

<<會計資訊系統課程講義>>

Access 2007 DBMS

~資料庫正規化~

周國華

屏東商業技術學院會計系

初版：2007/11/23

本次修正：2010/11/17

智慧財產權聲明

- 本文件係由周國華老師獨自撰寫，除引用之概念屬於原文作者外，其餘文字及圖形內容之智慧財產權當然屬於周老師獨有。
- 任何機構或個人，在未取得周老師同意前，不得直接以本文件做為學校、研究機構、企業、會計師事務所、政府機關或財團法人機構舉辦教學或進修課程之教材，否則即屬侵權行為。
- 任何機構或個人，在未取得周老師同意前，不得在自行編撰的教材中直接大量引用本文件的內容。若屬單頁內部分內容之引用，亦請註明出處。

大雜燴資料表

- 假設某家出版社的訂單資料表結構如下：

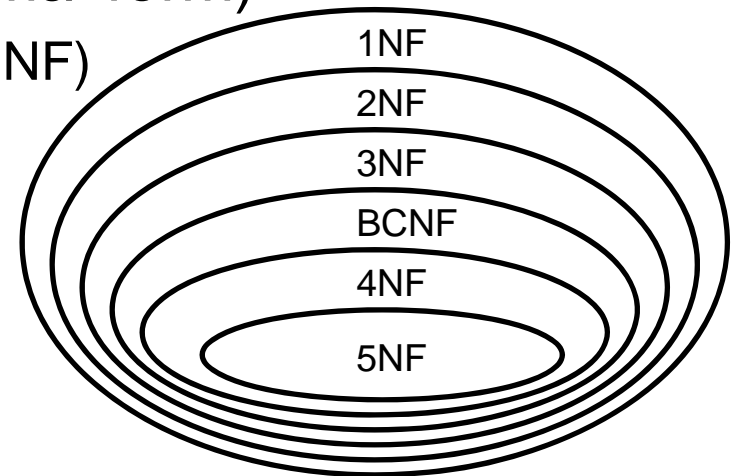
訂單編號	訂單日期	客戶名稱	送貨地址	客戶電話	書籍編號	書籍名稱	訂購數量	銷貨定價	收款條件
------	------	------	------	------	------	------	------	------	------

- 上述資料表結構的潛在問題如下：
 - 一張訂單可能包含多種書籍，則書籍編號及其後數個欄位會出現多重值。
 - 若將同一訂單的每種書籍分割在不同資料列中，則相同訂單編號會重複出現，不符合主索引的規範。
 - 許多欄位的值需一再重複輸入，不但浪費儲存空間，也很容易出現輸入內容不一致的情況。
- 這種大雜燴資料表所潛藏的問題，通稱為更新時的異常現象(update anomalies)，可透過資料庫正規化程序來排除。

資料庫正規化

- 資料庫正規化(normalization)是從關聯式資料模式衍生出來的理論，目的是為大雜燴型資料表的分割提供一套可遵循的模式。它可分成六種形式(normal form)：

- 第一正規化形式(First Normal Form, 1NF)
- 第二正規化形式(2NF)
- 第三正規化形式(3NF)
- Boyce/Codd 正規化形式(BCNF)
- 第四正規化形式(4NF)
- 第五正規化形式(5NF)



- 隨著正規化形式逐漸複雜(1NF→5NF)，被分割出來的資料表數量越多。且每一種較複雜的形式，都完全符合較簡單形式的規範(一個5NF，一定也是4NF...)，見上圖。

越分越細不一定好

- 資料表分割得越細，雖然可進一步減少更新時的異常現象，但使用者在查詢時，系統卻需要用到更多的合併(join)運算，而這種運算是很耗費系統資源的。
- 所以，在大部分資料庫管理實務上，資料表的正規化只需要達到**3NF**即可(有複合主索引的資料表需達到**BCNF**)，以便在「減少異常現象」與「確保執行效能」之間取得適當的平衡。

