DFG Priority Program SPP 2037: Scalable Data Management for Future Hardware

Kai-Uwe Sattler[1] Alfons Kemper[2] Thomas Neumann[2] Jens Teubner[3]

Abstract: The priority program 2037 "Scalable Data Management for Future Hardware" is funded by the DFG and comprises 10 projects from german universities. The program is based on the observation that the currently used database concepts and systems are not well prepared to support emerging application domains. At the same time current and future hardware trends provide new opportunities. In the following we give an overview of the overall goals and the projects funded within the program.

1 Program Goals and Structure

Over the past thirty years, database management systems have been established as one of the most successful software concepts. In today's business environment they constitute the centerpiece of almost all critical IT systems. The reasons for this success are manyfold. On the one hand, such systems provide abstractions hiding the details of underlying hardware or operating systems layers. This covers the existence of a memory hierarchy, memory organization, data representation and efficient data access for multiple users or application developers. On the other hand, database management systems are ACID compliant, which enables them to represent an accurate picture of a real world scenario, and ensures correctness of the managed data even in extreme cases (e.g., a high number of concurrent database operations or possible system failures). Hence, there is a wide acceptance of architectural patterns for database systems which are based on assumptions of classic hardware setups.

Today, the application of database systems has moved beyond pure transaction-oriented scenarios. Instead they are more and more utilized as data integration platforms to realize a unified access model (perhaps limited to read operations) to heterogeneous or even distributed data. In addition, database technology in a broader sense is exploited in pure analytical applications (e.g., building models for data mining algorithms such as classification, clustering, recommendation, etc.). These analyses are based on clickstream data or experimental results in the scientific environment (e.g., protein analyses in micro biology, or galaxy detection in astro-physical research projects). For such applications, the rigid transaction-oriented architecture of classical database systems often proves to be too

[1] TU Ilmenau, kus@tu-ilmenau.de
[2] TU Munich, *lastname*@in.tum.de
[3] TU Dortmund, jens.teubner@cs.tu-dortmund.de

doi:10.18420/btw2019-ws-29

rigid, inflexible and not scalable to the required extent. During the consequently emerging diversification of data management solutions some of the well established functionalities of classical database systems fell by the wayside: For instance, consistency in eventually consistent system has to be realized at the application level (e.g. using versioning). At the same time, current and future hardware trends provide new opportunities such as:

- many-core CPUs: Next-generation CPUs will provide hundreds of compute cores already in the commodity range. In order to allow high degrees of parallelism some architectures already provide hardware support for the necessary synchronization, e.g. transactional memory. Utilizing this parallelism for database processing is still an open issue.
- co-processors like GPUs and FPGAs: Special-purpose computing units such as GPUs and FPGAs allow for parallelism at much higher degrees accelerating compute-intensive tasks significantly – even for database tasks. Moreover, heterogeneous hardware designs such as coupled CPU-FPGA and CPU-GPU architectures represent a trend of close integration between classic hardware and emerging hardware which could be beneficial in data management.
- novel storage technologies like NVRAM and SSD: Modern in-memory database system solutions still rely mostly on block-based media for ensuring persistence of data. Emerging memory technologies such as non-volatile memory (NVRAM) promise byte-addressable persistence with latencies close to DRAM requiring to revisit memory and storage hierarchies in data management systems.
- high-speed networks: Both, in scale-up and scale-out scenarios efficient interconnects play a crucial role. Today, some network technologies based on Gbit Ethernet or InfiniBand already support Remote DMA, i.e., direct access to memory of a remote node. But, to utilize this technology in database systems requires new concepts.

The goals of the DFG priority program on Scalable Data Management for Future Hardware are based on the observation that data management architectures will undergo a radical shift in the next years. This is driven by the fact that on the one hand, the range of applications requiring to handle large sets of data has significantly broadened, in particular those based on analytics/machine learning, and on the other hand, new trends in hardware as well as at operating system level offer great opportunities for rethinking current system architectures. Thus, the leitmotif of the first phase of the priority program can be formulated as “Scalability beyond current limits”.

The program is coordinated by Kai-Uwe Sattler (TU Ilmenau), Jens Teubner (TU Dortmund), Alfons Kemper (TU Munich), and Thomas Neumann (TU Munich). The coordinators are supported by an advisory board with highly experienced experts both from industry and academia: Peter Boncz (CWI/VU Amsterdam), Franz Färber (SAP SE), Goetz Graefe (Google, Madison), Theo Härder (TU Kaiserslautern), and Wolfgang Lehner (TU Dresden).

2 Funded Projects

The first phase of the priority program started in summer 2017 with a kickoff meeting at the VLDB in Munich. During this phase, 10 projects in the areas **Networking and RDMA**, **FPGA & Many-core CPUs**, **Memory and Storage**, and **SGX** are funded involving 22 researchers and 18 PIs. In detail, the following projects are part of the priority program.

