

第 12 章

資料正規化



本章提要

- 何謂正規化
- 正規化的步驟
- 欄位相依的用意
- 正規化的另類思考



實體 - 關係圖與正規化的關係

- 資料表的**正規化理論**
(Normalization Theory)
其實是一套資料表分割的法則。
- 正規化的**目的**：
 - ✓ 避免資料重複或相互矛盾的情形,
 - ✓ 使資料庫在使用時能更有效率、更容易維護。



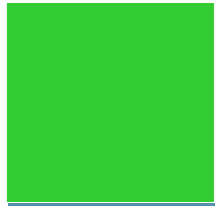
資料庫的正規化

- 資料庫的正規化共可分為 **1NF (Normal Form)**、**2NF**、**3NF**、**BCNF (Boyce-Codd Normal Form)**、**4NF**、**5NF** 等多個階段, 不過對於一般資料庫設計來說, 通常只需要執行到 **BCNF** 即可, 而其他更高階的正規化只有在特殊的情況下才用得到。
- 在資料表正規化的過程 (**1NF** 到 **BCNF**) 中, 每個階段都是以**欄位的相依性**, 做為分割資料表的依據之一。

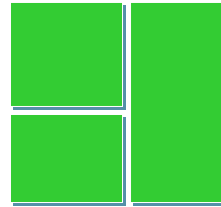


資料庫的正規化

單一的大資料表



多個獨立但相關
的小資料表



正規化的4個步驟

→ 1NF → 2NF → 3NF → BCNF →



5

12-2

欄位相依

- 在一個資料表中, 若乙欄位的值必須搭配甲欄位才有意義, 則我們說『乙欄位相依於甲欄位』。舉例來說, 一個員工資料表如下:

*員工編號	姓名	地址
101	陳鐵雄	...
102	王大明	...

員工編號欄為主鍵, 作為唯一辨識該筆記錄的欄位。

姓名欄必須相依於員工編號欄, 對此資料表來說, 姓名欄才有意義。

否則同名同姓的陳鐵雄, 都可對應到該筆記錄;

同理, 地址欄必須相依於員工編號欄, 才有意義。



6

欄位相依

* 課程編號	* 學 號	成 績
M102	903123	89
A205	903145	75

欄位相依

成績欄本身如果單獨存在是沒有意義的，因為不曉得是哪一門課、哪一位學生的成績。不過，當成績欄相依於課程編號及學號二欄時，就可以了解某個學生修某堂課的成績，這樣的成績資料才有意義。

7

12-3

第一階正規化(1NF)

■ 正規化基本概念

- ✓ 正規化的過程是循序漸進的，資料表必須在滿足第一階正規化的條件之下，才能進行第二階正規化。
- ✓ 也就是說，第二階正規化必須建立在符合第一階正規化的資料表上，依此類推。
- ✓ 而第一階正規化正是所有正規化的基礎。

8

第一階正規化的規則

- 如果資料表符合以下條件, 我們就說這個資料表符合**第一階正規化的形式 (First Normal Form, 簡稱 1NF)** :
 - ✓ 資料表中有主鍵, 而其他所有的欄位都相依於主鍵。
 - ✓ 每個欄位中都只有儲存單一值,
例如: 姓名欄位中不能存放 2 個人的姓名。
 - ✓ 資料表中沒有意義相同的多個欄位,
例如: 姓名1、姓名2...等重複的欄位。



9

未符合 1NF 資料表的缺點

選課資料表

課程編號	課程名稱	教師編號	教師姓名	學號	學生姓名	成績
CS101	資料庫系統	1001	曾曉明	91010	陳鐵雄	70
				91013	王大明	90
				91020	李珍珍	88
CS103	影像處理	1015	郝天才	91033	趙阿丁	86
				91045	鄭阿龍	93

- ✓ 學號、學生姓名及成績欄的長度無法確定
- ✓ 降低存取資料的效率, ex: 某位學生的成績



10

建構 1NF 資料表的方法

- 對於不具第一階正規化形式的資料表, 我們可將重複的資料項分別儲存到不同的記錄中, 並加上適當的主鍵:

*課程編號	*學號	課程名稱	教師編號	教師姓名	學生姓名	成績
CS101	91010	資料庫系統	1001	曾聰明	陳鐵雄	70
CS101	91013	資料庫系統	1001	曾聰明	王大明	90
CS101	91020	資料庫系統	1001	曾聰明	李珍珍	88
CS103	91033	影像處理	1015	郝天才	趙阿丁	86
CS103	91045	影像處理	1015	郝天才	鄭阿龍	93

11

12-4

第二階正規化 (2NF)

- 在執行符合一階正規化的資料表時, 應該會發覺: 我們輸入了許多重複的資料。
- 如此, 不但浪費儲存的空間, 更容易造成新增、刪除或更新資料時的異常狀況。
- 所以, 我們必須進行第二階正規化, 來消除這些問題。

12

第二階正規化的規則

- 如果資料表符合以下的條件, 我們說這個資料表符合第二階正規化的形式 (Second Normal Form, 簡稱 2NF)
 - ✓ 符合 1NF 的格式。
 - ✓ 各欄位與主鍵間沒有部分相依的關係。



13

第二階正規化的規則

選課資料表

只和學號有相依性

*課程編號	*學號	課程名稱	教師編號	教師姓名	學生姓名	成績
CS101	91010	資料庫系統	1001	曾聰明	陳鐵雄	70
CS101	91013	資料庫系統	1001	曾聰明	王大明	90
CS101	91020	資料庫系統	1001	曾聰明	李珍珍	88
CS103	91033	影像處理	1015	郝天才	趙阿丁	86
CS103	91045	影像處理	1015	郝天才	鄭阿龍	93

只和課程編號有相依性



14

部分相依的問題

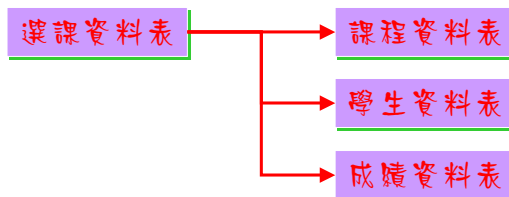
- 新增資料時：若有一個新來的轉學生『吳小安』，但還沒有選修任何課程，那麼它的資料將無法輸入。
- 更改資料時：當我們想要將課程名稱『資料庫系統』更改為『資料庫管理』時，必須搜尋整個資料庫並一一更改，非常沒有效率。
- 刪除資料時：由於陳鐵雄只修了『CS101』一門課，如果將該筆記錄刪除，那麼陳鐵雄的資料也就跟著消失了。



15

建構 2NF 資料表的方法

- 要除去資料表中的部分相依性，只需將部份相依的欄位分割成另外的資料表即可。
 - ✓ 例如：將修課資料表分割成 3 個較小的資料表 (加 "*" 號的欄位為主鍵)：



16

建構 2NF 資料表的方法

成績資料表

*課程編號	*學號	成績
CS101	91010	70
CS101	91013	90
CS101	91020	88
CS103	91033	86
CS103	91045	93

學生資料表

*學號	學生姓名
91010	陳鐵雄
91013	王大明
91020	李珍珍
91033	趙阿丁
91045	鄭阿龍

課程資料表

*課程編號	課程名稱	教師編號	教師姓名
CS101	資料庫系統	1001	曾曉明
CS103	影像處理	1015	郝天才

17

12-5

第三階正規化 (3NF)

- 經過了第二階正規化後的資料表, 其實還存在一些問題：
 - ✓ 在課程資料表中, 如果新來了一位教師『王小萍』, 在尚未安排他教授的課程之前, 我們無法輸入該教師的資料。
 - ✓ 若要刪除影像處理的課程, 勢必會將教授該門課程的教師一併刪除。
 - ✓ 若要更改影像處理課程的名稱, 則必須同時更改多筆記錄, 造成不便。(同一門課會有多筆記錄, 因為有許多老師都開這門課)

18

第三階正規化的規則

- 如果資料表符合以下條件, 我們就說這個資料表符合第三階正規化的形式
(Third Normal Form, 簡稱 3NF)
 - ✓ 符合 2NF 的格式。
 - ✓ 各欄位與主鍵間沒有間接相依的關係。



19

第三階正規化的規則

課程資料表

*課程編號	課程名稱	教師編號	教師姓名
CS101	資料庫原理	1001	曾曉明
CS103	影像處理	1015	郝天才

課程 (*課程編號, 課程名稱, 教師編號, 教師姓名)

