

Pemodelan Data Warehouse

Budi Susanto

Teknik Informatika
Universitas Kristen Duta Wacana
Yogyakarta

Tujuan

- Memahami konsep dasar data warehouse
- Memahami pemodelan berbasis dimensi

Pengantar

- Data warehouse mendukung keputusan bisnis dengan cara mengumpulkan, mengkonsolidasikan, dan mengorganisasikan data untuk pelaporan dan analisis.
 - Tool yang digunakan: OLAP dan Data Mining
- Walaupun data warehouse dibangun di atas database relasional, perancangannya harus berbeda dari database transaksional.

Data Warehouse vs OLTP

Data Warehouse	OLTP
Dirancang untuk analisis ukuran bisnis	Dirancang untuk operasi bisnis real-time
Dioptimalkan untuk load data yang besar, kompleks, query yang tidak terprediksi.	dioptimalkan untuk sekumpulan transaksi, pada umumnya CRUD.
Diisi dengan data yang sudah konsisten dan valid, tidak memerlukan validasi real time	Dioptimalkan untuk data yang masuk selama transaksi, menggunakan validasi terhadap data.
Mendukung beberapa pemakai dari sistem OLTP	mendukung banyak sekali user.

Mengapa data warehouse?

- Tanpa data warehouse untuk menyimpan informasi historical, data disimpan di media statik atau di akumulasikan di database OLTP.
- Jika data tersedia untuk cadangan, data tersebut tidak dapat digunakan untuk analisis dan pembuatan keputusan.
- Jika data diakumulasikan di database OLTP, maka akan membebani biaya query yang semakin lama semakin besar.

Mengapa data warehouse?

- Query dapat menjadi sangat kompleks terkait dengan skema database OLTP.
- Dengan pengorganisasian dan konsolidasi ke data warehouse:
 - memungkinkan OLTP tetap dapat berjalan untuk transaksi yang tinggi.
 - Query yang digunakan untuk analisis dapat lebih efisien dan sederhana.

OLAP

- OLAP dirancang untuk menyediakan kinerja yang baik untuk query business intelligent.
- OLAP mengorganisasikan data ke dalam kubus multidimensi.
- OLAP tidak dirancang untuk menyimpan data biner/teks yang besar.

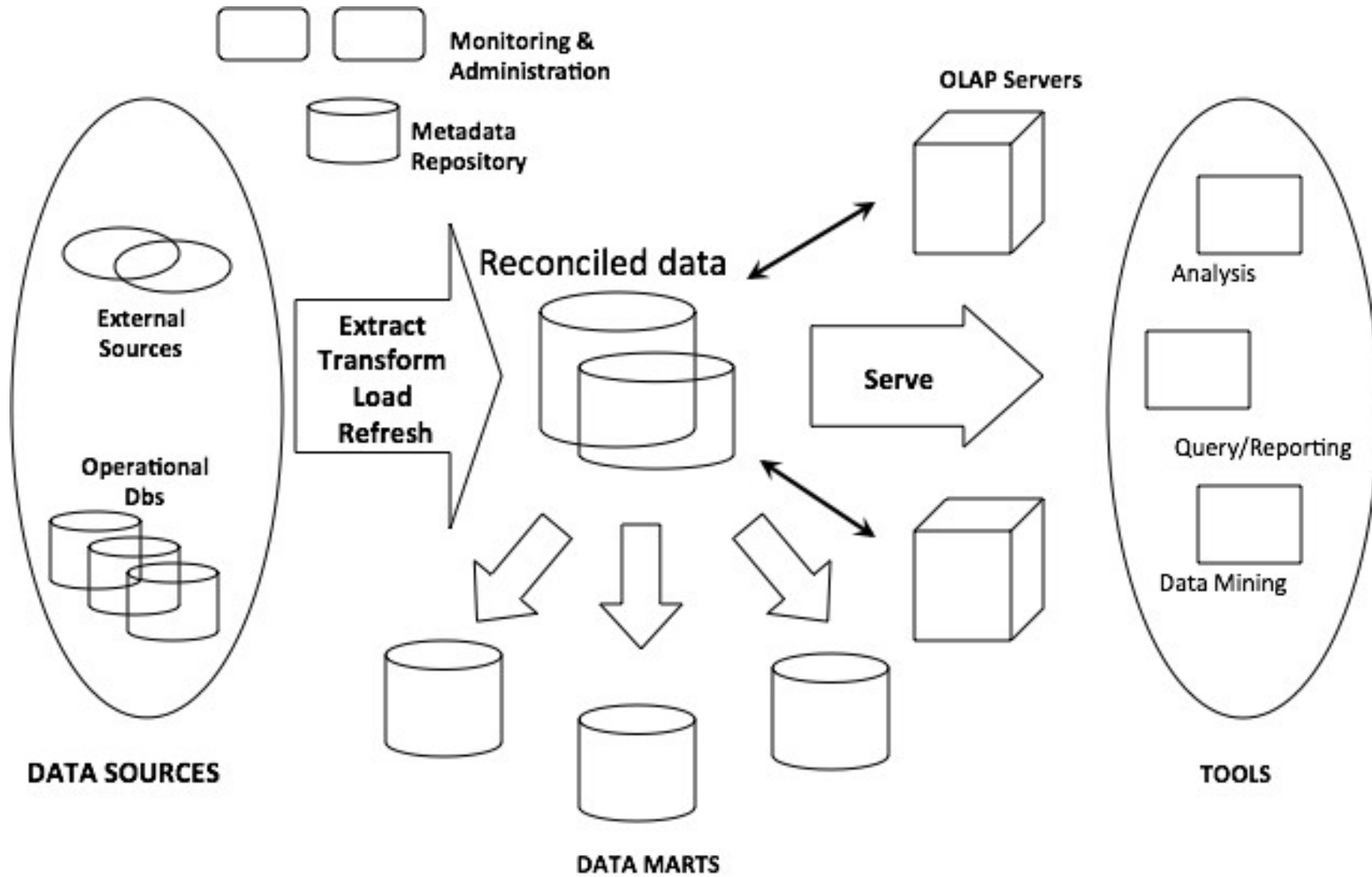
Data Mining (DM)

- DM merupakan teknologi yang menerapkan algoritma yang kompleks dan tepat untuk analisis dan menemukan informasi untuk pengambil keputusan.
- OLAP mengorganisasikan data dalam suatu model yang sesuai untuk analisis, sedangkan DM melakukan analisis data dan menyediakan hasilnya.
- OLAP mendukung model-driven analysis, dan DM mendukung data-driven analysis.

Tujuan Arsitektur Data Warehouse

- Data warehouse harus dirancang untuk memenuhi kebutuhan:
 - Memberikan pengalaman yang cukup penting kepada user.
 - Tidak perlu mengganggu OLTP
 - Menyediakan tempat penyimpanan data yang konsisten secara terpusat
 - Menjawab query kompleks secara cepat.
 - Menyediakan berbagai tool analisis, seperti OLAP dan Data Mining.

