

第二章 文獻回顧

本章的主要目的為回顧資料倉儲系統實作以及花卉拍賣所需的理論基礎。本章共分為六節，第 2.1 節敘述台灣花卉供應鏈運銷體系與拍賣流程，第 2.2 節為台北花卉拍賣市場的切花品名編碼說明，第 2.3 節說明拍賣制度，第 2.4 節說明資料倉儲技術的理論，第 2.5 節說明利用線上分析處理系統的架構，第 2.6 節則說明分析所用到的統計方法。

2.1 花卉供應鏈與拍賣流程

本節主要為描述花卉供應鏈的運銷體系以及花卉批發市場的拍賣作業流程，在 2.1.1 小節以台灣花卉市場運銷體系為例，介紹花卉供應鏈的關係，在 2.1.2 小節而是以台北花卉批發市場為例，介紹花卉拍賣之作業流程。

2.1.1 花卉供應鏈

花卉供應鏈是指花卉從上游生產者(產地花農、農會、產銷班、合作社、場等)，再經由物流運送以及交易動作來到中游的配銷中心(花卉拍賣市場、零批市場)與下游的消費者(市場承銷商、零批商、零售商等)再販賣到消費者的手中。以台灣花卉市場的運銷體系來說，從生產者到消費者之間所包含的階層大致分類如下：1.生產者 2.農會產銷班 3.合作社、場 4.產地販運商 5.拍賣、批發、零批市場 6.市場承銷商 7.零售商(花店、小販) 8.消費者(飯店、餐廳、一般消費者)[13]。其花卉運銷體系關係圖如圖 2.1 所示如下。

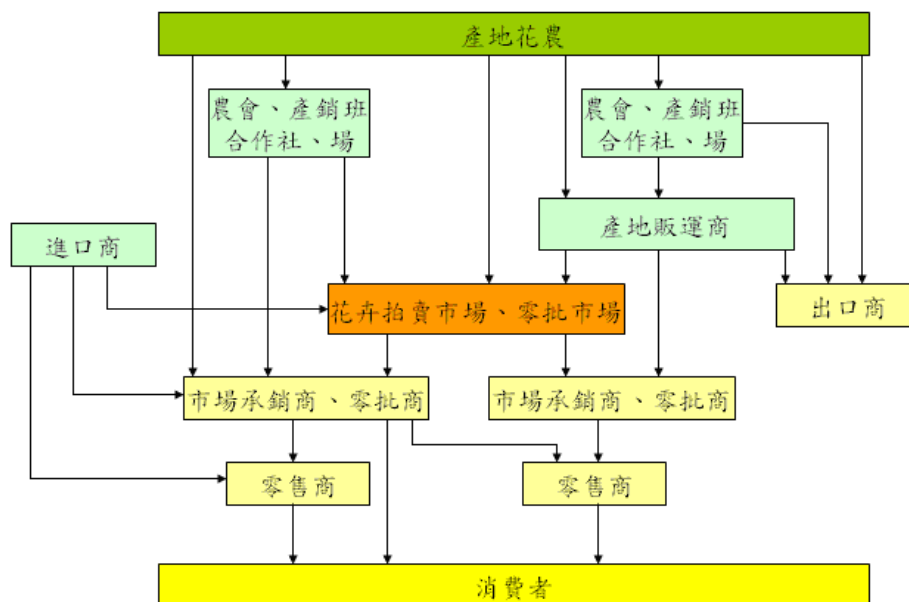


圖 2.1 花卉運銷體系的關係圖[13]

花卉的運送過程必須特別注重產品的保鮮與碰撞，在市場的交易制度也必須採取快速與公平的拍賣方式，以合理的價格銷售給承銷人，再由承銷人販賣給消費者。花卉與其他的農產品並不相同，因為花卉在運送過程的高受損率以及下游的需求常常因為外在的環

境、節日所影響，造成了花卉產品的單價波動較一般的農產品高。所以在花卉供應鏈中不但要注重生產產品的品質、運送過程的妥善、合理的拍賣制度外，更要熟悉利用行情報導與花卉批發資訊分享熱線等資訊系統，使生產更加貼近消費者的需求。

2.1.2 拍賣流程

花卉拍賣市場在花卉供應鏈體系中是佔有著相當重要的一席之地。而目前國內共有台北、台中、田尾、台南、高雄五大花市。其中台北花市更是佔了全國拍賣總量的40%以上，因此本章節特以台北花市來介紹花卉拍賣的作業流程[7]。

花卉的運送過程必須注意花卉的保鮮、保存問題，以及保護花農的權益，所以在拍賣制度上必須慎重的考慮。由於荷蘭式拍賣具有物品交易時間較短，以及連續拍賣時可使得價格核理化而提昇銷售價格的兩項特性，所以目前國內花卉批發市場正是使用荷蘭式拍賣，並搭配有線的拍賣系統進行拍賣。

圖 2.2 中，使用 IDEF0 技術來描述花卉拍賣市場的作業流程。拍賣的物流作業是由集貨(A1)、理貨(A2)、拍賣(A3)、分貨(A4)、領貨(A5)等五個作業方塊所完成。圖中每個方塊左方箭頭代表輸入，右方箭頭代表輸出，下方箭頭代表工作人員或設備，如拍賣員或是台車等；粗線代表物流如花卉，細線則為資訊流如花卉代號。各方塊中所包含之詳細作業可進一步在各方塊的展開圖中取得說明。

