

Normalization -- Practice

EGCI 321: LECTURE 13 (WEEK 9)

Redundancy

Redundancy is at the root of several problems associated with relational schemas:

- Redundant storage, insert/delete/update anomalies
- Integrity constraints, in particular *functional dependencies*, can be used to *identify schemas with* such problems and to suggest refinements
- Main refinement technique: *decomposition* (replacing ABCD with, say, AB and BCD, or ACD and ABD)

Data redundancy and update anomalies

Tables that contain redundant information may potentially **suffer from update anomalies**.

Types of update anomalies include

- Insertion
- Deletion
- Modification

Data redundancy and update anomalies

| staffNo | name | position | salary | branchNo | branchAddress | telNo |
|---------|---------------|------------|--------|----------|--------------------------------------|--------------|
| S1500 | Tom Daniels | Manager | 46000 | B001 | 8 Jefferson Way, Portland, OR 97201 | 503-555-3618 |
| S0003 | Sally Adams | Assistant | 30000 | B001 | 8 Jefferson Way, Portland, OR 97201 | 503-555-3618 |
| S0010 | Mary Martinez | Manager | 50000 | B002 | City Center Plaza, Seattle, WA 98122 | 206-555-6756 |
| S3250 | Robert Chin | Supervisor | 32000 | B002 | City Center Plaza, Seattle, WA 98122 | 206-555-6756 |
| S2250 | Sally Stern | Manager | 48000 | B004 | 16 – 14th Avenue, Seattle, WA 98128 | 206-555-3131 |
| S0415 | Art Peters | Manager | 41000 | B003 | 14 – 8th Avenue, New York, NY 10012 | 212-371-3000 |

Data redundancy and update anomalies (cont.)

Staff

| staffNo | name | position | salary | branchNo |
|---------|---------------|------------|--------|----------|
| S1500 | Tom Daniels | Manager | 46000 | B001 |
| S0003 | Sally Adams | Assistant | 30000 | B001 |
| S0010 | Mary Martinez | Manager | 50000 | B002 |
| S3250 | Robert Chin | Supervisor | 32000 | B002 |
| S2250 | Sally Stern | Manager | 48000 | B004 |
| S0415 | Art Peters | Manager | 41000 | B003 |

Branch

| branchNo | branchAddress | telNo |
|----------|--------------------------------------|--------------|
| B001 | 8 Jefferson Way, Portland, OR 97201 | 503-555-3618 |
| B002 | City Center Plaza, Seattle, WA 98122 | 206-555-6756 |
| B003 | 14 – 8th Avenue, New York, NY 10012 | 212-371-3000 |
| B004 | 16 – 14th Avenue, Seattle, WA 98128 | 206-555-3131 |

Functional Dependencies

Definition

- $x \rightarrow y$
- x functionally determines y in a relation R
- **OR** y is functionally dependent on x
- If each x -value in the relation is associated with only one y -value at any one time
- x and y may be composite attributes
- y may be associated with more than one x

Functional Dependencies

- Functional Dependency (FD)
 - Tool for analyzing relation schemas
 - A constraint among attributes in a relation based on the semantics of attributes
 - Identified by the database designer from a relation schema (not a relation state)
 - Used to determine what normal form a relation schema is in

Functional Dependencies

■ Example1 (2NF)

- EMP_PROJ (SSN, PNO, HOURS, ENAME, PNAME, PLOCATION)
- $SSN \rightarrow ENAME, PNO$
- $PNO \rightarrow PNAME, PLOCATION$
- $\{SSN, PNO\} \rightarrow HOURS$

■ Example2

- STUDENT_COURSE Relation
(SSN, COURSEID, SECTIONID, YEAR SNAME, SADDRESS, STATUS, CNAME, CDESC, GRADE)
- Functional Dependencies:
 $SSN \rightarrow SNAME, SADDRESS, STATUS$
 $\{COURSEID, YEAR\} \rightarrow CNAME, CDESC$
 $\{SSN, COURSEID, SECTIONID, YEAR\} \rightarrow GRADE$

Functional Dependencies

Transitive Functional Dependency

- y is transitively functionally dependent on x if x functionally determines z (not a candidate key or a subset) and z functionally determines y
- $x \rightarrow y$ if $x \rightarrow z$ and $z \rightarrow y$
- e.g. $SSN \rightarrow PNO$ and $PNO \rightarrow PLOCATION$, then $SSN \rightarrow PLOCATION$ (SSN transitively determines $PLOCATION$)

Multi-Valued (Is it 1NF?)

