Securing and Scaling Jupyter at Bloomberg: A Kubernetes-Native Approach

Bloomberg Engineering

JupyterCon 2025
November 4, 2025

Tucker Vento, Team Lead & Yifan Zhang, Software Engineer
Cloud Native Compute Services
AI Training & Notebooks Engineering Team

TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Our description

At Bloomberg, we operate a secure, enterprise-grade Jupyter platform supporting hundreds of data scientists, engineers, and researchers across the firm. This talk explores how we've built and evolved our internal managed notebooks infrastructure using open source principles and custom orchestration tools tailored for multi-tenant Kubernetes environments.

We'll dive into how we manage notebook lifecycle and orchestration via our Jupyter CRD and controller, and how we enforce strong authentication and authorization using Istio, Open Policy Agent, and Single Sign-On. We'll also share our strategies for enforcing service boundaries with Calico, reverse proxies, and other middleware integrations, as well as how we approach GPU management at scale.

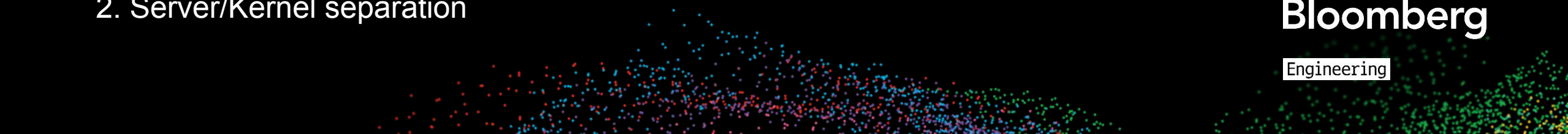
As security and usability go hand-in-hand, attendees will learn how we use a Vault-backed CSI secrets provider to securely inject Git credentials into user notebooks, and how we prevent sensitive data exposure using Fluent Bit filters. We'll discuss operational challenges like zombie processes and long-lived sessions, and look ahead at ideas like kernel sharing, checkpointing, and integrations with our AI model serving & inference platform.

Bloomberg

Engineering

outline

1. Introduction to DSP multi-tenant jupyter notebook as a service platform
2. JN CRD + Controller, how a notebook is launched
3. A notebook is ready, how does request come in? (Ingress)
4. Notebook users need to access external services, how to ensure safe network access? (egress)
5. Notebook developer experience
   1. Git credential injection based on CSI secrets provider
   2. Access to GPU
   3. Existing network storage solution
      * Object storage
      * HDFS
6. Interesting stories
   1. Zombie process
   2. BCS credential leakage in logs
7. Future
   1. Distributed persistent storage based on JuiceFS
   2. Server/Kernel separation



Bloomberg

Engineering

Agenda

- AI Infrastructure at Bloomberg
- Kubernetes 101
- Jupyter Notebook Platform Authn/Authz
- Platform Developer Experience
- Operational Challenges
- Future Plans

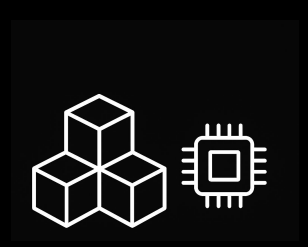
TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

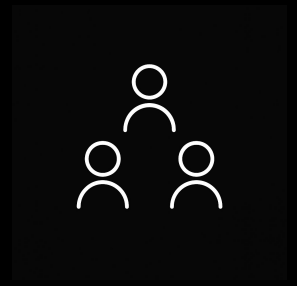
Bloomberg

Engineering

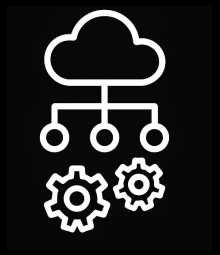
Bloomberg's Data Science Platform



Infrastructure Optimized for ML



ML Developers



ML Services & Tools

Data Gathering & Data Exploration → Build Model & ML Experimentation → Model Deployment & Serving → Prod Model Maintenance (Monitor & Update)

Jupyter Notebooks

Distributed Training & HPC

Model Serving

ML Pipelines

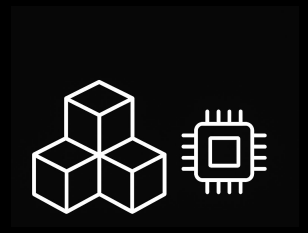
TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

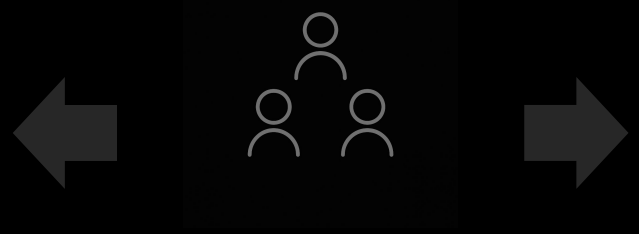
Bloomberg

Engineering

Bloomberg's Data Science Platform



Infrastructure Optimized for ML



ML Developers



ML Services & Tools

Data Gathering & Data Exploration → Build Model & ML Experimentation → Model Deployment & Serving → Prod Model Maintenance (Monitor & Update)

Jupyter Notebooks

Distributed Training & HPC

Model Serving

ML Pipelines

TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

What is Kubernetes?

Kubernetes (K8s) is an open source platform that automates deploying, scaling, and managing containers — like an operating system for the cloud ☁️.

📈 Scalability – Automatically adds or removes workloads

🔁 Self-healing – Restarts failed apps automatically

🔒 Isolation – Keeps workloads secure and organized

🧠 Automation – Declarative config: “what you want,” not “how to do it”

TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Why Kubernetes?

- Runs Jupyter notebooks reliably across thousands of users
- Automatically balances workloads (so no one machine gets overloaded)
- Makes upgrades, scaling, and recovery seamless

© 2025 Bloomberg Finance L.P. All rights reserved.

Kubernetes Glossary

- **Pod:** Smallest deployable unit – one or more containers that share network/storage
- **Service:** Stable virtual endpoint (DNS/IP) that load-balances to matching Pods, even as Pods churn
- **Namespace:** Logical partition of the cluster for isolation, naming, and policy scoping
- **Node:** A worker machine (VM or physical) where Pods run
- **Custom Resource Definition (CRD):** A way to define new Kubernetes resource kinds (schema, validation, versions) that appear in the API like native types
- **Controller:** A control loop that watches a Kubernetes resource (any kind – built-in or custom) and reconciles the real cluster state toward the resource's desired state; it continuously: *observes* → *compares* → *acts*
- **Custom Resource (CR):** An instance of a CRD (your desired state for that new kind)

Kubernetes Building Blocks – Node



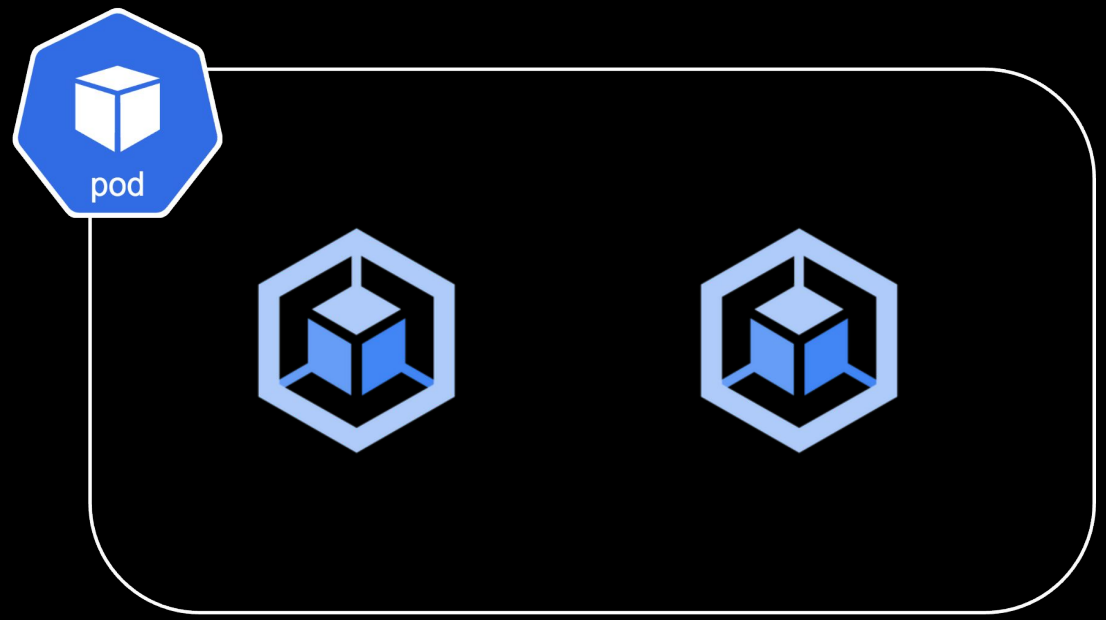
📦 Where pods actually run

© 2025 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Kubernetes Building Blocks – Pod