功能相依性

- 從1NF到BCNF等四種正規化形式，都是以功能相依性 (functional dependency) 的概念做為思考起點。
- 功能相依性：在R資料表中的A、B兩個欄位，若A欄位的一個值在任何時點下可對應至B欄位的唯一一個值，則稱B功能相依於A (B is functionally dependent on A)，或A在功能上決定B (A functionally determines B)。
 - 例如：一個學號(A)對應至唯一一個姓名(B)，但反之不然。
- 資料表經過正規化程序後，主索引欄位在功能上決定所有其他欄位，或者，所有其他欄位功能相依於主索引欄位。
 - 主索引可能由兩個或多個欄位構成(複合主索引)，因此上述的A，也可以由兩個或多個欄位所構成(：A欄位的一組值可決定B欄位的唯一一個值)。

1NF

- **1NF**定義：一份資料表內所有資料列的每個欄位，都只包含一個資料項，則該資料表即達到**1NF**。
- 以前述訂單資料表為例，一張訂單若包含多種書籍，則每種書籍需個別儲存在不同資料列內。
- 前述訂單資料表轉成**1NF**後，訂單編號這個主索引會出現重複值，不符合主索引的條件，因此需要增加一個欄位來構成複合主索引(**composite primary key**)。
 - 新的(複合)主索引：訂單編號 + 書籍編號

訂單編號	書籍編號	訂單日期	客戶名稱	送貨地址	客戶電話	書籍名稱	訂購數量	銷貨定價	收款條件
------	------	------	------	------	------	------	------	------	------

多重值欄位

- 大部分資料庫軟體都不允許在資料列的個別欄位內儲存多重值，因此資料表在一開始就符合**1NF**規範。
- 在某些特殊情況下，使用者可能需要在單一儲存格中存放或閱覽多個資料項，**Access 2007**及**2010**針對此需求有兩項新設計：
 - 附件資料類型：某個欄位的資料類型若設定為「附件」，即可在該欄位中儲存多個**OLE**資料檔案。
 - 多重值欄位：欄位若設定為「附件」以外的資料類型，就只能儲存一筆資料項。但透過「查閱精靈」的設定，讓特定欄位轉成查閱欄後，就可在資料表的單一欄位內呈現多個資料項的內容。

1NF的問題

- 資料表達到1NF後，雖然消除了多重值，卻產生一個大問題：某些其他欄位只相依於部分主索引欄位。
 - 例如，在前述訂單資料表中，訂單日期、客戶名稱、送貨地址、客戶電話、收款條件等欄位只相依於訂單編號，書籍名稱、銷貨定價等欄位只相依於書籍編號。
- 1NF所產生的部分相依性(partial dependency) 會阻礙欄位資料的新增及刪除。
 - 例如，想要新增書籍的相關資料(書籍編號、書籍名稱、銷貨定價)，必須等到拿到訂單後才能進行；想要刪除特定書籍的資料，會連帶把訂單內的其他珍貴資料也刪除掉。
- 此外，許多敘述性欄位(客戶名稱、地址、書籍名稱等)必須重複輸入資料，不但浪費空間也容易發生錯誤。

2NF

- **2NF定義**：一份已符合**1NF**規範的資料表，若其主索引以外欄位均完全相依於主索引欄位，該資料表即達到**2NF**。
- **從1NF→2NF的步驟**：
 - 將主索引分解成部分集合，做為新資料表的主索引。
 - 將**1NF**中的部分相依欄位抽離至對應的新資料表內。
- 按照上述步驟，前述訂單資料表將分割成以下三份表格：

訂單編號	訂單日期	客戶名稱	送貨地址	客戶電話	收款條件
------	------	------	------	------	------

書籍編號	書籍名稱	銷貨定價	訂單編號	書籍編號	訂購數量
------	------	------	------	------	------

- 資料表達到**2NF**後，前述因部分相依所產生的新增及刪除資料問題即迎刃而解，書籍名稱的敘述性內容也不再需要重複輸入。

2NF的問題

- 資料表達到2NF後，仍然存在一個問題：非主索引欄位之間有相依性。
 - 例如，前頁以訂單編號為主索引的資料表中，送貨地址、客戶電話兩個欄位相依於客戶名稱。

訂單編號	訂單日期	客戶名稱	送貨地址	客戶電話	收款條件
------	------	------	------	------	------

- 在非主索引欄位之間有相依性，使2NF的資料表仍然有敘述性內容會重複輸入的情況。
 - 例如，相同客戶每次訂書時，都得重新輸入客戶名稱、送貨地址、電話等資料。
- 此外，如果業務人員要事先建立客戶資料，也無法記錄在此份資料表中，而必須等到客戶下訂單後才能輸入。

