Guest Editors' Introduction: Special Issue on Big Data Systems on Emerging Architectures

Bingsheng He, *Senior Member, IEEE*, Yunji Chen, *Member, IEEE*, and Jingren Zhou, *Fellow, IEEE*

Index Terms—Big data systems, emerging architectures, GPU, FPGA, high performance computing

BIG data has become a buzz word in recent years. Among various big-data challenges, high performance is a must, not an option. We are facing the challenges (and also opportunities) at all levels ranging from sophisticated algorithms and procedures to mine the gold from massive data to high-performance computing (HPC) techniques and systems to get the value of the data in time. Big data systems have been a fruitful research area, and many systems have been designed and developed for different kinds of big data including relational data, graphs, and data in other forms. Along the journey, open-source systems have been a major driving force in wide adoption of big data systems, for example, Apache Hadoop, Spark, Storm and Flink. The success of those systems has lowered the bar of handling big data, but even led to more prosperous big data applications and ecosystem.

For building high-performance big data systems, there are basically two lines of design principles: *scale out* and *scale up*. The scale-out approach is to aggregate many relatively lower-performance machines to collectively address big data challenges, and the scale-up approach is to equip each single machine with more powerful hardware and infrastructures. They have strength and weakness in their design spaces of performance, energy efficiency and cost. The continued evolution of computing hardware and infrastructure imposes new challenges and opportunities in scaling big data systems in both scale-out and scale-up manners. Over the last few years, there has been a renewed interest in the area of (big) data systems on emerging hardware. Emerging hardware and infrastructures have been researched and even integrated into different scales, from a single machine to thousands of machines. Still, a lot of technical challenges exist. The need for effectively utilizing computing resources creates new technologies and research directions: from conventional ones (e.g., cluster computing, in-memory computing), to more recent ones (e.g., GPGPU, many-core processors, FPGA, and NVRAM). Thus, there is a need to fundamentally address all the above-mentioned issues in big data systems on emerging architectures.

IEEE Transactions on Big Data has recognized those important and timely concerns in this special issue. This special issue aims to enhance the architecture-awareness in current and future big data systems, and to compile research efforts that fundamentally addresses the technical challenges. We received an overwhelming response from the community of researchers from industry and academia, which are from many different countries. Each submission has undergone serious reviews with at least three expert reviewers. Each accepted paper has at least two rounds of reviews. Eventually, we have accepted 7 papers. The accepted papers reflect the diversity of the research areas for big data systems and can be grouped under three different categories, viz., Architecture-aware system design, Big data application frontiers and New architecture design.

The first category (*Architecture-aware system design*) has two papers that focus on accelerating big data processing with emerging hardware. Cai et al. developed in-memory data management system named MemepiC which unifies both online data query and data analytics into a single in-memory system, and the system is further optimized with RDMA awareness. Maziar Goudarzi conducted an extensive survey on how heterogeneous architectures (such as GPUs and FPGAs) can be integrated into batch processing platforms such as MapReduce.

The second category (*Big data application frontiers*) has three papers addressing various technical challenges from new big data applications. Wang et al. proposed an integrated infrastructure to handle huge price data in the smart power grid for price prediction. Liu et al. studied an automated deployment framework for streaming applications in the cloud. The increasing amount of geographically distributed massive data is pushing industries and academia to rethink the current big-data processing systems. Thus, Dolev et al. reviewed extensively on the work of big data processing in geographically distributed environment.

The last category (*New architecture design*) has two papers, covering the exploration of new processor and memory design for big data workloads. Jia et al. conducted comprehensive evaluations using representative data analytics workloads on multi-core and many-core processors, and further analyzed the performance, power, energy efficiency and performance-cost efficiency of different processor designs (including brawny and wimpy designs). Edstrom

- *B. He is with the School of Computing, National University of Singapore, Singapore 119077. E-mail: hebs@comp.nus.edu.sg.*
- *Y. Chen is with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China. E-mail: cyj@ict.ac.cn.*
- *J. Zhou is with Alibaba Group, Hangzhou 311121, China. E-mail: jingren.zhou@alibaba-Inc.com.*

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TBDATA.2019.2898067

et al. developed low-cost self-recovery video storage system by investigating meaningful data patterns hidden in big video data. Interestingly, they introduced data mining techniques to the hardware design process, which leads to a low-cost and efficient SRAM design.

We believe that big data systems on emerging hardware will continue to be an important and challenging research area, as new hardware and new applications will be developed. The accepted papers in this special issue represent some recent efforts towards this endeavour. We call for actions from relevant research communities such as database, architecture and systems to work together to address the research challenges.

ACKNOWLEDGMENTS

The guest editors would like to thank all the researchers (either in the role of an author or a reviewer) who had responded to this call. They also thank the Editor-in-Chief Professor Qiang Yang for supporting this timely topic as a special issue. Finally, they thank the IEEE journal coordinator and administrator Ms. Leigh Ann Testa and Ms. Kathy Santa Maria for their support in the peer review as well as publication of this special issue!

Bingsheng He received the bachelor's degree in computer science from Shanghai Jiao Tong University, 1999-2003, and the PhD degree in computer science from the Hong Kong University of Science and Technology, 2003-2008. He is an associate professor with the School of Computing of National University of Singapore. His research interests are high performance computing, distributed and parallel systems, and database systems. He is a senior member of the IEEE.

Yunji Chen is currently a full professor at Institute of Computing Technology (ICT), Chinese Academy of Sciences. Currently, he leads the Intelligent Processor Research Center of ICT to develop the Cambricon family of deep learning processors. He has authored or coauthored 1 book and more than 100 papers on various conferences (including ISCA, HPCA, MICRO, ASPLOS, ICSE, ISSCC, Hot Chips, IJCAI, FPGA, and SPAA) and jounals (including the *IEEE Journal of Solid-State Circuits*, the *IEEE Transactions on Computers*, the *IEEE Transactions on Parallel and Distributed Systems*, the *IEEE Transactions on Image Processing*, the *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, the *ACM Transactions on Computer Systems*, the *ACM Transactions on Intelligent Systems and Technology*, the *ACM Transactions on Architecture and Code Optimization*, the *ACM Transactions on Design Automation of Electronic Systems*, *ACM Computing Surveys*, and *IEEE Micro*). He is a member of the IEEE.

Jingren Zhou received the PhD degree in computer science from Columbia University. He is vice president at Alibaba Group. He is responsible for driving data intelligence infrastructure and several key data-driven businesses at Alibaba. Specifically, he leads work to develop cloud-scale distributed computing platform, data analytic products, and various business solutions. He also leads work to develop advanced techniques for personalized search, product recommendation, and advertisement at Alibaba's e-commerce platforms, including Taobao and Tmall. His research interests include cloud-computing, databases, and large scale machine learning systems. He is a fellow of the IEEE.