

行政院衛生署疾病管制局

建立符合 HIPAA 標準規範之全國預防接種資料庫

(倉儲)與管理決策輔助先導性系統之研究

成 果 報 告 書

計畫執行單位：輔仁大學、元智大學

計畫主持人：邱瑞科 博士

共同主持人：詹前隆 博士

共同主持人：翁頌舜 博士

協同主持人：張啟明 主任

協同主持人：顏哲傑 組長

研究助理：陳盛儀、陳聖棋、簡榮廷
郭建吾、江鴻屏、許默炫

中華民國九十一年十二月三十日

目錄

目錄.....	1
壹、前言	3
貳、相關研究探討	7
一、資料倉儲系統.....	7
二、資料倉儲建置策略.....	13
三、線上分析處理技術.....	18
四、知識探索技術.....	20
五、決策支援系統與統計預測模式.....	28
六、類神經網路.....	30
七、資料庫效能測試與效益分析.....	33
參、醫療決策支援系統	46
一、醫療決策支援系統基礎架構.....	46
二、資訊科技在醫學上應用之困難.....	49
三、醫療電腦化之趨勢.....	50
肆、整合型中央 NIIS 中央資料庫之建置	52
一、資料庫建置相關研究.....	52
二、分散式資料庫相關研究.....	53
三、資料庫設計方法.....	57
伍、整合型中央資料倉儲之建置及應用分析	71
一、NIIS 中央資料倉儲建立的目標.....	71
二、整合型中央版預防接種資料庫與資料倉儲之架構.....	73
三、整合型中央版預防接種資料庫與資料倉儲之差異.....	74
四、中央資料倉儲的建置效益.....	75
五、資料倉儲應用工具之線上分析處理.....	77
六、預防接種資料倉儲之應用.....	78
陸、資料倉儲之決策支援及分析應用	82
一、預防接種決策問題與資料探勘技術之結合.....	82
二、疫苗管理 OLAP 需求分析.....	86
三、運用灰色理論預測疫苗劑量.....	94

四、運用資料探勘分群演算法預測施打疫苗失約特性.....	106
附錄.....	119
附錄一：訪談記錄.....	119
附錄二：欄位轉換.....	121
參考文獻	123

摘要

企業資料倉儲在於整合企業作業與管理資訊，以完整的資訊視野平台提供相關的資訊應用與增值服務；利用線上分析(OLAP, On-Line Analytical Processing)加速決策資訊取得；利用資料挖掘的技術找出疫苗使用特性，提供催種政策的參考。透過疫苗預測分析模式輔助疫苗劑量的採購最佳化，相關的資訊應用將有效提高決策品質並降低決策失誤的風險。

本專題計畫針對資料倉儲及輔助決策支援系統建立之技術與方法深入探討，並以行政院衛生署疾病管制局(Center for Disease Control)之全國性預防接種資訊管理系統(NIIS, National Immunization Information System)之中央資料庫及資料倉儲之建置經驗為例，以線上分析處理機制加強預防接種疫苗管理功能中異常運作之管控模式，以分群理論找出預種特徵；以灰色理論建立疫苗預測模型，三方面的應用作為企業資料倉儲化決策系統實作案例，驗證企業資料倉儲化技術與方法應用在預防接種相關業務決策輔助上的可行性。

線上分析處理可加速疫苗管理資料的分析及運用。透過動態且多維度的分析方式，經由數學邏輯運算及函數等方式進行維度重組，讓使用者更快且更精確的描述出疫苗異動中異常的部份。進一步分析歷史資料，進行深入預測及推估，以推測疫苗的採購處理狀況，加強疫苗採購與使用的效能，進而提高疫苗採購之決策品質。最後則利用資料挖掘的技術，針對四個地區作顧客分群的動作，找出接受疫苗接種的出生嬰兒母親人口統計相關變數的型樣 (pattern)，提供決策單位實施催種作業的參考，以減少預種失約的人數。以上應用透過先導研究以系統建置與實作的經驗，作為疾病管制局運用資料倉儲建立決策系統之參考。

關鍵詞：資料倉儲化、全國性預防接種資訊管理系統、線上分析處理、灰色理論、分群。

Abstract

Enterprise data warehousing integrates enterprise operation and management information to provide a broad-vision platform for decision makers. The speed for accessing critical information for decision making can be accelerated by OLAP (On-Line Analytic Processing). Data mining technique can be applied to NIIS (National Immunization Information System) database to discover hidden knowledge related to policy making for compliance of vaccine injection and optimal vaccine purchasing amount. These information applications can effectively improve decision making quality and reduce the risk for making biased decisions.

In this project, we have conducted an in-depth study of techniques for building enterprise data warehouses and applied these techniques into NIIS (National Immunization Information System) of CDC (Center for Disease Control). By conducting an experimental implementation of CDC's central database and data warehouse for NIIS, we verified the feasibility of the following proposed technologies and methods. OLAP (On-Line Analytical Processing) can be used to accelerate the speed of the deployment of vaccine management. Prediction model adopting Grey System Theory can be used to predict the optimal vaccine purchasing quantity. Clustering algorithm and statistical methods can be used to find inoculation pattern of new-born baby's mothers' miss-appointment for vaccine injection. Therefore, decision makers for vaccine injection policy can easily observe and analyze the current status of national vaccine injection by using the above multi-dimensional and dynamic decision support system. In the end, it will improve the vaccine-injection compliance rate and the decision for vaccine purchasing amount and provide a good support for vaccine-injection policy making. The final report of this research incorporating with a real experience of system implementation can give a good example for enterprises in establishing their own decision support systems with data warehousing.

Keywords: Data Warehousing, National Immunization Information Ssystem, On-Line Analytic Processing, Grey Theory, Clustering.

壹、前言

近年來，資訊系統在醫療上的應用越來越普及，由於資訊系統所能帶來的自動化與便利性，改善了許多人工流程上的效率問題，但由於醫療本身的特殊性，使的醫療資訊系統的發展受到限制，不如一般商業上的使用，但就未來長遠的眼光看來，醫療資訊系統的確是可行的解決之道。

由於醫療資訊系統不再僅是免除人工的自動化的設計而已，對於醫療機構而言，更需要能提供決策輔助的決策支援系統，這樣的一個系統建立，促使國內醫療程序再造工程(medical process reengineering)的風潮，強調決策支援的醫療資訊系統，不但對於醫療機構也所助益，對於廣大的民眾來說，更能提供越趨近以病人為導向(patient-oriented)的服務，提供社會大眾更好的疾病預防及更佳的健康照護及醫療保健服務。

因此，衛生署疾病管制局(Center of Disease Control, CDC)委託輔仁大學及元智大學進行整合型中央版預防接種系統最適模式之規劃，發展中央型資料庫及資料倉儲系統，以與目前正在建置的地方版預防接種系統在未來能夠有效地銜接，並提供全國性預種統計分析作業的資訊化管理功能及國內各相關研究單位所需的預種資料庫。本文擬就下列幾個方向進行規劃與研究，而本篇報告便為規劃與研究後之初步成果：

- 系統資料庫/資料倉儲最適模式
- 預種決策支援功能
- 資料倉儲應用及分析
- 預種資料探勘

有別於傳統一般輔助作業的醫療資訊系統設計，本計畫強調在醫療資訊系統的設計上，導入資料倉儲化(Data Warehousing)，配合線上即時分析處理(OLAP)

的方式，提供即時性的統計分析資料；透過歷史性的醫療資訊查詢，將有助於相關各階層醫事人員進行決策的判斷依據。

內文主要分為三個部分：

1. 第一部份將對於現今常見且實用的決策方法，做簡單的介紹，透過資料倉儲 (Data ware House)、知識探索 (Knowledge Discovery in DataBase) 及一些演算法的介紹，有助於輔助醫療決策的制訂。最後並加上資料庫的效能評估方法，以期中央資料庫的設計能符合效能 (Effective) 與效率 (Efficient)。(詳見第二章)
2. 第二部份主要以建立醫療決策支援系統的角度 (詳見第三章)，分別就 NIIIS 中央資料庫 (詳見第四章) 及中央資料倉儲 (詳見第五章) 建立的目標及建立的模式提出建置重點，並對於中央資料庫與中央資料倉儲的發展所會面臨的困難與問題提出我們相關的建議，最後提出我們所建議的預種醫療決策系統做整體的效益評估。
3. 第三部分納入資料探勘技術之探討，內容包含資料探勘方法論、人工類神經網路，與基因演算法，並以決策支援的觀點將這些方法應用於預種資料的分析上。(詳見第六章)

本文的實質意義在於欲利用多樣化的探勘技術，挖掘出預種資料庫中所隱藏之知識及決策訊息，以利日後輔助主管機關提供新的醫療服務並加強服務的品質。因此，我們根據實際系統 (整合型中央版預防接種系統) 規格書配合訪談決策者，我們根據實際訪問計畫所規劃的系統 – 整合型中央版預防接種系統之可能使用者 (CDC 預種組、衛生局、衛生所)，提出系統功能之初步設計 (詳見附錄一)。本文的成果不僅可提供作為行政院衛生署疾病管制局在政策方針上的界定，亦可作為業界在進行醫療資訊整合過程中的一項參考，還望先進能針對不足之處予以指教。

貳、相關研究探討

為了能夠順利的進行這項計劃，本文搜集了相當多的相關文獻及資料，輔助計劃之進行，並試圖採用這些先進的資訊科技做為未來研究以及實行的技術，希望能夠有效提高醫療的品質及支援預種的多樣化需求及目標。討論的內容包括資料倉儲概念及架構、資料倉儲建置策略、即時線上分析處理、知識探索技術、決策支援系統及統計預測模式、類神經網路以及資料庫效能分析等主題。

一、資料倉儲系統

資料倉儲系統可視為一切資料之匯總，但是它又比資料庫具有更多更詳盡以及更強大的功能，所以在這裡根據它的基本觀念、架構以及建置策略做一說明。

(一) 資料倉儲之介紹

就概念上而言，一個資料倉儲包含了一系列的關鍵性資訊，它可以用來管理並導引企業走向最可能的獲利之路[14]。對於資料倉儲的定義，各家學說都大同小異，分述如下：

- 1) 資料倉儲不僅包含了分析所需的資料，而且包含了處理資訊所需的應用程式，這些程式包括了將資料由外部媒體轉入資料倉儲的應用程式，也包含了將資料加以分析並呈現給使用者的應用程式[1]。
- 2) 資料倉儲可以說是專為查詢和分析設計，用來複製交易資料的一種架構。資料倉儲是將企業內異質性的資料加以合併，將歷史資料取出，來輔助決策分析[13]。
- 3) 資料倉儲化是一種動作及程序，而不是一種產品，其以企業的角度，將不同資料來源的資料整合於一個單一旦詳細的視界中[62]。

但對於各家所作的定義，仍有需要補充的地方，即是「資料倉儲中大部分(95%~98%)是交易資料，但是仍有少數不是交易資料」。如此，對資料倉儲的定義

能更趨於完整。

資料倉儲是一個整合性的、主題導向的、隨時間變動以及非揮發性的資料集合，其主要目的是在提供管理者作決策的參考依據[1, 42, 49, 77]：

- 1) 整合性(Integrated)：將不同來源的資料以統一的規範，經過整理後儲存在一起，所謂統一的規範指的是相同的資料型態、格式或度量等。
- 2) 主題導向(Subject-oriented)：資料倉儲儲存的是與某一主題相關的資料，主題可以不只一個，但是不需要儲存與主題無關的資料。
- 3) 隨時間變動(Time-variant)：資料倉儲的資料都是過去所發生的事實資料，但是並非儲存過去的所有資料，而是在某一時間範圍內的資料。
- 4) 非揮發性(Non-volatile)：當我們將過去的資料轉入資料倉儲後，就不應該加以變更，但是可以以固定且週期性的方式定期加入新的歷史資料。

本文於表一整理了關聯式資料庫與維度資料倉儲之比較分析，希望有助於對資料倉儲的概念有進一步的了解[9]。

表一、關聯式資料庫與維度式資料倉儲之比較

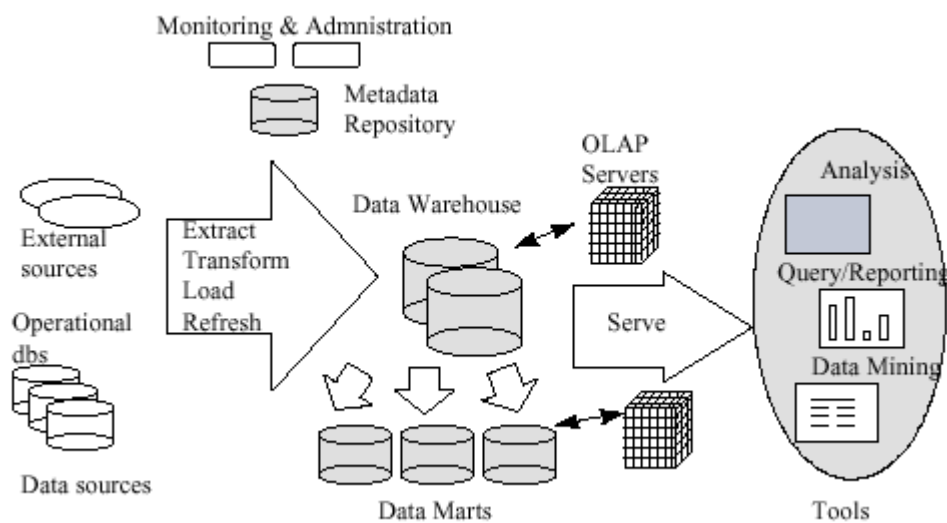
特性	關聯式資料庫	維度式資料倉儲
資料的時間性	當時的運算資料	經過處理的歷史資料
資料庫的綱要設計	個體-關係模式且正規化	多維度模式
資料特性	無重覆儲存	大量重覆儲存並預先加總
異動的頻率	經常異動(故稱 OLTP)	少有異動，大多為查詢
異動的資料數量	平時均有大量異動處理	定期大量載入，並聚合加總
查詢的頻率	少量的需求，且少有分析	大量需求(故稱 OLAP)，並予分析
查詢複雜度	較單純	相當複雜
所含資料量	Megabytes	Gigabytes
整合性	依功能分資料庫，未整合	整個組織的資料完全整合
主題性	依功能導向區分	依主題導向區分

暫存性	只保留目前最新的資料	完整保留所有的歷史資料
-----	------------	-------------

(二) 資料倉儲之架構

資料倉儲的資料從同質或異質、外部或內部的作業性資料庫萃取出來，經過資料轉換服務(DTS, Data Transformation Services)進行資料轉換或是將資料匯入資料倉儲，而資料轉換服務(DTS)包含了資料驗證(Data Validation)、資料搬移(Data Migration)、資料淨化(Data Scrubbing)、資料轉換(Data Transformation)的四個程序，如圖一所示[13]。其中 Metadata 扮演著監看與管理資料轉換的功能。

資料進入資料倉儲後，依不同的資訊需求加以分類，從資料倉儲中切分出數個資料市集(Data Marts)，透過前端的線上分析處理軟體(OLAP Services)提供不同應用層面的使用者獲得所想要的資訊[41, 62]。



圖一、資料倉儲的架構

(三) 資料倉儲之架構分析

架構分析包括了資料倉儲系統的處理程序介紹、系統的管理以及資料架構的設計。。系統的管理則有載入管理、倉儲管理、查詢管理。

1) 資料倉儲系統的處理程序

處理程序計有資料的萃取與載入、資料的整理與轉換[13]。

1. 資料的萃取與載入(Extract and Load)

資料萃取是將資料由來源系統中拿出來並使得它們可以為資料倉儲所使用；資料載入是將資料萃取步驟所得到的資料實際載入資料倉儲。重要的是，我們並不是要把所有在一般作業性資料庫的資料都存入資料倉儲中，而是儲存我們所需要的資料。此外，為了使得資料倉儲有更好的執行效能，我們也可以將一般作業性資料庫中的資料加以處理，而以衍生資料的型式儲存在資料倉儲中。

當資料由來源系統(Information Sources)萃取出來後，並不是直接放在資料倉儲中，而是儲存在暫存媒體(Staging Server)中，以便執行資料整理、轉換以及更正錯誤的工作，因為資料整理是全面性的，所以應該等到所有的資料都載入暫存媒體後再開始執行資料整理的工作，因為局部性的資料正確並不表示在整體的資料架構中是正確的。

2. 資料的整理與轉換(Organize and Transfer)

資料整理與轉換程序主要是將載入的資料組織成適於查詢的結構，以增加查詢的執行效能。在資料整理與轉換程序中包含了以下步驟：

- (1) 整理並轉換載入的資料為可加速查詢的資料結構。
- (2) 將資料加以分割，以加速查詢速度及最佳化硬體執行效能。
- (3) 建立彙總資料以加速一般的查詢。

在資料倉儲中，因為事實資料必須是正確的，為了提供輔助各種使用者的決策依據，所以在資料的整理、驗證與轉換的過程是很重要的。

2) 系統的管理

系統管理由多個管理程式所構成，區分如下[13]：

1. 載入管理(Load Manage)

載入管理的主要工作是支援資料萃取與載入的工作，它可以由一些外購的軟體工具、針對特殊的需要而撰寫的程式。載入管理的功能與複製度會因資料倉儲

的資料來源而異，若資料來源的資料重疊性高，則載入管理的複雜度也越高。

2. 倉儲管理(Warehouse Manage)

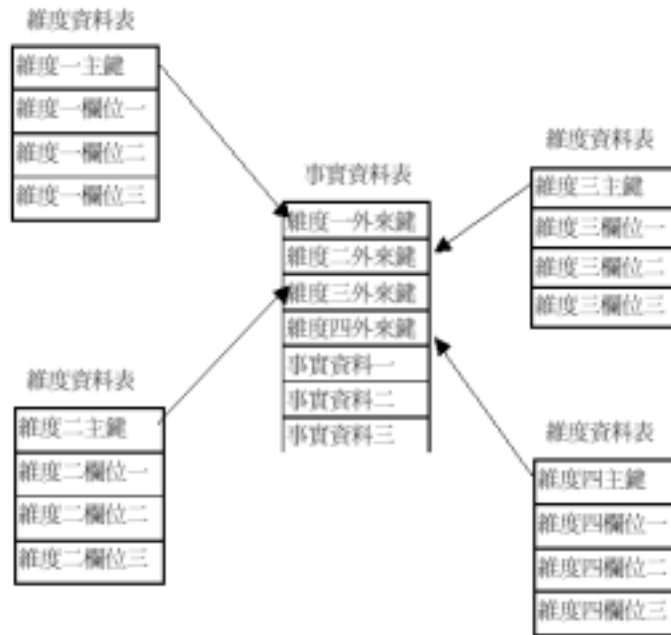
倉儲管理的主要工作是執行管理一資料倉儲的所有必要程序，包含：檢驗各欄位相互之間的關聯性與一致性、對倉儲的資料新增索引或資料分割、依需要產生新的彙總資訊、備份資料倉儲、備份資料倉儲過時的資料等。

3. 查詢管理(Query Manage)

查詢管理的主要工作是將查詢導引至正確的資料表或為所有的使用者查詢排程。查詢管理將其對查詢統計分析的結果存入查詢概觀檔，以提供倉儲管理使用，以決定要為那些項目執行資料彙總的工作。

3) 資料架構的設計(Data Schema Design)

資料倉儲的資料包括事實(Fact)資料和維度(Dimension)資料，就一項統計分析的範疇而言，它包含了一個事實資料表(Fact Table)和多個維度資料表(Dimension Table)，一個星狀綱要(Star-Schema)的維度資料表只會和事實資料表產生關聯性，維度資料表和維度資料表之間並不會產生關聯性，除非是雪花 Schema 或星狀雪花 Schema 的資料表架構，一般而言，一個資料倉儲的最適用綱要類型是將反正規化的星狀 Schema 與正規化的雪花式 Schema 合併在一起使用，而成為星狀雪花 Schema。圖二顯示了星狀 Schema 資料表架構的事實資料表和維度資料表[13]。



圖二、星狀綱要(Star-schema)的資料表架構[13]

(四) 資料倉儲與決策支援系統

一般來說，決策支援的目的是希望能在策略的規劃上產生可行方案，以作為制定決策的依據。所以，決策支援的有效與否，取決於「資料品質」的好壞，而高品質的資料來源則有賴於資料倉儲的建立。一個決策支援系統的架構，包含了以下四個部份[9]：

(一) 資料儲存體(Data Store Component)

即是資料倉儲中所置放的資料，其中包括了由運算資料所彙總過來的資料，以及來自其他資料來源的整合資料。此外，也有用來描述行為的模式庫資料。

(二) 資料萃取與過濾模組(Data Extracting and Filtering Component)

用來萃取、過濾並驗證那些來自傳統資料庫中的資料，以取得可以產生決策的資料。這些模組需能夠支援批次或可排程的資料萃取作業，同時也要能夠支援不同型式的資料轉換功能。

(三) 使用者查詢工具(End-User Query Tool)

用來協助使用者建立查詢。

(四) 使用者展示工具(End-User Presentation Tool)

協助使用者選擇最適合呈現資料分析結果的型式，如圓餅圖、長條圖等。

二、資料倉儲建置策略

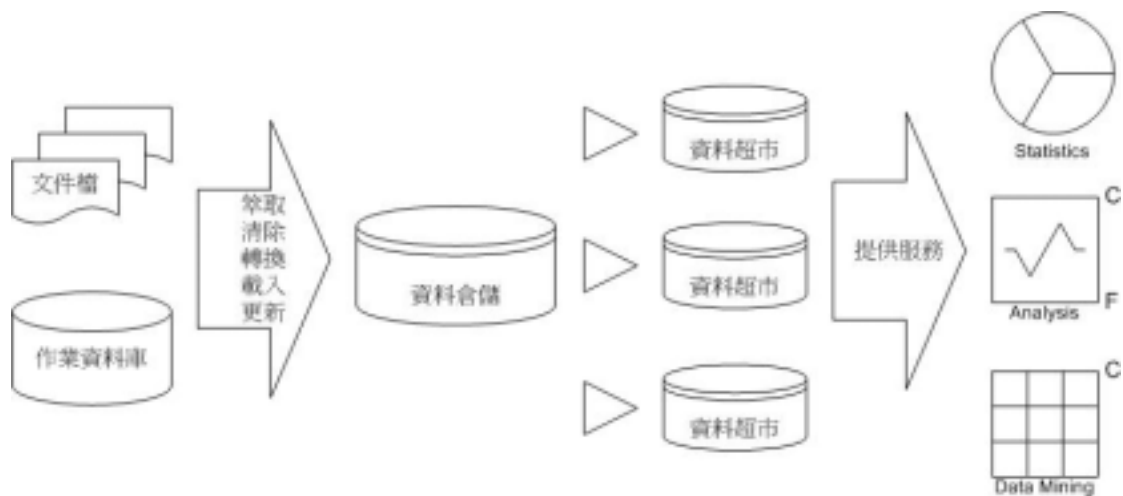
資料倉儲建置時，必須考量各項企業因素，如企業大小、初期欲投入的成本花費、查詢需求的多樣性、未來擴充需求、轉置成本、維護成本、建置時程及建置風險的控管等等，所以會衍生出多種建置資料倉儲之架構及策略。另外除了考慮建置策略之外，為了擁有更全面性的規劃及思考，還可以進一步的探討如何達到資料倉儲的效能最佳化，加速日後查詢需求的無限增長，而能快速反應使用者對於資料倉儲的多維度查詢、檢視及分析的需求。資料倉儲系統常用的設計架構在本研究蒐集眾多文獻之後，歸納出下列三種模式：三層式、兩層式及混合漸進式三種。本節將就此三種架構進行優缺點說明[5][62]及提出綜合比較與分析。其結果可提供作為建置 NIIS 中央型資料倉儲時之參考。

(一) 三層式架構

理想的資料倉儲系統在設計上大多採用三層式建置架構，如圖所示[1]。資料來源端是企業營運資料的來源處，例如，用戶基本資料、報表資料、作業性資料庫系統及用戶問卷資料等等。這些資料可能分散在各個不同的地理區域，或公司的各部門。將這些來自不同來源端的資料進行資料的前置處理(data preprocessing)的動作，即萃取(Extract)、淨化(Cleaning)、轉化(Transform)、載入(Load)及更新(Refresh)等動作，如此則可確保資料的純化與正確性，然後整合匯入資料倉儲系統之中。所以，資料倉儲系統可以說是整個企業資料的彙總處，提供一個日後線上分析處理的基礎。

然而，基於企業不同事業單位的個別需要（如生產、行銷、業務各部門）以及各區處用戶之區隔，可將資料倉儲系統的資料再分成若干個「資料超市(Data Mart)」。如此一來，在資料倉儲部份著重於來源資料的淨化及整合，在資料超市

部份則著重於使用者未來分析查詢的需要。這樣的做法有下列幾項優點：



圖三、三層式資料倉儲建置架構

(1) 加快前端資料分析速度(performance)

由於資料超市乃是針對企業各單位本身的需求設計，且資料量較少，系統負荷也較資料倉儲系統為輕。此外，資料超市根據單位功能劃分，分別存放在各部門內，如此一來便縮短傳輸上的距離，進而加速查詢速度。

(2) 滿足多樣化的查詢需求(Scalability)

透過三層式的架構，不同需求的查詢可以在不同的資料超市中實現。建立不同性質的資料超市，不僅可以滿足不同類型使用者的查詢需求，再加上由於是三層式的架構，如果因為新的需求而必需增加新的資料超市時，整體資料倉儲系統的架構並不需要太大的改變。

(3) 切合組織的需要(Flexibility)

特別是大企業，將資料倉儲系依據組織功能架構來劃分建立資料超市，避免因集中管理而產生無法有效配合各單位需求的情形；不僅在使用上更容易存取，並且更能反應出系統的彈性。

但是三層式的架構也有其缺點。最主要的缺點在於系統建置初期的成本過

高。所以三層式架構較不適合小規模的企業，以及一些先期的開發研究。

（二）兩層式架構

除了三層式架構之外，兩層式資料倉儲架構也列入許多企業在發展資料倉儲系統時喜歡採用的方案。與三層式架構不同的地方是，兩層式架構中沒有中央的資料倉儲，如圖所示[1]。就成本的角度來看，它是非常具有吸引力的解決方案，其所擁之優點為：

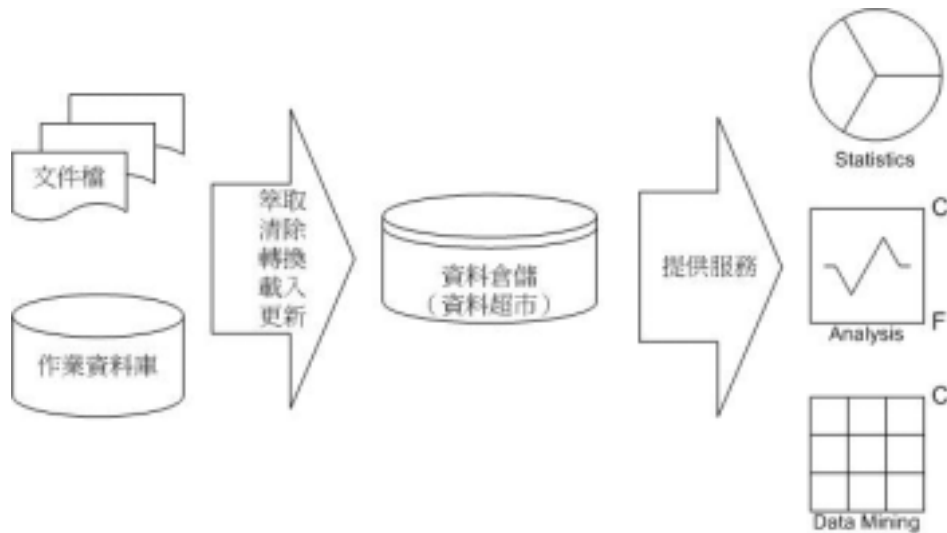
（1）降低硬體需求

由於資料超市的硬體需求較小，企業在初期的投資成本比較低，所以在企業規模不大或是資料量不大時，比較符合經濟效益。

（2）較容易建置

由於系統架構比較簡單，系統開發的時間可以縮短，技術問題也較容易克服。

然而，兩層式的資料倉儲架構雖然比較適合小規模的企業，但由於其日後擴充規模的彈性不大，將使得企業在系統設計上面臨兩難的局面。例如，一旦有新的需求需要整合到系統中，既有資料超市的軟硬體處理能力由於在事前沒有適當的規劃考量，預留其擴充性，它將難以勝任資料整合時造成的負荷。此時，若想只靠增加一個新的資料超市來滿足這些需求，勢必造成資料源系統的負荷，而且，新資料超市與既有資料超市的資料整合也將是一大問題。



圖四、二層式資料倉儲建置架構

(三) 混合漸進式架構

混合漸進式的資料倉儲系統架構相較於前面介紹之二者，顯然具有未來系統擴充的彈性。當然，它除了保有日後升級的彈性之外，更重要的是它的初期成本與兩層式架構相差不大。換言之，如果企業日後確實有擴大其資料倉儲系統的可能性，或是為減少系統開發初期風險的情形下，發展混合漸進式資料倉儲系統是一個十分適合的解決方案。站在系統開發的角度，混合漸進式架構也較適合「雛型法」的精神，畢竟在時間與系統品質的要求下，貿然直接進行複雜的系統開發，所負擔的風險非常高。



圖五、混合漸進式資料倉儲建置架構

如圖五所示[1]，基本上，混合漸進式的資料倉儲系統同樣採用兩層式的架構，但是在第二層架構並非只是資料超市，而是事先根據未來可能擴充之資料，規劃完整之資料超市與資料倉儲雛型，並在初期之時先行發展各部門之資料超市，等到發展成熟之後，再根據之前規劃之相關雛型及整合方案，進行全面化之資料倉儲系統，提供做為公司整體決策及分析之用。這樣的設計相較於二層式架構多考慮了日後調整架構的可能性。根據此一架構之設計，我們可以發現，混合漸進式資料倉儲架構有以下幾項優點：

- (1) 節省系統開發初期成本
- (2) 提供日後擴充的彈性
- (3) 系統維護較三層式容易
- (4) 適合成長中企業使用以及系統開發時的雛型
- (5) 縮短發展時程
- (6) 降低系統發展的風險

(四) 建置策略之綜合比較分析

在介紹完三項資料倉儲建置策略之後，本文分別就其優點、缺點及適用對象作一客觀性之比較，製表如表二。

表二、資料倉儲建置策略之綜合比較及分析

	優點	缺點	適用對象
三層式架構	<ol style="list-style-type: none"> 1. 加快前端資料分析速度 2. 滿足多樣化的查詢需求 3. 切合組織的需要 	<ol style="list-style-type: none"> 1. 系統建置初期的成本過高 	大型企業

二層式架構	<ol style="list-style-type: none"> 1. 降低硬體需求 2. 較容易建置 	<ol style="list-style-type: none"> 1. 擴充規模的彈性不大 2. 新資料超市與既有資料超市的資料整合 	中小型企業
混合式架構	<ol style="list-style-type: none"> 1. 節省系統開發初期成本 2. 提供日後擴充的彈性 3. 系統維護較三層式容易 4. 適合成長中企業使用以及統開發時的雛型 5. 縮短發展時程 6. 降低系統發展的風險 	N/A	成長型企業

我們可以從上表發現，混合型之建置策略擁有相當多的優點，無論是在初期成本、擴充性、發展時程、系統風險等都有不錯的表現。

三、線上分析處理技術

線上分析處理(On-line Analytical Processing, OLAP)系統是一個用來幫助使用者能有效率並且有效的輕易完成資訊的維度結構(Dimensional structure)分析工作。OLAP 允許使用者自行選擇在樹狀資料結構中，展開同一階層資料的明細或點選更細一層的資料明細或樞紐，並且 OLAP 還支援某些階層的資料計算能力。而資料倉儲(Data Warehouse)與資料市集(Data Mart)則是 OLAP 的資料來源。OLAP 通常與查詢的功能息息相關，經由複雜的計算能力(資料比對、資料擷取等)來提供各種分析模式 [1]。