- 🧩 Shared Environment
- 🤝 Tight Coupling
- ⚙️ Ephemeral

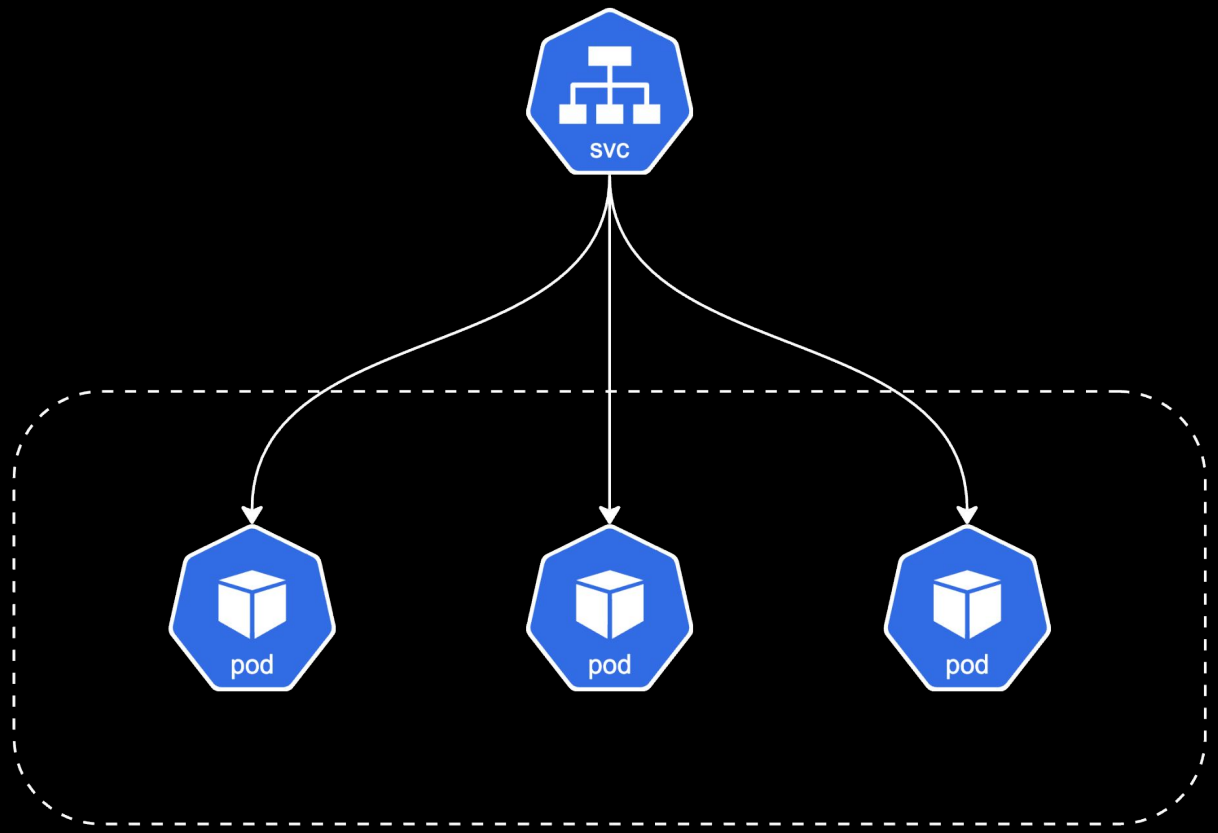
TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Kubernetes Building Blocks – Service



- 🎯 Stable Access Point
- ⚖️ Load Balancing
- 🧭 Service Discovery
- 🔒 Decoupling

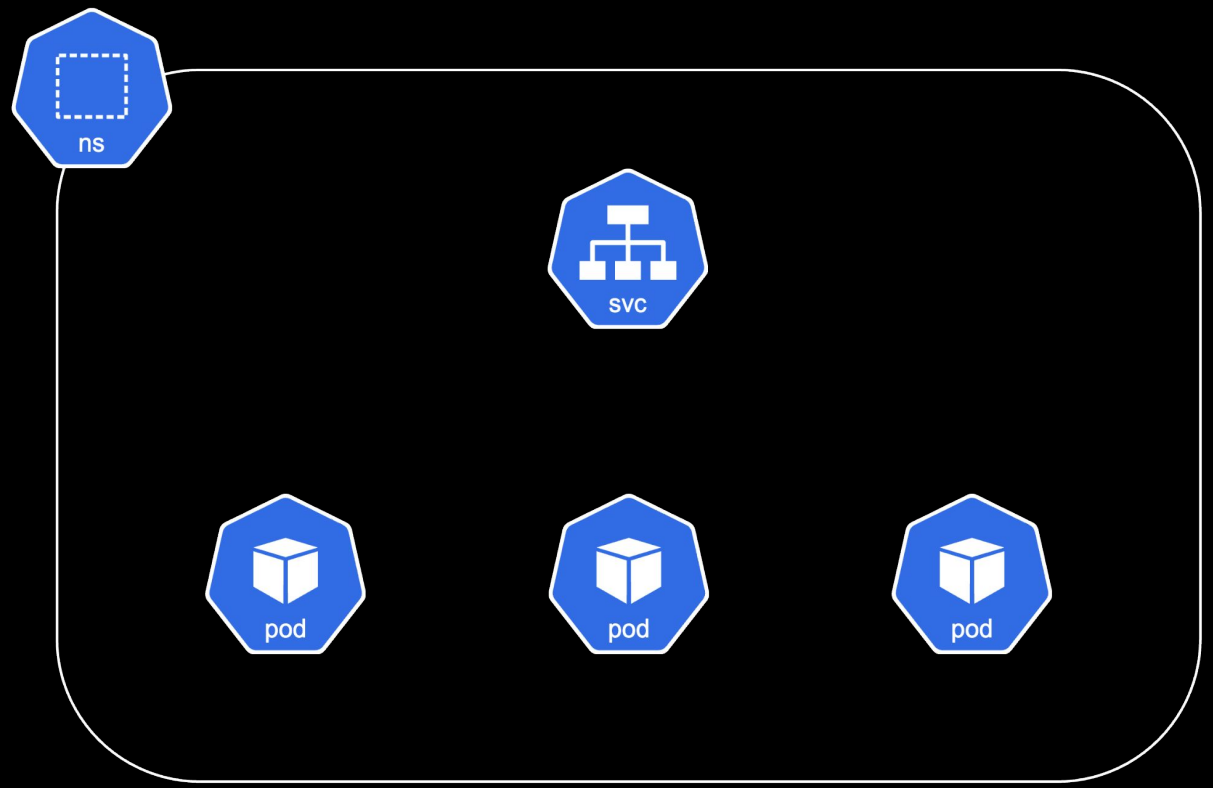
TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Kubernetes Building Blocks – Namespace



- 📦 Logical Partition
- 🔒 Isolation
- 🧭 Scoping
- 🧮 Multi-tenancy

TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Kubernetes Building Blocks – Custom Resource Definition (CRD)



JupyterNotebook

🧱 Built-ins: Pods, Services, Deployments, etc.

🧩 CRDs: Define new resources

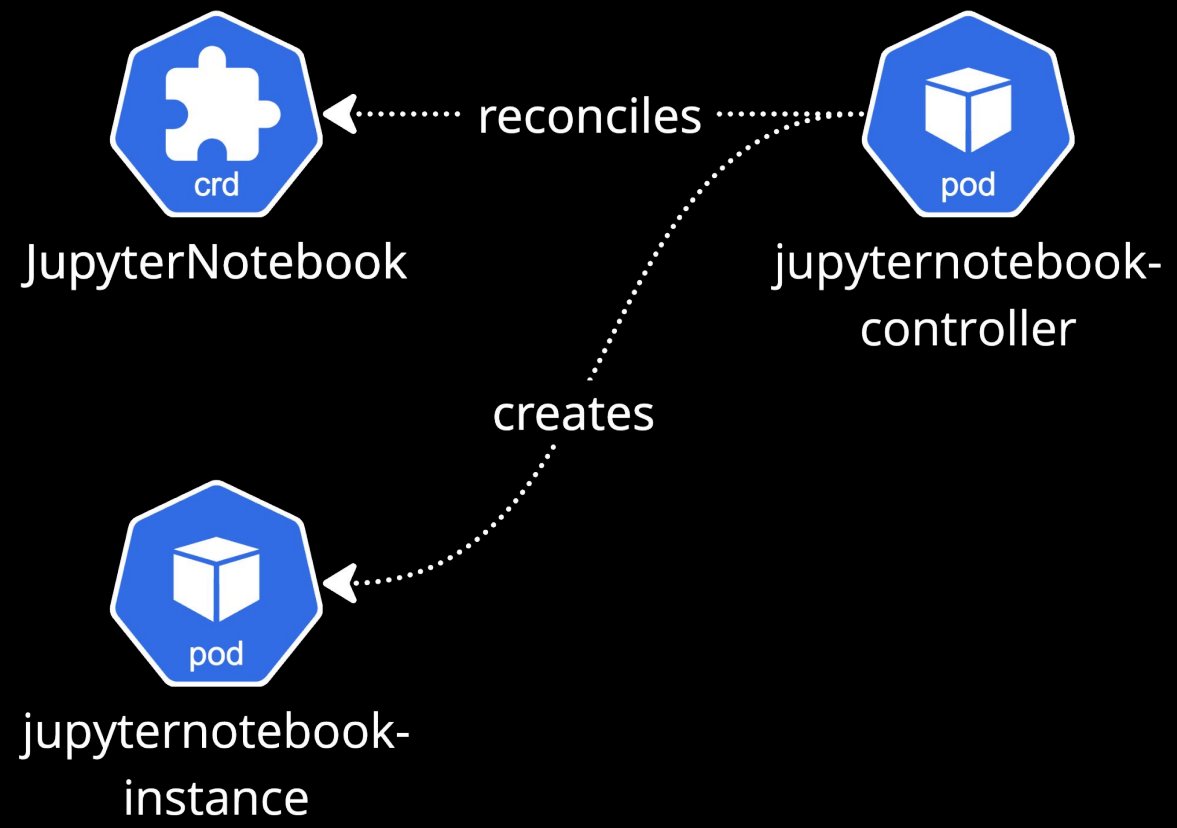
💡 Once created, they behave like native objects

© 2025 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Kubernetes Building Blocks – Controller



