



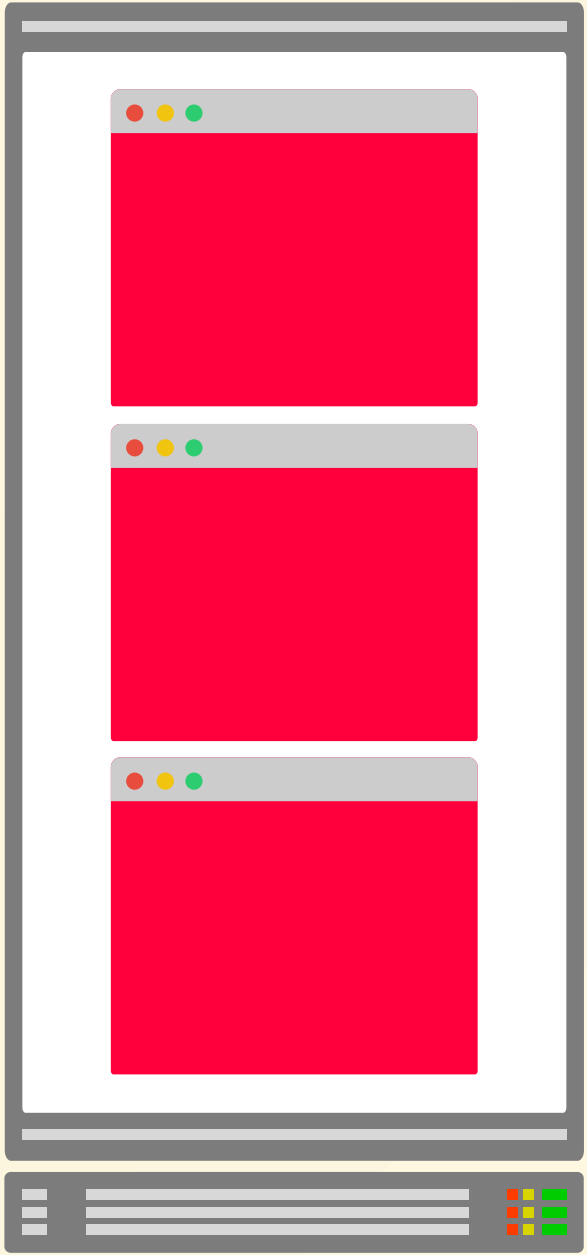
# Proactive cluster autoscaling in Kubernetes

**Chris Nesbitt-Smith**

- **Learnk8s - Instructor+consultant**
- **Crown Prosecution Service (UK gov) - Consultant**



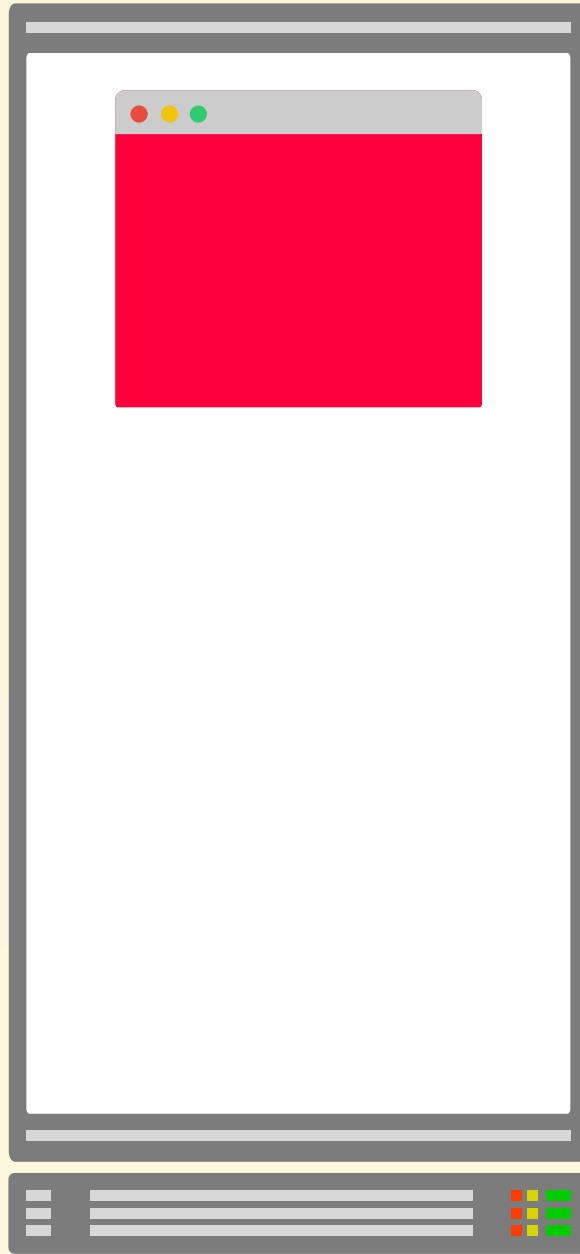
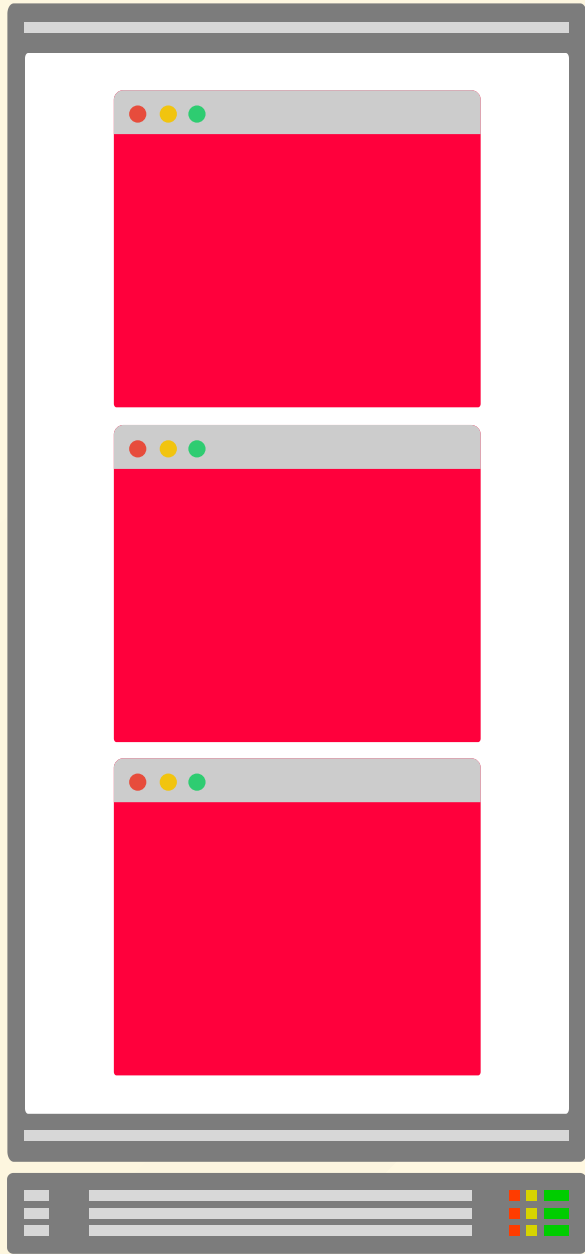


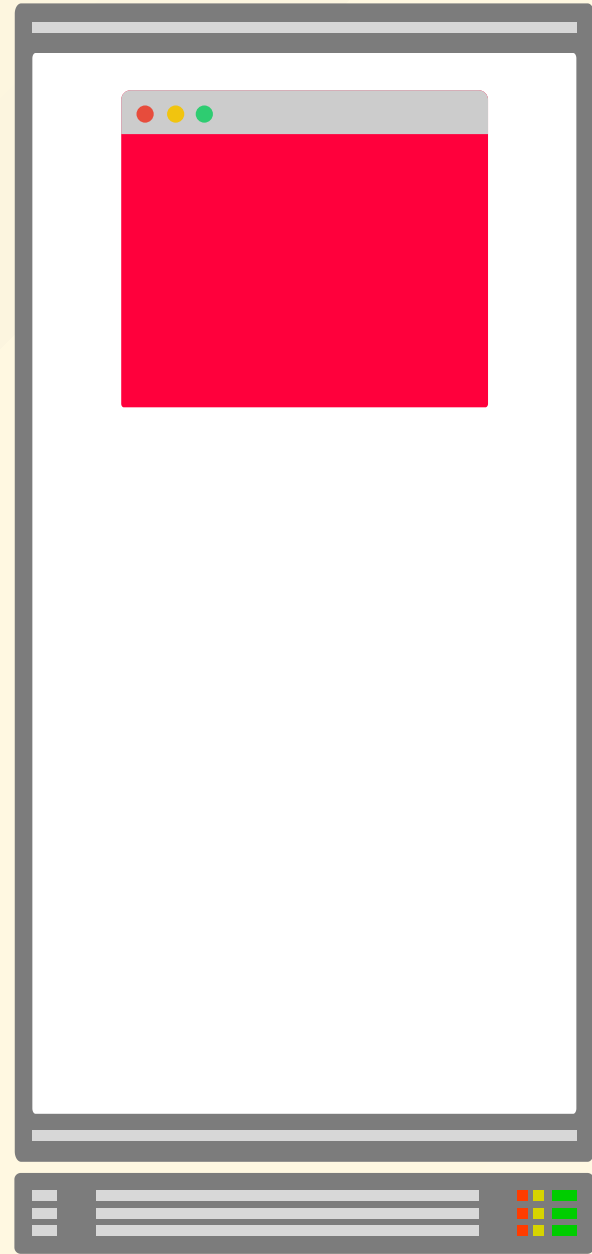
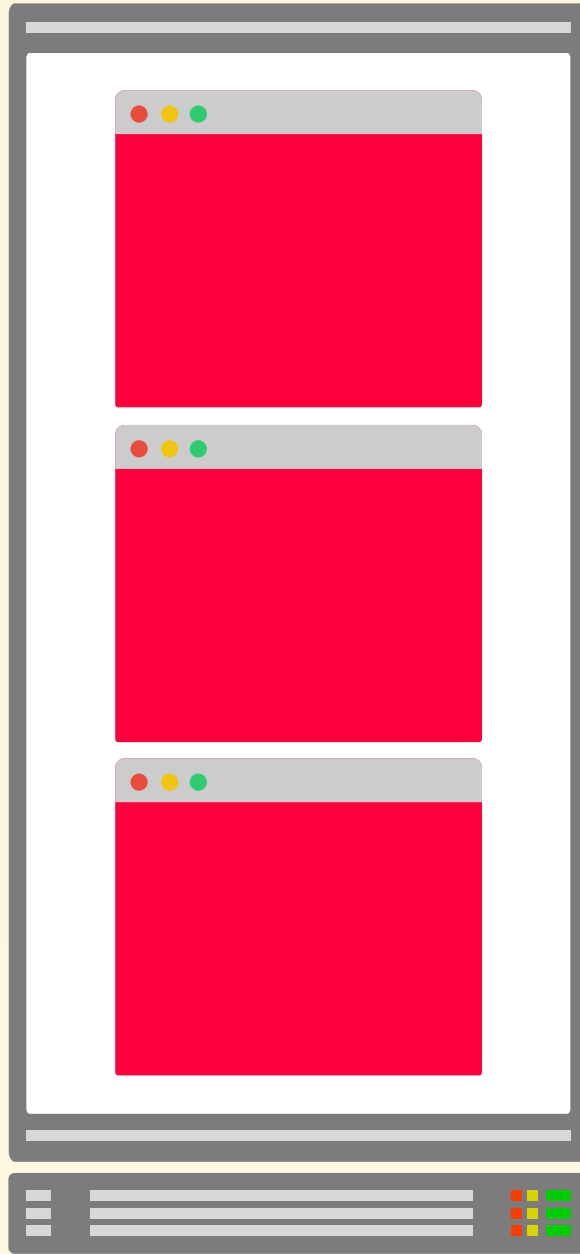
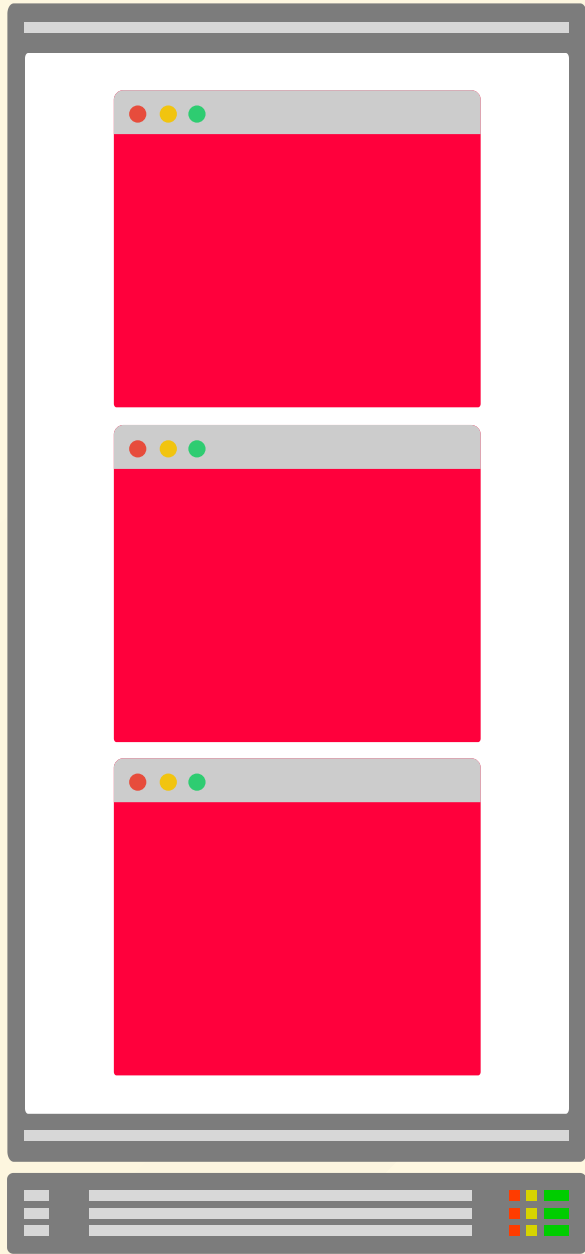


# Cluster AutoScaler

[github.com/kubernetes/autoscaler/tree/master/cluster-autoscaler](https://github.com/kubernetes/autoscaler/tree/master/cluster-autoscaler)







# Cluster AutoScaler

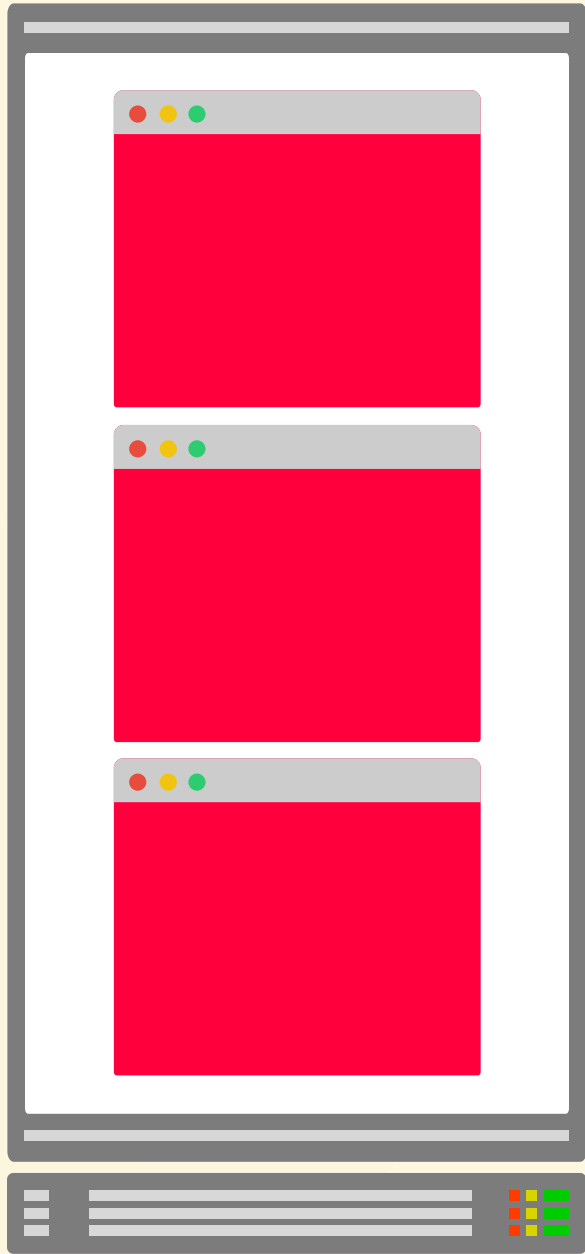
- 1. Memory Utilization**
- 2. CPU Utilization**
- 3. Pending pods**



# Cluster AutoScaler

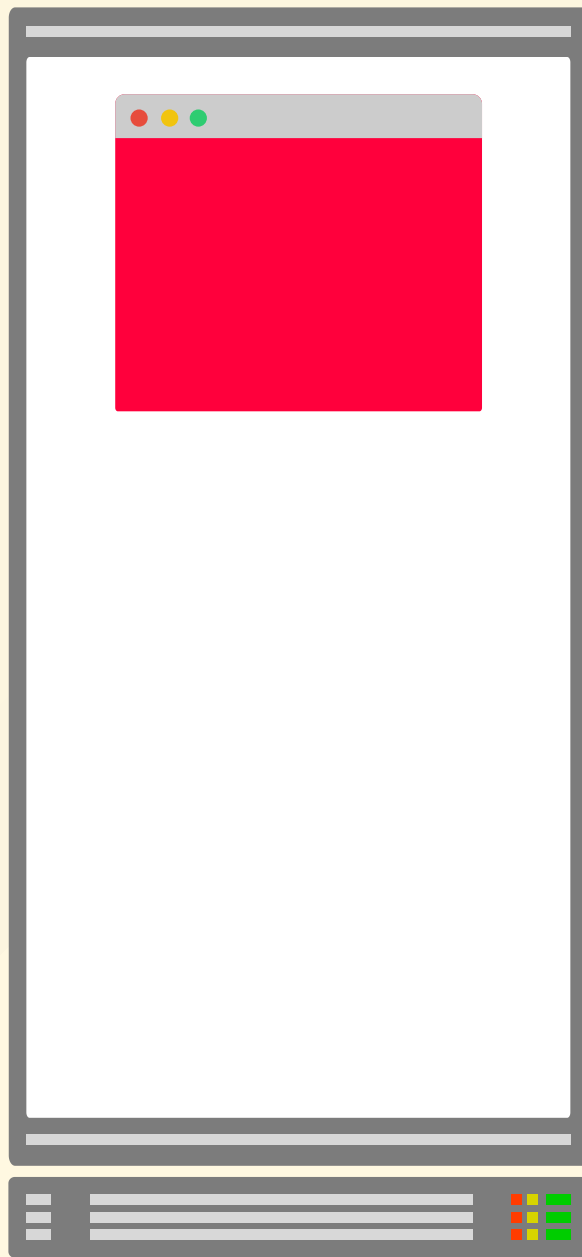
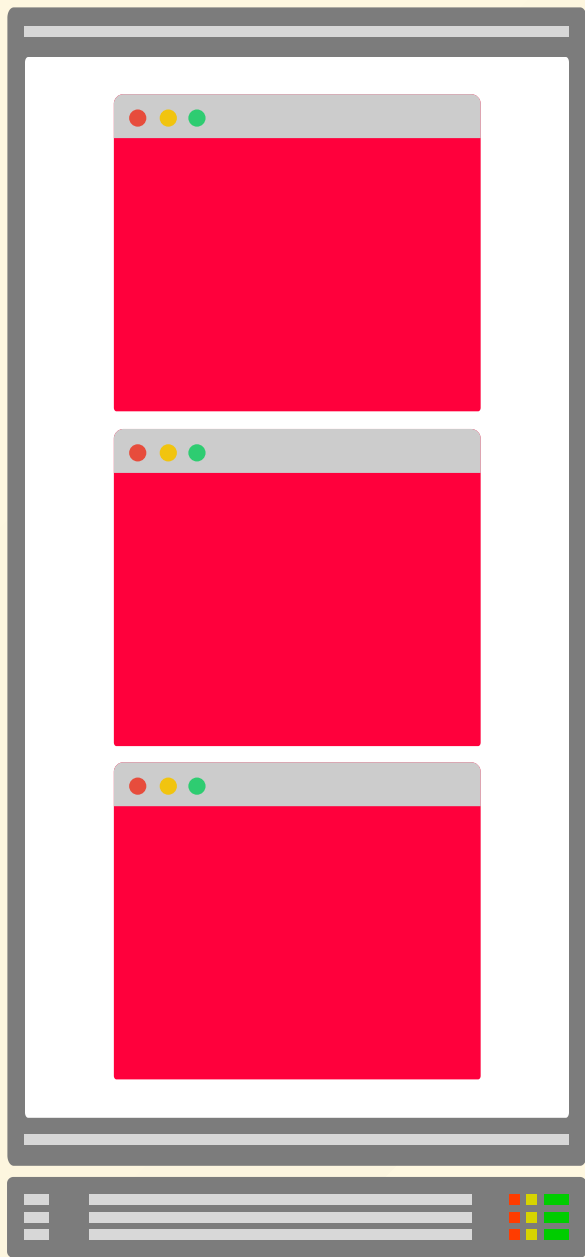
1. ~~Memory Utilization~~
2. ~~CPU Utilization~~
3. Pending pods





