

Akshat Sharma

☎ +91-7014721442 | ✉ workakshat22@gmail.com | 💻 sharmaakshat37 | 🎧 Screach37

WORK EXPERIENCE

AppPerfect

DevOps Engineer

Apr 2025 – Present

Udaipur

- Built and operated large-scale **Linux GPU compute infrastructure** and **Kubernetes clusters** across **AWS, GCP, OCI**, and other CSPs, supporting production AI training and inference workloads on **NVIDIA H100, B200, B300, GB200, and GB300** platforms.
- Provisioned and automated Kubernetes and GPU infrastructure using **Terraform, Packer, Ansible**, and **cloud-init**, standardizing machine images, cluster bootstrap, and node lifecycle management across multi-cloud environments.
- Deployed and administered production **Slurm** clusters using **AWS ParallelCluster, Slinky**, and **Soperator**, integrating compute, networking, storage, scheduling, and workload orchestration for large-scale AI/HPC platforms.
- Configured and validated **RDMA-enabled, SR-IOV**, and **multi-NIC** GPU environments, troubleshooting high-performance networking and GPU communication using **NCCL, NVBandwidth**, and distributed workload validation.
- Worked with the **NVIDIA GPU Operator** ecosystem and collaborated on validation and troubleshooting of the **Kubernetes Dynamic Resource Allocation (DRA)** driver, investigating GPU resource allocation, **IMEX, ResourceClaims**, stale allocations, and multi-node scheduling behavior.
- Validated AI infrastructure using **NCCL, IOR**, checkpointing, and **LLM inference benchmarking** with **TRT-LLM** and **SGLang**, identifying GPU, storage, and network bottlenecks across distributed GPU clusters.
- Integrated and optimized high-performance storage solutions including **Lustre, EFS, Filestore**, and **Parallelstore** for distributed AI and HPC workloads.
- Developed reusable **Kubernetes operational tooling** using privileged DaemonSets for cluster diagnostics, host-level troubleshooting, and GPU/RDMA node validation across production environments.
- Implemented secure platform access using **Teleport, Tailscale**, external **IdP** integration, and Kubernetes **RBAC**, strengthening operational security and multi-tenant access management.
- Developed automated incident remediation workflows integrated with **Prometheus, Grafana**, and **PagerDuty**, reducing **Mean Time to Recovery (MTTR) by 65%** through proactive detection and automated recovery of production infrastructure issues.
- Provided **24x7 production support**, Root Cause Analysis (RCA), and platform engineering for large-scale AI infrastructure while contributing to architecture discussions and mentoring junior engineers.

Zelta Tech

Associate DevOps Engineer

Apr 2024 – Mar 2025

Remote

- Administered a self-hosted **GitLab** platform with **Git LFS** backed by **GCP Cloud Storage**, while developing and maintaining **GitLab CI/CD** pipelines for automated build, testing, and deployment workflows.
- Deployed and maintained production workloads on **AWS** and **GCP** using **Kubernetes, Docker**, and **Helm**, ensuring scalable and reliable application delivery.
- Engineered a multi-user **Zero-to-JupyterHub** platform on Kubernetes using **NativeAuthenticator** and **DockerSpawner**, providing isolated notebook environments for concurrent users.
- Deployed and managed **Apache Kafka** clusters on **Google Kubernetes Engine (GKE)**, configuring Kubernetes workloads, persistent storage, and networking for reliable messaging services.
- Automated Linux administration and operational workflows using **Bash** scripting, improving deployment consistency and reducing repetitive manual tasks.
- Implemented centralized observability using the **EFK Stack, Prometheus**, and **Grafana** for infrastructure and application monitoring.

Celebal Technologies

DevOps Intern

Jun 2023 – Aug 2023

Jaipur

- Provisioned and configured **Azure** infrastructure components including Virtual Machines, Load Balancers, and Application Gateway while supporting Git-based deployment workflows and Linux server administration.

PROJECTS

EstateGuard - DevOps | GCP, Docker, GitLab CI/CD, NGINX

- Built and deployed containerized application infrastructure on Google Cloud Platform using Docker and Compute Engine virtual machines.
- Designed GitLab CI/CD pipelines to automate application build, deployment, and release workflows.
- Configured NGINX reverse proxy, SSL/TLS, DNS, IAM, and Cloud Storage integrations to deliver secure production-ready infrastructure.

EDUCATION

SKIT, Jaipur	2020 – 2024
Bachelor of Technology in Computer Science	CGPA – 9.23
• Coordinator – Tech Innova Club • Linux Training Certification – Spoken Tutorial, IIT Bombay	

SKILLS

- Cloud Platforms:** AWS, GCP, OCI, Azure, Nebius
- Containerization & Orchestration:** Kubernetes, Docker, Helm
- Infrastructure as Code:** Terraform, Packer, Ansible
- CI/CD & GitOps:** GitLab CI/CD, Git
- HPC Technologies:** Slurm, AWS ParallelCluster, Slinky, Soperator, Lustre, EFS, Filestore, Parallelstore, NFSoRDMA, GDS, RDMA, RoCE, InfiniBand
- Performance & Benchmarking:** NCCL, IOR, NVBandwidth, Inference Max
- Observability:** Prometheus, Grafana, DCGM Exporter, Elasticsearch, Fluentd, Kibana, PagerDuty
- Identity & Access:** Teleport, RBAC, Tailscale, SSO, IdP Integration
- Programming & Scripting:** Python, Bash, Shell Scripting