

Dibya Darshan Khanal

Cincinnati, OH 45220 | 646-842-4641 | dibya.ddk@gmail.com
dibyadarshankhanal.com.np | linkedin.com/in/dibya-darshan-khanal | github.com/Dibae101

SUMMARY

Site Reliability and DevOps Engineer with 5+ years building and operating production-grade cloud infrastructure on AWS and Azure, sustaining 99.95%+ uptime for distributed systems serving 100K+ users. Specialized in CI/CD automation, Kubernetes orchestration, Infrastructure as Code, and full-stack observability, with proven impact cutting MTTR by 25% through AIOps, chaos engineering, and automated incident response.

SKILLS

Cloud & Infrastructure: AWS (EC2, S3, EBS, EFS, Lambda, ECS, EKS, RDS, Auto Scaling, API Gateway, CloudFront, CloudTrail), Azure (VMs, VNet, Load Balancer, Monitor), GCP (Compute Engine, GKE, Cloud Run, Vertex AI, Cloud Storage, IAM, Cloud Monitoring), Nginx, DNS, SSL/TLS
CI/CD, IaC & Automation: Jenkins, GitHub Actions, GitLab CI/CD, ArgoCD, GitOps, Terraform, Ansible, CloudFormation, Docker, Kubernetes, Helm
Monitoring & Observability: Datadog, Dynatrace, Prometheus, Grafana, OpenTelemetry, ELK Stack, Splunk, CloudWatch, PagerDuty, Distributed Tracing
Reliability & Operations: Chaos Engineering (Gremlin, Chaos Mesh), Incident Management, SLO/SLI Design, Error Budgets, On-Call Rotation, Capacity Planning
Languages & Databases: Python, Bash, JavaScript, Java, SQL, PostgreSQL, MongoDB, Redis
AI/ML & Data: LLMs, RAG, LangChain, Transformers, Prompt Engineering, Vector Databases, MLOps, SageMaker, TensorFlow, ETL Pipelines
Tools & Platforms: Git, GitHub, GitLab, Jira, Confluence, Linux (Ubuntu, RHEL), PyTest, HashiCorp Vault, OPA, Agile/Scrum

PROFESSIONAL EXPERIENCE

Research Assistant – AI Operations Engineer (R&D)

July 2025 – Present

P&G Digital Accelerator @ University of Cincinnati, Cincinnati, OH

Part-time

- Designed and developed an **agentic LLM-based application** using **generative AI** and **prompt engineering** to extract, structure, and query knowledge from large volumes of unstructured text data using Python and Streamlit.
- Implemented **Retrieval-Augmented Generation (RAG)** pipelines with **vector databases** and LangChain, significantly improving response relevance and factual grounding across enterprise knowledge bases.
- Architected client-isolated deployments on **Microsoft Azure**, leveraging autoscaling, **immutable infrastructure**, and load balancing to ensure reliability under variable workloads.
- Developed an end-to-end **data processing and ETL pipeline in Databricks**, supporting scalable ingestion, transformation, and orchestration of unstructured data for downstream analytics and **MLOps** workflows.
- Designed a custom **authentication and usage-tracking wrapper**, enabling **zero-trust** secure user access, fine-grained IAM control, and per-user LLM cost attribution across multiple tenants.
- Built **automated performance and load testing** harnesses to validate system behavior under stress, identifying bottlenecks and ensuring SLO compliance before production rollouts.

Site Reliability Engineer

December 2022 – January 2025

UBA Solutions Pvt Ltd., Kathmandu, Nepal

Full-time

- Built and optimized a comprehensive **observability stack** using Datadog, ELK, Grafana, Splunk, and **OpenTelemetry distributed tracing**, improving alert accuracy by **40%** and reducing MTTR by **25%**.
- Architected and maintained **AWS multi-region failover** infrastructure with **immutable deployments**, supporting zero-downtime rollouts and improving production reliability across 3 regions.
- Maintained **99.95% uptime** (SLO/SLI compliance) for a high-traffic SaaS platform serving over 100K users per hour through proactive autoscaling, health checks, and automated recovery.
- Orchestrated containerized microservices on **Amazon EKS** with **Istio service mesh**, managing cluster scaling, networking policies, and **canary deployments** to support **50+ production workloads**.
- Implemented **agent-assisted root cause analysis (RCA)** and **blameless postmortem** workflows, leveraging **AIOps** and observability data to improve incident transparency and system resilience.
- Developed automated **runbooks and self-healing scripts** for **toil reduction** in Python and Bash, reducing manual intervention during on-call rotations by 35%. Received the **2023 Above and Beyond Award**.
- Defined and tracked **Service Level Indicators (SLIs)** and **error budgets** across distributed services, establishing data-driven reliability targets and automated alerting thresholds via Prometheus and Grafana.
- Led **change management processes** including creation and review of **Method of Procedures (MOPs)**, standard operating procedures (SOPs), and emergency operating procedures (EOPs) for mission-critical maintenance windows.
- Conducted **chaos engineering experiments** using Gremlin and custom fault-injection scripts to simulate network disruptions, hardware failures, and system overloads, validating failover mechanisms and improving system resilience.
- Managed **on-call rotations** and **incident escalation workflows** integrated with PagerDuty, ensuring rapid response to unplanned events and maintaining 24/7 operational coverage for mission-critical infrastructure.

