

ISMAIL KATTAKATH

Full Stack Developer | Tech Lead | LLMOps

📞 +1 (647) 225-2878 📩 ismail@kattakath.com 📍 24-242 John Garland Blvd, Toronto, ON M9V 1N8

🌐 ismail.kattakath.com

LinkedIn: linkedin.com/in/ismailkattakath GitHub: github.com/ismail-kattakath

Summary

Principal Software Engineer with deep expertise in full-stack development and cloud infrastructure, delivering high-impact, scalable systems at scale. Currently architecting and operating production-grade GenAI inference platforms using vLLM, Kubernetes, and multi-GPU clusters, achieving 25% cost reduction and 40% latency improvement. Previously rescued an abandoned healthcare MEAN-stack application and transformed it into a nationally deployed platform with 99.5% uptime; led the AngularJS-to-Next.js modernization, implemented enterprise federated authentication (OAuth 2.0 + SAML 2.0) across 5+ partners, and established DevOps practices that reduced deployment time by 92% and production incidents by 65%. Proven ability to drive end-to-end architecture, performance optimization, security/compliance, and measurable business outcomes in distributed cloud environments.

Education

Bachelor's Degree in Computer Science and Engineering

[Mahatma Gandhi University](#)

Apr, 2010

Work Experience

Silver Creek Insights Inc.

Aug, 2024 - Present

Principal Software Engineer

Aloshy.ai: AI-as-a-Service platform providing B2B APIs for media manipulation and custom-trained generative AI pipelines tailored to specific business needs.

- ⦿ Architected scalable GenAI inference infrastructure using Kubernetes/Containerization, vLLM etc. supporting multiple Diffusion and LLM models, and deployed on GCP/GKE and AWS/EKS for production.
- ⦿ Implemented advanced techniques such as KV-cache, Distributed Inference, Multi-GPU/Node Scaling, CUDA optimization and memory/load management to reduce inference latency by 40%, and cutting operational costs by 25%.
- ⦿ Designed and documented GPU cluster management playbooks including model versioning/rollbacks, utilization monitoring, failure recovery, and load balancing strategies.
- ⦿ Built enterprise-grade secured API architecture using centralized gateway/aggregator for serving internal and curated third-party MCP servers/tools.

Homewood Health Inc.

Oct, 2017 - Jul, 2024

Senior Software Engineer

Homeweb.ca: Canada's leading digital mental health provider, delivering EFAP and CBT solutions to hundreds of thousands of employees nationwide.

- ⦿ Inherited and revitalized an abandoned, undocumented MEAN stack application, transforming it into a maintainable, production-ready web platform with 99.5% uptime.
- ⦿ Rebuilt authentication system implementing OAuth 2.0 PKCE flow for SSO integration with Microsoft Identity Platform.
- ⦿ Collaborated with federated vendors to configure and troubleshoot SAML 2.0 authentication, enabling seamless user login across 5+ partner organizations.
- ⦿ Led complete frontend migration from AngularJS to Next.js upon framework EOL, preserving existing UI/UX and API compatibility.
- ⦿ Established comprehensive DevOps practices including automated testing, CI/CD pipelines (reducing deployment time from 4 hours to 20 minutes), containerization, monitoring, and backup systems.
- ⦿ Led infrastructure migration from DigitalOcean to AWS, orchestrating transition of EC2, RDS, S3, CloudFront, and VPC components using

AI/ML Stack

Generative AI Pipelines, RAG Systems, Multi-Agent LLM Workflows, MCP Server Gateways, Kubernetes/Containerization, vLLM, TGI, Triton Inference Server, CUDA Optimization, TensorRT, ONNX, Google Vertex AI, Hugging Face Transformers, Knowledge Graphs, NER Models, Claude Code, Gemini CLI, Cursor

Cloud Services

Google Cloud (Vertex AI, GKE), AWS (EKS, SageMaker, EC2 P4/P5, Inferentia, Lambda, S3), Vercel, DigitalOcean, MongoDB Atlas, Cloudflare Workers/R2

Authentication & Security

OAuth 2.0, PKCE, SAML 2.0, OpenID Connect (OIDC), JWT, Bearer Token, Auth0, Keycloak, Basic Auth

DevOps & CI/CD

Kubernetes, Docker, Docker Compose, GitHub Actions, GitLab CI, Terraform, Grafana, Ansible, Vagrant, Act

Backend & APIs

Node.js, Express, RESTful APIs, Microservices, Swagger/OpenAPI, GraphQL, SOAP/XML

Programming / Scripting

JavaScript, TypeScript, Python, Bash

Databases

PostgreSQL, MongoDB, MySQL, MS SQL, Redis, Supabase, Qdrant, Chroma, Neo4j

Protocols

HTTP/HTTPS, WebSocket, Server-Sent Events, WebRTC

Web & Mobile UI

Next.js, ReactJS, Angular, TailwindCSS, React Native, Expo, Flutter

Languages

English

Terraform for infrastructure-as-code, achieving zero-downtime deployment.

Etuper Technologies Pvt. Ltd.

Apr, 2016 - Aug, 2017

Lead Software Engineer

Brilliant Rewards: Complete hospitality software suite—dashboard, mobile and web apps—featuring loyalty programs, marketing automation, booking integration, and analytics.

- ⦿ Collaborated with stakeholders to gather requirements, create user stories, design wireframes, and develop functional prototypes.
- ⦿ Managed sprint planning, daily standups, backlog prioritization, and blocker resolution to ensure continuous delivery.
- ⦿ Designed normalized database schemas, optimized SQL queries and indexing strategies, and implemented stored procedures with comprehensive test coverage.
- ⦿ Led development of scalable microservices architecture with containerized, discoverable APIs following OpenAPI specifications.
- ⦿ Delivered multiple client applications including AngularJS web platforms and cross-platform mobile solutions.
- ⦿ Introduced development standards encompassing coding conventions, linting, code reviews, documentation, and incident management processes, reducing production incidents by 65%.

RM plc.

Jun, 2011 - Mar, 2016

Software Engineer

RM Integris: School MIS system featuring student management, grade tracking, attendance monitoring, parent portals, and performance analytics.

- ⦿ Implemented OpenAPI/Swagger bridging layer to modernize legacy SOAP APIs for RESTful client consumption.
- ⦿ Prototyped AngularJS interface to decouple presentation layer from legacy Struts JSP architecture.
- ⦿ Maintained load-balanced Linux VM clusters hosting distributed web, application, and database tiers.
- ⦿ Mentored interns and onboarded new team members through technical guidance and knowledge transfer.

Posibolt Solutions Pvt. Ltd.

Jul, 2010 - Jan, 2011

Junior Software Engineer

Posibolt: Adaptable cloud ERP solution serving key retail businesses across global industry segments.

- ⦿ Created UI mockups based on SRS specifications and converted approved designs into responsive HTML/CSS templates.
- ⦿ Developed integration test cases and implemented browser automation using Selenium WebDriver.
- ⦿ Authored user documentation including help articles, user guides, FAQs, and product website content.