

SHIVAM GUPTA

STAFF ENGINEER | DEVOPS | CLOUD PLATFORM | AI INFRASTRUCTURE

Bangalore, Karnataka | shivamgupta211@gmail.com | shivamgupta211.github.io

PROFESSIONAL SUMMARY

Staff Engineer with 10+ years of experience designing, building, and operating highly available cloud platforms across AWS, GCP, and Azure. Expertise in Kubernetes, EKS/GKE, CI/CD, Terraform, observability, networking, FinOps, and DevSecOps. Experience extending platform engineering into AI infrastructure, including AI Gateway architecture, LLM/API routing, Amazon Bedrock integration, centralized access controls, and AI service observability.

CORE SKILLS

Cloud & Platform: AWS, GCP, Azure, EKS, GKE, cloud architecture, high availability

AI Infrastructure: AI Gateway, LLM routing, Amazon Bedrock, model/API integration, AI observability

Kubernetes: Kubernetes, Helm, Kustomize, Jsonnet, Istio, Cilium/eBPF, NGINX, Traefik

Infrastructure as Code: Terraform, Atlantis, Ansible, AWX, Docker, Linux, Bash

CI/CD & GitOps: Jenkins, Argo CD, Spinnaker, GitHub Actions, CircleCI

Observability: Prometheus, Grafana Cloud, EFK, OpenTelemetry, Datadog, New Relic

Programming: Python, Go, Node.js

Data, Security & Cost: Redis, PostgreSQL, MySQL, MongoDB, RDS, DevSecOps, FinOps, Databricks

PROFESSIONAL EXPERIENCE

FalconX | Staff Engineer September 2021 – Present

Bengaluru, Karnataka

- Owned reliability, maintenance, and upgrades of the high-frequency trading platform, improving uptime from 99% to 99.9%.
- Designed and implemented AI infrastructure capabilities, including an AI Gateway providing centralized and observable access to LLM and AI services.
- Integrated Amazon Bedrock into the AI platform to enable managed foundation-model access with centralized routing, authentication, usage controls, and observability.
- Built in-house MCP server to explore devops tools, observability, etc using natural language
- Improved CoreDNS reliability by reducing DNS resolution errors to 0.01% through caching and prefetching.
- Designed and deployed a scalable IPsec VPN mesh across AWS, GCP, and Azure for secure cross-cloud connectivity.
- Reduced overall CI/CD time by 40% by optimizing CI pipelines and migrating continuous delivery from Jenkins to Argo CD.
- Integrated RunOn to run self-hosted github runners into our own infra and decreased build time by at least 20%
- Migrated multiple public EKS clusters to private clusters with no downtime and standardized Kubernetes operations.
- Consolidated EKS version upgrades into controlled annual upgrade windows instead of quarterly upgrade cycles.
- Established Ansible AWX to enable security teams to perform patching and upgrades at scale across multiple cloud accounts.
- Built multi-cluster, multi-cloud observability using Grafana Cloud.
- Attributed more than 90% of cloud costs to teams using FinOps and tag compliance at scale; Built CXO/team dashboards and reporting using Databricks; Integrated Databricks' Genie AI to get analytics around the cost data using natural language
- Integrated Atlantis with GitHub to automate Terraform infrastructure provisioning.
- Improved Kubernetes network performance by approximately 20% using Cilium and eBPF.
- Migrated proxies running on ec2 instance to EKS reduced overall cost by 70%

ShareChat & Moj | Sr. DevOps Engineer July 2020 – August 2021

Bengaluru, Karnataka

- Spearheaded running Redis locally within the cluster to reduce network latency.
- Introduced Terraform Cloud to automate infrastructure provisioning.
- Managed 150+ production GKE clusters including maintenance and upgrades.
- Scaled Spinnaker CD to support production scale and reduced overall CD time by 25%.
- Implemented Prometheus federation, recording rules, and remote write to Elasticsearch for long-term metric retention.

Rentomojo | Sr. Software Engineer (DevOps) June 2019 – July 2020

Bengaluru, Karnataka

- Deployed and self-managed multiple Kubernetes clusters using KOPS.
- Deployed Istio service mesh for routing and cluster observability.
- Reduced overall AWS cost by 30% through AWS resource optimization.
- Built and managed CI/CD pipelines on CircleCI.
- Integrated monitoring solutions for Kubernetes clusters and databases.

GoVocal.AI | Sr. Software Engineer July 2017 – May 2019

Gurugram, Haryana

- Handled the complete backend web platform.
- Managed CI/CD pipelines for Azure and GCP.
- Managed Docker Swarm clusters hosted on Microsoft Azure.
- Managed microservices hosted on Google Cloud App Engine.
- Led a team of 2 experienced software developers.
- Technologies: Node.js, Python, LoopBack, MongoDB, Redis, Docker, Docker Swarm, Azure, and GCP.

AreteCode Tech. Pvt. Ltd. | Full Stack Web Developer July 2016 – June 2017

Noida, U.P.

- Worked with HTML, CSS, JavaScript, jQuery, AJAX, UnderscoreJS, REST APIs, ReactJS, NodeJS, and MySQL.

EDUCATION

B.Tech, 2016 | Harcourt Butler Technological Institute, Kanpur