PERIODIC & ACCUMULATING SNAPSHOT

Topik

- Jenis Measures
- Jenis Fact Table

Fact Table

- Fact table → menyimpan nilai yang akan dianalisa
- Dimension table → menyimpan data deskriptif terkait nilai pada fact table
- Fact table → primary key + foreign keys + measures
- Measures: nilai numerik di dalam baris fakta yang bisa dihitung
- Data text umumnya tidak disimpan dalam fact table → dipindah ke dimension table, dikonversi ke numerik, atau dihapus

Additive Measure

- Measure yang dapat dijumlah berdasarkan **seluruh** dimension pada fact table

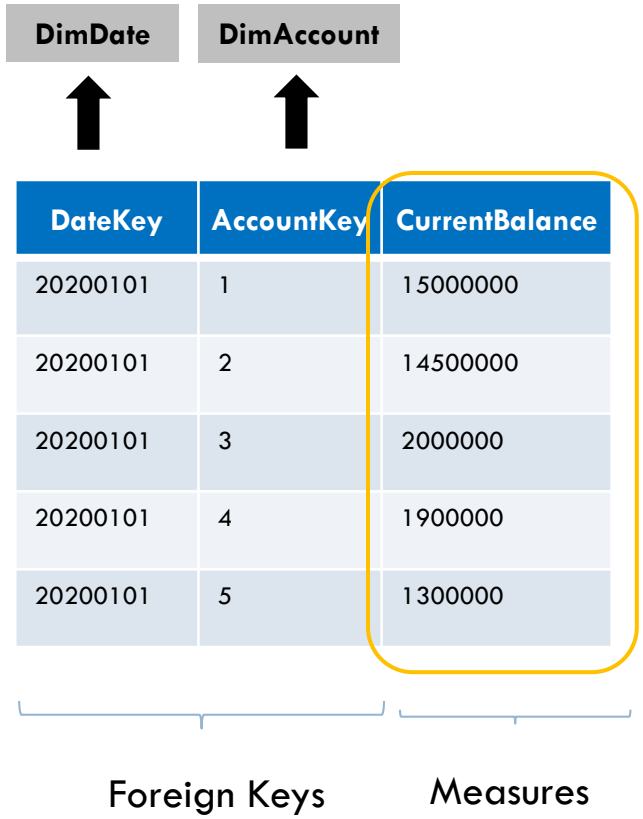
OrderDetail Key	CustomerKey	OrderDate Key	CurrencyKey	ProductKey	PromotionKey	Quantity	Total
1	1	20200101	1	201	1	2	100000
2	1	20200101	1	305	1	1	50000
3	2	20200105	1	11	1	3	10000
4	2	20200105	2	200	2	4	25000
5	2	20200105	2	7	3	1	2000000

Foreign Keys

Measures

Semi-additive Measure

- Measure yang dapat dijumlah berdasarkan **beberapa** (tidak semua) dimension pada fact table



DateKey	Account Key	Current Balance
20200101	1	15000000
20200101	2	14500000
20200101	3	2000000
20200101	4	1900000
20200101	5	1300000
20200102	1	20000000
20200102	2	12000000
20200102	3	1500000
20200102	4	1000000
20200102	5	1000000
.....		
.....		
20200131	1	30000000
20200131	2	1000000
20200131	3	1500000
20200131	4	1000000
20200131	5	1400000

Current Balance untuk seluruh account
= CB Acc1 + CB Acc2 + CB Acc3 + … + CB Acc(n)