3NF

- **3NF**定義：一份已符合**2NF**規範的資料表，若其主索引以外的欄位之間不存在功能相依性，該資料表即達到**3NF**。
- 從**2NF**→**3NF**的步驟：
 - 將彼此間具有功能相依性的欄位移出，成為獨立的資料表。
 - 若被相依的欄位(前例為客戶名稱)是新資料表的主索引，則該欄位保留在原資料表內做為外來鍵；若新資料表內另設主索引欄位，則以該主索引納入原資料表內做為外來鍵。
- 按上述步驟，前例資料表將分割成以下兩份表格：

訂單編號	訂單日期	客戶編號	收款條件	客戶編號	客戶名稱	送貨地址	客戶電話
------	------	------	------	------	------	------	------

- 資料表達到**3NF**後，因非主索引欄位間存在相依性所引發的問題即迎刃而解。

BCNF

- 在大部分資料庫管理實務上，資料表的正規化只需要達到3NF即可。如果資料表的主索引是由多個欄位組成，該資料表須進一步達到BCNF (Boyce/Codd NF)。
- **BCNF**定義：一份已符合3NF規範的資料表，若其主索引內的個別欄位並不相依於主索引以外的任何欄位，該資料表即達到**BCNF**。
- 前述書籍訂單資料表分割後所產生的唯一複合主索引資料表，其主索引內的個別欄位(訂單編號、書籍編號)均不相依於主索引以外的欄位(訂購數量)，因此該資料表已達到**BCNF**。


