



Pertemuan #12

Desain Data Storage



Pengertian

- **Fungsi data storage (penyimpanan data)**

- ☐ mengelola bagaimana data disimpan dan ditangani oleh program yang menjalankan sistem

- **Tujuan desain penyimpanan data**

- ☐ Pengambilan data yang efisien (waktu respons yang cepat)
- ☐ Akses ke informasi yang dibutuhkan pengguna

Format Data Storage

- **File:** daftar data elektronik yang dioptimalkan untuk melakukan transaksi tertentu
- **Database (Basis Data):** kumpulan pengelompokan informasi yang saling berhubungan
- **Database Management System (DBMS)** adalah perangkat lunak yang membuat dan memanipulasi basis data

Contoh File Appointment

Appointment Date	Appointment Time	Duration	Reason	Patient ID	First Name	Last Name	Phone Number	Doctor ID	Doctor Last Name
11/23/2003	2:30	.25 hour	Flu	758843	Patrick	Dennis	548-9456	V524625587	Vroman
11/23/2003	2:30	1 hour	Physical	136136	Adelaide	Kin	548-7887	T445756225	Tantalo
11/23/2003	2:45	.25 hour	Shot	544822	Chris	Pullig	525-5464	V524625587	Vroman
11/23/2003	3:00	1 hour	Physical	345344	Felicia	Marston	548-9333	B544742245	Brousseau
11/23/2003	3:00	.5 hour	Migraine	236454	Thomas	Bateman	667-8955	V524625587	Vroman
11/23/2003	3:30	.5 hour	Muscular	887777	Ryan	Nelson	525-4772	V524625587	Vroman
11/23/2003	3:30	.25 hour	Shot	966233	Peter	Todd	667-2325	T445756225	Tantalo
11/23/2003	3:45	.75 hour	Muscular	951657	Mike	Morris	663-8944	T445756225	Tantalo
11/23/2003	4:00	1 hour	Physical	223238	Ellen	Whitener	525-8874	B544742245	Brousseau
11/23/2003	4:00	.5 hour	Flu	365548	Jerry	Starsia	548-9887	V524625587	Vroman
11/23/2003	4:30	1 hour	Minor surg	398633	Susan	Perry	525-6632	V524625587	Vroman
11/23/2003	4:30	.5 hour	Migraine	222577	Elizabeth	Gray	667-8400	T445756225	Tantalo
11/24/2003	8:30	.25 hour	Shot	858756	Elias	Awad	663-6364	T445756225	Tantalo
11/24/2003	8:30	1 hour	Minor surg	232158	Andy	Ruppel	525-9888	V524625587	Vroman
11/24/2003	8:30	.25 hour	Flu	244875	Rick	Grenci	548-2114	B544742245	Brousseau
11/24/2003	8:45	.5 hour	Muscular	655683	Eric	Meier	667-0254	T445756225	Tantalo
11/24/2003	8:45	1 hour	Physical	447521	Jane	Pace	548-0025	B544742245	Brousseau
11/24/2003	9:30	.5 hour	Flu	554263	Trey	Maxham	663-8547	V524625587	Vroman

Contoh Database Appointment

Appointment Date	Appointment Time	Duration	Reason	Patient ID	Doctor ID
11/23/2003	2:30	.5 hour	Flu	758843	V524625587
11/23/2003	2:30	1 hour	Physical	136136	T445756225
11/23/2003	2:45	.25 hour	Shot	544822	V524625587
11/23/2003	3:00	1 hour	Physical	345344	B544742245
11/23/2003	3:00	.5 hour	Migraine	236454	V524625587
11/23/2003	3:30	.5 hour	Muscular	887777	V524625587
11/23/2003	3:30	.25 hour	Shot	966233	T445756225
11/23/2003	3:45	.75 hour	Muscular	951657	T445756225
11/23/2003	4:00	1 hour	Physical	223238	B544742245
11/23/2003	4:00	.5 hour	Flu	365548	V524625587
11/23/2003	4:30	1 hour	Minor surg	398633	V524625587
11/23/2003	4:30	.5 hour	Migraine	222577	T445756225
11/24/2003	8:30	.25 hour	Shot	858756	T445756225
11/24/2003	8:30	1 hour	Minor surg	232158	V524625587
11/24/2003	8:30	.25 hour	Flu	244875	B544742245
11/24/2003	8:45	.5 hour	Muscular	655683	T445756225
11/24/2003	8:45	1 hour	Physical	447521	B544742245
11/24/2003	9:30	.5 hour	Flu	554263	V524625587

