

國 立 交 通 大 學

工 業 工 程 與 管 理 學 系

碩士論文

花卉產業的第二次台北資料超市再設計

以解決其維度緩慢改變問題

Secondary Redesign of Taipei Data Mart for Solving Its Slowly

Changing Dimension Problem in Floral Industry

1896

研究生：方本欣

指導教授：梁高榮 博士

中華民國一百零一年六月

花卉產業的第二次台北資料超市再設計

以解決其維度緩慢改變問題

Secondary Redesign of Taipei Data Mart for Solving Its Slowly  
Changing Dimension Problem in Floral Industry



A Thesis Submitted to  
Department of Industrial Engineering and Management  
College of Management

National Chiao Tung University

In Partial Fulfillment of the Requirements  
For the Degree of Master of Engineering in  
Industrial Engineering and Management

2012

Hsinchu, Taiwan, Republic of China

# 花卉產業的第二次台北資料超市再設計以解決其維度緩慢改變問題

研究生：方本欣

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

維度緩慢改變問題是指在多維度資料超市或倉儲中，部份維度資料的意義有緩慢改變現象。本論文針對花卉產業中台北花市的維度緩慢改變問題進行研究。由於花卉資料的意義源自變動快速的花卉產業環境，研究結果顯示維度緩慢改變問題將必然存在，且應持續投入維護成本以修正此問題；而不同維度的改變速度也不盡相同，改變速度越慢的維度將越難察覺此問題。為解決現階段發現的特定緩慢改變維度問題，本論文利用錯誤的統計資料為基礎，對資料提出系統性判斷與修正方法。為了證明此方法的可行性，利用 SQL Server 整合服務實作 ETL 系統。

關鍵字：

花卉產業(Floral Industry)

緩慢改變維度問題(Slowly Changing Dimension Problem)

資料超市再設計(Data Mart Redesign)

萃取－轉換－載入(Extract-Transform-Load)

SQL Server 整合服務(SQL Server Integration Services)

Secondly Redesign of Data Marts for Solving  
Its Slowly Changing Dimension Problem in Taiwan Floral Industry

Student: Pen-Hsin Fang

Advisor: Dr. Gau-Rong Liang

Department of Industrial Engineering and Management  
National Chiao-Tung University

英文摘要

## **Abstract**

Slowly Changing Dimension (SCD) problem in a multi-dimensional data mart or warehouse refers to the phenomenon that some dimensions owns data with a slowly changing meaning. In this thesis, the SCD problem of floral industry in Taipei has been studied. Since the meaning of floral data evolves with the dynamic development of the floral industry, the research results show the inevitable of the SCD problem and the necessity of a persistent maintenance cost for fixing its dimensions. Moreover, each dimension has its own speed to change the inherent meaning, and a more SCD is harder to be found than a faster one. Accordingly, for solving the specific SCD problem we identified at the current stage, a systematic diagnosis method based on the faulty statistic we have known is proposed to locate the troublesome dimensions. Also for showing the feasibility of this diagnosis method, an Extract-Transform-Load (ETL) software based on SQL Server Integration Services (SSIS) has been coded.

**Keywords:**

Floral Industry

Slowly Changing Dimension

Data Mart Redesign

Extract-Transform-Load

SQL Server Integration Services

## 誌謝

本篇論文得以完成，首先最感謝的是我的指導教授 梁高榮博士的費心指導。在寫作過程中，不論是論文方向的調整，研究方法的探討與修正，以至整體文章的邏輯架構，皆經過老師不辭辛勞的耐心指導，使我漸漸熟悉撰寫論文的技巧。另外也特別感謝陳文智老師與王志軒老師對論文的詳加審閱，並提供許多寶貴意見，使本論文更趨完備。而在論文資料的部分，特別感謝台北花卉公司張堂穆總經理與黃綉蓉課長，提供了許多花卉拍賣的相關資訊，使論文中的數據與資料更加確實。

兩年的研究所生活中，除了學習論文的撰寫，更重要的是習得如何架構與維護資料倉儲。在此特別感謝連惠珍學姐與彭思瑜學姐傾囊相授，並留下許多重要的資料以供參考。此外也相當感謝范植宇學長、劉思宇學長、鄭仲元學長及朱明典學長的課業指導，使我得以順利的修習老師開授的課程。

最後感謝陪伴我渡過研究所生涯的同學：辜婉琪、張舜龍、徐宏智、張弘與林哲慧，與你們一起在實驗室打拼的日子，將會是這兩年最鮮明的回憶。其中特別感謝張弘與我討論 SSIS 的實作技巧，對於論文完成有很大的幫助。另外也要感謝學妹王敏與學弟陳書皓，你們為實驗室帶來許多歡笑，讓這段光陰更加難忘。

謹以本論文獻給親愛的父母、妹妹以及陪伴我度過這段歲月的老師和朋友，感謝你們的一切鼓勵與呵護。



# 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
<b>第一章 緒論 .....</b>	<b>1</b>
1.1 研究動機 .....	1
1.2 問題界定及研究目的 .....	3
1.3 研究方法及論文架構 .....	5
<b>第二章 文獻探討 .....</b>	<b>6</b>
2.1 資料倉儲之理論與應用 .....	6
2.1.1 線上分析處理 .....	6
2.1.2 維度資料模型 .....	7
2.2 台灣花卉資料超市 .....	9
2.2.1 資料超市維度結構 .....	9
2.2.2 資料超市維度 .....	9
2.3 維度緩慢改變問題 .....	11
2.3.1 緩慢改變的維度 .....	11
2.3.2 第一次資料超市維度再設計 .....	12
2.3.3 花卉倉儲其他潛在緩慢改變維度 .....	13
2.4 SQL Server 整合服務 .....	15
2.4.1 SSIS 的發展與其內容 .....	15
2.4.2 SSIS 與 DTS 資料轉換內容與比較 .....	16
<b>第三章 資料超市故障分析 .....</b>	<b>20</b>
3.1 資料超市故障來源分析 .....	20
3.2 供應單位維度故障問題分析 .....	21
3.2.1 供應單位不連續問題 .....	21
3.2.2 供應單位品牌辨識問題 .....	22
<b>第四章 供應單位維度的再設計 .....</b>	<b>24</b>
4.1 供應單位不連續問題再設計 .....	24
4.2 供應單位品牌辨識問題再設計 .....	26
4.3 使用 SQL Server 2005 再設計維度資料模型 .....	28
<b>第五章 花卉資料超市 ETL 系統再設計與實作 .....</b>	<b>30</b>
5.1 SSIS 元件設定 .....	30
5.1.1 SSIS 連接管理員 .....	30
5.1.2 SSIS 資料流程工作設定 .....	32

5.2	資料庫封裝再設計與實作 .....	41
5.2.1	花市原始資料 .....	41
5.2.2	資料庫封裝設計標準流程 .....	41
5.2.3	資料庫封裝實作 .....	43
5.3	資料超市封裝再設計與實作 .....	47
<b>第六章</b>	<b>結論與未來研究方向 .....</b>	<b>50</b>
6.1	結論 .....	50
6.2	未來研究方向 .....	51
<b>參考文獻</b>	<b>52</b>	
<b>附錄 A.1</b>	<b>: 修正後資料型別維度資料表 .....</b>	<b>54</b>
<b>附錄 A.2</b>	<b>: 資料庫封裝設計 .....</b>	<b>56</b>
A.2.1	資料庫封裝其餘資料轉換 .....	56
A.2.2	資料庫封裝判斷檔案指令碼 .....	59
<b>附錄 A.3</b>	<b>: 資料超市封裝設計 .....</b>	<b>60</b>



# 圖目錄

圖 1.1	農產品交易行情站與花卉資料倉儲資料對照.....	1
圖 1.2	花卉資料倉儲架構.....	3
圖 1.3	花卉資料倉儲運作流程.....	4
圖 1.5	研究方法.....	5
圖 2.1	超方體架構.....	7
圖 2.2	維度資料模型綱要圖.....	8
圖 2.3	複合鍵與替代鍵混合使用的作業流程.....	12
圖 2.4	複合鍵與替代鍵混合使用結果.....	13
圖 2.5	承銷地區維度屬性值.....	14
圖 2.6	DTS 的腳本語言設計畫面 .....	15
圖 2.7	SSIS 流程示意 .....	16
圖 2.8	DTS 資料轉換 .....	17
圖 2.9	SSIS 資料轉換 .....	18
圖 2.10	SSIS 錯誤訊息 .....	18
圖 3.1	供應單位維度資料錯誤 .....	21
圖 3.2	供應單位不連續問題 .....	22
圖 4.1	供應單位不連續問題修正 .....	24
圖 4.2	供應代碼狀態變化 .....	26
圖 4.1	花卉資料超市星狀綱要圖 .....	28
圖 5.1	SQL Server 連接設定 .....	30
圖 5.2	dBASE 連接設定 .....	31
圖 5.3	一般檔案連接管理員設定 .....	32
圖 5.4	OLE DB 設定 .....	33
圖 5.5	資料轉換設定 .....	34
圖 5.6	查閱設定 .....	35
圖 5.7	條件式分割設定 .....	36
圖 5.8	聯集全部設定 .....	37
圖 5.9	衍生的資料行設定 .....	37
圖 5.10	緩時變維度設定 .....	38
圖 5.11	OLE DB 命令設定 .....	39
圖 5.12	OLE DB 目的地設定 .....	40
圖 5.13	台北花市原始資料 .....	41
圖 5.14	維度資料轉換標準流程 .....	42
圖 5.15	事實資料轉換標準流程 .....	42
圖 5.16	標記結束品牌實作結果 .....	43
圖 5.17	供應單位轉換實作 .....	44
圖 5.18	第一組測試資料實作結果 .....	46

圖 5.19	第二組測試資料實作結果.....	46
圖 5.20	新增供應團體流程.....	47
圖 5.21	供應單位維度轉換實作.....	48
圖 5.22	供應維度轉換實作結果.....	49
圖 A.1	拍賣資料轉換.....	56
圖 A.2	承銷資料轉換.....	57
圖 A.3	花卉資料轉換.....	58
圖 A.4	判斷檔案程式碼.....	59
圖 A.5	花卉種類維度轉換.....	60
圖 A.6	拍賣事實表轉換.....	61
圖 A.7	承銷單位維度轉換.....	62



## 表目錄

表 1.1	資料倉儲錯誤因素影響範圍 .....	2
表 2.1	總倉儲及各家花市維度比較表 .....	9
表 2.2	緩慢改變維度方法二範例 .....	11
表 2.3	緩慢改變維度方法三範例 .....	11
表 2.4	維度緩慢改變性質 .....	14
表 3.1	資料超市維度性質整理 .....	20
表 3.2	供應單位資料品牌意義 .....	23
表 3.3	供應單位資料查詢結果 .....	23
表 4.1	星狀綱要與雪花綱要資料筆數比較表 .....	29
表 5.1	SSIS 資料流程工作 .....	32
表 5.2	供應單位測試資料—第一組 .....	45
表 5.3	供應單位測試資料—第二組 .....	45
表 6.1	供應單位資料表範例 .....	50
表 A.1	供應單位維度資料表 .....	54
表 A.2	花卉種類維度資料表 .....	54
表 A.3	花卉等級維度資料表 .....	54
表 A.4	拍賣線維度資料表 .....	55
表 A.5	容器維度資料表 .....	55
表 A.6	承銷地區維度資料表 .....	55

# 第一章 緒論

本章主要目的為闡述本論文的研究目的、方法與論文整體架構。以下共分成四節，1.1 節為說明研究動機，1.2 節為問題界定及研究目的，1.3 節為研究方法及論文架構。

## 1.1 研究動機

台灣花卉產業自 2002 年開始，即藉由建置地方花卉資料倉儲(Data Warehouse)[28]來降低資訊成本並幫助決策進行[4]。2002 年台北花卉資料倉儲[6]成功建立後，彰化花市資料倉儲[7]也於 2003 年完成建置，其後在 2004 完成台中及台南花市資料倉儲[2][15]。2006 年陳家瑜[9]完成台灣花卉資料倉儲的整合，上述四座區域性花卉資料倉儲改稱為資料超市(Data Mart)[28]，並同年加入高雄花卉資料超市[14]。五家資料超市與總資料倉儲，再加上線上分析處理(Online Analytical Processing, OLAP)[25]系統，即構成「批發資訊分享熱線」(Wholesale Information Sharing Hotline, WISH)[4]系統。透過批發資訊分享熱線提供的線上分析處理功能，各地的花卉交易資料得以互相交流，促進價格透明化。對於學術界而言，台灣花卉資料倉儲保存多年花卉拍賣的資料，相當適合進行實證研究；而對於產業界而言，利用這些資料進行資料挖掘、銷售趨勢分析等，可在決策時擁有更宏觀的思維，並追求利益極大化。

在各地花卉資料倉儲建置過程中，於 2004 年進行了第一次再設計(Redesign)[16]。然而近期有使用者[12]反應其中的資料不正確，影響決策進行。除了批發資訊分享熱線外，提供花卉交易資料查詢的尚有行政院農業委員會農糧署建置的農產品交易行情站[18]，其資料為官方認可，具有一定可信程度。因此將批發資訊分享熱線與農產品交易行情站[18]進行比較，如圖 1.1，可發現兩者在 2006 年的資料呈現完全一致，然而觀察 2010 年的資料，單日拍賣把數的誤差可達百把，以批發資訊分享熱線的歷年拍賣均價每把 48 元來計算，單月拍賣金額誤差達百萬之多。由此可知花卉資料倉儲的錯誤已嚴重影響數據分析的準確性。

日期	交易量	總成交量(把)	進口花	國產花	議價花卉	花卉種類
095.01.02	81812	作數值				
095.01.03	57432	2,406	79,406	0	81,812	
095.01.04	73004	1,088	56,344	0	57,432	
095.01.05	88938	1,635	71,369	0	73,004	
095.01.06	107204	2,379	86,559	0	88,938	
		2,945	104,259	0	107,204	

(a) 民國九十五年(2006)資料(左為農產品交易行情站，右為花卉資料倉儲)

日期	交易量	拍賣量(把)	進口花	國產花	花卉種類
099.01.01	42578	作數值			
099.01.02	37744	2010年01月01日	1,612	40,925	42,537
099.01.04	81646	2010年01月02日	1,139	36,575	37,714
099.01.05	48763	2010年01月04日	1,159	80,377	81,536
099.01.06	52534	2010年01月05日	920	47,778	48,698
		2010年01月06日	2,432	50,028	52,460

(b) 民國九十九年(2010)資料對照(左為農產品交易行情站，右為花卉資料倉儲)

圖 1.1 農產品交易行情站與花卉資料倉儲資料對照

台灣花卉資料倉儲採用維度資料模型(Dimensional Data Modeling)[28]建置，其中造成資料錯誤的可能因素有三個：(1)資料倉儲的維度設計錯誤，導致資料無法經由線上分析處理取得(2)用以清理及轉換資料的 ETL(Extract-Transform-Load)[30]系統運行流程錯誤，例如判斷流程不正確造成誤刪資料，或是轉換資料不完全等，進而影響倉儲內部資料的正確性(3)花市提供的原始資料有缺誤，但因批發資訊分享熱線功能複雜，維護困難，造成錯誤未修正。表 1.1 為各因素產生影響的範圍。

表1.1 資料倉儲錯誤因素影響範圍

	(1)維度設計錯誤	(2)ETL 運行流程錯誤	(3)錯誤未修正
花市提供原始資料	正確	正確	錯誤
資料倉儲內部資料	正確	錯誤	錯誤
線上分析處理呈現資料	錯誤	錯誤	錯誤

經過本研究分析發現供應單位維度屬於緩慢改變維度(Slowly Changing Dimension, SCD)[29]，而第一次再設計使此維度產生了供應單位不連續問題與供應單位品牌辨識問題，分別屬於類型(2)和類型(1)錯誤，影響線上分析處理呈現資料的正確性。本論文將針對供應單位維度進行第二次再設計以改善這些問題。因維度再設計會影響 ETL 系統的架構，因此 ETL 系統也將進行再設計，才能正確轉入資料。本論文希望透過以上再設計的過程，改善花卉資料倉儲的資料錯誤問題，以期批發資訊分享熱線能更廣泛的應用於產業界與學界。



## 1.2 問題界定及研究目的

資料倉儲作為決策支援的工具，其內部所儲存資料必須正確才有參考價值。目前台灣花卉資料倉儲因供應單位維度具有緩慢改變現象，造成錯誤資料的產生。然而台灣花卉資料倉儲採取由下而上的建構流程，故遵循由下而上的流程進行改善，才能根除問題。

台灣花卉資料倉儲的結構分為三層，如圖 1.2，由下而上為資料庫、資料超市及資料倉儲。資料庫共有五座，分別對應至五家花卉批發市場，其存放原始拍賣資料，以提供資料超市使用。資料庫設計上相對簡單，僅將供應單位資料、承銷單位資料、花卉資料及拍賣資料分為四張表格存放；另將 ETL 系統清理出來的缺漏資料分別存放於另外四張表格，共八張資料表。資料庫不存在維度設計問題，而其 ETL 系統僅用於過濾缺漏資料以及載入資料庫。

資料超市同樣有五座，其存放的資料來自資料庫，必須經過 ETL 系統轉換成適當的型態才能存入。資料超市的維度結構與資料倉儲相同，但只存放單一地方花市的資料。而資料倉儲又稱為總倉儲，儲存五家花卉批發市場匯總後的資料。

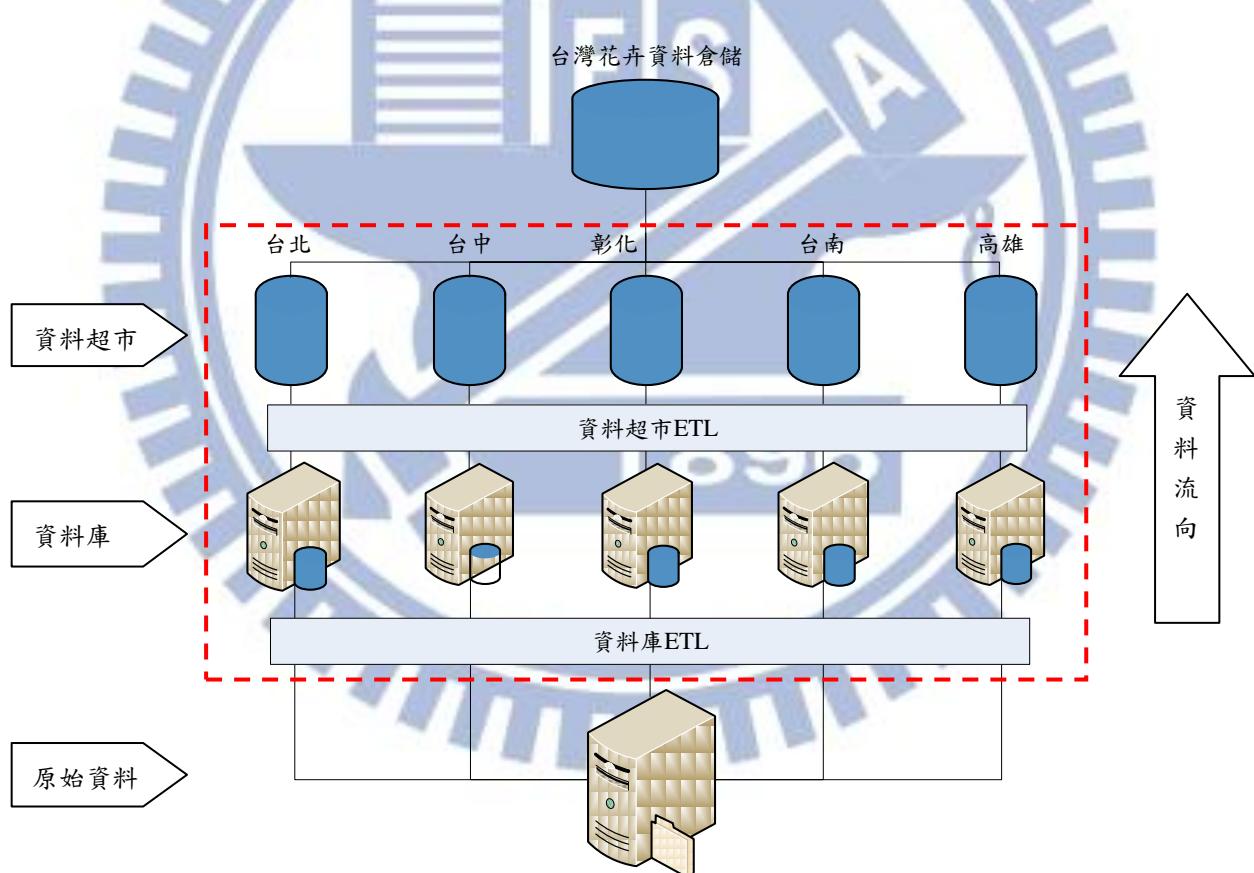


圖1.2 花卉資料倉儲架構

花卉資料倉儲資料傳輸的運作流程如圖 1.3。先將批發市場的原始資料傳送至位於交通大學的伺服器，再透過五次資料轉換服務(Data Transformation Services, DTS)[5]把資料匯入總資料倉儲。第一次資料轉換服務(DTS-1)目的為清理原始資料，使其格式統一，再匯入資料庫；第二次資料轉換服務(DTS-2)將資料庫的資料匯入資料超市的維度與事實資料表；

第三次資料轉換服務(DTS-3)處理供應人與承銷人的登入權限；第四次資料轉換服務(DTS-4)將資料超市的資料轉入總資料倉儲，完成倉儲的更新；第五次資料轉換服務(DTS-5)將資料超市及資料倉儲的審計日誌分別篩選淨化，並轉入審計資料庫之中。而上述流程利用 SQL Server 代理程式(SQL Server Agent) [5]排程，可於設定的時間自動進行作業。

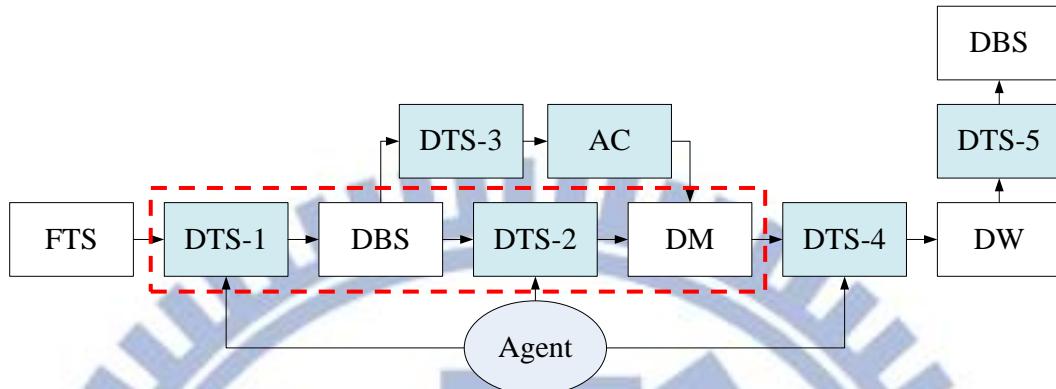


圖1.3 花卉資料倉儲運作流程

由於花卉資料倉儲的改善需由下而上進行，本論文主要針對資料超市層級的供應單位維度及原始資料至資料超市的 ETL 系統進行再設計。就資料倉儲架構而言，範圍如圖 1.2 虛線框內所示；就 ETL 系統而言，範圍為圖 1.3 的 DTS-1 與 DTS-2 包含的內容。而資料超市到總倉儲的部份，以及 DTS-4 的內容由張弘[10]完成。

本論文的研究目的在於修正花卉資料超市供應單位維度緩慢改變的問題，以確保資料超市提供正確無誤的資料給總倉儲。針對供應單位維度緩慢改變，將利用維度再設計來處理；ETL 系統的再設計部分則透過 SQL Server 整合服務(SQL Server Integration Services, SSIS)[3]進行。

