



KubeCon



CloudNativeCon

Europe 2019

Resize Your Pods w/o Disruptions

aka How to Have a Cake and Eat a Cake

Karol Gołąb
Beata Skiba

kgolab@google.com
bskiba@google.com

Size your workload



KubeCon



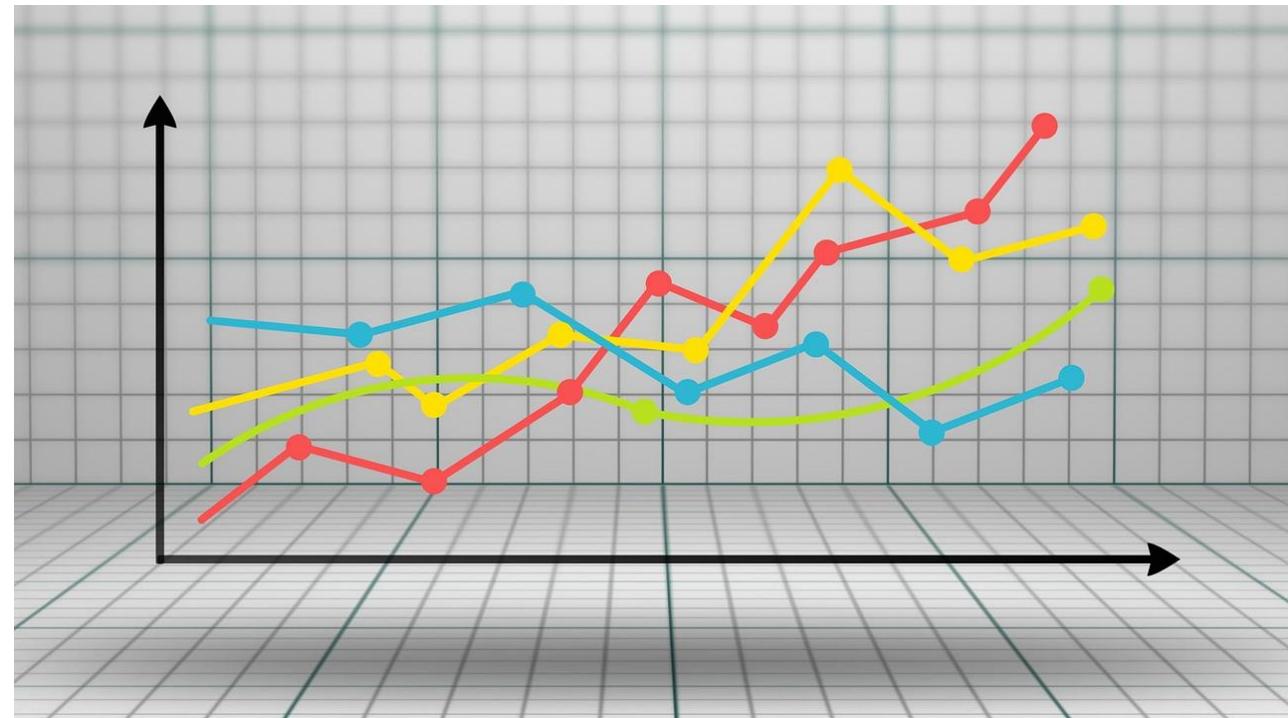
CloudNativeCon

Europe 2019

```
apiVersion: apps/v1
kind: Deployment
...
template:
  metadata:
    labels:
      app: worker
  spec:
    containers:
      - name: worker
        resources:
          requests:
            memory: "100M"
            cpu: "250m"
```

Everything is changing

- ▶ Daily/weekly traffic patterns
- ▶ User base growing over time
- ▶ App lifecycle phases with different resource needs



But why should you care?



