Normalisation 2

Chapter 4.2
V3.0

Copyright © Napier University
Dr Gordon Russell
Normalisation 2

- Overview
  - normalise a relation to Boyce Codd Normal Form (BCNF)
  - An example
Boyce-Codd Normal Form (BCNF)

• When a relation has more than one candidate key, anomalies may result even though the relation is in 3NF.
• 3NF does not deal satisfactorily with the case of a relation with overlapping candidate keys
  – i.e. composite candidate keys with at least one attribute in common.
• BCNF is based on the concept of a determinant.
  – A determinant is any attribute (simple or composite) on which some other attribute is fully functionally dependent.
• A relation is in BCNF is, and only if, every determinant is a candidate key.
The theory

• Consider the following relation and determinants.
  \( R(a,b,c,d) \)
  \[ a,c \rightarrow b,d \]
  \[ a,d \rightarrow b \]
  
• To be in BCNF, all valid determinants must be a candidate key. In the relation \( R \), \( a,c \rightarrow b,d \) is the determinate used, so the first determinate is fine.

• \( a,d \rightarrow b \) suggests that \( a,d \) can be the primary key, which would determine \( b \). However this would not determine \( c \). This is not a candidate key, and thus \( R \) is not in BCNF.
Example 1

<table>
<thead>
<tr>
<th>Patient No</th>
<th>Patient Name</th>
<th>Appointment Id</th>
<th>Time</th>
<th>Doctor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>John</td>
<td>0</td>
<td>09:00</td>
<td>Zorro</td>
</tr>
<tr>
<td>2</td>
<td>Kerr</td>
<td>0</td>
<td>09:00</td>
<td>Killer</td>
</tr>
<tr>
<td>3</td>
<td>Adam</td>
<td>1</td>
<td>10:00</td>
<td>Zorro</td>
</tr>
<tr>
<td>4</td>
<td>Robert</td>
<td>0</td>
<td>13:00</td>
<td>Killer</td>
</tr>
<tr>
<td>5</td>
<td>Zane</td>
<td>1</td>
<td>14:00</td>
<td>Zorro</td>
</tr>
</tbody>
</table>
Two possible keys

• DB(Patno, PatName, appNo, time, doctor)
• Determinants:
  – Patno -> PatName
  – Patno, appNo -> Time, doctor
  – Time -> appNo

• Two options for 1NF primary key selection:
  – DB(Patno, PatName, appNo, time, doctor) (example 1a)
  – DB(Patno, PatName, appNo, time, doctor) (example 1b)
Example 1a

- **DB(Patno, PatName, appNo, time, doctor)**
- No repeating groups, so in 1NF
- **2NF** – eliminate partial key dependencies:
  - **DB(Patno, appNo, time, doctor)**
  - **R1(Patno, PatName)**
- **3NF** – no transient dependences so in 3NF
- Now try BCNF.
BCNF Every determinant is a candidate key

DB(Patno, appNo, time, doctor)
R1(Patno, PatName)

• Is determinant a candidate key?
  – Patno -> PatName
    Patno is present in DB, but not PatName, so irrelevant.
Continued...

\[ DB(\text{Patno,appNo, time, doctor}) \]
\[ R1(\text{Patno,PatName}) \]

- \( \text{Patno,appNo} \rightarrow \text{Time, doctor} \)
  All LHS and RHS present so relevant. Is this a candidate key? Patno,appNo IS the key, so this is a candidate key.

- \( \text{Time} \rightarrow \text{appNo} \)
  Time is present, and so is appNo, so relevant. Is this a candidate key? If it was then we could rewrite DB as:

\[ DB(\text{Patno,appNo, time, doctor}) \]
This will not work, so not BCNF.
Rewrite to BCNF

- DB(Patno, appNo, time, doctor)
  R1(Patno, PatName)
- BCNF: rewrite to
  DB(Patno, time, doctor)
  R1(Patno, PatName)
  R2(time, appNo)
- time is enough to work out the appointment number of a patient. Now BCNF is satisfied, and the final relations shown are in BCNF
Example 1b

- DB(Patno, PatName, appNo, time, doctor)
- No repeating groups, so in 1NF
- 2NF – eliminate partial key dependencies:
  - DB(Patno, time, doctor)
  - R1(Patno, PatName)
  - R2(time, appNo)
- 3NF – no transient dependences so in 3NF
- Now try BCNF.
BCNF Every determinant is a candidate key

DB(Patno, time, doctor)
R1(Patno, PatName)
R2(time, appNo)

• Is determinant a candidate key?
  – Patno -> PatName
    Patno is present in DB, but not PatName, irrelevant.
  – Patno, appNo -> Time, doctor
    Not all LHS present so not relevant
  – Time -> appNo
    Time is present, but not appNo, so not relevant.
  – Relations are in BCNF.
Summary - Example 1

This example has demonstrated three things:

- BCNF is stronger than 3NF, relations that are in 3NF are not necessarily in BCNF
- BCNF is needed in certain situations to obtain full understanding of the data model
- there are several routes to take to arrive at the same set of relations in BCNF.
  - Unfortunately there are no rules as to which route will be the easiest one to take.
Example 2

Grade_report(StudNo, StudName, (Major, Adviser, (CourseNo, Ctitle, InstrucName, InstructLocn, Grade)))

- Functional dependencies
  - StudNo -> StudName
  - CourseNo -> Ctitle, InstrucName
  - InstrucName -> InstructLocn
  - StudNo, CourseNo, Major -> Grade
  - StudNo, Major -> Advisor
  - Advisor -> Major
Example 2 cont...

