

12

Database Design Theory & Normalization (part 2)

CSF2600700 - BASIS DATA



The background image shows an aerial view of the Universitas Indonesia campus, featuring several large, modern buildings with tiered, dark roofs and extensive green landscaping.

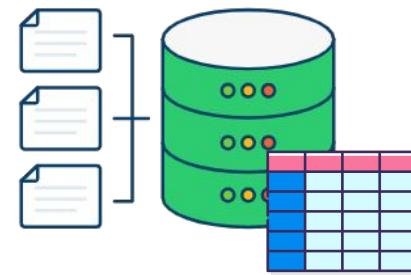
Acknowledgements

This slide is a modification to a slide titled
Database Design Theory & Normalization (part 1) used in “Basis Data” course in academic years
2019/2020 in the Faculty of Computer Science, Universitas Indonesia.

Tujuan Pembelajaran

Mengevaluasi **model database relasional** yang telah dibuat berdasarkan teori desain database relasional sehingga dapat menghasilkan skema database yang “baik” dalam level conceptual maupun physical.

Implicit goals: **Information preservation** and **minimum redundancy**.



Outline

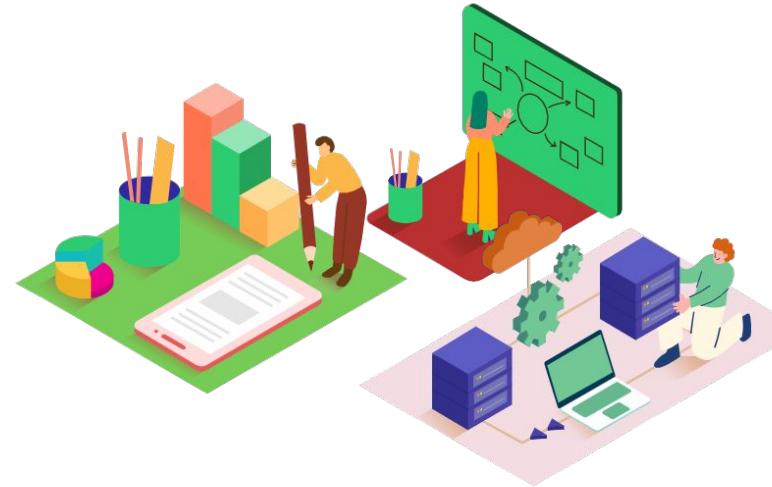
1. Panduan Informal dalam Merancang
Basis Data Relasional

2. Functional Dependency

3. Normalisasi Berdasarkan Primary Key

4. General Normal Form

5. Boyce Codd Normal Form



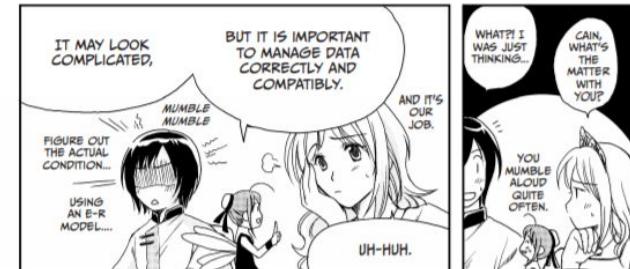
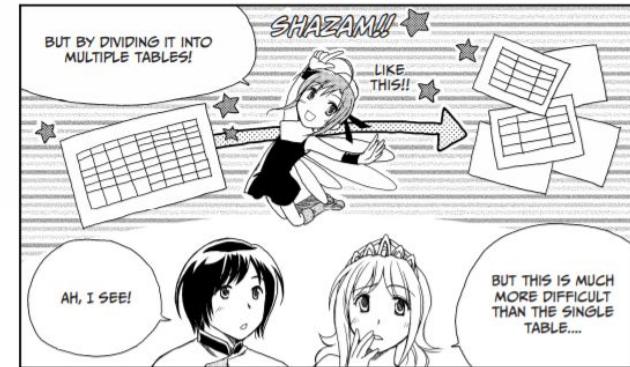
Normalisasi

Normalisasi (Normalization)

Proses dekomposisi relasi yang masih “buruk” dengan **memecah atribut-atributnya** untuk membentuk beberapa relasi

Bentuk Normal (Normal Form)

Kondisi (dengan menggunakan FD dan key) yang menentukan apakah suatu skema relasi **memenuhi kriteria tertentu**



