

資料庫正規化

正規化的概念

何謂**正規化**(Normalization)？就是結構化分析與設計中，**建構「資料模式」**所運用的一個技術，其目的是為了**降低資料的「重覆性」**與**避免「更新異常」**的情況發生。

因此，就必須將整個資料表中**重複性的資料**剔除，否則在關聯表中會造成**新增異常、刪除異常、修改異常**的狀況發生。

一般而言，**正規化的精神**就是讓資料庫中**重複的欄位資料減到最少**，並且能**快速的找到資料**，以**提高關聯性資料庫的效能**。

【目的】

1. **降低資料重覆性**(Data Redundancy)。
2. **避免資料更新異常**(Anomalies)。

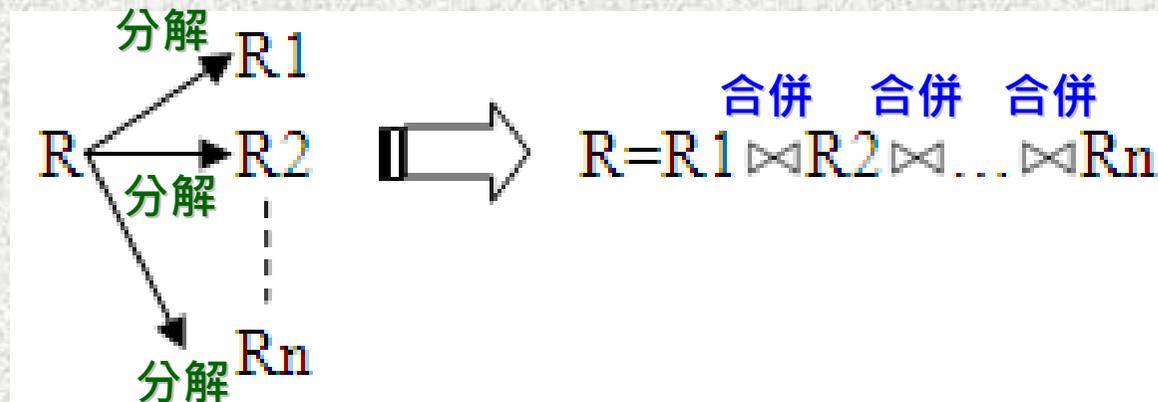
資料庫正規化(Normalization)

【定義】

是指將原先關聯(表格)的所有資訊，在「分解」之後，仍能由數個新關聯(表格)中經過「合併」得到相同的資訊。即所謂的「無損失分解 (Lossless decomposition)」的觀念。