Tables related using patient id

Tables related using doctor id

Patient ID	First Name	Last Name	Phone Number
136136	Adelaide	Kin	548-7887
222577	Elizabeth	Gray	667-8400
223238	Ellen	Whitener	525-8874
232158	Andy	Ruppel	525-9888
236454	Thomas	Bateman	667-8955
244875	Rick	Grenci	548-2114
345344	Felicia	Marston	548-9333
365548	Jerry	Starsia	548-9887
398633	Susan	Perry	525-6632
447521	Jane	Pace	548-0025
544822	Chris	Pullig	525-5464
554263	Trey	Maxham	663-8547
655683	Eric	Meier	667-0254
758843	Patrick	Dennis	548-9456
858756	Elias	Awad	663-6364
887777	Ryan	Nelson	525-4772
951657	Mike	Morris	663-8944
966233	Peter	Todd	667-2325

Doctor ID	Last Name
B544742245	Brousseau
T445756225	Tantalo
V524625587	Vroman



Attribut pada File

- File berisi informasi yang diformat untuk transaksi tertentu
- Biasanya diatur secara berurutan
- Pointer digunakan untuk mengaitkan satu record dengan record lainnya
- Linked List adalah file dengan record yang dihubungkan bersama menggunakan pointer



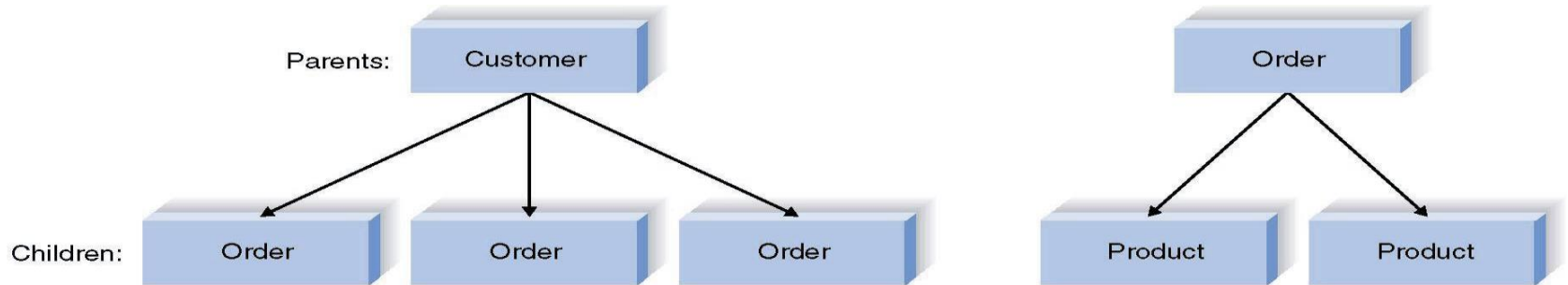
Type File

- File master - menyimpan data inti, informasi penting
- Look-up file - menyimpan nilai statis
- File transaksi - menyimpan informasi yang memperbarui file master
- File audit - merekam sebelum dan sesudah versi data
- File riwayat (arsip) - menyimpan informasi masa lalu

Type Database

- Legacy database (basis data lama)
 - Hierarchical (menggambarkan hubungan orangtua-anak menggunakan pohon terbalik)
 - Jaringan/*network* (menggambarkan asosiasi nonhierarkis menggunakan pointer)
- Database relasional
- Database objek
- Database multidimensi

Contoh Hierarchical Database



Notice how Order serves as a child to Customer and a parent to Product.