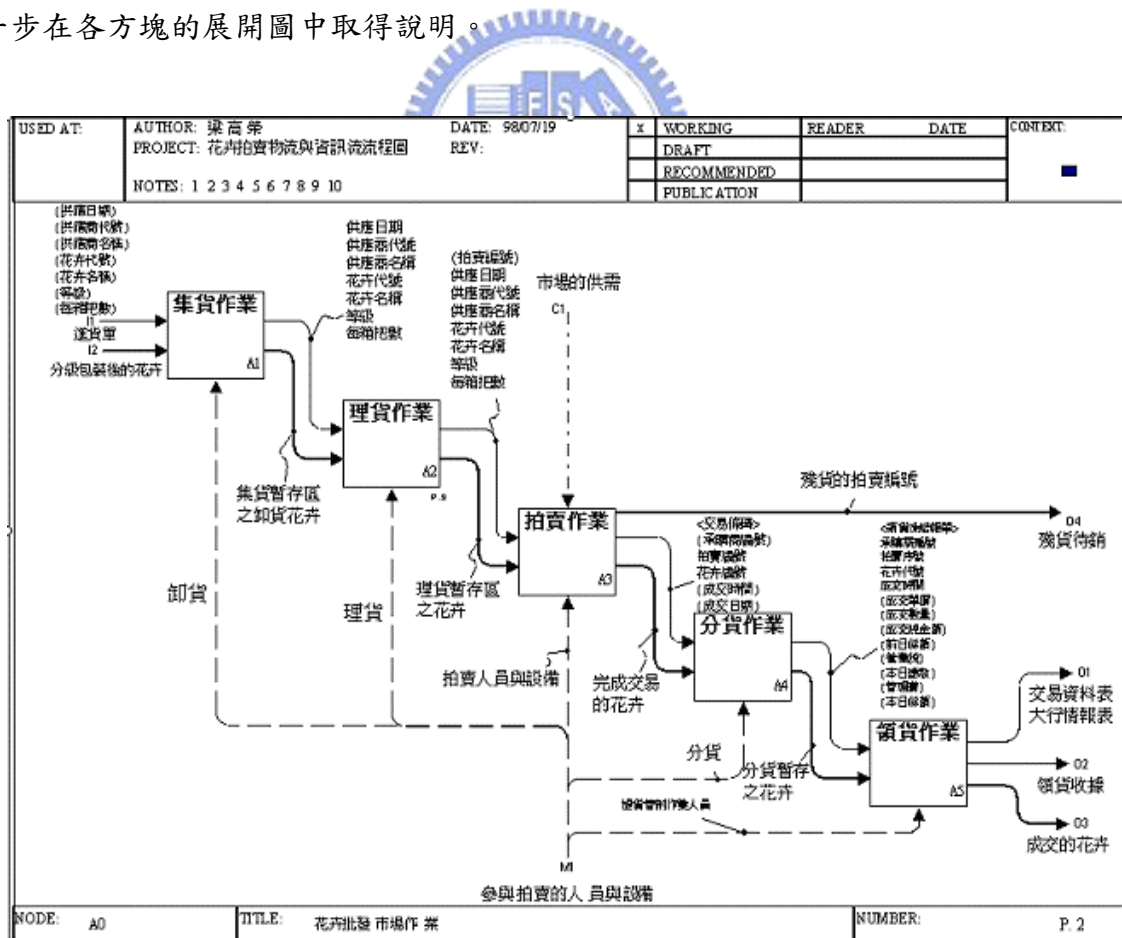


圖 2.2 花卉批發市場拍賣作業流程[7]

集貨作業(A1)為貨車將花卉由運銷合作社運至批發市場後進行卸貨作業，同時將進貨

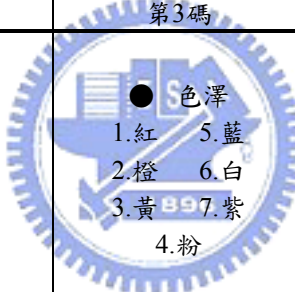
單送至下一個作業點作為進貨清點、登記的依據。理貨作業(A2)為當花卉卸貨搬入理貨區後，理貨員根據收到的進貨單內容將花卉依其類別、等級整理成不同的批賣批次，同時賦予每箱花卉一組拍賣序號。在拍賣作業(A3)裡，當理貨作業完成後，所有須拍賣的花卉資料皆送至資訊管理室。等待拍賣作業正式開始時，花卉便會被工作人員從台車或拍賣暫存區依進貨、理貨順序搬運到拍賣輸送帶上進行拍賣。當主拍賣員唱名到某拍賣序號後，資訊管理部門便會將該組序號的花卉進貨資料顯示在主拍賣員的終端機以及拍賣現場的拍賣鐘上。當花卉透過輸送帶來到主拍賣員面前時，主拍賣員會取出箱內最底層的一把花卉，加以描述給承銷人參考。此時主拍賣員會依據當日拍賣狀況及個人經驗對於花卉定價供承銷人競價。由於花卉拍賣是採取荷蘭式拍賣，在此拍賣法中商品的價格是隨拍賣時間進行而由高至低遞減，所以主拍賣員通常會以高於平均價格 50%左右開價。對於成交的花卉，條碼機會同步列印出出貨條碼，出貨條碼上紀錄拍賣序號、承銷人代碼、成交順序、成交時間等，再由工作人員將其黏貼於貨品包裝箱上。隨後拍賣流程進入分貨作業(A4)，分貨人員以出貨條碼上所登錄的承銷人代碼為分貨依據，將成交的花卉先堆置於承銷人個別的貨品暫存區，待承銷人完成所有的拍賣作業後領取。在領貨作業(A5)裡，領貨管制人員依據承銷人代號，核對領貨內容及與資訊管理部門所提供的領貨單資料是否無誤，若核對無誤，則完成出貨動作。



2.2 台北花卉拍賣市場切花品名

本節為對於系統中切花品名作編碼方式的介紹，以台北拍賣市場的切花品名為例，花卉種類共分為 48 個花卉品類，而花卉代碼共有五碼。花卉代碼的第一碼代表花卉的產地，F 表國產花、I 表進口花卉。第二至第五碼則依據花卉品類的不同，其編碼方式皆有些許之差異，但仍可歸納出部分規則。以康乃馨、火鶴、大菊等花卉為例，代碼之第二碼 A、B、C 等表其花卉品項，第三碼表其色澤，第四、五碼則表其品種。而葵百合、姬百合、香水百合等花卉代碼之第二碼 K、L、S 等同樣表示其花卉品項，但第三碼代表其品種，第四碼代表色澤，第五碼代表花型。以 FS444 香水百合馬可波羅粉四朵、FH053 水仙百合黃色、FR102 玫瑰成功紅三者為例，FS444 香水百合馬可波羅粉四朵第一碼為 F 表國產花，第二碼 S 表香水百合，第三碼 4 表品種(馬可波羅)，第四碼 4 表色澤(粉)，第五碼 4 表花型(4 朵)，而 FH053 水仙百合黃色第一碼 F 表國產花，第二三四碼 H05 表水仙百合，第五碼 3 表色澤(黃色)，而 FR102 玫瑰成功紅第一碼 F 表國產花，第二碼 R 表玫瑰，第三碼 1 表色澤(紅)，第四五碼 02 表品種(成功)。其他花卉品類的編碼方式，整理如下以表 2.1 說明其花卉品名編碼原則，所有花卉品名於附錄一。

表 2.1 切花品名代號編碼原則

第1碼	第2碼	第3碼	第4碼	第5碼
F 國產花 I 進口花	A 康 乃 馨 B 火 鶴 C 大 菊 D 小 菊 E 非 洲 菊 G 劍 蘭 R 玫 瑰		◎ 品種	◎ 品種
	A8康乃馨中輪 A9小 可 愛 W1星 辰 花 W2孔 雀 花		● 色澤 1.紅 5.藍 2.橙 6.白 3.黃 7.紫 4.粉	◎ 品種
	H05水仙百合 H31大 理 花 H44小 蒼 蘭 O56石 斛 蘭	H07石 蒜 花 H36金 魚 草 H52迴 香 O60蝴 蝶 蘭 W56滿 天 星	H11天 堂 鳥 H41愛 麗 絲 H61鬱 金 香 W47卡 斯 比 亞	● 色澤 1.紅 5.藍 2.橙 6.白 3.黃 7.紫 4.粉
		H23夜 來 香 H45紫 羅 蘭 O33文 心 蘭	H29向 日 葵 O20虎 頭 籃	○花型 1.大輪 2.中小輪 1.單瓣 2.雙瓣 1.大型 2.小型
		H131鶴 蕉 H251秀 線	H135垂 蕉 H330雞 冠 花	H156野 薑 花 H375飛 燕 草 H201貓 柳 Y991青 竹

	K 葵百合 L 姬百合 S 香水百合 T 鐵砲百合 U 洋桔梗	◎ 品種	● 色澤	○ 花型 1.單朵 2.雙朵 3.三朵 4.四朵 5.五朵 1.單瓣 2.雙瓣
	O 蘭花類 W 重量計花卉	◎ 品種	● 色澤	◎ 品種
F 國產花	H 其他花卉	◎ 品種	◎ 品種	● 色澤
I 進口花	Y 切葉類	◎ 品種	◎ 品種	◎ 品種

由於花卉品類眾多，在高達48種花卉品類中找出具有代表性意義的花卉品類進行分析成為本論文的前置作業之一。利用現有的台北花卉批發資訊分享熱線系統，根據近三年的資料找出具有高總成交量的十種國產花花卉品類，作為代表性的花卉，代表性花卉為FB火鶴花、FS香水百合、FR玫瑰、FD小菊、FE非洲菊、FC大菊、FG劍蘭、FK葵百合、FA8康乃馨中輪、FU洋桔梗。以下表2.2顯示其總成交量數據。

