CSC 742
Database Management Systems

Dr. Peng Ning

Materials developed based on lecture notes from Dr. Munindar Singh
About TA

- Pai Peng, PhD student of computer science
- ppeng@eos.ncsu.edu
- Office hours:
  - TBD

Online Course Materials

- Course website:
  - http://courses.ncsu.edu/csc742/lec/001/
- Message board:
  - http://courses.ncsu.edu/csc742/
- Course mailing list:
  - csc742-001@wolfware.ncsu.edu
Prerequisites

- Graduate standing in computer science or computer engineering at NCSU.
- Knowledge of discrete mathematics and predicate logic.
- CSC 431 (File Organization and Processing) or 541 (Advanced Data Structures), or instructor's approval.
- Sufficient ability to program in Java or a willingness to acquire it through self-study.

Course Texts

- Required text
- Reference texts
  - Java Tutorial: JDBC Database Access
Assignments

- Eight assignments
  - All work is to be performed individually unless otherwise specified.
  - For the collaborative problems, you are encouraged to form teams of 1-3 members (of students in this class) to cooperate *only* on those problems. After discussing the problems, please write up your answers individually. Indicate the names of the other members in your team, if any.
  - You get no extra credit for working alone where collaboration is permitted.

Project

- All students are required to complete a course project.
- Please read the details on the course website.
Grading

- Assignments 5%
- Project 30%
- Midterm 30%
- Final 30%
- Class participation 5%.

Self-Study Responsibilities

- Some of the topics are important but are either quite straightforward or not a main focus of this course.
- These topics are identified as self-study topics above.
- Your knowledge of them will be evaluated as appropriate through exams, homework, programming assignments, or the project.
Rules

- The NC State University and Department of Computer Science rules regarding academic honesty apply

Important Dates

- Monday, January 14, 2002
  - Last day to add a course without permission of instructor.

- Friday, January 18, 2002
  - Last day to register (includes payment of tuition and fees) or to add a course. Last day to drop a course, or change from credit to audit with a refund or reduction.

- Friday, March 8, 2002
  - Last day to withdraw or drop a course without a grade at the 500-900 level. Last day to change from credit to audit at the 500-900 level.
Scope of this Course

- Directed at computer science graduate students
- Emphasizes concepts and theory
- Requires design and development of a database application
- Includes little DBMS-specific details—you learn those on your own
- Intensive!

Contents

- Concepts and architecture
  - Traditional software systems
  - Database Management System (DBMS)
  - Languages and interfaces
  - Conceptual modeling
Contents (Cont’d)

- DB programming basics
  - DBMS: Sybase with JDBC front-end
  - Development environment
  - Application servers

Contents (Cont’d)

- Data modeling
  - Conceptual models
    - Entity-relationship approach
    - Enhanced entity-relationship approach
  - Computational model
    - Relational model
      - Keys
      - Constraints
      - Operations
Contents (Cont’d)

- Database design
  - Design of relational schemas
  - EER to relational mapping
  - Information redundancy and anomalies
    - Functional dependencies
    - Normalization
      - 1NF
      - 2NF
      - 3NF
      - BCNF
  - Design process

Contents (Cont’d)

- Data Manipulation Language (DML)
  - Relational calculus
  - Relational algebra
  - SQL
    - Queries
    - Updates
    - Views
    - Constraints
    - Embedded SQL
Contents (Cont’d)

- Transaction
  - ACID properties
  - Schedules
  - Recoverability
  - Serializability
- Transaction programming
  - Three-tier architecture
  - Component programming
  - JDBC

Contents (Cont’d)

- Concurrency control
  - Locking
  - Timestamps
  - Optimistic techniques
  - Granularity
Contents (Cont’d)

- Recovery
  - Concepts
  - Deferred update
  - Immediate update
  - Shadow paging
  - Backup and recovery

Contents (Cont’d)

- Query processing and optimization
  - SQL to relational algebra
  - Basic algorithms
  - Heuristics
  - Semantic query optimization
Topic #1:
A Brief Introduction to DBMS

Database

- Coherent collection of data with inherent meaning
  - Random assortment of data is not a database.
- About an aspect of the universe of discourse
  - Changes in the universe of discourse are reflected in the database
- Fit to use for its intended purpose
  - Somebody is going to use the database
DB Lifecycle

- Define or model
- Construct or populate
- Manipulate: retrieve and update
- ...
- Redefine

DBMS: 1

Tool suite for DB lifecycle

- Recording the meaning of data
  - catalog
  - metadata
  - data types
- Storing the data elements
DBMS: 2

- Facilitating access and managing the effects of change
  - data abstraction
  - views

DBMS: 3

- Managing processes
  - encapsulating operations
  - supporting concurrency
  - protection against application failure
- System functions
  - backup
  - recovery from system failure
Players

- Database administrator (DBA)
- *Database designer
- Users
  - casual—need ad hoc queries
  - *naive—need canned transactions (supply parameters)
  - standalone—have personal DBs as well
- *System analysts
- *Application programmers
  (*: roles you will play in this course)

What databases can do

- Reduce redundancy
- Help keep the data consistent
- Give clean models of the data
- Help access the data
- Help manipulate the data
- Provide backup and recovery
Classifying DBMSs

- Data model: relational, network, hierarchical, O-O, O-R
- Number of users
- Centralized vs. distributed
- Homogeneous vs. heterogeneous