Sample Records:

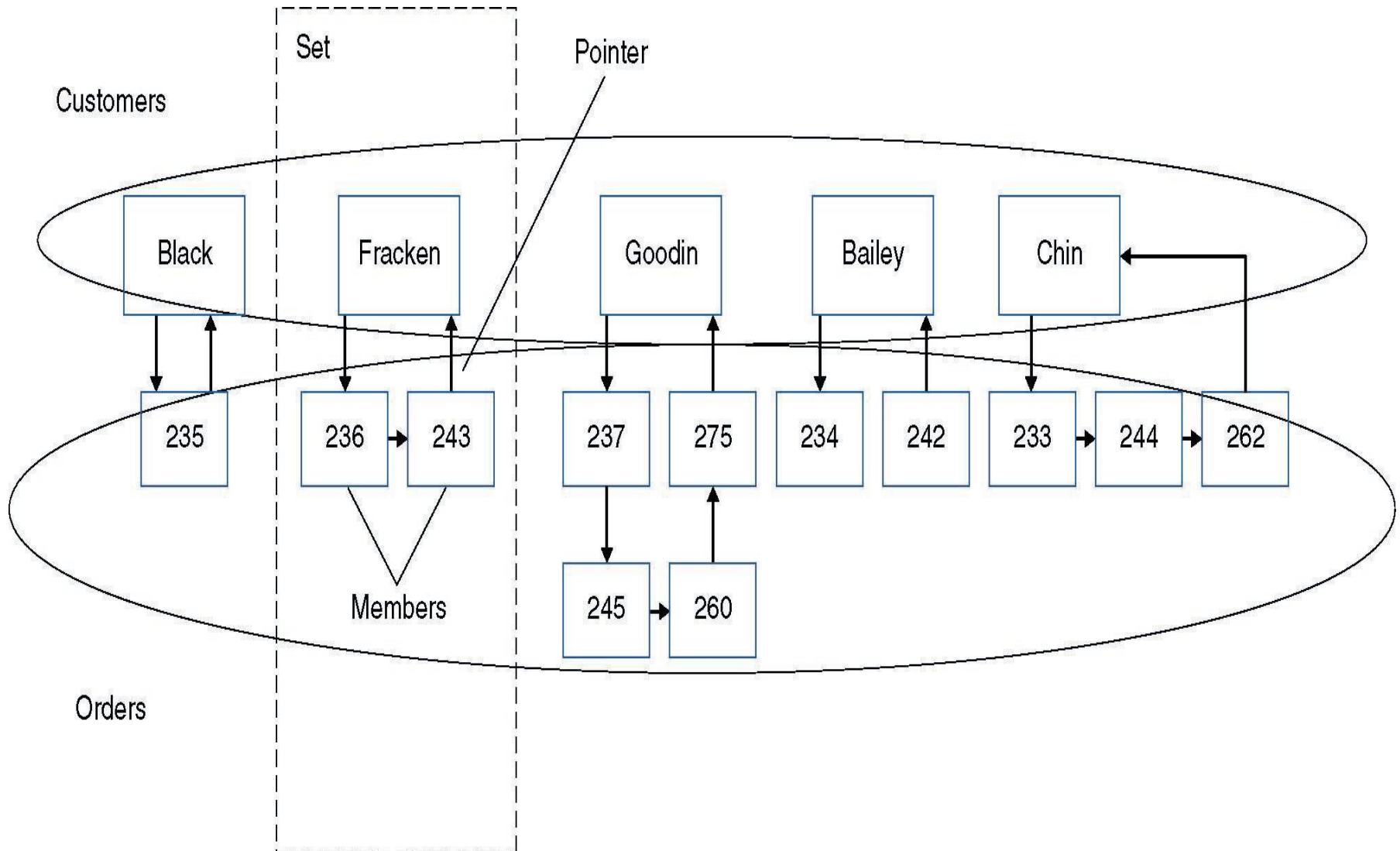
Customer as parent

1035 Black	...
235	11/23/00 ...
1556 Fracken	...
236	11/23/00 ...
243	11/26/00 ...
2274 Goodin	...
237	11/23/00 ...
245	11/26/00 ...
260	11/30/00 ...
275	12/7/00 ...
4254 Bailey	...
234	11/23/00 ...
242	11/26/00 ...
9500 Chin	...
233	11/23/00 ...
244	11/26/00 ...
262	11/30/00 ...

