

# 第 10 章

## 關聯式資料庫的功能相依性 與正規化

### 學習重點

- 關聯網要的非正式設計原則
- 關聯屬性的語意
- 值組中的資料重複和更新異常情況
- 值組中的空值
- 假值組
- 功能相依性
- 正規化的一般程序
- 關聯的正規化

## 學習重點

---

- 正規化的實際使用
- 鍵值與屬性的定義
- 第一正規化形式
- 第二正規化形式
- 第三正規化形式
- 第三正規化的一般化定義
- BCNF (Boyce-Codd Normal Form)
- 藉由分解動作達到BCNF

3

## 關聯網要的非正式設計原則 (1)

---

- 什麼是關聯式資料庫設計？  
如何將屬性分組以形成「良好的」關聯網要
- 關聯網要的兩個層次
  - 邏輯的“使用者視界”層
  - 儲存的“基底關聯”層
- 本章所說明的設計理論主要是應用在基底關聯上
- 所謂「好的」基底關聯有什麼條件？

4

## 關聯網要的非正式設計原則 (2)

- 我們首先討論良好關聯式設計的非正式設計原則
- 接著討論功能相依性和正規化的正式概念
  - 1NF (First Normal Form, 第一正規化形式)
  - 2NF (Second Normal Form, 第二正規化形式)
  - 3NF (Third Normal Form, 第三正規化形式)
  - BCNF (Boyce-Codd Normal Form, BCNF 正規化形式)

5

## 關聯屬性的語意

原則1：盡量不要將多個實體類型和關係類型的屬性結合在同一個關聯中

- 不同實體的屬性 (EMPLOYEE、DEPARTMENT、PROJECT) 不要混雜在同一個關聯裡
- 應該只有外來鍵才會被用來參考其他屬性
- 實體與關係屬性應該要盡量分離

目標: 設計關聯網要時應該盡量要讓每個關聯的意義容易解釋，屬性的意義也應該很容易了解

6

圖 10.1  
COMPANY 關聯式資料庫綱要的簡化版本

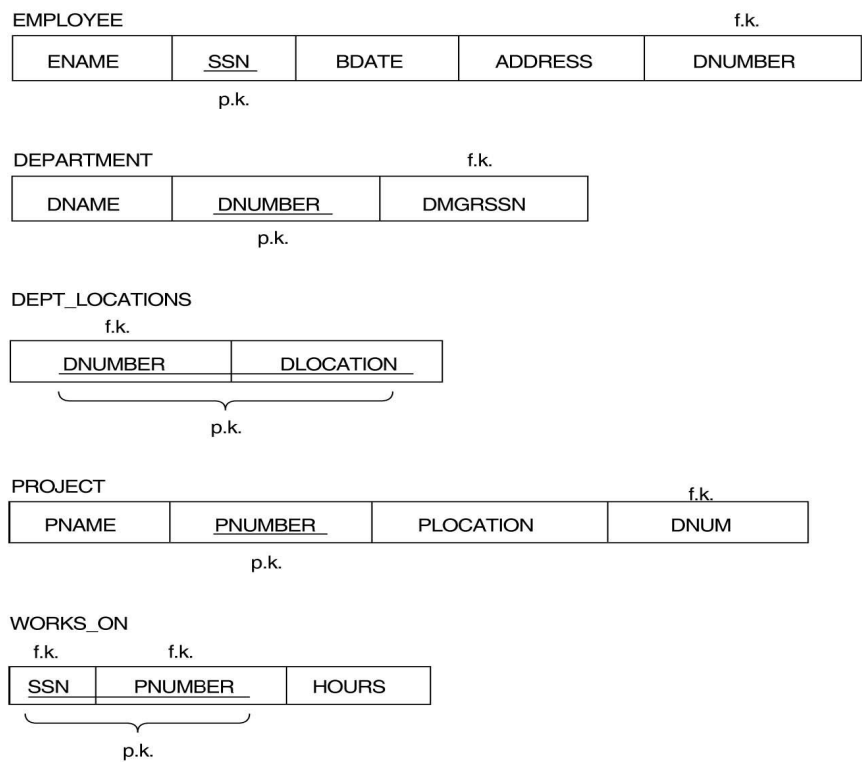


圖 10.1 COMPANY 關聯式資料庫綱要的簡化版本

## 值組中的資料重複和更新異常情況

- 不同實體的屬性混雜在一起可能會出問題
- 資訊重複儲存而浪費空間
- 更新異常的問題
  - 新增異常 (Insertion anomalies)
  - 刪除異常 (Deletion anomalies)
  - 修改異常 (Modification anomalies)