- Unnormalised
  Grade_report(StudNo, StudName, (Major, Advisor,
                      (CourseNo, Ctitle, InstrucName, InstructLocn, Grade)))

- 1NF Remove repeating groups
  - Student(StudNo, StudName)
  - StudMajor(StudNo, Major, Advisor)
  - StudCourse(StudNo, Major, CourseNo,
                Ctitle, InstrucName, InstructLocn, Grade)
Example 2 cont...

- 1NF
  Student(StudNo, StudName)
  StudMajor(StudNo, Major, Advisor)
  StudCourse(StudNo, Major, CourseNo, Ctitle, InstrucName, InstructLocn, Grade)

- 2NF Remove partial key dependencies
  Student(StudNo, StudName)
  StudMajor(StudNo, Major, Advisor)
  StudCourse(StudNo, Major, CourseNo, Grade)
  Course(CourseNo, Ctitle, InstrucName, InstructLocn)
Example 2 cont...

- 2NF
  Student(StudNo, StudName)
  StudMajor(StudNo, Major, Advisor)
  StudCourse(StudNo, Major, CourseNo, Grade)
  Course(CourseNo, Ctitle, InstrucName, InstructLocn)

- 3NF Remove transitive dependencies
  Student(StudNo, StudName)
  StudMajor(StudNo, Major, Advisor)
  StudCourse(StudNo, Major, CourseNo, Grade)
  Course(CourseNo, Ctitle, InstrucName)
  Instructor(InstructName, InstructLocn)
Example 2 cont...

• BCNF Every determinant is a candidate key
  – Student: only determinant is StudNo
  – StudCourse: only determinant is StudNo,Major
  – Course: only determinant is CourseNo
  – Instructor: only determinant is InstrucName
  – StudMajor: the determinants are
    • StudNo,Major, or
    • Advisor
    Only StudNo,Major is a candidate key.
Example 2: BCNF

- BCNF

Student(StudNo, StudName)
StudCourse(StudNo, Major, CourseNo, Grade)
Course(CourseNo, Ctitle, InstrucName)
Instructor(InstructName, InstructLocn)
StudMajor(StudNo, Advisor)
Adviser(Adviser, Major)
Problems BCNF overcomes

<table>
<thead>
<tr>
<th>STUDENT</th>
<th>MAJOR</th>
<th>ADVISOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>PHYSICS</td>
<td>EINSTEIN</td>
</tr>
<tr>
<td>123</td>
<td>MUSIC</td>
<td>MOZART</td>
</tr>
<tr>
<td>456</td>
<td>BIOLOGY</td>
<td>DARWIN</td>
</tr>
<tr>
<td>789</td>
<td>PHYSICS</td>
<td>BOHR</td>
</tr>
<tr>
<td>999</td>
<td>PHYSICS</td>
<td>EINSTEIN</td>
</tr>
</tbody>
</table>

- If the record for student 456 is deleted we lose not only information on student 456 but also the fact that DARWIN advises in BIOLOGY.
- We cannot record the fact that WATSON can advise on COMPUTING until we have a student majoring in COMPUTING to whom we can assign WATSON as an advisor.
Split into two tables

In BCNF we have two tables

<table>
<thead>
<tr>
<th>STUDENT</th>
<th>ADVISOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>EINSTEIN</td>
</tr>
<tr>
<td>123</td>
<td>MOZART</td>
</tr>
<tr>
<td>456</td>
<td>DARWIN</td>
</tr>
<tr>
<td>789</td>
<td>BOHR</td>
</tr>
<tr>
<td>999</td>
<td>EINSTEIN</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ADVISOR</th>
<th>MAJOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>EINSTEIN</td>
<td>PHYSICS</td>
</tr>
<tr>
<td>MOZART</td>
<td>MUSIC</td>
</tr>
<tr>
<td>DARWIN</td>
<td>BIOLOGY</td>
</tr>
<tr>
<td>BOHR</td>
<td>PHYSICS</td>
</tr>
</tbody>
</table>
Returning to the ER Model

• Now that we have reached the end of the normalisation process, you must go back and compare the resulting relations with the original ER model.

  - You may need to alter it to take account of the changes that have occurred during the normalisation process. Your ER diagram should always be a perfect reflection of the model you are going to implement in the database, so keep it up to date!
  - The changes required depends on how good the ER model was at first!
Video Library Example

- A video library allows customers to borrow videos.
- Assume that there is only 1 of each video.
- We are told that:
  
  ```
  video(title, director, serial)
  customer(name, addr, memberno)
  hire(memberno, serial, date)
  
  title->director, serial
  serial->title
  serial->director
  name, addr -> memberno
  memberno -> name, addr
  serial, date -> memberno
  ```
What NF is this?

- No repeating groups therefore at least 1NF
- 2NF – A Composite key exists:
  hire(memberno,serial,date)
  – Can memberno be found with just serial or date?
  – NO, therefore the relations are already in 2NF.
- 3NF?
Test for 3NF

- video(title, director, serial)
  - title -> director, serial
  - serial -> director

- Director can be derived using serial, and serial and director are both non keys, so therefore this is a transitive or non-key dependency.

- Rewrite video…
Rewrite for 3NF

- video(title, director, serial)
  - title -> director, serial
  - serial -> director

- Becomes:
- video(title, serial)
- serial(serial, director)
Check BCNF

- Is every determinant a candidate key?
- `video(title,serial)` - Determinants are:
  - `title->director,serial` Candidate key
  - `serial->title` Candidate key
  - `video` in BCNF
- `serial(serial,director)` Determinants are:
  - `serial->director` Candidate key
  - `serial` in BCNF
• customer(name, addr, memberno) Determinants are:
  – name, addr -> memberno  Candidate key
  – memberno -> name, addr  Candidate key
  – customer in BCNF
• hire(memberno, serial, date) Determinants are:
  – serial, date -> memberno  Candidate key
  – hire in BCNF
• Therefore the relations are also now in BCNF.