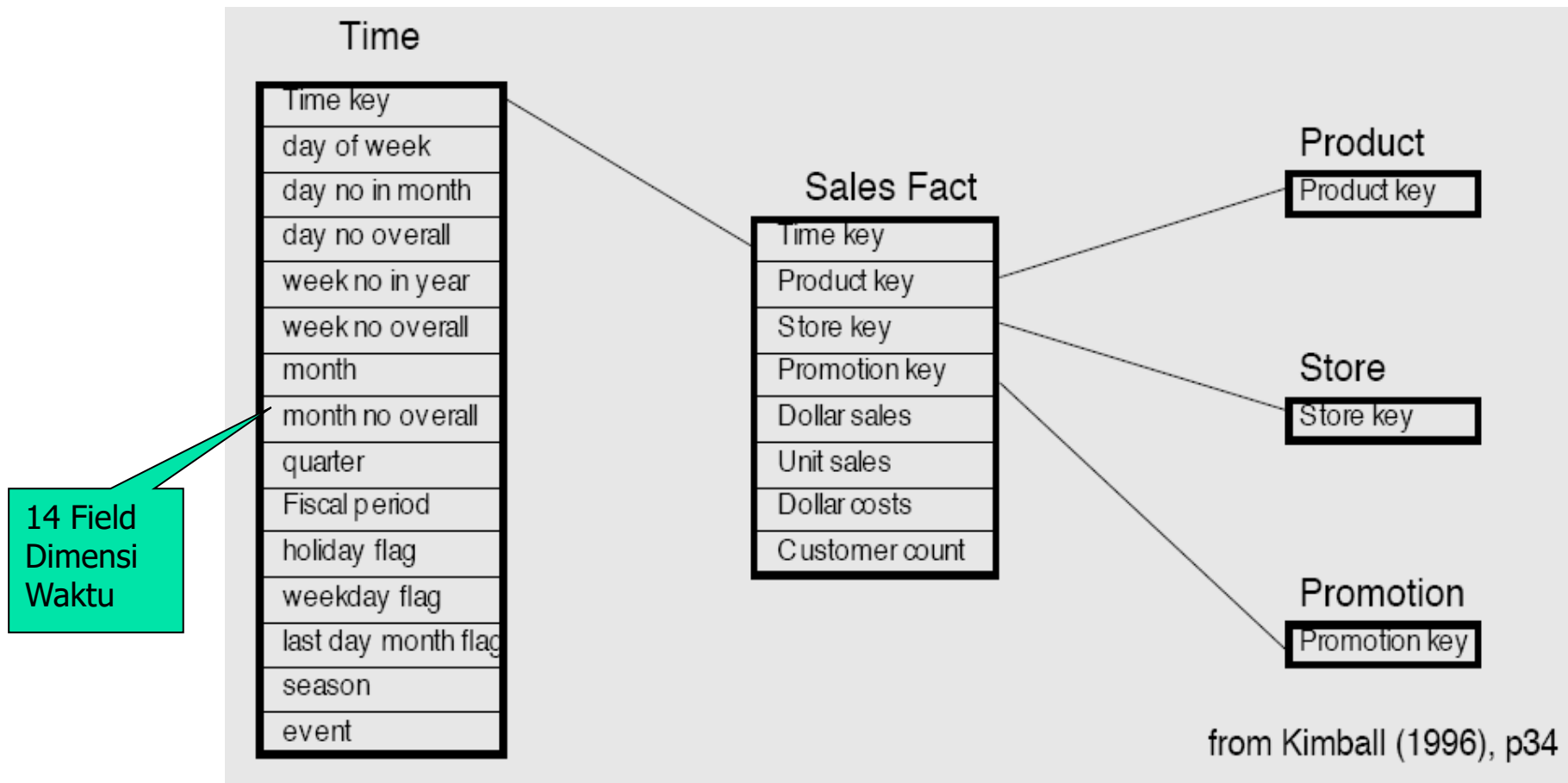
# Desain Data Warehouse Usaha Retail

- Fakta/informasi untuk menjawab permasalahan



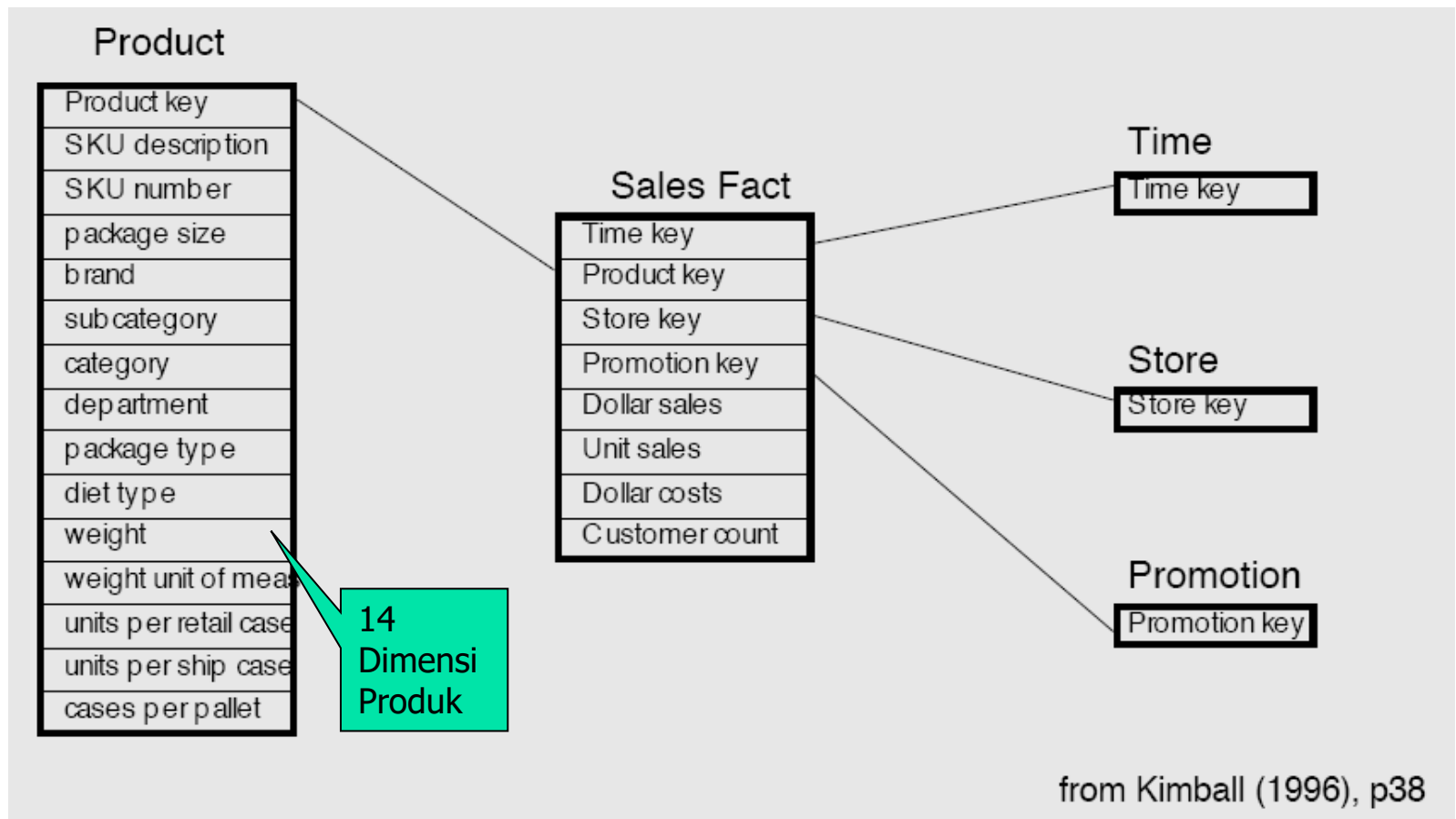
# Design Data Warehouse Usaha Retail

## ■ Dimensi Waktu



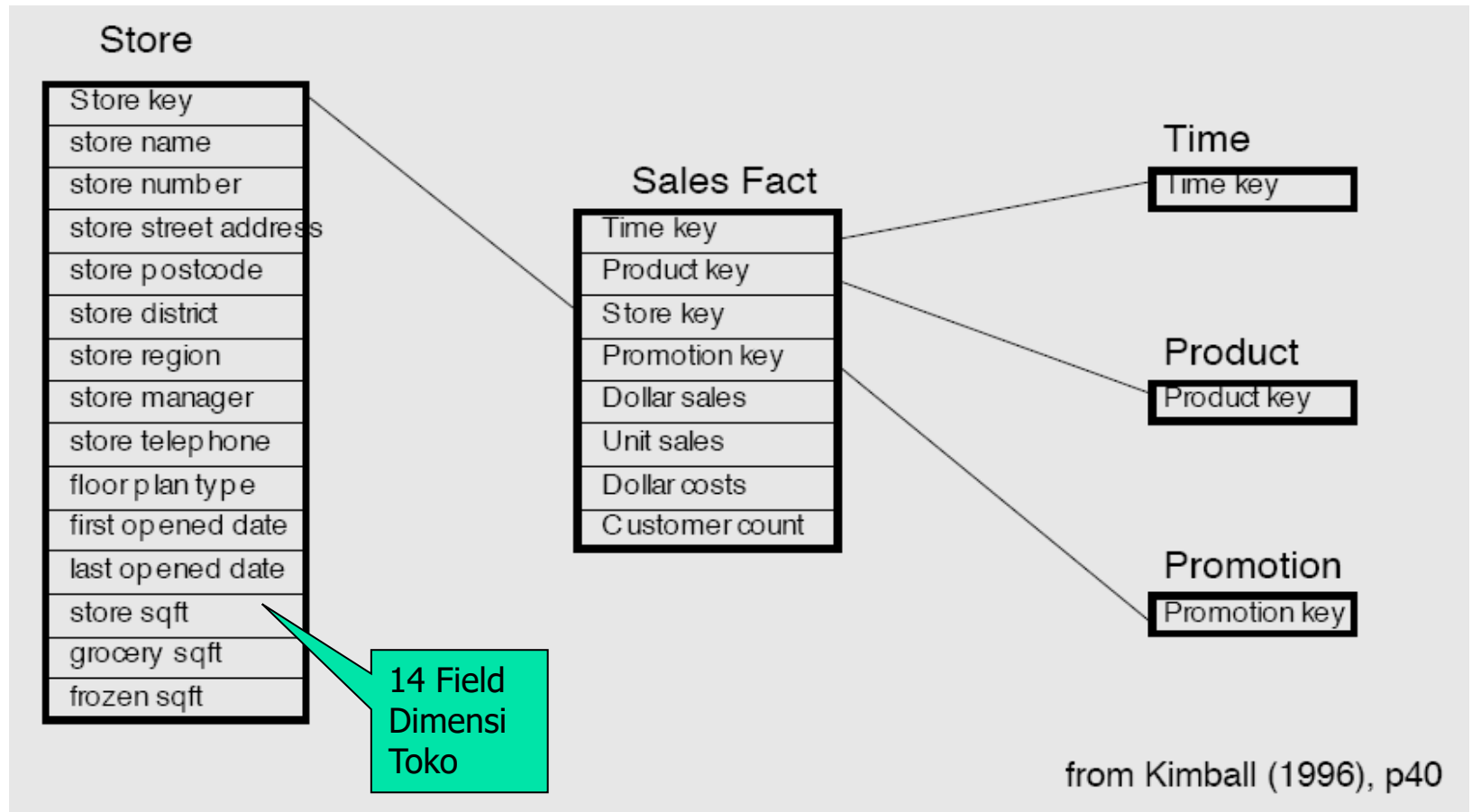
# Design Data Warehouse Usaha Retail

## ■ Dimensi Produk



# Design Data Warehouse Usaha Retail

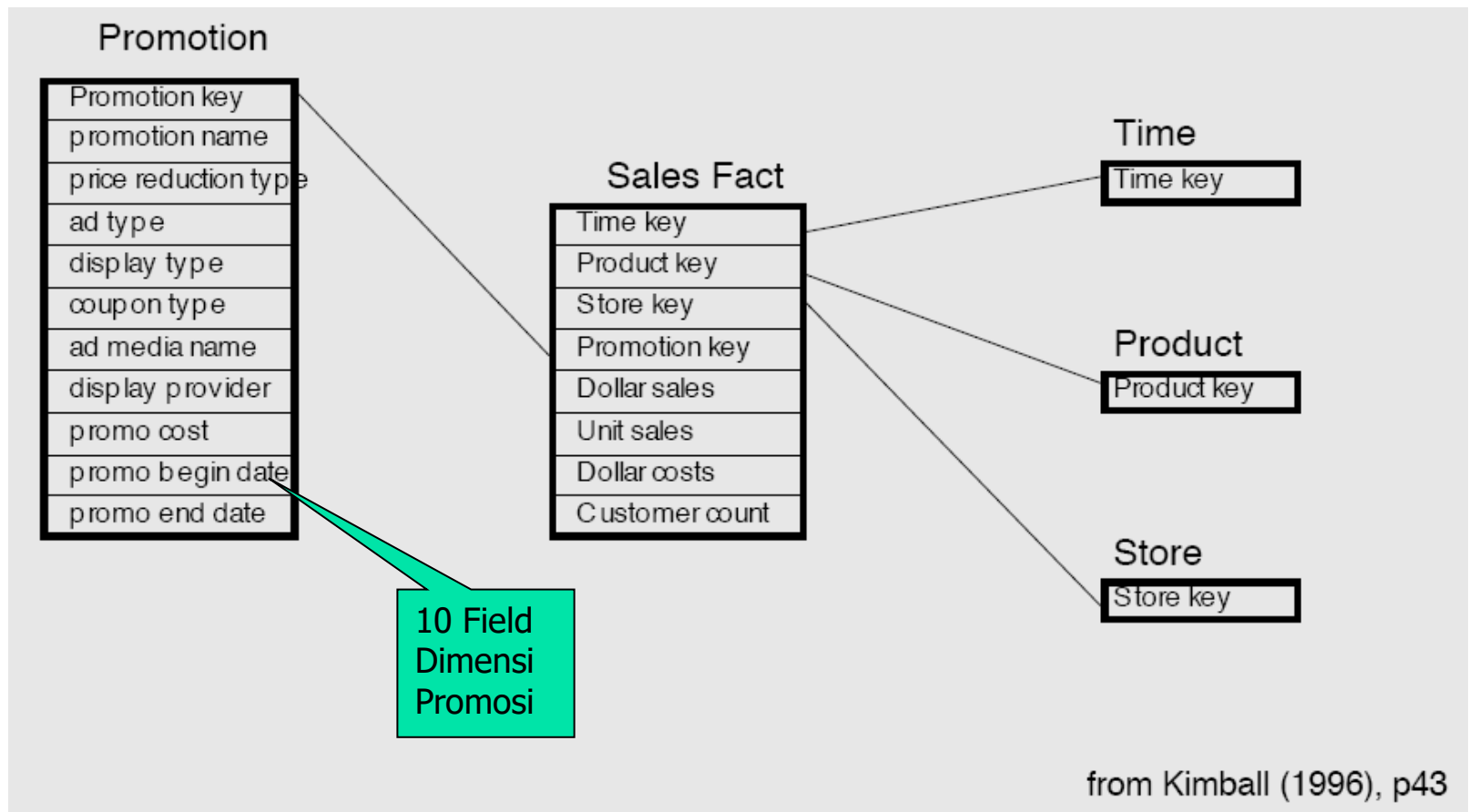
## ■ Dimensi Toko





# Usaha Retail: Dimensi Promosi

## ■ Dimensi Promosi





# Kondisi Database Data Warehouse

---

- Kombinasi dimensi mempunyai kecenderungan dalam jumlah besar
- Berefek pada fakta/informasi yang disimpan juga menjadi sangat besar dengan pengali sejumlah dimensi
- Dalam contoh di atas:
  - Dimensi Waktu sejumlah 14! Kombinasi
  - Dimensi Produk sejumlah 14! Kombinasi
  - Dimensi Toko sejumlah 14! Kombinasi
  - Dimensi Promosi sejumlah 10! Kombinasi
  - Fakta/Informasi sejumlah  $14! \times 14! \times 14! \times 10!$  Kombinasi



# Kompleksitas Diperlukan?

---

- Jawabannya: tergantung tujuan dan kebutuhan yang diungkapkan manajemen dalam pengambilan keputusan yang perlu dukungan Data Warehouse
- Pemilihan variable dimensi perlu untuk mendapatkan perhatian yang sangat serius, khususnya melihat pada kebutuhan pihak manajemen
- Batasan lainnya yang ada dalam pengembangan Data Warehouse:
  - Ketersediaan data dari OLTP/Sistem Informasi
  - Sistem OLTP/Sistem Informasi harus disediakan atau diredesign untuk bisa memenuhi kebutuhan data di Data Warehouse

# Disain Database Sistem Infomasi Usaha Retail

ID Hari Libur
Tanggal
Nama Hari Libur

ID Event
Tanggal
Nama Event

ID Toko
Store Name
Store Number
Store Street Address
Store Post Code
Store District
Store Region
Store Manager
Store Telephone
Floor Plan Type
First Opened Date
Last Opened Date
Store sqft
Grocery sqft
Frozen sqft

