

國立交通大學
工業工程與管理學系

碩士論文

台灣花卉供應鏈的
資料倉儲設計與量測變數迴歸應用
Data Warehouse Design for Flower
Supply Chains in Taiwan with an
Instrumental Variable Regression Application

研究生：陳家瑜

指導教授：梁高榮博士

中華民國九十五年六月

研究生：陳家瑜

指導教授：梁高榮博士

國立交通大學工業工程與管理學系

中文摘要

總資料倉儲已經由四個資料超市整合建置完成，用來分析台灣地區的花卉交易情形。總資料倉儲在本質上有三個顯著的特徵。第一、採用競逐模型建置資料倉儲可減少工程上的風險。亦即每一條供應鏈就建置一個資料超市，然後資料超市中的花卉交易資料，經由資料轉換服務被轉入資料倉儲。第二、在資料倉儲中新增一個節慶維度，提供國曆節慶與傳統農曆節慶的花卉市場資訊。節慶維度的新增，在許多資料超市使用者的強烈建議下誕生，用來幫助花卉運銷的決策制訂。第三、應用量測變數迴歸技術，在某些假設之下，來找尋文心蘭的價量關係。若滿足這些假設，則此方法的成功應用，將可擴展至其他花卉上。



關鍵詞：

資料超市(Data Mart)

資料倉儲 (Data Warehouse)

資料轉換服務(Data Transformation Service, DTS)

量測變數 (Instrumental Variable, IV)

Data Warehouse Design for Flower Supply Chains in Taiwan
with an Instrumental Variable Regression Application

Student : Chia-Yu Chen

Advisor : Dr.Gau-Rong Liang

Department of Institute of Industrial Engineering & Management
National Chiao Tung University

Abstract

A data warehouse has been established from four existed data marts for analyzing flower transactions in Taiwan. Essentially there are three salient features in this establishment. First Kimball model was proposed to construct the data warehouse for reducing engineering risk. It means to construct a data mart for each flower supply chain. Then the flower transactions in data marts are filtered into the data warehouse through Data Transformation Services (DTS) technology. Second a new festival dimension is added in the data warehouse for providing flower marketing information during national and traditional holidays. The addition of festival dimension was strongly suggested by many data mart users for helping decision making in flower marketing. Third an instrumental variable regression method is applied to finding a price-quantity relationship for oncidiums under some assumptions. The successful application of this method can be extended to other flowers if the assumptions are satisfied.

Keywords :

Data Mart

Data Warehouse

Data Transformation Service

Instrumental Variable

誌謝

本論文得以順利完成，首先要感謝恩師梁高榮老師的細心指導與諄諄教誨。就讀研究所這兩年的時間，老師不僅教導我課業上的知識，也時常授與實務上的經驗，加上平時相處時的身教言教中傳達的追根究底的研究精神與重視規矩的做事方法都讓我獲益良多。使我更有勇氣面對未來人生的挑戰，僅在此致上由衷的感激。此外，還要特別感謝唐麗英老師與張永佳老師對於論文的詳加審閱，並提供寶貴的意見使本論文更臻完備，在此亦要致上最誠摯的感激。

研究所兩年就讀的期間，感謝同窗好友小胖、小仙、喬 AA 在課業上的幫助，一起互相砥礪。還有室友幫妹、包包跟乖狗 SEVEN、K 隆星人小 KERO，給予我精神上的鼓勵。實驗室學弟新凱、炯堂、阿牛、阿端、BE 製造歡樂氣氛，使在學的兩年雖然辛苦但是仍然充滿歡樂。在此深致謝意。

本人在就讀研究所期間，曾經參予農委會之「花卉資料倉儲與標準化交易作業系統軟體之整合與開發」計畫（計畫編號：94I819），獲得許多幫助，特此致謝。

最後，要感謝一路陪伴我支持我完成碩士學業的奶奶、爸媽，有你們的支持我才有動力完成學習。還有哥哥均成，我們從小到大念的學校都一樣，祝福你在澎湖當兵順利，早點回新竹打拚。以及陪伴我度過難關的阿雞，謝謝你一直支持我並與我分享喜悲。所有陪我度過碩士階段的師長與朋友們，感謝各位在求學過程中給我的一切鼓勵與支持。



目錄

中文摘要.....	i
英文摘要.....	ii
誌謝.....	iii
目錄.....	iv
圖目錄.....	vi
表目錄.....	viii
第一章 緒論.....	1
1.1 研究動機.....	1
1.2 問題界定.....	2
1.3 研究目的.....	4
1.4 研究流程.....	6
1.5 論文架構.....	7
第二章 文獻回顧.....	8
2.1 資料倉儲.....	8
2.1.1 資料倉儲架構.....	8
2.1.2 資料倉儲之時間性.....	14
2.1.3 線上分析處理.....	15
2.2 節慶維度相關介紹.....	16
2.3 量測變數迴歸技術.....	18
2.3.1 線性迴歸基本原理.....	18
2.3.2 量測變數迴歸技術方法.....	19
2.3.3 時間序列資料.....	20
2.3.4 單位根檢定.....	21
第三章 台灣區花卉資料倉儲資料轉換服務實作.....	23
3.1 系統架構與資料倉儲架構.....	23
3.1.1 系統架構.....	23
3.1.2 資料倉儲架構.....	28
3.2 總資料倉儲整合工作.....	30
3.2.1 總資料倉儲架構.....	30
3.2.2 資料整合工作.....	30
3.3 資料轉換服務.....	33
3.3.1 實體關係模式至多維度模型轉換.....	34
3.3.2 資料轉換服務.....	36
3.3.3 自動化設計.....	43
3.4 資料倉儲遠端維修方式.....	46
第四章 節慶維度設計.....	52
4.1 維度內容介紹.....	52
4.1.1 節慶相關資訊.....	52



4.1.2 節慶日期重複問題.....	54
4.1.3 節慶日期排序問題.....	55
4.1.4 節慶產生的空值問題.....	55
4.2 節慶維度設計架構.....	56
4.2.1 維度設計方案一.....	56
4.2.2 維度設計方案二.....	57
4.2.3 維度設計方案三.....	58
4.2.4 維度設計方案四.....	59
4.3 設計方案的比較.....	63
4.3.1 線上顯示設計模式.....	63
4.3.2 優缺點比較.....	64
第五章 量測變數迴歸應用.....	67
5.1 分析流程.....	67
5.2 資料收集與分析.....	69
5.2.1 原始資料收集.....	69
5.2.2 單位根檢定.....	70
5.2.3 迴歸方程式分析.....	71
5.3 量測變數迴歸.....	73
5.3.1 外來變數資料分析.....	73
5.3.2 量測變數迴歸之測定.....	74
5.3.3 結果分析.....	76
5.4 資料倉儲節省時間之估算.....	77
第六章 結論與未來研究方向.....	78
6.1 結論.....	78
6.2 未來研究方向.....	79
參考文獻.....	80



圖目錄

圖 1.1 總資料倉儲系統架構圖	2
圖 1.2 量測變數迴歸技術方法與界定	3
圖 1.3 研究流程圖	6
圖 2.1 資料項目集合(DIS)	9
圖 2.2 般默資料倉儲螺旋式發展方式	10
圖 2.3 星狀綱要圖	11
圖 2.4 雪花綱要圖	12
圖 2.5 星座綱要圖	12
圖 2.6 超方體架構	15
圖 2.7 農民曆三資料表關聯圖	16
圖 2.8 農民曆三資料表資料內容	17
圖 3.1 總倉儲實體系統架構圖	24
圖 3.2 SQL Server連線群組(a)	25
圖 3.3 SQL Server連線群組(b)	25
圖 3.4 SQL Server連線群組(c)	26
圖 3.5 SQL Server連線群組(d)	26
圖 3.6 SQL Server連線群組(e)	27
圖 3.7 網路磁碟機畫面	27
圖 3.8 檔案複製到網路磁碟機語法畫面	27
圖 3.9 資料轉換時刻圖	28
圖 3.10 台灣區花卉資料倉儲競筭架構圖	29
圖 3.11 花卉批發資訊分享熱線架構圖	29
圖 3.12 四家花市聯集示意圖	31
圖 3.13 自動化更新轉換程序	33
圖 3.14 實體關係至多維度模型架構圖	34
圖 3.15 彰化資料超市內部轉換關係圖	35
圖 3.16 拍賣資料表與總倉儲交易資料表實體關聯圖	35
圖 3.17 總資料倉儲雪花綱要關聯圖	36
圖 3.18 拍賣資料轉換工作	37
圖 3.19 台北拍賣資料轉換工作	37
圖 3.20 台北拍賣轉換資料工作來源查詢	37
圖 3.21 台北拍賣轉換資料工作屬性	38
圖 3.22 台北拍賣轉換資料工作轉換語法	38
圖 3.23 訂貨資料轉換工作	39
圖 3.24 台北訂貨資料轉換工作	40
圖 3.25 殘貨資料轉換工作	41
圖 3.26 台北殘貨資料轉換工作	41
圖 3.27 議價資料轉換工作	42

圖 3.28 台北議價資料轉換工作	42
圖 3.29 自動化Macro程式	43
圖 3.30 自動化步驟語法編輯	44
圖 3.31 自動化排程設定	44
圖 3.32 自動化排程重複執行設定	45
圖 3.32 VNC伺服器與檢視器圖示	47
圖 3.33 VNC伺服器主要屬性選項	47
圖 3.34 VNC檢視器連線視窗	48
圖 3.35 VNC檢視器連線選項	48
圖 3.36 VNC檢視器連線狀態	49
圖 3.37 VNC驗證	49
圖 3.38 遠端桌面視窗	49
圖 3.39 資料轉換服務錯誤畫面	50
圖 3.40 SQL Server視窗畫面	51
圖 3.41 Query Analyzer視窗畫面	51
圖 4.1 節慶重複情形示意圖	54
圖 4.2 一日期對多節慶示意圖	54
圖 4.3 節慶日期排序問題示意圖	55
圖 4.4 節慶資料表關聯圖-方案一	56
圖 4.5 節慶資料表實體資料-方案一	56
圖 4.6 節慶資料表關聯圖-方案二	57
圖 4.7 節慶資料表實體資料-方案二	57
圖 4.8 節慶二資料表關聯圖-方案三	58
圖 4.9 節慶二資料表實體內容-方案三	59
圖 4.10 節慶二資料表關聯圖-方案四	60
圖 4.11 節慶二資料表實體內容-方案四	61
圖 4.12 節慶層級概念	62
圖 4.13 方案二節慶類別排列方式	65
圖 5.1 文心蘭照片	67
圖 5.2 量測變數迴歸技術作業流程圖	68
圖 5.3 文心蘭總平均價對總成交量之散佈圖	69
圖 5.4 文心蘭的總成交量與總平均價之時間序列走勢圖	70
圖 5.5 獨立變數與殘差之散佈圖	71
圖 5.6 文心蘭總平均價與產量 2000 年至 2005 年月平均圖	73
圖 5.7 量測變數對獨立變數散佈圖	75
圖 5.8 量測變數Z與殘差E之散佈圖	75

表目錄

表 1.1 資訊分享示意表.....	4
表 1.2 舊維度與新維度比較表.....	4
表 2.1 殷默及競篋模式基本特性之比較.....	14
表 2.2 農曆對照表.....	16
表 2.3 農曆月.....	16
表 2.4 農曆年.....	17
表 3.1 各主機項目命名表.....	25
表 3.2 總倉儲主機負責服務整理表.....	28
表 3.3 總倉儲與各家資料超市維度比較表.....	30
表 3.4 整合對照表.....	32
表 3.5 總倉儲資料轉換程式.....	33
表 3.6 拍賣資料轉換工作步驟表.....	39
表 3.7 訂貨資料轉換工作步驟表.....	40
表 3.8 殘貨資料轉換工作步驟表.....	41
表 3.9 議價資料轉換工作步驟表.....	42
表 3.10 Ultr@VNC功能特點.....	46
表 4.1 國曆特殊節慶.....	53
表 4.2 農曆特殊節慶.....	53
表 4.3 節慶三大問題.....	55
表 4.4 節慶資料表-方案一.....	56
表 4.5 節慶資料表-方案二.....	57
表 4.6 方案二維度層級表.....	58
表 4.7 節慶範圍資料表-方案三.....	58
表 4.8 節慶類別資料表-方案三.....	58
表 4.9 方案三維度層級表.....	59
表 4.10 節慶範圍資料表-方案四.....	60
表 4.11 節慶類別資料表-方案四.....	60
表 4.12 節慶類別代碼對照表.....	61
表 4.13 方案四維度層級表.....	62
表 4.14 節慶四方案設計方式比較表.....	64
表 4.15 四方案改善情況表.....	66
表 5.1 資料分析時間長度.....	69
表 5.2 文心蘭的總成交量與總平均價資料之單位根檢定.....	70
表 5.2 文心蘭需求方程式之分析結果.....	71
表 5.3 屏東縣月均溫.....	74
表 5.4 量測變數與獨立變數相關性之分析結果.....	75
表 5.5 量測變數與殘差相關性之分析結果.....	76
表 5.6 兩階段最小平方法之分析結果.....	76

表 5.7 總資料倉儲線上分析處理動作表.....77
表 5.8 總資料倉儲節省時間表.....77



第一章 緒論

本章的主要目的在闡述本篇論文的研究目的、方法與架構，共分為五小節。第 1.1 節解釋研究的動機，第 1.2 節論述問題的界定，第 1.3 節說明研究目的，第 1.4 節介紹研究方法，第 1.5 節表達論文的架構。

1.1 研究動機

台灣的花卉產業揚名國際，在台北花卉批發市場成功地建置世界第一個花卉產業資料倉儲後[1][11]，緊接著彰化[7][8]、台中[4][3]、台南[12]三地的花卉批發市場也完成了資料倉儲的建置，將每日拍賣作業的交易資料，定時轉入資料倉儲系統中，放置資料於花卉批發資訊分享熱線（Wholesale Information Sharing Hotline, WISH）[6]，提供花卉相關作業人員存取資料與線上分析，此舉帶動花卉產業的供應鏈進入資訊化的階段，在資訊加值應用上，帶給花卉供應人、承銷人更大的收益，並提供政策制訂的參考方向。

由於各地花卉批發市場的資料倉儲系統皆已成熟化，但限於各地花卉相關人員只能參考各地的花卉資訊，以整個台灣的大環境來說，整合北中南的花卉供應鏈，可以為台灣整體的花卉經濟帶來莫大的收益，因此催促台灣區花卉總資料倉儲的誕生。各地花卉資料超市的使用者，將可以在總資料倉儲中，參考其他家花卉資料超市的資訊，經由比較與整合，瞭解整個大環境的趨勢。並且在下載資料的速度上更加提升，利於做更進一步的資料加值應用。

在花卉資料倉儲中，拍賣的交易資料為時間序列，從過去的歷史資料，往往可以觀察出花卉銷售與時間的相關性，原本台北、台中、彰化和台南四家資料超市中，已建立有時間、星期和農曆[3]三個時間維度，經由花卉拍賣工作人員的長期觀察，認為花卉拍賣的銷售在節慶期間有非常明顯的趨勢現象，因此原本三個時間維度已不敷使用，再建立一個節慶維度[9]，將可以增加資料倉儲在使用上的價值。但節慶牽涉到國曆與農曆的日期，每年的農曆節慶對應到的國曆日期都不一樣，且不同的節慶期間又會互相重複到，在建置維度上是一個有待解決的問題。

由於建置好的總資料倉儲，在預期上可使下載資料的效率提升，因此本研究利用總資料倉儲內的資料，來做資料挖礦技術的應用。資料挖礦技術在資料加值的應用上已經是個主要趨勢，利用資料倉儲的資料建立迴歸方程式時，有時會發現獨立變數與相依變數的相關性很高，但事實上獨立變數為一隨機變數，將導致迴歸方程式係數的估計不準確，由獨立變數與殘差相關性高可得知此一性質。此時可透過外來的量測變數(Instrumental Variable, IV)去估計獨立變數以降低其不確定性，若量測變數與獨立變數相關性高，與殘差無相關，且最後判斷出其與相依變數有高相關性，則找出此為一成功的量測變數。在花卉資料倉儲中，獨立變數為一隨機變數是個很容易出現的情況，因此找出特殊的量測變數來估計獨立變數，去除不確定性，將可以用來預測花卉的重要資訊。

1.2 問題界定

花卉供應鏈的決策者可分為三類，分別是供應端的花卉供應人、負責中介的花卉批發市場經營者和需求端的花卉承銷人。這些決策者分別對花卉的行情資訊有著不同的需求，花卉供應人希望透過資訊的蒐集來選擇批發市場以及供應數量；花卉批發市場經營者希望穩定市場價格以及減少殘貨數量；花卉承銷人希望以最有利的價格來購買相當品質的花卉。

為了提供更多資訊給全台灣的花卉供應鏈決策者，使之能獲取其他地區供應鏈的情報，提升台灣區花卉整體的經濟，本論文利用競箴 (R. Kimball) [14]提出的由下而上建置法 (Bottom up)，將台北、台中、彰化、台南、高雄五家花卉資料超市，整合為總資料倉儲系統，由於高雄資料超市尚未完工，因此本論文將不討論高雄花市之部分，圖 1.1 為總資料倉儲的系統架構，前段虛線框住的區域為本論文完成之部分，將台北、台中、彰化、台南四家資料超市之資料分別經由資料轉換服務，轉入總資料倉儲，並加入新的節慶維度與市場別維度，分別在本論文第三章與第四章詳述作法，後段無虛線框住的區域則為陳佳佑同學[10]於同年完成之部分。

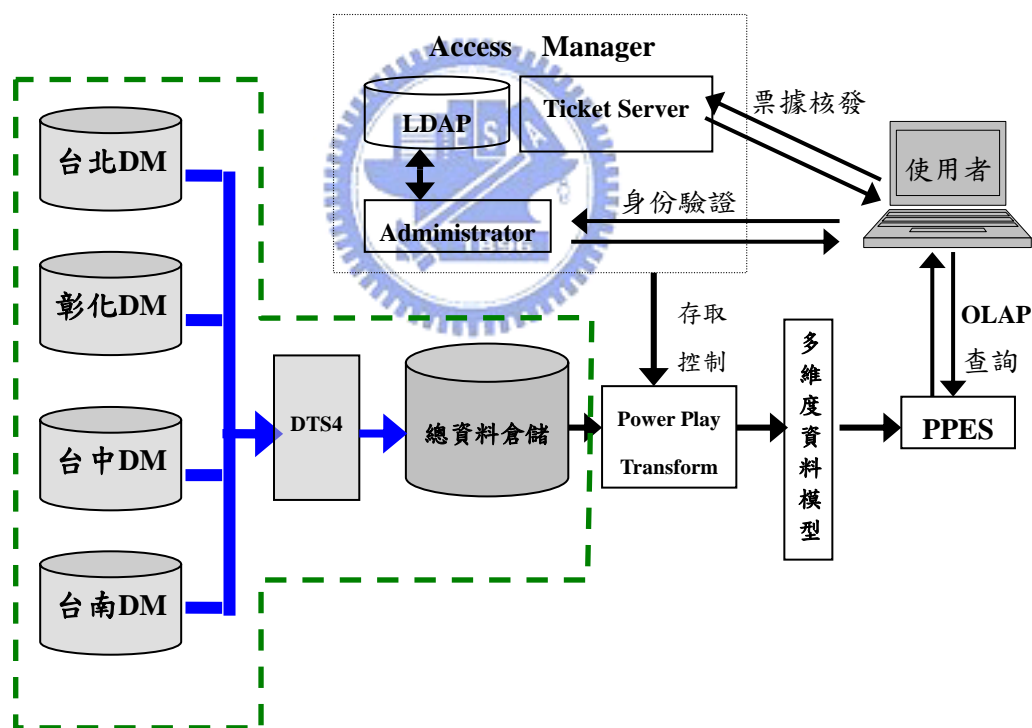


圖 1.1 總資料倉儲系統架構圖

供應鏈進行決策時，常需以定量的方式進行最佳化方案選定，在花卉供應鏈中，數量與價格常呈現反比關係，若可以知道其定量關係就可進行最佳化決策。如圖 1.2 所示，本論文在第五章部分，將利用總資料倉儲內的花卉資料，利用檢定與迴歸技術，選出一個具代表性的花卉，亦即其價格與交易量間具有高相關性的函數。接著去測定其交易量本質是否為隨機變數，若交易量之本質為隨機變數，則利用關聯式的發想，去找尋一外來之變數。測定外來變數之性質與交易量間成高相關性，與殘差無相關，且最後判斷出其與相依

變數有高相關性，則此一外來變數可作為成功降低獨立變數不確定性的量測變數。本論文使用時間序列工具來進行迴歸分析，先利用單位根檢定確認變數皆為統計非時變序列，再利用內建的兩階段最小平方法(2-stage Least Squares Method)[19]來做量測變數的迴歸估算。

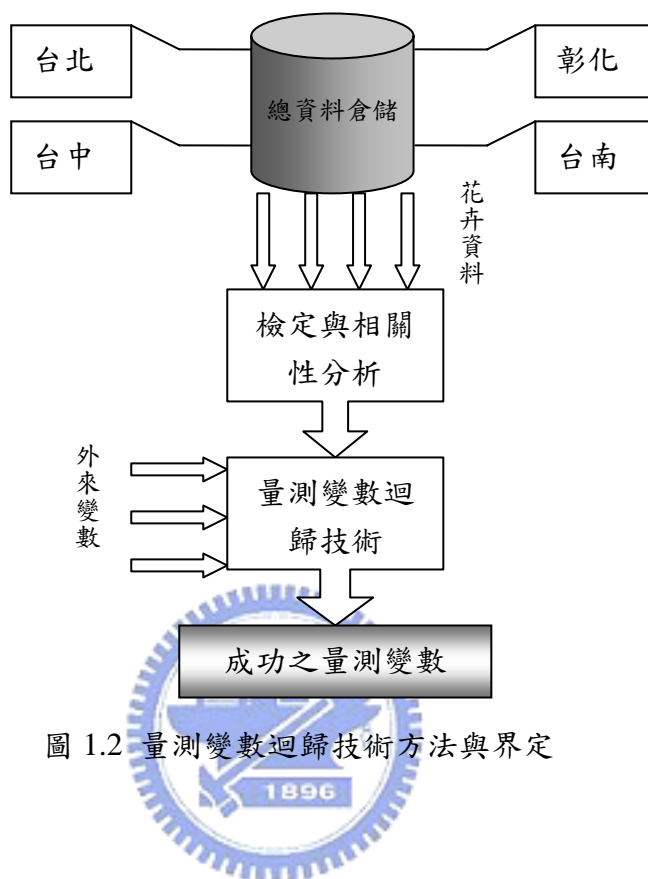


圖 1.2 量測變數迴歸技術方法與界定

1.3 研究目的

本論文的研究目的，在向上整合已建置好的各家花卉資料超市，成為完整的總資料倉儲，使花卉供應鏈的決策者能整合各家花市的資訊，來做最佳化決策，同時使各類決策者無法侵犯到其他決策者之權益與機密，使台灣的花卉產業能整體提升經濟效益。表 1.1 為總倉儲建置完成後，資訊的分享情形。

表 1.1 資訊分享示意表

資料倉儲 決策者	台北 DM	台中 DM	彰化 DM	台南 DM	高雄 DM	總倉儲
台北花市	◎					◎
台中花市		◎				◎
彰化花市			◎			◎
台南花市				◎		◎
高雄花市						◎

在過去各家花卉資料超市中，並沒有節慶期間的相關查詢，為配合花卉供應鏈決策者的需求，加入新的節慶維度，幫助決策者在選取報表資訊時，能有更彈性的選擇空間，表 1.2 為過去與加入節慶維度後，能使用的維度查詢項目比較。

表 1.2 舊維度與新維度比較表

資料倉儲 維度	各家資料超市	總資料倉儲
星期維度	◎	◎
供應類別維度	◎	◎
花卉種類維度	◎	◎
容器維度	◎	◎
農曆維度	◎	◎
日期維度	◎	◎
供應地區維度	◎	◎
承銷地區維度	◎	◎
花卉等級維度	◎	◎
拍賣線維度	◎	◎
市場別維度		◎
特殊節慶維度		◎

利用總資料倉儲來下載花卉資料之速度，較分別由各資料超市下載快，因此下載總資料倉儲裡的資料為輸入。由於資料倉儲內之資料，有時會發現其本質上是隨機變數，故其值含有不確定性。在進行迴歸方程式係數的估算時，若獨立變數是隨機變數，這時估計出

來的係數就會有很大的不確定性。獨立變數的不確定性可透過外來的量測變數來降低，此時迴歸方程式的係數就可估的更準確。因此藉由量測變數的找尋，可以提高花卉供應鏈中，預測銷售的準確性。

本研究將對以下幾個問題提出解決方法：

- 1.整合四家花卉資料超市，建置總資料倉儲系統。
- 2.分析台灣地區節慶特性，加入節慶維度於總資料倉儲。
- 3.找尋成功之量測變數，提高花卉需求的預測準確度。



1.4 研究流程

本論文的研究方法及步驟如圖 1.3 研究流程圖所示：

1. 業界訪談：第一批訪談對象為台北、台中、彰化、台南以及高雄各家花卉批發市場管理經營者，確認其對於總資料倉儲之需求，以確定衡量值及維度、權限之內容；第二批訪談對象為各家花卉資料超市設計者與管理者，透過訪談及溝通的過程，了解原始各家資料超市的系統架構，以進一步整合各家資料超市之資料。
2. 問題界定與分析：確定新增節慶維度之內容，包含節慶涵蓋期間的範圍，以及節慶的選擇。
3. 文獻回顧：參考資料倉儲以及時間序列相關文獻。
4. 倉儲系統設計與實作：利用過去所發展的花卉資料倉儲系統理論基礎，實作總資料倉儲系統。
5. 節慶維度實作：配合節慶的特殊性質，擬出幾個設計方案，比較過後選擇一方案實作。
6. 量測變數迴歸技術：利用花卉業務情報網進行資料蒐集，找尋與花卉相關之外在變數，利用時間序列工具進行分析。

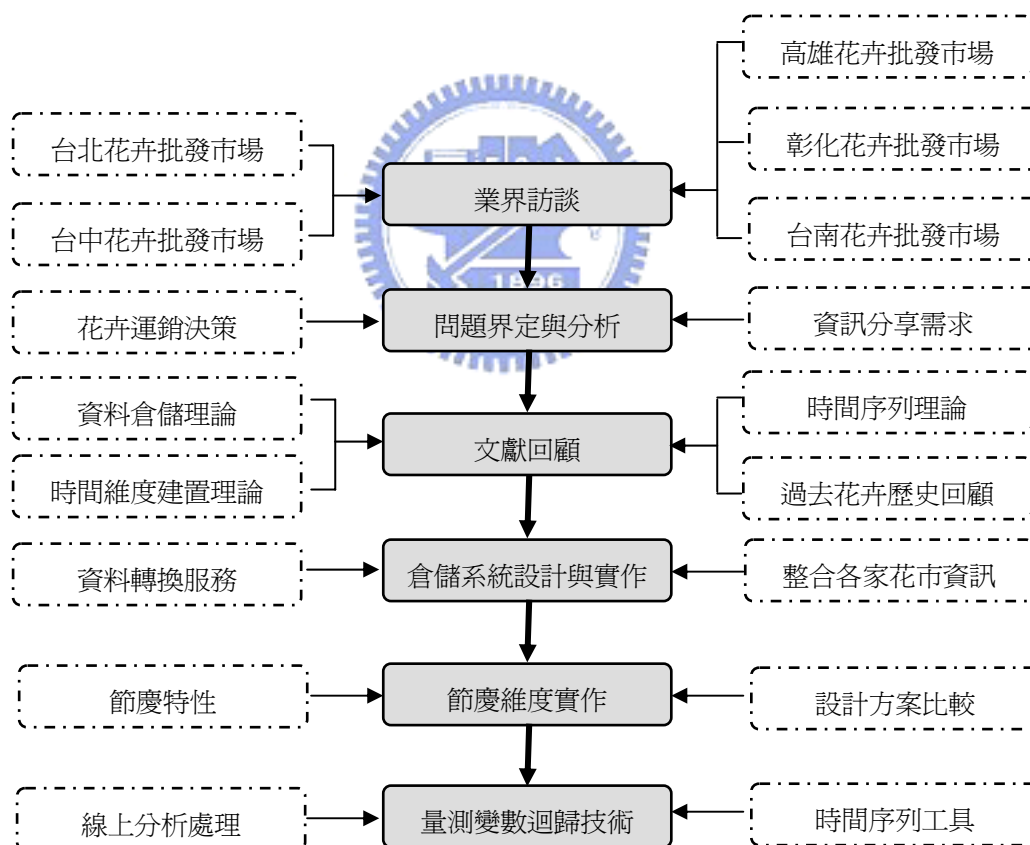


圖 1.3 研究流程圖

在量測變數迴歸技術部份，採用總資料倉儲為資料來源，一次下載可抓取所有花市的資料，節省資料分析時間。分析過程中，首先使用擴大迪-富氏檢定(Augmented Dicky-Fuller Test)技術[15]來進行資料變數的單位根檢定，再加入外在變數分析其與獨立變數和殘差之相關性，最後利用時間序列分析工具進行兩階段最小平方法，尋找出成功的量測變數。

