



第1章 資料庫的基礎

- 1-1 資料與資料處理
- 1-2 資料庫
- 1-3 資料管理系統
- 1-4 資料庫系統發展的歷史演進
- 1-5 資料庫技術的發展趨勢





1-1 資料與資料處理

- 1-1-1 資料與資訊的基礎
- 1-1-2 資料處理
- 1-1-3 資料階層





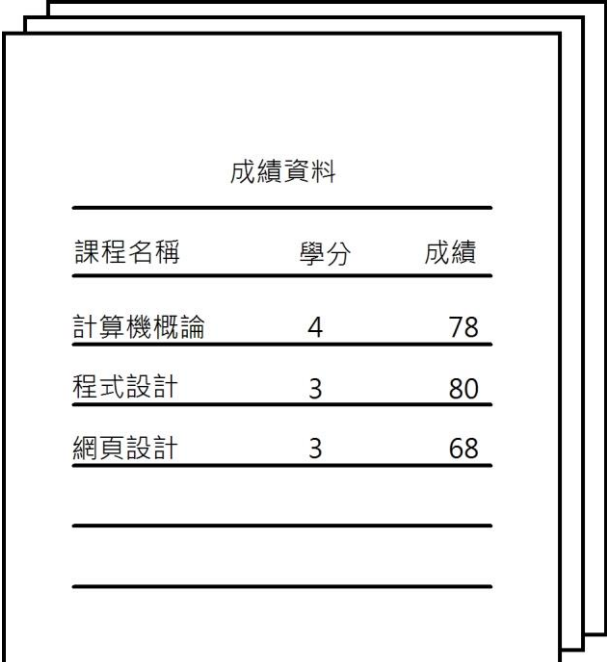
1-1-1 資料與資訊的基礎 – 資料(定義)

- 「資料」(Data)是指收集但沒有經過整理和分析的原始數值、文字或符號，屬於資訊的原始型態。「美國國家標準局」(American National Standards Institute; ANSI)定義的資料，如下所示：
 - 資料是使用定義語法或規則所描述的事實、概念或指令，可以適用在人類或程式間進行通訊、解釋和處理。
 - 資料代表一些特性或原始數值，換句話說，我們可以針對資料執行一些操作，將資料轉換成有意義的資訊，這就是「資料處理」(Data Processing)。



1-1-1 資料與資訊的基礎 – 資料(範例)

- 例如：學校整班學生必修課程的一疊成績資料，如下圖所示：



成績資料

課程名稱	學分	成績
計算機概論	4	78
程式設計	3	80
網頁設計	3	68



1-1-1 資料與資訊的基礎 – 資訊

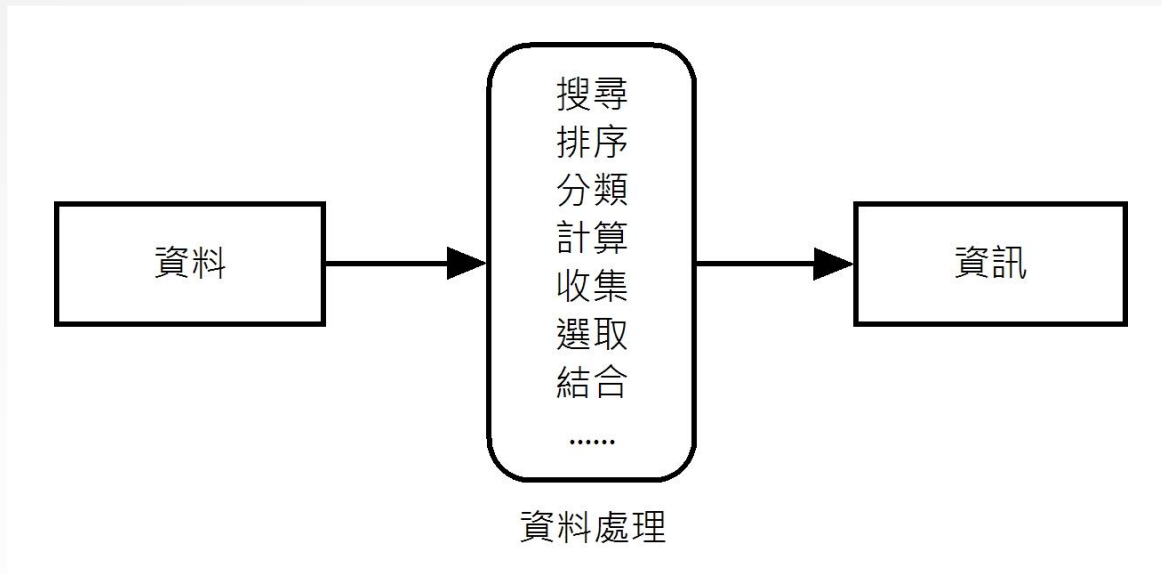
- 「資訊」 (Information) 是經過處理的資料，資料在經過整理和分析後，可以成為有用或可供決策的資訊。換句話說，資料是資訊的原始型態；資訊是處理後有實質意義的資料，如下圖所示：