ID Customer
Name
Address
Phone Number
Email

ID Transaksi
Tanggal Transaksi
ID Toko
ID Customer
ID Produk
ID Promotion
Volume
Harga Satuan
Total Penjualan
Biaya Transaksi
Biaya Pengiriman

ID Promotion
Promotion Name
Price Reduction Type
Ad Type
Display Type
Coupon Type
Ad Media Type
Display Provider
Promo Cost
Promo Begin Date
Promo End Date

ID Produk
Produk Name
SKU Description
SKU Number
Size
Brand
Sub Category
Category
Department
Package Type
Diet Type
Weight
Weight Unit of Measurement
Units per Retail Case
Units per Ship Case
Cases per Pallet



# Analisa Lanjutan Atas Design

---

- Ada beberapa dimensi yang terlalu detail sehingga pemanfaatannya untuk pengambilan keputusan diperkirakan tidak begitu banyak:
  - SKU Description, SKU Number, Weight Unit of Measurement
  - Store Name, Store Number, Store Street Address, Store Post Code, Store Manager, Store Telephone, First Opened Date, Last Opened Date
  - Promo Begin Date, Promo End Date
- **Transformasi data** dari data di Sistem Informasi menjadi data di Data Warehouse yang perlu dianalisa:
  - Perhitungan Dollar Sales dari data di Tabel Transaksi dan Tabel Dimensi
  - Perhitungan Unit Sales dari Tabel Transaksi dan Tabel Dimensi
  - Perhitungan Dollar Costs dari Tabel Transaksi dan Tabel Dimensi
  - Perhitungan Customer Count dari Tabel Transaksi dan Tabel Dimensi



# Extraction Transformation Loading (ETL)

---

Yudi Agusta, PhD

Kecerdasan Bisnis, Lecture 04 05 06 07

Copyright © Yudi Agusta, PhD 2024



# Contoh Transformasi

---

- <Tanggal> -> Hari
- <Tanggal> -> Hari Raya/Tidak
- <Alamat> -> Kawasan
- <Alamat> -> Regional
- <Kode Produk> -> Barang Sekali Pakai/Tahan Lama
- <Data Jumlah Penjualan> ->
  - Dollar Sales
  - Unit Sales
  - Dollar Costs
  - Customers Count

# Contoh Database Sistem Infomasi Sumber

ID Hari Libur
Tanggal
Nama Hari Libur

ID Event
Tanggal
Nama Event

ID Toko
Store Name
Store Number
Store Street Address
Store Post Code
Store District
Store Region
Store Manager
Store Telephone
Floor Plan Type
First Opened Date
Last Opened Date
Store sqft
Grocery sqft
Frozen sqft

ID Customer
Name
Address
Phone Number
Email

ID Transaksi
Tanggal Transaksi
ID Toko
ID Customer
ID Produk
ID Promotion
Volume
Harga Satuan
Total Penjualan
Biaya Transaksi
Biaya Pengiriman

Data  
Tanggal

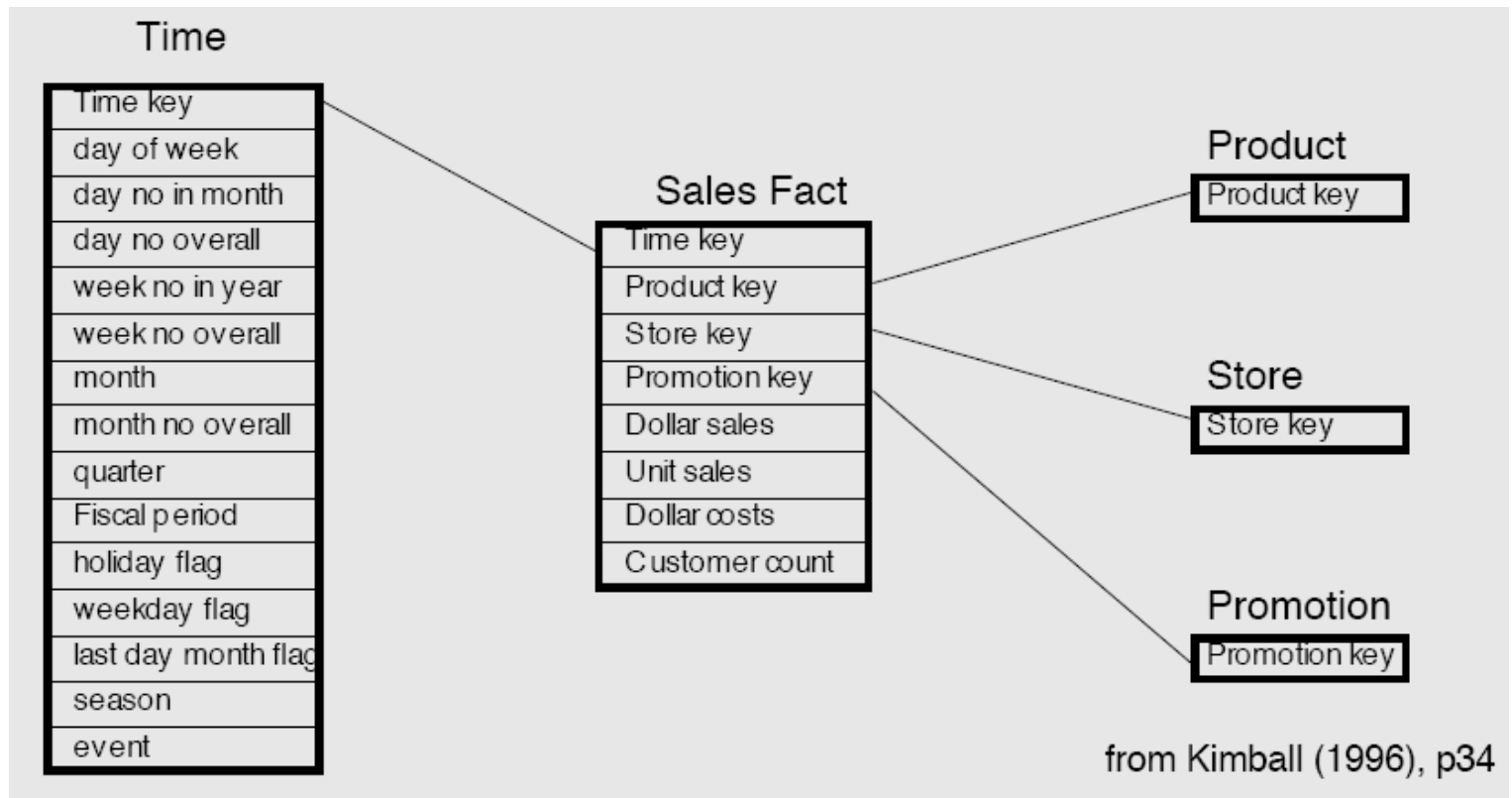
ID Promotion
Promotion Name
Price Reduction Type
Ad Type
Display Type
Coupon Type
Ad Media Type
Display Provider
Promo Cost
Promo Begin Date
Promo End Date

ID Produk
Produk Name
SKU Description
SKU Number
Size
Brand
Sub Category
Category
Department
Package Type
Diet Type
Weight
Weight Unit of Measurement
Units per Retail Case
Units per Ship Case
Cases per Pallet



# Database Design

## ■ Dimensi Waktu





# Transformasi Dimensi Waktu

---

- Hari Dalam Seminggu
- Nomor Hari Dalam Sebulan
- Nomor Hari Keseluruhan
- Nomor Minggu Dalam Setahun
- Nomor Minggu Keseluruhan
- Bulan
- Nomor Bulan Keseluruhan
- Nomor Triwulanan
- Masa Fiskal
- Flag Holiday
- Flag Hari Kerja
- Flag Hari Terakhir Dalam Sebulan
- Musim
- Event



# ETL Detail Algorithm

---

- <Tanggal> -> <Hari>
  - Tanggal Referensi
  - Tahun Kabisat
  - Jumlah Hari Dalam Setiap Bulan
  - Menghitung Jumlah Hari Sekarang Relatif Terhadap Tanggal Referensi
  - Membagi Jumlah Hari dengan 7
- Sekarang sudah banyak dipermudah hanya dengan menggunakan fungsi library



# Transformasi Dimensi Waktu

Input	Input Lainnya	Output (Dimensi)
Tanggal		Hari dalam Seminggu
Tanggal		Nomor Hari Dalam Sebulan
Tanggal		Nomor Hari Keseluruhan
Tanggal		Nomor Minggu Keseluruhan
Tanggal		Bulan
Tanggal		Nomor Bulan Keseluruhan
Tanggal		Nomor Triwulanan
Tanggal		Masa Fiskal
Tanggal	Data Hari Libur	Flag Holiday
Tanggal		Flag Hari Kerja
Tanggal		Flag Hari Terakhir Dalam Sebulan
Tanggal		Musim
Tanggal	Data Event	Flag Event



