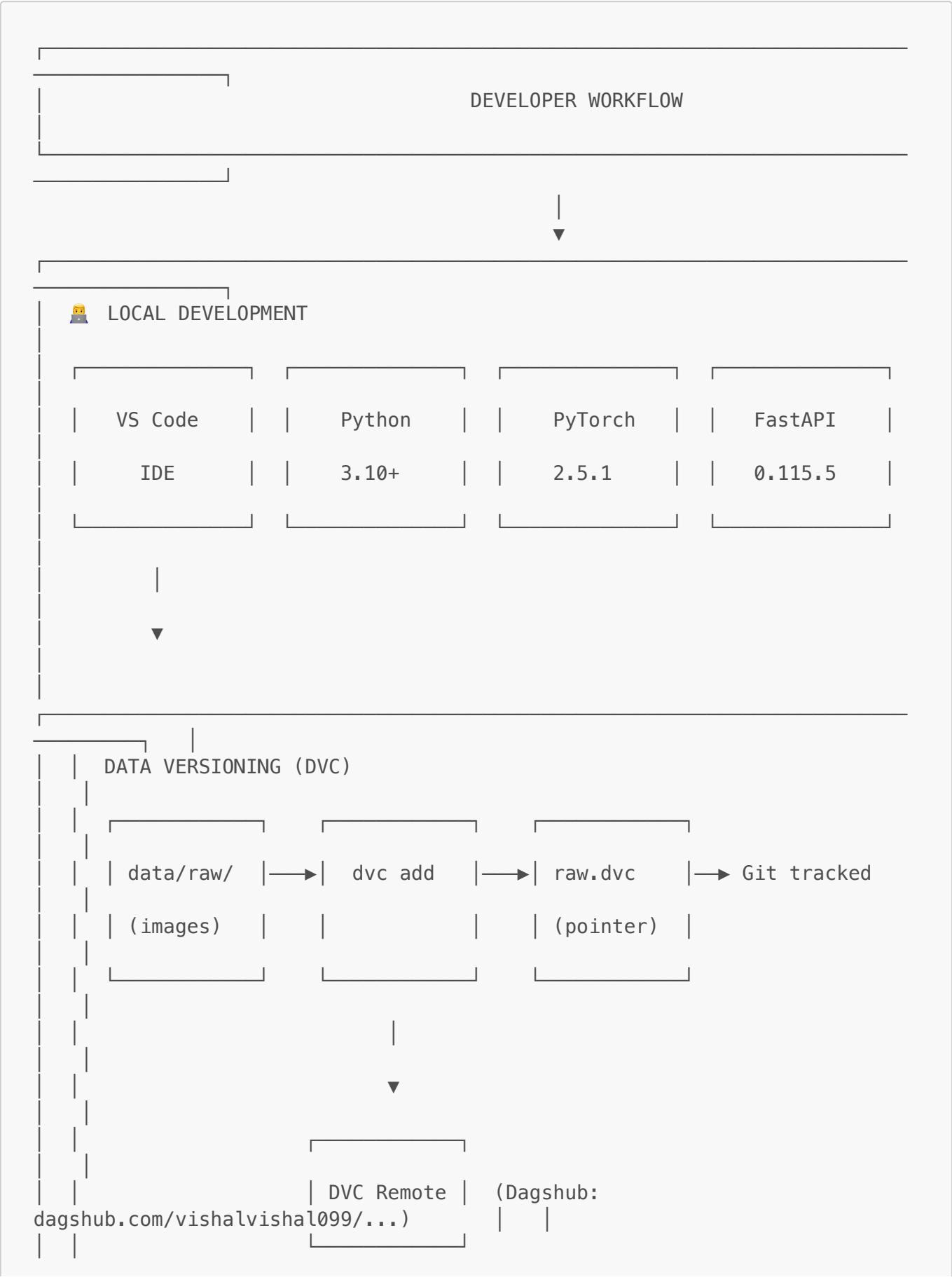
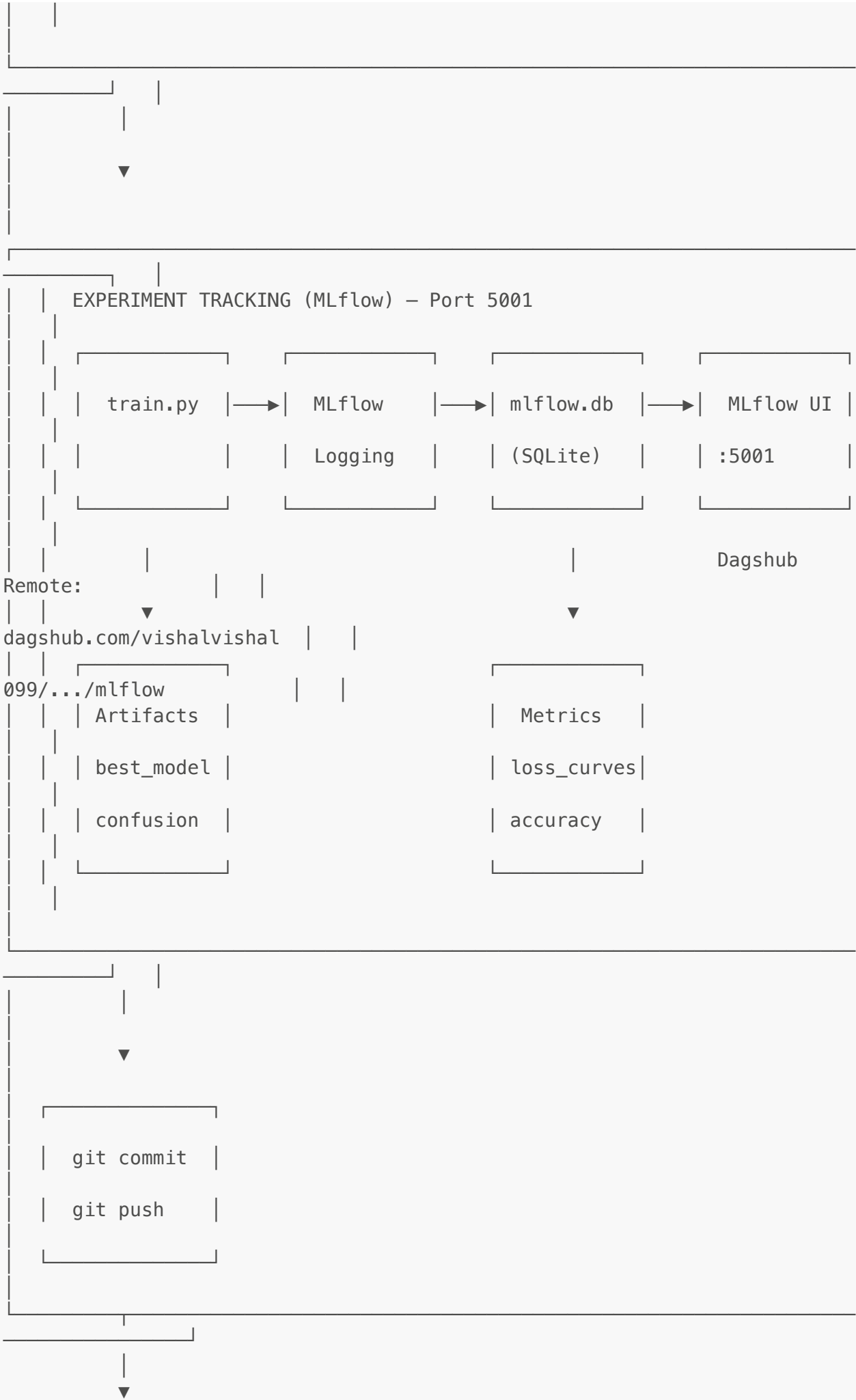
