Normalisation to 3NF

Database Systems Lecture 11
Natasha Alechina

In This Lecture
• Normalisation to 3NF
• Data redundancy
• Functional dependencies
• Normal forms
• First, Second, and Third Normal Forms
• For more information
  • Connolly and Begg chapter 13
  • Ullman and Widom ch.3.6.6 (2nd edition), 3.5 (3rd edition)

Redundancy and Normalisation

• Redundant data
  • Can be determined from other data in the database
  • Leads to various problems
    • INSERT anomalies
    • UPDATE anomalies
    • DELETE anomalies

• Normalisation
  • Aims to reduce data redundancy
  • Redundancy is expressed in terms of dependencies
  • Normal forms are defined that do not have certain types of dependency

First Normal Form

• In most definitions of the relational model
  • All data values should be atomic
  • This means that table entries should be single values, not sets or composite objects

A relation is said to be in first normal form (1NF) if all data values are atomic

Normalisation to 1NF

To convert to a 1NF relation, split up any non-atomic values

<table>
<thead>
<tr>
<th>Unnormalised</th>
<th>1NF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module</td>
<td>Dept</td>
</tr>
<tr>
<td>M1</td>
<td>D1</td>
</tr>
<tr>
<td>M2</td>
<td>D1</td>
</tr>
<tr>
<td>M3</td>
<td>D1</td>
</tr>
<tr>
<td>M4</td>
<td>D2</td>
</tr>
<tr>
<td>M5</td>
<td>D2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1NF</th>
<th>Module</th>
<th>Dept</th>
<th>Lecturer</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>D1</td>
<td>L1</td>
<td>T1</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>D1</td>
<td>L1</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>D1</td>
<td>L1</td>
<td>T1</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>D1</td>
<td>L1</td>
<td>T3</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>D1</td>
<td>L2</td>
<td>T4</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>D2</td>
<td>L3</td>
<td>T1</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>D2</td>
<td>L3</td>
<td>T5</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>D2</td>
<td>L4</td>
<td>T6</td>
<td></td>
</tr>
</tbody>
</table>

Problems in 1NF

• INSERT anomalies
  • Can’t add a module with no texts
• UPDATE anomalies
  • To change lecturer for M1, we have to change two rows
• DELETE anomalies
  • If we remove M3, we remove L2 as well
Functional Dependencies

- Redundancy is often caused by a functional dependency
- A functional dependency (FD) is a link between two sets of attributes in a relation
- We can normalise a relation by removing undesirable FDs

A set of attributes, A, functionally determines another set, B, or: there exists a functional dependency between A and B \((A \rightarrow B)\), if whenever two rows of the relation have the same values for all the attributes in A, then they also have the same values for all the attributes in B.

Example

- \(\{\text{ID, modCode}\} \rightarrow \{\text{First, Last, modName}\}\)
- \(\{\text{modCode}\} \rightarrow \{\text{modName}\}\)
- \(\{\text{ID}\} \rightarrow \{\text{First, Last}\}\)

<table>
<thead>
<tr>
<th>ID</th>
<th>First</th>
<th>Last</th>
<th>modCode</th>
<th>modName</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>Joe</td>
<td>Bloggs</td>
<td>G51PRG</td>
<td>Programming</td>
</tr>
<tr>
<td>222</td>
<td>Anne</td>
<td>Smith</td>
<td>G51DBS</td>
<td>Databases</td>
</tr>
</tbody>
</table>

FDs and Normalisation

- We define a set of ‘normal forms’
  - Each normal form has fewer FDs than the last
  - Since FDs represent redundancy, each normal form has less redundancy than the last
- Not all FDs cause a problem
  - We identify various sorts of FD that do
  - Each normal form removes a type of FD that is a problem
  - We will also need a way to remove FDs

Properties of FDs

- In any relation
  - The primary key FDs any set of attributes in that relation
  - \(K \rightarrow X\)
  - \(X\) is the primary key, \(X\) is a set of attributes
  - Same for candidate keys
  - Any set of attributes is FD on itself
  - \(X \rightarrow X\)

Rules for FDs

- Reflexivity: If \(B\) is a subset of \(A\) then \(A \rightarrow B\)
- Augmentation: If \(A \rightarrow B\) then \(A \cup C \rightarrow B \cup C\)
- Transitivity: If \(A \rightarrow B\) and \(B \rightarrow C\) then \(A \rightarrow C\)

FD Example

- The primary key is \(\{\text{Module, Text}\}\) so
  - \(\{\text{Module, Text}\} \rightarrow \{\text{Dept, Lecturer}\}\)
- ‘Trivial’ FDs, eg:
  - \(\{\text{Text, Dept}\} \rightarrow \{\text{Text}\}\)
  - \(\{\text{Module}\} \rightarrow \{\text{Module}\}\)
  - \(\{\text{Dept, Lecturer}\} \rightarrow \{\}\)

FD Example

- Other FDs are
  - \(\{\text{Module}\} \rightarrow \{\text{Lecturer}\}\)
  - \(\{\text{Module}\} \rightarrow \{\text{Dept}\}\)
  - \(\{\text{Lecturer}\} \rightarrow \{\text{Dept}\}\)
  - These are non-trivial and determinants (left hand side of the dependency) are not keys.
Partial FDs and 2NF

- Partial FDs:
  - A FD, $A \rightarrow B$ is a partial FD, if some attribute of $A$ can be removed and the FD still holds.
  - Formally, there is some proper subset of $A$, $C \subseteq A$, such that $C \rightarrow B$.
  - Let us call attributes which are part of some candidate key, key attributes, and the rest non-key attributes.

