

HEMANTH KARI

332-242-0668 | hkariusa@gmail.com | LinkedIn | GitHub

SUMMARY

Full Stack Engineer with mid-level experience delivering scalable online platforms and high-performance backend services. Expert in React, Next.js, TypeScript, Python, and AWS, with a track record of building microservices that process 50,000+ transactions daily, boosting Core Web Vitals by 40% and cutting API latency by 78%. Led the redesign of a quotation flow and order-fulfillment pipeline, reducing latency and improving conversion rates. Seeking to apply this expertise to drive efficient, reliable software for manufacturing firms.

TECHNICAL SKILLS

Frontend: React.js, Next.js, TypeScript, JavaScript (ES6+), Redux, Zustand, React Query, Tailwind CSS, HTML5, CSS3, Responsive Design, Server-Side Rendering (SSR), Static Site Generation (SSG), Web Performance Optimization

Backend: Python, Java, Kotlin, Node.js, Django, Spring Boot, Express.js, RESTful APIs, GraphQL, gRPC, Kafka, Redis, Celery

Infrastructure: AWS (EC2, EKS, Lambda, DynamoDB, S3, CloudFront), Docker, Kubernetes, Terraform, Vercel, CI/CD

Data: PostgreSQL, MongoDB, Redis, Elasticsearch, Prisma ORM, Query Optimization

ML: PyTorch, TensorFlow, HuggingFace, LLM Fine-Tuning (LoRA/QLoRA), Distributed Training

EXPERIENCE

Chewy | E-Commerce Platform

Dec 2025 – Present

Software Developer (Contract)

Remote

- Expanded the order fulfillment pipeline using Kotlin, Kafka, and DynamoDB to build event-driven microservices that process over 50,000 orders daily, cutting order dispatch time by 25% during peak holiday periods
- Built storefront sections with Next.js 14, React, and TypeScript, delivering server-side rendered product pages and checkout flows; SEO rankings improved and Largest Contentful Paint (LCP) increased by 40%
- Enhanced inventory synchronization by implementing SQS and Lambda for real-time updates across three fulfillment centers, and introduced Redis-based distributed locking to prevent overselling
- Developed React search components integrated with Elasticsearch, adding debounced autocomplete, faceted filtering, and infinite scroll, which lowered p95 latency from 450 ms to 120 ms
- Created a testing environment using LaunchDarkly feature flags and A/B testing, enabling 15+ checkout experiments that boosted conversion rates by 8%

Clearcover | Digital Auto Insurance

Jun 2024 – Nov 2025

Contract Software Engineer

Remote

- Led the move to microservices: planned the breakup of a large Django project on AWS EKS into domain-driven services and built an automated pipeline with Terraform and GitHub Actions, enabling daily releases and cutting deployment errors by 40%
- Redesigned the insurance quote wizard using React and TypeScript. The new interface reduced form abandonment by 35% and drove a 28% jump in mobile conversions.
- Solved N+1 query issues and added missing indexes to the main API. With these changes, the p95 latency dropped from 800ms to 180ms, which led to more customers buying from us.
- Built observability from the ground up by making bespoke dashboards and PagerDuty notifications to keep an eye on Prometheus and Grafana. It used to take 30 minutes to find faults, but now it only takes less than 5 minutes.
- Implemented CI/CD pipeline using Terraform and GitHub Actions, replacing manual weekly releases with automated daily deployments and enabling automatic rollbacks
- Implemented OAuth 2.0/JWT authentication in Node.js using AWS Cognito and Terraform, creating short-lived 15-minute tokens; this hardened the application against token replay attacks and helped pass the security audit
- Coached two junior engineers in weekly pairing sessions, raising test coverage from 65% to 92% and reducing post-release issues by 45%

Deepcloud | Cloud-Based Business Platform

Jan 2021 – Aug 2022

Software Development Engineer

Bangalore, India

- Developed a provisioning engine in TypeScript that autonomously managed the lifecycle of 500+ VMs, reducing manual effort by 90% and cutting setup time from two hours to 10 minutes
- Scaled the platform to support 50,000 daily users by implementing circuit breakers, retry logic, and graceful degradation in Python and Spring Boot microservices, achieving 99.5% uptime

- Optimized database performance by adding connection pooling, read replicas, and targeted indexing in PostgreSQL and MongoDB, lowering p99 latency by 38%
- Led on-call incident response as the designated DRI, creating runbooks and automated remediation scripts that reduced MTTR from 90 to 50 minutes (45% improvement)
- Created OpenAPI documentation for 30+ endpoints and defined an API versioning strategy, enabling partners to integrate 60% faster, reducing onboarding time from two weeks to five days

Purdue University

Oct 2022 – May 2023

Research Computing

West Lafayette, IN

- Develop a Django/PostgreSQL backend with CI/CD pipelines, supporting over 1,000 concurrent users and maintaining 99.9% uptime; implement caching and query optimization to improve page load speed by 40%
- Redesign the batch processing pipeline to use asynchronous workers and parallel execution, reducing processing time by 81% (from four hours to 45 minutes)
- Create React dashboards that pull data via Prisma ORM for over 2,000 faculty members, simplifying data access and lowering support requests by 35%

PUBLICATION

DeePromClass: Deep Neural Networks for Promoter Classification

IEEE/ACM TCBB, 2023

- DeePromClass: Deep Neural Networks for Promoter Classification. IEEE/ACM TCBB, 2023

PROJECTS

Llama 3.1 Pro Coder v1 | huggingface.co/HemanthKari/Llama-3.1-Pro-Coder-v1

2025

- Strict checkpoint evaluation was used to prevent overfitting; a model that is safe for commercial use and includes comprehensive documentation was released on Hugging Face.

Personal Portfolio | hemanthkari.hackthru.com

2024

- Using TypeScript, Next.js, and React Server Components, I created a portfolio that is compatible with all devices. The Google Lighthouse performance score was 100, the SEO score was 90, and the best practices score was 100.
- Added gradient effects, CSS animations, and a light and dark mode switch. Using Vercel edge deployment and next/image lazy loading, we accelerated global load times.

EDUCATION

Purdue University

August 2022–May 2024

Master of Science in Computer Science

GPA: 3.5/4.0

- coursework: computer security, distributed systems, machine learning, database systems, and algorithm design

KL University

July 2017–May 2021

B.Tech, Computer Science

India

ACHIEVEMENTS

Hack The Box Global Ranking: Elite security challenges and penetration testing for the top 100 out of more than 500,000 individuals worldwide

LLM Standard: With a HumanEval score of 68.3%, the Built Llama 3.1 Pro Coder outperformed the GPT-3.5 Turbo by more than 20 points.