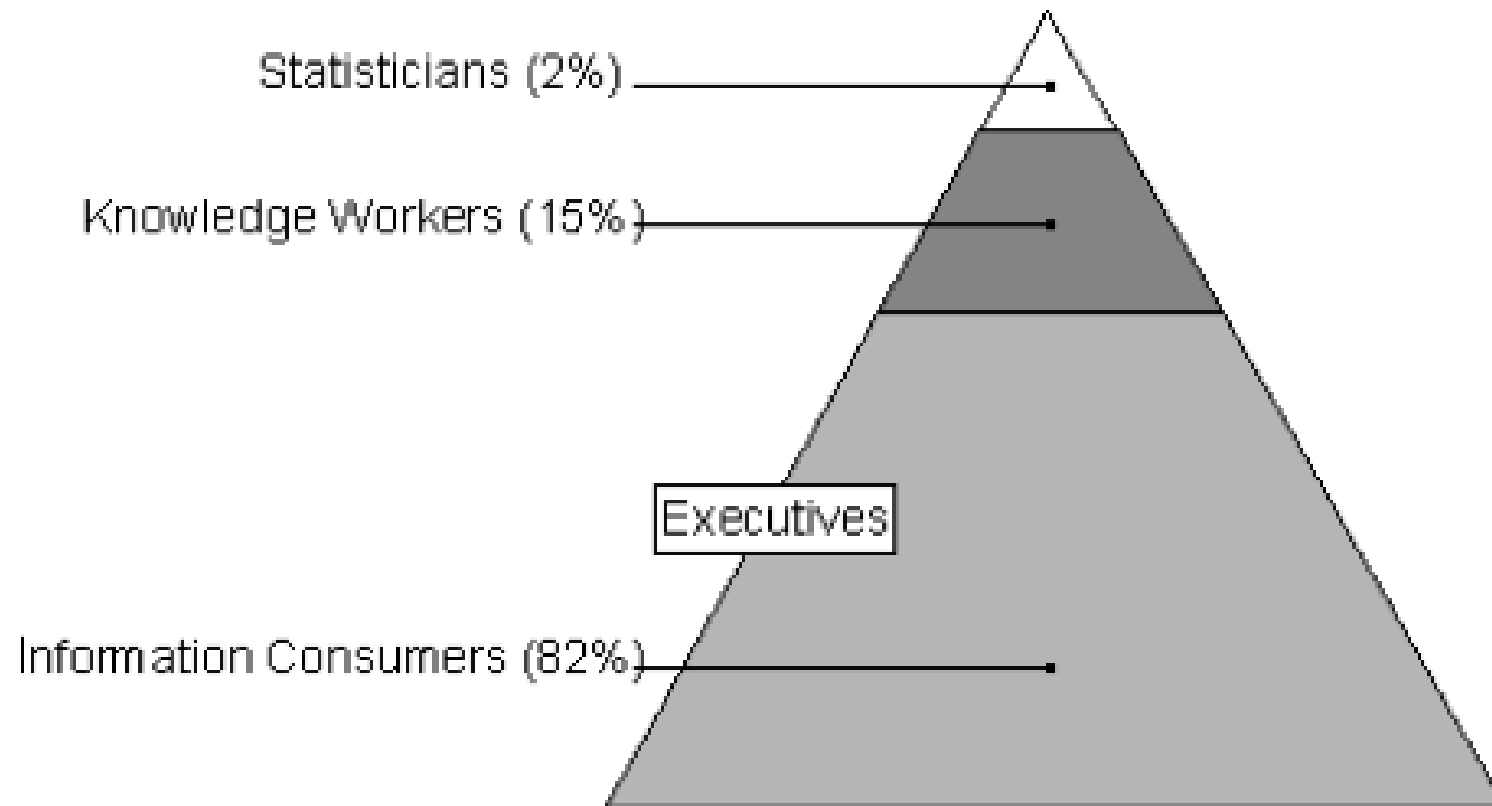
Arsitektur DW



Karakteristik Data Warehouse

- Di dasarkan pada model dimensi
- Berisi data historis
- Berisi data detil maupun ringkasan.
- Mengkonsolidasikan pemisahan data dari berbagai sumber selagi menjaga konsistensi
- Fokus pada subjek tunggal, seperti penjualan, inventori, atau keuangan.

Pemakai Data Warehouse



Pengembangan Data Warehouse

- Identifikasi kebutuhan
- Perancangan model dimensional
- Pengembangan arsitektur, termasuk Operational Data Store (ODS)
- Perancangan relational database dan OLAP
- Pengembangan aplikasi pemelihara data
- Pengembangan aplikasi analisis
- Pengujian dan pemasangan sistem

Identifikasi Kebutuhan

- Pahami bisnis sebelum diskusi dengan pemakai
- Interview dan bekerja dengan pemakai
- Temukan informasi apa yang mereka butuhkan
- Pemakai harus bekerja satu tim perancangan
- Interview dengan pakar data untuk menemukan data yang sudah ada dan dimana
- Komunikasikan dengan pemakai sesering mungkin.

Perancangan Model Dimensional

- Model harus memenuhi kebutuhan pemakai dan mudah digunakan
- Mudah untuk dipelihara
- Harus dapat diterapkan di relational database yang mendukung OLAP
- Model dimensional biasanya menggunakan star atau snowflake.

Model Dimensional

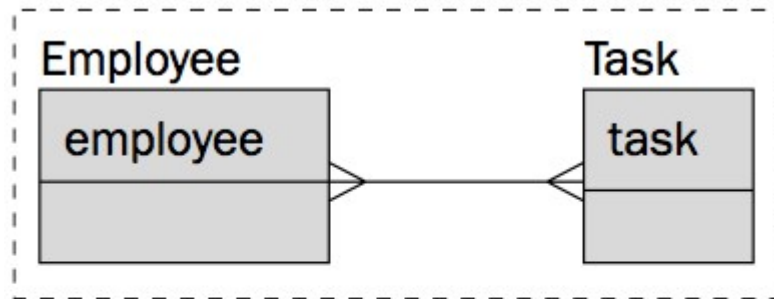
- Model yang terdapat tabel fact dan tabel dimensi.
- Tabel fact berisi transaksi historis
- Tabel dimensi berisi penjelasan dari fact.
- Dua skema
 - Skema star
 - Skema Snowflake

