

DATABASE SYSTEMS

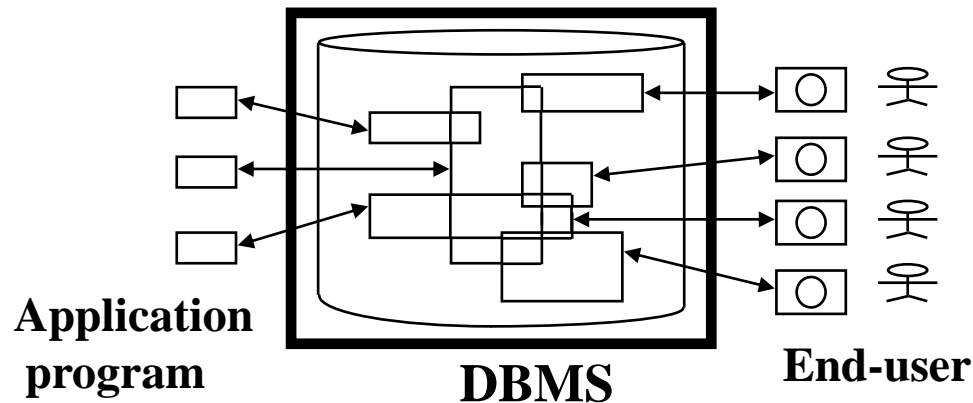
資料庫系統

現任：國立東華大學 副校長 特聘教授
曾任：國立交通大學 資訊科學系主任 所長 教授

楊維邦 教授
Prof. Wei-Pang Yang

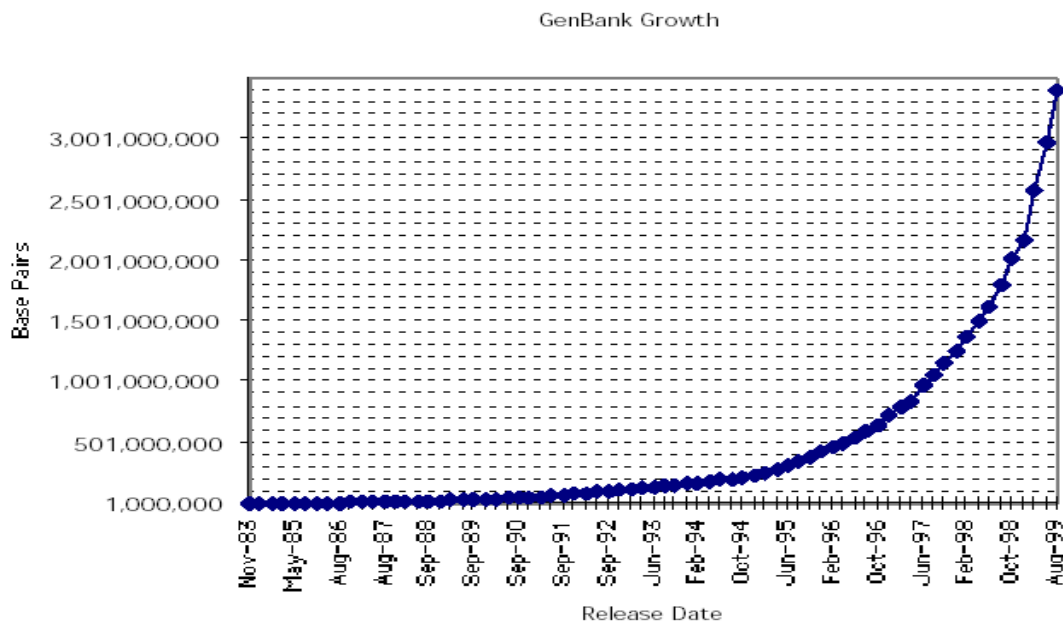
課程介紹

- ❑ 資料庫(Database)是一般資訊系統的核心資料，通常量很大
- ❑ 如何方便(convenient)又有效率的(efficient)管理存取資料庫是使用者最關心的二個要素
- ❑ 資料庫管理系統(Database Management System, 簡稱DBMS)
 - 是用來管理、儲存及查詢量很大的資料庫(database)之軟體
 - 其目標要使用者能夠方便的、有效率的存取及管理資料庫中的資料
- ❑ 這門課就是要澈底學會資料庫管理系統的實際操作及基本原理
- ❑ 所以這門課是資訊系統的核心基礎課程之一，非常重要

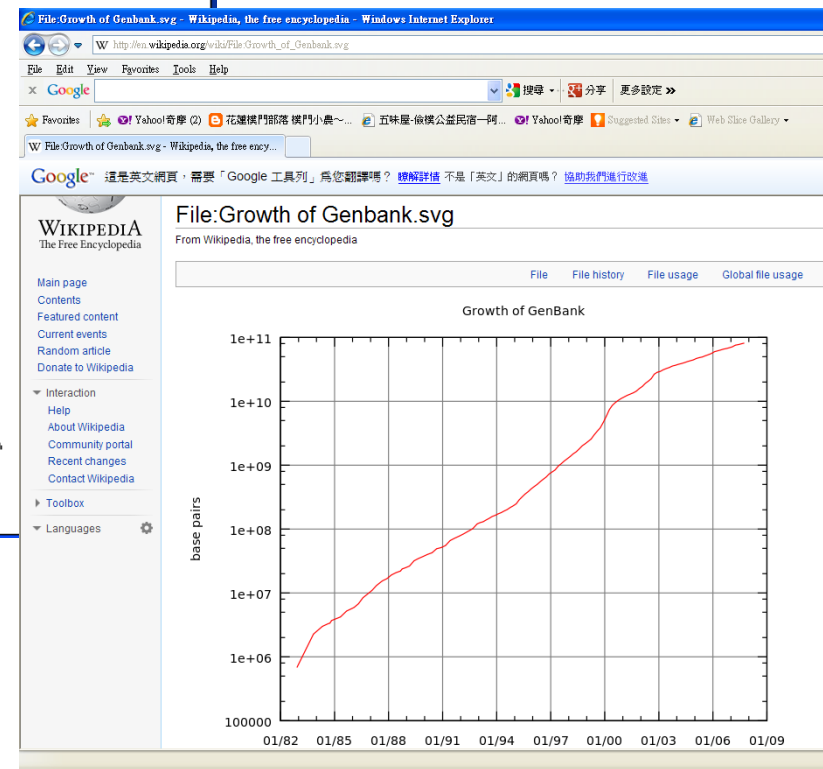


Database Growth Fast!

