Course Overview

Instructor: Sharma Chakravarthy
sharma@cse.uta.edu
The University of Texas at Arlington

Instructor/notes/project Information

- Instructor: Sharma Chakravarthy
- My course NOTES web site: https://wweb.uta.edu/faculty/sharma
- Blackboard (www.elearn.uta) will be used for assignments announcements, and project submissions as well as project discussions; No email submissions
- It is your responsibility to check for material (announcements, notes, home work, and quiz/exam details) added to the course web site or Blackboard!
- My Research web site: http://itlab.uta.edu/sharma
- My contact: Room: ERB 632  
  Email: sharma@cse.uta.edu, Phone: 817 272 2082

TA Information

- TA: Ms. Kanthi Komar
- Email: kanthisannappa.komar@mavs.uta.edu
- Instructor Office hours: ERB 632 
  Tu/Th: 11:15 am to 12:15 pm + by appointment
- TA office hours: M/W noon to 1:30pm + by appointment
- TA will deal with the project, home works
- Please send all correspondence to both

Other Information

- Each of you have to send me an email as follows: “For course CSE 5330 or CSE 3330, I will follow the UT Arlington honor code and all my submissions (projects and tests/exams) will conform to the UT Arlington standards for academic integrity (including UTA honor code)”
- Cheating, collusion, and plagiarism will be seriously dealt with (an automatic Fail grade)
- If you have difficulty, come see us but do not resort to the above
### Academic Honesty

#### What Constitutes Scholastic Dishonesty?

**Cheating**
- Copying another's test of assignment.
- Communication with another during an exam or assignment (i.e. written, oral or otherwise).
- Giving or seeking aid from another when not permitted by the instructor.
- Possessing or using unauthorized materials during the test.
- Buying, using, stealing, transporting, or soliciting a test, draft of a test, or answer key.

**Plagiarism**
- Using someone else's work in your assignment without appropriate acknowledgement.
- Making slight variations in the language and failing to give credit to the source.
- Copying materials from the Internet without citing the source.
- Using code/material from previous years without acknowledging the source.

**Collusion**
- Without authorization, collaborating with another when preparing an assignment or homework or other requirements of the course.

### Overview

- This is an introductory graduate (elective undergraduate) course on Database management systems at the CSE department at UTA.
- This course is about understanding a DBMS primarily from a users' perspective (as opposed to a system perspective).
  - Requirements analysis
  - Design and development of an application
  - Understanding the relational model
  - Understanding of DBMS architecture and functionality
  - Understanding SQL, creation and management of a database
  - Development of an application to interact with a DBMS
  - Overview of Storage management
  - Overview of query processing, and transaction management
- The emphasis of this course is on gathering application requirements, ER modeling, Relational model, SQL, NoSQL, and understanding the components of a DBMS to effectively use it for a variety of data intensive applications.
Organization of the course

- 3 modules
  - Need for a DBMS and Application modeling
  - Relational Concepts including Relational algebra and SQL
  - DBMS components: storage, Query processing and transaction management
  - Time permitting -- NoSQL DBMSs

- A semester long implementation project (Using Oracle on Omega and Java)

- 3 tests (in-class, closed book/notes)

- Home works are assigned (and graded if submitted) to help prepare for quizzes/exams

Project advise

- You will come up with an interesting problem in your domain/area of interest/expertise (see examples)
- This will be an application where we need to model the data aspect of the application (not the computation)
- Will develop an application with the DBMS as the backend
- Give enough thought to the overall project, make sure you understand all the steps, and have a management plan
- Should be able to ask queries, generate necessary reports from a biz perspective and management of data
- Need to use SQL (will be taught in the course) and as it is a different concept from a programming language, requires practice to get comfortable
- Please start on the project immediately (if we give 3 weeks, it means that it requires 3 weeks NOT 3 days)

Project details

- This is a semester-long project with five phases.
  - There is a deadline for each phase and each phase has a grade (except phase 1).
  - The grade for the project is a cumulative grade of the phases
  - As the phases are dependent on each other, you need to complete the previous phase in order to proceed to the next phase
  - Please make sure you submit each phase on time using bb; there are NO late submissions.
  - Take a look at http://infolab.stanford.edu/~ullman/fcdb.html#projects

  There are several examples. Use them to understand what you need to do, but do not copy them

  Phases are DUE as specified in the schedule by midnight on bb. Please submit it electronically in word or pdf or a scanned copy
**Topic suggestions**

- Organizing your music, video and others on your iPod
- Database for the athletics department
- Keeping track on inventory in an organization
- Library (or personal) book management
- Netflix video/users management
- International student organization requirements
- International students’ office requirements
- Inventory management for a grocery store
- Rental property management
- Management of accounts for the customers in a bank
- Management of games and statistics of players in a networked gaming environment
- Fraternity or sorority rush management
- Paper or web advertisement management
- Be creative, come with challenging and interesting problems

**Topic suggestions (Contd.)**

- Online Software rental and billing system
- Travel reservation system
- Hotel room rental and management system
- Management system for managing inventions
- System to associate a pharmacy and its inventory with various drug manufacturers.
- Auctioning system
- Management of the backend of social networks
- TA and I will make sure the problem is scoped properly for completion within the time available for this course
- Many students in the past have chosen problems of interest and continued it after the course
  - Modeling games
  - Modeling sorority rush
  - Modeling parent’s business
  - Problems that can be continued beyond this course

**Project details (contd.)**

- **Phase 2 [20%] EER diagram [Hard]**
  - Convert the problem statement to an EER diagram using the approach discussed in the class
  - Should contain at least 5 entities
  - Should contain 5 relationships (1:1, 1:n, and n:m)
  - Should have some attributes on relationships
  - Preferable to have at least one weak entity and relationships
  - Preferable to have set/subset relationships
- Due: as specified in the schedule on bb by midnight. Please submit electronically (only .doc or .pdf is accepted). You can draw and scan it and submit.
- Again, you may be asked to revise it and get approval before proceeding to the next step.

