Peter Bailis

535 Mission St. Suite 1100 Sisu Data San Francisco, CA 94105 http://bailis.org/ peter@sisudata.com *Revised 08/2020*

RESEARCH INTERESTS

Data management and analytics, data-intensive computing, distributed systems

EDUCATION

University of California, Berkeley PhD in Computer Science.	2011-2015
Advisors: Ali Ghodsi, Joseph M. Hellerstein, Ion Stoica	
Harvard College	2007-2011
AB in Computer Science.	
Mugnu cum uuue with Highest Honors in Computer Science	
CURRENT APPOINTMENT	
CEO and Founder	2018-present
Sisu Data	San Francisco, CA
ACADEMIC APPOINTMENTS	
Adjunct Professor	2020-present
Stanford University Computer Science Department	Stanford, CA
Assistant Professor	2016-2020
Stanford University Computer Science Department	Stanford, CA
NB: per Stanford rule on "2 in 7 years" academic leave, left tenure track to pursue Sisu full time	
Postdoctoral Researcher	Nov. 2015-Apr. 2016
MIT Computer Science and Artificial Intelligence Lab	Cambridge, MA
Graduate Student Researcher	2011–2015
UC Berkeley Computer Science Division	Berkeley, CA
Undergraduate Researcher	2009–2011
Harvard College	Cambridge, MA
AWARDS and HONORS	
Google Faculty Research Award, 2017, 2018, 2019, 2020	
Facebook Hardware & Software Systems Research Award, 2018	
ACM SIGMOD Jim Gray Doctoral Dissertation Award, 2017	
Okawa Foundation Grant, 2017	
NetApp Faculty Fellowship, 2017	

ACM Doctoral Dissertation Award, Honorable Mention, 2016

VISA Faculty Award, 2016

Facebook Faculty Award, 2016 VMWare Early Career Faculty Award, 2016 Stanford David Morganthaler II Faculty Fellow, 2016-2018 Forbes "30 Under 30" in Enterpise Technology, 2016 UC Berkeley EECS David J. Sakrison Memorial Prize, 2015 UC Berkeley Outstanding Graduate Student Instructor Award, 2014 National Science Foundation Graduate Research Fellowship, 2011-2016 Berkeley Fellowship for Graduate Study, 2011-2014 UC Berkeley EECS Chair's Excellence Award, 2011-2012 Computing Research Association Outstanding Undergraduate Researcher, 2011 Thomas Temple Hoopes Prize for Outstanding Senior Thesis, 2011 Phi Beta Kappa, Harvard College Chapter, inducted 2011 Derek Bok Center Certificate for Distinction in Teaching, 2009

Paper Awards

"Best of VLDB 2019" invitation to VLDB Journal Best Student Paper, ICLR DebugML Workshop 2019 "Best of SIGMOD 2017" invitation to ACM TODS "Best of SIGMOD 2014" citation in ACM TODS Communications of the ACM Research Highlight, 2014 "Best of VLDB 2012" citation in the VLDB Journal Best Student Paper, ANTS 2010

FUNDING

Raised approximately \$5.1M (directly in personal research funding, approximately 90% discretionary) from 2016-2019. Founded and led fundraising for >\$20M Stanford DAWN project.

PUBLICATIONS

Journal Articles

- Firas Abuzaid, Peter Kraft, Sahaana Suri, Edward Gan, Eric Xu, Atul Shenoy, Asvin Anathanarayan, John Sheu, Erik Meijer, Xi Wu, Jeffrey Naughton, Peter Bailis, and Matei Zaharia. DIFF: A relational interface for large-scale data explanation. *The VLDB Journal*, 2020. "Best of VLDB 2019" Special Issue.
- Clara E. Yoon, Karianne J. Bergen, Kexin Rong, Hashem Elezabi, William L. Ellsworth, Gregory C. Beroza, Peter Bailis, and Philip Levis. Unsupervised large-scale search for similar earthquake signalsunsupervised large-scale search for similar earthquake signals. *Bulletin of the Seismological Society of America*, June 2019.
- Firas Abuzaid, Peter Bailis, Jialin Ding, Edward Gan, Samuel Madden, Deepak Narayanan, Kexin Rong, and Sahaana Suri. MacroBase: Prioritizing attention in fast data. ACM TODS, 43(4), 12, 2018. "Best of SIGMOD 2017" Special Issue.
- Peter Bailis, Alan Fekete, Ali Ghodsi, Joseph M. Hellerstein, and Ion Stoica. Scalable atomic visibility with RAMP transactions. *ACM TODS*, 2016. "Best of SIGMOD 2014" Special Issue.