## 1.3 研究方法及論文架構

本論文主要針對台灣花卉資料倉儲的資料超市部分進行再設計，可分為五個步驟，如圖 1.4 所示：

**步驟 1—問題界定與分析：**分析影響花卉資料超市內部資料正確性的可能因素，以決定如何進行改善，以及確定必須優先改善的層級。

**步驟 2—文獻回顧：**參考文獻以了解花卉資料倉儲發展，並回顧維度緩慢改變問題與資料超市第一次再設計的結果，最後評估實作方面相關技術。

**步驟 3—資料超市故障分析：**分析資料超市產生錯誤資料的原因。

**步驟 4—供應單位維度再設計：**對供應單位維度進行再設計，以解決維度緩慢改變問題。

**步驟 5—ETL 系統再設計與實作：**透過 SSIS 再設計與實作 ETL 系統的作業流程。

本論文共分為六個章節，各章節的內容大綱如下：

**第一章—緒論：**說明本論文的研究動機、問題界定、研究目的、研究方法及論文架構。

**第二章—文獻探討：**簡述資料倉儲的理論及台灣花卉資料倉儲採用的模型，分析資料超市維度之結構及其存在問題，並介紹 SQL Server 整合服務的架構。

**第三章—資料超市故障分析：**分析資料超市的故障來源，並進行故障原因的分析。

**第四章—供應單位維度的再設計：**透過維度再設計改善供應單位維度的緩慢改變問題。

**第五章—ETL 系統再設計及實作：**使用 SSIS 進行 ETL 系統再設計及實作，並與再設計後的維度進行整合。

**第六章—結論與未來方向：**整理實作再設計之結果，並給予結論。最後建議未來繼續研究的方向。

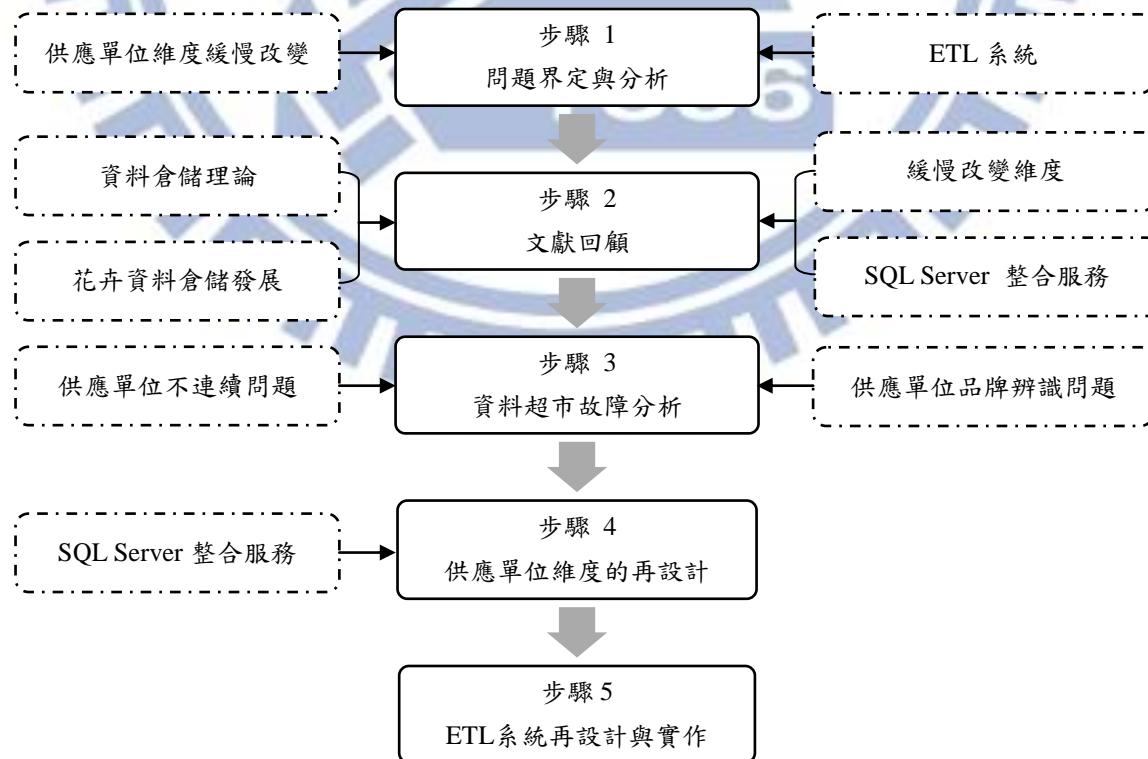


圖 1.4 研究方法

## 第二章 文獻探討

本章分為四個部分進行探討，分別是資料倉儲相關文獻、台灣花卉資料倉儲、維度緩慢改變問題及 ETL 相關文獻。資料倉儲的建置主要有兩種模式，分別是由上而下(Top-down)建置的殷默模型[27]與由下而上(Bottom-Up)建置的競簇(R. Kimball)模型[28]，由於花卉資料超市是採用由下而上方式建置，本章僅探討競簇模型的內容。以下分為四小節，2.1 節介紹線上分析處理的概念，以及競簇模型的特色；2.2 節回顧台灣花卉資料倉儲的建構以及資料超市維度分析；2.3 節說明維度緩慢改變現象及第一次再設計的方法；2.4 節說明 ETL 系統所使用的 SSIS 技術。

### 2.1 資料倉儲之理論與應用

資料倉儲是建構批發資訊分享熱線的重要技術，其概念於 1992 年由殷默(B. Inmon)[27]首先提出。殷默將資料倉儲定義為具有主題導向性(Subject-oriented)、整合性(Integrated)、非揮發性(Non-volatile)與時變性(Time-variant)四大特性的資料集合體。

1993 年，E. F. Codd 觀察到關聯式資料庫(Relational Database)[30]的結構難以滿足使用者對於整合、歸納及分析資料越來越強烈的需求[25]，因此提出了線上分析處理的概念，主要提供多維度資料分析(Multidimensional Data Analysis)的功能。而競簇(R. Kimball) [28]由維度的概念著手，於 1996 年提出了維度資料模型(Dimensional Data Modeling)。台灣花卉資料倉儲即採用此模型建構。

本節分為兩小節，於 2.1.1 節先說明線上分析處理的主要概念，再於 2.1.2 節說明維度資料模型的建構方式。

#### 2.1.1 線上分析處理

線上分析處理是一種可以直接存取資料來源如資料倉儲、資料庫等，以進行資料分析的應用程式。其可幫助使用者簡化複雜的查詢工作、製作報表、進行資料加總計算、過濾資料並分割成更具商業意義的細部資料，便於使用者進行決策。

線上分析處理最主要的概念是多維度資料分析，維度是指特定檢視資料的角度。舉例來說，假設有十筆交易資料，透過時間維度檢視，可知道這些交易是何時發生的；透過地區維度檢視，則可知道這些交易發生在何地。多維度資料分析即是同時應用多個維度以獲得特定的分析資料。

多維度資料可以超方體(Cube)[28]的形式表現，圖 2.1 就是一個由時間、供應地區及產品三個維度構成的超方體。超方體中的任何一點都是特定時間、供應地區與產品的交集，且維度中存在階層式的觀念。例如時間維度的階層由上而下可分為年、月及日三個階層，或是年、季、月、旬及日五個階層，端看使用者需要；地區維度由上而下也可分成北中南區、縣市以及鄉鎮市區三個階層。

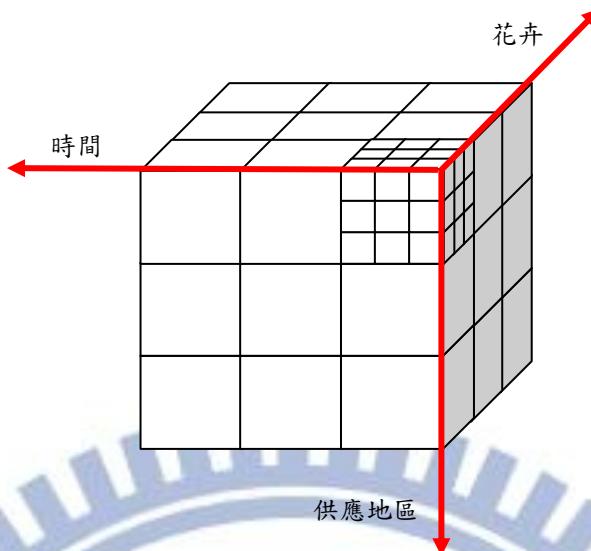


圖2.1 超方體架構

線上分析處理有十種常見的操作[4]，與維度有關的操作是轉軸(Pivoting)，把水平維度與垂直維度做交換；與層級有關的操作有上捲(Roll-Up)與下挖( Drill-Down)，其功能分別是將維度的單位放大與縮小，例如檢視的時間單位為月，可上捲至季，也可下挖至旬；與顯示範圍有關操作有切片(Slicing)及切丁(Dicing)，切片是限制單維度的範圍，例如僅檢視2010年的資料，就是對時間維度進行切片；切丁是限制多維度的範圍，例如檢視2010年北區的資料；與衡量值有關的操作有篩選>Selecting)、排序(Sorting)及計算(Computing)，篩選是限制衡量值的範圍，排序是將衡量值由大至小或由小至大排列，計算則是對衡量值進行計算；最後有輸出(Output)和視覺化(Visualize)，功能分別是將衡量值另外儲存供加值應用以及把數值化的衡量值以圖形方式表達。

## 2.1.2 維度資料模型

維度資料模型的建構，與關聯式資料庫由實體關聯圖(Entity-Relationship Diagrams)[30]開始的建構方式不同，是以資料表作為建構的基礎[24]。在維度資料模型中，資料表分為事實資料表(Fact Table)與維度資料表(Dimensional Table)。事實資料表存放的內容主要為企業使用的衡量值，屬於計量資料，例如單筆銷售金額、單筆銷售數量等；另外也存放事實資料表和維度資料表作關聯的外鍵(Foreign Key)[30]。

維度資料表主要存放文字型態資料，表內欄位包含主鍵(Primary Key)[30]及其他描述事實資料性質的屬性欄位，這些欄位用來管理維度的階層關係，並提供使用者利用維度屬性查詢事實資料。若維度屬性有著階層關係，則使用者可在這些維度中進行上捲及下挖的功能，如時間維度有年、月、日的階層關係，使用者就可從「年」的層級下挖到「月」的層級。

競賽的維度資料模型，依結構常分為兩種不同綱要(Schema)：星狀綱要(Star Schema)及雪花綱要(Snowflake Schema)[31]。星狀綱要是最常見的資料倉儲綱要，由一個事實資料表與多個維度資料表組成。其中多個維度資料表圍繞著事實資料表，如星形一般，因此稱

為星狀綱要，如圖 2.2(a)所示。雪花綱要的結構與星狀綱要類似，兩者差別在於雪花綱要中的維度資料表有進行三階正規化(Third Normal Form, 3NF)[30]，目的是減少資料的重覆性，然而會影響資料查詢速度，如圖 2.2(b)所示。



圖2.2 維度資料模型綱要圖

## 2.2 台灣花卉資料超市

本節主要說明花卉資料超市的維度結構與內涵，共分為兩小節，2.2.1 節說明資料超市維度結構，以及與資料倉儲維度的差異；2.2.2 節說明各維度的意義及架構。

### 2.2.1 資料超市維度結構

台灣地方花市資料超市採用維度資料模型建構，分為台北花市、台中花市、彰化花市、台南花市及高雄花市，共五座。其中分別存放各家花市的交易資料。由於資料超市與總倉儲規模不同，使用者也不同，採用的維度也有所差異，主要有以下三個部分：(1)資料超市在供應單位維度可查詢至供應單位層級，然而供應單位層級的資料量非常大，以台北花市為例，供應類別中僅個人花農的數量就有近千筆，而總倉儲是五座資料超市的加總，如提供查詢至供應單位維度，記憶體將無法負荷，因此基於執行效能考量，總倉儲不提供查詢至供應單位層級；(2)總倉儲新增市場別維度，供使用者查詢個別市場的資料，而資料超市並不需要此維度；(3)總倉儲新增節慶維度，便於使用者查詢，資料超市並無此維度。表 2.1 為地方資料超市與總倉儲的維度比較表。

表2.1 總倉儲及各家花市維度比較表

維度	地方資料超市	總倉儲
供應單位	有供應單位層級	無供應單位層級
供應地區	有	有
日期	有	有
星期	有	有
農曆	有	有
節慶	無	有
市場別	無	有
拍賣線別	不開放	不開放
花卉種類	有	有
花卉等級	有	有
容器	有	有
承銷地區	有	有
	共 10 個維度	共 12 個維度

### 2.2.2 資料超市維度

資料超市共有十個維度，分別是：供應單位、供應地區、日期、星期、農曆、拍賣線、花卉種類、花卉等級、容器及承銷地區。以下將針對單一維度進行說明。

#### 供應單位

供應單位維度分為兩個階層。第一層的屬性為供應單位類別，有個人花農、農會合作

社及公所、供應團體所屬花農、進口商、行口商、農企業、花班及其他；第二層的屬性為各供應單位的名稱。供應單位維度具有緩慢改變的現象，即維度的屬性資料會隨著時間改變，維度緩慢改變將於 2.3 節詳細說明。

#### 星期

星期維度屬性為一個星期的天數。

#### 供應地區

供應地區維度的屬性為行政區域，分為兩個層級，第一層為縣市等級，第二層為鄉鎮市區等級。

#### 日期

日期維度分為五個層級，由上到下分別為年、季、月、旬及日。此維度由線上分析軟體自動產生。

#### 拍賣線

花卉拍賣一般有多線同時進行，不同拍賣線負責不同種類的花卉，拍賣線維度屬性即為拍賣的線別。此維度涉及花卉公司商業機密，並不開放給終端使用者查詢。

#### 農曆

農曆維度類似日期維度，以農曆日期作為依據，分為年、月、日三層。

#### 花卉種類

花卉種類維度分為三個層級，第一層的屬性為花卉產地，分為國產花及進口花，第二層為花卉品類，第三層則為花卉品種。例如國產花下挖可得到康乃馨品類，由康乃馨再下挖即得到不同顏色、品種的康乃馨拍賣資料。

#### 花卉等級

花卉等級維度分為兩個層級，第一層的屬性為花卉品質等級，有特級、優級、良級三種；第二層為花卉長度等級。

#### 容器

容器維度的屬性為容器種類，分為紙箱及立式容器。

#### 承銷地區

承銷地區維度的屬性為大略的承銷地區，共分為九區，分別是：台北市、北部、台中、彰化、雲嘉、台南、高雄市、高屏及未分區。

## 2.3 維度緩慢改變問題

本節首先說明緩慢改變維度的定義，接著回顧因供應單位維度緩慢改變而進行的資料超市第一次維度再設計；最後比較本研究與第一次維度再設計的差異，並探討資料超市是否存在其他緩慢改變維度的可能。本節分為三小節，2.3.1 節說明緩慢改變維度定義與常用解法，2.3.2 節回顧資料超市第一次維度再設計，2.3.3 節比較本研究與先前再設計的差異，並分析其餘維度改變的可能。

### 2.3.1 緩慢改變的維度

資料倉儲具備非揮發性的特質，其記錄的資料不隨時間產生變化，因此可作為分析使用。當資料倉儲以資料維度模型建立時，其維度屬性若隨著時間會產生變化，就稱為緩慢改變維度[29]，例如商品維度中某商品的價格因通貨膨脹而進行調整，然而原有的商品價格也必須保存以對應至歷史資料，這就是很典型緩慢改變維度的例子。此現象於 1996 年由競筈[29]提出。

針對此問題，競筈提出三種解決方法[29]：(1)覆寫(2)新增資料列(3)建立「目前使用」欄位。方法(1)相當直觀，即是直接用新的資料覆蓋舊資料。此方法的優點是簡單且快速，無須針對維度資料表進行處理，缺點是歷史資料將無法保留。此方法適用於資料不具歷史意義，無須保存的欄位。

方法(2)是將變更後的資料當作一筆新的紀錄，並對資料表進行新增動作，如表 2.2 所示，供應商 A101 本來的所在位置為台北市，遷移到新北市後就在資料表新增一筆資料以代表現在的位置。此方法的優點是可以完整保留所有資料，並可以追蹤所有的變更紀錄；缺點是空間的使用量相對較大，且資料表的主鍵必須使用替代鍵。

表2.2 緩慢改變維度方法二範例

供應商主鍵	供應商代碼	供應商名稱	供應商地區	版本別
123	A101	XYZ Co.	台北市	0
124	A101	XYZ Co.	新北市	1

方法(3)是使用一個欄位存放目前的值，而用另一個欄位儲存修改前的值，如表 2.3 所示。此方法適用於必須同時關聯新舊資料的狀況，其缺點在於只能保存最近兩個版本的資料，如果要再往前回溯變更的紀錄，就必須再新增欄位。

表2.3 緩慢改變維度方法三範例

供應商主鍵	供應商代碼	供應商名稱	供應商地區	供應商目前地區
123	A101	XYZ Co.	台北市	新北市

### 2.3.2 第一次資料超市維度再設計

台灣花卉產業中，供應單位須使用五家花市統一的供應代碼進行拍賣，且此代碼並非唯一，以下兩種情況會更換其供應單位：(1)供應單位辦理註銷代碼後，其親屬或繼承者申請優先選用該代碼(2)代碼長時間無人使用，可重新分發給其他單位。此處由於五家花市使用共同代碼，須經過市場共同清查，確定供應人並未在任何市場供貨才會刪除並重新發放[13]。因此對於資料超市而言，每組供應代碼對應的供應單位名稱會隨著時間改變，進而產生維度緩慢改變現象。

2004 年溫師翰[16]提出第一次維度再設計以修正此問題，其方法是將複合鍵(Composite Key)與替代鍵(Surrogate Key)[30]混合運用，說明如下：利用供應代碼與供應單位名稱兩欄位形成複合鍵，再使用資料庫自動編號功能產生替代鍵作為資料表的主鍵，並新增「使用狀態」欄位判斷此供應單位是否處於供貨狀態。當讀取到供應代碼相同而名稱不同的供應單位時，就將原供應單位的使用狀態改為 N，並新增一筆供應單位資料，將其使用狀態設為 Y。轉入拍賣資料時，就連結狀態為 Y 的供應單位。其作業流程如圖 2.3，作業結果如圖 2.4。

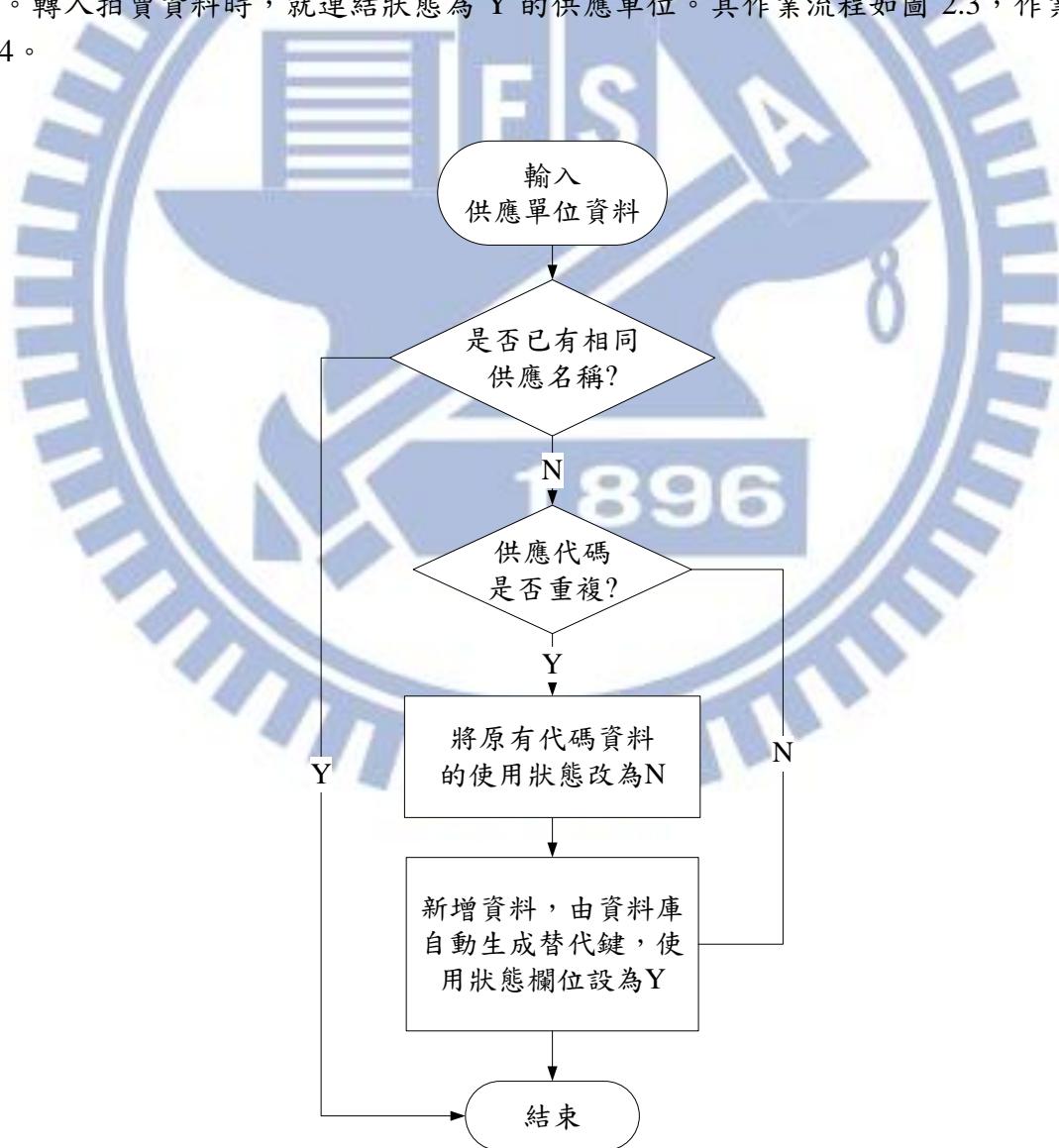


圖2.3 複合鍵與替代鍵混合使用的作業流程

Figure 2.4 illustrates the mixed use of composite keys and substitute keys through three tables:

自動編號(PK)	供應代碼	供應單位名稱	使用狀態
1	CFA01	王華	Y

↓ 王大華代碼註銷

自動編號(PK)	供應代碼	供應單位名稱	使用狀態
1	CFA01	王華	N

↓ 李明開始供貨

自動編號(PK)	供應代碼	供應單位名稱	使用狀態
1	CFA01	王華	N
2	CFA01	李明	Y

圖2.4 複合鍵與替代鍵混合使用結果

第一次再設計的第一個問題在於利用「使用狀態」記錄供應單位，無法得知每個供應單位供應的時期，假設交易資料未依時序匯入，就可能對應至錯誤的供應單位，造成資料錯亂。第二個問題則此再設計是無法反應供應單位存在的繼承特色，這將造成線上分析系統查詢結果失真。以上問題將於 3.2.1 節與 3.2.2 節深入探討，並為本次再設計的重點。

### 2.3.3 花卉倉儲其他潛在緩慢改變維度

花卉資料超市共有十個維度：供應單位、供應地區、日期、星期、農曆、拍賣線、花卉種類、花卉等級、容器及承銷地區。排除供應單位維度後，將剩下的維度依內容分類，可分為時間相關維度、地區相關維度及花卉相關維度。以下將針對各類別維度進行分析，以判斷其產生緩慢改變的可能性。

時間相關維度包括日期維度、星期維度與農曆維度。由於日期與農曆維度存放的值為每一天的國曆及農曆日期，是不會隨著時間發生改變的資料；星期維度存放一個星期中每一天的名稱，然而這些名稱使用已久，可視為不會改變的定值。總結以上性質，本論文將時間相關維度歸類為不緩慢改變的維度，然而需定期以人工方式新增資料。

地區相關維度包括供應地區與承銷地區維度。供應地區維度採用行政區域作為分區標準，存放縣市與鄉鎮市區的名稱與對應代號，當行政區域更動時，其值就會發生改變。例如 2010 年縣市改制直轄市，部分縣市合併為直轄市，所屬的鄉鎮市改為區[19]，造成供應地區維度所存放的地區名稱產生改變。

承銷地區維度採用較為粗略的分區方式，大致依照縣市名稱分區，如圖 2.5 所示，且承銷地區維度也可能受到行政區域更名的影響。地區相關維度都有緩慢改變的可能，但因其值與行政區域相關，改變的可能性相對較低，本論文將其歸類為隱性緩慢改變維度，並判斷使用人工方式修正及新增。

未分區
台北市地區
北部地區
台中地區
彰化地區
雲嘉地區
台南地區
高雄市地區
高屏地區

圖2.5 承銷地區維度屬性值

花卉相關維度共有四個，包括花卉種類、花卉等級、容器與拍賣線。花卉種類維度存放花卉的產地、品類與品名資料。當供應單位培育出新品種並到達一定的供貨量時，可向花市申請新增品名[20]。花市因使用上的需求，會不定期進行花卉品名微調，例如中英文品名並用，因此本論文將花卉種類維度歸類為顯性緩慢改變的維度。然而花卉品名的微調不具歷史意義，可採用 2.3.1 節提到的覆蓋方式，並由 ETL 系統處理。

花卉等級與容器維度存放內容都是花卉與容器的標準，這些標準由花市制定。因此當標準改變時，其內容也有改變的可能。然五家花市採用同一標準，更改的可能性較低，可列為隱性的緩慢改變維度。而拍賣線維度存放拍賣線的編號，也由花市決定是否改變，因此也歸類為隱性緩慢改變維度，並由人工修正與新增。各類別維度的緩慢改變性質可整理如表 2.4。