1.5 論文架構

本論文的內容編排如下：

- 第一章：緒論—說明本論文之研究動機、問題界定、研究目的以及研究方法。
- 第二章：文獻回顧—包含資料倉儲的觀念，及量測變數迴歸技術之相關理論。
- 第三章：總資料倉儲系統實作—先介紹總資料倉儲之系統架構，接著整合各家資料超市之資料內容，將其資料利用資料轉換服務轉入總資料倉儲系統。
- 第四章：節慶維度設計—先介紹節慶維度之內容與特性，接著介紹幾種設計方案，最後比較選出最佳方案進行實作。
- 第五章：量測變數迴歸技術—找尋特定花卉的需求方程式，及外來的量測變數，進行二階段最小平方法，最後找尋出成功的量測變數以供預測。
- 第六章：結論—心得、分析及未來研究方向。



第二章 文獻回顧

本章主要的目的在於回顧資料倉儲的理論與架構，並介紹關於量測變數迴歸技術的發展歷史與理論基礎，及特殊時間性質對預測影響的相關文獻。本章共分為三節，第 2.1 節介紹資料倉儲的相關理論與技術及線上分析處理的基本動作；第 2.2 節為節慶維度的相關介紹；第 2.3 節說明量測變數的理論。

2.1 資料倉儲

隨著資訊科技的時代起飛與商業活動的熱絡發展，人們每天接收到大量的訊息，資料倉儲便是一個幫助人們從中擷取有價值資訊的有效工具。資料倉儲廣泛的被使用於各行業，透過擷取、清理的動作將傳統資料庫的資料加以轉換與整合，提供後端決策支援系統的應用。

「資料倉儲」一詞起源於 1990 年，有資料倉儲之父之稱的殷默 (B. Inmon) 所提出，並於 1992 年定義資料倉儲為具有主題導向 (Subject-oriented)、整合性 (Integrated)、隨時間變化 (Time-Variant) 及非揮發性 (Non-Volatile) 四種特性之資料庫[22]。此四種特性簡述如下：

主題導向之意為資料倉儲是建立在特定的主題上，無關主題的資料會被剔除。例如台北花卉批發市場建立的資料倉儲，是以在台北拍賣的銷售為主，並不會存在台中拍賣的資料；整合性意指資料倉儲內的資料都是經過整理合併過的結果，從資料倉儲上擷取的資訊都是整合過的；隨時間變化便是指資料倉儲上的資料為時間序列，隨著日常的資料作業新增，需定時更新與整理，因此說資料倉儲中的資料是會隨時間變化的；非揮發性指的是資料一旦進入資料倉儲，便會永久保存不能刪除。

本節共分成三小節，第 2.1.1 節說明資料倉儲的兩種基本架構；第 2.1.2 節介紹在資料倉儲中，時間維度的重要性；第 2.1.3 節介紹線上分析處理的模式。

2.1.1 資料倉儲架構

當代資料倉儲界的兩大巨擘殷默與競箴，分別在 1990 年與 1996 年提出兩種不同的資料倉儲建立模式：由上而下 (Top-down) 與由下而上 (Bottom-up) 的建構方法。顧名思義，殷默所倡導的由上而下法[22]，以傳統關聯式資料庫工具建立的泛企業資料倉儲為頂點，向下發展出各別部門的資料庫，以提供大部分的決策需求；而競箴提倡的由下而上法[23][24]，將每一種商業程序建立一個資料超市 (Data Mart)，藉由資料通道(Data bus) 和制式化維度(Conformed Dimensions)的概念，將所有資料超市結合為一個總資料倉儲。將兩種資料倉儲的架構模式[14]分述如下：

殷默模式：

殷默所提倡的資料倉儲架構環境，是由組織內所有的資訊系統與資料庫所組成，稱之為為“企業資訊工廠”(Corporate Information Factory, CIF)。並將組織中的整個資料庫環境區分成四個層級：

1. 操作性系統(Operational)：由歷史及交易資料組成，支援組織每日的作業。
2. 原子資料倉儲(Atomic Data Warehouse)：資料由操作性系統彙總粹取而來。
3. 部門(Departmental)：從原子資料倉儲中擷取轉存所需的資料。
4. 個人(Individual)：由部門資料庫將所需的資料暫存於使用者電腦中。

上述的最後三個層級構成資料倉儲，而第一層級的操作性系統由歷史及交易資料組成，支援組織每日的作業，換句話說，也就是支援所有的交易程序，這個層級的資料會由操作性系統轉換到第二層的原子資料倉儲來。於原子資料倉儲中，是將儲存於操作性系統資料庫中的資料，進行彙總與粹取的動作。比較此二層級的資料有何不同，舉個例子說明，將顧客視為實體 (Entity)，而最令人感興趣的屬性為顧客的信用等級，操作性系統的資料庫中就儲存了顧客目前的信用等級、借貸情況、地址等一筆一筆的資料，相反地，原子資料倉儲中儲存的是顧客的信用歷史紀錄，以及經過彙整的年資料。在第三層部門資料庫中，主要將組織中各部門所需要的資料，由原子資料倉儲中擷取轉存。亦即在部門資料庫中儲存著較原子資料倉儲中更少量且高度彙總的資料。在第四層級中，個別使用者可由部門資料庫中，擷取及建立所需的資料集暫存於使用者電腦之中。

在般默架構中，一個原子資料倉儲可用來建立多個部門資料庫，此舉可以有效避免資料產生不一致的情況。般默並提出下列三層式資料模型來協助原子資料倉儲及部門資料庫的建立：

1. 第一層級資料模型：實體關聯圖(Entity Relationship Diagrams, ERD)。在第一層級的資料模型中，各企業首先建立各部門預期使用的資料倉儲實體關聯圖，並藉由各部門的實體關聯圖的合併來完成整體企業的實體關聯圖。
2. 第二層級資料模型：資料項目集合(Data Item Set, DIS)。其表示法如圖 2.1 所示，每一個長方格代表一個部門或團體的邏輯資料表 (Logic Table)，方格間的線段代表連結，在資料項目集合中共包含四個主要元件，分述如下：
 - (1) 主要資料群(Primary Data Grouping)

在每個主題區域中只存在一次。亦即該部門中最主要的資料群。
 - (2) 次要資料群(Secondary Data Grouping)

與主要資料群間相關的資料群，在 DIS 中以位於右邊的方格代表之。
 - (3) 連結(Connector)

在資料表間的關連以線段的連結來表示。
 - (4) 類似型態資料(“Type of” Data)

指一些與次要資料群相關的資料群，在 DIS 中以方格表現在次要資料群方格的右邊分枝。

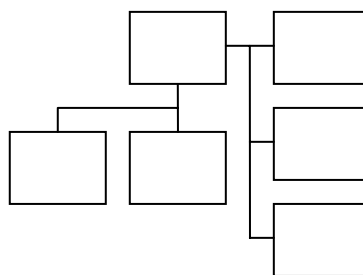


圖 2.1 資料項目集合(DIS)

3. 第三層級資料模型：實體模型(Physical)。在第三層級資料模型中，主要是將第二層級的資料模型加入鍵值及其模型的實體特性。

般默提出以螺旋式發展方法(Spiral Development Approach)，來建構資料倉儲。上述三層資料模型的設計為建置資料倉儲的先決要件之一。螺旋式發展方式如圖 2.2[14]所示，模式步驟以 DSS 為名，有決策支援 (Decision Support) 之意。

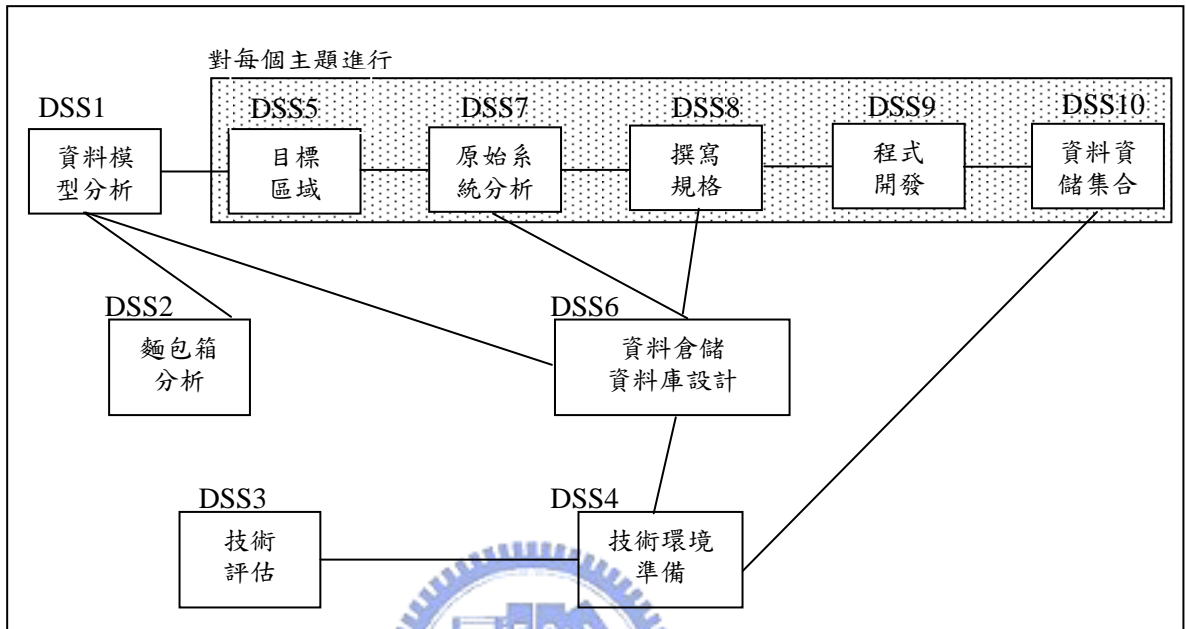


圖 2.2 般默資料倉儲螺旋式發展方式

在螺旋式發展方法過程中，首先企業需分析現有資料系統的資料模型，以幫助企業資料倉儲的建置，這步驟稱為資料模型(Data Model)分析流程（參照圖 2.2 中的 DSS1）。在完成三層式資料模型分析後，以資料模型為流程的輸入，接著進行下一步驟，要衡量資料倉儲內資料的詳細程度，此動作稱為麵包箱(Breadbox)分析(DSS2)。以交易資料為例，企業需評估資料倉儲內部是否要包含每日的交易資料以及彙總式資料。一完成麵包箱分析之後，主要目標區域就被選取(DSS5)，此部分將成為部門的資料庫。接著企業需分析主要區域中的系統來源(DSS7)，並撰寫規格 (DSS8)，及進程式開發(DSS9)，並將完成的資料庫進行總體集合(DSS10)，原子資料倉儲資料庫的設計同時開始 (DSS6)，當有足夠資訊著手上述步驟之後，可以進行技術性的評估 (DSS3)，技術評估工作主要在確保資料倉儲內部資料的可存取性以及是否好管理，也就是說在資料倉儲的設計及建置過程中，相關硬體、軟體、介面及存取軟體是否可以相互配合。在完成技術評估後，依著技術評估步驟的評估報告結果，企業開始著手進行技術環境準備(DSS4)的工作流程，來支援整體資料倉儲的建立。在技術環境準備流程中主要工作在於確定資料倉儲相關軟硬體設備的建立，如部門資料庫間的網路連結、資料儲存硬體設施的建置及作業系統及操作介面的相容性。

當部門資料庫都成功的建置完成，會影響到原子資料倉儲。圖 2.2 中可以看到線連結不同的步驟，代表模型中反覆的觀點，線連在原始系統分析 (DSS7)、撰寫規格 (DSS8) 與原子資料倉儲設計 (DSS6) 之間，表示在每一次新的部門資料庫被建置後，原子資料倉儲設計的步驟都會被再重覆一次，這是由於原子資料倉儲的資料為各部門資料庫的主要資

料來源，因此在資料倉儲資料庫設計的工作流程中，需避免原子資料倉儲與部門資料庫間資料不相容的問題。因此在資料倉儲資料庫設計流程（DSS6）中，對於原子資料倉儲的設計上需與資料模型分析（DSS1）、技術環境準備（DSS4）、原始系統分析（DSS7）及規格撰寫（DSS8）四個流程相互配合。連結在部門資料庫集合（DSS10）、技術評估（DSS3）、資料倉儲資料庫設計（DSS6）與技術環境準備（DSS4）的線，也顯示出螺旋式發展方法自然的反覆性，就是說技術環境準備流程的進行，除了需要技術評估流程的支援外，還需要與資料倉儲資料庫設計及資料倉儲集合流程進行配合。

般默的模式採取傳統的資料庫工具來建構資料庫，像是實體關連圖、資料項目集合，因此必須維持資料的一致性，達到以最小的記憶體來儲存資料，這使得查詢時間變快，但企業內的資料通常有幾千幾百個實體，因此建構之初複雜性很高。在由上而下的建構過程中，系統開發者對組織架構及系統流程必須有相當程度的瞭解，終端使用者不易瞭解此模式，也不容易使用查詢、操作的功能，因此般默的模式比較適合專業的 IT 人員與高階管理者來使用。建置一個般默模式的資料倉儲，需要長時間（約 4~9 個月）的運作，初期需要較高的成本，以儲存大量資料，適合資料倉儲專家用於泛企業需要的大型專案。

競簇模式：

競簇的模式架構以維度資料建模法(Dimensional Data Modeling)來建立資料倉儲。在維度資料建模法中，採用資料表做為建構的基礎，而非傳統的實體關聯圖。資料表包含事物表(Fact Table)及維度資料表(Dimensional Table)。

事物表內存放大量且不重覆的資料列，可能有上千萬或上億筆歷史資料，欄位相對之下較少，資料為數值型態，只有少數做為關聯索引的文字資料，表內存在許多和維度資料表做關連的外鍵(Foreign Key)。資料內容屬於靜態資料，一但載入便不再做任何修改異動。而維度資料表與事物表比較起來，資料量極小，可能只有上百筆或上千筆資料，大部份為文字型態，表內欄位較多，儲存事物表內所有資料的屬性，用來管理維度裡的階層關係，屬於可以修改異動的動態資料型態。

除此之外，競簇架構的維度資料模型可進一步區分為三種主要的綱要結構：星狀綱要(Star Schema)、雪花綱要(Snowflake Schema)、星座綱要(Fact Constellation Schema)[17][20]。其中星狀綱要為其中最常見的結構，其結構如圖 2.3 所示：

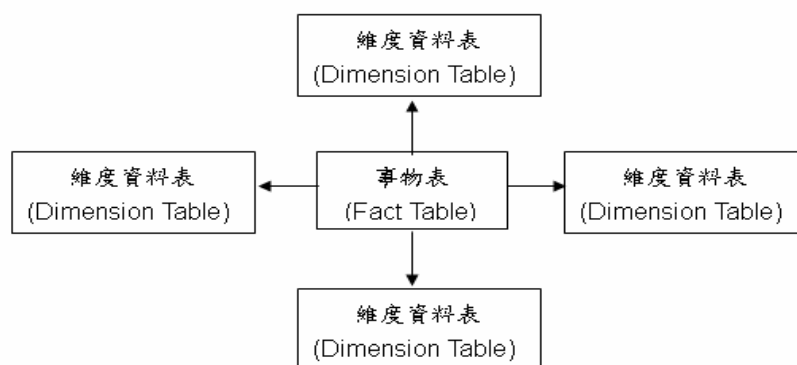


圖 2.3 星狀綱要圖

星狀網要以事物表為中心連結相關的維度資料表，其形狀如星狀呈發射狀。而雪花網要與星狀網要圖類似，主要的差異在於雪花網要對於維度資料表進行更進一步的三階正規化工作，以避免重覆性資料及減少資料的儲存空間，其形狀如雪花故稱之雪花網要圖。圖 2.4 為雪花網要之示意圖。

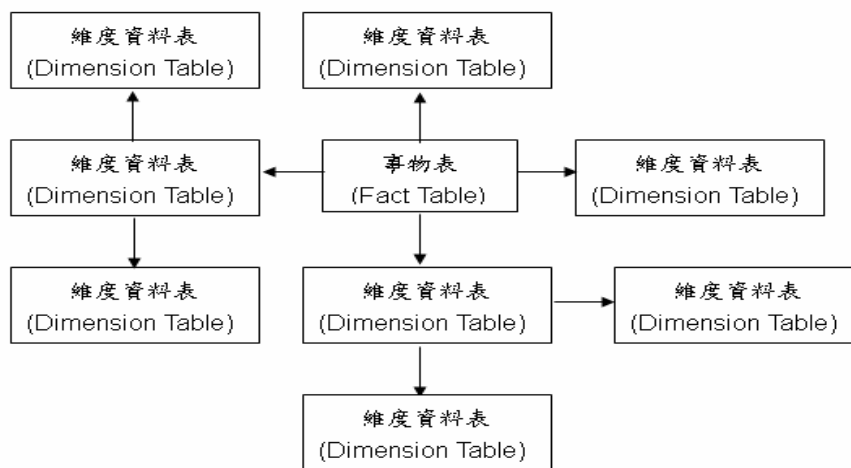


圖 2.4 雪花網要圖

第三種網要架構為星座網要，又稱為銀河網要(Galaxy Schema)。星座網要是某些維度資料表被多個事物表共同參考，如圖 2.5 所示。

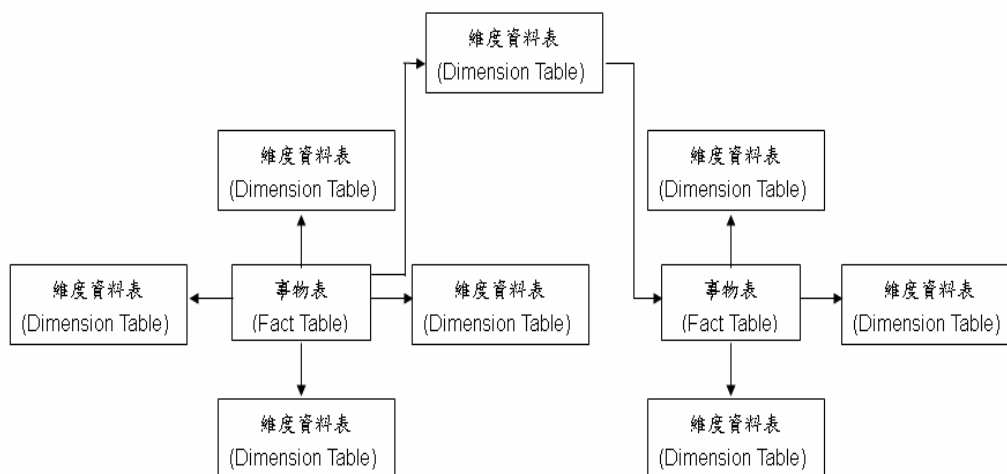


圖 2.5 星座網要圖

競簇模式以由下而上的方式建構資料倉儲，一個資料倉儲的建立是基於單一的商業流程，競簇稱此種架構的資料倉儲為資料超市 (Data Mart)。對於企業組織而言，內部可能同時包含多個商業流程，因此在資料倉儲的建置過程中，需為每個商業流程建立資料超市，然後將數個資料超市向上整合為一個整體的資料倉儲。在競簇模式架構中，資料是由操作來源系統複製到集結區域，此資料經由 ETL 的過程 (Extract Transform Load process)，符合一致性並易於終端使用者查詢，而資料會從集結區域被下載到資料超市。在完成資料

超市的建立後，系統開發者利用資料通道(Databus)及制式化維度(Conformed Dimensions)的概念，將數個資料超市整合為一個整體的資料倉儲。資料通道的架構是指所有的資料超市都必須使用標準制式化的維度，制式化維度的基本需求，即是資料表中的鍵值(Key)、欄位名稱、屬性定義與屬性的值必須保持跨商業流程的一致性。

競箴建議以一次建置一個資料超市的方式為其發展方法論，並針對單一資料超市的建構，提出四個維度化的設計流程步驟：

1. 選擇商業流程：競箴給予商業流程很廣泛的定義，舉例包含零售的銷售點、存貨、訂貨、運送等，這在大部分的企業組織中都是跨部門的。例如一個訂貨流程會牽涉到銷售、市場行銷、財務和存貨控制人員。建議選擇一個最有影響的流程作為資料倉儲的第一個商業程序，最有影響指的是選擇此流程可以解決最迫切的商業問題，並且已經準備好作資料的取出。
2. 定義資料細度：定義資料細度是指決定資料倉儲中資料詳細程度的流程。最低層級的資料細度稱為原子(Atomic)，表示為不能再被分割的資料。選擇原子層級的資料細度是非常重要的，因為使用者總是需要整合他們所需的資料，當選擇一個彙總的層級當作原子，表示資料倉儲將不能滿足在此層級之下的查詢。例如在資料倉儲內的最低層級為月資料，則使用者將無法查詢到日資料。
3. 選擇維度：在完成資料細度的定義後，下一步驟選擇維度資料表。維度資料表內包含大量的屬性，主要提供使用者對於資料倉儲中的資料有更多方向的查詢。
4. 確認事物表中的衡量值：最後一個步驟為決定事物表中的衡量值。衡量值是指經過計算或非經計算的數值資料，一般都是透過維度資料表來查詢衡量值。簡單來說，維度資料表的用意是將查詢分類化，而衡量值則是經由查詢出的結果。

競箴提倡的維度資料建模法，以資料表為架構，包含事物表及維度資料表，由於維度資料表的去正規化，因此資料沒有不一致性的問題，多維度的建構使得終端使用者可藉由任何一個維度查詢到資料，在建構與使用上的複雜性均低。競箴著重於流程導向的資料模式，試圖去找出不同商業流程間的交互作用，也就是找出資料倉儲中重要的衡量值(Facts)與屬性(Dimensions)。建構一個競箴模式的資料倉儲，花的時間比般默模式較少，且初期不需投入太多的成本，以每個主題的資料超市而言，成本大約是相同的，適合用於組織內的一般小團體來做專題型的資料倉儲開發。

儘管般默和競箴的模式有諸多不同，但基本的資料來源都是採用時間序列的資料，『時間』為資料倉儲最重要的特徵，藉由時間的不同可以比較與分析出各式各樣的關連性，創造隱含的價值；而另一項共同點，是兩種模式都使用 ETL 程序來進行資料轉換，也就是將原始資料由作業系統移至儲存空間，然後將資料做轉換，轉換可以是刪除、轉換表格等等的複合作業，轉換後的資料將儲存於資料倉儲中，這可以保證資料倉儲的整合性，但也是建構時最花心力的部分。表 2.1 為般默及競箴模式基本特性之比較[14]。

表 2.1 般默及競篋模式基本特性之比較

	般默模式	競篋模式
理論與架構(Methodology and architecture)		
整體方法	由上而下	由下而上
建構架構	以泛企業(原子)資料倉儲提供部門資料庫。	以單一商業流程建立一個資料超市，透過資料通道及制式化維度，將各資料超市達成一泛企業資料倉儲。
方法複雜度	相當複雜	頗為簡單
建構方法論之比較	使用螺旋式發展方式	使用四步驟流程建立一部門資料庫
實體設計	相當完整	相當輕微
資料建模(Data Modeling)		
資料導向	主題或資料導引	流程導向
工具	傳統方式(ERDs,DISs)	維度建模法
終端使用者存取性	低	高
哲學(Philosophy)		
主要觀眾	IT 專家	終端使用者
組織中之位置	整體企業資料工廠(CIF)	操作性資料的轉換與保留者
目的	利用經驗証的資料庫理論或技術基礎，來提供一個合理的技術解決方法	提供一個解決方法，使得終端使用者能輕易的存取資料，並獲取合理的回應時間

2.1.2 資料倉儲之時間性

在每個資料超市或是資料倉儲中，時間維度是獨一無二並且有強大功能的維度。雖然多維度模型的其中一個宗旨，便是要一致性的對待所有維度，但事實上，時間維度具有特殊性，必須以不同於其他維度的方式處理，以下將介紹幾個競篋於 1997 年提出[25]，設計時間維度時必須注意的觀念與重點。

實際上，每個資料超市都是由時間序列所構成，建立時間維度的初始理念，是源於資料超市需要維度資料表的基本理由：維度資料表提供對來源的限制以及來源資料的報表表頭，有好的維度資料表才會有好的資料超市。由於不同企業的日程表都有個別化的成分在，例如季節性、財務日程等，因此直接建立屬性於時間的資料表是最容易也最符合個別化的方法。

而建立時間維度，只需要在試算表 (Spread Sheet) 上建立即可。與其他維度不同的是，時間維度可以一次在設計者的個人電腦中完成，只要將之上傳到所有資料超市的機器上，便可組成於泛企業的資料倉儲。有些資料超市另外要追蹤一天的資料到分鐘甚至是秒的情形，並不適合將一天的時間都建在同一張時間資料表上，這種情形最好是將分與秒的值，建在另一張資料表。

時間維度容易造成互不相容 (Incompatible) 的上捲 (Rollup) 情形，最明顯的例子是星期和月份，這是由於每個月份中的星期並不規律，舉個例子來說，一月份的第一天必定為一月一日，但一月份的第一天未必為星期一，所以通常的作法是將不相容的時間分開為另外的維度，星期、農曆、節慶都是這種情況。

最具有彈性的基底是天 (Daily Time Grain)，天可以上捲至任何可能的日程，即便是無法相容的節慶期間，如復活節假期、聖誕節假期...等等，但需注意節慶期間無法再上捲至其他日程，因此只要一開始的基本單位是天，要加註任何季節性的解釋到日程表都是可以的。

2.1.3 線上分析處理

線上分析處理(On-Line Analytical Processing, OLAP)源自於 1993 年 Codd E.F 所提出 [20]，為一種可以直接存取資料倉儲做為資料分析的應用程式，並非單純的報表工具。線上分析處理能簡化一些使用著的複雜的查詢、製作報表、對資料的計算加總、將資料過濾后分割更具商業意義的細部資料。就功能面來說，線上分析處理可以說是整個資料倉儲/商業智慧的核心引擎，它彙整資料庫裡的原始資料，並將之轉成多維度的分析模組，將原始零散的歷史資料加值成有意義的資訊，便於決策者做決策分析。

資料倉儲裡資料是以超方體(Cube)的形式存在，透過線上分析處理對維度及衡量值的操作，即可對資料做加值應用，獲取所欲得到的資訊。在每個維度裡，資料是有層級式的繼承觀念。例如，在時間的維度裡可以分成年、季、月三層，亦或是年、季、月、旬、日五層，可依需求而設定；在地區的維度裡可以分成北中南區，再細分各區的縣市別，最後一層則是鄉鎮市。超方體裡維度的架構如圖 2.6。

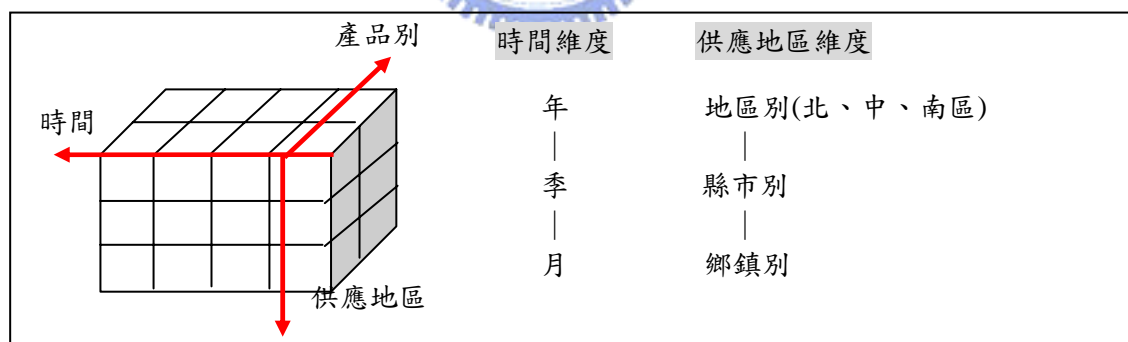


圖 2.6 超方體架構

線上分析處理的操作可分為十種，與維度有關的操作有：上捲(Roll-Up)、下挖(Drill-Down)、切片(Slicing)、切丁(Dicing)與轉軸(Pivoting)五種。其中上捲是指將維度的單位放大；下挖是將維度的單位縮小；切片是限制某維度的上下限範圍；切丁是限制多維度的上下限範圍；轉軸是把水平維度與垂直維度做交換。

線上分析處理的操作中與衡量值有關的操作有：視覺化(Visualize)、篩選(Selecting)、輸出(Output)、排序(Sorting)及計算(Computing)五種。其人視覺化是把數值化的衡量值用圖形表達；篩選是限制衡量值的範圍；輸出是將衡量值另外儲存供加值應用；排序是將衡量值由大至小或由小至大排列；計算則是用程式來修改衡量值。

