CrossMark

Special Issue: Modern Hardware

Peter Boncz[1] · Wolfgang Lehner[2] · Thomas Neumann[3]

While database systems have long enjoyed a "free ride" with ever-increasing clock cycles of the CPU, in the last decade this increase stalled. On the computational side, we have seen an ever-increasing number of cores as well as the advent of specialized computing units ranging from GPUs via FPGA to chips with specific extensions. On the memory side, we not only observe a significant growth of the capacity of main memory, but a continued large performance impact of RAM latency on data access cost, recently aggravated by increasing NUMA effects. Storage-wise we have witnessed the introduction of NAND devices (e.g., SSDs) impacting the established role of magnetic disk drive. These advances taken together impact current database architectures and ask for adjustments, extensions or even a complete re-write in order to establish a scalable, affordable, and flexible foundation for data management systems of the future.

This special issue focuses on conceptual and systems-architecture research related to the exploitation of modern hardware infrastructure for data management tasks. The five papers we finally selected for this special issue all went through a major revision in April–May 2015, and then a minor revision in July–August 2015, before being accepted in October–November 2015. We next present a brief summary of the accepted papers.

✉ Peter Boncz
p.boncz@cwi.nl

✉ Wolfgang Lehner
wolfgang.lehner@tu-dresden.de

✉ Thomas Neumann
neumann@in.tum.de

[1] CWI, Amsterdam, The Netherlands

[2] Technical University of Dresden, Dresden, Germany

[3] Technische Universität München, München, Germany

The paper "Characterization of the Impact of Hardware Islands on OLTP" by Danica Porobic, Ippokratis Pandis, Miguel Branco, Pinar Tozun, and Anastasia Ailamaki of EPFL presents extensive experimentation of controlled computational locality in OLTP. The idea is to recognize the non-uniform memory access cost and also memory consistency management cost (CPU cache consistency protocol) in multi-socket systems, and to partition computation to occur on certain subsets of the multi-cores. So-called hardware islands are formed by these subsets that have low-cost communication and synchronization between the cores. The experimentation is performed on 2-, 4- and 8-socket Intel Xeon machines, and the transcational evaluation is run on both ShoreMT and Silo, showing that such hardware-aware partitioning makes sense.

The paper "Exploiting SSDs in Operational Multi-version Databases" by Mohamman Sadoghi and Kenneth Ross from Columbia University, and Mustafa Canim and Bishwaranjan Bhattacharjee from IBM's TJ Watson Research Center, focuses on improving transactional workloads on multi-version database systems, such as those that allow temporal queries, by integrating SSDs as memory layer between RAM and disk. Specifically, the paper proposes to use SSDs to store an indirection map that translates logical row IDs into physical row IDs, as such a map has only a fraction of the size of B+-tree indexes that contain row IDs. Modifying data on SSD is much cheaper than modifying disk-resident B+-trees, while the additional read cost for traversing this indirection is low thanks to the fast reads of SSDs. Further, the paper contributes a delta format that allows to read the last k versions of a tuples in one disk I/O, while strongly reducing the storage cost of multi-version updates that comes with storing a full new copy of a new tuple version by storing only the changed parts.

🖄 Springer

The paper "Flash as Cache Extension for On-Line Transactional Workloads" by Woon-Hak Kang and Sang-Won Lee of Sungkyunkwan University, and Bongki Moon of Seoul National University introduces FaCE (Flash as Cache Extension). FaCE is a new design that extends a DBMS buffer manager with a flash cache, where its buffer management algorithm not only optimizes for transactional throughput but also improved recovery time, by leveraging the persistent nature of the flash cache.

A third paper on the topic of SSDs is "Read/Write-Optimized Tree Indexing for Solid State Drives" by Peiquan Jin, Chencheng Yang, Christan Jensen, and Puyuan Yang Lihua Yue of the University of Science and Technology of China. This paper introduces the BloomTree, a B+-tree that reduces the amount of node splits caused by updates, and hence reduces random writes and NAND write amplification, by using overflow buffers inside leaf nodes with a bloom filter to keep leaf lookups fast.

Finally, the paper "GPU-accelerated string matching for database applications" by Evangelia Sitaridi and Kenneth Ross of Columbia University takes on the problem of handling string matching tasks in database query processing using GPUs. The authors propose a GPU-optimized Knuth-Morris-Pratt algorithm that splits strings in multiple fixed-sized segments to account for their variable length and stores these segments decomposed and interleaved in smaller fixed-size "pivots" to create opportunity for parallel work. The successful results of these techniques will likely alter current thinking on GPUs and string query predicate processing, which until now has been thought of as an area where GPUs have a hard time to compete with CPUs. While this special issue offers some of the latest developments on data management for modern hardware, this area of work by its nature continues to evolve.

With CPU-based servers crossing into hundreds of cores, and ever widening of their SIMD instructions to 512 bits and beyond, and enhancements to their instruction sets with e.g., scatter/gather on the one hand, and GPU technology becoming more programmable, more adaptive and allowing for multitasking on the other hand; an integration of GPU and CPU in terms of software programming environments as well as hardware is coming to fruitition.

The next decade is also expected to bring the introduction of byte-addressable persistent memories, possibly replacing DRAM, which will force to rethink data and indexing structures, and generally speaking, all methods for consistency and durability in transaction processing.

Therefore, we think new hardware architectures in the next decade are poised to continue to change in significant ways the computational properties in which data management software operates and thus will continue to be a productive research area.