



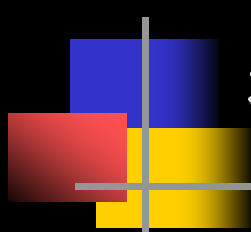
San Francisco, USA by Frank S.C. Tseng
(<http://www2.nkfust.edu.tw/~imfrank>)

第十四章 異質性分散式資料庫管理系統



本章內容

- 14.1 前言
- 14.2 異質性資料庫系統簡介
- 14.3 建構一個異質性資料庫系統
- 14.4 異質性資料庫系統中的綱要整合問題
- 14.5 異質性資料庫的查詢語言
- 14.6 異質性資料庫中的資料與綱要整合
- 14.7 異質性資料庫系統中的異動管理問題



異質性分散式資料庫系統

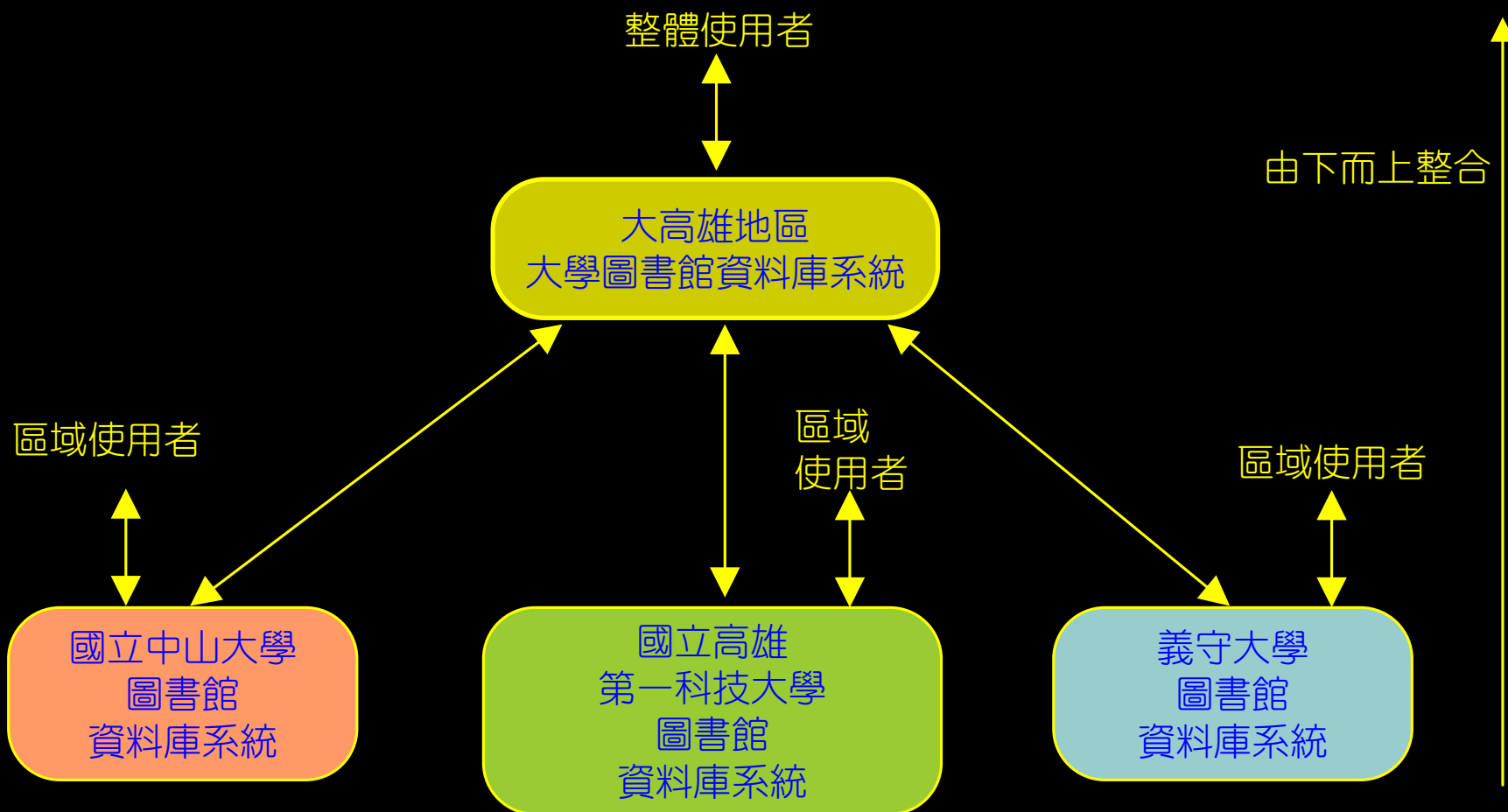
- 電腦系統發展的趨勢
 - 開放性（技術開放、連結性開放）
 - 異質性（各自發展不同的系統與介面）
 - 整合性（不同平臺可以互相整合）
- 銀行、電信、圖書，等服務業會漸漸朝整合型態發展
- 企業的購併風潮也促使異質性資料庫系統的整合



全球資訊互通的發展歷程

- World-Wide Telephony System
- World-Wide Computer Network System:
Internet—“Network as a Computer”
重點在於 “系統的整合”（整合現有系統）
- Heterogeneous Database Systems—
“Network as a Global Database”
重點在於 “資料的整合”（整合現有資料）

異質性分散式資料庫系統(續)





異質性資料庫上的異質性

- Computer System: Hardware, Software, OS
- Network Protocol: TCP/IP, Token Ring, ...
- Database Systems: (可以藉制定標準統一)
 - Data Model (Object-Oriented vs. Relational)
 - Product/System (Sybase, Oracle, MS SQL Server...)
 - Query Language (SQL-89, SQL-92, Quel, ...)
 - Data Encodinig (ASCII, BIG-5, GB, Unicode, ...)
- Logical Database Schemas (雖然有工具可以協助，但是目前還是需要人的參與) 目前有研究人員嘗試將 Schema Matching 自動化，但僅有初步成果



異質性資料站的自主性 (Autonomy)

- **Design Autonomy** : Design strategy, data model, query language, ...
- **Communication Autonomy** : *not forced to supply data.*
- **Execution Autonomy** : 各資料站可以 delay/reject/abort 任何的整體異動
- **Control Autonomy** : 區域異動不能因整體異動而被撤回
- **Association Autonomy** : 決定與誰相連



