

CS257 Spring 2024 Sec1 Home Page/Syllabus

Database System Principles

Room change to ISB 876!

Instructor: [Chris Pollett](#)

Office: MH 214

Phone Number: (408) 924 5145

Email: chris@pollett.org

Office Hours: MW 3:00-4:15pm in MH214

Class Meets:

Sec1 MW 1:30-2:45pm in ISB 876

Prerequisites

To take this class you must have taken:

[CS157B](#)

with a grade of C- or better.

Texts and Links

Required Texts:	Principles of Database Management. Wilfred Lemahieu, Seppe Vanden Broucke, Bart Baesens. Cambridge. 2018.
Online References and Other Links:	PostgresSql. BaseX. MongoDB. Cassandra. Neo4j.

Description

According to the catalog, this course covers: Design management and performance issues on: file organization and access methods, buffer management and storage management. Query processing and query optimization, transaction management, recovery, and concurrency control techniques. Reliability, protection and integrity techniques. Extensive programming project. **This course is currently under revision to fit into the context of our three other database classes:** CS157A, CS157B, and CS157C. CS157A covers the design and theory of databases for use in a relational database system, queries in languages like SQL for such systems, and simple transactions for these system. CS157B covers how file systems and database systems are built, how query compilation and optimization work in these systems, concurrency control, recovery, and simple data warehousing. Finally, CS157C covers a quick overview of key-value, column-oriented, document-oriented, and graph stores, replication and sharding, and then looks at MongoDB and Cassandra. CS257 builds on these courses by considering semi-structured data in more detail such as XML and XML query languages and web and enterprise search. We also discuss and deploy a graph database. CS 257 then reviews NoSQL databases and considers more advanced aspects of physical database organization. We review data warehousing of structured data and then look at how to handle data lakes and unstructured data warehouses. We look at business intelligence techniques related

to these. We consider different models for measuring data quality and processes for data management. We look at concrete big data stacks such as Hadoop and implement a non-trivial map reduce project using it. Finally, we consider different analytic processing models and machine learning techniques for analyzing big data and data related to social networks.

Course Learning Outcomes (CLOs)

By the end of this course, a student should be able to:

CLO1 -- Be able to build and deploy a database making use of XML columns as well as suitable queries to those columns for a concrete use case

CLO2 -- Be able to deploy and understand the use cases of a graph-based database

CLO3 -- Be able to code an application that makes use of a database API to a NoSQL store.

CLO4 -- Code a basic inverted index capable of performing conjunctive queries or code another advanced DBMS data structure such as Bloom Filters, or Log-structured Merge Trees.

CLO5 -- Be able to deploy a concrete business intelligence system making use of OLAP features in SQL

CLO6 -- Be able to evaluate the data governance of a concrete hypothetical organization according to a data management framework such as TDQM (Total Data Quality Management) or CMMI (Capability Maturity Model Integration)

CLO7 -- Be able to code or analyze a common clustering algorithm.

CLO8 -- Be able to implement in a big data stack one non-trivial Map Reduce algorithm (for example to calculate page rank).

Course Schedule

Below is a tentative time table for when we'll do things this quarter:

Week 1:Jan 22, Jan 24	Start Ch 10.
Week 2:Jan 29, Jan 31	Ch 10 XML Databases
Week 3:Feb 5, Feb 7	Finish Ch 10
Week 4:Feb 12, Feb 14(Hw1)	Start Ch 11 NoSQL Databases
Week 5:Feb 19, Feb 21	More NoSQL Databases
Week 6:Feb 26, Feb 28	Finish Ch 11
Week 7:Mar 4(Hw2), Mar 6	Ch 12 Physical File Organization and Indexing
Week 8:Mar 11, Mar 13(Midterm)	Review
Week 9:Mar 18, Mar 20	Finish Ch 12
Week 10:Mar 25, Mar 27(Hw3)	Ch 13 Physical Database Organization
Week 11:Apr 1, Apr 3	Spring Recess
Week 12:Apr 8, Apr 10	Finish Ch 13.
Week 13:Apr 15, Apr 17(Hw4)	Ch 17 Datawarehousing and Business Intelligence.
Week 14:Apr 22, Apr 24	Ch 18 Data Integration, Quality, and Governance.
Week 15:Apr 29, May 1	Ch 19.

Week 16:May 6, May 8	Ch 20 Analytics
Week 17:May 13(Hw5)	Review
	The final will be Thursday, May 16 12:15-2:30 PM

Grading

HWs and Quizzes	50%
Midterm	20%
Final	30%
Total	100%

Grades will be calculated in the following manner: The person or persons with the highest aggregate score will receive an A+. A score of 55 will be the cut-off for a B-. The region between this high and low score will be divided into five equal-sized regions. From the top region to the low region, a score falling within a region receives the grade: A, A-, B+, B, B-. If the boundary between an A and an A- is 85, then the score 85 counts as an A-. Scores below 55 but above 50 receive the grade D. Those below 50 receive the grade F.

If you do better than an A- in this class and want me to write you a letter of recommendation, I will generally be willing provided you ask me within two years of taking my course. Be advised that I write better letters if I know you to some degree.

