International University, VNU-HCMC

School of Computer Science and Engineering

Lecture 8: Normalization

Instructor: Nguyen Thi Thuy Loan

nttloan@hcmiu.edu.vn, nthithuyloan@gmail.com
https://nttloan.wordpress.com/

Acknowledgement

- The following slides are references from Northeastern University.
- Other slides have been created based on the Database system concepts book, 7th Edition.

International University, VNU-HCMC

Recap: Lecture 7

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Functional Dependencies
- Keys/Super keys
- Attribute closure
- Minimal cover

4

International University, VNU-HCMC

Purpose of the Lecture

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Understand why normalization is needed in database design.
- Learn the different normal forms (1NF, 2NF, 3NF, BCNF, 4NF).
- Identify and eliminate data anomalies (insertion, update, deletion).
- Apply normalization techniques to improve data consistency and reduce redundancy.

5

Warm-up question

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

If you design a table for storing student information (Name, ID, Course, Instructor, Instructor Phone), what kind of problems might happen when:

- A student takes multiple courses,
- An instructor changes their phone number, or
- A course has no enrolled students?

Outline

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Introduction
- Data Anomalies
- Concept of Normal Forms
- Examples
- Advantages of Normalization
- Limitations and Trade-offs

International University, VNU-HCMC

Introduction

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

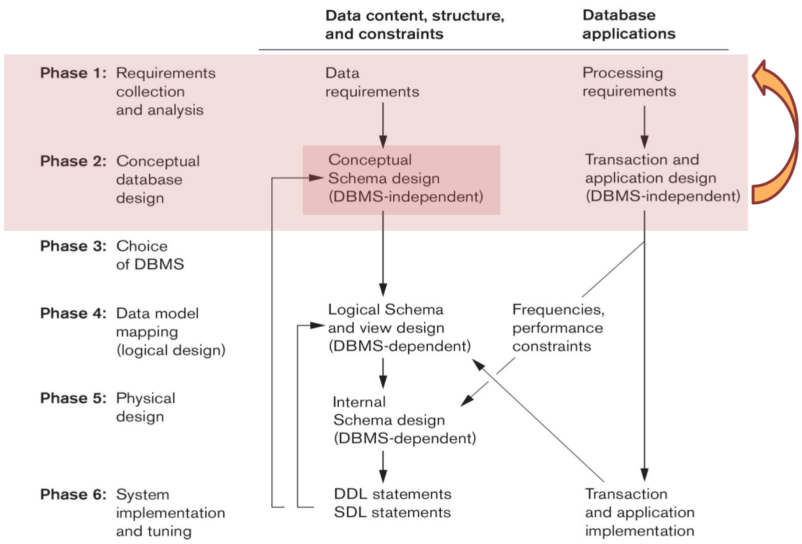
- Databases must be carefully structured to avoid problems.
- Poor design leads to data redundancy and anomalies (insertion, update, deletion).
- Normalization is the process of organizing data into well-structured tables.
- Goal: ensure consistency, efficiency, and integrity of data.

8

International University, VNU-HCMC

Database Design and Implementation Process

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

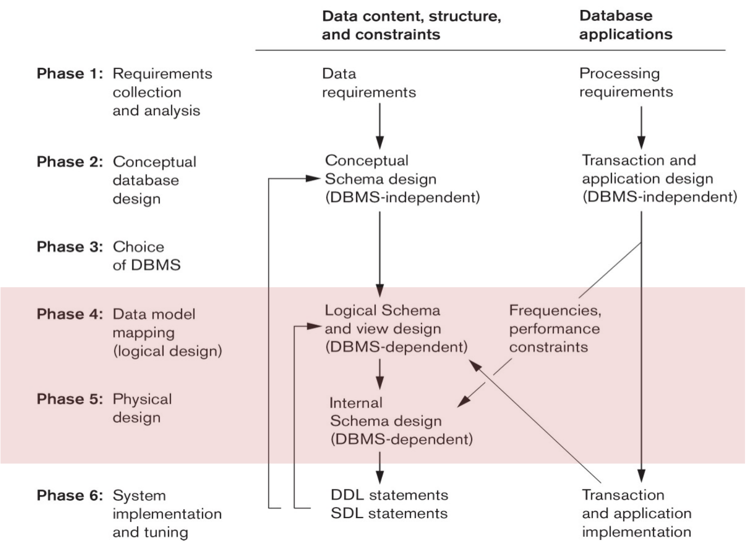


9

International University, VNU-HCMC

Database Design and Implementation Process

Assoc. Prof. Nguyen Thi Thuy Loan, PhD



International University, VNU-HCMC

Data Anomalies

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Redundancy

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

International University, VNU-HCMC

Data Anomalies

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

EMPLOYEE

Ename	Ssn	Bdate	Address	Dnumber
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4
Wallace, Jennifer S.	987654321	1941-06-20	291Berry, Bellaire, TX	4
Narayan, Ramesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1

DEPARTMENT

Dname	Dnumber	Dmgr_ssn
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

International University, VNU-HCMC

Data Anomalies

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Poorly designed tables cause problems:

- Insertion Anomaly: Cannot add new data without having to include unnecessary or missing information.
- Update Anomaly: Need to change the same data in multiple places; if missed, data becomes inconsistent.
- Deletion Anomaly: Removing one record may unintentionally remove valuable related data.

