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
   • Why ?
   • 1 NF
   • Functional dependency (FD), Full functional dependency (FFD)
   • Super key, candidate key, prime, non-prime and non-key attributes
   • 2 NF
   • 3 NF
   • BCNF

2. Publications database example

3. Street database – BCNF normalize

4. More exercises on normalization
Normalization

• Normalization? Why to normalize?
• 1 NF
• Functional dependency (FD), Full functional dependency (FFD)
• Super key, candidate key, non-key
• 2 NF
• 3 NF
• BCNF
First Normal Form, 1 NF

1NF: Relations should not have multivalued attributes or nested relations; only atomic values are allowed

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

<table>
<thead>
<tr>
<th>product</th>
<th>supplier</th>
<th>price</th>
<th>city</th>
<th>population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cars</td>
<td>Volvo</td>
<td>100000</td>
<td>Torslandan</td>
<td>80000</td>
</tr>
<tr>
<td>Cars</td>
<td>SAAB</td>
<td>150000</td>
<td>Södertälje</td>
<td>50000</td>
</tr>
<tr>
<td>Trucks</td>
<td>SAAB</td>
<td>400000</td>
<td>Södertälje</td>
<td>50000</td>
</tr>
<tr>
<td>Aspirin</td>
<td>Astra</td>
<td>10</td>
<td>Södertälje</td>
<td>50000</td>
</tr>
</tbody>
</table>

FD: FD4 (it will not hold if one removes product)
FFDs: FD1, FD2, FD3
Candidate key, prime, non-key attribute

• **Super key of a relation:** A set of attributes so that there are no two distinct tuples (rows) that have the same values for the attributes in this set
• **Candidate key of a relation:** a minimal super key for the relation
• **Prime attribute:** an attribute that is a member of any candidate key
• **Non-prime (non-key) attribute:** attribute that is not a member of any candidate key
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.

**Employee**

<table>
<thead>
<tr>
<th>department</th>
<th>pnumber</th>
<th>daddress</th>
<th>name</th>
<th>projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informational Technology</td>
<td>1234</td>
<td>Polack</td>
<td>Sara</td>
<td>Project_IT1</td>
</tr>
<tr>
<td>Engineering</td>
<td>4567</td>
<td>Ång</td>
<td>Erik</td>
<td>Project_EE0</td>
</tr>
<tr>
<td>Informational Technology</td>
<td>4567</td>
<td>Polack</td>
<td>Erik</td>
<td>Project_IT2</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>supplier</th>
<th>city</th>
<th>population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volvo</td>
<td>Torslanda</td>
<td>80000</td>
</tr>
<tr>
<td>SAAB</td>
<td>Södertälje</td>
<td>50000</td>
</tr>
<tr>
<td>Astra</td>
<td>Södertälje</td>
<td>50000</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).
Boyce-Codd Normal Form - BCNF

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

**Teach** (Each department is located at several addresses)

<table>
<thead>
<tr>
<th>department</th>
<th>course</th>
<th>teacher</th>
<th>daddress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computing Science</td>
<td>Database 1</td>
<td>Sara S</td>
<td>A1</td>
</tr>
<tr>
<td>Computing Science</td>
<td>Programming 1</td>
<td>Anna L</td>
<td>A3</td>
</tr>
<tr>
<td>Engineering</td>
<td>Signals and Systems</td>
<td>Peter E</td>
<td>E2</td>
</tr>
<tr>
<td>Computer Systems</td>
<td>Database 1</td>
<td>Sven P</td>
<td>A2</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
   • Why???
   • 1 NF
   • Functional dependency (FD), Full functional dependency (FFD)
   • Super key, candidate key, prime, non-prime and non-key attributes
   • 2 NF
   • 3 NF
   • BCNF

2. Publications database example

3. Street database – BCNF normalize

4. More exercises on normalization
Is *Publications* in 1NF, 2NF or 3NF? Why or why not? How would you normalize it so that it is in 3NF?

- **Area**: Scientific area, e.g. Computer Science, Mathematics, Chemistry, etc. A journal has at least one area.
- **Paper**: Title of a paper. A paper from the same author can be published in one journal.
- **Author**: Name of the main author of a paper, which is unique.
- **Journal**: Title of a scientific journal where papers can be published.
- **A_address**: Address 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>A_address</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>