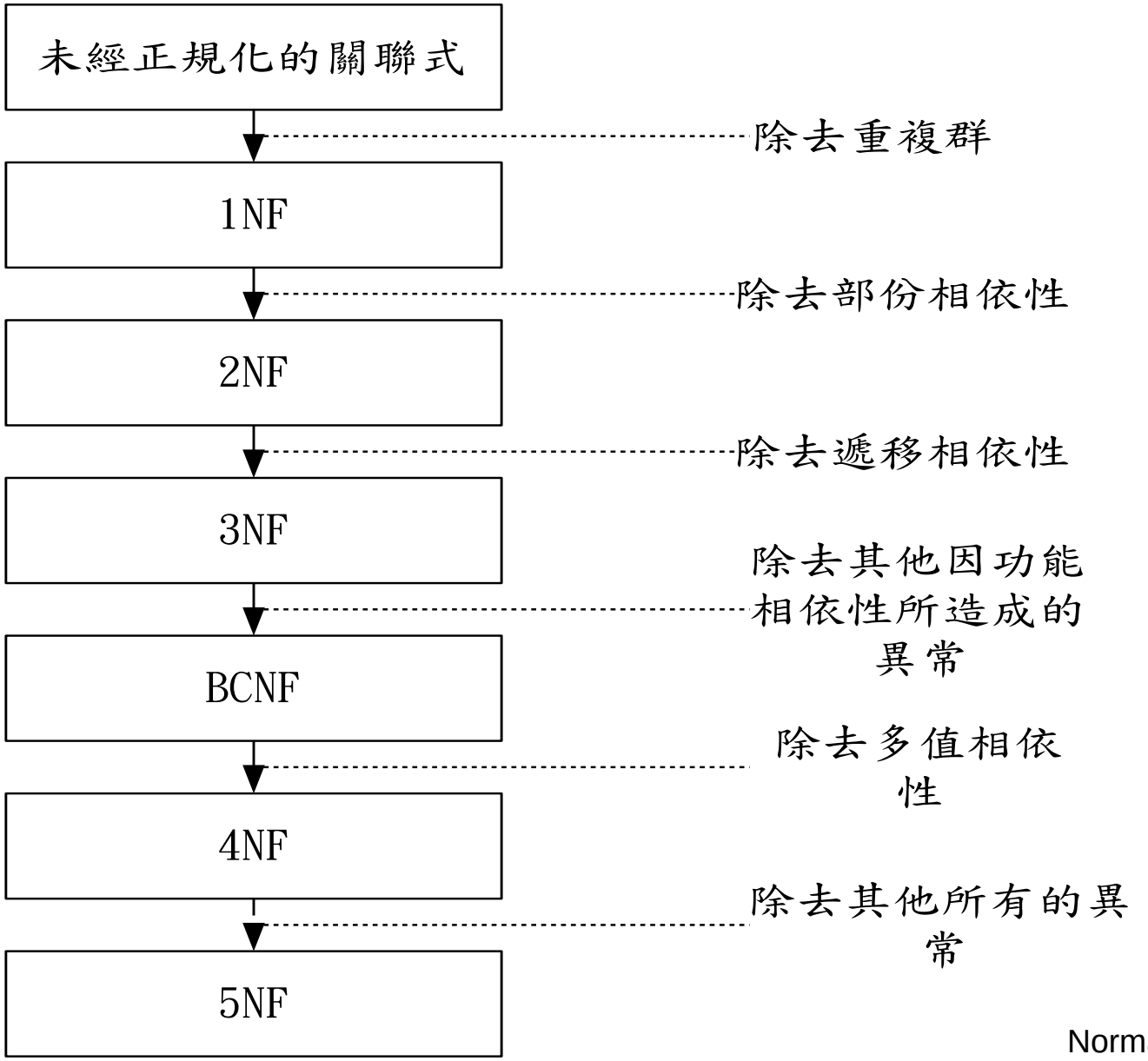
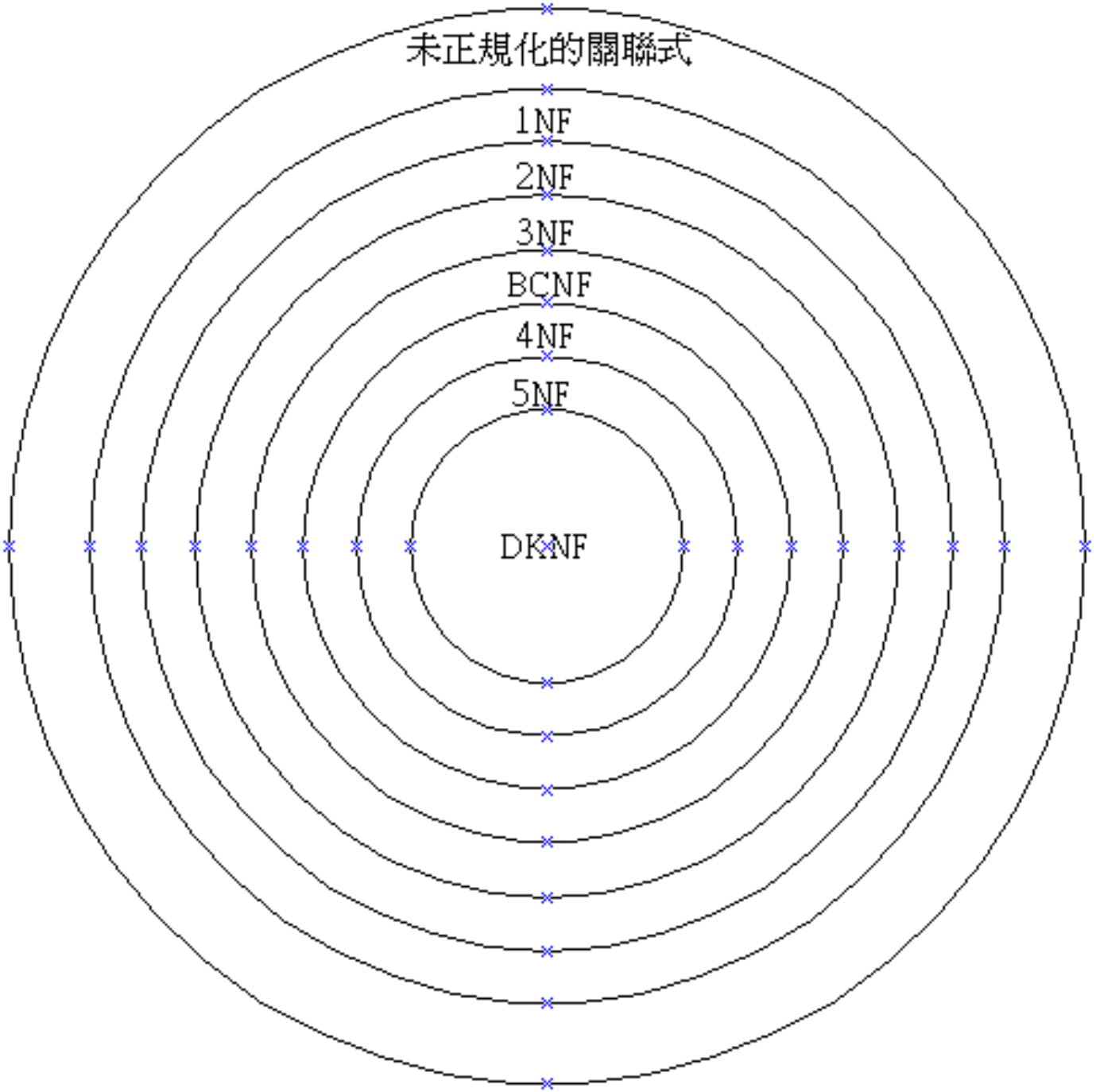
Normalization




未正規化的關聯式
1NF
2NF
3NF
BCNF
4NF
5NF
DKNF

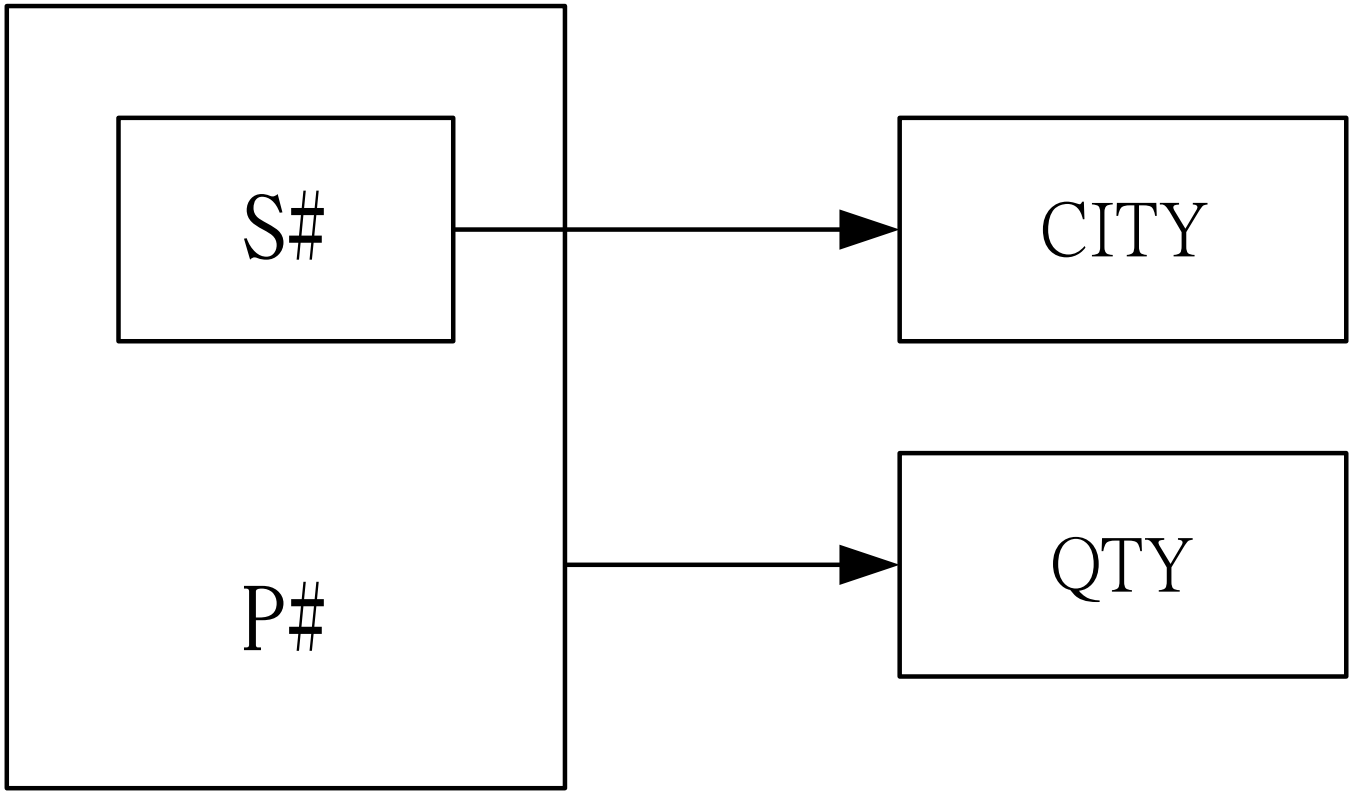
1. a high degree of redundancy.
2. The redundancy leads to a number of problems ◦ (after a update; cascade update?)

