

# 第 8 章 正規化



授課教師：\_\_\_\_\_

**Database Systems Core Theory and Practice**



## 第 8 章 . 正 規 化

2

### 8 - 2 好的 關 聯 模 式 資 料 庫 綱 目 的 特 性

編撰: 黃三益

Copyright 黃三益2023  
資料庫的核心理論與實務第八版



# 好的關聯網目特性

## 意義明確

- 記載迷你世界裡的同一類實體或關係
- 以下的關聯包括現實世界裡的交易和會員實體，所以意義不明確

TransactionMember

交易編號 <u>tNo</u>	交易方式 method	交易時間 transTime	會員編號 mId	會員ID pId	姓名 name	生日 birthday
91100	cart	2005-02-02:18:30:00	a0911234	A122555888	Tony	1980-12-12
92666	cart	2005-10-10:22:10:30	c0927777	B123123123	Su	1982-06-06
91888	fax	2005-09-10:10:10:00	a0910001	A220123456	Jenny	1979-01-01
92333	email	2005-10-15:09:00:00	c0927777	B123123123	Su	1982-06-06
90111	cart	2005-05-05:12:30:30	b0905555	C200456789	Jennifer	1974-03-04
92555	cart	2005-11-11:09:10:00	b0922468	R100200300	Jackson	1980-03-30

(b)





# 好的關聯網目特性 (Cont.)

## 資料不重複儲存

- 空間浪費
- 更新異常

以下關聯 (OrderProduct) 裡「任賢齊專輯三」的資料重複儲存



OrderProduct	<u>會員編號</u> mId	<u>購物車產生時間</u> cartTime	<u>商品編號</u> pNo	訂購數量 amount	商品名稱 pName	定價 unitPrice	種類 category
--------------	--------------------	----------------------------	--------------------	----------------	---------------	-----------------	----------------

(a)

OrderProduct	<u>交易編號</u> tNo	<u>購物車產生時間</u> cartTime	<u>商品編號</u> pNo	訂購數量 amount	商品名稱 pName	定價 unitPrice	種類 category
	a0911234	2005-02-02:18:00:30	b30999	1	資料庫理論與實務	500	Book
	b0905555	2005-05-05:12:00:00	v01888	3	哈利波特：混血王子的背叛	450	DVD
	c0927777	2005-10-10:22:00:00	d11222	1	任賢齊專輯三	300	CD
	c0927777	2005-10-10:22:00:00	d20777	1	蔡依林專輯二	350	CD
	c0927777	2005-10-10:22:00:00	v00111	2	英雄	400	DVD
	b0922468	2005-11-11:09:00:30	b10234	5	管理資訊系統概論	600	Book
	b0922468	2005-11-11:09:00:30	b40555	10	系統分析理論與實務	550	Book
	a0910001	2005-09-09:10:00:10	d11222	1	任賢齊專輯三	300	CD

(b)



# 好的關聯網目特性 ( Cont.)

## 有三種更新異常

### ○ 新增異常

```
INSERT INTO TransactionMember(tNo, mId, name)
VALUES ( '93000' , 'c0927777' , 'Chen' );
```

### ○ 刪除異常

```
DELETE TransactionMember
WHERE tNo = '90111' ;
```

### ○ 修改異常

```
UPDATE TransactionMember
SET name = 'Huang'
WHERE tNo = '92333';
```

TransactionMember	交易編號	交易方式	交易時間	會員編號	會員ID	姓名	生日
	tNo	method	transTime	mId	pId	name	birthday
	91100	cart	2005-02-02:18:30:00	a0911234	A122555888	Tony	1980-12-12
	92666	cart	2005-10-10:22:10:30	c0927777	B123123123	Su	1982-06-06
	91888	fax	2005-09-10:10:10:00	a0910001	A220123456	Jenny	1979-01-01
	92333	email	2005-10-15:09:00:00	c0927777	B123123123	Su	1982-06-06
	90111	cart	2005-05-05:12:30:30	b0905555	C200456789	Jennifer	1974-03-04
	92555	cart	2005-11-11:09:10:00	b0922468	R100200300	Jackson	1980-03-30



(b)



# 好的關聯網目特性 (Cont.)

## ∞ 屬性值很少有空值



- 一個設計不良的網目可能會造成有些屬性有許多空值

### ○ 缺點

- 浪費空間
- 彙總函數應用在有空值的屬性時，其意義不明確
- 空值的含義不唯一

ProductAuthor

商品編號 pNo	商品名稱 pName	定價 unitPrice	種類 category	創作者名稱 1 authorName1	創作者名稱 2 authorName2	創作者名稱 3 authorName3	創作者名稱 4 authorName4
b30999	資料庫理論與實務	500	Book	Huang	null	null	null
d11222	任賢齊專輯三	300	CD	William	null	null	null
b20666	OLAP 進階	500	Book	Sandra	null	null	null
b10234	管理資訊系統概論	600	Book	Lin	null	null	null
b40555	系統分析理論與實務	550	Book	Wu	null	null	null
d20777	蔡依林專輯二	350	CD	Jolin	null	null	null
v01888	哈利波特：混血王子的背叛	450	DVD	J.K.	null	null	null
d03333	5566 專輯	450	CD	Jackey	David	Tom	null
b51111	電子商務理論與實務	700	Book	Lai	Huang	Lin	null
v00111	英雄	400	DVD	null	null	null	null

