



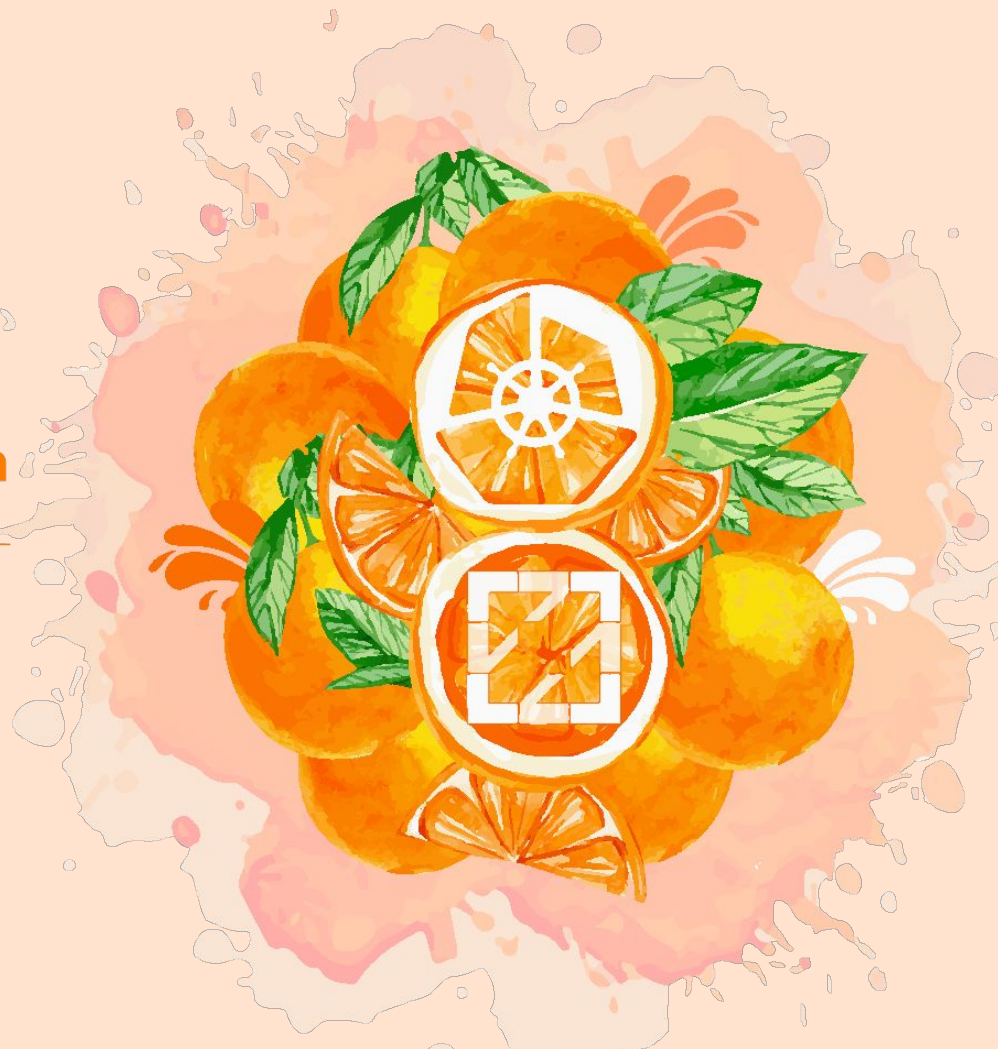
KubeCon



CloudNativeCon

Europe 2022

WELCOME TO VALENCIA





KubeCon



CloudNativeCon

Europe 2022

Scaling K8s Nodes Without Breaking the Bank or Your Sanity

Brandon Wagner & Nick Tran, AWS



Scaling K8s Nodes Without Breaking the Bank or Your Sanity



KubeCon



CloudNativeCon

Europe 2022

Agenda

- What is Spot? 🐕
- Best Practices 🏆
- K8s & Spot ❤️
- Autoscaling your nodes 🐳
 - Cluster Autoscaler
 - Karpenter
- Demo 🎁



Brandon Wagner
Software Engineer
AWS



Nick Tran
Software Engineer
AWS



KubeCon

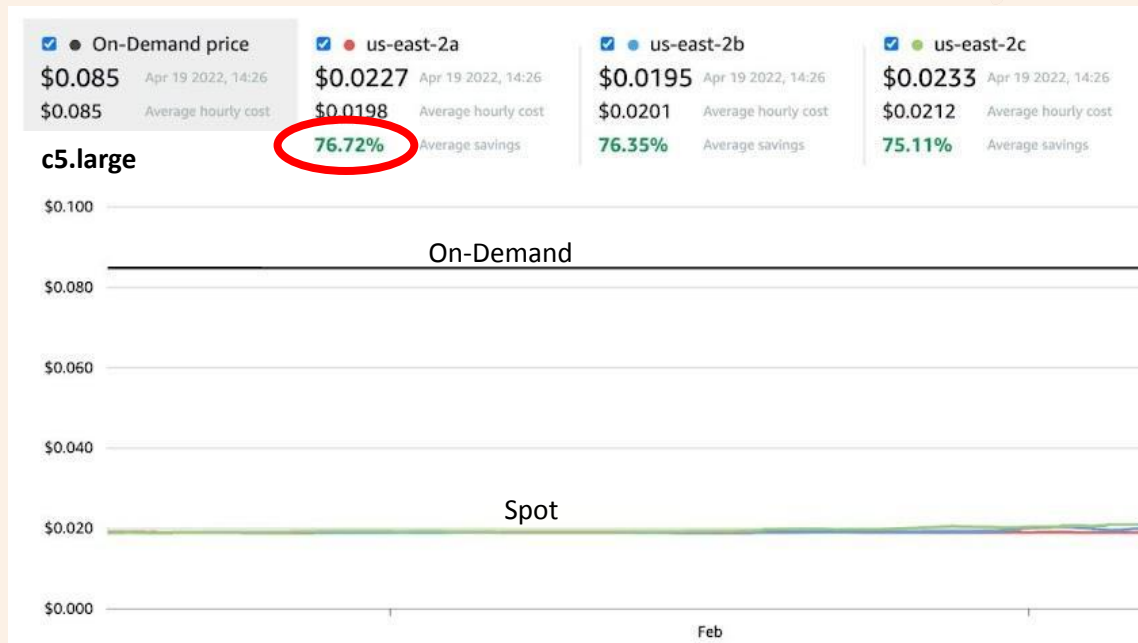


CloudNativeCon

Europe 2022

What is EC2 Spot?

- Spare VM Capacity
- Available at a discount





KubeCon



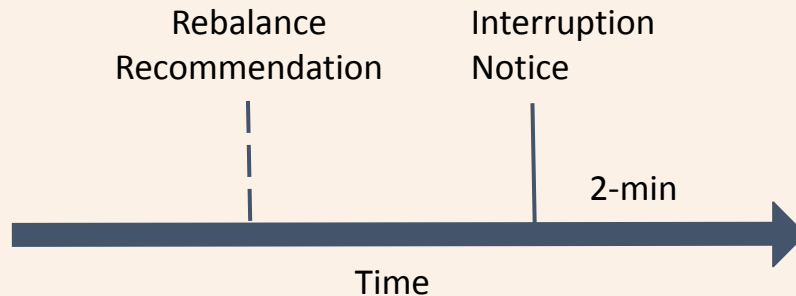
CloudNativeCon

Europe 2022

What is EC2 Spot?

