CSC 560: Special Topics in Databases - Modern DBMS Architectures
Fall 2012
Course Syllabus

September 17, 2012

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

What	When		Where
Lecture	TR	9:10 – 11:00pm	14-232B (Frank Pilling)
Final Exam	Dec 6 (Thursday)	10:10 - 1:00pm	14-232B

Office Hours

	When	Where
Tuesday	11:10am - 12:00pm	14-215
Wednsday	9:10am - 12:pm	14-215
Friday	9:10 - 10:00pm	14-215

Description

The past decade has given us more new DBMS architectures than the previous thirty years combined. Relational DBMS are still dominant on the market, but new emerging DBMS architectures prove to be more appropriate for specific applications. In this course we will study the current DBMS architectures used actively in modern software applications: both relational, and the newly emerging, so called "NoSQL" DBMS. We will discuss their internal organization, the trade-offs posed by each architecture and the applications for which each DBMS architecture is appropriate.

Textbook, Readings

The class does not have an official textbook. The first part of the course is a rehash of selected topics from CSC 468 (which, unfortunately will not be taught for a while, so this is our chance), and is taught from the CSC 365/366/468 textbook:

> *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.

In preparation for the remainder of the course I used the following books:

> *NoSQL Distilled: a brief guide to the emerging world of polyglot persistence*, P.J. Sadalage, M. Fowler, 2012, Pearson Education (Addison Wesley). ISBN: 978-0-321-82662-6.
>
> *Seven Databases in Seven Weeks: a guide to modern databases in the NoSQL movement*, E. Redmond, J.R. Wilson, 2012, Pragmatic Programmers. ISBN: 978-1-93435-692-0

The remainder of topics will be covered via a combination of academic papers, white papers, outside speakers and sheer luck.

Topics

The approximate list of topics is:

No.	Topic	Duration (weeks)
1.	ACID properties of DBMS and Relational DBMS	2-3
2.	Distributed Data Management (MapReduce)	1
3.	NoSQL DBMS: Overview	1
4.	NoSQL DBMS: Key-Value DBMS	1
5.	NoSQL DBMS: Document DBMS	1
6.	NoSQL DBMS: Colum-Family DBMS	1-2
7.	NoSQL DBMS: Graph DBMS	1-2

The order of topics is not fixed and will depend on a number of circumstances.

Outside speakers

A significant portion of this course will, most likely, be outsourced to outside speakers. We are expected to have anywhere between three and six different speakers presenting to the class. Some speakers will talk about specific data management applications and their approaches to success. Some speakers will present specific DBMS architectures and discuss their uses. The speakers will come both from academia and industry.

Some of the speakers will present you with a problem/challenge to be addressed as a course assignment. We plan to have at least two such assignments.

Grading

Research Project	30%
Team Assignments	40%
Literature Survey	10-20%
Final Presentation	10-15%
Class Participation	5%

Course Policies

Exams

We have a scheduled time to meet for the final exam (Thursday, Dec 6, 10:10am - 1:00pm). There will be no written exam for the course. However, we will use this time for team presentations on a selected topic (this is listed as *"Final Presentation"* above). The presentation topic will either be the main course research project, or one of the team assignments. The final presentation will be scored separately from the other deliverables for the appropriate assignment (which will be written reports and/or software deliverables/demos).

Project

The course will have two types of team-based activities: the course project and team assignments. The course project is a research project that each student team selects for itself and works on throughout the quarter. The subject matter of the project is limited only be the following requirements:

- the project must have a research component, i.e., the key goal of the project must be discovery of new knowledge and/or replication of some existing work (which is a form of new knowledge);
- the project must have data management as a key component.

Specifics for the research assignment will be handed out in the first two-three weeks of the course. The project may have multiple deliverables (proposal, final report, system demo, peer evaluations). Additionally, the project may be selected for the final presentation (the *Final Presentation* deliverable with a separate score).

Team Assignments

We are planning to offer a number (up to two, at this point) assignments brought to you by our invited outside speakers. Each assignment is to be executed over a period of a few (2-4) weeks. The assignments are to be completed by teams (usually, the same teams on the course project). One assignment is intended to give you a chance to use some of the material discussed in class to design a data management solution/architecure for a real problem. The other assignment is designed to let you practice software development skills with a popular open-source NoSQL DBMS.

Each assignment will come with a specific set of deliverables (a written report, a software demo, etc). Additionally, it is possible that one of the assignments will be selected for the final course presentation.

Literature Surveys

This is the only major compenent of the course that comes as an individual assignment. Each student will propose a database-related topic for literature survey, and, upon getting it approved, will select a number of papers on this topic, and prepare an academic literature survey. Because the course is devoted to essentially cutting-edge database research, your literature survey will have to include papers that have been published in major database conferences in the past five years (i.e., we are interested in problems that the database research community is working on now, rather than the ones it was addressing 10-20 years ago).

Late Submissions

There is very little room in the course for late submissions. Most written deliverables will have deadlines that are very hard to violate (e.g., last class of the quarter). Most in-person deliverables (demos, presentations) will have to happen on specific days or else.

Please don't make me come up with complicated rules for what happens to late submissions by not submitting anything late.

Web Page

Class web page can be found at

http://www.csc.calpoly.edu/∼dekhtyar/560-Fall2012

Through this page you will be able to access all class handouts including homeworks, project information, reading materials and lecture notes (should the latter be written).

Mailing List

The course has a mailing list:

csc-560-01-2128@calpoly.edu

I tend to use mailing lists actively during the quarter (especially in upper-division/graduate courses), so, please, make sure, you are monitoring the email subscribed to this mailing list.

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

First, all traditional warnings concerning cheating apply in this course. In particular, solicitation of help from people not involved in the course and submission of materials/code etc.. not developed by you are absolutely prohibited. Any outside materials used in preparation of homeworks, reports, project assignments must be properly documents. For example, you must properly cite all papers you refer to, all web resources used in perparation. You must also note any open source, off-the-shelf, etc... software or code fragments that you have incorporated in your solution. If you have questions concerning allowable use of such materials, please consult me **in advance**.

For example, if an assignment is to design and implement an XML parser, you are supposed to build one from scratch and not use any available parser code (which is plentiful). On the other hand, if you want to use an open-source library, or some code developed by one of the team members prior to the course as part of a project solution, this may qualify as allowable use, if the code is used in support of the main tasks of the project.