

Call for papers: PDSW'20

The 5th International Parallel Data Systems Workshop

<http://www.pdsw.org/>

(Virtual Workshop) Thursday, November 12, 2020 10:00am - 6:30pm (EST)

Held in conjunction with SC20

In cooperation with: IEEE TCHPC

General Chair

Philip Carns, Argonne National Laboratory

Program Chair

Shadi Ibrahim, Inria, France

Publicity Chair

Thomas Lambert, Inria, France

Program Vice-Chair

Kento Sato, Riken, Japan

Web and Proceedings Chair

Joan Digney, Carnegie Mellon University

Reproducibility Co-Chairs

Carlos Maltzahn, University of California, Santa Cruz

Ivo Jimenez, University of California, Santa Cruz

Important Dates

Regular Papers and Reproducibility Study Papers

Submissions due: Sep. 6, 2020, 11:59 PM AoE

Paper Notification: Sep. 28, 2020

Pre-recorded presentation due: Oct. 7, 2020, 11:59 PM AoE

Camera ready due: Oct. 18, 2020, 11:59 PM AoE

Work in Progress (WIP)

Submissions due: Sep. 27, 2020, 11:59 PM AoE

WIP Notification: On or before Oct. 1, 2020

Pre-recorded presentation due: Oct. 7, 2020

Workshop Abstract:

We are pleased to announce the 5th International Parallel Data Systems Workshop (PDSW'20). PDSW'20 will be hosted in conjunction with SC20: The International Conference for High Performance Computing, Networking, Storage and Analysis.

Efficient data storage and data management are crucial to scientific productivity in both traditional simulation-oriented HPC environments and Big Data analysis environments. This issue is further exacerbated by the growing volume of experimental and observational data, the widening gap between the performance of computational hardware and storage hardware, and the emergence of new data-driven algorithms in machine learning.

The goal of this workshop is to facilitate research that addresses the most critical challenges in scientific data storage and data processing. We therefore encourage the community to submit original manuscripts that:

- introduce and evaluate novel algorithms or architectures,
- inform the community of important scientific case studies or workloads, or
- validate the reproducibility of previously published work

Special attention will be given to issues in which community collaboration is crucial for problem identification, workload capture, solution interoperability, standardization, and shared tools. We also strongly encourage papers to share complete experimental environment information (software version numbers, benchmark configurations, etc.) to facilitate collaboration.

Topics of interest include the following:

- Scalable architectures for data storage, archival, and virtualization
- Performance benchmarking, resource management, and workload studies
- Programmability of storage systems
- Parallel file systems, metadata management, and complex data management
- Alternative data storage models, including object stores and key-value stores
- Programming models and frameworks for data intensive computing
- Techniques for data integrity, availability, reliability, and fault tolerance
- Productivity tools for data intensive computing, data mining, and knowledge discovery
- Application of emerging big data frameworks towards scientific computing and analysis
- Enabling cloud and container-based models for scientific data analysis
- Data filtering/compressing/reduction techniques
- Tools and techniques for managing data movement among compute and data intensive components
- Integrating computation into the memory and storage hierarchy to facilitate in-situ and in-transit data processing

Regular Paper Submissions

All papers will be evaluated by a competitive peer review process under the supervision of the workshop program committee. Selected papers and associated talk slides will be made available on the workshop web site. The papers will also be published by the IEEE TCHPC.

Authors of regular papers are strongly encouraged to submit Artifact Description (AD) Appendices that can help to reproduce and validate their experimental results. While the inclusion of the AD Appendices is optional for PDSW'20, submissions that are accompanied by AD Appendices will be given favorable consideration for the *PDSW Best Paper award*.

PDSW'20 follows the [SC20 reproducibility and transparency initiative](#). For Artifact Description (AD) Appendices, we will use the format of the SC20 for PDSW'20 submissions. The AD should include a field for one or more links to data (zenodo, figshare, etc.) and code (github, gitlab, bitbucket, etc.) repositories. For the Artifacts that will be placed in the code repository, we encourage authors to follow the [PDSW'20 Artifact Packaging Guidelines](#) on how to structure the artifact, as it will make it easier to the reviewing committee and readers of the paper in the future.

Submit a not previously published paper as a PDF file, indicate authors and affiliations. Papers must be up to 5 pages, not less than 10 point font and not including references and optional reproducibility appendices. Papers must use the IEEE conference paper template available at:

<https://www.ieee.org/conferences/publishing/templates.html>

Work-in-progress (WIP) Submissions

There will be a WIP session where presenters provide brief 5-minute talks on their on-going work, with fresh problems/solutions. WIP content is typically material that may not be mature or complete enough for a full paper submission and will not be included in the proceedings. Submit a one-page abstract or a 5-minute pre-recorded presentation.