Scalable Data Management in the Presence of High-Speed Networks (TU Darmstadt). The goal of this project is to develop abstractions for remote direct memory access (RDMA) in distributed databases. These abstractions shall support a wide range of different workloads ranging from traditional applications (OLAP and OLTP) to more complex workloads (e.g., machine learning). The experimental evaluation will address large-scale deployments with up to 100 nodes.

Distributed, fault-tolerant in-place consensus sequence on innovative hardware as a building block for data management (ZIB Berlin). This project deals also with RDMA technology in combination with NVRAM to manage distributed shared states. The main goal of this project is to extend the Paxos protocol to support a sequence of consensus decisions in-place.

Interactive Big Data Exploration on Modern Hardware (TU Munich). This project aims to provide interactive response times for database queries even on big data. This is achieved by deeply integrating big data exploration functionality into the core of the system. In particular, latest and emerging hardware trends to scale database systems as well as techniques such as query compilation and micro adaptivity are exploited.

Query Compilation for the Heterogeneous Many Core Age (TU Berlin). The goal of this project is to allow database systems to adapt themselves automatically to heterogeneous, previously unknown processors, and in this way, avoiding manual per-processor tuning. This is achieved by introducing variations to database operators (e.g., code optimizations, data structures and parallelization strategies), which allows to generate custom implementations of database operators for each processor.

ReProVide: Query Optimisation and Near-Data Processing on Reconfigurable SoCs for Big Data Analysis (University Erlangen-Nuremberg). In this project, a novel FPGA-based System-on-Chip (SoC) architecture called ReProVide (Reconfigurable Data ProVider) for near-data processing will be designed and developed. The goal is to provide query-specific accelerator datapaths and filter functions on-demand by exploiting the fact that the hardware of an FPGA may be dynamically reconfigured.

Adaptive Data Management in Evolving Heterogeneous Hardware/Software Systems (Uni Magdeburg). This project aims to develop integration concepts for diverse operators and heterogeneous hardware devices in adaptive database systems. In particular, optimization strategies for exploiting individual device-specific features but also the inherent cross-device parallelism in multi-device systems are investigated. The complexity of the query optimization design space incurred by the parallelism is handled by a distributed optimization approach as well as a set on cross-layer optimizations strategies incorporating learning-based techniques.

MxKernel: A Bare-Metal Runtime System for Database Operations on Heterogeneous Many-Core Hardware (TU Dortmund). As part of this projecz a bare-metal runtime system called MxKernel is developed. MxKernel provides very lightweight resource management for database system. For this purpose, heterogeneity and parallelism become first-class citizens. This is achieved by an abstraction for work items called MxTask which represents a unit of work for which atomic execution is guaranteed.

High-Performance Event Processing on Modern Hardware (Uni Marburg). This project deals with low latency requirements of Complex Event Processing (CEP) and high-throughput analysis of event data. The main goals are to develop new indexing techniques for CEP and analysis of historical event analysis exploiting modern storage technologies such as SSDs. In particular, new loading strategies for multiversion indexes and storage layouts for processing queries like pattern matching are considered.

Transactional Stream Processing on Non-Volatile Memory (TU Ilmenau). The project addresses the challenges of transactional stream processing, i.e. a combination of data stream processing with transactional guarantees such as ACID, exactly-once and ordered execution, by exploiting opportunities of modern hardware technology – in particular NVRAM. For this purpose, data structures for managing operator states taking into account the specific properties of NVRAM are developed and evaluated.

Scalable Hardware-Aided Trusted Data Management (TU Braunschweig/University of Applied Sciences Harz). The goal of this project is to exploit recent hardware security technologies, in particular Intel Software Guard Extensions (SGX), for scalable data management. In particular, the project aims to deriving an architecture model for document-based DBMSs with functionality tailored to hardware encryption support (e.g., confidentiality and integrity protection) and security awareness.

3 Activities and Outlook

In addition to the ordinary program meetings and workshops, the members of the priority program have organized several scientific activities, e.g., a Dagstuhl seminar on “Database Architectures for Modern Hardware (Seminar 18251)” [BGH+18] as well as a special issue of the german database jourrnal Datenbank-Spektrum [SKH18].

The first funding period ends in summer 2020. The call for the second phase will be published in September 2019.

References

[BGH+18] Peter A. Boncz, Goetz Graefe, Bingsheng He, Kai-Uwe Sattler: Database Architectures for Modern Hardware (Dagstuhl Seminar 18251). Dagstuhl Reports 8(6): 63-76 (2018).

[SKH18] Kai-Uwe Sattler, Alfons Kemper, Theo Härder: Editorial. Datenbank-Spektrum 18(3): Special Issue on Data Management on New Hardware (2018).