CS 4604: Introduction to Database Management Systems

From the Undergraduate Catalog: Emphasis on introduction of the basic database models, corresponding logical and physical data structures, comparison of models, logical data design, and database usage. Terminology, historical evolution, relationships, implementation, database personnel, future trends, applications, performance considerations, data integrity. Pre: CS 2604 and senior standing with a major or minor in CS. (3H, 3C). I, II.

Instructor's Perspective: The purpose of CS 4604 is to offer the student an introduction to the design and programming of database systems. In particular, we will cover OO (object-oriented) and ER (entity-relationship) approaches to data modeling, the relational model of database management systems (DBMSs), and the use of query languages such as SQL. We will also cover relational algebra, Datalog, and the use of SQL in a programming environment. We will touch upon query processing in a skeletal manner and the role of transaction management will be briefly mentioned. Issues clearly outside the scope of this course are how to implement a DBMS, active database elements, and object oriented DBMSs. We will also devote a class or two to futuristic prospects such as data warehousing and data mining. This is intended to be a first course in database systems for advanced undergraduates in computer science.

Meeting Times	MWF 9:05-9:55am, Pamplin 32
Instructor	Naren Ramakrishnan, 1-8451, McBryde 629
	naren@cs.vt.edu, http://www.cs.vt.edu/∼ramakris
Office Hours	Tuesdays, Thursdays 10am-12pm,
	or walk in any time.
Teaching Assistant	Feng Min, McBryde 1st floor
	fmin@vt.edu
Office Hours	Mondays, Wednesdays, 2-4pm.
Listserv	CS4604_91403@listserv.vt.edu
	(yes, the name is rather long winded.)
Course Web Page	http://courses.cs.vt.edu/∼cs4604

If you are unable to make the above times 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, or physical), 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.

Pre-requisites: CS 4604 is intended for senior CS majors and minors only. Force-adds will be very difficult for others. The CS 2604 pre-requisite will be rigorously enforced for all students. On the first day of classes, every student must fill out a pre-requisite form, and must return it to the professor in order to remain enrolled in this class. Remaining enrolled in this course without having the pre-requisite course is an honor code violation.

Approach: Focus is balanced between establishing a theoretical foundation and a pragmatic application in a real world DBMS environment. Particular emphasis is placed ont he logical design of relational database systems. Lectures will be reinforced by a significant, semester-long project in which students will apply concepts presented in class and obtain practical, hands-on experience. Because students are more likely to use rather than implement database systems, this introductory

course covers logical aspects (*externals*) of database design and implementation, not construction of database management systems (*internals*).

Required Textbook and References: The FCDB textbook is required – 'A Fist Course in Database Systems,' by Jeffrey D. Ullman and Jennifer Widom, Prentice Hall, Second Edition, 2002. **Make sure you get the second edition!** The authors have a webpage for the book at http://www-db.stanford.edu/~ullman/fcdb.html. Your local bookstore will have copies for purchase. Lecture notes to supplement this book are posted on the class webpage. Since we will utilize the ORACLE Enterprise Edition system, you may also wish to purchase one of the hundreds of ORACLE books/manuals (don't get frightened by the huge tomes!). Many online resources will also be maintained at the course webpage.

Evaluation: There will be 8-10 homeworks, a midterm, a final (both are closed book and closed notes), and a semester-long group project. Projects will be done in groups of 2-5 students (more details later) and will involve designing and programming a real-world database application. You will do some work on the project every now and then, beginning with modeling your database, designing it in relational terms, loading data into a real DBMS, and finally writing some cute SQL queries, embedding SQL queries into host language programs, and tinkering with a lot of fascinating features. The grades are broken down as: homeworks (20%), project (30%), midterm exam (20%), final (30%).

The homeworks expand on material taught in class, teach you some design and modeling, and will improve your vocabulary. Many of them will be in tune with an appropriate step of your database project. The project will help you to use the concepts learnt in realistic situations and encourage the use of programming skills in a realistic computing environment. The two exams will test a variety of skills.

All the assignments (homeworks, projects, midterm, final) are designed by the instructor. In addition, the instructor grades both the exams individually. The homeworks and projects are graded by the teaching assistant. If you have an exam, homework, or projet 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.

