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, AI and distributed systems. Demonstrated success building secure, multi-tenant cloud-native solutions leveraging AWS, Azure, and Kubernetes.

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 big data clusters using Spring microservices, reducing query processing time by **70%** and **saving enterprise customers over $1M in cloud expenditure**
- 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 and rapid issue resolution
- Enhanced observability of distributed data processing using Spring Actuator and propagating telemetry data to **Prometheus and Datadog**, ensuring comprehensive visibility into microservices and database operations
- Organized technical design reviews and **cross-functional collaboration** with product and design teams, driving feature releases from design to production
- Engineered a comprehensive API suite using Spring MVC to enable programmatic configuration of big data services (HDFS, Spark, Hive), reducing manual post-deployment configuration and **improving cluster setup time by 20%**

Projects

GhastlyDB, a key-value vector store | *Go, LSM Trees, Distributed Systems, DB Internals* 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
- Engineered query optimization mechanisms and a robust similarity search engine supporting multiple distance metrics (cosine, L2, dot product) for efficient nearest neighbor lookup in high-dimensional vector space

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
- Engineered high-performance vector retrieval system using **PostgreSQL and PgVector** for efficient similarity search, with RESTful APIs for document ingestion and query processing

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, C/C++, Rust, TypeScript
Frameworks: Spring, FastAPI, Echo, Cobra, JUnit, Mockito, Tokio, Reactor, Next.js, SvelteKit
Databases & Big Data: HDFS, Spark, Hive, PostgreSQL, MySQL, MongoDB, ElasticSearch, PgVector, LanceDB
DevOps & Cloud: Linux, Docker, Kubernetes, CI/CD, AWS, Azure, Google Cloud
Observability: Prometheus, Datadog, Spring Actuator