

# NISHANK KOUL

+91 9873446506 koulnishank5@gmail.com nishankkoul nishank-koul nishank

## SUMMARY

**DevOps Engineer** with expertise in automating infrastructure, streamlining deployments, and delivering scalable, resilient cloud-native solutions.

## EDUCATION

**PES University - Bengaluru, India** Dec 2021 - May 2025  
*Bachelor of Technology: Computer Science*

## TECHNICAL SKILLS

**Languages:** Python, Bash, C++  
**Cloud Platforms:** Amazon Web Services (AWS), Google Cloud Platform (GCP)  
**CI/CD Tools:** Jenkins, GitHub Actions  
**Containerization:** Docker, Kubernetes  
**Monitoring:** Prometheus, Grafana, ELK Stack  
**Infrastructure as Code:** Terraform, Ansible

## EXPERIENCE

**Stringify AI** | *DevOps Engineer (Remote, USA)* Feb 2025 - Present

- Designed and containerized cloud-native applications using **multi-stage Docker builds**, and deployed them on **Google Cloud**, resulting in a **30% reduction in image size** and significantly improving application startup time and resource efficiency.
- Configured a **Global HTTP(S) Load Balancer** with **backend services as Network Endpoint Groups (NEG)s** and **CDN cache activated**, resulting in up to **60% reduction in latency** and **40% improvement in page load times** for global users.
- Provisioned a **production-grade PostgreSQL database on GCP Compute Engine** with SSH access restricted via a bastion host; implemented **SSH tunneling on PgAdmin4** for secure local access, improving database security posture by **50%** and reducing manual connection errors by **30%**.
- Streamlined CI/CD processes using GitHub Actions, **accelerating deployment cycles by 40%** while ensuring consistency, reliability, and faster time-to-market.
- Adopted **DevSecOps principles** by integrating **SonarQube (SAST)** and **Trivy** into CI/CD pipelines, reducing critical vulnerabilities by **40%** and enhancing secure software delivery.
- Activated Cloud Audit Logging in the production environment to ensure **comprehensive audit trails** for administrative activities and data access, enhancing security posture and enabling **100% compliance visibility** in Google Cloud.

**Bimaplan** | *DevOps Engineer Intern (On-site Bangalore, India)* Sep 2024 - Feb 2025

- Crafted Python scripts for AWS Lambda functions to automatically shut down EC2 instances in the Dev and UAT environments during non-business hours, leading to a **25% reduction in overall cloud costs** by optimizing resource utilization and minimizing idle time.
- Orchestrated **zero-touch deployment** by automating 90% of AWS infrastructure provisioning using Terraform; enabled Disaster Recovery through cross-region replication to ensure business continuity.
- Executed the setup of a **read replica for the RDS Database** to enhance availability and scalability, **improving read query performance by 40%** and reducing downtime risks.
- Refined Jenkins CI/CD pipelines** across Dev, UAT, and Prod by integrating Terraform, ensuring **100% consistency** in provisioning. **Established backup strategies** for pipeline code and statefiles, reducing rollback time by **60%**.
- Delivered an efficient **API Gateway rate-limiting strategy** based on traffic analysis to enhance performance and prevent abuse; integrated **CloudWatch alarms** with Slack for real-time alerting on HTTP 429 errors.

## PROJECTS

**Celestia Validator Node Deployment on Mocha-4 Testnet** | *Blockchain, Ansible, AWS EC2, Prometheus, Grafana* |

- Built an end-to-end Ansible playbook to **automate Celestia validator node** provisioning, **reducing manual setup time by 80%** and ensuring consistent deployments with zero configuration drift.
- Configured a **Grafana-based monitoring system with custom dashboards** to track node performance metrics, including block height, sync status, and resource utilization in real-time, enhancing operational visibility and reducing incident resolution time by **50%**.
- Developed industry-standard security protocols by applying encryption and access **restrictions for sensitive credentials using Ansible Vault** and designed rollback mechanisms, **reducing downtime risk by 30%** and improving validator resilience.

**Scalable LLM Inference Service with Ollama** | *LLMs, Flask, Docker, AWS EKS, K6.io, GitHub Actions* |

- Engineered a **scalable LLM inference service** with Ollama, integrating the moondream model. Developed a Dockerized Flask API and orchestrated the service on AWS EKS for **high availability and scalability**.
- Accelerated application performance by identifying and resolving memory allocation bottlenecks during **Load Testing with K6.io**, resulting in a **35% reduction in response time** and **50% improvement in container accessibility under peak load**.
- Executed auto-scaling strategies, increasing the successful request response rate from **53.66% to 85.49%**.

## CERTIFICATION

**GCP Certified Associate Cloud Engineer** Achieved in May 2025  
**AWS Certified Cloud Practitioner** Achieved in Oct 2024