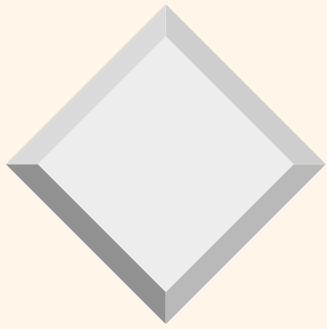




關聯表的正規化 Normalization

Chapter 5

補充

關聯網要(結構)設計的一般原則一

- 關聯網要(結構)應該很清楚呈現其所代表的意義(語意) *不要把不同東西放在一起, ER要分開*
- 勿將數個資料實體或關係的屬性放在同一關聯資料表中
 - 因為由單一實體建構的關聯資料表, 可以很清楚了解其所代表的意義(語意)
 - 反之, 若將數個資料實體或關係的屬性放在同一關聯資料表中, 則意義將混淆 *造成資料重複*

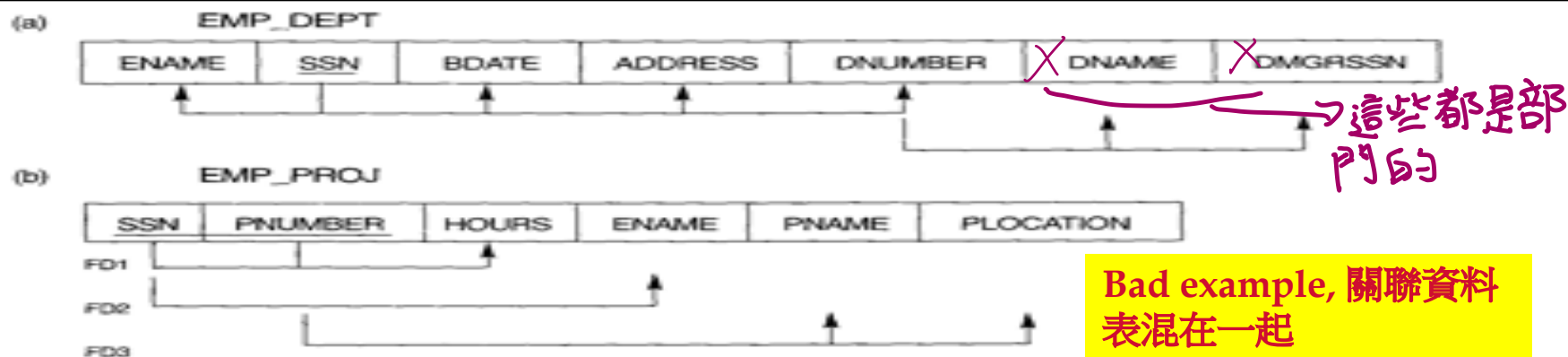


Figure 14.3 Two relation schemas and their functional dependencies. Both suffer from update anomalies. (a) The EMP_DEPT relation schema. (b) The EMP_PROJ relation schema.

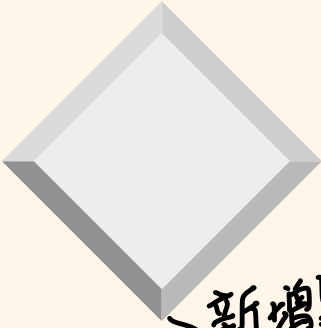
EMP_DEPT

ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

EMP_PROJ

SSN	PNUMBER	HOURS	ENAME	PNAME	PLOCATION
123456789	1	32.5	Smith, John B.	ProductX	Bellaire
123456789	2	7.5	Smith, John B.	ProductY	Sugarland
666884444	3	40.0	Narayan, Ramesh K.	ProductZ	Houston
453453453	1	20.0	English, Joyce A.	ProductX	Bellaire
453453453	2	20.0	English, Joyce A.	ProductY	Sugarland
333445555	2	10.0	Wong, Franklin T.	ProductY	Sugarland
333445555	3	10.0	Wong, Franklin T.	ProductZ	Houston
333445555	10	10.0	Wong, Franklin T.	Computerization	Stafford
333445555	20	10.0	Wong, Franklin T.	Reorganization	Houston
999887777	30	30.0	Zelaya, Alicia J.	Newbenefits	Stafford
999887777	10	10.0	Zelaya, Alicia J.	Computerization	Stafford
987987987	10	35.0	Jabbar, Ahmad V.	Computerization	Stafford
987987987	30	5.0	Jabbar, Ahmad V.	Newbenefits	Stafford
987654321	30	20.0	Wallace, Jennifer S.	Newbenefits	Stafford
987654321	20	15.0	Wallace, Jennifer S.	Reorganization	Houston
888665555	20	null	Borg, James E.	Reorganization	Houston

Figure 14.4 Example relations for the schemas in Figure 14.3 that result from applying NATURAL JOIN to the relations in Figure 14.2. These may be stored as base relations for performance reasons.



原則二、三

新增異常：有必要的資料無法取得致無法新增

□ **Guideline 2**：設計關聯網要(資料表)時，避免發生新增、刪除、更新等異常狀況(查詢不會有異常)
(*anomalies*)
刪除列
其它東西
↳ 有重複資料，致無法完全更新

— 若所設計的關聯網要(資料表)，無法保證不會發生異常狀況，則必須寫程式使系統避免發生異常狀況

□ **Guideline 3**：設計關聯資料表的屬性欄位儘可能不會常出現空值(null)

Guideline 3 – 範例

- 若僅有10%的員工有獨立的研究室
 - 若在EMPLOYEE 關聯資料表中多建一屬性欄位 OFFICE_NUMBER，則此欄位90%為空值
- 建議
 - 增建一關聯資料表EMP_OFFICES (ESSN, OFFICE_NUMBER)
 - 此資料表內僅存放有獨立研究室的員工資料，即可避免出現空值欄位

不好

類似多值做法

原: Employee(ssn, Name, Addr., phone, office No.)

但 90% 情况为 null

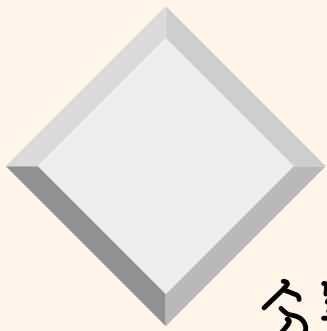
↓

Employee(ssn, Name, Addr., phone)

Emp-offC (ssn, office-no)

FK

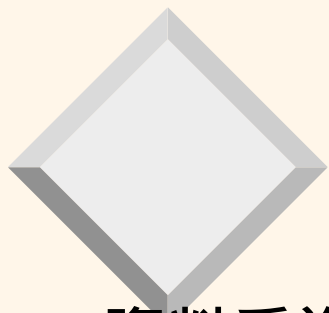
→ 主索引 | 設 Foreign Key



原則四

分割成2資料表示，將來若要合併則資料量要一樣

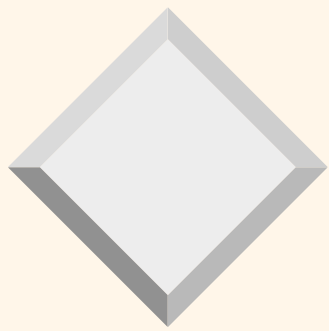
- 設計關聯資料表時，盡量使關聯資料表間以主索引、外來鍵(*primary keys, foreign keys*)作為相等關係的合併(*JOINED*)條件
 - 避免合併時產生多餘的資料(假值組)
 - 兩關聯資料表進行合併運算時，結合的屬性欄位若不是主索引、外來鍵的關係，常會產生多餘的資料(假值組)



