Hariharan (Hari) Devarajan

PERSONAL INFORMATION

Email : hariharandev1@llnl.gov

Phone : +1(312)383-9795

Website : https://hariharan-devarajan.github.io/

Social Media : https://www.linkedin.com/in/dhariharan/

Repos : https://github.com/hariharan-devarajan | https://github.com/llnl-asr | https://bitbucket.org/scs-io/

RESEARCH STATEMENT

Storage and data movement are becoming increasingly complex in modern supercomputing environments due to heterogeneous hardware, diverse programming environments, and rapidly evolving scientific workloads. Modern workflows combine large-scale simulations, data analytics, and AI pipelines, all of which move massive volumes of data across layered software stacks and heterogeneous storage infrastructures. Understanding and optimizing data movement in such environments is critical for enabling scalable and reliable scientific discovery.

Challenge 1: How do we understand workflow data movement and performance across heterogeneous tasks, programming environments, software stacks, compute devices, and storage systems?
Challenge 2: How do we connect performance signals across these heterogeneous components to track the lineage and lifecycle of scientific data?
Challenge 3: How can we use this information to enable targeted optimizations instead of today's trial-and-error tuning approaches?

My research aims to address these challenges by developing systems that combine performance monitoring, data lineage analysis, and intelligent optimization for scientific workflows. First, we develop **eBPF-based telemetry mechanisms** that capture fine-grained performance signals enriched with workload context, enabling scalable monitoring and explainability of I/O and storage behavior across heterogeneous HPC systems. Then, we establish **causal relationships between workflow characteristics and system performance**, enabling tracking of data lineage and lifecycle across complex software and hardware layers. Finally, we build **runtime optimization pipelines** that analyze workflow telemetry to identify bottlenecks and apply targeted optimizations during execution.

By enabling explainable monitoring, data lineage tracking, and targeted optimization, this research aims to transform how scientists interact with data-intensive workflows on future supercomputing systems. The ability to understand and optimize data movement across heterogeneous infrastructures will improve workflow efficiency, enhance reproducibility, and accelerate data-driven scientific discovery across simulation, AI, and analysis workloads.

SUMMARY

- Computer scientist specializing in data movement, storage systems, and performance analysis for large-scale scientific and AI workflows in high-performance computing (HPC).
- Research focuses on explainable I/O, workflow-aware data systems, and end-to-end performance optimization, enabling scientists to understand and optimize data movement across heterogeneous compute, storage, and software stacks.
- Develops workload-aware monitoring, data lineage tracking, and intelligent runtime optimization systems to improve scalability, reproducibility, and efficiency of data-intensive scientific workflows.
- Creator and lead developer of open-source research infrastructure including DLIO Benchmark and DFTracer, widely used for benchmarking and analyzing deep learning I/O performance in HPC environments.
- Published extensively in top HPC venues (SC, IPDPS, HPDC, CCGrid, Cluster) with multiple best paper awards, and actively contributes to the community through peer review, conference service, and collaborative research initiatives.

RESEARCH AREAS

- **Explainable I/O.** Explaining *why* I/O is slow by connecting low-level storage telemetry with workflow context and the anatomy of the storage stack: DataCrumbs (eBPF-based multi-layer I/O telemetry), Metis (agentic knowledge synthesis for explainable I/O, REX-IO'26), and the XIO vision (SSDBM'25).
- **Performance Monitoring and Analysis.** Scalable tracing and automated analysis of workflow data movement: DFTracer (SC'24), DFAnalyzer, WisIO (ICS'25), HORATIO (SSDBM'26), IOMax (SC-W'23), Recorder, and SmartIO (PDSW'25).
- **Provenance for Scientific Workflows (new).** Moirai (LDRD, PI), a provenance tool that tracks the lineage and lifecycle of data across workflow tasks, software layers, and storage systems.
- **Deep Learning I/O Characterization and Optimization.** From monitoring to characterization to optimization of AI training pipelines: DLIO (CCGrid'21, Best Paper; MLPerf Storage), GLANCED-IO (HPDC'26), CoDL (REX-IO'26), Stimulus (CCGrid'22), DYAD (SBAC-PAD'24), and checkpointing for large language models.
- **Workload-aware Middleware and Storage Libraries.** Hermes (HPDC'18), UnifyFS (IPDPS'26; R&D 100), DYAD (SBAC-PAD'24), WADO (SSDBM'26), HCL (Cluster'20), LABIOS (HPDC'19, Best Paper), and intent-driven I/O (Mimir, H5Intent, IFlux).

EDUCATION

Illinois Institute of Technology — Chicago, IL, USA
Ph.D. in Computer Science. CGPA: 3.66/4 — 2016–2021

National Institute of Technology — Jamshedpur, India
B.Tech (Hons) in Computer Science and Engineering. CGPA: 8.61/10 — 2009–2013

PROPOSALS AWARDED

- **Co-Principal Investigator**, "DLDL: Deep Learning Data Loading."
- **Principal Investigator**, "Trusted Data Provenance for HPC Workflows (Moirai)," LDRD Exploratory Research 27-ERD-025 (3 years).
- **Thrust Lead for Data Management**, "FRACTALE: multi-cluster scheduling and data management," LDRD Strategic Initiative (from Oct 2023).
- **Thrust Lead for Data Management**, "Heuretes: cross-site data movement under time constraints," LDRD Strategic Initiative.