Staff

| <u>ENO</u> | Name | Dno | DeptName | ProjNo | ProjName |
|-------------------|-------------|------------|------------------|---------------|----------------------------------|
| E001 | Somchai | D01 | Physic | P01, P02 | NMR, Laser |
| E002 | Sompong | D01 | Physic | P03 | Medical Image processing |
| E003 | Somchay | D02 | Computer Science | P04, P05 | Voice ordering, Speech Coding |
| E004 | SomSiri | D02 | Computer Science | P04, P06 | Voice ordering, Speech Synthesis |

Multi-Valued (cont.)

| <u>ENO</u> | Name | Dno | DeptName | <u>ProjNo</u> | ProjName |
|------------|---------|-----|------------------|---------------|--------------------------|
| E001 | Somchai | D01 | Physic | P01 | NMR |
| E001 | Somchai | D01 | Physic | P02 | Laser |
| E002 | Sompong | D01 | Physic | P03 | Medical Image processing |
| E003 | Somchay | D02 | Computer Science | P05 | Voice ordering |
| E003 | Somchay | D02 | Computer Science | P04 | Speech Coding |
| E004 | SomSiri | D02 | Computer Science | P04 | Speech Coding |
| E004 | SomSiri | D02 | Computer Science | P06 | Speech Synthesis |

Insert a new project but there is still has problem


Second Normal Form (2NF)

A relation is in first normal form if and only if

- It is in 1NF
- Every **non-key** attribute is dependent on **all parts of the primary key**.

2NF ?

Staff



| <u>ENO</u> | Name | Dno | DeptName | <u>ProjNo</u> | ProjName |
|------------|---------|-----|------------------|---------------|--------------------------|
| E001 | Somchai | D01 | Physic | P01 | NMR |
| E001 | Somchai | D01 | Physic | P02 | Laser |
| E002 | Sompong | D01 | Physic | P03 | Medical Image processing |
| E003 | Somchay | D02 | Computer Science | P05 | Voice ordering |
| E003 | Somchay | D02 | Computer Science | P04 | Speech Coding |
| E004 | SomSiri | D02 | Computer Science | P04 | Speech Coding |
| E004 | SomSiri | D02 | Computer Science | P06 | Speech Synthesis |



KEY = ENO + ProjNo

Answer is No. Because
ProjName is dependent on
ProjNo. (not all part of Key)

Problem

| <u>ENO</u> | Name | Dno | DeptName | <u>ProjNo</u> | ProjName |
|------------|---------|-----|------------------|---------------|--------------------------|
| E001 | Somchai | D01 | Physic | P01 | NMR |
| E001 | Somchai | D01 | Physic | P02 | Laser |
| E002 | Sompong | D01 | Physic | P03 | Medical Image processing |
| E003 | Somchay | D02 | Computer Science | P05 | Voice ordering |
| E003 | Somchay | D02 | Computer Science | P04 | Speech Coding |
| E004 | SomSiri | D02 | Computer Science | P04 | Speech Coding |
| E004 | SomSiri | D02 | Computer Science | P06 | Speech Synthesis |

We can not insert Project if they have not yet assigned project to any employee

Solution

Remove the attribute involved

Take its determinant with it

Normalize

| <u>E_{NO}</u> | Name | D _{no} | DeptName | <u>ProjNo</u> | ProjName |
|-----------------------|---------|-----------------|------------------|---------------|--------------------------|
| E001 | Somchai | D01 | Physic | P01 | NMR |
| E001 | Somchai | D01 | Physic | P02 | Laser |
| E002 | Sompong | D01 | Physic | P03 | Medical Image processing |
| E003 | Somchay | D02 | Computer Science | P05 | Voice ordering |
| E003 | Somchay | D02 | Computer Science | P04 | Speech Coding |
| E004 | SomSiri | D02 | Computer Science | P04 | Voice ordering |
| E004 | SomSiri | D02 | Computer Science | P06 | Speech Synthesis |

Result: 2NF

PERSON

| <u>ENO</u> | Name | Dno | DeptName |
|------------|---------|-----|------------------|
| E001 | Somchai | D01 | Physic |
| E003 | Somchay | D02 | Computer Science |
| E004 | SomSiri | D02 | Computer Science |