資料重複引發的問題

The Evils of Redundancy

- 資料重複(**Redundancy**)是關聯綱要(結構)許多問題的根源
 - 浪費多餘的儲存空間、發生新增、刪除、更新異常
- 整合性限制條件，尤其是功能性相依 (**functional dependencies**)，可以用來指出資料綱要(結構)的問題，並提出改善的意見 FD
 - 主要採用的方法是分割關聯資料表(**decomposition**)，例如：將關聯資料表 ABCD 分割成 AB、BCD, 或 ACD、ABD
- 進行分割要非常小心：
 - 是否因違反某正規化規則，所以要分割？
 - 分割後是否會發生另外的問題？



範例:實體上的限制條件

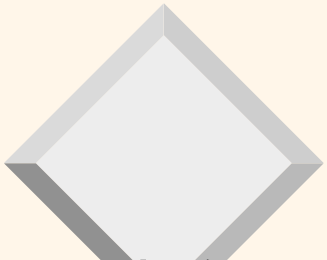
- 假設計時制員工資料實體 Hourly_Emps:
 - Hourly_Emps (ssn, name, lot, rating, hrly_wages, hrs_worked)
 - 一般以SNLRWH簡化表示
 - 欄位的集合 {S,N,L,R,W,H}.
 - 有時用 Hourly_Emps 表示 SNLRWH
 - 在Hourly_Emps上的一些功能性相依(FDs):
 - 若ssn 是索引(key)則: $S \xrightarrow{\text{implies}} \text{SNLRWH}$ 我如果知道某一欄 就可知所有欄
 - rating 決定 hrly_wages: $R \xrightarrow{\text{implies}} W$ 若我知S,就可推知 SNLRWH
- (等級可知道時薪)
- ①找主索引
②找有沒有什麼相依的

範例 (Contd.)

S	N	L	R	W	H
123-22-3666	Attishoo	48	8	10	40
231-31-5368	Smiley	22	8	10	30
131-24-3650	Smethurst	35	5	7	30
434-26-3751	Guldu	35	5	7	32
612-67-4134	Madayan	35	8	10	40

□ 由於 R W產生的問題：

- 重複儲存空間(Redundant storage) : $\langle \text{rating}, \text{wage} \rangle = \langle 8, 10 \rangle$ 出現3次
- 更新異常(Update anomaly)(資料重複造成更新不完全) :
若第1筆資料的hourly_wages值更新，但第2筆未更新，則
.....
- 新增異常(Insertion anomaly)(必要資料無法取得以致無法新增) : 除非知道某員工之rating值，則無法新增此員工資料
- 刪除異常(Deletion anomaly)(刪除時，將重要資料一併刪除) : 若刪除第3, 4筆資料，剛好刪除所有 rating 5的資料，同時喪失rating 5的wage資料



正規化的定義 Normalization

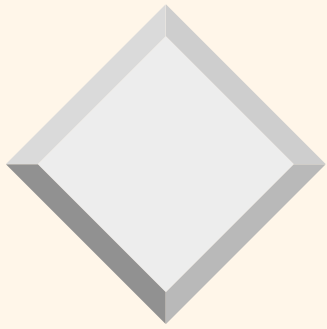
- 正規化可以分析所設計的關聯式資料庫是否有問題，不但提供**找出問題**的方法，也提供**改正**的方法
- 整個正規化流程的目的是建構**不會發生更新異常、和原來關聯資料表效果相同，至少需3階正規化**
- 正規化正式的分析工作，主要是根據索引(**keys**)及功能性相依(**functional dependencies**)的特性
- 正規化是一連串關聯資料表的檢查工作(評估步驟)，將不符合某階層正規化要求的資料表加以分割成數個較小的關聯資料表
非正規化沒有合併，合併：反正規化 *A implies BCD*
- 正規化過程必須保有：
A is the determinant BCD相依於A 決定者
 - 屬性欄位不可減少
 - FD不可遺漏
 - 分割後未來(查詢)若有需要合併，不可增加多餘的資料

功能性相依

Functional Dependencies (FDs)

- 存在關聯資料表R中的功能性相依 $X \rightarrow Y$ ，表示對R中每一筆資料r，均存在：
 - 即在關聯資料表R中，若X已知，則Y值可得，也就是Y值由X決定
- 若A值可以決定唯一的B值，則稱B功能性相依於A (functionally dependent)，記為 $A \rightarrow B$
- 若K 是 R的候選鍵，則存在 $K \rightarrow R$

⊙ R:所有欄位



功能性相依(cont.)

- 關聯資料表中的關聯性相依(FD)，是由應用系統本身的意義找出的
- 定義：若屬性欄位 A 為關聯資料表R的主索引(primary key) 則：
 1. 關聯資料表R中所有欄位均功能性相依於A
 2. A屬性欄位的部份集合不具備性質1(即其他欄位未必相依於部分索引欄位) (主索引是超索引的最小欄位集合)

功能相依的封閉性

□ F包含所有功能相依，及所有由F推論得到的功能性相依之集合，稱為功能性相依的封閉性，記為 F^+

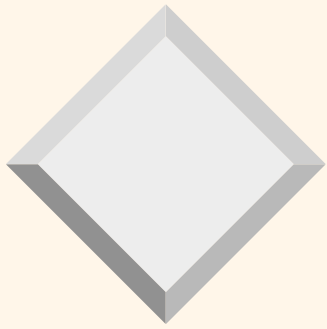
□ $F = \{ \text{Ssn} \rightarrow \{ \text{Ename}, \text{Bdate}, \text{Address}, \text{Dnumber} \}, \text{Dnumber} \rightarrow \{ \text{Dname}, \text{Dmgr_ssn} \} \}$

可推導出：

$\text{Ssn} \rightarrow \{ \text{Dname}, \text{Dmgr_ssn} \}$

$\text{Ssn} \rightarrow \text{Ssn}$

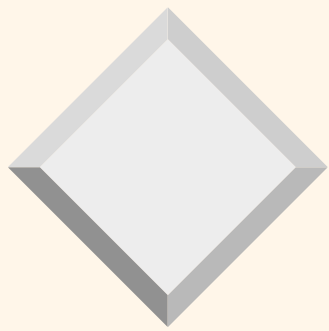
$\text{Dnumber} \rightarrow \text{Dname}$



功能性相依的推理規則

Inference Rules for FDs

- IR1(反身性規則reflexive rule): If $X \supseteq Y$, then $X \rightarrow Y$ 包含 $X \rightarrow X$ 也是成立的
- IR2(擴充性規則augmentation rule): $\{X \rightarrow Y\} \models XZ \rightarrow YZ$ 則 因Y可夾入
- IR3(遞移性規則transitive rule): $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$ 則
- IR4(分割或投射規則decomposition or projective rule):
 $\{X \rightarrow YZ\} \models X \rightarrow Y$ 則 $\models X \rightarrow Z$ 也成立
- IR5(聯集或加法規則union or additive rule):
 $\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow YZ$
- IR6(虛擬遞移規則pseudotransitive rule):
 $\{X \rightarrow Y, WY \rightarrow Z\} \models WX \rightarrow Z$



複習專有名詞

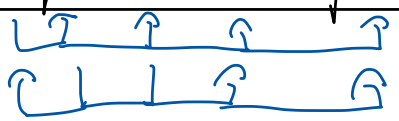
Review some terms

- 若一關聯資料表有一個以上的索引(key)，每一索引均稱為候選鍵(candidate key)，再從其中選出一個當作主索引(primary key)，其他未選上的候選鍵稱為次索引(secondary keys)
- 資料表中的欄位若是某候選鍵的一部分，則稱此欄位為主要欄位(prime attribute)，否則稱為非主要欄位(nonprime)
- 若某欄位不是主索引的一部分稱為非索引(鍵)欄位(non-key attribute)