- Spare VM Capacity
- Available at a discount
- Interruptible
 - 2-min notice
 - Rebalance recommendation

Spot Instance Lifecycle



Common Workloads



KubeCon



CloudNativeCon

Europe 2022

- Quick Continuous Integration
- Batch processing
- Stateless APIs
- Dev Workloads



Spot Best Practices

- Don't set Spot max price
- Flexible instance type requests
- Rebalance Recommendations



KubeCon



CloudNativeCon

Europe 2022

Spot Best Practices - Don't Set Max Price

- Spot pricing model overhaul
- Long-term supply and demand





KubeCon



CloudNativeCon

Europe 2022

Spot Best Practices - Flexible Instance Types

- Increases Spot instance availability
- Capacity pools
- Extend instance runtime
 - w/ capacity-optimized

C5	1a	1b	1c	On-demand
8XL	\$0.28	\$0.27	\$0.29	\$1.76
2XL	\$0.08	\$0.07	\$0.08	\$0.44
L	\$0.01	\$0.01	\$0.04	\$0.11

Example Hourly Prices



KubeCon

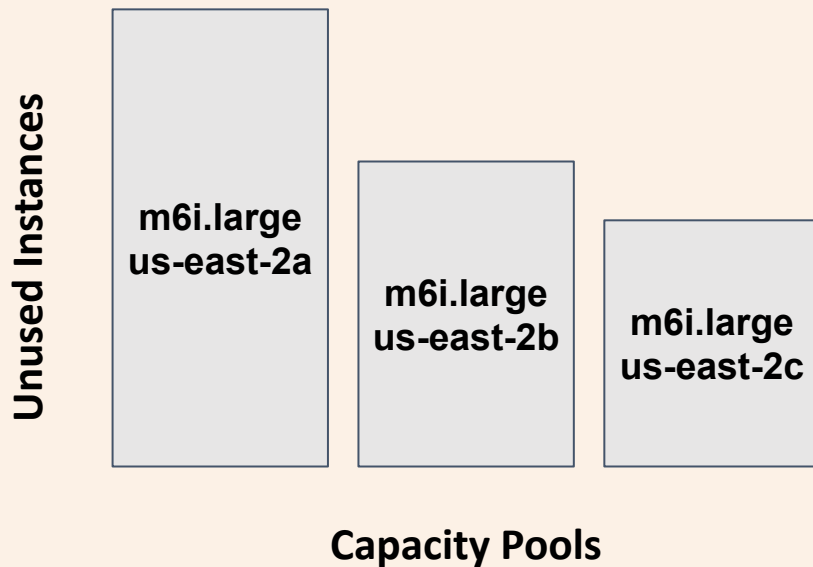


CloudNativeCon

Europe 2022

Spot Best Practices - Flexible Instance Types

- Increases Spot instance availability
- Capacity pools
- Extend instance runtime
 - w/ capacity-optimized





KubeCon



CloudNativeCon

Europe 2022

Spot Best Practices - Flexible Instance Types

- Increases Spot instance availability
- Capacity pools
- Extend instance runtime
 - w/ capacity-optimized

Instance Type	vCPU	Memory GiB	Savings over On-Demand*	Frequency of interruption
r6g.large	2	16	78%	10-15% ■ □ □ □ □
m4.large	2	8	81%	<5% □ □ □ □ □
c6g.large	2	4	71%	5-10% ■ □ □ □ □
t3.medium	2	4	70%	<5% □ □ □ □ □
im4gn.large	2	8	70%	5-10% ■ □ □ □ □
is4gen.large	2	12	70%	5-10% ■ □ □ □ □
m5ad.large	2	8	81%	<5% □ □ □ □ □
c6i.large	2	4	76%	<5% □ □ □ □ □

<https://aws.amazon.com/ec2/spot/instance-advisor>



KubeCon



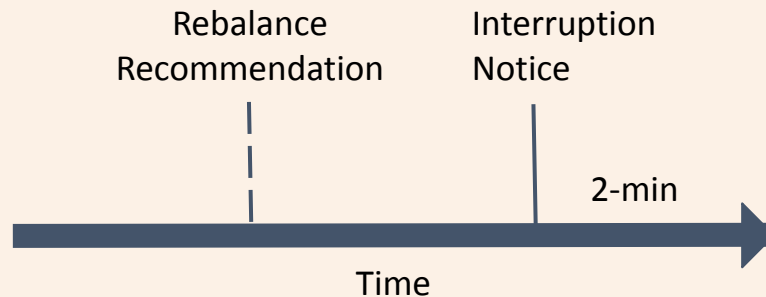
CloudNativeCon

Europe 2022

Spot Best Practices - Rebalance Recommendations

- Warns that interruption is coming
- Vended through Instance Metadata Service (IMDS)

Spot Instance Lifecycle





KubeCon



CloudNativeCon

Europe 2022

K8s and Spot

- github.com/aws/aws-node-termination-handler
 - Interruption Notifications
 - Rebalance Recommendations
- Pod Disruption Budgets (PDBs)





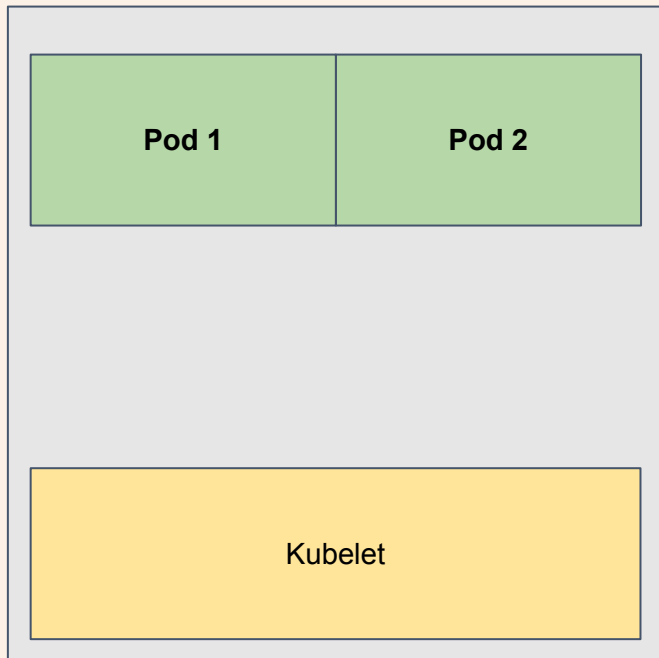
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API



Node



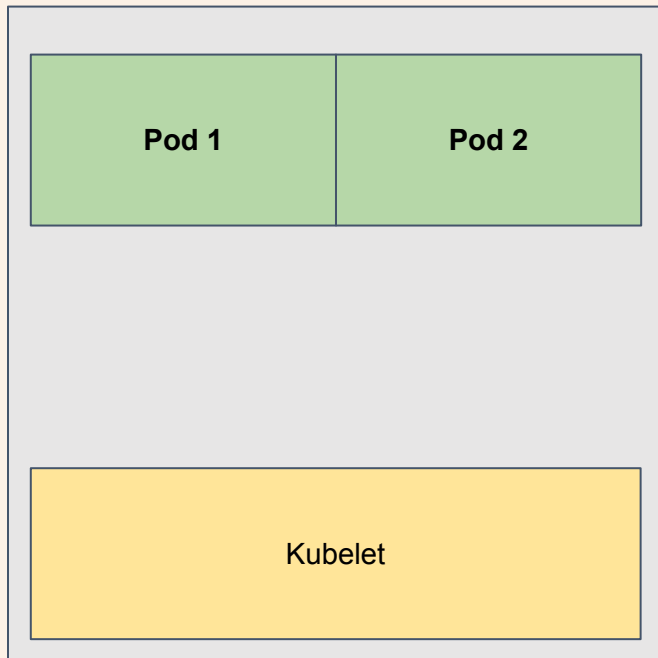
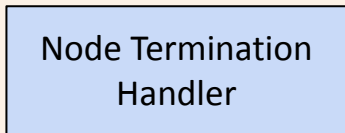
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API



