

2023 IEEE International Conference on Cluster Computing (CLUSTER 2023)

**Santa Fe, New Mexico, USA
31 October – 3 November 2023**



**IEEE Catalog Number: CFP23235-POD
ISBN: 979-8-3503-0793-1**

**Copyright © 2023 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:	CFP23235-POD
ISBN (Print-On-Demand):	979-8-3503-0793-1
ISBN (Online):	979-8-3503-0792-4
ISSN:	1552-5244

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

2023 IEEE International Conference on Cluster Computing (CLUSTER) **CLUSTER 2023**

Table of Contents

Welcome Message from the IEEE Cluster 2023 General Co-Chairs	xi
<i>Scott Pakin (Los Alamos National Laboratory, USA) and Antonio J. Peña (Barcelona Supercomputing Center (BSC), Spain)</i>	
Welcome Message from the IEEE Cluster 2023 Program Chairs	xii
<i>Sunita Chandrasekaran (University of Delaware, USA) and Frank Mueller (North Carolina State University, USA)</i>	
Cluster 2023 Committees	xiv
<i>Scott Pakin (Los Alamos National Laboratory, USA) and Antonio J. Peña (Barcelona Supercomputing Center (BSC), Spain)</i>	

Conference Papers

Accelerating Distributed ML Training via Selective Synchronization	1
<i>Sahil Tyagi (Indiana University, USA) and Martin Swamy (Indiana University, USA)</i>	
PredictDDL: Reusable Workload Performance Prediction for Distributed Deep Learning	13
<i>Kevin Assogba (Rochester Institute of Technology, USA), Eduardo Lima (Rochester Institute of Technology, USA), M. Mustafa Rafique (Rochester Institute of Technology, USA), and Minseok Kwon (Rochester Institute of Technology, USA)</i>	
Exact Distributed Stochastic Block Partitioning	25
<i>Frank Wanye (Virginia Tech, USA), Vitaliy Gleyzer (MIT Lincoln Laboratory, USA), Edward Kao (MIT Lincoln Laboratory, USA), and Wu-chun Feng (Virginia Tech, USA)</i>	
DEHype: Retrofitting Hypervisors for a Resource-Disaggregated Environment	37
<i>Taehoon Kim (Electronics and Telecommunications Research Institute (ETRI), Republic of Korea), Kwangwon Koh (Electronics and Telecommunications Research Institute (ETRI), Republic of Korea), Changdae Kim (Electronics and Telecommunications Research Institute (ETRI), Republic of Korea), Eunji Pak (Electronics and Telecommunications Research Institute (ETRI), Republic of Korea), Yeonjeong Jeong (Electronics and Telecommunications Research Institute (ETRI), Republic of Korea), and Sang-Hoon Kim (Ajou University, Republic of Korea)</i>	