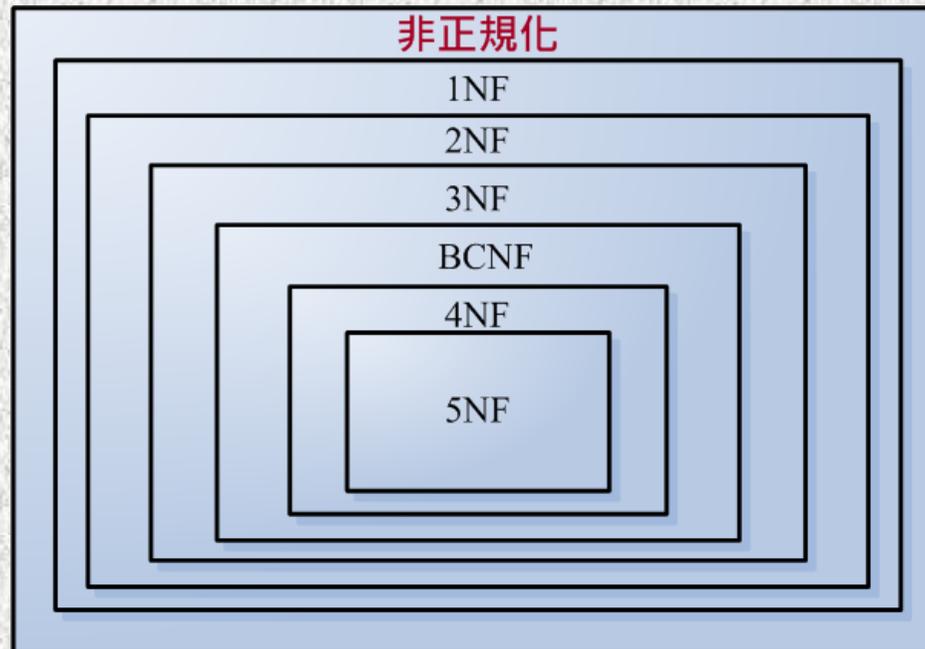
【無損失分解觀念】

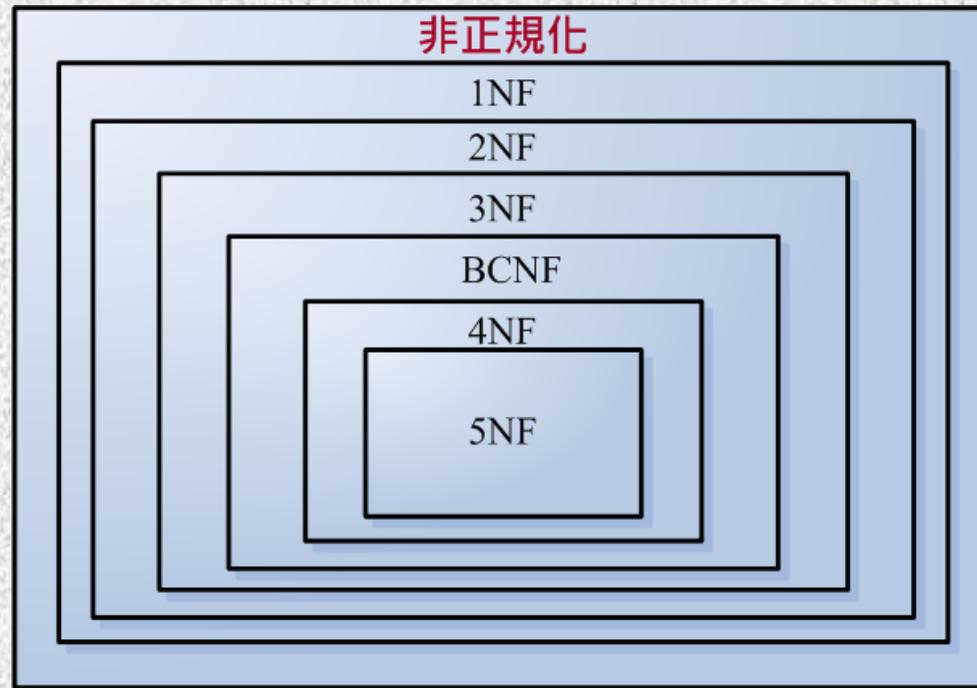
當關聯表R被「分解」成數個關聯表R1, R2, ..., Rn時，則可以再透過「合併」 $R1 \bowtie R2 \bowtie \dots \bowtie Rn$ 得到相同的資訊R。如下圖所示。



正規化的規則

資料庫在正規化時會有一些規則，並且**每條規則**都稱為「**正規形式**」。如果**符合第一條規則**，則資料庫就稱為「**第一正規化形式(1NF)**」。如果**符合前二條規則**，則資料庫就被視為屬於「**第二正規化形式(2NF)**」。雖然資料庫的**正規化最多**可以進行到**第五正規化形式**，但是在**實務上**，**BCNF**被視為大部分應用程式所需的**最高階正規形式**。