Second normal form:

- A relation is in second normal form (2NF) if it is in 1NF and no non-key attribute is partially dependent on a candidate key.
- In other words, no $C \rightarrow B$ where $C$ is a strict subset of a candidate key and $B$ is a non-key attribute.

Second Normal Form

- 1NF is not in 2NF
  - We have the FD $(\text{Module, Text}) \rightarrow (\text{Lecturer, Dept})$
  - But also $(\text{Module}) \rightarrow (\text{Lecturer, Dept})$
  - And so Lecturer and Dept are partially dependent on the primary key

Removing FDs

- Suppose we have a relation $R$ with scheme $S$ and the FD $A \rightarrow B$ where $A \cap B = \{\}$.
- Let $C = S - (A \cup B)$.
- In other words:
  - $A$ – attributes on the left hand side of the FD
  - $B$ – attributes on the right hand side of the FD
  - $C$ – all other attributes

- It turns out that we can split $R$ into two parts:
  - $R_1$, with scheme $C \cup A$
  - $R_2$, with scheme $A \cup B$
  - The original relation can be recovered as the natural join of $R_1$ and $R_2$.
  - $R = R_1 \text{ NATURAL JOIN } R_2$

Problems Resolved in 2NF

- Problems in 1NF:
  - INSERT – Can’t add a module with no texts.
  - UPDATE – To change lecturer for M1, we have to change two rows.
  - DELETE – If we remove M3, we remove L2 as well.

- In 2NF the first two are resolved, but not the third one.

Problems Remaining in 2NF

- INSERT anomalies
  - Can’t add lecturers who teach no modules.
- UPDATE anomalies
  - To change the department for L1 we must alter two rows.
- DELETE anomalies
  - If we delete M3 we delete L2 as well.
Transitive FDs and 3NF

- Transitive FDs:
  - A FD, \( A \rightarrow C \), is a transitive FD if there is some set \( B \) such that \( A \rightarrow B \) and \( B \rightarrow C \) are non-trivial FDs.
  - \( A \rightarrow B \) non-trivial means \( B \) is not a subset of \( A \).
  - We have \( A \rightarrow B \rightarrow C \).

- Third normal form:
  - A relation is in third normal form (3NF) if it is in 2NF and no non-key attribute is transitively dependent on a candidate key.

Third Normal Form

- 2NF
  - We have the FDs:
    - \( \{\text{Module}\} \rightarrow \{\text{Lecturer}\} \)
    - \( \{\text{Lecturer}\} \rightarrow \{\text{Dept}\} \)

  - So there is a transitive FD from the primary key \( \{\text{Module}\} \) to \( \{\text{Dept}\} \).

2NF to 3NF – Example

<table>
<thead>
<tr>
<th>Module</th>
<th>Dept</th>
<th>Lecturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>D1</td>
<td>L1</td>
</tr>
<tr>
<td>M2</td>
<td>D1</td>
<td>L1</td>
</tr>
<tr>
<td>M3</td>
<td>D1</td>
<td>L2</td>
</tr>
<tr>
<td>M4</td>
<td>D2</td>
<td>L3</td>
</tr>
<tr>
<td>M5</td>
<td>D2</td>
<td>L4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lecturer</th>
<th>Dept</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>D1</td>
</tr>
<tr>
<td>L2</td>
<td>D2</td>
</tr>
<tr>
<td>L3</td>
<td>D2</td>
</tr>
<tr>
<td>L4</td>
<td>D2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Module</th>
<th>Lecturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>L1</td>
</tr>
<tr>
<td>M2</td>
<td>L1</td>
</tr>
<tr>
<td>M3</td>
<td>L2</td>
</tr>
<tr>
<td>M4</td>
<td>L3</td>
</tr>
<tr>
<td>M5</td>
<td>L4</td>
</tr>
</tbody>
</table>

Normalisation and Design

- Normalisation is related to DB design:
  - A database should normally be in 3NF at least.
  - If your design leads to a non-3NF DB, then you might want to revise it.

- When you find you have a non-3NF DB:
  - Identify the FDs that are causing a problem.
  - Think if they will lead to any insert, update, or delete anomalies.
  - Try to remove them.

Problems Resolved in 3NF

- Problems in 2NF:
  - INSERT – Can't add lecturers who teach no modules.
  - UPDATE – To change the department for L1 we must alter two rows.
  - DELETE – If we delete M3 we delete L2 as well.

- In 3NF all of these are resolved (for this relation – but 3NF can still have anomalies!).

Next Lecture

- More normalisation:
  - Lossless decomposition; why our reduction to 2NF and 3NF is lossless.
  - Boyce-Codd normal form (BCNF).
  - Higher normal forms.
  - Denormalisation.

- For more information:
  - Connolly and Begg chapter 14.
  - Ullman and Widom chapter 3.6.