# Contoh Transformasi Waktu

---

## ■ Hari Dalam Seminggu

```
In [1]: 1 import datetime
```

```
In [2]: 1 date_string = 'October 12, 2022'
```

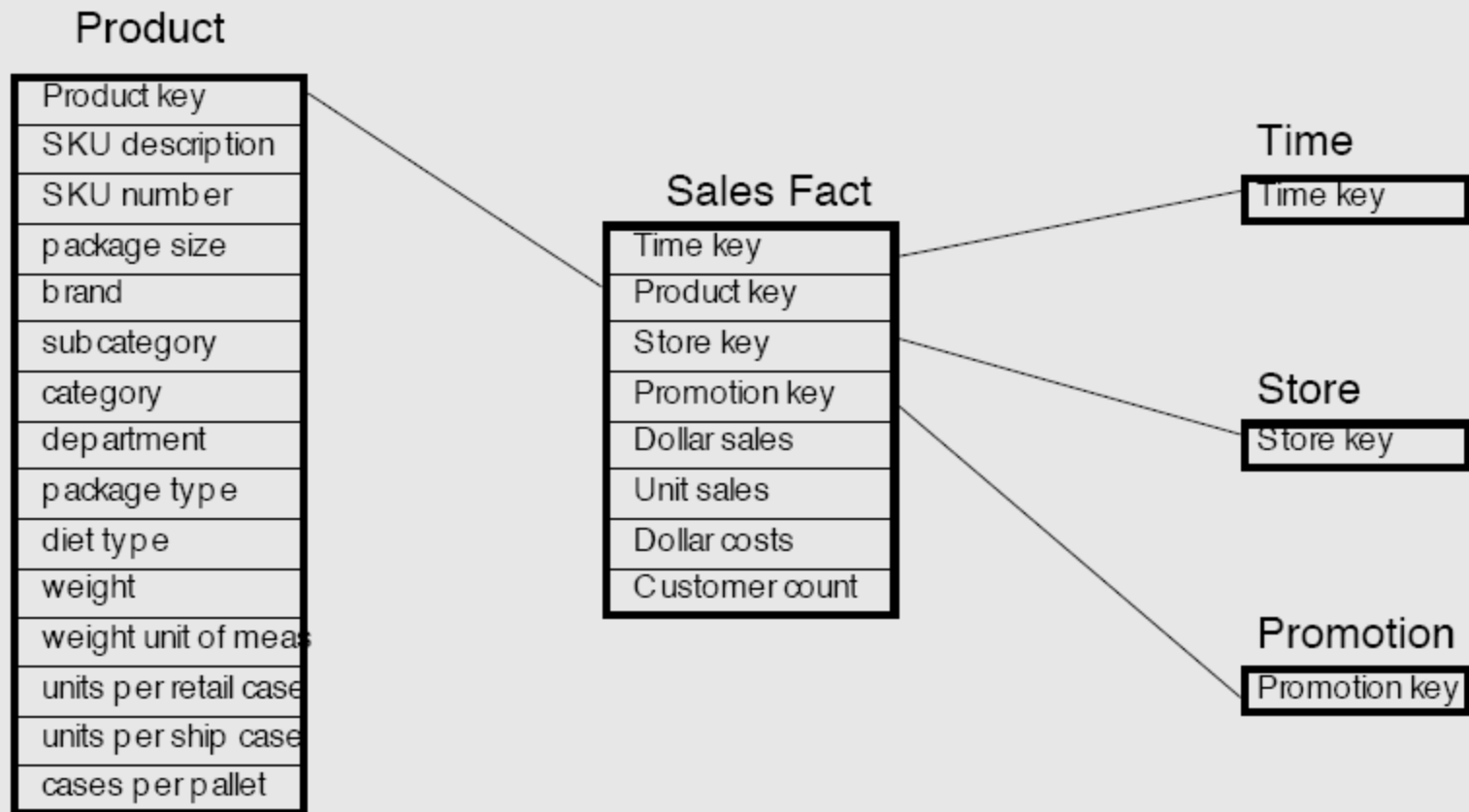
```
In [3]: 1 hari = datetime.datetime.strptime(date_string, '%B %d, %Y').strftime('%A')
```

```
In [4]: 1 hari
```

```
Out[4]: 'Wednesday'
```

```
In [ ]: 1
```

# Database Design (Dimensi Produk)





# Transformasi Dimensi Produk

---

- Data Spesifik Produk
  - Deskripsi SKU
  - Nomor SKU
- Data Non Spesifik Produk
  - Ukuran Paket
  - Brand
  - Sub Kategori
  - Kategori
  - Departemen
  - Tipe Paket
  - Tipe Makanan
  - Berat
  - .....

