

Scale your (AKS) cluster, Luke!

Alessandro Melchiori

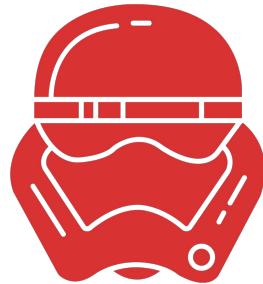
Founder & Software architect @ CodicePlastico
alessandro@codiceplastico.com | @amelchiori

Container &
DevOps Day



aspitalia.com

Talking about



SCALING PODS



SCALING NODES

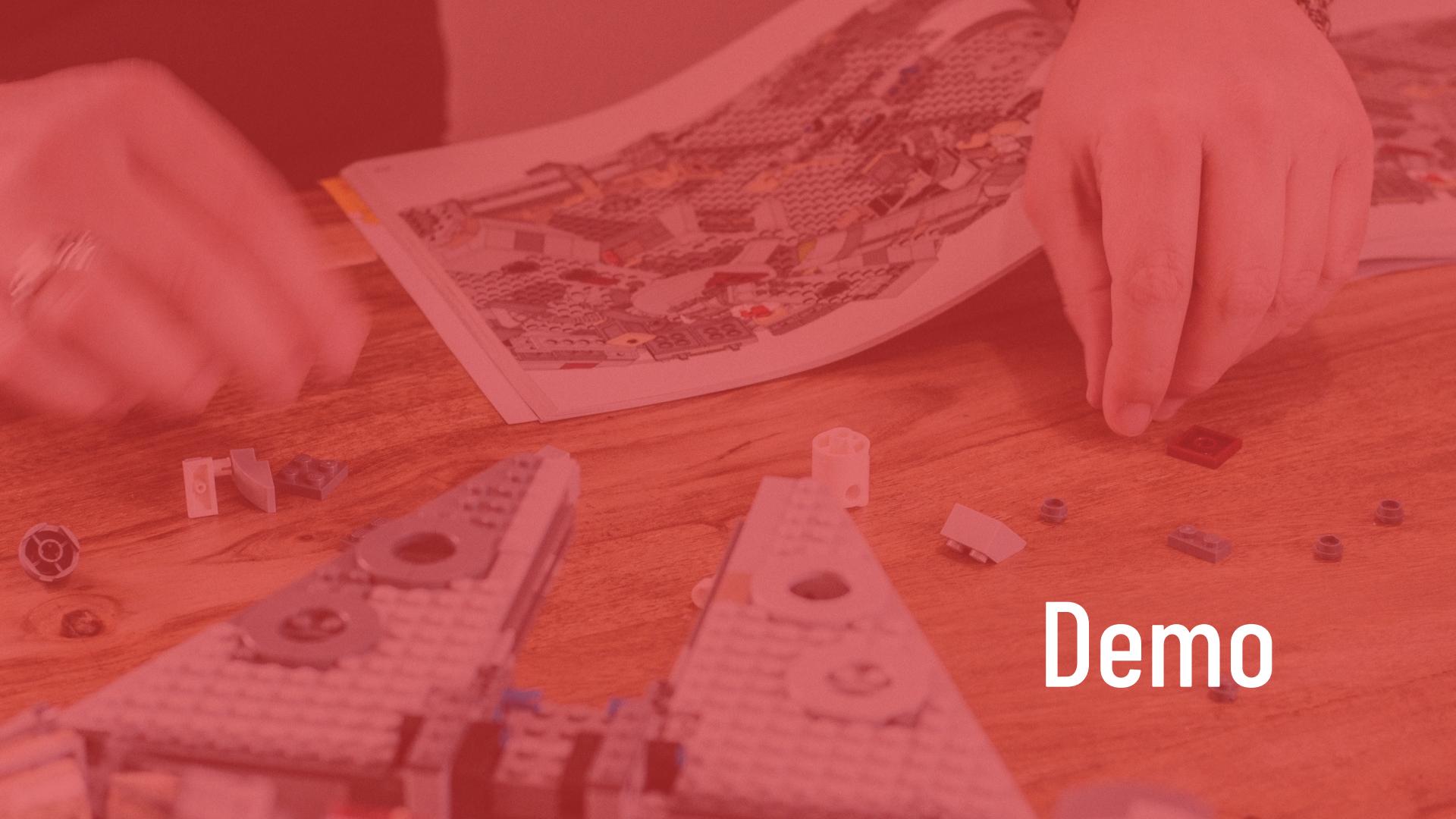




Scaling Pods



Manual Scaling



Demo

Metrics Server is a **scalable, efficient** source of container resource metrics for Kubernetes built-in autoscaling pipelines.

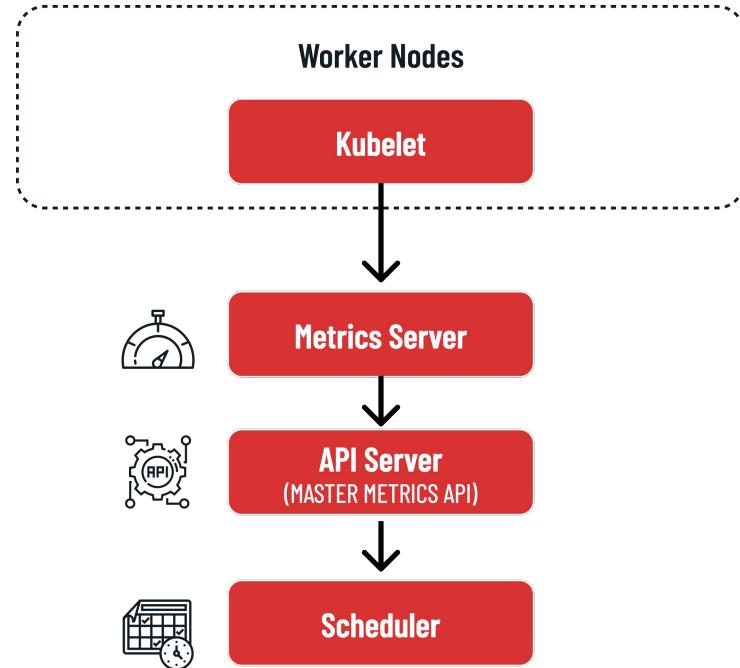
Metrics Server is not meant for non-autoscaling purposes.

