Teaching Experience Sharing of Database Systems (資料庫系統)

Hao-hua Chu (朱浩華)

i-Space Labs

Computer Science & Information Engineering Department

Institute of Networking & Multimedia

National Taiwan University

(臺灣大學資訊工程)

Who am I?

- Left Taiwan at 13
- Cornell, B.S., Computer Science
- University of Illinois at Urbana Champaign, PhD, Computer Science

- Xerox
- Intel
- NTT DoCoMo
- Came back to Taiwan at 33
- NTU, Associate Professor
- A member of i-space Labs

My research interests

- Systems & networking background
- Ubiquitous & pervasive computing
  - Indoor location systems
  - Ubiquitous healthcare
  - Social computing
  - Personal experience computing
- http://mll.csie.ntu.edu.tw/
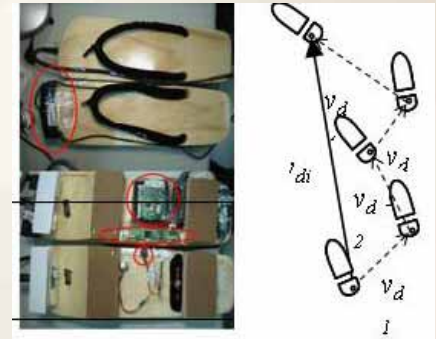
i-space labs

- Vision
  - Fantasy of being able to live a quality, yet effortless life.
  - Empower seemingly senseless space by the right combination of hardware and software.

- Mission
  - Conduct research on aspects of intelligent space, including automated control, artificial intelligence, computer vision, system & middleware, networking, and system architecture.
  - Develop practical solutions to real-life problems by collaborating with experts of diverse background, including medicine, psychology, social science, art, and others.

How do I fill the time?

- My meandering experiences on teaching database
- Overview on well-organized curriculum resources provided by MSDN academic alliance
- Comparison between my course and MSDN curriculum
- Q & A Long one :)

Course Objective

- First course in database systems for CS majors
- Objectives:
  - ¼: How to use a relational database? (ER model, SQL)
  - ¾: How to build a relational database? (Storage, Indexing, Query Optimization, Concurrency, Transaction)
- Emphasize on hands-on learning & on systems aspects
  - Build key components of a relational database system
- Objectives may be different for IM majors

Student Composition

- The only required course for undergraduate senior students
  - It is also open to all under/grad students
- 2003 Fall Semester
  - 98 (1 EE senior, 1 CSIE junior)
- 2004 Fall Semester
  - 111 (1 Math senior, 1 EE senior, 2 CSIE sophomores)

Course Prerequisite

- Data structure and algorithms
- Good at C++ programming
  - Considerable amount of coding in C++.
  - Students already know or learn C++ on their own.
- Taught (almost completely) in English
  - Students can ask questions in Chinese.

Textbook

- "Database Management Systems", 3rd Edition, by Ramakrishnan & Gehrke.
- Other good textbooks:
  - "Database System Concepts" by Silberschatz, Korth, & Sudarsan
  - "Database Systems: The Complete Book", by Garcia-Molina, Ullman, & Widom
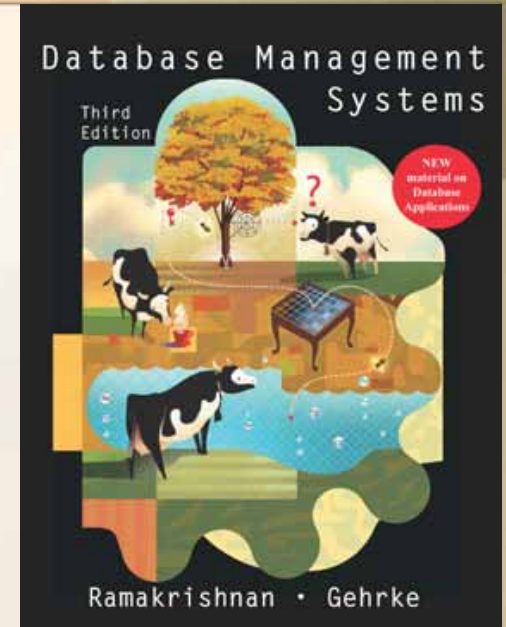
Ramakrishnan textbook

