Alexey Shvechkov

Greater Boston, MA | +1.774.278.1743 | GitHub: github.com/shvechkov | LinkedIn: www.linkedin.com/in/alexey-shvechkov

Summary

Technical leader & Principal Architect with 20+ years designing and delivering distributed systems, cloud-native platforms, and AI-driven solutions. Led $20M+ enterprise product portfolios from strategy to delivery, accelerating time-to-market by 40% and influencing architecture for teams of up to 100 engineers, QA, and support staff. Hands-on leader skilled in C++, Go, Rust, and Python, combining deep technical expertise with strategic vision. Experienced in AI/ML integration, infrastructure modernization, and mentoring global teams to deliver customer-focused, scalable solutions. Open to principal-level roles that demand direct coding contributions alongside architectural leadership, driving innovation in distributed platforms, cloud infrastructure, and AI solutions.

Core Competencies

- **Cloud & Distributed Systems:** AWS (EC2, S3, Lambda, API Gateway), Kubernetes, Docker, Terraform, scalable microservices, high-availability architectures, CI/CD pipelines, fault tolerance
- **Programming Languages:** C++, Go, Rust, Python, Perl, Bash
- **AI/ML:** Generative AI, RAG, transformers (LoRA fine-tuning), PyTorch, TensorFlow, Hugging Face, LangChain, ONNX Runtime, distributed training, anomaly detection
- **Systems Programming:** Linux/Windows kernel drivers (eBPF, KMDF/WDM, minifilters), file systems, performance tuning
- **Databases & Storage:** PostgreSQL, MySQL, Redis, Cassandra, object storage systems
- **Leadership & Strategy:** Technical strategy, agile execution, product ownership, cross-functional influence, mentoring (10+ years), roadmap planning, cost optimization
- **Frontend & APIs:** React, Next.js, REST API design, rapid prototyping

Professional Experience

Arcserve (spun out from CA Technologies) — Greater Boston, MA
Technical Director, Software Engineering | 2014–Present
- Directed technical strategy and architecture for multiple enterprise products (Replication & High Availability, OneXafe, CloudDirect) generating $20M+ annual revenue.
- Reduced delivery cycles by 40% through agile practices and direct coding contributions, enabling a Linux-based immutable object-store to ship 6 months earlier.
- Provided architectural guidance influencing 100+ engineering, QA, and support staff; directly managed teams of up to 10 engineers across geographies.
- Architected and coded components and features for high-availability replication, data deduplication, and agentless protection for AWS/Azure workloads, using C++ and Go.
- Provided technical leadership for cloud-native and AI initiatives, reporting to CTO/VP, and influenced M&A projects through rapid prototyping.
- Developed and optimized transformer-based LLMs for edge inference with ONNX Runtime (for anomaly detection and agentic features), integrating RAG-based chatbots with legacy systems via REST and MCP protocols.
- Won multiple internal hackathons for AI/infrastructure prototypes, helping accelerate the innovation pipeline.

CA Technologies (formerly Computer Associates) — Greater Boston, MA
Principal Engineer | 2006–2014
- Led post-acquisition integration of XOsoft's replication technologies into CA's enterprise portfolio.
- Designed and developed high-availability replication, deduplication, and backup features using C/C++ for Windows and Linux.
- Built a Unix/Linux build automation framework with web UI, reducing manual labor by 50%.

XOsoft *(Acquired by CA Technologies in 2006)* — Israel
Lead Software Engineer | 2000–2006

- Designed and developed CDN components, including network management and monitoring tools, and high-traffic HTTP/FTP caching proxies, along with replication solutions.
- Led the development of kernel-mode drivers and file system modules in C/C++, improving system performance and reliability.

Yandex — Moscow, Russia
Software Engineer | 1998–2000
- Developed high-traffic search, advertising network and web applications using C++ and Perl, optimizing performance for large-scale systems.

Open-Source Projects/POCs

- s3stor – deduplicating archiving /backups into S3 compatible storage (Go) - github.com/shvechkov/s3stor
- gos3rve – exposing local file systems via s3 APIs (Go) - github.com/shvechkov/gos3rve
- ufc – fast unique file copy/ indexes and stores unique files (Rust) - github.com/shvechkov/ufc

Education

- M.Sc., Applied Mathematics | Institute of Telecommunications and Computer Science, Russia | 1996
- Machine Learning Specialization (Stanford Online) | 2024 (Certificate)
- Holder of multiple U.S. patents in storage and infrastructure technologies