International University, VNU-HCMC

Make the Schema Clear and Simple

- Design tables so their meaning is easy to understand.
- Do not mix attributes from different entities or relationships into a single table; this can cause confusion and ambiguity.
- Normalized tables and their relationships accurately reflect real-world concepts and their connections.

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

14

International University, VNU-HCMC

Example

What is this table about?

- Employees? Departments?

Redundancy

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

15

International University, VNU-HCMC

Reduce Data Redundancy

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Redundancy

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

16

International University, VNU-HCMC

Reduce Data Redundancy

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Avoid repeating the same information across multiple rows (duplication).
- Too much duplication leads to:
  - More storage use
  - Inconsistencies when data changes
  - Difficult maintenance
- Avoid excessive NULL values (in wide “fat” tables):
  - Wastes space
  - Makes queries more complicated and more error-prone

17

International University, VNU-HCMC

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Insertion Anomaly

Redundancy

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

- Sometimes it is difficult or impossible to add new data.

Examples:

- Adding a new employee without knowing the manager or department can result in blocked or incorrect data entry.
- Adding a new department requires at least one employee and cannot be stored alone.

18

International University, VNU-HCMC

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Update Anomaly

Redundancy

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

- Updating data in one place may require multiple changes elsewhere.
- If not all rows are updated, it leads to inconsistent information.
- Example: Changing the department name or manager must be updated in every related employee record.

19

International University, VNU-HCMC

Deletion Anomaly

Redundancy

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

- Deleting one fact may cause loss of other valuable information.
- Example: If we delete employee James E. Borg, we also lose the record of the Headquarters department (Dept 1)

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

20

International University, VNU-HCMC

Avoid Spurious Tuples

- Ensure that joins are made only on valid key relationships (Primary Key and Foreign Key).
- Joining on unrelated attributes can create false or meaningless rows (spurious tuples).
- Good schema design prevents these errors.

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

21

International University, VNU-HCMC

Bad Decomposition

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

CAR

ID	Make	Color
1	Toyota	Blue
2	Audi	Blue
3	Toyota	Red

CAR1

ID	Color
1	Blue
2	Blue
3	Red

CAR2

Make	Color
Toyota	Blue
Audi	Blue
Toyota	Red

Association between Color and Make is lost.

22

International University, VNU-HCMC

Bad decomposition

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

ID	Make	Color
1	Toyota	Blue
1	Audi	Blue
2	Toyota	Blue
2	Audi	Blue
3	Toyota	Red

CAR1

ID	Color
1	Blue
2	Blue
3	Red

CAR2

Make	Color
Toyota	Blue
Audi	Blue
Toyota	Red

Join returns more rows than the original relation

23

International University, VNU-HCMC

Additive Decomposition

CAR

ID	Make	Color
1	Toyota	Blue
2	Audi	Blue
3	Toyota	Red

JOIN

ID	Make	Color
1	Toyota	Blue
1	**Audi**	**Blue**
2	**Toyota**	**Blue**
2	Audi	Blue
3	Toyota	Red

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

24

International University, VNU-HCMC

Concept of Normal Forms

- Normal forms are rules to structure relations properly.
- Each higher normal form removes more redundancy and anomalies.
- Goal: Ensure data consistency and simpler queries.

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

25

First Normal Form (1NF)

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- No repeating groups or arrays.
- Each cell must hold a single atomic value.
- Each row is unique.

Example: A student table should not store multiple phone numbers in one column.

International University, VNU-HCMC

Second Normal Form (2NF)

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Must first be in 1NF.
- Remove partial dependencies (no attribute should depend on part of a composite key).

Example: In Enrollment(<u>StudentID, CourseID</u>, Grade),

StudentID, CourseID → Grade

⇒ Grade should rely on StudentID and CourseID.

International University, VNU-HCMC

Third Normal Form (3NF)

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Must first be in 2NF.
- Remove transitive dependencies (non-key attributes should depend only on the key, not on other non-key attributes).

Example:

COURSE(<u>CourseID</u>, CourseName)

CourseID → CourseName

⇒ The CourseName depends on the CourseID.

International University, VNU-HCMC

Boyce–Codd Normal Form (BCNF)

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Stronger version of 3NF.
- For every functional dependency X → Y, X must be a superkey.
- Eliminates anomalies not covered by 3NF.

Example: C_Textbook(<u>Textbook</u>, Course)

Textbook → Course

International University, VNU-HCMC

Summary of Normal Forms

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- 1NF: Eliminate repeating groups.
- 2NF: Eliminate partial dependencies.
- 3NF: Eliminate transitive dependencies.
- BCNF: Every determinant is a superkey.