▲ 圖 8-3 ProductAuthor 關聯



## 好的關聯網目特性 ( Cont.)

- ∞ 當關聯裡的屬性太多時，就可能造成以上三種特性被違反
- ∞ 但將一個關聯切割成兩個時，不可以亂切
  - 如下頁圖關聯的RecordProduct關聯不可以切成Record2和Product2兩個關聯
  - 將Reocrd2和Product2關聯JOIN後產生兩筆虛假序列值 ( 以藍色底色顯示 )



# 好的關聯網目特性 (Cont.)



RecordProduct	交易編號	商品編號	交易數量	售價	商品名稱	定價	種類
	tNo	pNo	amount	salePrice	pName	unitPrice	category
	91100	b30999	1	450	資料庫理論與實務	500	Book
	92555	b10234	5	3000	管理資訊系統概論	600	Book
	92666	d11222	1	300	任賢齊專輯三	300	CD

(a)

Record2	交易編號	商品編號	交易數量	售價	種類
	tNo	pNo	amount	salePrice	category
	91100	b30999	1	450	Book
	92555	b10234	5	3000	Book
Product2	92666	d11222	1	300	CD
	商品名稱	定價	種類		
	pName	unitPrice	category		
	資料庫理論與實務	500	Book		
	管理資訊系統概論	600	Book		
	任賢齊專輯三	300	CD		

(b)

RecordProduct2	交易編號	商品編號	交易數量	售價	商品名稱	定價	種類
	tNo	pNo	amount	salePrice	pName	unitPrice	category
	91100	b30999	1	450	資料庫理論與實務	500	Book
	91100	b30999	1	450	管理資訊系統概論	600	Book
	92555	b10234	5	3000	資料庫理論與實務	500	Book
	92555	b10234	5	3000	管理資訊系統概論	600	Book
	92666	d11222	1	300	任賢齊專輯三	300	CD

(c)

▲ 圖 8-4 (a) RecordProduct 關聯；(b) 分解成不良的兩個關聯：Record2 和 Product2；(c) Record2 \* Product2，裡頭包含了兩筆虛假序列值（淡藍網底）





# 好的關聯網目特性 ( Cont.)

- ∞ 若將一個關聯網目分解成數個，分解的方式要注意
- ∞ **無損JOIN的分解** ( Lossless join decomposition )：切割之後的兩個關聯，其共同屬性必須是其中一個關聯的主鍵(或次要建)
  - 上例的分解 ( Product2和Record2關聯 ) 不滿足無損JOIN的分解，因為共同屬性「category」在兩個關聯裡既非主鍵也非次要鍵



## 第 8 章 . 正 規 化

10

### 8 - 3 函 數 相 依



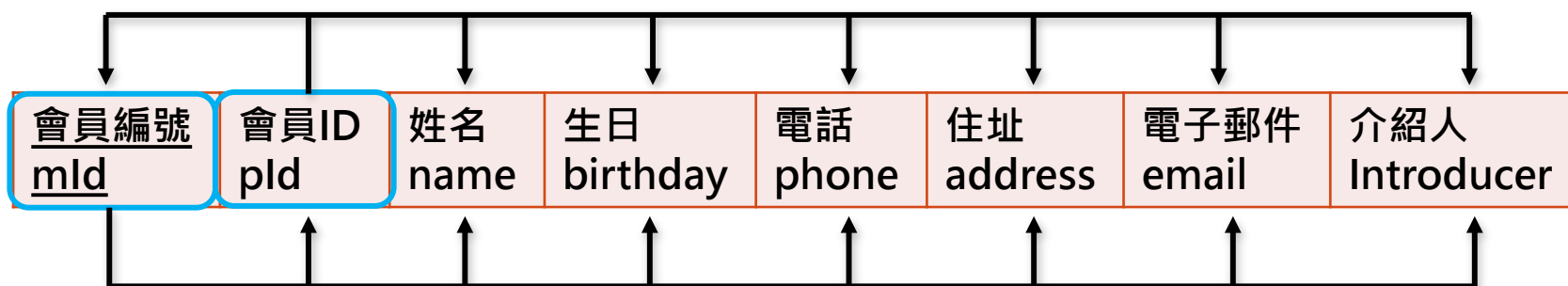
# 函數相依

- ∞ 定義：兩個（群）屬性間存在的一種類似函數（function）裡定義域（domain）和對應域（co-domain）的關係
  - $\{mld\} \rightarrow \{name, birthday\}$ 
    - 每個mld剛好有一組{name, birthday}
    - 不會有一個mld有兩組或以上的{name, birthday}
  - 請注意{name}  $\rightarrow$  {mld}不成立
    - 因為可能有同名同姓的會員
- ∞ 對於函數相依 $X \rightarrow Y$ ，我們稱 X決定Y 或Y 函數相依於X



