

福特馬食品連鎖經銷商店資料倉儲之研究

曹文瑜 楊惠貞

勤益技術學院資管科

摘 要

未來誰能掌握正確即時的資訊，誰就是市場上的贏家。本文試著以福特馬食品連鎖經銷商的歷史性資料，成長到某個程度，交易所產生的資料就變的很可觀，如何從不同的角度讓不同的主管可得到最快速的查詢分析及做決策和趨勢的判斷，由資料倉儲管理系統來處理是再恰當不過的。資料倉儲可以彌補非資訊管理工作人員來做大量資料的查詢分析的工作，主管可以根據自己的需求，自行選定項目，可快速而自動的萃取出來自不同資料來源，並且產生易閱讀、易於了解的圖表以協助決策的制定，並可解決市面上一般使用關聯式資料庫管理系統會留下遺產系統所產生的問題。

關鍵字:。 資料倉儲，關聯式資料庫管理系統，遺產系統

一、前言

本研究針對福特馬食品連鎖經銷商的歷史資料依主管不同角度能快速的查詢、分析、統計及趨勢判斷建立一個資料倉儲系統，試著以 MS SQL Server 7.0 來完成。資料倉儲是一種能正確的組合與管理不同資料來源的技術，在一般資料庫管理系統會常留下遺產系統的問題，舉例說明：資料重複(例 ROC, R.O.C., Republic Of China)，資料欄的定義不一致(例 A 資料庫某欄位用字串定義，而在 B 資料庫則用數值)，資料不一致(例 A 資料庫的資料欄 0 代表男，1 代表女，但在 B 資料庫中卻以 0 代表女生，1 代表男生)，資料倉儲可以解決以上的問題，用法為在主管和遺產系統中多了一個界面，此界面朝二步驟進行資料的分析工作，第一步驟將很多的遺產系統加以整合與轉換，第二步驟是提供相關的統計分析，本研究是以 star schema 來為福特馬食品連鎖經銷商店做資料的分析，並提供主管做為決策之用。

二、作法

先說明福特馬食品連鎖經銷商店的歷史資料表(表 1)，共有 store, customers... 若干關聯表。作法如下：

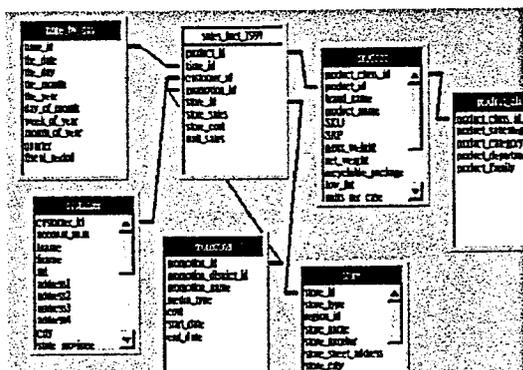
第一步是將交易系統轉入多維式資料庫系統，移轉時必須要注意資料的完整性和正確性，資料倉儲的資料是由很多不同的交易系統所匯集而成，先注意資料是否有衝突，重複和參考完整性的問題，緊接著要注

意單位的問題，同樣是金額欄位在 A 交易系統中以千元為單位，但在交易系統中卻是以萬元為單位，資料長度和數值精確度也必須一併考量，再來就是同義查詢的問題，例如：台中市和臺中市都是指同樣的地方，但卻有不同的二筆資料，最後就是空值的處理了，某個資料可能是 Null, 空字元或 0，在轉入過程中必須小心的轉換、捨棄或保留 SQL SERVER 7.0 截取舊系統料並轉入資料倉儲的資料庫，在轉換之前都要驗證資料的正確性，此時 SQL SERVER 7.0 提供資料轉換工具 DTS (Data Transformation Services Import and Export)，不但可以資料的移轉，並且將資料清離。先畫出多維式資料庫分析角度圖(圖 1)，此圖為星狀結合圖，它只有一個中心，是根據主管思考的角度及每一個角度的層級關係，其中以銷售情況的歷史資料為中心 sales_fact_1997(稱為 Fact table)，存放所有數值化的測量值，這些測量值可以由維度交叉或合併(Join)的過程而產生若干個資料表 customer, store... (稱為 Dimension table)，分別為產品類別(product family)，時間(year)，地區(customers)，在 Fact table 中存放著所有主管所要觀察的測量值，其中在銷售情況表中的資料是累積以往銷售的記錄通常是非常龐大的。

