ULAM LABS

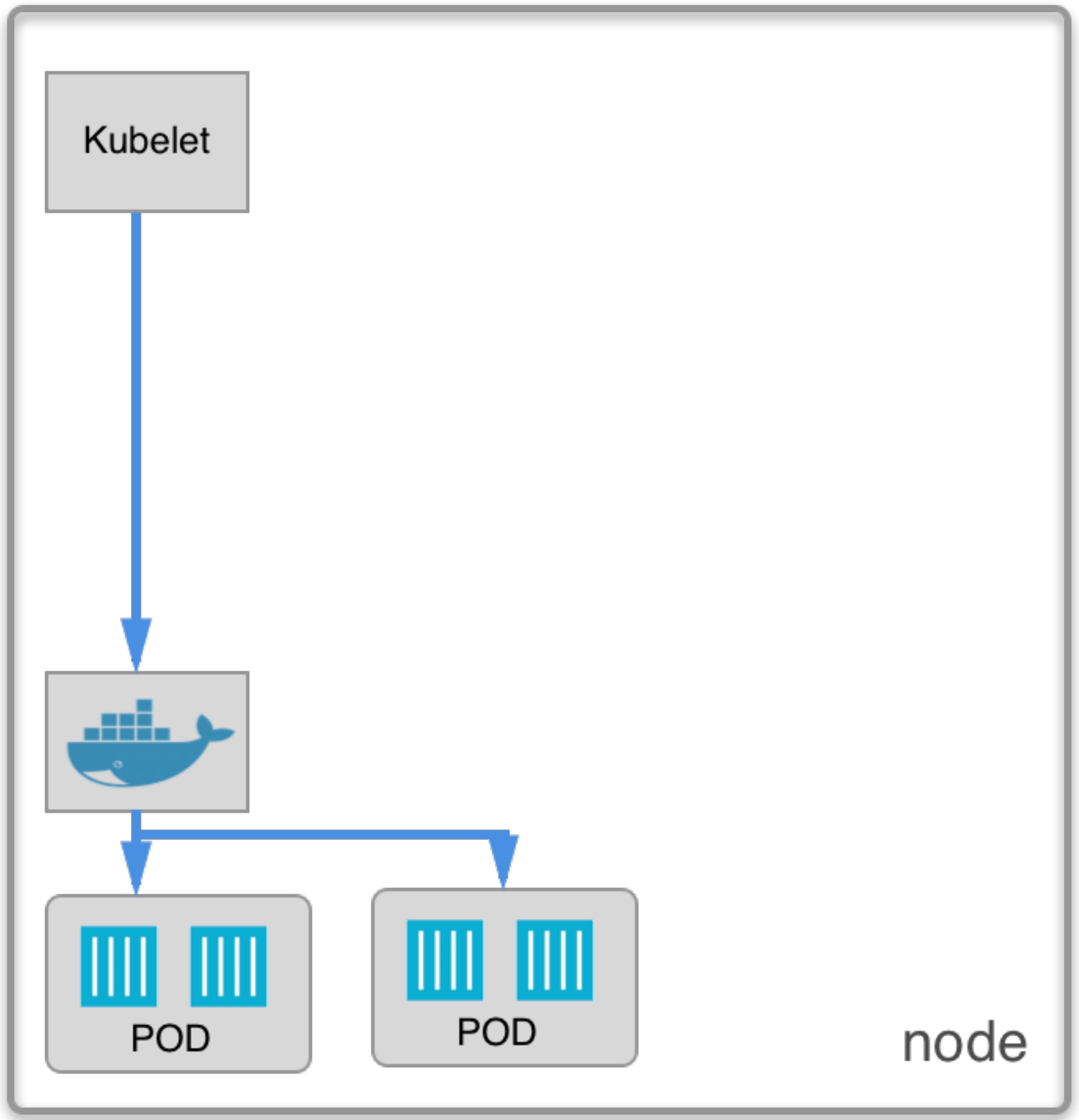
KUBERNETES FROM SCRATCH

Konrad Rotkiewicz
Owner @ Ulam Labs
konrad@ulam.io

AGENDA

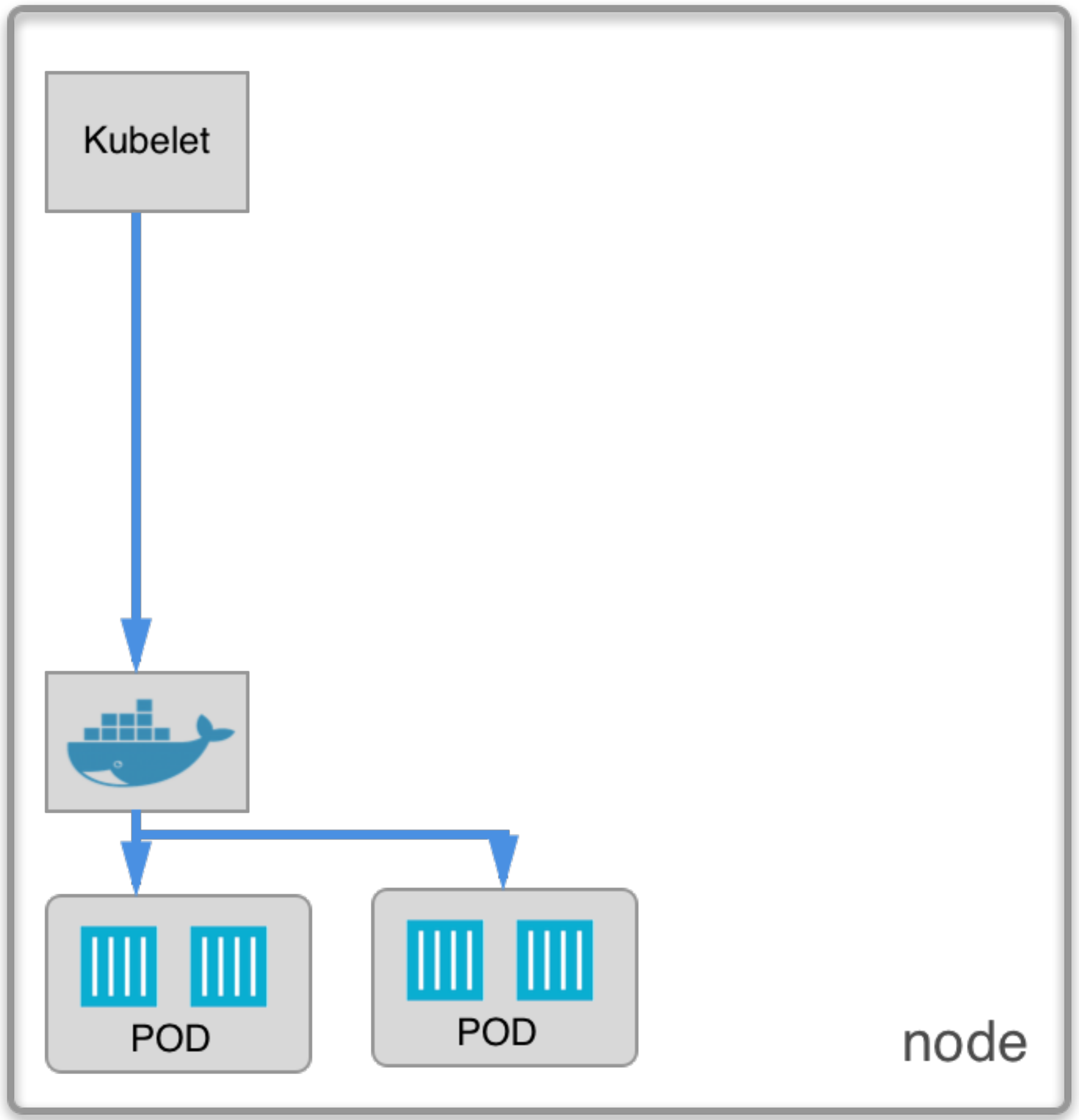
- kubelet
- api server
- etcd
- scheduler
- controller manager
- proxy
- networking
- IaaS integration
- Q&A

KUBELET



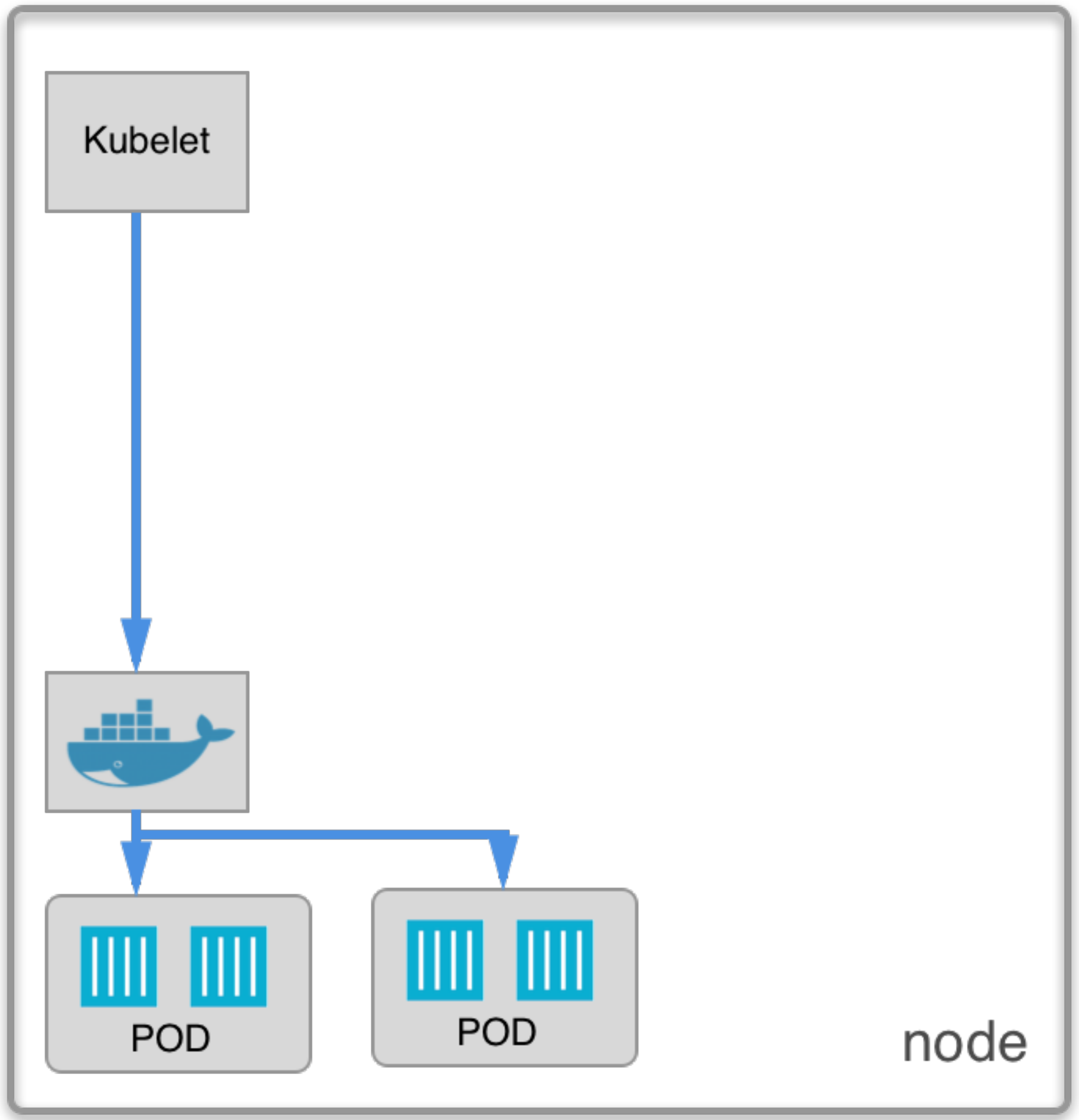
KUBELET

- the “worker”



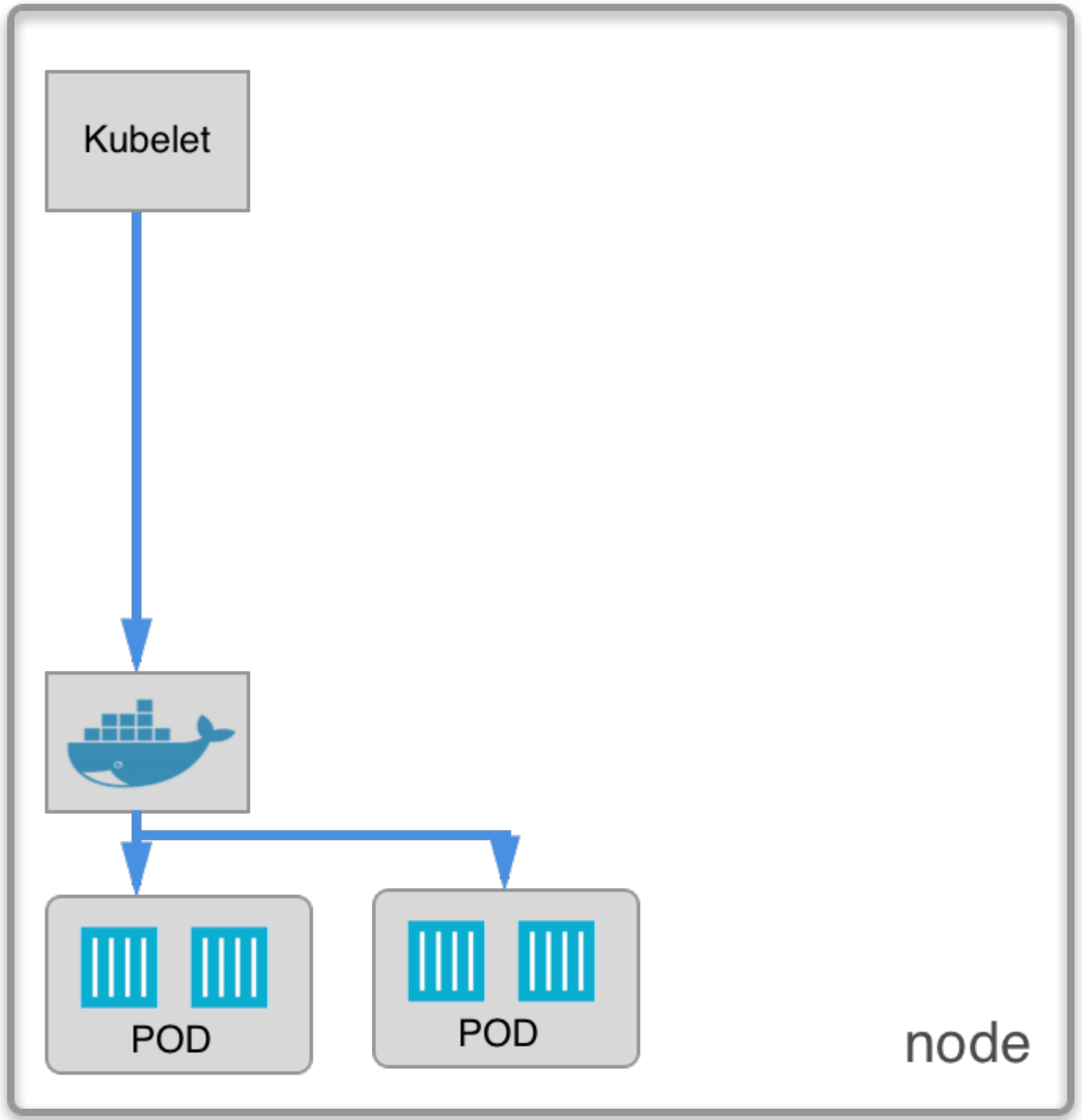
KUBELET

- the “worker”
- reads pod’s specification in YAML/JSON (from directory, own api or etcd)



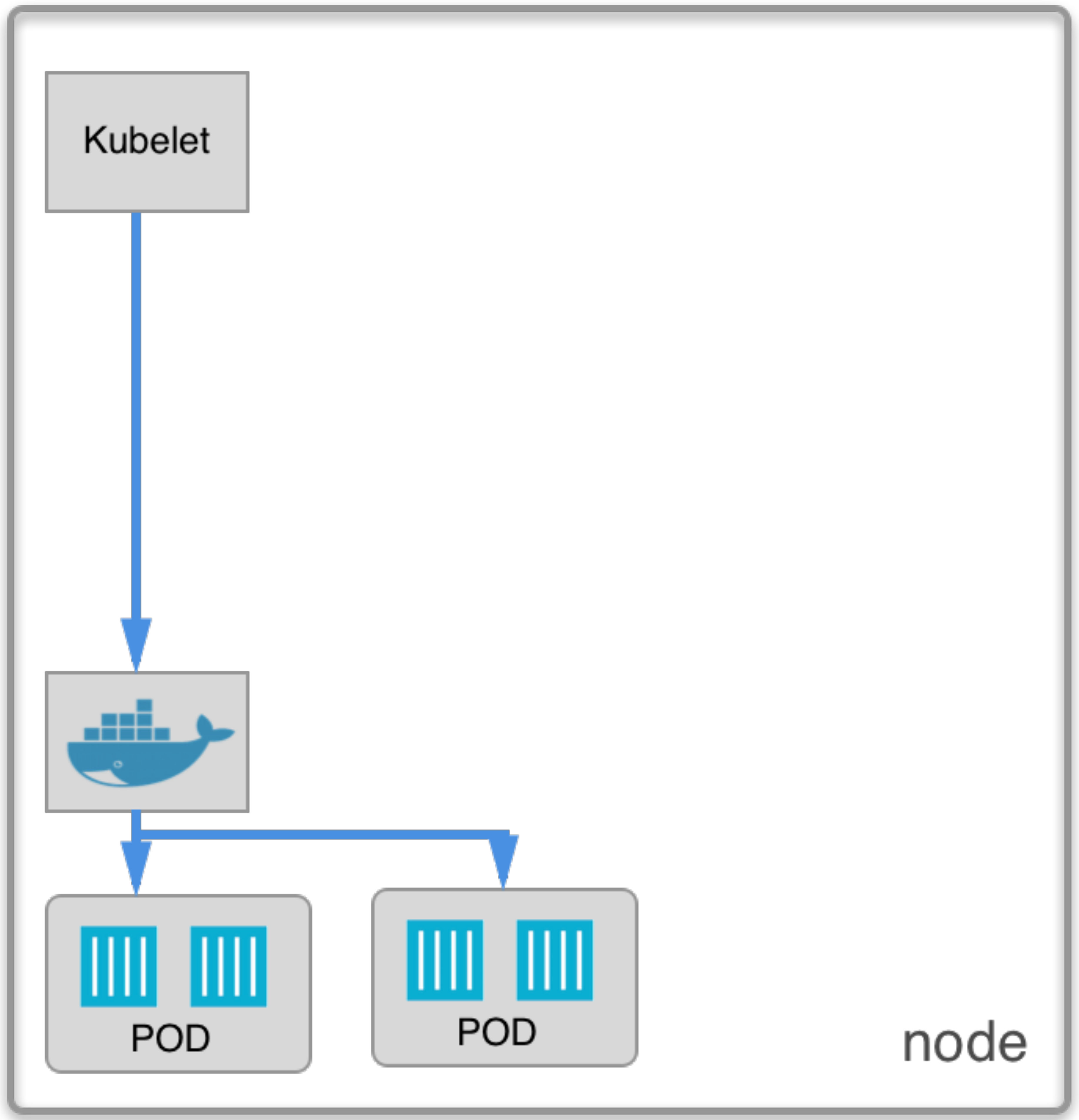
KUBELET

- the “worker”
- reads pod’s specification in YAML/JSON (from directory, own api or etcd)
- talks to the Docker



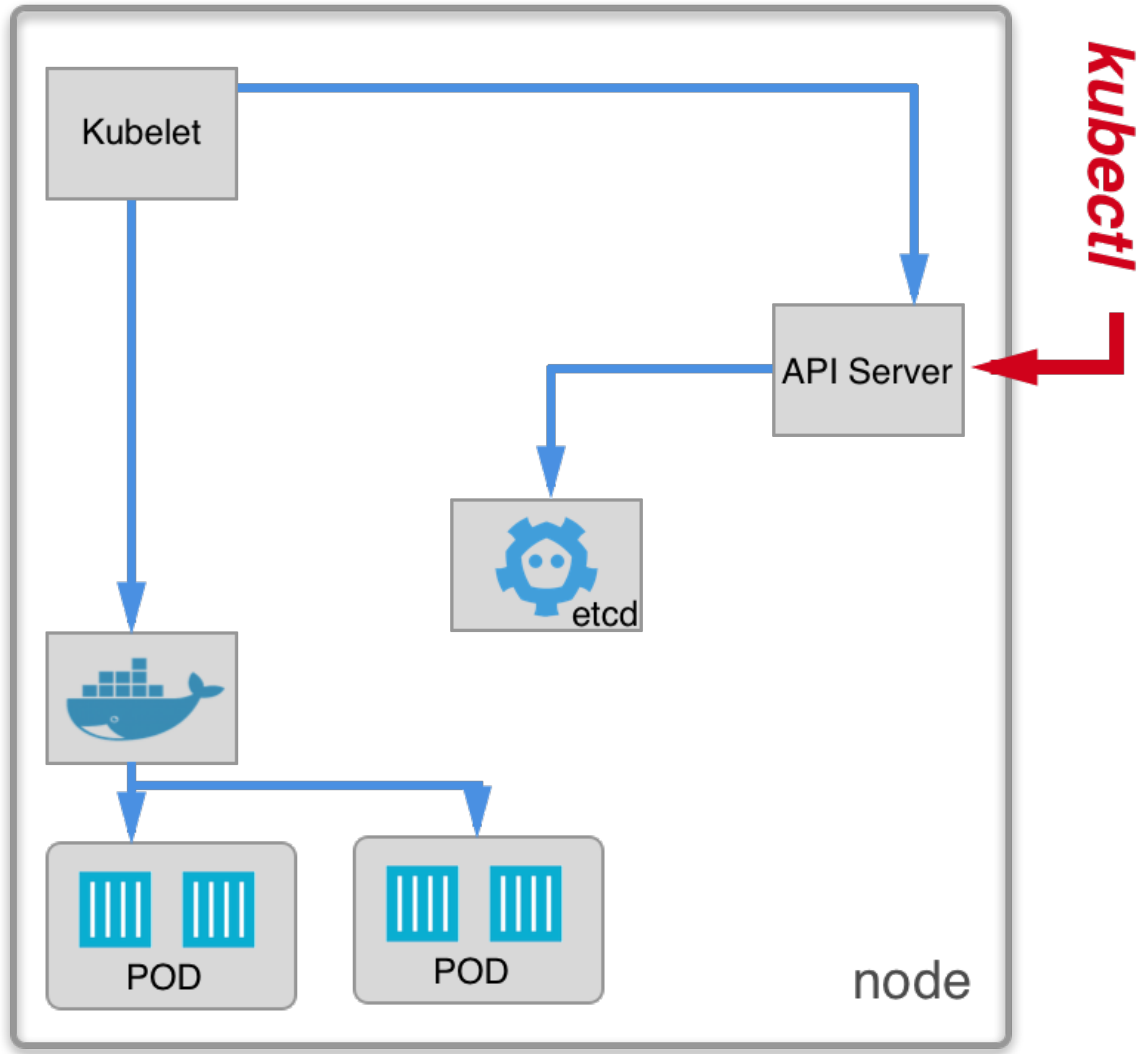
KUBELET

- the “worker”
- reads pod’s specification in YAML/JSON (from directory, own api or etcd)
- talks to the Docker
- creates pods / containers