線上分析處理(OLAP)系統主要將資料倉儲中的資料轉換至多維面(Multi-dimension)結構，如圖所示[10]，並且呼叫 Cube(OLAP 的主要物件，包含多維面結構的資訊，每個 Cube 是由一群 Dimensions 及 Measures 所組成)，來執行有效及非常複雜的運算查詢。

多維面的 OLAP 主要提供一個簡化複雜問題與查詢困難度的服務。當那些常被使用者查詢的問題都能夠以有規則性的呈現出這些資料的數量時，這大概表示 OLAP 已經趨近完善。要強調的是，不同的 OLAP 設計模式源自於不同的使用者需求。以下為 OLAP 的特性[12]：

1. 歷史性：OLAP 資料來自於資料倉儲，其最大的特色在於資料具有歷史性，且通常涵蓋相當長的一段時間。
2. 唯讀性：由於資料倉儲的資料是歷史性的，這意謂著資料也具有唯讀性。一般而言，只有在原始資料來源有錯，才會變動資料倉儲的資料。
3. 群組分類性：OLAP 資料強調所有細節的資料，並重視資料的分類以利於統計分析。
4. 合併性：將企業內所有的異質性資料，經過資料驗證、資料遷移、資料篩選及資料轉換的方式全部彙集合併在資料倉儲中。
5. 一致性：包含了資料格式一致與資料單位一致，如此以方便於資料的分析與比較。

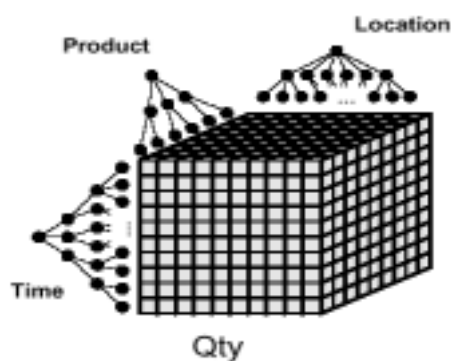
OLAP 資料模型 (Data Model)

在 OLAP 資料模型中，資訊是以 cubes 來表示其包含了以下二個項目：

- Dimensions：是一敘述性項目，例如時間、地點、產品、部門 等。
- Measures：是一量化的項目，例如銷售額、單位銷售量、存貨量、收入 等。

在多維面的資料模型中，可以簡化一些使用者的複雜的查詢、製作報表、對資料的計算加總、將資料過濾後分割更具商業意義的細部資料。

在 OLAP 資料模型的每個維面 (dimension) 中，資料可以組織成階層式的繼承觀念。舉個例來說，在時間維面中，您可以將它分成年、月、日，在地理域上，您可以將它分成國家、地區、州 / 省以及城市，在每個維面階層中都會有其分析過後的數值，使用者可以用很簡單方式在這些維面階層中移動來看各階層的分析數值。



圖六、多維面的資料結構

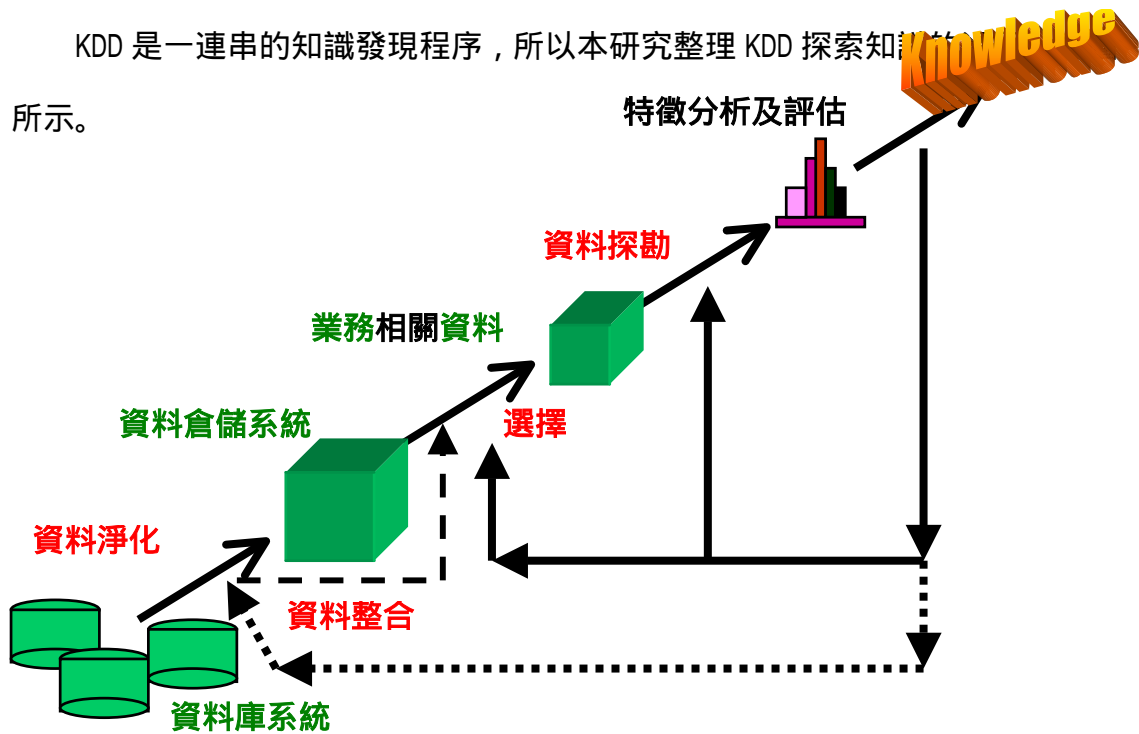
因此，醫療資訊使用者透過線上分析處理工具的使用，可彈性化訂製其所需的報表，藉由多維面的資料架構，以快速取得具結構性的資料。

四、知識探索技術

知識探索 (Knowledge Discovery in Databases, KDD) 是指從資料中發現有效，嶄新，有潛在效益知識的一個非細瑣 (nontrivial) 流程，其最終目標是瞭解資料的特徵 (Pattern)。KDD 是知識發現的一連串程序，Data Warehouse 及 Data Mining 只是 KDD 的一個重要程序 (Process)，其最終目的乃為組織取得決策支援所需的資訊。

(一) KDD Process

KDD 是一連串的知识发现程序，所以本研究整理 KDD 探索知识的过程如图所示。



圖七、知識探索過程

在我們擁有企業資料庫及資料倉儲的前提之下，我們可以進一步快速地從資料來源端取得所要分析及探索之資料，經由資料挖掘技術及特徵評估二階段，進一步獲得我們所想要的知識，所以在下一個小節將介紹何謂資料挖掘技術及其使用之效益。

(二) 資料探勘

資料探勘技術是在龐大的資料庫或資料倉儲之中尋找出有價值的隱藏事件，如趨勢(trend)、特徵(pattern)及相關性(relationship)，並且加以分析；或者是從資料庫或資料倉儲中獲取有意義的資訊及對資料歸納出有結構的模

式，以作為企業在進行決策時之參考依據。Berry and Linoff (1997) 認為「資料探勘」就是針對大量的資料，利用自動學習或配合經驗的方式予以分析解釋，以找出有意義的關係或法則[40]。Cabena et al. (1997) 認為「資料探勘」是將未知且有效的資訊從大型資料庫抽出的過程，並且將所獲得的資訊提供給決策者做為決策的參考依據。Grupe and Owrang (1995) 認為「資料探勘」的運用在於從既有的資料中分析出及發現尚未被找到的關係與共識[51]。

其運用電腦儲存運算能力及使用各種工具，其中亦包括傳統的統計方法。與資料挖掘習習相關的技術包括：

1. 資料庫系統、資料倉儲、OLAP
2. 機器學習 (Machine learning)
3. 統計及資料分析方法 (Statistical and data analysis methods)
4. 視覺化技術 (Visualization)
5. 數學規劃 (Mathematical programming)
6. 高效能計算 (High performance computing)

Data mining 的方法主要有以下五種方式，分別是分類(classification)、推估(Estimation)、預測(Predication)、關聯式法則(association rule)及分群(clustering)。分別簡述如下：

1. 分類 (Classification)
 - 1) 按照分析對象的屬性分門別類加以定義，建立類組 (Class)。
 - 2) 例：將信用申請者的風險屬性，區分為高度風險申請者，中度風險申請者及低度風險申請者。
 - 3) 使用的技巧有決策樹(Decision Tree)，記憶基礎推理(Memory-Based

Reasoning)。

2. 推估 (Estimation)

- 1) 根據既有連續性數值之相關屬性資料，以獲致某一屬性未知之值。
- 2) 例：按照信用申請者之教育程度、行為別來推估其信用卡消費量。
- 3) 使用的技巧包括統計上之相關分析、迴歸分析及類神經網路方法。

3. 預測 (Prediction)

- 1) 根據對象屬性之過去觀察值來推估該屬性未來之值。
- 2) 例：由顧客過去刷卡消費量預測其未來刷卡消費量。
- 3) 使用的技巧包括迴歸分析、時間數列分析及類神經網路方法。

4. 關聯式法則 (Association Rule)

- 1) 從所有物件決定那些相關物件應該放在一起。
- 2) 例：超市中相關之盥洗用品(牙刷、牙膏、牙線)，放在同一貨架上。
- 3) 在客戶行銷系統上，此種功能係用來確認交叉銷售(Cross-Selling)的機會以設計出吸引人的產品群組。

5. 分群 (Clustering)

- 1) 將異值母體中區隔為較具同質性之群組(Cluster)，其目的是要將組與組之間的差異辨識出來，並對個別組內之相似樣本進行挑選。
- 2) 同質分組(Clustering) 相當於行銷術語中的區隔化(Segmentation)，但是，此為假定事先未對區隔加以定義，而讓資料自然產生區隔。
- 3) 使用的技巧包括 K-Means 法及 Agglomeration 法。

資料挖掘的常見應用有以下三方面：

1. Customer Profiling

- Data Mining 可以從現有客戶資料中找出他們的特徵，再利用這些特徵到潛在客戶資料庫裡去篩選出可能成為我們客戶的名單，作為行銷人員

推銷的對象。

- 只針對這些名單寄發廣告資料，可降低成本，也提高行銷的成功率。

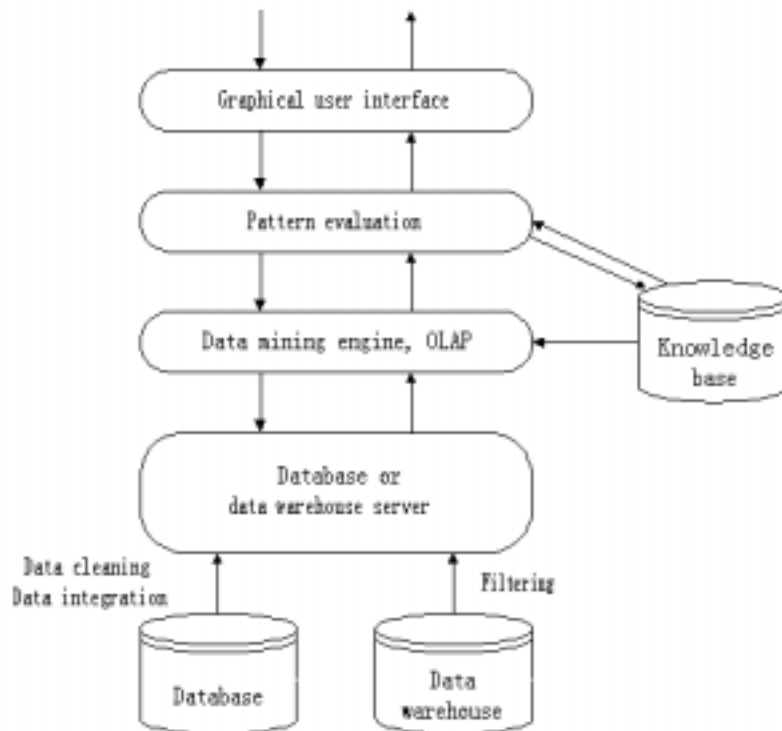
2. Targeted Marketing

3. Market-Basket Analysis

- 主要是用來幫助零售業者瞭解客戶的消費行為，譬如哪些產品客戶會一起購買，或是客戶在買了某一樣產品之後，在多久之內會買另一樣產品等等。
- 利用 Data Mining，零售業者可以更有效的決定進貨量或庫存量，或是在店裡要如何擺設貨品，同時也可以用來評估店裡的促銷活動的成效。

資料探勘的進行步驟及架構大略可以由下列八個步驟來組成：

1. 理解資料與進行的工作
2. 獲取相關知識與技術(Acquisition)
3. 融合與查核資料(Integration and checking)
4. 去除錯誤或不一致的資料(Data cleaning)
5. 發展模式與假設(Model and hypothesis development)
6. 實際資料探勘工作
7. 測試與檢核所探勘的資料(Testing and verification)
8. 解釋與使用資料(Interpretation and use)



圖八、典型的資料探勘系統架構

資料探勘建置的四個階段大致如下。

1. 目標(target)設定

分析現有的 business process 以確認 data mining 可以應用之領域，這些領域可能包括行銷、銷售、顧客服務等等。另外，亦要在此階段，從事使用者訪談、蒐集資料等工作。其次，將資料按可能使用的模型如以 clearing、filtering 及 transformation。

在此階段應產生下列各項相關文件資料：

- 有關 data mining 先導計畫實施目的之說明
- 評估計畫之評估準則
- 資料整理結果及初步分析報告
- 計畫時程
- 最後目標之大綱

2. Prototyping

從第一階段獲致結果，使用合適的軟硬體從事系統 prototype 模型之開發。在開發 prototype 的過程中，而同時修正資料之整理方法模型之建立。在本階段應完成下列各項工作：

- Prototype 模型開發系統
- data mining 技術及工具之評估
- 調整 business process 及 data mining 系統整合之計畫
- data mining 環境設定及開發計畫

3. 系統建置

在此階段應產生下列各項：

- 資料擷取及整理之程序及軟體
- data mining 模型開發系統
- data mining 上線之第一個版本
- solution 的執行及 migration plan

4. 系統移植(migration)

Data mining 文化的建立及使用者的訓練，在此階段應產生下列項目：

- data mining 之上線環境
- data mining 使用者清單
- business process 對 data mining 之回應
- 系統改善計畫
- Data Mining 計畫的擬定
- 問題的確認

以銀行活期存款帳戶流失率之估計為例，銀行希望能夠預估客戶之流失，而在尚未太遲的情況下瞭解潛在的問題。

1. 資料來源

一般的交易資料可能不足以用來估計銀行活期存款帳戶之流失率，必須再蒐集資料，以瞭解客戶流失之原因。

2. 資料需求的界定

找出針對與特定問題相關原因與象徵之資訊。

3. 訪談人員需求

訪談之被訪人可能包括服務中心人員，分行經理、及行銷分析人員等。從事訪談的人員則以從事流失模型建立之分析為宜。

4. 模型建立

模型的種類可以涵蓋簡單的 OLAP，以致複雜的 neural network。

5. 資料整理

不同的模型有不同的資料需求，資料整理方式也不同。例如在 neural network 模型的情況，可能要將原始資料轉換成以 0 至 1 為範圍之數列。

6. 軟體需求

利用的原有的交易資料及額外蒐集的資料後必須利用專業的軟體建立模型。所需要的軟體可能包括 SQL queries 及特殊的分析軟體。

7. 資料倉儲的支援

在 Data mining 的應用上必須以資料倉儲作為支援。因此，在建置資料倉儲時必須考慮到 Data mining 的應用，例如具有關聯性及下拉式的 OLAP 核心對 Data mining 有相當大的助益。

(三) 資料探索應用在醫學研究上必須具備下列條件

一般來說，應具備以下五項條件。

1. 優良的演算結果：演算法必須能夠在可獲得的資料中選用有意義的資訊。而診斷結果的準確性必須要越高越好，至少不可低於具有相同資訊的醫生的準確度。
2. 要有解釋能力：在演算法做出診斷結果之後必須要能夠對醫生解釋，診斷為該結果的原因。否則要是醫生無法得知該演算法之所以會有診斷結果的因素的話，此種診斷結果是不被醫生所接受的。
3. 診斷知識的透明化：經由演算法所產生的知識必須要能夠對醫生透明讓醫生瞭解診斷結果的原因，在理想上甚至希望經由演算法所產生的知識能讓醫生對這些疾病能夠有新的觀點，並且能夠以最為詳盡的形式顯示出醫生之前不

曾遭遇到的新的相互關係或是規則性。

4. 有能力處理遺失或是雜亂的資料：因為當病人在看診時本身會忽略敘述某病徵或是醫生在詢問病人因當時的清況而對病人的情形有所遺漏，另外在記錄上也許會記載與病情無關的紀錄，此時所選用的演算法必須要有能力處理這類型的問題。
5. 能減少演算法學習以及測試的次數：由於所有的資料都收集是來自病人本身，因此在考量成本、時間外，就是某些檢查對病人是有害的，因此在測試資料的來源不易之下演算法本身應該能夠盡量的減少訓練或是學習的樣本數。

五、決策支援系統與統計預測模式

一般來說，決策支援的目的是希望能在策略的規劃上產生可行方案，以作為制定決策的依據。所以，決策支援的有效與否，取決於「資料品質」的好壞，而高品質的資料來源則有賴於資料倉儲的建立。一個決策支援系統的架構，包含了以下幾個部份[9]：

1. 資料儲存體(Data Store Component)

即是資料倉儲中所置放的資料，其中包括了由運算資料所彙總過來的資料，以及來自其他資料來源的整合資料。此外，也有用來描述行為的模式庫資料。

2. 資料萃取與過濾模組(Data Extracting and Filtering Component)

用來萃取、過濾並驗證那些來自傳統資料庫中的資料，以取得可以產生決策的資料。這些模組需能夠支援批次或可排程的資料萃取作業，同時也要能夠支援不同型式的資料轉換功能。

3. 使用者查詢工具(End-User Query Tool)

用來協助使用者建立查詢。

4. 使用者展示工具(End-User Presentation Tool)

協助使用者選擇最適合呈現資料分析結果的型式，如圓餅圖、長條圖等。

表三、統計預測法之分類

種類	方法
時間序列	簡單迴歸分析 (LR) 移動平均/指數平滑 (MV, WMV, EXPOS) 古典分解/普查 X-11 Holt's-Winters 傅立葉序列 ARIMA(Box-Jenkins)
因果法	多元迴歸 (MLR) 經濟計量 循環 多變數 ARIMA(Box-Jenkins) 狀況空間 向量自動迴歸
定性 / 科技的	專家觀點 銷售力組合 DELPHI (德菲法) 歷史比對 S-成長曲線 AHP
其它的定量法	市場研究 (MR) 管理科學/作業研究 (OR) 專家系統 (ES) 人工類神經網路 (ANN) 遺傳演算 (GA, GP) 組合法

六、類神經網路

一般而言類神經網路有以下分類，各有其運用的領域。

1. 監督式：利用訓練範例學習，找出其映對規則，應用於新案例，適用於分類、預測
2. 非監督式：利用訓練範例之內在聚類規則，適用於聚類
3. 聯想式：利用內在記憶應用於新範例，適用於雜訊過濾、資料擷取
4. 最適化應用：與聯想式類似，適用於設計、排程

一般而言，類神經網路是由許多個運算元所組成。一個基本的類神經網路架構是由運算元之連結方式、轉換函數和學習法則所決定。本文在此對類神經網路之架構說明如下。

(一) 運算元(Processing Element, PE)

亦可稱為處理單元，為組成類神經網路之最基本單元。運算元主要是接受輸入值後彙總將其輸出，其作用可用三種函數來加以說明。

(1) 集成函數(summation function)

集成函數是用以將從其他處理單元之輸出透過網路連結傳來的訊息加以綜合。如加權乘積和、歐式距離等.....。

(2) 作用函數(activity function)

其目的將經由集成函數所傳來的值和該運算元目前的狀態加以彙總。作用函數之功能一般來說並不明顯，所以一般類神經網路大多採取集成函數之輸出。

(3) 轉換函數(transfer function)

轉換函數的目的在把作用函數之輸出值轉換成運算元之輸出值。如：屬於線

性函數的Perception 函數(或稱階梯函數step function)、Hopfield-Tank 函數、Signum 函數及非線性函數的雙彎曲函數、雙曲線正切函數等等...都是在類神經網路上在使用的轉換函數。

(二) 層(Layer)

由數個功能相同的運算元組成，可區分為三種，分別是輸入層、隱藏層及輸出層。層本身也有三種功能，描述如下：

(1) 正規化輸出

正規化輸出之目的在於將位於同一層中的處理單元的原始輸出值所組成的向量加以正規化，成為單位長度的向量後，再作為層的輸出。

(2) 競爭化輸出

競爭化輸出的目的在於將位於同一層中之處理單元的原始輸出值所組成的向量中，選擇一個或數個較強值的處理單元，令其值為1，其餘值為0後，再作為此層的輸出。這些輸出值為1的處理單元稱為優勝單元(winner)。

(3) 競爭化學習

競爭化學習將位於同一層中之處理單元的原始輸出值所組成的向量中，選擇一個或數個優勝單元，而後網路只調整這些與優勝單元相連的下層網路連結。

(三) 連接鍵(connection)

為類神經網路基本要素之一，其主要作用為連接層與層之間的運算元，做為其傳遞訊號路徑之用。

(四) 網路(network)

將層與層中之運算元利用連接鍵相互連接而成網路。根據處理單元連接的方式，可分為向前式(feed forward)和回饋式(feedback)兩種架構。向前式架構的運算元分層排列，形成輸入層、隱藏層和輸出層，每一層只接受前一層的輸出做

為輸入；回饋式的輸出層之輸出會回饋到輸入層，或者層內各處理單元間有連結者，或者運算元不分層排列，只有一層。類神經網路中網路的運作模式可以分成學習過程與回想過程兩種。

(1) 學習過程(learning)

不同的網路模式有不同的學習法則，在學習過程中，不斷的調整網路中的權數，若網路達到穩定狀態，學習過程就可以中斷。而學習的演算法又可分為三種。

- 監督式學習(supervised learning)

在學習的過程中，每一個輸入之資料均有一個期望之輸出值對應，學習的目標在調整權數以降低網路輸出值與期望值之差距，由於需要不斷的學習、調整。一般而言，監督式學習在學習過程中需要花較多時間才能得到一個較好的結果。

- 非監督式學習(unsupervised learning)

在非監督式學習網路中，每一個輸出單元代表一個聚類結果，因此網路學習的目的在降低網路優勝單元的連結加權值所構成的向量與輸入向量間之差距。學習之品質和監督式學習一樣可以用能量函數來表示。利用此種學習方式的網路有自適應共振網路等。

- 聯想式學習(associative learning)

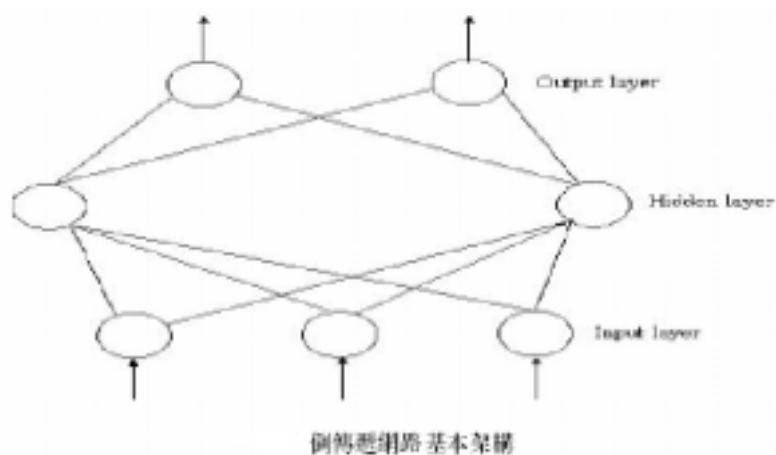
此種學習方法是利用運算元組成的向量來表示模型(pattern)，網路學習旨在使從一開始所組成的初始樣式，經過聯想的過程得到最終樣式。此種學習方式又可分為自聯想學習與異聯想學習兩種，例如：霍普菲爾網路與雙向聯想記憶網路等。

(2) 回想過程(recall)

經過學習過程的學習之後，類神經網路會將學習之結果儲存在網路的連接權數

中，因此我們將資料輸入網路之後，經過一些數學運算即可獲得輸出結果，此過程稱之為回想過程。

在所有的類神經方法中，倒傳遞神經網路(Back-Propagation Network)是應用最普遍的演算法，他的成功案例多、學習精度高、回想速度快，加上理論簡明使的常為大家所使用。一般而言，倒傳遞網路常被應用於幾個領域，包括專家系統、樣本識別、適應控制、雜訊過濾及資料壓縮 等。



圖九、倒傳遞類神經網路基本架構

倒傳遞網路分為兩個階段，第一階段是所謂的學習階段，依學習演算法，從範例中學習，找出其網路連結加權值的過程，然後進入第二階段，回想階段，以輸入資料決定網路輸出資料的過程。

七、資料庫效能測試與效益分析

一、資料庫效能指標測試

本小節將介紹資料庫管理系統績效評估之基本概念及相關組成要素，並在文中針對目前較常用來評量績效之理論及方法深入介紹，並於本章結尾提出一綜合性之方法比較。希望此章之內容有助於 NIIS 系統在選擇中央資料庫規格時可以

有一個較為客觀且公平之參考。

(一) 績效評估的定義

績效評估是一套評估效能的方法或技術。通常績效評估定義一組規則用來比較相同系統或不同系統下的效能，因此我們可以定義績效評估是：一組標準的指令用來比較不同系統的相對績效[50]。

一般而言，用來測試效能的規則稱之為標竿測試 (Benchmark) 或績效評估 (Performance Evaluation)。績效評估提供了系統建置及使用的參考標準，其主要目的是提供不同系統間績效評估的標準，或是廠商用來彰顯其系統的特色，亦或是用以設計某個可預測系統績效表現的模式。除此之外，Benchmark 也可提供系統修改前後所產生的績效影響之相關數據，作為評量其改善成效之參考。就績效評估的實際用途而言，可將其分為四類，如表四所示：

表四、績效評估的用途分類

1	比較同一個應用軟體在不同的軟硬體平台的執行效能
2	在同一台機器上測試不同的應用軟體的執行效能
3	相容之不同機器的效能
4	比較在同一機器執行的不同版本的應用軟體之效能

針對一套資料庫系統，以一套規定好的規則去進行測試，不同種類的測試有不同的比較指標 (Metric)，此即是資料庫績效評估。例如有些績效評估以一定時間內所完成的異動數為指標，或以秒異動數、甚或以價格/成本比為衡量指標，這些都端視系統的需求而定，然而，重要的是這些衡量指標沒有所謂的絕對標準存在，當我們在進行比較的動作時，只比較兩者在設定指標上的相對數字表現。

(二) 績效評估的特色

由於沒有一個簡單的指標能夠衡量電腦系統在所有的應用程式上的績效。系統的表現可從一個應用領域到另一個，而每個系統是用來設計解決一些特定的 (Domain Specific) 問題領域，因此極有可能無法解決其他的任務。例如在資

料庫系統領域裡，分成了各式各樣不同的問題領域，如某系統可在對線上的、簡單的、多維更新動作上表現優良，但卻在複雜查詢動作上表現較差，反之亦是。領域專屬的標竿 (Domain-Specific Benchmark) 也就因應而生，每個此類的標竿皆規範了一個綜效的工作量 (Synthetic Workload)，可用來特徵化該問題領域的標竿問題。領域專屬的績效評估在不同系統上的執行，以得出在相同領域裡不同系統的表現，然而，領域專屬之標竿需達成四個重要的特色，如表五所示：

表五、領域專屬標竿的特色

相關性 (Relevance)	在某問題領域中能衡量其最佳表現和價格/表現比
可攜性 (Portability)	在不同架構下能輕易的運用
延展性 (Scalability)	能同時適用於小型至大型的系統
簡單性 (Simplicity)	必須是易於了解的，太困難則缺乏可信度

(三) 績效評估的目標

績效評估的目標主要是做為選購軟硬體的指標，如果沒有績效評估的話，就沒有比較公正信服的基準，消費者只能由廠商的廣告憑空猜測誰的系統效能較佳。績效評估的主要目標如表六所示：

表六、績效評估常見的目標

1	彰顯某系統的特色所在，如資料庫的設計使得查詢的速度較快
2	比較不同系統間的效能表現，可以比較不同廠牌的資料庫
3	在系統配置新硬體、修改演算法或是增加資料結構後，測試系統前後的相對績效表現
4	衡量系統的參數 (如快取記憶體大小) 對系統的績效影響
5	確認系統的瓶頸所在 (使用在內部測試)
6	設計某個可用來預測績效表現的模型

(四) 資料庫管理系統績效評估

績效評估在電腦界所測試的指標，可以指的是硬體的運作績效，也可以是某應用程式的效能。就本研究而言，績效評估所測試的是資料庫管理系統的績效，一個完整的績效評估，除了需有明確的達成目標之外，必需具備有下列幾個組成要素：

1. 自變數(Experimented Factors)

測試的產出結果，可予以改變的參數或系統組態，舉凡資料管理系統的大小、查詢的複雜度，系統的配置狀況等。

2. 績效指標(Performances Metrics)

測試的產出結果，以指標形式呈現以供比較。一般而言，績效指標可以是單位時間完成量(Throughput)、回應時間(Response Time)及價格/表現比(Price/Performance Ratio)等。

3. 實驗設計與衡量方式的選擇(Experimental Design & Measurement Techniques)

在這個階段，績效評估必須產生工作量(Workload)的模式。所謂的工作量的定義，即是在特定時段內，指定要完成的工作。每套績效評估必定擁有並詳細載明其專屬的工作量模式。工作量在其內容上可分為三種，分別是人造的工作量(Synthetic Workload)、實際的工作量(Empirical Workload)與混合工作量(Mixed Workload)。在工作量模式的建立上，又分為兩類情形，一是硬性規定輸入的工作量之細節，另一是使用者根據自己的需要決定輸入的工作量；在後者來說，為了符合完全揭露原則，結果報表上必須詳載其造出的工作量的一切細節。

4. 實驗的討論與分析(Analysis)

在衡量結束，數據出來之後，針對結果加以解釋。所有的績效評估報告，需遵守完全揭露原則，需詳細描述測試的環境，如軟硬體規格、系統的組態及所有自變數的值。

藉由標竿(Benchmark)，可評估資料庫管理系統的效能，目前已存在且為公認的標準有：第一個屬關連式查詢標竿的 Wisconsin Benchmark、以及屬複雜化

且深入的關連式查詢標準的 AS³AP Benchmark。另外就是 TPC 測試，將於下面二節逐一介紹之。

(五) Wisconsin 績效評估及 AS³AP 績效評估

1. Wisconsin 績效評估

Wisconsin 是早期第一個發展測試出來的資料庫管理系統績效評估方法 [14]，其主要的目標是在關聯式資料庫管理系統中，測量資料存取方式和查詢的效能。Wisconsin 績效評估以 32 個關連式查詢為主要的測量對象，其測量指標為測試每個查詢所花費的時間 [21]。雖然 Wisconsin 廣泛地應用於各領域中，然而其本身有設計上的限制存在，如單人系統、測試資料庫管理系統大小為 5 megabytes，所以在目前的應用上較不常見。

2. AS³AP 績效評估

AS³AP (ANSI SQL Standard Scalable and Portable) 績效評估方法主要提供一個全面、易處理的測試，以評量資料庫管理系統的處理效能，此外此方法具有可攜性 (Portability) 及延展性 (Scalability)，並提供 Equivalent Database Size 為資料處理衡量的指標，此指標主要用來衡量以最大資料庫的空間及 12 小時內的限制下，單人環境與多人環境可完成多少的測試。AS³AP 的測試可分成兩個模組 (Module)：單人測試與多人測試 [65]。

(1) 單人測試

主要包含載入、重整資料庫及基本查詢 (如 Joins, Projections, Aggregates, Bulk Updates 等) 測試。

(2) 多人測試

在多人測試的模組下，可區分成數個不同類型的資料庫管理系統工作量模式，主要有線上交易處理工作量 (OLTP) 資訊檢索工作量 (Information Retrieve) 及混合工作量模式 (Mix)。

