

# **Build School Blog**

Build School 是軟體人才教育的先鋒,為你打造更有競爭力的軟體工程師職涯

網誌

關於開發的那些小事

首百

f





2022-07-22

### 資料庫正規化筆記

#### 熱門文章與頁面:

- 資料庫正規化筆記
- [2024冬季班10/23 先修課 起-AI/LLM課程全面增加] Build School 軟體人才就業培訓簡介 及FAQ - 最專業微軟 .NET 全端 軟體工程師培訓
- C#及 HTML/CSS 初學者線上 學習資源
- C#學習筆記—物件與類別(2)
- [2024冬季班10/23 先修課 起-AI/LLM課程全面增加] Build School 軟體人才就業培訓簡介 及FAQ - 最專業微軟 .NET 全端 軟體工程師培訓
- C#學習筆記-物件與類別(4): 靜態類別與靜態成員





Photo by luis gomes on Pexels.com



Build School 粉絲專頁





- 先把資料儲存在資料表內(未正規化的資料表)
  - 未符合第一正規化造成的問題
- 第一正規化 (1NF)
  - 定義
  - 作法
  - 回到選課表
- 第二正規化 (2NF)
  - 定義
  - 作法
  - 回到選課表
- 第三正規化 (3NF)
  - 定義
  - 作法
  - 將有遞移相依的欄位分割出去
- 正規化結果
- Boyee-Codd正規化(BCNF)
  - 定義
  - 作法
  - 確認資料表是否都有符合BCNF
- Schema
- 結論
- 參考資料

## 正規化概念

資料庫正規化,又稱正規化、標準化,是資料庫設計的一系列原理和技術,以減少資料庫中數據冗餘,增進數據的一致性。

現在資料庫設計最多滿足3NF,普遍認為正規化過高,雖然具有對數據關係更好的約束性,但也 導致數據關係表增加而令資料庫IO更易繁忙。

#### 搜尋

搜尋...

#### 文章分類

- Build School 2017 (35)
- Build School 2018 (20)
- Build School 2019 (19)
- Build School 2020 (5)
- <u>Build School 2020台北班</u> (7)
- <u>Build School 2020新竹班</u> (4)
- 開發工具 (3)
- 關於開發的那些小事 (19)
- 行銷類課程 (2)
- 軟實力 (11)
- <u>台灣青築未來協會Build</u> the Future (23)
- 台北公開班 (6)
- 微軟認器 (5)
- 課程諮詢或協助? 請留訊...

校友幫幫忙分享 (17)

### 定義

將原先關聯表格的所有資訊,在分解成其他表格後,仍可以透過「合併」新關聯表格的方式得到相同的資訊,亦為「無損失分解」(Lossless Decomposition)。

### 不當設計造成的異常

• 新增異常:新增時,資料不齊全。

• 刪除異常:刪除時,造成資料遺失。

• 更新異常:更新時,可能會漏改。

### 小結

正規化的目的就是為了刪除「重複的資料」(Data Redundancy)及「避免更新異常」
(Anomalies)、每個階段都是以「欄位的相依性」作為分割資料表的依據、在實務上通常以滿足 3NF即可。

## 來看個例子



	某某科技大學【學生選課資料表】										
學號:001	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:	李碩崴	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					

## 不當設計的影響

- 新增異常:
  - o 新增學生,但未選課
  - o 新增課程,但未被選課
  - o 新增老師,但未授課
- 刪除異常:
  - o 刪除課程時,學生的成績資料遺失
- 更新異常:
  - o 更新課程資料時,需一次修改多筆資料



o 更新老師資料時,需一次修改多筆資料

## 正規化步驟

## 先把資料儲存在資料表內(未正規化的資料表)

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, C002	程式語言, 網頁設計	4, 3	必, 選	74, 93	T001, T002	李安, 張三
002	李碩崴	男	C002, C003, C005	網頁設計, 計概, 網路教學	3, 2, 1	選, 必, 選	63, 82, 94	T002, T003, T005	張三, 李四, 王五

#### 未符合第一正規化造成的問題

單一個欄位內有超過1個以上的值。

欄位長度無法確定(可能很多,也可能很少),必須預留空間,造成儲存空間的浪費。

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, C002	程式語言, 網頁設計	4, 3	必, 選	74, 93	T001, T002	李安, 張三
002	李碩崴	男	C002, C003, C005	網頁設計, 計概, 網路教學	3, 2, 1	選, 必, 選	63, 82, 94		張三, 李四, 王五

## 第一正規化(1NF)

#### 定義

- 1. 要排除重複群的出現,要求每一列資料的欄位值都只能是單一值(基元值)。
- 2. 沒有任何兩筆以上的資料完全重複。



- 3. 資料表中需有主鍵(唯一值),其他所有欄位都相依於主鍵。
- 4. 同一張資料表內,不建議用多個欄位表達同一個事情(如下圖)。

-- -- -- --

人	喜歡的顏色	不喜歡的食物 (1)	不喜歡的食物 (2)	不喜歡的食物 (3)
Jim	Green	Liver	Goat's cheese	
Alice	Fuchsia	Broccoli		
Norman	Blue	Pheasant	Liver	Peas
Emily	Yellow			

