CSE532
Theory of Database Systems
Course Information

CSE 532, Theory of Database Systems

Stony Brook University

http://www.cs.stonybrook.edu/~cse532

Course Description

- "The *3 credits* course will cover advanced topics in modern database systems, including object-oriented databases, rule-based databases, temporal and active databases, parallel and distributed databases, distributed object model, data mining, on-line analytical processing, data warehousing, multimedia databases.."

- Prerequisites:
  - CSE 305 or equivalent,
  - You will need knowledge of Java, JDBC, servlets, and JSP,
  - You should have working knowledge of mathematical induction, logic, and related material.

General Information

- Meeting Information:
  - Lectures: TuTh 8:20AM - 9:40AM, Harriman 116
- Course Web page: http://www.cs.stonybrook.edu/~cse532
- Blackboard will be used for assignments, grades and course material

Instructor Information

- Dr. Paul Fodor
  1437 Computer Science Building
- Office hours: Tuesdays 10:00AM-12:00PM
  - I am also available by appointment
- Email: pfodor (at) cs (dot) stonybrook (dot) edu
- TAs: TBD
- Please include "CSE 532" in the email subject and your name in your email correspondence

Course Description (cont.)

- Grades will be based on homework submissions, projects and exams according to the following formula:
  - Homework assignments, quizes and projects -- 35%
  - Midterm exam -- 30%
  - Final exam -- 35%

Getting Help

- Keep up with the reading assignments and homeworks!

- Don't be afraid to see me and the TAs

- Form a study group and work together to learn the material – **not for copying homework solutions**

- If you are having personal problems, help is available. Call the Counseling Center, 632-6720. They also have workshops on building better study habits.

Textbook

- Database Systems: An Application Oriented Approach, 2nd Edition (Compete Version).

- Authors: Michael Kifer, Arthur Bernstein and Philip M. Lewis.

- Publisher: Addison Wesley; 2nd edition (March 26, 2005).

- ISBN-10: 0321268458.

- ISBN-13: 978-0321268457.

