

Database Management Systems

Vijaya Saradhi

IIT Guwahati

Mon, 03rd Feb 2020

Hospital Form Example - 01

Post Surgery Form

Hospital Number: H17	Hospital Name: St Vincent's	Operation Number: 48			
Hospital Category: P		Contact at Hospital: Fred Fleming			
Operation Name: Heart Transplant		Operation Code: 7A	Procedure Group: Transplant		
Surgeon Number: S15	Surgeon Specialty: Cardiology		Total Drug Cost: \$75.50		
Drug Code	Full Name of Drug	Manufacturer	Method of Admin.	Cost of Dose (\$)	Number of Doses
MAX 150mg	Maxicillin	ABC Pharmaceuticals	ORAL	\$3.50	15
MIN 500mg	Minicillin	Silver Bullet Drug Co.	IV	\$1.00	20
MIN 250mg	Minicillin	Silver Bullet Drug Co.	ORAL	\$0.30	10

Hospital Form Example - 02

Observations

- One form is filled for each operation
- Each hospital is given a unique number
- All hospitals are prefixed by "H"
- Operations numbers are assigned sequentially
- Hospital categories {Teaching, Public, Private}
- Operation Code - standard international code is adopted
- Surgeon name - recorded in terms of surgeon number
- Total drug cost
- Drug Code, name, method of administration

Hospital Form Example - 03

Initial Data Model

```
OPERATION(Hospital Number, Operation Number,  
  Hospital Name, Hospital Category, Contact Person,  
  Operation Name, Operation Code, Procedure Group,  
  Surgeon Number, Surgeon Speciality, Total Drug Cost,  
  Drug Code 1, Drug Name 1, Manufacturer 1, Method of Administration 1, Dose Cost 1, Number of Doses 1,  
  Drug Code 2, Drug Name 2, Manufacturer 2, Method of Administration 2, Dose Cost 2, Number of Doses 2,  
  Drug Code 3, Drug Name 3, Manufacturer 3, Method of Administration 3, Dose Cost 3, Number of Doses 3,  
  Drug Code 4, Drug Name 4, Manufacturer 4, Method of Administration 4, Dose Cost 4, Number of Doses 4  
)
```

Hospital Form Example - 04

One Fact Per Column

- Drug Code Multivalued - holds short form of name of drug and dosage size
- Dosage Size holds two facts **numeric size** and units
- Three facts are recorded into one column of Drug Code
- Hospital Category: Multivalued - **Private, Public, teaching or non-teaching**
- **1NF**: tells us not to include multivalued attributes
- These attributes to be split from the **OPERATION** table

Hospital Form Example - 04

Hidden Data

- Say hospital files the operations forms as per a sequence
- This information that hospital files the forms in sequence is not explicitly recorded in the form
- It must be included in the OPERATION table
- If the forms are color coded: red for EMERGENCY operations, blue for elective operations etc. These facts must be recorded in the OPERATION table

Hospital Form Example - 04

Derivable Data, Primary Key

- Total Cost a derivable column. Its presence must be argued
- Primary key: (Hospital Number, Operation Number)
- Split the repeating columns/multivalued columns from OPERATION table

Hospital Form Example - 05

Removing Repeating Groups

- Put the data in table form by identifying and eliminating repeating groups
- Remove each **separate set of repeating group** into a new table
- In the new table, each of occurrence of the group becomes a row
- Include key of the original table into each new table
- Name each of the new table
- Identify the primary key for each of the new table

Hospital Form Example - 06

Repeating Groups in OPERATION table

- Drug Short Name
- Drug Name
- Manufacturer
- Size of Dose
- Unit of Measure
- Method of Administration
- Dose Cost
- Number of Doses
- Place these into a new table named DRUG ADMINISTRATION
- With key (Hospital Name, Operation Number, Drug Short Name, Size of Dose, Unit of Measure, Method of Administration)
- Rest are non-repeating attributes are retained in the original table