表 2.2 總成交量數據

總成交量(把)	2002 年	2003 年	2004 年	總和
FB 火鶴花	4430059	4137805	4371939	12939803
FS 香水百合	3368768	3216848	3355294	9940910
FR 玫瑰	2539691	3015866	2774314	8329871
FD 小菊	1705983	1987137	1912841	5605961
FE 非洲菊	1756676	1822237	1821858	5400771
FC 大菊	894034	1147856	1125271	3167161
FG 劍蘭	906147	1047549	933546	2887242
FK 葵百合	957032	884844	751863	2593739
FA8 康乃馨中輪	707994	769805	728140	2205939
FU 洋桔梗	733717	742914	680828	2157459

2.3 拍賣制度

根據已知的文獻記載，目前最早的拍賣行為出現在西元前500年的古巴比倫時代。當時的婚姻市場(Babylonian Marriage Market)制度[33]，將每年的適婚少女聚集於廣場上拍賣，許配給出價最高者，如果拍賣的底價(Reserve Price)為負的則視為少女的嫁妝。到了古羅馬時代拍賣的流程已呈現嚴謹的組織結構，並有專門提供給拍賣使用的廣場，而拍賣物品更逐漸的多元化，食物、工藝品、古物，甚至官位都可以拍賣。在描述拍賣由來之後，接下來的兩小節將介紹拍賣的定義以及拍賣形式。[7]

2.3.1 拍賣定義

拍賣(Auction)來自於拉丁語根auctus，在拉丁文裡，語根auctus有著上昇的意思，而英文的涵義則含有「拍賣」及「拍買」兩個意思[7]。所以所謂的拍賣(Auction)是一種在市場上用來決定產品價格的經濟機制，不論是賣或是買。清楚的來說，就是市場上有一獨賣者(獨買者)以及一群買者(賣者)在進行交易以決定物品的價格。由於獨賣者(獨買者)不清楚買者(賣者)對產品的需求情形，因此將買者(賣者)集合起來彼此競買或競賣，以決定議價之物品價格及資源配置的一種交易方式。

而過去的學者也為拍賣下了一些定義，如下表2.3所示：

表 2.3 拍賣定義

拍賣的定義	出處
拍賣是在明確的規則下，基於市場參與者的競標以決定資源分配與價格的一種市場制度。[32]	McAfee and McMillan (1987)
拍賣基本上是在不確定狀態下的一種對稀有資源進行分配與定價的機制。[28]	Cramton (1998)
拍賣是一個買賣雙方針對特定銷售產品進行價格競爭的資源分配制度。[25]	Bierman and Fernandez (1998)
拍賣是一種交易過程的形式化，即指交易雙方透過特別機制的管轄下進行出價及實現交易。[29]	Kumar & Feldman(1999)

2.3.2 拍賣形式

所謂拍賣的形式指的就是拍賣活動進行的規則制度，這些規則制度包含出價的方式、資源的配置、以及最終價格的決定等。由於拍賣物品種類以及拍賣目的的不同，所以也隨著出現不同的拍賣形式。以下介紹幾種較為常見的拍賣形式[7]。

在市場需求的機制下，拍賣形式可以分為單邊拍賣(One-sided Auction)以及雙邊拍賣(Two-sided Auction)。單邊拍賣指的是在單一賣方多位買方以及單一買方多位賣方的拍賣情形下，雙邊拍賣則是在買賣雙方皆有多人的情況下的拍賣方式。在單邊拍賣中又可以因為價格公開與否、價格的決定方式不同分為荷蘭式拍賣(Dutch Auction)、英國式拍賣(English Auction)、首價閉式拍賣(First-price Sealed-bids Auction)、次價閉式拍賣(Second-price

seal-bids Auction)四種。而多邊拍賣由於執行的方式可分為標單導向(Order driven)、報價導向(Quote driven)、混合制(Hybrid Approach)三種。其中標單導向的雙邊拍賣還可分為集合拍賣(Call Auction)、複式拍賣(Double Auction)、混合式三種。以下將清楚說明各種拍賣。

- 荷蘭式拍賣(Dutch Auction)：

起源於荷蘭的農業批發市場。拍賣的起始價格較高，通常是大於所有人的購買意願。當隨著拍賣的進行價格逐漸降低，直到某位出價者接受價格，交易才算完成。
- 英國式拍賣(English Auction)：

為最常見的一種拍賣方式。拍賣物品由一個相當低的價格或是底價(reserve price)開始競價，其後的出價皆要大於前一個價格，直到只剩一位最高的出價者，即為得標者。
- 首價閉式拍賣(First-price sealed-bids Auction)：

又稱為日本式拍賣(Japanese Auction)。拍賣方式為確定一個開標時間，在開標之前接受大家的秘密標價。在活動期間所有的出價者的出價皆為保密，到了開標時間同時公開所有出價，出價最高者即為得標者。
- 次價閉式拍賣(Second-price seal-bids Auction)：

由William Vickrey所提出，所以又稱為維式拍賣(Vickrey Auction)。拍賣過程與首價閉式拍賣大致相同，得標者為出價最高者，但是得標金額為出價次高者的出價。
- 集合拍賣(Call Auction)：

集合拍賣是把所有供應人以及承銷人的標單收齊後，再依據總需求量以及總供給量來決定其成交價格。
- 複式拍賣(Double Auction)：

所謂的複式拍賣是指把拍賣時間分成數個拍賣時段，在每個時段內，買賣雙方都可以自由地重複出價，直到成交為止。是種在多人競價時，效率很高的拍賣方式。

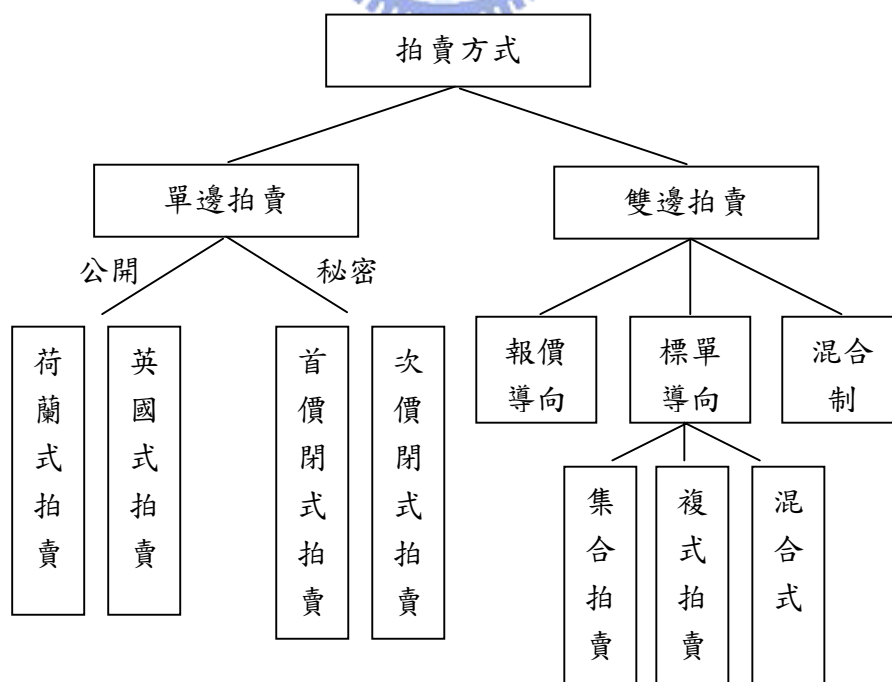


圖 2.3 拍賣形式[7]

