



redhat.[®]

Extending Kubernetes 101

Michael Hausenblas **@mhausenblas**

Developer Advocate, Red Hat

2018-11-15, ContainerConf, Mannheim

\$ whois mhausenblas

- Developer Advocate @ Red Hat (Go, Kubernetes, OpenShift)
- Developer Advocate @ Mesosphere (Mesos, DC/OS, Kubernetes)
- Chief Data Engineer @ MapR (HDFS, HBase, Drill, etc.)
- Applied research (4y in Ireland, 7y in Austria)
- Nowadays mainly developing tools in Go (Python, Node, Java, C++)
- Kinda developer turned ops (aka appops)



admin

developer

architect



SRE

infosec

PM

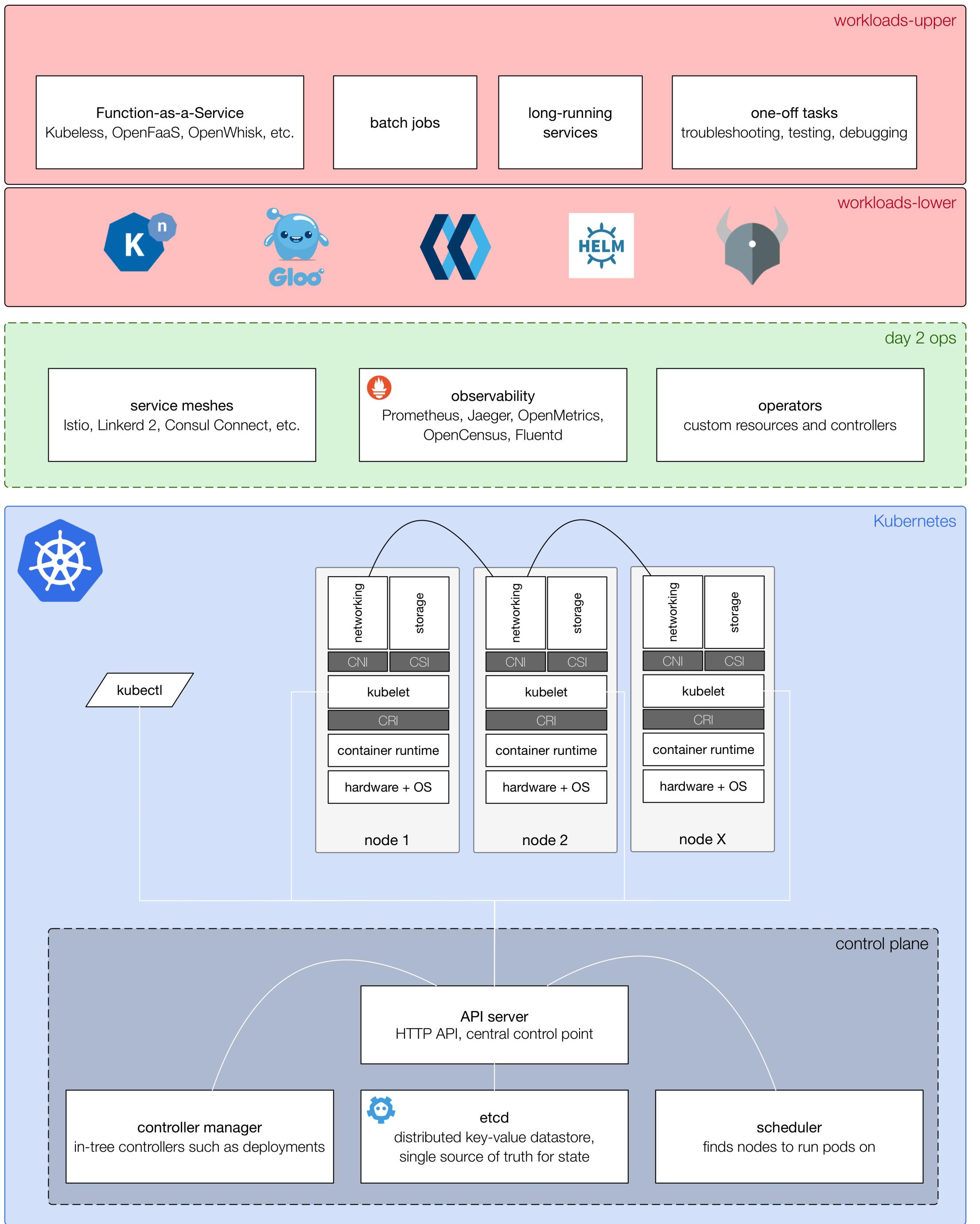
Kubernetes 101



“Begin at the beginning,” the King said, very gravely, “and go on till you come to the end: then stop.”