SP	S#	CITY	P#	QTY
	S1	London	P1	300
	S1	London	P2	200
	S1	London	P3	400
	S1	London	P4	200
	S1	London	P5	100
	S1	London	P6	100
	S2	Paris	P1	300
	S2	Paris	P2	400

功能相依性：Functional Dependency

- 關連R中若其屬性B功能相依於R中的屬性A，記為R.A→R.B。A的屬性值可以決定唯一的B屬性值。
- A,B可以是複合屬性。
- 若屬性B功能相依於複合屬性 A，但不功能相依於A的子集，則稱B『完全功能相依』於A。
- A稱為決定性屬性，B稱為相依性屬性。
- 若A為R中的一個候選鍵，則每個R中的屬性都會功能相依於A。
- 多個『功能相依性』可合併成『功能相依圖』。

畫出SP（S#, CITY, P#,QTY）的功能相依圖



關聯 SP 中的功能相依圖

- EMPLOYEE

ENAME	SSN	BDATE	ADDRESS	DNUMBER

- DEPARTMENT

DNAME	DNUMBER	DMGRSSN

- DEPT_LOCATIONS

DNUMBER	DLOCATION

- WORKS_ON

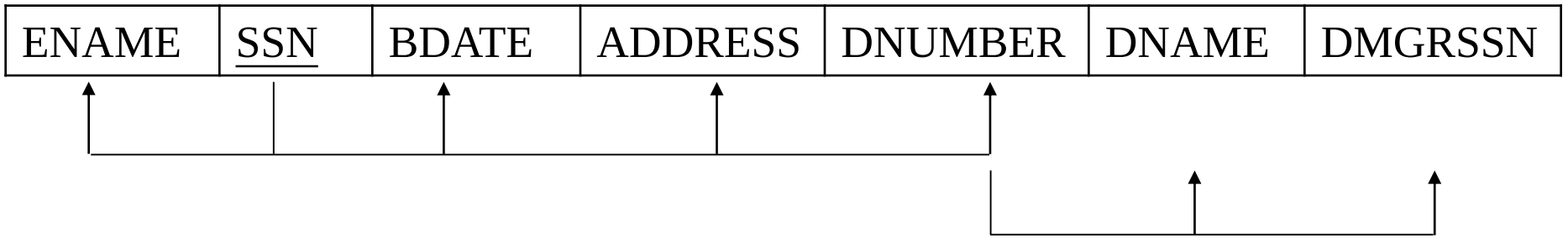
SSN	PNUMBER	HOURS

- PROJECT

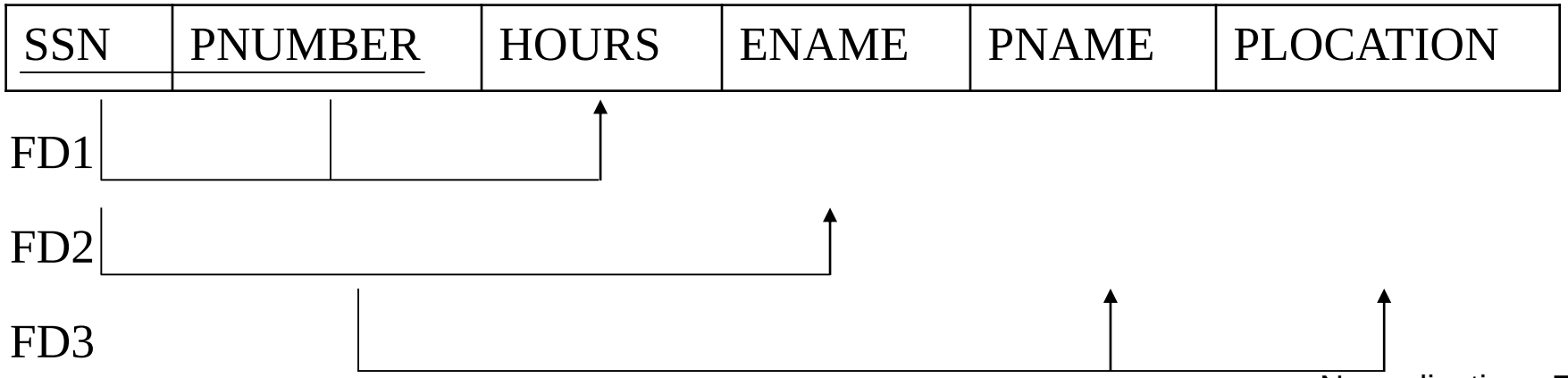
PNAME	PNUMBER	PLOCATION	DNUM

功能相依性的例子

(a) EMP_DEPT



(b) EMP_PROJ



B-EMP_DEPT與EMP_PROJ關連的Schema

EMP_DEPT

ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
Smith	1234	1965/01/09	731 Fondren	5	Research	33344555
English	4534	1972/07/31	5631 Rice	5	Research	33344555

EMP_PROJ

SSN	PNUMBER	HOURS	ENAME	PNAME	PLOCATION
1234	1	32.5	Smith	X	Taipei
1234	2	7.5	Smith	Y	Kaohsiung
4534	1	20.0	English	X	Taipei
4534	2	20.0	English	Y	Kaohsiung

-EMP_LOCS及EMP_PROJ1來表示的EMP_PROJ

(a)

EMP_LOCS

ENAME	PLOCATION

p.k.

EMP_PROJ1

SSN	PNUMBER	HOURS	PNAME	PLOCATION

p.k.

(b)

EMP_LOCS

ENAME	PLOCATION
Smith	Taipei
Smith	Kaohsiung
English	Taipei
English	Kaohsiung

EMP_PROJ1

SSN	PNUMBER	HOURS	PNAME	PLOCATION
1234	1	32.5	X	Taipei
1234	2	7.5	Y	Kaohsiung
4534	1	20.0	X	Taipei
4534	2	20.0	Y	Kaohsiung

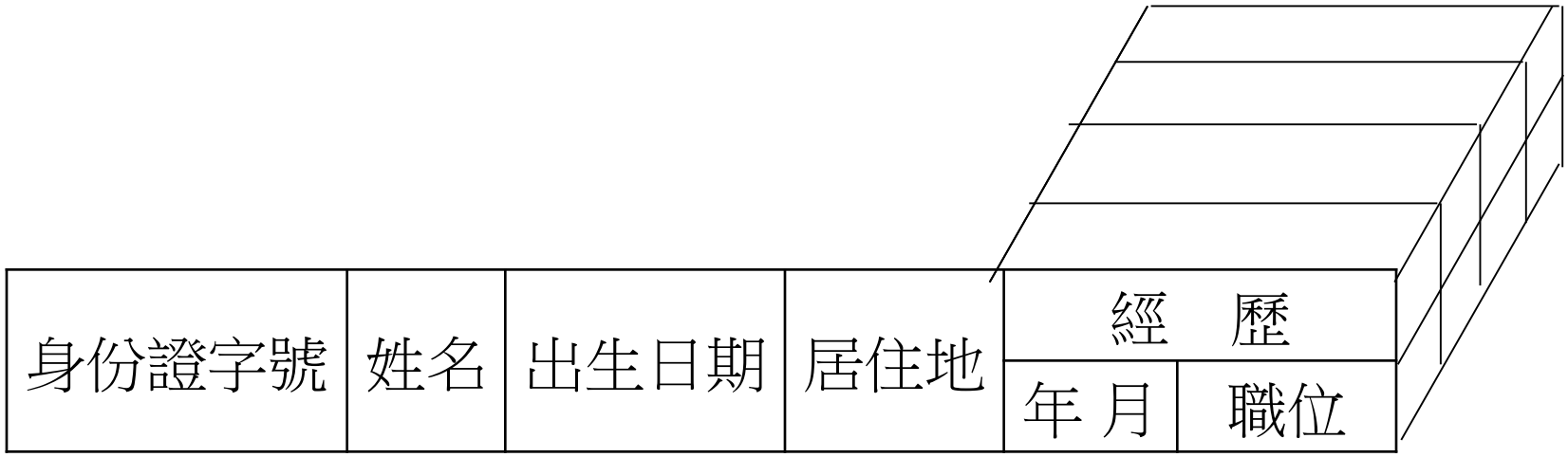
- 將PARTUAL JOIN操作應用在EMP_PROJ1與EMP_LOCS上的結果，其中標示了星號(*)代表假值組。

SSN	PNUMBER	HOURS	PNAME	PLOCATION	ENAME
1234	1	32.5	X	Taipei	Smith
*1234	1	32.5	X	Taipei	English
1234	2	7.5	Y	Kaohsiung	Smith
*1234	2	7.5	Y	Kaohsiung	English
*4534	1	20.0	X	Taipei	Smith
4534	1	20.0	X	Taipei	English
*4534	2	20.0	Y	Kaohsiung	Smith
4534	2	20.0	Y	Kaohsiung	English