Course Requirements, Homework, Quiz Info, and In-class exercises

This semester we will have five homeworks, weekly quizzes, and weekly in-class exercises.

Every Monday this semester, except the first day of class, the Midterm Review Day, and holidays, there will be a quiz on the previous week's material. The answer to the quiz will either be multiple choice, true-false, or a simple numeric answer that does not require a calculator. Each quiz is worth a maximum of 1pt with no partial credit being given. Out of the total of thirteen quizzes this semester, I will keep your ten best scores.

On Wednesday's, we will spend 15-20 minutes of class on an in-class exercise. You will be asked to post your solution to these exercises to the class discussion board. Doing so is worth 1 "insurance point" towards your grade. A "insurance point" can be used to get one missed point back on a midterm or final, up to half of that test's total score. For example, if you scored 0 on the midterm and have 10 insurance points, you can use your insurance points, so that your midterm score is a 10. On the other hand, if you score 18/20 on the midterm, you can use at most 1 insurance point since half of what you missed (2pts) on the midterm is 1pt. In addition, to the weekly in-class exercises, one insurance point is available if in the week before the midterm you can convince me I know your name, and in the week before the final, I still know your name (Please help me improve my memory).

Links to the current list of homeworks and quizzes can be found on the left hand side of the class homepage. After an assignment has been returned, a link to its solution (based on the best student solutions) will be placed off the assignment page. Material from assignments may appear on midterms and finals. **For homeworks you are encouraged to work in groups of up to three people. Only one person out of this group needs to submit the homework assignment; however, the members of the group need to be clearly identified in all submitted files.**

Homeworks for this class will be submitted and returned completely electronically using the Canvas link for the name of the homework. Hardcopies or e-mail versions of your assignments will be rejected and not receive credit. Homeworks will always be due by midnight according to the Canvas server on the day their due. Late homeworks will not be accepted and missed quizzes cannot be made up; however, your lowest score amongst the first four homeworks and your quiz total will be dropped. Homework 5 can't be substituted for.

When doing the programming part of an assignment please make sure to adhere to the specification given as closely as possible. Names of files should be as given, etc. Failure to follow the specification may result in your homework not being graded and you receiving a zero for your work.

Classroom Protocol

I will start lecturing close to the official start time for this class modulo getting tangled up in any audio/visual presentation tools I am using. Once I start lecturing, please refrain from talking to each other, answering your cell phone, etc. If something I am talking about is unclear to you, feel free to ask a question about it. Typically, on practice tests days, you will get to work in groups, and in so doing, turn your desks facing each other, etc. Please return your desks back to the way they were at the end of class. This class has an online class discussion board which can be used to post questions relating to the homework and tests. Please keep discussions on this board civil. This board will be moderated. Class and discussion board participation, although not a component of your grade, will be considered if you ask me to write you a letter of recommendation.

Exams

The midterm will be during class time on: Mar 13.

The final will be: Thursday, May 16 12:15-2:30 PM.

All exams are closed book, closed notes and in this classroom. You will be allowed only the test and your pen or pencil on your desk during these exams. The final will cover material from the whole semester although there will be an emphasis on material after the last midterm. No make ups will be given. The final exam may be scaled to replace a midterm grade if it was missed under provably legitimate circumstances. These exams will test whether or not you have mastered the material both presented in class or assigned as homework during the quarter. My exams usually consist of a series of essay style questions. I try to avoid making tricky problems. The week before each exam I will give out a list of problems representative of the level of difficulty of problems the student will be expected to answer on the exam. Any disputes concerning grades on exams should be directed to me, Professor Pollett.

Regrades

If you believe an error was made in the grading of your program or exam, you may request **in person** a regrade from me, Professor Pollett, during my office hours. **I do not accept e-mail requests for regrades.** A request for a regrade must be made no more than a week after the homework or a midterm is returned. If you cannot find me before the end of the semester and you would like to request a regrade of your final, you may see me **in person** at the start of the immediately following semester.

University Policies and Procedures

SJSU adheres to required safety measures from the California Department of Public Health and the Santa Clara County Public Health Department. Please refer to our [SJSU Health Advisories website](#) for the latest information and updates.

Per [University Policy S16-9](#), relevant university policy concerning all courses, such as student responsibilities, academic integrity, accommodations, dropping and adding, consent for recording of class, etc. and available student services (e.g. learning assistance, counseling, and other resources) are listed on [Syllabus Information web page](#) (<https://www.sjsu.edu/curriculum/courses/syllabus-info.php>). Make sure to visit this page to review and be aware of these university policies and resources. Below are some brief comments on some of these policies as they pertain to this class.

Academic Integrity

For this class, you should obviously not cheat on tests. For homeworks, you should not discuss or share code or problem solutions between groups! At a minimum a 0 on the assignment or test will be given. Faculty members are required to report all infractions to the Office of Student Conduct and Ethical Development.

Accommodations

If you need a classroom accommodation for this class, and have registered with the [Accessible Education Center](#), please come see me earlier rather than later in the semester to give me a heads up on how to be of assistance.