ArgoCon NA 2024

Dog Food Delight:

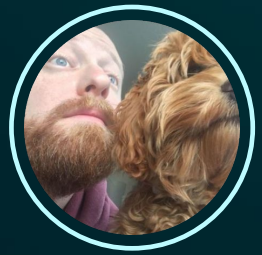
How Argo Workflows Eats Its Own CI

Tim Collins – Pipekit
Denise Schannon - Loft Labs

November 12, 2024

Intros

pipekit



Tim Collins

Staff Infrastructure Engineer @ Pipekit.io

- Argo Maintainer. Lives in the CNCF Argo Slack
  - Member of the Pipekit Services team
- The beard distracts you from his bald spot

loft

Denise Schannon

VP of Engineering @ Loft Labs

- Leads the Loft engineering team
  - Not meant to design slides

About Pipekit

Scale Argo & Kubernetes with Pipekit

- Direct support from 40% of the global active Argo Workflows maintainers
- Save engineering time and up to 60% on compute costs
- Add 3 Argo maintainers and 7 Argo contributors to your team
- Serving startups & Fortune 500 enterprises since 2021:

> **Enterprise Support for Argo:**
> Ideal for Platform Eng teams scaling with Argo

> **Control Plane for Argo Workflows:**
> Ideal for data teams, granular RBAC, and multi-cluster architectures

About Loft Labs

vCluster made by loft

Simplify Kubernetes with Loft Labs

- Kubernetes experts building tools for platform engineers
- Creators of open-sources projects:
  - vCluster, DevPod, DevSpace and JsPolicy
- 17.3k Stars on GitHub and 3.6k Slack members
- $24m Series A in 2024 with a team of 40+

Backed by khosla ventures

Fusion Fund · SURFACE VENTURES · EMERGENT VENTURES · Berkeley SKYDECK

Dogfooding: Why do we want to do this?

1

Improve the DevEx that upstream currently provides

- The Github actions don't run on forked copies without manual intervention
  - Put control of testing back onto the fork
  - Decrease the noise in upstream PRs

2

Test the container images and Kubernetes RBAC

- Modify the test framework so that we test the artifacts we intend to release.
- Run tests as they would operate in real usage.

3

Demonstrate that Argo Workflows can be used for CI

- Use the migration as a learning opportunity

Our goal

- Use Argo Workflows to drive CI for Argo Workflows
  - Build and push Argo Workflows images
  - Run Unit tests
  - Run End to End tests
- Don’t modify the existing tests
- Keep CI separate from the upstream branch
- Run the tests in real K8s clusters, using the built images, in parallel, as fast/reliably as possible.
- Stick to Open Source solutions

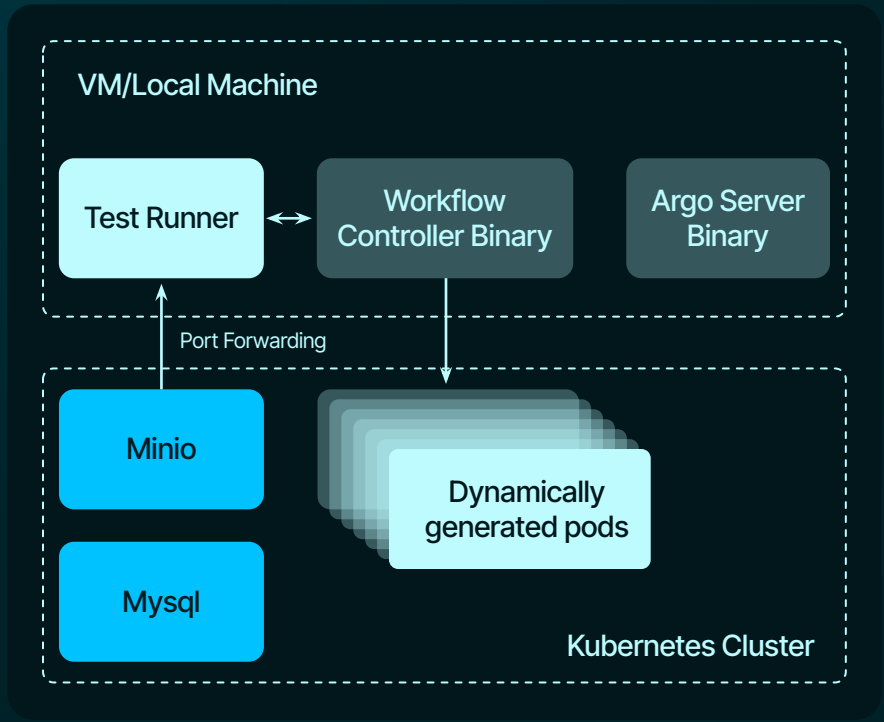
The End-to-end test suite

Upstream CI

- Runs Argo binaries outside of Kubernetes
  - Not like the real-world
  - Binaries have admin permission on the cluster