## 函數相依 (Cont.)

☞ 根據定義，一個關聯的任何關聯鍵皆可決定其所有屬性



☞ 屬性和屬性間的函數相依是一種語意的關係，必須由人工仔細推敲來訂定

☞ 但有些函數相依仍然可以有規則可以推導而得



# 函數相依 (Cont.)

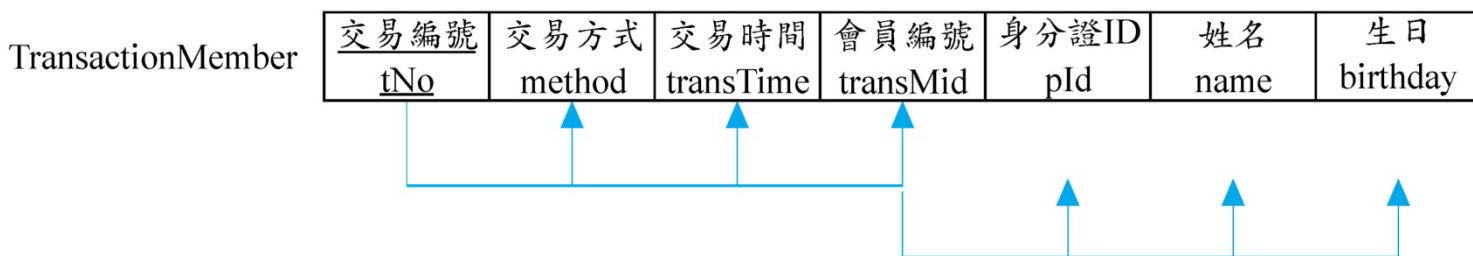
## 函數相依的推導規則

- $IR1 : X \supseteq Y \Rightarrow X \rightarrow Y$
- $IR2 : X \rightarrow Y \Rightarrow XZ \rightarrow YZ$
- $IR3 : X \rightarrow Y, Y \rightarrow Z \Rightarrow X \rightarrow Z$

一個屬性集合  $S$  的泛封閉集合  $S^+$ ：給定一些函數相依，我們可以推導出  $S$  屬性可以決定的所有屬性，稱為泛封閉集合  $S^+$



# 函數相依 (Cont.)



○ 圖8-7(a)

▲ 圖 8-7 (a) TransactionMember 關聯網目的函數相依

$\{tNo\}^+ = \{tNo, transMid, method, transTime, pId, name, birthday\}$

$\{transMid\}^+ = \{transMid, pId, name, birthday\}$



▲ 圖 8-7 (b) OrderProduct 關聯網目的函數相依

○ 圖8-7(b)

$\{pNo\}^+ = \{pNo, pName, unitPrice, category\}$

$\{mId, cartTime, pNo\}^+ = \{mId, cartTime, pNo, amount, pName, unitPrice, category\}$



## 第 8 章 . 正 規 化

15

### 8 - 4 基 本 的 正 規 化



# 正規化

∞ 定義：將一個關聯綱目轉化另一個滿足某種正規式的關聯綱目之過程

∞ 基本正規式

- 第一正規式
- 第二正規式
- 第三正規式

∞ 高等正規式

- BCNF
- 第四正規式
- 第五正規式





# 第一正規式

- ∞ R滿足第一正規式(1NF)如果
  - R的每個屬性都是簡單且單值
- ∞ 每個合法的關聯綱目都滿足第一正規式
- ∞ 如何分解多值屬性
  - 處理方式一：
    - ✓ 維持同樣的屬性，但主鍵要加上該多值屬性
  - 處理方式二：
    - ✓ 產生另一個關聯，此關聯包括原主鍵和該多值屬性，而主鍵則為所有屬性



# 多值屬性處理方式一

Transaction\_Products

交易編號 tNo	會員編號 transmId	交易方式 method	交易時間 transTime	商品 products
91100	a0911234	cart	2005-02-02:18:30:00	{b30999}
92666	c0927777	cart	2005-10-10:22:10:30	{d11222, d20777, v00111}
92333	c0927777	email	2005-10-15:09:00:00	{b51111}
91888	a091001	fax	2005-09-10:10:10:00	{b40555, d03333}
90111	b0905555	cart	2005-05-05:12:30:30	{v01888}
92555	b0922468	cart	2005-11-11:09:10:00	{b10234, b40555}

(a)

Transaction\_Product

交易編號 tNo	會員編號 transmId	交易方式 method	交易時間 transTime	商品 product
91100	a0911234	cart	2005-02-02:18:30:00	b30999
92666	c0927777	cart	2005-10-10:22:10:30	d11222
92666	c0927777	cart	2005-10-10:22:10:30	d20777
92666	c0927777	cart	2005-10-10:22:10:30	v00111
92333	c0927777	email	2005-10-15:09:00:00	b51111
91888	a0910001	fax	2005-09-10:10:10:00	b40555
91888	a0910001	fax	2005-09-10:10:10:00	d03333
90111	b0905555	cart	2005-05-05:12:30:30	v01888
92555	b0922468	cart	2005-11-11:09:10:00	b10234
92555	b0922468	cart	2005-11-11:09:10:00	b40555