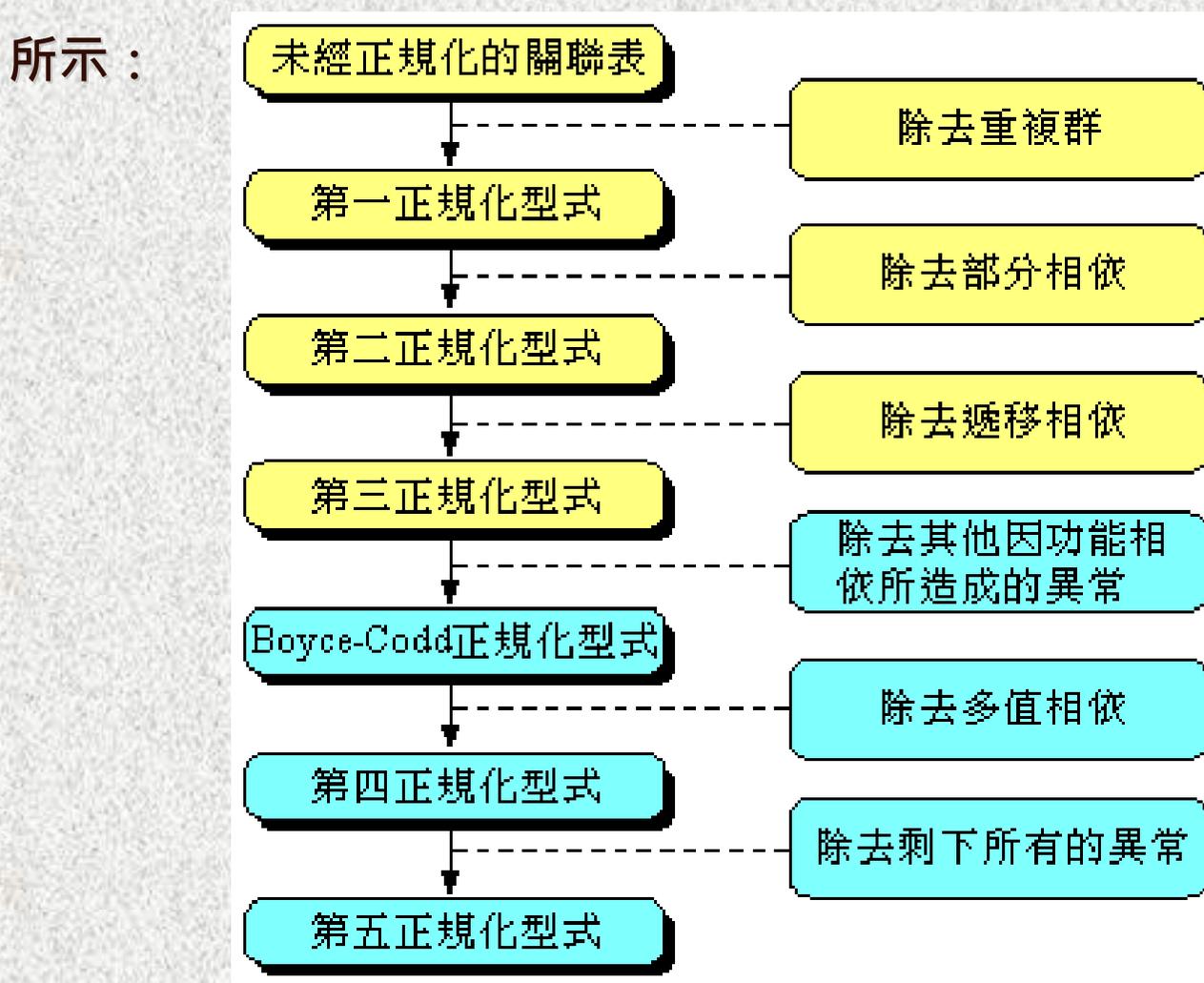




從上圖中，我們可以清楚得知，正規化是循序漸進的過程，亦即資料表必須滿足第一正規化的條件之後，才能進行第二正規化。換言之，第二正規化必須建立在符合第一正規化的資料表上，依此類推。

➤ 正規化步驟

在資料表正規化的過程 (1NF 到 BCNF) 中, 每一個階段都是以欄位的「相依性」, 做為分割資料表的依據之一。其完整的正規化步驟如下圖所示：



第一正規化(1NF)

【定義】

是指在資料表中的所有記錄之屬性內含值都是基元值(Atomic Value)。

亦即無重覆項目群。

【實例】假設現在有一份某某科技大學的學生選課資料表，如下表所示：**學生選課資料表**

某某科技大學【學生選課資料表】

=====						
學號：001		姓名：李碩安			性別：男	
課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
C001	程式語言	4	必	74	T001	李安
C002	網頁設計	3	選	93	T002	張三
學號：002		姓名：李碩崑			性別：男	
課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
C002	網頁設計	3	選	63	T002	張三
C003	計 概	2	必	82	T003	李四
C005	網路教學	4	選	94	T005	王五

我們可以將原始資料利用二維表格來儲存

學生選課資料表

某某科技大學【學生選課資料表】						
學號：001		姓名：李碩安			性別：男	
課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
C001	程式語言	4	必	74	T001	李安
C002	網頁設計	3	選	93	T002	張三
學號：002		姓名：李碩崑			性別：男	
課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
C002	網頁設計	3	選	63	T002	張三
C003	計 概	2	必	82	T003	李四
C005	網路教學	4	選	94	T005	王五

