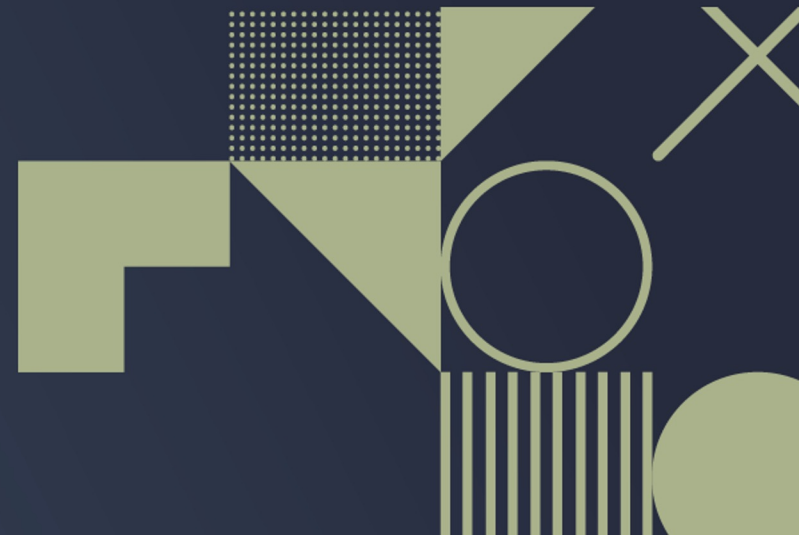




PDSW 2021

6th International Parallel Data Systems Workshop

Shadi Ibrahim, General Chair
Kento Sato, Program Co-Chair
Amelie Chi Zhou, Program Co-Chair



Welcome!

- ▶ *The goal of PDSW is to facilitate research that addresses the most critical challenges in scientific data storage and data processing.*
- ▶ PDSW brings together experts from several overlapping communities:
 - ▶ HPC
 - ▶ Big data
 - ▶ Analytics
- ▶ PDSW 21 is a hybrid event.

We continue to build on the tradition established by previous Data Intensive Scalable Computing and Parallel Data Storage workshops dating back to 2006.



PDSW 2021: 6th International Parallel Data Systems Workshop
<https://www.pdsw.org>

Program highlights

(Full program at [pdsw.org](https://www.pdsw.org))

- ▶ Invited talks
 - ▶ *Marcos K. Aguilera is a principal researcher at VMware*
 - ▶ *Alex Davies is a production engineering team leader at Jump Trading*
- ▶ Technical presentations
 - ▶ 7 full paper presentations
 - ▶ 5 work in progress (WIP) presentations
- ▶ Networking space (in Zoom meeting) is available for further discussion
 - ▶ During Morning coffee break, Lunch break and Afternoon coffee break
 - ▶ After the PDSW workshop (~5:30pm)
 - ▶ Please go to the PDSW page and click "Zoom link" to join

10:00am - 10:30am **Morning Coffee Break**
Virtual Chat Time: Live discussion open to all Workshop Attendees
Zoom link



PDSW 2021: 6th International Parallel Data Systems Workshop
<https://www.pdsw.org>

This wouldn't be possible without the combined efforts of this year's workshop team:

Jay Lofstead:

Steering Committee Chair

Dean Hildebrand:

Steering Committee Vice Chair

Shadi Ibrahim:

General Chair

Kento Sato:

Program Co-Chair

Amelie Chi Zhou:

Program Co-Chair

Carlos Maltzahn and Alexandru Uta:

Reproducibility Co-Chairs

Thomas Lambert:

Publicity Chair

Joan Digney:

Web and Publications Chair

And of course: thank you to everyone who contributed research papers and WIP presentations for sharing your work with the community!



PDSW 2021: 6th International Parallel Data Systems Workshop

<https://www.pdsw.org>

We also owe a big thanks to the program committee:

These subject matter experts are not just gatekeepers for PDSW. They provide constructive guidance to our community to help make our research stronger.