## 更新異常的範例 (1)

---

以這個關聯為例：

EMP\_PROJ ( Emp#, Proj#, Ename, Pname, No\_hours)

- 更新異常：變更計畫編號P1的計畫名稱，從“Billing”改成“Customer-Accounting”可能會導致必須修改在計畫P1上工作的所有100位員工的資料

9

## 更新異常的範例 (2)

---

- 新增異常：若要新增一個計畫就一定要指派員工  
相反的 - 若要新增員工就一定要為他指派計畫
- 刪除異常：如果要刪除某個計畫，將會導致在該計畫上工作的所有員工也會被刪除。反過來看，如果某位員工是該計畫唯一的工作人員，則刪除該位員工將會導致該計畫也會被刪除

10

圖10.3 這兩個關聯網要都有更新異常的問題

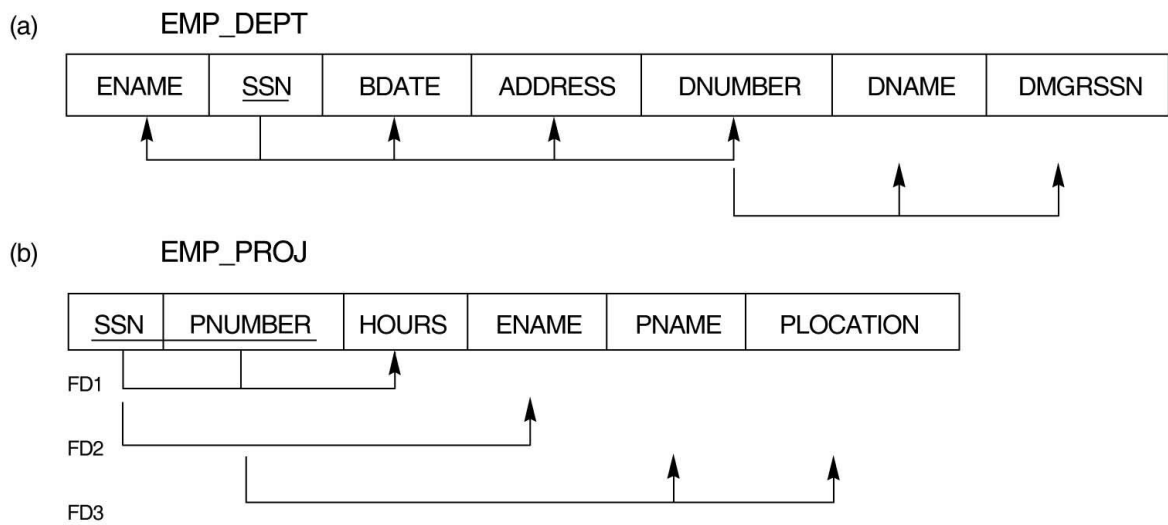


圖 10.3 這兩個關聯網要都有更新異常的問題

EMP_DEPT					重複	
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 FireOak, 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		重複		重複	
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

圖 10.4 對圖 10.2 執行NATURAL JOIN動作所產生的EMP\_DEPT和EMP\_PROJ關聯狀態範例。為了效能原因可能會將它們儲存成其底關聯

## 針對資料重複和更新異常的原則

- 原則2：設計基底關聯時必須避免發生新增、刪除或修改異常的情形。假如發現任何異常情況，要清楚的記錄下來，並確定程式能正確更新資料庫

13

## 值組中的空值

原則3：盡量別讓基底關聯中的屬性值經常為空值

- 內容值經常是NULL的屬性應該要放在另外分開的關聯中(還有主鍵)
- 空值可以有許多種解釋：
  - 這個值組沒有這個屬性值
  - 這個值組的這個屬性值不知道是否存在
  - 已知這個屬性值存在，但不知道它的值