1-1-2 資料處理

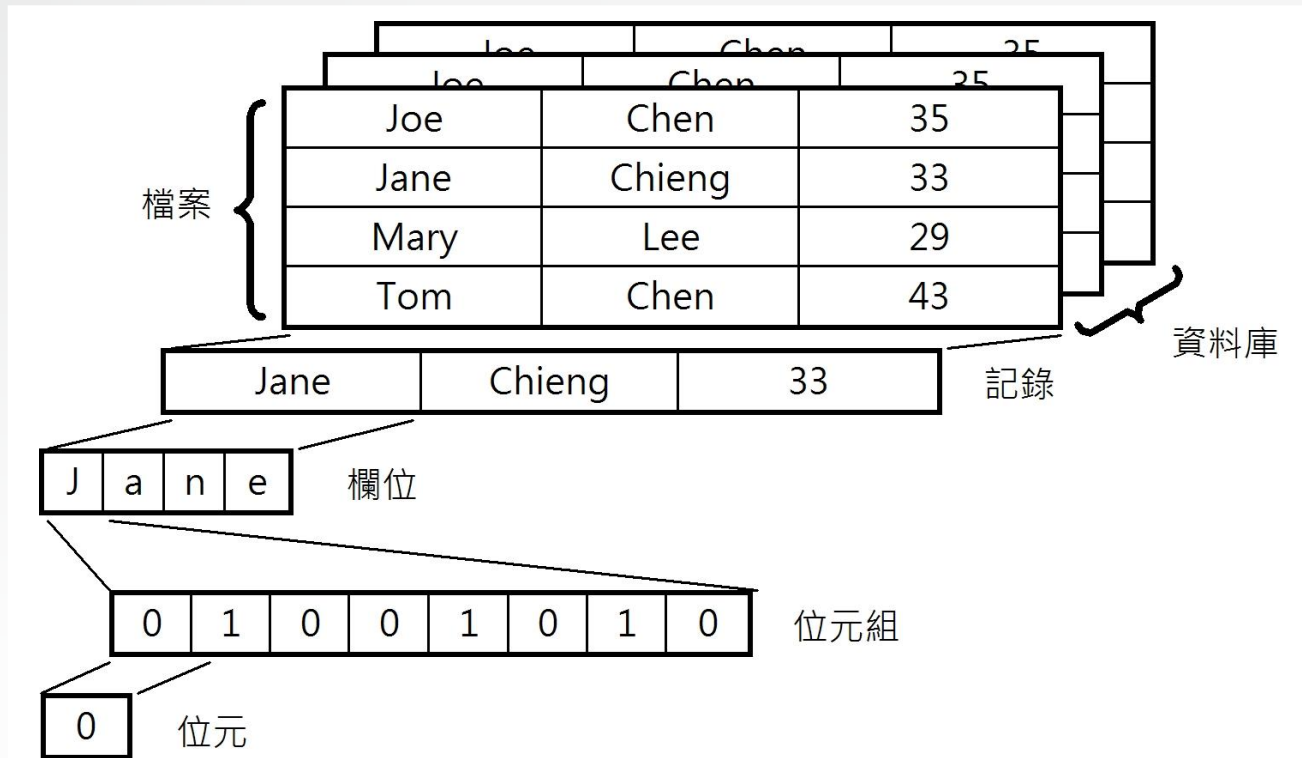
- 「資料處理」(Data Processing) 是使用特定方法將資料轉換成資訊的過程，資料可以進行搜尋、排序、分類、計算、收集、選取或結合等操作，以便產生所需的資訊，如下圖所示：





1-1-3 資料階層 – 說明

- 資料階層共分成六個階層：位元、位元組、欄位、記錄、檔案和資料庫，如下圖所示：





1-1-3 資料階層 – 六個階層

- 資料階層的最小儲存單位是位元，8個位元組成一個位元組，也就是ASCII碼的字元。數個位元組結合成欄位，多個欄位組成記錄，最後將一組記錄儲存成檔案，資料庫就是一組相關檔案的集合。
 - 第一階層：位元（Bits）
 - 第二階層：位元組（Bytes）
 - 第三階層：欄位（Fields）
 - 第四階層：記錄（Records）
 - 第五階層：檔案（Files）
 - 第六階層：資料庫（Database）



1-2 資料庫

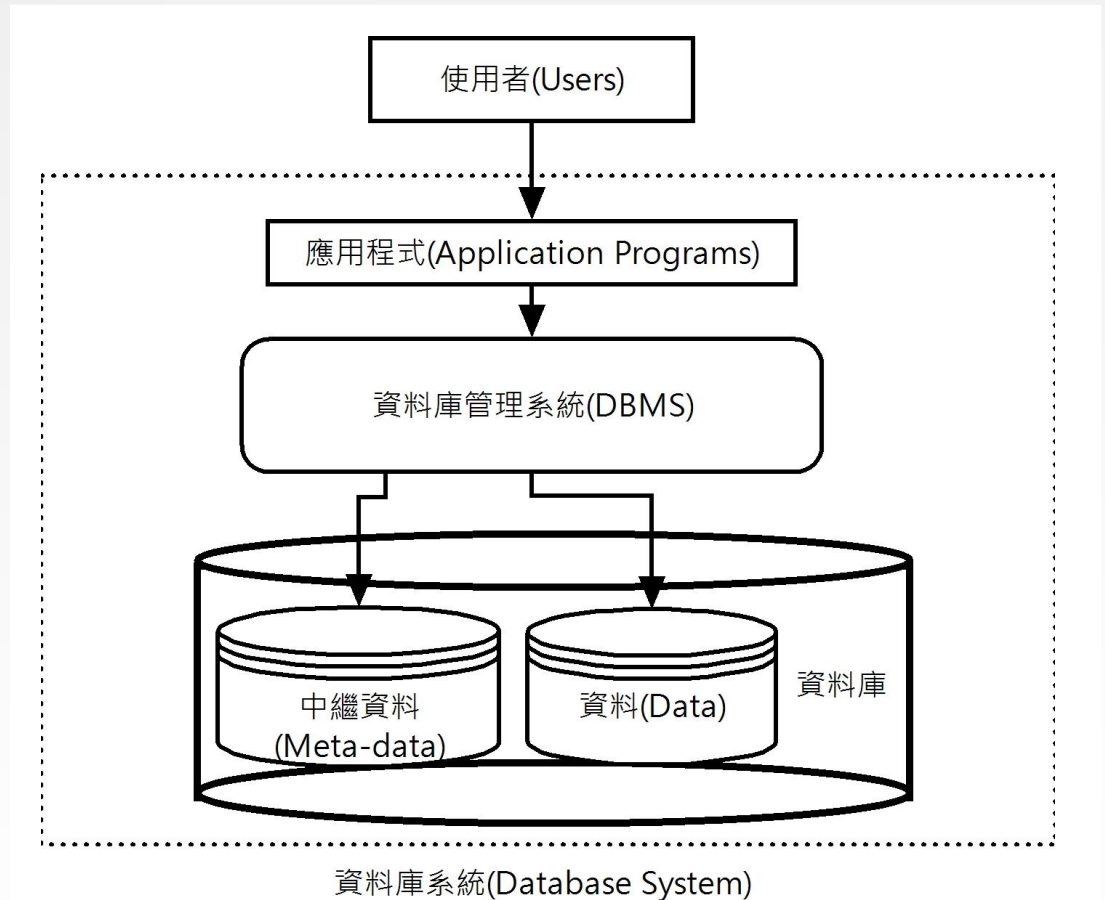
- 1-2-1 資料庫的定義
- 1-2-2 資料塑模
- 1-2-3 資料庫資料的三個層次