2.4 資料倉儲系統

資料倉儲技術(Data Warehouse)是一種能夠整合不同來源資料庫資料，並將其資料進行適當的整合或轉換以方便管理的儲存技術。而倉儲內的資料再搭配線上分析處理工具後便可提供使用者作為決策輔助及統計分析的依據。相對於傳統系統的被動式查詢—需要查詢時才作資料的運算；資料倉儲技術的主動式查詢—當來源資料變動，及作出相對應的資料運算，實為一種能提供即時且彈性資訊的輔助工具。本節共分為三小節，第 2.4.1 節介紹資料倉儲的定義與特性；第 2.4.2 節說明資料倉儲的架構；第 2.4.3 節則討論資料倉儲與傳統關聯式資料庫的比較。

2.4.1 資料倉儲的定義與特性

有資料倉儲之父稱號的殷默(B. Inmon)在 1992 年首次對資料倉儲作了以下的定義 [30]：是個以主題為導向(Subject-Oriented)、整合的(Integration)、隨著時間變化的(Time Variation)及非揮發性 (Non Volatilization) 的資料蒐集系統，用來支援管理決策的制定。對資料倉儲的主題導向性、整合性、時變性、非揮發性四大特性，分述如下：

1. 主題導向性(Subject-Oriented)：

有別於一般作業系統採用的功能或處理導向(Function or Process Oriented)；資料倉儲是採用主題導向，著重於快速、彈性的滿足使用者對於資訊的需求。因此資料倉儲內的資料是依某種主題而整合在一起，與主題無相關的資料便可從資料倉儲中去除。以台北花卉資料倉儲系統為例，是以花卉銷售數量、價格為主題，因此供應人與承銷人的電話地址以及其他不相關之資料便不加於儲存於倉儲中。

2. 整合性(Integration)：

資料倉儲是一個將不同資料庫中資料整合後集中儲存的資料庫。也就是將企業中不同的部門所各自不同格式的資料，經過萃取、轉換、淨化、載入等動作整合儲存在同一資料庫內，來自於異質資料來源的資料在此資料庫內將擁有一致性的欄位名稱、資料屬性與編碼方式。

3. 時變性(Time Variation)：

資料倉儲內儲存的資料除了目前以及以往的歷史資料外，對於未來作業系統產生的新資料也必須加以儲存。因此資料倉儲需要定時的擷取更新資料，以確保資料的一致性，因此資料倉儲內的資料是會隨著時間而變化的。

4. 非揮發性 (Non Volatilization)：

在資料倉儲內資料只能隨著時間定期的被加入，不因新資料而有任何異動的可能，以及資料倉儲所儲存的是長期的歷史資料，因此也不允許有刪除的動作，即表示加入的資料擁有保留與唯讀的特性。

2.4.2 資料倉儲的架構

資料倉儲的建構方式可分為兩種模式[30,31]：分別為殷默模型(Inmon's model)[30]及競篋模型(Kimball's model)[31]。以下將各別敘述說明：

般默模型：

是一種由上而下(Top-Down)的資料倉儲架構，主要是由企業的資料系統以及資料庫所構成，稱之為”企業資訊工廠”(Corporate Information Factory, CIF)。般默模型將企業環境分為四個層級：操作性系統(Operational)、原子資料倉儲(Atomic Data Warehouse)、部門(Departmental)、個人使用者(Individual)。

操作性系統層級主要是由多個操作性的資料庫所組成，儲存企業的歷史資料以及交易資料，用以支援企業每日的操作性活動。在第二層級的原子資料倉儲則是利用操作性系統中所儲存的資料加以萃取彙總所得。在部門層級為更進一步的根據所需萃取彙總原子資料倉儲中的資料。而在個人使用者層級，使用者則是由部門層級資料庫中取得所需的資料暫存於個人電腦中。在這種架構下，一個原子資料倉儲可用來建置多個部門層級資料庫，可有效避免不一致資料的產生。

在般默模型建置的設計上，般默教授提出螺旋式發展方式(Spiral Development Approach)來建置資料倉儲，如圖 2.4 所示。

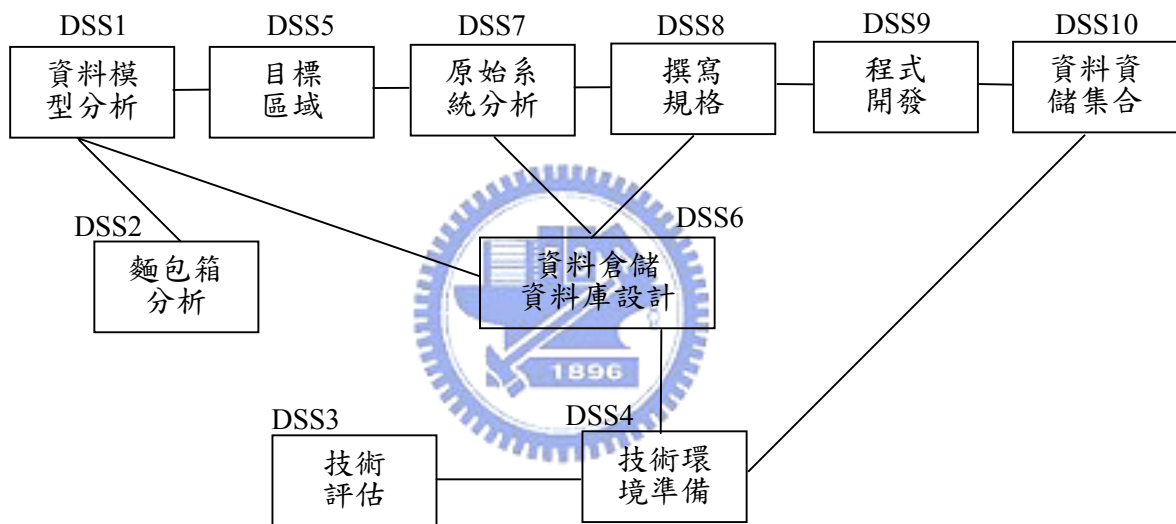


圖 2.4 螺旋式發展方式

螺旋式發展方式建置資料倉儲的步驟，須先由資料模型分析(DSS1)步驟進行現有的資料系統模型。接著進行麵包箱分析(DSS2)衡量資料倉儲所需的資料詳細程度。並展開技術評估(DSS3)的工作，評估相關技術的可行性。根據評估的結果著手於技術環境準備(DSS4)確認相關軟硬體之完善。便可進行目標區域(DSS5)的選擇，選定目標區域的建置處後分析目標區域原始系統(DSS7)並撰寫系統規格(DSS8)。在資料倉儲資料庫設計(DSS6)的步驟之前，需確認完成資料模型分析、技術環境準備、原始系統分析、撰寫規格四項前置作業，以避免資料倉儲與部門資料不相容的問題產生。在資料倉儲設計完成後，便可進行相關程式開發(DSS9)，利用開發的程式完成最後的資料倉儲集合(DSS10)。

