

USENIX Annual Technical Conference (USENIX ATC'23)

Boston, Massachusetts, USA
10-12 July 2023

Volume 1 of 2

ISBN: 978-1-7138-8395-1

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2023) by Usenix Association
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

For permission requests, please contact Usenix Association
at the address below.

Usenix Association
2560 Ninth Street, Suite 215
Berkeley, California, 94710

<https://www.usenix.org/>

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
Email: curran@proceedings.com
Web: www.proceedings.com

2023 USENIX Annual Technical Conference
July 10–12, 2023
Boston, MA, USA

Monday, July 10

Security and Privacy

- Bifrost: Analysis and Optimization of Network I/O Tax in Confidential Virtual Machines** 1
Dingji Li, *Institute of Parallel and Distributed Systems, SEIEE, Shanghai Jiao Tong University; Engineering Research Center for Domain-specific Operating Systems, Ministry of Education, China; MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University*; Zeyu Mi, Chenhui Ji, Yifan Tan, and Binyu Zang, *Institute of Parallel and Distributed Systems, SEIEE, Shanghai Jiao Tong University; Engineering Research Center for Domain-specific Operating Systems, Ministry of Education, China*; Haibing Guan, *Shanghai Key Laboratory of Scalable Computing and Systems, Shanghai Jiao Tong University*; Haibo Chen, *Institute of Parallel and Distributed Systems, SEIEE, Shanghai Jiao Tong University; Engineering Research Center for Domain-specific Operating Systems, Ministry of Education, China*
- SecretFlow-SPU: A Performant and User-Friendly Framework for Privacy-Preserving Machine Learning**17
Junming Ma, Yancheng Zheng, Jun Feng, Derun Zhao, Haoqi Wu, Wenjing Fang, Jin Tan, Chaofan Yu, Benyu Zhang, and Lei Wang, *Ant Group*
- Portunus: Re-imagining Access Control in Distributed Systems**..... 35
Watson Ladd, *Akamai*; Tanya Verma, *Cloudflare*; Marloes Venema, *University of Wuppertal*; Armando Faz-Hernández, *Cloudflare*; Brendan McMillion; Avani Wildani and Nick Sullivan, *Cloudflare*

Searching Graphs

- GLogS: Interactive Graph Pattern Matching Query At Large Scale** 53
Longbin Lai, *Alibaba Group, China*; Yufan Yang, *The Chinese University of Hong Kong, Shenzhen*; Zhibin Wang, *Nanjing University*; Yuxuan Liu and Haotian Ma, *The Chinese University of Hong Kong, Shenzhen*; Sijie Shen, Bingqing Lyu, Xiaoli Zhou, Wenyuan Yu, and Zhengping Qian, *Alibaba Group, China*; Chen Tian and Sheng Zhong, *Nanjing University*; Yeh-Ching Chung, *The Chinese University of Hong Kong, Shenzhen*; Jingren Zhou, *Alibaba Group, China*
- Cyclosa: Redundancy-Free Graph Pattern Mining via Set Dataflow** 71
Chuangyi Gui, *National Engineering Research Center for Big Data Technology and System/Service Computing Technology and System Lab/Cluster and Grid Computing Lab, Huazhong University of Science and Technology, China*; Zhejiang Lab, *China*; Xiaofei Liao, *National Engineering Research Center for Big Data Technology and System/Service Computing Technology and System Lab/Cluster and Grid Computing Lab, Huazhong University of Science and Technology, China*; Long Zheng, *National Engineering Research Center for Big Data Technology and System/Service Computing Technology and System Lab/Cluster and Grid Computing Lab, Huazhong University of Science and Technology, China*; Zhejiang Lab, *China*; Hai Jin, *National Engineering Research Center for Big Data Technology and System/Service Computing Technology and System Lab/Cluster and Grid Computing Lab, Huazhong University of Science and Technology, China*
- SOWalker: An I/O-Optimized Out-of-Core Graph Processing System for Second-Order Random Walks**..... 87
Yutong Wu, Zhan Shi, Shicai Huang, Zhipeng Tian, Pengwei Zuo, Peng Fang, Fang Wang, and Dan Feng, *Wuhan National Laboratory for Optoelectronics Huazhong University of Science and Technology*

