



KubeCon



CloudNativeCon

Europe 2019

Sig-Autoscaling Deep Dive

Aleksandra Malinowska, Google
Stawomir Chytek, Google





KubeCon



CloudNativeCon

Europe 2019

Vertical Pod Autoscaler

Vertical Pod Autoscaler



KubeCon



CloudNativeCon

Europe 2019

Do you set pod resource requests?

Vertical Pod Autoscaler



KubeCon



CloudNativeCon

Europe 2019

Are you confident your pod requests are correct?

Vertical Pod Autoscaler



KubeCon



CloudNativeCon

Europe 2019

Are you confident your pod requests are correct over time?

Vertical Pod Autoscaler



KubeCon



CloudNativeCon

Europe 2019

How Vertical Pod Autoscaler helps you?

- Hands free resource adjustments
- Save money
- Buy reliability

Vertical Pod Autoscaler



KubeCon



CloudNativeCon

Europe 2019

Recommended talk:

Resize Your Pods w/o Disruptions aka How to Have a Cake and Eat a Cake

Karol Gołqb & Beata Skiba

VPA — Components

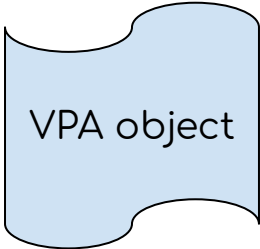


KubeCon



CloudNativeCon

Europe 2019

A light blue, irregularly shaped box with a black outline, representing a VPA object. The text "VPA object" is centered inside the box.

VPA object

VPA — Components

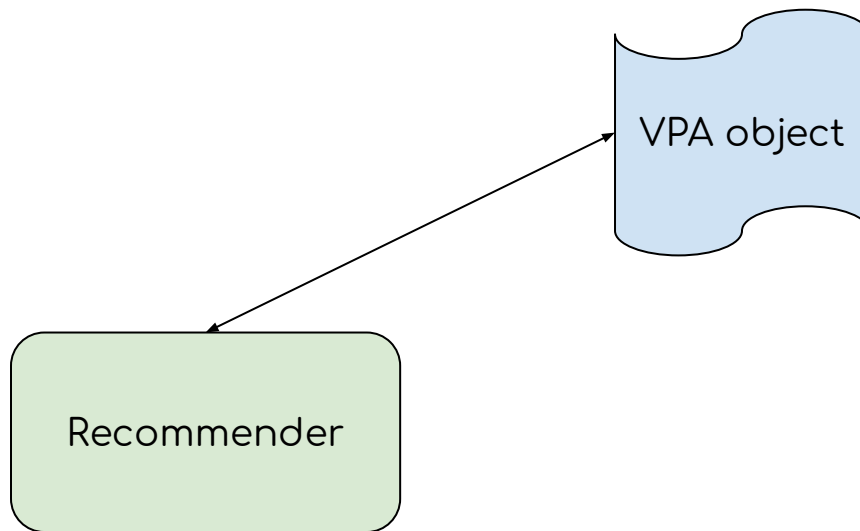


KubeCon



CloudNativeCon

Europe 2019



Compute
recommendations

VPA — Components

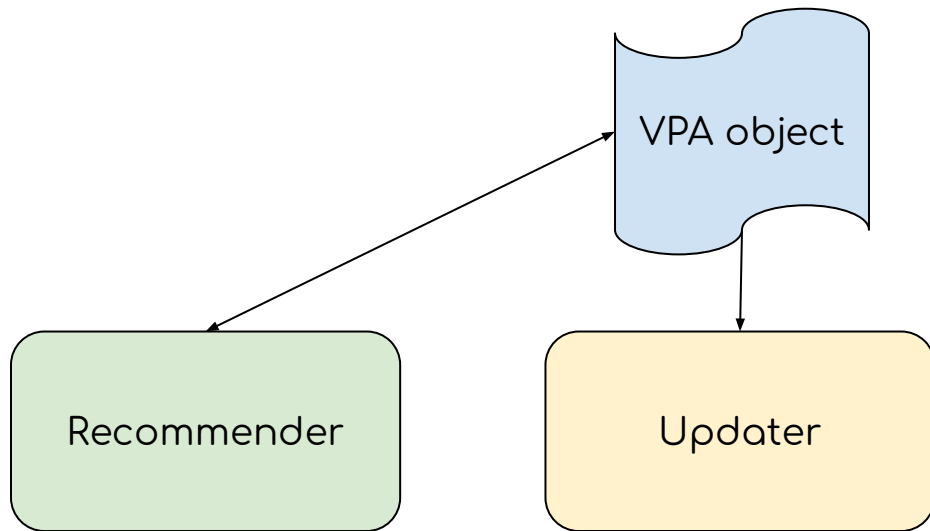


KubeCon



CloudNativeCon

Europe 2019



Compute
recommendations

Select pods to
actuate

VPA — Components

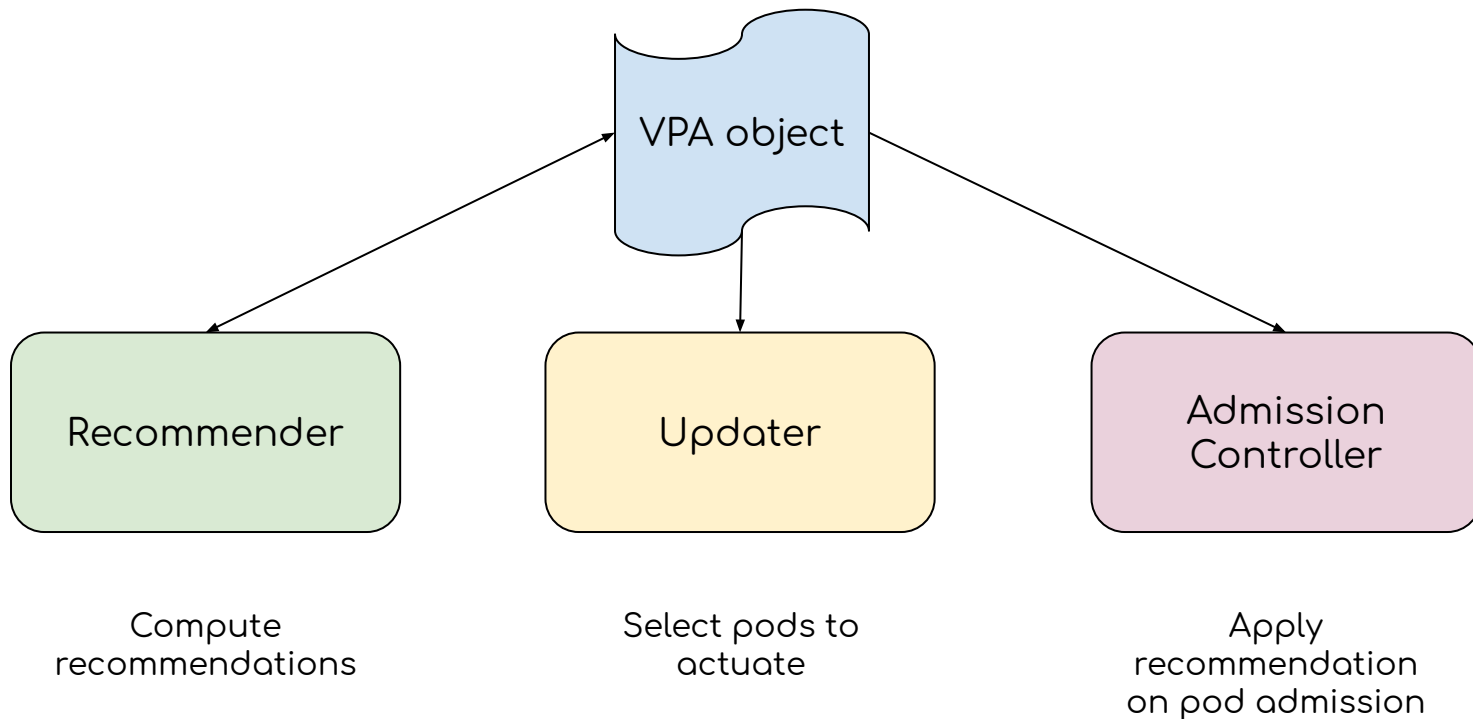


KubeCon



CloudNativeCon

Europe 2019



VPA — API



KubeCon



CloudNativeCon

Europe 2019

- VPA target
 - Deployment, DaemonSet, StatefulSet, Job
 - Everything with Scale sub-resource

VPA — API



KubeCon



CloudNativeCon

Europe 2019

- VPA target
 - Deployment, DaemonSet, StatefulSet, Job
 - Everything with Scale sub-resource
- Per container/resource configuration
 - On/Off
 - Min/Max

VPA — API



KubeCon



CloudNativeCon

Europe 2019

- VPA target
 - Deployment, DaemonSet, StatefulSet, Job
 - Everything with Scale sub-resource
- Per container/resource configuration
 - On/Off
 - Min/Max
- Modes of actuation
 - Off (Recommender)
 - Initial (Recommender + Admission Controller)
 - Auto / Recreate (Recommender + Updater + Admission Controller)

