

NIKHIL SINGH

DevOps & Backend Engineer

 +91-9027500166 |  singhnikhil2508@gmail.com |  resume.codenickk.com |  LinkedIn/NikhilSingh

PROFESSIONAL SUMMARY

DevOps + Backend engineer with 5 years building cost-optimized, high-throughput platforms across AWS and GCP. Works end to end — multi-cloud Kubernetes (EKS + GKE), GitOps CI/CD, Infrastructure as Code, and event-driven Go microservices — with a track record of cutting cloud spend, replacing brittle infrastructure with tooling teams can rely on, and root-causing incidents that span services and clouds.

CORE SKILLS

Cloud (AWS): EKS, ECS Fargate, EC2, Lambda, EventBridge, API Gateway, ALB, RDS (PostgreSQL/MySQL, IAM auth), ElastiCache (Valkey/Redis), MSK, S3, Route 53, Cognito, SQS/SNS, ECR, KMS, Secrets Manager, GuardDuty, IAM/IRSA

Cloud (GCP): GKE, Artifact Registry, Cloud SQL, Cloud Storage, BigQuery, Secret Manager, Cloud Build, Workload Identity

Kubernetes & Platform: Kubernetes (multi-cloud EKS + GKE), Helm (layered charts), ArgoCD (GitOps), Envoy Gateway (Gateway API, HTTPRoute, SecurityPolicy/JWT), Kustomize, cert-manager, OPA

IaC & CI/CD: Terraform (modules, remote state, drift control), AWS CDK, Serverless Framework, CloudFormation; GitHub Actions, Bitbucket Pipelines, Jenkins, ArgoCD, OIDC federation, immutable SHA-pinned images, Docker

Backend & Languages: Go (chi, gin, pgx, gRPC/protobuf, grpc-gateway), Python (FastAPI, boto3), Bash, PowerShell, SQL

Messaging & Streaming: NATS JetStream, Apache Kafka (Strimzi/MSK), RabbitMQ/AMQP, MQTT

Databases: PostgreSQL, MySQL, ClickHouse, Redis/Valkey, OpenSearch, DynamoDB; Flyway migrations, backup/restore automation, slow-query tuning

Observability: OpenTelemetry (OTLP, W3C/B3 propagation), Prometheus, Grafana, Loki, OpenObserve, CloudWatch, distributed tracing

Security: IAM least-privilege (IRSA), RDS/Valkey IAM auth, Envoy Gateway JWT/SecurityPolicy, Firebase Auth, Cognito (OIDC/JWT/JWKS/OAuth 2.0), KMS (customer-managed keys), Secrets Manager, WAFv2, ACM/TLS, multi-tenant isolation

PROFESSIONAL EXPERIENCE

Good AI Labs — DevOps & Backend Engineer

Apr 2026 – Present

- Own multi-cloud CI/CD and GitOps delivery for a high-throughput, real-time sports-odds & order-book trading platform (~19 Go microservices on AWS EKS + GCP GKE) — every merge builds an immutable sha-`<commit>` image that ships via ArgoCD, with a scoped, non-blocking prod-test deploy job and a GitHub Environments approval gate to promote the very same image forward.
- Made bet-placement rate limits operator-tunable at runtime — under a 10,000-wager burst test that was throttling ~64% of attempts, root-caused the limiter to the order-book service (not the edge gateway), moved its caps into a Backoffice settings surface that persists to Postgres and mirrors to Redis, and surfaced a standards-compliant Retry-After via gRPC RetryInfo.
- Designed operator-editable display names for leagues and markets that survive every vendor re-sync — override columns plus a market-template table, a COALESCE(override, name) read rule swapped in at the three data exits, and a tri-state dictionary API backed by a 15-second in-memory cache — verified propagating across 4,800+ leagues within seconds, with zero reindex machinery.
- Repaired the severed writer behind a blank Match Overview — built a durable stats/timeline projector off the NATS livescore stream with deterministic content-hash incident IDs (heal-on-conflict) and shipped safe in-place dedup migrations, cutting one page from 667 noisy rows to 54 event-grade rows and surviving a live rolling-deploy migration incident.
- Designed and implemented a move from static, non-rotating database passwords to short-lived (15-minute) AWS RDS IAM tokens across 9 Go services — minted per connection via a pgxpool hook with TLS enforced

end to end — collapsing three roles across four databases into a single least-privilege IAM role bound through IRSA.