Node



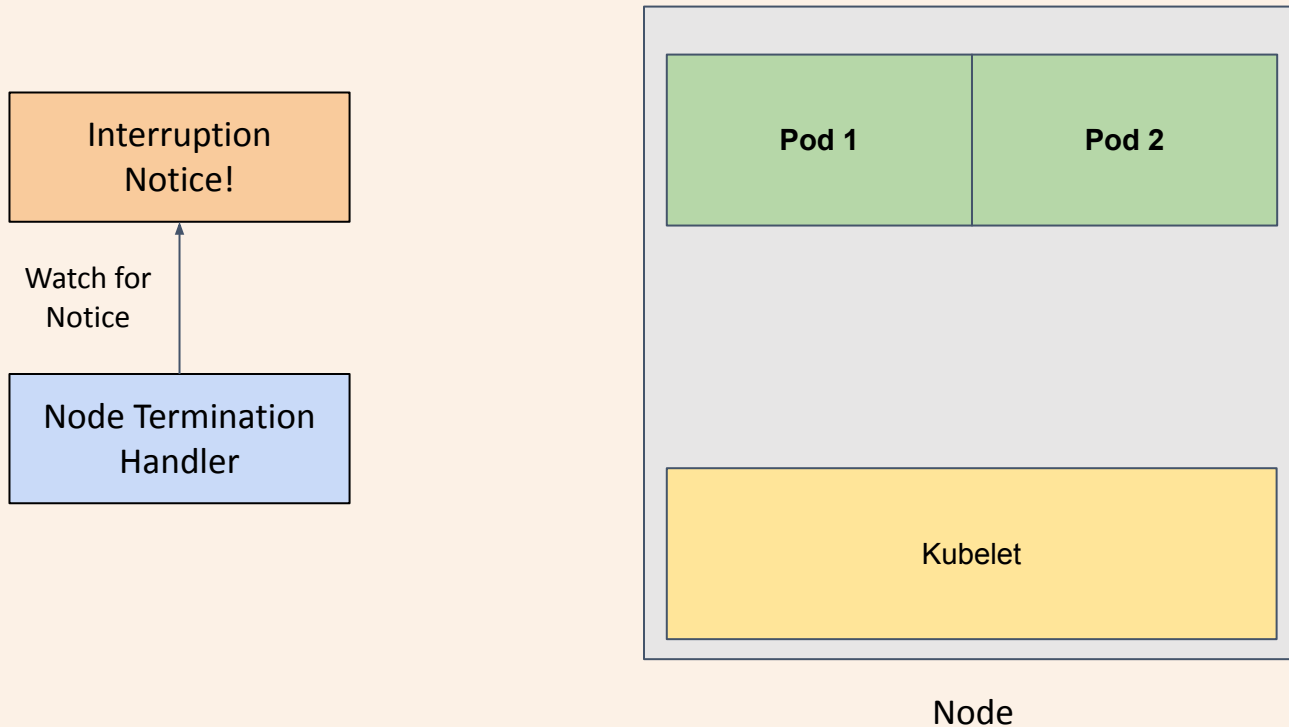
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API





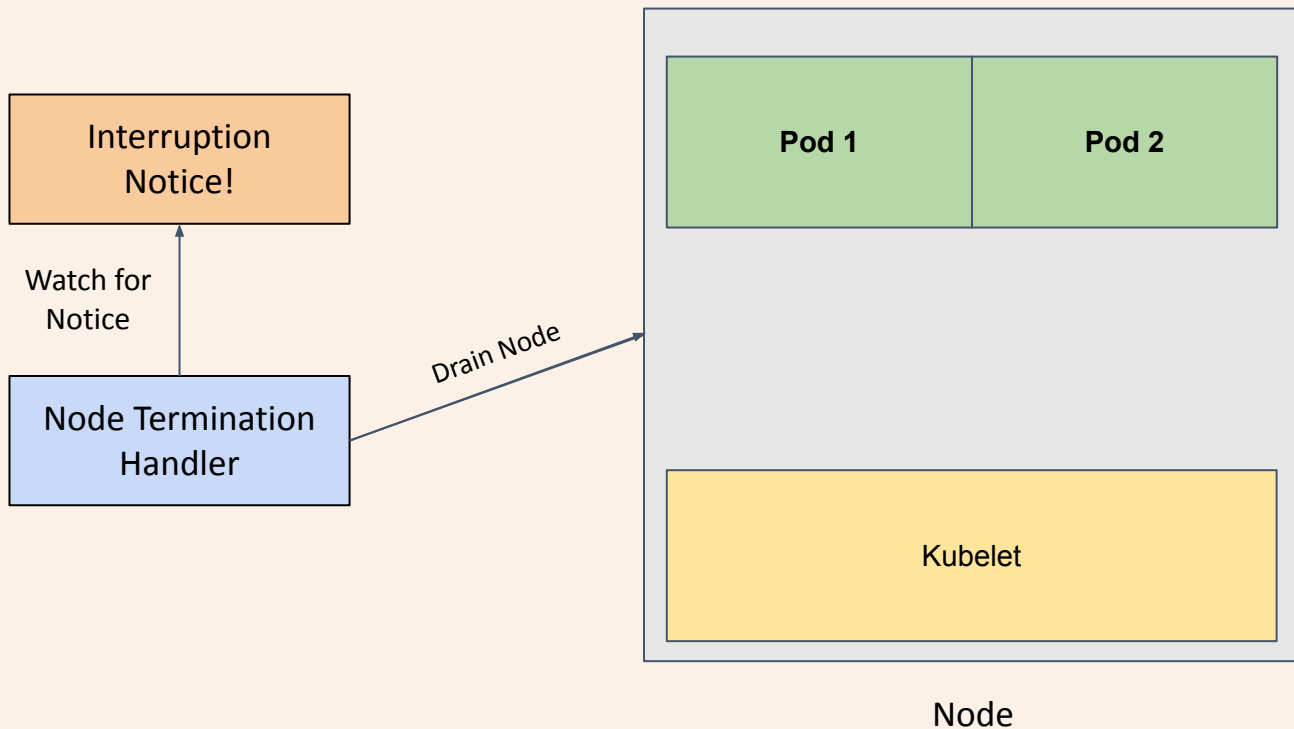
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API





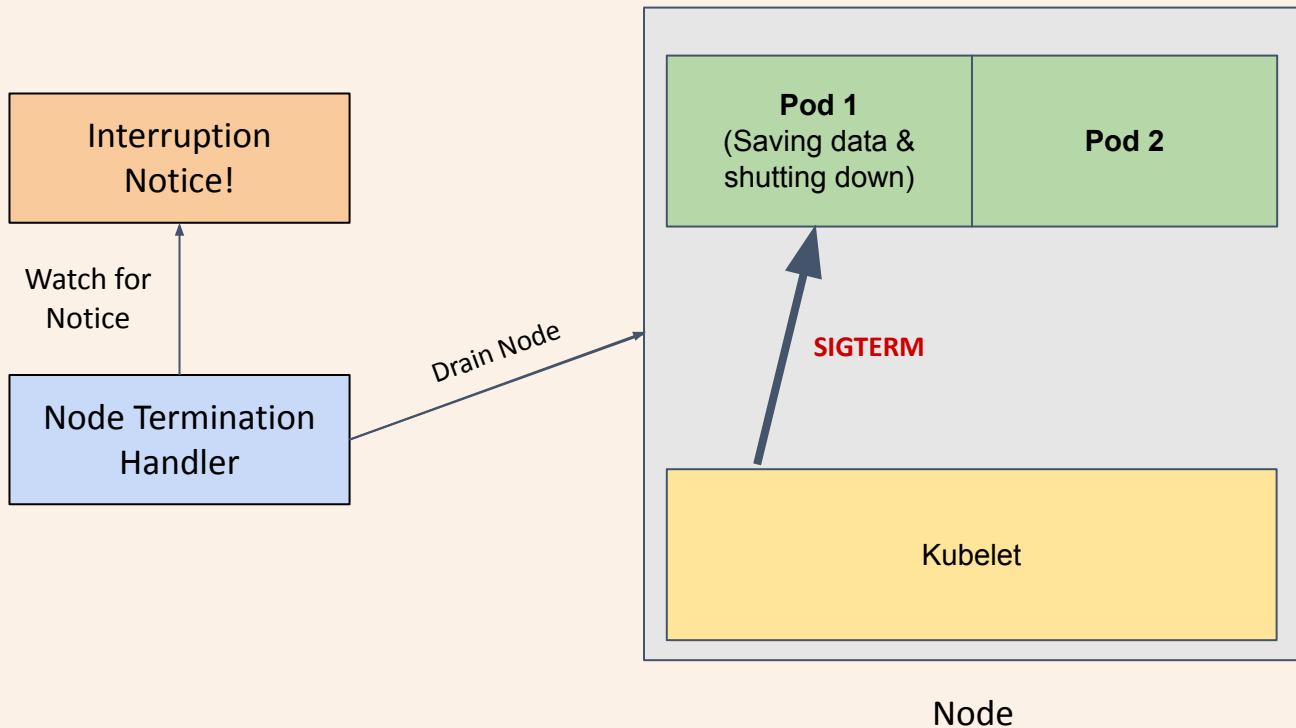
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API





