



# CS4420 Database System Implementation (Spring 2013)

---

| [General](#) | [Textbook](#) | [Administration](#) | [Description](#) | [Project](#) | [GT Calendar](#) | [Lecture Schedule](#) | [Readings](#) |

---

## General Information

**Instructor:** Professor [Ling Liu](#)

**Office:** KACB 3340, **Phone:** 5-1139, **Email:** [lingliu@cc.gatech.edu](mailto:lingliu@cc.gatech.edu)

**Lecture location :** CCB 16

**Lecture hours:** Tuesday and Thursday 12:05pm - 1:25pm (Jan 7 ~ April 26, 2013)

**Office hours:** Tuesdays and Thursdays 1:25pm - 2:15pm, or by appointment

### Course TAs:

Kisung Lee (kslee AT cc.gatech.edu), Office hour: TBA

**Prerequisite(s):** CS 4400 or equivalent

**All course materials will be accessible from T-Square, including** the newsgroup.

**Course Evaluation online** Click [Here](#).

---

## Course Materials

**TextBook:** Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom, *Database System Implementation*, Prentice Hall, 2000.

**Lecture Notes and Readings:** to be provided by the instructor on [T-Square](#) .

---

## Grading Policy

Reading Homework	10%	5-10 Reading-Critique Assignments
Quizzes	20%	Quiz 1 on Feb 28; Quiz 2 on April 9
Project (Group)	50%	Due midnight on April 26 (last day of Classes)
Final (Technology Review)	20%	Due midnight on the final exam day

**Re-examination:** No re-examination except as per regulations.

**On-line Information:** Course-related information, such as lecture notes, reading assignments, project descriptions, signup and presentation/demo schedule, will be accessible via CS4420-CS6422 on [T-Square](#). You are recommended to check the course schedule page on T-Square regularly.

**Policy on Collaboration:** You are allowed to discuss the project with each other. You should give credit to others for the ideas. But you should obey the [honor code](#) and DO NOT copy the code of others.

---

## Course Description:

In this course we will study four major topics relating to database system implementation. The emphasis is on the "systems" components of a database management system. To better understand these components, a database implementation project will be required where you will build some of the basic "system" components for a simple database management system. We start with a brief overview of the basic components of a database system and discuss a set of open issues in designing and implementing a database management system, including relational DBMS and NoSQL databasesystem before we detail the four core system components: Storage, Query Processing, Transaction Management and Distributed Data Management.

- **Storage management.** How data is stored (organized) on secondary storage plays an important role in processing database queries efficiently. We will examine the various file structure alternatives involving indexing and hashing.
- **Query processing and optimization.** In the context of relational databases, we are interested in two topics: transformations which are applied to a user query to make it execute more efficiently and algorithms which implement various relational algebra operators efficiently. Both of these topics fall within the realm of the query optimizer. We will also touch query processing issues in column stores and RDF stores.
- **Concurrency control and recovery.** We discuss protocols and algorithms to ensure multiple transactions to execute on a shared database concurrently and still see a consistent view of the data, as well as leave the database in a consistent state. We examine several concurrency control schemes and their tradeoffs. We will also discuss the recovery manager of a database system. The main concern is how the database system recovers from a failure, e.g., a transaction failure, a system crash, etc. We examine the advantages and disadvantages of several recovery schemes.
- **Distributed and parallel processing of data.** The fourth topic involves big data technologies, including big data storage, big data processing and big data analytics. We will introduce some NoSQL database systems, including column stores, RDF stores, Hbase and so forth. Hadoop MapReduce Algorithms for expensive queries over big data will be discussed as well. If time permits, we will discuss the various issues in database performance tuning and how parallel relational database systems can be used to improve the performance of query and transaction processing.

Topic	Chapter
Introduction to DBMS Implementation	1
Relational DB Review	2
Data Storage	3

Representing Data Elements	4
Index Structures	5
Query Execution	6
Query Compiler	7
Coping with System Failures	8
Concurrency Control	9
Transaction Management	10
Information Integration	11

## Project

The course project is a group project. Each group should consist of 3-4 students. You will choose one of the proposed projects listed in the project description document made available on T-Square under cs4420-cs6422.

You need to register your project team on TSquare in the 4<sup>th</sup> week (midnight on Thursday of Feb 1), including the following information about your group:

- Project team name
- the list of your group members (name, email)
- the group contact person and his/her email

Project Resources will be made available on T-Square. Here are some standard resources:

[C Tutorial 1](#) and [C Tutorial 2](#)  
[C++ Tutorial](#)  
[Java Tutorials](#)

## Project Schedule

Phases	Due Date	Location
Phase I Team Formation	1/24 (week 3)	TSquare (assignment)
Phase II Design Document	2/8 (week 5)	TSquare (assignment)
Phase III In Class Presentation (PPT)	4/16~4/25 (week 15-16)	TSquare (assignment)
Phase IV Project Demo	4/23~4/26 (week 16)	Instructor's office

## Grading

Phase II (design doc): 20% of the project grade

Phase III (ppt): 20% of the project grade

Phase IV (final report + demo + code package): 60% of the project grade



*Last updated on Dec. 28, 2011 by Ling Liu ([lingliu@cc.gatech.edu](mailto:lingliu@cc.gatech.edu))*