

Dibya Darshan Khanal

Cincinnati, OH | +1 646-842-4641 | 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 of experience building and operating production-grade cloud infrastructure on AWS and Azure, maintaining 99.95%+ uptime for distributed systems serving 100K+ users. Proven track record in CI/CD automation, Kubernetes orchestration, Infrastructure as Code, and full-stack observability — with a focus on reducing MTTR through AIOps, chaos engineering, and intelligent incident response.

SKILLS

Cloud & Infrastructure:	AWS (EC2, S3, Lambda, ECS, EKS, RDS, Auto Scaling, SNS, Secrets Manager), Azure (VMs, VNet, Load Balancer, DevOps, Monitor), GCP (Compute Engine), Apache, Nginx, DNS, SSL/TLS
CI/CD, IaC & Automation:	Jenkins, Terraform, Ansible, CloudFormation, CodePipeline, GitHub Actions, GitLab CI/CD, ArgoCD, GitOps, Docker, Kubernetes, ECS, EKS, Docker Swarm, Helm, Nexus Repository
Monitoring & Observability:	Datadog, CloudWatch, CloudTrail, Splunk, ELK Stack, Grafana, Prometheus, OpenTelemetry, Nagios, Site24x7, PagerDuty, Distributed Tracing, SLO/SLI Monitoring
Reliability & Operations:	Chaos Engineering (Gremlin, Chaos Mesh), Incident Management, Change Management, Capacity Planning, SLO/SLI Design, Error Budgets, On-Call Rotation, DCIM, CMMS
Languages & Databases:	Python, Shell/Bash, JavaScript, Groovy, SQL, PostgreSQL, MongoDB, SQLite, Redis
AI/ML & Data:	LLMs, LangChain, RAG, Prompt Engineering, TensorFlow, scikit-learn, SageMaker, Databricks, ETL Pipelines, Vector Databases, MLOps, Anomaly Detection
Tools & Platforms:	Git, GitHub, GitLab, Bitbucket, Jira, Confluence, Linux (Ubuntu, CentOS, RHEL), PyTest, HashiCorp Vault, OPA, Visio, DISA STIGs, RMF Compliance, 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

April 2020 – December 2022

Cloudlaya LLC, San Francisco, CA

Full-time

- 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.
- 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%**.
- Configured and managed **DNS, SSL/TLS certificates, Zimbra mail servers**, and Linux-based hosting environments, supporting seamless client onboarding and **capacity planning**.
- 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.

PROJECT EXPERIENCE

CARBON: Contextual Automated Reproduction of Bugs on Android

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.

Java Spring Boot CI/CD Pipeline with AWS EKS

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**.

Datadog Synthetic Monitor Validator

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.

Kubernetes Cluster Resilience Testing Framework

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.

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

Jan 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

Aug 2017 – Aug 2021

Tribhuvan University, Nepal

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