資料表	查詢	表單	報表
customer	days	inventory_fact_1997	
inventory_fact_1998		product	
product_class	promotion	region	
sales_fact_1997	sales_fact_1998	sales_fact_dec_1998	
store	time_by_day	warehouse	
warehouse_class			

表 1 歷史資料表

圖 1 多維式資料庫分析角度圖



接下來就是界定 Fact Table 內要有那些組成份子，思考模式如下：

- a. 有那些數據會拿來分析
- b. 主管會從那些觀點來分析這些資料不外乎何者銷售量(What)、何時 (When)、在那裏(Where)、發生原因(Why)、如何補救 (How)
- c. 儲存的最小單位
- d. 資料存放的週期為何?

緊接著界定 Dimension Table 的資料內容由多維式分析角度可以得知，主管至少會以產品類別(product)、時間(time_by_day)、地區(customers)等觀點來分析資料。

下一步是定義 Fact table 的測量值。利用 SQL Server 7.0 OLAP 設計與建立 OLAP 資料庫 FoodMart，使用 OLAP Manager 可以管理和維護，在 OLAP Server

內的每一個資料倉儲是由 Cube, Virtual Cube 及 Library 所組成的。

Cube 內主要包含資料定義，例如資料來源，維度定義，測量值定義、分割方式、規則、Virtual Cube、Library。

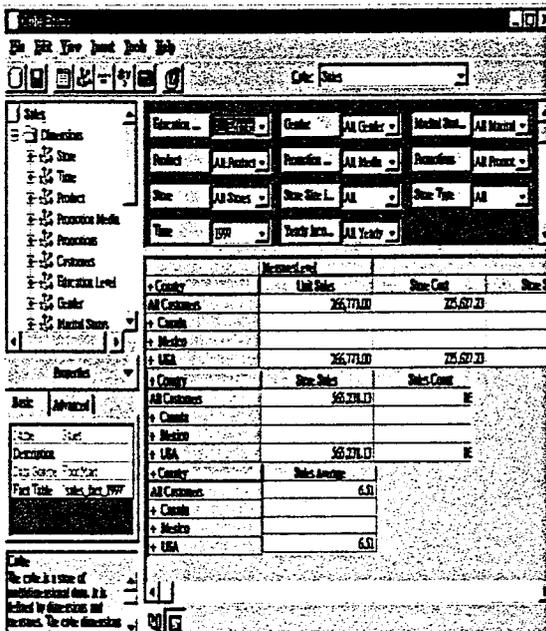
現在要定義 Cube，先要將 Fact table 中用得到的 Measurement 找出來，例如建立銷售分析的 Cube(sales)，會用到成本與價格及什麼時候在那個門市出售那項產品，在定義 Dimension Table 時要詳細的釐清階層關係例如：Time Dimension 中要從年別季別，月別，週數等由上下階層性的逐一定義，主管在查詢問題時一方面可給予高階層彙總性答案，例如：主管可查詢過去 5 年的銷售業績，發現第三年的業績比較差。再依主管的需求進行細部資料的查詢，查出第三年的那一季或那一月的業績出問題，找出業績較差的原因，其中 Private 只可讓該 Cube 使用而 Shared 可讓該資料倉儲中所有的 Cube 使用。OLAP 提供三種資料儲存的方式：

1. MOLAP(multiOLAP)
2. ROLAP(RelatioalOLAP)
3. HOLAP(HyperOLAP)為前面兩者混合體

我們把彙總性資料存放在 MOLAP，基本資料存在 ROLAP 中，當主管要查詢彙總資料時，會到 MOLAP 中找尋，要再細部查詢資料時會從 ROLAP 中找尋所要的資料。資料也可分割成不同的儲存策略來儲存資料。例如：有一個 Cube，存放了數年的銷售資料，可把今年的資料存放在 MOLAP 中把去年和前年的資料存放 HOLAP 中，至於三年以前的資料則以 ROLAP 的方式儲存，一方面可有最好的執行速度，另一方面有最省空間 OLAP 有前置性彙總運算，將前置運算所得的結果存在 Cube 的 Cell 中，當主管詢問 1997 年對 CA 地區 Customers(顧客)購買本公司