KubeCon



CloudNativeCon

Europe 2019



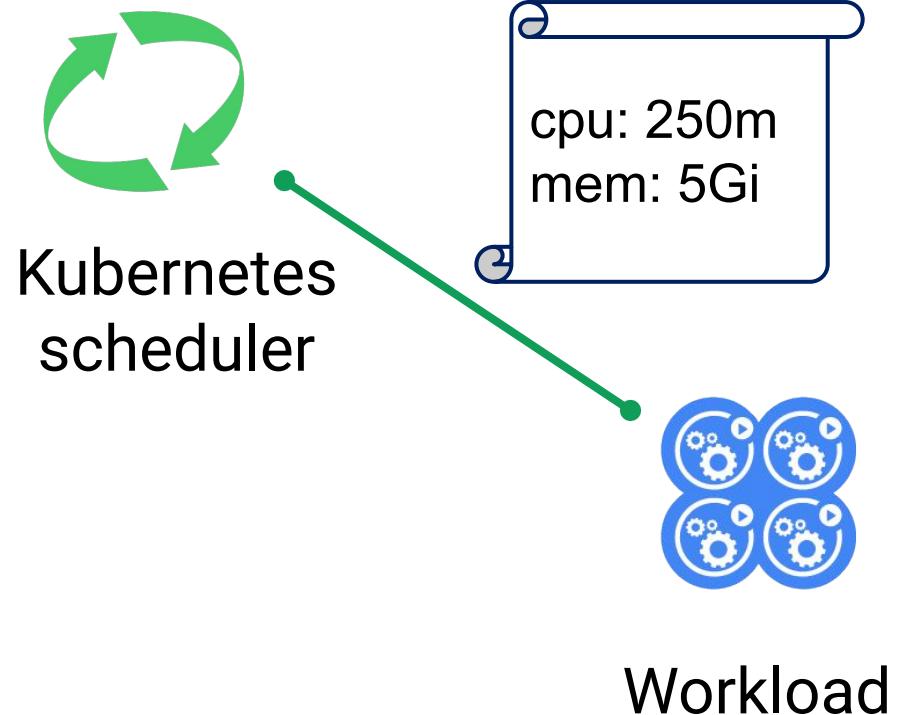
Contract with the scheduler



KubeCon

CloudNativeCon

Europe 2019



Scheduling in action



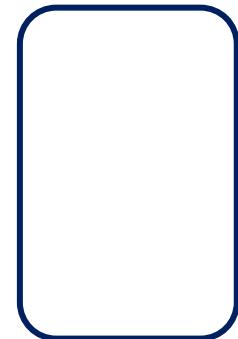
KubeCon



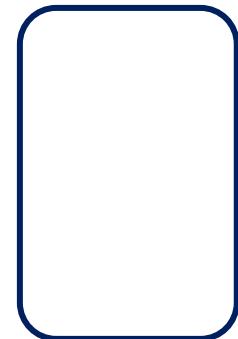
CloudNativeCon

Europe 2019

Nodes



cpu: 900m



cpu: 900m

Pods

cpu: 250m



cpu: 600m



Kubernetes
scheduler

Scheduling in action

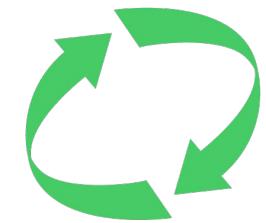


KubeCon



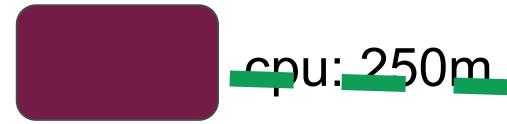
CloudNativeCon

Europe 2019



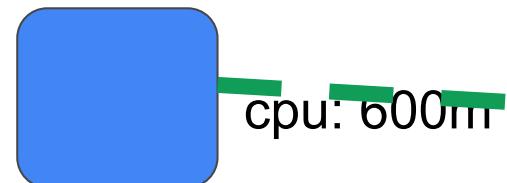
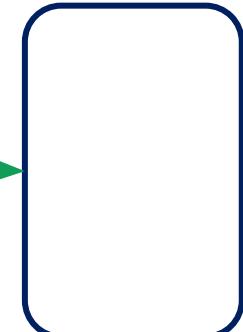
Kubernetes
scheduler