表2.4 維度緩慢改變性質

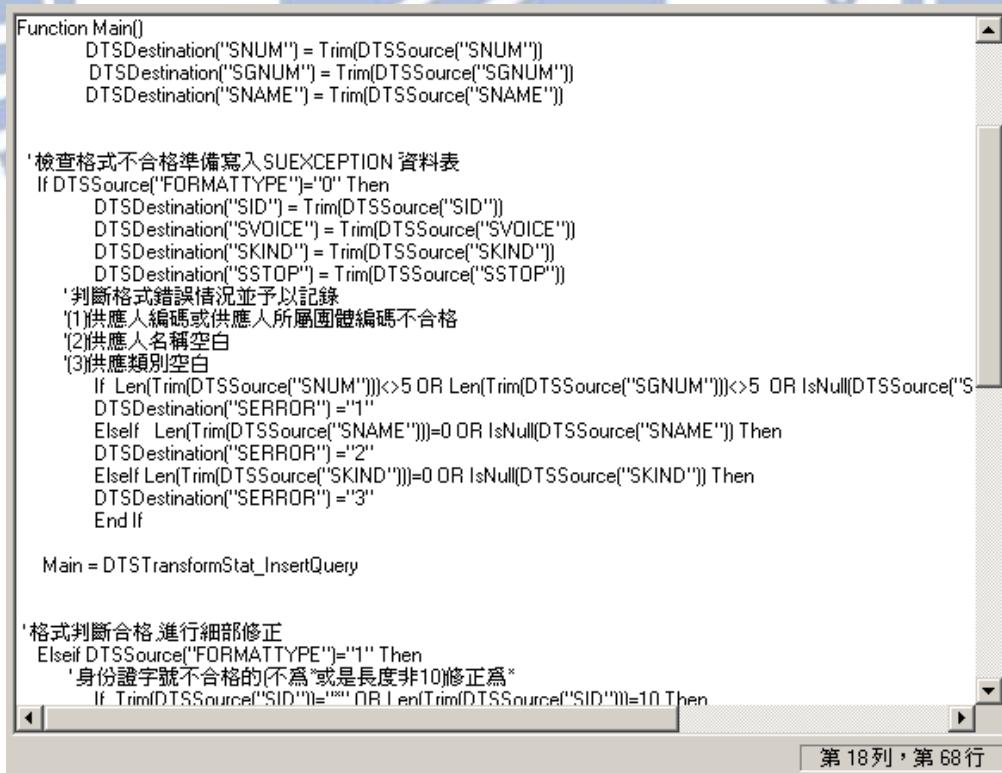
維度類別	維度名稱	緩慢改變性質	修正方式	新增資料方式
時間相關	日期	不改變	無	人工加入
	星期	不改變	無	人工加入
	農曆	不改變	無	人工加入
地區相關	供應地區	隱性	人工修正	人工加入
	承銷地區	隱性	人工修正	人工加入
花卉相關	花卉種類	顯性	覆蓋	ETL 匯入
	花卉等級	隱性	人工修正	人工加入
	容器	隱性	人工修正	人工加入
	拍賣線	隱性	人工修正	人工加入

## 2.4 SQL Server 整合服務

在花卉資料超市建立時，其 ETL 系統採用 SQL Server 2000 的 DTS[5]工具進行設計。然而微軟公司於 2005 年推出 SSIS[23]取代 DTS 成為搭配 SQL Server 2005 的 ETL 工具，因此本論文將比較 DTS 與 SSIS 何者較適合搭配本次維度再設計使用。本節首先簡略說明由 DTS 到 SSIS 發展的過程及 SSIS 核心內容，接著比較 SSIS 和 DTS 在處理資料轉換的不同之處。由於資料轉換是 ETL 系統相當重要的一環，本論文也將藉此決定採用何者進行 ETL 系統設計。以下共分為兩小節，2.4.1 節介紹 SSIS 的發展與內容，2.4.2 節比較 SSIS 與 DTS 的資料轉換。

### 2.4.1 SSIS 的發展與其內容

DTS 是微軟公司從 SQL Server 7.0 版開始內含的資料轉換工具，提供轉換資料工作(Transform Data Task)、批量插入工作(Bulk Insert Task)等[26]。當時其他提供類似功能的工具不是資料來源上有所限制，就是非常昂貴；而 DTS 不僅能在不同資料來源間轉換，且其包含在 SQL Server 軟體裡，不用另行購買，因而成為處理資料轉換的熱門選擇[26]。然而隨著資料量逐漸增加，使用者在利用 DTS 處理大量資料上遇到困難。且 DTS 的資料轉換功能以撰寫腳本(Script)語言為主，隨著轉換過程的複雜性增加，腳本語言介面將難以管理與維護，如圖 2.6，因此微軟公司決定推出一套更完善的 ETL 整合工具[26]。



```
Function Main()
    DTSDestination("SNUM") = Trim(DTSSource("SNUM"))
    DTSDestination("SGNUM") = Trim(DTSSource("SGNUM"))
    DTSDestination("SNAME") = Trim(DTSSource("SNAME"))

    '檢查格式不合格準備寫入SU_EXCEPTION 資料表
    If DTSSource("FORMATTYPE")="0" Then
        DTSDestination("SID") = Trim(DTSSource("SID"))
        DTSDestination("SVOICE") = Trim(DTSSource("SVOICE"))
        DTSDestination("SKIND") = Trim(DTSSource("SKIND"))
        DTSDestination("STOP") = Trim(DTSSource("STOP"))

    '判斷格式錯誤情況並予以記錄
    '(1)供應人編碼或供應人所屬團體編碼不合格
    '(2)供應人名稱空白
    '(3)供應類別空白
    If Len(Trim(DTSSource("SNUM")))<>5 OR Len(Trim(DTSSource("SGNUM")))<>5 OR IsNull(DTSSource("SNUM"))
        DTSDestination("SError") ="1"
        ElseIf Len(Trim(DTSSource("SNAME")))=0 OR IsNull(DTSSource("SNAME")) Then
            DTSDestination("SError") ="2"
            ElseIf Len(Trim(DTSSource("SKIND")))=0 OR IsNull(DTSSource("SKIND")) Then
                DTSDestination("SError") ="3"
            End If

    Main = DTSTransformStat_InsertQuery

    '格式判斷合格 進行細部修正
    Elseif DTSSource("FORMATTYPE")="1" Then
        '身份證字號不合格的(不為"或是長度非10修正為"
        If Trim(DTSSource("SID"))="" OR Len(Trim(DTSSource("SID")))=10 Then
```

第 18 列，第 68 行

圖 2.6 DTS 的腳本語言設計畫面

SSIS[3]即是微軟公司在推出 SQL Server 2005 時同時推出的資料整合工具，其用以設計、建立、存取和執行的單位是封裝(Package)，封裝內含多種物件可供運用，包括(1)工作(Task)(2)容器(Container)(3)優先順序約束條件(Precedence Constraint)[3]等。在流程的設計上，主要就利用這三種物件來進行，以下將分別說明其功能：

**工作**：封裝內用來執行各種功能的元件，一個封裝內可包含多個工作。工作的內容主要分為兩大類，分別是「資料流程工作」(Data Flow Task)[26]與「一般工作」。「資料流程工作」針對資料進行處理，包括擷取、轉換及載入等，也就是 ETL 系統需使用的功能。「一般工作」則用來執行寄送電子郵件、SQL 語法等輔助性質的任務。

**容器**：用來控制工作執行的流程，提供用迴圈或循序的方式執行。當特定工作須重複執行多次時，可將工作置於「迴圈容器」(Loop Container)[3]內；當某些工作需要一起執行，且以整體執行的結果判斷成功與否，則可將這些工作置於「循序容器」(Sequence Container)[3]。

**優先順序約束條件**：作為工作與工作之間的連結，並可設定工作執行的順序。

利用以上三種物件即可進行資料的處理與轉換，並控管轉換的流程。工作、容器及優先順序約束條件的關係示意如圖 2.7。

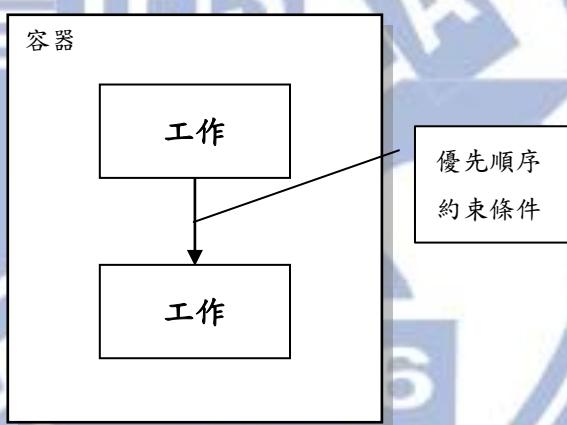


圖 2.7 SSIS 流程示意

#### 2.4.2 SSIS 與 DTS 資料轉換內容與比較

DTS 用來進行資料轉換的元件為「轉換資料工作」，外型為連結箭號，如圖 2.8(a)所示。在使用「轉換資料工作」前必須先建立兩個資料連結，以設定轉換的來源與目地資料表。「轉換資料工作」則對來源及目的資料表的欄位進行簡單對應，如圖 2.8(b)。如果需要篩選或轉換來源資料表的資料，必須使用腳本語言撰寫細部的設定，如圖 2.8(c)。

SSIS 進行資料轉換的工作為「資料流程工作」，其外型與其他工作相同，如圖 2.9(a)。「資料流程工作」可展開為獨立的子流程，包含許多用於資料轉換的子工作。子工作主要分為三大類：(1)來源(Sources)(2)轉換(Transformations)(3)目的地(Destinations)[26]，以下分別說明：

**來源**：此類型子工作可萃取不同格式的外部資料以進行轉換，包括資料庫、Excel 報表、文字檔案等。

**轉換**：此類型子工作提供更新、清除、合併與分割資料等功能，以輔助 ETL 的資料轉換過程。例如「衍生的資料行 (Derived Column)」[26]子工作可利用運算式更新資料值。

**目的地**：此類型子工作將資料載入不同類型的目的地，包括資料庫、一般檔案等。

以圖 2.9(a)的「資料流程工作」為例，其子流程中就包含來源、目的地及轉換三種類型的子工作。而這些子工作的內容都以圖形介面進行設定，如圖 2.9(b)。



(a) DTS 資料流程

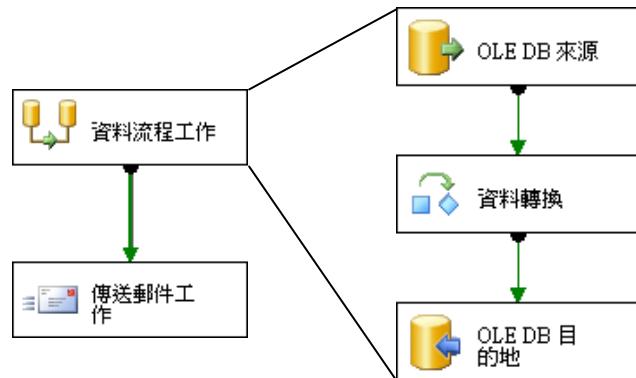


(b) 資料對應

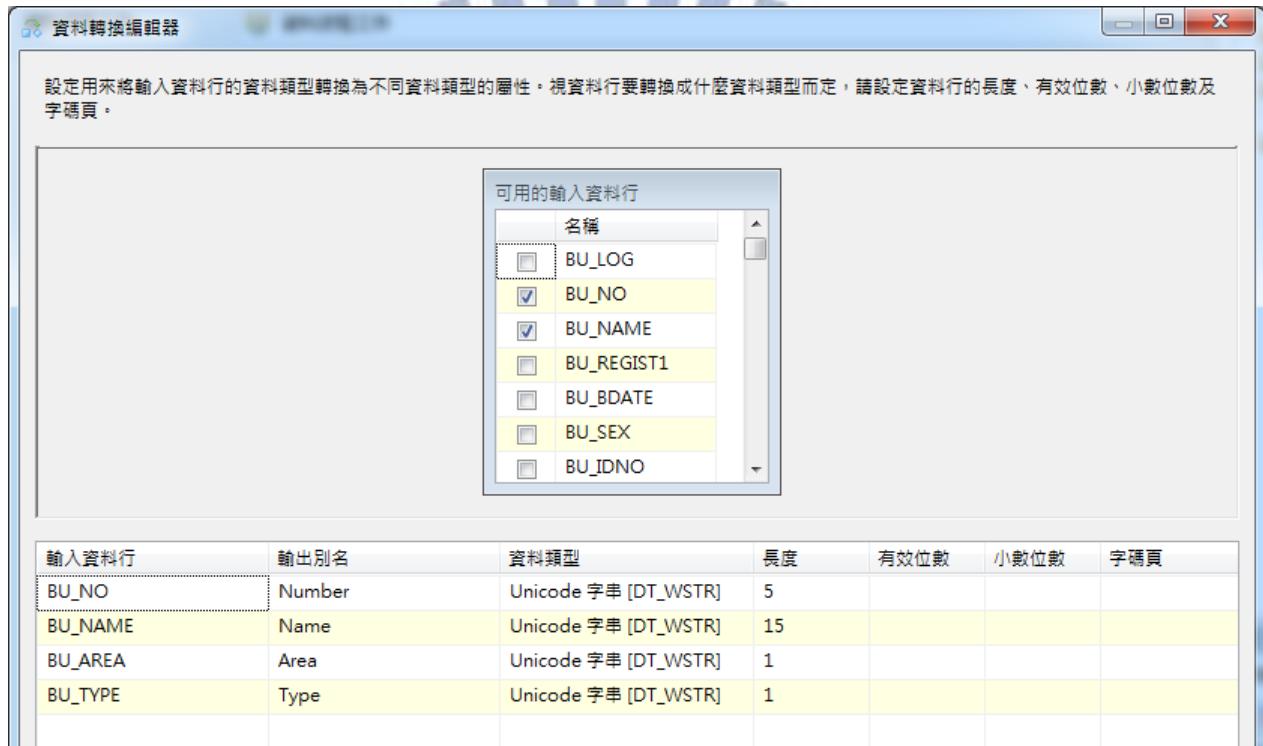
```
' Visual Basic Transformation Script
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
' 將每一個來源資料行複製到目的資料行
Function Main()
    DTSDestination("TP_BUY_NUM") = DTSSource("BNUM")
    DTSDestination("TP_BUY_NAME") = DTSSource("BNAME")
    DTSDestination("TP_BUY_AREA") = DTSSource("BAREA")
    DTSDestination("TP_BUY_ID") = DTSSource("BID")
    Main = DTSTransformStat_OK
End Function
```

(c) 腳本語言編輯畫面

圖2.8 DTS 資料轉換



(a) SSIS 資料轉換流程



(b) 資料轉換設計畫面

圖2.9 SSIS 資料轉換

```

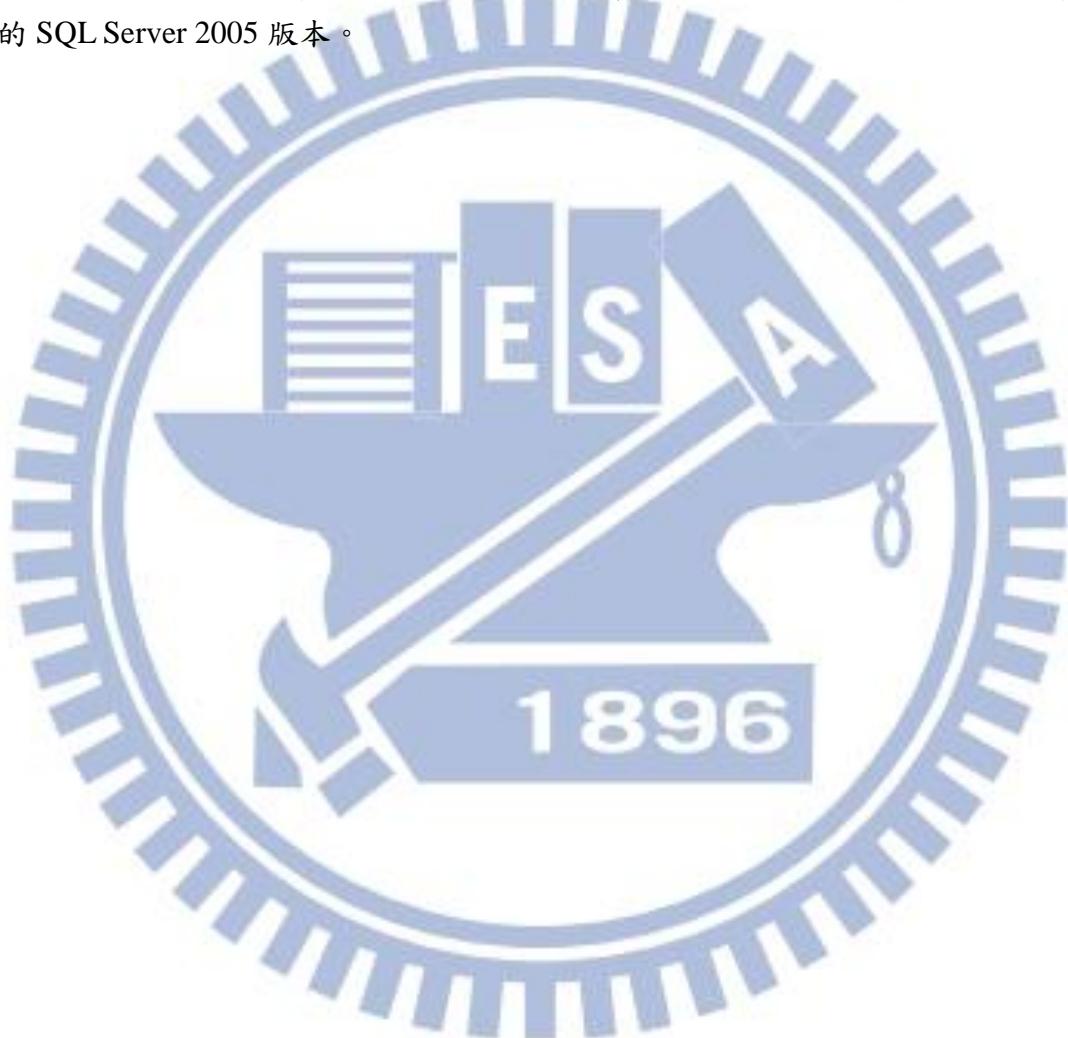
▶ [DTS.Pipeline] 資訊: 正在開始執行前階段。
▶ 進度: 執行前 - 0 百分比完成
▶ 進度: 執行前 - 9 百分比完成
▶ 進度: 執行前 - 18 百分比完成
▶ 進度: 執行前 - 27 百分比完成
▶ 進度: 執行前 - 36 百分比完成
▶ 進度: 執行前 - 45 百分比完成
▶ 進度: 執行前 - 54 百分比完成
▶ 進度: 執行前 - 63 百分比完成
▶ 進度: 執行前 - 72 百分比完成
▶ 進度: 執行前 - 81 百分比完成
▶ 進度: 執行前 - 90 百分比完成
▶ 進度: 執行前 - 100 百分比完成
▶ [DTS.Pipeline] 資訊: 正在開始執行階段。
① [供應團體代碼格式判斷 [661]] 錯誤: "輸出 \"團體所屬花農代碼格式錯誤\" (703)" 上的運算式 "((DT_I4
① [供應團體代碼格式判斷 [661]] 錯誤: SSIS 錯誤碼 DTS_E_INDUCEDTRANSFORMFAILUREONERROR
① 錯誤: SSIS 錯誤碼 DTS_E_PROCESSINPUTFAILED。在元件 "供應團體代碼格式判斷" (661) 上的 Proce
① 錯誤: SSIS 錯誤碼 DTS_E_THREADFAILED。"WorkThread0" 執行緒已結束，錯誤碼為 0xC0209029。

```

圖2.10 SSIS 錯誤訊息

DTS 與 SSIS 的資料轉換可從兩個層面進行比較：(1)設計方式(2)錯誤偵查。對於設計方式而言，DTS 的資料轉換以腳本語言設計，將所有轉換流程寫在一份文件內，如圖 2.6；SSIS 的資料轉換則使用圖形介面進行設計，並使用與一般工作一致的格式，如圖 2.9(a)。對於錯誤偵查而言，DTS 無法分析以腳本語言撰寫的資料轉換流程，因此必須利用人工檢視程式碼找出錯誤來源；在 SSIS 中，由於資料轉換流程由「工作」構成，會明確回報產生錯誤的「工作」，如圖 2.10，可知是「供應團體代碼格式判斷」工作項目發生錯誤。

經過以上整理與比較，本研究認為 SSIS 的圖形介面可幫助開發與維護 ETL 時，更清楚了解資料轉換的流程；而其偵錯能力不僅可加速開發時的除錯，也可降低往後維護難度。基於以上原因，本研究選擇 SSIS 作為 ETL 再設計的開發工具，資料庫軟體則選擇與 SSIS 相容的 SQL Server 2005 版本。



## 第三章 資料超市故障分析

本章首先進行資料超市維度的分析，找出造成資料錯誤的關鍵維度。分析後發現供應單位維度存在維度緩慢改變造成的問題，因此將深入分析問題來源，並針對問題提出再設計方法。由於本研究使用 SQL Server 2005 作為資料庫設計工具，維度模型必須進行調整，以配合新的資料庫軟體。本章共分為兩節，3.1 節進行資料超市的故障來源分析，找出故障來自於供應單位維度；3.2 節則深入分析供應單位維度的問題，分為供應單位不連續問題與供應單位品牌辨識問題。

### 3.1 資料超市故障來源分析

為了找出資料超市錯誤的原因，須針對各維度資料輸入其更新方式進行分析，以篩選出發生錯誤的維度。分析結果整理如表 3.1。

其中日期維度與星期維度透過 Cognos 商業智慧軟體[21]自動產生資料並更新，因此可將此二維度優先排除。其餘以人工輸入且不需更新的維度，在資料超市建置完成後即輸入資料並經過檢驗無誤，因此也可將其排除。農曆維度定期以人工輸入方式更新，僅可能發生人為錯誤，不在本研究討論範圍。

透過 ETL 轉入資料並更新的維度有供應單位維度與花卉種類維度。其中花卉種類維度存放花卉品名資料，花卉的品名由供應人向花市申請新增，而花市會進行審核，通過後即不會再行更改[20]。因此該維度雖會新增資料，卻不會產生維度緩慢改變現象。而供應單位維度則如於 2.3 節所述，具有緩慢改變維度現象，因此本研究選擇供應單位維度進行深入分析。

表3.1 資料超市維度性質整理

維度	初始輸入方式	資料是否持續更新	資料更新方式	緩慢改變現象	故障來源
供應單位	ETL 轉入	是	ETL 轉入	有	是
供應地區	人工輸入	否	無	無	否
日期	軟體自動產生	是	軟體自動更新	無	否
星期	軟體自動產生	否	無	無	否
農曆	人工輸入	是	人工輸入	無	否
拍賣線別	人工輸入	否	無	無	否
花卉種類	ETL 轉入	是	ETL 轉入	無	否
花卉等級	人工輸入	否	無	無	否
容器	人工輸入	否	無	無	否
承銷地區	人工輸入	否	無	無	否

## 3.2 供應單位維度故障問題分析

經過 3.1 節整理，本研究決定以供應單位維度為研究主體，分析其是否存在導致資料錯誤的問題，而分析結果發現供應單位維度存在兩個問題，分別是供應單位不連續問題與供應單位品牌辨識問題。本節共分為兩小節，3.2.1 節說明資料不一致問題，3.2.2 說明供應單位品牌辨識問題。

### 3.2.1 供應單位不連續問題

供應單位不連續問題如圖 3.1 所示，從資料庫可得知代碼 HHA05 先後由兩位不同林姓供應人使用，如圖 3.1(a)，且供應代碼更換使用者後並不能任意切換回前任使用者；然而在拍賣資料表部份卻發現連續日期的拍賣紀錄輪流連結至兩位林姓供應人，如圖 3.1(b)，形成供應單位的不一致。

TP_SUP_NUM	TP_SUP_ONUM	TP_SUP_NAME
5963	HHA05	HHA05 林榮
10386	HHA05	HHA05 林盛

(a) 供應單位維度資料表

TP_DT_AUCNUM	TP_DT_DATE	TP_DT_PR	TP_DT_SU
9511091769901	2006/11/9	FR408	5963
9511093004801	2006/11/9	FR408	5963
9511093004901	2006/11/9	FR408	5963
9511093004902	2006/11/9	FR408	5963
9511093005001	2006/11/9	FR408	5963
9511103003401	2006/11/10	FR408	10386
9511103003501	2006/11/10	FR408	10386
9511103003502	2006/11/10	FR408	10386
9511103003601	2006/11/10	FR408	10386
9511103003602	2006/11/10	FR408	10386
9511103003701	2006/11/10	FR408	10386
9511113002601	2006/11/11	FR408	5963
9511113002701	2006/11/11	FR408	5963
9511113002702	2006/11/11	FR408	5963
9511113002801	2006/11/11	FR408	5963
9511113002802	2006/11/11	FR408	5963
9511113002901	2006/11/11	FR408	5963