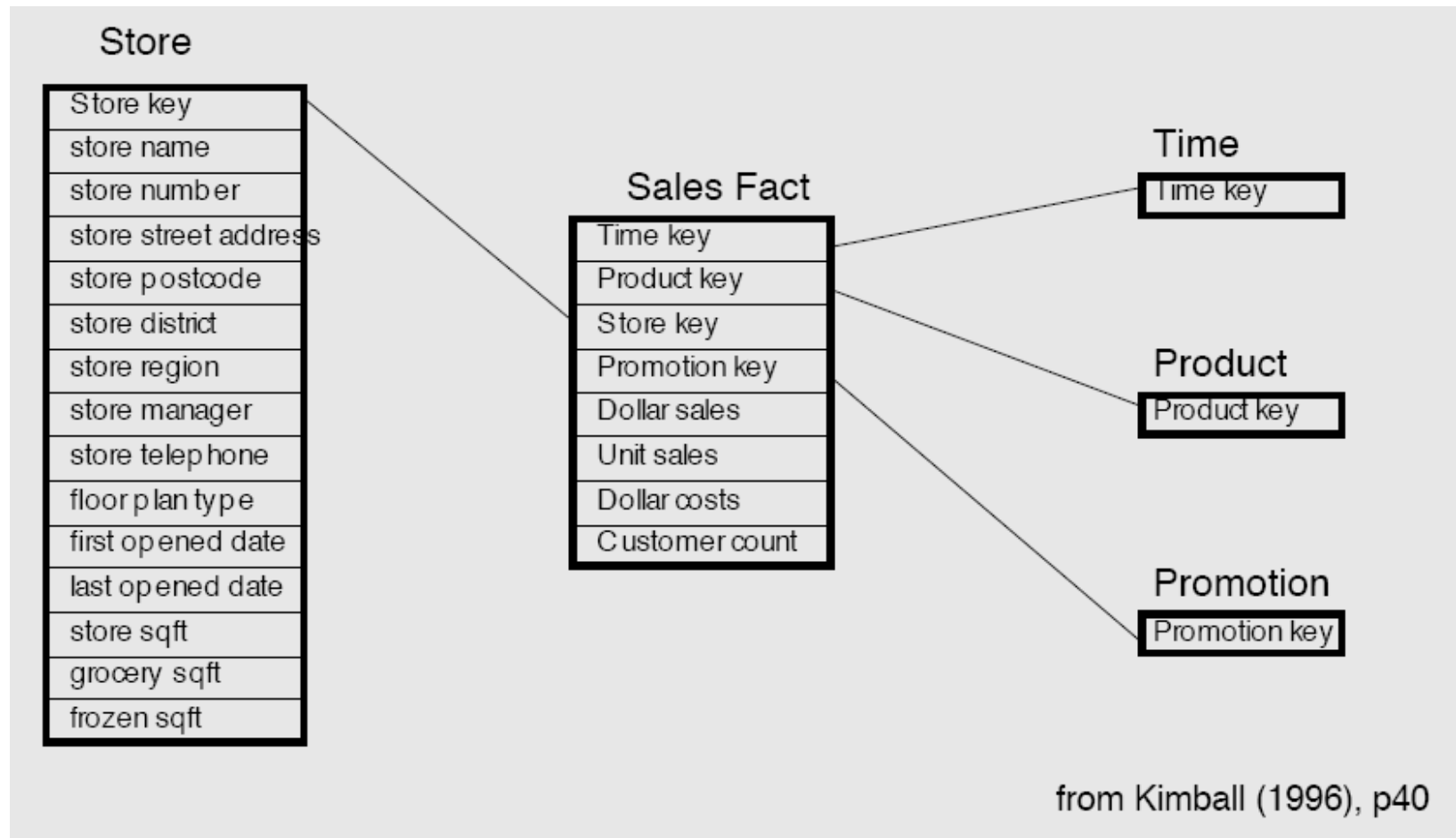


# Transformasi Dimensi Produk

Input	Input Lainnya	Output (Dimensi)
Ukuran	Ukuran untuk Kategori Kecil, Sedang, dan Besar	Ukuran (Dibagi menjadi 3 ukuran: Kecil, Sedang, Besar)
Brand		Brand
Sub Kategori		Sub Kategori
Kategori		Kategori
Departemen		Departemen
Tipe Paket		Tipe Paket
Tipe Makanan	Khusus untuk Produk Makanan	Tipe Makanan
Berat	Ukuran untuk Kategori Ringan, Sedang, dan Berat	Berat (Dibagi menjadi 3 ukuran: Ringan, Sedang, Berat)
Satuan Berat		Satuan Berat
Jumlah Per Kotak	Kategori <10, 10 – 20, >20 Per Kotak	Jumlah Per Kotak (Kategori <10, 10 – 20, >20 Per Kotak)
Jumlah Per Pengiriman	Kategori <10, 10 – 20, >20 Per Pengiriman	Jumlah Per Pengiriman (Kategori <10, 10 – 20, >20 Per Pengiriman)
Kotak Per Pengiriman	Kategori <10, 10 – 20, >20 Kotak Per Pengiriman	Kotak Per Pengiriman (Kategori <10, 10 – 20, >20 Kotak Per Pengiriman)



# Database Design (Dimensi Toko)



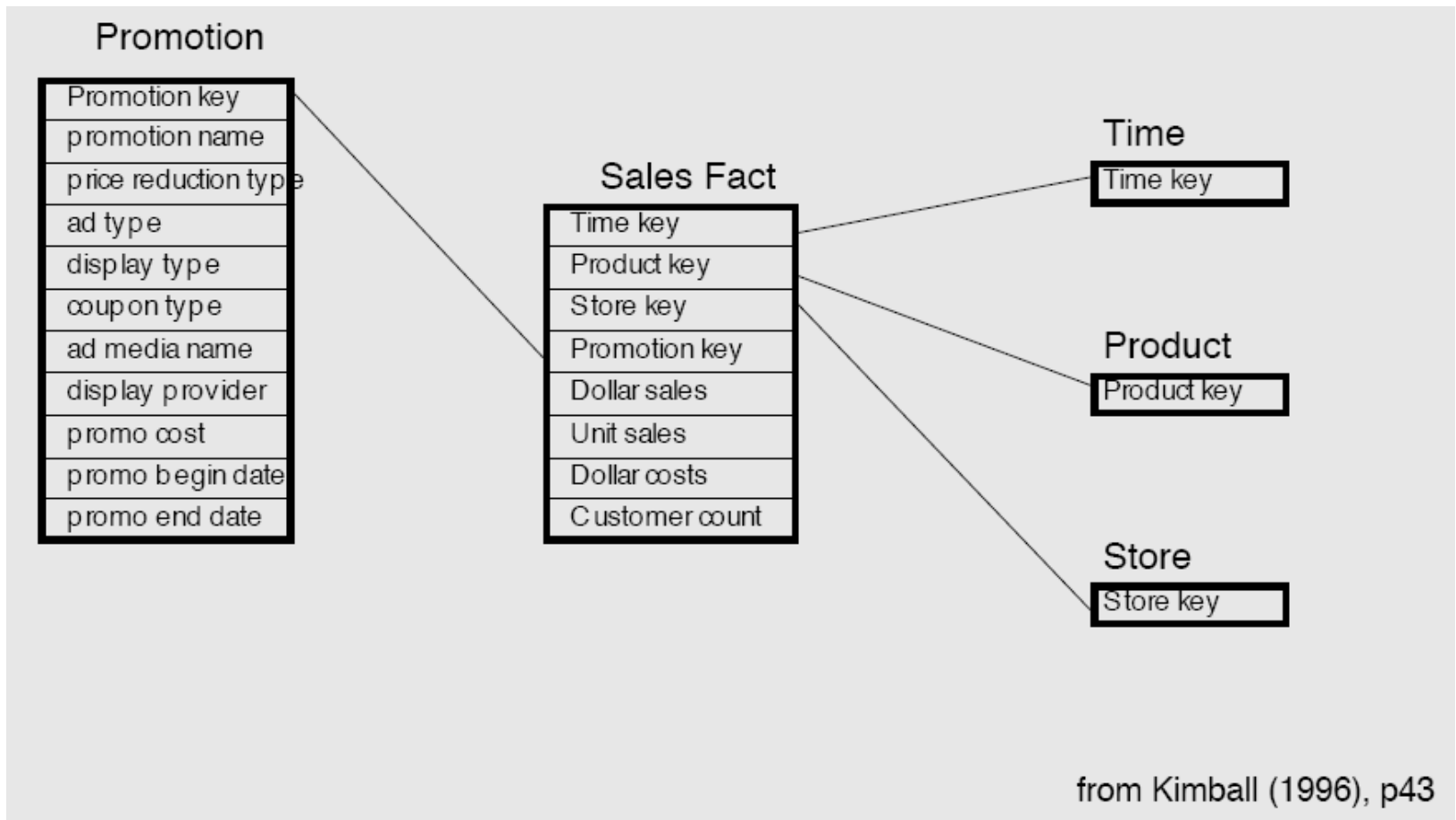


# Transformasi Informasi Retail

---

- Data Spesifik Toko
  - Nama
  - Nomor Toko
  - Manajer
  - Nomor Telpon
- Data Non Spesifik Toko
  - Nama Jalan
  - Kode Pos
  - Kabupaten
  - Provinsi
  - Tipe Pengaturan Lantai
  - Ukuran Toko
  - Ukuran Toko Untuk Barang Umum
  - Ukuran Toko Untuk Barang Didinginkan
  - .....