Bentuk Normal

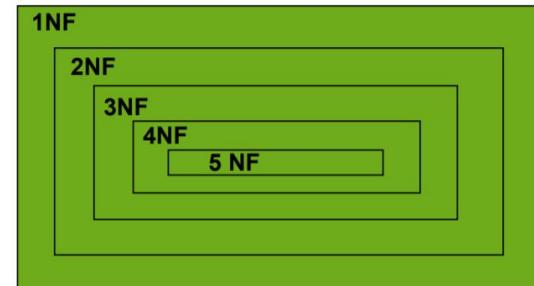
Ada beberapa bentuk normal berdasarkan sejumlah kriteria:

Primary keys (1NF, 2NF, 3NF)

All Candidate Keys (2NF, 3NF, BCNF)

Multivalued Dependencies (4NF)

Join Dependencies (5NF)



Penggunaan Bentuk Normal

Normalisasi perlu dilakukan agar rancangan basis data yang dihasilkan berkualitas baik dan memenuhi sifat yang diinginkan

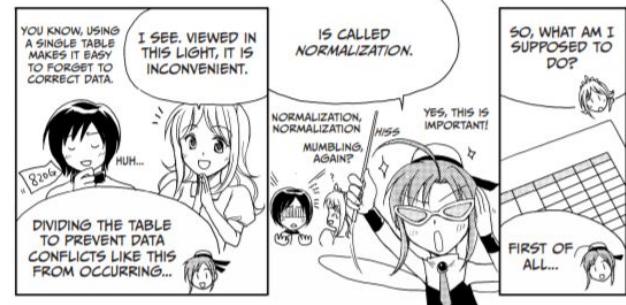
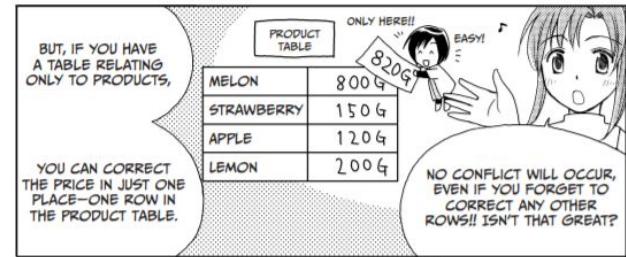
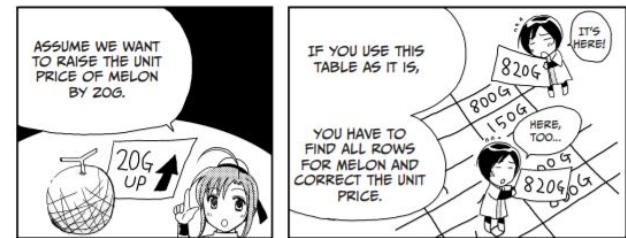
Normalisasi dalam praktiknya sulit dilakukan jika batasan-batasan yang menjadi dasar normalisasi sulit dimengerti atau sulit dideteksi

Perancang basis data tidak perlu melakukan normalisasi sampai bentuk tertinggi

Normalisasi biasanya dilakukan sampai mencapai 3NF atau BCNF

Denormalisasi

kebalikan dari proses normalisasi, yakni menggabungkan beberapa relasi, membawa ke bentuk normal yang lebih rendah



Key Attributes

Superkey dari relasi R

Himpunan attribute dari R yang dapat membedakan satu tuple dengan tuple lainnya pada R

Key K

Superkey minimal, sedemikian hingga penghapusan salah satu attribute dari K akan menyebabkan K tidak lagi menjadi key

Jika relasi punya beberapa key, masing-masing disebut **candidate key**. Salah satu dari candidate key dipilih menjadi **primary key**

Prime Attribute

Attribute yang menjadi anggota dari candidate key

Non Prime Attribute

Attribute yang bukan merupakan anggota candidate key manapun

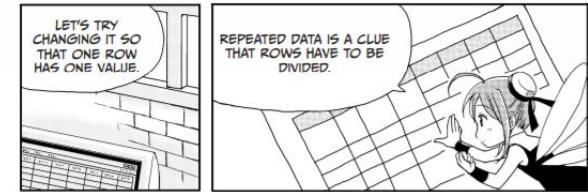
Bentuk Normal Pertama (1NF)

Bentuk Normal Pertama (1 NF)

merupakan bagian dari definisi relasi

Bentuk normal pertama tidak mengizinkan:

- Composite attributes,
- Multivalue attributes,
- Nested relations.



SO, I'LL DIVIDE IT INTO...

ONE TABLE WITH DATE, EXPORT DESTINATION CODE, AND EXPORT DESTINATION NAME...

AND ANOTHER TABLE WITH PRODUCT CODE, PRODUCT NAME, UNIT PRICE, AND QUANTITY.