若 a-b-c 為候選 key
a 為主索引

a	b	c	d	e

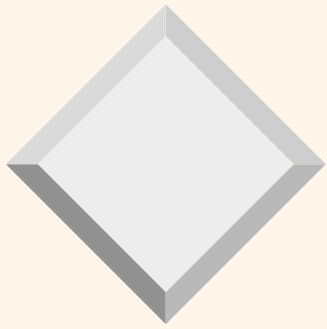
Prime Attr.: a, b, c
non Prime: d, e, f



完全相依

Primary key: a

non key: b, c, d, e



正規化

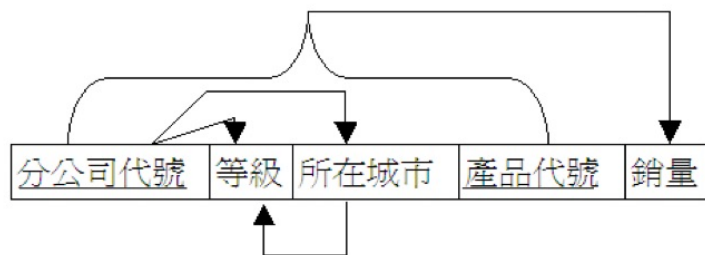
STEPS:

1. 先將所有資料列出, 找出相依 (參下頁)
2. 找出重複的欄位 (1NF)
3. 找出 key 與 non-key 的各項關係, 分割成不同資料表 (2NF)
4. 找出 non-key 之間的關係 (3NF)

正規化過程



- 假設有一個**非正規化表格**與它的屬性間之相依關係如下：

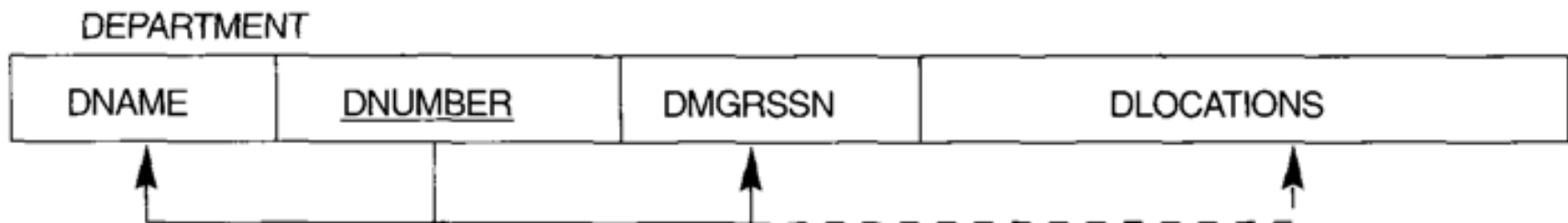


分公司代號	等級	所在城市	產品代號	銷量
S1	20	台北	P1	300
			P2	200
			P3	400
			P4	200
			P5	100
			P6	100
S2	10	高雄	P1	300
			P2	400
S3	10	高雄	P2	200
S4	20	台北	P2	200
			P4	300
			P5	400

一階正規化(1NF)

- 一階正規化(First Normal Form, 1NF) 的定義：一關聯資料表若為1NF，則其所有屬性欄位的定義域必須僅包含不可再分割的單值屬性(atomic、simple、indivisible)
 - 1NF 不允許某欄位有一組值(a set of values)
 - 也不允許一筆資料值(a tuple of values)(幾個欄位組成的直)
 - 即 1NF不允許「資料表中還有資料表(relations within relations)」或「欄位的值為一個關聯資料表(relations as attributes of tuples)」
 - 簡言之，1NF 就是不允許重複群組(repeating groups) 或多重值的欄位(multi-valued attribute)
- e.g. SOFTWARE(PACK-ID,PACK_NAME,TAG-NO,INS-DATE,SOFT-COST)
SOFTTAG(PACK-ID,TAG-NO,INS-DATE,SOFT-COST)
SOFTWARE(PACK-ID,PACK_NAME)

(a)



↑
多重值Multi-value

(b)

DEPARTMENT

DNAME	<u>DNUMBER</u>	DMGRSSN	DLOCATIONS
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

(c)

DEPARTMENT

DNAME	<u>DNUMBER</u>	DMGRSSN	<u>DLOCATION</u>
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

Figure 14.8 Normalization into 1NF. (a) Relation schema that is not in 1NF. (b) Example relation instance. (c) 1NF relation with redundancy.

(a)

EMP_PROJ

SSN	ENAME	PROJS	
		PNUMBER	HOURS

(b)

EMP_PROJ

SSN	ENAME	PNUMBER	HOURS
123456789	Smith, John B.	1	32.5
		2	7.5
666884444	Narayan, Ramesh K.	3	40.0
453453453	English, Joyce A.	1	20.0
		2	20.0
333445555	Wong, Franklin T.	2	10.0
		3	10.0
		10	10.0
		20	10.0
999887777	Zelaya, Alicia J.	30	30.0
		10	10.0
987987987	Jabbar, Ahmad V.	10	35.0
		30	5.0
987654321	Wallace, Jennifer S.	30	20.0
		20	15.0
888665555	Borg, James E.	20	null

1NF不允許複合式
多重值欄位(又稱
重複群組欄位
repeating group)

(c) 分割

EMP_PROJ1

SSN	ENAME
-----	-------

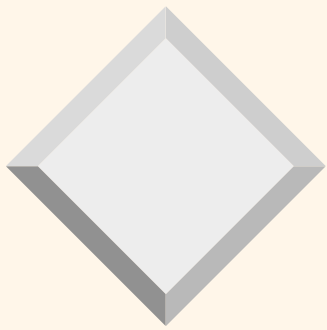
EMP_PROJ2

SSN	PNUMBER	HOURS
-----	---------	-------

拉出來與多重值
一樣手法

Figure 14.9

Normalizing nested relations into 1NF. (a) Schema of the EMP_PROJ relation with a “nested relation” PROJS. (b) Example extension of the EMP_PROJ relation showing nested relations within each tuple. (c) Decomposing EMP_PROJ into 1NF relations EMP_PROJ1 and EMP_PROJ2 by propagating the primary key.



將非1NF轉換成 1NF

□ 在多重值的案例：

- 將破壞1NF的屬性欄位 DLOCATIONS 移出，放入新建的關聯資料表 DEPT_LOCATIONS中，同時包含DEPARTMENT的主索引 DNUMBER(外來鍵，foreign key)，關聯資料表DEPT_LOCATIONS的主索引為 {DNUMBER, DLOCATION}.

Best way

- 將此多重值欄位與原資料表之主索引結合，形成新的主索引

Not so good

- 若此多重值欄位出現最多的次數已知，設為3，則建立相關的3個屬性欄位 DLOCATION1, DLOCATION2, DLOCATION3

□ 但若多數部門沒有那麼多地點，則會出現許多空值(null)

□ 在重複群組(repeating group)的案例：擴充主索引，以將非1NF轉換成 1NF (primary key+repeating group key)

- SOFTWARE(PACK-ID,PACK_NAME,TAG-NO,INS-DATE,SOFT-COST)

→ SOFTTAG(PACK-ID,TAG-NO,INS-DATE,SOFT-COST)

SOFTWARE(PACK-ID,PACK_NAME)

課本例子

Students

<u>sid</u>	name	c_no	title	eid	instructor	office	room
S001	陳會安	{ CS101, CS203, CS222, CS213, }	{ 計算機概論, 程式語言, 資料庫管理系統, 物件導向程式設計, }	{ E001, E003, E002, E003, }	{ 陳慶新, 李鴻章, 楊金欉, 李鴻章 }	{ CS-102, M-100, CIS-101, M-100 }	{ 180-M, 221-S, 100-M, 500-K }
S002	江小魚	{ CS222, CS203 }	{ 資料庫管理系統, 程式語言 }	{ E002, E003 }	{ 楊金欉, 李鴻章 }	{ CIS-101, M-100 }	{ 100-M, 221-S }
S003	張三丰	{ CS121, CS213 }	{ 離散數學, 物件導向程式設計 }	{ E002, E001 }	{ 楊金欉, 陳慶新 }	{ CIS-101, CS-102 }	{ 221-S, 622-G }
S004	李四方	CS222	資料庫管理系統	E002	楊金欉	CIS-101	100-M

