Normalization

Gary KL Tam

Department of Computer Science
Swansea University

Overview

Redundancy and Functional Dependencies

- Can be determined from other data in the database
- Leads to various problems
  - INSERT Problem
  - UPDATE Problem
  - DELETE Problem
- Redundancy is often caused by a functional dependency

Normalisation

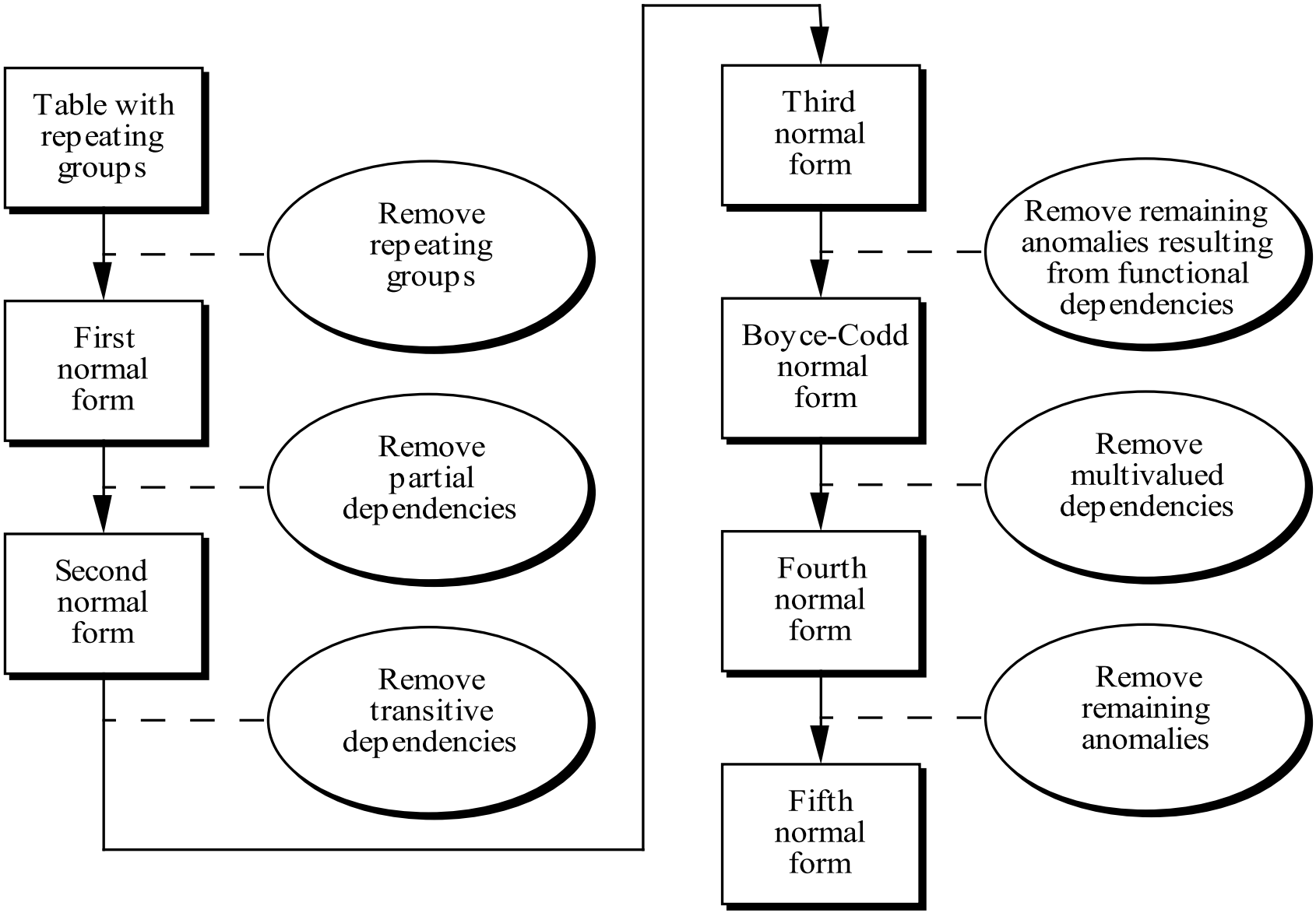
- Aims to reduce data redundancy
- By removing undesirable FDs