SciLance: Mitigate Load Imbalance for Parallel Scientific Applications in Cloud Environments	49
<i>Xinying Wang (University of Nevada, Reno), Lipeng Wan (Georgia State University), Scott Klasky (Oak Ridge National Laboratory), Dongfang Zhao (University of Washington, Tacoma), and Feng Yan (University of Houston)</i>	
Generalized Collective Algorithms for the Exascale Era	60
<i>Michael Wilkins (Northwestern University), Hanming Wang (Northwestern University), Peizhi Liu (Northwestern University), Bangyen Pham (Northwestern University), Yanfei Guo (Argonne National Laboratory), Rajeev Thakur (Argonne National Laboratory), Peter Dinda (Northwestern University), and Nikos Hardavellas (Northwestern University)</i>	
FEDGUARD: Selective Parameter Aggregation for Poisoning Attack Mitigation in Federated Learning	72
<i>Melvin Chelli (Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarland Informatics Campus, Germany), Cédric Prigent (University of Rennes, Inria, CNRS, IRISA - Rennes, France), René Schubotz (Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarland Informatics Campus, Germany), Alexandru Costan (University of Rennes, Inria, CNRS, IRISA - Rennes, France), Gabriel Antoniu (University of Rennes, Inria, CNRS, IRISA - Rennes, France), Loïc Cudennec (DGA Maîtrise de l'Information, France), and Philipp Slusallek (Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarland Informatics Campus, Germany)</i>	
Prophet: Fine-Grained Load Balancing for Parallel Training of Large-Scale MoE Models	82
<i>Wei Wang (National University of Defense Technology, China), Zhiquan Lai (National University of Defense Technology, China), Shengwei Li (National University of Defense Technology, China), Weijie Liu (National University of Defense Technology, China), Keshi Ge (National University of Defense Technology, China), Yujie Liu (National University of Defense Technology, China), Ao Shen (National University of Defense Technology, China), and Dongsheng Li (National University of Defense Technology, China)</i>	
HIOS: Hierarchical Inter-Operator Scheduler for Real-Time Inference of DAG-Structured Deep Learning Models on Multiple GPUs	95
<i>Turja Kundu (University of North Texas, USA) and Tong Shu (University of North Texas, USA)</i>	
FullRepair: Towards Optimal Repair Pipelining in Erasure-Coded Clustered Storage Systems	107
<i>Yuzuo Zhang (Huazhong University of Science and Technology, China), Xinyuan Tu (Huazhong University of Science and Technology, China), Lin Wang (Huazhong University of Science and Technology, China), Yuchong Hu (Huazhong University of Science and Technology, China), Fang Wang (Huazhong University of Science and Technology, China), and Ye Wang (Huazhong University of Science and Technology, China)</i>	

Performance Characterization of NVMe Flash Devices with Zoned Namespaces (ZNS)	118
<i>Krijn Doekemeijer (Vrije Universiteit Amsterdam, the Netherlands), Nick Tehrani (Vrije Universiteit Amsterdam; Delft University of Technology, the Netherlands), Balakrishnan Chandrasekaran (Vrije Universiteit Amsterdam, the Netherlands), Matias Bjørling (Western Digital, Denmark), and Animesh Trivedi (Vrije Universiteit Amsterdam, the Netherlands)</i>	
KV-CSD: A Hardware-Accelerated Key-Value Store for Data-Intensive Applications	132
<i>Inhyuk Park (SK hynix), Qing Zheng (Los Alamos National Laboratory), Dominic Manno (Los Alamos National Laboratory), Soonyeal Yang (SK hynix), Jason Lee (Los Alamos National Laboratory), David Bonnie (Los Alamos National Laboratory), Bradley Settlemyer (NVIDIA), Youngjae Kim (Sogang University), Woosuk Chung (SK hynix), and Gary Grider (Los Alamos National Laboratory)</i>	
Rethinking Virtual Machines Live Migration for Memory Disaggregation	145
<i>Xingguo Jia (Shanghai Jiao Tong University, China), Xingzi Yu (Shanghai Jiao Tong University, China), Yun Wang (Shanghai Jiao Tong University, China), Senhao Yu (Beijing Institute of Technology, China), and Zhengwei Qi (Shanghai Jiao Tong University, China)</i>	
Efficient Intra-Rack Resource Disaggregation for HPC Using Co-Packaged DWDM Photonics	158
<i>George Michelogiannakis (Lawrence Berkeley National Laboratory, USA), Yehia Arafa (New Mexico State University, USA), Brandon Cook (Lawrence Berkeley National Laboratory, USA), Liang Yuan Dai (Columbia University, USA), Abdel-Hameed Badawy (New Mexico State University, USA), Madeleine Glick (Columbia University, USA), Yuyang Wang (Columbia University, USA), Keren Bergman (Columbia University, USA), and John Shalf (Lawrence Berkeley National Laboratory, USA)</i>	
ExplSched: Maximizing Deep Learning Cluster Efficiency for Exploratory Jobs	173
<i>Hongliang Li (Jilin University; Key Laboratory of Symbolic Computation and Knowledge Engineering of the Ministry of Education, China), Hairui Zhao (Jilin University, China), ZheWen Xu (Jilin University, China), Xiang Li (Jilin University, China), and Haixiao Xu (Jilin University, China)</i>	
Hierarchical Resource Partitioning on Modern GPUs: A Reinforcement Learning Approach	185
<i>Urvij Saroliya (Technical University of Munich, Germany), Eishi Arima (Technical University of Munich, Germany), Dai Liu (Technical University of Munich, Germany), and Martin Schulz (Technical University of Munich, Germany)</i>	
Communication-Avoiding Recursive Aggregation	197
<i>Yihao Sun (Syracuse University, USA), Sidharth Kumar (University of Illinois at Chicago, USA), Thomas Gilray (University of Alabama at Birmingham, USA), and Kristopher Micinski (Syracuse University, USA)</i>	
HASpMV: Heterogeneity-Aware Sparse Matrix-Vector Multiplication on Modern Asymmetric Multicore Processors	209
<i>Wenxuan Li (China University of Petroleum-Beijing, China), Helin Cheng (China University of Petroleum-Beijing, China), Zhengyang Lu (China University of Petroleum-Beijing, China), Yuechen Lu (China University of Petroleum-Beijing, China), and Weifeng Liu (China University of Petroleum-Beijing, China)</i>	