Pods

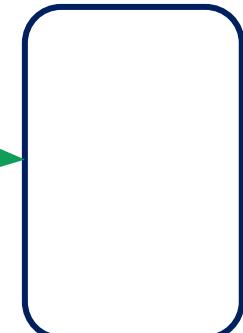


cpu: 250m

Nodes



cpu: 600m



Scheduling in action



KubeCon



CloudNativeCon

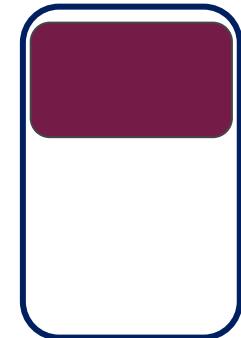
Europe 2019

Pods

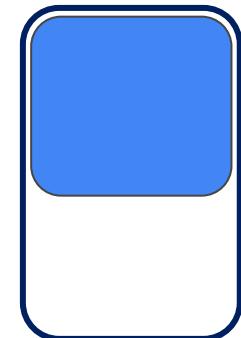
Nodes



Kubernetes
scheduler



cpu: 900m



cpu: 900m

Scheduling in action



KubeCon



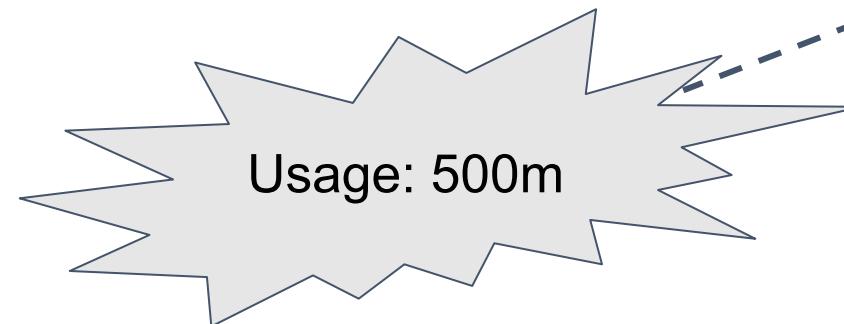
CloudNativeCon

Europe 2019

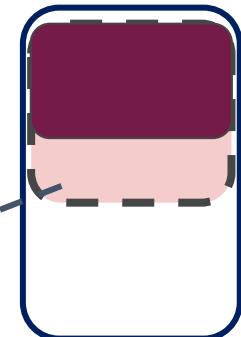


Kubernetes
scheduler

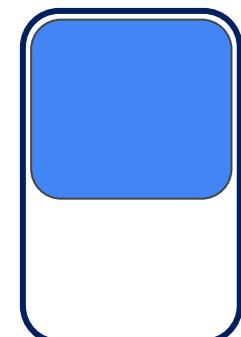
Pods



Nodes



cpu: 900m



cpu: 900m

Scheduling in action

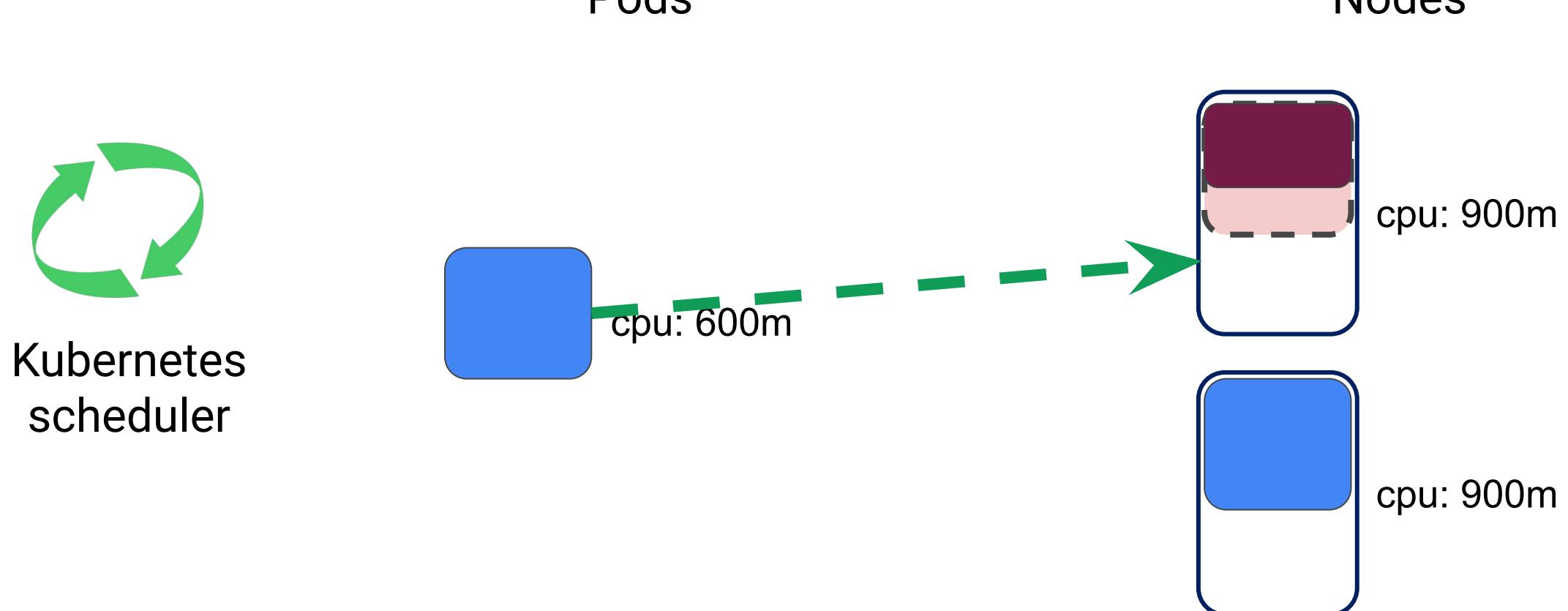


KubeCon



CloudNativeCon

Europe 2019



Scheduling in action

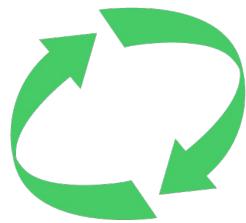


KubeCon



CloudNativeCon

Europe 2019

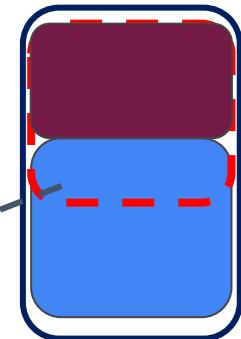


Kubernetes
scheduler

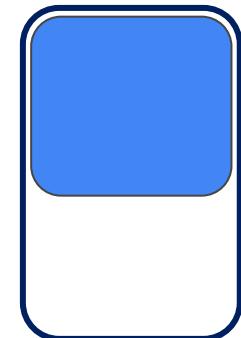
Pods



Nodes



cpu: 900m



cpu: 900m

Always up-to-date!