For example, don't use it to forward metrics to monitoring solutions, or as a source of monitoring solution metrics.

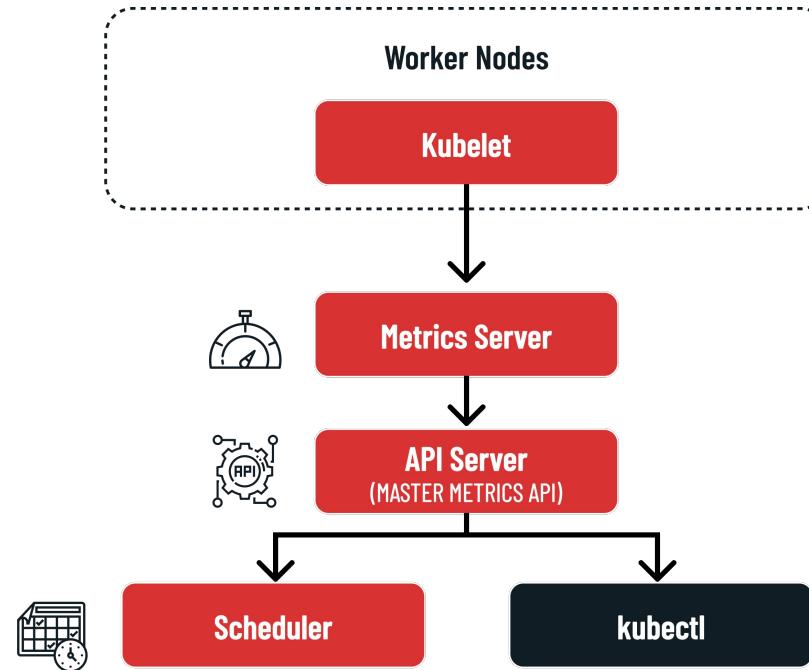
Metric Server



Metric server

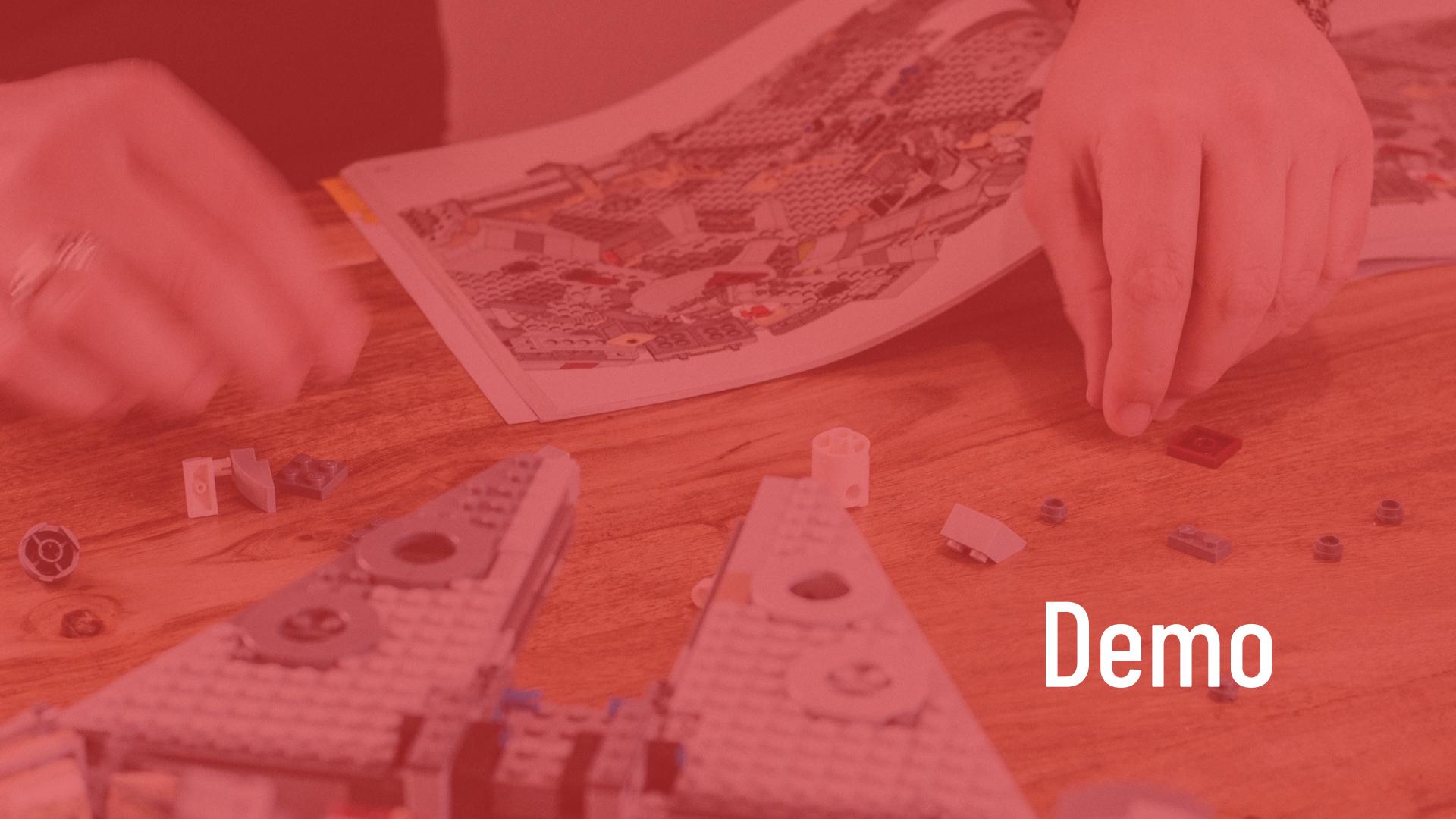


Metric server (kubectl top)