Deduplication

- Light-Dedup: A Light-weight Inline Deduplication Framework for Non-Volatile Memory File Systems**101
Jiansheng Qiu, Yanqi Pan, Wen Xia, Xiaojia Huang, Wenjun Wu, Xiangyu Zou, and Shiyi Li, *Harbin Institute of Technology, Shenzhen*; Yu Hua, *Huazhong University of Science and Technology*
- TiDedup: A New Distributed Deduplication Architecture for Ceph**.....117
Myoungwon Oh and Sungmin Lee, *Samsung Electronics Co.*; Samuel Just, *IBM*; Young Jin Yu and Duck-Ho Bae, *Samsung Electronics Co.*; Sage Weil, *Ceph Foundation*; Sangyeun Cho, *Samsung Electronics Co.*; Heon Y. Yeom, *Seoul National University*

LoopDelta: Embedding Locality-aware Opportunistic Delta Compression in Inline Deduplication for Highly Efficient Data Reduction	133
Yucheng Zhang, <i>School of Mathematics and Computer Sciences, Nanchang University and Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology</i> ; Hong Jiang, <i>Department of Computer Science and Engineering, University of Texas at Arlington</i> ; Dan Feng, <i>Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology</i> ; Nan Jiang, <i>School of Information Engineering, East China Jiaotong University</i> ; Taorong Qiu and Wei Huang, <i>School of Mathematics and Computer Sciences, Nanchang University</i>	

Structuring Graphs

TC-GNN: Bridging Sparse GNN Computation and Dense Tensor Cores on GPUs	149
Yuke Wang, Boyuan Feng, Zheng Wang, Guyue Huang, and Yufei Ding, <i>University of California, Santa Barbara</i>	

Legion: Automatically Pushing the Envelope of Multi-GPU System for Billion-Scale GNN Training	165
Jie Sun, <i>Collaborative Innovation Center of Artificial Intelligence, Zhejiang University, China</i> ; Li Su, <i>Alibaba Group</i> ; Zuocheng Shi, <i>Collaborative Innovation Center of Artificial Intelligence, Zhejiang University, China</i> ; Wenting Shen, <i>Alibaba Group</i> ; Zeke Wang, <i>Collaborative Innovation Center of Artificial Intelligence, Zhejiang University, China</i> ; Lei Wang, <i>Alibaba Group</i> ; Jie Zhang, <i>Collaborative Innovation Center of Artificial Intelligence, Zhejiang University, China</i> ; Yong Li, Wenyuan Yu, and Jingren Zhou, <i>Alibaba Group</i> ; Fei Wu, <i>Collaborative Innovation Center of Artificial Intelligence, Zhejiang University, China and Shanghai Institute for Advanced Study of Zhejiang University, China</i>	

Bridging the Gap between Relational OLTP and Graph-based OLAP	181
Sijie Shen, <i>Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University and Alibaba Group</i> ; Zihang Yao and Lin Shi, <i>Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University</i> ; Lei Wang, Longbin Lai, Qian Tao, and Li Su, <i>Alibaba Group</i> ; Rong Chen, <i>Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University and Shanghai AI Laboratory</i> ; Wenyuan Yu, <i>Alibaba Group</i> ; Haibo Chen and Binyu Zang, <i>Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University</i> ; Jingren Zhou, <i>Alibaba Group</i>	

Placement and Fault Tolerance

Comosum: An Extensible, Reconfigurable, and Fault-Tolerant IoT Platform for Digital Agriculture	197
Gloire Rubambiza, Shiang-Wan Chin, Mueed Rehman, Sachille Atapattu, José F. Martínez, and Hakim Weatherspoon, <i>Cornell University</i>	

oakestra: A Lightweight Hierarchical Orchestration Framework for Edge Computing	215
Giovanni Bartolomeo, Mehdi Yosofie, Simon Baurle, Oliver Haluszczynski, Nitinder Mohan, and Jörg Ott, <i>Technical University of Munich, Germany</i>	

