ICS 321 Fall 2009
Schema Refinement & Normal Forms (ii)

Asst. Prof. Lipyeow Lim
Information & Computer Science Department
University of Hawaii at Manoa
Two More Rules

- **Union**
  - If $X \rightarrow Y$ and $X \rightarrow Z$, then $X \rightarrow YZ$
  - Eg. $FLD \rightarrow A$ and $FLD \rightarrow T$, then $FLD \rightarrow AT$

- **Decomposition**
  - If $X \rightarrow YZ$, then $X \rightarrow Y$ and $X \rightarrow Z$
  - Eg. $FLD \rightarrow AT$, then $FLD \rightarrow A$ and $FLD \rightarrow T$

- **Trivial FDs**
  - Right side is a subset of Left side
  - Eg. $F \rightarrow F$, $FLD \rightarrow FD$
Closure

• **Implication**: An FD $f$ is *implied by* a set of FDs $F$ if $f$ holds whenever all FDs in $F$ hold.
  
  – $f=A \rightarrow C$ is implied by $F=\{A \rightarrow B, B \rightarrow C\}$ (using Armstrong’s transitivity)

• **Closure $F^+$**: the set of all FDs implied by $F$
  
  – **Algorithm**:
    
    • start with $F^+=F$
    
    • keep adding new implied FDs to $F^+$ by applying the 5 rules (Armstrong’s Axioms + union + decomposition)
    
    • Stop when $F^+$ does not change anymore.
Example: Closure

- Given FLD is the primary key and C → Z
- Find the closure:
  - Start with { FLD → FLDSCZT, C→Z }
  - Applying reflexivity, { FLD → F, FLD →L, FLD → D, FLD → FL, FLD → LD, FLD →DF, FLDSCZT → FLD, ...}
  - Applying augmentation, { FLDS → FS, FLDS → LS, ...}
  - Applying transitivity ...
  - Applying union ...
  - Applying decomposition ...
  - Repeat until F^+ does not change
Attribute Closure

• Computing the closure of a set of FDs can be expensive. (Size of closure is exponential in # attrs!)

• Typically, we just want to check if a given FD $X \rightarrow Y$ is in the closure of a set of FDs $F$. An efficient check:
  – Compute attribute closure of $X$ (denoted $X^+$) wrt $F$:
    • Set of all attributes $A$ such that $X \rightarrow A$ is in $F^+$
    • There is a linear time algorithm to compute this.
  – Check if $Y$ is in $X^+$

• Does $F = \{ A \rightarrow B, \; B \rightarrow C, \; C \; D \rightarrow E \}$ imply $A \rightarrow E$?
  – i.e., is $A \rightarrow E$ in the closure $F^+$? Equivalently, is $E$ in $A^+$?
Normal Forms

• Helps with the question: do we need to refine the schema?

• If a relation is in a certain normal form (BCNF, 3NF etc.), it is known that certain kinds of problems are avoided/minimized. This can be used to help us decide whether decomposing the relation will help.

• Role of FDs in detecting redundancy:
  – Consider a relation R with 3 attributes, ABC.
    • No FDs hold: There is no redundancy here.
    • Given $A \rightarrow B$: Several tuples could have the same A value, and if so, they’ll all have the same B value!
Boyce-Codd Normal Form (BCNF)

- Let R denote a relation, X a set of attributes from R, A an attribute from R, and F the set of FDs that hold over R.
- R is in **BCNF** if for all $X \rightarrow A$ in $F^+$,
  - $A \in X$ (trivial FD) or
  - $X$ is a superkey
- **Negation**: R is not in BCNF if there exists an $X \rightarrow A$ in $F^+$, such that $A \notin X$ (non-trivial FD) AND $X$ is not a key

The only non-trivial FDs that hold are key constraints.
Examples: BCNF

• Are the following in BCNF?

<table>
<thead>
<tr>
<th>Firstname</th>
<th>Lastname</th>
<th>DOB</th>
<th>Address</th>
<th>Telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>Smith</td>
<td>Sep 9 1979</td>
<td>Honolulu,HI</td>
<td>808-343-0809</td>
</tr>
</tbody>
</table>

F= \{ \text{FLD} \rightarrow \text{FLDAT} \}

<table>
<thead>
<tr>
<th>Firstname</th>
<th>Lastname</th>
<th>DOB</th>
<th>Street</th>
<th>CityState</th>
<th>Zipcode</th>
<th>Telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>Smith</td>
<td>Sep 9 1979</td>
<td>1680 East West Rd.</td>
<td>Honolulu,HI</td>
<td>96822</td>
<td>808-343-0809</td>
</tr>
</tbody>
</table>

F= \{ \text{FLD} \rightarrow \text{FLDSCZT}, \text{C} \rightarrow \text{Z} \}
Third Normal Form (3NF)

• Let R denote a relation, X a set of attributes from R, A an attribute from R, and F the set of FDs that hold over R.
• R is in 3NF if for all $X \rightarrow A$ in $F^+$,
  - $A \in X$ (trivial FD) or
  - $X$ is a superkey or
  - $A$ is part of some key
• **Negation**: R is not in 3NF if there exists an $X \rightarrow A$ in $F^+$, such that $A \not\in X$ (non-trivial FD) AND $X$ is not a key AND $A$ is not part of some key
• If R is in BCNF, obviously in 3NF.
• If R is in 3NF, some redundancy is possible. It is a compromise, used when BCNF not achievable (e.g., no “good” decomp, or performance considerations).
Example: 3NF

• Which of the following is in 3NF and which in BCNF?

<table>
<thead>
<tr>
<th>Firstname</th>
<th>Lastname</th>
<th>DOB</th>
<th>Address</th>
<th>Telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>Smith</td>
<td>Sep 9 1979</td>
<td>Honolulu,HI</td>
<td>808-343-0809</td>
</tr>
</tbody>
</table>

F = \{ FLD \rightarrow FLDAT \}

<table>
<thead>
<tr>
<th>Firstname</th>
<th>Lastname</th>
<th>DOB</th>
<th>Street</th>
<th>CityState</th>
<th>Zipcode</th>
<th>Telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>Smith</td>
<td>Sep 9 1979</td>
<td>1680 East West Rd.</td>
<td>Honolulu,HI</td>
<td>96822</td>
<td>808-343-0809</td>
</tr>
</tbody>
</table>

F = \{ FLD \rightarrow FLDSCZT, C \rightarrow Z \}

<table>
<thead>
<tr>
<th>Student</th>
<th>Course</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith</td>
<td>OS</td>
<td>Mark</td>
</tr>
</tbody>
</table>

F = \{ SC \rightarrow I, I \rightarrow C \}