FDs and Normalisation

Normal Forms

- Each normal form i has fewer FDs than the last i-1
- Each normal form i has less redundancy than the last i-1, because violating FDs (redundancy) are removed
- e.g. 2NF has fewer FDs and redundancy than 1NF
- Note: Not all FDs cause a problem:
  - Identify violating FDs
  - Each normal form removes a type of FD

Normalisation - In one slide



Normalisation Forms - Quick reminder

1-3NF keywords:

1. no repeating / be atomic
2. no partial
3. no transitive

More definition and details follow.

Normalisation Procedure - In another slide

Normalisation - 4 Steps:

1. Identify primary key
   - draw functional dependency diagram
2. Identify violating functional dependencies
3. Split tables using the violating functional dependencies
4. Define new primary and foreign keys
   - and references respective tables

First Normal Form

First Normal Form (1NF)

- A relation is in 1NF if all data values are atomic

- All data values should be atomic
- i.e., table entries should be single values, not sets or composite objects
- Simplifies queries and data comparisons

Normalisation to 1NF

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

Unnormalised

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

e.g. A typical excel spreadsheet may look like this

1NF

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

Second Normal Form

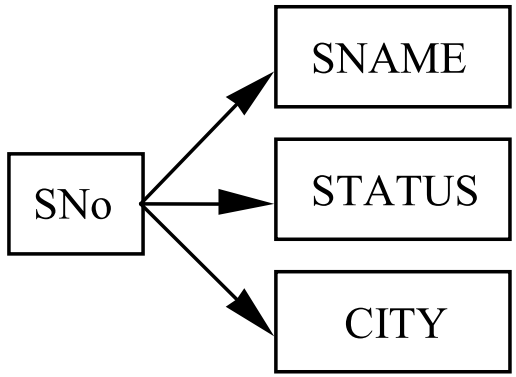
Second Normal Form (2NF)

A relation is in 2NF if it is

- in 1NF and
- every non-key attribute is fully functionally dependent on the primary key
- that's no partial functional dependency.

Supplier S

SNo	SName	Status	City
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clark	20	London
S5	Adams	30	Athens



- SNo is the primary key of S table.
- Attributes SName, Status and City are non-key attributes.
- SNo $\rightarrow$ SName, Status and City

Second Normal Form

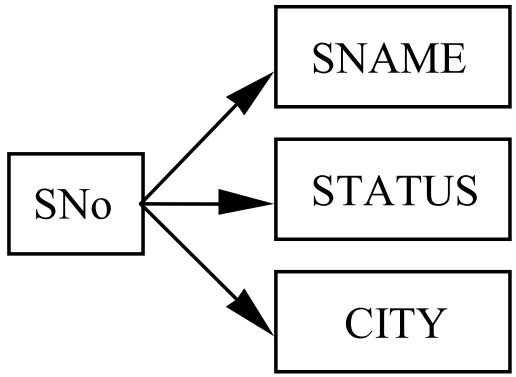
Second Normal Form (2NF)

A relation is in 2NF if it is

- in 1NF and
- every non-key attribute is fully functionally dependent on the primary key
- that's no partial functional dependency.

Supplier S

SNo	SName	Status	City
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clark	20	London
S5	Adams	30	Athens

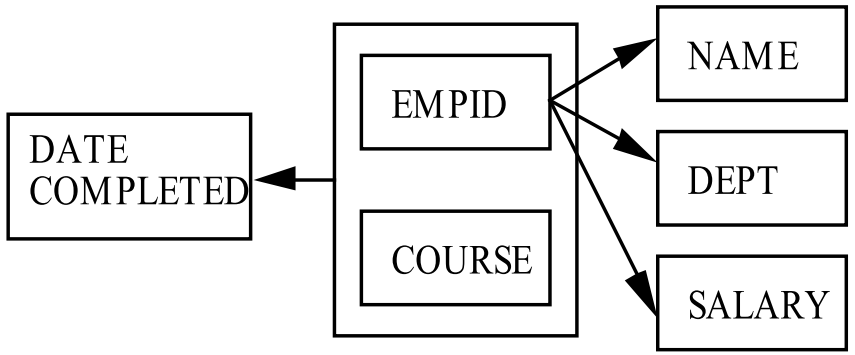


- SNo is the primary key of S table.
- Attributes SName, Status and City are non-key attributes.
- SNo $\rightarrow$ SName, Status and City ✓S is in 2NF.