對不同綱要整合的作法

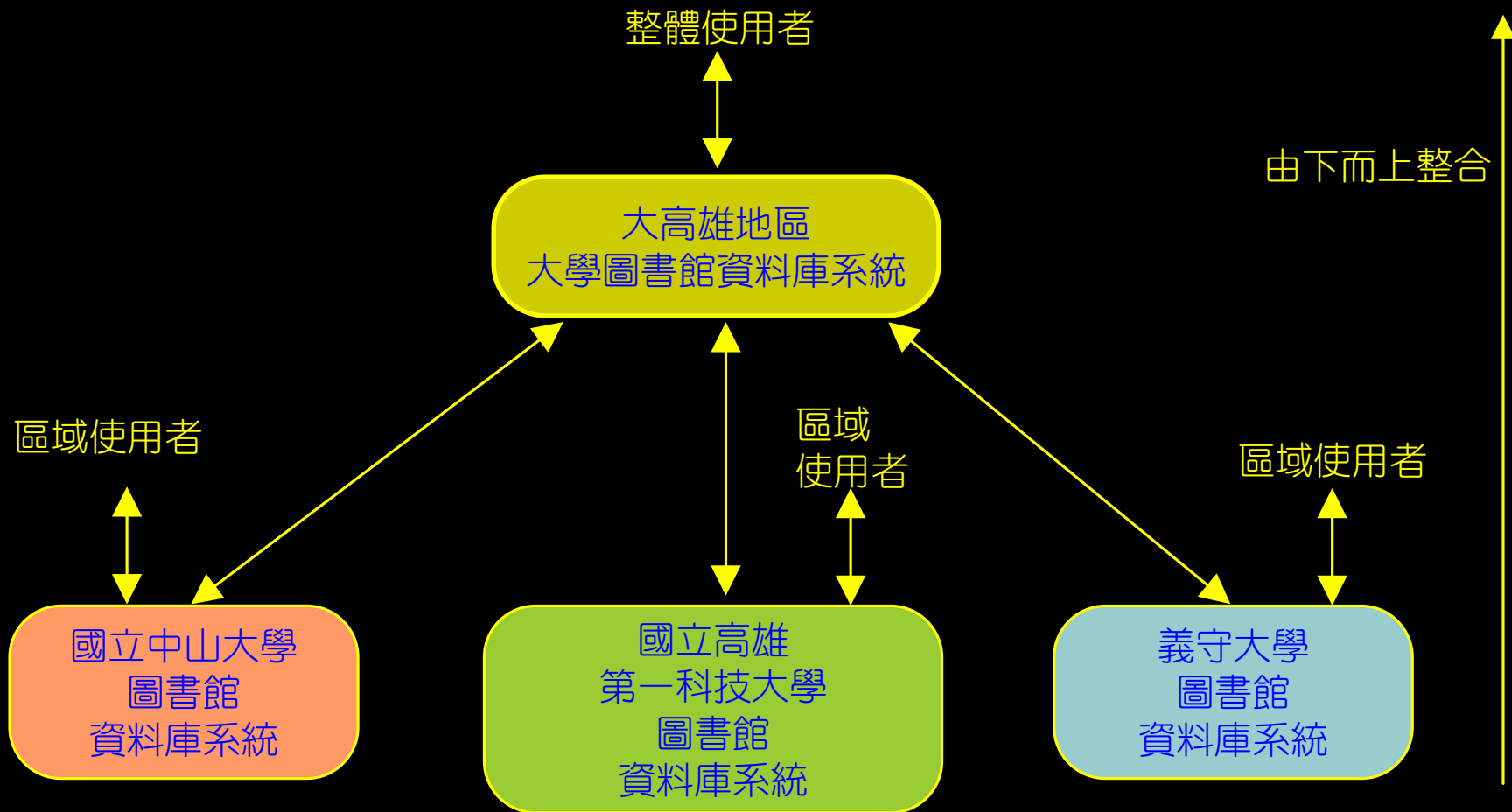
- 異質性結構化資料庫查詢語言 (MSQL) 或提供使用者一個圖型化介面來做整合
- 整體綱要整合 (Global Schema Approach)
- 聯邦式綱要整合 (Federated Database Approach) : 較好的 Approach
- 所以 Heterogeneous Database Systems 又稱為 **Federated Database Systems** 或 **Multidatabase Systems**



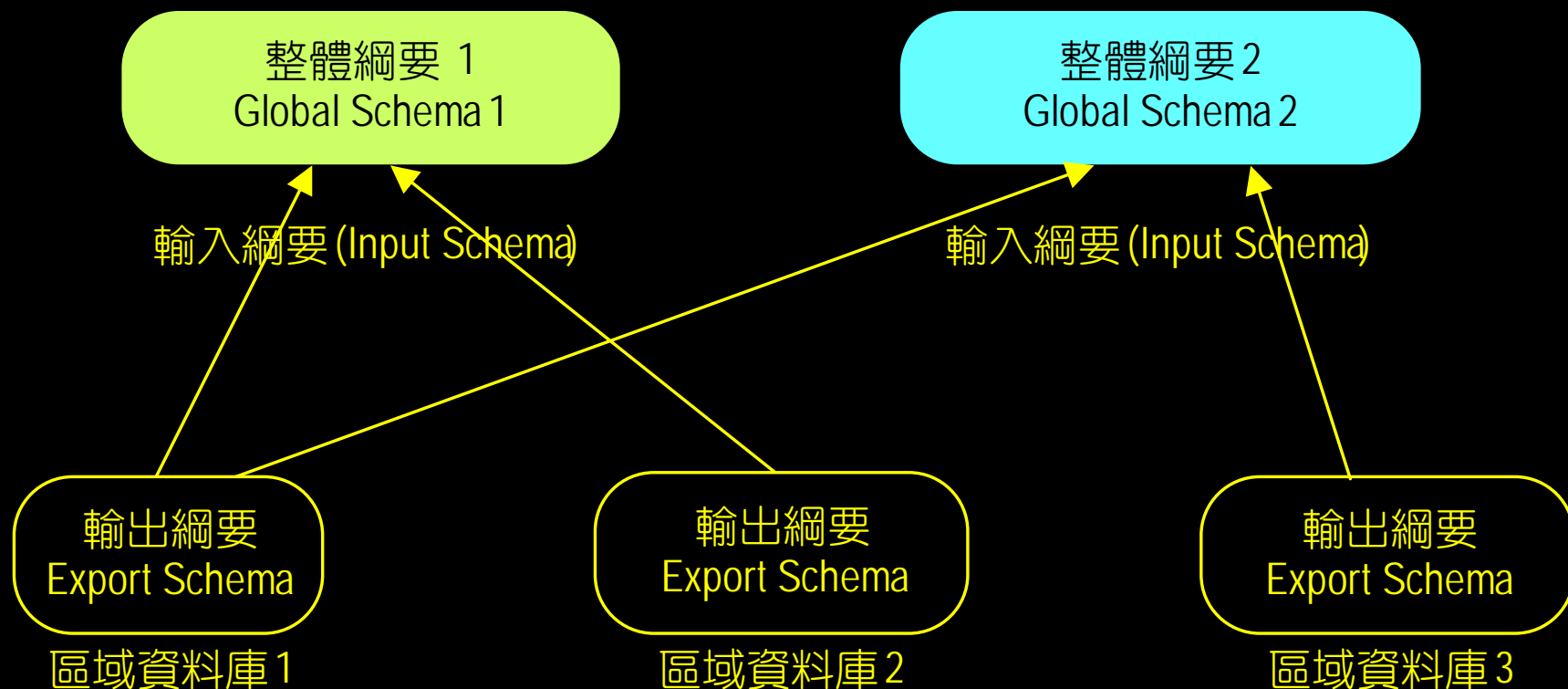
異質性資料庫的查詢語言

- 提供一套可以跨越不同資料站間的「異質性資料庫查詢語言」(Multi-database Query Language)
- 只能說是建置一個異質性資料庫系統的暫時性做法
- 使用者必須自行利用此套語言，自行負責整合的工作，所以使用上的親和力並不夠
- 讀者可以參考 Witold Litwin 等人所發展的 MultiDatabase SQL (MSQL)

整體綱要整合 (Global Schema Approach)



聯邦式異質性資料庫整合





對不同綱要整合的問題

- 資料命名上的衝突 (Data Naming Conflicts)
- 資料表示法上的衝突 (Data Representation Conflicts)
- 資料的編碼或資料內容的差異 (Data Encoding or Content Conflicts)
- 資料單位的衝突 (Data Scaling Conflicts)
- 不相容的資料 (Inconsistent Data)
- 資料缺項 (Missing Data)
- 角色互換的差異 (Role Conflicts)
- 綱要結構上的差異 (Structural Difference)



