Third Normal Form (3NF)

Zaki Malik

October 23, 2008

Third Normal Form - Motivation

- *R (A, B, C)*

  AB ->*C* and *C* ->*B*.

  – Example: *A* = street address, *B* = city, *C* = zipcode.
- What is the key?

  – There are two keys, {*A*,*B* } and {*A*,*C* }.
- *C* ->*B* is a BCNF violation, so we must decompose into *AC*, *BC*.
- These (and similar) structures of FD's cause trouble when we decompose.

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* ->*C* by checking FD's in these decomposed relations.
- Example with *A* = street, *B* = city, and *C* = zip on the next slide.

An Unenforceable FD

street **A**	zip **C**
545 Tech Sq.	02138
565 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
565 Tech Sq.	Cambridge	02139

Although no FD's were violated in the decomposed relations,
FD street city -> zip (AB ->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. $\mathtt{PSY} \rightarrow \mathtt{CD}$
  2. Each course is taught either in the fall every year or in the spring every year. $\mathtt{CD} \rightarrow \mathtt{S}$
- What are the keys? $\{P, S, Y\}$ and $\{C, D, P, Y\}$
- Decomposing using $\mathtt{CD} \rightarrow \mathtt{S}$ yields `Teach1(C, D, S)` and `Teach2(C, D, P, Y)`. Both are in BCNF?
- How do you enforce $\mathtt{PSY} \rightarrow \mathtt{CD}$?
- 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 $A_1\ A_2\ \ldots\ A_n \rightarrow B$ for R, one of the following two conditions is true:
  1. $\{A_1, A_2, \ldots, A_n\}$ 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* ->*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* ->*C* and *C* ->*B*, we have keys *AB* and *AC*.
- Thus *A*, *B*, and *C* are each prime.
- Although *C* ->*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 decompsition.
  - Explanation needs to wait for relational algebra.
- 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.
3. What is the decomposition algorithm? Outside the scope of CS4604.

Sample Problems (handout)

Zaki Malik

October 23, 2008