ETCD + API SERVER

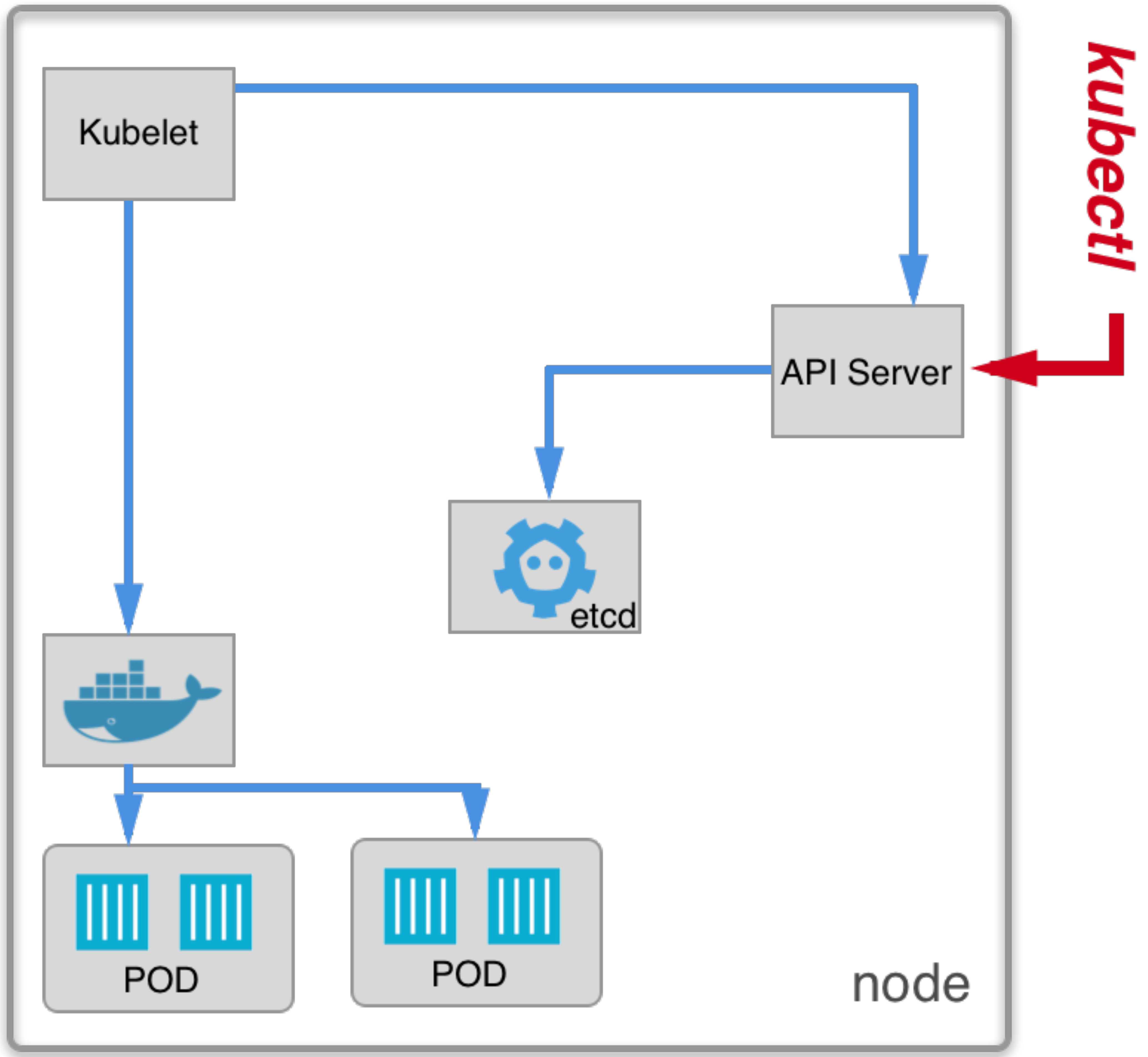
etcd



ETCD + API SERVER

etcd

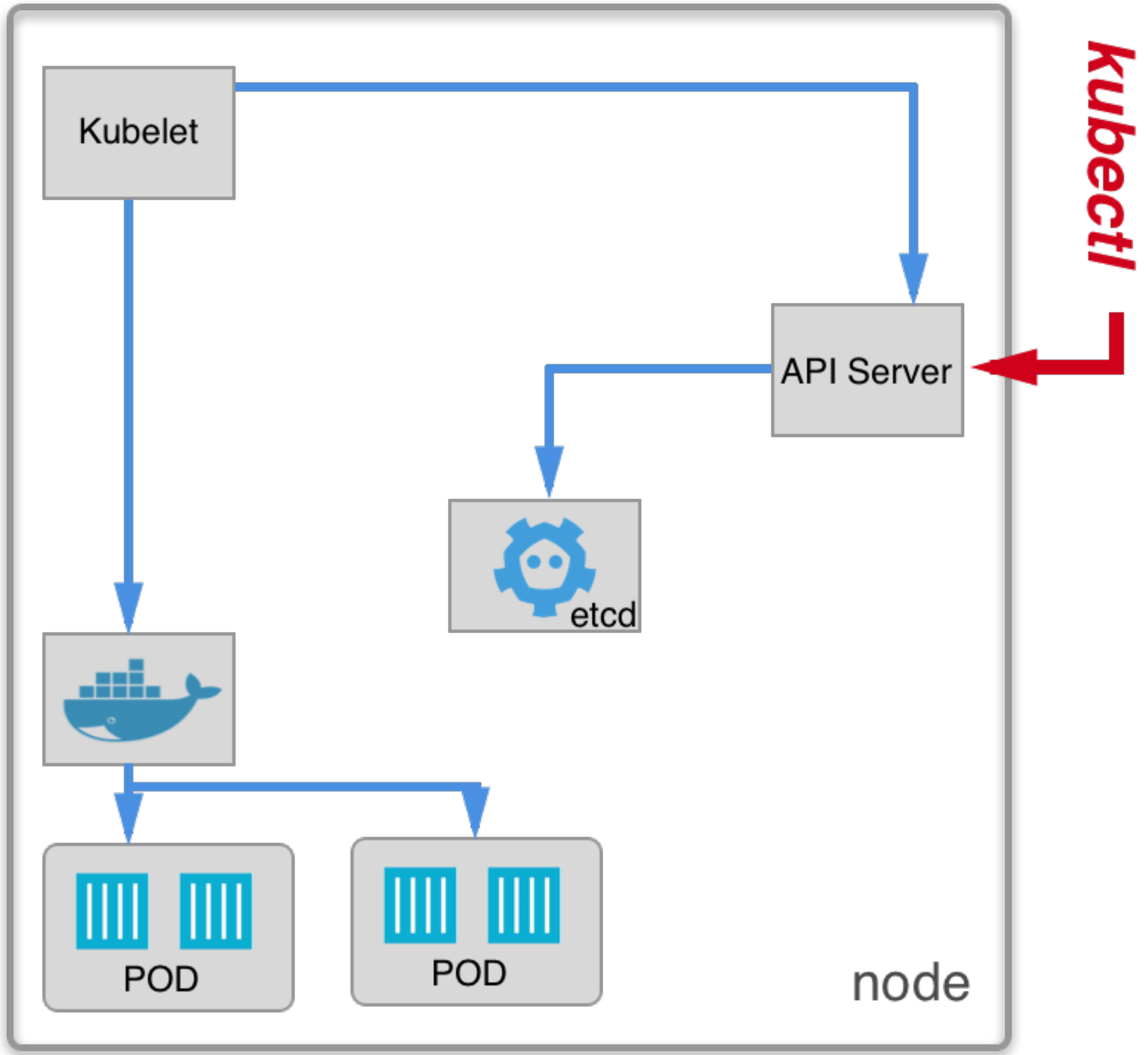
- key-value distributed database



ETCD + API SERVER

etcd

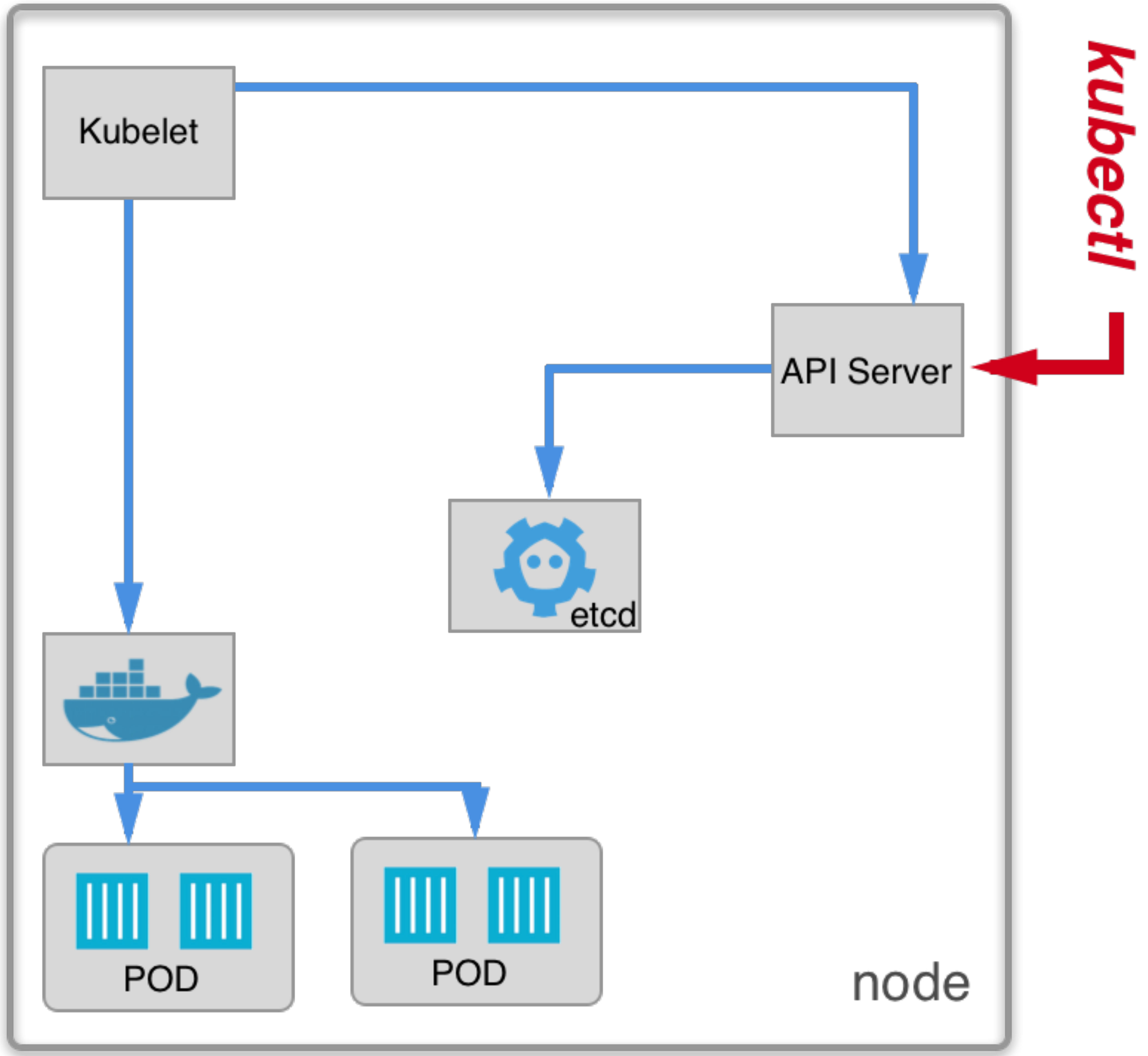
- key-value distributed database
- keys can be watchable



ETCD + API SERVER

etcd

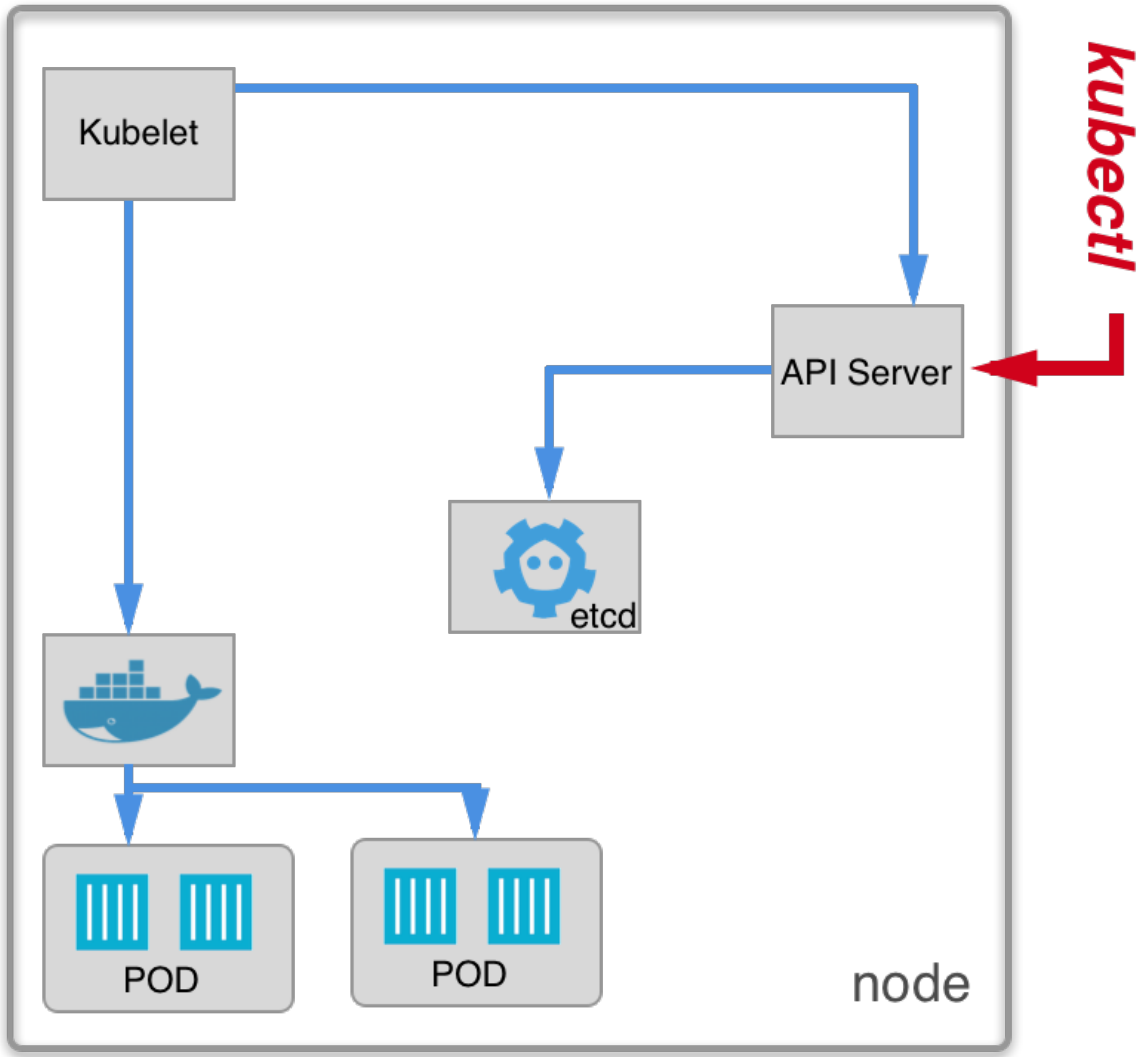
- key-value distributed database
- keys can be watchable
- “similar” to zookeeper or consul



ETCD + API SERVER

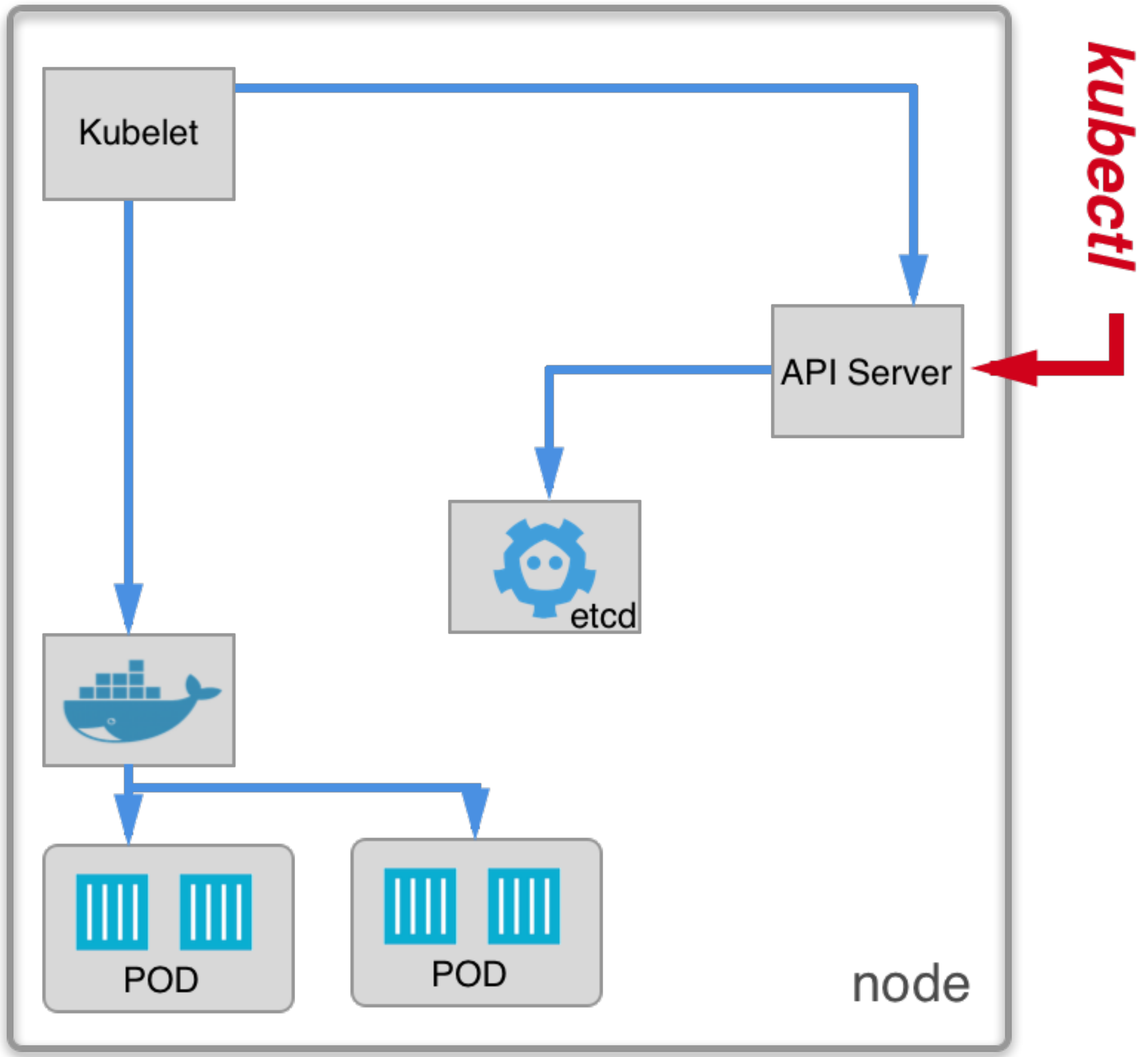
etcd

- key-value distributed database
- keys can be watchable
- “similar” to zookeeper or consul
- stores state of the cluster



ETCD + API SERVER

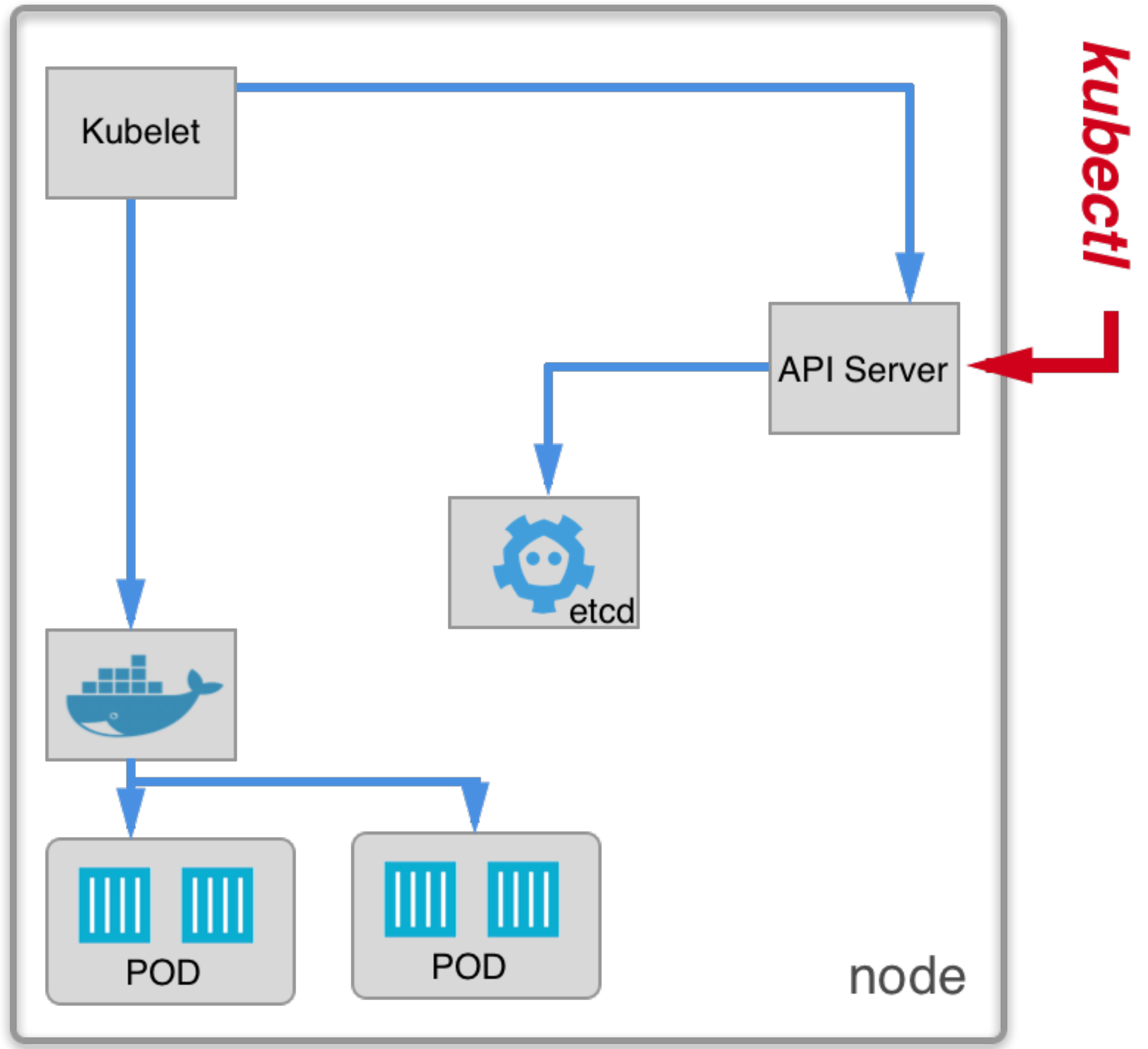
API Server



ETCD + API SERVER

API Server

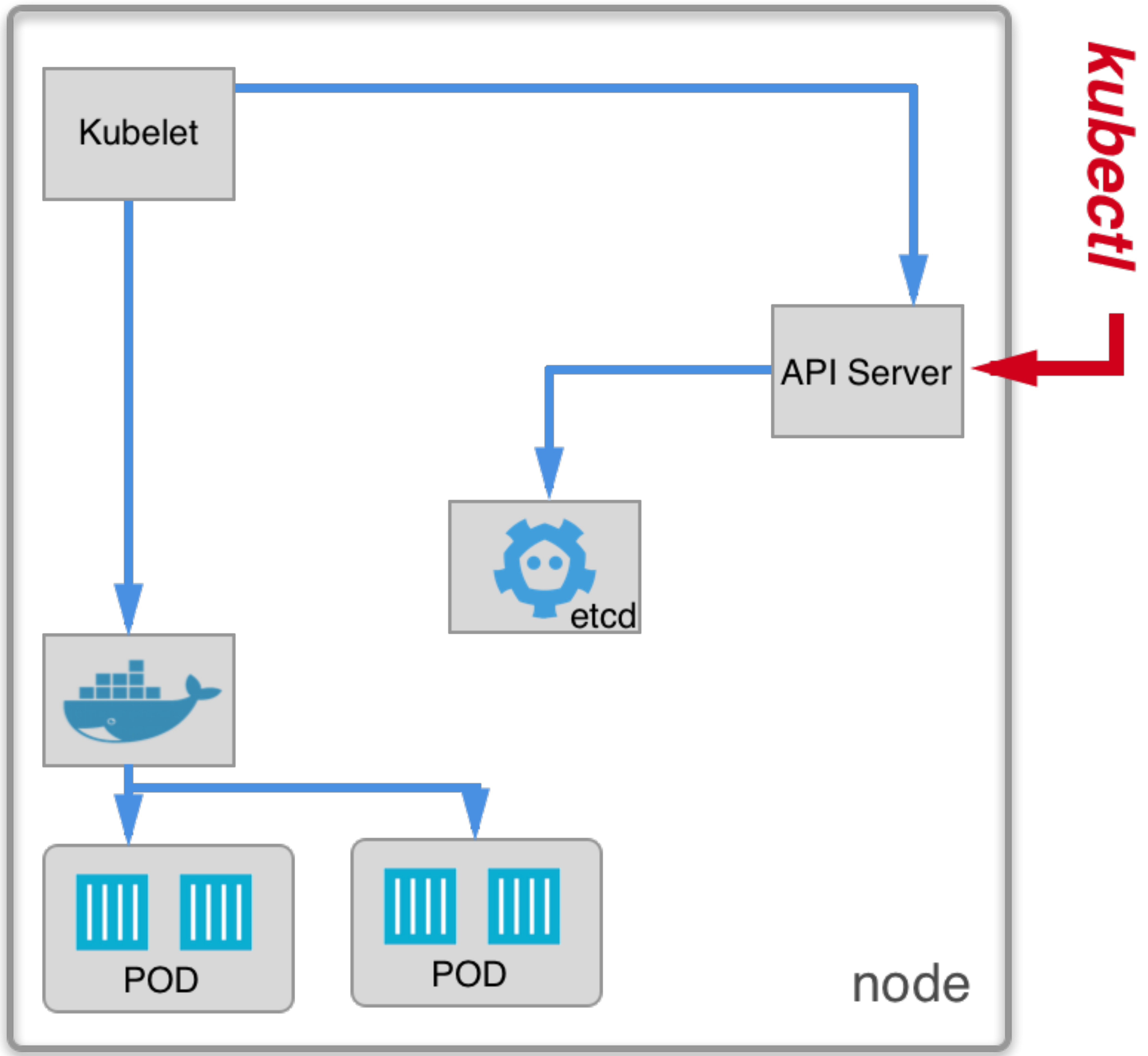
- “gateway” to etcd



ETCD + API SERVER

API Server

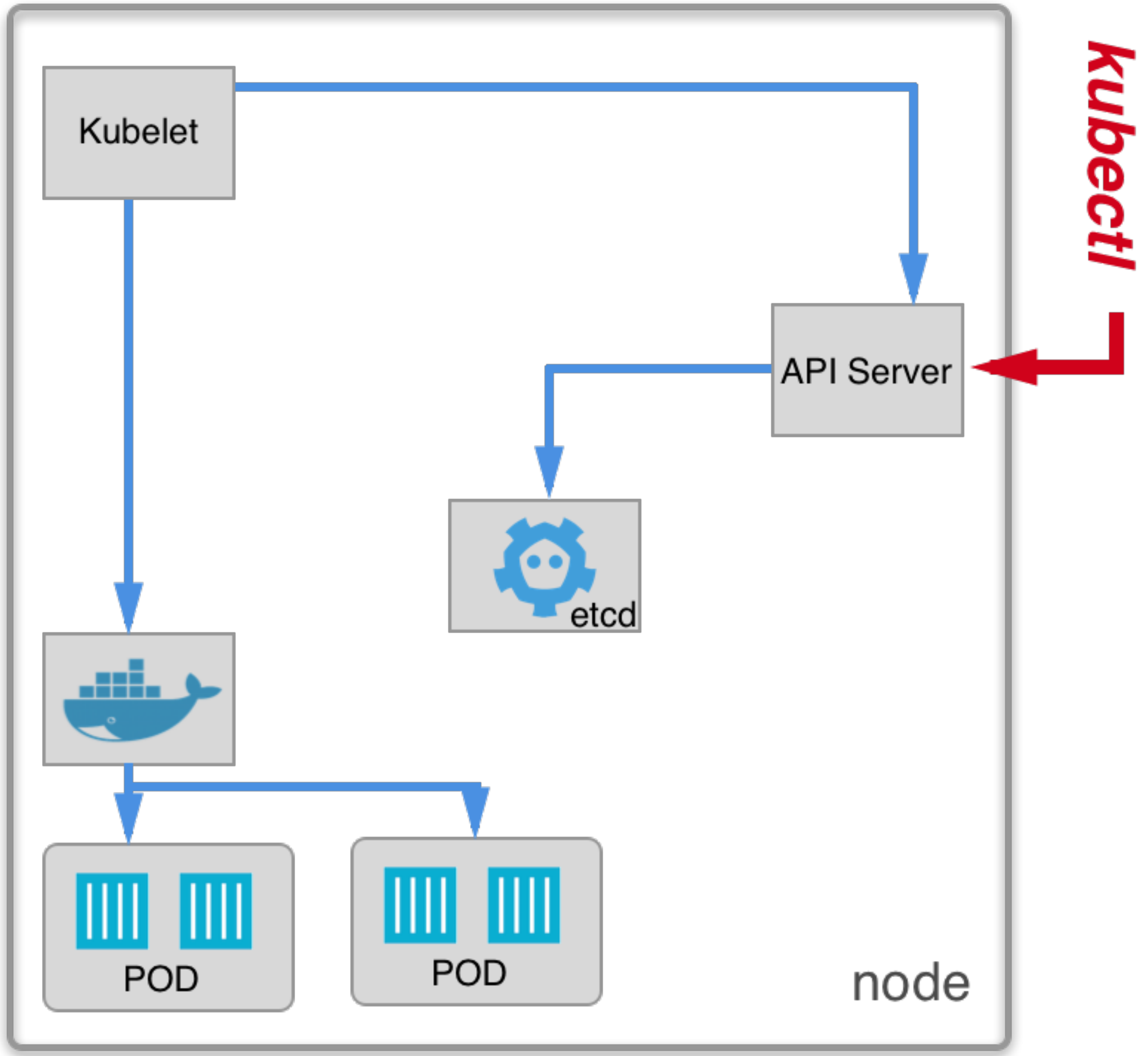
- "gateway" to etcd
- authenticates and authorizes requests



ETCD + API SERVER

API Server

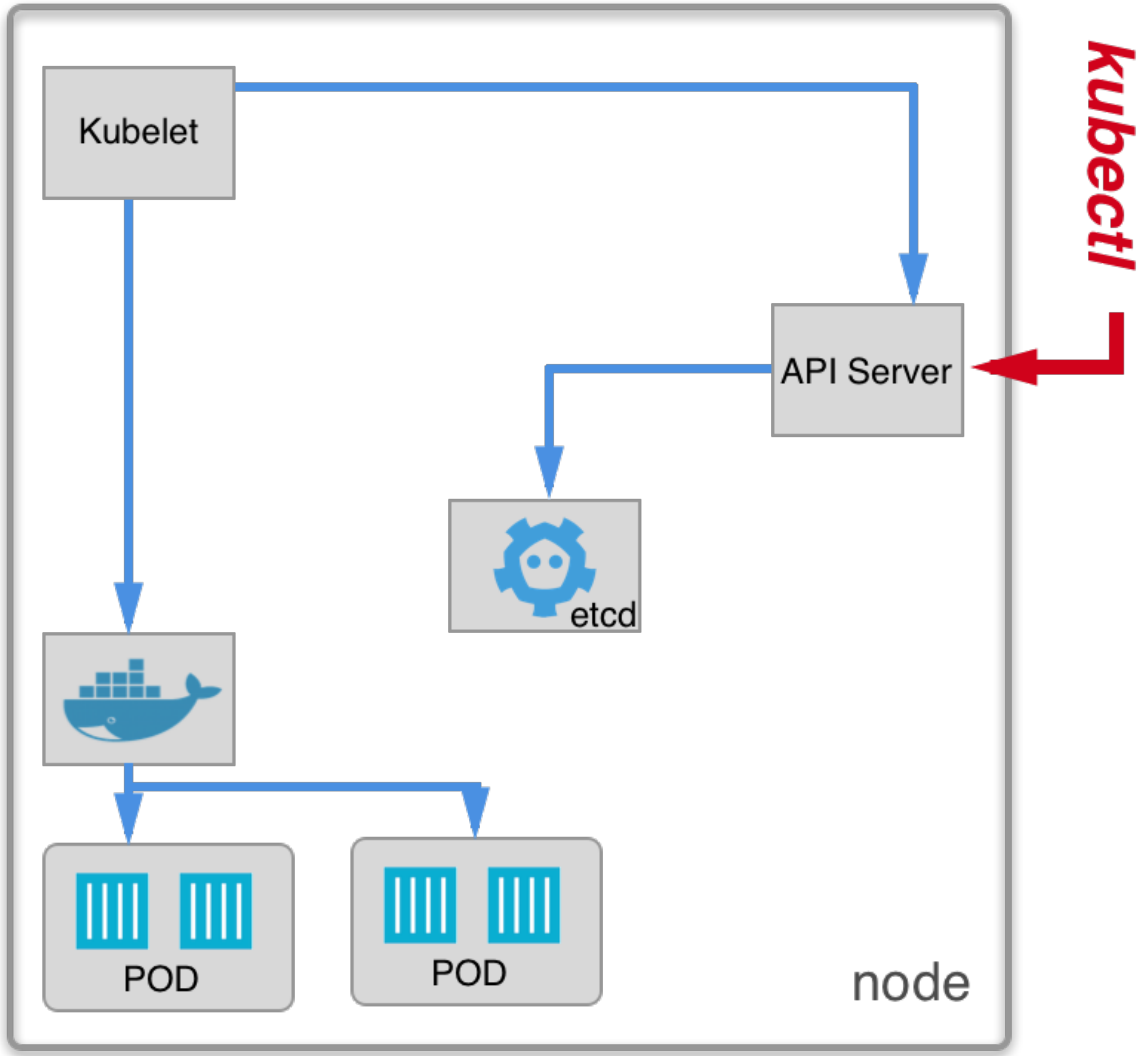
- "gateway" to etcd
- authenticates and authorizes requests
- all k8s components talk to it



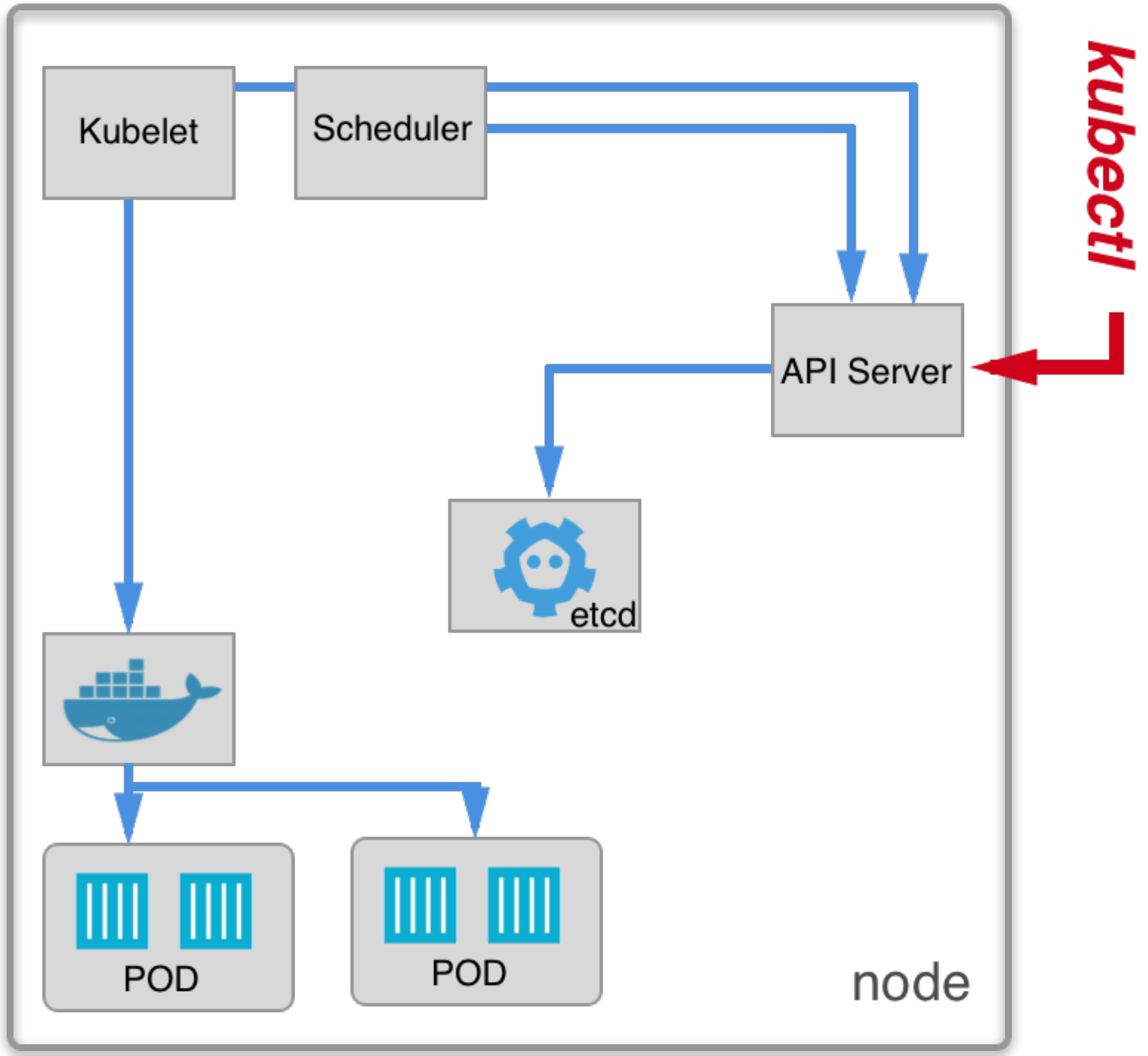
ETCD + API SERVER

API Server

- "gateway" to etcd
- authenticates and authorizes requests
- all k8s components talk to it
- kubectl talks to it