**PENDING** 🤔



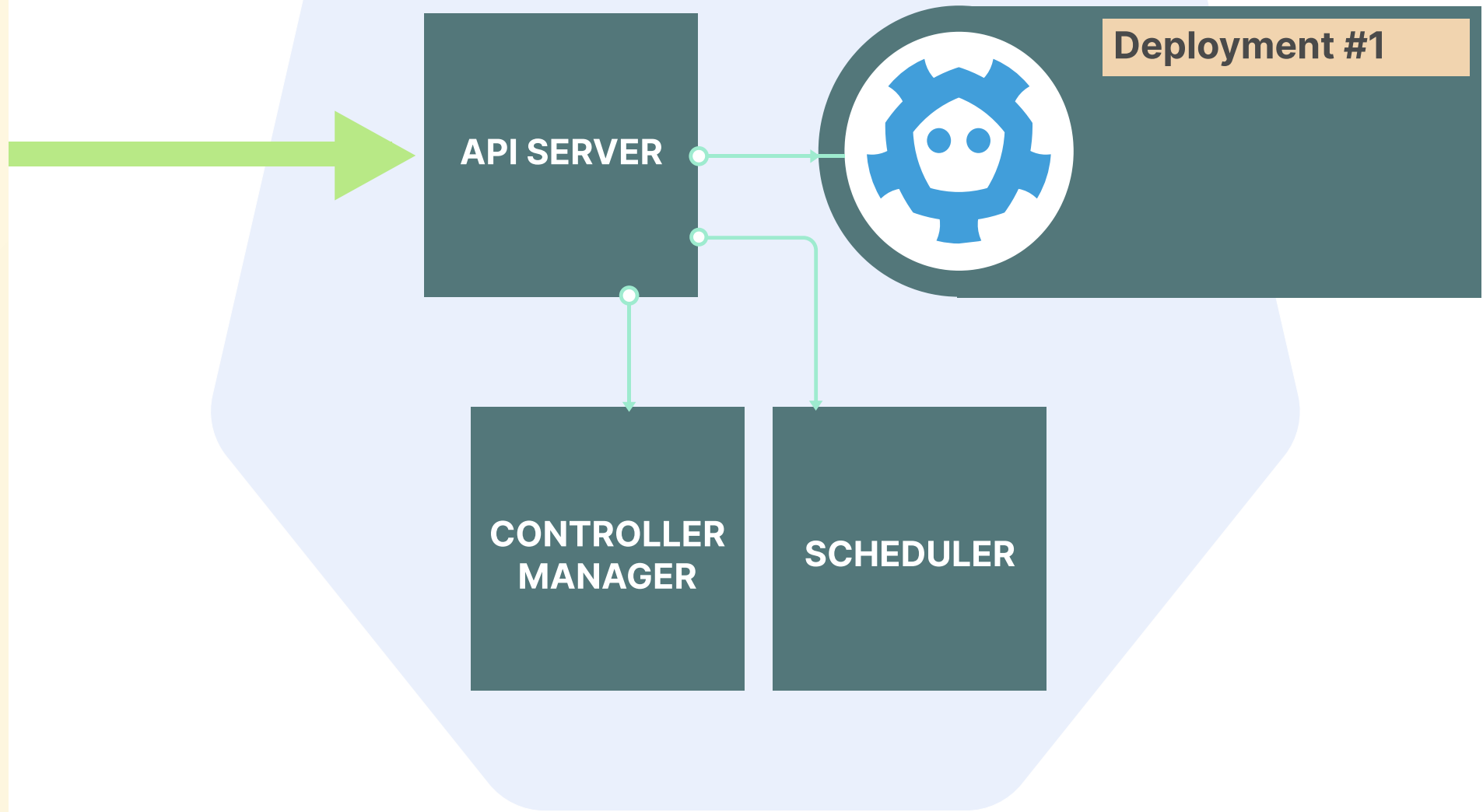


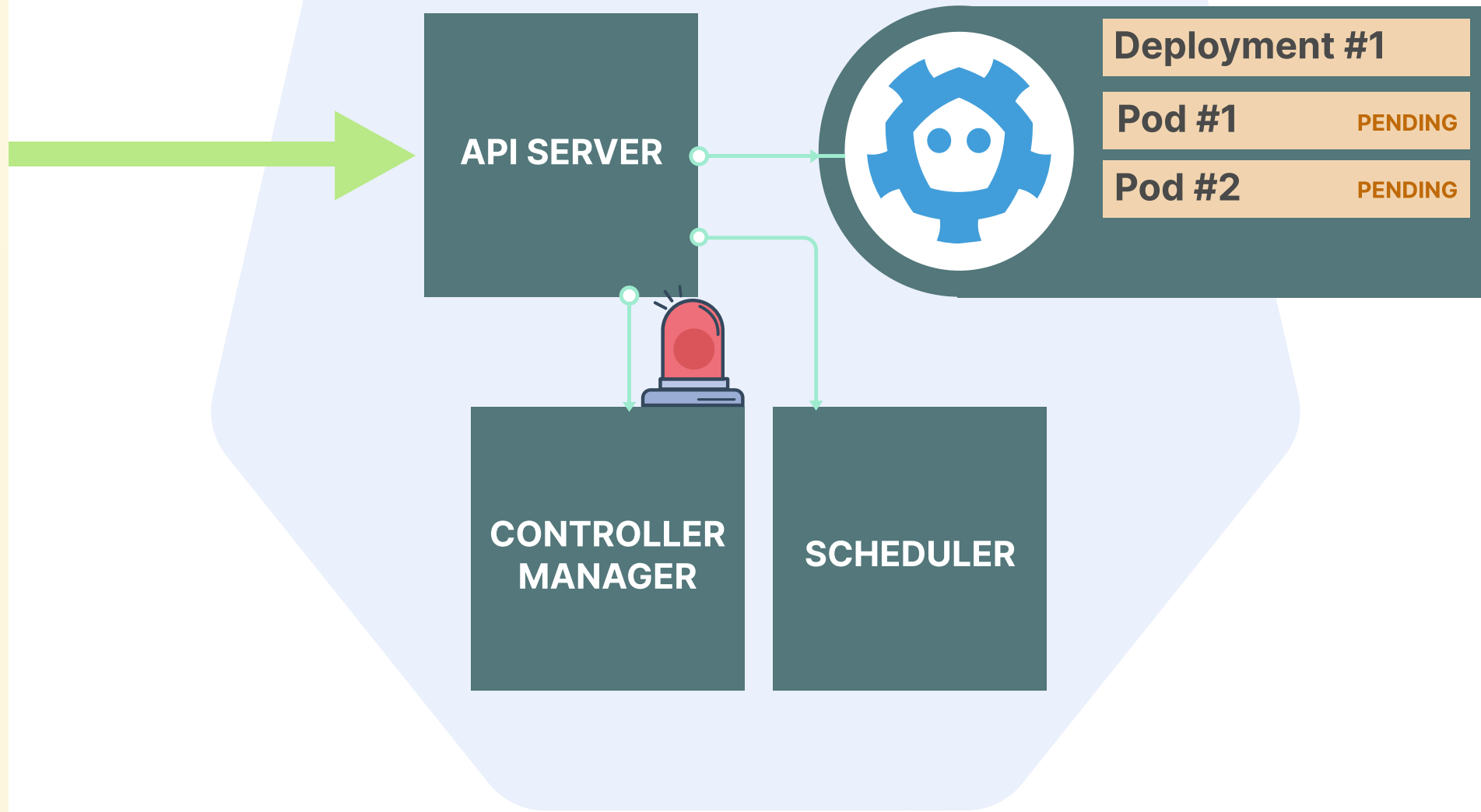
# Kubernetes Scheduler

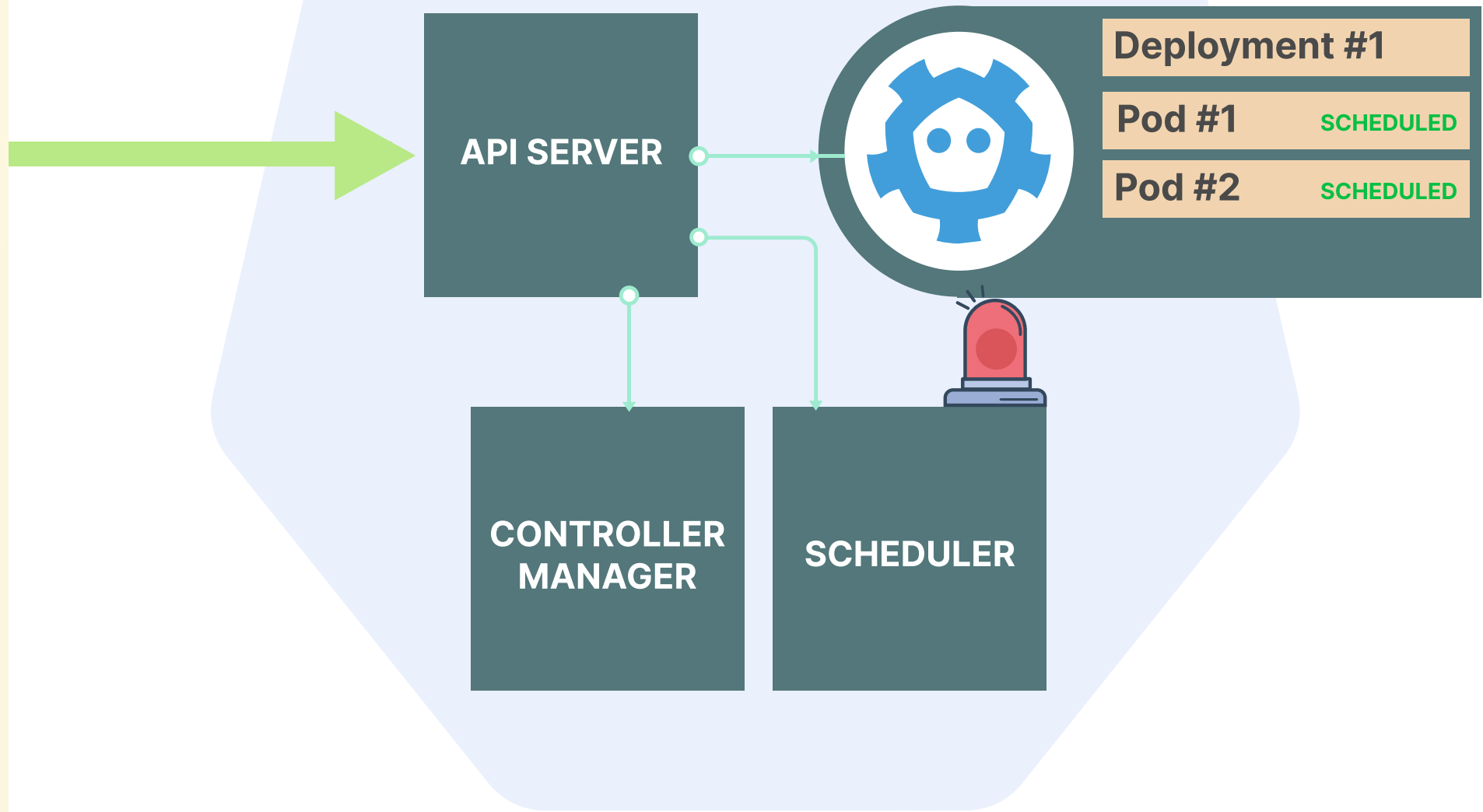


# Create a Deployment with 2 replicas









Queue

Filter

Score

(Notifier)

(Binding Policies)

**Scheduling**

Binding

**Binding**



# Requests & Scheduler



## Memory

2048MB

Limit

1

1024MB

Max memory available

512MB

Request

2

Process



---

resources:

  requests:

    memory: "64Mi"

    cpu: "250m"

  limits:

    memory: "128Mi"

    cpu: "500m"



