

David Quirós A.

MLOps / AI Infrastructure Engineer | DevOps / Site Reliability Engineer

San José, Costa Rica (Remote) | dquiros@itelecomcr.com | (506) 8790-0333

linkedin.com/in/davidquiros | github.com/dquiros1985 | resume.itelecomcr.com

Authorized to work in Costa Rica. Available worldwide as a contractor or via Employer of Record.

PROFESSIONAL SUMMARY

MLOps / AI Infrastructure Engineer with 15+ years in systems and cloud infrastructure, 8+ of them leading DevOps and SRE across AWS, Azure and GCP. I take machine-learning and LLM systems and make them cheap, fast and safe to run in production, the same discipline I have used to run Kubernetes, CI/CD and observability for retail, banking, telecom and a regulated clinical (GxP) platform. I lead and mentor engineering teams, I build infrastructure as code, and I pick up new stacks quickly.

TECHNICAL SKILLS

MLOps / AI: LLM serving (vLLM, TGI, Ollama), quantization (AWQ, GGUF, GPTQ), RAG and vector stores, tool-calling agents and MCP, LLM evaluation (promptfoo), LLM observability (Langfuse), AI security and guardrails (OWASP LLM Top 10)

Cloud: AWS, Microsoft Azure, Google Cloud Platform, AWS Bedrock, Azure OpenAI, Vertex AI

Containers and Orchestration: Kubernetes, Amazon EKS, Rancher, Docker, Helm, ArgoCD

Infrastructure as Code: Terraform, CloudFormation, ARM templates, Ansible

CI/CD: GitLab CI/CD, GitHub Actions, Jenkins, Bamboo, canary and blue/green deployment, DAST integration

Observability: Prometheus, Grafana, OpenTelemetry, Dynatrace, Datadog, CloudWatch, New Relic, PagerDuty

Security and Compliance: Trivy, HashiCorp Vault, IAM, PCI DSS, HIPAA, GDPR, GxP validation

Languages and Scripting: Python, Bash, Ruby, PowerShell, SQL

Databases: PostgreSQL, MySQL, Oracle, SQL Server, pgvector, Memcached

PROFESSIONAL EXPERIENCE

MLOps / AI Infrastructure Engineer | Independent (Remote) | Sep 2025 - Present

- Full-time self-directed transition into AI and ML engineering: LLM serving, RAG, agents and LLMOps, built on 15+ years of production infrastructure experience.
- Evaluating and serving open and hosted models (Llama-class, Anthropic, OpenAI) with vLLM, TGI and Ollama; quantization (GGUF, AWQ, GPTQ) and GPU throughput-versus-latency tuning.
- Building tool-calling agents using function calling and the Model Context Protocol, and RAG pipelines over vector stores, with automated evaluations (promptfoo) and LLM observability (Langfuse).
- Applying SRE discipline to the AI lifecycle: CI/CD for models, autoscaling GPU workloads on Kubernetes, cost and latency controls, and AI security and guardrails.

Senior DevOps Engineer | Clario SRL, Costa Rica | *Apr 2023 - Aug 2025*

- Rolled out the Dynatrace Operator across 30+ Kubernetes clusters running 2,000+ concurrent pods. Cut mean time to recovery from 4 hours to 2.2 on the 15-20 critical incidents handled each month.
- Tuned container-native GitLab CI/CD across 120+ microservices using Docker layer caching and Helm chart validation. Pipeline runtime fell from 40 minutes to 14, and cluster deployment latency dropped 30 percent.
- Built the ARB-approved dynamic GitOps pipeline: one reusable CI/CD framework with Helm Hub charts, ephemeral Kubernetes (Rancher and EKS) test namespaces, canary releases and DAST scanning. Onboarded 85 microservices, moved deployments from bi-weekly to 5+ per day, and brought change-failure rate from 14 percent to under 3 percent.
- Wrote 40+ GxP-compliant runbooks covering Kubernetes RBAC, network policies and cluster disaster recovery. Engineer onboarding went from 3 weeks to 4 days, with a 100 percent pass rate on external compliance audits.