(六) TPC 績效評估

目前國際上進行資料庫效能指標測試的機構並不多，較為有名的就是異動處

理效能評議委員會 (TPC, Transaction Processing Performance Council) 這個組織。它針對世界上佔有率較高之大型資料庫廠商,如 Sybase、IBM Db2、Oracle 及 MS SQL 等,於異質之平台上,根據其應用之範圍,分別採行 TPC-A、TPC-B、TPC-C、TPC-H、TPC-R 及 TPC-W 等測試。

在測試的同時,由於還必須考量工作平台及作業系統上的差異,所以測試的重點除了資料庫之效能外,還必須考慮之層面涉及系統平台之 I/O 測試及 CPU 測試。在這裡本文將介紹各種測試類別之詳細內容。

1. TPC-A 及 TPC-B 績效評估

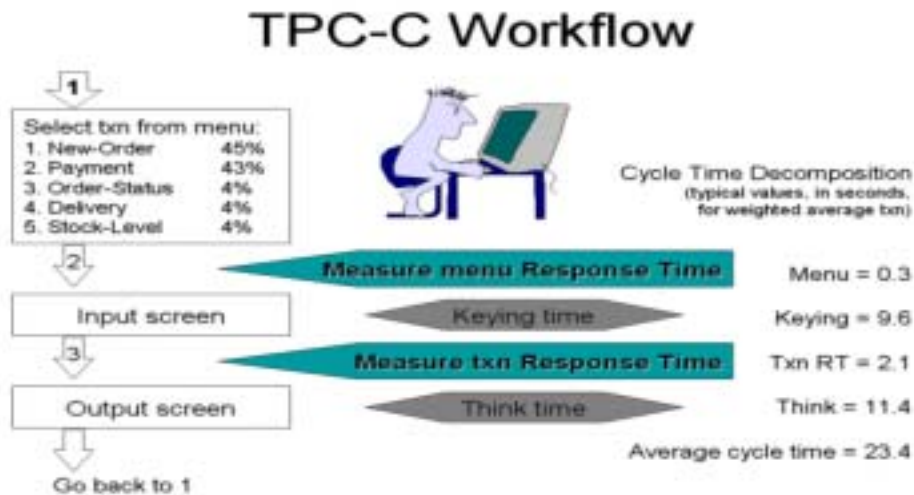
TPC 協會成立於 1988 年,由一群軟硬體廠商共同組成,此組織的目標主要是制定資料庫系統 (Database System) 及交易處理系統 (Transaction Processing) 的產業績效評估標準。TPC 協會於 1989 年發表 TPC-A 績效評估,TPC-A 是以線上交易處理系統 (OLTP, On-line Transaction Processing) 為主要的測試環境。OLTP 的環境特色為多終端機、資料輸入輸出、以交易為主等,其所測試的交易型態以單一、簡單及更新密集度高的交易為主,因此其工作量可反應出 OLTP 的交易環境,但卻無法對較複雜的交易型態進行評估。在多終端機的系統環境,TPC-A 績效衡量指標為每秒所完成的交易數 (tpsA, Transaction per Second)。之後 TPC 協會於 1990 年發表第二個績效評估 TPC-B,它並不是以 OLTP 主要的測試環境,我們可將其視為以資料庫壓力測試 (Database Stress test) 為主的績效評估,其衡量指標也是以每秒所完成的交易數 (tpsB) 為主 [68, 69]。

2. TPC-C 績效評估

TPC-C 模擬了一個完整的計算環境,這個環境中可以有大量的使用者在其上進行資料庫交易 (Database Transaction) [70]。指標測試 (Benchmark) 所測量的是交易形態的指令下達的環境。這些交易包括了登錄及傳送訂單、記錄費用、確認訂單的情形及監視資料倉儲中進貨的層級等等。當指標測試扮演的是批發提供者的活動時,TPC-C 並不限於某些個別的企業部門活動,更可表示任何

工業上的活動，例如管理、銷售販賣或配送產品或服務。

在圖中所顯示的是一個 TPC-C 測試的標準流程，整個測試的過程視為一週期性之循環，為了獲得測試的循環時間，我們可以將之解構成幾個流程模組。一開始測試者必須選用欲測試之 SQL 交易運算式，在下圖中，我們可以得知交易運算式共包含了五種類型之查詢語句，而每一類型之查詢語句又有其比重，即 New order 查詢佔了查詢總數的 45%、Payment 佔了 43%、Order-status 佔了 4%、Delivery 佔了 4%，及 Stock-level 佔了 4%。第二步測量的是 Menu 的回應時間，在進行測量回應時間之前還必須測量使用者輸入的時間，第三步則測量反應的時間，最後則是取得使用者思考的時間。整個流程即為 TPC-C 測試之流程。



圖十、TPC-C 測試工作流程圖

TPC-C 績效評估共包含了五個可以同時發生的交易混合模式，這五個交易可以是不同形態或不同複雜度，更可以是線上或延遲的佇列執行。它運用了系統元件的一瞬間來執行運作，而這些系統元件的環境可分類成以下幾項[44]：

- (1) 同時執行且跨越廣泛複雜度的多個交易形態
- (2) 線上及延遲的交易執行模式
- (3) 多個線上的終端機交談
- (4) 顯著的磁碟輸入和輸出
- (5) 交易完整性(ACID 特性)

(6) 經由 P.K 和 S.K 存取資料時所產生的變動分配(non-uniform distribution)

(7) 資料庫可包含許多的表格(tables) , 而表格可以具有不同的大小、屬性及關聯

(8) 允許資料存取和更新時的競爭

TPC-C 效能所測量的是每分鐘的新指令。主要的三個公制是交易速率 (tpmC)、價格和交易速率的比率 (\$/tpmC) 以及定價設定 (priced configuration) 的可用性三項。

3. TPC-D 績效評估

TPC-D 績效評估發展於 1995 年 , 其與上述前三個績效評估最大的不同點是以決策支援系統(DSS)為主要的測試環境, 因此其所測試的重點以複雜查詢(OLAP, Online Analytic Processing) 為主[71]。

4. TPC-H 績效評估

TPC-H 是一個決策支援的指標測試。他包括了一系列企業導向的及時 (ad-hoc) 查詢及同時的資料修改。資料庫中的查詢及資料則涉及了整個工業之間的活動。這項指標測試圖示了決策支援系統檢視大量資料的執行及具有高度複雜層級查詢的情形 , 並給予企業一個滿意的答覆[72]。

TPC-H 的效能公制稱為 TPC-H 混合的每小時查詢效能 (QphH@Size, TPC-H Composite Query-per-Hour performance Metric), 它可以反應系統進行查詢時的各種能力。這些能力包括了查詢或執行時選取資料庫的動作、經由單一資料流查詢處理的能力及在多個同步伺服器送出查詢時的單位時間完成量。TPC-H 的價格和效能公制可表示成 \$/QphH@Size。

5. TPC-R 績效評估

TPC-R 是類似 TPC-H 的決策支援指標測試 , 但是它允許額外先進查詢知識的最佳化。它包括了一系列的企業導向查詢及同步的資料修改[73]。

TPC-R 的效能公制稱為 TPC-R 混合的每小時查詢效能(QphR@Size, TPC-R

Composite Query-per-Hour performance Metric)，它可以反應系統進行查詢時的各種能力。這些能力包括了查詢或執行時選取資料庫的動作、經由單一資料流查詢處理的能力及在多個同步伺服器送出查詢時的單位時間完成量。TPC-R 的價格和效能公制可表示成 $\$/QphR@Size$ 。

6. TPC-W 績效評估

TPC-W 在設計時是模擬一個網路書籍零售商的所有活動。使用者可以在網站上瀏覽及下訂單。使用者設計成一個模擬者(Emulator)或遠端瀏覽模擬者(RBE, Remote Browser Emulator)，他們所處的環境也是在一般的 HTTP 網路環境下，所以可視為一個真實的瀏覽者。TPC-W 的詳細規格為 14 種不同的網頁形態。如果你想了解 TPC-W 績效測試的實際執行環境，可以想像你正在瀏覽一個電子商務網站，從事產品的瀏覽及交易。當你進入網站時，你所進入的第一個頁面是 Introduction 或 Home page，在這個頁面裡頭，它包含了公司的 Logo、促銷項目以及檢視選項，詳細項目可能包括了最佳銷售書籍、新書介紹、搜尋網頁、購物卡以及訂單狀況頁。

以下將介紹測試時的實際作業情況，RBE 將連結 SUT(System under Test) 並從 14 項不同的網頁亂數選取，進行瀏覽或交易動作，而 RBE 將遵從真實使用者的檢視行為，如先進行瀏覽，再進行交易。設計時，將有 80%的動作在存取 Home、New Products、Best Sellers 以及 Search Pages 等頁面，而其它 20%的動作則在存取 Shopping Cart、Order、Buy 及 Admin 頁面。在這 20%的動作之中，有 5%的動作在存取含有 SSL 加密的安全性網頁。

TPC-W 績效測試共有三種不同的公制，第一個是 WIPS(Web Interaction per Second)、再者是 WIPsb，著重於網頁瀏覽動作，最後是 WIPSo，則著重於使用者交易動作。WIPsb 和 WIPSo 的最大不同，在於瀏覽行為和交易行為的比重分配，WIPsb 中，有 95%的瀏覽行為和 5%的交易行為，而 WIPSo 中，瀏覽行為和交易行為則各佔 50%。

TPC-W 藉由修改瀏覽的速率來模擬三種不同的數據圖表：主要的採購動作

(WIPS, Primary shopping)、瀏覽動作(WIPSo)及網頁為基的訂單動作(WIPSo)。其主要的公制是 WIPS 速率、價格和 WIPS 速率比(\$/WIPS)以及定價設定(priced configuration) 的可用性三項。

TPC-W 是一個交易性網頁的指標測試。工作量產生於一個可控制的網際網路商業環境，這個環境模擬了企業導向交易伺服器的活動。它的工作量運用了系統元件的一瞬間來執行運作，而這些系統元件的環境可分類成以下幾項[44, 74]：

- (1) 多個線上交易的瀏覽交談。
- (2) 可經由資料庫存取及更新而產生動態的網頁
- (3) 一致性的網頁物件
- (4) 同時執行且跨越廣泛複雜度的多個交易形態
- (5) 線上交易執行模式
- (6) 交易完整性(ACID 特性)
- (7) 資料庫可包含許多的表格(tables)，而表格可以具有不同的大小、屬性及關聯
- (8) 允許資料存取和更新時的競爭

7. 各類型 TPC 效能指標測試比較

在介紹完 TPC 各項資料庫效能指標測試之後，我們得到一個匯總的綜合比較表，如表八。在比較表中，我們可以發現 TPC 從 1990 年發表 TPC-A 及 TPC-B 績效評估方法之後，又陸續發表 TPC-C、TPC-D、TPC-H 及 TPC-R 之後，最後於 1999 年發表 TPC-W，形成一跨多領域之多面向測試標準。

而這些資料庫效能測試指標包括了 OLTP、Data Stress、DSS(OLAP)、複雜 DSS、特殊 DSS 及電子商務網站等各類型之資料庫。形成了一系列相當完整的測試系統。

表七、TPC 指標測試綜合比較表

	發表年份	測試領域	測試指標公制
--	------	------	--------

TPC-A	1990	OLTP	tpsA
TPC-B	1990	Database Stress	tpsB
TPC-C	1992	Complex OLTP	tpmC
TPC-D	1995	DSS	QphD@Size
TPC-H	1998	Complex DSS	QphH@Size
TPC-R	1998	Specific DSS	QphR@Size
TPC-W	1999	Web Commerce	WIPS, WIPsb, WIPSo

(七) 資料庫績效評估方法之綜合性介紹

針對上述所討論的績效評估，在此作一整理，如表八所示。

表八、關聯式資料庫績效評估之介紹

標竿名稱	簡介
Winsconsin	第一個關連式資料庫績效評估
AS3AP (ANSI SQL Standard Scalable and Portable)	AS3AP 績效評估主要提供一個全面、易處理的測試，以評量資料庫的處理效能，此外具有可攜性 (Portability) 及延展性 (Scalability)，並提供 Equivalent Database Size 為績效衡量的指標
TPC Benchmark A(TPC A)	TPC-A 衡量的是專門處理線上交易處理、修改密集的資料庫環境，即一般的 OLTP 系統。
TPC Benchmark B(TPC B)	TPC-B 衡量指標與 TPC-A 相同。但不同的是 TPC-B 用不到任何的終端機、網路或 think-time。
TPC Benchmark C(TPC C)	TPC-C 衡量指標與 TPC-A 相同，可以適用於 OLTP 的應用環境；但 TPC-C 為多交易型態、複雜資料庫與執行環境為 TPC-A 的改良。
TPC Benchmark D(TPC D)	TPC-D 與上述績效評估最大的不同點在於其應用於決策支援系統的環境，其系統特色為資料量大、查詢較為複雜。
TPC Benchmark H(TPC H)	其也是以決策支援系統為主要的測試環境，其所

	測試的重點為複雜查詢及一致性資料更動測試。
TPC Benchmark R(TPC R)	TPC-R 所使用的資料庫與評估的環境都與上述的 TPC-H 相似，只不過兩者之間最大的不同處是其使用的目的：TPC-H 著重於複雜查詢的評估上；而 TPC-R 則在於企業日常的報表評估上。
TPC Benchmark W(TPC W)	此績效評估主要是應用在電子商務的資料庫績效評估上。

二、NIIS 中央資料庫之效能考量

為了支援日後資料倉儲建置及決策支援應用，必須考量 NIIS 系統中央資料庫建置時的效能是否足以支援。所以本文參考 TPC 網站所提供之 DBMS 效能指標測試，針對 TPC-C 之雙重指標，即 tpmC 及\$/tpmC 兩項指標說明。在評量資料庫採用方案時，必須有諸多考量，本文歸納如下：

(一) 延展性(Scalability)

由於企業是不斷變動的組織，所以會有成長上的考量，而資料庫之選用相對的必須考慮其能力是否具有足夠的擴充能力，以支援日後企業成長之需求及資料量成長的需求。

(二) 維護性(Maintenance)

資料庫有許多不同的類型，企業若本身不具備相關之能力，則必須考量其支援廠商是否有足夠能力進行資料庫之維護。

(三) 整合性(Integration)

企業若非新創之組織，必然具有歷史之資料，選用之資料庫必須能夠向下相容或提供快速之轉換機制，另外在前後端之整合，也必須考慮流程的順暢性及可行性。

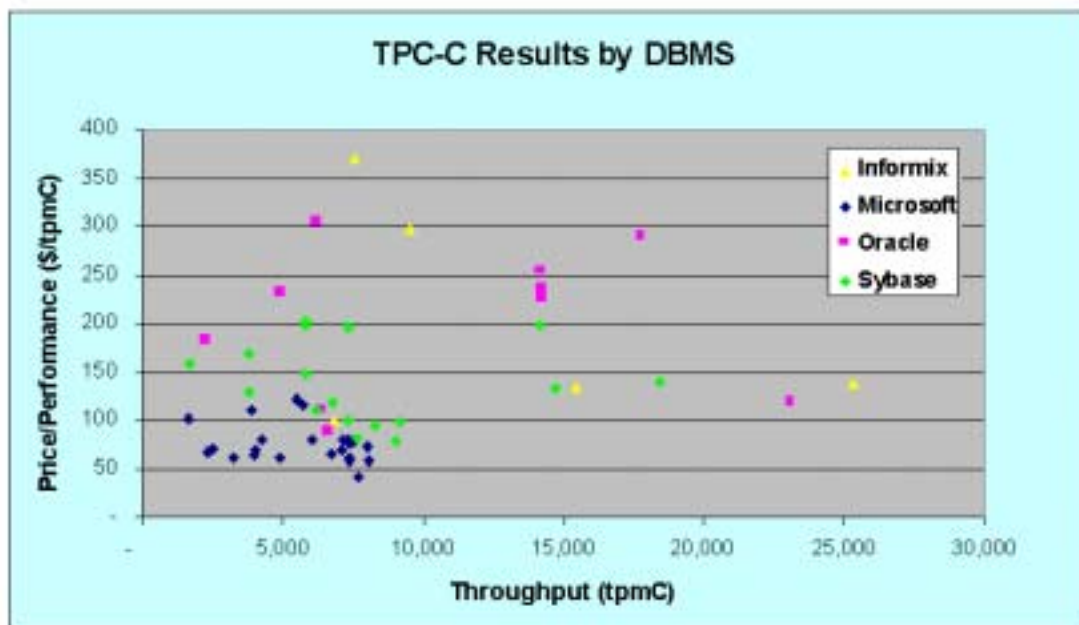
(四) 技術可行性

在建置資料庫時，必然有許多技術問題有待克服，在選用時，也必須考量是

否支援之廠商具有足夠之技術能力完成資料庫之相關建置問題。

(五) 投入成本(Invest Cost)

很重要的一點就是，除了選用良好之資料外，還必須考量投入成本是否有超出預算的問題，必須在二者之間取得一個平衡。



圖十一、資料庫管理系統之 TPC-C 效能指標測試

由圖十一可以得知[56]，Microsoft 提供 SQL Sever 在評估指標的表現中偏低，而較高階之 Sybase、Informix 及 Oracle 三類資料庫則具有良好之績效，所以 Microsoft SQL Server 顯然較適合用在不具成長性之中小型企業之上，而其它三者則較適合用在成長型中小企業及大型企業之中。

參、醫療決策支援系統

由於醫療本身的特殊性，使的資訊科技的應用也必須有所不同，必須更具備視覺化的設計，處理流程也更需要仔細考量，因此，醫療資訊系統所在意的並非只有資訊架構本身，所以在接下來的章節中，我們首先介紹現行醫療決策支援系統的基礎架構，之後在介紹資訊系統在醫學上應用的困難與醫療資訊系統的未來趨勢，藉由這個章節讓我們能更審慎評估仔細考量，免於隨波逐流、盲目跟從於追求資訊科技上。

一、醫療決策支援系統基礎架構

本文乃針對衛生醫療資訊系統在 Web-based 環境下的資訊架構，衍申發展其智慧型決策支援系統，望能提供予使用者更方便而快速的醫療決策資訊，並能提升醫療決策品質。

邱瑞科等[8]曾提出多層次架構符合近代分散式資訊及決策系統所需之參考架構，如圖十二所示。



圖十二、全國醫療資訊網整體架構

該參考架構中明確的分類，清楚地展現了一個層級關係：位於最底層的通訊管理包含了國家資訊基礎建設、全國醫療資訊網、衛生局所內部網路標準架構三個部份。位於第二層的系统可分為六類，分別是高階主管資訊系統、決策支援系統、醫學專家系統、公用性資訊系統、個別性資訊系統、衛生行政管理資訊系統。第三層的通用性連線服務基礎建設則可視為管理第一層及第二層等技術平台運用的角色。

所以，醫療資訊的網路化、電腦化，強調分散/整合/資訊共享，並進而提供醫護人員決策支援、病情的/病理的快速有效分析，已成為未來醫療資訊系統發

展的主要方向，各衛生醫療單位無不積極推動這些資訊系統的研究與發展。網路上以分散式、web-based 方式建立之系統強調利用網站伺服器提供使用者線上存取資料的服務，且利用資料格式的轉換及相關的傳輸機制做為地區與中央資料庫間線上資料的交換，如此可大幅減少資料傳輸的成本。配合瀏覽器作為單一性使用者介面，可以讓使用者不因應用程式邏輯的改變而需重新安裝應用軟體。分散式資料庫或伺服器的應用，更加強了網路傳輸的效能及功能架構的可擴充性，這些發展促使衛生醫療資訊及決策系統的改良並朝向以 web-based 在網路上利用分散多層式的架構發展。

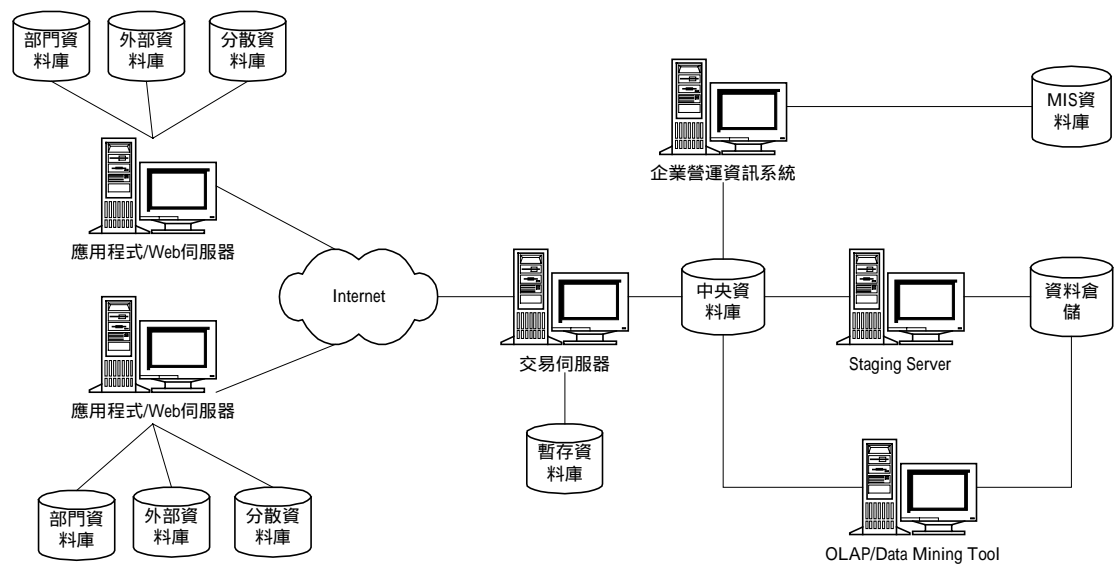
傳統 client/server 式的資訊系統架構因為擴充性不足而導致三層式架構的出現，如表九所示。此架構中強調在 client 端只需負責使用者介面之呈現，故稱為「thin client」[66]。隨著瀏覽器的普及化，以之作為 client 便成為了最佳的選擇。至於 AP Server 與 DB Server 之間的溝通，目前的技術有：CGI、Extensions to the Web Server (eg. ASP)、ODBC 及 CORBA 等，Client 與 AP Server 的溝通則可透過 Scripting Language (eg. JavaScript) [66]。所以醫療資訊的網路化、電腦化、共享 / 整合並進而提供醫護人員決策支援、病情 / 病理的快速有效分析，已成為未來醫療資訊系統發展的主要方向，各衛生醫療單位無不積極推動這些資訊系統的研究與發展。

表九、三層式架構

Clients	AP Servers	DB Servers
Windows 95、Windows 98、Windows Me、Windows 2000 professional	Windows 2000、IIS、HTTP、DCOM	Oracle、Unix、SQL Server、Windows 2000 Server、Linux

網路上以分散式、web-based 方式建立之系統強調利用網站伺服器提供使用者線上存取資料的服務，且利用資料格式的轉換及相關的傳輸機制做為分散與中央資料庫間線上資料的交換，如此可大幅減少資料傳輸的成本。配合瀏覽器作為

使用者介面，可以讓使用者不因應用程式邏輯的改變而需重新安裝應用軟體。分散式資料庫或伺服器的應用，更加強了網路傳輸的效能及功能架構的可擴充性，這些發展促使衛生醫療資訊及決策系統的改良並朝向以 web-based 網路化的架構發展。圖十三所示便為一個以分散式、web-based 方式建立之決策支援系統基礎資訊架構。



圖十三、醫療決策支援系統的基礎資訊架構

二、資訊科技在醫學上應用之困難

根據台北醫學院李有專教授的說法，電腦在醫學上之所以會產生應用上的困難，在於心裡、科技和人才三個方面。所謂的心理層面指的是醫療從業人員，尤其是醫師，對電腦常持有莫名之排斥感，這種排斥感，一半源自對電腦之不了解，一半則害怕自己的權威被電腦取代，使的電腦在醫學上的發展受到些許的阻礙，但所幸這項因素隨著電腦日趨普及，對於資訊科技逐漸有些認識，而有改善。

而所謂的科技因素是指由於過去電腦是昂貴、笨重、速度不快又非常不易操作使用的機器。除了少數行政工作 (Administration Overhead)過度龐大的醫院之外，少有人能供養得起一部電腦主機。導致對資訊科技上瞭解不易，使的資訊科技未能在醫院普及；更別說是設計出讓醫師能操作的電腦。然而由於電腦軟硬

體技術的快速進步，體積小、功能強且最重要的 - 容易使用的電腦已相繼問世，加上精心量身設計的軟硬體系統，使的醫師能使用相關的資訊科技，減少科技造成的不良因素。

最後是所謂的人才因素是指設計臨床資訊系統(Clinical Information System)時之困難常常在於醫師不了解資訊科學，而資訊人員又不了解醫學。這兩者皆是進入障礙頗大之學問。對非醫學領域之人而言，醫師通常是一種很難溝通的專業人員，溝通不良的結果，多半是一個徹底失敗的系統。這也是為甚麼多數醫院的資訊人員不敢輕易嘗試架設臨床資訊系統的原因，其實這方面的資訊技術早已存在。而解決的方法最好的方法莫過於培養醫學與資訊學兼修之人才，增加資訊與醫療互動的機會。

三、醫療電腦化之趨勢

近年來由於資訊產業快速發展，也逐漸應用在醫療行業中，大致而言，電腦在醫學應用可分四大部份：

(一)行政 (Administration) 部份

由於醫院的管理十分複雜，資訊科技最明顯的效益就是提高自動化及快速的計算能力，因此對電腦的需求也相對提高，在行政方面，資訊科技已經非常成功地應用於醫院及診所的行政管理上，包括掛號、批價、住院、出院、給藥等行政管理系統。

(二)改善醫療過程

包括利用所謂的 PICS 系統對於電腦斷層掃描或核磁共振影像影像處理系統，來輔助醫師作正確的診斷。其次在運用人工智慧 (AI) 或專家系統時，還可使用作為輔助診斷之用，利用資訊科技結合病患個人、電子化病歷及專業知識知識庫，進行系統性評估，即時 (Real-time) 提供醫護相關人員臨床診斷或醫療上的建議。可應用在用藥、預防保健、醫療診斷、特殊疾病、檢驗需求或結果解釋等方面。

(三)醫學教育

目前許多醫學相關資訊都已電腦化，最著名的是 Medline 醫學文獻資料庫，包括 1966 年以來絕大部份世界上的醫學文獻資料，打破了許多以往醫學資訊流通的限制，線上檢索 (On-line reference) 可以提供最新的資訊。

(四)遠距醫療及教學

如在醫師不易到達或醫師較少的偏遠地區，便可以利用遠距醫療的技術，醫師不必親臨現場，即可做種種的會診甚至執行醫療作業；此外，各種醫學新知亦可以上課或座談會的方式經由網際網路傳播

肆、整合型中央 NIIS 中央資料庫之建置

隨著地方版整合性衛生所資訊系統視窗版的推展與上線完成，衛生署疾病管制局計劃建立一套整合型中央版預防接種系統，藉由與地方版系統的相互運作、資訊交換及 IC 健保卡的推動，來建構一套完整性國家型預防接種系統，讓我國疾病管制-預防接種工作所需的資訊取得與運用能夠更有效、快速、正確。

此一資料倉儲的架構主要是利用分散式資料庫的特點所建立，強調各個地區資料庫都能保持其自主性(Local Autonomy)與平等的地位，而又能提供一致性、正確性高的資料作為參考的依據，因此整個資訊架構即是以分散式資料庫為基礎，配合代理伺服器提供各種應用。

一、資料庫建置相關研究

Roger 與 Dennis(1985)提出一個資料庫建置方法與模式來建構辦公室資訊系統(取名為 INSYDE)，此系統為一個具高互動性的系統，對於資料庫建置的程序與步驟配合資訊系統的應用頗為詳盡[60]。張明惠(1986)針對中央控制的資料庫管理系統研究進行建置與實作，透過星狀網路架構，設計一個分散式資料庫管理系統，由中央節點負責通訊的控制其它節點，而分散的節點具有完整的分散式資訊庫管理系統與辦公室自動化系統，在資料庫系統建置中考量了目錄管理、資料的分佈型態、協同控制與查詢工作處理，並且設計資料庫管理系統、遠地資料存取控制器、及協同控制器來進行中央與地方的資料交換[30]。林俊宏(1993)設計與實作一分散式異動伺服器，在分散式系統中，設計一協調者去整合分散在各節點的資料[28]。之後，倪正耀(1994)製作一分散式資料庫系統測試檯(取名為 DITSE)，提供結構性查詢語言讓使用者存取散佈於網路上不同節點的資料庫，DITSE 遵循分散式交易的兩階段委任協定，避免節點與節點間的溝通協調發生問題[29]。

多數相關文獻對於分散式資料庫系統建置著重資料交換協同控制及資料存取一

致性[59,71]，此外，以張明惠(1996)研究之中央控制的資料庫管理系統強調避免資料重覆配置以達成資料分享，與目前於各企業分散式資料庫建立整合性中央資料庫之架構不盡相同[30]。

二、分散式資料庫相關研究

一、分散式資料庫概論

Connolly 與 Begg(1999)定義了分散式資料庫為一群在電腦網路上具有資料共享、實體分散、邏輯交互關係的資料庫[43]；Tamer 與 Valduriez(1999)定義分散式資料庫為在電腦網路上集結多個且邏輯上相互運作的資料庫[67]；而曾守正等(1996)認為分散式資料庫是一個資料的集合體，這些資料在邏輯上可以看成一個單一的資料庫，而在實體上則是分散儲存在不同節點上，使用者可以透過網路查詢與更新維護，在每一個節點上的資料庫都有其本身區域的資料庫管理系統所控制，所以每個資料庫都有其個性的自治性，但需配合分散式資料庫管理系統來維護整個分散式資料庫的資料一致性[10]。此外，分散式資料庫管理系統則要提供使用者查詢的分解、位置的透通性、查詢修改位置的透通性、查詢語句執行的最佳化、分散式環境故障的恢復、交易的管理、以及資料安全性等功能來維護整個分散式資料庫系統的正常運作。綜合上述而言，分散式資料庫具有兩個特性[12]：(一) 資料庫中的資料屬於某一邏輯相關個體，此項特性表示這些資料彼此間具有邏輯關係存在；(二) 資料不是存放在同一地點，而是分散在各個不同的地點。

一般而言，一個完善的分散式資料庫系統必須符合十二個規則，此十二個規則即是分散式資料庫系統的目標，分述如下[5,50]：

- (一) 本地自治性(Local Autonomy)：分散式系統的每一節點必須有充分的自治權，本地性自治權是指某一節點的所有動作皆由該節點自行控制。而資料的安全性、完整性及儲存表示法仍由該節點控制。
- (二) 所有的資料站都具有平等的地位(No Reliance on a Central Site)：由

於本地自治性，也就是所有的節點皆一視同仁，並不需要依靠所謂的 ” 集中站 ” 來處理一些集中的服務，如集中式查詢處理或集中式異動管理。

- (三) 持續性操作(Continue Operation)：在新加入一個資料站時，並不需將系統暫時中止或關機，即使對系統做某些變更，如更新資料庫，系統也能繼續運作。
- (四) 位置無關性(Location Independence)：位置無關性的主要觀念為讓使用者不需要知道他們所處理的資料實際上存於何處，而在感覺上，該資料似乎是儲存在其所在的節點一樣。
- (五) 分段無關性(Fragmentation Independence)：基於執行效率的考量，有時我們需對資料做資料分段，使資料可被儲存最常被用到的位置，也使得大部份的動作像是區域性，且可減少網路上的交通流量。
- (六) 複製無關性(Replication Independence)：為了增加系統的可靠性及效率，必須將某些節點的資料拷貝到其他節點，然而還必須讓使用者察覺不到資料是複製的。
- (七) 分散式查詢處理(Distributed Query Processing)：由於資料是分散，所以系統必須提供分散式查詢處理的功能，使查詢可以正確執行在分散式資料庫系統中。
- (八) 分散式交易管理(Distributed Transaction Management)：分散式系統應該是提供分散式的並行控制(Concurrency Control)及回復控制(Recovery Control)，使得分散式交易處理能確保資料之一致性。
- (九) 硬體無關性(Hardware Independence)：各節點硬體設備的差異，必須與整個分散式資料庫沒有影響。
- (十) 作業系統無關性(Operating System Independence)：各節點作業系統的差異，必須與整個分散式資料庫沒有影響。
- (十一) 網路無關性(Network Independence)：各資料庫間通訊網路的差異，必須與整個分散式資料庫沒有影響。