# Database Design





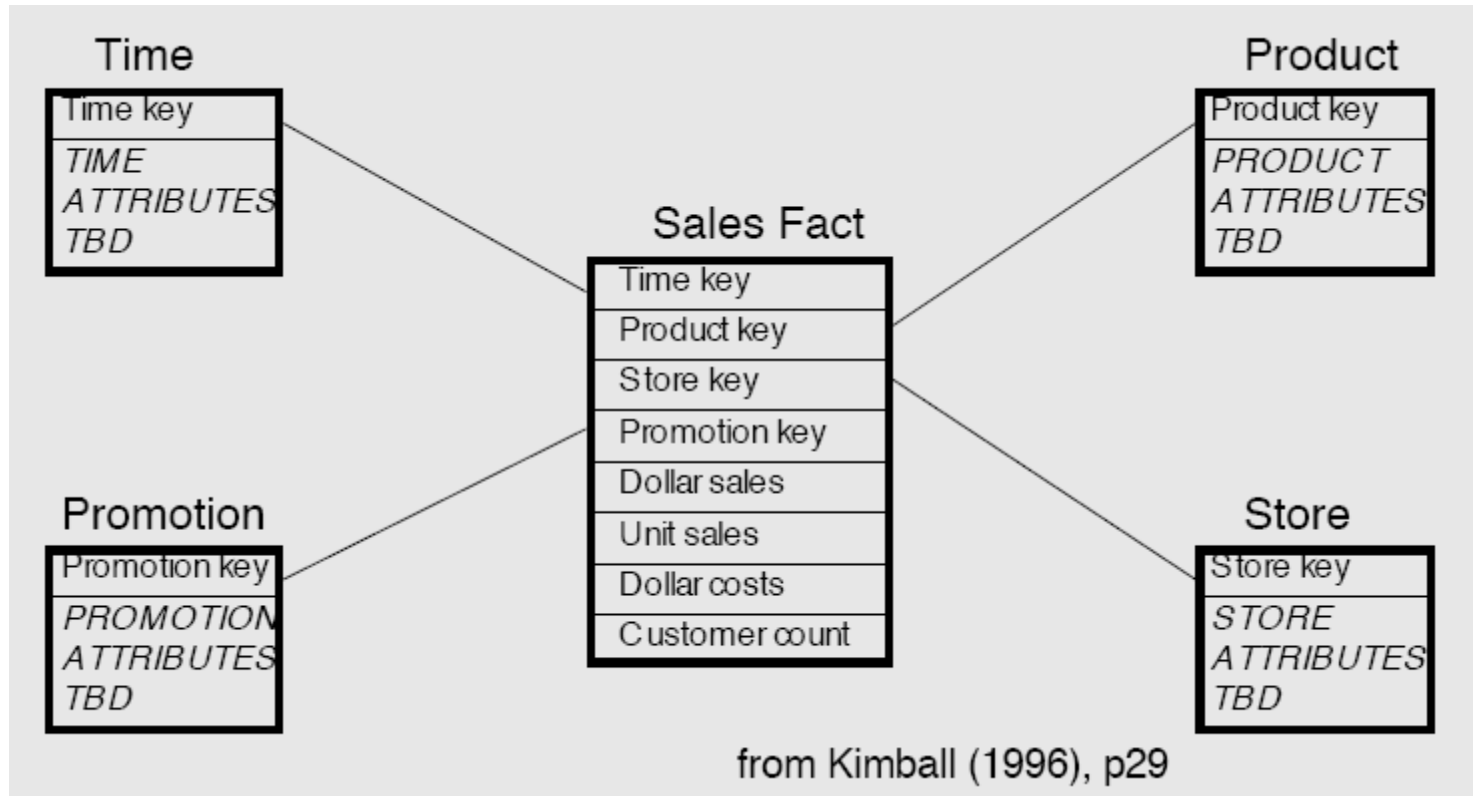
# Transformasi Promosi

---

- Data Spesifik Produk
  - Nama Promosi
- Data Non Spesifik Produk
  - Tipe Pengurangan Harga
  - Tipe Ad
  - Tipe Penataan
  - Tipe Kupon
  - Nama Media Ad
  - Perusahaan Penataan
  - Biaya Promosi
  - Tanggal Mulai Promosi
  - Tanggal Selesai Promosi

# Usaha Retail

- Memilih fakta terukur





# ETL Fakta Terukur

---

- ETL yang melakukan berbagai transformasi berdasarkan dimensi yang diperlukan
- Mencari nilai summary dari fakta terukur yang tercakup di dalam tabel fakta mengikuti fakta yang terdapat di dalam dimensinya termasuk
  - Dollar Sales
  - Unit Sales
  - Dollar Costs
  - Customers Count



# Contoh Transformasi Fakta Terukur

---

- Fakta Terukur: Dollar Sales, Unit Sales, Dollar Costs, Customers Count
- Dimensi Waktu: Weekday Flag
  - Nilai: Weekday, Weekend
- Dimensi Produk: Package Size (Ukuran)
  - Nilai: Small, Medium, Large
- Dimensi Toko: District (Kabupaten/Kota)
  - Jembrana, Tabanan, Badung, Gianyar, Klungkung, Bangli, Karangasem, Buleleng, Denpasar
- Dimensi Promosi: Promo Cost
  - Low, Medium, High



# Bentuk Penyimpanan Fakta Terukur

ID	Waktu	Produk	Toko	Promosi	Sales	Unit Sales	Costs	Customers Count
1	All	All	All	All	Rp. ...		Rp. ...	
2	Weekday	All	All	All	Rp. ...		Rp. ...	
3	Weekend	All	All	All	Rp. ...		Rp. ...	
4	All	Small	All	All	Rp. ...		Rp. ...	
5	All	Medium	All	All	Rp. ...		Rp. ...	
6	All	Large	All	All	Rp. ...		Rp. ...	
7	All	All	Jembrana	All	Rp. ...		Rp. ...	
...	...	...	...	...	Rp. ...		Rp. ...	
15	All	All	Denpasar	All	Rp. ...		Rp. ...	
16	All	All	All	Low	Rp. ...		Rp. ...	
17	All	All	All	Medium	Rp. ...		Rp. ...	
18	All	All	All	High	Rp. ...		Rp. ...	

Query semua dan query masing2 dimensi dan nilainya





# Bentuk Penyimpanan Fakta Terukur

ID	Waktu	Produk	Toko	Promosi	Sales	Unit Sales	Costs	Customers Count
19	Weekday	Small	All	All	Rp. ...		Rp. ...	
20	Weekday	Medium	All	All	Rp. ...		Rp. ...	
21	Weekday	Large	All	All	Rp. ...		Rp. ...	
22	Weekend	Small	All	All	Rp. ...		Rp. ...	
23	Weekend	Medium	All	All	Rp. ...		Rp. ...	
24	Weekend	Large	All	All	Rp. ...		Rp. ...	
25	All	Small	Jembrana	All	Rp. ...		Rp. ...	
...	...	...	...	...	Rp. ...		Rp. ...	
33	All	Small	Denpasar	All	Rp. ...		Rp. ...	
34	All	Medium	Jembrana	All	Rp. ...		Rp. ...	
...	...	...	...	...	Rp. ...		Rp. ...	
42	All	Medium	Denpasar	All	Rp. ...		Rp. ...	
...	...	...	...	...	Rp. ...		Rp. ...	

Query dengan dua nilai dimensi yang berbeda-beda



# Bentuk Penyimpanan Fakta Terukur

ID	Waktu	Produk	Toko	Promosi	Sales	Unit Sales	Costs	Customers Count
...	Weekday	Small	Jembrana	All	Rp. ...		Rp. ...	
...	...	...	...	All	Rp. ...		Rp. ...	
...	Weekday	Small	Denpasar	All	Rp. ...		Rp. ...	
...	Weekday	Medium	Jembrana	All	Rp. ...		Rp. ...	
...	...	...	...	All	Rp. ...		Rp. ...	
...	Weekday	Medium	Denpasar	All	Rp. ...		Rp. ...	
...	Weekday	Large	Jembrana	All	Rp. ...		Rp. ...	
...	...	...	...	...	Rp. ...		Rp. ...	
...	Weekday	Large	Denpasar	All	Rp. ...		Rp. ...	
...	Weekend	Small	Jembrana	All	Rp. ...		Rp. ...	
...	...	...	...	...	Rp. ...		Rp. ...	
...	Weekend	Small	Denpasar	All	Rp. ...		Rp. ...	
...	...	...	...	...	Rp. ...		Rp. ...	

Query dengan tiga nilai dimensi yang berbeda-beda



# Bentuk Penyimpanan Fakta Terukur

ID	Waktu	Produk	Toko	Promosi	Sales	Unit Sales	Costs	Customers Count
...	...	...	...	...	Rp. ...		Rp. ...	
26127350	Weekend	Large	Bangli	Low	Rp. ...		Rp. ...	
26127351	Weekend	Large	Bangli	Medium	Rp. ...		Rp. ...	
26127352	Weekend	Large	Bangli	High	Rp. ...		Rp. ...	
26127353	Weekend	Large	Karangasem	Low	Rp. ...		Rp. ...	
26127354	Weekend	Large	Karangasem	Medium	Rp. ...		Rp. ...	
26127355	Weekend	Large	Karangasem	High	Rp. ...		Rp. ...	
26127356	Weekend	Large	Buleleng	Low	Rp. ...		Rp. ...	
26127357	Weekend	Large	Buleleng	Medium	Rp. ...		Rp. ...	
26127358	Weekend	Large	Buleleng	High	Rp. ...		Rp. ...	
26127359	Weekend	Large	Denpasar	Low	Rp. ...		Rp. ...	
26127360	Weekend	Large	Denpasar	Medium	Rp. ...		Rp. ...	
26127361	Weekend	Large	Denpasar	High	Rp. ...		Rp. ...	

Sampai dengan query dengan empat nilai dimensi yang berbeda-beda



# Perhitungan Jumlah Record

---

- Dimensi Waktu:  $2! = 2$
- Dimensi Produk:  $3! = 6$
- Dimensi Toko:  $9! = 362880$
- Dimensi Promosi:  $3! = 6$
- Total Record di Tabel Fakta:
  - $\text{Dim Waktu} \times \text{Dim Produk} \times \text{Dim Toko} \times \text{Dim Promosi} + 1$
  - $2 \times 6 \times 362880 \times 6 + 1$
  - 26127361
- Tambahan 1 adalah untuk perhitungan query All Waktu, All Produk, All Toko, dan All Promosi



# Beberapa Kondisi Khusus

---

Yudi Agusta, PhD

Kecerdasan Bisnis, Lecture 04 05 06 07

Copyright © Yudi Agusta, PhD 2024



# Design Data Warehouse Usaha Layanan Keuangan

---

- Latar Belakang:
  - Umumnya bank besar
  - Layanan mencakup:
    - Cheque account, savings account, mortgage loans, investment loans, credit cards etc.
  - Goal
    - Untuk memasarkan produk ke setiap keluarga secara efektif
  - Membangun data warehouse keluarga untuk melacak accounts, pemilik account, dan pengelompokan keluarga tersebut