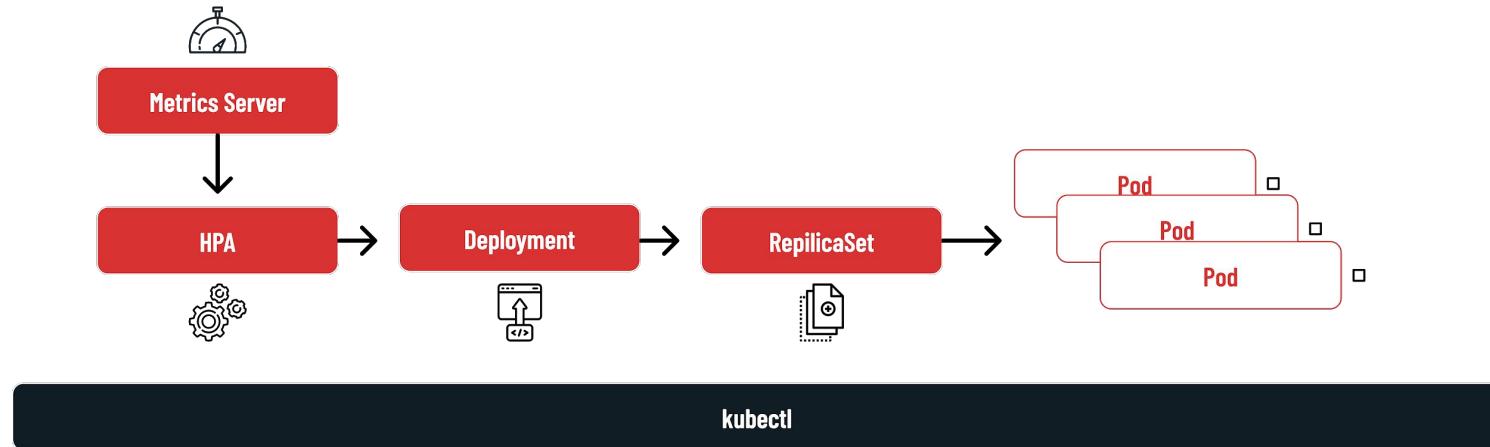


```
kubectl get --raw "/apis/metrics.k8s.io/v1beta1" | jq '..'
```