對於般默模型而言，由於建置複雜度相當高，系統開發者須對於組織架構與作業流程相當的了解，否則在系統建置與設計的過程中易遭受挫折。

競篋模型：

是種採用由下而上(Bottom-Up)的建構方式，整合系統各部門的資料超市(Data Mart)彙

總成資料倉儲。建構的方法採用維度資料模式(Dimensional data model)，此模式與傳統的實體關聯資料模式不同，是使用資料表作為建構的基礎。在維度資料模式中資料表分為兩大類：事物表(Fact table)與維度資料表(Dimensional table)。事物表存放實際資料中需要分析的數值型態資料，稱之為衡量值(Measure)以及可與其他維度作關聯設定的鍵值；維度資料表則用來儲存與事物表中數值型數據相關的資訊。事物表與維度資料表的相關說明如下表 2.4 所示：

表 2.4 事物表與維度資料表說明

事物表 (Fact table)	<ol style="list-style-type: none"> 1. 大量且不重覆的資料列，可能上千萬或億筆歷史資料。 2. 包含為數值型態的衡量值，以及和維度資料表做關聯的外鍵(Foreign Key)。 3. 靜態資料，一但載入便不再做任何修改異動。
維度資料表 (Dimensional table)	<ol style="list-style-type: none"> 1. 資料量遠小於事物表，可能只有上百筆或十來筆資料。 2. 大部份為文字型態的資料。 3. 欄位大多用來管理維度裡的階層關係以及作為分析事物表衡量值的窗口。 4. 動態資料，可以進行資料變更。

而維度資料模式又可分為三種不同的綱要架構來呈現：星狀綱要(Star schema)、雪花綱要(Snowflake schema)、星座綱要(Fact constellation schema)。

1. 星狀綱要(Star schema)：

為最常見的一種模型，架構如圖 2.5 所示。以事物表為中心，四周環繞維度資料表且每張維度資料表無法下挖。此種綱要以狀似星星故稱之為星狀綱要，也是最簡單的一種模型。

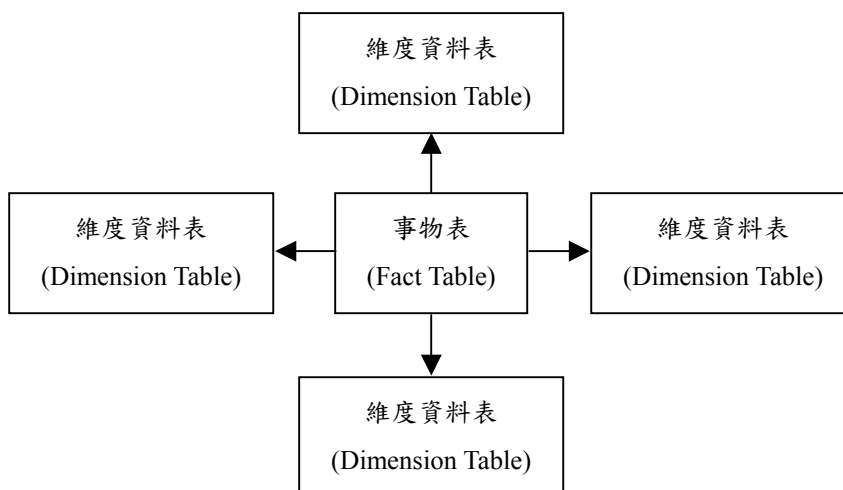


圖 2.5 星狀綱要

2. 雪花綱要(Snowflake schema)：

雪花綱要的架構觀念與星狀綱要近似，可說是星狀綱要的延伸。主架構與星狀綱要相同，維度資料表環繞在事物表的周圍；差異在於雪花綱要對於維度資料表進行三階正規化

動作，因此在雪花網要中同一維度不一定只存在一個維度架構。架構方式如圖 2.6 所示。

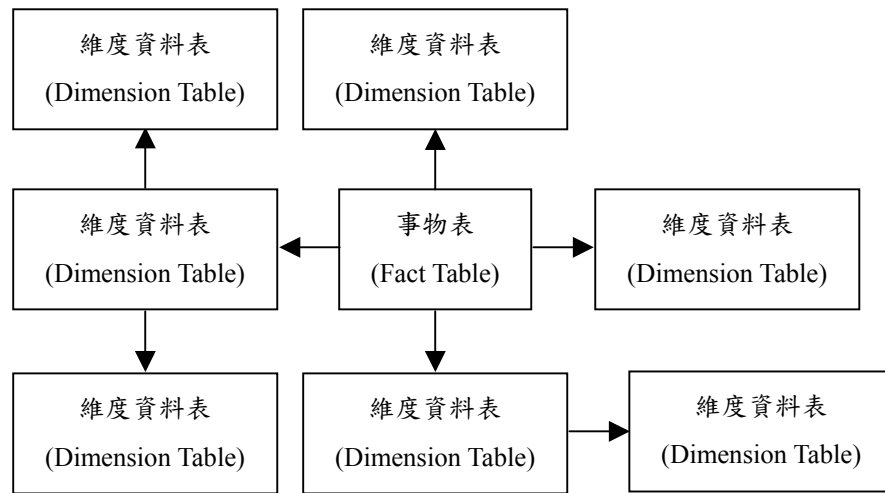


圖 2.6 雪花網要

3. 星座網要(Fact constellation schema)：

又可稱為銀河網要(Galaxy schema)，是一種較為複雜的模型。架構中包含的多個事物表參照同一維度資料表，其架構圖型如圖 2.7 所示。

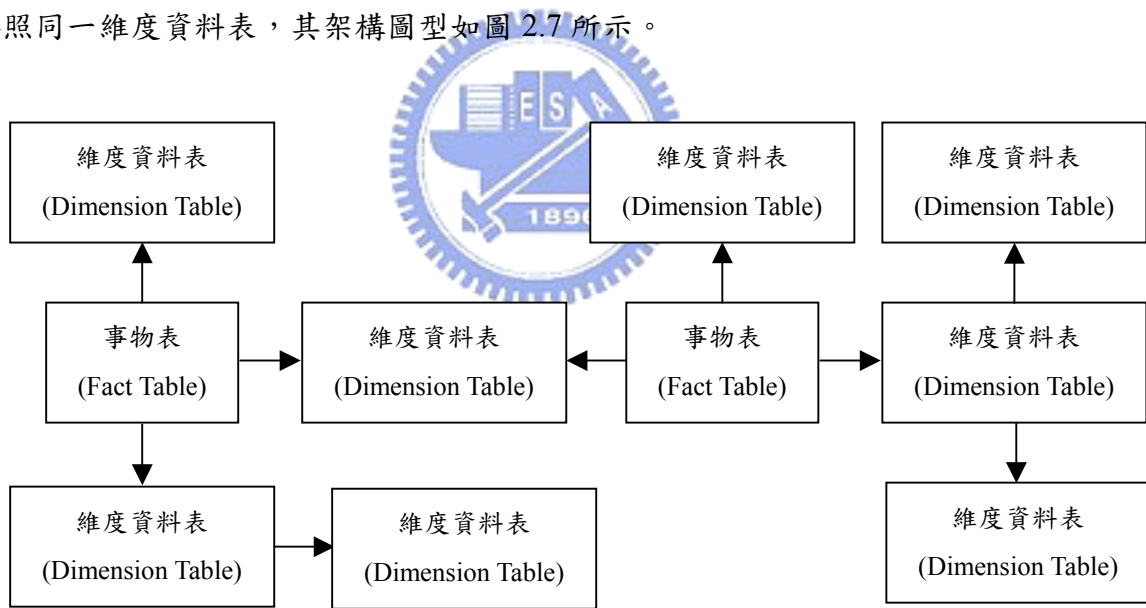


圖 2.7 星座網要

在競筭模型中由於採用事物表存放衡量值，若增加一個衡量值將大量的佔用儲存空間，影響系統的效能；但是衡量值的多寡也會影響查詢結果的豐富與否，因此在系統效能與查詢結果變化性的抉擇上也是系統開發者仔細考慮的地方。