(十二) 資料庫管理系統無關性(DBMS Independence)：分散式資料庫裡各資料庫管理系統可以是異質的,也就是可以不同版本的資料庫管理系統所組成。

二、分散式資料庫架構的優點

多數文獻對於分散式資料庫系統優缺點皆有著墨,針對分散式資料庫系統應用於本系統建置相關之優點分述整理如下[12,50]：

(一) 資料庫資料的分散式本質

資料可以依資料的性質與資料的種類而分散到不同的地方,這樣可以減少資料的傳輸或處理。

(二) 增加可靠性與有效性

在分散式資料庫系統中,存在有多個於網路節點上之資料庫,因此,即使其中一個資料庫發生了故障,整個系統也不致於不能運作,所以分散式資料庫系統的可靠性和有效性將比集中式資料庫高,因為集中式資料庫發生了故障,則整個系統則不能運作。而在資料分散且集中的環境,除了對於資料之可靠性與有效性有確實之保障外,另外也具有資料異地備援之特性。

(三) 允許資料共享並保有區域控制性

在分散式資料庫系統中,每個節點的資料庫可供其它節點存取使用,而達到資料共享,而同時每一個節點可以控制及設定其本地資料在分散式資料庫系統中別的節點的使用方式。

(四) 改善效能

由於資料庫分散在多個節點上,而每個節點都有其本地的資料庫管理系統,所以整體存取資料的負荷就會分散,此外,若有一個交易是涉及多個節點時,每一個節點可以平行處理,增加其處理速度,減少反應的時間。

(五) 系統易擴充

由於分散式資料庫系統中,資料庫的加入與刪除,只需要修改資料庫整體目錄,無關於其他資料庫的改變,所以在分散式資料庫系統中,系統是很容易來擴

充的。

三、分散式交易管理(Distributed Transaction Management)

在分散式資料庫系統中，因必須要作分散式交易平行處理，使得分散式交易管理在分散式資料庫系統中變成極重要的一個議題。交易是指一個邏輯單位的工作，此邏輯單位的工作可能由一連串執行操作所組成。一般而言，交易管理必須達到下列四個重要性質，簡稱 ACID 性質[1,5,50]：

- (一) 單一性(Atomicity)：一個交易是單一的運作(Operation)，也就是一個交易只分為有交易成功或交易失敗。
- (二) 一致性(Consistency)：每個交易都要從一個一致狀態轉移到另一個一致狀態，但中間過程並不需要保持一致性。
- (三) 隔離性(Isolation)：每一交易都是獨立，也就是一個交易不能影響另一個交易狀態。
- (四) 持續性(Durability)：一個交易完成，表示它的結果是永久的，如果想要改變它的結果，則需要另一個交易來改變它。

四、異動記錄(Transaction Log)

一般而言，異動記錄中通常會包含：恢復(Undo)與重做(Redo)所需的相關資訊，以便於資料庫管理系統可以在系統再次被啟動時，利用異動記錄來判斷那些異動必須要回復到原點，那些異動的結果必須要重做且將結果反應到資料庫中。而恢復與重做的運算方式則需有執行數次與執行一次的結果相同的特性。一筆異動記錄的內容通常會含有下列幾項資料[1]：

- 1、異動編號(Transaction id)：系統會給予每一個異動一個編號以示區別。
- 2、該異動記錄的編號(Transaction Log Record id)：系統會給予每一個異動記錄一個編號以示區別。
- 3、異動形式(Action Type)：包括委任、撤回、新增、刪除、更新等動作。
- 4、異動前的資料值：供恢復(Undo)時使用。
- 5、異動後的資料值：供恢復(Redo)時使用。

6、其它相關的輔助資訊。

以下幾個時間點會記錄上述的相關資訊到異動記錄中：

1、當一個具有編號為 `transaction_id` 異動開始執行時，在異動記錄中會記下一筆

```
begin transaction transaction_id
```

2、若該異動正常完成，且下達委任(`commit`)命令時，會在異動記錄中記下一筆

```
commit transaction_id
```

3、若該異動不正常完成，且下達撤回(`Abort` 或 `Rollback`)命令時，會在異動記錄中記下一筆

```
abort transaction_id (或 Rollback transaction_id)
```

為了防止以下兩個連續動作之間發生了系統錯誤，導致系統於恢復正常運作時無法正確地作判斷：1、寫入異動記錄的動作；2、將異動結果寫入資料庫的動作。

資料庫管理系統通常會採用所謂的「異動記錄優先記錄協定」(Log-write-ahead Protocol)，也就是先將異動記錄優先記錄完成後，再將異動記錄寫入資料庫中。

此外，異動記錄會一直持續地增長，所以最後會使磁碟空間達到飽和狀態，而使系統無法正常運作下去，所以，備份完畢後，也要定期去刪除異動記錄。通常，一個好的資料庫管理系統會定期提供使用者定義磁碟空間使用門檻(Threshold)的監督，一旦磁碟空間的使用超過門檻值的時候(例如，超過磁碟空間的 80%)，便依照使用者先前的定義去執行相關的動作，例如製作備份並將異動記錄刪除。

三、資料庫設計方法

資料庫設計的基本目的在於合理的時間內讓使用者獲取正確的且可用的資料以滿足其需求。在資料庫結構中，資料庫設計的方法來自於使用者所提供的需求[68]。一般而言，資料庫設計方法是一個動態的過程，必須適時地調整，而沒有最佳的設計方法，一個實用的資料庫設計方法必須根據所公認的原則而非取於僅可接受的狹礙觀點。Connolly 與 Begg 認為資料庫設計方法是使用程序 技術、

工具及文件等結構性的方式，以支援設計的過程[43]。而 Novak 與 Fry(1976)對於資料庫設計方法定義了四個基本的要件[56]：

- 一、含一系列結構化的設計步驟；
- 二、每一步驟的設計上需選擇一評估標準；
- 三、每一步驟的設計上需考量資訊的需求；
- 四、每一步驟的設計上需清楚描述資訊的輸出與輸入。

Teorey 與 Fry(1980)提出實用且適合於多數資料庫設計開發的逐步過程，分別為需求描述與分析、邏輯設計及實體設計與檔案設計[64]；Navathe(1992)認為資料庫模式設計過程需配合資料與應用程式的需求，並依循一結構化的程序與步驟來執行，包含概念設計、邏輯設計及實體設計[55]；此外，Lyon(1976)認為一個有意義的資料庫整體設計需兼顧彈性與整合性[54]；和 Roger 與 Dennis(1985)提出的資料模式規格及物件化階層程序的完整流程相符合[60]。因此，本研究將資料庫設計方法歸納並整理如下(表十)：

表十、資料庫設計方法
(資料來源：本研究整理)

資料庫設計方法	
程序	進行方式
一	需求搜集與分析
二	概念資料模式設計
三	邏輯資料模式設計
四	實體資料模式設計

一、需求搜集與分析

需求描述與分析包含了組織目標的建立、資料庫特定需求的起源等，透過與決策者、管理者、或主要相關業務人員進行訪視，並搜集相關資訊與需求，在了解組織流程及業務程序後，將所得資料轉換成為流程圖，以助於之後的資料庫模式建立。

二、概念資料模式設計

透過概念資料模式來表達使用者需求，將需求規格建構成實體關係模式。

二、邏輯關聯式資料庫設計

邏輯關聯式資料庫設計是建構一個以企業使用資訊為基礎的資料模式的過程，但並不包含特定的資料庫管理系統與實體設計的考量。其中應從使用者觀點建立邏輯概念資料模式、並且反覆與使用者檢視邏輯概念資料模式、而後建立並確認區域性資料模式與全域性資料模式(表十一)。

表十一、邏輯關聯式資料庫設計
(資料來源：本研究整理)

邏輯關聯式資料庫設計	
程序	進行方式
一	從使用者觀點建立邏輯概念資料模式
二	反覆與使用者檢視邏輯概念資料模式
三	建立並確認區域性資料模式
四	建立並確認全域性資料模式

三、實體關聯式資料庫設計

實體關聯式資料庫設計包含許多階段，從檔案設計到整體資料庫的結構設計都需考量。Teorey 與 Fry(1980)認為應先後以存取路徑的設計、叢集分析與設計、容量大小及完整性與安全限制為設計方式。其中亦依循了實體關聯式資料庫設計方式，將邏輯關聯式資料庫設計所產生的全域性資料模式轉換全域性邏輯資料模式至資料庫管理系統、設計與實行資料庫實體呈現、安全機制的設計與實行，並監視與調整整體系統的效能(表十二)[64]。

表十二、實體關聯式資料庫設計
(資料來源：本研究整理)

實體關聯式資料庫設計	
程序	進行方式
一	轉換全域性邏輯資料模式至資料庫管理系統
二	設計與實行資料庫實體呈現
三	設計與實行安全機制
四	監視與調整系統

此外，在分散式邏輯資料庫設計方面，亦應考量下列幾個問題[67]：

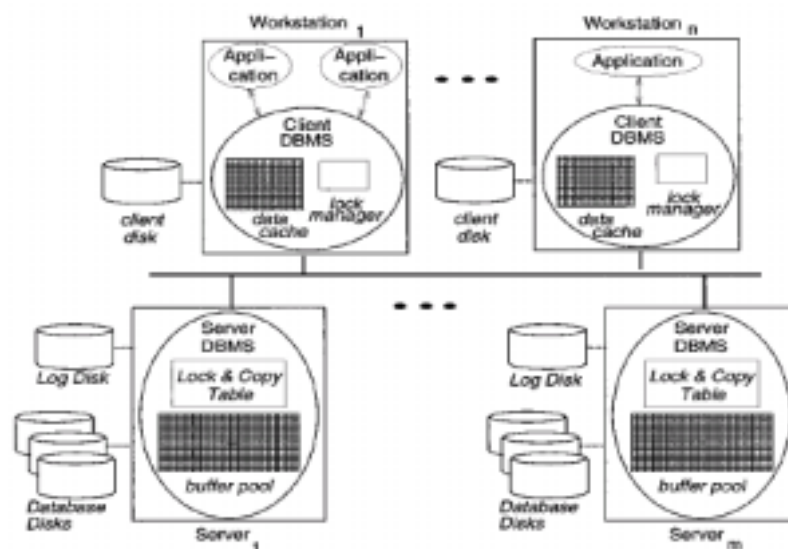
- 1、將資料分散到各個節點，讓各節點能夠充分平行處理資料的擷取與更新。

2、系統方面需考量整體查詢轉換到區域節點的子查詢，同時亦需考量各節點執行完子查詢後所傳回的結果必須整合成一個完整的整體結果。

然而，目前分散式資料庫系統尚無法滿足所有實用的需要，其中最主要且實用的問題在於在資料分配與設計[9]，若各地只存在其本身的資料，那麼在搜尋資料時可能會相當耗時；若利用完全複製的方式，則在資料更新時亦會耗費許多時間，因此，在設計分散式資料庫時，需評估並權衡其間的取捨，是一值得研究的問題。

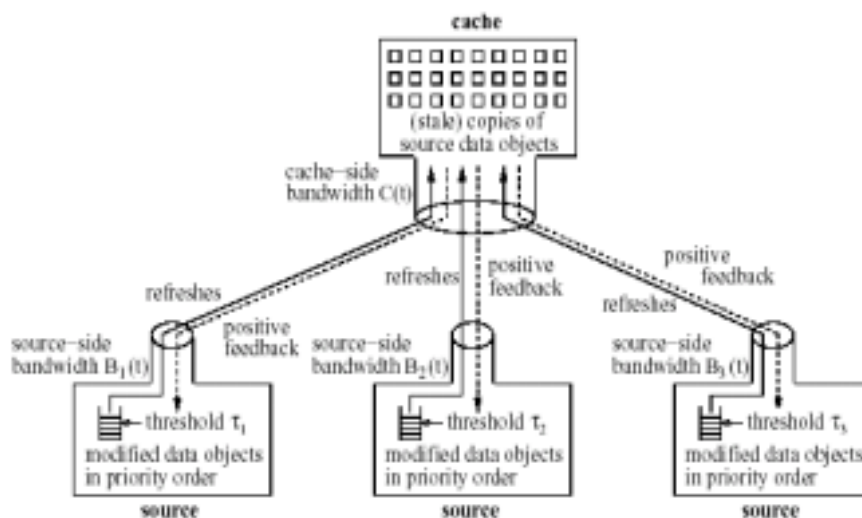
四、資料庫更新機制

現有相關文獻中強調資料更新的重要性[43,75,76,77]，但卻僅有少數研究針對此問題提出詳盡的解決方式，Adiba 與 Lindsay(1980)提出一個透過快照(snapshot)的機制來更新資料庫的內容，透過快照的更新來反應現有資料庫的狀態，對於複製資料的分散式資料庫尤具效益[38]。而 Lindsay(1986)以快照為基礎，提出一差異性更新演算法來減少更新過程中訊息所帶來的附加成本[53]。Franklin et al.(1997)提出分散式資料庫之資料一致性維護演算法(圖十四)，從主從式架構的觀點切入，各節點進行資料複製以改良資料可用性與系統整體效能，然而，由於分散的資料庫必須透過緩衝區與鎖定管理來控制資料的交換與存取，以確保各節點資料的完整性與一致性[57]。



圖十四：資料更新架構圖[26]

Olston 與 Widom(2002)研究作用於分散式系統間資料更新機制之同步化排程最適化(圖十五)，透過門檻值(threshold)的設定，解決節點間之協同控制易發生衝突及錯誤的問題，分別以優先更新次序、跨節點協同更新、緩衝區與儲存體之更新等考量進行架構建置[57]。



圖十五：資料最適化同步架構圖[57]

整合型預防接種資料庫之資料架構為分散且集中的環境，其原因為可由下列因素探討[7,48]：

(一) 資料庫的管理

在集中式的資料庫系統中，中央資料庫管理人員可以完全掌控管理資料庫。但是在分散式的資料庫系統中，各衛生局地方資料庫均有其區域資料管理人員(Local DBA)，同時在整體的環境下，可能亦有整體資料庫管理人員(Global DBA)。因此，整合型預防接種中央資料庫具有中央與地方的資料庫管理人員。

(二) 資料的獨立性

資料的獨立性在集中式資料庫的目的是要讓應用程式不受實際儲存結構的改變而影響。但在分散式資料庫的環境中，除了要達到前述的目的外，還要達到分散資料的通透性(Transparency)。各地方衛生局將其轄下各衛生所資料統一管理，各地方衛生所透過網路存取該地區預防接種資料，具有集中式資料庫管理的

特性，各地方衛生局將預防接種資料複製並存放於中央資料庫，使得中央與地方同時具有預防接種作業性資料，具有集中且分散之資料的管理機制。

（三）重覆資料的儲存

在集中式資料庫系統中，通常是儘量降低資料的重覆性，以方便維持一致性並節省儲存空間。但是在分散式的環境中，則反而要重覆將資料儲存在不同的資料庫中，以提高資料的存取速度，並增進可靠度與可用性。在整合型中央版預防接種系統中，由於依組織分權型態及管理上的需要，將預防接種資料重覆存放於中央與地方資料庫，以層層的組織架構而言，此重覆資料的儲存兼具集中與分散的特色。

（四）內部的儲存結構與快速存取

在集中式資料庫中，具有內部快速儲存結構的資訊，可以增進存取的速度。但在分散式的環境中，則很難建立與維持一個或多個資料庫的快速存取結構。在整合型預防接種中央資料庫中，除了資料異動與更新的存取效能尚需考量外，內部的快速存取可依資料結構的處理特性來評估。以各衛生局與衛生所間的作業性交易資料而言，一方面受到新一代全國性預防接種地方資料庫內部資料儲存結構的影響，另一方面則受到現有網路環境因素的影響。

（五）整合性、回復性與並行控制

在分散式資料庫系統中進行資料的處理，必須確保異動處理的特性及完整性，它的困難度與複雜度都比集中式的環境來得高。在預防接種系統中，由於各地區民眾具有身份唯一性，且該民眾不可能於同一日短時間內到不同的兩地接受疫苗注射，所以在資料的異動處理上較不具威脅性。此外，整合型預防接種中央資料庫除了便利中央處理整合性作業外，亦具有備援地方資料庫之功用，可避免地方因例外狀況而造成系統的永久損毀。

五、整合型中央版預防接種資料庫之系統發展與管理

總而言之，整合型中央版預防接種系統同時具有集中與分散之資料架構特

性，為一既集中且分散之系統。在此，我們亦需進一步考量此系統發展與管理的問題。

(一) 資料配置之集中方案評估

整合型預防接種中央資料庫藉由各衛生局地方資料庫所提供的區域性預防接種資料，有助於中央處理日常所需之業務，但是中央資料庫中所存放的資料內容為何？值得我們進一步評估。就分散式資料架構之資料配置(Data Allocation)策略方式，一般而言，可以有集中化、分割、完全複製、選擇複製等四種主要方式，以下就個別進行分析：

(一) 集中化(Centralized)

集中化的策略由單一資料庫及資料庫管理系統所組成，且置放於同一地點讓分散於各地的使用者透過網路存取。集中化的資料配置使分散於各地的使用者需要花費許多通訊成本來使用資料，一旦該中央資料庫發生錯誤，可能導致所有資料的損毀。

(二) 分割，或稱切割(Partitioned or Fragmented)

此策略將資料庫分割成不同的片段，而每個片段分別存放於不同的地點。不同的片段是依據不同地點對於該資料的高度使用性而分割。若片段依各地點需求適當地分割，其整體的效能將會較集中化的資料配置來得高。

(三) 完全複製(Complete Replication)

此策略是將所有的資料於每一地點皆存放一份，如此一來，無論是對於本地的使用方便性、可信賴性與效能皆是最佳化，但是在儲存及通訊成本卻相當地昂貴。

(四) 選擇複製(Selective Replication)

選擇複製的方式結合了集中化、分割、完全複製的特性。將資料依需求，選擇性切割或複製存放在不同的地點。這樣的方式因為較具彈性，所以常被一般所使用。

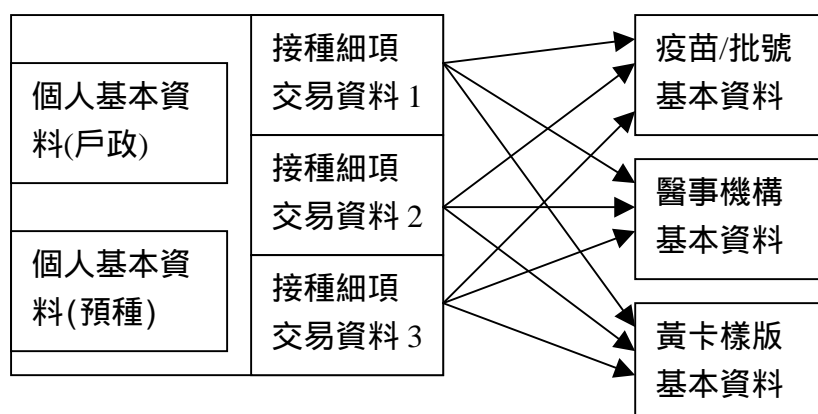
表十三、不同資料配置策略的比較[22]

	本地使用性	可信與有效性	效能	儲存成本	通訊成本	管理容易度
集中化	最低	最低	低	最低	最高	最易
分割	高	資料項目(低) 系統整體(高)	滿意度 高	最低	低	最不易
完全複製	最高	最高	讀取 (最佳)	最高	資料更新(高) 讀取(低)	不易
選擇複製	高	資料項目(低) 系統整體(高)	滿意度 高	普通	低	最不易

依不同的資料配置策略比較(表十三)[57]可看出,新一代全國性地方版預防接種系統於各衛生局地方資料庫為集中化之資料配置已成既定事實,於各衛生所通訊上需透過中華電信或專線等網路環境,進行資料之交換處理,故效能與通訊成本上較不具優勢。而整合型預防接種中央資料庫基於集中且分散之架構為考量,應採用地方分割再選擇複製於中央資料庫的折衷方式。換言之,各地方衛生局並不需要將所有資料完全複製到中央資料庫,而僅僅需要依中央的需求與政策,將資料選擇複製到中央資料庫,如此一來,不僅使地方與中央的資料可用性提高、系統整體的可靠與有效性提高,且達到效能、儲存成本與通訊成本的最適化考量。

(二) 資料有效性集中之策略

由於各地方衛生局採用選擇性複製資料至中央資料庫,因此,我們需考量現有新一代整合型預防接種系統之資料結構圖十六、資料庫 Schema 與衛生局上傳至中央的報表種類與內容[10]。各項目交互比對後,篩選出整合型中央版預防接種系統所需之適用資料,以作為第一次建置中央資料庫設計之最適化參考依據。



圖十六、預防接種之資料架構

該預防接種資料結構呈現出將舊有預防接種黃單之資訊化處理。我們可看出黃單主要由三部分構成，分別是個人戶政基本資料、個人預種基本資料及各項接種細項交易資料。其所參照的為疫苗/批號、醫事機構及黃卡樣版基本資料。而上傳的報表可分為工作量報表、完成率報表、衛生所疫苗庫存報表、學齡前國小新生類報表等四類。且根據整合型地方版預防接種資料庫 Schema(表十四)[10]及其資料關係類別圖，整理出整合型中央版預防接種資料庫所應具備之資料庫 Schema，因中央與地方需求的不同，中央所具有之資料庫 Schema 與各衛生局資料庫 Schema 亦具有部份差異，如表十五 [10]。

表十四、整合型地方版預防接種系統資料庫 Schema 一覽表

整合型地方版預防接種系統資料庫 Schema 一覽表			
英文名稱	中文名稱	英文名稱	中文名稱
Employee	預種工作人員	BBAR1	衛生局疫苗撥入單頭
PHConfig	衛生所組態資料	BBAR2	衛生局疫苗撥入單身
BAD2	疫苗批號	BBAE1	衛生局疫苗撥出單頭
BAD1	疫苗基本資料	BBAE2	衛生局疫苗撥出單身
BAI1	預種地點	BBA01	衛生局疫苗損毀單頭
BAI2	縣市鄉鎮代碼	BBA02	衛生局疫苗損毀單身
BAG1	監視片指數	BBAM1	衛生局疫苗退貨單頭

BAG2	高危險群註記	BBAM2	衛生局疫苗退貨單身
BAB3	老人流感接種記錄	BAN1	領用單頭
BAH3	老人流感健康評估	BAN2	領用單身
BAA1	嬰幼兒基本資料	BAM1	退貨單頭
BAB2	預防接種黃卡細項	BAM2	退貨單身
BAC1	下次預種日期	BAP1	耗損單頭
BAQ1	黃卡疫苗劑別	BAP2	耗損單身
BAD3	多合一疫苗	BAP3	上月結存量紀錄單頭
BAJ1	預種日期設定	BAP4	上月結存量紀錄單身
BAH1	禁忌檔健康評估	BA01	損毀單頭
BAH2	副作用調查	BA02	損毀單身
BAL1	假日延後天數	BAE1	撥出單頭
BCK1	孕婦 B 肝產前檢查登錄表	BAE2	撥出單身
BAF1	合約醫院疫苗領用單頭	BAR1	撥入單頭
BAF11	合約醫院疫苗領用單身	BAR2	撥入單身
BAF2	合約醫院疫苗退貨單頭		
BAF22	合約醫院疫苗退貨單身		
BAF3	合約醫院疫苗結存量回報單頭		
BAF33	合約醫院疫苗結存量回報單身		

表十五、整合型中央版預防接種系統資料庫 Schema 一覽表

整合型中央版預防接種系統資料庫 Schema 一覽表			
英文名稱	中文名稱	英文名稱	中文名稱
PHConfig	衛生所組態資料	BBAR1	衛生局疫苗撥入單頭
PBConfig	衛生局組態資料	BBAR2	衛生局疫苗撥入單身
BAA1	嬰幼兒基本資料	BBAE1	衛生局疫苗撥出單頭
BAD1	疫苗基本資料	BBAE2	衛生局疫苗撥出單身

BAB2	預防接種黃卡細項	BBA01	衛生局疫苗損毀單頭
BAD3	多合一疫苗	BBA02	衛生局疫苗損毀單身
BAI2	縣市鄉鎮代碼	BBAM1	衛生局疫苗退貨單頭
BAJ1	預種日期設定	BBAM2	衛生局疫苗退貨單身
BAL1	假日延後天數	BCAU1	中央疫苗領用單頭
BAQ1	黃卡疫苗劑別	BCAU2	中央疫苗領用單身
BCAD2	疫苗批號	BCAV1	中央疫苗退貨單頭
BCAB3	老人流感接種記錄	BCAV2	中央疫苗退貨單身
BCAG1	監視片指數	BCAW1	中央疫苗耗損單頭
BCAH2	副作用調查	BCAW2	中央疫苗耗損單身
BCAI1	預種地點	BCAW3	中央疫苗上月結存量單頭
BCCK1	孕婦 B 肝產前檢查紀錄	BCAW4	中央疫苗上月結存量單身
BCAF1	合約醫院疫苗領用	BCAX1	中央疫苗損毀單頭
BCAF2	合約醫院疫苗退貨	BCAX2	中央疫苗損毀單身
BCAF3	合約醫院疫苗結存量回報	BCAS1	中央疫苗撥出單頭
		BCAS2	中央疫苗撥出單身
		BCAT1	中央疫苗撥入單頭
		BCAT2	中央疫苗撥入單身

(三) 整合型中央與地方資料庫設計之差異性比較

(一) 預防接種中央資料庫設計目標

1. 集全國性衛生局、衛生所及合約醫院之統計資料為主。
2. 統一制定全國性預防接種相關業務之標準(如疫苗基本資料設定)。
3. 統一制定全國性預防接種資料傳輸之標準(如傳輸時間、參數設定)。
4. 統一制定全國性預防接種資料備份之機制(如資料備份、清除、回存等設定)。
5. 提供戶政資料擷取之作業，有助於各衛生局所線上作業之便利。
6. 透過中央交易伺服器取得更新異動資料，保持中央與地方之一致性。
7. 彙整之資料應具有全國性各單位及整合性統計報表列印功能。

8. 彙整之資料應具有全國性各單位及整合性疫苗管理現況資訊。
9. 提供有助於中央高階主管對於疫苗備用量、建議採購量之決策功能。

(二) 預防接種地方資料庫設計目標

1. 彙集地方性衛生所及合約醫院之統計資料為主。
2. 依循中央所制定預防接種之統一標準。
3. 透過轉介歸戶機制，節省人力、物力與時間之消耗。
4. 確保其轄下各衛生所日常預防接種交易之資料完整性。
5. 彙整之資料應具有地方性各單位及整合性統計報表列印功能。
6. 彙整之資料應具有地方性各單位及整合性疫苗管理現況資訊。
7. 提供地方各項有助於中央管理之資訊上傳回報功能。
8. 提供有助於地方單位高階主管對於疫苗備用量、建議採購量之決策功能。

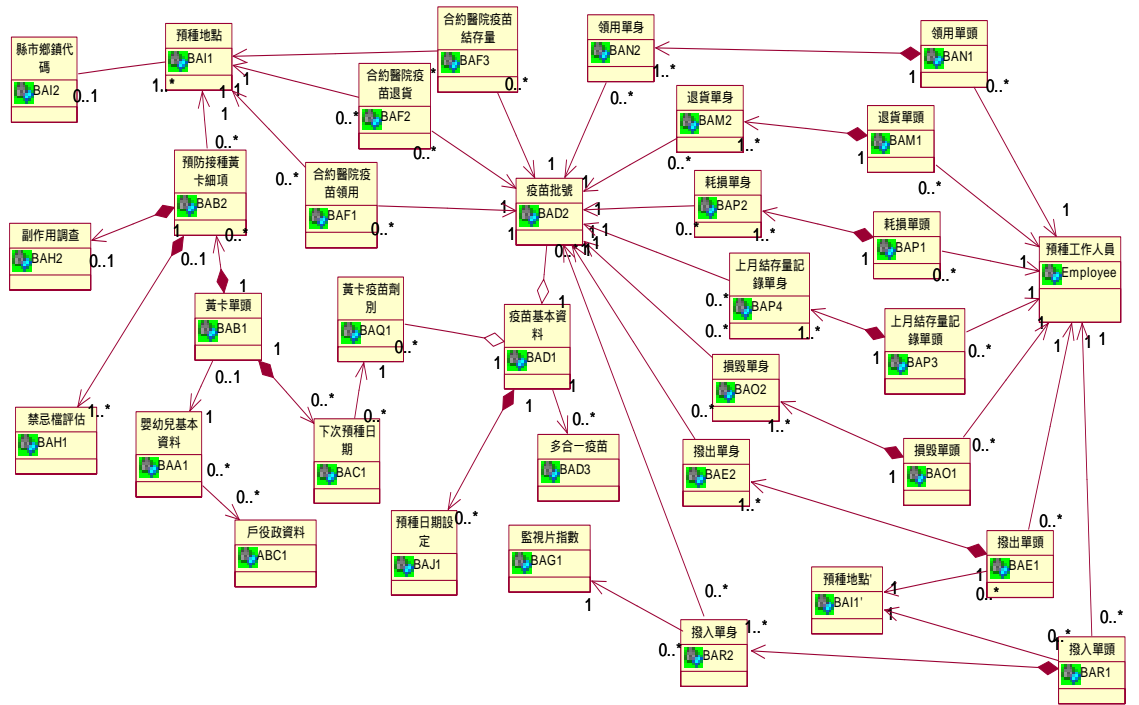
(四) 中央疾病管制局與地方衛生局所之資料交換與傳輸策略

衛生所醫療資訊使用人員透過預防接種線上作業存取衛生局資料庫，藉由預防接種伺服器元件，將該日具有在本戶籍地或使用轉介、歸戶機制的接受預防接種者記錄於資料異動日誌檔中。每日衛生局交易程式伺服器與中央交易處理元件進行訊息交換的確認後，開始進行最新資料的異動紀錄傳輸，待傳輸完成並確認無誤後，將此資料的異動紀錄載入中央資料庫。此一作法，將本地及外地之預防接種異動紀錄於中央資料庫進行更新存取的動作，可以使中央資料庫內的預防接種紀錄與各衛生局地方資料庫保持一致性。

此外，各地方衛生局所之疫苗管理亦需納入考量，醫療資訊處理人員透過疫苗管理子功能將各類疫苗之領用、撥入、撥出、損毀、結存量等狀況登錄，透過資料的傳輸與轉換，中央交易處理元件統計分析各地方疫苗使用情形，將有助於節省人工回報疫苗現況之人力、物力與時間，進一步讓中央即時獲得地方相關預防接種資訊，有助於中央對於地方異常狀況之處理與決策。