14

## 假值組

- 不好的關聯式資料庫設計可能會導致某些JOIN運算將產生錯誤的結果

原則4：在設計關聯網要時，若要使關聯資料可以針對屬性使用相等條件進行合併動作，該屬性就必須是主鍵或外來鍵，這樣就可以確保不會產生假值組。因此要盡量讓關聯的相配屬性是（外來鍵，主鍵）的組合，否則這樣的合併動作可能會產生假值組

## 功能相依性 (1)

- 所謂的功能相依性 (functional dependency, FD) 是指屬性間的限制，它是評估關聯網要中將屬性分組為關聯是否正確的主要方式
- FD和鍵值是用來定義關聯的正規化形式 (normal form)
- FD是指資料庫中屬性集合之間的意義與相互關係的限制
- 在兩組屬性集合X與Y之間，假如X的值可以唯一決定Y的值，則稱Y功能相依於X，表示成 $X \rightarrow Y$



## 功能相依性 (2)

- 假如任兩個值組的X值相同，Y值也就一定相同，就表示  $X \rightarrow Y$  成立
- 對於任何關聯實例  $r(R)$  的任何兩個值組  $t1$  和  $t2$  而言，假如  $t1[X] = t2[X]$  成立，則  $t1[Y] = t2[Y]$  也一定成立
- 在  $X \rightarrow Y$  中，屬性集合X稱為FD的左手邊 (left-hand side)，而屬性集合Y則稱為FD的右手邊
- $X \rightarrow Y$  在關聯網要圖中是以水平線連接箭頭表示
- FD是從屬性在真實世界上的限制所推導出來的

17

## FD限制的範例 (1)

- 從社會安全號碼可唯一決定員工姓名  
 $SSN \rightarrow ENAME$
- 專案編號可唯一決定專案名稱與專案地點  
 $PNUMBER \rightarrow \{PNAME, PLOCATION\}$
- 員工的社會安全號碼和專案編號可唯一決定員工每週在專案中工作的時數  
 $\{SSN, PNUMBER\} \rightarrow HOURS$

18

## FD限制的範例 (2)

- FD是關聯網要R的一種特性
- 這個限制必須在每個關聯實例 $r(R)$ 上都成立
- 假如K是R的一個鍵值，則從K可以唯一決定R中的所有屬性 (因為絕對不會有兩個不同的值組是  $t1[K]=t2[K]$ )

19

## 正規化的一般程序

- 通常正規化的程序可以分成幾個階段，每一階段各對應一種正規化形式：
  1. 第一正規化形式 (1NF)：目的是移除多值屬性 (又稱為重複群組)，使表格的每個行列交會點都是單一值 (可能為 null)
  2. 第二正規化形式 (2NF)：移除部份的功能相依性
  3. 第三正規化形式 (3NF)：移除可遞移相依性
  4. Boyce/Codd正規化形式 (BCNF)：移除那些因為功能相依性所產生的其餘異常
  5. 第四正規化形式 (4NF)：移除多值相依性
  6. 第五正規化形式 (5NF)：移除合併相依性

20

## 關聯的正規化

---

- 正規化 (normalization)：將不夠好的關聯其屬性分組成較小關聯的過程
- 正規化形式 (normal form, NF)：藉由關聯的鍵值和FD條件，判斷這個關聯網要是否符合某個正規化形式

21

## 正規化的實際使用

---

- 進行正規化動作的目的是要讓設計的品質更高，並且更符合需求
- 如果資料庫設計人員和使用者很難瞭解或偵測有這些限制，那麼正規化的效果就會大打折扣
- 資料庫設計人員不必一定要將關聯正規化到最高的正規化形式 (通常是到3NF、BCNF或4NF)
- 去正規化 (denormalization)：將符合較高正規化的關聯合併成符合較低正規化的基底關聯的過程

22

## 鍵值與屬性的定義 (1)

- 關聯網要  $R = \{A_1, A_2, \dots, A_n\}$  的超鍵 (superkey) 是一組屬性的集合  $S \subseteq R$ ，而且在  $R$  的任何合法關聯狀態  $r$  中，不會存在有任何兩筆值組  $t_1$  和  $t_2$ ，使得  $t_1[S] = t_2[S]$
- 假如鍵值 (key)  $K$  是一個超鍵，但若移除其中任何一個屬性它就不再是超鍵