產品的銷售情形是如何時，事實上是從年、度、產品、地區三個 Dimension table 中所對應到的方格中取出销售量。如何將福特馬食品連鎖經銷商店龐大的資料分析，讓這些資料更能顯示出此商店營運的狀態[4]。圖 2 將分析的結果呈現出來，但是沒有好用的圖表供主管參考，現將在以上的步驟轉向 excel 2000 來分析。

圖 2 分析圖

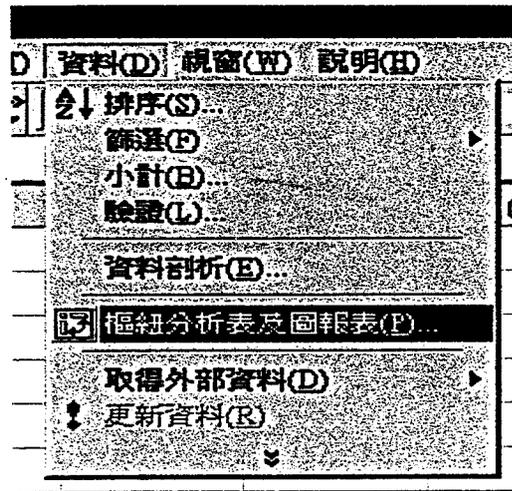


三、分析

現在就以 excel 2000 來分析。分析如下：

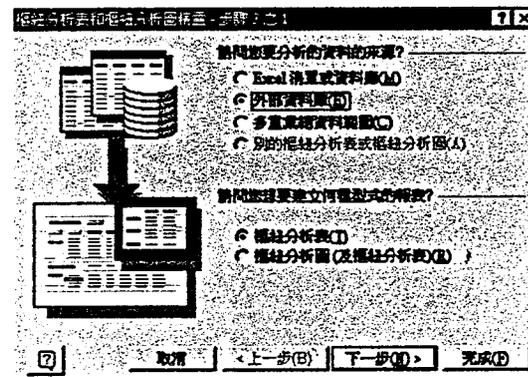
step1、進入 EXCEL，選功能表中資料選項中的樞紐分析表及圖報表(圖 3)

圖 3



Step2、出現圖 4，請問您要分析的資料來源，選擇外部資料庫(E)後，按下一步，出現圖 5。

圖 4



step3、點選“取得資料”，出現圖 6。

圖 5

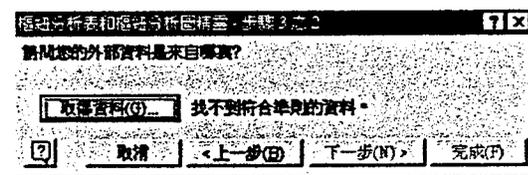
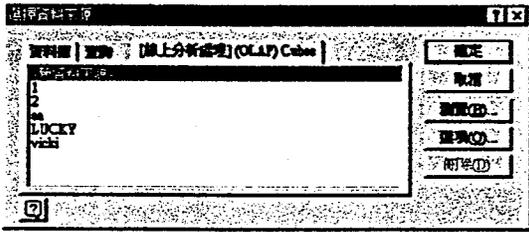
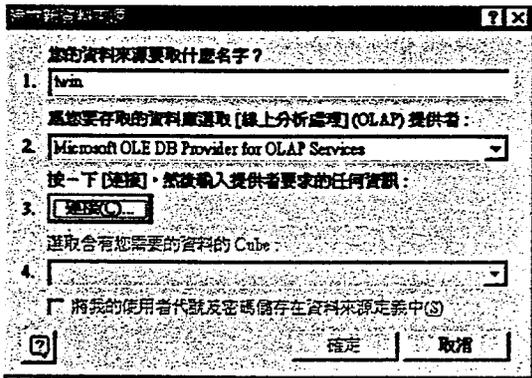


