

Invited Talk Holger Pirk, Imperial College London

TITLE: Hardware-conscious data processing systems

ABSTRACT:

Hardware-conscious database systems evaluate queries in milliseconds that take minutes in conventional systems, turning long-running jobs into interactive queries. However, the plethora of hardware-focused tuning techniques creates a design-space that is hard to navigate for a skilled performance engineer and even harder to exploit for modern, code-generating data processing systems. In addition, hardware-conscious tuning is often at odds with other design goals such as implementation effort, ease of use and maintainability -- in particular when developing code-generating database systems. Well-designed programming abstractions are essential to allow the creation of systems that are fast, easy to use and maintainable.

In my talk, I demonstrate how existing frameworks for high-performance, data-parallel programming fall short of this goal. I argue that the poor performance of many mainstream data processing systems is due to the lack of an appropriate intermediate abstraction layer, i.e., one that is amenable to both, traditional data-oriented as well as low-level hardware-focused optimizations.

To address this problem, I introduce Voodoo, a data parallel intermediate language that is abstract enough to allow effective code generation and optimization but low-level enough to express many common optimizations such as parallelization, loop tiling or memory locality optimizations. I demonstrate how we used Voodoo to build a relational data processing system that outperforms the fastest state-of-the-art in-memory database systems by up to five times. I also demonstrate how Voodoo can be used as a performance engineering framework, allowing the expression of many known optimizations and even enabling the discovery of entirely new optimizations.

BIO:

Holger is an Assistant Professor (Lecturer) in the Department of Computing at Imperial College London. As such, he is a member of the Large-Scale Data and Systems Group. Before that, he was a Postdoc at the Database group at MIT CSAIL. He spent his PhD years in the Database Architectures group at CWI in Amsterdam resulting in a PhD from the University of Amsterdam in 2015. He received a master's degree (Diplom) in computer science at Humboldt-Universität zu Berlin in 2010. His research interests lie in analytical query processing on memory-resident data. In particular, he studies storage schemes and processing models for modern hardware.

Wednesday, 23. May 2018, 13:00, room APB 3105