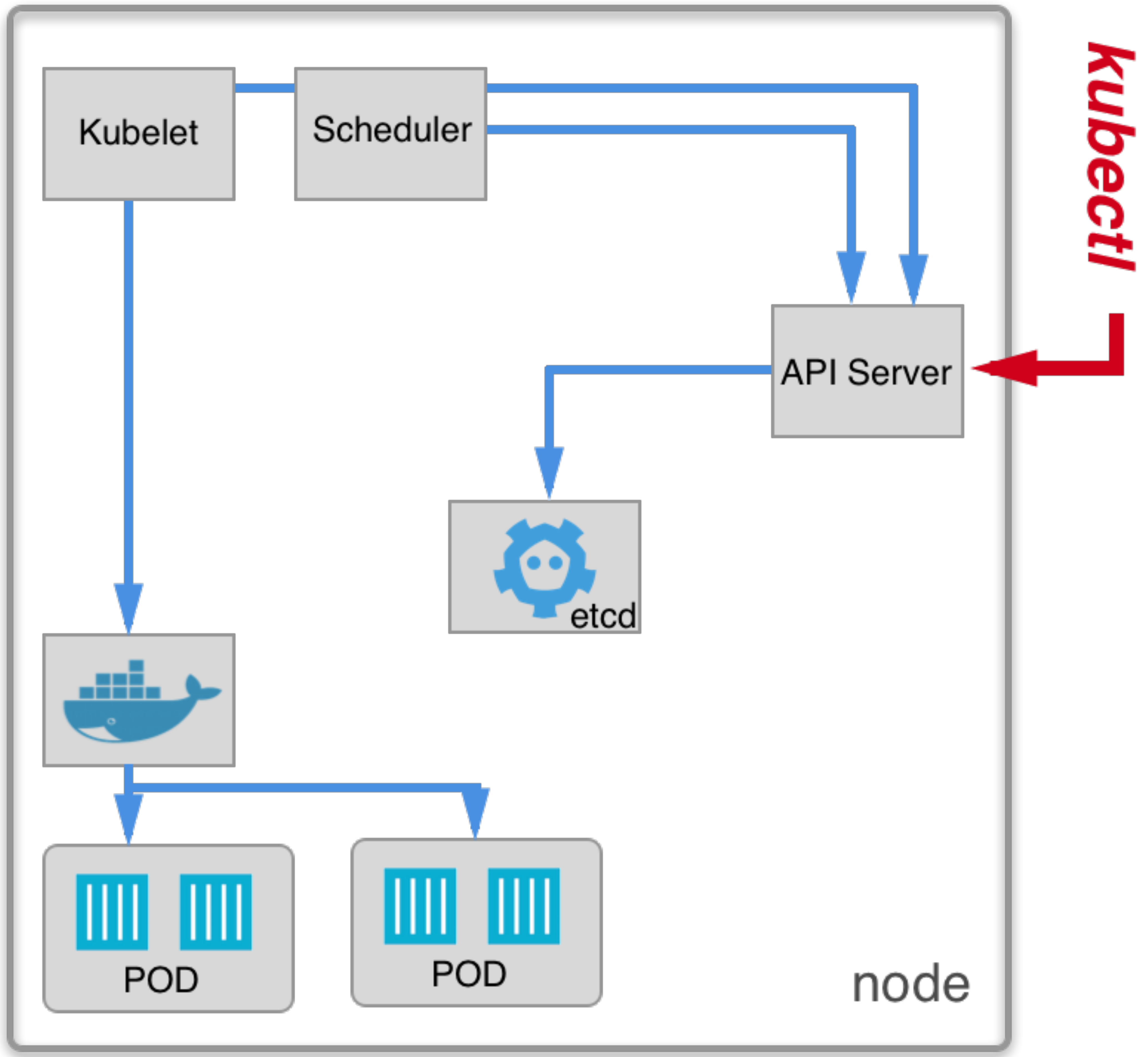


SCHEDULER



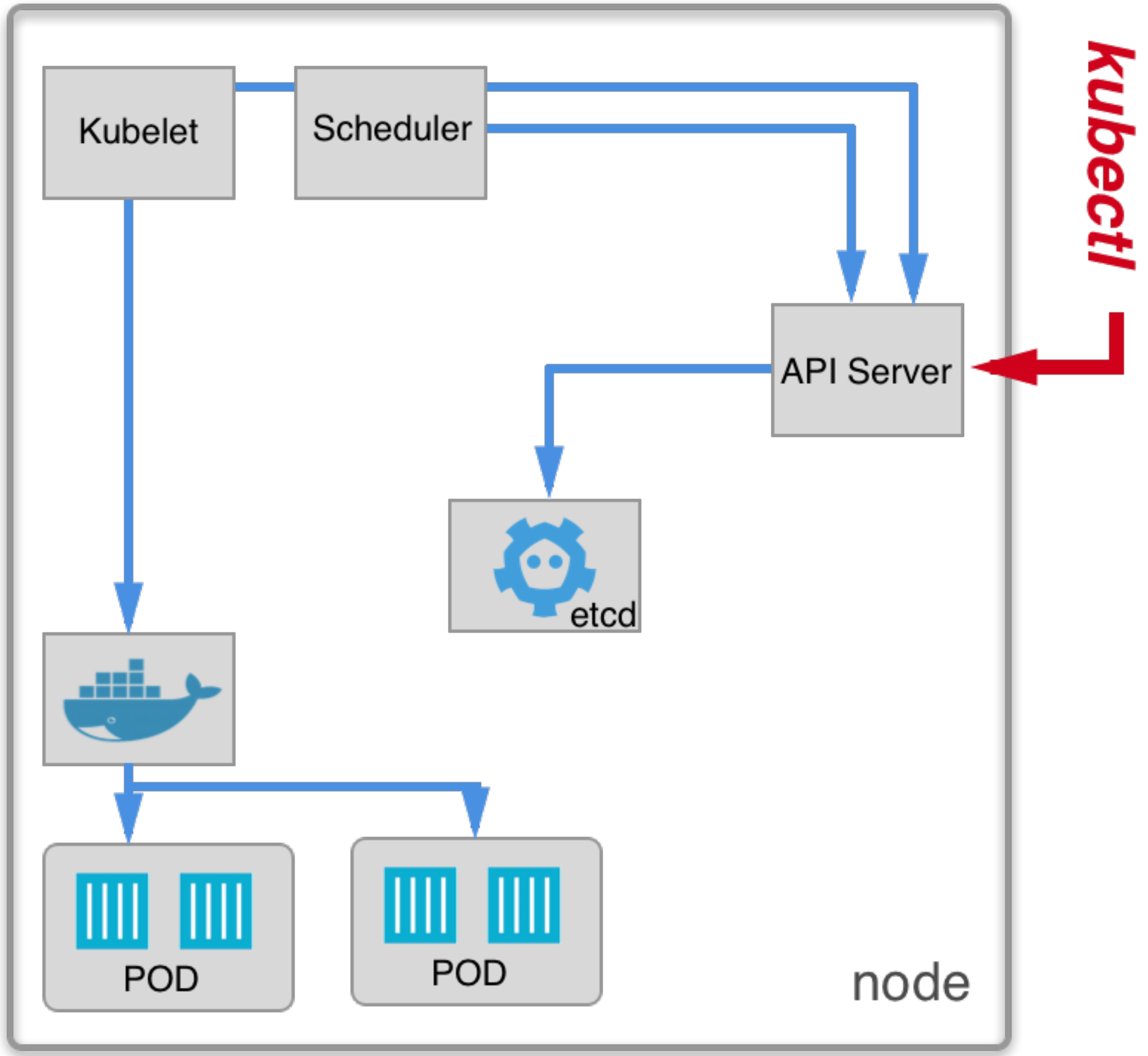
SCHEDULER

- responsible for scheduling pods on nodes



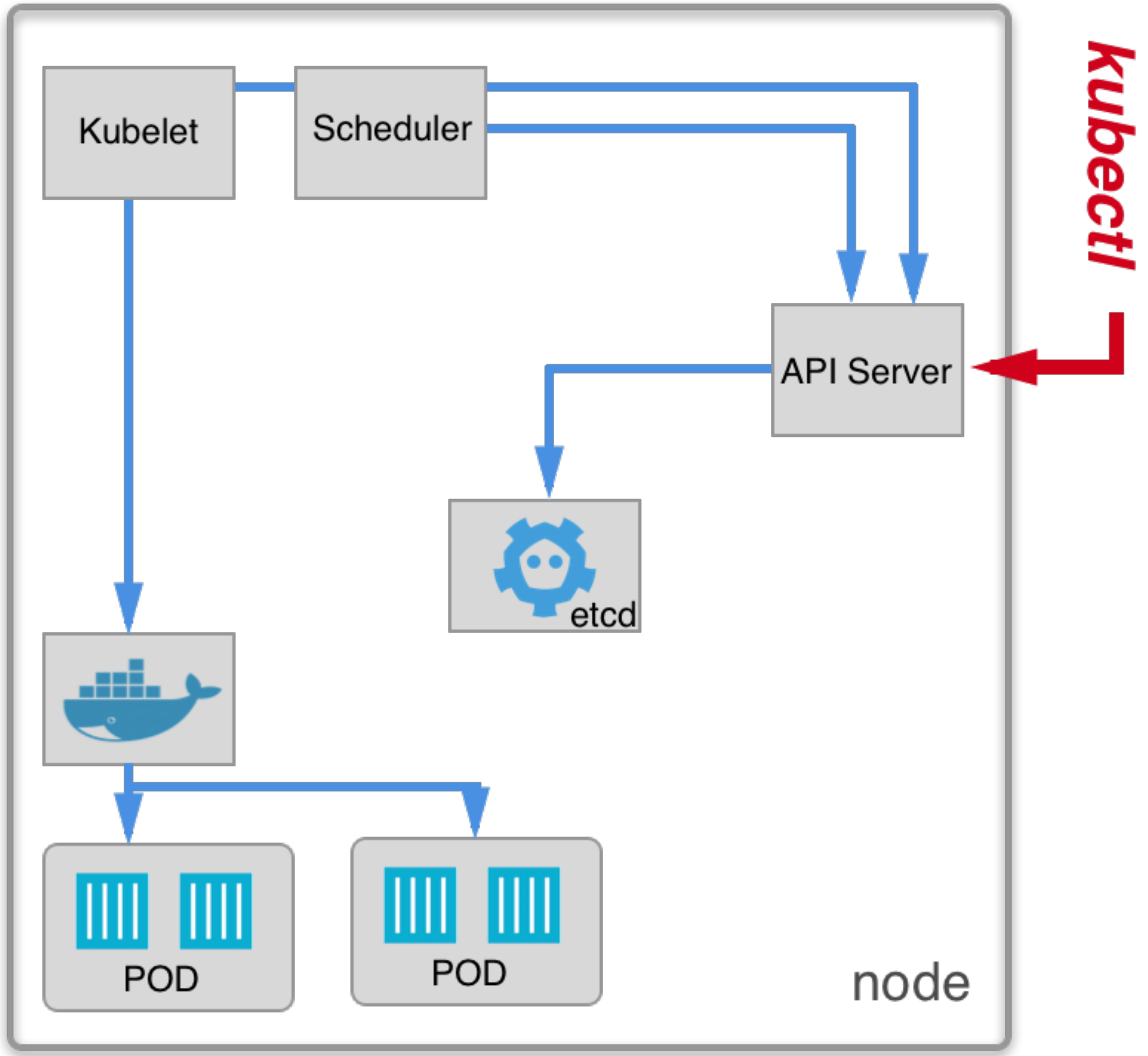
SCHEDULER

- responsible for scheduling pods on nodes
- connects to API Server

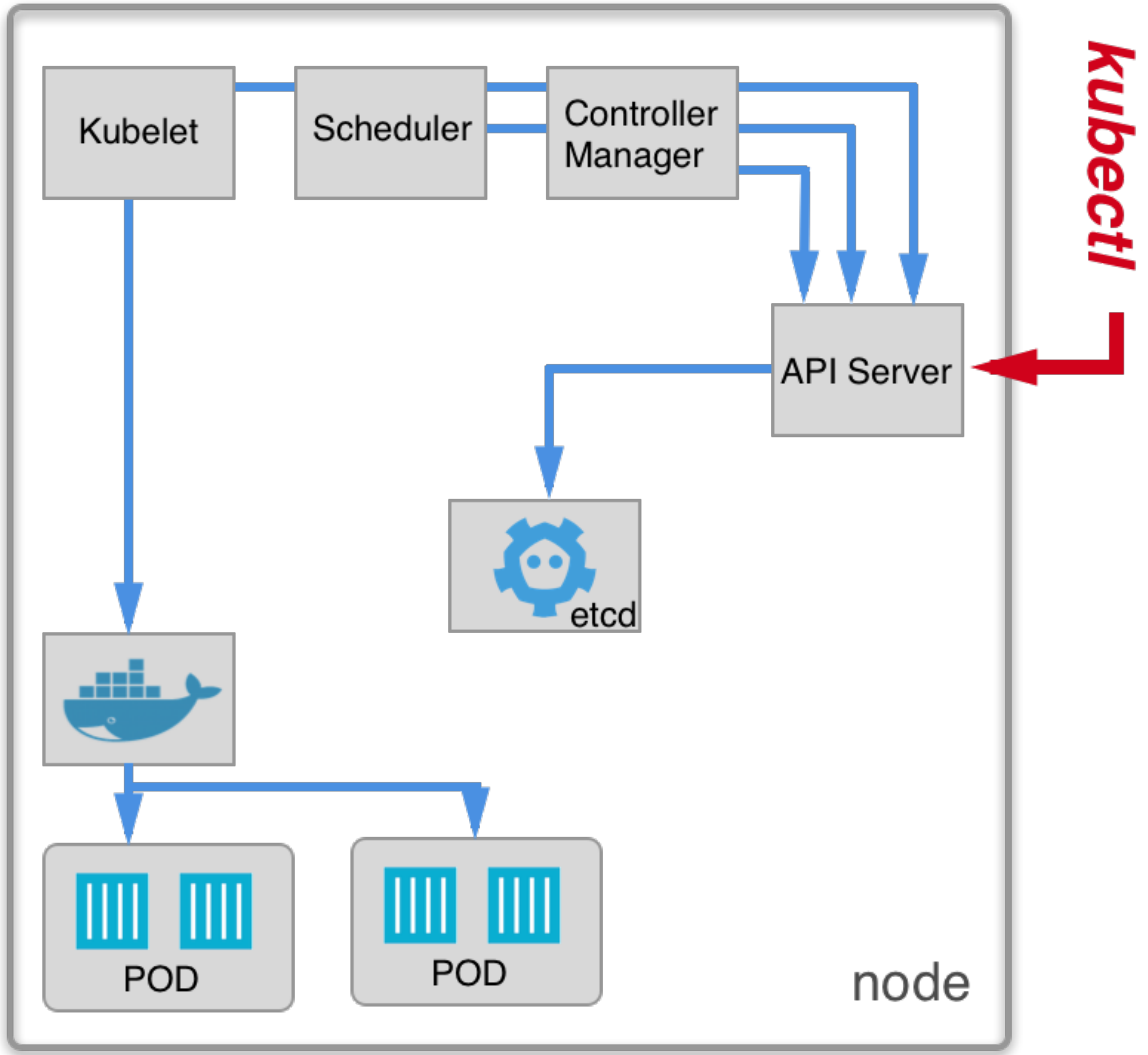


SCHEDULER

- responsible for scheduling pods on nodes
- connects to API Server
- watches for pods that aren't bound to a node and assigns one

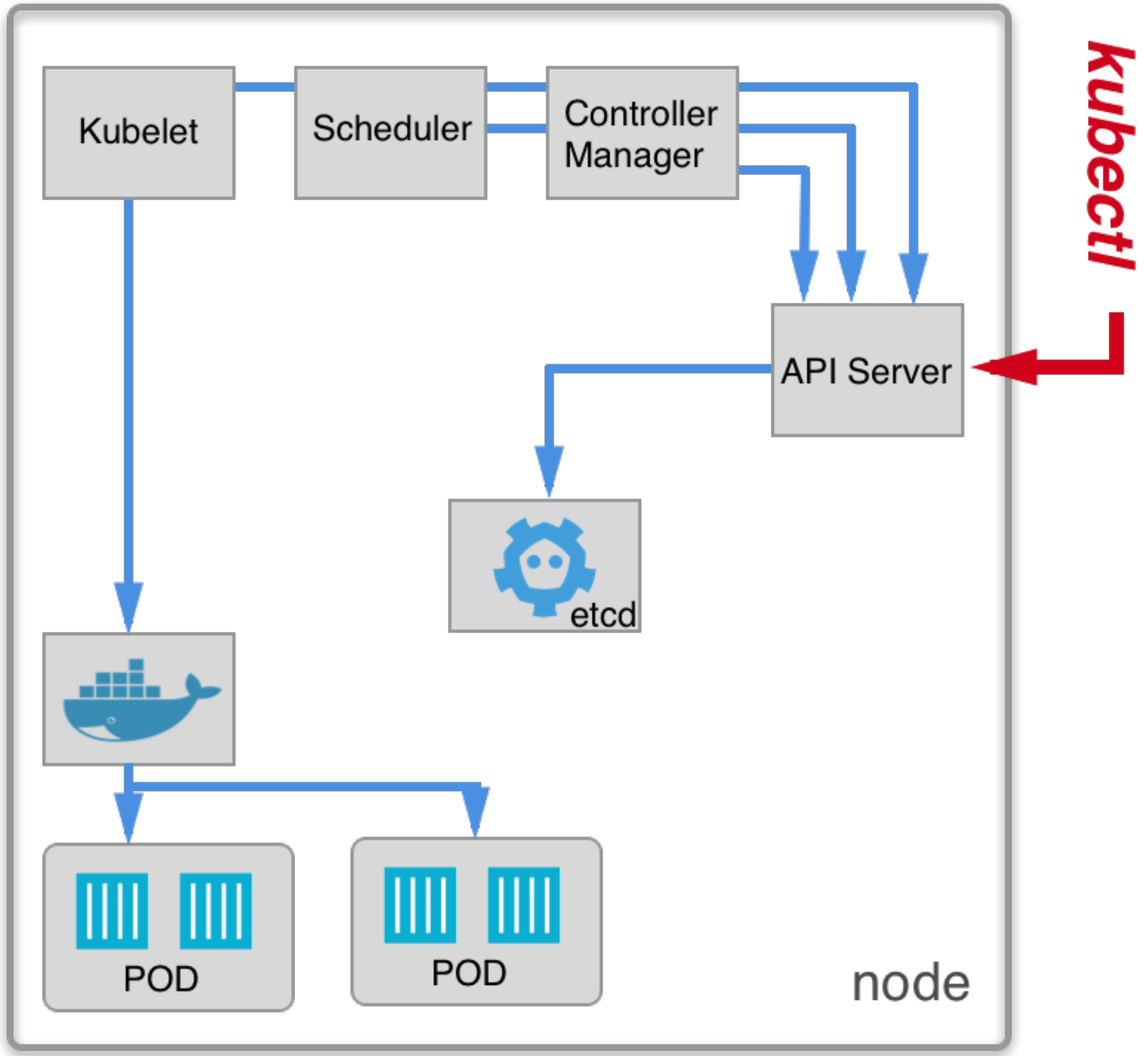


CONTROLLER MANAGER



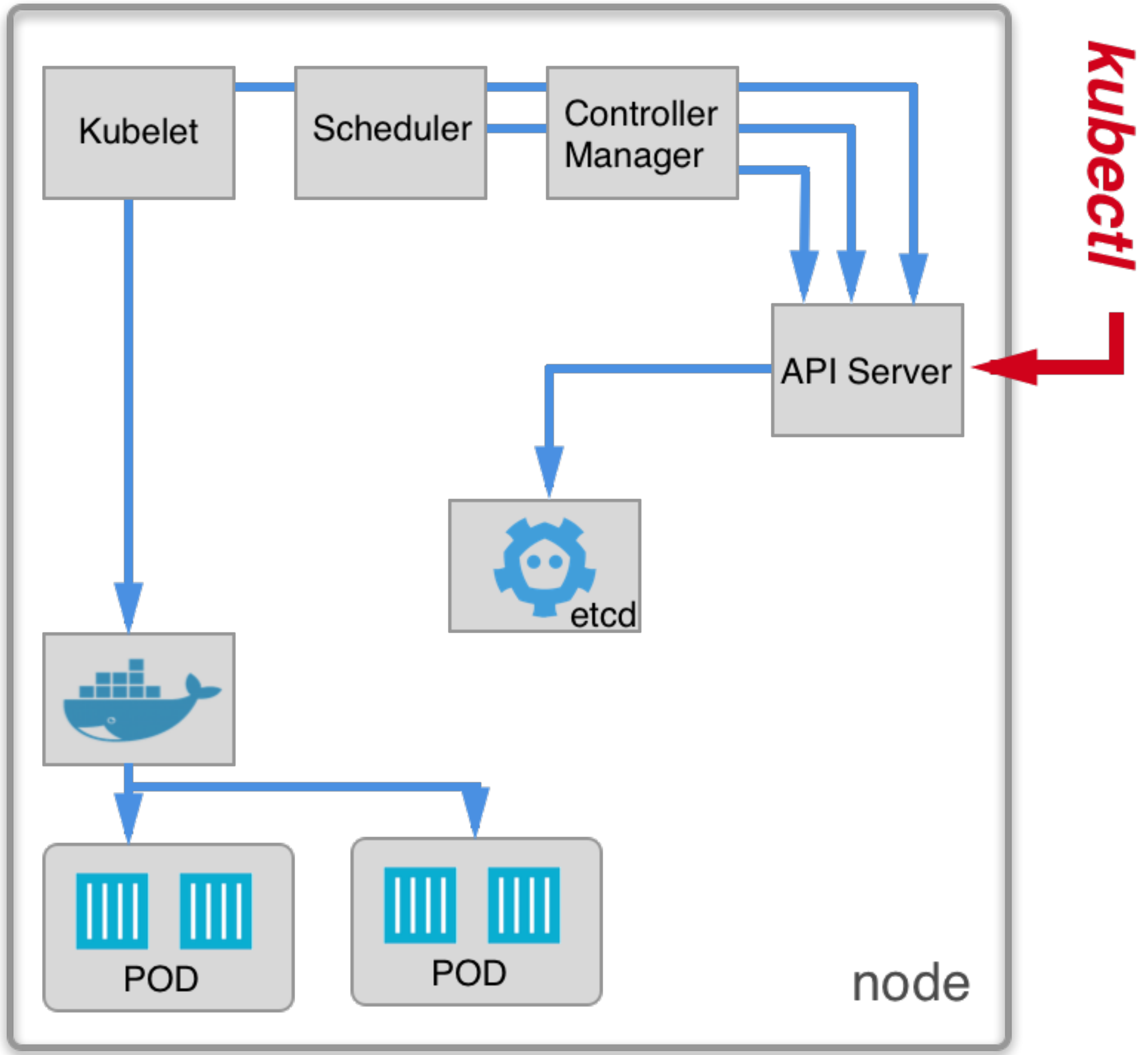
CONTROLLER MANAGER

- responsible for managing Deployments and ReplicaSets



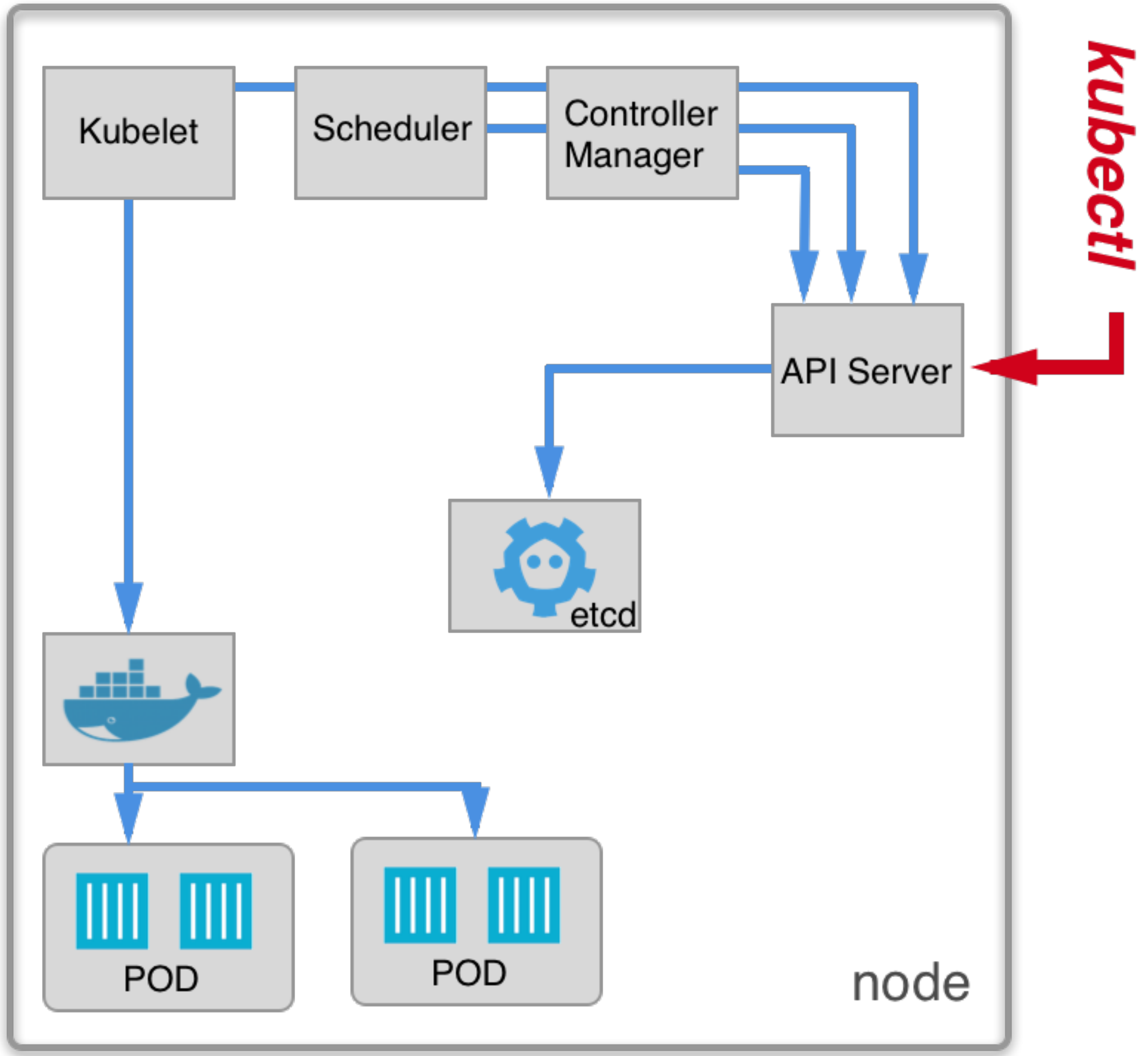
CONTROLLER MANAGER

- responsible for managing Deployments and ReplicaSets
- connects to API Server

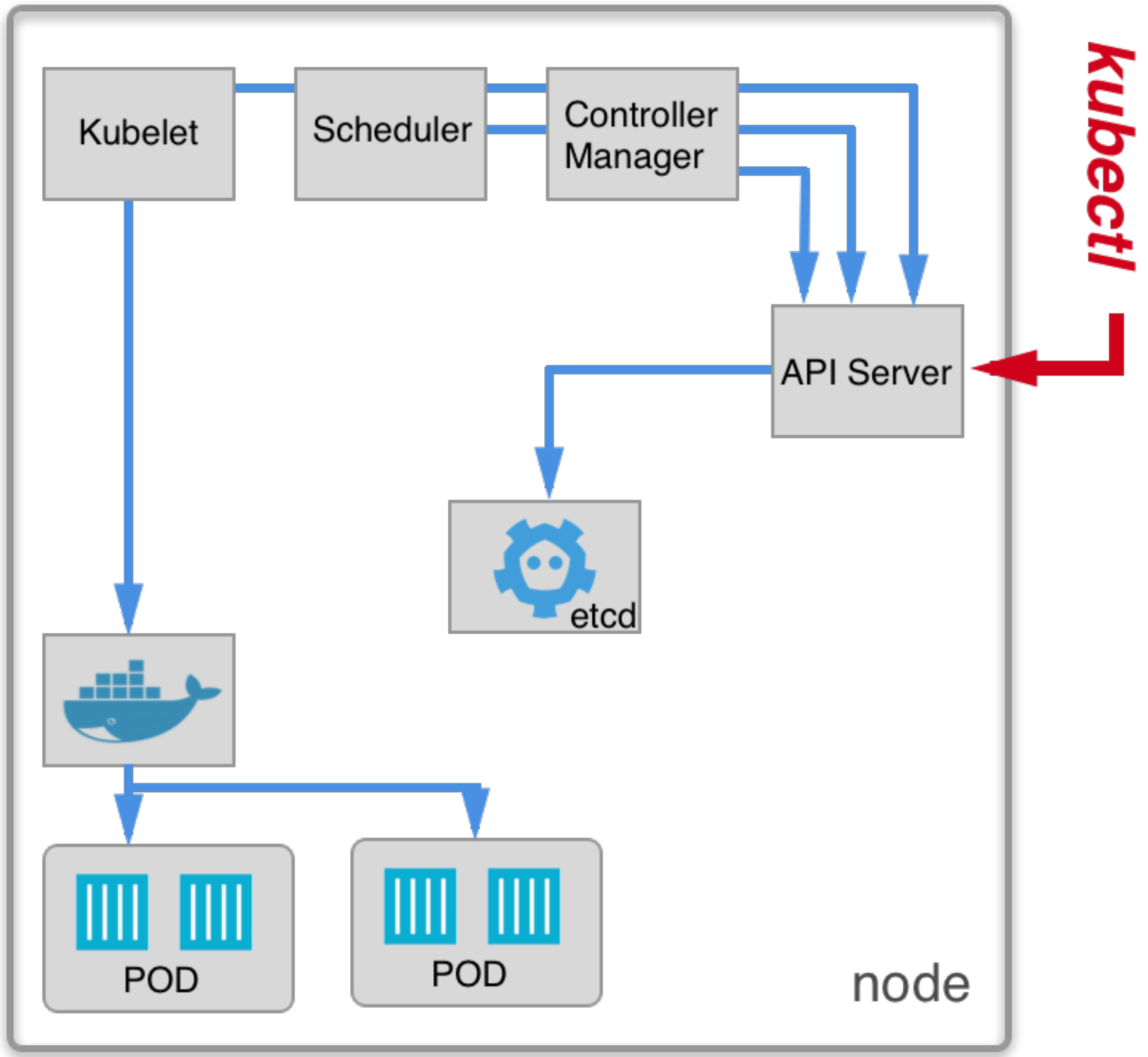


CONTROLLER MANAGER

- responsible for managing Deployments and ReplicaSets
- connects to API Server
- takes care of keeping pods counts at given number

