Schema Decomposition (II)

R&G Chapter 19

Announcement

- **Assignment 4 in Canvas**
  - Schema refinement
  - Due at 11:59pm, Dec 12.

Last Lecture: A Good Decomposition

- **A good decomposition is**
  - Lossless
  - Dependency preserving

In Last Lecture

- Lossless decomposition
  - The decomposition of R into X and Y is lossless with respect to F *if and only if* the closure of F contains:

$$X \cap Y \rightarrow X, \quad \textbf{or}$$
$$X \cap Y \rightarrow Y$$

- BCNF decomposition

Quiz of Last Lecture

Consider the relation R={CSJDPQV}:

–Its primary key is C;

–It has the following FDs: JP → C, SD → P, J → S.

- **Several dependencies may cause violation of BCNF. The order in which we ``deal with" them could lead to very different sets of relations!**

–We just tried the order of SD → P, J -> S, and JP ->C

–Now try starting from JP → C, then J → S, at last SD → P

Problem #2 of Decomposition

A	B	C
1	2	3
4	5	6
7	2	8

A	C
1	3
4	6
7	8

B	C
2	3
5	6
2	8

$A \rightarrow B; C \rightarrow B$

$X=\{A, C\}, Y=\{B, C\}, X \cap Y = \{C\}, \quad C\text{->}\{B, C\}$

Lossless decomposition!

A	C
1	3
4	6
7	8

Join

B	C
2	3
5	6
2	8

=

A	B	C
1	2	3
4	5	6
7	2	8

But, now we can't check $A \rightarrow B$ without doing a join!
(Problem #2 of decomposition!)

Task #2

- **How to decompose the table so that dependency checking does not need joins?**

Dependency Preserving Decomposition

- ***Projection of set of FDs F*** **:** If R is decomposed into X and Y, the projection of F on X (denoted F_X) is the set of FDs $U \rightarrow V$ in F^+ (*not F!*) such that all of the attributes U, V are in X. *(same holds for Y of course)*
  - E.g., $F^+ = \{A\text{->}B , A\text{->}C, D\text{->}E\}$
  - Two tables: X={A, B, C}, Y={C, D, E}
  - $F_X = \{A\text{->}B , A\text{->}C\}$,
  - $Fy = \{D\text{->} E\}$.

Dependency Preserving Decompositions (Contd.)

- **Decomposition of R into X and Y is *dependency preserving* if $(F_X \cup F_Y)^+ = F^+$**
  - i.e., consider only dependencies in the closure F^+ that can be checked in X without considering Y, and in Y without considering X, these imply all dependencies in F^+.

Example 1

- **R=ABC, F={A $\rightarrow$ B, B $\rightarrow$ C, C $\rightarrow$ A},**
- **R is decomposed into X={A,B} and Y={B,C}.**
- **Is this decomposition dependency preserving?**
- **Way of thinking:**
  - F^+= {A->B, B->C, C->A, A->C, B->A, C->B}
  - F_x={A->B, B->A}
  - F_y={B->C, C->B}
  - (Fx $\cup$ Fy $)^+$={A->B, B→C, B->A, C->B, A->C, C->A}
  - (Fx $\cup$ Fy $)^+$= F+, so it is dependency preserving.

Example 2

- **R(A, B, C, D, E)**
- **F={AB->C, C->E, B->D, E->A}**
- **R is decomposed into X=(B, C, D) and Y=(A, C, E)**
- **Is the decomposition dependency preserving?**
- **Way of thinking:**
  - F^+={AB->C, C->E, B->D, E->A, AB->E, C->A, AB->A}
  - F_x={B->D}
  - F_y={C->E, E->A, C->A}
  - $(Fx \cup Fy)^+$={B->D, C->E, E->A, C->A}
  - $(Fx \cup Fy)^+ != F^+$, so it is NOT dependency preserving.

Decomposition into 3NF

- The algorithm for BCNF decomposition does not ensure dependency preservation (it only assures lossless join)

- To ensure dependency preservation (and lossless-join)
  - Instead of the given set of FDs F, use a *minimal cover for F*.

Minimal Cover for a Set of FDs

- **G is the *minimal cover* of a set of FDs F if:**
  1. Right hand side (RHS) of each FD in G is a single attribute;
  2. $F^+ = G^+$; and
  3. G is minimal: if we modify G by deleting an FD in G, G^+ changes.

- **Intuitively, every FD in G is needed, and ``*as small as possible*'' in order to get the same closure as F.**

Method of Finding Minimal Cover

- **Three steps:**
  - Step 1: Change right-hand-side (RHS) of FDs so that they only contain single attributes
    - Change X->YZ to be X->Y and X-Z
  - Step 2: Minimize left-hand-side (LHS) if possible
    - If A->B, then replace ABX->Z with AX->Z
    - If AB->C and A->C, then remove AB->C and keep A->C.
  - Step 3: Remove redundant FDs
    - If X->Y can be inferred from other FDs, remove X->Y

Example 1

- **R(ABCDE)**
- **F={A->D, BC->AD, C->B, E->A, E->D}**
- **What is the minimal cover of F?**
- **Way of thinking:**
  - Step 1: Change RHS to be single attribute:
    - Split BC->AD to BC->A, BC->D
  - Step 2: Minimize LHS:
    - Since C->B, replace BC->A with C->A, and BC->D with C->D
  - Step 3: Remove redundant FDs
    - C->A, A->D: C->D, so remove C->D
    - E->A, A->D: E->D, so remove E->D
  - Minimal cover: {A->D, B->D, C->B, E->A}

Example 2

- FDs: {A $\to$ B, ABCD $\to$ E, EF $\to$ GH, ACDF $\to$ EG}
- What's its minimal cover?
  - Step 1: rewrite RHS
    - Rewrite ACDF $\to$ EG to ACDF $\to$ E and ACDF $\to$ G
    - Rewrite EF $\to$ GH to EF $\to$ G and EF $\to$ H
  - Step 2: minimize LHS
    - Since A->B, replace ABCD $\to$ E with ACD $\to$ E
  - Step 3: remove redundant FDs
    - ACDF $\to$ G is implied by ACD $\to$ E and EF $\to$ G. So delete it.
    - Similarly, delete ACDF $\to$ E.
  - The minimal cover: {A $\to$ B, ACD $\to$ E, EF $\to$ G, EF $\to$ H}