8 vCPU

4 vCPU

1 vCPU

CPU

0

2GiB

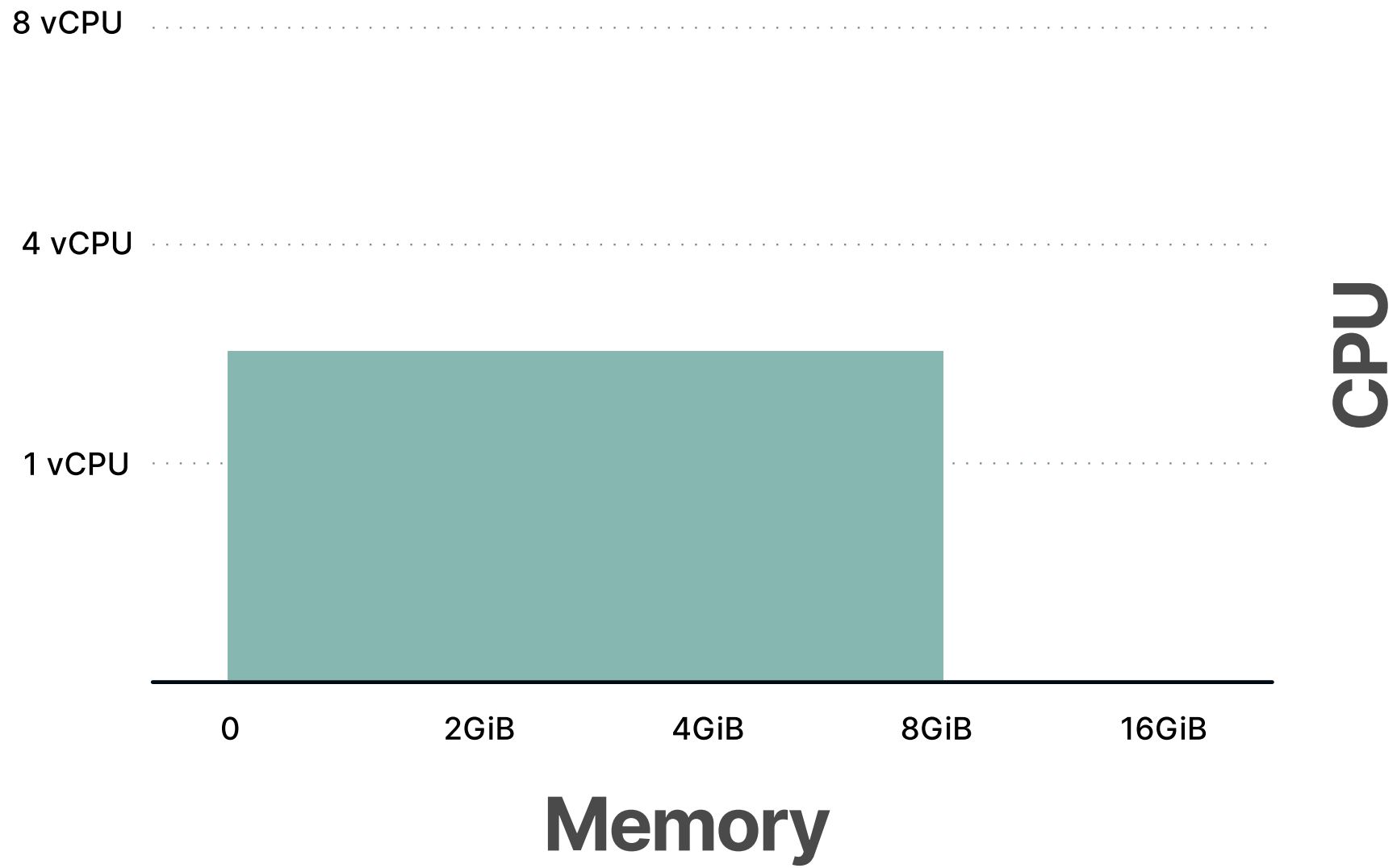
4GiB

8GiB

16GiB

Memory





0

2GiB

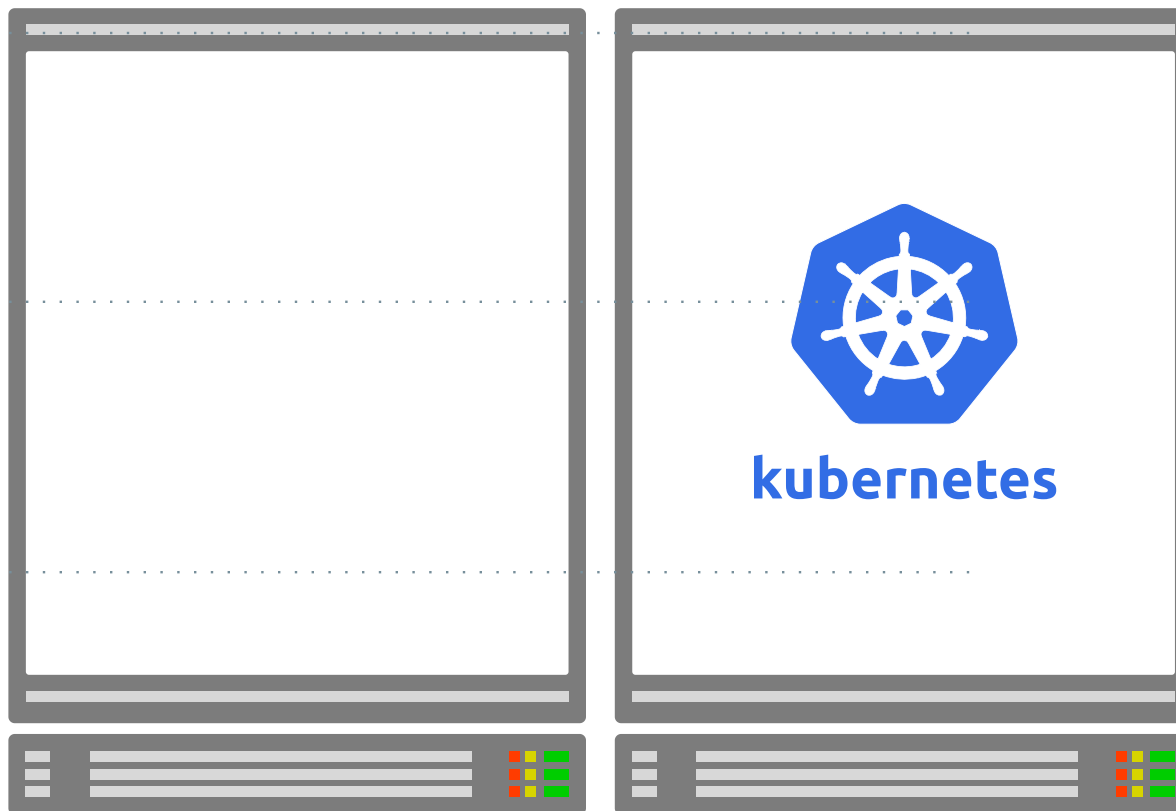
4GiB

8GiB

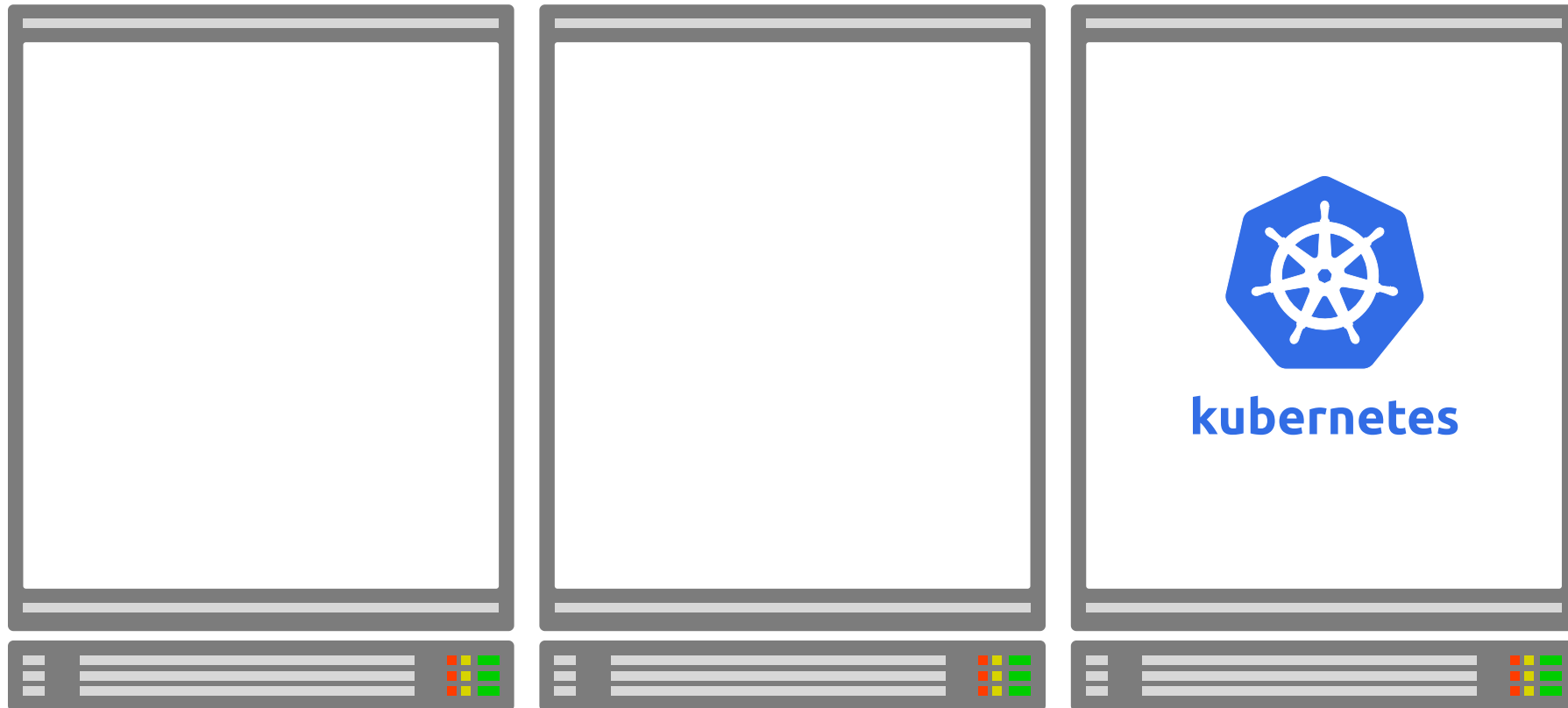
1 vCPU

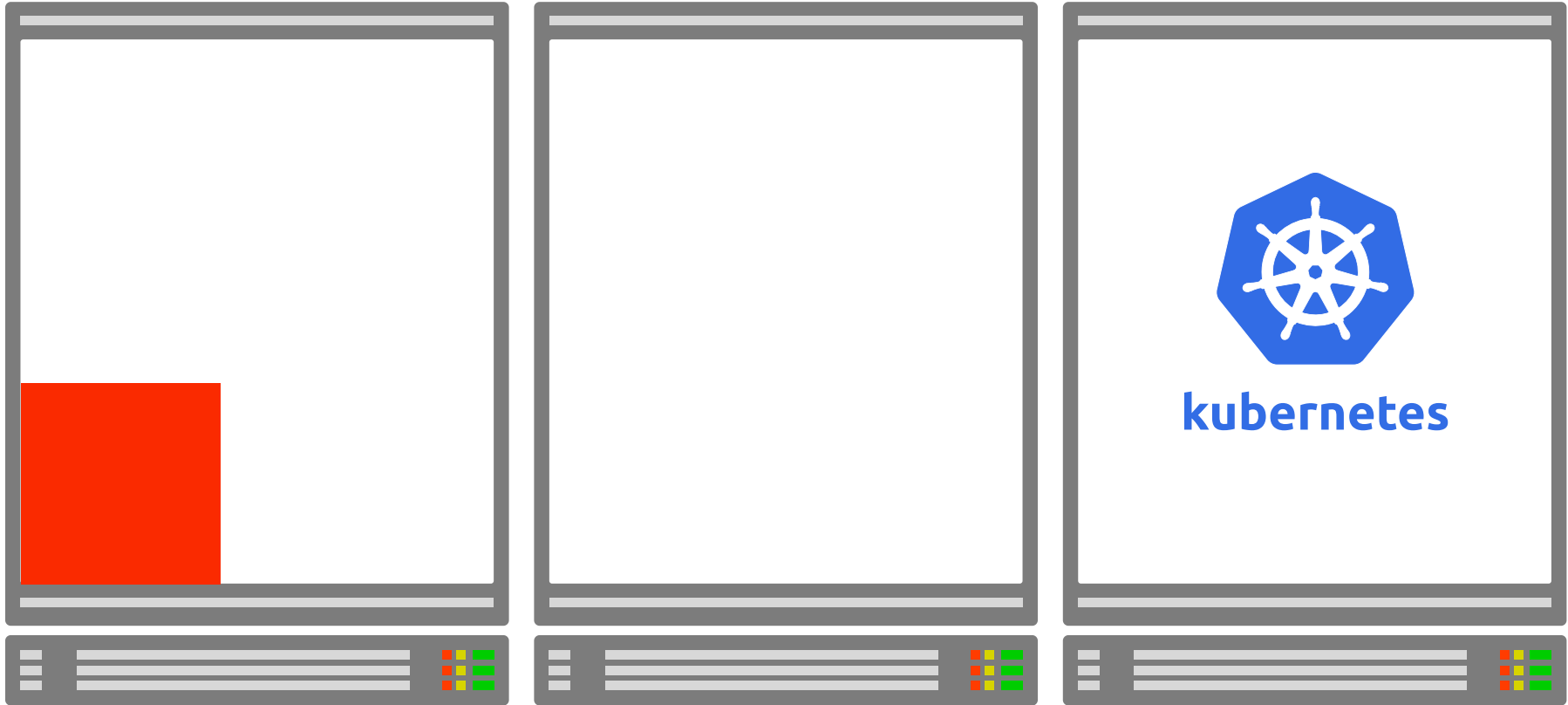
4 vCPU

8 vCPU

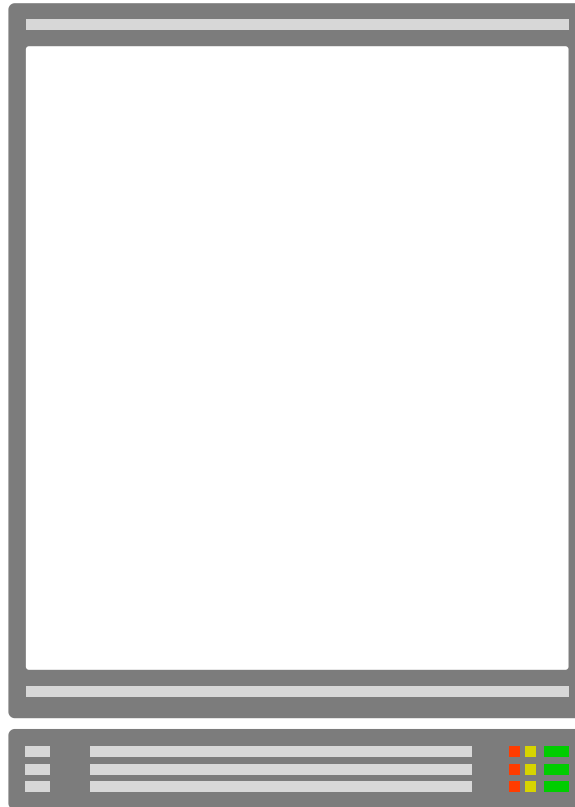
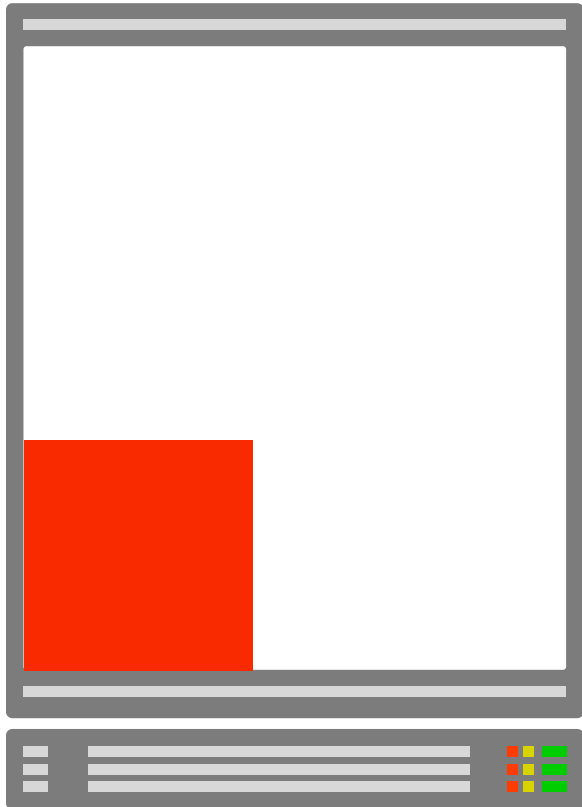


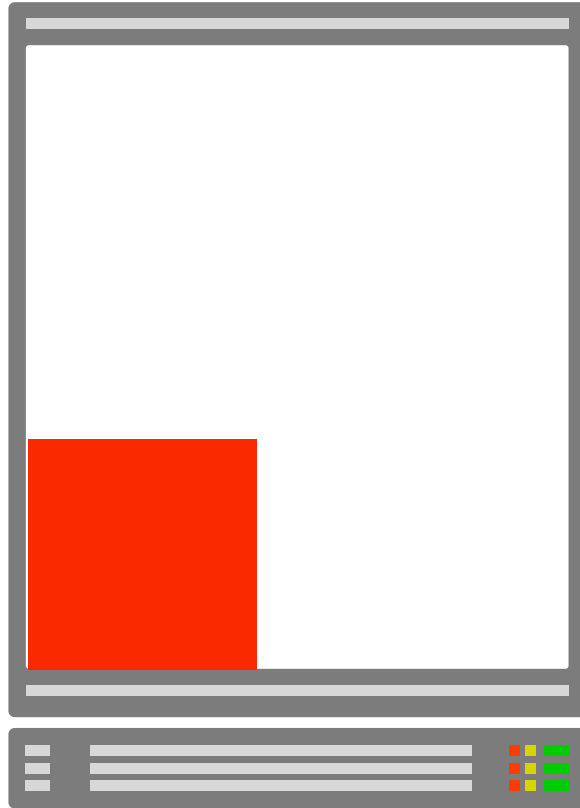
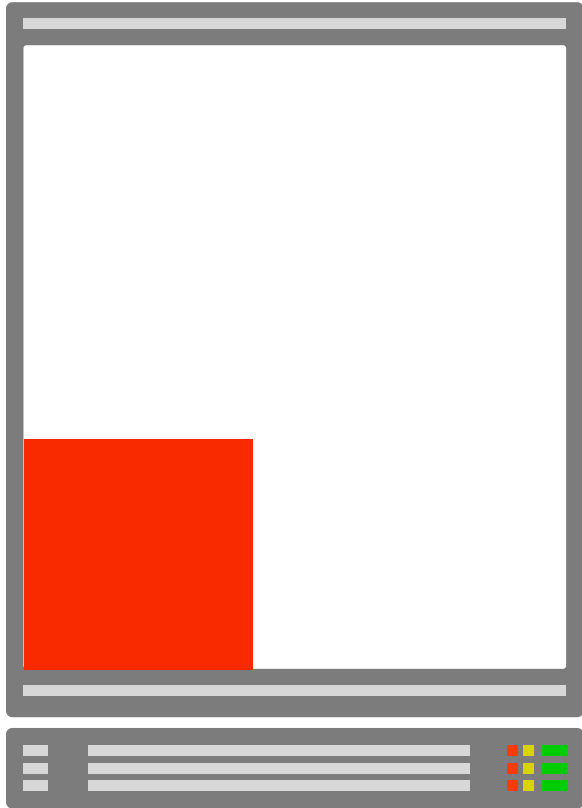
Next container



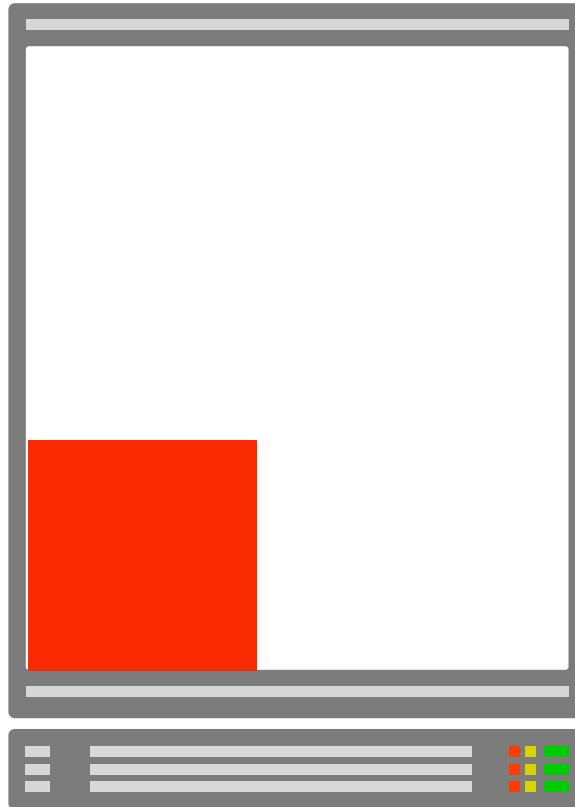
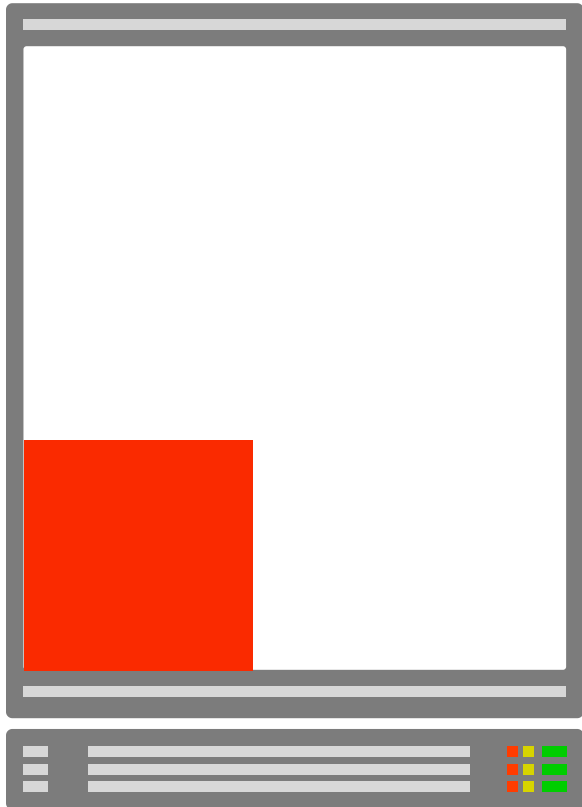


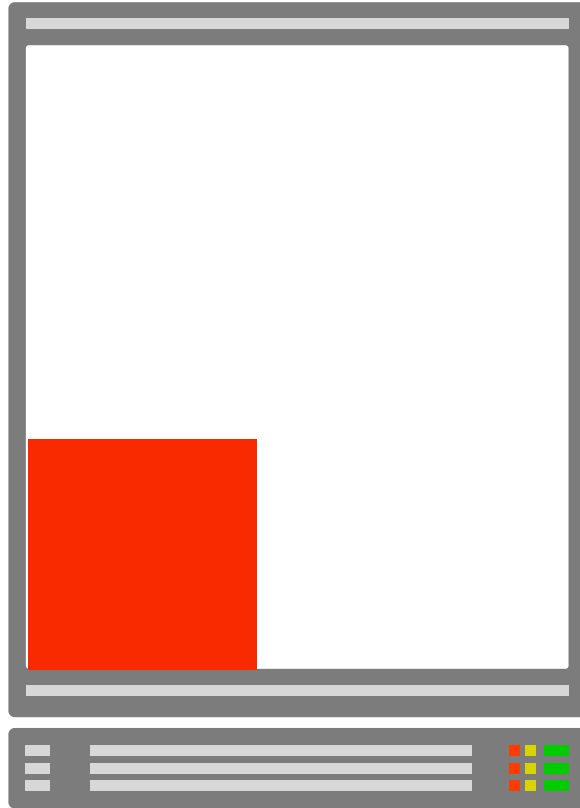
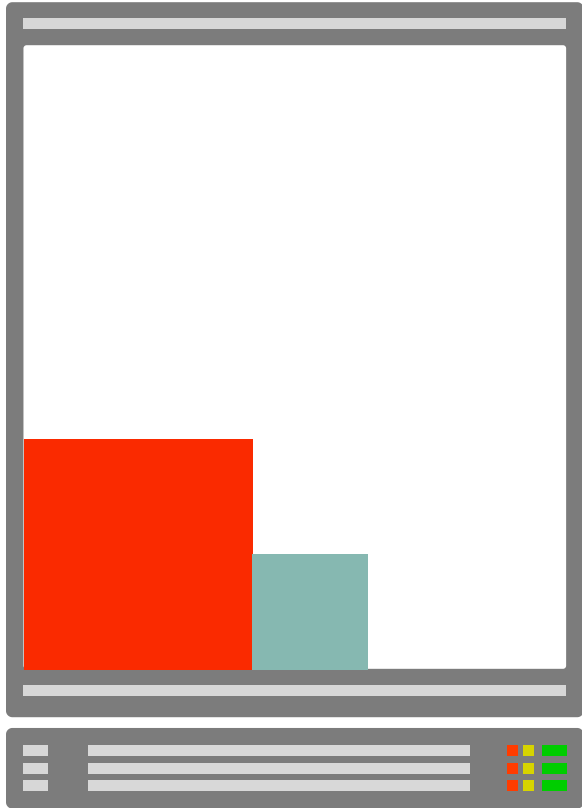
Next container



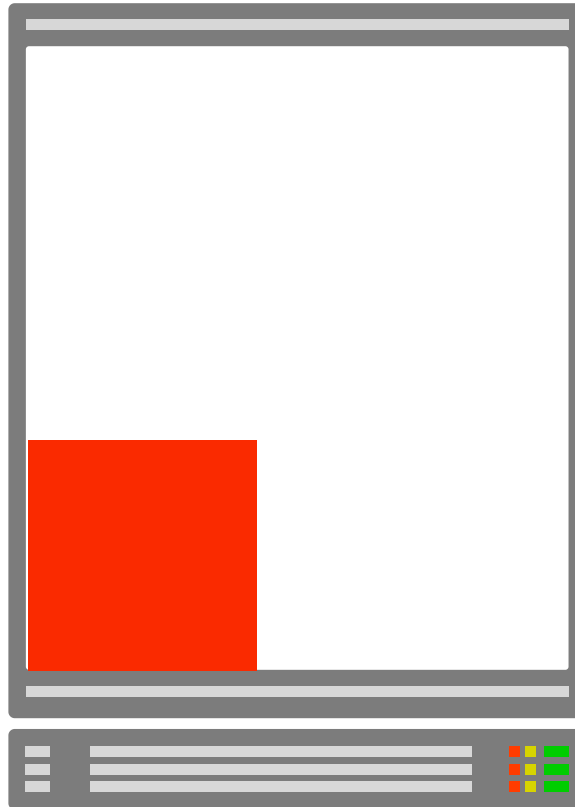
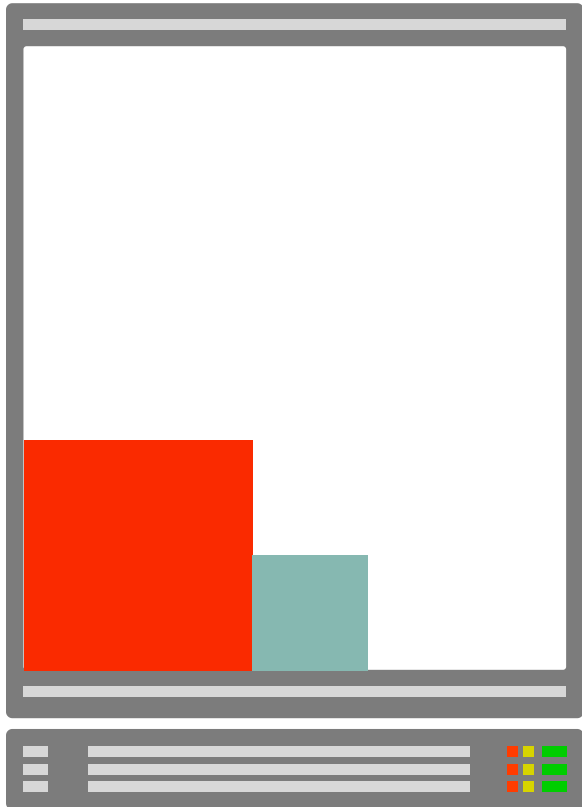