PROJECT

| <u>ProjNo</u> | ProjName |
|---------------|--------------------------|
| P01 | NMR |
| P02 | Laser |
| P03 | Medical Image processing |
| P04 | Speech Coding |
| P05 | Voice ordering |
| P06 | Speech Synthesis |

PERSON_PROJECT

| <u>ENO</u> | <u>ProjNo</u> |
|------------|---------------|
| E001 | P01 |
| E001 | P02 |
| E002 | P03 |
| E003 | P04 |
| E003 | P05 |
| E004 | P04 |
| E004 | P06 |

PERSON(ENO,NAME,Dno,DeptName)

PROJECT(ProjNo,ProjName)

PERSON_PROJ(ENO,ProjNo)

Third Normal Form

A relation is in 3NF if, and only if:


- It is in 2NF
- Every non-key attribute is functionally dependent upon the key (**No non-key attribute is functional dependent on another non-key attribute**)
- Or non-key attribute no **transitive dependent** on key

Transitive dependent

- $R(\underline{A}, B, C, D)$; A is Key, others are non- key
- If $A \rightarrow B$ and $B \rightarrow C$ can say
 $A \rightarrow B \rightarrow C$ (C transitive dependent on A)

3NF?

PERSON



| <u>ENO</u> | Name | Dno | DeptName |
|------------|---------|-----|------------------|
| E001 | Somchai | D01 | Physic |
| E003 | Somchay | D02 | Computer Science |
| E004 | SomSiri | D02 | Computer Science |

Answer is No
Because DeptName is dependent on Dno
(has transitive dependent on key)

PROJECT

| <u>ProjNo</u> | ProjName |
|---------------|--------------------------|
| P01 | NMR |
| P02 | Laser |
| P03 | Medical Image processing |
| P04 | Speech Coding |
| P05 | Voice ordering |
| P06 | Speech Synthesis |

PERSON_PROJECT

| <u>ENO</u> | <u>ProjNo</u> |
|------------|---------------|
| E001 | P01 |
| E001 | P02 |
| E002 | P03 |
| E003 | P04 |
| E003 | P05 |
| E004 | P04 |
| E004 | P06 |

Solution

- Remove the offending attributes
- Take the determinant along

Result

PERSON

| <u>ENO</u> | Name | Dno |
|------------|---------|-----|
| E001 | Somchai | D01 |
| E003 | Somchay | D02 |
| E004 | SomSiri | D02 |

DEPARTMENT

| <u>Dno</u> | DeptName |
|------------|------------------|
| D01 | Physic |
| D02 | Computer Science |

