



Introduction to the special issue on self-managing and hardware-optimized database systems 2020

Herodotos Herodotou¹ · Panos K. Chrysanthis² · Shimin Chen³ · Meichun Hsu⁴ · Khuzaima Daudjee⁵ · Yingjun Wu⁶ · Constantinos Costa²

Accepted: 2 April 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

During the past forty years, data management systems have grown in scale, complexity, and variety. There have been novel extensions to relational database management systems as well as fast evolution of big data systems, such as Key-Value stores, Document stores, Graph stores, Spark, MapReduce/Hadoop, Graph Computation Systems, and Data Stream Processing System. At the same time, hardware technology in processors, memory, storage, and networking is undergoing rapid changes, while administration and tuning of these systems has become very expensive, imposing new challenges and creating new opportunities for data management systems.

✉ Herodotos Herodotou
herodotos.herodotou@cut.ac.cy

Panos K. Chrysanthis
panos@cs.pitt.edu

Shimin Chen
chensm@ict.ac.cn

Meichun Hsu
meichun.hsu@oracle.com

Khuzaima Daudjee
kdaudjee@uwaterloo.ca

Yingjun Wu
yingjunwu@singularity-data.com

Constantinos Costa
costa.c@cs.pitt.edu

- ¹ Cyprus University of Technology, Limassol, Cyprus
- ² University of Pittsburgh, Pittsburgh, PA, United States
- ³ Chinese Academy of Sciences, Beijing, China
- ⁴ Oracle, California, United States
- ⁵ University of Waterloo, Waterloo, Canada
- ⁶ Singularity Data Inc., Beijing, China

Two specialized IEEE ICDE (International Conference on Data Engineering) workshops, namely SMDB (International Workshop on Self-Managing Database Systems) and HardBD&Active (Joint International Workshop on Big Data Management on Emerging Hardware and Data Management on Virtualized Active Systems), have provided two forums to examine the above system-related challenges from different angles. The SMDB workshop focuses on providing autonomic or self-* features in database and data management systems to support complex administrative tasks, while the HardBD&Active workshop is interested in exploiting hardware technologies for efficient data management. We are pleased to present the second special issue of DAPD entitled “Self-Managing and Hardware-Optimized Database Systems 2020” that features the best contributions in SMDB 2020 and HardBD&Active 2020 workshops.

In the following, we provide a brief overview of the four papers in this special issue.

In “PatchIndex—Exploiting Approximate Constraints in Distributed Databases”, Kläbe et al. introduce the concept of the PatchIndex structure, which enables database systems to define approximate constraints in order to handle exceptions to given constraints in uncleaned datasets. The authors present parallel approaches for index creation and optimization techniques for parallel queries using PatchIndexes. Finally, they present heuristics for automatically discovering PatchIndex candidate columns and experimentally demonstrate the performance benefits of using PatchIndexes.

In “Self-Adapting Data Migration in the Context of Schema Evolution in NoSQL Databases”, Hillenbrand et al. evaluate and compare different data migration strategies in NoSQL database systems, based on which, legacy data entities are migrated to a newer schema. As data migration strategies exhibit various tradeoffs, the authors explore a methodology of self-adapting data migration, which automatically adjusts migration strategies and their parameters, depending on the migration scenario and service-level agreements.

In “On the Necessity of Explicit Cross-Layer Data Formats in Near-Data Processing Systems”, Weber et al. motivate the necessity for data format pushdown in Near-Data Processing (NDP) systems in order to optimally utilize the hardware properties of the underlying NDP storage and compute elements. To this end, the authors discuss a type hierarchy for NDP operations, including creating and generating dedicated parsers and accessors. The performance benefit of their approach is evaluated on RocksDB and the COSMOS hardware platform.

In “Selective Caching: A Persistent Memory Approach for Multi-Dimensional Index Structures”, Jibril et al. present the Selective Caching technique for exploiting Persistent Memory properties for analytical index structures. The proposed technique is based on a mixture of dynamic and static caching of tree nodes in DRAM to reach near-DRAM access speeds for index structures stored in Persistent Memory. When Selective Caching is used (with a suitable replacement strategy) in the main-memory index structure Elf, the index performance is on par with pure DRAM storage while guaranteeing persistence.

Finally, we would like to thank all the authors who have contributed their papers for this issue. We own our sincere gratitude to the reviewers who contributed to assembling such a high-quality special issue. We are also indebted to the DAPD

Journal Editors, editorial office, and the publishing and production teams for their assistance in preparation and publication of this issue.

1 List of Accepted Papers

PatchIndex - Exploiting Approximate Constraints in Distributed Databases

Steffen Kläbe (TU Ilmenau), Kai-Uwe Sattler (TU Ilmenau), and Stephan Baumann (Actian Germany GmbH)

Self-Adapting Data Migration in the Context of Schema Evolution in NoSQL Databases

Andrea Hillenbrand (Darmstadt University of Applied Sciences), Uta Störl (Darmstadt University of Applied Sciences), Shamil Nabiyeu (Darmstadt University of Applied Sciences), and Meike Klettke (University of Rostock)

On the Necessity of Explicit Cross-Layer Data Formats in Near-Data Processing Systems

Lukas Max Weber (TU Darmstadt), Tobias Vincon (Reutlingen University), Christian Knödler (Reutlingen University), Leonardo Solis-Vasquez (TU Darmstadt), Arthur Bernhardt (Reutlingen University), Iliia Petrov (Reutlingen University), and Andreas Koch (TU Darmstadt)

Selective Caching: A Persistent Memory Approach for Multi-Dimensional Index Structures

Muhammad Attahir Jibril (TU Ilmenau), Philipp Götze (TU Ilmenau), David Broneske (OvG University Magdeburg & Anhalt University of Applied Science), and Kai-Uwe Sattler (TU Ilmenau)

2 Guest editors

Herodotos Herodotou (Lead Editor), Cyprus University of Technology, herodotos.herodotou@cut.ac.cy

Panos K. Chrysanthis, University of Pittsburgh, panos@cs.pitt.edu

Shimin Chen, Chinese Academy of Sciences, chensm@ict.ac.cn

Meichun Hsu, Oracle, meichun.hsu@oracle.com

Khuzaima Daudjee, University of Waterloo, kdaudjee@uwaterloo.ca

Yingjun Wu, Singularity Data Inc., yingjunwu@singularity-data.com

Constantinos Costa, University of Pittsburgh, costa.c@cs.pitt.edu

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.