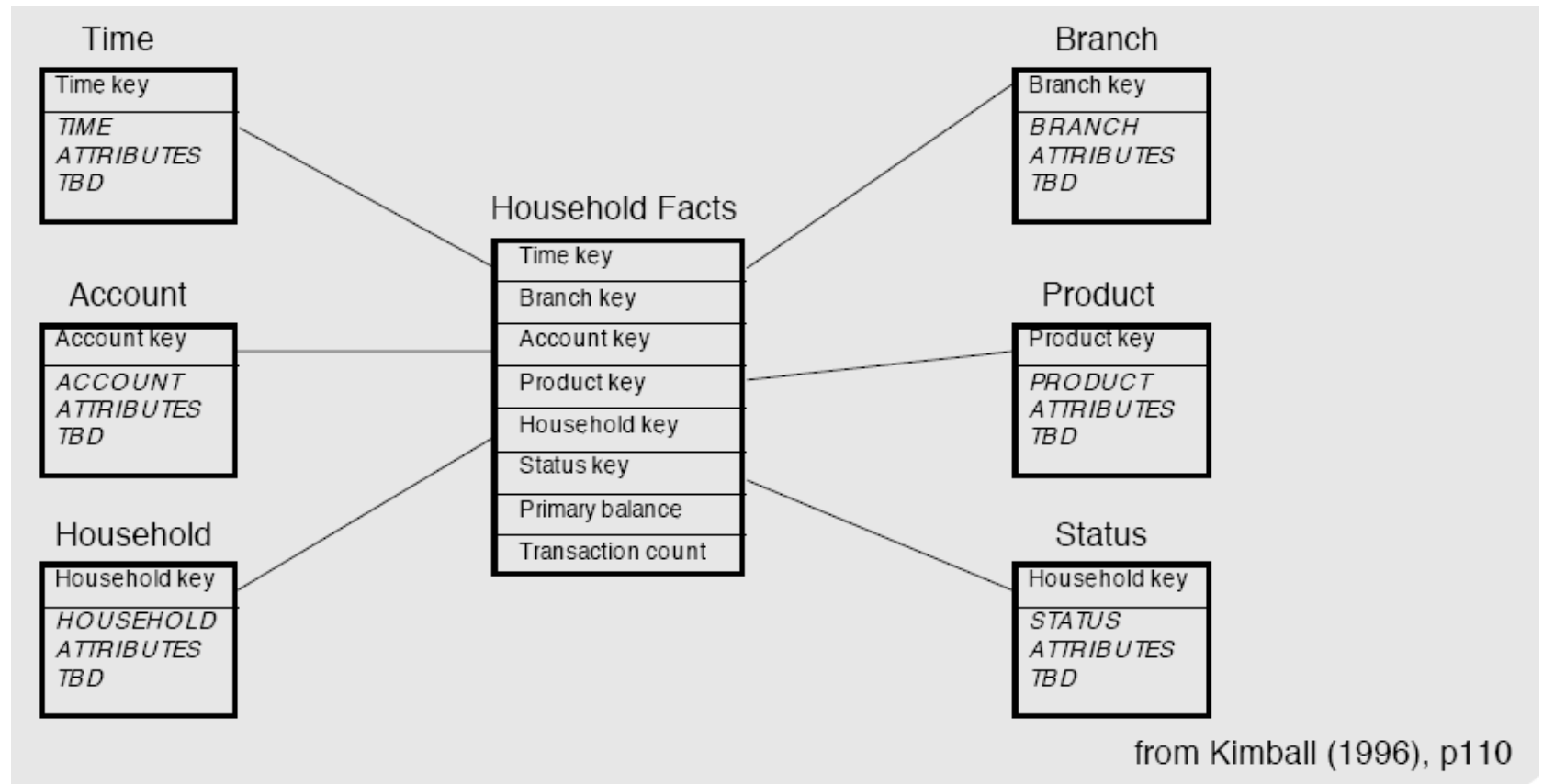
# Design Data Warehouse Usaha Layanan Keuangan

---

- Yang Diperlukan
  - Lima tahun dari data bulanan untuk setiap account
  - Untuk bulan sekarang harus diprotret dari hari sebelumnya
  - Setiap tipe account mempunyai atribut dan fakta numeric yang berbeda-beda
  - Setiap account terkait dengan suatu keluarga
  - Catatan nama dan alamat mengenai pemilik account mungkin berbeda untuk setiap account
  - Menarik untuk mencari penyebaran dan aktivitas dari setiap account

# Design Data Warehouse Usaha Layanan Keuangan

## ■ Memilih fakta terukur







# Produk Dengan Jenis Berbeda

---

- Merupakan situasi umum di perbankan
- Setiap tipe account mempunyai sejumlah fakta yang tidak terasosiasikan dengan tipe account lainnya
  - Savings
    - Jumlah bunga yang dibayarkan
  - Cheque
    - Overdraft limit
  - Credit cards
    - Credit limit
- Dalam keadaan seperti ini, rancangan logic dari tabel fakta mencakup semua fakta tambahan dan atribut dimensi di dalam tabel
- Dapat dengan mudah mencakup lusinan atau lebih atribut untuk setiap jenis account atau produk

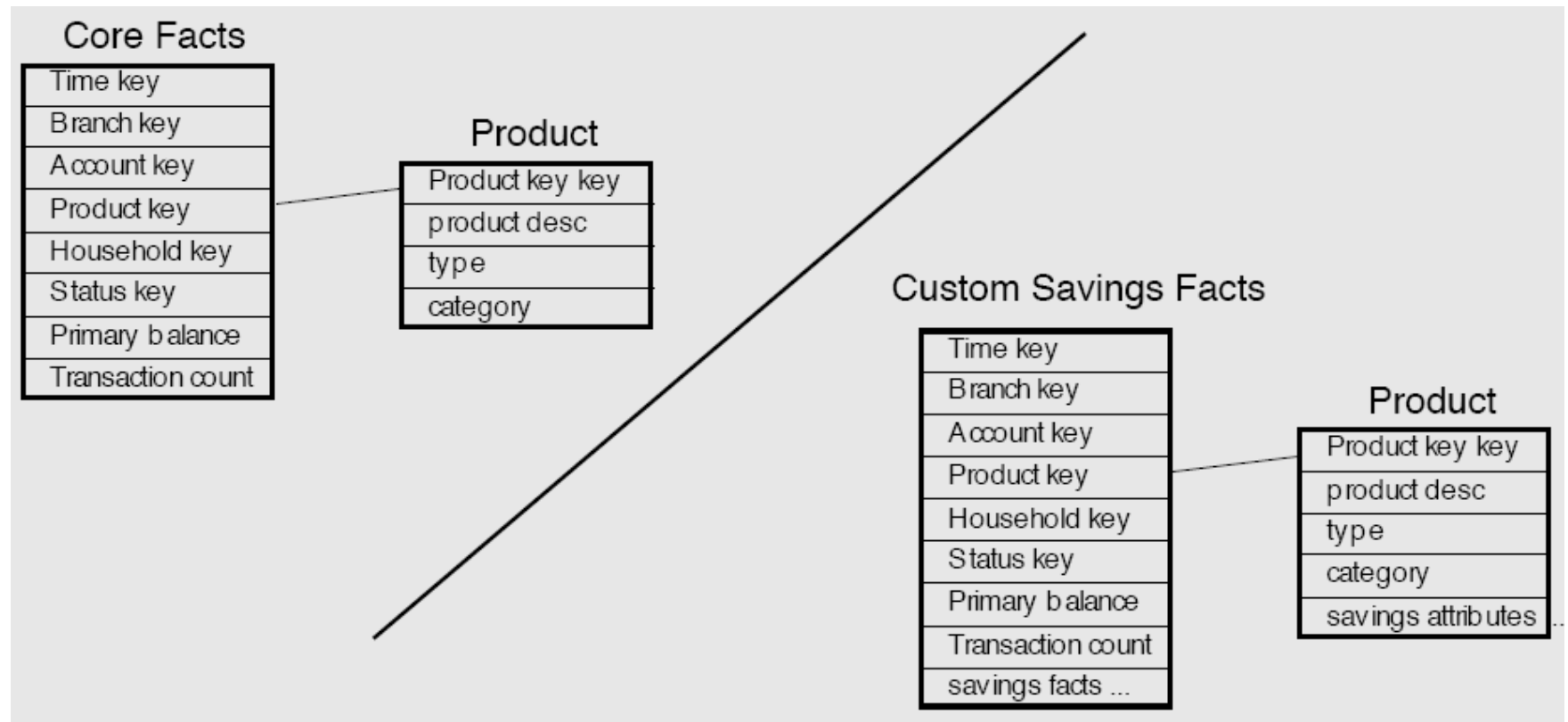


# Secara Riil - Bencana

---

- Setiap fakta dan tabel produk dapat mempunyai lebih dari 100 field
  - Terlalu mahal bila dihitung dari segi media penyimpan, kecepatan dan tingkat kekompleksan
- Kebanyakan fields dalam sebagian besar records akan kosong
- **Jawabannya** - memecah fakta dan tabel dimensi sesuai dengan tipe produk
- Pekerjaan tambahan bagi meta data!