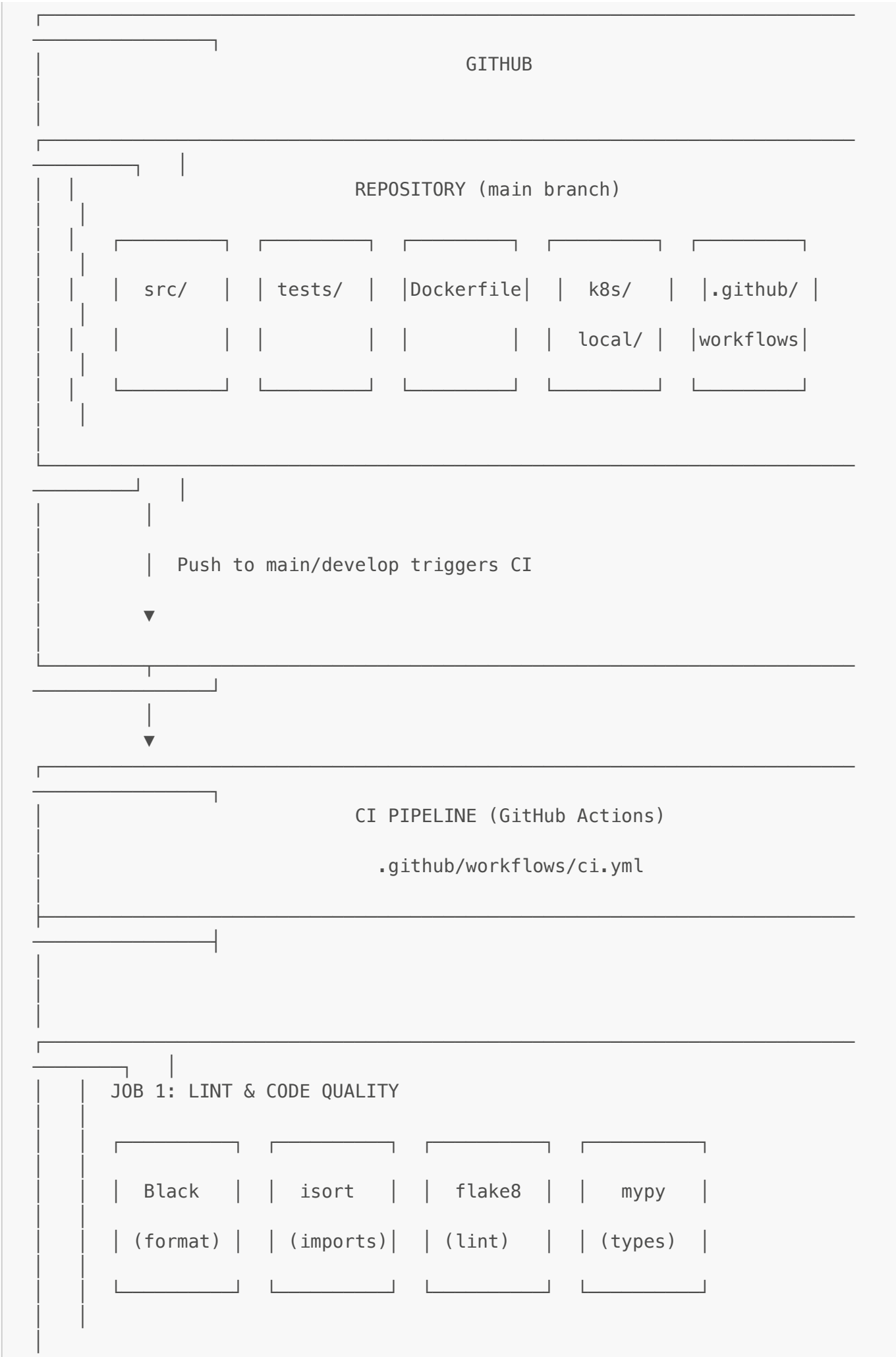