23

## 鍵值與屬性的定義 (2)

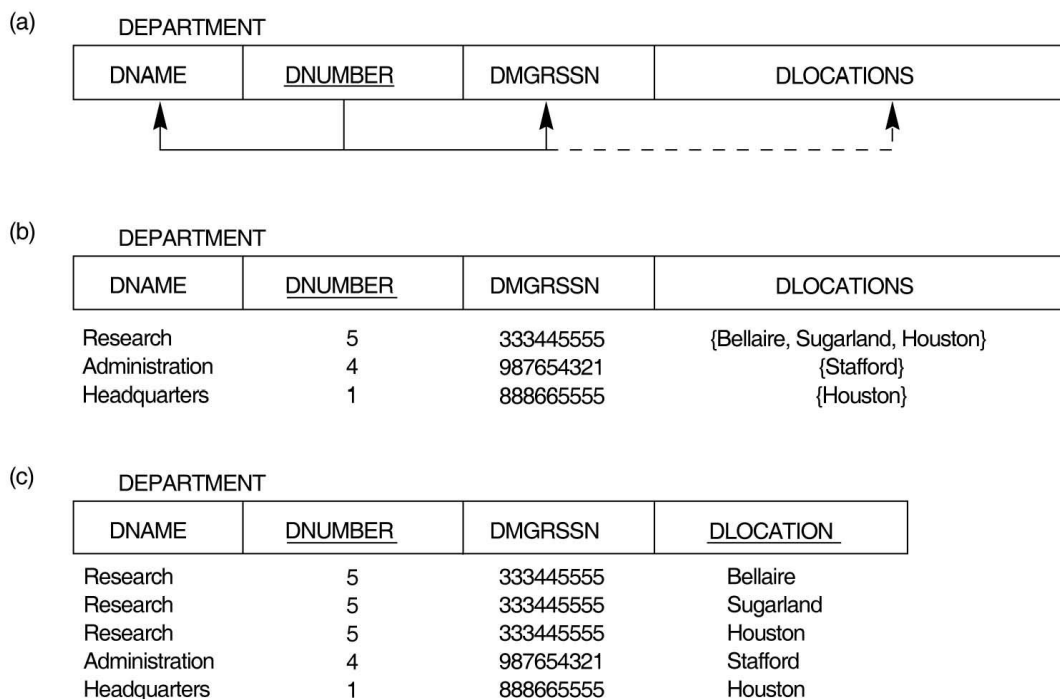
- 假如某個關聯網要有一個以上的鍵值，則每個都稱為候選鍵 (candidate key)。候選鍵中的其中一個可被指定為主鍵 (primary key)，其他稱為次要鍵 (secondary key)
- 在  $R$  中某個屬性如果是  $R$  的候選鍵之一，則稱之為關聯  $R$  中的主要屬性 (prime attribute)，其他則稱為次要 (nonprime) 屬性，也就是說它不是任何一個候選鍵

24

# 第一正規化形式

- 不允許多值屬性、複合屬性和巢狀關聯，在每筆值組中屬性的值必須是不可分割的 (non-atomic)
- 它被認為是關聯定義的一部分

25



◎ 圖 10.8 正規化至 1NF。 (a) 這個關聯網要不符合 1NF； (b) DEPARTMENT 關聯的範例狀態； (c) 同一個關聯的 1NF 版本，此時有重複資料

26

(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
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

(c) EMP\_PROJ1

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

EMP\_PROJ2

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

⊖ 圖 10.9 將巢狀關聯正規化至 1NF。 (a) EMP\_PROJ 關聯的綱要與一個「巢狀關聯」屬性 PROJS； (b) EMP\_PROJ 關聯的範例狀態，展示每個值組中的巢狀關聯； (c) 將 EMP\_PROJ 分解成 EMP\_PROJ1 和 EMP\_PROJ2 關聯並傳遞主鍵

27

## 第二正規化形式 (1)

### ■ 使用 FD 和主鍵的觀念

#### 定義：

- 主要屬性：此屬性是主鍵 K 的成員之一
- 完全功能相依 (full functional dependency)：假設功能相依  $X \rightarrow Y$  是完全功能相依，則由 X 移除任何屬性 A，則相依性就無法成立

範例：{SSN, PNUMBER}  $\rightarrow$  HOURS 是個完全 FD，因為無論是 SSN  $\rightarrow$  HOURS 或 PNUMBER  $\rightarrow$  HOURS 都不成立

