Edinburgh Napier UNIVERSITY

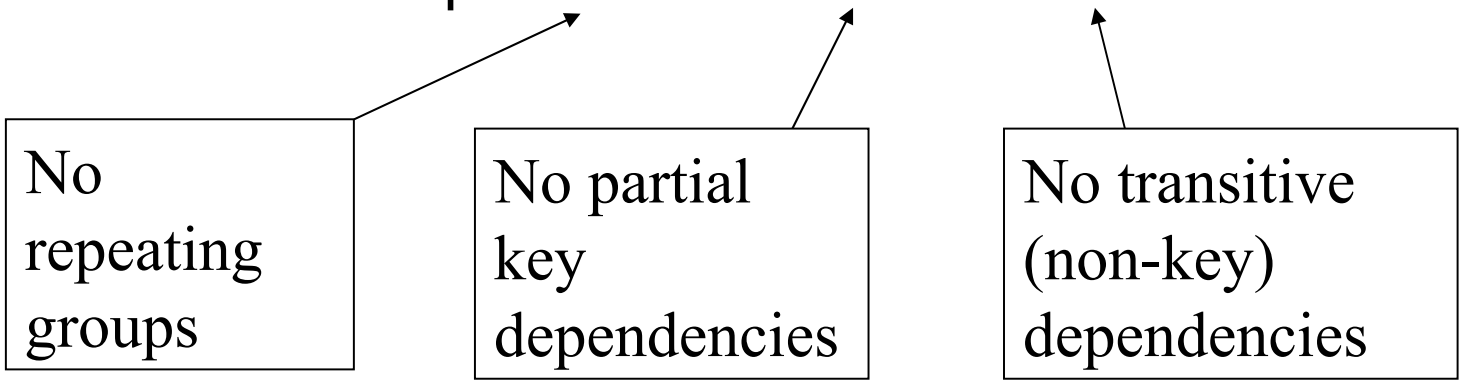
Normalisation 2

Chapter 4.2
V4.0

Edinburgh Napier UNIVERSITY

Normalisation 2

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



Edinburgh Napier UNIVERSITY

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 <u>empno</u>
name, DOB	=> empno	candidate key <u>name, DOB</u>
name	=> DOB	determinant is NOT a candidate -- *only 1 attrib on the RHS*

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

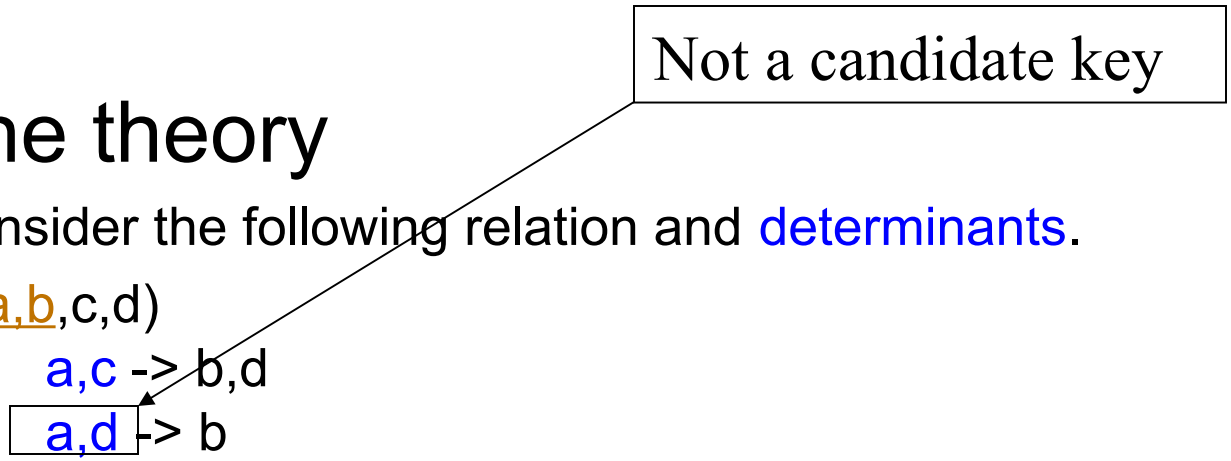
DETERMINANTS

Edinburgh Napier UNIVERSITY

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 -> b,d

  a,d -> b

- To be in BCNF, all valid determinants must be candidate keys. In the relation R, for a,c->b,d all of the non-key attributes can be determined, so the first determinant a,c 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.

