PDSW 2020
5th International Parallel Data Systems Workshop

Philip Carns, General Chair
Shadi Ibrahim, Program Chair
Kento Sato, Program Vice 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

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

► PDSW changes in 2020:
  ► The workshop is completely virtual!
  ► We’ve returned to the traditional “short paper” format to facilitate fresh ideas and discussion.

► PDSW constants over time:
  ► A strong community
  ► A positive, interactive culture

► How the audience can help:
  ► Take advantage of opportunities to be interactive.
  ► Give your program chairs a virtual pat on the back for adeptly navigating this year’s challenges!
Program highlights (Full program at pdsw.org)

- Keynote speech
  - Nitin Agrawal is a Principal Engineer at ThoughtSpot

- 8 full paper presentations (3 sessions)
- 3 work in progress (WIP) presentations (1 session)
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
Philip Carns: General Chair
Shadi Ibrahim: Program Chair
Kento Sato: Program Vice Chair
Carlos Maltzahn and Ivo Jimenez: 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!
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
► Raghunath Raja Chandrasekar, Amazon Web Services, USA
► Yong Chen, Texas Tech University, USA
► Yue Cheng, George Mason University, USA
► Jason Cope, DDN Storage, USA
► Toni Cortes, Universitat Politècnica de Catalunya, Spain
► Matthieu Dorier, Argonne National Laboratory, USA
► Lisa Gerhardt, Lawrence Berkeley National Laboratory, USA
► Elsa Gonsiorowski, Lawrence Livermore National Laboratory, USA
► Bingsheng He, National University of Singapore, Singapore
► Johann Lombardi, Intel Corporation, France
► Xiaoyi Lu, Ohio State University, USA
► Xiaosong Ma, Qatar Computing Research Institute, Qatar
► Diana Moise, Hewlett Packard Enterprise, Switzerland
► Anna Queralt, Barcelona Supercomputing Center, Spain
► Brad Settlemyer, Los Alamos National Laboratory, USA
► Xuanhua Shi, Huazhong University of Science and Technology, China
► Vasily Tarasov, IBM Corporation, USA
► Osamu Tatebe, University of Tsukuba, Japan
► Amelie Chi Zhou, Shenzhen University, China
We are also grateful for

φ

Philip C. Roth: SC workshops Chair
Jackie Kern: SC20 Virtual Team
Leah Glick and Taylor Carr: Support team behind Linklings
Manish Parashar and Siva Rajamanickam: IEEE TCHPC committee
Patrick Kellenberger: IEEE Computer Society
Logistics for speakers and Audience

- PDSW 2020 includes 5 sessions: Keynote speech, 3 sessions for paper presentations and 1 session for work in progress (WIP) presentations

- Talk’s Organization:
  - Pre-recorded presentation
  - Live Q&A session
    - The speaker and session chair can use audio and video during q&a
    - Attendees will be asking questions
      - Chat box
      - Question box (this one lets attendees vote too)
Keynote Talk

5th International Parallel Data Systems Workshop
Keynote Talk

Nitin Agrawal
Principal Engineer at ThoughtSpot

Sink or Swim: How Not to Drown in Colossal Streams of Data?