

應用資料探勘技術於智慧型電腦整合製造系統建構之研究- 以 IC 封裝測試業為例

研究生：黃振聲

指導教授：陳瑞順 博士

國立交通大學資訊管理研究所

摘要

本研究是應用資料探勘技術來將企業推向商業智慧化系統架構，建構一個有特定功能導向的電腦整合製造系統(Computer Integrated Manufacturing, CIM)。在本研究中詳細介紹一個智慧型電腦整合製造系統，這個系統中所整合了五個主要的技術分別是：電腦整合製造、資料倉儲、線上分析、資料探勘和人工智慧。

本系統架構是以 ERP 系統所積蓄的既有龐大資料為基礎，經由整合 CIM 系統與資料倉儲、整合資料倉儲與決策分析、整合決策分析與資料探勘系統等三階段的過程，實際建置涵蓋操作資料、倉儲資料、智慧應用等三個層級的系統架構。資料探勘系統則是應用決策樹演算法與分類預測的模型，以探勘 CIM 系統中隱含的、有意義的決策資訊。探勘到的知識規則，則以專家系統的法則式知識表現，搭配 WEB 化方式來展示。

本系統所提出的系統架構也可應用於傳統製造業，藉由行銷管理、目標市場分析、顧客關係管理等方面，協助企業提昇競爭力。

關鍵字：資料探勘，資料倉儲，決策樹，電腦整合製造。

Using Data Mining Technology to Build an Intelligent CIM System- A Case Study of IC Testing House

Student : Huang Chen-Sheng

Advisor : Dr. Ruey-Shun Chen

Institute of Information Management
National Chiao Tung University
Hsinchu, Taiwan, Republic of China

ABSTRACT

This paper is applied data mining technology to use a systematic framework of business intelligence to bring a specific CIM system. This paper explains an intelligent CIM system by integrating the following five major subject areas: computer integrated manufacturing, data warehouse, online analytical processing, data mining, and artificial intelligence. The framework of this system is provided by the massive amounts of data gathered by a ERP system. Through a three-step process of integrating CIM systems, data warehouses and decision analysis with data mining technology, a three-tiered systematic framework has been established. The three tiers of this practical framework consist of operational data, warehouse data, and intelligent application.

The data mining system makes use of the decision tree algorithm and classification model in exploring the meaningful information, which is useful in the process of decision-making. Subsequently, the rules discovered by the data mining system are expressed through the rule based knowledge presentation method of the expert system. These rules are exhibited by web-based framework.

The contribution can be help increase the business competitiveness, the analysis of customer characteristics, target market analysis, the analysis of customer loyalty, predicting the loss of customers and the segmentation of customers.

Keyword: Data Mining, Data Warehouse, Decision Tree, Computer Integrated Manufacturing.

誌謝

首先要感謝的是我的指導教授陳瑞順博士，在老師不辭辛勞地諄諄教誨之下，讓我能掌握本論文研究的方向與方法，並能整合與應用各個學域的理論知識。當面臨瓶頸而亟需突破時，老師傾囊相授的提示與指導，更是讓我能順利地完成研究與撰寫本論文的關鍵。

其次要感謝口試委員劉錦松教授、黃柏翔教授的指導，讓本論文的內容能更臻於完整。

另外要感謝公司 IT 的同仁，由於您們的努力將理論與實際得以整合與應証，並於專案推廣期間主動的克服困難，從實務上獲得寶貴的經驗與資料，讓本論文能夠在應用面上更具深度與價值。

最後要感謝我的內人，由於您的鼎力支持與悉心照顧小孩，才能讓我毫無顧慮地專心研究與撰寫本論文。



目錄

中文摘要	1
英文摘要	2
誌謝	3
目錄	4
表目錄	6
圖目錄	7
第一章 緒論	
1.1 研究背景	8
1.2 研究動機	9
1.3 研究目的	9
1.4 論文架構	11
第二章 文獻探討	
2.1 資料倉儲	12
2.1.1 資料倉儲的定義	12
2.1.2 資料倉儲的內涵	13
2.1.3 資料倉儲的架構	15
2.1.4 線上分析處理(OLAP)	17
2.2 資料探勘	19
2.2.1 資料探勘(DM)概論	19
2.2.2 資料探勘技術	21
2.2.3 資料探勘模組	23
2.3 IC 封裝與測試之生產流程概述	25
2.3.1 IC 封裝型態及製程	25
2.3.2 IC 晶圓測試製程	27
2.3.3 IC 成品測試製程	28
第三章 電腦整合製造問題定義	
3.1 電腦整合製造	32
3.1.1 電腦整合製造概述	32
3.1.2 半導體電腦整合製造	34
3.2 電腦整合製造遭遇的問題	35
第四章 智慧型電腦整合製造系統	
4.1 智慧型電腦整合製造系統設計架構	37
4.2 整合智慧型電腦整合製造系統與資料倉儲	39
4.2.1 建立資料倉儲系統	39
4.2.2 整合 CIM 系統與資料倉儲模式	40
4.3 整合資料倉儲與決策分析	42

4.3.1 設計資料倉儲的綱要	42
4.3.2 設計事實資料表	42
4.3.3 設計維度資料表	46
4.3.4 設計彙總資料表	47
4.3.5 設計中繼資料表	48
4.3.6 設計多重維度綱要	51
4.4 整合決策分析與資料探勘系統	52
4.5 設計資料探勘系統	53
4.5.1 資料分類與預測功能的處理流程	53
4.5.2 設計資料探勘引擎	54
第五章 智慧型 CIM 系統之實作與分析	
5.1 智慧型 CIM 系統之環境與架構	57
5.2 結果比較與分析	57
5.2.1 結果比較	58
5.2.2 效益分析比較	58
第六章 結論與未來研究方向	
6.1 結論	60
6.2 未來研究方向	61
參考文獻	62



表目錄

表一 資料倉儲資料層面意涵	14
表二 資料倉儲應用程式層面意涵	14
表三 資料倉儲分類表	15
表四 Data mining 分析工具	34
表五 CIM 功能需求表	37
表六 系統元件的設計概念與功能表	48
表七 應用於資料的轉換與載入的中繼資料架構	49
表八 應用於資料管理的中繼資料架構	50



圖目錄

圖一 資料倉儲的架構	16
圖二 OLAP 系統各元件關係圖	18
圖三 知識發現流程圖	21
圖四 資料探勘模組	25
圖五 積體電路(Integrated Circuits, IC)的製造過程	27
圖六 TSOP 封裝製程	28
圖七 晶圓測試製程	31
圖八 成品測試製程	33
圖九 CIM 基本功能示意圖	34
圖十 智慧型電腦整合製造系統架構設計圖	38
圖十一 整合 CIM 系統與資料倉儲的架構	41
圖十二 四重維度的資料立方體	51
圖十三 資料探勘演算步驟流程圖	56
圖十四 智慧型 CIM 系統環境架構圖	58

第一章 緒論

1.1 研究背景

廿一世紀，全球半導體產業更邁向「既競爭又合作」的時代，如何運用既有資源，並整合內部資源以建構新的競爭優勢，以因應產業版圖的重整、與經濟環境的改變，已成為半導體產業價值鏈重要的一環。其中資訊的有效整合將扮演著關鍵性的角色。國內的企業均擁有很好的資訊與網路的基礎建設，這些建設都有助於公司內部或企業集團間的資訊溝通，但是對於公司內主管作關鍵性決策時，卻總覺得想當做參考的資訊一直都不夠齊全，而無法做正確的分析與預測，常造成問題反應過慢與決策品質不佳等諸多問題，而傳統之管理資訊系統卻難以提出有效的解決方案。因此有效的整合資訊科技與演算法，並配合適當的決策模式，來彌補現有的管理資訊系統上的分析預測能力不足之處，將可協助企業提昇經營優勢。



然而當資訊應用系統需求日益增加，造成資訊環境也十分複雜，且資料分布在多個不同的資料庫中，要整合這些資料的工作將是相當的困擾，當決策者需要這些資料來輔助決策時，決策分析將是非常沒有效率的，所以資料倉儲(Data Warehouse)系統的觀念因應而生，就在於解決上述的問題，資料倉儲提供集中式的資料管理，根據分析的需求來收集各個資料庫的資料，經過篩選整理後集中放在資料倉儲中，以增進決策分析時快速取得資料之效益。

產業界在近十年的資訊化過程中，了解到妥善利用資訊科技所帶來的績效提昇，將是激烈的競爭中不能缺少的一部分，因此相關的技術因應而生，例如決策支援系統(Decision Support System, DSS)、專家系統(Expert System, ES)、企業資源規劃系統(Enterprise Resource Planning, ERP)等，均可提供決策人員所最迫切需要的資訊或建議。但也因為資訊系統的繁雜，各自的系統

對於問題的分析 and 資訊的提供各有其價值，為了提供真正具有洞察能力的資訊分析，於是 OLAP(Online Analytical Processing)技術就因此而誕生，為了提升決策人員的資訊品質，不會因為採用資訊的偏誤，而造成企業的損失。

1.2 研究動機

資訊與知識是新經濟時代中重要的資源，Nonaka[1]認為在變動的經濟環境中，唯一可確定的是環境充滿了不確定因素，要掌握競爭優勢必須先掌握知識。在市場變化、科技不斷推陳出新、競爭者愈來愈多，及產品生命週期愈來愈短的環境下，成功將屬於能不斷地創新知識、廣佈新知識於組織當中、並快速地吸收新科技與推陳出新產品的企業。

隨著企業全面的實施電腦化，各種應用系統的功能也不斷的增加，產生的資料量也就愈來愈龐大。然而資料量愈大的時候，它的價值與數量的比值會愈小。但經由有效的萃取，所獲得的資訊和知識數量雖少，但它的價值卻相對地提高，它能提供決策的參考價值也相對的增加。面對這些愈來愈龐大的資料量時，要如何才能有效且快速的從中取得更有價值的資訊，成為作業上或決策上可用的參考，是很重要的挑戰。

因此運用近年來發展熱門的資料庫知識探索(Knowledge Discovery in Database；KDD)，搭配資料探勘的技巧，可從大量資料中發覺出潛在而有用的型樣或規則等整合而成的知識，提供企業營運或決策支援使用。讓企業在知識經濟時代裡，真正掌握知識、運用知識，以提昇企業競爭力[1]。

1.3 研究目的

提出一套應用資料探勘技術可自行建構符合 IC 封裝測試業本身需求且是即

時正確的智慧型電腦整合製造系統系統。

期望達成的目標：

1. 此系統架構使用 Data Warehouse，並整合運用 OLAP 及 Data Mining 技術。

藉由資料倉儲系統整合前端的分析工具(例如：OLAP、報表工具)，讓使用者依據本身需求來瀏覽資料，提供動態且即時的產生所需的報表，並能以多維的分析方式了解與呈現企業經營活動的各項資訊。後端應用資料探勘的技術與功能，設計一個資料探勘引擎，以適合於從資料倉儲中挖掘出潛在而有用的規則或樣式等知識。

2. 此系統架構可整合企業內多個異質資料庫。

藉由資料倉儲整合後端的各種資料庫，可將企業各種不同來源，不同型態的資料，利用有效的方式彙整(例如：擷取、轉換、載入)之後，有組織地排列並集中儲存於資料倉儲內，使資訊的掌握遠比傳統資料庫系統更為完整。這些整合後的資料，可以讓使用者利用各種分析工具，將資料做更進一步地詮釋。

3. 此系統採用 Web Base 方式開發設計。

藉著企業內部資訊網路的發達，利用 Web-based 的 three-tire 架構，突破空間與時間的限制，讓需要資訊的人能隨時隨地藉著網路得到所需的資訊。

4. 此系統可以提供即時正確的整合性資料給決策者

在這些大量的資料中，通常都會隱藏著有用的資訊或知識，藉由整合資料探勘引擎與資料倉儲，利用資料探勘各種分析模式，在大量資料的資料倉儲中尋找到這些不易發現，又極具意義與價值的資訊，讓企業掌控先機和優勢。

5. 期盼能改善生產力，穩定的管理設備性，及品管解析來提高良率。

藉由資訊的整合與知識的發掘，能利用此一系統架構將有限的資源做最有效的利用，藉著資料探勘技術找出最佳的生產機制，及整合性的

資料分析來輔助決策。

1.4 論文架構

本論文內容共分為六章，其撰寫架構略述如下：

第一章：緒論。說明本論文的研究背景、動機、目的、方法及整個研究流程。

第二章：文獻探討。回顧整理既有的理論與相關文獻，主要包含半導體封裝測試流程、電腦整合製造、資料探勘模組與技術之研究。

第三章：問題定義與演算法。定義本論文研究欲解決之問題所在，並說明所使用之演算法。

第四章：系統架構設計。運用第三章所提之演算法建置一新的資料探勘系統架構。

第五章：系統實作及分析。以一實際半導體封裝測試廠為實驗對象，用以說明整個系統的執行方法，並做效益的分析與檢討。

第六章：結論與建議。對本研究方法作個總結，並對未來研究方向提出建議。最後為有關論文之參考文獻整理

第二章 文獻探討

2.1 資料倉儲

2.2.1 資料倉儲的定義

Inmon[14]認為資料倉儲是一種整合的(Integrated)、主題導向(Subject Oriented)、時間變動性(Time Variant)、以及非揮發性(Nonvolatile)的資料集合，用來支援管理決策。說明如下：

1. 整合的：

資料倉儲是一個將分散的資料加以整理合併後的集中儲存所，為避免資料的解讀無法一致，它將一個企業分散四處及不同格式的資料，依照既定規則加以轉換、合併與彙總集中，以達到一致化的目標。

2. 主題導向：

資料是依據主題而被組織起來，資料倉儲中的資料可以根據不同的主題或觀點來組合、彙總，以達到如何在最短的時間、最有彈性的方式，滿足管理者對資訊的需求。

3. 時間變動性：

資料倉儲中存放的資料，包括目前的資料、短期的歷史資料、以及長期的歷史資料。而且所儲存的資料通常會包含一個時間元素，以利資料同期或是不同時期的比較分析。當資料定期地彙集到資料倉儲後，原本在資料倉儲中與該資料有時間相依關係的所有彙整資料都必須重新計算，以保持資料的一致性。因此，資料倉儲中的資料會隨著時間而變化的。資料倉儲的歷史資料是用來對照、趨勢與預測，這些是不被更新的。

4. 非揮發性：

在資料倉儲中的資料，只能定期地被加進去，僅供讀取而不可以依需

要隨時對它進行新增、修改或刪除。資料倉儲中的資料一經快照(Snapshot)後，就不允許再修改，而且因為所儲存的資料均為長時期的歷史資料，所以也不能隨意刪除。因此，資料倉儲的資料環境相對是比較穩定，只會增加而不會減少資料。