2.4.3 資料倉儲與傳統關聯式資料庫的比較

傳統的關聯式資料庫(Relational Database Management System, RDBMS)通常是採用實體關聯資料模式(Entity-Relational Data Model)建構，資料網要是以實體與實體之間的關聯

組成。關聯式資料庫主要著重於線上即時交易與查詢的過程，是用來支援線上交易處理系統(On-Line Transaction Process, OLTP)而建置的，是以每天的交易作業處理為導向，如訂單處理、採購作業、會計作業等。

資料倉儲採用多維度模式(Multidimensional Model)架構，而資料綱要則有多種形式，如星狀綱要(Star Schema)、雪花綱要(Snowflake Schema)、星座綱要(Fact Constellation Schema)等。資料倉儲主要為提供使用者與管理者資料分析與決策制定的資訊，採主題導向，因此與主題無關的資料便不儲存於系統中。資料倉儲是適用於能夠因應不同使用者的需求來組織與呈現資料的彈性系統，所以用來支援線上分析系統(On-Line Analytical Process, OLAP)。

對於資料部份，關聯式資料庫由於需支援 OLTP，資料部份多為每日的交易資料，因此系統必須及時更新且資料須經過正規化過程減少重覆的資料以提供快速的反應效率。而資料倉儲部分，儲存的資料多半為長期的歷史資訊以及經過彙總後的資料；由於 OLAP 系統強調於可依使用者需求查詢改變資料倉儲維度，在反應速度上就沒有特別強求。

關聯式資料庫系統和資料倉儲系統比較可由表 2.5 清楚說明，但是資料倉儲並非為取代傳統關聯式資料庫而出現，而是在其基礎上進行資料的加值處理及應用。透過資料倉儲與外部的異質資料庫整合後，再利用線上分析處理加以分析查詢或製作報表，或是利用資料挖掘進行分析和決策輔助，因此資料倉儲和關聯式資料庫系統並非對立，而是相輔相成的。

表 2.5 關聯式資料庫系統和資料倉儲系統的比較[15]

	關聯式資料庫系統	資料倉儲系統
組織特性	採交易作業處理為導向。	採主題導向。
資料特性	及時更新、常更改，少歷史性及重複性質的資料。	為歷史性、不常更動且資料重複性高。
功能特性	供每日的交易資料儲存，及日常交易資料查詢。	可提供長期的資料分析，以供決策者的參考。
系統特性	回應速度敏感、但要有不同查詢須另外設計程式。	回應速度較慢、使用者可以依查詢需要改變資料倉儲維度。
資料庫設計模式	實體關係圖(ER model)、應用導向的 OLTP 線上交易處理。	星狀、雪花或星座綱要圖(schema)、主題分析導向的 OLAP 線上分析處理。
應用發展	及時資料的查詢與顯示，線上交易處理。	資料挖掘分析、客戶關係管理、決策支援系統和線上分析處理。

2.5 線上分析

由於資訊時代的高度發展，如何將資料做完好的儲存、管理並從其中找出有意義的資訊逐漸成為研究的重點。透過資料倉儲技術的整合、儲存資料後，再利用工具進行線上分析處理得到更進一步的資料提供給決策者，這些步驟形成了決策支援系統的主要資料來源。之前的文獻介紹過資料倉儲，接著本節繼續說明線上分析處理，在第 2.5.1 小節說明線上分析處理的意義與操作方式，第 2.5.2 小節說明線上分析處理的類別，第 2.5.3 小節說明在本文中線上分析處理所採用的工具。

2.5.1 線上分析處理意義與操作方式

線上分析處理(On-Line Analytical Processing, OLAP)的觀念最早源自於西元 1993 年 E.F. Codd 博士所提出。線上分析處理的是一種可以彙整資料倉儲裡的原始資料，並將其轉換為多維度分析模組，讓使用者可以透過網際網路在線上利用多種的分析處理操作方式去分析資料的應用程式。線上分析處理系統不但能提供快速的資料內容給使用者，更能支援多位使用者同時上線查詢。

以下介紹線上分析處理的操作方式：

1. 上捲(Roll-up)：將維度的單位放大，如圖 2.8 所示如下。

拍賣平均價格(元)	1 季	2 季	3 季	4 季
玫瑰花	17	17.5	16	18
波斯菊	23	22	20	22
滿天星	9	8	10	11

上捲

拍賣平均價格(元)	1 月	2 月	3 月
玫瑰花	18	16	17
波斯菊	22	26	21
滿天星	9	10	8

圖 2.8 上捲圖

2. 下挖(Drill-down)：將維度的單位縮小，如圖 2.9 所示如下。

銷售量(百萬)	1 月份
服飾	3.5
電器	5
食品	1.8

下挖

銷售量(百萬)	1 月份
冰箱	2.8
電視	2
冷氣	0.2

圖 2.9 下挖圖

3. 切片(Slicing)：限制某一維度的上下範圍，如圖 2.10 所示如下。

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
波斯菊	22	26	21	22	23	21	22	20	18	23	24	19
滿天星	9	10	8	8	7	9	11	10	9	13	11	9

↓對月份維度作限制，只要後半年。

	7月	8月	9月	10月	11月	12月
波斯菊	22	20	18	23	24	19
滿天星	11	10	9	13	11	9

圖 2.10 切片圖

4. 切丁(Dicing)：限制多個維度的上下範圍，如圖 2.11 所示如下。

地區電器銷售量(百萬)	1季	2季	3季	4季
台中縣	1.8	2.0	2.2	2.0
桃園縣	1.8	2.1	1.7	1.8
嘉義縣	1.1	0.9	1.0	1.2
彰化縣	1.2	1.0	1.3	1.1
台北市	3.1	2.9	3.8	3.2
台北縣	2.8	2.6	2.8	3.0