30

International University, VNU-HCMC

Fourth Normal Form (4NF)

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- A relation is in 4NF if:
- It is already in Boyce–Codd Normal Form (BCNF).
- It has no non-trivial multivalued dependencies (MVDs) other than a candidate key.

Meaning:

- An attribute should not determine independent sets of values within the same relation.
- If multivalued dependencies exist, they cause redundancy and should be separated into different tables.

31

International University, VNU-HCMC

Fourth Normal Form (4NF)

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Relation: COURSE(CourseID, Instructor, Textbook)
  - One course can have multiple instructors.
  - One course can also have multiple textbooks.
  - If stored together, every combination of instructor and textbook is listed, resulting in redundancy.
- Solution (4NF decomposition):
  - COURSE_INSTRUCTOR(CourseID, Instructor)
  - COURSE_TEXTBOOK(CourseID, Textbook)

32

International University, VNU-HCMC

Lossless join decomposition

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Decompose relation R into relations S and T
- Attrs(R) = attrs (S) ∪ attrs(T)
  - S = $\pi_{attrs(S)}(R)$
  - T = $\pi_{attrs(T)}(R)$
- The decomposition is a lossless join decomposition if, given known constraints such as FD's, we can guarantee that R = S $\bowtie T$

33

1NF – First Normal Form

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- A relation is in first normal form if every attribute in every row can contain only one single (atomic) value.

Examples: 1NF?

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Student(FirstName, LastName, Knowledge)

FirstName	LastName	Knowledge
Thomas	Mueller	Java, C++, PHP
Ursula	Meier	PHP, Java
Igor	Mueller	C++, Java

Problem:

- Attribute Knowledge contains multiple values, not atomic.
- Therefore, the relation is not in First Normal Form (1NF).

Examples: 1NF violation

FirstName	LastName	Knowledge
Thomas	Mueller	Java, C++, PHP
Ursula	Meier	PHP, Java
Igor	Mueller	C++, Java

FirstName	LastName	Knowledge
Thomas	Mueller	Java
Thomas	Mueller	C++
Thomas	Mueller	PHP
Ursula	Meier	PHP
Ursula	Meier	Java
Igor	Mueller	C++
Igor	Mueller	Java

Examples: 1NF?

- Assume a video library maintains a database of rented movies.
- Without normalization, all details are stored in one table:

Full names	Physical address	Movies rented	Salutation
Janet Jones	First Street Plot No 4	Pirates of the Caribbean; Clash of the Titans	Ms.
Robert Phil	3rd street 34	Forgetting Sarah Marshal; Daddy's Little Girls	Mr.
Robert Phil	5th Avenue	Clash of the Titans	Mr.

International University, VNU-HCMC

Examples: 1NF

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

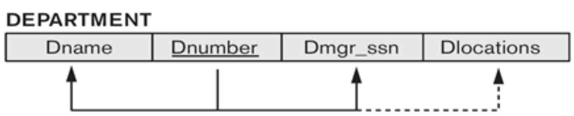
Full names	Physical address	Movies rented	Salutation
Janet Jones	First Street Plot No 4	Pirates of the Caribbean	Ms.
Janet Jones	First street Plot No 4	Clash of the Titans	Ms.
Robert Phil	3rd street 34	Forgetting Sarah Marshal	Mr.
Robert Phil	3rd Street 34	Daddy’s Little Girls	Mr.
Robert Phil	5th Avenue	Clash of the Titans	Mr

38

International University, VNU-HCMC

Examples 1NF?

Assoc. Prof. Nguyen Thi Thuy Loan, PhD



(b)

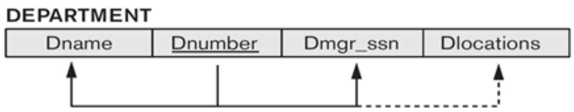
DEPARTMENT

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

39

1NF Violation

Assoc. Prof. Nguyen Thi Thuy Loan, PhD



(b)

DEPARTMENT

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

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

Important FD Definitions

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Trivial FD	$X \rightarrow Y, Y \subseteq X$
Non-prime attribute	An attribute that does not occur in any key (opposite: Prime)
Full FD	$X \rightarrow Y, \forall A \in X((X - \{A\}) \nrightarrow Y)$
Transitive FD	$X \rightarrow Y$ and $Y \rightarrow Z \therefore X \rightarrow Z$

International University, VNU-HCMC

Example 2NF?

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

StudentID	Course	StudentAddress
1	COMP570	555 Huntington
1	COMP285	555 Huntington
2	COMP570	610 Huntington
3	COMP355	Louis Prang
3	COMP553	Louis Prang

{StudentID, Course}→{StudentAddress}
{StudentID}→ {StudentAddress}

StudentID	StudentAddress
1	555 Huntington
2	610 Huntington
3	Louis Prang

StudentID	Course
1	COMP570
1	COMP285
2	COMP570
3	COMP355
3	COMP553

International University, VNU-HCMC

Examples 2NF?

