CSC 468: Database Management System Implementation
Spring 2016
Course Syllabus

March 27, 2016

Instructor: Alexander Dekhtyar
email: dekhtyar@csc.calpoly.edu
office: 14-215

What	When		Where
Lecture	MW	4:10 – 5:30am	14-256
Lab	MW	5:40 – 7:00pm	14-256
Final Exam	Friday, June 10	4:10 - 7:00pm	14-256

Office Hours

	When	Where
Monday	2:10pm - 3:00pm	14-215
Wednsday	2:10am - 3:00pm	14-215
Friday	8:30am - 10:00pm	14-215

Additional appointments can be scheduled by emailing the instructor at *dekhtyar@calpoly.edu*.

Description

The goal of the course is to give students hands-on knowledge of the organization and inner workings of modern database management systems. The course concentrates on issues of physical and logical organization of data in a DBMS, query processing, transaction management and concurrency control and crash recovery.

Learning Objectives

After taking the course the students are expected to

- understand how information is stored in DBMS
- be able to implement information indexing techniques
- understand query processing in DBMS, be able to solve query optimization problems
- be able to use their knowledge of concurrency control protocols to solve transaction scheduling problems
- be able to simulate the work of ARIES algorithm for crash recovery in relational databases.

Texbook

Database Systems. The Complete Book, 2nd Ed. H. Garcia-Molina, J.D. Ullman, J. Widom., 2009, Prentice Hall.

or

Database System Implementation, H. Garcia-Molina, J.D. Ullman, J. Widom, 2000, Prentice Hall.

Note: First edition of "The Complete Book" (Prentice Hall, 2002) is also fine.

Topics

No.	Topic	Duration (weeks)	Material
1.	Storage of information in DBMS	2	Chapters 13,14
2.	Query Processing	3	Chapters 15, 16
MIDTERM			
3.	Transaction Management	1	Chapter 18
4.	Concurrency Control	2	Chapter 18, 19
5.	Crash Recovery	1	Chapter 17
FINAL (conprehensive)			

Grading

Labs and Homeworks	10-15%
Project	30-40%
Midterm Exam	15 - 20%
Final Exam	25 - 35%

Course Policies

Prerequisites

The prerequisite for **CSC 468** is **CSC/CPE 365** or an *eqivalent* undergraduate course in databases from another school.

In this course, I assume that you are familiar with the main concepts studied in an undergraduate database course. In particular, I assume familiarity with ***Relational Database model***, ***relational algebra*** and **SQL**.

If you lack the prerequisite knowledge, this course **will not** help you gain it.

Course Structure

Because the field of databases and data management is experiencing rapid change at the moment, this version of the course sets forth two goals:

- **Learning the principles of organization of relational DBMS.** This is the original goal/purpose of the course. RDBMS is the most popular type of database management systems out there, learning how to build them both allows you to understand an incredibly important technology, as well as understand how to live without them.

- **Learning about the organization of modern non-relational DBMS.** Under normal circumstances we'd teach a whole new class on this topic. However, two things prevent this: (a) the amount of paperwork required to create a separate course, and (b) the need to know how RDBMS work (goal 1 of this course) in order to really appreciate the organization of modern non-relational DBMS.

In order to meet the two goals, we will adopt creative use of our meeting space and time. Recall that we meet for three academic hours twice a week for a total of six contact hours a week. Three of those are designated as "lecture" and three - as "lab. We will distribute course content into a number of activities as follows.

Study of organization of RDBMS. This is the traditional content of the course, and it will form the core of our lecture periods. Some of the lecture material has been updated to make it DBMS-general (not RDBMS-specific), however, as we proceed through the course, a number of RDBMS-specific topics (query processing and optimization, transaction management, crash recovery) will be covered.

Study of non-relational RDBMS. We will spend part of our lab periods on the study of non-relational DBMS technologies. At the beginning of the course, we will need to learn some specific material in order to prepare for the course project. Once project-specific material related to non-relational DBMS is covered, we will perform a number of case studies in which we will discuss different non-relational DBMS and DBMS technologies trying to pay specific attention on the similarities and differences between them and RDBMS. The case studies will be student-driven, i.e., each case study will be prepared and delivered by one student team.