(b) 拍賣資料表

圖 3.1 供應單位維度資料錯誤

欲分析此問題，首先須確認拍賣資料轉入資料超市的流程。由於花卉資料超市為適時(Right Time)系統[15]，以月份為單位轉入批次資料。其轉入資料的流程如下：(1)先檢查供應單位資料是否有更新，若花市有提供更新資料，則進行供應單位維度的更新(2)轉入拍賣資料，並將供應代碼取代為資料庫自動生成的供應單位編號(3)上期若有遺漏資料，則於此期轉入時補齊。

圖 3.2 說明問題產生的情境。由圖 3.2 可看出 3 月份的資料於 3 月 31 日轉入，3 月份並未有供應單位資料更新，因此直接進行拍賣資料的轉入，其中 3 月 4 日的資料因花市未傳送而產生缺漏；假設花市在 4 月補齊了 3 月 4 日的資料，並且更新供應單位資料，將代碼 AA001 的使用者從供應人甲改為供應人乙，則必須先進行供應單位資料表的更新，將供應人甲使用狀態改為 N，並新增供應人乙資料，再轉入 4 月份的拍賣資料。接著要補齊 3 月份缺漏資料時，會將 AA001 的供應代碼連結至供應人乙，就產生了資料不一致。

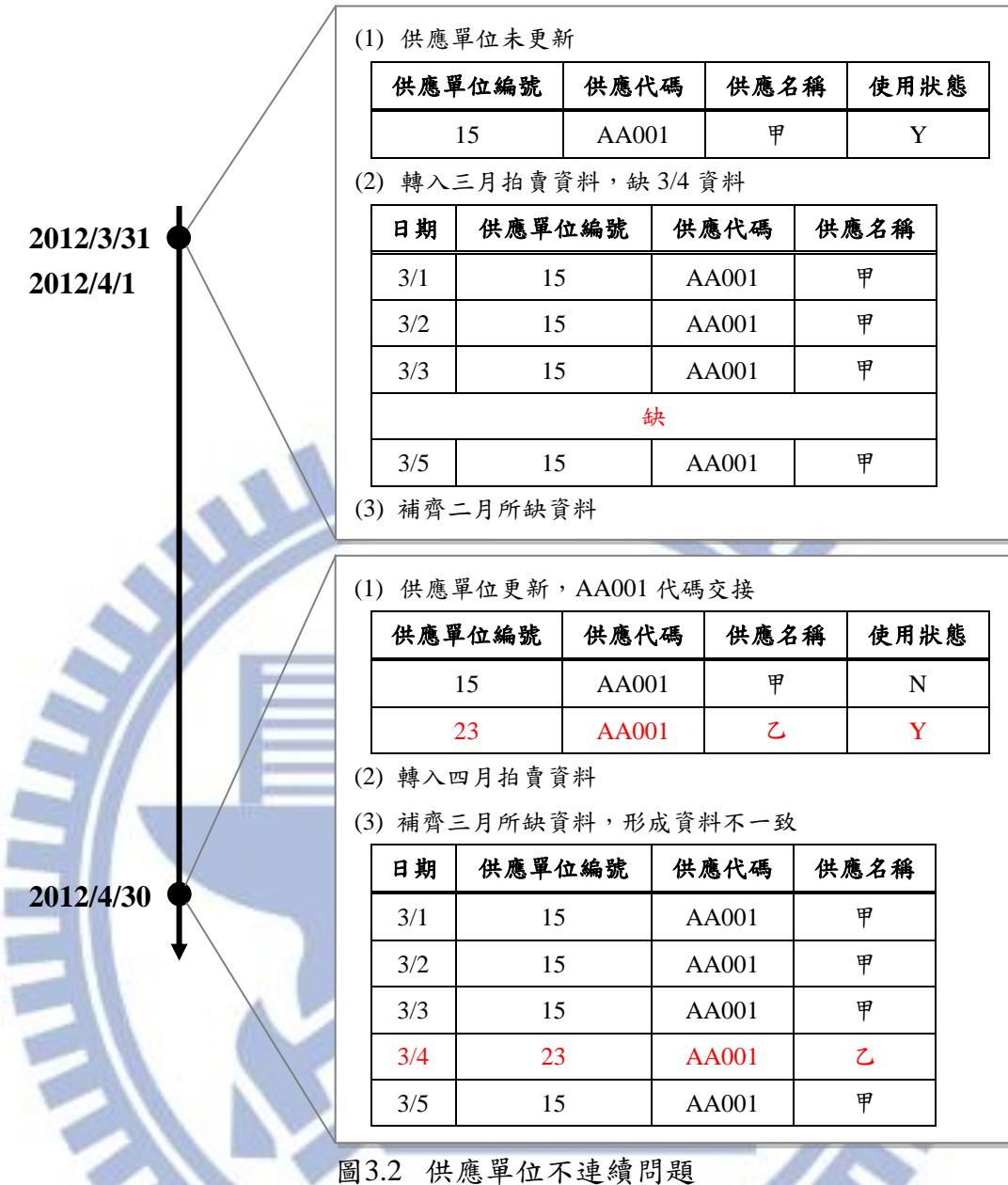


圖3.2 供應單位不連續問題

產生此問題的原因，是由於第一次再設計利用「使用狀態」欄位來紀錄目前使用代碼的供應單位，卻沒有紀錄代碼轉換的日期，造成補齊缺漏拍賣資料時直接連結狀態為 Y 的供應單位，形成資料不一致的現象。

### 3.2.2 供應單位品牌辨識問題

2.3 節提及供應單位主動註銷代碼時，可由原供應人的親屬或繼承者優先申請使用該代碼，此方式通常會在舊供應單位辦理註銷代碼後同步更名為新供應單位，且申請通過後可立即開始供貨[13]。舉例來說，假設供應人甲於 3 月 15 日供貨結束後註銷代碼，並同步更名為供應人乙，則供應人乙於 3 月 16 日開始即可繼續供貨。此種優先申請辦法之所以存在，是因為花市與承銷人透過供應代碼來辨識供應單位的品牌，而非透過供應名稱。採用此方式交接代碼的供應單位，其品牌形象具有延續性，不會因為供應名稱不同而有所改

變[11]。而經由市場清查刪除的代碼，再分配給供應單位時，其代碼所象徵的品牌意義則不會延續。

以拍賣市場的觀點來看，假設代碼 AA001 在 2005 年為供應人甲使用，其後因未繼續供貨而清查刪除；2010 年代碼 AA001 再度發放給供應人乙使用，而供應人乙在 2012 年 3 月中旬交接給供應人丙，則在 2012 年 4 月查詢供應單位資料，應呈現如表 3.2，然而利用線上分析處理查詢所得到的結果將如表 3.3，使用者無從得知供應人乙與丙具有相同的供應品牌，供應人甲則是獨立的供應品牌。

表3.2 供應單位資料品牌意義

供應單位 年度	AA001 甲	AA001 丙
2005	10086	0
2006	0	0
2007	0	0
2008	0	0
2009	0	0
2010	0	13190
2011	0	14875
2012	0	4421

\*資料單位：拍賣把數

表3.3 供應單位資料查詢結果

供應單位 年度	AA001 甲	AA001 乙	AA001 丙
2005	10086	0	0
2006	0	0	0
2007	0	0	0
2008	0	0	0
2009	0	0	0
2010	0	13190	0
2011	0	14875	0
2012	0	2651	1770

\*資料單位：拍賣把數

會產生此問題，是由於第一次再設計處理維度緩慢改變時，未考量供應代碼有品牌延續的特性，使用供應名稱作為唯一的判斷依據，且將名稱不同的供應單位視為獨立個體，造成資料超市無法反應真實情況，影響線上分析處理的查詢結果。

## 第四章 供應單位維度的再設計

針對 3.2 節提出的供應單位維度問題，本章使用新增欄位的再設計方式將不同問題個別解決。而由於本次維度再設計使用 SQL Server 2005 進行，將配合其調整維度模型與資料型態。本章分為三節，4.1 節為供應單位不連續問題的再設計；4.2 節為供應單位品牌辨識問題的再設計；4.3 節調整維度模型，並整理更新後的資料表。

### 4.1 供應單位不連續問題再設計



圖 4.1 供應單位不連續問題修正

3.2.1 節指出利用「使用狀態」來紀錄目前使用代碼的供應單位，會產生供應單位不連續的現象。針對此問題，本研究使用「時間戳記(Timestamp)」來解決。時間戳記為一組字

串，用以表達事件發生的日期與時間。以具有時間戳記性質的欄位取代「使用狀態」欄位，可更精確地界定不同供應時期。實作上在供應單位維度中新增「開始日期」與「結束日期」欄位，在新增供應資料時記錄開始日期，結束日期預設為一相隔較遠的日期，例如 2100 年 1 月 1 號；在交接供應單位時將結束日期改為更新供應單位前一天的日期，如此可得到供應單位的供應期間，以作為連結拍賣資料的判斷。3.2.1 節的案例經過再設計後，資料轉入流程將如圖 4.1 所示。

由圖 4.1 可看出 3 月份資料與前例相同，但「使用狀態」欄位改為「開始日期」與「結束日期」欄位，假設開始日期為 2010 年 5 月 1 號，結束日期為預設日期。而在 4 月 30 日更新供應單位資料時，先紀錄供應人甲的供應期限至 4 月 29 日，再加入供應人乙資料。接著補齊 3 月份缺漏資料時，就可利用「開始日期」與「結束日期」欄位作為判斷基準，將 3 月 4 日代碼 AA001 的拍賣資料連結至供應人甲，維持供應單位的一致性。



## 4.2 供應單位品牌辨識問題再設計

3.2.2 節提出當供應單位代碼相同而名稱不同時，在品牌意義上有具有兩種可能性：(1)前後供應單位屬於相同品牌(2)前後供應單位屬於不同品牌。本研究將(1)的交接方式定義為「繼承」。然而資料超市目前擁有的欄位不足以判斷代碼的交接是否屬於繼承方式，必須取得更多資訊輔助判斷。

經過分析花市提供的原始供應單位資料並與花市求證後，得到下列資訊可作為判斷的依據：(1)「停用狀態」欄位(2)資料是否刪除。前者紀錄供應代碼的狀態，數值為1代表此代碼正常供貨，數值為0則代表此代碼為停用狀態，供應單位沒有供貨[13]；後者是指透過市場清查確定代碼沒有在任何市場供貨，則花市會將內部資料庫中該代碼的供應資訊刪除，傳送到資料超市的原始資料也會同步更新。

整理以上資訊，供應代碼的狀態變化可以圖 4.2 表達。代碼的初始狀態為無供應單位，經過分發給供應單位，狀態則改為供應中。當供應代碼使用繼承方式交接時，花市會直接於資料庫更改供應名稱，代碼狀態不會改變[12]。要將代碼狀態改為停用，可透過下列方式：(1)供應單位主動停用代碼(2)長時間未供貨，由花市停用代碼[12]。當市場清查時，若該代碼在別家花市尚有供貨則不刪除，並保持停用狀態；若在所有花市都未供貨，則予以刪除，回復至無供應單位的狀態。

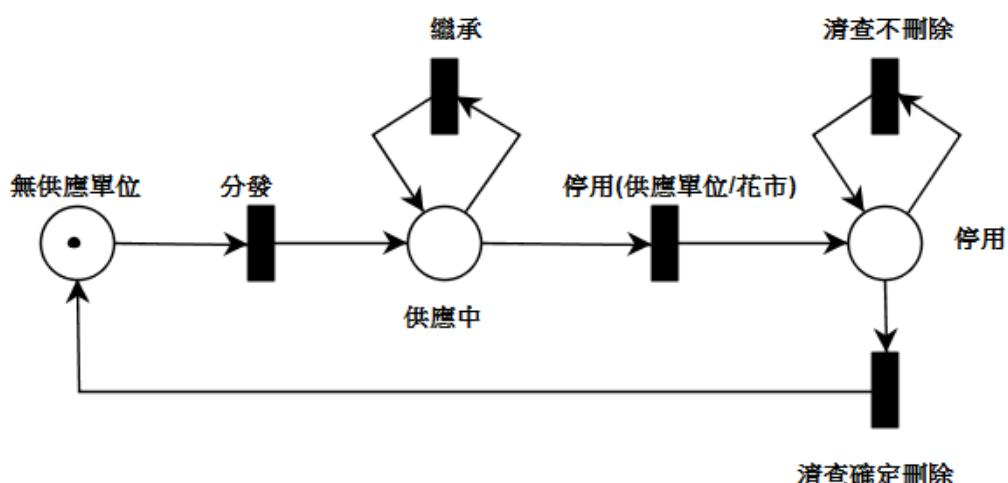


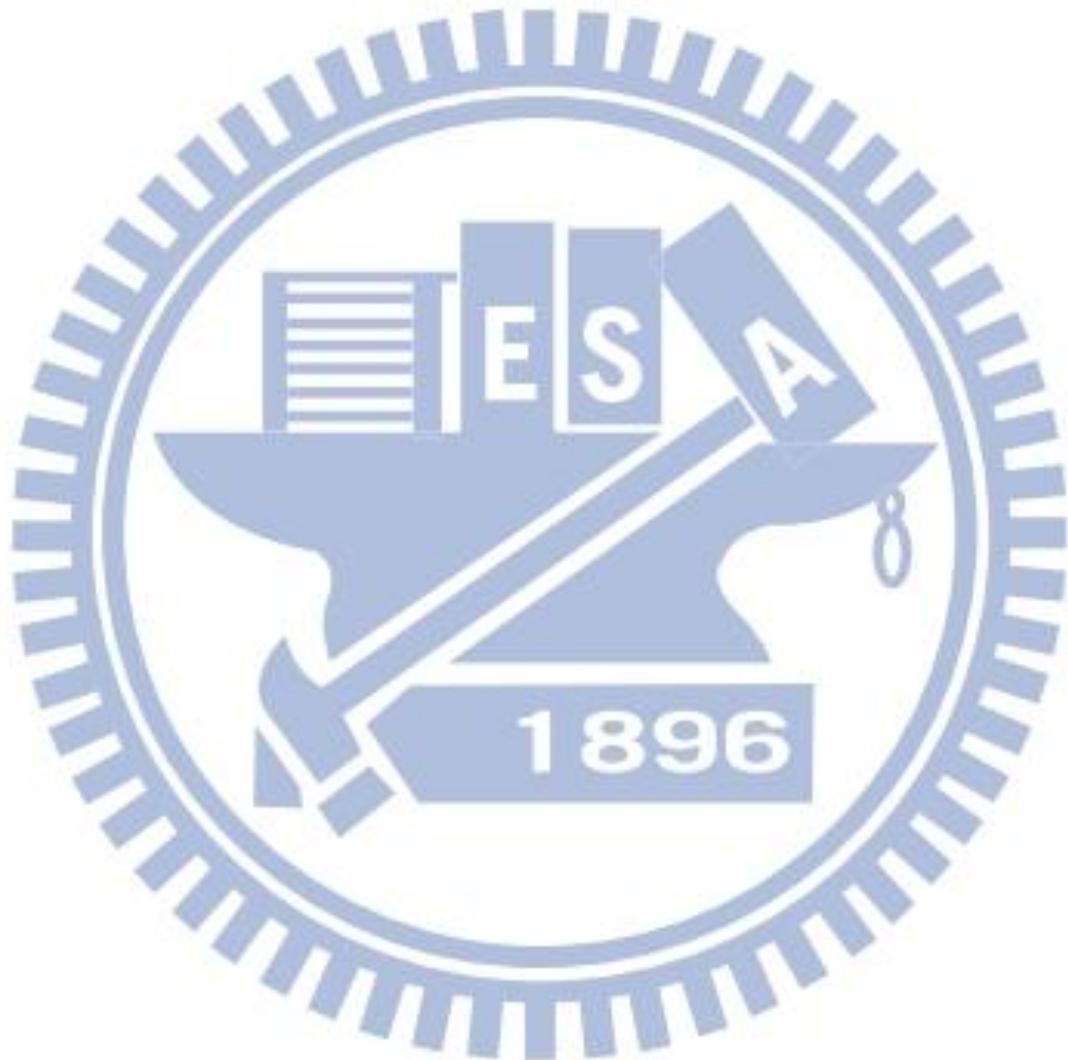
圖4.2 供應代碼狀態變化

透過以上分析，可發現代碼是否使用繼承方式交接，其中判斷的關鍵為有無回到「無供應單位」狀態。將花市傳送的原始供應單位資料與資料超市進行比對，發現有供應代碼遭到刪除，即代表該代碼已回到「無供應單位」狀態；而之後再度有供應單位使用該代碼時，就可判定前後任為不連續的品牌。反之未經歷「無供應單位」狀態的代碼，卻更新了供應名稱，則可判斷前後任供應單位是連續的品牌。

針對繼承判斷問題的再設計，本研究使用「旗標(Flag)」來處理。旗標在資料庫裡代表存放二元資料的欄位，用以表示資料的「是/否」等二元性質狀態。實作上以旗標性質「品牌結束」欄位來紀錄代碼的品牌是否已不再延續。紀錄的流程如下：「品牌結束」欄位預

設值為 0，代表品牌有繼承的可能；於更新供應單位資料時進行代碼比對，若有代碼遭到刪除，則將其「品牌結束」欄位值改為 1，以表示之後若到讀取到相同代碼的資料，就可直接視為新的品牌。

使用此方式的前提是供應單位資料維持每日更新的頻率，然而目前資料超市採用每月更新的方式，若供應代碼的狀態從「供應中」改為「停用」，並經過市場清查刪除回到「無供應單位」，最後分發給新的供應單位，整個變更過程所花的時間在一個月內，則此判斷方式會將其視為連續品牌。因此透過「品牌結束」欄位判斷是否為連續品牌，並配合每日更新資料，就能確保判斷的正確性，解決供應單位品牌辨識的問題。



## 4.3 使用 SQL Server 2005 再設計維度資料模型

由於本次再設計採用 SQL Server 2005 作為資料庫軟體，必須將維度資料模型由 SQL Server 2000 移植至 SQL Server 2005。在各維度結構不變的前提下，維度資料模型所選用的綱要圖設計將會影響其後 ETL 系統建立的複雜度[30]，由於資料超市會使用 SSIS 設計新的 ETL 系統，本節將重新評估欲採用的綱要模型。

台灣花卉資料倉儲過去採用雪花綱要模型，以減少資料表使用的空間，然而雪花綱要將維度進行三階正規化，ETL 系統轉換的過程就必須將資料重複轉入多個表格，並確定其一致性，增加了 ETL 系統設計的複雜度；而星狀綱要因結構簡單，ETL 系統的設計也相對簡單[31]。

針對星狀綱要使用儲存空間增加的問題，可比較雪花綱要與星狀綱要各維度屬性個數及資料筆數，如表 4.1。由表中可看出雖然星狀綱要屬性數雖然較少，資料表的去正規化(Denormalization)[30]的確會造成資料筆數增加。統計最早建立的台北花卉資料超市，其中最大的供應單位維度有三個階層，屬性數與資料表數如表 4.1，假設一個資料欄位使用 8 byte 的硬碟空間，則使用雪花綱要的供應單位維度資料表使用空間計算如(1)；使用星狀綱要的供應單位維度資料表使用空間計算如(2)。由(1)(2)可知就供應單位維度而言整體增加的容量不超過 500 MB，並不會對硬碟空間造成嚴重影響。

$$8 \times (5 \times 11768 + 3 \times 3558 + 2 \times 8) = 556240 \text{ byte} \approx 543 \text{ MB} \quad (1)$$

$$8 \times (11 \times 11768) = 1035584 \text{ byte} \approx 1011 \text{ MB} \quad (2)$$

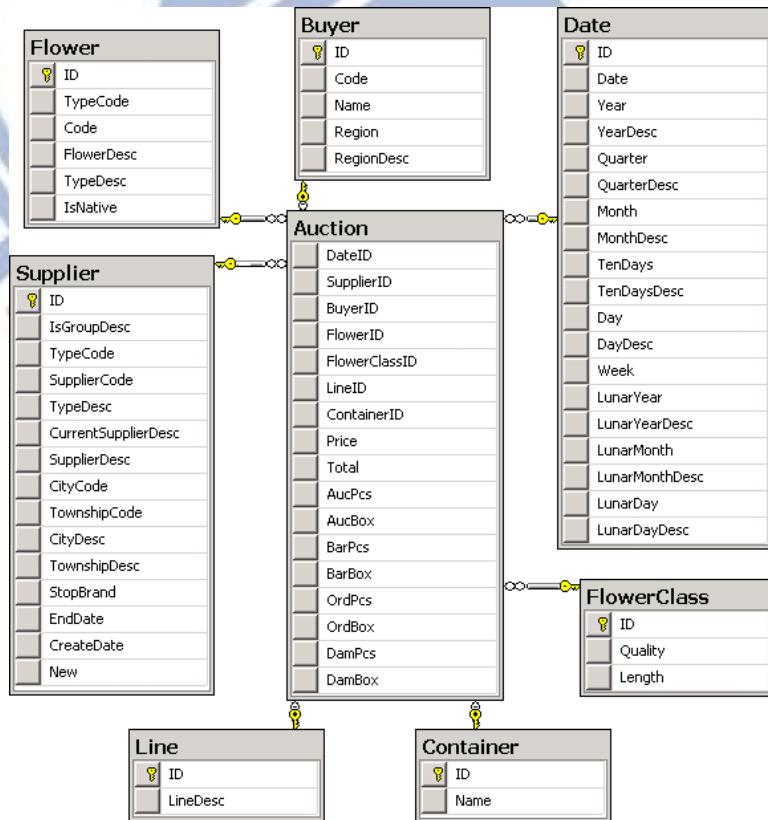


圖 4.1 花卉資料超市星狀綱要圖

因此本研究選用星狀綱要建立維度資料模型，以降低其後 ETL 系統設計的複雜度。SQL Server 2005 建立的花卉資料超市星狀綱要圖如圖 4.1，其中雪花綱要的「日期」維度將由線上分析系統軟體自動產生，故不列出。而由於 SQL Server 2005 與 SQL Server 2000 的資料型態有些微差異，也需進行調整。修正資料型態後的維度資料表見附錄 A.1。

表4.1 星狀綱要與雪花綱要資料筆數比較表

維度	雪花綱要		雪花綱要資料筆數	星狀綱要屬性數	星狀綱要資料筆數
	資料表	屬性數			
供應單位	供應單位	5	11768	11	供應單位數(萬)
	供應團體	3	3558		
	供應類別	2	8		
供應地區	一二級行政區	2	27		
	三級行政區	3	353		
星期	星期	2	7	2	7
農曆	農曆年	2	農曆年(十)	8	農曆日期(千)
	農曆月	3	12		
	農曆日期	4	農曆日期(千)		
拍賣線	拍賣線	2	拍賣線數	2	拍賣線數
花卉種類	花卉品種	5	花卉品種數(千)	10	花卉品種數(千)
	花卉品類	3	花卉品類數(百)		
	產地	2	2		
花卉等級	等級	3	39	4	39
	等級略表	2	3		
容器	容器	2	2	2	2
承銷地區	承銷人	3	承銷人數(百)	5	承銷人數(百)
	承銷地區	2	8		
總數	50		較少	33	較多

# 第五章 花卉資料超市 ETL 系統再設計與實作

經過供應單位維度再設計後，維度資料表的欄位有所變化，因此 ETL 系統也需配合維度資料表進行再設計。本章將針對原花卉資料倉儲系統的 DTS-1 與 DTS-2 進行再設計，並根據 SSIS 的執行單位，將 DTS-1 更名為資料庫封裝，DTS-2 更名為資料超市封裝，以便分辨其目的。而第四章提出的維度再設計實作部份，將分散於資料庫封裝與資料超市封裝中進行修正。本章共分為三小節：5.1 節說明實作使用的 SSIS 元件設定；5.2 節說明資料庫封裝的再設計與實作；5.3 節說明資料超市封裝的再設計與實作。

## 5.1 SSIS 元件設定

本節說明實作 ETL 系統須使用的 SSIS 元件及其設定方式，所使用的元件包括連結外部資料的「連接管理員」與進行資料轉換的「資料流程工作」。本節共分為兩小節，5.1.1 節說明連接管理員的設定；5.1.2 節說明資料流程工作的設定。

### 5.1.1 SSIS 連接管理員

連接管理員是 SSIS 連接外部資料的工具，提供許多不同來源的連接。設計資料庫封裝與資料超市封裝時主要使用「OLE DB 連接管理員」與「一般檔案連接管理員」。

OLE DB(Object Linking and Embedding, Database)是微軟公司用以連接不同來源資料的標準應用程序介面[22]。在資料庫封裝中，OLE DB 連接管理員可用於連接 SQL Server 資料庫與台北花市原始檔所使用的 dBASE 檔案。連接 SQL Server 的設定如圖 5.1，提供者選擇 SQL Native Client，表示對象為 SQL Server 原生資料庫；接著選取資料庫所在的伺服器名稱及驗證方式，通過驗證後即可選擇欲連接的資料庫，並完成連接。



