

## 資料環 - 一個支援資料立方體的操作架構

### Data Rin - A Structure for Data Cube Operation

邱明政 謝金雲<sup>1</sup> 鄭有進

國立台北科技大學電子工程系

{hsieh,yccheng}@en.ntut.edu.tw

周忠信

東海大學資訊科學系

jwo@mail.thu.edu.tw

#### 摘要

線上分析處理(OLAP)所分析的資料一般以資料立方體(Data Cube)的型態儲存於系統，而資料立方體之訴求重點則是多維的資料顯示效能。惟目前線上分析處理系統所提供的資料操作方式主要以表格為主，即分析者先在表格中進行資料瀏覽，俟觀察到欲分析的資料時，再利用圖表顯示工具將資料製作成統計圖表；若圖表不符所求，再回頭利用表格進行資料分析。此一分析方式不僅過程繁複，而且分析者也不易從表格中得知資料的趨勢。此外，一般的統計圖表也不易達到多維的資料顯示效果。

本研究提出的資料環模型，其目的即是為資料立方體提供一個更有效益的作業環境。此模型之主要特色在於將以往以表格為主，圖表為輔的交互式分析方式，改變為以資料環為主體之整合式分析環境直接進行資料之瀏覽、操作及顯示作業。資料環之資料展現方式非但克服了一般線上分析處理系統無法顯示多維度資料圖表的問題，並讓分析者憑藉直觀即可輕易進行資料分析及了解資料趨勢。

關鍵詞：資料環、線上分析處理、資料立方體

#### 1. 簡介

線上分析處理(OLAP)[1]是近年來資料倉儲(Data Warehouse)的重要分析工具，至於以往資料分析方式大抵都是藉人工利用 SQL 語言產生資料報表，再由各類的報表整合達到資料分析的目的。線上分析處理，顧名思義是一種即時的線上資料庫處理方式，利用線上分析處

理，使用者不必再進行複雜的資料報表整理及分析，即可從線上分析處理系統經由資料操作(Data Manipulation)得到分析結果。

線上分析處理將資料型態以表及關係的方式轉換為階層式的多維資料型態，讓使用者可以從不同的角度觀察資料，並藉由資料的階層化及分類化，讓使用者可觀察不同層級的資料內容，此多維階層資料型態稱為資料立方體[2]。

目前線上分析處理所提供的資料操作方式主要以表格為主，分析者在表格中以 Drill Down 及 Roll U 方式進行資料瀏覽，俟觀察到欲分析的資料時，再利用圖表顯示工具將資料製作成統計圖表，若圖表不符合要求，再利用表格進行資料分析。然而，以表格方式顯示之資料為數值或文字資料，分析者不易從中得知資料的趨勢，若能經由圖表方式顯示其效果必然更佳。一般的統計圖表所能表達的資料維度大多為二維及三維，因此必須藉由表格的資料操作，將表格的資料維度降至二維或三維，再經由統計圖表將資料顯示。由於資料立方體所強調的重點在於多維的資料顯示，而一般統計圖表卻不易達到多維的資料顯示，因此，針對能同時提供資料顯示及資料操作的多維資料模型顯得格外重要。

在本論文中，我們提出了支援資料立方體的一種操作架構，稱為資料環(Data Ring)。資料環以直條圖的顯示方式為基礎，透過圓環的串接，衍生出多維的資料模型架構。分析者可從圓環看出資料的分類及群組，並瀏覽各資料統計內容。我們針對資料環制訂各式符合資料立方體之操作方式，並以直觀及便利為訴求重點。其

<sup>1</sup> 主要聯絡人

操作方式包括資料環結構操作、資料環標記及資料環分析等。以往線上分析處理是以表格操作為主、圖表顯示為輔的交互方式進行資料分析，資料環則是將資料操作及顯示合而為一，達到一致的分析環境。資料環提供具體的圖形顯示方式，不僅更符合資料立方體的特性，並帶給分析者更快速便捷的分析效果。

