Introduction to Databases

Syllabus

Web Page

http://www.cs.northwestern.edu/~pdinda/db (Note: may change with reorganization of ECE/CS)

Instructor

Peter A. Dinda
Technological Institute L463
847-467-7859
pdinda@cs.northwestern.edu
Office hours: Thursdays, 2-4pm, or by appointment

Teaching assistant

Ananth Sundararaj
Ford 2-221
847-467-4708
ais@cs.northwestern.edu
Recitation section: Mondays, 6-7pm, Tech A110
Office hours: Wednesdays, Fridays, 11-12 (after class)
or by appointment

Location and Time

Technological Institute, M152, 10-10:50am

Prerequisites

Required	CS 311 or equivalent data structures course
Required	CS 213 or equivalent computer systems course
Recommended	Familiarity with concepts from discrete math such as set theory
Recommended	Some familiarity with Perl or other scripting language
Recommended	Familiarity with C/C++

Textbook and other readings

Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer D. Widom, *Database Systems: The Complete Book*, Prentice Hall, 2001 (Textbook)

- An in-depth introduction to databases and database implementation

Phillip Greenspun, *SQL for Web Nerds*, http://philip.greenspun.com/sql/. (Required)

- A great introduction to RDBMS systems from the perspective of a web application developer.

Joe Celko, *SQL for Smarties: Advanced SQL Programming*, 2nd edition, Morgan Kaufman, 1999. (Useful)

- A collection of wisdom on how working developers get useful things done in SQL.

Jim Gray, Andreas Reuter, *Transaction Processing: Concepts and Techniques*, Morgan Kaufman, 1993. (Related)

- Definitive book on transactions, a very important component of any modern database system.

Larry Wall, Tom Christiansen, Jon Orwant, *Programming Perl*, 3rd Edition, O'Reilly and Associates, 2000. (Useful)

- A detailed introduction to the Perl language. Your web-oriented projects in this class will be based on Perl CGI. You will need to know (or learn) only limited amounts of Perl.

Objectives, framework, philosophy, and caveats

This course introduces the underlying concepts behind data modeling and database systems using relational database management systems (RDBMS), the structured query language (SQL), and web applications (Perl DBI in CGI) as examples.

You will learn:

- How to model your data using the entity-relationship model
- How to design a normalized schema in the relational data model
- How to implement your schema using SQL
- How to keep your data consistent and safe with your schema using the ACID properties that a modern RDBMS gives you
- How to query your data using SQL
- How to interface to a modern RDBMS from a modern programming language.
- How such interfaces are used to create web applications
- How an RDBMS provides quick access to your data using indices, and how indices are implemented.
- How an RDBMS manages the storage hierarchy.
- How an RDBMS optimizes and execute your queries using the relational algebra, the theoretical underpinning of database systems.
- The history of database systems, including old ideas, like hierarchical databases, that are seeing a resurgence of interest today in the context of XML and LDAP.

- Current database-related hot topics in peer-to-peer systems, including search on distributed hash tables.

The textbook I have chosen is actually a combination of two books, an introduction to the concepts and use of databases and an introduction to the implementation of RDBMS systems. We will cover mostly the former. However, this is a very useful and essentially timeless book to have on your bookshelf for both elements.

This is a learn-by-doing kind of class. You will dive right in and modify a small database-based web application, a web log. Next, you will propose, design, and implement your own database-based web application. Finally, you will implement an on-disk BTree index. The majority of the programming in this class will be from scratch. We will use SQL, Perl, and C/C++ on Linux systems.

Projects

At the beginning of the course, I will provide you with a simple web application, a tiny web log ("blog"). Microblog is based on an Oracle database and provides a web interface using a CGI application written in Perl that talks to the database via DBI. This is a very common form of web application. You will spend three weeks learning how Microblog works and extending it in several simple ways. The goal is to immediately introduce you to SQL right away and bring you up to speed on the programming elements of the course that you'll need for the second project.

The second project is self-defined project. You will propose your own problem and spend the remainder of the quarter designing and implementing your solution. This project will take four weeks and the end result will be another web application.

The final project is to implement an on-disk BTree index, a key data structure used by a modern database system. This is lower level programming project that you will do in C or C++. This project will take three weeks.

Detailed descriptions of the requirements of the projects are available on the course web site. Please be sure to read them now so that you know what you're getting yourself into.

Homework

There will be three (maybe four) homework problems sets that will be periodically assigned to help you improve your understanding of the material.

Exams

There will be a midterm exam and a final exam. The midterm exam will take place in the evening outside of class. The final exam will not be cumulative.

Grading

10 % Dry-run project (Microblog)
25 % Self-defined project (Your own web application)
15% Implementation project (Btree index)
20 % Midterm
20 % Final
10 % Homework

Final grades will be computed in the following way. A final score from 0 to 100 will be computed as a weighted sum of each of the projects, the homeworks, and the exams. Scores greater than 90 or greater than 90th percentile will be assigned As, scores greater than 80 or greater than 80th percentile will be assigned Bs, scores greater than 70 or greater than 70th percentile will be assigned Cs, scores greater than 60 or greater than 60th percentile will be assigned Ds, and the remainder will be assigned Fs. Notice that this means that if everyone works hard and gets >90, everyone gets an A. Please choose wisely where you put your time.