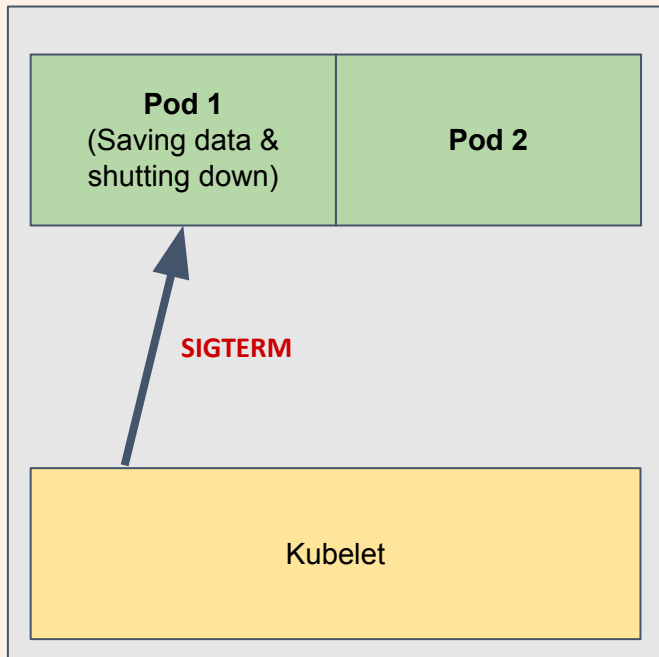
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API





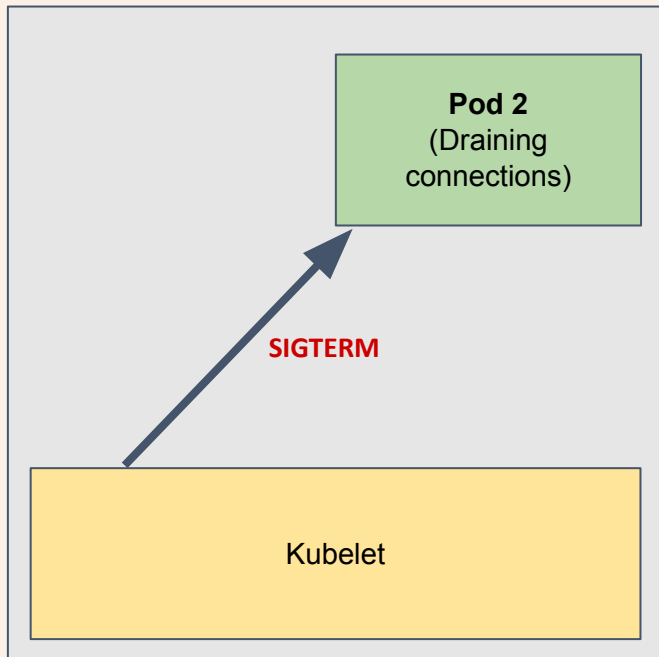
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API



**A FEW
MINUTES
LATER...**



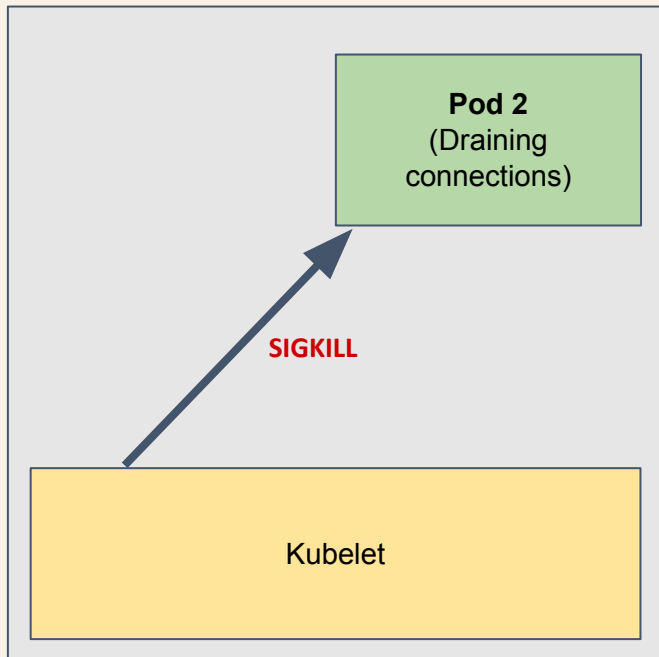
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API





KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API





KubeCon



CloudNativeCon

Europe 2022

Autoscaling your Cluster

- Pod Autoscaling
 - Horizontal Pod Autoscaler (HPA)
 - Vertical Pod Autoscaler (VPA)
- Node Autoscaling
 - Cluster Autoscaler
 - Karpenter



KubeCon

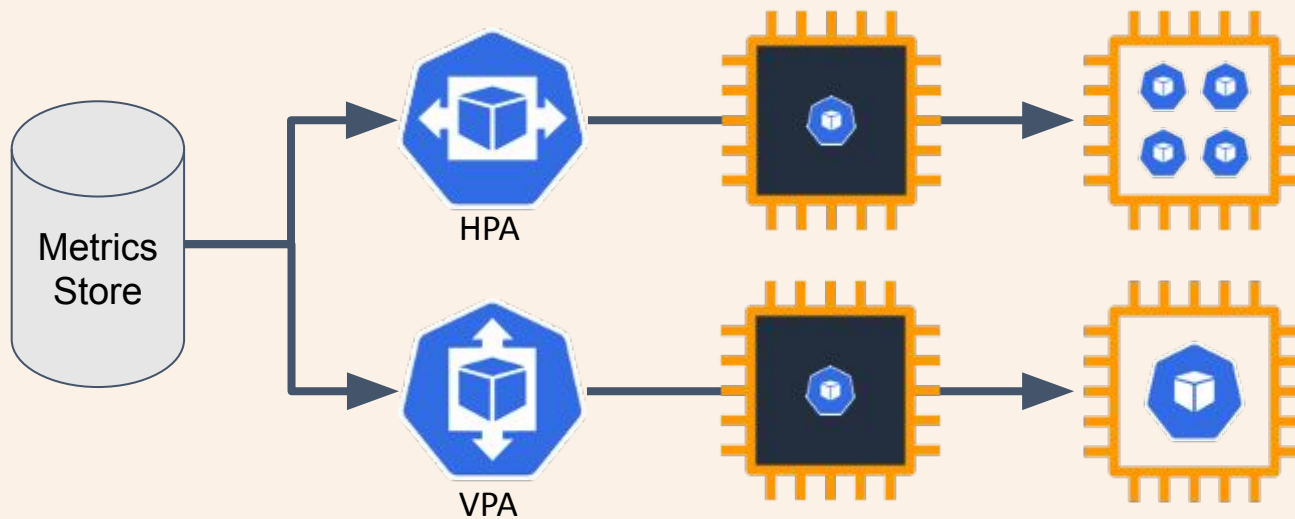


CloudNativeCon

Europe 2022

HPA & VPA

- Horizontally scale: adjust pod replicas
- Vertically scale: adjust resources requests of pods