VPA — Recommendations

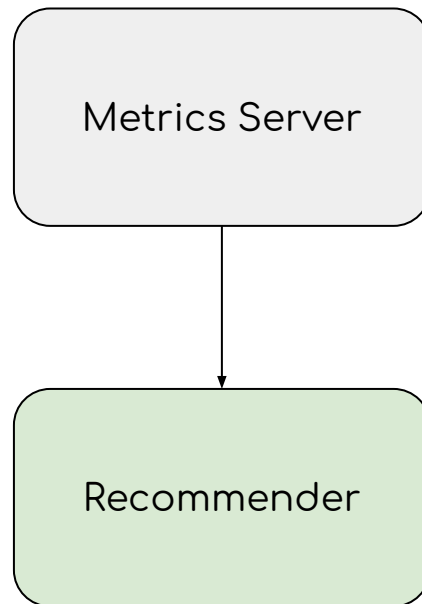


KubeCon



CloudNativeCon

Europe 2019



VPA — Recommendations

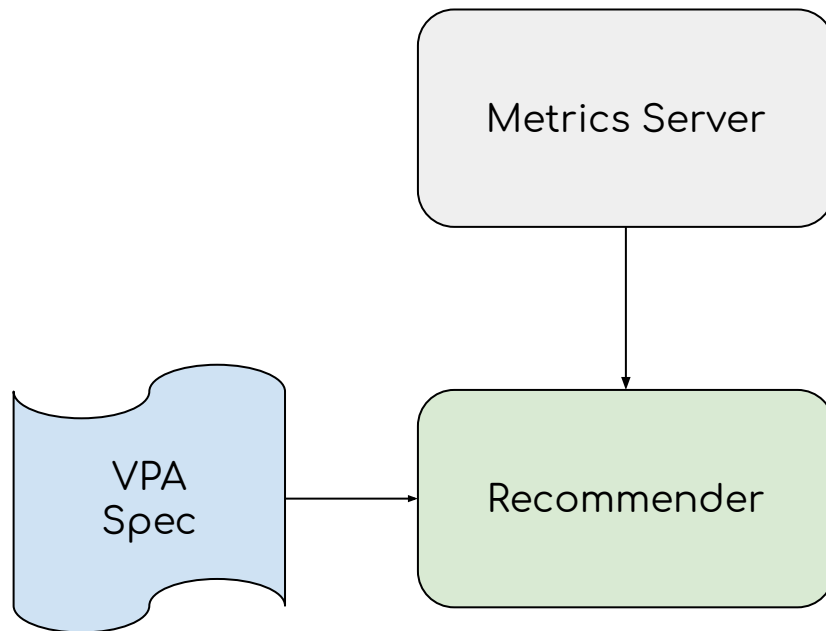


KubeCon



CloudNativeCon

Europe 2019



VPA — Recommendations

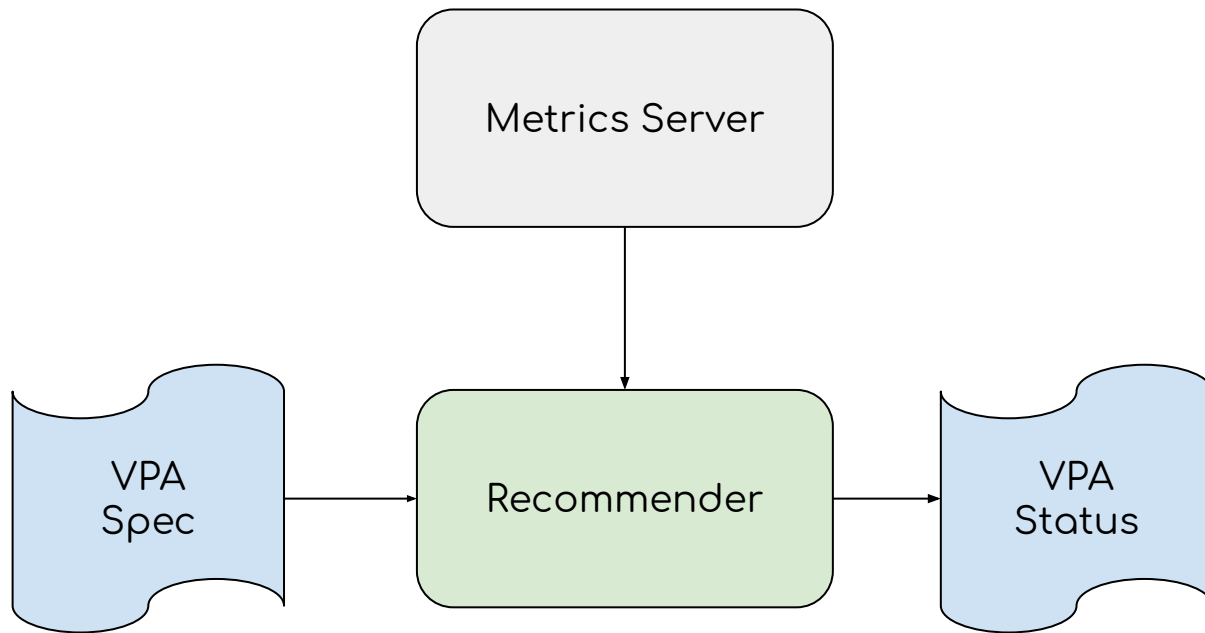


KubeCon



CloudNativeCon

Europe 2019



VPA — Recommendations



KubeCon



CloudNativeCon

Europe 2019

Status:

Recommendation:

Container Recommendations:

Container Name: app

Lower Bound:

Cpu: 381m

Memory: 262144k

Target:

Cpu: 587m

Memory: 262144k

Uncapped Target:

Cpu: 587m

Memory: 262144k

Upper Bound:

Cpu: 141467m

Memory: 2771500k

VPA — Recommendations



KubeCon



CloudNativeCon

Europe 2019

Status:

Recommendation:

Container Recommendations:

Container Name: app

Lower Bound:

Cpu: 381m

Memory: 262144k

Target:

Cpu: 587m

Memory: 262144k

Uncapped Target:

Cpu: 587m

Memory: 262144k

Upper Bound:

Cpu: 141467m

Memory: 2771500k

VPA — Recommendations



KubeCon



CloudNativeCon

Europe 2019

Status:

Recommendation:

Container Recommendations:

Container Name: app

Lower Bound:

Cpu: 381m

Memory: 262144k

Target:

Cpu: 587m

Memory: 262144k

Uncapped Target:

Cpu: 587m

Memory: 262144k

Upper Bound:

Cpu: 141467m

Memory: 2771500k

VPA — Recommendations



KubeCon



CloudNativeCon

Europe 2019

Status:

Recommendation:

Container Recommendations:

Container Name: app

Lower Bound:

Cpu: 381m

Memory: 262144k

Target:

Cpu: 587m

Memory: 262144k

Uncapped Target:

Cpu: 587m

Memory: 262144k

Upper Bound:

Cpu: 141467m

Memory: 2771500k

Actuate when out of range



VPA — Recommendations



- Computing recommendations — as of today:
 - Decaying histogram of weighted samples
 - Newer samples have higher weight (decaying)
 - Cover 8 days of history
 - Recommendation:
 - CPU ~ 90-th percentile
 - Memory ~ max over window
 - Lower/Upper bound
 - Different percentiles (50-th, 95-th)
 - Confidence factor (more samples -> closer to target)
 - Safety margins

VPA — Recommendations



KubeCon



CloudNativeCon

Europe 2019

- Computing recommendations — as of today:
 - Recommendations are silently computed for every possible target
 - Creating VPA object is just surfacing the recommendation
 - Surfaced recommendations are checkpointed
 - To gain restart stability

VPA — Updater

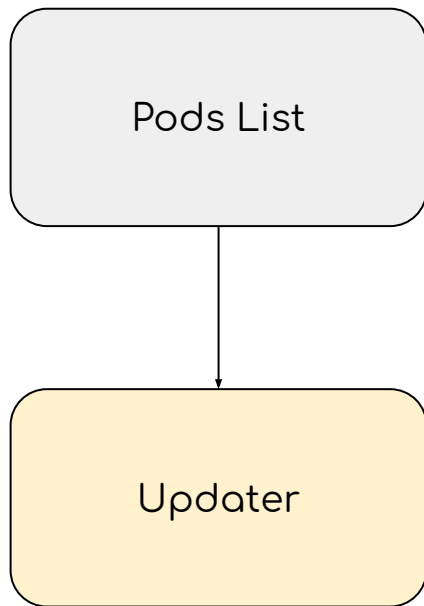


KubeCon



CloudNativeCon

Europe 2019



VPA — Updater

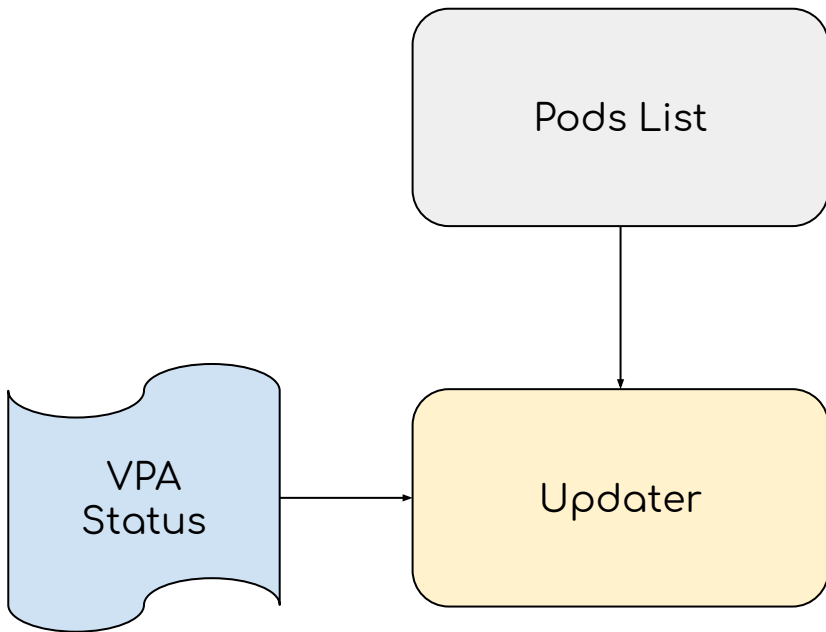


KubeCon



CloudNativeCon

Europe 2019



VPA — Updater

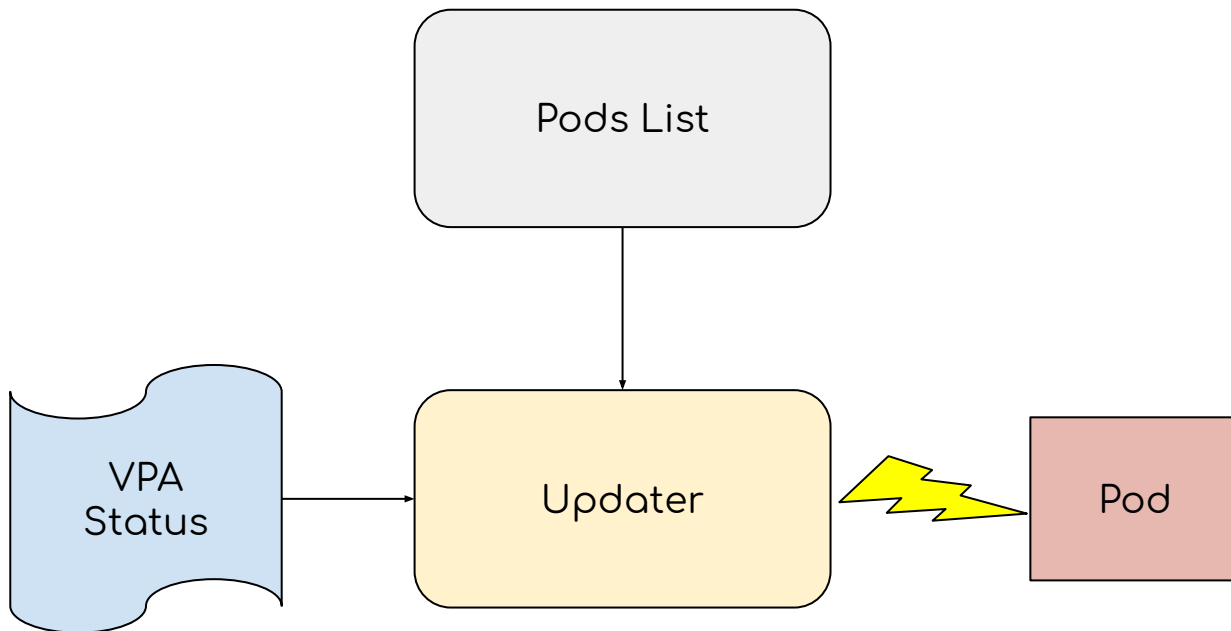


KubeCon



CloudNativeCon

Europe 2019



VPA — Updater



KubeCon



CloudNativeCon

Europe 2019

- Uses eviction API -> Pod Disruption Budget (PDB) is respected
- Additional restrictions:
 - Min number of replicas (default 2)
 - Eviction tolerance (default: 50% of replicas)
- Pod eviction priority
 - Recent OOMs
 - Pods most offending requests