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Students(IDSt, StudentName, IDProf, ProfessorName, Grade)
  F={IDProf → ProfessorName; IDSt → StudentName; IDSt, IDProf →Grade}
  The attributes IDSt and IDProf are the identification keys.

Students

IDSt	StudentName	IDProf	ProfessorName	Grade
1	Mueller	3	Schmid	5
2	Meier	2	Borner	4
3	Tobler	1	Bernasconi	3

Examples 2NF?

- All attributes a single valued (1NF).

Students

IDSt	StudentName
1	Mueller
2	Meier
3	Tobler

Professors

IDProf	ProfessorName
1	Bernasconi
2	Borner
3	Schmid

Grade

IDSt	IDProf	Grade
1	3	5
2	2	4
3	1	6

International University, VNU-HCMC

Examples 2NF?

- A school wants to store data of teachers and the subjects they teach.
- One teacher can teach multiple subjects, so the table repeats teacher information:

Teacher

Teacher_id	Subject	Teacher_age
111	Maths	38
111	Physics	38
222	Biology	38
333	Physics	40
333	Chemistry	40

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Examples 2NF?

Teacher(Teacher_id, Subject, Teacher_age)

F={Teacher_id, Subject → Teacher_age; Teacher_id → Teacher_age}

- Only key is: {Teacher_id, Subject}

Teacher

Teacher_id	Subject	Teacher_age
111	Maths	38
111	Physics	38
222	Biology	38
333	Physics	40
333	Chemistry	40

International University, VNU-HCMC

Examples 2NF?

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- To make the table complies with 2NF we can break it in two tables like this.

Teacher

Teacher_id	Teacher_age
111	38
222	38
333	40

Teacher_Subject

Teacher_id	Subject
111	Maths
111	Physics
222	Biology
333	Physics
333	Chemistry

2NF Can Suffer Update Anomalies

Year	Winner	Nationality
1994	Miguel Indurain	Spain
1995	Miguel Indurain	Spain
1996	Bjarne Riis	Denmark
1997	Jan Ullrich	Germany

- Relation is in 2NF?
  - Trivially true (why?)
- List all non-trivial FDs for this relation state

  {Year}→{Winner, Nationality}

  {Winner} →{Nationality}
- What if we insert (1998, Jan Ullrich, USA)?

Exercise 2NF?

Patients(StaffNo, ApptDate, ApptTime, DentistName, PatientNo, PatientName, SurgeryNo)

F={StaffNo, ApptDate, ApptTime → PatientNo, PatientName; StaffNo →DentistName; PatientNo → PatientName,SurgeryNo; StaffNo, ApptDate → SugeryNo; ApptDate, ApptTime, PatientNo →StaffNo, DentistName}

Exercise 2NF?

R(ABCDEGH)

F = {ABC → EG, A → D, E → GH, AB → H, BCE → AD}

SA: BC

IA: AE

Keys: ABC, BCE

R1(AD) F1 = {A → D}

R2(ABCEGH)

F2 = {ABC → EG, E → GH, AB → H, BCE → A}

Keys: ABC, BCE

Exercise 2NF?

R2(ABCEGH)

F2 = {ABC → EG, E → GH, AB → H, BCE → A}

Keys: ABC, BCE

R21(EGH) F21 = {E → GH}

R22(ABCE) F22 = {ABC → E}

3NF – Third Normal Form

- It is already in 2NF
- Every non-key attribute depends directly on the key, not on another non-key attribute.
- Rule of Thumb:
  "A non-key attribute must depend on the key, the whole key, and nothing but the key."

Fix:

- Decompose the table into smaller relations.
- Remove transitive dependencies (when one non-key attribute determines another).

International University, VNU-HCMC

3NF – Third Normal Form

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

A table is in 3NF if:

- It is already in 2NF, and
- For every functional dependency $X \rightarrow Y$, at least one of these holds:
  - X is a superkey of the table
  - Y is a prime attribute (part of some candidate key)

Note:

- A prime attribute = an attribute that belongs to key(s).
- A non-prime attribute = an attribute not in any key(s).

International University, VNU-HCMC

3NF Example

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

F={Year → Winner, Nationality; Winner → Nationality}

Year	Winner	Nationality
1994	Miguel Indurain	Spain
1995	Miguel Indurain	Spain
1996	Bjarne Riis	Denmark
1997	Jan Ullrich	Germany

Year	Winner
1994	Miguel Indurain
1995	Miguel Indurain
1996	Bjarne Riis
1997	Jan Ullrich

Winner	Nationality
Miguel Indurain	Spain
Bjarne Riis	Denmark
Jan Ullrich	Germany

56

International University, VNU-HCMC

Examples: 3NF

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Suppose a company wants to store the complete address of each employee, they create a table named employee_details that looks like this:

Employees(Emp_id, Emp_Name, Emp_zip, Emp_state, Emp_city, Emp_district)

F={Emp_zip → Emp_state, Emp_city, Emp_district; Emp_id → Emp_zip; Emp_id → Emp_Name}

Only key is {Emp_id}

57

International University, VNU-HCMC

Examples: 3NF

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Non-prime attributes (Emp_state, Emp_city, & emp_district) are transitively dependent on the key (Emp_id) through Emp_zip. This violates the rule of 3NF.