解法 (最佳解 , 以重複群組的方法解)

Students

<u>sid</u>	name
S001	陳會安
S002	江小魚
S003	張三丰
S004	李四方

Classes

<u>sid</u>	<u>c_no</u>	title	<u>eid</u>	instructor	office	room
S001	CS101	計算機概論	E001	陳慶新	CS-102	180-M
S001	CS203	程式語言	E003	李鴻章	M-100	221-S
S001	CS222	資料庫管理系統	E002	楊金欉	CIS-101	100-M
S001	CS213	物件導向程式設計	E003	李鴻章	M-100	500-K
S002	CS222	資料庫管理系統	E002	楊金欉	CIS-101	100-M
S003	CS203	程式語言	E003	李鴻章	M-100	221-S
S003	CS121	離散數學	E002	楊金欉	CIS-101	221-S
S003	CS213	物件導向程式設計	E001	陳慶新	CS-102	622-G
S004	CS222	資料庫管理系統	E002	楊金欉	CIS-101	100-M

解法 (不佳, ~~分割成值組~~, 形成多筆資料)

Students

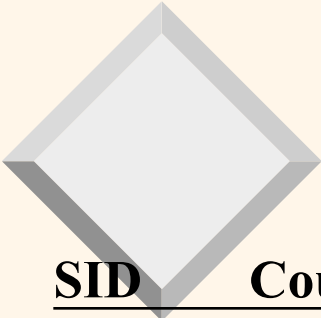
<u>sid</u>	name	<u>c_no</u>	title	<u>eid</u>	instructor	office	room
S001	陳會安	CS101	計算機概論	E001	陳慶新	CS-102	180-M
S001	陳會安	CS203	程式語言	E003	李鴻章	M-100	221-S
S001	陳會安	CS222	資料庫管理系統	E002	楊金欉	CIS-101	100-M
S001	陳會安	CS213	物件導向程式設計	E003	李鴻章	M-100	500-K
S002	江小魚	CS222	資料庫管理系統	E002	楊金欉	CIS-101	100-M
S002	江小魚	CS203	程式語言	E003	李鴻章	M-100	221-S
S003	張三丰	CS121	離散數學	E002	楊金欉	CIS-101	221-S
S003	張三丰	CS213	物件導向程式設計	E001	陳慶新	CS-102	622-G
S004	李四方	CS222	資料庫管理系統	E002	楊金欉	CIS-101	100-M

解法 (不佳，~~分割成不同屬性~~，建立多個欄位)

Students

<u>sid</u>	name	c_no1	title1	c_no2	title2
S001	陳會安	CS101	計算機概論	CS203	程式語言
S002	江小魚	CS222	資料庫管理系統	CS203	程式語言
S003	張三丰	CS121	離散數學	CS213	物件導向程式設計
S004	李四方	CS222	資料庫管理系統	NULL	NULL

1NF (Contd.)



<u>SID</u>	<u>Course No</u>	Course	Teacher	Room	Grade
38214	IS350	DATA BASE	CODD	F101	A
38214	IS465	SYS ANAL	KEMP	F103	C
69173	IS465	SYS ANAL	KEMP	F103	A
69173	PM300	PROD MGT	LEWIS	H101	B
69173	OM440	OP RES	KEMP	F103	C

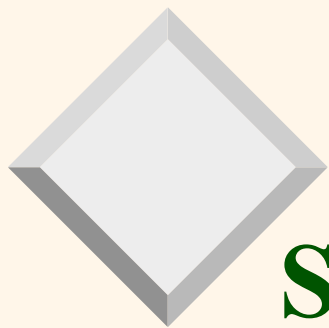
* 1NF的更新異常(Update anomalies)

1. 更新：若要更新 Course IS465 的值，必須進行數次的更新
2. 資料不一致：無法控制 IS465 有不同的值
3. 新增：若要新增一個新課程(course)，但卻尚未被學生選上，則無法新增(SID= ?)
4. 刪除：若刪除 SID 38214的學生資料，卻會將課程 Course IS350的資料同時刪除(若僅有 SID38214 選此課程)

1NF 多值方法：把資料拉出來

A	B	C	D
1		1111 1112	A B
2		222 223	C D
3		331 332	E F

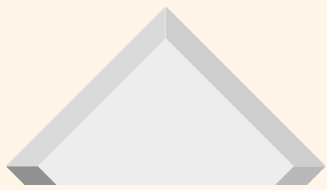
		D
1	1111	A
1	1112	B
2	222	C
2	223	D
3	331	E
3	332	F



二階正規化 (不能部份相依)

Second Normal Form(2NF)

- 定義：一個關聯資料表若已經是1NF，且所有非主要欄位 (nonprime attribute) A 均完全功能性相依(fully functionally dependent)於任何的索引(any key of R) (包含主索引、候選鍵)，則稱此關聯資料表已具2NF
 - 所謂完全功能性相依 $X \rightarrow Y$ 表示，若將X中某欄位移除，則此功能性相依即不存在，反之，若存在，表示部份相依於索引X
- 2NF 一般性定義 (include candidate key)：一個關聯資料表若已經是1NF，且所有非主要欄位(nonprime attribute) A 沒有任何部分功能性相依(fully functionally dependent)於任何的索引(any key of R) (包含主索引)，則稱此關聯資料表已具2NF



由 1NF轉成 2NF的方法

部分相依

(pack-id ,tag-no ,pack-name , comp-id , ins-date, soft-cost)

部分相依

1. 找出原來主索引的所有部份集合，並以此為主索引，各建一關聯資料表

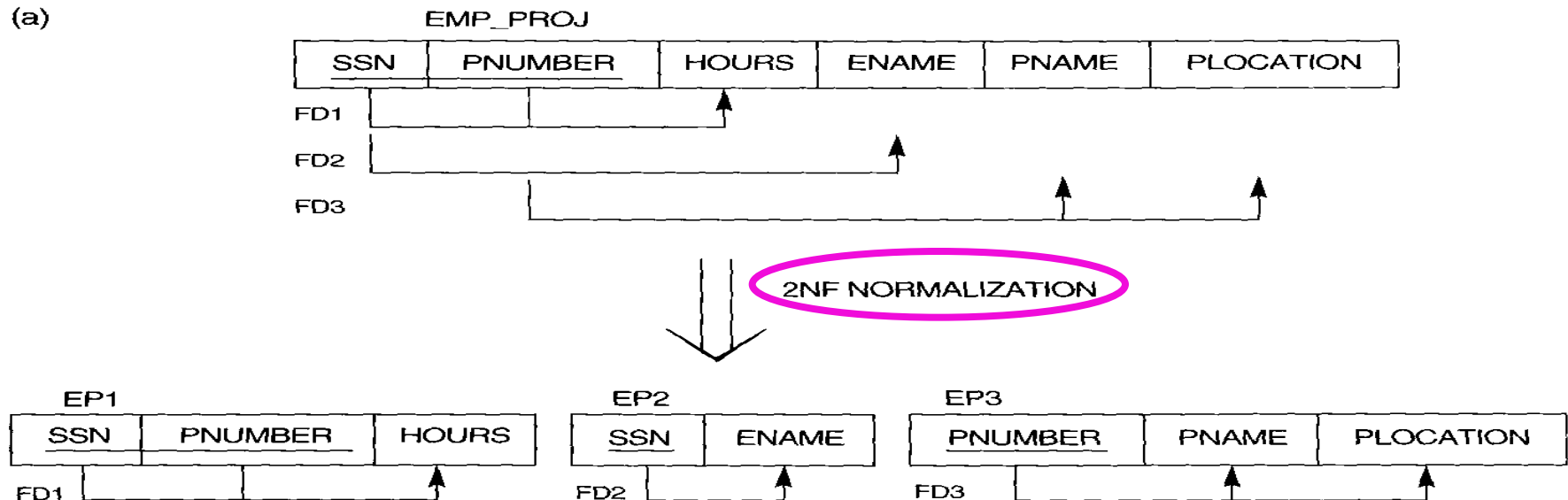
(PACK-ID,
TAG-NO,
(PACK-ID,TAG-NO

2. 再將其他相依於各主索引的屬性欄位分配至各資料表中

(PACK-ID,PACK-NAME)
(TAG-NO,COMP-ID)
(PACK-ID,TAG-NO,INS-DATE,SOFT-COST)