GenBank Database



Source: http://en.wikipedia.org/wiki/File:Growth_of_Genbank.svg



雲端、巨量資料(Big Data) vs. 資料庫

- Big Data (巨量資訊、超大量的資料、海量資料)
 - 是指對於大量資料處理的工具、程序、方法和流程等的集合
 - 資料本身在沒有做任何處理以前，也許沒有價值，但經過適當的分析萃取才能顯出價值
- 「巨量資訊可以分析未來、創造商機！」
- 資料類型多樣: Big Data 的議題會受到企業重視，是因為企業所收集的資料不只是文字的類型，還包括有影音和圖像等多樣類型
- 資料來源多元: 除了傳統的人工輸入和系統計算產生的資料以外，還包含網路上每日產生的大量資料，其產生的速度超過人工和現行資料庫所能處理的能力
- 歷史資料借鏡: 資訊化多年後，企業也許累積龐大的資料量，企業希望能夠從這些超大量資料中，透過一些方法和工具能夠快速取出可以幫助企業參考應變的資訊哩！
- 了解資料庫也可幫助了解巨量資料



圖片來源 R-Blogger.com

課程目標

□ 會使用資料庫系統

- 學員會使用資料庫管理系統(DBMS) 來處理及存取大量的資料
- 而且是方便的、有效率的
- 再加上是能保持資料正確的、保護資料安全的利用

□ 能了解資料庫系統

- 探討資料庫管理系統內部底層的設計、運作方式與基本原理
- 培育學生具有能力成為"完全內行"的資料庫管理師(DBA)之基本功力。

□ 所以本課程之目標不只教學員學會用資料庫系統

- 好比不只教會開車(一般大學部開的[資料庫管理](#)課程--只學會使用某種資料庫系統)；
- 還要探討汽車內部之架構，至少能分辨引擎蓋下之三油三水之位置及功能(即了解資料庫管理系統底層內部的設計及運作方式與原理)。
- 將來運用時或遇到瓶頸時才能"知其所以然"而解決問題。

Example: A Simple Query Processing

Query in SQL :

```
SELECT CUSTOMER. NAME
FROM CUSTOMER, INVOICE
WHERE REGION = 'N.Y.' AND
      AMOUNT > 10000 AND
      CUTOMER.C#=INVOICE.C
```

Internal Form :

$\Pi(\sigma(S \bowtie SP))$

Operator :

```
SCAN C using region index, create C
SCAN I using amount index, create I
SORT C?and I?on C#
JOIN C?and I?on C#
EXTRACT name field
```

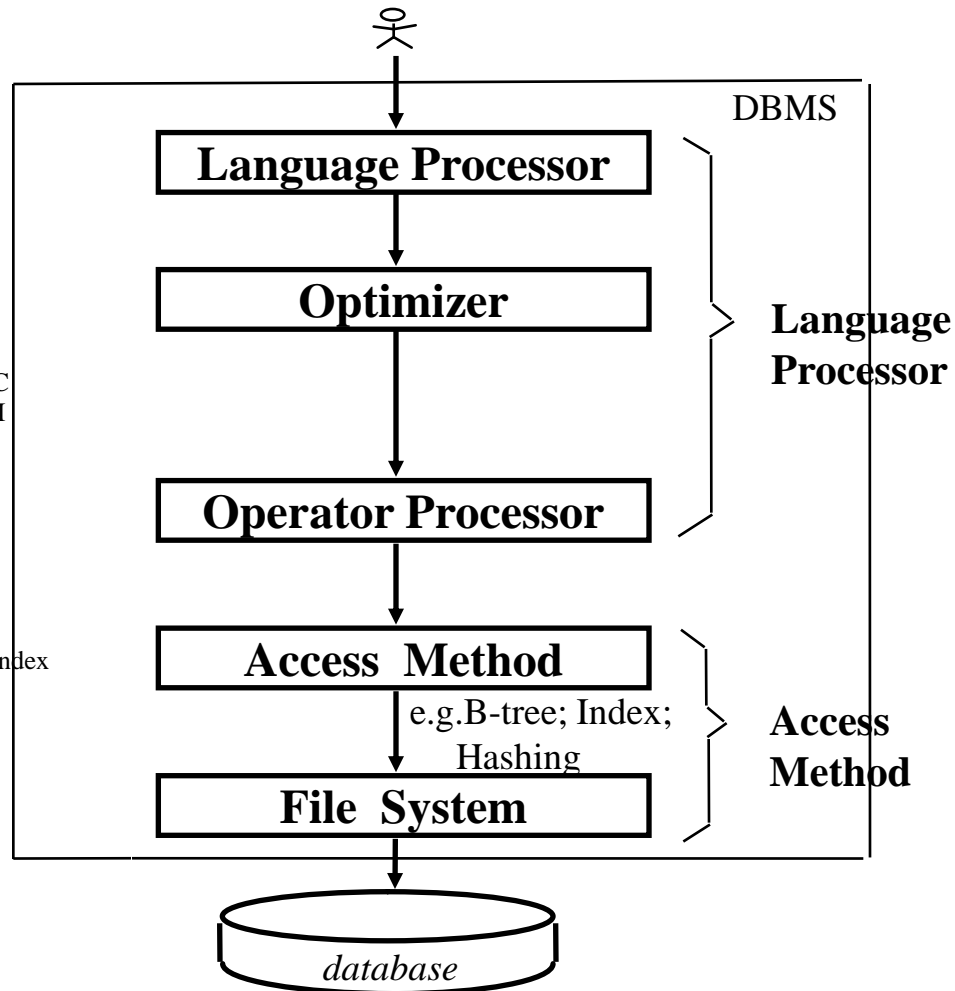
Calls to Access Method :

```
OPEN SCAN on C with region index
GET next tuple
```