- ▶ Olivier Beaumont, Inria, France
- ▶ Jalil Boukhobza, University of Western Brittany, France
- ▶ Suren Byna, Lawrence Berkeley National Laboratory, USA
- ▶ Yong Chen, Texas Tech University, USA
- ▶ Hank Childs, University of Oregon, USA
- ▶ Matthieu Dorier, Argonne National Laboratory (ANL), USA
- ▶ Bogdan Ghit, Databricks, USA
- ▶ Bingsheng He, National University of Singapore, Singapore
- ▶ Youngjae Kim, Sogang University, Korea
- ▶ Johann Lombardi, Intel Corporation, USA
- ▶ Xiaoyi Lu, University of California, Merced, USA
- ▶ Xiaosong Ma, Qatar Computing Research Institute, Qatar
- ▶ Kathryn Mohror, Lawrence Livermore National Laboratory, USA
- ▶ Diana Moise, Hewlett Packard Enterprise, USA
- ▶ Sarah Neuwirth, Habilitation Candidate at Goethe University
- ▶ M. Mustafa Rafique, Rochester Institute of Technology, Germany
- ▶ Raghunath Raja Chandrasekar, Enfabrica, USA
- ▶ Michael Schöttner, Duesseldorf University, Germany
- ▶ Tong Shue, Southern Illinois University Carbondale, USA
- ▶ Domenico Talia, University of Calabria, Italy
- ▶ Vasily Tarasov, IBM Corporation, USA
- ▶ Osamu Tatebe, University of Tsukuba, Japan
- ▶ Bing Xie, Oakridge National Laboratory, USA
- ▶ Haiying Xu, NCAR, USA

We are also grateful to

Anshu Dubey and Mohamed Wahib:

Leah Glick and Taylor Carr:

Manish Parashar and Siva Rajamanickam:

Patrick Kellenberger:

Workshop Chair

Support team behind Linklings

IEEE TCHPC committee

IEEE Computer Society



Association for
Computing Machinery



PDSW 2021: 6th International Parallel Data Systems Workshop
<https://www.pdsw.org>

Logistics for Speakers and Audience

- ▶ PDSW 2021 includes 6 sessions: Invited talk, Invited Industry Talk, 4 sessions for paper presentations and work in progress (WIP) presentation
 - ▶ Program: <http://www.pdsw.org/index.shtml>
- ▶ Talk's Organization
 - ▶ Presentations:
 - ▶ On-site presenters give live presentations
 - ▶ For remote presenters, pre-recorded videos are played
 - ▶ Live Q&A session
 - ▶ Both remote and on-site attendees can ask question via Sli.do
 - ▶ The session chair ask questions on behalf of attendees
 - ▶ Please note that there will be no microphones at the convention center for in person questions! QR codes is available to ask questions on site
 - ▶ **Please note that this workshop is all recorded**

Best Paper

- ▶ We accepted 7 full workshop papers and 5 WiP talks
- ▶ During the PC meeting, we also selected the best paper addition to paper acceptance discussion
- ▶ This paper presents a preliminary evaluation of the VAST storage system which incorporates NVRAM and QLC flash storage and discuss effective benchmarking methods to properly characterize such new storage architectures with excellent artifact description
- ▶ In reviewing process, we recognized their evaluations and discussion on benchmarking methods are outstanding and this paper received the highest review score

To respect their contributions to the communities,
we would like to present the best paper award to ...



PDSW 2021: 6th International Parallel Data Systems Workshop
<https://www.pdsw.org>

Best Paper

PDSW 2021:
6th International Parallel Data Systems Workshop

Best Paper Award

Presented to
Glenn K. Lockwood, Alberto Chiusole, Nicholas J. Wright
Lawrence Berkeley National Laboratory

For their paper entitled
New Challenges of Benchmarking All-Flash Storage for HPC

Congratulations !!

