MLOps, k8s, GitOps and other acronyms

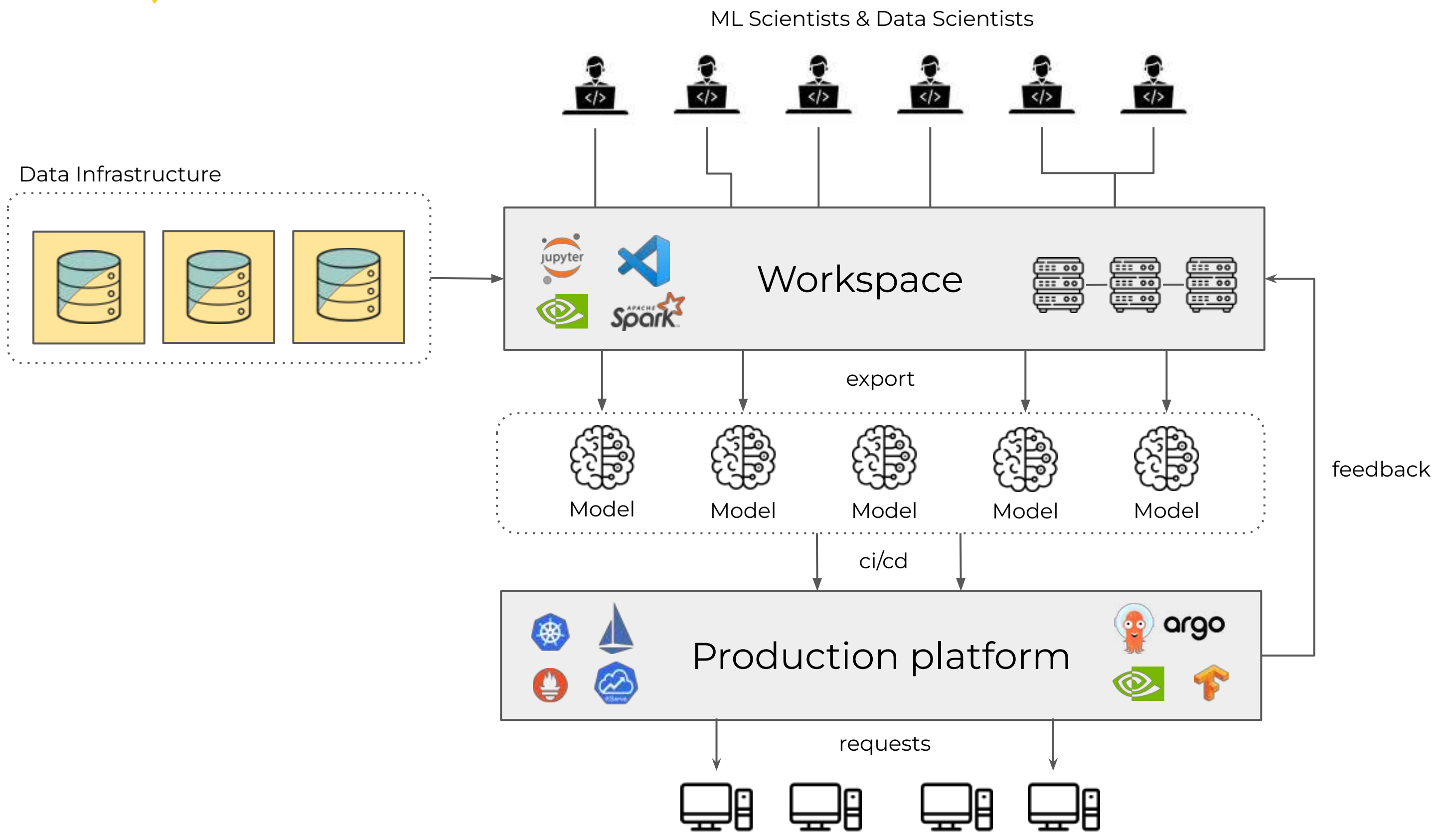
Automating multi-cluster kubernetes environments for ML tasks and services

March 2024

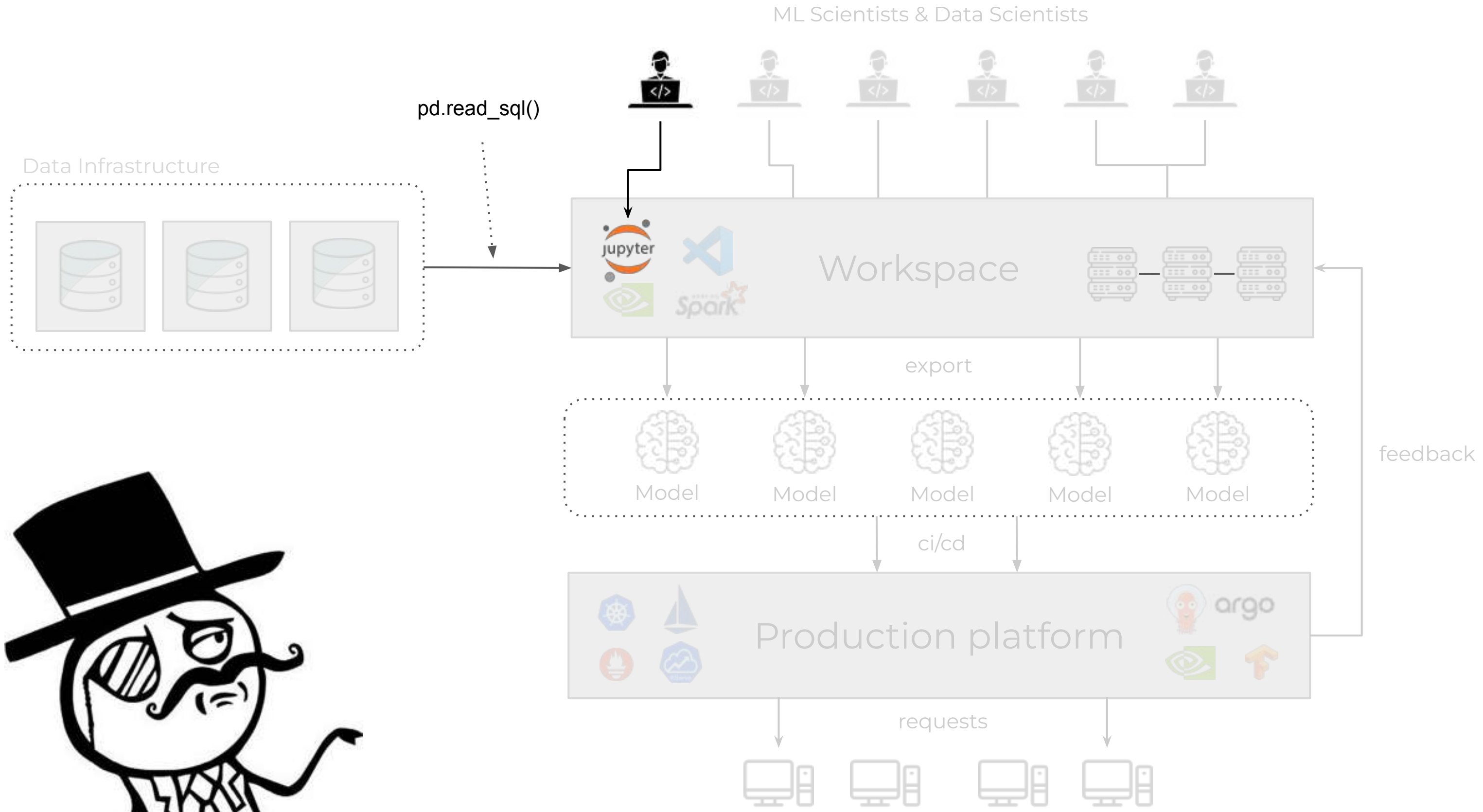
Gleb Vazhenin

- Born in Cherepovets, Russia
- Bachelor of science in Nuclear Physics (NRNU MEPHI)
- 8 Years of DS/MLE/SRE/DevOps experience
- Staff MLE @ Bumble
- Wish to become chess grandmaster