BUT THE REPORT CODE IS PROVIDED IN BOTH TABLES, ISN'T IT?

HUH.

YES, THAT WAY YOU CAN IDENTIFY IF THERE IS AN ASSOCIATION BETWEEN THE TWO TABLES.

REPORT_CODE	DATE	EXPORT_DEST_CODE	EXPORT_DEST_NAME
1101	3/5	12	THE KINGDOM OF MINAMI
1102	3/7	23	ALPHA EMPIRE
1103	3/8	25	THE KINGDOM OF RITO
1104	3/10	12	THE KINGDOM OF MINAMI
1105	3/12	25	THE KINGDOM OF RITO

REPORT_CODE	PRODUCT_CODE	PRODUCT_NAME	UNIT_PRICE	QUANTITY
1101	101	MELON	800G	1,100
1101	102	STRAWBERRY	150G	300
1102	103	APPLE	120G	1,700
1103	104	LEMON	200G	500
1104	101	MELON	800G	2,500
1105	103	APPLE	120G	2,000
1105	104	LEMON	200G	700

Normalisasi dari Multivalued ke 1NF

(a)

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocations

(b)

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

(c)

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocation
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

Figure 14.9

Normalization into 1NF. (a) A relation schema that is not in 1NF. (b) Sample state of relation DEPARTMENT. (c) 1NF version of the same relation with redundancy.

Sample Solution for Example 1

1st Option

Dname	<u>Dnumber</u>	Dmgr_ssn	<u>Dlocations</u>
-------	----------------	----------	-------------------

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	<u>Dlocations</u>
-------	----------------	----------	-------------------

```
graph TD; DEPARTMENT[Dname | Dnumber | Dmgr_ssn | Dlocations] --> DEPT1[DEPT_1]; DEPARTMENT --> DEPT2[DEPT_2]; DEPARTMENT --> DEPT2
```

2nd Option

DEPT_1

Dname	<u>Dnumber</u>	Dmgr_ssn
-------	----------------	----------

DEPT_2

<u>Dnumber</u>	<u>Dlocation</u>
----------------	------------------

3rd Option

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocation1	Dlocation2	Dlocation3
-------	----------------	----------	------------	------------	------------

Normalisasi dari Nested Relation ke 1NF

(a)

EMP_PROJ

Projs			
Ssn	Ename	Pnumber	Hours

(b)

EMP_PROJ

Ssn	Ename	Pnumber	Hours
123456789	Smith, John B.	1	32.5
		2	7.5
666884444	Narayan, Ramesh K.	3	40.0
		2	20.0
453453453	English, Joyce A.	1	20.0
		2	20.0
333445555	Wong, Franklin T.	2	10.0
		3	10.0
		10	10.0
		20	10.0
		30	30.0
999887777	Zelaya, Alicia J.	10	10.0
		30	35.0
987987987	Jabbar, Ahmad V.	30	5.0
		20	15.0
987654321	Wallace, Jennifer S.	30	20.0
888665555	Borg, James E.	20	NULL

Figure 14.10

Normalizing nested relations into 1NF.
 (a) Schema of the EMP_PROJ relation with a *nested relation* attribute PROJS. (b) Sample extension of the EMP_PROJ relation showing nested relations within each tuple.
 (c) Decomposition of EMP_PROJ into relations EMP_PROJ1 and EMP_PROJ2 by propagating the primary key.

(c)

EMP_PROJ1

Ssn	Ename
-----	-------

EMP_PROJ2

Ssn	Pnumber	Hours
-----	---------	-------

Bentuk Normal Kedua (2NF)

Bentuk Normal Kedua (2 NF)

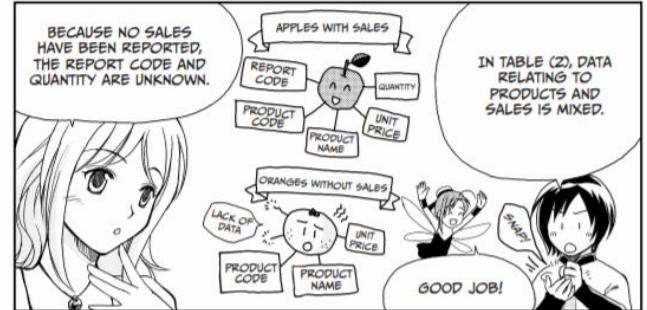
Menggunakan konsep FD dan primary key

Definisi:

Full functional dependency: suatu FD $X \rightarrow Y$ sedemikian hingga jika salah satu attribute dari X dibuang, maka FD tersebut tidak ada lagi.

