Schema Refinement: Other Dependencies and Higher Normal Forms

Grant Weddell

Cheriton School of Computer Science
University of Waterloo

CS 348
Introduction to Database Management
Winter 2017

Outline

1. Multivalued Dependencies
   - Reasoning about MVDs
   - Lossless-Join Decompositions
   - Fourth Normal Form
2. Other Dependencies

Beyond Functional Dependencies

There exist anomalies/redundancies in relational schemas that cannot be captured by FDs.

Example: consider the following table:

Course	Teacher	Book
Math	Smith	Algebra
Math	Smith	Calculus
Math	Jones	Algebra
Math	Jones	Calculus
Advanced Math	Smith	Calculus
Physics	Black	Mechanics
Physics	Black	Optics

There are no (non-trivial) FDs that hold on this scheme; therefore the scheme (Course, Set-of-teachers, Set-of-books) is in BCNF.

Multivalued Dependencies (MVD)

- CTB table contains redundant information because:

  whenever $(c, t_1, b_1) \in CTB$ and $(c, t_2, b_2) \in CTB$
  then also $(c, t_1, b_2) \in CTB$
  and, by symmetry, $(c, t_2, b_1) \in CTB$
- we say that a **multivalued dependency** (MVD)

  $C \twoheadrightarrow T$ (and $C \twoheadrightarrow B$ as well)

  holds on CTB.

> given a course, the set of teachers and the set of books are uniquely determined and independent.

Another Example

Course	Teacher	Hour	Room	Student	Grade
CS101	Jones	M-9	2222	Smith	A
CS101	Jones	W-9	3333	Smith	A
CS101	Jones	F-9	2222	Smith	A
CS101	Jones	M-9	2222	Black	B
CS101	Jones	W-9	3333	Black	B
CS101	Jones	F-9	2222	Black	B

- FDs:
  $C \rightarrow T$, $CS \rightarrow G$, $HR \rightarrow C$, $HT \rightarrow R$, and $HS \rightarrow R$
- MVDs:
  $C \twoheadrightarrow HR$

Axioms for MVDs

1. $Y \subset X \Rightarrow X \twoheadrightarrow Y$ (reflexivity)
2. $X \twoheadrightarrow Y \Rightarrow X \twoheadrightarrow (R - Y)$ (complementation)
3. $X \twoheadrightarrow Y \Rightarrow XZ \twoheadrightarrow YZ$ (augmentation)
4. $X \twoheadrightarrow Y, Y \twoheadrightarrow Z \Rightarrow X \twoheadrightarrow (Z - Y)$ (transitivity)
5. $X \rightarrow Y \Rightarrow X \twoheadrightarrow Y$ (conversion)
6. $X \twoheadrightarrow Y, XY \rightarrow Z \Rightarrow X \rightarrow (Z - Y)$ (interaction)

Theorem:

Axioms for FDs (1)-(6) are sound and complete for logical implication of FDs and MVDs.

Example

In the $CTHRSG$ schema, $C \twoheadrightarrow SG$ can be derived as follows:

1. $C \twoheadrightarrow HR$
2. $C \twoheadrightarrow T$ (from $C \rightarrow T$)
3. $C \twoheadrightarrow CTSG$ (complementation of (1))
4. $C \twoheadrightarrow CT$ (augmentation of (2) by C)
5. $CT \twoheadrightarrow CTSG$ (augmentation of (3) by T)
6. $C \twoheadrightarrow SG$ (transitivity on (4) and (5))

Dependency Basis

Definition:

A **dependency basis** for X with respect to a set of FDs and MVDs F is a partition of $R - X$ to sets $Y_1, \ldots, Y_k$ such that $F \models X \twoheadrightarrow Z$ if and only if $Z - X$ is a union of some of the Y_is.

- unlike for FDs we can't split right-hand sides of MVDs to single attributes (cf. minimal cover).
- the dependency basis of X w.r.t. F can be computed in PTIME [Beeri80].
- The dependency basis of $CTHRSG$ with respect to C is $[T, HR, SG]$

Lossless-Join Decomposition

- similarly to the FD case we want to decompose the schema to avoid anomalies

  $\Rightarrow$ a lossless-join decomposition (R_1, R_2) of R
  with respect to a set of **MVDs** F:

  $$F \models (R_1 \cap R_2) \twoheadrightarrow (R_1 - R_2)$$

  or, by symmetry

  $$F \models (R_1 \cap R_2) \twoheadrightarrow (R_2 - R_1)$$

- this condition implies the one for FDs (in only FDs appear in F).

Fourth Normal Form (4NF)

Definition:

Let R be a relation schema and F a set of FDs and MVDs.
Schema R is in **4NF** if and only if

whenever $(X \twoheadrightarrow Y) \in F^+$ and $XY \subseteq R$, then either

- $(X \twoheadrightarrow Y)$ is trivial ($Y \subseteq X$ or $XY = R$), or
- X is a superkey of R

A database schema $\{R_1, \ldots, R_n\}$ is in 4NF if each relation schema R_i is in 4NF.

$\Rightarrow$ use BCNF-like decomposition procedure to obtain a lossless-join decomposition into 4NF.

Example

The CTB schema can be decomposed to 4NF (using $C \twoheadrightarrow T$) as follows:

Course	Teacher
Math	Smith
Math	Jones
Physics	Black
Advanced Math	Smith

Course	Book
Math	Algebra
Math	Calculus
Physics	Mechanics
Physics	Optics
Advanced Math	Calculus

$\Rightarrow$ no FDs here!

Other Dependencies

- **Join Dependency** on R

  $\Rightarrow \bowtie [R_1, \ldots, R_k]$ holds if $I = \pi_{R_1}(I) \bowtie \ldots \bowtie \pi_{R_k}(I)$

  $\Rightarrow$ generalization of an MVD

  $X \twoheadrightarrow Y$ is the same as $\bowtie [XY, X(R - Y)]$

  $\Rightarrow$ **cannot** be simulated by MVDs

  $\Rightarrow$ no axiomatization exists

  $\Rightarrow$ Project-Join NF (5NF)

  $\bowtie [R_1, \ldots, R_k]$ implies R_i is a key.

- **Inclusion Dependency** on R and S

  $\Rightarrow R[X] \subseteq S[Y]$ holds if $\pi_X(I_R) \subseteq \pi_Y(I_S)$

  $\Rightarrow$ relates **two** relations

  foreign-key relationships