Local testing

- Slow
- 1 Kubernetes cluster
- Not the same coverage



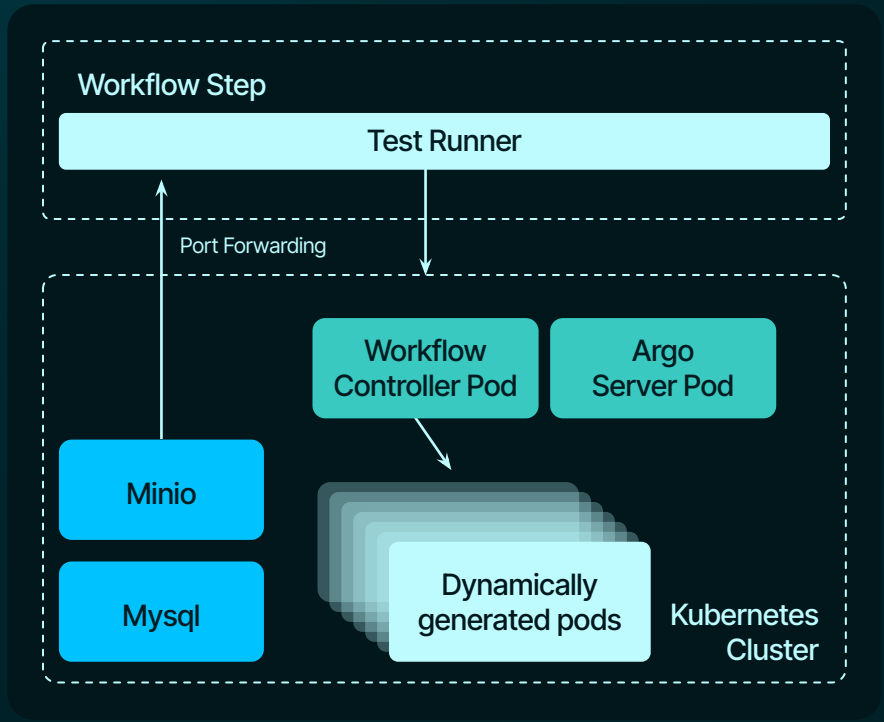
What do we want?

Testing the images that will ultimately be released by running them in a K8s cluster as they would in the wild

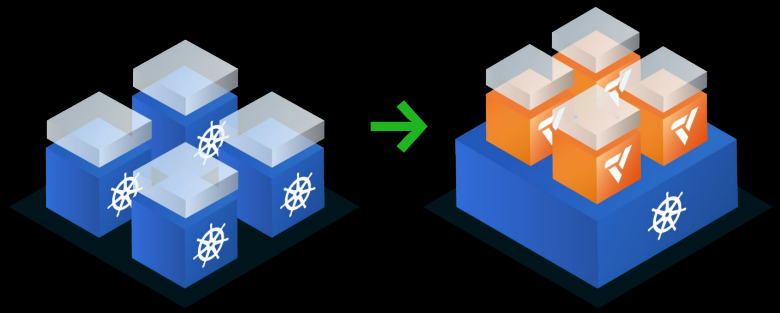
- Kubernetes RBAC that represents real world usage

An external Argo Workflows to orchestrate the tests

Match Github Actions with 10 Kubernetes clusters running at the same time allowing us to test in parallel



= Virtual Clusters



Virtual clusters run as containers
inside namespaces of a “real” cluster

vCluster
made by loft

Virtual Kubernetes Clusters

vCluster is the only certified Kubernetes distro
for creating virtual Kubernetes clusters

6,500+
GitHub Stars

github.com/loft-sh/vcluster | www.vcluster.com

Multi-tenant clusters help consolidate infrastructure

Provisioning a separate K8s cluster for every team, customer & env is costly.

- Resource and financial waste
- Operational complexity and overhead
- Requires lots of fleet management tooling
- Takes 30+ minutes to provision

Enhance efficiency and security with Multi-Tenancy

- Quickly provision virtual clusters
- Stronger tenant isolation than namespaces
- Much cheaper than separate clusters
- Tenants are admins inside their virtual cluster

Maximize efficiency & cut cost with virtual clusters.