圖 5.1 SQL Server 連接設定

dBASE[17]是一種關聯式資料庫系統，多用於磁碟檔案系統(Disk Operating System, DOS)。連接 dBASE 檔案時，提供者須選取 Microsoft Jet 4.0 OLE DB Provider，並將資料庫檔名設為 DBF 檔所在的資料夾路徑，如圖 5.2(a)。接著選取「全部」屬性，將 Extended Properties 設為 dBASE 5.0，如圖 5.2(b)，即可完成連接。



(a) dBASE 資料庫設定



(b) dBASE 進階設定

圖5.2 dBASE 連接設定

一般檔案連接管理員可連接TXT檔與CSV檔，台北以外花市所傳送的原始檔都是TXT檔，因此採用一般檔案連接管理員。其設定方式如圖 5.3，輸入檔案所在位置，並針對檔案內容進行格式調整。若資料的首行為欄位名稱，則須勾選下方「第一個資料列的資料行名稱」，才可新增為欄位名稱。

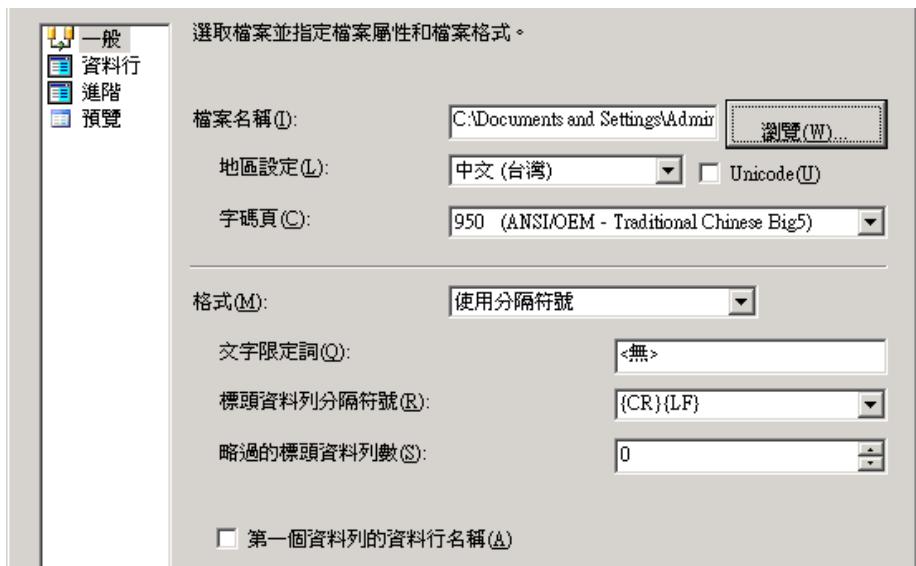


圖5.3 一般檔案連接管理員設定

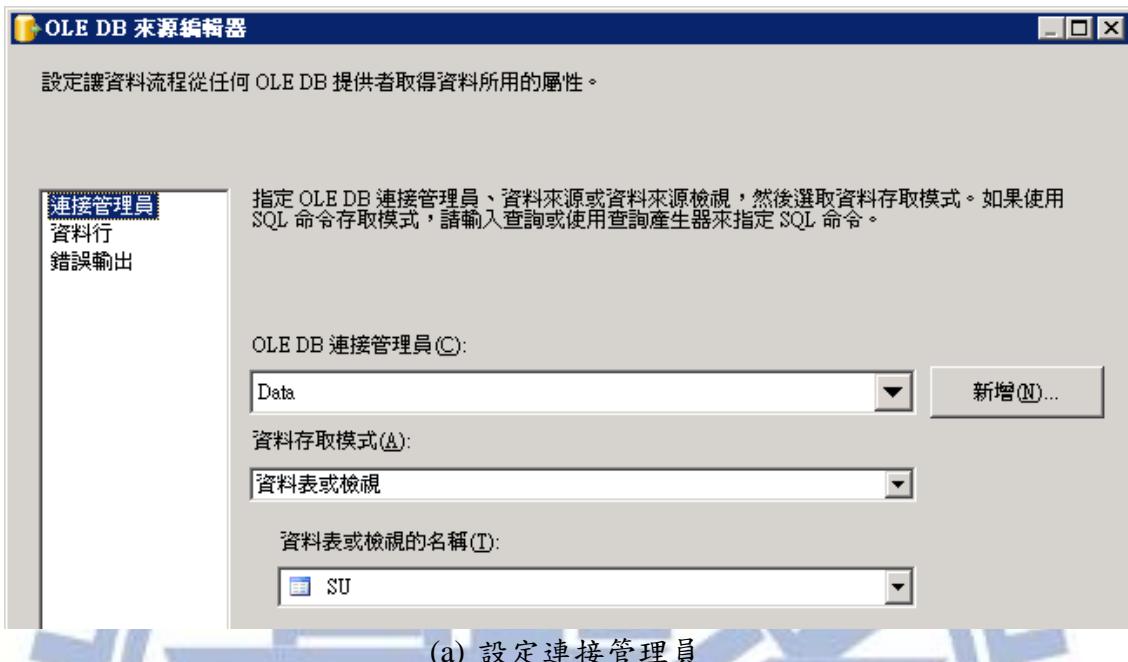
### 5.1.2 SSIS 資料流程工作設定

SSIS 提供了相當多樣化的資料流程工作，以下僅說明資料庫封裝與資料超市封裝所使用的資料流程工作如何設定，常用的資料流程工作如表 5.1。

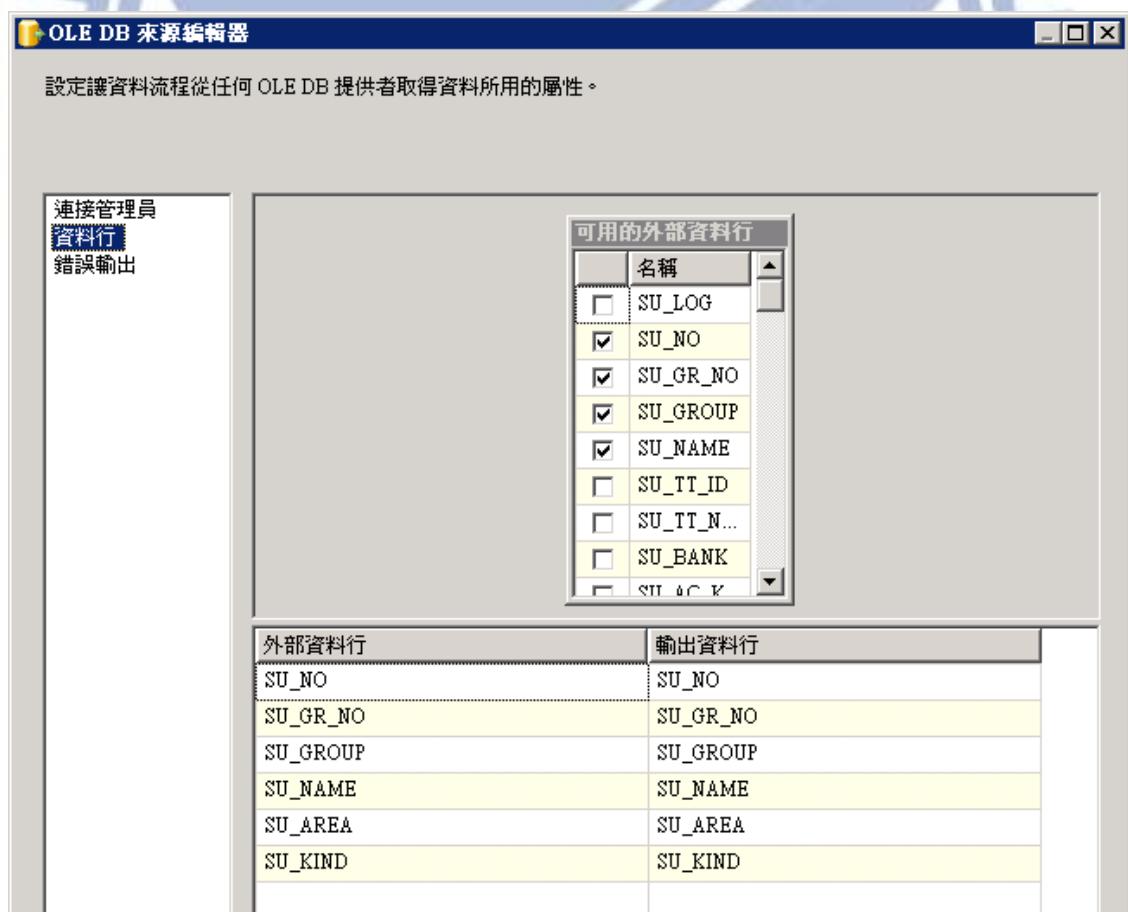
表5.1 SSIS 資料流程工作

圖示	工作名稱	功能說明
	OLE DB 來源	從關聯式資料庫擷取資料，對象不必限定為微軟公司的 SQL Server。
	資料轉換	將資料轉換為不同型別，並視為新的資料行輸出。
	查閱	將輸入資料與參考資料利用相等聯結(Equi-Join)[30]進行完整比對。
	條件式分割	利用條件限制，將輸入資料導向多個輸出。
	聯集全部	合併多個資料輸入。
	衍生的資料行	使用運算式更新資料的值或新增資料。
	緩時變維度	更新緩慢改變的維度。
	OLE DB 命令	針對每一資料列執行 SQL 命令。
	OLE DB 目的地	將資料匯入關聯式資料庫，對象不必限定為微軟公司的 SQL Server。

「OLE DB 來源」透過連接管理員連結至不同資料來源，以取得欲轉換的資料，如圖 5.4(a)。並可從來源資料中直接選取欲使用的資料行進行輸出，以減少後續轉換處理的資料量，如圖 5.4(b)所示。



(a) 設定連接管理員



(b) 選取資料行

圖5.4 OLE DB 設定

「資料轉換」處理資料的型別轉換，並輸出為新的資料列。從圖 5.5(a)可知輸入資料的型別為 DT\_WSTR，也就是 unicode 編碼字串。在「資料轉換」中選取欲轉換型別的資料，並更改其資料類型與輸出別名，如圖 5.5(b)，則輸出後會新增型別轉換後的資料，如圖 5.5(c)。

名稱	資料類型	有效位數	小數位數	長度
SU_NO	DT_WSTR	0	0	5
SU_GR_NO	DT_WSTR	0	0	5
SU_GROUP	DT_WSTR	0	0	30
SU_NAME	DT_WSTR	0	0	30
SU_AREA	DT_WSTR	0	0	3
SU_KIND	DT_WSTR	0	0	1

(a) 輸入資料型別

資料轉換編輯器

設定用來將輸入資料行的資料類型轉換為不同資料類型的屬性。視資料行要轉換成什麼資料類型而定，請設定資料行的長度、有效位數、小數位數及字碼頁。

可用的輸入資料行

名稱
<input checked="" type="checkbox"/> SU_NO
<input checked="" type="checkbox"/> SU_GR_NO
<input type="checkbox"/> SU_GROUP
<input type="checkbox"/> SU_NAME
<input checked="" type="checkbox"/> SU_AREA
<input checked="" type="checkbox"/> SU_KIND

輸入資料行	輸出別名	資料類型	長度	有效位數	小數位數	字碼頁
SU_NO	Code	字串 [DT_STR]	5			950 (ANSI)
SU_GR_NO	GroupCode	字串 [DT_STR]	5			950 (ANSI)
SU_AREA	Region	字串 [DT_STR]	2			950 (ANSI)
SU_KIND	Type	單一位元組不帶正負...				

(b) 資料轉換設定

名稱	資料類型	有效位數	小數位數	長度
SU_NO	DT_WSTR	0	0	5
SU_GR_NO	DT_WSTR	0	0	5
SU_GROUP	DT_WSTR	0	0	30
SU_NAME	DT_WSTR	0	0	30
SU_AREA	DT_WSTR	0	0	3
SU_KIND	DT_WSTR	0	0	1
Code	DT_STR	0	0	5
GroupCode	DT_STR	0	0	5
Region	DT_STR	0	0	2
Type	DT_UI1	0	0	0

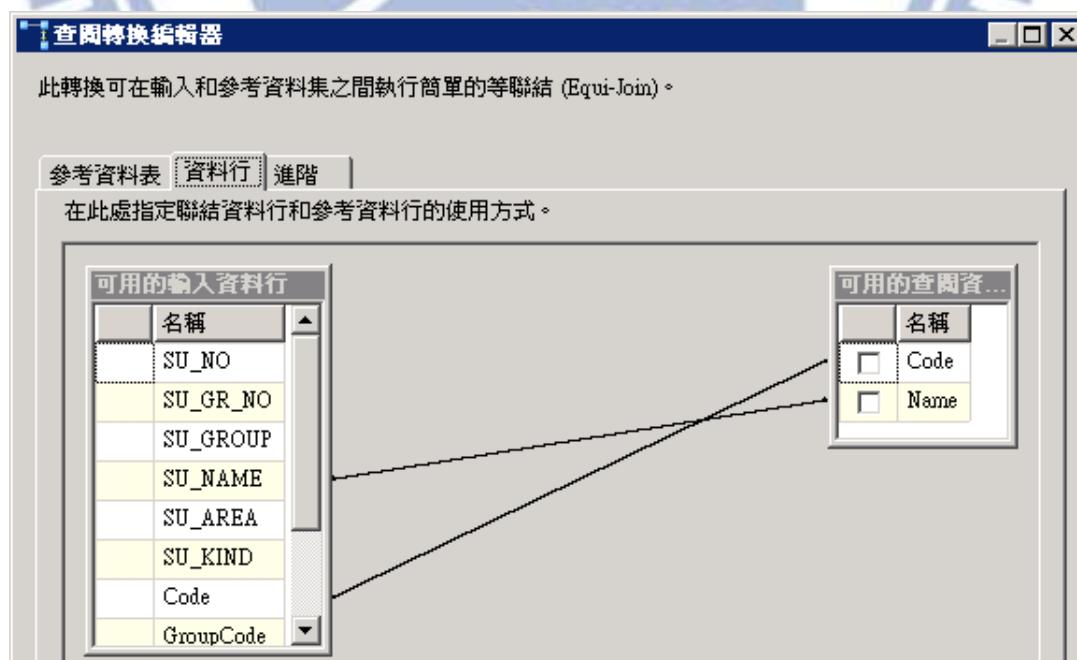
(c) 輸出資料型別

圖5.5 資料轉換設定

「查閱」比對輸入資料與參考資料的差異，並將相符與不符的資料分開輸出。以供應單位資料為例，首先連接資料庫，並利用 SQL 語法選取供應單位資料表的代碼與名稱欄位作為參考資料，如圖 5.6(a)；再建立參考資料與輸入資料的欄位對應，如圖 5.6(b)，即可利用欄位分辨相符與不符的資料。



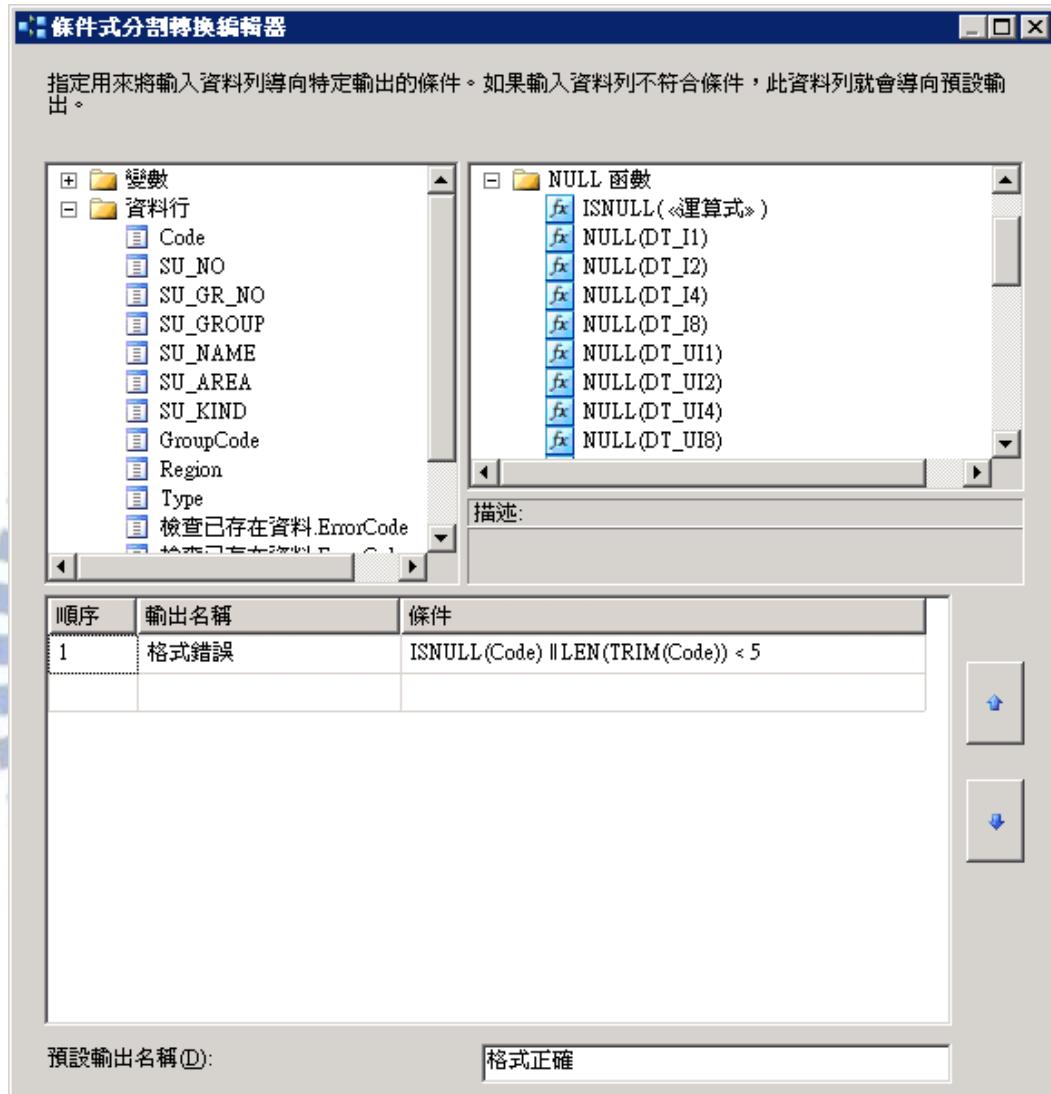
(a) 設定參考資料表



(b) 設定輸入資料與參考資料對應

圖5.6 查閱設定

「條件式分割」利用條件式進行判斷，將輸入資料根據不同條件導向多個不同輸出。條件式由資料行與 SSIS 提供的函數組成，如圖 5.7(a)所示。圖 5.7(a)目的為判斷供應代碼格式是否正確，其格式錯誤的條件有兩項：(1)代碼為空值(2)代碼位數小於五碼。只要符合其中一項條件，即被分類至「格式錯誤」輸出；兩項條件都不符合的資料，則被分類至「格式正確」輸出，輸出結果如圖 5.7 所示。



(a) 條件式輸入畫面



(b) 條件式分割輸出結果

圖5.7 條件式分割設定

「聯集全部」將多個輸入合併為一個輸出，如圖 5.8，並可新增或刪除用不到的欄位，以及更改輸出資料行的名稱。若某個輸入缺少對應的輸出資料行，則會自動補上空值。

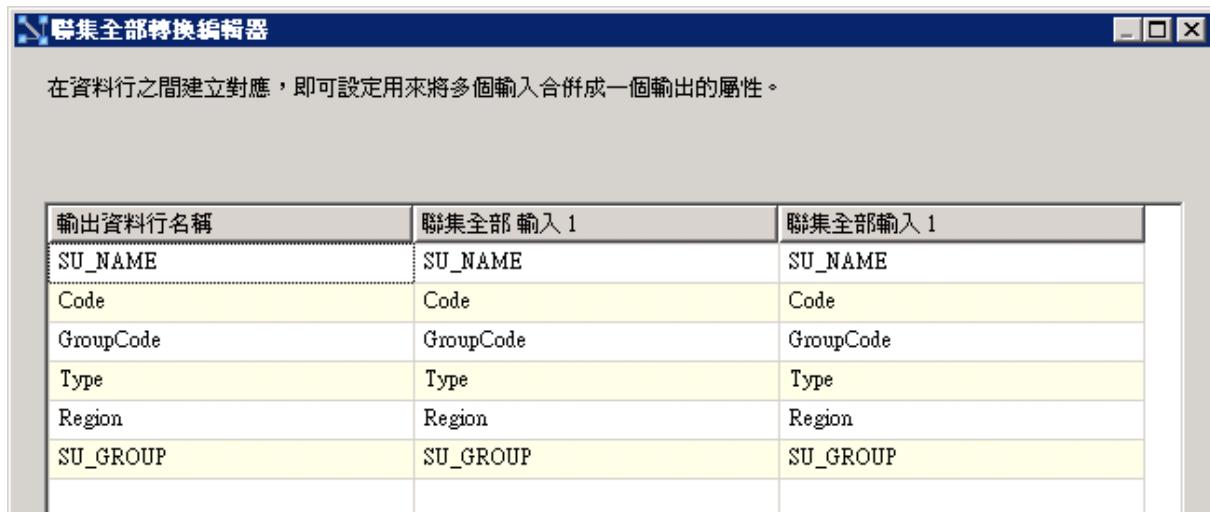


圖 5.8 聯集全部設定

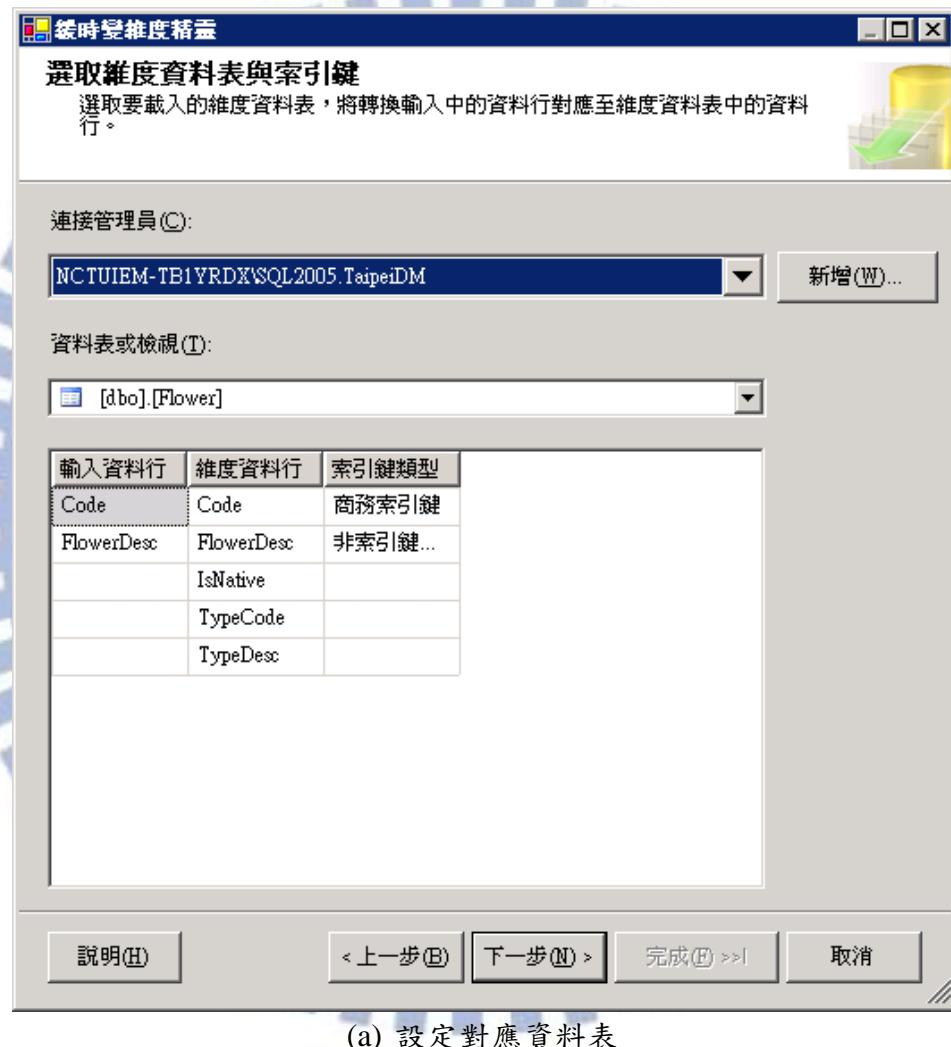
「衍生的資料行」可新增資料行並根據運算式加入值，也可利用運算式取代原有資料行的值，然而取代的情況下無法更改資料行的型別或長度，新增的資料行則可修改。運算式由資料行與函數組成，如圖 5.9。