🔁 Reconciliation Loop

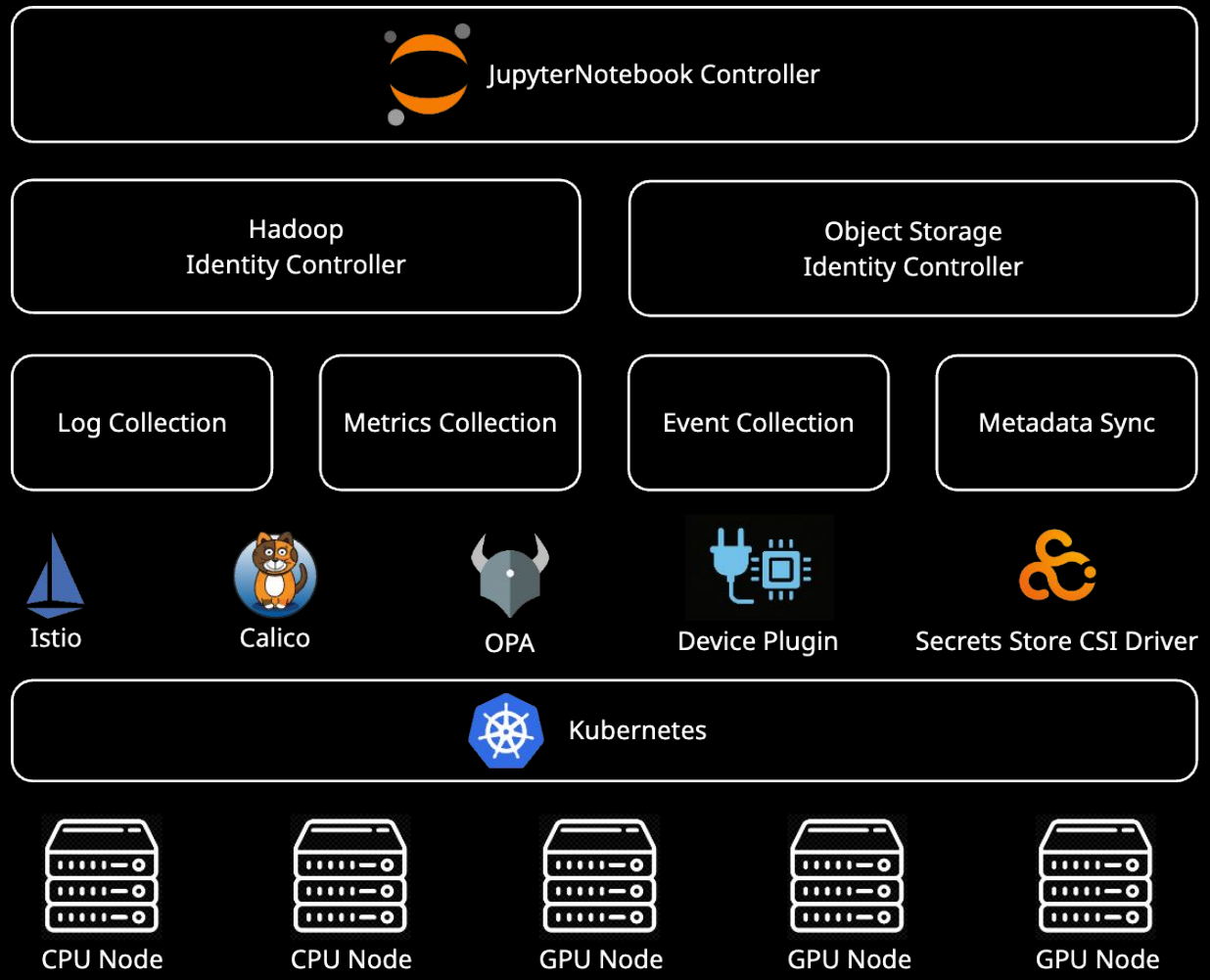
🧠 Declarative Model

© 2025 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Jupyter Notebooks as a Service at Bloomberg



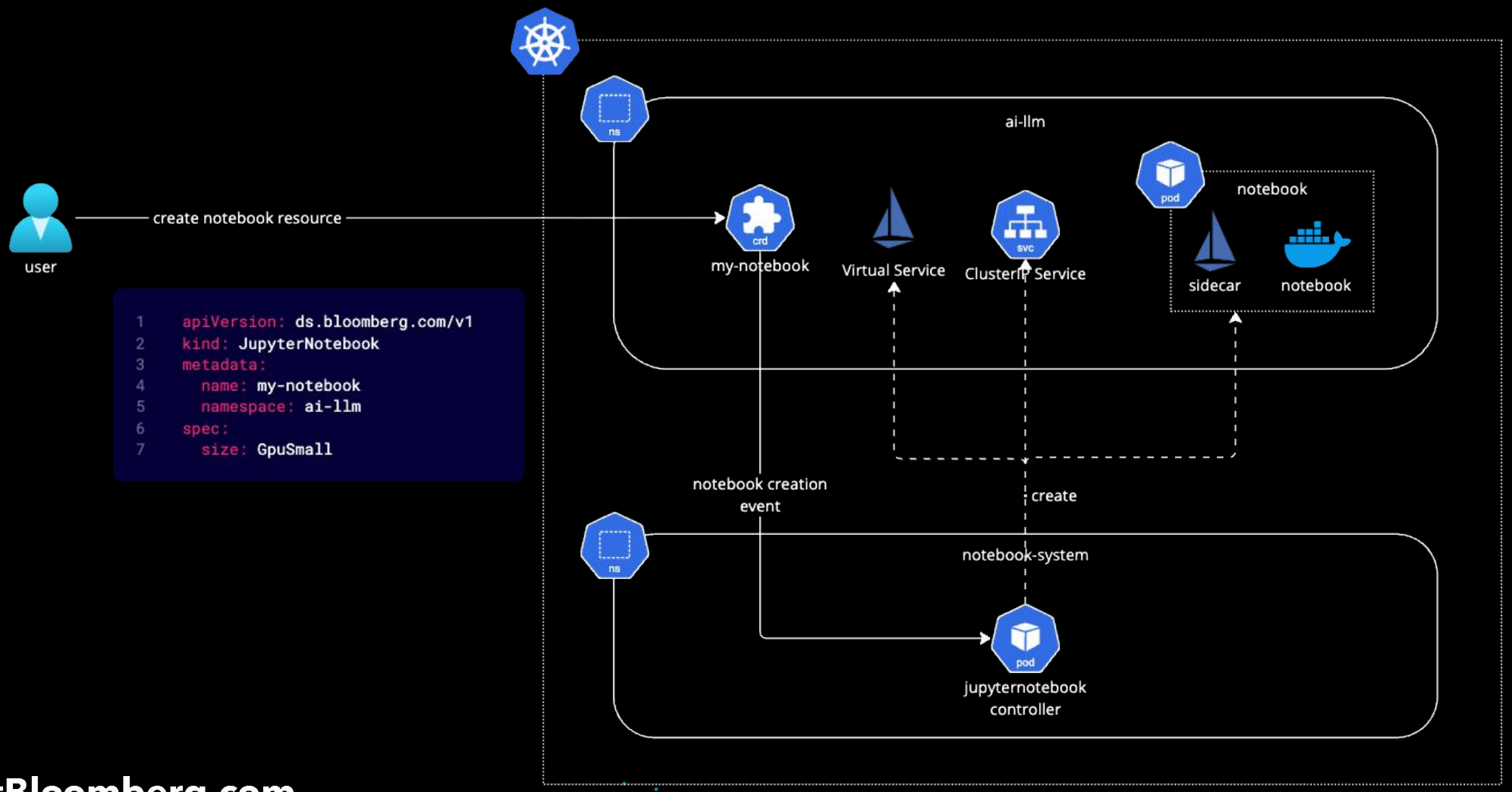
TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

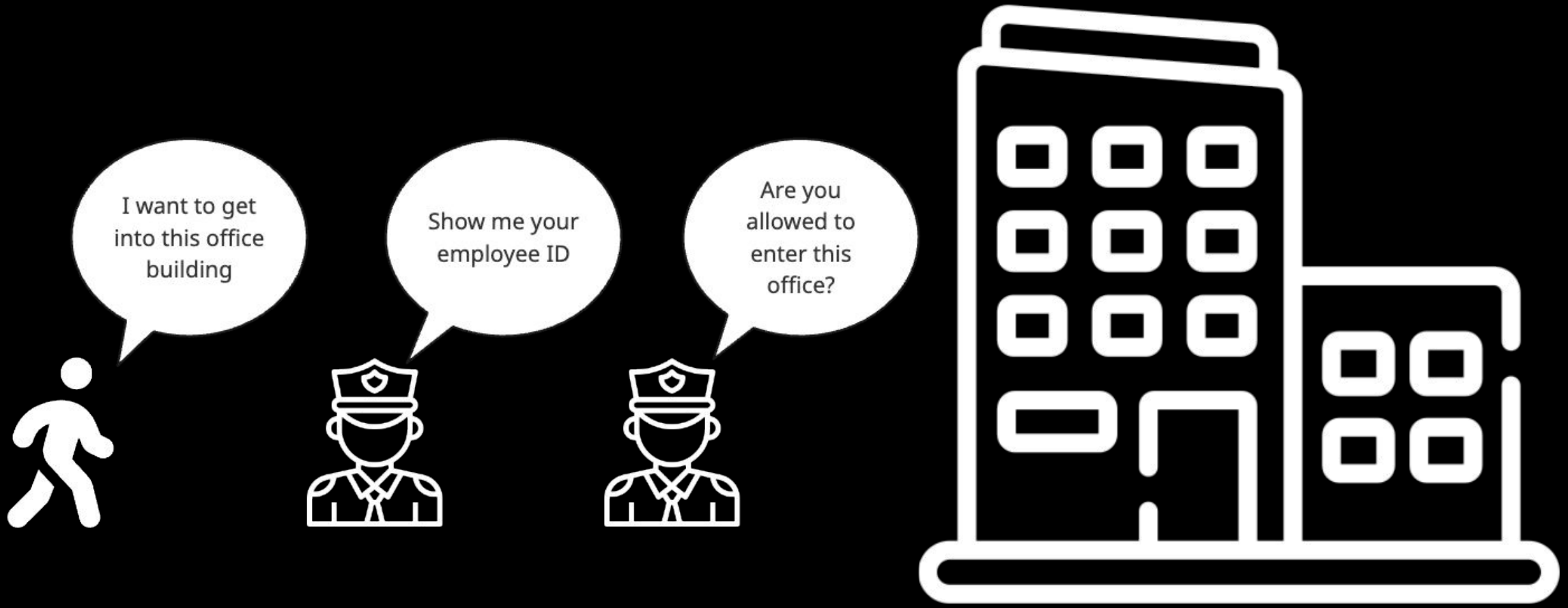
What Happens When You Launch a Notebook?



TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

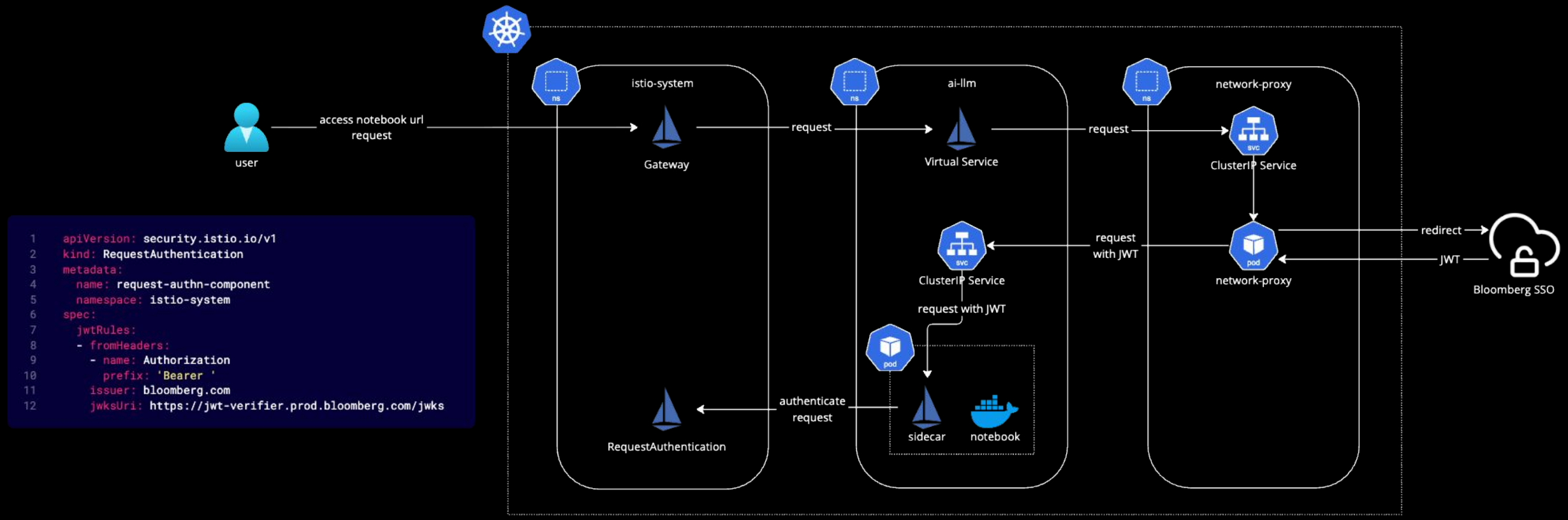
Ingress – Who Can Enter?



TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Ingress – Authentication

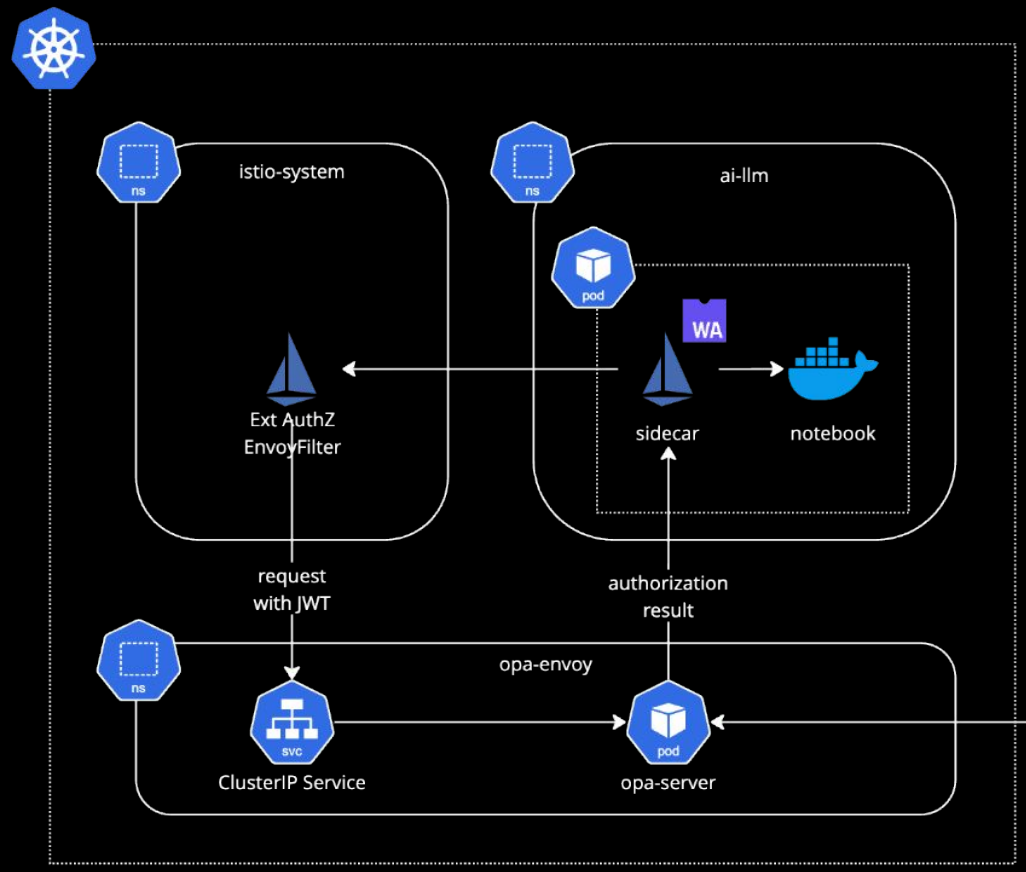


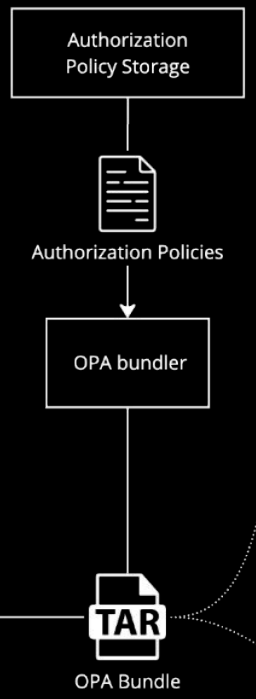

```yaml
apiVersion: security.istio.io/v1
kind: RequestAuthentication
metadata:
  name: request-authn-component
  namespace: istio-system
spec:
  jwtRules:
  - fromHeaders:
    - name: Authorization
      prefix: 'Bearer '
    issuer: bloomberg.com
    jwksUri: https://jwt-verifier.prod.bloomberg.com/jwks
```

TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Ingress – Authorization






```yaml
principal:
    ldap_group:notebook_user
action:
    access
resource:
    cluster:notebook-cluster:namespace:ai-llm:*
```

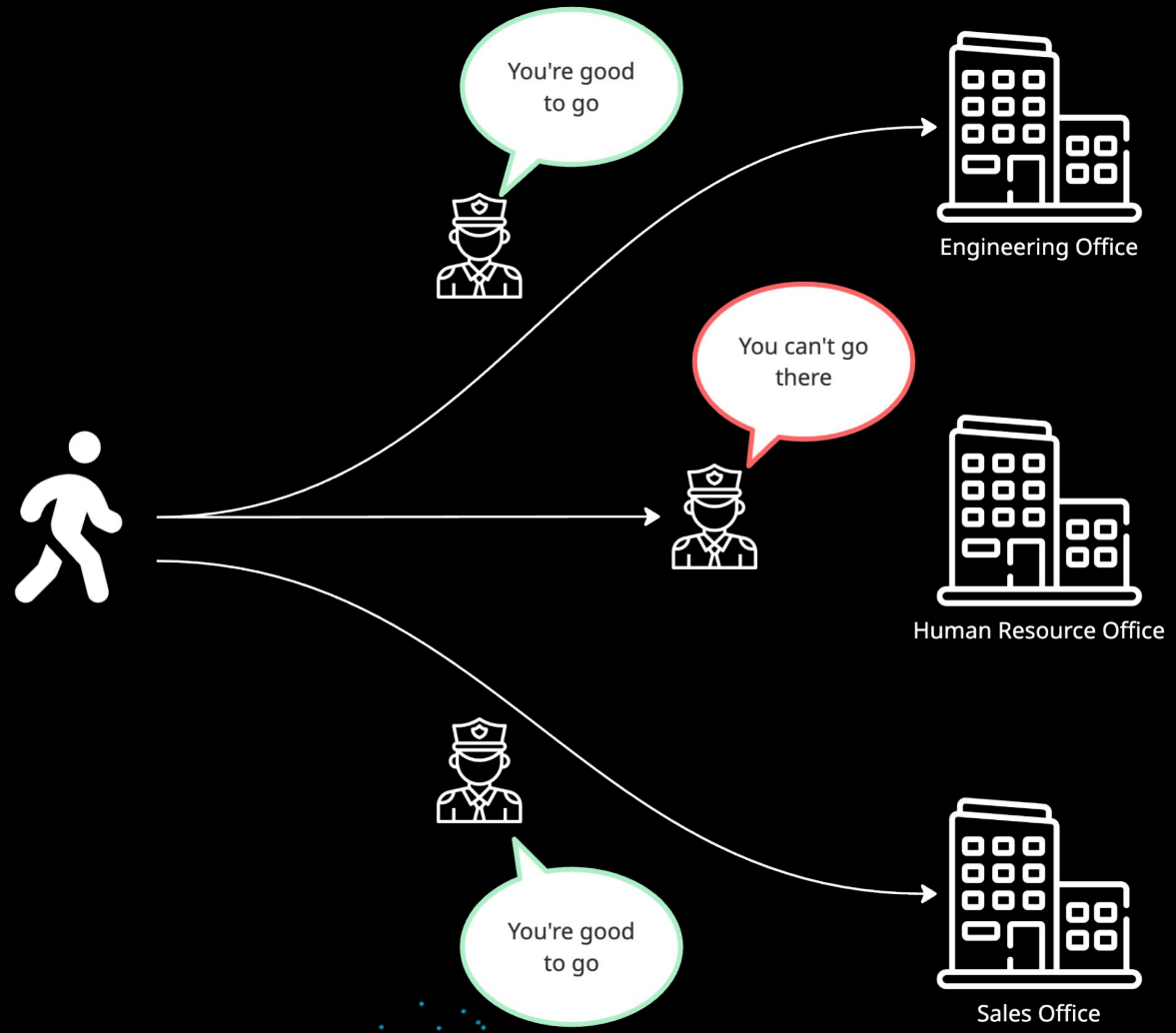
```
"ldap_user:yifan": {
    "access": [
        "cluster:notebook-cluster:namespace:ai-llm:*"
        ]
}
```

```
default allow = false
allow {
    policies := authorized_users
    principal := http_request.headers.principal
    action := http_request.headers.action
    resource := http_request.headers.resource
    some r
    glob.match(policies[principal][action][r], [":"], resource)
}
```

TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

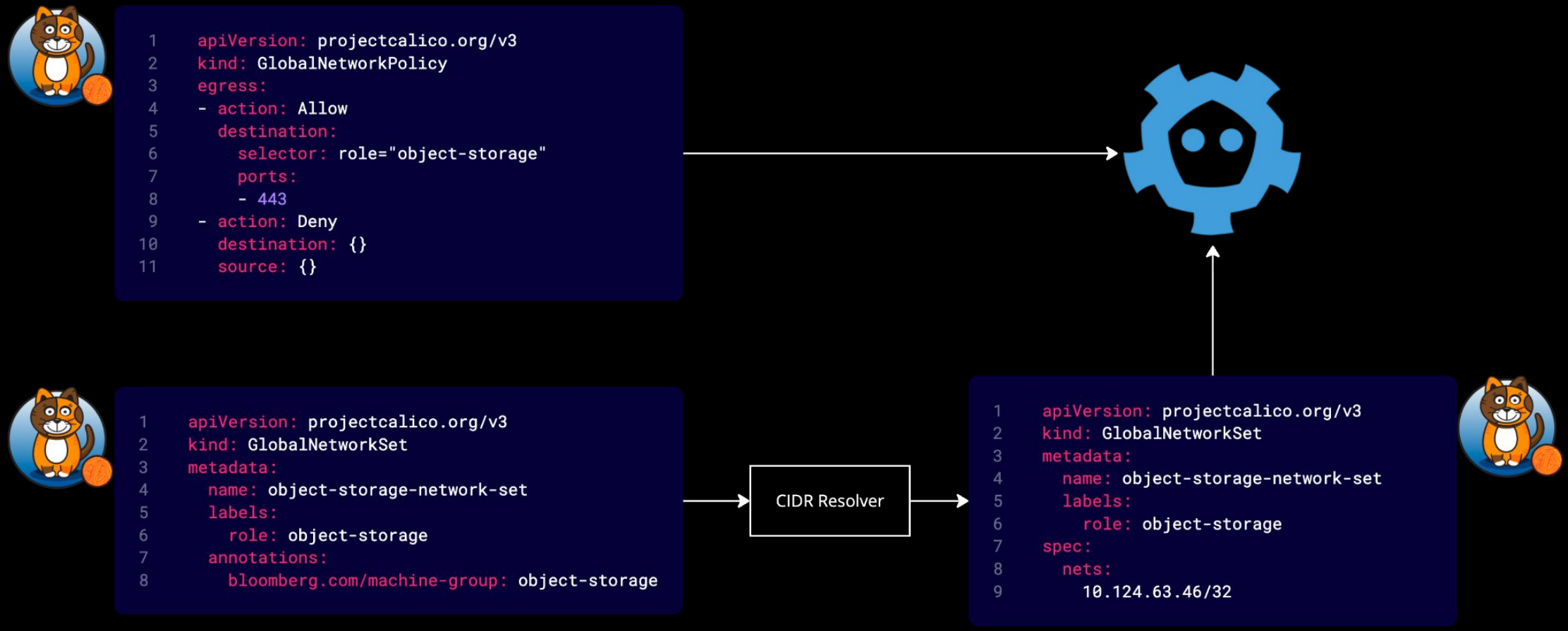
Egress – Where You Can Go Once Inside?



TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Egress



```yaml
apiVersion: projectcalico.org/v3
kind: GlobalNetworkPolicy
egress:
- action: Allow
  destination:
    selector: role="object-storage"
    ports:
    - 443
- action: Deny
  destination: {}
  source: {}
```

```yaml
apiVersion: projectcalico.org/v3
kind: GlobalNetworkSet
metadata:
  name: object-storage-network-set
  labels:
    role: object-storage
  annotations:
    bloomberg.com/machine-group: object-storage
```

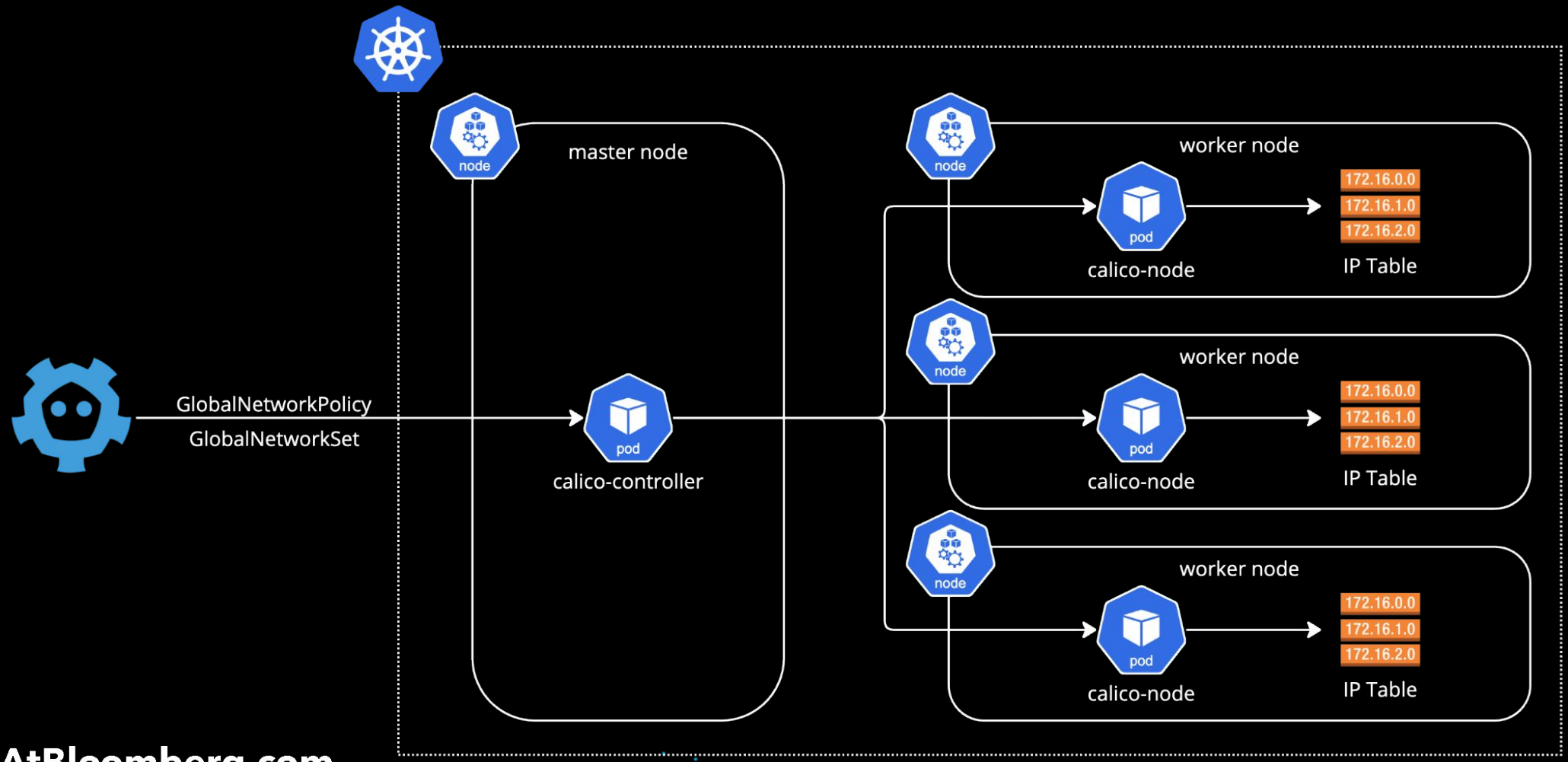


```yaml
apiVersion: projectcalico.org/v3
kind: GlobalNetworkSet
metadata:
  name: object-storage-network-set
  labels:
    role: object-storage
spec:
  nets:
    10.124.63.46/32
```

TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

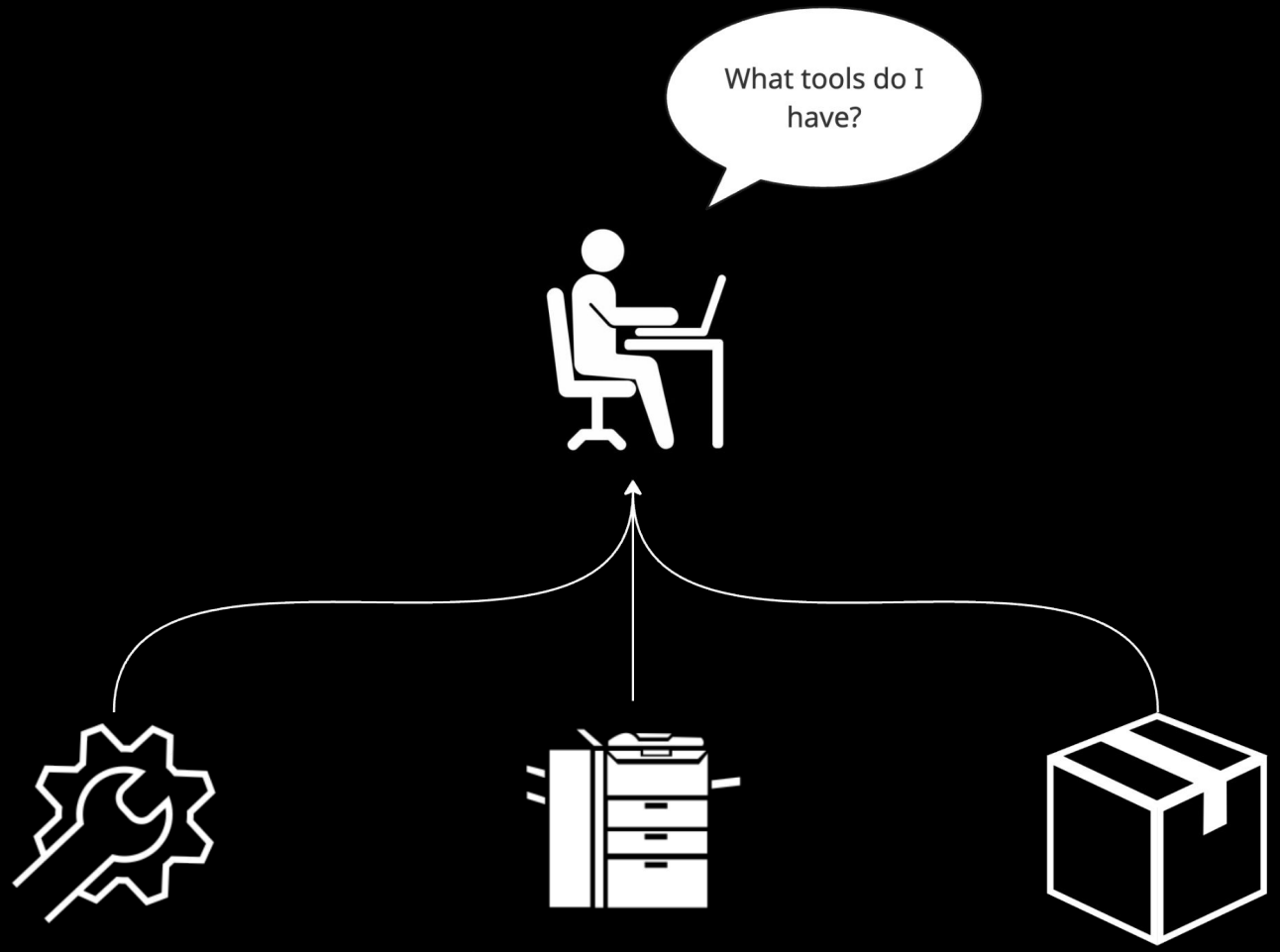
Egress



TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Developer Experience – What Tools Do I Have at my Desk?



TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Git Credential Injection

🎯 Goal

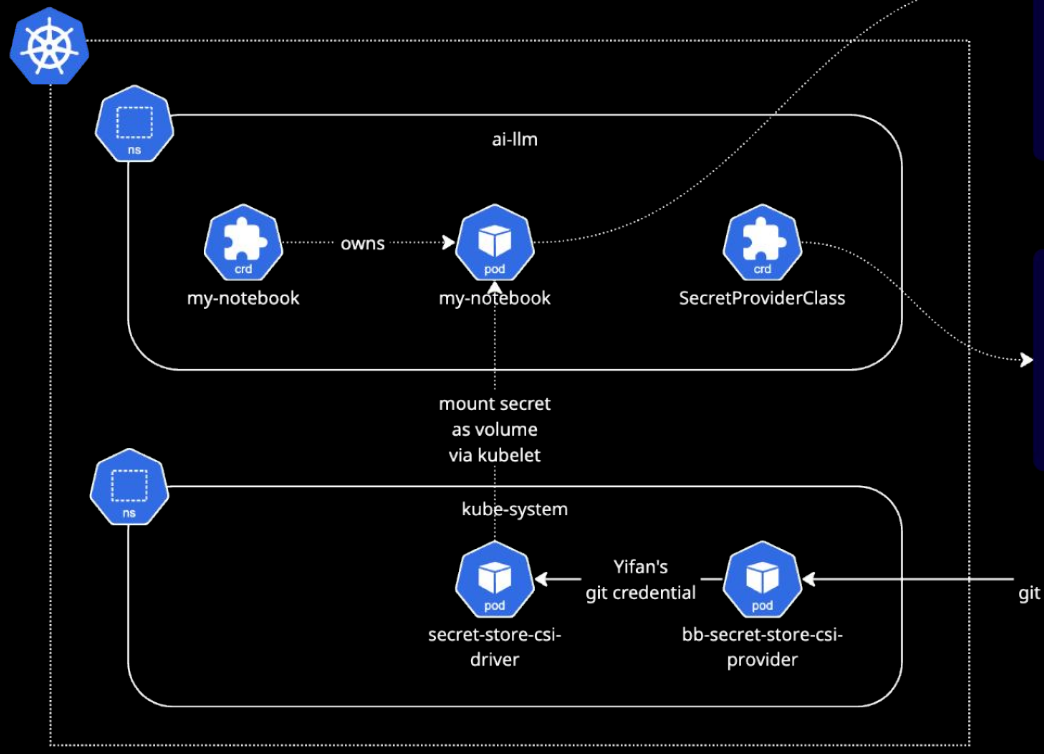
Provide notebook users with seamless access to internal commercial version control system – no manual credential setup required

🔐 How We Achieve It

- Secret Store CSI Driver
- Internal Secret Store CSI Provider

© 2025 Bloomberg Finance L.P. All rights reserved.

Git Credential Injection




```yaml
apiVersion: v1
kind: Pod
metadata:
  name: my-notebook
  annotations:
    createdBy: yifan
spec:
  volumes:
  - csi:
      driver: secrets-store.csi.k8s-io
      readOnly: true
      volumeAtributes:
        secretProviderClass: yifan-notebook-secret-provider-class
```

refers to

```yaml
apiVersion: secrets-store.csi.x-k8s.io/v1
kind: SecretProviderClass
metadata:
  name: yifan-notebook-secret-provider-class
spec:
  provider: bb-secret-store
  options: '{"secrets":[{"path":"notebook-cluster/ai-llm/git-credential-yifan"}]}'
```

```
notebook-cluster
- ai-llm
  - git-credential-yifan
```

TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Remote Storage Access

🎯 Goal

Enable notebook users to securely and effortlessly access company-wide storage systems – including HDFS, object storage, and other internal data services

🔐 How We Achieve It

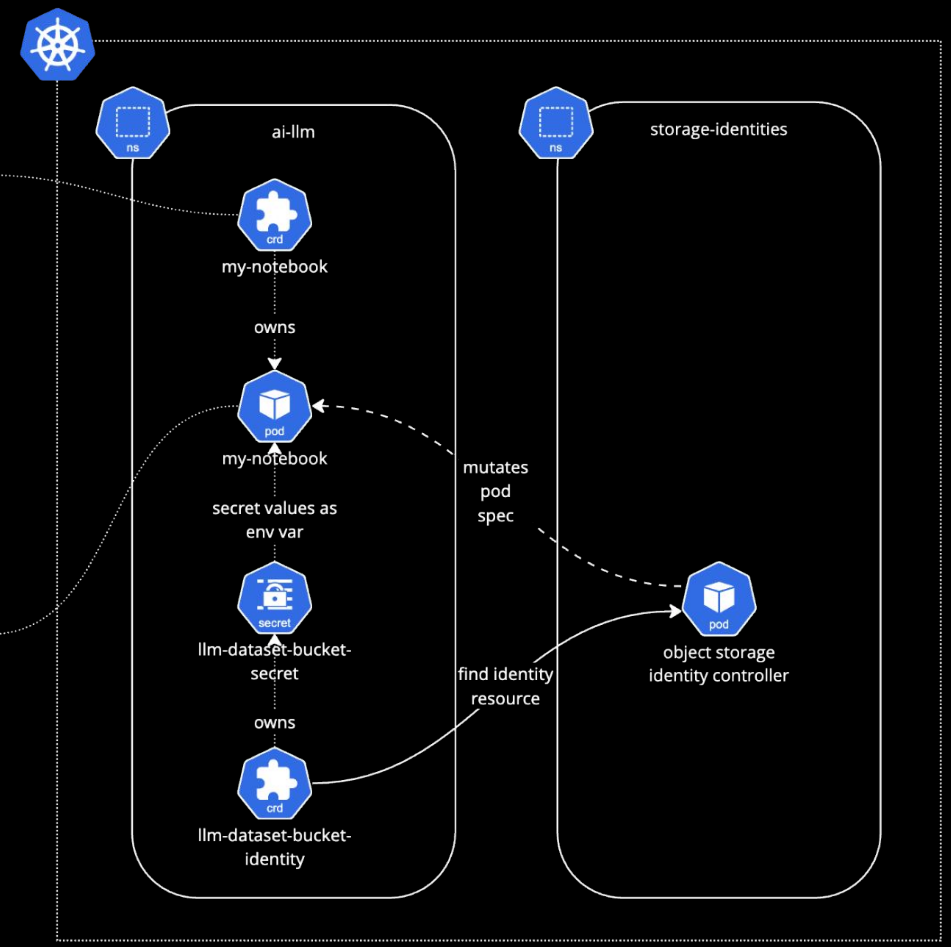
Use a combination of Kubernetes CRDs and Secrets to define, configure, and inject storage connection credentials directly into notebook Pods

© 2025 Bloomberg Finance L.P. All rights reserved.