訂單編號	書籍編號	訂購數量
------	------	------

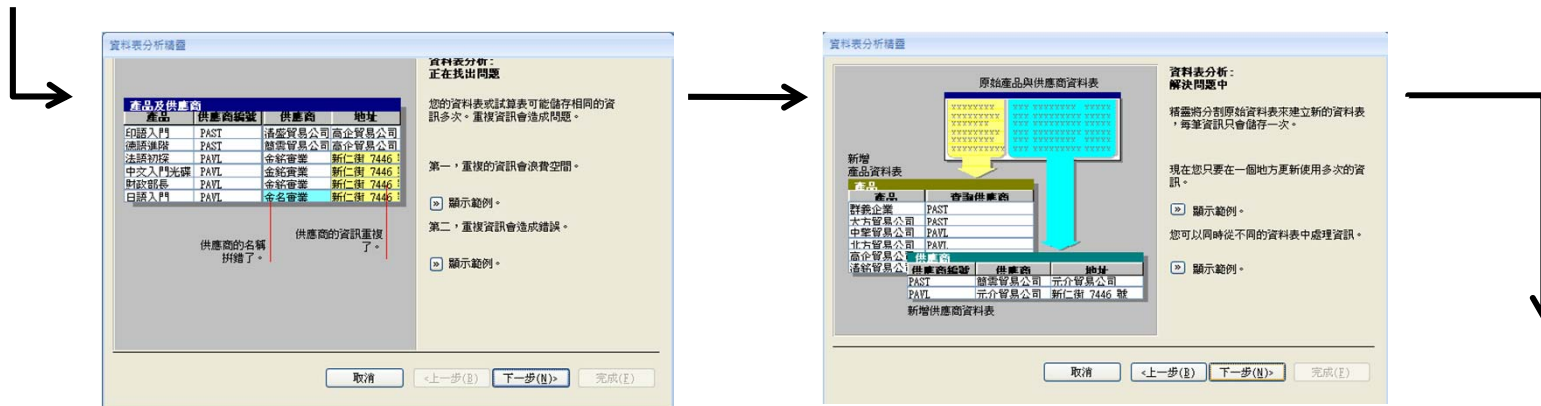
資料表分析精靈

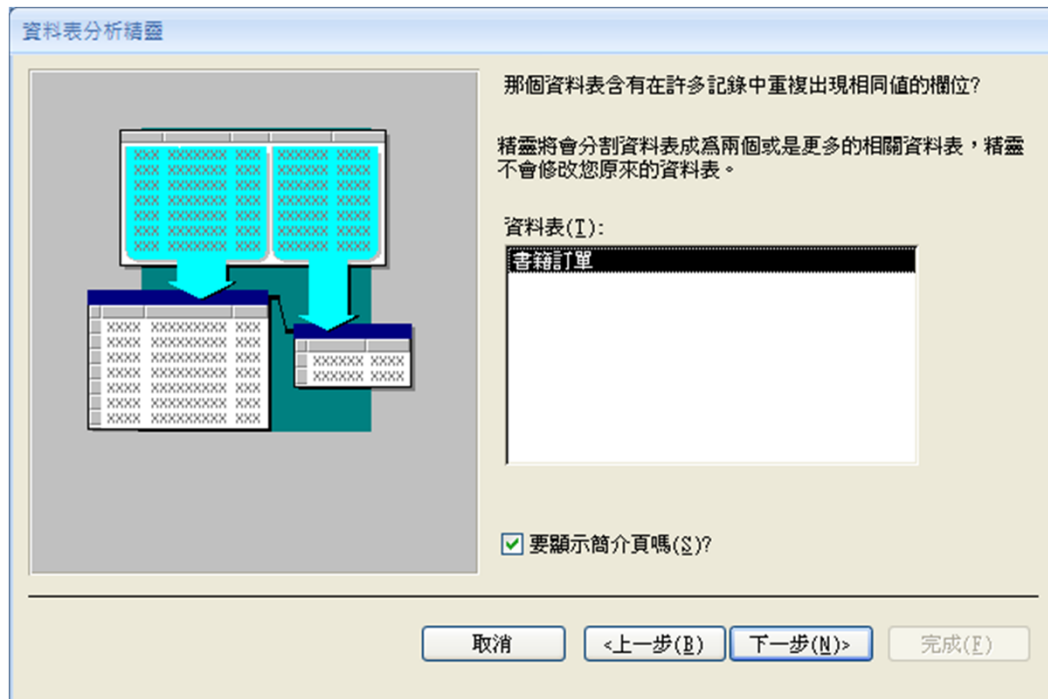
- Access 2007內建的「資料表分析精靈」，可協助使用者將大雜燴式資料表分割成符合正規化的形式。
- 解說範例：書籍訂單資料表

訂單編號	書籍編號	訂單日期	客戶名稱	送貨地址	客戶電話	書籍名稱	訂購數量	銷貨定價	收款條件
S00001	A00001	2007/11/14	大方書局	屏東市民生路101	08-7231234	會計資訊系統	10	NT\$500.00	2/10, n/30
S00001	A00002	2007/11/14	大方書局	屏東市民生路101	08-7231234	網際網路會計應付	15	NT\$450.00	2/10, n/30
S00002	A00001	2007/11/15	小氣書局	高雄市十全路1號	07-8235465	會計資訊系統	30	NT\$500.00	貨到付款
S00002	A00003	2007/11/15	小氣書局	高雄市十全路1號	07-8235465	會計資料庫管理	8	NT\$630.00	貨到付款
S00003	A00004	2007/11/16	東北書局	潮州鎮健康路890	08-4567890	中級會計學	5	NT\$650.00	貨到付款
S00004	A00002	2007/11/16	東西書局	台南市王府路56號	06-6543210	網際網路會計應付	15	NT\$450.00	2/10, n/60
S00004	A00005	2007/11/16	東西書局	台南市王府路56號	06-6543210	財務報告分析	30	NT\$580.00	2/10, n/60
S00005	A00006	2007/11/17	西南書局	高雄縣鳳山市中區	07-7229988	高級會計學	9	NT\$560.00	貨到付款
S00006	A00003	2007/11/22	小氣書局	高雄市十全路1號	07-8235465	會計資料庫管理	10	NT\$630.00	貨到付款
S00006	A00007	2007/11/22	小氣書局	高雄市十全路1號	07-8235465	初級會計學	50	NT\$450.00	貨到付款
S00007	A00008	2007/11/23	中央書局	屏東市和生路909	08-7235654	財務管理	8	NT\$800.00	2/10, n/30
S00008	A00008	2007/11/24	東北書局	潮州鎮健康路890	08-4567890	財務管理	5	NT\$800.00	2/10, n/30
S00009	A00004	2007/11/25	大方書局	屏東市民生路101	08-7231234	中級會計學	6	NT\$650.00	2/10, n/30
S00009	A00007	2007/11/25	大方書局	屏東市民生路101	08-7231234	初級會計學	20	NT\$450.00	2/10, n/30
S00010	A00009	2007/11/26	小氣書局	高雄市十全路1號	07-8235465	投資學	13	NT\$650.00	2/10, n/30