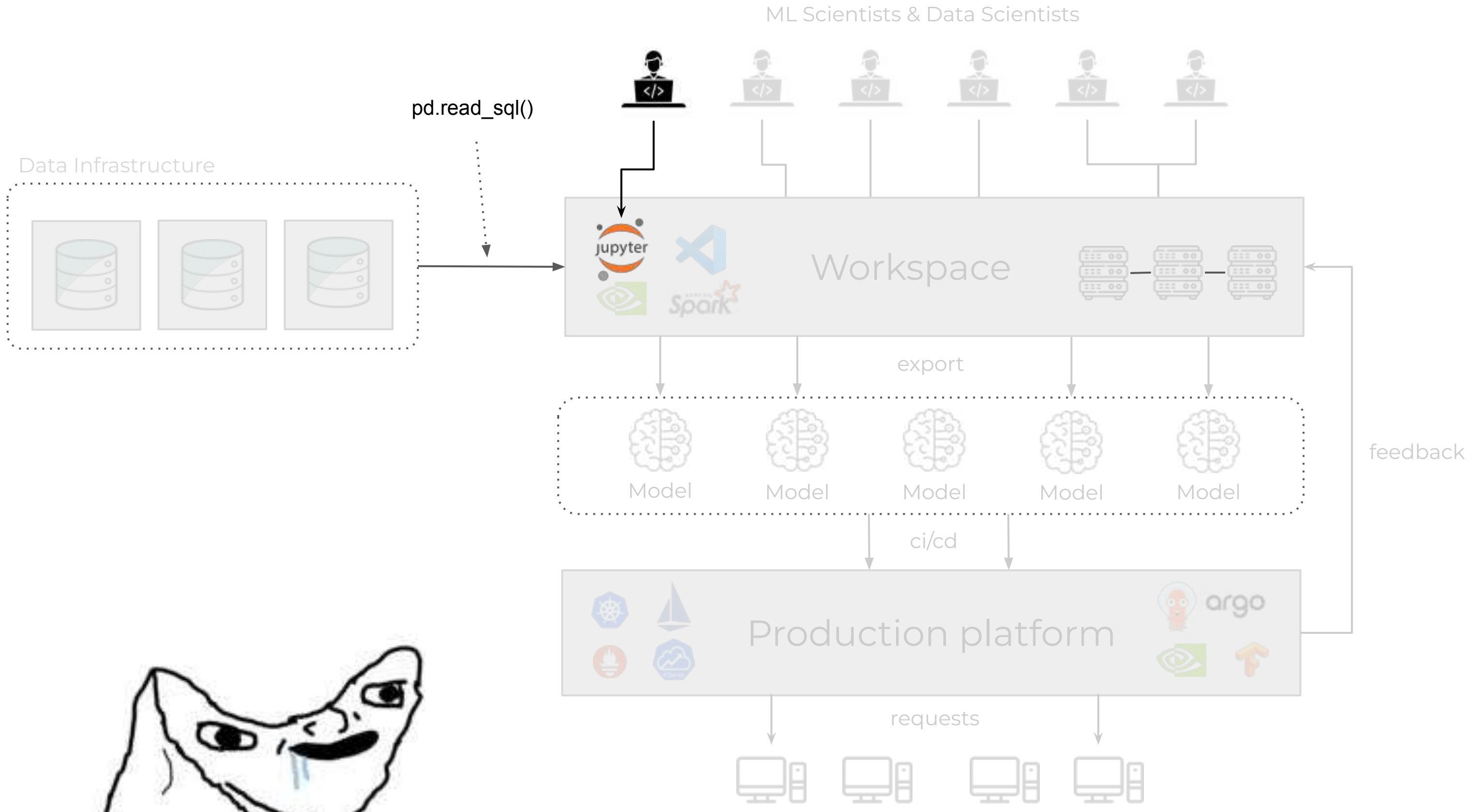
MLOps


ML Scientists & Data Scientists
Data Infrastructure
Workspace
jupyter
Spark
export
Model
Model
Model
Model
Model
ci/cd
feedback
Production platform
argo
requests

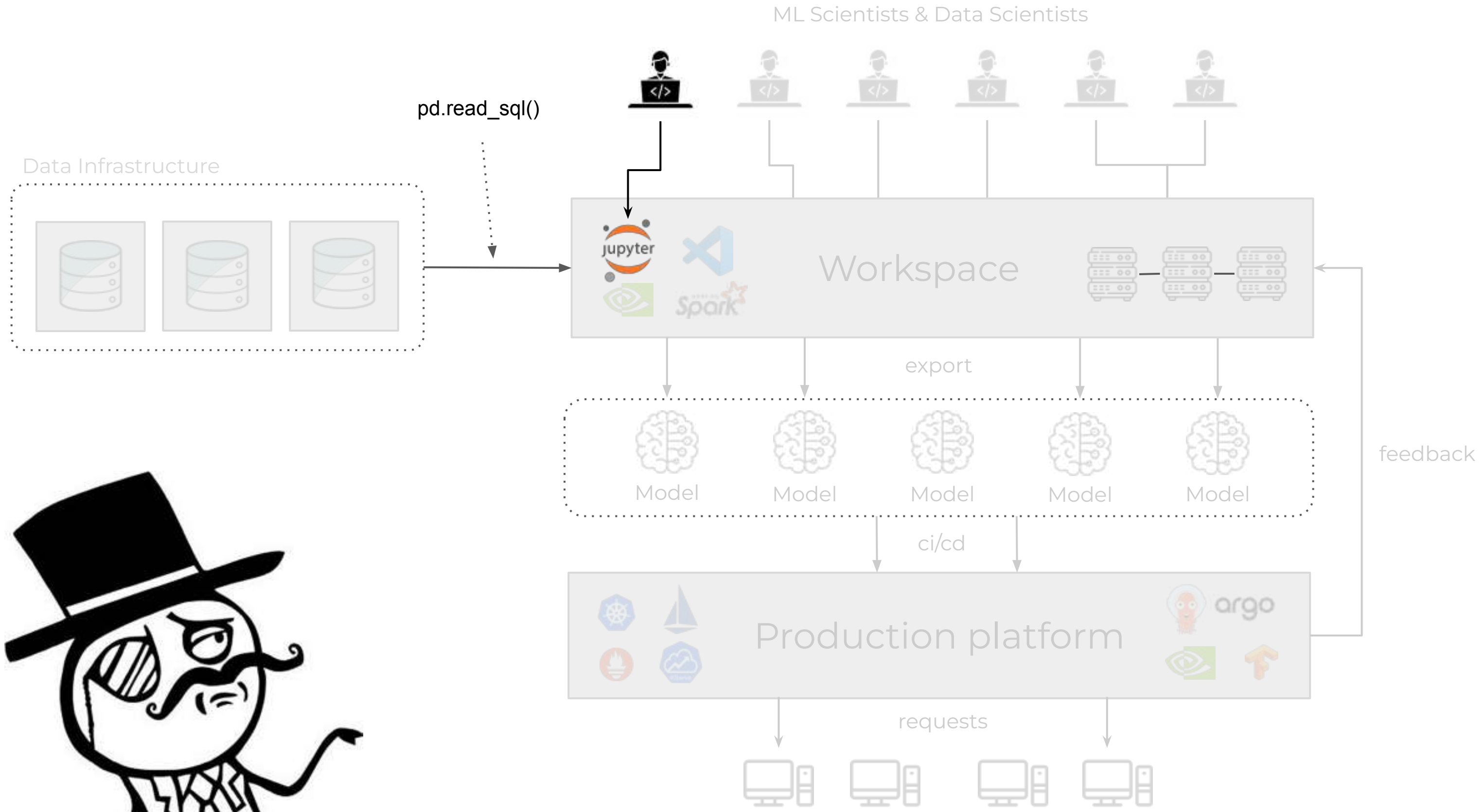
MLOps?


ML Scientists & Data Scientists
pd.read_sql()
Data Infrastructure
jupyter
Spark
Workspace
export
Model
Model
Model
Model
Model
feedback
ci/cd
Production platform
argo
requests

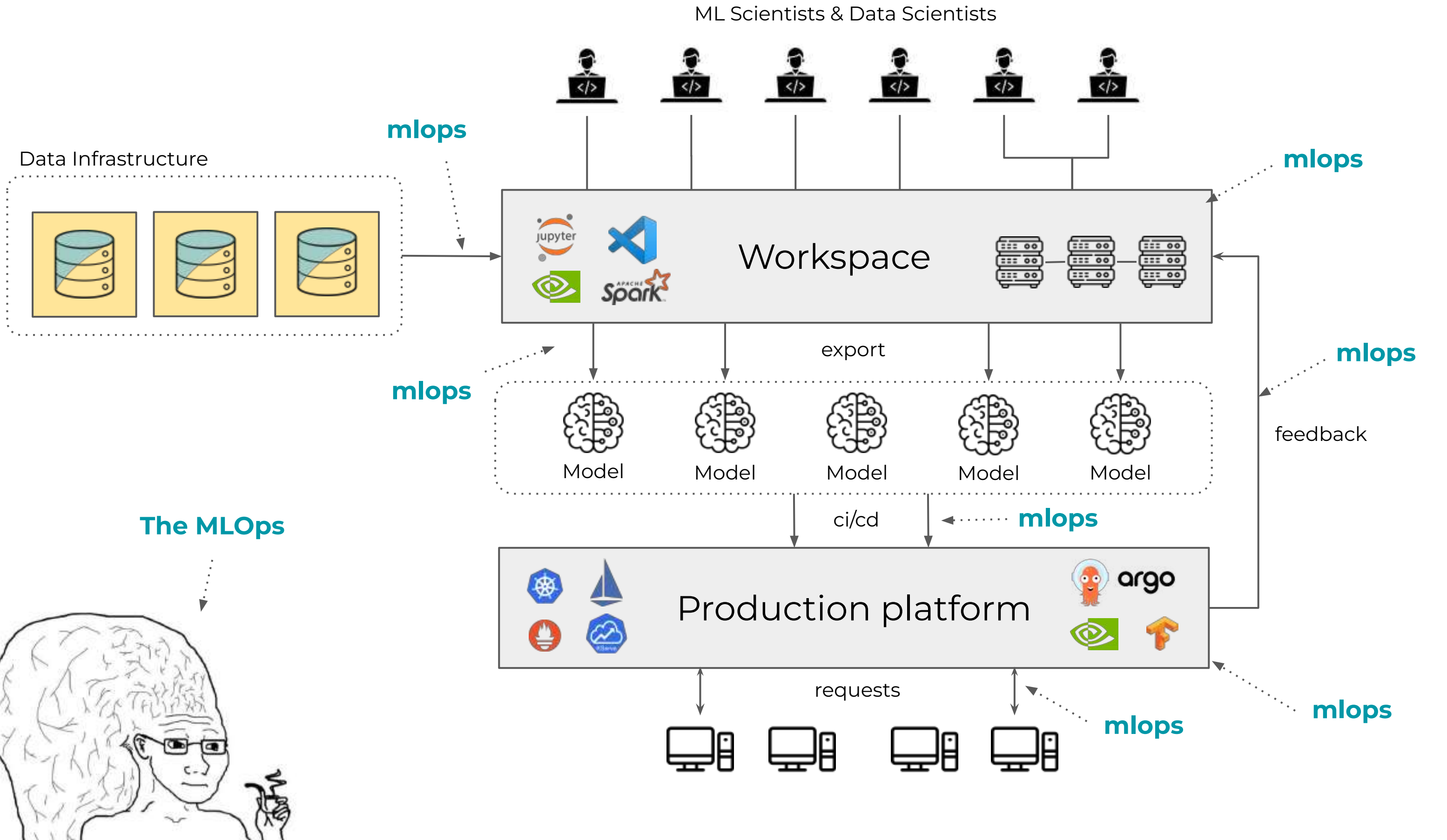
MLOps?