1-2 資料庫

- 資料庫系統是由「資料庫」（Database）和「資料庫管理系統」（Database Management System，DBMS）所組成，如右圖所示：





1-2-1 資料庫的定義 – 通用定義

- 資料庫正式的定義有很多種，比較通用的定義，如下所示：

定義1.1： 資料庫（Database）是一個儲存資料的電子文件檔案櫃（An Electronic Filing Cabinet）。

- 上述電子文件檔案櫃是儲存結構化（Structured）、整合的（Integrated）、相關聯（Interrelated）、共享（Shared）和可控制（Controlled）的資料。



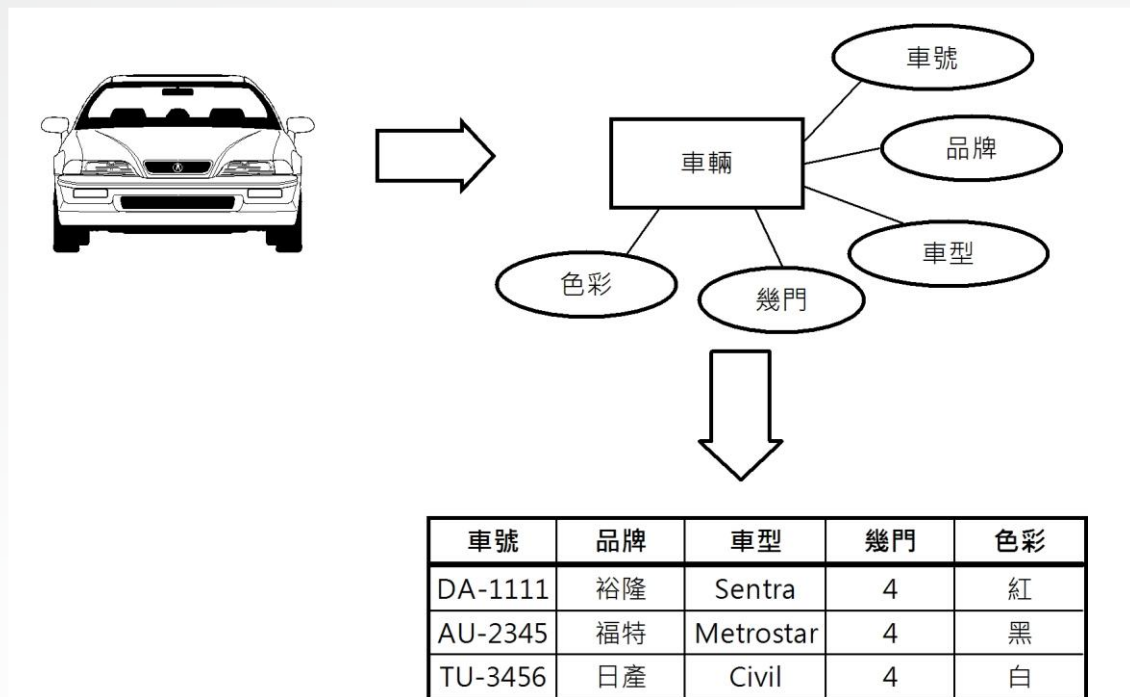
1-2-1 資料庫的定義 – 長存資料

- 在企業或組織的資料庫中儲存的大量資料，並非是一種短暫儲存的暫時資料，而是一種長時間存在的資料，稱為「長存資料」(Persistent Data)，它是維持企業或組織正常運作的重要資料，如下所示：
 - 在組織中的資料需要一些操作或運算來維護資料。
 - 資料是相關聯的。
 - 資料不包含輸出資料、暫存資料或任何延伸資訊。



1-2-2 資料塑模 – 說明

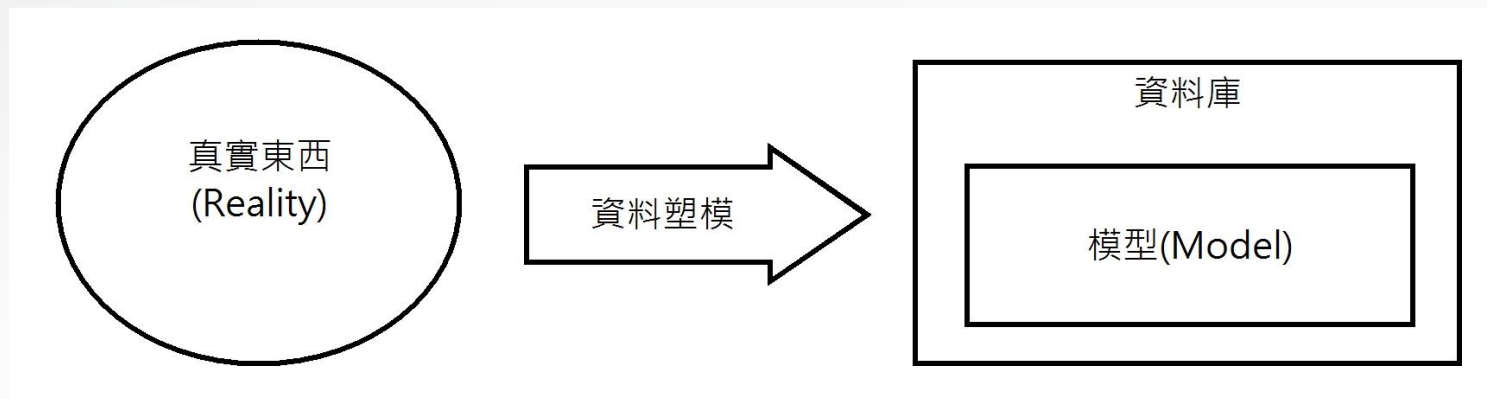
- 在資料庫儲存是結構化收集的「實體」(Entity)資料，實體是現實生活中存在的東西，我們可以將它塑模 (Modeling) 成資料庫儲存的結構化資料，如下圖所示：