Order as parent

233	11/23/00 ...	444	Wine Gift Pack
222	Bottle Opener	555	Cheese Tray
234	11/23/00 ...	222	Bottle Opener
235	11/23/00 ...	555	Cheese Tray
222	Bottle Opener	236	11/23/00 ...
333	Jams & Jellies	222	Bottle Opener
237	11/23/00 ...	111	Wine Guide
242	11/26/00 ...	444	Wine Gift Pack
243	11/26/00 ...	333	Jams & Jellies
222	Bottle Opener	555	Cheese Tray

Contoh Network Database





Konsep Database Relational

- Populer, mudah digunakan untuk pengembangan
- Primary dan foreign key digunakan untuk mengidentifikasi dan menautkan tabel
- Integritas referensial memastikan sinkronisasi tabel yang benar dan valid
- Structured Query Language (SQL) - bahasa standar untuk mengakses data

Contoh Database Relational

Cust ID is the primary
key of Customer

Customer

Cust ID	Last Name	First Name
4254	Baily	Ryan
9500	Chin	April
1556	Fracken	Chris
2487	Hancock	Bill
2243	Harris	Linda
1035	Black	John
1123	Williams	Mary
9501	Kaplan	Bruce
4453	Min	Julie
9505	Marvin	Sandra
2282	Lau	Mark
5927	Lee	Diane
2241	Jones	Chris
2242	DeBerry	Ann
2274	Goodin	Dan
9507	Nelson	Dave
2264	White	Anthony

Order

Order Number	Date	Cust ID	Amount	Payment Type
234	11/23/00	4254	\$30.00	MC
235	11/23/00	9500	\$20.00	VISA
236	11/23/00	1556	\$20.00	VISA
237	11/23/00	2487	\$60.00	AMEX
238	11/23/00	2243	\$50.00	MC
239	11/23/00	1035	\$50.00	AMEX
240	11/23/00	1556	\$20.00	VISA
241	11/23/00	1123	\$40.00	MC
242	11/24/00	9501	\$30.00	VISA
243	11/24/00	4453	\$30.00	VISA
244	11/24/00	9505	\$20.00	VISA
245	11/24/00	2282	\$20.00	AMEX
246	11/24/00	5927	\$60.00	MC
247	11/24/00	2241	\$50.00	VISA
248	11/24/00	4254	\$50.00	AMEX
249	11/24/00	2242	\$50.00	AMEX
250	11/24/00	2274	\$20.00	MC
251	11/24/00	9507	\$10.00	MC
252	11/24/00	2487	\$60.00	MC
253	11/24/00	2264	\$40.00	AMEX

Cust ID is a foreign key in Order

Payment Type is a
foreign key in Order

Referential integrity:

- All Payment Type values in Order must exist first in the Payment Type table
- All Cust ID values in Order must exist first in the Customer table