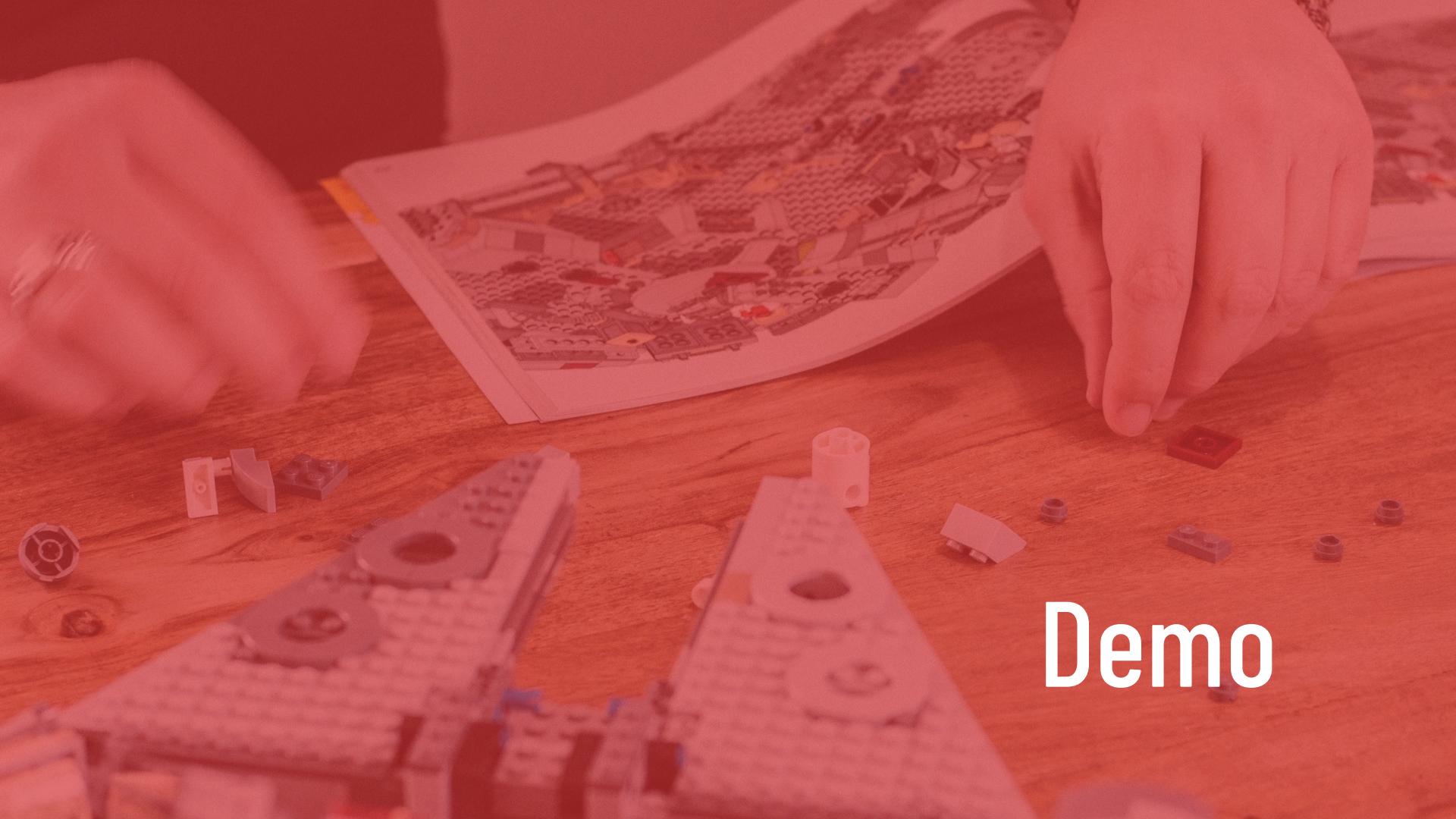
Demo

Horizontal Pod Autoscaler





HPA



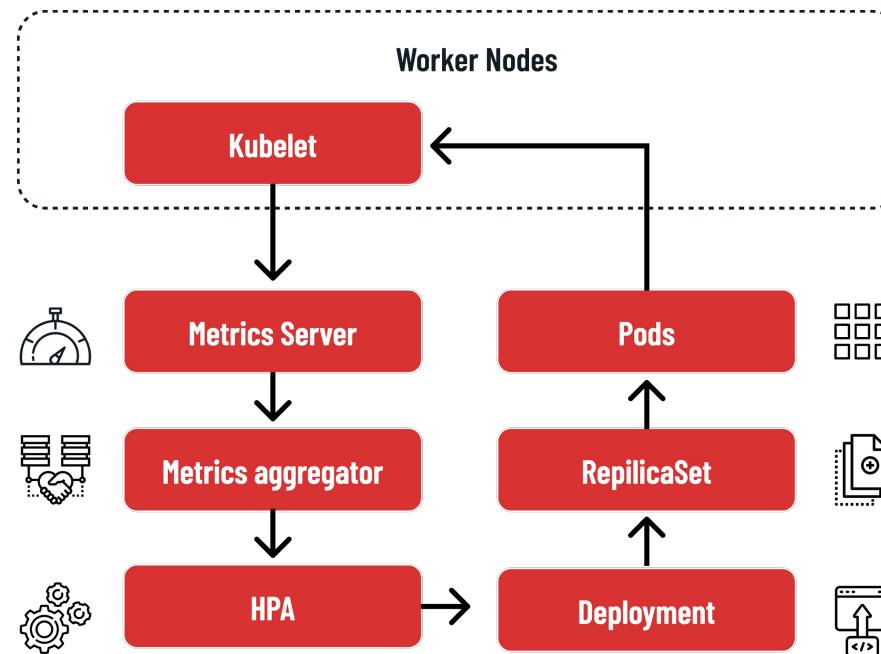
Demo

A Stormtrooper LEGO minifigure is standing on the right side of the image. It is holding a red speech bubble that contains the text "Custom Metrics".

Custom Metrics

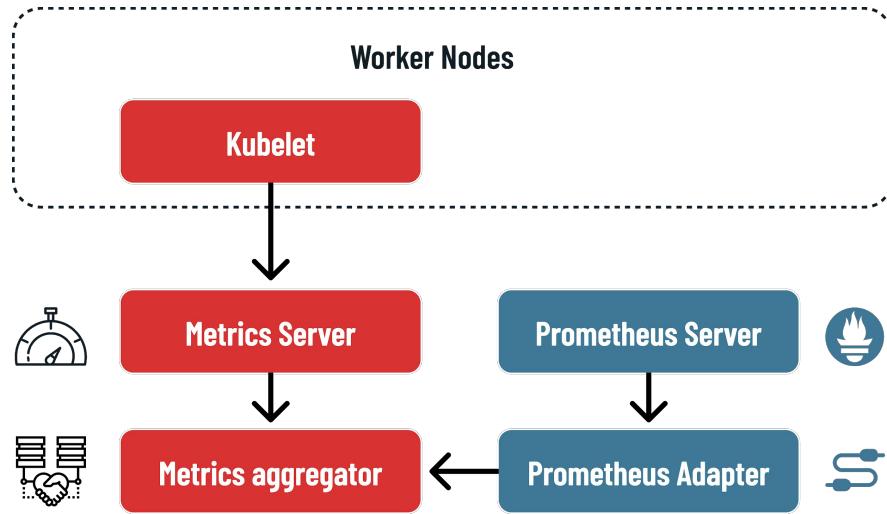
```
kubectl get --raw "/apis/custom.metrics.k8s.io/v1beta1" | jq '..'
```

Horizontal Pod Autoscaler (HPA)

