

# ASUTOSH PANDA

📍 Brahmapur, India • ✉ [asutosh.pda@gmail.com](mailto:asutosh.pda@gmail.com) •  <https://linkedin.com/in/75asu> •  <https://75asu.pages.dev> •  <https://github.com/75asu>

## SKILLS

**GPU & AI Infrastructure:** NVIDIA GPU Operator, DCGM, drivers/CUDA, MIG, vLLM, KubeRay/Ray Serve, LiteLLM, Vertex AI, GPU SLOs  
**Infrastructure & Networking:** Kubernetes (GKE, EKS, bare-metal), Helm, Kustomize, FluxCD, ArgoCD, Terraform/OpenTofu, Ansible, Docker/Compose, Traefik, AWS ALB/NLB, Linkerd, GCP, AWS, Cloudflare, Hetzner, OVH  
**Secrets and Auth:** Workload Identity Federation, IRSA + EKS Pod Identity, External Secrets Operator (ESO), GCP + AWS Secret Manager, cert-manager, Pomerium, Cloudflare Zero Trust (Tunnel + Access)  
**CI/CD and Observability:** GitHub Actions, self-hosted ARC runners, Trivy, SonarQube, TruffleHog, Renovate; Prometheus, Grafana, Mimir, Loki, Tempo, Grafana Alloy, grafana-operator, Alertmanager  
**Development & Data:** Golang, Python, FastAPI, Bash, Node.js, TypeScript, PostgreSQL, AWS Aurora, Redis, GCP Pub/Sub, Linux

## PROFESSIONAL EXPERIENCE

### ONE2N

Pune, Maharashtra, India

#### Senior SRE & Team Lead · Client: Fravity AI (Remote)

May 2026 - Present

- Stood up the model-inference platform on the org's new **AWS EKS GPU** cluster via **Flux** GitOps: **KubeRay** for **vLLM** serving, an HA **LiteLLM** gateway on Aurora + Redis (secrets via **ESO**), and a **DCGM** + **Alloy** observability spoke to the central **LGTM**.
- Built **fravity-spine**, the production multi-tenant GKE platform for an AI product: private control plane, CMEK etcd, Workload Identity Federation, OpenTofu across **6** state layers, FluxCD add-ons, Pomerium SSO, and identity-based **kubectrl** (Connect Gateway) that retired bastions and shared SSH.
- Led multi-cloud governance and cost visibility across **5** GCP + **5** AWS accounts with the **2-engineer** team I lead: audited posture (**19** service-account keys, ~\$**26,520/yr** zombie spend), shipped an org-wide **FinOps** board (per-account cloud, SaaS + AI spend in Grafana), migrated all human + **10** workload identities to Terraform IAM-as-code, and rebuilt CI on keyless Workload Identity Federation.
- Delivered the central **LGTM** observability platform: Mimir + Loki + Tempo (GCS-backed, multi-tenant), Grafana unified alerting to Slack, and dashboards-as-code. A single **grafana/alloy** agent unifies the GKE spine/spokes and the first VM into one label-differentiated view (host/container/log).
- Stood up a **5-tool** SAST stack on self-hosted ARC runners and shipped **scan-aggregator** (normalises **7** scanners to a per-repo audit); delivered a customer security review surfacing **76** H/C findings across **4** images.
- Delivered staging as a commit-to-URL GitOps pipeline (Flux image-automation, live TLS, Slack alerts), **keyless** via per-app Workload Identity; onboarded the full app tier including the **workflow-engine** with attribute-filtered PubSub isolation, and locked the database to **automation-only** writes via **Atlas** migrations.
- Shipped the first client MVP on a hardened, fully **IaC** GCE VM (no public IP, **IAP**-only SSH, least-priv SA) behind a **Cloudflare** tunnel, with a keyless **WIF** build-to-deploy **Ansible** pipeline, the org's first reusable VM-deploy blueprint in an all-Kubernetes estate.

### LUXOR TECHNOLOGY

Seattle, WA, US

#### DevOps Engineer (Remote)

June 2023 - April 2026

- Operated a multi-cluster Kubernetes platform for **20+** teams and **3,000+** enterprise customers: **8+** clusters across Hetzner, OVH, and GCP with **24/7** uptime for L1/L2 validator and RPC workloads, including **2** Hetzner bare-metal HA clusters (KubeOne, OpenEBS) built from scratch. Standardised on a 7-layer FluxCD GitOps pipeline with **Linkerd** mTLS, migrating **70+** secrets to Workload Identity.
- Owned the on-call and incident platform: migrated **5** teams from Opsgenie to BetterStack, led response through real outages (a Litecoin hard fork with a **zero-downtime** deploy, a kubelet TLS expiry), and shipped x509 monitoring across **10** clusters.
- Owned CI/CD for **50+** repositories on self-hosted ARC runners. Ported a failing Playwright E2E suite from Mac hosted runners (**\$5,000/mo**) to Linux ARC, cutting runtime **40 to 15 min** and cost to **\$600/mo**. Built code-signed Windows + Mac release pipelines and automated FluxCD deployments to GCP Artifact Registry.
- Built org-wide platform tooling: **KubeOrbit** (Go CLI for kubeconfig provisioning across **20+** clusters) and **Memphis** (Python SOC1/2/SOX compliance binary, solo in **4 weeks** for the IPO). Vendored Trivy into a composite action across **9** repos, sidestepping the later CVE-2026-33634 in aquasecurity/trivy-action.
- Proposed and built **BVT**, the org's first zero-hardcoded-secrets cluster: replaced SOPS-in-git with ESO + GCP Secret Manager + Workload Identity Federation; the pattern became the gold standard for new clusters. Built custom Python tooling against the Hetzner + OVH firewall APIs for secure cross-cloud bare-metal peering where no Terraform modules existed.

### TATA CONSULTANCY SERVICES





Hyderabad, Telangana, India

#### Systems Engineer (Remote)

October 2020 - June 2023

- Migrated **2,500+** repositories (GitLab, Bitbucket, SVN) to GitHub Enterprise for **120** teams and built their CI/CD pipelines, resolving self-hosted-API pagination limits across the platforms and saving **15 weeks** of manual effort.
- Built PowerShell + Python automation (server monitoring, alerting, log rotation, Oracle DB replication, MS SQL updates) across **22** RHEL/WebLogic servers and **3** projects, saving **380+ man-hours/month**.
- Shipped a **Golang** CLI that cut HP storage provisioning from **30 man-hours** across **30** monthly tickets to a single command, adopted by the storage team as their primary provisioning interface.

## PROJECTS

- **Frit** (May 2026 - Present)  - GPU SRE lab: vLLM on NVIDIA GPU Operator; OSS PRs to **dcgm-exporter** + **gpu-operator**.
- **Platform Zero** (Apr 2026 - Present)  - **4** SRE labs (k8slab, ansiblelab, awslab, gcplab); **0-to-100** in one command.
- **Truss** (Jun 2026 - Present)  - Open-source self-hostable **Supabase/Appwrite** alternative, with an AI-agent **MCP server**.
- **ducat** (Jun 2026 - Present)  - Open-source **Prometheus** cost exporter: cloud, SaaS + AI spend in one Grafana view (**FOCUS** spec).

## TEACHING EXPERIENCE

- Full-Stack Instructor, **4** cohorts (2022-2026), taught NextJS, FastAPI, and DevOps to BSc CS students.

## EDUCATION

SASTRA DEEMED UNIVERSITY

Master of Computer Applications (MCA) - Chennai, India - 2021 - 2023