Payment Type

Payment Type	Description
MC	Mastercard
VISA	VISA
AMEX	American Express

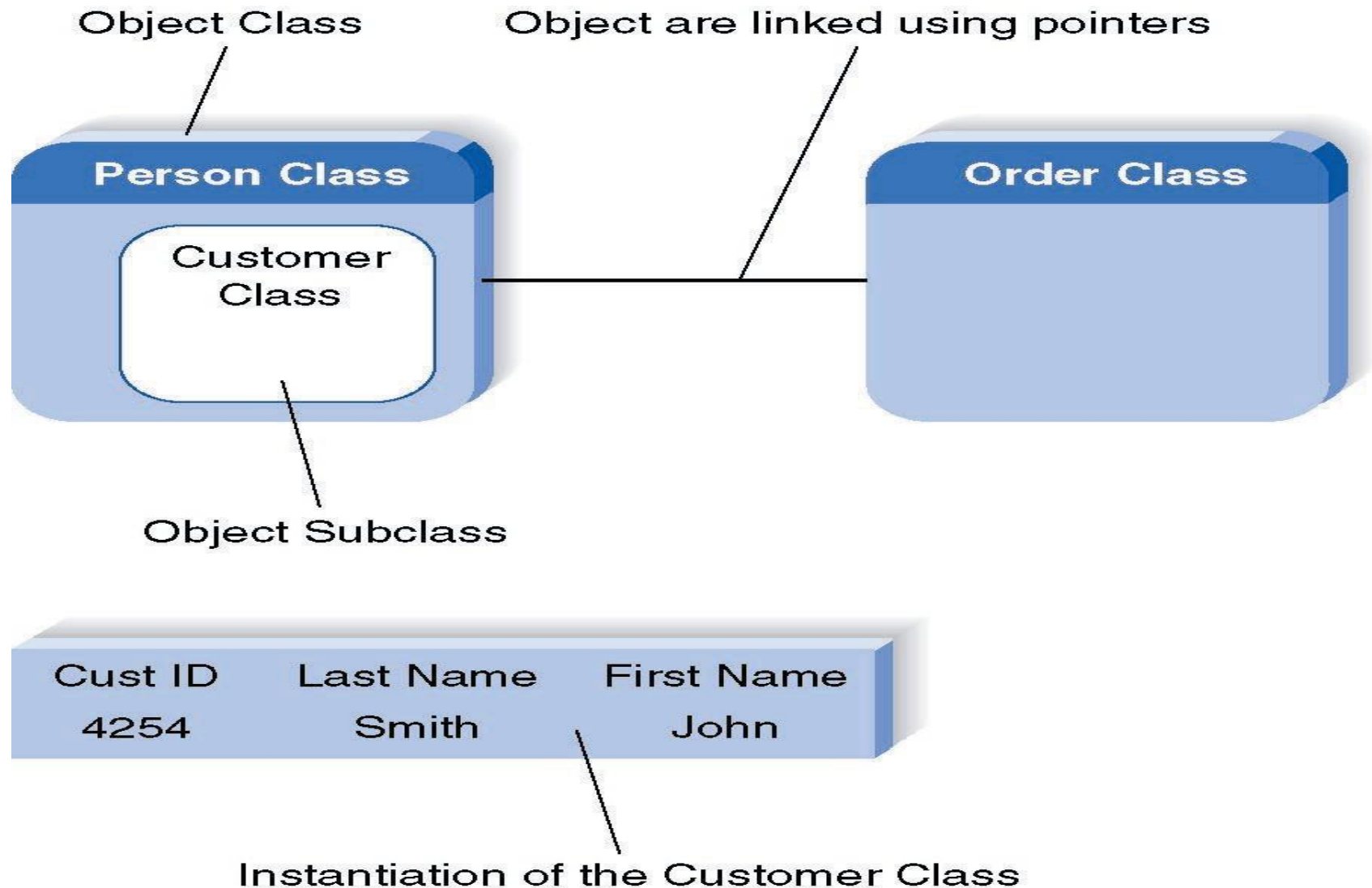
Payment Type is
the primary key of the
Payment Type table



Konsep Database Berbasis Object

- Dibangun di sekitar objek yang terdiri dari data dan proses
- Objek dienkapsulasi (mandiri)
- Kelas objek - kategori objek utama
- OODBMS - digunakan terutama untuk aplikasi dengan data multimedia atau data kompleks
- OODBMS hibrid - fitur objek dan relasional