PUBLICATIONS

Conferences:

- Olga Kogiou, **Hariharan Devarajan**, Chen Wang, Weikuan Yu, and Kathryn Mohror. “iCR: Intent-Driven Adaptable Checkpoint/Restart”. *2026 IEEE International Conference on Cluster Computing (CLUSTER’26)*. Alexandria, VA, USA: IEEE, Sept. 2026.
- Karim Youssef, **Hariharan Devarajan**, Nikoli Dryden, and Roger Pearce. “WADO: A Distributed WORM Storage Service for Asynchronous Data Operations”. *Proceedings of the 38th International Conference on Scalable Scientific Data Management (SSDBM’26)*. ACM, Aug. 2026, pp. 1–12.
- Ray A. O. Sinurat, **Hariharan Devarajan**, Nikoli Dryden, Haryadi S. Gunawi, and William Nixon. “HORATIO: Bridging Management and Analysis of Traces at Scale”. *Proceedings of the 38th International Conference on Scalable Scientific Data Management (SSDBM’26)*. ACM, Aug. 2026, pp. 1–12.
- Ray A. O. Sinurat, William Nixon, Philip Carns, Huihuo Zheng, Sandeep Madireddy, Sam Foreman, Troy Arcomano, Robert B. Ross, Haryadi S. Gunawi, and **Hariharan Devarajan**. “GLANCED-IO: Taming I/O Optimization for Deep Learning at Scale”. *Proceedings of the 35th International Symposium on High-Performance Parallel and Distributed Computing (HPDC’26)*. ACM, July 2026, pp. 305–318.
- Izzet Yildirim, **Hariharan Devarajan**, Anthony Kougkas, Xian-He Sun, and Kathryn Mohror. “WisIO: Automated I/O Bottleneck Detection with Multi-Perspective Views for HPC Workflows”. *Proceedings of the 39th ACM International Conference on Supercomputing (ICS’25)*. ACM, June 2025, pp. 749–763.
- Sarah Neuwirth, **Hariharan Devarajan**, Chen Wang, and Jay Lofstead. “XIO: Toward eXplainable I/O for HPC Systems”. *Proceedings of the 37th International Conference on Scalable Scientific Data Management (SSDBM’25)*. ACM, June 2025, pp. 1–6.
- Yiheng Xu, Pranav Sivaraman, **Hariharan Devarajan**, Kathryn Mohror, and Abhinav Bhatele. “ML-based Modeling to Predict I/O Performance on Different Storage Sub-systems”. *31st IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC)*. IEEE, Dec. 2024.
- **Hariharan Devarajan**, Loic Pottier, Kaushik Velusamy, Huihuo Zheng, Izzet Yildirim, Olga Kogiou, Weikuan Yu, Anthony Kougkas, Xian-He Sun, Jae Seung Yeom, and Kathryn Mohror. “DFTracer: An Analysis-Friendly Data Flow Tracer for AI-Driven Workflows”. *SC24: International Conference for High Performance Computing, Networking, Storage and Analysis*. Atlanta, GA: IEEE, Nov. 2024.
- **Hariharan Devarajan** and Kathryn Mohror. “TailorFS: An Adaptive file system to support dynamic I/O requirements of HPC workloads”. *2024 IEEE 36th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*. IEEE, Nov. 2024, pp. 81–92.
- **Hariharan Devarajan**, Ian Lumsden, Chen Wang, Konstantia Georgouli, Tom Scogland, Jae-Seung Yeom, and Michela Taufer. “DYAD: Locality-aware Data Management for accelerating Deep Learning Training”. *2024 IEEE 36th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*. IEEE, Nov. 2024, pp. 13–24.
- **Hariharan Devarajan** and Kathryn Mohror. “Mimir: Extending I/O Interfaces to Express User Intent for Complex Workloads in HPC”. *2023 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. St. Petersburg, Florida: IEEE, May 2023.
- **Hariharan Devarajan** and Kathryn Mohror. “Extracting and characterizing I/O behavior of HPC workloads”. *2022 IEEE International Conference on Cluster Computing (CLUSTER’22)*. Heidelberg, Germany, Sept. 2022.
- **Hariharan Devarajan**, Anthony Kougkas, Huihuo Zheng, Venkatram Vishwanath, and Xian-He Sun. “Stimulus: Accelerate Data Management for Scientific AI applications in HPC”. *2022 IEEE/ACM International Symposium in Cluster, Cloud, and Internet Computing (CCGrid’22)*. Taormina, Italy, May 2022.
- Jaime Cernuda Garcia, **Hariharan Devarajan**, Luke Logan, Keith Bateman, Neeraj Rajesh, Jie Ye, Anthony Kougkas, and Xian-He Sun. “HFlow: A Dynamic and Elastic Multi-Layered Data Forwarder”. *2021 IEEE International Conference on Cluster Computing (CLUSTER’21)*. Portland, OR: IEEE, Sept. 2021.
- Neeraj Rajesh, **Hariharan Devarajan**, Jaime Cernuda Garcia, Keith Bateman, Luke Logan, Jie Ye, Anthony Kougkas, and Xian-He Sun. “Apollo: An ML-assisted Real-Time Storage Resource Observer”. *Proceedings of the 30th International Symposium on High-Performance Parallel and Distributed Computing (HPDC’21)*. New York, NY, USA: Association for Computing Machinery, June 2021, pp. 147–159.
- **Hariharan Devarajan**, Huihuo Zheng, Anthony Kougkas, Xian-He Sun, and Venkatram Vishwanath. “DLIO: A Data-Centric Benchmark for Scientific Deep Learning Applications”. *2021 21st IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*. IEEE, May 2021. **Best Paper**.
- **Hariharan Devarajan**, Anthony Kougkas, and Xian-He Sun. “HReplica: A Dynamic Data Replication Engine with Adaptive Compression for Multi-Tiered Storage”. *2020 IEEE International Conference on Big Data (Big Data)*. Atlanta, Georgia, USA, Dec. 2020.
- **Hariharan Devarajan**, Anthony Kougkas, Keith Bateman, and Xian-He Sun. “HCL: Distributing Parallel Data Structures in Extreme Scales”. *2020 IEEE International Conference on Cluster Computing (CLUSTER)*. IEEE, Sept. 2020.
- **Hariharan Devarajan**, Anthony Kougkas, Luke Logan, and Xian-He Sun. “HCompress: Hierarchical Data Compression for Multi-Tiered Storage Environments”. *2020 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. New Orleans, Louisiana, USA, May 2020.
- **Hariharan Devarajan**, Anthony Kougkas, and Xian-He Sun. “HFetch: Hierarchical Data Prefetching for Scientific Workflows in Multi-Tiered Storage Environments”. *2020 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. New Orleans, Louisiana, USA, May 2020.
- Anthony Kougkas, **Hariharan Devarajan**, and Xian-He Sun. “Bridging Storage Semantics using Data Labels and Asynchronous I/O”. *ACM Transactions on Storage (TOS)* 1.1 (2020).

