Normalisation 2

Chapter 4.2

V3.0

INVESTOR IN PEOPLE

NAPIER UNIVERSITY

EDINBURGH

Normalisation 2

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

NAPIER UNIVERSITY
EDINBURGH

INVESTOR IN PEOPLE

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.

NAPIER UNIVERSITY
EDINBURGH

The theory

- Consider the following relation and determinants.

  R(a,b,c,d)

    a,c -> b,d
    a,d -> b

- To be in BCNF, all valid determinants must be a candidate key. In the relation R, a,c->b,d is the determinate used, so the first determinate is fine.

- a,d->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

Patient No	Patient Name	Appointment Id	Time	Doctor
1	John	0	09:00	Zorro
2	Kerr	0	09:00	Killer
3	Adam	1	10:00	Zorro
4	Robert	0	13:00	Killer
5	Zane	1	14:00	Zorro

NAPIER UNIVERSITY
EDINBURGH

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(<u>Patno,</u>PatName,<u>appNo,</u>time,doctor) (example 1a)
  - DB(<u>Patno,</u>PatName,<u>appNo,</u><u>time,</u>doctor) (example 1b)

NAPIER UNIVERSITY
EDINBURGH

Example 1a

- DB(<u>Patno,</u>PatName,<u>appNo</u>,time,doctor)
- No repeating groups, so in 1NF
- 2NF – eliminate partial key dependencies:
  - DB(<u>Patno,appNo</u>,time,doctor)
  - R1(<u>Patno</u>,PatName)
- 3NF – no transient dependences so in 3NF
- Now try BCNF.

NAPIER UNIVERSITY
EDINBURGH

BCNF Every determinant is a candidate key

DB(<u>Patno,appNo</u>,time,doctor)
R1(<u>Patno</u>,PatName)

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

Continued…

DB(<u>Patno,appNo</u>,time,doctor)
R1(<u>Patno</u>,PatName)

- Patno,appNo -> 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.

- Time -> 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(Patno,appNo,<u>time</u>,doctor)
  This will not work, so not BCNF.

NAPIER UNIVERSITY
EDINBURGH

Rewrite to BCNF

- DB(<u>Patno,appNo</u>,time,doctor)
  R1(<u>Patno</u>,PatName)
- BCNF: rewrite to
  DB(<u>Patno,time</u>,doctor)
  R1(<u>Patno</u>,PatName)
  R2(<u>time</u>,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

NAPIER UNIVERSITY
EDINBURGH

INVESTOR IN PEOPLE

Example 1b

- DB(<u>Patno,</u>PatName,appNo,<u>time</u>,doctor)
- No repeating groups, so in 1NF
- 2NF – eliminate partial key dependencies:
  - DB(<u>Patno,time,</u>doctor)
  - R1(<u>Patno,</u>PatName)
  - R2(<u>time,</u>appNo)
- 3NF – no transient dependences so in 3NF
- Now try BCNF.

NAPIER UNIVERSITY
EDINBURGH

BCNF Every determinant is a candidate key

DB(<u>Patno</u>,<u>time</u>,doctor)

R1(<u>Patno</u>,PatName)

R2(<u>time</u>,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.

NAPIER UNIVERSITY

EDINBURGH

Summary - Example 1

This example has demonstrated three things:

- BCNF is stronger than 3NF, relations that are in 3NF are not necessarily inBCNF
- 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.

NAPIER UNIVERSITY

EDINBURGH

Example 2

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

- Functional dependencies
  - StudNo -> StudName
  - CourseNo -> Ctitle,InstrucName
  - InstrucName -> InstrucLocn
  - StudNo,CourseNo,Major -> Grade
  - StudNo,Major -> Advisor
  - Advisor -> Major

NAPIER UNIVERSITY
EDINBURGH

Example 2 cont...