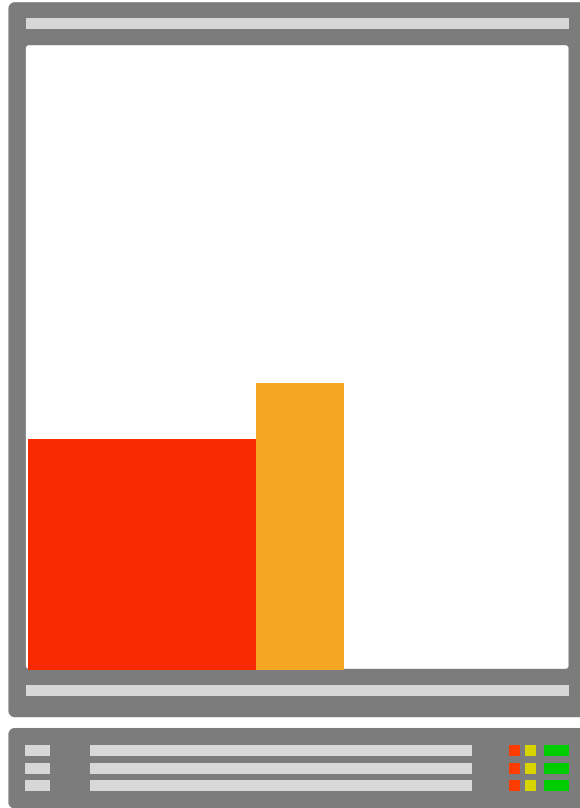
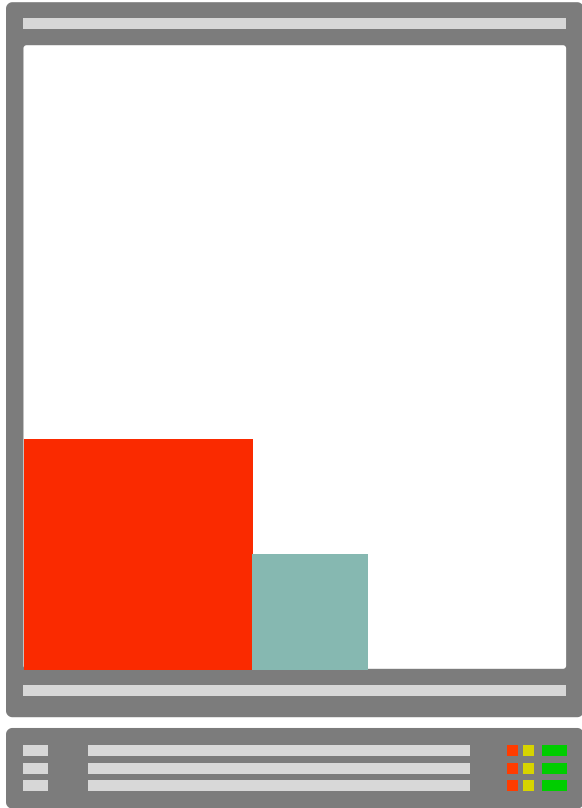
Next container



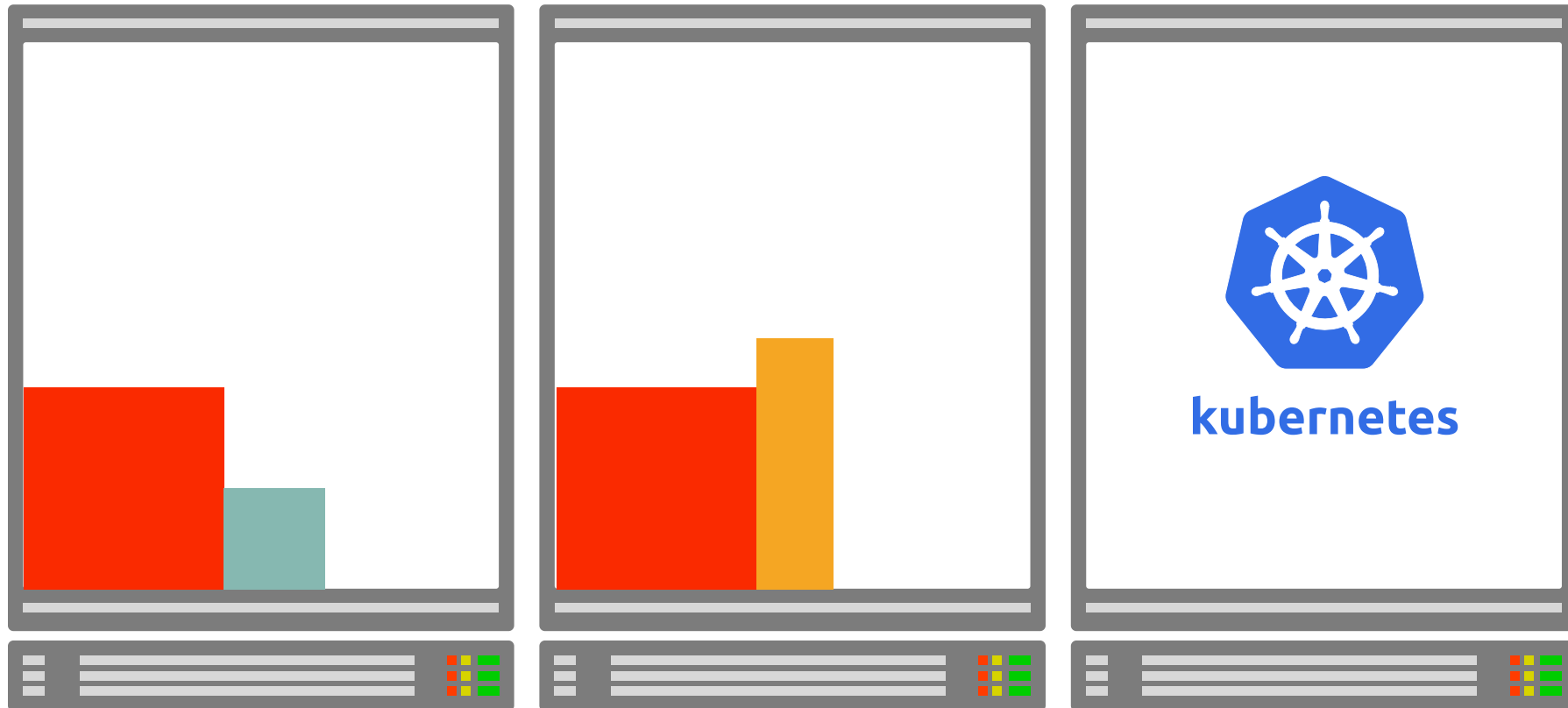


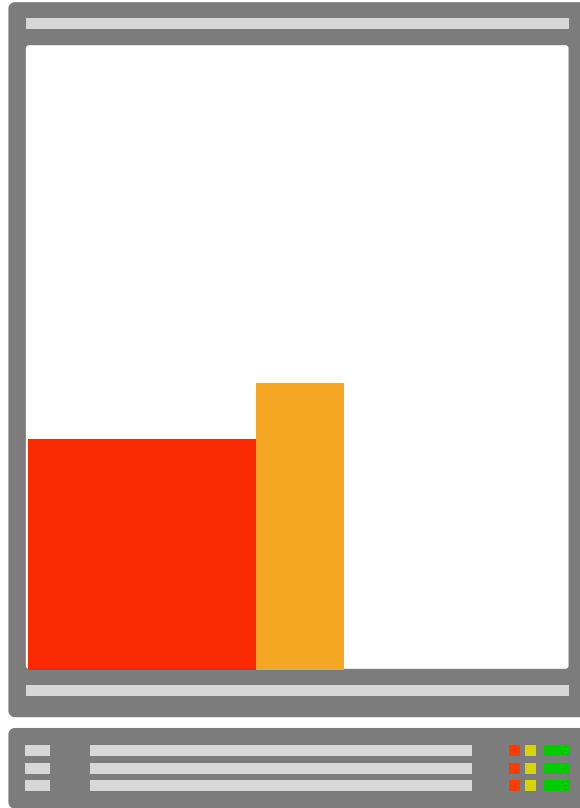
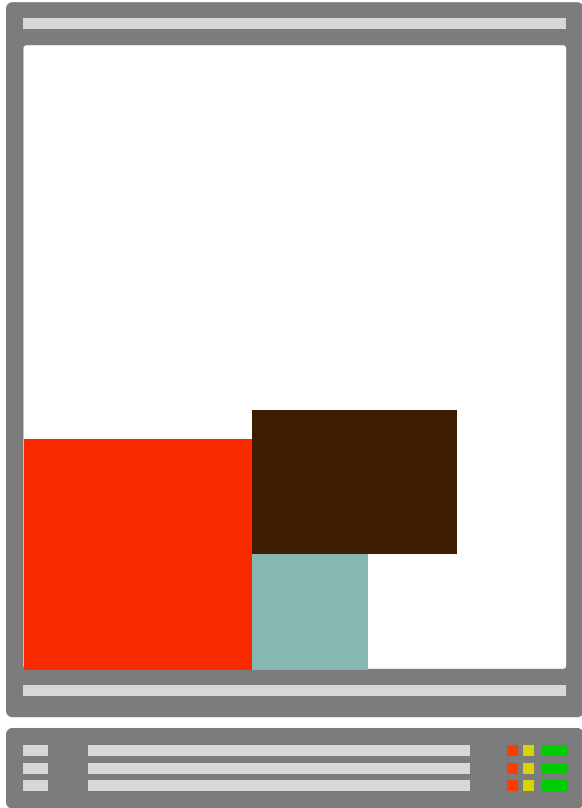
Next container



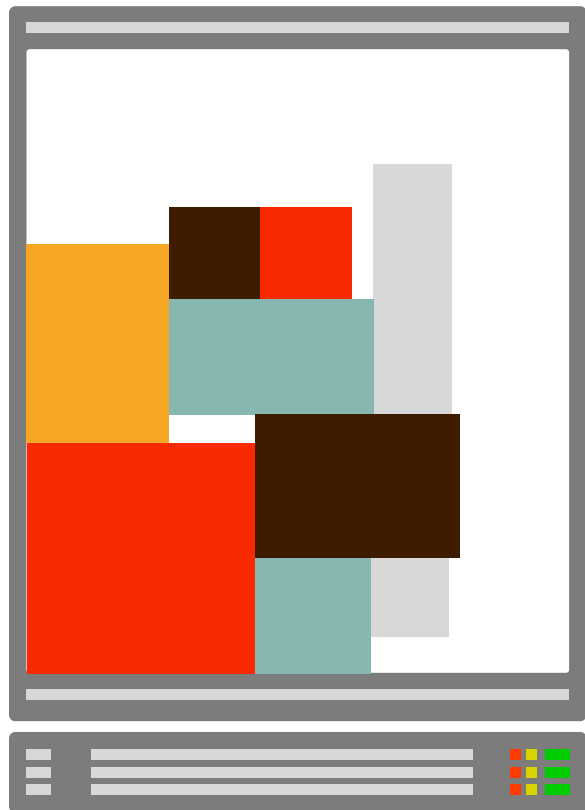


Next container

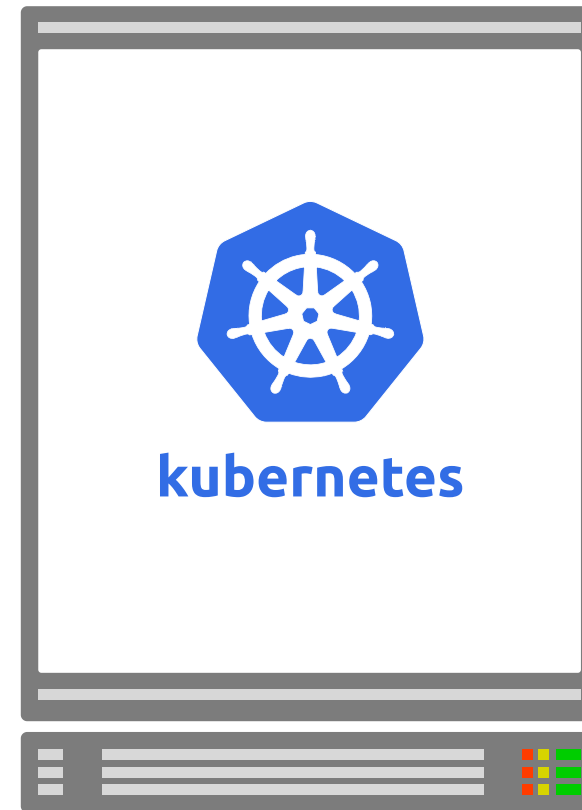
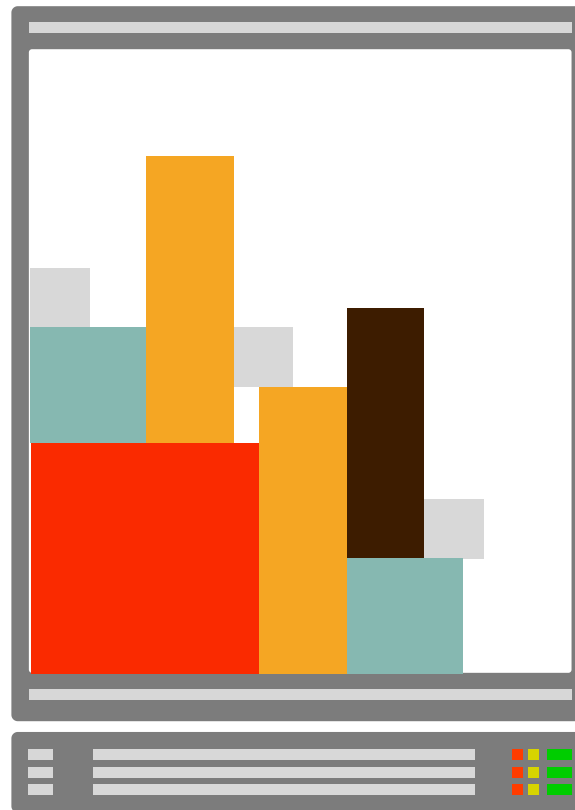