Workload: The course will appear deceptively simple but history shows that it is possible to get a variety of grades. The course also moves at a very fast pace and the workload is generally high! I am assuming that you will relish a design-oriented course and have the initiative to learn about various aspects of database design and programming. If you would like a career involving database technology (i.e., how to implement databases, not just use them effectively) we recommend, in addition, that you take a more advanced course in database system implementation or pursue a graduate program in this area (as mentioned before, we will not cover these aspects).

Keepin' in Touch: Please use the listserv actively for discussions and exchanging ideas. Since it is created automatically by a central university system, any student registered in CS 4604 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 address is properly recorded in the university system (the system only records @vt.edu addresses, not, for instance, @myfavoriteisp.com).

Syllabus: Here's a weekly-wise listing of the schedule (of course, this is subject to being pushed around):

- Week 1: Introduction to database systems, logistics, concrete examples (Chapter 1)
- Weeks 2 and 3: Data Modeling (Chapter 2, 4.1-4.3)
- Weeks 4, 5, 6, and 7: The Relational Model (Chapter 3, 4.4)
- Week 8: Review Classes, Introduction to ORACLE
- Week 9: Midterm Exam, More ORACLE
- Week 10: Introduction to SQL, Data Definition in Query Languages (Chapter 6.6, 7.1)
- Weeks 11, 12, and 13: Relational Algebra, Datalog, more on SQL (Chapter 5.1-5.3, 6.1-6.4, 10.1-10.2)
- Week 14: Misc. Topics, Systems Programming with SQL (Chapter 8, rest of Chapters 5, 6, 7, 10)
- Week 15: Cool Stuff Week (Data Warehousing, XML, Data Mining)
- Week 16: Schedule for Slippage

Electronic Accounts: You are expected to have an account in the undergraduate PC lab in McBryde 1st floor that will give you access to ORACLE and languages such as C/C++. If you require accounts, please contact the lab administrator(s) in rooms McBryde 116/118/124. Use your home computer at your own risk. We cannot guarantee that we will be able to grade work done at home (on other versions of the software, file transfer problems etc.).

Moreover, we assume familiarity with high-level languages such as C/C++ and scripting languages such as Perl/TCl. These languages will not be taught in this course but you will be required to use them for some of your assignments (particularly toward the end of the course) and your project. If you have never used these facilities before, we expect that you would either (i) obtain the necessary expertise outside the classroom (and the scope of CS 4604) or (ii) come and talk to the instructor about it (in which case, you will most probably be advised to drop this course). We strongly believe this background is essential to fully appreciate this course and that such a background would have been acquired by the end of the junior year.

Honor Code: The honor code is completely in effect. You are expected to do your own work. No one may give you the answers to assignments. You may discuss project concepts with colleagues but you may not discuss specific details nor do any joint work resulting in code or solutions. In other words, students are encouraged to communicate about general principles of the course, but all assigned work must be done on an individual basis. Assistance may be obtained from the instructor or GTA. You may not exchange any code or solutions, either in pieces or in entirety, by any electronic means or hard copy.

Of course, copying answers or code from anyone is considered cheating. If you suspect, or allow someone to obtain answers or code from you and that person is caught cheating, you may be held as an accomplice. Thus, if you suspect someone cheating, you should discuss this with the instructor;

anything you say will be held confidential. Students suspected of cheating will be sent directly to the higher authorities in the university; in turn, if they determine that the student was cheating, the student will receive a failing grade.

Things to do on the First Day:

- Sign the pre-req sheet and return it to the instructor. This is compulsory and IMPORTANT!
- Ensure that you have an account on the 116/118/124 lab network that gives you access to ORACLE. **ORACLE accounts will be distributed later in the semester; for now, just make sure that you have login accounts on the PC machines.** It is your responsibility to get this matter settled by the end of the first week of classes.

More Info. about CS 4604:

- CS 4604 is not easy. **There is typically no correlation between the number of years of experience with ORACLE/MySQL/substitute-your-favorite-db-here and your grade in CS 4604**.
- For most of the assignments, there is more than one correct solution. Solution sketches posted by the instructor cannot, hence, cover all possible answers. Moreover notice that many questions are designed to address sources of learning/design difficulties.
- Since you are students at or above the senior level, you are expected to demonstrate initiative in learning the various software systems. A one-semester course such as this cannot cover all nuances, nor should you as computer science students get bogged down by the details of different implementations. Finally, realize that this is a team effort (you plus us), so the better our interaction, the more effective the experience for both sides.