Peter Bailis, Shivaram Venkataraman, Michael J. Franklin, Joseph M. Hellerstein, and Ion Stoica. Quantifying eventual consistency with PBS. *The VLDB Journal*, 23(2):279–302, 2014. "Best of VLDB 2012" Special Issue.

Refereed Conference Proceedings

- Firas Abuzaid, Srikanth Kandula, Ishai Menache, Peter Bailis, and Matei Zaharia. Contracting wide-area network topologies to solve flow problems quickly. *NSDI*, 2021.
- Cody Coleman, Stephen Mussmann, Baharan Mirzasoleiman, Peter Bailis, Percy Liang, Jure Leskovec, and Matei Zaharia. Selection via proxy: efficient data selection for deep learning. *ICLR*, 2020.
- Edward Gan, Peter Bailis, and Moses Charikar. Storyboard: optimizing precomputed summaries for aggregation. *VLDB*, 2020.
- Daniel Kang, Peter Bailis, and Matei Zaharia. Blazeit: optimizing declarative aggregation and limit queries for neural network-based video analytics. *VLDB*, 2020.
- Daniel Kang, Edward Gan, Peter Bailis, Tatsunori Hashimoto, and Matei Zaharia. Approximate selection with guarantees using proxies. *VLDB*, 2020.
- Daniel Kang, Ankit Mathur, Teja Veeramacheneni, Peter Bailis, and Matei Zaharia. Jointly optimizing preprocessing and inference for DNN-based visual analytics. *VLDB*, 2020.
- Daniel Kang, Deepti Raghavan, Peter Bailis, and Matei Zaharia. Model Assertions for Monitoring and Improving ML Models. *MLSys*, 2020.
- Peter Kraft, Daniel Kang, Deepak Narayanan, Shoumik Palkar, Peter Bailis, and Matei Zaharia. Willump: A statistically-aware end-to-end optimizer for machine learning inference. *MLSys*, 2020.
- Peter Mattson et al. MLPerf Training Benchmark. MLSys, 2020.
- Kexin Rong, Yao Lu, Peter Bailis, Srikanth Kandula, and Philip Levis. Approximate partition selection for big-data workloads using summary statistics. *VLDB*, 2020.
- Sahaana Suri, Raghuveer Chanda, Neslihan Bulut, Pradyumna Narayana, Yemao Zeng, Peter Bailis, Sugato Basu, Girija Narlikar, and Christopher Re. Leveraging organizational resources to adapt models to new data modalities. *VLDB*, 2020.
- Firas Abuzaid, Peter Kraft, Sahaana Suri, Edward Gan, Eric Xu, Atul Shenoy, Asvin Anathanarayan, John Sheu, Erik Meijer, Xi Wu, Jeffrey Naughton, Peter Bailis, and Matei Zaharia. DIFF: A relational interface for large-scale data explanation. *VLDB*, 2019.
- Firas Abuzaid, Geet Sethi, Peter Bailis, and Matei Zaharia. To index or not to index: Optimizing exact maximum inner product search. *ICDE*, 2019.
- Animesh Koratana, Daniel Kang, Peter Bailis, and Matei Zaharia. LIT: Block-wise intermediate representation training for model compression. *ICML*, 2019.
- Vatsal Sharan, Kai Sheng Tai, Peter Bailis, and Gregory Valiant. Compressed factorization: Fast and accurate low-rank factorization of compressively-sensed data. *ICML*, 2019.
- Paris Siminelakis, Kexin Rong, Peter Bailis, Moses Charikar, and Philip Levis. Rehashing kernel evaluation in high dimensions. *ICML*, 2019.
- Kai Sheng Tai, Peter Bailis, and Gregory Valiant. Equivariant transformer networks. ICML, 2019.
- Cody Coleman, Deepak Naraynanan, Daniel Kang, Tian Zhao, Jian Zhang, Luigi Nardi, Peter Bailis, Kunle Olukotun, Chris Re, and Matei Zaharia. Analysis of the time-to-accuracy metric and entries in the dawnbench deep learning benchmark. *NeurIPS Systems for ML Workshop*, 2018.