即銷售量。

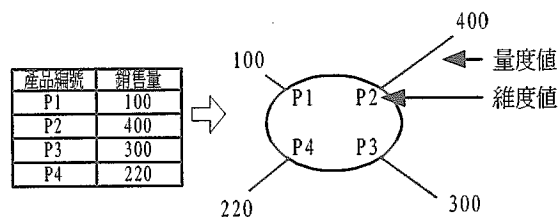


圖 1 資料環之基本圖形範例

## 2. 資料環之視覺表達方式

有關資料環之基本元件及建構方式，極易藉例子來說明。圖 1 左側為一簡單的資料表格，其內容為產品編號及銷售量的統計結果。根據資料立方體的定義，產品編號為維度(dimension)，銷售量為量度(measure)。圖 1 右側即是該表格所對應之資料環，環內所標示內容為維度值，即產品編號，環上的直線條長度代表量度值，

通常我們會根據多個資料欄位產生資料報表，例如若於圖 1 資料報表中增加一庫位欄位如圖 2 左側所示，則維度為庫位及產品編號，量度為銷售量。其對應之資料環結構只需新生一庫位資料環，並將分別代表各庫位產品銷售情況之四個小環串起構即可，如圖 2 右側所示。

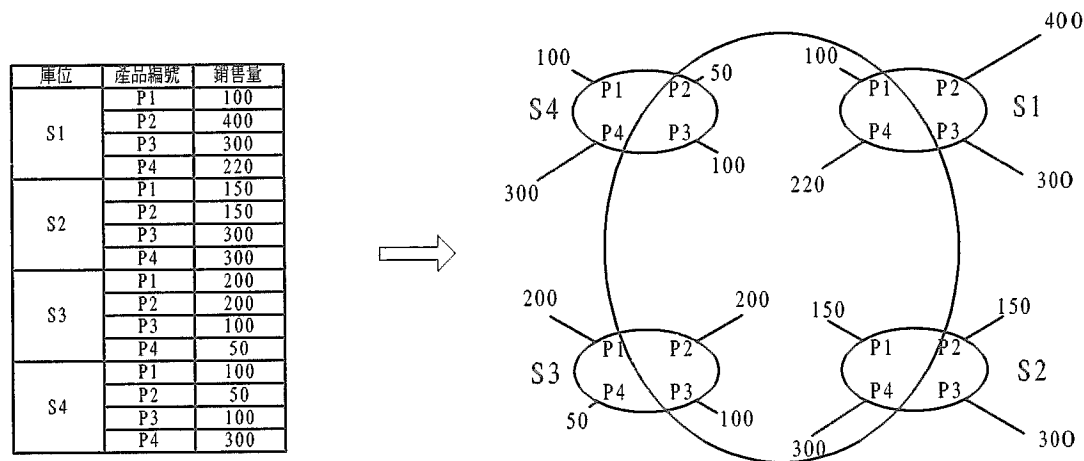


圖 2 根據庫位、產品編號及銷售量所構成之資料環

資料環所對應之資料報表其欄位由左至右分成兩群，一是維度群，另一是量度群，如圖 3(a)所示。圖 3(b)所示為各維度欄位與資料環之相對位置關係。資料環架構中，每一個環即代表了一個維度，環中顯示所代表之資料。最大環（即最內層者）稱為根環(Root Ring)，它代表最左邊的維度欄，而向外層延伸之各環則依序代表由左至右之其他維度欄；最小環上的直條線即是量度。相鄰兩層資料環中，居內層者稱為父環(Super Ring)，居外層者稱為子環(Sub Ring)。由此可明顯看出資料環本身

是以一致的方式顯示資料報表的內容，並且透過大環串接小環的方式達成多維的資料顯示。

資料環在多個量度上顯示方式略有不同，圖 1 及圖 2 所示皆為單一量度，在單一量度時，資料環每一條直線都代表相同之量度，而資料環顯示多個量度時，則每一個直線所代表的量度值則不同，我們藉由增加一子環，環上直線代表所有的量度值，因此資料若有多個量度，則環會多增加一個，以利顯示，如 4 所示即為一多量度的範例。