圖 6



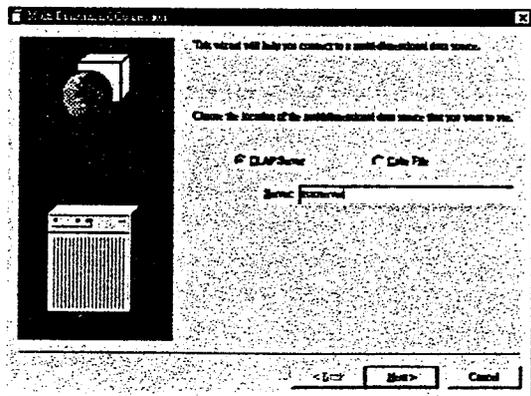
step4、選擇“線上分析處理”頁籤，選擇<新資料來源>後按確定，出現圖 7。

圖 7



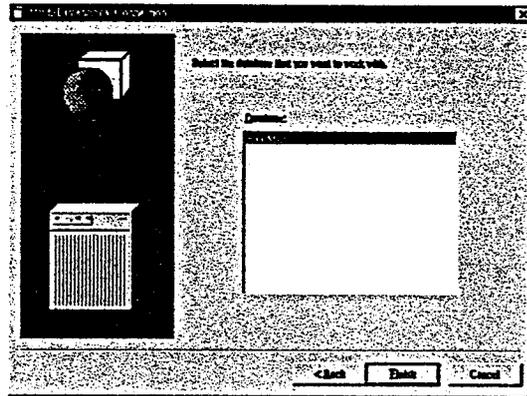
step5、在第一列中填入連接資訊的名稱 twin，在第二列中選擇連接後端資料庫的介面後，按“連接”，出現圖 8。

圖 8



step6、點選 OLAP server，輸入 server 名稱後按 next，出現圖 9。

圖 9

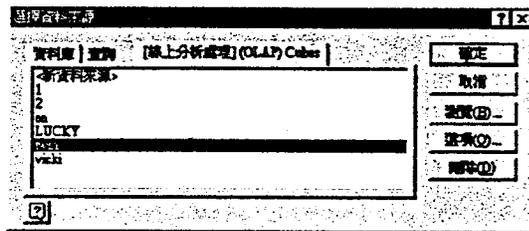


step7、系統會將 OLAP Server 中可使用的資料倉儲資料庫資訊顯示出來，此處選擇 FoodMart 倉儲資料庫，然後點選 Finish 鍵。

step8、會回到 step5 的圖示(圖 7)，在第四列中選取您需要的資料的 cube.後按確定。

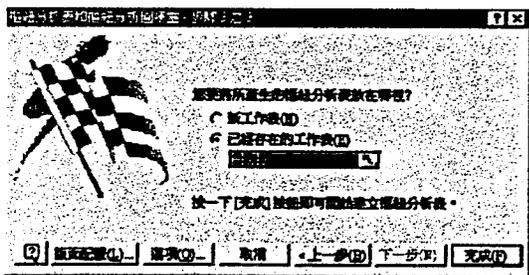
step9、會出現圖 10 所設連接資訊的名稱後，按確定。

圖 10



step10、會回到 step3 的圖示(圖 5)，按下一步。選擇樞紐分析表的位置，出現圖 11。

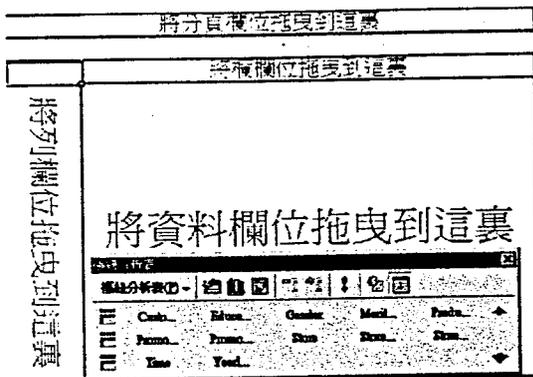
圖 11



step11、按完成，出現圖 12，其中在樞紐分析表所看到的是可供測量的 Measurement 及可供分析

的 Dimension 名稱，”將分頁欄位拖曳到這裏”是做為資料分頁選擇標準的依據，”將欄欄位拖曳到這裏”及”將列欄位拖曳到這裏”是做為各 Dimension 不同角度樞紐分析的地方，至於”將資料欄位拖曳到這裏”則是做為放置要分析的 Measurement 地方。

