KubeCon | CloudNativeCon

Europe 2018

@jsloyer @dancberg

Not one Size Fits All - How to Size Kubernetes Clusters

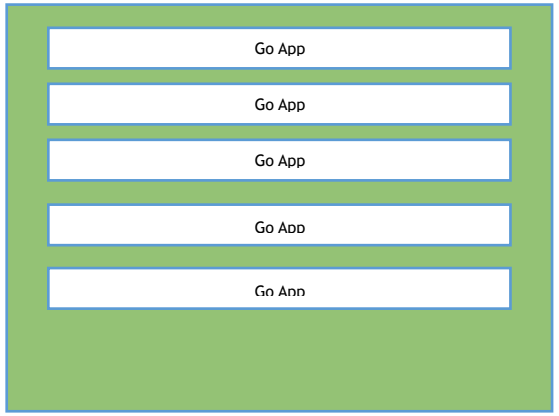
Jeff Sloyer - Dev Lead - IBM Kubernetes Service

Dan Berg - Architect - IBM Kubernetes Service

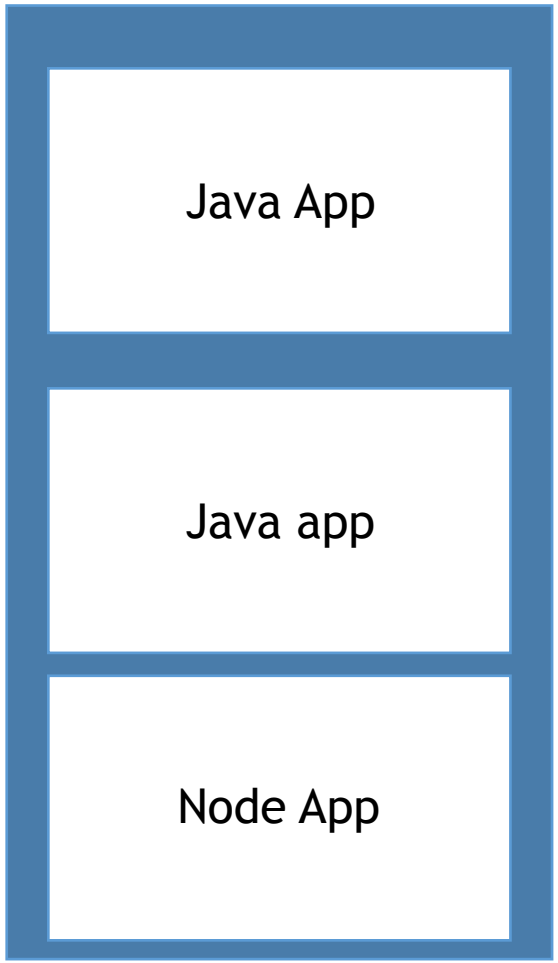
Cornucopia of Node Types

- Have multiple node types
- CPU vs memory
- Understanding makeup of your apps

4x16



4x16



16x64

@jsloyer @dancberg

Limits & Requests

```
spec:
  containers:
  - image: nginx
    imagePullPolicy: Always
    name: qos-demo-2-ctr
    resources:
      limits:
        memory: 200Mi
        cpu: 3
      requests:
        memory: 100Mi
        cpu: .5
...
  qosClass: Guaranteed
```

- Improve Kubernetes scheduler efficiency
- Requests determine minimum cpu/memory required by a container
- Limits set the max cpu/memory allowed to be used
- Use LimitRange to set defaults

Name	CPU Requests	CPU Limits	Memory Requests	Memory Limits
----	----	----	-------	-------
api-74cf7d665d-2qrg8	500m (12%)	3 (75%)	100Mi (0%)	0 (0%)
api-74cf7d665d-mdwkx	500m (12%)	3 (75%)	100Mi (0%)	0 (0%)

@jsloyer @dancberg

Selectors

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: nginx
  labels:
    env: test
spec:
  containers:
  - name: nginx
    image: nginx
    imagePullPolicy: IfNotPresent
  nodeSelector:
    disktype: ssd
```

- Lock an app to a particular node type
- Helpful for performance

Anti-Affinity Policies

```
...
    matchLabels:
      app: store
  replicas: 3
  template:
    metadata:
      labels:
        app: store
    spec:
      affinity:
        podAntiAffinity:
          requiredDuringSchedulingIgnoredDuringExecution:
          - labelSelector:
              matchExpressions:
              - key: app
                operator: In
                values:
                - store
            topologyKey: "kubernetes.io/hostname"
```

- Used to spread PODs across worker nodes
- Use "preferredDuringSchedulingIgnoredDuringExecution"
- Ensures availability if there is a worker node failure

@jsloyer @dancberg

Taints & Tolerations


```
kubectl taint nodes mygpunode special=true:NoSchedule
```

```
tolerations:

- key: "gpu"

  operator: "Equal"

  value: "true"

  effect: "NoSchedule"
```

- Dedicate node/nodes for certain uses
- Prevents scheduling on the node
- Only can schedule with a toleration