

DevOps Engineer Exclusive job

Job Information

Recruiter
NEXUS Corporation

Job ID
1616376

Industry
IT Consulting

Job Type
Permanent Full-time

Location
Tokyo - 23 Wards

Salary
Negotiable, based on experience

Refreshed
November 4th, 2026 06:00

General Requirements

Minimum Experience Level
Over 3 years

Career Level
Mid Career

Minimum English Level
Business Level

Minimum Japanese Level
Business Level

Minimum Education Level
Bachelor's Degree

Visa Status
Permission to work in Japan required

Job Description

We are seeking a DevOps Engineer with strong experience in Azure, Kubernetes, Terraform, CI/CD, and LLMOps to design and operate AI-focused delivery platforms. The role focuses on infrastructure, Kubernetes, Terraform, CI/CD, LLMOps, security, observability, reliability, and production deployment, while driving platform engineering standards and supporting multiple engineering teams.

Key Responsibilities:

- Design, provision, and operate infrastructure supporting AI application workloads across development, staging, and production environments.
- Own Infrastructure-as-Code definitions using Terraform, including module design, state management, environment segregation, and drift detection.
- Design and operate Kubernetes workloads, including namespace strategy, resource governance, autoscaling, and network policy enforcement.
- Define environment topology, environment promotion strategy, and configuration management across environments.
- Design, implement, and maintain CI/CD, including reusable workflow libraries and shared templates.
- Define branching, versioning, tagging, and release management strategy across multiple repositories and teams.

- Implement automated quality gates covering linting, unit and integration testing, static analysis, vulnerability scanning, and policy checks.
- Implement progressive delivery patterns, including blue/green and canary deployments, feature-flagged rollout, and automated rollback.
- Own release readiness verification, including rollback rehearsal and post-rollback regression validation for both new and existing user paths.
- Reduce build and deployment cycle time through caching, parallelisation, and pipeline instrumentation.
- Build and operate deployment pipelines for LLM-integrated applications, RAG services, and vector store dependencies.
- Automate AI evaluation pipelines (RAGAs, Langfuse, Phoenix) as pre-deployment gates, with defined thresholds for retrieval quality and response consistency.
- Integrate guardrail and grader checks into CI/CD so that Responsible AI validation is enforced automatically rather than manually.
- Own secrets and credential management, including rotation policy and least-privilege access.
- Implement identity and access controls.
- Integrate static code analysis and open-source vulnerability scanning into pipelines, including agentic AI scanning tooling used on the program.
- Maintain audit-ready evidence of deployments, approvals, and access changes to satisfy client governance requirements.
- Implement and maintain observability standards across logging, metrics, tracing, and alerting.
- Own backup, restore, and disaster recovery design, including periodic restore validation.
- Define and enforce platform engineering standards, deployment checklists, and change management discipline across teams.
- Produce and maintain architecture diagrams, runbooks, and operational documentation.
- Support client-facing technical discussions on release governance, security posture, and operational readiness.

Required Skills

What are we looking for:

- Minimum of 5 years of hands-on DevOps, Platform Engineering, or Site Reliability Engineering experience.
- Experience supporting multiple engineering teams from a shared platform or centre-of-excellence model.
- Exposure to AI-enabled, data-intensive, or high-throughput API workloads is strongly preferred.
- Strong hands-on depth in Azure Kubernetes Service (AKS), including upgrade strategy, node pool design, and workload isolation.
- Working expertise across Azure Container Registry, Azure Key Vault, Azure Monitor and Log Analytics, Application Gateway / Front Door, Azure Storage, and Azure networking.
- Strong understanding of Microsoft Entra ID, Azure RBAC, managed identities, and service principal governance.
- Strong hands-on experience with GitHub Actions, including reusable workflows, composite actions, self-hosted runners, and environment protection rules.
- Strong hands-on experience with Azure DevOps, including Pipelines (YAML), Repos, Artifacts, and Environments with approval gates.
- Experience with artifact management, versioned releases, and dependency promotion across environments.
- Strong hands-on expertise in Terraform, including module authoring, remote state, workspaces, and policy-as-code.
- Strong scripting capability in Python and Bash for automation and tooling.
- Strong working knowledge of container fundamentals, image optimisation, and multi-stage builds.
- Hands-on experience with Docker-based workflows.
- Strong Kubernetes operational capability, including deployments, services, ingress, secrets, config maps, probes, RBAC, HPA, and resource quotas.
- Understanding of LLM application architectures, RAG pipelines, and vector database deployment and operational characteristics.
- Experience integrating AI evaluation frameworks (RAGAs, Langfuse, Phoenix) into automated pipelines.
- Understanding of prompt versioning, prompt registry patterns, and configuration-driven prompt deployment.
- Understanding of guardrail enforcement and Responsible AI validation within a delivery pipeline.
- Strong understanding of IAM, OAuth2 / OIDC, and secure API design.
- Experience with secrets management, credential rotation, and least-privilege enforcement.
- Experience integrating SAST, dependency, container, and IaC scanning into pipelines.
- Hands-on experience with Prometheus, Grafana, and OpenTelemetry.
- Strong incident management capability, including structured root cause analysis and preventive action tracking.
- Familiarity with data pipeline deployment and scheduled workload orchestration.
- Understanding of application architecture patterns, including microservices, REST APIs, asynchronous processing, and streaming, sufficient to troubleshoot across the stack.
- Strong ownership and accountability for platform stability and delivery outcomes.
- Ability to communicate technical trade-offs, risk, and operational impact to both engineering and business stakeholders.
- Strong collaboration across multiple engineering pods, with the maturity to standardise rather than customise per team.
- Reliability and safety mindset, with a bias towards reversible, verifiable change.
- Openness to feedback and continuous improvement.

Language Requirements:

- English: Working proficiency required.
- Japanese: Working proficiency required.

Company Description