Contoh Database Object

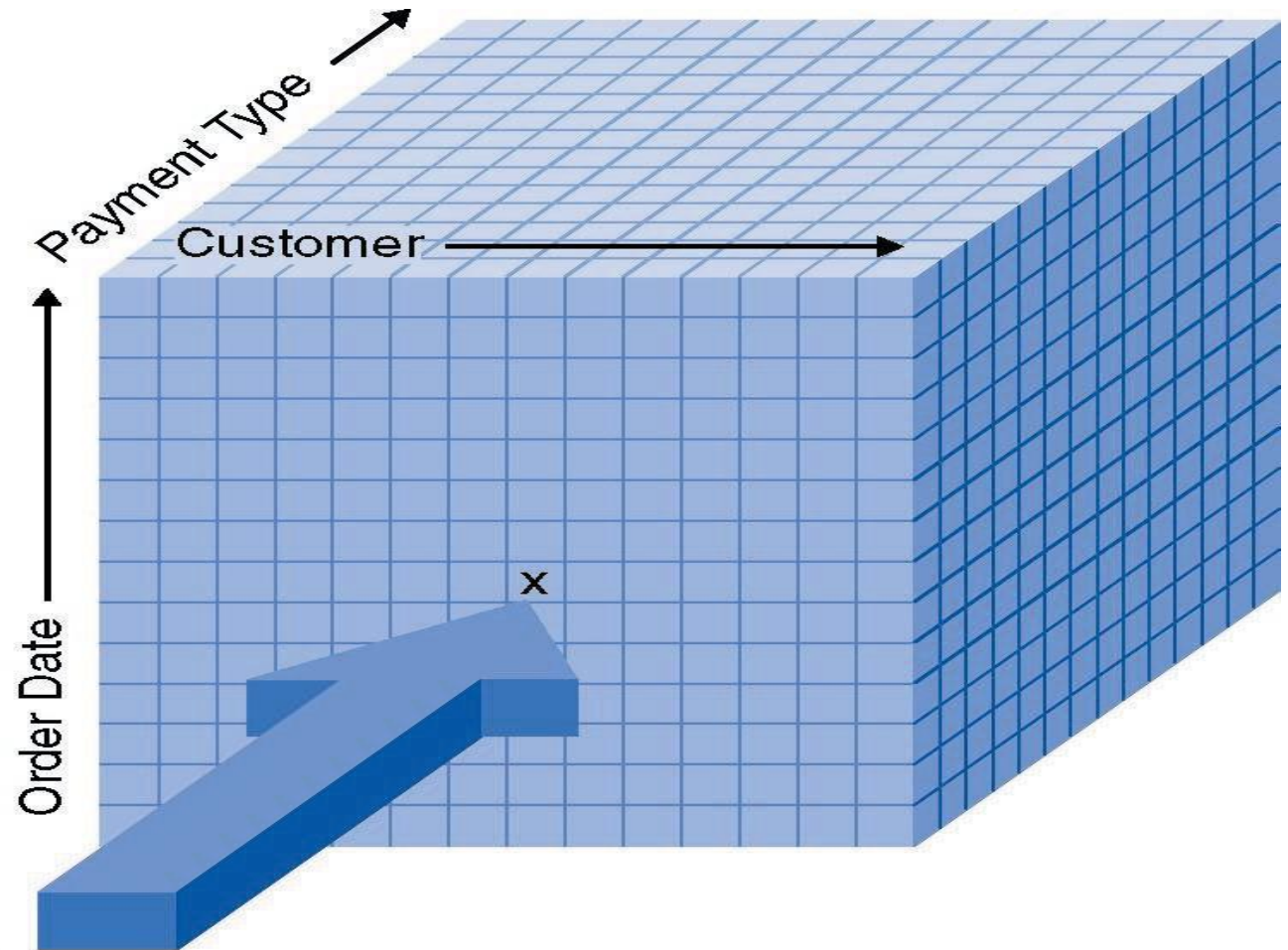




Konsep Database Multidimensi

- Menyimpan data untuk kemudahan agregasi dan manipulasi di banyak dimensi
- Digunakan untuk gudang data (data warehouse)
- Data ringkasan sudah dihitung sebelumnya dan disimpan untuk kecepatan akses

Contoh Database Multidimensi



Last quarter, how many customers placed more than one order using an American Express card?

Optimalisasi Data Storage

- 2 Tujuan (tetapi saling bertentangan):
 - **Efisiensi penyimpanan** (meminimalkan ruang penyimpanan)
 - **Kecepatan akses** (meminimalkan waktu untuk mengambil informasi yang diinginkan)
- **Efisiensi penyimpanan**
 - Minimalkan nilai null dan redundansi
 - Kurangi anomali pembaruan
 - Proses normalisasi mengoptimalkan desain penyimpanan data untuk efisiensi penyimpanan
- **Optimalisasi kecepatan akses**
 - Denormalisasi
 - Clustering (Intrafile dan Interfile)
 - Pengindeksan (indexing)