ML Scientists & Data Scientists
pd.read_sql()
Data Infrastructure
jupyter
Spark
Workspace
export
Model
Model
Model
Model
Model
feedback
ci/cd
argo
Production platform
requests

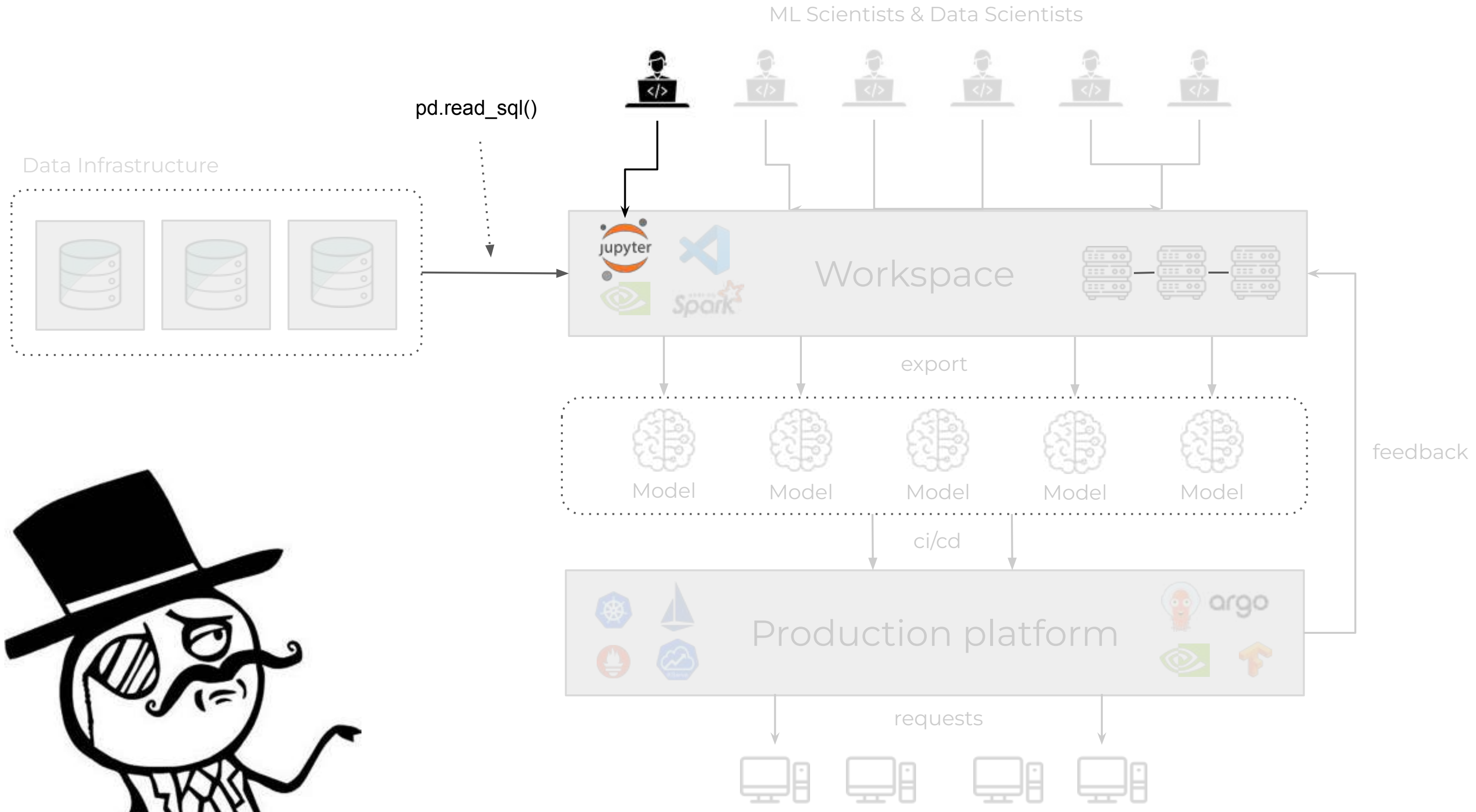
MLOps?


ML Scientists & Data Scientists
pd.read_sql()
Data Infrastructure
jupyter
Spark
Workspace
export
Model
Model
Model
Model
Model
feedback
ci/cd
argo
Production platform
requests

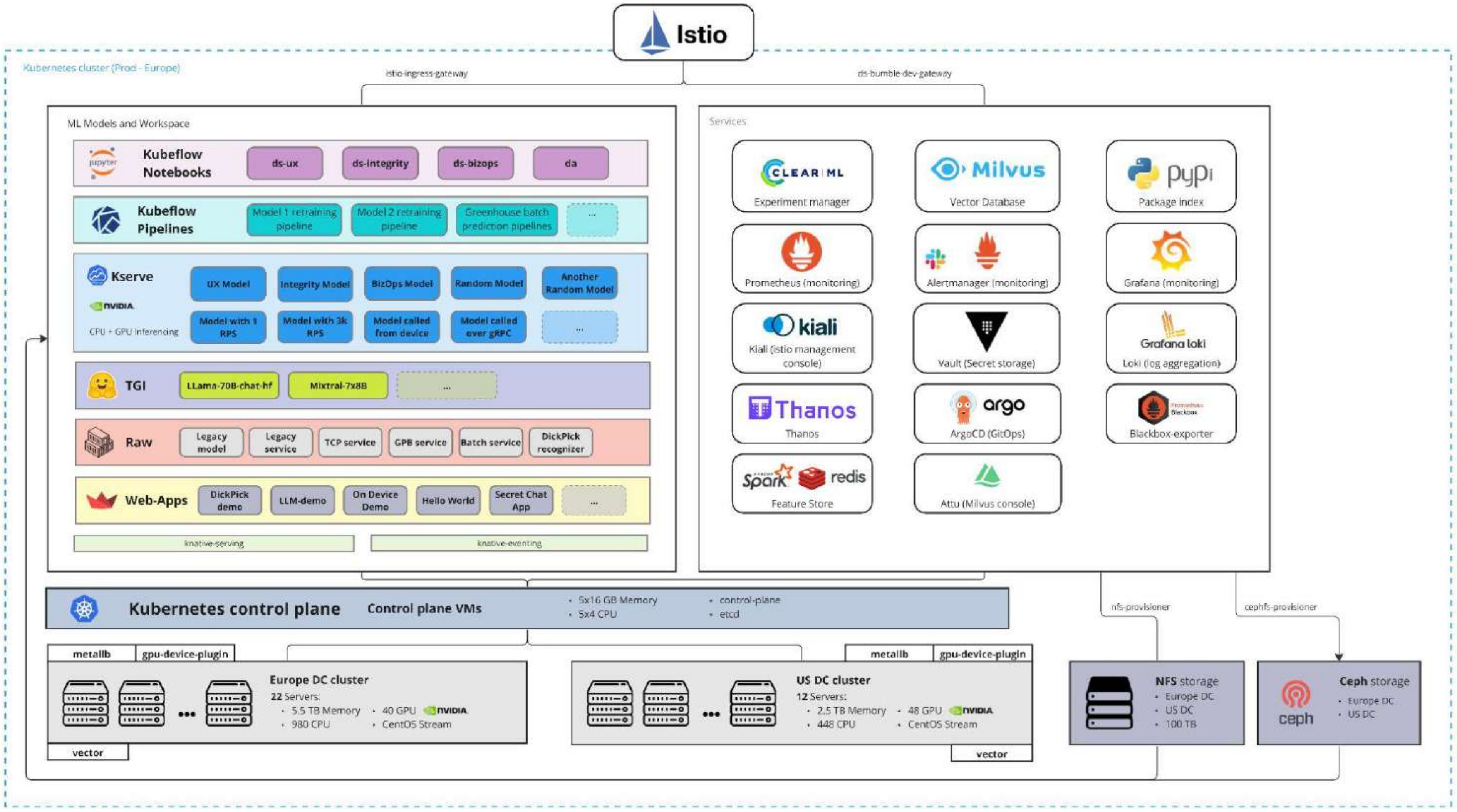
MLOps - is the core function of the Data Science team