+50%
Cost Savings

Key Features

Sleep Mode

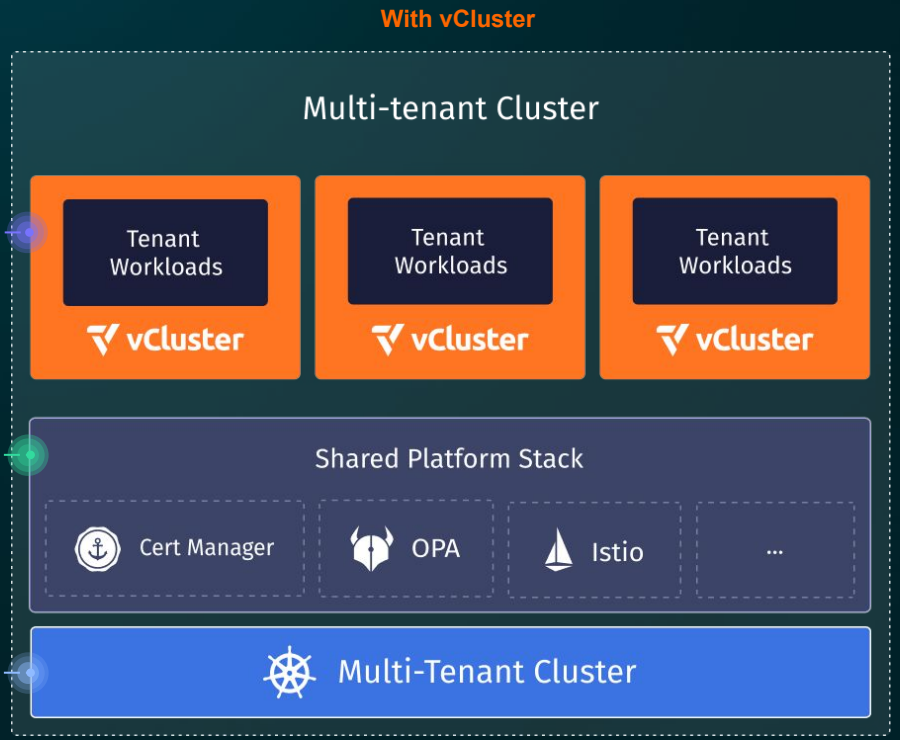
Reduce idle workloads cost by putting clusters to sleep when not in use.

Shared Platform Stack

Run fewer duplicates of tools like Istio, Cert Manager, OPA, Datadog and Vault.

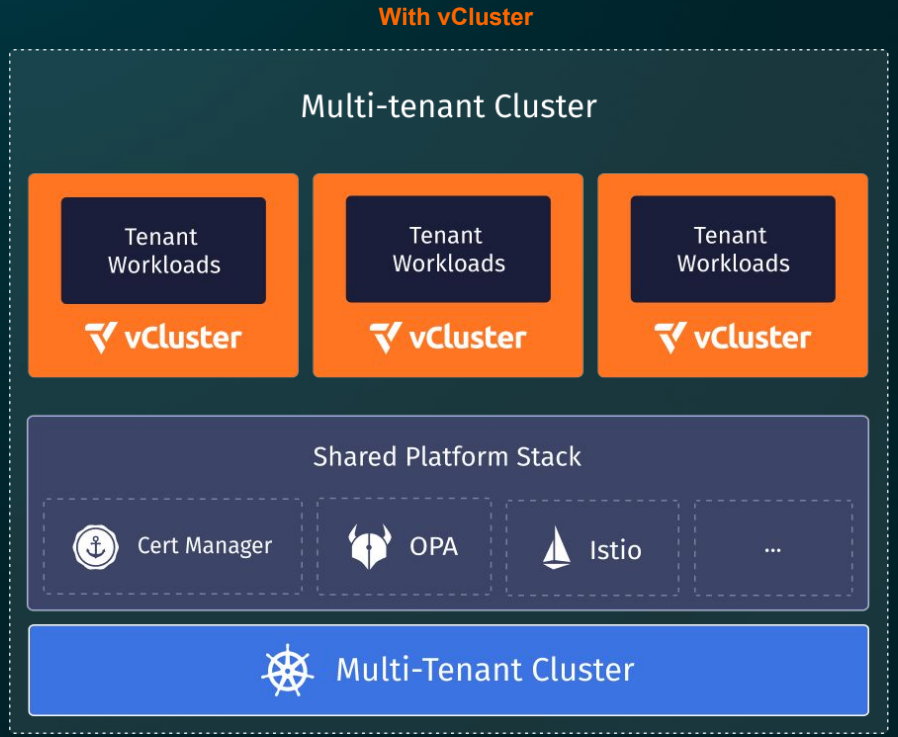
Multi-Tenant Clusters

Run fewer clusters by isolating multiple tenant workloads within a single cluster.



Why vCluster for this challenge?