綱要整合問題（李強教授的分法）

- Value-to-Value Conflicts. (資料值間的衝突)
 - Data Representation Conflicts (資料表示法衝突)
 - Data Scaling Conflicts (資料單位衝突)
 - Inconsistent Data (資料不一致)
- Value-to-Attribute Conflicts (資料與屬性的衝突)
- Value-to-Table Conflicts (資料與關聯表的衝突)
- Attribute-to-Attribute Conflicts (屬性間的衝突)
 - 對應到前面的 Data Naming Conflicts
- Attribute-to-Table Conflicts (屬性與關聯表的衝突)
- Table-to-Table Conflicts (關聯表間的衝突)
 - 對應到前面的 Missing Data

資料命名上的衝突

| Books | | | |
|-----------|-------------|---------------|------------------|
| <i>no</i> | <i>name</i> | <i>author</i> | <i>publisher</i> |
| 1 | 鹿鼎記 | 金庸 | 遠流出版社 |
| 2 | 水滸傳 | 施耐庵 | 中庸出版社 |
| 3 | 紅樓夢 | 曹雪芹 | 春秋出版社 |
| 4 | 西遊記 | 吳承恩 | 聊齋出版社 |
| 5 | 水經注 | 酈道元 | 易經出版社 |
| 6 | 道德經 | 老子 | 大唐出版社 |
| 放在資料站 A | | | |

| Booklist | | | | |
|-----------|-----------------|---------------|--------------|------------------|
| <i>id</i> | <i>bookname</i> | <i>author</i> | <i>price</i> | <i>publisher</i> |
| 1 | 三國演義 | 羅貫中 | 120 | 古文出版社 |
| 2 | 水滸傳 | 施耐庵 | 170 | 中庸出版社 |
| 3 | 紅樓夢 | 曹雪芹 | 170 | 春秋出版社 |
| 4 | 西遊記 | 吳承恩 | 140 | 聊齋出版社 |
| 5 | 水經注 | 酈道元 | 120 | 易經出版社 |
| 6 | 道德經 | 老子 | 190 | 大唐出版社 |
| 放在資料站 B | | | | |

資料表示法上的衝突：如果在 **Books** 中的屬性 *no* 之型態為 **int**，但是在 **Booklist** 中的屬性 *id* 為 **char(5)**。

資料的編碼或資料內容的差異

Books

| <i>No</i> | <i>name</i> | <i>author</i> | <i>publisher</i> |
|-----------|-------------|---------------|------------------|
| 1 | 鹿鼎記 | 金庸 | 遠流 |
| 2 | 水滸傳 | 施耐庵 | 中庸 |
| 3 | 紅樓夢 | 曹雪芹 | 春秋 |
| 4 | 西遊記 | 吳承恩 | 聊齋 |
| 5 | 水經注 | 酈道元 | 易經 |
| 6 | 道德經 | 老子 | 大唐 |

放在資料站 A

Booklist

| <i>Id</i> | <i>bookname</i> | <i>author</i> | <i>price</i> | <i>publisher</i> |
|-----------|-----------------|---------------|--------------|------------------|
| 1 | 三國演義 | 羅貫中 | 120 | 古文出版社 |
| 2 | 水滸傳 | 施耐庵 | 170 | 中庸出版社 |
| 3 | 紅樓夢 | 曹雪芹 | 170 | 春秋出版社 |
| 4 | 西遊記 | 吳承恩 | 140 | 聊齋出版社 |
| 5 | 水經注 | 酈道元 | 120 | 易經出版社 |
| 6 | 道德經 | 老子 | 190 | 大唐出版社 |

放在資料站 B

Representation Conflicts

Books

| <i>No</i> | <i>name</i> | <i>author</i> | <i>publisher</i> |
|-----------|-------------|---------------|------------------|
| 1 | 鹿鼎記 | 金庸 | YL |
| 2 | 水滸傳 | 施耐庵 | CY |
| 3 | 紅樓夢 | 曹雪芹 | TC |
| 4 | 西遊記 | 吳承恩 | LT |
| 5 | 水經注 | 酈道元 | YC |
| 6 | 道德經 | 老子 | DT |

放在資料站 A

Booklist

| <i>Id</i> | <i>bookname</i> | <i>author</i> | <i>price</i> | <i>publisher</i> |
|-----------|-----------------|---------------|--------------|------------------|
| 1 | 三國演義 | 羅貫中 | 120 | 古文出版社 |
| 2 | 水滸傳 | 施耐庵 | 170 | 中庸出版社 |
| 3 | 紅樓夢 | 曹雪芹 | 170 | 春秋出版社 |
| 4 | 西遊記 | 吳承恩 | 140 | 聊齋出版社 |
| 5 | 水經注 | 酈道元 | 120 | 易經出版社 |
| 6 | 道德經 | 老子 | 190 | 大唐出版社 |

放在資料站 B

不同的日期使用法，如：DD/MM/YYYY、DD/MM/YY、MM/DD/YY，它們之間的差異也會造成困擾。

資料單位的衝突 (連續型資料)

| Books | | | | |
|-----------|-------------|---------------|--------------|------------------|
| <i>no</i> | <i>name</i> | <i>author</i> | <i>price</i> | <i>publisher</i> |
| 1 | 鹿鼎記 | 金庸 | 4 | 遠流出版社 |
| 2 | 水滸傳 | 施耐庵 | 5 | 中庸出版社 |
| 3 | 紅樓夢 | 曹雪芹 | 5 | 春秋出版社 |
| 4 | 西遊記 | 吳承恩 | 4.5 | 聊齋出版社 |
| 5 | 水經注 | 酈道元 | 4 | 易經出版社 |
| 6 | 道德經 | 老子 | 5.3 | 大唐出版社 |
| 放在資料站 A | | | | |

| Booklist | | | | |
|-----------|-------------|---------------|--------------|------------------|
| <i>no</i> | <i>name</i> | <i>author</i> | <i>price</i> | <i>publisher</i> |
| 1 | 三國演義 | 羅貫中 | 120 | 古文出版社 |
| 2 | 水滸傳 | 施耐庵 | 170 | 中庸出版社 |
| 3 | 紅樓夢 | 曹雪芹 | 170 | 春秋出版社 |
| 4 | 西遊記 | 吳承恩 | 140 | 聊齋出版社 |
| 5 | 水經注 | 酈道元 | 120 | 易經出版社 |
| 6 | 道德經 | 老子 | 190 | 大唐出版社 |
| 放在資料站 B | | | | |

Books 的 *price* 是以美金為單位 (可能是負責國外訂購的部門)，而 **Booklist** 的 *price* 卻是以新臺幣為單位。

資料單位的衝突 (離散型資料)

Bookstores

| <i>No</i> | <i>name</i> | <i>location</i> |
|-----------|-------------|-----------------|
| 1 | 元智書苑 | 中壢市 |
| 2 | 誠品書店 | 台北市 |
| 3 | 儒林出版社 | 台北市 |
| 4 | 松崗出版社 | 台北市 |
| 5 | 格致出版社 | 台北市 |
| 6 | 全民書局 | 新竹市 |

放在資料站 A

BookCompany

| <i>Id</i> | <i>title</i> | <i>location</i> |
|-----------|--------------|-----------------|
| 1 | 金石堂書店 | 松山區 |
| 2 | 誠品書店 | 大安區 |
| 3 | 儒林出版社 | 中正區 |
| 4 | 松崗出版社 | 大安區 |
| 5 | 格致出版社 | 文山區 |
| 6 | 陽明書局 | 士林區 |

