

Normalisation

G51DBI – Databases and Interfaces

Yorgos Tzimiropoulos

yorgos.tzimiropoulos@nottingham.ac.uk

This Lecture

- Normalisation
 - Data Redundancy
 - Functional Dependencies
 - Normal Forms
 - First, Second and Third Normal Forms
- Further reading
 - The Manga Guide to Databases, Chapter 3
 - Database Systems, Chapter 14

2

Design by Decomposition

- Normalisation is about Database Design
- Given some data/attributes we want to come up with the tables (and the schemas) for our Database
- Approach: start with a **Mega Table** that contains all data
- This Mega Table is not the best data structure because there are a lot of redundancies creating INSERT Anomalies, UPDATE Anomalies DELETE Anomalies
- To remove redundancies we first quantify them in terms of functional dependencies. Then we remove functional dependencies by splitting appropriately the Mega Table into smaller tables

3

Design by Decomposition

1NF

| Module | Dept | Lecturer | Text |
|--------|------|----------|------|
| M1 | D1 | L1 | T1 |
| M1 | D1 | L1 | T2 |
| M2 | D1 | L1 | T1 |
| M2 | D1 | L1 | T3 |
| M3 | D1 | L2 | T4 |
| M4 | D2 | L3 | T1 |
| M4 | D2 | L3 | T5 |
| M5 | D2 | L4 | T6 |

- Start with a Mega Table that contains all data
- There is a lot of data redundancy
- Because of redundancies there are
- INSERT Anomalies
 - Can't add a lecturer with no texts
- UPDATE Anomalies
 - To change the lecturer for M1, we will need to update two rows
- DELETE Anomalies
 - If we remove M3, we will remove L2 and T4 as well

4

First Normal Form

- In most definitions of the relational model
 - All data values should be atomic
 - This means that table entries should be single values, not sets or composite objects
 - Simplifies queries and data comparisons
- A relation is said to be first normal form (1NF) if all data values are atomic

5

Normalisation to 1NF

- To convert any relation into 1NF, split any non-atomic values

Unnormalised

| Module | Dept | Lecturer | Texts |
|--------|------|----------|--------|
| M1 | D1 | L1 | T1, T2 |
| M2 | D1 | L1 | T1, T3 |
| M3 | D2 | L2 | T4 |
| M4 | D2 | L3 | T1, T5 |
| M5 | D2 | L4 | T6 |

1NF

| Module | Dept | Lecturer | Text |
|--------|------|----------|------|
| M1 | D1 | L1 | T1 |
| M1 | D1 | L1 | T2 |
| M2 | D1 | L1 | T1 |
| M2 | D1 | L1 | T3 |
| M3 | D1 | L2 | T4 |
| M4 | D2 | L3 | T1 |
| M4 | D2 | L3 | T5 |
| M5 | D2 | L4 | T6 |

6

Problems with 1NF

1NF

| Module | Dept | Lecturer | Text |
|--------|------|----------|------|
| M1 | D1 | L1 | T1 |
| M1 | D1 | L1 | T2 |
| M2 | D1 | L1 | T1 |
| M2 | D1 | L1 | T3 |
| M3 | D1 | L2 | T4 |
| M4 | D2 | L3 | T1 |
| M4 | D2 | L3 | T5 |
| M5 | D2 | L4 | T6 |

- INSERT Anomalies
 - Can't add a lecturer with no texts
- UPDATE Anomalies
 - To change the lecturer for M1, we will need to update two rows
- DELETE Anomalies
 - If we remove M3, we will remove L2 and T4 as well

7

Functional Dependencies?

1NF

| Module | Dept | Lecturer | Text |
|--------|------|----------|------|
| M1 | D1 | L1 | T1 |
| M1 | D1 | L1 | T2 |
| M2 | D1 | L1 | T1 |
| M2 | D1 | L1 | T3 |
| M3 | D1 | L2 | T4 |
| M4 | D2 | L3 | T1 |
| M4 | D2 | L3 | T5 |
| M5 | D2 | L4 | T6 |

- Find Primary key
- Identify all FDs up to this point:

8

Functional Dependencies?

