2022 IEEE/ACM 8th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD) | 978-1-6654-6337-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/DRBSD56682.2022.00004

Message from the DRBSD-8 Workshop Chairs

With an explosive growth of scientific datasets in diverse use-cases, efficient data reduction and analysis techniques are critical to the modern high performance computing (HPC) applications. For instance, an ever-growing disparity between scientific simulation execution speeds and data movement rates (such as I/O and communication) makes it increasingly infeasible for HPC applications to save all results for offline analysis. By 2024, computers are expected to compute at 10^{18} ops/sec but write to disk only at 10^{12} bytes/sec: a compute-to-output ratio 200 times worse than on the first petascale systems. In this new world, applications must increasingly perform online data analysis and reduction—tasks that introduce algorithmic, implementation, and programming model challenges that are unfamiliar to many scientists and that have major implications for the design of various elements of exascale systems.

This trend has spurred interest in high-performance online data analysis and reduction methods, motivated by a desire to conserve I/O bandwidth, storage, and/or power; increase accuracy of data analysis results; and/or make optimal use of parallel platforms, among other factors. This requires our community to understand a clear yet complex relationship between application design, data analysis and reduction methods, programming models, system software, hardware, and other elements of a next-generation High Performance Computer, particularly given constraints such as applicability, fidelity, performance portability, and power efficiency.

The goal of this workshop is to provide a focused venue for researchers in all aspects of data reduction and analysis to present their research results, exchange ideas, identify new research directions, and foster new collaborations within the community. The major topics that our community is striving to answer include (1) whether several orders of magnitude of data reduction is possible for exascale sciences; (2) understanding the performance and accuracy trade-off of data reduction; and (3) solutions to effectively reduce data while preserving the information hidden in large scientific data. Tackling these challenges requires expertise from computer science, mathematics, and application domains to study the problem holistically, and develop solutions and hardened software tools that can be used by production applications.

This year, we are excited to accept 8 high-quality papers, which address a wide range of data reduction issues for different scientific applications and use cases. Specifically, the research topics include portability issues (optimizing the data reduction operation in a heterogeneous environment), performance issues (characterizing the data reduction quality and the performance for additive manufacturing and data deduplication systems), data fidelity issues (analyzing the impact of data reduction on climate models and derivative-related metrics), and data reduction technology issues (estimating the compressibility of transform-based compression).

Jieyang Chen
Oak Ridge National Laboratory, USA
DRBSD-8

Sheng Di
Argonne National Laboratory, USA
DRBSD-8

Ana Gainaru
Oak Ridge National Laboratory, USA
DRBSD-8

Xin Liang
University of Kentucky, USA
DRBSD-8

Todd Munson
Argonne National Laboratory, USA
DRBSD-8