2.2 節慶維度相關介紹

台灣地區花卉的供給需求與節慶時節有相當大的關係，在節慶上使用的日曆與其他國家不同，有西曆與農曆之分，尤其台灣地區對於農曆的時令節氣相當重視，因此在花卉資料倉儲的節慶維度設計上，必須同時考量西曆與農曆的節慶，目前關於節慶維度的設計尚無前人之研究可以參考。

單純考量農曆的時間維度設計，在 2005 年國立交通大學的柯珮婕同學於其碩士論文中[3]，提出農曆的維度設計，其中提到農民曆一般都被稱為「陰曆」，但其實是一種「陰陽合曆」，與太陽及月亮的運動有關，因此每個農曆月的長短沒有規則可循，而只能由觀測太陽及月亮的運動所決定。如要完全正確的寫出西曆與農民曆的對照程式，則必需考慮到天文的數學式，故其採取折衷方法，使用中央氣象局天文台所提供之資料直接建構農曆維度，這可減少建構的時間。

關於其農民曆的維度架構，新增三個資料表，分別是農曆對照表、農曆月及農曆年。農曆對照表直接記錄農曆月日及西曆月日的對照，屬性欄位包含西曆日期、農曆日期、農曆日期的中文名稱，及農曆年月。農曆月資料表的屬性欄位則是農曆年月、農曆年月的中文名稱及農曆年。農曆年資料表的屬性欄位包含農曆年及農曆年的中文名稱。三個資料表的關聯如圖 2.7，最後再與資料倉儲的事物表(Fact Table)連結，便完成可下挖的農民曆維度。



圖 2.7 農民曆三資料表關聯圖

表 2.2、表 2.3、表 2.4 分別為其設計之農曆對照表、農曆月、農曆年三個資料表的屬性的詳細說明，包含資料表的主鍵、外鍵、資料欄名稱、資料型別、長度、可否為空值及相關說明。

表 2.2 農曆對照表

主鍵	外鍵	欄位名	資料型別	允許空值	長度	說明
◎		F_DATE	Datetime		8	西曆日期，ex: 2004/08/22
		C_DATE	Char		7	農民曆日期，ex: 0930707
		C_DNAME	Char		16	農民曆中文，ex: 農曆93年07月07日
	◎	C_MOON	Char		5	農民曆年月，ex: 09307

表 2.3 農曆月

主鍵	外鍵	欄位名	資料型別	允許空值	長度	說明
◎		C_MOON	Date time		5	農民曆年月，ex: 09307
		C_MNAME	Char		20	農民曆年月中文，ex: 農曆93年07月
	◎	C_YEAR	Char		3	農民曆年，ex: 093

表 2.4 農曆年

主鍵	外鍵	欄位名	資料型別	允許空值	長度	說明
◎		C_YEAR	Char		3	農民曆年，ex: 093
		C_YNAME	Char		10	農民曆年中文，ex: 農曆93年

詳細說明農曆維度的架構關係，在此以實際資料及關聯來說明，如圖 2.8。三個資料表皆有一個中文欄位名稱，目的是在日後線上分析處理時，資料能夠以較友善的中文方式呈現，如以「農曆年 83 年 12 月 01 日」表現而非「0831201」。另外利用紅色框線的欄位來當做關聯，使得農曆維度可以達到上捲與下挖等目的。

農曆對照表				農曆月			農曆年	
F_DATE	C_DATE	C_DNAME	C_MOO	C_MOON	C_MNAME	C_YEA	C_YEAR	C_YNAME
1995/1/1	0831201	農曆83年12月01日	08312	08312	農曆83年12月	083	083	農曆83年
1995/1/2	0831202	農曆83年12月02日	08312	08401	農曆84年01月	084	084	農曆84年
1995/1/3	0831203	農曆83年12月03日	08312	08402	農曆84年02月	084	085	農曆85年
1995/1/4	0831204	農曆83年12月04日	08312	08403	農曆84年03月	084	086	農曆86年
1995/1/5	0831205	農曆83年12月05日	08312	08404	農曆84年04月	084	087	農曆87年
1995/1/6	0831206	農曆83年12月06日	08312	08405	農曆84年05月	084	088	農曆88年
1995/1/7	0831207	農曆83年12月07日	08312	08406	農曆84年06月	084	089	農曆89年
1995/1/8	0831208	農曆83年12月08日	08312	08407	農曆84年07月	084	090	農曆90年
1995/1/9	0831209	農曆83年12月09日	08312	08408	農曆84年08月	084	091	農曆91年
1995/1/10	0831210	農曆83年12月10日	08312	08409	農曆84年09月	084	092	農曆92年

圖 2.8 農民曆三資料表資料內容



2.3 量測變數迴歸技術

資料倉儲裡的資料常常在本質上是隨機變數，故其值含有不確定性。進行迴歸方程式係數估算時，如果獨立變數是隨機變數，此時估計出的係數會有很大的不確定性，有時獨立變數可以透過量測變數 (Instrumental Variable, IV) [28] 來降低其不確定性，此時其迴歸方程式的係數將可估計的更準確。另外，在利用時間序列資料進行迴歸動作時，必須注意到自變數與因變數都必須為統計非時變序列，否則會出現假性迴歸的情形，採用單位根檢定法可檢定序列是否為統計非時變序列。本節共分四小節，第 2.3.1 節介紹線性迴歸基本原理；第 2.3.2 節說明量測變數迴歸技術之方法；第 2.3.3 節介紹時間序列資料；第 2.3.4 節介紹單位根檢定法。

2.3.1 線性迴歸基本原理

線性迴歸方法為計量經濟 (Econometrics) 領域最常見的分析方法之一[19]，主要在探討兩個或兩個以上變數之間的關係。而迴歸分析的目的在于了解並且建立一個因變數(Y)與一組自變數(X)間的關係。由一個因變數 Y 及 m 個自變數組成的多元線性迴歸模式，其形式表示為下列式子， β_0, \dots, β_m 為係數， ε 為誤差項。

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon$$

若有 n 個樣本資料，則可表示成

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \dots + \beta_m x_{m1} + \varepsilon_1 \\ \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \dots + \beta_m x_{m2} + \varepsilon_2 \\ \vdots \\ \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \dots + \beta_m x_{mn} + \varepsilon_n \end{bmatrix}$$

改以矩陣來表示，則可表示成

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{m1} \\ 1 & x_{12} & \dots & x_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & \dots & x_{mn} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$Y = X \cdot \beta + \varepsilon$$

[n × 1] [n × (m+1)] [(m+1) × 1] [n × 1]

此一多元迴歸模式 $Y = X \cdot \beta + \varepsilon$ 可用最小平方法來求迴歸係數 β 的數值。

最小平方法的目的在于找出未知係數的數值，使誤差平方和 (error sums-of squares, SSE) 為最小。可經由將下式對 β 偏微分得到。

$$SSE = \varepsilon' \cdot \varepsilon = (Y - X\beta)' \cdot (Y - X\beta)$$

其中， ε' 為 ε 之轉置矩陣。

此過程可得到 $m+1$ 個未知數的聯立方程式，稱為最小平方理論的常態方程式。以矩陣表示之常態方程式為： $(X'X)\beta = X'Y$ 。兩邊各乘以 $(X'X)$ 的反矩陣，即 $(X'X)^{-1}(X'X)\beta = (X'X)^{-1}X'Y$ 。故得到 $I\beta = (X'X)^{-1}X'Y$ ，即 $\beta = (X'X)^{-1}X'Y$ 。如此可求得迴歸係數如下。

$$\beta = (X'X)^{-1}X'Y$$

在一般的情況下，最小平方法能滿足高斯馬可夫定理[21]的五大假設，如下所示：

- (1) 常態性 (Normality)

- (2) 平均數為 0 (Zero mean)
- (3) 變異數齊一性 (Homoskedasticity)
- (4) 無自我相關性 (Nonautocorrelation)
- (5) 自變數 X 為非隨機 (Nostochastic X)

當誤差項與自變數有相關性時，最小平方方法違反高斯馬可夫定理的第五大假設，將不滿足一致性，因此估計出來的值會不精確，2.3.2 節介紹的量測變數迴歸技術，可以解決上述問題。

2.3.2 量測變數迴歸技術方法

量測變數迴歸的起源，在西元 1928 年時，美國政府欲調高進口黃油的關稅，因此藉由黃油在供給與需求的彈性分析，來獲取最大之利潤。但學者 P.G.Wright 與 S.Wright 卻發現黃油每個時間的交易點，並不在同一條供給曲線上。這些交易點是由不同的供給與需求曲線交集形成，因此無法藉由最小平方方法正確地估計迴歸方程式，做供給與需求的彈性分析。於是學者 P.G.Wright 與 S.Wright 發明了量測變數迴歸的方法，藉由外來變數來消弭供給曲線的隨機成分，利用兩階段最小平方方法得以成功的估計迴歸方程式。

量測變數迴歸的文獻，最早在 1928 年由 P. G. Wright 與 S. Wright 父子於”The Tariff on Animal and Vegetable Oils”的附錄 B 中發表[28]，證明了當普通最小平方估計式 (Ordinary Least Squares Estimator, OLS) 不可行時，量測變數可以用來估算內生變數 (Endogenous Regressor) 的係數。量測變數迴歸技術常與兩階段最小平方方法劃上等號，使用量測變數來估算隨機變數時，可得到除去隨機成分後的觀察值，此部分使用最小平方方法來估計，而用獨立變數來估算其迴歸方程式係數時，亦使用最小平方方法來估計，故此方法又稱為兩階段最小平方方法 (2-stage Least Squares Method, 2SLS)。

設模型為線性，其形式為

$$y = X\theta + \varepsilon \quad (1.1)$$

其中 θ 為自變數 X 的係數向量，當 X 存在與誤差項 ε 相關時，此時可利用量測變數 z 來代替 X 進行迴歸。以下進行量測變數之數學分析，假設迴歸模型為 $y = bx + \tau$ ，則自變數與誤差項存在相關，表示為 $E(x\tau) \neq 0$ ，以量測變數 z 估計 x，代表 $x = z + \eta$ ，此時變數 z 必須與殘差無相關，表示為 $E(z\eta) = 0$ ，因而推導出下列算式：

$$\begin{aligned} E(z\eta) &= E(z(x - z)) = E(zx) - E(z^2) = 0 \\ \Rightarrow E(z^2) &= E(zx) \neq 0 \end{aligned}$$

代表 x 與 z 高相關性，此稱作量測相關性。

將 z 對 y 作迴歸，可得出 $y = b(z + \eta) + \tau = bz + (b\eta + \tau)$ ，此時表示 z 與殘差無相關，表示為

$$\begin{aligned} E(z(b\eta + \tau)) &= 0 \\ \Rightarrow E(zb\eta) + E(z\tau) &= 0 \\ \because bE(z\eta) &= 0, \therefore E(z\tau) = 0 \end{aligned}$$

$E(z\tau) = 0$ 代表 z 與 τ 無相關性，此稱為量測外部性。一個好的量測變數必須具有上述兩大性質，也就是量測變數必須與自變數高度相關，並且與原始迴歸式之殘差項互不相關。

當迴歸模型為 (1.1) 式時，可以知道 $\lim_{N \rightarrow \infty} \frac{1}{N} Z^T \varepsilon = 0$, $\lim_{N \rightarrow \infty} \frac{1}{N} Z^T X \neq 0$ ，量測變數之估計式寫為

$$\begin{aligned}\hat{\theta} &= (Z^T X)^{-1} (Z^T y) - (Z^T X)^{-1} Z^T \varepsilon \\ &= (Z^T X)^{-1} Z^T (X\theta + \varepsilon) \\ &= \theta + (Z^T X)^{-1} Z^T \varepsilon\end{aligned}$$

如果 $\lim_{N \rightarrow \infty} \frac{1}{N} Z^T \varepsilon = 0$ 條件成立，當樣本數趨近於無窮大，下列式子可證明量測變數估計式符合一致性。

$$\begin{aligned}\lim_{N \rightarrow \infty} \hat{\theta} &= \theta + \lim_{N \rightarrow \infty} \frac{(Z^T X)^{-1}}{N} \lim_{N \rightarrow \infty} \frac{Z^T \varepsilon}{N} \\ &= \theta\end{aligned}$$

當量測變數估計式是建立在一個非隨機的工具上，則很容易證明他是無偏的，且具有下列變異數

$$\text{Var}(\hat{\theta}) = \sigma^2 \sum_{*t} z_{*t}^2 / \left(\sum x_{*t} z_{*t} \right)^2$$

其中 x_{*t} 和 z_{*t} 皆為該變數與其平均數之差。OLS 估計式 θ 的變異數為

$$\text{Var}(\theta) = \sigma^2 / \sum x_{*t}^2,$$

若 x_t 和 z_t 之間的相关係數為

$$r_{xz}^2 = \frac{\left(\sum x_{*t} z_{*t} \right)^2}{\sum x_{*t}^2 \sum z_{*t}^2}$$

則 $\hat{\theta}$ 的效率性為

$$\frac{\text{Var}(\theta)}{\text{Var}(\hat{\theta})} = r_{xz}^2$$

那麼 x_t 和其量測變數 z 之間的相关性愈高，量測變數估計式就越有效率。此結果實際上可推廣到更一般的情況，也就是具有隨機自變數的模型： x_t 和 z_t 的元素相关性愈高，量測變數的估計過程就愈有效率。另一方面，如果 $\lim_{N \rightarrow \infty} \frac{1}{N} Z^T X$ 愈接近 0 矩陣，該方法的估計結果就會很差。

2.3.3 時間序列資料

對於時間序列資料而言，一組觀測值 $\{X_1, X_2, \dots, X_N\}$ ，若沿著時間先後有順序地產生，則稱此組觀測值為一時間序列，而正整數 N 被稱為時間序列的長度。就資料特性而言，

時間序列資料可分為統計非時變程序及統計時變程序兩種。關於統計非時變程序的特性，定義如下：

嚴格統計非時變程序(Strictly Stationary)：

若一隨機過程 $\{X_t\}_{t=1}^{\infty}$ ，在任一 n 個時間點內 $(X_t, X_{t+1}, \dots, X_{t+n})$ 的聯合機率分配，與另一組 n 個經由平移 k 單位時間之 $(X_{t+k}, X_{t+k+1}, \dots, X_{t+k+n})$ 的聯合分配相同，則稱此隨機過程為嚴格統計非時變程序。

弱式統計非時變程序(Weakly Stationary)：

若一隨機過程 $\{X_t\}_{t=1}^{\infty}$ 滿足以下三個條件：

(1) $E(X_t) = \mu$ ，(2) $Var(X_t) = \gamma_0$ ，(3) $Cov(X_t, X_{t+m}) = \gamma_m$

在 $t=1, 2, \dots$ 均成立，則稱此隨機過程具弱式統計非時變程序。

傳統的迴歸分析中，是在假設資料為統計非時變序列的情況下進行估計，這樣的假設並未考慮到時間序列資料可能具有統計時變程序的特性。統計時變程序時間序列是指當時間序列若受到外生衝擊的影響，此影響會一直存留於時間序列中，不會因為時間的經過而消失。根據 1974 年葛蘭傑(C. W. J. Granger)與紐伯(P. Newbold)所提出的假性迴歸[18]指出，若迴歸模型中存在統計時變程序的變數，使用傳統的迴歸分析將產生假性迴歸(Spurious Regression)的問題。所謂假性迴歸是指當使用統計時變程序的變數進行傳統的迴歸分析時，會出現 t 統計量及 F 統計量很顯著，也就是以 R^2 來衡量有很高的契合度，但德-華氏檢定(Durbin-Watson Test)值卻非常低的現象，這表示殘差存在有自相關性。因此在處理時間序列資料時，必須先判定資料特性，而後才能決定資料分析及處理的方法。對於資料特性的判別，一般而言可以透過單位根檢定(Unit Root Test)來判定資料的特性及其整合階次(Order)。

2.3.4 單位根檢定

單位根檢定最先是由迪契(D. A. Dicky)與富樂(W. A. Fuller)於 1979 年所提出的[16]，其方法是透過蒙地卡羅法求出單位根檢定模型的機率分配。藉由他們所假設的模型並配合此分配表，令使用者可以迅速的檢測出時間序列中是否存在單位根。早期所提出的迪-富氏檢定(Dicky-Fuller Test, DF Test)主要在於檢定時間序列變數的殘差項是否為白噪音(White Noise)，但在原始的檢定模型中，卻忽略了資料中可能存在的序列相關(Serial Correlation)的問題。為了解決序列相關的問題，1981 年迪契及富樂進一步的提出了擴大迪-富氏檢定(Augmented Dicky-Fuller Test, ADF Test)。擴大迪-富氏檢定改善了迪-富氏檢定的缺點，其主要增加了更多前期的資料做為解釋變數，幫助減少誤差項的序列相關問題，使其加接近白噪音。

單位根檢定中一般常用的方法為擴大迪-富氏檢定，在擴大迪-富氏檢定中共包含了以下三種模式：

無漂浮項與無趨勢項之隨機漫步模型

$$\Delta Y_t = \beta Y_{t-1} + \sum_{i=1}^N \alpha_i \Delta Y_{t-i} + \varepsilon_t$$

有漂浮項但無趨勢項之隨機漫步模型

$$\Delta Y_t = \alpha_0 + \beta Y_{t-1} + \sum_{i=1}^N \alpha_i \Delta Y_{t-i} + \varepsilon_t$$

有漂浮項及趨勢項之隨機漫步模型

$$\Delta Y_t = \alpha_0 + \gamma t + \beta Y_{t-1} + \sum_{i=1}^N \alpha_i \Delta Y_{t-i} + \varepsilon_t$$

其中 Y_t 為要檢定是否具有單位根的變數， α_0 為截距項， t 為時間趨勢， β 為估計系數， N 為使殘差項近似白噪音的最適落遲期數(Lag)。在模型檢定中，其假設檢定如下：

H_0 : $\beta = 0$ (有單位根，資料為統計時變程序)

H_1 : $\beta \neq 0$ (無單位根，資料為統計非時變程序)

在虛無假設中假設 $\beta = 0$ ，亦即若檢定結果無法拒絕 H_0 ，則表示模型中存在單位根，資料具統計時變程序特性。反之若檢定結果為拒絕虛無假設，則表示模型中不存在單位根，資料為統計非時變程序。



第三章 台灣區花卉資料倉儲資料轉換服務實作

自從 2002 年世界第一座花卉資料倉儲系統-台北資料倉儲系統建立成功後，至 2004 年兩年間已完成了彰化、台中和台南共四家資料倉儲系統，使花卉產業向前邁進一大步，促進了花卉交易資訊的分享。

由於目前台北、彰化、台中、台南、高雄五家花市的資料倉儲系統是各自獨立的個體，各家花市資料倉儲的使用者不能互相擷取資訊，但對花農來說，整個台灣區的花卉市場都是他們能供應的範圍，因此整合五家花卉供應鏈的資訊是非常重要的工作，能為整個台灣地區花卉的生產者與消費者創造更多利機，於是在各個花卉資料倉儲的運作達成熟之際，台灣區花卉資料倉儲系統也應運而生。

資料倉儲之父殷默提出資料倉儲有四大特性，以主題為導向 (Subject-Oriented)、時間變化性 (Time Variation)、整合性 (Integration)、非揮發性 (Non Volatilization)，本系統架構特別著重在時間變化性與整合性。在時間變化性上，建立有日期、星期、農曆與節慶四個維度；以整合性來說，整合了台北、彰化、台中、台南與高雄五個資料超市。本章將說明台灣區花卉資料倉儲的系統架構與資料轉換服務的設計。

本章共分為四小節，第 3.1 節說明系統與資料倉儲的架構。第 3.2 節說明如何整合四家資料超市的檔案。第 3.3 節介紹資料轉換服務的設計。第 3.4 節介紹總資料倉儲遠端維修的方式。

3.1 系統架構與資料倉儲架構

本節將介紹花卉總資料倉儲建立的系統架構，與資料倉儲建構模式。第 3.1.1 節介紹系統建置在主機的方式；第 3.1.2 節介紹資料倉儲建置的架構模式。

3.1.1 系統架構

總資料倉儲的系統建立，硬體部分是將資料處理與檔案發佈分別於兩台 IBM e-server 處理，e-server 1-2 主要負責資料的傳輸與轉換，e-server 1-1 主要負責檔案發佈與權限控管。加上各家資料超市整個硬體設施如圖 3.1 所示，左邊三個方框由上而下分別代表存放台北資料超市之主機 e-server 2-1、存放彰化資料超市之主機 e-server 2-2、存放台中與台南資料超市之主機 e-server 2-3，而中間名為 e-server 1-2，主要存放總資料倉儲的資料庫檔案 *.mdf，右邊名為 e-server 1-1 的主機，則負責將總資料倉儲的檔案發佈於網路上。表 3.1 將各主機之命名及其他相關屬性做詳細的整理。

如圖 3.1 所示，各個主機以 Hub 方式接到防火牆(如圖 3.1 實線部分)，其中 e-server 1-2 透過 SQL Server 建立遠端伺服器連線，SQL Server 可透過防火牆內的網路，直接讀取 Domain Name 來達到對其他伺服器資料的存取，如圖 3.2 所示，SQL Server 的群組底下共建立四個伺服器連線，FI1 為本機伺服器，存放總資料倉儲，FLOWERDB2 為彰化資料超市之伺服器，FLOWERDB3 為台中與台南資料超市共用之伺服器，SCFS01 為台北資料超市之伺服器。e-server 1-2 每日自動化進行資料轉換服務，透過遠端伺服器連線(如圖 3.1 虛線部分)，可自動轉換四家資料超市的資料於總資料倉儲中，此部分連線透過資料轉換服務進行，如圖 3.3 為台北拍賣 DTS1-1 之連線畫面與設定、圖 3.4 為台中拍賣 DTS1-2 之

連線畫面與設定、圖 3.5 為彰化拍賣 DTS1-3 之連線畫面與設定、圖 3.6 為台南拍賣 DTS1-4 之連線畫面與設定。然後 e-server1-2 透過 cognos 套裝軟體建立報表檔案*.pyi 與超方體檔案*.mdc，將超方體檔案透過防火牆下之網路連線，複製存放於 e-server 1-1，此部分是由 e-server1-2 透過網路磁碟機的設定完成連線，如圖 3.7 所示名為 F 槽的即為網路磁碟機，Fdatabase 為 e-server 1-1 的 Domain Name；在自動化更新轉換程序當中，將此語法設定在自動化 Macro (DTS_5)，圖 3.8 為檔案複製到網路磁碟機之語法畫面；e-server 1-1 存放權限控管之檔案資料，以及 e-server 1-2 所產生之超方體檔案，透過 cognos 的套裝軟體 PPES (PowerPlay Enterprise Server)，將超方體發佈到網路上，使用者可透過網路進行線上分析處理。

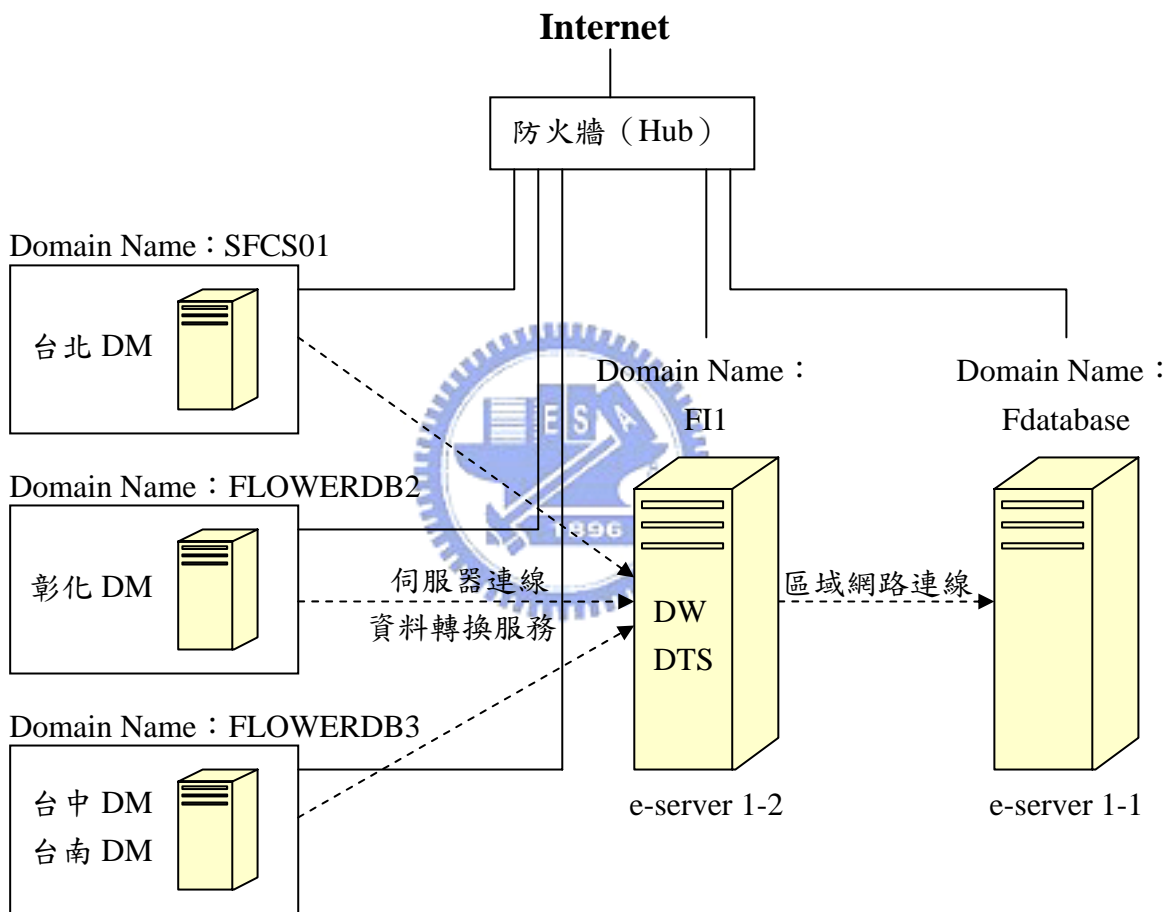


圖 3.1 總倉儲實體系統架構圖

表 3.1 各主機項目命名表

硬體代號	主機型號	IP 位置	Domain Name	資料庫名稱	功能
e-server 1-1	IBM e-server X series 236	140.113.58.186	Fdatabase		提供線上存取服務
e-server 1-2	IBM e-server X series 236	140.113.58.185	FI1	Flower2	總資料倉儲
e-server 2-1	IBM e-server X series 235	140.113.58.162	SFCS01	TPCDB	台北資料超市
e-server 2-2	IBM e-server X series 235	140.113.58.180	FLOWERDB2	TenWay1	彰化資料超市
e-server 2-3	IBM e-server X series 235	140.113.58.181	FLOWERDB3	flower3	台中資料超市
				flower4	台南資料超市

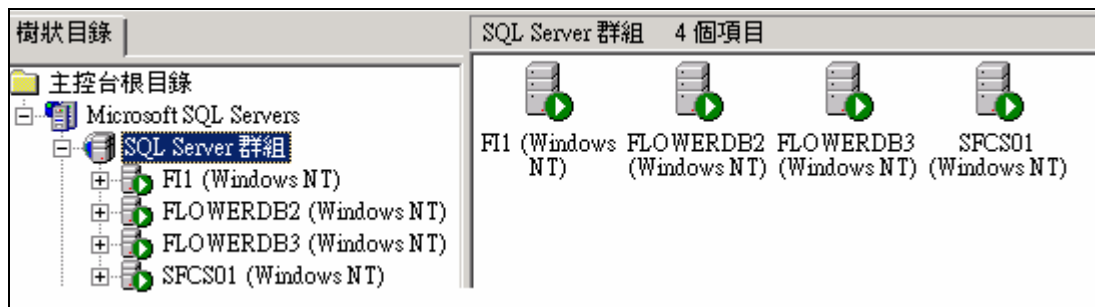


圖 3.2 SQL Server 連線群組(a)

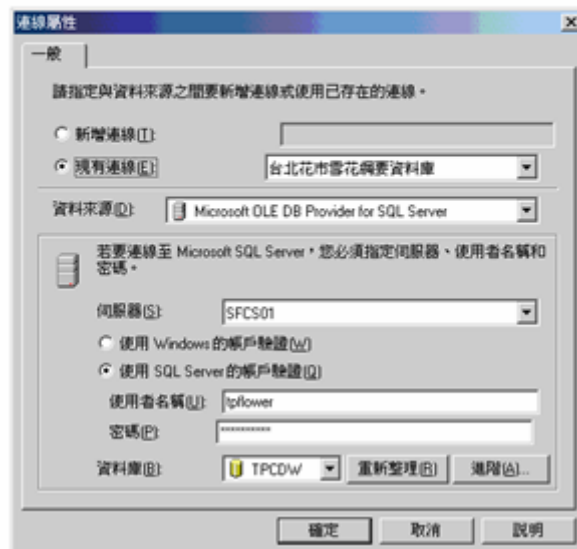


圖 3.3 SQL Server 連線群組(b)



台中花市雪花網... 總倉儲雪花網要...



圖 3.4 SQL Server 連線群組(c)



彰化花市雪花網... 總倉儲雪花網要...