- Anthony Kougkas, **Hariharan Devarajan**, Keith Bateman, Jaime Cernuda, Neeraj Rajesh, and Xian-He Sun. “ChronoLog: A Distributed Shared Tiered Log Store with Time-based Data Ordering”. *Proceedings of the 36th International Conference on Massive Storage Systems and Technology (MSST 2020)*. 2020.
- K. Feng, **H. Devarajan**, A. Kougkas, and X.-H. Sun. “NIOBE: An Intelligent I/O Bridging Engine for Complex and Distributed Workflows”. *IEEE International Conference on Big Data*. Dec. 2019.
- Anthony Kougkas, **Hariharan Devarajan**, Jay Lofstead, and Xian-He Sun. “LABIOS: A Distributed Label-Based I/O System”. *Proceedings of the ACM 28th International Symposium on High-Performance Parallel and Distributed Computing (HPDC’19)*. June 2019. **Best Paper**.
- **Hariharan Devarajan**, Anthony Kougkas, and Xian-He Sun. “An Intelligent, Adaptive, and Flexible Data Compression Framework”. *Proceedings of the IEEE/ACM International Symposium in Cluster, Cloud, and Grid Computing (CCGrid’19)*. May 2019.
- **Hariharan Devarajan**, Anthony Kougkas, Prajwal Challa, and Xian-He Sun. “Vidya: Performing Code-Block I/O Characterization for Data Access Optimization”. *Proceedings of the IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC)*. Dec. 2018.
- Anthony Kougkas, **Hariharan Devarajan**, Xian-He Sun, and Jay Lofstead. “Harmonia: An Interference-Aware Dynamic I/O Scheduler for Shared Non-Volatile Burst Buffers”. *Proceedings of the IEEE International Conference on Cluster Computing (Cluster’18)*. Sept. 2018.
- Anthony Kougkas, **Hariharan Devarajan**, and Xian-He Sun. “Hermes: A Heterogeneous-Aware Multi-Tiered Distributed I/O Buffering System”. *Proceedings of the ACM 27th International Symposium on High-Performance Parallel and Distributed Computing (HPDC’18)*. June 2018.
- Anthony Kougkas, **Hariharan Devarajan**, and Xian-He Sun. “IRIS: I/O Redirection via Integrated Storage”. *Proceedings of the 32nd ACM International Conference on Supercomputing (ICS’18)*. June 2018.

Journals:

- **Hariharan Devarajan**, Vanessa Sochat, Daniel J. Milroy, and Tom Scogland. “IFlux: Intent-Aware Storage Tiers & Software Scheduling for HPC Systems”. *IEEE Transactions on Parallel and Distributed Systems* (2026), pp. 2236–2251.
- **Hariharan Devarajan**, Gerd Heber, and Kathryn Mohror. “H5Intent: Autotuning HDF5 with User Intent”. *IEEE Transactions on Parallel and Distributed Systems* (2024).
- Anthony Kougkas, **Hariharan Devarajan**, and Xian-He Sun. “I/O Acceleration via Multi-Tiered Data Buffering and Prefetching”. *Journal of Computer Science and Technology* (2019).