綜合上述的定義可歸納得知，資料倉儲乃是整合了分散在不同地方、不同作業環境下之資料庫的資料，整合進資料倉儲的資料將定期更新並不得任意更動。日常線上交易處理(OLTP)的資料經過擷取、清理、轉換、整併、載入後，統一儲存在倉儲資料庫中，讓使用者可以依特定主題導向，從各種不同的角度來使用資料。資料倉儲提供使用者可以更方便地存取不同的資料庫，使資訊的掌握遠比傳統資料庫系統更為完整，並能提供後續決策分析的需求。

2.1.2 資料倉儲的內涵



資料倉儲不僅包含了分析所需的資料，而且包含了處理資料所需的應用程式，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，也包括了將資料加以分析並呈現給使用者的應用程式[2]。

首先可以大致可分為資料層面、應用程式層兩個角度來說明。就資料層面而言，可分為事實資料、中繼資料、維度資料和彙總資料，下表一為資料倉儲在資料層面的意涵。其次就應用程式層面而言，資料倉儲包含了載入管理員、倉儲管理員和查詢管理員，下表二為資料倉儲在應用程式層面的意涵。

表一 資料倉儲資料層面意涵

分類	說明
事實資料 (Fact Data)	是最原始的資料，由 OLTP 系統轉入，能反應一項已經發生過的事實，例如：一筆訂單。而且它是不可變更的。
中繼資料 (Meta Data)：	描述整個資料倉儲系統各個部份的描述性資料。
維度資料 (Dimension Data)	為了增加查詢速度而建立的，它是可以變更的。
彙總資料 (Aggregation Data)	以精簡的方式來儲存使用者經常查詢的資料。

表二 資料倉儲應用程式層面意涵

分類	說明
載入管理員 (Load Manager)	自來源系統萃取資料、將萃取的資料快速載入暫存媒體、執行簡易的資料轉換工作。
倉儲管理員 (Warehouse Manager)	檢驗資料的關連性與一致性、將暫存媒體中的資料加以轉換並載入資料倉儲、新增索引或 view、資料分割、依需要進行資料反正規化、依需要產生新的彙總資訊、備份與備存資料。
查詢管理員 (Query Manager)	將查詢導引到正確的資料表、為所有使用者查詢排程。

資料倉儲當然有其資料來源，但是它的資料模式並不是來自終端使用者 (End User) 每天所輸入的資料。資料倉儲的資料來源包括下列兩項：

1. 企業的 OLTP 資料：

資料倉儲的資料大部份都是來自於本身所擁有的 OLTP 資料，但是處理上並不是以日為單位，資料倉儲資料的保存時效會比較長。此種保存資料的時距稱為資料倉儲時距，而新的歷史資料納入資料倉儲的週期時間稱為資料倉儲週期。

2. 由外部單位購買的資料：

企業可向外部來源購買資料，例如 IDC 或 Dataquest，以進行資料分析之用，

譬如分析明年個人電腦的全球性銷售量預估。

資料倉儲以其規模與應用層面來加以區分，大致可區分為標準資料倉儲、資料超市、多層資料倉儲和聯合式資料倉儲幾種，下表三為各種資料倉儲分類及簡單說明。

表三. 資料倉儲分類表

分類	說明
標準資料倉儲	以整個企業的需求為著眼點而建構出來，所以其資料都是有關整個企業的資料，使用者可以從中得到整個組織運作的統計分析資訊。
資料超市 (Data Mart)	針對某一個主題或是某一個部門而建構的資料倉儲。
多層資料倉儲 (Muti-tier Data Warehouse)	標準資料倉儲與資料超市的組合應用，由上層的標準資料倉儲提供資料給下層的資料超市。
聯合式資料倉儲 (Federated Data Warehouse)	包含多重的資料倉儲或資料超市，也可包括多層的資料倉儲，但只有一個資料倉儲資料的提供者。

2.1.3 資料倉儲的架構

資料倉儲的主要精神乃是將關聯式資料庫立體化，讓使用者可以依特定主題導向，從各種不同的角度來使用資料。目前的倉儲資料庫依其儲存資料的方式可分為關連式資料庫(Relational Database；RDB)與多維矩陣式資料庫(Multi-Dimensional Database；MDDB)，關聯式資料庫以行和列組成的表格方式來儲存資料(所以亦可視為二維矩陣)，而多維矩陣式資料庫則以矩陣的方式來儲存資料(可以是三維、四維等多維的矩陣)。以多維度(Multi-Dimensional)的資料型式[3]儲存於資料倉儲系統中，這種新的資料儲存結構，一般稱為資料立方體(Data

Cube)[4]。

我們可以把資料倉儲想像成是一種儲存各類整合過資料的超大型資料庫，只要我們事先訂定好它的篩選條件，就可透過事先的排程而自動的到指定的位置去搜集資料。換言之，資料倉儲系統可從多個分散式(Distributed)、自主性(Autonomous)及異質性(Heterogeneous)的資料來源中，搜集並維護相關的資訊。當資料的來源被修改時，對應的資料可由來源擷取、轉換成所需型態後，再與已存在的資訊互相整合[1]。

Chen[15]認為一個典型的資料倉儲架構是由下列幾項元件所組合：

1. 資料倉儲伺服器：儲存企業整體的大量資料。
2. 資料超市：儲存各部門的資料，依各部門所需的主題而定。
3. 後端工具：提供資料萃取、轉換、清洗、載入之功能。
4. 前端使用者工具：例如 OLAP 分析工具。
5. 中繼資料：描述其他資料的資料。

綜合以上說明，資料倉儲的架構可如圖 1 所示。

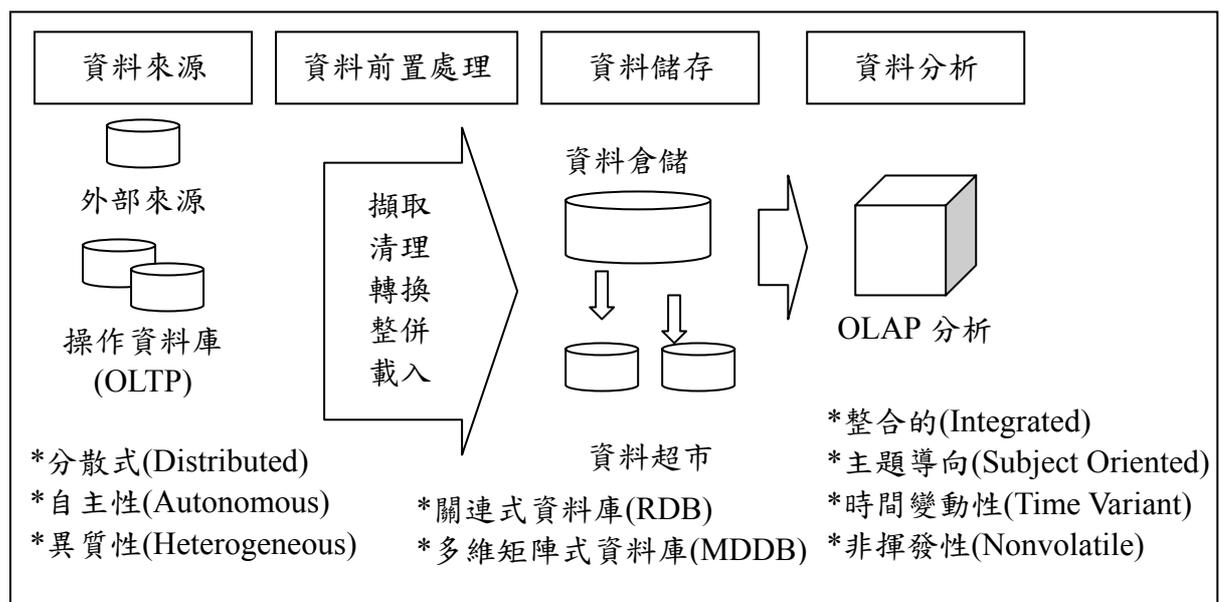


圖 1 資料倉儲的架構

2.1.4 線上分析處理(OLAP)

線上分析處理系統是用來幫助使用者，能有效率、且輕易完成商業資訊的多維結構分析工作，可將資料倉儲的資料，加以篩選、分析、彙總而產生實體資料，而以各種資料模式呈現給查詢者，讓使用者可以不同的主題和角度，依專業的直覺，透過複雜的查詢能力，資料比對、資料採掘和報表來提供不同層次的分析；透過 OLAP，當使用者有需求時，只需利用工具就能找到各種資料，例如管理角度的交叉分析、資料排名、預算及實際值的比較等，並不需要再排隊等待程式設計師撰寫程式，以簡單的操作方法，獲得想要查詢的資料[5]。

OLAP 為一種軟體技術，其目標為滿足決策支援或多維度環境特定的查詢和報表需求。它的技術核心為維度這個概念，正由於使用者可以從多維度來觀察和分析資料，因此又稱多維度分析。OLAP 也可以說是多維度資料分析工具的集合。OLAP 的功能結構是三層客戶伺服器結構，此結構由三個伺服器組件組成：即資料儲存伺服器、OLAP 伺服器、及用戶描述伺服器。其實體結構有多維資料儲存器及關連式資料儲存器。因而 OLAP 也就形成了多維度資料庫導向的 MOLAP 和關連式資料庫導向的 ROLAP。分別說明如下：

1. 多維度 OLAP：MOLAP 使用多維度資料庫管理系統來管理所需的資料或資料倉儲。各個 OLTP 資料庫中的資料，透過萃取、淨化、轉換與綜合等步驟後，提交給多維度資料庫。多維度資料庫依靠維度來形成超立方體結構(Hyper-Cube Structure)，而產生旋轉、上翻等操作。使用者透過客戶端的應用軟體介面，向 OLAP 伺服器提交分析要求，由 MOLAP 伺服器檢索 MDDB 資料庫，然後將得到的結果顯示給用戶。
2. 關連式 OLAP：ROLAP 以關連式資料庫為重心，在 ROLAP 結構中，用 RDBMS 來代替 MDBMS。ROLAP 將多維度資料庫的多維結構劃分為兩種列表：一類為事實資料表，用來儲存資料和維度關鍵字。另一種是維度

表，即對每個維度至少使用一個表來存放維度的階層、成員類別等維度的描述資訊。它是一種有組織的分類階層架構。例如 geography 維度可能包含 country、region、state 或 province 以及 city 層級。且兩者透過主鍵和外在鍵結合起來。使用者可透過客戶端的工具，向 OLAP 伺服器提交分析要求，OLAP 伺服器動態地將這些要求轉換成結構化查詢語言(Structural Query Language；SQL)執行，結果將以多維度視窗方式，展示給客戶。

OLAP 系統大致可分為以下幾個元件：

1. 資料來源: 資料庫與其他資料的傳統來源，包含了必須轉換到資料倉儲與資料超市的 OLAP 資料。
2. 中間資料中心: 合併的資料儲存區域與處理皆在此開發、淨化並轉換資料成為可用的 OLAP 資料。
3. 倉儲伺服器: 倉儲伺服器是指執行關聯式資料庫的電腦，此資料庫包含資料倉儲與資料超市的資料，以及管理 OLAP 資料的伺服器。
4. 商業情報: 一套工具與應用程式，可查詢 OLAP 資料並提供報表與資訊給企業的決策者。

各元件之關係如下圖 2 所示。

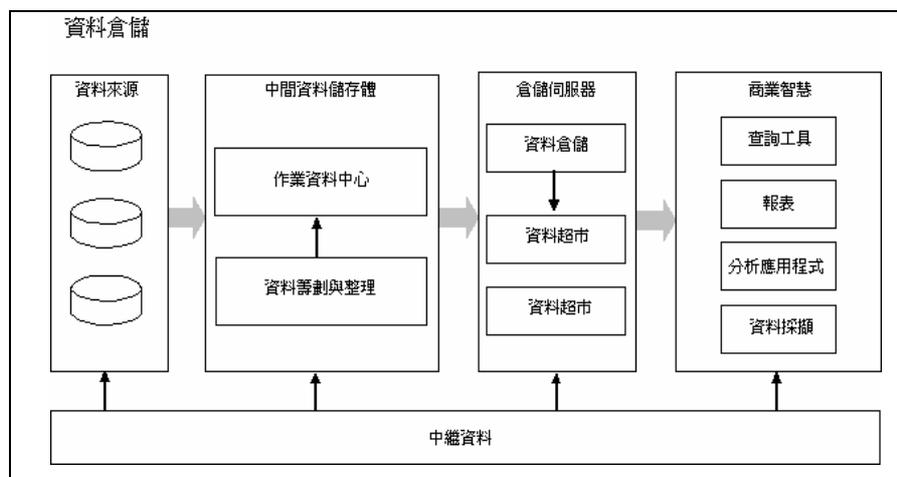


圖 2. OLAP 系統各元件關係圖

2.2 資料探勘

2.2.1 資料探勘(DM)概論

資料探勘是一個熱門的技術，資訊和知識從巨大的資料庫而來，它被許多研究者在資料庫系統和機器學習當作關鍵研究議題，而且也見件的被應用在企業體，當作主要利基的重要所在。有許多不同領域的專家，對資料探勘展現出極大興趣，例如在資訊服務業中，浮現一些應用，如在 Internet 之資料倉儲和線上服務，並且增加企業的許多生機。

資料探勘可以增加企業智慧，提昇企業競爭優勢，那到底應該如何進行呢？整個資料探勘過程中，是一個線性的過程，而這個過程我們也可以稱他為知識發現，在整個過程中，每個步驟都可返回，或是加入其它的步驟，這個過程並不是唯一的。這個過程包括了以下幾個步驟：

1. 資料選擇 (Data Selection)

在資料探勘的過程中，第一個步驟莫非要去選擇一個資料來探勘。

包括了選擇一個想要探勘的資料子集合，或是強調一個變數的子集合。

2. 資料清除 (Data Cleaning)

在資料庫中的資料不一定永遠是正確的、完整的，一定會包含些錯誤、遺失、不完整的資料，以上這些問題可用數學統計或模糊理論方法來推論與解決。此步驟包括要清除這些有雜訊的資料，並選擇對整個過程有需要的資料，決定一個策略去處理這些遺失資料。雖然資料探勘跟資料清除是兩個不同的學科，不過他們有很多相同之處，比如可用資料探勘的特徵辨識 (Pattern Recognition) 將演算法應用在資料清除上。

3. 資料減少與量化 (Data reducing and Coding)

