INFO-445: Advanced Database Design, Management, and Maintenance

Spring 2013
B.S. Informatics
Information School
University of Washington

Advanced perspectives on DBMS theory, architecture and implementation. Conceptual, logical and physical modeling. Index structures, query optimization and performance tuning, relational algebra, transaction processing and concurrency control. Operational databases, decision support systems and data warehousing. Projects in database implementation and integration. Social implications of large distributed database systems. *Prerequisite*: INFO 340, CSE 373.

Course website, Listserv and Dropbox

Course website: http://courses.washington.edu/info445/

Dropbox: info445a_sp13@uw.edu
(Archive: https://mailman1.u.washington.edu/mailman/admin/info445a_sp13)
Registered students are subscribed automatically using their UW mail account.

Dropbox: https://catalyst.uw.edu/collectit/dropbox/dhendry/26782

Credit Hours

5 (5 lecture hours)

Meeting times

Lecture	Wednesday/Friday 3:30 – 5:20, CMU 228
Studio	Wednesday 5:30-6:20, MGH 430

Instructor

David Hendry, Associate Professor
330P Mary Gates Hall
dhendry@u.washington.edu | http://faculty.washington.edu/dhendry

Office hours: by appointment.

Teaching assistant

To be decided

Student services

Dowell Eugenio, Student Services Administrator
470E Mary Gates Hall
deugen3@u.washington.edu

Please note: If you have any concerns about a course or the TA, please see the TA about these issues as soon as possible. If you are not comfortable talking with the TA or not satisfied with the response that you receive, you may contact the instructor of the course. If you are still not satisfied with the response that you receive, you may contact Joseph Janes, the Associate Dean for Academics in 370 Mary Gates Hall, by phone at : (206) 616-0987, or by e-mail at jwj@u.washington.edu. You may also contact the Graduate School at G-1 Communications Building, by phone at (206) 543-5900, or by e-mail at efeetham@u.washington.edu

Overview

In this course, we shall as one major question: What are the merits of different kinds of data models for representing, storing, and accessing information? To engage this question we shall engage the following questions through the quarter:

1. What the central problems that all database models must address? How do database systems enable structured, unstructured and semi-structured data to be represented, accessed, and updated? What are the key issues and approaches to data storage and access?
2. Why did object-oriented databases emerge and then evidently fade away? Why have "NoSQL" databases emerged? How are they better (or worse) than the relational model?
3. What is required of a database so that XML documents can be stored, updated, and accessed efficiently? How does XML integrate with relational databases? How are fundamental database issues, such as data integrity, query processing, concurrency, and efficient update and access, addressed with object-oriented and XML approaches?
4. What is the cloud? How do developments in cloud data services and computing improve upon the relational model (if at all)?
5. How does an application designer decide on a particular data model?

By answering these and similar questions, you will expand your knowledge for databases. In addition, you will be better equipped to tackle more complex data modeling problems and to learn more about databases on your own.

We shall seek to answer these questions by first deepening our knowledge by covering the following advanced topics: 1) Rules for transforming conceptual models to relational schemas (conceptual and logical database design); 2) Performance monitoring and tuning (physical database design); 3) Transaction management; and 4) Query Processing. These topics shall be considered within a database design methodology, which covers conceptual, logical, and physical database design through a serious of stepwise refinements.

Then, in the second part of the course, we shall turn to an examination of the Cloud, NoSQL databases, and other related data models. Throughout the course we shall study the concepts that underpin these models and then examine the practical aspects of their use through weekly activities. Finally, you will have an opportunity to engage a substantial database problem.

Textbooks and readings

The textbook for this is course is:

> Connolly, T. M. & Begg, C. E. (2009). *Database Systems: A Practical Approach to Design, Implementation, and Management* (5th Edition). New York: Addison-Wesley Publishing.

In addition, the textbook will be supplemented with several readings, which will be posted on the course website.

Learning

Aims

The general aims of this course are to:

1. Expand and deepen knowledge for the relational database model
2. Develop knowledge for alternative data models, including XML/DB integration, NoSQL databases, and cloud data services.

Objectives

On the successful completion of this course, you should be able to:

1. Apply a stepwise refinement methodology for transforming a complex conceptual model into a database schema
2. Discuss the major issues, design approaches, and implementation strategies for storing and searching structured, unstructured, and semi-structured data
3. Implement a three-tier database system
4. Describe how the performance of a database can be monitored and improved and use EXPLAIN diagnostically to improve the performance of a database system
5. Discuss and apply approaches for improving a database performance, including denormalizat-ion, use of indexes, and data partitioning
6. Discuss the major features of NOSQL databases and cloud computing services
7. Design and implement data storage and access solutions involving structured, semi-structured, and unstructured data on the cloud
8. Demonstrate skills for design and implementing database systems.