Workshops:

- **Hariharan Devarajan**, Sarah Neuwirth, Chen Wang, Ray A. O. Sinurat, Neeraj Rajesh, Jay Lofstead, Vanessa Sochat, Daniel Milroy, and Thomas Scogland. “DataCrumbs: Efficient Cross-Layer Data Capture for Explainable I/O Performance in HPC Storage Stacks”. *Proceedings of the SC’26 Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC Workshops’26): 8th Workshop on Programming and Performance Visualization Tools (ProTools’26)*. Chicago, IL, USA: IEEE, Nov. 2026.
- **Hariharan Devarajan**, Ray A. O. Sinurat, Izzet Yildirim, Olga Kogiou, Aniruddha Marathe, and Nikoli Dryden. “ORION: Observability Framework for Hybrid Workflow-Centric I/O Monitoring in HPC Systems”. *Proceedings of the SC’26 Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC Workshops’26): 8th Workshop on Programming and Performance Visualization Tools (ProTools’26)*. Chicago, IL, USA: IEEE, Nov. 2026.
- Karim Youssef, Sarah Neuwirth, Neeraj Rajesh, and **Hariharan Devarajan**. “Metis: Agentic Knowledge Synthesis for Explainable I/O Performance in HPC Systems”. *6th Workshop on Re-envisioning Extreme-Scale I/O for Emerging Hybrid HPC Workloads (REX-IO’26)*. ACM, July 2026.
- Druva Dhakshinamoorthy, Ray A. O. Sinurat, Nikoli Dryden, Arnab K. Paul, and **Hariharan Devarajan**. “CoDL: A Framework for Studying Cross-Component Interference in Deep Learning Training Pipelines”. *6th Workshop on Re-envisioning Extreme-Scale I/O for Emerging Hybrid HPC Workloads (REX-IO’26)*. ACM, July 2026.
- **Hariharan Devarajan**, Brian Behlendorf, Blake Devcich, Dean Roehrich, Matt Richerson, Tony Floeder, Brent Petit, Don Boyd, Trace Hamada, Tim McCree, Gian-Carlo Defazio, James Corbett, Elsa Gonsiorowski, Cameron Stanavige, Nikoli Dryden, Roger Pearce, Kathryn Mohror, and Bronis R. de Supinski. “Chasing the Rabbit: A Systematic Exploration of Software-Defined HPC Storage”. *Proceedings of the 7th Workshop on Extreme-Scale Storage and Analysis (ESSA’26), held in conjunction with the 40th IEEE International Parallel and Distributed Processing Symposium (IPDPS’26)*. IEEE, May 2026.
- Hammad Bin Ather, Chen Wang, **Hariharan Devarajan**, Hank Childs, and Kathryn Mohror. “SmartIO: A Lightweight End-to-End Workflow for Runtime I/O Optimization of HPC Systems”. *Proceedings of the Parallel Data Storage Workshop (PDSW’25), held in conjunction with SC25: The International Conference for High Performance Computing, Networking, Storage and Analysis*. ACM, Nov. 2025, pp. 1396–1405.
- Ian Lumsden, **Hariharan Devarajan**, Izzet Yildirim, Stefano Markidis, Andong Hu, Ivy Bo Peng, Luca Pennati, Dewi Yokelson, Stephanie Brink, Olga Pearce, Tom Scogland, Bronis R. de Supinski, Gian Luca Delzanno, Anthony Kougkas, Xian-He Sun, and Michela Taufer. “Optimizing I/O for an Exascale Implicit Kinetic Plasma Simulation using the Rabbit Storage System”. *5th Workshop on Re-envisioning Extreme-Scale I/O for Emerging Hybrid HPC Workloads (REX-IO’25), 2025 IEEE International Conference on Cluster Computing Workshops (CLUSTER Workshops)*. IEEE, Sept. 2025, pp. 1–6. **LLNL Director’s Institutional Award**.
- Olga Kogiou, **Hariharan Devarajan**, Chen Wang, Weikuan Yu, and Kathryn Mohror. “Understanding Highly Configurable Storage for Diverse Workloads”. *4th Workshop on Re-envisioning Extreme-Scale I/O for Emerging Hybrid HPC Workloads (REX-IO)*. IEEE, Sept. 2024.

- **Hariharan Devarajan**, Adam Moody, Donglai Dai, Cameron Stanavige, Elsa Gonsiorowski, Marty McFadden, Olaf Faaland, Greg Kosinovsky, and Kathryn Mohror. "The impact of asynchronous I/O in checkpoint-restart workloads". *5th Workshop on Extreme-Scale Storage and Analysis (ESSA)*. IEEE, May 2024.
- Ian Lumsden, **Hariharan Devarajan**, Jack Marquez, Stephanie Brink, David Boehme, Olga Pearce, Jae-Seung Yeom, and Michela Taufer. "Empirical Study of Molecular Dynamics Workflow Data Movement: DYAD vs. Traditional I/O Systems". *23rd IEEE International Workshop on High Performance Computational Biology (HICOMB)*. IEEE, May 2024.
- Izzet Yildirim, **Hariharan Devarajan**, Anthony Kougkas, Xian-He Sun, and Kathryn Mohror. "IOMax: Maximizing Out-of-Core I/O Analysis Performance on HPC Systems". *Workshops of the International Conference on High Performance Computing, Network, Storage, and Analysis (SC-W 2023)*. Denver, CO, USA: ACM, Nov. 2023, pp. 1209–1215.
- Anthony Kougkas, **Hariharan Devarajan**, and Xian-He Sun. "Enosis: Bridging the Semantic Gap between File-based and Object-based Data Models". *Proceedings of the ACM SIGHPC Datacloud'17, 8th International Workshop on Data-Intensive Computing in the Clouds*. Nov. 2017.
- Anthony Kougkas, **Hariharan Devarajan**, and Xian-He Sun. "Syndesis: Mapping Objects to Files for a Unified Data Access System". *Proceedings of the ACM SIGHPC MTAGS'17, 10th International Workshop on Many-Task Computing on Clouds, Grids, and Supercomputers, in conjunction with SC'17*. Nov. 2017.
- **Hariharan Devarajan**, Anthony Kougkas, Hsing-Bung Chen, and Xian-He Sun. "Open Ethernet Drive: Evolution of Energy-Efficient Storage Technology". *Proceedings of the ACM SIGHPC Datacloud'17, 8th International Workshop on Data-Intensive Computing in the Clouds, in conjunction with SC'17*. Nov. 2017.

