

#### 四、資料倉儲與資料超市

➤ 資料倉儲 (Data Warehouse)：組織中資料集中儲存的倉庫，管理者可以透過資料分析工具動態即時檢視其中的資料，以支援決策制定。

##### 1. 特性 (整 no 差主)：

- (1) 整合性 (Integrated)：整合組織內外、不同時間/來源/型態的資料。
- (2) 不變動性 (Non-Volatile)：資料一旦存入後便不可更改。
- (3) 時間差異性 (Time Variance)：強調隨時間變化的動態資料，目前與歷史資料並存。
- (4) 主題導向 (Subject Orient)：根據決策主題設計資料架構。

##### 2. 資料前置處理：

- (1) 資料整合 (Integration)：解決多重資料來源的整合問題
  - a. 消除不一致：數值不一致、綱要不一致
  - b. 消除重複性：數值重複、綱目重複
- (2) 資料清洗 (Cleansing)：對資料進行修正與填補
  - a. 正確性修正：
    - (a) 數值內容有效、合理
    - (b) 主鍵唯一性
    - (c) 參考完整性
  - b. 完整性修正：
    - (a) 確認數值資料是否滿足探勘需求
    - (b) 是否可鑽研詳細資料
  - c. 遺失填補：直接忽略、人工填補、自動填補、推論
  - d. 雜訊修正：消除雜訊、平順化 (smoothing) 雜訊
- (3) 資料轉換 (Transformation)
  - a. 資料統整 (aggregation)：加總、統計或是建立資料方塊 (data cube)
  - b. 資料一般化 (generalization)：資料的概念階層向上提升
  - c. 建立新屬性 (attribute construction)：利用舊屬性將探勘所需的新屬性建立

##### 3. 資料分析：線上分析處理 (On-Line Analytical Processing, OLAP)

- (1) 定義：是一套以多維度方式分析資料，呈現整合性決策資訊的方法或工具。其主要的功能，在於方便大規模數據分析及統計計算，提供資料以利決策。
- (2) 操作：
  - a. Roll-up：藉由提升資料階層、維度刪減獲取總和性資料。
  - b. Drill-down：從大範圍資料，鑽研到更詳細的細節。
  - c. Slide and Dice：取出某一個維度或某一部分資料進行分析。
  - d. Pivot (Rotate)：旋轉 Cube 或將其視覺化 (Visualization)。

版權所有，重製必究

4. 邏輯設計：一個資料倉儲是由事實資料與維度資料所組成的，事實資料是能夠反應過去事實的資料，而維度資料則是為了使查詢更加快速而建立的索引參考資料。設計可分為3種：

(1) 星狀綱要 (Star Schema)：最常見的類型，每個維度都儲存在單一資料表中。一個中心事實表格有大量不重複的資料，以及較小的附屬維度表格。

(2) 雪花式綱要 (Snowflake Schema)：是星狀綱要的變形，部份維度表格經正規化分裂成新的維度表格。與星狀綱要最主要的差異是維度表格被正規化，容易維護及節省儲存空間。

