

CS 617 Databases, Data Mining, and Big Data

Spring 2017 Syllabus and Information

General Information

Contact Your Instructor

Name: Jeff Kinne

Email: kinne@cs.indstate.edu

Phone: 812-237-2136

Office: Root Hall, room A-120

Lecture, Exam, Office Hours

Lecture: Mon/Wed/Fri 11-11:50am in Science 134.

Exam: Friday Dec 15 from 10-11:50am.

Instructor Office Hours: I am generally in my office and available most MWF's from about 8:30am-4pm except when in other meetings. My official office hours are Wednesdays 9-10am, 1-3pm.

Unix lab: See <http://cs.indstate.edu/info/labs.html>

Website: this google doc, or find a link from kinnejeff.com

Prerequisites

CS graduate student or approval of instructor.

After this Course

This counts as a Data Management and Analysis course in the professional concentration of the CS MS degree. See <http://cs.indstate.edu/info/programs-grad.html> for more information.

Recommended text

There is no required textbook. Much of the information required for the course is available online. Sources that can be used for the course include the following.

- Lecture notes from a similar course from 2003 at MIT - <http://ocw.mit.edu/courses/sloan-school-of-management/15-062-data-mining-spring-2003/>
- Lecture slides from a similar course from 2016 at Stanford - <http://web.stanford.edu/class/cs246/handouts.html> And [videos](#)
- Lecture slides from a similar course at Purdue - <https://www.cs.purdue.edu/homes/clifton/cs490d/>

Textbooks that align well with this course include *Mining Massive Datasets* by Anand Rajaraman and Jeffrey Ullman (available for free download from <http://www.mmds.org/>) and *Data Mining: Concepts and Techniques* by Jiawei Han and Micheline Kamber.

Course Announcements

Announcements regarding the course will be made both during class and via email to your @sycamores.indstate.edu email address. You should regularly check this email account or have it forwarded to an account that you check regularly. You can set the account to forward by logging into your indstate.edu email from Internet Explorer (the "light" version of the webmail client that opens up from Firefox or Chrome does not give the option to forward email).

Classroom conduct

You may not use cell phones, iPods/music players, etc. during class. You should be civil and respectful to both the instructor and your classmates, and you should arrive to class a few minutes before the scheduled lecture so you are ready for lecture to begin on time. You may use your computer during class if you are using it to follow along with the examples that are being discussed. You may not check email, facebook, work on other courses, etc. during class.

Course Description

This course introduces students to ways to store, process, and analyze data. First the traditional framework of relational databases is studied. Relational databases are the standard used on websites and in businesses for storing client/customer data. When datasets become too large to store on a single server different strategies must be used, and the remainder of the course studies techniques and challenges related to extremely large datasets.

The official catalog description for the course is

“The course introduces students to using relational databases to accomplish common data

storage and analysis needs. The course also studies the techniques and tools available to deal with data that is too large to store in a database running on a single server. Topics in this part of the course may include data warehousing, map reduce, association rules, machine learning, and streaming algorithms.”

This course gives students exposure to using a variety of database tools and techniques for storing and analyzing data, skills that are important to managing the computing needs of most businesses.

Course Outline

1. Relational Databases
 - a. Overview of operators and syntax
 - b. Case study - mySql
2. Very large datasets
 - a. Overview of challenges
 - b. Case studies - the Internet, SETI@home
3. Data mining with very large datasets
 - a. Data warehousing
 - b. Association rules
 - c. Map reduce
 - d. Machine learning
 - e. Streaming algorithms
 - f. Tools - Hadoop

Grading and Assignments

The students of this course have the following responsibilities: read assigned readings before lecture, attend lecture, complete homework assignments, take in-class quizzes, take exams, and complete a project. You will have a grade for attendance, homeworks + quizzes, exams, and the project.

Your “total” or “overall” grade will be the minimum of these. There may also be assignments or quizzes labeled as “checkpoints” - which will mean that if you do not complete them you either cannot pass the class, or cannot get a grade of C (the assignment will specify whether it is a checkpoint for passing or a checkpoint for earning a C).

- **Project:** graded based on correctness, proper coding style, completed sub-tasks I asked you to, good explanation of it during your presentation.
- **Homeworks and Quizzes:** average of all homework and quiz grades, computed as a weighted average so that assignments worth more points count more. Quizzes cannot

be “made up”; homeworks can be completed for 50% late credit.

- **Exams:** 3 exams total, with the exam grade calculated as
 $\max(.2*\text{exam1} + .3*\text{exam2} + .5*\text{exam3},$
 $.4*\text{exam2} + .6*\text{exam3},$
 $\text{exam3})$
- **Class Attendance:** computed as $.5*\text{lecture} + .5*\text{lab}$ attendance. Note that you are required to spend 5 hours per week in the unix lab and logged in and working. For unix lab attendance, one week will be dropped.

