

# **19th USENIX Conference on File and Storage Technologies (FAST'21)**

Online  
23 - 25 February 2021

ISBN: 978-1-7138-2422-0

**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© (2021) by Usenix Association  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

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

# 19th USENIX Conference on File and Storage Technologies (FAST '21)

February 23–25, 2021

## Tuesday, February 23

### Indexing and Key-Value Store

- ROART: Range-query Optimized Persistent ART** ..... 1  
Shaonan Ma and Kang Chen, *Tsinghua University*; Shimin Chen, *SKL of Computer Architecture, ICT, CAS, and University of Chinese Academy of Sciences*; Mengxing Liu, Jianglang Zhu, Hongbo Kang, and Yongwei Wu, *Tsinghua University*
- SpanDB: A Fast, Cost-Effective LSM-tree Based KV Store on Hybrid Storage** .....17  
Hao Chen, *University of Science and Technology of China & Qatar Computing Research Institute, HBKU*; Chaoyi Ruan and Cheng Li, *University of Science and Technology of China*; Xiaosong Ma, *Qatar Computing Research Institute, HBKU*; Yinlong Xu, *University of Science and Technology of China & Anhui Province Key Laboratory of High Performance Computing*
- Evolution of Development Priorities in Key-value Stores Serving Large-scale Applications:  
The RocksDB Experience.** ..... 33  
Siying Dong, Andrew Kryczka, and Yanqin Jin, *Facebook Inc.*; Michael Stumm, *University of Toronto*
- REMIX: Efficient Range Query for LSM-trees** ..... 51  
Wenshao Zhong, Chen Chen, and Xingbo Wu, *University of Illinois at Chicago*; Song Jiang, *University of Texas at Arlington*

### Advanced File Systems

- High Velocity Kernel File Systems with Bento.** ..... 65  
Samantha Miller, Kaiyuan Zhang, Mengqi Chen, and Ryan Jennings, *University of Washington*; Ang Chen, *Rice University*; Danyang Zhuo, *Duke University*; Thomas Anderson, *University of Washington*
- Scalable Persistent Memory File System with Kernel-Userspace Collaboration** ..... 81  
Youmin Chen, Youyou Lu, and Bohong Zhu, *Tsinghua University*; Andrea C. Arpaci-Dusseau and Remzi H. Arpaci-Dusseau, *University of Wisconsin–Madison*; Jiwu Shu, *Tsinghua University*
- Rethinking File Mapping for Persistent Memory** ..... 97  
Ian Neal, Gefei Zuo, Eric Shiple, and Tanvir Ahmed Khan, *University of Michigan*; Youngjin Kwon, *School of Computing, KAIST*; Simon Peter, *University of Texas at Austin*; Baris Kasikci, *University of Michigan*
- pFSCK: Accelerating File System Checking and Repair for Modern Storage.** ..... 113  
David Domingo and Sudarsun Kannan, *Rutgers University*
- Pattern-Guided File Compression with User-Experience Enhancement for Log-Structured File System on Mobile Devices.** ..... 127  
Cheng Ji, *Nanjing University of Science and Technology*; Li-Pin Chang, *National Chiao Tung University, National Yang Ming Chiao Tung University*; Riwei Pan and Chao Wu, *City University of Hong Kong*; Congming Gao, *Tsinghua University*; Liang Shi, *East China Normal University*; Tei-Wei Kuo and Chun Jason Xue, *City University of Hong Kong*

## Wednesday, February 24

### Transactions, Deduplication, and More

- ArchTM: Architecture-Aware, High Performance Transaction for Persistent Memory** .....141  
Kai Wu and Jie Ren, *University of California, Merced*; Ivy Peng, *Lawrence Livermore National Laboratory*; Dong Li, *University of California, Merced*
- SPHT: Scalable Persistent Hardware Transactions** ..... 155  
Daniel Castro, *INESC-ID & Instituto Superior Técnico*; Alexandro Baldassin, *UNESP - Universidade Estadual Paulista*; João Barreto and Paolo Romano, *INESC-ID & Instituto Superior Técnico*
- The Dilemma between Deduplication and Locality: Can Both be Achieved?** .....171  
Xiangyu Zou and Jingsong Yuan, *Harbin Institute of Technology, Shenzhen*; Philip Shilane, *Dell Technologies*; Wen Xia, *Harbin Institute of Technology, Shenzhen, and Wuhan National Laboratory for Optoelectronics*; Haijun Zhang and Xuan Wang, *Harbin Institute of Technology, Shenzhen*

**Remap-SSD: Safely and Efficiently Exploiting SSD Address Remapping to Eliminate Duplicate Writes** . . . . . 187  
You Zhou, Qiulin Wu, and Fei Wu, *Huazhong University of Science and Technology*; Hong Jiang, *University of Texas at Arlington*; Jian Zhou and Changsheng Xie, *Huazhong University of Science and Technology*

**CheckFreq: Frequent, Fine-Grained DNN Checkpointing** . . . . . 203  
Jayashree Mohan, *UT Austin*; Amar Phanishayee, *Microsoft Research*; Vijay Chidambaram, *UT Austin and VMware research*