- Supervised **infrastructure redundancy configurations** (N+1, 2N, ATS/STS failover scenarios) and performed regular **capacity planning** reviews to ensure headroom for traffic spikes and growth projections.

Associate Cloud & DevOps Engineer

August 2021 – December 2022

Cloudlaya LLC, Kathmandu, Nepal

Full-time

- Automated **CI/CD pipelines** for 50+ applications using Jenkins and AWS CodePipeline with **shift-left testing**, reducing deployment effort and turnaround time by over **80%**.
- Migrated **20+ services to AWS ECS Fargate** using Terraform and CloudFormation (**Infrastructure as Code**), lowering infrastructure costs by **30%** while improving reliability and scalability.
- Refactored legacy applications (Spring Boot, PHP, Node.js) applying **DevSecOps** best practices and contributed to new feature development using Django and Vue.js, reducing security risks by **40%**.
- Developed **automated test suites** integrated into CI/CD pipelines to validate infrastructure changes and application deployments, ensuring reliability with every release and reducing post-deployment incidents by **20%**.
- Created and maintained **Infrastructure as Code** modules in Terraform with reusable templates, enabling consistent provisioning of VPCs, security groups, load balancers, and auto-scaling groups across multiple environments.
- Implemented **branching strategies** and Git workflows (GitFlow) across development teams, standardizing source code management and improving collaboration for 15+ concurrent projects.
- Supported **compliance and accreditation initiatives** including SOC2 and security audits, documenting infrastructure configurations and access controls to meet organizational security standards.

Cloud & DevOps Engineer Intern

April 2020 – July 2021

Cloudlaya LLC, Kathmandu, Nepal

Internship

- Administered **100+ production web applications** on Linux-based hosting (WHM/cPanel), achieving **99.9% availability** through proactive monitoring, secure configurations, and rapid **incident response** with sub-15-minute MTTA.
- Configured and managed **DNS, SSL/TLS certificates, Zimbra mail servers**, and Linux-based hosting environments, supporting seamless client onboarding and **capacity planning**.

PROJECT EXPERIENCE

LLM-Droid-Tester: Contextual Automated Reproduction of Bugs on Android

December 2025 – Present | GitHub

- Built an **LLM-driven bug reproduction system** for Android that connects to an emulator, captures annotated screenshots with color-coded bounding boxes, and uses **Gemini 2.5 Pro** to iteratively follow reproduction steps from natural-language bug reports.
- Engineered support for **complex gestures** (pinch-to-zoom, double-tap, edge swipes, coordinate-based interactions) and **vision-grounded prompt engineering**, enabling reproduction of bugs that prior tools could not handle.

Kubernetes Cluster Resilience Testing Framework

March 2026 | GitHub

- Built a **chaos engineering toolkit** for Kubernetes clusters that simulates pod failures, network partitions, and resource exhaustion scenarios to validate application resilience and auto-recovery mechanisms.
- Integrated resilience tests into **CI/CD pipelines** with automated pass/fail gates based on SLO thresholds, ensuring deployments meet reliability standards before reaching production.

Datadog Synthetic Monitor Validator

October 2024 | GitHub

- Developed an automated **health-check validation and alerting tool** using the Datadog API with **anomaly detection** to cross-verify synthetic monitor results, reducing false-positive alerts and improving on-call efficiency.
- Implemented configurable thresholds, **error budget** tracking, and automated **incident triage logic**, integrating with PagerDuty and Slack for streamlined **alert routing** across distributed services.

Java Spring Boot CI/CD Pipeline with AWS EKS

April 2022 | GitHub

- Designed and implemented an end-to-end **CI/CD pipeline** for deploying Java Spring Boot microservices to AWS EKS using Jenkins, **SonarQube** (static analysis), Docker, Terraform, and Kubernetes.
- Automated **Infrastructure as Code** provisioning with Terraform and containerized deployments via Helm charts, enabling repeatable, zero-downtime **blue-green/canary deployments** with **container security scanning**.

CERTIFICATIONS & ACHIEVEMENTS

- AWS Certified Solutions Architect – Associate, Amazon Web Services
- AWS Community Builder, Cloud Operations Track
- 2023 Above and Beyond Award, UBA Solutions Pvt Ltd. (Recognized for exceptional contributions to SRE practices)
- Graduate Student Judge, CEAS Expo 2026, University of Cincinnati

EDUCATION

Master of Science in Computer Science

January 2025 – Present

University of Cincinnati, Cincinnati, OH

Relevant Coursework: Cloud Computing, Advanced Algorithms, Machine Learning, Artificial Intelligence, Data Analysis

Bachelor of Science in Computer Science and Information Technology

August 2017 – August 2021

Tribhuvan University, Nepal

Relevant Coursework: Operating Systems, Computer Networks, Database Management, Data Structures