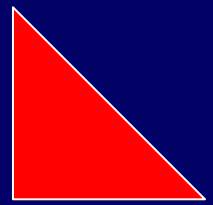
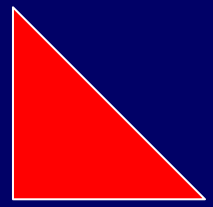


Normalisation 2

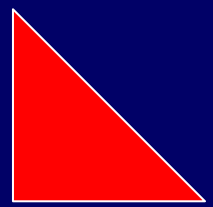
Unit 3.2

Normalisation 2



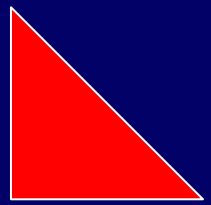
- Overview
  - normalise a relation to Boyce Codd Normal Form (BCNF)
  - normalise a relation to forth normal form (4NF)
  - normalise a relation to fifth normal form (5NF)

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



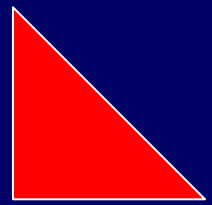
- Consider the following relation and determinants.

  R(<u>a,b</u>,c,d)

  a,c -> b,d

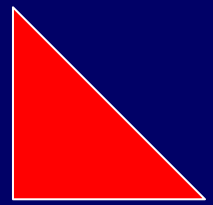
  a,d -> b
- To be in BCNF, all valid determinants must be a candidate key. In the relation R, a,c->b,d is the determinate used, so the first determinate is fine.
- a,d->b suggests that a,d can be the primary key, which would determine b. However this would not determine c. This is not a candidate key, and thus R is not in BCNF.

Example 1



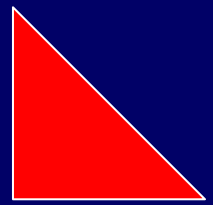
Patient No	Patient Name	Appointment Id	Time	Doctor
1	John	0	09:00	Zorro
2	Kerr	0	09:00	Killer
3	Adam	1	10:00	Zorro
4	Robert	0	13:00	Killer
5	Zane	1	14:00	Zorro

Two possible keys



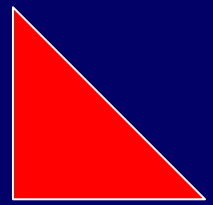
- DB(Patno,PatName,appNo,time,doctor)
- Determinants:
  - Patno -> PatName
  - Patno,appNo -> Time,doctor
  - Time -> appNo
- Two options for 1NF primary key selection:
  - DB(Patno,PatName,appNo,time,doctor) (example 1a)
  - DB(Patno,PatName,appNo,time,doctor) (example 1b)

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.

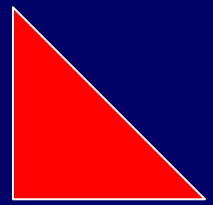
BCNF Every determinant is a candidate key



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

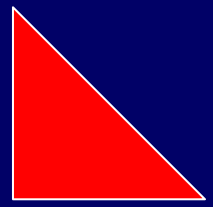
- Is determinant a candidate key?
  - Patno -> PatName
    Patno is present in DB, but not PatName, so not relevant.
  - Patno,appNo -> Time,doctor
    All LHS and RHS present so relevant. Is this a candidate key? Patno,appNo IS the key, so this is a candidate key.
  - Time -> appNo
    Time is present, and so is appNo, so relevant. Is this a candidate key? If it was then we could rewrite DB as:
    DB(Patno,appNo,time,doctor)
    This will not work, so not BCNF.

Rewrite to BCNF



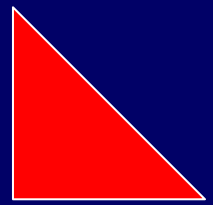
- DB(Patno,appNo,time,doctor)
  R1(Patno,PatName)
- BCNF: rewrite to
  DB(Patno,time,doctor)
  R1(Patno,PatName)
  R2(time,appNo)
- time is enough to work out the appointment number of a patient. Now BCNF is satisfied, and the final relations shown are in BCNF

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



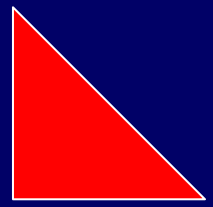
DB(Patno,time,doctor)