1NF

| Module | Dept | Lecturer | Text |
|--------|------|----------|------|
| M1 | D1 | L1 | T1 |
| M1 | D1 | L1 | T2 |
| M2 | D1 | L1 | T1 |
| M2 | D1 | L1 | T3 |
| M3 | D1 | L2 | T4 |
| M4 | D2 | L3 | T1 |
| M4 | D2 | L3 | T5 |
| M5 | D2 | L4 | T6 |

- Find Primary key
- Identify all FDs up to this point:

PK = M,T
FDs are
M,T → D,L
M → D, L
L → D

9

Functional Dependencies

1NF

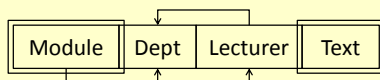
| Module | Dept | Lecturer | Text |
|--------|------|----------|------|
| M1 | D1 | L1 | T1 |
| M1 | D1 | L1 | T2 |
| M2 | D1 | L1 | T1 |
| M2 | D1 | L1 | T3 |
| M3 | D1 | L2 | T4 |
| M4 | D2 | L3 | T1 |
| M4 | D2 | L3 | T5 |
| M5 | D2 | L4 | T6 |

- The Primary Key is {Module, Text} so
 - {Module, Text} → {Dept, Lecturer}
- Other FDs are
 - {Module} → {Lecturer}
 - {Module} → {Dept}
 - {Lecturer} → {Dept}

10

FD Diagrams

- Rather than an entire table, FDs can be represented simply using the headings:



- {Module, Text} is a candidate key, so we put a double box around them
- {Lecturer} → {Dept}, so we have an arrow from Lecturer to Dept
- {Module} → {Dept} and {Module} → {Lecturer}, so we have {Module} → {Dept, Lecturer}

Note: Trivial FDs and FDs dependent on an entire candidate key are not included

11

Second Normal Form

- Partial FDs:
 - A FD, $A \rightarrow B$ is a partial FD, if some attribute of A can be removed and the FD still holds
 - Formally, there is some proper subset of A, $C \subset A$, such that $C \rightarrow B$
- Second normal form:
 - A relation is in second normal form (2NF) if it is in 1NF and no non-key attribute is partially dependent on a candidate key
 - In other words, no $C \rightarrow B$ where C is a strict subset of a candidate key and B is a non-key attribute.

12

Second Normal Form

More practical:

Assume A,C is a key, and B is a non-key attribute then if

$A, C \rightarrow B$ and

$A \rightarrow B$

then the relation is not in the 2nd normal form.

To bring it to 2nd form split the table into 2 tables:

??

13

Second Normal Form

More practical:

Assume A,C is a key, and B is a non-key attribute then if

$A, C \rightarrow B$ and

$A \rightarrow B$

then the relation is not in the 2nd normal form.

To bring it to 2nd form split the table into 2 tables:

AUB

AUCUOthers

14

Second Normal Form

Partial FDs:

- A FD, $A \rightarrow B$ is a partial FD, if some attribute of A can be removed and the FD still holds
- Formally, there is some proper subset of A, $C \subset A$, such that $C \rightarrow B$

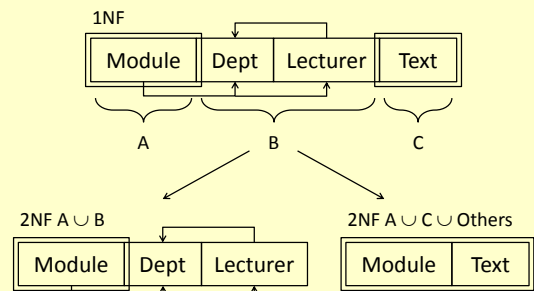
Second normal form:

- A relation is in second normal form (2NF) if it is in 1NF and no non-key attribute is partially dependent on a candidate key
- In other words, no $C \rightarrow B$ where C is a strict subset of a candidate key and B is a non-key attribute.