### 作法

- 1. 確認是否有重複表達的欄位。
- 2. 將欄位內重複的資料分別存為不同的Data Row資料。

### 回到選課表



Step1.消除	單一欄位內重	直複資料(將	欄位內的值捷	斥開,變成其	(他筆資料)				
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	李碩崴	男	C002	網頁設計	3	選	63	T002	張三
002	李碩崴	男	C003	計概	2	必	82	T003	李四
002	李碩崴	男	C005	網路教學	1	選	94	T005	王五
Step2. 確認 學號	Step2. 確認主鍵 (每筆資料間,欄位的值為不重複的) 學號 姓名 性別 課程代碼 課程名稱					必選修	成績	老師編號	老師姓名
001	李碩安	男	C001	程式語言	學分數 4	必	74	T001	李安
001	李碩安	男	C002	網頁設計	3	選	93	T002	張三
002	李碩崴	男	C002	網頁設計	3	選	63	T002	張三
002	李碩崴	男	C003	計概	2	必	82	T003	李四
002	李碩崴	男	C005	網路教學	1	選	94	T005	王五

可以將學號&課程代碼組合成唯一值(Key),透過學號&課程代碼,可以修改成績。

## 第二正規化(2NF)

### 定義

- 1. 符合1NF。
- 2. 消除「部分功能相依」,每一個非鍵欄位必須完全相依主鍵(學號=>學生,課程=>學分...),通常「主鍵有多個欄位」組成時會發生「部分功能相依」。

### 作法

- 1. 檢查是否存在「部分功能相依」(可從多個欄位組成的主鍵開始檢查)。
- 2. 將「部分功能相依」的欄位分割出去,另外組成新的資料表。



### 回到選課表

1NF 後的	<b>的資料表</b>								
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	李碩崴	男	C002	網頁設計	3	選	63	T002	張三
002	李碩崴	男	C003	計概	2	必	82	T003	李四
002	李碩崴	男	C005	網路教學	1	選	94	T005	王五

主鍵是由「學號」與「課程代碼」組成:

- 姓名、性別,相依於「學號」
- 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」

存在「部分功能相依」,需要將部分功能相依欄位分割(2NF),以下為分割後結果



學生資料表					
學號	姓名	性別			
001	李碩安	男			
001	李碩安	男			
002	李碩崴	男			
002	李碩崴	男			
002	李碩崴	男			
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
C001	程式語言	4	必	T001	李安
C002	網頁設計	3	選	T002	張三
C002	網頁設計	3	選	T002	張三
C003	計概	2	必	T003	李四
C005	網路教學	1	選	T005	王五
成績資料表					
學號	課程代碼	成績			
001	C001	74			
001	C002	93			
002	C002	63			
002	C003	82			
002	C005	94			



發現「重複的資料」,因為需符合1NF,所以將重複的資料去除

學生資料表					
學號	姓名	性別			
001	李碩安	男			
002	李碩崴	男			
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
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	1	選	T005	王五
成績資料表					
學號	課程代碼	成績			
001	C001	74			
001	C002	93			
002	C002	63			
002	C003	82			
002	C005	94			

紅底的欄位為「主鍵」

第三正規化 (3NF)



### 定義

- 1. 符合2NF。
- 2. 各欄位之間沒有存在「遞移相依」的關係,也就是與「主鍵」無關的相依性。

#### 作法

- 1. 檢查是否存在「遞移相依」的欄位。
- 2. 將「遞移相依」的欄位分割出去,另外組成新的資料表。

#### 將有遞移相依的欄位分割出去

1. 檢查是否存在遞移相依(與主鍵無關的相依性)。

課程資料表(不符合3NF,課程編號與老師姓名存在遞移相依)								
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	1	選	T005	王五			

- 「老師編號」相依於「課程代碼」而且「老師姓名」相依於「老師編號」,所以「老師姓名」與「課程編號」(主鍵)為無關的相依。
- 「老師姓名」與主鍵存在著「遞移相依」的關係。
- 2. 分割欄位



課程資料表(符合3NF)					老師資料表(符合3
課程代碼	課程名稱	學分數	必選修	老師編號	老師編號 老師如
C001	程式語言	4	必	T001	T001 李安
C002	網頁設計	3	選	T002	T002 張三
C003	計概	2	必	T003	T003 李四
C005	網路教學	1	選	T005	T005 王五

正規化結果



學生資料表	(符合3NF)			
學號	姓名	性別		
001	李碩安	男		
002	李碩崴	男		
成績資料表	(符合3NF)			
學號(FK)	課程代碼(FK)	成績		
001	C001	74		
001	C002	93		
002	C002	63		
002	C003	82		
002	C005	94		
課程資料表	(符合3NF)			
課程代碼	課程名稱	學分數	必選修	老師編號(FK)
C001	程式語言	4	必	T001
C002	網頁設計	3	選	T002
C003	計概	2	必	T003
C005	網路教學	1	選	T005
老師資料表	(符合3NF)			
老師編號	老師姓名			
T001	李安			
	_			