VPA — Admission Controller

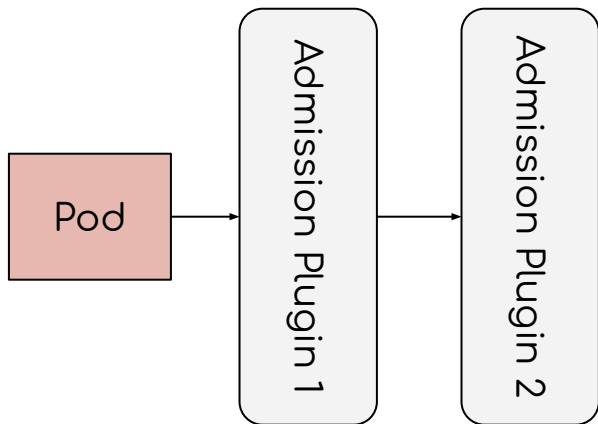


KubeCon



CloudNativeCon

Europe 2019



VPA — Admission Controller

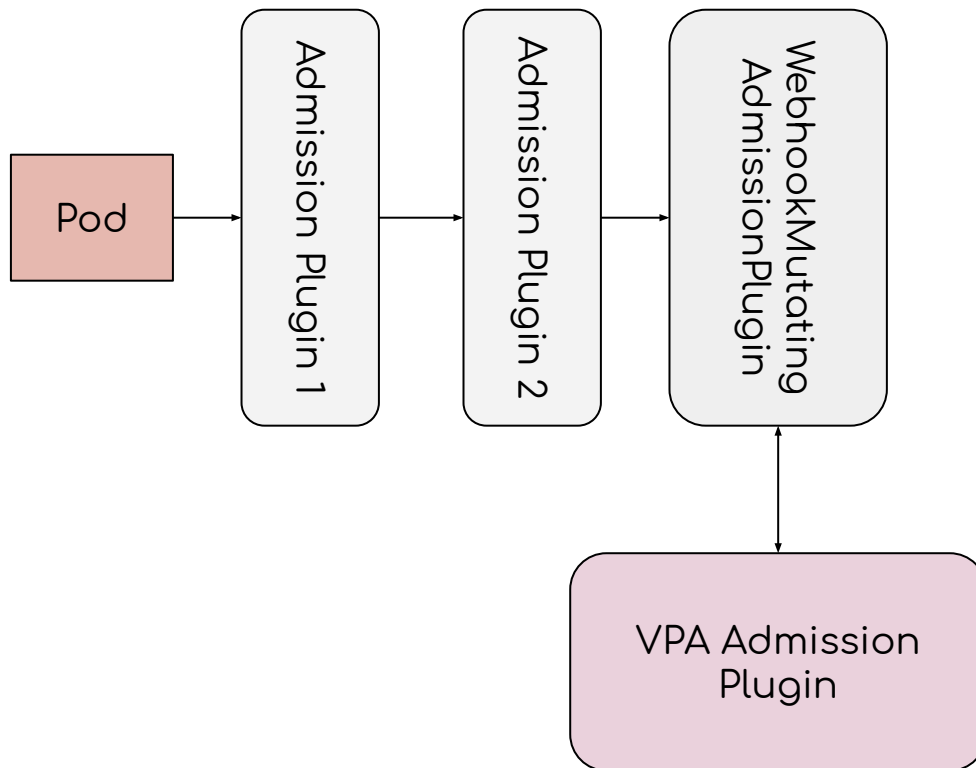


KubeCon



CloudNativeCon

Europe 2019



VPA — Admission Controller

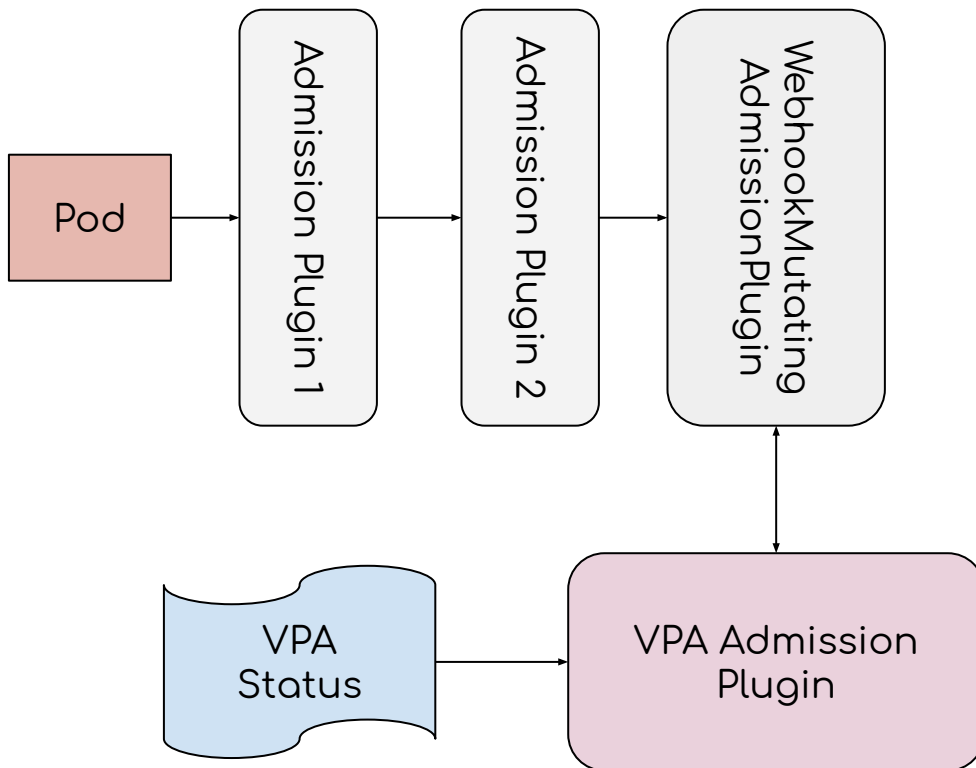


KubeCon



CloudNativeCon

Europe 2019



VPA — Admission Controller

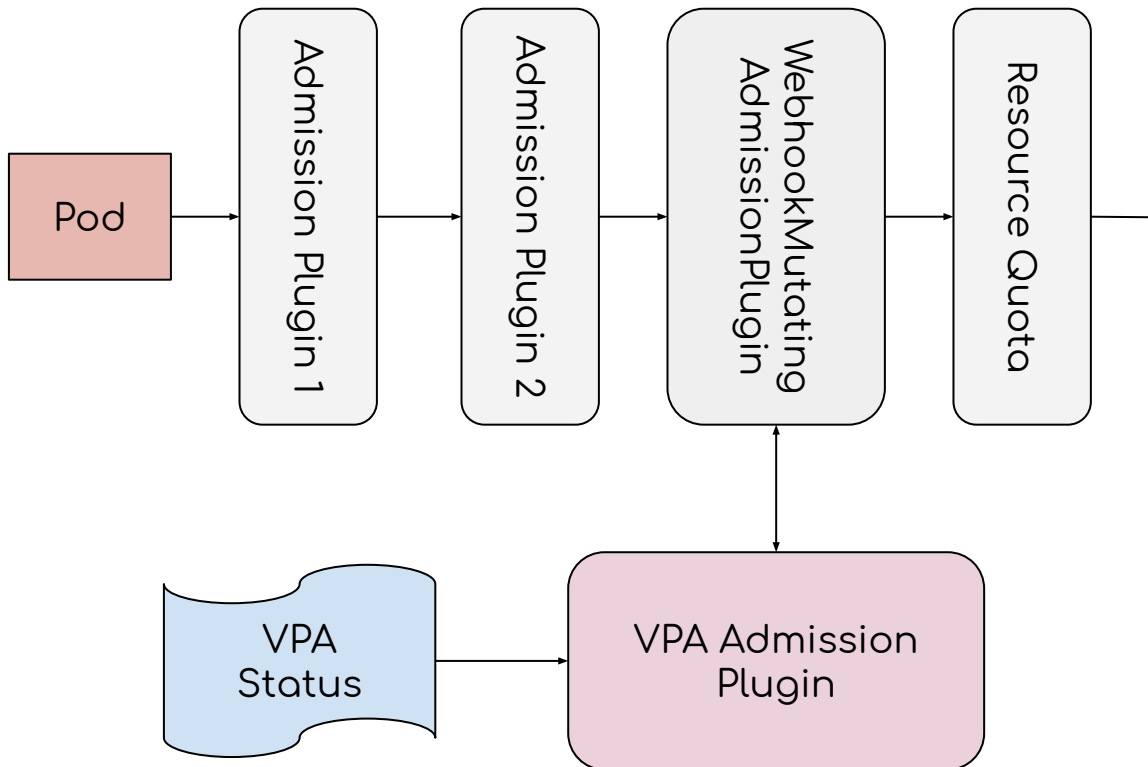


KubeCon



CloudNativeCon

Europe 2019



VPA — Admission Controller



KubeCon



CloudNativeCon

Europe 2019

- Operates in "Auto"/"Init" modes
- Applies Target Recommendation to requests
- Annotates Pod ("vpaUpdates")

VPA — Best practices for Auto mode



KubeCon



CloudNativeCon

Europe 2019

- Start with "Off" mode enabled for 1 week to gain confidence
- Define PDB
- Define min/max resources
- VPA adopts slowly to new usage characteristics
 - e.g. numbers of replicas changed
- Mixing with HPA only when you know what you are doing
 - e.g. HPA based on QPS or absolute value of CPU usage
- Enable Cluster Autoscaler

VPA — Status



KubeCon



CloudNativeCon

Europe 2019

- API in Beta2
- Used in production clusters
- Collecting feedback
- Next steps:
 - Limits
 - GA VPA
 - In-place update



KubeCon



CloudNativeCon

Europe 2019

Cluster Autoscaler

Cluster Autoscaler



KubeCon



CloudNativeCon

Europe 2019

Is your cluster large enough to fit all workloads?

Cluster Autoscaler



KubeCon



CloudNativeCon

Europe 2019

Are your nodes underutilized?

NOT metric based



KubeCon



CloudNativeCon

Europe 2019



Naive solution:

calculate desired number of nodes based on utilization

Horizontal scaling



KubeCon



CloudNativeCon

Europe 2019

Application



91%



94%



93%

Horizontal scaling



KubeCon



CloudNativeCon

Europe 2019

Application



72%



70%



73%



65%

Horizontal scaling



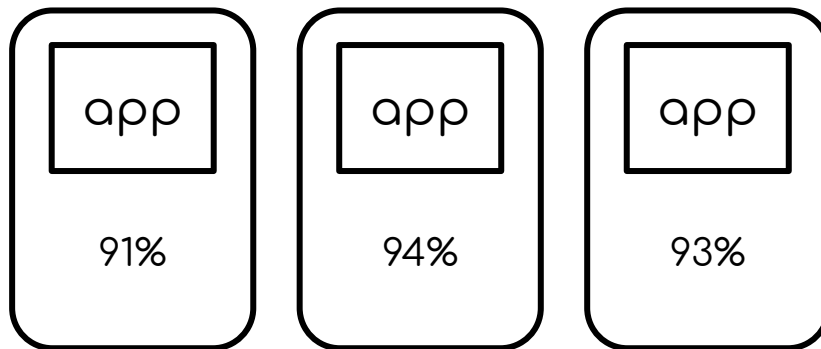
KubeCon



CloudNativeCon

Europe 2019

Instance group



Horizontal scaling



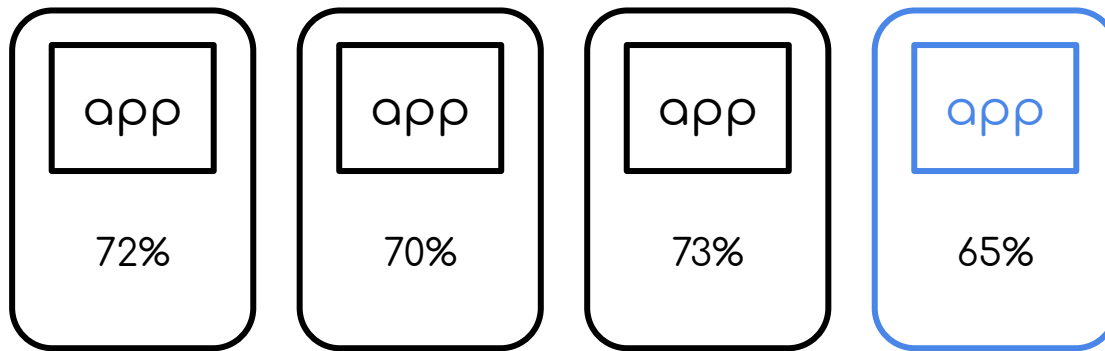
KubeCon



CloudNativeCon

Europe 2019

Instance group



Horizontal scaling



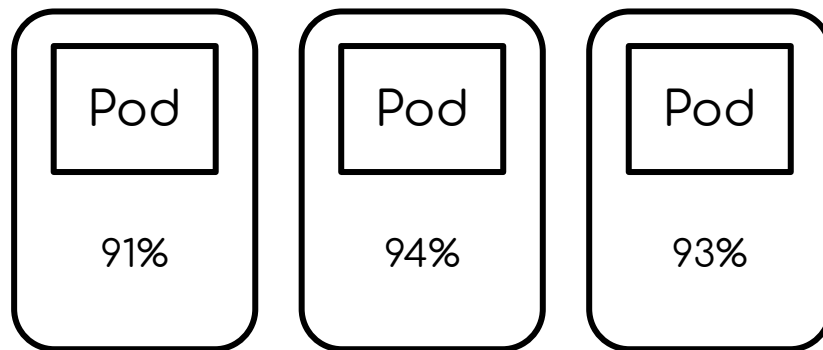
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Node scaling?