1NF

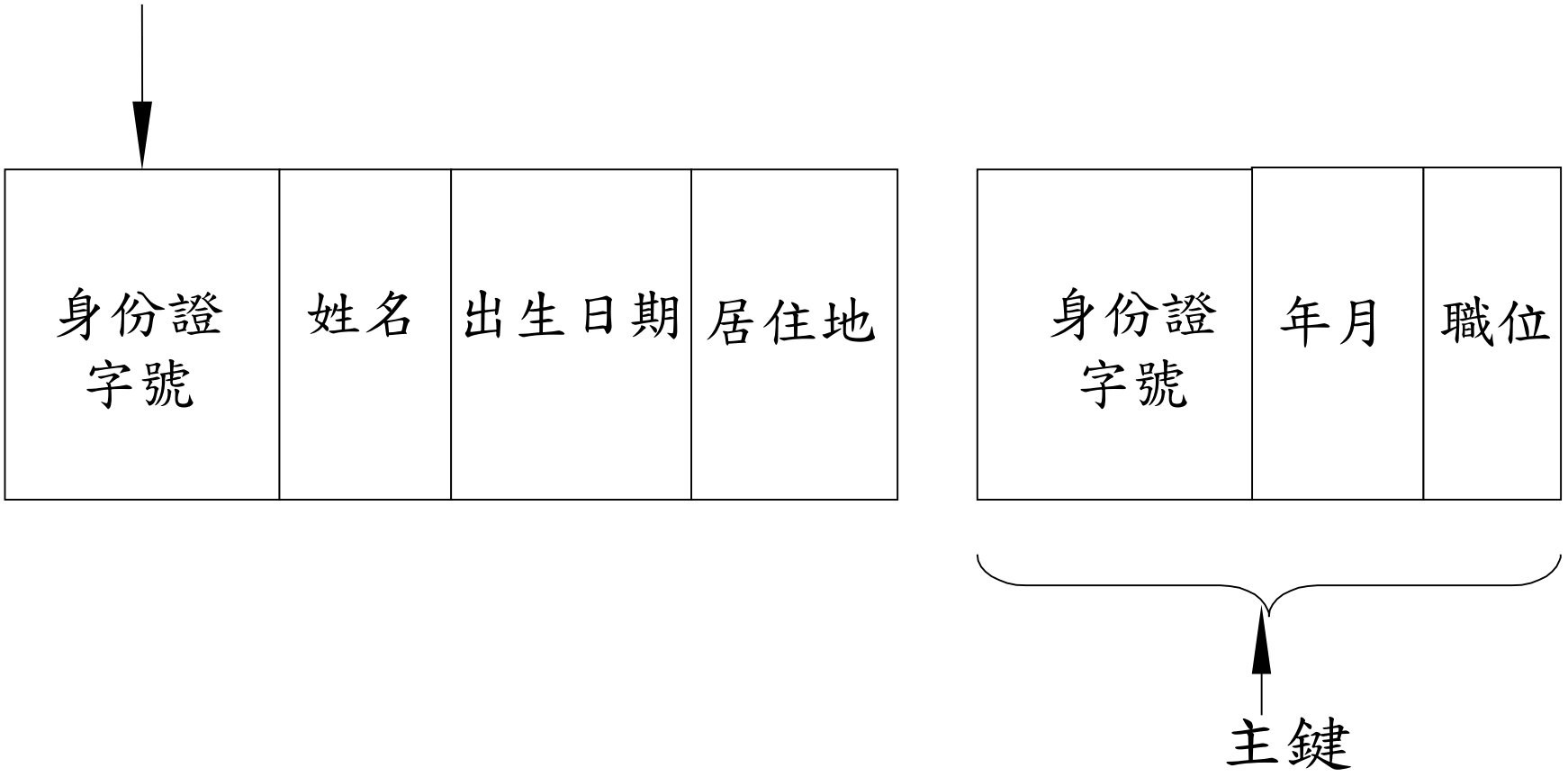
- 一個關連R為1NF　其所有屬性的定義域（Domain）僅含基元值。
- 所有重複的群組，即同一值組的所有欄位不可儲存一個以上的資料值。
- How to adjust
  1. 由R中除去重複的群組。
  2. R中的主鍵加上重複的群組。

例子：



非第一正規式的檔案

主鍵

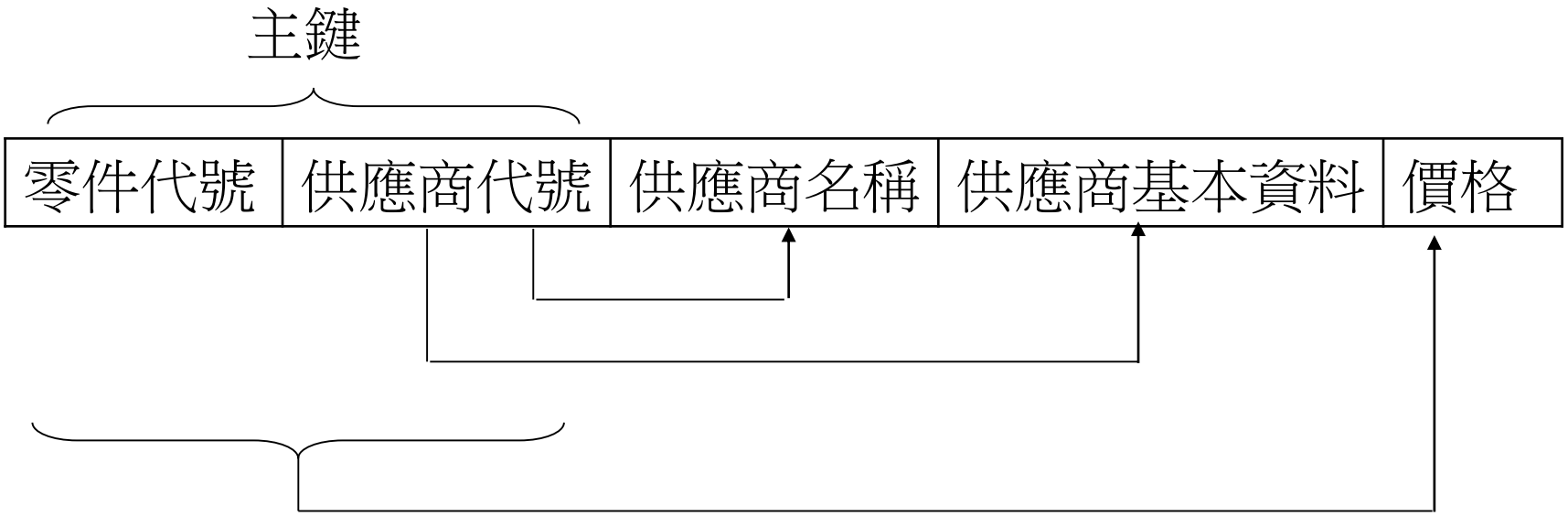


拆開後兩者都是第一正規式

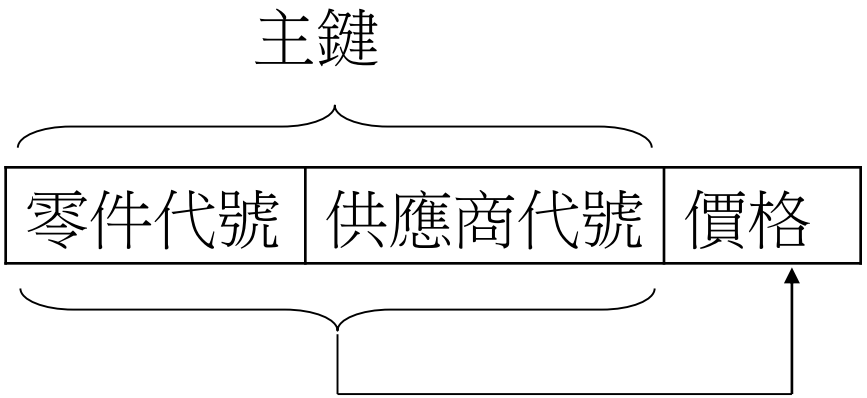
2NF

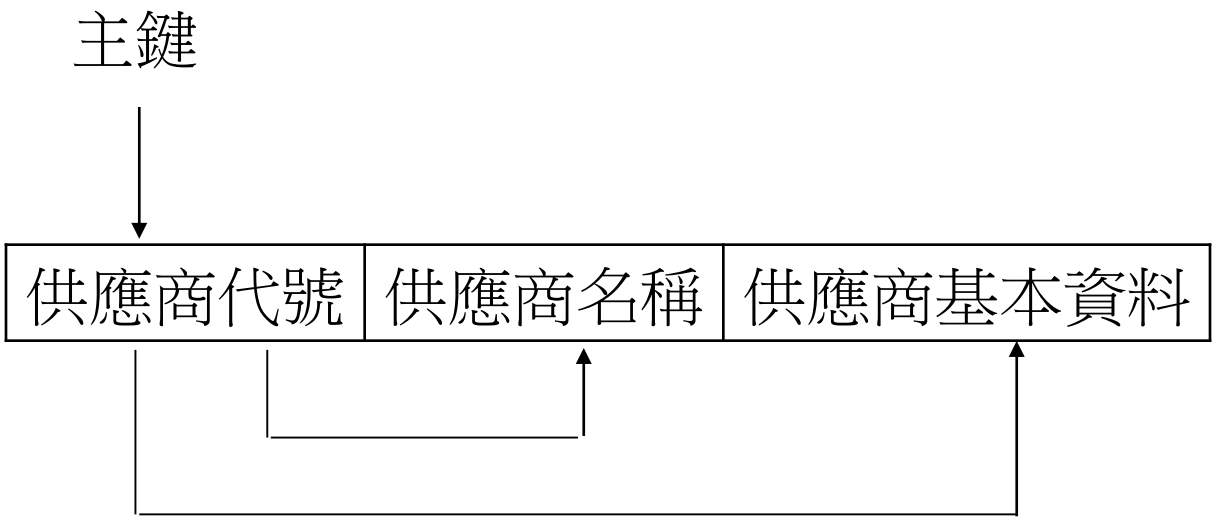
- 一個關連 R 為 2NF
  1. R為1NF且
  2. 所有非主鍵的屬性***完全功能相依*** 於主鍵的屬性。
- How to adjust
  1. 和主鍵非完全功能相依的屬性和其完全 功能相依的屬性組成關連。
  2. 原關連剩下的屬性完全功能相依於主鍵。

