

Normalisasi Bag. 1

Sistem Basis Data
Cosmas Haryawan

1

Berapa tabel yang seharusnya kita buat?

The image shows two database tables. The first table, 'SKU_DATA', has columns: SKU, Description, Department, and Buyer. The second table, 'ORDR_ITEM', has columns: OrderNumber, SKU, Quantity, Price, and ExtendedPrice.

SKU	Description	Department	Buyer
100100	Std. Scuba Tank, Yellow	Water Sports	Pete Hansen
100200	Std. Scuba Tank, Magenta	Water Sports	Pete Hansen
101100	Dive Mask, Small Clear	Water Sports	Nancy Meyers
101200	Dive Mask, Med Clear	Water Sports	Nancy Meyers
201000	Half-dome Tent	Camping	Cindy Lo
202000	Half-dome Tent Footprint	Camping	Cindy Lo
301000	Light Fly Climbing Harness	Climbing	Jerry Martin
302000	Lodging cabinier, Oval	Climbing	Jerry Martin

OrderNumber	SKU	Quantity	Price	ExtendedPrice
3000	100200	1	300	300
2000	101100	4	50	200
3000	101100	2	50	100
2000	101200	2	50	100
3000	101200	1	50	50
1000	201000	1	300	300
1000	202000	1	130	130

2

Karakteristik Relasi (review)

- Tiap baris berisi data tentang sebuah entity
- Kolom-kolom berisi data tentang attribute dari suatu entity
- Satu sel pada tabel berisi satu data tunggal (atomic)
- Seluruh entry pada satu kolom memiliki domain yang sama
- Tiap kolom memiliki nama yang unik
- Urutan kolom tidak penting
- Urutan baris tidak penting
- Tidak ada 2 baris yang identik

3

Apakah ini Relasi ?

NIM	Nama	Mata Kuliah
001	Totok	Kalkulus, Aljabar Linier
002	Titik	Basis Data, Kalkulus
003	Andi	Basis Data, Kalkulus, Pemrograman Web

- TIDAK!
– Terdapat Cell Data yang tidak atomic/tunggal

4

Apakah Ini Relasi ?

IDM	Nama	Mata Kuliah
001	Totok	Kalkulus
		Aljabar Linier
002	Titik	Basis Data
		Kalkulus
003	Andi	Basis Data
		Kalkulus
		Pemrograman Web

- TIDAK!
 - Terdapat Primary Key yang NULL
 - terdapat Cell data yang NULL

5

Apakah Ini Relasi ?

EmployeeNumber	FirstName	LastName	Department	Email	Phone
100	Jerry	Johnson	Accounting	JJ@somewhere.com	236-9987
200	Mary	Abernathy	Finance	MA@somewhere.com	444-8898
300	Liz	Smathers	Finance	LS@somewhere.com	777-0098
400	Tom	Caruthers	Accounting	TC@somewhere.com	236-9987
500	Tom	Jackson	Production	TJ@somewhere.com	444-9980
600	Eleanore	Caldera	Legal	EC@somewhere.com	767-0900
700	Richard	Bandalone	Legal	RB@somewhere.com	767-0900

- YA!

6

Normalisasi

- Definisi
 - Normalisasi merupakan teknik analisis data yang mengorganisasikan atribut-atribut data dengan cara mengelompokkan sehingga terbentuk entitas yang non-redundant, stabil, dan fleksible.
- Keterangan
 - Normalisasi dilakukan sebagai uji coba pada suatu relasi secara berkelanjutan untuk menentukan apakah relasi itu sudah baik, yaitu dapat dilakukan proses insert, update, delete, dan modifikasi pada satu atau beberapa atribut tanpa mempengaruhi integritas data dalam relasi tersebut.

7

Mengapa Normalisasi perlu?

- Karena adanya struktur database yang kurang bagus
 - Data yang sama tersimpan di beberapa tempat (file atau record)
 - Integritas data yang tidak terjamin
 - Terjadi kehilangan informasi
 - Terjadi adanya **redundansi** (pengulangan) atau **duplikasi** data sehingga memboroskan ruang penyimpanan dan menyulitkan saat proses updating data
 - Adanya **NULL VALUE**
 - Adanya atribut yang tidak perlu disimpan (*derive attribut*)

8

Tujuan Normalisasi

1. Untuk menghilangkan kerangkapan data
2. Untuk mengurangi kompleksitas
3. Untuk mempermudah pemodifikasian data

9

Proses Normalisasi

- Data diuraikan dalam bentuk tabel, selanjutnya dianalisis berdasarkan persyaratan tertentu ke beberapa tingkat.
- Apabila tabel yang diuji belum memenuhi persyaratan tertentu, maka tabel tersebut perlu dipecah menjadi beberapa tabel yang lebih sederhana sampai memenuhi bentuk yang optimal.

