Hari Sandeep Reddy Nandyala

sandeepreddyn08@gmail.com | (716) 705-1836 | linkedin.com/in/hari-sandeep-nandyala/

SUMMARY

Full-Stack Software Engineer with over three years of experience architecting scalable event-driven microservices and interactive frontend solutions across fintech, telecom and real estate domains. Skilled in Java, Python, Go and JavaScript with expertise in Spring Boot, Node.js, React and integrating AI/LLM features and automating CI CD pipelines to deliver robust cloud-native applications.

EDUCATION

University at Buffalo, The State University of New York **December 2024**
Master of Science in Computer & Information Sciences (GPA 3.7) *Buffalo, NY*

SKILLS

Programming Languages: Java, Python, Go, C++, SQL, Bash, TypeScript, JavaScript
Frontend: HTML5, Tailwind CSS, Angular, React, Node.js, Next.js, Material UI, Redux, RTK Query
Backend: Spring Boot, Fast API, RESTful APIs, gRPC, Apache Kafka, Redis, PostgreSQL, Maven, Hibernate, Mockito, GraphQL
Data & Cloud Infrastructure: AWS, Azure, Kubernetes, Docker, DevOps
DevOps & Observability: GitHub Actions, CI/CD Pipelines, Prometheus, Grafana, Junit, Jenkins, Jira, Scrum, Agile
Other: Distributed Systems Design, Microservices, Event-Driven Architecture, LangChain, OpenAI APIs, Prompt Engineering

EXPERIENCE

PilProgram **January 2025 – Current**
Founding Engineer *Buffalo, NY*

- Architected a modular React portal with Tailwind CSS and Material UI, leveraging Redux Toolkit slices, RTK Query and Firebase Authentication to streamline realtor onboarding and data flow.
- Designed and implemented a Node.js and Express.js microservices architecture with MongoDB and Kubernetes orchestration for Parliamade's real estate platform, enabling horizontal scaling and reducing API response times by 30%.
- Engineered a real-time analytics pipeline to replace slow DB queries by streaming like/save events via Kafka into Node.js microservices, aggregating counts in Redis for sub-second REST endpoints and chart renders.
- Used AWS S3 with presigned URLs and CloudFront caching to securely store and serve realtor media with low latency.

Invesco **January 2022 – July 2023**
Software Engineer *Hyderabad, India*

- Built gRPC transaction handlers in Java with Spring Boot to ensure atomic consistency in distributed fund-reconciliation workflows, reducing end-of-day processing time by 25%.
- Developed a document processing service using LangChain, OpenAI APIs, and React.js to automate extraction and summarization of contract data, saving over 100 hours for teams across multiple departments.
- Refactored PL/SQL stored procedures for portfolio and trade data pipelines, reducing daily investment report generation time by 45% and boosting data availability for analysts.
- Implemented Kafka consumers in Java Spring Boot microservices and Hazelcast IMap caching for in-memory dashboard snapshots, enabling sub-second REST queries and reducing API latency by 15%.
- Migrated legacy portfolio dashboards from Angular to React with Material UI and RTK Query for efficient API calls and caching, reducing page load time by 40 % and improving developer productivity.
- Developed JUnit and Mockito tests covering edge-case scenarios to validate error handling and improve service reliability.

Adroit **November 2020 - December 2021**
Software Engineer - Backend *Hyderabad, India*

- Developed standalone Spring Boot microservices to expose ODA reporting metrics via REST APIs, enhancing backend maintainability and enabling real-time operational visibility for analytics workflows.
- Implemented a GraphQL gateway on top of those microservices to consolidate data fetching, reduce redundant API calls by 8%, and enhance overall application performance.
- Built a Kafka-backed metrics ingestion pipeline to validate and aggregate over 1 M daily ODA events, achieving data accuracy and slashing downstream processing times by 30%.
- Designed & deployed CI/CD pipelines using Azure DevOps (Pipelines, Artifacts, Repos), reducing release cycles by 60%.

PROJECTS

Resume-to-AI Chatbot (RAG Pipeline) | *Python, LangChain, OpenAI API, FAISS, FastAPI, React* **May 2025**

- A context-aware chatbot that answers questions about any resume by combining semantic search with GPT-3.5 Turbo.
- Chunked resumes into semantically coherent fragments using LangChain's RecursiveCharacterTextSplitter to ensure relevant context retrieval.
- Generated embeddings for each fragment via OpenAI's text-embedding-ada-002 model and indexed them in FAISS, achieving sub-100 ms similarity searches.
- Developed a Fast API RetrievalQA microservice with a React.js chat UI and containerized the full stack in Docker for seamless, production-ready deployment.