Edinburgh Napier UNIVERSITY

Example 1

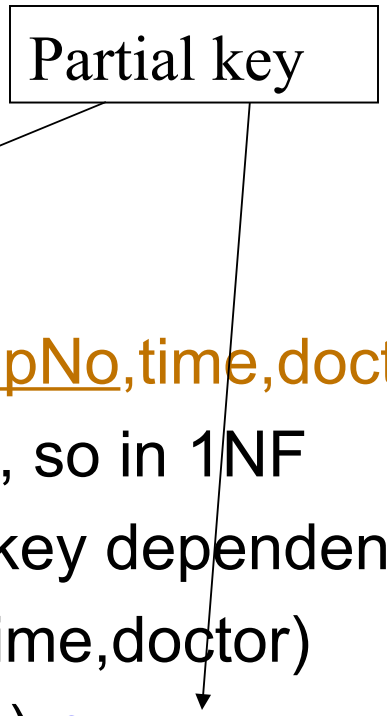
Patient No	Patient Name	Appointment No.	Time	Doctor
1	John	0	09:00	Zorro
2	Kerr	0	09:00	Killer
3	Adam	1	10:00	Zorro
4	Robert	0	13:00	Killer
5	Zane	1	14:00	Zorro

Edinburgh Napier UNIVERSITY

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)



Edinburgh Napier UNIVERSITY

Example 1a

- DB (Patno,PatName,appNo,time,doctor)
  - No repeating groups, so in 1NF
- 2NF – eliminate partial key dependencies (PKDs):
  - DB (Patno, appNo, time,doctor)
  - R1 (Patno, PatName) [from **Patno -> PatName**]

- 3NF – no transitive dependencies so in 3NF
- Now try BCNF.

Edinburgh Napier UNIVERSITY

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(Patno,appNo,time,doctor)

- 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
  Both Time and appNo, are present in DB so relevant. Is this a candidate key? NO. Time does not determine all attributes, so DB is not in BCNF. => Fix =>

Fix: Decompose to BCNF

- DB(Patno,appNo,time,doctor)
  R1(Patno,PatName)
- BCNF: rewrite to
  DB(Patno,time,doctor)
  R1(Patno,PatName)
  R2(time,appNo) **from 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

(another example) MatricNo would substitute for StudName in Napier's database because MatricNo => StudName

- Time -> appNo so Time has substituted for appNo in the composite key of DB.

Edinburgh Napier UNIVERSITY

DB (Patno, appNo, time, doctor) =>

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

PatNo		AppNo	Time	Doctor
1		0	09:00	Zorro
2		0	09:00	Killer
3		1	10:00	Zorro
4		0	13:00	Killer
5		1	14:00	Zorro

TIME	APPNO
09:00	0
10:00	1
11:00	2
13:00	0
14:00	1
15:00	2
16:00	3

Edinburgh Napier UNIVERSITY

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.

Edinburgh Napier UNIVERSITY

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

Edinburgh Napier UNIVERSITY

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.

Edinburgh Napier UNIVERSITY

Example 2

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

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

Edinburgh Napier UNIVERSITY

Example 2 cont...

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

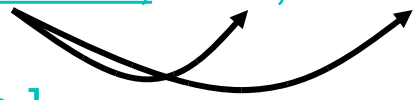
- 1NF Remove repeating groups
  - Student (StudNo, StudName) [studno => studname]
  - StudMajor (StudNo, Major, Advisor) [studno, major => advisor]
  - StudCourse (StudNo, Major, CourseNo,
    Ctitle, InstrucName, InstructLocn, Grade)

    [studno, major, courseno => grade]

Example 2 cont...

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

  [CourseNo => Ctitle, InstrucName]



- 2NF Remove partial key dependencies

  Student (StudNo, StudName)

  StudMajor (StudNo, Major, Advisor)

  StudCourse (StudNo, Major, CourseNo, Grade)

  Course (CourseNo, Ctitle, InstrucName, InstructLocn)

InstrucName -> InstrucLocn

Edinburgh Napier UNIVERSITY

Example 2 cont...

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

  [InstructName -> 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 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)

Advisor cannot be the key for the whole relation

Edinburgh Napier UNIVERSITY

Example 2: BCNF

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

Advisor -> Major

Because Advisor determines Major

- StudMajor (StudNo, Advisor)
  Advisor (Advisor, Major)

Problems that BCNF overcomes

StudMajor (StudNo,Major,Advisor)

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.

Edinburgh Napier UNIVERSITY

Split into two tables

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

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

ADVISOR	MAJOR
EINSTEIN	PHYSICS
MOZART	MUSIC
DARWIN	BIOLOGY
BOHR	PHYSICS

WATSON	COMPUTING

Now we can delete 456 and or add Watson advising Computing

Edinburgh Napier UNIVERSITY

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!

Edinburgh Napier
UNIVERSITY

Edinburgh Napier UNIVERSITY

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

Edinburgh Napier UNIVERSITY

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?

Edinburgh Napier UNIVERSITY

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…

Edinburgh Napier UNIVERSITY

Rewrite for 3NF

- Video (title, director, serial)
  - Title -> director, serial
  - Serial -> director
- Becomes:
- Video (title, serial)
- **Serial** (serial, director)

Edinburgh Napier UNIVERSITY

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