放在資料站 B



不相容的資料

| Books | | | | |
|-----------|-------------|---------------|--------------|------------------|
| <i>no</i> | <i>name</i> | <i>author</i> | <i>price</i> | <i>publisher</i> |
| 1 | 鹿鼎記 | 金庸 | 180 | 遠流出版社 |
| 2 | 水滸傳 | 施耐庵 | 153 | 中庸出版社 |
| 3 | 紅樓夢 | 曹雪芹 | 153 | 春秋出版社 |
| 4 | 西遊記 | 吳承恩 | 126 | 聊齋出版社 |
| 5 | 水經注 | 酈道元 | 108 | 易經出版社 |
| 6 | 道德經 | 老子 | 171 | 大唐出版社 |
| 放在資料站 A | | | | |

| Booklist | | | | |
|-----------|-------------|---------------|--------------|------------------|
| <i>no</i> | <i>name</i> | <i>author</i> | <i>price</i> | <i>publisher</i> |
| 1 | 三國演義 | 羅貫中 | 120 | 古文出版社 |
| 2 | 水滸傳 | 施耐庵 | 170 | 中庸出版社 |
| 3 | 紅樓夢 | 曹雪芹 | 170 | 春秋出版社 |
| 4 | 西遊記 | 吳承恩 | 140 | 聊齋出版社 |
| 5 | 水經注 | 酈道元 | 120 | 易經出版社 |
| 6 | 道德經 | 老子 | 190 | 大唐出版社 |
| 放在資料站 B | | | | |



資料缺項 (Missing Data)

| Books | | | |
|---------------|------|-------|-----------|
| ISBN | name | pages | publisher |
| 975-723-044-1 | 鹿鼎記 | 350 | 遠流出版社 |
| 975-86-7532-2 | 水滸傳 | 490 | 中庸出版社 |
| 975-86-7532-3 | 紅樓夢 | 580 | 春秋出版社 |
| 975-86-7532-4 | 西遊記 | 480 | 聊齋出版社 |
| 975-86-7532-5 | 水經注 | 370 | 易經出版社 |
| 975-86-7532-6 | 道德經 | 235 | 大唐出版社 |
| 放在資料站 A | | | |

沒有 year 欄位

| Booklist | | |
|---------------|------|------|
| ISBN | name | year |
| 975-86-7532-1 | 三國演義 | 1980 |
| 975-86-7532-2 | 水滸傳 | 1992 |
| 975-86-7532-3 | 紅樓夢 | 1994 |
| 975-86-7532-4 | 西遊記 | 1995 |
| 975-86-7532-5 | 水經注 | 1990 |
| 975-86-7532-6 | 道德經 | 1992 |
| 放在資料站 B | | |

沒有 pages, publisher 欄位

角色互換的差異

每隊只存其對手是誰而已

Games

| <i>date</i> | <i>opponent</i> |
|-------------|-----------------|
| 10/5/93 | 興農牛 |
| 10/6/93 | 兄弟象 |

在統一獅的資料庫裡

Games

| <i>date</i> | <i>opponent</i> |
|-------------|-----------------|
| 10/5/93 | 統一獅 |
| 10/6/93 | 兄弟象 |

在興農牛的資料庫裡

Games

| <i>date</i> | <i>home</i> | <i>visitor</i> |
|-------------|-------------|----------------|
| 10/5/93 | 興農牛 | 統一獅 |
| 10/6/93 | 興農牛 | 兄弟象 |
| 10/6/93 | 統一獅 | 兄弟象 |

在整體資料站上

但整體站台
要看對戰雙方的
資料

關聯表與關聯表間的差異

Stock_Markets

| date | stock | index |
|---------|---------|-------|
| 10/5/93 | Tokyo | 24600 |
| 10/5/93 | Taipei | 4120 |
| 10/5/93 | Bangkok | 780 |
| 10/6/93 | Tokyo | 26065 |
| 10/6/93 | Taipei | 4321 |
| 10/6/93 | Bangkok | 803 |

放在資料站 A

Stock_Markets

| <i>date</i> | <i>Tokyo</i> | <i>Taipei</i> | <i>Bangkok</i> |
|-------------|--------------|---------------|----------------|
| 10/5/93 | 24600 | 4120 | 780 |
| 10/6/93 | 26065 | 4321 | 803 |

放在資料站 B

Tokyo

| <i>date</i> | <i>index</i> |
|-------------|--------------|
| 10/5/93 | 24600 |
| 10/6/93 | 26065 |

Taipei

| <i>date</i> | <i>index</i> |
|-------------|--------------|
| 10/5/93 | 4120 |
| 10/6/93 | 4321 |

Bangkok

| <i>date</i> | <i>index</i> |
|-------------|--------------|
| 10/5/93 | 780 |
| 10/6/93 | 803 |

存在資料站 C



異質性資料與綱要整合運算

- 在異質性資料庫中做資料與綱要的整合，
 - 通常都是使用所謂的 “外部合併” (Outerjoin) 運算，
 - 搭配上上述的各種轉換運算，包括：名稱轉換、型態轉換、單位轉換等
- 對於 “不相容資料” 與 “資料缺項” 兩大問題，會使得整合後的資料形成「不精確」(Imprecise) 的情況。我們將會說明一種特殊的「不確定值」一稱為「部份值」(Partial Values)，它可以用來做為整合處理時用以代表不精確資料的表示法。



外部合併運算 (Outerjoin)

- 由關聯式代數中的 Join 所延伸而得，
- 相對於外部合併運算來說，第五章中所定義的合併運算 (Join) 也可以稱為「內部合併」(Inner-join)
- 觀念在於：
 - 先將兩關聯表的所有相同欄位全部做 Inner Join 的動作，
 - 如果合併完後兩個關聯表中有某些記錄並沒有成為合併結果時，也要將它們全部收納到合併結果中，
 - 對於在一個關聯表中的各筆記錄，如果在另一個關聯表中找不到相對應的屬性時，則以虛值 (NULL) 將它填補
- 整個做法上有點像「聯集」(Union) 運算，所以才有整合的效果

看一個例子就知道

Books

| <i>ISBN</i> | <i>name</i> | <i>pages</i> | <i>publisher</i> |
|---------------|-------------|--------------|------------------|
| 975-723-044-1 | 鹿鼎記 | 350 | 遠流出版社 |
| 975-86-7532-2 | 水滸傳 | 490 | 中庸出版社 |
| 975-86-7532-3 | 紅樓夢 | 580 | 春秋出版社 |
| 975-86-7532-4 | 西遊記 | 480 | 聊齋出版社 |
| 975-86-7532-5 | 水經注 | 370 | 易經出版社 |
| 975-86-7532-6 | 道德經 | 235 | 大唐出版社 |

可整合

Booklist

| <i>ISBN</i> | <i>name</i> | <i>year</i> |
|---------------|-------------|-------------|
| 975-86-7532-1 | 三國演義 | 1980 |
| 975-86-7532-2 | 水滸傳 | 1992 |
| 975-86-7532-3 | 紅樓夢 | 1994 |
| 975-86-7532-4 | 西遊記 | 1995 |
| 975-86-7532-5 | 水經注 | 1990 |
| 975-86-7532-6 | 道德經 | 1992 |

可整合

| <i>ISBN</i> | <i>name</i> | <i>pages</i> | <i>publisher</i> | <i>year</i> |
|---------------|-------------|--------------|------------------|-------------|
| 975-86-7532-1 | 三國演義 | <NULL> | <NULL> | 1980 |
| 975-723-044-1 | 鹿鼎記 | 350 | 遠流出版社 | <NULL> |
| 975-86-7532-2 | 水滸傳 | 490 | 中庸出版社 | 1992 |
| 975-86-7532-3 | 紅樓夢 | 580 | 春秋出版社 | 1994 |
| 975-86-7532-4 | 西遊記 | 480 | 聊齋出版社 | 1995 |
| 975-86-7532-5 | 水經注 | 370 | 易經出版社 | 1990 |
| 975-86-7532-6 | 道德經 | 235 | 大唐出版社 | 1992 |

整合後
資料互補

用來整合的
合併屬性是
(*ISBN*, *name*)

Outerjoin 的定義 (Full Outerjoin)