Assessment

Work	Worth
Project	50%
Implementation activities	30%
Weekly DB questions	20%

Project

The project will give you an opportunity to develop your system design, modeling, and implementation skills. The project brief and deliverables are found on a separate handout.

Individual Assignments

The individual assignments are intended to help you develop your technical skills in databases. You will complete four individual assignments. Assignment details can be found on separate handouts.

Weekly DB Questions

The weekly DB questions are intended to position you to understand the weekly readings and be ready to engage key concepts during class. You will complete about 5 weekly DB question sheets. The questions and due dates are found on a separate handouts. *Please note*: Late submissions will normally not be graded.

Standard cover sheet

For the individual assignments and for the project deliverables if you like please included a cover sheet with the following information:

- Course name,
- Quarter, program, department, and university
- Assignment name
- Your name and e-mail address
- A date
- A web site address (if relevant).

Grading criteria

Work in this course will be graded to criteria. In other words, you won't be graded on a curve. Each deliverable is designed to test your achievement against one or more of the learning objectives. Different assignments emphasize different learning objectives. The meanings of grades are described below.

General grading information for the University of Washington is available at:

- http://www.washington.edu/students/gencat/front/Grading_Sys.html

The iSchool has adopted its own criteria for grading graduate courses. The grading criteria used by the iSchool are available at:

- http://depts.washington.edu/grading/practices/guidelin.htm

Grade	Performance Quality*
3.9 - 4.0	Superior performance in all aspects of the course with work exemplifying the highest quality. Unquestionably prepared for subsequent courses in field.
3.5 - 3.8	Superior performance in most aspects of the course; high quality work in the remainder. Unquestionably prepared for subsequent courses in field.
3.2 - 3.4	High quality performance in all or most aspects of the course. Very good chance of success in subsequent courses in field.
2.9 - 3.1	High quality performance in some of the course; satisfactory performance in the remainder. Good chance of success in subsequent courses in field.
2.5 - 2.8	Satisfactory performance in the course. Evidence of sufficient learning to succeed in subsequent courses in field.
2.2 - 2.4	Satisfactory performance in most of the course, with the remainder being somewhat substandard. Evidence of sufficient learning to succeed in subsequent courses in field with effort.
1.9 - 2.1	Evidence of some learning but generally marginal performance. Marginal chance of success in subsequent courses in field.

*Taken from Faculty Resource on Grading

Late policy

1. For the individual assignments and project deliverables, if you will miss the deadline, you should inform the instructor as soon as you can, indicating when you will submit the work. The instructor will try to accommodate your needs. You should use this clause only for extraordinary personal reasons.
2. It is at the instructor's discretion to accept late work or assign late penalties (see 1 above). Typically, for any late assignment, 10% will be taken off your work per day. After five days, your work will not be accepted.

Work that is handed in late is penalized for two reasons. First, to be fair, all students should be given the same time limits. Second, if you spend too much time on one assignment, it is quite likely that you will have insufficient time to spend on subsequent assignments.

Right to Revise

The instructor reserves the right to revise this syllabus.

Re-grading policy

To have work re-graded, you must submit a Re-grade Request within five days of when your work was returned. The request must be a single page, printed on paper or sent by e-mail. It should contain the following information:

- Re-grade Request
- The information contained on the standard cover sheet
- An explanation for why you believe you deserve a higher grade.

The instructor, possibly in collaboration with the teaching assistant, will consider your request. If the instructor is convinced by your argument, your work will be re-graded. If not, the instructor will send you e-mail explaining why. No re-grades will be considered for late work.

Guidelines on Using E-mail

When communicating with the instructor or teaching assistant, please follow these guidelines:

- You are welcome to give feedback to the instructor and teaching assistant about the course, to ask a question about an assignment, to share an interesting article or resource, to report that you will be absent from a class/lab, to request additional time for an assignment (because of significant health, personal, or educational matter), or similar communication ;
- Whenever appropriate, please copy the class listserv with your question or comment;
- If your e-mail concerns your grade, please follow the re-grading policy (see above);
- E-mail that is sent on Friday afternoon or over the weekend might not be replied to until Monday or Tuesday of the following week;
- If you don't receive a reply within 2 days or so, please resend your e-mail or ask about it during class or lab.

Academic Accommodations

To request academic accommodations due to a disability, please contact Disabled Student Services: 448 Schmitz, 206-543-8924 (V/TTY). If you have a letter from DSS indicating that you have a disability which requires academic accommodations, please present the letter to me so we can discuss the accommodations you might need in the class.