⋮

Calls to file system :

```
GET10th to 25th bytes from
block #6 of file #5
```



Querying and Data Storage

□ Components of Database System

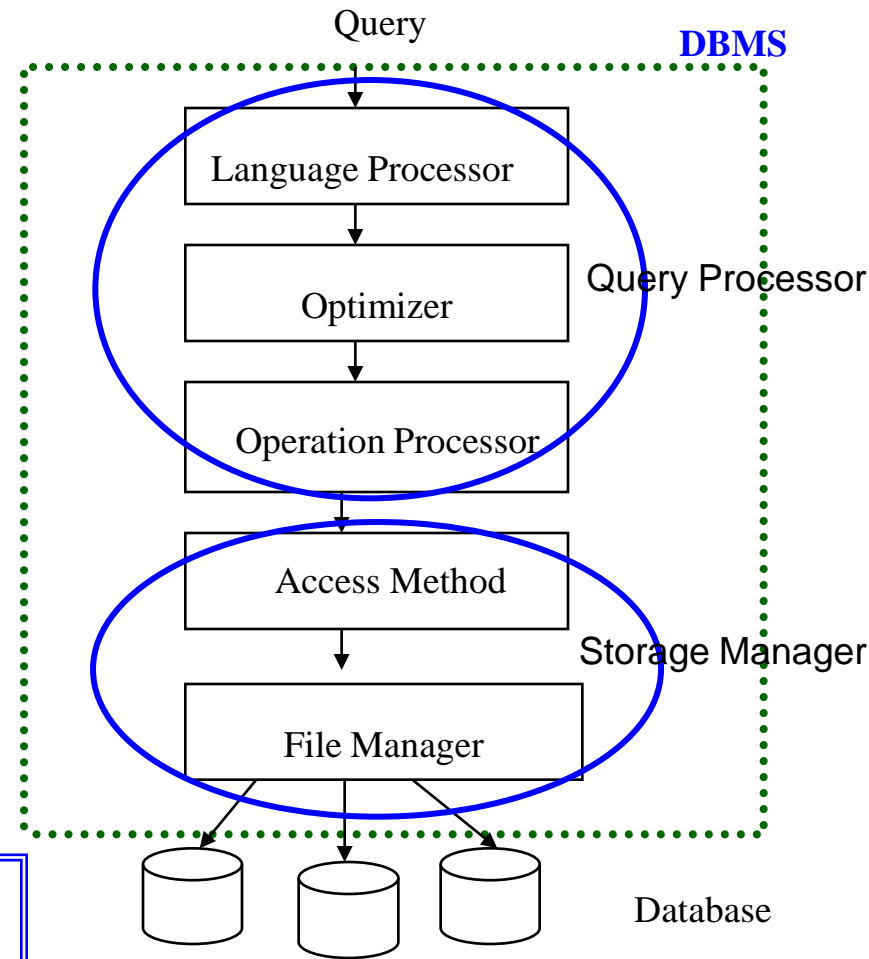
■ Query Processor

- Helps to simplify to access data
- High-level view
- Users are not be burdened unnecessarily with the physical details

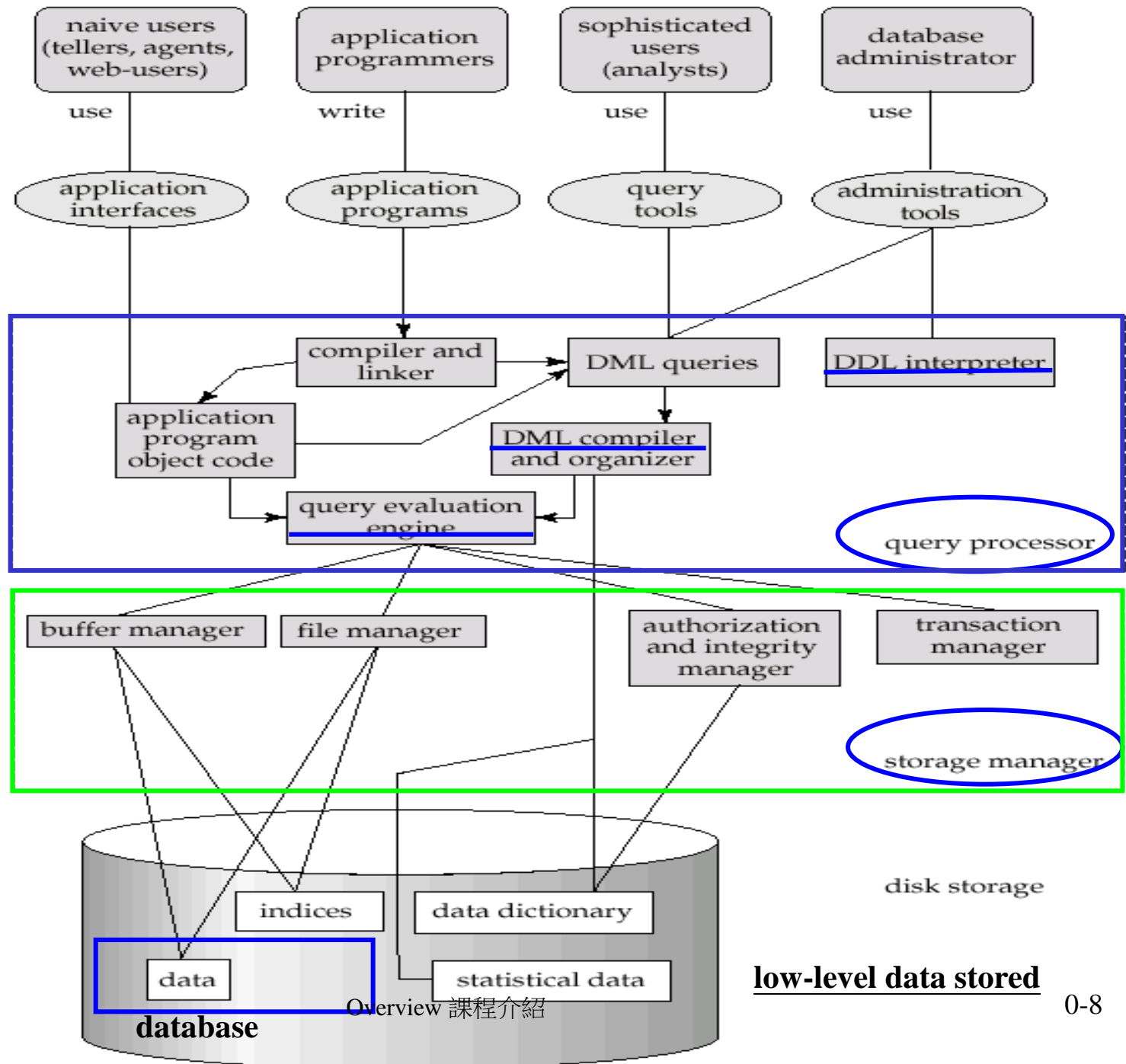
■ Storage Manager

- Require a large amount of space
- Can not store in main memory
- Disk speed is slower
- Minimize the need to move data between disk and main memory

Goal of a DBMS: provides a way to store and retrieve data that is both *convenient* and *efficient*.