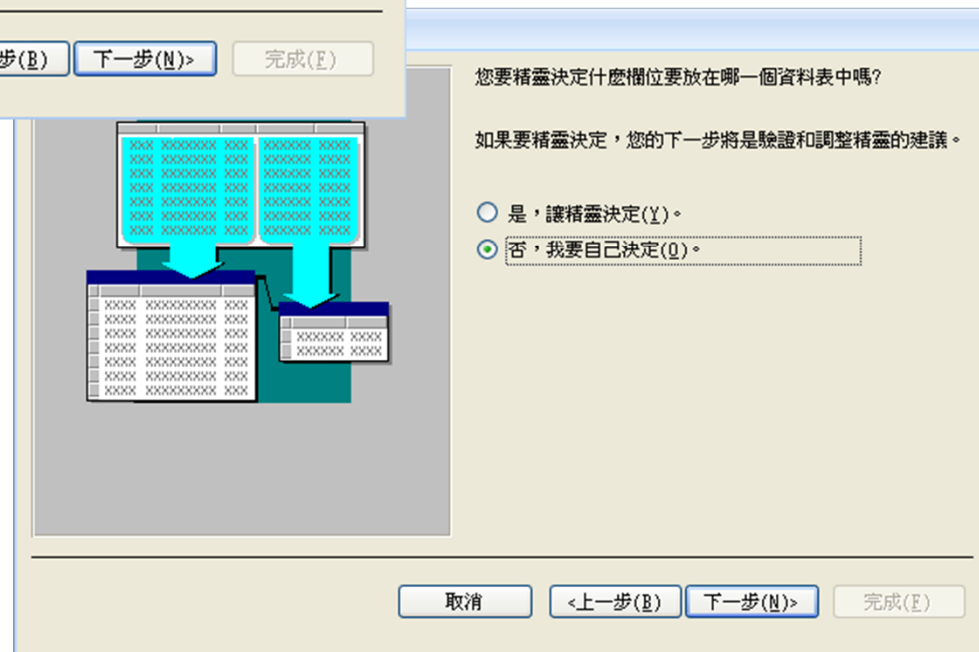
分析步驟

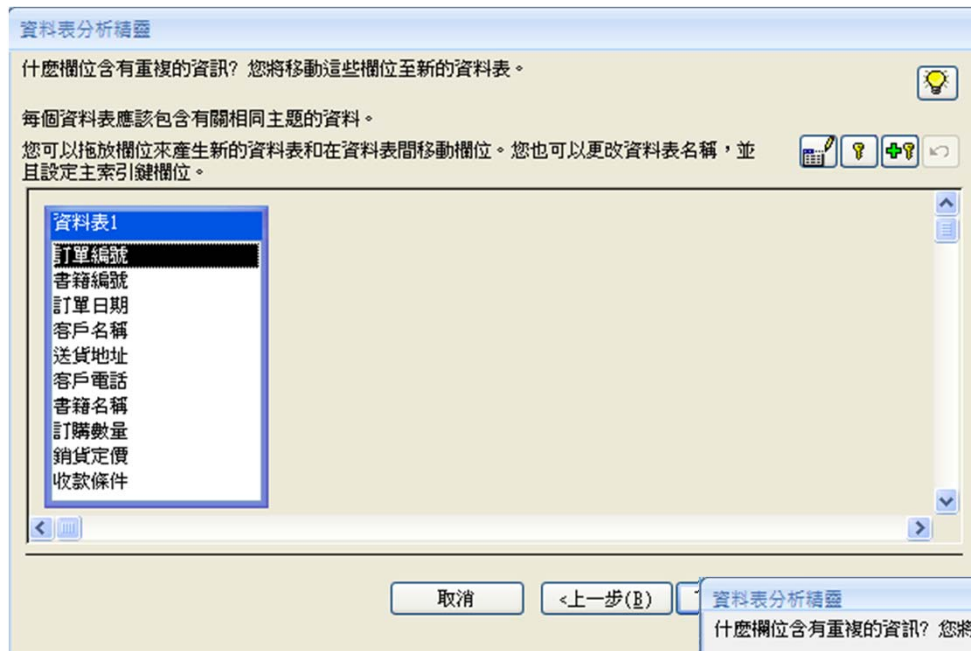
- 開啟資料庫檔案後，按下功能區的「資料庫工具」索引標籤，再按  按鈕，出現資料表分析精靈後，按兩次「下一步」略過前兩頁..