Denormalisasi

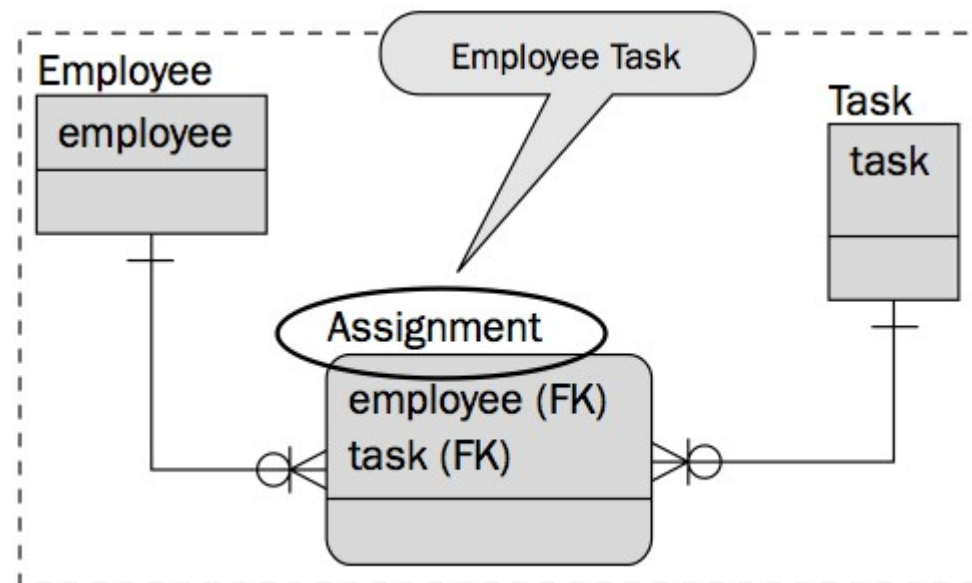
- Lawan dari Normalisasi.
- Digunakan untuk membuat model data warehouse
- Biasanya mencoba untuk melakukan granularity terbalik.