- 定義 14.1 : [外部合併] (Outerjoin)—記做 “ ∞ ”，假定有兩個關聯表 $R(X, Y)$ 與 $S(X, Z)$ ，則令 T 為 R 與 S 做合併的結果，也就是說：

$$T = (R \bowtie_{R.X \theta S.X} S) ;$$

另外，令

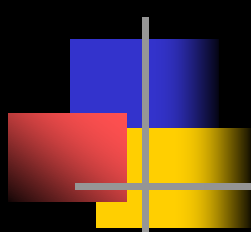
$$U = R - \pi_{X,Y} (R \bowtie_{R.X \theta S.X} S)$$

$$V = S - \pi_{X,Z} (R \bowtie_{R.X \theta S.X} S)$$

則 $R(X, Y)$ 與 $S(X, Z)$ 的外部合併結果

$R(X, Y) \infty_{R.X \theta S.X} S(X, Z)$ 定義為：

$$T \cup (U \times \{<NULL>\}) \cup (\{<NULL>\} \times V)$$



Full Outerjoin 產生資訊互補

- 透過完整的外部合併後，由於雙方互補不足而得到更多的資訊了，
 - 像“西遊記”一書本來在 **Books** 中只知道 ISBN 編號 *ISBN*、書名 *name*、頁數 *pages*，以及出版社 *publisher* 而已，不知道它的出版年份 *year*；
 - 經過外部運算後我們發現 **Booklist** 中的資料正好補足了此項缺憾。
- 所以我們知道外部運算很適合用來做異質性資料庫中的資料整合，不但可以將資料融合在一起，同時還可以因為互補的關係，導出更多的資訊出來

左方外部合併 (Left Outer Join)

Books

| ISBN | name | pages | publisher |
|---------------|------|-------|-----------|
| 975-723-044-1 | 鹿鼎記 | 350 | 遠流出版社 |
| 975-86-7532-2 | 水滸傳 | 490 | 中庸出版社 |
| 975-86-7532-3 | 紅樓夢 | 580 | 春秋出版社 |
| 975-86-7532-4 | 西遊記 | 480 | 聊齋出版社 |
| 975-86-7532-5 | 水經注 | 370 | 易經出版社 |
| 975-86-7532-6 | 道德經 | 235 | 大唐出版社 |

Booklist

| ISBN | name | year |
|---------------|------|------|
| 975-86-7532-1 | 三國演義 | 1980 |
| 975-86-7532-2 | 水滸傳 | 1992 |
| 975-86-7532-3 | 紅樓夢 | 1994 |
| 975-86-7532-4 | 西遊記 | 1995 |
| 975-86-7532-5 | 水經注 | 1990 |
| 975-86-7532-6 | 道德經 | 1992 |

這筆留下

可整合

←這筆丟棄

可整合

| ISBN | name | pages | publisher | year |
|--------------------------|-----------------|--------|-----------|-----------------|
| 975-86-7532-1 | 三國演義 | <NULL> | <NULL> | 1980 |
| 975-723-044-1 | 鹿鼎記 | 350 | 遠流出版社 | <NULL> |
| 975-86-7532-2 | 水滸傳 | 490 | 中庸出版社 | 1992 |
| 975-86-7532-3 | 紅樓夢 | 580 | 春秋出版社 | 1994 |
| 975-86-7532-4 | 西遊記 | 480 | 聊齋出版社 | 1995 |
| 975-86-7532-5 | 水經注 | 370 | 易經出版社 | 1990 |
| 975-86-7532-6 | 道德經 | 235 | 大唐出版社 | 1992 |

整合後
資料互補

用來整合的
合併屬性是
(ISBN, name)

右方外部合併 (Right Outer Join)

Books

| ISBN | name | pages | publisher |
|---------------|------|-------|-----------|
| 975-723-044-1 | 鹿鼎記 | 350 | 遠流出版社 |
| 975-86-7532-2 | 水滸傳 | 490 | 中庸出版社 |
| 975-86-7532-3 | 紅樓夢 | 580 | 春秋出版社 |
| 975-86-7532-4 | 西遊記 | 480 | 聊齋出版社 |
| 975-86-7532-5 | 水經注 | 370 | 易經出版社 |
| 975-86-7532-6 | 道德經 | 235 | 大唐出版社 |

Booklist

| ISBN | name | year |
|---------------|------|------|
| 975-86-7532-1 | 三國演義 | 1980 |
| 975-86-7532-2 | 水滸傳 | 1992 |
| 975-86-7532-3 | 紅樓夢 | 1994 |
| 975-86-7532-4 | 西遊記 | 1995 |
| 975-86-7532-5 | 水經注 | 1990 |
| 975-86-7532-6 | 道德經 | 1992 |

這筆丟棄

可整合

←這筆留下

可整合

| ISBN | name | pages | publisher | year |
|--------------------------|----------------|----------------|------------------|-------------------------|
| 975-86-7532-1 | 三國演義 | <NULL> | <NULL> | 1980 |
| 975-723-044-1 | 鹿鼎記 | 350 | 遠流出版社 | <NULL> |
| 975-86-7532-2 | 水滸傳 | 490 | 中庸出版社 | 1992 |
| 975-86-7532-3 | 紅樓夢 | 580 | 春秋出版社 | 1994 |
| 975-86-7532-4 | 西遊記 | 480 | 聊齋出版社 | 1995 |
| 975-86-7532-5 | 水經注 | 370 | 易經出版社 | 1990 |
| 975-86-7532-6 | 道德經 | 235 | 大唐出版社 | 1992 |

整合後
資料互補

用來整合的
合併屬性是
(ISBN, name)

SQL 中的 Full OuterJoin 運算

- `select isnull(Books.ISBN, Booklist.ISBN) as ISBN,
isnull(Books.name, Booklist.name) as name,
Books.pages, Books.publisher, Booklist.year
from Books Full Outer Join Booklist
on Books.ISBN = Booklist.ISBN`

| <i>ISBN</i> | <i>name</i> | <i>pages</i> | <i>publisher</i> | <i>year</i> |
|---------------|-------------|--------------|------------------|-------------|
| 975-86-7532-1 | 三國演義 | <NULL> | <NULL> | 1980 |
| 975-723-044-1 | 鹿鼎記 | 350 | 遠流出版社 | <NULL> |
| 975-86-7532-2 | 水滸傳 | 490 | 中庸出版社 | 1992 |
| 975-86-7532-3 | 紅樓夢 | 580 | 春秋出版社 | 1994 |
| 975-86-7532-4 | 西遊記 | 480 | 聊齋出版社 | 1995 |
| 975-86-7532-5 | 水經注 | 370 | 易經出版社 | 1990 |
| 975-86-7532-6 | 道德經 | 235 | 大唐出版社 | 1992 |

SQL 中的 Left OuterJoin 運算

- `select Books.*, Booklist.year`
`from Books Left Outer Join Booklist`
`on Books.ISBN = Booklist.ISBN`

| <i>ISBN</i> | <i>name</i> | <i>pages</i> | <i>publisher</i> | <i>year</i> |
|--------------------------|-------------|-------------------------|-------------------------|-----------------|
| 975-86-7532-1 | 三國演義 | <NULL> | <NULL> | 1980 |
| 975-723-044-1 | 鹿鼎記 | 350 | 遠流出版社 | <NULL> |
| 975-86-7532-2 | 水滸傳 | 490 | 中庸出版社 | 1992 |
| 975-86-7532-3 | 紅樓夢 | 580 | 春秋出版社 | 1994 |
| 975-86-7532-4 | 西遊記 | 480 | 聊齋出版社 | 1995 |
| 975-86-7532-5 | 水經注 | 370 | 易經出版社 | 1990 |
| 975-86-7532-6 | 道德經 | 235 | 大唐出版社 | 1992 |