1-2-2 資料塑模 – 基礎

- 「資料塑模」(Data Modeling) 是將真實東西轉換成模型，這是一種分析客戶需求的技術。其目的是建立客戶所需資訊和商業處理的正確模型，將需求使用圖形方式來表示，其塑模過程如下圖所示：





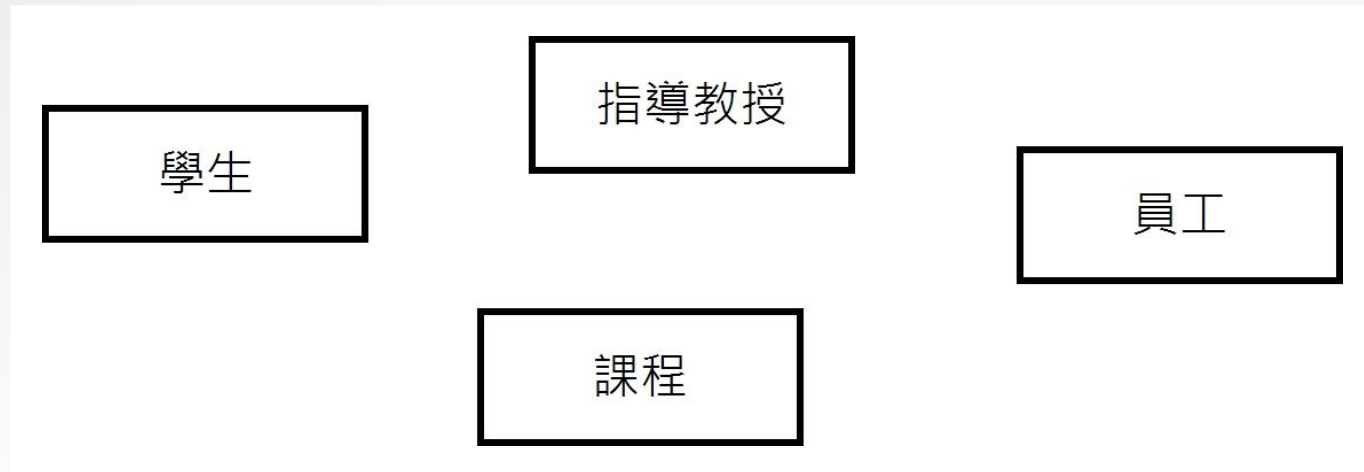
1-2-2 資料塑模 – 邏輯關聯資料

- 資料庫是將真實東西轉換成模型定義的資料結構
 - 例如：塑模一間大學或技術學院，也就是從大學或技術學院儲存的資料中識別出：
 - 實體
 - 屬性
 - 關聯性



1-2-2 資料塑模 – 邏輯關聯資料(實體)

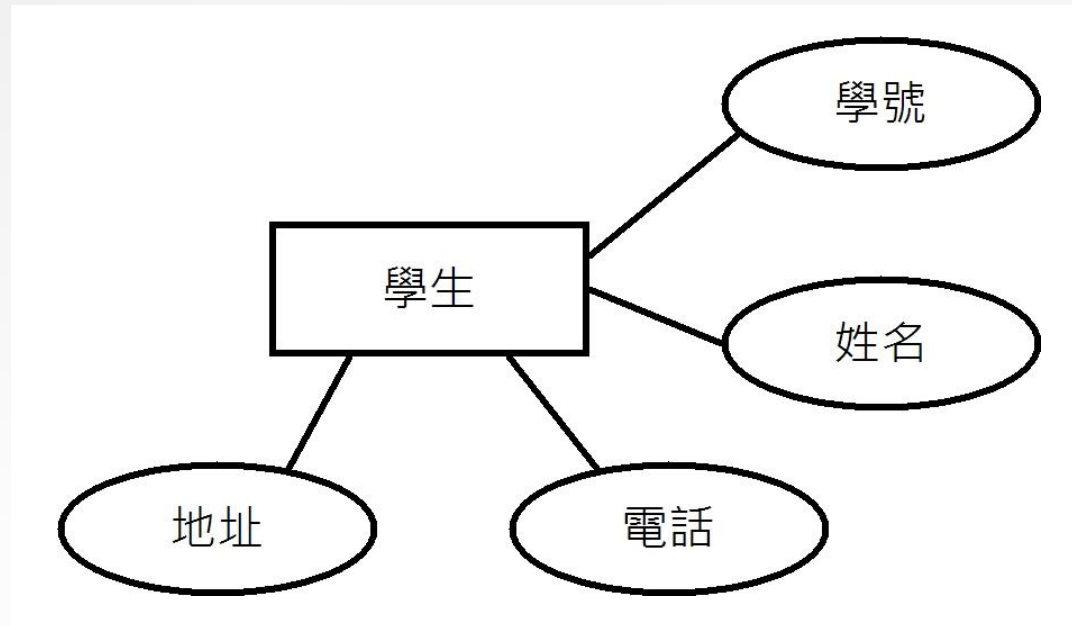
- 實體（Entities）是在真實世界識別出的東西。例如：從大學和技術學院可以識別出學生、指導教授、課程和員工等實體，如下圖所示：





1-2-2 資料塑模 – 邏輯關聯資料(屬性)