KubeCon

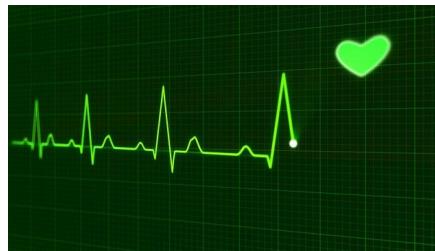


CloudNativeCon

Europe 2019

Vertical Pod Autoscaler

Observes
usage



Recommends
resources



Updates
resources

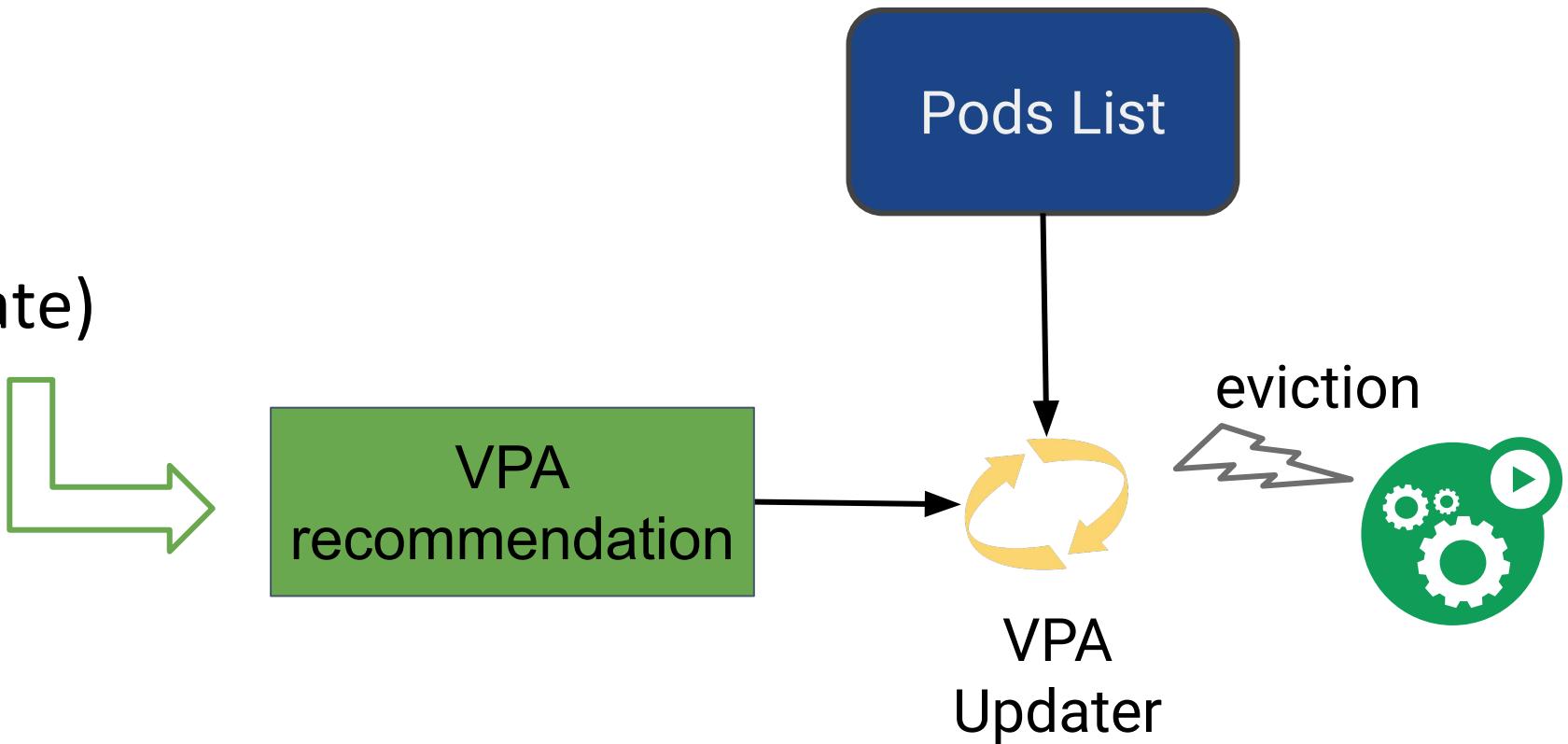


VPA updater modes



Europe 2019

- ▶ Off
- ▶ Initial
- ▶ Auto (Recreate)



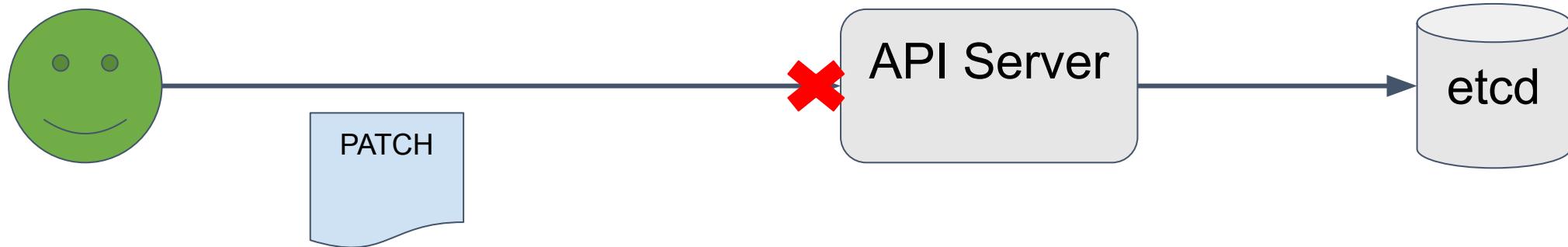
Why though?



KubeCon

CloudNativeCon

Europe 2019



```
apiVersion: core/v1
kind: Pod
...
spec:
  containers:
    - name: worker
  resources:
    requests:
      memory: "300M"
      cpu: "500m"
```

```
apiVersion: core/v1
kind: Pod
...
spec:
  containers:
    - name: worker
  resources:
    requests:
      memory: "100M"
      cpu: "250m"
```

One Simple Trick



KubeCon



CloudNativeCon

Europe 2019

The problem:



immutable PodSpec

One Simple Trick



KubeCon



CloudNativeCon

Europe 2019

The solution:

- ▶ **mutable** PodSpec,
- ▶ update resources **in-place**.

Caveat:

- ▶ a **breaking** change.