(b)

▲ 圖 8-8 第一正規化的第一種方式：(a) Transaction\_Products 表格不滿足 1NF；(b) Transaction\_Product 關聯滿足 1NF



# 多值屬性處理方式二

**Product**

商品編號 pNo	商品名稱 pName	定價 unitPrice	作者 authors	
			作者名稱 name	職稱 title
b30999	資料庫理論與實務	500	Huang	Prof.
d11222	任賢齊專輯三	300	William	Mr.
b20666	OLAP進階	500	Sandra	Prof.
b10234	管理資訊系統概論	600	Lin	Prof.
b40555	系統分析理論與實務	550	Wu	Prof.
d20777	蔡依林專輯二	350	Jolin	Ms.
v01888	哈利波特：混血王子的背叛	450	J.K.	Mrs.
d03333	5566專輯	450	Jackey David Tom	Mr. Mr. Mr.
b51111	電子商務理論與實務	700	Lai Huang Lin	Dr. Prof. Prof.
v00111	英雄	400	NULL	NULL

**Product**

商品編號 pNo	商品名稱 pName	定價 unitPrice
b30999	資料庫理論與實務	500
d11222	任賢齊專輯三	300
b20666	OLAP進階	500
b10234	管理資訊系統概論	600
b40555	系統分析理論與實務	550
d20777	蔡依林專輯二	350
v01888	哈利波特：混血王子的背叛	450
d03333	5566專輯	450
b51111	電子商務理論與實務	700
v00111	英雄	400

**Author**

商品編號 pNo	作者名稱 name	職稱 title
b30999	Huang	Prof.
d11222	William	Mr.
b20666	Sandra	Prof.
b10234	Lin	Prof.
b40555	Wu	Prof.
d20777	Jolin	Ms.
v01888	J.K.	Mrs.
d03333	Jackey	Mr.
d03333	David	Mr.
d03333	Tom	Mr.
b51111	Lai	Dr.
b51111	Huang	Prof.
b51111	Lin	Prof.



## 第二正規式

第二正規式（簡稱2NF）是根據**完全函數相依**的概念

- 一個函數相依 $X \rightarrow Y$ ，如果我們可以從 $X$ 裡找出部分元素 $X'$ ，使得 $X' \rightarrow Y$ 仍然成立，則稱 $X \rightarrow Y$ 為一個部分函數相依
- 否則稱 $X \rightarrow Y$ 為完全函數相依
- 下圖OrderProduct關聯裡

### OrderProduct

<u>會員編號</u> mId	<u>購物車產生時間</u> carTime	<u>商品編號</u> pNo	訂購數量 amount	商品名稱 pName	定價 unitPrice	種類 catalog
			↑	↑	↑	↑

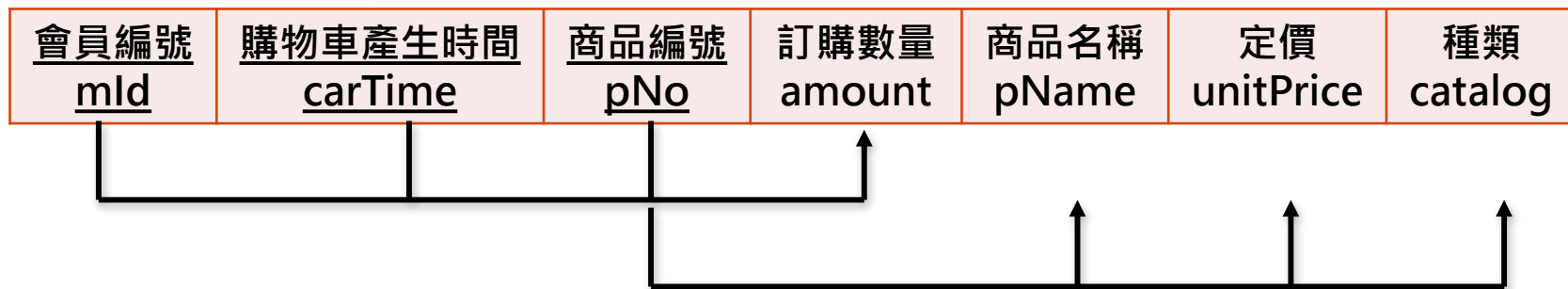
- $\{mId, carTime, pNo\} \rightarrow \{pName\}$ 是一個部分函數相依
  - 因為存在 $\{pNo\} \rightarrow \{pName\}$
- $\{mId, carTime, pNo\} \rightarrow \{amount\}$ 是一個完全函數相依



## 第二正規式 ( Cont. )

- ☞ 一個關聯網目R滿足第二正規式(2NF)如果
  - 對於R的每一個關聯鍵K，所有非鍵屬性都完全函數相依於K
- ☞ 下圖的OrderProduct關聯網目不滿足2NF，因為pName部分函數相依於主鍵
  - Because {pNo} → {pName}