- 屬性（Attributes）是每一個實體擁有的特性。例如：學生擁有學號、姓名、地址和電話等屬性，如下圖所示：



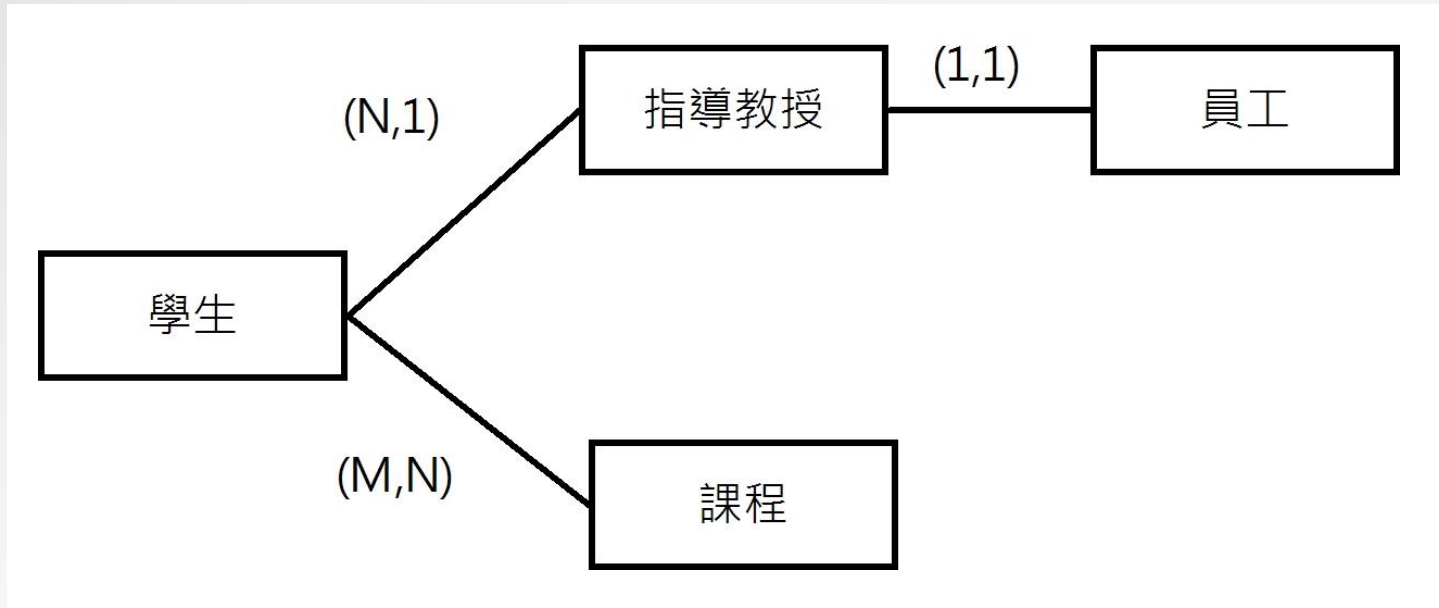


1-2-2 資料塑模 – 邏輯關聯資料(關聯性)

- 關聯性 (Relationships) : 在二個或多個實體間擁有的關係，主要分為三種，如下所示：
 - 一對一 (1:1) : 指一個實體只關聯到另一個實體。例如：指導教授是一位學校員工，反過來，此員工就是指這位指導教授。
 - 一對多 (1:N) : 指一個實體關聯到多個實體。例如：學生寫論文時可以找一位指導教授，而一位指導教授可以同時收多位學生。
 - 多對多 (M:N) : 指多個實體關聯到多個其他實體。例如：一位學生可以選修多門課程，反過來，同一門課程可以讓多位不同學生來選修。



1-2-2 資料塑模 – 邏輯關聯資料(關聯性圖例)





1-2-3 資料庫資料的三個層次 – 說明

- 在資料庫儲存的資料是使用模型找出的實體和屬性所轉換成的資料，可以分成三個層次，如下：
 - 資料模型（**Data Model**）：將真實東西轉換成資料模型的實體、屬性和關聯性，使用圖形化的高階模型來描述這些資料，通常使用在資料庫設計階段來分析資料庫儲存的資料。
 - 中繼資料（**Meta-data**）：這是用來描述資料庫儲存的是什麼樣的資料，定義資料列（**Rows**）或記錄（**Record**）型態，也就是定義各資料欄（**Columns**）或資料項目（**Data Item**）的型態。
 - 資料（**Data**）：資料庫實際儲存的資料列（**Rows**），或稱為記錄（**Records**）。



1-2-3 資料庫資料的三個層次 - 圖例

中繼資料

車號	品牌	車型	幾門	色彩
----	----	----	----	----

資料欄(Column)



DA-1111	裕隆	Sentra	4	紅
AU-2345	福特	Metrostar	4	黑
TU-3456	日產	Civil	4	白

資料列(Row)

資料



1-3 資料管理系統

- 1-3-1 資料管理系統的基礎
- 1-3-2 使用檔案處理方式
- 1-3-3 使用資料庫方式





1-3-1 資料管理系統的基礎

- 檔案處理和資料庫系統都屬於「資料管理系統」（Data Management System）的一環，資料管理是在探討組織、存取、更新和保存資料的方法，如下所示：
 - 結構（Structuring）：組織資料建立其資料模型（Data Model）或綱要（Schema），也就是儲存的資料結構。
 - 儲存（Storing）：依據建立的資料模型來儲存資料。
 - 取出（Retrieving）：取出資料以便進一步執行資料處理。
 - 更新（Updating）：更新儲存的資料。
 - 保存（Archiving）：長時間保存資料。



