Vamshi Krishna Ginna

+1 443-710-1662 | vk.ginna@gmail.com | LinkedIn | Medium | Portfolio

SUMMARY

Full Stack Engineer with **5+ years** architecting large-scale AI platforms and cloud-native solutions across **Azure** and **AWS**. Specializes in building production **RAG systems**, **conversational AI**, and **microservices** architectures using **Java/J2EE**, **Python**, **React**, and **Semantic Kernel** for federal agencies and enterprise clients. Recognized for driving **innovation** in NLP, **recommendation engines** and **data privacy** centric applications, and for mentoring teams to deliver production ready systems. **Microsoft Azure AI Engineer and Data Scientist certified** with proven track record of transforming complex challenges into scalable business solutions.

SKILLS

- **Languages:** Java, Python, JavaScript, SQL, C++
- **Cloud:** Azure (App Service, Functions, VM, Event grid, Monitor, AI Foundry, AKS, ARM, ML, Language, Vision, VNet, AI Search, Document Intelligence, Intra ID, Bot Service), AWS (EC2, Lambda, ECS, RDS, S3, CloudWatch, Lex, SageMaker, ELB, Bedrock)
- **Frameworks/Libraries:** Spring Boot, FastAPI, Node.js, React.js, GraphQL, Hibernate, Sequelize, SQLAlchemy, Pytest, Junit
- **Database:** MySQL, PostgreSQL, MongoDB, Elastic Search, Azure AI Search, Azure SQL, FAISS, Redis, Mem0
- **DevOps:** Docker, Kubernetes, Jenkins, Azure DevOps, Linux, Maven, Gradle, Git
- **AI/ML:** RAG systems, Semantic Kernel, LangChain, LLM fine-tuning, prompt engineering, agentic AI, NLP, PyTorch, TensorFlow
- **Others:** Postman, Swagger, GitHub Copilot, CodeRabbit, Figma, Power Automate, Kafka, OAuth 2.0, Microsoft Visio

EXPERIENCE

Full Stack Developer | *LCG Inc, USA* — **Aug 2024 – Present**

- Architected enterprise-scale conversational AI platform using **Azure OpenAI** and **Semantic Kernel**, processing 100K+ monthly conversations with 99.9% uptime and 200ms response times.
- Engineered production-ready **multi-modal agentic RAG system** integrating **Azure Document Intelligence, Vision Services, Azure AI Search** with custom multi-agent architecture, enabling real-time ingestion and analysis of documents, audio, video, and SQL data.
- Designed secure microservices architecture using **Azure Kubernetes Service**, **Azure App Services**, and containerization with **VNet** isolation, **private endpoints**, and **Microsoft Defender** compliance.
- Developed **high-performance APIs** using Java Spring Boot and Python FastAPI with Test-Driven Development (Junit, Pytest), async processing and Redis caching, achieving 60% improvement in data retrieval performance.
- Implemented DevOps automation with **Azure DevOps CI/CD pipelines** and **ARM templates**, reducing deployment time from 2 hours to 15 minutes and enabling zero-downtime releases.
- Invented a novel **agentic data retriever** with query rewriting, semantic re-ranking, and automated metadata filtering, building a context engineering pipeline that increased retrieval accuracy by 40% and reduced response latency by 25%.
- Built **recommendation engine** using transformer embeddings and vector similarity search, increasing user engagement by 35%.
- Led frontend modernization with **React**, **TypeScript**, **Tailwind CSS**, and **USWDS**, ensuring accessibility compliance with U.S. Web Design Standards and reducing UI load times by 30%.
- Delivered AI-powered inquiry management using **Azure Bot Services** and **Language Services**, handling 10K+ monthly inquiries with 90% automated resolution rate.

Full Stack Developer | *Ardent Security LLC, USA* — **Aug 2023 – May 2024**

- Led transformation of legacy system to multi-tenant SaaS platform using **Node.js, Express.js**, **React.js, and TypeScript**, supporting 100+ clients with 80% reduction in database query times through intelligent caching strategies.
- Built secure **RESTful** and **GraphQL** APIs using Test-Driven Development with JWT authentication, achieving 40% improvement in data efficiency and enhanced security compliance.
- Developed **NLP**-powered data discovery system using **Python** and **LangChain**, improving privacy risk identification accuracy by 25% and automating PII detection workflows.
- Engineered real-time **WebSocket** notification system serving 500+ concurrent users with distributed architecture and event-driven.
- Built **generative AI** powered **Text-to-SQL** system enabling natural language queries, cutting manual query workload by 60%.

Software Engineer, Associate | *Tata Consultancy Services, India* — **Jan 2021 – Aug 2022**

- Built **Java Spring Boot microservices** with JPA and Hibernate for robust data management, cutting API query latency by 20%.
- Architected **AWS** cloud-native infrastructure using **ECS**, **Lambda**, **RDS**, and **S3**, implementing auto-scaling policies and load balancing that reduced operational costs by 15% while handling 3x traffic growth.
- Engineered high-performance database solutions by fine-tuning **PostgreSQL** queries and implementing **MongoDB** indexing, improving throughput by 30% for applications processing 1M+ records daily.
- Automated data transfers and system monitoring with **Python** scripts and **AWS CloudWatch**, reducing manual intervention by 45% and increasing synchronization reliability across distributed systems.

Software Engineer | *Ezerka Technology Solutions, India* — **Jun 2019 – Dec 2020**

- Co-founded and led technical development of inventory management SaaS platform, architecting 100+ REST APIs with Node.js, building real-time React.js dashboards, and implementing automated data pipelines serving MongoDB and GCP storage systems.

EDUCATION

- **Master of Science in Computer Science** | University of Maryland, Baltimore County, USA — **May 2024**
- **Bachelor of Technology in Computer Science** | Jawaharlal Nehru Technological University Hyderabad, India — **Aug 2020**