

2016 1st Joint International Workshop on Parallel Data Storage and Data Intensive Scalable Computing Systems (PDSW-DISCS 2016)

**Salt Lake City, Utah, USA
14 November 2016**



**IEEE Catalog Number: CFP16A46-POD
ISBN: 978-1-5090-5217-2**

**Copyright © 2016 by the Institute of Electrical and Electronics Engineers, Inc
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP16A46-POD
ISBN (Print-On-Demand):	978-1-5090-5217-2
ISBN (Online):	978-1-5090-5216-5

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2016 1st Joint International Workshop on Parallel Data Storage and data Intensive Scalable Computing Systems

PDSW-DISCS 2016

Table of Contents

Workshop Organization	v
-----------------------------	---

I/O Insight

Scientific Workflows at DataWarp-Speed: Accelerated Data-Intensive Science	
Using NERSC's Burst Buffer	1
<i>Andrey Ovsyannikov, Melissa Romanus, Brian Van Straalen, Gunther H. Weber, and David Trebotich</i>	
Parallel I/O Characterisation Based on Server-Side Performance Counters	7
<i>Salem El Sayed, Matthias Bolten, Dirk Pleiter, and Wolfgang Frings</i>	
Replicating HPC I/O Workloads with Proxy Applications	13
<i>James Dickson, Steven Wright, Satheesh Maheswaran, Andy Herdman, Mark C. Miller, and Stephen Jarvis</i>	

Data Insight

Can Non-volatile Memory Benefit MapReduce Applications on HPC Clusters?	19
<i>Md. Wasi-ur-Rahman, Nusrat Sharmin Islam, Xiaoyi Lu, and Dhabaleswar K. (DK) Panda</i>	
FatMan vs. LittleBoy: Scaling Up Linear Algebraic Operations in Scale-Out	
Data Platforms	25
<i>Luna Xu, Seung-Hwan Lim, Ali R. Butt, Sreenivas R. Sukumar, and Ramakrishnan Kannan</i>	
Klimatic: A Virtual Data Lake for Harvesting and Distribution of Geospatial	
Data	31
<i>Tyler J. Skluzacek, Kyle Chard, and Ian Foster</i>	

Performance and Testing Insight

Get out of the Way! Applying Compression to Internal Data Structures	37
<i>Rob Latham, Matthieu Dorier, and Rob Ross</i>	
Towards Energy Efficient Data Management in HPC: The Open Ethernet Drive Approach	43
<i>Anthony Kougkas, Anthony Fleck, and Xian-He Sun</i>	
A Generic Framework for Testing Parallel File Systems	49
<i>Jinrui Cao, Simeng Wang, Dong Dai, Mai Zheng, and Yong Chen</i>	
A Bloom Filter Based Scalable Data Integrity Check Tool for Large-Scale Dataset	55
<i>Sisi Xiong, Feiyi Wang, and Qing Cao</i>	
Author Index	61