Contoh:

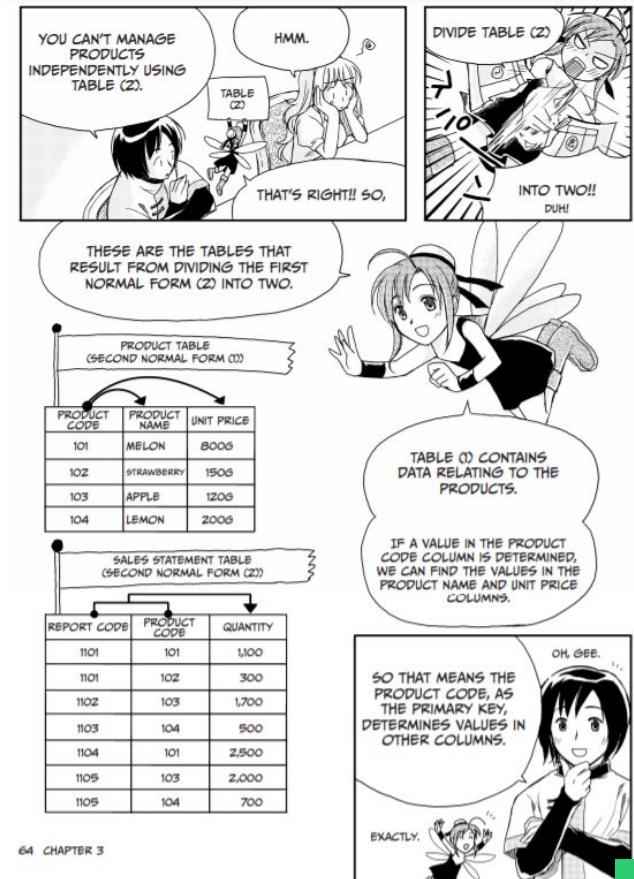
- $\{\text{SSN}, \text{PNUMBER}\} \rightarrow \text{HOURS}$ merupakan full FD, karena tidak ada $\text{SSN} \rightarrow \text{HOURS}$ dan $\text{PNUMBER} \rightarrow \text{HOURS}$
- $\{\text{SSN}, \text{PNUMBER}\} \rightarrow \text{ENAME}$ merupakan partial dependency, bukan full FD karena terdapat dependency $\text{SSN} \rightarrow \text{ENAME}$



Bentuk Normal Kedua (2NF) (Cntd.)

Suatu relasi R berada dalam bentuk normal kedua (2NF) jika R berada dalam **1NF** dan setiap non prime attribute A dalam relasi R bersifat **full functional dependent terhadap primary key**.

Suatu relasi R dapat didekomposisi ke relasi relasi yang memenuhi 2NF melalui proses normalisasi 2NF



Contoh Normalisasi 2NF

(a)

EMP_PROJ

Ssn	Pnumber	Hours	Ename	Pname	Plocation
FD1			↑		
FD2				↑	
FD3					↑

- ENAME fully dependent on SSN
- PNAME, PLOCATION fully dependent on PNUMBER

2NF Normalization

EP1

Ssn	Pnumber	Hours
FD1		↑

EP2

Ssn	Ename
FD2	↑

EP3

Pnumber	Pname	Plocation
FD3	↑	↑

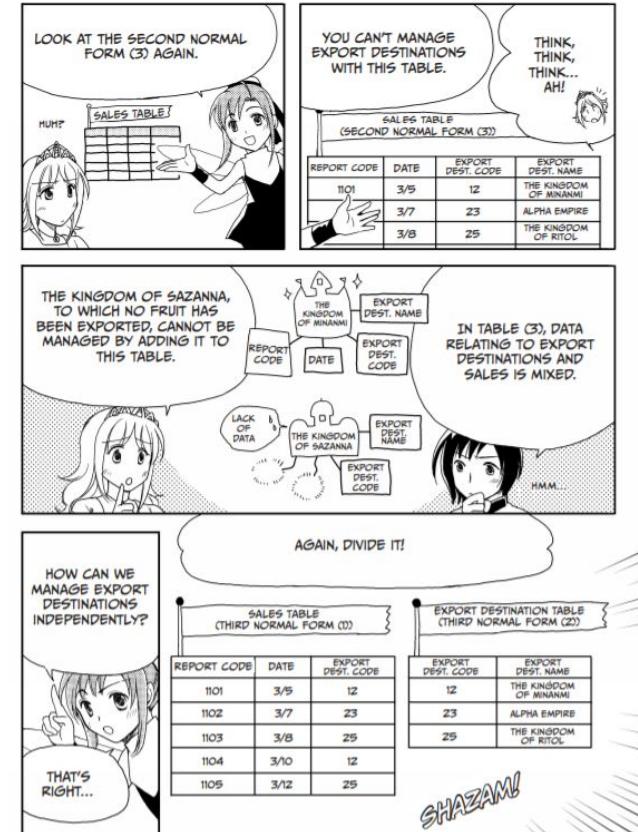
Bentuk Normal Ketiga (3NF)

Suatu relasi R berada dalam bentuk normal ketiga (3NF) jika R berada dalam 2NF dan tidak ada nonprime attribute A di R yang memiliki dependensi transitif terhadap primary key.