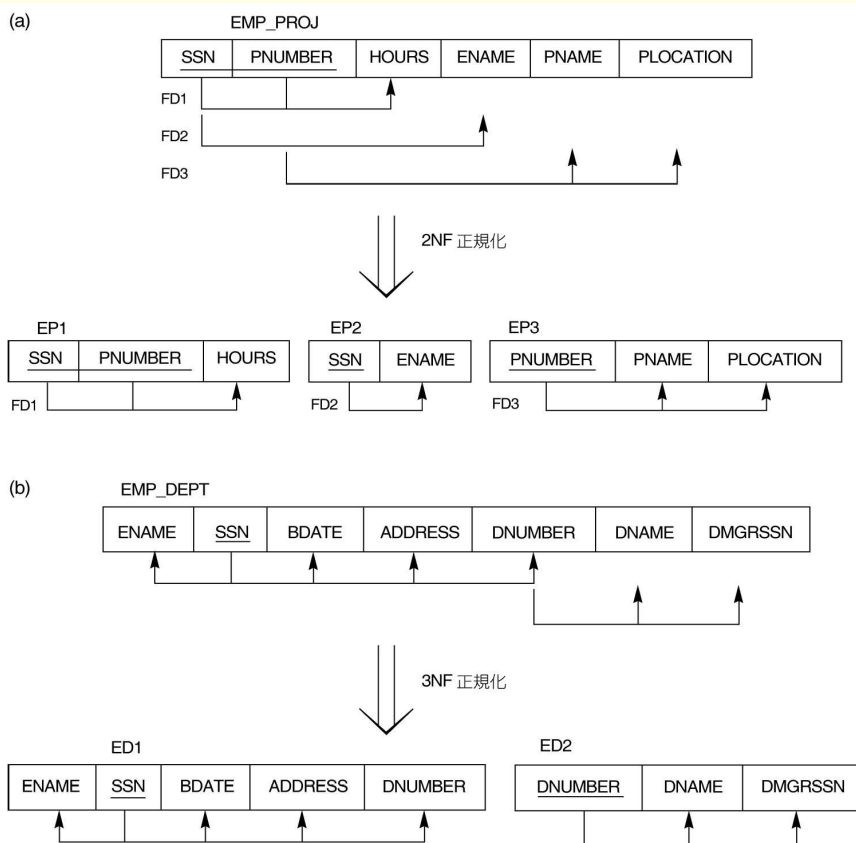
- {SSN, PNUMBER}  $\rightarrow$  ENAME 不是個完全 FD，因為 SSN  $\rightarrow$  ENAME 也成立。這稱作部分功能相依 (partial functional dependency)

28

## 第二正規化形式 (2)

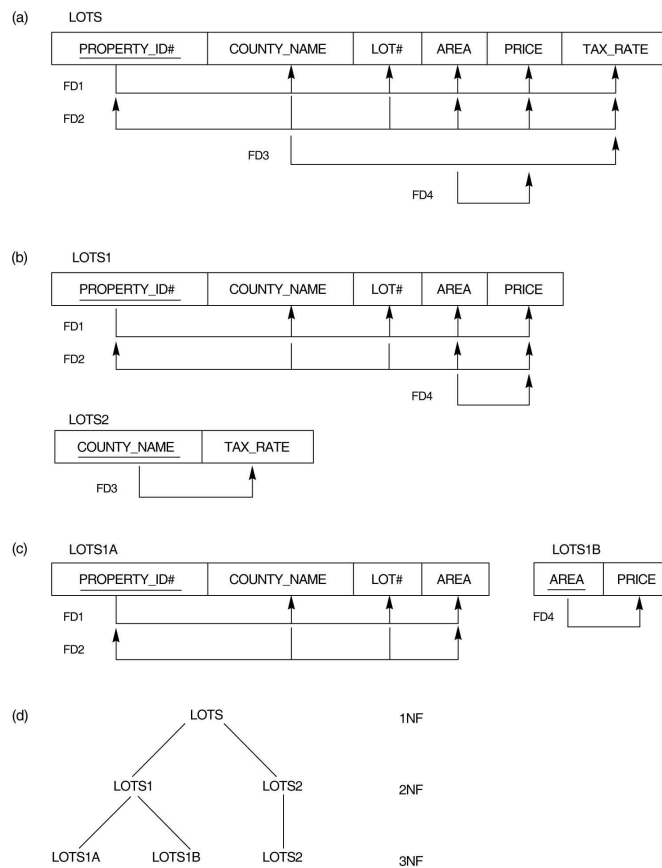
- 若關聯R中每個非主鍵的屬性A都完全功能相依於R的主鍵，則關聯R符合2NF
- 假如關聯不符合2NF，則可將關聯分解成多個關聯，使得每個關聯中非主鍵的屬性都完全功能相依於主鍵，如此便可符合2NF