In 2NF?

empid	name	dept	salary	course	date_completed
100	Margaret Simpson	Marketing	42000	SPSS	19/6/96
100	Margaret Simpson	Marketing	42000	Surveys	20/7/96
140	Alan Beeton	Accounting	39000	Tax Acc	21/8/96
110	Chris Lucero	IT	41500	SPSS	22/9/96
110	Chris Lucero	IT	41500	C++	23/10/96
190	Lorenzo Davis	Finance	38000	Investments	24/11/96
150	Susan Martin	Marketing	38500	SPSS	25/11/96
150	Susan Martin	Marketing	38500	TQM	12/1/97

- empid $\rightarrow$ name, dept, salary
- empid, course $\rightarrow$ date_completed
- candidate key = (empid, course)



Think

Is it in 2NF?

Normalisation to 2NF

Is it in 1NF?

Normalisation to 2NF

Is it in 1NF?

- ✓, there is no non-atomic values

Is it in 2NF?

Normalisation to 2NF

Is it in 1NF?

- ✓, there is no non-atomic values

Is it in 2NF?

- it is in 1NF (above)

Normalisation to 2NF

Is it in 1NF?

- ✓, there is no non-atomic values

Is it in 2NF?

- it is in 1NF (above)
- are all non-key attributes fully dependent on the primary key?

Normalisation to 2NF

Is it in 1NF?

- ✓, there is no non-atomic values

Is it in 2NF?

- it is in 1NF (above)
- are all non-key attributes fully dependent on the primary key?
- What is the Primary key?

Normalisation to 2NF

Is it in 1NF?

- ✓, there is no non-atomic values

Is it in 2NF?

- it is in 1NF (above)
- are all non-key attributes fully dependent on the primary key?
- What is the Primary key? (empid, course)

Normalisation to 2NF

Is it in 1NF?

- ✓, there is no non-atomic values

Is it in 2NF?

- it is in 1NF (above)
- are all non-key attributes fully dependent on the primary key?
- What is the Primary key? (empid, course)

- ✗, it is not in 2NF.

Normalisation to 2NF

Is it in 1NF?

- ✓, there is no non-atomic values

Is it in 2NF?

- it is in 1NF (above)
- are all non-key attributes fully dependent on the primary key?
- What is the Primary key? (empid, course)

- ✗, it is not in 2NF.
- Why? There are non-key attributes partially functionally dependent on the primary key.
- Violating FDs: empid $\rightarrow$ name, dept, salary

General Procedure to 2NF

R(A, B, C, D) PRIMARY KEY (A, B)

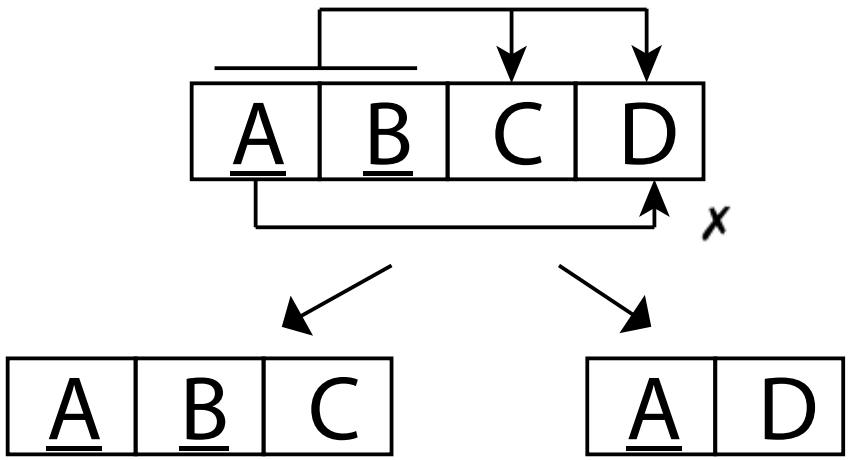
- R.(A,B) $\rightarrow$ R.C
- R.A $\rightarrow$ R.D

