

RICHARD OLANIYAN

Toronto, Canada

Email: rich.olaniyan@gmail.com **Phone:** 514-677-8414

GitHub: github.com/rolani **LinkedIn:** linkedin.com/in/richard-olaniyan

Professional Summary

Senior Infrastructure Engineer with over 8 years of experience building scalable, reliable AI and cloud infrastructure systems. Proven expertise designing and operating large-scale distributed systems with Kubernetes, GPU/TPU workloads, and multi-cloud environments (AWS, GCP). Track record of 40–50% improvements in system reliability, performance optimization, and infrastructure efficiency. Deep experience with observability systems (Prometheus, Grafana, Splunk), Infrastructure as Code (Terraform), data pipelines (Airflow, Spark), and developer productivity tooling. Published researcher with patent in distributed AI training systems. Strong technical leadership, stakeholder collaboration, and passion for supporting research and engineering teams to achieve frontier model development.

Technical Skills

Cloud & Infrastructure: AWS, GCP (Compute, GKE, BigQuery), Azure, Kubernetes, Infrastructure as Code (Terraform, Helm), Distributed Systems

ML Infrastructure: GPU Cluster Management, Distributed Training Systems, AI Model Serving, ML Pipelines

Data Infrastructure: Airflow, Spark, Hadoop/MapReduce, ETL Pipelines, Data Lakes, Redshift

Observability & Monitoring: Prometheus, Grafana, Datadog, OpenTelemetry, Alerting, Dashboards

Programming: Python, Bash, Java, Rust (learning), SQL

DevOps & Automation: GitOps (ArgoCD, Flux), CI/CD, Jenkins, Ansible, Karpenter, Service Mesh

Systems & Networking: Linux Kernel Tuning, TCP/IP, Load Balancing, High-Performance Computing

Security: IAM, RBAC, TLS, Vault, Zero Trust, Security Best Practices

Operations: Incident Response, On-Call Rotations, SLO/SLI Development, Reliability Engineering

Professional Experience

OneVest Inc.

Toronto, Canada

Senior Platform Engineering Developer (Tech Lead)

Dec 2023 – Present

- Led infrastructure strategy for cloud platform serving enterprise customers, mentoring 4 engineers and driving technical architecture decisions for AWS/Kubernetes-based distributed systems at scale.
- Built observability infrastructure (Prometheus, Grafana, Datadog) and established incident response processes including postmortem reviews and on-call rotations, reducing MTTR by 50% through automated alerting and proactive monitoring.
- Developed scalable Infrastructure as Code (Terraform) for AWS multi-account architecture, data lake systems, and Kubernetes autoscaling with Karpenter, reducing provisioning time 40% and compute spend 20% while enforcing security best practices.
- Launched developer productivity initiatives and collaborated with cross-functional stakeholders to understand infrastructure needs, reducing onboarding time 40% and accelerating deployment velocity across engineering teams.

SociVolta Inc.

Montreal, Canada

Software Engineer, Cloud Infrastructure & Data Systems

Oct 2021 – Nov 2023

- Deployed and scaled AI/ML models on Kubernetes clusters with GPU-enabled workloads and built large-scale distributed data systems across AWS and GCP, ensuring reliable model serving and optimizing for performance and cost efficiency.
- Built data infrastructure and ETL pipelines using Airflow, Spark, and AWS EMR (MapReduce) supporting ML feature engineering, improving data freshness SLAs by 35% and partnering with data science teams to understand compute needs.
- Implemented GitOps-driven CI/CD pipelines with Flux and Helm, increasing deployment frequency by 35% and reducing rollback incidents by 25% through automated testing and gradual rollouts.
- Developed comprehensive observability solutions using Grafana and Prometheus with SLO-based alerting and dashboards, cutting incident triage time by 35% and improving system visibility across distributed infrastructure.

Ericsson Canada
Cloud Software Engineer, AI Infrastructure

Montreal, Canada
Apr 2020 – Jul 2021

- Designed and implemented distributed AI training infrastructure across edge and cloud environments, improving training throughput by 30% through optimized resource scheduling, synchronization algorithms, and networking optimizations.
- Benchmarked and optimized Kubernetes clusters for GPU/TPU-intensive AI workloads, increasing cluster utilization by 20% through performance tuning, resource allocation strategies, and monitoring tools for ML infrastructure.
- Developed synchronization algorithms for large-scale distributed training systems, resulting in patent: *“Edge device, edge server and synchronization thereof for improving distributed training of an AI model in an AI system”* (US Patent Application).
- Collaborated with research teams to understand ML training requirements and translate them into scalable infrastructure solutions, enabling data-driven optimization of training pipelines and system performance.

McGill University
Cloud Research Assistant & PhD Researcher

Montreal, Canada
Jan 2016 – Mar 2020

- Conducted research on distributed systems, cloud computing, and large-scale infrastructure optimization, publishing findings in IEEE and ACM peer-reviewed venues.
- Designed and built distributed simulators for hybrid and fog cloud architectures, benchmarking system performance at scale.
- Developed IoT synchronization algorithms for distributed data collection systems, improving throughput by 30% through optimized coordination protocols.

Key Projects

Large-Scale GPU Cluster Automation for ML Training

Technologies: AWS ECS/EKS, Kubernetes, Python, Terraform

Developed automation framework and Python libraries for provisioning and managing GPU clusters on AWS, accelerating ML experimentation and enabling researchers to scale training workloads efficiently. Implemented monitoring and cost tracking for multi-user GPU infrastructure.

Data Infrastructure Cost Optimization & Scaling

Technologies: Kubernetes, Apache Spark, EMR, Terraform, Karpenter, Spot Instances

Architected autoscaling data infrastructure combining Karpenter with EMR integration, reducing compute costs by 35% while improving analytics performance. Implemented governance policies for efficient resource utilization across data pipelines.

Education

McGill University
PhD in Computer Science (Distributed Systems & Cloud Computing)

Montreal, Canada
2021

University of Edinburgh
MSc in Computer Science

Edinburgh, UK
2015

Obafemi Awolowo University
BSc in Computer Engineering

Ile-Ife, Nigeria
2011

Certifications

HashiCorp Certified: Terraform Associate (003)
AWS Certified Solutions Architect – Associate
Professional Scrum Master I (PSM I)

Oct 2024
Dec 2022
Sep 2021