Denormalisasi 3NF

- many-to-many

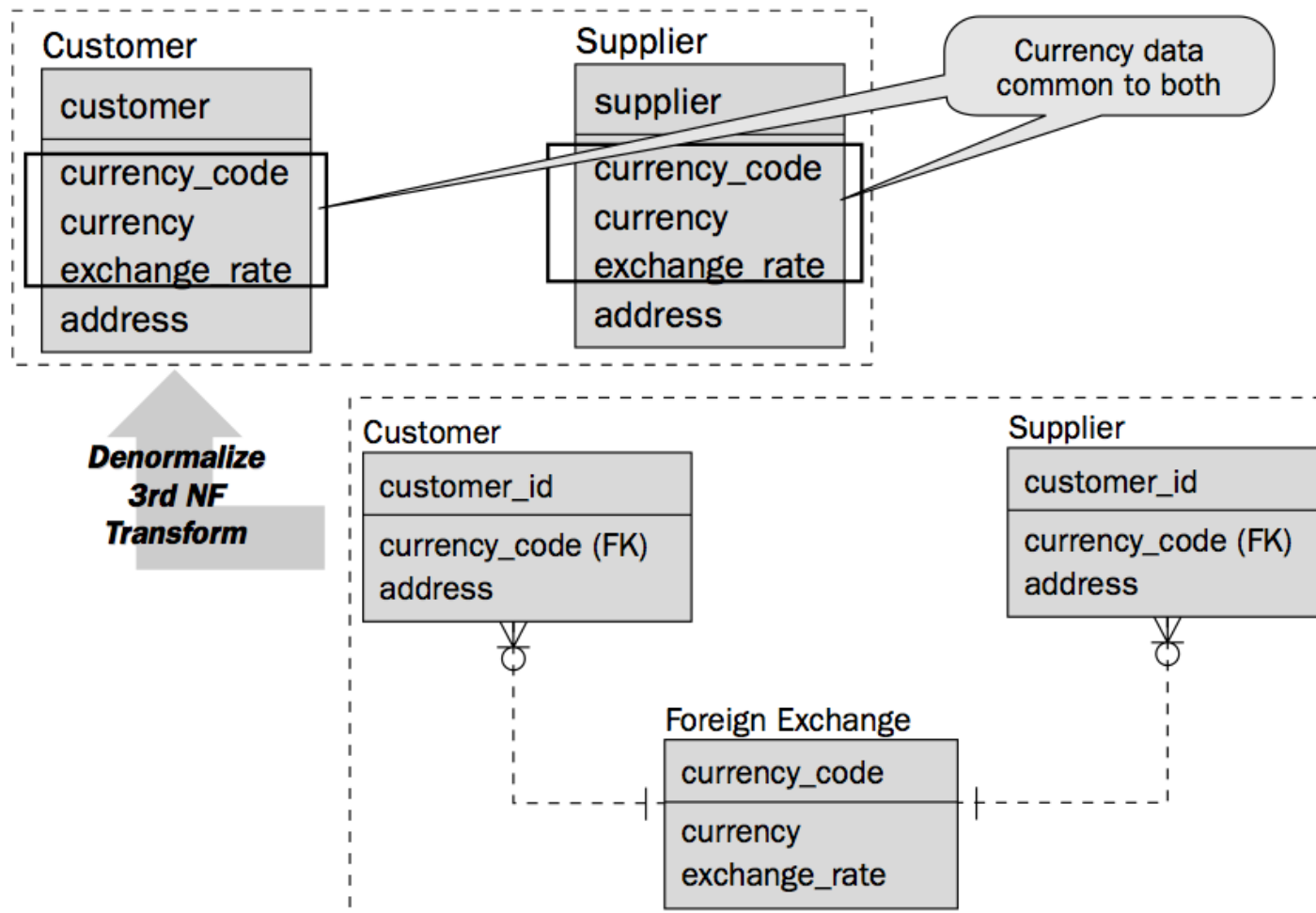


Denormalize
3rd NF
Transform



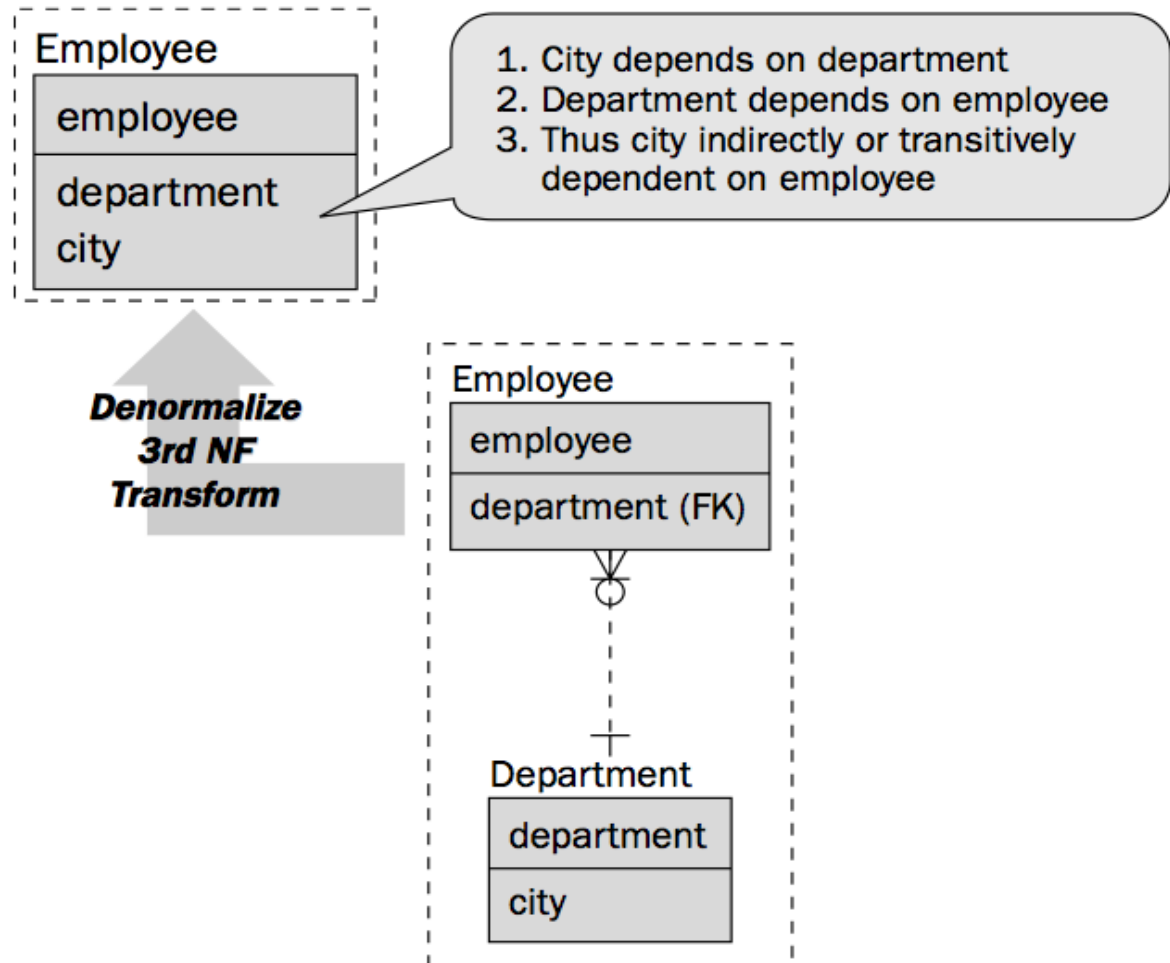
Denormalisasi 3NF

- Penggabungan atribut dalam tabel terpisah



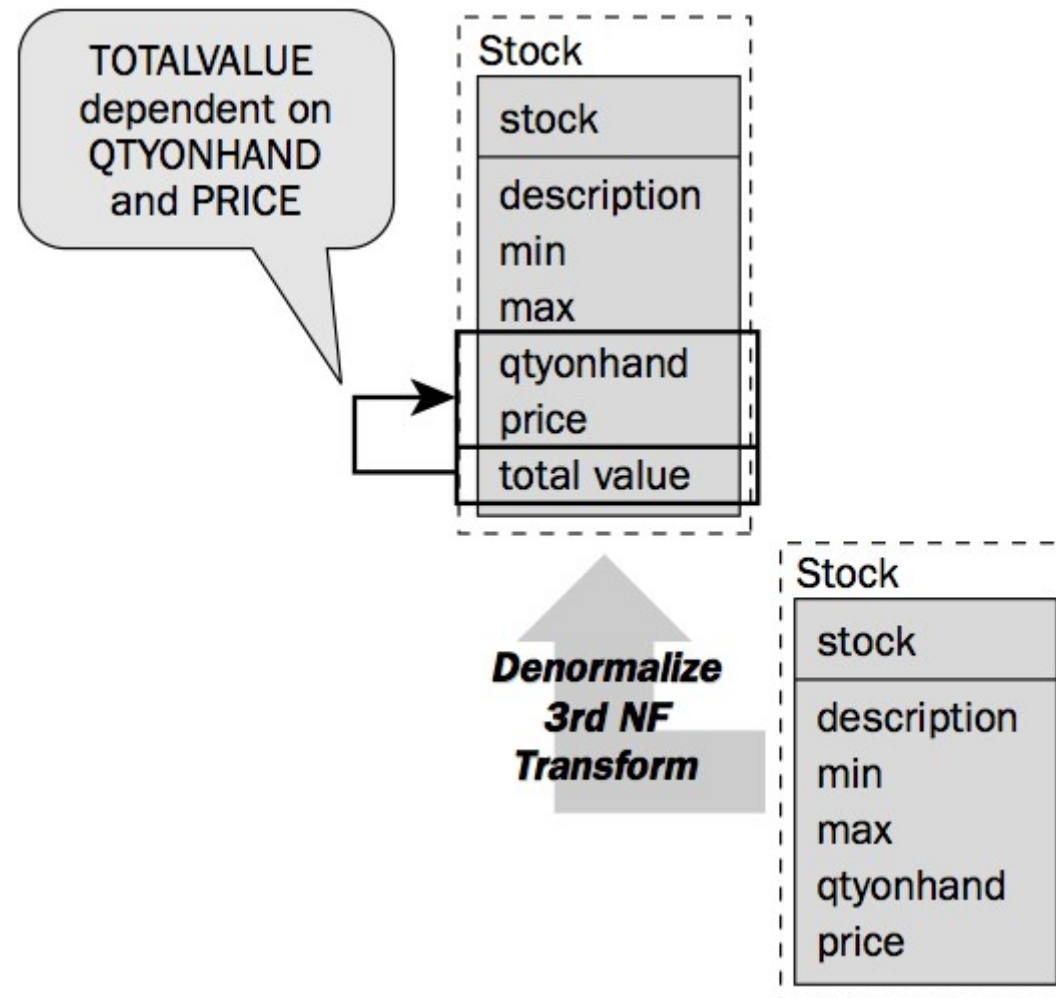
Denormalisasi 3NF

- Transitif Dependence



Denormalisasi 3NF

