Sun XH, Yu W. Special section on selected I/O technologies for high-performance computing and data analytics. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 35(1): 1–3 Jan. 2020. DOI 10.1007/s11390-020-0001-9

Preface

It is our great pleasure to announce the publication of this special section in JCST, Selected I/O Technologies for High-Performance Computing and Data Analytics.

With the explosive growth of colossal data from various academic and industrial sectors, many High-Performance Computing (HPC) and data analytics systems have been developed to meet the needs of data collection, processing and analysis. Accordingly, many research groups around the world have explored unconventional and cutting-edge ideas for the management of storage and I/O.

For the I/O research community to get a global picture on the current state-of-the-art and vibrant progress, we have taken an initiative to invite submissions from research groups with a reputable I/O track record and proven technologies. After two rounds of thorough review and revisions, we end up with a special section including eight high-quality papers from China, Europe, Japan, and the United States.

Our special section begins with a review paper from an international team that presents the results from the Dagstuhl Seminar 17202 "Challenges and Opportunities of User-Level File Systems for HPC". This paper provides an in-depth discussion on the design strategies for ad hoc file systems using node-local storage media. Using three representative ad hoc file systems BeeOND, GekkoFS, and BurstFS, the authors have discussed various interfaces and semantics issues in the design of user-level file and storage systems to exploit the performance capacity of the underlying devices.

Then we are fortunate to have two papers from China presenting the design and implementation of storage and file systems to enable two different lines of supercomputers: Tianhe-2 and Sunway Taihu Light. A collaborative team consisting of researchers from the National Supercomputer Center and National University of Defense Technology in China describes the first paper from China: "Design and Implementation of the Tianhe-2 Data Storage and Management System". This paper covers the team's effort in addressing data management challenges at three different levels including the user applications, to middleware libraries and the underlying parallel file system, for the same TianHe-2 supercomputer to support distinctive applications on HPC, big data and artificial intelligence. The second paper from China is titled as "Lessons Learned from Optimizing the Sunway Storage System for Higher Application I/O Performance", which covers an effort on the Sunway storage system to pinpoint the performance issues, and accordingly develop optimizations at the I/O forwarding layer and the parallel file system layer to shorten the I/O path, mitigate the I/O interference and performance variations for a highly performant system that supports challenging N-N I/O patterns.

Next we have a paper from the University of Tsukuba in Japan titled as "Gfarm/BB — Gfarm file system for node-local burst buffer". Tatebe *et al.* present their recent effort to enrich their Gfarm file system and integrate node-local storage devices as burst buffers. The team recognizes the temporary nature of burst buffers and pursues their enhancement by optimizing both data and metadata accesses. For the latter, the authors have omitted the persistency and redundancy, a strategy that is often acceptable for temporary file systems. For the former, the team has leveraged remote direct memory access (RDMA) to speed up data read/write accesses.

Enabling high-performance burst buffers is also recognized by a team of German and Spanish researchers in Europe with a paper titled as "GekkoFS — A temporary burst buffer file system for HPC applications". In view of the distinctive requirements between a generic file system and a temporary burst buffer system, the team has developed GekkoFS as a file system that can adapt to the special needs of different application use cases and support scalable I/O performance. Similar to Gfarm/BB, GekkoFS also relaxes the consistency of POSIX I/O operations for scalable and high-performance I/O.

