

# Teaching Statement

Andrew Pavlo  
pavlo@cs.brown.edu

As a Ph.D. student, I have taught two different graduate-level seminar courses on advanced database management system (DBMS) topics. I also have worked as a teaching assistant for three semesters of DBMS courses. I believe that I learned several lessons from these endeavors. My experiences from them provides me with keen insight on how to approach my future undergraduate and graduate courses on database management, data science, and distributed systems.

The first seminar course that I taught under my advisor's supervision was in 2011 on non-standard, Web-scale DBMSs (i.e., *NoSQL*). I designed the course such that the first half of the semester focused on the academic literature that provides the architectural foundation for many NoSQL systems (e.g., BigTable, Dynamo). In the latter half, I had the students study real-world implementations of these designs. The students were each assigned a system to research. They then had to write a short report about that system and then teach it to the class. In addition to this, each student implemented a benchmark driver for their system using a common framework that I created and we held a contest to see who could get their system to go the fastest. For this part of the course, I received an educational grant from Amazon for students to deploy their systems on the EC2 cloud-computing platform. I encouraged students to contact the NoSQL developers directly to assist them when the system did not perform as expected.

As a follow-up to the NoSQL course, in 2012 I taught a seminar course on modern transaction processing DBMSs (i.e., *NewSQL*). I based the course around the idea of having the students read papers on each of the key components that one would have to implement in order to build a large-scale, data-intensive relational DBMS. For example, one week we read papers on how to implement fast logging in a DBMS, and then the following week we explored techniques for improving the two-phase commit protocol. For the project portion of the course, I was not able to do the same kind of the benchmarking assignment that I did in the NoSQL course because many of the NewSQL systems were either not open source or required specialized hardware. Instead, I had the students select project topics that I developed based on the H-Store system. This proved to be more challenging (and time consuming for me) than I had anticipated, but most students submitted projects that surpassed my expectations.

The key lesson that I learned from these experiences is that holding a friendly competition in the course caused student to be more engaged. It is one thing to have a student complete an assignment only to see that it works correctly, but it is more fulfilling if the student makes an effort to improve their implementation. In the case of the NoSQL benchmark projects, most students spent the time to try to better understand the underlying principles of the DBMS in order to improve its performance and increase their standing in the class leader-board, rather than merely stopping once their project was "correct" (i.e., it executed the right queries for the benchmark). Hence, I plan to extend this concept to my future introduction to databases courses. For example, for projects on concurrency control in DBMSs, I will have students implement either a two-phase locking or multi-version scheme in a simple DBMS and then maintain a ranking of who has the fastest implementation in the class. I will then use these results as a teaching tool to discuss with the class why one approach (or implementation) performed better than another.

With so much activity taking place in the commercial database systems space in the last five years, it is important that students understand how the material fits into the broader context. For my first seminar course, I plan on teaching NewSQL systems and will once again invite guest speakers from database companies to discuss the types of problems that they are solving for customers. Students continually tell me that this is the part of my courses that they have enjoyed the most.

Overall, I found teaching to be a gratifying experience, since it forced me to focus on the scientific fundamentals of databases and distributed systems. As a professor, I will design the next generation of data management courses that will also incorporate elements from information retrieval, data mining, and parallel computing fields.