

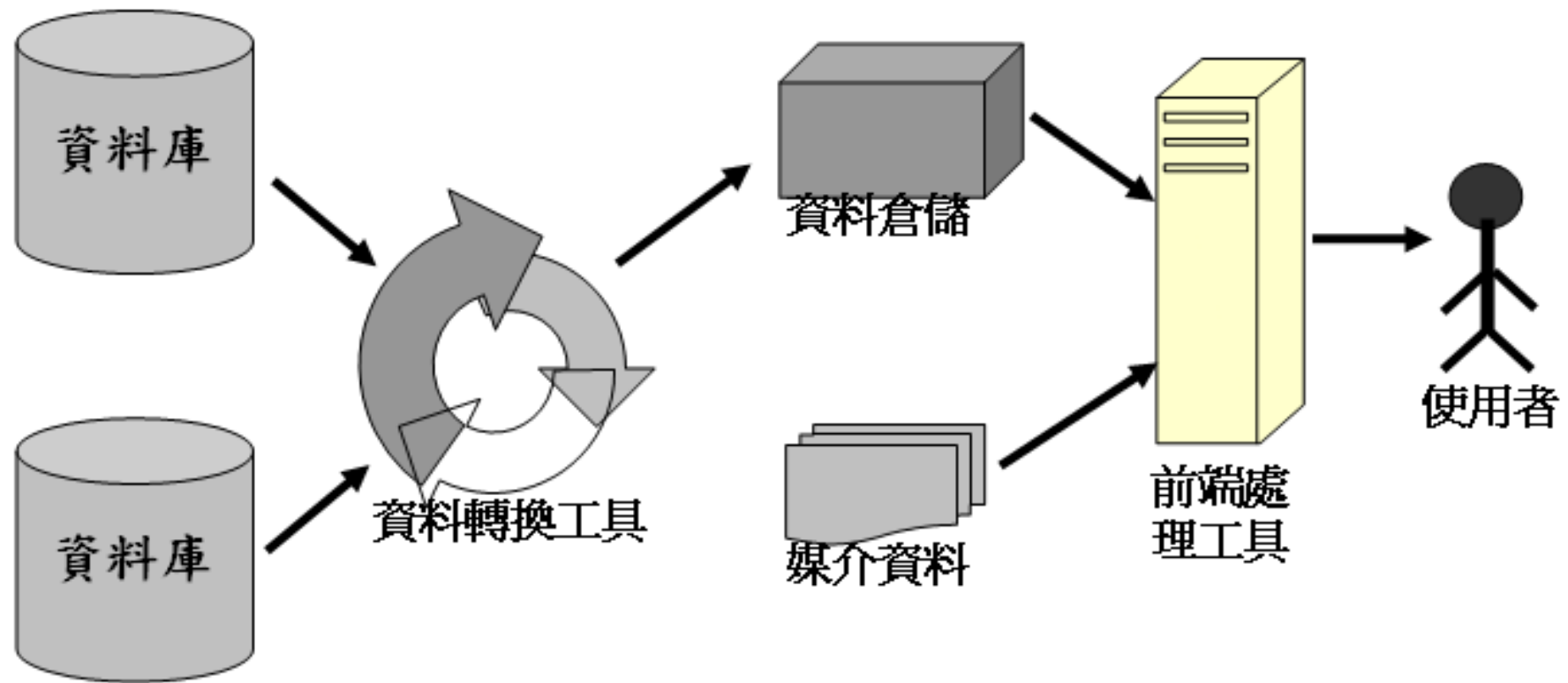
# 現今資料常遇到的問題

1. 巨量的記錄，流通筆數已大到令人無法想像。
2. 高維度的資料，資料的維度，增加傳統分析技術的難度。
3. 蒐集到的眾多資料，卻只採用一小部分來分析（5%~10%）。
4. 蒐集資料的過程中，本身並不具有延伸性的探討特性，以導致常常忽略其未來潛在的重要性。