**Worker Node**



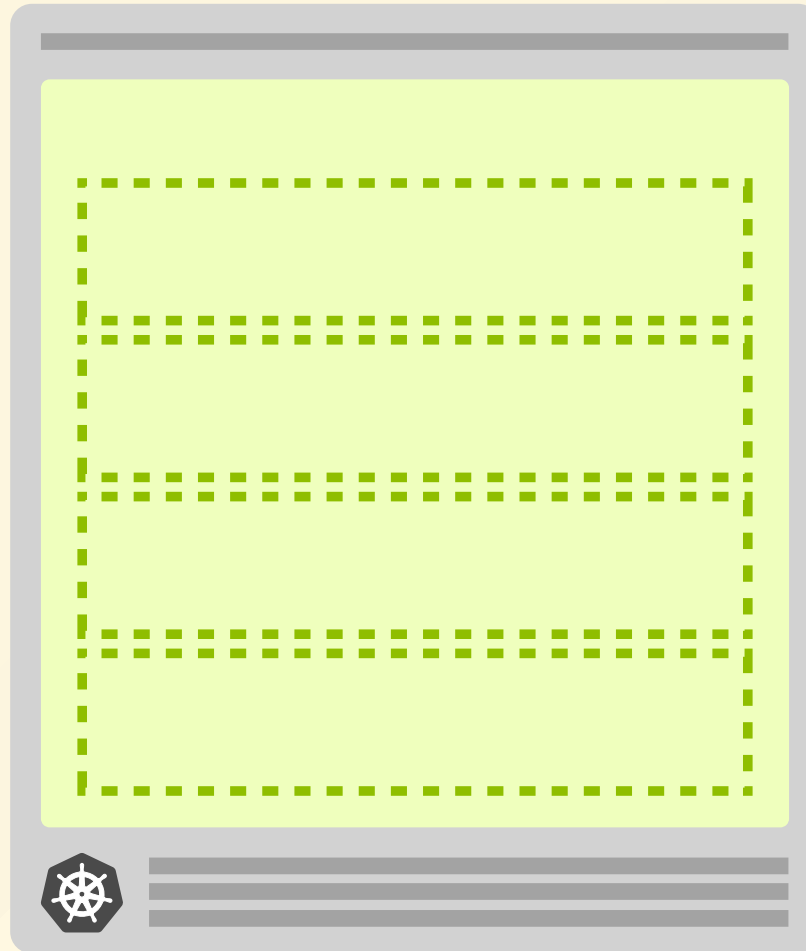
**Worker Node**



# Requests & Limits

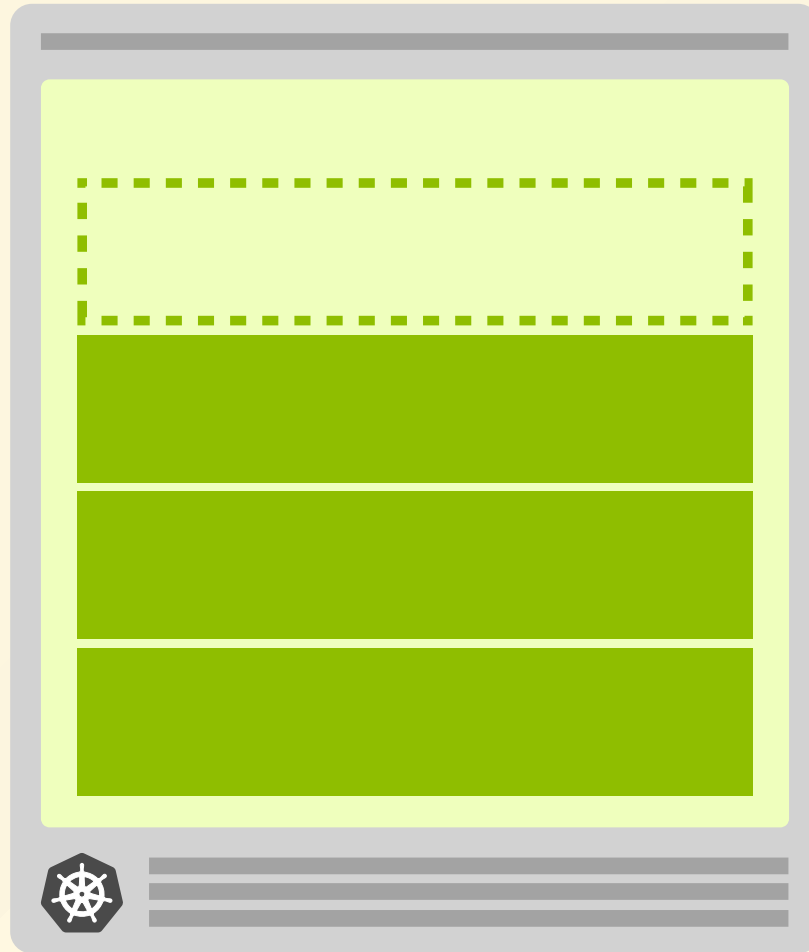


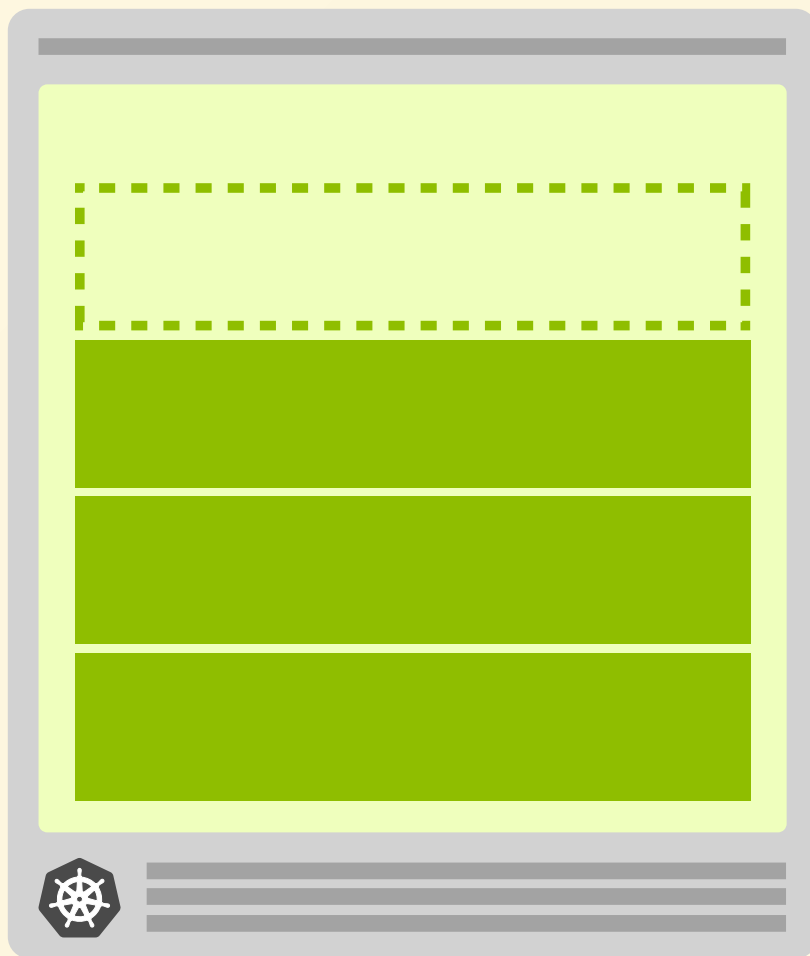
**1.5GB memory  
0.25 vCPU**



**8GB memory  
2 vCPU**

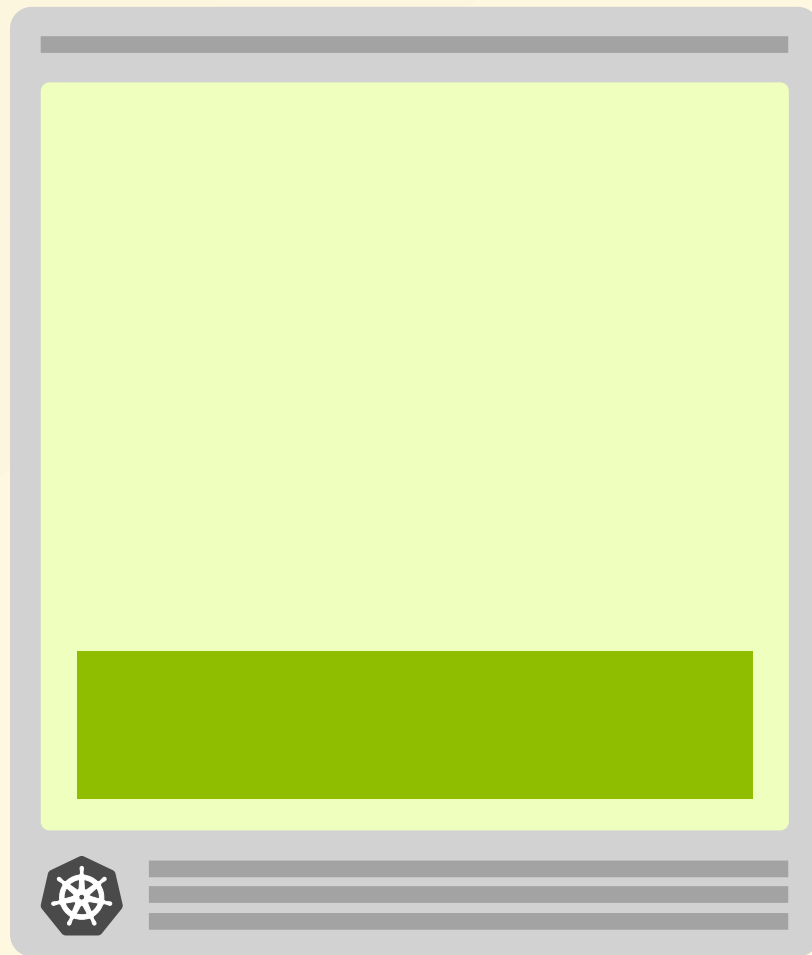
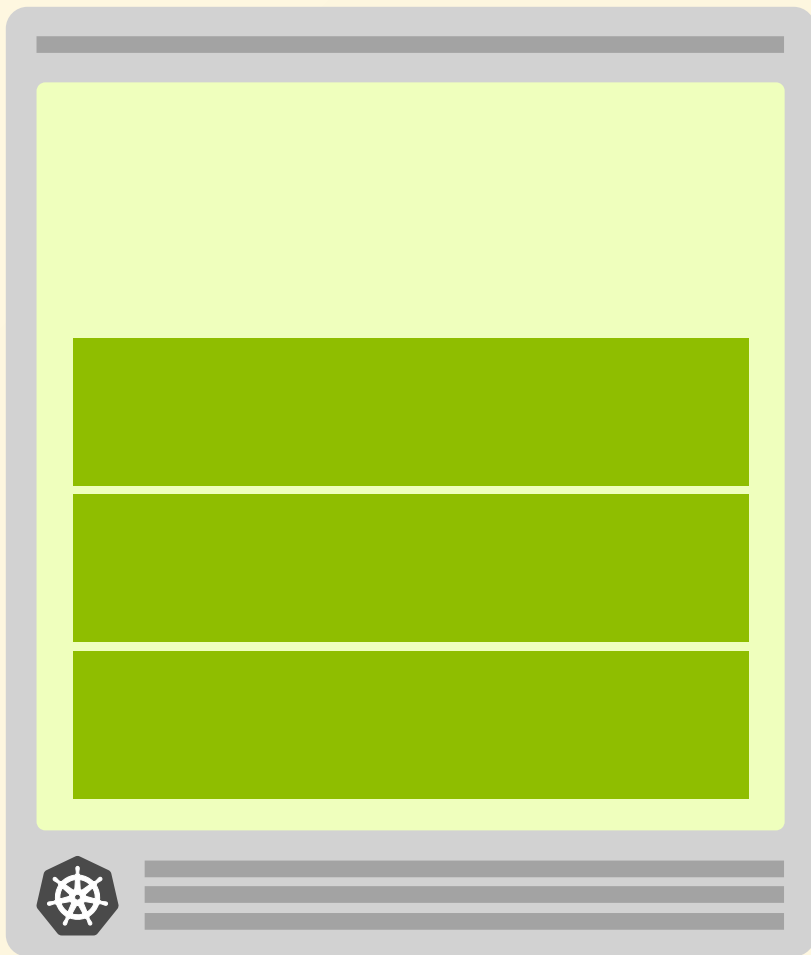