Fixed-Size Containers



KubeCon

CloudNativeCon

Europe 2019

```
apiVersion: v1
kind: Pod
...
containers:
- image: my-java-app
  env:
  - name: "JAVA_OPTS"
    value: "-XX:UseContainerSupport"
resources:
  requests:
    memory: "1024M"
  limits:
    memory: "1024M"
```



Two Scenarios

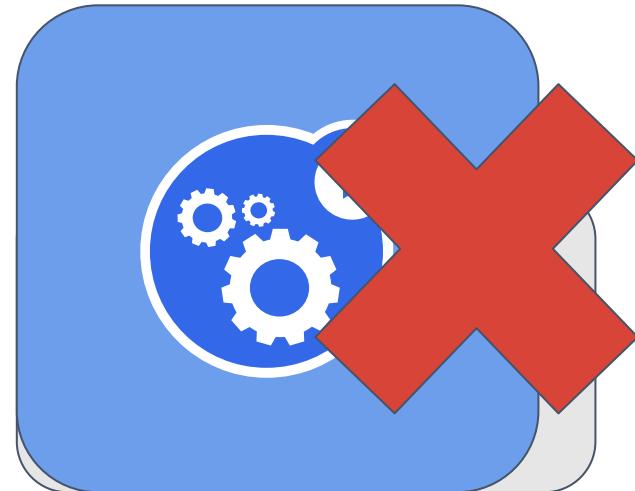
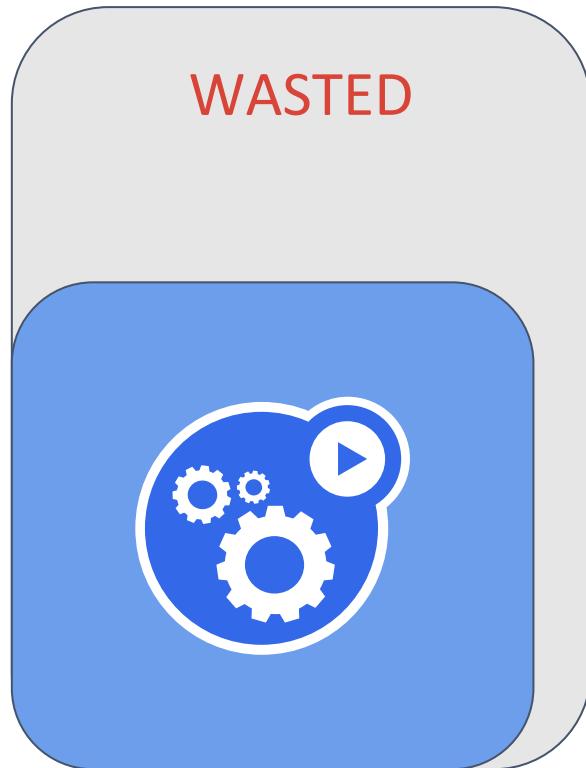
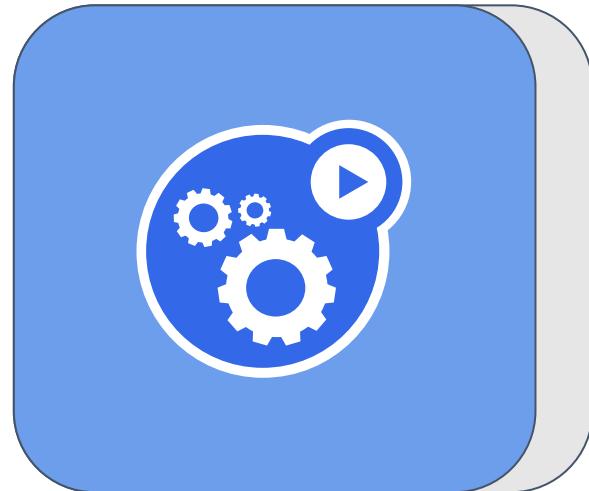