Overall System Structure



本課程講授內容

- **PART I: 入門與導論**
 - Overview
 - DB2系統及SQL語言
 - 闡述關連式資料模型(The Relational Model)
 - 階層式資料模型(The Hierarchical Model)簡介
 - 網狀式資料模型(The Network Model)簡介
- **PART II: 資料庫設計 (Database Design)**
 - 資料庫問題分析與 E-R Model
 - 資料庫的表格正規化
 - 設計介面增刪查改資料庫
- **PART III: 進階探討**
 - 快速存取方法(Access Methods)
 - 資料庫回復(Database Recovery)
 - 協同控制(Concurrency Control)
 - 資料安全與資料正確(Security and Integrity)
 - 查詢最佳化(Query Optimization)
 - 分散式資料庫系統(Distributed Database)

PART I: 入門與導論

- ❑ DB2系統及SQL語言:
 - 介紹最多人使用的查詢語言SQL
 - 配合實作習題，先試用一個免費的DBMS系統MySQL
 - 好比要學開車可先在大停車場讓你繞一圈
 - 這樣我們在介紹下單元關連式資料模型的設計原理時會比較容易想像
- ❑ 關連式資料模型(The Relational Model):
 - 闡述使用者視資料庫為許多表格(tables)組成的關連式資料庫之原始設計原理
 - 這是關連式資料庫的理論基礎
- ❑ 階層式資料模型(The Hierarchical Model)及網狀式資料模型(The Network Model):
 - 這是最早的二個資料模型
 - 介紹這二模型將有助於我們對資料庫模型更深入了解，並知其來龍去脈
 - 我們將簡要的說明其原始設計原理

Data Models

- ❑ Hierarchical Data Model
- ❑ Network Data Model
- ❑ Relational Data Model
- ❑ Object-Oriented Data Model
- ❑ ...

Database Technology Trends

	1960s to Mid-1970s	1970s to Mid-1980s	Late 1980s	Future
Data Model	Network Hierarchical	Relational	Semantic Object-oriented Logic	Merging data models, knowledge representation, and programming languages
Database Hardware	Mainframes	Mainframes Minis PCs	Faster PCs Workstations Database machines	Parallel processing Optical memories
User Interface	None Forms	Query languages - SQL, QUEL	Graphics Menus Query-by-forms	Natural language Speech input
Program Interface	Procedural	Embedded query language	4GL Logic programming	Integrated database and programming language
Presentation and display processing	Reports Processing data	Report generators Information and transaction processing	Business graphics Image output Knowledge processing	Generalized display managers Distributed knowledge processing

Contents of PART I: 入門與導論

- ❑ **Unit 1 Introduction to DBMS**
- ❑ **Unit 2 DB2 and SQL**
- ❑ **Unit 3 The Relational Model**
- ❑ **Unit 4 The Hierarchical Model**
- ❑ **Unit 5 The Network Model**

PART II: 資料庫設計 (Database Design)

□ 資料庫問題分析與架構規劃:

- 若有一大量資料想利用DBMS建資料庫來管理。第一步要分析問題，找到使用者需求
- 實體-關係模型(Entity-Relationship Model,簡稱E-R Model)是一套資料庫的設計工具。我們可以利用E-R Model分析資料庫問題。它可以把真實世界中複雜的問題中的事物和關係轉化為資料庫中的資料架構
- 由於利用實體-關係模型設計資料庫時,並不會牽涉到資料庫的操作、儲存方式等複雜的電腦運作。所以,我們會把心力放在需求分析去規劃想要的資料庫,並以實體-關係圖(E-R Diagram)來呈現

□ 資料庫的表格正規化:

- 實體-關係圖很容易轉化為表格(Tables),而資料庫就是由許多表格(tables)組成的
- 這些表格要正規化(Normalization)才能避免將來操作時的異常現象發生

□ 設計介面增刪查改資料庫:

- 如何方便、又有效率的管理存取資料庫是使用者最關心的二個要素
- 良好的介面設計,可以讓使用者方便的查詢、方便的新增、方便的刪除、方便的修改的處理資料庫

Database Design

- **Database Design** - The process of designing the general structure of the database:
 - Logical Design
 - Physical Design
- **Logical Design** – Deciding on the database schema.
 - To find a “good” collection of relation schemas.
 - Business decision – What attributes should we record in the database?
 - Computer Science decision – What relation schemas should we have and how should the attributes be distributed among the various relation schemas?
- **Physical Design** – Deciding on the physical layout of the database

Design Process

- **Phase I**
 - Specification of user requirement (with domain experts)
- **Phase II**
 - Conceptual design (unit 6)
 - Choose a data model
 - Design tables
 - Normalization (unit 7)
- **Phase III**
 - Specification of functional requirements
- **Phase IV**
 - User interface design (unit 8)
 - Implementation

