DATABASE DESIGN I - 1DL300

Autumn 2012

An Introductory Course on Database Systems

http://www.it.uu.se/edu/course/homepage/dbastekn/ht12/

Uppsala Database Laboratory
Department of Information Technology, Uppsala University,
Uppsala, Sweden
Normalization Example

Elmasri/Navathe ch 14
Padron-McCarthy/Risch ch 11

Silvia Stefanova

Department of Information Technology
Uppsala University, Uppsala, Sweden
Outline

1. Normalization – Summary
   • Good database design ???
   • Redundancy, Update anomalies, NULL values, spurious tuples
   • Functional dependency (FD), Full functional dependency (FFD)
     • 1 NF
     • 2 NF
     • 3 NF
     • BCNF

2. Publications database example

3. Street database – BCNF normalize

4. More exercises on normalization
Normalization – Summary

• “Good” database design ???
• Redundancy, Update anomalies, NULL values, spurious tuples
• Functional dependency (FD), Full functional dependency (FFD)
  • 1 NF
  • 2 NF
  • 3 NF
  • BCNF
First Normal Form, 1 NF

1NF: Relations should not have multivalued attributes or nested relations.

<table>
<thead>
<tr>
<th>Ssn</th>
<th>Ename</th>
<th>Pnumber</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234</td>
<td>Smith</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>4534</td>
<td>Wong</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>26</td>
</tr>
</tbody>
</table>

Normalization:
Form new relations for each multivalued attribute or nested relation.
FD and FFD

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>A</th>
<th>P</th>
<th>XY</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Y1</td>
<td>A1</td>
<td>P1</td>
<td>X1Y1</td>
</tr>
<tr>
<td>X1</td>
<td>Y2</td>
<td>A1</td>
<td>P2</td>
<td>X1Y2</td>
</tr>
<tr>
<td>X2</td>
<td>Y3</td>
<td>A2</td>
<td>P3</td>
<td>X2Y3</td>
</tr>
</tbody>
</table>

FDs: FD1, FD2

FFDs: FD3
Second Normal Form - 2NF

**2NF:** A relation is in 2NF if:

– It is in 1NF
– Every non-key attribute in the relation is FFD of each candidate key.

**Normalization:**
Decompose the relation, set up a new relation for each partial key with its dependent attribute(s).
### Third Normal Form - 3NF

**3NF:** A relation is in 2NF if:
- It is in 2NF
- No non-key attribute in a relation is allowed to be FFD on other non-key attribute.

#### Subdivision

<table>
<thead>
<tr>
<th>subdivision</th>
<th>department</th>
<th>num_emp</th>
<th>Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Systems</td>
<td>Informational Technology</td>
<td>400</td>
<td>Project_IT_CS</td>
</tr>
<tr>
<td>Systems and Control</td>
<td>Informational Technology</td>
<td>400</td>
<td>Project_IT_SysC</td>
</tr>
<tr>
<td>Electrical Engineering</td>
<td>Engineering</td>
<td>1000</td>
<td>Project_E_EE</td>
</tr>
</tbody>
</table>

#### Normalization:
Decompose the relation, set up a new relation including the non-key attribute(s) that is FD on other non-key attribute(s).

- transitive FD on the primary key
Boyce-Codd Normal Form - BCNF

**BCNF:** A relation is in BCNF if:
- It is in 1NF
- Every determinant is a candidate key.

### Teach

<table>
<thead>
<tr>
<th>department</th>
<th>course</th>
<th>teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computing Science</td>
<td>Database 1</td>
<td>Sara S</td>
</tr>
<tr>
<td>Computer Systems</td>
<td>Database 1</td>
<td>Sara S</td>
</tr>
<tr>
<td>Engineering</td>
<td>Signals and Systems</td>
<td>Peter E</td>
</tr>
<tr>
<td>Computing Science</td>
<td>Database 2</td>
<td>Sven P</td>
</tr>
</tbody>
</table>

### Normalization:
Decompose the relation so that after joining the new relations *spurious tuples* will not be generated (*lossless join decomposition*)
Outline

1. Normalization – Summary
   • Good database design ?
   • Redundancy, Update anomalies, NULL values, spurious tuples
   • Functional dependency (FD), Full functional dependency (FFD)
     • 1 NF
     • 2 NF
     • 3 NF
     • BCNF

2. Publications database example

3. Street database – BCNF normalize

4. More exercises on normalization
The Database *Publications*

Is *Publications* in 1NF, 2NF or 3NF? Why or why not? How would you normalize it?

- **Area**: Scientific area, e.g. Computer Science, Mathematics, Chemistry, etc
- **Paper**: Title of a paper. A paper can be published in one journal.
- **Author**: Name of the main author of a paper.
- **Journal**: Title of a scientific journal where papers can be published. A paper can be published in only one journal.
- **Country**: Country of origin for the main author.
- **Language**: Publication language for a journal

### Publications

<table>
<thead>
<tr>
<th>Area</th>
<th>Author</th>
<th>Paper</th>
<th>Journal</th>
<th>Country</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Normalization to BCNF

Is *Street* in BCNF? Why or why not? How would you normalize it?

<table>
<thead>
<tr>
<th>street</th>
<th>city</th>
<th>length</th>
<th>zipcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rydsvägen</td>
<td>Linköping</td>
<td>19</td>
<td>58248</td>
</tr>
<tr>
<td>Mårdtorpsgatan</td>
<td>Linköping</td>
<td>0.7</td>
<td>58248</td>
</tr>
<tr>
<td>Storgatan</td>
<td>Linköping</td>
<td>1.5</td>
<td>58223</td>
</tr>
<tr>
<td>Storgatan</td>
<td>Gnesta</td>
<td>0.014</td>
<td>64631</td>
</tr>
</tbody>
</table>
Summary

- Normalization
- Redundancy
- Functional dependency (FD)
- Full functional dependency (FFD)
- 1 NF
- 2 NF
- 3 NF
- BCNF