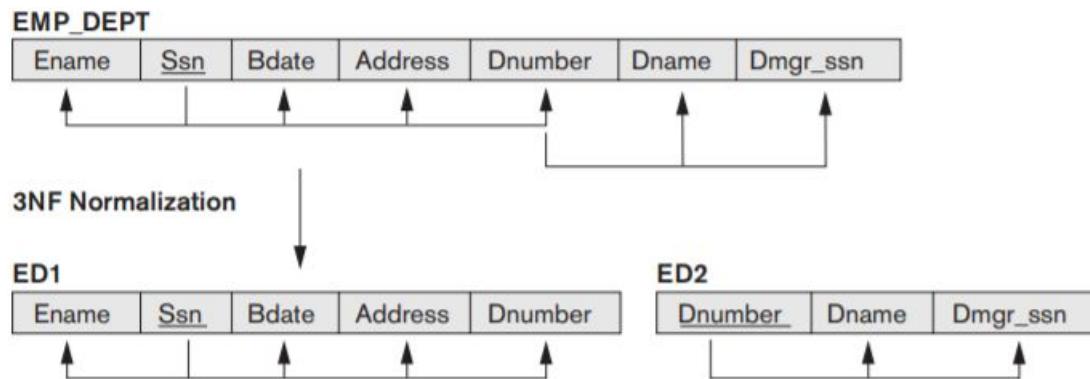
Suatu relasi R dapat didekomposisi ke relasi-relasi yang memenuhi 3NF melalui proses normalisasi 3NF

Catatan:

- Pada FD $X \rightarrow Y$ dan $Y \rightarrow Z$, dengan X sebagai primary key, kita mempertimbangkan ini sebagai sebuah problem hanya jika Y bukan sebuah candidate key.
- Jika Y merupakan candidate key, tidak ada masalah dengan dependensi transitif ini
- Contoh: Relasi EMP (SSN, Emp#, Salary) tidak melanggar 3NF karena meskipun terdapat FD $SSN \rightarrow Emp\# \rightarrow Salary$, Emp# merupakan candidate key



Contoh Normalisasi 3NF



Bentuk Normal Didefinisikan secara Informal

1st normal form

All attributes depend on the key

2nd normal form

All attributes depend on the whole key

3rd normal form

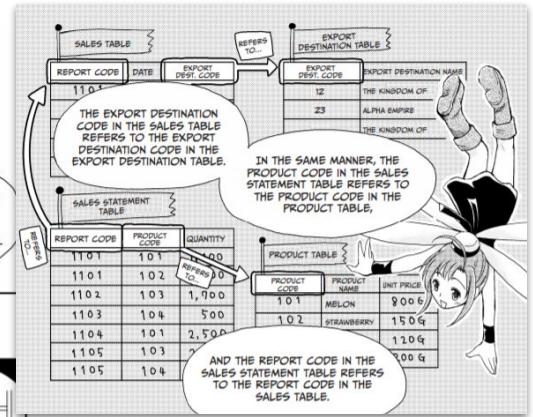
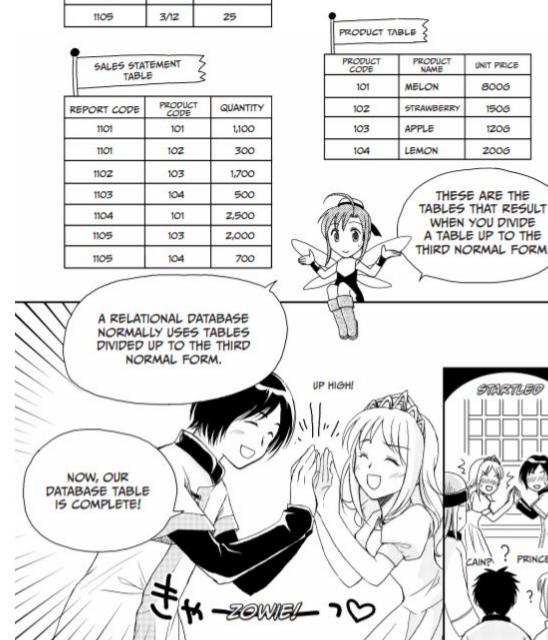
All attributes depend on nothing but the key