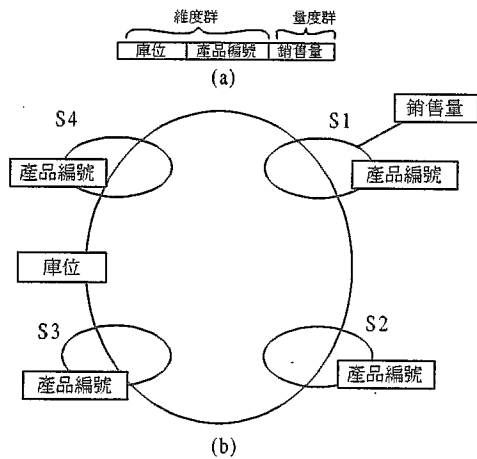


圖 3 欄位表頭與資料環的對應關係

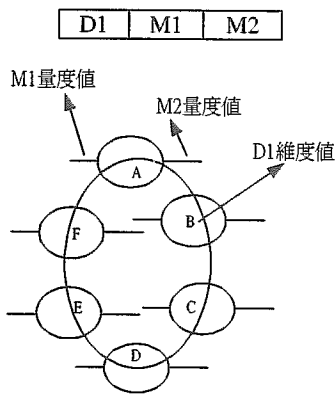


圖 4 由 D1、M1 及 M2 組成之多量度資料環範例

### 3. 資料環操作方式

資料環除提供資料的顯示方式外，也提供各式以簡便、速捷為訴求重點的操作方法，包括資料環結構操作、資料環標記、資料環分析等，以達成分析處理機能。

#### 3.1 資料環結構操作

本節針對資料環結構的變化說明各項操作方式，其結構的主要變化在於維度及量度的增減；透過此結構操作，可讓資料環根據維度及量度的變化決定顯示的內容。

##### ● 增加維度

增加維度即於資料環中增加一個新的維度，此新加入的維度可放置於維度群中的任意位置。在增加維度時，可在維度內加入限制條件，其限制條件可如 SQL 之

WHERE 子句敘述，例如  $R_i$  為年度欄位，則可限制“ $R_i = 1999$ ”或“ $R_i$  Between 1997 and 1999”。另外，在每一個維度的階層中，屬於同一個維度階層中的其它維度，如時間維度中的年、季、月及日等，可同時放置於維度群的不同位置。例如可將年與季同時放置於維度群中並列顯示。圖 5 所示為一個增加維度示意圖，由一原始維度為 D2、D1 及量度為 M1 的資料報表範例，在加入 D3 維度於維度群的第一個位置後，成為一個其維度為 D3、D2、D1，量度為 M1 的資料報表。增加維度其目的在於透過新增加的維度將資料劃分的更加精細，讓使用者可得到更詳細的資料。

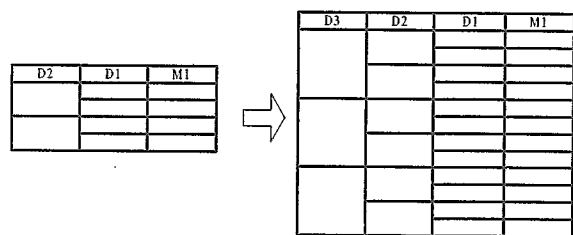


圖 5 增加維度示意圖

##### ● 刪除維度

刪除維度即於維度群中刪除某一維度，被刪除的維度可為維度群中的任一個。圖 6 所示為一原始維度為 D3、D2、D1 及量度為 M1 的資料報表範例，在刪除 D3 維度後，成為一個其維度為 D2、D1，量度為 M1 的資料報表。刪除維度之目的在於呈現一個較精簡的資料，以便讓使用者可快速了解資訊內容。

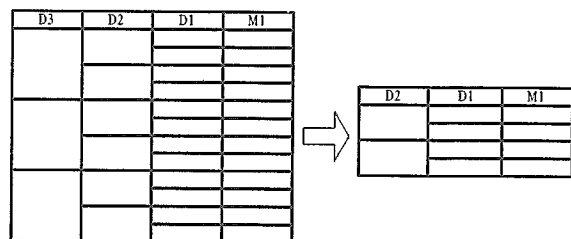


圖 6 刪除維度示意圖

##### ● 增加量度

為了讓使用者同時可觀察多個量度，可於量度群中增加一個新的量度。此新加入的量度可放置於量度群中的任意位置。圖 7 所示為一原始維度為 D2、D1 及量度為 M1 的資料報表範例，於量度群的第 2 位置增加 M2 量度後，成為一個其維度為 D2 及 D1，量度為 M1、

