Normalisation 2

- Overview
  - Define Boyce Codd Normal Form (BCNF)
  - normalise a relation to BCNF
  - more examples 1NF => 2NF => 3NF => BCNF

- No partial key dependencies
- No transitive (non-key) dependencies
- No repeating groups
Candidate Key

- An attribute (or set of attributes) in a functional dependency (determinant) which may be chosen as the Primary key of a relation.
- To be a Primary key it must uniquely identify every row and must include all the attributes.
- Some relations may have more than one candidate key e.g. R (empno, name, dob)

- empno => name, DOB candidate key empno
- name, DOB => empno candidate key name, DOB
- name => DOB determinant is NOT a candidate

For any given relation, we choose one of the candidate keys to be the primary key – others, if any, are called alternate keys.

DETERMINANTS
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 the determinant. (LHS of a functional dependency)
  – A determinant is any attribute (simple or composite) on which some other attribute is fully functionally dependent.
• A relation is in BCNF if, 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 candidate keys. In the relation \( R \), for \( a, c \rightarrow b, d \) all of the non-key attributes can be determined, so the first determinant \( a, c \) 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 No.</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

- We have 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 (\textit{Patno}, \textit{PatName}, \textit{appNo}, \textit{time}, \textit{doctor})
  - No repeating groups, so in 1NF
- 2NF – eliminate partial key dependencies (PKDs):
  - DB (\textit{Patno}, \textit{appNo}, \textit{time}, \textit{doctor})
  - R1 (\textit{Patno}, \textit{PatName})  \textbf{[from Patno -> PatName]}  

- 3NF – no transitive dependencies so in 3NF
- Now try BCNF.
BCNF Every determinant is a candidate key

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

• Is every determinant a candidate key?
  – Patno -> PatName

  1. DB relation: Patno (LHS) is present but PatName (RHS) is not, so not relevant to DB

  2. R1 Relation: Patno is the key to R1 so Patno is a candidate key.
     As it is the only determinant, every determinant IS a candidate key, and R1 is in BCNF
Continued…

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

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

- \text{Time} -> \text{appNo}
  Both Time and appNo, are present in DB so \textit{relevant}. Is this a candidate key? \textit{NO}. Time does not determine all attributes, so \textit{DB is not in BCNF}. => Fix =>
Fix: Decompose to BCNF

- $DB(\text{Patno}, \text{appNo}, \text{time}, \text{doctor})$
- $R1(\text{Patno}, \text{PatName})$

BCNF: rewrite to
- $DB(\text{Patno, time, doctor})$
- $R1(\text{Patno, PatName})$
- $R2(\text{time, appNo})$ from $\text{Time} \rightarrow \text{appNo}$

- time is enough to work out the appointment number of a patient. Now BCNF is satisfied, and the final relations shown are in BCNF

- Time $\rightarrow$ appNo so Time has substituted for appNo in the composite key of DB.

(Another example)
MatricNo would substitute for StudName in Napier’s database because MatricNo $\Rightarrow$ StudName
DB (Patno, appNo, time, doctor) =>

R2 (time, appNo)
DB (Patno, appNo, doctor)
DB (Patno, time, doctor)  substitute from Time -> appNo

<table>
<thead>
<tr>
<th>PatNo</th>
<th>AppNo</th>
<th>Time</th>
<th>Doctor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>09:00</td>
<td>Zorro</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>09:00</td>
<td>Killer</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>10:00</td>
<td>Zorro</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>13:00</td>
<td>Killer</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>14:00</td>
<td>Zorro</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TIME</th>
<th>APPNO</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00</td>
<td>0</td>
</tr>
<tr>
<td>10:00</td>
<td>1</td>
</tr>
<tr>
<td>11:00</td>
<td>2</td>
</tr>
<tr>
<td>13:00</td>
<td>0</td>
</tr>
<tr>
<td>14:00</td>
<td>1</td>
</tr>
<tr>
<td>15:00</td>
<td>2</td>
</tr>
<tr>
<td>16:00</td>
<td>3</td>
</tr>
</tbody>
</table>
Example 1b a different approach – same result

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

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

• Is each determinant a candidate key (DB relation)?
  1. Patno -> PatName
     Patno is present in DB, but not PatName. Not relevant.
  2. Patno, appNo -> Time, doctor
     Not all of LHS present (appNo) so Not relevant
  3. Time -> appNo
     Time is present, but not appNo, so Not relevant.