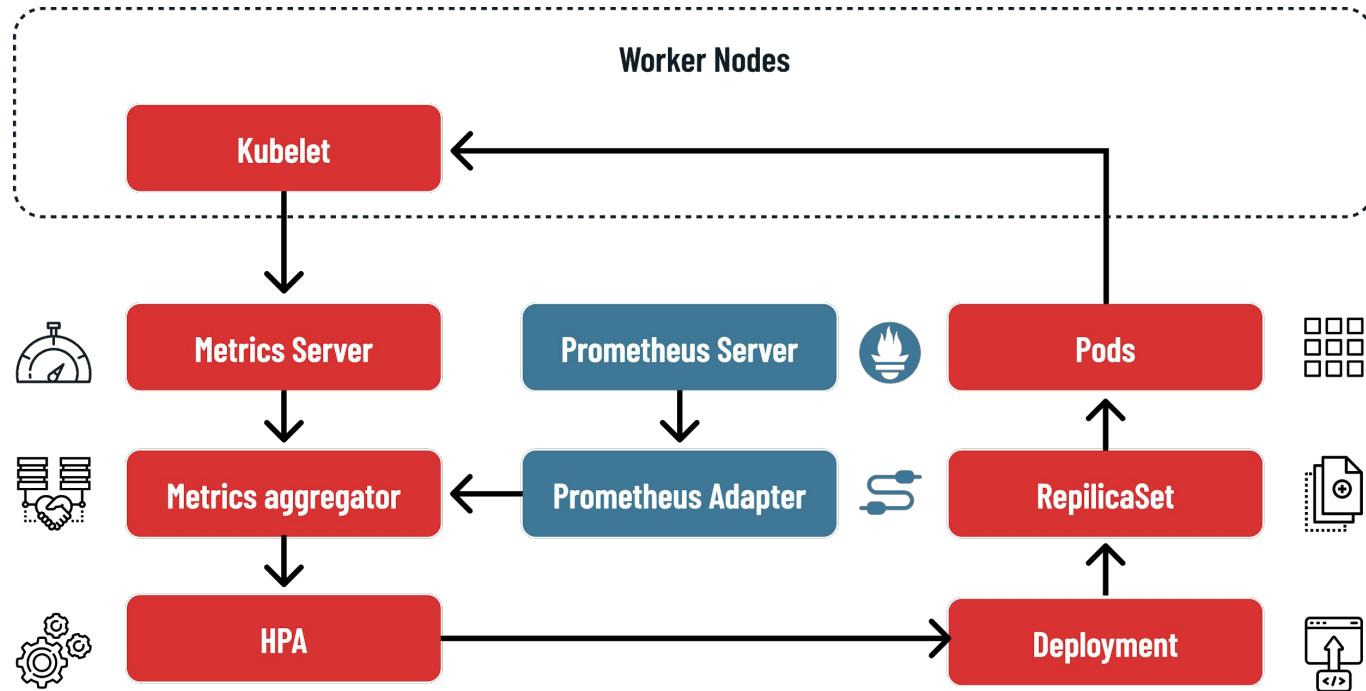


Custom Metrics

From now on, **Metrics Aggregator** contains not only the data from the **Metrics Server** but also those from the **Prometheus Adapter** (for example) which, in turn, is fetching metrics from **Prometheus Server**.



HPA





AGENT

KEDA activates and deactivates Kubernetes Deployments **to scale to and from zero on no events.**