Posters and Short Papers:

- Ian Lumsden, Stefano Markidis, Andong Hu, Ivy Bo Peng, Luca Pennati, Dewi Yokelson, Stephanie Brink, Olga Pearce, Tom Scogland, **Hariharan Devarajan**, Bronis R. de Supinski, Gian Luca Delzanno, and Michela Taufer. "Performance Optimization of an Exascale Implicit Kinetic Plasma Simulation on El Capitan". *2025 IEEE International Conference on eScience (eScience)*. IEEE, Sept. 2025, pp. 319–320.
- Olga Kogiou, **Hariharan Devarajan**, Chen Wang, Weikuan Yu, and Kathryn Mohror. "I/O characterization and performance evaluation of large-scale storage architectures for heterogeneous workloads". *2023 IEEE International Conference on Cluster Computing Workshops (CLUSTER Workshops)*. IEEE, Oct. 2023, pp. 44–45.

Preprints:

- Chen Wang, Izzet Yildirim, **Hariharan Devarajan**, Kathryn Mohror, and Marc Snir. *Recorder: Comprehensive Parallel I/O Tracing and Analysis*. Jan. 2025. arXiv:2501.04654.

Under Review and In Preparation:

- "FluxFiction." Under review at the Workshop on Performance Modeling, Benchmarking and Simulation (PMBS) at SC26.
- "A4X." Under review at the 27th ACM/IFIP International Middleware Conference.
- "Ferry." Under review at the USENIX Annual Technical Conference.
- "Cladia," "SigML," "GraphIOX," and "LiveFlow" (in submission).

Publication Distinctions:

- Best Student Paper Award from the Computer Science Department at Illinois Tech 2021.
- Best Paper Award at the 21st IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing CCGrid (May'21)
- Karsten Schwan Best Paper Award at the 28th International ACM Symposium on High-Performance Parallel and Distributed Computing – HPDC (June'19)
- Citation Rate:
  - All citations: 432 with h-index: 12

PEER REVIEW

- **Journals:** 20 reviews for IEEE TPDS, IEEE TKDE, ACM Computing Surveys, and JCST over 5 years (2020–2024).
- **Conferences:** 150 reviews for SC, IPDPS, HPDC, IEEE Cluster, CCGrid, ICS, and ICPP over 6 years (2021–2026).
- **Workshops:** 47 reviews for REX-IO, ESSA, ReWords, Practically FAIR, PERMAVOST, and LLMxHPC over 6 years (2021–2026).

Review Board:

- Review Board Member for Transactions on Parallel and Distributed Systems.
- Co-editor of "Advancements in Extreme-Scale I/O, Storage Systems, and Data Analytics" Issue in Frontiers Journal.