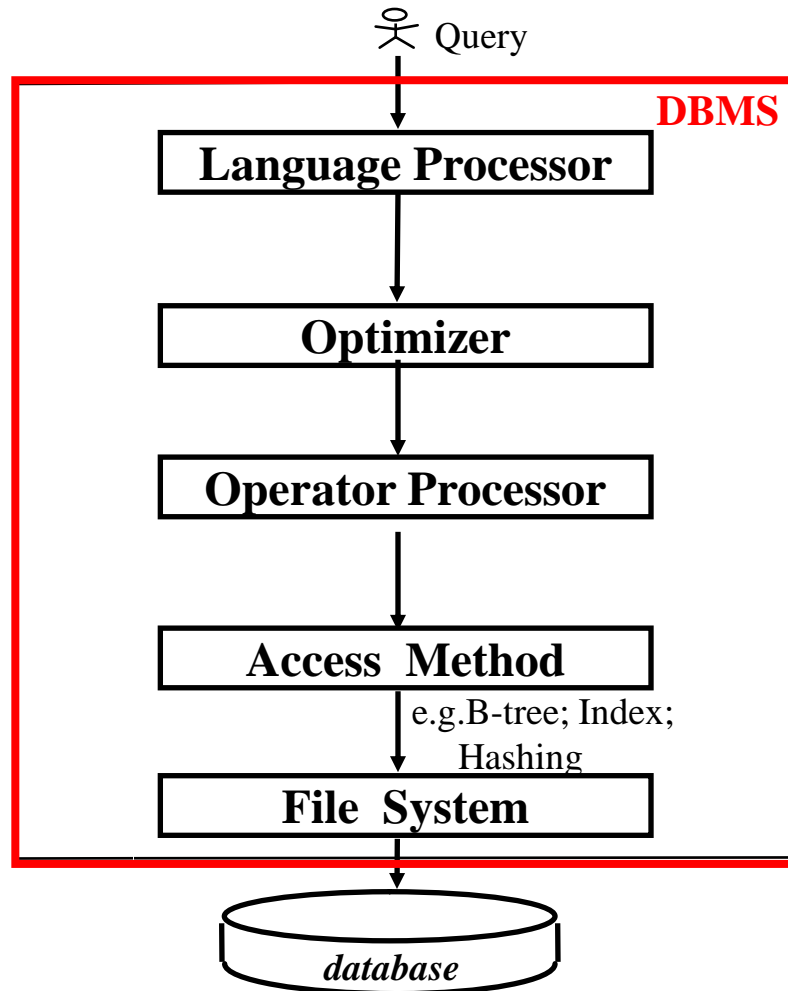
Contents of PART II: 資料庫設計

- ❑ Unit 6 Database Design and the E-R Model
- ❑ Unit 7 Normalization (表格正規化)
- ❑ Unit 8 User Interfaces (使用者介面)
- ❑ Unit 9 實作範例一：
- ❑ Unit 10 實作範例二：