Hospital Form Example - 07

Convert to 1NF

- OPERATION(Hospital Number, Operation Number, Hospital Name, Hospital Category, Teaching Status, Contact Person, Operation Name, Operation Code, Procedure Group, Surgeon Number, Surgeon Speciality, Total Drug Cost)
- HOSPITAL(Hospital Number, Operation Number, Drug Name, Size of Dose, U
Method of Administration, Dose Cost, Number of Doses)

Hospital Form Example - 08

Identify & Eliminate Redundancy

- Hospital Number \rightarrow { Hospital Name, Hospital Type, Teaching Status, Contact Person }
- Remove this FD from the OPERATION table resulting in three tables
- OPERATION(Hospital Number, Operation Number, Hospital Name, Hospital Category, Contact Person, Operation Name, Operation Code, Procedure Group, Surgeon Number, Surgeon Speciality, Total Drug Cost)
- HOSPITAL(Hospital Number, Hospital Name, Hospital Category, Contact Person)
- DRUG
ADMINISTRATION(Hospital Number, Operation Number, Drug Name, Drug Dose Cost, Number of Doses)

Hospital Form Example - 09

Identify FDs

- Identify FDs in a given relation and place them in separate table
- Remove the above FD columns from the original table
- FD 2: Surgeon: (Hospital Number, Operation Number) \rightarrow Surgeon Name
- FD 3: Operation Type: Operatoion Code \rightarrow (Operation Name, Procedure Group)
- FD 4: Standard Drug Dosage
(Drug Short Name, Size of Dose, Unit of Measure, Method of Administration) \rightarrow Standard Dose Cost
- FD 5: Drug: Drug Short Name \rightarrow (Drug Name, Manufacturer)

Hospital Form Example - 10

Summary - Original Table

OPERATIONS(Hospital Number, Operation Number,
 Hospital Name, Hospital Category, Contact Person,
 Operation Name, Operation Code, Procedure Group,
 Surgeon Number, Surgeon Speciality, Total Drug Cost,
 Drug Code 1, Drug Name 1, Manufacturer 1, Method of Administration 1, Dose Cost 1, Number of Doses 1,
 Drug Code 2, Drug Name 2, Manufacturer 2, Method of Administration 2, Dose Cost 2, Number of Doses 2,
 Drug Code 3, Drug Name 3, Manufacturer 3, Method of Administration 3, Dose Cost 3, Number of Doses 3,
 Drug Code 4, Drug Name 4, Manufacturer 4, Method of Administration 4, Dose Cost 4, Number of Doses 4

Tables after eliminating redundancy

- OPERATION(Hospital Number, Operation Number, Operation Code, Surgeon Name)
- SURGEON((Hospital Number, Operation Number), Surgeon Name)
- OPERATION TYPE(Operatoion Code, (Operation Name, Procedure Group))
- STANDARD DRUG DOSAGE((Drug Short Name, Size of Dose, Unit of Measure, Method of Administration), Standard Dose Cost)
- DRUG(Drug Short Name, (Drug Name, Manufacturer))
- HOSPITAL(Hospital Number, Hospital Name, Hospital Type, Teaching Status, Contact Person)
- DRUG ADMINISTRA-
 TION(Hospital Number, Operation Number, Drug Name, Drug Short Name, Size of Dose, Unit of Measure, Metho
Dose Cost, Number of Doses))

BCNF

Introduction

- A simple condition under which the update anomalies can be guaranteed not to exist.
- The Boyce-Codd Normal Form (BCNF)

BCNF

Definition

A relation R is in BCNF if and only if whenever there is a **non-trivial FD** $A_1A_2 \cdots A_n \rightarrow B$ for R , it is the case that $\{A_1A_2 \cdots A_n\}$ is a **superkey**

BCNF

Note

- Left side of every non-trivial FD must be a super key
- Superkey need not be minimal
- Equivalent statement of BCNF is that left side of every non-trivial FD must contain a key

BCNF

Note

- Alternate definition of BCNF
- Look for a set of FDs with **common left side**
- At least one of which is **non-trivial and violates the BCNF condition**

BCNF

Definition

Relation R is in BCNF if and only if whenever there is a non-trivial FD $A_1A_2 \cdots A_n \rightarrow B_1B_2 \cdots B_m$ for R , it is the case that $\{A_1, A_2, \cdots, A_n\}$ is a superkey

