← Devops Engineer Assignment

Assignment

Assignment: Kubernetes MLOps Infrastructure Implementation Task

Problem Statement: Design and implement a mini multi-tenant ML platform on AWS (or mocked locally), using infrastructure-as-code. You must deploy two isolated tenant environments running a simple ML inference API and demonstrate automation, security, and monitoring.

You are allowed to use any open-source tools (Terraform modules, Helm charts, Prometheus operator, Karpenter, MLflow, FastAPI model servers, etc.)

What You Need to Deliver

1. Infrastructure (Local Only)

Using Terraform or manifests:

- Deploy local KIND/Minikube cluster resources
- Create two tenant namespaces
- Add RBAC + NetworkPolicies
- Deploy MinIO as S3 alternative
- Use Kubernetes secrets or Vault

2. ML Inference Service (Containerized)

Deploy a small containerized model server (choose one):

- FastAPI + a small sklearn model
- MLflow model server
- Hugging Face small model (CPU-only allowed)

Devops Engineer Assignment

- tenant-a
- tenant-b

Each tenant must have:

- Its own Deployment + Service
- Tenant-based isolation via RBAC or NetworkPolicy

3. CI/CD Pipeline

Create a GitHub Actions pipeline that:

- Builds Docker image
- Runs a basic open-source security scan (Trivy / Snyk OSS)
- Pushes image to registry (Docker Hub or ECR)
- Deploys to EKS using kubectl/helm

4. Autoscaling (Real GPU not required.)

Implement one of the following:

- Kubernetes HPA (CPU-based)
- Document GPU autoscaling using Karpenter / Cluster Autoscaler

5. Monitoring & Alerting

Use any open-source stack:

- Deploy Prometheus + Grafana

Devops Engineer Assignment

- 1 dashboard (basic request/latency metrics)
- 1 alert rule (high latency or pod restarts)

Deliverables:

- README File
- Architecture diagram (simple PNG or draw.io)
- Short Explanatory Video
- Notes on Mocked vs Real Services

Note: This assignment must be completed using a local Kubernetes setup (KIND/Minikube) and open-source tools only. You do not need to use any cloud services or paid platforms. However, if you prefer to use your own AWS account or any other cloud provider, you are free to do so.