從一個巨大的資料庫中去發現有用的資訊，是一件非常困難的事，所以必須適時縮減資料量。這包括依據任務目的不同，去發現有用的特

色而呈現的資料。使用多次元（Dimensionality）縮減、轉換或編碼的方法去減少有效的變數或資料。比如可把年齡一到一百歲編碼成一到十。或著錢的數量可以千元為單位等。

4. 選擇資料探勘演算法（Choosing the Data Mining Algorithm）

這步驟包括去選擇適當演算法，在資料中去尋找有用的特徵，比如決定模式和參數是否適當。還有在整個資料探勘的所有過程中，是否能選擇一個適合資料探勘的方法。

5. 資料探勘(Data Mining)

在整個過程中，最重要的步驟莫過於此。包括去探勘有用、有興趣的特徵或資料，以一個特別的形式呈現，包括了分類規則、決策樹、統計迴歸、群聚方法、線性分析等演算法。

6. 詳細報告（Reporting）

最後把這些探勘出來的特徵或資料，可用一些報告方法或圖形工具，轉換成可讓人輕易了解的圖示或報表，以提供決策支援（DSS or EIS）之用。

從六個步驟來看，資料探勘牽涉大量的規劃與準備，專家聲稱高達 80% 的過程花在準備資料階段，這包括表格的整合以及可能相當大量的資料轉換。從這個角度看，資料探勘只是知識發掘過程中的一個步驟而已，而達到這個步驟前還有許許多多的工作要完成。

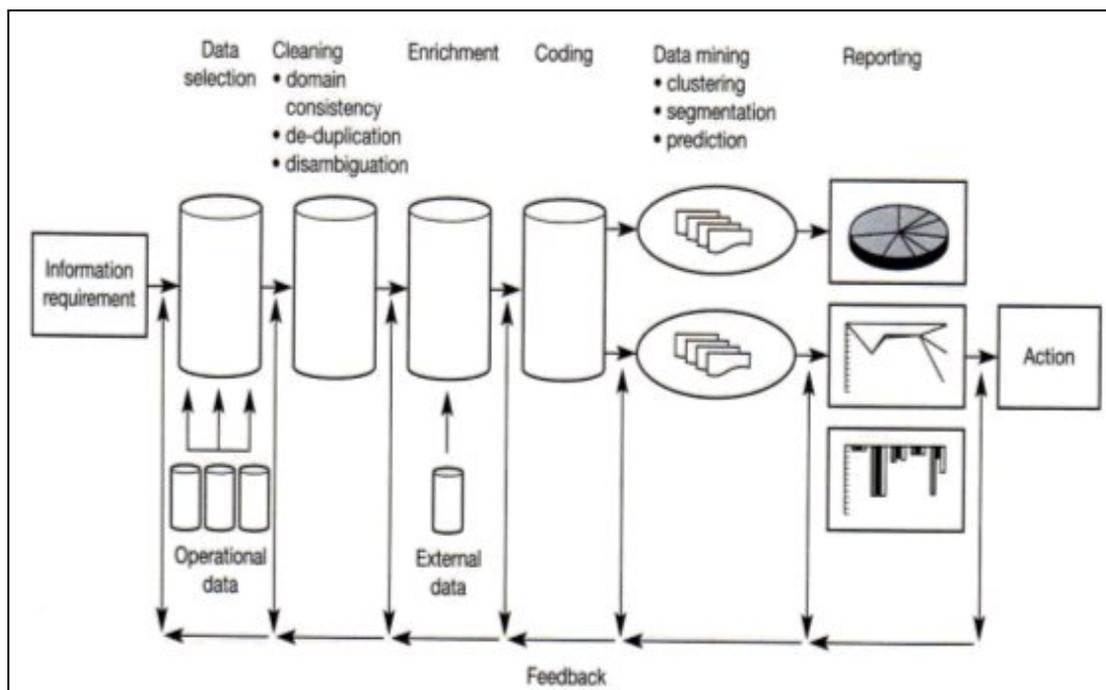


圖 3 知識發現流程圖

2.2.2 資料探勘技術

資料探勘的工具是利用資料來建立一些模擬真實世界的模式 (Model)，利用這些模式來描述資料中的特徵 (Patterns) 以及關係 (Relations)。這些模式有兩種用處，第一，瞭解資料的特徵與關係可以提供你做決策所需要的資訊，譬如 Association Model 可以幫助超級市場或百貨店規畫如何擺設貨品。第二，資料的特徵可以幫助你做預測，例如你可以從一份郵寄名單預測出哪些客戶最可能對你的推銷做回應，所以你可以只對特定的對象做郵購推銷，而不必浪費許多印刷費郵寄費而只得到很少的回應。

隨著資訊科技超乎想像的進展，許多新的電腦分析工具問世，例如關聯式資料庫、模糊計算理論、基因演算法則以及類神經網路等，使得從資料中發掘寶藏成為一種系統性且可實行的程序。一般而言，資料探勘的理論技術可分為傳統技術與改良技術兩支。傳統技術以統計分析為代表，舉凡統計學內所含之敘述統

計、機率論、迴歸分析、類別資料分析等皆屬之，尤其資料探勘對象多為變數繁多且筆數龐大的資料，是以高等統計學裡所包括之多變量分析中用來綜簡變數的因素分析（Factor Analysis）、用來分類的判別分析（Discriminant Analysis），以及用來區隔群體的分群分析（Cluster Analysis）等，在 Data Mining 過程中特別常用。

在改良技術方面，應用較普遍的有決策樹理論（Decision Trees）、類神經網路（Neural Network）以及規則歸納法（Rules Induction）等。

1. 決策樹是一種用樹枝狀展現資料受各變數的影響情形之預測模型，根據對目標變數產生之效應的不同而建構分類的規則，一般多運用在對顧客資料的區隔分析上，例如針對有回函與未回函的郵寄對象找出影響其分類結果的變數組合，常用分類方法為 CART（Classification and Regression Trees）及 CHAID（Chi-Square Automatic Interaction Detector）兩種。
2. 類神經網路是一種模擬人腦思考結構的資料分析模式，由輸入之變數與數值中自我學習並根據學習經驗所得之知識不斷調整參數以期建構資料的型樣（patterns）。類神經網路為非線性的設計，與傳統迴歸分析相比，好處是在進行分析時無須限定模式，特別當資料變數間存有交互效應時可自動偵測出；缺點則在於其分析過程為一黑盒子，故常無法以可讀之模型格式展現，每階段的加權與轉換亦不明確，是故類神經網路多利用於資料屬於高度非線性且帶有相當程度的變數交感效應時。
3. 規則歸納法是知識採礦的領域中最常用的格式，這是一種由一連串的「如果.../則...（If/Then）」之邏輯規則對資料進行細分的技術，在實際運用時如何界定規則為有效是最大的問題，通常需先將資料中發生數太少的項目先剔除，以避免產生無意義的邏輯規則。

隨著上述的技術不同，資料探勘工具只在某些領域下有特別的效能，也就是說尚無適用所有業種、用途的工具問世。大致分為 Case-Based Reasoning、Data

Visualization、Fuzzy Query and Analysis、Knowledge Discovery、Neural Networks
五類。

2.2.3 資料探勘模組

運用各式理論技術，資料探勘可以建立六種模式：Classification、Regression、Time Series、Clustering、Association、以及 Sequence。Classification 以及 Regression 主要是用來做預測，而 Association 與 Sequence 主要是用來描述行為。Clustering 則是二者都可以用的上。

1.分類(Classification)

分類是根據一些變數的數值做計算，再依照結果作分類。（計算的結果最後會被分類為幾個少數的離散數值，例如將一組資料分為"可能會回應"或是"可能不會回應"兩類）。分類常常被用來處理如前面說到的郵寄對象篩選的問題。我們會用一些已經分類的資料來研究它們的特徵，然後再根據這些特徵對其他未經分類或是新的資料做預測。這些我們用來尋找特徵的已分類資料可能是來自我們的現有的歷史性資料，或是將一個完整資料庫做部份取樣，再經由實際的運作來測試；譬如利用一個大的郵寄對象資料庫的部份取樣來建立一個 Classification Model，以後再利用這個 Model 來對資料庫的其他資料或是新的資料作預測。

2.迴歸預測(Regression)

迴歸預測是使用一系列的現有數值來預測一個連續數值的可能值。若將範圍擴大亦可利用 Logistic Regression 來預測類別變數，特別在廣泛運用現代分析技術如類神經網路或決策樹理論等分析工具，推估預測的模式已不在止於傳統線性的侷限，在預測的功能上大大增加了選擇工具的彈性與應用範圍的廣度。

3.時間順序預測(Time-Series Forecasting)

時間順序預測與迴歸預測很像，只是它是用現有的數值來預測未來的數值。時間順序預測的不同點在於它所分析的數值都與時間有關。時間順序預測的工具可以處理有關時間的一些特性，譬如時間的階層性（例如每個禮拜五個或六個工作天）、季節性、節日、以及其他的一些特別因素如過去與未來的關連性有多少。

4. 叢集(Clustering)

叢集是將資料分為幾組，其目的是要將組與組之間的差異找出來，同時也要將一個組之中的成員的相似性找出來。叢集與分類不同的是，你不曉得它會以何種方式或根據什麼來分類。所以你必須要有一個分析師來解讀這些分類的意義。

5. 關聯性(Association)

關聯性是要找出在某一事件或是資料中會同時出現的東西。關聯性主要是要找出下面這樣的資訊：如果 Item A 是某一事件的一部份，則 Item B 也出現在該事件中的機率有 X%。（例如：如果一個顧客買了低脂乳酪以及低脂優酪乳，那麼這個顧客同時也買低脂牛奶的機率是 85%。）

6. 循序性(Sequence Discovery)

循序性與關聯性關係很密切，所不同的是循序性中相關的 Item 是以時間區分開來。

目前資料探勘的技術已趨向成熟且應用廣泛，例如：製造商預測可能造成的生產停頓問題，政府部門從金融交易資料的分析中找出金融犯罪形態，或從犯罪的資料中找出可能的犯罪趨勢，企業預測顧客的購買行為等。在商業應用上，資料探勘使企業的決策者和行銷管理人員可以利用這些新的資訊科技，來整合管理企業內部的知識，提供更好的商業智慧，讓企業能夠對客戶提供更好的產品與服務，以提高公司形象與營運效率。

2.3 IC 封裝與測試之生產流程概述

封裝(Assembly or Packaging)與測試(Final Test) [17][18]是屬於 IC 製程的後段(Back End)製程(圖 4)，許多測試廠或封裝廠為保有競爭力，多半向前整合封裝廠或向後整合測試廠，使得後段製程可以在同一個廠中完成，提供一元化(Turn Key)服務，以降低運輸成本。

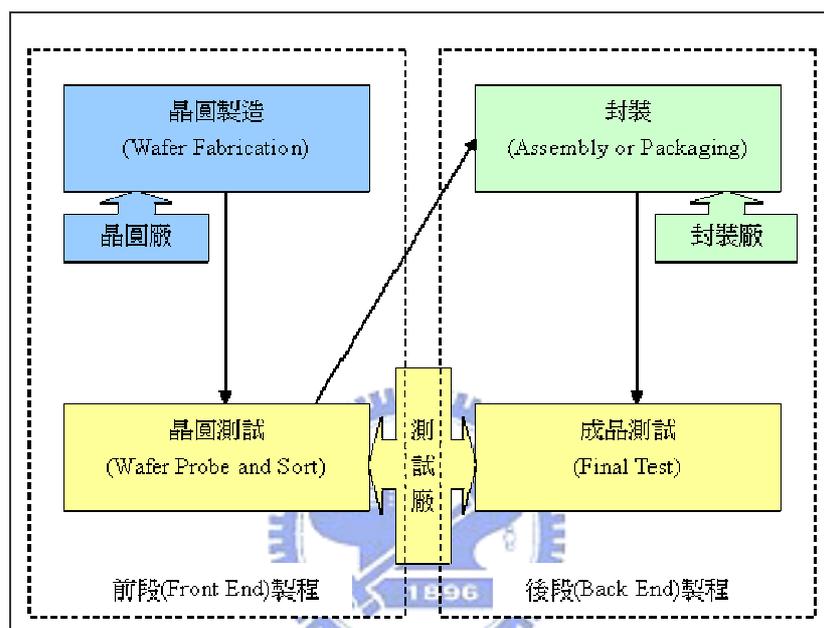


圖 4 積體電路(Integrated Circuits, IC)的製造過程

2.3.1 IC 封裝型態及製程

本文以 TSOP 為例，一般 IC 封裝製程如圖 5。以下僅以此類產品的封裝流程的幾項重要作業程序做一介紹：

1. 晶圓黏貼(Wafer Mount)：將晶圓背面貼上藍膜(Blue Tape)並置於鋼製的框架上，避免晶片和膠帶間有氣泡產生。
2. 晶圓切割(Die Saw)：將晶圓片上面的晶粒(Die)切割分離。
3. 晶粒黏貼(Die Bond)：將晶粒置於導線架(Lead Frame)上並使用銀膠(Epoxy)加以黏貼固定。
4. 鐳線(Wire Bond)：將晶粒上的接點以極細的金線連接到導線架之內引腳，

以便將 IC 晶粒之電訊號傳輸至外界。

5. 斷、短路測試(O/S Test)：測試晶粒上接點到導線架之間的接線是否有問題，而進行斷，短路測試，以確保接線的正常。
6. 壓模(Molding)：將導線架置框架上並預熱，再將框架置於壓模機上的封裝模上，再以半融化的樹脂充填，待樹脂硬化後便可取出封膠後的成品。
7. 去框(Singulation)：在封膠後將不需要的凸出樹脂切除。
8. 蓋印(Mark)：將商品規格及製造者等資訊印於封裝完的 IC 膠體之上，蓋印有雷射蓋印及油墨蓋印兩種方式。
9. 目視抽驗(V/M QC Gate)：依抽驗規則所決定的抽樣數來進行抽樣檢查。
10. 目視檢驗(V/M)：使用人工目視的方式，來進行外觀上的檢驗，以便在進行下一作業程序前，先去除掉一些不合格的。其檢驗項目包含：外引腳之平整性、共面度(Coplanarity)、腳距、蓋印之清晰與否與膠體是否有損傷等外觀上能檢驗的項目。



