

# PDSW'24 WELCOME & PROGRAM HIGHLIGHTS

Bing Xie, General Chair

Suren Byna, Program Co-Chair

Anthony Kougkas , Program Co-Chair



# WELCOME !

- **The goal of PDSW is to facilitate research and development that addresses the most critical challenges in large-scale data storage and data processing.**
- **In 2016, PDSW (Parallel Data Systems Workshop) was established by combining two predecessor workshops**
  - **The Petascale Data Storage Workshop (PDSW, 2006-2015)**
  - **The Data Intensive Scalable Computing Systems workshop (DISCS 2012-2015)**
- **This joint workshop, PDSW, brings together experts from several overlapping communities from HPC, Big data and Data analytics**
  - **It has been continuing for 9 years from 2016 to 2024**

# WHAT'S NEW IN PDSW 2024 ?

- PDSW2024 extends the scope to align with new technologies:
  - **Cross-cloud data management**
  - Storage system optimization and data analytics with **AI**
  - Innovative techniques and performance evaluation for **new memory and storage systems**

# THIS WOULDN'T BE POSSIBLE WITHOUT THE COMBINED EFFORTS OF THIS YEAR'S WORKSHOP TEAM:

<b>Bing Xie</b>	<b>General Chair</b>
<b>Suren Byna Anthony Kougkas</b>	<b>Program Co-chairs</b>
<b>Jean Luca Bez Radita Liem</b>	<b>Reproducibility Co-chairs</b>
<b>Qian Gong</b>	<b>Publicity Chair</b>
<b>Joan Digney</b>	<b>Web chair</b>
<b>Jay Lofstead</b>	<b>Steering Committee Chair</b>
<b>Dean Hildebrand</b>	<b>Steering Committee Vice-chair</b>

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

# PROGRAM HIGHLIGHTS

## (FULL PROGRAM AT PDSW.ORG)

- **Invited Talk**
  - *Bridging the Data Gaps in Computing for Science, Education and Society*, by **Dr. Ilkay ALTINTAS, University of California, San Diego**
- **Technical presentations**
  - **6 full paper presentations (9 submissions; 67% acceptance rate)**
    - **At least 3 reviews for submitted paper,**
  - **13 work in progress (WIP) presentations (15 submissions; 87% acceptance rate)**
    - **WIP abstracts are reviewed by program co-chairs**
- **Panel discussion: Data, Data Everywhere (4:35 pm to 5:30 pm)**
  - **Moderated by Kathryn Mohror**
  - **Panelists: Laura Biven, Eli Dart, Sarp Oral, Manish Parashar, and Adam Thompson**

# FULL AGENDA

Time slot	Title
9:00 to 9:10	PDSW 2024 Welcome
9:10 to 10:00	Invited Talk: Bridging the Data Gaps in Computing for Science, Education and Society, Ilkay Altıntaş
10:00 to 10:30	<b>Coffee Break</b>
10:30 to 11:00	Fault-Tolerant Deep Learning Cache with Hash Ring for Load Balancing in HPC Systems
11:00 to 11:30	MOSAIC: Detection and Categorization of I/O Patterns in HPC Applications
11:30 to 12:00	Scalable RPC Layer Towards Millions of IOPS per Server
12:00 to 12:30	WIP Session 1 – 6 talks, 5 min each
12:30 to 2:00	<b>Lunch Break</b>
2:00 to 2:30	Initial Experiences with DAOS Object Storage on Aurora
2:30 to 3:00	Understanding and Predicting Cross-Application I/O Interference in HPC Storage Systems
3:00 to 3:30	<b>Coffee Break</b>
3:30 to 4:00	Copper: Cooperative Caching Layer for Scalable Data Loading in Exascale Supercomputers
4:00 to 4:35	WIP Session 1 – 7 talks, 5 min each
4:35 to 5:30	Panel: Data, Data Everywhere, Moderator: Kathryn Mohror Panelists: Laura Biven (Jefferson Lab), Eli Dart (LBNL), Sarp Oral (ORNL), Manish Parashar (University of Utah), and Adam Thompson, NVIDIA