未正規化的資料表：學生選課資料報表

二維表格來儲存

學號	姓名	性別	課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
001	李碩安	男	C001	程式語言	4	必	74	T001	李安
			C002	網頁設計	3	選	93	T002	張三
002	李碩崑	男	C002	網頁設計	3	選	63	T002	張三
			C003	計 概	2	必	82	T003	李四
			C005	網路教學	4	選	94	T005	王五

因此，我們發現有許多屬性的內含值都具有二個或二個以上的值(亦稱為重複資料項目)，其原因：尚未進行第一正規化。

未正規化的資料表：學生選課資料報表

學號	姓名	性別	課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
001	李碩安	男	C001	程式語言	4	必	74	T001	李安
			C002	網頁設計	3	選	93	T002	張三
002	李碩歲	男	C002	網頁設計	3	選	63	T002	張三
			C003	計 概	2	必	82	T003	李四
			C005	網路教學	4	選	94	T005	王五

重複資料項目

■未符合 1NF 資料表的「缺點」

以上資料表中的『課程代碼』、『課程名稱』、『學分數』、『必選修』、『成績』、『老師編號』及『老師姓名』欄位的長度無法確定，因為學生要選修多少門課程，無法事先得知(李碩安同學選了2門，李碩歲同學選了3門)，因此，必須要預留很大的空間給這七個欄位，如此反而造成儲存空間的浪費。

➤ 第一正規化的規則

1. 每一個欄位只能有一個基元值(Atomic)即單一值。

例如：課程名稱欄位中不能存入兩科或兩科以上的課程名稱。

2. 沒有任何兩筆以上的資料是完全重覆。

3. 資料表中有主鍵, 而其他所有的欄位都相依於「主鍵」。

例如1：姓名與性別欄位都相依於「學號」欄位。

例如2：課程名稱、學分數、必選修、老師編號及老師姓名相依於「課程代碼」欄位。

例如3：「成績」欄位相依於「學號」與「課程代碼」欄位。

➤ 第一正規化的作法：

【作法】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。

未經正規化前的學生選課表

重複資料項目

學號	姓名	性別	課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
001	李碩安	男	C001	程式語言	4	必修	74	T001	李安
			C002	網頁設計	3	選	93	T002	張三
002	李碩歲	男	C002	網頁設計	3	選	63	T002	張三
			C003	計 概	2	必修	82	T003	李四
			C005	網路教學	4	選	94	T005	王五

步驟二：將重複資料項分別儲存到不同的記錄中, 並加上適當的主鍵

未經正規化前的學生選課表

學號	姓名	性別	課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
001	李碩安	男	C001	程式語言	4	必	74	T001	李安
			C002	網頁設計	3	選	93	T002	張三
002	李碩歲	男	C002	網頁設計	3	選	63	T002	張三
			C003	計 概	2	必	82	T003	李四
			C005	網路教學	4	選	94	T005	王五

重複資料項

經過正規化後的學生選課表(1NF)

儲存到不同的記錄

學號	姓名	性別	課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
001	李碩安	男	C001	程式語言	4	必	74	T001	李安
001	李碩安	男	C002	網頁設計	3	選	93	T002	張三
002	李碩歲	男	C002	網頁設計	3	選	63	T002	張三
002	李碩歲	男	C003	計 概	2	必	82	T003	李四
002	李碩歲	男	C005	網路教學	4	選	94	T005	王五

經過正規化後的學生選課表(1NF)

The diagram shows a table with 10 columns. The first two columns, '學號' and '課程代碼', are highlighted with red boxes. A blue arrow points from the '學號' column to the '成績' column. Another blue arrow points from the '課程代碼' column to the '成績' column. The '成績' column is also highlighted with a blue box.

學號	姓名	性別	課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
001	李碩安	男	C001	程式語言	4	必	74	T001	李安
001	李碩安	男	C002	網頁設計	3	選	93	T002	張三
002	李碩歲	男	C002	網頁設計	3	選	63	T002	張三
002	李碩歲	男	C003	計 概	2	必	82	T003	李四
002	李碩歲	男	C005	網路教學	4	選	94	T005	王五

在經由第一正規化之後，使得每一個欄位內只能有一個資料(基元值)。雖然增加了許多記錄，但每一個欄位的「長度」及「數目」都可以固定，而且我們可用「課程代碼」欄位加上「學號」欄位當作主鍵，使得在查詢某學生修某課程的「成績」時，就非常方便而快速了。

