

Vrushab Ghodke

Bangalore, Karnataka, India

+91 9850214244 — vrushab.ghodke@gmail.com — linkedin.com/in/vrushab-bit — github.com/vrushab-bit

Skills

Core Languages: Go, Python, Java, JavaScript (ES6+), SQL, Lua, HTML/CSS

Web Frameworks: Node.js, React, Next.js, Express, Flask, FastAPI, Gin

AI & LLMs: LlamaIndex, Microsoft AutoGen, Langchain, RAGFlow, Vertex AI, Agentic AI, Embeddings

Cloud Services: GCP (GKE, Artifact Registry), AWS (Lambda, API Gateway, Route 53, S3, EC2, EKS)

DevOps & IaC: Docker, Kubernetes, Terraform, Terragrunt, Jenkins, GitHub Actions, Linux (Ubuntu)

Observability: Prometheus, Grafana, Loki

Security & Auth: AWS IAM, JWT, RBAC, OAuth

Databases & Storage: MongoDB, PostgreSQL, OpenSearch (Vector), Redis, MinIO

Architecture: Microservices, Serverless, Event-Driven Architecture

Experience

OneAssure

Lead AI Platform Engineer

Dec 2024 – Present

Bangalore, Karnataka

- Agentic AI Architecture:** Architected a multi-agent ecosystem using **Microsoft AutoGen**, enabling specialized agents to autonomously collaborate on complex tasks such as policy comparison and claims eligibility.
- Dynamic MCP Gateway:** Developed a Model Context Protocol (MCP) gateway using **FastMCP** to interface with the AutoGen multi-agent cluster, implementing dynamic configuration fetching from the database to ensure real-time agent adaptability.
- Enterprise Vertex AI Integration:** Engineered specialized AutoGen classes to integrate Google Gemini models via Vertex AI, creating a fully OpenAPI-compliant, enterprise-grade agent system capable of secure, scalable deployment.
- Advanced RAG Architecture:** Re-engineered the retrieval pipeline using **RAGFlow** to leverage deep document understanding for superior similarity and vector search. Extended the open-source codebase to support **Google Cloud Vertex AI** and **Gemini embedding models**.
- Temporal Workflow Orchestration:** Re-architected critical business flows (KYC, Payments) using Temporal, compressing product integration timelines from **3 months to 3 weeks**.
- Go Ecosystem Acceleration:** Engineered a standardized Go/gRPC boilerplate, slashing microservice development cycles by 50% and reducing developer onboarding time by 30%.
- Cloud Migration & Cost Optimization:** Orchestrated the infrastructure migration from AWS to GCP using Terraform/Terragrunt, resulting in a **30% reduction in cloud costs** and a 40% improvement in scalability.
- GKE Orchestration:** Managed the containerization and deployment of 10+ microservices on Google Kubernetes Engine (GKE), optimizing resource allocation by 25% while maintaining **99.99% system uptime**.
- Dynamic API Gateway:** Integrated Lua scripting for dynamic parameter injection and payload transformation, significantly enhancing system adaptability without requiring backend code changes.

SwaLay

Software Engineering Intern

Feb 2024 – Jul 2024

Pune, Maharashtra

- API & Database Design:** Designed scalable Node.js APIs for third-party integrations and architected the comprehensive database infrastructure to support the music distribution platform.
- CI/CD Implementation:** Built and maintained CI/CD pipelines using GitHub Actions, automating testing and deployment workflows to reduce manual release overhead.

Open Source Contributions

LlamaIndex

Contributor

- Developed & contributed a Mongo Chat Store, enhancing LlamaIndex's capabilities for persistent chat history management. 
- Added functionality to Postgres Chat Store to support different schemas. 

Education

Vellore Institute of Technology

Master's in Computer Applications

Aug. 2023 – March 2025

Vellore, Tamil Nadu

Pune University

Bachelor's in Computer Applications

Aug. 2019 – March 2022

Pune, Maharashtra

Honors & Awards

- Bolt 2.0 Hackathon (2024):** Developed an AI-powered learning assistant with personalized video context; focused on EdTech accessibility.
- HackBattle (2023):** Built a patient data management platform for healthcare organizations (IEEE CS Club).