限制時間維度為後半年。
區域為台中縣、台北市。

地區電器銷售量(百萬)	3季	4季
台中縣	2.2	2.0
台北市	3.8	3.2
台北縣	2.8	3.0

圖 2.11 切丁圖

5. 轉軸(Pivoting)：將水平維度與垂直維度作交換，如圖 2.12 所示如下。

銷售量(百萬)	1月	2月	3月
服飾	3.5	3	4
電器	5	4.5	6
食品	1.8	3	2.5

轉軸

銷售量(百萬)	服飾	電器	食品
1月	3.5	5	1.8
2月	3	4.5	3
3月	4	6	2.5

圖 2.12 轉軸圖

6. 視覺化(Visualize)：為將數值化的衡量值以圖形方式表達，如圖 2.13 所示如下。

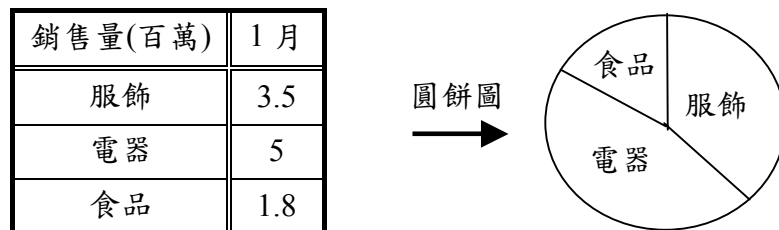


圖 2.13 視覺化圖

7. 篩選(Selecting)：限制衡量值的範圍，如圖 2.14 所示如下。

單日拍賣量(把)	3 點	4 點	5 點	6 點
玫瑰花	0	41	38	0
波斯菊	12	15	11	0
滿天星	8	11	9	0

去零值 →

單日拍賣量(把)	3 點	4 點	5 點
玫瑰花	0	41	38
波斯菊	12	15	11
滿天星	8	11	9

圖 2.14 篩選圖

8. 輸出(Output)：將衡量值另外儲存以供利用，如圖 2.15 所示如下。

銷售量(百萬)	1 月	2 月	3 月
服飾	3.5	3	4
電器	5	4.5	6
食品	1.8	3	2.5

輸出 →

銷售量.xls

圖 2.15 輸出圖

9. 排序(Sorting)：將衡量值做大小的排序，如圖 2.16 所示如下。

拍賣均價(元)	1 季	2 季	3 季	4 季
滿天星	9	8	10	11

↓ 排序

拍賣均價(元)	4 季	3 季	1 季	2 季
滿天星	11	10	9	8

圖 2.16 排序圖

10. 計算(Computing)：訂定規則來改變衡量值，如圖 2.17 所示如下。

產品所得(百萬)	1 季	2 季	3 季	4 季
產品 A	2	3	2	4
產品 B	1	1	1	2
產品 C	5	4	6	4

建立規則，使上半年等於 1、2 季相加，下半年為 3、4 季相加。 →

產品所得(百萬)	上半年	下半年
產品 A	5	6
產品 B	2	3
產品 C	9	10

圖 2.17 計算圖

2.5.2 線上分析處理類別

線上分析處理在設計上主要有三種不同的模式：多維度型(MOLAP)、關聯型(ROLAP)、混合型(HOLAP)。表 2.6 為三種模式的比較。

1. 多維度型(Multidimensional - OLAP, MOLAP)：

此模式適用於經常使用以及需要快速回應的系統。由於此模式是將事先需要運算的資料進行彙總與原資料加入超方體中，並把超方體放在資料庫伺服器前端。雖然會需要耗費大量儲存空間，卻可提供最快速的查詢。

2. 關聯型(Relation - OLAP, ROLAP)：

此種模式是將資料儲存在一個強大的中介資料(Mata Data)儲藏區並利用 SQL 引擎在關聯式資料庫(RDBMS)中查詢而不另外架構超方體。使用者要看報表或多維度分析時，便會根據中介資料儲藏區的訊息，進行 SQL 語法查詢。

3. 混合型(Hybrid – OLAP, HOLAP)：

此種模式為兩種模式的混合型。原資料依舊儲存於關聯式資料庫中，而彙總後資料則用來建構超方體。此模式的好處是簡化多維度型的大量儲存動作以及改善關聯型的查詢時間。

表 2.6 MOLAP、ROLAP、HOLAP 之比較

	多維度型(MOLAP)	關聯型(ROLAP)	混合型(HOLAP)
特色	1.多維度線上分析處理。 2.資料彙整會加入超方體。 3.建置超方體放置於伺服器前端。	1.關聯式線上分析處理。 2.資料儲存在關聯式資料庫中(RDBMS)。 3.不建超方體。	1.多維度與關聯式混合型分析處理。 2.原資料儲存在關聯式資料庫中(RDBMS)。 3.彙整後資料加入超方體。
優點	1.硬體設備的需求較低。 2.查詢速度最快。 3.簡單易學,使用者不需資訊背景即可輕鬆上手。 4.分析、評比、數學的功能強大。	1.靈活性佳且變更設計容易,可支援中大型資料倉儲需求。 2.節省空間,建檔速度快。 3.開放性技術,開發人員與工具較好找。 4.可大範圍查詢	1.查詢速度介於多維式與關聯式線上兩者之間。 2.建檔速度快,擴展性佳,可支援大型資料庫。 3.資料模組設計彈性佳,適用元件實體模型(ER Model)。
缺點	1.資料建構速度慢。 2.需要大量的儲存空間。 3.架構缺乏彈性、不易更改。	1.查詢速度較慢。 2.硬體設備的需求較高。 3.SQL 查詢對非資訊人員來說,不易上手。	1.微觀查詢速度慢。 2.SQL 有些先天限制,難以執行許多複雜的查詢。

2.5.3 線上分析處理工具

在本文中所使用的線上分析處理工具是使用由加拿大 Cognos 公司所提出的商業情報系統,此套系統的主要架構如圖 2.18 所示。系統中包含了 Impromptu、PowerPlay Web Server 及 Visualizer(VIP)三套軟體。

Impromptu 軟體為資料庫查詢以及報表產生工具。使用者可利用 Impromptu Administrator 建立與資料庫的連結以及其中設定各欄位的關聯性,完成以上動作後可將其設定建立目錄(Catalog)並存成*.cat 檔以作為製作報表的基礎。產生的報表可存成*.imr 檔(Impromptu Report)以供使用者瀏覽。將其報表內容還可另存為*.iqd 檔(Impromptu Query Definition)以供作為製成超方體的主要來源。

PowerPlay Web Server 軟體為主要線上分析處理的工具，可提供決策者多角度的綜觀資料，或是逐層分析各筆交易明細資料。系統管理者使用 PowerPlay Transformer 製作超方體並將其存成*.mdc 檔(Multi-Dimension Cube)，再利用 PowerPlay User 或是 PowerPlay Web Server 來進行多維度分析。若並不為 PowerPlay Transformer 建立的超方體，也可利用 PowerPlay Connect 來作連結轉換，同樣也可為 PowerPlay User 或是 PowerPlay Web Server 所接受。

Visualizer 軟體主要為提供高階決策人員所需的企業績效衡量指標(Key Performance Indicator, KPI)。Visualizer 以視覺化的方式，如燈號、儀表版等圖形界面呈現公司資訊，讓高階決策人員可以一目了然的掌握公司的營運狀況，進而快速的分析與決策。由於在本文中並沒有使用到 Visualizer 軟體部分，所以在實作情形部分，不作額外的介紹。

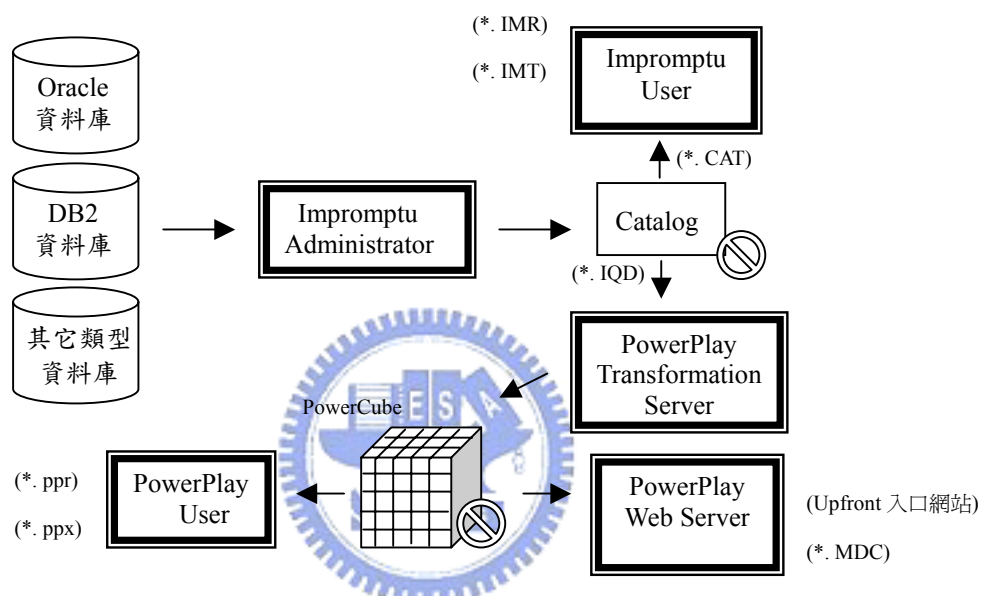


圖 2.18 Cognos 商業情報系統架構[27]