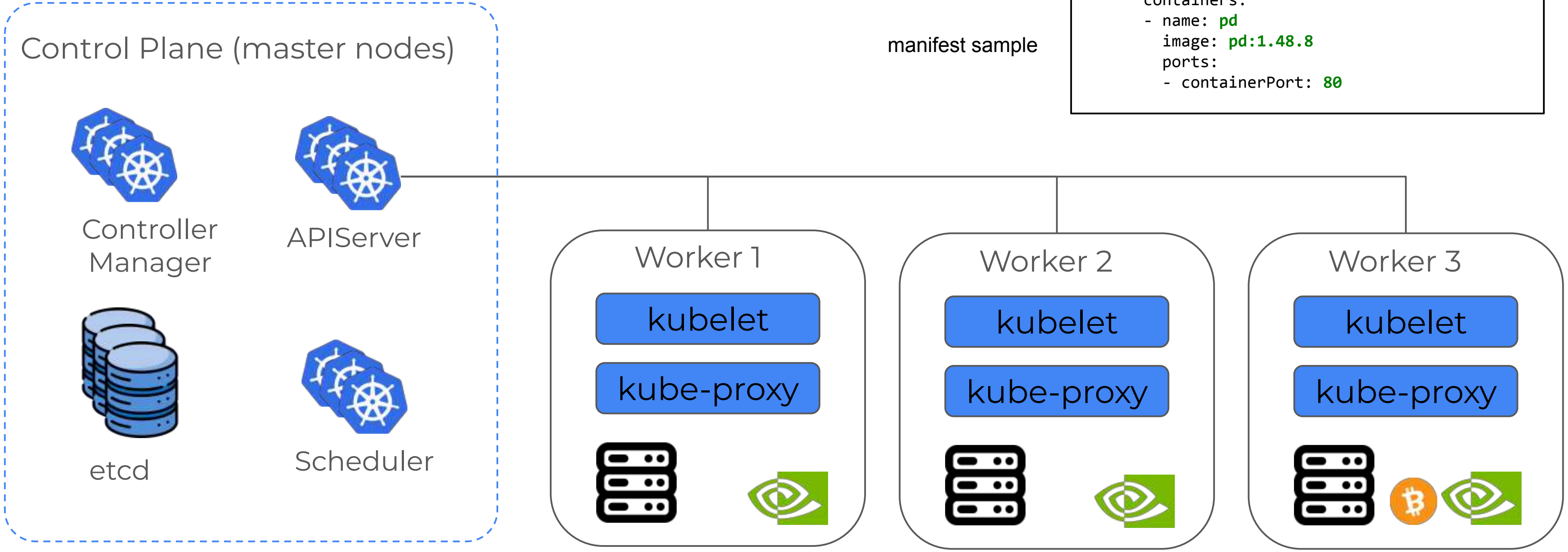
That's the desired way!



What if there are so many different services and tools, that you don’t want to hire a specialist to maintain each one?


Istio
Kubernetes cluster (Prod - Europe)
istio-ingress-gateway
ds-bumble-dev-gateway
ML Models and Workspace
Kubeflow Notebooks
ds-ux
ds-integrity
ds-bizops
da
Kubeflow Pipelines
Model 1 retraining pipeline
Model 2 retraining pipeline
Greenhouse batch prediction pipelines
...
Kserve
NVIDIA
CPU + GPU inferencing
UX Model
Integrity Model
BizOps Model
Random Model
Another Random Model
Model with 1 RPS
Model with 3k RPS
Model called from device
Model called over gRPC
...
TGI
LLama-70B-chat-hf
Mixtral-7x8B
...
Raw
Legacy model
Legacy service
TCP service
GPB service
Batch service
DickPick recognizer
Web-Apps
DickPick demo
LLM-demo
On Device Demo
Hello World
Secret Chat App
...
knative-serving
knative-eventing
Services
CLEAR ML
Experiment manager
Milvus
Vector Database
PyPI
Package index
Prometheus (monitoring)
Alertmanager (monitoring)
Grafana (monitoring)
kiali
Kiali (istio management console)
Vault (Secret storage)
Grafana loki
Loki (log aggregation)
Thanos
Thanos
argo
ArgoCD (GitOps)
Prometheus Blackbox
Blackbox-exporter
Spark
redis
Feature Store
Attu (Milvus console)
Kubernetes control plane
Control plane VMs
• 5x16 GB Memory
• 5x4 CPU
• control-plane
• etcd
nfs-provisioner
cephfs-provisioner
metallb
gpu-device-plugin
Europe DC cluster
22 Servers:
• 5.5 TB Memory
• 980 CPU
• 40 GPU NVIDIA
• CentOS Stream
vector
metallb
gpu-device-plugin
US DC cluster
12 Servers:
• 2.5 TB Memory
• 448 CPU
• 48 GPU NVIDIA
• CentOS Stream
vector
NFS storage
• Europe DC
• US DC
• 100 TB
ceph
Ceph storage
• Europe DC
• US DC