(若此1NF關聯之主索引是由單一欄位組成，則關聯即為2NF)

2NF 範例



3NF 範例

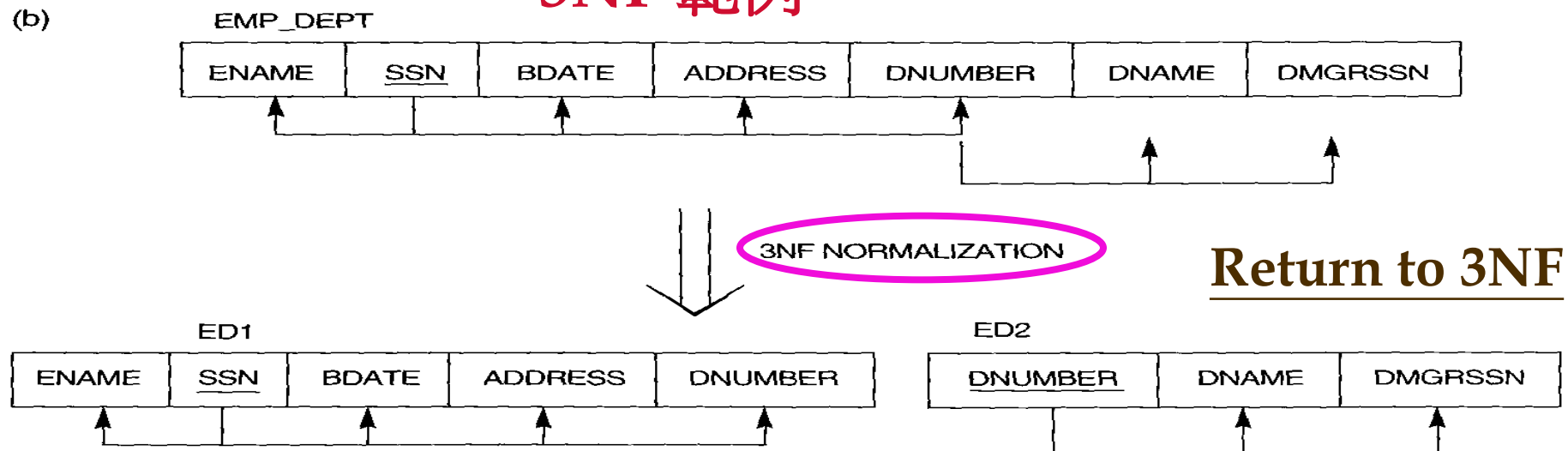


Figure 14.10 The normalization process. (a) Normalizing EMP_PROJ into 2NF relations. (b) Normalizing EMP_DEPT into 3NF relations.

5-3-2 課本範例

- 當執行學生選課資料關聯表的第一階正規化後，目前關聯表已經分割成Students和Classes兩個關聯表。接著繼續執行Classes關聯表的第二階正規化

Classes

<u>sid</u>	<u>c_no</u>	title	<u>eid</u>	instructor	office	room
S001	CS101	計算機概論	E001	陳慶新	CS-102	180-M
S001	CS203	程式語言	E003	李鴻章	M-100	221-S
S001	CS222	資料庫管理系統	E002	楊金欉	CIS-101	100-M
S001	CS213	物件導向程式設計	E003	李鴻章	M-100	500-K
S002	CS222	資料庫管理系統	E002	楊金欉	CIS-101	100-M
S003	CS203	程式語言	E003	李鴻章	M-100	221-S
S003	CS121	離散數學	E002	楊金欉	CIS-101	221-S
S003	CS213	物件導向程式設計	E001	陳慶新	CS-102	622-G
S004	CS222	資料庫管理系統	E002	楊金欉	CIS-101	100-M

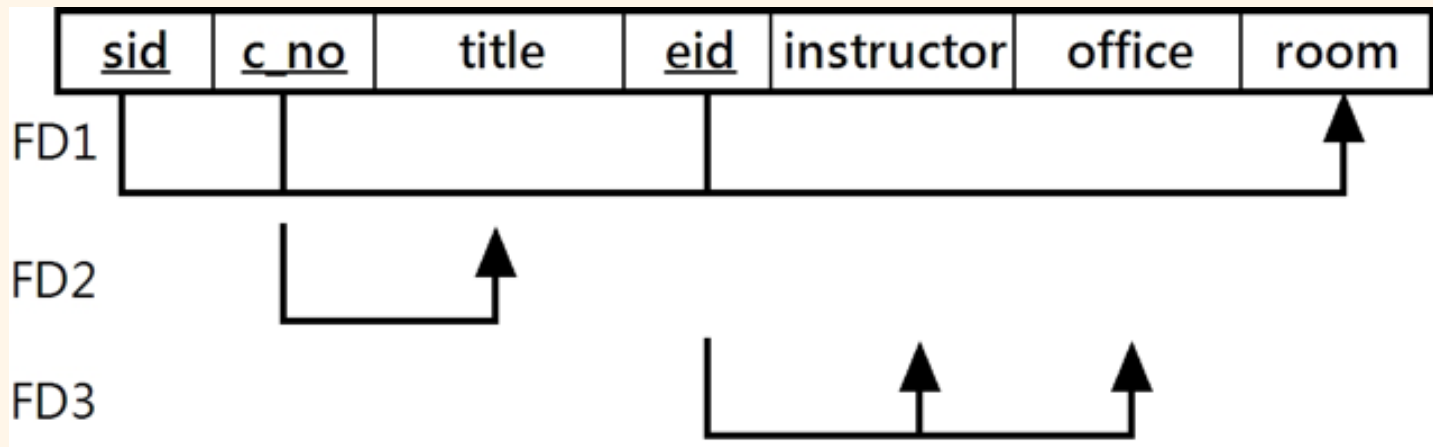
5-3-2 第二階正規化型式 – 2NF (功能相依)

- Classes關聯表的主鍵是(sid, c_no, eid)，關聯表已知的功能相依，如下所示：

FD1 : { sid, c_no, eid } → room

FD2 : c_no → title

FD3 : eid → { instructor, office }



5-3-2 第二階正規化型式 – 2NF (正規化結果)

考試不用寫資料

- 將功能相依 $c_no \rightarrow title$ 和 $eid \rightarrow \{ instructor, office \}$ 兩邊屬性獨立成關聯表，左邊剩下的屬性就是新關聯表的主鍵，如下圖所示：

Classes

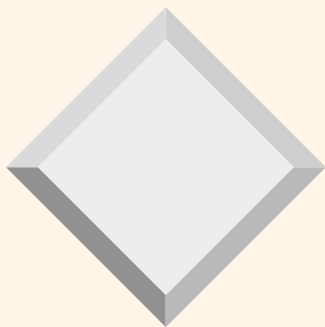
<u>sid</u>	<u>c_no</u>	<u>eid</u>	room
S001	CS101	E001	180-M
S001	CS203	E003	221-S
S001	CS222	E002	100-M
S001	CS213	E003	500-K
S002	CS222	E002	100-M
S003	CS203	E003	221-S
S003	CS121	E002	221-S
S003	CS213	E001	622-G
S004	CS222	E002	100-M

