

## Call for papers: PDSW'19

### The 4th International Parallel Data Systems Workshop

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

Monday, November 18, 2019 9:00am - 5:30pm  
SC'19 Workshop  
Denver, CO

#### General Chair

Suzanne McIntosh, New York University

#### Publicity Chair

Adrian Jackson, University of Edinburgh

#### Web and Proceedings Chair

Joan Digney, Carnegie Mellon University

#### Program Co-Chairs

Glenn K. Lockwood, Lawrence Berkeley National Laboratory  
Philip Carns, Argonne National Laboratory

#### Reproducibility Co-Chairs

Carlos Maltzahn, University of California, Santa Cruz  
Ivo Jimenez, University of California, Santa Cruz

#### Works in Progress Chair

Jay Lofstead, Sandia National Laboratories

### Important Dates

#### Regular Papers and Reproducibility Study Papers

Submissions due: Sep. 1, 2019, 11:59 PM AoE  
Paper Notification: Sep. 29, 2019  
Camera ready due: Oct. 11, 2019, 11:59 PM AoE  
Slides due: Nov. 10, 2019, 11:59 PM AoE

#### Work in Progress (WIP)

Submissions due: Nov. 3, 2019, 11:59 PM AoE  
WIP Notification: Nov. 10, 2019

#### Workshop Abstract:

We are pleased to announce the 4th International Parallel Data Systems Workshop (PDSW'19). PDSW'19 will be hosted in conjunction with SC19: 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 are also strongly encouraged to automate the reproducibility and validation of their experimental results. Submissions that are accompanied by URLs to resources that allow reviewers to repeat automatic validation will be given favorable consideration for the PDSW Best Paper award. The PDSW reproducibility initiative ([bit.ly/pdsw-automatic](http://bit.ly/pdsw-automatic)) will do their best to provide infrastructure and resources to support automated validation. PDSW reviewers, while appreciative, might not be able to validate non-automated artifact descriptions and evaluations included in (optional) reproducibility appendices.

Detailed information on the PDSW reproducibility initiative will be available on the workshop website on July 1, 2019.

Submit a not previously published paper as a PDF file, indicate authors and affiliations. Papers must be between 6 and 10 pages long including references, but not including 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. A one-page abstract is required.