第二正規化(2NF)

在完成了第一正規化之後，是否發現在資料表中產生許多重複的資料。如此，不但浪費儲存的空間，更容易造成新增、修改及刪除資料時的異常狀況，說明如下。

(1) 新增異常檢查(Insert Anomaly)

記錄	學號	姓名	性別	課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
#1	001	李碩安	男	C001	程式語言	4	必	74	T001	李安
#2	001	李碩安	男	C002	網頁設計	3	選	93	T002	張三
#3	002	李碩崴	男	C002	網頁設計	3	選	63	T002	張三
#4	002	李碩崴	男	C003	計 概	2	必	82	T003	李四
#5	002	李碩崴	男	C005	網路教學	4	選	94	T005	王五

×
無法
新增

例如：鍵入#6 筆記錄，如下所示：

記錄	學號	姓名	性別	課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
#6	NULL			C004	系統分析			NULL		

無法先新增課程資料，如「課程代碼」及「課程名稱」，要等選課之後，才能新增。

原因：以上的新增動作違反「實體完整性規則」，因為，主鍵或複合主鍵不可以為空值 NULL。

(2)修改異常檢查(Update Anomaly)

記錄	學號	姓名	性別	課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
#1	001	李碩安	男	C001	程式語言	4	必	74	T001	李安
#2	001	李碩安	男	C002	網頁設計	3	選	93	T002	張二
#3	002	李碩崴	男	C002	網頁設計	3	選	63	T002	張三
#4	002	李碩崴	男	C003	計 概	2	必	82	T003	李四
#5	002	李碩崴	男	C005	網路教學	4	選	94	T005	王五

「網頁設計」課程重覆多次，因此，修改「網頁設計」課程的成績時，可能有些記錄未修改到，造成資料的不一致現象。

例如：有選「網頁設計」課程的同學之成績各加5分，可能會有些同學有加分，而有些同學卻沒有加分，導致資料不一致的情況。

(3)刪除異常檢查>Delete Anomaly)

記錄	學號	姓名	性別	課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
#1	001	李碩安	男	C001	程式語言	4	必	74	T001	李安
#2	001	李碩安	男	C002	網頁設計	3	選	93	T002	張三
#3	002	李碩崴	男	C002	網頁設計	3	選	63	T002	張三
#4	002	李碩崴	男	C003	計 概	2	必	82	T003	李四
#5	002	李碩崴	男	C005	網路教學	4	選	94	T005	王五

當刪除#4學生的記錄時，同時也會刪除課程名稱、學分數及相關的資料。

所以導致「計概」課程的2學分數也同時被刪除了。

綜合上述的三種異常現象，所以，我們必須進行「第二階正規化」，來消除這些問題。

➤ 第二正規化的規則

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

☑每一非鍵屬性(如：姓名、性別...)必須「完全相依」於主鍵(學號)；即不可「部分功能相依」於主鍵。

換言之，「部分功能相依」只有當「主鍵」是由「多個欄位」組成時才會發生(亦即複合主鍵)，也就是當某些欄位只與「主鍵中的部分欄位」有「相依性」，而與另一部分的欄位沒有相依性。

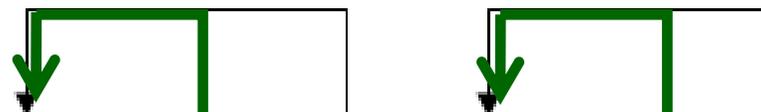
➤ 第二正規化的作法

☑ 分割資料表；亦即將「部分功能相依」的欄位「分割」出去，

再另外組成「新的資料表」。其步驟如下：

步驟一：檢查是否存在「部分功能相依」

「姓名」只相依於「學號」 「課程名稱」只相依於「課程代碼」



記錄	學號	姓名	性別	課程代碼	課程名稱	學分數	必選修	成績	老師編號	老師姓名
#1	001	李碩安	男	C001	程式語言	4	必	74	T001	李安
#2	001	李碩安	男	C002	網頁設計	3	選	93	T002	張三
#3	002	李碩歲	男	C002	網頁設計	3	選	63	T002	張三
#4	002	李碩歲	男	C003	計 概	2	必	82	T003	李四
#5	002	李碩歲	男	C005	網路教學	4	選	94	T005	王五