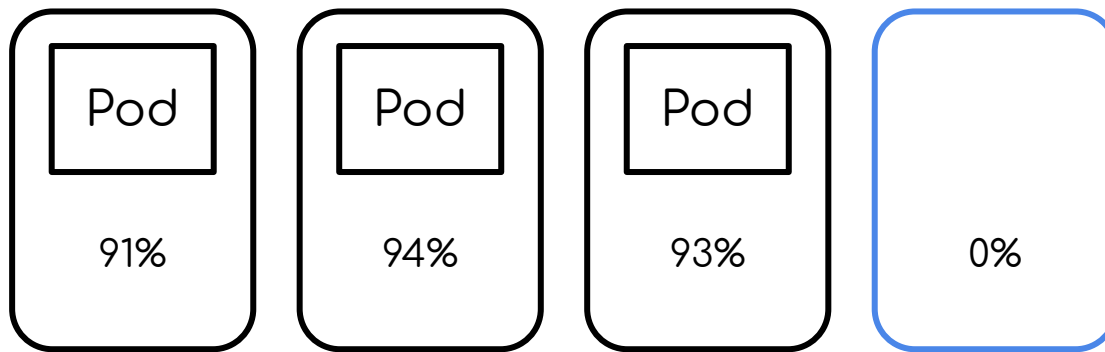
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Pod scaling



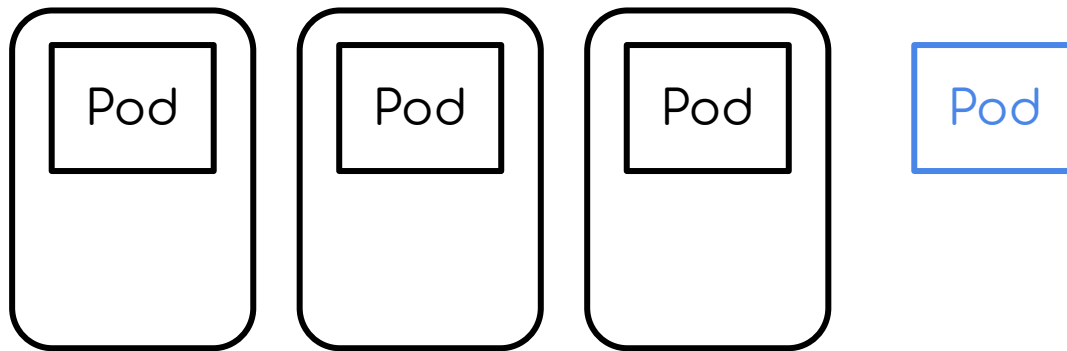
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Pod scaling



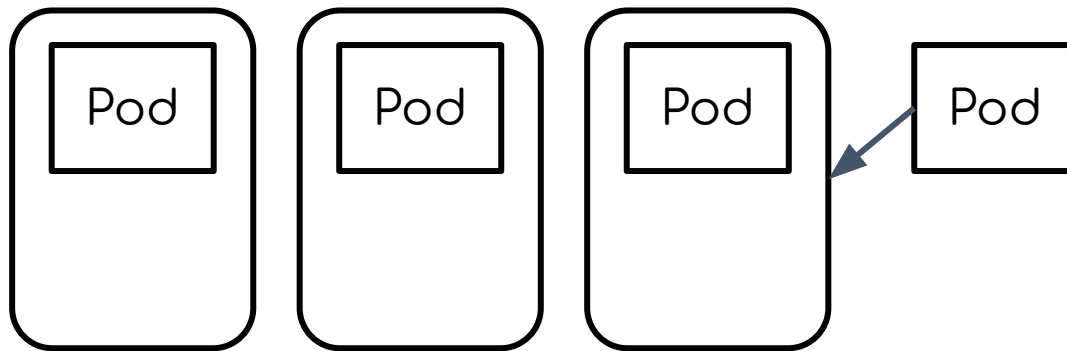
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Pod scaling



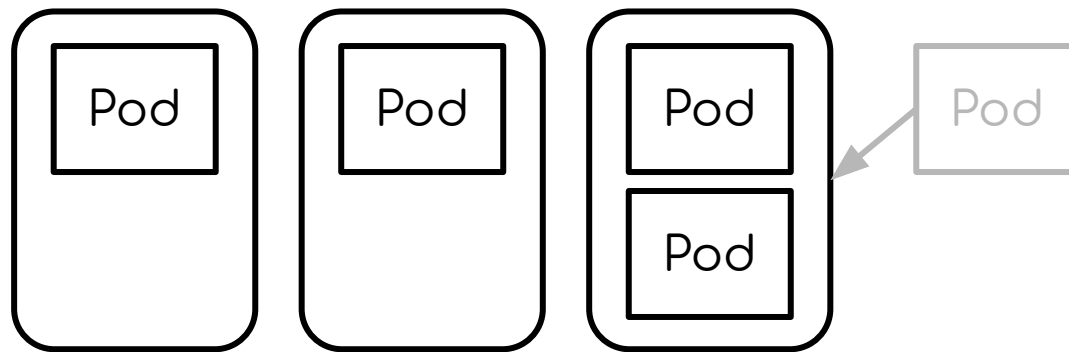
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Pending pods



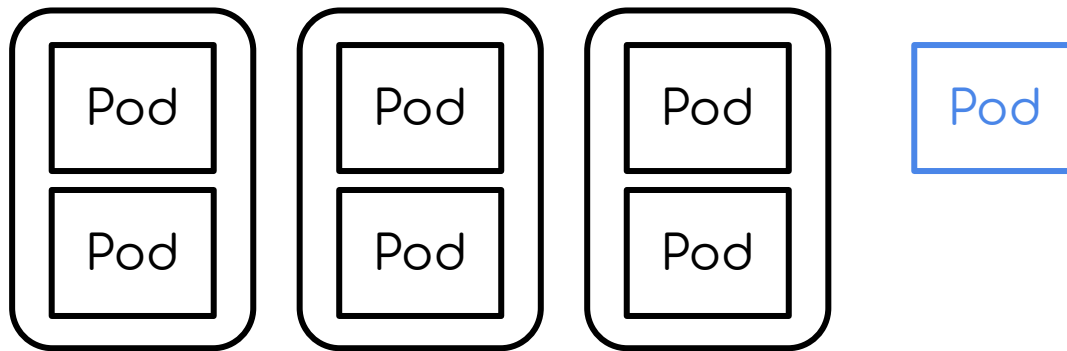
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Pending pods



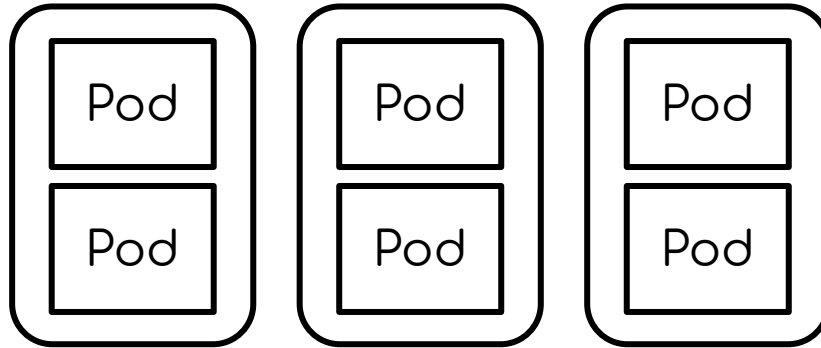
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Pending pods

