William C. Benton

willb@acm.org
willb

Select Professional Experience

September 2020– **Principal Architect**, *NVIDIA Corporation*.
Provided technical and strategic guidance across product group functions. Prototyped product features that were adopted by engineering teams. Enabled technology integrations behind joint go-to-market efforts with MLOps ecosystem partners. Led NVIDIA's *data science blueprints* program, which produced didactic example accelerated machine learning systems that show NVIDIA technologies, partner technologies, and ecosystem projects working together to solve real business problems. Designed experiments and developed arguments to show the value of accelerated computing and NVIDIA DGX for LLM inference. Represented NVIDIA in industry consortia, in high-level customer and partner discussions, and at peer-reviewed conferences.

May 2019– September 2020 **Engineering Manager and Senior Principal Software Engineer**, *Red Hat, Inc.*, Office of the CTO, AI Center of Excellence.
Founded and led a team with the mission to make it easier to build machine learning systems on hybrid clouds. Managed a distributed team of engineers and data scientists. Collaborated with leadership in Red Hat's services organization to define services offerings for machine learning (ML) on Kubernetes. Developed training content on data engineering, ML workflows on Kubernetes, and challenges of ML systems and platforms; presented peer-reviewed talks and tutorials at industry and academic conferences. Served as a ML and AI expert in customer executive briefings. Helped customers put ML into production on Kubernetes in industries including insurance, energy, and financial services. Built open-source MLOps tooling to streamline the experiment-to-production lifecycle on Kubernetes.

October 2016– April 2019 **Senior Principal Software Engineer**, *Red Hat, Inc.*.
Developed and popularized the *intelligent applications* concept for ML systems on Kubernetes, which influenced strategy and execution across multiple business units in Red Hat. Served as overall engineering lead for the *radanalytics.io* open-source community, which enabled intelligent applications, machine learning frameworks, and data engineering workloads on Kubernetes and OpenShift. Implemented and refined groupwide engineering processes (e.g., Kanban and retrospectives), coached and mentored across and beyond the group, and directed the efforts of a global team of data scientists and engineers. Served as a data science and ML subject matter expert on interview panels for roles across Red Hat.

April 2015– October 2016 **Principal Software Engineer**, *Red Hat, Inc.*.
Led a team focused on emerging technology and data science consulting; worked with product and engineering management to develop and refine AI/ML strategy.

September 2008 – March 2015 **Senior Software Engineer**, *Red Hat, Inc.*.
Led a data science team focused on internal consulting. Led emerging-technology efforts around open-source analytics and ML frameworks. Contributed to many open-source projects including Apache Spark, the SBT build tool, and the HTCondor resource manager. Designed and implemented numerous open-source projects related to distributed application development, distributed configuration management, probability and simulation, and distributed ML and data engineering. Mentored new contributors to the Fedora Project.

Summer 2005 – Fall 2007 **Teaching Assistant/Lecturer**, *University of Wisconsin*, Computer Sciences.
Lectured and developed course materials for introductory and senior-level courses.
Awarded departmental (2006) and university-wide (2007) teaching honors.

October 2005 – April 2006 **Independent Consultant**.
Prototyped a novel audio synthesis technique for a music-technology startup.

Fall 2000 – Spring 2003 **Research Assistant**, *University of Wisconsin*, Computer Sciences.
Supervised by Prof. Barton Miller. Assisted with development and maintenance of the Paradyn performance analysis tool and the Dyninst dynamic instrumentation library. Ported Paradyn from userspace threads to POSIX threads. Identified a bug in the SGIIRIX kernel.

Select Publications

2023 Benton, William C. and Sophie Watson. "Distributed anomaly detection using combinable measurement value summaries." United States Patent 11,719,563.

Benton, William Christian and Erik Jordan Erlandson. "Generation of predictive model metrics using intermediate operations." United States Patent 11,810,011.

Benton, William Christian and Erik Jordan Erlandson. "Usage-based software library decomposition." United States Patent 11,550,553.

2022 Benton, William and Sophie Watson. "Determining validity of multipart branching literate programs." United States Patent 11,416,222.

2020 Benton, William C. "Machine Learning Systems and Intelligent Applications." In *IEEE Software* 37.4, pp. 43–49.

2019 Benton, William Christian. "Distributed function generation with shared structures." United States Patent 10,241,767.

Benton, William Christian and Erik Jordan Erlandson. "Repeatable distributed pseudorandom number generation." United States Patent 10,474,432.

2011 Benton, William C, Robert H Rati, and Erik J Erlandson. "Wallaby: a scalable semantic configuration service for grids and clouds." In *SC'11: Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*.

2009 Benton, William C. and Charles N. Fischer. "Mostly-functional behavior in Java programs." In *International Conference on Verification, Model Checking, and Abstract Interpretation*.

2008 Benton, William Christian. *Fast, effective program analysis for object-level parallelism*. PhD dissertation. University of Wisconsin–Madison.

2007 Benton, William C. and Charles N. Fischer. "Interactive, scalable, declarative program analysis: from prototype to implementation." In *Proceedings of the 9th ACM SIGPLAN International Conference on Principles and Practice of Declarative Programming*.

Select Professional Service

2018– **Program Committee**, *Berlin Buzzwords*.

2014 **Birds-of-a-Feather Committee**, *The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC14)*.

Education

PhD (computer sciences), *University of Wisconsin–Madison*.
My dissertation focused on identifying implicit module-level parallelism in Java programs. I developed (1) a declarative framework for developing static program analyses, (2) a type-and-effect system and analysis to infer methods that can safely execute in parallel, and (3) an evaluation of dynamic opportunities to exploit such parallelism.

BA with honors (philosophy, music, computer science), *St. Olaf College*.