SQL 中的 Right OuterJoin 運算

- select isnull(Books.ISBN, Booklist.ISBN) as ISBN,
isnull(Books.name, Booklist.name) as name,
Books.pages, Books.publisher, Booklist.year
from Books Right Outer Join Booklist
on Books.*ISBN* = Booklist.*ISBN*

| <i>ISBN</i> | <i>name</i> | <i>pages</i> | <i>publisher</i> | <i>year</i> |
|--------------------------|-------------|----------------|------------------|-------------------------|
| 975-86-7532-1 | 三國演義 | <NULL> | <NULL> | 1980 |
| 975-723-044-1 | 鹿鼎記 | 350 | 遠流出版社 | <NULL> |
| 975-86-7532-2 | 水滸傳 | 490 | 中庸出版社 | 1992 |
| 975-86-7532-3 | 紅樓夢 | 580 | 春秋出版社 | 1994 |
| 975-86-7532-4 | 西遊記 | 480 | 聊齋出版社 | 1995 |
| 975-86-7532-5 | 水經注 | 370 | 易經出版社 | 1990 |
| 975-86-7532-6 | 道德經 | 235 | 大唐出版社 | 1992 |



加上 Where 子句

- Outer join 合併條件是放在 on 子句中，若有其它的過濾條件要加在 where 子句中：

```
select isnull(Books.ISBN, Booklist.ISBN) as ISBN,  
       isnull(Books.name, Booklist.name) as name,  
       Books.pages, Books.publisher, Booklist.year  
from Books Right Outer Join Booklist  
on Books.ISBN=Booklist.ISBN  
where year > 1970
```



部份值 (Partial Values)

- 最早是由 J. Grant [J. Grant (1979)] 提出來的，後來由 L.G. DeMichiel [L.G. DeMichiel (1989)] 將它應用在異質性資料庫整合上面
- 定義 5.2：[部份值] (Partial Values)——一個部份值 $\eta = [v_1, v_2, \dots, v_n]$ 是一個不確定的值，以一個集合 $[v_1, v_2, \dots, v_n]$ 來表示，而集合中的元素 v_i 有一個是該部份值的真正值，但我們不知道是那一個。該集合稱為該部份值的值域 (Domain)，我們以 $v(\eta) = \{v_1, v_2, \dots, v_n\}$ 表示之。
- 一個 Applicable Null Value 可以看成一個部份值，此部份值的值域便是其屬性值域 (Attribute Domain)

舉例說明

- 「三國演義」的出版年份確定一定是 1980 或 1985 其中一年，只是我們不知道是那一年

Booklist

| <i>ISBN</i> | <i>name</i> | <i>year</i> |
|---------------|-------------|--------------|
| 975-86-7532-1 | 三國演義 | [1980, 1985] |
| 975-86-7532-2 | 水滸傳 | 1992 |
| 975-86-7532-3 | 紅樓夢 | 1994 |
| 975-86-7532-4 | 西遊記 | 1995 |
| 975-86-7532-5 | 水經注 | 1990 |
| 975-86-7532-6 | 道德經 | 1992 |

← 部份值

異質性整合的例子

CS_Researchers

| <i>Name</i> | <i>region</i> | <i>specialty</i> | <i>age</i> | <i>degree</i> |
|-------------|---------------|------------------|------------|---------------|
| Andy | 臺灣 | AI | (Null) | MS |
| Frank | 臺灣 | DB | 26 | PhD |
| Jesse | 臺灣 | SE | 30 | MS |
| Mike | 日本 | DB | 32 | PhD |
| John | 韓國 | SE | 28 | PhD |

放在資料站A

Taiwan_Researchers

| <i>Name</i> | <i>city</i> | <i>specialty</i> | <i>age</i> | <i>affiliation</i> |
|-------------|-------------|------------------|------------|--------------------|
| Andy | 台北 | CS | 25 | NTU |
| Frank | 新竹 | CS | 28 | NCTU |
| Annie | 高雄 | CS | 27 | NSYSU |
| Paul | 台北 | EE | 30 | NTU |
| Lisa | 新竹 | IE | 26 | NTHU |

放在資料站 B

異質性整合的例子 (續)

TW_CS_Researchers

| <i>Name</i> | <i>region</i> | <i>specialty</i> | <i>age</i> | <i>degree</i> |
|-------------|---------------|------------------|------------|---------------|
| Andy | 臺灣 | AI | (Null) | MS |
| Frank | 臺灣 | DB | 26 | PhD |
| Jesse | 臺灣 | SE | 30 | MS |

放在資料站 A

CS_TW_Researchers

| <i>Name</i> | <i>city</i> | <i>specialty</i> | <i>age</i> | <i>affiliation</i> |
|-------------|-------------|------------------|------------|--------------------|
| Andy | 台北 | CS | 25 | NTU |
| Frank | 新竹 | CS | 28 | NCTU |
| Annie | 高雄 | CS | 27 | NSYSU |

放在資料站 B

異質性整合的例子 (續)

TW_CS_Researchers_2

| <i>name</i> | <i>city</i> | <i>specialty</i> | <i>age</i> | <i>degree</i> |
|-------------|--------------|------------------|------------|---------------|
| Andy | [台北, 新竹, 高雄] | AI | (Null) | MS |
| Frank | [台北, 新竹, 高雄] | DB | 26 | PhD |
| Jesse | [台北, 新竹, 高雄] | SE | 30 | MS |

放在資料站 A

CS_TW_Researchers_2

| <i>Name</i> | <i>city</i> | <i>Specialty</i> | <i>age</i> | <i>affiliation</i> |
|-------------|-------------|------------------|------------|--------------------|
| Andy | 台北 | [AI, DB, SE] | 25 | NTU |
| Frank | 新竹 | [AI, DB, SE] | 28 | NCTU |
| Annie | 高雄 | [AI, DB, SE] | 27 | NSYSU |

放在資料站 B



異質性整合的例子 (續)

- 使用 Outerjoin 運算來整合 (但必須在 ORDBMS 下)

Taiwan_CS_Researchers

| <i>Name</i> | <i>city</i> | <i>specialty</i> | <i>age</i> | <i>degree</i> | <i>affiliation</i> |
|-------------|-------------|------------------|------------|---------------|--------------------|
| Andy | 台北 | AI | 25 | MS | NTU |
| Frank | 新竹 | DB | [26, 28] | PhD | NCTU |
| Jesse | [台北,新竹,高雄] | SE | 30 | MS | (Null) |
| Annie | 高雄 | [AI,DB,SE] | 27 | (Null) | NSYSU |

異質性整合的例子 (續)

- “請問年齡大於 28 歲或研究資料庫系統的研究人員有那些？”

$\sigma_{\text{age} > 28 \text{ or specialty} = \text{"DB"}} (\text{Taiwan_CS_Researchers})$

| Name | city | specialty | age | degree | affiliation | 可靠度 |
|-------|------------|------------|----------|--------|-------------|-------|
| Frank | 新竹 | DB | [26, 28] | PhD | NCTU | True |
| Jesse | [台北,新竹,高雄] | SE | 30 | MS | (Null) | True |
| Annie | 高雄 | [AI,DB,SE] | 27 | (Null) | NSYSU | Maybe |

- 也可以加上機率來推導與計算可能的程度
(是很好的論文題目)

異質性整合的例子 (續)

- “請問年齡大於 28 歲或研究資料庫系統的研究人員有那些？”