KubeCon

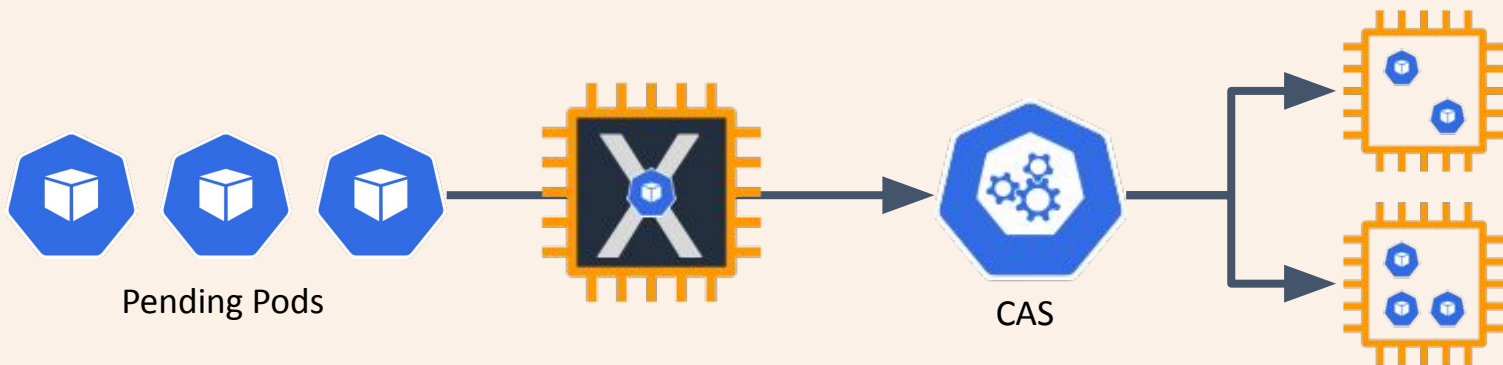


CloudNativeCon

Europe 2022

Cluster Autoscaler

- Simple interface between EC2 AutoScaling Groups (ASGs)
- Increments desired capacity in response to pending pods
- Need to create resource templates per type of pod resource request





KubeCon



CloudNativeCon

Europe 2022

Cluster Autoscaler

- Externally Managed Infrastructure
- An ASG for each instance shape, capacity-type, and zone

AZ 1

On-Demand
2 VCPUs
4 GiB RAM

Spot
2 VCPUs
4 GiB RAM

On-Demand
8 VCPUs
61 GiB RAM
NVIDIA V100

AZ 2

On-Demand
2 VCPUs
4 GiB RAM

Spot
2 VCPUs
4 GiB RAM

On-Demand
8 VCPUs
61 GiB RAM
NVIDIA V100



KubeCon



CloudNativeCon

Europe 2022

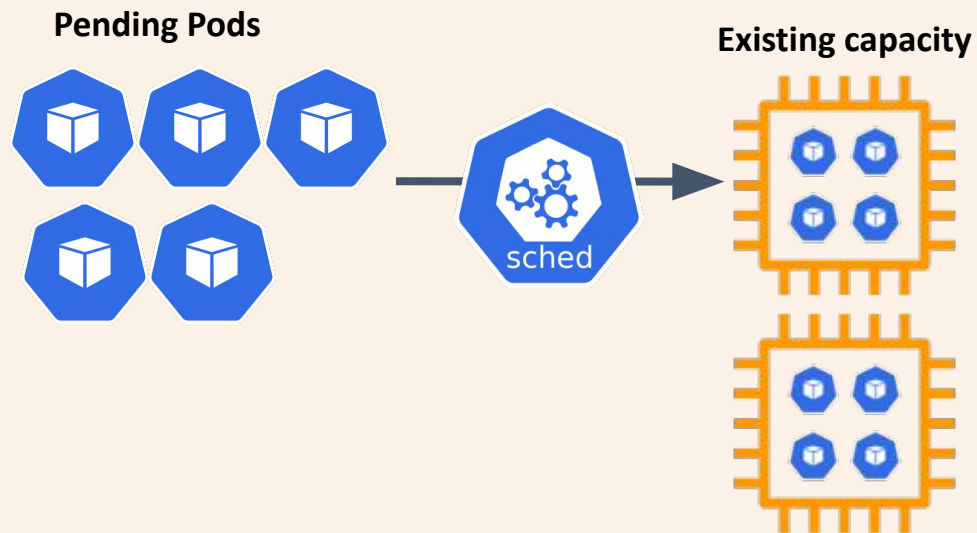
Karpenter

- Groupless Node Autoscaler
- Just-in-Time Provisioning
 - Pending Pods
- Designed to be vendor neutral with a Cloud Provider interface



github.com/aws/karpenter

Karpenter



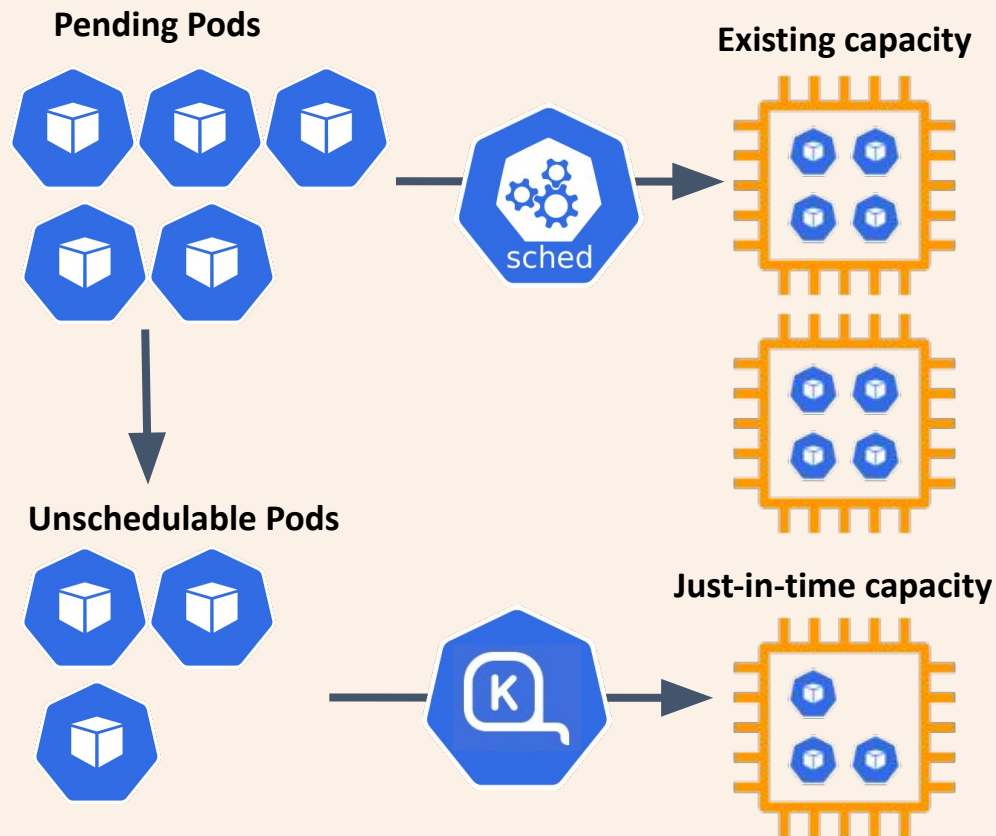
KubeCon



CloudNativeCon

Europe 2022

Karpenter





KubeCon



CloudNativeCon

Europe 2022

Karpenter

- Provisioner CRD
- Requirements
 - K8s Scheduling Constraints
 - Well Known Labels
 - Capacity Type
- Cloud Provider

```
apiVersion: karpenter.sh/v1alpha5
kind: Provisioner
metadata:
  name: default
spec:
  ttlSecondsAfterEmpty: 60
  ttlSecondsUntilExpired: 525600 # ~6 days
  requirements:
    - key: "kubernetes.io/arch"
      operator: In
      values: ["arm64", "amd64"]
    - key: "karpenter.sh/capacity-type"
      operator: In
      values: ["spot", "on-demand"]
  provider:
    kind: AWS
    securityGroupSelector:
      'karpenter.sh/discovery': my-cluster
    subnetSelector:
      'karpenter.sh/discovery': my-cluster
```