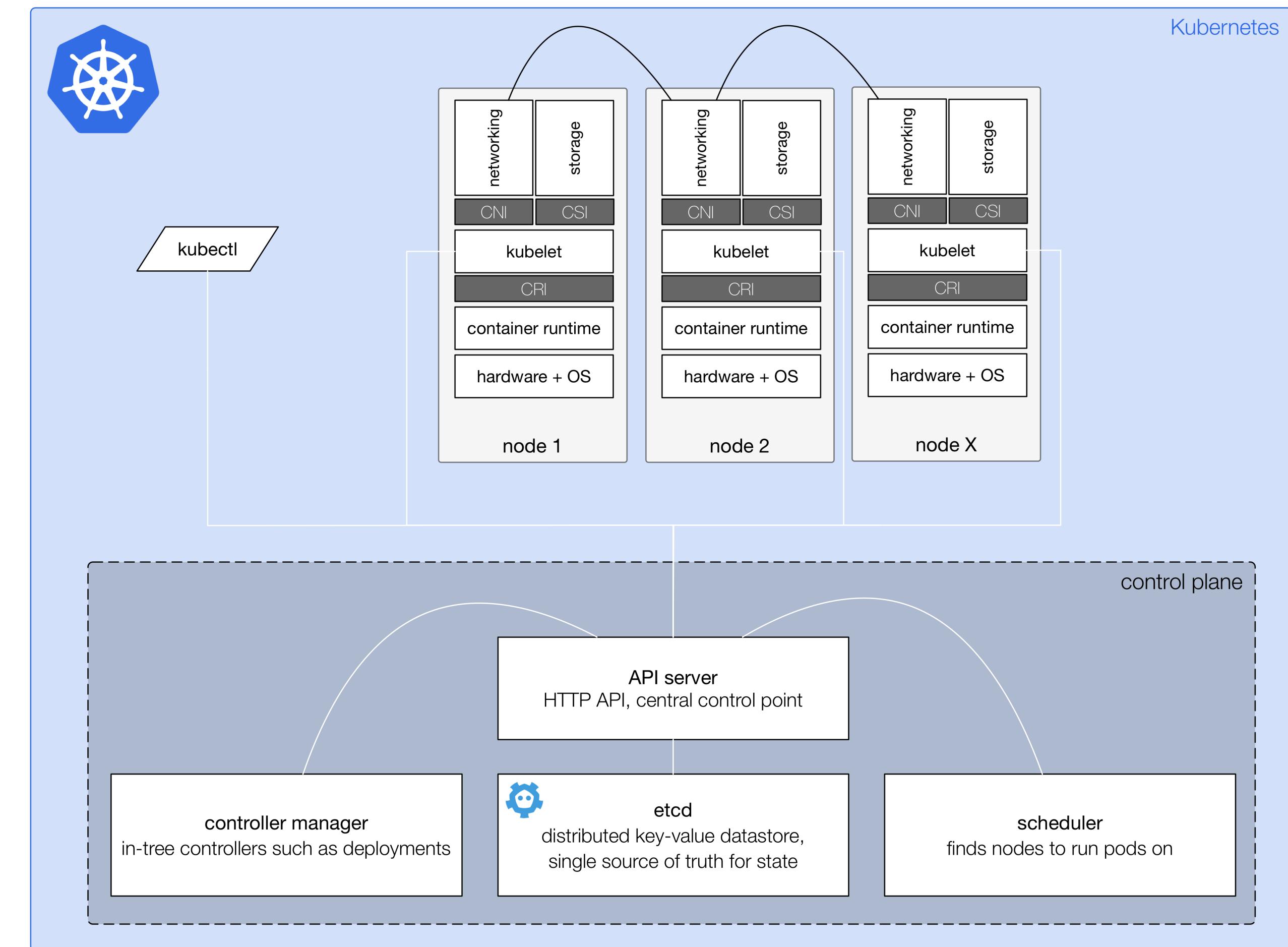
— Lewis Carroll, *Alice in Wonderland*

tags: humor



Kubernetes

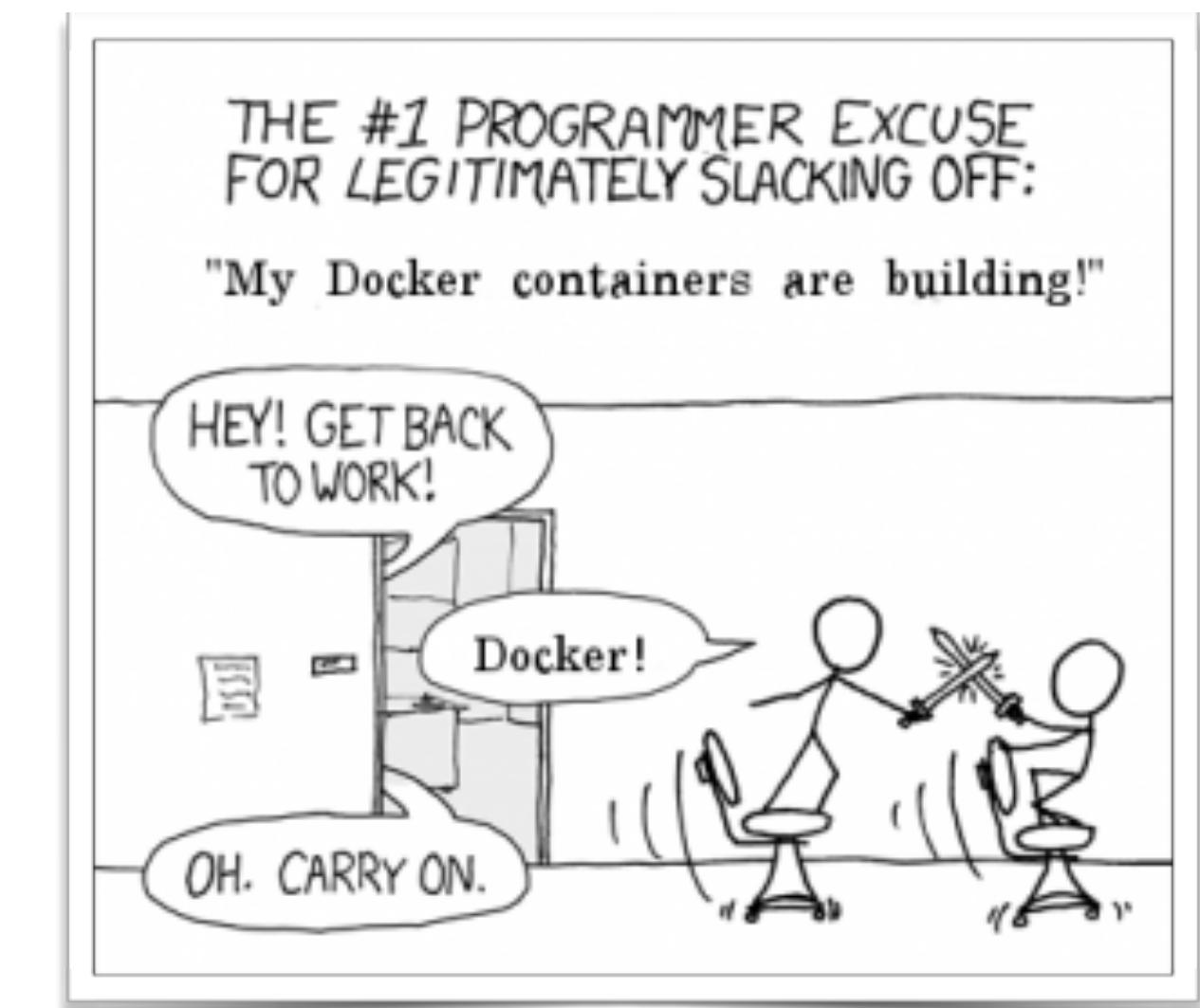
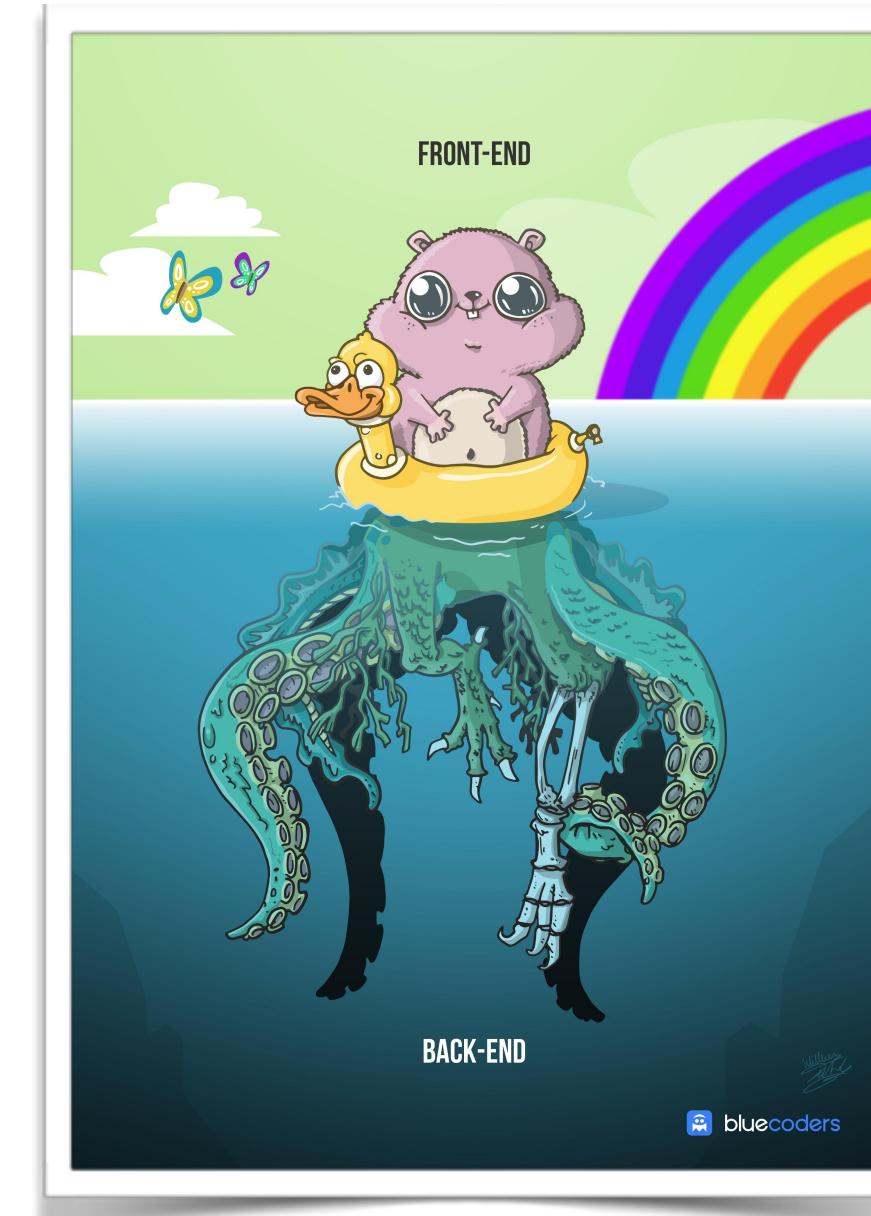
- Container lifecycle management
- Declarative API + control loops
- Robust, flexible, scalable
- Extensible



kubernetes.io

Roles and responsibilities

- infrastructure admin
- namespace admin
- developer



How can I customize Kubernetes?



Customization options in principle

- in-tree (upstream) via SIG or direct PR
- maintain your own fork
- built-in customization approaches



Customization approaches

- configuration files and flags (kubelet, kube-apiserver, etc.)
- extension points
 - cloud providers I
 - kubelet (plugins for network/devices/storage and container runtimes) I
 - kubectl plugins I
 - access extensions in the API server A
 - custom resources/controllers A
 - extension API servers A
 - scheduler extensions I

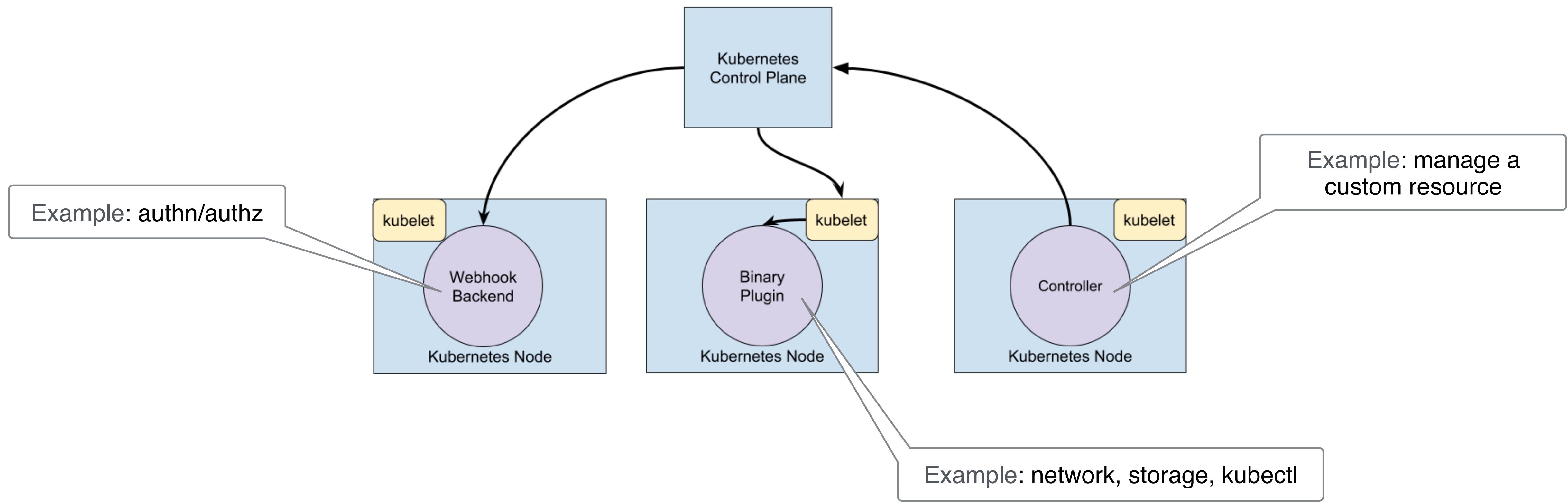
I

A

infrastructure

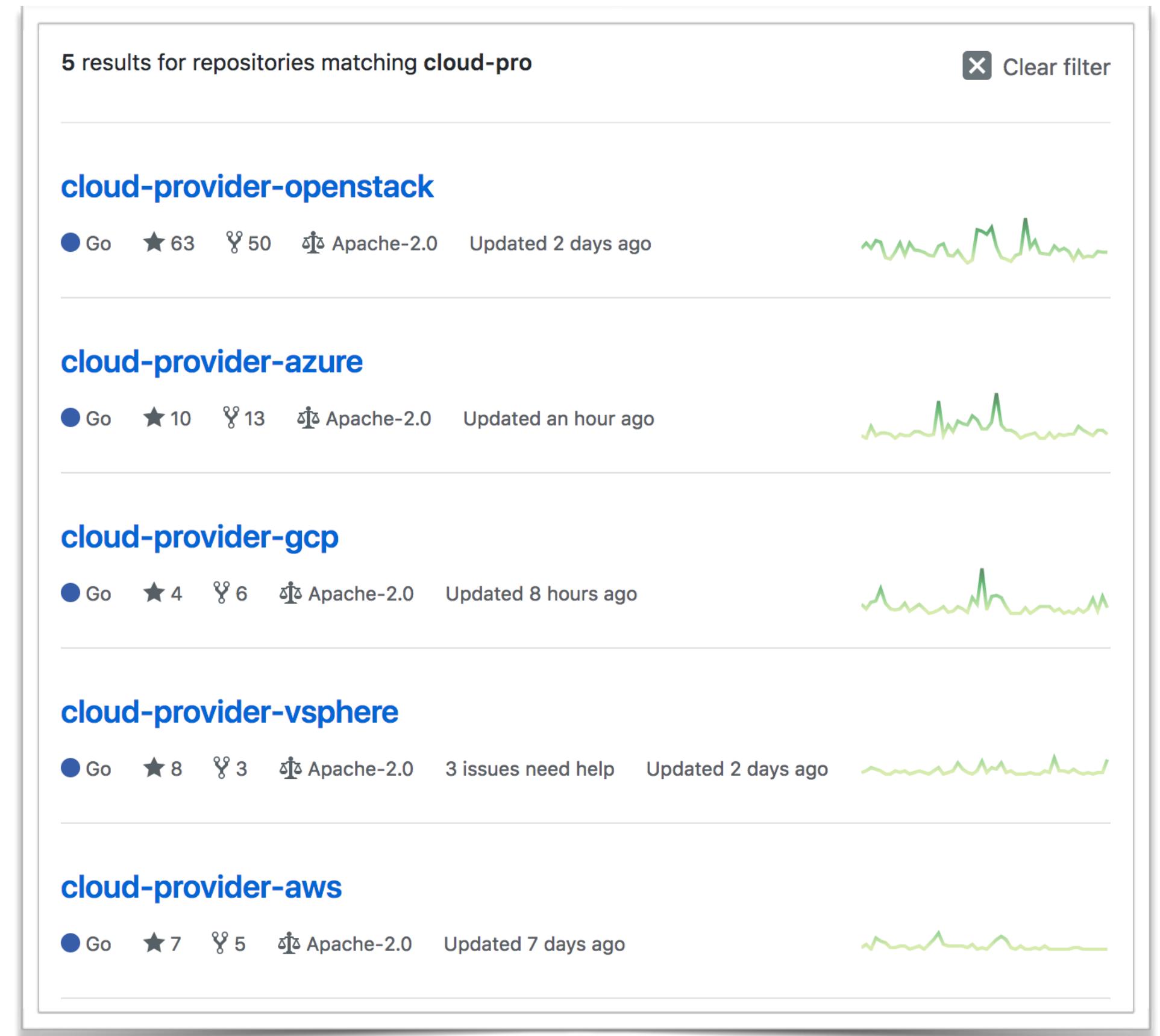
API

Extension patterns



Cloud providers

- in-tree libraries/controller manager
- interfaces for things like:
 - load balancers
 - network routes
 - nodes/VMs



github.com/kubernetes

kubelet: network/device/storage plugins

- Network—standard: CNI

FEATURE STATE: Kubernetes v1.12 alpha

github.com/containernetworking/cni

kubernetes.io/docs/concepts/extend-kubernetes/compute-storage-net/network-plugins

- Devices—GPUs, FPGAs, etc.