10

Kebergantungan

- Sebelum masuk ke proses normalisasi harus memahami istilah kebergantungan / dependency
- Terdapat 2 yang penting yaitu :
 - Kebergantungan Fungsional (Functional Dependency / FD)
 - Kebergantungan Transitif (Transitif Dependency)

11

Functional Dependency

- Diberikan sebuah relasi R, atribut Y dari R adalah bergantung fungsi pada atribut X dari R jika dan hanya jika setiap nilai X dalam R punya hubungan dengan tepat satu nilai Y dalam R
- Digambarkan : $X \rightarrow Y$
 - Sisi kiri (X) disebut sebagai determinan atau penentu
 - Untuk setiap nilai X tertentu pasti akan didapatkan nilai Y yang sama
 - Secara fungsional : X menentukan Y, atau Y tergantung pada X
- Notasi: $X \nrightarrow Y$ atau $X \not\rightarrow Y$
 - Adalah kebalikan dari Notasi FD di atas

12

Functional Dependency

Contoh tabel nilai

Namakul	Nrp	namaMhs	NiHuruf
Struktur Data	980001	Ali Akbar	A
Struktur Data	980004	Indah Susanti	B
Basis Data	980001	Ali Akbar	
Basis Data	980002	Budi Haryanto	
Basis Data	980004	Indah Susanti	
Bahasa Indonesia	980001	Ali Akbar	B
Matematika I	980002	Budi Haryanto	C

13

FD dari Tabel Nilai

- **Nrp \rightarrow namaMhs**
 - Karena untuk setiap nilai nrp yang sama, maka nilai namaMhs juga sama.
- **{Namakul, nrp} \rightarrow NiHuruf**
 - Karena atribut Nihuruf tergantung pada Namakul dan nrp secara bersama-sama. Dalam arti lain untuk Namakul dan nrp yang sama, maka NiHuruf juga sama, karena Namakul dan nrp merupakan key (bersifat unik).
- **Namakul \nrightarrow nrp**
- **Nrp \nrightarrow NiHuruf**

14

Kebergantungan Fungsional Penuh

- Full Functional Dependency
- Atribut Y pada relasi R dikatakan tergantung fungsional penuh pada atribut X pada relasi R, jika Y tidak tergantung pada subset dari X (bila X adalah key gabungan).
- Suatu atribut Y mempunyai dependensi sepenuhnya terhadap atribut X jika
 - Y mempunyai dependensi terhadap X
 - Y **tidak** mempunyai dependensi terhadap bagian dari X

15

Contoh Full FD

- Tabel Kirim(kodesup,namasup,kodebrg,jumlah)

Kodesup	Namasup	Kodebrg	Jumlah
P01	Bahana	B01	1000
P01	Bahana	B02	1400
P01	Bahana	B03	2000
P02	Sinar Mulia	B03	1000
P03	Harapan	B02	2000

- Full FD:
 - Kodesup \rightarrow namasup
 - {kodebrg, kodesup} \rightarrow jumlah
- Bukan Full FD
 - {kodebrg,kodesup,namasup} \rightarrow jumlah
 - Karena jumlah hanya tergantung pada kodebrg+kodesup, jadi tidak perlu ada namasup

16

Kebergantungan Transitif / Transitif Dependency

- Atribut Z pada relasi R dikatakan tergantung transitif pada atribut X , jika atribut Y tergantung pada atribut X pada relasi R dan atribut Z tergantung pada atribut Y pada relasi R. ($X \rightarrow Y, Y \rightarrow Z$, maka $X \rightarrow Z$)
- Suatu atribut Z mempunyai dependensi transitif terhadap X jika:
 - Y memiliki dependensi terhadap X dan
 - Z memiliki dependensi terhadap Y