圖 12



範例說明

若主管想知道福特馬食品連鎖商店在 1997 年對 CA 地區 Customers(顧客)購買本公司產品的銷售情形是如何，我們以地區 customers 為分頁欄位，先找出 CA 地區，再以時間 time 為列欄位、銷售數量 sales count 為分析的資料欄位、產品類別 product Family 為欄位，如表 2。

Customers	USA				
Sales Count	Product Family				
Year	Drink	Food	Non-Consumable	總計*	
1997		7978	62445	16414	86837

表 2

主管若想細看 USA 地區的三個州銷售的情形，在地區 customers 下拉式選項中可選取所要分析區域的顧客，此處 USA 有三州顧客，分別是 CA,OR,WA 我們選取 USA 中的 CA，如表 3。

Customers	USA				
Sales Count	Product Family				
Year	Drink	Food	Non-Consumable	總計*	
1997		7978	62445	16414	86837

表 3

主管若想細看年度的銷售情形

在 Year 的下拉選項中可勾選所要分析的年度 (亦可選擇 Quarter 或月份)，如表 4。

Customers	CA		
Sales Count	Product Family		
Year	Drink		
<input checked="" type="checkbox"/> 1997			2291
<input checked="" type="checkbox"/> 1998			
確定		取消	

表 4

若在 “1997” 欄位中點二下則會顯示四個等份 Q1、Q2、Q3、Q4，更可以進一步的得知各個等份的銷售情形，如表 5。

Customers	CA			
Sales Count	Product Family			
Year	Quarter	Drink	Food	
1997	Q1		535	
	Q2		520	
	Q3		571	
	Q4		665	
1997 小計*			2291	

表 5

甚至於在 Q1 欄位中點二下則會顯示一季中的一月到三月三個月份的銷售情形，如表 6

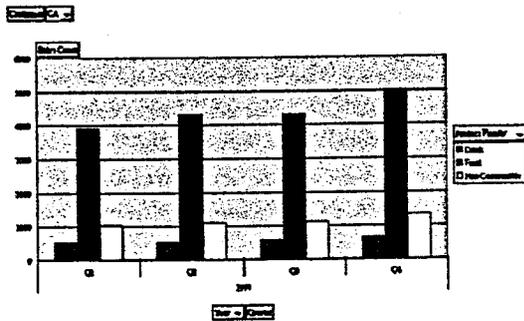
Customers	CA			
Sales Count	Product Family			
Year	Quarter	Month	Drink	
1997	Q1	1	169	
		2	202	
		3	164	
Q1 小計*			535	
	Q2		520	

表 6

別為 Q1、Q2、Q3、Q4 四個等份，以 Q4 這個時期的總銷售為 7015 單位最高，在整個 1997 年度 Drink、Food、Non-Consumable 三種產品中又以 Food 產品類別 17556 單位銷售最好。

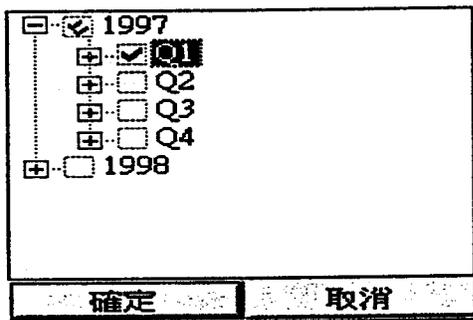
由表 10 可做出圖 16

圖 16



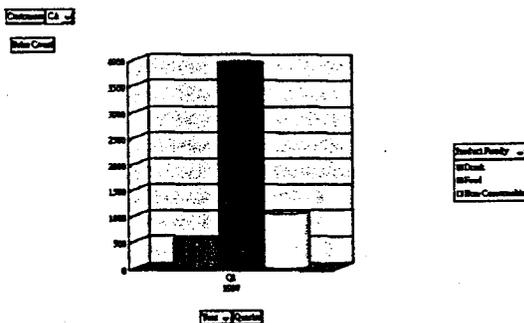
上圖中如欲查看單一時期，則可按 Year 下拉式選單，例如：要查看 Q1，則在 YEAR 下拉式選單中(如圖 17)。