# Solusi: Fakta Utama dan Yang Dibuat



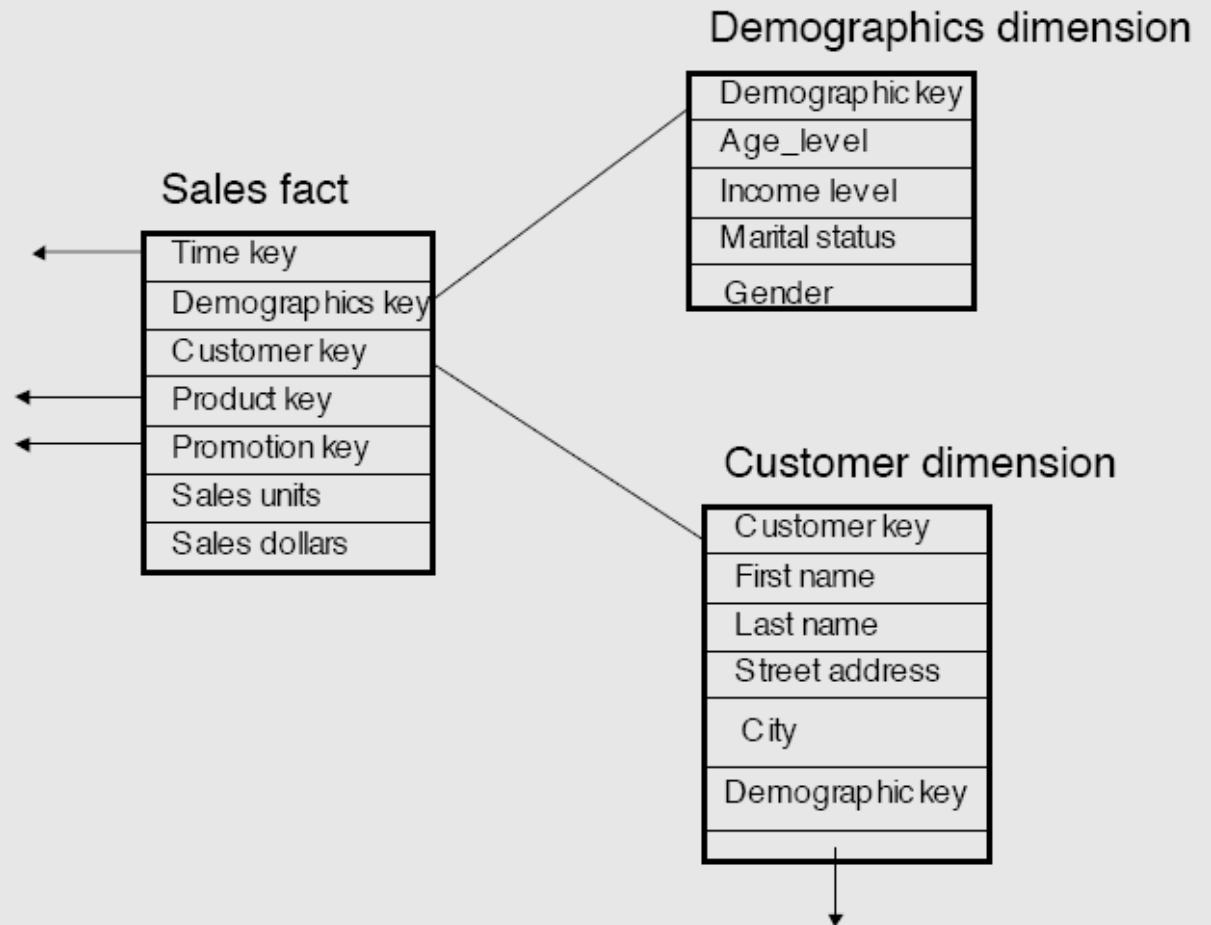


# Penggunaan Dimensi Mini

---

- Umumnya, digunakan untuk melihat sebaran atau kategori
- Membuat link antara tabel dimensi
- Digunakan untuk mengelompokkan nilai yang bersifat continuous ke dalam grup
  - Contoh: Pendapatan, Umur dll
- Mungkin mempunyai lebih dari 1 tabel dimensi mini
  - 2-3 tabel dimensi mini bukan tidak biasa

# Dimensi Mini





# Studi Kasus Kegiatan Inventory (Tugas)

---

Yudi Agusta, PhD

Kecerdasan Bisnis, Lecture 04 05 06 07

Copyright © Yudi Agusta, PhD 2024



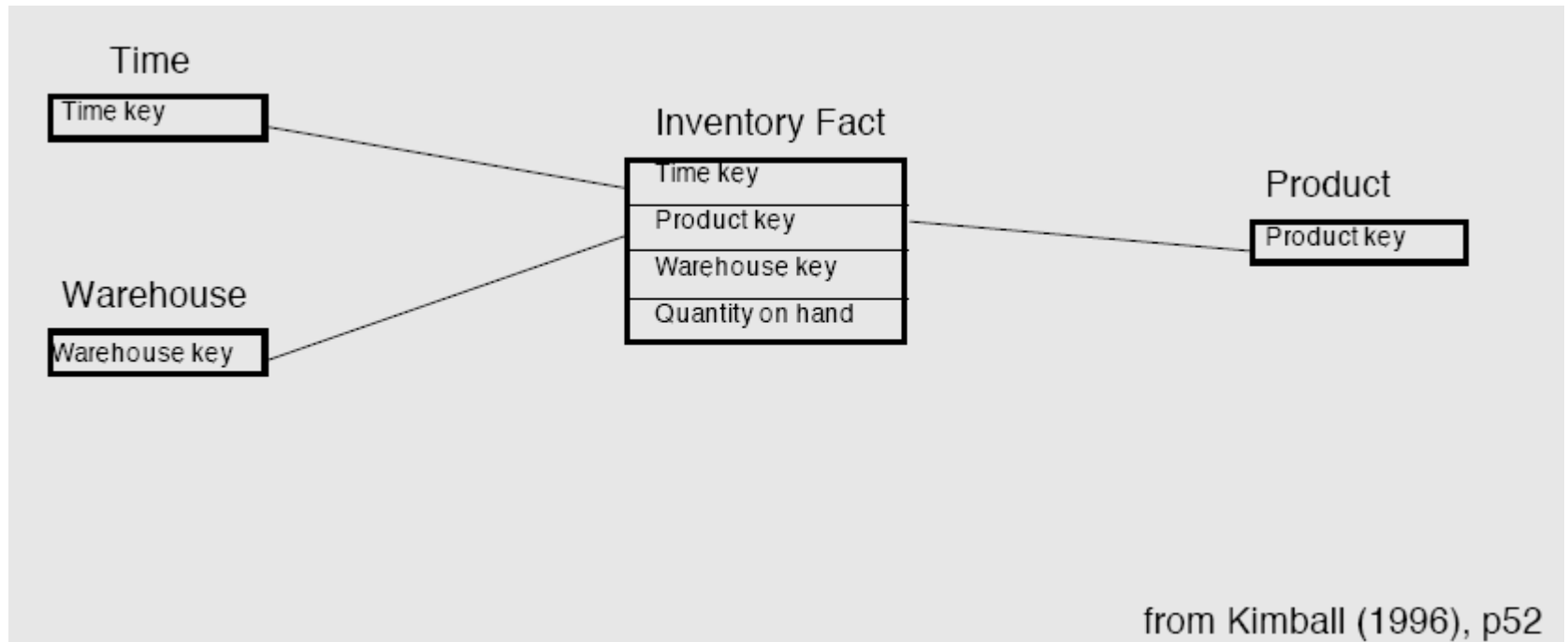
# Inventory

---

- Latar Belakang:
  - Sistem inventory memberikan servis sebagai 'penengah' antara perusahaan manufaktur dan perusahaan retail
  - Proses penambahan nilai
  - Ada tiga jenis model inventory
    - Potret Inventory
    - Status Pengiriman
    - Transaksi

# Model Potret Inventory

- Untuk jangka waktu tertentu, level inventory diukur dan dicatat







# Inventory

---

- Lengkapi field-field di dalam tabel dimensi
- Susun disain sistem informasi sebagai sumber datanya
- Analisa proses ETL yang diperlukan
- Presentasikan



# Summary Lecture 08 09

---

Yudi Agusta, PhD  
Kecerdasan Bisnis, Lecture 04 05 06 07

Copyright © Yudi Agusta, PhD 2024



# Summary Lecture 08 09

---

- Contoh Sistem Dashboard
- Identifikasi Fasilitas Dalam Sistem Dashboard
- Latihan Mendisain Sistem Dashboard