KubeCon



CloudNativeCon

Europe 2022

Karpenter

- Provisioner CRD
- Requirements
 - K8s Scheduling Constraints
 - Well Known Labels
 - Capacity Type
- Cloud Provider

```
apiVersion: karpenter.sh/v1alpha5
kind: Provisioner
metadata:
  name: default
spec:
  ttlSecondsAfterEmpty: 60
  ttlSecondsUntilExpired: 525600 # ~6 days
  requirements:
    - key: "kubernetes.io/arch"
      operator: In
      values: ["arm64", "amd64"]
    - key: "karpenter.sh/capacity-type"
      operator: In
      values: ["spot", "on-demand"]
  provider:
    kind: AWS
    securityGroupSelector:
      'karpenter.sh/discovery': my-cluster
    subnetSelector:
      'karpenter.sh/discovery': my-cluster
```




KubeCon



CloudNativeCon

Europe 2022

Karpenter

- Provisioner CRD
- Requirements
 - K8s Scheduling Constraints
 - Well Known Labels
 - Capacity Type
- Cloud Provider

```
apiVersion: karpenter.sh/v1alpha5
kind: Provisioner
metadata:
  name: default
spec:
  ttlSecondsAfterEmpty: 60
  ttlSecondsUntilExpired: 525600 # ~6 days
  requirements:
    - key: "kubernetes.io/arch"
      operator: In
      values: ["arm64", "amd64"]
    - key: "karpenter.sh/capacity-type"
      operator: In
      values: ["spot", "on-demand"]
  provider:
    kind: AWS
    securityGroupSelector:
      'karpenter.sh/discovery': my-cluster
    subnetSelector:
      'karpenter.sh/discovery': my-cluster
```

Spot Best Practices

- Don't set Spot max price
- Flexible instance type requests
- Rebalance Recommendations



KubeCon



CloudNativeCon

Europe 2022

Karpenter

The Provisioner CRD

```
apiVersion: karpenter.sh/v1alpha5
kind: Provisioner
metadata:
  name: default
spec:
  ttlSecondsAfterEmpty: 30
  ttlSecondsUntilExpired: 525600 # ~6 days
  requirements:
    - key: kubernetes.io/arch
      operator: In
      values: ["amd64", "arm64"]
    - key: karpenter.sh/capacity-type
      operator: In
      values: ["spot", "on-demand"]
    - key: node.kubernetes.io/instance-type
      operator: NotIn
      values: ["r5dn.large"]
  provider:
    subnetSelector:
      karpenter.sh/discovery: "my-cluster"
    securityGroupSelector:
      karpenter.sh/discovery: "my-cluster"
```



KubeCon



CloudNativeCon

Europe 2022

Karpenter

The Provisioner CRD

- Flexibility
 - Architecture

```
apiVersion: karpenter.sh/v1alpha5
kind: Provisioner
metadata:
  name: default
spec:
  ttlSecondsAfterEmpty: 30
  ttlSecondsUntilExpired: 525600 # ~6 days
  requirements:
    - key: kubernetes.io/arch
      operator: In
      values: ["amd64", "arm64"]
    - key: karpenter.sh/capacity-type
      operator: In
      values: ["spot", "on-demand"]
    - key: node.kubernetes.io/instance-type
      operator: NotIn
      values: ["r5dn.large"]
  provider:
    subnetSelector:
      karpenter.sh/discovery: "my-cluster"
    securityGroupSelector:
      karpenter.sh/discovery: "my-cluster"
```



KubeCon



CloudNativeCon

Europe 2022

Karpenter

The Provisioner CRD

- Flexibility
 - Architecture

```
apiVersion: v1
kind: Pod
metadata:
  name: my-arm-app
spec:
  containers:
    - name: my-arm-app
      image: "my-arm-app:v0.1"
      resources:
        requests:
          cpu: 100m
          memory: 100M
  nodeSelector:
    kubernetes.io/arch: arm64
```



KubeCon



CloudNativeCon

Europe 2022

Karpenter

The Provisioner CRD

- Flexibility
 - Architecture
 - Capacity Type

```
apiVersion: karpenter.sh/v1alpha5
kind: Provisioner
metadata:
  name: default
spec:
  ttlSecondsAfterEmpty: 30
  ttlSecondsUntilExpired: 525600 # ~6 days
  requirements:
    - key: kubernetes.io/arch
      operator: In
      values: ["amd64", "arm64"]
    - key: karpenter.sh/capacity-type
      operator: In
      values: ["spot", "on-demand"]
    - key: node.kubernetes.io/instance-type
      operator: NotIn
      values: ["r5dn.large"]
  provider:
    subnetSelector:
      karpenter.sh/discovery: "my-cluster"
    securityGroupSelector:
      karpenter.sh/discovery: "my-cluster"
```



KubeCon



CloudNativeCon

Europe 2022

Karpenter

The Provisioner CRD

- Flexibility
 - Architecture
 - Capacity Type
 - Instance Type

```
apiVersion: karpenter.sh/v1alpha5
kind: Provisioner
metadata:
  name: default
spec:
  ttlSecondsAfterEmpty: 30
  ttlSecondsUntilExpired: 525600 # ~6 days
  requirements:
    - key: kubernetes.io/arch
      operator: In
      values: ["amd64", "arm64"]
    - key: karpenter.sh/capacity-type
      operator: In
      values: ["spot", "on-demand"]
    - key: node.kubernetes.io/instance-type
      operator: NotIn
      values: ["r5dn.large"]
  provider:
    subnetSelector:
      karpenter.sh/discovery: "my-cluster"
    securityGroupSelector:
      karpenter.sh/discovery: "my-cluster"
```



KubeCon



CloudNativeCon

Europe 2022

Karpenter

The Provisioner CRD

- Flexibility
 - Architecture
 - Capacity Type
 - Instance Type
 - GPU

```
apiVersion: v1
kind: Pod
metadata:
  name: cuda-vector-add
spec:
  containers:
    - name: cuda-vector-add
      image: "k8s.gcr.io/cuda-vector-add:v0.1"
      resources:
        limits:
          nvidia.com/gpu: 1
```