In 1NF, but not in 2NF
violating FD: R.A $\rightarrow$ R.D

Split to give:

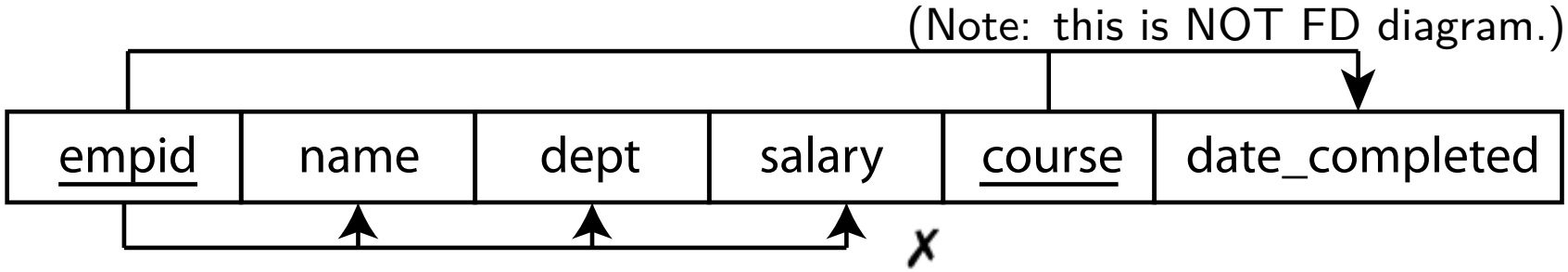
- R1(A,D) PRIMARY KEY (A)
- R2(A,B,C) PRIMARY KEY (A,B) FOREIGN KEY (A) REFERENCE R1

foreign key : integrity check.



Note: this is NOT FD diagram.

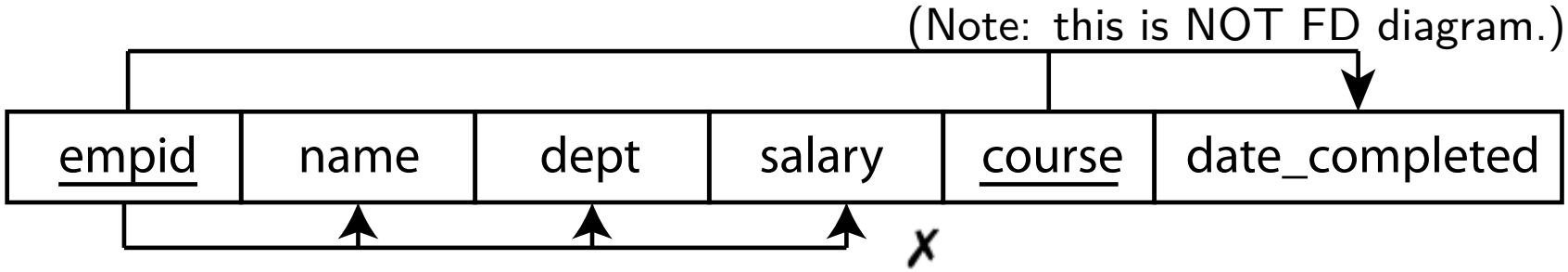
Example



- EMP.empid $\rightarrow$ EMP.(name,dept,salary)
- EMP.(empid,course) $\rightarrow$ EMP.date_completed

Using the 2NF splitting rule:

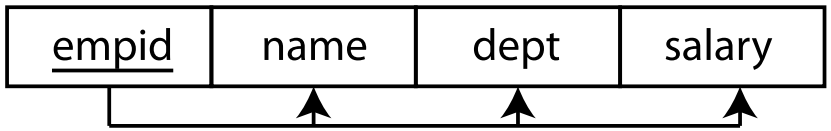
Example

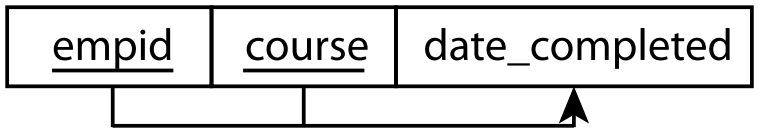


- EMP.empid → EMP.(name,dept,salary)
- EMP.(empid,course) → EMP.date_completed

Using the 2NF splitting rule:

- EMP(empid, name, dept, salary)
- COURSES(course, empid, date completed)
  FOREIGN KEY (empid) REFERENCE EMP (empid)





2NF, still bad

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

- course $\rightarrow$ staff, room
- staff $\rightarrow$ room

Is it in 2NF?

2NF, still bad

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

- course $\rightarrow$ staff, room
- staff $\rightarrow$ room

Is it in 2NF?

- It is in 1NF

2NF, still bad

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

- course $\rightarrow$ staff, room
- staff $\rightarrow$ room

Is it in 2NF?

- It is in 1NF
- There are NO non-key attributes partially functionally dependent on the primary key.

2NF, still bad

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

- course $\rightarrow$ staff, room
- staff $\rightarrow$ room

Is it in 2NF?

- It is in 1NF
- There are NO non-key attributes partially functionally dependent on the primary key.
- Note: staff is not part of the primary key

2NF, still bad

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

- course $\rightarrow$ staff, room
- staff $\rightarrow$ room

Is it in 2NF?

- It is in 1NF
- There are NO non-key attributes partially functionally dependent on the primary key.
- Note: staff is not part of the primary key
- ✓, it is in 2NF.

- The problems we saw earlier are due to the relation not being in third normal form.

Third Normal Form

Third Normal Form (3NF)

- A relation is in third normal form (3NF) if
  - it is in 2NF and
  - no transitive dependencies exist.
- A transitive dependency in a relation is a functional dependency between two (or more) non-key attributes

Supplier S

SNo	SName	Status	City
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clark	20	London
S5	Adams	30	Athens

SNo $\rightarrow$ SName, Status, City

Is it in 3NF?

- It is in 1NF.

Third Normal Form

Third Normal Form (3NF)

- A relation is in third normal form (3NF) if
  - it is in 2NF and
  - no transitive dependencies exist.
- A transitive dependency in a relation is a functional dependency between two (or more) non-key attributes

Supplier S

SNo	SName	Status	City
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clark	20	London
S5	Adams	30	Athens

SNo $\rightarrow$ SName, Status, City

Is it in 3NF?

- It is in 1NF.
- It is in 2NF.

Third Normal Form

Third Normal Form (3NF)

- A relation is in third normal form (3NF) if
  - it is in 2NF and
  - no transitive dependencies exist.
- A transitive dependency in a relation is a functional dependency between two (or more) non-key attributes

Supplier S

SNo	SName	Status	City
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clark	20	London
S5	Adams	30	Athens

SNo $\rightarrow$ SName, Status, City

Is it in 3NF?

- It is in 1NF.
- It is in 2NF.
- There is no transitive dependency.

Third Normal Form

Third Normal Form (3NF)

- A relation is in third normal form (3NF) if
  - it is in 2NF and
  - no transitive dependencies exist.
- A transitive dependency in a relation is a functional dependency between two (or more) non-key attributes

Supplier S

SNo	SName	Status	City
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clark	20	London
S5	Adams	30	Athens

SNo $\rightarrow$ SName, Status, City

Is it in 3NF?

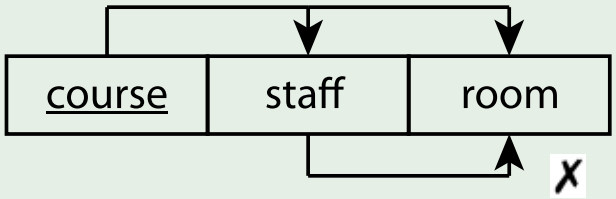
- It is in 1NF.
- It is in 2NF.
- There is no transitive dependency.
- ✓, S is in 3NF.