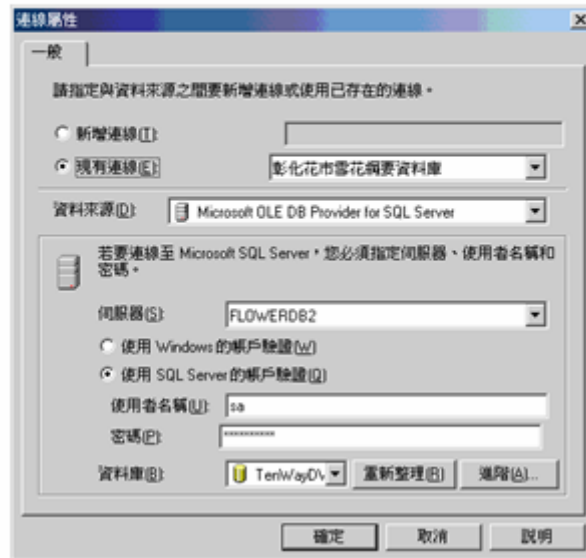


圖 3.5 SQL Server 連線群組(d)

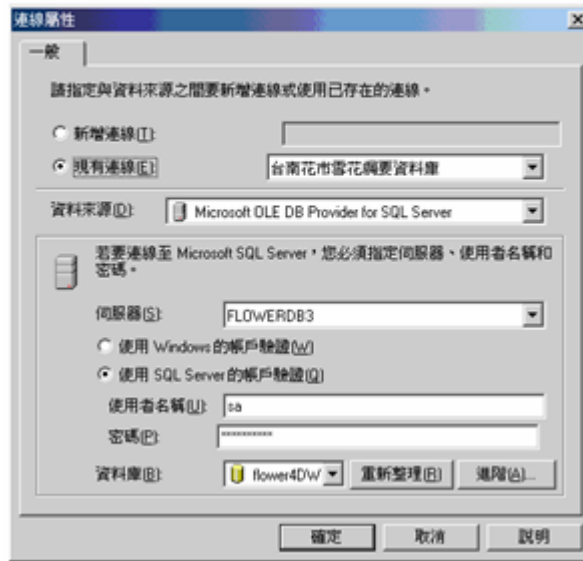


圖 3.6 SQL Server 連線群組(e)



圖 3.7 網路磁碟機畫面

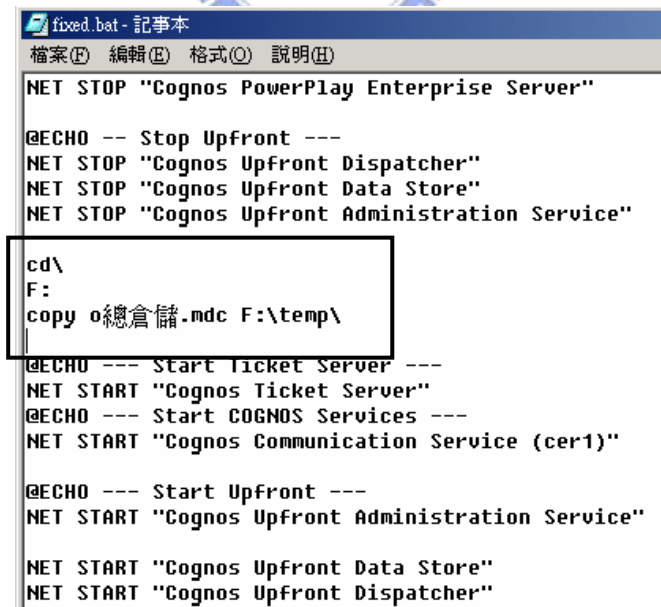


圖 3.8 檔案複製到網路磁碟機語法畫面

從圖 3.1 中，由四個資料超市至第二台主機的所有資料轉換流程，皆透過每日進行的自動化資料轉換服務完成，在時間進行上可參照圖 3.9 資料轉換時刻圖，每日於四家資料超市個別執行完資料轉換服務後，才於 e-server 1-2 進行總資料倉儲的自動化資料轉換服

務。表 3.2 將兩台負責總資料倉儲的 IBM 主機，做存放檔案以及分別執行之服務的整理。

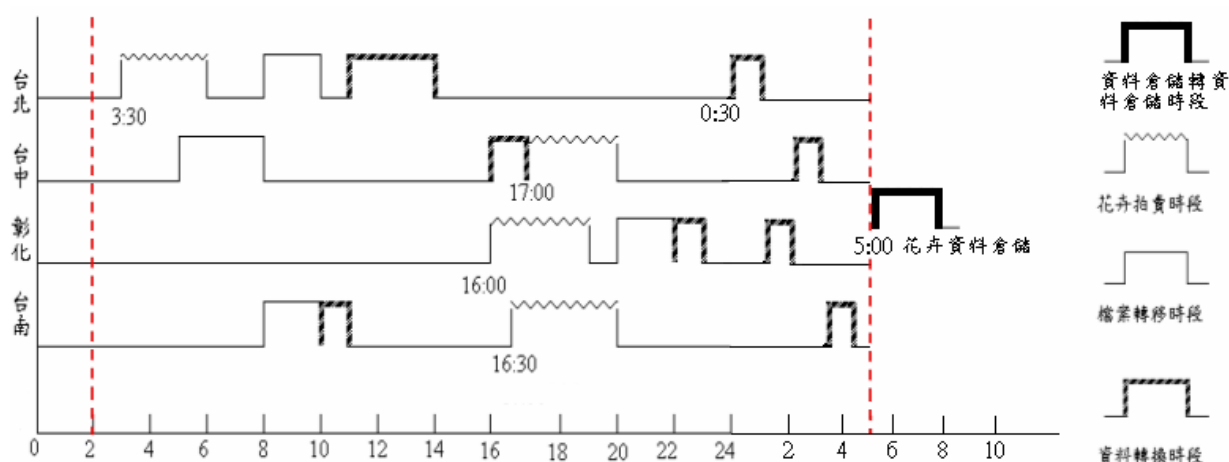


圖 3.9 資料轉換時刻圖

表 3.2 總倉儲主機負責服務整理表

主機名稱	第一台 IBM e-server	第二台 IBM e-server
存放檔案	<ul style="list-style-type: none"> · 資料庫檔案*.mdf · DTS 檔案*.dts · 報表檔案*.pyi · 超方體檔案*.mdc 	<ul style="list-style-type: none"> · 權限控管檔案*.lae · 超方體檔案*.mdc
執行服務	總資料倉儲之自動化轉換服務 <ul style="list-style-type: none"> · 資料轉換服務 · 建立超方體 · 將超方體複製於第二台主機 	提供線上分析處理服務 <ul style="list-style-type: none"> · 權限控管 · 檔案發佈於網路上

3.1.2 資料倉儲架構

台灣區花卉資料倉儲，採取競簇的資料倉儲架構[23]，由台北、台中、彰化、台南和高雄五家資料超市由下而上建立而成，基本架構如圖 3.10 台灣區花卉資料倉儲競簇架構圖所示，其中高雄花卉資料超市尚未完工，因此本系統架構將不考慮高雄花卉資料超市。

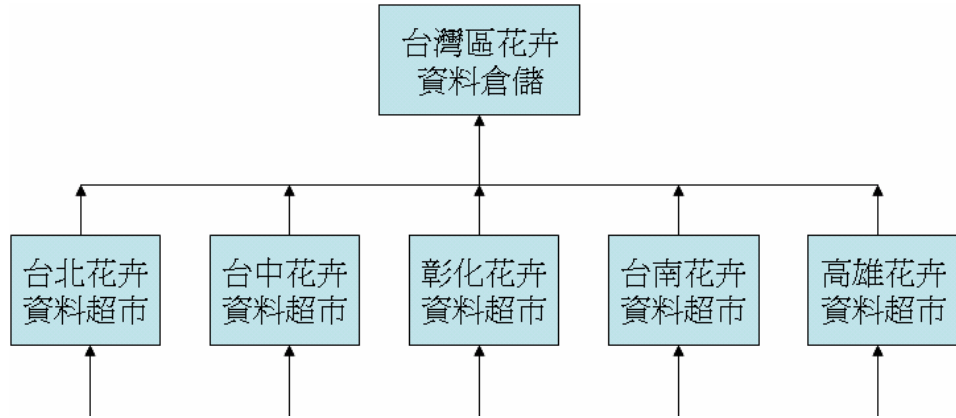


圖 3.10 台灣區花卉資料倉儲競筭架構圖

資料倉儲架構如圖 3.11 花卉批發資訊分享熱線架構圖所示，本章將著重在右邊粗虛線框內部分的資料轉換服務，除去尚未完工的高雄花市，分別從各家花市進行資料轉換服務，彙整為總資料倉儲。

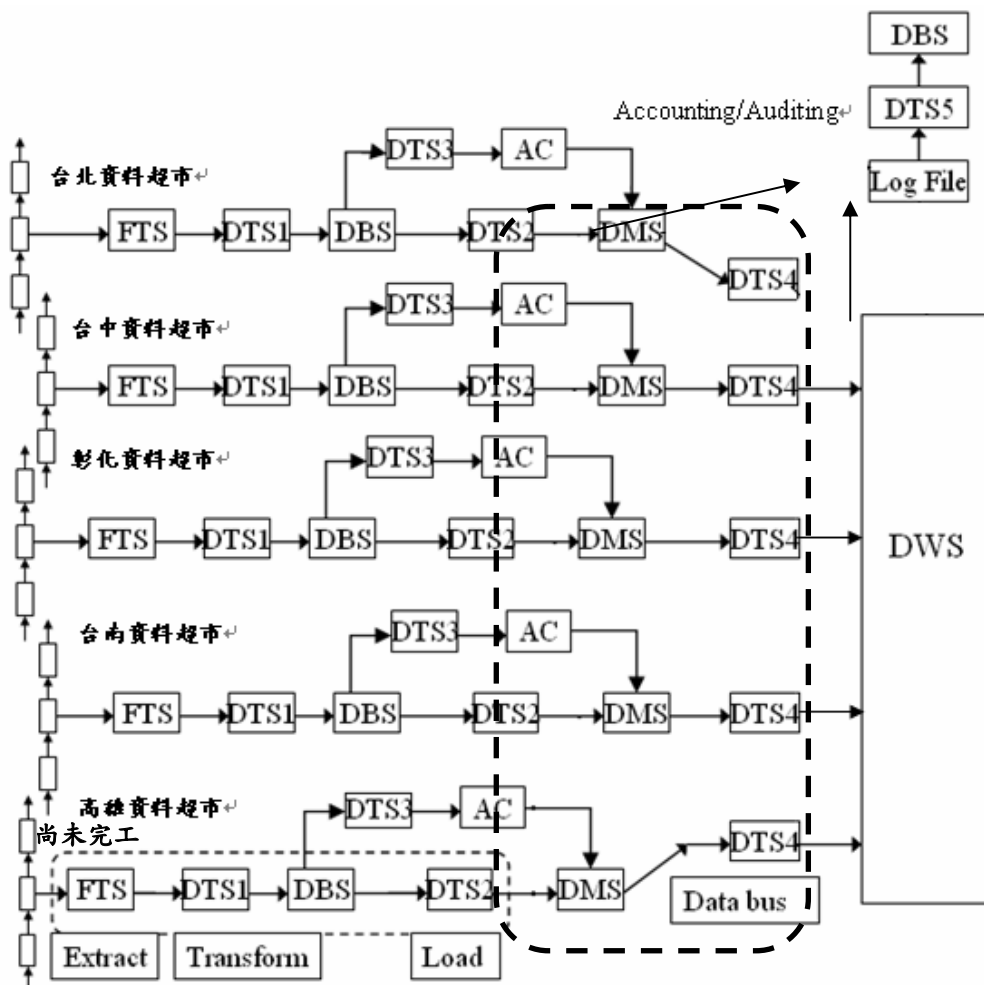


圖 3.11 花卉批發資訊分享熱線架構圖

3.2 總資料倉儲整合工作

總資料倉儲的設計重點，在於能夠同時顯示台北、台中、彰化和台南四家資料超市的資訊，但礙於四家超市為不同設計者所設計，且各家資料超市的資料內容有些許差異，因此欄位名稱以及資料表內容都必須先行整合，才能使總資料倉儲的資料轉換工作繼續進行。本節共分兩小節，第 3.2.1 節介紹總資料倉儲的設計架構；第 3.2.2 節說明資料整合的結果。

3.2.1 總資料倉儲架構

總資料倉儲為了避免私密性的資料外洩，以及考慮資料傳輸速度和使用者訴求，必須減少不必要的維度，以及新增維度。和之前各家資料超市相比較，總資料倉儲多了特殊節慶以及市場別維度，而在供應單位中減少供應人層級，以避免個人資料外洩，表 3.3 為總倉儲與各家資料超市維度架構之比較。

表 3.3 總倉儲與各家資料超市維度比較表

維度	總倉儲	各家資料超市
供應單位	無供應人層級	有供應人層級
供應地區	有	有
日期	有	有
星期	有	有
農曆	有	有
節慶	有	無
市場別	有	無
拍賣線	有	有
花卉種類	有	有
花卉等級	有	有
容器	有	有
承銷地區	有	有
	共 12 個維度	共 10 個維度

3.2.2 資料整合工作

進行資料轉換服務的前置工作，就是必須將台北、台中、彰化、台南四家資料超市的資料表做整合，以制訂共同標準。維度當中除了日期、星期、農曆、花卉等級四個維度的資料表欄位內容不需整合，以及新增的節慶、市場別以外，其他維度皆採取聯集方式進行整合。

聯集方式如圖 3.12 四家花市聯集示意圖所示，以 A 代表台北、B 代表台中、C 代表彰化、D 代表台南，取共同的聯集，以集合符號表示為：

$$A \cup B \cup C \cup D = A + B + C + D - A \cap B - B \cap C - C \cap D - D \cap A - B \cap D - C \cap A \\ + A \cap B \cap C + B \cap C \cap D + C \cap D \cap A + A \cap B \cap D - A \cap B \cap C \cap D$$

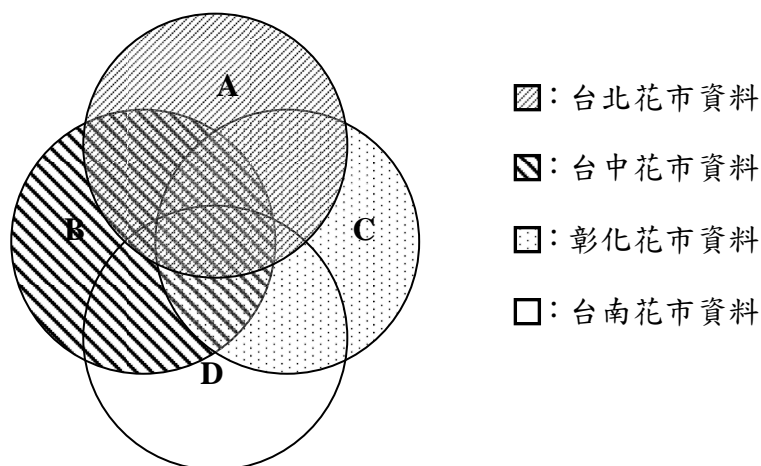


圖 3.12 四家花市聯集示意圖

表 3.4 整合對照表顯示各家資料超市的資料表整合後的情形，欄位一表示為各個維度，在最末一列另外加入 Fact Table，並非維度之意；欄位二內容為各家資料超市內的原始資料表，以統一名稱代替，不另贅述出自哪家花市；欄位三為總資料倉儲內新建的資料表，是經過聯集整合後的資料；欄位四為新建在每家資料超市的轉換資料表，以台北資料超市為例，新增了台北供應類別對應表，做為總資料倉儲內供應單位與台北資料超市內供應類別的對照表，以方便進行資料轉換的動作。

表 3.4 整合對照表

維度	各花市原始資料表	總倉儲整合資料表	各花市轉換資料表
供應單位	供應團體資料表、 供應類別	供應單位	(台北、台中、彰化、台南) 供應類別對應表
供應地區	供應行政區資料表、 供應縣市資料表、 供應人資料表	供應縣市、 供應鄉鎮	(台北、台中、彰化、台南) 供應鄉鎮對應表
日期			
星期	星期表	星期	
農曆	農曆對照表、 農曆年、農曆月	農曆對照表、 農曆年、農曆月	
節慶		節慶範圍資料表、 節慶類別資料表	
市場別		市場別	
拍賣線	拍賣線資料表	拍賣線	
花卉種類	花卉資料表、 花卉品名資料表、 花卉種類	花卉種類、 花卉品類、 花卉品名	
花卉等級	花卉等級略表、 花卉等級資料表	花卉等級、 花卉等級二碼	
容器	容器資料表	容器	
承銷地區	承銷人資料表、 承銷區域資料表	承銷地區	(台北、台中、彰化、台南) 承銷地區對應表
Fact Table	拍賣資料表	總倉儲交易資料表	

3.3 資料轉換服務

台灣區花卉資料倉儲的資料轉換模組，將交易程序分為拍賣、訂貨、殘貨與議價四種分類，分別從台北、台中、彰化和台南四個雪花網要資料庫，轉換過來總倉儲的雪花網要資料庫，本節說明此模組的資料轉換步驟。

完整的資料轉換模組如圖 3.13 自動化更新轉換程序所示，由五個資料轉換程式所構成，其中拍賣資料轉換工作、訂貨資料轉換工作、殘貨資料轉換工作、議價資料轉換工作為自動化更新轉換程序底下的四個封包，再加上自動化 Macro 的動作，就組成自動化更新轉換程序，透過資料轉換服務，可將四家資料庫的交易資料三階正規化，並自動轉入資料倉儲。本節將自動化更新轉換程序編號為 DTS_0，拍賣資料轉換工作編號為 DTS_1，訂貨資料轉換工作編號為 DTS_2，殘貨資料轉換工作編號為 DTS_3，議價資料轉換工作編號為 DTS_4，自動化 Macro 編號為 DTS_5，如表 3.5 總倉儲資料轉換程式所示，此外，表 3.4 還列出此四個封包之下的其他子資料轉換服務程式，在此給予 DTS 子編號以清楚識別。

本節分為三小節，第 3.3.1 節介紹實體資料如何轉換為多維度模型；第 3.3.2 節說明如何進行資料轉換服務；第 3.3.3 節介紹如何將資料倉儲的建置自動化。



圖 3.13 自動化更新轉換程序

表 3.5 總倉儲資料轉換程式

DTS 編號	拍賣資料 轉換工作 DTS_1	訂貨資料 轉換工作 DTS_2	殘貨資料 轉換工作 DTS_3	議價資料 轉換工作 DTS_4	自動化 Macro DTS_5
清空總倉 儲交易資 料表	DTS_1-0				
台北	DTS_1-1	DTS_2-1	DTS_3-1	DTS_4-1	
台中	DTS_1-2	DTS_2-2	DTS_3-2		
彰化	DTS_1-3	DTS_2-3	DTS_3-3		
台南	DTS_1-4	DTS_2-4	DTS_3-4		
高雄	尚未進行	尚未進行	尚未進行		

3.3.1 實體關係模式至多維度模型轉換

總倉儲的實體關係模式至多維度模型轉換，見圖 3.14 實體關係至多維度模型架構圖呈現出本節的模型架構。由家花市已存在的拍賣資料表，經由內部轉換，將資料轉入總倉儲的總倉儲交易資料表，總倉儲交易資料表再加上建好的整合維度資料表，便組成總倉儲雪花網要關聯圖，有了這個架構想法，經由資料轉換工作的實行，便可完成總倉儲完整的資料轉換。

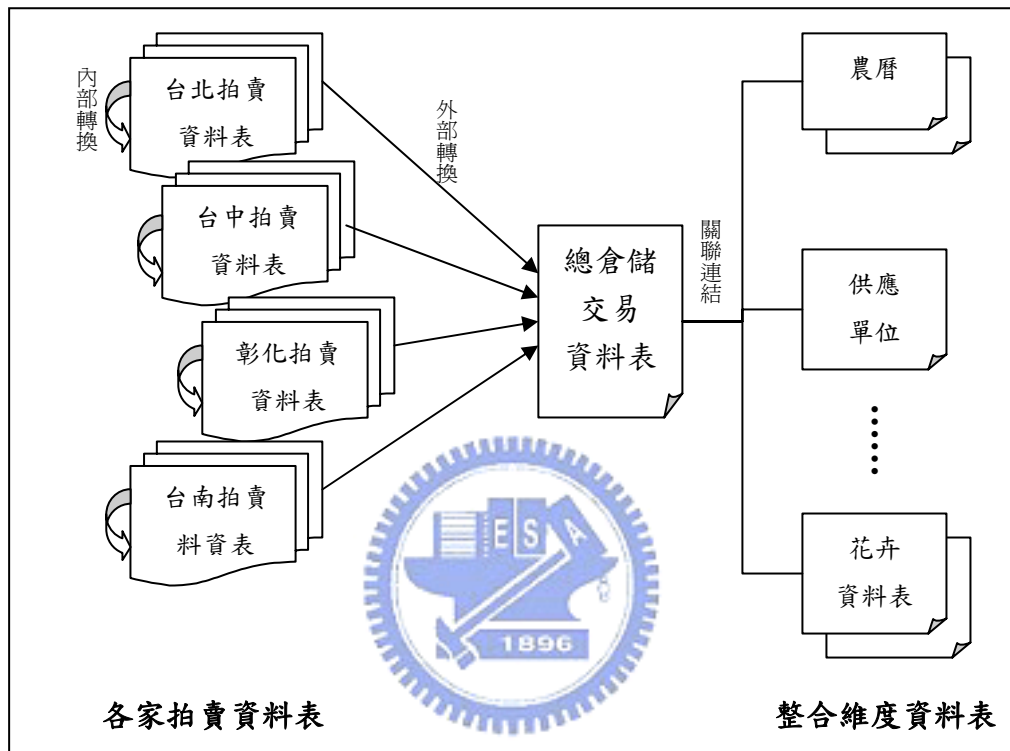


圖 3.14 實體關係至多維度模型架構圖

以彰化資料倉儲為例，在彰化資料倉儲的資料庫內，新建三個整合用的資料表，分別為彰化承銷地區對應表、彰化供應鄉鎮對應表、彰化供應類別對應表，這三個資料表用途為與供應人資料表、承銷人資料表做資料整合，如圖 3.15 彰化資料超市內部轉換關係圖右上方虛線框住部分，從拍賣資料表的供應人代號欄位 TEN_AUC_SUPPLYNUM，對應到供應人資料表的供應人代號 TEN_SUP_NUM，供應人資料表的鄉鎮代碼 TEN_SUP_LOCATIONNUM，對應到彰化供應鄉鎮對應表的彰化供應鄉鎮代碼，供應人資料表的供應團體代碼 TEN_SUP_GROUP，對應到供應團體資料表的供應團體代碼 TEN_SUPGROUP_GROUPNUM，供應團體資料表的供應團體代碼 TEN_SUPGROUP_GROUPNUM，再對應至彰化供應類別對應表的彰化供應類別代碼；圖 3.15 左下方虛線框住部分，從拍賣資料表的承銷人代號欄位 TEN_AUC_BUYNUM，對應到承銷人資料表的承銷人代碼 TEN_BUY_BUYNUM，再由承銷人資料表的承銷地區代碼 TEN_BUY_AREANUM，對應到彰化承銷地區對應表的承銷地區代碼。

彰化資料超市的拍賣資料表在內部轉換資料後，經過資料轉換工作，資料加入到總倉儲交易資料表中，如圖 3.16 拍賣資料表與總倉儲交易資料表實體關聯圖所示，黑線連結部

分表示欄位資料轉換，便完成了實體連結，其他台北、台中、台南模式同彰化模式，經由四家花市資料表的轉換後，便轉為雪花綱要架構圖，如圖 3.17 總資料倉儲雪花綱要關聯圖。

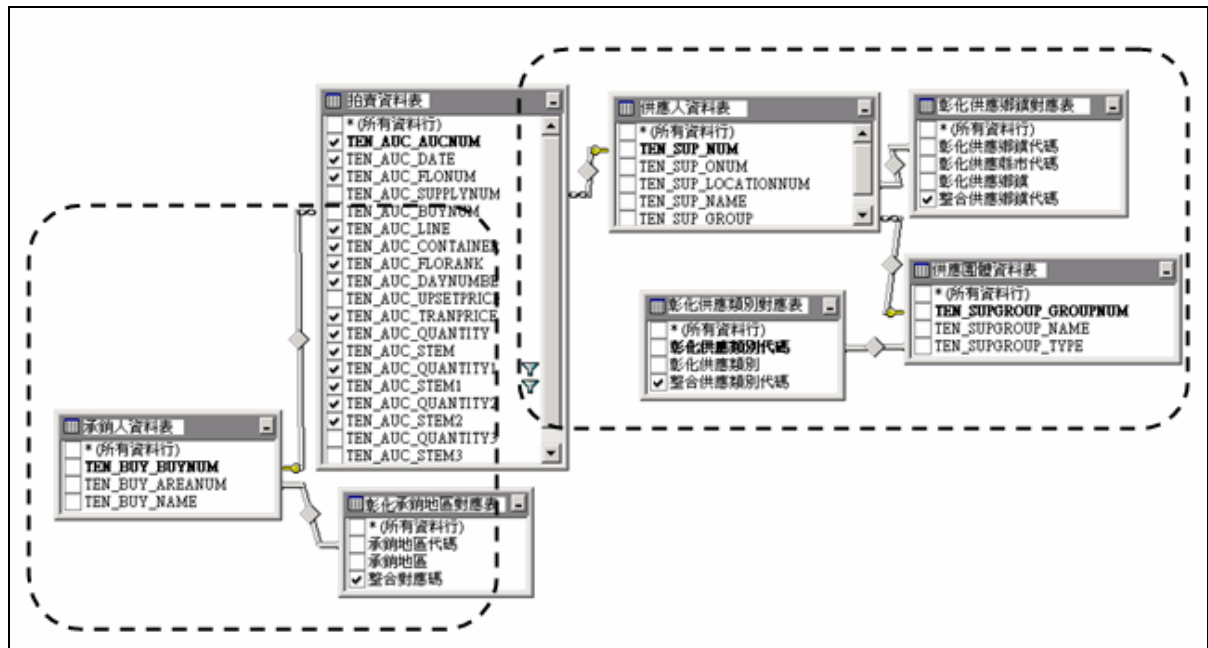


圖 3.15 彰化資料超市內部轉換關係圖

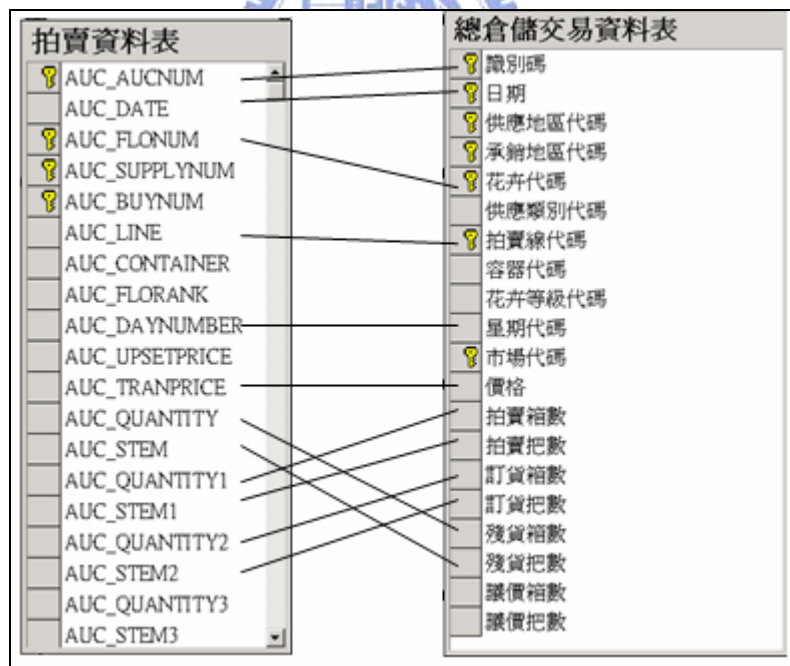


圖 3.16 拍賣資料表與總倉儲交易資料表實體關聯圖

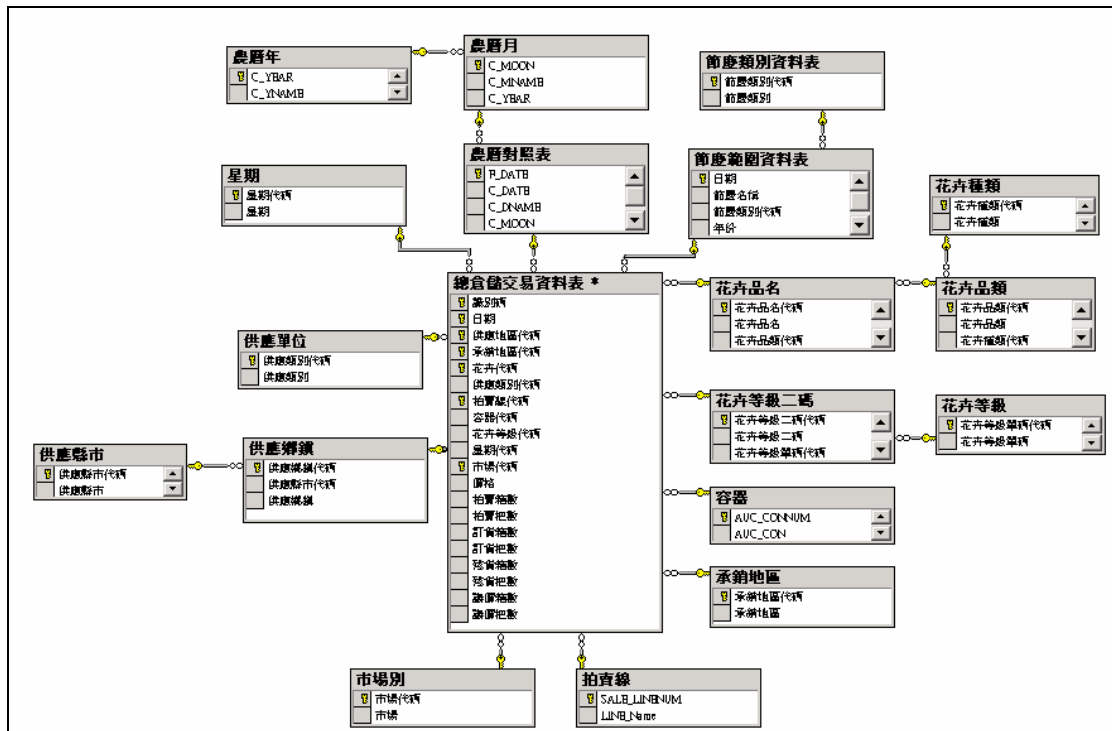


圖 3.17 總資料倉儲雪花網要關聯圖

3.3.2 資料轉換服務

完整的自動化更新轉換程序，是按照拍賣、訂貨、殘貨、議價順序進行資料轉換工作，當各家花市的四個原始交易檔案（註一），經由個別的自動化資料轉換服務轉入資料超市後，首先進行拍賣資料轉換工作，如圖 3.18 拍賣資料轉換工作所示，總倉儲伺服器連線至總倉儲的雪花網要資料庫，需先成功清空交易資料表，再依台北、台中、彰化、台南的順序進行拍賣資料轉換工作，以台北拍賣為例子，如圖 3.19 台北拍賣資料轉換工作，先連線至台北資料超市的雪花網要資料庫，利用查詢方式將拍賣資料表當中的拍賣資料，轉換至總倉儲雪花資料庫中的總倉儲交易資料表，此部分 SQL 查詢語法建立如圖 3.20 台北拍賣轉換資料工作來源查詢所示，另外，圖 3.21 台北拍賣轉換資料工作屬性，可看出來源與目的地之轉換，此部分的轉換語法如圖 3.22 台北拍賣轉換資料工作轉換語法所示，以 VB script 撰寫而成。