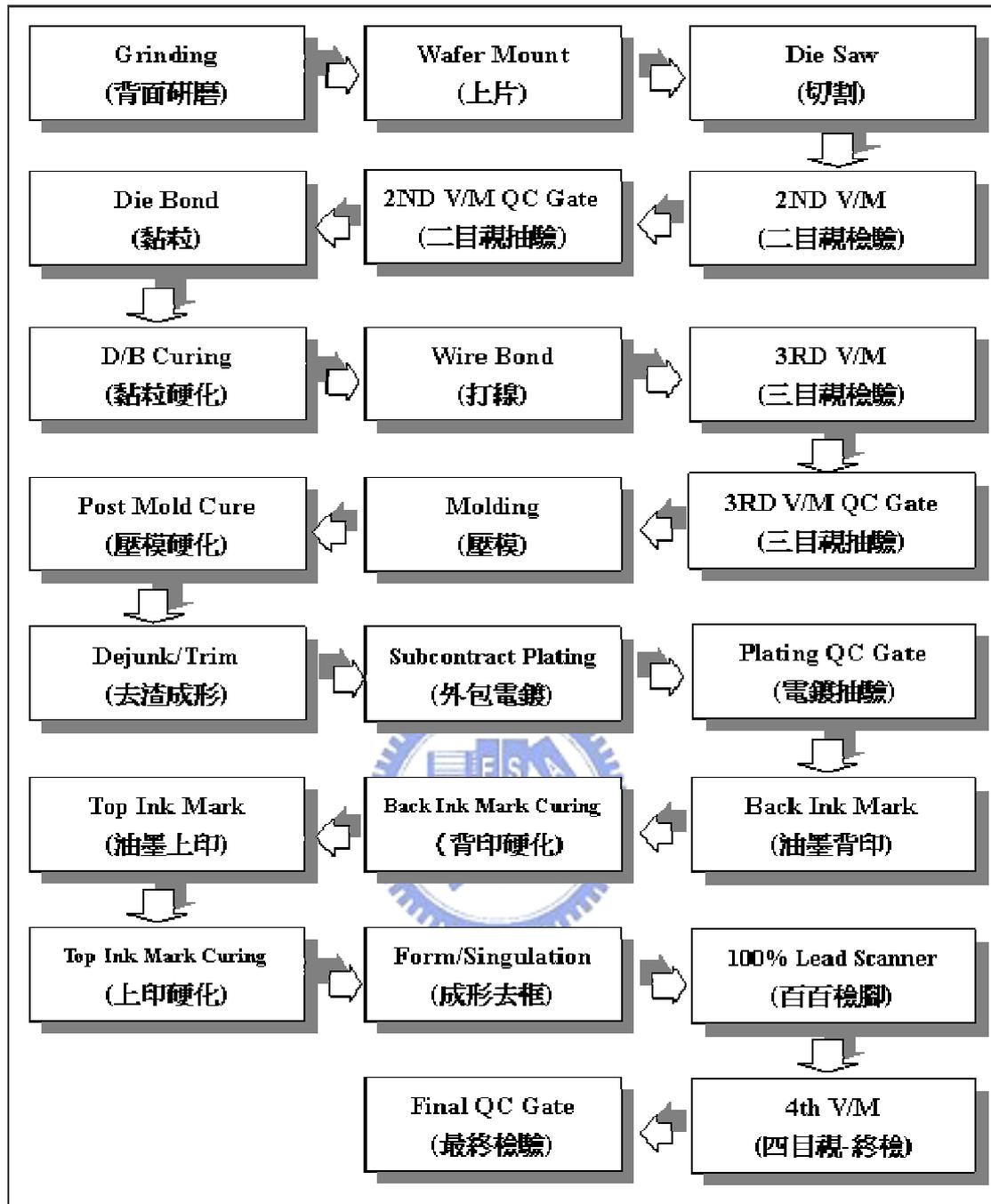


圖 5 TSOP 封裝製程

2.3.2 IC 晶圓測試製程

晶圓片由晶圓廠產出後，須先在測試廠進行晶圓測試後，再到封裝廠進行封裝。所以晶圓測試(Circuit Probing)是在 IC 封裝前，先對晶圓進行測試，以便篩選掉一些不合格的晶圓片，進而減少封裝材料的損失。

IC 晶圓測試的製程較為單純，可分為幾下幾道作業程序：

1. 收貨:接收從晶圓廠送來的晶圓片。
2. 進貨檢驗:檢驗晶圓片的外觀上是否有缺損。
3. 晶圓測試:進行晶圓測試作業。
4. 出貨檢驗:於出貨前再確認。
5. 出貨:出貨到封裝廠進行 IC 封裝作業。

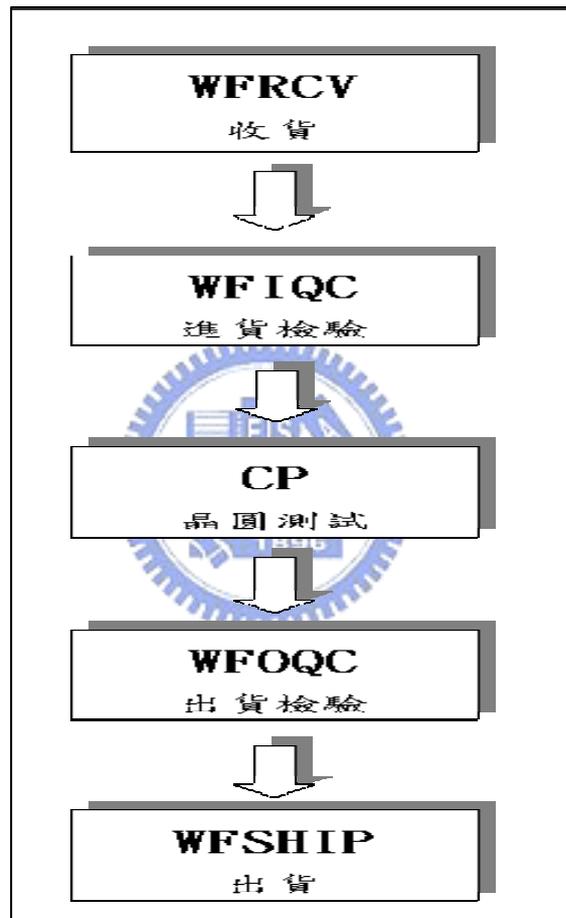


圖 6 晶圓測試製程

2.3.3 IC 成品測試製程

測試製程乃是於 IC 封裝後，測試封裝完成的產品之電性功能以保證出廠 IC 功能上的完整性，並對已測試的產品依其電性功能作分類（即分 Bin），作為 IC 不同等級產品的評價依據。

一般 IC 成品測試製程如圖 7。以下將就測試流程的幾項重要作業程序做一介紹：

1. 上線備料

上線備料的用意是將預備要上線測試的待測品，自上游廠商送來的包箱內拆封，並一顆顆的放在一個標準容器（幾十顆放一盤，每一盤可以放的數量及其容器規格，依待測品的外形而有不同）內，以利在上測試機台（Tester）時，待測品在分類機（Handler）內可以將待測品定位，而使其內的自動化機械機構可以自動的上下料。

2. 測試機台測試（FT1、FT2、FT3）

待測品在入庫後，經過入庫檢驗及上線備料後，再來就是上測試機台去測試；如前述，測試機台依測試產品的電性功能種類可以分為邏輯 IC 測試機、記憶體 IC 測試機及混合式 IC（即同時包含邏輯線路及類比線路）測試機三種。測試機的主要功能在於發出待測品所需的電性訊號並接受待測品因此訊號後所回應的電性訊號並作出產品電性測試結果的判斷，當然這些在測試機台內的控制細節，均是由針對此一待測品所寫之測試程式（Test Program）來控制。

3. 預燒（Burn-In）（測試記憶體 IC 才有此一程序）

在測試記憶體性產品時，在第一道機台測試(FT1)之後，待測品都會上預燒爐裡去預燒(Burn In)，其目的在於提供待測品一個高溫、高電壓、高電流的環境，使生命週期較短的待測品在預燒(Burn In)的過程中提早的顯現出來，在預燒(Burn In)後必需在 96 個小時內待測品預燒(Burn In)物理特性未消退之前完成後續測試機台測試的流程，否則就要將待測品送回預燒爐去重新預燒(Burn In)。在此會用到的配件包括 Burn-In Board 及 Burn In Socket..等。

4. 電性抽測

在每一道機台測試後，都會有一個電性抽測的動作（俗稱 QC 或 Q

貨)，此作業的目的在於將此完成測試機台測試的待測品中抽出一定數量，重回測試機台，在測試程式、測試機台、測試溫度都不變的情況下，看看其測試結果是否與之前上測試機台的測試結果相同，若不相同，則有可能是測試機台故障、測試程式有問題、測試配件損壞、測試過程有瑕疵..等原因，若原因比較輕微者，則需回測試機台重測；原因比較嚴重者，則要將此批待測品轉為待確認狀態(Hold)，等待工程師、生管人員與客戶協調後再作決策。

5. 標籤掃描 (Mark Scan)

利用機械視覺設備對待測品的產品上的產品標籤(Mark)作檢測，內容包括 Mark 的位置歪斜度及內容的清晰度..等。

6. 人工檢腳或機器檢腳

檢驗待測品 IC 的接腳的對稱性、平整性及共面度等，這部份作業有時會利用雷射掃描的方式來進行，也會有些利用人力來作檢驗。

7. 檢腳抽檢與彎腳修整

針對彎腳品進行彎腳品的修復作業，然後再利用人工進行檢腳的抽驗。

8. 加溫烘烤 (Baking)

在所有測試及檢驗流程之後，產品必須進烘烤爐中進行烘烤，將待測品上之水氣烘乾，使產品在送至客戶手中之前不會因水氣的腐蝕而影響待測品的品質。

9. 包裝 (Packing)

將待測品依其客戶的指示，將原來在標準容器內的待測品的分類包裝成客戶所指定的包裝容器內，並作必要的包裝容器上之商標粘貼等程序。

10. 出貨(Shipping)

由於最終測試是半導體 IC 製程的最後一站，所以許多客戶就把測試廠當作他們的成品倉庫，以避免自身工廠的成品存放的管理，另一方面也減少不必要的成品搬運成本，因此針對客戶的要求，測試廠也提供了所謂的「Door to Door」的服務，即幫助客戶將測試完成品送至客戶指定的地方(包

括客戶的產品買家)。

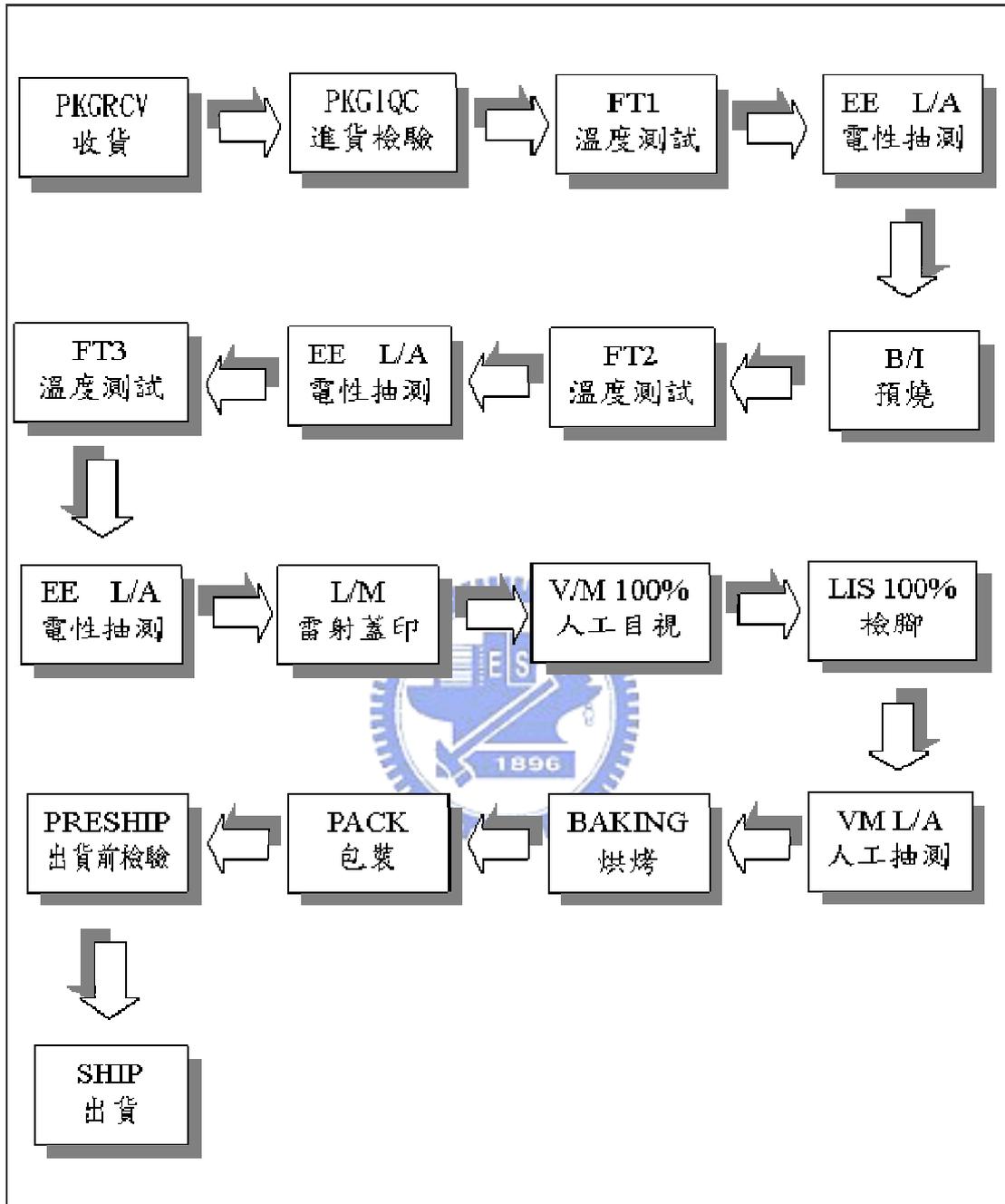


圖 7 成品測試製程

第三章 電腦整合製造問題定義

3.1 電腦整合製造

3.1.1 電腦整合製造概述

電腦整合製造系統的目的，就是在將工廠作業中的三個基本階段—產品設計、生產規畫和製造，利用電腦與設備間相互交換資訊來整合完成。CIM 代表的是一連串的技术整合，不僅是設備、儀器的整合，更是資訊的整合。CIM 的範圍，是從供應商的管理、產品的設計、製造，一直到把產品交到顧客手中，其中包括了生產、排程、庫存管理以及財務資源管理。

在一個生產製造的環境中，包含了兩個川流不息的網路，一個是看得見摸得到的物流(Material Flow)，一個則是看不見的資訊流(Information Flow)。物流是指由材料的購入、發料、加工、組裝、檢驗到成品出貨；資訊流則指倉儲管理、銷售管理、品管、研發、生產計畫與控制、物料管理及會計財務。CIM 便是將物流網路及資訊流加以自動化，並且相互配合以達到快速而經濟的方式，來生產高品質的成品。

