

SAGYAM THAPA

+977 986-229-9137 | sagyamthapa32@gmail.com | linkedin.com/in/sagyam | github.com/sagyam | sagyamthapa.com.np

SUMMARY

DevOps engineer with 4+ years in cloud, self-hosted infrastructure, and full-stack development. Built 20+ Kubernetes clusters across AWS and Azure and a highly available healthcare SaaS, and trains engineers in Kubernetes, IaC, and DevSecOps. Writes interactive CS explainers; one drew 162 points on Hacker News.

EXPERIENCE

DevOps Engineer

2025 – Present

Ascend Cloud

Contract

- Designed and operated 20+ Kubernetes clusters on AWS and Azure for 10+ clients
- Cut monthly cloud bills by 12–35% per project by right-sizing nodes, running non-critical workloads on spot instances, and moving suitable workloads to bare metal or serverless
- Caught an AWS secret key being shipped in a client's Next.js bundle, then rolled out GitGuardian secret scanning across the organization
- Built 30+ GitHub Actions pipelines with Helm and secure IaC templates for a retail client's 11 microservices, supporting 50+ PR merges a day
- Monitored all 11 services with Prometheus, Grafana, and Loki, with dashboards and Teams/email alerting; added OpenTelemetry tracing to 3 of them
- Evaluated commercial and open-source ASR models (NeMo, Parakeet, Whisper) and fine-tuned on a few hundred hours of noisy, code-switched speech with crosstalk, reducing WER by 20–25%

DevOps Trainer

2026 – Present

Tech Axis

Contract

- Mentored 20+ DevOps students over 6 months, guiding them one-on-one through capstone projects
- Designed the curriculum and authored all labs and slide decks: Linux, Docker, CI/CD, Kubernetes, AWS, Terraform, GitOps, observability, and DevSecOps
- Built hands-on labs for Argo CD + Flagger progressive delivery on EKS, an LGTM + Jaeger observability stack, and deliberately broken Kubernetes stacks for troubleshooting practice
- Co-designed an agentic AI course covering transformer fundamentals, open-weight models, RAG and vector databases, tool use and agent harnesses, prompt caching, cost and benchmarks, and model gateways (OpenRouter, Cloudflare, Vercel)

Full Stack Engineer

2024

Bitmosys

Full-time

- Architected the ERP behind a UK nurse staffing SaaS serving 300+ nurses: 3-node HA PostgreSQL cluster, Redis caching, and Nginx API gateway
- Designed a highly available permission system with audit logging for every action across 1M+ objects, so nurses could always log shifts, expenses, and progress
- Built a constraint-based shift planner (staff preferences, availability, UK bank holidays) that replaced fully manual scheduling
- Ran a Linkerd service mesh and a self-managed Grafana, Prometheus, and Loki stack; deployed with Docker, Kubernetes, and CI/CD
- Chose the frontend stack and set performance, code-quality, and design-language guidelines; mentored a junior developer and reviewed PRs

Full Stack Developer

2023 – 2024

Cango

Contract

- Developed dynamic dashboards with React, React Query, Zustand, and Tailwind CSS
- Built a reusable component library for a political survey platform
- Integrated backend APIs with PostgreSQL and Redis, and created data visualizations with Apache ECharts
- Refactored a legacy codebase into modular TypeScript for maintainability

Frontend Developer

2022 – 2023

Matrice AI

Full-time

- Cut Next.js landing page time to interactive from ~2 s to under 500 ms, fixing poor FCP and layout shift
- Rebuilt the annotation dashboard in React and Tailwind CSS to align with the new UI/UX
- Added responsive layouts, accessibility, and internationalization across the dashboard
- Developed backend services for the image annotation platform on AWS, with real-time monitoring hooks

PROJECTS

Labeling Harness | *Python, FastAPI, TypeScript, PostgreSQL, MinIO* github.com/Sagyam/labeling-harness

- Human-in-the-loop labeling tool for a Nepali–English code-switched ASR corpus, used to label 20+ hours of training data
- Fuses three independent ASR systems into seed transcripts and ranks segments for triage by model disagreement (1.7 s median per segment)
- Includes speaker diarization, overlap detection, forced alignment, Devanagari transliteration, and versioned dataset exports; 1,800+ tests

Linite | *Next.js, TypeScript, Drizzle, Turso, Vercel* linite.sagyamthapa.com.np

- Bulk installer for 200+ curated apps across 14 distros and 11 package managers, from APT to Flatpak
- Shareable app collections, cross-device install tracking, and daily metadata sync from Flathub, Snapcraft, AUR, and Repology; 34 GitHub stars

InfraWise | *Next.js, TypeScript, shadcn/ui, Recharts* infrawise.sagyamthapa.com.np

- Models cloud vs. on-prem total cost of ownership across 8 cost categories, including security and compliance
- Multi-year projections with inflation, breakeven, and sensitivity analysis over 80+ variables; 14 GitHub stars

WRITING

Technical Blog | *Interactive computer science explainers* blog.sagyamthapa.com.np

- [An Interactive Guide to Rate Limiting](#): 162 points and 46 comments on [Hacker News](#)
- Interactive guides to Bloom filters, HyperLogLog, Count-Min Sketch, and caching strategies
- Hands-on tutorials on OpenTelemetry tracing for Node.js and scaling PostgreSQL on Kubernetes

EDUCATION

Kantipur Engineering College

2018 – 2022

Bachelor of Engineering in Computer Engineering

Lalitpur, Nepal

TECHNICAL SKILLS

Languages & Frameworks: TypeScript, JavaScript, Python, SQL, Bash, React, Next.js, FastAPI

Cloud & DevOps: AWS, Azure, Linux, Docker, Terraform, Ansible, GitHub Actions, Azure Pipelines

Kubernetes: Helm, Argo CD, Flagger, Linkerd, Istio, Harbor, CloudNativePG

Security: Trivy, Falco, GitGuardian

Observability: Prometheus, Grafana, Loki, OpenTelemetry, Jaeger

Data & ML: PostgreSQL, MySQL, MongoDB, Redis, Kafka, PyTorch, NeMo, Whisper