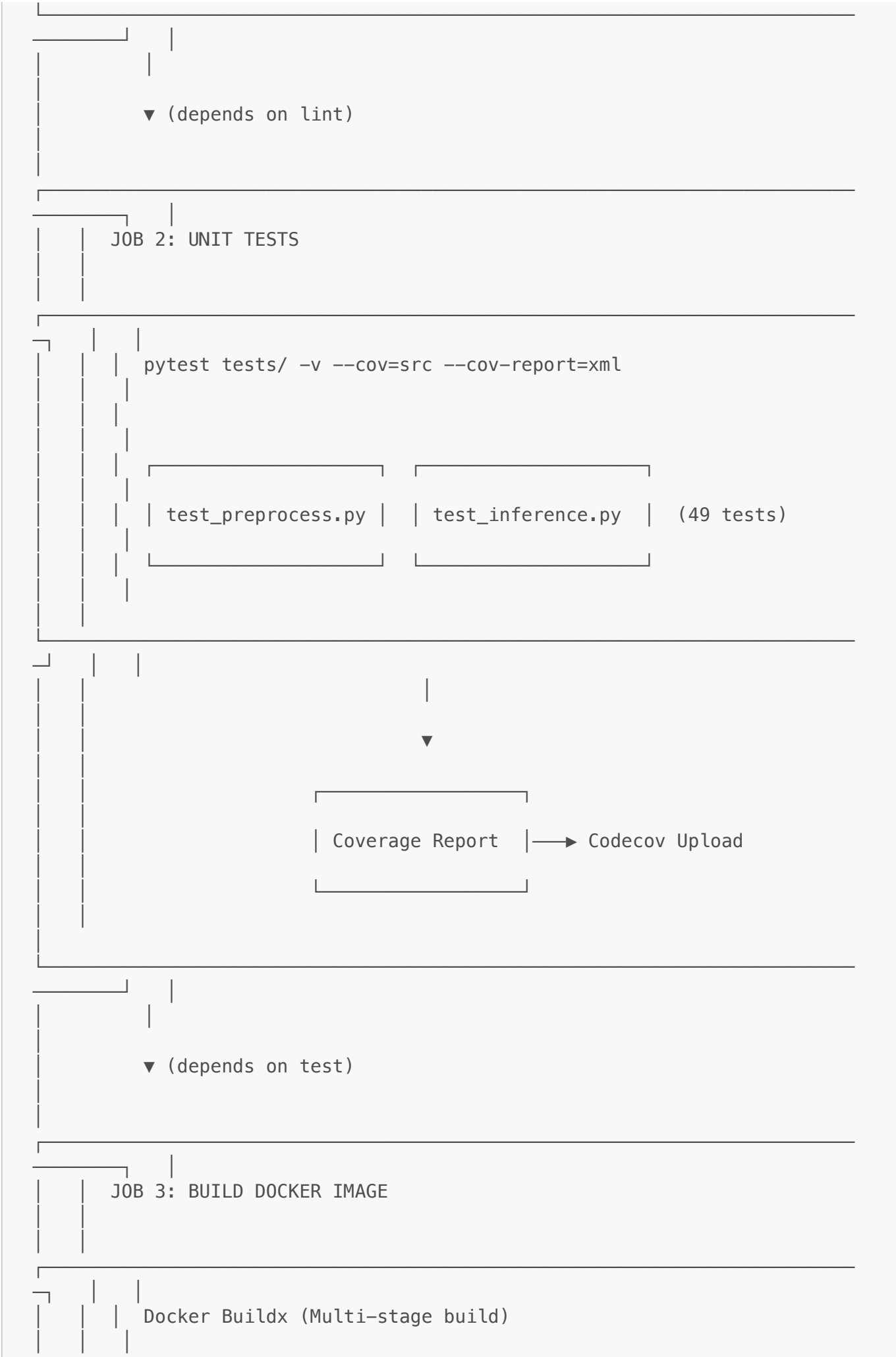
CI/CD Architecture Diagram

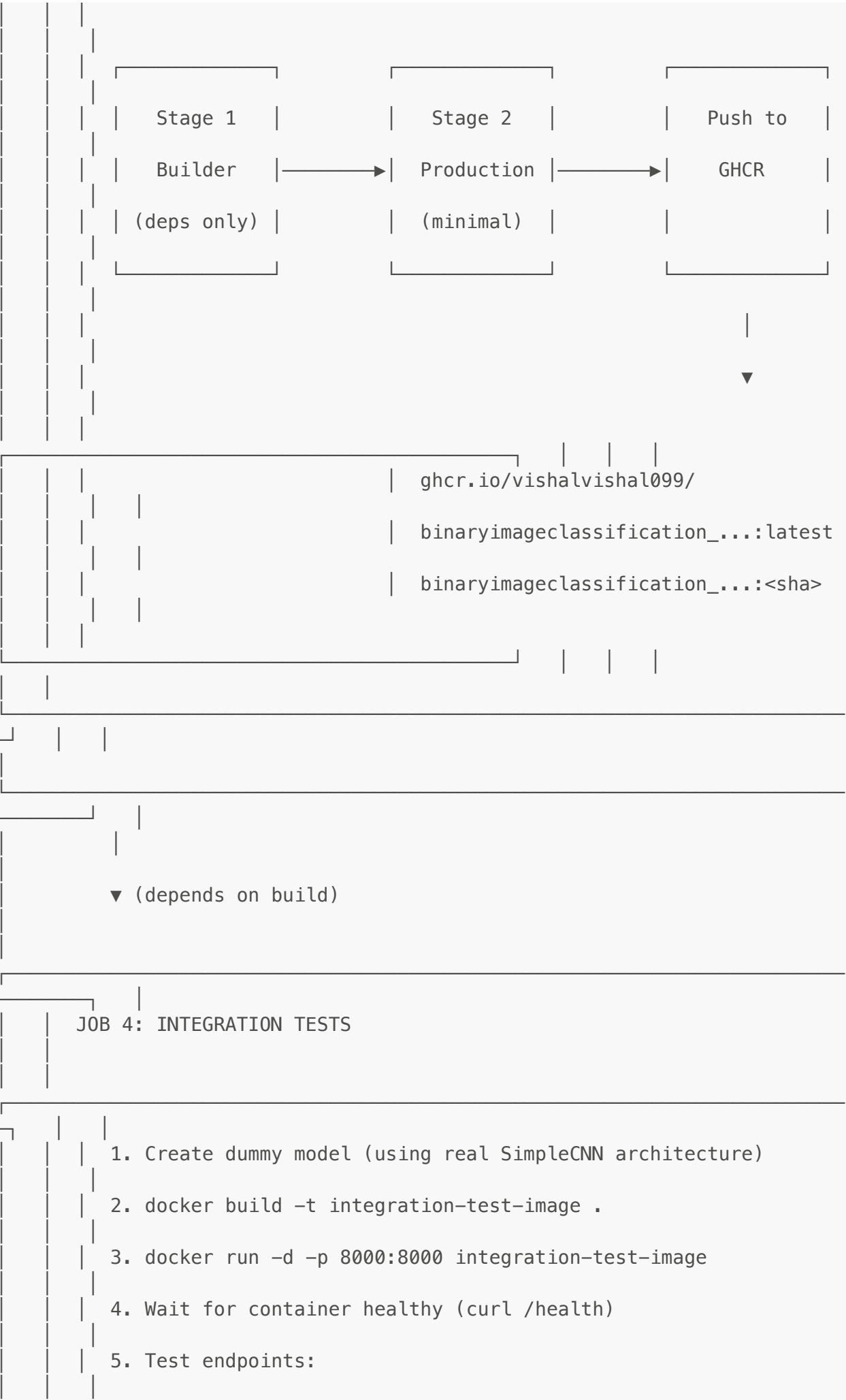
Complete MLOps Pipeline Architecture