R1(Patno,PatName)

R2(time,appNo)

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

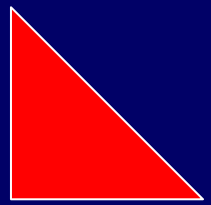
Summary - Example 1



This example has demonstrated three things:

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

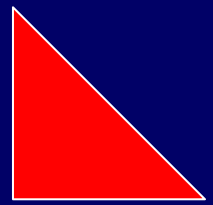
Example 2



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

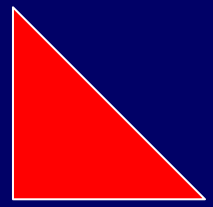
- Functional dependencies
  - StudNo -> StudName
  - CourseNo -> Ctitle,InstrucName
  - InstrucName -> InstrucLocn
  - StudNo,CourseNo,Major -> Grade
  - StudNo,Major -> Advisor
  - Advisor -> Major

Example 2 cont...



- Unnormalised
  Grade_report(StudNo,StudName,(Major,Adviser,
  (Course,Ctitle,InstrucName,InstructLocn,Grade)))
- 1NF Remove repeating groups
  - Student(StudNo,StudName)
  - StudMajor(StudNo,Major,Adviser)
  - StudCourse(StudNo,Major,Course,
    Ctitle,InstrucName,InstructLocn,Grade)

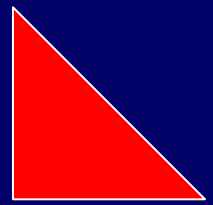
Example 2 cont...



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

- 2NF Remove partial key dependencies
  Student(StudNo,StudName)
  StudMajor(StudNo,Major,Adviser)
  StudCourse(StudNo,Major,Course,Grade)
  Course(Course,Ctitle,InstrucName,InstructLocn)

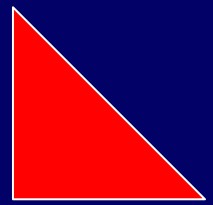
Example 2 cont...



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

- 3NF Remove transitive dependencies
  Student(StudNo,StudName)
  StudMajor(StudNo,Major,Adviser)
  StudCourse(StudNo,Major,Course,Grade)
  Course(Course,Ctitle,InstrucName)
  Instructor(InstructName,InstructLocn)

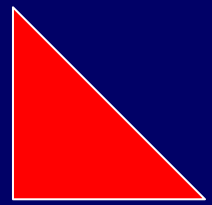
Example 2 cont...



- BCNF Every determinant is a candidate key
  - Student : only determinant is StudNo
  - StudCourse: only determinant is StudNo,Major
  - Course: only determinant is Course
  - Instructor: only determinant is InstrucName
  - StudMajor: the determinants are
    - StudNo,Major, or
    - Adviser

    Only StudNo,Major is a candidate key.

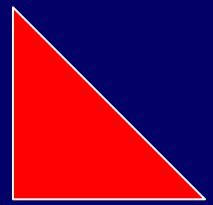
Example 2: BCNF



- BCNF

  Student(StudNo,StudName)
  StudCourse(StudNo,Major,Course,Grade)
  Course(Course,Ctitle,InstrucName)
  Instructor(InstructName,InstructLocn)
  StudMajor(StudNo,Adviser)
  Adviser(Adviser,Major)

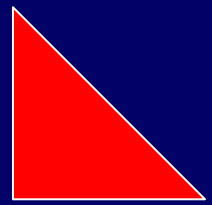
Problems BCNF overcomes



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.

Split into two tables

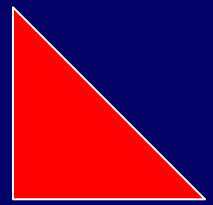


In BCNF we have two tables

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

ADVISOR	MAJOR
EINSTEIN	PHYSICS
MOZART	MUSIC
DARWIN	BIOLOGY
BOHR	PHYSICS

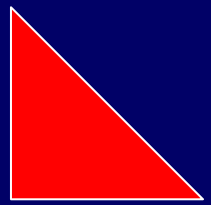
Fourth Normal Form



Under 4NF, a record type should not contain two or more independent multi-valued facts about an entity.