- **Open-Source**: Certified K8s Cluster Distro
- **Faster**: Virtual clusters start in <30 seconds
- **Cheaper**: +50% Cost Savings
- **Secure**: Isolated Control Plane for each virtual cluster
- **Reliable + Production-Grade**: Battle-tested by 100s of companies



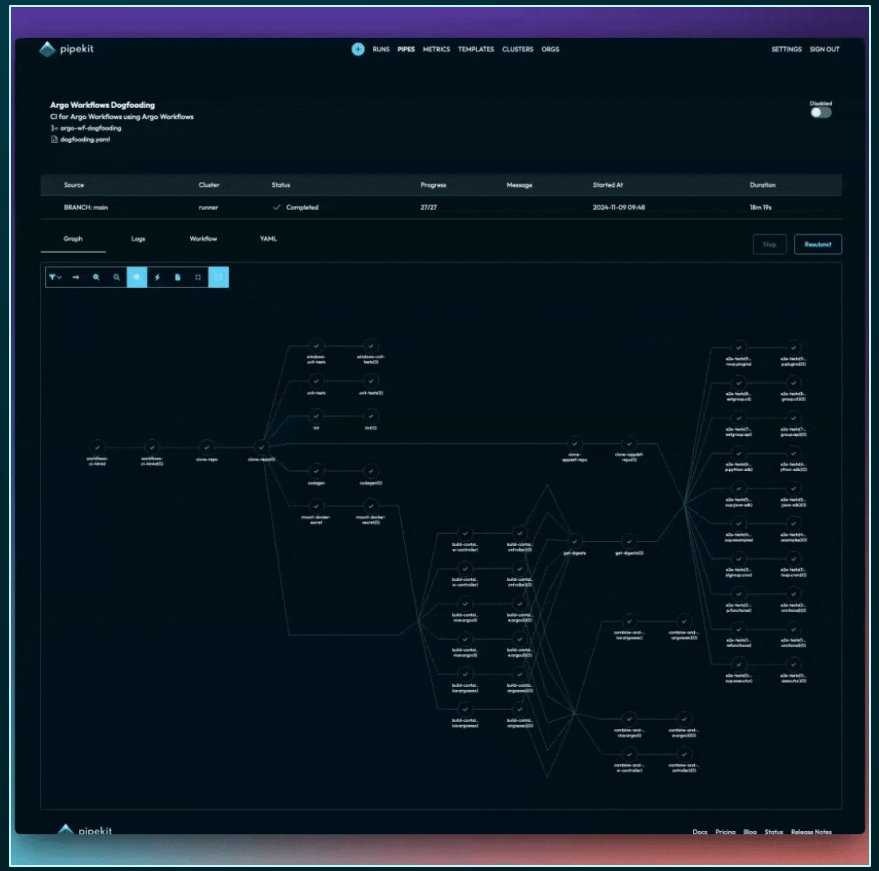
We did the thing

Run-time ~18-20 minutes

- Comparable to upstream CI (~20-25 mins)
- Faster than local test

Cost estimate

- ~$2.6 per run
  - Ondmd, AMD64 nodes, gp3 storage.



Learning - Infrastructure considerations

- How to efficiently pass data between steps?
- v3.5.x+ workflow steps hanging?
- Karpenter / Cluster Autoscaling
- Mastering Argo Workflows at Scale
- Observability
  - "Conntrack getting close to the limit"
  - Free Workflow Metrics by Pipekit
  - Cluster-wide logging solution
- Spegel - stateless cluster local OCI registry mirror
- Kube Janitor - clean things up based on a ttl

Bumpy road - Upstream

Crashing 3rd party tools (minio, mysql)

- Set resource requests on everything in your cluster.
- Have strong cluster observability

Uncovered issues that upstream tests weren't finding

Test flakiness

Next steps

Open source what we have done

Make it a viable option to replace Github Actions upstream

- Refactor framework for robustness.
  - Is there community interest?
  - Targeting 30% increase in speed
  - Targeting run cost below $1
- Hosting considerations

Free stuff!

Slide Deck:

github.com/pipekit/talk-demos

Free workflow metrics by Pipekit:

pipekit.io/metrics-signup

Office Hours (free Argo consultation!):

pipekit.io/office-hours

Free Argo/Infrastructure Help & Advice:

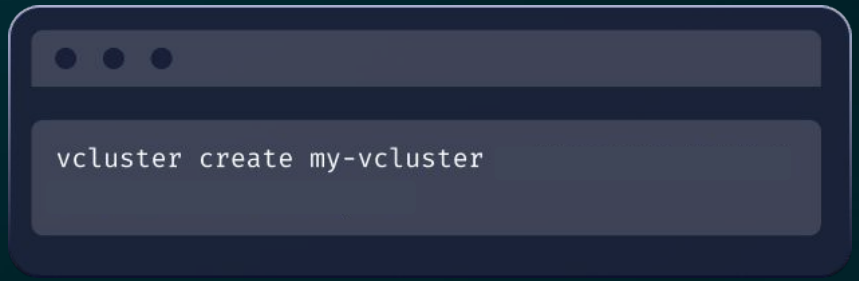
Booth T33

vCluster made by loft

Get Started For FREE @

vCluster.com/install

Or run:



Free vCluster Help & Advice:

Booth A6