The K8s



manifest sample

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: private-detector
  labels:
    app: pd
spec:
  replicas: 3
  template:
    metadata:
      labels:
        app: pd
    spec:
      containers:
      - name: pd
        image: pd:1.48.8
        ports:
        - containerPort: 80
```

Why K8s?

- **Scalability**
- Fault Tolerance and High Availability
- Declarative Configuration and Automation
- Portability
- Ecosystem and Community
- Containers
- Rolling upgrades and rollbacks
- Service discovery and load balancing

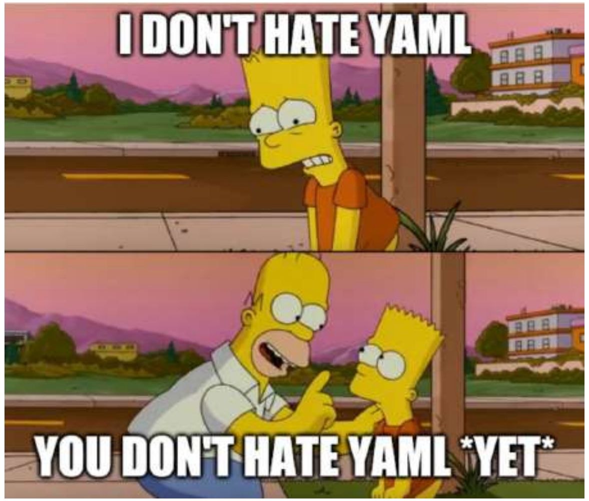
You can't have scalability problems if nobody uses your app

Why K8s?

- Scalability
- **Fault Tolerance and High Availability**
- Declarative Configuration and Automation
- Portability
- Ecosystem and Community
- Containers
- Rolling upgrades and rollbacks
- Service discovery and load balancing

Why K8s?

- Scalability
- Fault Tolerance and High Availability
- **Declarative Configuration and Automation**
- Portability
- Ecosystem and Community
- Containers
- Rolling upgrades and rollbacks
- Service discovery and load balancing

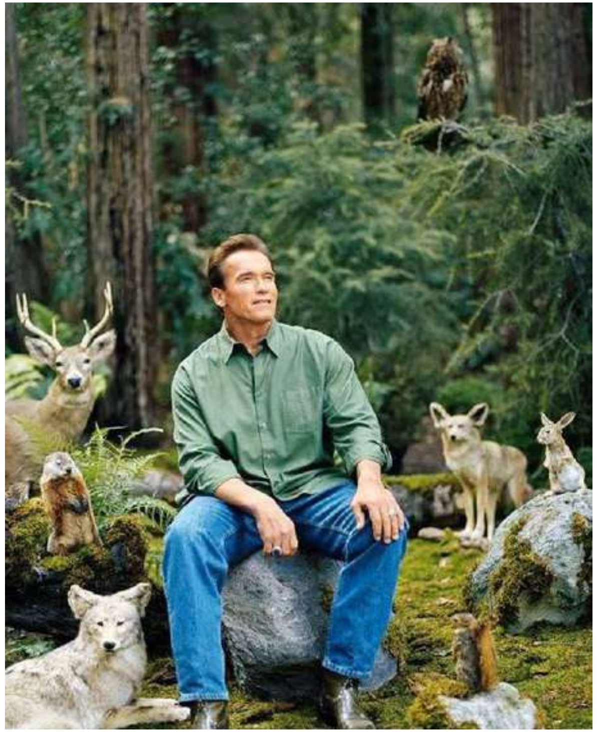


Why K8s?

- Scalability
- Fault Tolerance and High Availability
- Declarative Configuration and Automation
- **Portability**
- Ecosystem and Community
- Containers
- Rolling upgrades and rollbacks
- Service discovery and load balancing

Why K8s?

- Scalability
- Fault Tolerance and High Availability
- Declarative Configuration and Automation
- Portability
- **Ecosystem and Community**
- Containers
- Rolling upgrades and rollbacks
- Service discovery and load balancing



Why K8s?

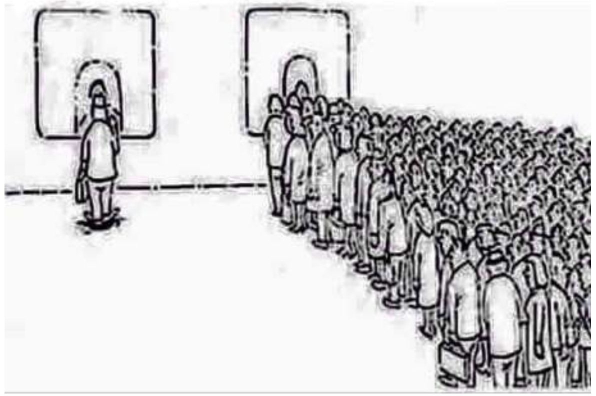
- Scalability
- Fault Tolerance and High Availability
- Declarative Configuration and Automation
- Portability
- Ecosystem and Community
- **Containers**
- Rolling upgrades and rollbacks
- Service discovery and load balancing

Why K8s?

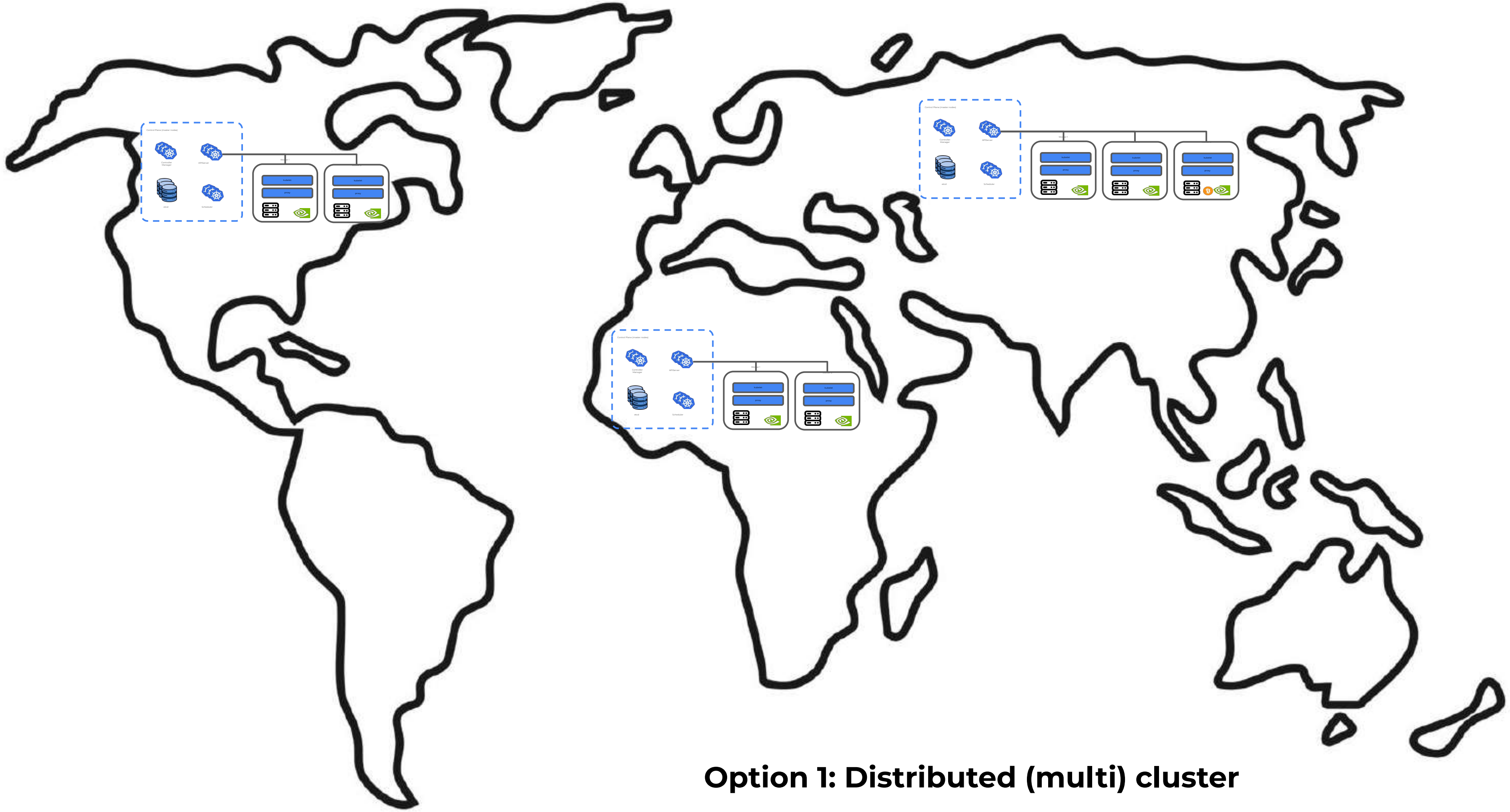
- Scalability
- Fault Tolerance and High Availability
- Declarative Configuration and Automation
- Portability
- Ecosystem and Community
- Containers
- **Rolling upgrades and rollbacks**
- Service discovery and load balancing

Why K8s?

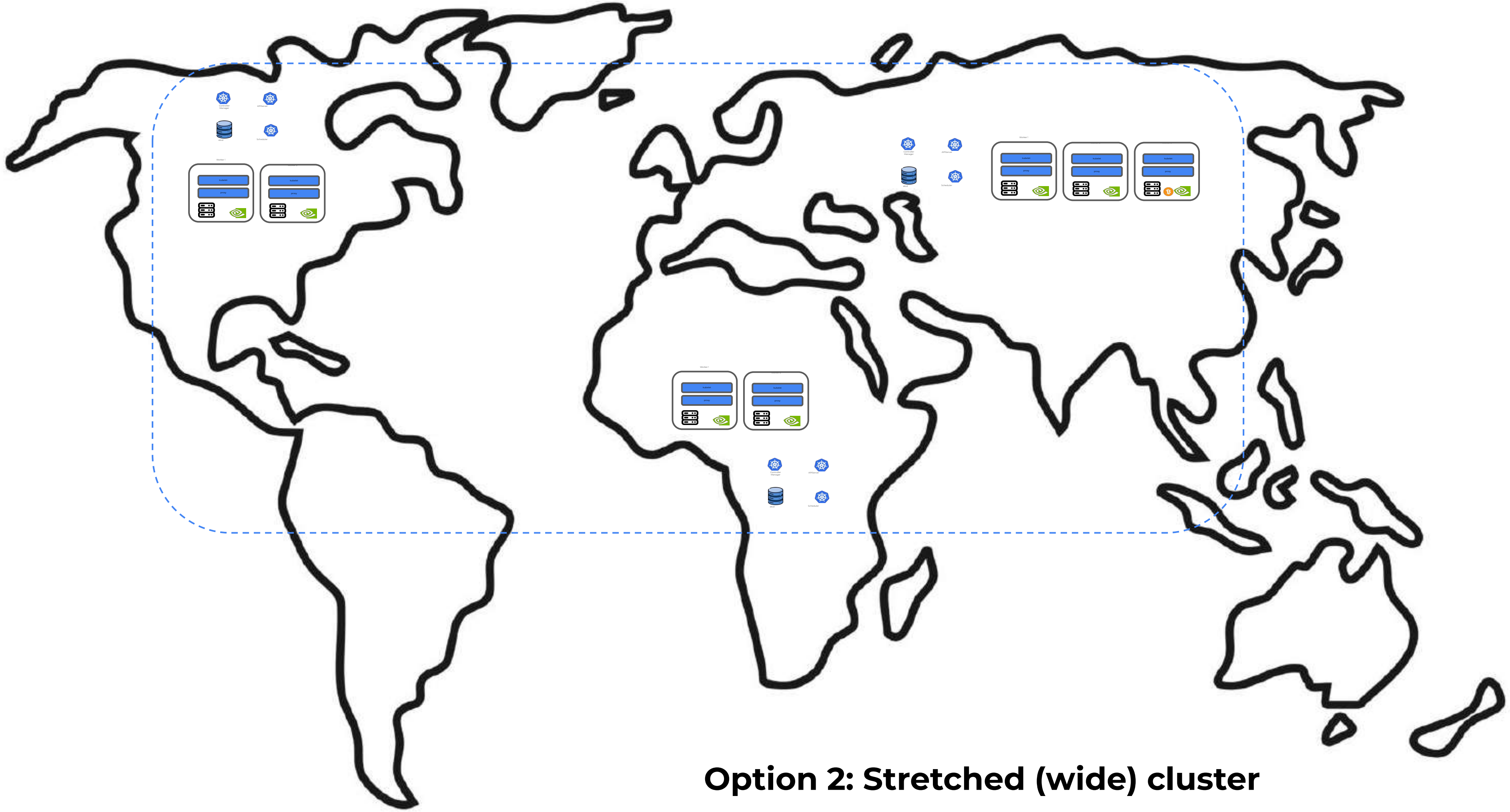
- Scalability
- Fault Tolerance and High Availability
- Declarative Configuration and Automation
- Portability
- Ecosystem and Community
- Containers
- Rolling upgrades and rollbacks
- **Service discovery and load balancing**



What if the DC is so cheap, that there’s a disaster in electricity every other week?



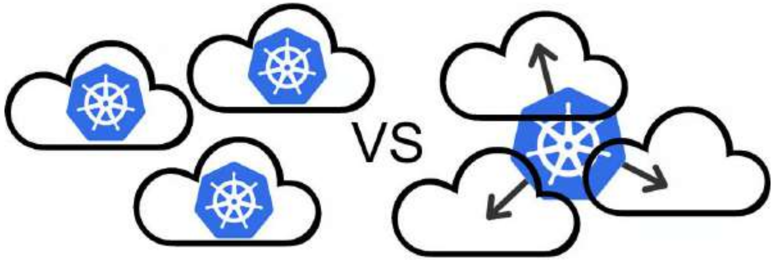
Option 1: Distributed (multi) cluster



Option 2: Stretched (wide) cluster

Wide vs Multi - Key points

- **Redundancy** - same
- **Scalability -** same
- **Latency**
  - Wide **- higher** latencies
  - Multi **- lower** latencies
- **Isolation**
  - Wide **-** same cluster, **less** isolation
  - Multi **-** different clusters, **strong** isolation by design
- **Management complexity**
  - Wide - **easier** to maintain (same cluster), though not out-of-the-box
  - Multi - **harder** to maintain



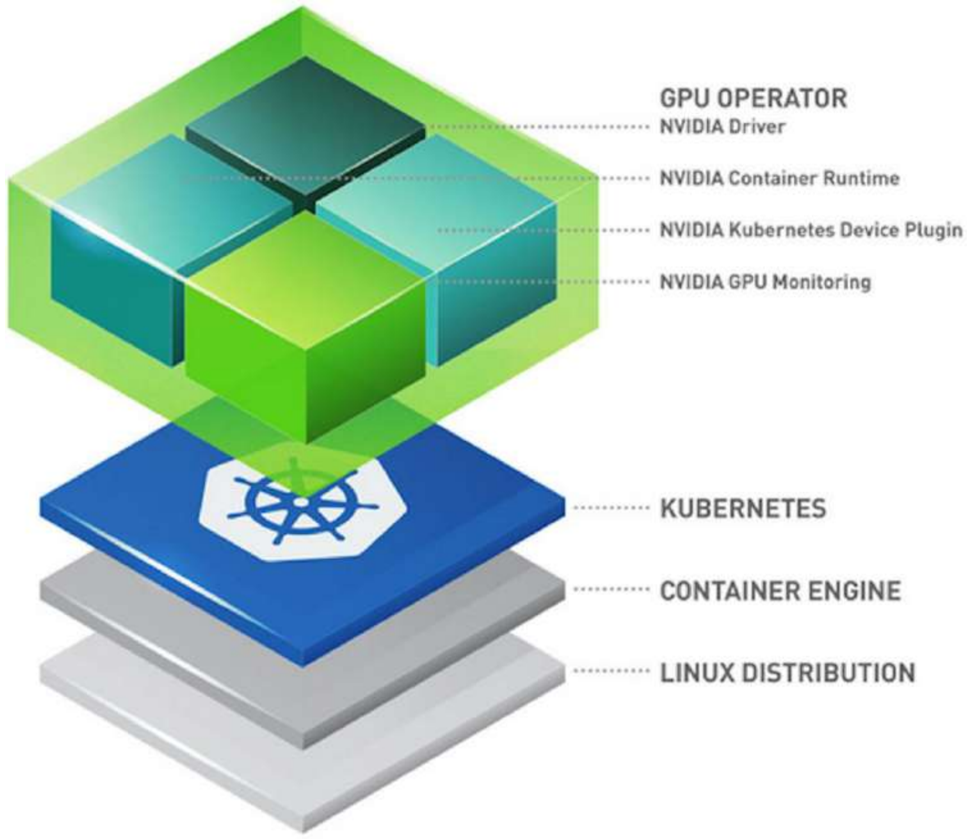
GPUs on k8s

March 2024

NVIDIA GPU Operator

Components:

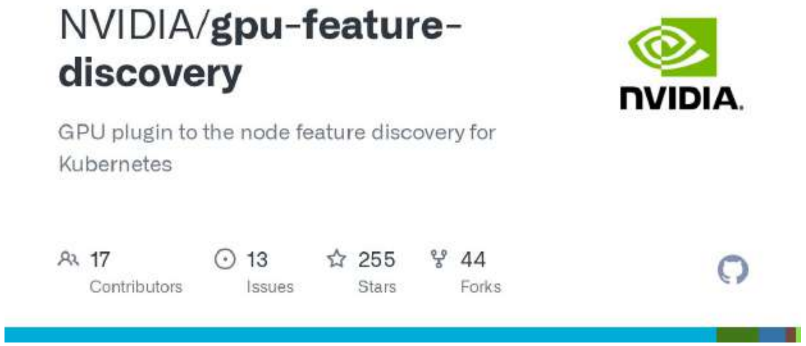
- GPU Feature discovery
- Nvidia Container Runtime
- K8s Device Plugin
- DCGM Exporter
- Driver Manager
- MIG Manager



NVIDIA GPU Operator

Components:

- **GPU Feature discovery**
- Nvidia Container Runtime
- K8s Device Plugin
- DCGM Exporter
- Driver Manager
- MIG Manager


NVIDIA/**gpu-feature-discovery**

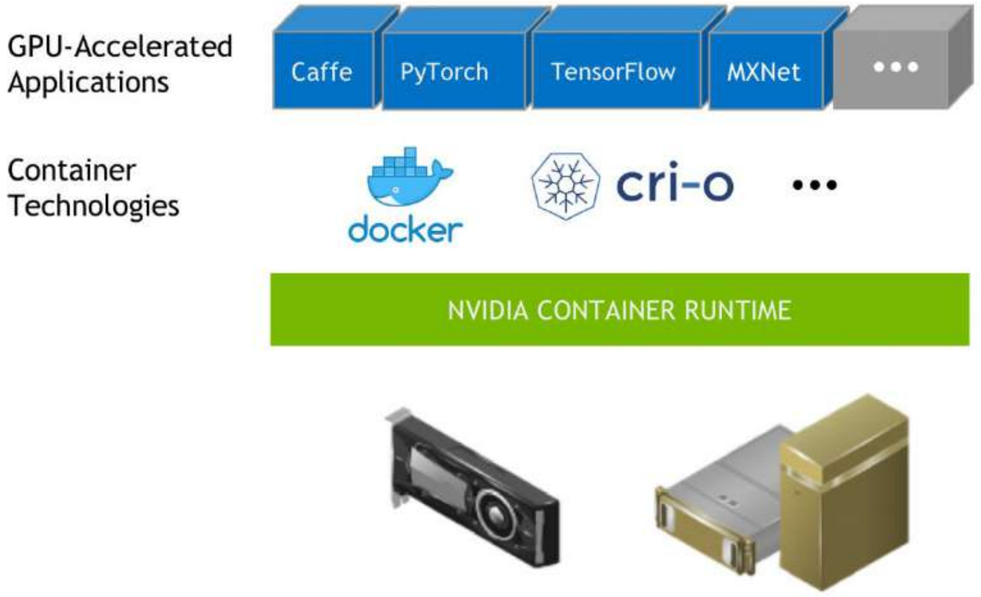
GPU plugin to the node feature discovery for Kubernetes