T002	張三		
T003	李四		
T005	王五		

## Boyee-Codd正規化(BCNF)

視情況使用,實務上大多都只會做到3NF。

### 定義

- 3NF的改良式(必須滿足3NF)。
- 主鍵中的各欄位(單獨看)不可以相依於其他非主鍵的欄位。

### 作法

- 1. 確認由多個欄位組成的主鍵是否有其他欄位相依於主鍵中的每個欄位。
- 2. 如有獨立相依的情況,新增一個獨立的主鍵欄位。

### 確認資料表是否都有符合BCNF

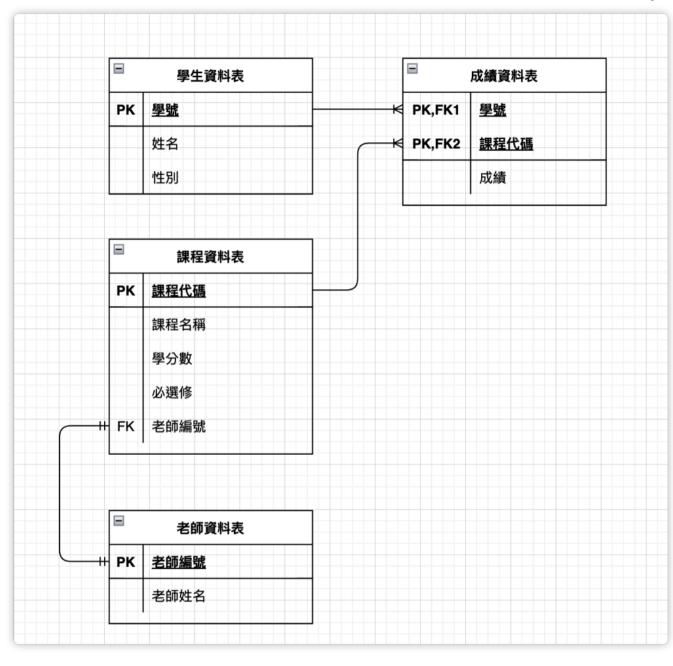
- 學生資料表
  - 3NF
  - o 單一主鍵
  - o 符合BCNF
- 課程資料表



- 3NF
- o 單一主鍵
- o 符合BCNF
- 老師資料表
  - o 3NF
  - o 單一主鍵
  - o 符合BCNF
- 成績資料表
  - 3NF
  - 主鍵由兩個欄位組成,但:
    - 成績相依於「學號」及「課程代碼」欄位
    - 成績未相依於「學號」(必須要有課程代碼才知道是哪堂課的分數)
    - 成績未相依於「課程代碼」(必須要有學號才知道是誰的分數)
  - o 符合BCNF

### Schema





結論



- 每個階段的正規化,都會產生一個要解決的問題
  - o 1NF:去除重複資料。
  - o 2NF:去除部分功能相依。
  - o 3NF:去除遞移相依。
  - o BCNF:去除因功能相依產生的異常。
- 實務上我們都會做到第三正規化(3NF)。
- 設計不當的資料(未正規化),會導致異動操作可能會有問題。
  - o 新增異常
  - o 刪除異常
  - o 更新異常
- 適當的正規化,可以確保數據資料的一致性。
- 當遇到資料處理速度上的困擾,或是因分割後造成的損失時,會將資料表「反正規化」。
  - o 商品的庫存
  - o 剩餘點數
  - 0 ...
- 在效能的考量之下的反正規化設計,需要小心控制「資料重複性的問題」及「更新的異常」。

## 參考資料

<a href="http://cc.cust.edu.tw/~ccchen/doc/db\_04.pdf">http://cc.cust.edu.tw/~ccchen/doc/db\_04.pdf</a>



- <a href="http://debussy.im.nuu.edu.tw/sjchen/Database/Final/Ch05.pdf">http://debussy.im.nuu.edu.tw/sjchen/Database/Final/Ch05.pdf</a>
- https://ithelp.ithome.com.tw/articles/10223029

本文同步發布在 <u>資料庫正規化筆記</u> 若對內容有任何疑問或內容有錯誤,

歡迎來信: andyhuang@build-school.com



Andy Huang

PM / Software Engineer / Student Coach @Build School



Medium

- ┗ 關於開發的那些小事
- ◆ DATABASE 
  ◆ DEVELOPER 
  ◆ NORMALIZATION 
  ◆ RDBMS 
  ◆ T-SQL







## Andy Hunag

PM · Software Engineer · Student Coach @BuildSchool



Next >

Build School 2022 專題 Demo 簡介及影片

用生動的圖表說故事 – 活用 Microsoft Power BI 成為資料分析英雄 [0729 台北市就 業服務處-創業精進講座x台灣微軟 Microsoft」

Build School is a pioneer of software developer education and career builder for engineers

Build School 青杉人才有限公司 | 青群科技 | 社團法人台灣青築未來協會 Build the Future 台北: 台北市忠孝東路三段96號11F-1 (北科大對面/忠孝新生站3號出口) | 台北新竹招生中 contact@build-school.com