Courses

<u>c_no</u>	title
CS101	計算機概論
CS203	程式語言
CS222	資料庫管理系統
CS213	物件導向程式設計
CS121	離散數學

Instructors

<u>eid</u>	instructor	office
E001	陳慶新	CS-102
E002	楊金欉	CIS-101
E003	李鴻章	M-100

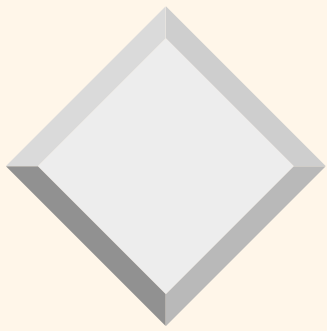


2NF 的問題

PC(TAG-NO,COMP-ID,EMP-NO,EMP-NAME,LOCATION)
TAG-NO --> COMP-ID,EMP-NO,EMP-NAME,LOCATION
EMP-NO --> EMP-NAME

1. 更新：更改 emp-name 的值，必須至少改兩次(相同值出現)
2. 資料不一致：同一員工可能因更新不完全出現兩個名字
3. 新增：必須先給一台 PC 才能新增員工資料
4. 刪除：若刪除 PC 32808 的資料，同時尚失 employee 611 的資料

定義：可以決定其他欄位值的屬性欄位稱為決定者欄位(**determinant**)
主索引、候選鍵都是決定者欄位(候選鍵會引發問題)



3NF

□ 定義：若一關聯資料表已是2NF，且僅有候選鍵為決定者則稱此關聯資料表已具3NF

或 若一關聯資料表已是2NF，且沒有非主要欄位遞移性相依(transitively **dependent**)於任何索引(包含主索引)則稱此關聯資料表已具3NF

– 若某欄位集合Z(不是任意索引的部份集合)，若 $A \rightarrow X$ 且 $X \rightarrow Z$ 則 $A \rightarrow Z$ 稱為遞移性相依。

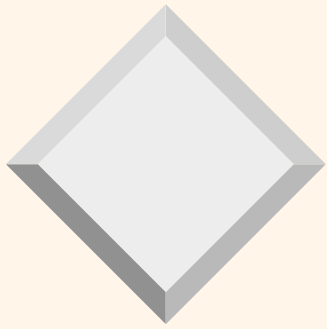
□ 3NF的一般性定義 (**include candidate key**)

若關聯資料表 R 是 3NF，則不存在多餘的 $X \rightarrow A$ ，即若 $X \rightarrow A$ 存在，則必定：

- (a) X 為 R 的超索引(superkey)(即X為包含索引的欄位集合) 或
- (b) A 為R的主要屬性欄位(**prime attribute**)

non prime 不能為決定者

Return BCNF



將2NF轉換為 3NF

1. 對每一個不是候選鍵的決定者，建一新關聯資料表，包含相依於此決定者的欄位
2. 設定此決定者欄位為此新關聯資料表的主索引

e.g. PC(TAG-NO,COMP-ID,EMP-NO,EMP-NAME,LOCATION)

Replaced by:

PC(TAG-NO,COMP-ID,EMP-NO, LOCATION)

EMPLOYEE(EMP-NO,EMP-NAME)

non Prime 的 EMP-NO 為 EMP Name 的決定者 ⇒ 3NF 不可

See example

5-3-3 課本範例

- 例如：請繼續第5-3-2節的關聯表，執行Instructors關聯表的第三階正規化

Instructors

<u>eid</u>	instructor	o_no	office
E001	陳慶新	CS-102	網路研究室
E002	楊金欉	CIS-101	資料庫中心
E003	李鴻章	M-100	數學系電腦中心
E004	王陽明	M-100	數學系電腦中心

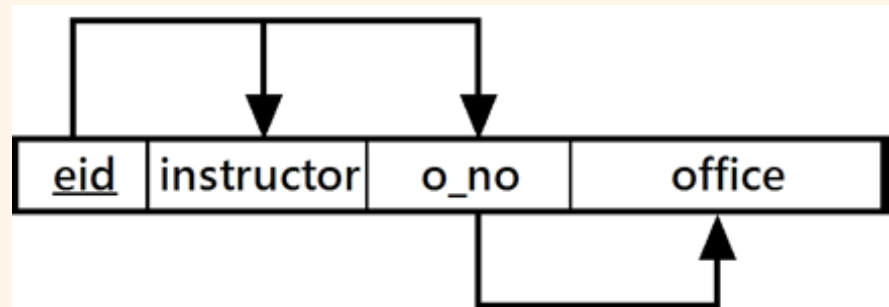
5-3-3 第三階正規化型式 – 3NF (找出遞移相依)

- Instructors 關聯表已知的功能相依，如下所示：

$eid \rightarrow office$

$eid \rightarrow o_no$

$o_no \rightarrow office$



- ~~□ 上述第office屬性功能相依於eid主鍵，它是藉由倒數第o_no屬性功能相依於eid和office屬性功能相依於o_no屬性所得，所以 $eid \rightarrow office$ 是遞移相依，如下圖所示：~~

5-3-3 第三階正規化型式 – 3NF (正規化結果)

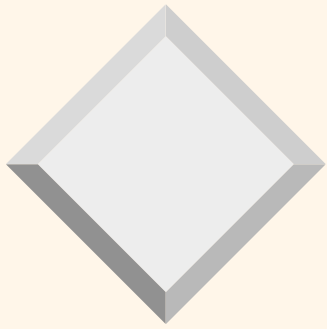
- 在執行第三階正規化時，就是將造成遞移相依的 $A \rightarrow B$ 功能相依兩邊的屬性獨立成關聯表，左邊的屬性就是新關聯表的主鍵

Instructors

<u>eid</u>	instructor	o_no
E001	陳慶新	CS-102
E002	楊金欉	CIS-101
E003	李鴻章	M-100
E004	王陽明	M-100