```
$ kubectl get nodes --show-labels

NAME       STATUS   ROLES     AGE    VERSION    LABELS
ds-node1   Ready    worker    210d   v1.24.3    ...,nvidia.com/gpu.product=NVIDIA-A100,nvidia.com/gpu.replicas=2
ds-node2   Ready    worker    210d   v1.24.3    ...,nvidia.com/gpu.product=Tesla-T4,nvidia.com/gpu.replicas=4
ds-node3   Ready    worker    210d   v1.24.3    ...,nvidia.com/gpu.product=Tesla-T4,nvidia.com/gpu.replicas=4
```

NVIDIA GPU Operator

Components:

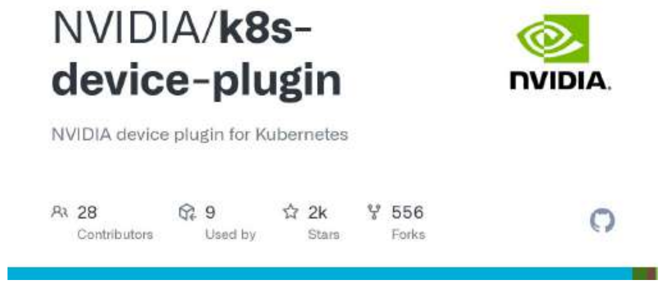
- GPU Feature discovery
- **Nvidia Container Runtime**
- K8s Device Plugin
- DCGM Exporter
- Driver Manager
- MIG Manager



NVIDIA GPU Operator

Components:

- GPU Feature discovery
- Nvidia Container Runtime
- **K8s Device Plugin**
- DCGM Exporter
- Driver Manager
- MIG Manager



NVIDIA/**k8s-device-plugin**

NVIDIA device plugin for Kubernetes

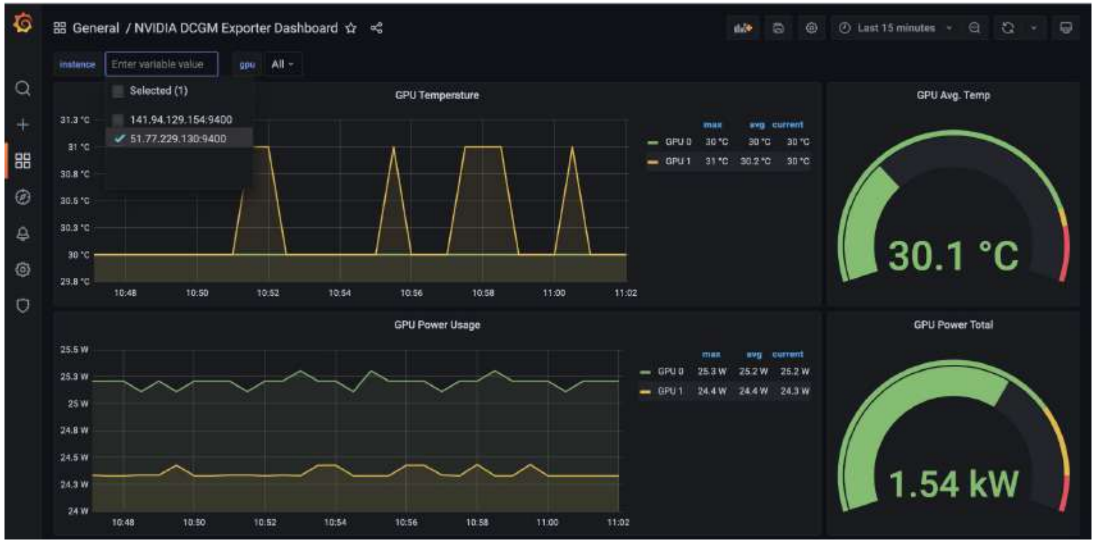
28 Contributors · 9 Used by · 2k Stars · 556 Forks

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: private-detector
  labels:
    app: pd
spec:
  replicas: 3
  template:
    metadata:
      labels:
        app: pd
    spec:
      containers:
      - name: pd
        image: pd:1.48.8
        ports:
        - containerPort: 80
         resources:
          limits:
            cpu: 2
            memory: 16Gi
            nvidia.com/gpu: 2
```

NVIDIA GPU Operator

Components:

- GPU Feature discovery
- Nvidia Container Runtime
- K8s Device Plugin
- **DCGM Exporter**
- Driver Manager
- MIG Manager



NVIDIA GPU Operator

Components:

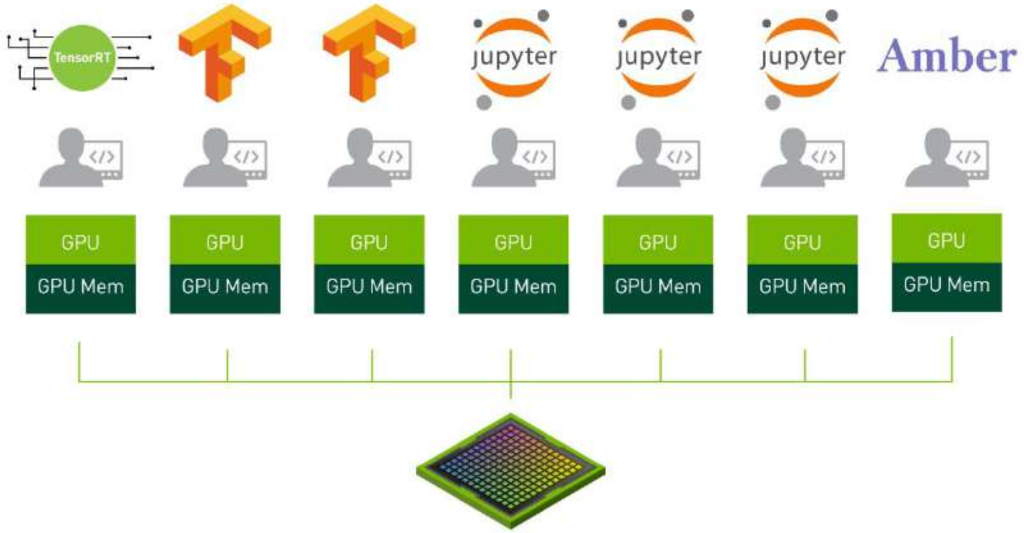
- GPU Feature discovery
- Nvidia Container Runtime
- K8s Device Plugin
- DCGM Exporter
- **Driver Manager**
- MIG Manager

```
Fri Dec 30 10:57:52 2022
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 525.60       Driver Version: 525.60       CUDA Version: 12.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  NVIDIA A100-PCI...  Off  | 00000000:86:00.0 Off |                    0 |
| N/A   30C    P0    36W / 250W |      0MiB / 40960MiB |      0%      Default |
|                               |                      |             Disabled |
+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

NVIDIA GPU Operator

Components:

- GPU Feature discovery
- Nvidia Container Runtime
- K8s Device Plugin
- DCGM Exporter
- Driver Manager
- **MIG Manager**



NVIDIA GPU Operator

Components:

- GPU Feature discovery
- Nvidia Container Runtime
- K8s Device Plugin
- DCGM Exporter
- Driver Manager
- **MIG Manager**

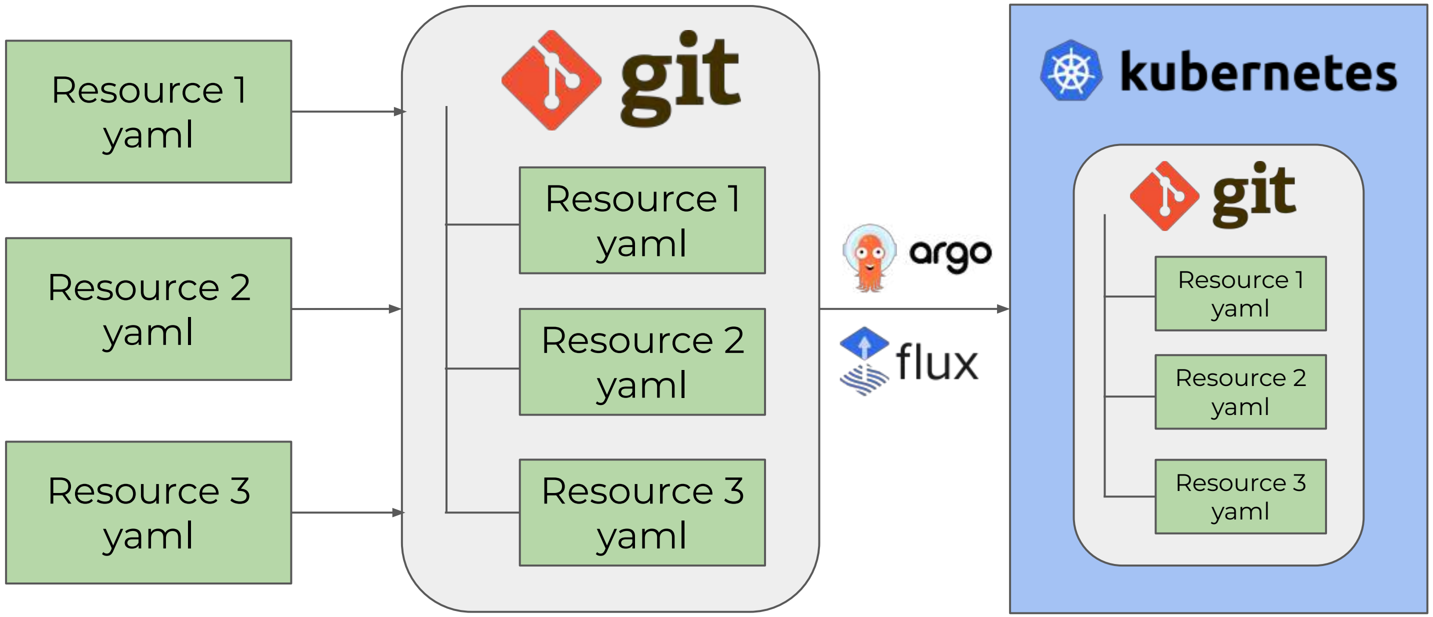
```
kubectl label nodes ds-node1 nvidia.com/mig.config=all-1g.10gb

kubectl label nodes ds-node1 nvidia.com/mig.config=all-1g.5gb

kubectl label nodes ds-node1 nvidia.com/mig.config=all-3g.40gb
```

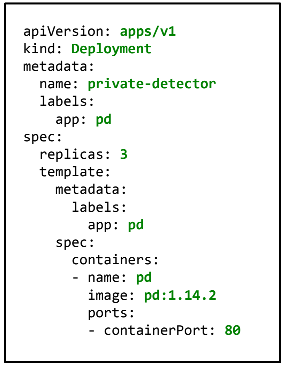
```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: private-detector
  labels:
    app: pd
spec:
  replicas: 3
  template:
    metadata:
      labels:
        app: pd
    spec:
      containers:
      - name: pd
        image: pd:1.76.9
        ports:
        - containerPort: 80
         resources:
          limits:
            cpu: 1
            memory: 2Gi
            nvidia.com/mig-1g.5gb: 1
```

What if I need to manage 50 GPU clusters simultaneously, taking into account different service configurations?

GitOps

GitOps modernises software management by allowing Engineers to declaratively manage infrastructure and software code using a single source of truth — typically a Git repository.

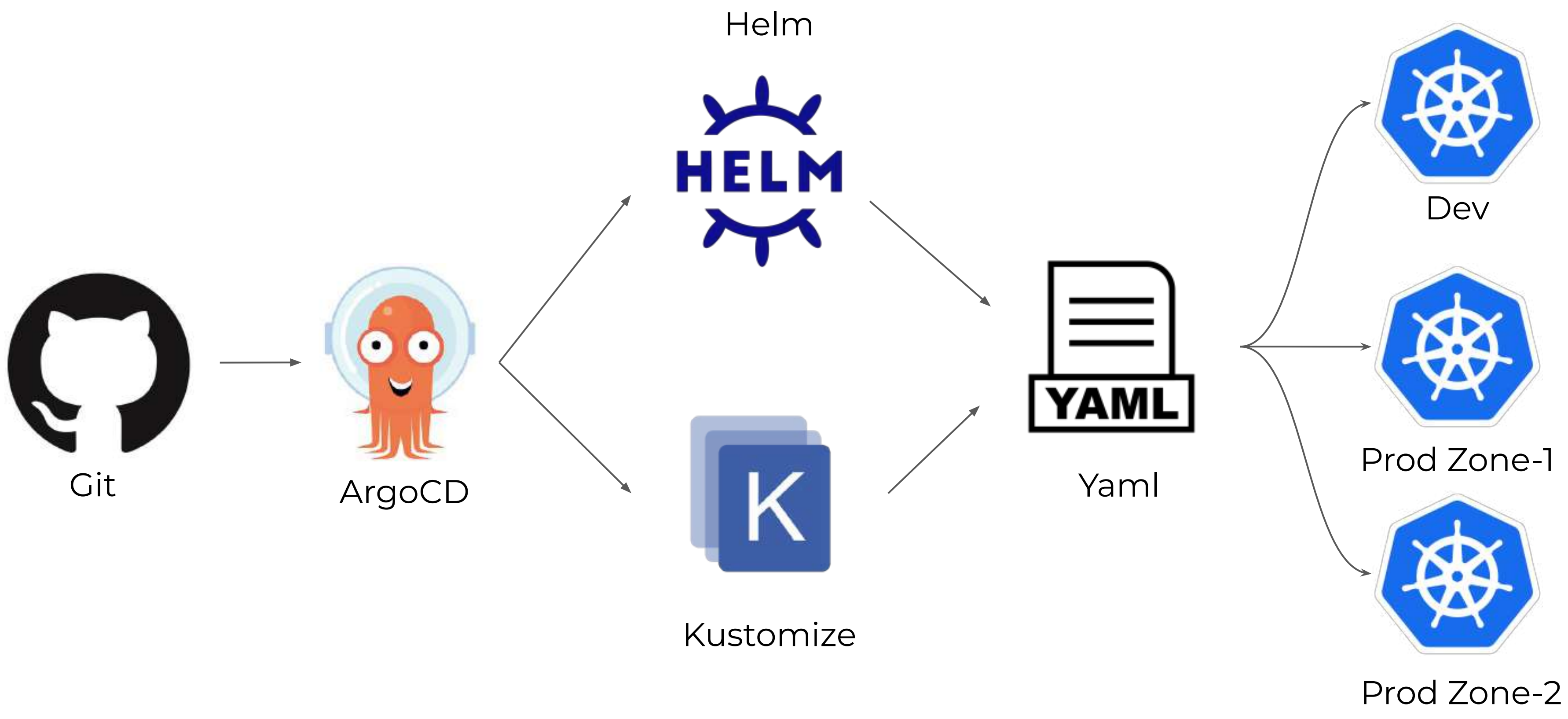