To make the table comply with 3NF we decompose it into two tables to remove the transitive dependency:

Employees_zip (Emp_zip, Emp_state, Emp_city, Emp_district)

F1={Emp_zip → Emp_state, Emp_city, Emp_district}

Employees (Emp_id, Emp_name, Emp_zip)

F2 ={Emp_id →Emp_zip, Emp_Name}

58

International University, VNU-HCMC

Examples: 3NF

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

A bank uses the following relation:

Vendors(ID, Name, Account_No, Bank_Code_No, Bank)

F={ID → Name, Account_No, Bank_Code_No; Bank_Code_No → Bank}

Only key is {ID}

59

Examples: 3NF

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Non-prime attributes (Bank) is transitively dependent on key (ID) through Bank_Code_No. This violates the rule of 3NF.
- To make the table comply with 3NF we decompose it into two tables to remove the transitive dependency:
- Vendors1(Bank_Code_No, Bank)

  F1={Bank_Code_No $\rightarrow$ Bank}

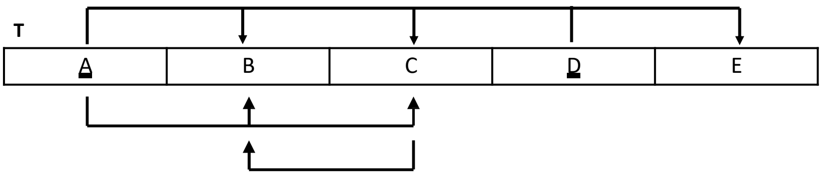
- Vendors2(ID, Name, Account_No, Bank_Code_No)

  F2={ID $\rightarrow$ Name, Account_No, Bank_Code_No}

International University, VNU-HCMC

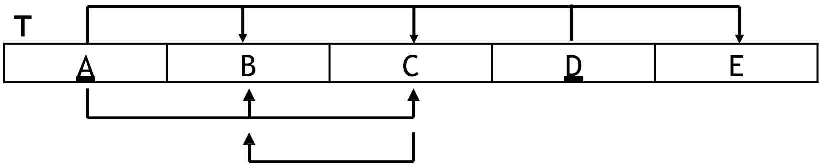
Exercises: 3NF

Assoc. Prof. Nguyen Thi Thuy Loan, PhD



- Consider the schema for relation T, as well as all FDs.
- Determine the highest normal form of relation T.
- If T is not in 3NF, decompose it into relations that:
  - Satisfy all functional dependencies
  - Preserve the primary key
  - Avoid spurious tuples (lossless join)
- Show and explain the steps of your analysis and decomposition.

Answer (1)



List non-trivial FDs

$AD \rightarrow BCE$

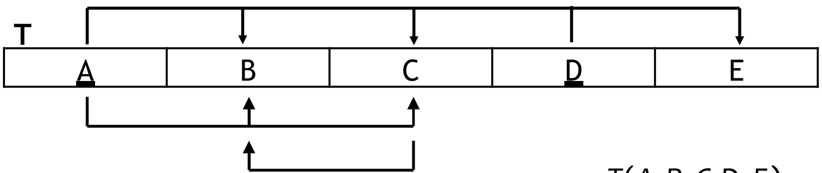
$A \rightarrow BC$

$C \rightarrow B$

Written algebraically

$T(\underline{A}, B, C, \underline{D}, E)$

Answer (2)



T is in … 1NF

- Both B & C are FD on A
  - Thus not fully FD on PK (AD)

Decompose!

$T(\underline{A}, B, C, \underline{D}, E)$

$AD \rightarrow BCE$

$A \rightarrow BC$

$C \rightarrow B$

International University, VNU-HCMC

Answer (3)

T1

<u>A</u>	<u>D</u>	E

T2

<u>A</u>	B	C

$T1(\underline{A}, \underline{D}, E)$
$T2(\underline{A}, B, C)$
$AD \rightarrow E$
$A \rightarrow BC$
$C \rightarrow B$

T1 is in… .3NF
- 2NF: E is fully FD on AD
- 3NF: No transitive FDs (trivially true)

T2 is in … 2NF
- 2NF: B and C fully FD on A (trivially true)
- !3NF: B is transitively FD on A [via C]

Decompose!

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

64

International University, VNU-HCMC

Answer (4)

T1

<u>A</u>	<u>D</u>	E

T2_1

<u>A</u>	C

T2_2

<u>C</u>	B

$T1(\underline{A}, \underline{D}, E)$
$T2\_1(\underline{A}, C)$
$T2\_2(\underline{C}, B)$
$AD \rightarrow E$
$A \rightarrow C$
$C \rightarrow B$

Database is in 3NF
- Why?