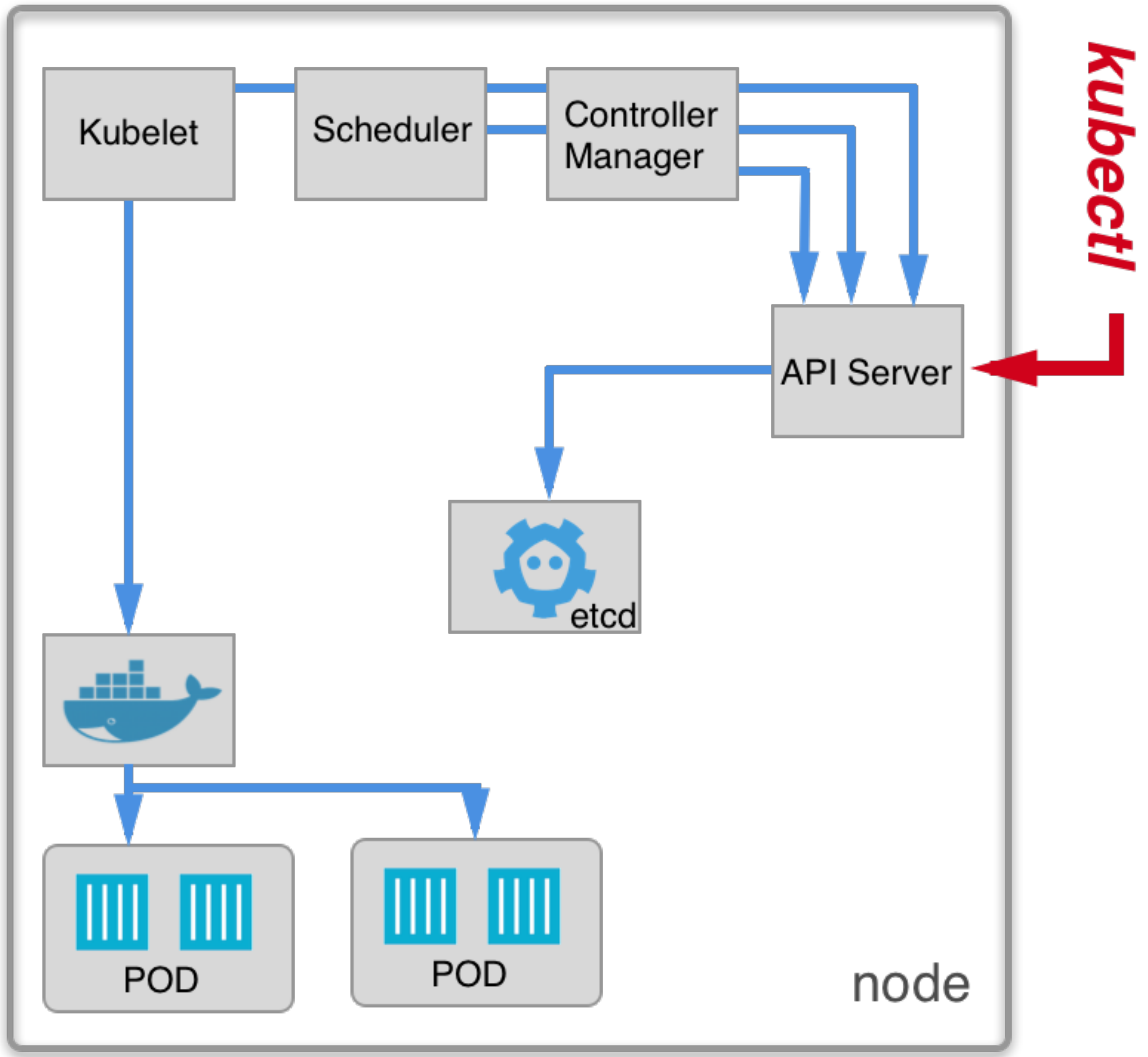


BASIC FLOW


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
POD
POD
node

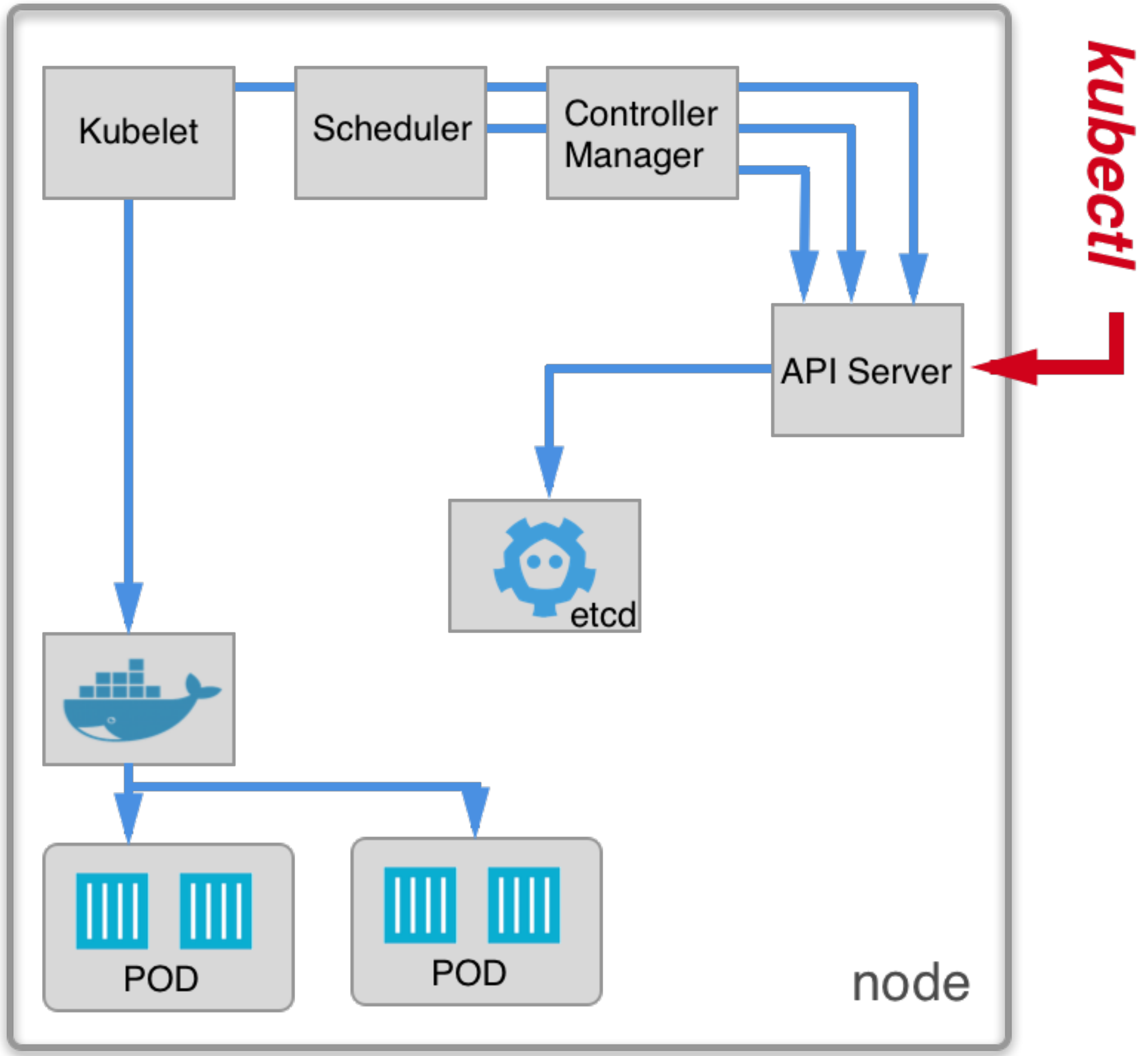
BASIC FLOW

- kubectl apply -f deploy.yaml (5 replicas)



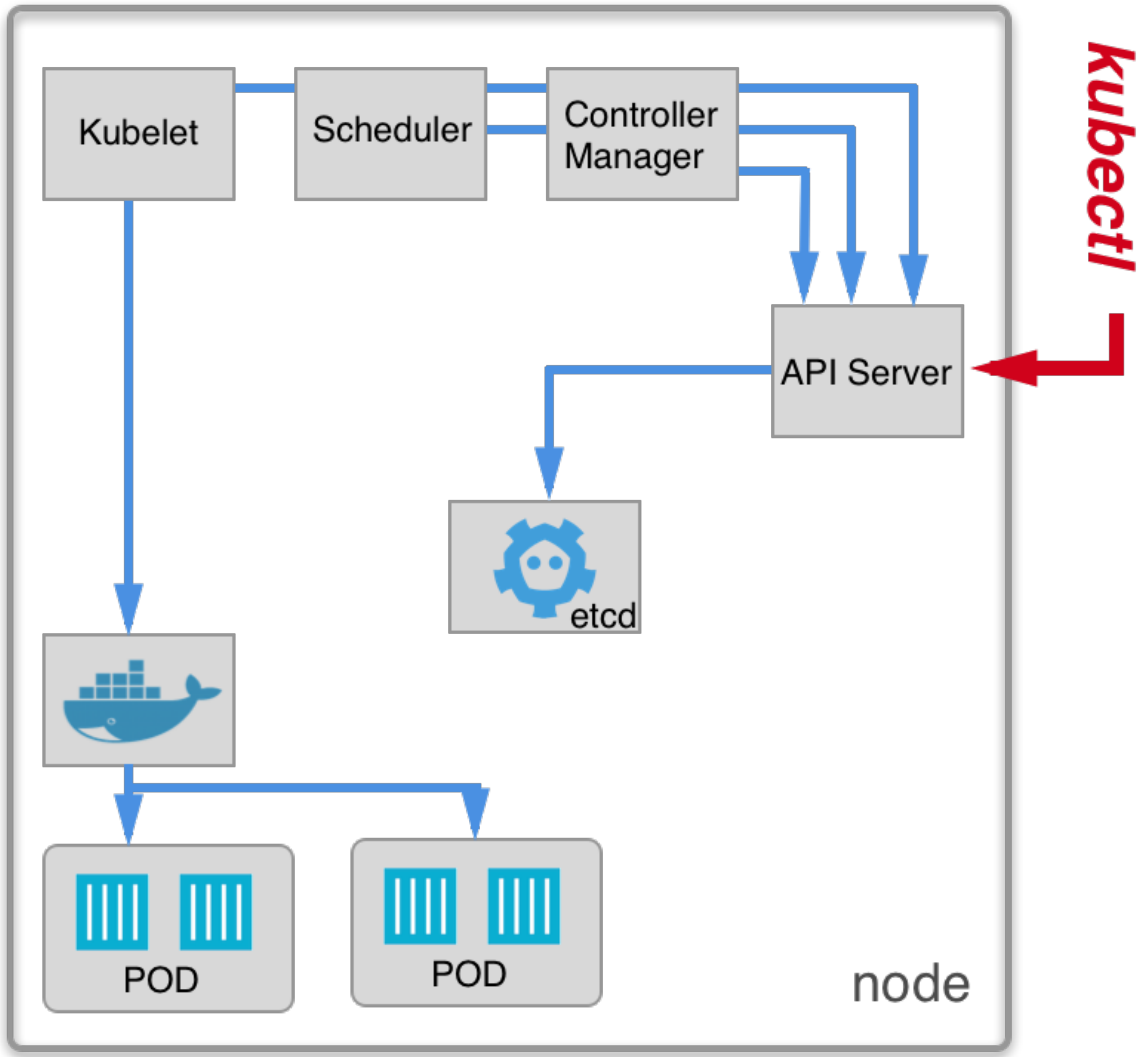
BASIC FLOW

- kubectl apply -f deploy.yaml (5 replicas)
- CM creates replicaset object



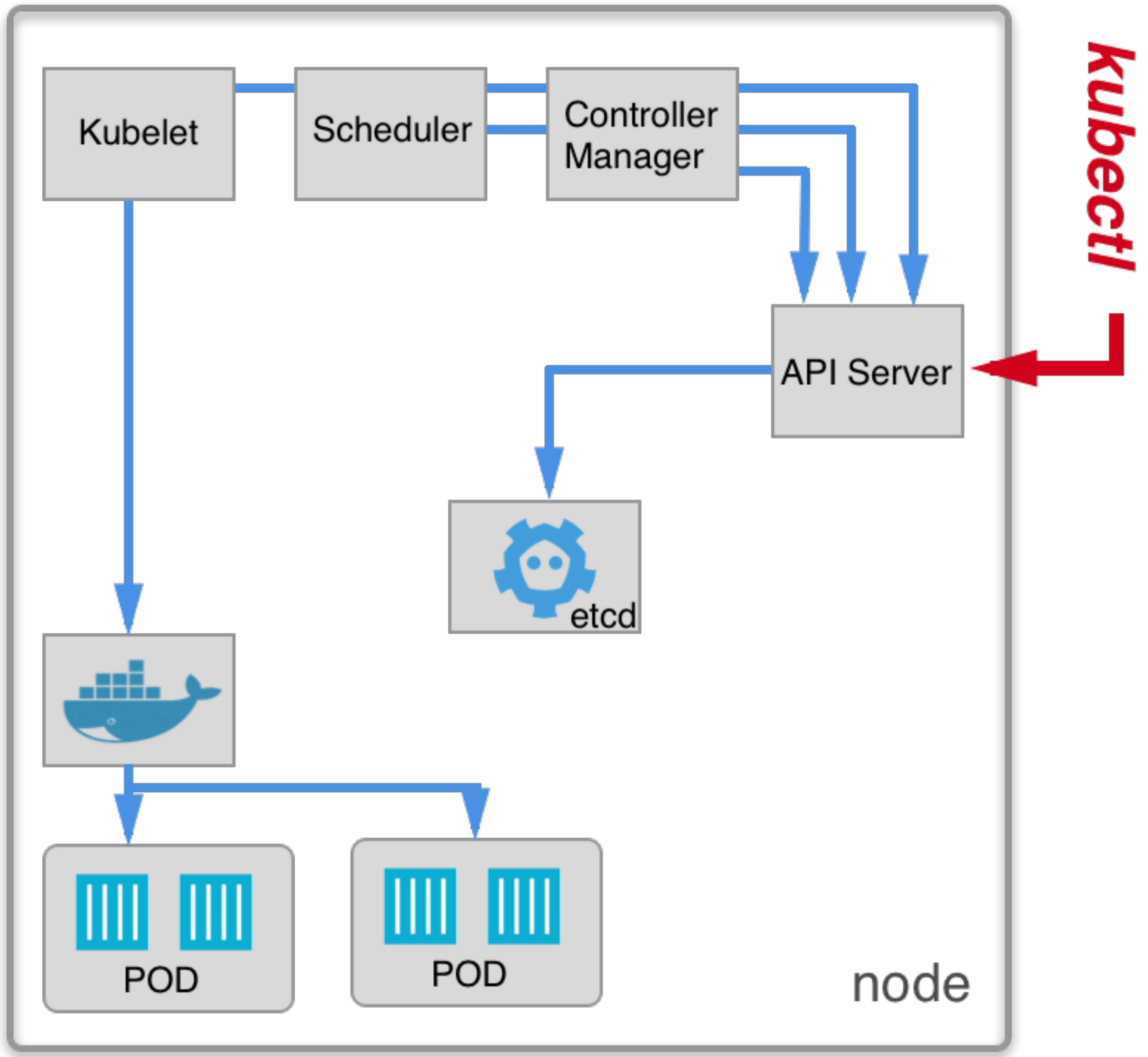
BASIC FLOW

- kubectl apply -f deploy.yaml (5 replicas)
- CM creates replicaset object
- CM creates 5 pods' objects



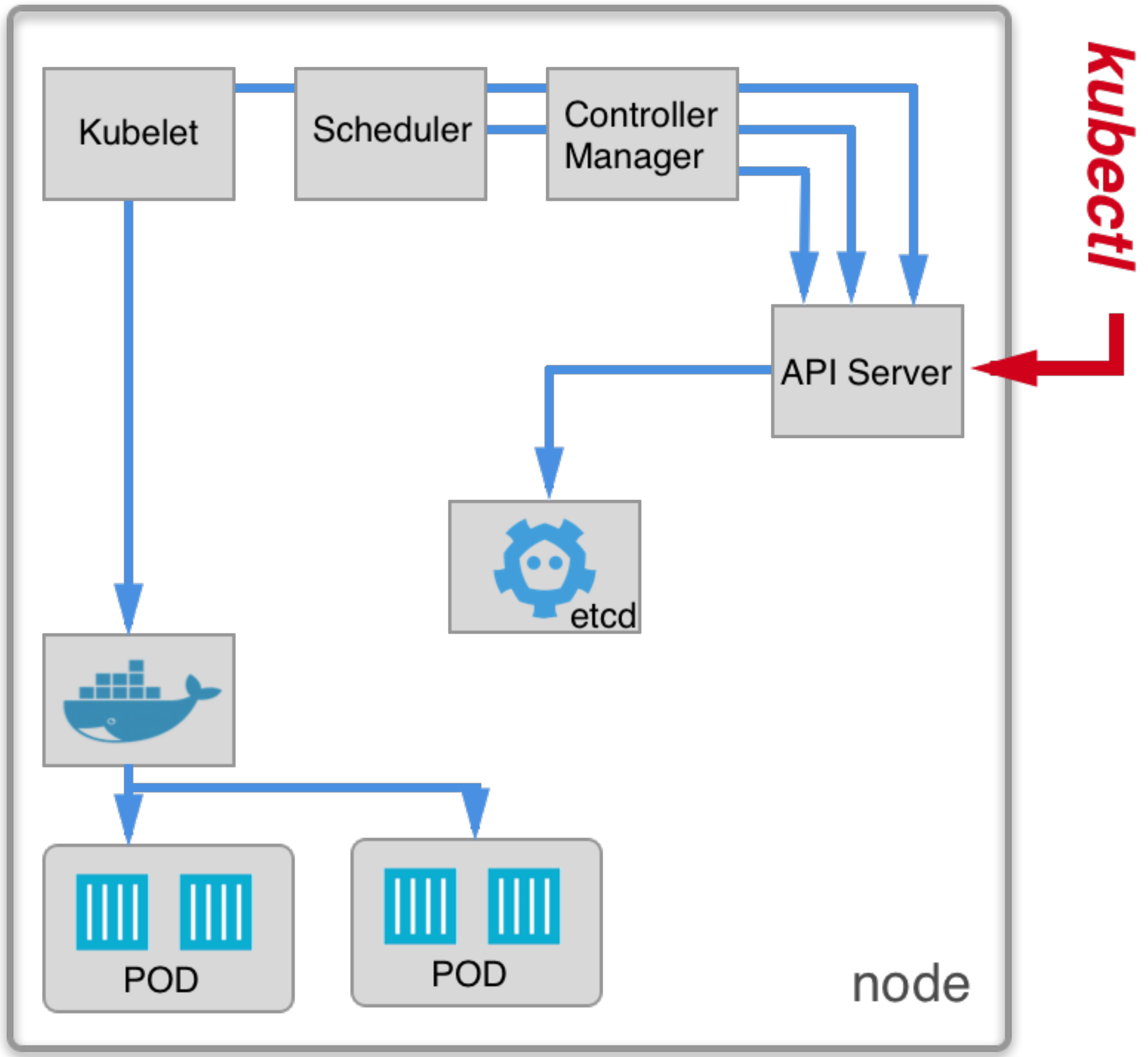
BASIC FLOW

- kubectl apply -f deploy.yaml (5 replicas)
- CM creates replicaset object
- CM creates 5 pods' objects
- Scheduler assigns nodes

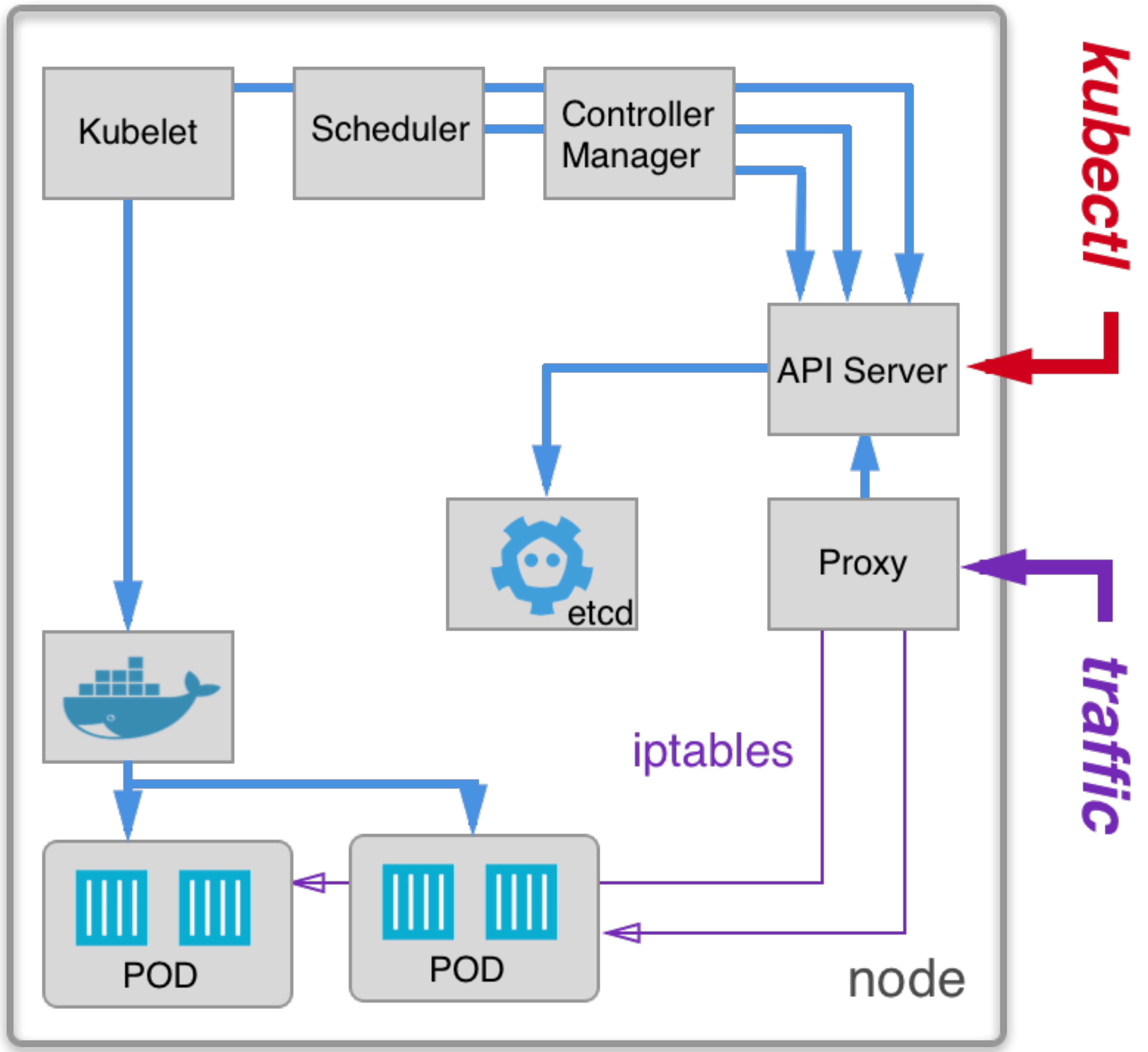


BASIC FLOW

- kubectl apply -f deploy.yaml (5 replicas)
- CM creates replicaset object
- CM creates 5 pods' objects
- Scheduler assigns nodes
- Kubelets create containers