(3) 事實星座綱要 (Fact Constellation Schema)：以多個的事實表格共用維度表格，使用到多個事實表格及需要不同階層的彙總資料。

5. 實體儲存架構：

(1) ROLAP (Relational OLAP)：利用傳統的關聯式資料庫來存放彙總資料。

a. 優點：彈性佳、建檔速度快、適用於中大型資料倉儲。

b. 缺點：查詢速度慢、需要使用 SQL 查詢、硬體設備要求高。

(2) MOLAP (Multidimensional OLAP)：真正的多維度資料庫結構，可以存放彙總資料。

a. 優點：查詢速度快、硬體設備需求低、較簡單好學、分析能力強、易於維護。

b. 缺點：資料建構速度慢、架構缺乏彈性。

(3) HOLAP (hybrid OLAP)：為前兩者的混和體，在前兩者的優缺點之間取得平衡。

➤ 子集合：資料超市 (Data Mart)：

1. 定義：提供特定部門或定使用者使用的高度專注資料，為資料倉儲的子集合。

2. 建置原因：

(1) 降低成本與風險：規模較資料倉儲小，建置成本與難度較低。

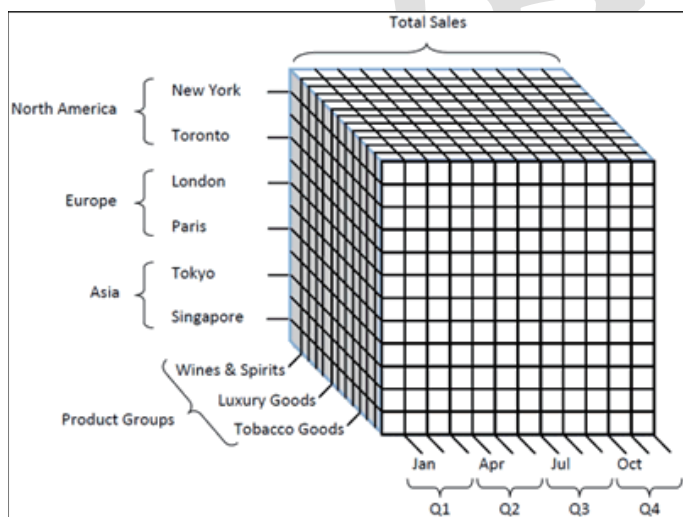
(2) 提升執行效率：只針對企業組織內的某子公司或部門的需要，建置時間較短、執行速度較快。

(3) 先導經驗：可以累積使用者管理資料倉儲的經驗，以利日後可以較快進入正式的资料倉儲投資的先導計畫。

➤ 資料庫、資料倉儲與資料超市比較表：

	資料庫	資料倉儲	資料超市
主要功能	提供操作交易資料 儲存	提供全組織決策分 析	提供個人或功能部 門決策分析
使用者	操作者、部門管理 者、客戶	高階主管	部門管理者、知識工 作者

使用時機	交易發生或查詢時	偶發性組織決策	偶發性部門決策
時間性	最新的操作資料	歷史資料	歷史資料
建置時間	1 到 2 個月	12 到 36 個月	4 到 12 個月
資料來源	交易處理系統	資料超市； 內部資料庫、外部資料庫、網際網路	資料倉儲； 內部資料庫、外部資料庫、網際網路
資料涵蓋	功能部門下特定交易處理系統的輸入	組織內外資料	特定功能部門或人員所需資料
資料型態	格式、型態固定	來自異質平台，需經過轉化才能使用	來自異質平台，需經過轉化才能使用
資料產出	報表、動態網頁	多維度資訊	多維度資訊
風險成本	低	極高	高
處理方式	SQL	OLAP	OLAP



圖片：Data Cube 提供多維度的資料分析，如將總銷售區分為時間、產品與區域等構面（來源：<http://zeesql.wordpress.com>）

### 範例 1

- (一) 企業在維護來自企業內外部的操作性資料時，經常發現資料存在許多問題，以致難以運用，請說明資料的問題有那些？（10 分）
- (二) 為使資料倉儲中的資料正確並符合所需，在資料倉儲中，資料的前置處理技術有那些？（15 分）

101 鐵路 4

解答：

- (一) 從資料前置處理需要處理的問題切入
- (二) 資料整合、資料清洗、資料轉換

範例 2

何謂資料倉儲？資料倉儲可能的資料來源有那些？在把資料存放於資料倉儲之前，需要進行那兩種資料處理動作？又何謂資料超市？為何需要建立資料超市？（30分）

99 地特

(一)何謂資料倉儲？請詳述之。(5分)

(二)何謂資料超市？請詳述之。(5分)

105 身心 4

解答：

(一) 資料倉儲定義

(二) 資料來源通常包含組織內部與外部的各種資訊系統所產生的有用資料

1. 內部系統：TPS、MIS、DSS、ERP、SCM、CRM…等。

2. 外部系統：

(1) 外部資料庫，如市場調查、人口統計資料

(2) 網際網路、Extranet。

(三) 資料清洗、資料轉換

(四) 資料超市定義

(五) 資料超市建置原因

範例 3

請說明何謂資料超市 (Data Mart) 及資料倉儲 (Data Warehouse)？並請詳細比較資料超市、資料倉儲及資料庫。(25分)

104 高考 2

請說明什麼是資料倉儲 (data warehouse) 以及資料超市 (data mart)，並說明若只建立資料超市是否可以解決資訊整合的問題？(30分)

97 地特 4

試定義資料倉儲 (Data Warehouse) 與資料市集 (Data Mart) 並說明兩者之間的差異性。(25分)

99 身心

解答：

(一) 資料倉儲定義

(二) 資料超市定義

(三) 資料超市只能解決部分資訊整合的問題；資料倉儲與資料超市以「優缺點」比較