電腦整合製造不僅強調生產設備的自動化，更強調生產環境資訊的連貫整合，使得資訊的流通更加迅速、確實。如此可以減少相同資料的一再重複輸入，以達到省時、省錢及正確決策的目的。所以 CIM 系統強調的是資訊的自動化，以快速且精確掌握產品在世界各地的生產與銷售狀況。透過資訊自動化，企業可以結合自動化孤島，連結經營系統，除可提升企業的生產力及效率，也可提高產品的品質，減少生產成本，增加市場競爭力。總括而言，CIM 系統具有下列優點：

1. 提高生產力

使用 CIM 系統，能夠減少直接或間接人力之需求，降低成本支出，並

且可依照更為緊密的排序排程來縮短生產時間，因而顯著地提高生產力。

2. 改善產品品質

在 CIM 系統中，從產品的設計、生產製造到測試工作，均由整合性的電腦系統加以支援，因此能夠生產出規格與可靠度均符合顧客需求的高品質產品。

3. 減少在製品的庫存數量

大量的在製品對企業是一筆龐大的成本負擔，因此各企業莫不期望能有效降低在製品數量。利用電腦整合製造系統，可以有效控制等待和設置時間，進而縮短生產及製造的前置時間，從而降低在製品的庫存數量。

此外，CIM 系統還有提升企業形象、改善顧客服務效率與增加企業長期利潤等優點。圖 8 及圖 9 分別為 CIM 系統之基本功能示意圖及 CIM 整體架構圖

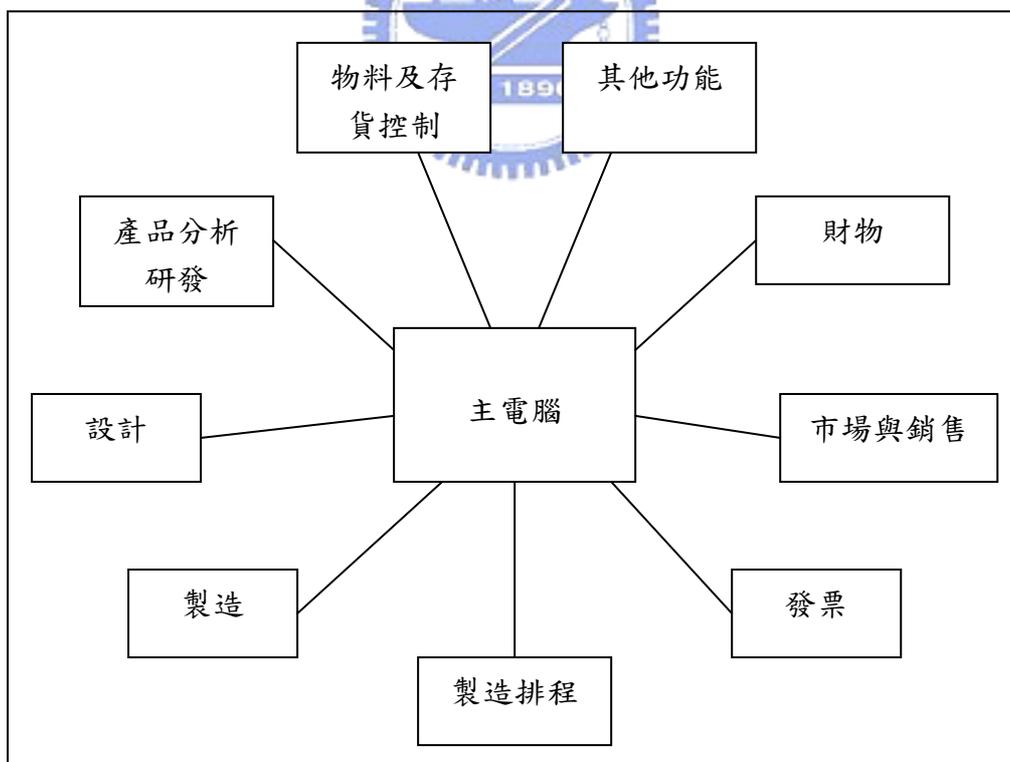


圖 8 CIM 基本功能示意圖

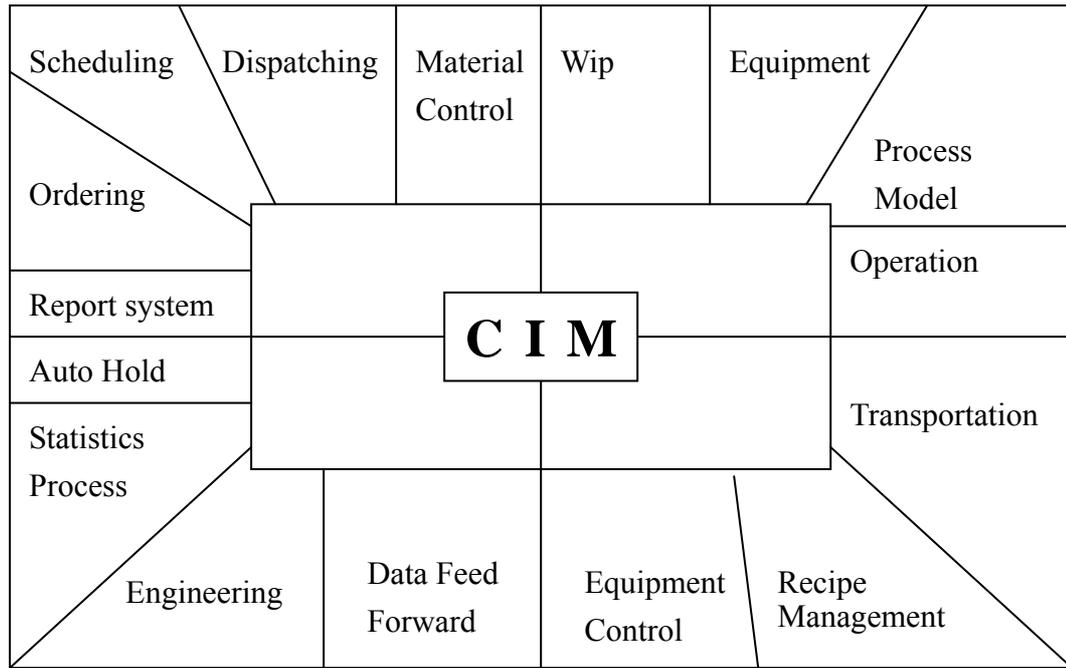


圖 9 CIM 架構圖

3.1.2 半導體電腦整合製造

半導體廠商基於公司內部機密之考量，很少將其系統內容公開，因此就一般可以取得的文獻資料來看，德州儀器公司（Texas Instrument）所發展之 Microelectronics Manufacturing Science and Technology（MMST）系統，在半導體製造業之 CIM 上提供了最完整的參考架構與實作成效，此一系統事實上是結合了製程管制、生產規劃、排程、生產模擬等功能，形成一完整的製造管理環境，因此成為許多廠商發展類似系統的借鏡。下表為以 MMST 之 CIM 架構為例之 CIM 功能需求表。

表 4 CIM 功能需求表

功能	說明
工廠管理	此功能透過整個系統架構之連接，可以掌握全廠訊息，系統之設定、啟動、關閉、作業管理、在製品追蹤、效能監控等都透過此功能完成。
工廠模擬	模擬是在下真正的指令前，預先推演可能之結果，找出最佳之輸出結果。好的模擬器必須要有很好的內部推演機制，並結合專家知識才能準確預知系統之行為。

生產規劃	生產規劃是根據工廠產能與機台能力限制，決定是否接單，並對所承接之訂單排定最佳之生產計劃與交期。半導體製造由於 Cycle time 長，生產期間之變因多，因此其生產規劃之困難度亦相對較高，新一代的生產規劃則期望能有較動態的規劃能力。
生產排程	生產排程承接生產計劃，因此是一種與資源密切結合，並力求資源要素平衡之執行性計劃，因此可以視為一種短期規劃。主要工作內容包括現場排程、投料控制、工作指派等等，在進行排程時。
規範管理	製程步驟、流程順序、控制條件等工程資訊，為了使得生產能更有效率，應將這些規範予以電腦化管理，因此在產品正式投料生產時，即可透過電腦將這些資訊叫出，並遵循所記載之順利與控制條件，以避免人為之錯誤。
通用設備模型	通用設備模型是對新一代半導體生產機台設備所定義的通訊模式，使得機台間的操作如物料傳送、機器排程、晶圓資料收集、生產效率監控、預防維護保養管理、遠端機台監控等都能有效執行，而對於 GEM 則提供了轉換器的功能。
機台控制	機台控制屬於最下層的作業控制，主要負責按照所接收到之工作指令，執行機台之作業順序、參數設定、機台校驗等工作。
製程控制 1	製程控制的目的是當製程有錯誤或偏移發生時，能夠在最短的時間內即時修正以避免不良品的繼續產生，以確保製程品質的一致與穩定，進而提昇產品良率。

3.2 電腦整合製造遭遇的問題

在生產製造的過程中，如果可以藉由在生產的過程中記錄一些資料，是否也可以採擷出一些有價值的資訊。我們首先必須要找出資料具有哪些特性，以及定義資料的記錄方式。本論文會將焦點集中在半導體封裝製程之資料探勘，在半導體封裝的製作過程中具有下列之特性：

1. 生產流程有著一定的模式
2. 每個生產步驟會有一部至多部的機器
3. 在生產的過程中每一批原料所經過的機器組合不盡然相同

以生產 DDR RAM 之顆粒，其中必然要經過一些步驟才能完成生產，但是每一批原料在每一個步驟所經過的機器則不盡然會相同，而且最重要的是不同的組合可能會造成不同的良率。也因此本文中假設：不同的機器組合，是否會影響到生產過程的良率。



第四章 智慧型電腦整合製造系統

4.1 智慧型電腦整合製造系統設計架構

智慧型電腦整合製造系統主要著重於解決以下幾個問題：智慧型電腦整合製造系統，資料倉儲，整合資料倉儲與決策分析，整合決策分析與資料探勘，設計資料探勘引擎。其系統架構設計如圖 10 所示。其系統元件的設計概念如下表 5。

表 5 系統元件的設計概念與功能表

目標	設計概念
專家提供專業知識	資料收集、資料的前處理及資料屬性的選擇，以及應該選用哪一種或多種資料探勘的方法來處理最適當。
資料規格化	資料來源及取得方法確定後，對資料的類型給予規格化，使後續的資料收集及資料前處理能有一致性，並提供資料屬性選擇上的基礎。
資料產生與收集	資料的收集方法需要考慮如何從現有資訊應用系統中的舊有資料作轉換的問題，要如何擷取出所要的部份，資料量要多少才足夠，資料分散及不同資料型態整合的問題
資料前置處理	在將資料存入資料倉儲之前，我們收集到的資料中某些欄位可能會有遺失內容、不正確的干擾資料、資料內容不一致性的問題。還有為了提高處理的效率與準確性，我們可能會先將資料給予整合、轉換、精煉或過濾的處理。
資料庫系統及資料倉儲的建構	由於利用資料探勘的技術來找出隱藏在資料庫中的資訊，所要處理的資料可能分散在不同資料庫，若是資料未經前置處理，沒有系統化的存放在資料庫中，則在萃取知識前將花費大量的時間在資料處理上，因此我們採用資料倉儲來克服這些問題。資料倉儲將資料再加以分類和整合後

	再存進資料倉儲中，如此可減少資料探勘在對資料進行搜尋的時間。這是資料探勘配合使用資料倉儲最大的好處。
屬性的挑選	在一般情況下，資料收集所考慮到的屬性都是廣泛的，當進行資料分析時才由其中挑選合適的屬性所構成的資料集進行分析。通常這樣的屬性挑選動作，都是由該領域的專家，負責針對特定的分析目標選擇適當的屬性集。
資料探勘引擎	資料探勘引擎是整個架構的核心，也是最為重要的環節。經常被使用的方法或技術，如前面章節所討論。
結果評估	關連法則中的支持度及可靠度的值、組群化中的組群個數K 的值及分類中的階層數，才能使挖掘的結果更有可用性與解釋性。
視覺化	將所探勘到的結果以視覺化的方式來展示。
知識庫	知識庫用來儲存專家的專業知識以及探勘所獲得的可用知識，這可隨時的更新，做為以後應用的參考，以及提供各項決策的依據。

由表 5 中可以清楚的了解到系統設計中各個元件的主要設計理念以及所希望達成的目標，由圖 10 中更可以清楚的看出系統中各元件之間的關聯。

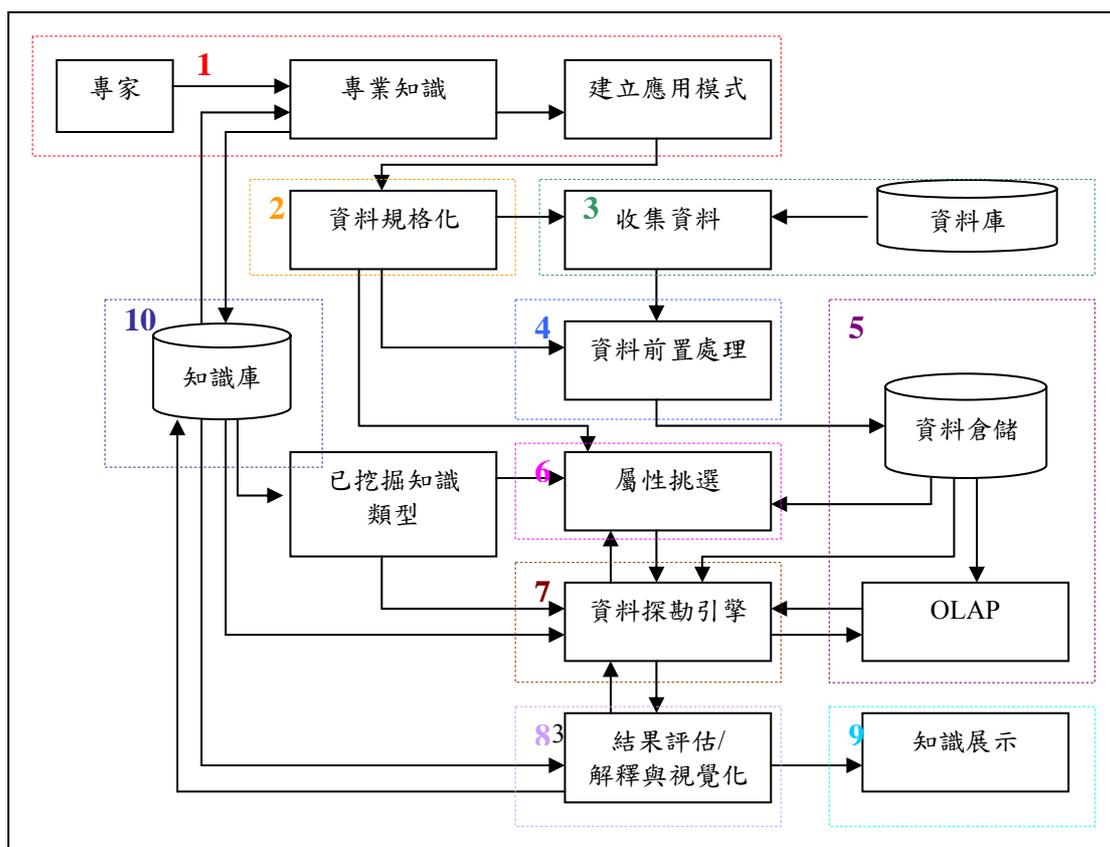


圖 10. 智慧型電腦整合製造系統架構設計圖

4.2 整合智慧型電腦整合製造系統

4.2.1 資料倉儲系統

隨著資料倉儲技術的發展，建立資料倉儲的模式，已從原先的「由上而下」模式，發展到下列六種模式[5]：

1. 由上而下模式：

是指從原來分散儲存在企業各處的線上交易處理資料庫(OLTP)中的有用資料，透過萃取(extraction)、淨化(clean)、轉換(transformation)、聚集(aggregation)等處理步驟建立一個整體性資料倉儲。