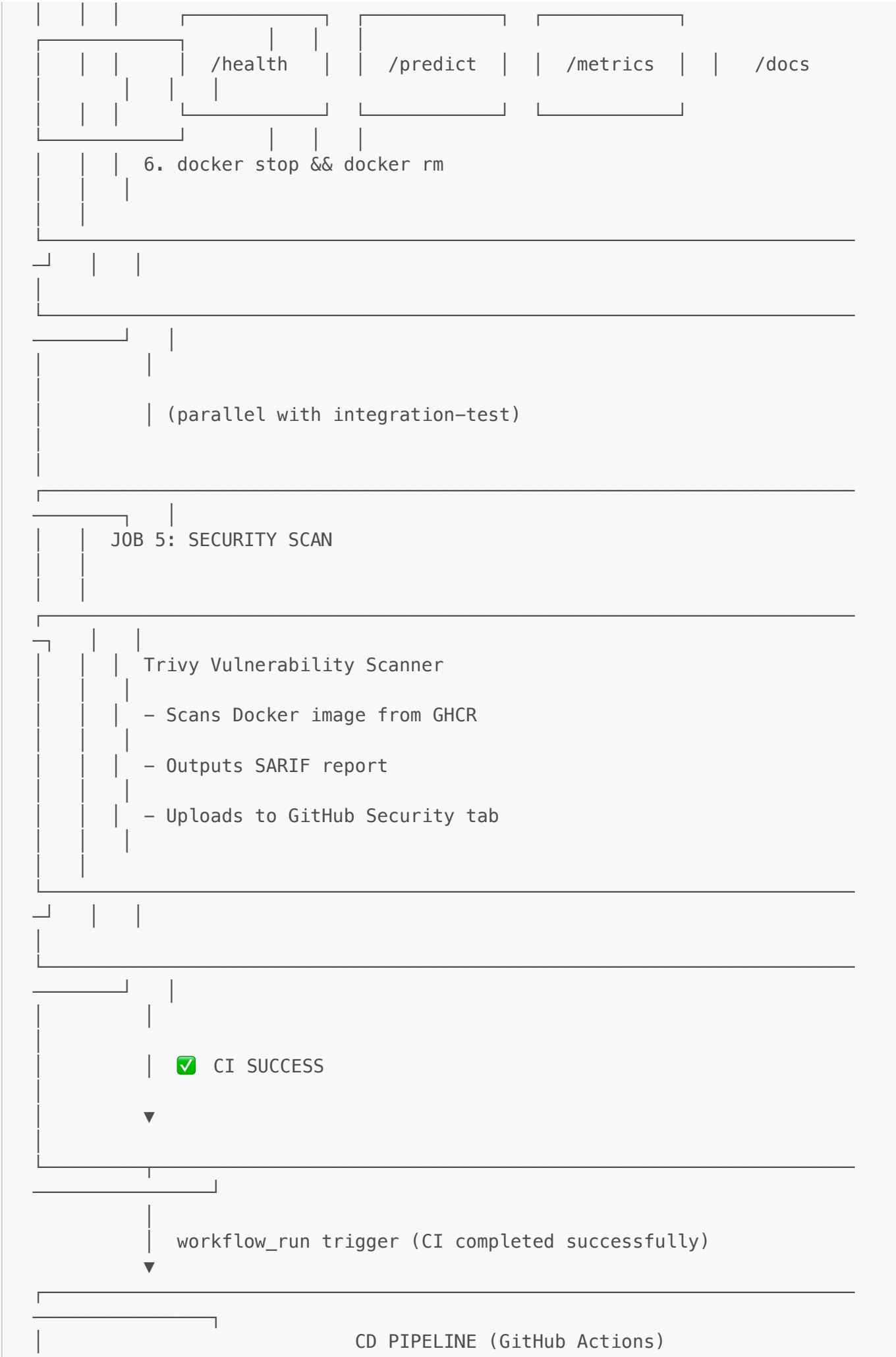


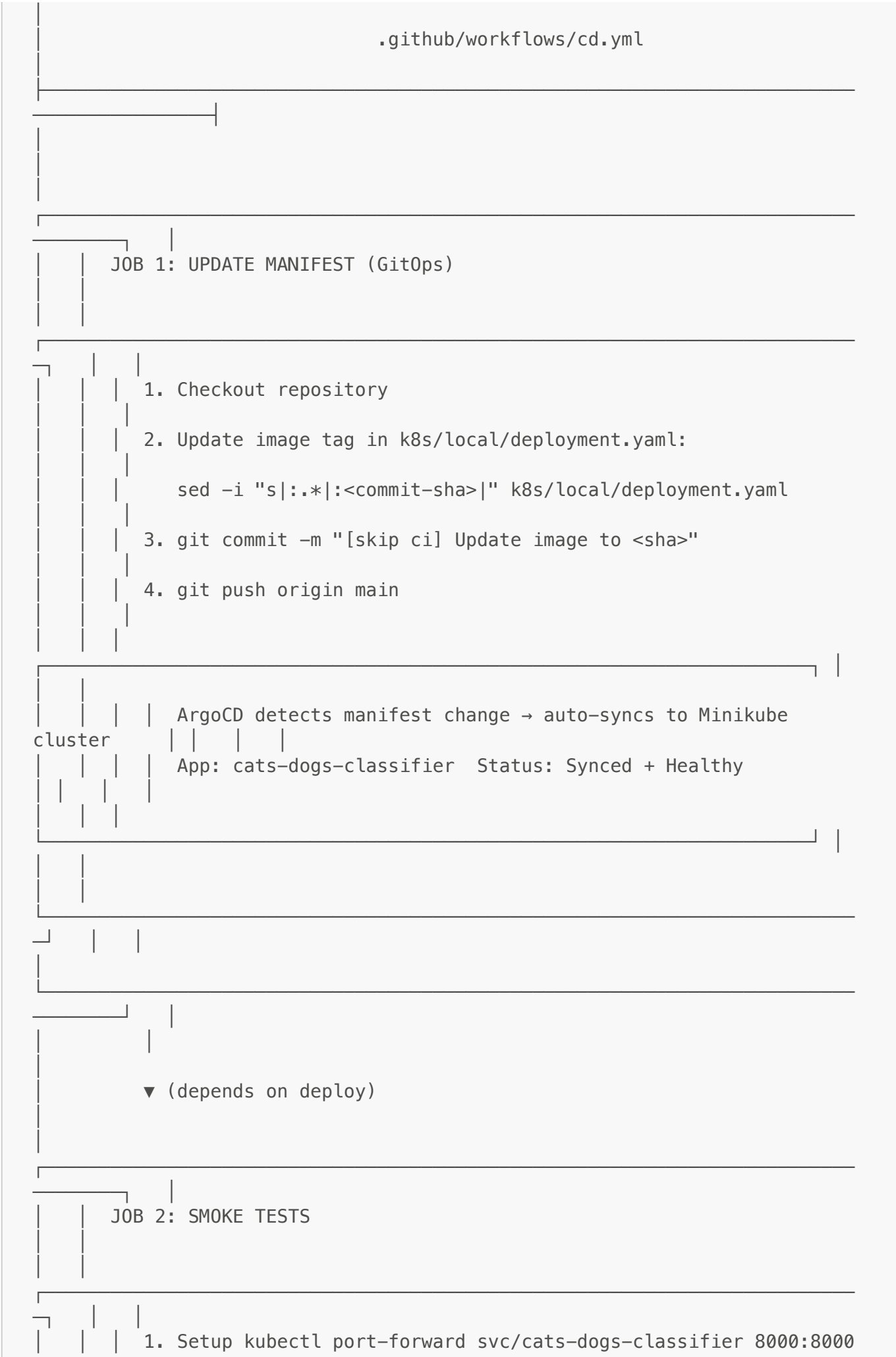


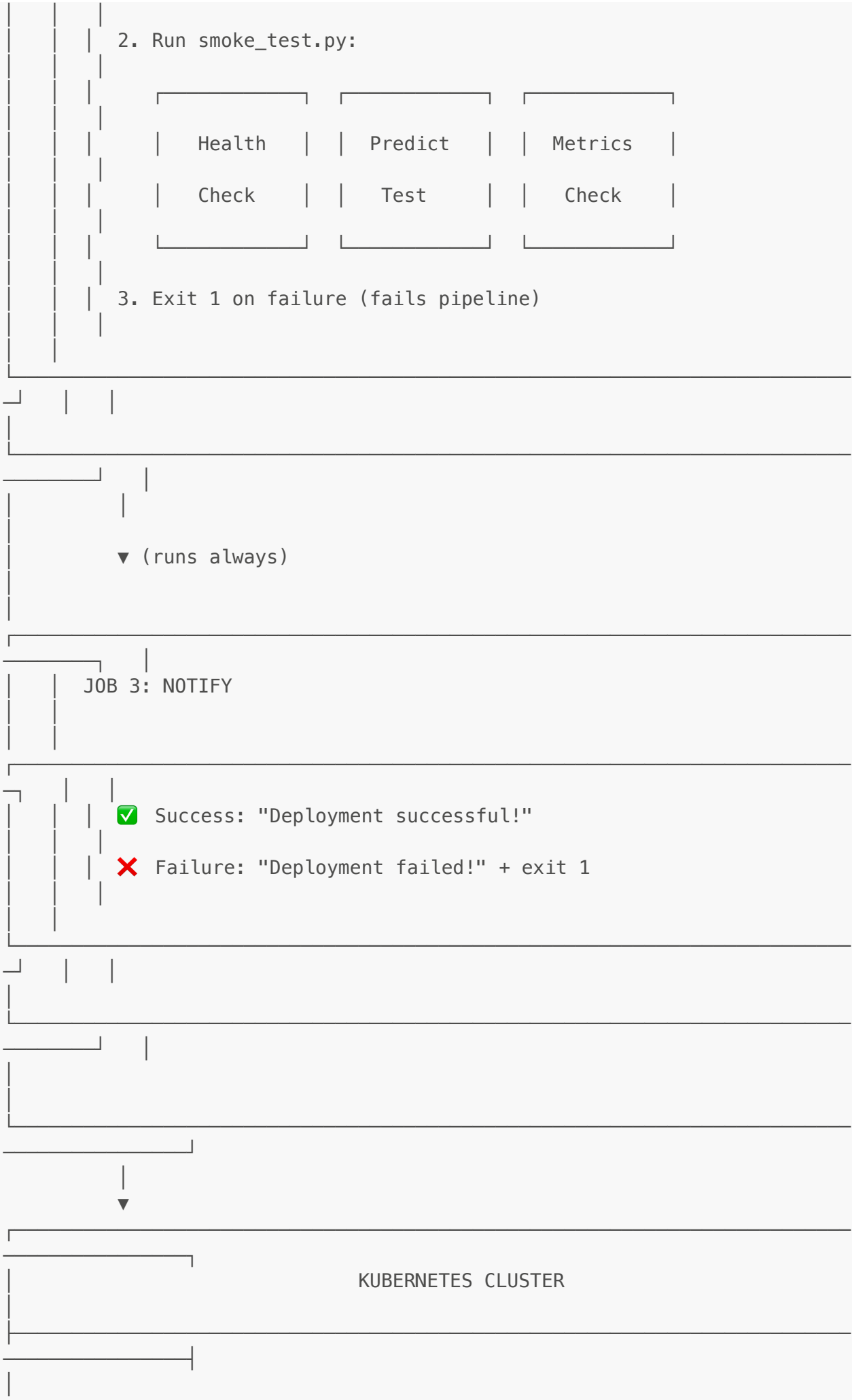