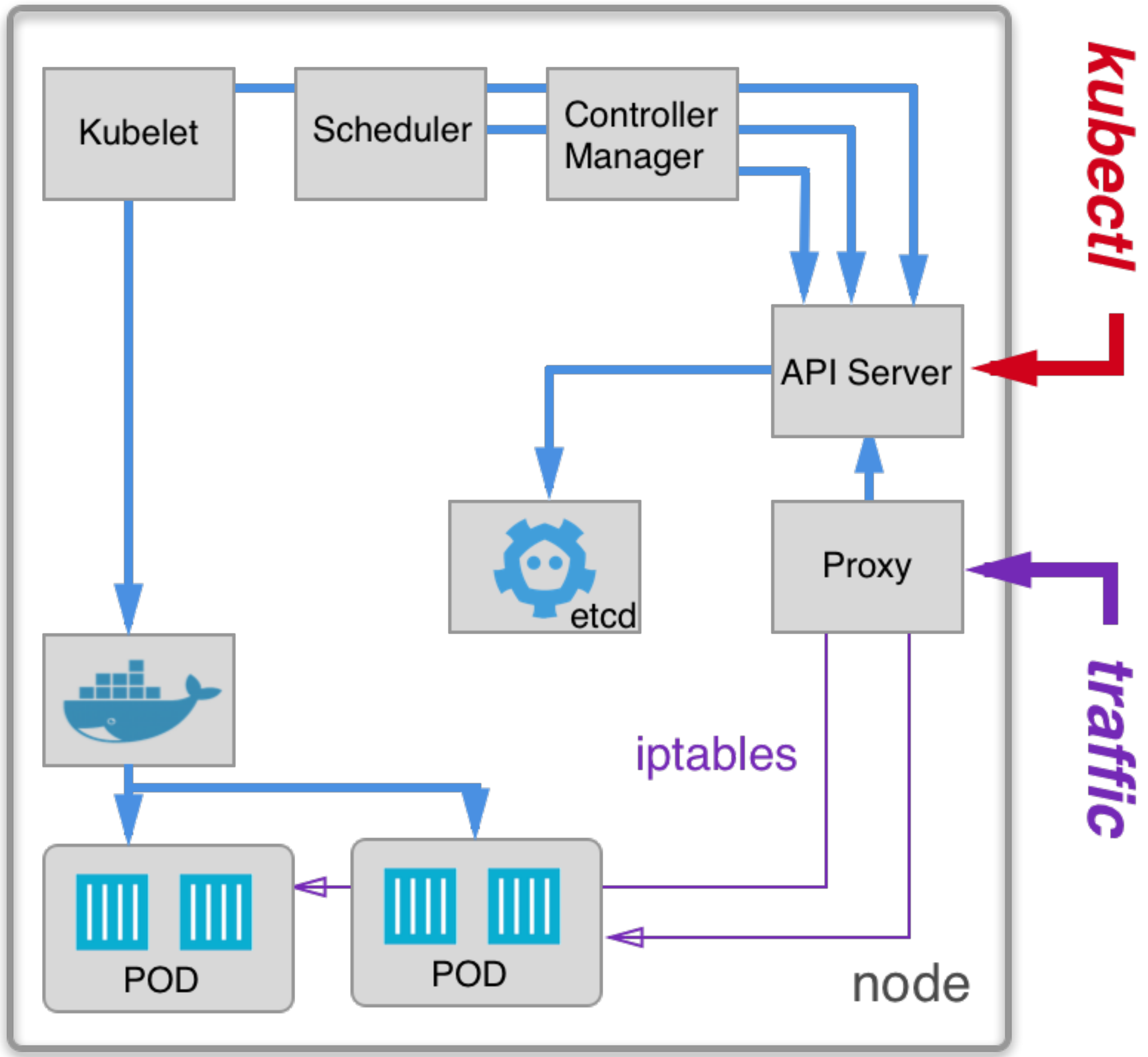


PROXY


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
iptables
POD
POD
node

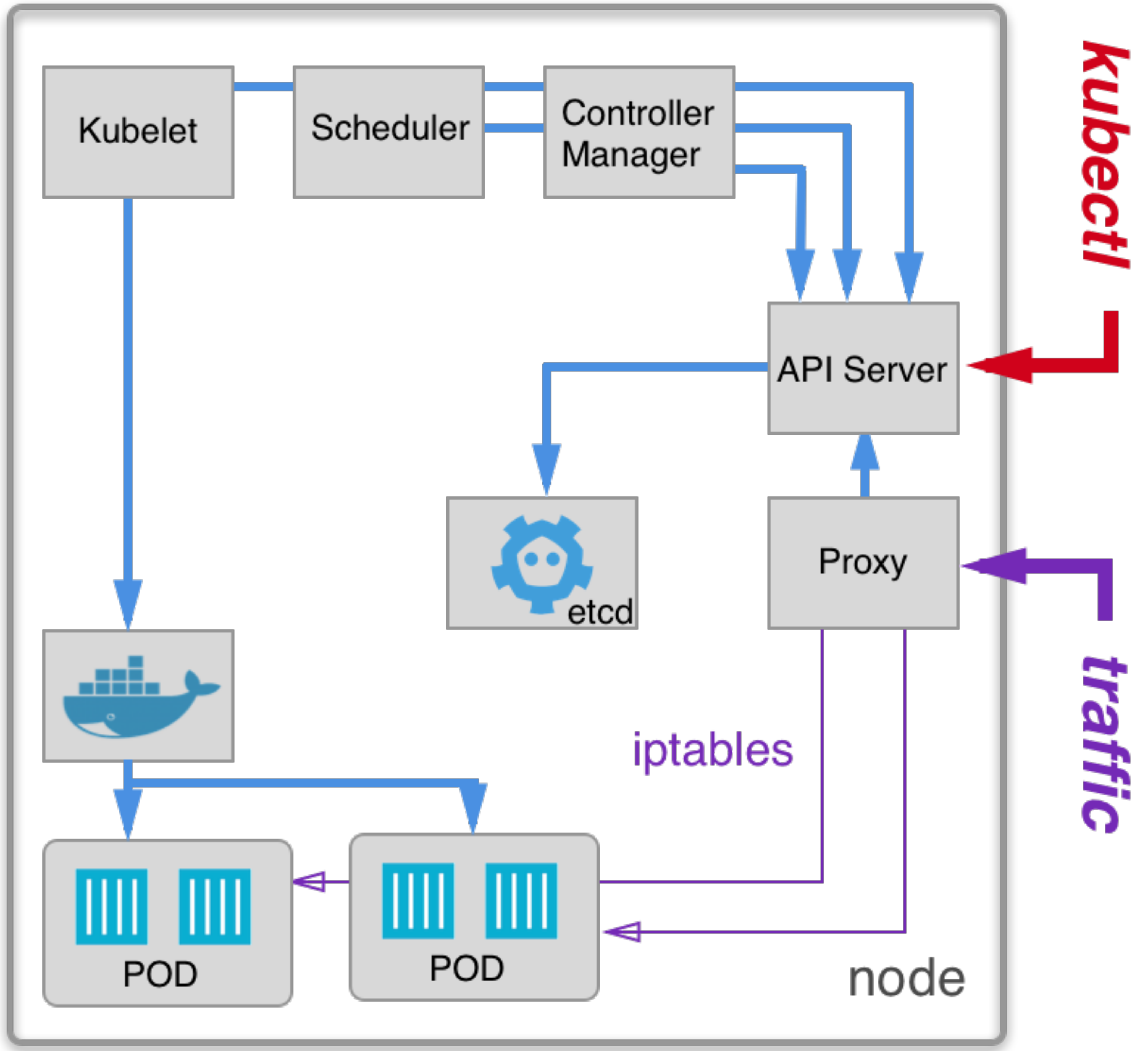
PROXY

- responsible for managing Services



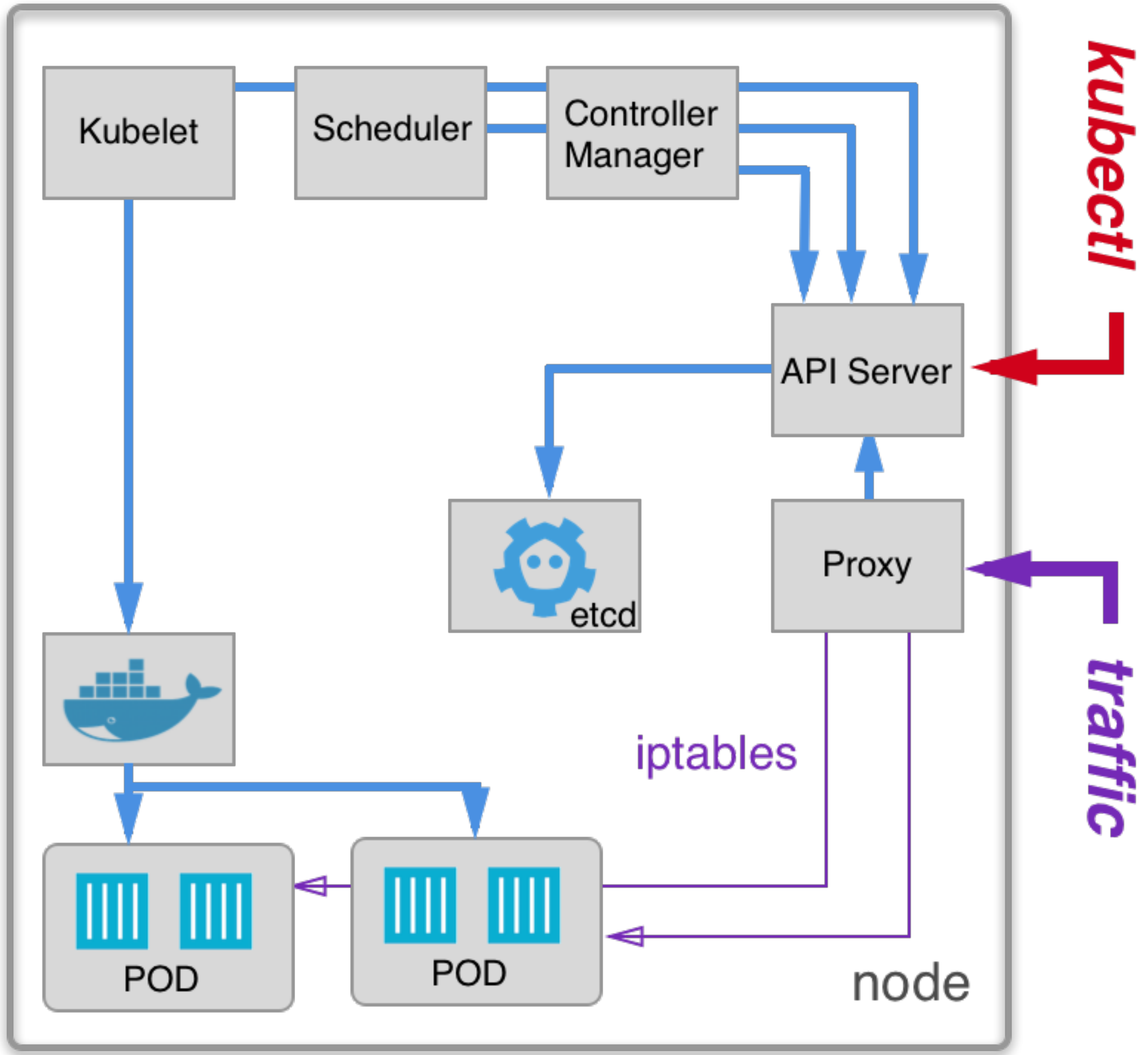
PROXY

- responsible for managing Services
- accepts external traffic on NodePorts and distributes it to pods



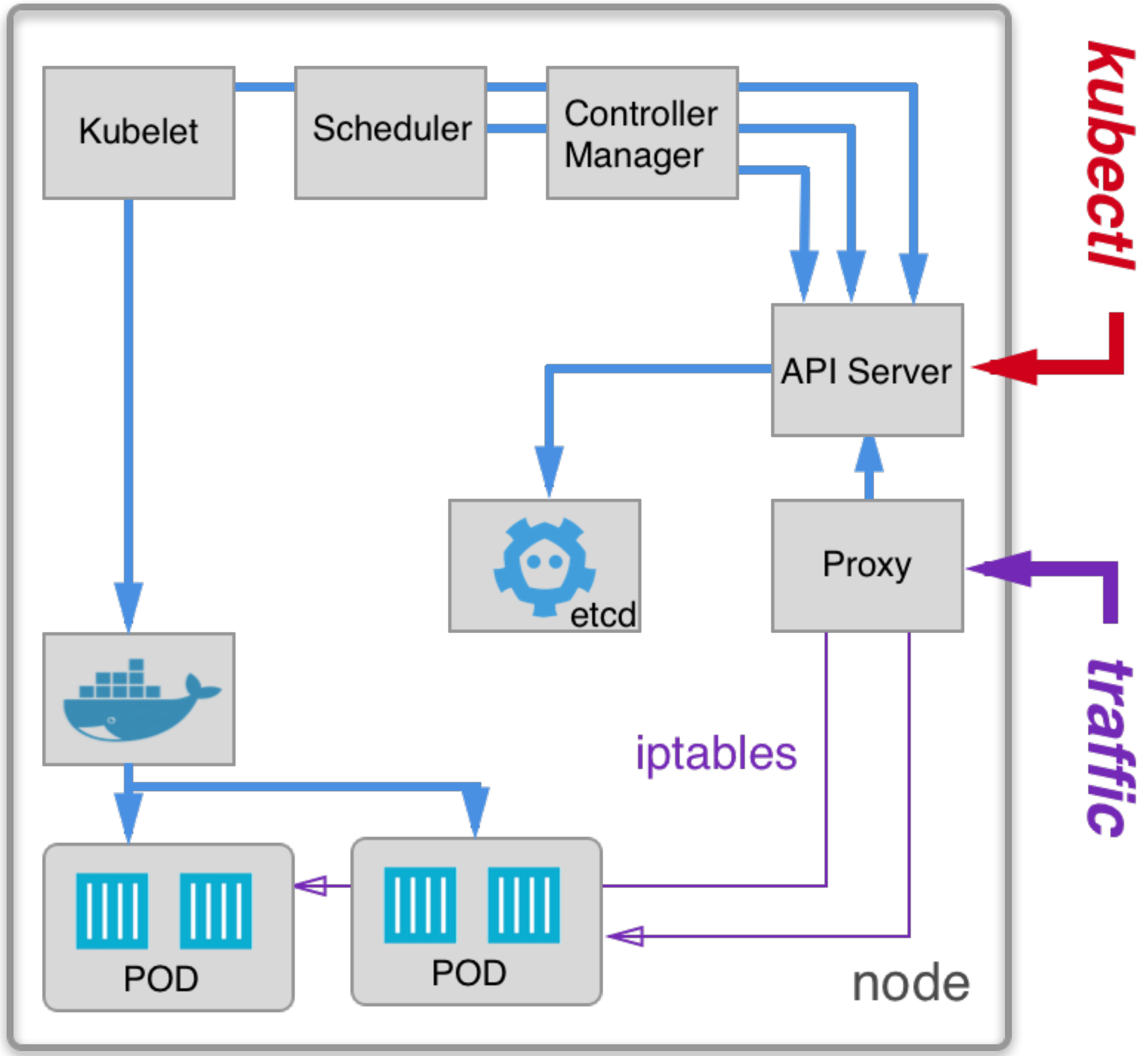
PROXY

- responsible for managing Services
- accepts external traffic on NodePorts and distributes it to pods
- load balances traffic between Service's pods



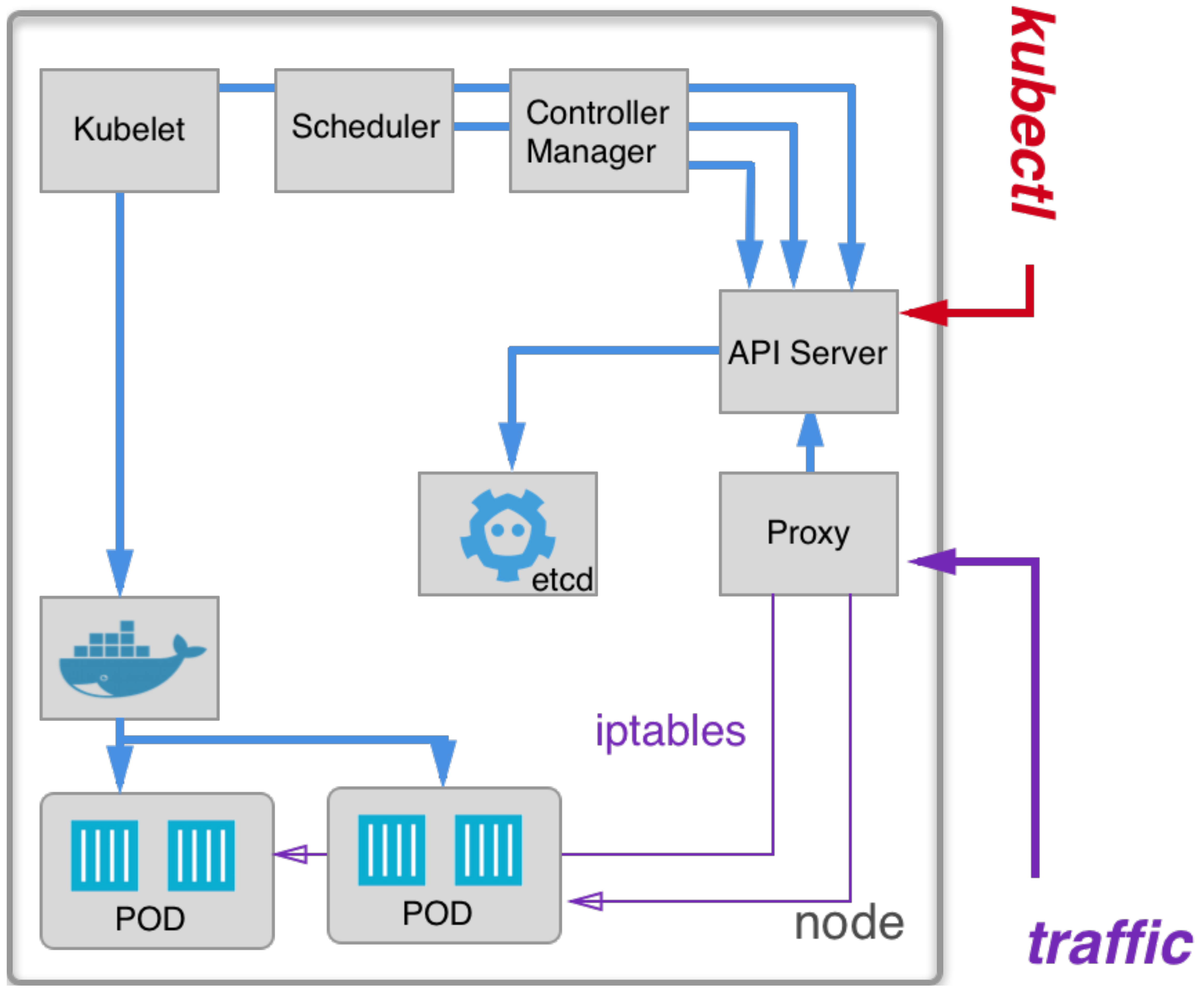
PROXY

- responsible for managing Services
- accepts external traffic on NodePorts and distributes it to pods
- load balances traffic between Service’s pods
- everything is done using *iptables*

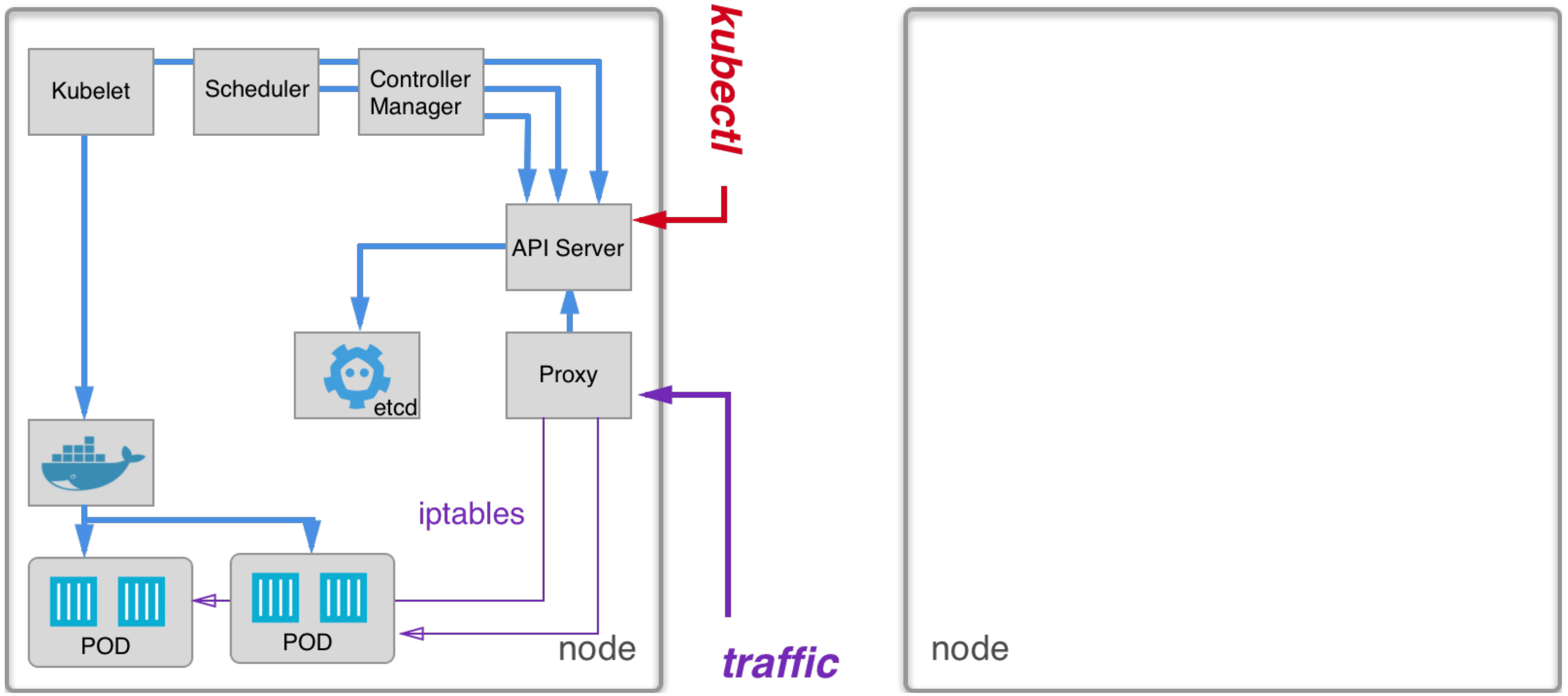


DEMO

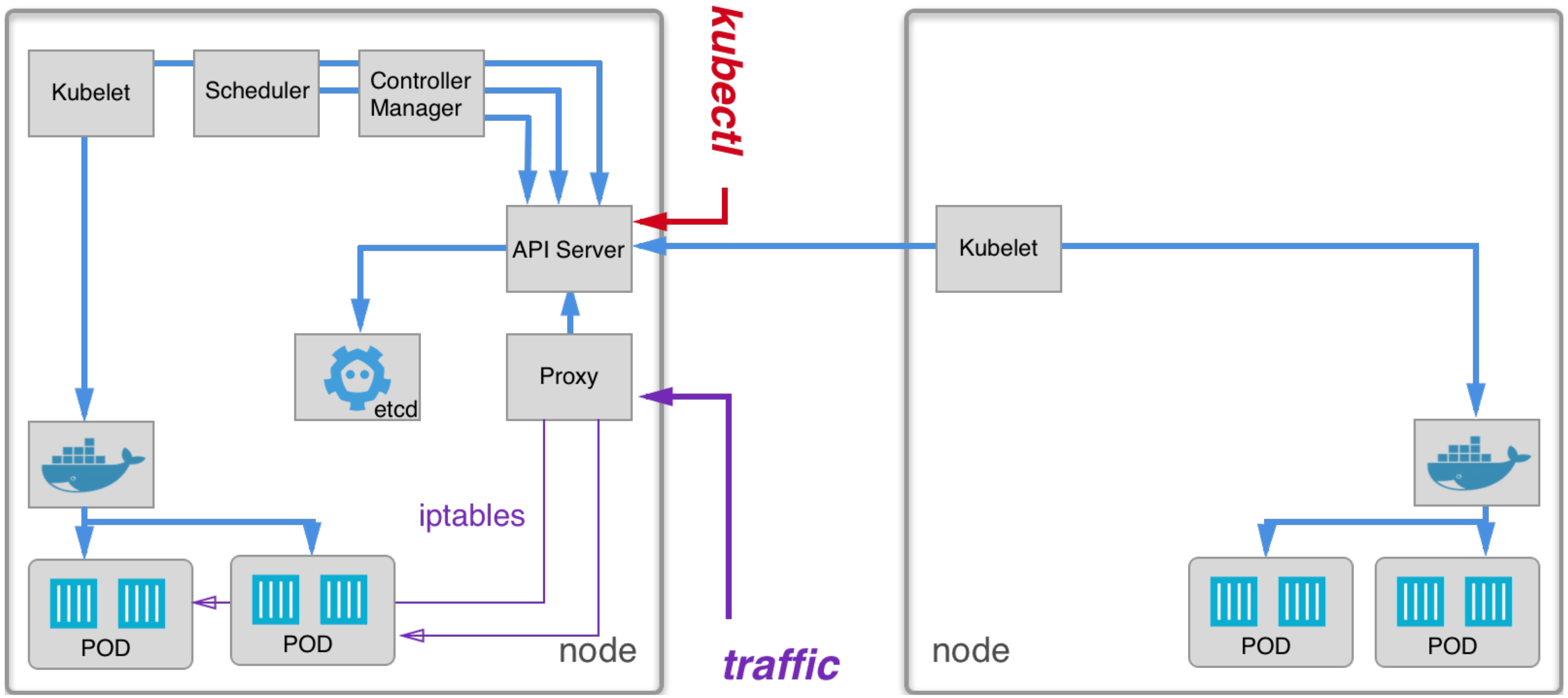
NETWORKING


Kubelet
Scheduler
Controller Manager
API Server
etcd
Proxy
POD
POD
iptables
node
kubectl
traffic

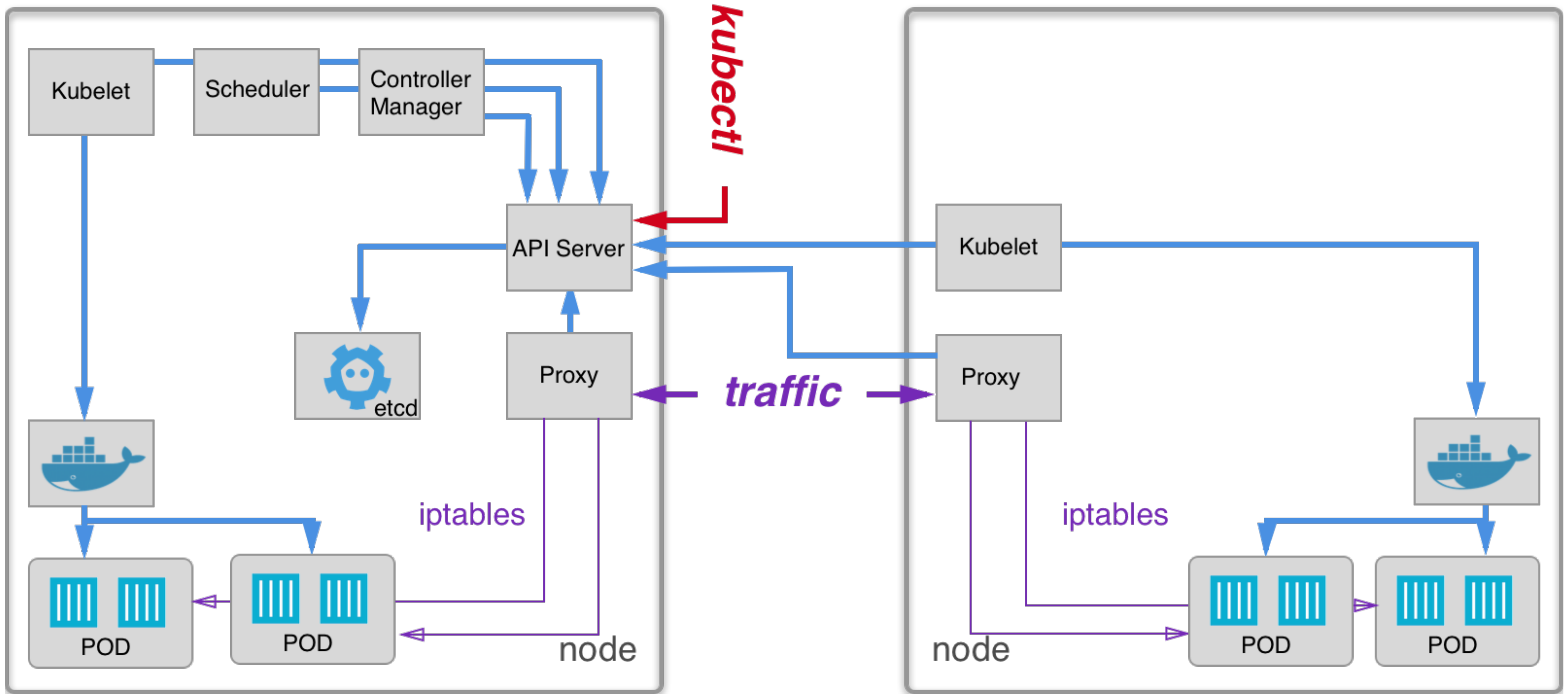
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
iptables
POD
POD
node
traffic
node

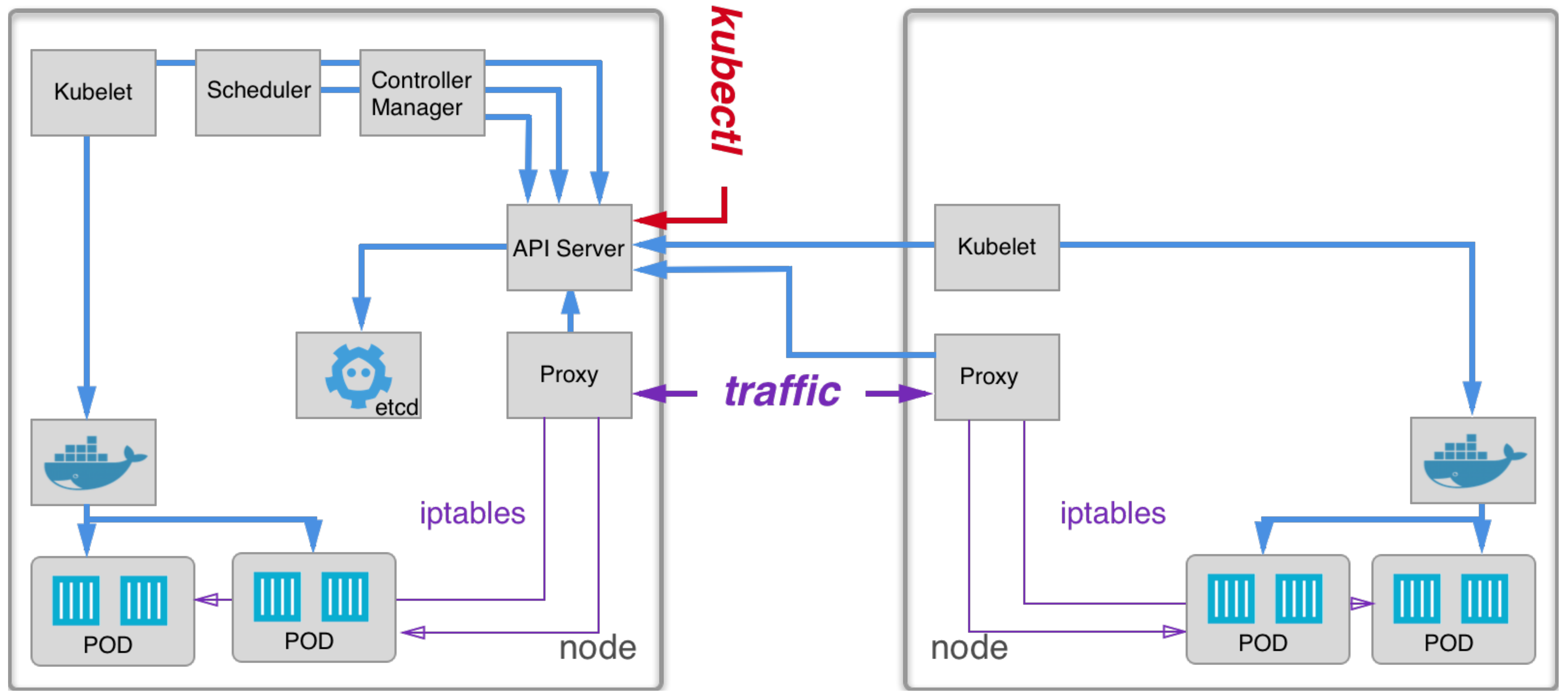
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
Kubelet
etcd
Proxy
iptables
POD
POD
node
traffic
node
POD
POD