Lead DevOps / Platform Engineer | Confidential, Fortune 500 enterprise software company | *Sep 2019 - Mar 2023*

- Led a cross-functional team of 8 engineers running 45 Kubernetes clusters and 1,200+ worker nodes. FinOps work cut annual cloud spend 28 percent while platform uptime moved from 99.9 to 99.99 percent SLA.
- Built cloud CI/CD pipelines and provisioning toolsets that automated build, test, integration and deployment for enterprise software at cluster scale.
- Owned monitoring, alerting and daily administration of production cloud environments, including incident response and release execution with development teams.
- Set the team automation and integration standards across source control, CI, infrastructure and deployment automation, containers, orchestration and cloud.

Principal DevOps Architect | Equifax Inc., Corporate Services | *Nov 2018 - Sep 2019*

- Managed multi-cloud (AWS, Azure, GCP) infrastructure with Chef and Ansible; part of the team that built the PCI-compliant cloud-based Java CI/CD pipeline for financial tools.
- Deep AWS delivery across VPC, EC2, S3, ELB, Auto Scaling, EBS, RDS, IAM, CloudFormation, Route 53, CloudWatch, CloudFront and CloudTrail, building secure network zones with security groups, ACLs and route tables.
- Authored CloudFormation templates and Terraform IaC for staging and production; enforced IAM roles and MFA account-wide; deployed monitoring and APM with Datadog, AppDynamics, Nagios and CloudWatch.
- Defined the Costa Rica business unit cloud strategy, migration and application consolidation; ran Lean and Agile delivery with Confluence, Jira, Bitbucket and Bamboo.

Senior DevOps Engineer | Snap Finance, USA / UK | *Jan 2018 - Nov 2018*

- Managed AWS infrastructure with automation and orchestration tooling (Chef, Ansible); part of the team building the PCI-compliant cloud-based Java CI/CD pipeline for financial tools.
- Built load-balanced, auto-scaled, cost-efficient high-availability environments; S3 lifecycle archiving, RDS snapshots and launch-configuration images.
- Authored CloudFormation and Terraform templates for staging and production; Route 53 DNS for highly available applications; production monitoring and alerting via CloudWatch.

DevOps / SRE Consultant | iTelecom Costa Rica (own company) | *Jan 2017 - Dec 2017*

- Migrated two complete on-premise infrastructure stacks to the cloud end to end.
- Advised on design, build and rollout of cloud-based DevOps frameworks supporting migration of high-visibility customer-facing products.
- Provided technical guidance, knowledge transfer and mentorship; managed relationships with multiple suppliers and internal teams.

Cloud Project Manager | Hewlett Packard Enterprise | *Jun 2016 - Dec 2016*

- Managed all phases of cross-functional cloud implementation projects to time and budget; developed the Master Implementation Plan with Workstream Leads and the PMO.
- Identified high-complexity integration and conflict issues including cost-benefit analysis and resource estimating, and escalated with analyzed solution options.

Systems, Virtualization and Cloud Roles | Various | *2008 - 2016*

- Linux/Unix administration (Ubuntu, RHEL/CentOS, Debian), enterprise Windows (Server 2008-2019, Active Directory, Exchange, IIS, SCCM), and VMware ESXi/vSphere (HA/DRS clusters, vMotion, P2V/V2V migrations) before specializing in DevOps. Detail available on request.

SELECTED PROJECTS

- Terraform module for GPU inference (github.com/dquiros1985/terraform-aws-gpu-inference): one apply provisions EKS with Karpenter-managed GPU capacity, the NVIDIA GPU Operator, vLLM serving and Prometheus/Grafana observability.
- LLM Serving Platform: Llama 3 8B served on vLLM behind a FastAPI endpoint, with automated evaluations (promptfoo), latency and token-cost observability (Langfuse), and GPU pods on Kubernetes using quantization and autoscaling for cost control.
- RAG Document Assistant: ingestion, chunking, embeddings and pgvector retrieval producing grounded, cited answers, with retrieval-quality and hallucination checks via promptfoo.
- Résumé Site (resume.itelecomcr.com): Terraform-provisioned static site on private S3 with CloudFront Origin Access Control and ACM DNS-validated TLS, deployed by script with cache invalidation.

CERTIFICATIONS

Certified Kubernetes Administrator (CKA); HashiCorp Certified: Terraform Associate; AWS Solutions Architect - Associate; Google Associate Cloud Engineer (ACE); Microsoft Azure Administrator Associate (AZ-104); Linux Server Administrator - Advanced.

Additional: PMI Certified Associate in Project Management (CAPM); Cisco Certified Network Associate (CCNA); Cisco Unified Communications Manager; Scrum Fundamentals Certified.

LANGUAGES

English (Advanced). Spanish (Native).