圖 17



分析結果為圖 18

圖 18



Year	Quarter	Month	Drink	Food	Non-Consumable	總計*
1997	Q1	1	169	1241	321	1731
		2	232	1405	360	1967
		3	164	1278	358	1800
	Q1小計*		533	3924	1034	5491
1997小計*			2291	17556	4991	24442

表 11

表 11 為地一第份 Q1 再細分為 month(月份)，分別為 1 月、2 月、3 月、以 2 月總銷售 19767 單位為最高，在三種產品類別中以 Food 總銷售值 3924 單位最高，即銷售最好，依次為 Non-Consumable，銷售最差的是 Drink。

由表 11 可做出圖 19

圖 19

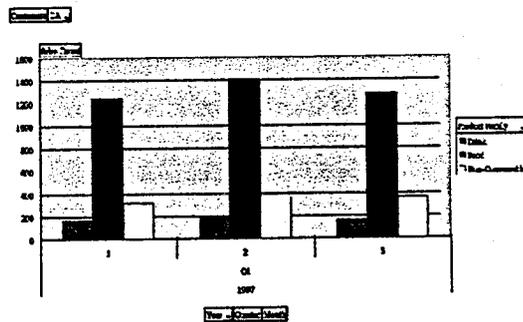


表 12 是分析福特馬在“CA”地區的顧客在 1997 整個年度中對飲料產品購買的情況，此處我們勾選飲料產品類別細分的三部分：Alcoholic Beverages(酒類飲料)、Beverages(非酒類飲料)、Dairy(奶類飲料)作為分析，從表中我們即能得知飲料的銷售量為酒類飲料 1252 單位最好，依次為非酒類飲料 632 單位、奶類飲料 407 單位，對於擁有此份分析我們即能在作生產決策時能有所參考。由表 12 可做出圖 20

Year	Alcoholic Beverages	Beverages	Dairy	Drink小計*	總計*
1997	632	1252	407	2291	24442

表 12

圖 20

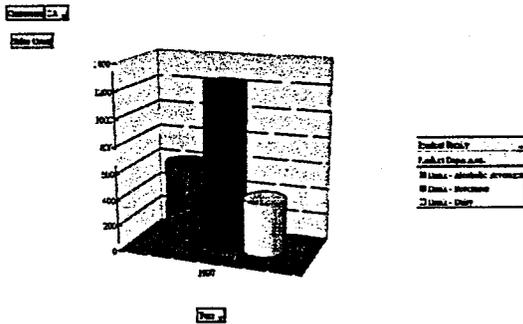
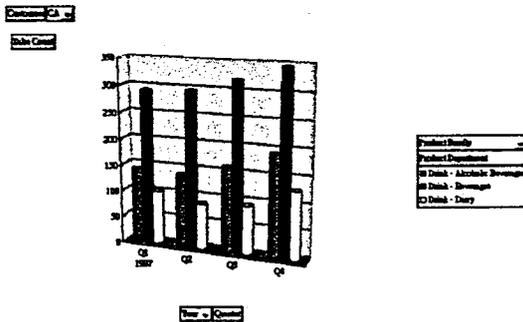


表 13 是將 1997 年細分為四個 Quarter(等份)加以分析每個時期飲料的銷售量, 每個時期中含有 3 個月份, Q1 為 1 月、2 月、3 月, Q2 為 4 月、5 月、6 月, 其餘 Q3、Q4 以此類推, 由表 13 中我們能得知更詳細的資料, 例如: 在 Q2(4 月、5 月、6 月這三個月)等份時, 銷售量為酒類飲料 296 單位最好, 依次為非酒類飲料 138 單位、奶類飲料 86 單位。表 13 可做出圖 21