Example

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

L(course, staff, room)

- course → staff, room
- staff → room



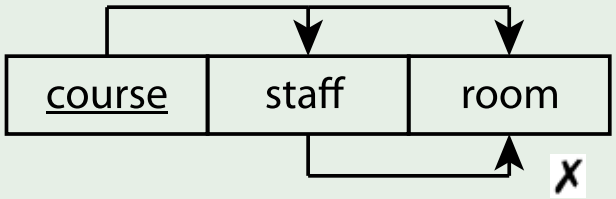
Is it in 3NF?

Example

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

L(course, staff, room)

- course → staff, room
- staff → room



Is it in 3NF?

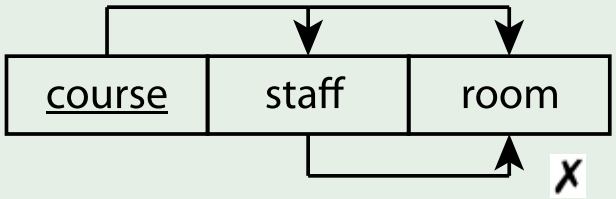
- it is in 1NF.

Example

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

L(course, staff, room)

- course → staff, room
- staff → room



Is it in 3NF?

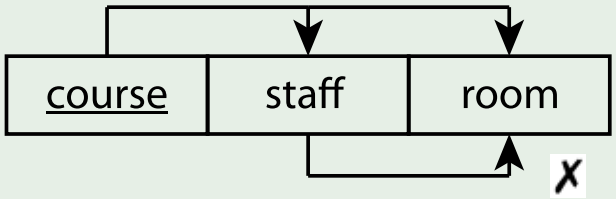
- it is in 1NF.
- it is in 2NF.

Example

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

L(course, staff, room)

- course → staff, room
- staff → room



Is it in 3NF?

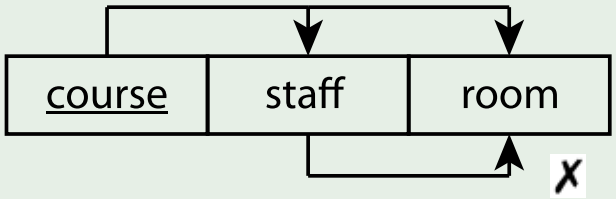
- it is in 1NF.
- it is in 2NF.
- There is a transitive dependency in the relation. ✗, it is NOT in 3NF.

Example

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

L(course, staff, room)

- course → staff, room
- staff → room



Is it in 3NF?

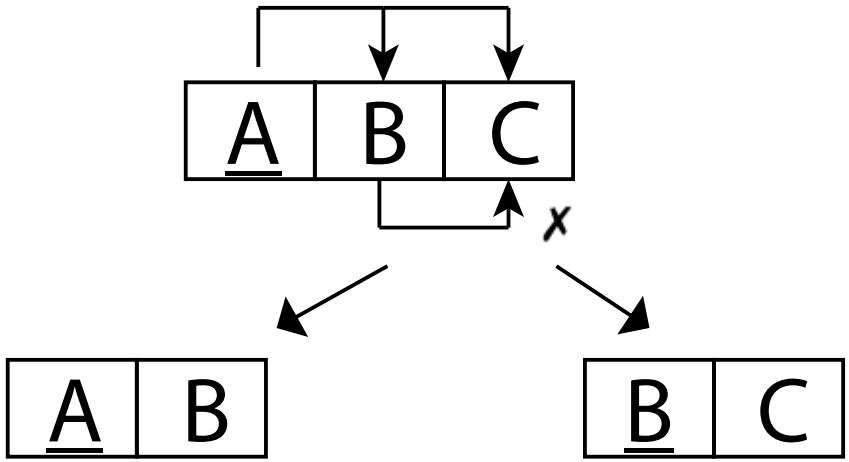
- it is in 1NF.
- it is in 2NF.
- There is a transitive dependency in the relation. ✗, it is NOT in 3NF.
- violating FD: staff → room

General Procedure to 3NF

R(A, B, C) PRIMARY KEY (A)

- R.(A) $\rightarrow$ R.(B,C)
- R.B $\rightarrow$ R.C

A could be composite.
2NF Relation, not in 3NF
violating FD: R.B $\rightarrow$ R.C



Note: this is NOT FD diagram.

