

CMPSC 380
Principles of Database Systems
Fall 2014

Syllabus

Course Instructor

Dr. Gregory M. Kapfhammer
Office Location: Alden Hall 108
Office Phone: +1 814-332-2880
Email: gkapfham@allegheny.edu
Twitter: @GregKapfhammer
Web Site: <http://www.cs.allegheny.edu/sites/gkapfham/>

Instructor's Office Hours

- Monday: 2:30 pm – 4:30 pm (15 minute time slots)
- Tuesday: 2:30 pm – 4:30 pm (15 minute time slots)
- Wednesday: 10:00 am – 12:00 noon (15 minute time slots) *and*
4:30 pm - 6:00 pm (15 minute time slots)
- Thursday: 10:30 am - 11:00 am (10 minute time slots)
- Friday: 11:00 am – 12:00 noon (10 minute time slots) *and*
3:30 pm – 4:30 pm (10 minute time slots)

To schedule a meeting with me during my office hours, please visit my Web site and click the “Schedule” link in the top right-hand corner. Now, you can browse my office hours or schedule an appointment by clicking the correct link and then reserving an open time slot.

Course Meeting Schedule

Lecture, Discussion, Presentations, and Group Work: Tuesday and Thursday, 11:00 am – 12:15 pm
Laboratory Session: Wednesday, 2:30 pm – 4:20 pm
Final Examination: Friday, December 12, 2014 at 9:00 am

Course Catalogue Description

A study of the design and implementation issues in database management systems. Topics include data models, logical/physical database design, data access/search techniques, normalization theory, mappings from logical to physical structures, storage, and utilization. Additional topics include database reorganization, migration, database integrity, consistency, privacy and security, distributed database systems, architecture of knowledge-based systems, and intelligent query interfaces. One laboratory per week. Prerequisite: Computer Science 112. Offered in alternate years.

Course Objectives

The essence of the discipline of computer science is algorithms; this course will introduce students to the principles of data management using algorithms. We will investigate some of the key techniques that scientists use to manage data. Areas of discussion include, but are not limited to, relational databases and query languages, object-oriented data storage, encoding data in the eXtensible Markup Language (XML), low-level data storage, transactions and concurrency control, data warehousing and mining, and the implementation and testing of database applications.

The course will introduce students to the theory and practice of data management while covering both the well-established and the cutting-edge areas of the discipline. The course also invites students to assess the correctness of their implementations and conduct both analytical and empirical evaluations of the performance of data management techniques. Moreover, the course will ask students to implement small- and medium-scale data management systems and to install and use a wide variety of support tools. In addition to improving their teamwork skills, students will enhance their ability to write and speak about software in a clear and concise fashion.

Performance Objectives

At the completion of this class, a student must be comfortable with fundamental data management topics and be aware of current research in the area. When given a new data management problem, students should be able to select proper data management tools and implement a complete application that uses them to solve the stated problem. Students also must develop a toolkit of data management concepts that they can use in the context of the solutions to real-world problems. Finally, students must develop and apply a strong knowledge of analytical and empirical techniques that they can use to characterize and predict the performance of data management systems.

Students should also be able to handle many of the important, yet accidental, aspects of implementing programs with modern programming languages and data management systems. In addition to being comfortable with program editors, compilers, debuggers, testing tools, virtual machines, database management systems, and query languages, students should also be able to understand the purpose of shell environment variables such as the `PATH` and the `CLASSPATH`.

Required Textbooks

Database Systems Concepts. Avi Silberschatz, Henry F. Korth, and S. Sudarshan. Sixth Edition, ISBN: 0-07-352332-1, 2011.

(References to the textbook are abbreviated as “DSC” in the syllabus and on the Web site).

Students who want to improve their technical writing skills may consult the following books.

BUGS in Writing: A Guide to Debugging Your Prose. Lyn Dupré. Second Edition, ISBN-10: 020137921X, ISBN-13: 978-0201379211, 704 pages, 1998.

Writing for Computer Science. Justin Zobel. Second Edition, ISBN-10: 1852338024, ISBN-13: 978-1852338022, 270 pages, 2004.

Along with reading the required books, you will be asked to study many additional articles from a wide variety of conference proceedings, journals, and the popular press.

Class Policies

Grading

The grade that a student receives in this class will be based on the following categories. All percentages are approximate and, if the need to do so presents itself, it is possible for the assigned percentages to change during the academic semester.

Class Participation and Instructor Meetings	5%
First Examination	15%
Second Examination	15%
Final Examination	20%
Laboratory and Homework Assignments	30%
Final Project	15%

These grading categories have the following definitions:

- *Class Participation and Instructor Meetings:* All students are required to actively participate during all of the class sessions. Your participation will take forms such as answering questions about the required reading assignments, asking constructive questions of the other members of the class, giving presentations, and leading a discussion session. Moreover, all students are required to meet with the course instructor during office hours for a total of thirty minutes during the Fall 2014 semester. These meetings must be scheduled through the course instructor's reservation system and documented on a meeting record that you submit on the day of the final examination. A student will receive an interim and final grade for this category.
- *First and Second Examinations:* The first and second interim examinations will cover all of the material in their associated module(s). While the second examination is not cumulative, it will assume that a student has a basic understanding of the material that was the focus of the first examination. The date for the first and second examinations will be announced at least one week in advance of the scheduled date. Unless prior arrangements are made with the course instructor, all students will be expected to take these examinations on the scheduled date and complete the tests in the stated period of time.
- *Final Examination:* The final examination is a three-hour cumulative test. By enrolling in this course, students agree that, unless there are extenuating circumstances, they will take the final examination at the time stated on the first page of the syllabus.
- *Laboratory and Homework Assignments:* These assignments invite students to explore the concepts, tools, and techniques associated with the management of data. All of the laboratory assignments require the use of the provided tools to design, implement, and evaluate systems that solve data management problems. To ensure that students are ready to develop software in both other classes at Allegheny College and after graduation, the instructor will assign individuals to teams for some of the laboratory assignments. Unless specified otherwise, each laboratory assignment will be due at the beginning of the next laboratory session. Many of the laboratory assignments in this course will expect students to give both a presentation and a demonstration of the software that they created to manage a collection of data.

- *Final Project:* This project will present you with the description of a problem and ask you to implement a full-featured solution using one or more programming languages and a wide variety of data management techniques. The final project in this class will require you to apply all of the knowledge and skills that you have accumulated during the course of the semester to solve a problem and, whenever possible, make your solution publicly available as a free and open-source tool. The project will invite you to draw upon both your problem solving skills and your knowledge of programming languages and data management systems. The final project will be completed in groups assigned by the course instructor.

Assignment Submission

All assignments will have a stated due date. The printed version of the assignment is to be turned in at the beginning of the class on that due date; the printed materials must be dated and signed with the Honor Code pledge of the student(s) completing the work. Late assignments will be accepted for up to one week past the assigned due date with a 15% penalty. All late assignments must be submitted at the beginning of the session that is scheduled one week after the due date. Unless special arrangements are made with the course instructor, no assignments will be accepted after the late deadline. For any assignment completed in a group, students must also turn in a one-page document that describes each group member's contribution to the submitted deliverables.

Attendance

It is mandatory for all students to attend the class and laboratory sessions. If you will not be able to attend a session, then please see the course instructor at least one week in advance to describe your situation. Students who miss more than five unexcused classes, laboratory sessions, or group project meetings will have their final grade in the course reduced by one letter grade. Students who miss more than ten of the aforementioned events will automatically fail the course.

Use of Laboratory Facilities

Throughout the semester, we will experiment with many different tools that data managers use during the phases of the data management process. The course instructor and the department's systems administrator have invested a considerable amount of time to ensure that our laboratories support the completion of both the laboratory assignments and the final project. To this end, students are required to complete all assignments and the final project while using the department's laboratory facilities. The course instructor and the systems administrator will only be able to devote a limited amount of time to the configuration of a student's personal computer.

Class Preparation

In order to minimize confusion and maximize learning, students must invest time to prepare for class discussions and lectures. During the class periods, the course instructor will often pose demanding questions that could require group discussion, the creation of a program or test suite, a vote on a thought-provoking issue, or a group presentation. Only students who have prepared for class by reading the assigned material and reviewing the current assignments will be able to effectively participate in these discussions. More importantly, only prepared students will be able to acquire the knowledge and skills that are needed to be successful in both this course and the field of data

management. In order to help students remain organized and effectively prepare for classes, the course instructor will maintain a class schedule with reading assignments and presentation slides. During the class sessions students will also be required to download, use, and modify programs, diagrams, and data sets that are made available through the course Web site. Students who are not comfortable with compiling, editing, and running Java programs should see the course instructor.

Email

Using your Allegheny College email address, I will sometimes send out class announcements about matters such as assignment clarifications or changes in the schedule. It is your responsibility to check your email at least once a day and to ensure that you can reliably send and receive emails. This class policy is based on the following statement in *The Compass*, the college's student handbook.

“The use of email is a primary method of communication on campus. . . . All students are provided with a campus email account and address while enrolled at Allegheny and are expected to check the account on a regular basis.”

Disability Services

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. Students with disabilities who believe they may need accommodations in this class are encouraged to contact Disability Services at 332-2898. Disability Services is part of the Learning Commons and is located in Pelletier Library. Please do this as soon as possible to ensure that approved accommodations are implemented in a timely fashion.

Honor Code

The Academic Honor Program that governs the entire academic program at Allegheny College is described in the Allegheny Course Catalogue. The Honor Program applies to all work that is submitted for academic credit or to meet non-credit requirements for graduation at Allegheny College. This includes all work assigned for this class (e.g., examinations, laboratory assignments, and the final project). All students who have enrolled in the College will work under the Honor Program. Each student who has matriculated at the College has acknowledged the following pledge:

I hereby recognize and pledge to fulfill my responsibilities, as defined in the Honor Code, and to maintain the integrity of both myself and the College community as a whole.

Welcome to an Adventure in Data Management

In reference to software, Frederick Brooks, Jr. wrote in *The Mythical Man Month*, “The magic of myth and legend has come true in our time.” Software—and the data that frequently drives it—is a pervasive aspect of our society that changes how we think and act. High quality software and well-designed databases also have the potential to positively influence the lives of people. Moreover, the collection and maintenance of data is an exciting and rewarding activity! At the start of this class, I invite you to pursue this adventure in data management with enthusiasm and vigor.