FEATURE STATE: Kubernetes v1.12 beta

kubernetes.io/docs/concepts/extend-kubernetes/compute-storage-net/device-plugins

- Storage—20+ in-tree, up-and-coming standard: CSI

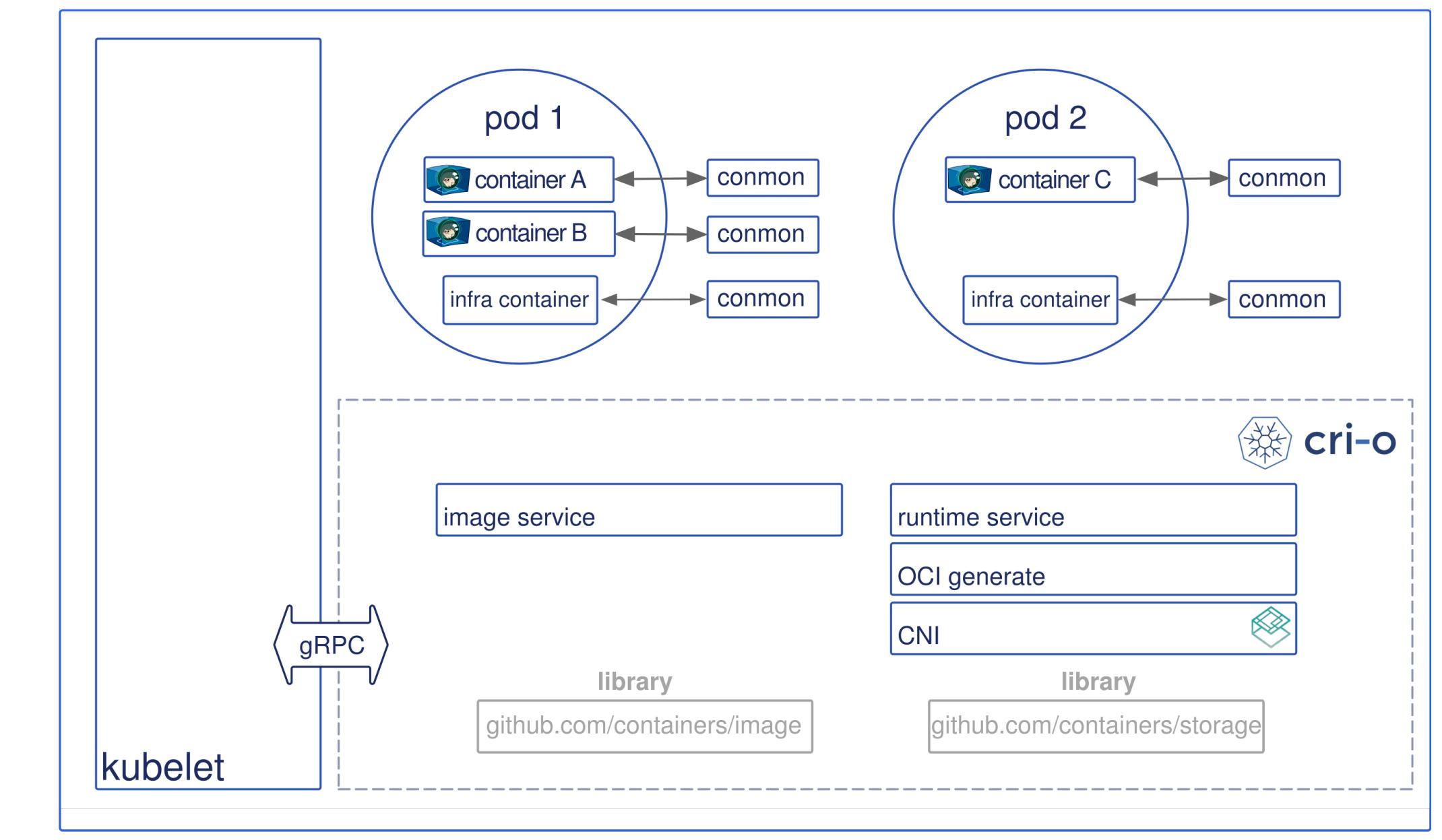
FEATURE STATE: Kubernetes v1.10 □ beta

kubernetes.io/docs/concepts/storage/volumes/#types-of-volumes

kubernetes.io/blog/2018/04/10/container-storage-interface-beta

kubelet: container runtimes

- Container runtime – standard: CRI (since Kubernetes 1.5)
kubernetes.io/blog/2016/12/container-runtime-interface-cri-in-kubernetes
- Nowadays multiple options:
 - runc
 - containerd
 - Kata containers
 - gVisor
 - hyper.sh



cri-o.io



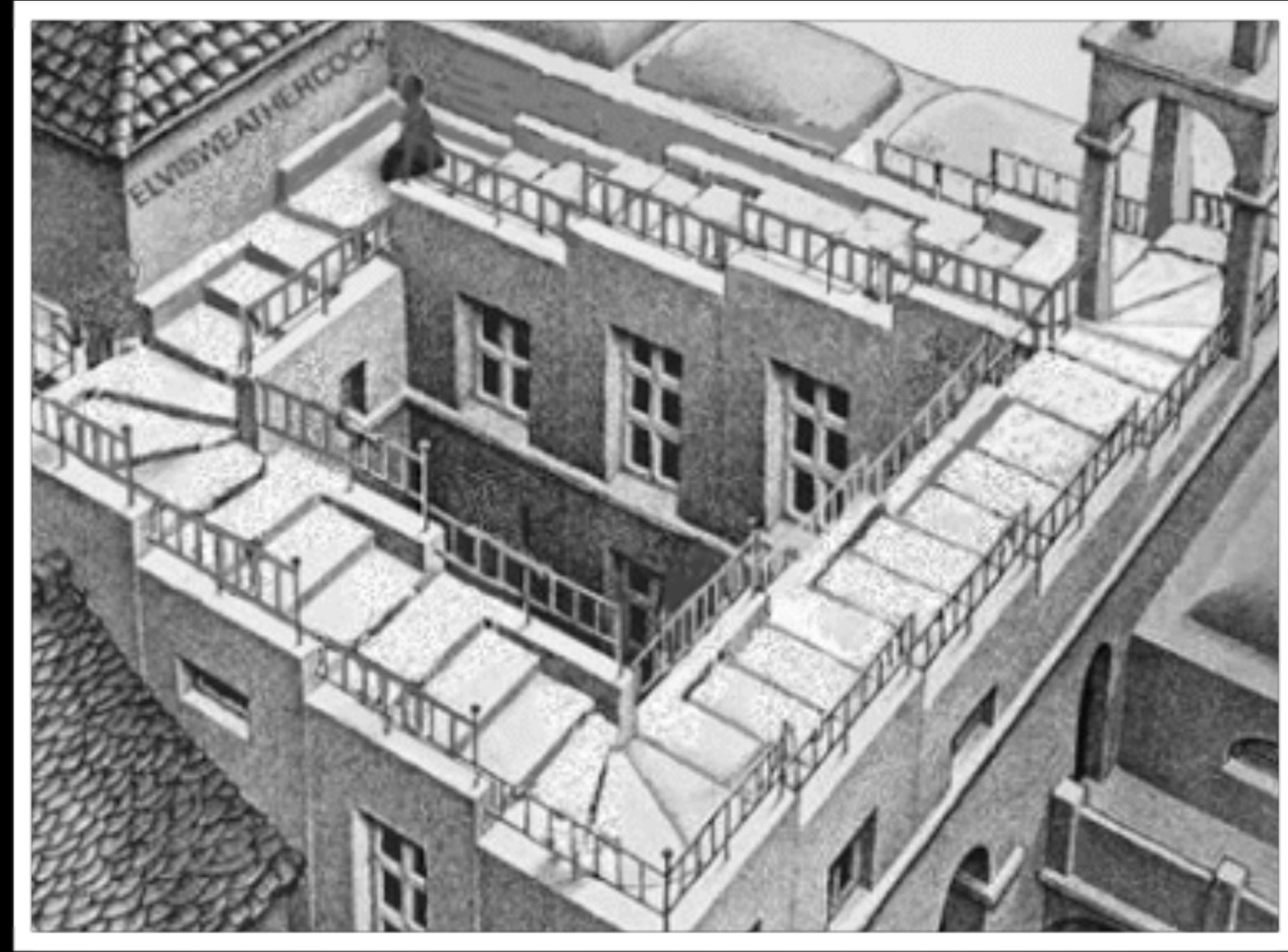
kubectl plugins

- Extend the set of commands

FEATURE STATE: Kubernetes v1.12 alpha

kubernetes.io/docs/tasks/extend-kubectl/kubectl-plugins

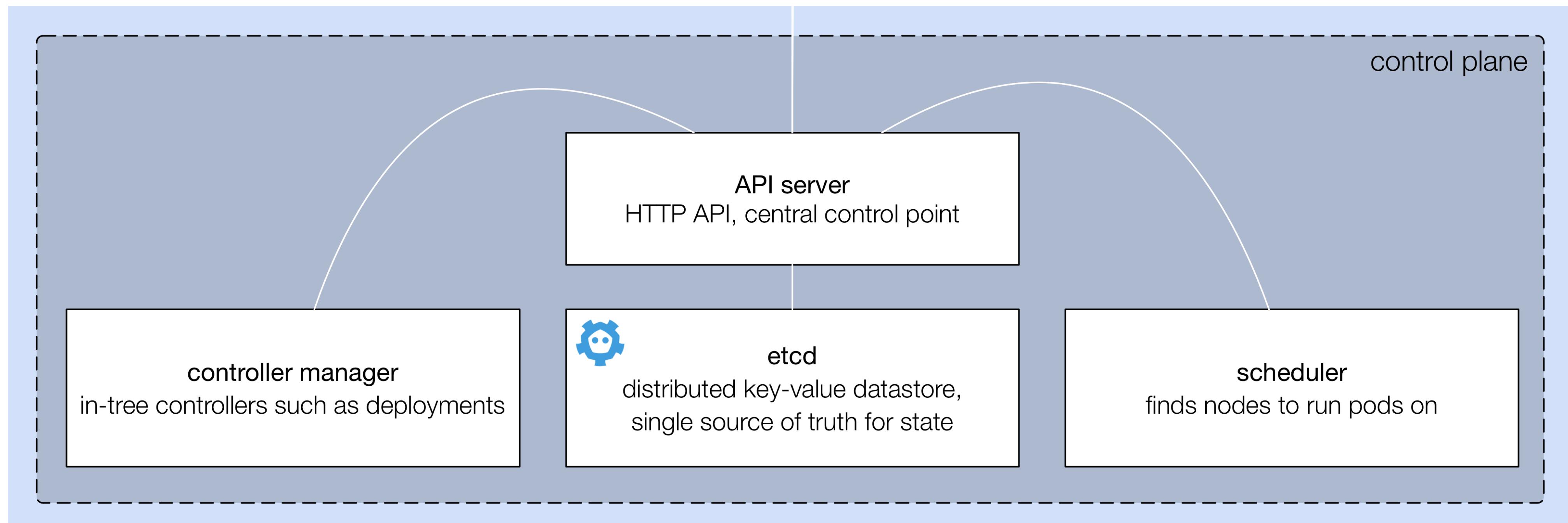
- Write in any programming language (note: these are binary extensions)
- Examples: context control, service catalog, user verification