Candidate Keys are minimal!!

15

Normalising to 2NF



16

Normalising to 2NF

| Module | Dept | Lecturer |
|--------|------|----------|
| M1 | D1 | L1 |
| M2 | D1 | L1 |
| M3 | D1 | L2 |
| M4 | D2 | L3 |
| M5 | D2 | L4 |

| Module | Text |
|--------|------|
| M1 | T1 |
| M1 | T2 |
| M2 | T1 |
| M2 | T3 |
| M3 | T4 |
| M4 | T1 |
| M4 | T5 |
| M5 | T6 |

17

Problems Resolved in 2NF

INSERT Anomalies

- We can now add a lecturer without texts
- However, we cannot add a lecturer without teaching a module

UPDATE Anomalies

- We only need to change a single row when changing the lecturer of a module
- However, we need to change 2 rows to update the department for L1

Delete Anomalies

- If we delete M3 then all information about L2 is lost

| Module | Text |
|--------|------|
| M1 | T1 |
| M1 | T2 |
| M2 | T1 |
| M2 | T3 |
| M3 | T4 |
| M4 | T1 |
| M4 | T5 |
| M5 | T6 |

| Module | Dept | Lecturer |
|--------|------|----------|
| M1 | D1 | L1 |
| M2 | D1 | L1 |
| M3 | D1 | L2 |
| M4 | D2 | L3 |
| M5 | D2 | L4 |

18

Functional Dependencies?

2NFa

| Module | Dept | Lecturer |
|--------|------|----------|
| M1 | D1 | L1 |
| M2 | D1 | L1 |
| M3 | D1 | L2 |
| M4 | D2 | L3 |
| M5 | D2 | L4 |

FDs so far?

19

Functional Dependencies?

2NFa

| Module | Dept | Lecturer |
|--------|------|----------|
| M1 | D1 | L1 |
| M2 | D1 | L1 |
| M3 | D1 | L2 |
| M4 | D2 | L3 |
| M5 | D2 | L4 |

FDs so far?

$M \rightarrow D, L$

$L \rightarrow D$

which imply

$M \rightarrow L \rightarrow D$

Notice that M is a key for the sub-table

20

Transitive FDs and 3NF

- Transitive FDs:
 - A FD, $A \rightarrow C$ is a transitive FD, if there is some set B such that $A \rightarrow B$ and $B \rightarrow C$ and these are non-trivial FDs
 - $A \rightarrow B$ non-trivial means: B is not a subset of A
 - Essentially $A \rightarrow B \rightarrow C$
- Third normal form
 - A relation is in third normal form (3NF) if it is in 2NF and no non-key attribute is transitively dependent on a candidate key

21

Third Normal Form

More practical:

Assume A is a key, and B, C are non-key attributes then if

$A \rightarrow B \rightarrow C$ and

$A \rightarrow C$

then the relation is not in the 3rd normal form.

To bring it to 3rd form split the table into 2 tables:

??

22

Third Normal Form

More practical:

Assume A is a key, and B, C are non-key attributes then if

$A \rightarrow B \rightarrow C$ and

$A \rightarrow C$

then the relation is not in the 3rd normal form.

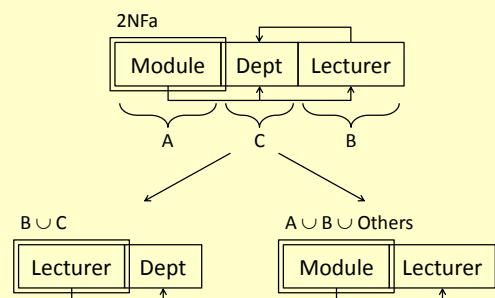
To bring it to 3rd form split the table into 2 tables:

