TEACHING STATEMENT

Lin Ma (lin.ma@cs.cmu.edu)

I am interested in teaching both system courses in databases and applied courses on data science topics. I want to guide students in applying their knowledge to real-world problems and innovating new solutions to solve future challenges. My previous experiences as a course instructor and teaching assistants have prepared me with skills and insights to achieve such goals.

Teaching Experience

I have formal teaching experiences with two different roles.

Instructor

I was a co-instructor of the upper level "15-445/645 Introduction to Database Systems" class at CMU in 2021[1]. I helped plan the course materials. And I was responsible for delivering half of the course lectures. The topics included query optimization, concurrency control, logging and recovery, and distributed databases. I also managed several TAs and more than 100 enrolled students with the other instructor. I found a few teaching methods that were effective in this course, which I plan to continue in my future courses:

1. *Have an organized course structure with a unified theme.* – It is one thing to teach students about the individual database components and algorithms, but it is more beneficial if the students understand the connection between these techniques' designs. I discussed each component in the context of the larger DBMS architecture and how they work collaboratively.

2. *Make intuitive explanations with simple terms.* – I believe even the most complex concepts should have clear intuitions that the students can appreciate. And the instructor should concisely deliver the essence of the material. I focused on teaching the core insights in database techniques with graphical illustrations in my slides.

3. *Connect the course material with real-world applications.* – Like many areas of computer science, there is no "correct" solution to a database problem. The students should realize the trade-offs among alternative solutions. I related the course material to the design choices made by real-world DBMSs and the reasonings behind them. I think this helped students identify the proper solution for a particular problem based on its context.

I also sent feedback forms to the students during the semester to help improve my teaching. There were several constructive suggestions, such as giving more demonstrations during the lecture. I want to continue to make efforts in these areas to become a better teacher. I also tried to have an observer from the Eberly Center at CMU to evaluate my lectures. But they were unable to come in person because of the pandemic. I want to take advantage of similar resources at my next university.

Teaching Assistants

I was a teaching assistant in the upper-level "15-445/645 Introduction to Database Systems" class at CMU in 2018[2], as well as the PhD-level "15-721 Advanced Database Systems" class at CMU in 2019[3]. In both courses, I helped design homeworks and projects and held office hours to answer students' questions. We designed all projects for 15-721 on the DBMS that I was building, and I worked closely with students to help them with technical problems. An important lesson I learned is to design the assignments with the right level of difficulty. Thus, the students could evaluate their understanding of the course material and sufficiently exercise what they have learned within a reasonable time frame. I found that discussing the pros and cons of different approaches with the students sharpened my understanding of the material and helped my teaching as well.

[1] https://15445.courses.cs.cmu.edu/fall2021/
[2] https://15445.courses.cs.cmu.edu/fall2018/
[3] https://15721.courses.cs.cmu.edu/spring2019/

Teaching Interests

Given my teaching and research experiences, I am eager to teach both undergraduate- and graduate-level courses in database systems, database applications, and data science.

For undergraduate students, I want to teach a course on how to implement high-performance database systems. Many design decisions and algorithms from the early databases in the 1970s still influence the implementation of modern DBMSs. But shifts in applications and computing environments (e.g., emerging serverless DBMSs in the cloud) have also fostered many new database architecture developments in the last decade. I will discuss both the traditional methods and new techniques in databases, as well as how DBMS technologies evolve over time as workloads and hardware change. These insights will prepare students for the data management challenges they may encounter in the future.

For graduate-level courses, I am particularly interested in teaching state-of-the-art research topics at the intersection of databases, machine learning (ML), and artificial intelligence (AI). There are many open questions in how to use ML/AI to automate database administration (e.g., autonomous databases) and enhance DBMS components (e.g., learned cardinality estimators). Using database techniques to improve the efficiency of ML systems is also an interesting topic to discuss. I want to make students aware of the exciting challenges and opportunities in this interdisciplinary area.

Research Mentoring

Mentoring is essential in helping students to achieve their career goals. I fortunately had the opportunity to mentor several junior students as a PhD student and postdoc. I try to guide the students that I mentor to have meaningful participation in the projects and learn new research skills. In particular, I collaborated with two undergraduate students for almost two years:

1. I have been working with one student to design a modeling framework for self-driving DBMSs. He is strong on system programming but relatively new to research. I helped him formulate the research problem and identify the challenges that we needed to solve. He developed a set of specialized data generators that provides the foundations for our modeling process. He became the second author of our SIGMOD paper and joined a database start-up after graduation. Starting spring 2022, he will return to CMU as a PhD student.

2. I have been mentoring another student to develop a Monte-Carlo tree search based framework to select actions for self-driving DBMSs. She did not have much experience in working with large systems at the beginning. I guided her to learn system programming skills while developing the search framework. She eventually integrated this advanced framework into our DBMS. She became the third author of our SIGMOD paper and now continues to work with me on a follow-up research project as a senior student.

In both cases, I regularly met with the students to discuss their progress and helped them iteratively improve their solutions, design experiments, and analyze the results.