COMPLETE VERSION

Database Systems

An Application-Oriented Approach

SECOND EDITION

Michael **KIFER**   Arthur **BERNSTEIN**   Philip M. **LEWIS**
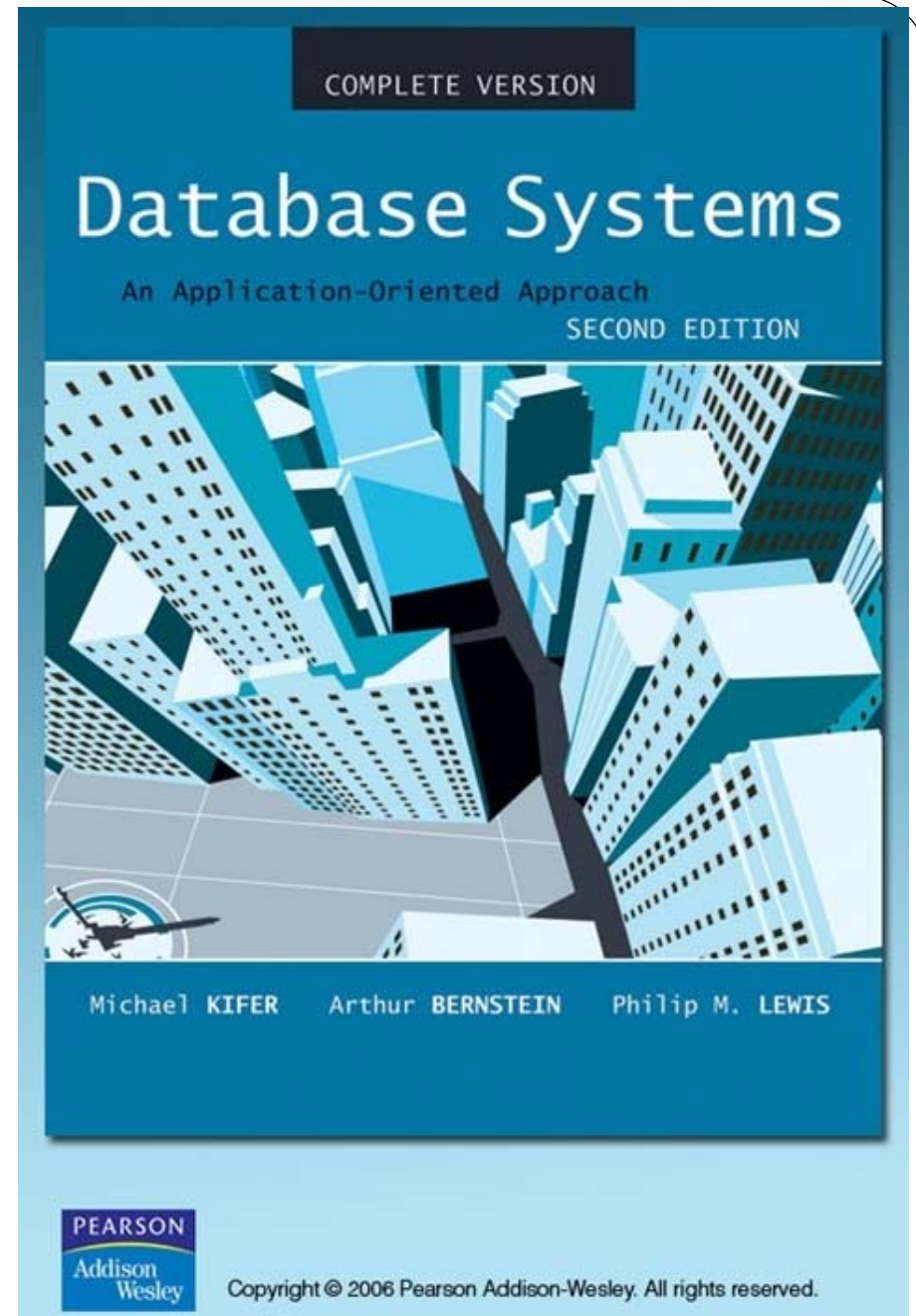
Important Dates

- Midterm Exam: Th 3/22, in class

    (8:20AM - 9:40AM, Harriman 116)

- Final Exam: Fr 5/11, 2:15PM - 4:45PM in classroom

    (Harriman 116)

    - **http://www.stonybrook.edu/registrar/finals.shtml**

Assignment Submission

- All assignments should be submitted paper printed at the beginning of class on due date

- I use Blackboard to send emails to the class, so make sure that your email address in Blackboard is up-to-date.

- Projects are due on Blackboard.

(c) Pearson and P.Fodor (CS Stony Brook)

Academic Integrity

- You can discuss general assignment concepts with other students

- You MAY NOT share assignments, source code or other answers
  - Assignments are subject to manual and automated similarity checking

- If you cheat, you MAY be brought up on academic dishonesty charges without warning - we follow the university policy:
  - http://www.stonybrook.edu/uaa/academicjudiciary

Disability

- If you have a physical, psychological, medical or learning disability, contact the DSS office at Room 128 ECC. Phone 632-6748/TDD

- If you are planning to take an exam at DSS office, you need to tell me ahead of time for every exam. Otherwise you may not be able to take it there

- **All documentation of disability is confidential**

Please

- Please be on time

- Please show respect for your classmates

- Please turn off (or use vibrate for) your cellphones

…

- On-topic questions are welcome

Textbook Chapter 1:
What is a Database?

- Collection of data central to some enterprise
- Essential to operation of enterprise
  - Contains the only record of enterprise activity
- An asset in its own right
  - Historical data can guide enterprise strategy
  - Of interest to other enterprises
- State of database mirrors state of enterprise
  - Database is persistent

# What is a Database Management System?

- A Database Management System (DBMS) is a program that manages a database:
  - Supports a high-level access language (e.g. SQL).
  - Application describes database accesses using that language.
  - DBMS interprets statements of language to perform requested database access.

