CMPSC 312
Database Systems
Syllabus

Spring 2022

Course Instructor

Dr. Oliver BONHAM-CARTER (said and written as "Bonham-Carter," not "Carter")
Email: obonhamcarter@allegheny.edu
Web Site: http://www.cs.allegheny.edu/sites/obonhamcarter/
Class and lab meeting place: Alden 101
Final Exam Code: D
Final deliverable due: 18^{th} May 2022 at 9:00am
Distribution Requirements: QR, SP
Syllabus updated on: February 21, 2022

Instructor's Office Hours

- Monday and Wednesday: 11:00am – 12:00pm (10 minute time slots)
- Tuesday and Thursday: 2:00pm – 4:00pm (10 minute time slots)
- By appointment

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 view my calendar or by clicking "schedule an appointment" link browse my office hours and schedule an appointment by clicking the correct link to reserve an open time slot.

Technical Leaders

- https://www.cs.allegheny.edu/teaching/technicalleaders/

Course Meeting Schedule

Lecture, Discussion, Presentations, and Group Work:
Duration: 21 Feb 2022 - 20 May 2022
T/Th 10:20 AM - 11:50 AM, Alden Hall, 101

Laboratory Session:
Duration: 21 Feb 2022 - 20 May 2022
M 2:50 PM - 4:40 PM Alden Hall, 101

Calendar

The calendar link is provided below to allow you to add the course and lab meeting times into your own Google calendar. Note, the whole link fits onto one line.

`https://calendar.google.com/calendar/u/0?cid=`
`Y19hc2sxdW00MnNmaTdvcTk4YmwxOHE0M2xrNEBncm91cC5jYWxlbmRhci5nb29nbGUuY29t`

Note: When copying and pasting the above hyperlink for the address, there are no spaces in the link.

Discord Channel

The below link will expire in 7 days from 21^{st} Feb 2022

`https://discord.gg/X7XXfvbX`

The `ClassDocs/` Repository

All materials given out in class will be accessible using the `classDocs/` repository. Note: The HTTP link works in absence of SSH keys.

Main site on GitHub:

- `https://github.com/Allegheny-ComputerScience-312-S2022/classDocs`

HTTPS:

- `git clone https://github.com/Allegheny-ComputerScience-312-S2022/classDocs.git`

SSH:

- `git clone git@github.com:Allegheny-ComputerScience-312-S2022/classDocs.git`

Academic Bulletin 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. Distribution Requirements: QR (Quantitative Reasoning), SP (Scientific Process and Knowledge).

Distribution Requirements

The following definitions were taken from the *Distribution Requirements: Learning Outcomes* website, `https://sites.allegheny.edu/registrar/academic-policies/graduation-requirements/distribution-requirement/distribution-requirements-learning-outcomes/`.

- *Quantitative Reasoning (QR).* Quantitative Reasoning is the ability to understand, investigate, communicate, and contextualize numerical, symbolic, and graphical information towards the exploration of natural, physical, behavioral, or social phenomena.
  - Learning Outcome: Students who successfully complete this requirement will demonstrate an understanding of how to interpret numeric data and/or their graphical or symbolic representations.
- *Scientific Process a Knowledge (SP).* Courses involving Scientific Process and Knowledge aim to convey an understanding of what is known or can be known about the natural world; apply scientific reasoning towards the analysis and synthesis of scientific information; and create scientifically literate citizens who can engage productively in problem solving.
  - Learning Outcome: Students who successfully complete this requirement will demonstrate an understanding of the nature, approaches, and domain of scientific inquiry.

This course meets the course distribution requirements of QR (Quantitative Reasoning) and SP (Scientific Process and Knowledge) for its use of applying concepts of computer programming to the design and creation databases which are tested on public data from real-world applications. In addition, the class aims to introduce an component of ethical reasoning in the design, maintenance and application of database systems for potentially sensitive data.

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. Ethical discussions are also introduced into the course to introduce students to the concepts of responsible computing.

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 will be working with some Python programming where code will be provided to be modified.

Required Textbooks

- *Database Systems Concepts, Fifth or Sixth Edition.* Avi Silberschatz, Henry F. Korth, and S. Sudarshan.

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	10%
First Examination	10%
Second Examination	10%
Laboratory Assignments	40%
Final Project	30%

These grading categories have the definitions which are defined below.

- *Class Participation*: 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, completing in-class exercises, asking constructive questions of the other members of the class, giving presentations, leading a discussion session in class and in the course's Slack channels.

- *First and Second Examinations*: The first and second 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.

- *Laboratory 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. Some of the laboratory assignments in this course will expect students to give both a short 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.

Assignment Submission

We will be using GitHub Classroom to collect all assignments. It is expected that you are able to effectively use `git` to submit your work. If you require help, please see your peers, the Technology Leaders, or your instructor.

