

Rui Liu

[Google Scholar](#) | [LinkedIn](#) | [GitHub](#) | Email: [rui.liu\[at\]lbl.gov](mailto:rui.liu[at]lbl.gov) | Website: csruiliu.github.io

EDUCATION

- Ph.D. in Computer Science** Sept 2017 – Dec 2023
- University of Chicago, with CERES Graduate Scholar Fellowship
 - Advised by Aaron J. Elmore and Michael J. Franklin
 - Close collaboration with Sanjay Krishnan
- M.Phil. in Computer Science** Oct 2012 – Dec 2015
- Hong Kong Polytechnic University, with PolyU Research Postgraduate Scholarship
 - Advised by Jiannong Cao
- B.Eng. in Software Engineering** Sept 2008 – June 2012
- Northeastern University (China), Graduated with Honors

EXPERIENCE

- Lawrence Berkeley National Laboratory** Dec 2024 – Present
- *NESAP Postdoctoral Fellow* at National Energy Research Scientific Computing Center
 - Working on performance modeling and prediction for workloads on supercomputers.
 - Supervisors: Brian Austin and Richard Shane Canon
- Microsoft** June 2022 – Sept 2022
- *Research Intern* at Gray Systems Lab
 - Worked on feature engineering optimization in feature stores for AI workloads
 - Mentors: Kwanghyun Park, Jesús Camacho Rodríguez, and Yuanyuan Tian
- DocuSign** July 2021 – Sept 2021
- *Data Science Intern* at AI Engineering Team
 - Worked on optimizing AI infrastructure and refining training pipelines for AI workloads
 - Mentor: David Wong
- National University of Singapore** Aug 2016 – July 2017
- *Research Assistant* at Data System Research Group
 - Worked on blockchain, data integration, and distributed data management
 - Mentor: Beng Chin Ooi
- The Chinese University of Hong Kong** Mar 2016 – Aug 2016
- *Research Assistant* at Department of Information Engineering
 - Worked on data sharing in mobile sensing systems
 - Mentor: Kehuan Zhang
- Career Breaks – Family Caregiving and Personal Travel** Multiple
- June 2012 – Oct 2012 | Jan 2016 – Feb 2016 | July 2017 – Sept 2017 | Jan 2024 – Dec 2024

PROJECTS

Selected projects in reverse-chronological order.

- Performance Modeling for GPU-intensive Workloads on Supercomputers at NERSC [Ongoing]**
Role: Driver. **Affiliation:** Lawrence Berkeley National Laboratory.
 - *Objective:* Designing models that predict application runtime on future systems from a limited set of hardware counters collected on an existing system plus published specifications of the target hardware.