NAMESPACE: cats-dogs-classifier

DEPLOYMENT: cats-dogs-classifier

POD

CONTAINER: cats-dogs-api

Image: ghcr.io/.../binaryimageclassification:\$SHA

Port: 8000

FastAPI Application

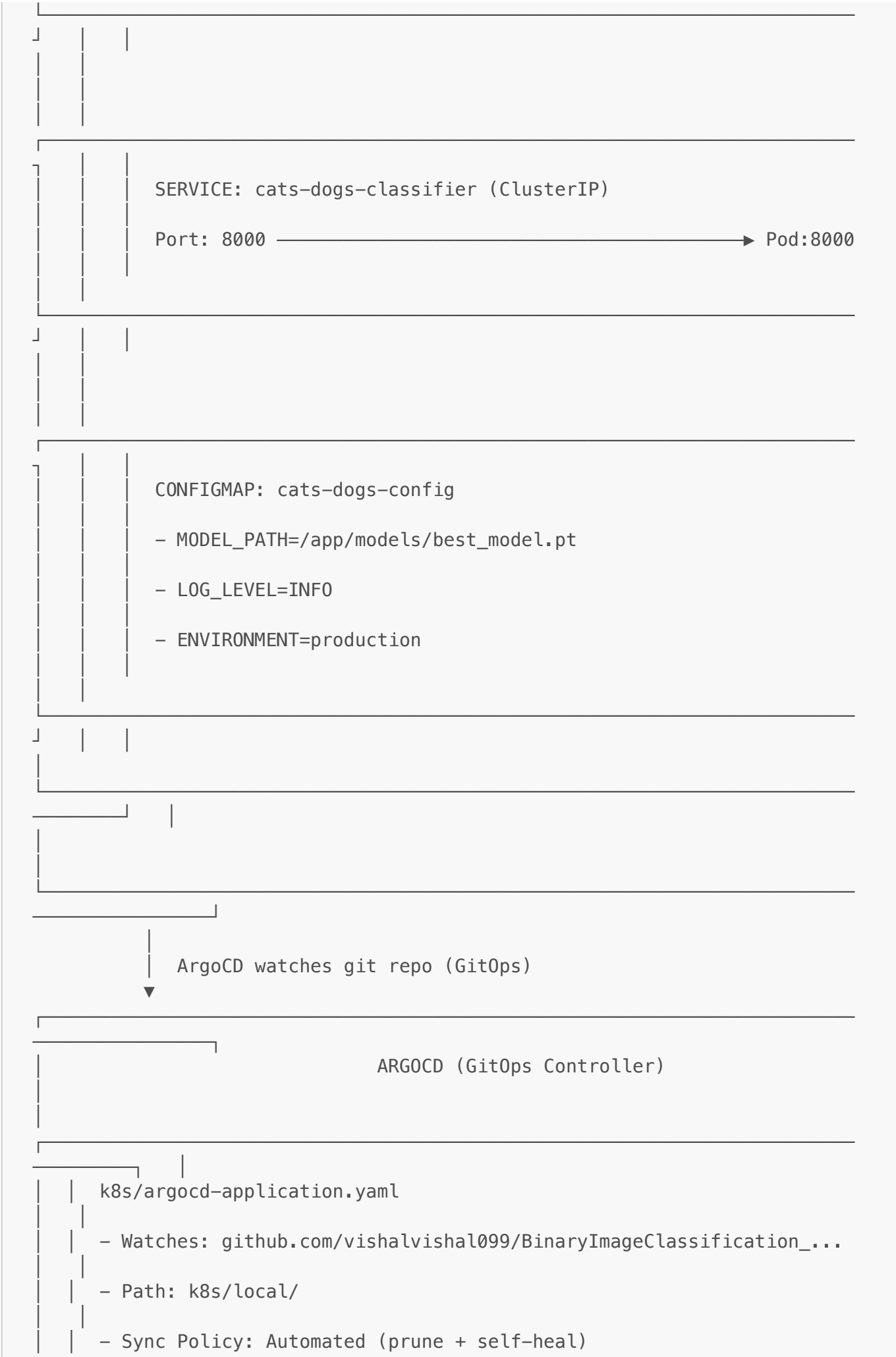
/health

/predict

/metrics

PyTorch Model (SimpleCNN)