在進行完台北拍賣之後，便進行台中拍賣的資料轉換，接著彰化拍賣、台南拍賣，便完成了拍賣資料轉換工作。其他三個資料轉換動作同台北拍賣一樣，只是來自不同的連線資料庫，表 3.6 拍賣資料轉換工作步驟表將拍賣資料轉換工作步驟詳細描述。

註一：四個資料表為供應人資料檔、承銷人資料檔、花卉資料檔、拍賣資料檔，請參閱『花卉批發資訊分享熱線的設計與實作-以台中花市為例』柯珮婕著，內文第 32 頁。

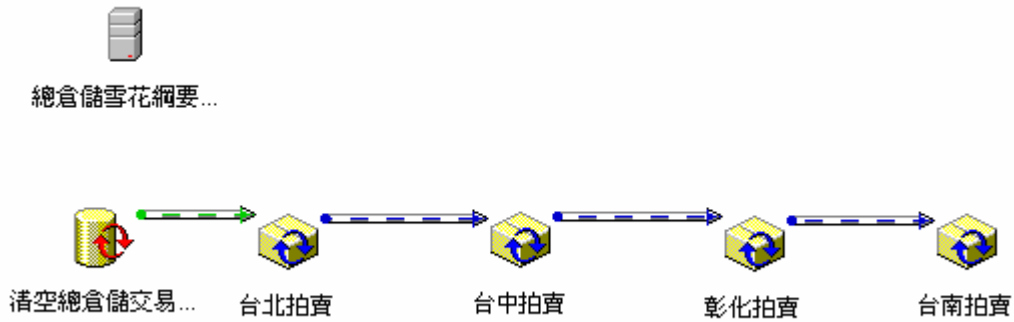


圖 3.18 拍賣資料轉換工作



圖 3.19 台北拍賣資料轉換工作

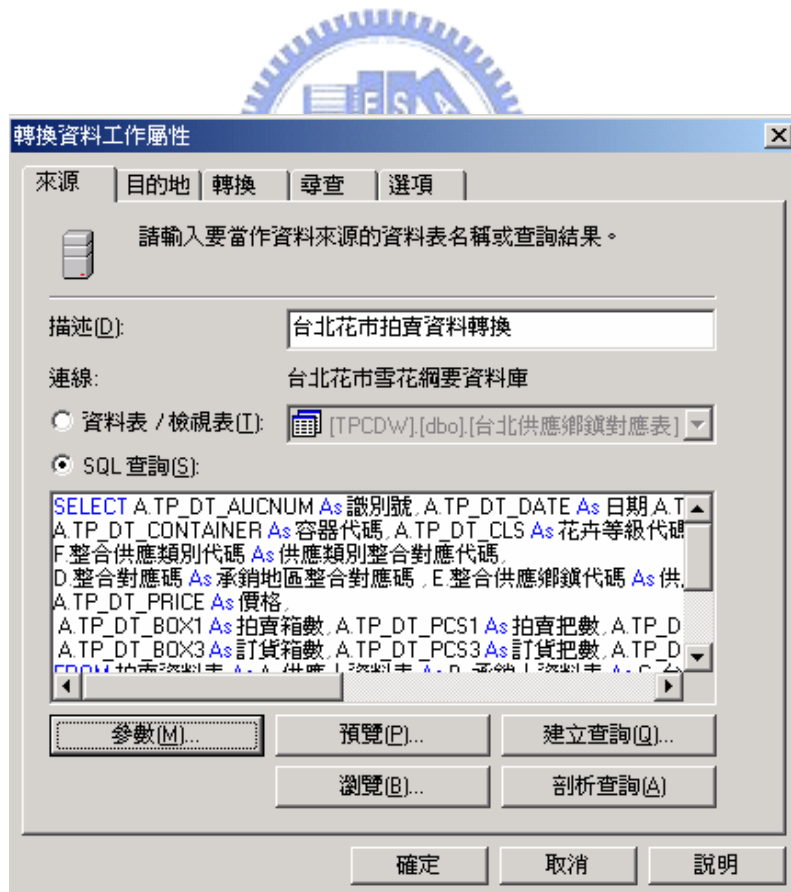


圖 3.20 台北拍賣轉換資料工作來源查詢



圖 3.21 台北拍賣轉換資料工作屬性



圖 3.22 台北拍賣轉換資料工作轉換語法

表 3.6 拍賣資料轉換工作步驟表

步驟	工作名稱	資料轉換動作
DTS_1-0	清空總倉儲交易資料表	清空交易資料表
DTS_1-1	台北拍賣	1. 連線台北花市的雪花網要資料庫與總倉儲雪花網要資料庫 2. 將台北花市拍賣資料表的拍賣資料轉入總倉儲交易資料表
DTS_1-2	台中拍賣	1. 連線台中花市的雪花網要資料庫與總倉儲雪花網要資料庫 2. 將台中花市拍賣資料表的拍賣資料轉入總倉儲交易資料表
DTS_1-3	彰化拍賣	1. 連線彰化花市的雪花網要資料庫與總倉儲雪花網要資料庫 2. 將彰化花市拍賣資料表的拍賣資料轉入總倉儲交易資料表
DTS_1-4	台南拍賣	1. 連線台南花市的雪花網要資料庫與總倉儲雪花網要資料庫 2. 將台南花市拍賣資料表的拍賣資料轉入總倉儲交易資料表

在完成拍賣資料轉換工作後，緊接著進行訂貨資料轉換工作，由於一開始進行拍賣資料轉換工作時，便已經先清空總倉儲交易資料表，所以在訂貨資料轉換工作便不需要做清空的動作，如圖 3.23 訂貨資料轉換工作所示，按照台北訂貨、台中訂貨、彰化訂貨、台南訂貨的順序，便完成了訂貨資料轉換工作。

各家訂貨的詳細程序如圖 3.24 台北訂貨資料轉換工作所示，先連線至台北資料超市的雪花網要資料庫，將拍賣資料表當中的訂貨資料，轉換至總倉儲雪花資料庫中的總倉儲交易資料表，同樣地，完成之後再連線至台中資料超市的雪花網要資料庫，將拍賣資料表當中的訂貨資料，轉換至總倉儲雪花資料庫中的總倉儲交易資料表，接著彰化、台南也都是是一致的動作，表 3.7 訂貨資料轉換工作四步驟表將訂貨資料轉換工作步驟詳細描述。

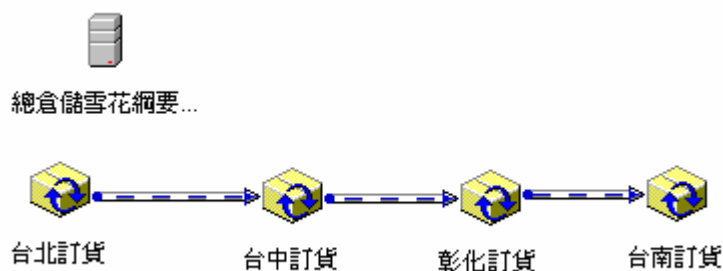


圖 3.23 訂貨資料轉換工作



圖 3.24 台北訂貨資料轉換工作

表 3.7 訂貨資料轉換工作步驟表

步驟	工作名稱	資料轉換動作
DTS_2-1	台北訂貨	<ol style="list-style-type: none"> 1. 連線台北花市的雪花網要資料庫與總倉儲雪花網要資料庫 2. 將台北花市拍賣資料表的訂貨資料轉入總倉儲交易資料表
DTS_2-2	台中訂貨	<ol style="list-style-type: none"> 1. 連線台中花市的雪花網要資料庫與總倉儲雪花網要資料庫 2. 將台中花市拍賣資料表的訂貨資料轉入總倉儲交易資料表
DTS_2-3	彰化訂貨	<ol style="list-style-type: none"> 1. 連線彰化花市的雪花網要資料庫與總倉儲雪花網要資料庫 2. 將彰化花市拍賣資料表的訂貨資料轉入總倉儲交易資料表
DTS_2-4	台南訂貨	<ol style="list-style-type: none"> 1. 連線台南花市的雪花網要資料庫與總倉儲雪花網要資料庫 2. 將台南花市拍賣資料表的訂貨資料轉入總倉儲交易資料表

在完成訂貨資料轉換工作後，緊接著進行殘貨資料轉換工作，如圖 3.25 殘貨資料轉換工作所示，同於 DTS_1 與 DTS_2 的模式，按照台北殘貨、台中殘貨、彰化殘貨、台南殘貨的順序，便完成了殘貨資料轉換工作。

各家殘貨的詳細程序如圖 3.26 台北殘貨資料轉換工作所示，先連線至台北資料超市的雪花網要資料庫，將拍賣資料表當中的殘貨資料，轉換至總倉儲雪花資料庫中的總倉儲交易資料表，同樣地，完成之後再連線至台中資料超市的雪花網要資料庫，將拍賣資料表當中的殘貨資料，轉換至總倉儲雪花資料庫中的總倉儲交易資料表，接著彰化、台南也都是是一致的動作，表 3.8 殘貨資料轉換工作四步驟表將殘貨資料轉換工作步驟詳細描述。

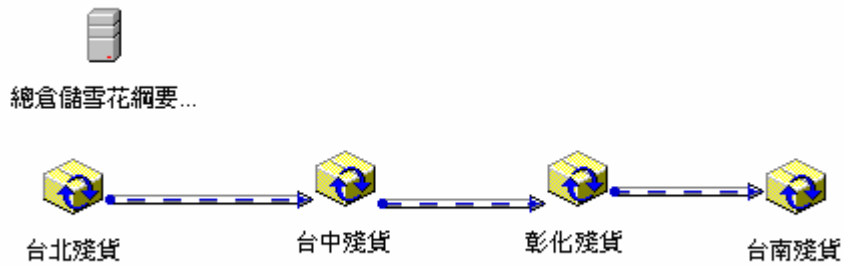


圖 3.25 殘貨資料轉換工作



圖 3.26 台北殘貨資料轉換工作

表 3.8 殘貨資料轉換工作步驟表

步驟	工作名稱	資料轉換動作
DTS_3-1	台北殘貨	<ol style="list-style-type: none"> 1. 連線台北花市的雪花綱要資料庫與總倉儲雪花綱要資料庫 2. 將台北花市拍賣資料表的殘貨資料轉入總倉儲交易資料表
DTS_3-2	台中殘貨	<ol style="list-style-type: none"> 1. 連線台中花市的雪花綱要資料庫與總倉儲雪花綱要資料庫 2. 將台中花市拍賣資料表的殘貨資料轉入總倉儲交易資料表
DTS_3-3	彰化殘貨	<ol style="list-style-type: none"> 1. 連線彰化花市的雪花綱要資料庫與總倉儲雪花綱要資料庫 2. 將彰化花市拍賣資料表的殘貨資料轉入總倉儲交易資料表
DTS_3-4	台南殘貨	<ol style="list-style-type: none"> 1. 連線台南花市的雪花綱要資料庫與總倉儲雪花綱要資料庫 2. 將台南花市拍賣資料表的殘貨資料轉入總倉儲交易資料表

在完成殘貨資料轉換工作後，接著進行議價資料轉換工作，如圖 3.27 議價資料轉換工作所示，由於議價交易行為只有在台北花市進行，因此資料轉換工作也只有台北議價一個封包，如圖 3.28 台北議價資料轉換工作所示，連線至台北花市雪花綱要資料庫後，將拍賣資料表當中的議價資料，轉換至總倉儲雪花資料庫中的總倉儲交易資料表後，便完成了議價資料轉換工作，表 3.9 殘貨資料轉換工作步驟表將殘貨資料轉換工作步驟詳細描述。



圖 3.27 議價資料轉換工作



圖 3.28 台北議價資料轉換工作

表 3.9 議價資料轉換工作步驟表

步驟	工作名稱	資料轉換動作
DTS_4-1	台北議價	<ol style="list-style-type: none"> 1. 連線台北花市的雪花綱要資料庫與總倉儲雪花綱要資料庫 2. 將台北花市拍賣資料表的議價資料轉入總倉儲交易資料表

在 DTS_5 中，自動化 Macro 程式利用 DOS 程式的撰寫，使 cognos 套裝軟體能自動開啟。將總資料倉儲轉換為多維度模型，使總資料倉儲能每日定時更新資料。自動化 Macro 程式所挾帶的檔案為 MS-DOS 的批次檔案，程式撰寫如圖 3.29 所示。

```

@ECHO -- Stop PowerPlay Enterprise Server ---

@ECHO -- Start PowerPlay Enterprise Server ---

start /w C:\Program\1\Cognos\cer1\bin\runmac32.exe D:\Cognos3\TENWAY.MCX /w

@ECHO -- Stop PowerPlay Enterprise Server ---
NET STOP "Cognos PowerPlay Enterprise Server"

@ECHO -- Stop Upfront ---
NET STOP "Cognos Upfront Dispatcher"
NET STOP "Cognos Upfront Data Store"
NET STOP "Cognos Upfront Administration Service"

cd\
D:
copy o總倉儲.mdc D:\cube\

@ECHO --- Start Ticket Server ---
NET START "Cognos Ticket Server"
@ECHO --- Start COGNOS Services ---
NET START "Cognos Communication Service (cer1)"

@ECHO --- Start Upfront ---
NET START "Cognos Upfront Administration Service"

NET START "Cognos Upfront Data Store"
NET START "Cognos Upfront Dispatcher"

@ECHO --- Start PowerPlay Enterprise Server ---
NET START "Cognos PowerPlay Enterprise Server"

@ECHO --- Start Visualizer ---
NET START "Cognos Visualizer Web Edition"

exit

```

圖 3.29 自動化 Macro 程式

3.3.3 自動化設計

在拍賣資料轉換工作、訂貨資料轉換工作、殘貨資料轉換工作、議價資料轉換工作完成後，必須建立自動化的的工作排程，讓資料不需手動能自動轉到資料倉儲裡，以節省人力與時間。

在完成自動化更新程序的資料轉換服務工作以後，利用 SQL Server2000 Agent 進行排程，首先編輯作業步驟，如圖 3.30 自動化步驟語法編輯採用作業系統命令語法，代表去路徑下執行自動化更新程序.dts 的檔案，接著再編輯作業排程，如圖 3.31 自動化排程設定，命名此排程為總資料倉儲轉換，選擇重複執行方式便可以設定每日執行時間，如圖 3.32 所示，便完成了自動化的設計。

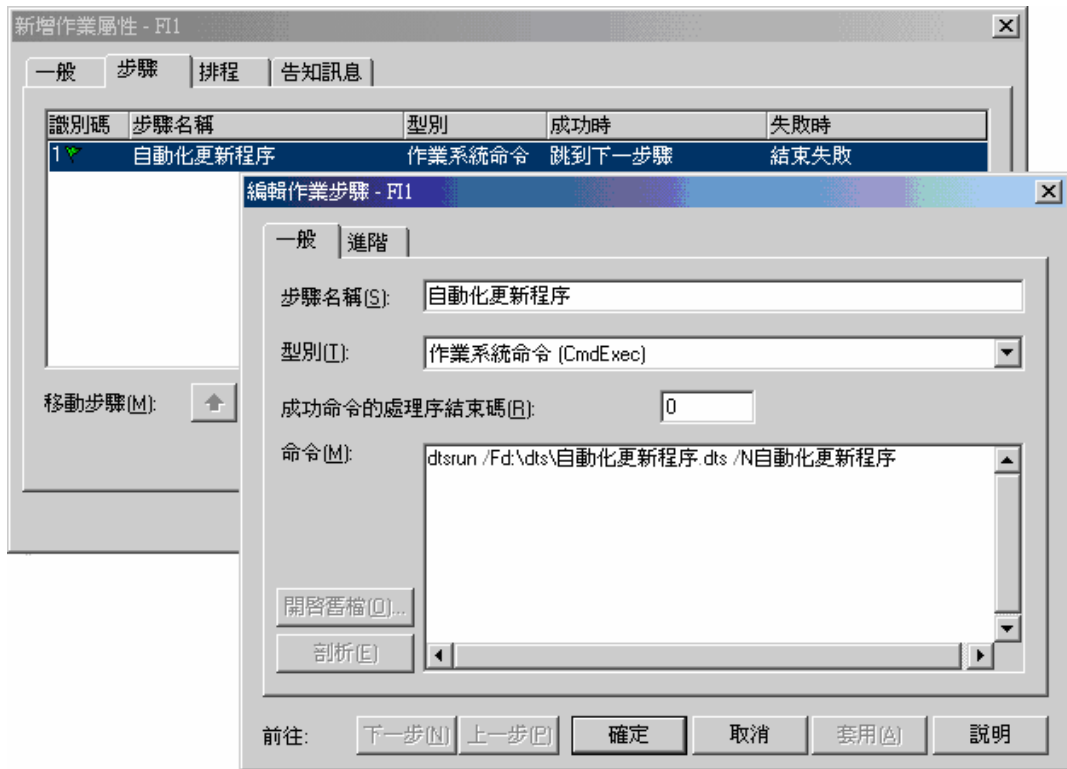


圖 3.30 自動化步驟語法編輯

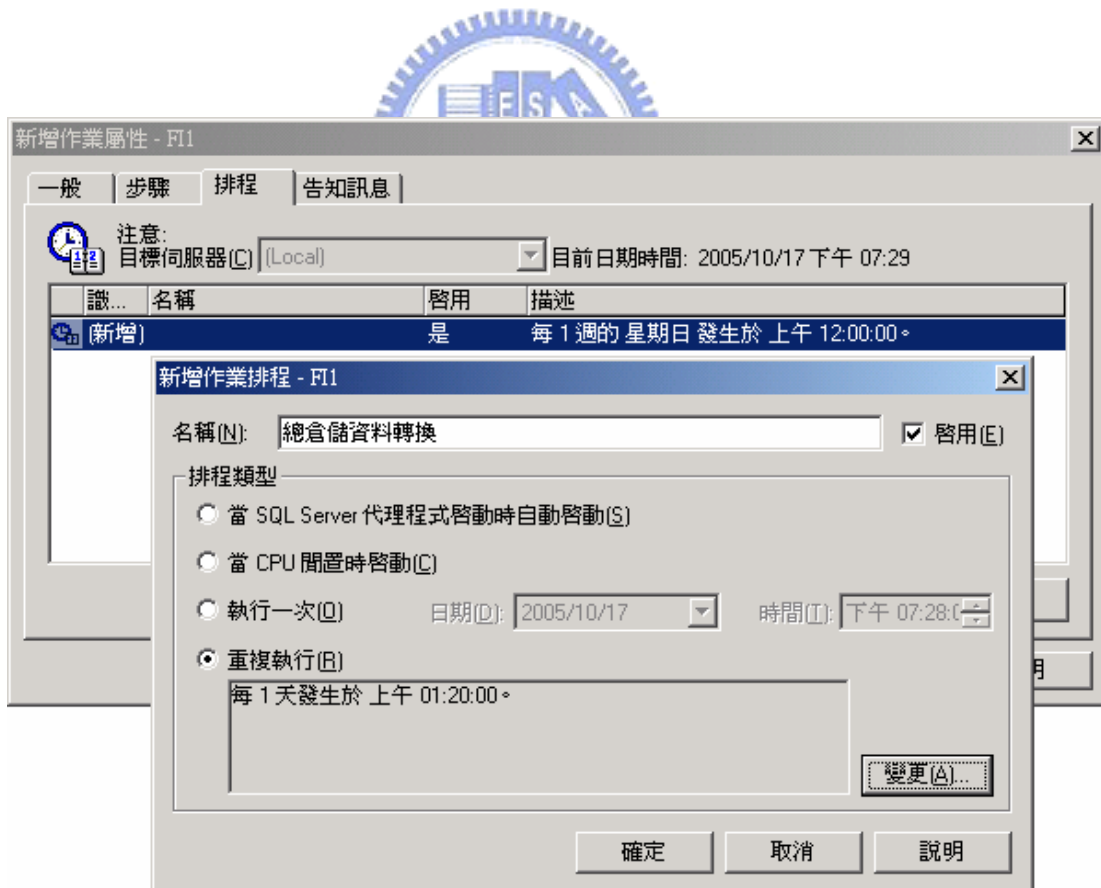


圖 3.31 自動化排程設定

編輯重複執行作業排程 - FI1

作業名稱: 自動化更新程序 snow3

發生頻率

每天(D) 每日
 每隔(N) 1 天

每週(W)
 每月(M)

每日頻率

執行一次於(U): 上午 01:20:00
 重複執行於每(B): 1 小時 開始時間(U): 上午 01:00:00
 結束時間(N): 下午 11:59:59

持續時間

開始日期(S): 2005/10/17 結束日期(E): 2006/ 1/ 9
 沒有結束日期(A)

確定 取消 說明

圖 3.32 自動化排程重複執行設定



3.4 資料倉儲遠端維修方式

對於資料倉儲的管理者而言，一套完善的資料倉儲系統必須使資料的轉換運作流暢，才能提供使用者每日更新的資訊。因此當資料倉儲發生了不可預期的錯誤時，管理者必須盡可能在最短的時間內進行維修。由於存放系統的伺服器是擺在固定的地方，而管理者不可能隨時待在伺服器旁，因此採用遠端遙控軟體的維修方式。只要有網路的地方，就可以進行維修的工作，目的在節省維修的前置時間，也就是減少管理者往來維修的人力與運輸成本。資料倉儲伺服器的管理者可以透過遠端連線對伺服器進行重新開機，或者是手動執行資料轉換服務的工作。進行維修時常用於資料轉換服務後的偵錯，配合SQL Query Analyzer的語法查詢，找出資料錯誤之處。第 3.4.1 節介紹遠端遙控軟體Ultr@VNC架構。第 3.4.2 節介紹VNC的操作方式。第 3.4.3 節舉一例子說明如何進行維修。

3.4.1 VNC 架構

VNC (Virtual Network Computing) 軟體為英國劍橋大學 ATT 實驗室 (Association of Taxation Technicians) 所發展出來的開放原始碼軟體，其中 Ultr@VNC 便是 UltrVNC Team 以 VNC 為基礎所發展出來的一套免費遠端遙控軟體。其最強大的功能在於能透過網路連接分隔兩地的電腦，使用者可透過該軟體直接操控遠端電腦的桌面，也可以直接用瀏覽器當成操控的介面，表 3.10 為 Ultr@VNC 的功能特點。

表 3.10 Ultr@VNC 功能特點

項目	功能特點
1	提供自動偵測與快速設定的選項。
2	工具列上提供主要的功能按鈕。
3	執行及傳輸速度快。
4	內建檔案傳輸功能，可直接在兩台電腦間對傳檔案。
5	支援 MS 的登入與 NT 安全性設定，並可記錄使用者的行為。
6	提供減低傳輸資料量的選項。例如：直接縮小主機提供的畫面、快取管理、本機游標管理，與調整顯示顏色等。
7	開放外掛程式介面，可外加各種加密、權限管理等功能。
8	可直接用瀏覽器開啟遠端桌面遙控。

Ultr@VNC 可分為 VNC 伺服器與檢視器兩個部分，圖 3.32VNC 伺服器與檢視器圖示所示為此兩個子部分在電腦中的圖示，其中 VNC 伺服器必須安裝在被遠端遙控的電腦。在 VNC 伺服器的主要屬性選項中，必須設定好連線密碼，才能讓本地的使用者連線。VNC 伺服器的主要屬性選項如圖 3.33VNC 伺服器主要屬性選項所示。



圖 3.32 VNC 伺服器與檢視器圖示

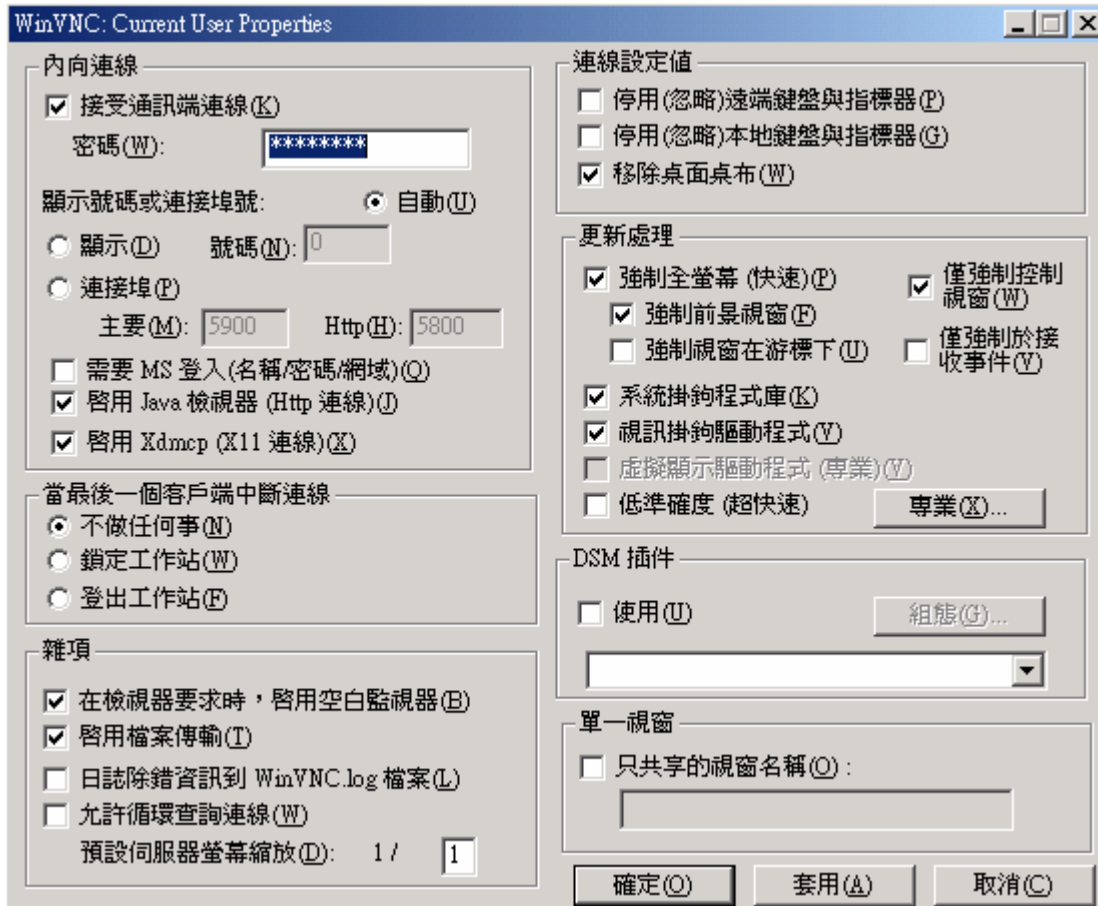


圖 3.33 VNC 伺服器主要屬性選項

3.4.2 VNC 的操作

VNC 檢視器則安裝在使用者欲使用的本地電腦，使用者透過 VNC 檢視器來建立與遠端電腦的連線。圖 3.34 所示為連線設定的視窗，使用者在 VNC 伺服器選項的空格中輸入遠端電腦的 IP 位置；另外在選項按鈕中，可做其他連線的設定，如圖 3.35 所示。