## Cloud and Distributed Systems

**Facebook's Tectonic Filesystem: Efficiency from Exascale** . . . . . 217  
Satadru Pan, *Facebook, Inc.*; Theano Stavrinos, *Facebook, Inc. and Princeton University*; Yunqiao Zhang, Atul Sikaria, Pavel Zakharov, Abhinav Sharma, Shiva Shankar P, Mike Shuey, Richard Wareing, Monika Gangapuram, Guanglei Cao, Christian Preseau, Pratap Singh, Kestutis Patiejunas, and JR Tipton, *Facebook, Inc.*; Ethan Katz-Bassett, *Columbia University*; Wyatt Lloyd, *Princeton University*

**Exploiting Combined Locality for Wide-Stripe Erasure Coding in Distributed Storage** . . . . . 233  
Yuchong Hu, Liangfeng Cheng, and Qiaori Yao, *Huazhong University of Science & Technology*; Patrick P. C. Lee, *The Chinese University of Hong Kong*; Weichun Wang and Wei Chen, *HIKVISION*

**On the Feasibility of Parser-based Log Compression in Large-Scale Cloud Systems** . . . . . 249  
Junyu Wei and Guangyan Zhang, *Tsinghua University*; Yang Wang, *The Ohio State University*; Zhiwei Liu, *China University of Geosciences*; Zhanyang Zhu and Junchao Chen, *Tsinghua University*; Tingtao Sun and Qi Zhou, *Alibaba Cloud*

**CNSBench: A Cloud Native Storage Benchmark** . . . . . 263  
Alex Merenstein, *Stony Brook University*; Vasily Tarasov, Ali Anwar, and Deepavali Bhagwat, *IBM Research–Almaden*; Julie Lee, *Stony Brook University*; Lukas Rupprecht and Dimitris Skourtis, *IBM Research–Almaden*; Yang Yang and Erez Zadok, *Stony Brook University*

**Concordia: Distributed Shared Memory with In-Network Cache Coherence** . . . . . 277  
Qing Wang, Youyou Lu, Erci Xu, Junru Li, Youmin Chen, and Jiwu Shu, *Tsinghua University*

## Thursday, February 25

### Caching Everywhere

**eMRC: Efficient Miss Ratio Approximation for Multi-Tier Caching** . . . . . 293  
Zhang Liu, *University of Colorado Boulder*; Hee Won Lee, *Samsung Electronics*; Yu Xiang, *AT&T Labs Research*; Dirk Grunwald and Sangtae Ha, *University of Colorado Boulder*

**The Storage Hierarchy is Not a Hierarchy: Optimizing Caching on Modern Storage Devices with Orthus** . . . . . 307  
Kan Wu, Zhihan Guo, Guanzhou Hu, and Kaiwei Tu, *University of Wisconsin–Madison*; Ramnathan Alagappan, *VMware Research*; Rathijit Sen and Kwanghyun Park, *Microsoft*; Andrea C. Arpaci-Dusseau and Remzi H. Arpaci-Dusseau, *University of Wisconsin–Madison*

**A Community Cache with Complete Information** . . . . . 325  
Mania Abdi, *Northeastern University*; Amin Mosayyebzadeh, *Boston University*; Mohammad Hossein Hajkazemi, *Northeastern University*; Emine Ugur Kaynar, *Boston University*; Ata Turk, *State Street*; Larry Rudolph, *TwoSigma*; Orran Krieger, *Boston University*; Peter Desnoyers, *Northeastern University*

**Learning Cache Replacement with CACHEUS** . . . . . 341  
Liana V. Rodriguez, Farzana Yusuf, Steven Lyons, Eysler Paz, Raju Rangaswami, and Jason Liu, *Florida International University*; Ming Zhao, *Arizona State University*; Giri Narasimhan, *Florida International University*

### The SSD Revolution Is Not Over

**FusionRAID: Achieving Consistent Low Latency for Commodity SSD Arrays** . . . . . 355  
Tianyang Jiang, Guangyan Zhang, and Zican Huang, *Tsinghua University*; Xiaosong Ma, *Qatar Computing Research Institute, HBKU*; Junyu Wei, Zhiyue Li, and Weimin Zheng, *Tsinghua University*

**Behemoth: A Flash-centric Training Accelerator for Extreme-scale DNNs** . . . . . 371  
Shine Kim, *Seoul National University and Samsung Electronics*; Yunho Jin, Gina Sohn, Jonghyun Bae, Tae Jun Ham, and Jae W. Lee, *Seoul National University*

<b>FlashNeuron: SSD-Enabled Large-Batch Training of Very Deep Neural Networks. ....</b>	<b>387</b>
<i>Jonghyun Bae, Seoul National University; Jongsung Lee, Seoul National University and Samsung Electronics;</i>	
<i>Yunho Jin and Sam Son, Seoul National University; Shine Kim, Seoul National University and Samsung Electronics;</i>	
<i>Hakbeom Jang, Samsung Electronics; Tae Jun Ham and Jae W. Lee, Seoul National University</i>	
<b>D2FQ: Device-Direct Fair Queueing for NVMe SSDs .....</b>	<b>403</b>
<i>Jiwon Woo, Minwoo Ahn, Gyusun Lee, and Jinkyu Jeong, Sungkyunkwan University</i>	
<b>An In-Depth Study of Correlated Failures in Production SSD-Based Data Centers .....</b>	<b>417</b>
<i>Shujie Han and Patrick P. C. Lee, The Chinese University of Hong Kong; Fan Xu, Yi Liu, Cheng He, and Jiongzhou Liu,</i>	
<i>Alibaba Group</i>	