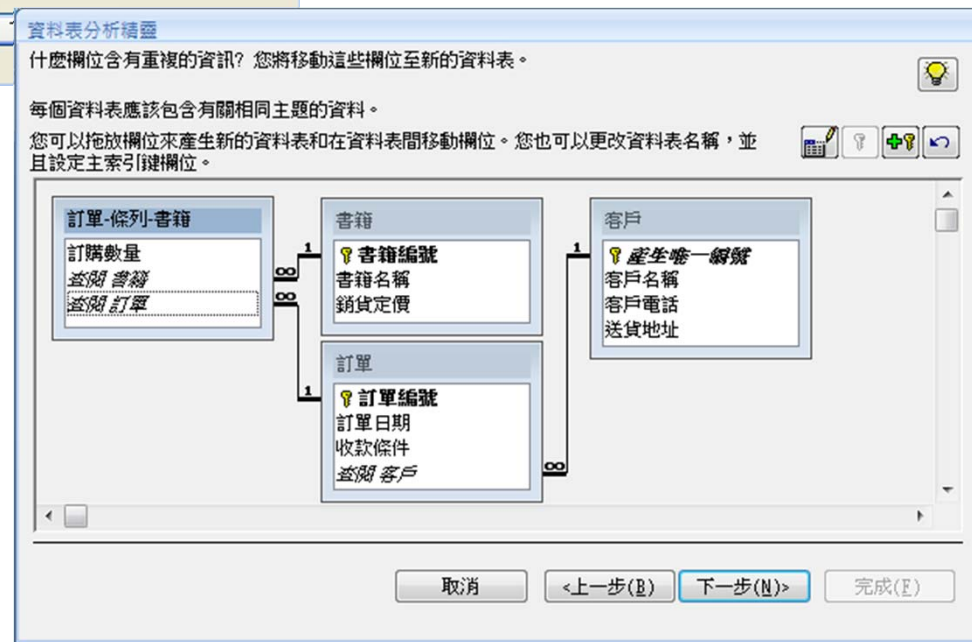


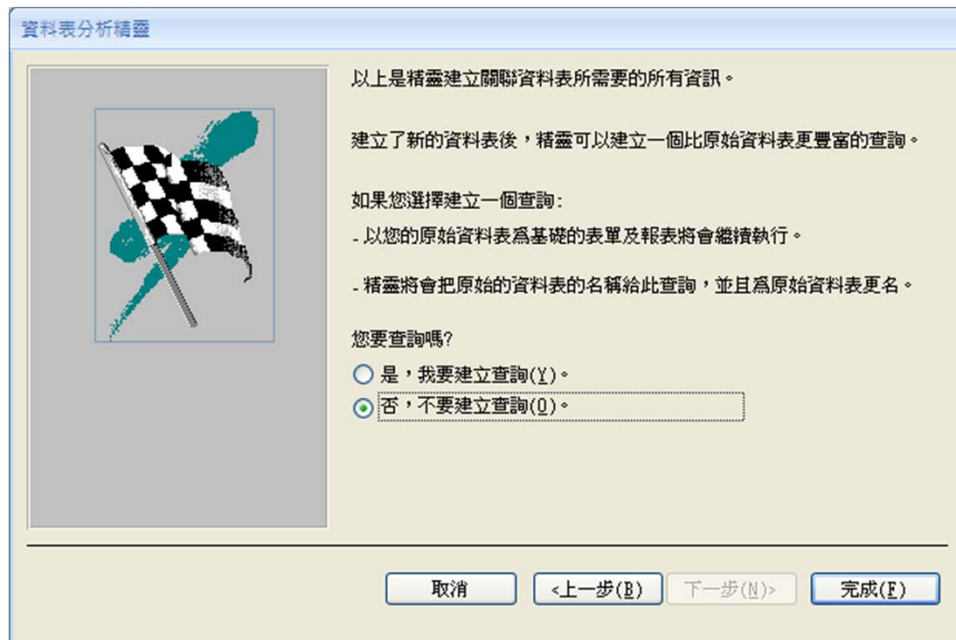
- 在第三個頁面(上圖)中選擇欲分析的資料表。
- 在第四個頁面(右圖)中選擇「否，我要自己決定」(: 因為精靈不夠聰明)。