圖 5.9 衍生的資料行設定

「緩時變維度」比對現有資料表與輸入資料，針對有資料更動的欄位，可選擇選擇覆蓋舊資料或新增資料列等緩慢改變維度處理方式。其設定方式如下：首先選擇欲更新的資料表，將輸入資料與資料表的欄位進行對應，並決定「商務索引鍵」以作為資料對應的參考，如圖 5.10(a)。

對於對應的資料行，可選擇以下三種處理方式：(1)固定屬性，表示該欄位的值不允許變更(2)變更屬性，表示以輸入資料行的值覆蓋資料表(3)歷程記錄屬性，表示將有更動的資料儲存為新的資料列，並在舊的資料列進行標記。最後「緩時變維度」工作將會根據不同的處理方式產生多個輸出，如圖 5.10(b)。



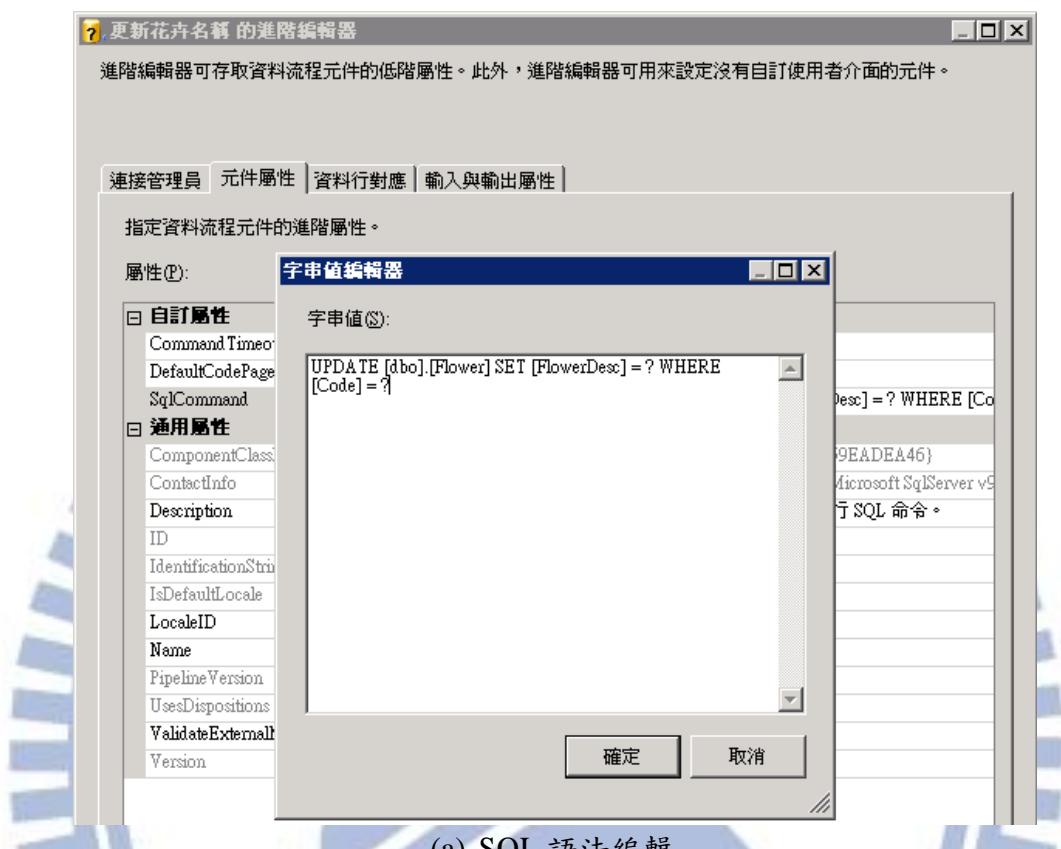
(a) 設定對應資料表



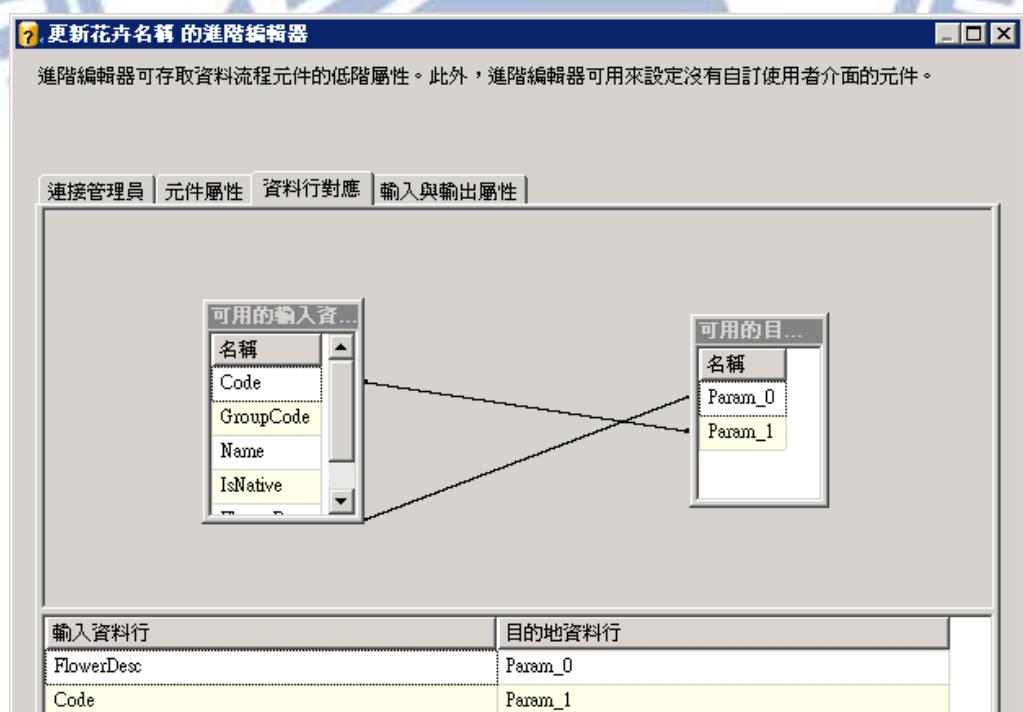
(b) 緩時變維度輸出

圖5.10 緩時變維度設定

「OLE DB 命令」可在指定的資料庫內執行 SQL 命令，並利用輸入資料產生值。如所示，在 SQL 語法中加入問號，表示該數值由輸入資料提供，再透過資料行對應數值，即可對資料表執行指令，並使用輸入資料作為數值參考。



### (a) SQL 語法編輯



(b) 資料行對應

圖5.11 OLE DB 命令設定

「OLE DB 目的地」可將轉換好的資料匯入 SQL Server 資料庫。與「OLE DB 來源」相同，「OLE DB 目的地」也必須先設定連接管理員，接著進行輸入資料與資料庫欄位的對應，如圖 5.12 所示。在對應時必須注意輸入資料行與目的地資料行的資料型態是否相同，不相容的型別會導致錯誤警告。



圖5.12 OLE DB 目的地設定

## 5.2 資料庫封裝再設計與實作

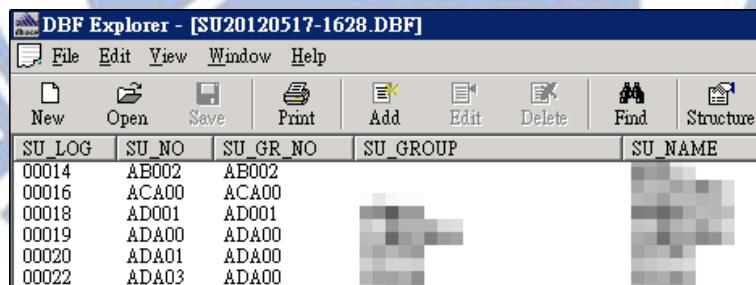
花卉資料庫的目的在於存放清理過的花卉資料，以供資料超市直接取用，因此資料庫封裝最重要的功能即是進行資料的清理與篩選，並配合第四章提出的維度再設計進行 ETL 再設計。由於花市提供資料的性質會影響資料庫封裝設計，本節將對花市原始資料進行分析，本節分為三小節，5.2.1 節說明花市提供原始資料特色；5.2.2 節提出資料庫封裝設計的標準流程，5.2.3 節以供應單位資料為例說明實作 ETL 方法。

### 5.2.1 花市原始資料

花市提供的原始資料分為四個部份：供應單位資料、承銷單位資料、花卉資料與日交易資料，檔名依序為 SU、BU、PR 與 DTyyymmdd，yy 為交易日的民國年份，mm 為月份，dd 為日期，例如民國 97 年 1 月 1 日的交易資料檔名為 DT970101。原始資料的檔案格式根據不同花市有所不同，台北花市使用 DBF 檔，屬於 dBASE 資料庫系統的檔案，如圖 5.13；其餘花市使用 TXT 文字檔。

原始資料中日交易資料為事實資料，其他三者則為不同的維度資料。維度資料更新頻率不定，因此花市在資料有所更動時才會傳送檔案，其內容為未更動的舊資料加上更改的新資料，因此在設計 ETL 時必須考慮資料是否會重複匯入的問題。

原始資料內的欄位又可分為兩種：必要欄位與次要欄位。必要欄位的資料如有遺漏，則必須向花市索取，無法利用其他欄位加以補齊，如「供應代碼」欄位即無法透過其他欄位得知；次要欄位的資料遺漏時可利用其他欄位加以補齊，如「供應地區代碼」欄位可參考供應代碼前兩碼。



The screenshot shows a DBF Explorer window titled "DBF Explorer - [SU20120517-1628.DBF]". The menu bar includes File, Edit, View, Window, and Help. The toolbar contains New, Open, Save, Print, Add, Edit, Delete, Find, and Structure buttons. The main area displays a table with columns: SU\_LOG, SU\_NO, SU\_GR\_NO, SU\_GROUP, and SU\_NAME. The SU\_NAME column contains blurred data. The data in the SU\_LOG column is as follows:

SU_LOG	SU_NO	SU_GR_NO	SU_GROUP	SU_NAME
00014	AB002	AB002		
00016	ACA00	ACA00		
00018	AD001	AD001		
00019	ADA00	ADA00		
00020	ADA01	ADA00		
00022	ADA03	ADA00		

圖 5.13 台北花市原始資料

### 5.2.2 資料庫封裝設計標準流程

根據 5.2.1 節整理的原始檔案資訊，資料庫封裝中維度資料轉換的標準流程如圖 5.14。首先匯入原始資料，由於原始資料匯入後預設為字串型別，必須進行資料型別轉換，以方便後續作業。轉換後將原始資料與資料庫進行比對，過濾已存在的資料，並減少後續流程處理資料筆數，進而增加轉換速度。篩選出的新增資料首先進行必要欄位檢查，將必要欄位有所缺誤的資料移至例外資料表；再進行次要欄位檢查，並將空缺的欄位補齊。最後彙整所有資料，匯入資料庫。

事實資料轉換的標準流程與維度轉換大致相同，但較為簡化，其差異處有兩點：(1)事實資料不需與資料庫進行比對，因為每日的交易資料不會重複；(2)事實資料中所有欄位都是必要欄位。資料庫封裝事實資料轉換的標準流程設計如圖 5.15。

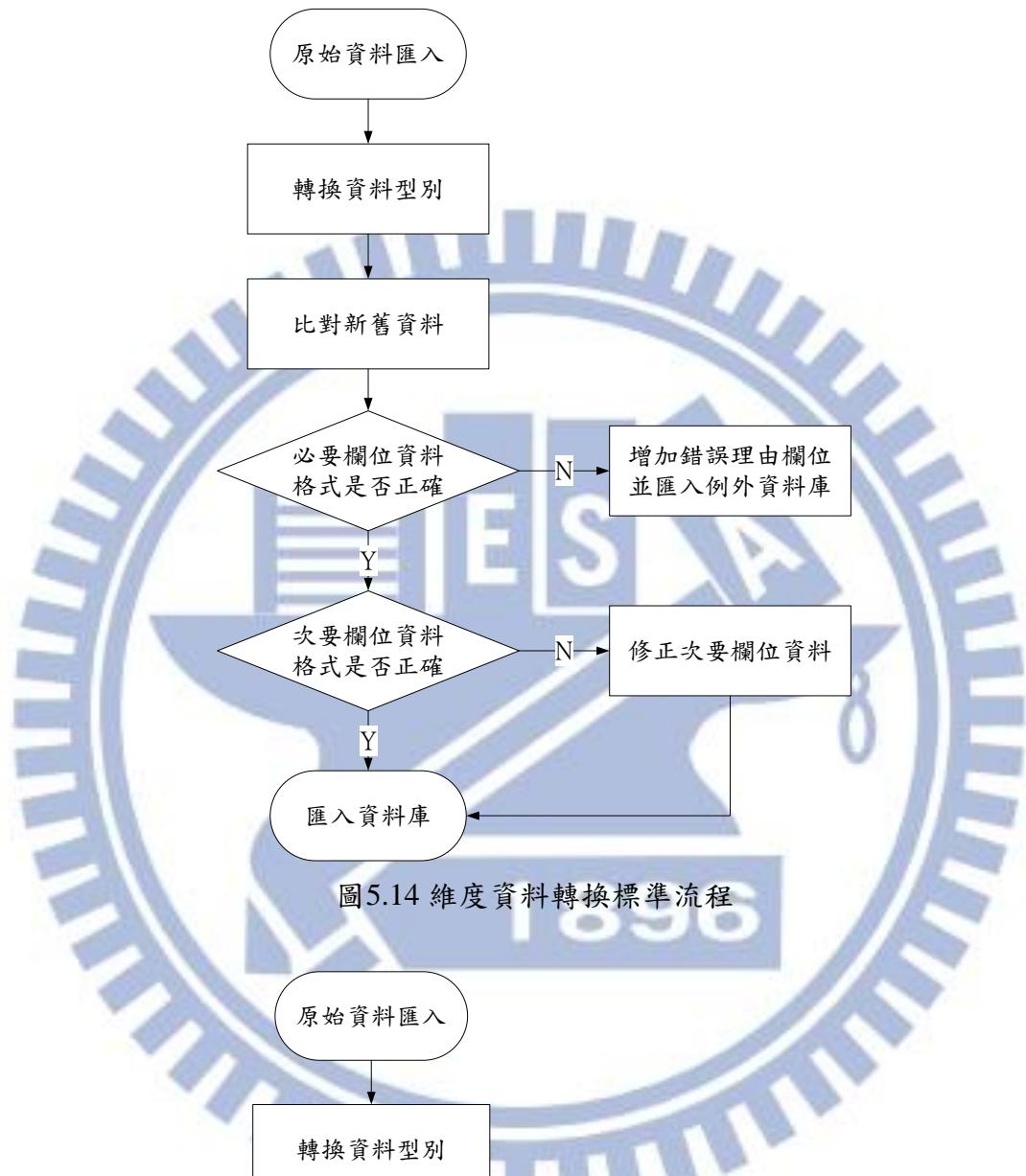


圖5.14 維度資料轉換標準流程

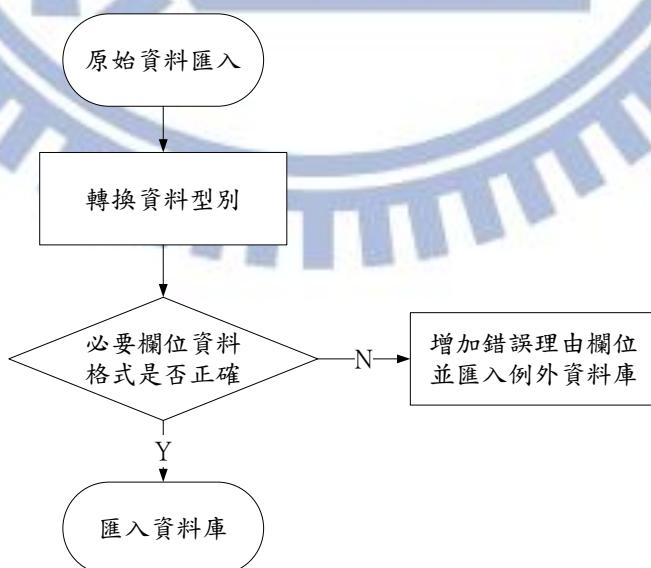


圖5.15 事實資料轉換標準流程

### 5.2.3 資料庫封裝實作

本節以台北花市的供應資料為例，說明資料庫封裝的實作，對於供應資料而言，在進行資料轉換前，依據 4.2 節提出的供應品牌辨識問題再設計，應先比對原始資料與資料超市的供應單位資料表，並將已結束的品牌進行標記。結束品牌標記的實作結果如圖 5.16，在更改品牌狀態的同時也加上結束供應的時間戳記，以配合 4.1 節提出的供應單位不連續問題再設計。

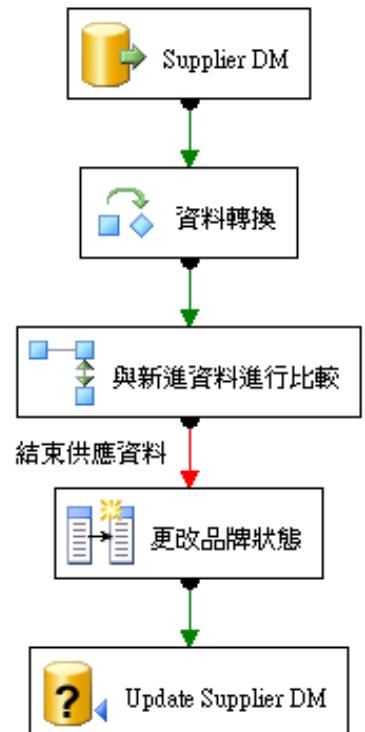


圖5.16 標記結束品牌實作結果

根據 5.2.2 節的設計標準流程，可將供應單位資料轉換分為兩個部份，第一部分處理資料型態轉換、比對新舊資料與必要欄位的篩檢，第二部份處理次要欄位的修正。

供應單位資料轉換第一部分的實作如圖 5.17 虛線上半部。由圖中可看出將原始資料匯入後，首先進行資料型別的轉換，再與資料表及例外資料表進行比對，篩選出新增的資料。接著依照供應代碼、供應名稱及供應類別的順序，進行必要欄位資料的檢查。當資料的必要欄位為空值或格式錯誤時，會將這些資料分離出來，增加欄位以記錄錯誤原因，並匯入例外資料表。供應單位轉換的第二部份實作如圖 5.17 虛線下半部，主要進行次要欄位的檢查與修正，依序為團體代碼及地區，最後匯入供應單位資料表。

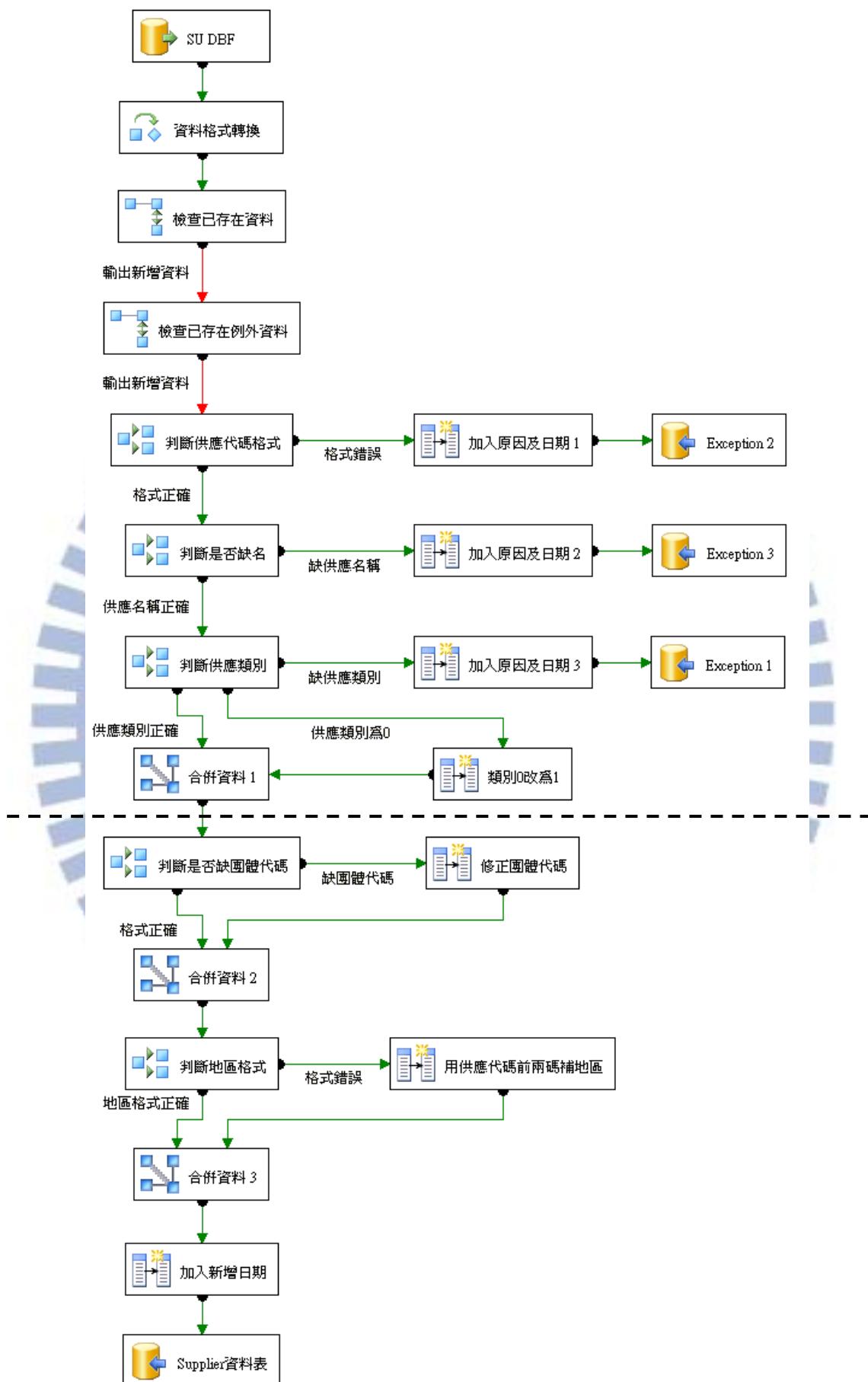


圖 5.17 供應單位轉換實作

為驗證供應單位資料轉換的結果，本研究設計了兩組測試資料，如表 5.2 與表 5.3。測試資料包括個別供應單位六筆，團體及所屬供應單位七筆，共十三筆。資料依序缺少特定欄位資料，以測試資料清理能力。第二組資料與第一組的差別在於代碼 AG004 與 AGA04 反映了繼承狀態，代碼 AG006 與 AGA06 品牌已結束，代碼 AG007 與 AGA07 則為新增的資料。表中的預期狀態代表根據前述判斷邏輯，所預期的資料流向。

表5.2 供應單位測試資料—第一組

供應代碼	供應團體代碼	供應名稱	供應地區	供應類別	供應型態	預期狀態
	AG001	鄧柏翰	AG	3	個別花農	進入例外資料表
AG002		許柏廷	AG	3	個別花農	進入資料表
AG003	AG003		AG	3	個別花農	進入例外資料表
AG004	AG004	黃典強		3	個別花農	進入資料表
AG005	AG005	周怡菁	AG		個別花農	進入例外資料表
AG006	AG006	蔡明珠	AG	3	個別花農	進入資料表
AGA00	AGA00	三星地區農會	AG	1	供應團體	進入資料表
	AGA00	蘇雅惠	AG	2	團體所屬花農	進入例外資料表
AGA02		洪冠伶	AG	2	團體所屬花農	進入資料表
AGA03	AGA00		AG	2	團體所屬花農	進入例外資料表
AGA04	AGA00	鄧志瑋		2	團體所屬花農	進入資料表
AGA05	AGA00	陳希富	AG		團體所屬花農	進入例外資料表
AGA06	AGA00	林宜琪	AG	2	團體所屬花農	進入資料表