KubeCon

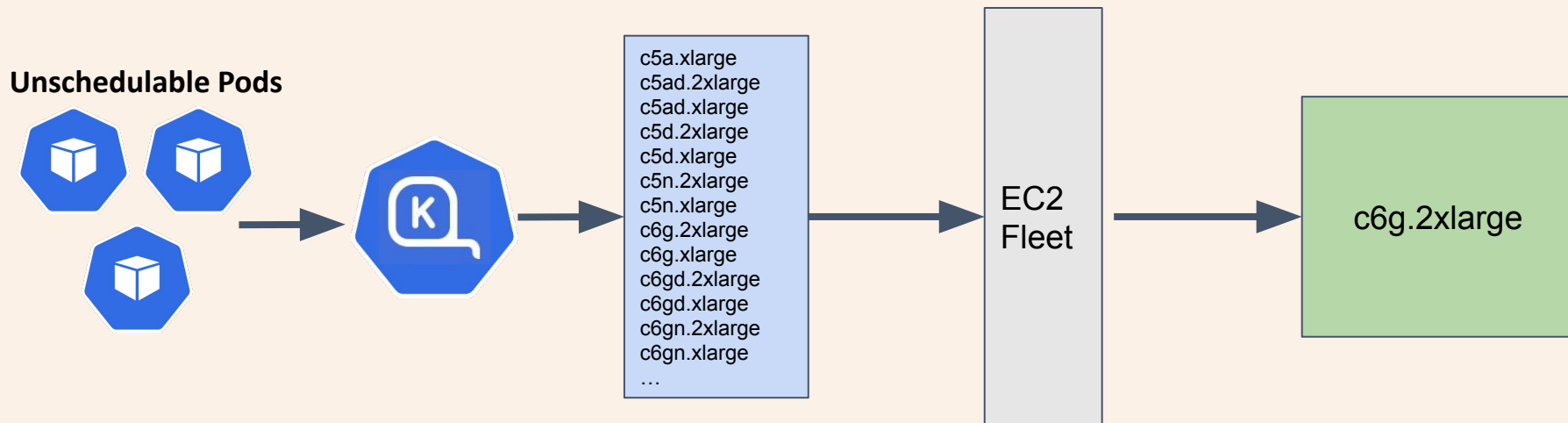


CloudNativeCon

Europe 2022

Karpenter - AWS Cloud Provider

- EC2 Fleet API
 - Flexible to many instance types
 - Chooses optimal AZ and instance type





KubeCon



CloudNativeCon

Europe 2022

Karpenter

```
apiVersion: karpenter.sh/v1alpha5
kind: Provisioner
metadata:
  name: default
spec:
  ttlSecondsAfterEmpty: 30
  ttlSecondsUntilExpired: 525600 # ~6 days
  requirements:
    - key: kubernetes.io/arch
      operator: In
      values: ["amd64", "arm64"]
    - key: karpenter.sh/capacity-type
      operator: In
      values: ["spot", "on-demand"]
    - key: node.kubernetes.io/instance-type
      operator: NotIn
      values: ["r5dn.large"]
  provider:
    subnetSelector:
      karpenter.sh/discovery: "my-cluster"
    securityGroupSelector:
      karpenter.sh/discovery: "my-cluster"
```

c6g.2xlarge
arm64
spot

g4dn.16xlarge
NVIDIA T4 GPU
spot

r6i.32xlarge
x86_64
on-demand

g4ad.8xlarge
AMD GPU
on-demand



KubeCon



CloudNativeCon

Europe 2022

Karpenter

- Scaling down
 - ttlSecondsAfterEmpty
 - ttlSecondsUntilExpired
- Gracefully Drains Pods

```
apiVersion: karpenter.sh/v1alpha5
kind: Provisioner
metadata:
  name: default
spec:
  ttlSecondsAfterEmpty: 30
  ttlSecondsUntilExpired: 525600 # ~6 days
  requirements:
    - key: kubernetes.io/arch
      operator: In
      values: ["amd64", "arm64"]
    - key: karpenter.sh/capacity-type
      operator: In
      values: ["spot", "on-demand"]
    - key: node.kubernetes.io/instance-type
      operator: NotIn
      values: ["r5dn.large"]
  provider:
    subnetSelector:
      karpenter.sh/discovery: "my-cluster"
    securityGroupSelector:
      karpenter.sh/discovery: "my-cluster"
```



KubeCon



CloudNativeCon

Europe 2022

Karpenter

- AWS Node Termination Handler
- Interruption Handling
- Rebalance Handling

```
apiVersion: karpenter.sh/v1alpha5
kind: Provisioner
metadata:
  name: default
spec:
  ttlSecondsAfterEmpty: 30
  ttlSecondsUntilExpired: 525600 # ~6 days
  requirements:
    - key: kubernetes.io/arch
      operator: In
      values: ["amd64", "arm64"]
    - key: karpenter.sh/capacity-type
      operator: In
      values: ["spot", "on-demand"]
    - key: node.kubernetes.io/instance-type
      operator: NotIn
      values: ["r5dn.large"]
  provider:
    subnetSelector:
      karpenter.sh/discovery: "my-cluster"
    securityGroupSelector:
      karpenter.sh/discovery: "my-cluster"
```



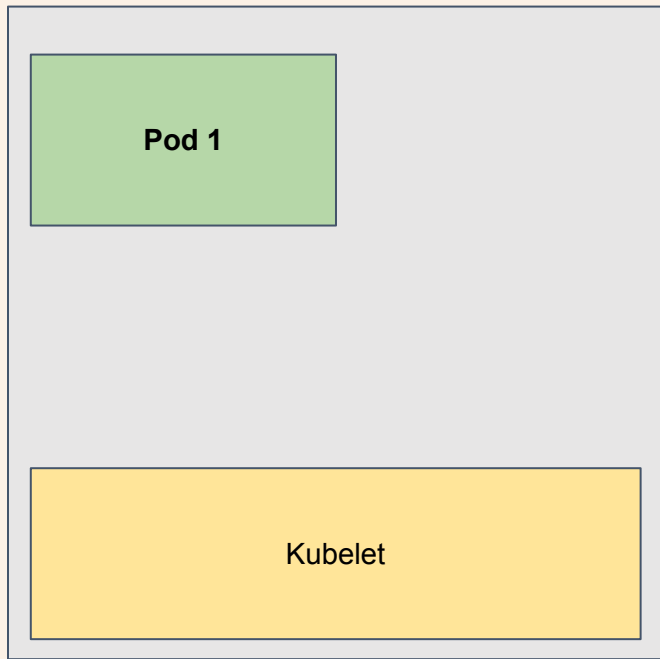
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API



Node



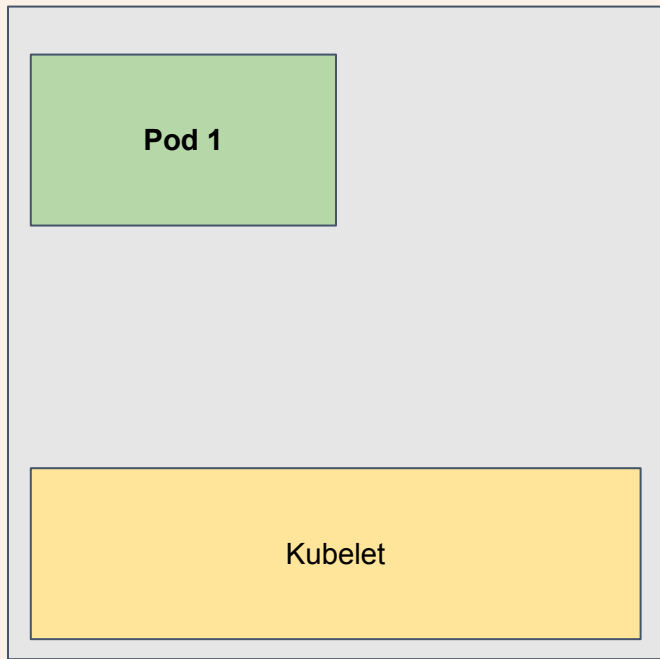
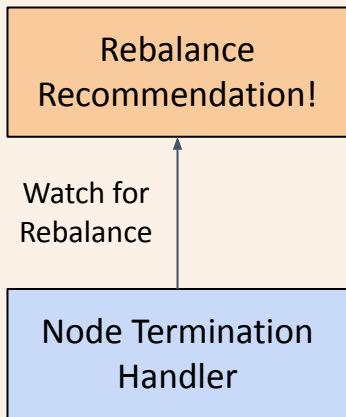
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API