$\sigma_{\text{age} > 28 \text{ or specialty} = \text{"DB"}} (\text{Taiwan_CS_Researchers})$

| <i>Name</i> | <i>city</i> | <i>specialty</i> | <i>age</i> | <i>degree</i> | <i>affiliation</i> | 可靠度 |
|-------------|---|---|------------------------|---------------|--------------------|------|
| Frank | 新竹 | DB | $[26^{1/2}, 28^{1/2}]$ | PhD | NCTU | 1 |
| Jesse | $[\text{台北}^{1/3}, \text{新竹}^{1/3}, \text{高雄}^{1/3}]$ | SE | 30 | MS | (Null) | 1 |
| Annie | 高雄 | $[\text{AI}^{0.2}, \text{DB}^{0.5}, \text{SE}^{0.3}]$ | 27 | (Null) | NSYSU | 0.35 |

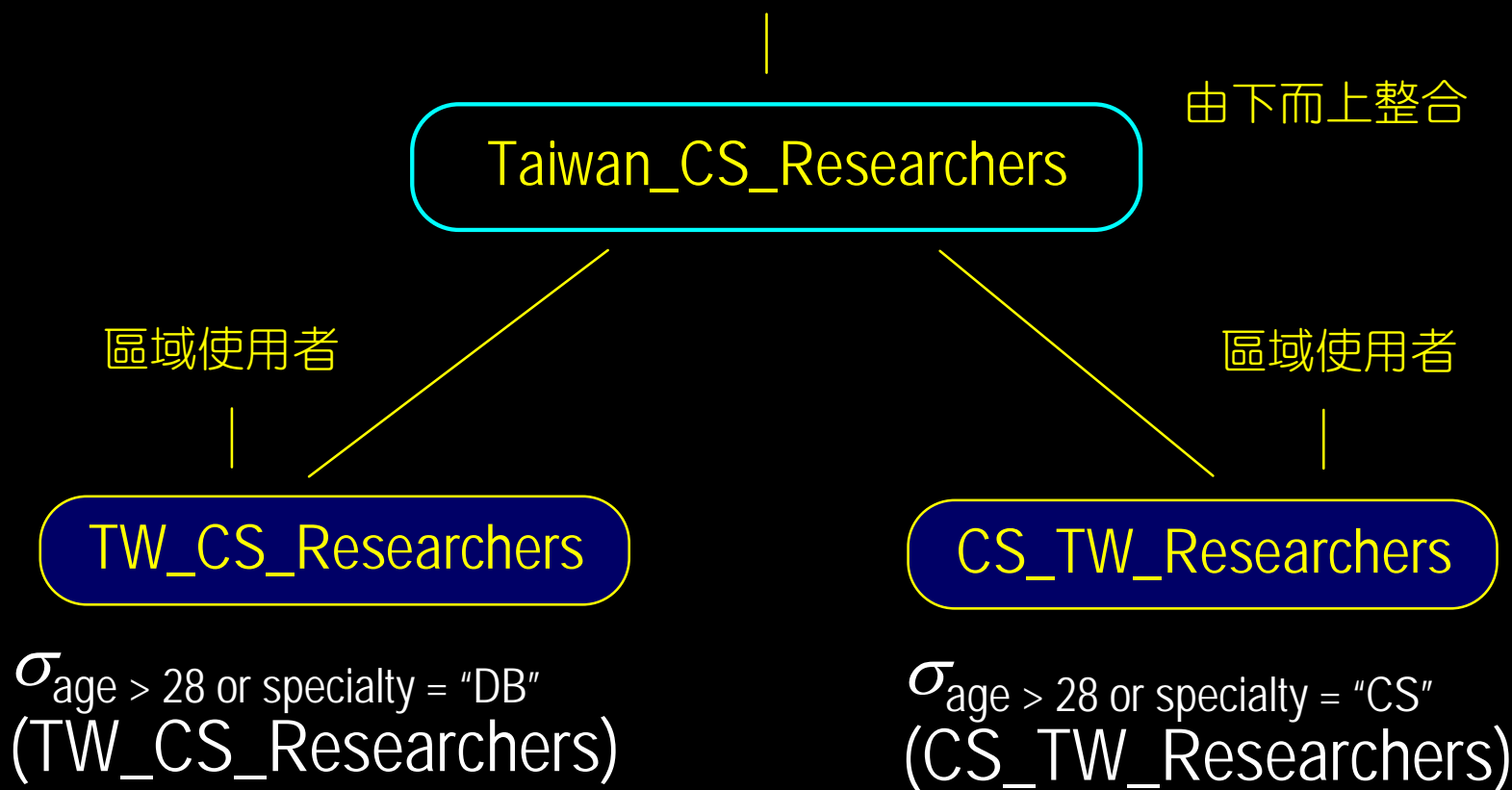


查詢的拆解 (Query Decomposition)

- 由於整體端並沒有真正存放資料，所以查詢必須要經過拆解後，分別傳向區域端去執行，整體端取得資料後再整合
- $\sigma_{\text{age} > 28 \text{ or specialty} = \text{"DB"}}(\text{Taiwan_CS_Researchers})$ 要拆解成
 - $\sigma_{\text{age} > 28 \text{ or specialty} = \text{"DB"}}(\text{TW_CS_Researchers})$ 與
 - $\sigma_{\text{age} > 28 \text{ or specialty} = \text{"CS"}}(\text{CS_TW_Researchers})$

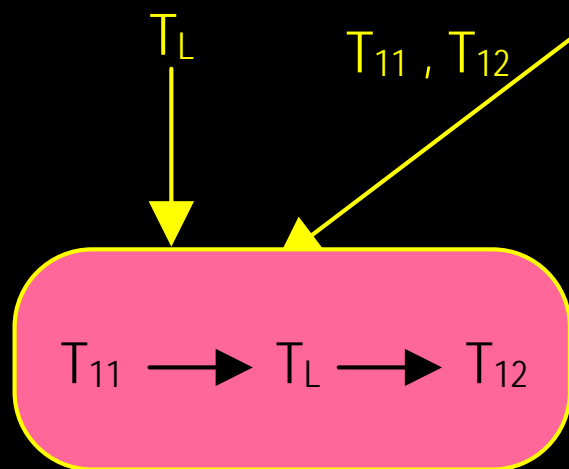
查詢拆解圖解

$\sigma_{\text{age} > 28 \text{ or specialty} = \text{"DB"}} (\text{Taiwan_CS_Researchers})$