INVITED TALKS AND PANELS

- 2021 Dagstuhl Seminar on Data Management
- Data management workshop in Advanced Scientific Computing Research (ASCR) 2021.
- “Challenges in Characterizing I/O in Large Scientific Workflows” in the 2nd Workshop on Re-envisioning Extreme-Scale I/O for Emerging Hybrid HPC Workloads (REX-IO) 2022
- “DLIO Benchmark: Data-driven benchmark for deep learning workload in HPC” in storage seminar at Lawrence Berkley National Laboratory 2022.
- “DLIO Benchmark: Data-driven benchmark for deep learning workload in HPC” in storage seminar at Centre for High Performance Computing, South Africa National Meeting 2023
- Talk on Intent driven HDF5 optimization at The HDF User Group Meeting 2023
- Invited talk at ALCF Seminar at Argonne National Laboratory 2024
- “DLIO Benchmark and Profiler” in storage seminar at NCSA 2024
- Talk for AI4S seminar at 2024 Scientific Machine Learning Research Group, Rutherford Appleton Laboratory, Science and Technology Facilities Council (STFC), UK
- “Optimizing Molecular Dynamics AI model using HDF5 and DYAD” at The HDF User Group Meeting (HUG) 2024
- Talk on “Analyzing I/O in Deep Learning Workloads” at Birds of Feather session panel in SC’24.
- “XIO: Towards eXplainable I/O using multi-layer monitoring solution for HPC Systems.” in storage seminar at Centre for High Performance Computing, South Africa National Meeting 2024
- Talk on “HDF5 Explainability with eBPF” at Birds of Feather session panel in SC’24.
- Panelist, “Storage Architectures and I/O Optimizations for AI Applications,” PDSW’25 at SC’25
- Keynote, “Why is my I/O Slow? Rethinking Storage Systems for Explainability,” 6th Workshop on Re-envisioning Extreme-Scale I/O for Emerging Hybrid HPC Workloads (REX-IO’26) at HPDC 2026, Cleveland, OH, July 13, 2026.
- Panelist, “Data Centers: Information, Misinformation, and the Future of Large-Scale Computing,” 35th ACM International Symposium on High-Performance Parallel and Distributed Computing (HPDC 2026), Cleveland, OH, July 15, 2026 (moderator: A. Maccabe).
- Invited talk on Rabbit software-defined storage, Livermore Computing seminar, LLNL, 2026.
- Mini-symposium talk on Rabbit software-defined storage, SIAM Conference on Computational Science and Engineering (CSE27).
- Contributor, DOE ASCR Principal Investigator Meeting, 2026.
- **Hariharan Devarajan**. “Skills Before Smarts: Harness Engineering for Robust Software-Development Agents.” Talk, 1st Workshop on ML for Assisting Code Quality (MLAC 2026) at the International Conference on Parallel Architectures and Compilation Techniques (PACT 2026), Discovery Partners Institute, Chicago, IL, Oct. 19, 2026.
- Panelist, “Who Gets to Do Computing Research in 2036?” International Conference on Parallel Architectures and Compilation Techniques (PACT 2026), Discovery Partners Institute, Chicago, IL, Oct. 20, 2026 (moderator: M. Taufer; panelists: E. J. Park, V. Pascucci, H. Devarajan, T. Malik, and L. Rauchwerger).
- **Hariharan Devarajan**, Ray A. O. Sinurat, Izzet Yildirim, Olga Kogiou, Aniruddha Marathe, and Nikoli Dryden. “ORION: Observability Framework for Hybrid Workflow-Centric I/O Monitoring in HPC Systems.” Paper presentation, 8th Workshop on Programming and Performance Visualization Tools (ProTools’26) at SC26, Chicago, IL, Nov. 16, 2026.
- **Hariharan Devarajan**, Sarah Neuwirth, Chen Wang, Ray A. O. Sinurat, Neeraj Rajesh, Jay Lofstead, Vanessa Sochat, Daniel Milroy, and Thomas Scogland. “DataCrumbs: Efficient Cross-Layer Data Capture for Explainable I/O Performance in HPC Storage Stacks.” Paper presentation, 8th Workshop on Programming and Performance Visualization Tools (ProTools’26) at SC26, Chicago, IL, Nov. 16, 2026.
- **Hariharan Devarajan**, Sarah Neuwirth, and Jay Lofstead. “Explaining HDF5 Performance through the I/O Stack.” Talk, HDF5 Birds of a Feather session, SC26: International Conference for High Performance Computing, Networking, Storage and Analysis, Chicago, IL, Nov. 2026.
- **Hariharan Devarajan**. “Chasing the Rabbit: Software-Defined Storage in Action.” Demonstration, U.S. Department of Energy booth, SC26: International Conference for High Performance Computing, Networking, Storage and Analysis, Chicago, IL, Nov. 2026.

PROFESSIONAL SERVICE

- Registration and Online Chair for IEEE Cluster Conference 2021.
- Session Chair for IEEE Cluster Conference 2021.
- Poster Judge for IEEE International Parallel and Distributed Processing Symposium 2023.
- Session Chair for IEEE Cluster Conference 2023.
- Web Chair for IEEE Cluster Conference 2023.
- Poster Judge for PhD Forum in IEEE International Parallel and Distributed Processing Symposium 2024.
- Session Chair for IEEE CCGRID Conference 2024.
- Session Chair for IEEE IPDPS Conference 2024.
- Session Chair for IEEE SBACPAD Conference 2024
- Program Committee Member for ICS, SC, IEEE Cluster, HPDC, CCGrid, and ICPP 2025.
- Program Committee Member for PDSW 2025 (at SC25).
- Judge, PhD Forum, 40th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2026).
- Judge, ACM Student Research Competition, International Conference on Parallel Architectures and Compilation Techniques (PACT).
- Program Committee Member for IPDPS 2026, HPDC 2026, IEEE Cluster 2026, CCGrid 2026, ICS 2026, ICPP 2026, ESSA 2026, and REX-IO 2026.
- Posters Co-Chair, 35th ACM International Symposium on High-Performance Parallel and Distributed Computing (HPDC 2026), with S. Neuwirth.
- Posters Co-Chair, IEEE International Conference on Cluster Computing (Cluster 2026), with A. K. Paul.
- Program Co-Chair, 38th International Conference on Scalable Scientific Data Management (SSDBM 2026), San Diego, CA, with D. Moise.