All assignments will have a stated due date. The electronic version of the class assignments are to be turned in at the beginning of the lab session on that due date.

Submissions after the beginning of class are counted as being 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.

Extensions

Unless special arrangements are made with the course instructor, no assignments will be accepted after the late deadline. If you are requesting extensions for a lab assignment, then you are to email me with your request and also provide a *valid reason* for your extension. This request must come before the due date of the lab and not on the due date. Requests will not be granted where the reason appears to be insignificant. Extensions are 24 hours of extra time (after the original due date) and are given out at my discretion. The decision to provide you with an extension (or not) will be weighed in light of fairness to your peers who are still able to complete their labs, regardless of their own busy schedules.

The submission of homework comprises the Honor Code pledge of the student(s) completing the work. For any assignment completed in a group, students must also turn in a one-page reflection that describes each group member's contribution to the submitted deliverables.

Attendance

Classes will be attended by in-person and online students. Each class will be recorded to produce videos for online students and to enhance learning for the class.

If you will not be able to attend your session, then please email 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.

Labs: The laboratory sessions will be held online and therefore, it is the student's responsibility to check up on materials for lab and to ask questions when necessary to ensure comprehension of deliverables.

Bring your own computer to class

The classrooms in the Departement of Computer Science no longer provide machines for student use. You are to bring your own wifi-ready device to class to be able to follow along with course material. If the class is meeting online using Zoom, then please be sure that you machine is configured correctly to use these services to connect you to the class. As it is your responsibility to maintain your machine, please perform online research to determine how to configure your machine accordingly, or to install any necessary software to enable online meetings.

During the semester, you will be told which software to install on your machine to be prepared for class. Some of the prominant software that we may be using include;

- Git and GiHub (a software development software system): `https://github.com/`
- Atom (an editor): `https://atom.io/`
- Docker (a software container system): `https://www.docker.com/`
  - Installing Docker: `https://docs.docker.com/get-docker/`
  - Basic tutorial from Docker: `https://www.docker.com/101-tutorial`
  - Play with Docker: `https://labs.play-with-docker.com/`
- SQLite (a database system): `https://www.sqlite.org/index.html`
  - Some machines have differing methods of installing this database software. Please complete online research to read documentation to determine how to get your machine ready to run this software.

Assignment Completion

All assignments will have a stated due date. To accommodate for unforeseen life events, each student will be given an option of dropping one assignment grade at the end of the semester. The dropped grade cannot include the final proposal assignment. Otherwise, unless severe extenuating circumstances have been presented to the instructor, no assignments will be accepted after the deadline.

Extensions

Unless special arrangements are made with the course instructor, no assignments will be accepted after the late deadline. If you are requesting extensions for an assignment, then you are to email me with your request and also provide a *valid reason* for your extension. This request must come before the due date of the lab and not on the due date. Requests will not be granted where the reason appears to be insignificant. Extensions are 24 hours of extra time (after the original due date) and are given out at my discretion. The decision to provide you with an extension (or not) will be weighed in light of fairness to your peers who are still able to complete their labs, regardless of their own busy schedules.

A Note on extenuating circumstances

If you should find yourself in difficult circumstances that significantly interfere with your ability to prepare for this class and to complete assignments, please inform me immediately so that we can work something out together! Do not wait until the last day of class to ask for exceptions to what is stated in this syllabus. In such a situation, you may also find it helpful to contact one of the available resources on campus:

- The Maytum Learning Commons, Library/Academic Commons, `http://sites.allegheny.edu/learningcommons/tutoring/`, 814-332-2898
- Counseling & Personal Development Center, `https://sites.allegheny.edu/counseling/`, 814-332-2105
- Winslow Health Center, `https://sites.allegheny.edu/healthcenter/`, 814-332-4355

Communication

Various digital channels will be used in this course for communication, including email, Discord, and the GitHub issue tracker. It is strongly advised for the student to install the Discord app on their computer and smart-phone to be sure to receive all communications from the instructor, as well as, the other members of the class.

Additionally, the course website will be used to store the syllabus, course schedule and information about the `classDocs/` repository using the GitHub. Your grades will be communicated to you by a Gradebook GitHub repository.

Special Needs and 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.*

It is recognized that an important part of the learning process in any course, and particularly one in computer science, derives from thoughtful discussions with teachers and fellow students. Such dialogue is encouraged. However, it is necessary to distinguish carefully between the student who discusses the principles underlying a problem with others and the student who produces assignments that are identical to, or merely variations on, someone else's work. While it is acceptable for students in this class to discuss their programs, technical diagrams, proposals, paper reviews, presentations, and other items with their classmates or other individuals, deliverables that are nearly identical to the work of others will be taken as evidence of violating the Honor Code.