That is: $A_1A_2 \cdots A_n \rightarrow B_i \ \forall i = 1, 2, \cdots, m$ there must be at least one B_i is not among A_j 's

BCNF

Example - 01

Movies(title, year, length, fileType, studioName, starName)
is not in BCNF

- Any set of attributes containing these three is a **superkey**
- No set of attributes that does not include these three could be a superkey
- (title, year, starName) is the only superkey for the Movies relation
- FD: (title, year) \rightarrow length fileType studioName
- Left side is not a superkey;
- There exists an FD that violates BCNF. Therefore Movies is not in BCNF

BCNF

Example - 02

Movies1(title, year, length, fileType, studioName) is in BCNF

BCNF

Example - 03

Any two attribute relation is in BCNF

Example - 03

No non-trivial FDs That is $\{A, B\}$ is the only key. Then surely the BCNF condition must hold as only non-trivial FDs can violate this condition

$A \rightarrow B$ hold but $B \not\rightarrow A$. A is the only key. Each non-trivial FD contains A on left side. Thus there is no violation of BCNF condition

$B \rightarrow A$ hold but $A \not\rightarrow B$. Identical argument as the above

$A \rightarrow B$ & $B \rightarrow A$ Then both A and B are keys. Any FD has at least one of these on the left. So no BCNF violation

Introduction

To Multi-valued dependencies

- Concept of FDs proved to be useful in the design and analysis of relational databases
- Multi-valued dependencies are generalization of functional dependencies
- They significantly extend the understanding of logical database design
- They lead to a new fourth normal form for relational databases

Introduction

Example

S(EMPLOYEE, SALARY, CHILD)		
EMPLOYEE	SALARY	CHILD
E1	\$40K	e1_c1
E2	\$50K	e2_c1
E2	\$50K	e2_c2
E3	\$20K	e3_c1

Introduction

Example

- S obeys the TD $\text{EMPLOYEE} \rightarrow \text{SALARY}$
- That is Employee has exactly one salary
- For every pair of rows in S agreeing on EMPLOYEE also agree on SALARY

What are Multi-valued dependencies

Example

- The multivalued dependency $\text{EMPLOYEE} \twoheadrightarrow \text{SALARY}$
- Read as **EMPLOYEE multidetermines SALARY** holds for S^*
- As the FD $\text{EMPLOYEE} \rightarrow \text{SALARY}$ holds
- Similarly, $\text{EMPLOYEE} \twoheadrightarrow \text{CHILD}$ holds for S^* as the FD $\text{EMPLOYEE} \rightarrow \text{CHILD}$ holds
- Employee's **set of children is** completely determined by employee

Introduction

Example - 02

S(EMPLOYEE, CHILD, SALARY, YEAR)			
EMPLOYEE	CHILD	SALARY	YEAR
E1	e1_c1	\$35K	1976
E1	e1_c1	\$40K	1976
E2	e2_c1	\$40K	1975
E2	e2_c1	\$50K	1976
E2	e2_c2	\$40K	1975
E2	e2_c2	\$50K	1976
E3	e3_c1	\$15K	1975
E3	e3_c1	\$20K	1976

Introduction

Example - 02

- The relation $T^*(\text{EMPLOYEE}, \text{CHILD}, \text{SALARY}, \text{YEAR})$
- T^* has no functional dependencies
- However, it has multi-valued dependencies (MVD)
- MVD 01: $\text{EMPLOYEE} \twoheadrightarrow \text{CHILD}$
- MVD 02: $\text{EMPLOYEE} \twoheadrightarrow \{\text{SALARY}, \text{YEAR}\}$
- Employee's salary history is completely determined by employee alone

Introduction

Note

- MVDs provide necessary and sufficient condition for relation to be decomposable into two
- The relation $T^*(\text{EMPLOYEE}, \text{CHILD}, \text{SALARY}, \text{YEAR})$ can be decomposed into
 - T^* is in 3NF and stronger BCNF
 - As the key is set of all columns
 - However, T^* has redundancy and is not in 4NF
 - Therefore it is necessary to decompose T^*
 - $T_1^*(\text{EMPLOYEE}, \text{CHILD})$
 - $T_2^*(\text{EMPLOYEE}, \text{SALARY}, \text{YEAR})$