Academic accommodations due to disability will not be made unless the student has a letter from DSS specifying the type and nature of accommodations needed.

For additional information, see *Statements to Ensure Equal Opportunity and Reasonable Accommodation*, downloaded March 5, 2003, http://www.washington.edu/admin/eoo/eoost.html

Academic Honesty

The essence of academic life revolves around respect not only for the ideas of others, but also their rights to those ideas and their promulgation. It is therefore essential that all of us engaged in the life of the mind take the utmost care that the ideas and expressions of ideas of other people always be appropriately handled, and, where necessary, cited. For writing assignments, when ideas or materials of others are used, they must be cited. The format is not that important–as long as the source material can be located and the citation verified, it's OK. What is important is that the material be cited. In any situation, if you have a question, please feel free to ask. Such attention to ideas and acknowledgment of their sources is central not only to academic life, but life in general.

Please acquaint yourself with the University of Washington's resources on academic honesty:

http://depts.washington.edu/grading/issue1/honesty.htm

Students are encouraged to take drafts of their writing assignments to the Writing Center for assistance with

using citations ethically and effectively. Information on scheduling an appointment can be found at:

http://www.uwtc.washington.edu/resources/eiwc/

Copyright

All of the expressions of ideas in this class that are fixed in any tangible medium such as digital and physical documents are protected by copyright law as embodied in title 17 of the United States Code. These expressions include the work product of both: (1) your student colleagues (e.g., any assignments published here in the course environment or statements committed to text in a discussion forum); and, (2) your instructors (e.g., the syllabus, assignments, reading lists, and lectures). Within the constraints of "fair use", you may copy these copyrighted expressions for your personal intellectual use in support of your education here in the iSchool. Such fair use by you does not include further distribution by any means of copying, performance or presentation beyond the circle of your close acquaintances, student colleagues in this class and your family. If you have any questions regarding whether a use to which you wish to put one of these expressions violates the creator's copyright interests, please feel free to ask the instructor for guidance.

Privacy

To support an academic environment of rigorous discussion and open expression of personal thoughts and feelings, we, as members of the academic community, must be committed to the inviolate right of privacy of our student and instructor colleagues. As a result, we must forego sharing personally identifiable information about any member of our community including information about the ideas they express, their families, life styles and their political and social affiliations. If you have any questions regarding whether a disclosure you wish to make regarding anyone in this course or in the iSchool community violates that person's privacy interests, please feel free to ask the instructor for guidance.

Knowing violations of these principles of academic conduct, privacy or copyright may result in University disciplinary action under the Student Code of Conduct.

Student Code of Conduct

Good student conduct is important for maintaining a healthy course environment. Please familiarize yourself with the University of Washington's Student Code of Conduct at:

http://www.washington.edu/students/handbook/conduct.html

Class Schedule

Please review chapters 1-5 and chapters 12-13.

Week 1 (04/01): Overview

Read *Claremont Report* (on website)
L1 Greetings and Introduction
L2 The relational model
Lab Technology set-up I

Week 2 (04/08): Design Techniques

Read *Review chapters 12-13*
L1 Entity-Relationship models
L2 Guest lecturer
Lab Technology set-up II

Week 3 (04/15): Methodology - Conceptual and Logical Database Design

Read *Chapters 15 & 16*
L1 Stepwise refinement
L2 Deriving relations
Lab The project: Team meetings/presentations

Week 4 (04/22): Methodology - Physical Database Design

Read *Chapters 17 & 18*
L1 Query processing & indexes
L2 Monitoring and tuning
Lab The project: Team meetings/presentations

Week 5 (04/29): Query Processing

Read *Chapter 23*
L1 File organizations
L2 The planner and EXPLAIN
Lab The project: Team meetings/presentations

Week 6 (05/06): The Cloud, I

Read TBD
L1 Introduction and use scenarios
L2 TBD
Lab The Project: meetings/presentations

Week 7 (05/13): The Cloud, II

Read TBD
L1 TBD
L2 TBD
Lab The Project: meetings/presentations

Week 8 (05/20): NoSQL Databases, I

Read TBD
L1 Introduction and use scenarios
L2 Alternative storage and access methods
Lab The Project: meetings/presentations

Week 9 (05/27): NoSQL Databases, II

Read TBD
L1 Introduction and use scenarios
L2 TBD
Lab The Project: meetings/presentations

Week 10 (06/03): The Projects!

Read No readings
L1 Presentations
L2 Presentations
Lab The project: Open wildcard

Notes:
[1]Chapters refer to the course textbook
[2]*PostgreSQL* refers to sections from the PostgeSQL 9.1 documentation, which is available here: http://www.postgresql.org/docs/9.1/static/