Split to give:

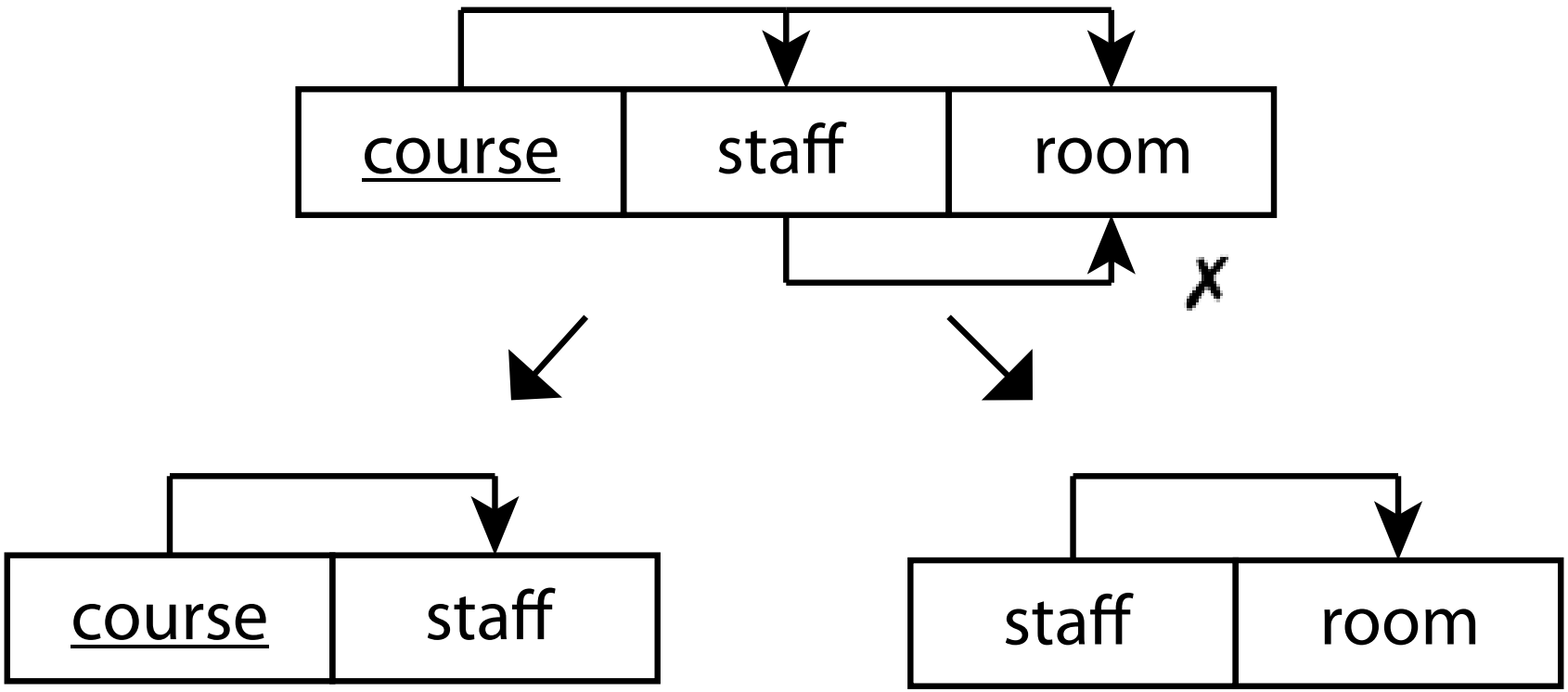
- R1(B, C) PRIMARY KEY (B)
- R2(A, B) PRIMARY KEY (A)
  FOREIGN KEY (B) REFERENCES
  R1 (B)

foreign key : integrity check.

Example

L(course, staff, room)

- course $\rightarrow$ staff, room
- staff $\rightarrow$ room



Note: this is NOT FD diagram.