Efisiensi Penyimpanan

CUSTOMER ORDER

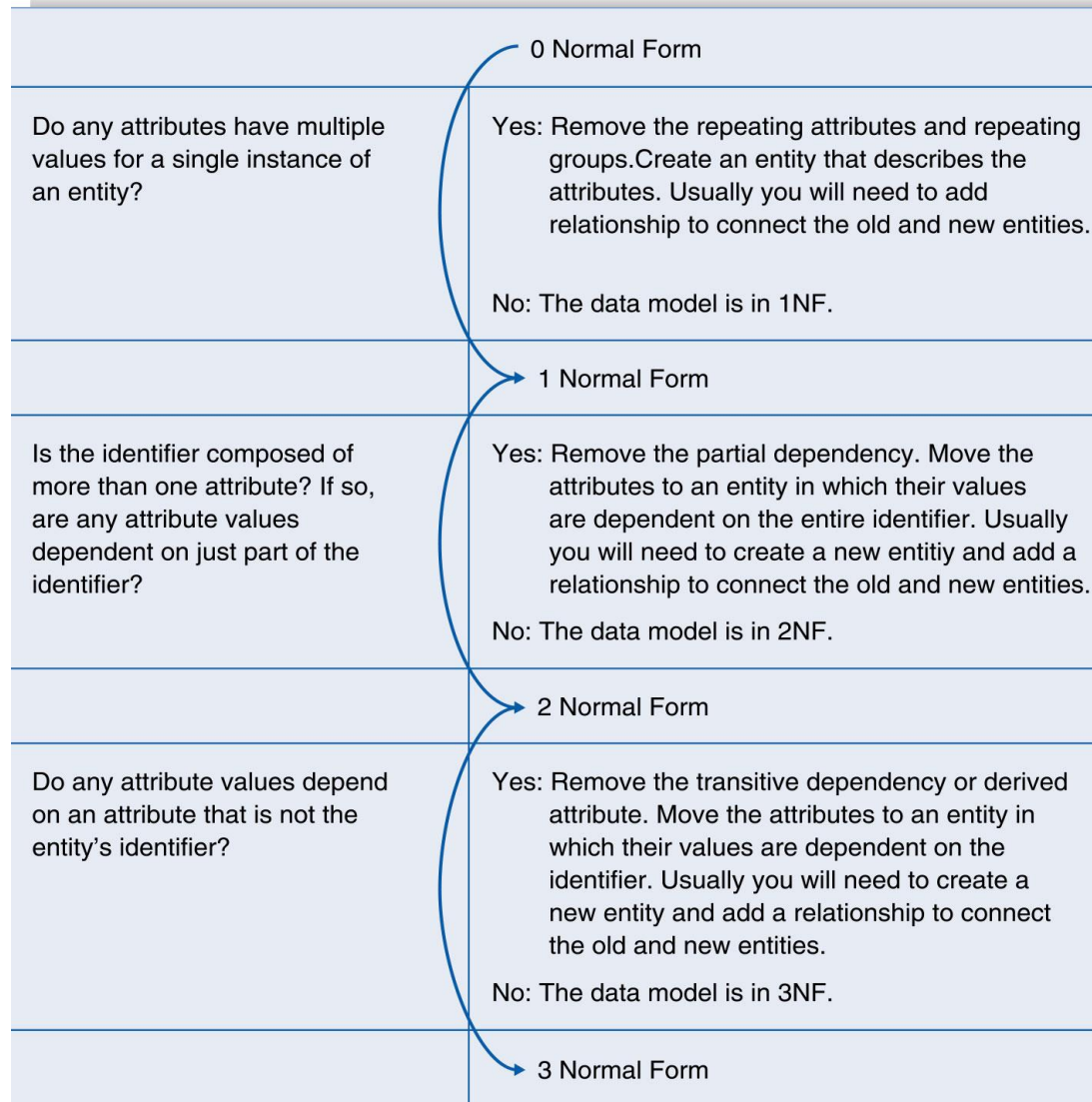
Order Number
Date
Cust ID
Last Name
First Name
State
Amount
Tax Rate
Product 1
Product Description 1
Product 2
Product Description 2
Product 3
Product Description 3