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

65

Answer (5)

TSupplies

SupplierID (A)	Status (B)	City (C)	PartID (D)	Qty (E)

Supplier_Parts

SupplierID	PartID	Qty

Suppliers

SupplierID	City

Cities

City	Status

$\{SupplierID, PartID\} \rightarrow \{Qty\}$
$\{SupplierID\} \rightarrow \{City\}$
$\{City\} \rightarrow \{Status\}$

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Boyce-Codd Normal Form (BCNF)

- BCNF is an advanced version of 3NF.
- A relation is in BCNF if:
  - It is already in 3NF, and
  - For every functional dependency $X \rightarrow Y$, X must be a superkey of the table.

Key Idea:

- In BCNF, every determinant is a superkey.

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

International University, VNU-HCMC

Boyce-Codd Normal Form (BCNF)

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

A relation R is in BCNF if:

- For every non-trivial functional dependency $X \rightarrow Y$ that holds in R, X must be a superkey.

Notes:

- Non-trivial FD: Y is not part of X.
- Superkey: Any set of attributes that uniquely identifies a tuple (may include extra attributes).

68

International University, VNU-HCMC

Examples: BCNF

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Drinkers(<u>name</u>, addr, <u>beersLiked</u>, manf, favBeer)

FD's: F= {name → addr, favBeer;
beersLiked → manf}

Only key is {name, beersLiked}

- In both FDs, the left-hand side is not a superkey.
- Therefore, these dependencies violate BCNF.

Conclusion:

- The relation Drinkers is not in BCNF.

69

International University, VNU-HCMC

Another Example

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Beers(name, manf, manfAddr)

FD's: F={name → manf, manf → manfAddr}

Only key is {name}

Name → manf does not violate BCNF, but manf → manfAddr does

70

International University, VNU-HCMC

Decomposition into BCNF

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Given: A relation R with functional dependencies F.
- Step 1: Check the given FDs for a BCNF violation.
  - For any FD X → Y, if X is not a superkey, it violates BCNF.
- Step 2: Compute X^+ (closure of X).
  - If X^+ does not contain all attributes of R, then X is not a superkey, confirming a violation.
- Step 3: Decompose R into two relations:
  - R1 = X ∪ Y
  - R2 = R – (Y – X)
- Repeat until all relations are in BCNF.

71

International University, VNU-HCMC

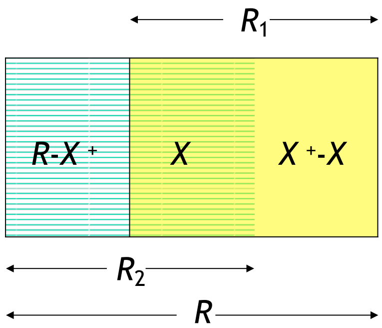
Decompose *R* Using X → Y

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Replace *R* by relations with schemas:

1. $R_1 = X^+$
2. $R_2 = R - (X^+ - X)$

Project given FD's *F* onto the two new relations



72

International University, VNU-HCMC

Examples: BCNF?

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Let's take R = {A,B,C,D,E,G} and

  F = {BC → D, CD → E}

73

Examples: BCNF?

- Let's take R = {A,B,C,D,E,G} and F = {BC → D, CD → E}
- Candidate Key: {A, B, C, G}

Step 1: Pick a violating FD

- Use FD BC → D (BC is not a superkey).

Step 2: Compute closure

- X = {BC}, X^+ = {B, C, D, E}

Step 3: Decompose

- R_1 = {B, C, D, E}
- R_2 = {A, B, C, G}

Step 4: Check intersection

- $R_1 \cap R_2$ = {B, C} = X

This ensures a lossless decomposition.

Examples: BCNF Decomposition

Drinkers(<u>name, beersLiked</u>, addr, manf, favBeer)

F= {name → addr, name → favBeers,
beersLiked → manf}

International University, VNU-HCMC

Examples: BCNF Decomposition

Drinkers(<u>name, beersLiked</u>, addr, manf, favBeer)

$F = \{name \rightarrow addr,\ name \rightarrow favBeers,$
$beersLiked \rightarrow manf\}$

Step 1: Identify BCNF violation

- FD name→ addr violates BCNF (since name is not a superkey)

Step 2: Compute closure

- $\{name\}^+ = \{name, addr, favBeer\}$

Step 3: Decompose relation

- Drinkers1(<u>name</u>, addr, favBeer)
  $F1 = \{name \rightarrow addr,\ name \rightarrow favBeers\}$
- Drinkers2(<u>name, beersLiked</u>, manf)
  $F2 = \{beersLiked \rightarrow manf\}$

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

76

International University, VNU-HCMC

Examples: BCNF Decomposition

- After decomposition, we must check both Drinkers1 and Drinkers2 for BCNF.
- Projecting FDs onto each relation:

For Drinkers1(name, addr, favBeer):

- $F1 = \{ name \rightarrow addr,\ name \rightarrow favBeer \}$
- key = {name}

Since {name} is a key, Drinkers1 is in BCNF.

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

77

International University, VNU-HCMC

Examples: BCNF Decomposition

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

For Drinkers2(name, beersLiked, manf):

- F2 = {beersLiked → manf}
- key: {name, beersLiked}

Check BCNF:

- {beersLiked} is not a superkey → violation of BCNF.