RESEARCH EXPERIENCE

Lawrence Livermore National Laboratory — Livermore, CA, USA

Computer Scientist — Sep 2023 – Current

- **Provenance (Moirai) – Principal Investigator, LDRD**
  - Leading a new effort to track the lineage and lifecycle of scientific data across workflow tasks and storage systems.
- **Deep Learning I/O (DLDL, DLIO, GLANCED-IO, CoDL) – Co-PI, DLDL**
  - Designed workload-driven benchmark models capturing dataset structure, access patterns, preprocessing, and checkpointing behavior.
  - Enabled systematic evaluation of storage systems and I/O pipelines for AI workloads across HPC platforms.
  - Adopted in multiple HPC centers and integrated into the MLCommons Storage Benchmark suite.
  - Delivered a deployable DFTracer module on Livermore Computing and generated the I/O characterization dataset for the Genesis project.
- **Scheduling and Data Management (Heuretes; FRACTALE, past) – Data Management Thrust Lead**
  - Delivered IFlux, DataCrumbs, and Metis, completing the I/O-intent component of the FRACTALE multi-cluster scheduling stack.
  - Investigating multi-cluster and cross-site scheduling and data management.
- **Explainable I/O (XIO, DataCrumbs, Metis)**
  - Built eBPF-based instrumentation to monitor I/O behavior across heterogeneous storage and compute layers.
  - Enriched low-level telemetry with contextual workflow information to enable explainable performance analysis.
  - Developed agentic knowledge synthesis (Metis) to explain I/O performance in HPC systems.
- **Performance Monitoring and Analysis (DFTracer, DFAnalyzer, WisIO, HORATIO)**
  - Lead multi-layer data-flow tracing and scalable trace analysis for AI and scientific workflows.
  - Automated I/O bottleneck detection and trace management at scale.
- **Rabbit Accelerator Performance Characterization and Optimization**
  - Published the first systematic evaluation of Rabbit software-defined storage, establishing the lab-wide performance baseline on El Capitan-class systems (up to 25× speedup on I/O benchmarks); results transferred to DLDL, FRACTALE, and CODA.
  - Optimized I/O of the exascale implicit kinetic plasma code iPIC3D on El Capitan using Rabbits (LLNL Director's Institutional Award).
  - Characterized performance behavior across diverse workloads including simulations and AI training pipelines.
  - Provided insights to guide system configuration and integration of accelerator-backed storage in HPC environments.
- **Asynchronous Data Services (CODA, WADO)**
  - Built WADO, a distributed WORM storage service for asynchronous data operations, and explored DPU offloading.
- **Scalable Checkpoint and Restart (SCR)**
  - Investigating the performance of checkpoint and restart in large language models.
  - Investigating the configurability of SCR with complex multi-stage workflows.

Lawrence Livermore National Laboratory — Livermore, CA, USA

Post Doctoral Researcher — June 2021 – Sep 2023

- **I/O Performance and Portability (IOPP)**
  - Experience with characterizing the I/O requirement of AI application in HPC systems.
  - Experience with building a comprehensive taxonomy of AI applications.
  - Experience with designing I/O systems which utilizes user's preferences and transparently optimizes I/O in a multi-tenant HPC system.
  - Invented I/O intents for storage system and demonstrated its benefits in designing workload-aware storage systems.
- **Scalable Checkpoint and Restart (SCR)**
  - Experience with the internal architecture of SCR.
  - Study of the impact of Async I/O on applications using SCR
- **DYnamic and Asynchronous Data Streamliner (DYAD)**
  - Experience with sharing data across producer-consumer workflows.
  - Demonstrating the need for efficient caching through DYAD in deep learning workloads.
- **GOTCHA**
  - Maintainer of the GOTCHA toolkit.
  - Experience with application compilation, execution, and assembly-level instructions.
  - Contributed to the GOTCHA toolkit to support newer Glibc kernes.

Illinois Institute of Technology — Chicago, IL, USA

Research Assistant — Aug 2016 – May 2021

- **AI in HPC:** Accelerating data management for Scientific AI workloads.
  - Experience with characterizing I/O behavior of Scientific AI application in ALCF.
  - Experience with data management in existing AI frameworks such as TensorFlow and PyTorch.
  - Experience with profiling I/O and data operations in AI frameworks.
  - Currently working on accelerating data ingestion in AI workloads using deep memory and storage hierarchy.
- **I/O Integration:** Convergence of the semantic gap between the **Cloud and HPC** ecosystems.
  - Experience with various data representations of storage solutions for applications.
  - Designed novel transformations between object and file data representations.
  - Deployed various software solutions such as OrangeFS, Lustre, Hyperdex, MongoDB, and redis.
- **Multi-tiered storage**: Design **I/O optimizations** for new **multi-tiered** environments.
  - Experience with various storage devices along with kernel models to deploy NVMe-over-fabric.
  - Deployed I/O hierarchical buffering VOL software for HDF5 library on Cori supercomputer.
  - Designed a data-centric hierarchical prefetching solution for HPC environments.
- LABIOS: A Label-based I/O System
  - Experience with building decoupled architecture and deploying it at scale.
  - Designed and developed the label based system and deployed it on Chameleon Cloud.
- **Data Compression**: Study the variability of **compression performance** for different data characteristics and requirements.
  - Experience with many lossless data compression libraries and their usage.
  - Unified all compression libraries under a single framework with a novel decision engine.
  - Deployed the compression framework on multiple nodes on Theta supercomputer.
- **Profiling**: Design an **application-profiler** to estimate the application's I/O based on its **source code**.
  - Experience with open compilers such as GCC and LLVM.
  - Designed python based tool to extract I/O features from application source code.
  - Developed a linear regression model to predict I/O intensity based on source code features.
- Designed and developed **a distributed data structure library** (HCL).
  - Experience with RDMA technologies such as IBVerbs, RoCE, NVMe over fabric.
  - Experience with parallel programming languages (e.g., Chappel and UPC++).

Argonne National Laboratory — Lemont, IL, USA

Research Aid — May 2020 – Aug 2020

- Understanding data access and processing API in TensorFlow and PyTorch
- Explored various profiling tools for high-level framework profiling (e.g., TensorBoard) and low-level I/O profiling (e.g., Darshan)
- Characterized I/O behavior for AI applications in ALCF.
- Built an analysis tool VaniDL which combines Darshan Traces with TensorBoard and provides a high-level analysis of I/O behavior.

The HDF Group — Champaign, IL, USA

Research Intern — May 2018 – Aug 2018

- Designed framework for Hierarchical Buffering Platform as a VOL Plugin in HDF5.
- Tested the new software on **Cori Supercomputer**.

