Reddy Harsha Vardhan

Software Engineer

harsha.r@mymailshub.com | +1 253-362-9609 | LinkedIn | GitHub | Portfolio | USA (Open to Relocate)

PROFESSIONAL SUMMARY

Software Engineer with 3+ years of experience designing and deploying scalable full-stack and AI-powered systems for global financial institutions. Proficient in Java, Spring Boot, React, Angular and Node.js, with expertise in microservices and event-driven architecture. Delivered trade settlement platforms processing over 1.5 million transactions daily and AI-driven analytics pipelines handling 4.2 TB of market data weekly, reducing insight turnaround time from 90 minutes to under 20. Experienced in implementing Transformer-based LLMs, GPT, and LangChain for automated insight generation, and skilled in AWS deployments, CI/CD automation, and observability using Prometheus and Grafana

SKILLS

Languages: Java, TypeScript, JavaScript, Python, SQL, C++
Frontend: React.js, Next.js, Angular, Redux Toolkit, Tailwind CSS, Material UI, HTML5, CSS3, SASS, Responsive Design
Backend: Spring Boot, Node.js, Express.js, GraphQL, RESTful APIs, Hibernate, JPA, Apache Kafka, RabbitMQ, Event-Driven Architecture
Cloud / DevOps: AWS (EC2, S3, Lambda, RDS), Docker, GCP, Terraform, Jenkins, GitHub Actions, GitLab CI/CD, Kubernetes, Prometheus, Grafana
Databases: PostgreSQL, MongoDB, MySQL, DynamoDB, Redis, ElasticSearch
Testing & Tools: JUnit, Mockito, Selenium, Cypress, Jest, Chai, Postman, Enzyme
Tools & Others: Git, Jira, VS Code, Webpack, Microservices Architecture, OAuth 2.0/JWT, System Design, API Gateway, DSA, OOPS
AI & LLMs: Generative AI, Large Language Models (LLMs), LangChain, Hugging Face Transformers, GPT, BERT, MLflow, Prompt Engineering

PROFESSIONAL EXPERIENCE

PNC Financial **Sep 2024 – Present | United States**
AI Software Engineer

- Engineered Java Spring Boot microservices for PNC's AI-assisted portfolio intelligence platform, supporting over 350 institutional portfolios and optimizing model-driven trade execution workflows through event-driven architecture and parallel processing pipelines.
- Developed React + Express-based analytics console with modular TypeScript components and secure API integrations, allowing investment strategists to visualize portfolio exposure, model drift, and rebalancing insights across multiple asset strategies.
- Integrated Python FastAPI services with LangChain and document-embedding pipelines to process 250K+ investment research files, enabling automated insight extraction and reducing manual analysis overhead by thousands of pages per quarter.
- Built scalable streaming infrastructure using Apache Kafka and Airflow to orchestrate real-time data ingestion from market feeds and internal trade systems, processing 4.2 TB of raw data each week for downstream AI pipelines.
- Collaborated with data science teams to fine-tune Transformer-based NLP models hosted on Hugging Face and deployed through AWS SageMaker, generating contextual summaries and cutting insight turnaround from 90 minutes to under 20.
- Implemented low-latency model-serving endpoints using Amazon DynamoDB and async I/O patterns, achieving sub-120 ms response times for predictive analytics APIs consumed by risk dashboards.
- Automated end-to-end build, validation, and deployment processes with Jenkins, Docker, and Kubernetes, maintaining continuous delivery pipelines for 30+ AI microservices hosted on AWS.

Goldman Sachs **Apr 2021 – Jul 2023 | India**
Software Engineer

- Constructed and deployed Java Spring Boot microservices for the internal trade settlement platform handling 1M+ daily transactions, improving throughput by restructuring message queues and optimizing serialization logic.
- Formed Angular-based dashboards using modular TypeScript components and RxJS data streams that provided real-time visibility into trade positions across 12 global trading desks, enhancing desk coordination and operational clarity.
- Established GraphQL APIs integrated with PostgreSQL and Redis caching layers, supporting high-frequency trade queries with a consistent 80–90ms average latency during peak hours.
- Programmed data validation and reconciliation pipelines through Python-based ETL scripts, replacing legacy manual checks and saving ~20 analyst-hours per trading day.
- Embedded AWS SageMaker models with Lambda for real-time trade anomaly detection, allowing risk teams to flag potential outliers 3× faster and significantly improving audit readiness.
- Designed and maintained Angular-Node.js-based internal tools for portfolio exposure reporting, improving collaboration between risk and compliance teams.
- Led migration of legacy MySQL databases to PostgreSQL, introducing partitioned tables and optimized indexing strategies that improved batch reconciliation stability across 14 systems.
- Set up Prometheus and Grafana stack to monitor service health and latency, enabling anomaly alerts and reducing on-call escalations.
- Executed centralized logging via the ELK stack (Elasticsearch, Logstash, Kibana), simplifying debugging for over 40 microservices and shortening incident analysis cycles.
- Streamlined CI/CD workflows using Jenkins, Docker, and Kubernetes, automating deployments and cutting build-to-release steps from 14 to 6.
- Partnered with the security engineering team to establish OAuth 2.0 and JWT-based authentication standards across all APIs, ensuring compliance with Goldman Sachs' internal security protocols.

ACADEMIC PROJECTS

CentrX – Student & Club Engagement Platform | Node.js | MongoDB | JavaScript | Hugging Face | LangChain | SLU SSO |

- Architected and deployed RESTful APIs supporting over 2,000 student interactions per semester by enabling RSVP tracking, real-time updates, and personalized content recommendations using Hugging Face Transformer models.
- Incorporated SLU Single Sign-On with OAuth2/JWT and Applied GPT-powered summarization via LangChain, generating automated event briefs and engagement insights that reduced manual reporting time by several hours weekly.

EDUCATION

Master of Science in Computer Science
Saint Louis University, Saint Louis, Missouri