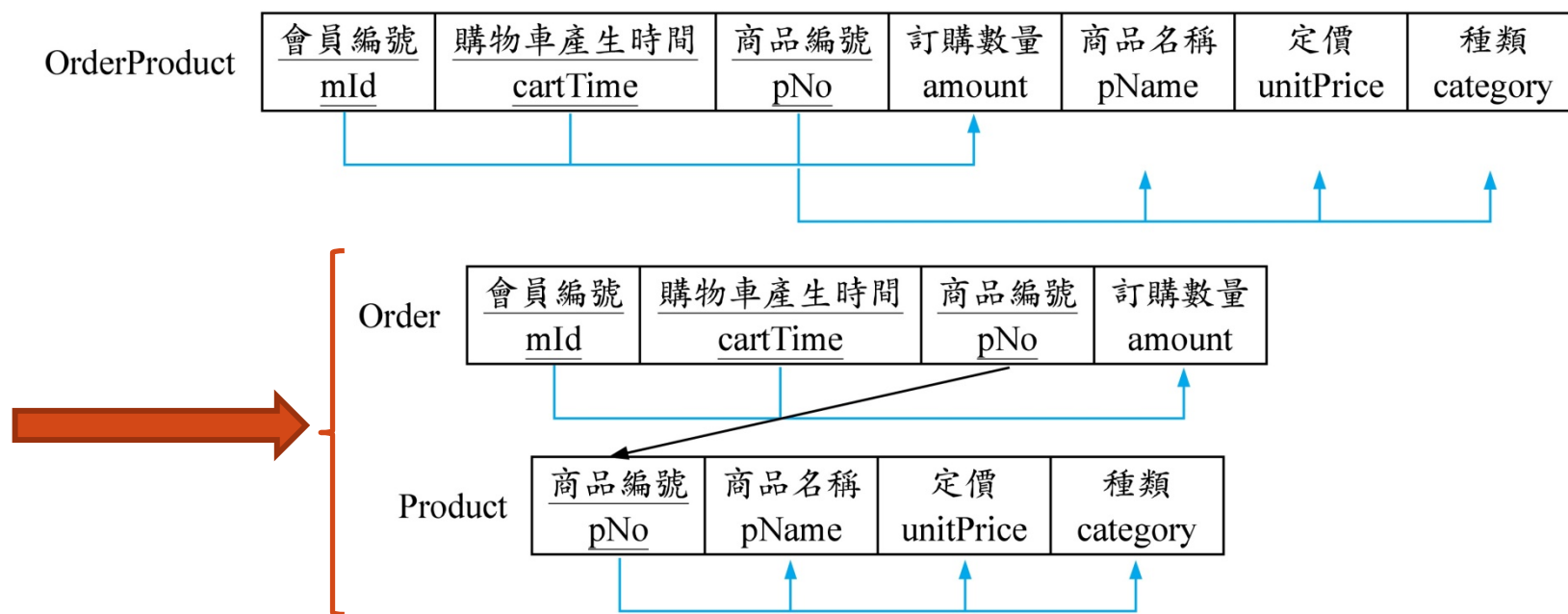
### OrderProduct





## 第二正規式 ( Cont. )

可以將OrderProduct關聯分解成Order和Product兩個關聯，便滿足2NF。如下：

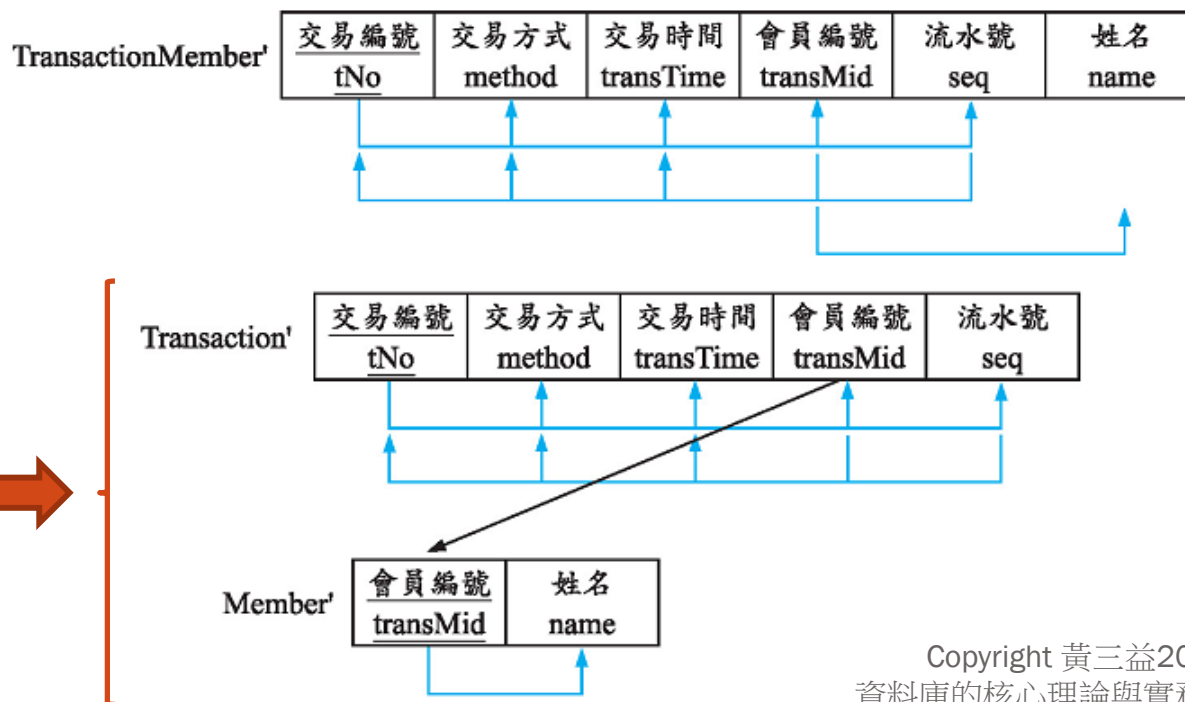




## 第二正規式 ( Cont. )

∞ 下圖的TransactionMember' 關聯有一個次要鍵  
( transMid, seq )

- {name}部分函數相依於{transMid, seq}，因此不滿足2NF
- 可分解成Transaction' 和Member' 兩個關聯，以滿足2NF



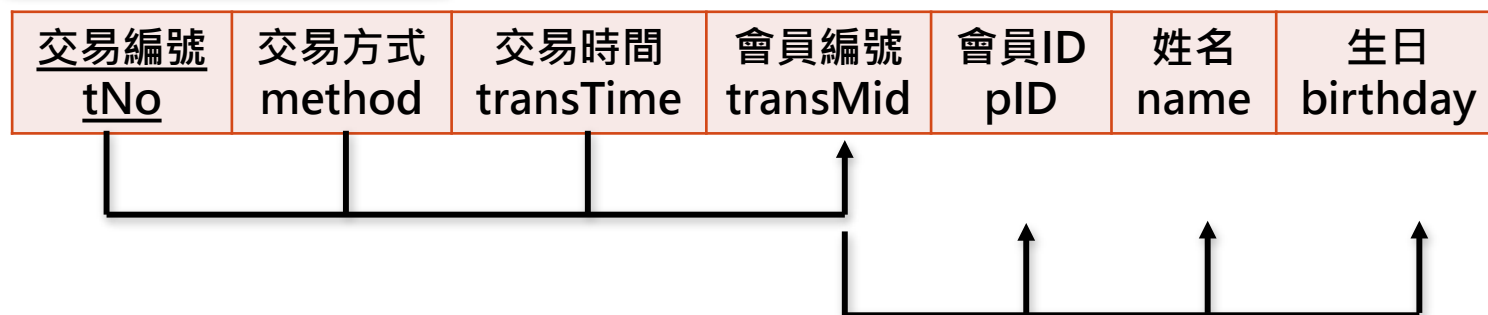


## 第三正規式

第三正規式（簡稱3NF）是根據遞移函數相依的概念

- 對於一個函數相依 $X \rightarrow Y$ ，如果存在著另外兩個函數相依： $X \rightarrow Z$ 和 $Z \rightarrow Y$ ，且 $Z$ 不為超級鍵，則稱 $X \rightarrow Y$ 為遞移函數相依
- 下圖TransactionMember關聯

TransactionMember



- $\{tNo\} \rightarrow \{name\}$ 是一個遞移函數相依
  - 因為 $\{tNo\} \rightarrow \{transMid\}$ 且 $\{transMid\} \rightarrow \{name\}$ ，且 $transMid$ 不為超級鍵





## 第三正規式 (Cont.)

☞ 以下的關聯綱目裡，雖然 $\{mld\} \rightarrow \{pld\}$ 且 $\{pld\} \rightarrow \{name\}$ ，但 $\{mld\} \rightarrow \{name\}$ 不為遞移相依，因為 $\{pld\}$ 為超級鍵。





## 第三正規式 (Cont.)

- ☞ 一個關聯網目R滿足3NF，如果
- R滿足2NF。
  - 對於每一個關聯鍵K，R的每一個非鍵屬性都不可遞移函數相依於K
  - TransactionMember不滿足3NF
    - $\{tNo\} \rightarrow \{name\}$  是一個遞移函數相依，而tNo為主鍵