$A \cup B \cup \text{Others}$

$B \cup C$

23

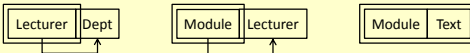
Normalising to 3NF



24

Normalising to 3NF

| 3NFa | | 3NFb | | 2NFb | |
|----------|------|--------|----------|--------|------|
| Lecturer | Dept | Module | Lecturer | Module | Text |
| L1 | D1 | M1 | L1 | M1 | T1 |
| L2 | D1 | M2 | L1 | M1 | T2 |
| L3 | D2 | M3 | L2 | M2 | T1 |
| L4 | D2 | M4 | L3 | M3 | T4 |
| | | M5 | L4 | M4 | T1 |
| | | | | M4 | T5 |
| | | | | M5 | T6 |



25

Problems Resolved in 3NF

- Problems resolved in 3NF
 - INSERT – We can now add Lecturers who don't teach any modules
 - UPDATE – We need only change a single row to update the department for L1
 - DELETE – We can delete M3 while preserving L2

| 3NFa | | 3NFb | |
|----------|------|--------|----------|
| Lecturer | Dept | Module | Lecturer |
| L1 | D1 | M1 | L1 |
| L2 | D1 | M2 | L1 |
| L3 | D2 | M3 | L2 |
| L4 | D2 | M4 | L3 |
| | | M5 | L4 |

26

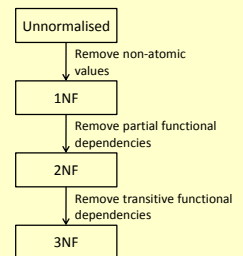
Normalisation and Design

- Normalisation is related to Database design
 - A database should normally be in 3NF at least
 - If your design leads to a non-3NF database, then you might want to revise it
- When you find you have a non-3NF database
 - Identify the FDs that are causing a problem
 - Think if they will lead to any insert, update, or delete anomalies
 - Try to remove them

27

Summary

- Normalisation
 - Data Redundancy
 - Functional Dependencies
 - Normal Forms
 - First, Second and Third Normal Forms



28

Another Example

| Unnormalised | | | | | | |
|--------------|-----------|------------|------------|------------|---------------|--------------|
| orderID | orderDate | customerID | cAddress | stockNos | stockQuant | stockPrices |
| 100152 | 12-11-10 | C1035 | 5 Ar... | 10,98,14 | 1,10,2 | 9.99,4.99... |
| 100236 | 19-11-10 | C1011 | 7 Be... | 59,13,... | 1,1,2,1,1,... | 0.99,3.99... |
| 101562 | 01-02-11 | C2693 | Flat 1a... | 7,45,9,... | 10,10,1,... | 2.99,3.49... |
| 102648 | 26-02-11 | C1011 | 7 Be... | 59,56,... | 1,5,3,4,6,... | 0.99,4.9... |

| orderID | orderDate | customerID | cAddress | stockNos | stockQuant | stockPrices |
|---------|-----------|------------|----------|----------|------------|-------------|
|---------|-----------|------------|----------|----------|------------|-------------|

Currently the only candidate key is {orderID}

29

Example

| 1NF | | | | | | |
|---------|-----------|------------|----------|---------|------------|------------|
| orderID | orderDate | customerID | cAddress | stockNo | stockQuant | stockPrice |
| 100152 | 12-11-10 | C1035 | 5 Ar... | 10 | 1 | 9.99 |
| 100152 | 12-11-10 | C1035 | 5 Ar... | 98 | 10 | 4.99 |
| 100152 | 12-11-10 | C1035 | 5 Ar... | 14 | 2 | 6.99 |
| 100236 | 19-11-10 | C1011 | 7 Be... | 59 | 1 | 0.99 |
| 100236 | 19-11-10 | C1011 | 7 Be... | 13 | 1 | 3.99 |
| 100236 | 19-11-10 | C1011 | 7 Be... | 4 | 2 | 3.49 |

| orderID | orderDate | customerID | cAddress | stockNo | stockQuant | stockPrice |
|---------|-----------|------------|----------|---------|------------|------------|
|---------|-----------|------------|----------|---------|------------|------------|