29



◎ 圖 10.10 正規化至 2NF 和 3NF。(a) 將 EMP\_PROJ 正規化成為 2NF 的關聯；(b) 將 EMP\_DEPT 正規化成為 3NF 的關聯

30



◎ 圖 10.11 正規化至 2NF 和 3NF。 (a) LOTS 關聯與它的功能相依性 FD1 到 FD4； (b) 分解成 2NF 關聯 LOTS1 與 LOTS2； (c) 將 LOTS1 分解成 3NF 關聯 LOTS1A 與 LOTS1B； (d) LOTS 正規化過程的示意圖

## 第三正規化形式 (1)

### 定義：

- 遞移相依性 (transitive dependency)：從  $X \rightarrow Y$  和  $Y \rightarrow Z$  這兩個 FD 可以推導出 FD  $X \rightarrow Z$

### 範例：

- $SSN \rightarrow DMGRSSN$  是個遞移 FD，因為  $SSN \rightarrow DNUMBER$  和  $DNUMBER \rightarrow DMGRSSN$  都成立
- $SSN \rightarrow ENAME$  是非遞移的，因為並沒有一組屬性  $X$  可以讓  $SSN \rightarrow X$  和  $X \rightarrow ENAME$  都成立



## 第三正規化形式 (2)

- 關聯網要R如果已經是2NF，而且沒有任何R的非主要屬性是遞移相依於主鍵，則R符合3NF
- R透過3NF正規化程序可以被分解成3NF關聯

注意：

假設 $X \rightarrow Y$ 且 $Y \rightarrow Z$ 都成立，其中X是主鍵，此時只有當Y不是候選鍵時才会有問題。如果Y是候選鍵，遞移相依性就沒有問題

以EMP (SSN, Emp#, Salary) 為例

此例 $SSN \rightarrow Emp\# \rightarrow Salary$ ，而且Emp# 是個候選鍵

33

## 第三正規化的一般化定義 (1)

- 以上的定義都是只針對主鍵
- 接下來的定義會將關聯的所有候選鍵也納入考慮
- 假如在R中的每一個非主要屬性A都不是部份相依於R的任何鍵值，則稱關聯R符合2NF

34

## 第三正規化的一般化定義 (2)

定義：

- 關聯網要R的超鍵 (superkey)：一組包含R的鍵值的屬性集合S
- 假如在關聯R中有任何功能相依性 $X \rightarrow A$ 成立，而且
  - (a) X是R的超鍵，或者
  - (b) A是R的主要屬性

這兩個條件其中之一成立，則稱關聯R符合3NF

注意：BCNF不允許條件 (b) 成立

35

## Boyce-Codd正規化形式 (BCNF)

- 假如在關聯R中有任何功能相依性 $X \rightarrow A$ 成立，而且X是R的超鍵，則稱關聯網要R符合BCNF
- 每一種正規化形式都是比前一種更強的正規化形式
  - 每個2NF關聯也一定符合1NF
  - 每個3NF關聯也一定符合2NF
  - 每個BCNF關聯也一定符合3NF
- 有些關聯符合3NF但不符合BCNF
- 目標是讓每個關聯都符合BCNF (或3NF)