ProvLight: Efficient Workflow Provenance Capture on the Edge-to-Cloud Continuum	221
<i>Daniel Rosendo (University of Rennes, Inria, CNRS, IRISA - Rennes, France), Marta Mattoso (Federal University of Rio de Janeiro, Brazil), Alexandru Costan (University of Rennes, Inria, CNRS, IRISA - Rennes, France), Renan Souza (Oak Ridge National Laboratory, USA), Débora Pina (Federal University of Rio de Janeiro, Brazil), Patrick Valduriez (University of Montpellier, Inria, CNRS, LIRMM - Montpellier, France), and Gabriel Antoniu (University of Rennes, Inria, CNRS, IRISA - Rennes, France)</i>	
Optimizing HPC I/O Performance with Regression Analysis and Ensemble Learning	234
<i>Zhangyu Liu (Northwest University, China), Cheng Zhang (Northwest University, China), Huijun Wu (National University of Defense Technology, China), Jianbin Fang (National University of Defense Technology, China), Lin Peng (National University of Defense Technology, China), Guixin Ye (Northwest University, China), and Zhanyong Tang (Northwest University, China)</i>	
A Lightweight, Effective Compressibility Estimation Method for Error-Bounded Lossy Compression	247
<i>Arkaprabha Ganguli (Argonne National Laboratory, USA), Robert Underwood (Argonne National Laboratory, USA), Julie Bessac (National Renewable Energy Laboratory; Virginia Polytechnic Institute and State University, USA), David Krasowska (Clemson University; Northwestern University, USA), Jon C. Calhoun (Clemson University, USA), Sheng Di (Argonne National Laboratory, USA), and Franck Cappello (Argonne National Laboratory, USA)</i>	
A Dynamic Network-Native MPI Partitioned Aggregation Over InfiniBand Verbs	259
<i>Yiltan Hassan Temuçin (Queen's University, Canada), Scott Levy (Sandia National Laboratories, USA), Whit Schonbein (Sandia National Laboratories, USA), Ryan E. Grant (Queen's University, Canada), and Ahmad Afsahi (Queen's University, Canada)</i>	
DoW-KV: A DPU-Offloaded and Write-Optimized Key-Value Store on Disaggregated Persistent Memory	271
<i>Yiwen Zhang (Huazhong University of Science and Technology, China), Guokuan Li (Huazhong University of Science and Technology, China), Jiguang Wan (Huazhong University of Science and Technology, China), Junyue Wang (Huazhong University of Science and Technology, China), Jun Li (Huazhong University of Science and Technology, China), Ting Yao (Huawei Technologies Co., Ltd, China), Huatao Wu (Huawei Technologies Co., Ltd, China), and Daohui Wang (Huawei Technologies Co., Ltd, China)</i>	
Uniform Algorithms for Reduce-Scatter and (most) Other Collectives for MPI	284
<i>Jesper Larsson Träff (TU Wien, Austria), Sascha Hunold (TU Wien, Austria), Ioannis Vardas (TU Wien, Austria), and Nikolaus Manes Funk (TU Wien, Austria)</i>	