Explore Data Placement Algorithm for Balanced Recovery Load Distribution.	233
Yingdi Shan, <i>Zhongguancun Laboratory and Tsinghua University</i> ; Kang Chen and Yongwei Wu, <i>Tsinghua University</i>	

Updating Code

LUCI: Loader-based Dynamic Software Updates for Off-the-shelf Shared Objects.	241
Bernhard Heinloth, Peter Wägemann, and Wolfgang Schröder-Preikschat, <i>Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany</i>	

MELF: Multivariant Executables for a Heterogeneous World	257
Dominik Töllner, <i>Leibniz Universität Hannover</i> ; Christian Dietrich, <i>Hamburg University of Technology</i> ; Illia Ostapysyn, Florian Rommel, and Daniel Lohmann, <i>Leibniz Universität Hannover</i>	

APRON: Authenticated and Progressive System Image Renovation	275
Sangho Lee, <i>Microsoft Research</i>	

zpoline: a system call hook mechanism based on binary rewriting.	293
Kenichi Yasukata, Hajime Tazaki, and Pierre-Louis Aublin, <i>IIJ Research Laboratory</i> ; Kenta Ishiguro, <i>Hosei University</i>	

Tuesday, July 11

Serverless

Sponge: Fast Reactive Scaling for Stream Processing with Serverless Frameworks 301
Won Wook Song, *Seoul National University*; Taegeon Um, *Samsung Research*; Sameh Elnikety, *Microsoft Research*;
Myeongjae Jeon, *UNIST*; Byung-Gon Chun, *Seoul National University and FriendlyAI*

On-demand Container Loading in AWS Lambda 315
Marc Brooker, Mike Danilov, Chris Greenwood, and Phil Piwonka, *Amazon Web Services*

Decentralized and Stateful Serverless Computing on the Internet Computer Blockchain 329
Maksym Arutyunyan, Andriy Berestovskyy, Adam Bratschi-Kaye, Ulan Degenbaev, Manu Drijvers, Islam El-Ashi,
Stefan Kaestle, Roman Kashitsyn, Maciej Kot, Yvonne-Anne Pignolet, Rostislav Rumenov, Dimitris Sarlis, Alin Sinpalean,
Alexandru Uta, Bogdan Warinschi, and Alexandra Zapuc, *DFINITY, Zurich*

Troubleshooting and Measurement

PINOLO: Detecting Logical Bugs in Database Management Systems with Approximate Query Synthesis 345
Zongyin Hao and Quanfeng Huang, *School of Informatics, Xiamen University*; Chengpeng Wang, *The Hong Kong University of Science and Technology*; Jianfeng Wang, *University of Southern California*; Yushan Zhang, *Tencent Inc.*; Rongxin Wu, *School of Informatics, Xiamen University*; Charles Zhang, *The Hong Kong University of Science and Technology*

AutoARTS: Taxonomy, Insights and Tools for Root Cause Labelling of Incidents in Microsoft Azure 359
Pradeep Dogga, *UCLA*; Chetan Bansal, Richard Costleigh, Gopinath Jayagopal, Suman Nath, and Xuchao Zhang, *Microsoft*

Avoiding the Ordering Trap in Systems Performance Measurement 373
Dmitry Duplyakin and Nikhil Ramesh, *University of Utah*; Carina Imburgia, *University of Washington*; Hamza Fathallah Al Sheikh, Semil Jain, Prikshit Tekta, Aleksander Maricq, Gary Wong, and Robert Ricci, *University of Utah*