CS Course Policies

Note that this course follows all standard CS course policies. In particular, (a) cheating/plagiarism by graduate students (for courses with graduate students) results in an F in the course, (b) missing 20% of the classes (lecture attendance) results in an F for any student, and (c) there will be no makeup exams. See <http://cs.indstate.edu/info/policies.html> for details.

Late Homeworks

All homework assignments will be given a preferred due date. Assignments can be turned in past the preferred due date, but any assignments turned in late will have their value multiplied by 50% (so the highest grade you can get on a late assignment is 50%). Some assignments might have a “final due date” past which no credit will be given.

Start Homeworks Early

I suggest attempting a homework assignment the day it is given, or the day after, so that if you have a problem you can ask early. If you continue to have problems in trying to complete the assignment, you will have time to ask again. Many of the homework assignments require thought and problem solving, which takes “time on the calendar” not just “time on the clock”. By that I mean that spending an hour on 3 consecutive days is likely to be more productive than trying to spend 3 hours at once on the assignment.

Expected Amount of Work

I hope that doing well in this course ends up taking about 10 hours/week for most students.

Grade Cutoffs

I will try design homework assignments and exams so that a standard cutoff for grades will be close to what you deserve. After the first exam I will create a grade in Blackboard called “Letter

Grade” that is what your letter grade would be if the semester ended today. Initially, I will likely assign the following grades: 93-100 A, 90-93 A-, 87-90 B+, 83-87 B, 80-83 B-, 77-80 C+, 73-77 C, 70-73 C-, 67-70 D+, 63-67 D, 60-63 D-, 0-60 F

My goal is that the different grades have the following rough meaning.

A

You understand everything and probably could teach the course yourself.

B+/A-

You understand most of what you should, and should be all set to use this knowledge in other courses or in a job.

B-/B

Some things you understand well and others you don't (more towards the former for a B and more towards the latter for a B-).

C/C+

You did put some effort in, and understand many things at a high level, but you haven't mastered the details well enough to be able to use this knowledge in the future.

F

Normally, students that get an F simply stopped doing the required work at some point.

Blackboard

The course has a blackboard site. Click [here](#) to go to blackboard. You should see this course listed under your courses for the current term. The blackboard site is only used for giving you your grades (go to the course in blackboard, then click “My Tools”, and then “My Grades”). All course content, schedule, etc. is kept in this google doc (which you are currently viewing).

Academic Integrity

Follow the standard CS course policies in terms of what is and is not allowed on assignments: <http://cs.indstate.edu/info/policies.html>

Please ask the instructor if you have doubts about what is considered cheating in this course.

Special Needs / Student Disabilities

Standard language included in the syllabi for ISU courses.

Indiana State University recognizes that students with disabilities may have special needs that must be met to give them equal access to college programs and facilities. If you need course adaptations or accommodations because of a disability, please contact us as soon as possible in a confidential setting either after class or in my office. All conversations regarding your disability will be kept in strict confidence. Indiana State University's Student Support Services (SSS) office coordinates services for students with disabilities: documentation of a disability needs to be on file in that office before any accommodations can be provided. Student Support Services is located on the lower level of Normal Hall in the [Center for Student Success](#) and can be contacted at 812-237-2700, or you can visit the ISU website under A-Z, [Disability Student Services](#) and submit a Contact Form. Appointments to discuss accommodations with SSS staff members are encouraged.

Once a faculty member is notified by Student Support Services that a student is qualified to receive academic accommodations, a faculty member is obligated to provide or allow a reasonable classroom accommodation under ADA.

Disclosures Regarding Sexual Misconduct

Standard language included in the syllabi for ISU courses.

Indiana State University fosters a campus free of sexual misconduct including sexual harassment, sexual violence, intimate partner violence, and stalking and/or any form of sex or gender discrimination. If you disclose a potential violation of the sexual misconduct policy I will need to notify the Title IX Coordinator. Students who have experienced sexual misconduct are encouraged to contact confidential resources listed below. To make a report or the Title IX Coordinator, visit the Equal Opportunity and Title IX website:

<http://www.indstate.edu/equalopportunity-titleix/titleix>.

The ISU Student Counseling Center – HMSU 7th Floor | 812-237-3939 | www.indstate.edu/cns

The ISU Victim Advocate – Trista Gibbons, trista.gibbons@indstate.edu

HMSU 7th Floor | 812-237-3939 (office) | 812-230-3803 (cell)

Campus Ministries - United Campus Ministries | 812-232-0186

<http://www2.indstate.edu/sao/campusministries.htm>

www.unitedcampusministries.org | ucmminister2@gmail.com

321 N 7th St., Terre Haute, IN 47807

For more information on your rights and available resources

<http://www.indstate.edu/equalopportunity-titleix/titleix>