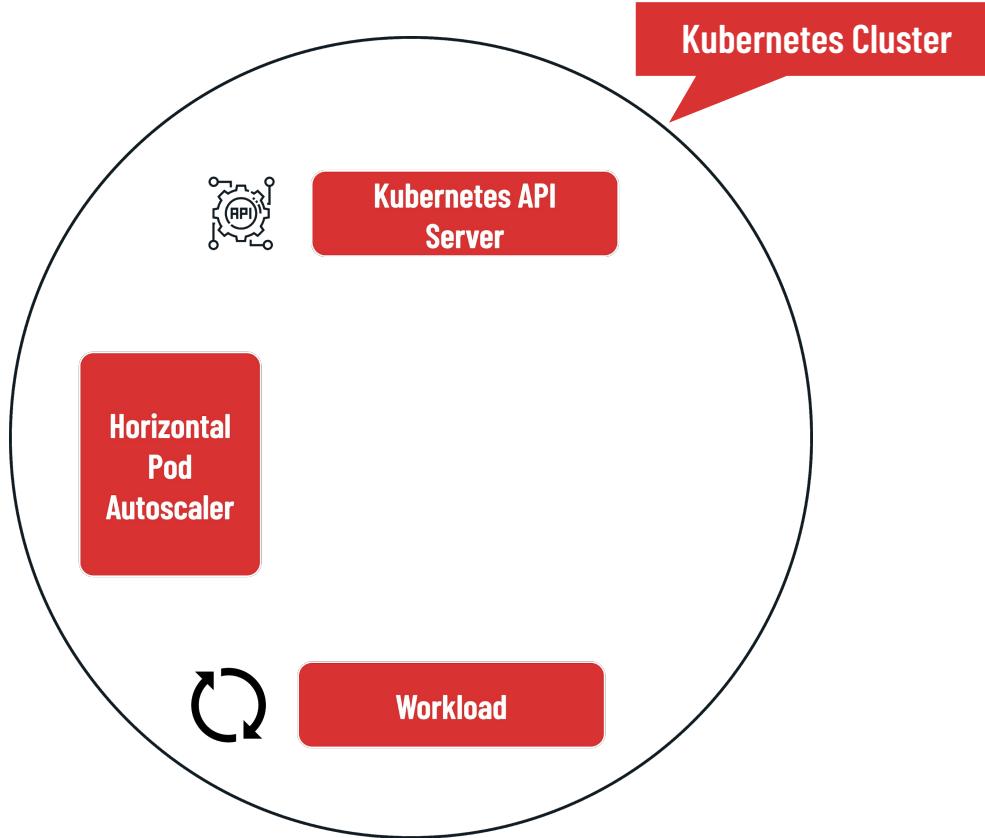


Kubernetes Event-driven Autoscaling



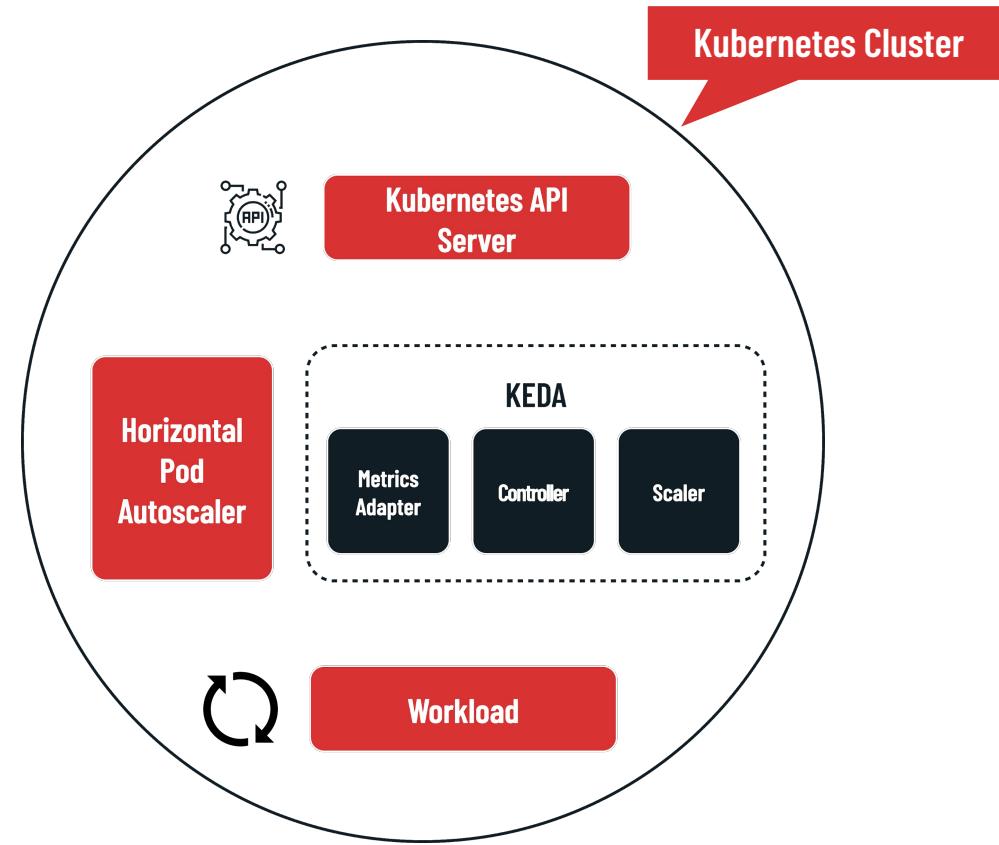
METRICS

KEDA acts as a Kubernetes metrics server that exposes **rich event data** like queue length or stream lag to the Horizontal Pod Autoscaler to drive scale out