- Built an automated cloud-cost watchtower on AWS Lambda + EventBridge that pulls Cost Explorer data weekly and monthly, renders a per-service PDF with color-coded deltas, and posts it to Slack — then used it to root-cause a tripling monthly AWS bill to cross-AZ data transfer and to catch a runaway GCP logging sink quietly burning ~\$2K/month.

Enetro (Readywire Subsidiary) — Cloud Automation & AI Platform Engineer

Jan 2026 – Jul 2026

- Architected a reusable, module-driven AWS platform in Terraform (VPC, ALB, ECS Fargate, API Gateway, RDS, Route 53) with isolated per-environment state as blast-radius boundaries — eliminating infrastructure drift across dev, beta, sandbox, and prod, and underpinning 40+ event-driven Lambda microservices (Python, boto3, Serverless Framework).
- Engineered zero-credential CI/CD on Bitbucket Pipelines via OIDC federation — secure ECR-to-ECS rolling deployments with changeset-gated builds and Secrets Manager injection, cutting a code push to a hands-off production rollout in ~5 minutes.
- Designed a multi-service ECS Fargate architecture splitting a request-rate-scaled FastAPI tier (behind ALB) from an SQS-queue-depth-scaled Celery worker tier, enabling independent scaling and zero-downtime rolling deploys.
- Built an in-house observability platform on AWS serverless primitives (Lambda OTLP proxy, OpenTelemetry Collector on Fargate) processing 10M+ events/month for under \$80 — with end-to-end distributed tracing (W3C + B3 propagation) across FastAPI, Celery, and SQS, replacing Datadog/New Relic.
- Built multi-tenant isolation on AWS Cognito — Terraform-managed user pools, JWT enrichment via Pre-Token-Generation Lambda, and a custom API Gateway authorizer validating signatures against JWKS — over a least-privilege IAM, customer-managed KMS, and Secrets Manager security baseline.
- Operated a production AI pipeline (transcription via AWS Transcribe / SarvamAI, translation, LLM feature extraction) with LiteLLM for provider-agnostic routing, Instructor for schema-validated output, and Langfuse for per-tenant token-cost attribution — catching a prompt bug consuming 4× average tokens.

Readywire — DevOps Engineer

Jun 2023 – Jan 2026

- Engineered an event-driven RDS optimization pipeline using AWS Lambda to dynamically toggle slow-query logging on CloudWatch CPU-spike alarms (ApplyMethod=immediate, no DB restart), enabling root-cause analysis that improved database performance by 80%.
- Accelerated application delivery by 75% (3 hours to 45 minutes) by architecting an end-to-end Jenkins CI/CD pipeline for Acumatica using PowerShell, Batch scripting, and REST APIs.
- Reduced manual database operations by 65% by orchestrating secure MySQL backup pipelines in Bash — encrypted credential handling, on-the-fly 7z compression, schema validation with retries, and hands-off S3 routing with one-day pre-signed download URLs.
- Created an operational tooling library (Python/Bash): cross-account S3 migrator with dry-run and auto-generated bucket policies, parallel mydumper/myloader RDS shuttle (16 threads/side with human-readable error mapping), and a GuardDuty weekly findings reporter delivered to the security team.
- Hardened production resilience by automating RDS snapshot backups with retention and scheduled restore drills, alongside unattended EC2 patching and volume lifecycle management — ensuring recoverability and cutting manual maintenance toil.

Tata Consultancy Services — Associate Engineer

Aug 2021 – Jun 2023

- Automated user-onboarding and account provisioning in Python, replacing static Excel trackers with a self-service workflow and cutting new-hire setup from hours to minutes.
- Provided L1/L2 production support on a rotational on-call rota — resolving incidents within SLA and escalating critical issues via ServiceNow with clear root-cause documentation, and building Python/Excel automation that turned support data into hands-off daily status reports.

EDUCATION

B.Sc. Computer Science
Invertis University, Bareilly

2018 – 2021