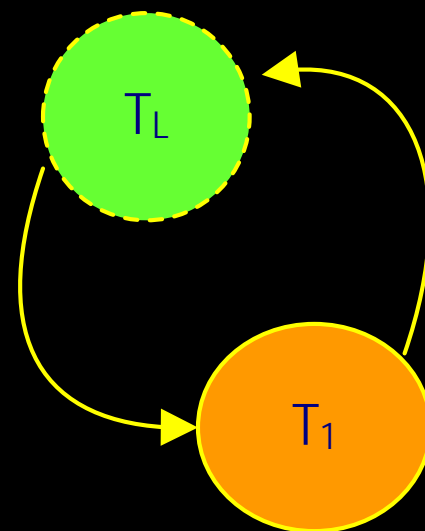


異質性資料庫系統的異動處理

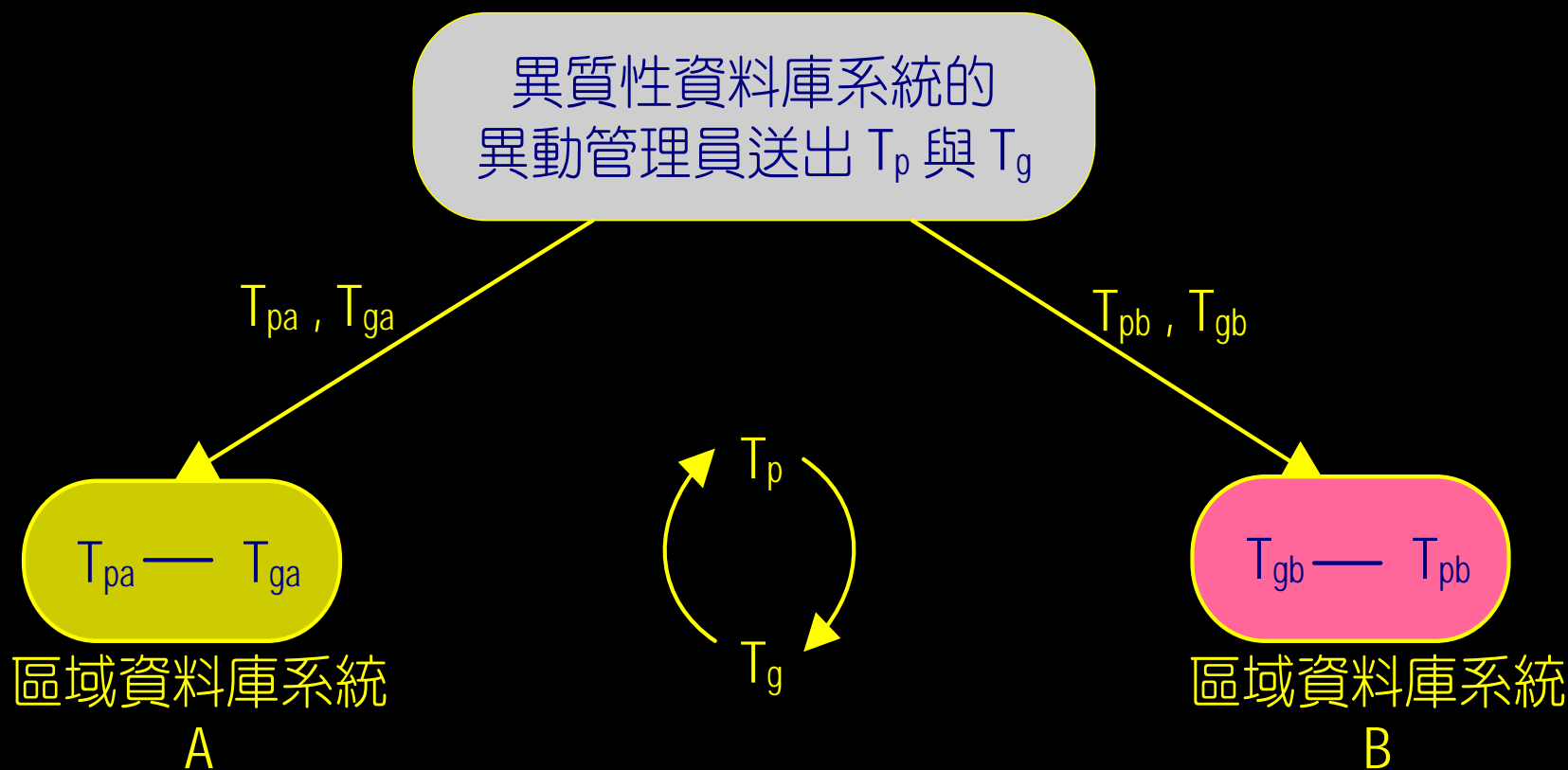
異質性資料庫系統的
異動管理員送出 T_1



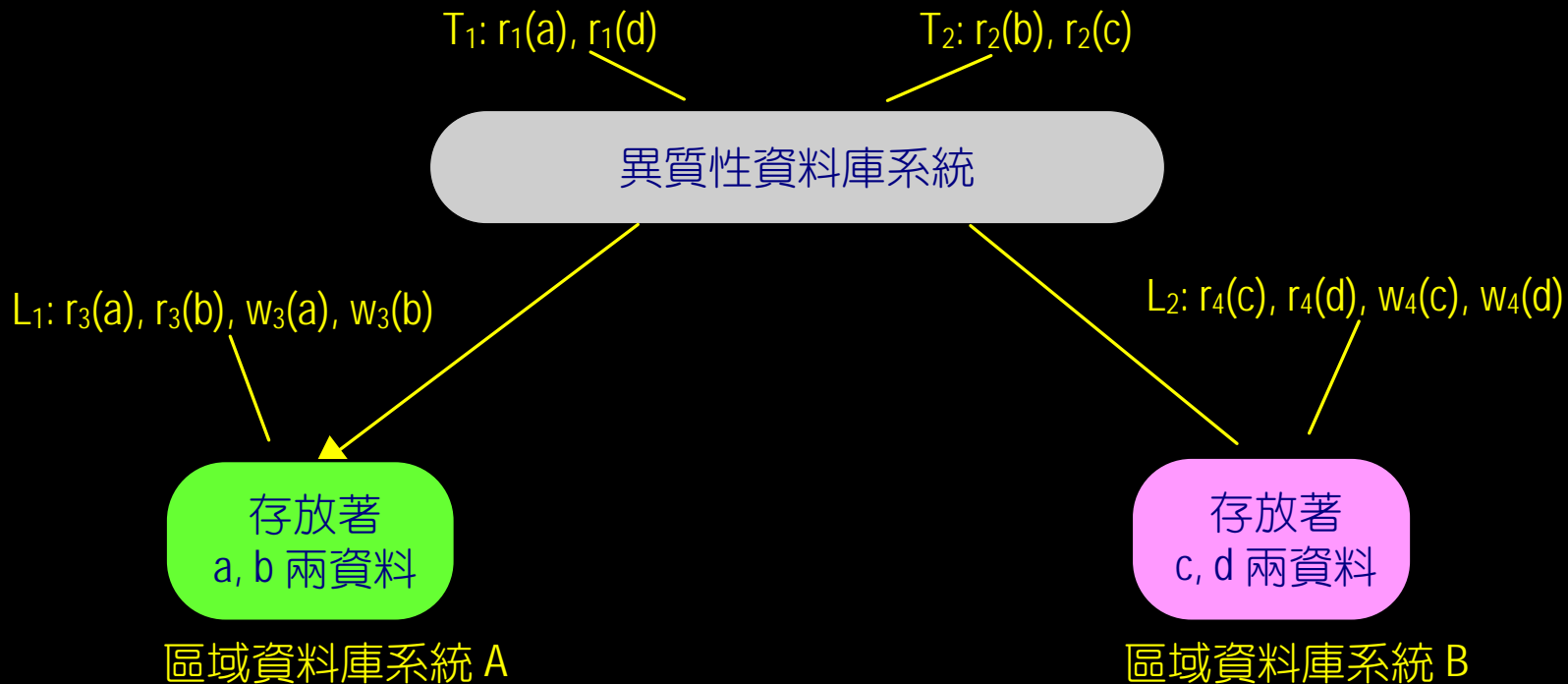
區域資料庫系統



異質性資料庫系統的異動處理



區域異動的干擾所造成的困擾



在資料站 A: $r_1(a), | r_3(a), r_3(b), w_3(a), w_3(b), | r_2(b)$

在資料站 B: $r_2(c), | r_4(c), r_4(d), w_4(c), w_4(d), | r_1(d)$

$T_1 \Rightarrow L_1 \Rightarrow T_2$

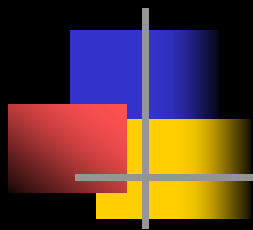
$T_2 \Rightarrow L_2 \Rightarrow T_1$

Cycle !!!



異質性資料庫系統 vs. 工作流程

- 異質性資料庫系統—各類資料庫系統的**縱向整合**（**資料整合**）
- 工作流程軟體—各類資料庫系統的**橫向整合**（**程序整合**）。如：IBM FlowMark, Lotus Notes Domino, HP WorkManager, Webflow, ...



本章結束
The End.