M2 的資料報表。

| D2 | D1 | M1 |
|----|----|----|
|    |    |    |
|    |    |    |
|    |    |    |
|    |    |    |

| D2 | D1 | M1 | M2 |
|----|----|----|----|
|    |    |    |    |
|    |    |    |    |
|    |    |    |    |
|    |    |    |    |

圖 7 增加量度示意圖

● 刪除量度

刪除量度表示不再觀察該量度，我們可以於量度群中刪除某一量度，被刪除的量度可為量度群中的任意一個。圖 8 所示為一原始維度為 D2 及 D1 與量度為 M1 及 M2 之資料報表範例，在刪除 M2 欄位後，成為一個其維度為 D2、D1，量度為 M1 的資料報表。

| D2 | D1 | M1 | M2 |
|----|----|----|----|
|    |    |    |    |
|    |    |    |    |
|    |    |    |    |
|    |    |    |    |

| D2 | D1 | M1 |
|----|----|----|
|    |    |    |
|    |    |    |
|    |    |    |
|    |    |    |

圖 8 刪除量度示意圖

3.2 資料環標記方式

資料環標記方式主要目的是讓使用者可以自由選擇欲處理的維度或量度以便進行資料分析。基本標記方式有三種，分別是量度標記、維度標記及維度個別資料標記。

● 量度標記

量度標記是讓使用者指出欲顯示或處理之單一或多個量度。量度標記後，代表於分析時處理該量度，未標記量度則不處理。圖 9 所示為一維度為 D3、D2 及 D1，量度為 M1、M2 及 M3，並於 M1 及 M3 標記量度之資料報表範例。

| D3 | D2 | D1 | M1 | M2 | M3 |
|----|----|----|----|----|----|
|    |    |    |    |    |    |
|    |    |    |    |    |    |
|    |    |    |    |    |    |
|    |    |    |    |    |    |

圖 9 量度標記示意圖

● 維度標記

維度標記表示標記該維度之所有資料，其功用是讓使用者指出欲處理之維度。維度被標記表示使用者於分析時以被標記之維度為限，即根據被標記之維度計算其量度。而於維度小計總合運算則表示該維度以“所有”代表之。圖 10 所示為一維度為 D3、D2 及 D1，量度為 M1、M2 及 M3，並於 D2 標記維度之資料報表範例。

| D3 | D2 | D1 | M1 | M2 | M3 |
|----|----|----|----|----|----|
|    |    |    |    |    |    |
|    |    |    |    |    |    |
|    |    |    |    |    |    |
|    |    |    |    |    |    |

圖 10 維度標記示意圖

● 維度個別資料標記

維度個別資料標記是讓使用者標記某維度內的若干個別資料，以表示於分析時以被標記之資料為限，亦即僅根據被標記之資料計算其量度。圖 11 所示為一維度為 D3、D2 及 D1，量度為 M1、M2 及 M3，並於 D2 維度上標記個別資料之資料報表範例。

| D3 | D2 | D1 | M1 | M2 | M3 |
|----|----|----|----|----|----|
|    |    |    |    |    |    |
|    |    |    |    |    |    |
|    |    |    |    |    |    |
|    |    |    |    |    |    |

圖 11 維度個別資料標記示意圖

3.3 資料環之分析方式

資料環分析包含一般的集合(Aggregation)運算及資料立方體之維度小計總合運算，當使用者在資料環上完成標記動作之後，便可在所選定之資料上執行上述運算，並顯示其運算結果。透過不斷的操作及運算，我們即可於資料環中獲得各種統計資訊。以下說明各運算之操作方式及其意義。

● 根據維度標記套用集合運算

使用者於標記維度並指定作用於量度之集合函式之後，系統便可根據被標記之維度以及套用該集合函式的運算結果，產生一個新的資料環。設有一原始維度為

D3、D2 及 D1，量度為 M1 之資料報表如圖 12 所示，則於標記維度 D1 並套用集合函式後，即產生維度為 D1，量度為 M1 的資料報表。由圖中可看出 D1 包含兩個不同項目的資料，因此於 M1 指定集合函式運算後，其結果為根據該兩項資料，分別計算其對應之 M1 欄的量度結果，如圖 12 右所示。