例子：



非第二正規式的檔案



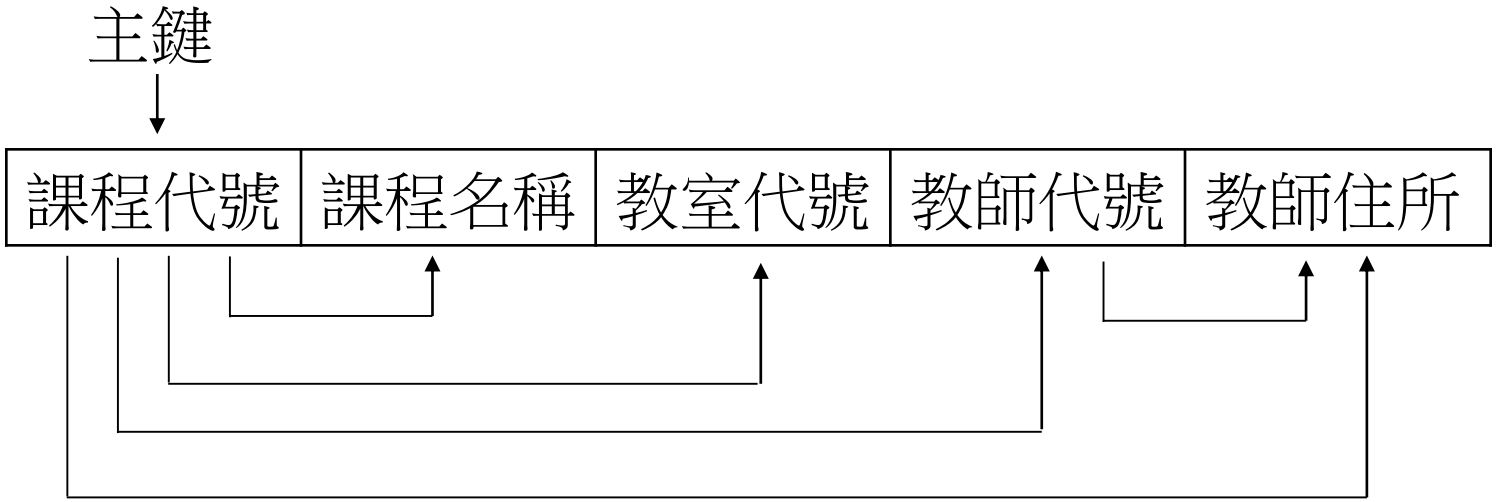


拆開後兩者都是第二正規式

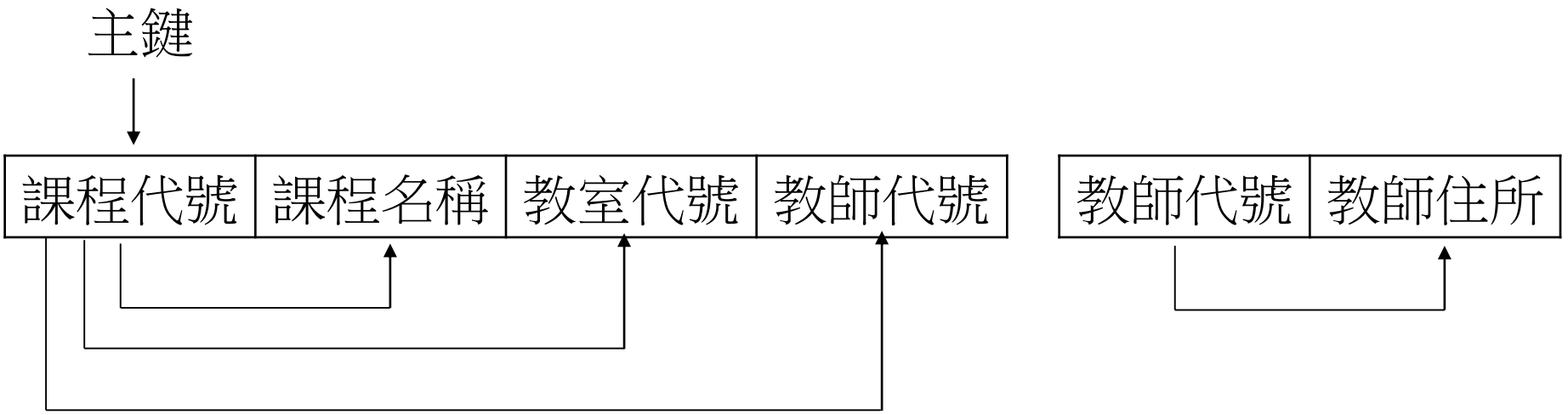
3NF

- 一個關連R為3NF
  1. R為2NF。
  2. 非主鍵屬性不能有遞移相依。
- How to adjust
  1. 保留沒有遞移相依的非主鍵屬性與原主鍵形成關連。
  2. 將形成遞移相依的非主鍵屬性及其決定性屬性形成另一關連，並以後者為主鍵。

例子：



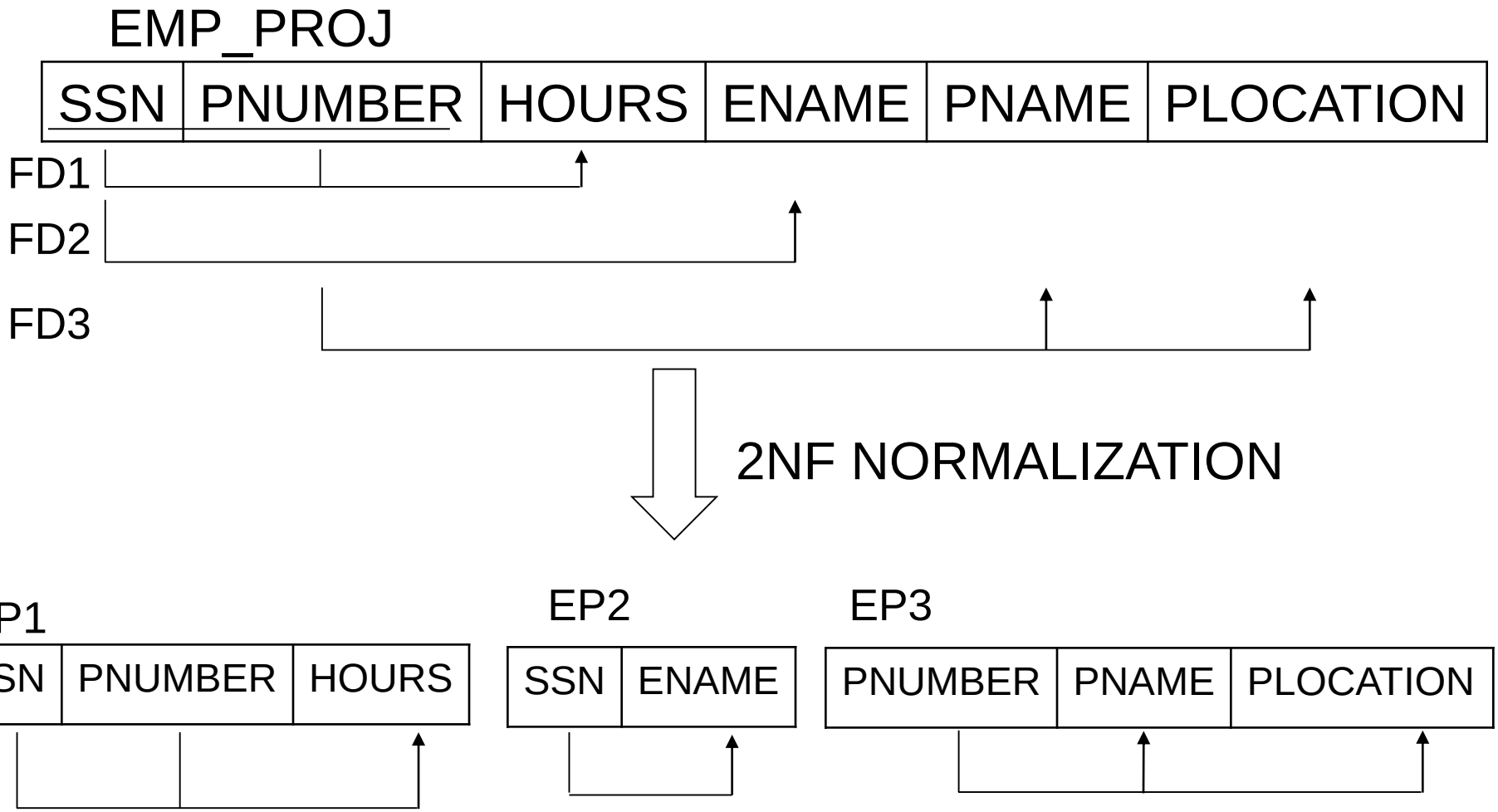
非第三正規式檔案



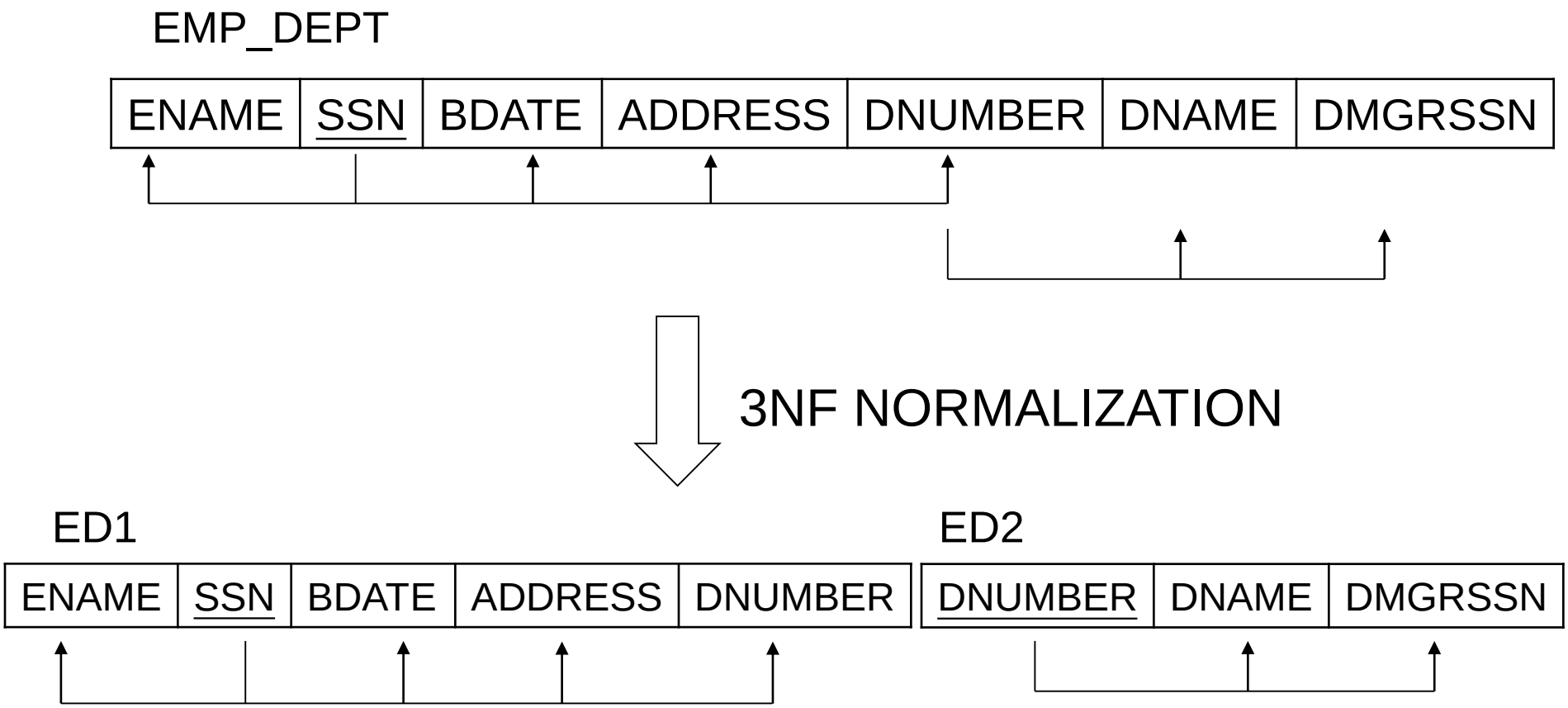
拆成兩個檔案都是第三正規式

正規化過程

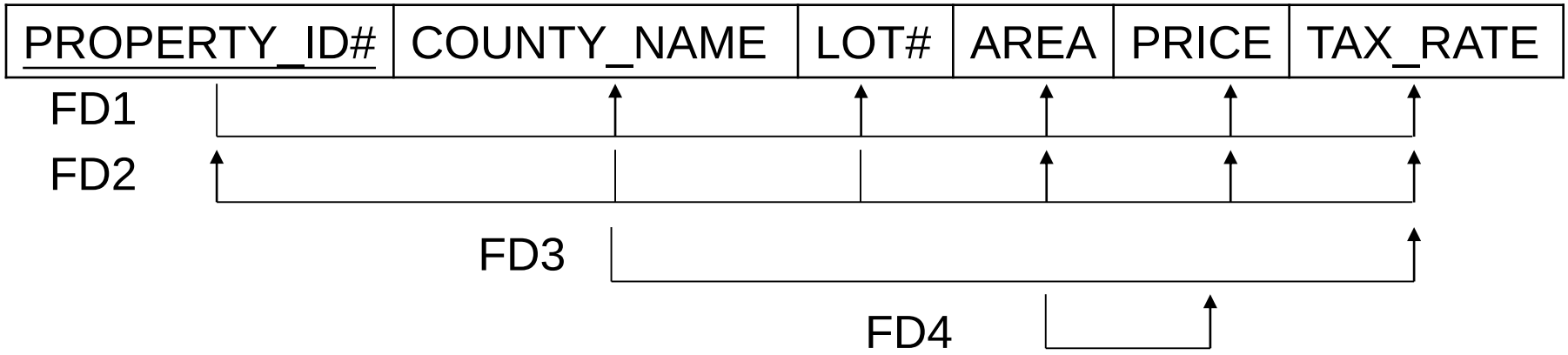
(a)


