PDSW 2019
4th International Parallel Data Systems Workshop

Suzanne McIntosh, General Chair

Jay Lofstead, Vice General Chair

Glenn Lockwood, Program Co-chair

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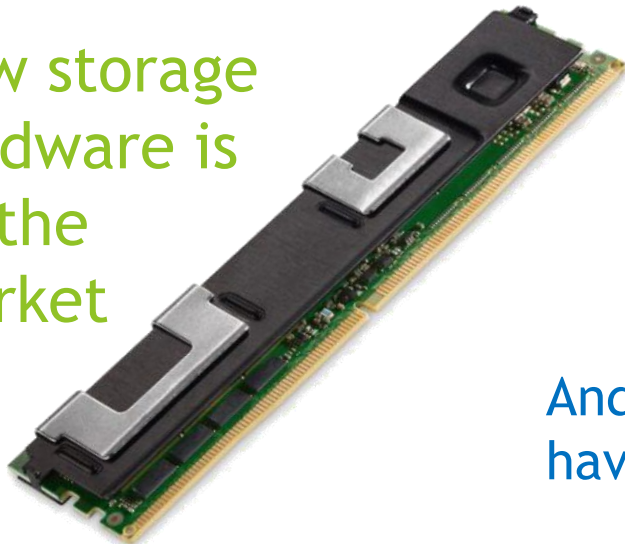
   ▶ Analytics

PDSW 2019: 4th International Parallel Data Systems Workshop
https://www.pdsw.org

Data management is always evolving!
What's new in our field for 2019?

New world's
fastest computer
is in production

New storage
hardware is
on the
market

Machine learning
continues to take
over the world

And much more – just
have a look at today's program!

PDSW 2019: 4th International Parallel Data Systems Workshop
https://www.pdsw.org

Program highlights
(Full program at pdsw.org)

▶ Keynote speech

  ▶ *Haoyuan (H.Y.) Li, founder of Alluxio*

▶ Panel discussion

  ▶ *A House Divided: Why Don't Cloud Storage and HPC Storage Share More Technology?*

▶ 8 full paper presentations (3 sessions)

▶ 12 work in progress (WIP) presentations (3 sessions)

PDSW 2019: 4th International Parallel Data Systems Workshop
https://www.pdsw.org

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

Suzanne McIntosh:	General Chair
Jay Lofstead:	Vice General, WIP Chair
Glenn Lockwood and Philip Carns:	PC Co-Chairs
Carlos Maltzahn and Ivo Jimenez:	Reproducibility Co-Chairs
Adrian Jackson:	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 2019: 4th International Parallel Data Systems Workshop
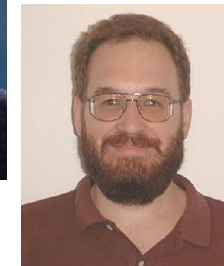https://www.pdsw.org

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.

- Yong Chen, Texas Tech University
- Yue Cheng, George Mason University
- Jason Cope, DDN Storage
- Stratos Efstathiadis, New York University
- Lisa Gerhardt, Lawrence Berkeley National Laboratory
- Elsa J. Gonsiorowski, Lawrence Livermore National Laboratory
- Jian Huang, University of Illinois
- Shadi Ibrahim, French Institute for Research in Computer Science and Automation (INRIA)
- Sidharth Kumar, University of Alabama

- Julian Kunkel, University of Reading
- Johann Lombardi, Intel Corporation
- Xiaoyi Lu, Ohio State University
- Pierre Matri
- Ron Oldfield, Sandia National Laboratories
- Sangmi Pallickara, Colorado State University
- Vasily Tarasov, IBM
- Osamu Tatebe, University of Tsukuba
- Gala Yadgar, Technion - Israel Institute of Technology
- Amelie Chi Zhou, Shenzhen University