下表 2.7 為說明 Impromptu、PowerPlay Web Server 及 Visualizer(VIP)三套軟體的版本與其用途的比較。

表 2.7 Impromptu、PowerPlay、Visualizer 的版本與功能比較[27]

比較項目	Impromptu	PowerPlay	Visualizer	
產品定位	互動式報表查詢工具	線上分析處理工具	視覺化分析處理工具	
產品主要用途	ADM 版	製作 Catalog 及權限管理	製作超方體及間接權限管理	無
	USER 版	報表的瀏覽及製作	超方體瀏覽(多維度分析與處理)	做視覺化的地圖，可供使用者來瀏覽資訊
	WEB 版	可將報表的瀏覽提升至 Internet 層次	可將超方體的瀏覽提升至 Internet 層次	可將地圖的瀏覽提升至 Internet 層次
使用人員	ADM 版	資訊中心之人員(開發版)	資訊中心之人員(開發版)	無
	USER 版	各階層之人員	各階層之人員	各階層之人員
	WEB 版	各階層之人員	各階層之人員	各階層之人員

2.6 統計方法

對於本論文所進行的統計分析所使用的統計方法，將介紹如下，在 2.6.1 節說明兩組獨立樣本之 t 檢定的意義與檢定方法、使用時機[6]；在 2.6.2 節說明成對變異數分析的意義與檢定方法、使用時機[6]。

2.6.1 兩組獨立樣本之 t 檢定

所謂兩組獨立樣本是指分別從兩獨立常態母體中隨機抽出的兩組樣本，兩組樣本中並無關聯性。而從兩組樣本中檢定兩獨立母體平均數是否有顯著差異性的問題，可使用兩組獨立樣本的 t 檢定來處理。檢定的方法視情況分為以下幾種：

- 兩獨立母體，母體變異數未知但相等(σ_1^2 、 σ_2^2 未知但 $\sigma_1^2 = \sigma_2^2$)：

統計假設	檢定統計量	拒絕域
$H_0 : u_1 - u_2 \leq C$ $H_1 : u_1 - u_2 > C$	$t^* = \frac{(\bar{x}_1 - \bar{x}_2) - C}{\sqrt{S_p^2 \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$	$t^* > t_{\alpha, n_1 + n_2 - 2}$
$H_0 : u_1 - u_2 \geq C$ $H_1 : u_1 - u_2 < C$		$t^* < -t_{\alpha, n_1 + n_2 - 2}$
$H_0 : u_1 - u_2 = C$ $H_1 : u_1 - u_2 \neq C$		$t^* > t_{\alpha/2, n_1 + n_2 - 2}$ $t^* < -t_{\alpha/2, n_1 + n_2 - 2}$

其中 C 為常數； u_1 表第一組樣本母體平均數； u_2 表第二組樣本母體平均數； n_1 表第一組樣本的個數， \bar{x}_1 表第一組樣本的平均數； S_1^2 表第一組樣本變異數； n_2 表第二組樣本的個數， \bar{x}_2 表第二組樣本的平均數； S_2^2 表第二組樣本變異數； S_p^2 表兩組獨立樣本合併後的變異數(Pooled Sample Variance)，

$$S_p^2 = \frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}。$$

- 兩獨立母體，母體變異數未知且不相等(σ_1^2 、 σ_2^2 未知且 $\sigma_1^2 \neq \sigma_2^2$)：

統計假設	檢定統計量	拒絕域
$H_0 : u_1 - u_2 \leq C$ $H_1 : u_1 - u_2 > C$	$t^* = \frac{(\bar{x}_1 - \bar{x}_2) - C}{\sqrt{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right]}}$	$t^* > t_{\alpha, v}$
$H_0 : u_1 - u_2 \geq C$ $H_1 : u_1 - u_2 < C$		$t^* < -t_{\alpha, v}$
$H_0 : u_1 - u_2 = C$ $H_1 : u_1 - u_2 \neq C$		$t^* > t_{\alpha/2, v}$ $t^* < -t_{\alpha/2, v}$

$$\text{其中 } \nu \text{ 表檢定統計量之自由度， } \nu = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right]^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}。$$

2.6.2 變異數分析

變異數分析(Analysis of Variance, ANOVA)是分析實驗數據的一種工具，是兩組獨立樣本 t 檢定(假設兩個母體變異數相等)的延伸。當數據為三組或三組以上獨立樣本時，可用此法來比較分析樣本母體的平均數間是否有顯著的差異；例如：比較三所大學學生每天讀書所花的時間或是四種品牌洗衣機的清潔度是否有顯著差異的問題。

在變異數分析中有其三大基本假設：

- 常態性假設(Normality)：假設依變項 k 各母體分配均為常態分配。可利用常態機率圖(Normal Probability Plot)來檢定殘差是否成常態分配。
- 同質性假設(Homogeneity)：假設 k 個常態母體的變異數均相等。可用 Bartlett's test、Cochran's test 和 Hartley's test 等來檢定母體變異數是否相等。
- 獨立性假設(Independency)：假設 k 個常態母體分配間相互獨立。可透過殘差對 \hat{y}_{ij} 的圖形來檢測抽樣方法是否隨機。

變異數分析是將樣本中各觀測值之離中差的總平方和(總變異)，按照變異發生的原因分解為個別原因所引起的平方和，然後將這些個別原因的平方和除以其自由度，得到變異數之估計值，在取成 F 統計量。統計假設為在 K 個變異數母體相等，以及其虛無假設($H_0: \mu_1 = \mu_2 = \dots = \mu_k$)成立的情況下，F 統計量值會等於 1。當對立假設(H_1 : 母體平均數間有顯著差異)成立時，F 統計量值會顯著地大於 1；因此在變異數分析中，F 統計量值是否顯著的大於 1 可用來判斷虛無假設是否成立的依據。

若影響觀測值發生變異的原因只有一個時，稱之為一因子變異數分析；若影響觀測值發生變異的原因不只一個時，稱之為多因子變異數分析。由於本論文所牽涉到的變異數分析，僅為一因變異數分析，故以下說明僅只介紹一因子變異數分析的統計模型。

因子的不同程度狀況可稱為水準(level)，而不同因子的水準組合則稱為處理(treatment)。若比較因子的 k 種水準或 k 種處理間是否有顯著差異時，可用以下的的線性統計模型(linear statistical model)來描述， $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ ， $\begin{cases} i = 1, 2, \dots, k \\ j = 1, 2, \dots, n \end{cases}$ ，其中 y_{ij} 表示利用在第 i 個水準之處理下第 j 個觀測值， μ 表全體平均數， τ_i 表第 i 個水準之處理效應(treatment effect)， ε_{ij} 為隨機誤差，假設為 *i.i.d.* $N(0, \sigma^2)$ 。

分析的目的為檢驗個母體平均數間是否有顯著差異，其統計假設如下：

$H_0: \mu_1 = \mu_2 = \dots = \mu_a = 0$ (各處理之平均數無差異)

$H_1: \tau_1, \tau_2, \dots, \tau_k$ 不全為 0 (至少有一處理的平均數與其他處理的平均數有明顯差異)