

Shreeraj Shetty K

📍 Bangalore, India ✉️ rajshetty727@gmail.com ☎️ +91 8660095291 in shreeraj-shetty-k
🌐 ShreerajShettyK

Profile

Experienced Golang developer specializing in cloud-native applications and GenAI solutions. Proven expertise in building scalable microservices, implementing AI-powered systems and deploying cloud infrastructure on AWS and Azure. Strong background in API development and distributed systems with hands-on experience in LLM integration.

Education

N.M.A.M. Institute of Technology, Nitte Udupi, India
BE in Computer Science and Engineering Sept 2019 – June 2023
◦ GPA: 9.12/10.0

Expert PU College Mangalore, India
XII Science June 2017 – May 2019
◦ Score: 88% in 2nd PU Exam

Poorna Prajna Education Centre Bangalore, India
X High School June 2005 – March 2017
◦ Score: 91.68% in 10th Standard

Work Experience

Associate Software Engineer Bangalore, India
Relanto Sept 2024 – Present
◦ Developed and maintained cloud-native Golang microservices serving high-traffic applications
◦ Designed and implemented RESTful APIs and gRPC services for distributed systems architecture
◦ Built GenAI-powered automation tools and intelligent chatbot systems using LLM integration

Technical Trainee Bangalore, India
Relanto Feb 2024 – Sept 2024
◦ Gained hands-on experience in Golang development, focusing on concurrent programming.
◦ Implemented cloud solutions using AWS services and also aware of Terraform, Jenkins and Docker
◦ Participated in code reviews and agile development practices, contributing to team velocity improvements

Software Intern Bangalore, India
Viyaat Agro LLP LTD July 2023 – December 2023
◦ Worked on an agricultural ERP system used to manage 500 acres of farmland with active users
◦ Migrated the web app from Next.js 12 to Next.js 13 (App router)
◦ Implemented a new feature that allows agronomists to update the admin about tasks assigned to them, including image upload functionality where images are uploaded to Minio storage using external APIs

Projects

Cisco Learning Platform - Learning Assistant Oct 2024 – Present
◦ Developed an intelligent chatbot platform that provides expert-level answers to networking and Cisco product-related queries using advanced LLM models
◦ Built with agentic architecture featuring an orchestrator agent that coordinates multiple specialized micro-agents for enhanced query processing and response accuracy
◦ Implemented intelligent follow-up question suggestions to guide users through complex networking concepts and troubleshooting scenarios

- Technology stack: Golang for backend services, PostgreSQL for persistent data storage, Redis for high-performance caching, and GORM for efficient database operations
- Designed to accelerate learning and problem-solving for network engineers and IT professionals working with Cisco technologies

Azure APIM Gateway for LLM Load Balancing

May 2023 – July 2023

- Designed and implemented Azure API Management (APIM) service to intelligently load-balance requests across multiple Azure OpenAI models deployed in different regions (e.g., GPT-3.5-turbo, GPT-4)
- Developed custom routing policies within the APIM gateway to handle failover scenarios and distribute traffic based on priority. Also configured managed identity for secure authentication
- Integrated backend pools with priority-based routing - prioritizing specific regions until exhaustion, then distributing equally among fallback regions for optimal performance and cost efficiency
- Automated the entire deployment and configuration process using Golang with Azure SDK, enabling programmatic provisioning of APIM resources, policy updates, and backend configurations
- Technology stack: Azure API Management, Azure OpenAI Services, Azure Managed Identity, Azure SDK

Vipasana-Enterprise Resource Planning

July 2023 – Present

- Agricultural ERP system designed to optimize various aspects of agricultural operations, from farm management to food production and distribution
- This web-app is actively serving the needs of managing 500 acres of farmland, with a significant number of active users relying on its features
- Technology stacks: Next.js, React.js, Node.js, Prisma ORM and Minio object storage (similar to Amazon S3)

Skills

Programming Languages: Golang (Expert), C++

GenAI & AI/ML: Large Language Models (LLMs), Agentic Architecture, Azure OpenAI Services, AI Gateway Patterns, Prompt Engineering, Model Integration

Cloud Platforms: AWS (EC2, VPC, S3, Lambda, Bedrock), Azure (APIM, AI Services, Managed Identity, Resource Manager), Multi-cloud Architecture

Backend & APIs: RESTful APIs, gRPC, Microservices Architecture, WebSockets, API Gateway Design

Databases & Storage: PostgreSQL, MongoDB, MySQL, Redis (Caching), GORM, Database Optimization, Distributed Data Systems

DevOps & Infrastructure: Docker, Kubernetes, Jenkins, Terraform Infrastructure as Code (IaC), CI/CD Pipelines, Monitoring (Splunk)

Web Technologies: React.js, Next.js, Node.js, Angular, HTML5, CSS3, TypeScript

Certifications

- "AWS Cloud Practitioner Essentials" authorized by Amazon Web Services (AWS) (Dec. 2023)

Leadership & Organizations

Vice President

Sept 2021 – July 2022

Association of Computer Engineers

Executive Member

Sept 2021 – July 2022

Computer Society Of India

Extra-Curricular Activities

- Smart India Hackathon - Shortlisted in Internal Hackathon conducted by our college (Mar 2022)
- Organized a technical event for National level techno-cultural fest (Incridea, NMAMIT)