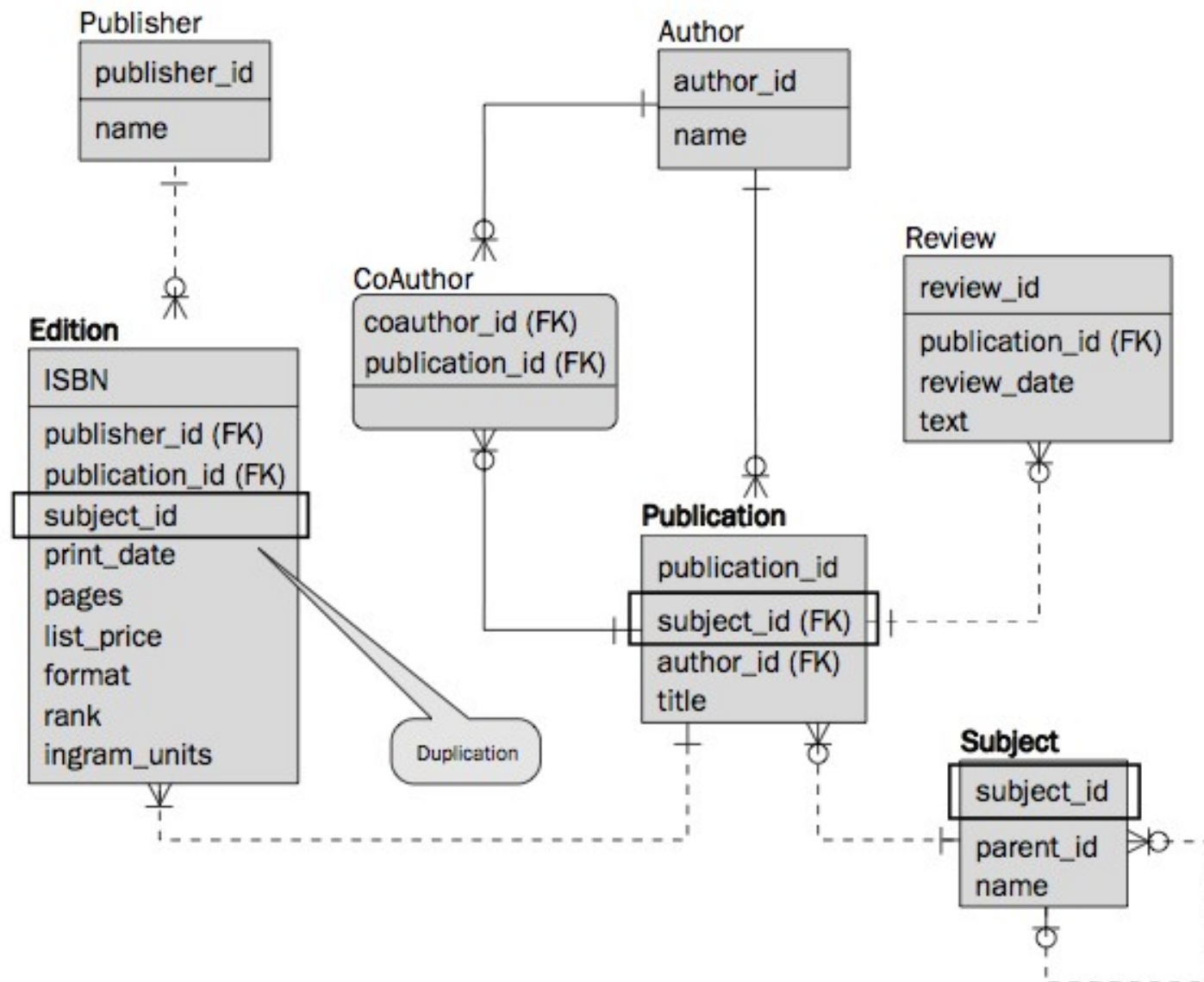
- Derived Value



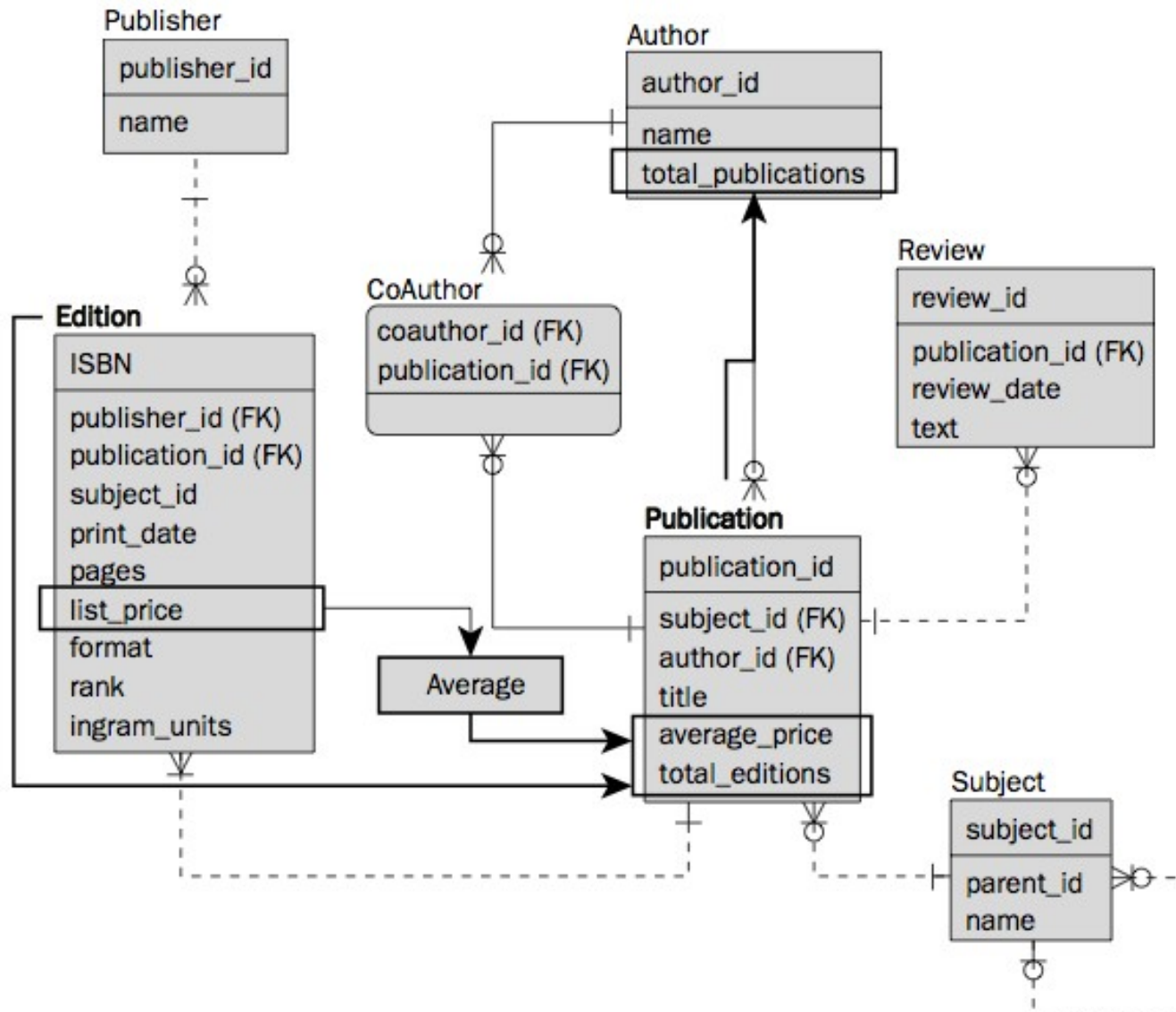
Trik Denormalisasi

- Pisahkan antara data aktif dan nonaktif.
- Copy field antar tabel.
- Field-field ringkasan diletakkan pada tabel induk.
- Pisahkan field yang jarang dan sering diakses.
 - Efek dari pemisahan data aktif dan nonaktif.