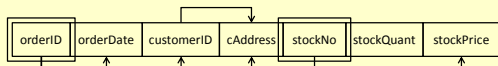
The only candidate key is now {orderID, stockNo}

30

Example

1NF

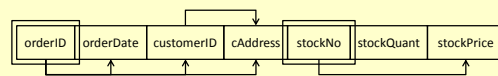
| orderID | orderDate | customerID | cAddress | stockNo | stockQuant | stockPrice |
|---------|-----------|------------|----------|---------|------------|------------|
| 100152 | 12-11-10 | C1035 | 5 Ar... | 10 | 1 | 9.99 |
| 100152 | 12-11-10 | C1035 | 5 Ar... | 98 | 10 | 4.99 |
| 100152 | 12-11-10 | C1035 | 5 Ar... | 14 | 2 | 6.99 |
| 100236 | 19-11-10 | C1011 | 7 Be... | 59 | 1 | 0.99 |
| 100236 | 19-11-10 | C1011 | 7 Be... | 13 | 1 | 3.99 |
| 100236 | 19-11-10 | C1011 | 7 Be... | 4 | 2 | 3.49 |



31

Example

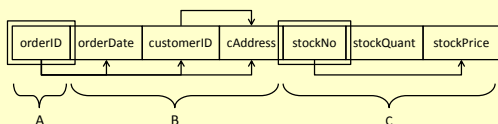
- This database does not adhere to 2NF
 - There are non-key attributes partially dependent on a candidate key



32

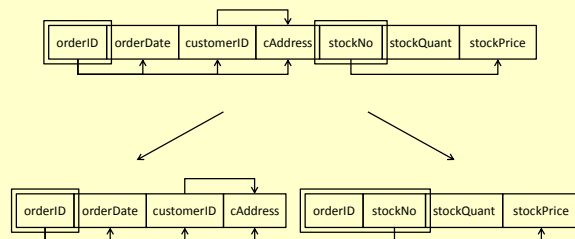
Example

- To remove the FD $A \rightarrow B$, where C is all other attributes
 - Create two new relations $A \cup B$ and $A \cup C$



33

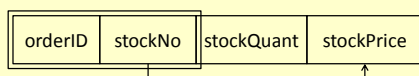
Example



34

Example

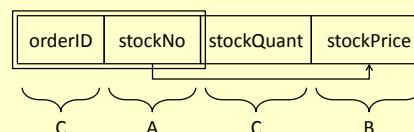
- One of the relations is still not in 2NF
 - $\{\text{stockPrice}\}$ is partially dependent on $\{\text{orderID}, \text{stockNo}\}$



35

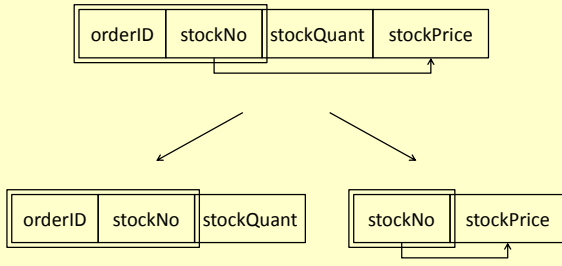
Example

- One of the relations is still not in 2NF
 - As before, we need to create two new relations $A \cup B$ and $A \cup C$



36

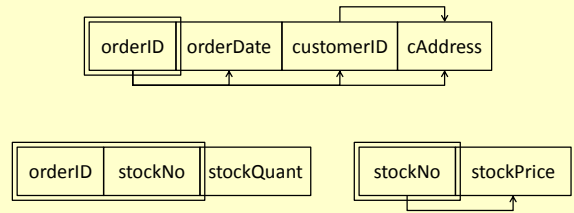
Example



37

Example

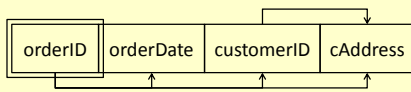
- This database is now in 2NF, but it isn't in 3NF
 - A transitive functional dependency exists



