STEPHEN ONYANGO ROY
DevOps & Platform Engineer
Nairobi, Kenya | +254 706 519 258 | steve.onyangoroy@gmail.com | Medium | LinkedIn | Github

PROFESSIONAL SUMMARY

DevOps & Platform Engineer with 5+ years running production workloads on Kubernetes, AWS, and Azure across Kenya, Uganda, and Burkina Faso, now specializing in self-hosted LLM infrastructure. Built and deployed an LLM serving and RAG stack (Ollama, LangChain, LangGraph, ChromaDB) with Helm and Prometheus/Grafana monitoring. Reduced cloud spend by 66% and led Terraform and GitOps adoption.

PROFESSIONAL EXPERIENCE

DevOps Engineer — Living Goods | *Apr 2022 – Present*

- Reduced AWS expenditure from approximately $60K to below $20K/month (over 66% reduction, representing more than $480K in annualised savings) through resource right-sizing, workload scheduling and ongoing FinOps practices.
- Led organisation-wide adoption of Terraform, migrating 1,000+ AWS resources (VPC, EKS, EC2 and networking) across development, staging, production and data warehouse environments into Infrastructure as Code through import workflows, with minimal configuration drift or production disruption.
- Designed, operated and maintained production cloud infrastructure across AWS, Azure, Kubernetes, Linux and Windows, supporting business-critical workloads across multiple environments and three countries, including EC2, Lambda, RDS, S3, IAM, VPC, CloudWatch, SQS and SNS.
- Led migration from self-managed Kubernetes (kOps) to Amazon EKS, improving cluster management, operational visibility and infrastructure resilience.
- Led migration from a single AWS account to a multi-account AWS Organization, implementing AWS Control Tower, Service Control Policies (SCPs) and AWS Config for centralised governance and compliance monitoring.
- Implemented GitOps practices using ArgoCD to manage Kubernetes cluster state consistently across development, staging, and production environments.
- Eliminated reliance on long-lived AWS access keys by migrating to SSO-based access, reducing persistent credential exposure risks.
- Designed and implemented an integrated observability platform combining Prometheus (EKS/EC2 metrics via node exporters), Loki/Promtail (centralised logging), Uptime Kuma (availability monitoring) and AWS Cost and Usage Reports (CUR) for cost visibility, avoiding additional CloudWatch agent and Cost Explorer API charges.
- Automated infrastructure and operational workflows using Terraform, AWS Lambda, EventBridge, API Gateway, Python, Bash, GitHub Actions (OIDC) and ArgoCD, reducing manual intervention and improving deployment repeatability.
- Diagnosed and resolved complex production incidents across applications, APIs, Kubernetes, cloud infrastructure, networking and databases, conducting root-cause analysis and implementing corrective actions.

DevOps Engineer | Project Nautilus — KodeKloud | *Feb 2020 – Mar 2022*

- Administered multi-server Linux environments and built Docker and Kubernetes workloads, troubleshooting networking, storage, and runtime issues.
- Automated infrastructure with Ansible and built Jenkins CI/CD pipelines.

Customer Technical Support Engineer — Raj Ushanga House Limited | *Dec 2018 – Mar 2022*

- Provided L2/L3 support for AWS- and Azure-hosted IoT platforms across Linux systems, cloud infrastructure, networking, and APIs.
- Investigated and resolved production issues involving service availability, performance, and distributed system dependencies.
- Built runbooks and troubleshooting procedures adopted across the support team to improve consistency and reduce resolution time.

Software Engineer | Andela | *February 2019 – March 2020*

- Developed and maintained software applications based on functional and technical requirements.
- Wrote, tested, debugged, and maintained application code to support software functionality and reliability.
- Investigated software defects, identified root causes, and implemented fixes and improvements.
- Collaborated with other developers and technical stakeholders throughout the software development lifecycle.

PROJECTS

Self-Hosted LLM Serving & RAG Platform (Demo Prototype) | llm project
Engineered a production-grade, self-hosted LLM inference and RAG platform on Kubernetes using Ollama, FastAPI, ChromaDB, and LangChain — with Helm packaging, Prometheus/Grafana observability, and a full CI pipeline delivering zero-cloud-dependency AI infrastructure deployable on 8 GB RAM.

Cloud Command Hub (Demo Prototype) | monitoring project
Built a full-stack AWS multi-account observability platform — single-pane-of-glass dashboard covering EC2/EKS fleet health, IAM posture, GuardDuty threat detection, FinOps cost attribution, and SLA tracking, backed by a Python/FastAPI async collector, deployed on EKS via GitHub Actions CI/CD with Prometheus/Loki for metrics and log aggregation.

CORE COMPETENCIES

- **Cloud & Infrastructure:** AWS (EC2, EKS, Lambda, RDS, S3, VPC, IAM, CloudWatch, SQS, SNS), Azure, GCP — multi-account/multi-cloud production operations
- **Kubernetes & Containers:** Amazon EKS, Docker, ArgoCD, GitOps, workload & service configuration
- **IaC & Automation:** Terraform, CloudFormation, Ansible, Python, Bash, GitHub Actions, OIDC
- **Reliability & Observability:** Incident response, root-cause analysis, Prometheus, Grafana, CloudWatch, centralized logging, alerting
- **Integration & APIs:** REST APIs, service-to-service integration, API Gateway, authentication, event-driven workflows
- **Security & Networking:** IAM, RBAC, VPC, bastion access, secrets management, Vault
- **Languages:** Python, Bash, JavaScript, TypeScript

EDUCATION & CERTIFICATIONS

BSc, Information Technology — Jomo Kenyatta University of Agriculture and Technology (JKUAT)

AWS Certified DevOps Engineer – Professional · Certified Kubernetes Administrator (CKA) · AWS Certified Developer – Associate · AWS Certified Cloud Practitioner · KCNA · ITIL Foundations