17

Contoh Dependensi Transitif

Kuliah	Ruang	Tempat	Waktu
Jaringan Komputer	Merapi	Gedung Utara	Senin, 08.00-09.50
Pengantar Basis Data	Merbabu	Gedung Utara	Selasa, 08.00-09.50
Matematika I	Rama	Gedung Selatan	Rabu, 10.00-11.50
Sistem Pakar	Sinta	Gedung Selatan	Kamis, 08.00-09.50
Kecerdasan Buatan	Merapi	Gedung Utara	Selasa, 10.00-11.50

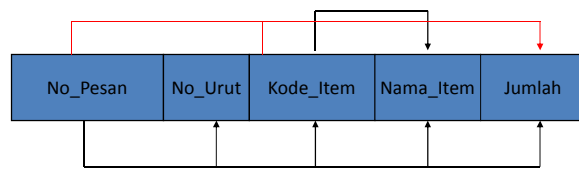
$Kuliah \rightarrow \{ Ruang, Waktu \}$
 $Ruang \rightarrow Tempat$

} $Kuliah \rightarrow Ruang \rightarrow Tempat$

18

Contoh Lain Dependensi Transitif

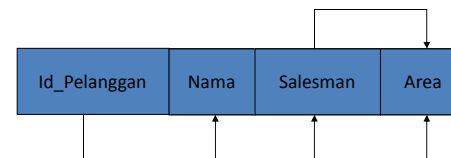
No_Pesan	No_Urut	Kode_Item	Nama_Item	Jumlah
06008	1	P1	Pensil	5
06008	2	P2	Buku Tulis	10
06008	3	P3	Penggaris	6
06008	4	P4	Penghapus	4
06009	1	P3	Penggaris	1
06009	2	P5	Pulpen	10
06009	3	P6	Spidol	5
06010	1	P1	Pensil	4
06010	2	P2	Buku Tulis	10



19

Contoh Lain Dependensi Transitif

Id_Pelanggan	Nama	Salesman	Area
A-001	Andi	Farkan	Jateng
A-002	Kurnia Jati	Dian	Jabar
B-001	Fika Dewi	Joned	Jatim
B-002	Gani Wirawan	Farkan	Jateng
C-001	Cici Kusuma	Joned	Jatim



$Idpelanggan \rightarrow \{ nama, salesman \}$
 $Salesman \rightarrow Area$
 $IdPelanggan \rightarrow Salesman \rightarrow Area$

20

Problem pada Dependensi Transitif

Id_Pelanggan	Nama	Salesman	Area
A-001	Andi	Farkan	Jateng
A-002	Kurnia Jati	Dian	Jabar
B-001	Fika Dewi	Joned	Jatim
B-002	Gani Wirawan	Farkan	Jateng
C-001	Cici Kusuma	Joned	Jatim

- Anomali Penyisipan
 - Seorang salesman baru yang bertugas di Jateng tidak dapat dimasukkan dalam tabel sampai salesman tersebut mendapatkan seorang pelanggan
 - Karena Salesman Bergantung Fungsional pada Id-Pelanggan

21

Problem pada Dependensi Transitif

Id_Pelanggan	Nama	Salesman	Area
A-001	Andi	Farkan	Jateng
A-002	Kurnia Jati	Dian	Jabar
B-001	Fika Dewi	Joned	Jatim
B-002	Gani Wirawan	Farkan	Jateng
C-001	Cici Kusuma	Joned	Jatim

- Anomali Penghapusan
 - Jika pelanggan A-002 dihapus, informasi bahwa Dian menangani daerah Jabar ikut hilang
 - Karena Salesman Bergantung Fungsional pada Id-Pelanggan

22

Problem pada Dependensi Transitif

Id_Pelanggan	Nama	Salesman	Area
A-001	Andi	Farkan	Jateng
A-002	Kurnia Jati	Dian	Jabar
B-001	Fika Dewi	Joned	Jatim
B-002	Gani Wirawan	Farkan	Jateng
C-001	Cici Kusuma	Joned	Jatim

- Anomali Pengupdatean
 - Jika katakanlah Farkan mendapat penugasan baru untuk menangani daerah Kalimantan, maka sejumlah baris harus ikut diupdate agar data tetap konsisten (disebut *dependency preservation*)
 - Karena Area bergantung fungsional pada Salesman dan Salesman Farkan terdapat lebih dari 1 baris

23

Tabel Universal

- Tabel Universal (*Universal / Star Table*) → sebuah tabel yang merangkum semua kelompok data yang saling berhubungan, bukan merupakan tabel yang baik.