- Good:
  - Come with programming assignments that implement algorithms described in the text book
- Bad:
  - Ambiguous writing, inconsistent wordings, etc.
  - Require certain amount of clarifications from lecturers
  - "More like an experience report from researchers rather than introductory textbooks for beginners"
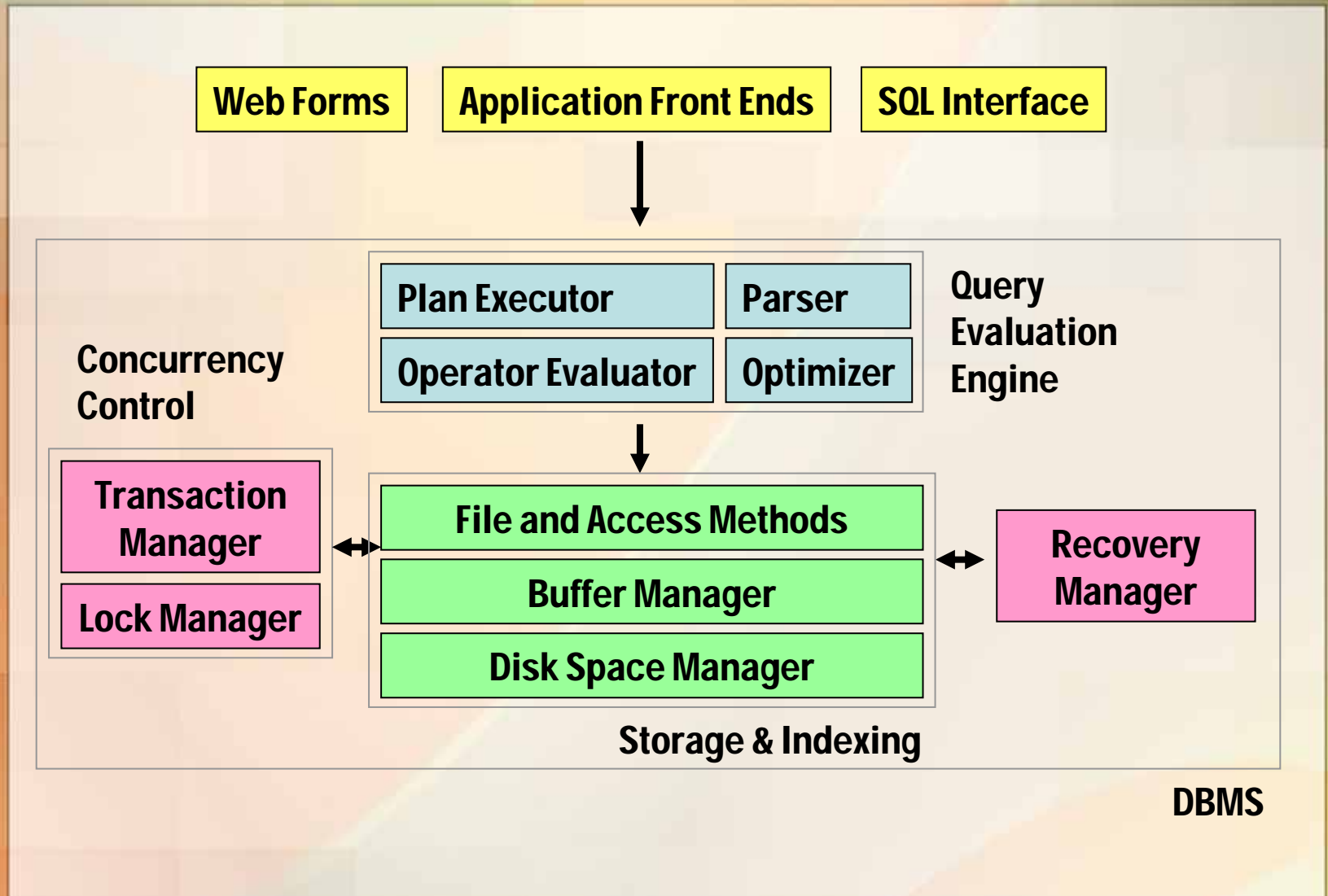
Course Format