Project. The course will have a quarter-long team project devoted to an implementation of a simple (but real) non-relational DBMS. The project will be organized in a number of stages, each stage will be graded and will receive a score that counts towards the total project score, although opportunities to fix problems with earlier deliverables will be given. Opportunities for team meetings during lab periods will be given (especially during the second half of the course), but otherwise, project-related assignments are designated as take-home tasks.

Project background. Another part of lab periods, especially at the beginning of the course, will be spent discussing the relevant background for the course project. This quarter, we will work on a joint project with CPE 458: Distributed Systems course. The tentative goal of the overarching project is to develop a distributed peer-to-peer file sharing system (think BitTorrent). The role of our course is in developing a light-weight special-purpose DBMS that can be deployed with each peer-to-peer system client to store and manage the data related to the file exchange activities in the system. The DBMS we develop will not necessarily be distributed (that is, a single peer-to-peer client will work with a single DBMS server installed on a single machine), but the collection of the DBMS servers viewed holistically will work as a distributed DBMS.

Labs. Labs will be used to facilitate the course project and to test some of the knowledge obtained in lectures. In the first half of the course labs will be devoted to project setup and preparation issues: study of the packages, technologies, APIs and query languages needed for the project. Prior to the direct use of this information in the project, a lab will be offered to get you familiarized with the necessary concepts. In the later parts of the course, labs will target some of the theoretical and practical material learned in the lectures (such as query optimization and transaction processing). Most of the labs will be team-based (to be performed by the project teams).

For the most part, labs are treated as take-home assignments. Some time during the lab periods for team meetings to discuss the assignments will be given, but do not expect to complete any labs in class.

Homeworks. Paper-and-pencil homeworks will be assigned to prepare you for the exams. The main purpose of the homeworks is preparation for the exams. By tradition, homeworks are not graded, but credit is given for submitting a completed homework.

Late Submissions

Typically, paper-and-pencil homeworks are due in class, labs are due at the end of the lab period on the due date and project stages are due at midnight of the due date. Exact assignment/project submission instructions will be included with the assignments. Be sure to follow exactly the submission procedures. Homework/projects submitted later than indicated will be considered *late submissions.*

Late paper-and-pencil ***homework*** *submissions* will not be accepted.

Late ***lab*** *and* ***project*** *submissions* are strongly discouraged. A penalty of 10 - 30% will be assessed for any submissions that are late by less than 24 hours. No credit will be given for any later submissions. You are encouraged to submit your code on time even if it is not perfect. You can then debug your code and submit a fixed version late, subject to the abovementioned rules. When more than one submission is present, we will independently grade two submissions: (i) the latest on-time submission and (ii) the latest late submission for which non-zero credit can be assessed. Your grade for the project will be the **maximum** of the two grades.

Communication

The class will have an official mailing list. The email address for the mailing list is *cpe-468-05-2164@calpoly.edu*. All students enrolled in the class are automatically subscribed to the mailing list (using the email addresss that the CS department has on file). You are allowed to tranfer the subscription to another account (e.g., CSLAB account). However, you must stay subscribed to the mailing list for the entire duration of the class.

I encourage questions during classtime and questions via email. My answers to email questions may be broadcast to the entire class via the mailing list, if the answer may be relevant to everyone (e.g. a correction in a text of a handout, or a clarification of a homework problem), and may also appear on the web page. The questions can also be posted to the mailing list directly. The mailing list will also be used for all annoucements related to the course. It is your responsibility to read your class-related email. Failure to read email posted to the mailing list cannot be used as an excuse in the class.

Web Page

Class web page can be found at

http://www.csc.calpoly.edu/∼dekhtyar/468-Spring2016

Through this page you will be able to access all class handouts including homeworks, lab assignments, project information, lab/project data and lecture notes.

Links to additional information, and notes and announcements will also be posted.

Strike

As you might know, Cal Poly faculty, as well as the faculty at other CSU campuses are represented for the purposes of collective bargaining by California Faculty Association (CFA), an affilliate union of SEIU, American Association of University Professors (AAUP), National Education Association (NEA) and California Teachers Association (CTA).