\*表內姓名皆由姓名產生器產生

表5.3 供應單位測試資料—第二組

供應代碼	供應團體代碼	供應名稱	供應地區	供應類別	供應型態	預期狀態
	AG001	鄧柏翰	AG	3	個別花農	不重複進入資料表
AG002		許柏廷	AG	3	個別花農	不重複進入資料表
AG003	AG003		AG	3	個別花農	不重複進入資料表
AG004	AG004	黃郁婷		3	個別花農	進入資料表
AG005	AG005	周怡菁	AG		個別花農	不重複進入資料表
AG007	AG007	邱婷名	AG	3	個別花農	進入資料表
AGA00	AGA00	三星地區農會	AG	1	供應團體	不重複進入資料表
	AGA00	蘇雅惠	AG	2	團體所屬花農	不重複進入資料表
AGA02		洪冠伶	AG	2	團體所屬花農	不重複進入資料表
AGA03	AGA00		AG	2	團體所屬花農	不重複進入資料表
AGA04	AGA00	鄧振財		2	團體所屬花農	進入資料表
AGA05	AGA00	陳希富	AG		團體所屬花農	不重複進入資料表
AGA07	AGA00	曹柏宇	AG	2	團體所屬花農	進入資料表

Code	GroupCode	Name	Region	Type
AG002	AG002	許柏廷	AG	3
AG004	AG004	黃典強	AG	3
AG006	AG006	蔡明珠	AG	3
AGA00	AGA00	三星地區農會	AG	1
AGA02	AGA00	洪冠伶	AG	2
AGA04	AGA00	鄧志偉	AG	2
AGA06	AGA00	林宜琪	AG	2

(a) 供應單位資料表結果

Code	GroupCode	Name	Region	Type	Reason
NULL	AG001	鄧柏翰	AG	3	供應代碼格式錯誤
AG003	AG003	NULL	AG	3	缺供應名稱
AG005	AG005	周怡菁	AG	NULL	缺供應類別
NULL	AGA00	蘇雅惠	AG	2	供應代碼格式錯誤
AGA03	AGA00	NULL	AG	2	缺供應名稱
AGA05	AGA00	陳希富	AG	NULL	缺供應類別

(b) 例外資料表結果

圖 5.18 第一組測試資料實作結果

以第一組測試資料得到的實際結果如圖 5.18(a)(b)，可看出與預期的結果相符。而第二組測試資料得到的結果如圖 5.19(a)(b)，符合預期結果，也可由圖 5.19(b)中看出代碼 AG006 與 AGA06 的供應單位於「StopBrand」欄位標記已結束品牌，且「EndDate」欄位也更改為結束日期。由此可知資料庫封裝中供應單位資料可成功運作。

Code	GroupCode	Name	Region	Type
AG002	AG002	許柏廷	AG	3
AG004	AG004	黃典強	AG	3
AG004	AG004	黃郁婷	AG	3
AG006	AG006	蔡明珠	AG	3
AG007	AG007	邱婷名	AG	3
AGA00	AGA00	三星地區農會	AG	1
AGA02	AGA00	洪冠伶	AG	2
AGA04	AGA00	鄧志偉	AG	2
AGA04	AGA00	鄧振財	AG	2
AGA06	AGA00	林宜琪	AG	2
AGA07	AGA07	曹柏宇	AG	2

(a) 供應單位資料表結果

ID	SupplierCode	CurrentSupplierDesc	StopBrand	EndDate
5	AG002	AG002 許柏廷	0	2030/1/1 上午 12:00:00
6	AG004	AG004 黃典強	0	2030/1/1 上午 12:00:00
7	AG006	AG006 蔡明珠	1	2012/3/9 下午 07:51:06
1	AGA00	AGA00 三星地區農會	0	2030/1/1 上午 12:00:00
2	AGA02	AGA02 洪冠伶	0	2030/1/1 上午 12:00:00
3	AGA04	AGA04 鄧志偉	0	2030/1/1 上午 12:00:00
4	AGA06	AGA06 林宜琪	1	2012/3/9 下午 07:51:06

(b) 供應單位維度結果

圖 5.19 第二組測試資料實作結果

## 5.3 資料超市封裝再設計與實作

資料超市封裝的目的在於完成維度資料表的建立，也就是利用資料庫的資料，將維度資料表的欄位填滿。因此資料超市封裝的轉換設計原則較資料庫封裝來得簡單，即是依序建立各欄位值。然而因各維度資料表需求不同，在實作上反而較為複雜。本節以供應單位維度為例，說明資料超市封裝的實作過程。供應單位維度轉換實作的重點有二：(1)供應團體的對應(2)繼承狀態的判斷，以下將分別說明兩者的實作方法。

供應單位在性質上可分為三類，分別是：(a)獨立存在的供應單位(b)屬於某個供應團體的供應單位(c)供應團體。對於原始資料而言，(a)與(c)的供應代碼與團體供應代碼相同，如表 5.2 的前七列；(b)的供應代碼與團體代碼則不同。對於供應單位維度的「供應團體名稱」欄位而言，(a)和(c)的供應團體名稱等於本身供應單位名稱，然而(b)的供應團體名稱須對應至(c)的供應單位名稱。這造成「供應團體名稱」欄位來源複雜，可能來自資料本身、其他資料列或已存在於維度資料表內的資料。

因此在實作上本研究採取統一來源的做法，先新增「供應團體資料表」，僅存放供應代碼與名稱，並於供應單位維度轉換前，新增另一轉換將(a)與(c)的資料轉入，如圖 5.20。之後轉換供應單位維度資料時，就以此表作為「查閱」的參考資料表，並匯入資料以作為「供應團體名稱」欄位來源，如圖 5.21 下方框內所示。

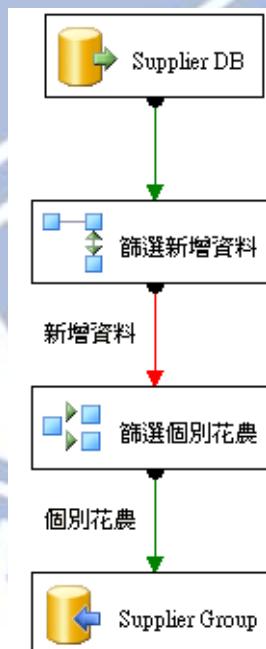
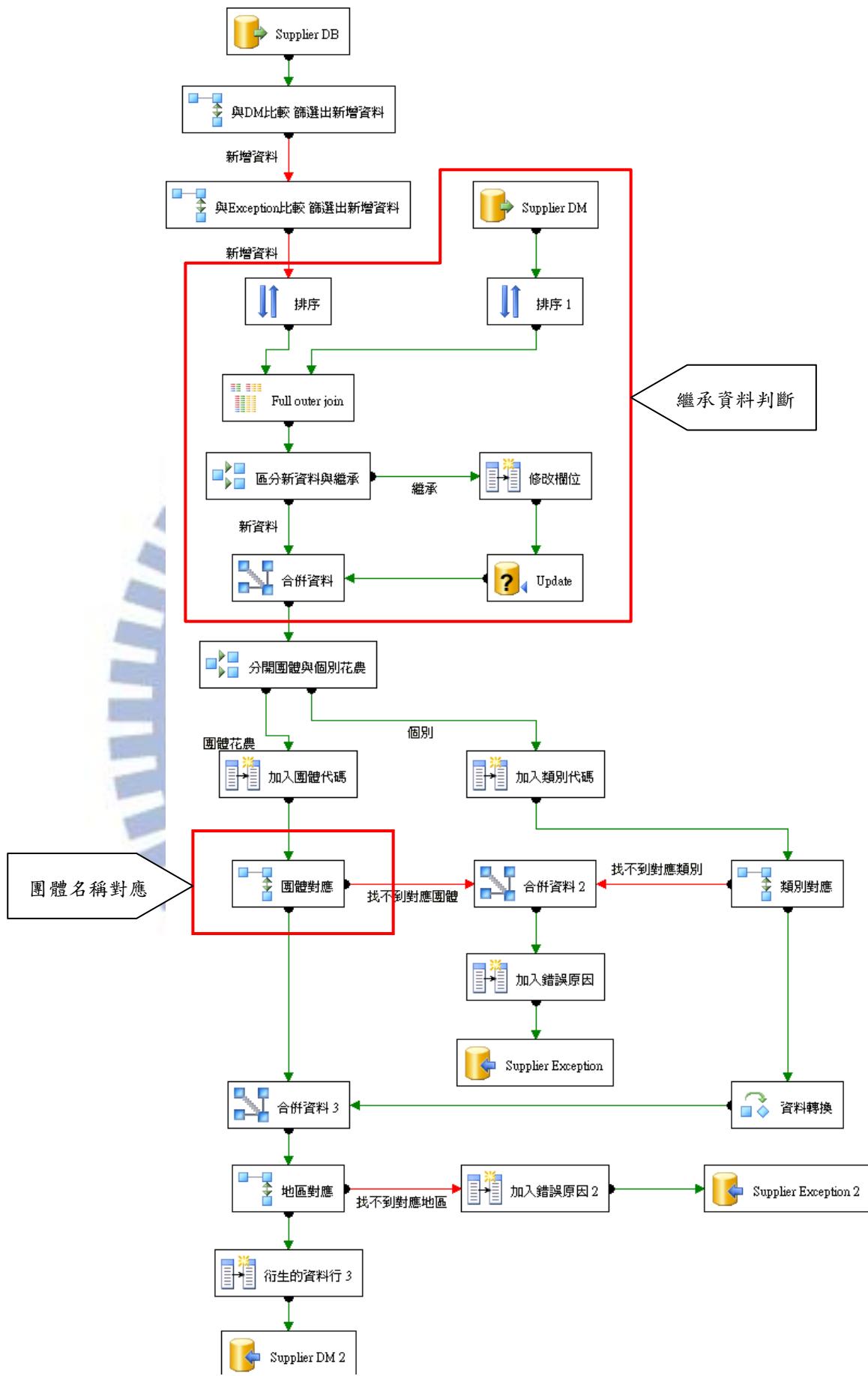


圖5.20 新增供應團體流程

繼承狀態的判斷是透過供應代碼連接(Join[30])維度資料表，當符合以下條件：(1)成功連接，即維度資料表存在相同供應代碼的資料(2)供應名稱不相等(3)「結束品牌」欄位不為1，即可判斷此資料為繼承資料。實作如圖 5.21 上方框內所示，將維度內舊的供應單位加上時間戳記表示供應已終止，並新增新的供應單位。完整的供應單位維度轉換過程如圖 5.21。



為驗證供應單位維度轉換的結果，同樣採用表 5.2 與表 5.3 的資料進行測試。第一組測試資料得到的實際結果如圖 5.22(a)，可看出圖 5.18(a)的七筆資料全數轉入。而第二組測試資料轉換結果如圖 5.22(b)，可由圖中看出代碼 AG004 與 AGA04 的資料各有兩筆，其中各有一筆更新了結束日期，但並未標記品牌結束，且兩筆資料的現任供應單位欄位 (CurrentSupplierDesc)相同，代表下一任使用者為繼承；而代碼 AG007 與 AGA07 的資料也成功新增。由此可知資料超市封裝中供應單位維度轉換可成功運作。

ID	TypeCode	SupplierCode	TypeDesc	CurrentSupplierDesc	SupplierDesc	StopBrand	StartDate	EndDate
5	00003	AG002	個人花農	AG002 許柏廷	許柏廷	0	2012/3/1 上午 12:00:00	2030/1/1 上午 12:00:00
6	00003	AG004	個人花農	AG004 黃典強	黃典強	0	2012/3/1 上午 12:00:00	2030/1/1 上午 12:00:00
7	00003	AG006	個人花農	AG006 蔡明珠	蔡明珠	0	2012/3/1 上午 12:00:00	2030/1/1 上午 12:00:00
1	AGA00	AGA00	三星地區農會	AGA00 三星地區農會	三星地區農會	0	2012/3/1 上午 12:00:00	2030/1/1 上午 12:00:00
2	AGA00	AGA02	三星地區農會	AGA02 洪冠伶	洪冠伶	0	2012/3/1 上午 12:00:00	2030/1/1 上午 12:00:00
3	AGA00	AGA04	三星地區農會	AGA04 鄭志偉	鄭志偉	0	2012/3/1 上午 12:00:00	2030/1/1 上午 12:00:00
4	AGA00	AGA06	三星地區農會	AGA06 林宜琪	林宜琪	0	2012/3/1 上午 12:00:00	2030/1/1 上午 12:00:00

(a) 第一組資料供應維度轉換結果

ID	TypeCode	SupplierCode	TypeDesc	CurrentSupplierDesc	SupplierDesc	StopBrand	StartDate	EndDate
5	00003	AG002	個人花農	AG002 許柏廷	許柏廷	0	2012/3/1 上午 12:00:00	2030/1/1 上午 12:00:00
6	00003	AG004	個人花農	AG004 黃郁婷	黃典強	0	2012/3/1 上午 12:00:00	2012/3/9 下午 08:03:22
19	00003	AG004	個人花農	AG004 黃郁婷	黃郁婷	0	2012/3/10 下午 08:03:22	2030/1/1 上午 12:00:00
7	00003	AG006	個人花農	AG006 蔡明珠	蔡明珠	1	2012/3/1 上午 12:00:00	2012/3/9 下午 07:51:06
18	00003	AG007	個人花農	AG007 邱婷名	邱婷名	0	2012/3/10 下午 08:03:22	2030/1/1 上午 12:00:00
1	AGA00	AGA00	三星地區農會	AGA00 三星地區農會	三星地區農會	0	2012/3/1 上午 12:00:00	2030/1/1 上午 12:00:00
2	AGA00	AGA02	三星地區農會	AGA02 洪冠伶	洪冠伶	0	2012/3/1 上午 12:00:00	2030/1/1 上午 12:00:00
3	AGA00	AGA04	三星地區農會	AGA04 鄭振財	鄭志偉	0	2012/3/1 上午 12:00:00	2012/3/9 下午 08:03:22
17	AGA00	AGA04	三星地區農會	AGA04 鄭振財	鄭振財	0	2012/3/10 下午 08:03:22	2030/1/1 上午 12:00:00
4	AGA00	AGA06	三星地區農會	AGA06 林宜琪	林宜琪	1	2012/3/1 上午 12:00:00	2012/3/9 下午 07:51:06

(b) 第二組資料供應維度轉換結果

圖5.22 供應維度轉換實作結果

# 第六章 結論與未來研究方向

## 6.1 結論

花卉產業利用花卉資料超市所儲存的大量歷史交易資料，可進行各種分析以協助決策的進行。2004 年曾發現供應單位維度有緩慢改變的現象，並進行了第一次再設計以修正此問題。然而分析花卉資料發現仍有錯誤出現，因此本研究針對資料超市所有維度進行了系統性分析，並得到以下結論：

### 一、花卉資料超市維度具有不同的緩慢改變型態

各維度的緩慢改變速度根據其儲存內容而有所不同，可分為以下三類：(1)顯性緩慢改變維度(2)隱性緩慢改變維度(3)不緩慢改變維度。顯性維度包括供應單位維度、花卉種類維度，由於這些維度資料變動性高，較容易發生緩慢改變現象，也較容易觀察到；隱性維度包括地區相關維度與其他花卉相關維度，其改變的速度相對緩慢，不易察覺；而時間相關維度則不會緩慢改變。

### 二、花卉資料超市的維度緩慢改變不會終止

由於花卉產業的持續變動，資料超市的維度緩慢改變將不會終止，且可能隨著產業變化而產生新的問題。因此資料超市與倉儲也必須持續維護，才能保持其可用性，不至與現實脫節。

本研究也針對現階段資料超市的問題加以解決，包括供應單位不連續與供應單位品牌辨識。針對供應單位不連續問題，本研究在供應單位維度加入具有時間戳記性質「開始日期」與「結束日期」欄位，記錄供應單位的供應期間。並做對應交易資料時的判斷條件。而實作上分為兩部份處理：在資料超市封裝新增供應資料時標記開始日期；於資料庫封裝處理結束品牌，及資料超市封裝處理繼承時加入結束日期。

針對供應單位品牌辨識問題，本研究於供應單位維度加入了「結束品牌」欄位。此欄位具有旗標的性質，以 0 與 1 表示品牌結束與否。實作上於資料庫封裝更新供應單位資料前先比對原始資料與維度資料，判斷品牌是否結束。針對已標記結束的供應單位資料，下一筆相同代碼的資料新增時，將不會更改其現任供應單位的值，以作為品牌之間的區隔，如表 6.1 所示。

表6.1 供應單位資料表範例

供應代碼	現任供應單位名稱	供應單位名稱	結束品牌
AA001	乙	甲	1
AA001	乙	乙	1
AA001	丁	丙	0
AA001	丁	丁	0

本研究透過維度與 ETL 的再設計，修正了現階段資料超市的維度緩慢改變問題，使線上分析系統得以提供正確結果輔助決策進行。未來可能產生的其他維度緩慢改變，則可成為後續研究發展的方向。

## 6.2 未來研究方向

本研究針對資料超市的維度與 ETL 系統進行了再設計，解決了現階段維度緩慢改變問題。未來研究可針對三個方向進行：其餘維度緩慢改變的修正、資料超市與線上分析系統的整合及花卉產業供應層面分析。以下為三個方向的詳述。

### 一、其餘緩慢改變維度的修正

本研究分析發現其餘維度也有緩慢改變的可能，且供應單位維度可能出現新的問題。針對不同問題提出新的再設計方法，為未來研究的方向之一。

### 二、資料超市與線上分析系統的整合

從資料中找尋有意義的資訊，是資料存在的價值。花卉資料超市已針對資料正確性進行改善，可供後續加值運用。然而經過維度再設計，資料超市的結構有所調整，也採用不同的技術實作，如何將資料整合進線上分析系統軟體，是相當值得探討的問題。

### 三、花卉產業供應層面分析

本研究於供應層面加入了品牌資訊，未來研究可針對此資訊進行分析。品牌之於花卉產業是否有其特殊意義，價格是否受到品牌影響，又或品牌的繼承是否影響供貨水準，都是可深入探討的議題。



## 參考文獻

- [1] 何致億，SQL Server 2005 資料庫管理，精誠資訊，2006 年。
- [2] 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與管理研究所碩士論文，2004 年。
- [3] 胡百敬、姚巧玲，SQL Server 2005 SSIS 整合服務，精誠資訊，2006 年。
- [4] 梁高榮，花卉產業利基發現系統，國立交通大學工業工程管理學系，2007 年。
- [5] 章立民，SQL Server 2000 資料轉換服務(DTS)，碁峰資訊股份有限公司，2001 年。
- [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與管理研究所碩士論文，2003 年。
- [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與管理研究所碩士論文，2003 年。
- [8] 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與管理研究所碩士論文，2006 年。
- [9] 陳家瑜，「台灣花卉供應鏈的資料倉儲設計與量測變數迴歸應用」，國立交通大學工業工程與管理研究所碩士論文，2006 年。
- [10] 張弘，「透過刀鋒伺服器及 SSIS 技術降低台灣花卉資料倉儲的維護成本」，國立交通大學工業工程與管理研究所碩士論文，2012 年。
- [11] 張堂穆，個人通信，2010 年。
- [12] 黃綉蓉，個人通信，2010 年。
- [13] 黃綉蓉，個人通信，2012 年。
- [14] 黃俊端，「設計與實作花卉批發資訊分享熱線的整合型審計資訊系統」，國立交通大學工業工程與管理研究所碩士論文，2006 年。
- [15] 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與管理研究所碩士論文，2004 年。
- [16] 溫師翰、梁高榮，「供應鏈資訊系統再工程問題的案例分析」，機械工業，五月，204-214 頁，2004 年。
- [17] 鄒柏瑜，dBase 5 for Windows 中文版程式設計，1995 年。
- [18] 農產品交易行情站，<http://amis.afa.gov.tw>。
- [19] 縣市改制直轄市資訊網，[http://www.moi.gov.tw/countryreform/news\\_content.aspx?id=1](http://www.moi.gov.tw/countryreform/news_content.aspx?id=1)。
- [20] 蘇石長，花卉分級包裝（七），台北花卉產銷股份有限公司，2001 年。
- [21] Cognos 商業智慧與財務績效管理，<http://www-01.ibm.com/software/tw/data/cognos/>。
- [22] Microsoft 技術支援：什麼是 MDAC、DA SDK、ODBC、OLE DB、ADO、RDS 以及 ADO/MD，<http://support.microsoft.com/kb/190463/zh-tw>。
- [23] SQL Server 7.0 以及 SQL Server 2000 的主流支援服務終止時間，<http://www.microsoft.com/taiwan/sql/solutions/upgrade/nonservices.mspx>。
- [24] Breslin, Mary, "Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models," Business Intelligence Journal, pp.6-20, Winter, 2004.
- [25] Codd, E. F., Codd, S. B., and Salley, C. T., "Providing OLAP to User-Analysts: An IT

- Mandate,” Codd & Date, Inc., 1993.
- [26] Haselden, Kirk, Microsoft® SQL Server 2005 Integration Services, Sams Publishing, 2006.
  - [27] Inmon, W. H., Building the Data Warehouse, John Wiley & Sons, Inc., 2002.
  - [28] Kimball, R., The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, John Wiley & Sons, Inc., 1996.
  - [29] Kimball, R., “Slowly Changing Dimensions,” DBMS, 1996.
  - [30] Kroenke, D. M. and Auer, D. J., Database Processing: Fundamentals, Design, and Implementation, 11th ed., Pearson Education, Inc., 2010.
  - [31] Mundy, J. and Thorntwaite, W., The Microsoft Data Warehouse Toolkit : with SQL Server 2005 and the Microsoft Business Intelligence Toolset, John Wiley & Sons, Inc., 2006.



## 附錄 A.1：修正後資料型別維度資料表

由於維度綱要的調整，維度資料表也需重新設計；並配合 SQL Server 2005 修改資料型別。本節附錄列出修改資料型別的各維度資料表。

表A.1 供應單位維度資料表

主鍵	欄位名稱	資料型別	長度	允許空值	說明
◎	SupplierID	int			供應單位主鍵
	IsGroupDesc	nvarchar	10		是否為供應團體之描述
	TypeCode	char	5		供應類別/供應團體代碼
	SupplierCode	char	5		供應單位代碼
	TypeDesc	nvarchar	30		供應類別/供應團體名稱
	CurrentlySupplierDesc	nvarchar	30		現任供應單位代號加名稱
	SupplierDesc	nvarchar	30		供應單位名稱
	CityCode	char	1		二級行政區代碼
	TownshipCode	char	2		三級行政區代碼
	CityDesc	nvarchar	30		二級行政區名稱
	TownshipDesc	nvarchar	30		三級行政區名稱
	StopBrand	tinyint			是否結束品牌
	StartDate	datetime			開始供應日期
	EndDate	datetime			結束供應日期

表A.2 花卉種類維度資料表

主鍵	欄位名稱	資料型別	長度	允許空值	說明
◎	FlowerID	int			花卉種類維度主鍵
	IsNative	nvarchar	5		是否為國產花
	TypeCode	char	5		花卉品類代碼
	TypeDesc	nvarchar	30		花卉品類名稱
	Code	char	5		花卉品名代碼
	FlowerDesc	nvarchar	30		花卉品名名稱

表A.3 花卉等級維度資料表

主鍵	欄位名稱	資料型別	長度	允許空值	說明
◎	FlowerClassID	int			花卉等級維度主鍵
	Quality	char	1		花卉等級代碼
	QualityDesc	nchar	5		花卉等級名稱
	Length	varchar	2		花卉等級與長度

表A.4 拍賣線維度資料表

主鍵	欄位名稱	資料型別	長度	允許空值	說明
◎	LineID	int			拍賣線維度主鍵
	LineDesc	nvarchar	5		拍賣線名稱

表A.5 容器維度資料表

主鍵	欄位名稱	資料型別	長度	允許空值	說明
◎	ContainerID	int			容器維度主鍵
	ContainerDesc	nvarchar	5		容器名稱