Closure:

- $\{beersLiked\}^+ = \{beersLiked, manf\}$

Decompose Drinkers2:

- Drinkers3(beersLiked, manf)
- F3 = {beersLiked → manf}
- Drinkers4(name, beersLiked); F4 = {∅}

Now all relations are in BCNF

78

International University, VNU-HCMC

Examples: BCNF Decomposition

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- The resulting decomposition of Drinkers:
- Drinkers1(name, addr, favBeer)
- Drinkers3(beersLiked, manf)
- Drinkers4(name, beersLiked)
- Interpretation:
- Drinkers1: stores information about drinkers.
- Drinkers3: stores information about beers and their manufacturers.
- Drinkers4: stores the relationship between drinkers and the beers they like
- Compare with running example:
  1. Drinkers(name, addr, phone)
  2. Beers(name, manf)
  3. Likes(drinker,beer)

79

Exercises: BCNF

- Suppose there is a company wherein employees work in more than one department. They store the data like this:
- Employees (Emp_id, Emp_Nationality, Emp_Dept, Dept_type, Dept_no_of_emp)
- F = {Emp_id → Emp_Nationality; Emp_Dept → Dept_type, Dept_no_of_emp}
- Only key is {Emp_id, Emp_dept}

International University, VNU-HCMC

BCNF– Motivation

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

There is one structure of FD's that causes trouble when we decompose

$AB \rightarrow C$ and $C \rightarrow B$

Example:

A = street address; B = city; C = post code

There are two keys, $\{A,B\}$ and $\{A,C\}$

$C \rightarrow B$ is a BCNF violation, so we must decompose into AC, BC

International University, VNU-HCMC

We Cannot Enforce FD's

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

The problem is that if we use *AC* and *BC* as our database schema, we cannot enforce the FD $AB \rightarrow C$ by checking FD's in these decomposed relations

Example with *A* = street, *B* = city, and *C* = post code on the next slide

83

International University, VNU-HCMC

An Unenforceable FD

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

street	post
Campusvej	5230
Vestergade	5000

city	post
Odense	5230
Odense	5000

Join tuples with equal post codes

street	city	post
Campusvej	Odense	5230
Vestergade	Odense	5000

No FD's were violated in the decomposed relations and FD street, city → post holds for the database as a whole

84

An Unenforceable FD

street	post
Hjallesevej	5230
Hjallesevej	5000

city	post
Odense	5230
Odense	5000

Join tuples with equal post codes

street	city	post
Hjallesevej	Odense	5230
Hjallesevej	Odense	5000

Although no FD's were violated in the decomposed relations, FD street, city → post is violated by the database as a whole.

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Another Unenforceable FD

Departures(time, track, train)

F={time, track → train; train → track}

Two keys, {time,track} and {time,train}

train → track is a BCNF violation, so we must decompose into

Departures1(time, train)
Departures2(track,train)

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Another Unenforceable FD

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

time	train
19:08 19:16	ICL54 IC852

tracktrain	
4 3	ICL54 IC852

Join tuples with equal train code

time	track	train
19:08 19:16	4 3	ICL54 IC852

No FD's were violated in the decomposed relations,
FD time, track → train holds for the database as a whole

International University, VNU-HCMC

Another Unenforceable FD

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

time	train
19:08 19:08	ICL54 IC 42

Tracktrain	Train
4 4	ICL54 IC 42

Join tuples with equal train code

time	track	train
19:08 19:08	4 4	ICL54 IC 42

Although no FD's were violated in the decomposed relations, FD time, track → train is violated by the database as a whole.

International University, VNU-HCMC

Examples: Decomposition into BCNF

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

1. Let's take R(ABCDE), and FD's
F = {A → BC, C → DE}

2. Let's take R(ABCD) and FD's
F = {AB → C, B → D; C → A}

89

International University, VNU-HCMC

Multivalued dependencies

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- A multivalued dependency (MVD) has the from X ↠ Y, where X and Y are sets of attributes in a relation R.
- X ↠ Y means that whenever two rows in R agree on all the attributes of X, then we can swap their Y components and get two rows that are also in R

X	Y	Z
a	b1	c1
a	b2	c2
a	b2	c1
a	b1	c2
…	…	…

92

International University, VNU-HCMC

MVD examples

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

User (uid, gid, place)
- uid ↠ gid
- uid ↠ place
  - Intuition: given uid, gid, and place are "independent"
- uid, gid ↠ place
  - Trivial: **LHS ∪ RHS = all attributes of R**
- uid, gid ↠ uid
  - Trivial: **LHS ⊇ RHS**

93

International University, VNU-HCMC

Complete MVD + FD rules

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

FD reflexivity, augmentation, and transitivity
- MVD complementation:
  - If $X \twoheadrightarrow Y$, then $X \twoheadrightarrow attrs\ R - X - Y$
- MVD augmentation:
  - If $X \twoheadrightarrow Y$ and $V \subseteq W$, then $XW \twoheadrightarrow YV$
- MVD transitivity:
  - If $X \twoheadrightarrow Y$ and $Y \twoheadrightarrow Z$, then $X \twoheadrightarrow Z - Y$