```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: private-detector
  labels:
    app: pd
spec:
  replicas: 3
  template:
    metadata:
      labels:
        app: pd
    spec:
      containers:
      - name: pd
        image: pd:1.14.2
        ports:
        - containerPort: 80
```

yaml sample

Kustomize & Helm are the tools used to manage (template) kubernetes manifests



Smooth GitOps experience: Overlays

```
apps
└── private-detector
    ├── base
    │   └── deployment.yaml
    ├── bumble
    │   ├── deployment_patch.yaml
    │   └── kustomization.yaml
    ├── zone-1-dev
    │   ├── deployment_zone_1_dev_patch.yaml
    │   └── kustomization.yaml
    ├── zone-1-prod
    │   ├── deployment_zone_1_prod_patch.yaml
    │   └── kustomization.yaml
    └── zone-2-prod
        ├── deployment_zone_2_patch.yaml
        └── kustomization.yaml
```

apps/private-detector/base/deployment.yaml

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: private-detector
  labels:
    app: pd
spec:
  replicas: 3
  template:
    metadata:
      labels:
        app: pd
    spec:
      containers:
      - name: pd
        image: pd:1.14.2
        ports:
        - containerPort: 80
```

Shipped like this by private-detector community

apps/private-detector/bumble/deployment_patch.yaml

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: private-detector
spec:
  template:
    spec:
      image: pd:1.13.2
```

apps/private-detector/bumble/kustomization.yaml

```yaml
resources:
- ../base/deployment.yaml

patchesStrategicMerge:
- deployment_patch.yaml
```

Bumble cluster-agnostic label patch

Smooth GitOps experience: Overlays

```
└── apps
    └── private-detector
        ├── base
        │   └── deployment.yaml
        ├── bumble
        │   ├── deployment_patch.yaml
        │   └── kustomization.yaml
        ├── zone-1-dev
        │   ├── deployment_zone_1_dev_patch.yaml
        │   └── kustomization.yaml
        ├── zone-1-prod
        │   ├── deployment_zone_1_prod_patch.yaml
        │   └── kustomization.yaml
        └── zone-2-prod
            ├── deployment_zone_2_patch.yaml
            └── kustomization.yaml
```

apps/private-detector/**zone-1-dev**/deployment_zone_1_dev_patch.yaml

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: private-detector
spec:
  replicas: 1
  template:
    spec:
      containers:
      - name: pd
        resources:
          limits:
            nvidia.com/mig-1g.5gb: 1
```

apps/private-detector/**zone-1-dev**/kustomization.yaml

```yaml
resources:
- ../bumble

patchesStrategicMerge:
- deployment_zone_1_patch.yaml
```

Smooth GitOps experience: Two-level overlays

- **apps**
  - **private-detector**
    - base
      - deployment.yaml
    - bumble
      - deployment_patch.yaml
      - kustomization.yaml
    - zone-1-dev
      - deployment_zone_1_dev_patch.yaml
      - kustomization.yaml
    - **zone-1-prod**
      - **deployment_zone_1_prod_patch.yaml**
      - **kustomization.yaml**
    - zone-2-prod
      - deployment_zone_2_patch.yaml
      - kustomization.yaml

apps/private-detector/**zone-1-prod**/deployment_zone_1_prod_patch.yaml

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: private-detector
spec:
  replicas: 10
  template:
    spec:
      nodeSelector:
        nvidia.com/gpu.product: NVIDIA-A100
      containers:
      - name: pd
        resources:
          limits:
            nvidia.com/gpu: 1
```

apps/private-detector/**zone-1-prod**/kustomization.yaml

```yaml
resources:
- ../../../base/istio/bumble

patchesStrategicMerge:
- deployment_zone_2_patch.yaml
```

Smooth GitOps experience: Two-level overlays

base manifest

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: private-detector
  labels:
    app: pd
spec:
  replicas: 3
  template:
    metadata:
      labels:
        app: pd
    spec:
      containers:
      - name: pd
        image: pd:1.14.2
        ports:
        - containerPort: 80
```

zone-1-dev complete manifest

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: private-detector
spec:
  replicas: 1
  template:
    metadata:
      labels:
        app: pd
    spec:
      containers:
      - name: pd
        image: pd:1.13.2
        ports:
        - containerPort: 80
        resources:
          limits:
            nvidia.com/mig-1g.5gb: 1
```

zone-1-prod complete manifest

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: private-detector
spec:
  replicas: 10
  template:
    metadata:
      labels:
        app: pd
    spec:
      containers:
      - name: pd
        image: pd:1.13.2
        ports:
        - containerPort: 80
        resources:
          limits:
            nvidia.com/gpu: 1
```

GitOps: Transparency

```
apps
    private-detector
        base
        bumble
        zone-1-dev
        zone-1-prod
    toxicity-detector
        base
        bumble
        zone-1-dev
        zone-2-prod
mapping
    zone-1-dev-kustomization.yaml
    zone-1-prod-kustomization.yaml
    zone-2-prod-kustomization.yaml
```

- **Transparency:** Each cluster has a list of the resources intended to run there.

mapping/zone-1-dev-kustomization.yaml

```yaml
kind: Kustomization
resources:
- apps/private-detector/zone-1-dev
- apps/toxicity-detector/zone-1-dev
```

mapping/zone-1-prod-kustomization.yaml

```yaml
kind: Kustomization
resources:
- apps/private-detector/zone-1-prod
# - apps/toxicity-detector/zone-1-prod
```

mapping/zone-2-prod-kustomization.yaml

```yaml
kind: Kustomization
resources:
# - apps/private-detector/zone-2-prod
- apps/toxicity-detector/zone-2-prod
```

GitOps: Extensibility

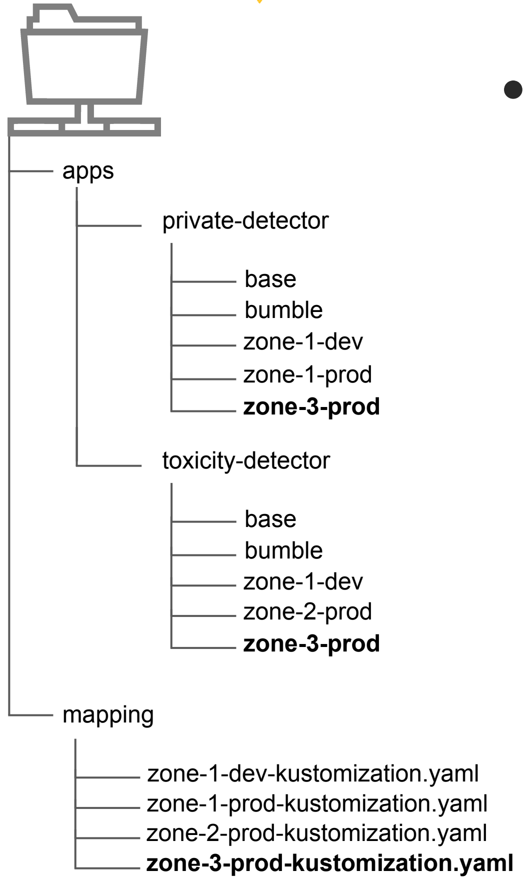
- apps
  - private-detector
  - toxicity-detector
  - **scam-detector**
  - **fake-detector**
- mapping
  - zone-1-dev-kustomization.yaml
  - zone-1-prod-kustomization.yaml
  - zone-2-prod-kustomization.yaml

- **Transparency:** The process of adding new resources is simple and clear

mapping/zone-1-dev-kustomization.yaml

```yaml
kind: Kustomization
resources:
- apps/private-detector/zone-1-dev
- apps/toxicity-detector/zone-1-dev
- apps/scam-detector/zone-1-dev
- apps/fake-detector/zone-1-dev
```

GitOps: Scalability



```
apps
    private-detector
        base
        bumble
        zone-1-dev
        zone-1-prod
        zone-3-prod
    toxicity-detector
        base
        bumble
        zone-1-dev
        zone-2-prod
        zone-3-prod
mapping
    zone-1-dev-kustomization.yaml
    zone-1-prod-kustomization.yaml
    zone-2-prod-kustomization.yaml
    zone-3-prod-kustomization.yaml
```

- **Scalability:** New environment could be added easily

mapping/zone-3-prod-kustomization.yaml

```
kind: Kustomization
resources:
- apps/private-detector/zone-3-prod
- apps/toxicity-detector/zone-3-prod
```

GitOps: Convenience

- apps
  - private-detector
    - **base**
    - bumble
    - zone-1-dev
    - zone-1-prod
    - zone-2-prod
  - toxicity-detector
    - **base**
    - bumble
    - zone-1-dev
    - zone-1-prod
    - zone-2-prod
- mapping
  - zone-1-dev-kustomization.yaml
  - zone-1-prod-kustomization.yaml
  - zone-2-prod-kustomization.yaml

- **Convenience:** Resource upgrades require a single commit to be rolled out to all overlay clusters; overlay-specific parameters are declaratively defined.

GitOps: Collaboration

- apps
  - private-detector
    - base
    - **bumble**
    - **zone-1-dev**
    - **zone-1-prod**
    - **zone-2-prod**
  - toxicity-detector
    - base
    - **bumble**
    - **zone-1-dev**
    - **zone-1-prod**
    - **zone-2-prod**
- mapping
  - zone-1-dev-kustomization.yaml
  - zone-1-prod-kustomization.yaml
  - zone-2-prod-kustomization.yaml

- **Collaboration:** Both cluster-specific and cluster-agnostic changes to a common base are described in a declarative way.

Thank you!

March 2024