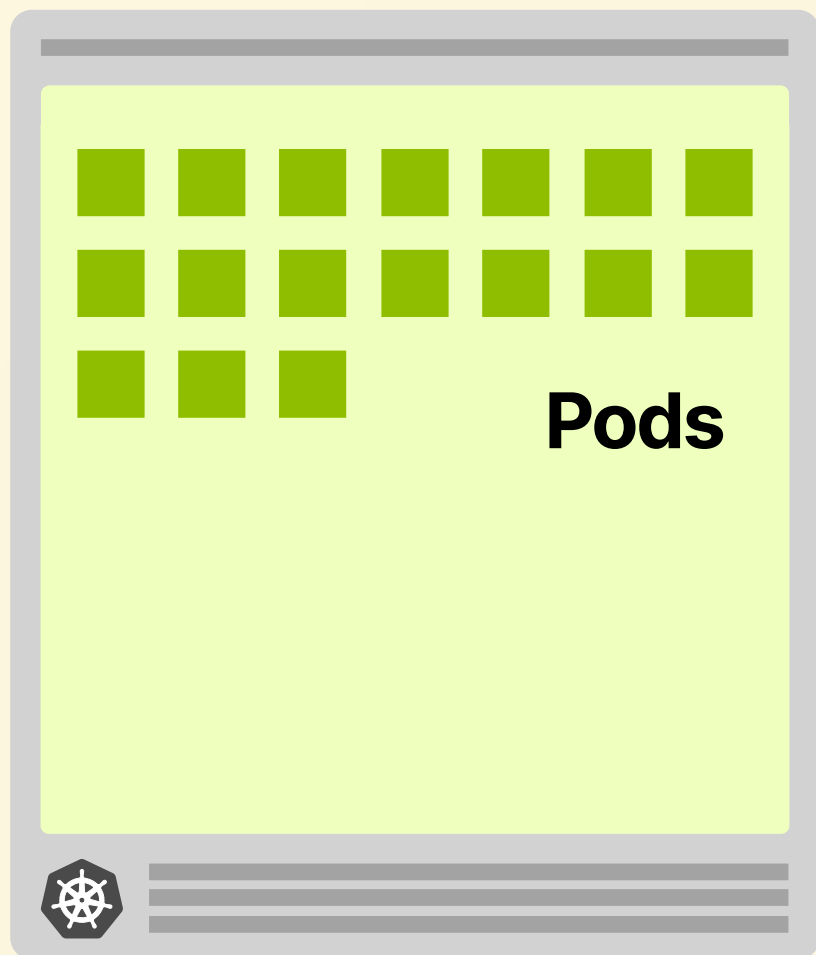
**Pending**

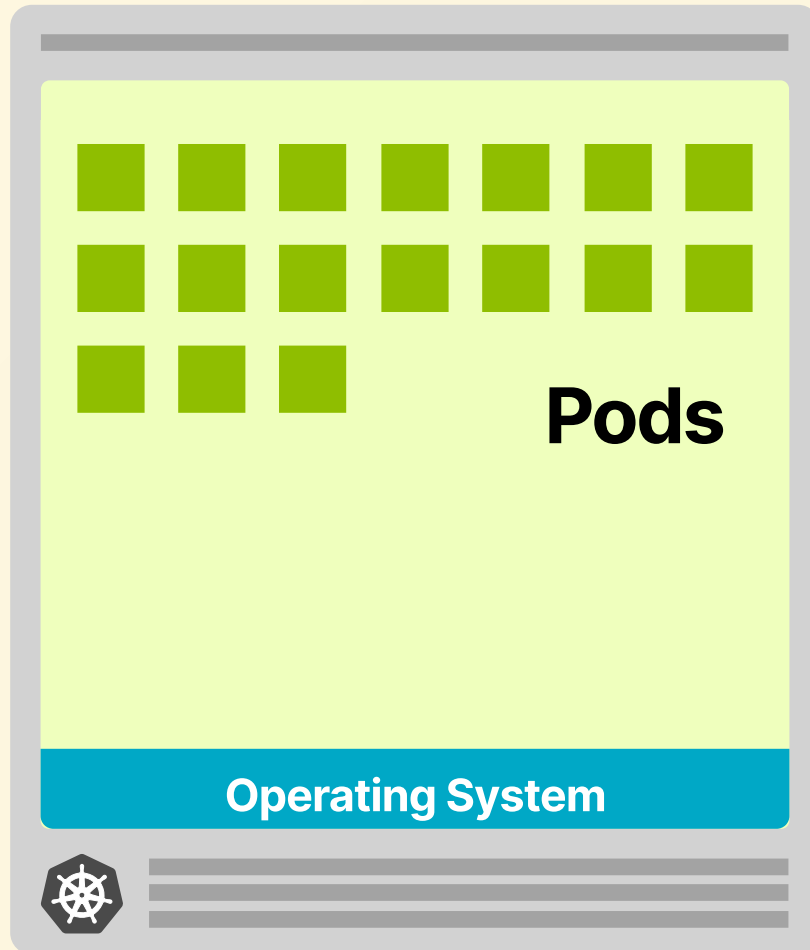




# Node instances



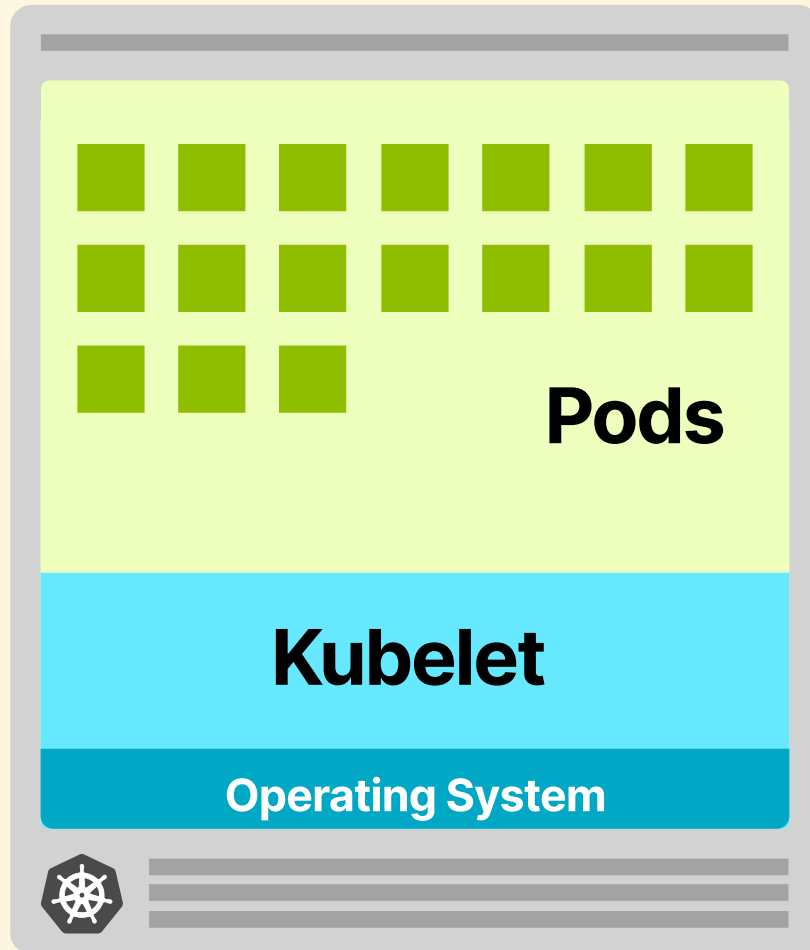




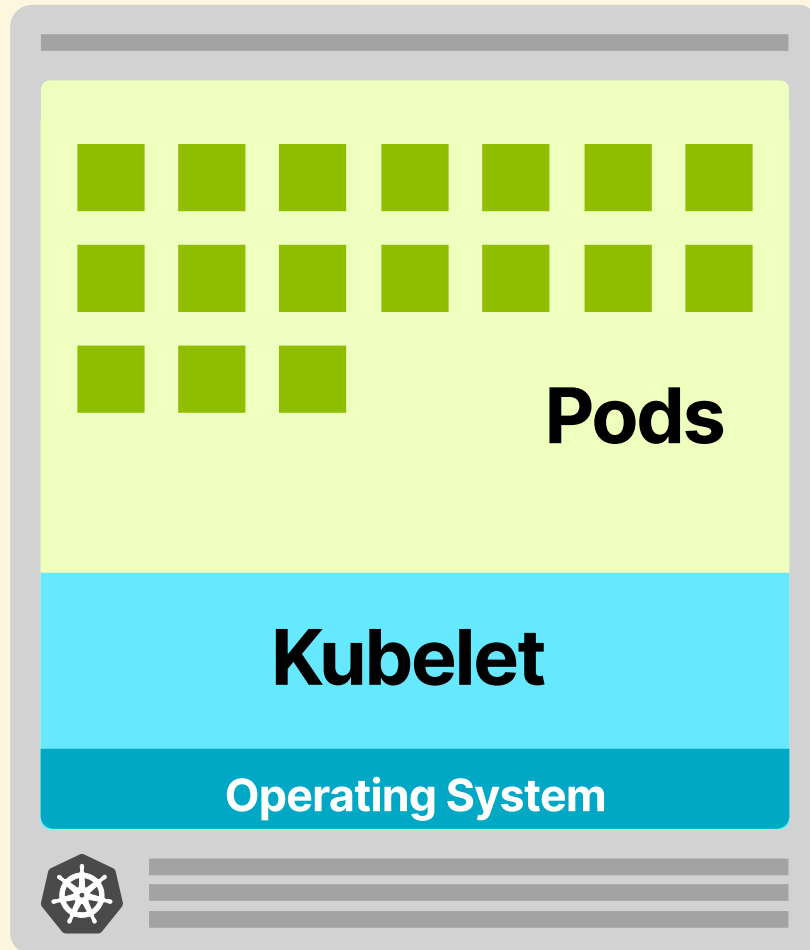
**4**

**Memory and CPU  
reserved to the OS**





- 3 Memory and CPU reserved to the kubelet
- 4 Memory and CPU reserved to the OS



1

2

**Memory and CPU  
left to Pods**

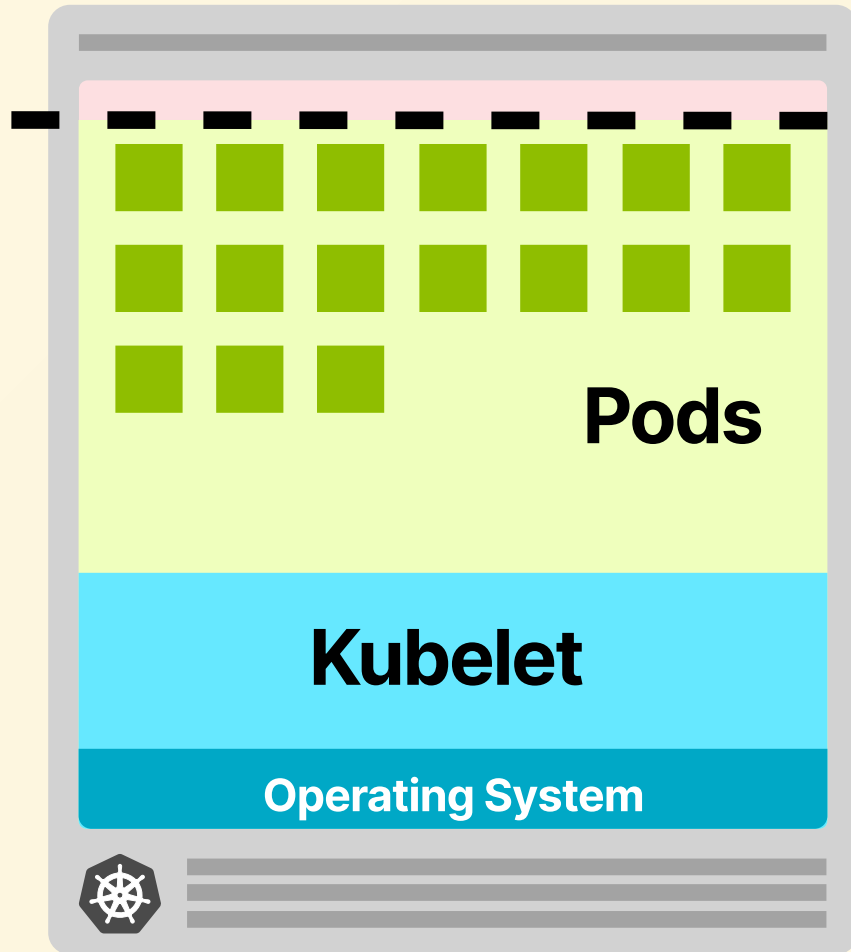
3

**Memory and CPU  
reserved to the  
kubelet**

4

**Memory and CPU  
reserved to the OS**





**1** Eviction threshold

**2** Memory and CPU left to Pods

**3** Memory and CPU reserved to the kubelet

**4** Memory and CPU reserved to the OS



**Kubelet reserved memory**

**+ 255MiB**



## **Kubelet reserved memory**

---

**+ 255MiB**

**+ 11MiB \* MAX\_PODS**

---



## **Kubelet reserved memory**

**+ 255MiB**

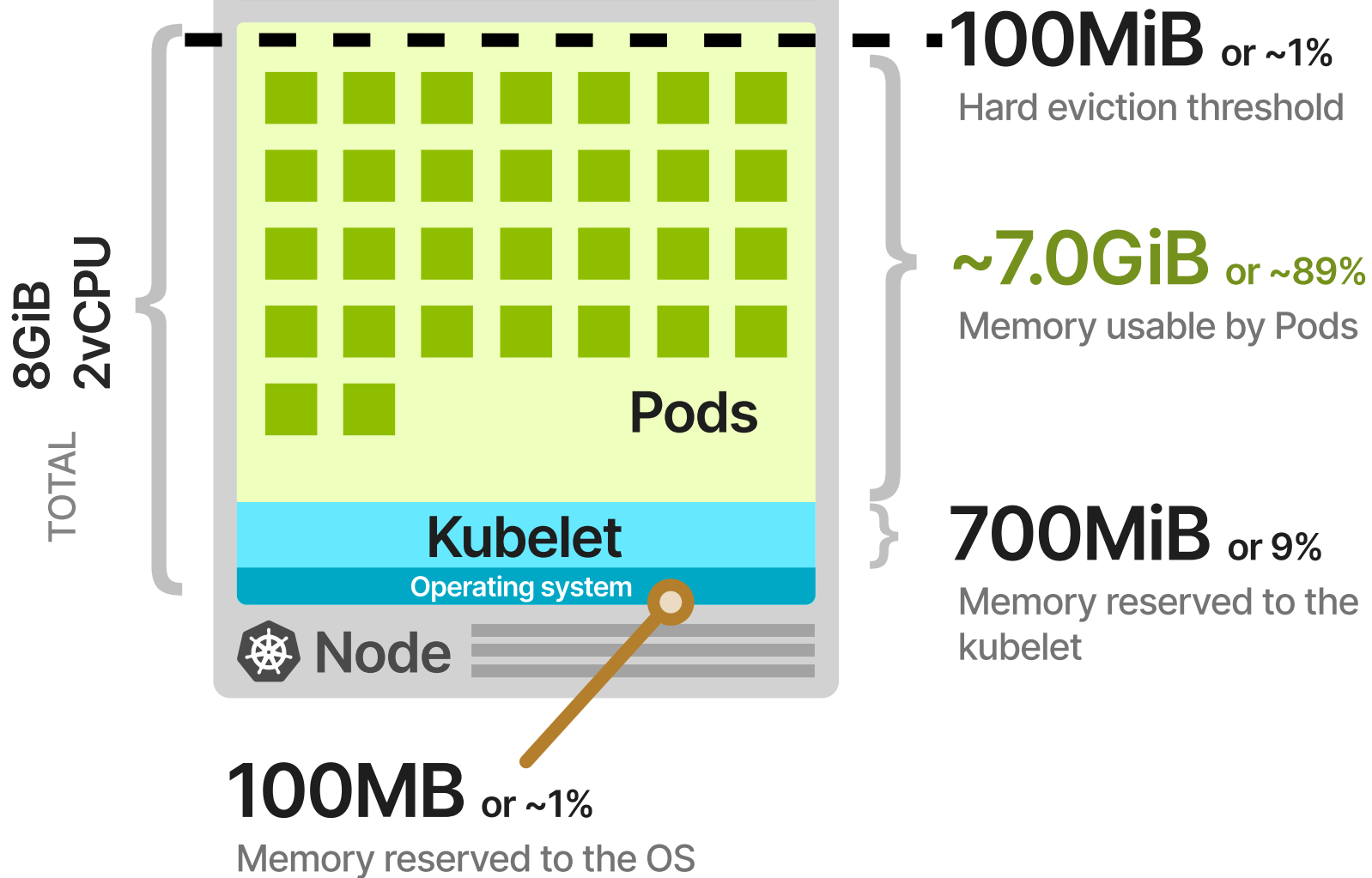
**+ 11MiB \* MAX\_PODS**

---

**= Reserved memory**



m5.large

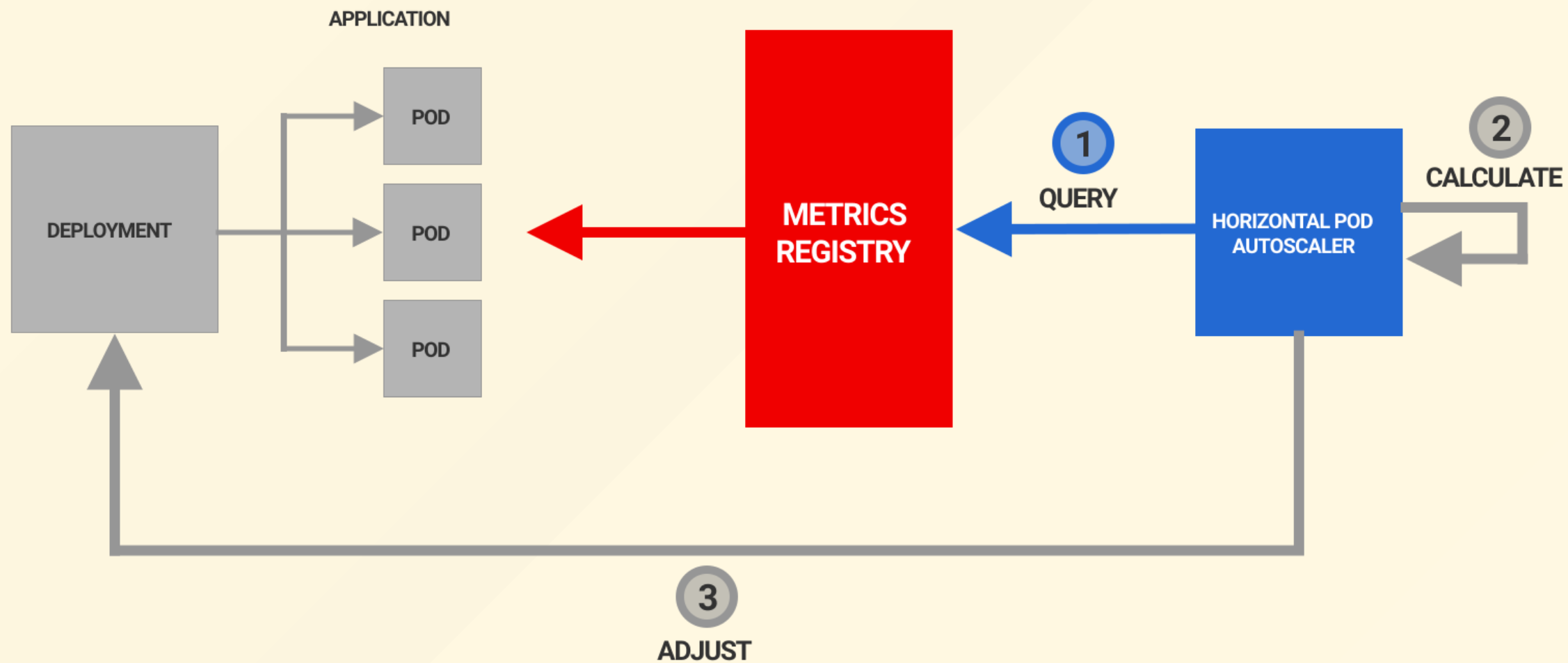


# Cluster AutoScaler

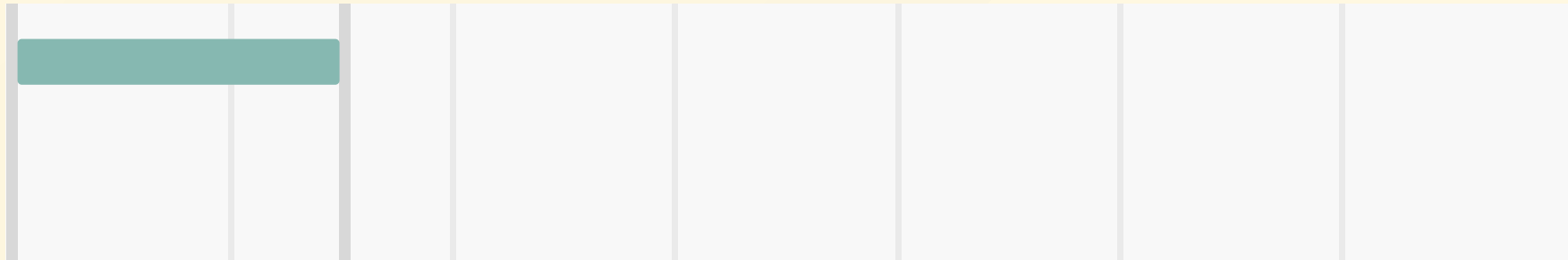
## Lead Time



## CLUSTER

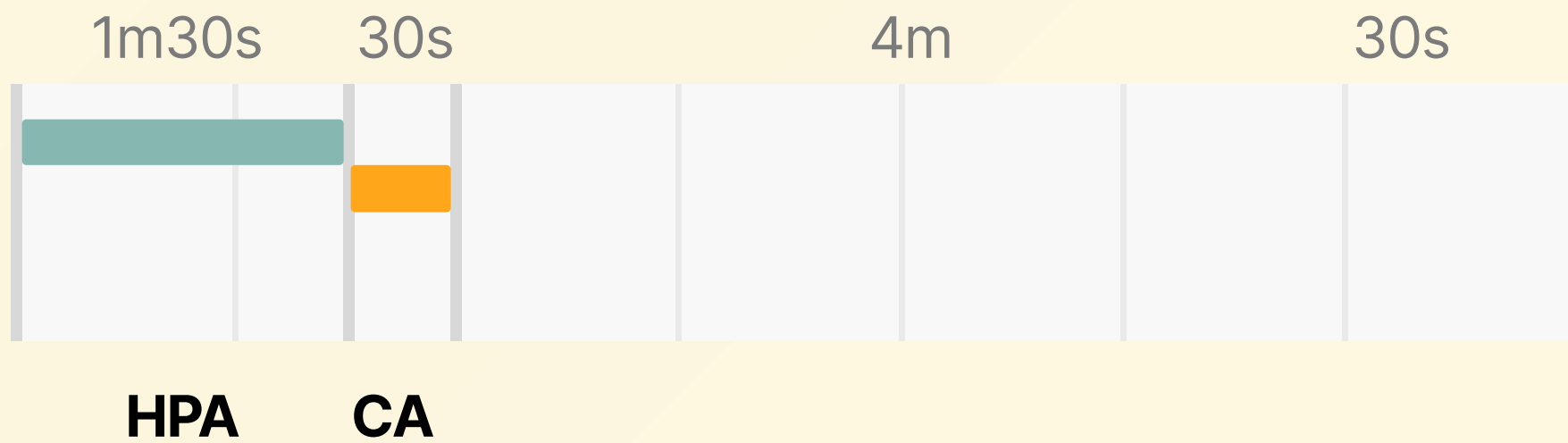


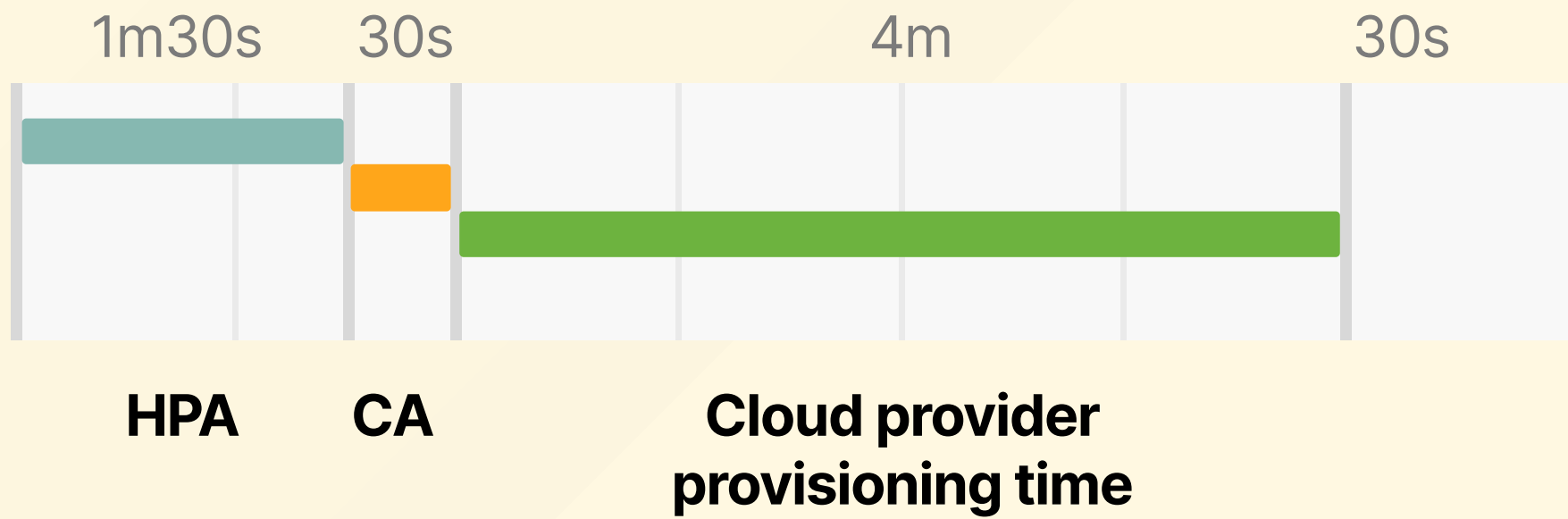
1m30s

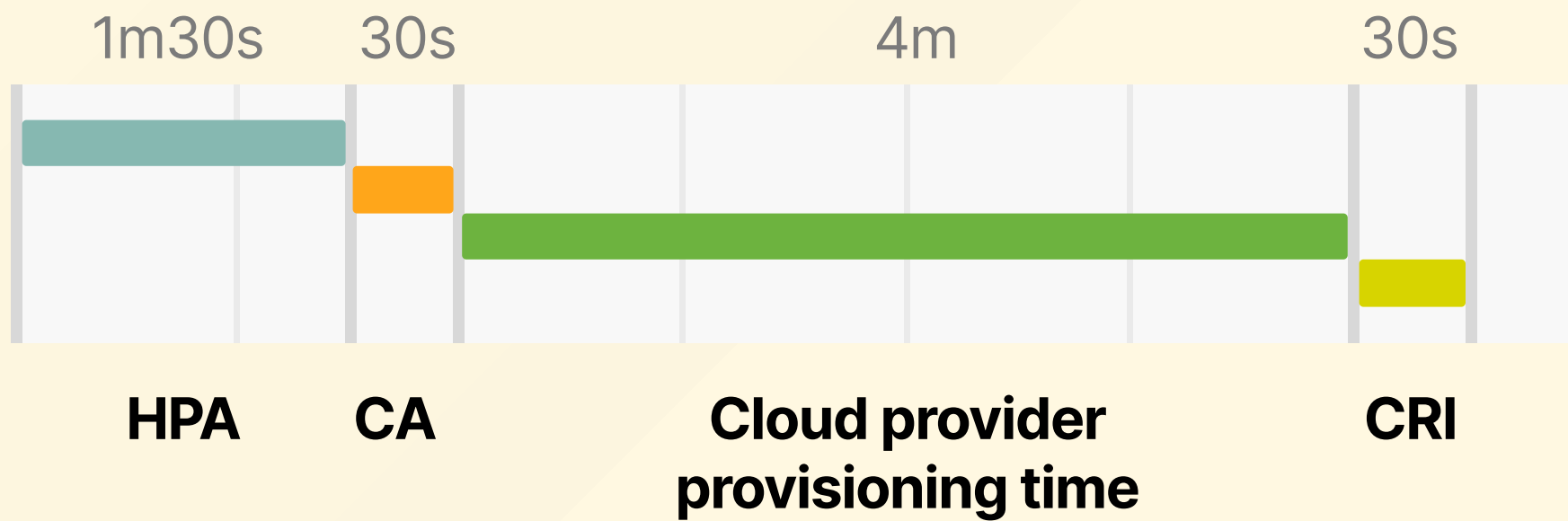