Cloud and Microservices

AWARE: Automate Workload Autoscaling with Reinforcement Learning in Production Cloud Systems 387
Haoran Qiu and Weichao Mao, *University of Illinois at Urbana-Champaign*; Chen Wang, Hubertus Franke, and Alaa Youssef, *IBM Research*; Zbigniew T. Kalbarczyk, Tamer Başar, and Ravishankar K. Iyer, *University of Illinois at Urbana-Champaign*

Nodens: Enabling Resource Efficient and Fast QoS Recovery of Dynamic Microservice Applications in Datacenters ..403
Jiuchen Shi, Hang Zhang, Zhixin Tong, Quan Chen, Kaihua Fu, and Minyi Guo, *Department of Computer Science and Engineering, Shanghai Jiao Tong University*

Lifting the veil on Meta’s microservice architecture: Analyses of topology and request workflows 419
Darby Huye, *Tufts University, Meta*; Yuri Shkuro, *Meta*; Raja R. Sambasivan, *Tufts University*

Distributed Storage

Tectonic-Shift: A Composite Storage Fabric for Large-Scale ML Training 433
Mark Zhao, *Stanford University and Meta*; Satadru Pan, Niket Agarwal, Zhaoduo Wen, David Xu, Anand Natarajan, Pavan Kumar, Shiva Shankar P, Ritesh Tijoriwala, Karan Asher, Hao Wu, Aarti Basant, Daniel Ford, Delia David, Nezih Yigitbasi, Pratap Singh, and Carole-Jean Wu, *Meta*; Christos Kozyrakis, *Stanford University*

Calcspar: A Contract-Aware LSM Store for Cloud Storage with Low Latency Spikes 451
Yuanhui Zhou and Jian Zhou, *WNLO, Huazhong University of Science and Technology, Wuhan, Hubei, China*; Shuning Chen, *PingCAP, China*; Peng Xu, *Research Center for Graph Computing, Zhejiang Lab, Hangzhou, Zhejiang, China*; Peng Wu, *WNLO, Huazhong University of Science and Technology, Wuhan, Hubei, China*; Yanguang Wang and Xian Liu, *PingCAP, China*; Ling Zhan, *Division of Information Science and Technology, Wenhua University, Wuhan, China*; Jiguang Wan, *WNLO, Huazhong University of Science and Technology, Wuhan, Hubei, China*

Adaptive Online Cache Capacity Optimization via Lightweight Working Set Size Estimation at Scale 467
Rong Gu, Simian Li, Haipeng Dai, Hancheng Wang, and Yili Luo, *State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China*; Bin Fan, *Alluxio Inc*; Ran Ben Basat, *University College London*; Ke Wang, *Meta Inc*; Zhenyu Song, *Princeton University*; Shouwei Chen and Beinan Wang, *Alluxio Inc*; Yihua Huang and Guihai Chen, *State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China*

Hardware and Software for Security and Performance

SAGE: Software-based Attestation for GPU Execution 485

Andrei Ivanov and Benjamin Rothenberger, *ETH Zürich*; Arnaud Dethise and Marco Canini, *KAUST*; Torsten Hoefler and Adrian Perrig, *ETH Zürich*

Confidential Computing within an AI Accelerator 501

Kapil Vaswani, Stavros Volos, Cédric Fournet, Antonio Nino Diaz, and Ken Gordon, *Microsoft*; Balaji Vembu, *Meta*; Sam Webster and David Chisnall, *Microsoft*; Saurabh Kulkarni, *Lucata Systems*; Graham Cunningham, *XTX Markets*; Richard Osborne, *Graphcore*; Daniel Wilkinson, *Imagination Technologies*

Arbitor: A Numerically Accurate Hardware Emulation Tool for DNN Accelerators 519

Chenhao Jiang and Anand Jayarajan, *University of Toronto and Vector Institute*; Hao Lu, *University of Toronto*; Gennady Pekhimenko, *University of Toronto and Vector Institute*