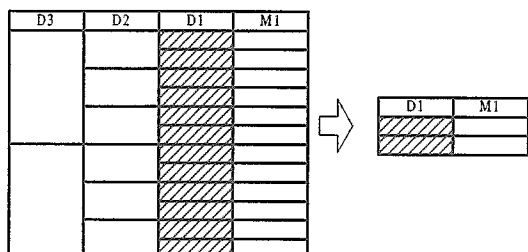


圖 12 標記維度並套用集合函式之示意圖

● 根據維度個別資料標記套用集合運算

同理使用者亦可根據維度個別資料標記方式，標記該維度中的個別資料並指定作用於量度之集合函式，以產生一個新的資料環。設有一原始維度為 D3、D2 及 D1，量度為 M1 之資料報表如圖 13 所示，則於 D1 標記其第一項資料並套用集合函式後，即可產生維度為 D1，量度為 M1 的資料報表。惟此新資料報表中 D1 維度僅包含被標記的資料項目，即未被標記的另一項資料不會出現。

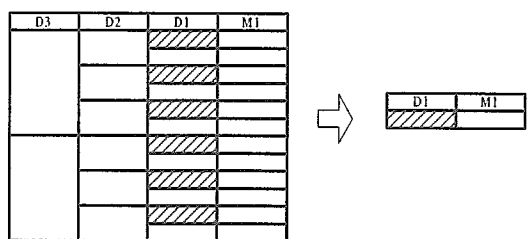


圖 13 標記維度個別資料並套用集合函式之示意圖

● 維度小計總合運算

本節列舉資料環的維度小計總合運算，資料環可根據標記的維度執行該維度之小計運算。如圖 14(a)所示，資料環有年度、產品編號及銷售量；年度維度之原始資料為 1996、1997、1998 及 1999；產品為 P1 及 P2。每一項產品之各年份皆有一銷售量，若執行年份維度的小計總合運算，則年份維度會以所有年份取代，並求其

P1 及 P2 的銷售量，亦即 1996 至 1999 的 P1 及 P2 的個別總銷售量，如圖 14(b)所示。圖 14(b)中的年度以 'ALL' 代替，表示該資料報表的產品編號及銷售量已不需根據年度維度劃分群組，因此可以將該年度維度刪除，如圖 14(c)所示。

| 年度   | 產品編號 | 銷售量 |
|------|------|-----|
| 1996 | P1   |     |
|      | P2   |     |
| 1997 | P1   |     |
|      | P2   |     |
| 1998 | P1   |     |
|      | P2   |     |
| 1999 | P1   |     |
|      | P2   |     |

(a)



| 年度  | 產品編號 | 銷售量 |
|-----|------|-----|
| ALL | P1   |     |
|     | P2   |     |

(b)



| 產品編號 | 銷售量 |
|------|-----|
| P1   |     |
| P2   |     |

(c)

圖 14 在年份維度上執行小計總合的結果

圖 15 為一原始維度為 D3、D2 及 D1，量度為 M1 之資料報表，若於 D1 維度上標記並執行資料小計總合運算後，其 D1 維度以 "所有" 代替，因此略去 D1 維度，留下 D3、D2 維度及 M1 量度之新資料報表。

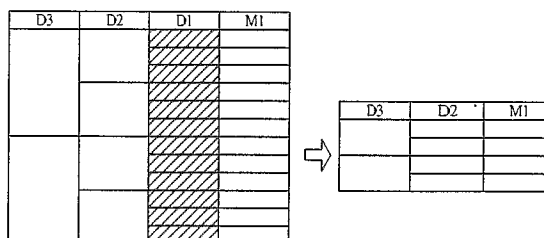


圖 15 於 D1 套用小計總合運算之示意圖

● 各式套用集合運算之資料環顯示範例

本節列出上述分析方式之對應資料環範例，以便能更清楚了解資料表格及資料環的對應關係。

圖 16 所示為一原始維度為 D2、D1，量度為 M1 之資料報表及資料環範例，資料環中粗線部分表示於 D1 維度上選擇某項資料並套用集合運算，其結果為以 D1 維度內該資料項目為依據，所產生之新資料報表及資料環。