# 資料倉儲 ( Data Warehouses )

- 資料倉儲 ( Data Warehouses ) 本身是一個非常大的資料庫，儲存著由企業作業型資料庫中整合而來的不同資料，特別是從線上交易處理系統 ( OLTP ) 所得的資料與從分散、異質資料來源傳遞而來的資料
- 建置資料倉儲的**最終目的**就是讓組織能夠有效地得到有關的資料，進而從這些分析中**增加**本身的**競爭力**
- 定義：「一份為了**查詢**與**分析**為目的建置的系統。系統中包括所有對組織做決定有用處且經過整合的資料。」

# 資料倉儲架構



# 資料倉儲系統

- 在以下的情況下一個組織需要建置一個資料倉儲系統：
  - 組織內**資訊**沒有**有效地共享**。
  - 不同部門產出**互相抵觸**的**報表**。
  - 報表獲取的步驟很**複雜**。
  - 報表無法靈活運用。
  - 需要歷史記錄的報表不易產生
- **資料倉儲系統的利益**
  - 將不同來源的資料整合
  - 將線上操作系統與決策輔助系統分開
  - 處理大量的資料
  - 提供一個提取資料的平台
  - 在符合安全考量限制之下開放資料獲取的管道

# 資料倉儲的建置流程

1. **萃取**—依限定條件將資料由來源資料庫取出，轉換成資料倉儲之架構。
2. **合併**—將不同來源資料合併為單一資料庫，以利分析的進行。
3. **過濾**—挑選實際使用的資訊，避免重複及多餘的發生。找出相關的資料，並且刪除重複的項目。
4. **清理**—提昇資料的精確性與可靠度。同一資料在不同的資料庫可能具有不同的名稱，清理的工作就是為了避免此種情形的發生。
5. **轉換**—將原始資料傳至資料倉儲模型中，並轉換成資料倉儲之格式。
6. **整合**—計算產生倉儲系統所有之總結與衍生性資料，並以視覺化介面表達。將資料區分之後，合併成可以進行資料分析的格式。

# 線上交易處理

- 線上交易處理(On-Line Transaction Processing , OLTP)是指經由**資訊網路**與**資料庫**或**檔案**的結合，以交易資料進行即時處理，有別於傳統的批次處理。
- OLTP典型用在自動化的資料處理工作，其主檔案龐大、交易數量頻繁，常用於訂單輸入、銀行業務上，性質是**結構化**且**反覆性**。

# 線上分析處理

- **線上分析處理**(Online Analytical Proccession, OLAP)意指一些使用者在線上完成分析的作業，例如：在線上透過試算表與圖片的DSS 模式。OLAP與線上交易的不同，在於OLAP簽涉到許多關係複雜的資料項目。OLAP的目標之一即是分析這些關係，並找出模式、趨勢以及例外的條件

# 資料倉儲系統例子

- 通訊業

- 將所有顧客依照**貢獻度**區隔
- 找出即將離開的高貢獻度顧客，且主動地做出挽留的動作
- 依不同族群的消費習性來設計出對各族群的**促銷行動**
- 找出**高信用風險**的族群
- 瞭解**線路負荷**狀況

- 零售業

- 找出顧客群的消費趨勢，並對此趨勢做出適當的反應
- 瞭解每一間店的表現
- 增加產品庫存管理的效率
- 找出對顧客最有價值的商品
- 商店空間設計
- 每間商店進貨選擇
- 產品銷售分析



# 資料倉儲系統例子 (Cont.)

- **金融業**

- 列出顧客與組織中所有部門的交易以及接觸的歷史記錄。
- 找出顧客在投資、資產分配以及借貸狀態裏任何不均衡的情況，以降低組織的風險。
- 將所有顧客依據交易的記錄以及本身的屬性分為不同族群。
- 不同產品的銷售分析
- 分行營運分析

- **保險業**

- 保險經紀人管理
- 成本分析及控制
- 預測未來理賠的金額與數量
- 找出對現有顧客賣更多保單的機會
- 新產品設計
- 理賠分析