Networking

oBBR: Optimize Retransmissions of BBR Flows on the Internet 537

Pengqiang Bi, Mengbai Xiao, Dongxiao Yu, and Guanghui Zhang, *Shandong University*; Jian Tong, Jingchao Liu, and Yijun Li, *BaishanCloud*

Bridging the Gap between QoE and QoS in Congestion Control: A Large-scale Mobile Web Service Perspective.553

Jia Zhang, *Tsinghua University, Zhongguancun Laboratory, Beijing National Research Center for Information Science and Technology*; Yixuan Zhang, *Tsinghua University, Beijing National Research Center for Information Science and Technology*; Enhuan Dong, *Tsinghua University, Quan Cheng Laboratory, Beijing National Research Center for Information Science and Technology*; Yan Zhang, Shaorui Ren, and Zili Meng, *Tsinghua University, Beijing National Research Center for Information Science and Technology*; Mingwei Xu, *Tsinghua University, Quan Cheng Laboratory, Beijing National Research Center for Information Science and Technology*; Xiaotian Li, Zongzhi Hou, and Zhicheng Yang, *Meituan Inc.*; Xiaoming Fu, *University of Goettingen*

FarReach: Write-back Caching in Programmable Switches 571

Siyuan Sheng and Huancheng Puyang, *The Chinese University of Hong Kong*; Qun Huang, *Peking University*; Lu Tang, *Xiamen University*; Patrick P. C. Lee, *The Chinese University of Hong Kong*

Memory-Related Hardware and Software

CXL-ANNS: Software-Hardware Collaborative Memory Disaggregation and Computation for Billion-Scale Approximate Nearest Neighbor Search 585

Junhyeok Jang, *Computer Architecture and Memory Systems Laboratory, KAIST*; Hanjin Choi, *Computer Architecture and Memory Systems Laboratory, KAIST and Panmnnesia, Inc.*; Hanyeoreum Bae and Seungjun Lee, *Computer Architecture and Memory Systems Laboratory, KAIST*; Miryeong Kwon and Myoungsoo Jung, *Computer Architecture and Memory Systems Laboratory, KAIST and Panmnnesia, Inc.*

Overcoming the Memory Wall with CXL-Enabled SSDs. 601

Shao-Peng Yang, *Syracuse University*; Minjae Kim, *DGIST*; Sanghyun Nam, *Soongsil University*; Juhung Park, *DGIST*; Jin-yong Choi and Eyee Hyun Nam, *FADU Inc.*; Eunji Lee, *Soongsil University*; Sungjin Lee, *DGIST*; Bryan S. Kim, *Syracuse University*

STRYX: Exploiting SmartNIC Capability to Reduce Datacenter Memory Tax. 619

Houxiang Ji, *University of Illinois Urbana-Champaign*; Mark Mansi, *University of Wisconsin-Madison*; Yan Sun, *University of Illinois Urbana-Champaign*; Yifan Yuan, *Intel Labs*; Jinghan Huang and Reese Kuper, *University of Illinois Urbana-Champaign*; Michael M. Swift, *University of Wisconsin-Madison*; Nam Sung Kim, *University of Illinois Urbana-Champaign*

Deployed Networking

Change Management in Physical Network Lifecycle Automation 635

Mohammad Al-Fares, Virginia Beauregard, Kevin Grant, Angus Griffith, Jahangir Hasan, Chen Huang, Quan Leng, Jiayao Li, and Alexander Lin, *Google*; Zhuotao Liu, *Tsinghua University*; Ahmed Mansy, *Google*; Bill Martinusen, *Formerly at Google*; Nikil Mehta, Jeffrey C. Mogul, Andrew Narver, and Anshul Nigam, *Google*; Melanie Obenberger, *Formerly at Google*; Sean Smith, *Databricks*; Kurt Steinkraus, Sheng Sun, Edward Thiele, and Amin Vahdat, *Google*