Offices

<u>o_no</u>	office
CS-102	網路研究室
CIS-101	資料庫中心
M-100	數學系電腦中心



特殊的案例 – 候選鍵的功能性相依 FD on candidate key

2NF

由此可知, county-name-lot
為候選鍵 => 部分相依

3NF

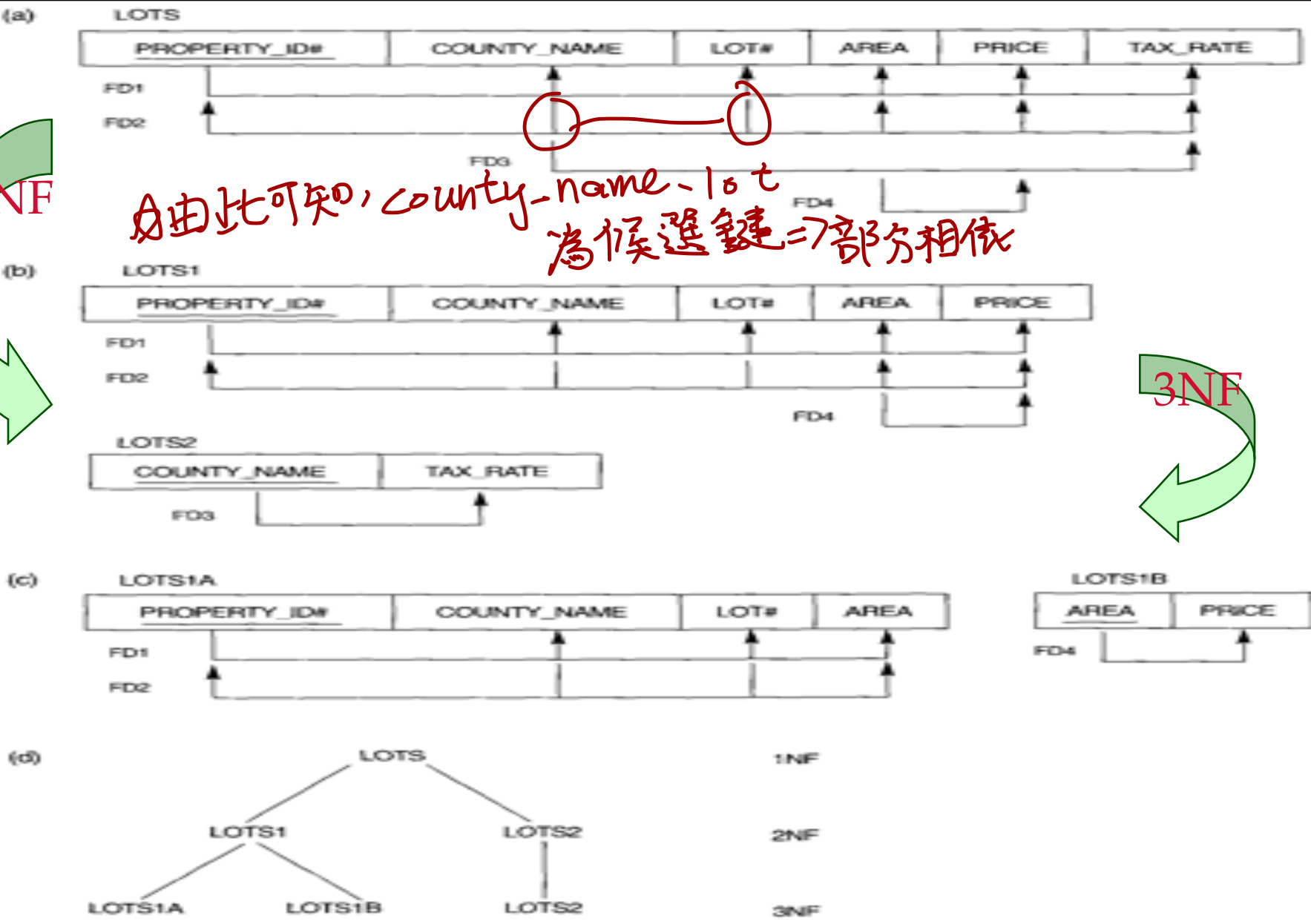


Figure 14.11 Normalization to 2NF and 3NF. (a) The LOTS relation schema and its functional dependencies FD1 through FD4. (b) Decomposing LOTS into the 2NF relations LOTS1 and LOTS2. (c) Decomposing LOTS1 into the 3NF relations LOTS1A and LOTS1B. (d) Summary of normalization of LOTS.

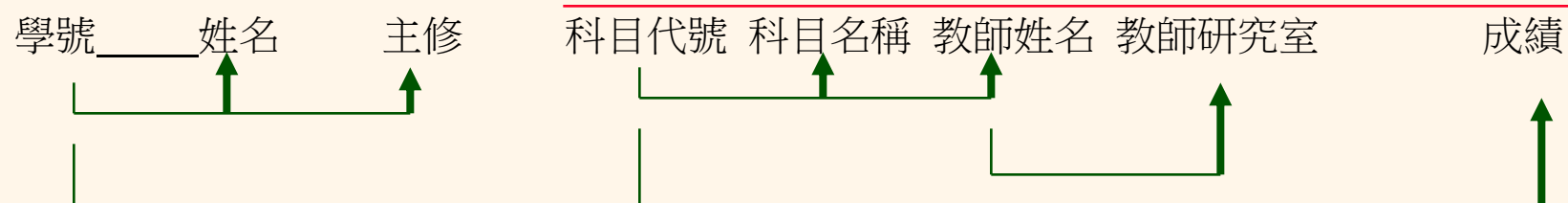
正規化例子

將GRADE_REPORT三階正規化，並說明理由。

GRADE REPORT

學號	姓名	主修	科目代號	科目名稱	教師姓名	教師研究室	成績
38214	BRIGHT	IS	IS350	DATABASE	CODD	F101	A
			IS465	SYS ANAL	KEMP	F103	C
69173	SMITH	PM	IS465	SYS ANAL	KEMP	F103	A
			PM300	PROD MGT	LEWIS	H101	B
			QM440	OP RES	KEMP	F103	C

其中欄位關係如下：



學號和姓名及主修是1-1，而學號和科目間是1-M

正規化例子(續)

□ 第一 正規化: 移除repeating group

學號	姓名	主修	科目代號	科目名稱	教師姓名	教師研究室	成績
----	----	----	------	------	------	-------	----



<u>STUDENT</u>	<u>學號</u>	姓名	主修
	38214	BRIGHT	IS
	69173	SMITH	PM

STUDENT COURSE

<u>學號</u>	<u>科目代號</u>	科目名稱	教師姓名	教師研究室	成績
38214	IS350	DATA BASE	CODD	F101	A
38214	IS465	SYS ANAL	KEMP	F103	C
69173	IS465	SYS ANAL	KEMP	F103	A
69173	PM300	PROD MGT	LEWIS	H101	B
69173	OM440	OP RES	KEMP	F103	C

.....

正規化例子(續)

□ 第二正規化: 本身 為1NF, 並移除部份相依

STUDENT COURSE

<u>學號</u>	<u>科目代號</u>	科目名稱	教師姓名	教師研究室	成績
-----------	-------------	------	------	-------	----

REGISTRATION

<u>學號</u>	<u>科目代號</u>	成績
38214	IS350	A
38214	IS465	C
69173	IS465	A
69173	PM300	B
69173	OM440	C

COURSE INSTRUCTION

<u>科目代號</u>	科目名稱	教師姓名	教師研究室
IS350	DATA BASE	CODD	F101
IS465	SYS ANAL	KEMP	F103
PM300	PROD MGT	LEWIS	H101
OM440	OP RES	KEMP	F103

正規化例子(續)

□ 第三正規化: 本身已是2NF且去除遞移性相依

COURSE INSTRUCTION

<u>科目代號</u>	科目名稱	教師姓名	教師研究室
-------------	------	------	-------

COURSE

<u>科目代號</u>	科目名稱	教師姓名
IS350	DATA BASE	CODD
IS465	SYS ANAL	KEMP
PM300	PROD MGT	LEWIS
OM440	OP RES	KEMP

INSTRUCTOR

<u>教師姓名</u>	教師研究室
CODD	F101
KEMP	F103
LEWIS	H101

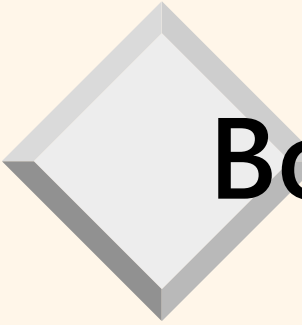
正規化例子(續)--總結

<u>STUDENT</u>	<u>學號</u>	姓名	主修	STUDENT(<u>學號</u> , 姓名, 主修)
	38214	BRIGHT	IS	
	69173	SMITH	PM	

<u>COURSE</u>			<u>INSTRUCTOR</u>	
<u>科目代號</u>	科目名稱	教師姓名	<u>教師姓名</u>	教師研究室
IS350	DATA BASE	CODD	CODD	F101
IS465	SYS ANAL	KEMP	KEMP	F103
PM300	PROD MGT	LEWIS	LEWIS	H101
OM440	OP RES	KEMP		

COURSE(科目代號, 科目名稱, 教師姓名) INSTRUCTOR(教師姓名, 教師研究室)

<u>REGISTRATION</u>			
<u>學號</u>	<u>科目代號</u>	成績	
38214	IS350	A	
38214	IS465	C	
69173	IS465	A	
69173	PM300	B	
69173	OM440	C	REGISTRATION(<u>學號</u> , <u>科目代號</u> , 成績)



Boyce-Codd 正規化 (BCNF)

- Boyce-Codd正規化(normal form, BCNF) 比 3NF簡單但比3NF嚴謹，所以，若某關聯資料表為 BCNF 一定是 3NF，3NF的關聯資料表未必是 BCNF.
- 定義：一個關聯資料表R，任意非多餘功能相依 $X \rightarrow A$ 表示，X必為超索引(superkey) (即X為包含索引的欄位集合)
 - BCNF與 3NF 的區別，是BCNF不具 3NF的(b)性質：允許A是主要欄位(3NF允許，但BCNF不允許)
 - BCNF不允許有相依於非主要欄位(non-prime)的相依關係存在



Figure 14.12 Boyce-Codd normal form. (a) BCNF normalization with the dependency of FD2 being “lost” in the decomposition. (b) A relation R in 3NF but not in BCNF.