圖 3.34 VNC 檢視器連線視窗

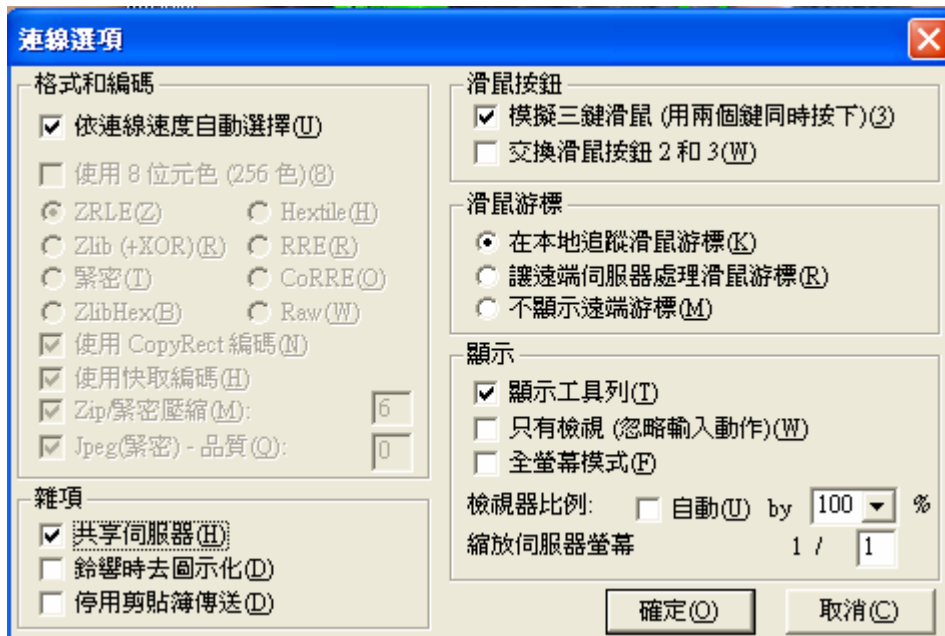


圖 3.35 VNC 檢視器連線選項

設定好連線屬性後，便可啟動與遠端電腦的連線。此時會跳出 VNC 檢視器連線狀態與 VNC 驗證的視窗，如圖 3.36、圖 3.37 所示。在 VNC 驗證的視窗中，輸入設定於遠端 VNC 伺服器的密碼，便可成功的產生遠端桌面連線的視窗，如圖 3.38 所示。本地使用者可在視窗內進行遠端遙控的動作，其中在視窗的上排為快速工具列。較常使用的為左一的按鈕，發送 Ctrl+Alt+Del 的動作，因此本地使用者也可以透過遠端遙控電腦重新開機。



圖 3.36 VNC 檢視器連線狀態

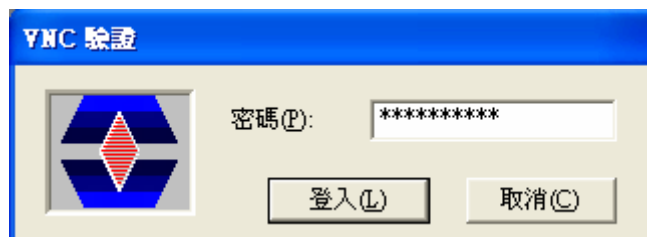


圖 3.37 VNC 驗證

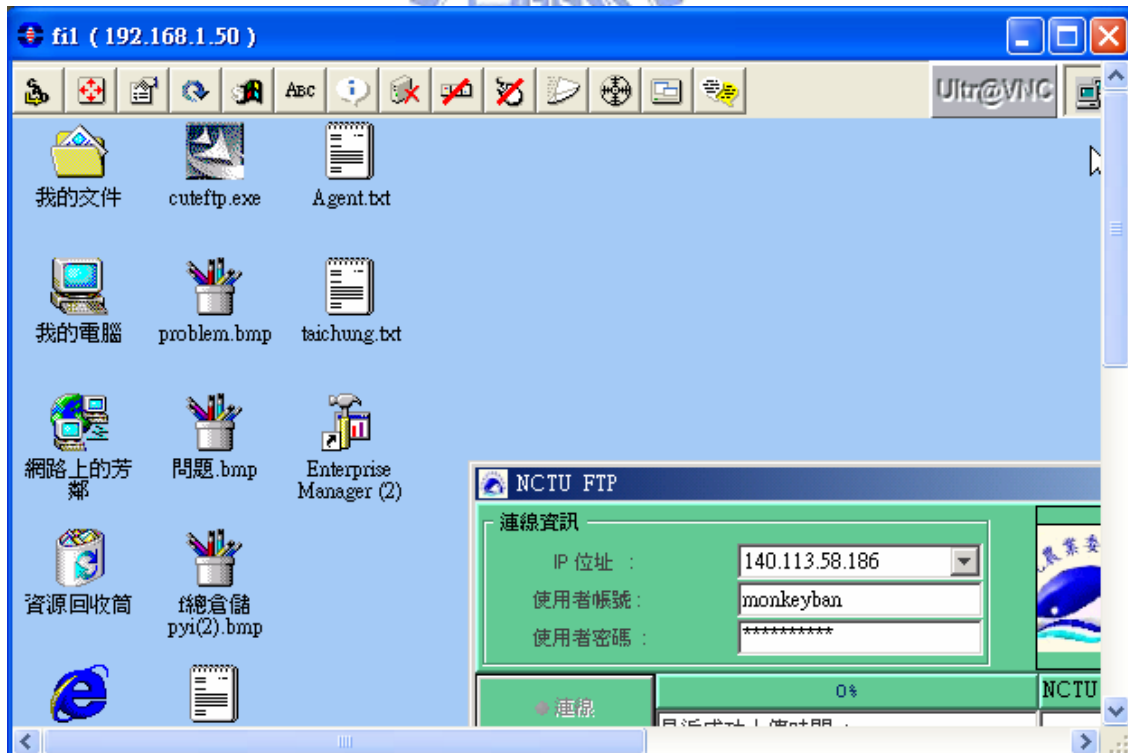


圖 3.38 遠端桌面視窗

3.4.3 維修案例說明

以下舉資料轉換服務錯誤的維修案例，以總資料倉儲為維修對象，說明資料倉儲伺服器的管理者如何透過遠端連線進行維修。

當資料倉儲的管理者發現資料倉儲的資料，從某日開始沒有定期更新，而管理者本身無法立即前往伺服器所在位置進行維修工作時，便可利用遠端遙控方式開始偵錯與維修。首先管理者必須啟動 VNC 檢視器，啟動方式為上述 3.4.2 節的操作步驟，輸入 IP 位置連線之後，再進行密碼驗證，請參考圖 3.34 與圖 3.37。接著便進入遠端桌面的視窗畫面，如圖 3.38。

視窗內的畫面即為資料倉儲伺服器的桌面，此時先手動開啟資料轉換服務進行偵錯的工作，以確認資料轉換服務於何處發生錯誤。圖 3.39 為手動資料轉換服務完成後，發現錯誤的畫面。本範例啟動的資料轉換服務為自動化更新程序 DTS_0，此錯誤訊息陳述資料轉換衝突發生於拍賣資料轉換工作 DTS_1 中的台北拍賣 DTS_1-1 處，該衝突原因在於「總倉儲交易資料表」與「花卉品名」資料表間的主外鍵無法產生一對一的關聯；換句話說，可以推斷由台北資料超市的「拍賣資料表」要轉入「總倉儲交易資料表」中的花卉品名代碼，並不存在於總倉儲的「花卉品名」資料表中。

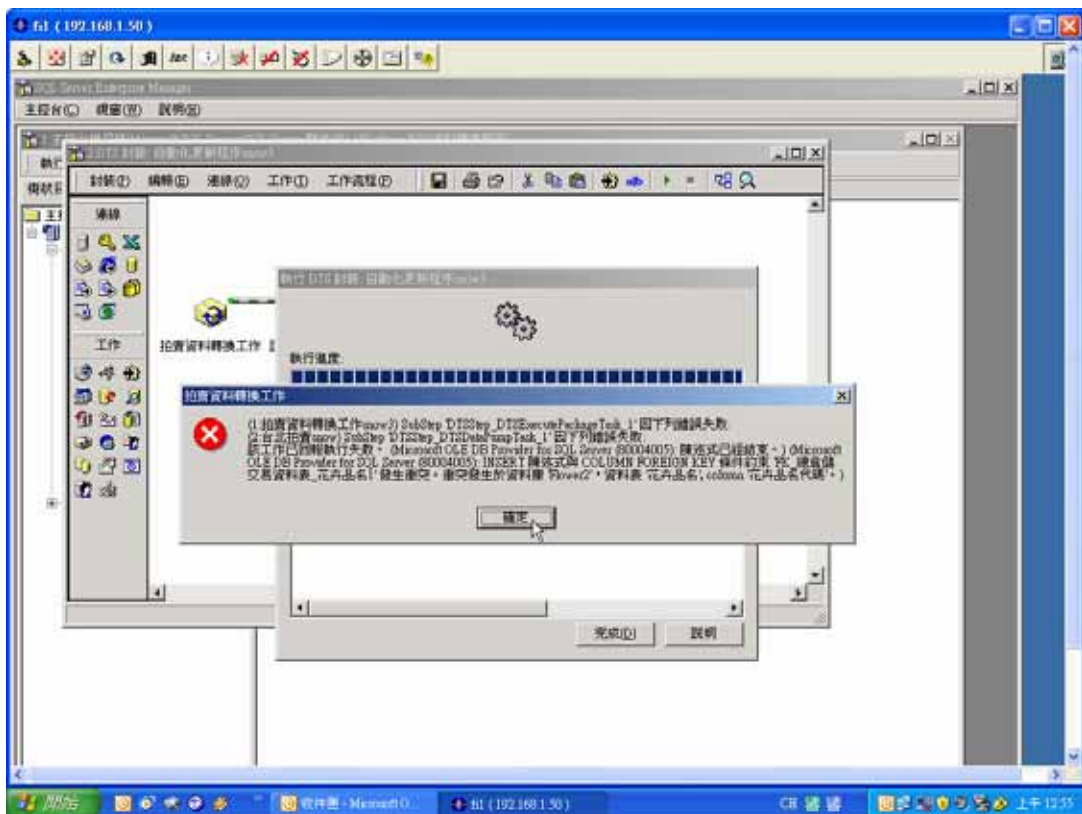


圖 3.39 資料轉換服務錯誤畫面

因此管理者必須找出在台北資料超市中，哪幾筆資料的花卉品名不存在於總資料倉儲的「花卉品名」資料表，然後將這幾筆可能是台北資料超市新增的花卉品名加入總資料倉儲的「花卉品名」資料表。

接著將台北資料超市的「花卉資料表」匯入總資料倉儲的 flower2 資料庫，取名為「花

卉資料表 1」，如圖 3.40 中方框內所示。接著開啟工具 SQL Query Analyzer，寫入 SQL 查詢語法進行資料比對找出錯誤。圖 3.41 方框內為進行資料比對的 SQL 語法，意思是找出在 flower2 資料庫中，該幾筆資料存在於「花卉資料表 1」的 TP_PR_NUM 欄位，卻不存在於「花卉品名」資料表的花卉品名代碼欄位中。最後將查詢的結果新增於「花卉品名」資料表中，此後資料轉換服務即可順利每日運作。

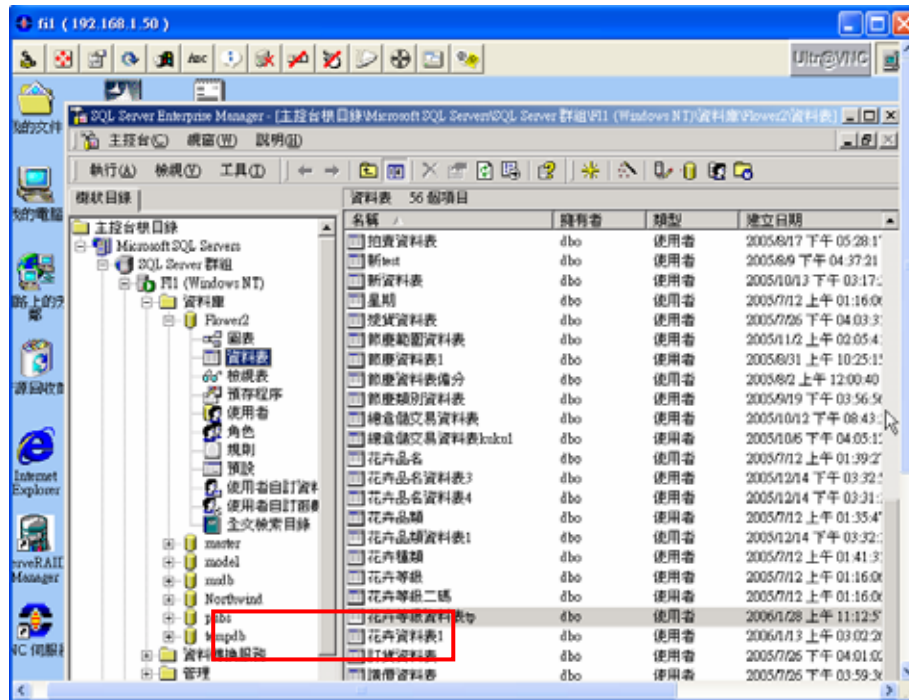


圖 3.40 SQL Server 視窗畫面

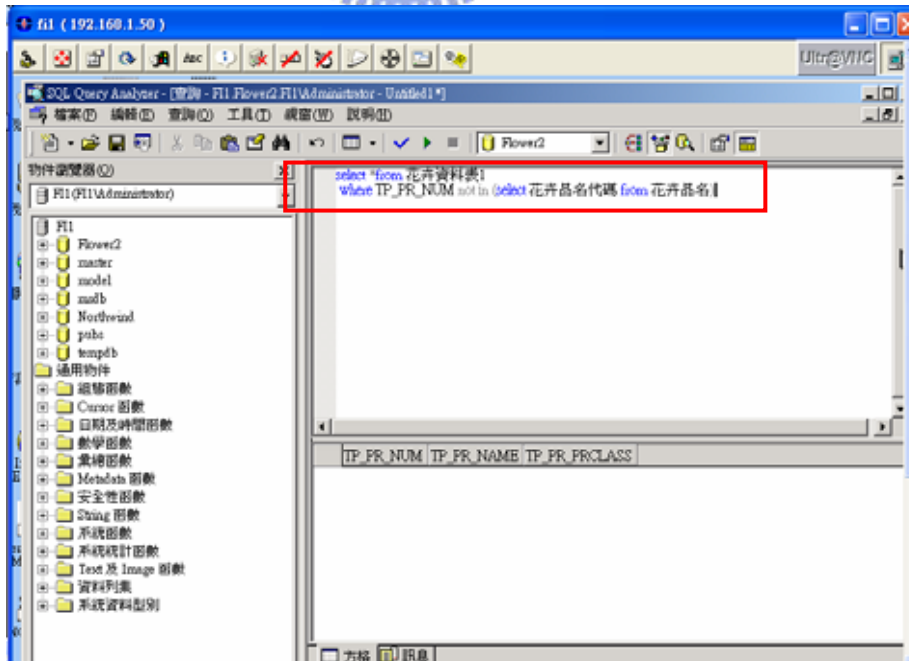


圖 3.41 Query Analyzer 視窗畫面

第四章 節慶維度設計

對於使用資料倉儲來分析的資訊人員來說，有價值的資訊才是觀察的重點，長期在地方花卉批發市場工作的人員，觀察到一個明顯的現象，就是花卉的交易與節慶日期之間有很高的相關性，以普遍的例子來說，在情人節前後玫瑰花的銷售量和價格都比一般日還高出許多，因此重要節日前後幾天的銷售數據，絕對是做為資料分析的重要指標，在台灣區花卉資料倉儲的架構上，加入節慶維度之後，將使得花卉資訊價值的應用更加多元化。

本章第 4.1 節主要說明維度內容與節慶設計的特殊情況。第 4.2 節介紹四種不同的維度設計方式。第 4.3 節探討四種方案的優缺點。

4.1 維度內容介紹

總資料倉儲共包含供應單位、供應地區、日期、星期、農曆、節慶、市場別、拍賣線、花卉種類、花卉等級、容器、承銷地區共十二個維度，其中日期、星期、農曆與節慶為與時間相關的維度，本節重點將放在節慶維度的介紹，將說明特殊節慶選定的衍生性問題，採用哪些特殊節慶做為維度內容，以及節慶個數多寡的選擇，都會影響後續的設計方式，第 4.1.1 節介紹節慶選定的基本資訊；第 4.1.2 節介紹節慶在日期上的重複問題；第 4.1.3 節介紹節慶在日期上的排序問題；第 4.1.4 節介紹節慶可能產生數據為空值的問題。

4.1.1 節慶相關資訊

以台北、台中、彰化、台南四家資料超市來說，最早的交易資料從西元 1996 年開始，因此節慶日期的開始時間遵循交易資料起始日期，從西元 1996 年開始轉入節慶資料。

特殊節慶的選擇則按照花卉批發市場工作人員的意見，分為國曆節慶與農曆節慶兩種，國曆特殊節慶有開國紀念日、西洋情人節、白色情人節、清明節、母親節、父親節、教師節、國慶紀念日、聖誕節共九個節日；農曆特殊節慶包含有農曆春節、天公生、元宵節、土地公生、觀音生、三日節、媽祖生、浴佛節、端午節、觀音得道、關公生、七夕中國情人節、中元節、中秋節、重陽節、觀音出家、尾牙、送神共十八個節日，每年共包含二十七個重要節慶，如表 4.1 和表 4.2 所示。

表 4.1 國曆特殊節慶

NO	國曆日期	節日名稱
1	1月1日	開國紀念日
2	2月14日	西洋情人節
3	3月14日	白色情人節
4	4月5日	清明節
5	5月8日	母親節
6	8月8日	父親節
7	9月28日	教師節
8	10月10日	國慶紀念日
9	12月25日	聖誕節

表 4.2 農曆特殊節慶

NO	農曆日期	節日名稱
1	1月1日	農曆春節
2	1月9日	天公生
3	1月15日	元宵節
4	2月2日	土地公生
5	2月19日	觀音生
6	3月3日	三日節
7	3月23日	媽祖生
8	4月8日	浴佛節
9	5月5日	端午節
10	6月19日	觀音得道
11	6月23日	關公生
12	7月7日	七夕中國情人節
13	7月15日	中元節
14	8月15日	中秋節
15	9月9日	重陽節
16	9月19日	觀音出家
17	12月16日	尾牙
18	12月24日	送神

此外，節慶影響的日期範圍，按照地方花卉批發市場工作人員的意見統計過後，決定將節慶擴及的範圍設為前七天與後五天。

4.1.2 節慶日期重複問題

特殊節慶的總個數有二十七個，橫跨範圍為前七天後五天，但一年有三百六十五天，不可避免的不同節慶期間會有重複到同一日的情形，加上節慶有國曆農曆之分，節慶日期愈相近，重複的情形就愈顯著，這使得節慶維度設計的工作變的複雜，圖 4.1 節慶重複情形示意圖所示，呈現出節慶日期重複的情形，如箭頭所指處 2005/2/8 重複在農曆春節與西洋情人節兩個節慶期間。

2005/2/6	送神後四天	農曆春節前三天
2005/2/7	送神後五天	農曆春節前二天 西洋情人節前七天
2005/2/8	農曆春節前一天	西洋情人節前六天
2005/2/9	農曆春節	西洋情人節前五天
2005/2/10	農曆春節後一天	西洋情人節前四天 天公生前七天

圖 4.1 節慶重複情形示意圖

考慮到一個日期對應到多個節慶名稱的情形，如圖 4.2 一日期對多節慶示意圖所示，cognos 軟體會重複計算衡量值，因此最初的節慶欄位設計，必須遵照資料庫連結一對一的準則，決定將多個節慶名稱放在同一欄位，如圖 4.1 所呈現樣子。

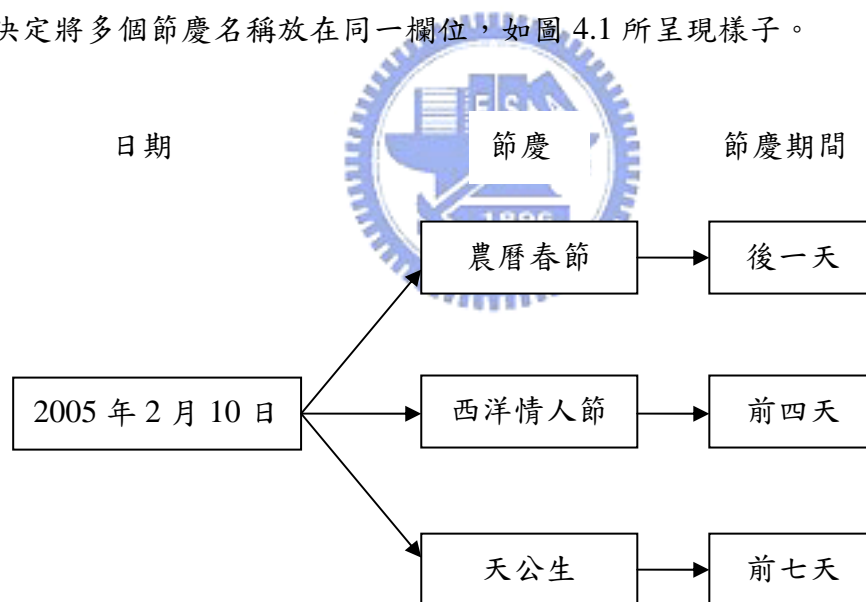


圖 4.2 一日期對多節慶示意圖

此外，農曆日期對應到的國曆日期逐年不同，可以想見的是隨著不同年份，節慶重複的形式也都不相同，增加往後建立節慶資料時的困難度。

4.1.3 節慶日期排序問題

節慶維度設計給使用者使用時，若按照日期排序，可增加使用上的友善度，但日期並不會顯現於節慶維度層級之中（註一），當線上處理並未展開時間維度時，會由於不同年份的關係，使得節慶日期的排序造成同個節慶卻放在不同位置上的混亂，這個問題在線上呈現節慶維度上，有很大的問題。

舉一個例子詳細解釋，如圖 4.3 在 2000/8/12 為『七夕中國情人節後五天、父親節後三天、中元節前三天』，應置於 2002/8/12 為『父親節後四天、七夕中國情人節前三天』之前，但資料倉儲只會顯示『』中的內容，因此使用者在操作資料倉儲時，會對於『七夕中國情人節後五天』擺在『七夕中國情人節前三天』之前感到邏輯錯亂，實際上這是由於選取的日期維度層級在『年』的關係，只要選取不同年份就可以解決邏輯上的錯亂，但考慮到使用的容易程度，在設計上仍然要對這個問題做改善。

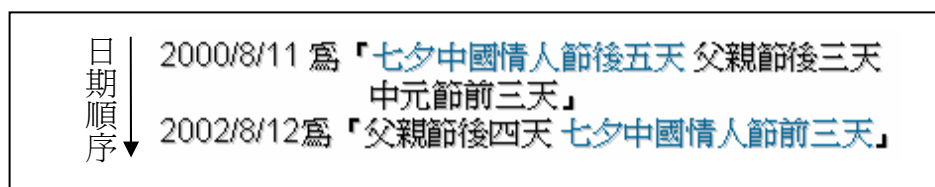


圖 4.3 節慶日期排序問題示意圖

4.1.4 節慶產生的空值問題

此節所要介紹的節慶產生空值問題，是由第 4.1.3 節中年份所造成的排序問題衍生而來，由於節慶維度在呈現上並不會拉出日期的層級，因此在節慶維度中，特殊節慶並不會被限定在特定的日期範圍內，因此就算在日期維度選擇特定時間範圍，不在範圍內的特殊節慶還是會顯現出來，而顯示該數值為 0 的情況，這種情形容易造成使用者的混亂，因為無從去判斷該節慶範圍是屬於哪一個日期。表 4.3 整理出節慶三個特殊問題的相關資訊。

表 4.3 節慶三大問題

節慶問題	發生原因	解決方法
重複	節慶日期相近	1. 從設計上著手
排序	維度內部不受日期限制	1. 操作外部日期維度來限制 2. 從設計上著手
空值	維度內部不受日期限制	1. 使用『清零』功能 2. 從設計上著手

註一：日期在資料倉儲的設計上，原本就是一個維度，所以節慶維度上並不會顯示出日期這個欄位。

4.2 節慶維度設計架構

在維度設計上，可以從資料表的欄位設計、資料表的內容、資料表的數目...等等做不同的改變，不同的設計方式造就的結果也不相同，以下介紹四種不同的節慶維度設計方式。第 4.3.1 節介紹第一種維度設計方式；第 4.3.2 節介紹第二種維度設計方式；第 4.3.3 節介紹第三種維度設計方式；第 4.3.4 節介紹第四種維度設計方式。

4.2.1 維度設計方案一

節慶維度設計方案一的架構中，只以一個節慶資料表與事物表 (Fact Table) 做連結，其欄位屬性包括日期、節慶名稱。與事物表 (Fact Table) 關聯如圖 4.4 節慶資料表關聯圖-方案一。

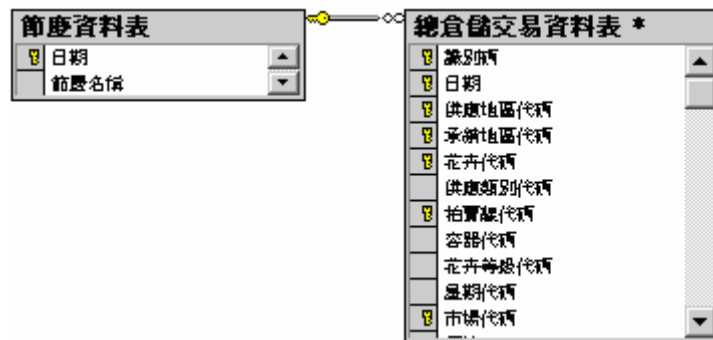


圖 4.4 節慶資料表關聯圖-方案一

表 4.4 節慶資料表-方案一為節慶資料表屬性的詳細說明，包含資料表的主鍵、外鍵、資料欄名稱、資料型別、長度、可否為空值及相關說明。實際資料表顯示如圖 4.5 節慶資料表實體資料-方案一。

表 4.4 節慶資料表-方案一

主鍵	外鍵	欄位名	資料型別	允許空值	長度	說明
◎		日期	datetime		8	西曆日期，ex: 2005/09/26
		節慶名稱	varchar		30	節慶前七天後五天期間，ex: 教師節前一天

日期	節慶名稱
▶ 1996/1/1	開國紀念日
1996/1/2	開國紀念日後一天
1996/1/3	開國紀念日後二天
1996/1/4	開國紀念日後三天
1996/1/5	開國紀念日後四天
1996/1/6	開國紀念日後五天
1996/1/7	一般日
1996/1/8	一般日
1996/1/9	一般日
1996/1/10	一般日

圖 4.5 節慶資料表實體資料-方案一

4.2.2 維度設計方案二

節慶維度設計方案二的架構中，包含一個節慶資料表，其欄位屬性為日期、節慶名稱與節慶類別。與事物表（Fact Table）關聯如圖 4.6 節慶資料表關聯圖-方案二。

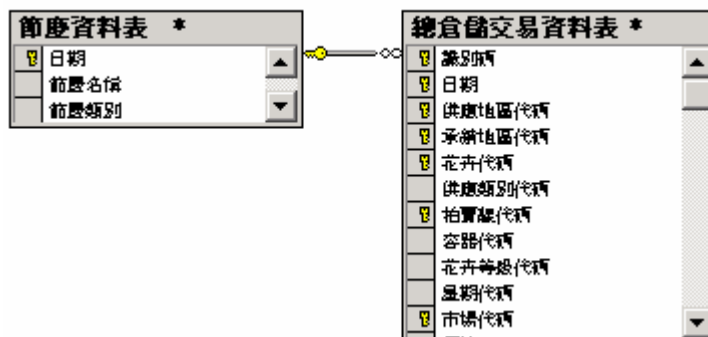


圖 4.6 節慶資料表關聯圖-方案二

表 4.5 節慶資料表-方案二為節慶資料表屬性的詳細說明，包含資料表的主鍵、外鍵、資料欄名稱、資料型別、長度、可否為空值及相關說明。實際資料表顯示如圖 4.7 節慶資料表實體資料-方案二。在建立維度層級時，將節慶類別放置於第一層、節慶名稱放置為第二層，如表 4.6 方案二維度層級表所示。