36

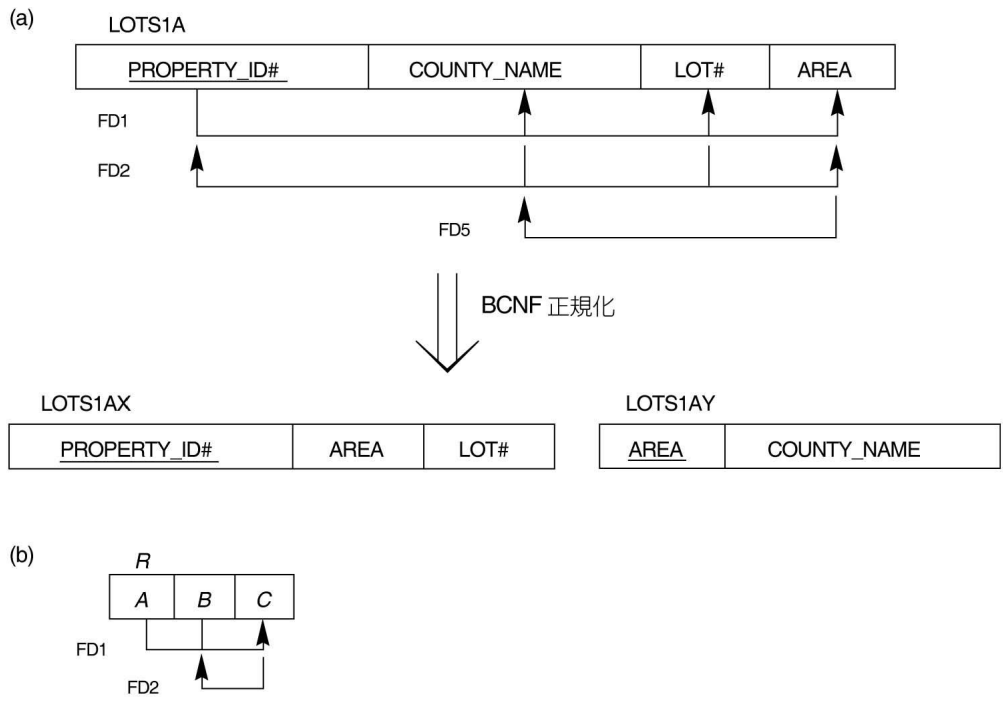


圖 10.12 Boyce-Codd正規化形式。(a) 對LOTS1A關聯進行BCNF正規化，功能相依性FD2 在分解過程中將失去；(b) 某個關聯與其FD，它符合 3NF，但不符合BCNF

### 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	Omiecinski
Zelaya	Database	Navathe

圖 10.13 這個TEACH關聯符合 3NF但不符合BCNF

## 藉由分解動作達到BCNF (1)

- 假設在關聯TEACH中存在兩個FD：  
fd1: { student, course } -> instructor  
fd2: instructor -> course
- {student, course} 是候選鍵，而其相依性如圖10.12(b)所示。因此這個關聯符合3NF但不符合BCNF
- 不符合BCNF的關聯應該要加以分解，同時要盡量保存所有的功能相依性

39

## 藉由分解動作達到BCNF (2)

- 關聯TEACH有以下3種可能的分解方式：
  1. {student, instructor} and {student, course}
  2. {course, instructor} and {course, student}
  3. {instructor, course} and {instructor, student}
- 以上這3種分解方式都會遺失功能相依性FD1
- 只有第3種在關聯合併後不會產生假值組

40

## 學習評量

- 請說明新增、刪除和修改異常。為什麼它們被認為是不好的？請舉例說明。
- 為何在一個關聯中有許多空值被認為是不好的？請說明假值組的問題，以及如何避免它發生。
- 何謂功能相依性？
- 何謂非正規化關聯？由第一正規化到Boyce-Codd正規化的發展過程為何？

41

## 學習評量

- 假設有個全域 (universal) 關聯  $R = \{A, B, C, D, E, F, G, H, I, J\}$ ，其功能相依性集合  $F = \{\{A, B\} \rightarrow \{C\}, \{A\} \rightarrow \{D, E\}, \{B\} \rightarrow \{F\}, \{F\} \rightarrow \{G, H\}, \{D\} \rightarrow \{I, J\}\}$ 。請問何者為  $R$  的鍵值？請將  $R$  分解為 2NF，然後再分解成 3NF 的關聯。
- 重複上題，但使用不同的功能相依性集合  $G = \{\{A, B\} \rightarrow \{C\}, \{B, D\} \rightarrow \{E, F\}, \{A, D\} \rightarrow \{G, H\}, \{A\} \rightarrow \{I\}, \{H\} \rightarrow \{J\}\}$

42