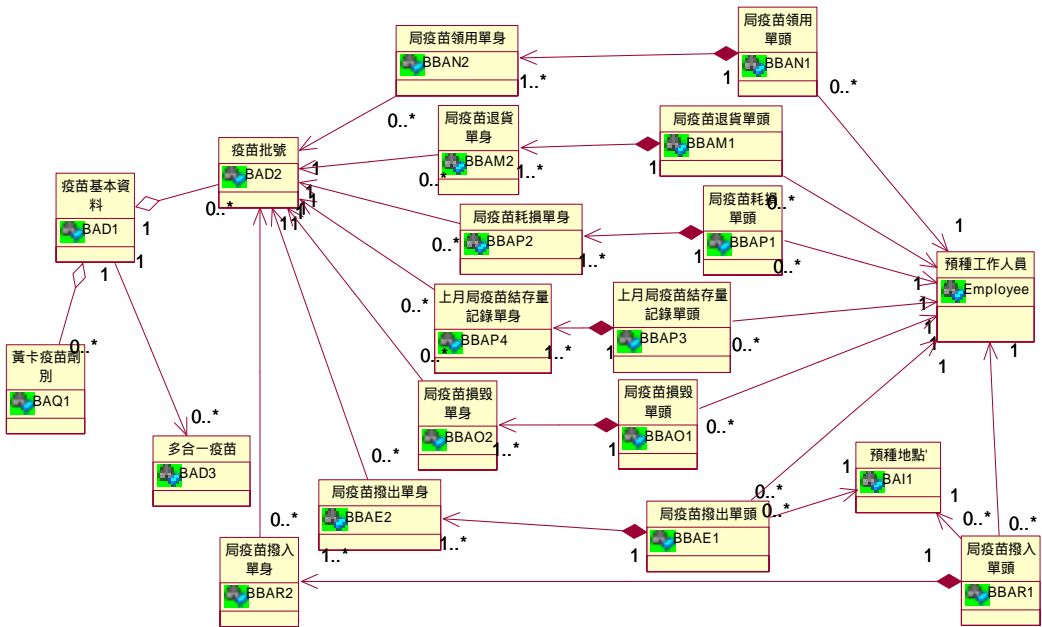
(五) 資料庫模式設計

衛生所



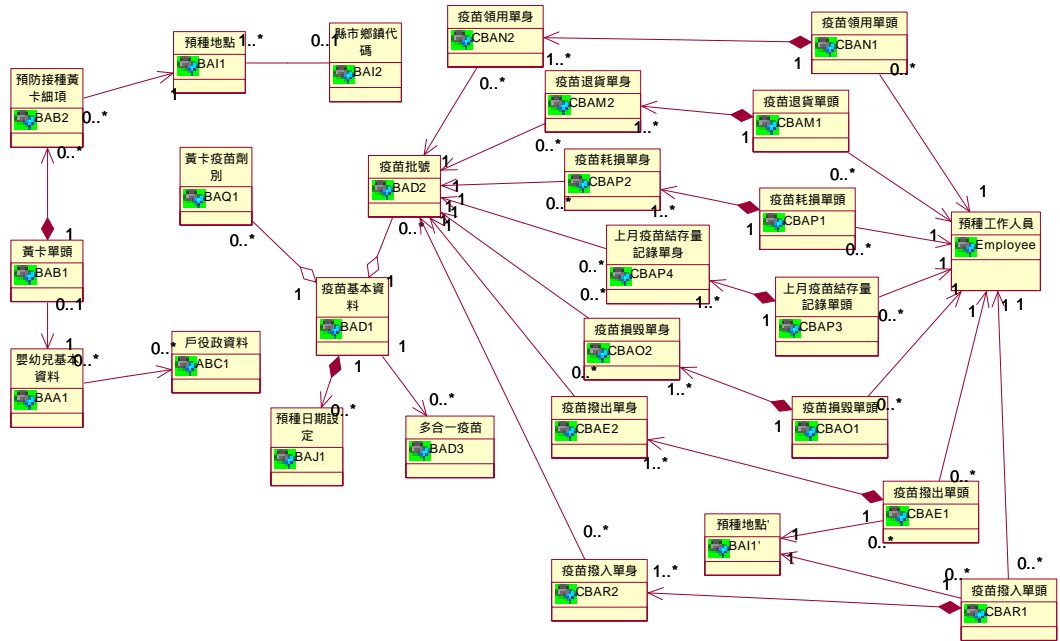
圖十七 衛生所 CD 圖

衛生局



圖十八 衛生局 CD 圖

中央疾病管制局



圖十九 中央疾病管制局 CD 圖

伍、整合型中央資料倉儲之建置及應用分析

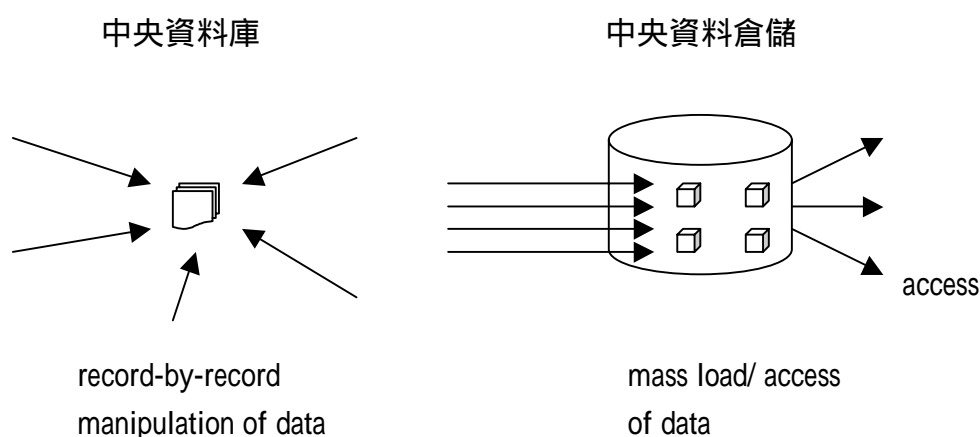
衛生署疾病管制局委託輔仁大學及元智大學進行整合型中央版預防接種系統最適模式之規劃，以與目前正在建置的地方版預防接種系統在未來能夠有效地銜接，並提供全國性預種統計分析作業的資訊化管理功能及國內各相關研究單位所需的預種資料庫。於此目的下，預計規劃一個資料倉儲作為整合型中央版預防接種系統之資料來源，以利用這些經由資訊科技所產生的龐大資料，提供全國性衛生資料的查詢。

當整合型中央資料庫建置完成後，系統中多半已經累積了不少預防接種相關資料。然而中央與地方資料庫可運用資料的方式仍尚嫌不足，原因在於中央的決策者對於資訊的要求愈來愈多樣化，而中央資料庫的內部結構不足以應付多角度且彈性的分析需求。造成這種現象的主要原因是由於預防接種中央資料庫資訊系統的資料匯集往往是片段而零散地存放在不同的資料檔案上，這些資料必須透過複雜的演算法與報表程式的處理，才能轉換成對管理者有意義的資訊。但是隨著醫療資訊的使用愈趨於專業化，近幾年來，逐漸有許多企業開始試圖將部份產生資訊的工作直接交由使用者來執行。透過資料庫與圖形化的報表產生器，由使用者自行定義資訊需求，自己來完成資料查詢、整理、及分析的工作。因此，資料倉儲化(Data Warehousing)這一新興的資訊技術在資訊業界逐漸佔有可觀的一席之地。而建構資料倉儲(Data Warehouse)在醫療應用策略上已逐漸受到重視，因為它不僅僅是現代醫療資訊發展的助力，更是處理未來醫療龐大資訊的必備利器。

一、NIIS 中央資料倉儲建立的目標

中央資料倉儲為中央資料庫的應用面衍生而來。藉由中央資料庫取得全國性預種資料，經過資料驗證、轉換、萃取與淨化等資料倉儲化過程進入中央資料倉

儲，並透過線上分析處理（OLAP, On-line Analytic Process）工具的使用，以時間、地點、年齡層、性別與疫苗種類等維度（Dimension）建構而成的 cube 具有快速的查詢效率，可使彈性化的報表查詢工作更具即時性，如圖二十所示。



圖二十、中央資料庫與中央資料倉儲的關係

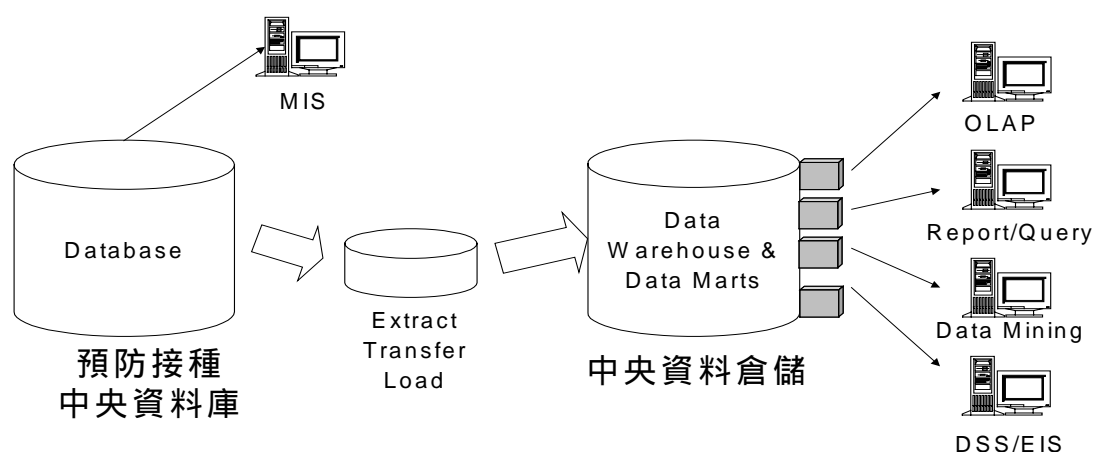
在使用效益上，對於資料倉儲的建立而言，需要的是以宏觀的角度去分析資料，因此較重視整合性報表的列印和趨勢分析。其中有三個主要的決策因子：第一是接受預種者（嬰幼兒）；第二是疫苗種類；第三是衛生單位（地點），此三者的交互關係是產生出整合式報表的關鍵。以接受預種者（嬰幼兒）、疫苗種類和所處區域衛生局的地域資料的交叉比對，產生出預種管理的項目，而利用疫苗種類和地域的關係產生了疫苗管理的項目，利用轉介歸戶管理的基本資料，產生出轉介歸戶的趨勢管理，因此主要能有利於疫苗管理、預種管理及趨勢分析等三方面的使用。

在決策支援的應用上，可充份掌控經營及醫療品質控管相關資訊，利用資訊科技（如資料探勘技術）自動產生相關需要的報表，供管理階層做研判分析，目前主要應用在醫院內醫療品質的控管，如入院死亡率、開刀失敗率、原因的探討等統計數字及意義分析。此外，預種組做疫苗管理（撥出、撥入、結存量等）決策的判斷模式亦可由此提供。

在了解病患的需求上，透過資料倉儲化的建立，能對病患的基本資料如病歷、年齡及地區做樞紐分析，就可以知道某地區病患對某種疾病的發病機率，作罹患相關疾病的比率分析。在醫療資源充分利用上，於相關資料能夠落實紀錄的前提下，可以透過用藥次數及種類的交叉分析中，了解各衛生單位的用藥情況，而透過資料庫的建置，也能夠了解醫療諮詢的使用狀況。在疾病通報的應用上，透過各地區預種完全率的長期觀察，中央可將疫苗種類與疾病發生的可能性作一比對，以預測與抑制各種傳染病的發生。

二、整合型中央版預防接種資料庫與資料倉儲之架構

預防接種中央資料庫將 25 個縣市衛生局地方資料庫整合後，透過中央交易伺服器擷取最新異動資料，透過管理資訊系統伺服器(MIS Server)可處理中央作業性彙整統計資料，提供了報表整合、疫苗管理、基本預種相關資料設定等功能。作業性資料經過資料轉換服務(萃取、轉化、載入)的處理，進入中央資料倉儲，而前端以了各類型伺服器提供不同的使用者進行整合及彈性的資訊分析工作，如圖二十一中的 OLAP、Data Mining、DSS/EIS 等分析工具。



圖二十一、預防接種中央資料庫與中央資料倉儲的架構

三、整合型中央版預防接種資料庫與資料倉儲之差異

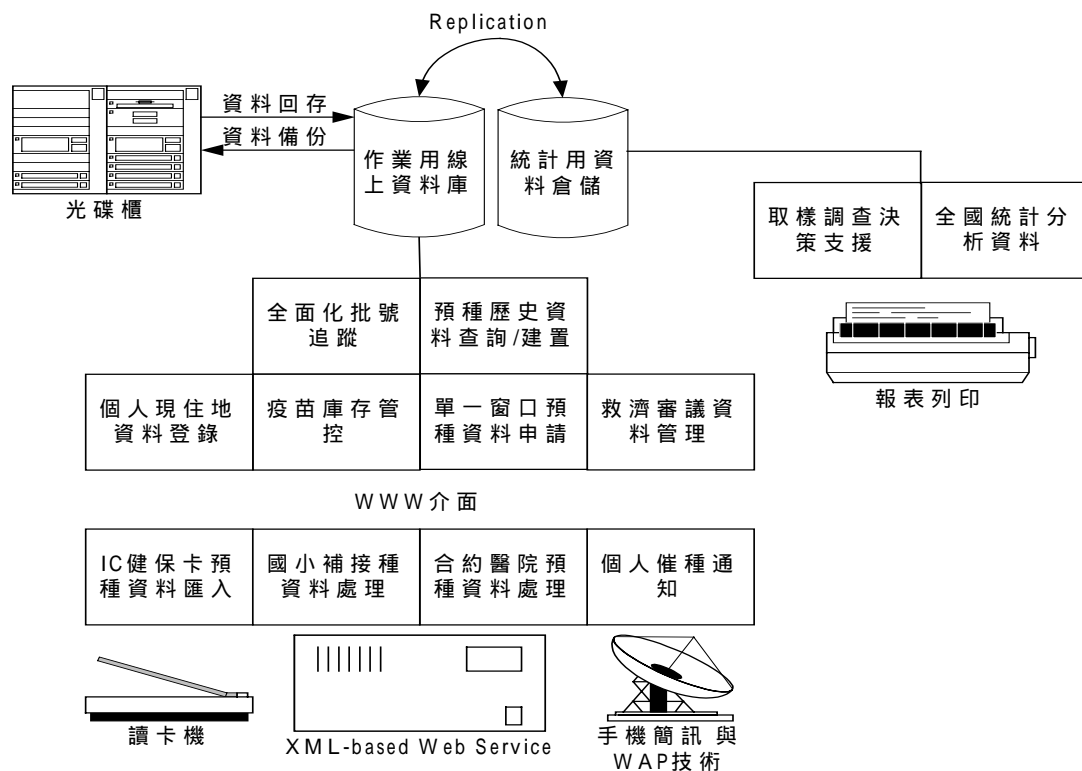
(一) 預防接種中央資料庫具有以下特性：

1. 是針對醫療業務自動化而設計的。
2. 為了協助醫療資訊使用者執行既有的活動。
3. 儲存的相關醫療資料比較偏重細節。
4. 資料的來源是日常工作的輸入。
5. 資料庫的內容會隨時被新增、刪除以及變更。
6. 資料處理是以交易為單位。
7. 設計資料庫是採用正規化的設計方式。

(二) 中央資料倉儲具有以下特性：

1. 是針對醫療資料分析需求而設計的。
2. 為了協助醫療資訊管理者執行管理決策。
3. 儲存的醫療資料偏重較高層級的彙總資訊。
4. 資料的來源是一般的作業性資料庫(中央與地方資料庫)。
5. 資料庫的內容是歷史性的資料，不會隨時變更。
6. 資料處理不是以交易為單位。
7. 設計資料庫採用非正規化的設計方式。

綜合上述而言，未來整合型中央資料庫架構(圖二十二)中應以一般作業性線上資料庫與統計分析性之資料倉儲為主。以作業性線上資料庫而言，除了允許多人同時修改資料之一般性功能外，全面化疫苗批號追蹤、預防接種歷史資料查詢與建置、個人現住地資料登錄、疫苗庫存管控與單一窗口之預種證明申請等便民服務，讓民眾透過友善地介面進一步快速地得到個人化相關資訊。以統計分析性之資料倉儲而言，透過快速地查詢與統計效能，配合決策支援系統的彈性取樣，產生全國性研究用預種報表，除了可使高階主管即時了解預防接種相關資訊，亦可便於提供學術單位取得預種統計分析資料。



圖二十二、整合型中央資料庫功能性架構[3]

四、中央資料倉儲的建置效益

整合型中央版預防接種資訊系統在未來應以資料倉儲化 (Data Warehousing) 為考量。資料倉儲是一種兼具效率及彈性的輔助決策工具，將一般存放於各部門、分院或來自於外在醫療單位資料庫中的分散資料需經過加總、轉換、過濾及萃取等過程轉而存放於資料倉儲中，成為對決策者有效的資訊。透過線上分析處理的快速性、分析性、共享性、多維面及資訊化等特性，配合決策支援系統 (DSS, Decision Support Systems) 提供的一般報表到更高階的決策支援功能，將使得醫療決策應用的層面更具效益。透過資料採礦 (Data Mining) 的技術，在資料倉儲中尋找出有具價值的隱藏資訊以及對資料歸納出具有結構的模式，並且加以分析，一方面作為醫療高階主管者在進行決策時之參考依據；另一方面，則為其他的醫療資訊使用者帶來更多且更有意義的資訊。因此，當整合型中央版預防接種資訊系統於資料倉儲化之始，應該先確定各種醫療資訊使用者

的需求，制定各種不同的維面資料分析模式，然後訂定資訊系統的應用工具，最後再將建立整合性且集中化的資料倉儲架構，如此一來，才能達到資料倉儲化的最高效益。

在應用層面而言，整合性的預防接種資訊系統，透過中央資料庫及各種資訊工具來輔助預防接種作業的進行，應具有下列功能：

1. 統計分析報表：

提供決策支援、研究用預種資料下載、全國及地方預防接種報表統計分析，以符合中央與地方的不同需求。

2. 便民服務

提供單一窗口預種證明書申請、預種資料的備份與回存、嬰幼兒催種通知及戶政資料的即時取得。

3. 疫苗庫存管理

提供疫苗運送路線配置、各地方即時疫苗庫存查詢、衛生局所疫苗調撥作業及疫苗廠商批號管理。

4. 提高預種完成率

透過預防接種中央資料庫，對於轉介歸戶之回報情形進行預防接種資料分析與催種通知，以提昇預種完成率。

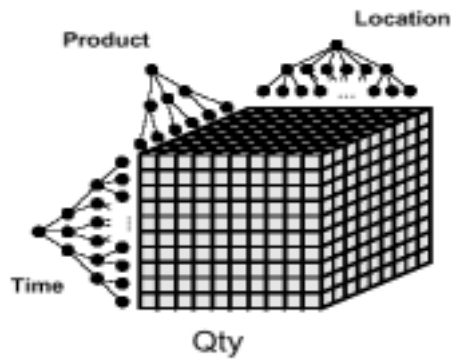
透過資訊科技的提昇，醫療院所的服務品質與效率都會有所提高，過去醫院以醫病為訴求的時代已經過去了，取而代之的是強調健康諮詢和顧客導向的服務，使得醫院不再只是醫病的場所，這些種種都是倚賴資訊科技和醫療間的相互配合，且透過資訊科技與醫療的結合，醫生能更加了解病患的需要，醫院管理者也能提供更合適的服務，也更能貼近病友的需要，對於醫療資源也更能有效利用。因此透過此類資料倉儲的建立，可減少以前所需花費的成本及時間，以獲得更高的顧客滿意度，並提供更有效的醫療資源分配。

五、資料倉儲應用工具之線上分析處理

線上分析處理(OLAP, On-line Analytical Processing)系統是一個用來幫助使用者能有效率並且有效的輕易完成資訊的維面結構(Dimensional structure)分析工作。OLAP 允許使用者自行選擇在樹狀資料結構中，展開同一階層資料的明細或點選更細一層的資料明細或樞紐，並且 OLAP 還支援某些階層的資料計算能力。而資料倉儲(Data Warehouse)與資料市集(Data Mart)則是 OLAP 的資料來源。OLAP 通常與查詢的功能息息相關，經由複雜的計算能力(資料比對、資料擷取等)來提供各種分析模式[1, 12, 13, 14, 41]。

線上分析處理(OLAP)系統主要將資料倉儲中的資料轉換至多維面(Multi-dimension)結構(圖二十三)[41]，並且呼叫 Cube(OLAP 的主要物件，包含多維面結構的資訊，每個 Cube 是由一群 Dimensions 及 Measures 所組成)，來執行有效及非常複雜的運算查詢。多維面的 OLAP 主要提供一個簡化複雜問題與查詢困難度的服務。當那些常被使用者查詢的問題都能夠以有規則性的呈現出這些資料的數量時，這大概表示 OLAP 已經趨近完善。要強調的是，不同的 OLAP 設計模式源自於不同的使用者需求。以下為 OLAP 的特性[12, 13]：

- (一) 歷史性：OLAP 資料來自於資料倉儲，其最大的特色在於資料具有歷史性，且通常涵蓋相當長的一段時間。
- (二) 唯讀性：由於資料倉儲的資料是歷史性的，這意謂著資料也具有唯讀性。一般而言，只有在原始資料來源有錯，才會變動資料倉儲的資料。
- (三) 群組分類性：OLAP 資料強調所有細節的資料，並重視資料的分類以利於統計分析。
- (四) 合併性：將企業內所有的異質性資料，經過資料驗證、資料遷移、資料篩選及資料轉換的方式全部彙集合併在資料倉儲中。
- (五) 一致性：包含了資料格式一致與資料單位一致，如此以方便於資料的分析與比較。



圖二十三、預種系統多維面資料結構

因此，醫療資訊使用者透過線上分析處理工具的使用，可彈性化訂製其所需的報表，藉由多維面的資料架構，以快速取得具結構性的資料。

六、預防接種資料倉儲之應用

整合型中央版預防接種資訊系統在未來應以資料倉儲化(Data Warehousing)為考量。資料倉儲是一種兼具效率及彈性的輔助決策工具，將一般存放於各部門、分院或來自於外在醫療單位資料庫中的分散資料需經過加總、轉換、過濾及萃取等過程轉而存放於資料倉儲中，成為對決策者有效的資訊。透過線上分析處理(OLAP)的快速性、分析性、共享性、多維面及資訊化等特性，配合決策支援系統(DSS, Decision Support Systems)提供的一般報表到更高階的決策支援功能，將使得醫療決策應用的層面更具效益。透過資料採礦(Data Mining)的技術，在資料倉儲中尋找出有具價值的隱藏資訊以及對資料歸納出具有結構的模式，並且加以分析，一方面作為醫療高階主管者在進行決策時之參考依據；另一方面，則為其他的醫療資訊使用者帶來更多且更有意義的資訊。因此，當整合型中央版預防接種資訊系統於資料倉儲化之始，應該先確定各種醫療資訊使用者的需求，制定各種不同的維面資料分析模式，然後訂定資訊系統的應用工具，最後再將建立整合性且集中化的資料倉儲架構，如此一來，才能達到資料倉儲化的最高效益。

在應用層面而言，整合性的預防接種資訊系統，透過中央資料庫及各種資訊

工具來輔助預防接種作業的進行，應具有下列功能：

(一) 統計分析報表：提供決策支援、研究用預種資料下載、全國及地方預防接種報表統計分析，以符合中央與地方的不同需求。

(二) 便民服務：提供單一窗口預種證明書申請、預種資料的備份與回存、嬰幼兒催種通知及戶政資料的即時取得。

(三) 疫苗庫存管理：提供疫苗運送路線配置、各地方即時疫苗庫存查詢、衛生局所疫苗調撥作業及疫苗廠商批號管理。

(四) 提高預種完成率：透過預防接種中央資料庫，對於轉介歸戶之回報情形進行預防接種資料分析與催種通知，以提昇預種完成率。

透過資訊科技的提昇，醫療院所的服務品質與效率都會有所提高，過去醫院以醫病為訴求的時代已經過去了，取而代之的是強調健康諮詢和顧客導向的服務，使得醫院不再只是醫病的場所，這些種種都是倚賴資訊科技和醫療間的相互配合，且透過資訊科技與醫療的結合，醫生能更加了解病患的需要，醫院管理者也能提供更合適的服務，也更能貼近病友的需要，對於醫療資源也更能有效利用。因此透過此類資料倉儲的建立，可減少以前所需花費的成本及時間，以獲得更高的顧客滿意度，並提供更有效的醫療資源分配。

中央資料庫及資料倉儲一旦建置完成，可以滿足下列幾項系統需及應用的效益。在使用效益上，對於資料倉儲的建立而言，需要的是宏觀的角度去分析資料，因此較重視整合性報表的列印和趨勢分析。其中有三個主要的決策因子：第一是接受預種者（嬰幼兒）；第二是疫苗種類；第三是衛生單位（地點），此三者的交互關係是產生出整合式報表的關鍵。以接受預種者（嬰幼兒）、疫苗種類和所處區域衛生局的地域資料的交叉比對，產生出預種管理的項目，若利用疫苗種類和地域的關係產生了疫苗管理的項目，利用轉介歸戶管理的基本資料，產生出轉介歸戶的趨勢管理，因此主要能有利於疫苗管理、預種管理及趨勢分析等方面的使用。

在決策支援的應用上，可充份掌控經營及醫療品質控管相關資訊，利用資訊

科技（如資料探勘）會自動產生相關需要的報表，供管理階層做研判分析，目前主要應用在醫院內醫療品質的控管，如入院死亡率、開刀失敗率、原因的探討等統計數字及意義分析。此外，預種組做疫苗管理（撥出、撥入、結存量等）決策的判斷模式亦可由此提供。以下就三個與決相關的分析性系統加以闡述：

（一）疫苗配給量的預測分析

此功能在於應用整體性疫苗庫存的功能，對於各個衛生局、衛生所與合約醫療院所使用疫苗的狀況做一觀察及歷史記錄的追蹤，以建立疫苗建議採購量決策分析，作為疫苗管理的判斷準則。應用例如有衛生局、所疫苗庫存相關資料（領用、撥入、撥出、退貨、耗損、結存量之管理）等。

（二）疫苗施打與副作用及疫苗施打與疾病預防之關聯性

此功能在於利用全部的地方預種資料中的副作用與疫苗基本資料的相關資訊利用關聯分組的觀念，試圖找出該疫苗可能發生的副作用，以增加藥品使用的安全性，或是配合疾病通報系統資料庫，將疾病流行地區與先關施打疫苗比率或是施打疫苗種類作關聯分組，能分析是否該疫苗是否能夠成功抑制疾病，或是分析出疾病發生與疫苗完成率的關聯性，訂定疫苗施打完成率標準，加強醫療資源的應用，而這些訊息也可提供產學界作研究參考，加速醫療的發展。

（三）施打疫苗的行為分析

此功能在於利用各種預防接種的執行情況與戶政資料作配合，找出未按時或不接受施打疫苗的行為模式，透過分群的方式，將此族群區隔出來，這樣一來不但有利於未來預防接種政策的推動，更能減少催種行政費用及衛生人員的工作量，加強醫療行政資源的有效使用效果。

在了解病患的需求上，透過資料倉儲化的建立，能對病患的基本資料如病歷、年齡及地區做樞紐分析（cube），就可以知道某地區病患對某種疾病的發病

機率，做罹患相關疾病的比率分析。在醫療資源充分利用上，於相關資料能夠落實紀錄的前提下，可以透過用藥次數及種類的交叉分析中，了解各衛生單位的用藥情況，而透過資料庫的建置，也能夠了解醫療諮詢的使用狀況。

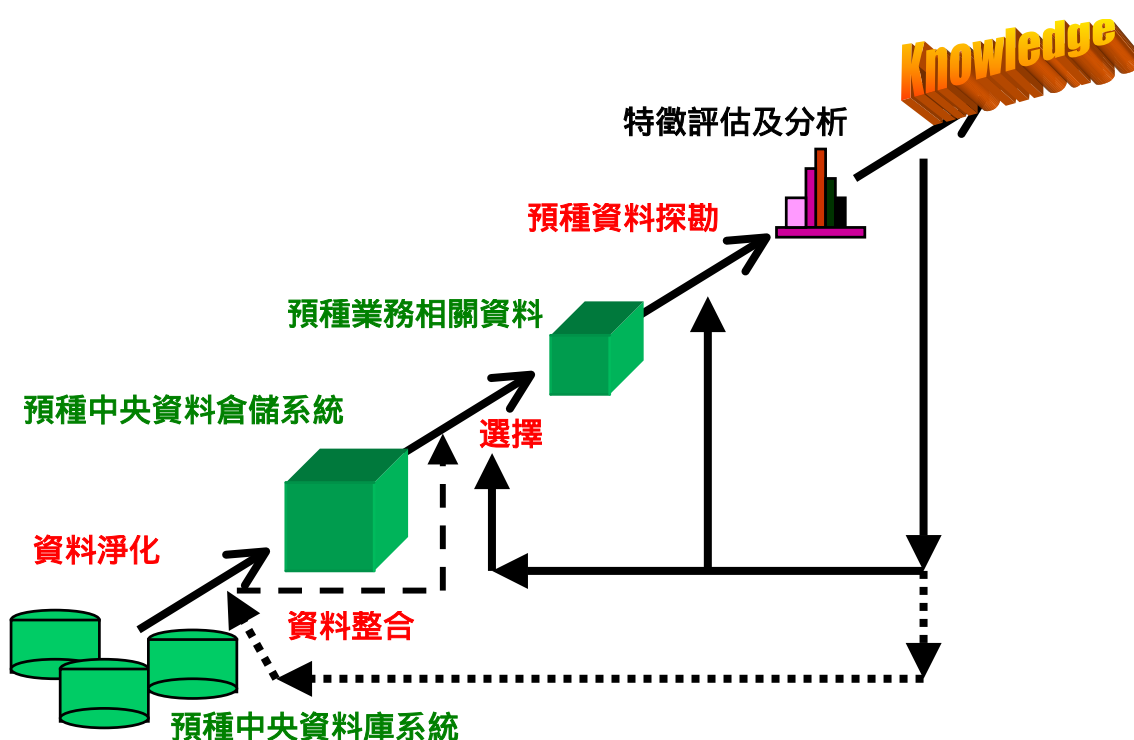
在疾病通報的應用上，透過各地區預種完全率的長期觀察，中央可將疫苗種類與疾病發生的可能性作一比對，以預測與抑制各種傳染病的發生。衛生署疾病管制局為了建立全國性之預防接種資料庫及網路服務系統，擬結合原有衛生局所之預防接種系統，運用現代化的網路架構，提供即時性及完整性的資料庫管理與網路服務。本節從地方版 PHIS 預防接種系統引申建立整合型中央版預防接種系統 [3]，配合中央資料庫及資料倉儲的逐漸建立完成可有效來達成。

陸、資料倉儲之決策支援及分析應用

未來完成預種資料倉儲建置之後，即會面臨各級決策之運用，其中包括作業決策、高等決策等等。故本計畫擬先針對未來可能面臨之決策問題，規畫相關之決策應用，並在計畫的進行過程，逐步建構最適化之決策支援模型。目前本計畫所規畫之決策需求包括疫苗管理相關 OLAP 分析、預防接種決策問題與資料探勘技術之結合以及運用類神經網路法於施打疫苗失約人數之預測等三個主要項目。

一、預防接種決策問題與資料探勘技術之結合

經由先前的文獻中，KDD Process 是一連串的知識探索過程，而 Data Mining 只是其中的一個過程，但是這個過程也是最重要的階段之一，套用於目前預種資料庫及資料倉儲，我們可以得到一個簡單的示意圖，如圖二十四所示。



圖二十四、預防接種系統知識探索過程

(一) 決策支援系統與資料探勘技術

在預防接種的業務支援上，除了原有管理資訊系統 (Management Information System) 所含之結構化問題，亦可以決策支援系統 (Decision Support System) 輔助非結構化的問題或突發的決策需求，降低不確定性 (Uncertainty)，並利用模擬 (Simulation) 或模式 (Model) 來作預測或推測。通常一個政策要推行之初，可以根據資料或數據的分析並以系統化或科學化的方法，提高決策的品質。然而，資訊系統仍然要靠具有專業知識與嫺熟業務的專家，提供模式建構的基礎。預防接種業務經常面臨關鍵性的決策，這些決策靠非常有經驗與專業背景的專家制定，如果能透過知識探索 (Knowledge Discovery in Databases, KDD) 的過程，將其發展為決策支援系統，不僅可以輔助決策，更是作好知識管理的重要關鍵。

由於資料探勘過程是循環的，因此有一連串的工作要進行。這些工作多半是在假設檢定與知識探索兩者間交換。其中假設檢定的階段需要有 OLAP 及其他查詢工具，如：交叉分析表、監督式購物籃分析及傳統統計。知識探索的階段則有賴於群集偵測、決策樹、類神經網路、連結分析、記憶基礎理解及非監督式購物籃分析。

要決定在這其中的某一個時點採用何種資料探勘技術，端看特定資料探勘工作要達成的目標及所要分析資料的取得性來決定。我們選擇資料探勘技術的方式有兩個步驟：

1. 將預防接種決策問題解讀成一連串的資料探勘工作
2. 了解可取得資料的內容、型態以及各記錄間的關聯性

(二) 將預防接種決策問題轉換為資料探勘工作

預防接種資訊與決策系統的功能經本研究歸納後，主要有相關的決策包括：

1. 單劑量或多劑量疫苗之最佳採購模式：

疫苗之採購通常根據經驗來決定單(多)劑量之採購量,單劑量與多劑量有其不同的成本、效益與耗損率,如何採購以達最佳效益,可以建立一個決策模式,設定相關參數,如體內抗體存量等以輔助決策。