During the 2015-2016 academic year CFA has engaged in collective bargaining activities with the CSU adminstration on behalf of the CSU faculty. The CFA has asked for a 5% salary increase for all CSU faculty to help if not offset, then at least somewhat minimize the financial hardships faced by the CSU faculty over the past eight years of no stagnant salaries (complete with a 10% furlough during one of these years).

At present the CFA is at an impasse with the CSU administration regarding the salary negotiations. While this impasse can be resolved at any time, CFA has used its right to call for a concerted action. **As things stand now, CFA has called for a CSU-wide strike to take place during the following days:**

- April 13 — April 15
- April 18 — April 19

We have two classes scheduled during the announced strike days: on April 13 (Wednesday) and April 18 (Monday).

If CFA and the CSU administration reach an agreement on faculty salaries, the strike will be called off and the classes on those days will take place as usual.

If **the agreement is not reached** and **the strike is on**, you should expect the following to happen:

- **No classes** on April 13 and April 18.
- Because the faculty who go on strike must stop **all work-related activities**, if I join the strike I will not respond to any work-related emails, and in fact, I will be unable to check my calpoly.edu email account. (Please note that April 16 and April 17 are **not** strike days, so during these two days, I may be able to check email and respond to it).
- Faculty who go on strike will not be present on campus, except for picket lines.
- **You are not on strike** in your capacity as Cal Poly students. What this means is that while I may not be able to hold the classes, there may be assignments that will have to be done by you during the strike days. I promise **no due dates** for course assignments on any of the strike days, but if the strike takes place, one or more assignments may be issued before the strike commences with due dates after the strike is over. The expectation is that you will spend some of your time during the strike days

working on these assignments, even if I am unable to answer the questions you might have about them. (We will try to make sure you have sufficient time to resolve any questions related to these assignments prior to the start of the strike).

- Classes will resume as usual on Wednesday, April 20.

I understand how disruptive a week-long strike will be on your education and I apologize in advance for the negative impact on your education the strike might have. At the same time, as far as preventing the strike, the ball is squarely in the court of the CSU administration. If you want the CSU administration to know your thoughts concerning the faculty strike, I would like to encourage you to convey those thoughts - either directly, or through Cal Poly administration.

For more information concerning the strike, please see

http://www.calfac.org/item/faq-cfa-bargaining-possible-strike

and

http://www.calfac.org/strike-411-students

Academic Integrity

University Policies

Cal Poly's Academic Integrity policies are found at

http://www.academicprograms.calpoly.edu/academicpolicies/Cheating.htm

In particular, these policies define *cheating* as (684.1)

> "... *obtaining or attempting to obtain, or aiding another to obtain credit for work, or any improvement in evaluation of performance, by any dishonest or deceptive means. Cheating includes, but is not limited to: lying; copying from another's test or examination; discussion of answers or questions on an examination or test, unless such discussion is specifically authorized by the instructor; taking or receiving copies of an exam without the permission of the instructor; using or displaying notes, "cheat sheets," or other information devices inappropriate to the prescribed test conditions; allowing someone other than the officially enrolled student to represent same.*"

Plagiarism, per University policies is defined as (684.3)

> "... the act of using the ideas or work of another person or persons as if they were one's own without giving proper credit to the source. Such an act is not plagiarism if it is ascertained that the ideas were arrived through independent reasoning or logic or where the thought or idea is common knowledge. Acknowledgement of an original author or source must be made through appropriate references; i.e., quotation marks, footnotes, or commentary."

University policies state (684.2): "Cheating requires an "F" course grade and further attendance in the course is prohibited." (appeal process is also outlined, see the web site above for details.). Plagiarism, per university policies (684.4) can be treated as a form of cheating, although a level of discretion is given to the instructor, allowing the instuctor to determine the causes of plagiarism and effect other means of remedy. It is the obligation of the instructor to inform the student that a penalty is being assessed in such cases.

Course Policies

All homeworks are to be completed by each student **individually**. Lab assignments are to be completed by the appropriate units (individual, pair, group), and no code/solution-sharing between units is permitted. Students are encouraged to discuss class content among themselves but NOT in a manner that constitutes plagiarism and cheating as defined above (e.g., you can solve together a problem from the textbook that had not been assigned in the homework, but you should solve assigned problems individually).