



SCHOOL OF COMPUTER SCIENCE

Faculty Candidate

Andrew Pavlo
Brown University

Everything I Know About Fast Databases I Learned at the Dog Track

An emerging class of distributed database management systems (DBMS), known as NewSQL, provide the same scalable performance of NoSQL systems while maintaining the consistency guarantees of a traditional, single-node DBMS. These NewSQL systems achieve high throughput rates for data-intensive applications by storing their databases in a cluster of main memory partitions. This partitioning enables them to eschew much of the legacy, disk-oriented architecture that slows down traditional systems, such as heavy-weight concurrency control algorithms, thereby allowing for the efficient execution of single-node transactions.

But many applications cannot be partitioned such that all of their transactions execute in this manner; these multi-node transactions require expensive coordination that inhibits performance. Thus, without intelligent methods to overcome these impediments, a NewSQL DBMS will scale no better than a traditional DBMS.

In this talk, I will present our research on integrating machine learning techniques to improve the performance of fast database systems that is inspired by my adventures at greyhound racing tracks. In particular, I will discuss my work on the H-Store parallel, main memory transaction processing system. I will first describe the Houdini framework that uses Markov models to predict transactions' behaviors to allow a DBMS to selectively enable runtime optimizations. I will then present Hermes, a method for the deterministic execution of speculative transactions whenever a DBMS stalls because of distributed transactions. Together, these projects enable H-Store to support transactional workloads that are beyond what single-node systems can handle.



Andy Pavlo is Ph.D. candidate at Brown University working on database management systems under the circumspect guidance of Stan Zdonik and Michael Stonebraker. His most recent work is focused on the research and development of the H-Store distributed transaction processing system (since commercialized as VoltDB). Before this, he was a systems programmer for the Condor Project at the University of Wisconsin-Madison with Miron Livny.

Thursday, March 21st
1:00 p.m. GHC 6115

Host: Garth Gibson

For Appointments: Contact Jenn Landefeld jennsbl@cs.cmu.edu, x8-4740