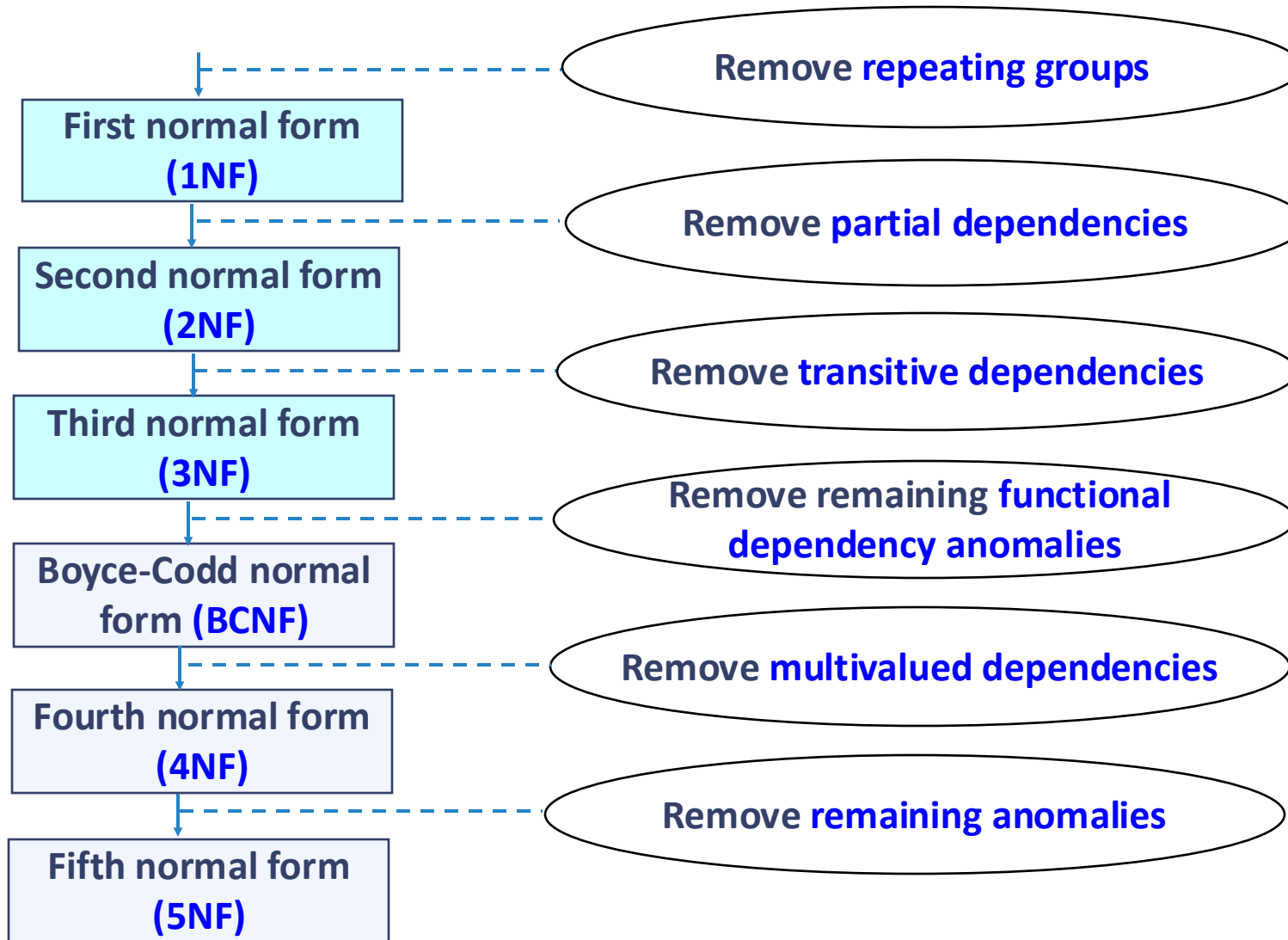
PROJECT

| <u>ProjNo</u> | ProjName |
|---------------|--------------------------|
| P01 | NMR |
| P02 | Laser |
| P03 | Medical Image processing |
| P04 | Speech Coding |
| P05 | Voice ordering |
| P06 | Speech Synthesis |

PERSON_Proj

| <u>ENO</u> | <u>ProjNo</u> |
|------------|---------------|
| E001 | P01 |
| E001 | P02 |
| E002 | P03 |
| E003 | P04 |
| E003 | P05 |
| E004 | P04 |
| E004 | P06 |

Stages of Normalisation



BCNF (Boyce-Codd Normal Form)

A relation schema R is in **Boyce-Codd Normal Form (BCNF)** if whenever an FD $X \rightarrow A$ holds in R , then X is a superkey of R

- Each normal form is strictly stronger than the previous one:
 - ▶ Every 2NF relation is in 1NF
 - ▶ Every 3NF relation is in 2NF
 - ▶ Every BCNF relation is in 3NF
- There exist relations that are in 3NF but not in BCNF
- The goal is to have each relation in BCNF (or 3NF)

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 |

BCNF

3NF Functional Dependencies

- $\{\text{Student}, \text{Course}\} \rightarrow \text{Instructor}$
- $\text{Instructor} \rightarrow \text{Course}$
- Decomposing into 2 schemas of BCNF (Which one is good?)
 - 1) $\{\underline{\text{Student}}, \text{Instructor}\} \{\underline{\text{Student}}, \text{Course}, \text{Instructor}\}$
 - 2) $\{\underline{\text{Instructor}}, \text{Course}\} \{\underline{\text{Student}}, \text{Course}, \text{Instructor}\}$
 - 3) $\{\underline{\text{Instructor}}, \text{Course}\} \{\underline{\text{Instructor}}, \text{Student}\}$

Reference

1. Ramakrishnan R, Gehrke J., Database management systems, 3rd ed., New York (NY): McGraw-Hill, 2003.