MVD

Notation

- Let $R(X_1, \dots, X_m, Y_1, \dots, Y_n, Z_1, \dots, Z_r)$ be a relation
- R has $(m + n + r)$ column names
- Let $\mathbf{X} = \{X_1, \dots, X_m\}$; so is \mathbf{Y}, \mathbf{Z}
- if $X_1 = x_1, \dots, X_m = x_m$ then $\mathbf{x} = (x_1, \dots, x_m)$; so is \mathbf{y}, \mathbf{z}

MVD

Define

$$Y_{xz} = \{y : (x, y, z) \in R\}$$

is non-empty if and only if x and z appear together in a tuple of R

MVD

Definition

The MVD $\mathbf{X} \twoheadrightarrow \mathbf{Y}$ holds for $R(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ if $\mathbf{Y}_{x,z}$ depends only on x
That is $\mathbf{Y}_{x,z} = \mathbf{Y}_{x,z'}$ for each x, z, z' such that $\mathbf{Y}_{x,z}$ and $\mathbf{Y}_{x,z'}$ are nonempty

MVD

S(EMPLOYEE, CHILD, SALARY, YEAR)

EMPLOYEE	CHILD	SALARY	YEAR
E1	e1.c1	\$35K	1976
E1	e1.c1	\$40K	1976
E2	e2.c1	\$40K	1975
E2	e2.c1	\$50K	1976
E2	e2.c2	\$40K	1975
E2	e2.c2	\$50K	1976
E3	e3.c1	\$15K	1975
E3	e3.c1	\$20K	1976

- $\text{EMPLOYEE} \twoheadrightarrow \text{CHILD}$ holds for T
- $\text{CHILD}_{E2, \$40K, 1975} = \{e2.c1, e2.c2\}$
- $\text{CHILD}_{E2, \$50K, 1976} = \{e2.c1, e2.c2\}$
- Therefore $\text{CHILD}_{E2, \$40K, 1975} = \text{CHILD}_{E2, \$50K, 1976}$

MVD

S(EMPLOYEE, CHILD, SALARY, YEAR)

EMPLOYEE	CHILD	SALARY	YEAR
E1	e1.c1	\$35K	1976
E1	e1.c1	\$40K	1976
E2	e2.c1	\$40K	1975
E2	e2.c1	\$50K	1976
E2	e2.c2	\$40K	1975
E2	e2.c2	\$50K	1976
E3	e3.c1	\$15K	1975
E3	e3.c1	\$20K	1976

- Delete the row (E2, e2.c2, \$50K, 1976) from T; Let this relation be T'
- Then $\text{EMPLOYEE} \twoheadrightarrow \text{CHILD}$ holds for T'
- $\text{CHILD}_{E2, \$40K, 1975} = \{e2.c1, e2.c2\}$
- $\text{CHILD}_{E2, \$50K, 1976} = \{e2.c1\}$
- Therefore $\text{CHILD}_{E2, \$40K, 1975} \neq \text{CHILD}_{E2, \$50K, 1976}$

FDs in terms of MVD

Definition

- The FD $\mathbf{X} \rightarrow \mathbf{Y}$ holds for $R(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ if $\mathbf{Y}_{x,z}$ depends only on x AND
- $\mathbf{Y}_{xz} = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}, \mathbf{z}) \in R\}$ is non-empty if and only if x and z appear together in a tuple of R AND
- The set $\mathbf{Y}_{x,z}$ is nonempty and contains **at most one member**

FDs and MVDs

Proposition - 01

If \mathbf{X} and \mathbf{Y} are disjoint, and if the FD $\mathbf{X} \rightarrow \mathbf{Y}$ holds for a relation R , then the MVD $\mathbf{X} \twoheadrightarrow \mathbf{Y}$ also holds for R

FDs and MVDs

Discussion

Proposition 01 implies that if the FD $\mathbf{X} \rightarrow \mathbf{Y}$ holds for a relation every instance of R, then the MVD $\mathbf{X} \twoheadrightarrow \mathbf{Y}$ necessarily holds