- One long 3 hours lecture each week (14 lectures)
- About 1 chapter per week
  - Total: 16 chapters or 500 textbook pages in 2004
- Course load:
  - 6 programming assignments
  - 4 non-programming exercises
- Midterm exam
- Final exam

Course Design

- Same difficult level & speed as some of top CS schools in the U.S.
- Challenge & push students beyond their current limit in learning
  - Challenging programming assignments
  - (I was told that I was to be blamed for low turnouts among NTU CSIE students in programming competitions)
- NTU CSIE does not produce too many "strawberries".
  - Better to produce bananas?

	Lecture Topics	Readings	Assignments
9/13	Introduction	R&G Chapters 1	
9/20	ER & Relational Models	R&G Chapter 2 (except 2.7)& 3	Assigment 1: ER Diagram
9/27	Relational Algebra and SQL	R&G Chapters 4.1~4.2 & 5	Assignment 2: Translating ER Model to Tables & Relational Algebra
10/4	Storage and Indexing	R&G Chapter 8	Assignment 3: SQL
10/11	Storage and Indexing	R&G Chapter 8	Assignment 4: Heap File Page
10/18	Storage and Indexing	R&G Chapter 9	
10/25	No class		Assignment 5: Buffer Manager
11/1	Storage and Indexing	R&G Chapter 10	Assignment 6: B+ Tree
11/8	Storage and Indexing	R&G Chapter 11	Assignment 7: Extendible Hashing and Linear Hashing
11/20	Midterm		Midterm Solution
11/22	Query Evaluation	R&G Chapter 12	
11/29	Query Evaluation	R&G Chapter 13	Assignment 8: Query Evaluation and External Sorting Deadline changed to 12/7
12/6	Query Evaluation	R&G Chapter 14	Assignment 9: Sort-Merge Join
12/13	Query Evaluation	R&G Chapter 15	
12/20	Transaction Management	R&G Chapter 16	
12/27	Transaction Management	R&G Chapter 17	Assignment 10: Transaction Simulation & Crash Recovery
1/3	Transaction Management	R&G Chapter 18	
	Final Exam		

Content Map (from R&G book)

Web Forms	Application Front Ends	SQL Interface

Query Evaluation Engine

Plan Executor	Parser
Operator Evaluator	Optimizer

Concurrency Control

Transaction Manager
Lock Manager

File and Access Methods
Buffer Manager
Disk Space Manager

Recovery Manager

Storage & Indexing

DBMS

Course Contents (More Details)

Foundations:
- Database design
  - ER Model, Relational Model, Relational Algebra
- SQL
  - Select, Update, Aggregate, Constraints, Triggers, ..

Storage and Indexing:
- RAID (redundancy vs. reliability)
- Buffer replacement algs (LRU, MRU, FIFO, …) (performance)
- Indexing: ISAM tree, B+ tree, static & dynamic hashing (tables & queries vs. cost vs. concurrency)

Query Evaluation:
- External sorting algs
  - merge-sort
- Relational operator eval algs
  - projection, selection, join
  - File scan, index-based, sorting, in-memory hash, etc.
- Query optimizer

Transaction Management:
- Concurrency control
  - Serializability
  - Locking algs (2PL, Strict 2PL, Conservative 2PL), deadlock detection & prevention
- Crash recovery (ARIES)

Assignments

#1: ER Diagrams

#2: Translating ER diagrams to relational tables

#3: SQL Queries (Microsoft Access)
- Answer Trivia questions on a StarTrek database
- Find all characters that have been on all neutral planets.

