



第二章 文獻探討

第一節 資料庫廣為盛行的緣由

在八〇年代以前，電腦皆採用「檔案處理系統」的方法來處理資料。其處理方法是依據每一企業組織各自的需求來設計程式，再根據所寫的程式去設計所需要的檔案結構，而不考慮企業組織整體的需求。在此發展模式下，每一組程式和檔案自成一系統，和其他系統互不相關，所以各系統使用的程式語言和檔案結構亦可能不同，因而增加系統維護的困難度。而且在此發展模式下，每當有新需求產生，便撰寫新的程式和建立新的檔案，而不管該檔案資料是否早已存在於檔案處理系統中，造成資料重複和資料不一致的問題。

現在採用「資料庫方式」來處理資料已相當普遍。企業組織在發展一個系統時，分析考量整體需求，將相關資料都用相同的資料結構儲存在資料庫之中，讓不同使用者皆可利用資料庫來開發所需的應用程式。此可避免資料重複和資料不一致的情況。且利用資料庫提供的管理功能可對不同使用者設定不同使用權限，加強資料安全性。利用其資料檢查功能，易於維護資料的正確性。利用 SQL 查詢功能，可直接擷取所需資料，不需額外撰寫程式。有上述諸多好處，故資料庫廣為盛行[2]。

第二節 何謂資料倉儲

資料倉儲(Data Warehousing) 係運用新資訊科技所提供的大量資料儲存、分析能力，將以往無法深入整理分析的客戶資料建立成為一個強大的客戶關係管理系統，以協助企業訂定精準的營運決策。「資料倉儲」對於企業的貢獻在於「效果」(Effectiveness)，能適時地提供高階主管最需要的決策支援資訊，做到「在適當的時間將正確的資訊傳遞給適當或需要的人。」簡單地說，就是運用資訊科技將寶貴的營運資料，建立成為協助主管做出各種管理決策的一個整合性「智庫」，利用這個「智庫」，企業可以靈活地分析所有詳細深入的客戶資料，以建立強大的「客戶關係管理」優勢。以下參考[3]說明資料倉儲各種相關內容。

一、資料倉儲的特性

資料倉儲與傳統的資料庫是有所不同的。資料庫是未經整理後的一大堆資料集結；而資料倉儲是從資料庫中萃取出來，經過整理、規劃、建構而成的一個有系統的資料庫的子集合。資料倉儲具有下列幾種特性：

1. 主題導向(Subject Oriented)

資料倉儲的資訊系統，資料建立的著重點在於以重要的主題元件為核心，做為建構的方向。資料需求者只要把要研究的相關主題資料，從資料庫中擷取、整合之後就可以做研究分析。

2. 具整合性(Integrated)

各應用系統的資料須經過整合，以便利執行相關分析作業。

3. 具長期性(Time Variant)

資料倉儲系統，為了執行趨勢的分析，常須保留 1 至 10 年的歷史資料。這與資料庫為日常性的資料有所不同。

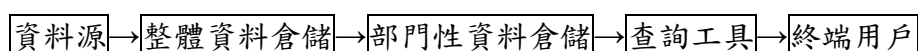
4. 具少變性(Non-Volatile)

資料庫 (Database) 的資料可以隨時被更動，但是資料倉儲的資料，並非日常性的資料而是歷史性的資料，通常做為長期性分析用途，只有內部相關人員會定期性的修改資料結構，但頻率不會太高，資料倉儲並不允許使用者去做更新的動作，所以其資料是較少有變動。

由於資料倉儲內的資料，具備上述特性，故須藉由一連串的程序(配合良好的軟硬體設備)始可建置完成，而非一個即買即可使用的產品。

二、資料倉儲的架構

建置資料倉儲(Data Warehousing)是一種能正確地組合與管理不同資料來源的技術，其目的在於回答業務經營上的問題以便做出正確決定。資料倉儲的處理架構如下：



三、建置資料倉儲的基本過程

專業顧問透過與企業需求訪談，建立資料倉儲的 Model，然後將企業內各種資料整合於資料庫中，並建置前端分析資料的工具以及管理工具，如此的過程即為建置資料倉儲的基本過程，區分為以下幾個基本元件說明：

1. Design :

即資料倉儲之資料 Model 的設計，這部分是最重要的，若 Model 設計的不夠周延或不理想，儘管之後的報表設計如何精美，也有可能跑出錯誤的資訊，這也是需要選擇有經驗的專業顧問建置資料倉儲的一個重要原因。

2. Integration :

即資料的整合轉換過程，包含資料擷取 (Data Extraction)、資料轉換 (Data Transformation)、資料清理 (Data Cleaning)、資料載入 (Data Loading)，而將各種來源之資料整合轉換載入資料倉儲中，資料轉換的程式撰寫不易，自動化處理困難，經常要人工參與作業，約占資料倉儲專案之 60% 至 70% 的人力及時間。

3. Management :

即資料倉儲之中心-是一個容量巨大及提供 ad hoc 查詢的資料庫。

4. Visualization :

即前端呈現給使用者觀看之型式，例如資料採礦 (Data Mining) 及線上及時分析 (On-Line Analytical Processing 簡稱 OLAP) 工具，用以呈現分析過之資料型式。