SALES TABLE		
REPORT CODE	DATE	EXPORT DEST. CODE
1101	3/5	12
1102	3/7	23
1103	3/8	25
1104	3/10	12
1105	3/12	25

SALES STATEMENT TABLE		
REPORT CODE	PRODUCT CODE	QUANTITY
1101	101	1,000
1101	102	300
1102	103	1,700
1103	104	500
1104	101	2,500
1105	103	2,000
1105	104	700

EXPORT DESTINATION TABLE	
EXPORT DEST. CODE	EXPORT DESTINATION NAME
12	THE KINGDOM OF MINAMI
23	ALPHA EMPIRE
25	THE KINGDOM OF RITOL

PRODUCT TABLE		
PRODUCT CODE	PRODUCT NAME	UNIT PRICE
101	MELON	8006
102	STRAWBERRY	1506
103	APPLE	1206
104	LEMON	2006



Rangkuman Bentuk Normal Berdasarkan Primary Key

Table 14.1 Summary of Normal Forms Based on Primary Keys and Corresponding Normalization

Normal Form	Test	Remedy (Normalization)
First (1NF)	Relation should have no multivalued attributes or nested relations.	Form new relations for each multivalued attribute or nested relation.
Second (2NF)	For relations where primary key contains multiple attributes, no nonkey attribute should be functionally dependent on a part of the primary key.	Decompose and set up a new relation for each partial key with its dependent attribute(s). Make sure to keep a relation with the original primary key and any attributes that are fully functionally dependent on it.
Third (3NF)	Relation should not have a nonkey attribute functionally determined by another nonkey attribute (or by a set of nonkey attributes). That is, there should be no transitive dependency of a nonkey attribute on the primary key.	Decompose and set up a relation that includes the nonkey attribute(s) that functionally determine(s) other nonkey attribute(s).

Outline

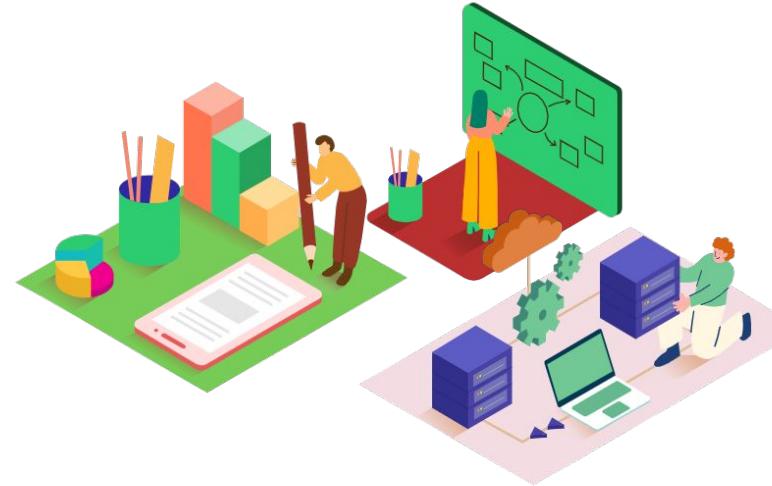
1. Panduan Informal dalam Merancang
Basis Data Relasional