表 4.5 節慶資料表-方案二

主鍵	外鍵	欄位名	資料型別	允許空值	長度	說明
◎		日期	datetime		8	西曆日期，ex: 2005/09/26
		節慶名稱	varchar		30	節慶前七天後五天期間，ex: 教師節前一天
		節慶類別	int		4	節慶類別名稱，ex: 教師節

	日期	節慶名稱	節慶類別
▶	1996/1/1	開國紀念日	開國紀念日
	1996/1/2	開國紀念日後一天	開國紀念日
	1996/1/3	開國紀念日後二天	開國紀念日
	1996/1/4	開國紀念日後三天	開國紀念日
	1996/1/5	開國紀念日後四天	開國紀念日
	1996/1/6	開國紀念日後五天	開國紀念日
	1996/1/7	一般日	無節慶
	1996/1/8	一般日	無節慶
	1996/1/9	一般日	無節慶
	1996/1/10	一般日	無節慶
	1996/1/11	一般日	無節慶
	1996/1/12	一般日	無節慶

圖 4.7 節慶資料表實體資料-方案二

表 4.6 方案二維度層級表

層級	層級名稱	舉例說明
第一層	節慶類別	ex:教師節
第二層	節慶期間	ex:教師節前一天

4.2.3 維度設計方案三

節慶維度設計方案三的架構中，包含節慶範圍資料表與節慶類別資料表，節慶範圍資料表的屬性欄位為日期、節慶類別代碼與節慶名稱，節慶類別資料表的欄位屬性為節慶類別代碼、節慶類別。與事物表（Fact Table）關聯如圖 4.8 節慶二資料表關聯圖-方案三。



圖 4.8 節慶二資料表關聯圖-方案三

表 4.7 節慶範圍資料表-方案三、表 4.8 節慶類別資料表-方案三分別為節慶範圍資料表、節慶類別資料表屬性的詳細說明，包含資料表的主鍵、外鍵、資料欄名稱、資料型別、長度、可否為空值及相關說明。

表 4.7 節慶範圍資料表-方案三

主鍵	外鍵	欄位名	資料型別	允許空值	長度	說明
◎		日期	datetime		8	西曆日期，ex: 2005/09/26
		節慶名稱	varchar		30	節慶前七天後五天期間，ex: 教師節前一天
	◎	節慶類別代碼	int		4	節慶類別代碼1~69，ex: 教師節代碼為56

表 4.8 節慶類別資料表-方案三

主鍵	外鍵	欄位名	資料型別	允許空值	長度	說明
◎		節慶類別代碼	int		4	節慶類別代碼1~69，ex: 教師節代碼為56
		節慶類別	varchar		30	節慶類別名稱，ex: 教師節

詳細說明節慶維度設計方案三的架構關係，在此以實際資料表來說明關聯性，如圖 4.9 節慶二資料表實體內容-方案三。虛線框住部分顯示兩個資料表以節慶類別代碼作為關聯，目的在使節慶類別與節慶範圍能達到上捲下挖的功能，在建立維度層級時，將節

慶類別放置於第一層、節慶名稱放置為第二層，如表 4.9 方案三維度層級表所示，在 OLAP 技巧中，是依層級順序做下挖的處理動作。

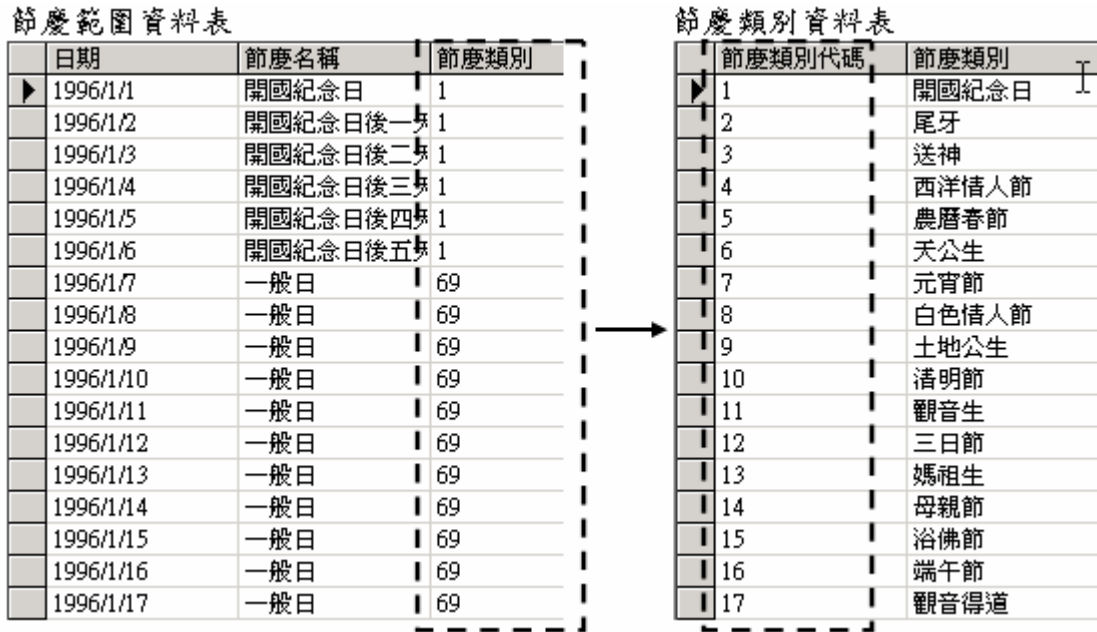


圖 4.9 節慶二資料表實體內容-方案三

表 4.9 方案三維度層級表

層級	層級名稱	舉例說明
第一層	節慶類別	ex:教師節
第二層	節慶期間	ex:教師節前一天

4.2.4 維度設計方案四

節慶維度設計方案四的架構中，延續 4.2.3 維度設計方案三中的設計，保留節慶範圍資料表與節慶類別資料表，但將節慶範圍資料表的欄位以及節慶類別資料表的資料內容做改變。

在節慶範圍資料表中，新增了『年份』欄位，因此其屬性欄位共包含有節慶日期、節慶名稱、節慶類別代碼及節慶年份。而節慶類別資料表的屬性欄位為節慶類別代碼和節慶類別，節慶類別欄位的資料則按照日期、農曆節慶、國曆節慶的優先順序，重新編制類別代碼，並將代碼置於節慶類別資料之前。兩個資料表的關聯如圖 4.10，便完成了節慶維度。



圖 4.10 節慶二資料表關聯圖-方案四

表 4.10 節慶範圍資料表-方案四、表 4.11 節慶類別資料表-方案四分別為節慶範圍資料表、節慶類別資料表屬性的詳細說明，包含資料表的主鍵、外鍵、資料欄名稱、資料型別、長度、可否為空值及相關說明。表 4.12 節慶類別代碼對照表列出 69 個節慶類別代碼，可供給對照。

表 4.10 節慶範圍資料表-方案四

主鍵	外鍵	欄位名	資料型別	允許空值	長度	說明
◎		日期	datetime		8	西曆日期，ex: 2005/09/26
		節慶名稱	varchar		30	節慶前七天後五天期間，ex: 教師節前一天
	◎	節慶類別代碼	int		4	節慶類別代碼1~69，ex: 教師節代碼為56
		年份	char		4	西曆年份，ex: 2005

表 4.11 節慶類別資料表-方案四

主鍵	外鍵	欄位名	資料型別	允許空值	長度	說明
◎		節慶類別代碼	int		4	節慶類別代碼1~69，ex: 教師節代碼為56
		節慶類別	varchar		30	節慶代碼+節慶類別名稱，ex: 56教師節

表 4.12 節慶類別代碼對照表

NUM	F_NAME	NUM	F_NAME	NUM	F_NAME
1	1 開國紀念日	24	24 白色情人節 觀音生	47	47 父親節 七夕中國情人節 中元節
2	2 開國紀念日 尾牙	25	25 土地公生 白色情人節	48	48 父親節 中元節
3	3 尾牙	26	26 土地公生	49	49 七夕中國情人節
4	4 尾牙 送神	27	27 清明節	50	50 七夕中國情人節 中元節
5	5 尾牙 送神 西洋情人節	28	28 清明節 觀音生	51	51 中元節
6	6 送神	29	29 清明節 三日節	52	52 中秋節
7	7 送神 西洋情人節	30	30 清明節 媽祖生	53	53 中秋節 教師節
8	8 送神 西洋情人節 農曆春節	31	31 觀音生	54	54 中秋節 教師節 國慶紀念日
9	9 送神 農曆春節	32	32 三日節	55	55 中秋節 國慶紀念日
10	10 西洋情人節	33	33 媽祖生	56	56 教師節
11	11 西洋情人節 天公生	34	34 媽祖生 母親節	57	57 教師節 國慶紀念日
12	12 西洋情人節 天公生 元宵節	35	35 母親節	58	58 教師節 國慶紀念日 重陽節
13	13 西洋情人節 元宵節	36	36 母親節 浴佛節	59	59 教師節 重陽節
14	14 西洋情人節 土地公生	37	37 浴佛節	60	60 國慶紀念日
15	15 農曆春節 西洋情人節	38	38 端午節	61	61 國慶紀念日 重陽節
16	16 農曆春節 西洋情人節 天公生	39	39 觀音得道	62	62 國慶紀念日 觀音出家
17	17 農曆春節	40	40 觀音得道 父親節	63	63 重陽節 國慶紀念日 觀音出家
18	18 農曆春節 天公生	41	41 觀音得道 關公生	64	64 重陽節
19	19 天公生	42	42 觀音得道 關公生 父親節	65	65 重陽節 觀音出家
20	20 天公生 元宵節	43	43 關公生	66	66 觀音出家
21	21 元宵節	44	44 關公生 父親節	67	67 聖誕節
22	22 元宵節 白色情人節	45	45 父親節	68	68 聖誕節 開國紀念日
23	23 白色情人節	46	46 父親節 七夕中國情人節	69	69 一般日

節慶範圍資料表

日期	節慶名稱	節慶類別代碼	年份
1996/1/2	開國紀念日後一	1	1996
1996/1/3	開國紀念日後二	1	1996
1996/1/4	開國紀念日後三	1	1996
1996/1/5	開國紀念日後四	1	1996
1996/1/6	開國紀念日後五	1	1996
1996/1/7	一般日	69	1996
1996/1/8	一般日	69	1996
1996/1/9	一般日	69	1996
1996/1/10	一般日	69	1996
1996/1/11	一般日	69	1996
1996/1/12	一般日	69	1996
1996/1/13	一般日	69	1996
1996/1/14	一般日	69	1996
1996/1/15	一般日	69	1996
1996/1/16	一般日	69	1996
1996/1/17	一般日	69	1996

節慶類別資料表

節慶類別代碼	節慶類別
1	開國紀念日
2	尾牙
3	送神
4	西洋情人節
5	農曆春節
6	天公生
7	元宵節
8	白色情人節
9	土地公生
10	清明節
11	觀音生
12	三日節
13	媽祖生
14	母親節
15	浴佛節
16	端午節

圖 4.11 節慶二資料表實體內容-方案四

表 4.13 方案四維度層級表

層級	層級名稱	舉例說明
第一層	節慶類別	ex:56 教師節
第二層	年份	ex:2005 年
第三層	節慶期間	ex:教師節前一天

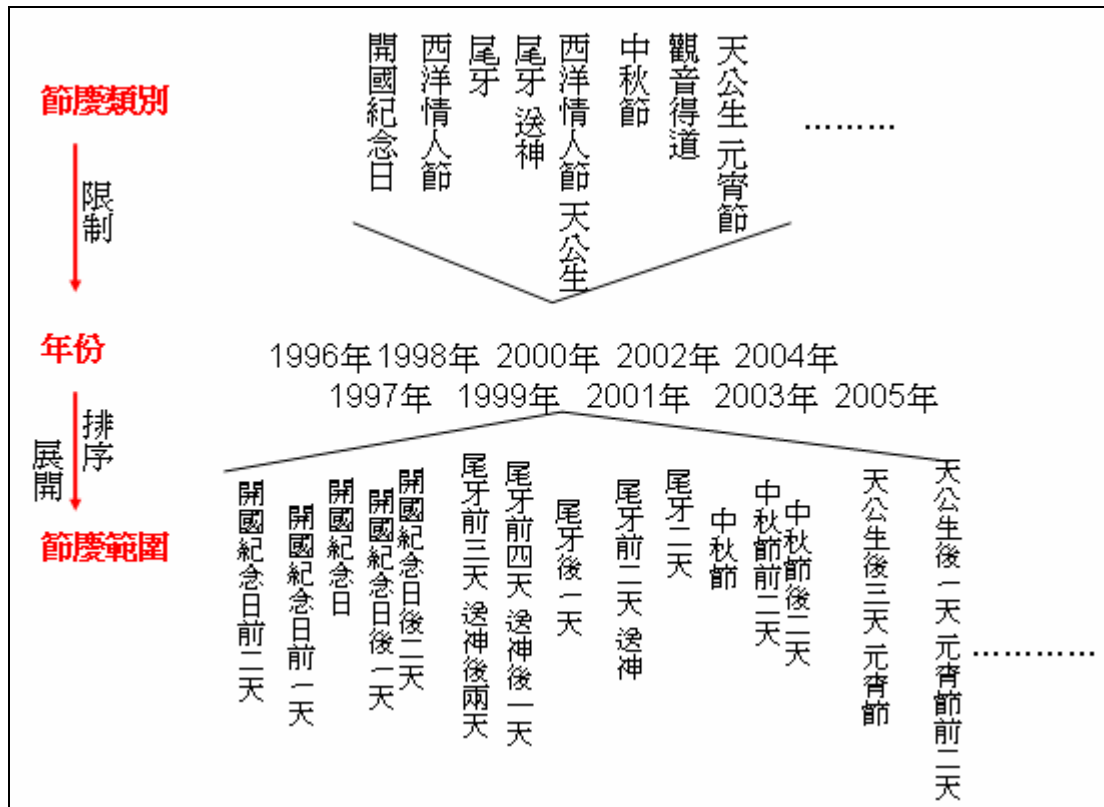


圖 4.12 節慶層級概念

4.3 設計方案的比較

各種節慶的設計方式都有優缺點，不管是讓資料讀取速度變快，或是讓畫面好看，最重要的是要讓使用者容易操作資料倉儲的功能，並能快速查詢到目標數據，本節分為兩小節。第 4.3.1 節探討四個方案在維度層級的線上設計方式；第 4.3.2 節討論四個方案設計上的優缺點。

4.3.1 線上顯示設計模式

在操作資料倉儲時，維度上面最重要的顯示部分就是層級，這包含該維度內有幾個層級以及層級內容的排列方式。

在維度設計方案一中，只包含節慶期間一個層級，所以並無上捲下挖的功能，因此節慶期間就依照日期順序排列顯示。

在維度設計方案二中，包含有節慶類別與節慶期間兩個層級，節慶類別與節慶期間皆依照日期順序排列顯示，但不同年份會出現不同的節慶類別，因此在類別的呈現上會出現第 4.1.3 節所敘述年份的邏輯混亂現象，在維度的第一層出現這種情形並不是好現象。

在維度設計方案三中，包含有節慶類別與節慶期間兩個層級，但節慶類別在此是以節慶類別代碼順序排列，能避開第 4.1.3 節所敘述年份的邏輯混亂現象，而節慶期間仍然是按照日期來排序。

在維度設計方案四中，包含有節慶類別、年份與節慶期間三個層級，節慶類別以節慶類別代碼順序排列，並在內容開頭加入代碼，讓使用者較容易記憶常使用的節慶類別，另外，將年份置於第二層級，使得第三層級的節慶期間按日期排序能避開第 4.1.3 節所敘述年份的邏輯混亂現象。

表 4.14 節慶四方案設計方式比較表當中，較清楚整理了節慶四個方案在資料表、欄位以及層級與排序的設計方式。

表 4.14 節慶四方案設計方式比較表

節慶設計方案	資料表	欄位	線上呈現方式	
			層級	排序
方案一	節慶資料表	日期	節慶期間	按日期
		節慶名稱		
方案二	節慶資料表	日期	節慶類別	按日期
		節慶名稱		
		節慶類別	節慶期間	按日期
方案三	節慶範圍資料表	日期	節慶類別 節慶期間	按節慶類別代碼 按日期
		節慶名稱		
		節慶類別代碼		
	節慶類別資料表	節慶類別代碼		
節慶類別				
方案四	節慶範圍資料表	日期	節慶類別 年份 節慶期間	按節慶類別代碼 按日期 按日期
		節慶名稱		
		節慶類別代碼		
		年份		
	節慶類別資料表	節慶類別代碼		
節慶類別				

4.3.2 優缺點比較

維度設計方案一的設計較為簡單，只有一個層級的方式讓使用者在邏輯上較易懂，但在使用資料倉儲的高階技巧上反而較為困難。

維度設計方案二的設計以節慶類別作為第一層級，在維度介面的呈現方式較好，但依照日期排序的情況，會造成使用者邏輯混亂，如圖 4.13 方案二節慶類別排列方式虛線框住部分所示，農曆春節排序於中秋節之下，原因在這個節慶類別名稱『送神 農曆春節』，最早是在西元 1997 年產生的類別，而『中秋節』這個類別在西元 1996 年便有了，因此在日期的順序上並沒錯。而第二層級的節慶期間由於按日期排列的關係，會出現第 4.1.3 節所敘述年份的邏輯混亂現象以及第 4.1.4 節所描述節慶產生的空值問題。

拍賣量(件)	作Y值	教師節 重慶節 前六天	教師節 重慶節 前五天	教師節 重慶節 前四天	教師節 重慶節 前三天	教師節 重慶節 前二天	教師節 重慶節 前一天	教師節 重慶節 前七天
FR玫瑰	506	1094	1095	1047	1093	0	0	0
FC大菊	252	242	266	345	423	0	0	0
FH33重瓣花	22	37	100	127	135	0	0	0
FD小菊	518	451	432	632	840	0	0	0
FB火鶴花	804	1467	979	1357	1600	0	0	0
FW5滿天星	28	22	28	25	20	0	0	0
FS香水百合	424	751	657	678	734	0	0	0
FK賽百合	135	286	162	131	151	0	0	0

圖 4.13 方案二節慶類別排列方式

維度設計方案三的設計以節慶類別作為第一層級，在設計上按照節慶類別代碼排序，解決了維度設計方案二當中節慶類別的問題，但第二層級的節慶期間由於按日期排列的關係，仍然沒有改善第 4.1.3 節所敘述年份的邏輯混亂現象以及第 4.1.4 節所描述節慶產生的空值問題。

維度設計方案四的設計以節慶類別作為第一層級，在設計上按照節慶類別代碼排序，解決了維度設計方案二當中節慶類別的問題，並在內容開頭加入代碼，讓使用者較容易記憶常使用的節慶類別，第二層級以年份限制住第三層級節慶期間的範圍，也就是做下挖動作時，必須先選取年份才能往下選取該年份的節慶期間，改善第 4.1.3 節所敘述年份的邏輯混亂現象，也減少第 4.1.4 節所描述節慶產生的空值問題。

表 4.15 四方案改善情況表，呈現出四種維度設計方案，在第 4.1 節所敘述的重複性、排序、空值三種特殊節慶問題上改善的狀況，並輔以使用的友善度做比較，從表上的分析可以得知，維度設計方案四能改善的地方最多，友善度也最高，因此決定採用維度設計方案四為最終的設計方式。

表 4.15 四方案改善情況表

節慶設計方案	改善情況			
	重複性	排序	空值	友善度
方案一	無改善	無改善	無改善	低
方案二	以類別層級減少介面複雜度	無改善	無改善	低
方案三	以類別層級減少介面複雜度	類別層級有較好排序	無改善	中
方案四	以類別層級減少介面複雜度，並加上代碼幫助記憶	1.類別層級有較好排序 2.用年份層級做限制，不會出現邏輯錯亂	以年份層級做選擇上的控制，發生情況較少	高

第五章 量測變數迴歸應用

在第 2.3.2 節所述的歷史案例當中，學者 P.G.Wright 與 S.Wright 找出「降雨量」為黃油交易量的量測變數。量測變數的找尋，是朝關聯性去發想，再藉由嚴密的數學去驗證而得。「降雨量」的發想便是藉由降雨多而草生長快，草生長快因此牛吃的多，牛吃的多而黃油便生產的多，黃油量多於是價格便低。將「降雨量」作為預想的外來變數，經由量測變數迴歸的方法去計算，確認其為黃油案例的成功之量測變數。

除此之外，國外也有許多著名的案例，例如美國研究發現啤酒的銷售量為尿布銷售之量測變數。由於在美國多為男性購買嬰兒尿布，而男性在購買尿布的同時，會順道購買啤酒。因此超市會將啤酒與尿布擺在相鄰的展示架上，使男性便利購買。這兩個看似毫無相關的產品，便是藉由創意的發想而來，並獲得數學之驗證。

在學術上，量測變數迴歸的方法普遍見於電機、機械工程的研究，本論文將之應用於花卉產業的供給需求狀況，利用上述之邏輯概念找尋花卉交易量的成功之量測變數。本論文選擇以文心蘭花卉作為研究之對象，其為節慶常使用之花卉，大量種植於台灣中部與南部，圖 5.1 為文心蘭照片[26]。



圖 5.1 文心蘭照片

本章第 5.1 節主要說明量測變數迴歸的分析流程。第 5.2 節介紹資料的收集與分析。第 5.3 節探討量測變數迴歸的成果。

5.1 分析流程

本研究在資料找尋過程中，採用試誤法不斷的找尋合適之資料。首先，考慮資料變數屬於統計時變程序時可能造成的假性迴歸問題。因此在進行迴歸分析之前，採用單位根檢定的方法，先判定資料變數是否具統計非時變程序特性。資料滿足統計非時變程序特性，才能進行迴歸分析。

當尋找到一個擬合度的迴歸方程式後，必須去檢驗其資料是否符合高斯馬可夫定理的五大假設。本研究特別著重於去解決資料不滿足高斯馬可夫定理的第五假設之狀況，此情形仍須利用試誤法去找尋在迴歸方程式中，誤差項與自變數有高相關性之資料。

下一步驟便利用聯想去找尋預想的外來變數。當驗證出此外來變數與原始迴歸式之殘差無相關，而與自變數有高相關時，再利用二階段最小平方法建立量測變數之迴歸，最終找尋出擬合度高的迴歸方程式，確認此為成功的量測變數。本研究分析流程如圖 5.2 所示。

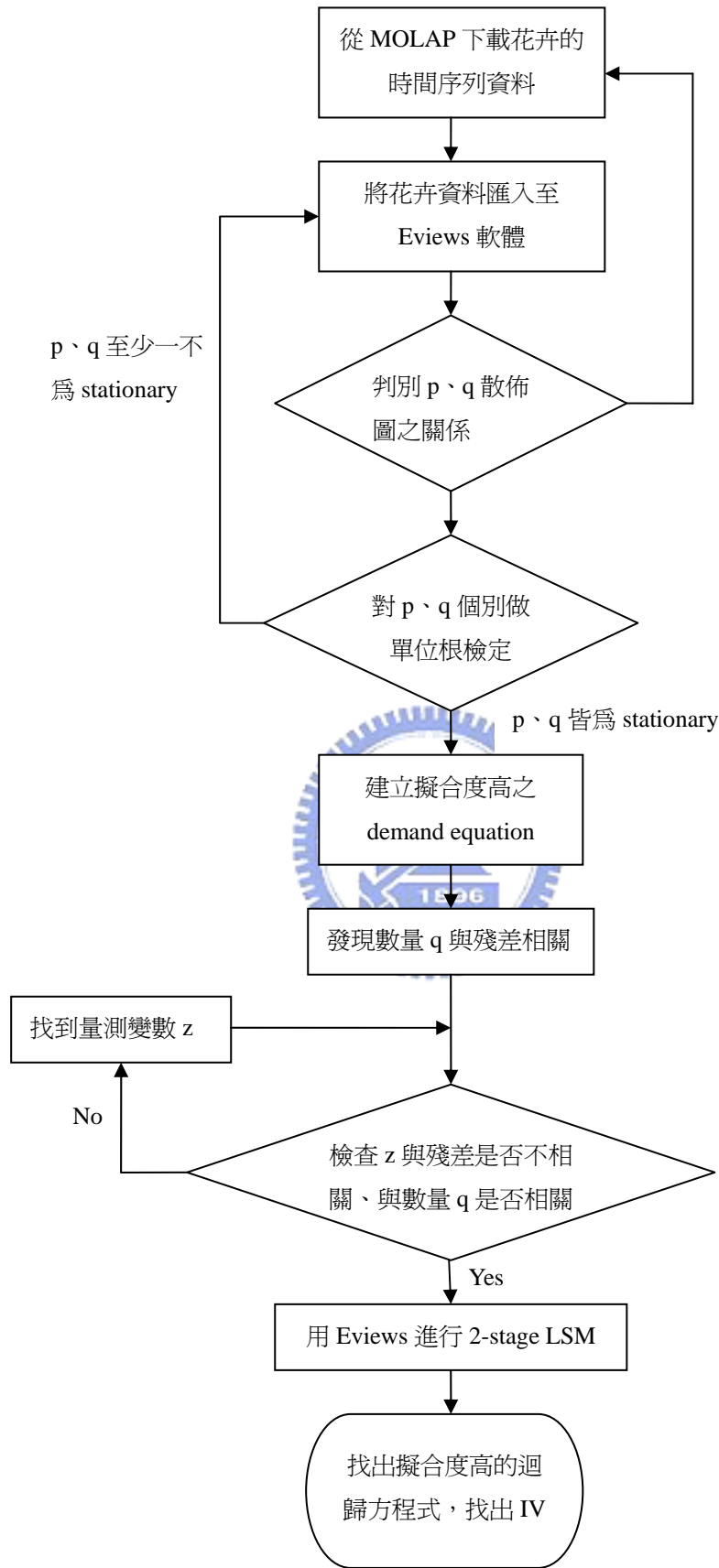


圖 5.2 量測變數迴歸技術作業流程圖

5.2 資料收集與分析

5.2.1 原始資料收集

本研究之資料收集，是透過花卉批發資訊分享熱線進行，於總資料倉儲中下載台北資料超市之花卉交易資料。針對文心蘭的總平均價與總成交量進行最小平方方法之迴歸分析，驗證文心蘭成交的總平均價與總成交量之相關性，並確認其迴歸方程式是否滿足高斯馬可夫定理的五大假設：

- (1) 常態性 (Normality)
- (2) 平均數為 0 (Zero mean)
- (3) 變異數齊一性 (Homoskedasticity)
- (4) 無自我相關性 (Nonautocorrelation)
- (5) 自變數 X 為非隨機 (Nostochastic X)

當誤差項與自變數有相關性時，最小平方方法違反高斯馬可夫定理第五項假設，將不滿足一致性。必須找尋量測變數，對其進行二階段最小平方方法之量測變數迴歸技術。

在本研究中，所需資料數據愈多進行迴歸之效果欲佳。考慮到花卉批發資訊分享熱線的建置時間，因此在資料的分析時間區間長度上，將資料分析時間長度設定為 2000 年 1 月至 2005 年 12 月，共 72 筆資料。而對於分析的資料方面，則是蒐集台北資料超市文心蘭的月總平均價及月總成交量時間序列資料進行分析。分析所使用的資料分析時間區間長度，如表 5.1 中所示。圖 5.3 為文心蘭總平均價對總成交量之散佈圖，可以初步觀察出其總平均價與總成交量間呈現相關性，當成交量愈高，總平均價也就愈低。

表 5.1 資料分析時間長度

市場別	花卉種類	資料類別	分析時間	資料筆數
台北花卉批發市場	文心蘭	總成交量	2000 年 1 月至 2005 年 12 月	72 筆
		總平均價	2000 年 1 月至 2005 年 12 月	72 筆

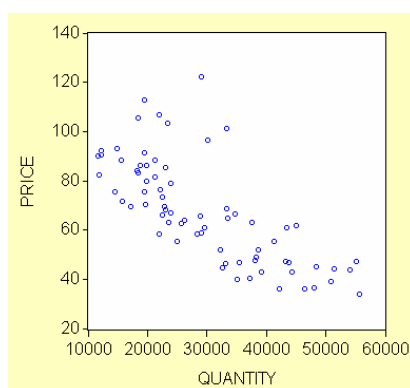


圖 5.3 文心蘭總平均價對總成交量之散佈圖

5.2.2 單位根檢定

藉由第 2.3.4 節中時間序列資料的單位根檢定，可以判斷一組時間序列是否具有統計非時變程序特性。本節將對文心蘭的總成交量與總平均價資料，自 2000 年 1 月至 2005 年 12 月做統計時變程序特性的探討。