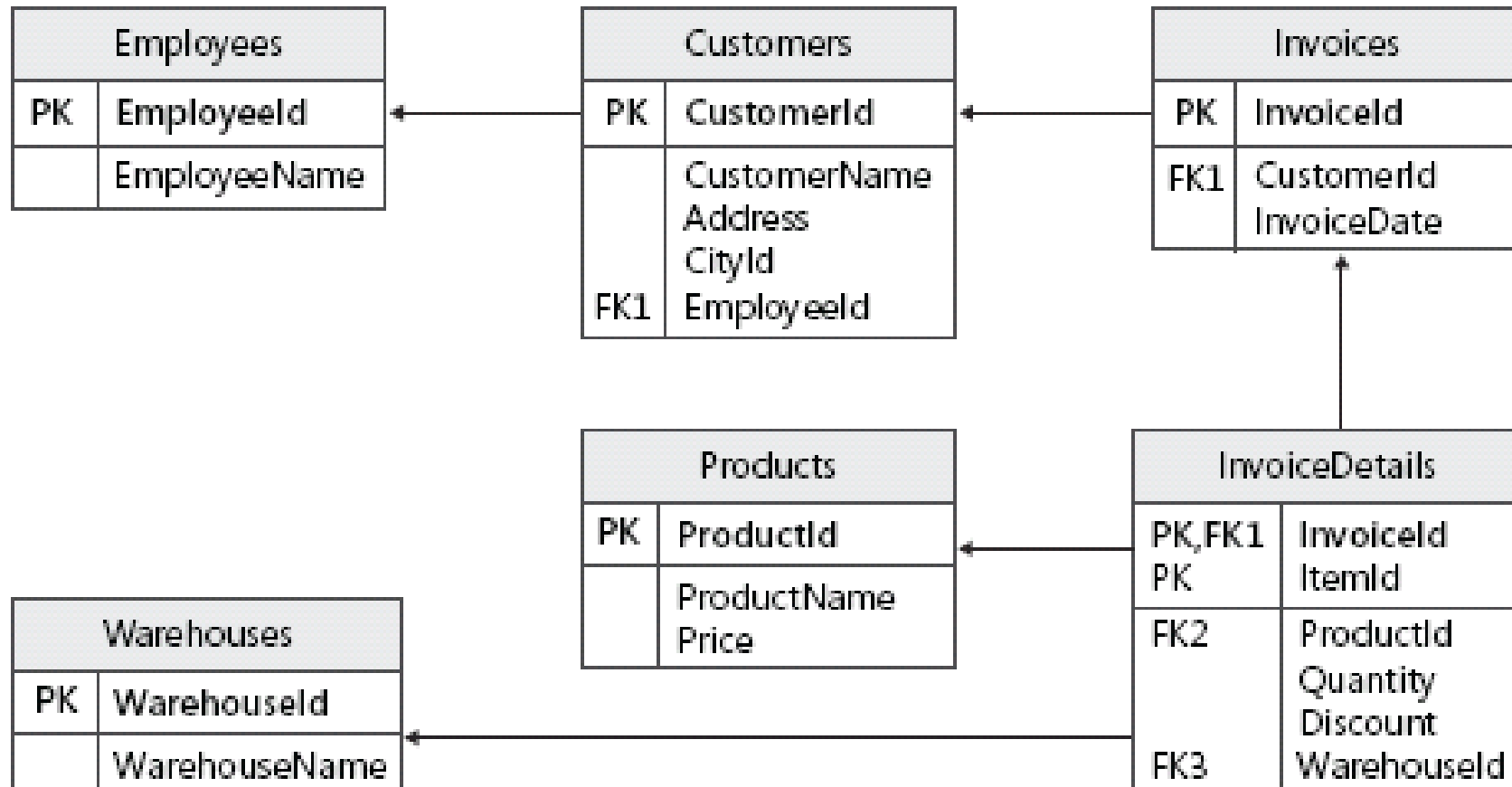
Contoh Copy field antar tabel



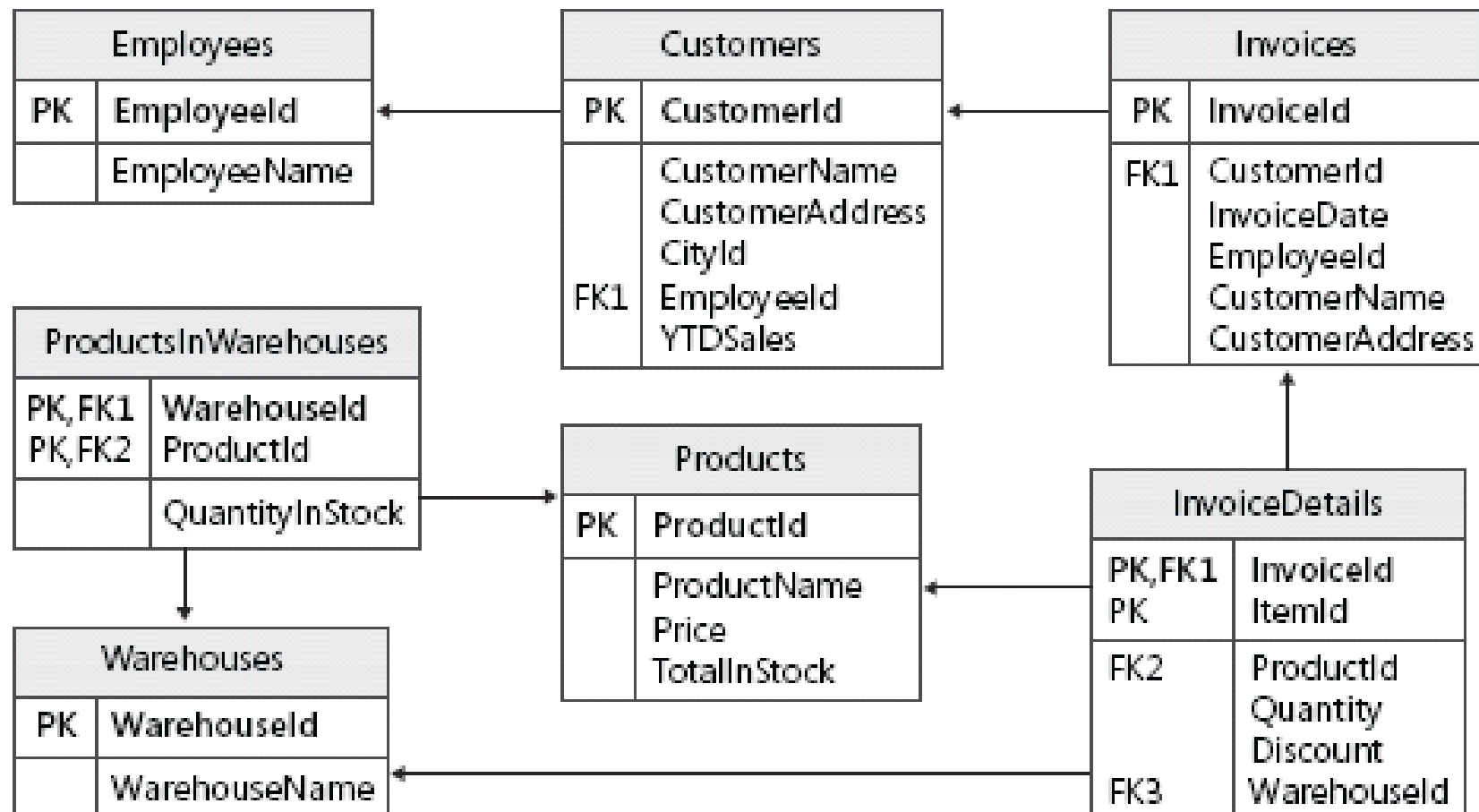
Field-field ringkasan



Sebelum Denormalisasi



Setelah Denormalisasi



Latihan Denormalisasi

- Dari 3NF ke 0NF
 - Seorang pelanggan dapat memiliki beberapa nomor telepon
 - CUSTOMER (CustomerId, CustomerName,...)
 - CUST_PHONE (CustomerId, Phone)
- Bentuk denormalisasi?