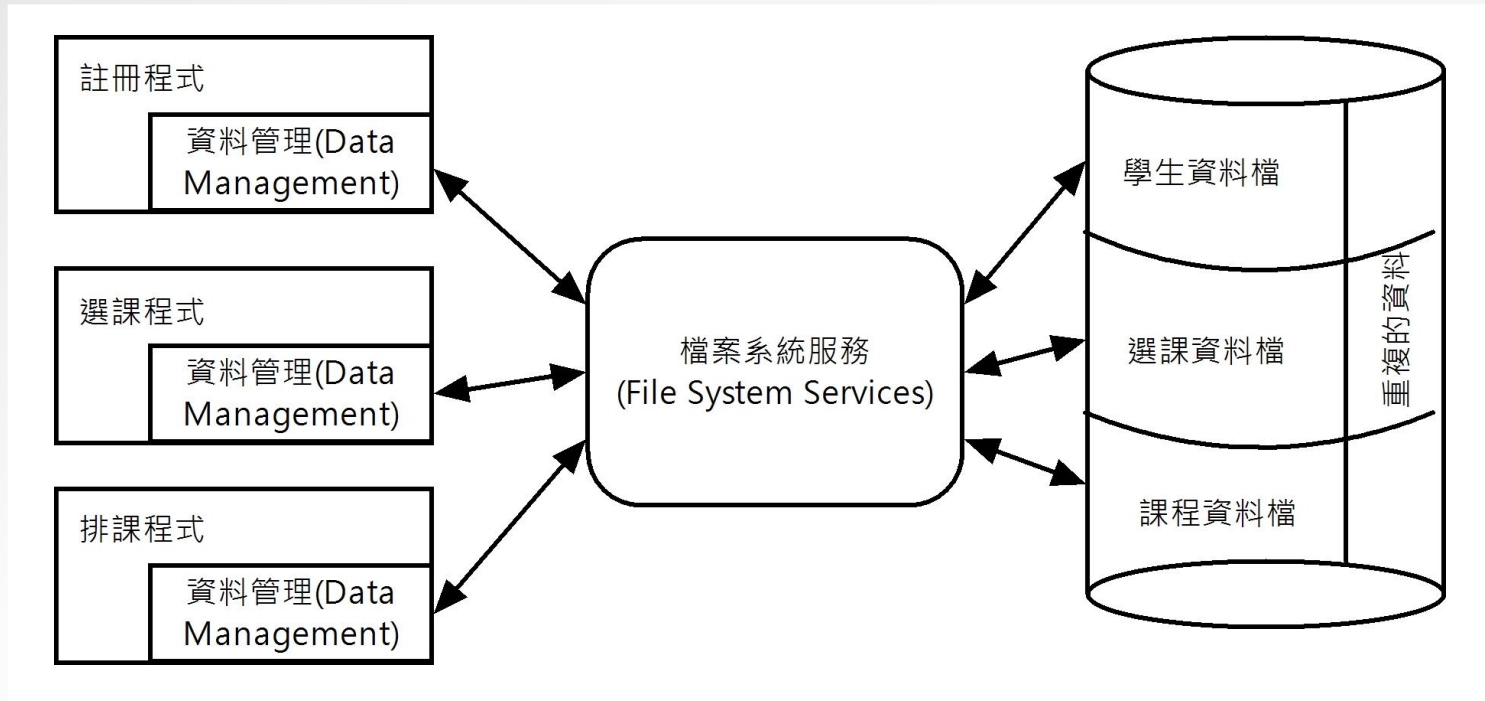
1-3-2 使用檔案處理方式 – 說明

- 「檔案系統」 (File System) 是一種非常原始的資料庫。不論是UNIX/Linux作業系統的檔案系統，或Windows作業系統的NTFS檔案系統，資料都是使用相同觀念，以檔案方式來儲存。程式設計師撰寫應用程式來處理檔案儲存的資料，稱為「檔案處理系統」 (File Processing System)。
- 檔案處理系統是使用檔案處理方式 (File Processing Approach) 建立的應用程式。早期在資料庫尚未出現的年代，大部分公司都是使用第三代程式語言建立應用程式，例如：COBOL語言，使用作業系統的檔案系統來儲存資料。



1-3-2 使用檔案處理方式 – 架構

- 傳統檔案處理系統一樣提供資料庫功能，只是處理的對象是儲存在檔案的資料，如下圖所示：





1-3-2 使用檔案處理方式 – 資料檔案的內容

- 檔案管理系統主要是處理邏輯檔案的資料，在學生註冊、選課和排課系統中的資料檔案是分散儲存在不同部門的電腦檔案，檔案格式可能是文字檔或試算表檔案。
- 實際儲存的文字檔案內容是以特殊分隔字元儲存欄位資料，如下所示：

S001:江小魚:中和景平路1000號:02-22222222:1978/2/2

S002:劉得華:桃園市三民路1000號:03-33333333:1982/3/3

S003:郭富成:台中市中港路三段500號:04-44444444:1978/5/5

S004:張學有:高雄市四維路1000號:05-55555555:1979/6/6



1-3-2 使用檔案處理方式 – 問題

- 結構與資料相關（Structural and Data Dependence）。
- 資料分隔與孤立（Data Separation and Isolation）。
- 資料沒有集中管理。
- 檔案格式不相容。
- 更新系統困難。
- 資料重複與不一致（Data Redundancy and Inconsistency）。
- 多使用者問題（Multiple Users Problems）。
- 安全問題（Security Problems）。
- 資料完整性問題（Data Integrity Problems）。



