Harshil Donga

harshildonga9@gmail.com| linkedin.com/in/harshil-donga/ | github.com/Harshil2120

(404)4960652 | Atlanta, GA

EDUCATION

Georgia State University

Atlanta, GA

Master of Science in Computer Science

Aug 2023 - May 2025

University Of Pune

Pune, India

Bachelors of Engineering in Computer Engineering

Aug 2017 - May 2021

TECHNICAL SKILLS

Front-End Skills: Html, Css, Javascript, ReactJS, NextJS, TailwindCss, Bootstrap, ChartJS

Back-End Skills: Java, Maven, Spring Boot, SpringMVC, NodeJS, Express, Python, Flask, Kafka, Redis, MySQL,

PostgreSQL, MongoDB, GraphQL, Nginx, Junit

Cloud & DevOps: AWS, Azure, Docker, Git, Github, Github Actions CI/CD, Grafana, Loki, Prometheus, K6 Testing

Machine Learning & AI: Ollama, Langchain, Langsmith, VectorDB, Neo4J

EXPERIENCE

AI Software Engineer

Aug 2023 - Apr 2025

SWAN-AI Research Group

Atlanta, GA

- Developed and architected an AI-powered RAG Chatbot application for individuals with anxiety disorders who can use the application to for their mental health, the application had features like user-authentication, real-time notifications, and interactive metrics visualization in the form of graphs.
- Gained experience in deploying **LLMs** on GPU-servers using **Ollama** and working with frameworks like **LangChain** for building AI-driven applications. Additionally, built data annotation tools using **Next.js**, **React Forms**, and **MongoDB** to assist PhD students in collecting high-quality datasets for AI/ML research.

Associate Software Engineer

Sep 2021 - Jun 2023

Accenture Pune, India

- Developed microservices using Test-Driven Development with Maven, Java, Spring Boot and JUnit.
 Integrated third-party APIs such as Twilio and Microsoft EWS API for sending emails and notifications.
- Optimized **REST APIs** performance by **reducing latency** from **295ms** to **85ms** through **Redis** caching. Additionally, documented it using **Swagger**, ensuring clear API specifications for integration and usage.
- Implemented observability using the **LGTM** (**Loki**, **Grafana**, **Tempo**, **Prometheus**) stack for centralized **logging**, **metrics**, and **distributed tracing**, which improved system monitoring, troubleshooting, and performance analysis by monitoring application performance using **Grafana dashboards**.

Projects

Implementation of Consistent Hashing for Distributed Cache Cluster using AWS

- Designed and implemented a customized consistent hashing algorithm as a backend service using Maven,
 Java, Spring Boot, and AWS SDK to distribute incoming requests across a distributed cache cluster for optimal load balancing.
- The service was deployed on **EC2** and integrated with **AWS ElastiCache** and **AWS DynamoDB.** Additionally, **AWS CDK** was used to automate the **deployment** of AWS Services.
- NextJS client-side application to send requests as a client and observe the distribution of traffic.

EaseMind (Mental Health Virtual Assistant)

- Deployed **LLMs (Llama, Deepseek, Llava)** on GPU servers using **Ollama** and integrated **Kafka** as a message queue to efficiently manage multiple user interactions. **WebSockets** were used for real-time communication.
- The chatbot UI was built using **Next.js**, designed with **Tailwind CSS** and **ShadcnUI**, while **Chart.js** was used to visualize metrics in the form of graphs. **NextAuth** was implemented for user authentication.
- **MongoDB** was used for the persistence of chat history between users and LLMs, and **BullMQ**, along with **workers**, was employed to send **real-time notifications** to admins for monitoring conversation toxicity.
- **Docker** was leveraged to containerize the applications, and a **GitHub Actions CI/CD pipeline** was used for deployment, triggered whenever changes were pushed to **GitHub**.

Additional

AWS Certified Cloud Practitioner Docker Foundations Professional Certificate MongoDB Associate Developer Certificate