/app/models/best_model.pt



- On git change → auto-sync to cluster

MONITORING & OBSERVABILITY

PROMETHEUS (metrics collection) – Port 9090

monitoring/prometheus.yml

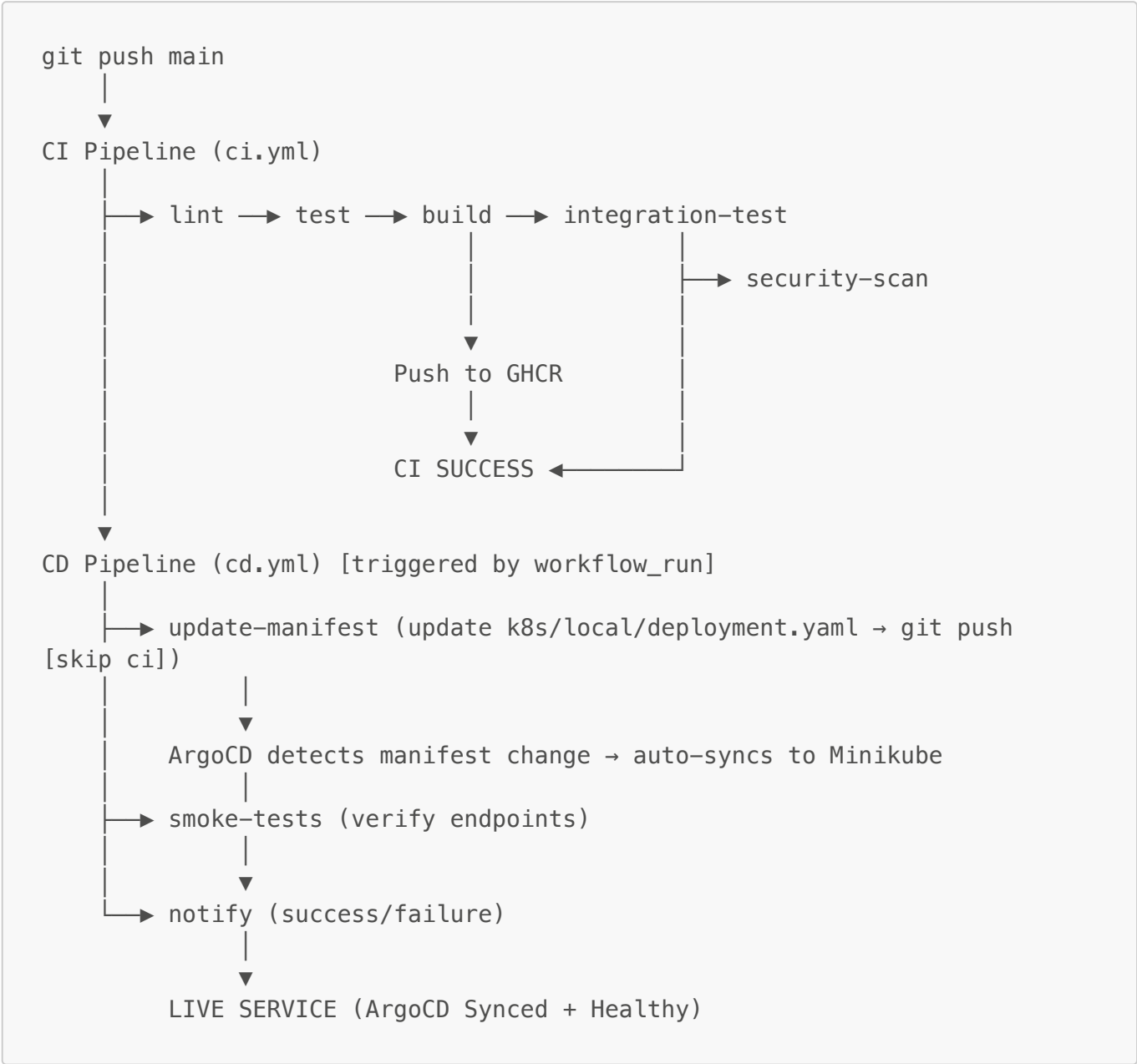
server)

- Scrapes: host.containers.internal:8081/metrics (metrics
- Scrapes: host.containers.internal:8000/metrics (FastAPI)
- URL: http://localhost:9090/graph
- Collects 31 metric families / 60+ time series:
 - api_requests_total, latency percentiles (p50/p90/p95/p99)
 - prediction_count, model accuracy/precision/recall/F1
 - error_count, system CPU/memory, batch inference