• Relations are already in BCNF. Every determinant is a candidate key
Summary - Example 1

This example has demonstrated three things:

• **BCNF is stronger than 3NF**, relations that are in 3NF are not necessarily in BCNF

• 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, Major, CourseNo -> 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) [studno => studname]
  – StudMajor (StudNo, Major, Advisor) [studno, major => advisor]
  – StudCourse (StudNo, Major, CourseNo, Ctitle, InstrucName, InstructLocn, Grade)
    [studno, major, courseno => grade]
Example 2 cont...

• **1NF**  
  Student \((\text{StudNo}, \text{StudName})\)  
  StudMajor \((\text{StudNo}, \text{Major}, \text{Advisor})\)  
  StudCourse \((\text{StudNo}, \text{Major}, \text{CourseNo}, \text{Ctitle}, \text{InstrucName}, \text{InstructLocn}, \text{Grade})\)  
  \[\text{CourseNo} \rightarrow \text{Ctitle, InstrucName}\]

• **2NF Remove partial key dependencies**  
  Student \((\text{StudNo, StudName})\)  
  StudMajor \((\text{StudNo, Major, Advisor})\)  
  StudCourse \((\text{StudNo, Major, CourseNo, Grade})\)  
  Course \((\text{CourseNo, Ctitle, InstrucName, InstructLocn})\)  
  InstrucName \rightarrow \text{InstrucLocn}\)
Example 2 cont...

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

  \[ \text{InstructName} \rightarrow \text{InstructLocn} \]

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

- **BCNF Every determinant must be a candidate key**
  - Student: only determinant is StudNo
  - StudCourse: only determinant is StudNo, Major, CourseNo
  - Course: only determinant is CourseNo
  - Instructor: only determinant is InstrucName
  - StudMajor: the determinants are
    - StudNo, Major [StudNo, Major -> advisor]
    - Advisor [Advisor -> Major]

Only StudNo, Major is a candidate key. I.e. StudMajor (StudNo, Major, Advisor)
Example 2: BCNF

- **BCNF remove non-candidate determinants**
- **Student** \((\text{StudNo}, \text{StudName})\)
  **StudCourse** \((\text{StudNo}, \text{Major}, \text{CourseNo}, \text{Grade})\)
  **Course** \((\text{CourseNo}, \text{Ctitle}, \text{InstrucName})\)
  **Instructor** \((\text{InstructName}, \text{InstructLocn})\)
  **StudMajor** \((\text{StudNo}, \text{Major}, \text{Advisor})\)

  \[\text{Advisor} \rightarrow \text{Major}\]

  Because Advisor determines Major

- **StudMajor** \((\text{StudNo}, \text{Advisor})\)
  **Advisor** \((\text{Advisor}, \text{Major})\)
Problems that BCNF overcomes

StudMajor (StudNo, Major, Advisor)

<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.
In BCNF we get two tables: StudMajor (StudNo, Advisor) & Advisor (Advisor, Major)

<table>
<thead>
<tr>
<th>StudNo</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>
<tr>
<td>WATSON</td>
<td>COMPUTING</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 using just serial or just date?
  - NO, therefore the relations are already in 2NF (i.e., no PKD possible).
- 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 is in BCNF
• Serial (serial, director) Determinants are:
  – serial -> director Candidate key
  – Serial is in BCNF
• Customer (name, addr, memberno) Determinants are:
  – name, addr -> memberno             Candidate key
  – memberno -> name, addr             Candidate key
  – Customer is in BCNF

• Hire (memberno, serial, date) Determinants are:
  – serial, date -> memberno            Candidate key
  – Hire is in BCNF

• Therefore the relations are also now in BCNF.

• END