24

nim	nama_mhs	alamat_mhs	tgl_lahir	Kode_kul	Nama_kul	sks	semester	waktu_ajar	tempat	Ruang	Nama_dos	Alamat_dos
20001	A. Pinar	J. Merdeka no 1 Jember 60132	01-01-1999	001	Matematika	3	1	08.00-10.00	Jember	Ruang 1	Dr. Nur Hafid	J. Garuda no 1, Jember 60132
20002	A. Pinar	J. Merdeka no 1 Jember 60132	01-01-1999	001	Matematika	3	2	08.00-10.00	Jember	Ruang 2	Dr. Nur Hafid	J. Garuda no 1, Jember 60132
20003	A. Pinar	J. Merdeka no 1 Jember 60132	01-01-1999	001	Matematika	3	2	08.00-10.00	Jember	Ruang 3	Dr. Nur Hafid	J. Garuda no 1, Jember 60132
20004	B. S	J. Jember no 2 Jember 60132	01-01-1999	001	Matematika	3	2	08.00-10.00	Jember	Ruang 3	Dr. Nur Hafid	J. Garuda no 1, Jember 60132
20005	B. S	J. Jember no 2 Jember 60132	01-01-1999	001	Matematika	3	1	08.00-10.00	Jember	Ruang 1	Dr. Nur Hafid	J. Garuda no 1, Jember 60132
20006	C. S	J. Jember no 2 Jember 60132	01-01-1999	001	Matematika	3	1	08.00-10.00	Jember	Ruang 1	Dr. Nur Hafid	J. Garuda no 1, Jember 60132
20007	C. S	J. Jember no 2 Jember 60132	01-01-1999	001	Matematika	3	1	08.00-10.00	Jember	Ruang 1	Dr. Nur Hafid	J. Garuda no 1, Jember 60132

- Dari table universal tersebut, dengan memperhatikan kesamaan dan ketidak samaan data diantara baris-baris data juga dengan memahami hubungan alamiah antar data, kita dapat membentuk FD sebagai berikut:

- $Nim \rightarrow nama_mhs$
- $Nim \rightarrow alamat_mhs$
- $Nim \rightarrow tgl_lahir$
- $Kode_kul \rightarrow nama_kul$
- $Kode_kul \rightarrow sks$
- $Kode_kul \rightarrow semester$
- $Kode_kul \rightarrow waktu$
- $Kode_kul \rightarrow tempat$
- $Kode_kul \rightarrow nama_dos$
- $Nama_dos \rightarrow alamat_dos$
- $Nim, kode_kul \rightarrow indeks_nilai$

26

3 Kelemahan dari tabel universal

- Pengulangan informasi (redundansi)
 - Yang terjadi pada atribut **nama_mhs**, **alamat_mhs** dan **tgl_lahir** yang dinyatakan berulang-ulang sesuai dengan data atribut **nim**, begitu juga dengan atribut **nama_kul**, **tempat**, **waktu** dan seterusnya.
- Potensi inkonsistensi data pada operasi pengubahan
 - Yang terjadi jika ada perubahan pada data **nama_mhs**, dimana perubahan ini harus dijalankan keseluruhan baris data pada table tersebut untuk **nim** yang sama. Jika perubahan ini tidak dilakukan, maka FD yang telah ditetapkan akan terganggu, karena kelak akan ada 2 row atau lebih dengan **nim** yang sama, tapi **nama_mhs** nya berbeda.

27

Kelemahan...

- Tersembunyiya informasi tertentu
 - Tabel universal dibangun atas dasar keterkaitan antar item-item data. Karena itu table semacam ini tidak akan mampu menampilkan informasi tentang item-item data yang kebetulan belum memiliki keterkaitan dengan item data yang lain. Misalnya, mata kuliah yang tidak diambil mahasiswa tidak akan pernah diketahui

Kelemahan-kelemahan tersebut mengiring kita untuk melakukan dekomposisi, yakni melakukan pemilihan table tersebut menjadi beberapa table dengan mempertimbangkan ketergantungan fungsional yang telah kita dapatkan, dekomposisi dilakukan agar setiap table yang kita gunakan hanya memiliki 1 (satu) FD saja, lebih tepatnya FD minimum

