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

Autumn 2010
B.S. Informatics
Information School
University of Washington


Course website & listserv
http://courses.washington.edu/info445/
info445a_au10@u.washington.edu
(Archive: https://mailman1.u.washington.edu/mailman/listinfo/info445a_au10)
Registered students are subscribed automatically using their UW mail account.

Dropbox for submitting assignments
https://catalyst.uw.edu/collectit/dropbox/dhendry/11939

Credit hours
5 (5 lecture hours)

Meeting times

Lecture Tuesday/Thursday 3:30 – 4:50, EEB 003
Lab/Studio Wednesday, 3:30-5:20, MGH 430

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

Office hours: Wednesday, 2:00 – 3:00 or by appointment.

Teaching assistant
Kristen Shinohara, Ph.D. Student, kshino@uw.edu

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

Please note: If you have any concerns about the 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 Matthew Saxton, the Associate Dean for Academics.
Overview

In this course, we shall engage one major question: What are the merits of different kinds of data models for representing, storing, and accessing information? More specifically, we shall seek answers to the following questions:

1. What are the central problems that all database models must address?
2. Why did object-oriented databases emerge? How are they better (or worse) than the relational model? How do Object-Relational models seek to integrate both the “object” and “relational” approaches to represent, store, and access data?
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. How do developments in cloud data services and computing, NoSQL databases, and flexible component-based data models improve upon or complement the relational database model?
5. How does an application designer decide on a particular data model?

By engaging these questions, you will expand your knowledge for databases and be better positioned to tackle more complex data modeling problems, and to learn more about databases on your own.

In the first part of the course, after reviewing the relational model, 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 examine a selection of alternative data models and access methods, including Object-Relational Databases, XML/DB integration, and cloud data services and computing. 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 course is:


In addition, several supplemental readings will be posted on the course website.
Learning
Aims
The general aims of this course are to:
1. Develop skills for professional written and oral communication
2. Expand and deepen knowledge for the relational database model
3. Develop knowledge for alternative data models, including object-oriented, object-relational, XML/DB integration, and cloud data services and computing.

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. Describe how the performance of a database can be monitored and improved
3. Discuss and apply approaches for identifying and improving a database’s performance, including denormalization, use of indexes, and data partitioning
4. Discuss concurrency control and describe major kinds of concurrency problems, and discuss database mechanisms for achieving Atomicity, Consistency, Isolation, and Durability. Employ the SQL commands BEGIN, COMMIT, and ROLLBACK to maintain consistency between the database and a world model
5. Describe the major features of Object-Oriented Database Management Systems
6. Discuss Object-Relational model and the problems it seeks to address
7. Create and query simple XML documents, stored in a relational database
8. Discuss the major approaches for integrating XML and design database applications that use both the relational and semi-structured data models
9. Discuss recent technologies in cloud computing and approaches for making the relational model less monolithic and more flexible.

Academic Conduct
The following paragraphs discuss academic integrity, copyright and privacy concerns governing student conduct in the iSchool and the University of Washington. They apply to all assignments and communications in this course.

Academic Integrity
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 it is consistent, the source material can be located and the citation can be verified. In any situation, if you have a question, please feel free to ask the instructor or teaching assistant. 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.

**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 download or copy slides, recordings or notes for your personal intellectual use in support of your education here in the iSchool. All of these examples are copyrighted expressions, and fair use by you does not include further distribution by any means of copying, performance or presentation beyond the circle of your student colleagues in this class. 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.

**Students with Disabilities**

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 the instructor so you 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.

**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.
**Assessment**

<table>
<thead>
<tr>
<th>Work</th>
<th>Worth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Six weekly activities</td>
<td>50%</td>
</tr>
<tr>
<td>Project</td>
<td>30%</td>
</tr>
<tr>
<td>Special topic report/presentation</td>
<td>10%</td>
</tr>
<tr>
<td>Participation</td>
<td>10%</td>
</tr>
</tbody>
</table>

**A: Weekly Activities**

Weekly activities will be due throughout the quarter, as described in the table below. Each activity is divided into two parts, with different due dates. Part I is intended to position you to understand the weekly readings for class. Part II is a focused implementation activity, which will allow you to engage a problem in some depth, typically with a programming task.

You will complete six activities, with the top five scores counting to your overall score. You may complete these activities alone or in groups of two. If you work in groups of two you are expected to work synthetically, where each member fully works on the whole assignment.