- *Contribution:* We built and validated an application-agnostic, cross-architecture GPU performance model (occupancy-aware, roofline-bounded, DCGM-counter-driven) that predicts various GPU architectures runtimes within $\pm 10\%$ in both directions on five NERSC-10 benchmarks and preserves correct target ranking across architectural discontinuities; applied it to a month-long production workload on Perlmutter for procurement. We also developed a companion MPI model that reconstructs message-size distributions from Slingshot NIC counters.
2. **Resource-adaptive Query Execution for Cloud-native Databases**
Role: Driver. **Affiliation:** University of Chicago.
- *Objective:* Designing and developing a resource-adaptive query execution framework to enable adaptive query suspension and resumption at various granularities when resources become unavailable or impractical during execution.
 - *Contribution:* We built a prototype system Riveter, which supports various suspension strategies: a redo strategy, a pipeline-level strategy (by modifying DuckDB), and a process-level strategy (by building on top of a Linux checkpoint and restore tool). Additionally, we devised a cost model to adaptively select the optimal suspension strategy for different data analytic queries. We evaluated Riveter with the TPC-H benchmark, demonstrating its adaptivity and effectiveness.
3. **Optimizing Data Pipelines for Machine Learning in Feature Stores**
Role: Core Member. **Affiliation:** Gray Systems Lab@Microsoft.
- *Objective:* Proposing and developing optimizations for data pipelines in feature stores, including feature computation reuse, adaptive data layouts, incremental feature refresh, feature maintenance scheduling, and compute engine selection.
 - *Contribution:* I implemented optimizations for point-in-time joins by modifying Feathr, a state-of-the-art open-source feature store widely used in production at LinkedIn. I evaluated these optimizations using use cases from the TPCx-AI benchmark and real-world online retail applications. Our experimental analysis shows that the optimizations can accelerate data pipelines by up to 3x compared to state-of-the-art baselines.
4. **Resource-arbitration for Progressive Iterative Analytics**
Role: Driver. **Affiliation:** University of Chicago.
- *Objective:* Proposing and developing a resource arbitration framework for progressive iterative analytic jobs to continuously prioritize the jobs and determine if/when to preempt and reallocate resources.
 - *Contribution:* We implemented two resource arbitration systems: (1) Rotary-AQP, a modification of Apache Spark for approximate query processing, and (2) Rotary-DLT, an extension of TensorFlow for deep learning training. We evaluated these systems using the TPC-H benchmark and a survey-based deep learning workload, respectively. The results demonstrate that both systems significantly outperform state-of-the-art methods and other commonly used baselines across various metrics, such as the number of attained jobs and the attainment progress achieved by each job.
5. **Accelerating Container-based Deep Learning Hyperparameter Optimization Workloads**
Role: Driver. **Affiliation:** AI Engineering Team@DocuSign.
- *Objective:* Accelerating deep learning hyperparameter optimization (HPO) workloads within container-based environments (e.g., Docker).
 - *Contribution:* We developed an end-to-end system, Relish, to efficiently segment long-running hyperparameter optimization (HPO) jobs into sub-jobs and distribute them across multiple devices for parallel execution. We evaluated Relish using a synthetic workload based on a three-month trace from a multi-tenant cluster at DocuSign. The results show that Relish significantly improves GPU utilization and accelerates workload execution.
6. **Understanding and Optimizing Packed Neural Network Training for Hyper-Parameter Tuning**
Role: Driver. **Affiliation:** University of Chicago.
- *Objective:* Investigating the benefits and challenges of optimizing resource utilization and sharing in model training by efficiently packing multiple models.
 - *Contribution:* We proposed a pack primitive and built a prototype system on top of TensorFlow to share training resources among deep learning models, incorporating a cost model to balance the benefits and limitations of packing models for simultaneous training. We conducted an ablation study to explore the trade-offs involved in sharing limited resources across multiple neural network models. Additionally, we applied the pack primitive

to Hyperband, a widely used hyperparameter tuning algorithm, and improved its performance by up to 2.7x through the packing operation.

7. Draining the Data Swamp

Role: Core Member. **Affiliation:** University of Chicago.

- ▶ *Objective:* Building data discovery and management tools for data swamps where data lakes have overwhelming heterogeneous and continuous data, lax data model design, poor metadata management, and dynamic user requirements.
- ▶ *Contribution:* I built a schema completion tool by re-implementing Google’s WebTables to identify data with similar schemas across two large datasets: the Google WebTables dataset, consisting of over 1 billion records, and the CDIAC dataset, containing 0.5 million files of environmental science data. The results indicate that the WebTables tool struggles to efficiently identify correct synonyms, even in relaxed cases like the Top-20.

8. Benchmarking Private Blockchains

Role: Core Member. **Affiliation:** National University of Singapore.

- ▶ *Objective:* Evaluating the performance of popular private blockchains and providing insights for further development and deployment of blockchains.
- ▶ *Contribution:* I was a major contributor to BlockBench, the first evaluation framework designed to analyze the performance of private blockchains. I deployed Hyperledger Fabric (versions 0.6.0 and 1.0.0) in a private network, implemented APIs, and conducted performance evaluations. Additionally, I made significant contributions to a comprehensive technical survey of private blockchain platforms, including Ethereum and Hyperledger, providing valuable insights into their architectures, capabilities, and trade-offs.

