

Third Normal Form (3NF)

Dr. Bing Zhou

Third Normal Form - Motivation

- $R(A, B, C)$
 $AB \rightarrow C$ and $C \rightarrow B$.
 - Example: A = street address, B = city, C = zipcode.
- What is the key?
 - There are two keys, $\{A, B\}$ and $\{A, C\}$.
- These (and similar) structures of FD's cause trouble when we decompose.
- $C \rightarrow B$ is a BCNF violation, so we must decompose into AC, BC .

We Cannot Enforce FD's

- 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 $A = \text{street}$, $B = \text{city}$, and $C = \text{zip}$ on the next slide.

An Unenforceable FD

street A	zip C
545 Tech Sq.	02138
545 Tech Sq.	02139

city B	zip C
Cambridge	02138
Cambridge	02139

Join tuples with equal zip codes.

street A	city B	zip C
545 Tech Sq.	Cambridge	02138
545 Tech Sq.	Cambridge	02139

Although no FD's were violated in the decomposed relations,
FD street city \rightarrow zip (AB \rightarrow C) is violated by the database as a whole.

Preserving FDs in a Decomposition

- ▶ Consider the relation Teach(CourseNumber, DepartmentName, Professor, Semester, Year)
- ▶ The relation models which courses a professor teaches in which semester.
- ▶ Do not assume that each course is taught by at most one professor.
- ▶ University introduces two new rules:
 1. Each professor teaches ≤ 1 course per semester. $PSY \rightarrow CD$
 2. Each course is taught either in the fall every year or in the spring every year. $CD \rightarrow S$
- ▶ What are the keys? $\{P, S, Y\}$ and $\{C, D, P, Y\}$
- ▶ Decomposing using $CD \rightarrow S$ yields Teach1(C, D, S) and Teach2(C, D, P, Y). Both are in BCNF?
- ▶ How do you enforce $PSY \rightarrow CD$?
- ▶ The BCNF decomposition algorithm does not preserve FDs.

Third Normal Form (3NF)

- ▶ A relation R is in *Third Normal Form* (3NF) if and only if for every non-trivial FD $A_1 A_2 \dots A_n \rightarrow B$ for R , one of the following two conditions is true:
 1. $\{A_1, A_2, \dots, A_n\}$ is a superkey for R or
 2. B is an attribute in some key.
- ▶ Teach(C, D, P, S, Y) has FDs $PSY \rightarrow CD$ and $CD \rightarrow S$
- ▶ Keys are $\{P, S, Y\}$ and $\{C, D, P, Y\}$.
- ▶ $CD \rightarrow S$ violates BCNF.
- ▶ However, Teach is in 3NF because S is a part of a key.

3NF Let's Us Avoid the FD Problem

- 3rd Normal Form (3NF) modifies the BCNF condition so we do not have to decompose in this problem situation.
- An attribute is *prime* if it is a member of any key.
- $X \rightarrow A$ violates 3NF if and only if X is not a superkey, and also A is not prime.

Example

- In our problem situation with FD's $AB \rightarrow C$ and $C \rightarrow B$, we have keys AB and AC .
- Thus A , B , and C are each prime.
- Although $C \rightarrow B$ violates BCNF, it does not violate 3NF.

What 3NF and BCNF Give You

- There are two important properties of a decomposition:
 1. *Recovery* : it should be possible to project the original relations onto the decomposed schema, and then reconstruct the original.
 2. *Dependency preservation* : it should be possible to check in the projected relations whether all the given FD's are satisfied.

3NF and BCNF, Continued

- We can get (1) with a BCNF decomposition.
- We can get both (1) and (2) with a 3NF decomposition.
- But we can't always get (1) and (2) with a BCNF decomposition.
 - street-city-zip is an example.

Decomposition into 3NF

1. We can always decompose a relational schema into a set of schemas that are *dependency-preserving*, i.e.,
 - ▶ the decomposition is lossless-join,
 - ▶ each resulting relation is in 3NF, and
 - ▶ for each FD, there is a relation that allows that FD to be checked.
2. However, the relations are not in BCNF and contain some redundancy.

First Normal Form

- Table faithfully represents a relation, primarily meaning it has at least one candidate key, and each attribute is atomic

Name	Gender	Contact	Interest
Neil	M	Email:neil@ee.net,phone:1222456	Reading;Guitar
Devin	M	Email:studyzy@163.net,phone:13934563456	Swimming
Neil	M	Email:neil@ee.net,phone:1222456	Reading;Guitar

UserId	Name	Gender	Email	Phone	Interest
1	Neil	M	neil@ee.net	1222456	Reading;Guitar
2	Devin	M	studyzy@163.net	13934563456	Swimming

First Normal Form

- Table faithfully represents a relation, primarily meaning it has at least one candidate key, and each attribute is atomic

UserId	Name	Gender	Email	Phone
1	Neil	M	neil@ee.net	1222456
2	Devin	M	studyzy@163.net	13934563456

UserId	Interest
1	Reading
1	Guitar
2	Swimming

Second Normal Form

- It has to be in the first normal form + No non-prime attribute in the table is functionally dependent on a proper subset of any candidate key

StudentId	CourseId	ChooseTime	ConfirmTime	CourseName
1	10	2013/8/26	2013/8/27	微积分
1	11	2013/8/27	2013/8/27	线性代数
2	10	2013/8/26	2013/8/27	微积分

StudentId	CourseId	ChooseTime	ConfirmTime
1	10	2013/8/26	2013/8/27
1	11	2013/8/27	2013/8/27
2	10	2013/8/26	2013/8/27

CourseId	CourseName
10	微积分
11	线性代数

Third Normal Form

- It has to be in the second normal form + the left hand side is a super key OR the right hand side is prime

StudentId	Name	DepartmentId	DepartmentName
1	Neil	21	Math
2	Devin	22	Computer

- StudentId -> Name DepartmentID DepartmentName
- DepartmentId -> DepartmentName
- T1={DepartmentId, DepartmentName}
- T2={StudentID, Name, DepartmentID}

BCNF

- The left hand side of every non-trivial functional dependency must be a superkey

Fourth Normal Form

- The left hand side of every non-trivial Multi-valued dependency must be a superkey

Fifth Normal Form

- Beyond the scope of this class
- Every relation can be further decomposed into smaller relations unless the decomposed relation will have the same key as the original table

kNFs

- ▶ First Normal Form: each attribute is atomic.
- ▶ Second Normal Form: No non-trivial FD has a left side that is a proper subset of a key.
- ▶ Third Normal Form: we just discussed it.
- ▶ Fourth Normal Form: we just discussed it.
- ▶ Fifth Normal Form: outside the scope
- ▶ Sixth Normal Form: different versions exist. One version is newly developed for temporal databases.
- ▶ Seventh Normal Form: your ticket to fame and fortune.

Apply(SSN, cName, date, major)

Can apply to each college once for one major ✓

Colleges have non-overlapping application dates ✓

FDs: SSN, cName → date, major date → cName

Keys: SSN, cName

BCNF: No. A1 (date, cName) A2 (SSN, date, major) (?)

Good design? Not necessarily. 3rd Normal Form