2. Functional Dependency

3. Normalisasi Berdasarkan Primary Key

4. General Normal Form

5. Boyce Codd Normal Form



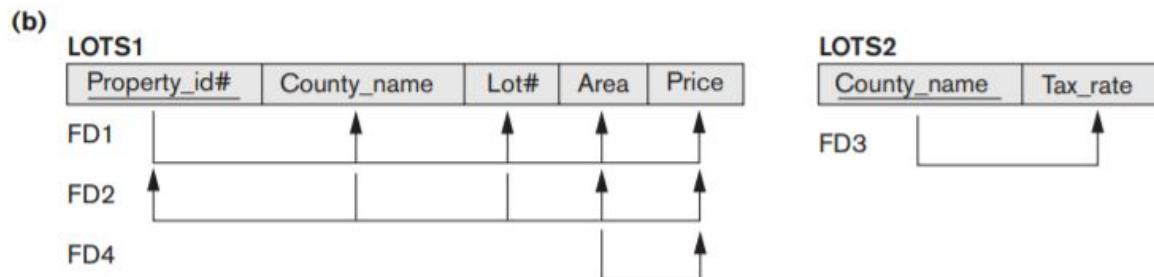
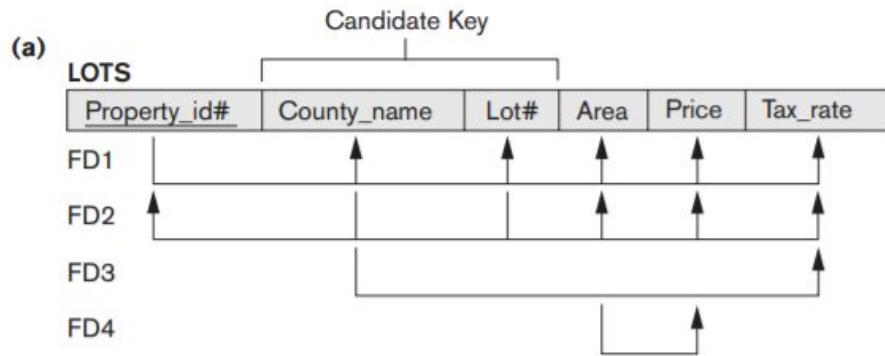
Definisi dalam General Normal Form

Definisi-definisi sebelumnya hanya mempertimbangkan primary key

Definisi selanjutnya mempertimbangkan semua candidate key pada suatu relasi

Skema relasi R berada pada 2NF (general definition) jika setiap nonprime attribute A pada R bersifat full functional dependent pada **setiap key** pada R

Contoh Normalisasi 2NF



Definisi dalam General Normal Form

Definisi:

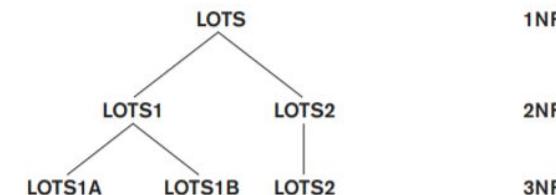
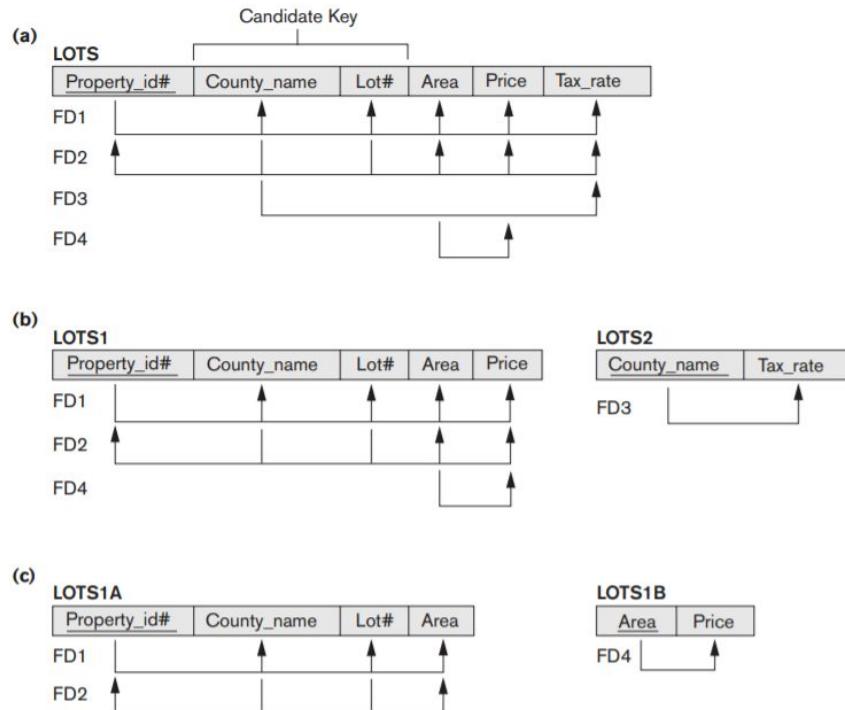
Superkey pada relasi R: himpunan attributes S dari R yang berisi key dari R

Skema relasi R berada dalam 3NF dengan syarat jika terdapat FD $X \rightarrow Y$ maka:

- (a) X merupakan superkey dari R atau
- (b) Y merupakan prime attribute dari R

BCNF tidak membolehkan kondisi (b) di atas

Contoh Normalisasi 3NF

**Figure 14.12**

Normalization into 2NF and 3NF. (a) The LOTS relation with its functional dependencies FD1 through FD4. (b) Decomposing into the 2NF relations LOTS1 and LOTS2. (c) Decomposing LOTS1 into the 3NF relations LOTS1A and LOTS1B. (d) Progressive normalization of LOTS into a 3NF design.