PUBLICATIONS

Conference Papers

- [1] **Rui Liu**, Paul Lin, Evan Weinberg, Dhruva Kulkarni, John Stile, Zhengji Zhao, Ian Karlin, Garima Kochhar, Brian Austin, “Counters Can Tell More: Workload-Scale Performance Modeling and Prediction using Passive Telemetry”, Under Preparing for **ISC High Performance**. 2027.
- [2] **Rui Liu**, Paul Lin, Evan Weinberg, Dhruva Kulkarni, Zhengji Zhao, Ian Karlin, Garima Kochhar, Brian Austin, “Towards MPI Communication Modeling From Passive Network Telemetry”, Workshop on HPC Operational Data Analytics (**HPC-ODA@SC**), Chicago, USA. 2026.
- [3] Peiyuan Dai, **Rui Liu**, Heng Zhang, “NanoCSV: A High-Performance CSV Processing Framework with Hierarchical Finite-State Transducer”, International Conference on Database Systems for Advanced Applications (**DASFAA**), Singapore. 2025.
- [4] **Rui Liu**, Aaron J. Elmore, Michael J. Franklin, Sanjay Krishnan, “Riveter: Adaptive Query Suspension and Resumption Framework for Cloud Native Databases”. IEEE International Conference on Data Engineering (**ICDE**), Utrecht, Netherlands. 2024.
- [5] **Rui Liu**, Jun Hyuk Chang, Riki Otaki, Zhe Heng Eng, Aaron J. Elmore, Michael J. Franklin, Sanjay Krishnan, “Towards Resource-adaptive Query Execution in Cloud Native Databases”, Conference on Innovative Data Systems Research (**CIDR**), Santa Cruz, USA. 2024.
- [6] **Rui Liu**, Kwanghyun Park, Fotis Psallidas, Xiaoyong Zhu, Jinghui Mo, Rathijit Sen, Matteo Interlandi, Konstantinos Karanasos, Yuan Yuan Tian, Jesús Camacho-Rodríguez, “Optimizing Data Pipelines for Machine Learning in Feature Stores”, VLDB Endowment, vol. 16, no. 13, pp. 4230–4239. Presentation at International Conference on Very Large Data Bases (**VLDB**), Guangzhou, China. 2024.
- [7] **Rui Liu**, Aaron J. Elmore, Michael J. Franklin, Sanjay Krishnan, “Rotary: A Resource Arbitration Framework for Progressive Iterative Analytics”, IEEE International Conference on Data Engineering (**ICDE**), Anaheim, USA. 2023.
- [8] Minzhao Liu, Junyu Liu, **Rui Liu**, Henry Makhanov, Danylo Lykov, Anuj Apte, Yuri Alexeev, “Embedding Learning in Hybrid Quantum-Classical Neural Networks”, IEEE International Conference on Quantum Computing and Engineering (**QCE**), Broomfield, USA. 2022.

- [9] **Rui Liu**, David Wong, Dave Lange, Patrik Larsson, Vinay Jethava, Qing Zheng, “Accelerating Container-based Deep Learning Hyperparameter Optimization Workloads”, Workshop on Data Management for End-to-End Machine Learning (**DEEM@SIGMOD**), Philadelphia, USA. 2022.
- [10] Heng Zhang, Lingda Li, Hang Liu, Donglin Zhuang, **Rui Liu**, Chengying Huan, Shuang Song, Dingwen Tao, Yongchao Liu, Charles He, Yanjun Wu, Shuaiwen Leon Song, “Bring Orders into Uncertainty: Enabling Efficient Uncertain Graph Processing via Novel Path Sampling on Multi-Accelerator System”, ACM International Conference on Supercomputing (**ICS**), Virtual. 2022.
- [11] **Rui Liu**, Sanjay Krishnan, Aaron J. Elmore, Michael J. Franklin, “Understanding and Optimizing Packed Neural Network Training for Hyper-Parameter Tuning”, Workshop on Data Management for End-to-End Machine Learning (**DEEM@SIGMOD**), Virtual. 2021.
- [12] Will Brackenbury, **Rui Liu**, Mainack Mondal, Aaron J. Elmore, Blase Ur, Kyle Chard, Michael J. Franklin, “Draining the Data Swamp: A Similarity-based Approach”, Workshop on Human-In-the-Loop Data Analytics (**HILDA@SIGMOD**), Houston, USA. 2018.
- [13] Tien Tuan Anh Dinh, Ji Wang, Gang Chen, **Rui Liu**, Beng Chin Ooi, Kian-Lee Tan, “BLOCKBENCH: A Framework for Analyzing Private Blockchains”, ACM International Conference on Management of Data (**SIGMOD**), Chicago, USA. 2017.
- [14] Zhe Zhou, Wenrui Diao, Xiangyu Liu, Zhou Li, Kehuan Zhang, **Rui Liu**, “Vulnerable GPU Memory Management: Towards Recovering Raw Data from GPU”, Privacy Enhancing Technologies Symposium (**PETS**), Minneapolis, USA. 2017.
- [15] **Rui Liu**, Jiannong Cao, Sebastian VanSyckel, Wenyu Gao, “PriMe: Human-centric Privacy Measurement based on User Preferences towards Data Sharing in Mobile Participatory Sensing Systems”, IEEE International Conference on Pervasive Computing and Communications (**PerCom**), Sydney, Australia. 2016.
- [16] **Rui Liu**, Jiannong Cao, Lei Yang, Kehuan Zhang, “PriWe: Recommendation for Privacy Settings of Mobile Apps Based on Crowdsourced Users’ Expectations”, IEEE International Conference on Mobile Services (**MS**), New York, USA. 2015.