Redundant data

Null cells

Order Number	Date	Cust ID	Last Name	First Name	State	Amount	Tax Rate	Product	Product Desc	Product	Product Desc	Product	Product Desc
239	11/23/00	1135	Black	John	MD	\$50.00	0.05	555	Cheese Tray				
260	11/24/00	1135	Black	John	MD	\$40.00	0.05	444	Wine Gift Pack				
273	11/27/00	1135	Black	John	MD	\$20.00	0.05	222	Bottle Opener				
241	11/23/00	1123	Williams	Mary	CA	\$40.00	0.08	444	Wine Gift Pack				
262	11/24/00	1123	Williams	Mary	CA	\$20.00	0.08	222	Bottle Opener				
287	11/27/00	1123	Williams	Mary	CA	\$20.00	0.08	222	Bottle Opener				
290	11/30/00	1123	Williams	Mary	CA	\$50.00	0.08	555	Cheese Tray				
234	11/23/00	2242	DeBerry	Ann	DC	\$50.00	0.065	555	Cheese Tray				
237	11/7/00	2242	DeBerry	Ann	DC	\$50.00	0.065	111	Wine Guide	444	Wine Gift Pack		
238	11/10/00	2242	DeBerry	Ann	DC	\$40.00	0.065	444	Wine Gift Pack				
245	11/11/00	2242	DeBerry	Ann	DC	\$20.00	0.065	222	Bottle Opener				
250	11/18/00	2242	DeBerry	Ann	DC	\$20.00	0.065	222	Bottle Opener				
252	11/22/00	2242	DeBerry	Ann	DC	\$60.00	0.065	222	Bottle Opener	444	Wine Gift Pack		
253	11/23/00	2242	DeBerry	Ann	DC	\$60.00	0.065	222	Bottle Opener	444	Wine Gift Pack		
297	11/24/00	2242	DeBerry	Ann	DC	\$30.00	0.065	333	Jams & Jellies				
243	11/11/00	4254	Bailey	Ryan	MD	\$50.00	0.05	555	Cheese Tray				
246	11/18/00	4254	Bailey	Ryan	MD	\$30.00	0.05	333	Jams & Jellies				
248	11/22/00	4254	Bailey	Ryan	MD	\$60.00	0.05	222	Bottle Opener	333	Jams & Jellies	111	Wine Guide
235	11/17/00	9500	Chin	April	KS	\$20.00	0.05	222	Bottle Opener				
242	11/23/00	9500	Chin	April	KS	\$30.00	0.05	333	Jams & Jellies				
244	11/24/00	9500	Chin	April	KS	\$20.00	0.05	222	Bottle Opener				
251	11/27/00	9500	Chin	April	KS	\$10.00	0.05	111	Wine Guide				

Normalisasi





Clustering

- Menempatkan record berdekatan secara fisik untuk data yang sering diakses sehingga mengurangi jumlah akses ke storage.
 - Pengelompokan Intrafile – record serupa dalam sebuah tabel disimpan bersama
 - Pengelompokan antarfile - menggabungkan record dari lebih dari satu tabel yang biasanya diambil bersama



Indexing

- Tabel kecil yang berisi nilai dari satu atau beberapa field dalam tabel dan lokasi nilai dalam tabel
- Mirip dengan indeks buku di perpustakaan

Contoh Indeks

PAYMENT TYPE INDEX

Payment Type	Pointer
AMEX	*
AMEX	*
AMEX	*
AMEX	*
AMEX	*
AMEX	*
MC	*
MC	*
MC	*
MC	*
MC	*
MC	*
MC	*
VISA	*
VISA	*
VISA	*
VISA	*
VISA	*
VISA	*
VISA	*

ORDER TABLE

Order Number	Date	Cust ID	Amount	Payment Type
234	11/23/00	4254	\$30.00	MC
235	11/23/00	9500	\$20.00	VISA
236	11/23/00	1556	\$20.00	VISA
237	11/23/00	2487	\$60.00	AMEX
238	11/23/00	2243	\$50.00	MC
239	11/23/00	1035	\$50.00	AMEX
240	11/23/00	1556	\$20.00	VISA
241	11/23/00	1123	\$40.00	MC
242	11/24/00	9501	\$30.00	VISA
243	11/24/00	4453	\$30.00	VISA
244	11/24/00	9505	\$20.00	VISA
245	11/24/00	2282	\$20.00	AMEX
246	11/24/00	5927	\$60.00	MC
247	11/24/00	2241	\$50.00	VISA
248	11/24/00	4254	\$50.00	AMEX
249	11/24/00	2242	\$50.00	AMEX
250	11/24/00	2274	\$20.00	VISA
251	11/24/00	9507	\$10.00	VISA
252	11/24/00	2487	\$60.00	VISA
253	11/24/00	2264	\$40.00	AMEX



Volumetrics – Estimasi Ukuran Data Storage

- Raw data - jumlah dari rata-rata lebar semua field dalam tabel
- Hitung persyaratan overhead berdasarkan rekomendasi vendor DBMS
- Perkirakan jumlah awal record
- Perkirakan tingkat pertumbuhan record

Contoh Estimasi Data Storage

Field	Average Size (Characters)
Order number	8
Date	7
Cust ID	4
Last name	13
First name	9
State	2
Amount	4
Tax rate	2
Record Size	49
Overhead (30%)	14.7
Total Record Size	63.7

Initial Table Size	50,000
Initial Table Volume	3,185,000
Growth/Month	1,000
Table volume @ 3 years	5,478,200