EMP_PROJ
SSN
PNUMBER
HOURS
ENAME
PNAME
PLOCATION
FD1
FD2
FD3
2NF NORMALIZATION
EP1
SSN
PNUMBER
HOURS
EP2
SSN
ENAME
EP3
PNUMBER
PNAME
PLOCATION

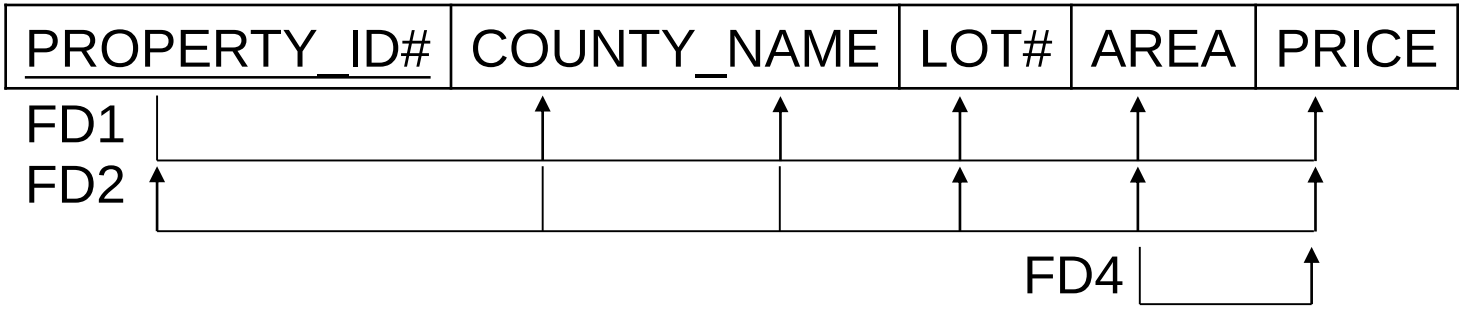
(b)


EMP_DEPT
ENAME
SSN
BDATE
ADDRESS
DNUMBER
DNAME
DMGRSSN
3NF NORMALIZATION
ED1
ENAME
SSN
BDATE
ADDRESS
DNUMBER
ED2
DNUMBER
DNAME
DMGRSSN

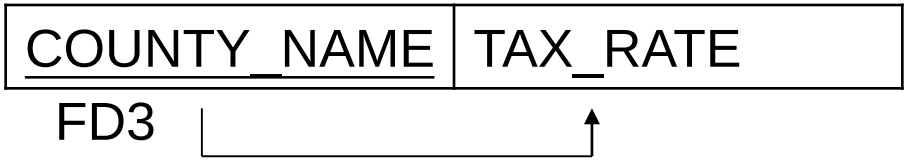
LOTS

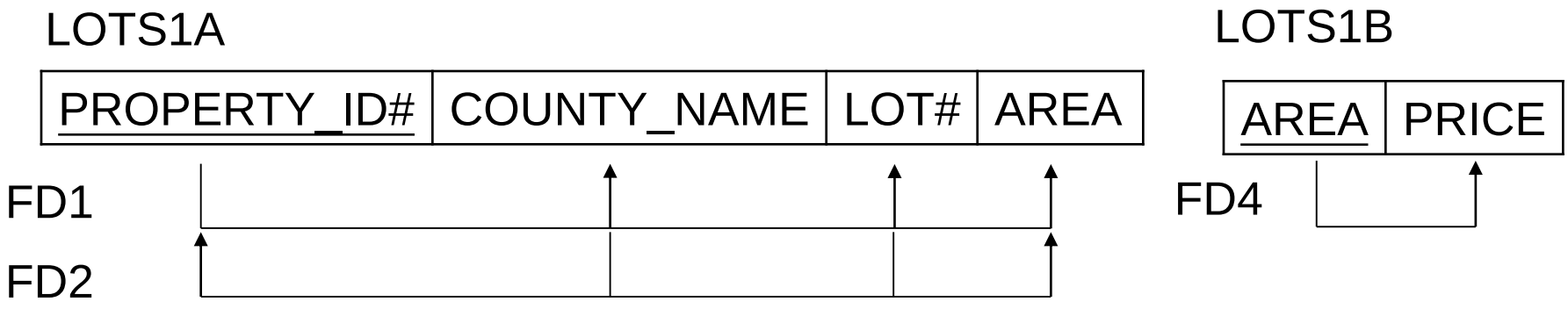


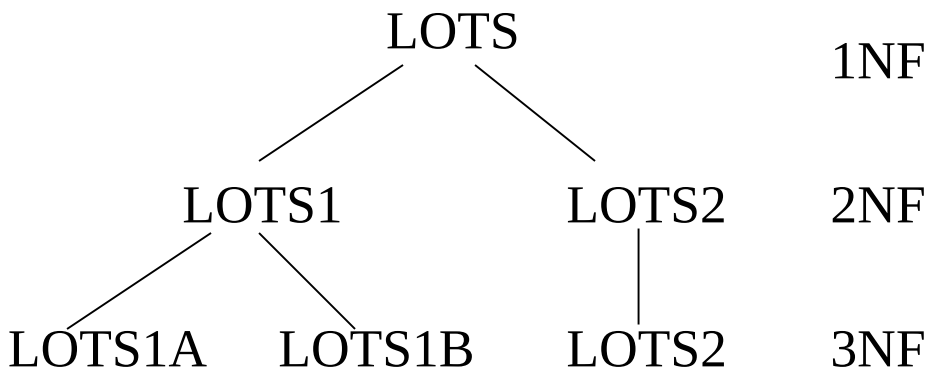
LOTS1



LOTS2




LOTS1A
PROPERTY_ID#
COUNTY_NAME
LOT#
AREA
FD1
FD2
LOTS1B
AREA
PRICE
FD4


LOTS
1NF
LOTS1
LOTS2
2NF
LOTS1A
LOTS1B
LOTS2
3NF

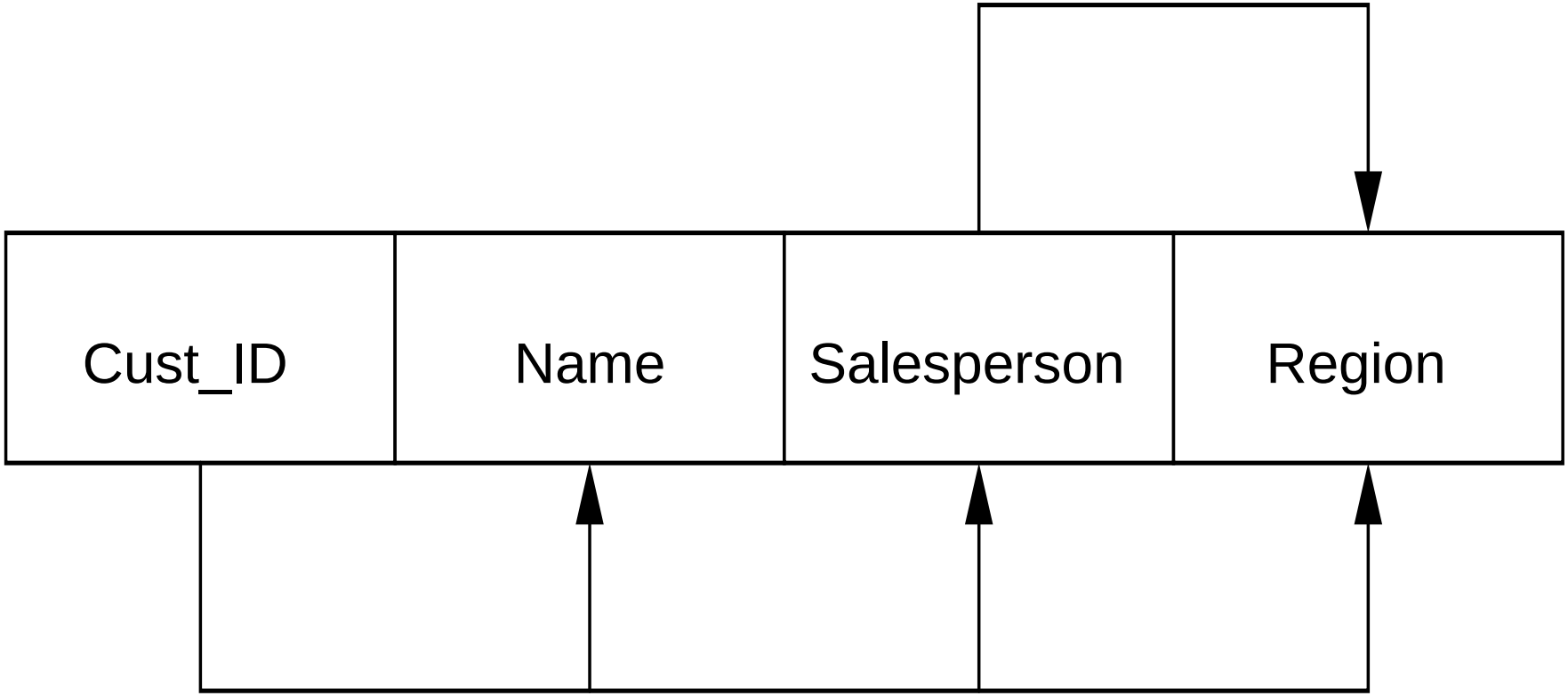
Relation with transitive dependency

a. SALES relation with sample data

SALES

Cust_ID	Name	Salesperson	Region
8023	Anderson	Smith	South
9167	Bancroft	Hicks	West
7924	Hobbs	Smith	South
6837	Tucker	Hernandez	East
8596	Eckersley	Hicks	West
7018	Arnold	Faulb	North