Journal Articles

- [1] Sanjay Krishnan, Aaron J. Elmore, Michael J. Franklin, John Paparrizos, Zechao Shang, Adam Dziedzic, **Rui Liu**, “Artificial Intelligence in Resource-Constrained and Shared Environments”, ACM SIGOPS Operating Systems Review (**OSR**), vol. 53, no. 1, pp. 1-6, 2019.
- [2] Ruiyun Yu, Jiannong Cao, **Rui Liu**, Wenyu Gao, Xingwei Wang, Junbin Liang, “Participant Incentive Mechanism towards Quality-Oriented Sensing: Understanding and Application”, ACM Transactions on Sensor Networks (**TOSN**), vol. 15, no. 2, pp. 21:1-21:25, 2019.
- [3] **Rui Liu**, Junbin Liang, Jiannong Cao, Kehuan Zhang, Wenyu Gao, Lei Yang, Ruiyun Yu, “Understanding Mobile Users’ Privacy Expectations: A Recommendation-based Method through Crowdsourcing”, IEEE Transactions on Services Computing (**TSC**), vol. 12, no. 2, pp. 304-318, 2019.
- [4] Tien Tuan Anh Dinh, **Rui Liu**, Meihui Zhang, Gang Chen, Beng Chin Ooi, Ji Wang, “Untangling Blockchain: A Data Processing View of Blockchain Systems”, IEEE Transactions on Knowledge and Data Engineering (**TKDE**), vol. 30, no. 7, pp. 1366–1385, 2018.
- [5] **Rui Liu**, Junbin Liang, Jiannong Cao, Weiping Zhu, Sebastian Vansyckel, Christian Becker, “UIO-based Testbed Augmentation for Simulating Cyber-Physical Systems”, **IEEE Intelligent Systems**, vol. 33, no. 6, pp. 69–86, 2018.
- [6] **Rui Liu**, Jiannong Cao, Kehuan Zhang, Wenyu Gao, Lei Yang, Junbin Liang, “When Privacy Meets Usability: Unobtrusive Privacy Permission Recommendation System for Mobile Apps based on Crowdsourcing”, IEEE Transactions on Services Computing (**TSC**), vol. 11, no. 5, pp. 864–878, 2018.
- [7] Wenrui Diao, **Rui Liu**, Xiangyu Liu, Zhe Zhou, Zhou Li, Kehuan Zhang, “Accessing Mobile User’s Privacy based on IME Personalization: Understanding and Practical Attacks”, Journal of Computer Security (**JCS**), vol. 26, no. 3, pp. 283–309, 2018.

- [8] Junbin Liang, Jiannong Cao, **Rui Liu**, Tao Li, “Distributed Intelligent MEMS: A Survey and a Real-time Programming Framework”, *ACM Computing Surveys (CSUR)*, vol. 49, no. 1, pp. 20:1–20:29, 2016.
- [9] Ruiyun Yu, **Rui Liu**, Xingwei Wang, Jiannong Cao, “Improving Data Quality with an Accumulated Reputation Model in Participatory Sensing Systems”, *Sensors*, vol. 14, no. 3, pp. 5573-5594, 2014.