Current Balance additive terhadap dimensi Account

Current Balance bulan Januari

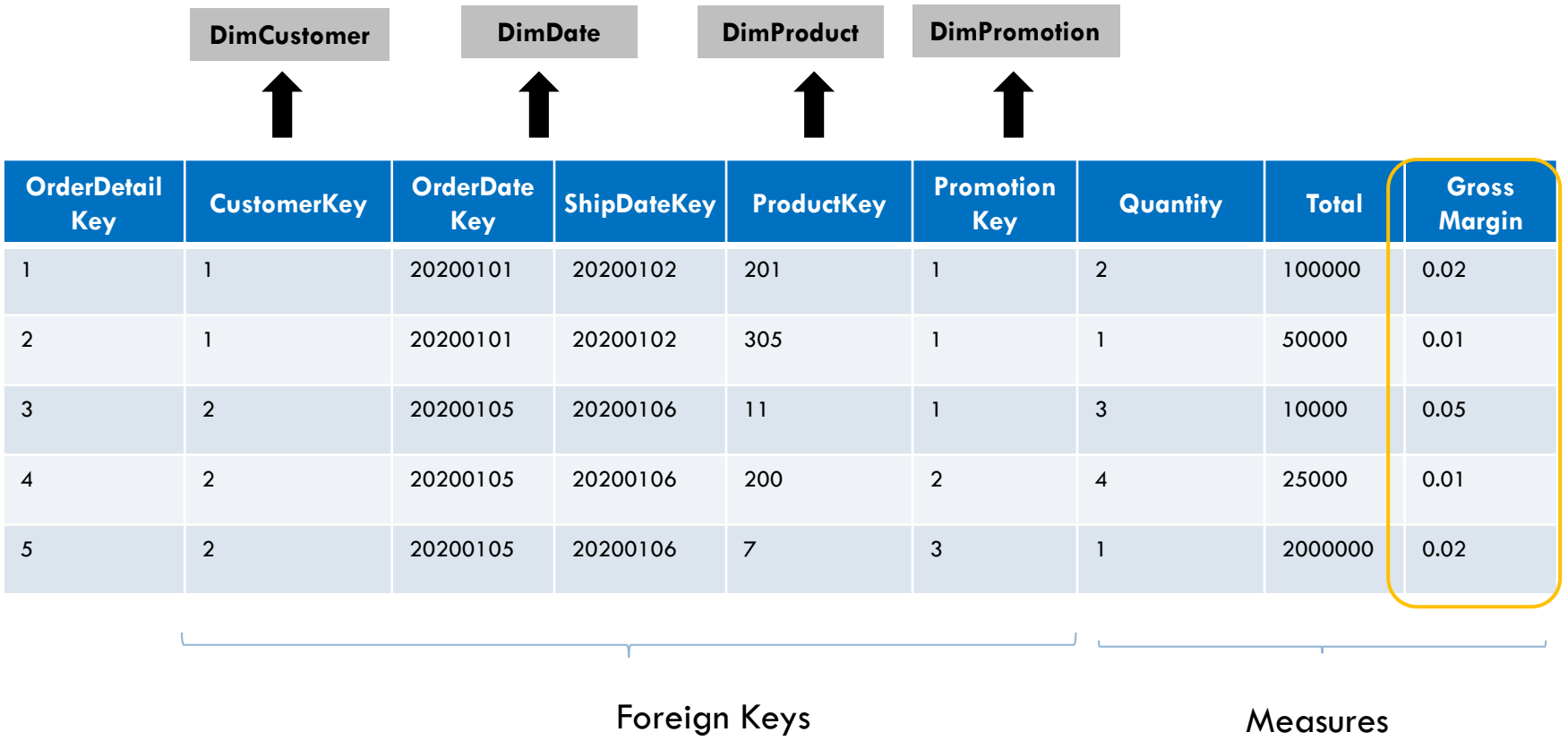
≠

CB tgl1 + CB tgl2 + CB tgl3 + … + CB tgl31

Current Balance tidak additive terhadap dimensi Date

Non-additive Measure

- Measure yang **tidak** dapat dijumlah berdasarkan **seluruh** dimension pada fact table



OrderDetail Key	CustomerKey	OrderDate Key	ShipDateKey	ProductKey	Promotion Key	Quantity	Total	Gross Margin
1	1	20200101	20200102	201	1	2	100000	0.02
2	1	20200101	20200102	305	1	1	50000	0.01
3	2	20200105	20200106	11	1	3	10000	0.05
4	2	20200105	20200106	200	2	4	25000	0.01
5	2	20200105	20200106	7	3	1	2000000	0.02

Null pada Fact Table

- Measure bernilai null tidak berdampak apapun terhadap hasil analisa → aggregate function dapat menghandle null dengan baik
- Sebisa mungkin, hindari null pada foreign key
  - Merusak hasil agregasi by dimension table → menurunkan kualitas data
  - Buat baris khsusus pada dimension table untuk mengasosiasikan nilai Unknown/Not Applicable

Jenis Fact Table

Kimball Methodology

- 4 Step Dimensional Design Process:
  - Select the business process
  - Declare the grain
  - Identify the dimensions
  - Identify the facts

Granularity

- Granularity adalah tingkat detail atau kehalusan data yang dicatat
- Semakin tinggi granularity, semakin rinci data yang tercatat
- Granularity tinggi
  - Mencatat setiap transaksi individual atau peristiwa
  - Volume data besar
  - Kedalaman analisis tinggi
- Granularity rendah
  - Agregat dari beberapa transaksi atau peristiwa
  - Volume data lebih kecil
  - Kedalaman analisis lebih rendah