- **Phase 3 [10%] RDB Schema [Easy]**
  - Once the EER diagram is approved, revise it according to suggestions and convert it into relations.
  - Identify all attributes and their types
  - Identify candidate, primary and foreign keys and other constraints that are relevant to the application
- DUE: as specified in the schedule on bb midnight. No late submissions.
Project details (contd.)

- Phase 4 [30%] Create the database and complex queries [easy, but needs to be done methodically, takes time]
  - Revise the schema according to suggestions and convert it into relations.
  - Identify all attributes and their types
  - Identify candidate, primary and foreign keys and other constraints that are relevant to the application
  - Generate schema for the Oracle database
  - Use assertions, check, and triggers where appropriate

- Phase 4 (contd.)
  - Create the relations in the database
  - Populate the relations using your own data
  - Run ad hoc queries
    - Queries should contain joins, group by, having by, and order by clauses
    - Use triggers to monitor the state of the database
  - Support insert/delete and update statements
  - Finally, demo this phase as specified

Project details (contd.)

- Phase 5 [35%] User interface [Moderate, takes JDBC understanding, has coding, takes time]
  - Create an user interface for your applications (can get prepared for this while other stages are ongoing)
  - Run interface based parameterized queries.
  - Use dynamic SQL
  - Use triggers to monitor the state of the database

  - Finally, demo the project by the last day of classes

SQL Home work

- Writing SQL Queries [5%] [Moderate, takes coding, takes time]
  - You will be asked to write a number of challenging SQL queries (not more than 10) on a database we will setup on Oracle on Omega
  - You will get to exercise a number of features that you might not be able to exercise as part of phase 4.
DBMS to be used

- We will use Oracle on Omega with a JDBC connection to the interface.
- You can log into Omega from anywhere.
- NH 239 has a number of terminals from which can be used for this course (open 24 hours)
- Using ssh you can log into Omega
- Projects will be accepted only electronically.
- Projects should run on Omega for grading purposes

What is assumed

- Some Discrete mathematics (functions, mapping)
- Set theory
- Knowledge of Java or C++
- JDBC and basic use of Linux/Unix

How to do well on the course

- Attend all the lectures
- Do follow up reading before and immediately after the lecture (not 1 day before the exam)
- Come prepared and ask questions in the class
- Make the class interactive
- There are NO dumb or trivial questions; all questions are important
- Solve all homework problems yourself and submit it
- Make use of my (and TA’s) office hours

Beyond this course …

- If you get excited about databases and related areas, there are a number of courses you can take beyond this course.
- If you are interested in doing a thesis (MS/PhD) in the general areas of Databases, social network analysis, cloud computing, information integration, mining, complex event and stream processing, information security – stop by and talk to me.
Information Technology Laboratory (ERB 514)

Prof. Sharma Chakravarthy (ERB 632)

Email: sharma@cse.uta.edu, URL: http://itlab.uta.edu/sharma

Funding Sources: NSF, Spawar, AFRL, Rome Lab, ONR, DARPA, TI, MCC

Select Projects
- Data Analytics Using Multi-layered graphs
- Scalable Graph Mining using Map/reduce (under
  joint proposal by V. Athitsos)
- MavVStream: video content analysis by querying
- Expertise identification in Q/A community
- Ranking in web databases
- Mining: Graph, Text, Assoc Rules
- Prediction of Event Patterns
- Information Search, Filtering, and Classification

Select Publications
5. Y. Cai and S. Chakravarthy, “Improving Answer Quality Prediction in Q/A Social Networks by Leveraging Temporal Features”, in IJNGC, March 2013

CSE 5331
- DBMS models and implementation techniques
- This course goes into the details of the components of each module
- System-oriented as opposed to user-oriented
- Hands on projects implementing the internals of a dbms (e.g., buffer manager, disk sorting, concurrent transactions, …)

CSE 6339
- Offered in Fall 20yy
- Data Stream Management Systems (DSMS)
- Stream data processing as opposed to stored data processing
- Sensors, RFID, and other new applications
- Importance of QoS and (near) real-time processing
- Scheduling, load shedding issues
- Commercial and research prototypes
  - Coral8, asper, streamBase, MavEstream, snoop, aleri, …

CSE 5334
- DATA MINING
- Preparing data for mining using preprocessing, data warehouses, OLAP
- Data mining primitives, languages, and system architecture;
- Data mining techniques including association rule mining, classification/prediction, and cluster analysis.
CSE 6331 (and others)

- Advanced topics in Database systems
- The topics may vary from offering to offering based on the instructor.
- Deals with new/advanced topics that are currently being researched
- I offer graph mining, stream processing, and cloud computing in spring
- Topics such as web db & XML, DB and information exploration have been offered
- I have offered data warehousing, data mining, and event processing as part of this course

CSE 6399 – Seminar course

- Advanced DB topics
- Typically a seminar course
- Reading and analyzing papers in new areas of research
- This semester I am offering this course on: Complex event & stream processing and information integration