AAsclepius: Monitoring, Diagnosing, and Detouring at the Internet Peering Edge	655
Kaicheng Yang and Yuanpeng Li, <i>National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University and Peng Cheng Laboratory, Shenzhen, China</i> ; Sheng Long, <i>National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University and Huawei Cloud Computing Technologies Co., Ltd., China</i> ; Tong Yang, <i>National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University and Peng Cheng Laboratory, Shenzhen, China</i> ; Ruijie Miao and Yikai Zhao, <i>National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University</i> ; Chaoyang Ji, Penghui Mi, Guodong Yang, Qiong Xie, Hao Wang, Yinhua Wang, Bo Deng, Zhiqiang Liao, Chengqiang Huang, Yongqiang Yang, Xiang Huang, Wei Sun, and Xiaoping Zhu, <i>Huawei Cloud Computing Technologies Co., Ltd., China</i>	
Deploying User-space TCP at Cloud Scale with LUNA	673
Lingjun Zhu, Yifan Shen, Erci Xu, Bo Shi, Ting Fu, Shu Ma, Shuguang Chen, Zhongyu Wang, Haonan Wu, Xingyu Liao, Zhendan Yang, Zhongqing Chen, Wei Lin, Yijun Hou, Rong Liu, Chao Shi, Jiaji Zhu, and Jiesheng Wu, <i>Alibaba Group</i>	
Key-Value Stores	
RubbleDB: CPU-Efficient Replication with NVMe-oF	689
Haoyu Li, Sheng Jiang, and Chen Chen, <i>Columbia University</i> ; Ashwini Raina, <i>Princeton University</i> ; Xingyu Zhu, Changxu Luo, and Asaf Cidon, <i>Columbia University</i>	
Distributed Transactions at Scale in Amazon DynamoDB	705
Joseph Idziorrek, Alex Keyes, Colin Lazier, Somu Perianayagam, Prithvi Ramanathan, James Christopher Sorenson III, Doug Terry, and Akshat Vig, <i>Amazon Web Services</i>	
Security: Attacks	
Prefix Siphoning: Exploiting LSM-Tree Range Filters For Information Disclosure	719
Adi Kaufman, <i>Tel Aviv University</i> ; Moshik Hershcovitch, <i>Tel Aviv University & IBM Research</i> ; Adam Morrison, <i>Tel Aviv University</i>	
EPF: Evil Packet Filter	735
Di Jin, Vaggelis Atlidakis, and Vasileios P. Kemerlis, <i>Brown University</i>	
Wednesday, July 12	
Virtual Machines	
Translation Pass-Through for Near-Native Paging Performance in VMs	753
Shai Bergman and Mark Silberstein, <i>Technion</i> ; Takahiro Shinagawa, <i>University of Tokyo</i> ; Peter Pietzuch and Lluís Vilanova, <i>Imperial College London</i>	
Efficient Memory Overcommitment for I/O Passthrough Enabled VMs via Fine-grained Page Meta-data Management	769
Yaohui Wang, Ben Luo, and Yibin Shen, <i>Alibaba Group</i>	
LPNS: Scalable and Latency-Predictable Local Storage Virtualization for Unpredictable NVMe SSDs in Clouds ...	785
Bo Peng, Cheng Guo, Jianguo Yao, and Haibing Guan, <i>Shanghai Jiao Tong University</i>	
Persistent Memory	
p²CACHE: Exploring Tiered Memory for In-Kernel File Systems Caching	801
Zhen Lin, <i>Binghamton University</i> ; Lingfeng Xiang and Jia Rao, <i>The University of Texas at Arlington</i> ; Hui Lu, <i>Binghamton University</i>	
Revisiting Secondary Indexing in LSM-based Storage Systems with Persistent Memory	817
Jing Wang, Youyou Lu, Qing Wang, Yuhao Zhang, and Jiwu Shu, <i>Department of Computer Science and Technology, Tsinghua University and Beijing National Research Center for Information Science and Technology (BNRist)</i>	
Zhuque: Failure is Not an Option, it's an Exception	833
George Hodgkins, <i>University of Colorado, Boulder</i> ; Yi Xu and Steven Swanson, <i>University of California, San Diego</i> ; Joseph Izraelevitz, <i>University of Colorado, Boulder</i>	