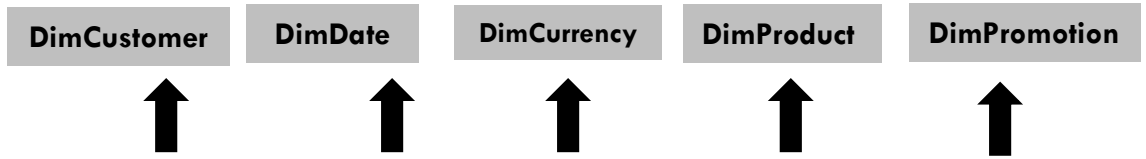
Jenis Fact Table

- Berdasarkan granularity:
  1. Transactional
  2. Periodic snapshots
  3. Accumulating snapshots

Transactional Fact Table

- Setiap row pada fact table berelasi dengan setiap peristiwa/transaksi
- Granularity → atomic transaction
- Highly dimensional
- Expressive, robust → enable maximum slicing & dicing

Transactional Fact Table



OrderDetail Key	CustomerKey	OrderDate Key	CurrencyKey	ProductKey	PromotionKey	Quantity	Total
1	1	20200101	1	201	1	2	100000
2	1	20200101	1	305	1	1	50000
3	2	20200105	1	11	1	3	10000
4	2	20200105	2	200	2	4	25000
5	2	20200105	2	7	3	1	2000000

Periodic Snapshot

- Setiap row pada fact table menyimpan data event pada periode waktu tertentu, misalnya hari, minggu, bulan, dst
- Granularity → agregat transaksi dalam periode waktu tertentu
- Tujuan: analisis tren/laporan berkala karena query bisa lebih cepat

Periodic Snapshot

Periodic Snapshot

Accumulating Snapshot

- Menyimpan status dari proses yang berlangsung dari awal hingga akhir
- Dapat diupdate seiring proses berlangsung
- Tidak periodik, tapi mengikuti lifecycle suatu proses
- Data diperbarui saat tahapan proses berubah
- Contoh:
  - Proses pengiriman
  - Tahapan klaim asuransi
  - Pemrosesan pesanan

