

# **2021 7th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD-7)**

**St. Louis, Missouri, USA  
14 November 2021**



**IEEE Catalog Number: CFP21W53-POD  
ISBN: 978-1-7281-8673-3**

**Copyright © 2021 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:	CFP21W53-POD
ISBN (Print-On-Demand):	978-1-7281-8673-3
ISBN (Online):	978-1-7281-8672-6

**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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# 2021 7th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD-7) **DRBSD-7 2021**

## Table of Contents

Workshop Organization .....	v
 <b>DRBSD-7 Technical Papers</b>	
Productive and Performant Generic Lossy Data Compression with LibPressio .....	1
<i>Robert Underwood (Clemson University, USA), Victoriana Malvoso (Clemson University, USA), Jon C. Calhoun (Clemson University, USA), Sheng Di (Argonne National Laboratory, USA), and Franck Cappello (Argonne National Laboratory, USA)</i>	
Mitigating Catastrophic Forgetting in Deep Learning in a Streaming Setting Using Historical Summary .....	11
<i>Sajal Dash (Oak Ridge National Laboratory, USA), Junqi Yin (Oak Ridge National Laboratory, USA), Mallikarjun Shankar (Oak Ridge National Laboratory, USA), Feiyi Wang (Oak Ridge National Laboratory, USA), and Wu-chun Feng (Virginia Tech, USA)</i>	
PyParSVD: A Streaming, Distributed and Randomized Singular-Value-Decomposition Library .....	19
<i>Romit Maulik (Argonne National Laboratory) and Gianmarco Mengaldo (National University of Singapore, Singapore)</i>	
Unbalanced Parallel I/O: An Often-Neglected Side Effect of Lossy Scientific Data Compression .....	26
<i>Xinying Wang (University of Nevada), Lipeng Wan (Oak Ridge National Laboratory), Jieyang Chen (Oak Ridge National Laboratory), Qian Gong (Oak Ridge National Laboratory), Ben Whitney (Oak Ridge National Laboratory), Jinzhen Wang (New Jersey Institute of Technology), Ana Gainaru (Oak Ridge National Laboratory), Qing Liu (New Jersey Institute of Technology), Norbert Podhorszki (Oak Ridge National Laboratory), Dongfang Zhao (University of Nevada), Feng Yan (University of Nevada), and Scott Klasky (Oak Ridge National Laboratory)</i>	

TributaryPCA: Distributed, Streaming PCA for In Situ Dimension Reduction with Application to Space Weather Simulations .....	33
<i>Yu Wang (University of Michigan), Natalie Klein (Los Alamos National Laboratory), Steven Morley (Los Alamos National Laboratory), Vania Jordanova (Los Alamos National Laboratory), Michael Henderson (Los Alamos National Laboratory), Ayan Biswas (Los Alamos National Laboratory), and Earl Lawrence (Los Alamos National Laboratory)</i>	
Understanding Effectiveness of Multi-Error-Bounded Lossy Compression for Preserving Ranges of Interest in Scientific Analysis .....	40
<i>Yuanjian Liu (University of Chicago, USA), Sheng Di (Argonne National Laboratory, USA), Kai Zhao (University of California, USA), Sian Jin (Washington State University, USA), Cheng Wang (Argonne National Laboratory, USA), Kyle Chard (University of Chicago, USA), Dingwen Tao (Washington State University, USA), Ian Foster (University of Chicago, USA), and Frank Cappello (Argonne National Laboratory, USA)</i>	
Exploring Lossy Compressibility through Statistical Correlations of Scientific Datasets .....	47
<i>David Krasowska (Holcombe Department of Electrical and Computing Engineering, Clemson University, USA), Julie Bessac (Mathematics and Computer Science Division Argonne National Laboratory, USA), Robert Underwood (Clemson University, USA), Jon C. Calhoun (Holcombe Department of Electrical and Computing Engineering, Clemson University, USA), Sheng Di (Mathematics and Computer Science Division Argonne National Laboratory, USA), and Franck Cappello (Mathematics and Computer Science Division Argonne National Laboratory, USA)</i>	
<b>Author Index .....</b>	<b>55</b>