b. Transitive dependency in SALES relation



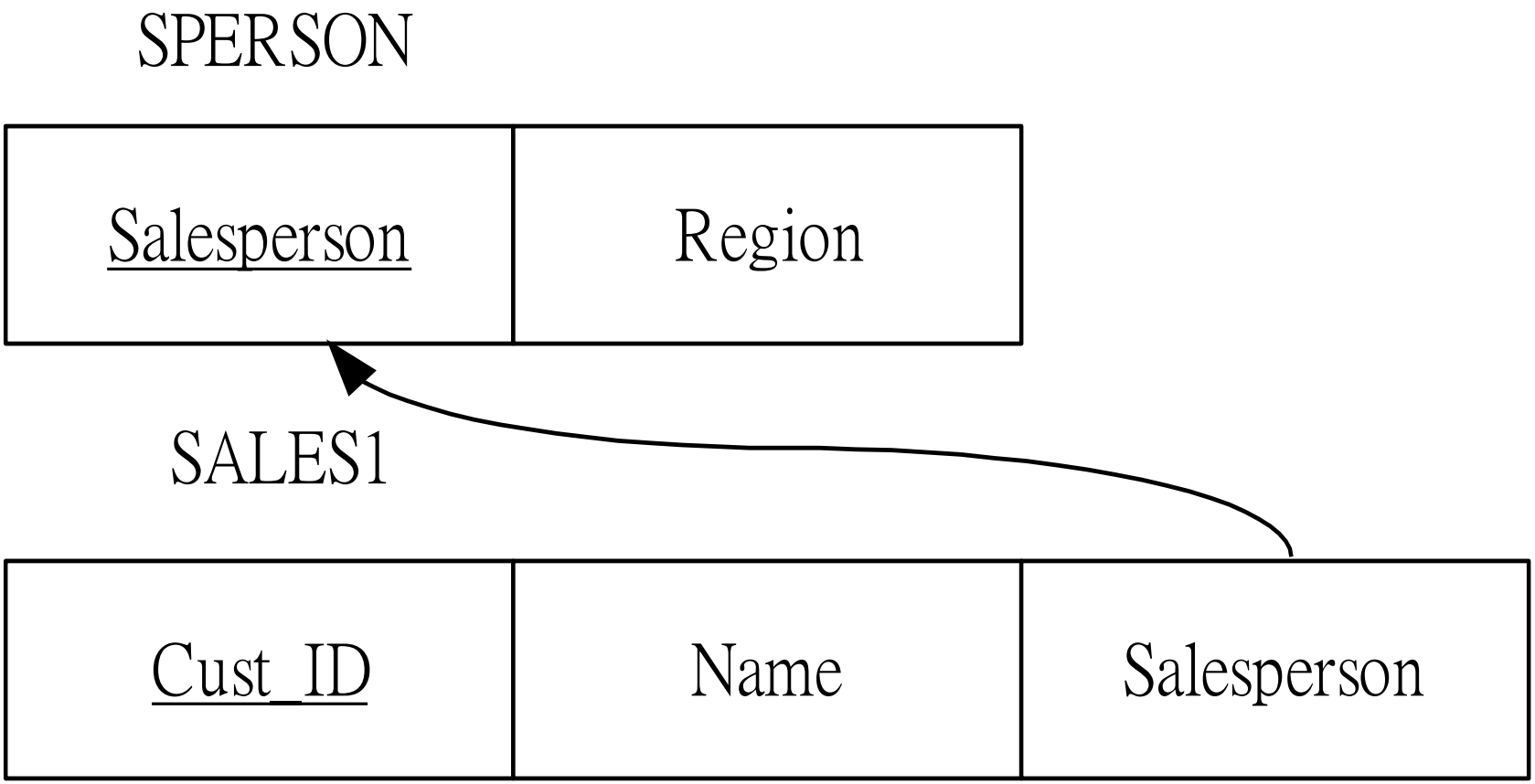
Removing a transitive dependency

a. Decomposing the SALES relation

Cust_ID	Name	Salesperson
8023	Anderson	Smith
9167	Bancroft	Hicks
7924	Hobbs	Smith
6837	Tucker	Hernandez
8596	Eckersley	Hicks
7018	Arnold	Faulb

Salesperson	Region
Smith	South
Hicks	West
Hernandez	East
Faulb	North

b. Relations in 3NF



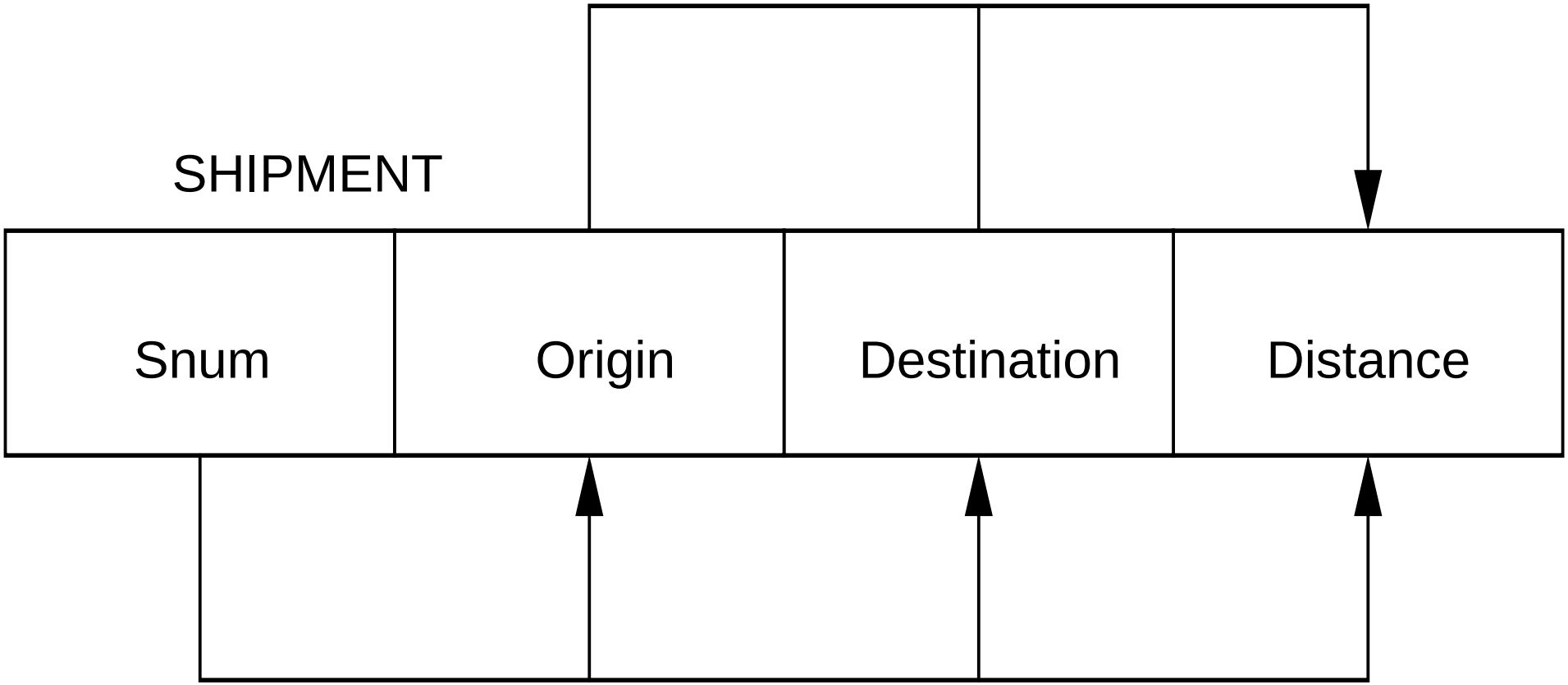
Another example of transitive dependencies

a. SHIPMENT relation with sample data

SHIPMENT

Snum	Origin	Destination	Distance
409	Seattle	Denver	1,537
618	Chicago	Dallas	1,058
723	Boston	Atlanta	1,214
824	Denver	Los Angeles	1,150
629	Minneapolis	St. Louis	587

b. Functional dependencies in SHIPMENT



c. Relations in 3NF

Snum	Origin	Destination
409	Seattle	Denver
618	Chicago	Dallas
723	Boston	Atlanta
824	Denver	Los Angeles
629	Minneapolis	St. Louis

SHIPTO

Origin	Destination	Distance
Seattle	Denver	1,537
Chicago	Dallas	1,058
Boston	Atlanta	1,214
Denver	Los Angeles	1,150
Minneapolis	St. Louis	587

DISTANCES

無損分解(nonloss/lossless decomposition)

- Relation經分解後，將分解後的Relation做Natural Join不會產生多餘的值組，則稱此分解為無損分解。 例子：

C#→ROOM#
C#→CREDIT

Course	C#	ROOM#	CREDIT
	C001	F8011	2
	C002	B5006	3
	C003	F8011	4

Case1

Course1(C#,ROOM#)及 COURSE2(C#,CREDIT)

→COURSE=COURSE1 JOIN COURSE2（保有原 FD）

Case2

COURSE1(C#,ROOM#)及 COURSE2(ROOM#,CREDIT)

→COURSE≠COURSE1 JOIN COURSE2

∴分解有損

- 例子：R(C#,R#,capacity)
  C# →R#
  R#→capacity
  C#→capacity　　∴遞移相移
- 例子：Case 1:
  by C#→capacity
  R1(C#,R#)及R2(C#,capacity)
  無損分解但喪失R#→capacity
  Case 2:
  R1(C#,R#)及R2(R#,capacity)
  無損分解且FD無損
  C#→R#
  R#→capacity　C#→capacity

3NF會引起更新異常的條件：當存在兩個或以上的候選鍵，且

1. 候選鍵為組合的(多個屬性)
2. 候選件的屬性重疊

Relation S(S#, sname, status, city)

Primary key (S#)

Alternative key (sname)

SSP(S#, sname, P#, QTY)