2. 疫苗可預防疾病發生流行之預測：

以歷史資料為基礎結合流行病學,建立傳染病預防及監控模式。對於像天花、流行性感冒等疾病,可以根據地區預防接種的完成率、體內抗體存量等數據以判斷疫情擴大的機率,以及如何適時注射疫苗以抑制疫情。

3. 施打疫苗之行為分析：

利用各種預防接種的執行情況與衛生人口資料作配合,找出未按時或不接受施打疫苗的行為模式,透過分群的方式,將此族群區隔出來,這樣一來不但有利於未來預防接種政策的推動,更能減少催種行政費用及衛生人員的工作量,加強醫療行政資源的有效使用效果。

若將預防接種決策問題的型態對映資料探勘的功能面,第一,我們可看出「單劑量或多劑量疫苗之最佳採購模式」是根據經驗,也就是其歷史資料而建立的一個決策模式。因此屬於資料探勘的預測(prediction)功能。第二,「疫苗可預防疾病發生流行之預測」是以歷史資料為基礎結合流行病學,建立傳染病預防及監控模式。因此也屬於資料探勘的預測(prediction)功能。第三,「施打疫苗之行為分析」是找出未按時或不接受施打疫苗的行為模式,以將此族群區隔出來。因此屬於資料探勘的群集化 (clustering)功能。

(三) 選擇正確的資料探勘技術

資料探勘的技術有以下幾種：

- 傳統統計(statistics)
- 購物籃分析(market basket analysis)

- 記憶基礎理解(memory based reasoning)
- 基因演算法(genetic algorithms)
- 群集偵測(clustering analysis)
- 連結分析(link analysis)
- 決策樹(decision tree)
- 類神經網路(neural network)

表十六、各項技術適用之任務

技術	分類	推估	預測	關聯分組	群集化	描述
傳統統計	✓	✓	✓	✓	✓	✓
購物籃分析			✓	✓	✓	✓
記憶基礎理解	✓		✓	✓	✓	
基因演算法	✓		✓			
群集偵測					✓	
連結分析	✓		✓	✓		
決策樹	✓		✓		✓	✓
類神經網路	✓	✓	✓		✓	

預防接種決策問題與各項技術適用之任務交叉來看，符合預測(prediction)與群集化 (clustering)的技術有傳統統計、購物籃分析、記憶基礎理解、決策樹與類神經網路。如此我們已確認處理預防接種決策問題可運用的資料探勘技術，但是如何在這些技術中選擇最適用的技術來做為決策支援系統之用，還須了解可取得資料的內容、型態以及各記錄間的關聯性，才能經過實際資料測試與評估，以做出最佳的系統規劃。

一旦確認了資料探勘的工作內容，並且經由工作的內容的確立，來縮小資料探勘工具的考慮範圍，事實上，可得資料的特色也可以進一步協助我們精確地選

擇工具。目前可考慮的資料探勘工具如 BusinessMiner (Business Objects)、Intelligent Miner (IBM)、PolyAnalyst (Megaputer)等，其對於各領域處理資料探勘在業界皆有不錯的評價。

二、疫苗管理 OLAP 需求分析

疫苗管理工作可以說是地方版預種資訊系統提供的重大功能之一，但是在地方版的系統之中，並不能快速的根據過去的疫苗使用量，分析各地區，及未來的疫苗消耗量，所以建置資料倉儲將能夠更快速而準確的以 Cube 及圖形化的方式顯示相關的資訊，加速有關主管及人員決策的時間。

Cube1: 疫苗量與時間之間的關係

本 Cube 的建置可用於疫苗管理的應用上，分析過去的需求，並預測未來可能的量。

(1) 維度設計

時間	疫苗種類
年	單劑量
季	多劑量
月	

(2) 測量值

疫苗使用量

Cube2: 疫苗量和地域及時間的關係評估預測

可用以找尋時間、地區及疫苗量之間的關係，更詳細的分析各地之需求。

(1) 維度設計

時間	地區	疫苗種類
年	縣市	單劑量
季	鄉	多劑量

月	村	
---	---	--

(2) 測量值

疫苗使用量

實例分析-----桃園縣復興鄉

本研究之資料倉儲及線上分析處理是以全國性預防接種資訊管理系統為範本建置，並以桃園縣復興鄉衛生所預防接種資料庫為資料來源，試圖規劃各項疫苗庫存管理功能，以疫苗撥入處理及疫苗領用處理二項功能得以規劃建置，然更突顯撥入及領用兩項處理之重要性。

資料轉換部分為顧及普及性，選擇 Microsoft Enterprise Server 2000 所提供之 Data Transformation Service(DTS)做為資料萃取、轉換以及載入之工具，同時設定自動化資料載入排程與撰寫詮釋性資料(Metadata)。線上分析處理建置則運用 Microsoft SQL Server 2000 之 Analysis Services 所提供之分析管理員。並以 Microsoft Excel 2000 作為統計數值、報表及各式視覺化圖表呈現，試圖以高階主管易於了解之方式提供資訊及輔助決策分析

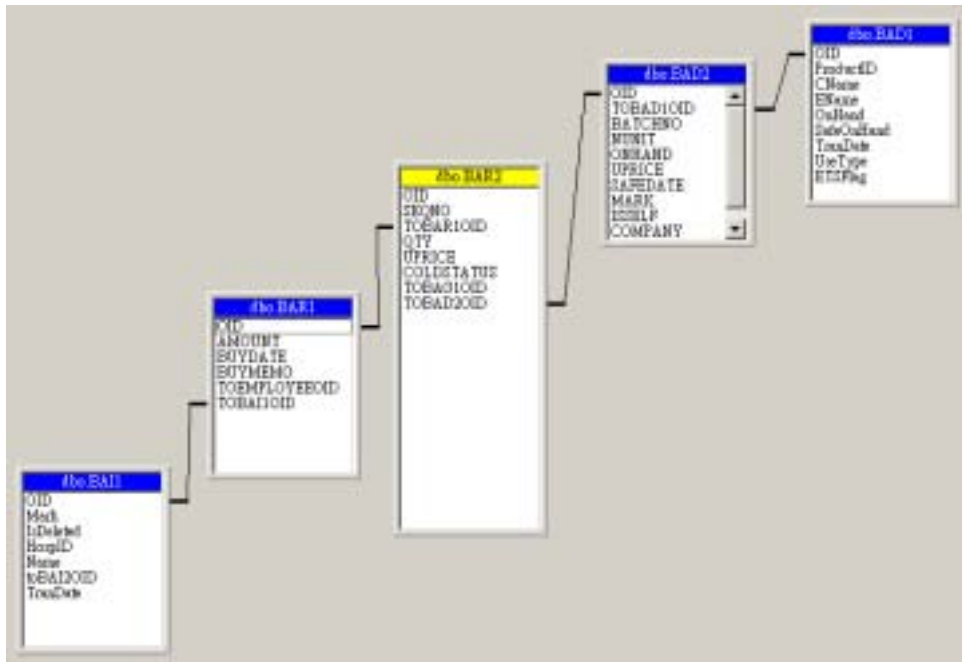
1. 疫苗撥入處理

在此為了能夠建置一個完善的疫苗撥入處理多維度立體方塊(Cube)，爾等必須詳細列出疫苗撥入處理 Cube 建置時之相關參數設定，其中包括 Fact Table、測量值、維度規劃、實體儲存設計以及效能改善比率等等，如下表所示。

疫苗撥入處理 Cube 建置參數表

Fact Table: BAR2
Measure: BAR2.Qty
時間維度: 星狀 Schema, BAR1, (BAR1.Buydate)
地點維度: 雪花狀 Schema, BAR1, BAI1, (BAI1.Name)
疫苗種類維度: 雪花狀 Schema, BAD2, BAD1, (BAD1.Cname)
實體儲存設計: MOLAP
改善效能: 50%

圖二十五為本 Cube 所引用之關聯資料表，其關係資料的取得，所以甚為重要。其中包含疫苗撥入單身(BAR2)、疫苗撥入單頭(BAR1)、預種地點(BAI1)、疫苗批號(BAD2)及疫苗基本資料(BAD1)。在 Cube 建置的過程之中，爾等必須在維度設定時額外帶入疫苗批號關聯資料表，如此才得以將疫苗撥入單身及疫苗基本資料作掛勾，從而取得疫苗之基本資料項目，在此提出以供參考。



圖二十五、疫苗撥入處理關聯資料表

Cube 瀏覽器 - 疫苗撥入處理

Measures: Qty

撥入時間: 所有的 撥入時間

Crname	Name	所有的 撥入地點	桃園縣復興鄉衛生所	桃園縣衛生局
所有的 疫苗類型		1,931.00	60.00	1,871.00
A型肝炎疫苗(小兒三劑裝)				
A型肝炎疫苗(小兒兩劑裝)		30.00	30.00	
B型肝炎血漿疫苗				
B型肝炎免疫球蛋白				
B型肝炎病毒工程疫苗		200.00		200.00
b型嗜血桿菌疫苗				
小兒麻痺口服疫苗		115.00		115.00
不活化小兒麻痺疫苗				
日本腦炎疫苗		117.00		117.00
水痘疫苗				
卡介苗		107.00		107.00
白喉百日咳破傷風混合疫苗		217.00	30.00	187.00
白喉破傷風非組胺性百日咳				
白喉破傷風混合疫苗		9.00		9.00
流行性感冒疫苗		392.00		392.00
破傷風減量白喉混合疫苗		62.00		62.00
麻疹疫苗		80.00		80.00
麻疹腮腺炎德國麻疹混合疫苗		602.00		602.00
德國麻疹疫苗				

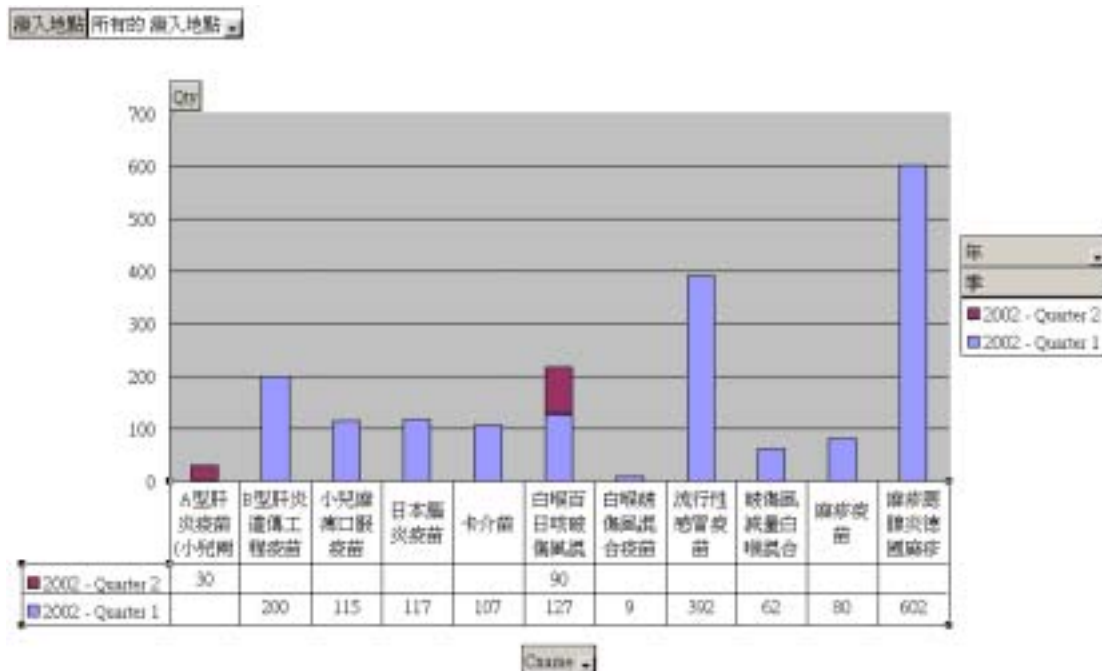
在成員上按兩下來將它向下切入或向上切出。

關閉 說明(S)

圖二十六、疫苗撥入處理多維度 OLAP 分析

在此所展示的是疫苗撥入處理狀況的 OLAP 樞紐分析，爾等以時間及疫苗種

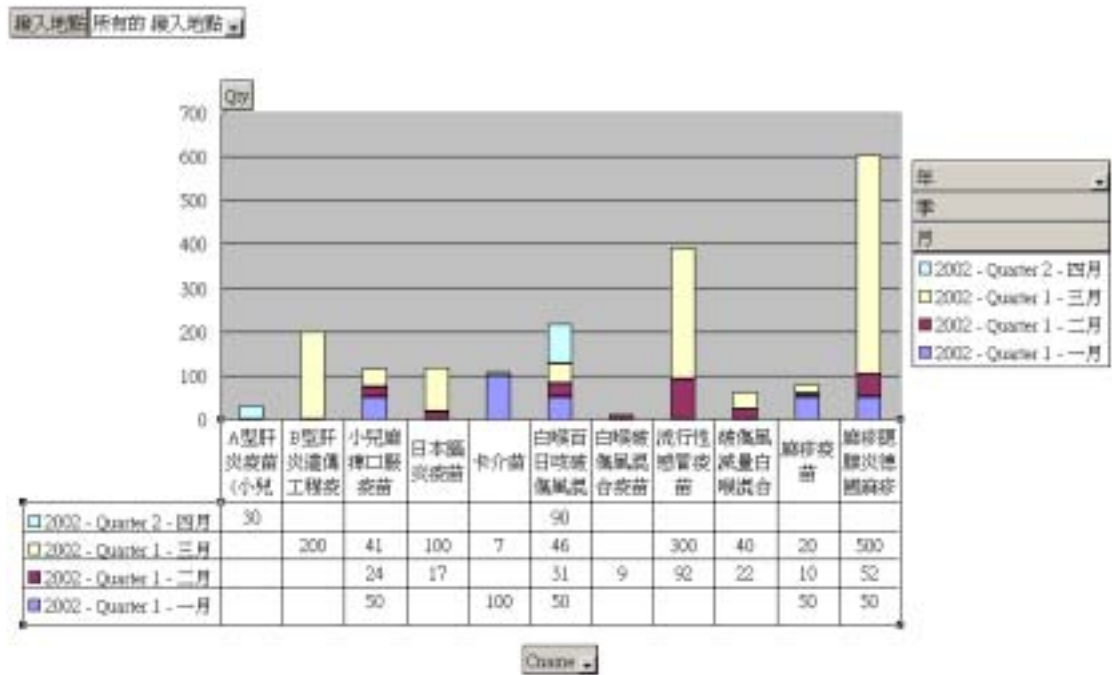
類兩個維度作交互集合，並以撥入地點為輔助選項，產生地區疫苗撥入量的圖形。可以從圖形中觀察到，於 2002 年度開始，桃園市復興鄉衛生所從各地撥入之疫苗異動者共佔十九項中的十一項，至於在時間的維度層級中，在圖一中展開到季的部份。所以在圖二中，爾等即展示將時間月份逐項展開(Drill-Down)的情形，這也是 OLAP 強大的功能之一，時間維度將以更詳細的月份圖形進行展示，圖形並以不同的顏色來區分不同的月份，讓使用者可以更一目瞭然的了解當年度各個月份的疫苗撥入量的差異。



圖二十七、疫苗撥入處理 OLAP 樞紐分析之一

從圖一展示的結果爾等可以得知，2002 年以來，桃園市復興鄉衛生所從各地撥入之疫苗中，麻疹腮腺炎德國麻疹相較於其它疫苗的量特別的大，共有 602 劑，其次依序為流行性感胃疫苗的 392 劑及 B 型肝炎遺傳工程疫苗的 200 劑。未來將此功能有效運用於中央型資料庫/倉儲將能有效的研判目前疫苗使用的趨勢及更進一步的計算疫苗使用量，有效的管控疫苗的撥入處理情形。

另外，另外一個訊息透露第一季的疫苗撥入量相較於第二季所佔的比重大，探討其原因時，爾等可以透過 OLAP 的 Drill Down 功能展開時間的層級，從月份的分析表來看，我們可以得知第一季共分為三個月份，第二季則只計算到四月，而在四個月之中，三月的疫苗撥入量尤其大，這可以透露出在三月時，民眾施打的概率較高，且衛生所的業務量也隨之增加，是否適時的提供人力上的輔助，將成為疫苗管理中的一個新議題。



圖二十八、疫苗撥入處理 OLAP 樞紐分析之二

2. 疫苗領用處理

Fact Table: BAN2

Measure: BAN2.Qty

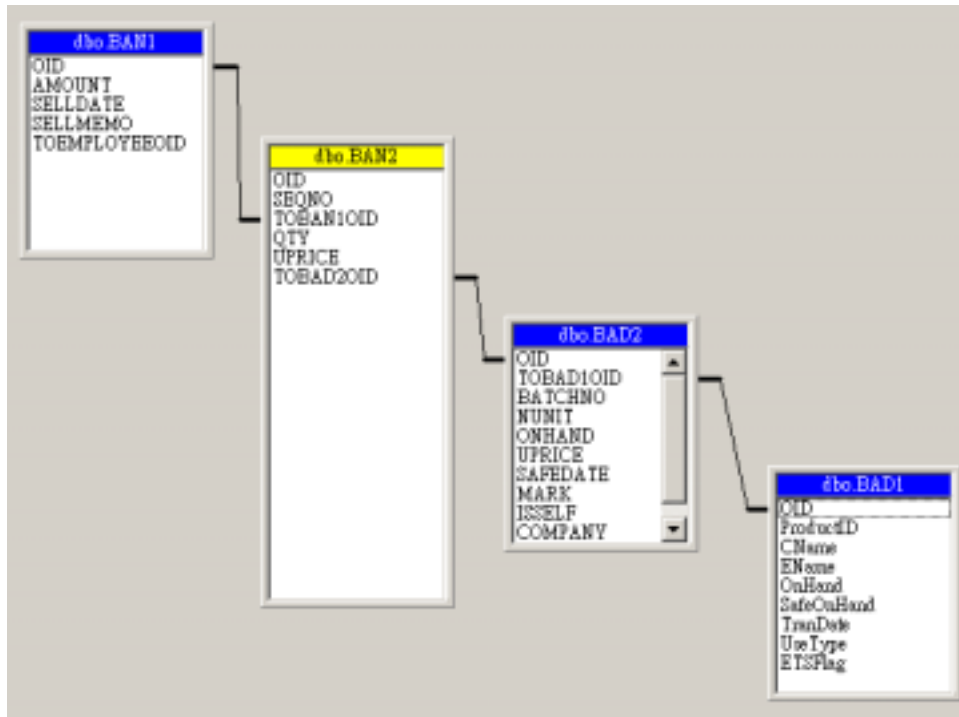
時間維度: 星狀 Schema, BAN1, (BAN1.Selldate)

地點維度: None

疫苗種類維度: 雪花狀 Schema, BAD2, BAD1, (BAD1.Cname)

實體儲存設計: MOLAP

改善效能: 50%



圖二十九、疫苗領用處理所使用之關聯資料表

Cube 瀏覽器 - 疫苗領入處理

Measures: Qty
 摺入時間: 所有的摺入時間

Cname	Name	所有的摺入地點	桃園縣復興鄉衛生所	桃園縣衛生局
所有的疫苗類型		1,931.00	60.00	1,871.00
A型肝炎疫苗(小兒三劑)		30.00	30.00	
B型肝炎血漿疫苗				
B型肝炎免疫球蛋白				
B型肝炎疫苗工程疫苗		200.00		200.00
b型肝炎疫苗				
小兒麻痺口服疫苗		115.00		115.00
不活化小兒麻痺疫苗				
日本腦炎疫苗		117.00		117.00
水痘疫苗				
卡介苗		107.00		107.00
白喉百日咳破傷風混合疫苗		217.00	30.00	187.00
白喉破傷風非細胞性百日咳				
白喉破傷風混合疫苗		9.00		9.00
流行性感冒疫苗		392.00		392.00
破傷風減量白喉混合疫苗		62.00		62.00
麻疹疫苗		80.00		80.00
麻疹腮腺炎德國麻疹混合疫苗		602.00		602.00
德國麻疹疫苗				

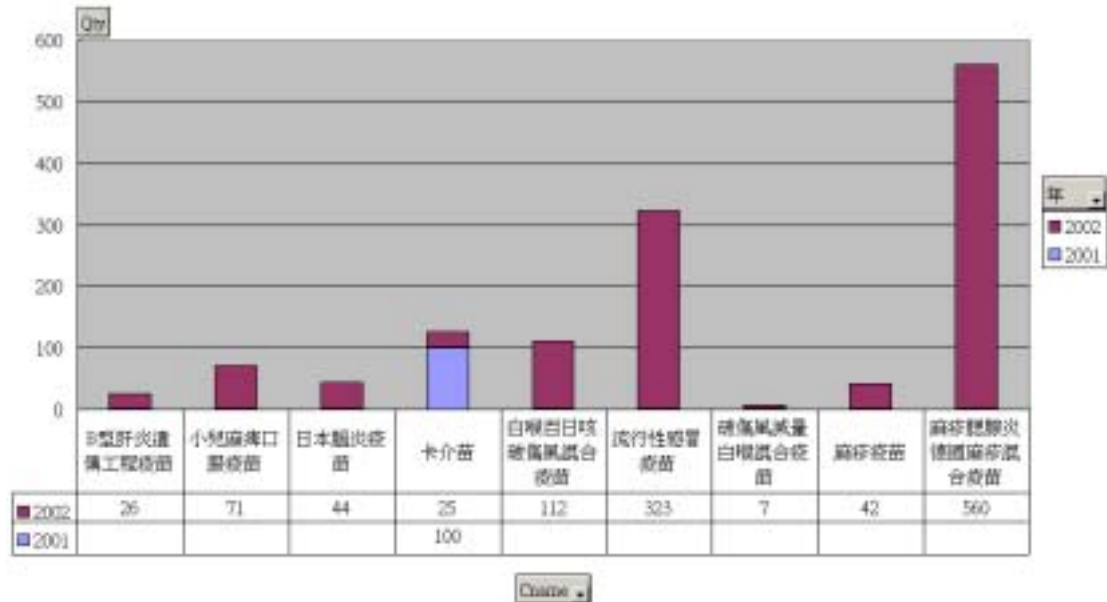
在成員上按兩下來將它向下切入或向上切出。

關閉 說明(S)

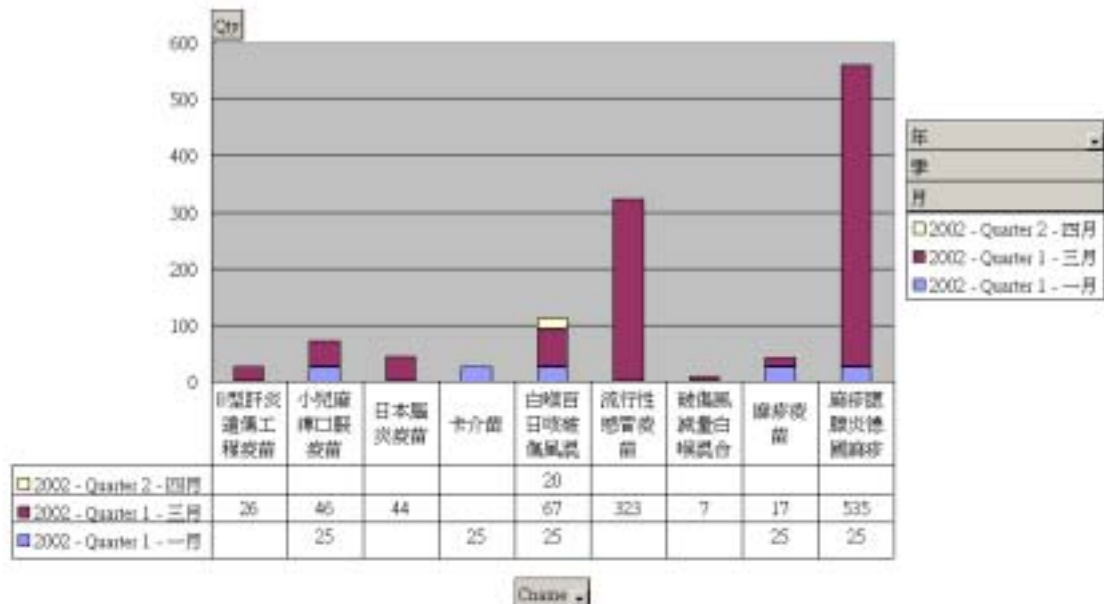
圖三十、疫苗領用處理多維度 OLAP 分析

疫苗領用處理

此處所展示的是桃園市復興鄉衛生所內部人員領用疫苗的情形，爾等可以在圖形中找到一些訊息，即疫苗領用的情形在 2002 年的比重較大，所以我們可以針對 2002 年進行 Drill Down 的工作，其中發現四個月中，有三個月發生領用的情形。其中尤以第三季的量特別的大，這樣的結果和之前於疫苗撥入情形中所推斷的，第三個月施打的民眾大量增加，且業務量呈大幅成長，結果不謀而合。



圖三十一、疫苗領用處理 OLAP 樞紐分析之一



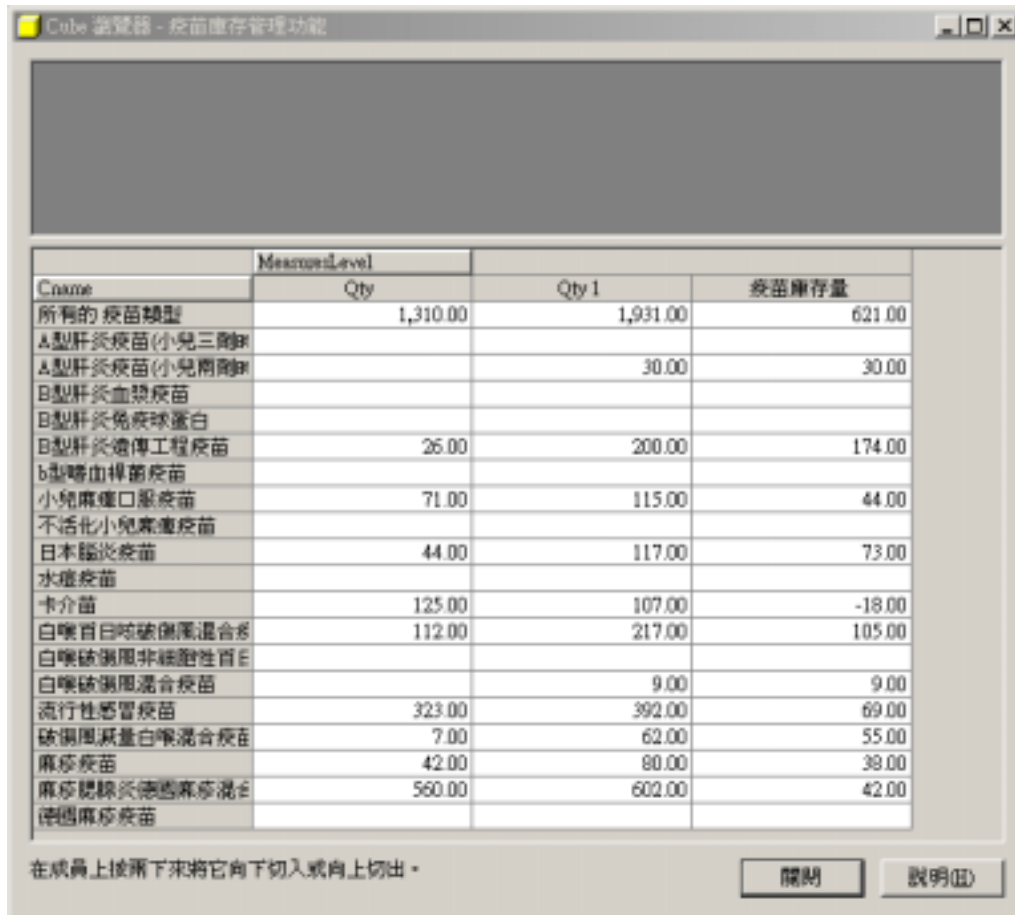
圖三十二、疫苗領用處理 OLAP 樞紐分析之二

3. 建立疫苗庫存管理虛擬立體方塊

Cube: 疫苗撥入處理, 疫苗領用處理

維度: 疫苗種類

導出成員建置: 疫苗庫存量 = 疫苗撥入處理.Qty - 疫苗領用處理.Qty



Name	MessagesLevel	Qty	Qty 1	疫苗庫存量
所有的疫苗類型		1,310.00	1,931.00	621.00
A型肝炎疫苗(小兒三劑)				
A型肝炎疫苗(小兒兩劑)			30.00	30.00
B型肝炎血漿疫苗				
B型肝炎免疫球蛋白				
B型肝炎遠傳工程疫苗		26.00	200.00	174.00
b型嗜血桿菌疫苗				
小兒麻痺口服疫苗		71.00	115.00	44.00
不活化小兒麻痺疫苗				
日本腦炎疫苗		44.00	117.00	73.00
水痘疫苗				
卡介苗		125.00	107.00	-18.00
白喉百日咳破傷風混合疫苗		112.00	217.00	105.00
白喉破傷風非細胞性百日咳				
白喉破傷風混合疫苗			9.00	9.00
流行性感冒疫苗		323.00	392.00	69.00
破傷風減量白喉混合疫苗		7.00	62.00	55.00
麻疹疫苗		42.00	80.00	38.00
麻疹腮腺炎德國麻疹混合疫苗		560.00	602.00	42.00
德國麻疹疫苗				

圖三十三、疫苗庫存量導出成員建立

計算完疫苗撥入及領用的資料之後，我們可以很快的透過虛擬 Cube 的建置，將疫苗撥入量減去疫苗領用量之後，即可得出目前最新的疫苗庫存量。在這裡疫苗庫存量之計算如此精簡之前提在於，目前資料庫並無其它疫苗處理之動作，如退貨、撥出、耗損及消耗等，所以能以簡單的數學運算結算出來。如果要考量其它的疫苗處理動作，則必須針對疫苗處理的關聯性深入探討。在圖五中所展示的即為目前桃園縣復興鄉衛生所目前的疫苗庫存情形。



圖三十四、疫苗庫存量

三、運用灰色理論預測疫苗劑量

灰色預測(Grey Prediction)是以 GM(1,1)模型為基礎對現有數據所進行的預測方法，找出某一數列間各個元素之未來動態狀況。是近年來快速發展、應用的理論，其預測能力相當的好，可由早期文獻發現。因此才廣受各界好評。將與灰色預測相關的重要文獻整理於下表，可發現都是將其與傳統統計方法作比較，結果皆顯示灰色預測有很好的效果，到了近期則多利用灰色預測於實際應用上的預測及如何改善其預測誤差。

表十七、灰色預測的實際應用

作者	研究對象	研究方法	與本文相關的研究結果
林房儋(1994)	針對 24 項人體計測資料進行分析	灰色預測模型、最小平方法	灰色預測模型的預測誤差相當低，且比被普遍採用的最小平方方法有更高的準確率
許巧鶯、溫裕弘(1997)	1980 年至 1990 年台灣地區國際航線出入境總旅客年運量	灰色預測模型、ARIMA 預測模式	證實以灰色模式建構之國際航空客運量預測模式確實比傳統迴歸分析模式與 ARIMA 模式較具解釋與預測能力

顏晴榮 (1997)	1989 年至 1992 年台電實際毛發電總量	灰色預測模式	不管採用幾筆數據，以預測一年的精確度最佳，其次發現，三年內的預測皆以七筆原始數據預測精確度為最高
許巧鶯、溫裕弘 (1998)	1984 年至 1997 年台灣某一國籍航空公司	灰色預測模式、灰色聚類	預測模式之運量預測結果對運量變化與未來發展的動態捕捉，並配合航線運量變動之形態設計結果驗證確為可行，且提供較高之決策彈性
劉定焜、施能仁 (1998)	1996 年台灣發行情加權股價收盤指數	灰色預測模式	灰色預測模型優於時間序列模型，有利於投資人進行準確的策略判斷，亦可提供投資人進入期貨與現貨市場之避險重要訊息
施東河、徐桂祥 (1999)	1983 年至 1992 年台灣地區壽險需求預測	模糊灰色預測模式	國民所得、平均壽命與死亡率為影響台灣地區壽險需求之主要解釋變數，且利用模糊灰色系統來作預測，準確度可達一定水準以上
何怡慧 (1999)	1993 年至 1996 年上市公司財務報表內 15 項重要財務比率	灰色預測模式、傅立葉級數殘差修正法、殘差 GM(1,1)模型	1. 灰色預測模式在短期財務績效指標預測之應用效果較傳統時間序列分析法或指數平滑法好 2. 利用傅立葉級數殘差修正法及殘差 GM(1,1)模型做殘差修正，可使原來運用灰色預測模式的預測準確度大有改善
李宗儒、鄭卉方 (2000)	1996 年紅豆價格	灰色預測模式	驗證灰色預測模式對於預測農作物價格之便利性，此資料亦讓生產者、批發商、及政府三方面在生產數量、訂價及制訂政策上有所助益
許哲強、陳家榮 (2000)	1994 年至 1997 年台灣中部地區尖峰負載預測	灰色預測模式	在台灣中部地區尖峰負載預測上有不錯的表現，可以精確了解未來總體電力負載形態，且可預知電力需求地