PROJECTS

Current Projects:

- **Moirai** – Principal Investigator, "Trusted Data Provenance for HPC Workflows" (LDRD ER 27-ERD-025). Provenance tool that tracks the lineage and lifecycle of data across scientific workflows.
- **DLDL** (Deep Learning Data Loading) – Co-Principal Investigator. Benchmarking, monitoring, and optimizing deep learning I/O pipelines on HPC systems through DLIO and DFTracer.
- **Heuretes** – Thrust Lead for all data management (LDRD Strategic Initiative). Cross-site data movement and scheduling under time constraints.
- **XIO** – I/O explainability and the anatomy of HPC storage stacks (DataCrumbs, Metis).
- **GOTCHA** – Lead developer and maintainer (since June 2023) of a mission-critical LLNL tool for function interposition, supporting applications, software, and tools across the DOE and NNSA ecosystem.

Past Projects:

- **FRACTALE** – Thrust Lead for Data Management (LDRD Strategic Initiative, from Oct 2023). Multi-cluster scheduling and data management; built IFlux (intent-aware storage tiers and scheduling in Flux) and DataCrumbs.
- **CODA** (Compute on Demand Asynchronous) – Built WADO, a distributed WORM storage service for asynchronous data operations, and explored offloading data services to DPUs.
- **IOPP** (I/O Performance and Portability) – Lead researcher (June 2021 – Sep 2023) for Dr. Kathryn Mohror's DOE ASCR Early Career Award; introduced I/O intents for workload-aware storage systems.

CODE DEVELOPMENT WITH IMPACT

DLIO Benchmark

44 GitHub Stars

https://github.com/argonne-lcf/dlio_benchmark

Jan 2021 – Current

- Used as a benchmark for **acceptance testing and debugging for Aurora** for AI workloads in ALCF.
- It is used a benchmark by many researchers in Germany (Radita Liam and Sarah Neuwirth)
  - Jean Luca Bez and Suren Byna Lawrence Berkeley National Laboratory, USA
  - Takaaki Fukai National Institute of Advanced Industrial Science and Technology (AIST), Tokyo
- It is part of **MLPerf benchmark suite for benchmarking Storage** (https://mlcommons.org/working-groups/benchmarks/storage/)

DFTracer: Data Flow Tracing Infrastructure for AI Workflows

https://github.com/llnl/dftracer

- Provides multi-layer tracing of deep learning I/O pipelines, capturing data movement across frameworks, libraries, and system calls.
- Used by researchers to analyze data movement in AI training pipelines and scientific workflows on large-scale HPC systems.
- Used by MLCommons MLPerf Storage working group for analyzing and understanding I/O performance of benchmark submissions.
- Adopted at Argonne Leadership Computing Facility (ALCF) and Oak Ridge Leadership Computing Facility (OLCF) for deep learning I/O performance analysis and workflow characterization.

Software Releases:

- DFTracer v1.0.0 – v1.0.4 (2024) and v2.0.2 (2026), including a deployable Livermore Computing module for turnkey multi-node I/O and compute tracing of production workflows.
- GOTCHA v1.0.6, v1.0.7 (2024), and v1.0.8 (2025).
- DLIO Benchmark v2.0.0 (2024).

PROFESSIONAL EXPERIENCE

Oracle Financial Services Software

Mumbai, India

Application Developer 2

Sept 2013 – July 2016

- Designed framework for **REST API** for product's middleware impacting cloud integrations.
- Developed utility for **inter-release** activities for clients enabling migration to a newer version in a day.
- Mentoring new employees on **product framework and design**, helping them deliver faster.
- Awarded "**We Applaud Award**" for designing a framework affecting the whole organization.

FUNDING AND MENTORING

Funding:

- **Co-Investigator**, LDRD Exploratory Research 25-ERD-042 with N. Dryden (three years, FY25–FY27).
- **LLNL Technical Point of Contact** for two university subcontracts: University of Tennessee, Knoxville (PI M. Taufer; student D. Dhakshinamoorthy) and Illinois Institute of Technology (PI X.-H. Sun; student N. Rajesh); one more pending for Fall 2026.
- Led the data management thrust (Thrust 2) of a DOE ASCR proposal with J. Hittinger (2025).

Mentoring:

- 2026: Mentored four summer interns, who presented five posters.
- 2025: Supervised two staff (K. Youssef and A. Marathe) and mentored six students (O. Kogiou, Harish, N. Rajesh, D. Dhakshinamoorthy, I. Yildirim, and R. A. O. Sinurat).
- 2024: Mentored five students, who submitted three SC posters.

HONORS AND AWARDS

- LLNL Director's Institutional Award for "Optimizing I/O for an Exascale Implicit Kinetic Plasma Simulation using the Rabbit Storage System" (iPIC3D on El Capitan, REX-IO 2025).
- R&D 100 Award for UnifyFS, 2024.
- Outstanding service Award at Cluster 2021 for organizing the virtual platform and managing registrations for the conference.
- Best Student Paper Award from the Computer Science Department at Illinois Tech 2021.
- Best Paper Award at the 21st IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing CCGrid (May'21)
- Best Poster Nominee at the Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis 2019.
- Karsten Schwan Best Paper Award at the 28th International ACM Symposium on High-Performance Parallel and Distributed Computing – HPDC (June'19)
- Awarded "We Applaud Award" for designing a framework affecting the whole organization at Oracle Financial Services Software.