- A relation is in 4NF if it is in BCNF and it contains no multi-valued dependencies.
- A multi-valued fact may correspond to a many-many relationship or to a many-one relationship.
- A multi-valued dependency exists when there are three attributes (e.g. A,B, and C) in a relation, and for each value of A there is a well-defined set of valued B and a well defined set of values C. However, the set of values of B is independent of set C and vice-versa.

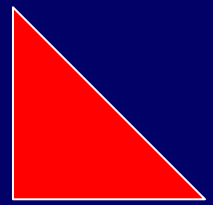
Example



- Consider information stored about movie stars. It includes details of their various addresses and the movies they starred in:
  - Name -> Address
  - Name -> Movie

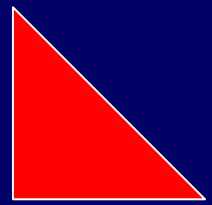
Name	Street	City	Title	Year
C. Fisher	123 Maple St.	Hollywood	Star Wars	1977
C. Fisher	5 Locust Ln.	Malibu	Star Wars	1977
C. Fisher	123 Maple St.	Hollywood	Empire Strikes Back	1980
C. Fisher	5 Locust Ln.	Malibu	Empire Strikes Back	1980
C. Fisher	123 Maple St.	Hollywood	Return of the Jedi	1983
C. Fisher	5 Locust Ln.	Malibu	Return of the Jedi	1983

Example cont...



- Carrie Fisher has two addresses and has been in three movies
- The only way to express the fact that addresses and movies are independent is to have each address appear with each movie
  - but this has introduced redundancy
- There is no BCNF violation
  - But the relation is not in 4NF
  - We need to break it up into 2 tables

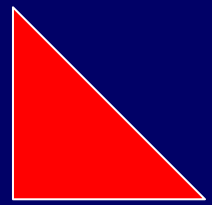
Example cont...



Name	Street	City
C. Fisher	123 Maple St.	Hollywood
C. Fisher	5 Locust Ln.	Malibu

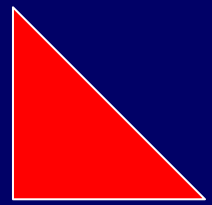
Name	Title	Year
C.Fisher	Star Wars	1977
C.Fisher	Empire Strikes Back	1980
C.Fisher	Return of the Jedi	1983

Fifth Normal Form



- 5NF is designed to cope with a type of dependency called join dependency
  - A relation that has a join dependency cannot be decomposed by a projection into other relations without spurious results
  - a relation is in 5NF when its information content cannot be reconstructed from several smaller relations
    - i.e. from relations having fewer attributes than the original relation

Join Dependency Decomposition

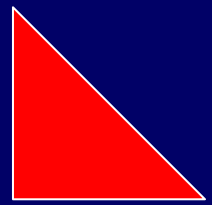


Name	Language	Hobby
C. Fisher	French	Cooks
C. Fisher	Spanish	Cooks
C. Fisher	English	Writes
M. Brown	Spanish	Read
M. Brown	Italian	Cook
K. Clark	Italian	Cook
K. Clark	Japanese	Decorating

Name	Language
C. Fisher	French
C. Fisher	Spanish
C. Fisher	English
M. Brown	Spanish
M. Brown	Italian
K. Clark	Italian
K. Clark	Japanese

Name	Hobby
C. Fisher	Cooks
C. Fisher	Writes
M. Brown	Read
M. Brown	Cook
K. Clark	Cook
K. Clark	Decorating

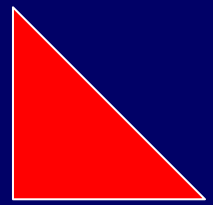
Spurious results



Name	Language	Hobby
C. Fisher	French	Cooks
C. Fisher	French	Writes
C. Fisher	Spanish	Cooks
C. Fisher	Spanish	Writes
C. Fisher	English	Cooks
C. Fisher	English	Writes
M. Brown	Spanish	Read

Name	Language	Hobby
M. Brown	Spanish	Cook
M. Brown	Italian	Read
M. Brown	Italian	Cook
K. Clark	Italian	Cook
K. Clark	Italian	Decorating
K. Clark	Japanese	Cook
K. Clark	Japanese	Decorating

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!