在上面的資料表中，主鍵是由「學號+課程代碼」兩個欄位所組成，但「姓名」和「性別」只與「學號」有「相依性」，亦即(姓名, 性別)相依於學號，而「課程名稱」只與「課程代碼」有

「相依性」，亦即(課程名稱, 學分數, 必選修, 老師編號, 老師姓名)相依於課程代碼。

因此，學號是複合主鍵(學號, 課程代碼)的一部份。 ∴ 存在部分功能相依。

步驟二：將「部分功能相依」的欄位**分割出去**，再另外**組成新的資料表**
我們將「選課資料表」分割成三個較小的資料表 (加「底線」的欄位為主鍵)：

一、學生資料表(學號，姓名，性別)

學號	姓名	性別
001	李碩安	男
002	李碩崑	男

二、成績資料表(學號，課程代碼，成績)

學號	課程代碼	成績
001	C001	74
001	C002	93
002	C002	63
002	C003	82
002	C005	94

三、課程資料表(課程代碼，課程名稱，學分數，必選修，
老師編號，老師姓名)

課程代碼	課程名稱	學分數	必選修	老師編號	老師姓名
C001	程式語言	4	必	T001	李安
C002	網頁設計	3	選	T002	張三
C003	計 概	2	必	T003	李四
C005	網路教學	4	選	T005	王五

在第二正規化之後，產生三個資料表，分別為學生資料表、成績資料表及課程資料表，除了「課程資料表」之外，其餘兩個資料表(學生資料表與成績資料表)都已符合2NF, 3NF及BCNF。

➤ 第三正規化的規則

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

☑各欄位與「主鍵」之間沒有「遞移相依」的關係。

【如何找遞移相依呢？】

若要找出資料表中各欄位與「主鍵」之間的遞移相依性, 最簡單的方法就是從左到右掃瞄資料表中各欄位有沒有『與主鍵無關的相依性』存在。

可能的情況如下：

1. 如果有存在時，則代表有「遞移相依」的關係
2. 如果有不存在時，則代表沒有「遞移相依」的關係

第三正規化的作法

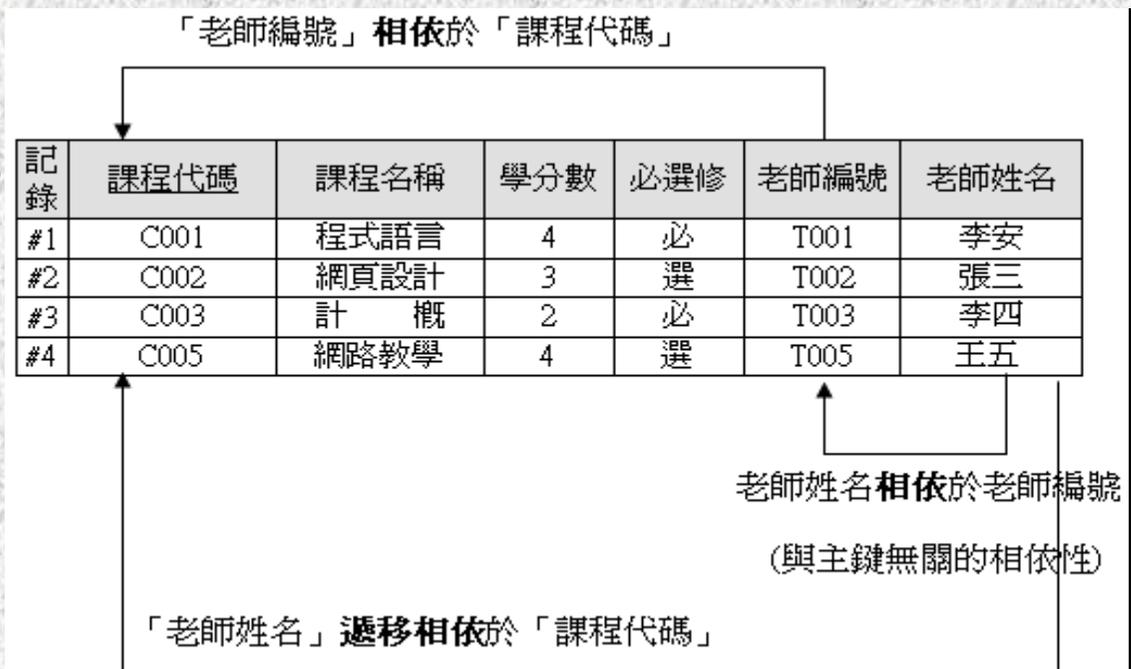
☑分割資料表；亦即將「**遞移相依**」或「**間接相依**」的欄位「**分割**」