表 13

圖 21



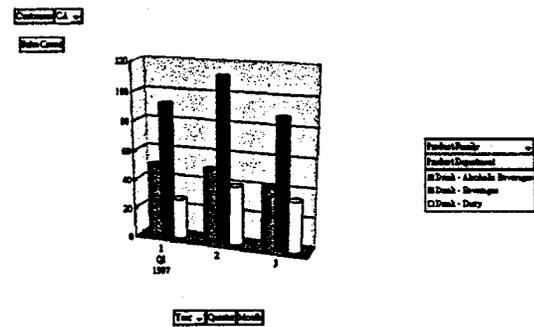
由表 14 是將 Q1 再細分為 1 月份、2 月份、3 月份對 Alcoholic Beverages(酒類飲料)、Beverages(非酒類飲料)、Dairy(奶類飲料)作為分析, 由表 14 即可知銷售量為酒類飲料 92 單位為最好, 依次為非酒類飲料 50 單位、奶類飲料 27 單位, 此細部資料顯示在冷的季節銷售量不

好, 此時少進奶類飲料貨品。由於有此種多重分析功能使我們更能容易的取得資訊, 更能幫助我們作決策。由表 14 可做出圖 22

Sales Group		Product Family		Product Department		Drink 小計 * 總計 *	
Year	Quarter	Month	Alcoholic Beverages	Beverages	Dairy		
1997	Q1	1	50	92	27	169	1731
		2	50	112	40	202	1967
		3	42	88	34	164	1800
		Q1 小計 *	142	292	101	535	5498
1997 小計 *			632	1252	407	2291	24442

表 14

圖 22



四、結論

本研究將傳統以交易為導向的資料庫管理系統及開發的理念調整為以查詢及趨勢分析為導向的資料倉儲系統, 對於大量資料和著重資料分析的企業公司都是一個非常好的競爭利器。

五、參考文獻

- [1] 李勁著, 1999, “Excel 2000 與 SQL Server 7.0 整合應用”, 第一版, 文魁
- [2] 李勁著, 1999, “Excel 2000 與 SQL Server 7.0 整合應用”, 松崗
- [3] 蕭凱文、薛志達、李政輝合著, 1999, “SQL Server 7.0 資料倉儲整合應用”, 微軟
- [4] 陳宗興編著, 1999, “SQL Server 7.0 資料庫管理與應用學習手冊”, 松崗

- 【5】 林存德，1999年1月，“資料倉儲觀念簡介(1)”，RUN PC，60期，P277~P288.
- 【6】 林存德，1999年2月，“資料倉儲觀念簡介(2)”，RUN PC，61期，P247~P256.
- 【7】 林存德，1999年3月，“資料倉儲實用應用”，RUN PC，62期，P241~P256.
- 【8】 林存德，1999年4月，“資料倉儲與VBA的前用應用-以EXCEL為例”，RUN PC，63期，P262~272.
- 【9】 林希孟，1999年2月，“資料倉儲檢視與展望”，0與1 BYTE科技雜誌，P166~173.
- 【10】 兆太科技提供，1999年6月，“資料倉儲的評估建置經驗-Hunt公司”，資訊與電腦，P130~133.
- 【11】 陳文華，1999年5月，“應用資料倉儲系統建立CRM”，資訊與電腦，P122~127.
- 【12】 陳文華，1999年5月，“應用資料倉儲系統建立CRM”，資訊與電腦，P122~127.
- 【13】 陳文華，1999年1月，“透視資料倉儲”，資訊與電腦，P98~103.
- 【14】 唐如君，1999年2月，“如何規畫資料倉儲”，資訊與電腦，P91~95.
- 【15】 陳文華，1999年3月，“架構資料倉儲的注意事項”，資訊與電腦，P94~99.
- 【16】 張庭禎，1999年4月，“SQL SERVER 7.0帶來新一代資料庫管理功能”，0與1 BYTE科技雜誌，P237~P241.
- 【17】 張庭禎，1999年5月，“新資料庫應用程式設計觀念與技術”，0與1 BYTE科技雜誌，P233~P238.
- 【18】 張庭禎，1999年6月，“SQL SERVER 資料倉儲的組織架構”，0與1 BYTE科技雜誌，P211~P214.