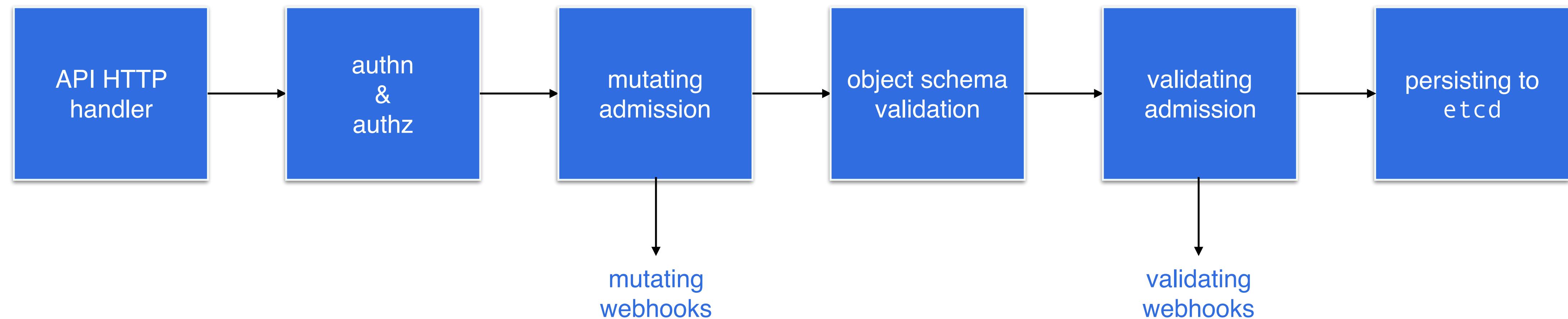
as simple plugin in action: `kubectl inspect`

Extending the Kubernetes API

Quick control plane refresher



The life of an API request



Flow diagram is based on [Extensible Admission is Beta](#) and [Kubernetes deep dive: API Server – part 1](#).

What are (in-tree) core resources?

Catalog	Kind	Group	1.5	Version	Group	1.6	Version	Group	1.7	Version	Group	1.8	Version	Group	1.9
Workloads	Container	Core	v1	v1	Core	v1	v1								
	CronJob	Batch	v2alpha1	v1beta1	Batch	v1beta1	v1beta1								
	DaemonSet	Extensions	v1beta1	Extensions	v1beta1	Extensions	v1beta1	Apps	v1beta1	Apps	v1beta2	v1beta2	Apps	v1	v1
	Deployment	Extensions	v1beta1	Apps	v1beta1	Apps	v1beta1	Apps	v1beta1	Apps	v1beta2	v1beta2	Apps	v1	v1
	Job	Batch	v1	Batch	v1	Batch	v1								
	Pod	Core	v1	Core	v1	Core	v1								
	ReplicaSet	Extensions	v1beta1	v1beta2	Extensions	v1beta2	v1								
	ReplicationController	Core	v1	Core	v1	Core	v1								
	StatefulSet	Apps	v1beta1	v1beta2	Apps	v1beta2	v1								

Catalog	Kind	Group	1.5	Version	Group	1.6	Version	Group	1.7	Version	Group	1.8	Version	Group	1.9
Workloads	Container	Core	v1	v1	Core	v1	v1								
	CronJob	Batch	v2alpha1	v1beta1	Batch	v1beta1	v1beta1								
	DaemonSet	Extensions	v1beta1	v1beta2	Extensions	v1beta2	v1								
	Deployment	Extensions	v1beta1	Apps	v1beta1	Apps	v1beta1	Apps	v1beta1	Apps	v1beta1	v1beta2	Apps	v1beta2	v1
	Job	Batch	v1	Batch	v1	Batch	v1								
	Pod	Core	v1	Core	v1	Core	v1								
	ReplicaSet	Extensions	v1beta1	v1beta2	Extensions	v1beta2	v1								
	ReplicationController	Core	v1	Core	v1	Core	v1								
	StatefulSet	Apps	v1beta1	v1beta2	Apps	v1beta2	v1								

Cluster	ClusterRole	RbacAuthorization	v1alpha1	RBAC	v1beta1	RBAC	v1beta1	RBAC	v1	RBAC	v1	RBAC	v1	RBAC	v1
	ClusterRoleBinding	RbacAuthorization	v1alpha1	RBAC	v1beta1	RBAC	v1beta1	RBAC	v1	RBAC	v1	RBAC	v1	RBAC	v1
	ComponentStatus	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1
	LocalSubjectAccessReview	Authorization	v1beta1	Authorization	v1	Authorization	v1	Authorization	v1	Authorization	v1	Authorization	v1	Authorization	v1
	Namespace	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1
	Node	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1
	PersistentVolume	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1
	ResourceQuota	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1
	Role	RbacAuthorization	v1alpha1	RBAC	v1beta1	RBAC	v1beta1	RBAC	v1	RBAC	v1	RBAC	v1	RBAC	v1
	RoleBinding	RbacAuthorization	v1alpha1	RBAC	v1beta1	RBAC	v1beta1	RBAC	v1	RBAC	v1	RBAC	v1	RBAC	v1
	SelfSubjectAccessReview	Authorization	v1beta1	Authorization	v1	Authorization	v1	Authorization	v1	Authorization	v1	Authorization	v1	Authorization	v1
	SelfSubjectRulesReview														
	ServiceAccount	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1	Core	v1
	SubjectAccessReview	Authorization	v1beta1	Authorization	v1	Authorization	v1	Authorization	v1	Authorization	v1	Authorization	v1	Authorization	v1
	TokenReview	Authentication	v1beta1	Authorization	v1	Authorization	v1	Authorization	v1	Authorization	v1	Authorization	v1	Authorization	v1
	NetworkPolicy	Extensions	v1beta1	Extensions	v1beta1	Networking	v1beta1	Networking	v1	Networking	v1	Networking	v1	Networking	v1

Access extensions in the API server

- Admission controllers (in-tree, via configuration of the API server)

<https://kubernetes.io/docs/reference/access-authn-authz/admission-controllers/>

- Dynamic Admission Control

<https://kubernetes.io/docs/reference/access-authn-authz/extensible-admission-controllers/>

- Admission Webhooks (beta)
- *Initializers (alpha)*

Custom resources

- Support for “known” resources beyond core resources
 - kubernetes.io/docs/concepts/extend-kubernetes/api-extension/custom-resources
 - blog.openshift.com/kubernetes-deep-dive-api-server-part-3a
- Use the API server to manage custom resources in etcd for you
- Custom resource definition (CRD) and instances
- Use the CLI to interact with custom resources in the usual way,
for example: `kubectl get mycustomresource`

Custom resource—example

```
1  apiVersion: apiextensions.k8s.io/v1beta1
2  kind:        CustomResourceDefinition
3  metadata:
4    name:       databases.example.com
5  spec:
6    group:      example.com
7    version:    v1
8    names:
9      kind:     Database
10     plural:   databases
11     scope:    Namespaced
```

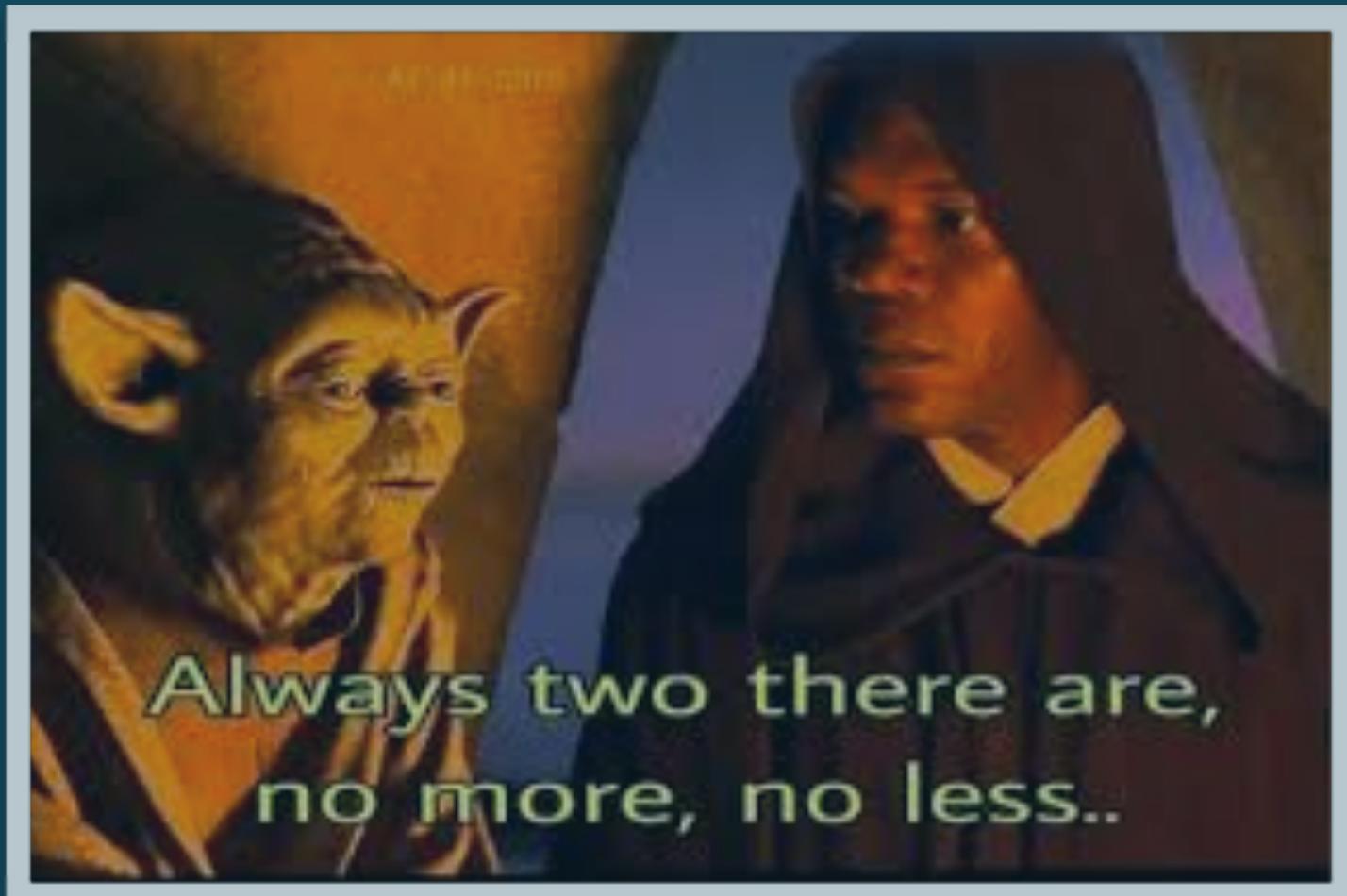
Custom controller

- Implement control loops beyond what thee (in-tree) controller manager supports
- Custom controller
 - dealing with core resources
github.com/kelseyhightower/secrets-controller
 - dealing with custom resources (aka operator)
github.com/kubernetes/sample-controller