#4: Heap File Structure (300)
- Organize and store pages, records, in fields in a heap file.

#5: Buffer Manager (260)
- Implement a page replacement alg on a buffer pool

More Assignments

#6: B+ tree index (850)
- Implement operations (search, insert, delete) B+ tree
- 850 lines

#7: Static & Extendible Hashing index

#8: Query Evaluation

#9: Sort-merge join
- Implement the algorithm.
- 80 lines

#10 Transaction simulation & ARIES crash recovery (120)
- Implement WAL & restart after crash

Mini-base versions

- Linux version (C++)
- Java version
- Visual C++ version

國立臺灣大學
92 學年度第 1 學期教學意見調查統計表
*** 修課學生的文字意見 ***

科目：資料庫系統　　　　課號：902 47000　　　班次：
任課教師：朱浩華　　　　　　　　統計日期：02/23/2004

我對這門課有以下的建議或意見(包括教師教學效果好或需要改進意見)：

Assignment 4 太難了啦，而且我覺得找網路上的解答，一點也不為過吧，既然題目都是一樣的，就要大方接受網路的解答 Q\#Q 氣死了，稿的助教也很為難，說不定以後沒有人敢當助教了.

老師教課認真 如果上課步調能再慢一點會更好

份量太重了

我在想是不是因為沒有期中考，似乎越來越多人沒有跟上進度來，後來就不知道在上什麼 ^^::::

不過老師真的教的很好很用心很認真，還有老師的聲音好好聽喔 :D

1.助教對於作業的內容一問三不知,既然連助教都不懂作業,要如何回答我們問題?
2.我覺得教授很固執,因為作業假如一起寫,或請教別人,本來就會有一部份相似,你今天連十幾%的相似度也要抓出來,我覺得根本就是…

… 要求太寬鬆的部分連用都不能用,連一點點接受別人的指導與意見都不行?我就是那個地方不會寫叫我不能問人是怎樣?可是偏偏你出的作業根本就是抄別人國外大學的作業,那不是抄襲得更嚴重許多?
另外,拜託,大四了很多人推甄沒上要準備研究所的......你這個必修課load也太重了吧!!!

老師人很好，教學認真

作業的份量太重
花在作業的時間上非常非常多
i think the load of the assignments is too much for students.

老師教太快了@@\#

我覺得老師非常認真,課程內容準備的非常好,對同學也很好^_^
我有一點小建議,我覺得這門課的programming的load太重了,同樣的課如果擺在大三,其實load還沒有實驗重,或許也沒有compiler重,但是到了大四,大家都選了專題,很多人也決定未來要走哪個方向了,這時候一個要做很多programming的課可能成為一個負擔,如果我想走比較系統的路線,這門課提供的hands-on experience應該會是一個非常棒的經驗;但是如果我想讀理論方面(雖然這樣的人佔非常少數),我覺得這門課的programming份量足以消耗我過多的時間,讓我不太能夠去鑽研我真正想要讀的領域;因此我覺得以這門課的份量,或許應該放在大二,大三比較好,或是把這門課放在選修,讓真正有興趣的同學們去修.
還有我覺得老師如果開一個老師自己的研究領域的課,會相當精采,也會有多很有興趣的同學去修.

老師人很好

以一學期所上的內容來說,份量實在太多了,一個禮拜接近一章的速度會令人吃不消,尤其大四許多人有其他的考量（準備研究所考試,GRE,預官考試等等）,如果許可的話,我覺得刪掉一些教材會比較好,也可以將節省下來的時間拿來講解比較難懂的部分…

作業太重了,大四總共不是只有3學分,也不是每個人都很輕鬆,但是作業的時間可以花掉半個星期的作業時間,助教出的作業根本可以當專題,一份從國外抓的 P R O J E C T 什麼 D O C U M E N T 都沒有要我們自己看,又不能用自己之前寫好的東西（也未必有用）,這樣不知浪費多少時間.希望教授重新考慮教學方法,助教出作業及對同學的方法也可以改進,不然對我這個還想考研究所的人根本就不想接受. T R A C E 別人的 C O D E 是 D A T A B A S E 嗎?我覺得不是,一點都不值得,所以提起興趣的部分是非常不同意,實際上原本對 D A T A B A S E 有興趣（因為需要用到）現在也很不想看.

