

RAVISH TIWARI

Solutions Architect · Cloud, DevOps & MLOps · Tech Leadership & Pre-Sales · GenAI / LLMOps
ravishktiwari[at]hotmail[dot]com | +91 | India | medium.com/@ravishtiwari
AWS CSA · AWS CDA · CKA · Azure AZ-301 · Zend ZCE

PROFESSIONAL SUMMARY

Technology leader with 15+ years spanning Solutions Architecture, DevOps, MLOps, Technical Leadership, and Pre-Sales, delivering production-scale platforms for clients across Europe and the US. Hands-on across all three major clouds (AWS, GCP, Azure, plus Azure Stack): landing zones, serverless and event-driven systems, enterprise Kubernetes, and end-to-end GenAI / GenMedia training and inference.

A genuine polyglot (Python, Node.js, PHP, Ruby on Rails and Go) who has shipped backend services, infrastructure-as-code, frontend, and hybrid-mobile applications. Core engineering philosophy: analyze the problem first and pick the most optimal tool or framework for it, rather than forcing a familiar stack onto every problem. Treats cost and reliability as first-class design constraints alongside performance.

CORE EXPERTISE

Disciplines: Solutions Architecture, DevOps, MLOps / LLMOps, Technical Leadership, Pre-Sales & Solutioning

Cloud Platforms: AWS, Google Cloud, Microsoft Azure, Azure Stack, DigitalOcean

AI / ML & GenAI: LLM & image-model inference (vLLM, FastAPI, JAX), fine-tuning (LoRA), MLOps lifecycle, Vertex AI, SageMaker, Amazon Bedrock, Cloud TPU, KubeRay, Weights & Biases

Languages & Frameworks: Go, Python, Node.js, PHP, Ruby on Rails; frontend and hybrid-mobile development

Containers & Orchestration: Docker, Podman, Kubernetes (self-hosted / kubeadm, GKE, AKS, EKS, Rancher RKE)

IaC & Automation: Terraform, AWS CDK, CloudFormation, Ansible

CI/CD: Jenkins, GitLab CI, GitHub Actions, AWS CodePipeline / CodeBuild, Azure DevOps

Serverless & Event-Driven: AWS Lambda, API Gateway, SNS / SQS, DynamoDB, CloudWatch, Google Cloud Functions

Security / DevSecOps: SonarQube, Sonatype Nexus IQ, HP Fortify, HP WebInspect, Palo Alto Prisma Cloud, Black Duck, Sophos Cloud Optix, Trivy

PROFESSIONAL EXPERIENCE

Senior DevOps & MLOps Architect

Apr 2021 – Present

<redacted>

- **MLOps platform ownership:** architect and operate ML pipelines from training to inference across AWS and GCP for GenAI and GenMedia workloads; lead PoCs for emerging tooling and drive cloud cost optimization.
- **LLM inference at scale:** deployed multi-GPU Llama2-based LLM serving on GKE using vLLM and FastAPI on A100 and L4 GPUs, with GPU autoscaling via KubeRay.
- **Generative media:** shipped SDXL text-to-image inference on GKE (JAX pipeline on Cloud TPU plus L4 GPUs), hosted large image models on Cloud TPU v5e, and deployed a Gemma3 LoRA fine-tuned model on Vertex AI.
- **Model fine-tuning & training:** fine-tuned Llama2-based models on custom datasets with ML teams; enabled PyTorch / TensorFlow training on SageMaker, EC2 Deep-Learning VMs, GCP Compute, and TPU v3; optimized training code for multi-cloud use with cost / ROI tracking.
- **Landing zones & IaC adoption:** designed multi-account (AWS) and multi-project (GCP) architecture; led an org-wide Terraform module rollout with a naming / tagging factory, network factory, Go-based unit tests, documentation, and project / module boilerplates.
- Stood up ML experiment tracking (Weights & Biases) on Kubernetes.

