CAS CS 460: Introduction to Database Systems

Boston University
Fall 2021

Class Syllabus

Course Description: Our everyday activities, our business and government management activities, and scientific discovery today are heavily based on generating, storing, managing, and accessing massive amounts of data. **We live in a *data-driven world*.** Database systems provide the necessary infrastructure to manage huge data collections. This class serves as a comprehensive introduction in the key concepts of the architecture of modern database systems. We will discuss both traditional approaches used modern trends that shape the data management industry today. The primary focus of the course will be on the core concepts of the *internals* of database systems, covering entity-relationship and relational data models, commercial relational query languages (SQL and relational algebra), file organization, storage and memory management, indexing and hashing, query optimization, query processing, transaction processing, concurrency control and recovery. Finally, we will cover new trends in data management including Big Data and NoSQL databases and data management on the Cloud, and we will discuss the history of database systems and their evolution over the years.

Prerequisites: CAS CS 112. Working knowledge of Python, Java, or C++ programming, data structures, and algorithms. CS 350 is recommended.

Instructor: Manos Athanassoulis (mathan@bu.edu)
office hours: Tue/Thu, 3:30-4:30 pm (after class)
office: MCS 106

Teaching Assistants: Aneesh Raman (aneeshr@bu.edu), Tarikul Islam Papon (papon@bu.edu), Subhadeep Sarkar (ssarkar1@bu.edu)
office hours: (check Piazza)

Meeting Times and Places
lectures: Tue/Thu, 2:00-3:15 pm, HAR 208
labs: Tue, 9:30-10:20, MCS B33 & 11:15-12:05, MCS B29 & 12:30-1:20, KCB 107 ***(tentative)***

Course Website: https://bu-disc.github.io/CS460/
All class assignments, schedules, and lecture notes can be found on this page. We will also use Piazza for discussions and other material distribution.

Required Textbook: R. Ramakrishnan and J. Gehrke. *Database Management Systems*. Third Edition. McGraw-Hill 2002. Throughout the class we will cover a few topics from recent research and survey papers.

Additional Reading Material: The following are excellent sources for additional reading.

- Architecture of a Database System, by J. Hellerstein, M. Stonebraker and J. Hamilton
- The Design and Implementation of Modern Column-store Database Systems, by D. Abadi, P. Boncz, S. Harizopoulos, S. Idreos, S. Madden
- Modern B-Tree Techniques, by Goetz Graefe, Foundations and Trends in Databases, 2011

Course Requirements
A. Collaborative Notes: Starting from class 3, the students will contribute towards a shared collaborative document. Each student will contribute notes to at least one class, and each class will have at least two contributors.
B. Programming Assignments: During the semester there will be two programming assignments. The first one will consist of building an application for managing data of a real-life application. The second project will help the students understand how query optimization works via hands-on experimentation with real-life database systems. The projects will be based on topics taught in class, and there will be additional hands-on support during the labs.
C. Written Assignments: Approximately every two weeks there will be one written assignment based on concepts taught in class. The assignments will be graded on completion basis, that is, you will receive full credit as long as you submit them on time and receive a grade above 70%.
D. Hands on tests: There will be two hands on tests for extra credit. In both cases, the students will have a short amount of time during class to write a few queries (SQL and Key-Value-based) based on query languages taught in class and discusses in labs.

Grading Policy: The course grade will break down as follows (minor alterations may occur):

- Class participation: 5%
- Written Assignments: 10%
- Programming Assignments: 30%
- Midterm: 20%
- Final: 35%
- Hands-on Test (bonus): 5% (SQL 3% & Key-Value Querying 2%)

Late Policy: Students needing additional time may submit **programming assignments** late. There will be no penalty for up to four (4) total late days in the semester. Any additional late days would result to a zero in the corresponding assignment. There are **no late days allowed for written assignments**.

Important Dates for all classes
September 16th, last day to add a class
October 7th, last day to drop (without a "W")
November 5th, last day to drop (with a "W") & to designate the course as Pass/Fail

Tentative Schedule

Week #	Topics	Readings
1	Introduction & Data Systems Architectures Essentials	Chapter 1
2	ER Model & Relational Model	Chapter 2, 3
3	Functional Dependencies & Schema Normalization	Chapter 19
4	Relational Algebra & SQL	Chapter 4, 5

5	File & Storage Organization	Chapter 8, 9
6	Indexing; Hashing and B-Trees	Chapter 10, 11
7	Advanced Indexing & External Sorting	Chapter 13
8	Review & Midterm 1	
9	Query Processing	Chapter 12, 14
10	Query Optimization	Chapter 15
11	Transactions	Chapter 16
12	Concurrency Control	Chapter 17
13	Recovery	Chapter 18
14	Cloud Data Management & NoSQL Key-Value Stores	paper-based
15	Research Topics & Review	

Collaboration Policy
You are strongly encouraged to collaborate with one another in studying the lecture materials and preparing for reviews and presentations.

You may discuss ideas and approaches to the projects with others (provided that you acknowledge doing so in your solution), but such discussions should be kept at a high level, and should not involve actual details of the code or of other types of answers. **You must complete the actual solutions on your own.**

Academic Misconduct
We will assume that you understand BU's Academic Conduct Code:
http://www.bu.edu/academics/policies/academic-conduct-code

Prohibited behaviors include:

- copying all or part of someone else's work, even if you subsequently modify it; this includes cases in which someone tells you what you should write for your solution
- viewing all or part of someone else's work
- showing all or part of your work to another student
- consulting solutions from past semesters, or those found online or in books
- posting your work where others can view it (e.g., online).

Incidents of academic misconduct will be reported to the Academic Conduct Committee (ACC). The ACC may suspend/expel students found guilty of misconduct. ***At a minimum, students who engage in misconduct will have their final grade reduced by one letter grade (e.g., from a B to a C).***

Course In-Person Attendance Policy
The class is now back to full in-person mode. Students and staff are expected to follow the BU mandates regarding public health. Specifically, students and staff are expected to wear masks during lectures. If the classroom allows for a 12-feet distance between the instructor and the students, the instructor may remove the mask to enhance the delivery of the lectures.