TransactionMember

交易編號 tNo	交易方式 method	交易時間 transTime	會員編號 transMid	會員ID pID	姓名 name	生日 birthday

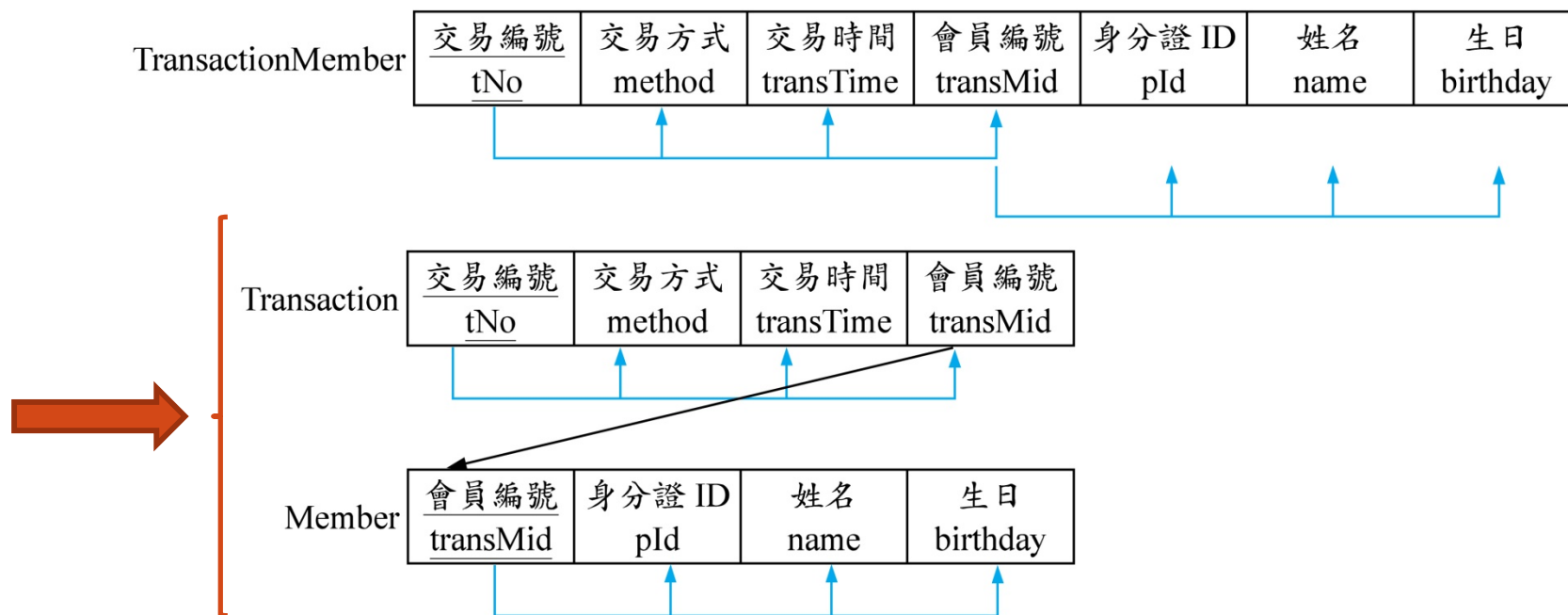
Diagram illustrating functional dependencies for TransactionMember:

- Primary key: tNo (indicated by a line from tNo to transMid).
- Transitive dependency: tNo  $\rightarrow$  transMid  $\rightarrow$  {pID, name, birthday} (indicated by arrows from transMid to pID, name, and birthday).



## 第三正規式 (Cont.)

可分解如下：



▲ 圖 8-11 第三正規化範例，TransactionMember 關聯網目的第三正規化



## 第 8 章 . 正 規 化

28

### 8 - 5 進 階 正 規 化



# Boyce-Codd 正規式

- ∞ 一個關聯網目R滿足BCNF，如果
  - 對於R的每一個函數相依 $A \rightarrow B$ ，**A都是超級鍵** (A具tuple區別性)
  - 發生條件：R有**複合候選鍵**，且複合候選鍵之間擁有重疊屬性，也就是說至少擁有一個相同屬性。
  
- ∞ 滿足BCNF的關聯網目必然也滿足2NF和3NF
  - 若R不滿足2NF，則對於一個關聯鍵X來說，存在著部分函數相依 $X \rightarrow Y$ ，也就是 $X' \rightarrow Y$ ， $X' \subset X$ 。所以Y可由非超級鍵 $X'$ 所決定 (e.g., OrderProduct) (**Y函數相依於多重主鍵X之部分主鍵 $X'$** )
  - 若R不滿足3NF，則存在著非鍵屬性Y是遞移函數相依於關聯鍵X。也就是有 $X \rightarrow Z$ 和 $Z \rightarrow Y$ ，且Z不是超級鍵。所以Y可由非超級鍵Z所決定 (e.g., TransactionMember) (**Y遞迴相依於非主鍵X之屬性Z**)



# Boyce-Codd 正規式 ( Cont. )

- 有些關聯網目滿足3NF但卻不滿足BCNF
- 複合候選鍵(tNo, pNo), (invNo, pNo)，且複合候選鍵之間擁有重疊屬性(pNo)，也就是說至少有一個相同屬性(pNo)。

Transaction

<u>交易編號</u> <u>tNo</u>	<u>產品編號</u> <u>pNo</u>	數量 amount	單價 unitPrice	報價單編號 invNo
c01	<u>bk001</u>	3	500	A
<u>c01</u>	<u>bk200</u>	2	400	D
<u>c03</u>	<u>cd100</u>	5	350	C
<u>c03</u>	<u>vcd888</u>	3	450	C
<u>f04</u>	<u>bk001</u>	1	250	B