Latihan Denormalisasi

- Dari 3NF ke 2NF
 - CUSTOMER(CustomerId, CustomerName)
 - ORDER (OrderId, OrderDate, DeliveryDate, Amount, CustomerId)
 - ORDERLINE (OrderId, ItemId, QtyOrdered, OrderPrice)
 - ITEM (ItemId, ItemDescription, CurrentPrice)
- Bentuk denormalisasi?

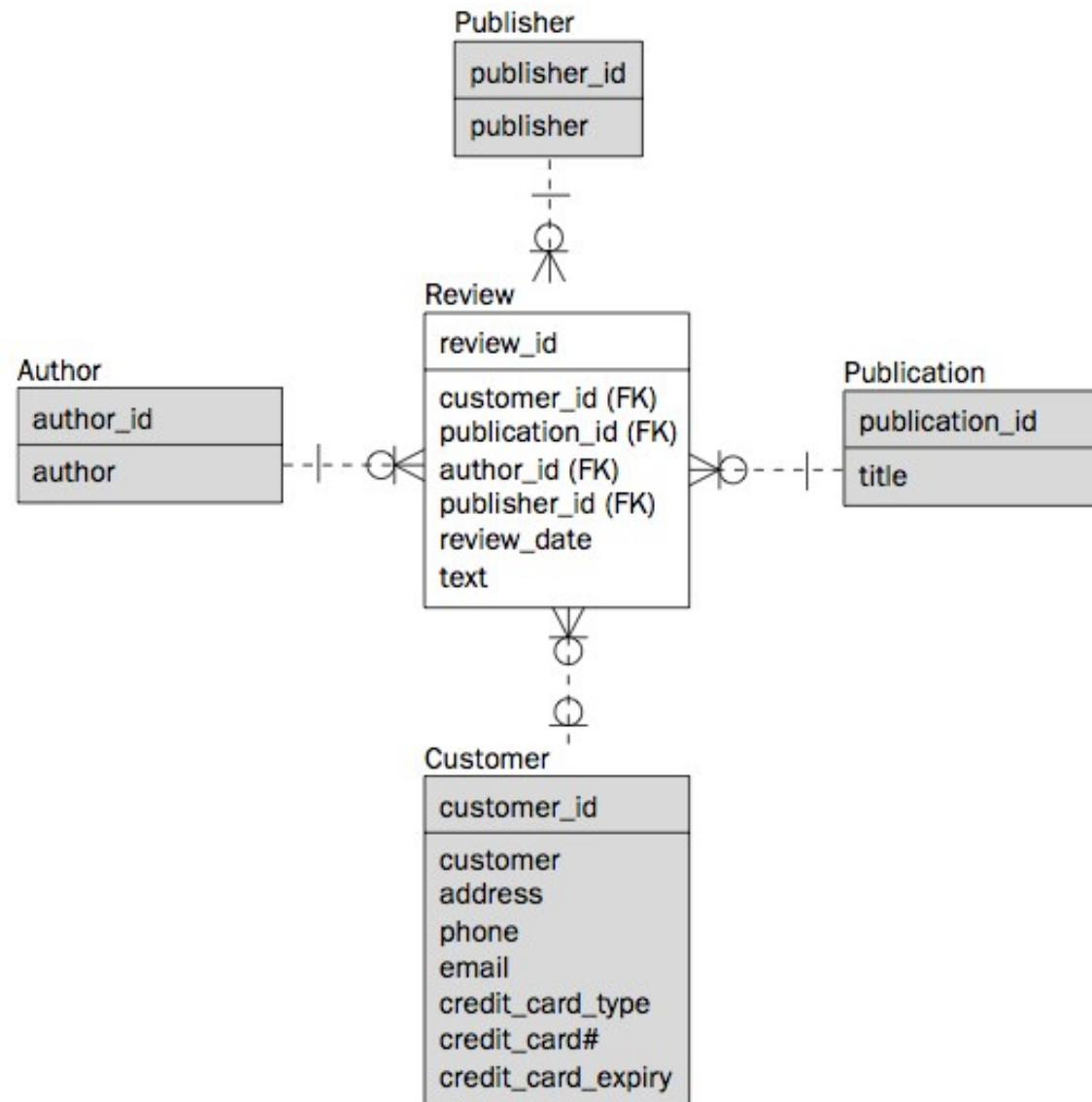
Latihan Denormalisasi

- Dari 3NF ke 2NF
 - SALESPERSON (SalespersonId, SalespersonName,...)
 - ORDER (OrderId, OrderDate, DeliveryDate, Amount, SalespersonId)
- Bentuk denormalisasi?