Remote Storage Access – Object Storage

```yaml
apiVersion: ds.bloomberg.com/v1
kind: JupyterNotebook
metadata:
  name: my-notebook
  namespace: ai-llm
spec:
  size: GpuSmall
  identities:
  - object-storage:
      id: llm-dataset-bucket
```

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: my-notebook
  namespace: ai-llm
  annotations:
    identities.bloomberg.com/object-storage: llm-dataset-bucket
spec:
  ...
  containers:
  - name: job
    env:
    - name: AWS_ACCESS_KEY_ID
      valueFrom:
        secretKeyRef:
          key: llm-datatset-bucket-secret
          name: access_key_id
    - name: AWS_SECRET_ACCESS_KEY
      valueFrom:
        secretKeyRef:
          key: llm-datatset-bucket-secret
          name: secret_access_key
```



TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Operational Challenges – Keep the Building Running





TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Zombie Process

🧩 Problem

- Alerts showed some machines filling up with hundreds of zombie processes

🔍 Observation

- The issue appeared mostly in long-running notebook instances
- All zombies were children of the Jupyter process
- They disappeared once the Jupyter container exited

© 2025 Bloomberg Finance L.P. All rights reserved.

Zombie Process



```
exec ${JUPYTER_HOME}/.venv/bin/jupyter lab \
  --ServerApp.password='' \
  --ContentsManager.root_dir=${NOTEBOOKS_DIR} \
  --LabApp.default_url=${JUPYTER_DEFAULT_URL} \
  ${JUPYTER_ARGS}
```

TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Credential Leakage in Logs

🧩 Problem

- Object storage credentials were appearing as plain text in platform logs

🔍 Observation

- Credentials appeared in startup logs before the notebook server launched
- Only notebooks using S3 content URI were affected

© 2025 Bloomberg Finance L.P. All rights reserved.

Credential Leakage in Logs

```
set -x
```



```
...

JUPYTER_ARGS="${JUPYTER_ARGS} --S3ContentsManager.access_key_id=${ AWS_ACCESS_KEY_ID }"
JUPYTER_ARGS="${JUPYTER_ARGS} --S3ContentsManager.secret_access_key=${ AWS_SECRET_ACCESS_KEY }"
...