BCNF

一個Relation為BCNF　　所有決定性屬性都是候選鍵。

a. 3NF正規化在此分解中會遺失fd2的相依性

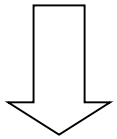
LOTS1

PROPERTY_ID#	COUNTY_NAME	LOT#	AREA

FD1

FD2

FD5



3NF NORMALIZATION

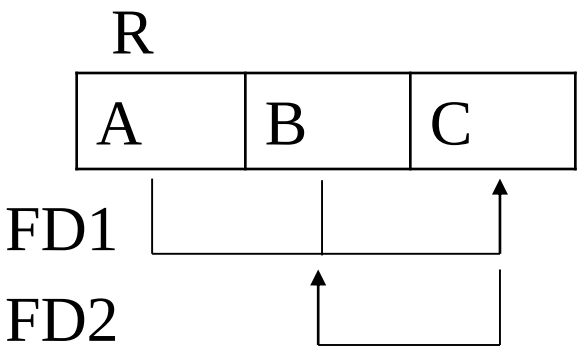
LOTS1AX

PROPERTY_ID#	AREA	LOTS

LOTS1AY

AREA	COUNTY_NAME

b. 一個屬於3NF但非BCNF的關聯R



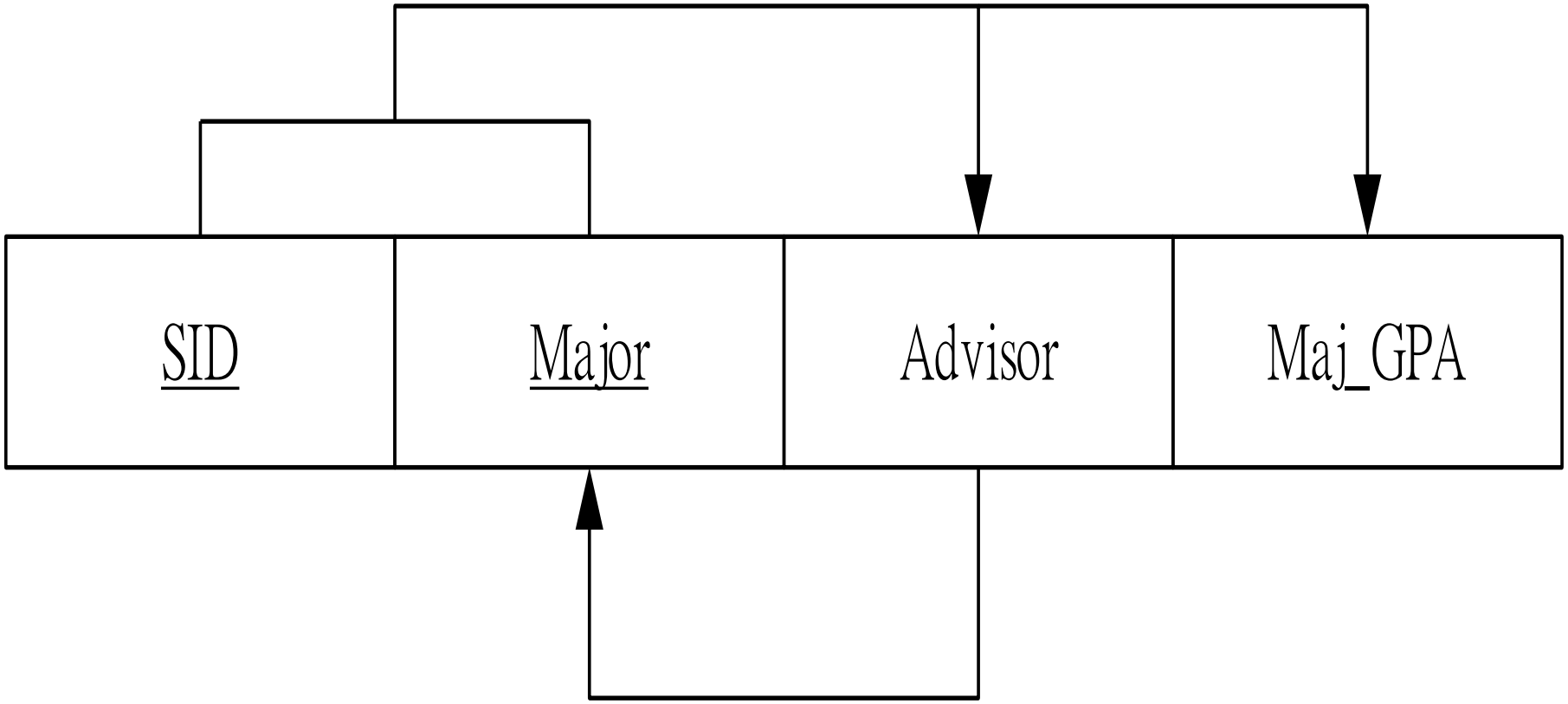
Relation in 3NF, but not BCNF

a. Relation with sample data

STUDENT_ADVISOR

SID	Major	Advisor	Maj_GPA
123	Physics	Hawking	4.0
123	Music	Mahler	3.3
456	Literature	Michener	3.2
789	Music	Bach	3.7
678	Physics	Hawking	3.5

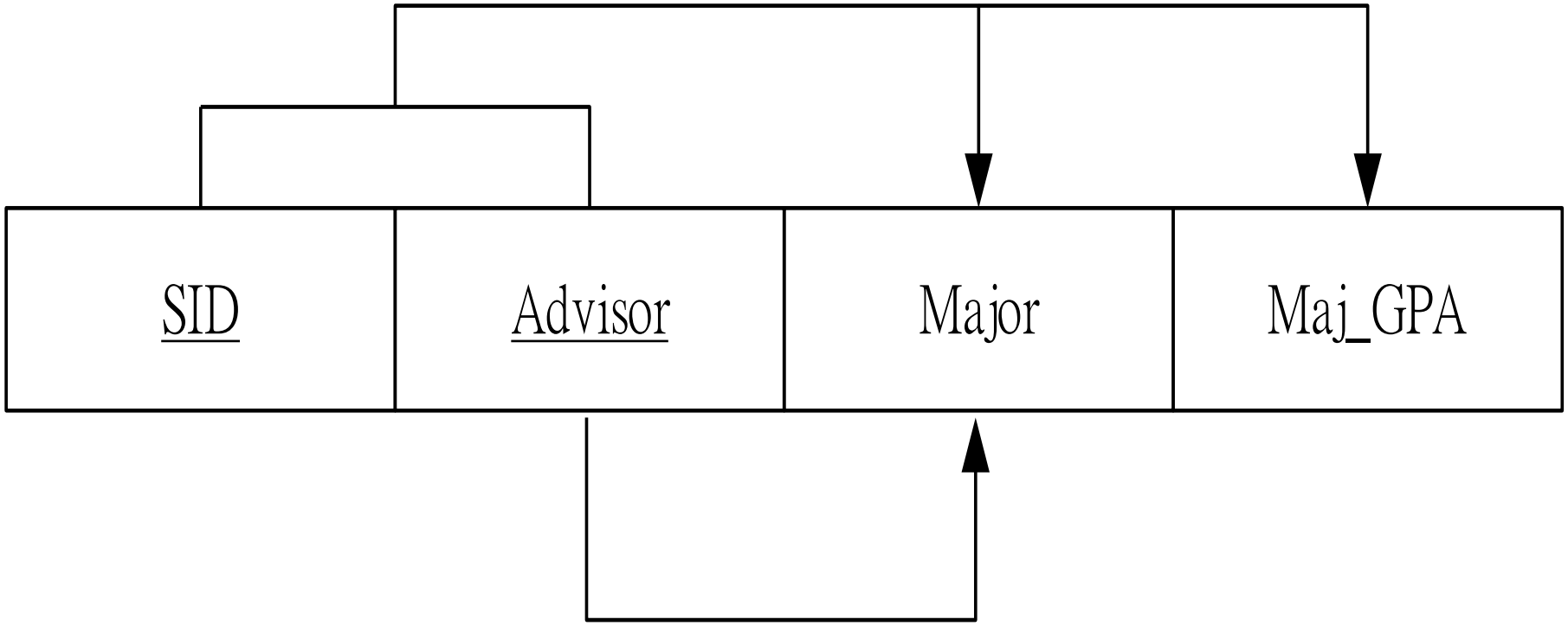
b. Functional dependencies in STUDENT_ADVISOR



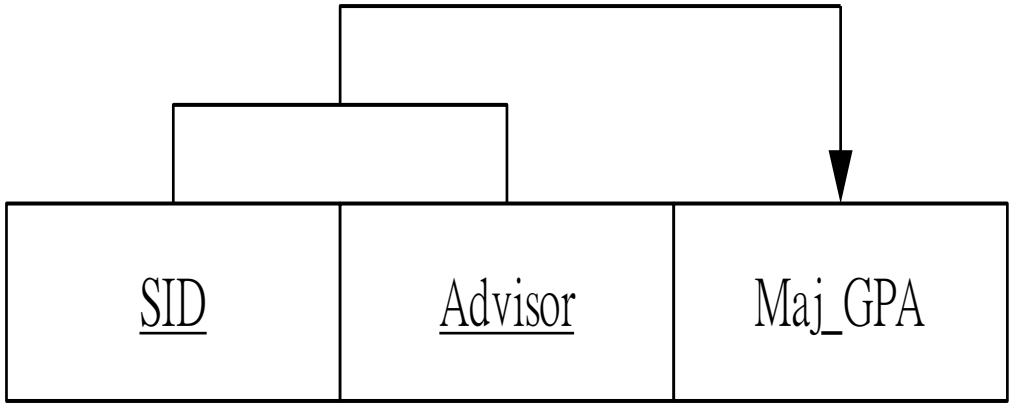
1. Suppose that in Physics the advisor Hawking is replaced by Einstein. This change must be made in two (or more) rows in the table (update anomaly).
2. Suppose we want to insert a row with the information that Babbage advises in Computer Science. This, of course, cannot be done until at least one student majoring in Computer Science is assigned Babbage as an advisors (insertion anomaly).
3. Finally, if student number 789 withdraws from school, we lose the information that Bach advises in Music (deletion anomaly).