KubeCon



CloudNativeCon

Europe 2019



Downscaling Is (NOT) Easy

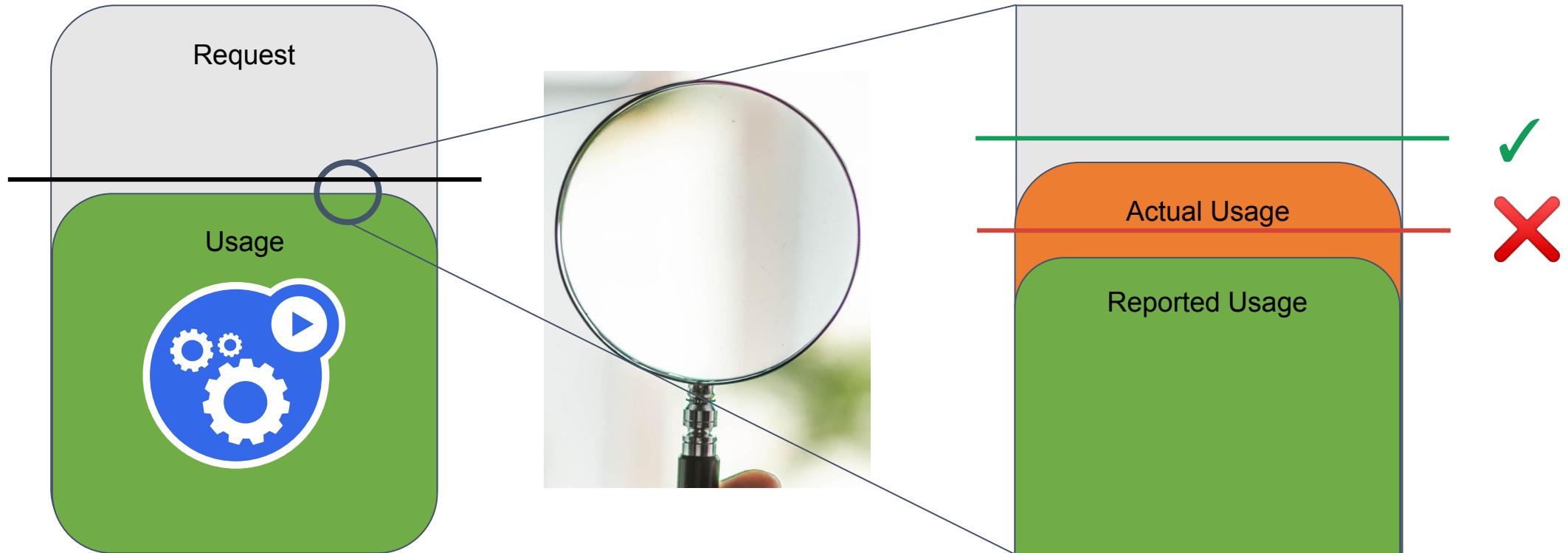


KubeCon



CloudNativeCon

Europe 2019



Change Is Coming



Europe 2019



Scheduler

- cache
- extra work



Kubelet



Quota

KEP - Main Change



Europe 2019

- ▶ PodSpec.Container.ResourceRequirements
 - **mutable**
 - **desired** state
- ▶ PodSpec.Container.ResourceAllocations
 - vetted state
- ▶ PodStatus.ContainerStatus.ResourcesAllocated
 - **actual** state
- ▶ new PodConditions
 - PodResizing
 - PodResizeSuccess