Trivial MVD

Trivial

- The MVD $\mathbf{X} \twoheadrightarrow \phi$
- The MVD $\mathbf{X} \twoheadrightarrow \mathbf{Y}$ necessarily hold for $R(\mathbf{X}, \mathbf{Y})$
- The MVD $\{A, B\} \twoheadrightarrow C$ holds for every relation $R(A, B, C)$ with exactly three columns A, B, C
- These are trivial MVDs

4th Normal Form

Definition

A relation schema R^* is in 4NF if, whenever a nontrivial MVD $\mathbf{X} \twoheadrightarrow \mathbf{Y}$ holds, then so does the *functional dependency* $\mathbf{X} \rightarrow A$ for every column A of R^* .

4th Normal Form

Definition

Intuitively all dependencies are the result of the keys

4NF and BCNF

Theorem

If a relation schema R^* is in 4NF, then it is in BCNF

4NF and BCNF

Proof - 01

- Assume R^* is in 4NF and **not** in BCNF
- As R^* **not in** BCNF, there is a nontrivial FD $\mathbf{X} \rightarrow \mathbf{Y}$
- This FD holds for R^* and there is a column A such that the FD $\mathbf{X} \not\rightarrow A$ hold

4NF and BCNF

Proof - 02

- Let $\mathbf{Y}_1 = \mathbf{Y} - \mathbf{X}$
- $\mathbf{X} \rightarrow \mathbf{Y}_1$ holds as $\mathbf{X} \rightarrow \mathbf{Y}$ holds
- \mathbf{X} and \mathbf{Y}_1 are disjoint
- The MVD $\mathbf{X} \twoheadrightarrow \mathbf{Y}_1$ holds from proposition 01

4NF and BCNF

Proof - 03

- The MVD is nontrivial
 - $\mathbf{Y}_1 \neq \phi$
 - \mathbf{X} and \mathbf{Y}_1 do not partition the columns of R^*
 - As $A \notin \mathbf{X}$ or $A \notin \mathbf{Y}_1$

4NF and BCNF

Proof - 05

- By definition of 4NF, since the MVD $\mathbf{X} \twoheadrightarrow \mathbf{Y}_1$ holds for R^*
- so does the FD $\mathbf{X} \rightarrow A$
- This is a contradiction
- Therefore the theorem holds

4NF Examples

When two (or more) many-to-many relations are placed in single table

Example - 01

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	
Smith	Type	
Smith		French
Smith		German
Smith		Greek

4NF Examples

Storage Format - Disjoint

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	
Smith	Type	
Smith		French
Smith		German
Smith		Greek

4NF Examples

Storage Format - Random Mix with minimal number of records

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	French
Smith	Type	German
Smith	Type	Greek

4NF Examples

Storage Format - Random Mix with NULL values

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	French
Smith	Type	German
Smith	⊥	Greek

4NF Examples

Storage Format - Random Mix - Unrestricted

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	French
Smith	Type	⊥
Smith	⊥	German
Smith	Type	Greek

4NF Examples

Storage Format - Cross product

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	French
Smith	Cook	German
Smith	Cook	Greek
Smith	Type	French
Smith	Type	German
Smith	Type	Greek

4NF Examples

When two (or more) many-to-many relations are placed in single table

Example - 01

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	
Smith	Type	
Smith		French
Smith		German
Smith		Greek

- MVD: EMPLOYEE \twoheadrightarrow SKILL
- MVD: EMPLOYEE \twoheadrightarrow LANGUAGE

4NF Examples

Example - 01

Decompose the original relation into two

- EMP_SKILL(EMPLOYEE, SKILL)
- EMP_LANGUAGE(EMPLOYEE, LANGUAGE)

4NF Examples

Example - 02

R^* (class, section, student, major, exam, year, instructor, rank, salary, text, day, room)

4NF Examples

FDs

FD #	X	→	Y
F1	{class, section}	→	instructor
F2	{class, section, day}	→	room
F3	student	→	{major, year}
F4	instructor	→	{rank, salary}