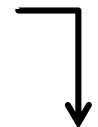
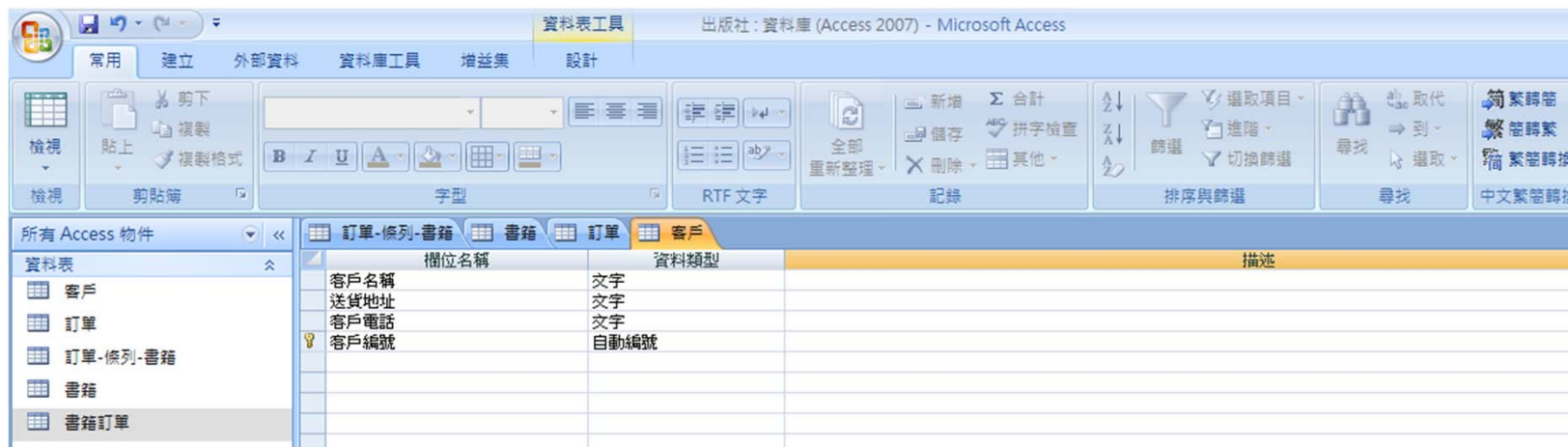


- 在第五個畫面(上圖)內，以手動方式將單一資料表分割成四個關聯資料表(右圖)。
 ※ 分割前須先將原資料表內所設定的主索引取消，否則將無法移出建立新表格。





- 在第六個畫面(左圖)中選擇「否，不要建立查詢」(:查詢為下一次課程主題)，按「完成」，系統即自動產生四個新的資料表(下圖)。
- 此時可繼續對各資料表內的欄位名稱進行修正，以符合正確語意。





- 整個範例分割完成後之資料庫關聯圖如上。
- 以上說明係使用者選擇由自己決定分割方式(分析精靈的第四個頁面)後之相關分析步驟。使用者也可選擇由精靈決定分割方式，如果分割結果與實況大致符合，就只需要做簡單微調即可。如果精靈分割結果與實況差異太大，需要做的調整過於複雜，可重新選擇由自己決定分割方式。

表單物件的必要性

- 大雜燴型資料表經過正規化後，消除了更新時的異常現象，但分割為數個資料表後，卻造成資料輸入的困擾。
 - 以訂單為例，通常至少需輸入客戶名稱、訂購品項、訂購數量、銷售價格、小計、總計等資訊，這些資訊項目或者散佈在不同資料表內，或者在原始資料表內並無該欄位、而須透過查詢物件間接取得數據內容。
- 因此，在資料庫管理實務上，除了必須完成資料表的正規化及關聯性設定外，也需要設定相關的查詢物件及表單物件。透過表單輸入資料，數據會自動儲存到對應的數個資料表內，可讓資料庫的操作更有效率。