2. 由下而上模式：

是從建構各部門或特定的企業問題的資料超市開始，而整體性資料倉儲是建立在這些資料超市的基礎上。其特點是初期投資少、見效快。

3. 平行開發模式：

是指在一個整體性資料倉儲的資料模型的指導下，資料超市的建立和整體性資料倉儲的建立同步進行，。因此可避免各部門在開發各自的資料超市時的盲目性，減少各個資料超市之間的資料冗餘和不一致性。

4. 有回饋的由上而下模式：

此模式中，使用者的新需求的回饋分為兩個階段。第一階段：用戶的新需求不斷地被回饋給部門的資料超市，部門資料超市根據使用者的新需求，產生自身的需求變化。第二階段：部門資料超市把自身的需求變化回饋

給整體性資料倉儲，整體性資料倉儲再作出相應的變化。

5. 有回饋的由下而上模式：

此模式先建構部門資料超市，再以各部門的資料超市為基礎，建構整體性資料倉儲的方式。因此資料超市較能滿足用戶的需求，在整體性資料倉儲建立好之後，需求的變化，將主要體現在資料超市與資料倉儲之間。如果各部門資料超市，在發展時注意保持相互之間的資料一致性，並能根據用戶的回饋資訊，不斷地調整自己，那麼以這種模式建立的資料倉儲，在投入使用之後，能減少因用戶的需求變化，所帶來的不便。

6. 有回饋的平行開發模式：

在開發的起始階段，開發人員主要是在整體性資料倉儲資料模型的指導下，建立部門資料超市。並將建立過程中，所遇到的問題、解決方案、客戶的意見等資訊，回饋給整體性資料倉儲資料模型。讓整體性資料模型，在指導部門資料超市建構的同時，也能根據這些回饋的資訊調整自己。

綜合上述六種資料倉儲的建立模式，以由上而下的模型，來作為建立資料倉儲系統之範例，並將此模型區分為三個層級。

4.2.2 整合 CIM 系統與資料倉儲模式

當要整合 CIM 系統與資料倉儲時，通常會遇到的問題是不一致的不完整的與重複的資料。因此，在整合 CIM 系統與資料倉儲時必須將原始資料的資料型態一併收集於資料籌劃區域，而且這些資料必須經過清除修改或是結合搬動重複的資料。經過處理過程之後，再儲存於簡報伺服器(Presentation Server)：一個放置經過組織、可直接供使用者查詢的資料的實體機器，整合 CIM 系統與資料倉儲的架構如圖 11 所示。

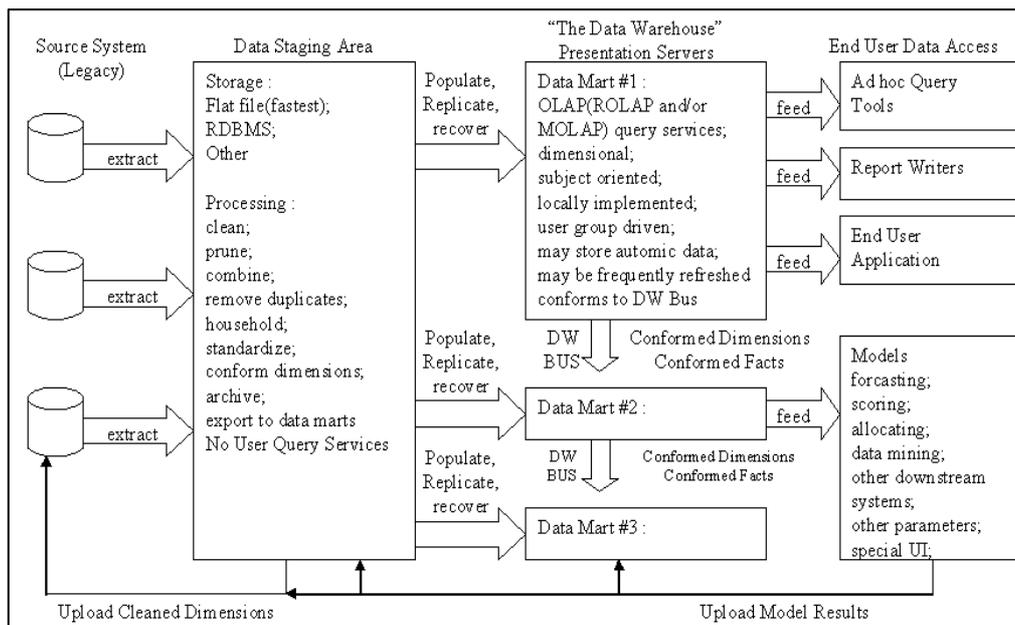


圖 11 整合 CIM 系統與資料倉儲的架構

整合 CIM 系統與資料倉儲有以下幾個步驟：

1. 收集(Collection):

將主要的資料收集到資料籌畫區以便後續的步驟使用

2. 轉換(Transformation):

這個步驟包含了修正資料的準確性與搬移不需要的資料

3. 建立索引>Loading an indexing):

轉換完成的資料存放在資料市場(data mart)中並且建立索引

4. 控管與查核(control check):

確保資料的品質

5. 發表/出版(Announcement or publication):

為了系統線上使用的準備工作

6. 更新(Renovation):

持續的修正不正確的資料

7. 搜尋(Search):

提供資料搜尋的服務

8. 查核與準備(Checks and preparations):

避免資料倉儲內的資料受到潛在的威脅

4.3 整合資料倉儲與決策分析

4.3.1 設計資料倉儲的綱要

我們可以使用下列三種的架構來建立一個資料倉儲的綱要：星狀綱要(Star Schema)、雪花式綱要(Snowflake Schema)、星狀雪花綱要(Starflake Schema)，此三種綱要都是以事實資料表為中心，它們的不同點只是在外圍維度資料表相互之間的關係，此三種綱要之建構方式，說明如下[6]：

1. 星狀綱要(Star Schema)：

一個星狀綱要的維度資料表只會與事實資料表產生關連性，維度資料表與維度資料表之間並不會產生任何的關連。

2. 雪花式綱要(Snowflake Schema)：

一個雪花式綱要是用來描述會合併在一起使用的維度資料，事實上維度資料表只與事實資料表相關連是反正規化後的結果。如果我們在作法上將時常合併在一起使用的維度加以正規化，也就是將維度局部正規化後的結果，這就是所謂的雪花式綱要。

3. 星狀雪花綱要(Starflake Schema)：

星狀雪花綱要則是將星狀綱要與雪花式綱要，合併在一起使用的結果。

4.3.2 設計事實資料表

設計一個資料倉儲綱要首先要面對的，就是從一個 OLTP 系統的眾多欄位中，決定那些是事實資料，而那些是屬於維度資料，其判定的步驟如下述：

1. 在整個 OLTP 系統中尋找最基本的交易，它們極可能是事實資料。
2. 決定搜尋每一事實資料的鍵值(Key)，它們極可能是維度資料。
3. 檢驗每一可能事實資料的欄位，確定它不是嵌入在事實資料中的維度資料。

4. 檢驗每一可能維度資料的欄位，確定它不是嵌入在維度資料中的事實資料。

事實資料表格中存放著我們所需要的真實資料，在此表格中的資料無法更改，只能新增資料到表格中，而且在表格中也包含了與維度表格相互連接的外鍵[3]。如果資料是因為公司的營運而產生出來的，而且這些資料是不會隨著時間而改變，則此資料即為事實表格資料。當要判斷一項資料是否為維度資料的時候，照理講透過維度資料應該可以找到一群的資料錄而不是只有單一的一筆資料錄。如果發覺可以使用一項資料找到很多筆的資料錄，則該項資料很可能就是一維度資料。

在設計一個事實資料表的時候，需考量下面的事項[6]：

1. 為每一項功能決定其資料倉儲時距：

一般而言，一個資料倉儲是公司詳細資料與彙總資訊的混合使用，以達到一個查詢速度與資料量的平衡點。因此，並不一定是資料倉儲時距越長越好、事實資料越詳細越好。就資料倉儲時距而言，每一個功能所要求的都不相同，例如某一功能希望能查詢去年度同一季及前年度同一季的資料，而另一功能則希望能查詢去年開始 15 個月份的資料。決定好資料倉儲的時距，將可協助了解全盤狀況，以及降低儲存在資料倉儲中的資料量。

2. 為每一項功能決定其統計採樣(Sampling)原則

如果 OLTP 系統擁有大量的資料，而且資料倉儲時距很長，而且只是想分析某一項目的趨勢，也許沒有必要將所有的事實資料都存入資料倉儲之中，因為此種情況下連計算彙總資訊都是一件很大的工程。因此，可以使用統計採樣的方式，僅將部份的事實資料納入資料倉儲，然後以採樣的資料為基準，再來計算彙總資訊。只要使用正確的採樣模式以及採用的資料具有代表性，則一定能符合統計分析需求。

3. 決定在事實資料表中應包含那些欄位

首先要排除 OLTP 系統中不必要的資料欄位，例如：顯示狀態的欄位、儲存中間結果的欄位、儲存彙總結果的欄位，以及某些內部使用的參考欄位，此都是應加以排除的對象。

4. 儘量縮減事實資料表中欄位的大小

在資料庫中每一項目都是有其資料型態的，而且每一類的資料型態也有許多類型，例如：整數資料有兩位元組整數、四位元組整數以及八位元組整數。我們可以對事實資料表中的每一個欄位加以分析，並且為它們訂定出佔用最少磁碟空間但是不影響精確度的資料型態，以有效節省資料空間。

5. 決定是否使用智慧型鍵值(Intelligent Key)

在一個事實資料表中，除了使用者所關心的事實資料之外，為了加快資料查詢的速度，我們為每一維度增加了一個欄位，而且該欄位是相對應維度資料表的外部索引鍵，它顯示了一個事實資料表與維度資料表之間的關連性。我們可以使用下列兩種方式來製作一事實資料表的外部索引鍵：

(1).使用智慧型鍵值：

智慧型鍵值指的是在真實世界中所能理解的數值，例如：台北市、電腦、主機板...。

(2).使用非智慧型鍵值：

因為智慧型鍵值會佔用較多的記憶體，所以我們可以在事實資料表中以一唯一數字，例如資料錄號碼來代表，而在維度資料表中新增一個參考至該筆資料的欄位。

6. 將時間因素納入事實資料表

時間也是一項重要的事實資料，我們應該將一件事實發生的時間納入事實資料表。時間也是一項維度資料，但是時間與一般的維度資料不大一樣，它具有固定的格式與階層架構。當我們將時間因素納入事實資料表時，可以考慮以下列的方式之一來儲存時間資料：

(1).儲存實際的時間：

是一種智慧型鍵值的作法。

(2).儲存距離一固定時間點的時距：

以一個固定的時間點為基準點(例如：2003 年 1 月 1 日)，而在事實資料表中儲存距離該時間點的一個差距值(Offset)。

(3).儲存一日期範圍：

在事實資料表中儲存一個時間範圍，例如產品 A 的庫存位置在今年度都是在某一個固定位置，則可將該產品儲存位置的時間設定為某一起始日至今年年底，而到了明年初再為產品 A 新增一筆庫存位置的資料錄。

7. 分割事實資料表

當我們設計好一個事實資料表之後，還要為它訂定分割策略，一般而言，分割事實資料表有下列幾種作法：

(1).水平式分割：

將事實資料表的特性加以分類，再水平的切割為數個區段，我們可以將水平式分割看作是在一 SELECT 敘述中所使用的 WHERE 子句，一般而言，可以下列的方式來作水平分割：

(A).以時間為基準將資料分割成相等的區段：使用時間來分割資

料，有一項最大好處，當資料倉儲週期到達時，我們只需將新的歷史資料建立成一個資料分割，然後再備存過時的資料分割就可以，其他的資料分割則不需變動。

(B).將資料分割成不相等的區段：例如將一年度的事實資料表劃分

為三種資料分割：最近三個月的資料每月一個分割、上一季的資料一個分割、上半年的資料一個分割。此分割方式的好處是當查詢常用的資料時不需要掃描大量的資料，缺點則是每當資料週期到達時，則必須重建所有的資料分割。

(C).以一維度為基準來分割資料：如果使用者的需求中一大部份是針對某一維度來分析資料，且該維度是不會經常改變的，則可以該維度為基準來分割事實資料表。

(D).將資料分割成相同大小的區段

(2).垂直式分割：

將事實資料表中的一個欄位或是某些欄位做縱向的篩選，而成為一個資料分割。

(3).硬體式分割：

將資料分割置於不同的工作平台之上。

4.3.3 設計維度資料表

維度表格的資料是用來參照事實表格內的資料，透過某些事件的需求將一些繁複敘述分割成幾個小部份，例如某一時點的品質資訊等，儘量藉由這些維度表格的設計減少對事實資料表格的負擔，其資料不像事實資料表格那麼多，但具有修改的特性[13]。但是要注意的是，在設計之初就要確定不會變更維度資料的主索引鍵，因為主索引鍵改變了，是不是事實資料表也要一起變更[14]。