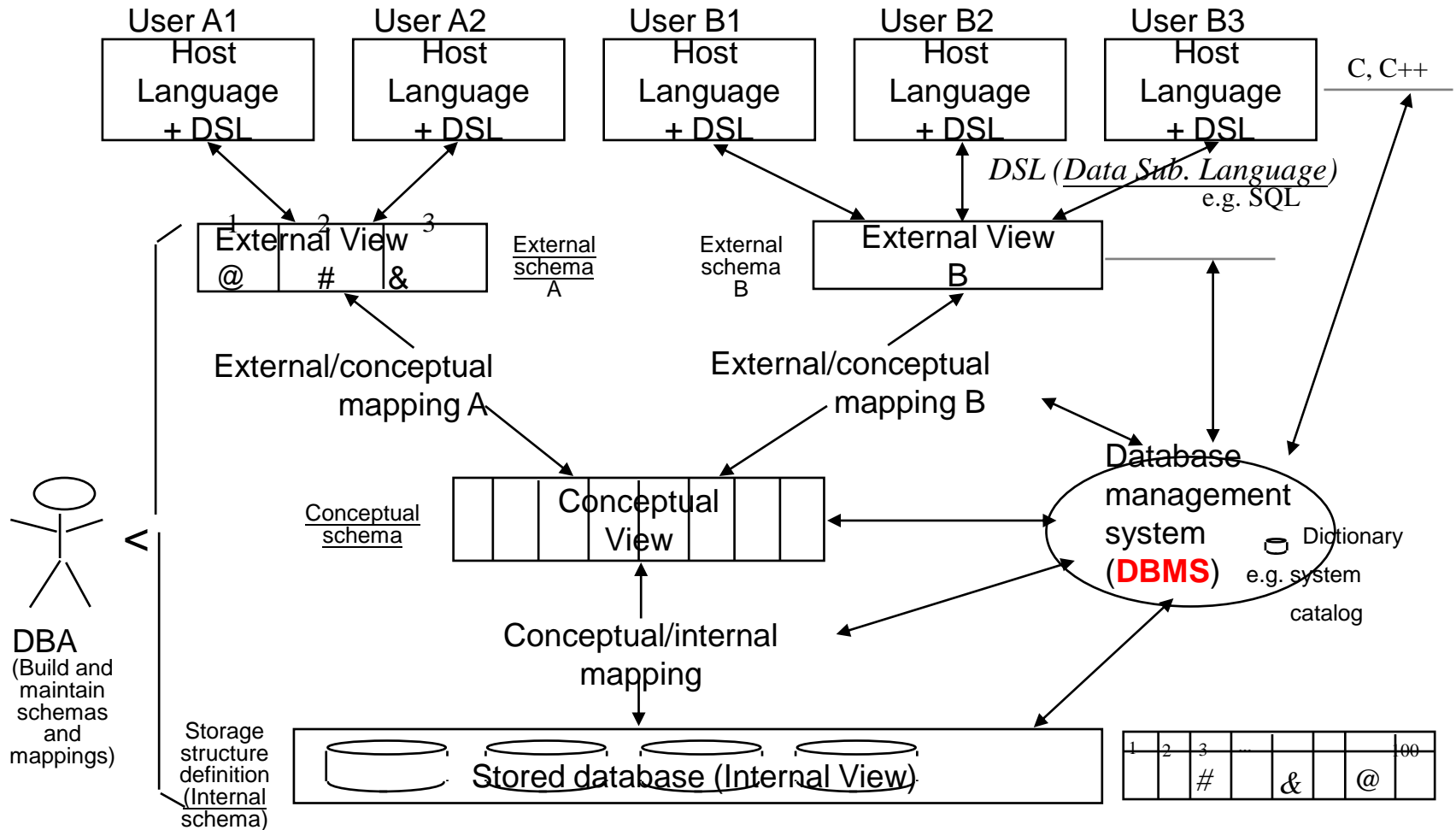
PART III: 進階探討

- ❑ 快速存取方法(Access Methods):介紹大量資料庫之所以可以快速存取的技巧及方法，包括最有名的索引系統 B-tree index.
- ❑ 資料庫回復(Database Recovery):介紹萬一系統出問題，導致資料不正確時，如何回復到某一正確點的相關議題。
- ❑ 協同控制(Concurrency Control):通常同一資料庫是允許多人同時共用的，本單元探討DBMS是如做控制使得多人共同而不互相干擾。
- ❑ 資料安全與資料正確(Security and Integrity): 資料安全是指保護資料不讓未授權的人偷竊、破壞;資料正確是指保護資料不讓已有授權的合法使用者誤用，例如把存款餘額變成負數。
- ❑ 查詢最佳化(Query Optimization):查詢資料庫某一問題可以有不同的查詢指令。不同的指令查詢時間可能相差好幾倍，本單元介紹撰寫最佳化的查詢語言指令的方法。
- ❑ 分散式資料庫系統(Distributed Database): 分散式資料庫是透過[電腦網路](#)將多個[資料庫](#)連線起來組成的資料庫。現今我們用的系統有太多的分散式資料庫系統，本單元將介紹之，探討它的優缺點。
- ❑ 擴充E-R Model: 一般不太複雜的問題，前面介紹的E-R Model基本上就夠用。但大系統或比較特殊的系統就必須有更多功能的擴充式E-R Model(Extended E-R Model,簡稱EER-model)可處理包括 Specialization、Generalization、Aggregation等問題。
- ❑ 進階表格正規化: 關連式資料庫是由許多表格(tables)組成的，表格要正規化。正規化分六級，一般也許到第三級正規化即可。理論上還有BCNF級、第四級、第五級的正規化。本單元要介紹這些特例。當年這些是博士論文級的問題探討。

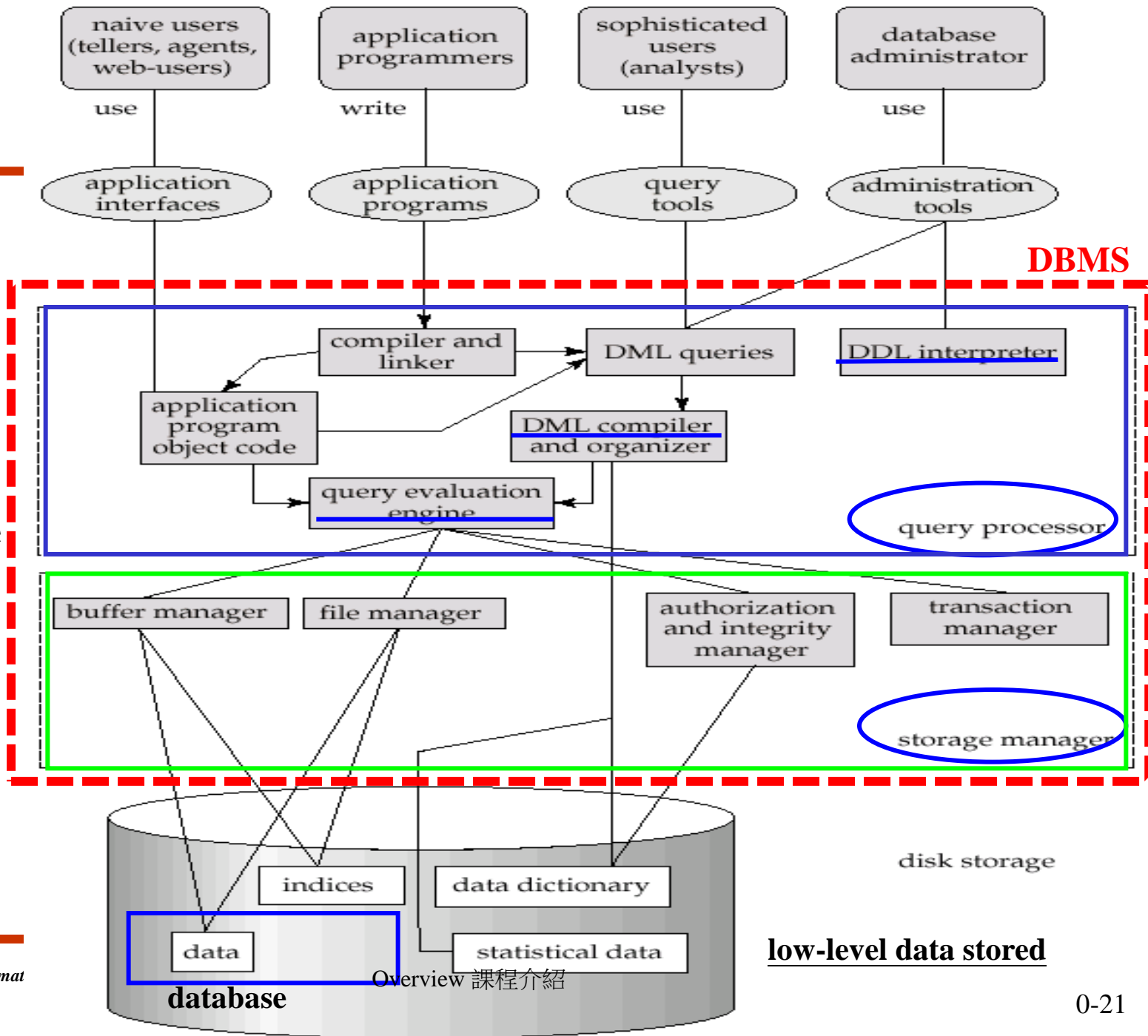
Architecture for a Database System: view 1