28

Kriteria Tabel Yang Baik (efisien)

- Sebuah tabel dikatakan baik (efisien) atau normal jika memenuhi 3 kriteria sbb:
 - Jika ada dekomposisi (penguraian) tabel, maka dekomposisinya harus dijamin aman (Lossless-Join Decomposition).
 - Terpeliharanya ketergantungan fungsional pada saat perubahan data (*Dependency Preservation*).
 - Tidak melanggar *Boyce-Codd Normal Form (BCNF)*.
- Jika kriteria ketiga (BCNF) tidak dapat terpenuhi, maka paling tidak tabel tersebut tidak melanggar bentuk normal tahap ketiga (*3rd Normal Form/3NF*).

29

Lossless Join Decomposition

- Dekomposisi merupakan upaya untuk mendapatkan table yang baik, tapi bila tidak berhati-hati upaya ini justru dapat menghasilkan kesalahan. Dekomposisi yang benar terjadi jika table-table hasil dekomposisi kita gabungkan kembali dapat menghasilkan table awal sebelum didekomposisi. Dekomposisi yang benar semacam ini disebut *Lossless-Join Decomposition* atau *Lossless Decomposition* (dapat di Indonesia-kan dengan istilah Dekomposisi Aman).

30

Lossless Join Decomposition

- Berikut sebuah contoh yang menghasilkan dekomposisi yang tidak aman (*Lossy-Join Decomposition*). Misalnya ada sebuah table ABC, yang didefinisikan dengan 2 buah FD yaitu $A \rightarrow B$ dan $B \rightarrow C$.
- Kedua FD tersebut diperoleh dari pengamatan terhadap data yang kurang memadai atau karena asumsi yang kurang tepat.

Tabel ABC

A	B	C
A1	100	C1
A2	200	C2
A3	300	C3
A4	200	C4

Memang dengan isi seperti itu, pernyataan FD yang kedua $B \rightarrow C$ tidak sepenuhnya tepat, karena pada row 2 dan row 4, dengan nilai untuk atribut B yang sama, nilai untuk atribut C nya berbeda.

31

Tapi yang ingin kita tekankan disini adalah adanya 2 buah FD itu, mendorong kita untuk medekomposisi table ABC menjadi 2 buah table, yaitu table AB dan Tabel BC sbb:

Tabel AB

A	B
A1	100
A2	200
A3	300
A4	200

Tabel BC

B	C
100	C1
200	C2
300	C3
200	C4

Hasil gabungan table AB dan BC

A	B	C
A1	100	C1
A2	200	C2
A2	200	C4
A3	300	C3
A4	200	C2
A4	200	C4

Jika table AB dan table BC digabungkan, hasilnya tidak menghasilkan table awal, sebelum dekomposisi, ini yang disebut dengan *Lossy Join Decomposition*, disini bisa disimpulkan bahwa kesalahan penentuan FD bisa mengakibatkan kesalahan dekomposisi

32

Dependency Preservation

- *Dependency preservation* (dapat di Indonesia-kan sebagai Pemeliharaan Ketergantungan) merupakan kriteria yang harus dicapai untuk mendapatkan tabel dan basis data yang baik.
- Ketika melakukan perubahan data, maka harus dapat dijamin agar perubahan tersebut tidak menghasilkan **inkonsistensi data** yang mengakibatkan FD yang sudah benar menjadi tidak terpenuhi tetapi dalam upaya untuk memelihara FD yang ada untuk tetap terpenuhi tersebut, prosesnya harus dapat dilakukan dengan efisien.

33

Dependency Preservation

- Jika ditinjau pada table universal yang telah digambarkan sebelumnya, sudah jelas sangat rapuh didalam memenuhi kriteria dependency preservation. Walaupun ingin dipaksakan (agar FD yang ada tetap dapat terjaga pada saat ada perubahan yang terjadi), maka upaya pemeliharaan FD tersebut akan berlangsung tidak efisien
- Katakanlah ada perubahan data **Alamat** untuk mahasiswa dengan nim=980001, maka perubahan ini harus juga dilakukan pada atribut **alamat_mhs** disemua row yang nilai atribut **nim_nya** berisi **980001**

34

Sekian

- Bagian 2 akan membahas tahapan-tahapan normalisasi

35