Custom resources and controllers

	resource		controller	
	core	custom	in-tree	custom
Kubernetes control plane	X		X	
simple controller	X			X
operator		X		X

Operators



Always two there are,
no more, no less..

Operators

operator = custom resource + custom controller

- Motivation: application lifecycle management
- Use one of 30+ available operators or write your own with:
 - Kubebuilder
 - Kubernetes Operator Kit
 - kutil
 - Metacontroller
 - Operator SDK

Operator use cases

- zero-downtime upgrades of the app the operator supervises
- workflow automations
- policy enforcement
- managing stateful workloads
 - resizing of followers in a distributed datastore
 - backup & restore of a database
 - re-balancing of a distributed message queue

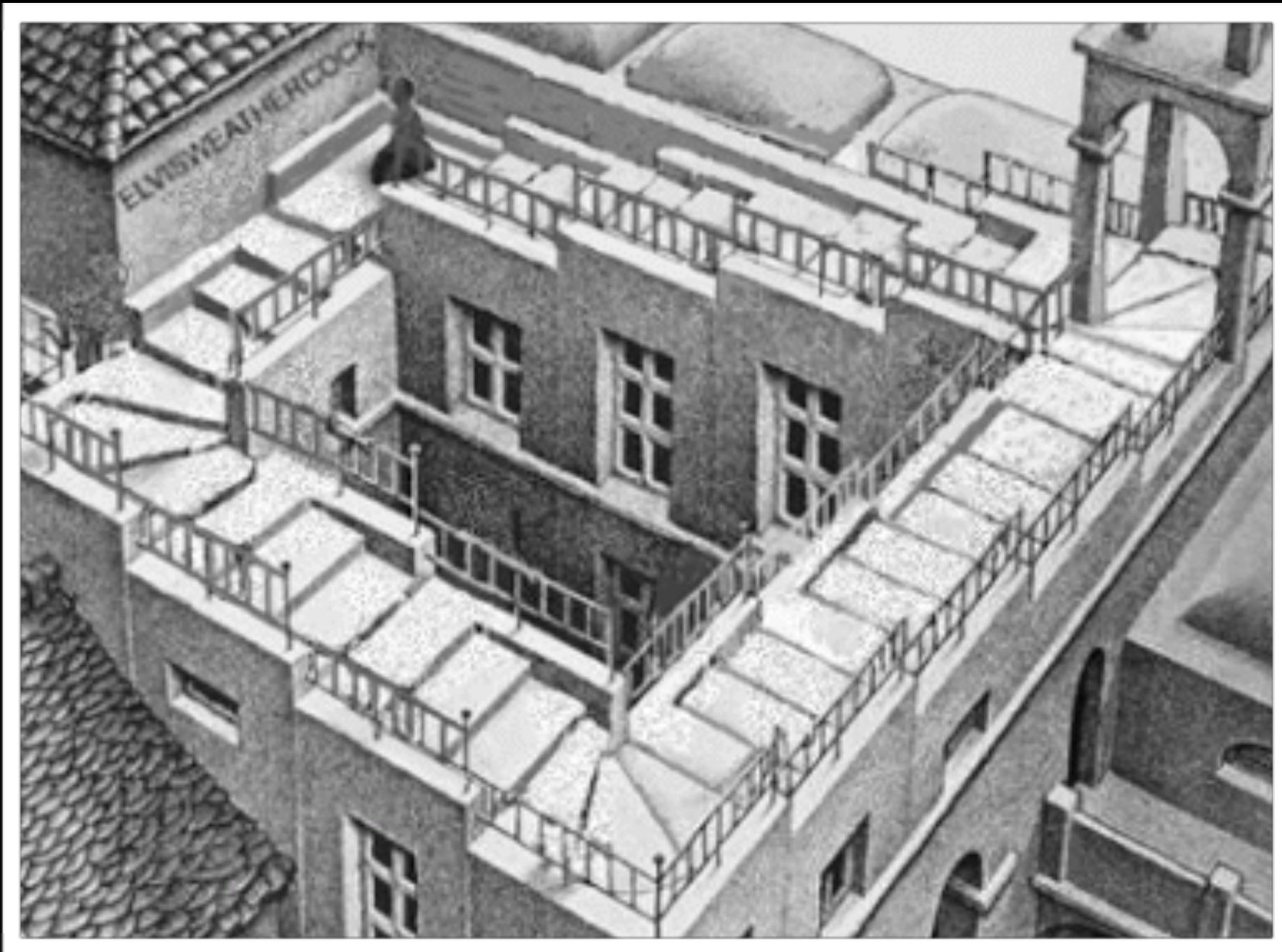
Operator examples

- etcd
- Prometheus
- Postgres
- Vitess MySQL
- MongoDB
- Couchbase
- Kafka

The screenshot shows the OpenShift Container Platform's Kubernetes Marketplace interface. The left sidebar includes links for Home, Operators (which is selected), Your Cluster Services, Catalog Sources, Subscriptions, Install Plans, Workloads, Networking, Storage, Builds, Service Catalog, Monitoring, and Administration. The main content area is titled "Kubernetes Marketplace" and displays "28 items". A callout highlights the "All Categories" section with the text "Expose ‘pending requests’ with info box." Another callout highlights the "View Filtered Results (2)" button with the text "Filter the offerings to see those pending requests.". Below these, there are sections for "Featured Services", "Databases", and "View All (12)". Each card includes a provider logo, name, and brief description.

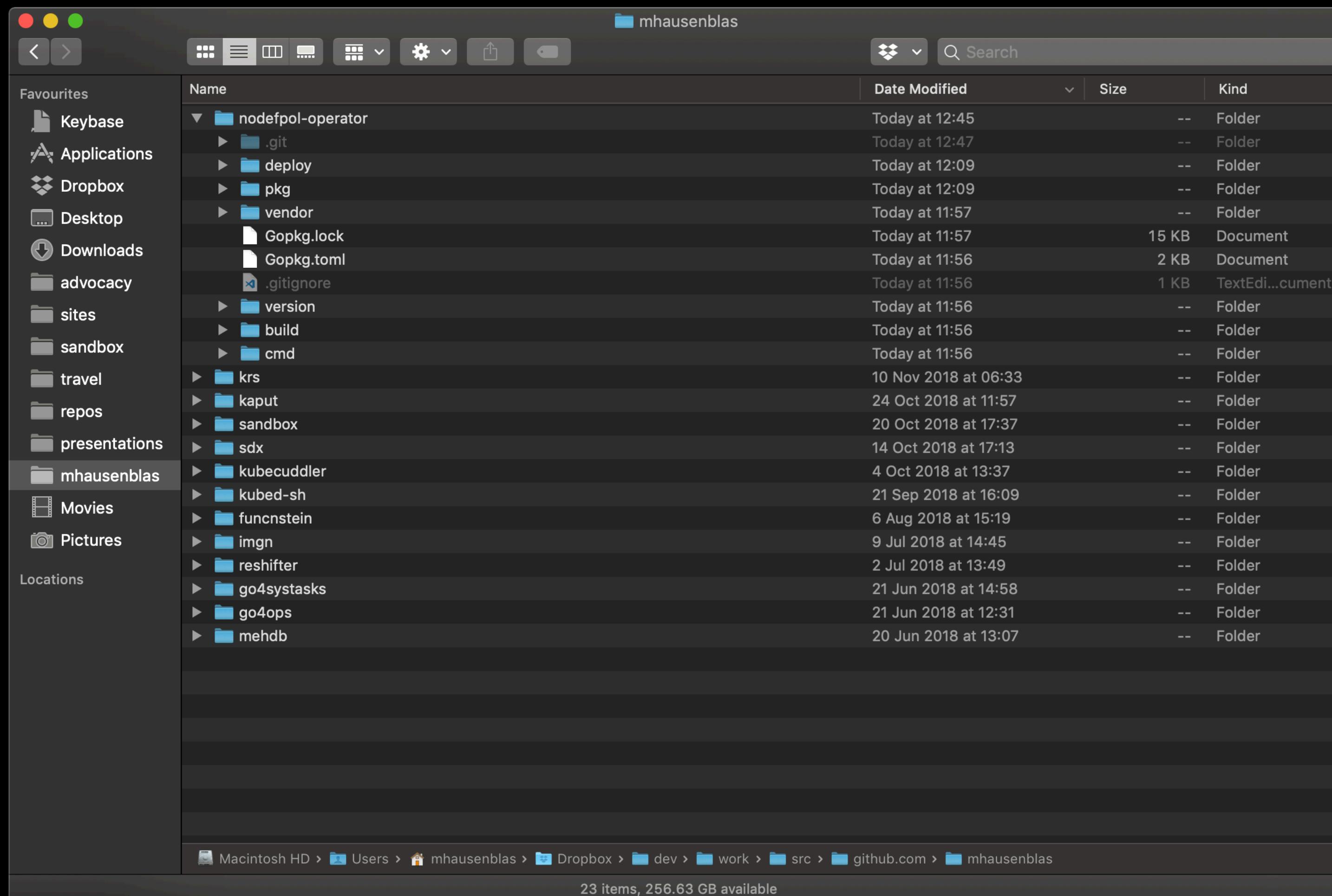
Category	Offering	Description
Featured Services	etcd	A distributed, reliable key-value store for the most critical data of a distributed system.
	Prometheus Operator	Manages Prometheus and Alertmanager clusters natively on Kubernetes.
	Vitess MySQL	Vitess is a database clustering system for horizontal scaling of MySQL.
Databases	Couchbase Enterprise	The system of engagement database for web, mobile and IoT.
	MongoDB Operator	NoSQL document-oriented database for content-driven applications.
	Vitess MySQL	Vitess is a database clustering system for horizontal scaling of MySQL.