- Edward Gan, Jialin Ding, Kai Sheng Tai, Vatsal Sharan, and Peter Bailis. Moment-based quantile sketches for efficient high cardinality aggregation queries. *VLDB*, 2018.
- Shoumik Palkar, Firas Abuzaid, Peter Bailis, and Matei Zaharia. Filter before you parse: Faster analytics on raw data with sparser. *VLDB*, 2018.
- Kexin Rong, Clara E. Yoon, Karianne J. Bergen, Hashem Elezabi, Peter Bailis, Philip Levis, and Gregory C. Beroza. Locality-sensitive hashing for earthquake detection: A case study of scaling data-driven science. *VLDB*, 2018.
- Kai Sheng Tai, Vatsal Sharan, Peter Bailis, and Gregory Valiant. Sketching linear classifiers over data streams. *SIGMOD*, 2018.
- Peter Bailis, Edward Gan, Samuel Madden, Deepak Narayanan, Kexin Rong, and Sahaana Suri. MacroBase: prioritizing attention in fast data. *ACM SIGMOD*, 2017.
- Peter Bailis, Edward Gan, Kexin Rong, and Sahaana Suri. Prioritizing Attention in Fast Data: Challenges and Opportunities. *CIDR*, 2017.
- Edward Gan and Peter Bailis. Scalable Kernel Density Classification via Threshold-Based Pruning. ACM SIGMOD, 2017.
- Daniel Kang, John Emmons, Firas Abuzaid, Peter Bailis, and Matei Zaharia. NoScope: Optimizing neural network queries over video at scale. *VLDB*, 2017.
- Kexin Rong and Peter Bailis. ASAP: Prioritizing attention via time series smoothing. VLDB, 2017.
- Todd Warzsawski and Peter Bailis. ACIDRain: concurrency-related attacks on database-backed web applications. ACM SIGMOD, 2017.
- Peter Bailis, Alan Fekete, Michael J. Franklin, Ali Ghodsi, Joseph M. Hellerstein, and Ion Stoica. Coordination avoidance in database systems. *VLDB*, 2015.
- Peter Bailis, Alan Fekete, Michael J. Franklin, Ali Ghodsi, Joseph M. Hellerstein, and Ion Stoica. Feral concurrency control: An empirical investigation of modern application integrity. *ACM SIGMOD*, 2015.
- Dan Crankshaw, Peter Bailis, Joseph E. Gonzalez, Haoyuan Li, Zhao Zhang, Michael J. Franklin, Ali Ghodsi, and Michael I. Jordan. The missing piece in complex analytics: Low latency, scalable model management and serving with Velox. *CIDR*, 2015.
- Peter Bailis, Aaron Davidson, Alan Fekete, Ali Ghodsi, Joseph M. Hellerstein, and Ion Stoica. Highly Available Transactions: Virtues and limitations. *VLDB*, 2014.
- Peter Bailis, Alan Fekete, Ali Ghodsi, Joseph M. Hellerstein, and Ion Stoica. Scalable atomic visibility with RAMP transactions. *ACM SIGMOD*, 2014.
- Peter Alvaro, Peter Bailis, Neil Conway, and Joseph M. Hellerstein. Consistency without borders. ACM SoCC, 2013. Vision track.
- Peter Bailis, Ali Ghodsi, Joseph M. Hellerstein, and Ion Stoica. Bolt-on causal consistency. ACM SIGMOD, 2013.
- Peter Bailis, Alan Fekete, Ali Ghodsi, Joseph M. Hellerstein, and Ion Stoica. The potential dangers of causal consistency and an explicit solution. *ACM SoCC*, 2012. Vision track.
- Peter Bailis, Shivaram Venkataraman, Michael J. Franklin, Joseph M. Hellerstein, and Ion Stoica. Probabilistically bounded staleness for practical partial quorums. *VLDB*, 2012.
- Peter Bailis, Vijay Janapa Reddi, Sanjay Gandhi, David Brooks, and Margo Seltzer. Dimetrodon: Processor-level preventive thermal management via idle cycle injection. *DAC*, 2011.

- Karthik Dantu, Bryan Kate, Jason Waterman, Peter Bailis, and Matt Welsh. Programming micro-aerial vehicle swarms with Karma. ACM SenSys, 2011.
- Peter Bailis, Radhika Nagpal, and Justin Werfel. Positional communication and private information in honeybee foraging models. *ANTS*, 2010.

Conference Demonstrations

- Daniel Kang, Peter Bailis, and Matei Zaharia. Challenges and opportunities in DNN-Based video analytics: a demonstration of the BlazeIt video query engine. *CIDR*, 2019.
- Peter Bailis, Edward Gan, Kexin Rong, and Sahaana Suri. Demonstration: MacroBase, A Fast Data Analysis Engine. *ACM SIGMOD*, 2017.
- Peter Bailis, Shivaram Venkataraman, Michael J. Franklin, Joseph M. Hellerstein, and Ion Stoica. PBS at work: Advancing data management with consistency metrics. *ACM SIGMOD*, 2013.