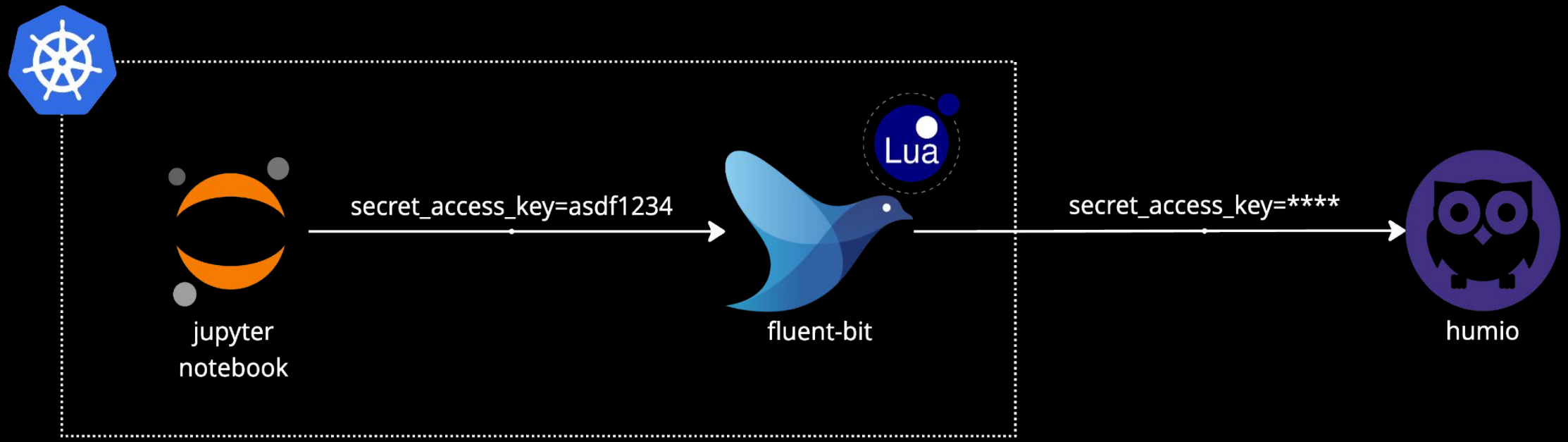
${JUPYTER_HOME}/.venv/bin/jupyter lab \
 --ServerApp.password='' \
 --ContentsManager.root_dir=${ NOTEBOOKS_DIR} \
 --LabApp.default_url=${ JUPYTER_DEFAULT_URL } \
 ${JUPYTER_ARGS}
```

TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Credential Leakage in Logs

- Removed `set -x`
- Implemented a fluent-bit lua filter to mask sensitive information



TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

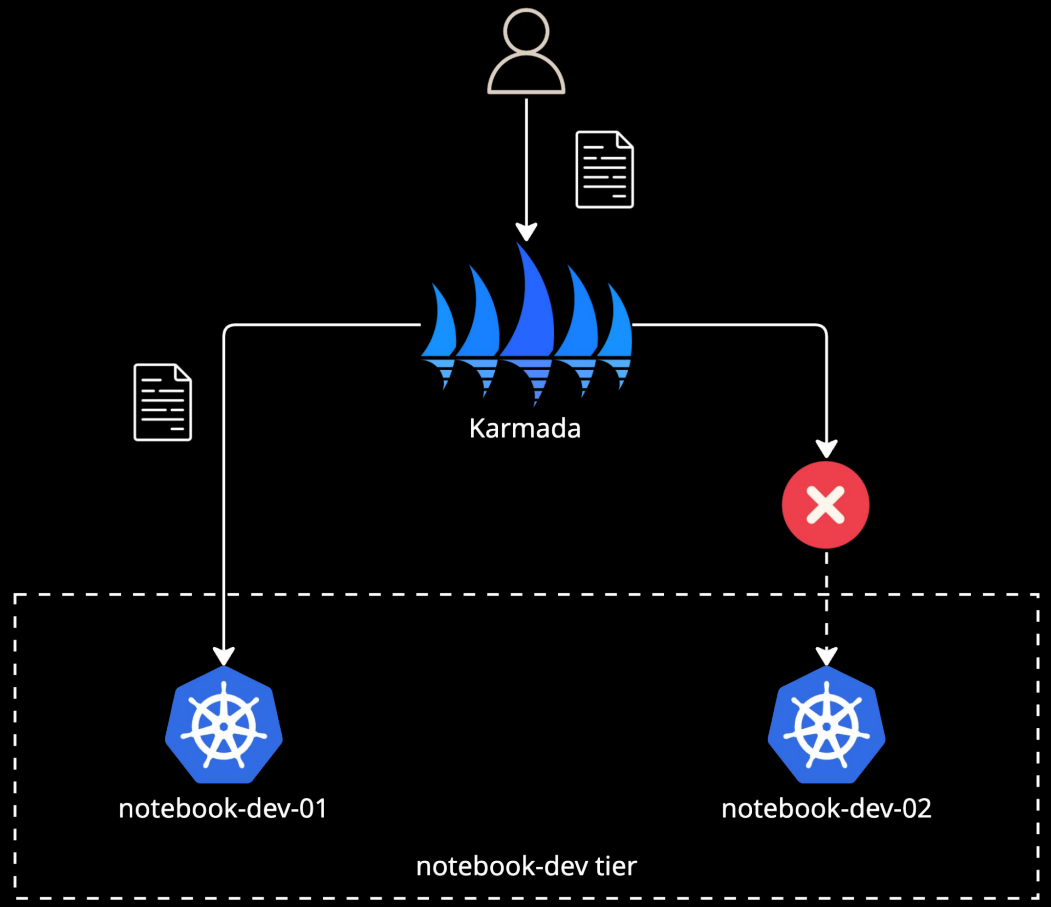
Federated Notebooks

🧩 Problem

- Infrastructure changes at the cluster level were visible to users, causing interruptions during maintenance or upgrades
- No automatic failover – if one cluster went down, users had to manually select other available clusters to launch new notebooks
- Uneven hardware utilization – some clusters were overloaded while others sat idle

© 2025 Bloomberg Finance L.P. All rights reserved.

Federated Notebooks



🔍 Transparent infrastructure change

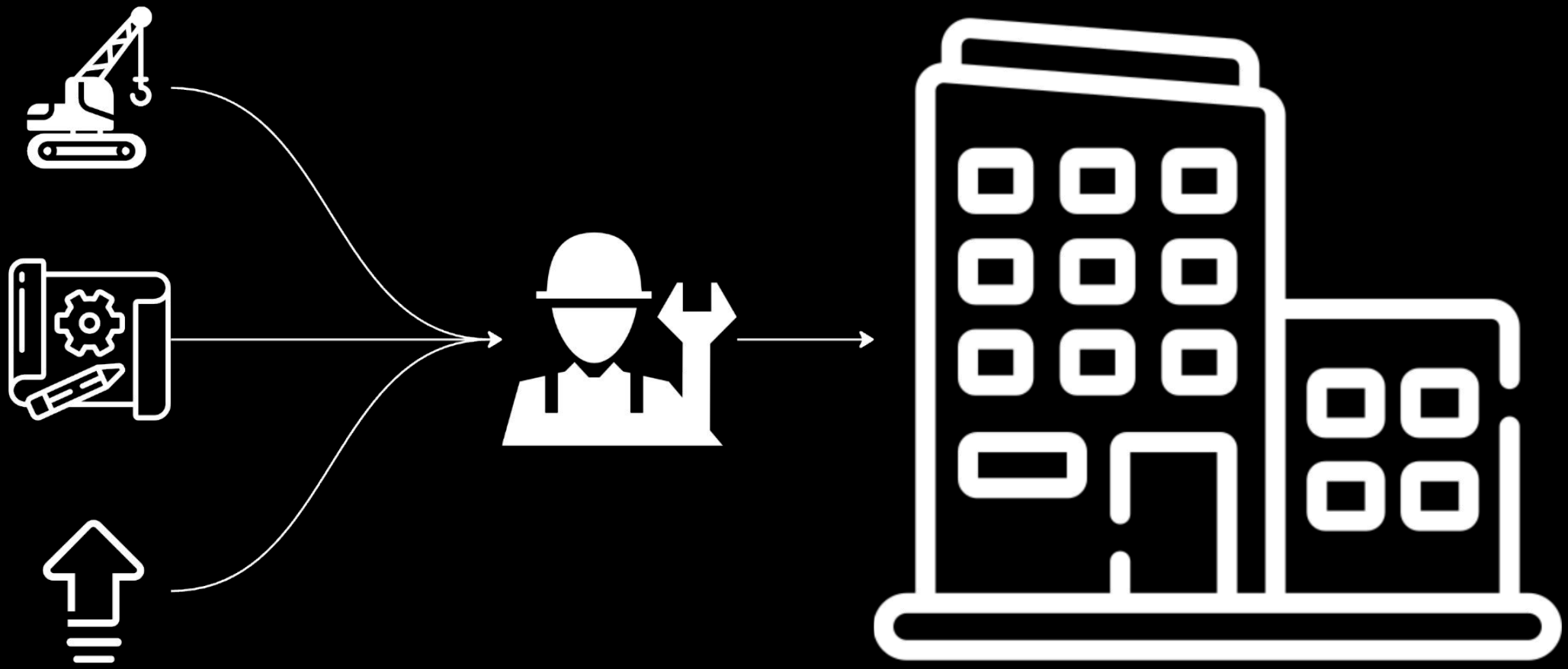
⚙️ Automatic failover

💪 High availability

TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

A Look Ahead

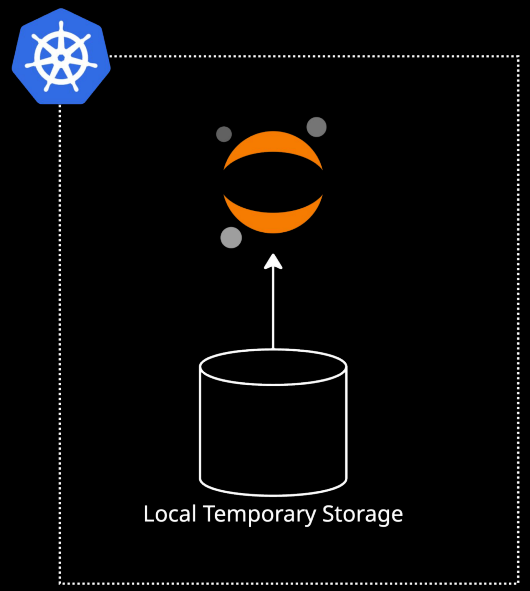


TechAtBloomberg.com

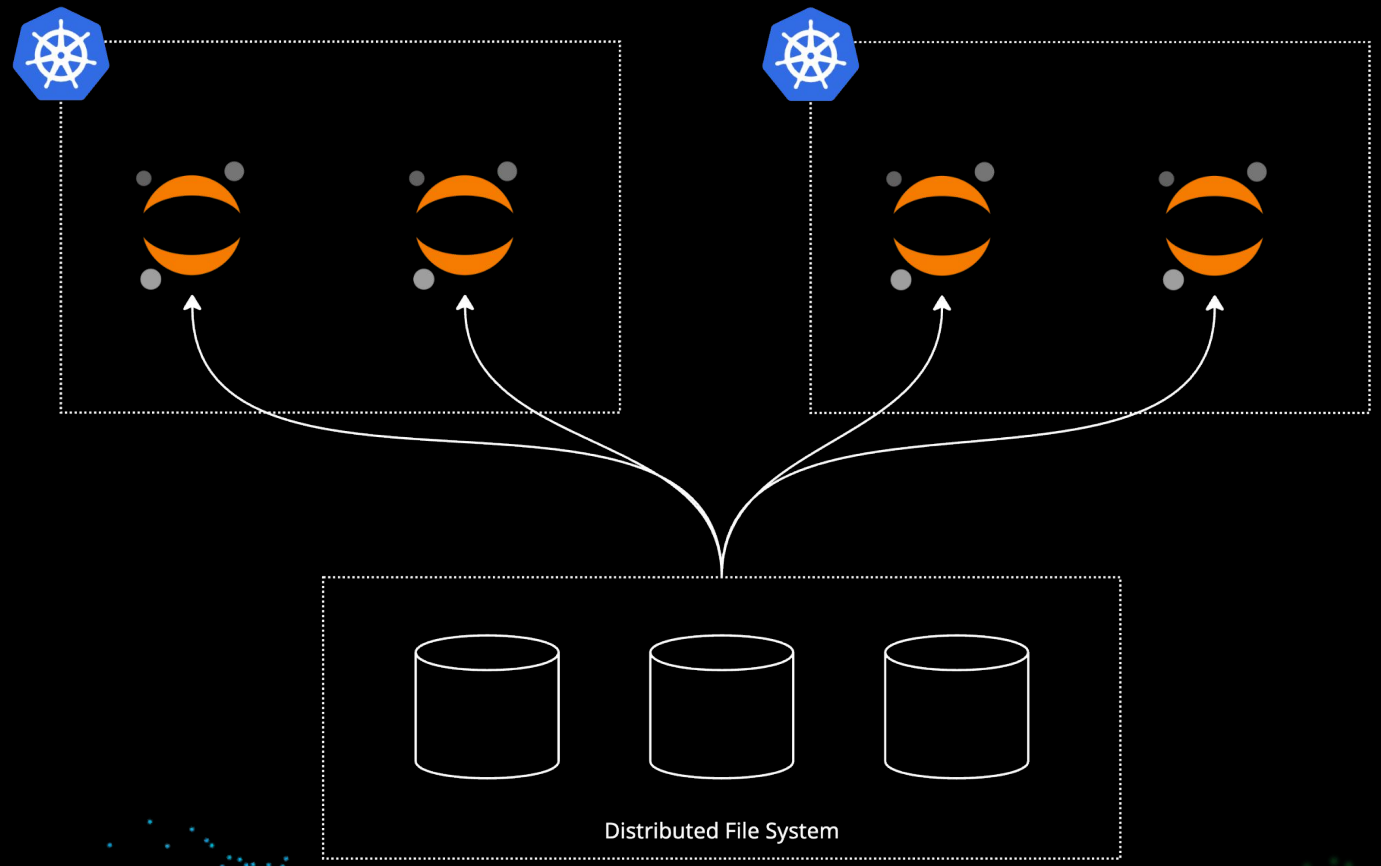
© 2025 Bloomberg Finance L.P. All rights reserved.

Distributed File System

Current State



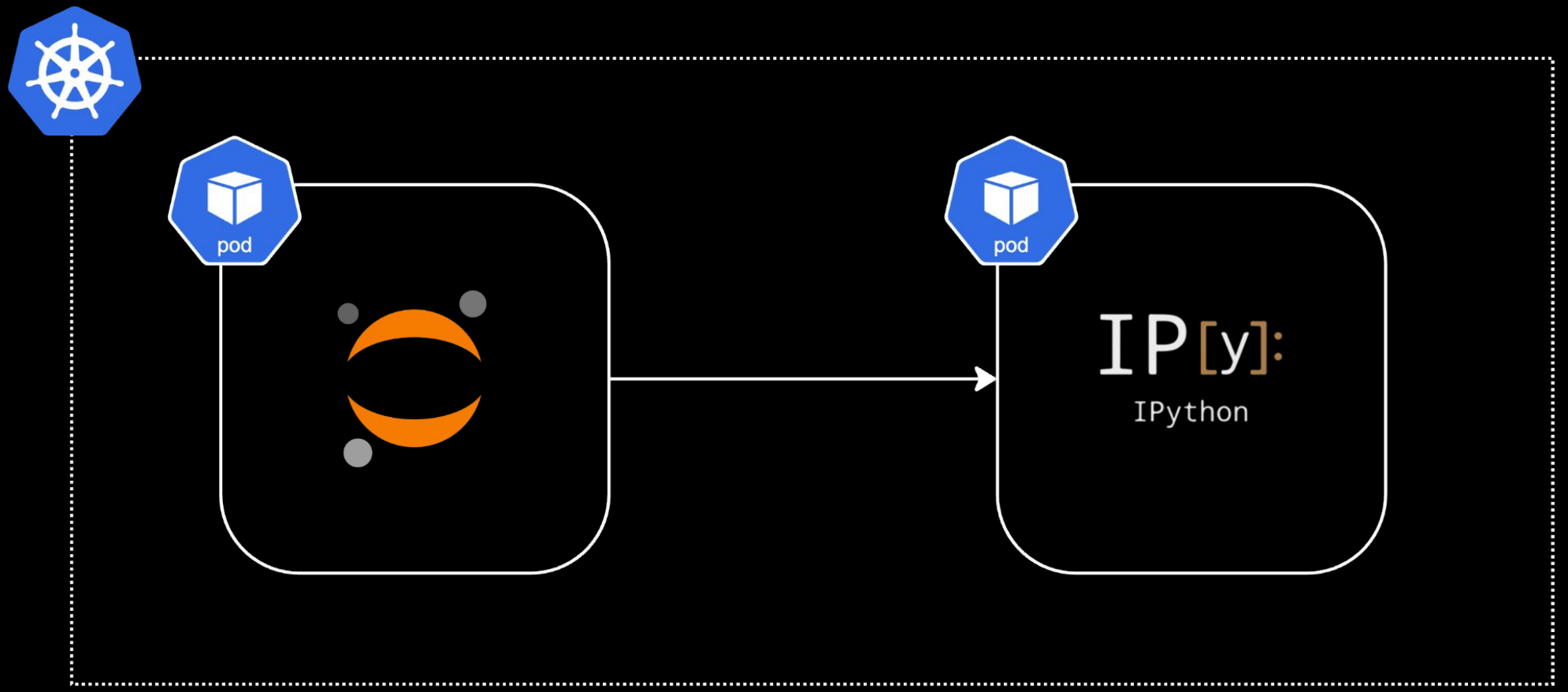
Future State



TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Server/Kernel Separation



TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.

Thank you!

Questions?

We are hiring: https://www.bloomberg.com/careers

https://bloomberg.com/company/values/tech-at-bloomberg/data-science/academic-engagement-programs/data-science-ph-d-fellowship/

Bloomberg

Engineering

TechAtBloomberg.com

© 2025 Bloomberg Finance L.P. All rights reserved.