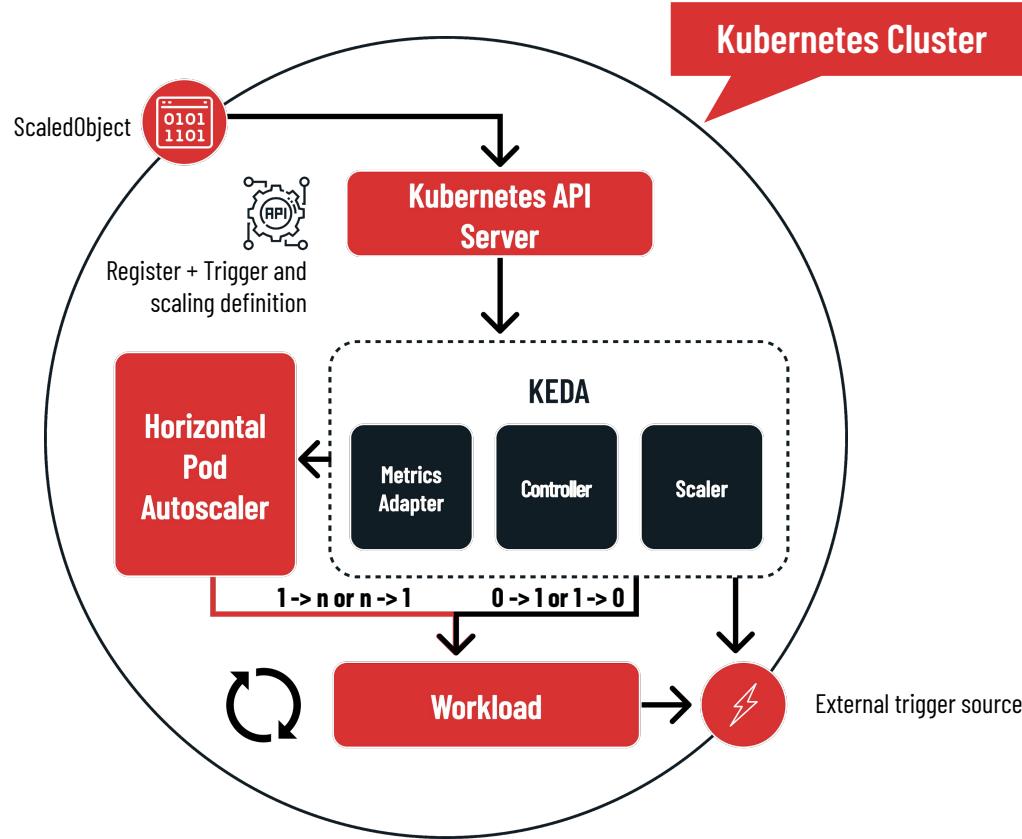
Keda architecture





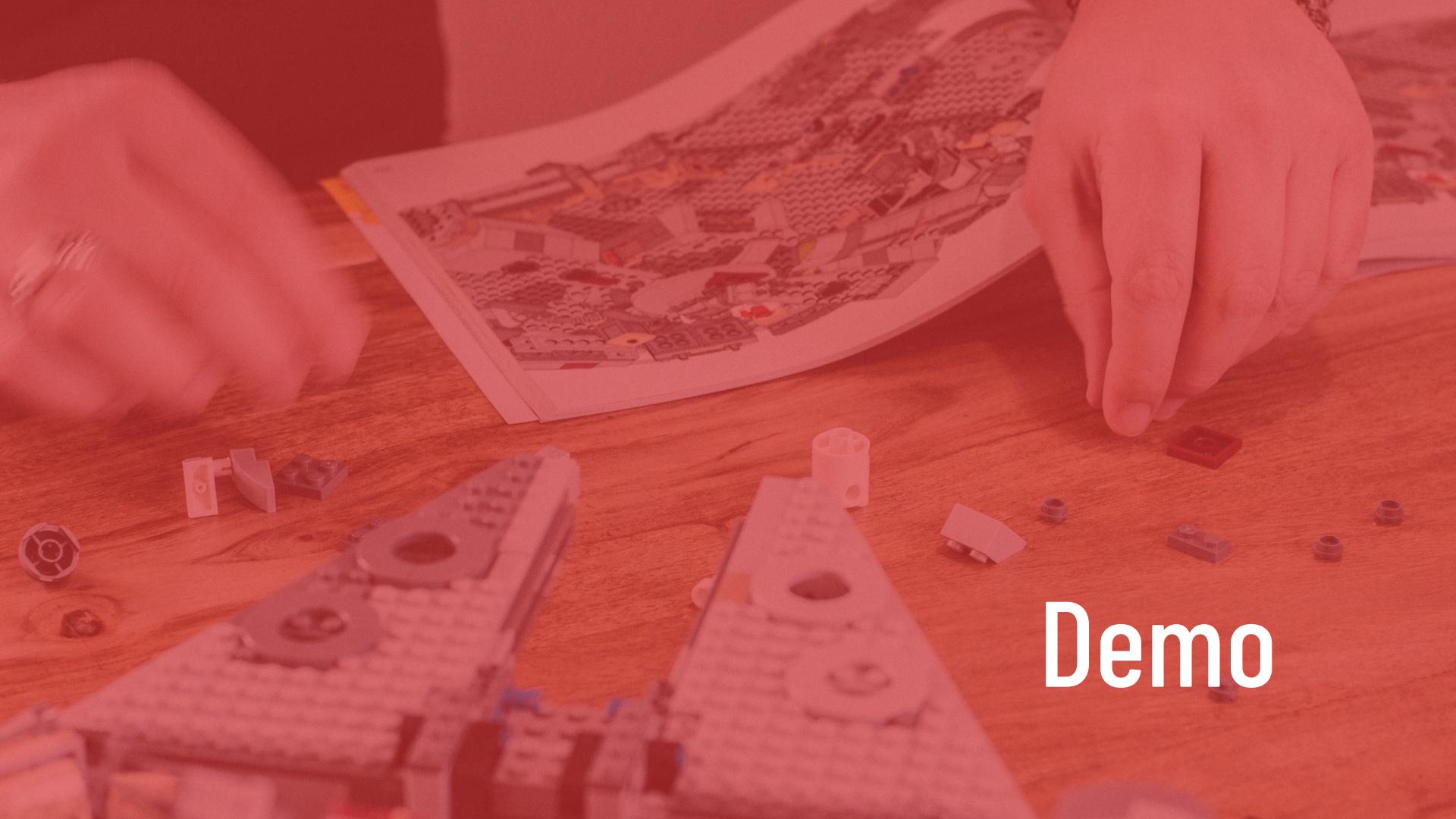
Keda architecture





Keda architecture





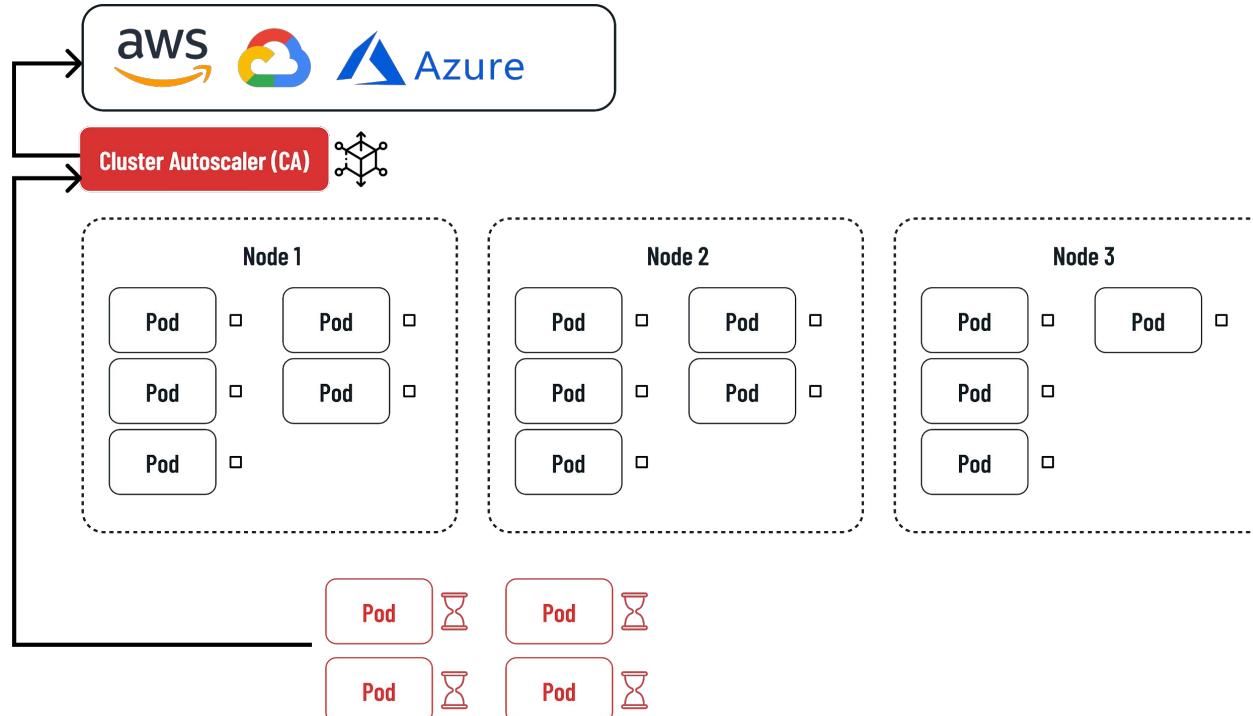
Demo



Scaling Nodes



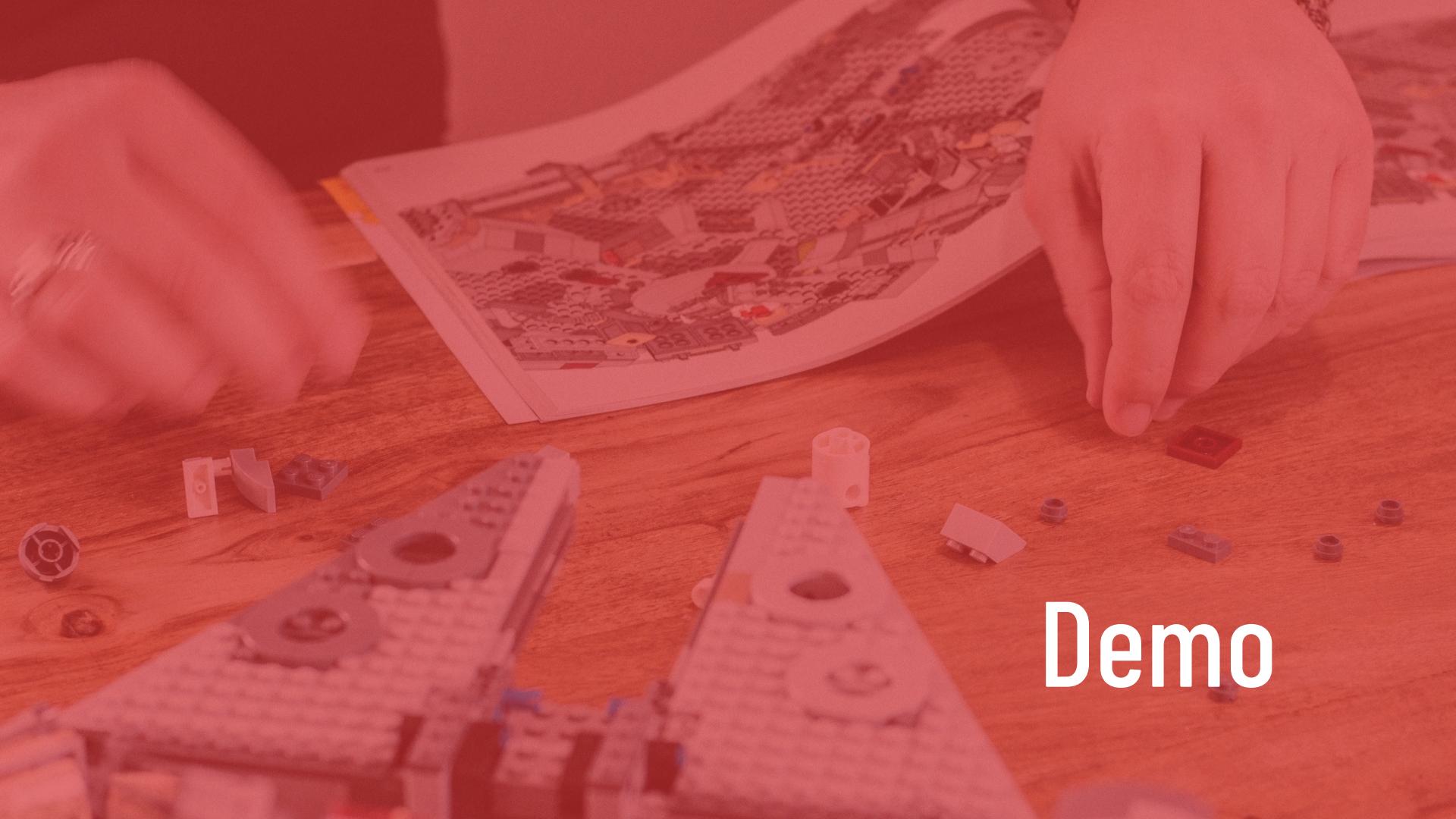
Cluster Autoscaler



```
az aks update \  
  --enable-cluster-autoscaler \  