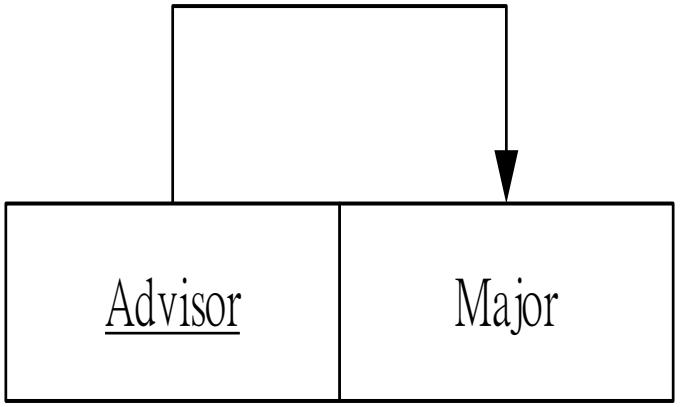
Converting a relation to BCNF relations

a. Revised STUDENT_ADVISOR relation(2NF)



b. Two relations in BCNF





c. Relations with sample data

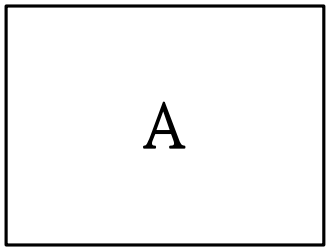
SID	Advisor	Maj_GPA
123	Hawking	4.0
123	Mahler	3.3
456	Michener	3.2
789	Bach	3.7
678	Hawking	3.5

STUDENT

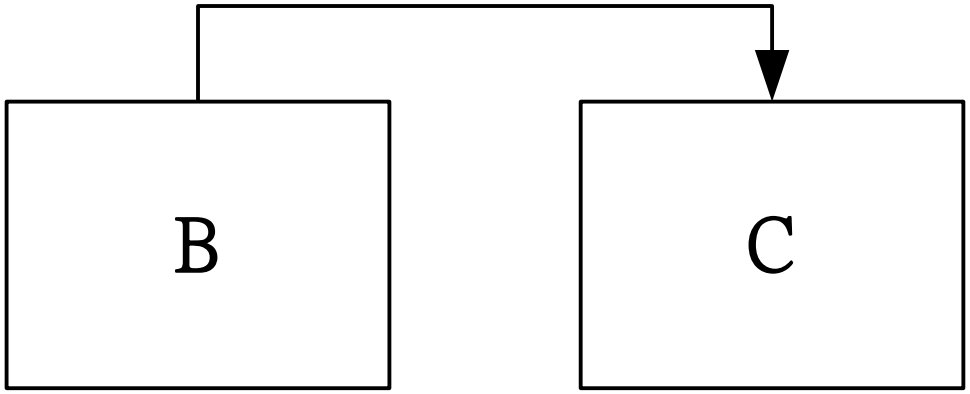
Advisor	Major
Hawking	Physics
Mahler	Music
Michener	Literature
Bach	Music

ADVISOR

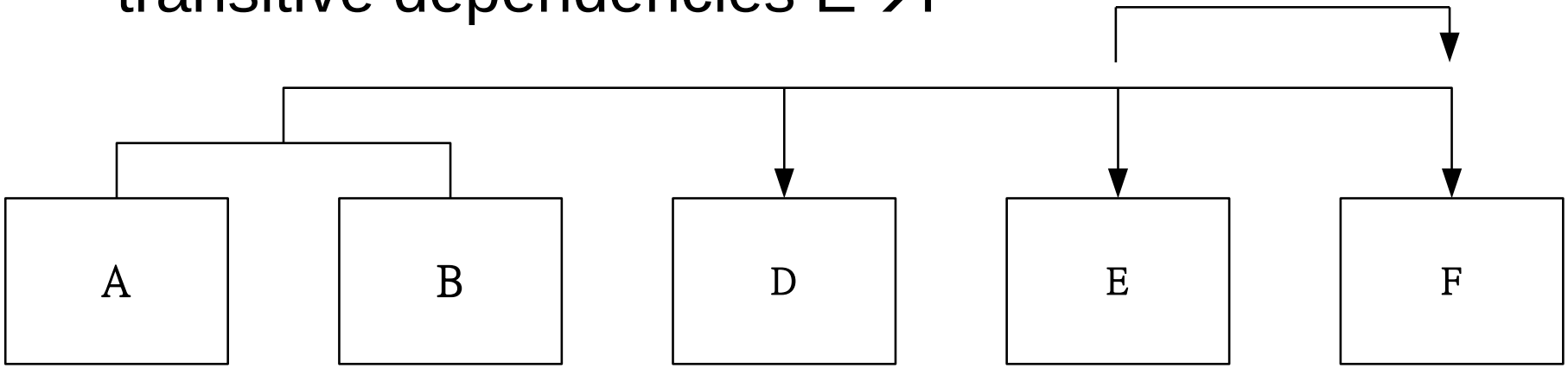
- NO attributes are dependent on A，and A nerver becomes the primary key for a table.



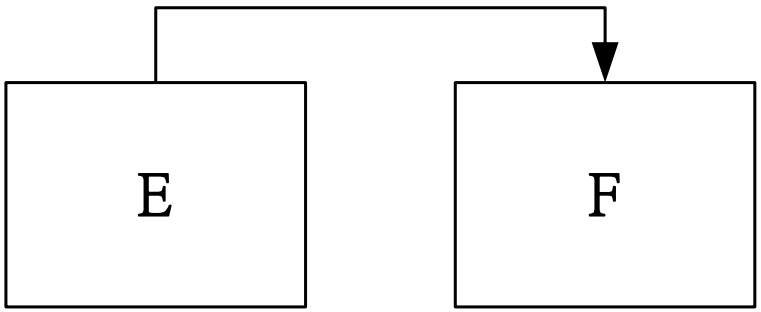
- This table is in 3NF because it is in 2NF (no partial dependencies) and it contains no transitive dependencies.



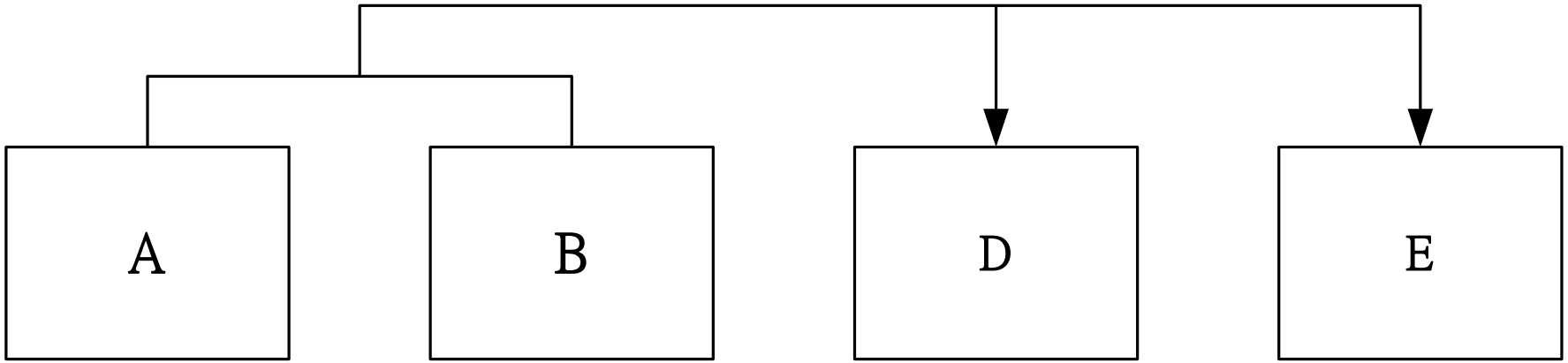
- This table is in 2NF because it contains a transitive dependencies E→F



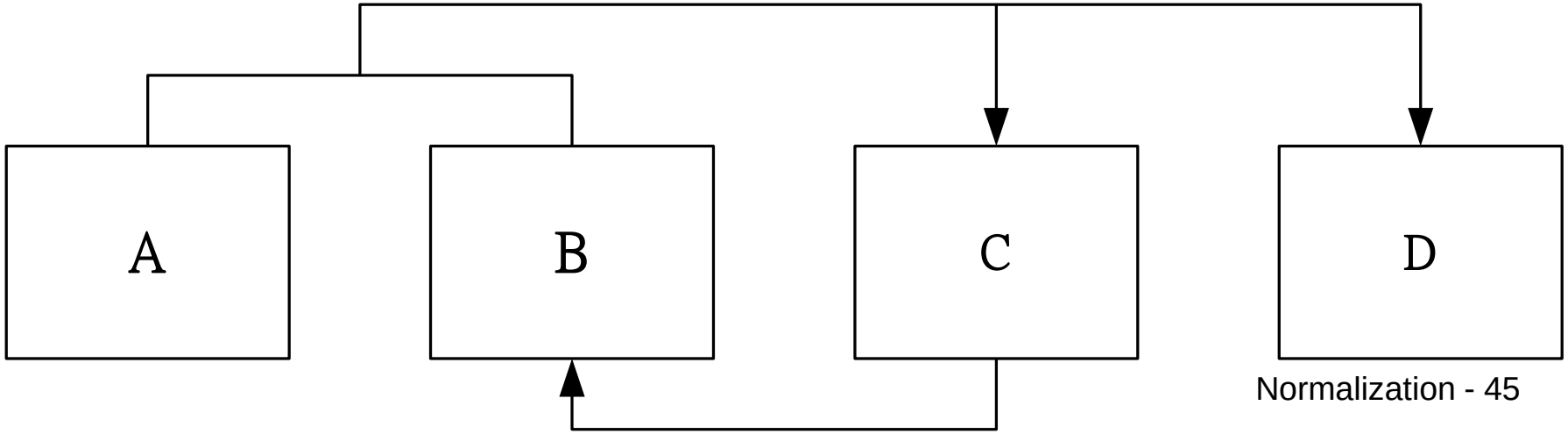
- Break off the E, F segment to form a new table, but keep E in the original table in order to create a foreign key link to the new table.



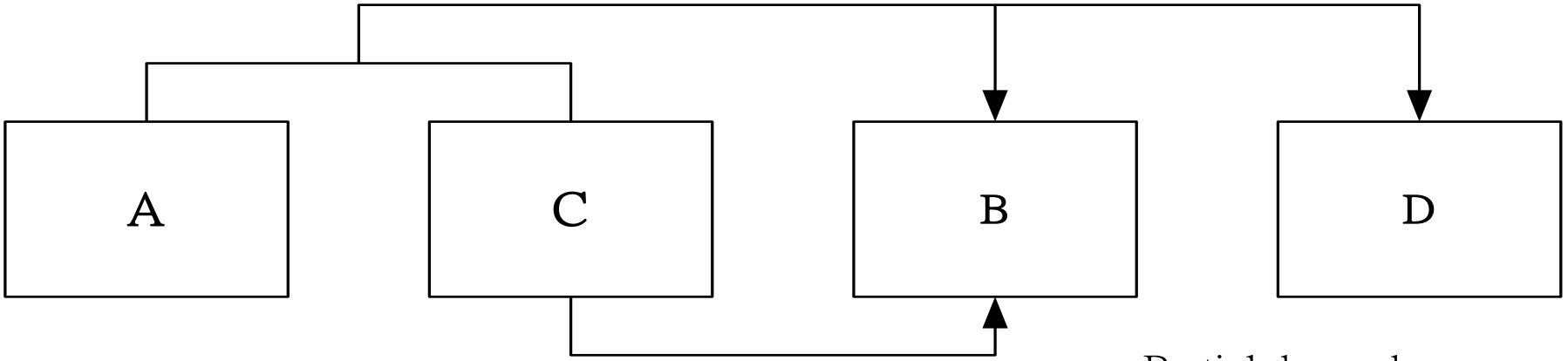
- Both tables are in 3NF；Neither partial nor transitive dependencies exist.



- 3NF, but not BCNF



2NF



Partial dependency

- 3NF and BCNF