設計維度資料表與傳統的 OLTP 系統不一樣，它是採用反正規化的方式來設計的。例如，如果要針對地域性來加以分析，可以新增一個地域維度資料表，而且在表格中納入其他相關資訊，譬如縣市名稱、里名稱。一個維度資料表會儲存許多的重複資料，這就是以空間換取時間的作法。我們可以將一個維度資料表看作是事先將所需資料加以合併並儲存起來的結果，這樣在使用時就不需要臨時再做合併的動作，而可以節省時間。在一個維度資料表中可能會有一些階層資料，在設計維度資料時我們應該把相關的階層資料納入同一個維度資料表，這樣可以增加查詢時細分資料的執行速度[6]。

4.3.4 設計彙總資料表

彙總資訊主要是為了加速一般查詢的執行，然而太多的彙總資訊會造成過高的營運成本，而太少的彙總資訊會造成執行效能低落，我們必須使得查詢執行效能與額外營運成本之間達到一個平衡點。訂定彙總資訊策略的基本假設是，一般查詢只會使用到彙總資訊或是詳細資訊的一個部份集合，而且管理決策系統是以維度來查詢趨勢性的資料。在資料倉儲之中，彙總資訊是與詳細資料分開儲存的，而且彙總資訊是會隨時間而改變的。星狀綱要的設計，使我們可以隨時變更或新增彙總資訊，而不會牽一髮而動全身，一個資料倉儲的彙總資料表範例，如圖 13 所示[6]。

設計一個彙總資料表與設計事實資料表不同，在一個彙總資料表中，除了要納入某一事實資料表欄位的加總結果之外，還要嵌入相關的索引資料，以降低因合併資料而產生的工作負擔。設計一彙總資料表的工作項目歸納如下[6]：

1. 決定要為那些維度產生彙總資訊：

彙總資訊是以維度為基準來加總資料，並納入所需的事實資料數值來整體加以考量，因此它是以不同的組合方式來檢視一項事實資料。針對一個具階層關係的維度，我們可由最低階開始加總，進而向上計算各個階層的彙總資訊。

2. 決定要為那些事實資料數值產生彙總資訊

3. 計算彙總資料並存入彙總資料表：

彙總資料不一定是加總(SUM)，它可以包含許多的統計運算函數，例如：MIN()、MAX()、AVERAGE()、COUNT()等。

4. 決定彙總資訊的階層關係：

彙總指的是以某一維度的某一階層來做資料的計算，所以在該基準點以下的階層是沒有彙總資料的。

5. 將維度資料嵌入至彙總資料表

6. 將時間維度加入至彙總資料表：

彙總資料表中最好永遠使用實際的日期，因為彙總資料表的資料筆數不會太多。

7. 為彙總資料表產生索引：

因為彙總資料表是經常要改變的，而且它們的資料筆數都不會太大，所以最好不要在彙總資料表中使用非智慧型鍵值，因為使用非智慧型鍵值的前提是不會變更唯一識別資料錄的主索引鍵數值。

4.3.5 中繼資料表

所謂的中繼資料指的是描述其它資料的資料，在一個資料倉儲系統中，中繼資料主要是用於下列的工作事項[6]：

1. 資料的轉換與載入：

利用中繼資料來定義下列的事項，包括資料來源、資料目的地、轉換方法。中繼資料有如一種組態資訊，它可以使得一項工作走向自動化之路。當我們設定好組態資訊後，如果工作內容變更改了，只需要修訂組態資訊即可。對於儲存轉換資訊的中繼資料而言，如果轉換公式是簡易的 SQL 敘述，則可以儲存實際的原始碼，如果是較複雜的計算或是經過編譯的大型程式碼，則可以儲存該模組的路徑名稱。一個資料倉儲的資料轉換與載入的中繼資料架構，如表 6 所示。

表 6 應用於資料的轉換與載入的中繼資料架構

應用面	資料來源	資料目的地	資料轉換
-----	------	-------	------

中 繼 資 料 架 構	識別碼	識別碼	識別碼
	名稱	名稱	模組名稱
	位置	位置	使用語言
	伺服器名稱	伺服器名稱	程式碼
	IP 位址	IP 位址	
	資料物件階層一	資料庫名稱	
	資料物件階層二	資料表名稱	
	
	欄位描述	欄位描述	
	資料欄一	資料欄一	
	資料欄二	資料欄二	
	

2. 資料管理：

中繼資料可以用來描述資料倉儲的資料，進而透過中繼資料而達到管理所有資料倉儲資料的目的。如果要使用軟體工具來管理資料倉儲的資料，則必須以中繼資料來描述每一個資料庫物件，例如：資料表、VIEW、索引、限制、彙總項目、資料分割等。一個資料倉儲的資料管理的中繼資料架構，如表 7 所示。

表 7 應用於資料管理的中繼資料架構

應用面	資料表	VIEW	索引
中 繼 資 料 架 構	識別碼	識別碼	識別碼
	表格名稱	VIEW 名稱	索引名稱
	資料筆數	資料欄一	索引欄一
	資料欄一	名稱	名稱
	名稱	來源欄位	資料型態
	資料型態	資料型態	上限值
	預設值	預設值	下限值

	唯一性 資料欄二 名稱 資料型態 預設值 唯一性	資料欄二 名稱 來源欄位 資料型態 預設值	索引欄二 名稱 資料型態 上限值 下限值
--	--	--	---

3.產生查詢：

查詢概觀檔就是一種為產生查詢而設計的中繼資料。凡是用來描述資料倉儲資料的中繼資料，都會被查詢管理員使用到，此外查詢管理員也會產生自己的中繼資料，以便產生有效的子查詢。當然這些中繼資料最後都可以回歸到查詢概觀檔。一個資料倉儲的產生查詢的中繼資料架構，如表 8 所示。



表 8 應用於產生查詢的中繼資料架構

應用面	查詢	執行時間	使用系統 資源	合併條件	彙總函數	GROUP 條件	排序條件
中 繼 資 料 架 構	使用者名稱 執行時間 使用系統資源 執行計劃 存取表格一 合併條件一	起始時間 完成時間	中央處理器 磁碟空間	表格名稱一 欄位名稱一 表格名稱二 欄位名稱二 執行運算	欄位名稱 函數類型	欄位名稱 公式	欄位名稱 排序方式

4.3.6 設計多重維度綱要

實際上當我們在分析資料時，並不是以單一的維度為基準，而是以多重維度為依據，所以我們應該將多個維度集合在一起成為一個單位而一併考量，這就是所謂的多重維度綱要(Multi-dimensional Schema)。因為一個三重維度綱要可以以立方體來表示，所以我們也稱呼多重維度綱要為一個立方體(Cube)。一個立方體(Cube)不一定要是一個三重維度綱要，它可以是二重維度綱要，也可以是N重維度綱要[6]。一個四重維度的資料立方體範例，如圖 12 所示[19]。

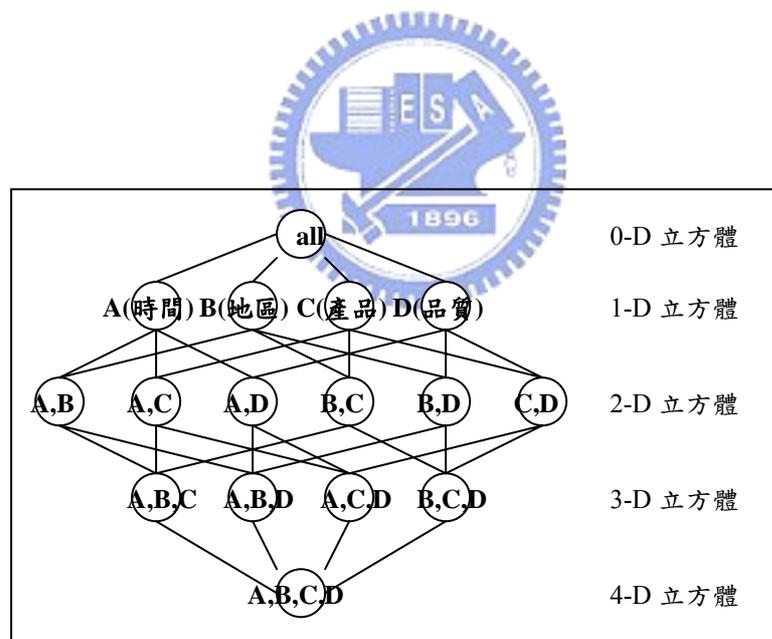


圖 12 四重維度的資料立方體

在原始資料經過整合與資料轉換、萃取後，將實際資料放入星狀綱要結構中，最後形成資料立方體(Data Cube)。資料立方體是指一個包含多維度資料模式的資料儲存體，其包含有事實表格與維度表格等相關資訊。每一個資料立方體擁有各式各樣的維度資訊，但有些維度則又具有層級關係，而每一個維度代表活動

事實的一種視野或觀點。由於資料倉儲的目的即是設計用來提供多元化的分析查詢，所以可能牽涉到的考量面很多，而利用資料立方體的方法，就可以快速地回應使用者所提出的要求[3]。一個資料倉儲系統可以包含許多個 Cube，每一個 Cube 是由不同的維度組合以及不同的事實資料表所構成的。一個資料倉儲的 Cube 組成架構範例，如圖 12 所示[6]。

每個資料立方體擁有自己本身的特性，而我們可以對資料立方體作各種角度的查詢與分析，資料立方體提供下列基本的瀏覽與查詢的機制：

1. 上捲(Roll-Up)：

可將目前的相關資訊，提昇一個層級，縮小成摘要性的資訊，讓使用者方便查尋與分析。

2. 下展(Drill-Down)：

這個動作與上捲相反，它將目前的資訊，向下一個層級，放大搜尋維度，讓使用者方便查詢與分析更細部的資訊。

3. 切片(Slice)：

將焦點集中在某一個觀點，來分析資料。

4. 轉換(Dice)：

轉換成不同的資料觀點，來做查詢與分析的動作。



4.4 整合決策分析與資料探勘系統

在完成了資料立方體(Data Cube)的設計建構之後，接著即可進行整合決策分析與資料探勘系統的工作，整合決策分析與資料探勘系統的方式，如圖 12 所示[15]。

首先，線上作業的資料探勘系統，分別接收與讀取來自決策分析系統所建構的資料立方體(Data Cube)的資料、批次作業的資料探勘系統的結果資料、或是

OLAP 決策分析的結果資料，然後資料探勘系統以其具有建立多種資料探勘功能的能力(例如：分類、群集等)，並運用多種資料探勘技術(例如：決策樹等)，以推演出有價值的隱藏資訊，最後將此有價值的資訊，透過使用者介面以視覺化的方式呈現。

在 OLAP 決策分析與資料探勘系統的整合方面，由於資料探勘是用來從資料中發掘出各種假設(Hypothesis)，但是它並不負責查證(Verify)假設，也無法確認資料的真實性，因此資料探勘所得的結果必須小心使用，不可完全依賴它。而 OLAP 決策分析則是著重於與客戶的互動(Interaction)，一般是由客戶預先假定一些假設(Assumption)，再進行資料的多維度分析。因此，兩者在處理層次上是不相同的。在整合資料探勘系統與 OLAP 決策分析的方式上，OLAP 的分析結果可以補充到資料探勘系統的知識庫中，給資料探勘系統提供分析資訊，並作為資料探勘的依據，OLAP 技術亦可將人的觀察力和智力融合入資料探勘系統中，以改善資料探勘系統的採掘速度與深度；而資料探勘所發現的知識，可以指導 OLAP 的分析處理，拓展 OLAP 分析的深度，以便發現 OLAP 所不能發現的，而且是更為複雜而細密的資訊。

4.5 設計資料探勘系統

4.5.1 資料分類與預測功能的處理流程

分類是在資料庫的物件集合中，按照分析對象的屬性分門別類加以定義，找尋共同性質，並建立類組(Class)的過程[7]。並且可以利用既有已經分類的歷史資料，來預測新的資料是屬於哪一個類別。

資料分類是一個二個步驟的處理流程，說明如下[7]。

1.學習：

從資料庫的物件集合中，找出一組訓練樣本資料集(Training Data Set)。根據資料分類的演算法加以分析，如決策樹(Decision Tree)、記憶基礎推理(Memory-Based Reasoning)、類神經網路等分類演算法，建立資料分類的模型又稱分類器(Classifier)。

2.分類：

從資料庫的物件集合中，找出一組有別於訓練樣本資料集的測試資料集(Testing Data Set)。將這組測試資料集送入分類器，以矯正資料分類模型的部份偏差後，再將未知的資料記錄(Unknow Record)。

4.5.2 設計資料探勘引擎

資料分類與預測模型所使用的方法為決策樹的 ID3 演算法，利用該演算法作為資料探勘引擎的核心，來分類隱藏在資料中的資訊，並且預測這些資訊。資料探勘引擎的演算步驟說明如下[19]：

1：訓練資料：

從資料庫的物件集合中，找出一組訓練樣本資料集(Training Data Set)。根據資料分類的演算法加以分析分類，將每一個類別的特徵定義清楚，並且透過預先已完成分類的訓練資料(Training Data)，為每一種類別建立出分類決策樹模型，利用這些模型來對未知的新資料進行分類。

2：建立決策樹節點：

選擇一個測試屬性，以該測試屬性為決策樹的單一節點，採由上而下的方式畫出分類決策樹。如果樣本資料皆是相同的類別，則此節點即成為一個葉節點，而且依此類別分類之。否則即利用資訊增益的衡量方式(又稱熵函數)，進行計算該測試屬性的資訊增益，計算方式如下述步驟3到步驟5。

3：計算分類樣本資料的期望資訊：

令 S 是由 s 樣本資料所組成的資料集，假設類別標籤屬性有 m 個不同的值，此 m 個不同的值定義 m 個不同的類別： C_i (for $i = 1, \dots, m$)。令 s_i 是 S 樣本資料集中，屬於類別 C_i 的樣本資料數。則此分類樣本資料的期望資訊為：

$$I(S_1, S_2, \dots, S_m) = - \sum_{i=1}^m P_i \log_2(P_i) \quad , P_i = s_i/S$$

公式中的 P_i 是屬於類別 C_i 中任意的樣本資料的機率。

4：計算所選擇的測試屬性的期望資訊：

令測試屬性 A 有 v 個不同的值： $\{a_1, a_2, \dots, a_v\}$ ，屬性 A 能將 S 分割成 v 個子資料集： $\{S_1, S_2, \dots, S_v\}$ ，式中 S_j 是包含 S 樣本資料中測試屬性 A 的 a_j 值。令 s_{ij} 是類別 C_i 中資料子集 S_j 的樣本資料數。則此測試屬性的期望資訊為：

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{S} I(s_{1j}, s_{2j}, \dots, s_{mj})$$