Outline

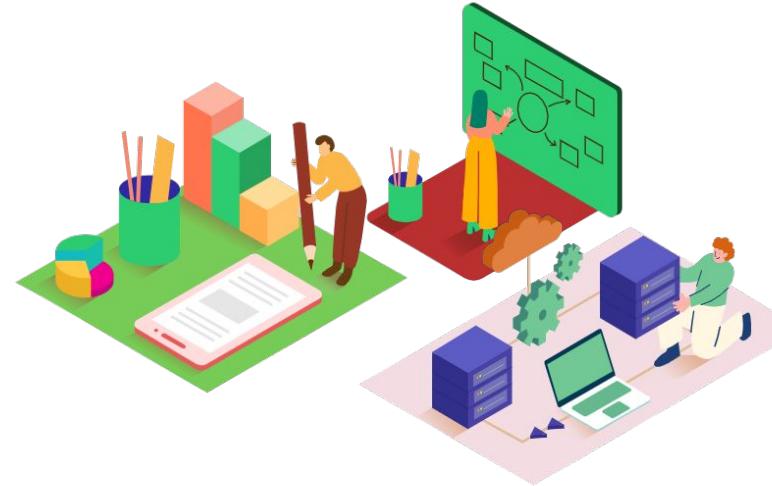
1. Panduan Informal dalam Merancang
Basis Data Relasional

2. Functional Dependency

3. Normalisasi Berdasarkan Primary Key

4. General Normal Form

5. Boyce Codd Normal Form



Boyce Codd Normal Form (BCNF)

Definisi:

Suatu skema relasi R berada pada BCNF dengan syarat jika terdapat FD $X \rightarrow Y$ pada R, maka X merupakan superkey dari R.

- Setiap BCNF pasti memenuhi 3NF
- Ada relasi yang berada pada 3NF namun tidak BCNF
- Tujuan normalisasi umumnya untuk mencapai 3NF atau BCNF

Contoh BCNF

(a)

LOTS1A

<u>Property_id#</u>	County_name	Lot#	Area
---------------------	-------------	------	------

FD1

FD2

FD5

BCNF Normalization

LOTS1AX

<u>Property_id#</u>	Area	Lot#
---------------------	------	------

LOTS1AY

Area	County_name
------	-------------

(b)

R

A	B	C
---	---	---

FD1

FD2

Contoh: Relasi pada 3NF Namun tidak BCNF

TEACH

Student	Course	Instructor
Narayan	Database	Mark
Smith	Database	Navathe
Smith	Operating Systems	Ammar
Smith	Theory	Schulman
Wallace	Database	Mark
Wallace	Operating Systems	Ahamad
Wong	Database	Omiecinski
Zelaya	Database	Navathe
Narayan	Operating Systems	Ammar

Figure 14.14

A relation TEACH that is in 3NF but not BCNF.

Normalisasi BCNF

Ada 2 FD pada relasi TEACH:

- FD1: { student, course } → instructor
- FD2: instructor → course

{student, course} merupakan candidate key untuk teach

Relasi ini berada pada 3NF namun tidak pada BCNF

Relasi yang belum BCNF dapat didekomposisi untuk mencapai BCNF, namun kadang-kadang dapat menghilangkan FD yang semula telah ada

TEACH

Student	Course	Instructor
Narayan	Database	Mark
Smith	Database	Navathe
Smith	Operating Systems	Ammar
Smith	Theory	Schulman
Wallace	Database	Mark
Wallace	Operating Systems	Ahamad
Wong	Database	Omiecinski
Zelaya	Database	Navathe
Narayan	Operating Systems	Ammar

Normalisasi BCNF (Cntd.)

Ada 3 dekomposisi yang mungkin untuk relasi TEACH:

- {student, instructor} and {student, course}
- {course, instructor} and {course, student}
- {instructor, course} and {instructor, student}

Semua dekomposisi tersebut kehilangan FD1 :
{ student, course} -> instructor

Semua FD yang ada pada relasi sedapat mungkin dijaga, namun dapat dimaklumi jika hilang pada normalisasi BCNF

Namun sifat lossless dan non additivity property setelah dekomposisi harus tetap dijaga

TEACH

Student	Course	Instructor
Narayan	Database	Mark
Smith	Database	Navathe
Smith	Operating Systems	Ammar
Smith	Theory	Schulman
Wallace	Database	Mark
Wallace	Operating Systems	Ahamad
Wong	Database	Omiecinski
Zelaya	Database	Navathe
Narayan	Operating Systems	Ammar

Q&A