# Boyce-Codd 正 規 式 ( Cont. )

Transaction

<u>交易編號</u> tNo	<u>產品編號</u> pNo	數量 amount	單價 unitPrice	報價單編號 invNo
c01	<u>bk001</u>	3	500	A
<u>c01</u>	<u>bk200</u>	2	400	D
<u>c03</u>	<u>cd100</u>	5	350	C
<u>c03</u>	<u>vcd888</u>	3	450	C
<u>f04</u>	<u>bk001</u>	1	250	B

Transaction2

<u>報價單編號</u> invNo	交易編號 tNo
A	c01
D	c01
C	c03
B	f04

Transaction1

<u>商品編號</u> pNo	<u>報價單編號</u> invNo	數量 amount	單價 unitPrice
bk001	A	3	500
bk200	D	2	400
cd100	C	5	350
vcd888	C	3	450
bk001	B	1	250



# Boyce-Codd 正規式 ( Cont. )

- ∞ 上頁Transaction關聯滿足3NF ( 但不滿足BCNF)
- ∞ 上頁Transaction1和Transaction2滿足BCNF
- ∞ 何者較好 ?
  - 考慮所有函數相依
  - 1. ( tNo, pNo ) → amount
  - 2. ( tNo, pNo ) → salePrice
  - 3. ( tNo, pNo ) → invNo
  - 4. invNo → tNo
  - 5. ( invNo, pNo ) → amount
  - 6. ( invNo, pNo ) → salePrice





# Boyce-Codd 正規式 ( Cont. )

- ☞ 只有Transaction關聯時，設定主鍵可決定
  - ( tNo, pNo )  $\rightarrow$  amount
  - ( tNo, pNo )  $\rightarrow$  salePrice
  - ( tNo, pNo )  $\rightarrow$  invNo
- ☞ 分解成Transaction1和Transaction2關聯時，設定主鍵可決定
  - ( invNo, pNo )  $\rightarrow$  amount
  - ( invNo, pNo )  $\rightarrow$  salePrice
  - invNo  $\rightarrow$  tNo
- ☞ 當分解成Transaction1和Transaction2關聯時，若仍想維持函數相依 ( tNo, pNo )  $\rightarrow$  amount ，則需做跨關聯的檢查
- ☞ 保留函數相依的分解：關聯網目分解後各函數相依仍可藉由檢查單一關聯來確定
- ☞ BCNF常無法保留所有函數相依



## 第四正規式



第四正規式（簡稱4NF）是源於多值相依的概念

- 多值相依 $X \twoheadrightarrow Y$ ：給定一個X的屬性值，便有一組Y的屬性值
  - $pNo \twoheadrightarrow author$
  - $pNo \twoheadrightarrow transaction$

<u>商品編號</u> pNo	<u>創作者</u> author	<u>交易</u> transaction
c01	ABC	50
c01	XYZ	42
c01	ABC	42
c01	XYZ	50
m05	DEF	51
m05	XYZ	33
m05	DEF	33
m05	XYZ	51



## 第四正規式 (Cont.)

ProductAuthor

<u>商品編號</u> pNo	<u>創作者</u> author
c01	ABC
c01	XYZ
m05	DEF
m05	XYZ

ProductTransaction

<u>商品編號</u> pNo	<u>交易</u> transaction
c01	50
c01	42
m05	51
m05	33

- 在ProductAuthor關聯裡的多值相依pNo→author和ProductTransaction關聯裡的多值相依pNo→transaction稱為微不足道的多值相依
- 一個關聯綱目R滿足4NF，如果
  - 每一個R的非微不足道的多值相依其左方都是超級鍵
- 滿足4NF也必定滿足BCNF



## 第五正規式

- ∞ 一個關聯網目R滿足5 NF：
  - R無法再被分解成數個關聯R1, R2, ..., Rk，使得 $R1 * R2 * ... * Rk = R$ ，或
  - R可以被分解成數個關聯R1, R2, ..., Rk，使得 $R1 * R2 * ... * Rk = R$ ，但是每一個Ri,  $1 \leq i \leq k$ ，都是R的超級鍵。
  - 多重主鍵R之部分主鍵Ri
- ∞ 5NF限制條件的發現和檢查不易，也因此**在實作上通常不予考慮**



# 包含相依

- 考慮兩個關聯R和S的兩個屬性集合X和Y，如果R(X)的任一屬性值必然都存在於S(Y)裡，則稱R(X)包含相依於S(Y)，以符號 $R(X) \subseteq S(Y)$ 表示。
- 考慮Browse(mld, pNo, bTime)和Purchase(mld, pNo, pTime)，若「會員必須瀏覽過商品才能購買該商品」則 $Purchase(mld, pNo) \subseteq Browse(mld, pNo)$
- 在實務上，包含相依可透過SQL的ASSERTION (Oracle) 或TRIGGER (MSSQL, Oracle...)來設定 (語意限制)



本章節講述到此結束..謝謝！

∞ 38 ∞