Finally, the special section includes three papers from research teams in the United States. The first paper "I/O Acceleration via Multi-Tiered Data Buffering and Prefetching" is contributed by a team from Illinois Institute of Technology. The paper details their recent I/O framework Hermes that can buffer dynamic I/O on heterogeneous and multi-tiered systems. Hermes features three novel data placement policies that can utilize heterogeneous devices across all layers and efficiently coordinate memory, metadata, and communication management. Hermes also includes a hierarchical data prefetcher that is designed to support server-driven data prefetching. The second U.S. paper is contributed by a team from Argonne National Laboratory. In their paper, Ross *et al.* present their recent efforts on "Mochi: Composing Data Services for High-Performance Computing Environments". Motivated by the diverse needs of HPC applications, the team has developed Mochi as a composition of diverse distributed data services that are each specified for different application's requirement. As part of the Mochi, the team has nicely integrated the concept of microservices for HPC applications. Lastly, but not the least, a team of researchers from Lawrence Berkeley National Laboratory presents their paper on "ExaHDF5: Delivering Efficient Parallel I/O on Exascale Computing Systems". In this paper, Byna *et al.* summarize their recent efforts to enrich Hierarchical Data Format version 5 (HDF5), a popular parallel I/O library for scientific applications. Specifically, the authors have presented a number of new features they have introduced to HDF5, including Virtual Object Layer (VOL), Data Elevator, asynchronous I/O, full-featured single-writer and multiple-reader (Full SWMR), and parallel querying. The authors further elaborate each feature in terms of its design, implementation, and performance benefits for representative HPC applications.

In summary, with its a unique collection of selected I/O technologies around the world, this special section offers a good glimpse of contemporary I/O technologies that have been developed, or still under active development, from various governmental and academic organizations. We hope that a great number of readers and users find this special section interesting and useful for their respective needs and endeavors.

Acknowledgement

We would like to take the opportunity to express our gratitude to all the authors for their contributions, including those whose submissions were not accepted due to various reasons. We are also very thankful to all the reviewers for their valuable time and efforts. We are particularly grateful to the Editor-in-Chief, Prof. Guo-Jie Li, for hosting this special section. We also thank Ms. Feng-Di Shu and the other editorial office staff, for their tireless efforts and extreme patience in coordinating the communication and revisions throughout the entire review process.

Leading Editor

Xian-He Sun, Distinguished Professor, Illinois Institute of Technology, USA, sun@iit.edu

Principal Guest Editor

Weikuan Yu, Professor, Florida State University, USA, yuw@cs.fsu.edu

- Ji-Wu Shu, Tsinghua University, China;
- Jin Xiong, Institute of Computing Technology, Chinese Academy of Sciences, China

Most of these board members are associate editors of JCST. Their effort and experience ensured the quality and success of this special section.

Xian-He Sun is a University Distinguished Professor of Computer Science at the Department of Computer Science in the Illinois Institute of Technology (IIT), Chicago. Before joining IIT, he worked at DoE Ames National Laboratory, at ICASE, NASA Langley Research Center, at Louisiana State University, Baton Rouge, and was an ASEE fellow at Navy Research Laboratories. Dr. Sun is an IEEE Fellow and is known for his memory-bounded speedup model, also called Sun-Ni's Law, for scalable computing. His research interests include data-intensive high-performance computing, memory and I/O systems, software system for big data applications, and performance evaluation and optimization. He has over 250 publications and 6 patents in these areas. He is the Associate Chief Editor of IEEE Transactions on Parallel and Distributed Systems, a Golden Core member of the IEEE CS society, and the past chair of the Computer Science Department at IIT. More information about Dr. Sun can be found at www.cs.iit.edu/∼sun/.

Weikuan Yu is a Professor in the Department of Computer Science at Florida State University, Tallahassee. He earned both his Ph.D. degree in computer science and Master's degree in neurobiology from the Ohio State University, Columbus. He also holds a Bachelor's degree in Genetics from Wuhan University, Wuhan. Yu's main research interests include big data management and analytics frameworks, parallel I/O and storage, GPU memory architecture, and high-performance networking. He has published more than 90 papers, many of which appeared in top conferences and journals such as IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Computers, Supercomputing, SigMetrics, PACT, ICS and IPDPS. Yu is currently serving as an Associate Editor for the IEEE Transactions on Parallel and Distributed Systems. He is a senior member of IEEE and a member of ACM, USENIX and AAAS.