TEACH

STUDENT	COURSE	INSTRUCTOR
Narayan	Database	Mark
Smith	Database	Navathe
Smith	Operating Systems	Ammar
Smith	Theory	Schulman
Wallace	Database	Mark
Wallace	Operating Systems	Ahamad
Wong	Database	Orniecinski
Zelaya	Database	Navathe

Figure 14.13 A relation TEACH that is in 3NF but not in BCNF.

多重值相依及 4NF

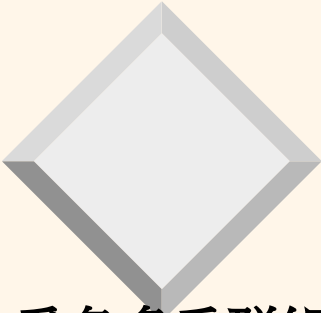
e.g. FACULTY(FAC-NO, FAC-NAME, STU-NO, STU-NAME, COMM-CODE, COMM-DESC)

FACULTY(FAC-NO, FAC-NAME, STU-NO, STU-NAME, COMM-CODE, COMM-DESC)

- 1.更新異常：欲更新委員會 fac-no 123的資料，必須更新2筆以上資料
- 2.新增異常：新faculty 666, 指導學生stu. 57384，但尚未加入任一委員會，則因為 COMM-CODE不詳而無法加入此筆資料
- 3.刪除異常：：刪除 fac-no 444的學生 57384，卻同時刪除 444's HSG 的資訊

定義：若欄位A和某特定B值同時出現(重複群組)而和C值無關，則稱為B多重值相依A(**multivalued dependence**)，記做 $A \twoheadrightarrow B$

定義：一關聯資料表若已經是3NF，且沒有任何多重值相依則稱此關聯資料表若已具4NF



轉換成 4NF

為各多重群組建一關聯資料表，並包含原來之主索引(外來鍵)，結合多重群組之索引與此外來鍵形成主索引)

FACULTY(FAC-NO,FAC-NAME,STU-NO,STU-NAME,COMM-CODE,COMM-DESC)

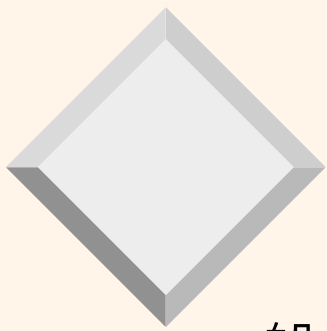
FACULTY(FAC-NO,FAC-NAME)

FACSTU(FAC-NO,STU-NO,STU-NAME)

FACCOMM(FAC-NO,COMM-CODE,COMM-DESC)

*如果在處理1NF時，已經考慮多個重複群組，則3NF就已經是4NF

Skip 5NF



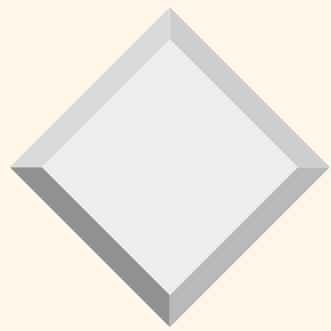
合併相依 Join Dependencies

- 一般將TABLE分割成2個TABLE再合併，可以與原TABLE相同, 但有些必須分割>2個TABLE, 再合併才可以與原TABLE相同
- 合併相依(join dependency, JD), 記為 $JD(R_1, R_2, \dots, R_n)$, 若關聯資料表R，原有資料r，分割為 $R_1, R_2, \dots, R_n$; 使用的

$$*(\Pi_{R_1}(r), \Pi_{R_2}(r), \dots, \Pi_{R_n}(r)) = r$$

即 $R_1, R_2, \dots, R_n$ 為R之projection, 則R等於 $R_1, R_2, \dots, R_n$ 之自然合併, $n > 2$ ，稱為合併相依

- 即若某資料表被分割成數個資料表，必須3個以上資料表合併才會恢復成原來資料表內容，稱為合併相依



第5階正規化 Fifth Normal form(5NF)

- 一個關聯綱要(結構) R 若為4NF且沒有合併相依的關係(JD)則具 5NF 或
- 一個關聯綱要(結構) R 是 5NF (或project-join normal form (PJNF)) 則其每一合併相依(join dependency)僅發生於候選鍵上

(a)

EMP

ENAME	PNAME	DNAME
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

(b)

EMP_PROJECTS

ENAME	PNAME
Smith	X
Smith	Y

EMP_DEPENDENTS

ENAME	DNAME
Smith	John
Smith	Anna

(c)

SUPPLY

SNAME	PARTNAME	PROJNAME
Smith	Bolt	ProjX
Smith	Nut	ProjY
Adamsky	Bolt	ProjY
Walton	Nut	ProjZ
Adamsky	Nail	ProjX
Adamsky	Bolt	ProjX
Smith	Bolt	ProjY

(d)

R1

SNAME	PARTNAME
Smith	Bolt
Smith	Nut
Adamsky	Bolt
Walton	Nut
Adamsky	Nail

R2

SNAME	PROJNAME
Smith	ProjX
Smith	ProjY
Adamsky	ProjY
Walton	ProjZ
Adamsky	ProjX

R3

PARTNAME	PROJNAME
Bolt	ProjX
Nut	ProjY
Bolt	ProjY
Nut	ProjZ
Nail	ProjX

JD(R1, R2, R3)

在任意兩關聯資料表間
使用自然合併(
NATURAL JOIN)都會
產生多出的資料,但將此
3個資料表一起使用自然
合併則不會產生多餘的
資料

Figure 15.4

Fourth and fifth normal form. (a) The EMP relation with two MVDs: ENAME $\twoheadrightarrow$ PNAME and ENAME $\twoheadrightarrow$ DNAME. (b) Decomposing EMP into two relations in 4NF. (c) The relation SUPPLY with no MVDs satisfies 4NF but does not satisfy 5NF if the JD(R1, R2, R3) holds. (d) Decomposing SUPPLY into three 5NF relations.

join dependency的範例

R1

sname	partname
Smith	Bolt
Smith	Nut
Adamsky	Bolt
Walton	Nut
Adamsky	Nail

R2

sname	projname
Smith	ProjX
Smith	ProjY
Adamsky	ProjY
Walton	ProjZ
Adamsky	ProjX

R3

partname	projname
Bolt	ProjX
Nut	ProjY
Bolt	ProjY
Nut	ProjZ
Nail	ProjX

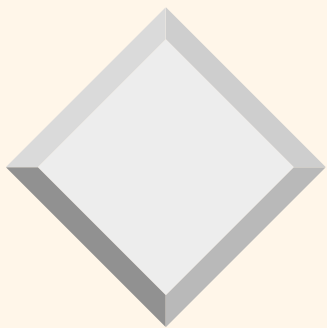
1

sname	partname	projname
Smith	Bolt	ProjX
Smith	Bolt	ProjY
Smith	Nut	ProjX
Smith	Nut	ProjY
Adamsky	Bolt	ProjY
Adamsky	Bolt	ProjX
Walton	Nut	ProjZ
Adamsky	Nail	ProjY
Adamsky	Nail	ProjX

2

非用候選鍵合併, 所以原檔非5NF, 所以必須將SUPPLY分割成R1,R2,R3

sname	partname	projname
Smith	Bolt	ProjX
Smith	Bolt	ProjY
Smith	Nut	ProjY
Adamsky	Bolt	ProjY
Adamsky	Bolt	ProjX
Walton	Nut	ProjZ
Adamsky	Nail	ProjX



關於 5NF

- 實務上這種功能性相依很難發現，所以，實務上很少考慮5NF