**HPA**









# github.com/appvia/cloud-spend-forecaster





# Strategies for faster scaling



# Strategies for faster scaling

**1. Don't scale!**

2. (Pre) scale



# Strategies for faster scaling

1. Don't scale!

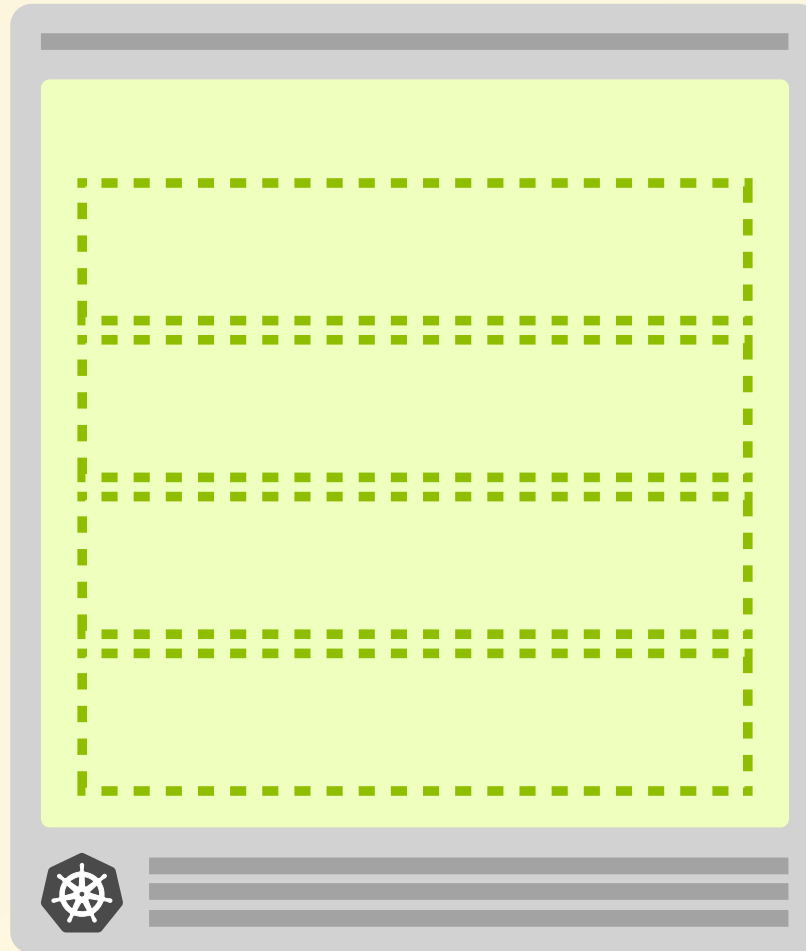
2. **(Pre) scale**



# Don't Scale!



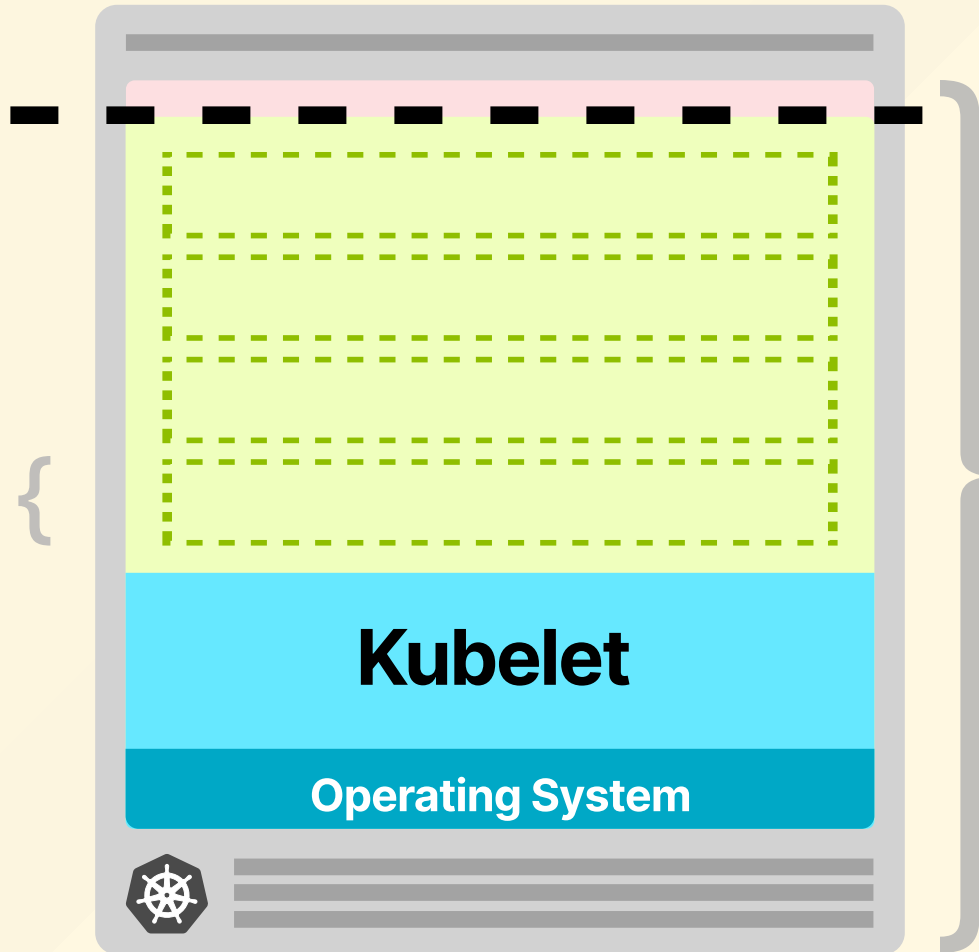
**1.5GB memory  
0.25 vCPU**



**8GB memory  
2 vCPU**



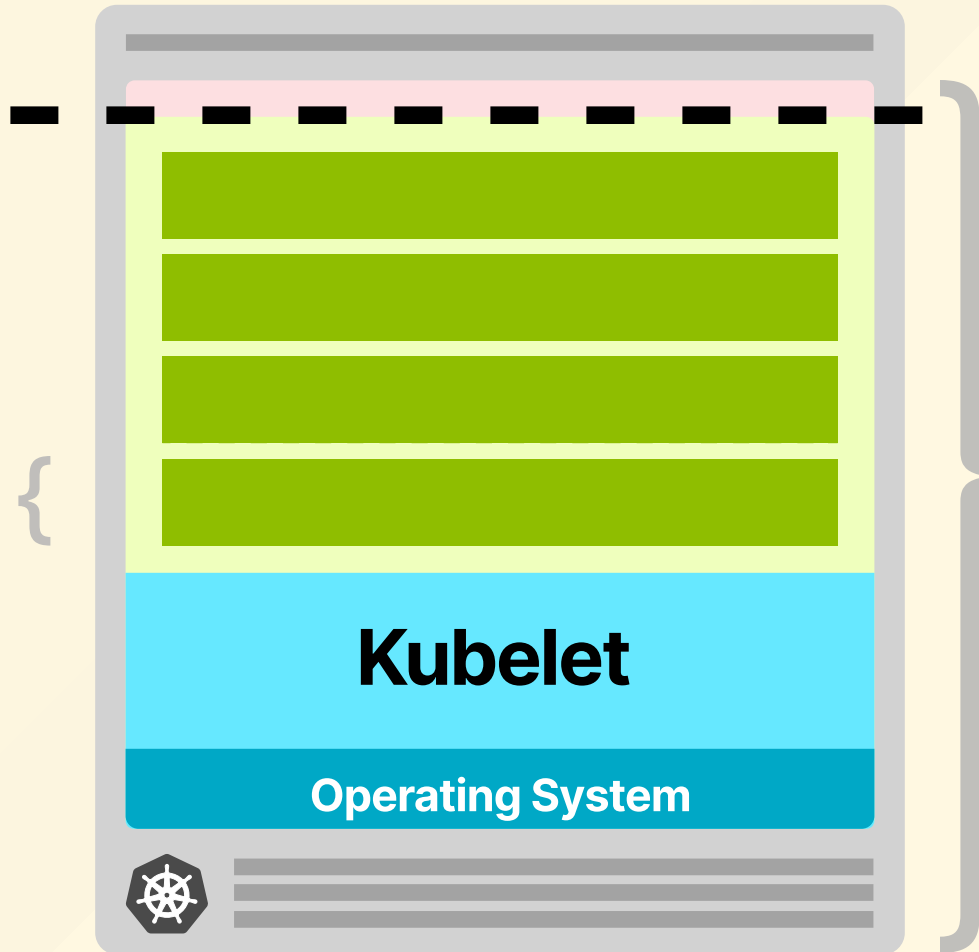
**1.2GB memory**  
**0.2 vCPU**



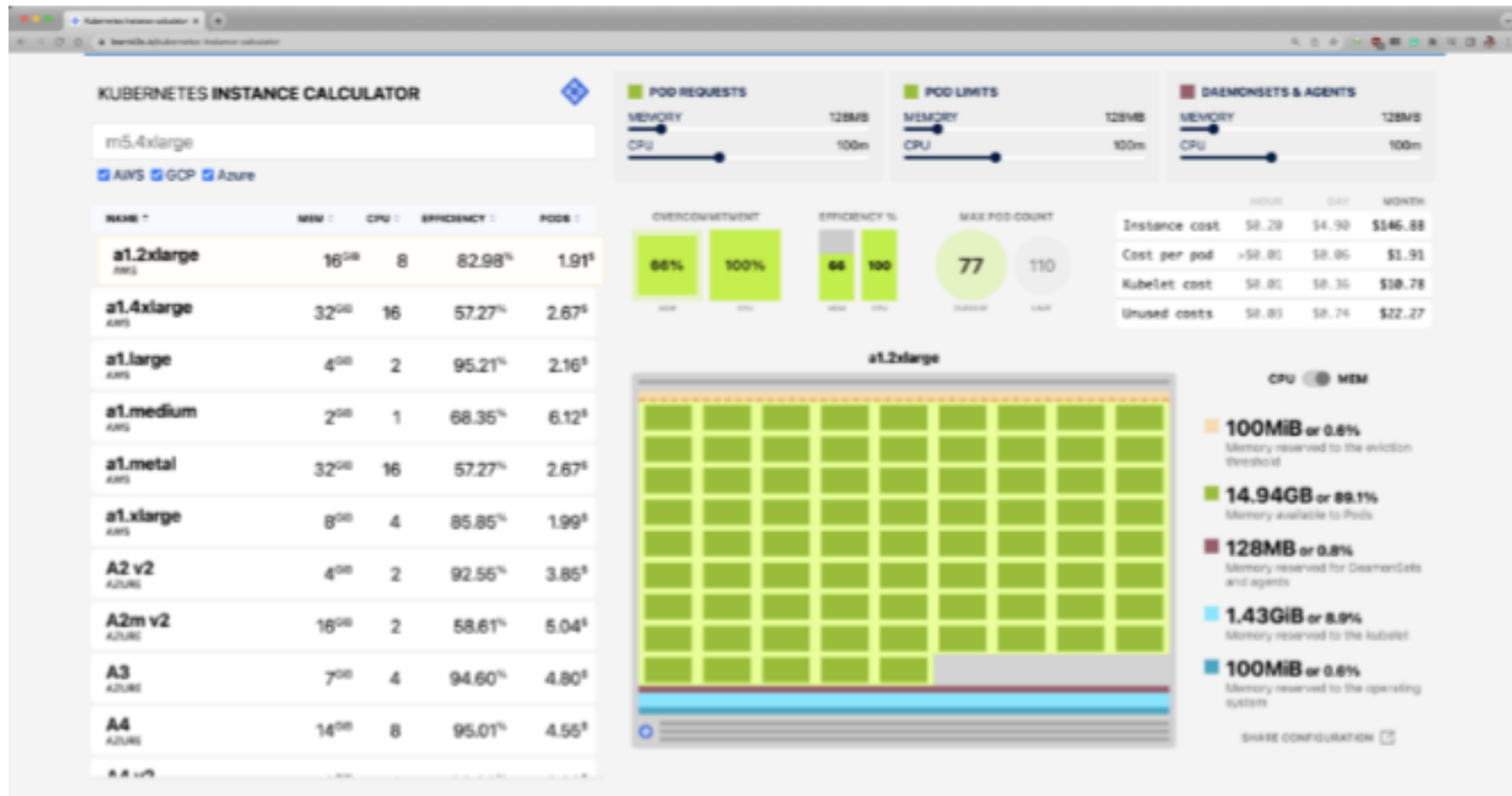
**8GB memory**  
**2 vCPU**



**1.2GB memory  
0.2 vCPU**



**8GB memory  
2 vCPU**



[learnk8s.io/kubernetes-instance-calculator](https://learnk8s.io/kubernetes-instance-calculator)



CONTACT US

|                                |    |                                  |     |
|--------------------------------|----|----------------------------------|-----|
| <input type="text" value="1"/> | GB | <input type="text" value="0.1"/> | CPU |
| <input type="text" value="1"/> | GB | <input type="text" value="0.1"/> | CPU |

|                                |    |                                |     |
|--------------------------------|----|--------------------------------|-----|
| <input type="text" value="0"/> | GB | <input type="text" value="0"/> | CPU |
|--------------------------------|----|--------------------------------|-----|