- Unnormalised
  Grade_report(StudNo,StudName,(Major,Advisor,
    (CourseNo,Ctitle,InstrucName,InstructLocn,Grade)))
- 1NF Remove repeating groups
  - Student(<u>StudNo</u>,StudName)
  - StudMajor(<u>StudNo,Major</u>,Advisor)
  - StudCourse(<u>StudNo,Major,CourseNo</u>,
      Ctitle,InstrucName,InstructLocn,Grade)

NAPIER UNIVERSITY
EDINBURGH

Example 2 cont...

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

- 2NF Remove partial key dependencies
  Student(<u>StudNo</u>,StudName)
  StudMajor(<u>StudNo,Major</u>,Advisor)
  StudCourse(<u>StudNo,Major,CourseNo</u>,Grade)
  Course(<u>CourseNo</u>,Ctitle,InstrucName,InstructLocn)

NAPIER UNIVERSITY
EDINBURGH

Example 2 cont...

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

- 3NF Remove transitive dependencies
  Student(<u>StudNo</u>,StudName)
  StudMajor(<u>StudNo,Major</u>,Advisor)
  StudCourse(<u>StudNo,Major,CourseNo</u>,Grade)
  Course(<u>CourseNo</u>,Ctitle,InstrucName)
  Instructor(<u>InstructName</u>,InstructLocn)

NAPIER UNIVERSITY
EDINBURGH

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(<u>StudNo</u>,StudName)
  StudCourse(<u>StudNo,Major,CourseNo</u>,Grade)
  Course(<u>CourseNo</u>,Ctitle,InstrucName)
  Instructor(<u>InstructName</u>,InstructLocn)
  StudMajor(<u>StudNo</u>,Advisor)
  Adviser(<u>Adviser</u>,Major)

Problems BCNF overcomes

STUDENT	MAJOR	ADVISOR
123	PHYSICS	EINSTEIN
123	MUSIC	MOZART
456	BIOLOGY	DARWIN
789	PHYSICS	BOHR
999	PHYSICS	EINSTEIN

- 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.

NAPIER UNIVERSITY
EDINBURGH

Split into two tables

In BCNF we have two tables

STUDENT	ADVISOR
123	EINSTEIN
123	MOZART
456	DARWIN
789	BOHR
999	EINSTEIN

ADVISOR	MAJOR
EINSTEIN	PHYSICS
MOZART	MUSIC
DARWIN	BIOLOGY
BOHR	PHYSICS

NAPIER UNIVERSITY
EDINBURGH

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 prefect 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!

NAPIER UNIVERSITY
EDINBURGH

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(<u>title</u>,director,serial)
customer(name,addr,<u>memberno</u>)
hire(memberno,<u>serial,date</u>)

     title->director,serial
     serial->title
     serial->director
     name,addr -> memberno
     memberno -> name,addr
     serial,date -> memberno

NAPIER UNIVERSITY
EDINBURGH

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?

NAPIER UNIVERSITY
EDINBURGH

INVESTOR IN PEOPLE

Test for 3NF

- video(<u>title</u>,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…

NAPIER UNIVERSITY
EDINBURGH

Rewrite for 3NF

- video(<u>title</u>,director,serial)
  - title->director,serial
  - serial->director
- Becomes:
- video(<u>title</u>,serial)
- serial(<u>serial</u>,director)

NAPIER UNIVERSITY
EDINBURGH

Check BCNF

- Is every determinant a candidate key?
- video(<u>title</u>,serial) - Determinants are:
  - title->director,serial **Candidate key**
  - serial->title **Candidate key**
  - video in BCNF
- serial(<u>serial</u>,director) Determinants are:
  - serial->director **Candidate key**
  - serial in BCNF

NAPIER UNIVERSITY
EDINBURGH
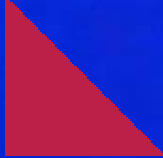
- customer(name,addr,<u>memberno</u>) Determinants are:
  - name,addr -> memberno   **Candidate key**
  - memberno -> name,addr   **Candidate key**
  - customer in BCNF
- hire(memberno,<u>serial,date</u>) Determinants are:
  - serial,date -> memberno   **Candidate key**
  - hire in BCNF
- Therefore the relations are also now in BCNF.

NAPIER UNIVERSITY
EDINBURGH