Aakash Raghunathan

aakshu.goku@gmail.com | +91 9987552227 | linkedin.com/in/aakshu

PROFESSIONAL SUMMARY

Backend Software Engineer with 2+ years of experience designing and deploying **scalable distributed systems**, **AI-powered applications**, and **cloud-native microservices**. Expertise in **Python**, **FastAPI**, event-driven architectures (Kafka), and search technologies (**Elasticsearch**, vector databases). Strong foundation in cloud infrastructure (AWS, GCP), containerization (**Docker**, **Kubernetes**), and AI-related technologies (**GenAI**, **Langchain**) from IBM and Carnegie Mellon's Master of Software Engineering program.

PROFESSIONAL EXPERIENCE

IBM India, Bangalore

Software Engineer - AI Centre of Excellence (AI-COE) | May 2024 – Present

Guide enterprise teams in building AI applications following best practices while experimenting with latest AI technologies. Architect and develop production-grade systems handling real-time data processing, feedback analytics, and intelligent debugging capabilities.

Feedback Hub - AI-Powered Analytics Platform

- Architected and developed end-to-end feedback analytics platform enabling teams to collect, analyze, and act on AI chatbot user feedback using Python, FastAPI, React, and **PostgreSQL**
- Designed **RESTful** API architecture with FastAPI handling 1,000+ requests/day with <100ms average response time, implementing async endpoints, request validation with Pydantic, and comprehensive error handling
- Integrated **LLM** models (GPT-4, Claude) for automated sentiment analysis, feedback categorization, and insight generation, reducing manual analysis time by 75% and providing actionable recommendations

Multi-Source RAG Chatbot - Enterprise Knowledge Assistant

- Built enterprise-grade **RAG** (Retrieval-Augmented Generation) chatbot with connectors for Slack conversations, Box Drive documents, and **GitHub** repositories, enabling natural language queries across organizational knowledge bases
- Architected scalable **data ingestion pipelines** processing 10GB+ of documents daily using concurrent processing patterns, achieving 3x throughput improvement through parallel document parsing and batch indexing
- Developed source-specific adapters handling OAuth authentication, API rate limiting, and incremental synchronization for real-time knowledge base updates across integrated platforms

Debug Assistant - AI Agent for Automated RCA

- Engineered **Agent-to-Agent (A2A**) debugging system with **Model Context Protocol (MCP**) integration to analyze large-scale log files (500MB+) and provide automated root cause analysis with actionable recommendations
- Designed **event-driven microservices** architecture using **Apache Kafka** for asynchronous message processing, enabling distributed log analysis across multiple services with guaranteed message delivery and fault tolerance
- Implemented resilient architecture with retry mechanisms, circuit breakers, and dead-letter queues for failed message handling

Carnegie Mellon University - TEEL Lab, Pittsburgh, PA

Research Assistant (Under Prof. Majd Sakr) | May 2023 – Aug 2023

- Designed comprehensive hands-on project demonstrating monolith-to-microservices migration, encompassing application decomposition, containerization with Docker, and production deployment on Kubernetes clusters
- Developed course curriculum on **Cloud Native Technologies** covering distributed systems patterns (**MapReduce**, **Apache Spark**), container orchestration, service mesh architectures, and cloud storage solutions
- Created practical lab exercises on Kubernetes deployments, **Helm** chart management, and **multi-cloud strategies** for high availability and disaster recovery

EDUCATION

Carnegie Mellon University - School of Computer Science | Pittsburgh, PA

Master of Software Engineering - Scalable Systems **| Dec 2023 | CGPA: 4.00/4.00**

Relevant Coursework: Cloud Computing, Distributed Systems, Software Architecture for Big Data, API Design & Implementation, DevOps & Continuous Integration

NMIMS University | Mumbai, India

Bachelor of Technology - Computer Engineering | MBA in Technology Management | May 2022

TECHNICAL SKILLS

Programming Languages: Python, Java, C++, SQL, Bash/Shell Scripting

Backend Technologies: FastAPI, Flask, RESTful API Design, GraphQL, WebSockets, Microservices Architecture

Databases & Search: PostgreSQL, MongoDB, Elasticsearch, Redis, Vector Databases, Database Optimization

Message Brokers & Streaming: Apache Kafka, Apache Samza, Event-Driven Architecture, Stream Processing

Cloud & Infrastructure: AWS (EC2, Lambda, S3, RDS, API Gateway, CloudFormation), IBM Cloud, GCP (GKE, GCR, Cloud Functions), Terraform, Infrastructure as Code

DevOps & Tools: Docker, Kubernetes, Helm, Git

AI/ML Technologies: LangChain, OpenAI API, Anthropic Claude, RAG Systems, Vector Embeddings, Sentence Transformers, TensorFlow, PyTorch

Software Engineering: System Design, Distributed Systems, Concurrency & Multithreading, Performance Optimization, Test-Driven Development, Agile/Scrum

SELECTED ACADEMIC PROJECTS

AI CoPilot for Finance Platform | Carnegie Mellon University | Sep 2023

- Built serverless AI chatbot on **AWS** leveraging **Lambda functions**, API Gateway, and WebSockets for real-time streaming responses, integrated with **LangChain** and **OpenAI** API for context-aware financial advisory
- Implemented sophisticated prompt engineering with few-shot learning and context-aware retrieval from MySQL database, achieving 85%+ accuracy in personalized financial recommendations
- Designed stateless architecture with **DynamoDB** for session management and S3 for conversation history, supporting concurrent users with auto-scaling Lambda configurations

Real-Time Taxi Matching System with Kafka & Samza | Carnegie Mellon University | Mar 2023

- Engineered distributed stream processing pipeline using Apache Kafka and Apache Samza on AWS to match riders with drivers in real-time based on geospatial proximity, ratings, and user preferences
- Developed Java-based Kafka producers simulating 10,000+ location update events per minute and ride requests, publishing to partitioned topics for parallel consumption
- Implemented stateful stream processing with windowing functions and Samza's key-value store for maintaining driver availability state, optimizing matching algorithms based on distance, gender preferences, ratings, and surge pricing

CERTIFICATIONS

AWS Certified Solutions Architect - Associate | Valid Until: March 2027