表A.6 承銷地區維度資料表

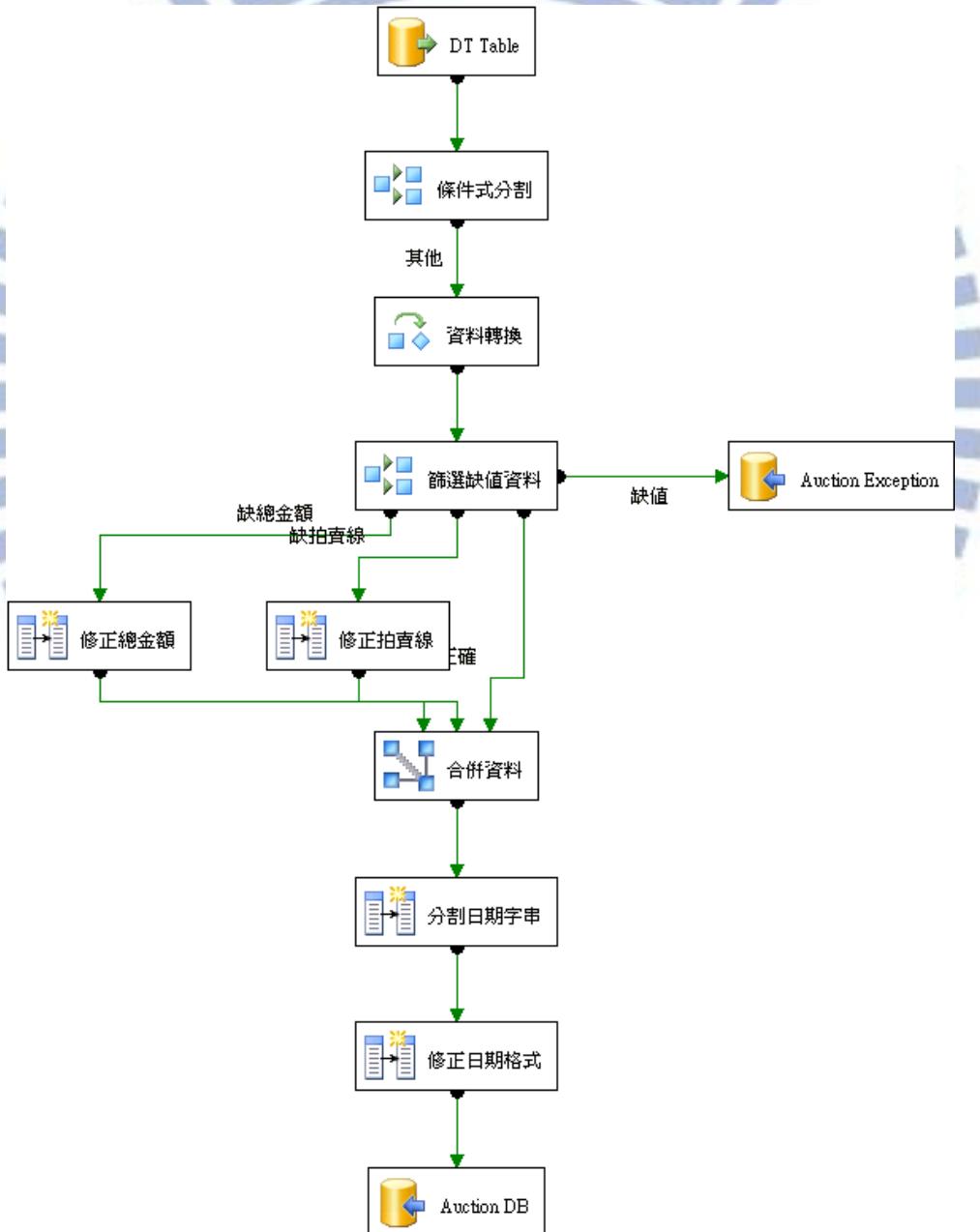
主鍵	欄位名稱	資料型別	長度	允許空值	說明
	BuyerID	int			承銷地區主鍵
	Code	char	5		承銷單位代碼
	Name	nverchar	30		承銷單位名稱
◎	Region	tinyint			承銷地區代碼
	RegionDesc	nvarchar	10		承銷地區名稱

## 附錄 A.2：資料庫封裝設計

資料庫封裝的供應單位轉換已於 5.2.3 節說明，本節附錄說明其餘轉換的設計，共分為兩小節：A.2.1 節說明其餘轉換的設計流程；A.2.2 節展示判斷原始資料是否存在的程式碼。

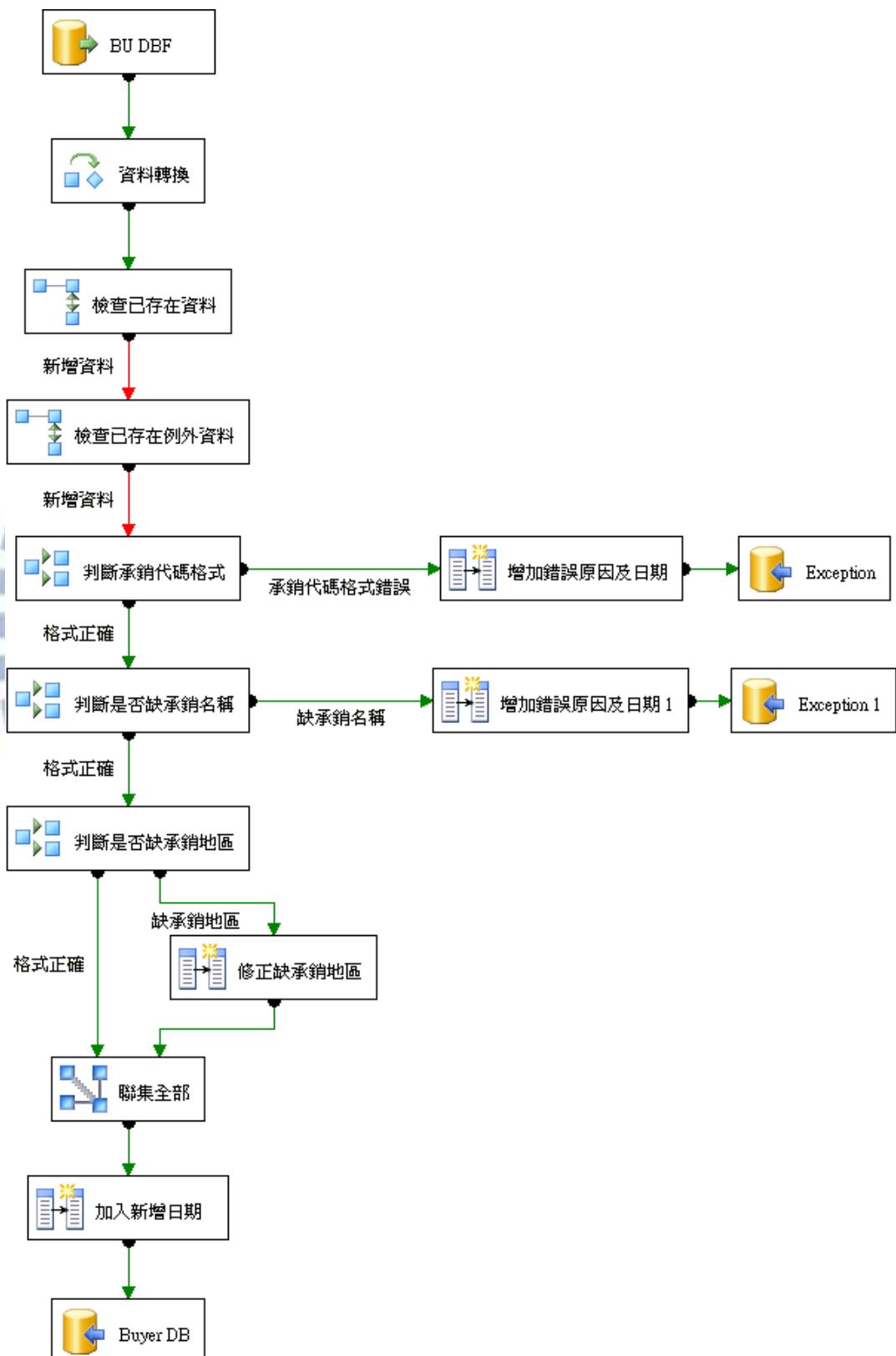
### A.2.1 資料庫封裝其餘資料轉換

本節附錄說明資料庫封裝中其餘轉換。圖 A.1 為花卉資料的轉換，首先排除進貨或取消交易的資料，並轉換資料型。接著判斷必要欄位的格式是否正確，其中未經過拍賣限的資料需要加以修正。最後修正日期格式，將民國轉為西元年，並匯入資料庫。



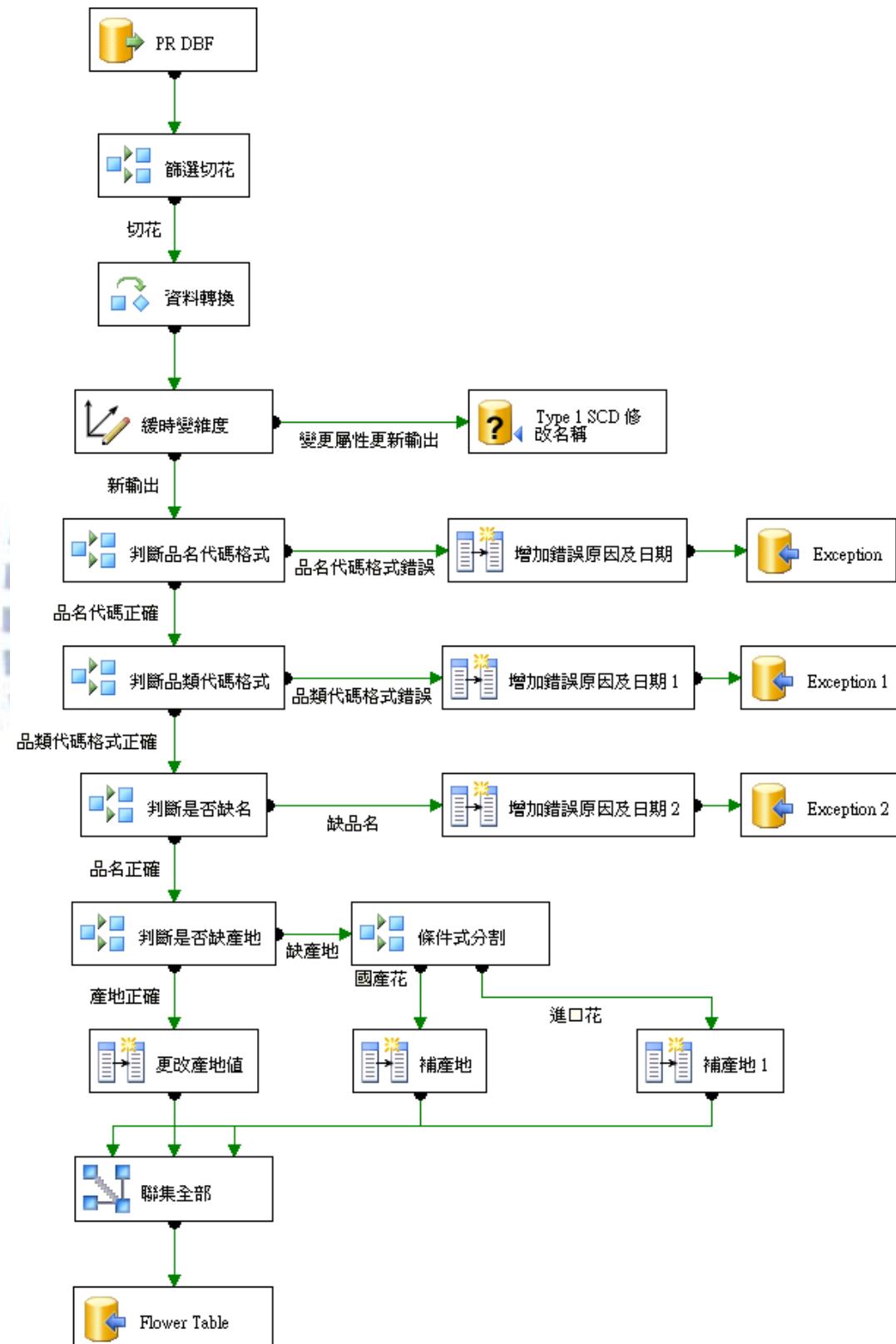
圖A.1 拍賣資料轉換

圖 A.2 為承銷資料的轉換，首先轉換資料型別，並與承銷資料表進行比對，將新增資料輸出。接著判斷必要欄位的格式是否正確，包括承銷代碼與承銷名稱。最後修正次要欄位，並匯入資料庫。



圖A.2 承銷資料轉換

圖 A.3 為花卉資料的轉換，首先篩選出切花，轉換資料型別，並更新名稱有所更改的花卉資料。接著判斷必要欄位的格式是否正確，包括品名代碼、品類代碼與花卉名稱。最後修正次要欄位，並匯入資料庫。



圖A.3 花卉資料轉換

## A.2.2 資料庫封裝判斷檔案指令碼

本節附錄說明在資料庫封裝中，檢查原始資料是否存在的程式碼。若檔案存在，即可繼續進行轉換工作；若檔案不存在則會執行失敗。

```
' Microsoft SQL Server Integration Services Script Task
' Write scripts using Microsoft Visual Basic
' The ScriptMain class is the entry point of the Script Task.

Imports System
Imports System.Data
Imports System.Math
Imports Microsoft.SqlServer.Dts.Runtime

Public Class ScriptMain

    ' The execution engine calls this method when the task executes.
    ' To access the object model, use the Dts object. Connections, variables, events,
    ' and logging features are available as static members of the Dts class.
    ' Before returning from this method, set the value of Dts.TaskResult to indicate success or failure.
    '
    ' To open Code and Text Editor Help, press F1.
    ' To open Object Browser, press Ctrl+Alt+J.

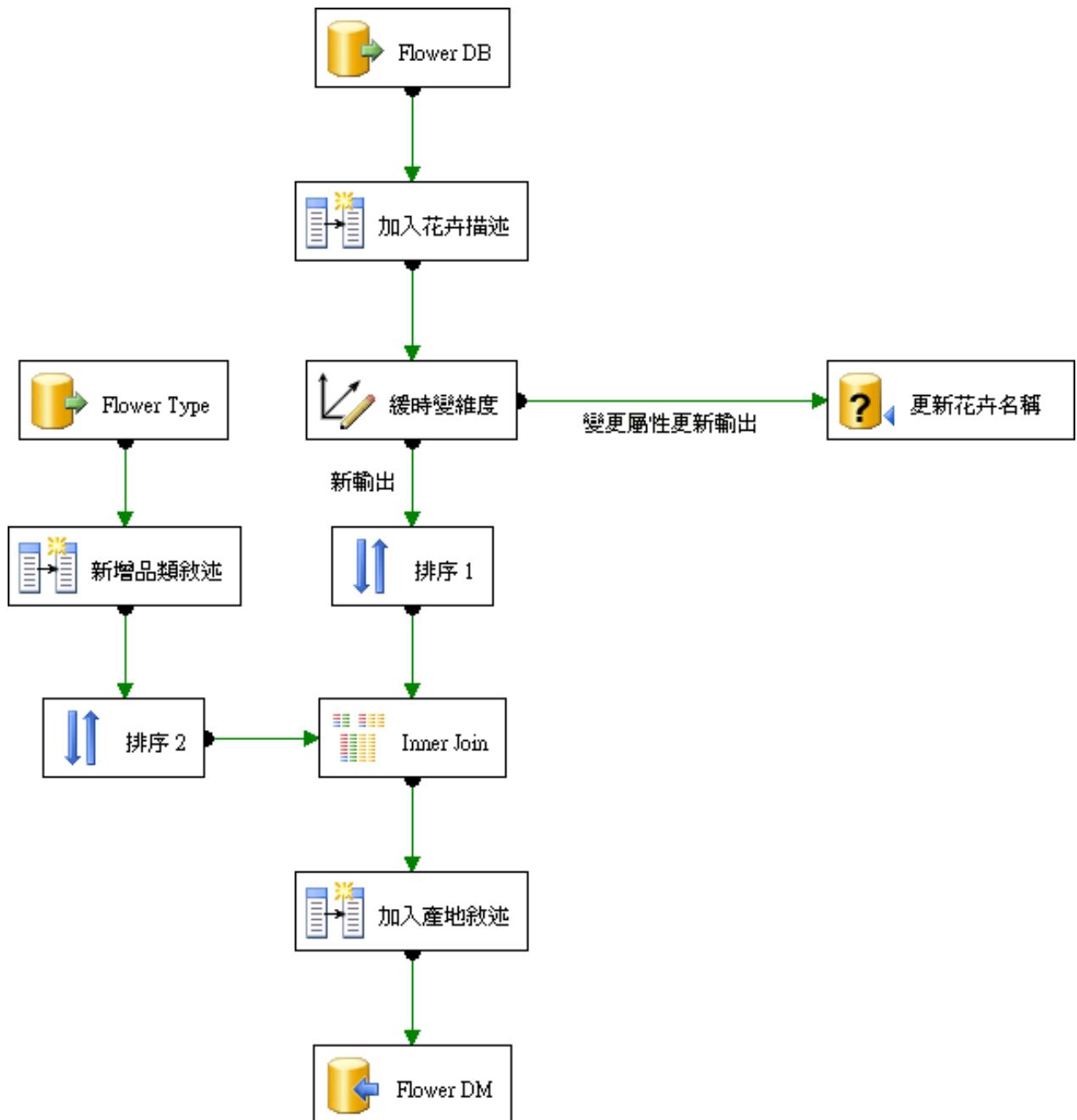
    Public Sub Main()
        '
        ' Add your code here
        '

        If My.Computer.FileSystem.FileExists("D://data/SU.DBF") Then
            Dts.TaskResult = Dts.Results.Success
        Else
            Dts.TaskResult = Dts.Results.Failure
        End If
    End Sub
End Class
```

圖A.4 判斷檔案程式碼

## 附錄 A.3：資料超市封裝設計

本節附錄說明資料超市封裝中其餘轉換。圖 A.5 為花卉種類維度的轉換，首先加入花卉名稱，並修正更新過的花卉資料。接著將花卉品類資料匯入後，利用內部連結(Inner Join[30])加入資料中，最後加入產地敘述，即可匯入資料超市。



圖A.5 花卉種類維度轉換

圖 A.6 為拍賣事實資料表的轉換，首先篩選出新增的資料，接著一一對應至維度資料，最後依照拍賣、訂貨、議價與殘貨分類交易資料，並匯入資料超市。

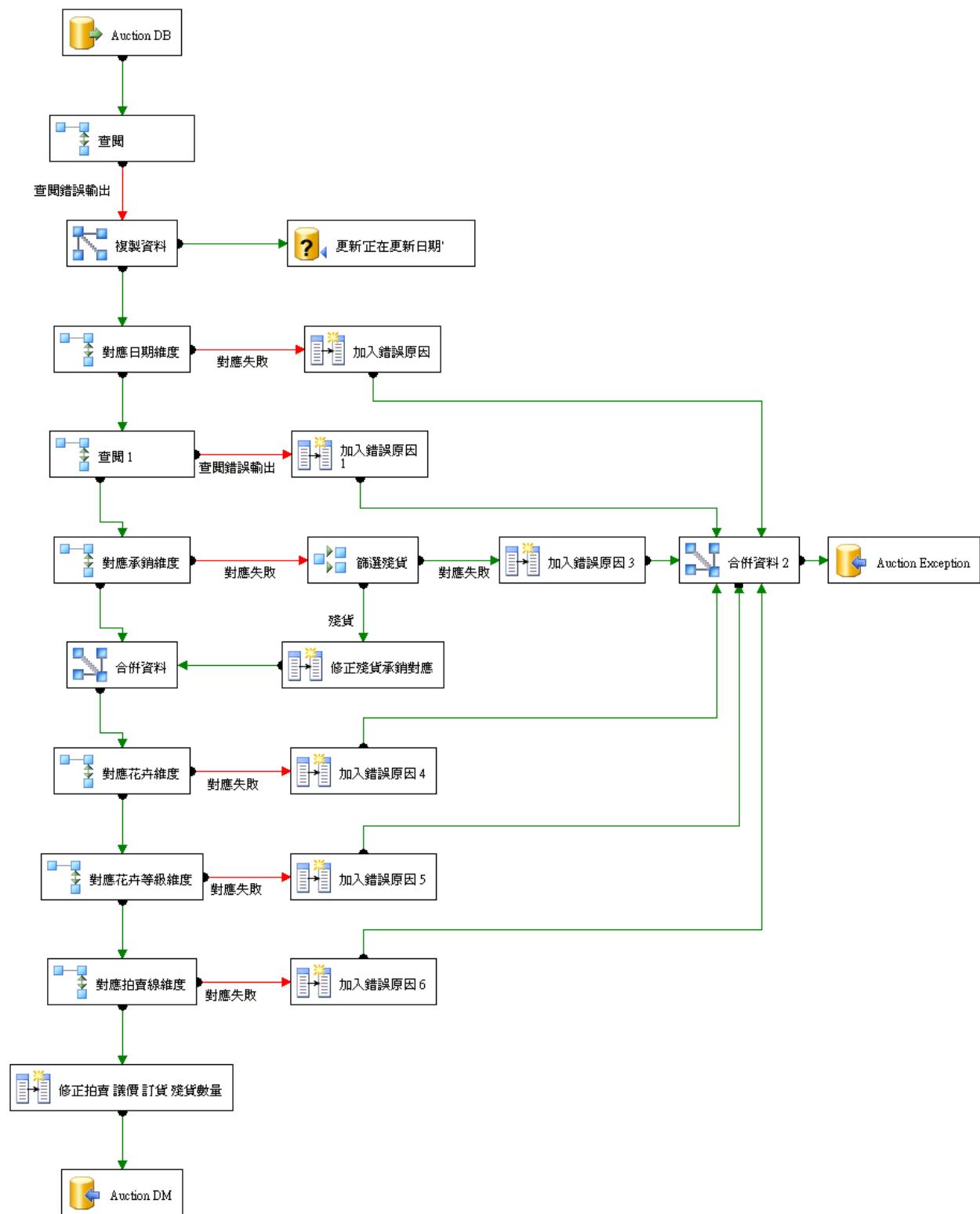
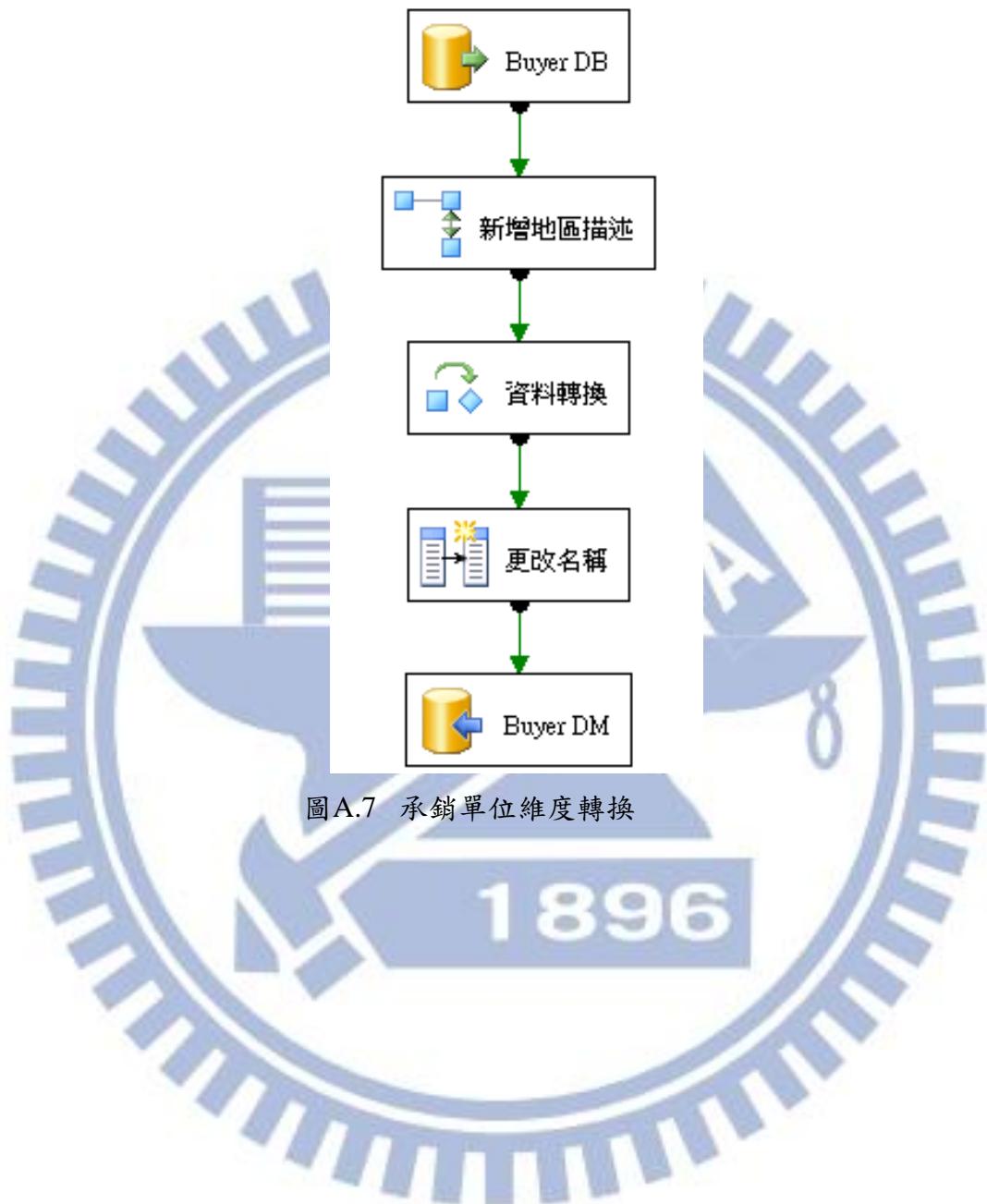


圖 A.6 拍賣事實表轉換

圖 A.7 為承銷單位維度的轉換，首先依照承銷地區代碼新增地區名稱，並修正承銷名稱表達方式，即可匯入資料超市。



圖A.7 承銷單位維度轉換