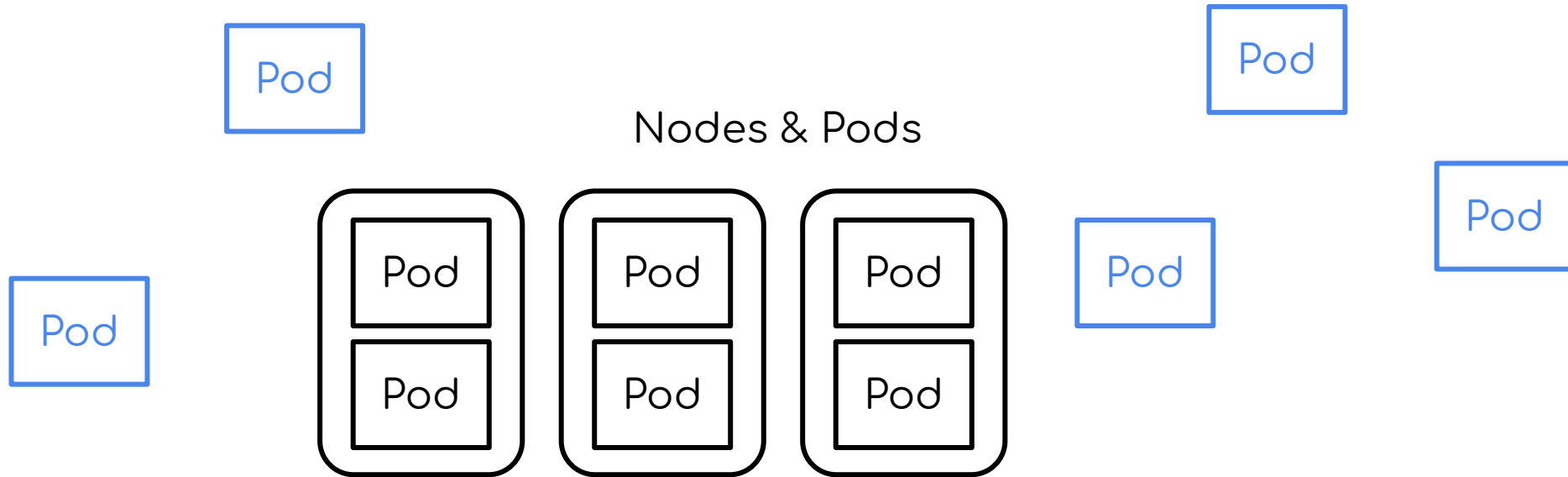


KubeCon



CloudNativeCon

Europe 2019



Pending pods

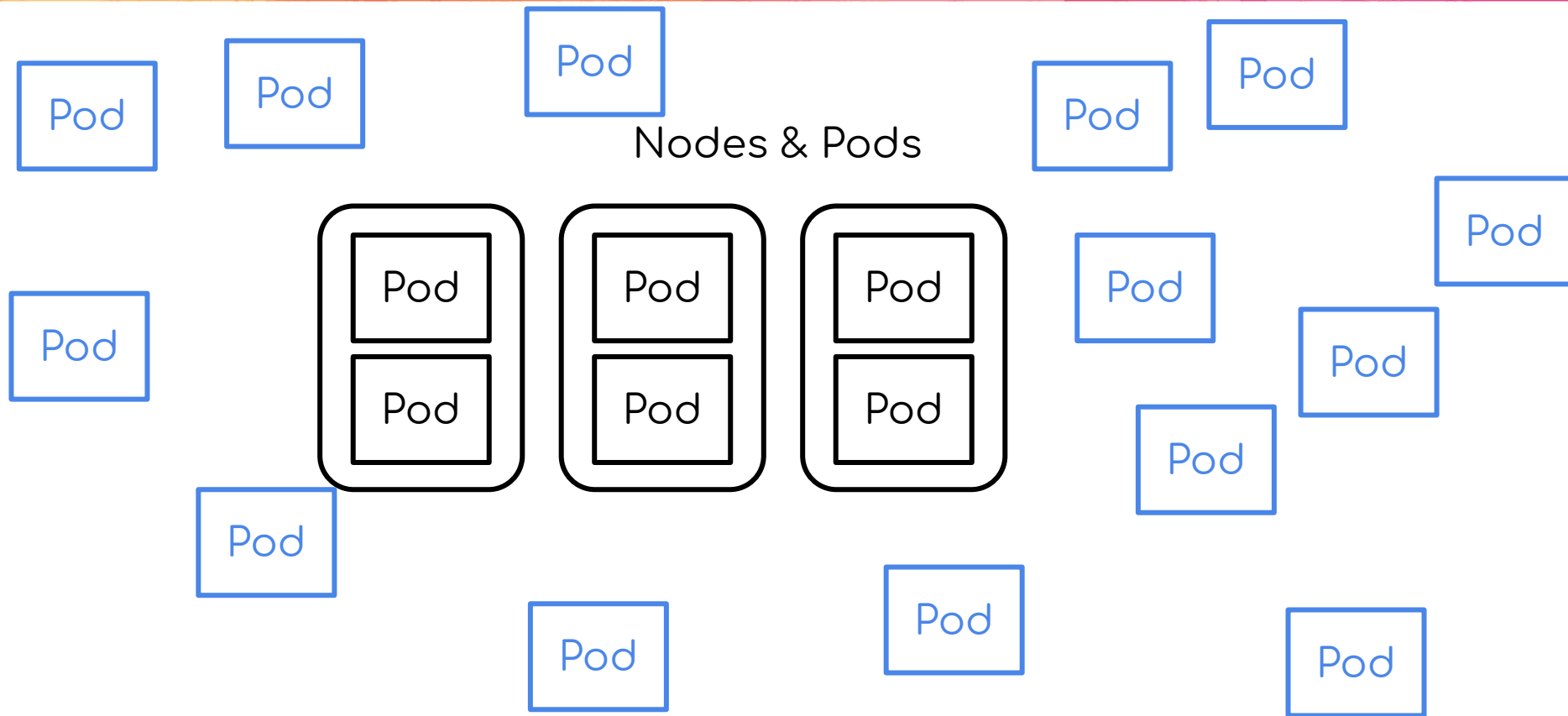


KubeCon



CloudNativeCon

Europe 2019



Pending pods

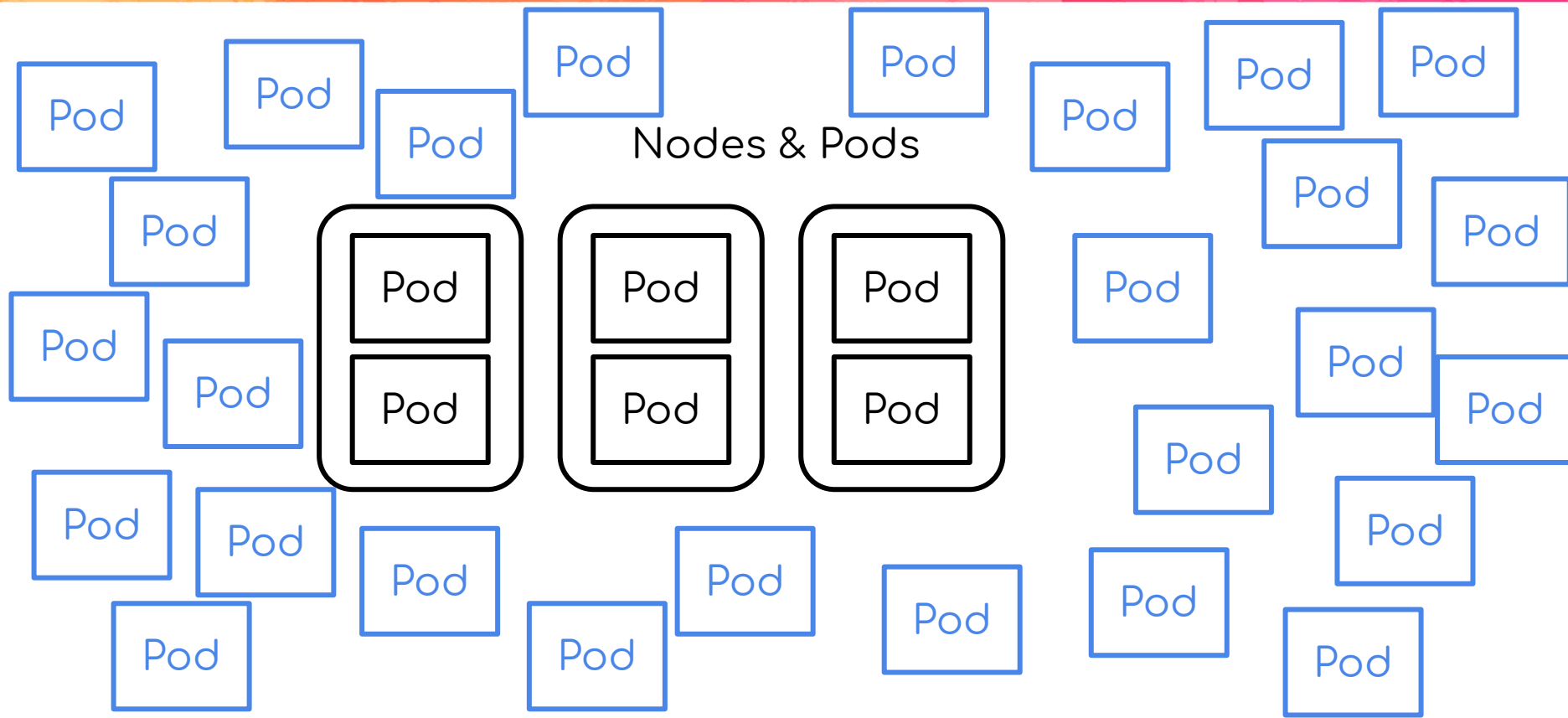


KubeCon



CloudNativeCon

Europe 2019



Pending pods

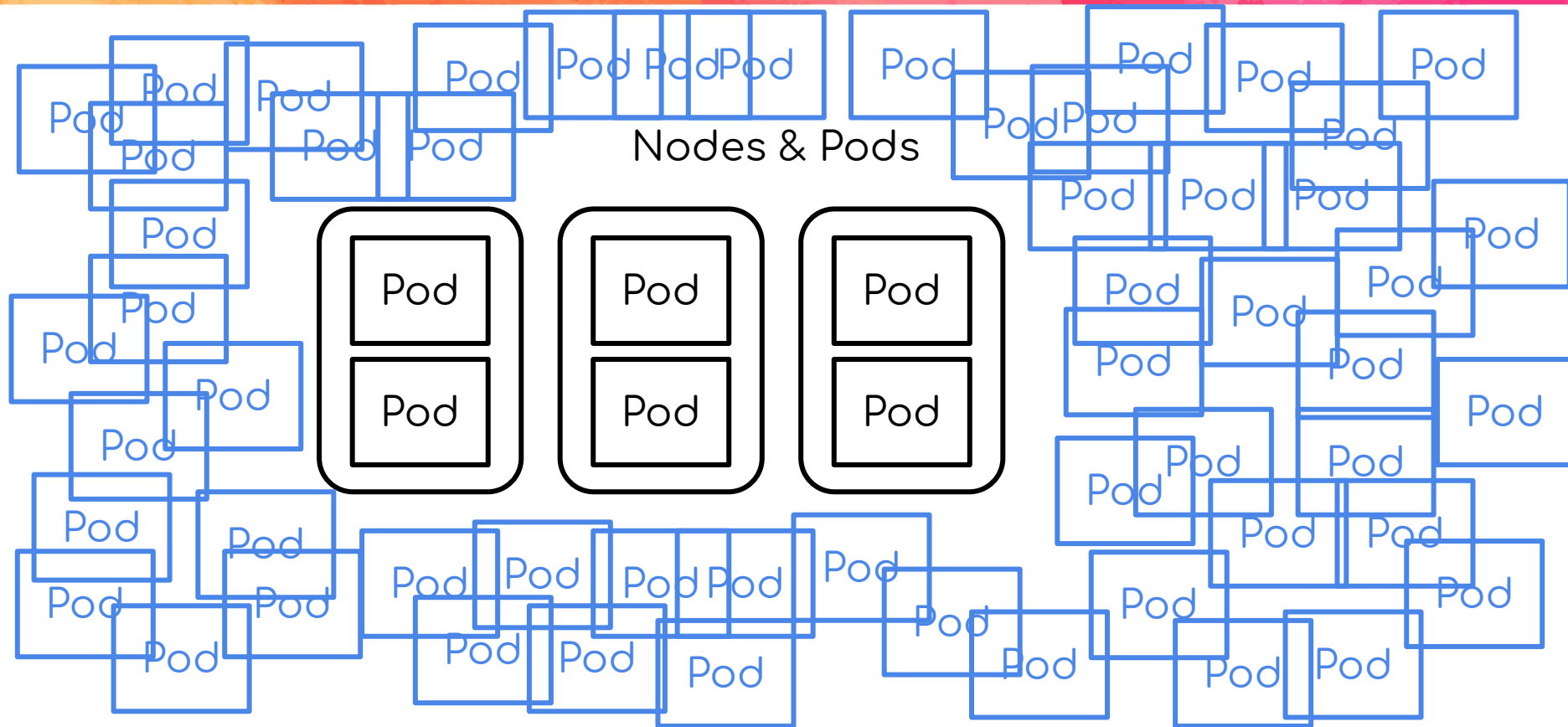


KubeCon



CloudNativeCon

Europe 2019



Pending pods



KubeCon



CloudNativeCon

Europe 2019



Solution:

add just enough nodes to make pods run

Bin packing



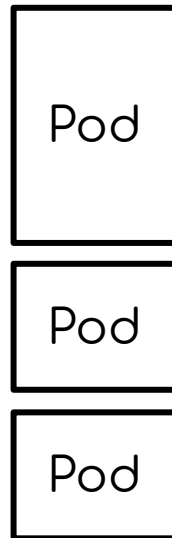
KubeCon



CloudNativeCon

Europe 2019

Pending Pods



New nodes

Bin packing



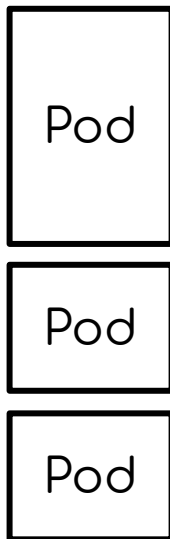
KubeCon



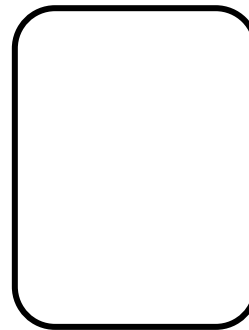
CloudNativeCon

Europe 2019

Pending Pods



New nodes



Bin packing



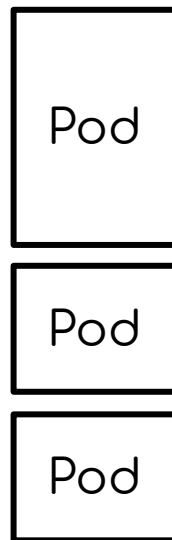
KubeCon



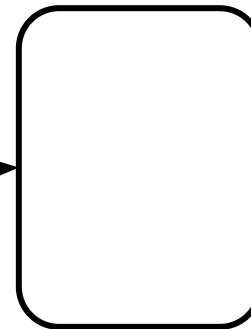
CloudNativeCon

Europe 2019

Pending Pods



New nodes



Bin packing



KubeCon



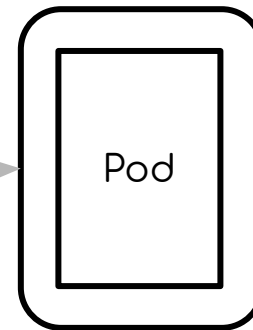
CloudNativeCon

Europe 2019

Pending Pods



New nodes



Bin packing



KubeCon

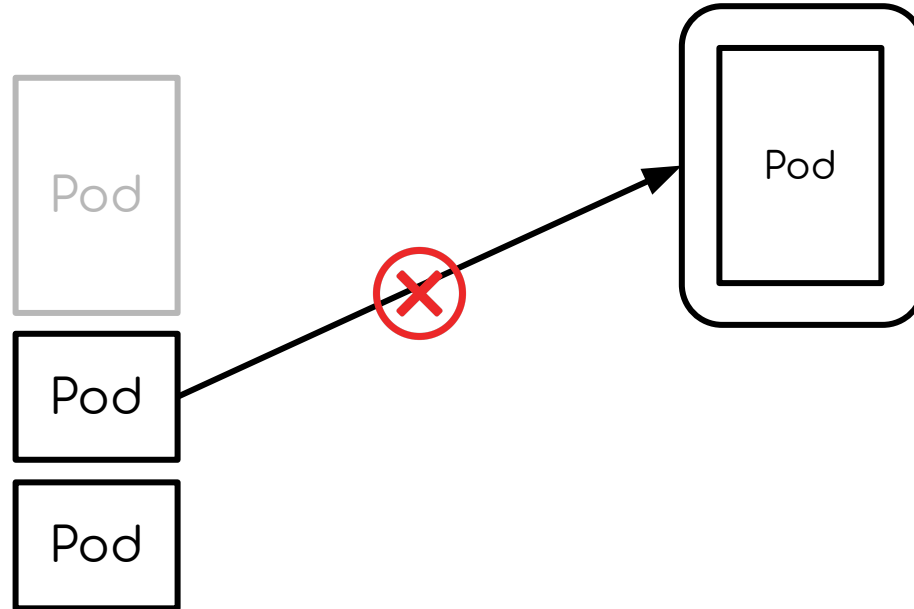


CloudNativeCon

Europe 2019

Pending Pods

New nodes



Bin packing



KubeCon



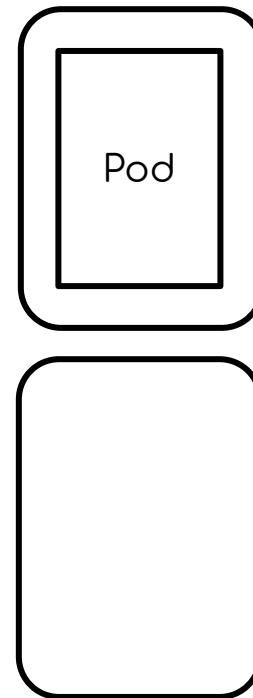
CloudNativeCon

Europe 2019

Pending Pods



New nodes



Bin packing



KubeCon



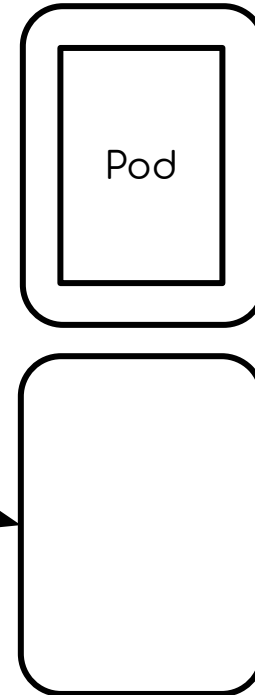
CloudNativeCon

Europe 2019

Pending Pods



New nodes



Bin packing



KubeCon



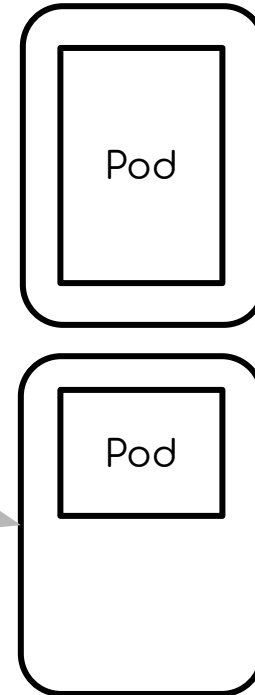
CloudNativeCon

Europe 2019

Pending Pods



New nodes



Bin packing



KubeCon



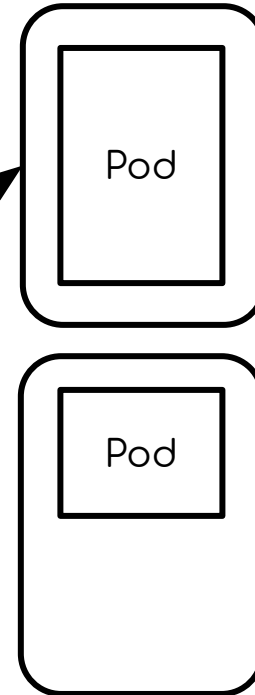
CloudNativeCon

Europe 2019

Pending Pods



New nodes



Bin packing



KubeCon



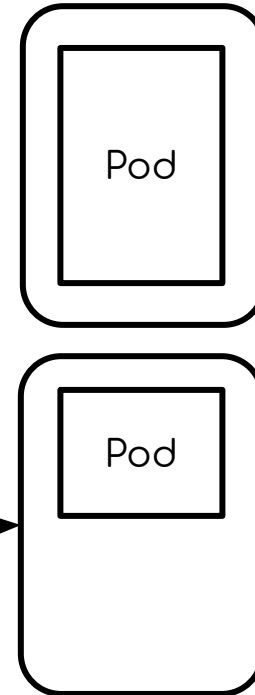
CloudNativeCon

Europe 2019

Pending Pods



New nodes



Bin packing



KubeCon



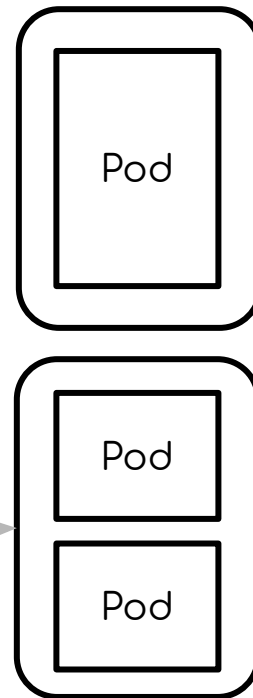
CloudNativeCon

Europe 2019

Pending Pods



New nodes



Bin packing



KubeCon



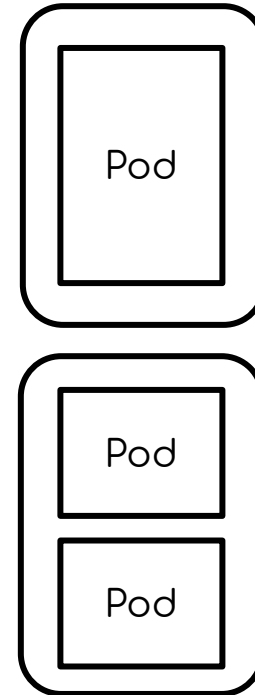
CloudNativeCon

Europe 2019

Pending Pods



New nodes



Bin packing



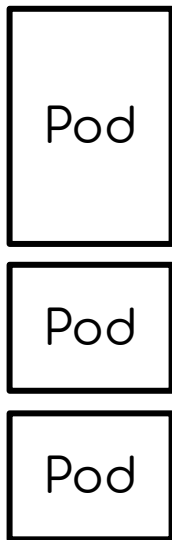
KubeCon



CloudNativeCon

Europe 2019

Pending Pods



New nodes



Decision: Add 2 nodes.



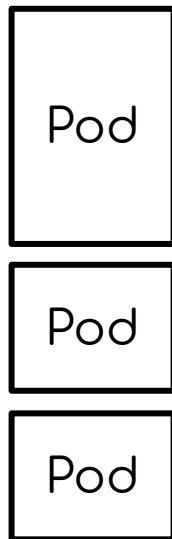
KubeCon



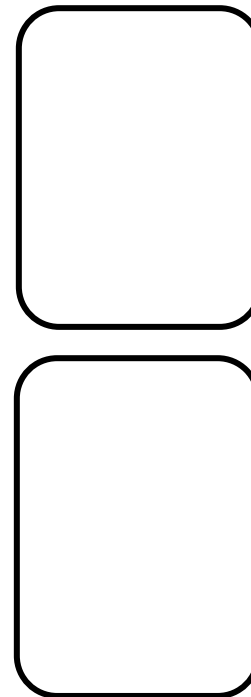
CloudNativeCon

Europe 2019

Pending Pods



New nodes



Unneeded nodes



KubeCon



CloudNativeCon

Europe 2019



Solution:

add just enough nodes to make pods run
remove nodes only if the pods can still run

Unneeded nodes



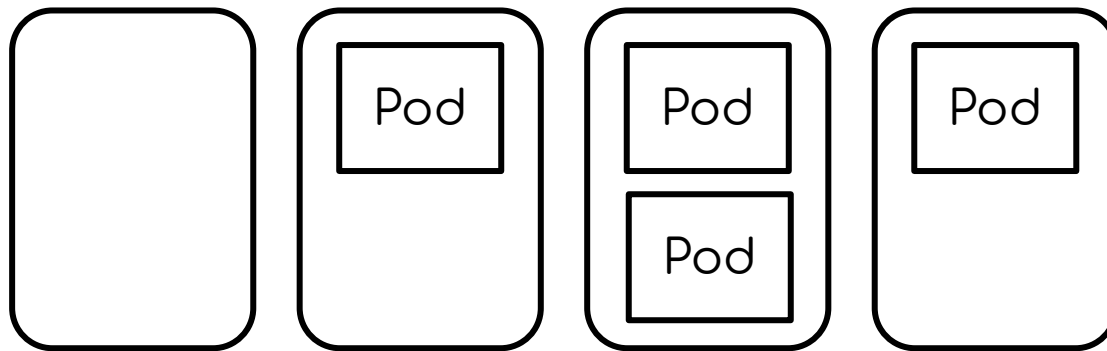
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Unneeded nodes



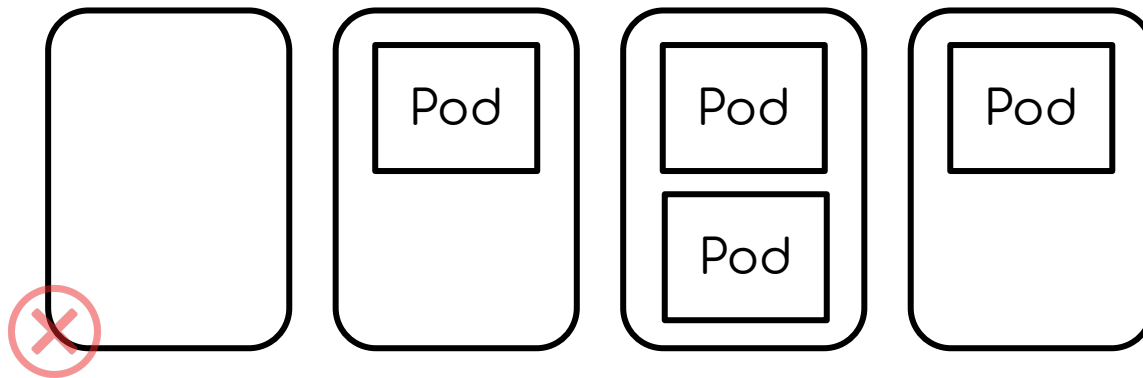
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Unneeded nodes



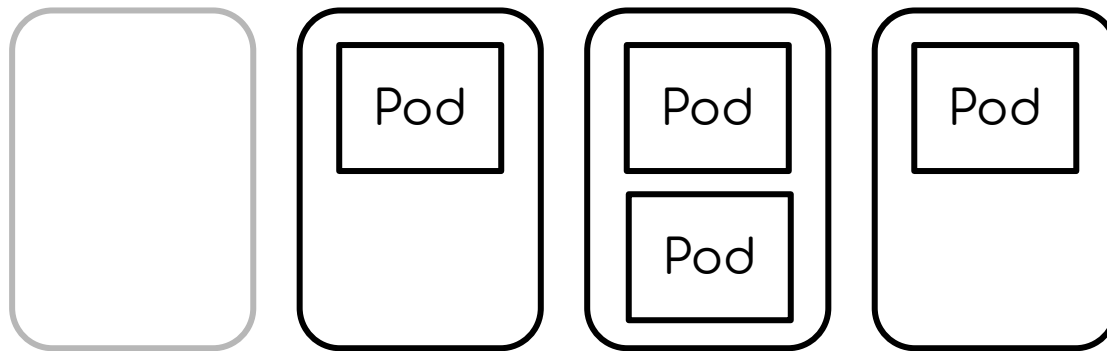
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Unneeded nodes