會講中文就用中文上課
用英文上課只有只會比較難理解
作業太多擠壓到念書時間
只有一門必修不代表只修一門課
作業拿國外大學的直接出
助教自己對作業都不了解
作業的reference也模模糊糊
… 非常大的困擾!

課業繁重 教學認真

作業十分繁重,對要準備研究所的同學來說,很吃力!

Page 1

- "Course load is too heavy"
- "English teaching"
- "Teaching staff copying programming assignments from abroad"

Evaluation II

- English communication is no problem.
  - Students do ask questions in Chinese & English
- Lots of complaints but students have handled challenges & pressures well
  - 95% completion rate across all assignments
  - avg final score: 76(2003)
  - Some problems with copying assignments (detected by MOSS)
- Overloading the 2 poor TAs before overloading the students
- Class attendance is a problem (33~50%)

Evaluation III

- Benefits to students
  - Gain Solid hands-on experiences with building systems
  - Learn how to read other people's code
  - Handle last minute deadline pressure
  - Teamwork
  - Maybe learned something about database

Overview of Microsoft Curriculum on Database

- MSDN Academic Alliance Developer Center Curriculum Depository
  - http://www.msdnaa.net/browse/
  - Computer Science Areas
    - Information management
- Developing SQL Server Apps
- (XML, ADO.NET, SQL Server CE, etc.)

Developing SQL Server Apps

- From DevelopMentor Inc.
- 15 Modules
- Emphasize on applications
  - How to use APIs provide by MS SQL server to develop applications?
  - Excellent hands-on lab assignments for each module

Developmentor Content Map I

From Developmentor Inc.

15 Modules:

#1: Architecture
- Client-server, 3 Tier, n-Tier
- Disconnected ops

#2: Database Objects
- Tables, views, indexes, stored procedures, triggers, etc.
- Built-in Data types, user-defined data types.

#3: SQL
- Specify Insert, update, delete, select statements
- Join, nested, aggregate
- Transaction , batch

#4: Stored Procedures
- Functions executed within dbms
- Interfaces to COM

Developmentor Content Map II

#5: Views & Triggers
- Specify views (tables whose rows are not explicitly stored).
- Specify Triggers (used to enforce constraints on the table.)

#6: Transactions
- Specify locking types (isolation levels vs. concurrency)
- Locking granularities
- Specify timer-based deadlock handling

#7: Distributed Transactions
- Access & modify different DBs on different servers
- Distributed transaction manager: coordinate locking, deadlock detection, commit/abort, etc.

#8 Declarative Transactions

Developmentor Content Map III

#9 Data Access Basics

#10 SQL server provider
- –OLE DB object model
- –ADO Object model

#11 XML & SQL server
- –XML, DTD,

#12 Managed Data Access
- –Access data from .NET execution engine

#13 Optimizing Database Objects
- –Specify indexing strategies (clustered, non-clustered)

#14 Optimizing Transactions
- –Bulk insertions and deletions
- –Minimizing traffic between client & server

#15 Optimizing Data Components
- –Processor objects
- –Caching, Load balancing

Developmentor Lab Assignments

- Relational
- DBObjects
- Transact SQL
- Procedures and Functions
- Views and Triggers
- Transactions, Isolation and Locking

- DTC
- COM+
- ADO
- SQLXML
- ADO.NET
- DB Perf
- T-SQL Perf

Comparisons

- MSDN materials much better organized and explained.
- Emphasize on how to use database and building DB applications
  - Less about Systems Implementation
- Excellent for IM students, or students who are prepared to go to IT industry
- Complementary materials for Systems-oriented courses:
  - Excellent case studies

Questions & Answers

Thank you