			區，使電力資源能有效運用
許哲強、賴正文、陳家榮(2000)	1981年至1998年台灣本島電力需求年資料	灰色預測模型、灰色殘差修正模型	改良型灰色殘差修正模型之各預測模式在事後預測精確度表現上明顯皆較傳統灰色預測模式為佳
H. Morita et al. (1996)	以1979年至1986年實際每年貨物最大需求來預測1987年至1992年需求量	Grey dynamic model、Grey system theory	1. 結果顯示灰色系統理論應用於貨物量的預測是一種新而有效的工具 2. 使用灰色動態模型來計算邊際能力的上下界限是很有效的方法

灰色系統理論

灰色系統理論(grey system theory)是由中國大陸華中理工大學鄧聚龍教授於1982年率先提出的理論[45,46,47]。其主要是針對現有的資訊不完整與系統不明確之狀況下進行系統的關聯分析(relational analysis)、建模(model)、預測(prediction)以及決策(decision)等。該理論的研究，已涵蓋各個領域，如農業、交通、氣象、工程、運輸、經濟、醫療、軍事、文化、教育、地質以及管理等等。灰色理論研究的項目可以歸納為：灰生成、灰關聯分析、灰建模、灰色預測、灰決策及灰控制六種。其中“灰”定義為訊息不完整或不確定，而其元素稱為“灰元”，介於“黑色”與“白色”之間，是一種相對的顏色。

灰色生成

在灰色理論中所謂的灰色生成技巧，事實上為補充訊息的數據處理方式，是一種就數找數的規律方法，利用此種方式，可以在一些欲處理但是雜亂又無章的數據中，設法將被掩蓋的規律及特加以淨現。換句話說，我們可以利用灰色生成的手段降低數據中的隨機性，並提昇數據的規律性。在灰色理論中常用的生成方法有：

1. 灰色關聯生成(Grey Relational Generating Operation, GRGO)：將數據依實際情形在不失真之下所做的數據處理。
2. 累加生成(Accumulated Generating Operation, AGO)：將數據依次累加。

3. 逆累加生成(Inverse Accumulated Generating Operation, IAGO)：累加生成的逆運算。
4. 插值生成(Interpolation Generating, IG)：利用現有數據及數學方法建立其間所缺失的數據。

灰色建模

利用生成過的數據建立一組灰差分方程與灰擬微分方程之模式，稱為灰色建模，一般可分成下面幾種：

1. GM(1,1)：表示一階微分，輸入變數一個，一般做預測用。
2. GM(1,N)：表示一階微分，輸入變數 N 個，一般做多變量關聯分析用。
3. GM(0,N)：是 GM(1,N)的特例，表示零階微分，輸入變數 N 個，一般做多變量關聯分析用。

灰色預測

灰色預測是以 GM(1,1)模型為基礎對現有數據所進行的預測方法，實際上則是找出某一數列中間各個元素之未來動態狀況，主要的優點為所需的數據不用太多及數學基礎相當簡單。灰色預測系統建立模型的思想是直接將序列轉化為微分方程，從而建立的是抽象系統發展變化的動態模型 - Grey Model (GM)。若以灰色預測的功能來區分，綜合文獻得知灰色預測主要有以下五種不同之用途[36]：

1. 數列預測：若是等時距取樣所得之數列，經過累加生成運算之數據建立灰微分方程，進而預測數列下一個或下幾個值為何。
2. 災變預測：預測在一定時間內是否會發生災變或某種異常值何時會再出現，又稱為異常預測。
3. 季節災變預測：是將災變預測數列的分辨率提高後得到的。
4. 拓撲預測：將已知數列連成曲線圖，在曲線上依某個值找出相交叉之時刻數列，再依 GM(1,1)模型預測未來出現的時刻，再將各未來發生之定值連成曲線，以掌握整個數據曲線未來的發展變化，又稱為圖形預測。
5. 系統預測：結合 GM(1,1)與 GM(1,N)模型，對系統中各變量同時進行預測，求出各變量間發展的相互關係。

本研究依其功能劃分屬於數列預測，其主要是以通過對原始數據處理來尋找數的規律，這叫作數的生成，這是一種就數找數的現實規律途徑。找出潛藏的內在規律，且關鍵是如何去挖掘它並且利用。此處使用 GM(1,1) 作為預測模型，其表示一階微分，且輸入變數為一個的 GM 模型，由於其計算簡單，適用範圍廣，亦稱為單序列一階線性動態模型。灰色預測模型的建立，少到只需四筆數據即可。

GM(1,1) 灰色預測模型及滾動模型之建模

(一) 擷取資料並進行級比檢驗 [27]

級比檢驗的目的在於檢驗各衛生所疫苗施打人數是否可作為灰色建模的依據。級比 $\lambda(k)$ (Class Ratio) 檢驗如下：

$$\lambda(k) = \frac{x(k-1)}{x(k)} \quad k \in \{2,3,\dots,n\} \quad (3-11)$$

當滿足 $\lambda(k) \in (0.135, 7.289), \forall k$ 時，表示原始序列 $x^{(0)}$ 可作 GM(1,1) 建模。

(二) 建立累加生成數列 (AGO : Accumulated Generating Operation)

$$\text{首先建立 GM}(1,1) \text{ 源模型為 } x^{(0)}(k) + az^{(1)}(k) = b \quad (3-12)$$

它雖然近似滿足微分方程構成條件，但它畢竟不是真正的微分方程，因此在灰色預測中使用一般微分方程

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (3-13)$$

$$\text{即建立累加生成數列 } X_{(k)}^{(1)} = \{X_{(1)}^{(1)}, X_{(2)}^{(1)}, \dots, X_{(k)}^{(1)}\} \quad (3-14)$$

來取代 GM(1,1) 源模型 $x^{(0)}(k) + az^{(1)}(k) = b$ 。這種取代不是用數學方法來推導的，而是一種“借用”及“白化”的手段，所以稱 $\frac{dx^{(1)}}{dt} + ax^{(1)} = b$ 為 GM(1,1)

源模型的白化方程或影子方程。兩者間對應關係如下：

$$\frac{dx^{(1)}}{dt} \Rightarrow x^{(0)}(k) \quad (3-15)$$

$$ax^{(1)} \Rightarrow az^{(1)}k \quad (3-16)$$

$$x^{(1)} = (x^{(1)}(1), \wedge, x^{(1)}(k)) \Rightarrow x^{(1)}(t) \quad (3-17)$$

$$x^{(0)} = (x^{(0)}(1), \wedge, x^{(0)}(k)) \Rightarrow x^{(0)}(t) \quad (3-18)$$

(三) 求均值生成

計算及辨識GM(1,1)參數a, b, 其中a為預測模型的發展係數, b為預測模型的灰作用量。令GM(1,1)源模型 $x^{(0)}(k) + az^{(1)}(k) = b$ 滿足序列 $x^{(1)}$ 與 $x^{(0)}$

$$x^{(1)} = (x^{(1)}(1), \wedge, x^{(1)}(n)) \quad (3-19)$$

$$x^{(0)} = (x^{(0)}(1), \wedge, x^{(0)}(n)) \quad (3-20)$$

$$x^{(1)}(k) = \sum_{m=1}^k x^{(0)}(m) \quad (3-21)$$

則有

$$x^{(0)}(2) + az^{(1)}(2) = b \quad (3-22)$$

$$x^{(0)}(2) + az^{(1)}(3) = b \quad (3-23)$$

.....

$$x^{(0)}(2) + az^{(1)}(n) = b \quad (3-24)$$

$$\text{其中 } z^{(1)}(k) = 0.5[x^{(1)}(k) + x^{(1)}(k-1)] \text{ 為均值生成} \quad (3-25)$$

(四) 建構數據矩陣, 利用最小平方法(least square method)求出待估參數a及b:

$$\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \wedge \\ x^{(0)}(n) \end{bmatrix} = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \wedge \\ -z^{(1)}(n) & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \quad (3-26)$$

得到GM(1,1)參數a, b的矩陣式為:

$$[a, b]^T = (B^T B)^{-1} B^T Y_N \quad (3-27)$$

$$\text{其中 } B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \wedge \\ -z^{(1)}(n) & 1 \end{bmatrix} \quad (3-28)$$

$$Y_N = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \wedge \\ x^{(0)}(n) \end{bmatrix} \quad (3-29)$$

$$\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} \quad (3-30)$$

(五) 解出灰微分方程式的累加預測式為：

$$\hat{x}(k+1) = \left[x^{(0)}(1) - \frac{b}{a} \right] e^{-ak} + \frac{b}{a} \quad (3-31)$$

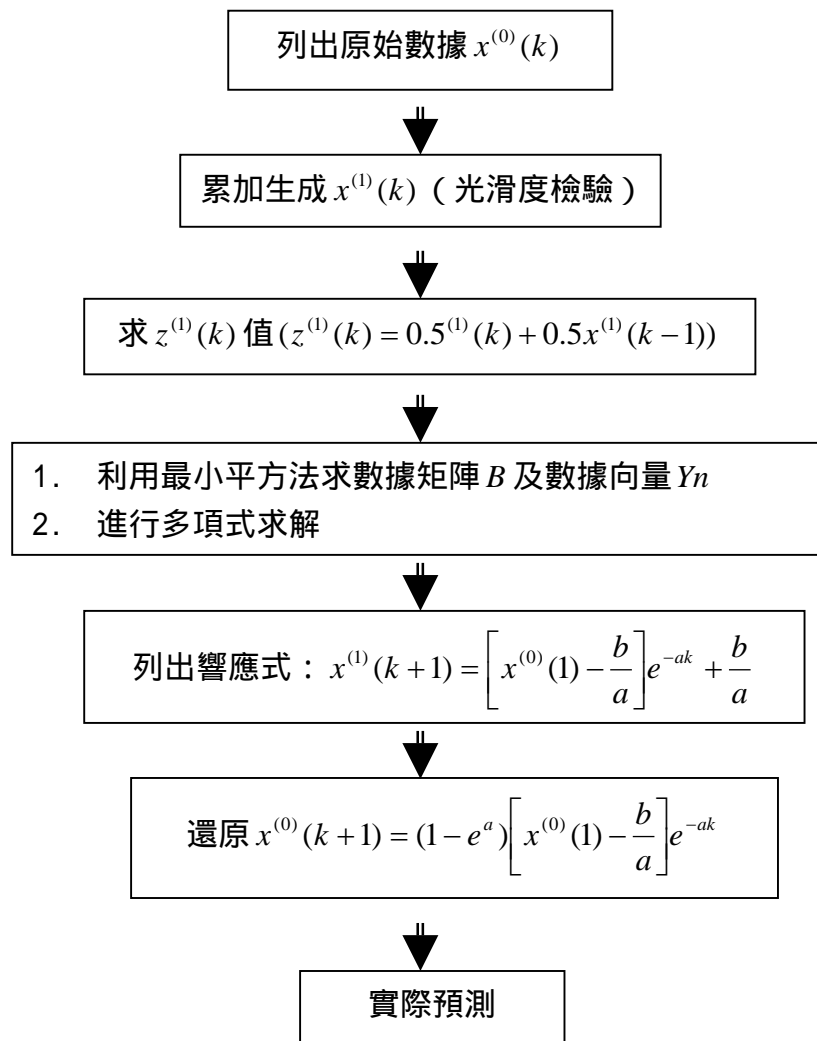
(六) 累減生成之形成 (IAGO : Inverse Accumulated Generating Operation) :

因預測模型並非對原始資料建模，所以要以累減生成來還原所預測之數據，用一次累減生成數列是以待運算數列之第一個元素作為新數列之第一個元素，以待運算數列之第二個元素減去第一個元素作為新數列之第二元素，依此類推，即：

$$\hat{X}^{(0)}(k) = \hat{X}^{(1)}(k) - \hat{X}^{(1)}(k-1) , \text{ 且 } \hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) \quad (3-32)$$

GM(1,1) 灰色預測模式建模生成流程圖

綜合上述相關步驟，本研究整理 GM(1,1) 模型之建模生成流程圖，提供一整體性架構以供參考[27]。



圖三十五、GM(1,1)建模生成流程

預測準確度檢驗

預測精確度為(Forecasting Accuracy)為評估預測模型所得預測值與實際值之間的差，也就是預測誤差，為決定一個預測模型成功或失敗的一種量度 [35]。以樣本外預測誤差作為預測績效評估的基礎。目前所常見的預測模型準確度分析方法很多，本文主要採用準確度比(Accuracy Ration, AR)及以誤差平均平方和之均方根(Root-Mean-Squared-Error, RMSE)、平均絕對誤差(Mean Absolute Error, MAE)、平均絕對誤差百分比(Mean Absolute Percent Error,

MAPE)，用以檢測預測值之可信度。

(一) 準確度比(Accuracy Ratio ; AR)

利用預測誤差來分析預測模型之準確度可採用絕對誤差的觀念，列出三種準確度比的型式：

其中 $\hat{x}(k)$ 為利用經濟預測模型所得到的預測數據序列

$q(k) = x(k) - \hat{x}(k); k = 1, 2, \dots, n$ 為預測誤差數據序列。

$$1. \text{ 平均準確度比 : } AR_1 = 1 - \frac{\sum |q(k)|}{\sum \hat{x}(k)} \quad (3-33)$$

$$2. \text{ 均方準確度比 : } AR_2 = 1 - \frac{(\frac{1}{n} \sum [q(k)]^2)^{1/2}}{\frac{1}{n} \sum \hat{x}(k)} \quad (3-34)$$

$$3. \text{ 最大準確度比 : } AR_3 = 1 - \frac{\max |q(k)|}{\frac{1}{n} \sum \hat{x}(k)} \quad (3-35)$$

以上所定義之各種準確度比(AR) 均為無因次(Dimensionless)，於進行預測模型之準確度分析比較時，可避免觀察事物單位的影響。換言之準確度比能提供一個更廣泛的準確度分析。當預測誤差越小，AR 就越接近 1；當預測誤差越大，AR 就越接近 0。

(二) 以誤差平均平方和之均方根(Root Mean Squared Error , RMSE)為衡量標準。RMSE 值若越接近 0，表示預測能力越好。

$$RSME = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{X}^{(0)}(k) - X^{(0)}(k))^2} \quad (3-36)$$

n：測試組的範例數目； $X^{(0)}(k)$ ；第 k 個範例目標值； $\hat{X}^{(0)}(k)$ ：第 k 個範例預測值。

(三) 平均絕對誤差(Mean Absolute Error, MAE) , 當 MAE 值越小 , 表示預測能力越好。

$$MAE = \sum \left| X^{(0)}(k) - \hat{X}^{(0)}(k) \right| \div n \quad (3-37)$$

(四) 平均絕對誤差百分比(Mean Absolute Percent Error, MAPE) , 當 MAPE 值越小 , 表示預測能力越好。

$$MAPE = \sum \left| (X^{(0)}(k) - \hat{X}^{(0)}(k)) \times 100 / X^{(0)}(k) \right| \div n \quad (3-38)$$

式中 $X^{(0)}(k)$ 表第 k 期之實際值 , $\hat{X}^{(0)}(k)$ 表第 k 期之預測值 , n 為預測期數。

應用灰色預測模型於疫苗採購量預測之應用

在未來，將應用灰色預測模型於疫苗採購量之預測，在此之前，本文先以桃園縣復興鄉衛生所為研究對象，進行疫苗使用量之預測，在未來中央資料庫建置完成之後，即可針對全國之疫苗使用量進行預測，從而推估全國疫苗採購量之建議值，以做為中央主管機關決策之參考。

一、研究說明

■ 研究對象

- 桃園縣復興鄉衛生所

■ 研究環境

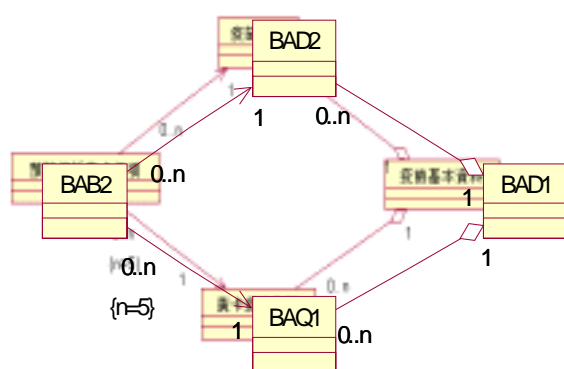
- Windows 2000 Server
- SQL Server 2000
- Grey Model (1 , 1) Version 1.0

■ 研究目的

- 預測當地 2001 年小兒麻痺口服疫苗使用量並進行後分析

二、疫苗工作量相關資料類別圖

本研究嘗試於使用 SQL 查詢語句，篩選出桃園縣復興鄉衛生所小兒麻痺口服疫苗之歷年使用量，因此首先調查有那些資料庫表格與疫苗工作量有所關聯，調查後得出共有下列四個表格具有關係：預防接種黃卡細項、疫苗批號、黃卡疫苗劑別、疫苗基本資料。



圖三十六、資料庫表格

三、疫苗基本資料

在疫苗基本資料部份，共有十九項不同的疫苗種類，B 型肝炎疫苗又分為四種不同種類的血清，A 型肝炎疫苗又分為二種不同種類的血清。

1	B 型肝炎免疫球蛋白
2	卡介苗
3	白喉百日咳破傷風混合疫苗
4	小兒麻痺口服疫苗
5	麻疹腮腺炎德國麻疹混合疫苗
6	B 型肝炎血漿疫苗
7	流行性感感冒疫苗
8	A 型肝炎疫苗(小兒兩劑時程)
9	不活化小兒麻痺疫苗
10	日本腦炎疫苗
11	破傷風減量白喉混合疫苗
12	麻疹疫苗
13	B 型肝炎遺傳工程疫苗
14	德國麻疹疫苗
15	白喉破傷風混合疫苗
16	A 型肝炎疫苗(小兒三劑時程)
17	白喉破傷風非細胞性百日咳混合疫苗
18	B 型嗜血桿菌疫苗
19	水痘疫苗

圖三十七、疫苗基本資料

四、小兒麻痺口服疫苗歷年施打工作量

經查詢資料庫後得知，小兒麻痺疫苗之歷年施打工作量如下表，我們可以發現在 1994 年以前之疫苗施打工作量特別的低，這表示之前的預防接種觀念並不是那麼的普及，至於 1998 後的工作量則呈現下降的趨勢，其原因乃在於出生人口數的減少所造成。

	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
1												70	75	75	76
2												89	59	59	81
3												114	89	78	75
4												108	68	46	54
5												119	102	92	98
6												147	103	92	71
7												145	84	76	105
8												111	119	75	87
9												97	75	74	70
10												88	110	88	75
11												89	78	69	66
12												83	92	74	21
Total	0	3	2	14	43	450	839	1116	1229	1201	1199	1260	1054	898	879

圖三十八、小兒麻痺口服疫苗歷年施打工作量

五、2001 年灰色預測結果分析

本研究將上述資料經過灰色預測模型之預測後，得到下表之結果，其中分別

以不同的資料筆數進行預測。其結果顯示以五筆資料來進行預測可以得最佳之啟發式解。另外亦可發現，在一般情況下，預測誤差會大於平均殘差，其原因在於平均殘差是以擬合度較高之預測值與原始值進行計算得來，故其差異性會較小，而預測誤差則決定於預測出來之未來數據與屆時發生之未來值所進行計算得來，故其差異會較大。由最後的預測結果得知，若以五筆數據為原始數據進行預測，可得到 2001 年之預測值為 861 劑，與 2001 年之實際值 879 相差緊 18 劑而以，預測誤差只有 2%而以，故其精確度可謂相當的準確。

原始數據筆數	年度區間	2001 實際值	2001 預測值	平均殘差	預測誤差
4	1997~2000	879	753.3997	0.4713	14.29
5	1996~2000	879	861.2409	5.2200	2.02
6	1995~2000	879	924.0855	5.5424	5.13
7	1994~2000	879	958.8601	5.5607	9.09
8	1993~2000	879	1015.6786	7.3733	15.55
9	1992~2000	879	1108.8497	12.5272	26.15
10	1991~2000	879	1230.8595	24.3925	40.03

圖三十九、2001 年灰色預測結果分析

六、未來研究方向

本文嘗試利用 GM(1,1) 灰色預測模型於桃園縣復興鄉衛生所之小兒麻痺口服疫苗施打量之預測上，未來還有以下幾項工作要進行。

1. 針對全國性之接種資料，進行各項疫苗之總和預測，而非僅以單一衛生所為預測對象進行預測。
2. 由於 GM(1,1) 僅考慮單一數列之發展趨勢，對於外界環境變數之影響並未考量，在未來不排除加入其他外部環境，如出生率、死亡率等因素，以取得更精確之預測結果。
3. 過去曾有不少學者提出提高灰色預測精確度的方法，未來將檢視過去之成效，嘗試利用不同模式提高灰色預測之精確度，期望能有更好的預測效果可以提供做為主管單位採購時之參考依據。

四、運用資料探勘分群演算法預測施打疫苗失約特性

因應資訊科技及網際網路的高度成長，國內醫療程序再造工程 (medical process reengineering) 的風潮已然起步，醫療資訊系統不再僅是免除人工的自動化的設計而已，對於醫療機構而言，更需要能提供決策輔助的決策支援系統，

這樣的一個系統建立，不但對於醫療機構也所助益，對於廣大的民眾來說，更能提供越趨近以病人為導向 (patient-oriented) 的服務，提供社會大眾更好的疾病預防及更佳的健康照護及醫療保健服務。

由於顧客失約會造成資源分配與管理的不便，要訂定準確的預約制度相對而研究相當困難，但由於顧客需求之不確定性較大，且需求量不易準確預測，為提升服務品質，企業往往採用預約制度，並且排定顧客接受服務之時間。為避免顧客失約造成服務資源損失，準確的預測失約顧客的特性以決定預種催注政策，因此分群的研究成為重要課題。我們期望透過分群的相關技術直接從顧客資料中找出相關顧客特性準確訂定顧客失約模型，利用失約特徵的建立有助我們對於再次失約的監控有直接的幫助，並可利用我們所建立出來的失約特性，對於事後催注政策的建立能有直接的成效。

一般而言，失約的影響有下列幾項：

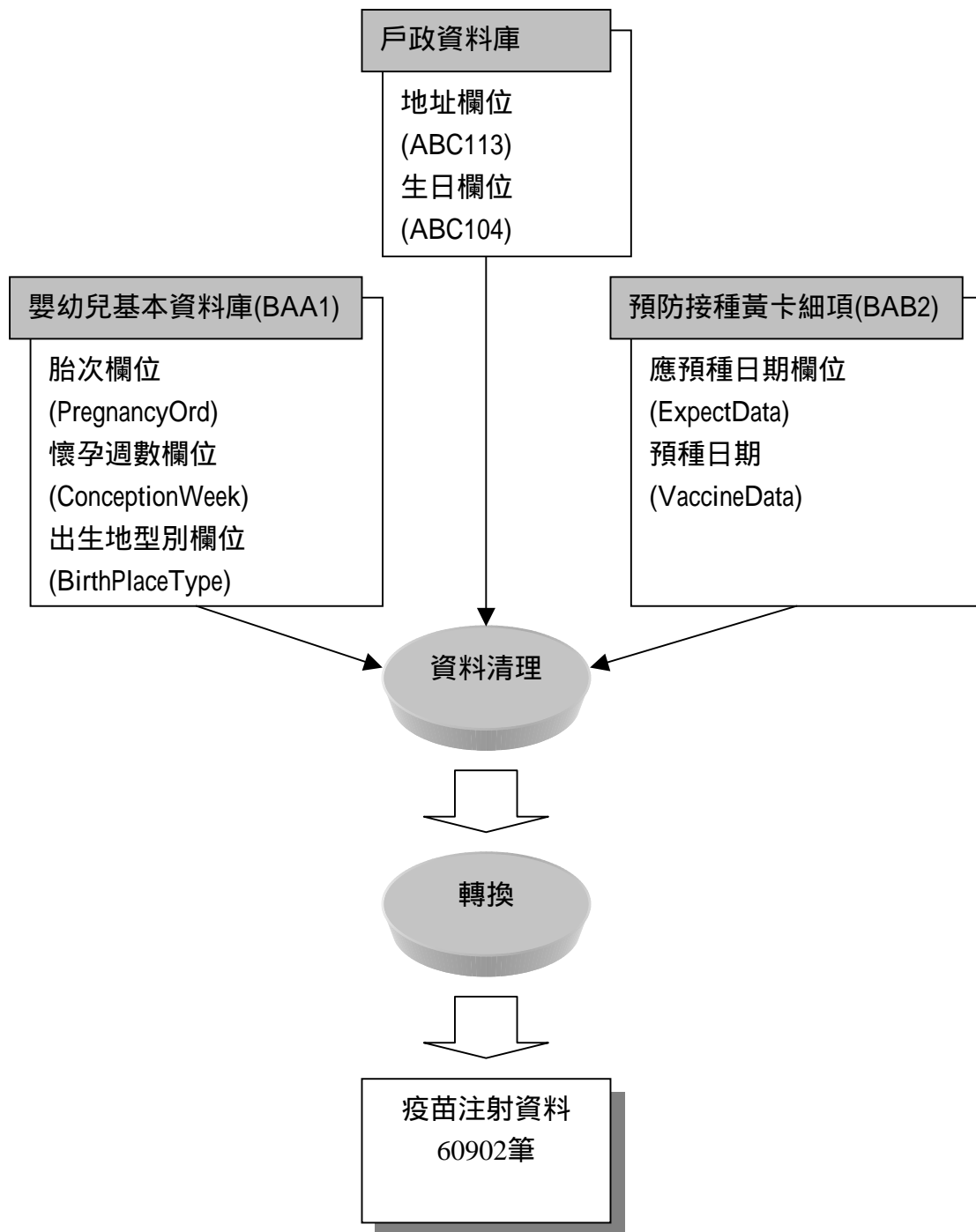
- 中斷病人持續性的照顧
- 降低原有疫苗的效用
- 增加疾病發生的風險
- 疫苗管理不易

失約的問題不但會影響醫療院所的成本計算及資源的浪費，對於病患本身也相對的影響，也是相當大的。因此本研究在於利用資料探勘的分群的演算法，對於失約顧客作分群，進而能提供政策制訂者產生更有效催注的方法，減少失約的現象。

本研究主要利用 IBM 的資料探勘工具進行分析，利用 IBM 之 intelligenceMiner 作為分析的工具，而資料來源主要是 NIIS 資料庫中的資料，分析對象主要是以桃園市、桃園縣復興鄉、嘉義市東區及嘉義縣中埔的卡介苗接種資料為主。

在分析過程中主要分為幾個步驟進行，第一個步驟就是資料萃取，第二步驟就是所謂的資料清理，第三步驟就是直接進行分析，最後就是解釋分析所得的分析結果。

在資料萃取的步驟中，我們主要是將儲存在 SQLSever 中的資料轉換成 IBM IntelligenceMiner 所符合的資料格式，我們先是將我們所要求的資料從既有的資料庫中取的，主要是在 BAB2（預防接種黃卡細項）中取的所有單一疫苗失約嬰兒的母親資料進行分析，我們主要的考量是以母親作為分析的出發點，因為我們認為嬰兒本身不具有行為能力，而親屬本身才是影響失約與否的主要關鍵，但由於大部分的接種都是以母親為主，接種記錄也是以母親為主，因此，我們於是選擇母親與嬰兒相關的欄位作為分析的重點。



圖四十、分群資料來源

經過主觀分析之後，我們在資料庫中選擇了五個欄位作為分群的依據，並將欄位作為合理的轉換，其中我們將母親的生日先計算成年齡，並將其區分為五組，將地址轉換成居住地區代碼，利用懷孕週數轉換成正常或異常兩類（詳見附錄二）。

表十八、欄位轉換表

表格名稱	欄位名稱	欄位內容	轉換內容
ABC	ABC104	出生日期	年齡層
ABC	ABC114	地址	居住地代碼
BAA1	PregnancyOrd	胎次	胎次
BAA1	ConceptionWeek	懷孕週數	懷孕類型
BAA1	BirthPlaceType	出生地型別	出生地型別
BAB2	VaccineDate	預種日期	是否失約
	ExpectDate	應預種日期	