What is a Transaction?

- When an event in the real world changes the state of the enterprise, a transaction is executed to cause the corresponding change in the database state
  - With an on-line database, the event causes the transaction to be executed in real time
- A transaction is an application program with special properties - discussed later - to guarantee it maintains database correctness

What is a Transaction Processing System?

- Transaction execution is controlled by a TP monitor
  - Creates the abstraction of a transaction, analogous to the way an operating system creates the abstraction of a process
  - TP monitor and DBMS together guarantee the special properties of transactions
- A Transaction Processing System consists of TP monitor, databases, and transactions
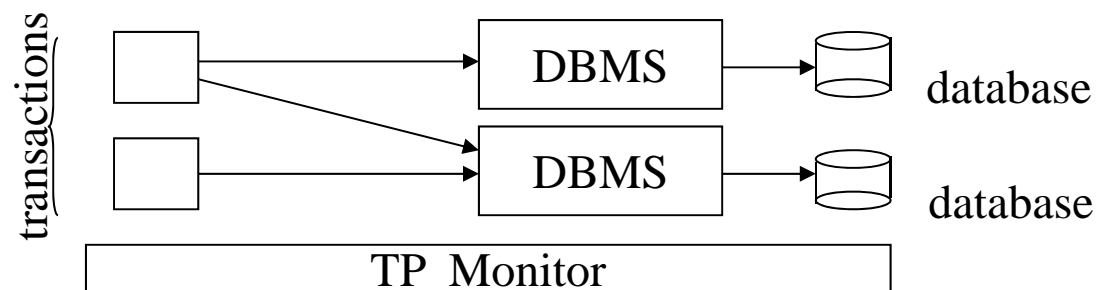
System Requirements

- **High Availability**: on-line => must be operational while enterprise is functioning
- **High Reliability**: correctly tracks state, does not lose data, controlled concurrency
- **High Throughput**: many users => many transactions/sec
- **Low Response Time**: on-line => users are waiting

System Requirements

- **Long Lifetime**: complex systems are not easily replaced
  - Must be designed so they can be easily extended as the needs of the enterprise change
- **Security**: sensitive information must be carefully protected since system is accessible to many users
  - Authentication, authorization, encryption

Roles in Design, Implementation and Maintenance of a TPS

- **System Analyst** - specifies system using input from customer; provides complete description of functionality from customer's and user's point of view

- **Database Designer** - specifies structure of data that will be stored in database

- **Application Programmer** - implements application programs (transactions) that access data and support enterprise rules

(c) Pearson and P.Fodor (CS Stony Brook)

Roles in Design, Implementation and Maintenance of a TPS

- **Database Administrator** - maintains database once system is operational: space allocation, performance optimization, database security

- **System Administrator** - maintains transaction processing system: monitors interconnection of HW and SW modules, deals with failures and congestion

OLTP vs. OLAP

- **On-line Transaction Processing** (OLTP)
  - Day-to-day handling of transactions that result from enterprise operation
  - Maintains correspondence between database state and enterprise state
- **On-line Analytic Processing** (OLAP)
  - Analysis of information in a database for the purpose of making management decisions

OLAP

- Analyzes historical data (terabytes) using complex queries
- Due to volume of data and complexity of queries, OLAP often uses a data warehouse
- **Data Warehouse –** (offline) repository of historical data generated from OLTP or other sources
- **Data Mining** - use of warehouse data to *discover* relationships that might influence enterprise strategy

Examples - Supermarket

- OLTP
  - Event is 3 cans of soup and 1 box of crackers bought; update database to reflect that event

- OLAP
  - Last winter in all stores in northeast, how many customers bought soup and crackers together?

- Data Mining
  - Are there any interesting combinations of foods that customers frequently bought together?