github.com/mhausenblas/operator-101

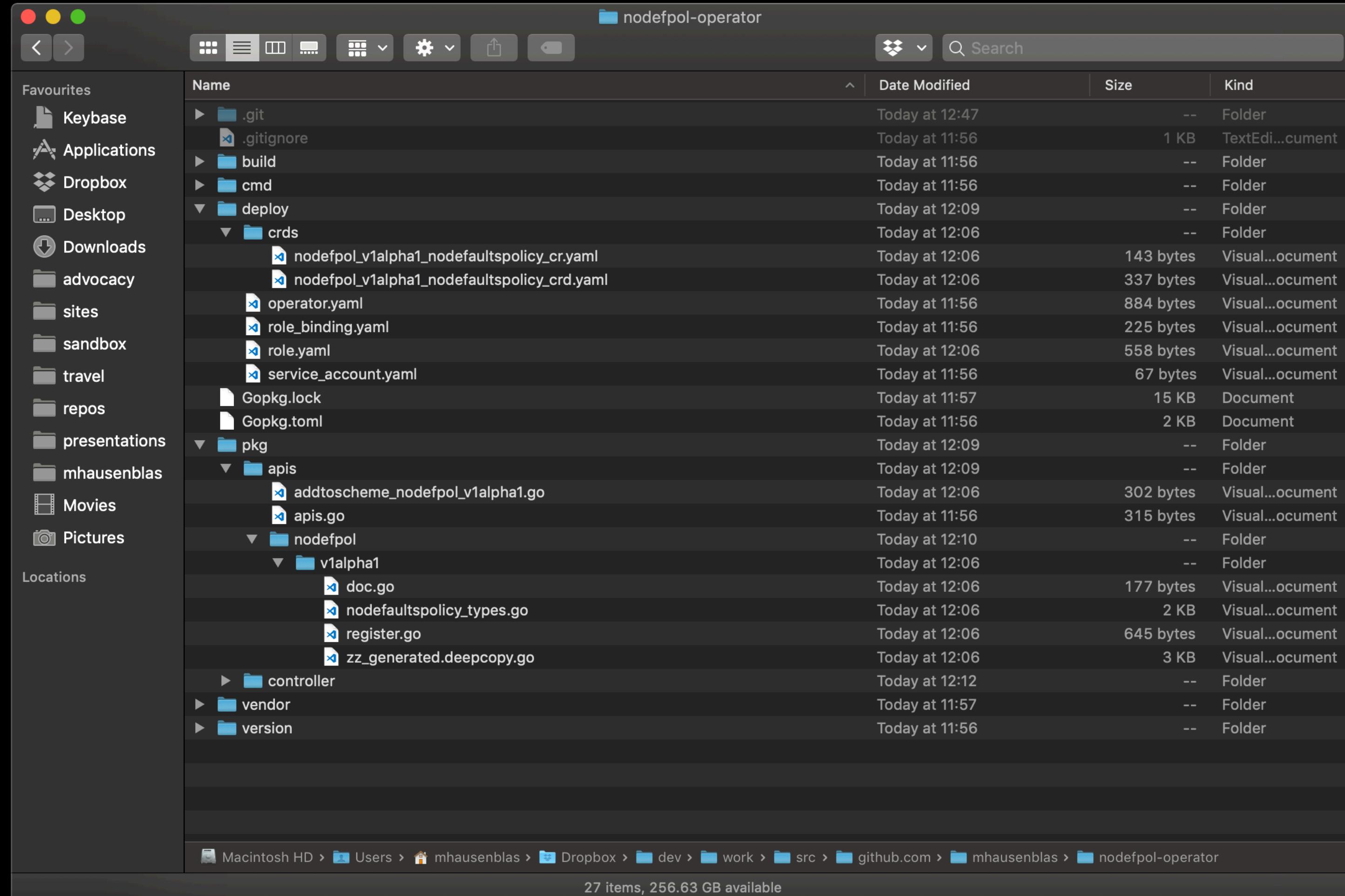


a simple operator in action: NoDefaultsPolicy

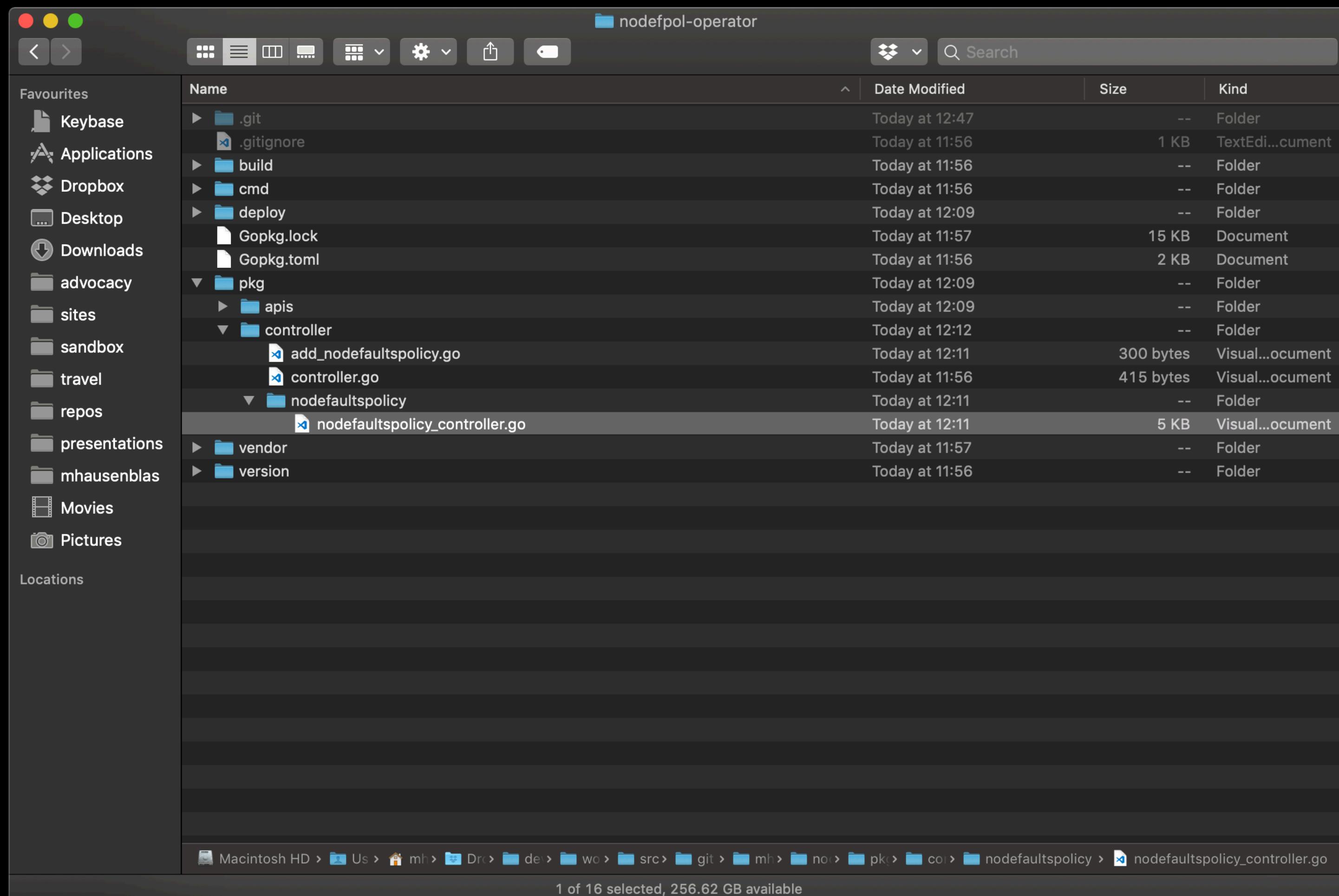
```
$ operator-sdk new nodefpol-operator
```



```
$ operator-sdk add api --api-version=nodefpol.k8space.io/v1alpha1 --kind=NoDefaultsPolicy
```



```
$ operator-sdk add controller --api-version=nodefpol.k8space.io/v1alpha1 --kind=NoDefaultsPolicy
```



```
$ kubectl -n ndp-demo apply -f deploy/crds/nodefpol_v1alpha1_nodefaultspolicy_crd.yaml  
$ OPERATOR_NAME=nodefpol-operator operator-sdk up local --namespace "ndp-demo"
```



1. tmux

```
2018/11/14 12:31:52 Go OS/Arch: darwin/amd64  
2018/11/14 12:31:52 operator-sdk Version: v0.1.1+git  
2018/11/14 12:31:52 Registering Components.  
2018/11/14 12:31:52 Starting the Cmd.  
2018/11/14 12:35:32 Reconciling NoDefaultsPolicy ndp-demo/example-nodefaultspolicy  
2018/11/14 12:35:32 Creating a new Pod ndp-demo/example-nodefaultspolicy-pod  
2018/11/14 12:35:32 Reconciling NoDefaultsPolicy ndp-demo/example-nodefaultspolicy  
2018/11/14 12:35:32 Skip reconcile: Pod ndp-demo/example-nodefaultspolicy-pod already exists  
2018/11/14 12:35:33 Reconciling NoDefaultsPolicy ndp-demo/example-nodefaultspolicy  
2018/11/14 12:35:33 Skip reconcile: Pod ndp-demo/example-nodefaultspolicy-pod already exists  
2018/11/14 12:35:33 Reconciling NoDefaultsPolicy ndp-demo/example-nodefaultspolicy  
2018/11/14 12:35:33 Skip reconcile: Pod ndp-demo/example-nodefaultspolicy-pod already exists  
2018/11/14 12:35:38 Reconciling NoDefaultsPolicy ndp-demo/example-nodefaultspolicy  
2018/11/14 12:35:38 Skip reconcile: Pod ndp-demo/example-nodefaultspolicy-pod already exists
```

```
~/Dropbox/dev/work/src/github.com/mhausenblas/nodefpol-operator/app-operator (master)*  
$ kubectl -n ndp-demo get crd  
NAME                                CREATED AT  
nodefaultspolicies.nodefpol.k8space.io  2018-11-14T11:23:35Z  
  
~/Dropbox/dev/work/src/github.com/mhausenblas/nodefpol-operator/app-operator (master)*  
$ kubectl -n ndp-demo get NoDefaultsPolicy  
NAME          AGE  
example-nodefaultspolicy  46s  
  
~/Dropbox/dev/work/src/github.com/mhausenblas/nodefpol-operator/app-operator (master)*  
$ kubectl -n ndp-demo describe NoDefaultsPolicy  
Name:           example-nodefaultspolicy  
Namespace:      ndp-demo  
Labels:         <none>  
Annotations:    kubectl.kubernetes.io/last-applied-configuration:  
                {"apiVersion":"nodefpol.k8space.io/v1alpha1","kind":"NoDefaultsPolicy","metadata":{"annotations":{},"name":"example-nodefaultspolicy","nam...  
API Version:   nodefpol.k8space.io/v1alpha1  
Kind:          NoDefaultsPolicy  
Metadata:  
  Creation Timestamp: 2018-11-14T11:35:32Z  
  Generation:       1  
  Resource Version: 230729  
  Self Link:        /apis/nodefpol.k8space.io/v1alpha1/namespaces/ndp-demo/nodefaultspolicies/example-nodefaultspolicy  
  UID:              651a1170-e801-11e8-87f9-ca17f3440361  
Spec:  
  Size:            3  
Events:         <none>  
  
~/Dropbox/dev/work/src/github.com/mhausenblas/nodefpol-operator/app-operator (master)*  
$
```

sandbox | 100% | 2018-11-14T12:36 0:bash* *minikube default

grep '//TODO(user)'