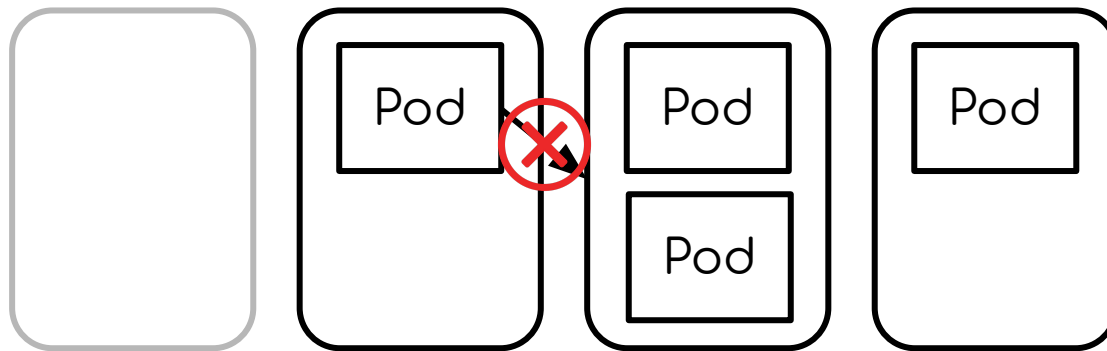
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Unneeded nodes



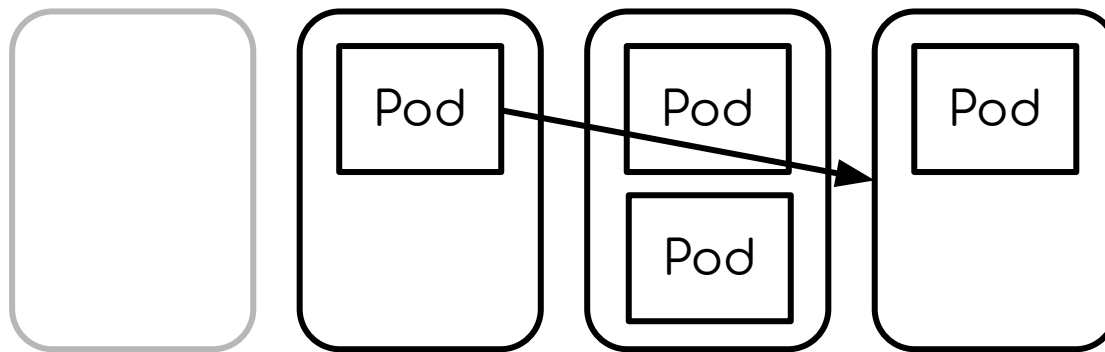
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Unneeded nodes



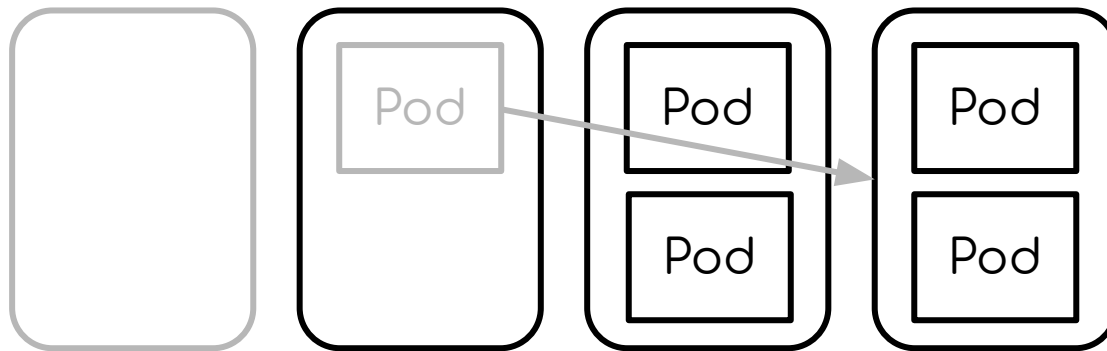
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Unneeded nodes



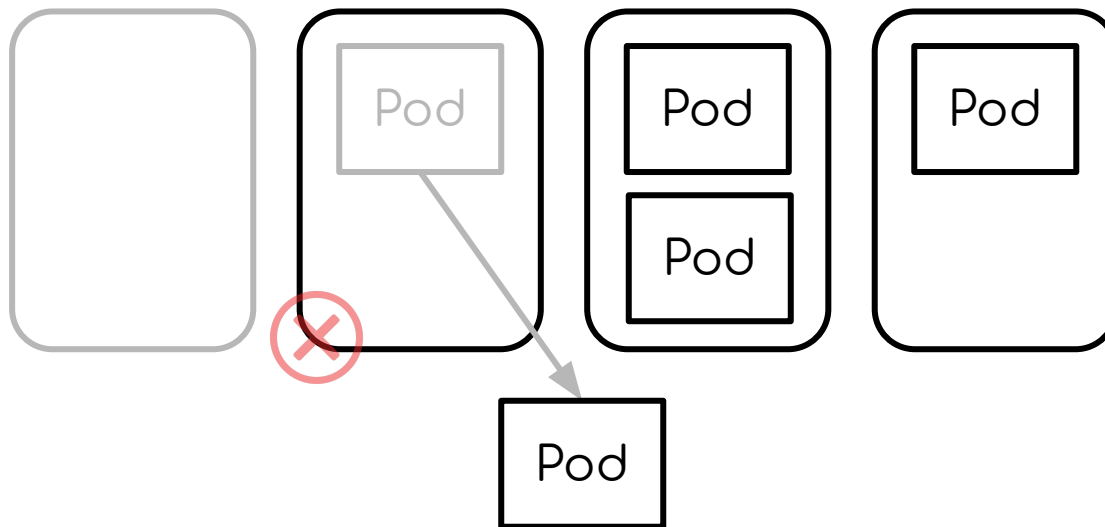
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Unneeded nodes



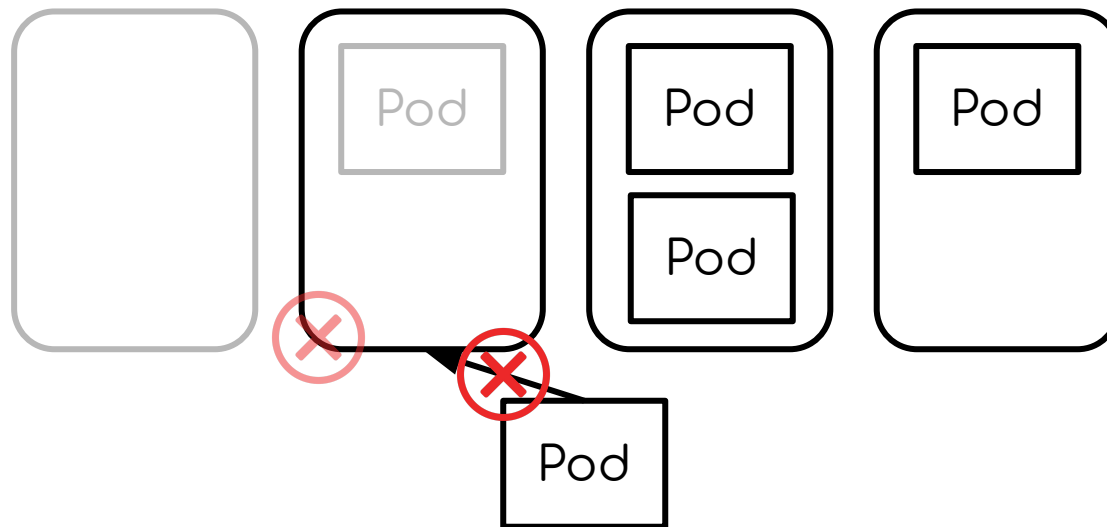
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Unneeded nodes



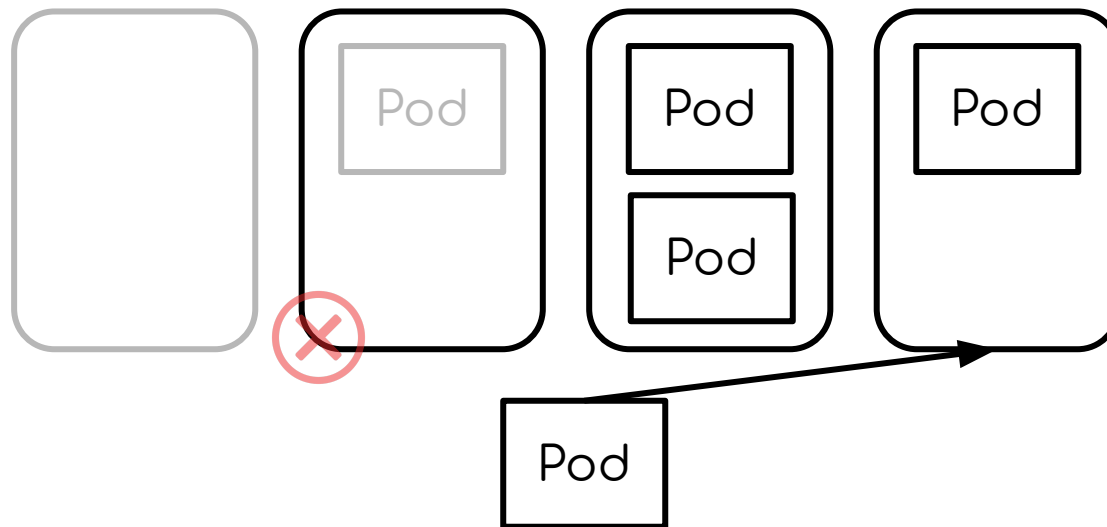
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Unneeded nodes



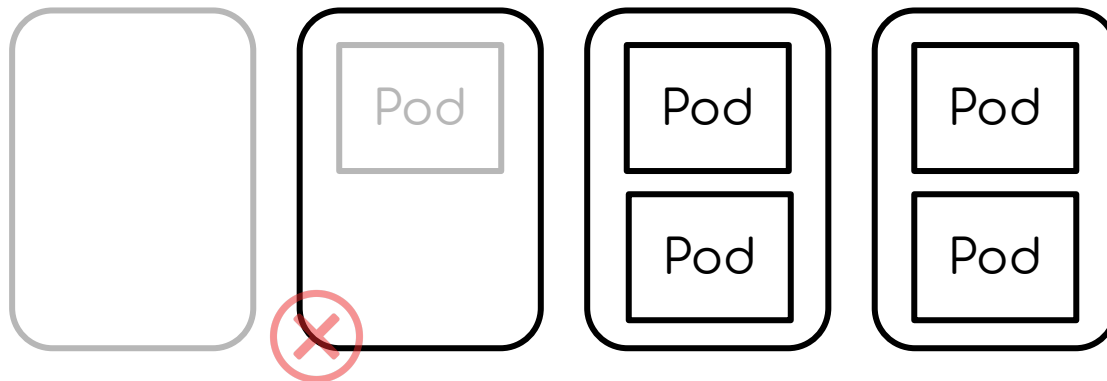
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Unneeded nodes



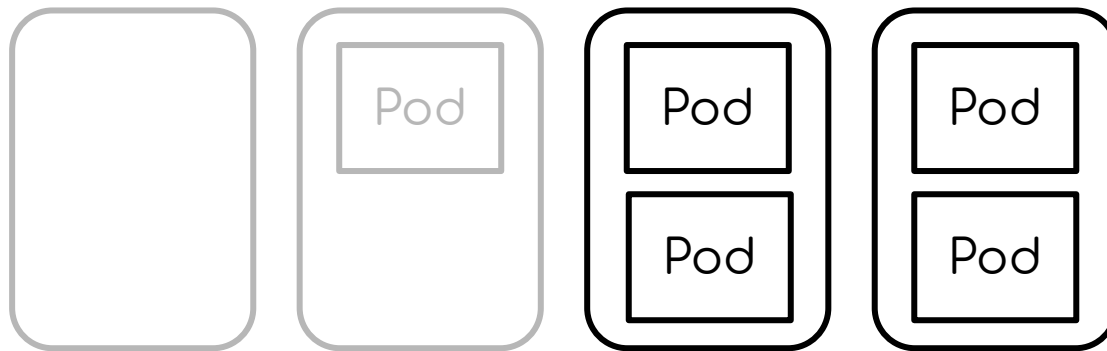
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Predict where pod can run



KubeCon



CloudNativeCon

Europe 2019



Naive solution:
check the pod's requests

Predict where pod can run



KubeCon



CloudNativeCon

Europe 2019



Naive solution:

check the pod's requests

check if pod tolerates node's taints

Predict where pod can run



KubeCon



CloudNativeCon

Europe 2019



Naive solution:

- check the pod's requests

- check if pod tolerates node's taints

- check the pod's node selector

Predict where pod can run



KubeCon



CloudNativeCon

Europe 2019



Naive solution:

- check the pod's requests

- check if pod tolerates node's taints

- check the pod's node selector

- and affinity...