  --min-count 1 --max-count 3 \  
  ...
```

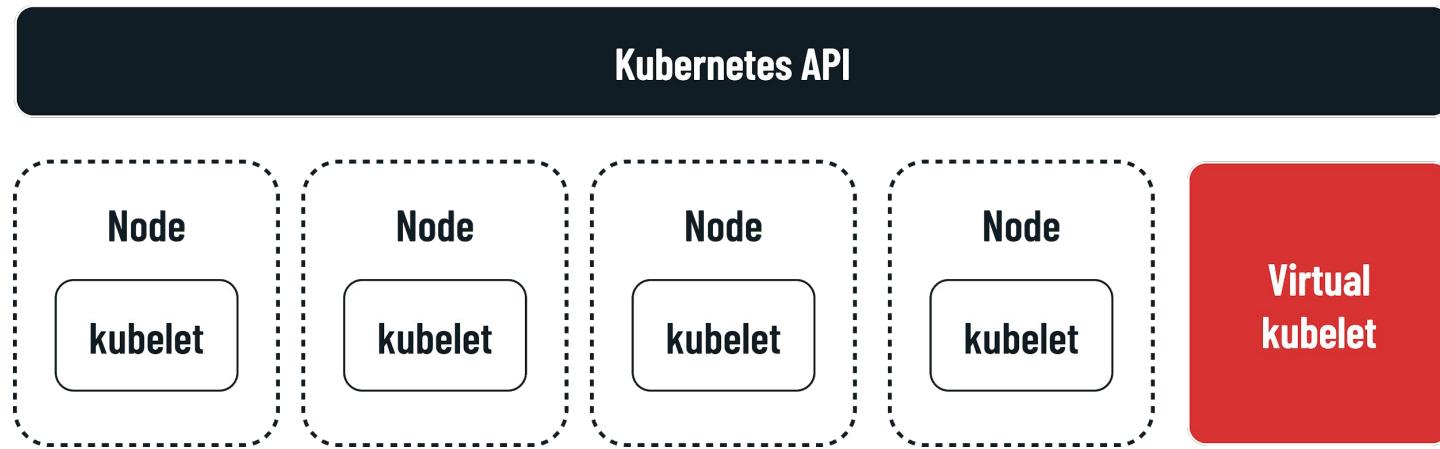


Cluster Autoscaler



Demo

Virtual Kubelet





Grazie!

FOLLOW US



www.codiceplastico.com



@amelchiori
alessandro@codiceplastico.com

Slide e materiale su
<https://aspit.co/ContainerDevOpsDay-21>

Container &
DevOps Day