If you work in groups of two **PLEASE** clearly indicate who your partner is on the assignment. Throughout the quarter, you may work with as many different partners as you like.

For all Part-I’s, please bring a printed copy of your assignment to class. You will need it for class discussion and you will hand it in to the instructor. In addition, you must upload the assignment to the course’s dropbox. For all Part-II’s, please upload the assignment to the course’s dropbox.

Please upload by Part-I and Part-II to the dropbox by 3:30 P.M. of the due date. A link to the dropbox can be found on the course webpage.

**Due dates for Activities, A1 – A6**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Part I</th>
<th>Part II</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1. Entity-Relationship Modeling, SQL, and other key concepts (Week #2)</td>
<td>Oct 05</td>
<td>Oct 14</td>
</tr>
<tr>
<td>A2. Conceptual and logical database design (Week #3)</td>
<td>Oct 12</td>
<td>Oct 21</td>
</tr>
<tr>
<td>A3. Performance monitoring and tuning (Week #5)</td>
<td>Oct 26</td>
<td>Nov 04</td>
</tr>
<tr>
<td>A4. Object-relational databases (Week #6)</td>
<td>Nov 02</td>
<td>Nov 11</td>
</tr>
<tr>
<td>A5. XML/SQL integration (Week #7)</td>
<td>Nov 09</td>
<td>Nov 18</td>
</tr>
<tr>
<td>A6. Cloud data services (Week #8)</td>
<td>Nov 16</td>
<td>Dec 02</td>
</tr>
</tbody>
</table>

Part I of each assignment, weighted at 20%, will be given one of the following five grades:

- **✓--** 2.0 Unsatisfactory (should be redone and resubmitted)
- **✓-** 2.5 Needs improvement
- **✓** 3.0 Satisfactory
- **✓+** 3.6 Very Good
- **✓++** 4.0 Outstanding (very hard to obtain)

Part II, weighted at 80%, will be given a numeric grade between 0.0 and 4.0.
D: Project
You will complete a project. To receive ongoing feedback, the project will be structured with the following deliverables. All deliverables should be submitted to the dropbox by 3:30 P.M. of the due date. For grading convenience, please bring a hardcopy of each deliverable to lab/class.

<table>
<thead>
<tr>
<th>Deliverables</th>
<th>Due</th>
<th>Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1. Draft proposal</td>
<td>Oct 13</td>
<td>3</td>
</tr>
<tr>
<td>D2. Final proposal</td>
<td>Nov 3</td>
<td>6</td>
</tr>
<tr>
<td>D3. Architectural design/implementation plan</td>
<td>Nov 3</td>
<td>6</td>
</tr>
<tr>
<td>D4. Implementation/testing plan</td>
<td>Nov 17</td>
<td>8</td>
</tr>
<tr>
<td>D5. Final presentation</td>
<td>TBD</td>
<td>10-11</td>
</tr>
<tr>
<td>D6. Final Report</td>
<td>Dec 9</td>
<td>11</td>
</tr>
</tbody>
</table>

The final report will be weighted 50%, the final presentation 25%, and with the other deliverables in total weighted at 25%.

T: Special topic presentation
You will select a special topic concerning some aspect of data management and lead the class in a brief discussion on the topic. You will work in groups of two.

The aim of the special topics is to delve into some aspect of databases and to discuss that area in-depth.

Possible topics include, but are not limited to: distributed databases, data warehousing, data mining, data analytics, object-oriented databases, semi-structured databases, database security, comparison of database implementations, history of databases, end-user programming of databases, user interfaces for databases, cloud data services and techniques in cloud computing, lightweight databases, NoSQL databases, databases for specific application areas, among others.

You will be asked to select an introductory reading on the topic, present an overview of the topic, and lead a brief discussion.

<table>
<thead>
<tr>
<th>Deliverables</th>
<th>Due</th>
<th>Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1. 1-page proposal</td>
<td>Oct 27</td>
<td>5</td>
</tr>
<tr>
<td>T2. Report and presentation</td>
<td>TBD</td>
<td>9</td>
</tr>
</tbody>
</table>

The presentation and report will each be weighted at 50%
P: Participation
It is important to the instructor that INFO-445 be challenging, relevant, and fun. With spirit and a professional manner, we can create a supportive and rewarding learning environment. Among the things you can do are:

1. Please focus on professionalism – use words carefully, be on-time with assignments, be concise and tidy in written and graphic work, speak clearly
2. Treat all with respect – be constructive in all discussions
3. Come to class prepared – read carefully and be ready for discussion
4. If you are ill, please stay at home and get better – the instructor and teaching assistant will understand and will try to accommodate your needs
5. Be an active listener – be attentive, be engaged, use in-class technology with discretion, leave Sudoku and crossword puzzles behind
6. Ask challenging questions, participate in discussion
7. Comment, build on, or clarify what others have done or said
8. Help your classmates use development tools and technologies – build infrastructure
9. Post useful or interesting information to the class discussion list
10. Visit the instructor during office hours to chat, to ask questions, or to give feedback
11. Talk to the teaching assistant about how the class is going.