94

International University, VNU-HCMC

Complete MVD + FD rules

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Replication (FD is MVD):
  - If $X \rightarrow Y$, then $X \twoheadrightarrow Y$
- Coalescence:
  - If $X \twoheadrightarrow Y$ and $Z \subseteq Y$ and there is some W disjoint from Y such that $W \rightarrow Z$, then $X \rightarrow Z$

95

International University, VNU-HCMC

An elegant solution: chase

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Given a set of FD's and MVD's $\mathcal{D}$, does another dependency d (FD or MVD) follow from $\mathcal{D}$?

Procedure

- Start with the "if-part" of d, and treat them as "seed" tuples in a relation
- Apply the given dependencies in $\mathcal{D}$ repeatedly
  - If we apply an FD, we infer equality of two symbols
  - If we apply an MVD, we infer more tuples

96

An elegant solution: chase

- If we infer the "then-part" of d, we have a proof
- Otherwise, if nothing more can be inferred, we have a counter example

International University, VNU-HCMC

Proof by chase

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- In R(A, B, C, D), does A ↠ B and B ↠ C imply that A ↠ C?

Have:

	A	B	C	D
	a	b_1	c_1	d_1
	a	b_2	c_2	d_2
$A \twoheadrightarrow B$	a	b_2	c_1	d_1
	a	b_1	c_2	d_2
$B \twoheadrightarrow C$	a	b_2	c_1	d_2
	a	b_2	c_2	d_1
$B \twoheadrightarrow C$	a	b_1	c_2	d_1
	a	b_1	c_1	d_2

Need:

A	B	C	D	
a	b_1	c_2	d_1	✌
a	b_2	c_1	d_2	✌

International University, VNU-HCMC

Another proof by chase

- In R(A, B, C, D), does A → B and B → C imply that A → C?

Have:

A	B	C	D
a	b_1	c_1	d_1
a	b_2	c_2	d_2

Need: $c_1 = c_2$ ✌

A → B $\quad$ $b_1 = b_2$

B → C $\quad$ $c_1 = c_2$

- In general, with both MVD's and FD's, chase can generate both new tuples and new equlities

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

99

International University, VNU-HCMC

Counterexample by chase

- In R(A, B, C, D), does A ↠ BC and CD → B imply that A → B?

Have:

A	B	C	D
a	b_1	c_1	d_1
a	b_2	c_2	d_2

$A \twoheadrightarrow BC$

A	B	C	D
a	b_2	c_2	d_1
a	b_1	c_1	d_2

Need: $b_1 = b_2$ 👎

Counterexample!

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

100

International University, VNU-HCMC

4NF

A relation R is in Fourth Normal Form (4NF) if

- For every non-trivial MVD $X \twoheadrightarrow Y$ in R, X is a superkey
- That is, all FD's and MVD's follow from "key → other attributes" (i.e., no MVD's and no FD's besides key functional dependencies)
- 4NF is stronger than BCNF, because every FD is also an MVD

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

101

International University, VNU-HCMC

4NF decomposition algorithm

- Find a 4NF violation: A non-trivial MVD $X \twoheadrightarrow Y$ in R where X is not a superkey
- Decompose R into R_1 and R_2, where
  - R1 has attributes X Y
  - R2 has attributes X Z (where Z contains R attributes not in X or Y)
- Repeat until all relations are in 4NF
- Almost identical to BCNF decomposition algorithm
- Any decomposition on a 4NF violation is lossless

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

102

4NF decomposition example

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

uid	gid	place
142	dps	Springfield
142	dps	Australia
456	abc	Springfield
456	abc	Morocco
456	gov	Springfield
456	gov	Morocco
…	…	…

User (*uid*, *gid*, *place*)

4NF violation: *uid* ↠ *gid*

Member (*uid*, *gid*)

4NF

uid	gid
142	dps
456	abc
456	gov
…	…

Visited (*uid*, *place*)

4NF

uid	place
142	Springfield
142	Australia
456	Springfield
456	Morocco
…	…

International University, VNU-HCMC

Summary

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

- Philosophy behind BCNF, 4NF: Data should depend on the key, the whole key, and nothing but the key!
  - You could have multiple keys though
- Other normal forms
  - 3NF: More relaxed than BCNF; will not remove redundancy if doing so makes FDs harder to enforce
  - 2NF: Slightly more relaxed than 3NF
  - 1NF: All column values must be atomic

Summary

- Normalization is the theory and process by which to evaluate and improve relational database design
  - Makes the schema informative
  - Minimizes information duplication
  - Avoids modification anomalies
  - Disallows spurious tuples
- Make sure all your relations are *at least* 3NF!
  - Higher normal forms exist
  - We may reduce during physical design

International University, VNU-HCMC

Assoc. Prof. Nguyen Thi Thuy Loan, PhD

Thank you for your attention!