JACO: Java Code Layout Optimizer Enabling Continuous Optimization Without Pausing Application Services	295
<i>Wenhai Lin (Zhejiang University, China), Jingchang Qin (Zhejiang University, China), Yiquan Chen (Zhejiang University; Alibaba Group, China), Zhen Jin (Zhejiang University, China), Jiexiong Xu (Zhejiang University, China), Yuzhong Zhang (Alibaba Group, China), Shishun Cai (Alibaba Group), Lirong Fu (Zhejiang University, China), Yi Chen (Zhejiang University, China), and Wenzhi Chen (Zhejiang University, China)</i>	
A Finite-Difference Time-Domain (FDTD) Solver with Linearly Scalable Performance in an FPGA Cluster	307
<i>Zhenyu Xu (University of Rhode Island, US), Miaoxiang Yu (University of Rhode Island,), Jillian Cai (University of Rhode Island, US), Qing Yang (University of Rhode Island, US), and Tao Wei (University of Rhode Island, US)</i>	
GPU Occupancy Prediction of Deep Learning Models Using Graph Neural Network	318
<i>Hengquan Mei (University of Science and Technology of China, China), Huaizhi Qu (University of Science and Technology of China, China), Jingwei Sun (University of Science and Technology of China, China), Yanjie Gao (Microsoft Research, China), Haoxiang Lin (Microsoft Research, China), and Guangzhong Sun (University of Science and Technology of China, China)</i>	
Reducing Data Motion and Energy Consumption of Geospatial Modeling Applications Using Automated Precision Conversion	330
<i>Qinglei Cao (Saint Louis University; University of Tennessee, USA), Sameh Abdullah (King Abdullah University of Science & Technology, KSA), Hatem Ltaief (King Abdullah University of Science & Technology, KSA), Marc G. Genton (King Abdullah University of Science & Technology, KSA), David Keyes (King Abdullah University of Science & Technology, KSA), and George Bosilca (University of Tennessee, USA)</i>	
SDT: A Low-Cost and Topology-Reconfigurable Testbed for Network Research	343
<i>Zixuan Chen (Fudan University, China), Zhigao Zhao (Fudan University, China), Zijian Li (Fudan University, China), Jiang Shao (Fudan University, China), Sen Liu (Fudan University, China), and Yang Xu (Fudan University; Peng Cheng Laboratory, China)</i>	
PiP-MColl: Process-in-Process-Based Multi-object MPI Collectives	354
<i>Jiajun Huang (University of California, Riverside), Kaiming Ouyang (NVIDIA Corporation), Yujia Zhai (University of California, Riverside), Jinyang Liu (University of California, Riverside), Min Si (Meta Platforms, Inc.), Ken Raffanetti (Argonne National Laboratory), Hui Zhou (Argonne National Laboratory), Atsushi Hori (National Institute of Informatics), Zizhong Chen (University of California, Riverside), Yanfei Guo (Argonne National Laboratory), and Rajeev Thakur (Argonne National Laboratory)</i>	
TopoCommit: A Topological Commit Protocol for Cross-Ledger Transactions in Scientific Computing	365
<i>Olamide Timothy Tawose (Lincoln University, USA), Lei Yang (University of Nevada, USA), and Dongfang Zhao (University of Washington, USA)</i>	

Author Index 377