在進行單位根檢定之前，必須對時間序列資料進行起始的檢定模型假設與分析，也就是必須先觀察時間序列資料的趨勢，來選擇假設的模型，以及最適落後期數。圖 5.4 為文心蘭的總成交量與總平均價之時間序列走勢圖。

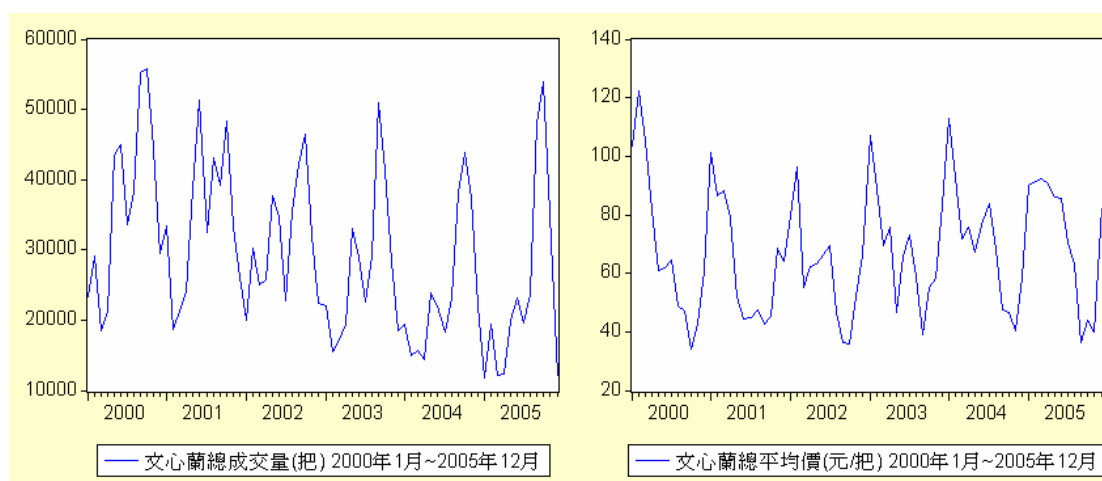


圖 5.4 文心蘭的總成交量與總平均價之時間序列走勢圖

由圖 5.3 中可看出，文心蘭的總成交量與總平均價兩組時間序列資料，皆無顯著的遞增或遞減趨勢。因此在本研究中，對於資料特性的單位根檢定，採第 2.3.4 節中擴大迪-富氏檢定法的模型一，假設文心蘭的總成交量與總平均價兩組時間序列資料皆為無截距項及趨勢項的隨機漫步模式。

除了檢定模型的選擇外，在進行時間序列資料的單位根檢定前，還需決定資料的最適落後期數。當選取資料的落後期數不足時，將可能造成模型的殘差項存在序列相關，造成估計偏誤；而過多的落後期數，則可能導致過度參數化。而富樂等[27]認為選取資料的落後期數不足，所造成估計偏誤問題，遠較因選取過多的落後期數而導致過度參數化的問題嚴重。因此建議採用赤池資訊準則值，做為最適落差期數選取準則。在本研究中，最適落後期數的選取，是採用赤池資訊準則值法做為最適落後期數的選取準則。

單位根檢定的虛無假設(H_0)為該時間序列資料存在單位根，亦即其具有統計時變程序特性；而對立假設(H_1)則為該時間序列資料不存在單位根，也就是該資料具統計非時變程序特性。表 5.2 為文心蘭的總成交量與總平均價資料的單位根檢定結果。

表 5.2 文心蘭的總成交量與總平均價資料之單位根檢定

分析資料 時間區段	變數	落後期數	t 檢定值	5%臨界值
文 心 蘭 2000.1~2005.12	總成交量	11	-2.3267	-1.95
	總平均價	9	-7.3682	-1.95

在表 5.2 中可看出，文心蘭的總成交量與總平均價兩組時間序列資料，在 5% 的顯著水準下，皆拒絕單位根檢定的虛無假設。因此可以判定該兩組時間序列資料皆呈現統計非時變程序特性。

5.2.3 迴歸方程式分析

經由 5.2.2 節確認文心蘭的總成交量與總平均價兩組時間序列資料，皆呈現統計非時變程序特性，皆下來便建立此兩組時間序列的迴歸方程式。y 代表文心蘭的總成交量(把)，x 代表文心蘭的總平均價(元/把)。下列式子便是文心蘭的需求方程式，對變數取 log 是為了要消除非線性關係。

$$\log(y) = -0.597185 \log(x) + 10.2647$$

文心蘭需求方程式的分析結果如表 5.3 所示。方程式的變數取 log 可消除非線性關係。R-square 值達 0.587867，表示文心蘭的總成交量與總平均價之間相關性高。而獨立變數 log(x) 的標準差為 0.059764，常數項標準差為 0.611052，相對於係數之下非常小，顯示出該迴歸方程式之準確性高。

表 5.2 文心蘭需求方程式之分析結果

相依變數：log(y)				
方 法：最小平方法				
樣 本：2000 年 1 月至 2005 年 12 月				
觀 察 值：72 筆				
變數	係數	標準差	t 檢定值	p 值
log(x)	-0.597185	0.059764	-9.992396	0.0000
c	10.26470	0.611052	16.79839	0.0000
R-square ：0.587867				

但文心蘭的迴歸方程式尚為確認其是否滿足高斯馬可夫定理第五項假設，也就是必須確認獨立變數與殘差是否具相關性。若為相關表示文心蘭之迴歸方程式無法作為合理之需求方程式。圖 5.5 為獨立變數與殘差之散佈圖，文心蘭的總平均價代表獨立變數 X，殘差代表 E。

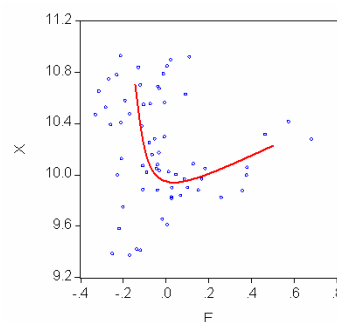


圖 5.5 獨立變數與殘差之散佈圖

從圖 5.5 中可以發現其間存在二次相關，表示文心蘭的需求函數存在內生關係。亦即獨立變數 x 為隨機變數，用來估計相依變數 y 有不確定性。此時可採用量測變數迴歸技術，藉由二階段最小平方法降低不確定性，以解決此一問題。



5.3 量測變數迴歸

量測變數之找尋必須考慮其關聯性，在本研究中尤其需要對文心蘭花卉的特性清楚，才能提高找尋量測變數之效率。而找出預想之外來變數，僅僅並表示其有可能作為量測變數，還須待進一步的數學驗證後，才能確認其是否適合作為本研究之成功的量測變數。

5.3.1 外來變數資料分析

經由花卉專家長期觀察文心蘭之產量分配[1]，可知文心蘭的產出有明顯的季節性，9月至11月為高峰期，5月至6月為次高峰，高峰期3個月之產出約可占全年之四成左右，而產期集中之結果通常會快速反應在價格之變化上。圖 5.6 所示為文心蘭 2000 年至 2005 年每月之平均價格與產量折線圖，橫軸代表每年之月份，左邊縱軸單位為文心蘭總平均價(元)，右邊縱軸為文心蘭產量(件)，線段上有點的代表文心蘭每月年產量，線段平整的代表文心蘭每月總平均價，每月線段的橫線為其月平均。從圖 5.6 可以觀察到在 9 月至 11 月之平均價格最低，但產量最高，其他月份亦有互補之情形。因此可以推論產量是影響文心蘭價格的主因。

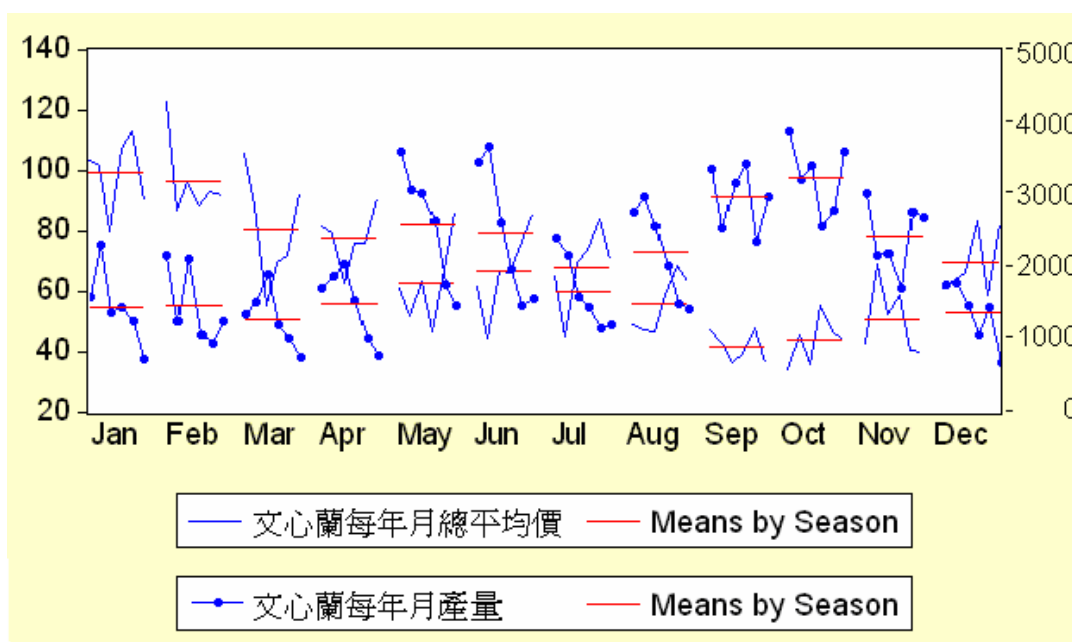


圖 5.6 文心蘭總平均價與產量 2000 年至 2005 年月平均圖

經由文獻的整理[13]發現，溫度普遍被認為是影響花期與品質的重要因素，文心蘭開花隨溫度之降低所需生育期愈長。目前文心蘭的主要產地為台中縣與屏東縣[1]，切花內外銷比例約為 49：51，內外銷比率有地區性之差異，中部幾乎以外銷為主，而內銷貨源則主要來自南部。因此本論文決定採用屏東縣恆春觀測站每月的氣溫平均資料[2]來作為預想之量測變數。

此外，由於文心蘭花序發育日期在 60 日~75 日之間[13]，因此將氣溫數據往前收集兩個月，得到資料如表 5.3 所示。

表 5.3 屏東縣月均溫

日期	平均氣溫	日期	平均氣溫	日期	平均氣溫
1999 年 11 月	24.1	2001 年 11 月	23.5	2003 年 11 月	24.8
1999 年 12 月	21.3	2001 年 12 月	21.6	2003 年 12 月	21.3
2000 年 01 月	21.1	2002 年 01 月	20.6	2004 年 01 月	20.4
2000 年 02 月	21.2	2002 年 02 月	20.7	2004 年 02 月	21.2
2000 年 03 月	23.3	2002 年 03 月	23.7	2004 年 03 月	22.9
2000 年 04 月	25.5	2002 年 04 月	26.4	2004 年 04 月	25.1
2000 年 05 月	26.8	2002 年 05 月	27.4	2004 年 05 月	28.2
2000 年 06 月	28.3	2002 年 06 月	28.5	2004 年 06 月	28.1
2000 年 07 月	27.7	2002 年 07 月	28.7	2004 年 07 月	27.7
2000 年 08 月	27.7	2002 年 08 月	28.4	2004 年 08 月	28.6
2000 年 09 月	27.2	2002 年 09 月	27	2004 年 09 月	27.3
2000 年 10 月	27	2002 年 10 月	26.9	2004 年 10 月	25.3
2000 年 11 月	25	2002 年 11 月	24.2	2004 年 11 月	24.5
2000 年 12 月	22.9	2002 年 12 月	22.5	2004 年 12 月	22.1
2001 年 01 月	21.5	2003 年 01 月	20	2005 年 01 月	19.8
2001 年 02 月	22	2003 年 02 月	21.9	2005 年 02 月	21.8
2001 年 03 月	23	2003 年 03 月	22.7	2005 年 03 月	21.1
2001 年 04 月	25.6	2003 年 04 月	26	2005 年 04 月	25.3
2001 年 05 月	27.2	2003 年 05 月	27.4	2005 年 05 月	27.6
2001 年 06 月	28.1	2003 年 06 月	27.3	2005 年 06 月	27.5
2001 年 07 月	28.1	2003 年 07 月	29.4	2005 年 07 月	28.5
2001 年 08 月	28.9	2003 年 08 月	28.3	2005 年 08 月	28.2
2001 年 09 月	26.6	2003 年 09 月	28	2005 年 09 月	27.9
2001 年 10 月	26.1	2003 年 10 月	26	2005 年 10 月	26.8

5.3.2 量測變數迴歸之測定

將屏東縣月均溫設為量測變數，其與獨立變數文心蘭的總平均價，相關性如圖 5.7 所示，設此屏東縣月均溫為量測變數數列 Z ，文心蘭的總平均價為獨立變數 X ，可觀察到其成正相關趨勢，而取 $\log(x)$ 與 z 之迴歸式，可發現其 R-squared 為 0.402561，代表 $\log(x)$ 與 z 有高相關性，因此可確認其達成量測變數兩大性質之一的量測相關性。細部數據如表 5.4 量測變數與獨立變數相關性之分析結果所示。

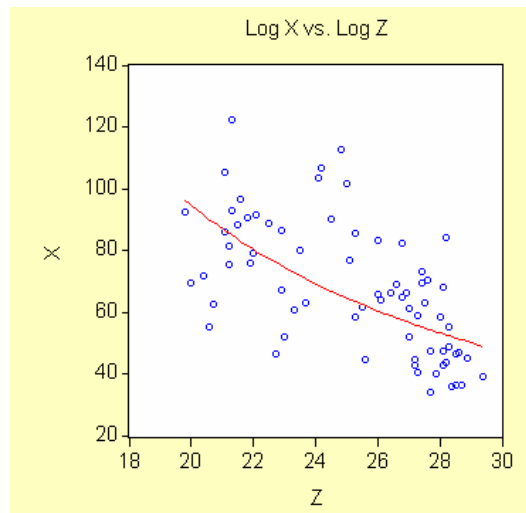


圖 5.7 量測變數對獨立變數散佈圖

表 5.4 量測變數與獨立變數相關性之分析結果

相依變數：z				
方 法：最小平方法				
樣 本：2000 年 1 月至 2005 年 12 月				
觀 察 值：72 筆				
變數	係數	標準差	t 檢定值	p 值
log(x)	-5.633842	0.820326	-6.867805	0.0000
c	48.66974	3.425372	14.20860	0.0000
R-square ：0.402561				

而圖 5.8 為量測變數 Z 與殘差 E 之散佈圖，經由觀察很明顯的確認其相關性甚低。可確認其達成量測變數兩大性質之一的量測外部性。細部數據如表 5.5 量測變數與殘差相關性之分析結果所示。

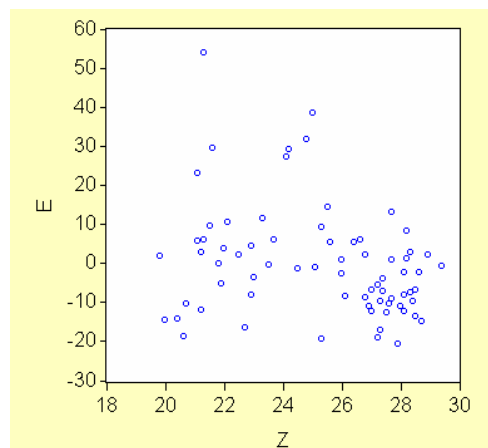


圖 5.8 量測變數 Z 與殘差 E 之散佈圖

表 5.5 量測變數與殘差相關性之分析結果

相依變數：Z				
方 法：最小平方法				
樣 本：2000 年 1 月至 2005 年 12 月				
觀 察 值：72 筆				
變數	係數	標準差	t 檢定值	p 值
E	-0.054691	0.022860	-2.392469	0.0194
c	25.21250	0.322638	78.14495	0.0000
R-square ：0.075589				

經由量測相關性與量測外部性的達成，接著利用兩階段最小平方法進行迴歸，得到數據如表 5.6 兩階段最小平方法之分析結果。經由兩階段最小平方法重估計的迴歸方程式如下式子所示，y 代表文心蘭的總成交量(把)，x 代表文心蘭的總平均價(元/把)。

$$\log(y) = -1.198876 \log(x) + 15.20804$$

表 5.6 兩階段最小平方法之分析結果

相依變數：log(y)				
方 法：兩階段最小平方法				
樣 本：2000 年 1 月至 2005 年 12 月				
觀 察 值：72 筆				
量測變數：z				
變數	係數	標準差	t 檢定值	p 值
log(x)	-1.198876	0.160440	-7.472441	0.0000
c	15.20804	0.668787	22.73975	0.0000
R-square ：0.559960				

R-square 值達 0.559960，表示經由量測變數降低文心蘭總平均價的不確定性之後，該方程式相關性高。而獨立變數 log(x)的標準差為 0.160440，常數項標準差為 0.668787，相對於係數之下非常小，顯示出該迴歸方程式之準確性高。可以說明以屏東縣月均溫足以做文心蘭總平均價的量測變數。

5.3.3 結果分析

本次所找出的量測變數用來產生迴歸方程式的相關係數為 0.559960，經二階段最小平方法做出來的效果顯示出屏東縣月均溫與文心蘭總平均價的密切程度，我們可以說以氣溫做為量測變數確有其重要性，或許在後續的研究之中可以對氣溫的指標再做進一步的探討。

5.4 資料倉儲節省時間之估算

由圖 5.1 的流程來看，整個量測變數迴歸的資料分析時間很長。尤其採用試誤法來選取資料，下載資料將耗費相當多的時間。本研究使用總資料倉儲作為資料來源，可同時比較各家花市之資料，與尚未建置總資料倉儲之前相比，可以節省許多時間。

以下載台北、台中、彰化、台南各家花市的文心蘭月交易資料為例，使用總資料倉儲下載資料的程序，需先登入帳號密碼，顯示隱藏需要的衡量值，將時間維度由年下挖至季，季下挖至月，再顯示隱藏所需的月範圍，然後嵌套市場別維度，最後下載資料於硬碟中。而使用各家資料超市來下載文心蘭的月交易資料，除了不需進行嵌套市場別維度這項動作外，所有動作與總資料倉儲相同，但每一家資料超市就要進行一次程序，因此共需重複四次程序。表 5.7 比較出使用總資料倉儲與各家資料超市為資料來源，所需進行的線上分析處理動作次數。

表 5.7 總資料倉儲線上分析處理動作表

資料來源	登入次數	下挖次數	顯示隱藏次數	嵌套次數	下載次數
總資料倉儲	1	2	2	1	1
各家資料超市	4	8	8	0	4

表 5.8 紀錄每個線上交易處理動作的所需時間，由表 5.8 來看總資料倉儲多出來的動作，嵌套只需花 19 秒，而整個總時間比使用各家資料超市下載資料少了 11 分 14 秒。此程序時間的計算僅包含線上交易處理動作，並沒考慮到網路處理的回應時間，因此若記考量整體所需時間的話，總資料倉儲能節省的時間將更多。本例子只以下載文心蘭花卉為例，實際上試誤法所下載的花卉資料包含所有其他花卉種類，所需花費之時間更多。由此可知，總資料倉儲的建立，能明顯提升資料分析之效率。

表 5.8 總資料倉儲節省時間表

資料來源	登入時間	下挖時間	顯示隱藏時間	嵌套時間	下載時間	總時間
總資料倉儲	1 分 8 秒	36 秒	1 分 51 秒	19 秒	16 秒	4 分 10 秒
各家資料超市	4 分 32 秒	2 分 24 秒	7 分 24 秒	0 秒	1 分 4 秒	15 分 24 秒

第六章 結論與未來研究方向

本論文的研究主要分為三個部分，第一、總資料倉儲的完成，採用資料轉換服務，成功整合了台北、台中、彰化與台南四個資料超市。第二、於總資料倉儲新建一個節慶維度，內容包含國曆與農曆節慶，使用者將可以直接查詢節慶期間的各類花卉交易量。第三、以總資料倉儲中的花卉資料為來源，應用量測變數迴歸技術來做資料挖礦的研究，提供花卉運銷的決策。本章第 6.1 節主要說明本研究完成後所帶來之效益。第 6.2 節提供未來的研究方向。

6.1 結論

在現代資訊發展快速的社會，人類的生活步調加快，凡事講求效率。利用高科技的資訊技術，人們可以迅速的獲取正確的資訊，以及在大量資料中挖掘出有價值的訊息。如何提升資訊的獲取效率，正是目前不可或缺的重要課題。

過去在總資料倉儲未建立以前，使用者若要分析不同花卉供應鏈之間的互動情形，必須先取得各家花卉資料超市的權限，分別至不同的資料超市蒐集資料後，才能進行資料的分析、估計與統計上的假設檢定。若是在處理資料的過程中，有資料收集不齊全或是無法通過檢定的情形，整個分析流程就要重新來過，直到滿足需求為止；因此使用者在分析不同花卉供應鏈的互動情形是非常耗費時間的。總資料倉儲在建置好之後，各個花卉供應鏈的資料可直接一次經由總資料倉儲的市場維度中下載；這使得整個資料分析作業的時間大為縮短。

另外，過去在節慶維度尚未建立之前，使用者欲分析各花卉供應鏈在節慶期間的行為時，必須透過人工查詢節慶日期，才能藉由時間維度觀察到資料；並且由於每年的節慶日期並不完全相同，有國曆節慶、農曆節慶之分，使得資料蒐集的工作更為複雜與困難。因此節慶維度的完成，節省了使用者更多的分析時間。總而言之，花卉總資料倉儲的完成，大量縮減了資料收集與分析所耗掉的時間，更加提升了使用者工作的效率。

在總資料倉儲完成後，資料的收集變得更加的容易。因此本論文透過總資料倉儲系統進行資料的蒐集，做一個資料挖掘的應用。經由大量資料的蒐集與統計假設檢定，找出文心蘭此一花卉在供給需求上，有不確定性的因素存在；透過量測變數迴歸技術的應用，發現氣溫可作為文心蘭在總成交量的量測變數。其中的關係在於氣溫影響文心蘭之產量，產量高則銷售量高，因此成交量也高。此一發現提供花卉運銷決策者，氣溫可作為文心蘭價格之預測來源。

6.2 未來研究方向

在未來的研究的方面，可朝五個方向去做研究：第一是銜接高雄資料超市於總資料倉儲，第二是改善節慶維度，第三是探討如何完善管理與維護總資料倉儲，第四是量測變數的再研究，第五是建立決策支援系統。以下為五點方向的描述。

一、銜接高雄資料超市於總資料倉儲

總資料倉儲的完成，成功地整合了台北、台中、彰化與台南四家資料超市，目前只剩高雄資料超市尚未建置完成，待完成後才能將之加入總資料倉儲的系統中。建議高雄資料超市在建立時，一併考慮總資料倉儲之銜接。將維度設計之欄位名稱，配合總資料倉儲已整合完成之維度欄位，保持一致性可加快高雄資料超市加入總資料倉儲之速度。

二、改善節慶維度

總資料倉儲的節慶維度為世界首創，雖克服了許多設計上的問題，但由於每年節慶日期可能不同，因此必須以人工方式建立節慶範圍資料表。以人工方式來建立節慶範圍資料表需耗費大量時間，且每年必須增加節慶資料量，因此維護需要花較多的心力。

未來若有後續的研究者，可考慮將國曆節慶與農曆節慶分開建立。由於國曆節慶每年有固定的日期，因此可建立為制式的資料表，不需每年更新；而農曆節慶若以農曆維度來連結，或許也可建立為制式的資料表。此方式可能存在其他建立上複雜的困難點，還需後進詳加研究。

三、探討如何完善管理與維護總資料倉儲

總資料倉儲是藉由轉換各家資料超市的拍賣資料表，來進行資料的轉換。因此下端的資料超市發生問題時，總資料倉儲便會受到波及；更由於各家資料超市的資料處理方式不同，因此衍生出許多繁雜的小問題。例如總資料倉儲與各家資料超市為不同管理者負責維護，當資料轉換中出錯時，必須有清楚的權責區分以及有效率的通報機制，才能縮短總資料倉儲以及資料超市當掉的時間。

再來總資料倉儲的管理者必須熟知各家資料超市的系統，未來可建立維修手冊，產生維修的機制，使未來的管理者容易接手總資料倉儲的管理。

四、量測變數再研究

目前本研究找出氣溫可作為文心蘭的量測變數，經由文心蘭生長探討文獻中，亦發現其生長與氣候的濕度亦有關，後續研究者可朝此一方向探討。

五、建立決策支援系統

目前在利用花卉資料倉儲進行資料挖礦的時候，仍不可避免必須利用各種軟體來作分析，若建立一套決策支援系統，將可縮短資料分析的時間。

參考文獻

- [1] 文心蘭花卉產銷班個案研究(一),
<http://ecaaser3.ecaa.ntu.edu.tw/weifang/lab551/flower/%A4%E5%A4%DF%C4%F5%AA%E1%A5c.htm> 柯珮婕、梁高榮,「台中花市的花卉交易資料超市設計」,機械工業,五月,84-95頁,2005。
- [2] 中央氣象局,<http://www.cwb.gov.tw/>
- [3] 柯珮婕,「花卉批發資訊分享熱線的設計與實作-以台中花市為例」,國立交通大學工業工程研究所碩士論文,2005。
- [4] 柯珮婕、梁高榮,「台中花市的花卉交易資料超市設計」,機械工業,五月,84-95頁,2005。
- [5] 郭軒豪、賴俊翰、梁高榮、江東陽,「網路型三階正規化切花/盆花資料庫的建立」,機械工業,十二月,218-235頁,2002。
- [6] 郭軒豪,「利用資料倉儲和 J2ME 技術設計花卉供應鏈系統的行動決策支援裝置」,國立交通大學工業工程研究所碩士論文,2003。
- [7] 陳楓凱,「高效能批發資訊分享熱線的建構」,國立交通大學工業工程研究所碩士論文,2004。
- [8] 陳楓凱、溫師翰、梁高榮、鍾國成,「彰化花市的花卉交易資料倉儲建構」,資訊科技在農業上之應用研討會,38-43頁,2003。
- [9] 陳家瑜、梁高榮、陳吉成,「花卉資料倉儲節慶維度與資料轉換服務的設計」,機械工業,四月,132-144頁,2006。
- [10] 陳佳佑,「花卉資料倉儲的三種改進方案: 權限控管、審計資訊與季節性資料分析」,國立交通大學工業工程研究所碩士論文,2006。
- [11] 曾世民、郭軒豪、梁高榮,「花卉供應鏈裡的資料倉儲技術介紹」,機械工業,六月,256-273頁,2003。
- [12] 黃彥修,「台南花市的資料倉儲建構及其共整合分析」,國立交通大學工業工程研究所碩士論文,2005。
- [13] 蘭花量產工程-環境溫度與文心蘭南茜開花品質,
http://bse.nchu.edu.tw/new_page_88.htm
- [14] Breslin, M., "Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models," Business Intelligence Journal, pp. 45-48, Spring 2004.
- [15] Dicky, D. A. and Fuller, W. A., "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," Econometrica, Vol. 49, pp. 1057-1072, 1981.
- [16] Dicky, D. A. and Fuller, W. A., "Distribution of the Estimations for Autoregressive Time Series with a Unit Root," Journal of the American Statistical Association, Vol. 74, pp. 427-431, 1979.
- [17] Gray, P. and Watson, H., Decision Support in the Data Warehouse, Prentice Hall, 1998.
- [18] Granger, C. W. J. and Newbold, P., "Spurious Regression in Econometrics," Journal of Econometrics, 2, pp. 111-120, 1974.
- [19] Harvey, A. C., The Econometric Analysis of Time Series, 1991.
- [20] Han, J. and Kamber, M., Data Mining: Concepts and Techniques, Academic Press, 2001. Codd, E. F., Codd, S. B., and Salley, C. T., "Providing OLAP to User Analysts: an IT Mandate," <http://www.arborsoft.com>, 1993.
- [21] Humberto Barreto and Frank M. Howland, Introductory Econometrics, 2005.
- [22] Inmon, W. H., Building the Data Warehouse, Addition-Wesley, 1992.

- [23] Kimball,R., “A Dimensional Modeling Manifesto,”DBMS, pp. 58-70, August 1997.
- [24] Kimball, R. and Ross. M., The Data Warehouse Toolkit : The Complete Guide to Dimensional Modeling (2nd Edition), John Wiley & Sons, 2000.
- [25] Kimball, R., ”It’s Time for Time,” DBMS, pp16-18, August 1997.
- [26] Oncidium splendidum, http://eorchids.org/orchids/onc_splendrdom.htm
- [27] Pantula, S. G., Gonzales-Farias G. and W. A. Fuller, “A Comparison of Unit-Root Test Criteria,” Journal of Business and Economic Statistics, Vol. 12, NO.4, pp. 449-459, 1994.
- [28] Wright, P.G., The Tariff on Animal and Vegetable Oils., New York: The Macmillan Company, 1928.