出去，再另外組成「**新的資料表**」。其步驟如下：

步驟一：檢查是否存在「**遞移相依**」

由於每一門課程都會有授課的老師,因此,「**老師編號**」相依於「**課程代碼**」。並且「**老師姓名**」相依於「**教師編號**」,因此,存在有『與主鍵無關的相依性』。亦即存在「**老師姓名**」與**主鍵(課程代碼)**無關的相依性。

∴存在遞移相依。

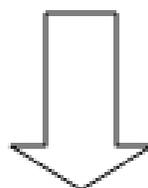




上述「課程資料表」中的[課程名稱]、[學分數]、[必選修]、[老師編號]都直接相依於主鍵[課程代碼](簡單的說，這些都是課程資料的必需欄位)，而[老師姓名]是直接相依於[老師編號]，然後才間接相依於[課程代碼]，它並不是直接相依於[課程代碼]，稱為「遞移相依」。

步驟二：將「遞移相依」的欄位「分割」出去，再另外組成「新的資料表」。因此，我們將「課程資料表」分割為二個資料表，並且利用外鍵(F.K.)來連接二個資料表。如下圖所示。

	課程代碼	課程名稱	學分數	必選修	老師編號	老師姓名
#1	A001	程式語言	4	必	T001	李安
#2	A002	網頁設計	3	選	T002	張三
#3	A003	計 概	2	必	T003	李四
#4	A005	網路教學	4	選	T005	王五



第三正規化，去除遞移相依

課程資料表

課程代碼*	課程名稱	學分數	必選修	老師編號#
A001	程式語言	4	必	T001
A002	網頁設計	3	選	T002
A003	計 概	2	必	T003
A005	網路教學	4	選	T005

符合 3NF, BCNF

老師資料表

老師編號	老師姓名
T001	李安
T002	張三
T003	李四
T005	王五

符合 3NF, BCNF

➤ 第三正規化後的四個表格

在我們完成第三正規化後，共產生了四個表格，如下表所示：

學生資料表

學號	姓名	性別
001	李碩安	男
002	李碩歲	男

符合 2NF, 3NF

成績資料表

學號	課程代碼	成績
001	A001	74
001	A002	93
002	A002	63
002	A003	82
002	A005	94

符合 2NF, 3NF

第二正規化產生的表格

課程資料表

課程代碼	課程名稱	學分數	必選修	老師編號#
A001	程式語言	4	必	T001
A002	網頁設計	3	選	T002
A003	計 概	2	必	T003
A005	網路教學	4	選	T005

符合 3NF

老師資料表

老師編號	老師姓名
T001	李安
T002	張三
T003	李四
T005	王五

符合 3NF

第三正規化產生的表格

BCNF正規化

是由Boyce和Codd於1974年所提出來的3NF的改良式。其條件比3NF更加嚴苛。因此每一個符合BCNF的關聯一定也是3NF。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就足夠了。

【適用時機】

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。

BCNF的規則

- ☑如果資料表的「主鍵」只由「單一欄位」組合而成, 則符合第三階正規化的資料表, 亦符合 BCNF(Boyce-Codd Normal Form)正規化。
- ☑如果資料表的「主鍵」由「多個欄位」組成(又稱為複合主鍵), 則資料表就必須要符合以下條件, 我們就說這個資料表符合BCNF(Boyce-Codd Normal Form)正規化的形式。
 - 1.符合3NF 的格式。
 - 2.「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

由於在我們完成第三正規化之後，已經分割成四個資料表，其中「成績資料表」的主鍵是由「多個欄位」組成(又稱為複合主鍵)。

因此，我們利用BCNF(Boyce-Codd Normal Form)正規化的條件，來檢驗「成績資料表」：

成績資料表(學號，課程代碼，成績)

<u>學號</u>	<u>課程代碼</u>	成績
001	C001	74
001	C002	93
002	C002	63
002	C003	82
002	C005	94

說明：

「成績」欄位相依於「課程代碼」及「學號」欄位，對「課程代碼」欄位而言，並沒有相依於「成績」欄位；對「學號」欄位而言，也沒有相依於「成績」欄位。所以成績資料表是符合「Boyce-Codd 正規化的形式」的資料表。