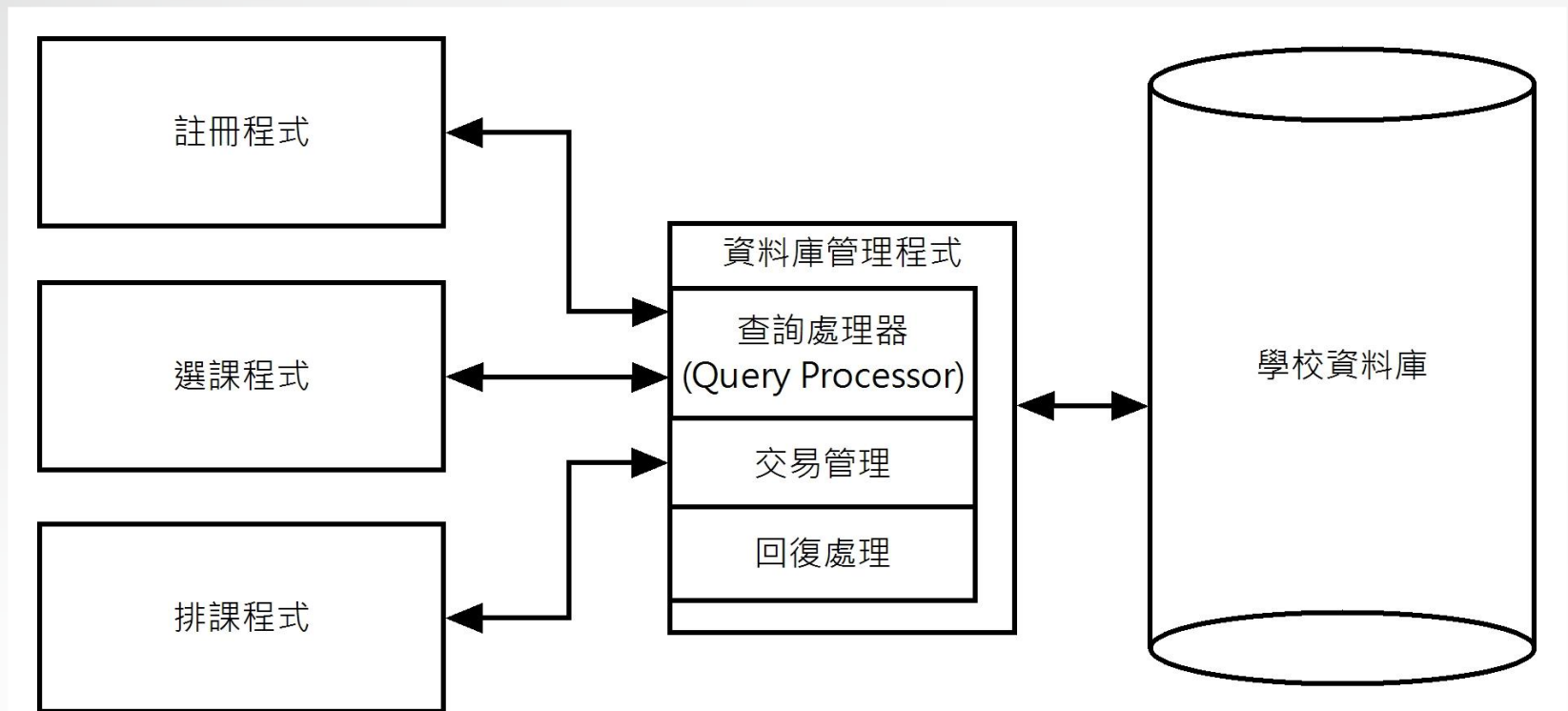
1-3-3 使用資料庫方式 – 說明

- 資料庫方式（**Database Approach**）建立的資料庫系統可以更有效率的管理和處理資料，解決檔案處理系統產生的問題。
- 在這一節筆者將比較檔案處理系統和資料庫系統架構的差異，詳細資料庫系統說明請參閱〈第2章：資料庫系統〉。



1-3-3 使用資料庫方式 – 架構

- 資料庫方式的學生註冊、選課和排課系統，如下圖所示：





1-3-3 使用資料庫方式 – 目的

- 在資料庫管理系統擁有多種程式模組：查詢處理模組、交易管理和回復處理等，可以進行資料庫的資料管理，將實際資料庫結構和存取都隱藏在資料庫管理系統之後，如此可以達到：
 - 多人使用，資料共享。
 - 資料一致和最少的資料重複。
 - 資料獨立（相反的是資料相關）。
 - 改進資料完整性問題。
 - 更佳的資料安全管理。
 - 同步與交易管理。
 - 資料備份與回復。