38

Example

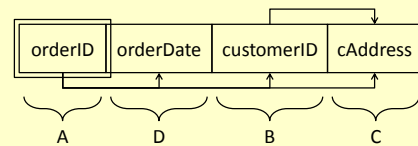
- This relation is not in 3NF
 - {cAddress} is transitively dependent on {orderID} via {customerID}



39

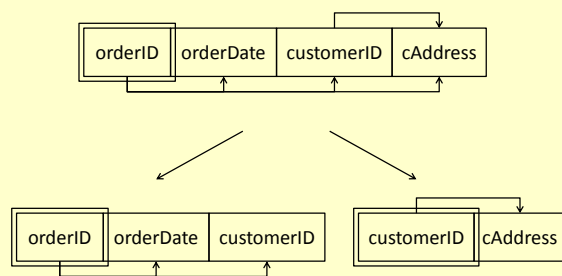
Example

- To remove the Transitive FD $A \rightarrow B \rightarrow C$, where D is all other attributes
 - Create two new relations $A \cup B \cup D$ and $B \cup C$



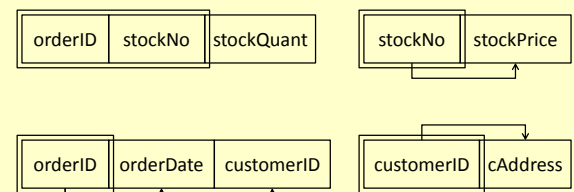
40

Example



41

3NF Database



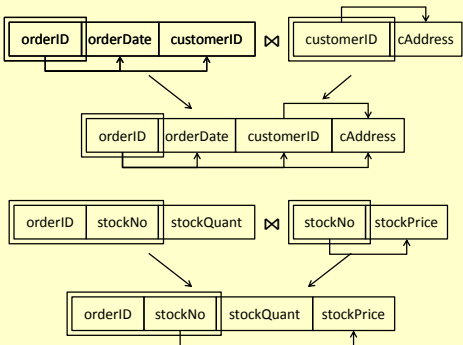
42

Lossless Decomposition

- Decomposition of tables is lossless if we can recover the original relation through a join
- A natural join is the most convenient way to do this, although most joins will work
- Lossless decomposition ensures that we haven't removed any data from our database
- All data can be retrieved again using joins if required

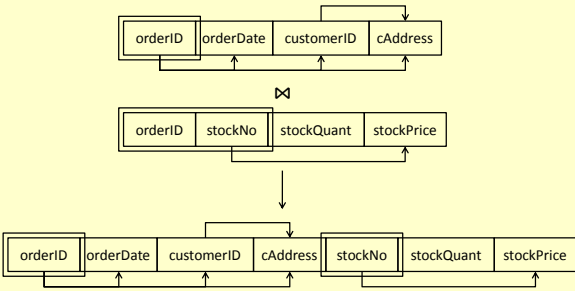
43

Lossless Decomposition



44

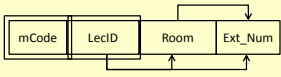
Lossless Decomposition



45

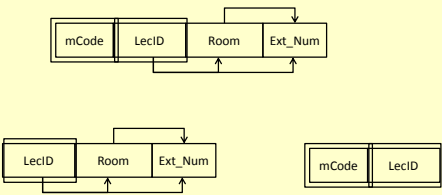
Example 2

Given the relation with the attributes below and the given functional dependencies determine which normal form it is in and convert it into Third Normal Form.