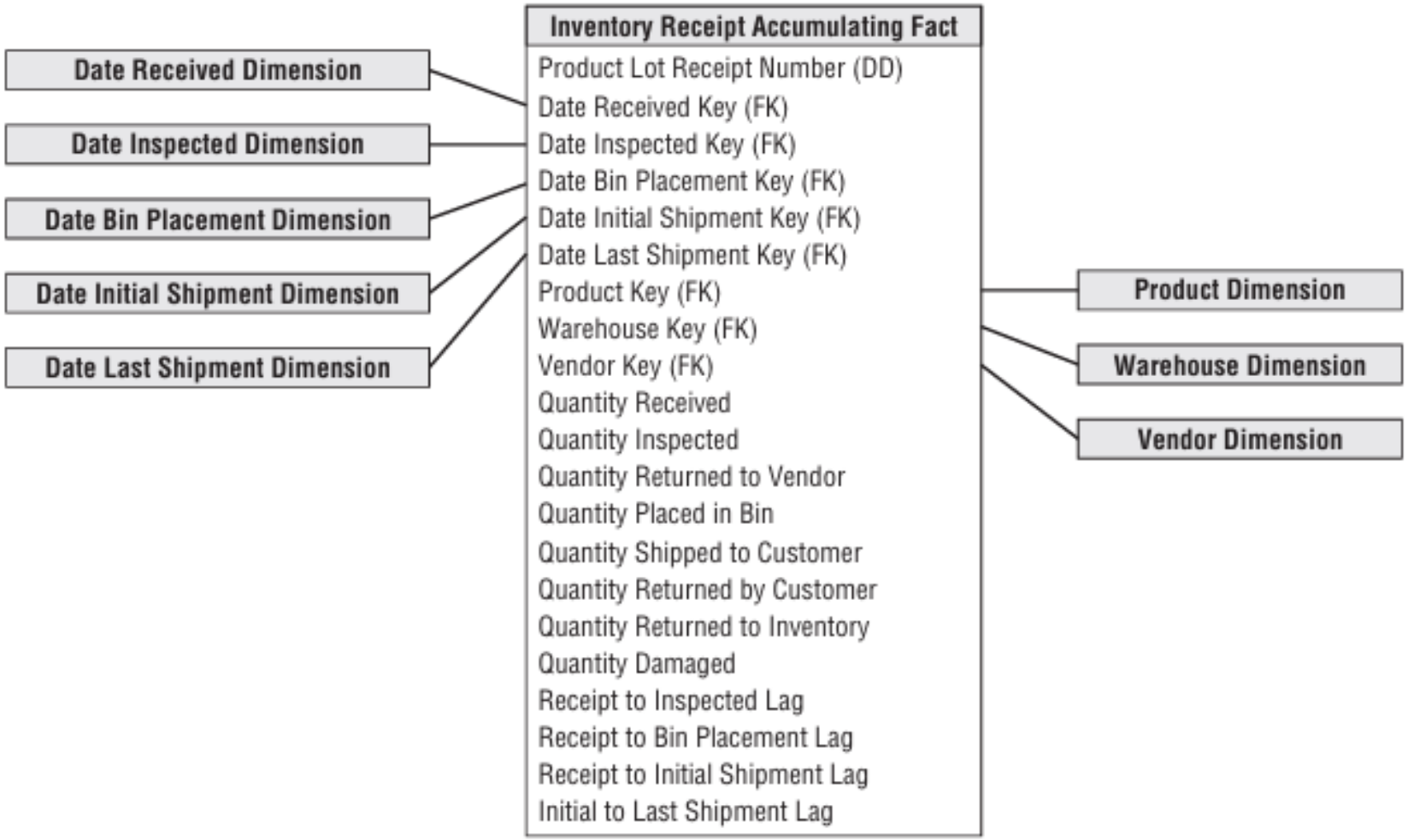
Transactional Fact Table

ClaimID	ProductID	Transaction Date	Status	Claim Amount
1001	1	20240115	Received	500000
1001	1	20240115	Verified	500000
1001	1	20240118	Evaluated	500000
1001	1	20240120	Settled	500000
1002	2	20240120	Received	200000
1002	2	20240120	Verified	200000
1003	3	20240201	Received	750000
1003	3	20240201	Verified	750000
1003	3	20240205	Evaluated	750000
1003	3	20240206	Settled	750000

Accumulating Snapshot

ClaimID	ProductID	Claim Date	Verified Date	Evaluated Date	Settled Date	Claim Amount	Status
1001	1	20241501	20241501	20241801	20242001	500000	Settled
1002	2	20242001	20242001	0	0	200000	Evaluated
1003	3	20240102	20240102	20240502	20240602	750000	Received

Accumulating Snapshot



Fact Table

- Transactional → melacak setiap transaksi atomic (individual) dengan detail
- Periodic Snapshot → ringkasan data pada titik waktu tertentu, misalnya total penjualan atau stok barang
- Accumulating Snapshot → melacak progres atau status dari suatu peristiwa yang melewati beberapa tahapan, seperti status pesanan atau status proyek
- Periodic dan accumulating snapshot sebaiknya dijadikan pelengkap, bukan pengganti transactional fact table