10:30am - 10:55am
HPC

(Best Paper presentation) New Challenges of Benchmarking All-Flash Storage for

Glenn K. Lockwood, Alberto Chiusole, Nicholas J. Wright



PDSW 2021: 6th International Parallel Data Systems Workshop
<https://www.pdsw.org>

Invited Talk

Session chair
Amelie Chi Zhou

6th International Parallel Data Systems Workshop

Invited Talk

Marcos K. Aguilera

Principal Researcher at VMware

Why Memory Is Your Next Bottleneck
And How To Overcome It



PDSW 2021

6th International Parallel Data Systems Workshop

Shadi Ibrahim, General Chair Kento Sato, Program Co-Chair Amelie Chi Zhou, Program Co-Chair

10:30am - 11:45am Session 2

Session Chair: Bing Xie, Oakridge National Laboratory



10:30am - 10:55am **(Best Paper)** New Challenges of Benchmarking All-Flash Storage for HPC
Glenn K. Lockwood, Alberto Chiusole, Nicholas J. Wright

10:55am - 11:20am Understanding the I/O Impact on the Performance of High-Throughput Molecular Docking
Stefano Markidis, Davide Gadioli, Emanuele Gadioli, Gianluca Palermo

11:20am - 11:45am I/O Bottleneck Detection and Tuning: Connecting the Dots using Interactive Log Analysis
Jean Luca Bez, Houjun Tang, Bing Xie, David Williams-Young, Rob Latham, Rob Ross, Sarp Oral, Suren Byna



PDSW 2021

6th International Parallel Data Systems Workshop

Shadi Ibrahim, General Chair Kento Sato, Program Co-Chair Amelie Chi Zhou, Program Co-Chair

11:45am - 12:30pm **Session 3**
Session Chair: Haiying Xu, NCAR



11:45am - 12:10pm **Data-Aware Storage Tiering for Deep Learning**
Cong Xu, Suparna Bhattacharya, Martin Foltin, Suren Byna, Paolo Faraboschi

12:10pm - 12:20pm **(WiP)** pMEMCPY: A Simple, Lightweight, and Portable
I/O Library for Storing Data in Persistent Memory
Luke Logan

12:20pm - 12:30pm **(WiP)** Optimising I/O using Non-Volatile Memory
Adrian Jackson



PDSW 2021

6th International Parallel Data Systems Workshop

Shadi Ibrahim, General Chair Kento Sato, Program Co-Chair Amelie Chi Zhou, Program Co-Chair

2:00pm - 2:30pm Session 4

Session Chair: Dean Hildebrand, Google



Invited Industry Talk

Jump's Archive for the Next Decade

Alex Davies (on-site)

Production engineering team leader at Jump Trading



PDSW 2021

6th International Parallel Data Systems Workshop

Shadi Ibrahim, General Chair Kento Sato, Program Co-Chair Amelie Chi Zhou, Program Co-Chair

2:30pm - 3:00pm Session 5

Session Chair: Jay Lofstead, Sandia National Laboratory



2:30pm - 2:40pm **(WiP)** Network-accelerated Distributed File Systems
Salvatore Di Girolamo

2:40pm - 2:50pm **(WiP)** High-throughput Small File Access for Large-scale Machine Learning Applications
Hiroki Ohtsuji

2:50pm - 3:00pm **(WiP)** Hyperconverged Storage for High Performance Data Analysis in High Energy Physics: A Case of Intel DAOS Deployment
Alexander Moskovsky



PDSW 2021

6th International Parallel Data Systems Workshop

Shadi Ibrahim, General Chair Kento Sato, Program Co-Chair Amelie Chi Zhou, Program Co-Chair

3:30pm - 4:45pm Session 6

Session Chair: Kathryn Mohror,
Lawrence Livermore National Laboratory



3:30pm - 3:55pm SCTuner: An Auto-tuner Addressing Dynamic I/O Needs on
Supercomputer I/O Sub-systems

Houjun Tang, Bing Xie, Suren Byna, Philip Carns, Quincey Koziol, Sudarsun Kannan, Jay Lofstead,
Sarp Oral

3:55pm - 4:20pm User-Centric System Fault Identification Using IO500 Benchmark
Radita Liem, Jay Loftstead

4:20pm - 4:45pm Verifying IO Synchronization from MPI Traces
Sushma Yellapragada (on-site), Chen Wang, Marc Snir