Participation will be graded based on: a) The frequency and quality of your contributions; b) Your ability to ask challenging questions and to clarify and build upon conversation; c) Your willingness to give room to everyone to contribute in class; and f) Your willingness to help build useful class infrastructure and share your knowledge. Optional – Write a 1-page personal statement on how you contributed to the class (optional) and upload it to the dropbox:

<table>
<thead>
<tr>
<th>Deliverables</th>
<th>Due</th>
<th>Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. Statement of participation (optional)</td>
<td>Dec 09</td>
<td>11</td>
</tr>
</tbody>
</table>
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

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

*Taken from Faculty Resource on Grading, downloaded March 5, 2003, http://depts.washington.edu/grading/practices/guidelin.htm

Standard cover sheet

To protect your privacy when exercises are returned and to facilitate communication, submitted work must have a cover sheet. The cover sheet should include the following information and be formatted nicely:
- Course name,
- Quarter, program, department, and university
- Assignment name
- Your name and e-mail address
- A date
- A web site address (if relevant).

Please staple together multiple pages.
Late policy
1. 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). For any late assignment, 10% will be taken off your work per day. After five days, your work will not be accepted.
3. Late work should be submitted to the dropbox and an email sent and, if possible, a paper copy should be handed to the instructor or teaching assistant.

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 should be a brief note along with the original assignment. 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. In general, late work will not be re-graded.

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;
- E-mail concerning assignments might not be replied to if it is sent within 36hr of the assignment due date;
- 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 is generally not replied to until Monday or Tuesday of the following week;
- If you do not receive a reply within 2 days or so, please resend your e-mail or ask about it during class or lab.
Class Schedule*

Week 1 (09/27): Overview
Read  Claremont Report (on website)
Lab  Greetings
L2  BIG questions

Week 2 (10/04): Relational Model
Read  Chapters 1-6 (Review)
L1  The relational data structure [A1.I]
Lab  Technology set-ups
L2  The relational operators

Week 3 (10/11): Design Techniques
Read  Chapters 11, 12 & 13
L1  Entity-Relationship models [A2.I]
Lab  Stored Proc./Team meetings [D0]
L2  Normalization [A1.II]

Week 4 (10/18): Methodology - Conceptual and Logical Database Design
Read  Chapters 15 & 16
L1  Stepwise refinement
Lab  Design Tools/Team meetings
L2  Deriving relations [A2.II]

Week 5 (10/25): Methodology - Physical Database Design
Read  Chapters 17 & 18 [A3.I]
L1  Query processing & indexes
Lab  Design Tools/Team meetings [T1]
L2  Monitoring and tuning

Week 6 (11/01): Object-Relational Databases
Read  Chapter 28
L1  The Third Manifesto
Lab  Demo/Team Meetings [D2|D3]
L2  PostgreSQL extensions [A3.II]

Week 7 (11/08): XML/DB Integration
Read  • Chapter 30
     • Beyer et al., 2006 (on website)
L1  Uses of XML [A5.I]
Lab  Demo/Team Meetings
L2  Hybrid approaches [A4.II]

Week 8 (11/15): Cloud Data Services
Read  Dean & Ghemawat, 2008 (on website)
L1  NoSQL [A6.I]
Lab  Demo/Team Meetings [D4]
L2  The MapReduce model [A5.II]

Week 9 (11/22): Special Topics
Read  No readings
L1  Student micro-presentations [T2]
Lab  Student micro-presentations [T2]
L2  No class – Thanksgiving Day

Week 10 (11/29): Project
Read  No readings
L1  Student presentations [D5]
Lab  Student presentations [D5]
L2  Student presentations [A6.II]

Week 11 (12/06): Projects
Read  No readings
L1  Student presentations [D5]
Lab  Student presentations [D5]
L2  Student presentations [D6|P]