Node



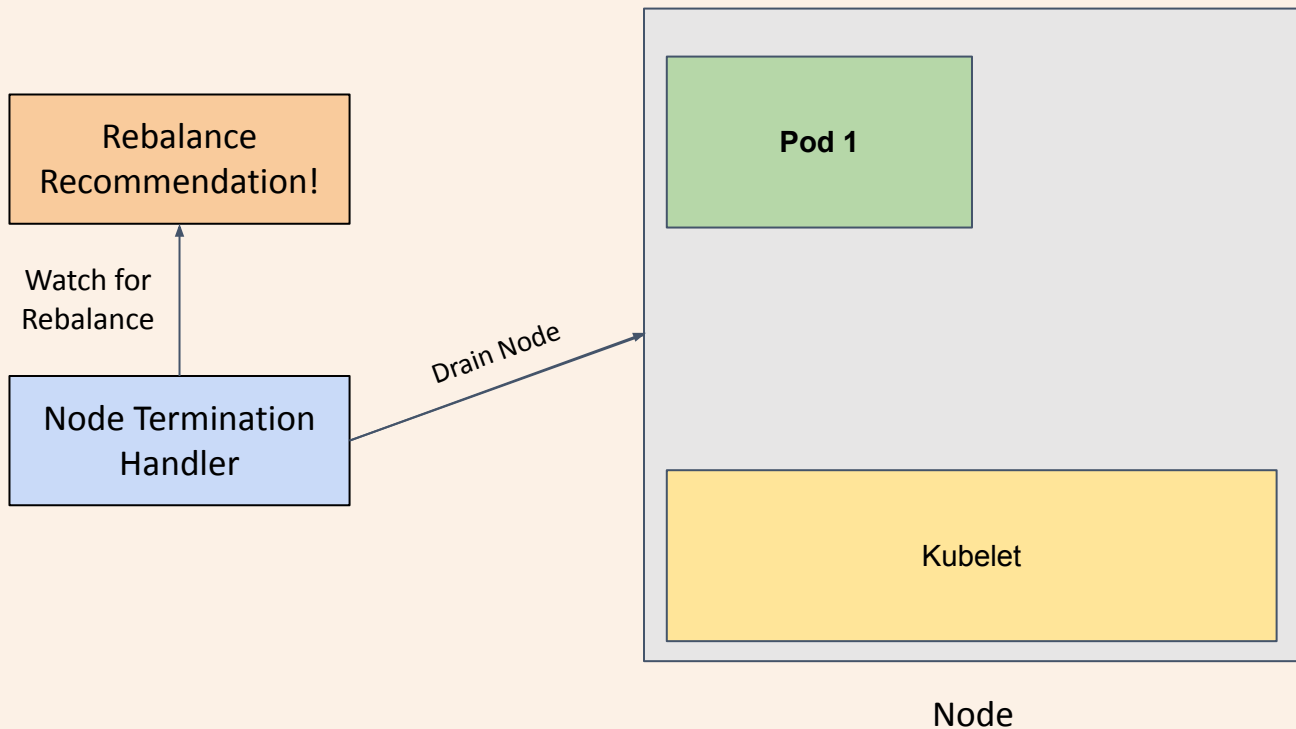
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API





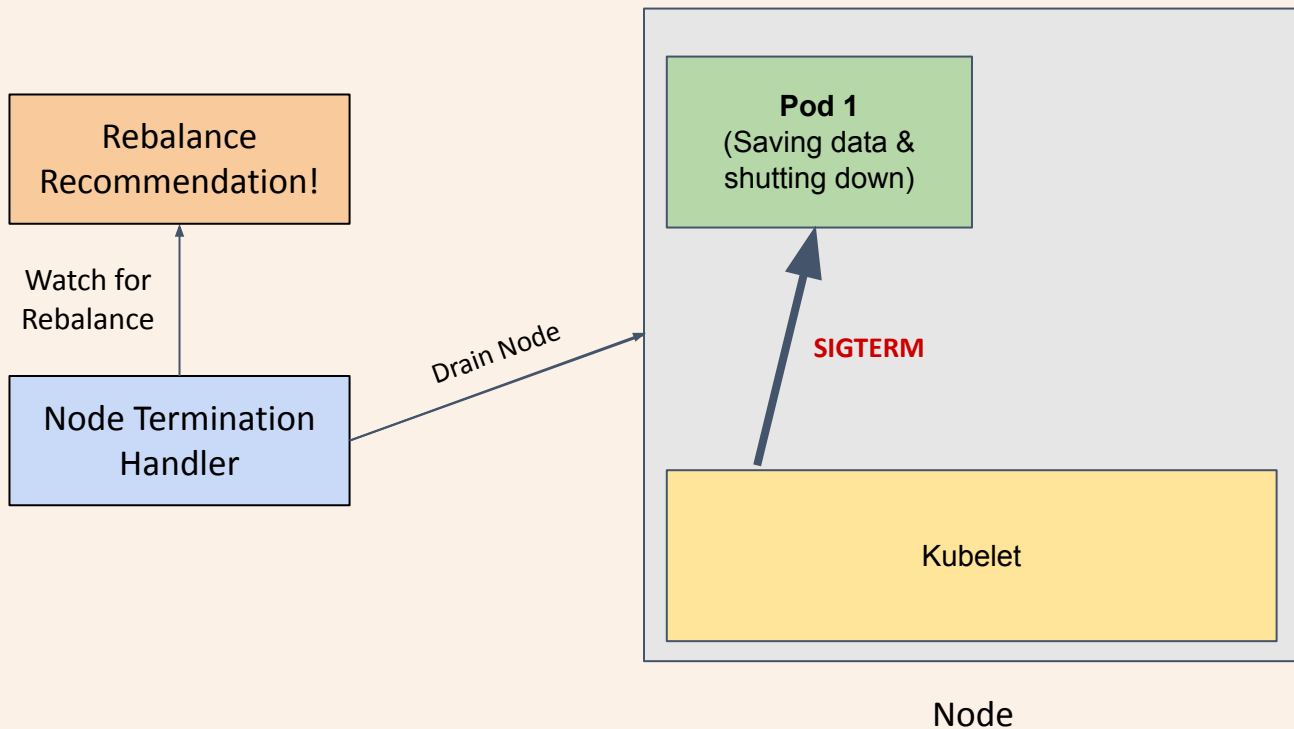
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API





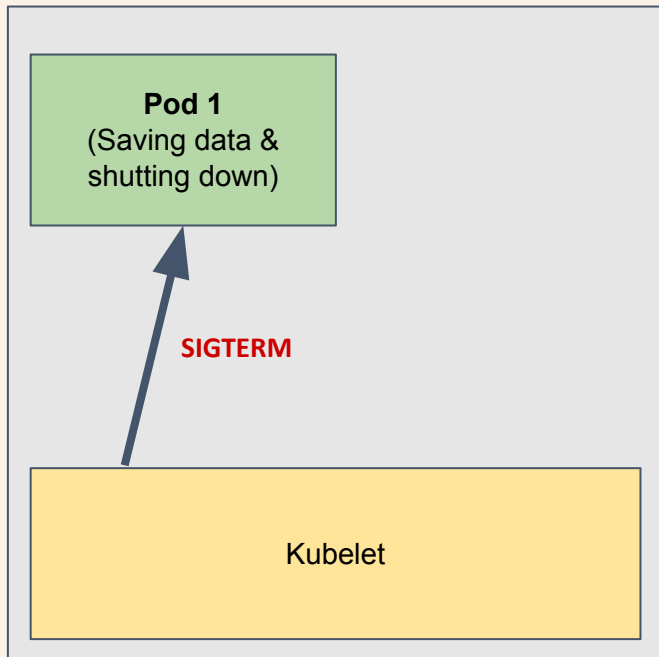
KubeCon



CloudNativeCon

Europe 2022

K8s and Spot - K8s Eviction API



Scaling K8s Nodes Without Breaking the Bank or Your Sanity



KubeCon



CloudNativeCon

Europe 2022

Agenda

- What is Spot? 🐕
- Best Practices 🏆
- K8s & Spot ❤️
- Autoscaling your nodes 🐳
 - Cluster Autoscaler
 - Karpenter
- Demo 🎪



Brandon Wagner
Software Engineer
AWS



Nick Tran
Software Engineer
AWS



KubeCon



CloudNativeCon

Europe 2022

DEMO