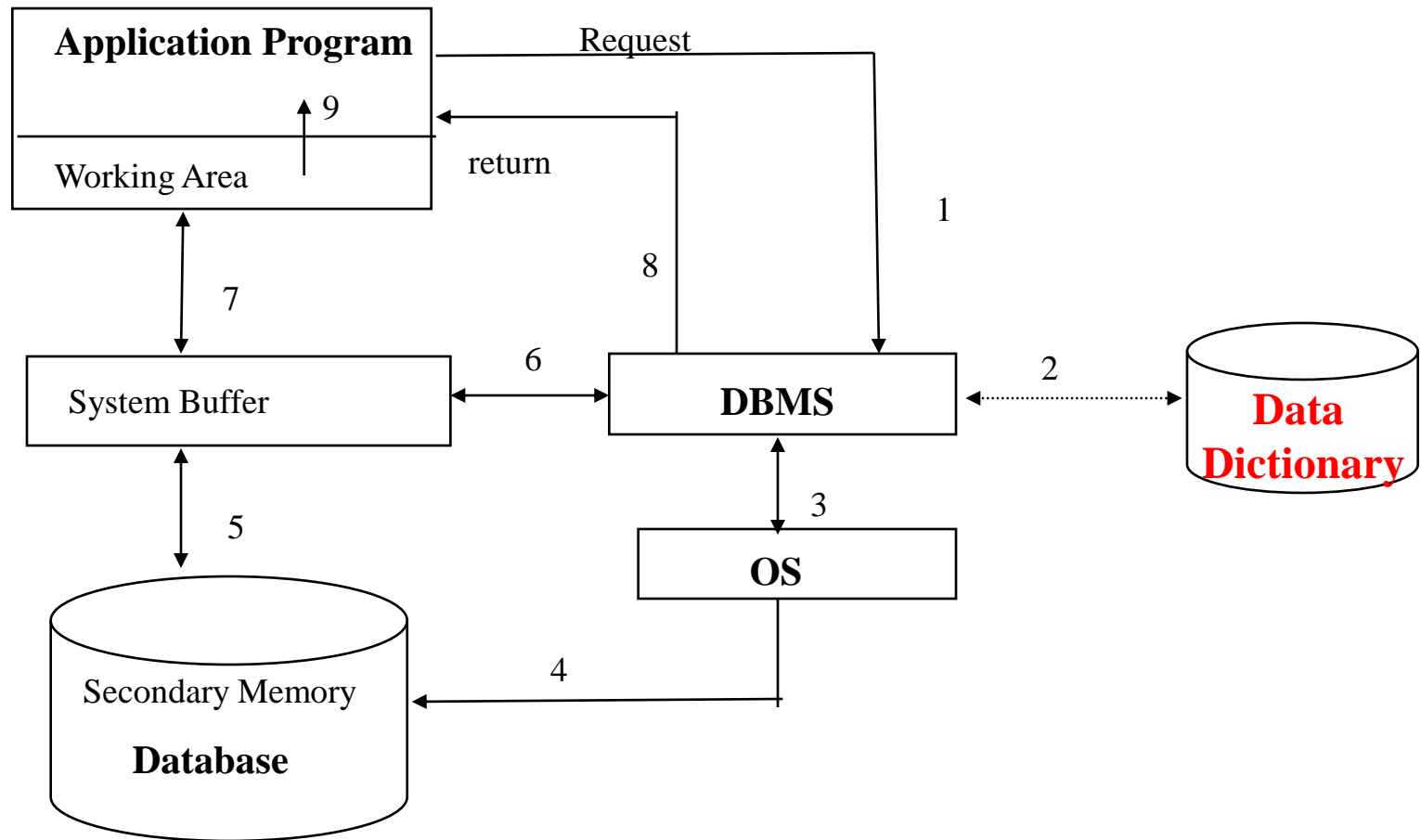
Architecture for a Database System: view 2



Overall System Structure



Data Dictionary in DBMS



Contents of PART III: 進階探討

- ❑ Unit 11 Access Methods
- ❑ Unit 12 Database Recovery
- ❑ Unit 13 Concurrency Control
- ❑ Unit 14 Security and Integrity
- ❑ Unit 15 Query Optimization
- ❑ Unit 16 Distributed Database
- ❑ Unit 17 More on E-R Model
- ❑ Unit 18 More on Normalization
- ❑ Unit 19 More on User Interfaces
- ❑ Unit 20 More on X?

實作練習與實作範例

□ 實作練習:

- 本課程有配合進度設計一系列的實作習題，學員要跟著照進度做，學習成效就會出來。
- MySQL是用一個免費的DBMS系統。用MySQL或你手邊有的任何系統來做我們設計的作業。
- 教育家杜威的Learning by doing (從做中學)對學電腦程式語言、任何軟體特別貼切。

□ 實作範例:

- 配合進度設計的一系列實作習題，也對應放了以前學員做的範例當參考。
- 每期有較好的作業也都有機會被放到這個實作範例池來。
- 大家可以互相觀摩

考古題與相關證照

□ 考古題:

- 收集各種考試的題目。學員可以試做來了解自己的能力
- 也可以模擬考來增加自己的實力，有時也會提高學習動機。

□ 相關證照:

- 了解業界認證考試的學習目標
- 可驗收所學「資料庫系統」知識是否達到業界基本門檻
- 學習到更多相關知識（課堂內容無法涵蓋所有知識）
- 取得業界認可的資料庫相關證照(TQC-SQL5)
- 參考範例試題：http://www.tqc.org.tw/TQC/exam_sample/MySQL5.pdf