NETWORKING

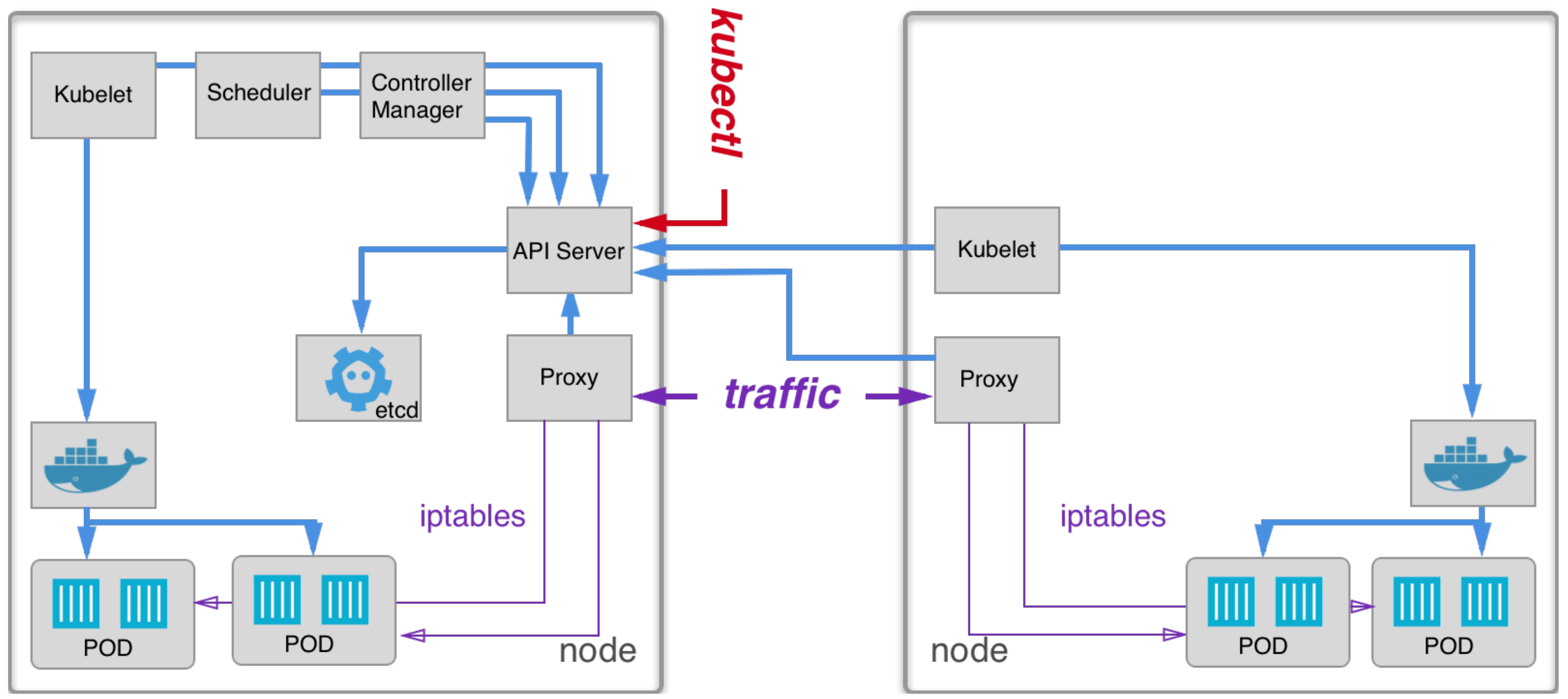

Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
iptables
POD
POD
node
Kubelet
Proxy
iptables
POD
POD
node

NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
Kubelet
etcd
Proxy
traffic
Proxy
iptables
iptables
POD
POD
node
node
POD
POD

NETWORKING

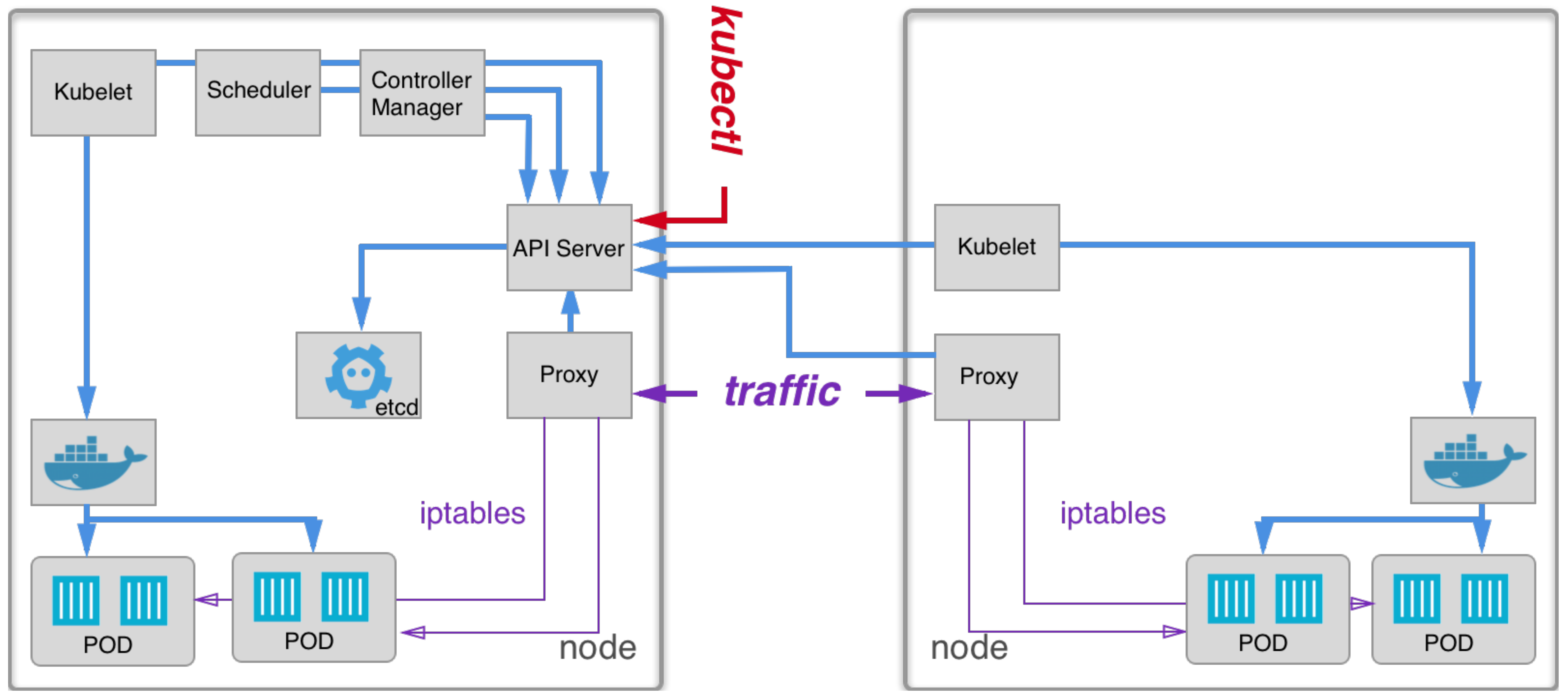
PROBLEMS:



NETWORKING

PROBLEMS:

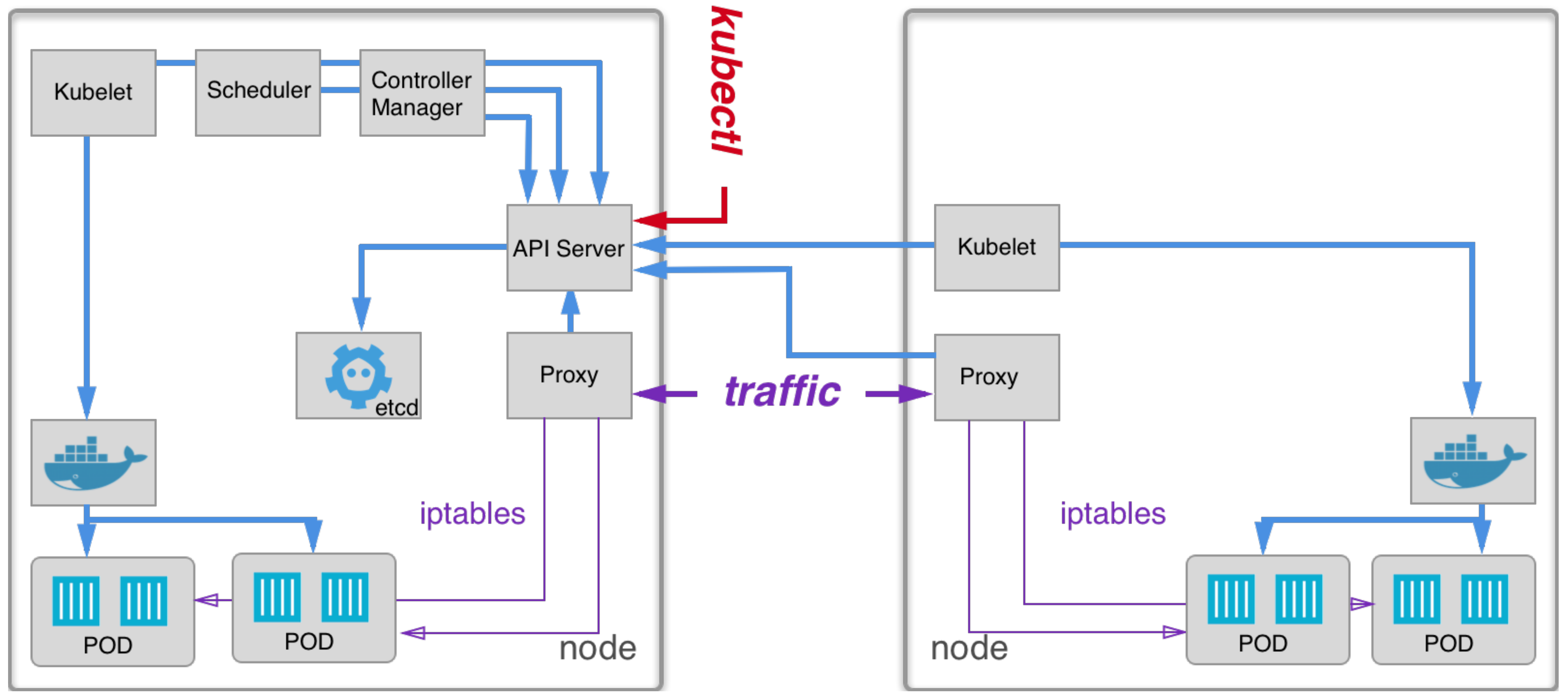
- no POD to POD communication



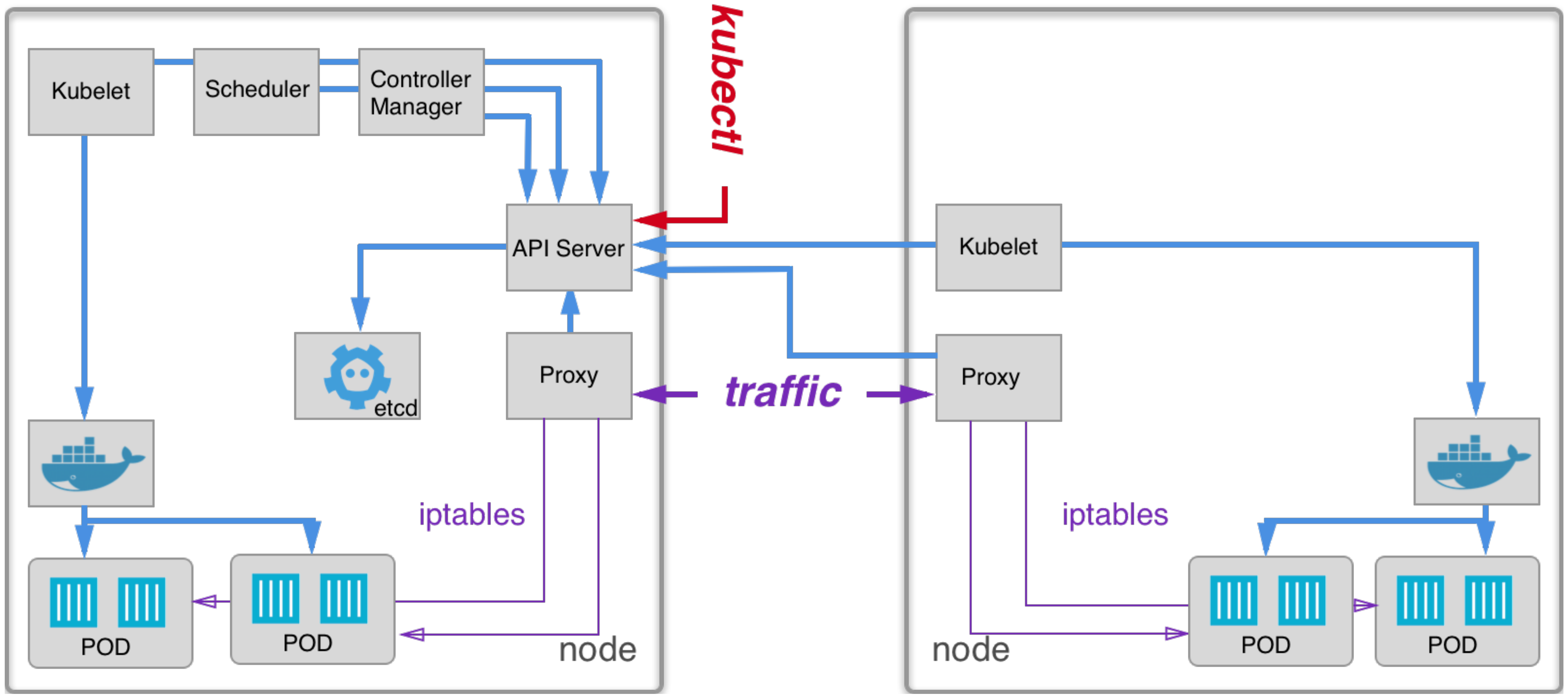
NETWORKING

PROBLEMS:

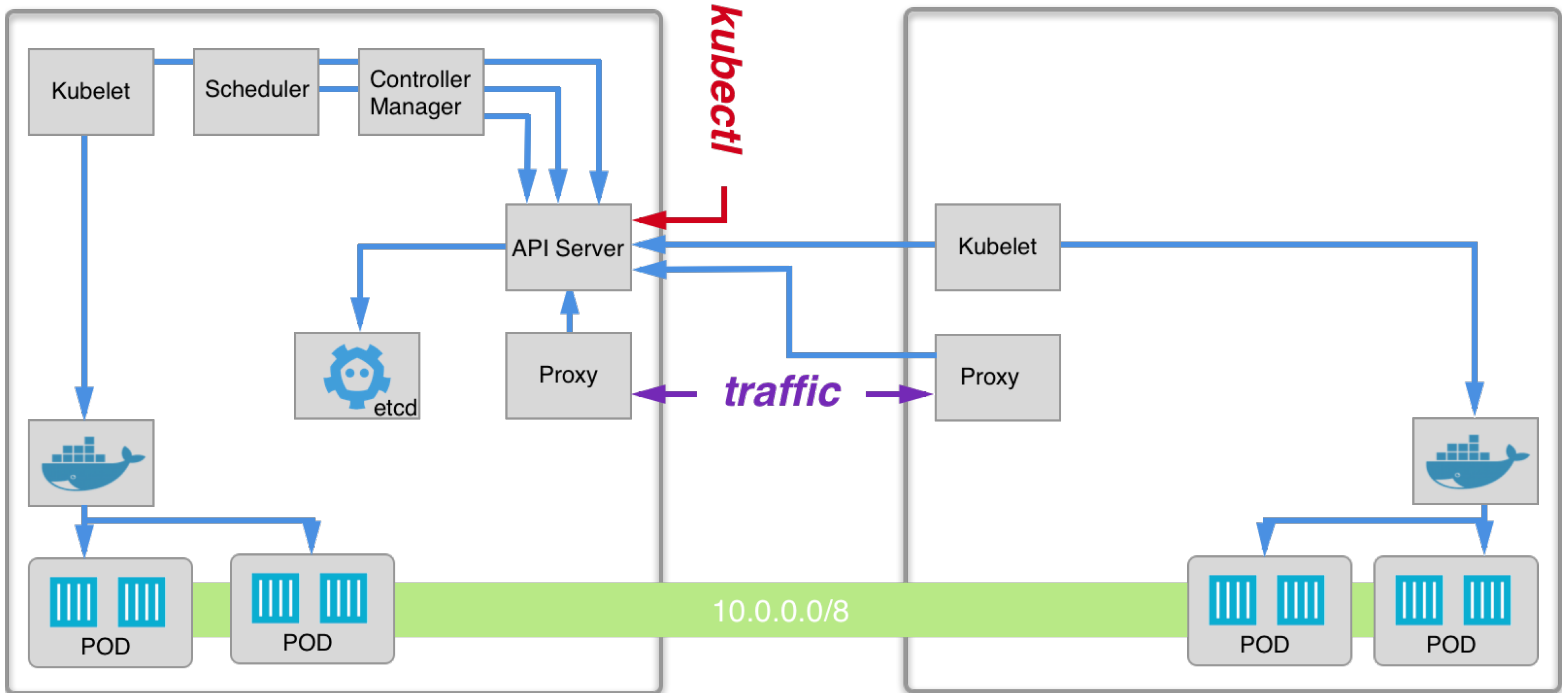
- no POD to POD communication
- traffic cannot be routed between nodes



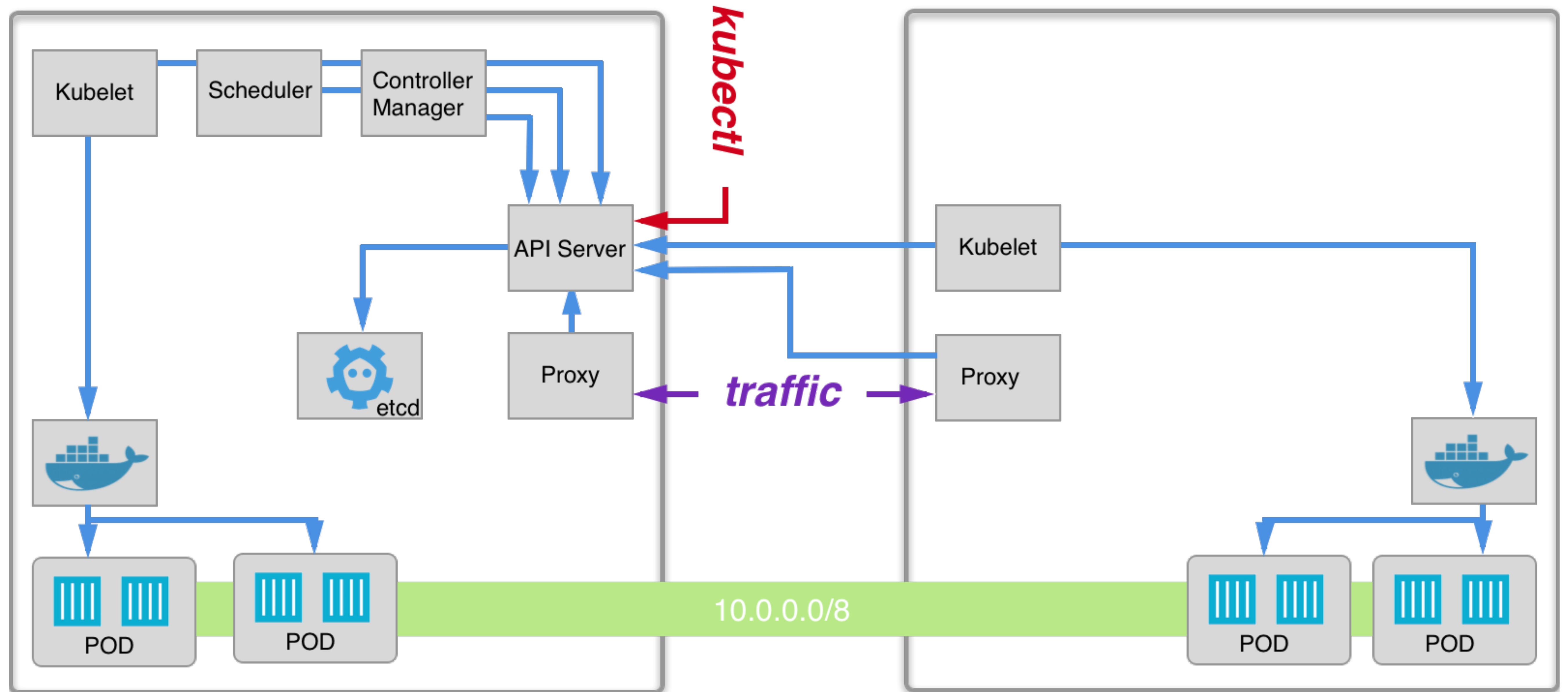
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
iptables
POD
POD
node
Kubelet
Proxy
iptables
POD
POD
node

NETWORKING

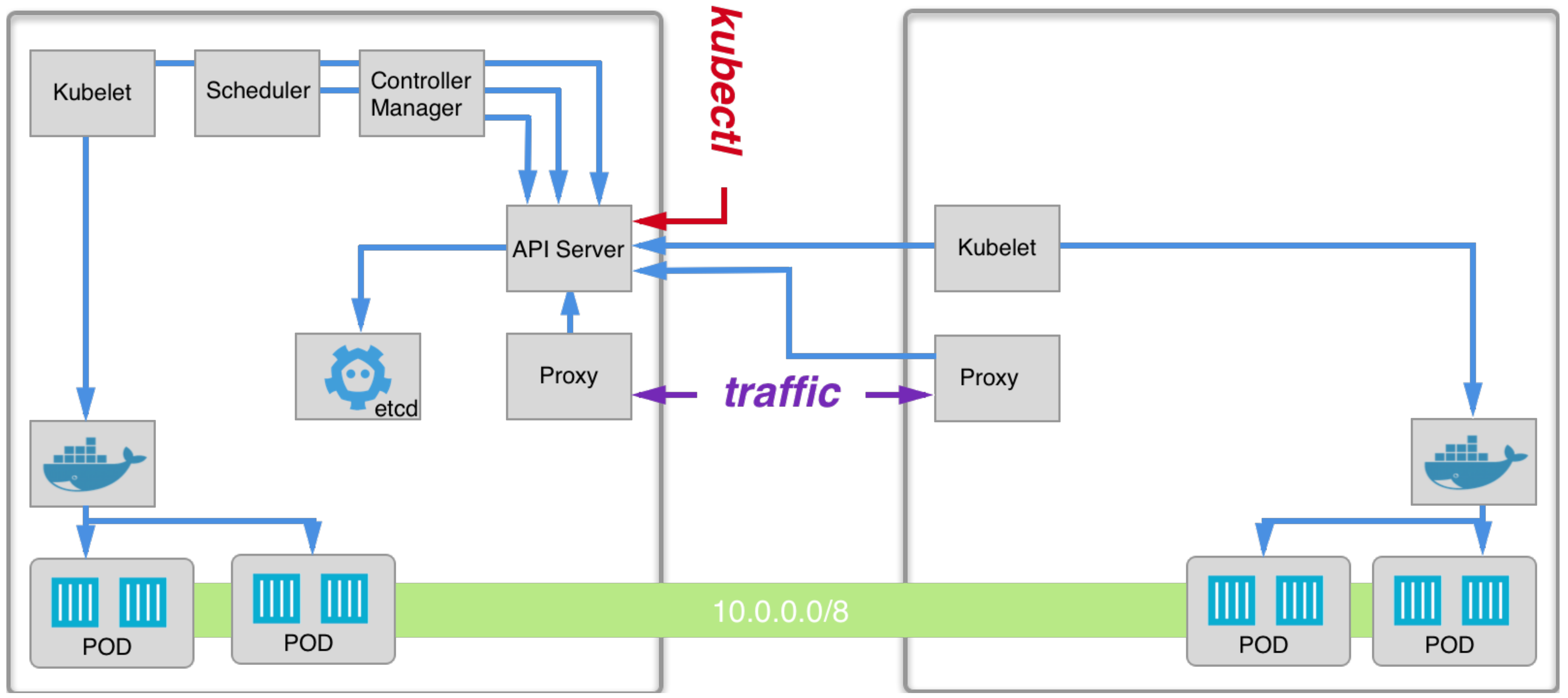

kubectl
Kubelet
Scheduler
Controller Manager
API Server
etcd
Proxy
traffic
Kubelet
Proxy
POD
POD
10.0.0.0/8
POD
POD

NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
Kubelet
Proxy
POD
POD
10.0.0.0/8
POD
POD

NETWORKING

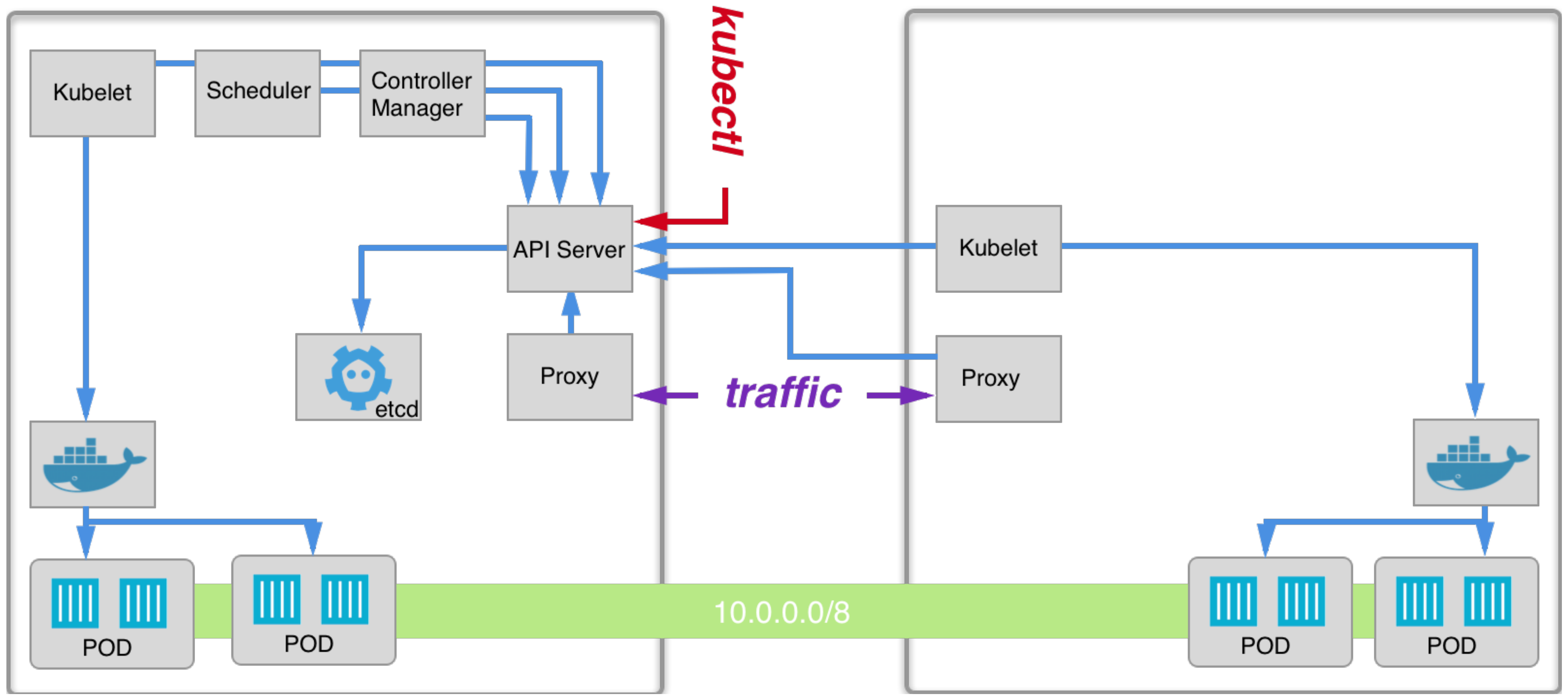
Kubernetes requirements:



NETWORKING

Kubernetes requirements:

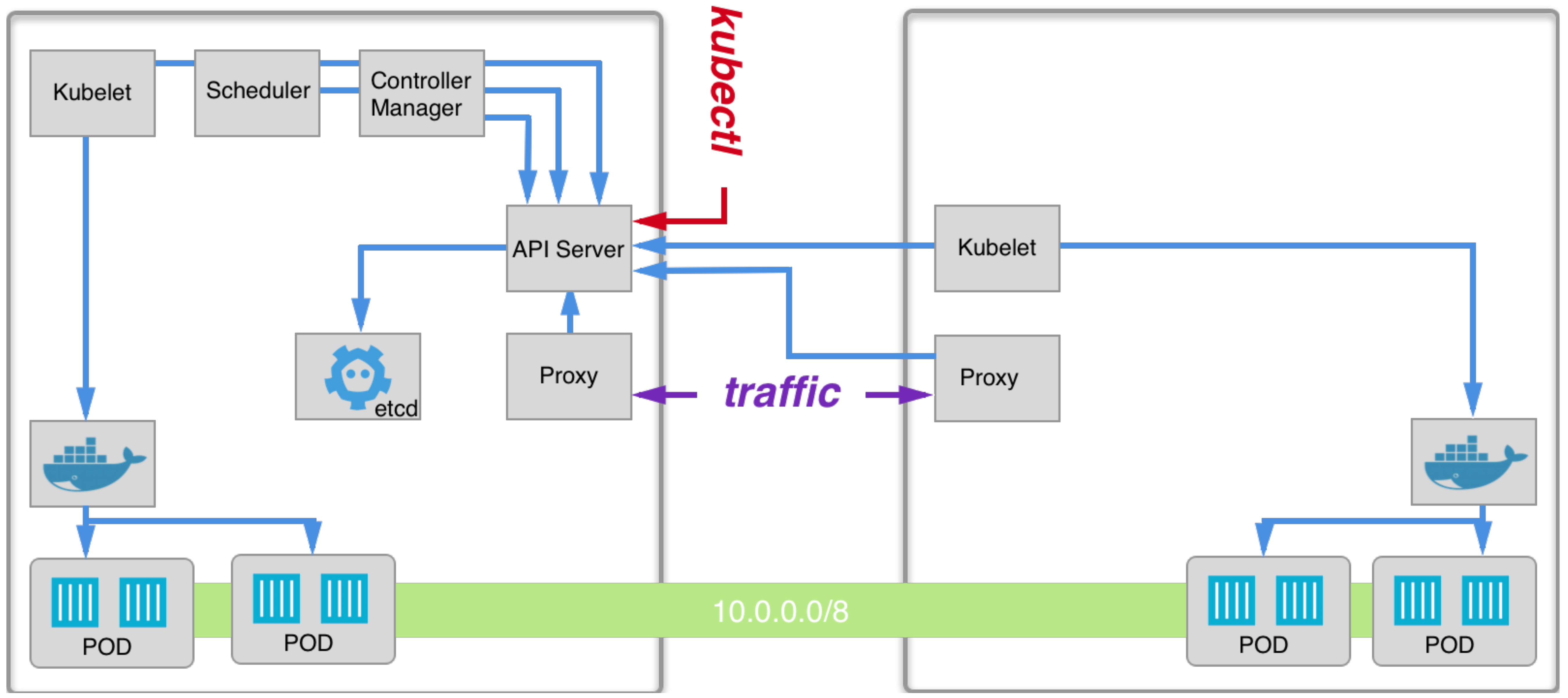
- each pod has it’s own IP



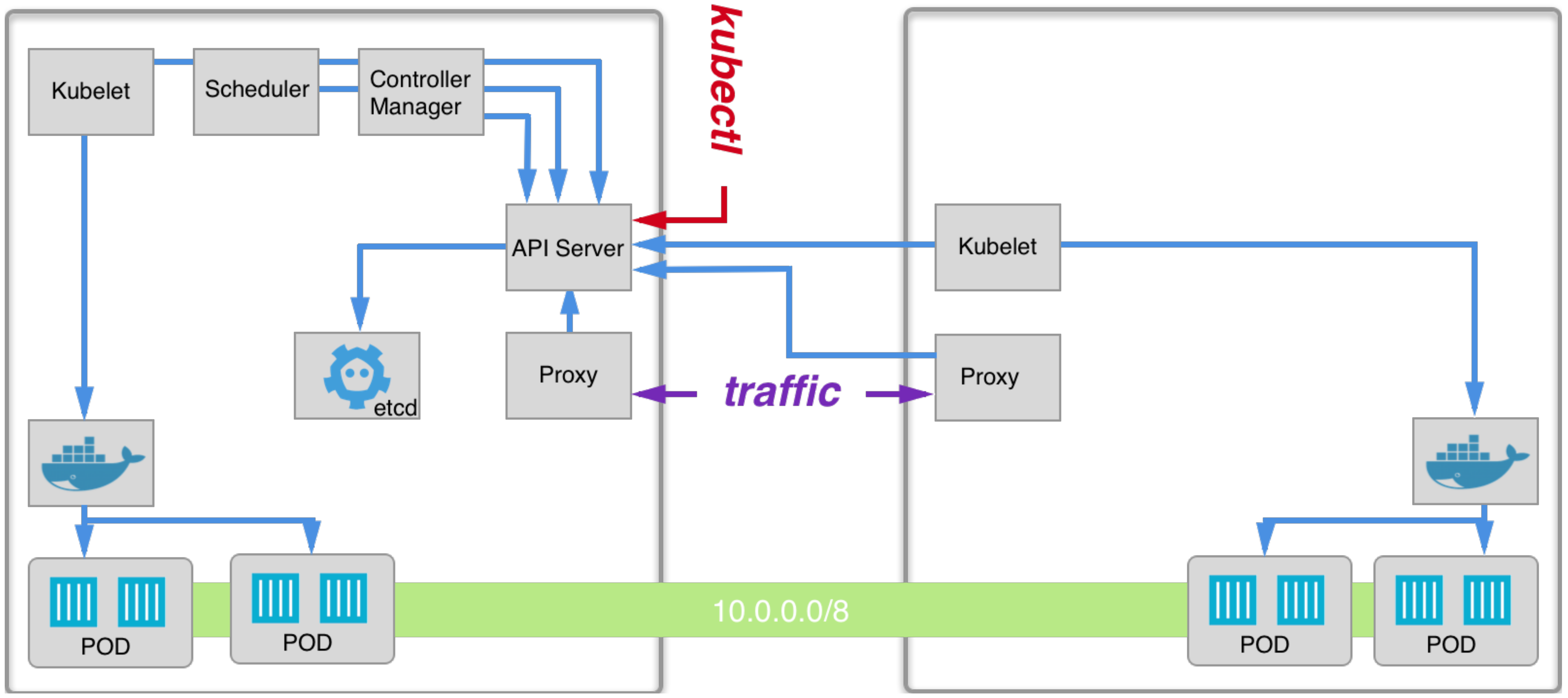
NETWORKING

Kubernetes requirements:

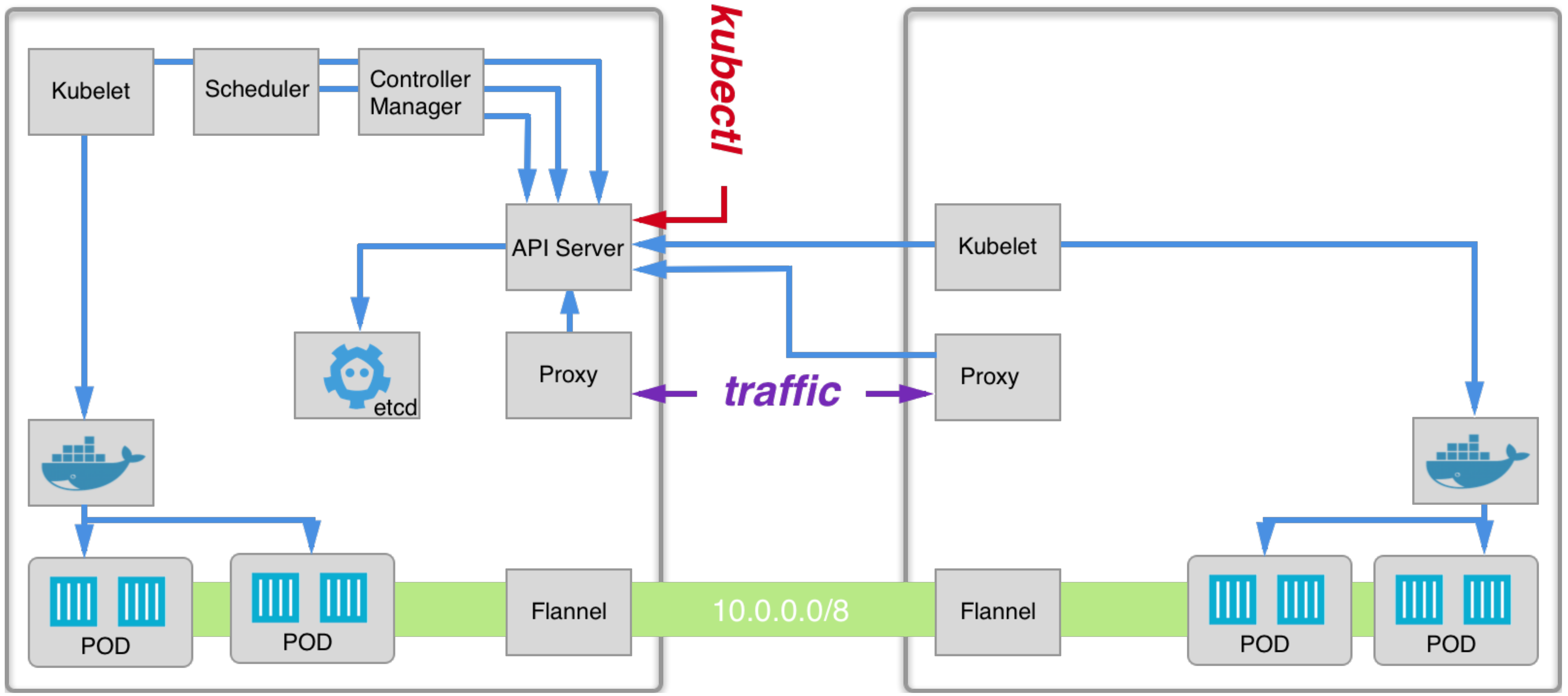
- each pod has it’s own IP
- all containers can communicate with all other containers using their IPs



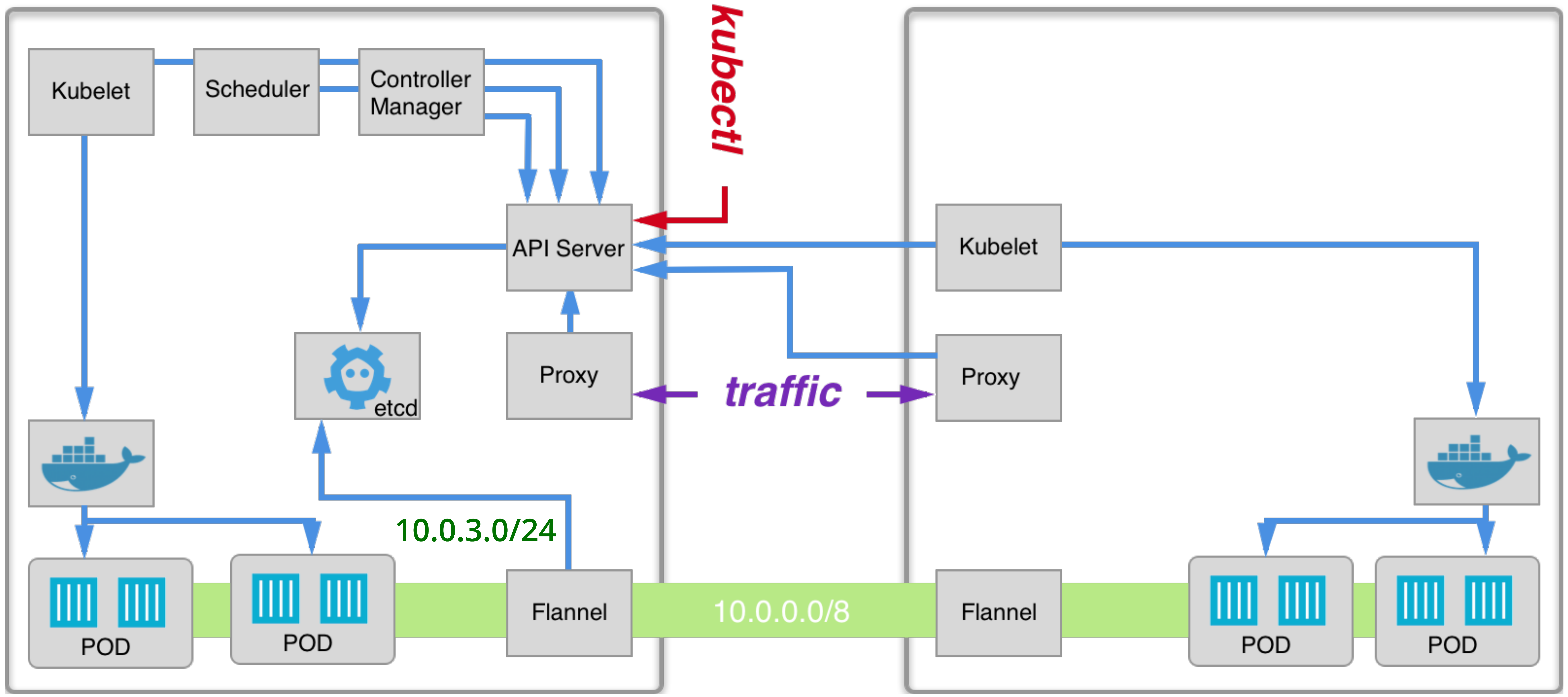
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
Kubelet
Proxy
POD
POD
10.0.0.0/8
POD
POD

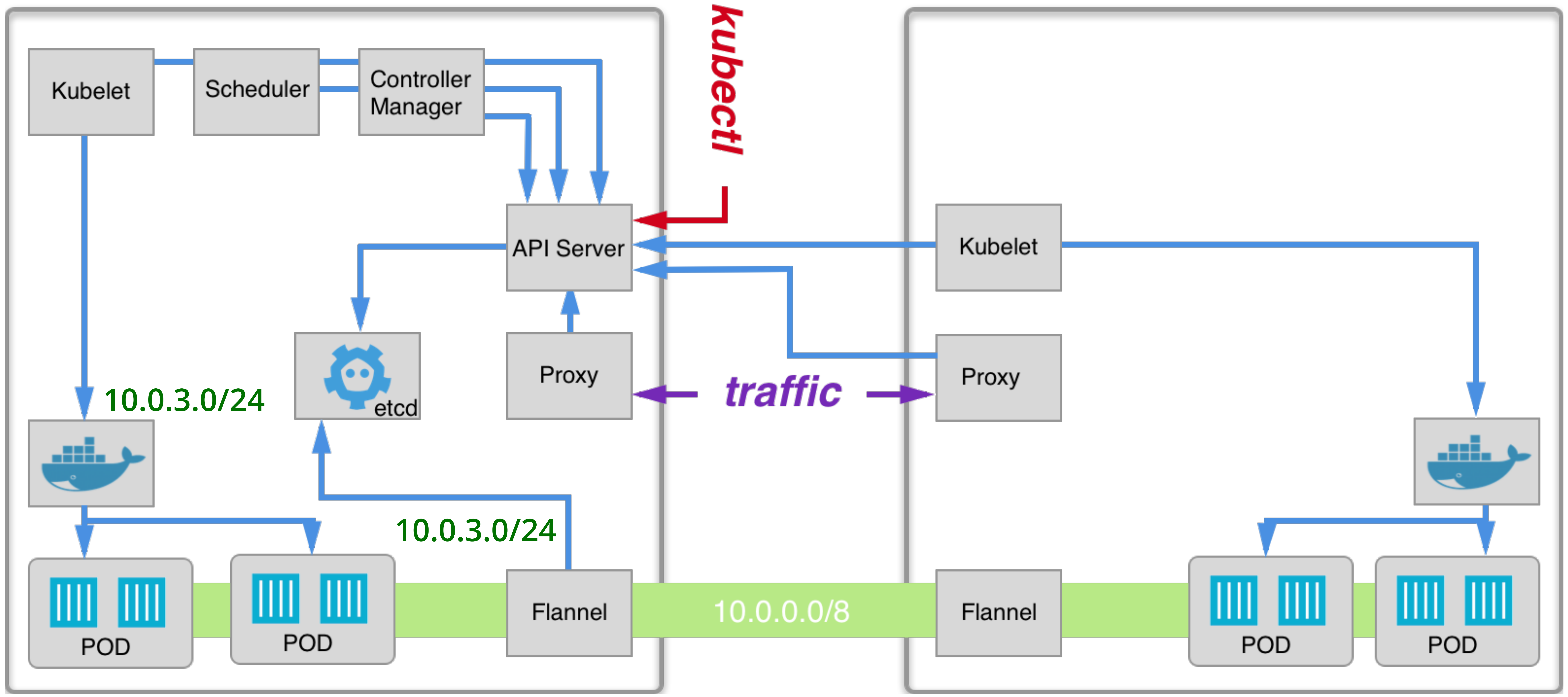
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
Kubelet
Proxy
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD

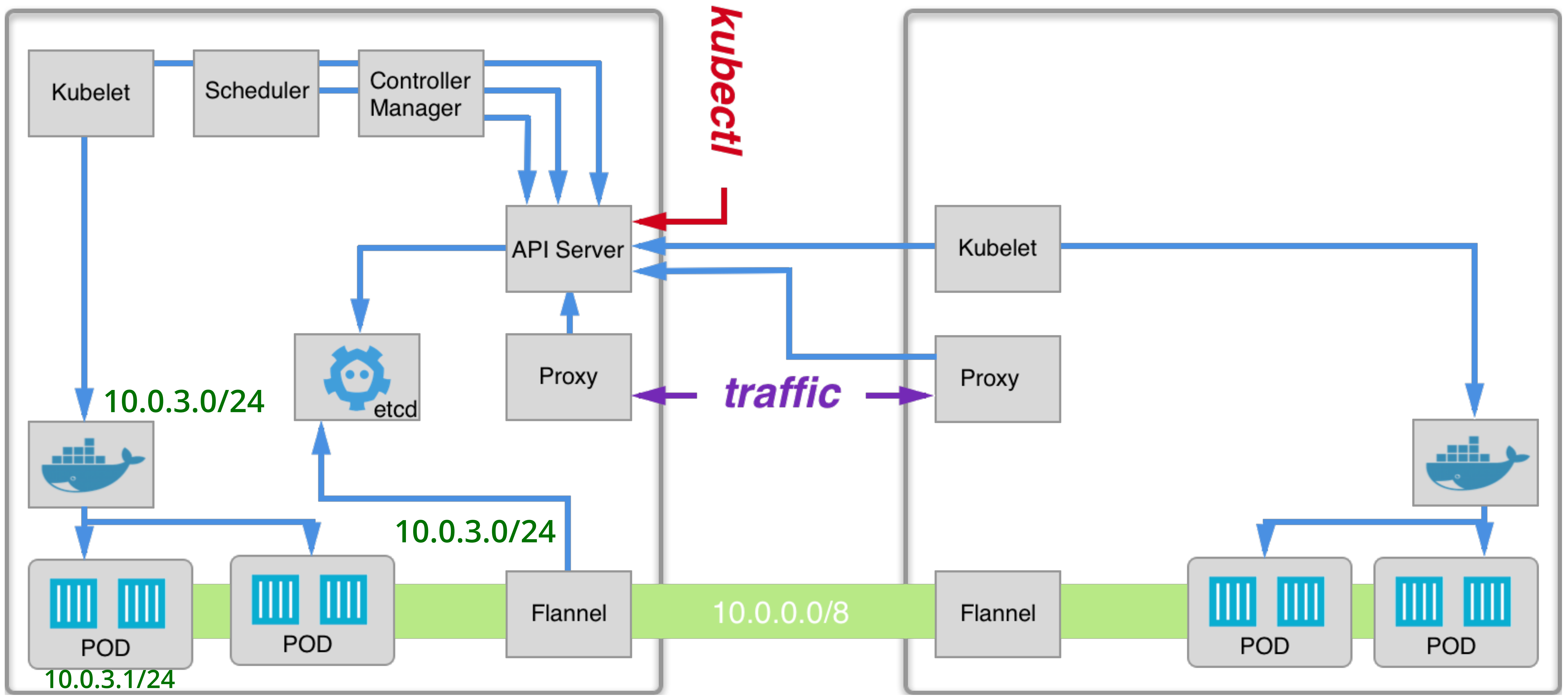
NETWORKING


kubectl
Kubelet
Scheduler
Controller Manager
API Server
etcd
Proxy
traffic
10.0.3.0/24
POD
POD
Flannel
10.0.0.0/8
Kubelet
Proxy
Flannel
POD
POD

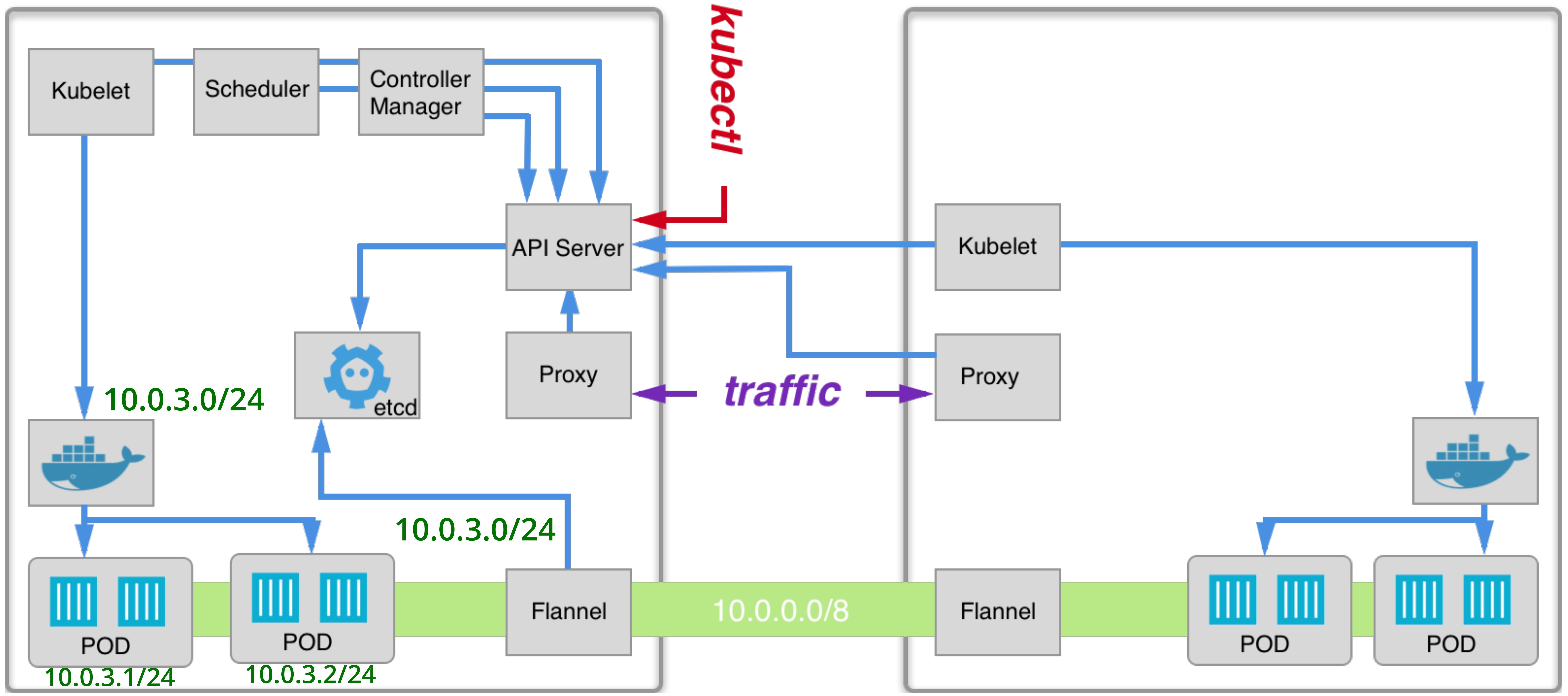
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
Kubelet
etcd
Proxy
traffic
Proxy
10.0.3.0/24
10.0.3.0/24
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD

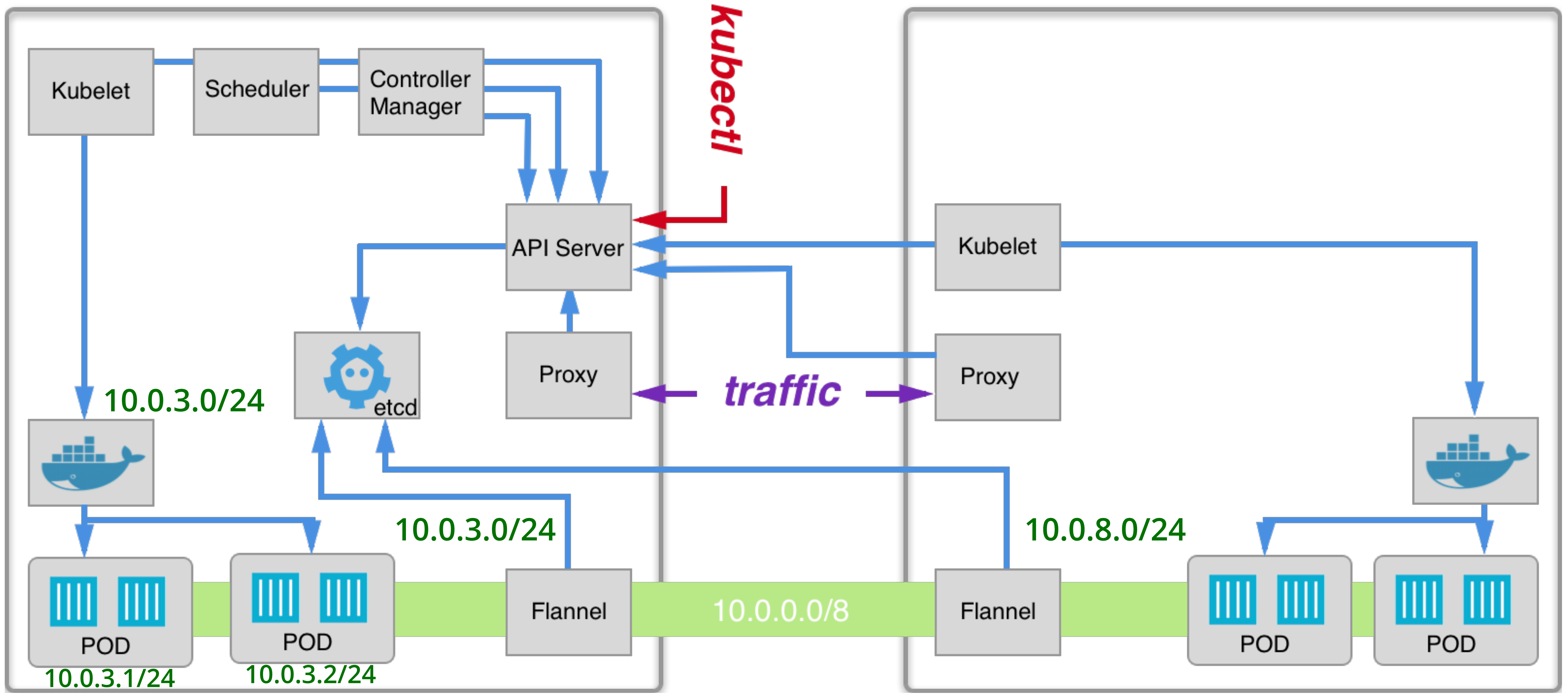
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
10.0.3.0/24
10.0.3.0/24
POD
POD
10.0.3.1/24
Flannel
10.0.0.0/8
Kubelet
Proxy
Flannel
POD
POD

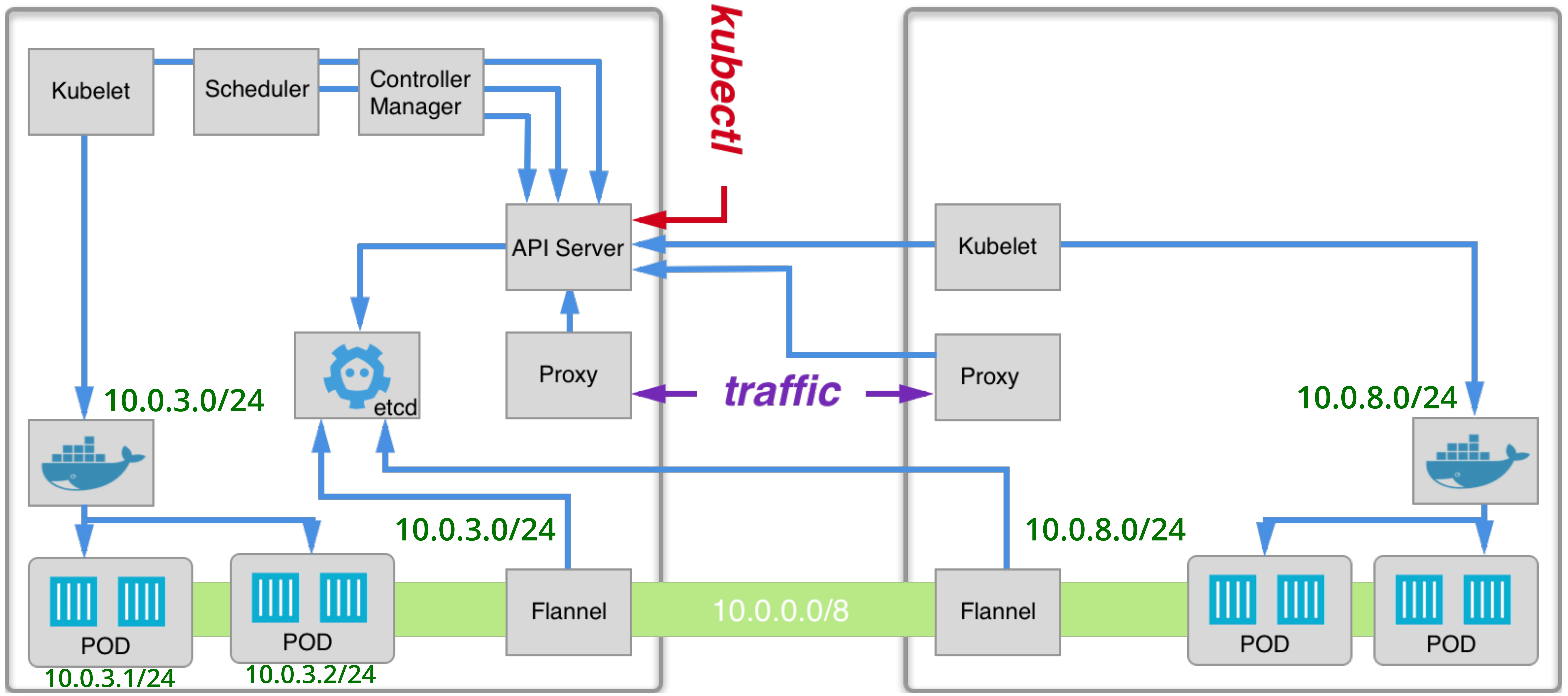
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
10.0.3.0/24
10.0.3.0/24
POD
POD
10.0.3.1/24
10.0.3.2/24
Flannel
10.0.0.0/8
Kubelet
Proxy
Flannel
POD
POD

