Schema Refinement: Other Dependencies and Higher Normal Forms

Spring 2018

School of Computer Science
University of Waterloo

Databases CS348
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:

<table>
<thead>
<tr>
<th>Course</th>
<th>Teacher</th>
<th>Book</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>Smith</td>
<td>Algebra</td>
</tr>
<tr>
<td>Math</td>
<td>Smith</td>
<td>Calculus</td>
</tr>
<tr>
<td>Math</td>
<td>Jones</td>
<td>Algebra</td>
</tr>
<tr>
<td>Math</td>
<td>Jones</td>
<td>Calculus</td>
</tr>
<tr>
<td>Advanced Math</td>
<td>Smith</td>
<td>Calculus</td>
</tr>
<tr>
<td>Physics</td>
<td>Black</td>
<td>Mechanics</td>
</tr>
<tr>
<td>Physics</td>
<td>Black</td>
<td>Optics</td>
</tr>
</tbody>
</table>

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

1. Multivalued Dependencies

2. Other Dependencies
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)
    - \(C \rightarrow\rightarrow T\) (and \(C \rightarrow\rightarrow B\) as well)
  - holds on CTB.

Given a course, the set of teachers and the set of books are uniquely determined and independent.
### Another Example

<table>
<thead>
<tr>
<th>Course</th>
<th>Teacher</th>
<th>Hour</th>
<th>Room</th>
<th>Student</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS101</td>
<td>Jones</td>
<td>M-9</td>
<td>2222</td>
<td>Smith</td>
<td>A</td>
</tr>
<tr>
<td>CS101</td>
<td>Jones</td>
<td>W-9</td>
<td>3333</td>
<td>Smith</td>
<td>A</td>
</tr>
<tr>
<td>CS101</td>
<td>Jones</td>
<td>F-9</td>
<td>2222</td>
<td>Smith</td>
<td>A</td>
</tr>
<tr>
<td>CS101</td>
<td>Jones</td>
<td>M-9</td>
<td>2222</td>
<td>Black</td>
<td>B</td>
</tr>
<tr>
<td>CS101</td>
<td>Jones</td>
<td>W-9</td>
<td>3333</td>
<td>Black</td>
<td>B</td>
</tr>
<tr>
<td>CS101</td>
<td>Jones</td>
<td>F-9</td>
<td>2222</td>
<td>Black</td>
<td>B</td>
</tr>
</tbody>
</table>

- **FDs:**  
  \[ C \rightarrow T, \quad CS \rightarrow G, \quad HR \rightarrow C, \quad HT \rightarrow R, \text{ and } HS \rightarrow R \]

- **MVDs:**  
  \[ C \rightarrow HR \]
Axioms for MVDs

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

Theorem:

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

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

1. $C \rightarrow \rightarrow HR$
2. $C \rightarrow \rightarrow T$ (from $C \rightarrow T$)
3. $C \rightarrow \rightarrow CTSG$ (complementation of (1))
4. $C \rightarrow \rightarrow CT$ (augmentation of (2) by $C$)
5. $CT \rightarrow \rightarrow CTSG$ (augmentation of (3) by $T$)
6. $C \rightarrow \rightarrow SG$ (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 $F \models X \rightarrow Z$ if and only if $Z - X$ is a union of some of the $Y_i$s.

- 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
  \[ R_1 \cap R_2 \rightarrow \rightarrow (R_1 - R_2) \]
  or, by symmetry
  \[ F \models (R_1 \cap R_2) \rightarrow (R_2 - R_1) \]
- 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 $(X \rightarrow \rightarrow Y) \in F^+$ and $XY \subseteq R$, then either
- $(X \rightarrow \rightarrow Y)$ 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.

⇒ use BCNF-like decomposition procedure to obtain a lossless-join decomposition into 4NF.
The $CTB$ schema can be decomposed to 4NF (using $C \rightarrow T$) as follows:

<table>
<thead>
<tr>
<th>Course</th>
<th>Teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>Smith</td>
</tr>
<tr>
<td>Math</td>
<td>Jones</td>
</tr>
<tr>
<td>Physics</td>
<td>Black</td>
</tr>
<tr>
<td>Advanced Math</td>
<td>Smith</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Course</th>
<th>Book</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>Algebra</td>
</tr>
<tr>
<td>Math</td>
<td>Calculus</td>
</tr>
<tr>
<td>Physics</td>
<td>Mechanics</td>
</tr>
<tr>
<td>Physics</td>
<td>Optics</td>
</tr>
<tr>
<td>Advanced Math</td>
<td>Calculus</td>
</tr>
</tbody>
</table>

$\Rightarrow$ no FDs here!
Outline

1 Multivalued Dependencies

2 Other Dependencies
Other Dependencies

- **Join Dependency** on $R$
  
  $\Rightarrow \Join [R_1, \ldots, R_k]$ holds if $I = \pi_{R_1}(I) \Join \ldots \Join \pi_{R_k}(I)$

  $\Rightarrow$ generalization of an MVD
  
  $X \leftrightarrow Y$ is the same as $\Join [XY, X(R - Y)]$

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

  $\Rightarrow$ no axiomatization exists

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

  $\Join [R_1, \ldots, R_k]$ 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