先修課程與適合學員

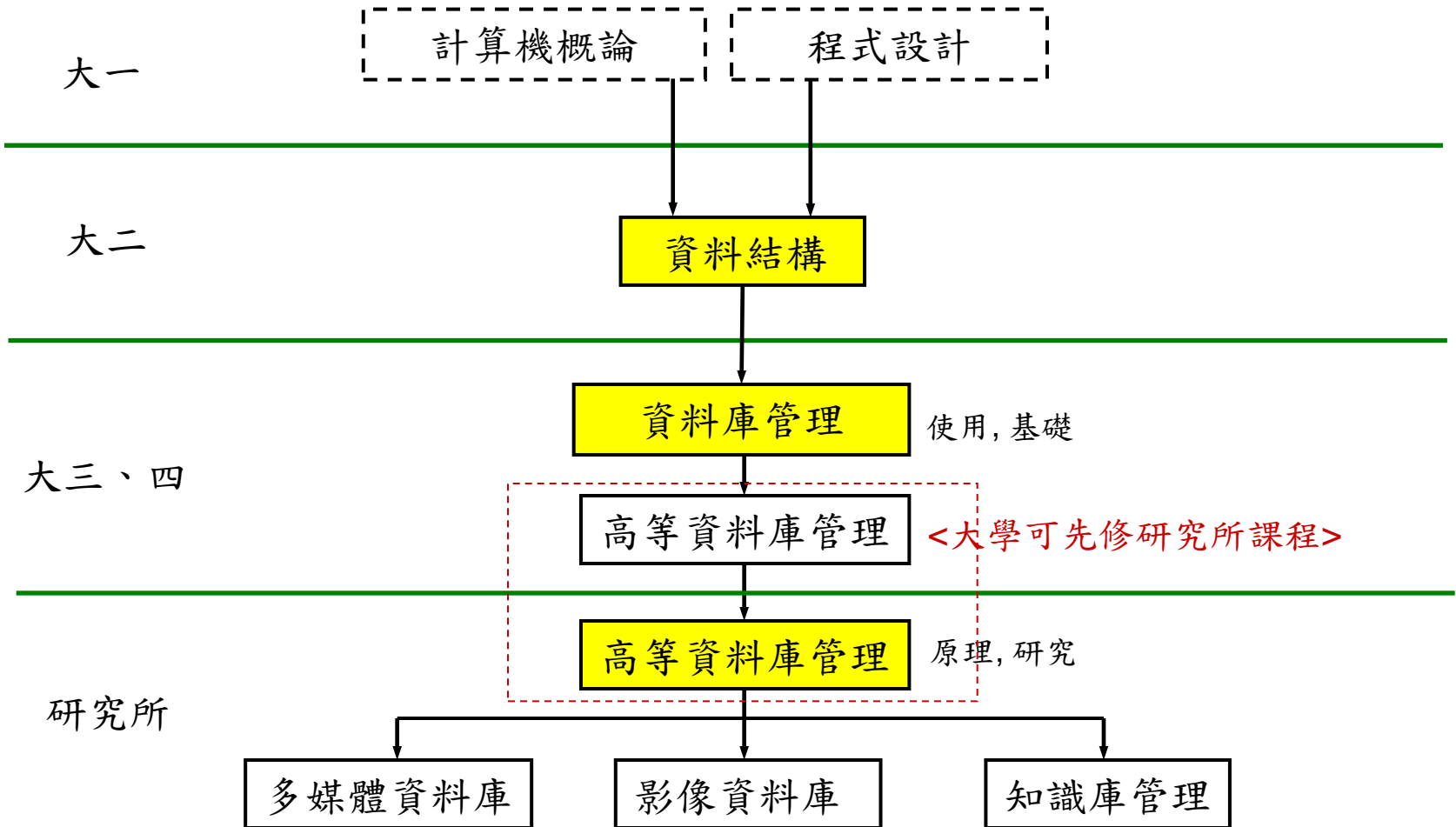
□ 先修課程

- 程式設計(任何語言)
- 資料結構基本概念

□ 適合學員

- 正在學校選讀資料庫相關課程者
- 具備程式設計(任何語言)、資料結構基本概念的同學、在職人士

Database 課程流程圖



Study and Research on Databases

Level 5: Doing Research

Level 4: Survey Papers: Special Topics (Unit 21 -)

**Advanced
DBMS**

Level 3: DBMS: Advanced Topics

(Unit 11 – 20)

Date, Vol. 1, 2

Ullman

Level 2: DBMS: Fundamentals

(Unit 1 – 10)

Date, Vol. 1

Using MySQL

Level 1: Using DBMS

Contents of PART VI: 主題研究

- ❑ Unit 21 Object-Oriented Database
- ❑ Unit 22 Logic-Based Database
- ❑ Unit 23 Image Database
- ❑ Unit 24 Multimedia Database
- ❑ Unit 25 Real-Time Database
- ❑ Unit 26 Parallel Database⁸¹
- ❑ Unit 27 Temporal Database
- ❑ Unit 28 Active Database
- ❑ Unit 29 Bioinformatic Database
- ❑ Unit 30

Where to find Database papers?

❑ Journals:

- ACM Trans. On Database Systems (TODS)
- IEEE Trans. On Software Engineering (TOSE)
- IEEE Trans. On Knowledge and Data Engineering
- ACM Trans. On Office Information Systems (TOOIS)
- **Journal of Information Science and Engineering**
- ...

❑ Conferences:

- VLDB (Very Large Data Base)
- ACM SIGMOD
- PODS (Principal of Database Systems)
- IEEE Data Engineering
- **NCS, ICS (National/International Computer Symposium)**
- ...

參考書目

- ❑ [楊維邦自編單元講義](#)(可以免費下載，若願意可自由樂捐[花蓮黎明教養院](#)):

我在哈佛大學做博士後研究時聽許玫君教授的資料庫課程。本講義就是參考許教授的自編講義，及參考下一本 C. J. Date, “An Introduction to Database Systems”再加上多年的教書經驗編輯而成。以深入淺出方式介紹資料庫領域的重點。

- ❑ C. J. Date, “An Introduction to Database Systems,” 8th edition, 2004 (最好買一本，幾乎可以用一輩子):

這是資料庫最有名的教科書，有時被譽為DBMS的聖經版參考書，與資料庫相關的問題幾乎都有解說，觀念又很清楚。本課程主要是參考這本書的說法與例子。值得買一本，配合我的自編講義，我會將書中的重要主題盡量闡述說明清楚。

- ❑ 曾守正與周韻寰編著，「資料庫系統之理論與實務 (SQL Server 2008 版)」，華泰文化公司總經銷，2010年二月出版。

原則上儘量買原文書(2)來參考，英文要越強越好。若一定要有一本中文的資料庫書籍，我推薦這本。

課程資料來源 (Course Materials)

- 本課程之資料主要由下面四個來源，建議學員可以購買其中之一。
- 建議下載我們的講義，隨著講義，一面觀看線上課程，一面研讀一本教科書，應會有很好的學習效果。
- 教科書及資料來源:
 1. C. J. Date, An Introduction to Database Systems, 8th edition, Addison-Wesley, 2004.
 2. A. Silberschatz, etc., Database System Concepts, 5th edition, McGraw Hill, 2006.
 3. J. D. Ullman and J. Widom, A First Course in Database Systems, 3rd edition, Prentice Hall, 2007.
 4. Cited papers (講義中提到之參考文獻)

end of unit 0