Late Policy

For each calendar day after the due date for a homework or a lab, 10% is lost. After 1 day, the maximum score is 90%, after 2 days, 80%, etc, for a maximum of 10 days.

Cheating

Since cheaters are mostly hurting themselves, we do not have the time or energy to hunt them down. We much prefer that you act collegially and help each other to learn the material and to solve development problems than to have you live in fear of our wrath and not talk to each other. Nonetheless, if we detect blatant cheating, we will deal with the cheaters as per Northwestern guidelines.

Schedule

Lecture	Date	Topics	Readings	Homework and Project
1	9/21	Class mechanics Introductory material, Web applications, client/server, and three-tier	GUW Intro; PG preface + 1	Project A (Microblog) out
If you're unfamiliar with Unix, now would be a good time to view the Unix introduction video available from the course web site.				
2	9/23	More introductory material: why a database is different from a filesystem and what it helps you with. Data modeling, transactions/ACID,	GUW 1; PG preface + 1	

		queries, abstracting storage+indices, some history lessons (Hierarchical, Network, Relational, Object, Object Relational, Hierarchical again). Hot stuff: P2P		
3	9/26	How web applications work. Apache, CGI, Perl, DBI, RDBMS, SQL in a nutshell	PG 1-7, Perl HO, Oracle HO	
4	9/28	SQL in a nutshell, Walk through Microblog (SQL) *Instructor out of town*	PG 1-7, Perl HO, Oracle HO	Note: you might find PG 10 useful reading
5	9/30	Perl in a nutshell *(Ananth Sundararaj)*	PG 1-7, Perl HO, Oracle HO	
6	10/3	Guest Lecture on P2P *(Ashish Gupta, David Huber, Jay Bruins)*	None	
7	10/5	More Perl	PG 1-7, Perl HO, Oracle HO	
8	10/7	Walk through Microblog (Perl)	PG 1-7, Perl HO, Oracle HO	
9	10/10	Walk through Microblog (Perl)	PG 1-7, Perl HO, Oracle HO	HW 1 (ER Modeling) out
10	10/12	Walk through Microblog (Perl). More SQL if time	PG 1-7, Perl HO, Oracle HO	
11	10/14	Data models and Data modeling: Why? Start Entity-Relationship: Entity sets, attributes, relationships, ER diagrams, instances, multiplicity, roles, multiway	GUW 2	Project A (Microblog) in. Project B (Self-defined) out
12	10/17	Entity-Relationship Model: conversion to binary relationships, subclassing, design principles	GUW 2	
13	10/19	Entity-Relationship Model: constraints, weak entity sets	GUW 2	HW 1 in, Project B proposal due, HW 2 out
14	10/21	Relational Data Model: basics, translating from ER to relational	GUW 3	
15	10/24	Relational Data Model: subclasses, functional	GUW 3	

		dependencies		
16	10/26	Relational Data Model: Schema design and normal forms	GUW 3	Project B spec/ER/relational due; HW 2 due
Midterm Exam, Thursday, 10/27, 6-7:30pm, Room TBD *Covers Everything Through Lecture 16, excluding P2P lecture*				
17	10/28	Relational Data Model: Multivalued dependencies	GUW 3	
18	10/31	Other data models: OO and XML	GUW 4	HW 3 out
19	11/2	Relational Algebra: Sets: union, intersection, difference, selection, projection, Cartesian product, and *spooky* joins;	GUW 5	
20	11/4	Relational Algebra: Bags, equivalent expressions, some extended operators	GUW 5	Project B DDL/DML/ Queries/App Logic Due
21	11/7	Relational Algebra: grouping, constraints, data-mining	GUW 5	
22	11/9	SQL: strings, regular expressions, date/time, nulls, 3-valued logic, explain plan, subqueries in/exists/>all/>any, correlation	GUW 6	Project B Final Handin; Project C out
23	11/11	SQL: insert/update/delete, multi-statement transactions using PL/SQL; create schemas: bit-fields, decimal, blob; drop, alter; indexes; views	GUW 6	
24	11/14	SQL: Constraints, Triggers, systems aspects.	GUW 7, 8.1, 8.2, 8.3, 8.4, 8.6	
25	11/16	Implementation: Storage	GUW 11	HW 3 due
26	11/18	Implementation: Representing Data	GUW 12	
27	11/21	Implementation: Indexes, Btrees	GUW 13, 14.4	
Thanksgiving Break				
28	11/28	Implementation: Indexes, Hashes	GUW 13, 14.4	
29	11/30	Implementation: Indexes, Bitmaps	GUW 13, 14.4	
30	12/2	Implementation: Transactions	GUW 17.1-4, 18.1-18.3,18.8	Project C due
Final Exam, Tuesday, 12/6, 9am-11am, Room TBD. Covers Lectures 16-30				

PG = Phillip Greenspun, *SQL for Web Nerds*
GUW = Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer D. Widom, *Database Systems: The Complete Book*