CAS CS 460: Introduction to Database Systems

Boston University
Fall 2020

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: **TBA**
office: MCS 106 (or in zoom link provided in class piazza)

Teaching Assistants: Dimitris Staratzis (dstara@bu.edu), Subhadeep Sarkar (ssarkar1@bu.edu), Tarikul Islam Papon (papon@bu.edu), Andy Huynh (ndhuynh@bu.edu)

Meeting Times and Places
lectures: Tue/Thu, 3:30-4:45 pm, CAS B12
labs: Tue, 9:30-10:20, 11:15-12:05, 12:30-1:20, PSY B35 ***(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

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

- Class participation: 5%
- Written Assignments: 20%
- Programming Assignments: 30%
- Midterm 1: 20%
- Midterm 2: 25%
- SQL Hands-on Test (bonus): 5%

Important Dates for all classes
September 16th, last day to add a class
October 7th, last day to drop (without a "W")
November 6th, last day to drop (with a "W")

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 & Recovery	Chapter 17, 18
13	BigData, NoSQL, and Key-Value Stores	paper-based
14	Research Topics	paper-based
15	Midterm 2 (Final)	

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 Recording Policy
All class sessions will be recorded for the benefit of registered students who are unable to attend live sessions (either in person or remotely) due to time zone differences, illness or other special circumstances. Recorded sessions will be made available to registered students ONLY via their password-protected piazza account. Students may not share such sessions with anyone not registered in the course and may certainly not repost them in a public platform. Students have the right to opt-out of being part of the class recording. Please contact your instructor or teaching assistant to discuss options for attending the course in such cases.

Synchronous Exam Policy
The student assessment via exams will take place via synchronous exams, which will require a stable internet connection and a working camera for every student. Please reach out to the instructor with any questions or requests for technological support in the beginning of the semester. Details about multiple exam options will be finalized upon receiving from the class roster more information about the time zone each student will be physically present.

Remote Class Attendance
A unique password-protected Zoom link for all class meetings (lecture and labs) will be generated and distributed to students via piazza in the beginning of the semester. All CS460 students can use this link to follow the class remotely. Students may not share this information with anyone not registered to the course.