Predict where pod can run



KubeCon



CloudNativeCon

Europe 2019



Naive solution:

- check the pod's requests

- check if pod tolerates node's taints

- check the pod's node selector

- and affinity...

- don't forget to account for host port conflicts

...

Predict where pod can run



KubeCon



CloudNativeCon

Europe 2019



Solution:

simulate scheduler's behavior by running default predicates

Predict where pod can run



KubeCon



CloudNativeCon

Europe 2019



Solution:

simulate scheduler's behavior by running default predicates

Caveat:

only supports fixed set of predicates

What will a new node look like?



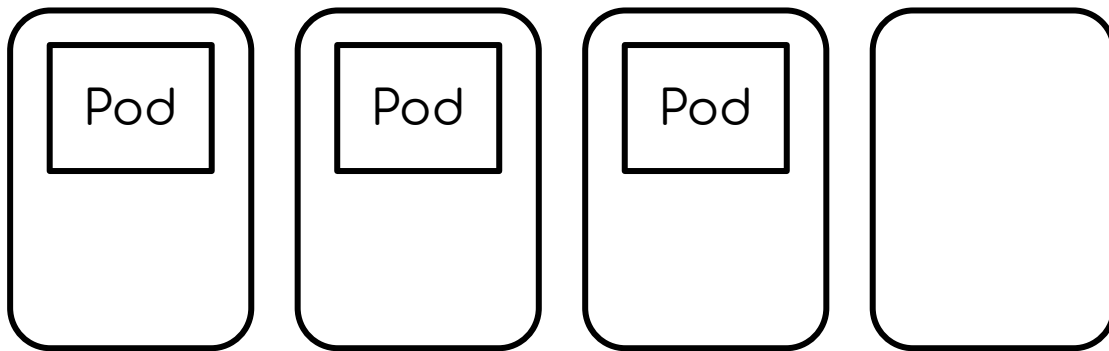
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



What will a new node look like?



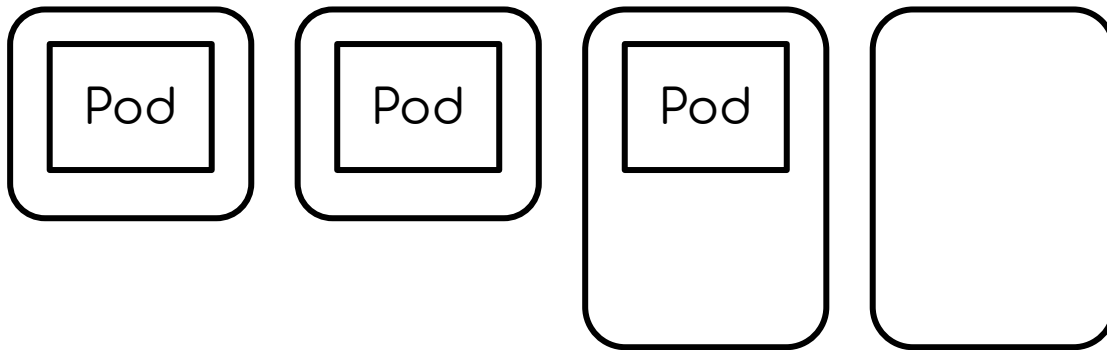
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



What will a new node look like?



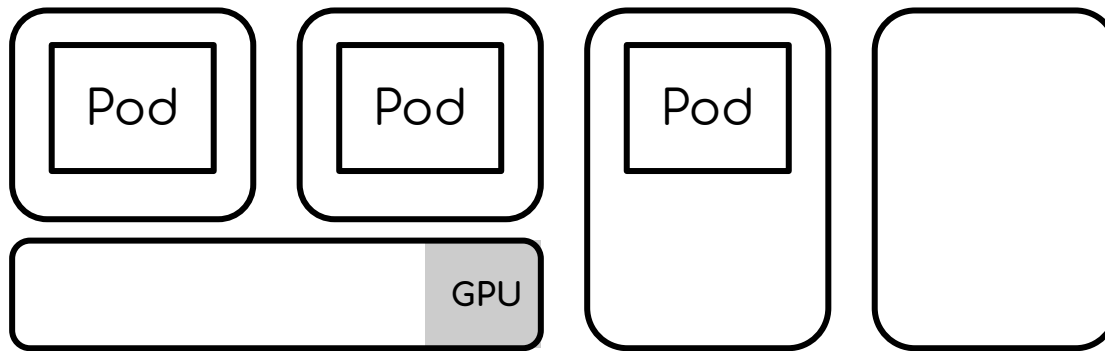
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



What will a new node look like?



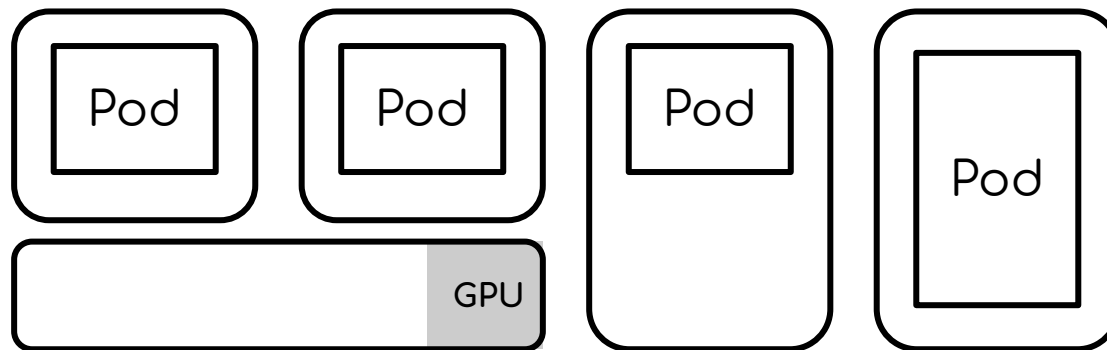
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



What will a new node look like?



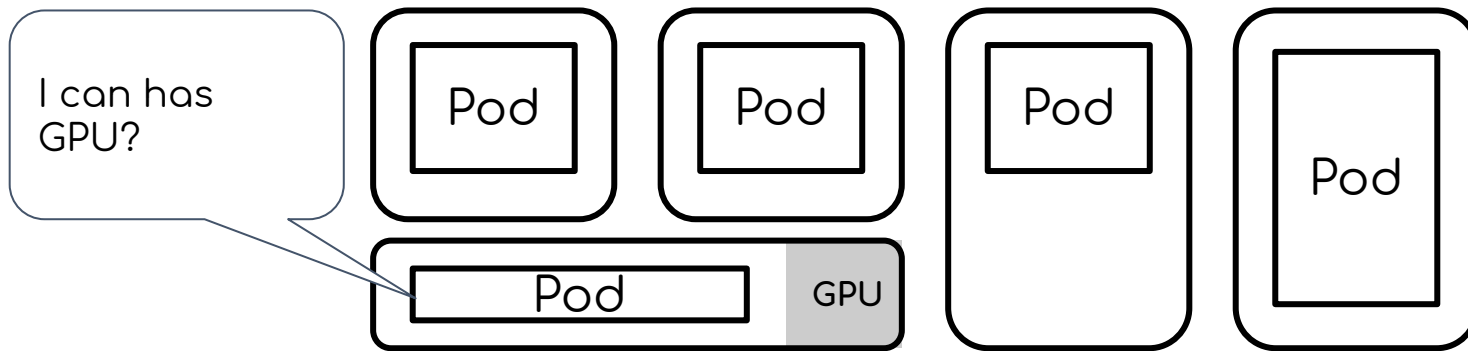
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Node groups



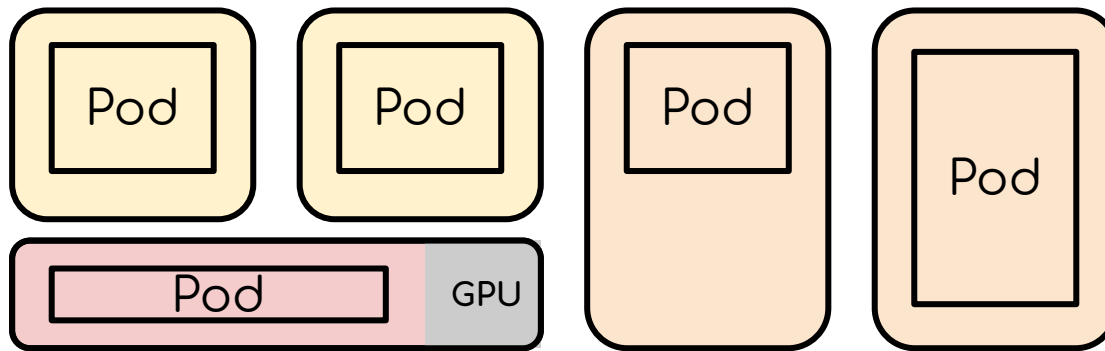
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Node groups



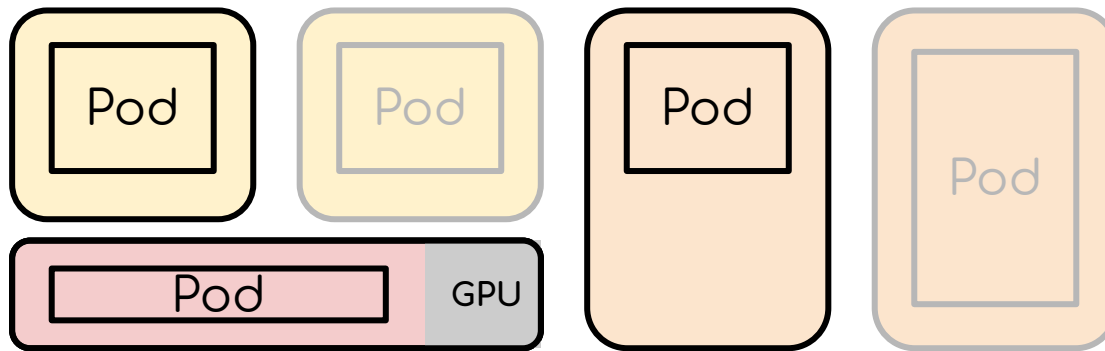
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods



Node groups



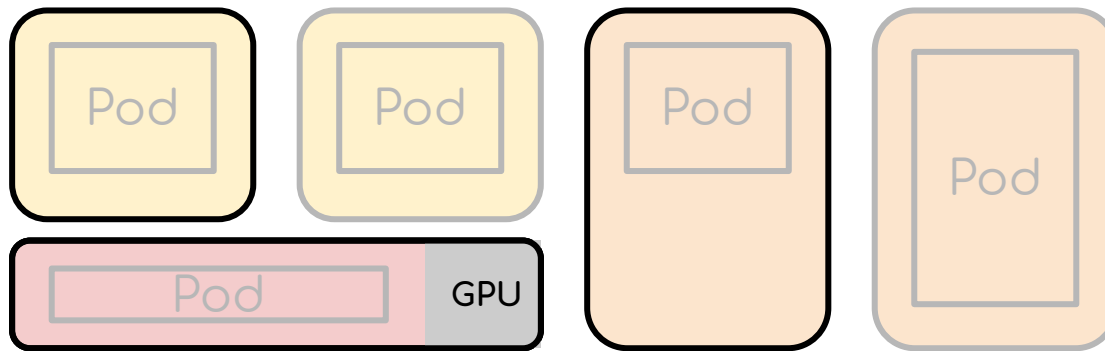
KubeCon



CloudNativeCon

Europe 2019

Nodes & Pods





What does CA do?

look for pending pods

simulate scheduler

add nodes by increasing node group size

delete particular nodes



What doesn't CA do?

- look at actual resource usage
- register nodes in Kubernetes
- configure nodes in any way
- put any labels or taints on new nodes
- support custom scheduling
- predictive autoscaling



KubeCon



CloudNativeCon

Europe 2019

Questions?



KubeCon



CloudNativeCon

Europe 2019