Flow Summary

Trigger Chain

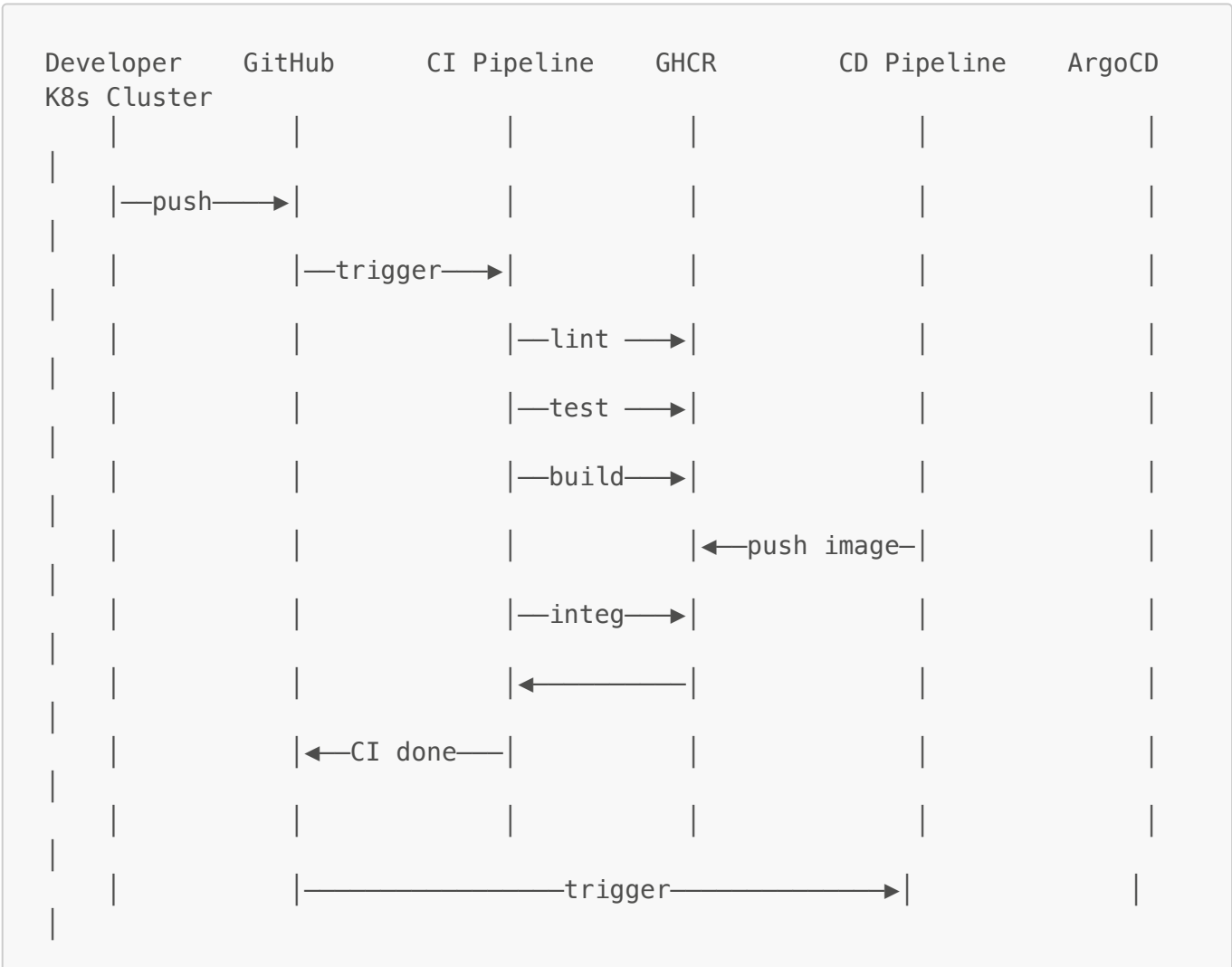


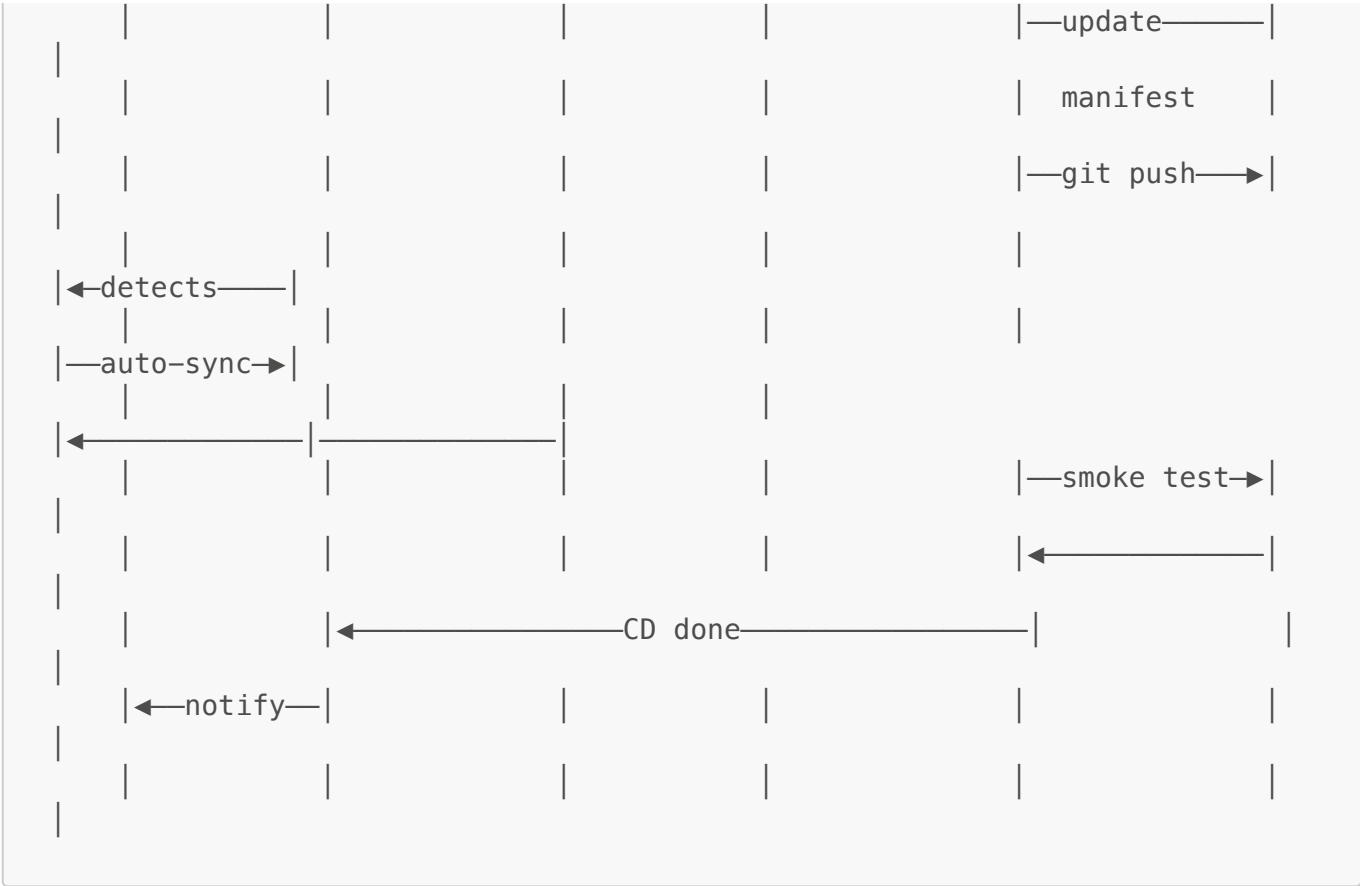
Tools & Technologies

Layer	Tool	Purpose
Version Control	Git + GitHub	Source code management
Data Versioning	DVC + Dagshub	Dataset versioning & cloud storage
Experiment Tracking	MLflow	Metrics, params, artifacts logging
CI/CD	GitHub Actions	Automated pipelines

Layer	Tool	Purpose
Containerization	Docker	Image building
Registry	GHCR	Container image storage
Orchestration	Kubernetes	Container orchestration
GitOps	ArgoCD	Declarative K8s deployment
API Framework	FastAPI	REST API service
ML Framework	PyTorch	Model training & inference
Monitoring	Prometheus	Metrics collection
Visualization	Grafana	Dashboards
Logging	structlog	Structured JSON logging
Testing	pytest	Unit & integration tests
Security	Trivy	Container vulnerability scanning
Code Quality	black, isort, flake8	Linting & formatting

 Sequence Diagram





Secrets Required

Secret Name	Purpose	Where Used
GITHUB_TOKEN	Auto-provided, GHCR auth	CI (build job)
KUBE_CONFIG	Base64-encoded kubeconfig	CD (deploy job)
SERVICE_URL	(Optional) External service URL	CD (smoke tests)

Key Files

```
.github/workflows/
├── ci.yml           # CI Pipeline definition
└── cd.yml          # CD Pipeline definition

k8s/local/
├── namespace.yaml  # K8s namespace
├── configmap.yaml  # Environment config
├── deployment.yaml # Pod deployment
└── service.yaml    # Service exposure

k8s/
└── argocd-application.yaml # ArgoCD GitOps config

monitoring/
├── prometheus.yaml # Prometheus scrape config
```

```
└─ grafana/
  └─ provisioning/
    └─ datasources/
      └─ datasource.yml # Grafana datasource

scripts/
├─ smoke_test.py      # Post-deploy verification
└─ simulate_performance.py # Load testing
```