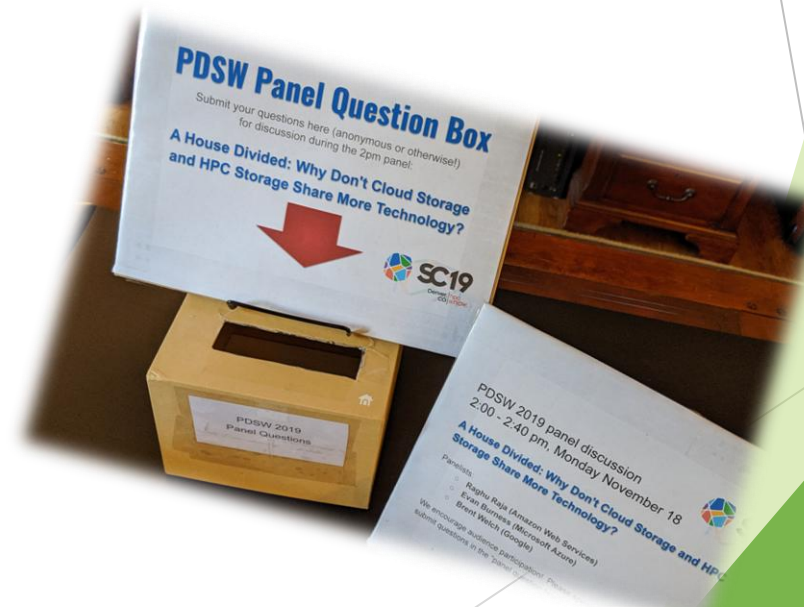**PDSW 2019:** 4th International Parallel Data Systems Workshop
https://www.pdsw.org

Logistics for speakers

▶ All presenters: please check in with your session chair before your presentation

  ▶ 10:30-11:20 papers: Yong Chen

  ▶ 11:40-12:30 papers: Ron Oldfield

  ▶ 3:30-5:10 papers: Brad Settlemyer

  ▶ All WIP papers: Jay Lofstead

▶ Paper presenters: please send a copy of your final presentation to Joan Digney (jdigney@cs.cmu.edu)

▶ WIP presenters: please send your slides to Jay Lofstead **ASAP** (before your talk) if you have not already (gflofst@sandia.gov)

PDSW 2019: 4th International Parallel Data Systems Workshop
https://www.pdsw.org

Please contribute to the panel discussion!

► The title of today's panel is "*A House Divided: Why Don't Cloud Storage and HPC Storage Share More Technology?*"

  ► *Panelists: Brent Welch, Raghu Raja, and Evan Burness*

► What would **you** like to discuss in this topic area?

  ► Questions? Observations? Discussion points? Objections to the premise itself?

► Two ways to contribute:

  ► Ad-hoc by participating in the live discussion

  ► By dropping suggestions in the "question box" at the back of the room by **2pm**

PDSW 2019: 4th International Parallel Data Systems Workshop
https://www.pdsw.org

Program can be found at pdsw.org

▶ 9:00 - 9:10 Welcome

▶ 9:10 - 10:00 Keynote talk by Haoyuan Li, Alluxio

▶ 10:00 - 10:30 Break

▶ 10:30 - 11:20 Paper Session 1
    Session chair: Yong Chen, Texas Tech University

▶ 11:20 – 11:40 WIP Session 1
    Session chair: Jay Lofstead, Sandia National Laboratories

▶ 11:40 – 12:30 Paper Session 2
    Session chair: Ron Oldfield, Sandia National Laboratories

Program can be found at pdsw-discs.org

- 12:30 – 2:00   Lunch (on your own)

- 2:00 - 2:40   Panel*: Brent Welch (Google), Raghu Raja (Amazon),*
  Evan Burness (Microsoft)

- 2:40 – 3:00   WIP session 2
  Session chair: Jay Lofstead, Sandia National Laboratories

- 3:00 – 3:30   Break

- 3:30 – 5:10   Paper Session 4
  Session chair: Brad Settlemyer, Los Alamos National Laboratories

- 5:10 – 5:30   WIP Session 3
  Session chair: Jay Lofstead, Sandia National Laboratories

PDSW 2019: 4th International Parallel Data Systems Workshop
https://www.pdsw.org