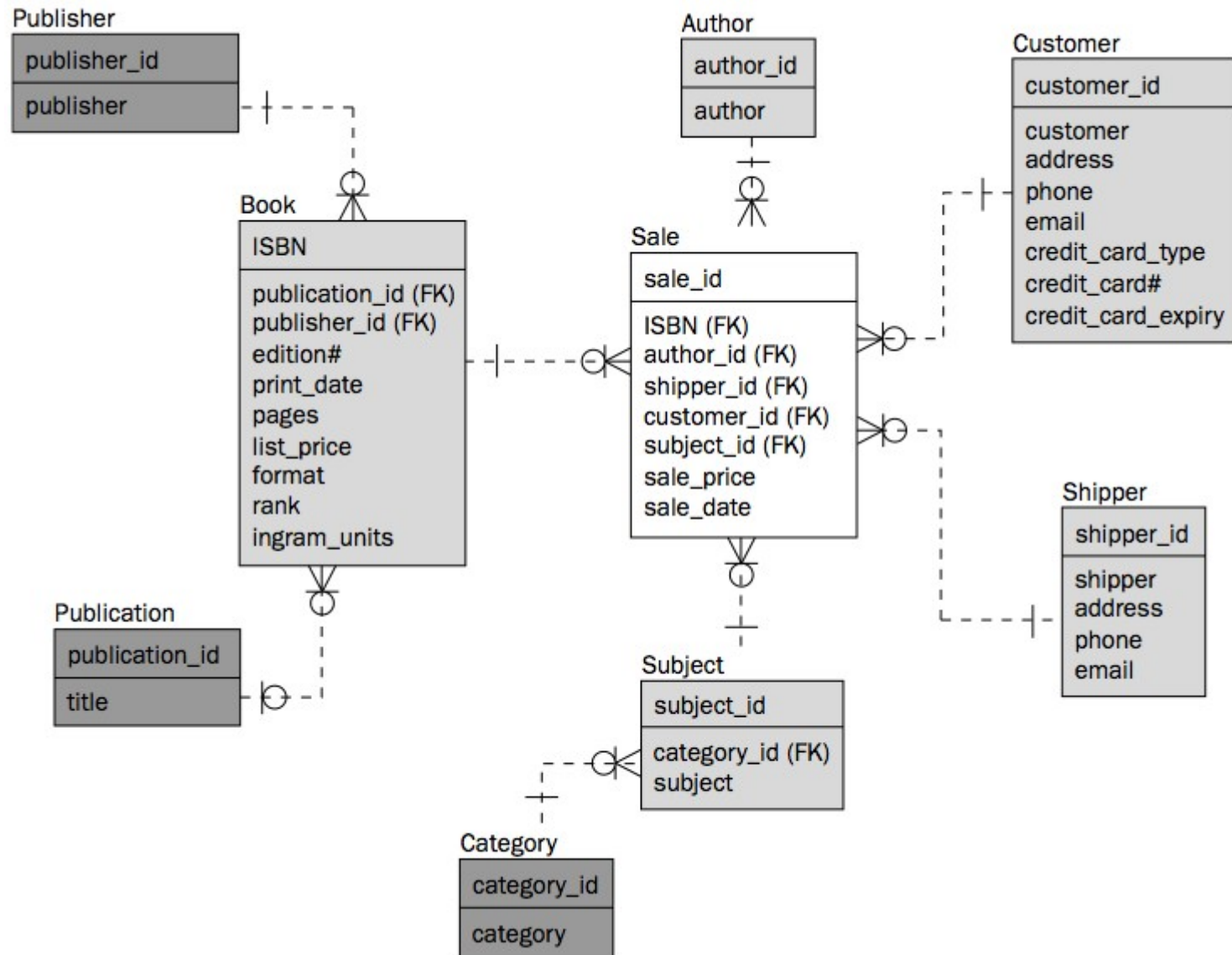
Surrogate Key dan Constraint di DW

- Surrogate key pada DW model mungkin akan semakin banyak digunakan.
- Referential integrity masih tetap dibutuhkan dalam pemodelan DW.

Skema Star



Skema Snowflake



Star atau Snowflake

- Perbedaannya hanya di penerapan fisiknya.
- Skema snowflake mendukung pemeliharaan dimensi yang lebih mudah.
- Star lebih mudah diakses pemakai dan mendukung query yang lebih sederhana dan efisien.
- Lebih baik membuat versi star dari snowflake untuk mempresentasikan kepada pemakai.

Tabel Dimensi

- Membungkus atribut yang terasosiasi dengan fact dan memisahkan atribut-atribut ini ke dalam kelompok berbeda secara logika
 - Contoh: waktu, geografi, produk, konsumen.
- Sebuah tabel dimensi dapat digunakan di beberapa tabel fact (skema).
 - Disebut *conforming dimension*.
- Jika atribut dimensi sering di update, maka lebih baik dibuat dimensi snowflake.

Tabel Dimensi

- Data dalam dimensi biasanya hirarki.
- Hirarki ditentukan oleh kebutuhan bisnis.
- Contoh:
 - Dimensi waktu berisi hirarki: (all time), Tahun, Kuartal, Bulan, hari
- OLAP tool tergantung pada hirarki untuk mengelompokkan data.

Dimensi Waktu dan Tanggal

- Granularity Waktu
 - Sebuah dimensi tanggal dengan satu record per hari sudah cukup jika tidak dibutuhkan rincian yang lebih detil (granularity fine) dari 1 hari.
 - Tabel dimensi waktu terpisah harus dibentuk jika detil waktu diperlukan, contoh: menit/detik.
 - Antara tanggal dan waktu lebih baik dipisahkan dalam dimensi yang berbeda.

Latihan!

Skenario

- Pabrik roti “Sarapan” adalah perusahaan pembuatan roti yang berlokasi di Sleman Yogyakarta.
- Produk-produk rotinya dijual di seluruh kabupaten di Yogyakarta.
- Pabrik ini memiliki unit penjualan di kota Wonosari, Kota Yogyakarta, dan Sleman.
- Direksi ingin informasi penjualan.

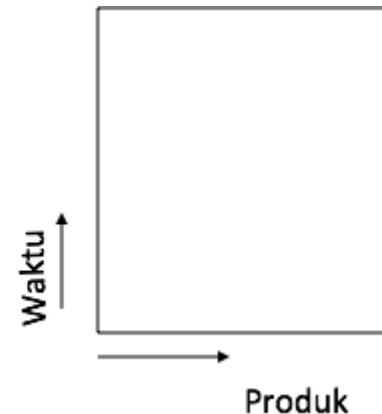
Informasi Penjualan

- Jumlah penjualan dari waktu ke waktu.

Januari	Februari	Maret	April
14	41	33	25

- Jumlah item yang dijual untuk setiap produk dan waktu.

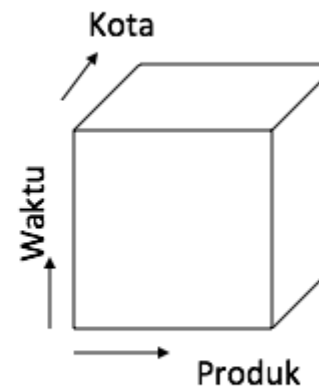
	Jan	Feb	Mar	Apr
Roti Tawar			6	17
Kue Keju	6	16	6	8
Swiss Rolls	8	25	21	



Informasi Penjualan

- Jumlah penjualan per item di tiap kota untuk setiap waktunya.

		Jan	Feb	Mar	Apr
Sleman	Roti Tawar			3	10
	Roti Keju	3	16	6	
	Swiss Rolls	4	16	6	
Wonosari	Roti Tawar			3	7
	Roti Keju	3			8
	Swiss Rolls	4	9	15	



Informasi Penjualan

- Jumlah penjualan item dan pendapatan tiap area untuk setiap item produk dan waktu.

		Jan		Feb		Mar		Apr	
		Rp	U	Rp	U	Rp	U	Rp	U
Sleman	Roti Tawar					7.44	3	24.80	10
	Roti Keju	7.95	3	42.40	16	15.90	6		
	Swiss Rolls	7.32	4	29.98	16	10.98	6		
Wonosari	Roti Tawar					7.44	3	17.36	7
	Roti Keju	7.95	3					21.20	8
	Swiss Rolls	7.32	4	16.47	9	27.45	15		

Ukuran Penjualan dan Dimensi

- Ukuran
 - Unit terjual
 - Jumlah
- Dimensi
 - Produk
 - Waktu
 - Area

Model Dimensi