The screenshot shows a code editor interface with a dark theme. A search bar at the top right contains the text 'nodefaultspolicy_controller.go — Untitled (Workspace)' and a search icon. Below the search bar is a breadcrumb navigation path: 'nodefpol-operator' > 'pkg' > 'controller' > 'nodedefaultspolicy' > 'nodefaultspolicy_controller.go' > '{} nodedefaultspolicy'. The main editor area displays a Go file named 'nodefaultspolicy_controller.go'. The code includes several TODO comments:

```
75 // Reconcile reads that state of the cluster for a NoDefaultsPolicy object and makes changes based on the state read
76 // and what is in the NoDefaultsPolicy.Spec
77 // [todo(user]: Modify this Reconcile function to implement your Controller logic... This example creates
78 // a Pod as an example
79 // Note:
80 // The Controller will requeue the Request to be processed again if the returned error is non-nil or
81 // Result.Requeue is true, otherwise upon completion it will remove the work from the queue.
82 func (r *ReconcileNoDefaultsPolicy) Reconcile(request reconcile.Request) (reconcile.Result, error) {
83     log.Printf("Reconciling NoDefaultsPolicy %s/%s\n", request.Namespace, request.Name)
84
85     // Fetch the NoDefaultsPolicy instance
86     instance := &nodefpolv1alpha1.NoDefaultsPolicy{}
87     err := r.client.Get(context.TODO(), request.NamespacedName, instance)
88     if err != nil {
89         if errors.NotFound(err) {
90             // Request object not found, could have been deleted after reconcile request.
91             // Owned objects are automatically garbage collected. For additional cleanup logic use finalizers.
92             // Return and don't requeue
93             return reconcile.Result{}, nil
94         }
95         // Error reading the object - requeue the request.
96         return reconcile.Result{}, err
97     }
98
99     // Define a new Pod object
100    nod := newPodForCR(instance)
```

At the bottom of the editor, a 'PROBLEMS' tab is active, showing 19 issues. The search results panel below the editor lists 'todo(user)' with 4 results in 2 files:

- nodefpol-operator/nodedefpolicy_controller.go: // TODO(user]: Modify this to be the types you create that are owned by the primary resource
- nodefpol-operator/controller_kind.go: // TODO(user): Modify this Reconcile function to implement your Controller logic. This example creates

The status bar at the bottom shows 'master*' and 'Run on Save done.' The bottom right corner displays the Red Hat logo.

Extension API servers

- Full control but a lot of effort and responsibility
kubernetes.io/docs/tasks/access-kubernetes-api/setup-extension-api-server
- Typically more LOC than an controller or operator
- You might end up to manage storage in etcd yourself
- And beyond: the Open Service Broker API and the service catalog

kubernetes.io/docs/concepts/extend-kubernetes/service-catalog

openservicebrokerapi.org

Scheduler extensions

A scheduler selects a node to run your pods on, based on resource requirements, QoS, affinity, etc.

jvns.ca/blog/2017/07/27/how-does-the-kubernetes-scheduler-work

- You can modify policies or run multiple schedulers (with pod opt-in)

kubernetes.io/docs/tasks/administer-cluster/configure-multiple-schedulers

embano1.github.io/post/sched-reconcile

- You can use a Webhook

github.com/kubernetes/community/blob/master/contributors/design-proposals/scheduling/scheduler_extender.md

Other stuff you can customize in Kubernetes

- Monitoring & alerting (Prometheus/Grafana), logging (ELK/EFK stack)
- Secret management (encryption at rest, Vault)

- Ingress

kubernetes.io/docs/concepts/services-networking/ingress

- DNS

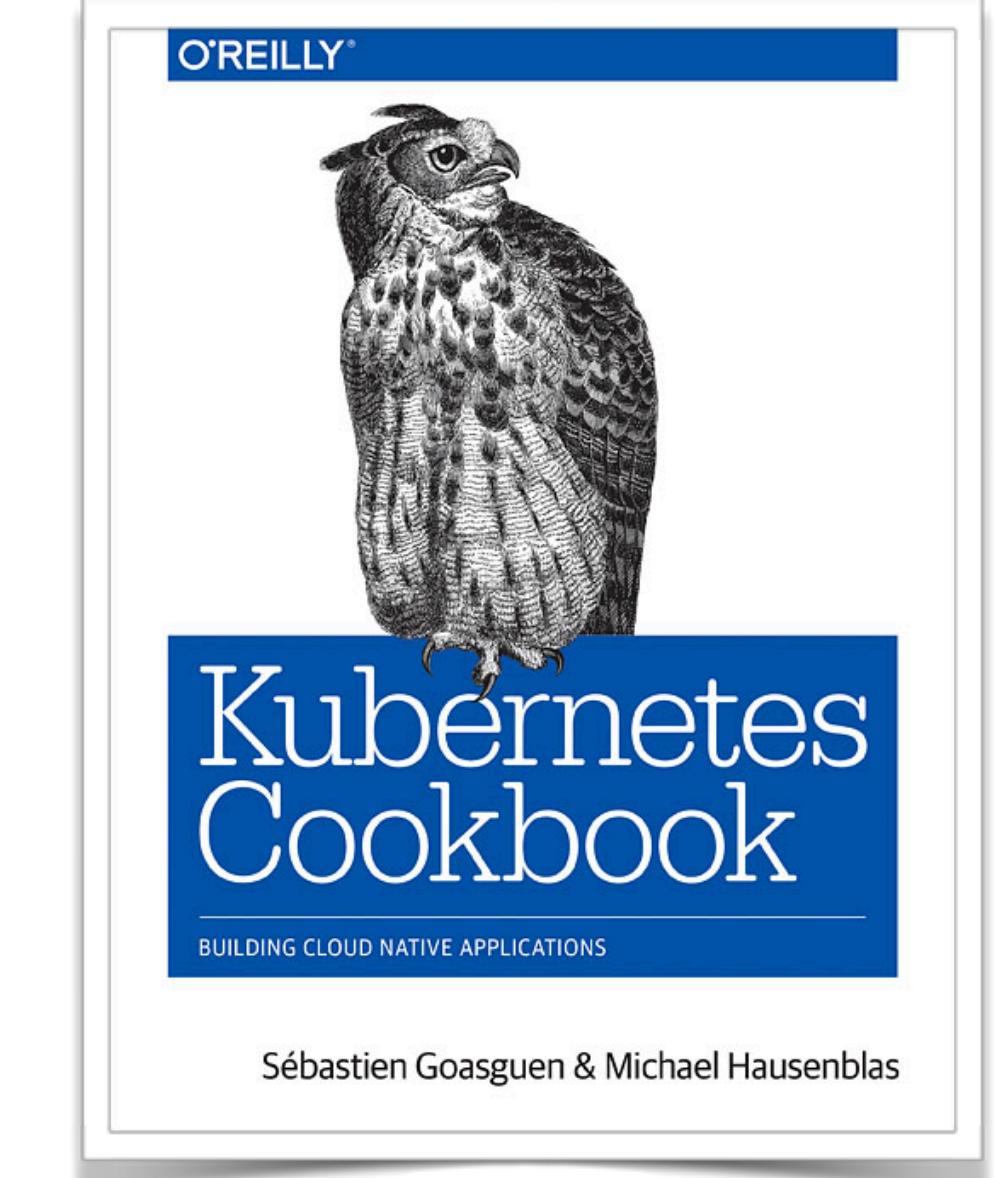
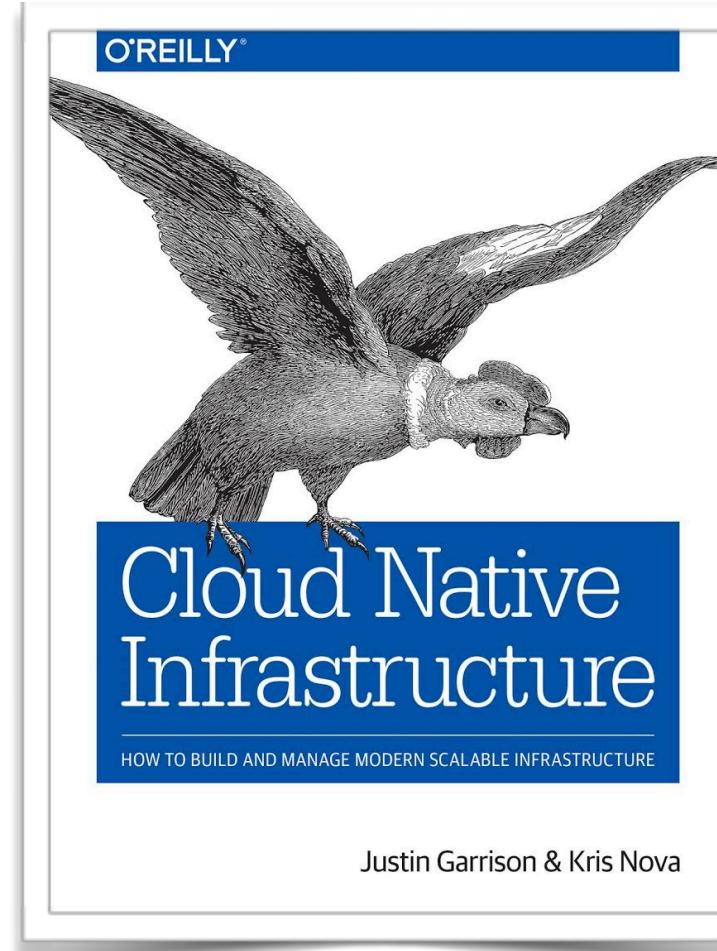
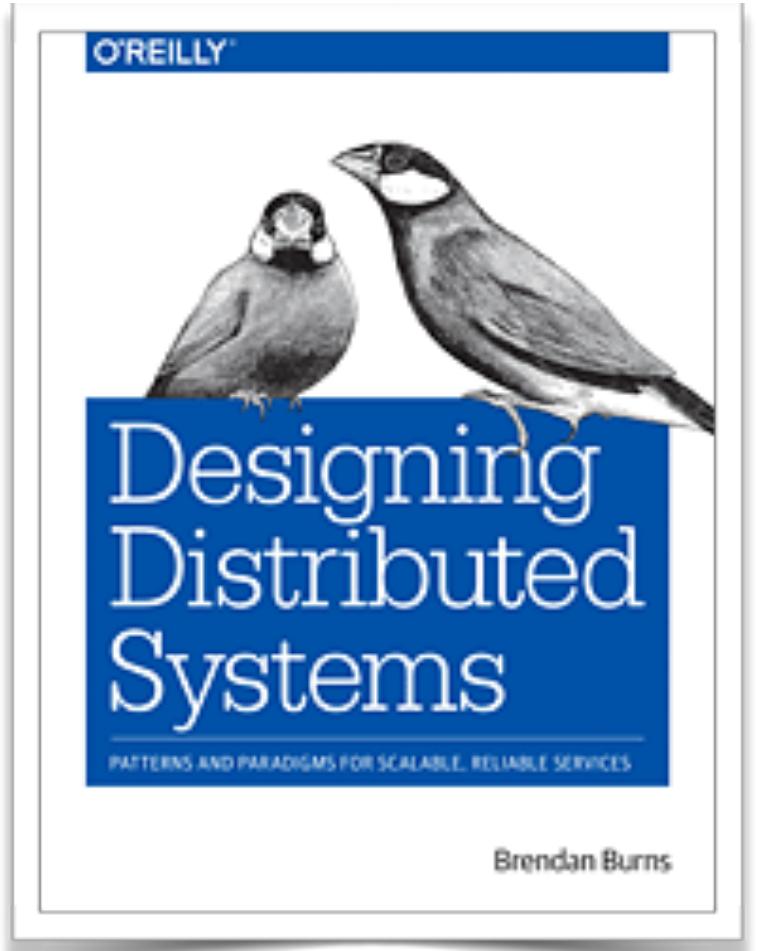
kubernetes.io/docs/tasks/administer-cluster/dns-custom-nameservers