MVDs

MVD #	X	→→	Y
M1	{class, section}	→→	{student, major, exam, year}
M2	{class, section}	→→	{instructor, rank, salary}
M3	{class, section}	→→	text
M4	{class, section}	→→	{day, room}
M5	{class, section, student}	→→	exam
M6	class	→→	text

4NF Examples - 02

R^* (class, section, student, major,
exam, year, instructor, rank, salary, text,
day, room)
Use

R_1^* (class, section, student, major, exam,
year)

R_2^* (class, section, instructor, rank, salary,
text, day, room)

MVD #	X	$\rightarrow\rightarrow$	Y
M1	{class, section}	$\rightarrow\rightarrow$	{student, major, exam, year}

to decompose R^*

4NF Examples - 02

R_1^* (class, section, student, major, exam, year)

R_2^* (class, section, instructor, rank, salary, text, day, room)

Use

MVD #	X		Y
M5	{class, section, student}	$\rightarrow\rightarrow$	exam

to decompose R_1^*

R_{11}^* (class, section, student, exam)

R_{12}^* (class, section, student, major, year)

R_2^* (class, section, instructor, rank, salary, text, day, room)

4NF Examples - 02

R_{11}^* (class, section, student, exam)

R_{12}^* (class, section, student, major, year)

R_2^* (class, section, instructor, rank, salary,
text, day, room)

Use

FD #	X	→	Y
F3	student	→	{major, year}

to decompose R_{12}^*

R_{11}^* (class, section, student, exam)

R_{121}^* (student, major, year)

R_{122}^* (class, section, student)

R_2^* (class, section, instructor, rank, salary,
text, day, room)

4NF Examples - 02

R_{11}^* (class, section, student, exam)

R_{121}^* (student, major, year)

R_{122}^* (class, section, student)

R_2^* (class, section, instructor, rank, salary,
text, day, room)

Use

R_{11}^* (class, section, student, exam)

R_{121}^* (student, major, year)

R_{122}^* (class, section, student)

R_{21}^* (class, section, instructor, rank, salary)

R_{22}^* (class, section, text, day, room)

MVD #	X	$\rightarrow\rightarrow$	Y
M2	{class, section}	$\rightarrow\rightarrow$	{instructor, rank, salary}

to decompose R_2^*

4NF Examples - 02

R_{11}^* (class, section, student, exam)

R_{121}^* (student, major, year)

R_{122}^* (class, section, student)

R_{21}^* (class, section, instructor, rank, salary)

R_{22}^* (class, section, text, day, room)

Use

FD #	X	→	Y
F4	instructor	→	{rank, salary}

to decompose R_{21}^*

R_{11}^* (class, section, student, exam)

R_{121}^* (student, major, year)

R_{122}^* (class, section, student)

R_{211}^* (instructor, rank, salary)

R_{212}^* (class, section, instructor)

R_{22}^* (class, section, text, day, room)

4NF Examples - 02

R_{11}^* (class, section, student, exam)

R_{121}^* (student, major, year)

R_{122}^* (class, section, student)

R_{211}^* (instructor, rank, salary)

R_{212}^* (class, section, instructor)

R_{22}^* (class, section, text, day, room)

Use

MVD #	X	$\rightarrow\rightarrow$	Y
M6	{class, section, student}	$\rightarrow\rightarrow$	exam

to decompose R_{22}^*

R_{11}^* (class, section, student, exam)

R_{121}^* (student, major, year)

R_{122}^* (class, section, student)

R_{211}^* (instructor, rank, salary)

R_{212}^* (class, section, instructor)

R_{221}^* (class, text)

R_{222}^* (class, section, day, room)

4NF Examples - 02

R_{11}^* (class, section, student, exam)

R_{121}^* (student, major, year)

R_{122}^* (class, section, student)

R_{211}^* (instructor, rank, salary)

R_{212}^* (class, section, instructor)

R_{22}^* (class, section, text, day, room)

Remove R_{122}^* as it is part of R_{11}^*

R_{11}^* (class, section, student, exam)

R_{121}^* (student, major, year)

R_{211}^* (instructor, rank, salary)

R_{212}^* (class, section, instructor)

R_{221}^* (class, text)

R_{222}^* (class, section, day, room)