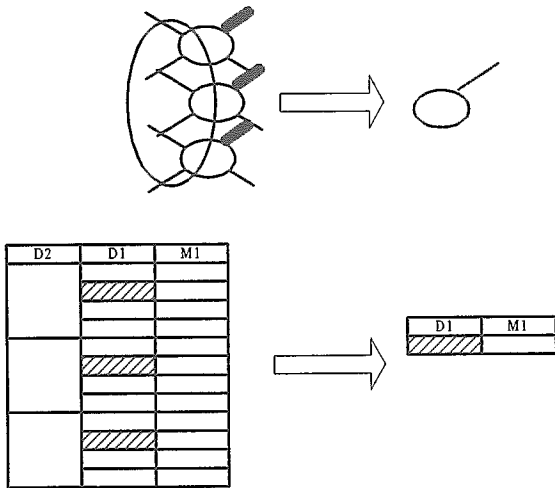


圖 16 標記單項資料並套用集合函式之資料環及資料報表示意圖

圖 17 所示為一原始維度為 D2、D1，量度為 M1 之資料報表及資料環範例，於資料環中標記 D2 維度的兩項資料並套用集合運算，其結果為以被標記之兩項資料為依據，所產生之新資料報表及資料環。

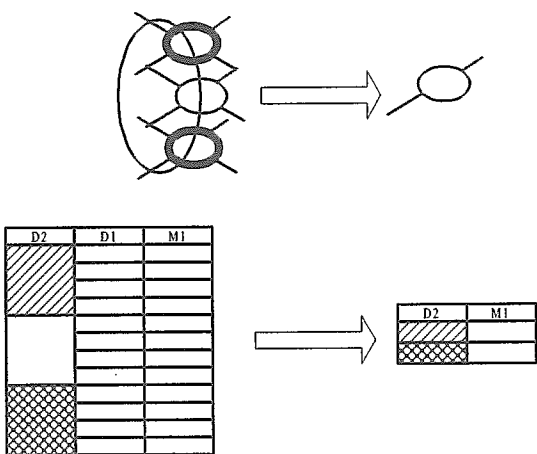


圖 17 標記多項資料並套用集合函式之資料環及資料報表示意圖

之資料報表及資料環範例，於其資料環中標記 D2 維度並套用集合運算，其結果為根據 D2 維度的資料，產生新資料報表及資料環。

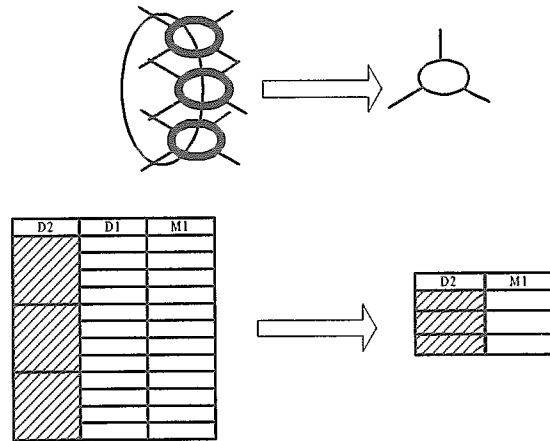


圖 18 標記 D2 維度並套用集合函式之資料環及資料報表示意圖

#### 4. 資料環模型及一般資料立方體顯示模式之比較

本節針對資料環模型及一般資料立方體顯示模式進行比較，包括多維統計顯示方式、資料操作方式及資料瀏覽方式等

##### 4.1 多維統計顯示方式

一般的統計圖表依資料維度大致區分為兩類，一類是平面統計圖表，另一類是立體統計圖表。平面統計圖表一般可顯示一個維度及一個量度所構成的統計資料，常見者如直條圖(Bar Chart)、曲線圖(Curve Chart)、圓餅圖(Pie Chart)等。立體統計圖表基本上是由多個平面統計圖表組合而成，一般能夠顯示兩個維度及單一量度的統計資料。若統計資料的維度及量度超過一般統計圖表所能顯示的範圍時，則必須藉由特殊的方式擴充可顯示維度或量度的個數，例如利用顏色或透明度、圖形大小、圖形旋轉角度代表各維度的資料。例如 DataMiner 3D[3]視覺化工具軟體，便是以此方式完成多維的資料顯示。DtatMiner 3 所呈現的統計畫面是一個三維的立體空間，可定義 X、Y 及 Z 三個維度，並利用一個正方體代表量度或擴充的維度，例如利用正方體之大小、顏色、透明度及正方體旋轉角度等來代表 X、Y 及 Z 維度外之