將所需欄位經過整理及轉換後進行資料分析的動作，所得的結果如下：我們可以看見最大群主要失約特徵。

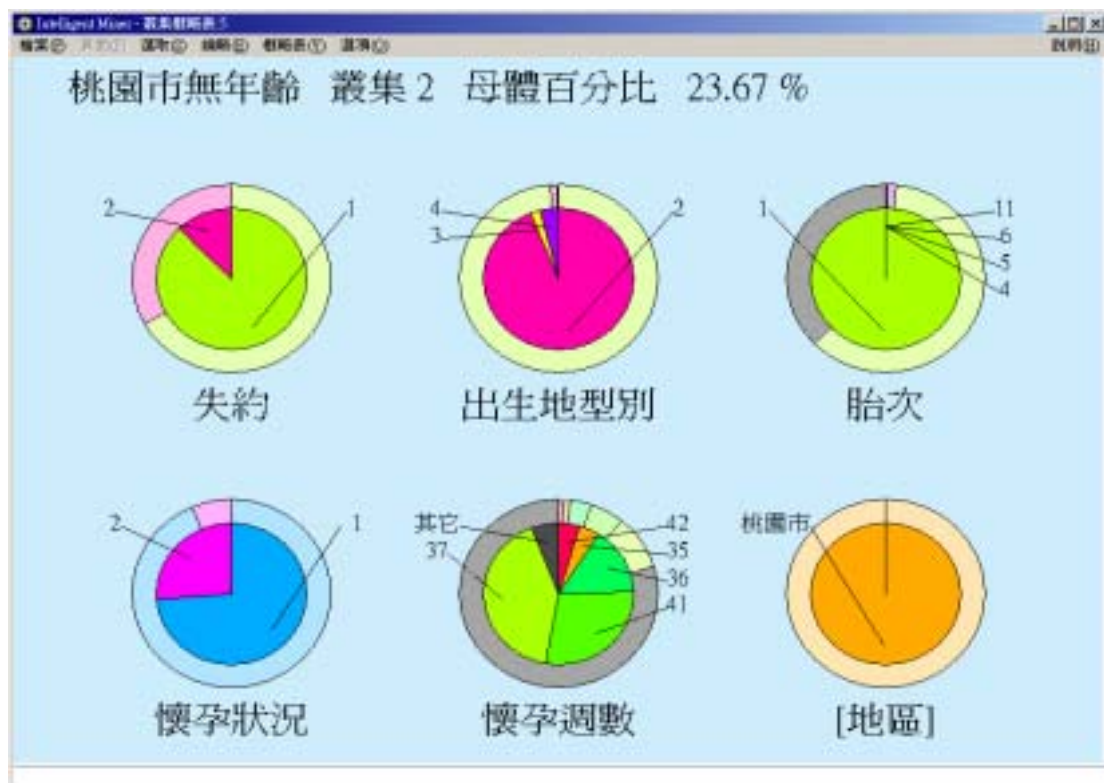


圖四十一、最大群失約叢集

根據分析出的結果指出，在所有的資料中，有百分之 23.05 的資料有著相同的特性，而這些特性主要是以圓餅圖的方式呈現，圖中內圈所涵蓋的區域表示此

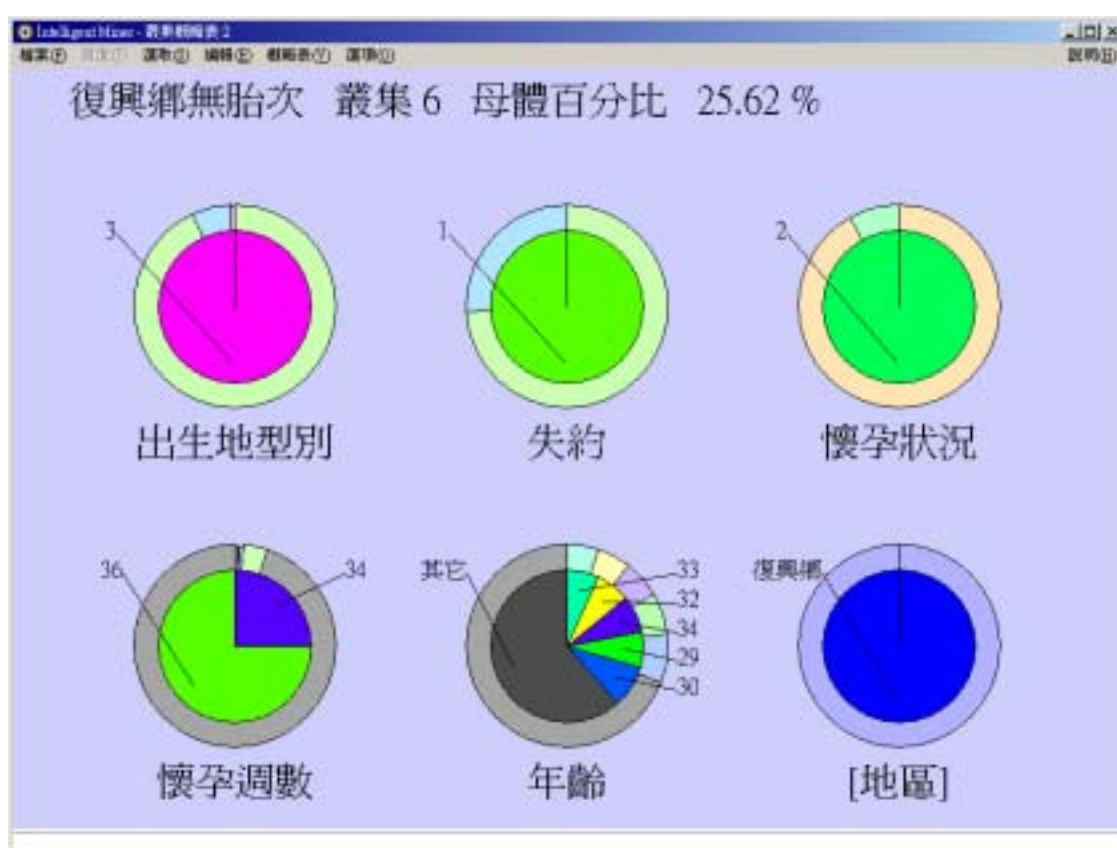
群組內資料的較顯著數值佔整個群組的比率，相對於外圍以整體資料的比例，可以互相察看，而第一群的特徵主要是以失約的資料為主，也就是表示此群為失約率高的群組，該群的主要特性包括懷孕狀況為正常，出生地型態大多為在診所出生，胎次為一胎，以居住桃園市為主，形成這群組中較明顯的特徵。而這些特徵值中，大多數並無法從中瞭解些什麼，因此，我們利用統計分析的方法，找出各個地區與我們選定的欄位間的顯著情況為何，我們在利用較為顯著的屬性，在進行分群。

(1) 在桃園市與屬性間的顯著關係下，我們發現除了年齡層較不顯著之外，其他的屬性皆達到顯著水準，因此，我們就利用其他的屬性進行分群。根據分析出的結果指出，在桃園市的資料中，有百分之 23.89 的資料有著相同的特性，而這一群的特徵主要是以失約的資料為主，也就是表示此群為履約率低的群組，懷孕狀況為正常，出生地型態大多為在診所出生，胎次為一胎，形成這群組中較明顯的特徵。



圖四十二、桃園市最大群叢集

(2) 在桃園縣復興鄉與屬性間的顯著關係下，我們發現除了胎次較不顯著之外，其他的屬性皆達到顯著水準，因此，我們就利用其他的屬性進行分群根據分析出的結果指出，在桃園縣復興鄉的資料中，有百分之 25.62 的資料有著相同的特性，而這一群的特徵主要是以失約的資料為主，也就是表示此群為履約率低的群組，懷孕狀況為異常，出生地型態大多為在診所出生形成這群組中較明顯的特徵。



圖四十三、復興鄉最大群叢集

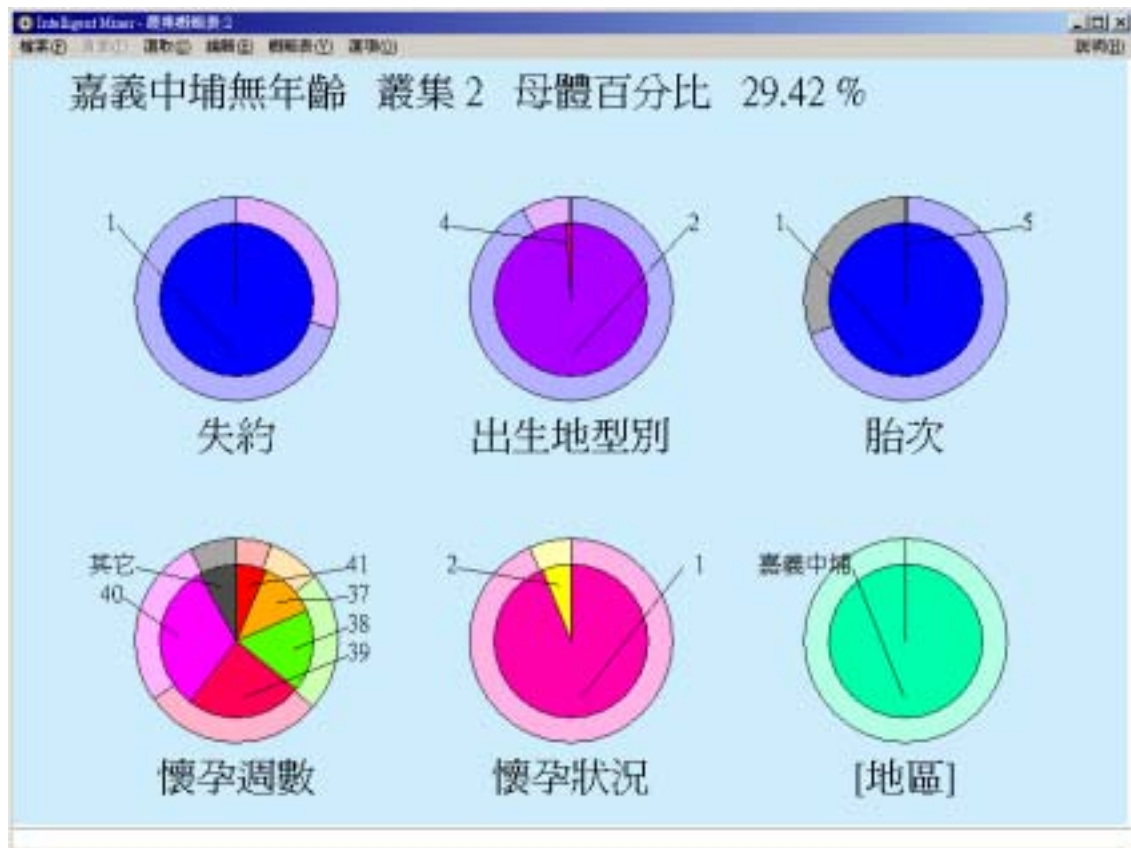
(3) 在嘉義市東區與屬性間的顯著關係下，我們發現除了所有的屬性皆為顯著，因此，我們就利用所有的屬性進行分群根據分析出的結果指出，在嘉義市東區的資料中，有百分之 62.16 的資料有著相同的特性，而這一群的特徵主要是以失約的資料為主，也就是表示此群為履約率低的群組，懷孕狀況為異

常，出生地型態大多為在診所及自宅出生形成這群組中較明顯的特徵。



圖四十四、嘉義東區最大群叢集

(4) 在嘉義縣中埔與屬性間的顯著關係下，我們發現除了年齡的屬性不顯著之外，其他的屬性皆為顯著，因此，我們就利用除了年齡之外的屬性進行分群。根據分析出的結果指出，在嘉義縣中埔的資料中，有百分之 29.42 的資料有著相同的特性，而這一群的特徵主要是以失約的資料為主，也就是表示此群為履約率低的群組，懷孕狀況為正常，出生地型態大多為在診所出生形成這群組中較明顯的特徵。



圖四十五、嘉義中埔最大群叢集

透過分群分析的研究，我們利用歷史預種資料，將卡介苗的接種記錄分區找出代表性高的群組，利用各群組的特性我們可以針對每一群利用不同的政策進行管理，但式分群的研究並非以預測為主要的出發點，而是一種以敘述資料特性為主的方法論，因此，若要進行預測的工作，應該尋求其他的方法（例如：類神經、決策樹），而透過分群的演算法期望能對於管理者能對資料特徵有深入的瞭解。

透過桃園市、桃園縣復興鄉、嘉義市東區及嘉義縣中埔的資料分析，我們發現出生地型態會是失約與否的影響因子，因為根據四各地區最大群的資料特徵中，都能明顯的發現出生地型別都是屬於診所，但是在全部的資料中非醫院出生的比率站母體的百分之 35 而已，卻是影響著失約特徵，而最後我們利用統計分析 crosstabulation 的結果，也發現失約與否與出生地型別的卻是呈現顯著的不同。因此，透過目前的分析結果，我們發現出生地型別的因素的卻會影響著名眾疫苗施打的失約與否，因此我們建議在催種政策的實行上，可以加強診所的催種

政策，以期催種成效能夠有較好的提升。

柒、結論

隨著健康醫療與生物科技的快速發展，數位化的醫療資訊技術也相對的呈現長足的進步，這樣的趨勢相對的也促進國內醫療發展，尤其在日新月異的資訊科技催化之下，醫療體系的資訊化也就因此如火如荼的進行中，其中醫療領域業務和網際網路功能的整合更是有長足的進步。網際網路的特性是任何人在任何時間及地點都能快速的取得和交換所要的資訊，這也帶動了醫療資訊快速交換的需求。現時最為普遍採用的電子資料交換機制在歐美先進國家也已有採用於醫療的建設發展之上。

要能快速達到交換的目的，我們必然需要去制定統一的標準以及共識，並且整合整個醫療體系間各個機構，所以本報告引入醫療上重要訊息交換標準 HL7，並且探討相關的議題。整個議題的面向涵括了資訊交換的標準及資訊安全相關議題，如加解密標準、密碼學的相關架構。其中很重要的是如何在 Web-based 平台下達成快速的訊息交換，卻又能保持資料的完整性、不可否認性及防止他人竊取私密性的個人健康資訊用於不正當的用途，而我們也根據上述要項統整歸納出了一個在網際網路上傳送醫療相關資訊的安全架構。

事實上，這些有關於隱私、安全及交換標準等議題，在美國的 HIPAA(Health Insurance Portability and Accountability Act)法案有相當篇幅的提及，而由於本研究的重點並不在此，所以相關資訊請參考相關網址 <http://www.aspe.os.dhhs.gov/admnsimp/>。同樣的在歐洲也有 SEISMED 這項法規在發展著。值得醫療及公共衛生主管機關的重視。

此外，為了呼應本計畫中，建立一個供全國性預種統計分析作業的資訊化管理功能及國內各相關研究單位所需的預種資料庫之需求，本報告亦探討了系統資料庫/資料倉儲最適模式，內容則含蓋了建立的模式、建立的目標及其可能的應用。在全國性預種統計分析作業的資訊化管理功能方面，本研究利用設計問卷及訪談使用者的方式，造模(modeling)了系統在作業流程及資料輸出二方面的需

求。

本研究強調將預防接種日常作業所得的各式資料經過萃取、淨化、以及彙整建立成預種資料倉儲，配合線上分析處理，及各式分析工具輔助預種決策的制訂上，並針對歷史性分析，預測未來趨勢。

現有全國性預防接種資訊管理系統之整合型中央資料庫之建構乃以整合型之應用預種資料庫與統計分析用之資料倉儲為主。以作業性線上資料庫而言，除了允許多人同時修改資料之一般性功能外，全面化疫苗批號追蹤、預防接種歷史資料查詢與建置、個人現住地資料登錄、疫苗庫存管控與單一窗口之預種證明申請等便民服務，讓民眾透過友善地介面進一步快速地得到個人化相關資訊。以統計分析性之資料倉儲而言，透過快速地查詢與統計效能，配合決策支援系統的彈性取樣，產生預防接種管理工作所需的預種報表，除了可使疾病管制局預種管理單位即時了解預防接種相關資訊，亦可便於提供學術單位取得預種統計分析資料，此一建置之經驗更可作為實務界進行類似建置之參考。

本研究的貢獻，除了提出了整合了電子資料交換和資訊安全機制的醫療傳輸模型，並針對敏感性隱私資訊在網際網路上傳送所需關切到的安全傳遞問題，提出相關的解決方案之外，期待本報告的粗淺介紹能提供 CDC 一個較為完整的觀點來看待醫療資訊傳遞安全議題。S/MIME 及 PGP/MIME 這兩個在網際網路上相當流行的安全郵件機制相當可能成為整合 HL7 的重要機制；另外，SSL 協定本身是處於應用層與傳輸層間的協定，與定義於應用層基礎上的 HL7 標準在整合上也將不會有太大的問題，而最後在安全機制的選擇上則可根據醫療資訊內容的特性，選擇適性化的方式來符合不同的需求；

同時在整個計畫中，中央資料庫及中央資料倉儲的規劃上，針對中央資料庫及中央資料倉儲的規劃，本報告提出建立一中央資料庫以之為中央資料倉儲資料來源的架構；系統功能方面，在作業流程上的需求是中央報表管理、中央戶政管理及中央傳輸代理等三項，資料輸出需求則計有整合型預種資料、戶政資料、預防接種嚴重不良反應及疫苗可預防疾病。以決策輔助雛形系統的展示，針對疫苗

失約特徵與疫苗管理機制，對於醫療資源的充分利用有直接的影響力，以期望利用圖形化的模式清楚的展示出醫療決策輔助系統的可行性與有用性，希望能利用資訊的角度去解決現存的問題，因此，冀望以上的研究成果，能對行政院衛生署疾病管制局在對傳送醫療資訊的安全觀點上及整合型中央版預防接種系統最適模式之規劃上能夠有所助益。

附錄

附錄一：訪談記錄

日期：8月29日 PM 2:00-3:00

地點：疾病管制局

訪談內容：了解預防接種組的疫苗管理相關業務流程

疫苗管理	預防接種組對於疫苗管理流程及辦法為何？ (ex. 疫苗運送路線配置、各地方即時疫苗庫存查詢、衛生局所疫苗調撥作業及疫苗廠商批號管理)
以疫苗管理來看，預防接種組著重庫存、消耗。 預防接種組將疫苗區分不同單位負責，各自管理。	

疫苗備用量 建議採購量	1. 以何標準判定？ 2. 單劑量或多劑量在備用量及採購量上有何不同？
疫苗備用量與採購量判定標準複雜。根據經驗是以1年量，分批交貨。若出生率有增減，也會隨之調整。 各疫苗的可能使用量不同，所以備用量與採購量也會不同。 單劑量與多劑量的成本有差，單劑量比多劑量貴，但好控管。	

疫苗配給量	根據各個衛生局、衛生所與合約醫療院所使用疫苗的狀況，中央如何規劃疫苗配給量？
預防接種組將採購的疫苗直接交給各地局來負責主要配給。 CDC 局 所 醫院（合約醫療診所） 局依所需求給予疫苗配給量，所在依醫院需求給予疫苗配給量。 假使 A 所疫苗有缺，必須回報給局，局再由 B 所調配給 A 所。 局以下醫療單位不可未經回報給局而私下互相配給疫苗。	

疫苗庫存	全國及地區庫存分布，對於通用、撥入、撥出...等疫苗管理方式有任務分配嗎？
CDC 無庫存，交由各衛生局配給。 但 CDC 負責施打國際疫苗（出/入境），在 CDC 一樓及各機場港口分局。	

回報疫苗現況	各地方衛生局所回報疫苗現況包括各類疫苗之領用、撥入、撥出、損毀、結存量等狀況登錄給中央，而各地方衛生局和中央何單位來負責？
--------	---

<p>回報疫苗現況有固定格式，而各自彙整後上傳並產生報表（ex. 完成率）。 CDC 局 所 醫院（合約醫療診所） 固定每月 10 號前回報一次疫苗現況。</p>	
疫苗基本資料	<p>1. 目前疫苗基本資料的記錄有何其標準，所必備的基本格式為何？</p> <p>2. 有包括疫苗供應商(疫苗廠商)來源、相關日期嗎？</p>
<p>目前疫苗基本資料，預種組有特定的資料格式，有關疫苗的項目皆有記錄。 然而地方單位與預種組的記錄可能不同，造成資料整理上的困擾，所以希望由中央建立檔案，以避免不一致產生。（ex. 地方施打疫苗可能有自費的情形）</p>	

疫苗批號	<p>根據疫苗批號，可提供何種資訊或其它資料所需？ （ ex. 衛生所疫苗庫存報表、完成率、工作量 ）</p>
<p>完成率：指的是固定範圍（ range ）內的完成接種比率 （ ex. 統計 1 月到 6 月份中，戶籍地為板橋市，在 2 月到 4 月出生的人，B 肝疫苗完成接種的比率。 ）</p> <p>工作量：指衛生所這月疫苗打了多少。</p>	

疫苗劑別	<p>黃卡上的疫苗劑別除了記錄接種項目外，還可提供何種資訊？</p>
<p>完成率報表對於各疫苗劑別有詳細區別。</p>	

多合一疫苗	<p>1. 多合一疫苗有可能是後續調配的嗎？</p> <p>2. 接種多合一疫苗時，資料庫是將可混合的疫苗同時記錄更新資料筆，還是只記錄多合一疫苗的資料筆？</p>
<p>預防接種組所購買的疫苗皆為單一疫苗，不購買多合一疫苗，所以資料庫中也不會有多合一疫苗的記錄。 多合一疫苗是地方依需求自行購買，接種者自費。也因此疫苗回報時會與中央不同。</p>	

預種組與疫苗管理	<p>希望報表可改進的地方？</p>
<p>希望報表能整合的呈現（ ex. 消耗量報表，瓶數、可用劑量 ） 可依需求調整，介面方便操作。 轉介與歸戶的資料如何於報表上呈現。</p>	

訪談後結論：

1. 未來中央資料庫有關疫苗的報表分析，需著重庫存與消耗
2. 可觀察出生率的增減而指示疫苗的備用量與採購量需調整
3. 完成率計算需注意所取的分母，工作量指的是衛生所

4. 中央資料庫應至少每月更新疫苗現況
5. 中央資料庫應考慮到地方自費的疫苗，此項目預種組無記錄
6. 分析報表中應對疫苗劑別詳細區分
7. 由衛生局上傳給中央資料庫的疫苗現況是已經彙整過的地方資料，對各個衛生所與合約醫療院所使用疫苗的狀況無法個別分析
8. 單劑量與多劑量疫苗消耗狀況難以追蹤
9. 轉介與歸戶的疫苗資料記錄待解決

OLAP 與疫苗的分析建議：

1. 疫苗種類和所處區域衛生局的地域資料的交叉比對，產生出預種管理的項目
2. 疫苗種類與疾病發生的可能性作一比對，以預測與抑制各種傳染病的發生

附錄二：欄位轉換

1. 履約狀況

代號	狀況
1	準時赴約
2	失約

2. 年齡層

代碼	年齡區域
1	20 歲以下
2	20 至 30 歲
3	30 至 40 歲
4	40 歲以上

3. 地區代碼

代碼	代表區域
1	復興鄉三民村
2	復興鄉三光村
3	復興鄉長興村
4	復興鄉澤仁村
5	復興鄉奎輝村
6	復興鄉高義村
7	復興鄉華陵村
8	復興鄉義盛村
9	復興鄉霞雲村

10	復興鄉羅浮村
11	其他(大溪鎮、復興鄉、楊梅,新竹,台南)

4. 懷孕類型

代碼	類型	懷孕週數
1	正常	37-42 週
2	異常	37 週以下或 42 週以上

5. 疫苗種類

代碼	疫苗種類	代碼	疫苗種類
1	B 型肝炎免疫球蛋白	11	破傷風減量白喉混合疫苗
2	卡介苗	12	麻疹疫苗
3	白喉百日咳破傷風混合疫苗	13	B 型肝炎遺傳工程疫苗
4	小兒麻痺口服疫苗	14	德國麻疹疫苗
5	麻疹腮腺炎德國麻疹混合疫苗	15	白喉破傷風混合疫苗
6	B 型肝炎血漿疫苗	16	A 型肝炎疫苗(小兒三劑時程)
7	流行性感冒疫苗	17	白喉破傷風非細胞性百日咳混合疫苗
8	A 型肝炎疫苗(小兒兩劑時程)		
9	不活化小兒麻痺疫苗	18	B 型嗜血桿菌疫苗
10	日本腦炎疫苗	19	水痘疫苗

參考文獻

1. 沈兆陽，資料倉儲與 Analysis Services：SQL Server 2000 的 OLAP 解決方案，台北市：文魁資訊有限公司，2001。
2. 周建成等，行政院衛生署疾病管制局預防接種系統設計規格書，財團法人資訊工業策進會電子商務應用推廣中心，2001 年 8 月。
3. 周建成，行政院衛生署疾病管制局預防接種系統專案執行報告，財團法人資訊工業策進會電子商務應用推廣中心，2001 年 6 月。
4. 周政宏，神經網路-理論與實務，松崗電腦圖書資料股份有限公司，1995 年 12 月，初版。
5. 林志麟，建立區域整體資源規劃資料庫先期研究，臺電工程月刊，第 662 期，89 年 6 月，pp73-83。
6. 何文雄等，行政院衛生署整合性衛生所 PHIS 視窗版發展計畫書，財團法人資訊工業策進會電子商務應用推廣中心，2000 年 11 月。
7. 何文雄、周建成等，行政院衛生署疾病管制局預防接種系統設計規格書，財團法人資訊工業策進會電子商務應用推廣中心，2001 年 9 月。
8. 邱瑞科等，整合性衛生所資訊系統視窗版之預防接種子系統設計發展審查、全面性導入計畫設計暨建立全國性預防接種管理資訊系統之規劃與評估期末報告，輔仁大學資訊管理所，2002 年 2 月。
9. 曾守正、周韻寰，資料庫系統進階實務 第二版，台北市：儒林圖書有限公司，2000 年 4 月。
10. 曾守正、周韻寰，資料庫系統進階實務，台北：儒林出版社，1996 年
11. 曾中文、郭修暉，拉式供應鏈(s,Q)存貨政策管理之探討，逢甲大學工業工程研究所碩士論文。
12. 蕭凱文等，“SQL Server 7.0 OLAP Services 設計與實務應用”，台北市：華彩軟體，1999。

13. 蕭凱文等, Microsoft SQL Server 7.0 資料倉儲整合應用, 台北市: 華彩軟體, 1999。
14. 蕭凱文等, SQL Server 2000 OLAP Services 設計與實務應用, 台北市: 華彩軟體, 2000。
15. 中華民國行政院衛生署疾病管制局, <http://www.cdc.gov.tw/>, 2002 年。
16. C. J. Date 編著, 黃加佩譯, 資料庫系統概論 第六版, 台北: 儒林出版社, 1997 年 3 月。
17. 施威銘, SQL Server 2000 管理實務, 台北: 旗標出版公司, 2001 年 10 月。
18. 施威銘, SQL Server 2000 設計實務, 台北: 旗標出版公司, 2002 年 1 月。
19. 宋鎧、范錚強等, 管理資訊系統 第二版, 台北: 華泰文化事業股份有限公司, 2001 年 2 月。
20. 李孟厚, 醫院門診作業系統規劃之研究, 中正大學企業管理研究所碩士論文, 1993。
21. 許純君, 預測的原則與應用, 台灣西書出版社, 1999, 初版。
22. 陳盈達、陳振明, 前置時間變動下配銷需求規劃最佳存貨政策之研究, 中央大學工業管理研究所碩士論文, 1996。
23. 陳記成、安揚龍、邱垂昱, 供應鏈存貨模式在不確定因素下的採購決策支援系統, 台北科技大學工程管理研究所碩士論文, 2001。
24. 陳慶餘、李龍騰、邱泰源, 候診民眾對門診醫療服務的期望性認知之調查, 中華民國家庭醫學會雜誌, 1991: 1(3), pp.119-123。
25. 葉宏謨, 企業資源規劃製造業管理篇, 松崗出版社, 2000。
26. 趙明彥、李知行, 類神經網路為基底的彈性門診預約系統, 華梵大學工業管理研究所碩士論文, 2002。
27. 蔡碩倉, 台灣地區農會信用部金融預警評等系統之研究, 國立中興大學農經所, 博士論文, 1998。
28. 林俊宏, 分散式異動伺服器的設計與製作, 國立交通大學資訊科學研究所碩

- 士論文，1993年6月。
29. 倪正耀，分散式資料庫系統測試檯，國立交通大學資訊科學研究所碩士論文，1994年6月。
 30. 張明惠，中央控制的分散式資料庫管理系統，國立成功大學電機工程研究所碩士論文，1986年6月。
 31. 湯維民，分散式企業資訊系統網路化效能評估之研究，私立天主教輔仁大學資訊管理研究所碩士論文，2002年7月。
 32. 王新富，行動代理者在分散式資料庫資料擷取之應用，國立成功大學工程科學研究所碩士論文，2000年6月。
 33. 顏哲傑、張啟明、邱瑞科等，NIIS全國預防接種資訊管理系統，MIST國際醫學資訊研討會，台灣台北，2002年10月。
 34. 顏哲傑 張啟明 邱瑞科等，企業資料倉儲化之決策應用-以NIIS為例，(CSIM)第八屆國際資訊管理研究暨實務研討會，台灣高雄，2002年11月。
 35. 唐淑娟，台灣地區鳳梨零售價格預測之研究 - 灰色預測、類神經網路與預測組，屏東科技大學農企業管理研究所碩士論文，2000。
 36. 江金山，吳佩玲等，灰色理論入門，高立圖書有限公司，1993。
 37. 池麗香、張銀益，電子商務環境下物流中心商品補貨模式之技術研究，輔仁大學資訊管理研究所碩士論文，2002。
 38. Adiba, M. E. and Lindsay, B. G., Database Snapshots, Proceedings 6th International Conference on Very Large Data Bases, Montreal, Canada, October 1980, pp.86-91.
 39. Albrecht J., Hummer W., Lehner W., and Schlesinger L., Query Optimization by Using Derivability in a Data Warehouse Environment, Communications of the ACM, November 2000, pp.46-56.
 40. Berry, M.J.A., Linoff, G., Data Mining Technique for Marketing, Sale, and Customer Support, Wiley company May 1997

41. Chaudhuri S. and Dayal U., An Overview of Data Warehousing and OLAP Technology, Communications of the ACM, Vol 26, 1997, pp.65-74.
42. Chen, P.Y. and Srivastava J., Warehouse Creation-A Potential Roadblock to Data Warehousing, IEEE Transaction on Knowledge and Data Engineering, Vol 11, February 1999, pp.118-126.
43. Connolly, T. and Begg, C., Database Systems: A Practical Approach to Design, Implementation and Management, 2nd edition, Addison-Wesley, 1998.
44. Dina Bitton, Turbyfill, Benchmarking Database Systems A Systemic Approach, <http://citeseer.nj.nec.com/bitton83benchmarking.html> 1983,
45. D. Julong, Control problems of grey system, System and Control Letters, pp.288-294, 1982.
46. D. Julong, Grey system and agriculture, Agriculture Science in Shanxi, pp.293-297, 1989.
47. D. Julong, Introduction to grey system theory, The Journal of Grey System, pp.1-24, 1989.
48. Franklin, M. J. et al., Transactional Client-Server Cache Consistency: Alternatives and Performance, ACM Transactions on Database Systems, Vol. 22, No.3, September 1997, pp.315-363.
49. George M. Marakas, Decision Support Systems in the Twenty-First Century, Prentice-Hall, Inc, 1999.
50. Gray, J., The Benchmark Handbook for Database and Transaction Processing Systems, Morgan Kaufmann INC. 1993.
51. Guape, F.H., Owrang, M.M., Database Mining Discovering New Knowledge

- and Cooperative Advantage, Information Systems Management, Vol.12, Fall 1995, pp.26-31.
52. J.K. Lee, and P.R. Larry, Operations Management: Strategy and Analysis, Fifth Edition, 1998.
53. Lindsay, B. G., et al., A Snapshot Differential Refresh Algorithm, ACM Proceedings SIGMOD, Vol. 15, No.2, 1986, pp.53-60.
54. Lyon, J. K., Database administrator, Wiley Interscience, New York, 1976.
55. Navathe, S. B., Evolution of Data Modeling for Databases, Communications of the ACM, Vol. 35, No. 9, September 1992, pp.112-123.
56. Novak, D. and Fry, J., The State of the Art of Logical Database Design, Proceeding 5th Texas Conference Computer-Systems(IEEE), Long Beach, California, 1976, pp.30-39.
57. Olston, C. and Widom, J., Best-Effort Cache Synchronization with Source Cooperation, ACM Proceedings SIGMOD, June 2002, pp.73-84.
58. Ozsu M. Tamer and Patrick V., Principles of Distributed Database Systems, 2nd edition, Prentice Hall International, Inc., 1999.
59. Ponniah P., Data Warehousing Fundamentals - A Comprehensive Guide for IT Professionals, Wiley Computer Publishing, 2001.
60. Roger, K. and Dennis, M., A Database Design Methodology and Tool for Information Systems, ACM Transaction on Office Information Systems, Vol. 3, No.1, January 1985, pp.2-21.
61. S. Chopra, P. Meindl, Supply Chain Management: Strategy, Planning and Operations, Prentice Hall, 2000.
62. Stephen R. Gardner, Building the Data Warehouse, Communications of the ACM, Vol 41, September 1998, pp.52-60.

63. Storey, V. C., Understanding Database Design Expertise, Data & Knowledge Engineering 16, March 1995, pp.97-124.
64. Teorey, T. J. and Fry, J. P., The Logical Record Access Approach to Database Design, ACM Computing Surveys, Vol. 12, No. 2, June 1980, pp.179-211.
65. The Benchmark Handbook, <http://www.benchmarkresources.com/handbook/>.
66. Thomas C. and Carolyn B., Database Systems: A Practical Approach to Design, Implementation and Management, 2nd edition, Addison-Wesley, 1999.
67. Tamer, M. and Valduriez, P., Principles of Distributed Database Systems 2nd, Prentice-Hall Inc.1999.
68. TPC-A, Transaction Processing Performance Council, TPC Benchmark A, Standard Specification, March 1992.TPC-A Standard Specification, 1990.
69. TPC-B, Transaction Processing Performance Council, TPC Benchmark B, Standard Specification, March, 1992.TPC-B Standard Specification, 1990.
70. TPC-C, Transaction Processing Performance Council, TPC Benchmark C, Standard Specification, August, 1992.
71. TPC-D, Transaction Processing Performance Council, TPC Benchmark C, Standard Specification, April, 1995.
72. TPC-H, Transaction Processing Performance Council, TPC Benchmark H, Standard Specification, Draft, 1998.
73. TPC-R, Transaction Processing Performance Council, TPC Benchmark R, Standard Specification, Draft, 1998.

74. TPC-W, Transaction Processing Performance Council, TPC Benchmark W, Standard Specification, Draft, 1999.
75. Traiger, I. L., et al., Transactions and Consistency in Distributed Database Systems, ACM Transactions on Database Systems, Vol. 7, No. 3, September 1982, pp.323-342.
76. Wang, H., Orłowska, M. and Liang, W. Efficient Refreshment of Materialized Views With Multiple Sources, ACM Proceedings SIGIR, 1999, pp.375-382.
77. W.H. Inmon, Building the Data Warehouse, 2nd edition, Wiley Computer Publishing, 1996.