

Resume (abridged)

Name: Dhruva Krishnamurthy
Email: dhruvakm@gmail.com
Phone: +1 (408) 623-0605
Location: California, United States

- Resume: <https://mechanicker.github.io/resume.html>
- GitHub profile: <https://github.com/mechanicker>
- LinkedIn profile: <https://www.linkedin.com/in/dhruvakm/>

Summary

Principal Software Engineer with over 20 years of expertise in designing, developing, and optimizing complex, large-scale distributed systems, particularly focused on high-performance storage solutions and IO-intensive applications. Proven ability to lead initiatives delivering substantial performance gains, cost savings, and enhanced system robustness in cloud (AWS) and on-premise environments. Deep experience in systems programming (C, C++, Go, Python), performance troubleshooting, distributed caching, file systems (NFS), and git internals at scale. Seeking roles leveraging deep technical skills in challenging distributed storage and performance-critical domains.

Skills

- Distributed Systems & Cloud-Native Microservices (AWS)
- High-Performance Storage Systems & File Systems (NFS, POSIX IO)
- IO-Intensive Application Development & Optimization
- Performance Analysis, Troubleshooting & Debugging (Unix/Windows, Distributed Systems)
- Systems Programming (C, C++, Go, Python, Java)
- Git & libgit2 Internals (Scaling, Performance, Robustness)
- Scalable Data Pipelines, CI/CD & Data Lifecycle Management
- Distributed Caching Strategies (Memcached, Redis, BerkeleyDB)

Experience

Principal engineer, [Atlassian](#)

Jan 2017 - Current

Focus Areas: Led initiatives enhancing Bitbucket's performance, scalability, and data management, specifically addressing challenges of distributed git repository storage (NFSv3) at petabyte scale on AWS. Deep involvement in IO performance optimization, distributed caching, large-scale data migration, and git internals.

Key Contributions:

- **Distributed Storage Performance & Scaling:**

- Architected and optimized storage solutions for large-scale monorepos, improving high-frequency git operation performance through custom caching and IO pattern analysis.
- Designed and implemented IO latency-based dynamic throttling and novel git cache invalidation techniques to enhance resource utilization and stability under heavy load.
- Led the design of git bundle-uri service, significantly reducing NFS load and costs by serving snapshots from AWS S3.
- Reduced latency for IO-intensive operations via custom `libgit2` caching backend (`memcached`), read-ahead logic, and improved `fscache` /page cache utilization.
- Prototyped `gRPC` consistency hashing service for highly available distributed git repository access.

- **Petabyte-Scale Data Management & Storage Optimization:**

- Architected and oversaw live migration of ~2.5 petabytes of git data across storage backends with minimal customer impact, leveraging AWS services.
- Designed unified repository archival/backup system using tiered storage (AWS S3). Led efforts moving ~70% of data to cheaper tiers.
- Implemented cross-region git repository replication meeting RPO SLAs.
- Designed scale-out storage sharding for large monorepos, increasing IO throughput.
- Proposed novel algorithms for fencing writes during repository migrations.

- **Git Internals & Troubleshooting:**

- Maintained Bitbucket's git fork, implementing enhancements for robustness, crash recovery, and observability in distributed storage environments.
- Contributed performance-related patches upstream to `libgit2` (e.g., `mmap` emulation for NFS consistency).
- Developed tools like `puntar` for parallel database restore, optimizing IO.

- **Mentoring & skill building:**

- Passionate about cultivating engineering talent, I led initiatives like the Systems Chapter to deepen expertise
- Created specialized hiring processes for systems roles, and actively coached junior engineers to accelerate their careers
- My focus is on empowering the next generation of engineers to innovate and assume greater responsibilities.

- **Other:** Developed scalable data pipelines for user search (GDPR compliance) and identity management systems.

Senior engineer, [NetApp](#)

Feb 2008 - Jan 2017

Focus Areas: Specialized in NFS server performance optimization, IO-intensive workload acceleration (SAP Hana), and architecting scale-out distributed NAS storage solutions. Deep C++/systems programming for performance analysis and improvement.

Key Contributions:

- **NFSv4 Performance & IO Optimization:**
 - Led initiative improving NFSv4 server performance, achieving ~40% IO performance gain via in-kernel buffer caching. Established performance baselines and identified optimization targets.
 - Designed and developed the `IOtrap` library for transparent `asyncio` interception, boosting IO throughput by ~50% for SAP Hana over NFSv4 workloads, critical for SAP certification.
- **Distributed Scale-Out NAS Storage (`Infinite Volume`):**
 - Core architect for `Infinite Volume`, contributing to object storage components and distributed file system metadata search using BerkeleyDB.
 - Designed and tuned a high-performance (achieved ~5x improvement) distributed query engine using Intel TBB for complex metadata searches.
 - Prototyped and benchmarked kernel-space porting of BerkeleyDB for performance analysis.
- **Storage Performance Analysis & Management:**
 - Implemented statistical performance analysis methods (C++ Boost) for NetApp Performance Advisor.
 - Developed a prototype for scale-out distributed storage management (scaling from 50 to 500+ controllers).

Technical Specialist, [McAfee](#)

Feb 2006 - Feb 2008

- Led performance engineering team (4 engineers), establishing performance measurement methodologies.
- Improved multi-threaded antivirus scan rate by ~25% by implementing custom memory allocators to reduce lock contention, leveraging learnings from Intel VTune training.

Technical Specialist, HP

Jan 2006 - Feb 2007

- Led the porting of Samba (CIFS file server) to VMS, implementing POSIX API emulation layer required for complex systems software.

Engineering Manager, [Bosch](#)

Dec 2004 - Dec 2005

- Led cross-site team developing data compiler pipeline for navigation systems, achieving significant processing time reduction (weeks to days) through optimizations.

Technical Lead, [Delmia \(Dassault Systemes\)](#)

Feb 1998 - Dec 2004

- Developed core C++ features for 3D simulation software (CATIA V5). Ported LXR code browsing tool to Windows.

Production Engineer, Wipro Fluid Power

Oct 1996 - Feb 1998

- Early career role applying software (C, dBase) to solve manufacturing problems, including CNC simulation and inventory tracking.

Awards and recognition

Patent 10812313: Federated namespace of heterogeneous storage system namespaces

- Covers federated file system namespace over distributed, heterogeneous storage with policy-based data lifecycle management. Directly relevant to distributed storage architecture.

Recognition at work

- Performance improvements using Async IO (~\$30k/month savings).
- Cloud-scale `git` object caching implementation lowering NFS IO.
- Improved cache utilization reducing server fleet costs.
- NetApp Hackathon 2nd place (filed a patent).
- Award for implementing embedded file system metadata search service (NetApp).

Education & certifications

- Bachelor of Engineering, Mechanical
- Intel VTune Performance Profiler Training (Applied at McAfee, NetApp)
- Unix & C Systems Programming Training

Other projects and links to referred articles

- Patent 10812313: <https://patents.justia.com/patent/10812313>
- `IOtrap` library (IO throughput improvement): <https://github.com/mechanicker/iotrap>
- `puntar` (Parallel unarchiving tool): <https://github.com/mechanicker/puntar>
- `cramp` (Win32 function call profiler): <https://github.com/mechanicker/cramp>
- NetApp best practices for SAP Hana using `asyncio`: <https://www.netapp.com/media/8991-tr4290.pdf>