PENDING APPROVAL

Flow Control

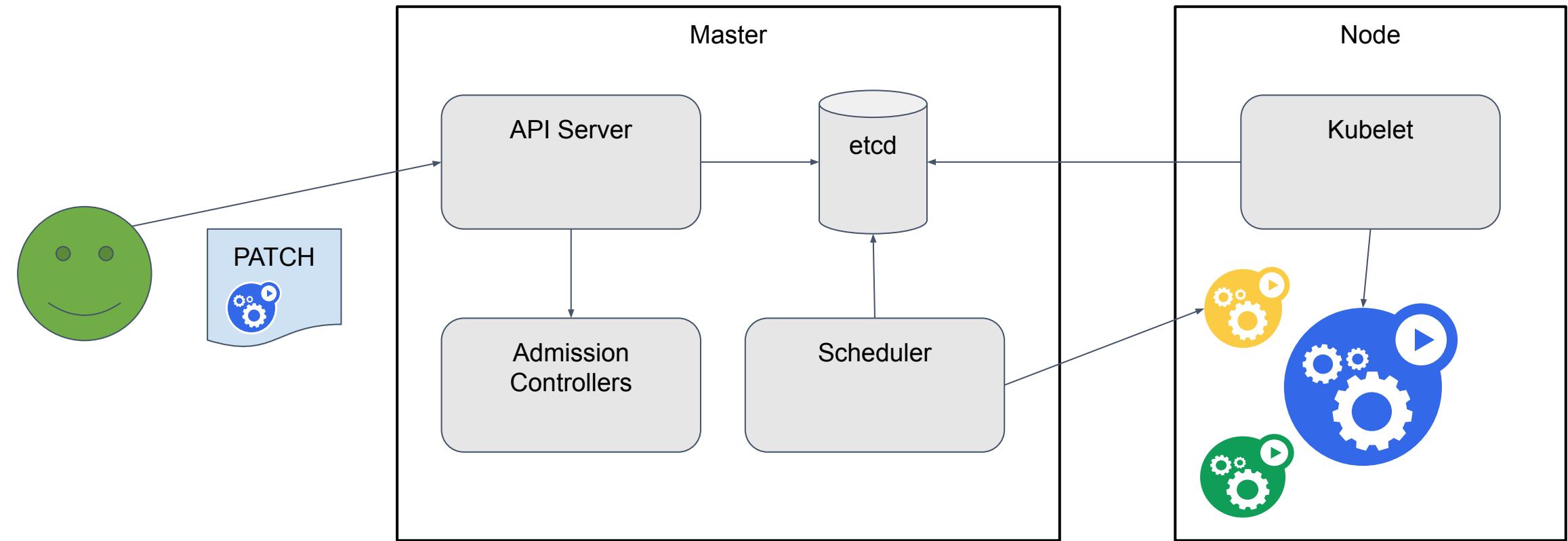


KubeCon



CloudNativeCon

Europe 2019



One Little Knob



KubeCon



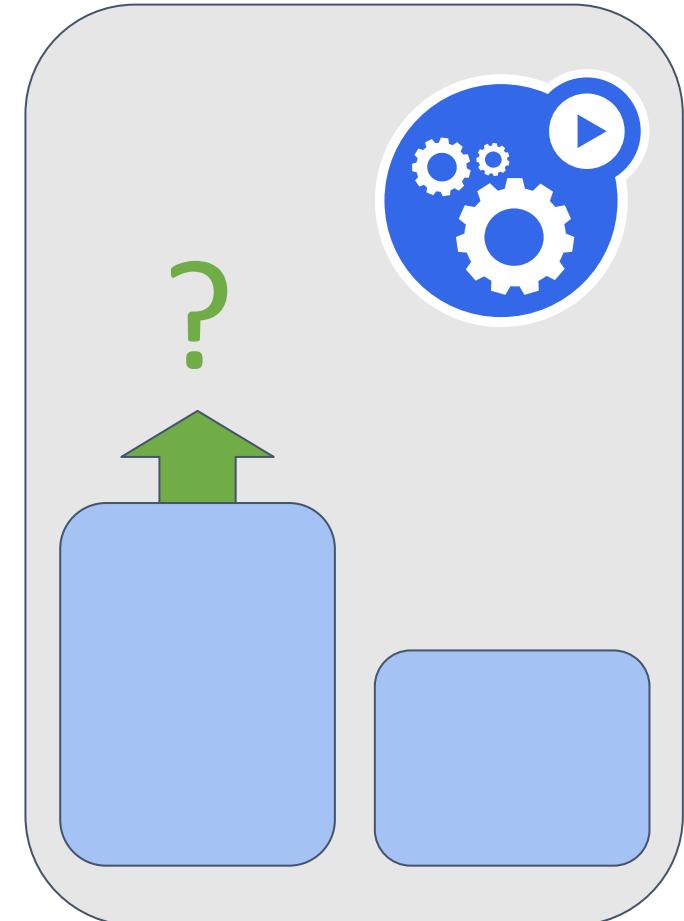
CloudNativeCon

Europe 2019



PodSpec.Container.ResizePolicy

- NoRestart - **default**
- RestartContainer
- RestartPod



Two Little Knobs



KubeCon



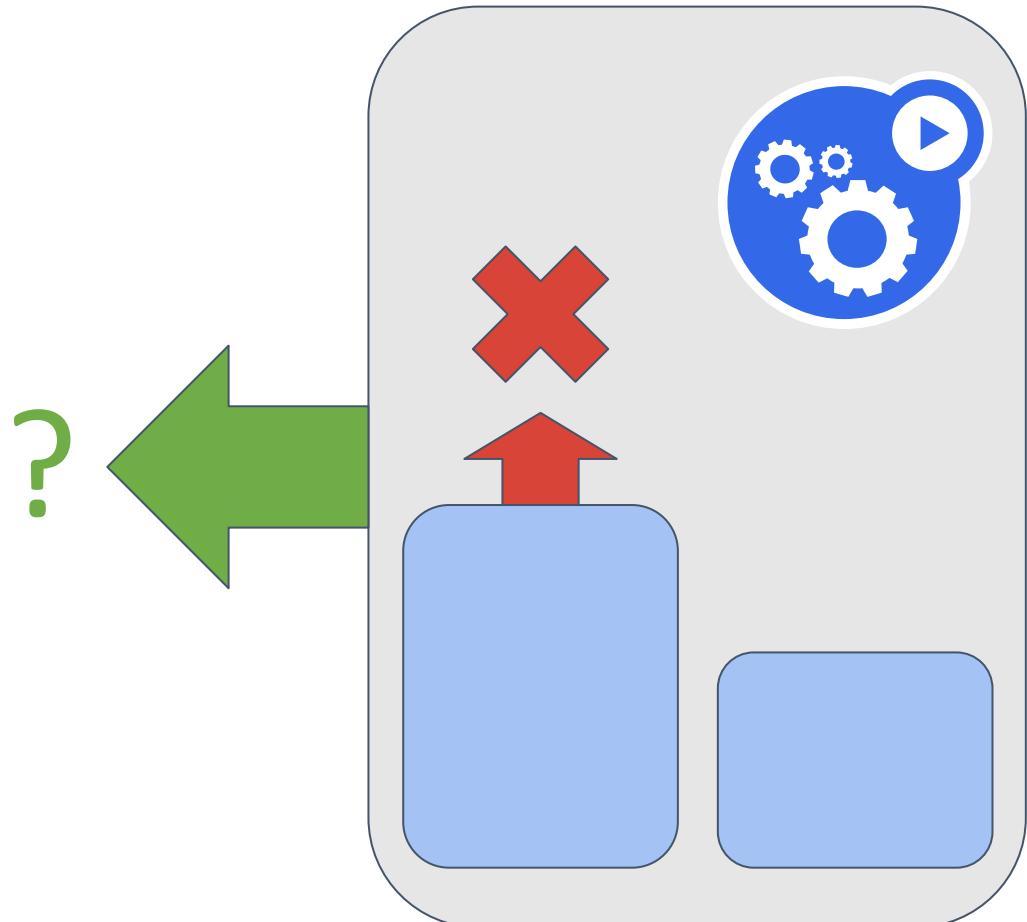
CloudNativeCon

Europe 2019



PodSpec.RetryPolicy

- NoRetry
- RetryUpdate
- Reschedule



Best of the two - eating the cake



KubeCon



CloudNativeCon

Europe 2019

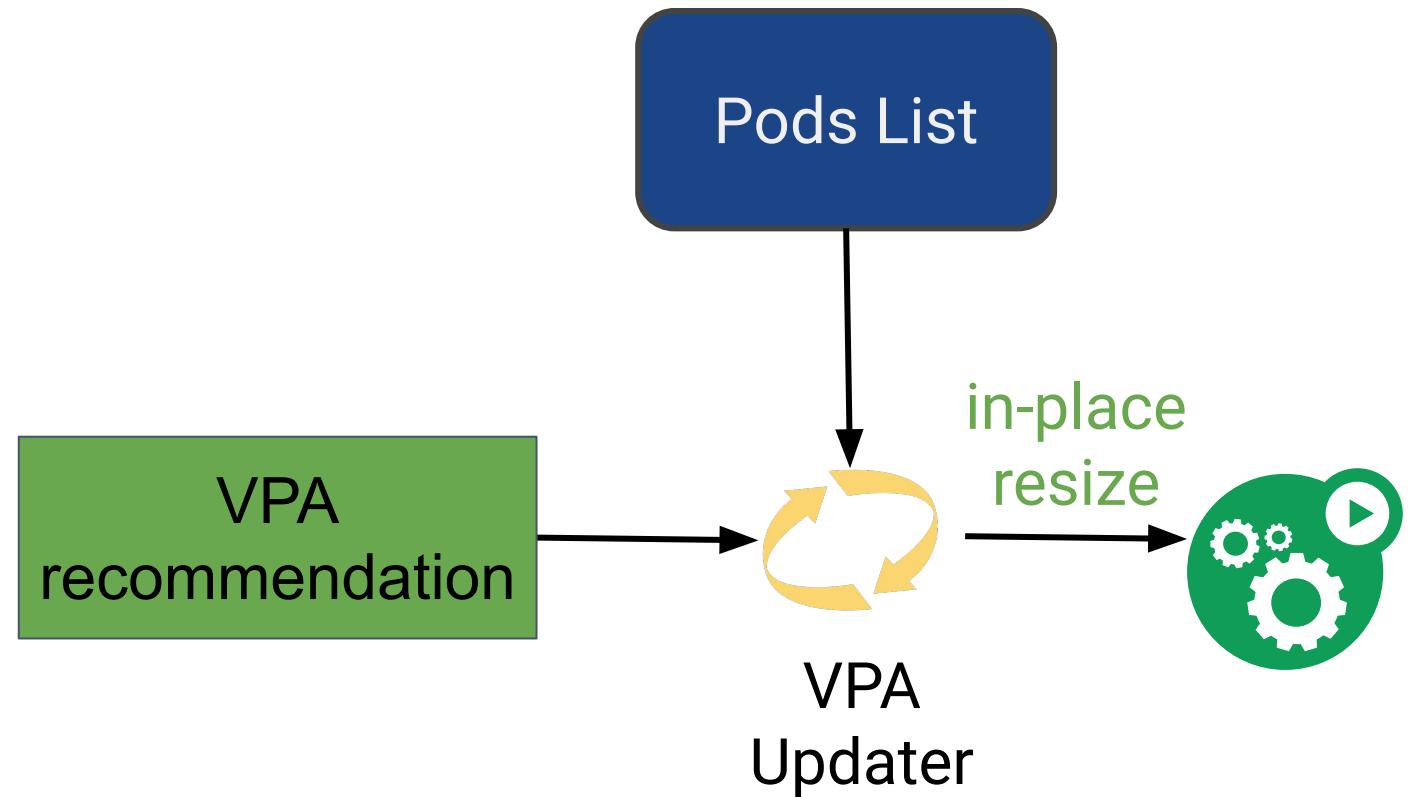
► Minimized disruption

+

► Better stability with up-to-date
resource requests



- ▶ Off
- ▶ Initial
- ▶ Recreate
- ▶ In-place only
- ▶ In-place preferred
(Auto)



Kubernetes Enhancement Proposal:

<https://github.com/kubernetes/enhancements/pull/686>

Thank you!



KubeCon



CloudNativeCon

Europe 2019

170+ comments!

Vinay Kulkarni wants to merge 11 commits into `kubernetes:master` from `vinaykul:master`

Conversation 174 Commits 11 Checks 0 Files changed 1 +423 -0

Vinay Kulkarni commented on 12 Jan First-time contributor

This PR moves Karol's [initial proposal](#) for in-place update of pod resources from k/community to k/enhancements as required by the new process.