1-4 資料庫系統發展的歷史演進

■ 1-4-1 1960年代：

網路與階層式資料庫模型

■ 1-4-2 1970年代：

關聯式資料庫模型與實體關聯模型

■ 1-4-3 1980年代：

關聯式資料庫與SQL結構化查詢語言

■ 1-4-4 1990年代：

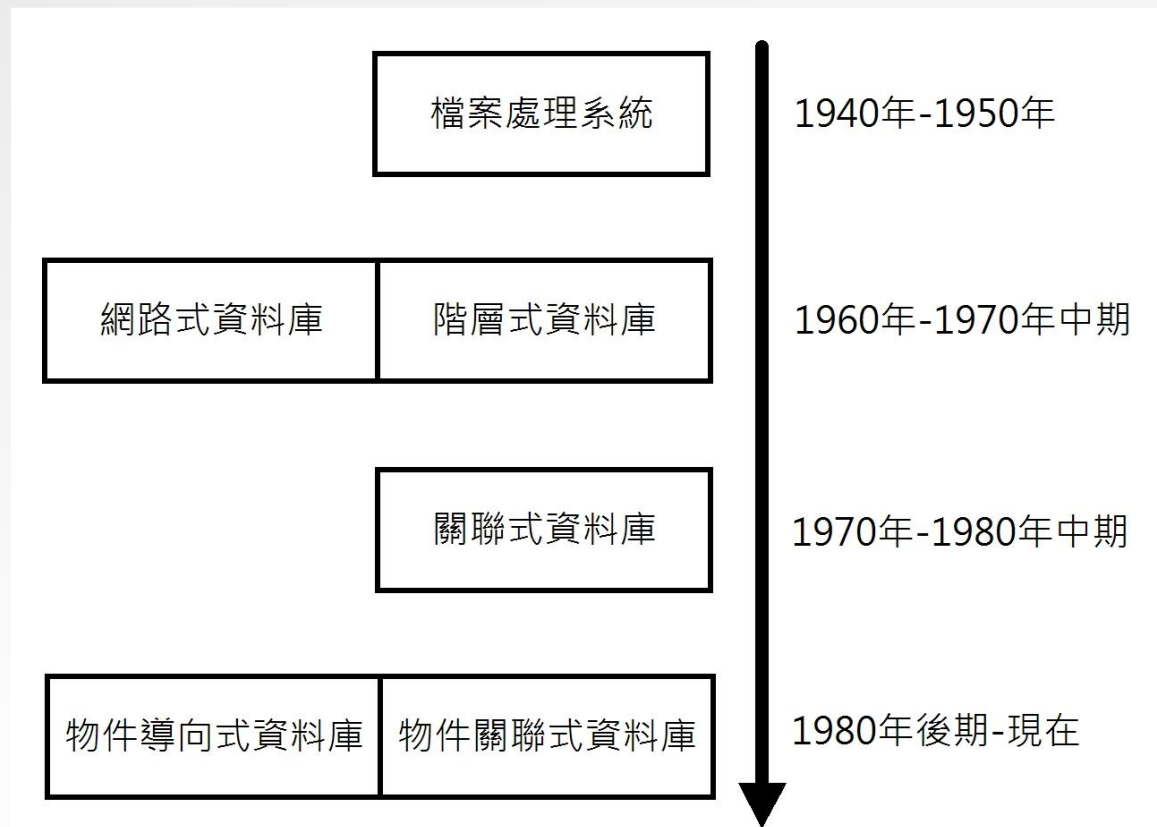
物件導向式資料庫模型與主從架構





1-4 資料庫系統發展的歷史演進

■ 資料庫系統演進的過程，如下圖所示：





1-4-1 1960年代：網路與階層式資料庫模型

- 1960年代是資料庫系統開始萌芽的年代，隨著磁碟的出現，檔案也從循序索引（Indexed-sequential）的循序存取轉變成「集合導向記錄模型」（Set-oriented Record Model）的直接存取。
 - **1961年**：Bachman替GE（General Electric）奇異設計第一套資料庫管理系統IDS（Integrated Data Store），1964年才廣泛的使用，這是一套使用「網路式資料庫模型」（Network Database Model）的資料庫。
 - **1965年**：IBM公司開發「IMS」（Information Management System）是使用「階層式資料庫模型」（Hierarchical Database Model）的資料庫。



1-4-2 1970年代：關聯式資料庫模型與實體關聯模型 - 1

- 1970年代是資料庫技術快速起飛的年代，資料庫管理系統已經成為大學學科和研究領域，眾多使用網路式和階層式資料庫模型的商用資料庫大量出現在市面。
 - 1970年：IBM研究科學家E. F. Codd博士發表「關聯式資料庫模型」（Relational Database Model）的重要論文。
 - 1976年：Peter Chen定義資料庫設計的「實體關聯模型」（Entity-Relationship Model），這是目前資料庫系統分析和設計的基礎。
 - 1978年：ANSI定義ANSI/SPARC三層資料庫系統架構。



1-4-2 1970年代：關聯式資料庫模型與實體關聯模型 - 2

- 1970年代的後期，有二個主要的關聯式資料庫研究計劃開始進行，如下所示：
 - **INGRES**：加州大學柏克萊分校的研究計劃，最後成立Ingres公司，這個研究計劃開發的資料庫系統使用QUEL查詢語言，它是Informix、Sybase和SQL Server資料庫系統的前身。
 - **System R**：IBM公司的研究計劃，最後成為IBM的DB2和Oracle資料庫的前身，使用的SEQUEL查詢語言就是第四篇SQL結構化查詢語言的前身。
- 資料庫查詢語言（Query Language）隨著上述研究計劃，在1970年代開始發展，例如：QUEL、SEQUEL、SQL和QBE查詢語言。

1-4-3 1980年代：關聯式資料庫與SQL結構化查詢語言



- 1980年代是商用關聯式資料庫大放異彩的年代，網路和階層式資料模型已經不再發展，不過直到現在仍有一些系統在使用。80年代初期已經開發超過100個ANSI/SPARC關聯式資料庫系統，例如：DB2、Oracle、Sybase和Informix等。
 - 1980年代中期：「SQL」（Structure Query Language）成為ISO標準的資料庫查詢語言，IBM DB2也成為IBM公司最重要的資料庫產品。
 - 在1980年代後期：專家系統（Expert Database System）、物件導向資料庫管理系統（Object-Oriented Database Management System）和主從架構分散式系統逐漸成為資料庫系統的未來趨勢。

1-4-4 1990年代：物件導向式資料模型與主從架構 - 1



- 1990年代關聯式資料庫的相關技術仍然持續的發展，隨著1990年代中期程式設計技術進入物件導向分析和設計，應用物件導向觀念的資料庫模型也逐漸成形，如下所示：
 - 物件導向式資料庫模型（Object-Orient Database Model）：這是使用物件（Object）觀念代替記錄儲存資料，以繼承減少資料重複，因為程式語言也支援物件導向，所以資料庫與程式語言可以使用一致的資料模型。
 - 物件關聯式資料庫模型（Object-Relational Database Model）：這是由Won Kim和Michael Stonebraker博士研究的資料庫模型，將物件導向的觀念整合至關聯式資料庫模型，強調這不是革命（Revolution），而是進化（Evolution）。



1-4-4 1990年代：物件導向式資料庫模型與主從架構 - 2

- 隨著Internet與WWW的興起和個人電腦的普及，集中處理的資料庫系統已經改為分散式主從架構（**Client/Server**）資料庫系統，將系統區分成兩個部分，如下所示：
 - **客戶端（Client）**：從端的應用程式負責使用者的資料輸入和顯示輸出的結果。
 - **伺服器端（Server）**：主端的資料庫系統是負責回應從端的請求，將查詢結果傳回從端的應用程式。
- 再加上平行資料庫處理（**Parallel Database Processing**）應用在關聯式資料庫上，可以將表格的資料水平或重直分割成多個資料庫且並行的進行資料處理，即「分散式資料庫系統」（**Distributed Database System**）。



1-5 資料庫技術的發展趨勢

	1960年-1970年中	1970年-1980年中	1980年之後	未來趨勢
資料模型 (Data Model)	網路式(Network)/ 階層式(Hierarchical)	關聯式(Relational)	物件導向式 (Object-oriented)	合併資料模型-物件關聯 式(Object-Relational)
資料庫硬體 (Hardware)	大型主機(Mainframes)	大型主機(Mainframes) /迷你主機(Minis)/ 個人電腦(PC)	工作站(Workstation)/ 快速個人電腦(Fast PC)	平行處理(Parallel Processing) /光學儲存媒體
系統架構 (System Architecture)	集中式(Centralized)	集中式(Centralized)	主從架構(Client/ Server)/分散式(Distributed)	異質分散/行動運算
使用介面 (User Interface)	沒有/ 表單(Forms)	查詢語言(Query Language)	圖形使用介面(GUI)	自然語言(Natural Language)/語音輸入
程式介面 (Program Interface)	程序式(Procedure)	內嵌查詢語言 (Embedded Query Language)	第四代程式語言(4GL)	整合資料庫和程式語言



End