圖 18 所示為一原始維度為 D1、D2，量度為 M1

其餘維度或量度。此種方式雖可達到多維的統計資料顯示目的，惟以各式之特殊方式來表達維度或量度時，使用者必須記住每一維度或量度是由何種方式來表達，這在統計資料的維度增多時，必將使得資料之顯示及理解益趨困難。

資料環則針對多維的資料提供一致性的資料顯示方式。在資料環結構中，環的層級（即大小）即代表不同的維度；相同維度的資料，其環的大小是相同，至於量度則是最小環上的直條線，使用者完全不必記憶維度及量度的代表圖形，因此不致因資料維度的多寡而影響使用者觀察資料的困難度。

#### 4.2 資料操作方式

以往傳統觀念的資料立方體由於缺乏一個多維的具體資料顯示模型，因此採用表格配合圖表方式作業，亦即透過表格的資料操作後再以圖表顯示其操作結果。資料環所提供的操作方式截然不同，它是以資料環為操作對象，藉由資料環的操作再觀察詳細資料。資料環提供多維的資料顯示，分析者可觀察到資料環 任意資料，並透過資料環標記方式直接瀏覽各子群組資料的詳細內容或執行集合運算。例如我們可以藉由資料環的標記方法，以表格方式顯示被標記資料的詳細內容，或根據標記維度個別資料重新執行算集合函式而得到新的量度結果。

#### 4.3 資料瀏覽方式

資料環將原始資料由表格轉換成圖形，藉由圖形的顯示，使用者可於顯示視窗上瀏覽整個資料內容，資料環實作方式可採用三維之虛擬實境方式，例如虛擬實境模式語言(VRML) [4]。採用 VRML 可讓資料環顯示於一個三維立體空間，使用者可於此立體空間內瀏覽資料內容，並利用類似一般操作物體的方式操作資料環，例如旋轉某維度之資料環，以便依序觀察資料環內容，或者在虛擬實境內漫步，瀏覽不同維度內的資料，如此的方式讓使用者觀察資料更加方便、直接。

藉由虛擬實境的瀏覽方式，可使一般在三維圖表中可能產生某些角度不易觀察的問題消失，例如在立體直條圖中，由於某些量度值過小，而使得該值被其他量

度值的直條方塊蓋住，此類問題都可以透過虛擬實境瀏覽的方式解決。

資料環是多維的資料顯示模型，因此也特別適用於立體的虛擬實境裡，目前網際網路上結合虛擬實境瀏覽已日益普及，我們可藉由此方式發展主從式架構之資料環線上分析處理，讓網際網路、企業網路(Intranet)及商際網路(Extranet)等使用者也可以得到資料環處理所帶來的好處。

#### 5. 結論

本論文提出資料環架構，其主要目的為提供一個資料立方體之顯示及操作工具，以改善資料立方體顯示多維資料之問題。此外，資料環操作方式，亦使得使用者可藉資料環直接進行資料分析。

資料環未來的發展將朝向於發展各式的線上分析處理伺服器之客戶端程式及更佳的資料環顯示方式，因此未來可針對各式的線上分析處理伺服器所提供的查詢方式，提供一個轉換方式讓資料環透過該伺服器提供的快速查詢服務。另外，在既定的資料環操作及資料結構下，未來可進一步發展更佳的資料環顯示方式，讓統計資料能夠更清楚且更明瞭的顯示，並獲取更精確的分析結果。

#### 參考文獻

- [1]E. F. Codd, S. B. Codd and C. T. Salley, "Providing OLAP (On-Line Analytical Processing) to user-analysts: An IT mandate," *Technical report*, E. F. Codd& Associates, 1993.
- [2]J. Gray, A. Bosworth, A. Layman, and H. Pirahesh, "Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals," *Proceedings of the 12th International Conference on Data Engineering*, pp.152-159, 1996.
- [3]Wired Analyzer of WIRED for OLAP, <http://www.appsource.com>
- [4]Virtual Reality Modelin Language, VRML, <http://www.vrml.org>