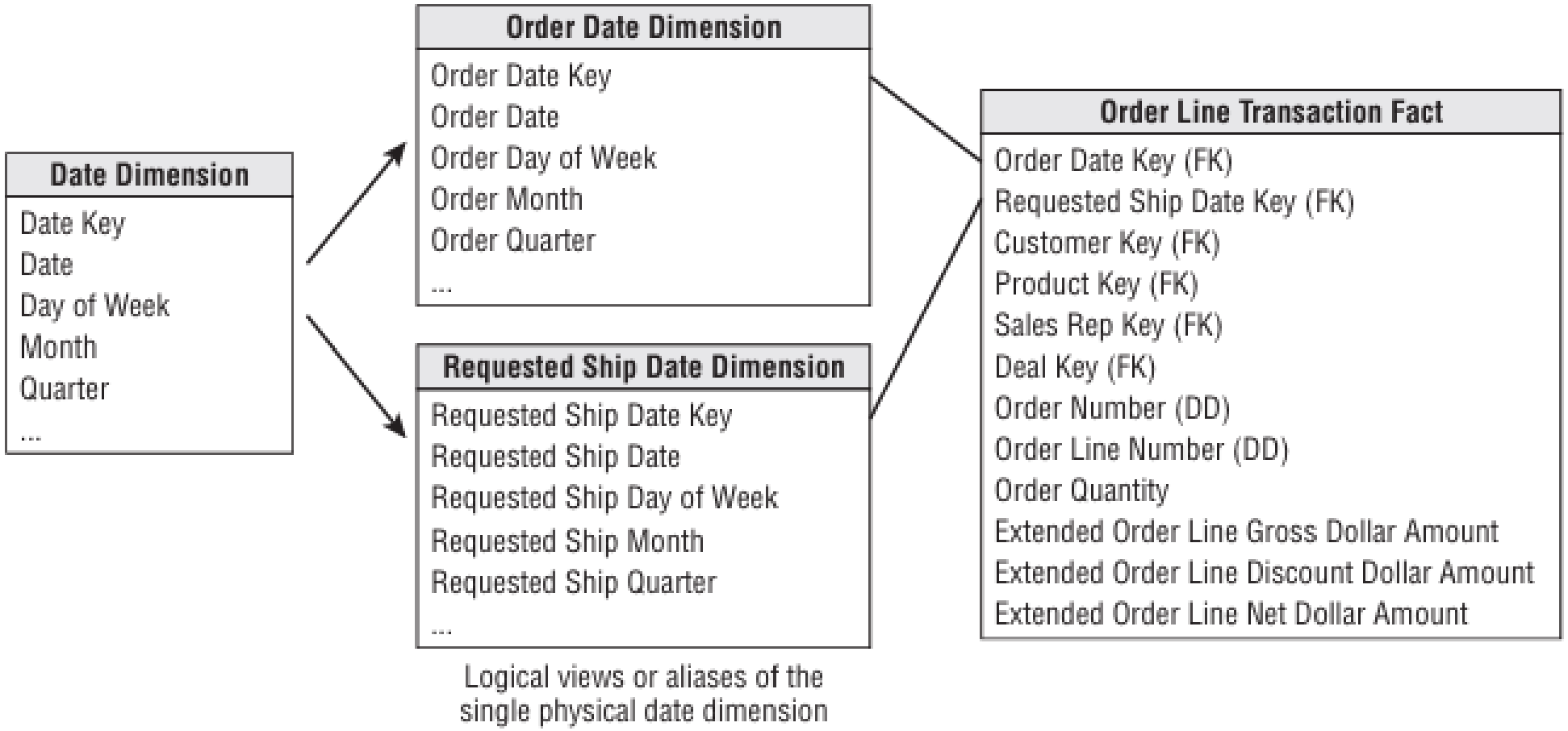
Bus Matrix

	Fact Type	Date	Position	Employee	Organization	Benefit
Hiring Processes						
Employee Position Snapshot	Periodic	X	X	Empl Mgr	X	
Employee Requisition Pipeline	Accumulating	X	X	Empl Mgr	X	
Employee Hiring	Transaction	X	X	Empl Mgr	X	
Employee "On Board" Pipeline	Accumulating	X	X	Empl Mgr	X	
Benefits Processes						
Employee Benefits Eligibility	Periodic	X		X	X	X
Employee Benefits Application	Accumulating	X		X	X	X
Employee Benefit Participation	Periodic	X		X	X	X
Employee Management Processes						
Employee Headcount Snapshot	Periodic	X		X	X	X
Employee Compensation	Transaction	X		X	X	X
Employee Benefit Accruals	Transaction	X		X	X	X
Employee Performance Review Pipeline	Accumulating	X		Empl Mgr	X	X
Employee Performance Review	Transaction	X		Empl Mgr	X	X
Employee Prof Dev Completed Courses	Transaction	X		X	X	
Employee Disciplinary Action Pipeline	Accumulating	X		Empl Mgr	X	
Employee Separations	Transaction	X		Empl Mgr	X	

Role-Playing Dimension

- Role-playing dimension adalah dimensi yang sama yang digunakan dalam beberapa konteks atau peran berbeda dalam data warehouse
- 1 physical dimension table di-refer berulang kali oleh fact table. Setiap reference merepresentasikan logical role yang berbeda
- Manfaat:
  - Mengurangi redundansi
  - Meningkatkan efisiensi
  - Memudahkan maintenance

Role-Playing Dimension



Sumber

- The Data Warehouse Toolkit 3rd ed, Ralph Kimball & Margy Ross
- Exam 70-463: Implementing a Data Warehouse with Microsoft SQL Server 2012

Tugas

- Buatlah masing-masing 1 contoh periodic dan accumulating snapshot fact table
- Note: proses bisnis bisa sama atau berbeda
- Buat lengkap dengan dimension tablenya