Offloading and Scheduling

- ENVPIPE: Performance-preserving DNN Training Framework for Saving Energy** 851
Sangjin Choi and Inho Koo, *KAIST*; Jeongseob Ahn, *Ajou University*; Myeongjae Jeon, *UNIST*; Youngjin Kwon, *KAIST*
- Decentralized Application-Level Adaptive Scheduling for Multi-Instance DNNs on Open Mobile Devices** 865
Hsin-Hsuan Sung and Jou-An Chen, *Department of Computer Science, North Carolina State University*; Wei Niu, Jiexiong Guan, and Bin Ren, *Department of Computer Science, William & Mary*; Xipeng Shen, *Department of Computer Science, North Carolina State University*
- UnFaaSener: Latency and Cost Aware Offloading of Functions from Serverless Platforms** 879
Ghazal Sadeghian and Mohamed Elsakhawy, *University of British Columbia*; Mohanna Shahradd, *McGill University*; Joe Hattori, *University of Tokyo*; Mohammad Shahradd, *University of British Columbia*

Kernel and Concurrency

- LLFREE: Scalable and Optionally-Persistent Page-Frame Allocation** 897
Lars Wrenger, Florian Rommel, and Alexander Halbuer, *Leibniz Universität Hannover*; Christian Dietrich, *Hamburg University of Technology*; Daniel Lohmann, *Leibniz Universität Hannover*
- SINGULARFS: A Billion-Scale Distributed File System Using a Single Metadata Server** 915
Hao Guo, Youyou Lu, Wenhao Lv, Xiaojian Liao, Shaoxun Zeng, and Jiwu Shu, *Tsinghua University*
- The Hitchhiker's Guide to Operating Systems** 929
Yanyan Jiang, *Nanjing University*

Optimizing ML

- Accelerating Distributed MoE Training and Inference with Lina** 945
Jiamin Li, *City University of Hong Kong*; Yimin Jiang, *ByteDance Inc.*; Yibo Zhu, *Unaffiliated*; Cong Wang, *City University of Hong Kong*; Hong Xu, *The Chinese University of Hong Kong*
- SMARTMoE: Efficiently Training Sparsely-Activated Models through Combining Offline and Online Parallelization** .. 961
Mingshu Zhai, Jiaao He, Zixuan Ma, Zan Zong, Runqing Zhang, and Jidong Zhai, *Tsinghua University*
- MSRL: Distributed Reinforcement Learning with Dataflow Fragments** 977
Huanzhou Zhu, *Imperial College London*; Bo Zhao, *Imperial College London and Aalto University*; Gang Chen, Weifeng Chen, Yijie Chen, and Liang Shi, *Huawei Technologies Co., Ltd.*; Yaodong Yang, *Peking University*; Peter Pietzuch, *Imperial College London*; Lei Chen, *Hong Kong University of Science and Technology*

GPU

- Beware of Fragmentation: Scheduling GPU-Sharing Workloads with Fragmentation Gradient Descent** 995
Qizhen Weng and Lingyun Yang, *Hong Kong University of Science and Technology*; Yinghao Yu, *Alibaba Group and Hong Kong University of Science and Technology*; Wei Wang, *Hong Kong University of Science and Technology*; Xiaochuan Tang, Guodong Yang, and Liping Zhang, *Alibaba Group*
- Towards Iterative Relational Algebra on the GPU** 1009
Ahmedur Rahman Shovon and Thomas Gilray, *University of Alabama at Birmingham*; Kristopher Micinski, *Syracuse University*; Sidharth Kumar, *University of Alabama at Birmingham*
- VectorVisor: A Binary Translation Scheme for Throughput-Oriented GPU Acceleration** 1017
Samuel Ginzburg, *Princeton University*; Mohammad Shahradd, *University of British Columbia*; Michael J. Freedman, *Princeton University*