- kube-proxy

kubernetes.io/blog/2018/07/09/ipvs-based-in-cluster-load-balancing-deep-dive

Resources

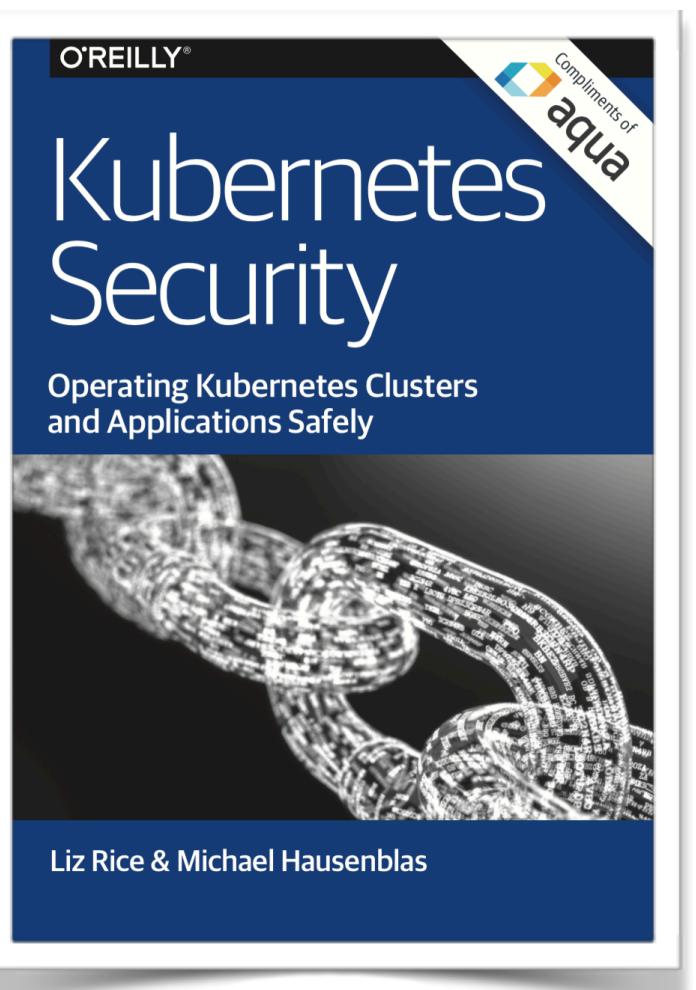
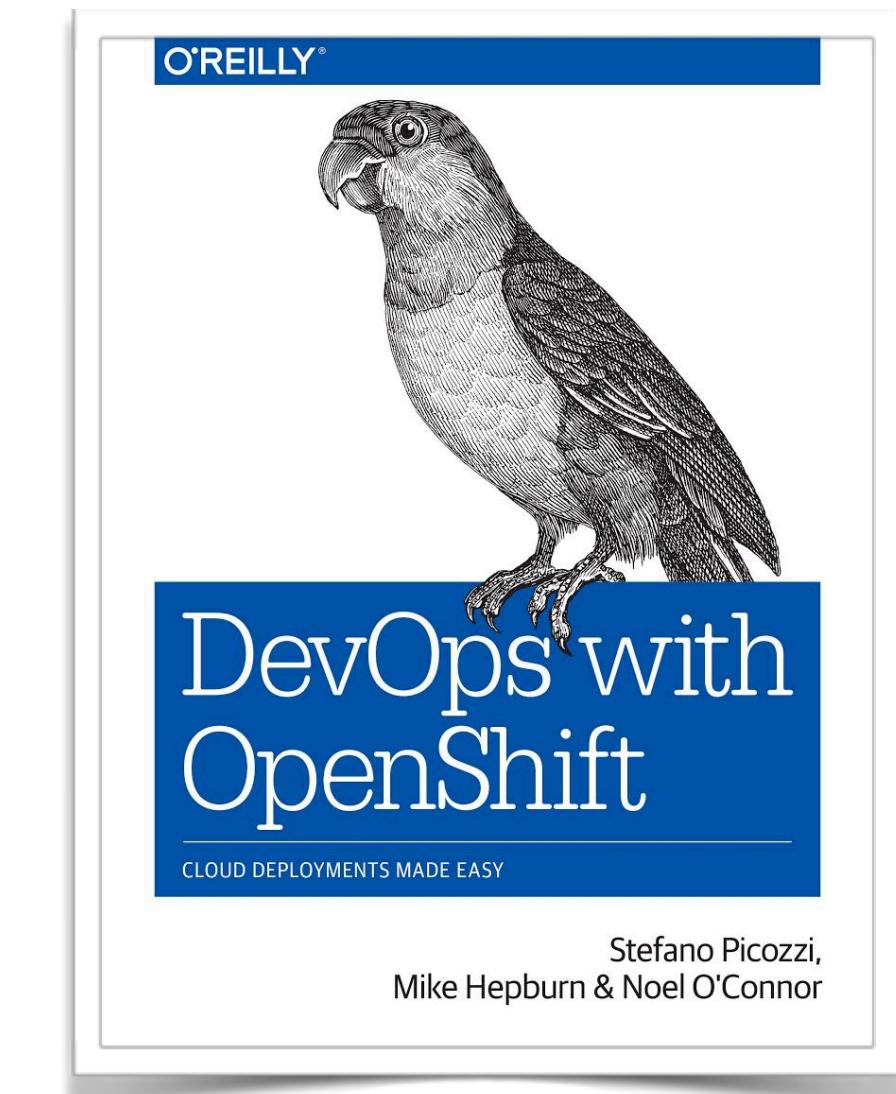




FREE eBook:
The Cloud Native Attitude

Everything you need to know about Cloud Native but were afraid to ask.

[DOWNLOAD NOW](#)



Articles and slide decks

- *Tim Hockin*—Kubernetes Extensibility
speakerdeck.com/thockin/kubernetes-extensibility
- *Jonathan Berkhahn & Carolyn Van Slyck*—Kubectl Plugins 101
kccnceu18.sched.com/event/DqwJ/kubectl-plugins-101-jonathan-berkhahn-ibm-carolyn-van-slyck-microsoft-intermediate-skill-level-slides-attached
- *Adrien Trouillaud*—Kubernetes Custom Resource, Controller & Operator Development Tools
admiralty.io/kubernetes-custom-resource-controller-and-operator-development-tools.html
- *Toader Sebastian*—A complete guide to Kubernetes Operator SDK
banzaicloud.com/blog/operator-sdk/
- *Rob Szumski*—Building an Kubernetes Operator for Prometheus and Thanos
robszumski.com/building-an-operator/

Repos, examples, tooling

- github.com/kubernetes/kubectl/tree/master/pkg/pluginutils
- github.com/carolynvs/kubectl-flags-plugin
- github.com/jordanwilson230/kubectl-plugins
- github.com/kelseyhightower/denyenv-validating-admission-webhook
- github.com/kubernetes-sigs/controller-tools
- github.com/kubernetes-sigs/kubebuilder
- metacontroller.app
- github.com/yaronha/kube-crd
- github.com/operator-framework/operator-sdk
- github.com/operator-framework/awesome-operators
- reactiveops.github.io/rbac-manager

Kubernetes docs and blog posts

- kubernetes.io/docs/concepts/extend-kubernetes/extend-cluster/
 - kubernetes.io/docs/concepts/extend-kubernetes/api-extension/custom-resources/
 - kubernetes.io/docs/concepts/extend-kubernetes/api-extension/apiserver-aggregation/
 - kubernetes.io/docs/tasks/access-kubernetes-api/setup-extension-api-server/
- kubernetes.io/docs/tasks/extend-kubectl/kubectl-plugins/
- kubernetes.io/docs/reference/access-authn-authz/webhook/
- kubernetes.io/docs/setup/scratch/#cloud-provider
- kubernetes.io/blog/2018/01/extensible-admission-is-beta/

Videos

- *Tim Hockin & Michael Rubin*—Kubernetes Distributions and ‘Kernels’
<https://www.youtube.com/watch?v=fXBjA2hH-CQ>
- *Stefan Schimanski*:
 - Kubernetes as a API driven platform, Reykjavík Kubernetes Meetup
<https://www.youtube.com/watch?v=BiE7oKeEzDU>
 - SIG API Machinery Deep Dive
<https://www.youtube.com/watch?v=XsFH7OEIIvl>
- *James Munnelly*—Extending the Kubernetes API: What the Docs Don’t Tell You
<https://www.youtube.com/watch?v=PYLFZVv68IM>



redhat

learn.openshift.com

The screenshot shows the homepage of learn.openshift.com. At the top, there's a navigation bar with links for PRODUCTS, LEARN, COMMUNITY, SUPPORT, FREE TRIAL, and SIGN IN. Below the navigation is a section titled "Interactive Learning Portal" with a subtext explaining that it provides pre-configured OpenShift instances for experimentation. There are six course cards arranged in two rows of three:

- Foundations of OpenShift** (START COURSE)
- Building Applications On OpenShift** (START COURSE)
- Subsystems, Components, and Internals** (START COURSE)

- OpenShift Playgrounds** (START COURSE)
- Service Mesh Workshop with Istio** (START COURSE)
- Serverless Scenarios with OpenShift Cloud Functions** (START COURSE)