Refereed Workshop Proceedings

- Justin Chen, Edward Gan, Kexin Rong, Sahaana Suri, and Peter Bailis. CrossTrainer: Practical domain adaptation with loss reweighting. *SIGMOD DEEM Workshop*, 2019.
- Sahaana Suri and Peter Bailis. DROP: A workload-aware optimizer for dimensionality reduction. SIGMOD DEEM Workshop, 2019.
- Daniel Kang, Deepti Raghavan, Sahaana Suri, Peter Bailis, and Matei Zaharia. Model assertions for debugging machine learning. *NeurIPS Systems for ML Workshop*, 2018.
- Shoumik Palkar, Sahaana Suri, Peter Bailis, and Matei Zaharia. Exploring the use of learning algorithms for efficient performance profiling. *NeurIPS ML for Systems Workshop*, 2018.
- Joseph E. Gonzalez, Peter Bailis, Michael J. Franklin, Joseph M. Hellerstein, Michael I. Jordan, and Ion Stoica. Asynchronous complex analytics in a distributed dataflow architecture. *NeurIPS Systems for ML Workshop*, 2015.
- Peter Bailis, Alan Fekete, Ali Ghodsi, Joseph M. Hellerstein, and Ion Stoica. HAT, not CAP: Towards highly available transactions. *HotOS*, 2013.

Invited Articles

- Peter Bailis et al. "Research for Practice" Column. *Communications of the ACM*, 2016-present. Bimonthly column featuring pragmatically-focused research highlights from all of Computer Science.
- Peter Bailis and Kyle Kingsbury. The network is reliable: An informal survey of real-world communications failures. ACM Queue, 12(7), 2014. Also appears in Communications of the ACM 57(9):48-55, September 2014.
- Peter Bailis, Shivaram Venkataraman, Michael J. Franklin, Joseph M. Hellerstein, and Ion Stoica. Quantifying eventual consistency with PBS. *Communications of the ACM*, 57:93–102, 8, 2014.
- Peter Bailis and Ali Ghodsi. Eventual Consistency today: Limitations, extensions, and beyond. *ACM Queue*, 11(3), 2013. Also appears in Communications of the ACM 56(3):55-63, May 2013.

Theses

- Peter Bailis. *Coordination Avoidance in Distributed Databases*, University of California, Berkeley, 2015. PhD Dissertation.
- Peter Bailis. *Dimetrodon: Processor-level Preventive Thermal Management via Idle Cycle Injection*, 2011. Harvard College Undergraduate Honors Thesis.

Books

Peter Bailis, Joseph M. Hellerstein, and Michael Stonebraker, editors. *Readings in Database Systems,* 5th Edition. 2015. URL: http://www.redbook.io/.

In Preparation and Under Review

- Peter Bailis, Kunle Olukuton, Christopher Re, and Matei Zaharia. Infrastructure for usable machine learning: The Stanford DAWN Project, Preprint.
- Cody Coleman, Christopher Yeh, Stephen Mussmann, Baharan Mirzasoleiman, Peter Bailis, Percy Liang, Jure Leskovec, and Matei Zaharia. Selection via proxy: Efficient data selection for deep learning.
- Daniel Kang, Peter Bailis, and Matei Zaharia. BlazeIt: fast exploratory video queries using neural networks.
- Peter Kraft, Daniel Kang, Deepak Narayanan, Shoumik Palkar, Peter Bailis, and Matei Zaharia. Willump: A statistically-aware end-to-end optimizer for machine learning inference.
- Alexander Ratner, Dan Alistarh, Gustavo Alonso, David G. Andersen, Peter Bailis, et al. SysML: The new frontier of machine learning systems.

SERVICE

Program Committees:

SysML (2021, 2020, 2019), CIDR 2021, ACM SIGMOD (2019, 2018, 2017, 2016), VLDB (2019, 2018, 2017), HPTS (2017), ACM SoCC (2015)

Editor:

Research for Practice, appearing in ACM Queue and CACM (2016-2019)

Editorial Board:

ACM Queue (2016-present) Foundations and Trends in Databases (2015-2019)

Organizing Committee:

SysML 2018

Steering Committee:

Workshop on Principles and Practice of Consistency (PaPoC) at Eurosys, 2015-2020

PhD STUDENTS and POSTDOCS

Firas Abuzaid, PhD CS, Stanford (co-advised with Matei Zaharia; 2015-present) Cody Coleman, PhD CS, Stanford (co-advised with Matei Zaharia; 2017-present) Edward Gan, PhD CS, Stanford (2016-2020; first employment; Databricks) Daniel Kang, PhD CS, Stanford (co-advised with Matei Zaharia; 2017-present) Fiodar Kazhamiaka, Postdoc CS, Stanford (co-advised with Matei Zaharia; 2019-present) Peter Kraft, PhD CS, Stanford (2018-present) Kexin Rong, PhD CS, Stanford (co-advised with Phil Levis; 2016-present) Sahaana Suri, PhD EE, Stanford (2016-present) Kai Sheng Tai, PhD CS, Stanford (2017-present)