This KEP intends to build upon the ideas in [proposal for live and in-place vertical scaling](#) and [Vertical Resources Scaling in Kubernetes](#).

This PR also updates the owning-sig to sig-autoscaling, and adds initial set of reviewers Babak "Bobby" Salamat , Derek Carr , Dawn Chen from sig-scheduling and sig-node where bulk of the anticipated code changes will happen.

The original pull request by Karol, and associated discussion is [here](#).

CC: Karol Gołęb Beata Skiba @schylek Babak "Bobby" Salamat Dawn Chen Derek Carr @karthickrajamani Youngjae Lee Harry Zhang

15+ reviewers!

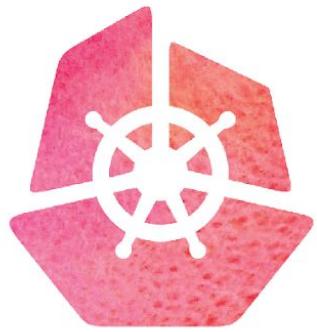
Reviewers

- ConnorDoyle
- kgolab
- jellonek
- hex108
- resouer
- dashpole
- ScorpioCPH
- derekwaynecarr
- PatrickLang
- bsalamat
- krmayankk
- schylek
- justaugustus
- DirectXMan12
- sjenning
- mwielgus

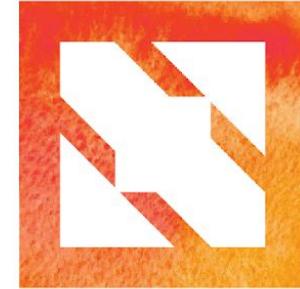
Kubernetes Prow... commented on 12 Jan Member

Thanks for your pull request. Before we can look at your pull request, you'll need to sign a Contributor License Agreement (CLA).

Please follow instructions at <https://git.k8s.io/community/CLA.md#the-contributor-license-agreement> to sign the CLA.



KubeCon



CloudNativeCon

Europe 2019