Guide

KUBERNETES  
INSTANCE CALCULATOR<sup>V2</sup>

POD ⓘ

Requests  
Limits

DAEMONSETS & AGENTS ⓘ

Reserved

CLOUD PROVIDER

AWS



☐ Prefix delegation ⓘ

INSTANCE TYPES

☐ ACCELERATOR OPTIMIZED



☒ COMPUTE OPTIMIZED



☒ GENERAL PURPOSE



☐ GPU OPTIMIZED



☐ HIGH PERFORMANCE



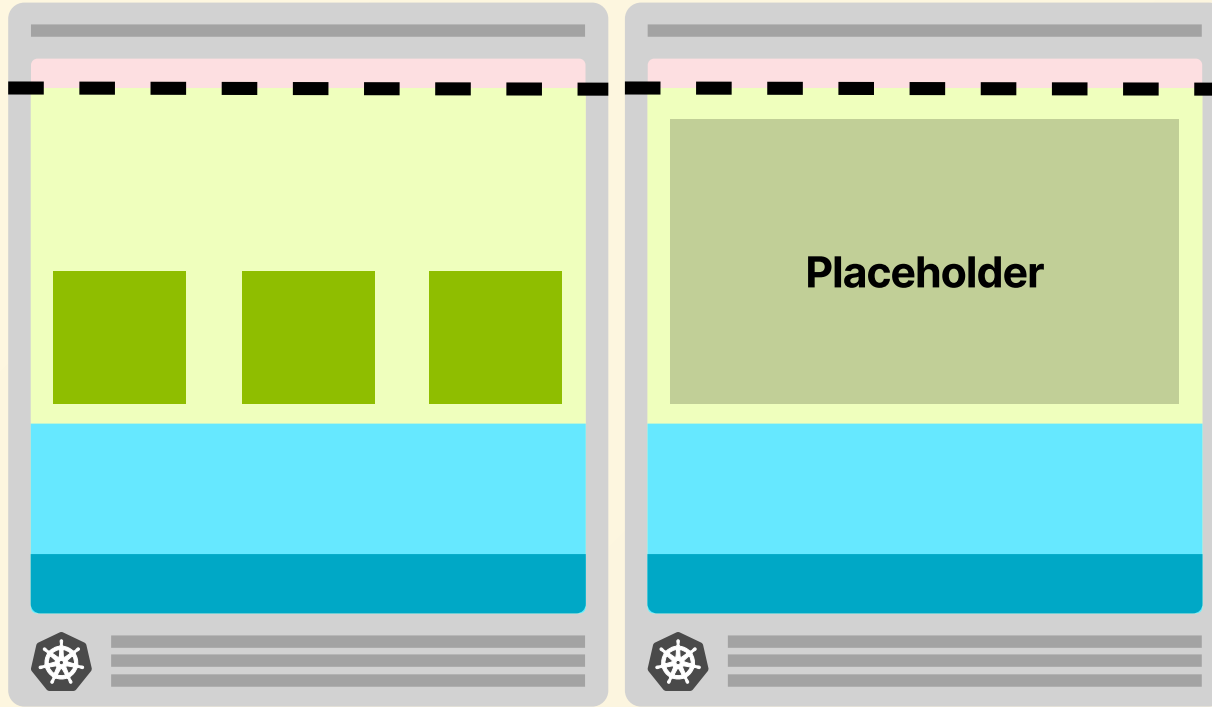
☒ MEMORY OPTIMIZED

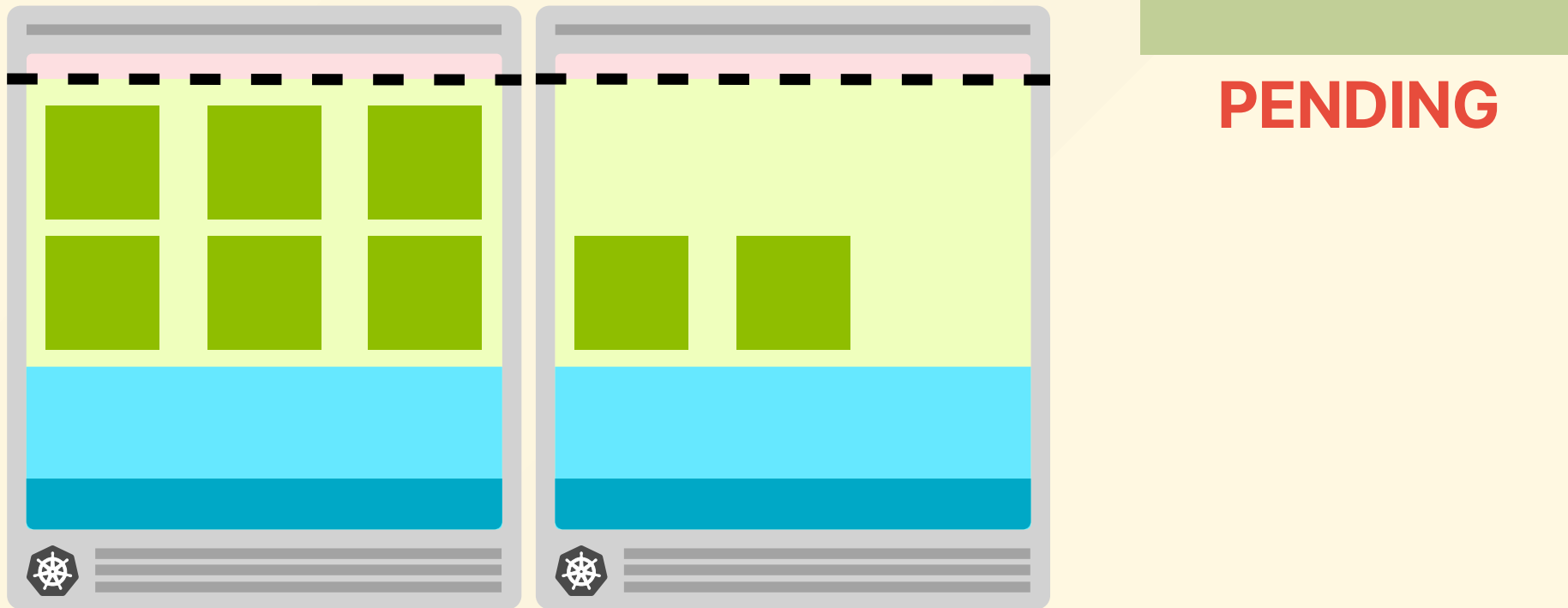


SORT INSTANCES BY

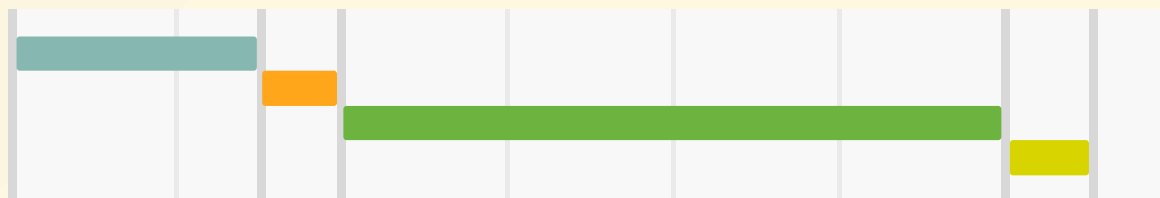
# Proactive scaling

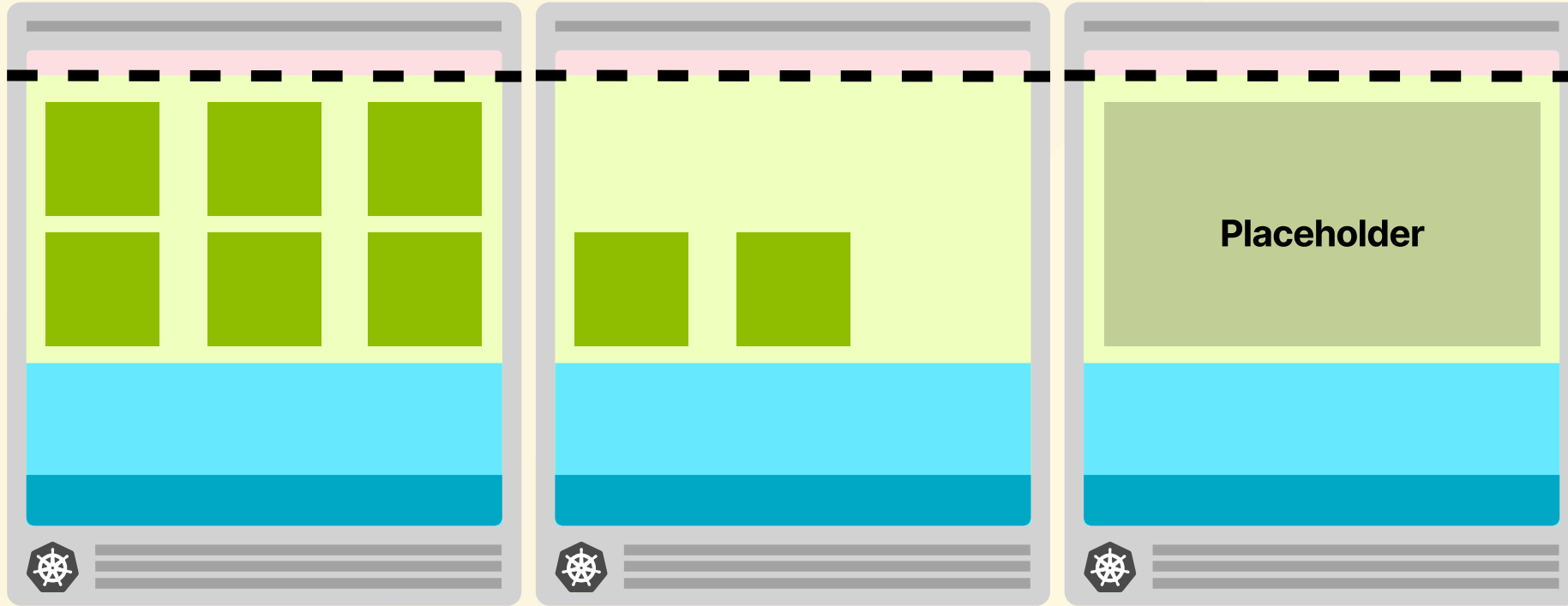






**Provisioning a node in the background**







# LIVE DEMO





|             |    |         |                         |                                   |  |
|-------------|----|---------|-------------------------|-----------------------------------|--|
| default     | 0s | Normal  | ScalingReplicaSet       | deployment/podinfo                | Scaled up replica set podinfo-8558cfdc5d to 5  |
| default     | 0s | Normal  | Scheduled               | pod/podinfo-8558cfdc5d-h4c5j      | Successfully assigned default/podinfo-8558cfdc5d-h4c5j to lke74013-115226-6331fbd57928   |
| default     | 0s | Normal  | SuccessfulCreate        | replicaset/podinfo-8558cfdc5d     | Created pod: podinfo-8558cfdc5d-h4c5j  |
| default     | 0s | Normal  | Scheduled               | pod/podinfo-8558cfdc5d-z9hvq      | Successfully assigned default/podinfo-8558cfdc5d-z9hvq to lke74013-115226-6331fbd57928   |
| default     | 0s | Normal  | Scheduled               | pod/podinfo-8558cfdc5d-d5khs      | Successfully assigned default/podinfo-8558cfdc5d-d5khs to lke74013-115226-6331fbd57928   |
| default     | 0s | Normal  | SuccessfulCreate        | replicaset/podinfo-8558cfdc5d     | Created pod: podinfo-8558cfdc5d-z9hvq  |
| default     | 0s | Normal  | SuccessfulCreate        | replicaset/podinfo-8558cfdc5d     | Created pod: podinfo-8558cfdc5d-d5khs  |
| default     | 0s | Warning | FailedScheduling        | pod/podinfo-8558cfdc5d-5x2v7      | 0/2 nodes are available: 1 Insufficient memory, 1 node(s) didn't match Pod's node affinity/selector.   |
| default     | 0s | Normal  | Pulling                 | pod/podinfo-8558cfdc5d-d5khs      | Pulling image "stefanprodan/podinfo"   |
| default     | 0s | Normal  | Pulling                 | pod/podinfo-8558cfdc5d-h4c5j      | Pulling image "stefanprodan/podinfo"   |
| default     | 0s | Normal  | Pulling                 | pod/podinfo-8558cfdc5d-z9hvq      | Pulling image "stefanprodan/podinfo"   |
| default     | 0s | Normal  | Pulled                  | pod/podinfo-8558cfdc5d-d5khs      | Successfully pulled image "stefanprodan/podinfo" in 1.708158076s   |
| default     | 0s | Normal  | Created                 | pod/podinfo-8558cfdc5d-d5khs      | Created container podinfo  |
| default     | 0s | Normal  | Started                 | pod/podinfo-8558cfdc5d-d5khs      | Started container podinfo  |
| default     | 0s | Normal  | Pulled                  | pod/podinfo-8558cfdc5d-h4c5j      | Successfully pulled image "stefanprodan/podinfo" in 3.104773992s   |
| default     | 0s | Normal  | Created                 | pod/podinfo-8558cfdc5d-h4c5j      | Created container podinfo  |
| default     | 0s | Normal  | Started                 | pod/podinfo-8558cfdc5d-h4c5j      | Started container podinfo  |
| default     | 0s | Normal  | NotTriggerScaleUp       | pod/podinfo-8558cfdc5d-5x2v7      | pod didn't trigger scale-up: 1 node(s) didn't match Pod's node affinity/selector   |
| default     | 0s | Normal  | Pulled                  | pod/podinfo-8558cfdc5d-z9hvq      | Successfully pulled image "stefanprodan/podinfo" in 8.442357379s   |
| default     | 0s | Normal  | Created                 | pod/podinfo-8558cfdc5d-z9hvq      | Created container podinfo  |
| default     | 0s | Normal  | Started                 | pod/podinfo-8558cfdc5d-z9hvq      | Started container podinfo  |
| default     | 0s | Warning | FailedScheduling        | pod/podinfo-8558cfdc5d-5x2v7      | 0/2 nodes are available: 1 Insufficient memory, 1 node(s) didn't match Pod's node affinity/selector.   |
| default     | 1s | Normal  | NodeHasSufficientMemory | node/lke74013-115226-6332129f84dd | Node lke74013-115226-6332129f84dd status is now: NodeHasSufficientMemory   |
| default     | 1s | Normal  | NodeHasNoDiskPressure   | node/lke74013-115226-6332129f84dd | Node lke74013-115226-6332129f84dd status is now: NodeHasNoDiskPressure   |
| kube-system | 0s | Normal  | SuccessfulCreate        | daemonset/calico-node             | Created pod: calico-node-tpxpn   |
| kube-system | 0s | Normal  | Scheduled               | pod/csi-linode-node-fvk46         | Successfully assigned kube-system/csi-linode-node-fvk46 to lke74013-115226-6332129f84dd  |
| kube-system | 0s | Normal  | Scheduled               | pod/calico-node-tpxpn             | Successfully assigned kube-system/calico-node-tpxpn to lke74013-115226-6332129f84dd  |
| kube-system | 0s | Normal  | SuccessfulCreate        | daemonset/csi-linode-node         | Created pod: csi-linode-node-fvk46   |
| kube-system | 0s | Normal  | Scheduled               | pod/kube-proxy-ds24f              | Successfully assigned kube-system/kube-proxy-ds24f to lke74013-115226-6332129f84dd   |
| kube-system | 0s | Normal  | SuccessfulCreate        | daemonset/kube-proxy              | Created pod: kube-proxy-ds24f  |
| default     | 0s | Normal  | Synced                  | node/lke74013-115226-6332129f84dd | Node synced successfully   |
| default     | 0s | Normal  | Starting                | node/lke74013-115226-6332129f84dd | Starting kubelet.  |
| default     | 0s | Normal  | NodeAllocatableEnforced | node/lke74013-115226-6332129f84dd | Updated Node Allocatable limit across pods   |
| default     | 0s | Normal  | NodeHasSufficientMemory | node/lke74013-115226-6332129f84dd | Node lke74013-115226-6332129f84dd status is now: NodeHasSufficientMemory   |
| default     | 0s | Normal  | NodeHasNoDiskPressure   | node/lke74013-115226-6332129f84dd | Node lke74013-115226-6332129f84dd status is now: NodeHasNoDiskPressure   |
| default     | 0s | Normal  | NodeHasSufficientPID    | node/lke74013-115226-6332129f84dd | Node lke74013-115226-6332129f84dd status is now: NodeHasSufficientPID  |
| default     | 0s | Normal  | RegisteredNode          | node/lke74013-115226-6332129f84dd | Node lke74013-115226-6332129f84dd event: Registered Node lke74013-115226-6332129f84dd in Controller  |
| kube-system | 1s | Normal  | Pulling                 | pod/kube-proxy-ds24f              | Pulling image "linode/kube-proxy-amd64:v1.23.10"   |
| kube-system | 1s | Normal  | Pulling                 | pod/calico-node-tpxpn             | Pulling image "docker.io/calico/cni:v3.22.1"   |
| kube-system | 0s | Normal  | Pulling                 | pod/csi-linode-node-fvk46         | Pulling image "bitnami/kubectrl:1.16.3-debian-10-r36"  |
| kube-system | 0s | Normal  | Pulled                  | pod/kube-proxy-ds24f              | Successfully pulled image "linode/kube-proxy-amd64:v1.23.10" in 3.544455967s   |
| kube-system | 0s | Normal  | Created                 | pod/kube-proxy-ds24f              | Created container kube-proxy   |
| kube-system | 0s | Normal  | Started                 | pod/kube-proxy-ds24f              | Started container kube-proxy   |
| default     | 0s | Normal  | Starting                | node/lke74013-115226-6332129f84dd |  |
| kube-system | 0s | Normal  | Pulled                  | pod/calico-node-tpxpn             | Successfully pulled image "docker.io/calico/cni:v3.22.1" in 8.06147771s  |
| kube-system | 1s | Normal  | Created                 | pod/calico-node-tpxpn             | Created container upgrade-ipam   |
| kube-system | 0s | Normal  | Started                 | pod/calico-node-tpxpn             | Started container upgrade-ipam   |
| kube-system | 0s | Normal  | Pulled                  | pod/calico-node-tpxpn             | Container image "docker.io/calico/cni:v3.22.1" already present on machine  |
| kube-system | 0s | Normal  | Created                 | pod/calico-node-tpxpn             | Created container install-cni  |
| kube-system | 0s | Normal  | Started                 | pod/calico-node-tpxpn             | Started container install-cni  |
| kube-system | 0s | Normal  | Pulling                 | pod/calico-node-tpxpn             | Pulling image "docker.io/calico/pod2daemon-flexvol:v3.22.1"  |
| kube-system | 0s | Normal  | Pulled                  | pod/csi-linode-node-fvk46         | Successfully pulled image "bitnami/kubectrl:1.16.3-debian-10-r36" in 11.657565568s   |
| kube-system | 0s | Normal  | Created                 | pod/csi-linode-node-fvk46         | Created container init   |
| kube-system | 0s | Normal  | Started                 | pod/csi-linode-node-fvk46         | Started container init   |
| kube-system | 0s | Normal  | Pulling                 | pod/csi-linode-node-fvk46         | Pulling image "linode/csi-node-driver-registrar:v1.3.0"  |
| default     | 0s | Warning | FailedScheduling        | pod/podinfo-8558cfdc5d-5x2v7      | 0/2 nodes are available: 1 Insufficient memory, 1 node(s) didn't match Pod's node affinity/selector.   |
| kube-system | 0s | Normal  | Pulled                  | pod/calico-node-tpxpn             | Successfully pulled image "docker.io/calico/pod2daemon-flexvol:v3.22.1" in 5.534993769s  |
| kube-system | 0s | Normal  | Created                 | pod/calico-node-tpxpn             | Created container flexvol-driver   |
| kube-system | 0s | Normal  | Started                 | pod/calico-node-tpxpn             | Started container flexvol-driver   |
| kube-system | 0s | Normal  | Pulling                 | pod/calico-node-tpxpn             | Pulling image "docker.io/calico/node:v3.22.1"  |
| default     | 0s | Normal  | NodeReady               | node/lke74013-115226-6332129f84dd | Node lke74013-115226-6332129f84dd status is now: NodeReady   |
| kube-system | 0s | Normal  | Pulled                  | pod/csi-linode-node-fvk46         | Successfully pulled image "linode/csi-node-driver-registrar:v1.3.0" in 4.741448054s  |
| kube-system | 0s | Normal  | Created                 | pod/csi-linode-node-fvk46         | Created container csi-node-driver-registrar  |
| kube-system | 0s | Normal  | Started                 | pod/csi-linode-node-fvk46         | Started container csi-node-driver-registrar  |
| kube-system | 0s | Normal  | Pulling                 | pod/csi-linode-node-fvk46         | Pulling image "linode/linode-blockstorage-csi-driver:v0.5.0"   |
| default     | 0s | Normal  | Scheduled               | pod/podinfo-8558cfdc5d-5x2v7      | Successfully assigned default/podinfo-8558cfdc5d-5x2v7 to lke74013-115226-6332129f84dd   |
| default     | 0s | Warning | FailedCreatePodSandbox  | pod/podinfo-8558cfdc5d-5x2v7      | Failed to create pod sandbox: rpc error: code = Unknown desc = failed to set up sandbox container "4d21d0493b5ac49eb7a173b9946f379f9e0d98b1d55ab13547ad08264284bc6d" network for pod "podinfo-8558cfdc5d-5x2v7": networkPlugin cni failed to set up pod "podinfo-8558cfdc5d-5x2v7" |
| default     | 0s | Normal  | TaintManagerEviction    | pod/podinfo-8558cfdc5d-5x2v7      | Cancelling deletion of Pod default/podinfo-8558cfdc5d-5x2v7  |
| default     | 0s | Normal  | SandboxChanged          | pod/podinfo-8558cfdc5d-5x2v7      | Pod sandbox changed, it will be killed and re-created.   |
| default     | 0s | Warning | FailedCreatePodSandbox  | pod/podinfo-8558cfdc5d-5x2v7      | Failed to create pod sandbox: rpc error: code = Unknown desc = failed to set up sandbox container "5d02b7c46e449134d4abb9e02dc77044a3163f97c0a46f47c398108ac3fa000" network for pod "podinfo-8558cfdc5d-5x2v7": networkPlugin cni failed to set up pod "podinfo-8558cfdc5d-5x2v7"  |
| default     | 0s | Normal  | SandboxChanged          | pod/podinfo-8558cfdc5d-5x2v7      | Pod sandbox changed, it will be killed and re-created.   |
| default     | 0s | Warning | FailedCreatePodSandbox  | pod/podinfo-8558cfdc5d-5x2v7      | Failed to create pod sandbox: rpc error: code = Unknown desc = failed to set up sandbox container "3447a34a5eda5501828a9c476e0d6d7077b7f895c8fbc7022af1643865241c6" network for pod "podinfo-8558cfdc5d-5x2v7": networkPlugin cni failed to set up pod "podinfo-8558cfdc5d-5x2v7"  |
| default     | 0s | Normal  | SandboxChanged          | pod/podinfo-8558cfdc5d-5x2v7      | Pod sandbox changed, it will be killed and re-created.   |
| default     | 0s | Warning | FailedCreatePodSandbox  | pod/podinfo-8558cfdc5d-5x2v7      | Failed to create pod sandbox: rpc error: code = Unknown desc = failed to set up sandbox container "77a8281777afade238f38dc9141ed2e52c78540ccb4ad7b296da3fc8c20a6ef" network for pod "podinfo-8558cfdc5d-5x2v7": networkPlugin cni failed to set up pod "podinfo-8558cfdc5d-5x2v7"  |
| default     | 0s | Normal  | SandboxChanged          | pod/podinfo-8558cfdc5d-5x2v7      | Pod sandbox changed, it will be killed and re-created.   |
| kube-system | 0s | Normal  | Pulled                  | pod/calico-node-tpxpn             | Successfully pulled image "docker.io/calico/node:v3.22.1" in 6.353089879s  |
| kube-system | 0s | Normal  | Created                 | pod/calico-node-tpxpn             | Created container calico-node  |
| kube-system | 0s | Normal  | Started                 | pod/calico-node-tpxpn             | Started container calico-node  |
| default     | 0s | Warning | FailedCreatePodSandbox  | pod/podinfo-8558cfdc5d-5x2v7      | Failed to create pod sandbox: rpc error: code = Unknown desc = failed to set up sandbox container "0fc48fe077b786da531ebc656016468d3ae3f41e6cf520e27b638dbfbab580e2" network for pod "podinfo-8558cfdc5d-5x2v7": networkPlugin cni failed to set up pod "podinfo-8558cfdc5d-5x2v7" |
| default     | 0s | Normal  | SandboxChanged          | pod/podinfo-8558cfdc5d-5x2v7      | Pod sandbox changed, it will be killed and re-created.   |
| default     | 0s | Warning | Unhealthy               | pod/calico-node-tpxpn             | Readiness probe failed: calico/node is not ready: BIRD is not ready: Error querying BIRD: unable to connect to BIRDv4 socket: dial unix /var/run/bird/birdctl: connect: no such file or directory  |
| default     | 0s | Normal  | Pulling                 | pod/podinfo-8558cfdc5d-5x2v7      | Pulling image "stefanprodan/podinfo"   |
| kube-system | 0s | Warning | Unhealthy               | pod/calico-node-tpxpn             | Readiness probe failed: calico/node is not ready: felix is not ready: readiness probe reporting 503  |
| kube-system | 0s | Normal  | Pulled                  | pod/csi-linode-node-fvk46         | Successfully pulled image "linode/linode-blockstorage-csi-driver:v0.5.0" in 6.754070918s   |
| kube-system | 0s | Normal  | Created                 | pod/csi-linode-node-fvk46         | Created container csi-linode-plugin  |
| kube-system | 0s | Normal  | Started                 | pod/csi-linode-node-fvk46         | Started container csi-linode-plugin  |
| kube-system | 0s | Warning | Unhealthy               | pod/calico-node-tpxpn             | Readiness probe failed: calico/node is not ready: BIRD is not ready: Error querying BIRD: unable to connect to BIRDv4 socket: dial unix /var/run/calico/birdctl: connect: connection refused   |

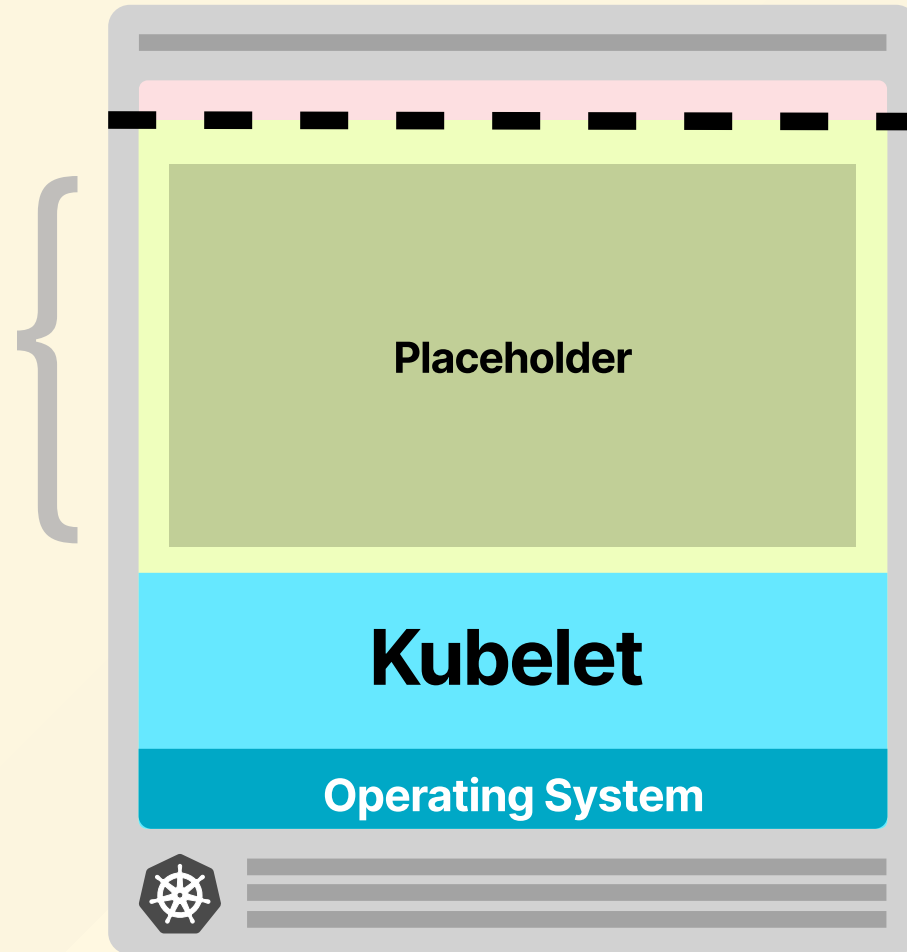




# Designing the right placeholder



LESS THAN  
**5.9GB memory**  
**1.73 vCPU**



**8GB memory**  
**2 vCPU**



```
apiVersion: scheduling.k8s.io/v1
kind: PriorityClass
metadata:
  name: overprovisioning
value: -1
globalDefault: false
```





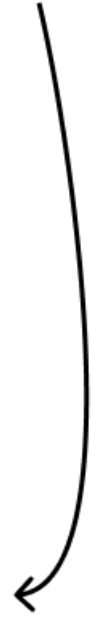
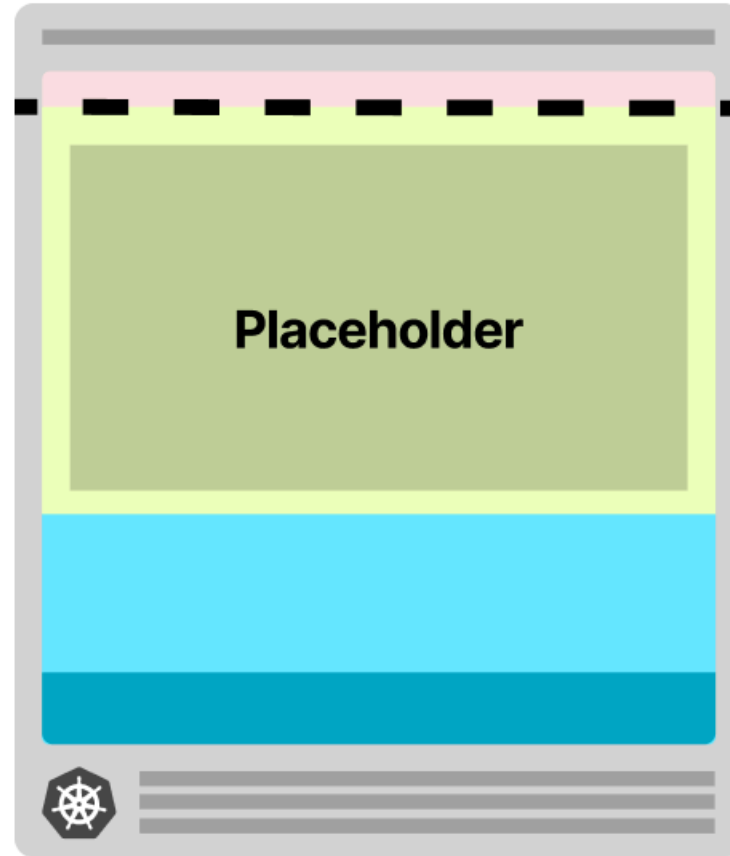
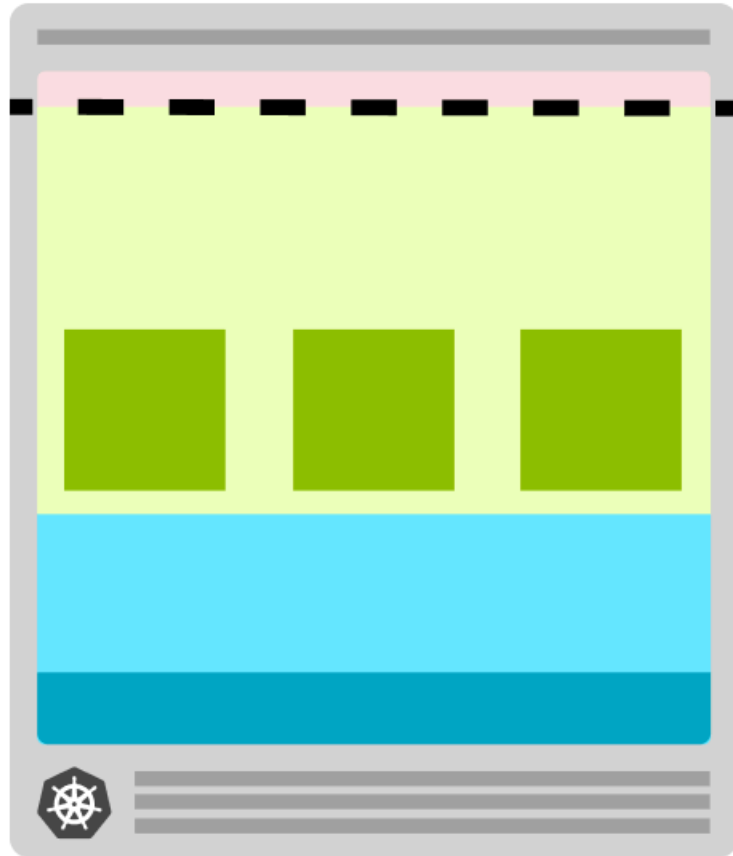
# DEMO



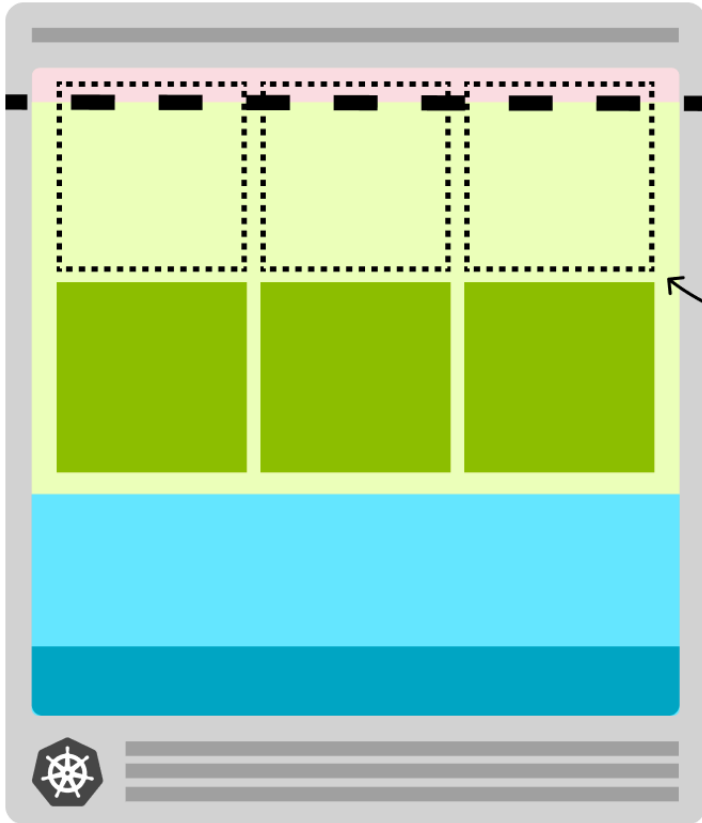




you are still billed for this node  
even if it does nothing



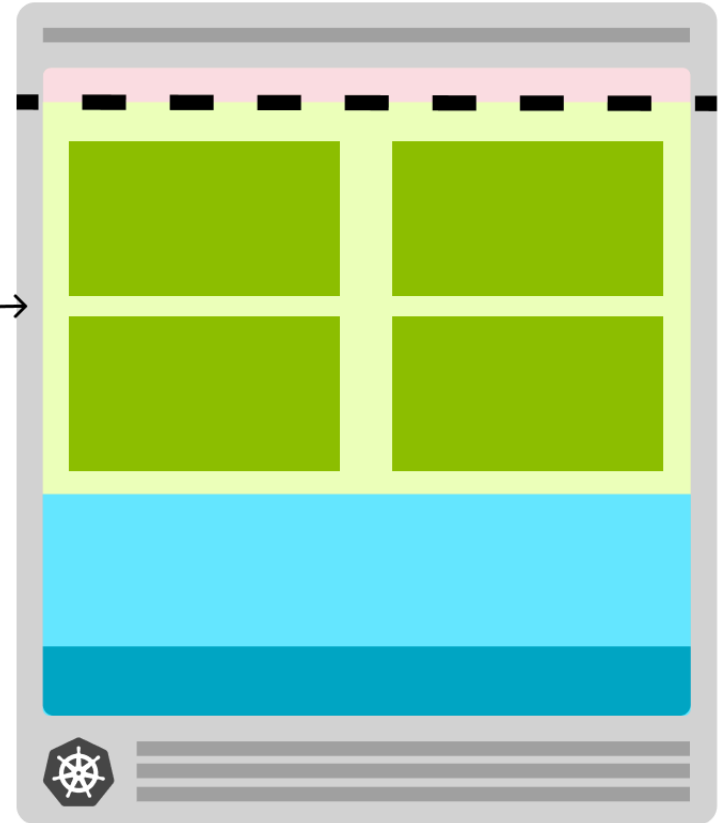
1



can't fit more pods.  
memory and CPU  
are not utilized in full

pods utilize all available  
resources

2



```
apiVersion: scheduling.k8s.io/v1  
kind: PriorityClass
```



🙏 Thanks 🙏

- [cns.me](https://cns.me)
- [talks.cns.me](https://talks.cns.me)

cns.me

Q&A 🙋 🙋 🙋

Chris Nesbitt-Smith