與主鍵無關的相依



20

建構 3NF 資料表的方法

- 要除去資料表中的間接相依性, 其方法和除去部分相依性完全相同。

✓ 例如: 課程資料表可再分割成兩個資料表:

課程資料表

*課程編號	課程名稱	教師編號
CS101	資料庫系統	1001
CS103	影像處理	1015

教師資料表

教師編號	教師姓名
1001	曾聰明
1015	郝天才



21

與直覺式的分割技巧做比較

正規化	功能相同的直覺式分割法
1NF: 有主鍵 欄位中只有一個單一值 沒意義相同的重複欄位	無
2NF: 去除『部分相依性』	分割『欄位一在重複』的欄位
3NF: 去除『間接相依性』	分割『與主鍵無關』的欄位



22

Boyce-Codd 正規化 (BCNF)

- 對於大部分資料庫來說, 通常只需要執行到第三階段的正規化即足夠了。
- 如果資料表的主鍵是由多個欄位組成的, 則必須再執行 **Boyce-Codd** 正規化。



Boyce-Codd 正規化的規則

- 如果資料表的主鍵只由**單一欄位**組成, 則符合第三階正規化的資料表, 亦符合 **Boyce-Codd** 正規化。
- 若資料表的主鍵由多個欄位組成, 則資料表只要符合以下條件, 我們就說這個資料表符合 **Boyce-Codd 正規化的形式** (**Boyce-Codd Normal Form**, 簡稱 **BCNF**) :
 - ✓ 符合 2NF 的格式。
 - ✓ 各欄位與主鍵沒有**間接相依**的關係。
 - ✓ 主鍵中的各欄位不可以相依於其他非主鍵的欄位。



檢驗成績資料表是否滿足 BCNF 規範

- 我們利用 Boyce-Codd 正規化的條件，來檢驗主鍵由多個欄位組成的成績資料表。

成績資料表

*課程編號	*學號	成績
CS101	91010	70
CS101	91013	90
CS101	91020	88
CS103	91033	86
CS103	91045	93

所有欄位都相依於主鍵,主鍵中的欄位亦無相依於其他非主鍵欄位

25

經過正規化產生的資料表

課程資料表

*課程編號	課程名稱	教師編號
CS101	資料庫系統	1001
CS103	影像處理	1015

教師資料表

教師編號	教師姓名
1001	曾曉明
1015	郝天才

成績資料表

*課程編號	*學號	成績
CS101	91010	70
CS101	91013	90
CS101	91020	88
CS103	91033	86
CS103	91045	93

學生資料表

*學號	學生姓名
91010	陳鐵雄
91013	王大明
91020	李珍珍
91033	趙阿丁
91045	鄭阿龍

26

正規化的另類思考

■ 不必要的分割

- ✓ 正規化的工作有時候不必做的太徹底，因為過多的資料表可能會降低系統執行的效能。
- ✓ 以學生資料表來看：
郵遞區號相依於『鄉鎮市區』。
因此，可以再分割，才符合正規化的要求。

學 生	
* 學 號	
性 名	
年 齡	
電 話	
鄉鎮市(區)	
街牌號碼	
郵遞區號	

27

正規化的另類思考

學 生	
* 學 號	
性 名	
年 齡	
電 話	
鄉鎮市(區)編號	
街牌號碼	

郵遞區號對應	
* 鄉鎮市(區)編號	
鄉鎮市(區)	
郵遞區號	

但實際作業上，我們一定是使用整筆的地址資料，如果將她們分割了，那麼每次查詢地址時，都要透過資料表的關聯，從郵遞區號對應資料表中，取得郵遞區號欄位。當資料增加到一定程度時，就會明顯感受到系統效能降低的事實，實非明智之舉。

28

正規化的另類思考

■ 人工的分割

- ✓ 有時為了增加資料處理的效率, 我們可以將已經符合 BCNF 的資料表再做分割。
- ✓ 例如: 一個資料表擁有很多欄位, 而其中又有
很多欄位根本很少用到。
那麼, 我們就可以將這些很少用到的欄位分離出來, 存放到另外一個資料表中。