DevOps & Cloud Architect

Apr 2018 – Mar 2021

<redacted>

- **Pre-Sales to delivery:** partnered with sales and Pre-Sales teams and engaged customers directly to convert pain points and constraints into feasible, cost-aware multi-cloud solutions across AWS, GCP, Azure, and Azure Stack.
- **400M-document serverless re-architecture:** re-platformed a high-GPU SageMaker Elasticsearch seeding workload onto AWS Lambda, reaching 2,000+ concurrency and cutting seeding of 400M+ searchable documents from roughly 6 hours to under 30 minutes; delivered the full pipeline as Terraform IaC and mentored data scientists in converting notebooks to Lambda.
- **Enterprise Kubernetes on Azure:** designed and deployed customer Kubernetes infrastructure on the Microsoft platform, including HA clusters via kubeadm with an external etcd cluster, HPA autoscaling, and Prometheus / Grafana monitoring.
- **Enterprise DevSecOps:** built a declarative Jenkinsfile pipeline on Azure with Ansible deployments, integrating SonarQube quality gates, Sonatype Nexus IQ, HP Fortify, and HP WebInspect; artifacts to Nexus and Docker registry with deployment to AKS.
- **Multi-tenant SaaS:** re-architected a Node.js / React application into a multi-tenant SaaS platform serving multiple customers from a single deployment at a fraction of the cost, decoupling the UI to S3 and CDN.
- **Monolith to serverless migration:** led Pre-Sales-to-implementation modernization: UI to static S3, REST platform to API Gateway, partner APIs to Lambda, and partner authentication to Amazon Cognito.
- **Serverless ML & automation:** served pickle ML models as Lambda REST APIs with on-demand model loading and made Python applications AWS IAM role-aware; built a scheduled-job solution with CloudWatch, SNS, SQS, Lambda, and DynamoDB, and Dockerized standalone services on ECS with Jenkins / ECR CI/CD and CloudWatch-to-Slack alerting.

Technical Lead

Dec 2011 – Apr 2018

<redacted>

- **FullBridge platform (Ruby on Rails):** served as Technical Lead for the FullBridge platform, owning architecture and delivery of the Ruby on Rails web application alongside the engineering team.
- **Time-Magazine-featured product:** delivered a hybrid iOS / Android mobile application featured in Time Magazine's 2017 list, with APIs on AWS and deployment automation via Jenkins (backend) and Fastlane (mobile).
- **Open API platform for partners:** delivered a partner Open API platform on AWS API Gateway and Lambda with CodePipeline CI/CD, helping customers reduce AWS spend on these services.
- **Reporting platform:** built a reading-list conversion and reporting platform with automated, notification-driven report delivery and flexible, customizable background-job workers.
- Partnered with technical architects across geographically distributed teams to address customer challenges and lead implementation with the development team.

Early Career: PHP / Web Developer

2006 – 2010

[PHPAuction.net \(Remote\)](#) · [eBiz.com Pvt. Ltd.](#)

- Built PHP and web solutions from problem briefs, demonstrated them to customers, and supported live customer sites; the hands-on engineering foundation for later architecture and leadership roles.

CERTIFICATIONS

- AWS Certified Solutions Architect, Associate
- AWS Certified Developer, Associate
- CNCF Certified Kubernetes Administrator (CKA)
- Microsoft Azure AZ-301, Certified Solutions Architect (Design)
- Zend Certified PHP Engineer (ZCE)

TECHNICAL WRITING & THOUGHT LEADERSHIP

- Publishes practitioner-focused engineering writing on Medium and LinkedIn for a technical audience.
- **Featured piece:** ["FastAPI vs Fiber: A Benchmark, Not an Opinion"](#), a data-driven Go-vs-Python comparison that reflects a measure-first approach to choosing tools.

AWARDS & RECOGNITION

- Technology Champion of the Year (2022–23), <redacted>
- Rising Star of the Year (2021–22), <redacted>
- Star Performer of the Quarter (x2) and Extra Mile Award (x2), <redacted>
- Annual WOW Award, plus a nomination for the Annual Agility Award, <redacted>

EDUCATION

Qualification	Board / University	Year	Location
Master of Computer Applications (MCA)	Sikkim Manipal University	2011	---

References available on request.