Banudeep Reddy Gade

Software Engineer

VA, USA | Mobile: (551) 465-6500 | gbanudeepreddy@gmail.com | LinkedIn | GitHub | Portfolio

Summary

Software Engineer with 4+ years of experience building Java/Spring Boot microservices and React frontends for banking and healthcare. Design and ship event-driven, cloud-native systems on AWS that handle high-volume transactions, improve fraud detection, and raise reliability with strong observability and CI/CD, including real-time voice agents and AI agents for conversational and analytics use cases.

Education

George Mason University — Fairfax, VA, USA
Master of Science in Computer Science — *Aug 2023 – May 2025*

Visvesvaraya National Institute of Technology — Nagpur, India
Bachelor of Technology in Computer Science Engineering — *Jul 2018 – Apr 2022*

Technical Skills

Languages: Java, JavaScript, TypeScript, SQL, Python
Backend: Spring Boot, FastAPI, Node.js, Spring MVC, Spring Security, Microservices, REST APIs, Apache Kafka
Frontend: React.js, Redux, React Hooks, HTML5, CSS3, Bootstrap, Tailwind
GenAI & LLMs: Prompt engineering, LangChain, LangGraph, RAG, MCP, vector search (Pinecone, Chroma), AI/voice agents
AI/ML & Data: TensorFlow, PyTorch, scikit-learn, Pandas, NumPy, feature engineering, model evaluation/tuning
Databases: MySQL, PostgreSQL, MongoDB
Cloud/DevOps: AWS (ECS/EKS, Lambda, S3, API Gateway, CloudWatch), Jenkins, Docker, Maven, Gradle
Testing/Quality: JUnit, Mockito, Selenium, Postman, Swagger/OpenAPI
Tools & Process: Git/GitHub, Jira, Agile/Scrum, SDLC, Linux/Windows
Certifications: AWS Certified Developer – Associate, AWS Certified AI Practitioner, Google Cloud Associate Cloud Engineer

Experience

Software Engineer — Aug 2024 – Present
Wells Fargo — *VA, USA*

- Owned banking features end-to-end by partnering with product owners and building Java/Spring Boot microservices with Kafka pipelines, cutting average transaction latency by ˜25% and stabilizing high-volume digital workflows against SLA breaches.
- Delivered responsive React.js + Redux dashboards for transaction monitoring and analytics, used by 30+ risk and operations users to track 15+ KPIs with real-time visibility into anomalies and SLA breaches.
- Strengthened fraud detection by wiring Python/Pandas-based anomaly checks into the data pipeline, computing risk features that reduced noisy alerts by ˜15% and helped compliance teams review suspicious activity ˜20% faster.
- Restructured schemas and tuned MySQL queries for payment settlements, increasing throughput and availability by ˜40% for critical transaction flows.
- Designed and deployed AWS architectures with API Gateway, ECS/EKS, and Lambda to securely handle millions of transactions per day, adding CloudWatch dashboards and alerts to surface bottlenecks and reduce time-to-detect for production issues.
- Improved release quality and speed by expanding JUnit-based regression coverage and standardizing CI/CD via GitHub workflows, cutting manual effort and stabilizing regulated release cycles.

Software Engineer — Sep 2020 – Jul 2023
Capgemini — *India*

- Developed Spring Boot services with Spring Security to enforce HIPAA-compliant authentication and authorization for core healthcare platforms, safeguarding access to records for 50k+ patients across multiple clinics.
- Built React healthcare portals with Redux and Hooks, enabling clinicians and patients to access dashboards, appointments, and records through responsive, accessible UIs.
- Created TensorFlow-based predictive models on tens of thousands of patient records to flag early disease-risk patterns, improving identification of at-risk cases by ˜20% and giving clinical teams earlier decision support for interventions.
- Optimized PostgreSQL performance for high-traffic workflows by refining queries, indexes, and connection pooling, reducing p95 query latency by ˜60% and improving responsiveness during peak usage windows.
- Containerized services with Docker and automated Jenkins pipelines with Postman/OpenAPI checks, reducing release-related incidents by ˜30% and improving interoperability with third-party healthcare systems.

Projects

Modular Voice Agent – Cascading STT → LLM → TTS Pipeline — GitHub

- Built a web-based voice assistant with a cascading STT → LLM → TTS pipeline at 16kHz, using AudioWorklet and WebSockets to support real-time conversations with live transcription and mid-response interruption (barge-in).
- Designed a provider-agnostic architecture that mixes Deepgram/Azure Speech for STT/TTS with OpenAI GPT-4 / Azure OpenAI for LLM processing, wired through a Python `aiohttp` backend and configurable via environment variables and system prompts.
- Containerized the service with Docker and deployed to Azure Container Apps on a single 8080 port for HTTP/WebSocket traffic, adding connection management, error handling, optional audio recording, and health checks for production-ready reliability.