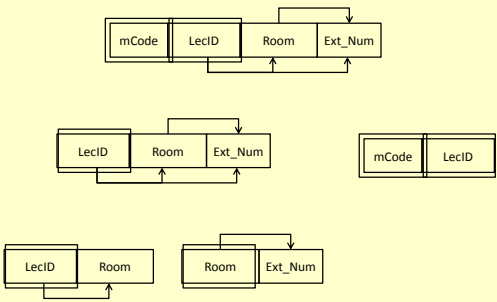
46

Example 2



47

Example 2



48

Example 3

1NF

| AdmNumb | AdmDate | WardID | WardName | PatNo | PatName |
|---------|----------|--------|------------|-------|---------|
| P001 | 03-02-13 | S123 | A & E | PT104 | Alex |
| P001 | 03-02-13 | S123 | A & E | PT105 | Chris |
| P002 | 10-06-13 | S234 | Urology | PT106 | Anne |
| P002 | 10-06-13 | S456 | Cardiology | PT106 | Anne |
| P003 | 10-11-13 | S123 | A & E | PT106 | Anne |
| P004 | 12-12-13 | S234 | Urology | PT105 | Chris |
| P004 | 12-12-13 | S123 | A & E | PT106 | Anne |

49

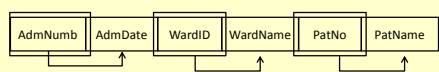
Example 3

1NF

| AdmNumb | AdmDate | WardID | WardName | PatNo | PatName |
|---------|----------|--------|------------|-------|---------|
| P001 | 03-02-13 | S123 | A & E | PT104 | Alex |
| P001 | 03-02-13 | S123 | A & E | PT105 | Chris |
| P002 | 10-06-13 | S234 | Urology | PT106 | Anne |
| P002 | 10-06-13 | S456 | Cardiology | PT106 | Anne |
| P003 | 10-11-13 | S123 | A & E | PT106 | Anne |
| P004 | 12-12-13 | S234 | Urology | PT105 | Chris |
| P004 | 12-12-13 | S123 | A & E | PT106 | Anne |

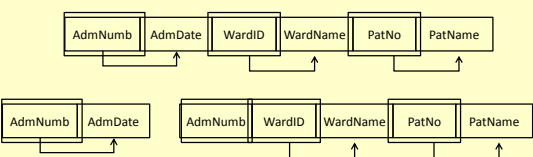
50

Example 3



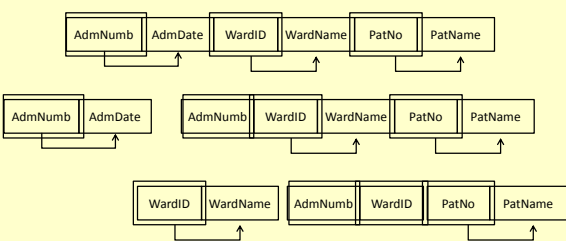
51

Example 3



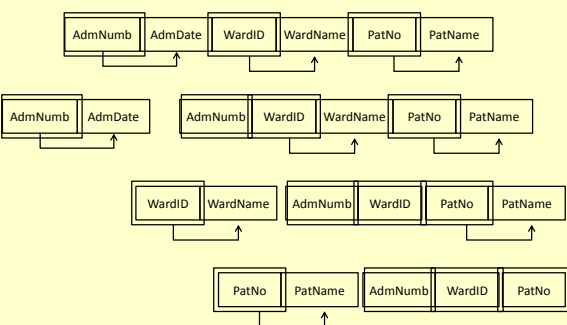
52

Example 3



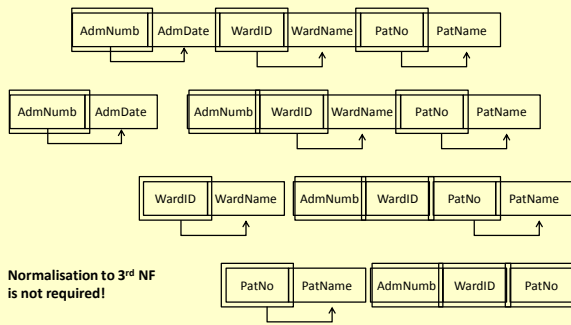
53

Example 3



54

Example 3



55

Take home messages

1. Data redundancy
2. Anomalies: Insert, Update, Delete
3. Functional Dependencies
4. Normal Forms
 - a. 1NF – data to be atomic
 - b. 2NF – remove partial FDs
 - c. 3NF – remove transitive FDs

56