Posters / Demonstrations

- [1] Heng Zhang, Lingda Li, Donglin Zhuang, **Rui Liu**, Shuang Song, Dingwen Tao, Yanjun Wu, Shuaiwen Leon Song, “An Efficient Uncertain Graph Processing Framework for Heterogeneous Architectures”, *ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP)*, Virtual. 2021. [Poster]
- [2] Shan Jiang, Junbin Liang, Jiannong Cao, **Rui Liu**, “An Ensemble-level Programming Model with Real-Time Support for Multi-Robot Systems”, *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Sydney, Australia. 2016. [Demo]

Patents

- [1] Jesus Camacho Rodriguez, Kwanghyun Park, Fotis Psallidas, Xiaoyong Zhu, Jinghui Mo, Rathijit Sen, Matteo Interlandi, Yuanyuan Tian, **Rui Liu**, Konstantinos Karanasos, “Feature Store Data Preparation Optimization”, *US Patent App*, 18/332,118. 2024.

Technical Reports

- [1] Yao Lu, Song Bian, Lequn Chen, Yongjun He, Yulong Hui, Matthew Lentz, Beibin Li, Fei Liu, Jialin Li, Qi Liu, **Rui Liu**, Xiaoxuan Liu, Lin Ma, Kexin Rong, Jianguo Wang, Yingjun Wu, Yongji Wu, Huanchen Zhang, Minjia Zhang, Qizhen Zhang, Tianyi Zhou, Danyang Zhuo, “Computing in the Era of Large Generative Models: From Cloud-Native to AI-Native”, Technical Report, <https://arxiv.org/pdf/2401.12230.pdf>. 2024.

RECOGNITION

- ✿ ATPESC Fellow, Argonne National Laboratory, 2026
- ✿ AI 2000 Most Influential Scholar Award Honorable Mention (Database Area), *AMiner*, 2022, 2025, 2026
- ✿ Travel Award of ICDE, National Science Foundation, 2023
- ✿ CERES Graduate Scholar Fellowship, University of Chicago, 2017
- ✿ Third Place, Oracle Java Programming Contest in China, 2015
- ✿ Honorable Mention, COMP Research Competition, Hong Kong Polytechnic University, 2013
- ✿ PolyU Research Postgraduate Scholarship, Hong Kong Polytechnic University, Oct 2012 – Dec 2015
- ✿ Outstanding Graduate, Liaoning Province of China, 2012
- ✿ Google Excellence Scholarship (China), Google, 2011
- ✿ First Class Scholarship, Northeastern University (China), 2011
- ✿ First Prize, China National Undergraduate Mathematical Contest in Modeling, 2010
- ✿ Third Place, IBM Mainframe Technology Application Contest in China, 2010
- ✿ Best Mainframe Application, Citi Finance and Information Technology Contest in China, 2009

TEACHING & MENTORSHIP

Mentorship

- | | |
|---|----------|
| ✿ Mentor , NERSC Open Hackathon, <i>National Energy Research Scientific Computing Center</i> | Aug 2026 |
| ✿ Project Lead , Intro to HPC Bootcamp, <i>Argonne National Laboratory</i> | Aug 2026 |

🔑 **Mentor**, TACC Open Hackathon, *Texas Advanced Computing Center*

Oct 2025

Teaching Assistantship

🔊 Introduction to Databases, <i>University of Chicago</i>	2019, 2020, 2021, 2022, 2023
🔊 Introduction to Data Science, <i>University of Chicago</i>	2018
🔊 Mobile Computing, <i>Hong Kong Polytechnic University</i>	2014
🔊 Introduction to Internet Computing, <i>Hong Kong Polytechnic University</i>	2013
🔊 Information Technology, <i>Hong Kong Polytechnic University</i>	2013

Invited Lecturer

🔊 Introduction to Big Data Platforms, <i>Hong Kong Polytechnic University</i>	Feb 2022
---	----------

MEDIA COVERAGE

- 🔊 [Disseminate: The Computer Science Research Podcast \(EP 49\).](#)
- 🔊 [The Morning Paper for Blockbench.](#)

Last Updated: September 7, 2026