

Aakash Shankar

☎ +1 (551)-375-9829 | @ aakashshankar217@gmail.com | linkedin.com/in/ahhcash | github.com/ahhcash | ahhcash.xyz

PROFESSIONAL SUMMARY

Software engineer with experience in fullstack engineering and distributed systems. Demonstrated success building high performance, robust, and scalable solutions for enterprise environments.

EXPERIENCE

Stealth Startup

June 2024 – Aug 2024

Software Engineer Intern

Remote, United States

- Engineered a system for customizing AI-generated data warehouse schema documentation, resulting in **90% improvement** in column definition accuracy and enhanced query comprehension
- Led vector database migration from ChromaDB to LanceDB, **reducing storage footprint by 40%** while enabling both semantic and keyword search capabilities for efficient data retrieval

Cloudera

Jan 2021 – Jul 2023

Software Engineer II

Bangalore, India

- Designed and implemented RESTful APIs for a rapid autoscaling system for **cloud-based (AWS, Azure, GCP)** big data clusters using **Java microservices**, reducing query processing time by **70%** and **saving enterprise customers over \$1M in cloud expenditure**
- Leveraged **CI/CD pipelines** (Jenkins/Gradle, Spinnaker) for automated build, testing, and deployment of microservices onto **AWS EKS**, improving deployment frequency and stability
- Developed an automated tracking system for cloud resource scaling, streamlining customer infrastructure management and **improving operational efficiency by 30%**
- Architected an **autonomous healing system** for big data clusters that automatically detected and replaced unhealthy nodes, ensuring consistent cluster performance and eliminating manual recovery procedures
- Led production support and incident response during **on-call rotations**, utilizing **Elasticsearch with KQL and Lucene queries** for efficient log analysis, performance troubleshooting, and rapid issue resolution
- Enhanced observability of distributed data processing using **Spring Actuator** and propagating telemetry data to **Prometheus and Datadog**, ensuring comprehensive visibility for troubleshooting microservices and database operations
- Authored **open source contributions to cloudbreak**, a platform to automate deployment and management of big data clusters on public cloud
- Engineered a comprehensive API suite using **Spring Boot** to enable programmatic configuration of big data services (HDFS, Spark, Hive), reducing manual post-deployment configuration and **improving cluster setup time by 20%**
- Organized technical design reviews and **cross-functional collaboration** with product and design teams, driving feature releases from design to production following Agile methodologies

PROJECTS

GhastlyDB, a key-value vector store 🗄️ | *Go, LSM Trees, Distributed Systems, DB Internals, gRPC*

2025

- Architected a **key-value vector database** using Log-Structured Merge (LSM) trees, achieving efficient data persistence with concurrent read-write operations and optimized storage management
- Implemented high-performance storage engine with **skiplist index structures** enabling fast lookup and range scan operations with transactional consistency guarantees

dRAG, a RAG-based documentation chatbot 🗄️ | *Python, Vector DBs, PgVector, Langchain, Prefect*

2025

- Developed a **documentation chatbot** enabling natural language interactions with technical documentation through **retrieval-augmented generation**
- Built a **scalable document processing pipeline** using Prefect workflow orchestration to efficiently chunk, index, and retrieve information from large-scale technical documentation

EDUCATION

New York University

New York, NY

Master of Science, Computer Science (GPA: 3.9 / 4.0)

Expected May 2025

Graduate Course Assistant - CSGY 9163 Application Security

SKILLS

Languages: Java, Go, Python, SQL, Rust, C/C++, TypeScript

Frameworks & Concepts: Spring, Hibernate, Gradle, Maven, FastAPI, Echo, Cobra, JUnit, Mockito, Tokio, REST, gRPC, Reactor, Next.js, React.js, SvelteKit

Databases & Big Data: PostgreSQL, MySQL, MongoDB, Elasticsearch, PgVector, LanceDB, HDFS, Spark, Hive

DevOps & Cloud: AWS, Azure, Google Cloud, Docker, Kubernetes, Linux, Jenkins, Spinnaker, Git, CI/CD

Observability: Prometheus, Datadog, Spring Actuator