# Proceedings

- <https://conferences.computer.org/sc-wpub>





# BIG THANKS TO THE PROGRAM COMMITTEE

Jean Luca Bez, Lawrence Berkeley National Laboratory

Jalil Boukhobza, ENSTA Bretagne

Wei Der Chien, University of Edinburgh

Dong Dai, University of North Carolina, Charlotte

Qian Gong, Oak Ridge National Laboratory

Luanzheng Guo, Pacific Northwest National Laboratory

Shadi Ibrahim, INRIA

Tanzima Islam, Texas State University

Anthony Kougkas, Illinois Institute of Technology

Quincey Koziol, Amazon Web Services

Michael Kuhn, Otto von Guericke University Magdeburg

Wei-keng Liao, Northwestern University

Johann Lombardi, DAOS Foundation

Xiaoyi Lu, University of California, Merced

Preeti Malakar, Indian Institute of Technology (IIT),  
Kanpur

Sarah M. Neuwirth, Goethe University Frankfurt, JSC

Line Pouchard, Brookhaven National Laboratory

M. Mustafa Rafique, Rochester Institute of Technology

Woong Shin, Oak Ridge National Laboratory

Masahiro Tanaka, Microsoft Corporation

Osamu Tatebe, University of Tsukuba

Chen Wang, Lawrence Livermore National Laboratory

Qing Zheng, Los Alamos National Laboratory



# WE ARE ALSO GRATEFUL TO

- SC24 Workshops organizers
  - Janine C. Bennett, SC24 Workshops Chair
  - Bruno Raffin, Workshops Vice Chair
- Linklings Support Team

# INVITED TALK

**Ilkay Altıntaş**



**Bridging the Data Gaps in Computing  
for Science, Education and Society**

# Closing

# Papers and Slides will be shared on the PDSW webpage

**9am-9:10am PDSW 2024 Welcome**  
Bing Xie, Microsoft

## **INVITED TALK:**

**9:10am- 10am Invited Speaker:**  
**Bridging the Data Gaps in Computing for Science, Education and Society**  
**Dr. Ilkay Altintas**, University of California, San Diego

[Slides](#)

## **MAIN SESSION:**

**10am- 10:30am Morning Break**

**10:30am- 11am Fault-Tolerant Deep Learning Cache with Hash Ring for Load Balancing in HPC Systems**

Seoyeong Lee, Sogang University  
Awais Khan, Oak Ridge National Laboratory (ORNL)  
Yoochan Kim, Sogang University, South Korea  
Junghwan Park, Sogang University, South Korea  
Soon Hwang, Sogang University, South Korea  
Jae-Kook Lee, Korea Inst of Science and Technology Information (KISTI)  
Taeyoung Hong, Korea Inst of Science and Technology Information (KISTI)  
Chris Zimmer, Oak Ridge National Laboratory (ORNL)  
Youngjae Kim, Sogang University, South Korea

[Paper | Slides](#)

**11am- 11:30am MOSAIC: Detection and Categorization of I/O Patterns in HPC Applications**

Théo Jolivel, French Institute for Research in Computer Science and Automation (INRIA)  
François Tessier, INRIA  
Julien Monniot, INRIA  
Guillaume Pallez, INRIA

[Paper | Slides](#)

# Send your feedback

- <https://sc24.conference-program.com/session/?sess=sess736>

## PDSW24: The 9th International Parallel Data Systems Workshop

**Session Chairs:** Dean Hildebrand - Google LLC, Suren Byna - The Ohio State University, Lawrence Berkeley National Laboratory (LBNL), Jay Lofstead - Sandia National Laboratories, University of New Mexico, Anthony Kougkas - Illinois Institute of Technology, Argonne National Laboratory (ANL), Bing Xie - Microsoft Corporation

**Description:** Efficient data storage and data management are crucial to scientific productivity in both traditional simulation-oriented HPC environments and Cloud, AI/ML/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 and development that addresses the most critical challenges in large-scale data storage and data processing. PDSW will continue to build on the successful tradition established by its predecessor workshops: the Petascale Data Storage Workshop (PDSW, 2006-2015) and the Data-Intensive Scalable Computing Systems (DISCS 2012-2015) workshop. These workshops were successfully combined in 2016, and the resulting joint workshop has attracted up to 38 full paper submissions and 195 attendees per year from 2016 to 2023.

Event Type: Workshop

+ Add to Schedule

give feedback

**Time:**

Sunday, 17 November 2024  
9am - 5:30pm EST

**Location:** B309

**Tags:**

Data Movement and Memory | I/O, Storage, Archive

**Registration Categories:**

W

# PDSW 2025

- **See you at PDSW 2025 in St. Louis**