CS 5614 Database Management Systems Naren Ramakrishnan

What?

From the Graduate Catalog

Emphasizes concepts, data models, mechanisms, and language aspects concerned with definition, organization, and manipulation of data at a logical level. Concentrates on relational model, plus some semantic, logic-based, object-oriented, and knowledge-based modeling. Functional dependency and normalization of relations. Relational algebra, relational calculus, mapping languages, graphical languages. Pre: Graduate standing; (3H,3C). I.

Instructor's Perspective

CS 5614 will represent something of a departure in the teaching of the introductory graduate course on database management systems. It will be taught from a practitioner's point of view and will cover all the three major aspects of the area - design, programming and implementation of DBMSs. The course develops into four natural modules: Data Modeling, Query Processing, Transaction Processing and Information Integration. We will begin with a historical perspective on database systems and identify the important components of database systems. We will then cover the ER approach to database modeling, the relational model, along with issues of normalization and redundancy elimination. Module 2 will cover three query languages - Relational Algebra, Datalog and SQL. We will also address implementation aspects such as query processing and physical database tuning for placement of indices. Module 3 covers several aspects of transaction management in detail. Specifically, we will deal with concurrency control, recovery, locking mechanisms, serializability and associated database elements. The final module will cover futuristic aspects such as information mediation, data warehousing, databases in a programming environment, web-enabled databases, visual image cartridges, and data mining. There will also be a cursory coverage of active and rule-based elements in a database system.

How is CS5614 different from CS4604?

Though both courses assume no prior background in database systems, CS 5614 will closely match CS4604 for only about a month (in total). The rest of the topics will provide an insider's view into DBs and help graduate students gain an appreciation for how and why things work, rather than merely knowing that certain features exist. In addition, it is expected to be the latest nightmare for graduate students who will spend 1-2 sleepless nights per week. Students are, moreover, cautioned that CS5614 is known to release intellectual stimulants often resulting in permanent addiction.

Who? Where?

Instructor: Naren Ramakrishnan (naren@cs.vt.edu, 231-8451, 629 McBryde). Teaching Assistant: Wei Yu (weyu@vt.edu). The class meets MWF 12:00-12:50 in RAND 206A. The office hours for the instructor are

MW 2-4pm and will be held in 629 McBryde Hall. The office hours for the TA will be announced in a day or so. If you are unable to make these hours and need to meet with us, you can setup an alternative time via email. If you need adaptations or accommodations because of a disability (learning disability, attention deficit disorder, psychological, physical, etc.), if you have emergency medical information to share with the instructor, or if you need special arrangements in case the building must be evacuated, please meet with the instructor ASAP.

Textbooks: Required and Recommended

There is no required textbook (only three recommended books). While this might seem like an inconvenience, the instructor intends to make course notes (and web material) available ahead of time that will cover all topics covered in class. The decision to not require a textbook was arrived at after careful discussions with other faculty members: (i) students typically come with diverse backgrounds into CS5614, (ii) different authors treat the same topic differently. For example, there is no standard treatment of E/R diagrams across textbooks, (iii) no textbook covers everything that we would like you to take away from this course.

As a result, we have narrowed down our selections to three recommended books (these will also be available in the Newman library reference desk). We recommend you to buy/have-access-to one of these depending on your individual needs. These are:

- (The heiro-book) Database System Implementation, by Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom, Prentice Hall, June 1999 (yes, brand new!) This book is ideal for students who have taken CS4604 at this institution and for whom CS5614 is a second course. The authors have a web page for this book at http://www-db.stanford.edu/~ullman/dbsi.html. For the rest of the students, who would like a book that covers (almost) everything, please use one of the following two.
- (The boat book) Database System Concepts, by Abraham Silberschatz, Henry F. Korth, and S. Sudarshan, McGraw-Hill, Third Edition, 1999. Notice the emphasis on the third edition (it is substantially revised).
- (The cow book) Database Management Systems, by Raghu Ramakrishnan, McGraw-Hill, 1998.

Lecture notes are posted on the class web page (http://courses.cs.vt.edu/~cs5614) on a 'module' basis. Since we will utilize the ORACLE DBMS (and possibly others), you may also wish to have access to one of the hundreds of ORACLE books/manuals (don't get frightened by the huge tomes on the software). The version of ORACLE that we will use in this course is Oracle8 Enterprise Edition 8.0.5.0.0. There are whole bookshelves devoted to such products in your local bookstore. A dynamic list of online web resources will also be linked in by the TA from the main 5614 web page. In addition, we will cover a few papers from the research literature for various topics.

Grading Policy and Assignments

There will be occasional homeworks (10 or so), a midterm and a final (both of these are closed book and closed notes). The homeworks will blur the distinction between traditional 'theory-type' assignments and programming tasks. They will contain a mix of both types and many individual questions will be mini-projects in their own right. The final grade is calculated as:

- 50% Homeworks
- 20% Midterm
- 30% Final

All the assignments (homeworks, midterm, final) are designed by the instructor. In addition, the instructor grades both the exams individually. The homeworks are graded by the TA. If you have an exam, homework or project that you feel has been graded incorrectly, you may request that it be regraded. Regrade requests

must be turned in to the instructor within one week of the date you received the graded assignment back. A regrade will result in the invalidation of the original grade and a complete re-evaluation of the entire assignment. This may increase or decrease your original score. The regraded score is final.

Keepin' in Touch

We will maintain a course listserv (cs5614@listserv.vt.edu) that students can use to exchange info. and voice questions about assignments etc. Since the listserv is created automatically by a central university system, any student registered in CS 5614 will be added to the mailing list. If you do not receive a test mail from the instructor by the end of the first week of classes, ensure that your email information is stored in your student record files.

Syllabus

Here's a 'module-wise' listing of the schedule (the relevant chapters from the recommended books will be highlighted as we go along; of course, this is subject to slight modifications and changes as the course gets underway):

- *Module 1*: Historical Trends, Conceptual Design, E/R Modeling, Relational Model, Normalization, 3NF, BCNF, 4NF (Properties and Guarantees), Functional Dependencies and Decompositions.
- *Module 2*: Query languages, Relational Algebra, Datalog, SQL, Logic and Databases, Physical Database Tuning, Query Compilation and Plan Selection.
- *Module 3*: Concurrency Control, Serializable Schedules, Conflicts, Locking Mechanisms, Properties of locking mechanisms, Recovery, Logging, Resolving Deadlocks, OLTP, Active and Rule-Based Elements.
- *Module 4*: Information Integration, Mediator-Based Approaches, Wrappers, Data Warehousing (including the CUBE operator), Online Analytical Processing (OLAP), Data Mining, DBMSs for Non-Traditional forms of data (images etc.)

Electronic Accounts

You are expected to have an account on the graduate lab network in McBryde 659. This will give you access to ORACLE and the other software systems that we will use in this course. Moreover, we assume familiarity with high level languages such as C/C++/Perl etc. and with both Windows and UNIX operating systems.

Workload

The course moves at an extremely fast pace! The idea is not to overburden you with needless homeworks but to give you a real appreciation for databases at the end of the semester. Please plan your course schedule accordingly.

Other Issues

- Enter your email address in the sheet that is passed around class on the first day. If you miss the first day of classes, send your email address to the TA.
- Ensure that you have an account on the graduate lab network that gives you access to ORACLE (Individual ORACLE accounts will be distributed in class later). It is your responsibility to get this matter settled by the end of the first week of classes.
- Some graduate students have printing quota restrictions on the departmental facilities. Please let the instructor know in advance if you would like readymade copies of lecture notes and reading material.