公式中的 $I(S_1, S_2, \dots, S_m) = - \sum_{i=1}^m P_i \log_2(P_i)$ ， $P_i = s_i/S$

公式中的 P_{ij} 是在 S_j 中屬於類別 C_i 中任意的樣本資料的機率。

5：計算所選擇的測試屬性的資訊增益：

$$\text{Gain}(A) = I(s_1, s_2, \dots, s_m) - E(A)$$

6：重複 2 到 5，直到完成計算所有測試屬性的資訊增益

7：選取最大資訊增益的測試屬性，作為決策樹的分割節點

8：依照上述步驟，依序找出下一階層的各節點的測試屬性，作成決策樹
上述步驟如圖 14 所示。

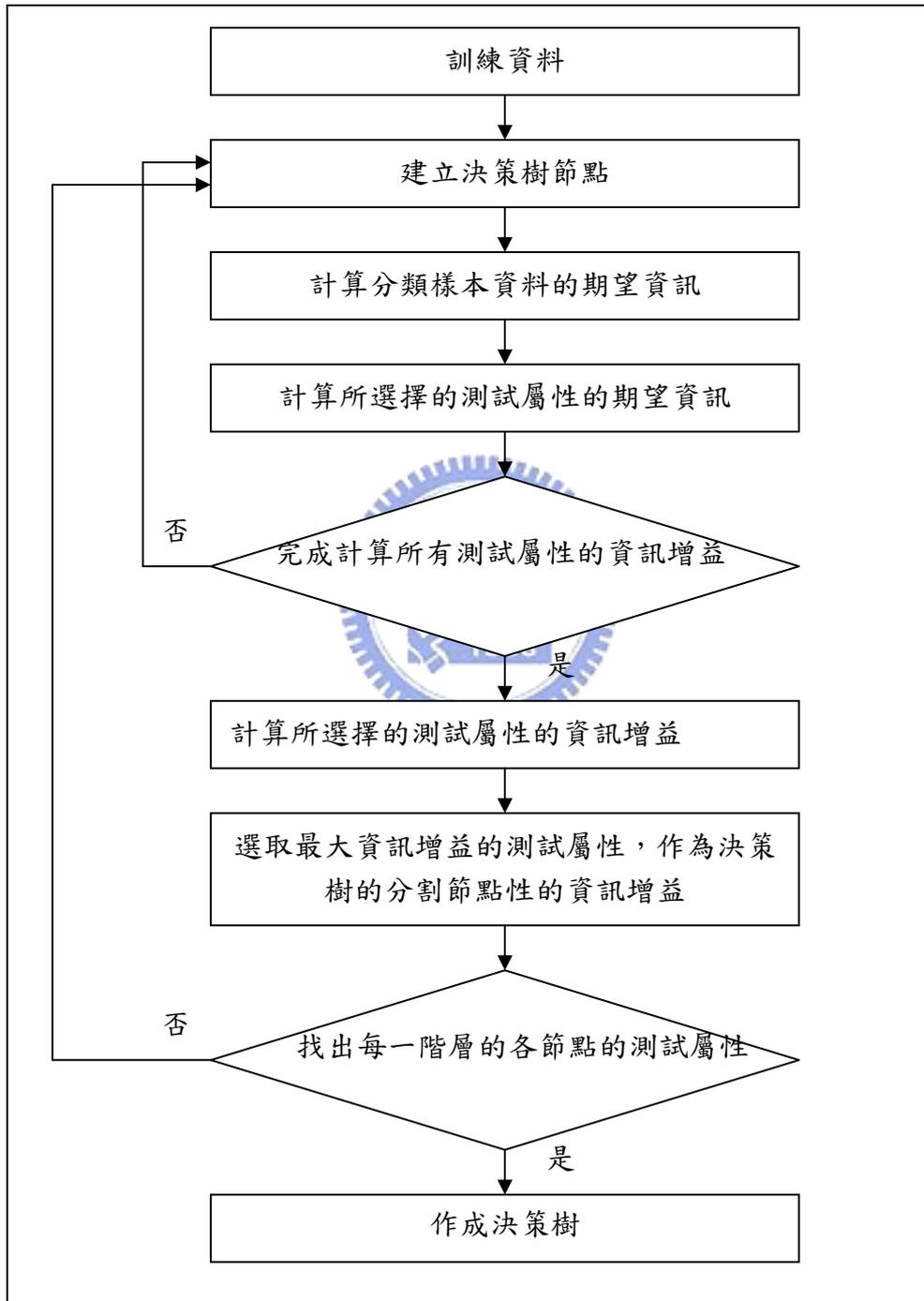


圖 13 資料探勘演算步驟流程圖

第五章 智慧型 CIM 系統之實作 與分析

5.1 智慧型 CIM 系統之環境與架構

智慧型 CIM 系統之環境與架構，如圖 14 所示。包括 CIM Server、Data Warehouse Server、Data Mining Server、Web Server 及前端 PC，其扮演的功能與角色，說明如下：

1. CIM Server：

包含了 CIM 系統、CIM 系統資料庫與應用程式三部份，主要是作為 CIM 資料庫管理系統、CIM 應用系統運作之環境，系統用以詳實記錄每日製造相關資訊，為 CIM 系統之核心。

2. Data Warehouse Server：

包含企業資料倉儲(BIDW)、企業資料超市(BIDM)及企業資料立方體(BIDC)運作之環境，資料倉儲用以收集、整合系統所有的歷史交易資料，資料超市將資料倉儲之資料作分割作為 OLAP 之資料來源，依企業各部門之需求功能作分割儲存，資料立方體則作為 OLAP 操作運算之環境。

3. Data Mining Server：

為資料探勘運作之核心所在，內含資料探勘演算法，主要是作為資料探勘系統運作之環境，透過各項資料探勘功能與技術的運作，從大量的歷史資料中分析統計後推算出預測結果。

4. Web Server：

主要是作為網際網路系統運作之環境，讓使用者可經由網際網路來操作

資料探勘系統，讓 Users 來瀏覽擷取出傳送至 Web Server 上之資料

5. 前端 PC：

使用者端的電腦，內含 WEB Browser 讓使用者能夠連結至 WEB Server 來執行資料探勘系統及 CIM 系統

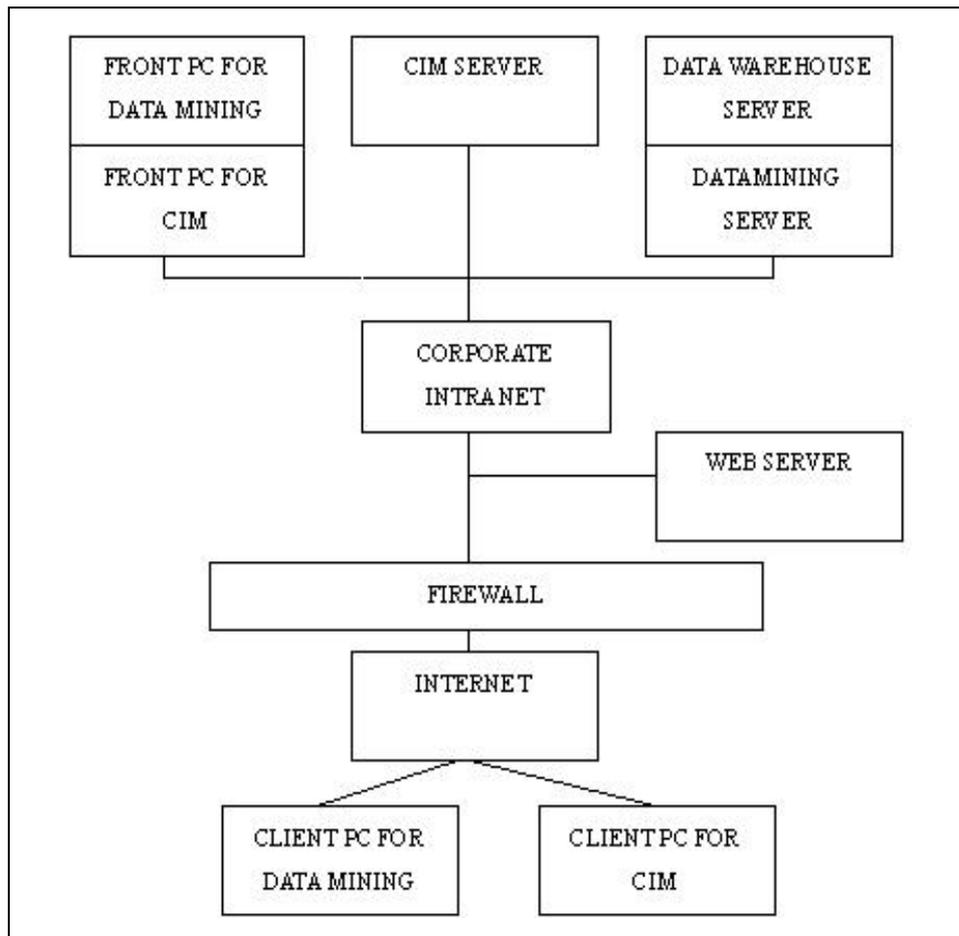


圖 14 智慧型 CIM 系統架構圖

5.2 結果比較與分析

5.2.1 效益分析比較

本研究之智慧型 CIM 系統，應用於半導體測試封裝的品質改善。製程領域中發生的問題，常常會導致降低產品良率，增加製程時間、人力與成本來解決品

質問題。本研究利用資料探勘系統，以製程過程中所產生的龐大品質記錄為基礎，來產生正確的品質情報知識規則，進而提昇整體營運績效。其提昇產品良率之效益說明如下：

1. 降低產品故障率

在使用智慧型 CIM 系統前的，平均的改善率為 30.0%。

2. 降低零件不良率：

本研究所應用之智慧型 CIM 系統，平均改善率為 36.0%。



第六章 結論與未來研究方向

6.1 結論

本研究中實際建立了一個智慧型 CIM 系統，其中包含了資料倉儲的應用，資料探勘引擎的應用及 CIM 流程的設計，並且將本研究所設計的系統實際應用在半導體封裝測試業。根據實際應用的經驗得到了以下的成果：

1. 建立了智慧型 CIM 系統的架構：

在智慧型 CIM 系統架構中，由 CIM 系統、資料探勘、資料倉儲、線上分析處理等組成，以企業資料庫所積蓄的既有龐大資料為基礎，從中獲取有價值的企業決策資訊。讓企業的電腦化應用，不會單純地侷限在資料處理層面，而能積極地邁向智慧型的資訊管理系統。

2. 整合 CIM 系統與資料探勘系統

整合 CIM 系統與資料倉儲，讓生產線上每日各項活動所累積的龐大資料，得以有效整合、保存至資料倉儲系統中，以作為決策輔助之珍貴資料來源。整合資料倉儲和決策分析讓管理者進行各種不同需求與面向的瀏覽與查詢，以獲取需要的決策分析的資訊。藉由整合決策分析與資料探勘系統，讓 OLAP 的分析結果可以補充到資料探勘系統的知識庫中，以作為資料探勘的依據。OLAP 技術亦可將人的觀察力和智力融合入資料探勘系統中，以改善資料探勘系統的採掘速度與深度。而資料探勘所發現的知識，可以拓展 OLAP 分析處理的深度，以便發現 OLAP 所不能發現的，而且是更為複雜而細密的資訊。

3. 使用智慧型 CIM 系統降低了產品不良率：

本研究將智慧型 CIM 系統，應用於製程中不良率的品質改善。本研究中以資料庫系統中所儲存的製程資料為基礎，經由智慧型 CIM 系統的資料

探勘引擎的挖掘，找出最佳的結果。在使用智慧型 CIM 系統前的產品故障率平均的改善率為 30.0%。另一方面，使用智慧型 CIM 系統也改善了零件不良率，平均改善率為 36.0%，

6.2 未來研究方向

在本研究中結合了數種企業的電腦化系統，在未來還有可以做後續的研究：

本研究提出了智慧型 CIM 系統的基礎架構，未來此架構可在各種面向整合其他企業內部的資訊系統，前端除了一般的 PC 外可在融入一些無線的 client，讓資訊的使用與取得更有彈性，後端可將挖掘出的資訊導入專家系統中，提高資訊的附加價值且充分的利用資訊，並且讓本系統的應用範圍更加的擴張。



參考文獻

- [1] 楊東麟、洪明傳，”資料探勘在資料倉儲的應用”，資訊與教育雜誌，2001/08
- [2] 沈兆陽，”資料倉儲與 Analysis Services：SQL Server 2000 OLAP 解決方案”，文魁資訊公司,2001/03
- [3] 黃士銘，”建置一網際網路資料倉儲系統”，資訊管理學報，第九卷第一期
- [4] 沈肇基、張慶賀，”淺談資料倉儲”，資訊與教育雜誌，2001/08
- [5] 林傑斌、劉明德、陳湘，”資料採掘與 OLAP 理論與實務”，文魁資訊公司，2002/05
- [6] 沈清正等，”資料間隱含關係的挖掘與展望”，資訊管理學報,第九卷專刊
- [7] 李金鳳，”資料探勘面面觀”，資訊與教育雜誌，2001/08
- [8] 陳安斌，”財務金融資訊管理與投資決策”，寶基資訊公司，2002/01
- [9] 蔡永順，”應用資料探勘技術於智慧型企業資源規劃系統”，國立交通大學資訊管理研究所碩士論文 2002
- [10] 范滄仁，”以全球資訊網為基礎之 CIM 資料探勘系統整合架構 - 以半導體封裝廠為例”，國立交通大學資訊管理研究所碩士論文，2002
- [11] 郭哲宏，”設計一種以 IC 封裝/測試業為主的自動化生產排程系統”，國立交通大學資訊管理研究所碩士論文，2001
- [12] 葉坤杰，”ERP 離型系統導入 IC 測試業之研究”，國立交通大學資訊管理研究所碩士論文，2001
- [13] Peter F. Drucker, Ikujiro Nonaka, David A. Garvin, ”Harvard business review on knowledge management”, Harvard Business School Press, 1998
- [14] W.H.Inmon, ”Building The Data Warehouse ”, John Wiley & Sons, 2002
- [15] Zhengxin Chen, ” Data Mining and Uncertain Reasoning：An Integrated Approach ”, John Wiley & Sons,2001
- [16] Alex Berson, Stephen Smith, Kurt Thearling, ”Building Data Mining

Applications for CRM ”, McGraw-Hill,2000

[17] Runyan, W. R. and Bean, K. E., “Semiconductor Integrated Processing Technology”, Addison Welsey, 1990.

[18] Sze, S., “VLSI Technology”, 2nd ed., McGraw Hill, 1988.

[19] Jiawei Han, Micheline Kamber, ”Data Mining : Concepts And Techniques”, Morgan Kaufmann Publishers, 2001.