5. Administration :

為管理的工具，例如：網路監控流量、security 管理等

四、超大型資料倉儲 (Very Large DataBase, 縮寫 VLDB) 之定義

VLDB 並無官方正式或標準的定義，它有時被用來描述資料庫使用磁碟儲存兆位元組 (terabyte, 縮寫 TB) 以上和包含十億筆以上資料記錄。通常是決策支援系統或使用者數量龐大的交易處理應用系統。原文引用 *While there is apparently no official or standard definition for the term Very Large Database, it is sometimes*

used to describe databases occupying magnetic storage in the terabyte range and containing billions of table rows. Typically, these are decision support systems or transaction processing applications serving large numbers of users. [4]

目前中華電信固網通聯三區資料庫合計約 5TB，最大資料表記錄總計約 90 億筆，符合上述超大型定義。

五、建置資料倉儲的各種技術

各種建置資料倉儲的技術其著眼點均在於如何支援使用者從龐大的資料中快速地找出其想要的答案，這和線上交易處理（On-Line Transaction Processing 簡稱 OLTP）是截然不同的。這些技術包括：

1. 快速且擴充性高的資料庫系統（High performance, high scalability database system）
2. 異質資料庫的連結（Heterogeneous database connectivity）
3. 資料萃取轉換與載入（Data extraction, transformation, and loading）
4. 多維度資料庫設計（Multi-dimension database design）
5. 大容量資料儲存系統（Mass storage system）
6. 高速網路（High speed network）
7. 支援隨意查詢（Ad hoc query support）
8. 友善的前端介面（User-friendly front end interface）
9. 資料挖掘（Data mining）

第三節 現有中華電信固網通聯系統研究

如圖 2.3.1 所示為一套 CDRS(全名 Call Detail Record System)架構（非全區），一個營運處一套，相同架構設備共 32 套，構成現行固網通聯系統（全區）。每套包括通話記錄收集萃取處理伺服器（HP 9000 主機），和資料庫載入處理暨

資料庫伺服器(HP 9000 主機，簡稱 DB Server—CPU 3 顆、RAM 1GB、資料庫 Oracle8)。通話記錄收集萃取處理伺服器將所負責各交換機各種不同格式資料，先初步整理萃取出相同格式且有意義資料，再送到資料庫伺服器，由資料庫伺服器再做進一步資料檢查、篩選及載入資料庫，也同時更新各實體視域。

CDRS 系統架構

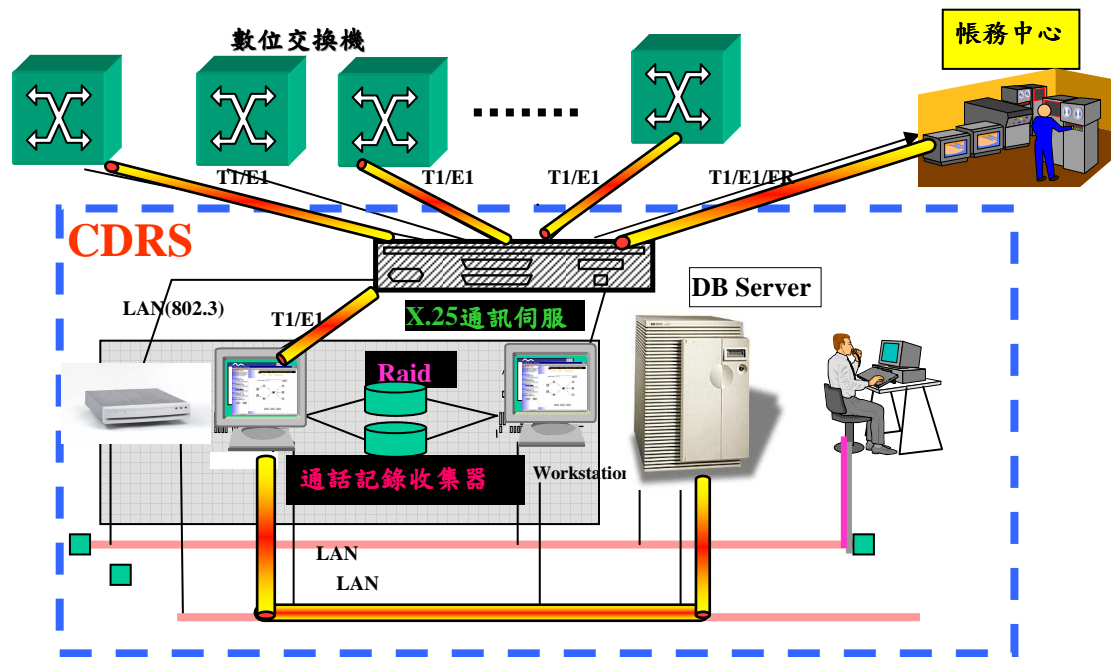


圖 2.3.1 現有中華電信固網通聯系統架構圖[中華電信內部資料]

除上圖所示相同設備共 32 套之外，尚有一台固網通聯查詢應用程式伺服器(簡稱 AP Server)及三台決策分析支援 Web 伺服器，連結此 32 台 DB Server 擷取資料。

現有系統效能除整合性資訊外，大致上還算不錯。但整合性之資訊取得(如全區或區分公司)，不僅查詢效能不太好，而且因資訊分散在 32 台 DB Server 整合不易，往往需撰寫大量且複雜的應用程式，造成維護成本增加。同時，無法即時提供臨時需要的資訊。更值得注意的是，同時管理 32 台 DB Server 著實不易，也缺乏完善的備援及回復機制，若發生任何 DB Server 掛掉，需重灌、重 Load，嚴重影響資訊提供的完整性。因此我們仔細觀察及研究原有系統，目的在瞭解其優缺點，希望優點得以延續，缺失得以改善。