

ANDREW JOHN KAISERAUER

ajkaiserauer@gmail.com | 734-957-2852 | Detroit, MI | linkedin.com/in/aj-kaiserauer | github.com/kornsour | akaiserauer.com

SUMMARY

Engineering leader who built a 33-person Platform Engineering organization from the ground up, serving thousands of developers across the Deloitte US Member Firm. I hire and coach tech leads, own platform architecture and delivery end to end, and stay hands-on in code. Ten years across platform engineering, cloud infrastructure, and AI-native engineering.

EXPERIENCE

Engineering Manager, Platform Engineering

Jan 2021 – Present

Deloitte US Member Firm | Detroit, MI

- Built and lead a four-squad Platform Engineering organization of 33 engineers, wearing the product-owner, engineering-manager, and lead-architect hats across all four squads; I hire and coach the tech leads, and leads own squad roadmaps while I set org vision and the architectural bar.
- Run a shared multi-tenant Kubernetes platform — an internal Vercel for every team in the department — serving 40+ microservices at ~99.95% uptime; standardized CI/CD across 15 teams with golden images and GitOps (ArgoCD), cutting deployment from hours to minutes and incidents ~30%.
- Built an automated platform-provisioning framework: SCIM-driven access to developer tooling, a policy enforcement engine that makes the secure path the default, health monitoring, adoption analytics, and cross-platform access intelligence for security and compliance — plus the frontend squad that fronts it.
- Own platform tooling and support across GitHub, Claude, Docker, and Postman for thousands of engineers firm-wide; cut operational burden ~60% through self-service tooling, standardized support lanes, and automation.
- Led org-wide Backstage adoption as the paved path for 50+ services, reducing onboarding from weeks to days; won over 15+ independent teams without direct authority via early-adopter demos and a hub-and-spoke model.
- Drive AI-native engineering workflows: enterprise GitHub Copilot rollout with automated code review on pull requests and agentic issue resolution; leading enterprise AI platform evaluation across ChatGPT, Gemini, and Anthropic Enterprise.

Senior Infrastructure Architect

Jan 2018 – Dec 2020

GE Aviation | Van Buren Township, MI

- Delivered 25+ cloud migrations and new builds producing ~\$1.9M in infrastructure cost savings across GE Aviation, Renewables, and Transportation.
- Built self-service AWS CloudWatch dashboards for 10+ application teams and designed an AWS-native CI/CD pipeline that cut website deployment from 1 hour to 2 minutes.

Enterprise Application Engineer

Jul 2017 – Jan 2018

GE Digital | Van Buren Township, MI

- Cut Digital-Healthcare AWS compute cost 6% (~\$300K/year) with a sleep-schedule service for non-production instances; built a web app aggregating cloud data from AWS, Chef, ServiceNow, and Qualys into actionable operational insight.

Cloud Automation Engineer

Jul 2016 – Jul 2017

GE Healthcare | Van Buren Township, MI

- Led 58+ application migrations and new cloud builds saving \$45K+ per month; increased team delivery speed 40% by coaching Agile teams on Scrum and Kanban.

Independent Engineering

- LLM inference platform on Kubernetes (KServe, vLLM): inference-aware autoscaling on queue depth and KV-cache utilization through a custom Go KEDA external scaler, scaling 1 to 5 replicas at ~2,360 tokens/sec aggregate for ~\$0.0015–0.005 per 1M tokens; GitOps-managed, with Prometheus and Grafana.
- Inference performance engineering (PyTorch, CUDA): KV-cache reuse for 4.8× faster decode ($O(T^2)$ to $O(T)$), request batching for 3.42× throughput at batch 16, int8 dynamic quantization for a 3.6× smaller model, a custom fused-RMSNorm CUDA kernel, Megatron-style tensor parallelism, and a p50/p95 benchmark harness.
- Agentic ops-triage system that handles IT tickets end to end behind human approval gates, with RBAC, a hash-chained audit trail, and an evaluation harness gating every release.

SKILLS

Leadership: Engineering org design and team leadership (33 reports), hiring and coaching tech leads, roadmap ownership, platform adoption strategy, cross-functional stakeholder engagement, Agile (Scrum, Kanban, SAFe)

Platform & Infrastructure: AWS, Azure, GCP, Kubernetes (EKS), Terraform, Docker, CI/CD, GitHub Actions/Apps, Helm, Backstage, GitOps (ArgoCD, Kustomize), Observability (Prometheus, Loki), Linux, Bash

AI & LLM: GitHub Copilot (enterprise adoption and enablement), Claude Code, Claude API integration, agentic workflow design, LLM inference infrastructure, enterprise AI platform evaluation

Languages: Python, Go, FastAPI, Flask, JavaScript/TypeScript, Node.js, React, PowerShell

Certifications: Certified SAFe 5 Product Owner/Product Manager (issued Feb 2023, expires Apr 2027); Deloitte Certified Architect Level 1 (issued Nov 2024); Microsoft Certified: Azure Fundamentals (issued May 2020); AWS Certified DevOps Engineer – Professional (issued Apr 2021, expired Apr 2024); AWS Certified Solutions Architect – Associate (expired Jan 2024)

EDUCATION

Michigan State University: BS, Media & Information | East Lansing, MI | May 2016 | Minors: Computer Science Engineering, Information Technology, Game Design & Development