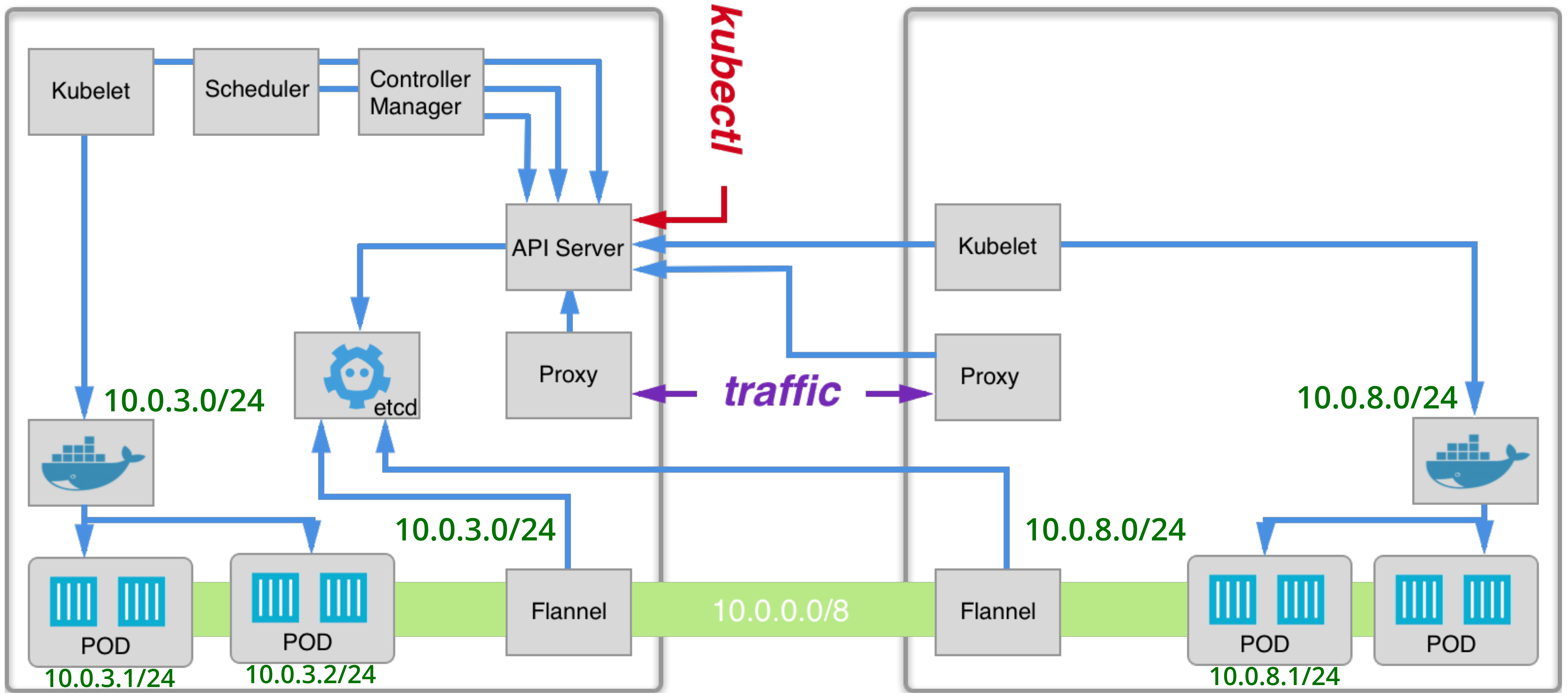
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
10.0.3.0/24
10.0.3.0/24
POD
POD
10.0.3.1/24
10.0.3.2/24
Flannel
10.0.0.0/8
Kubelet
Proxy
10.0.8.0/24
Flannel
POD
POD

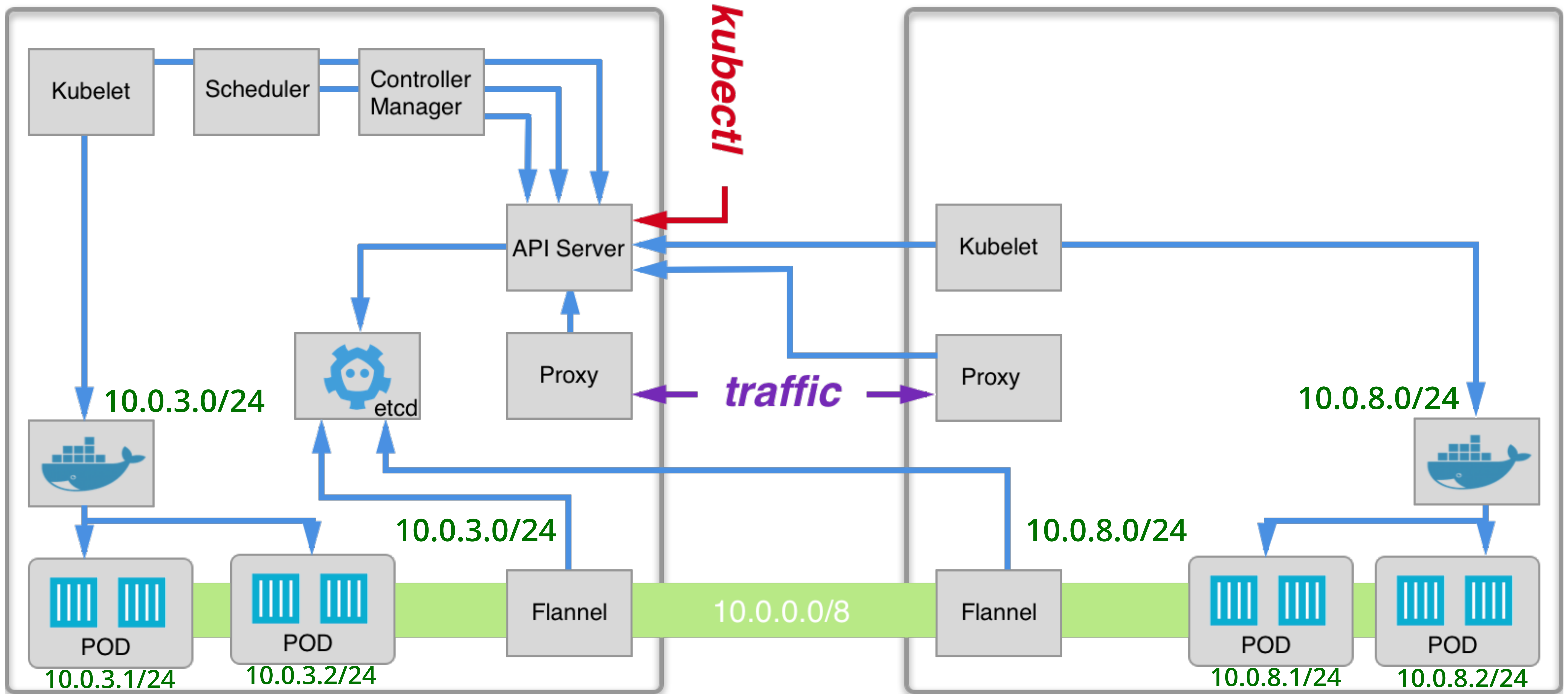
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
10.0.3.0/24
traffic
10.0.3.0/24
POD
POD
10.0.3.1/24
10.0.3.2/24
Flannel
10.0.0.0/8
Kubelet
Proxy
10.0.8.0/24
10.0.8.0/24
Flannel
POD
POD

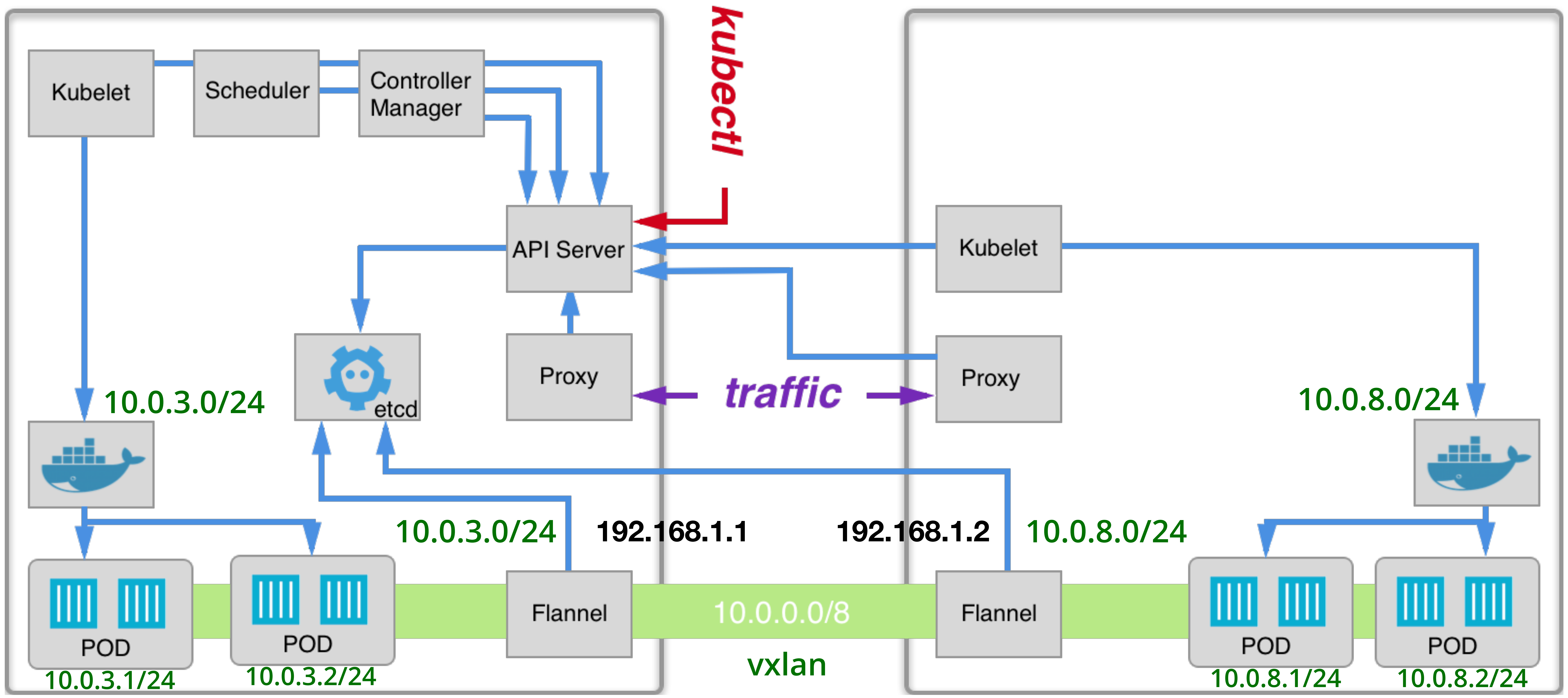
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
Proxy
Kubelet
10.0.3.0/24
10.0.8.0/24
10.0.3.0/24
10.0.8.0/24
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD
10.0.3.1/24
10.0.3.2/24
10.0.8.1/24

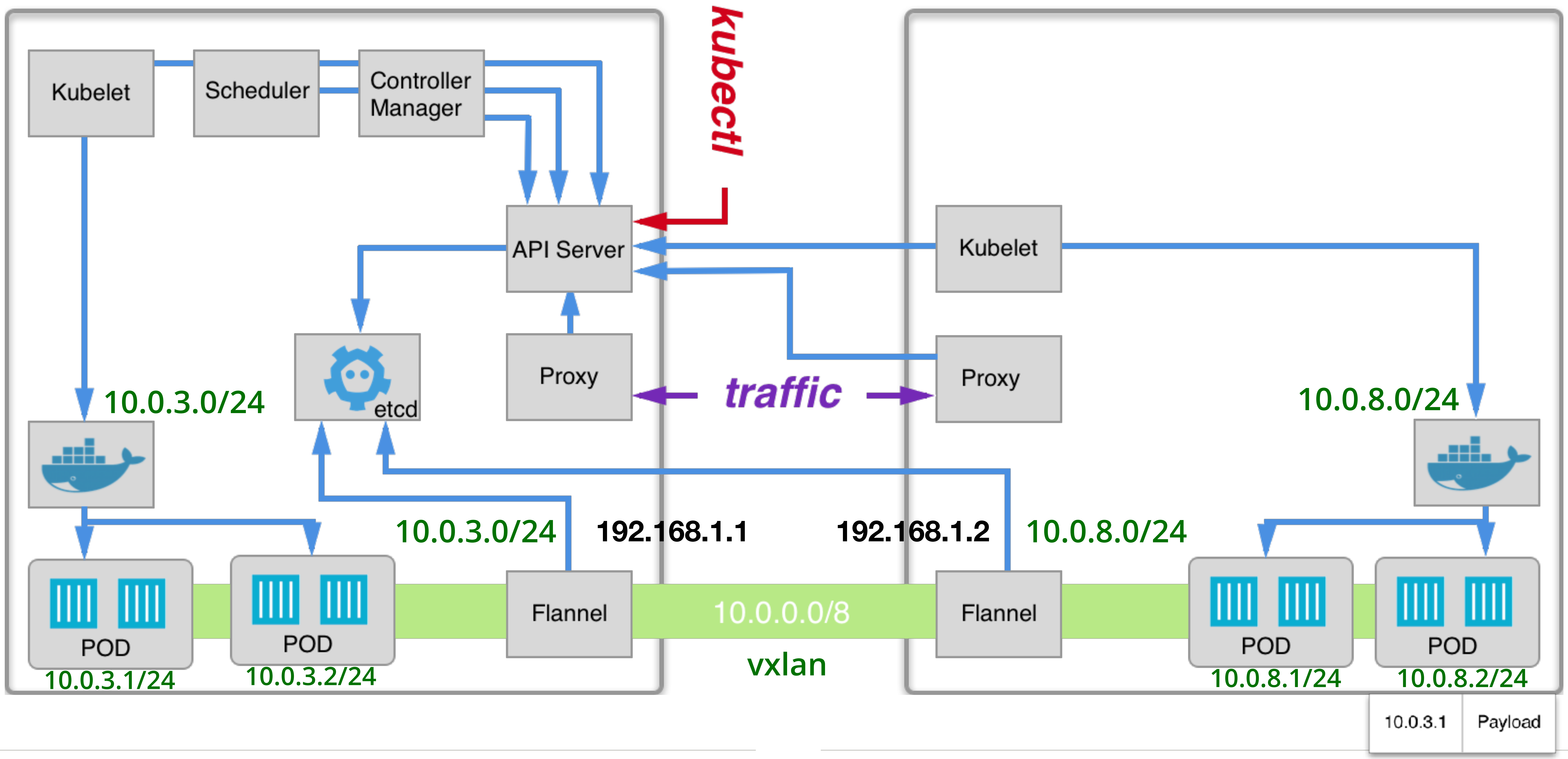
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
Kubelet
etcd
Proxy
traffic
Proxy
10.0.3.0/24
10.0.8.0/24
10.0.3.0/24
10.0.8.0/24
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD
10.0.3.1/24
10.0.3.2/24
10.0.8.1/24
10.0.8.2/24

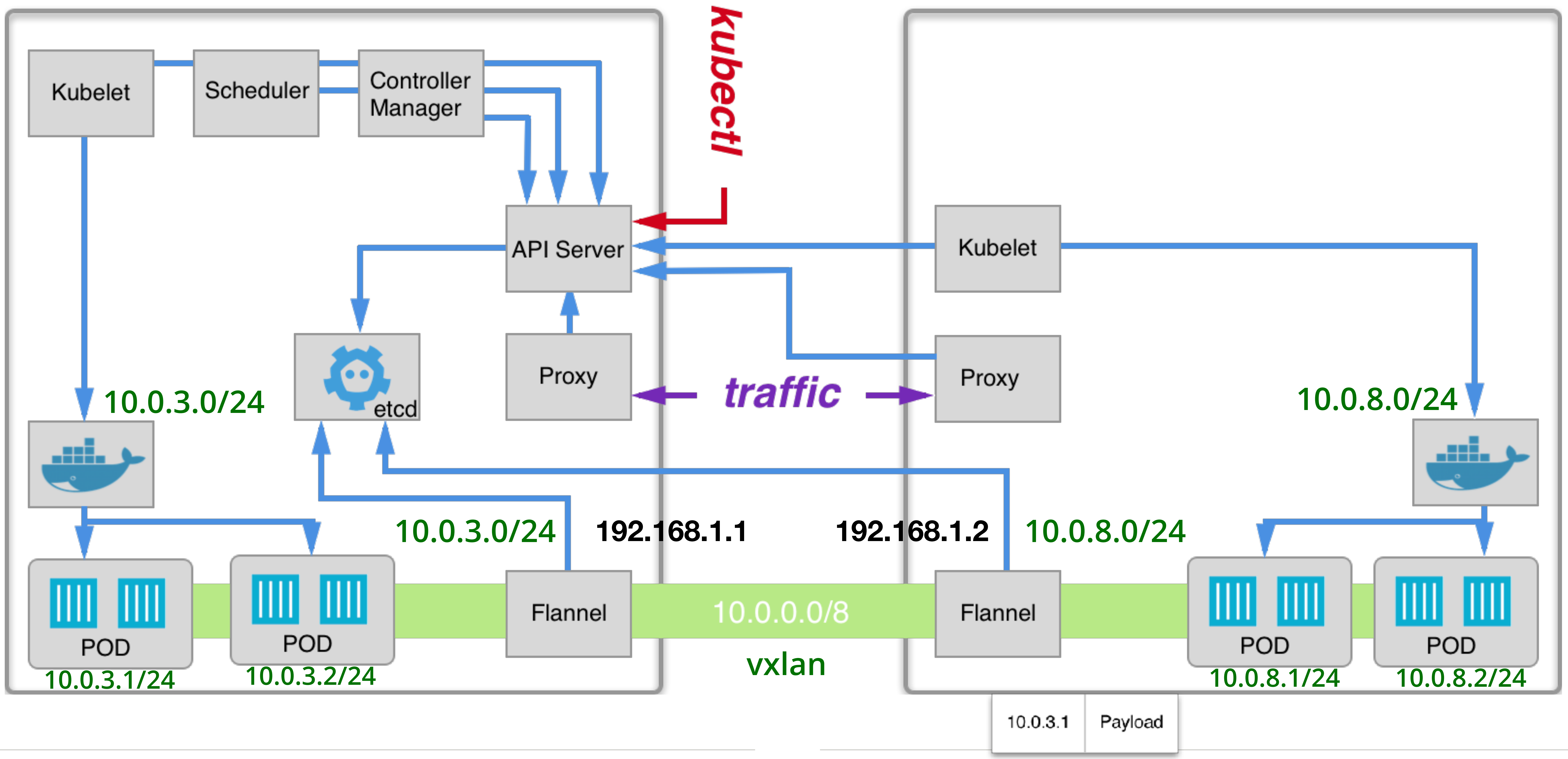
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
Kubelet
Proxy
10.0.3.0/24
10.0.8.0/24
10.0.3.0/24
192.168.1.1
192.168.1.2
10.0.8.0/24
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD
10.0.3.1/24
10.0.3.2/24
vxlan
10.0.8.1/24
10.0.8.2/24

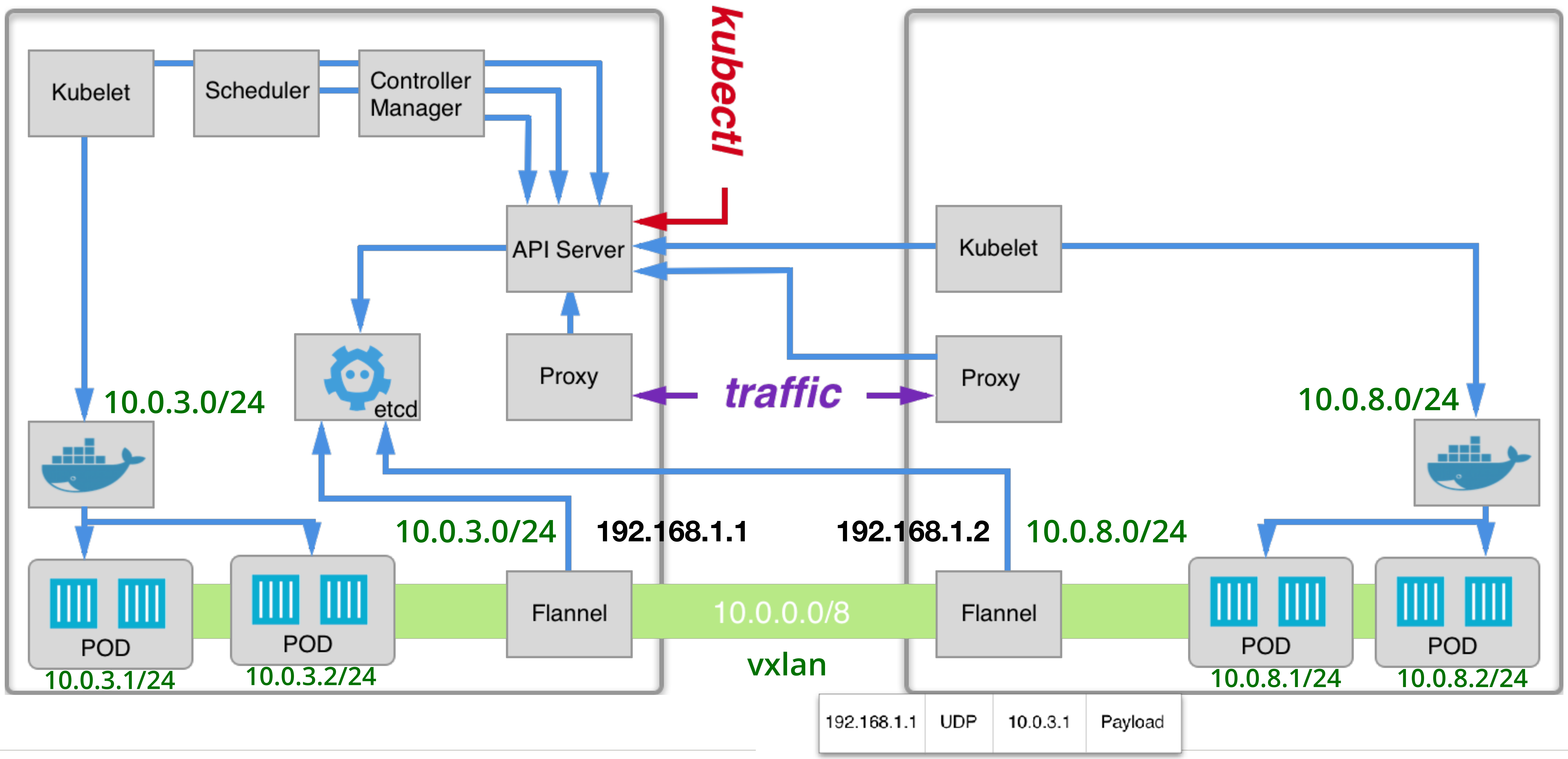
NETWORKING


kubectl
Kubelet
Scheduler
Controller Manager
API Server
etcd
Proxy
10.0.3.0/24
10.0.3.0/24
192.168.1.1
POD
POD
10.0.3.1/24
10.0.3.2/24
Flannel
traffic
10.0.0.0/8
vxlan
Kubelet
Proxy
10.0.8.0/24
192.168.1.2
10.0.8.0/24
Flannel
POD
POD
10.0.8.1/24
10.0.8.2/24
10.0.3.1
Payload

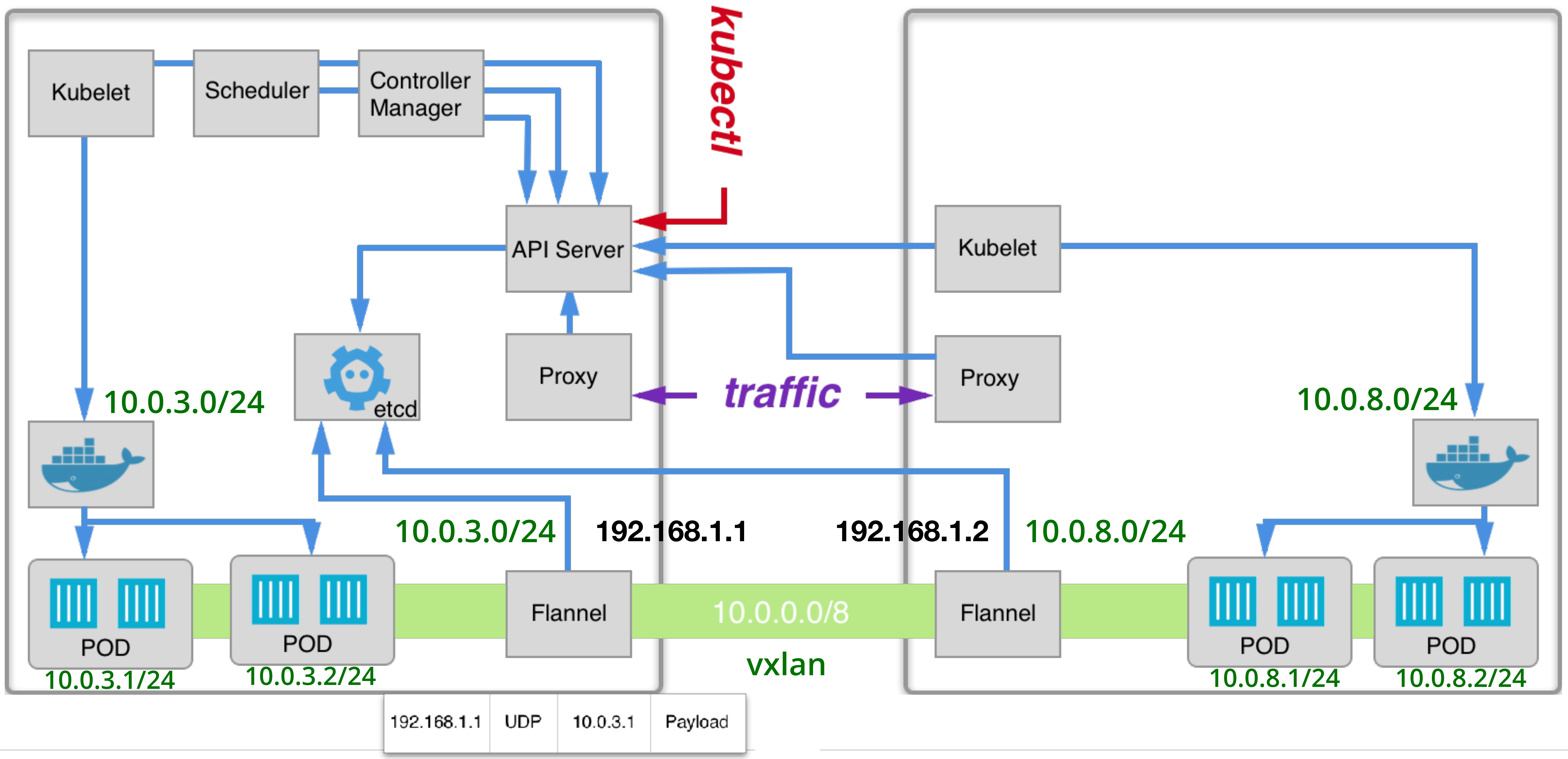
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
Kubelet
Proxy
10.0.3.0/24
10.0.8.0/24
10.0.3.0/24
192.168.1.1
192.168.1.2
10.0.8.0/24
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD
10.0.3.1/24
10.0.3.2/24
vxlan
10.0.8.1/24
10.0.8.2/24
10.0.3.1
Payload

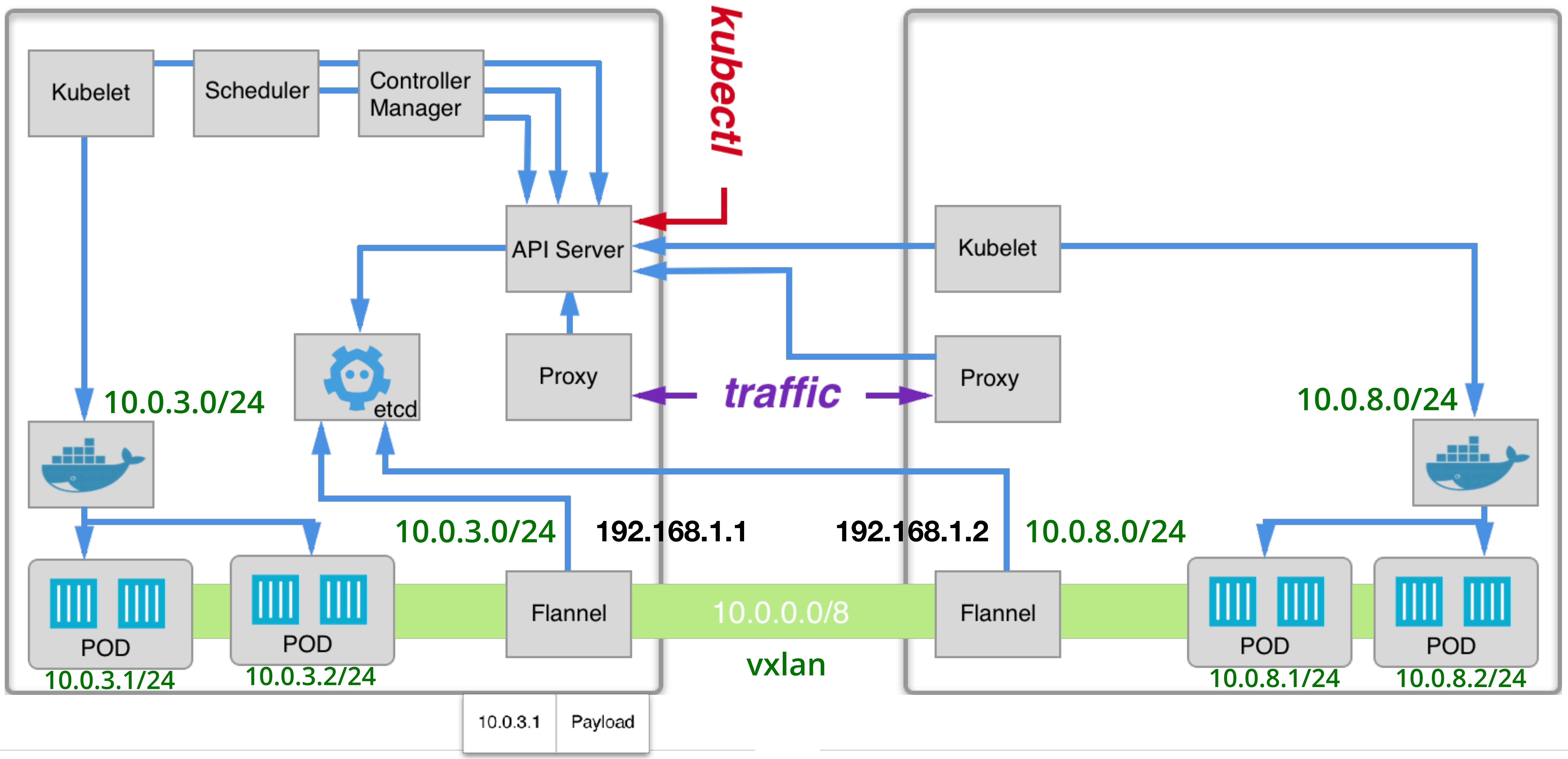
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
Kubelet
Proxy
10.0.3.0/24
10.0.8.0/24
10.0.3.0/24
192.168.1.1
192.168.1.2
10.0.8.0/24
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD
10.0.3.1/24
10.0.3.2/24
vxlan
10.0.8.1/24
10.0.8.2/24
192.168.1.1
UDP
10.0.3.1
Payload

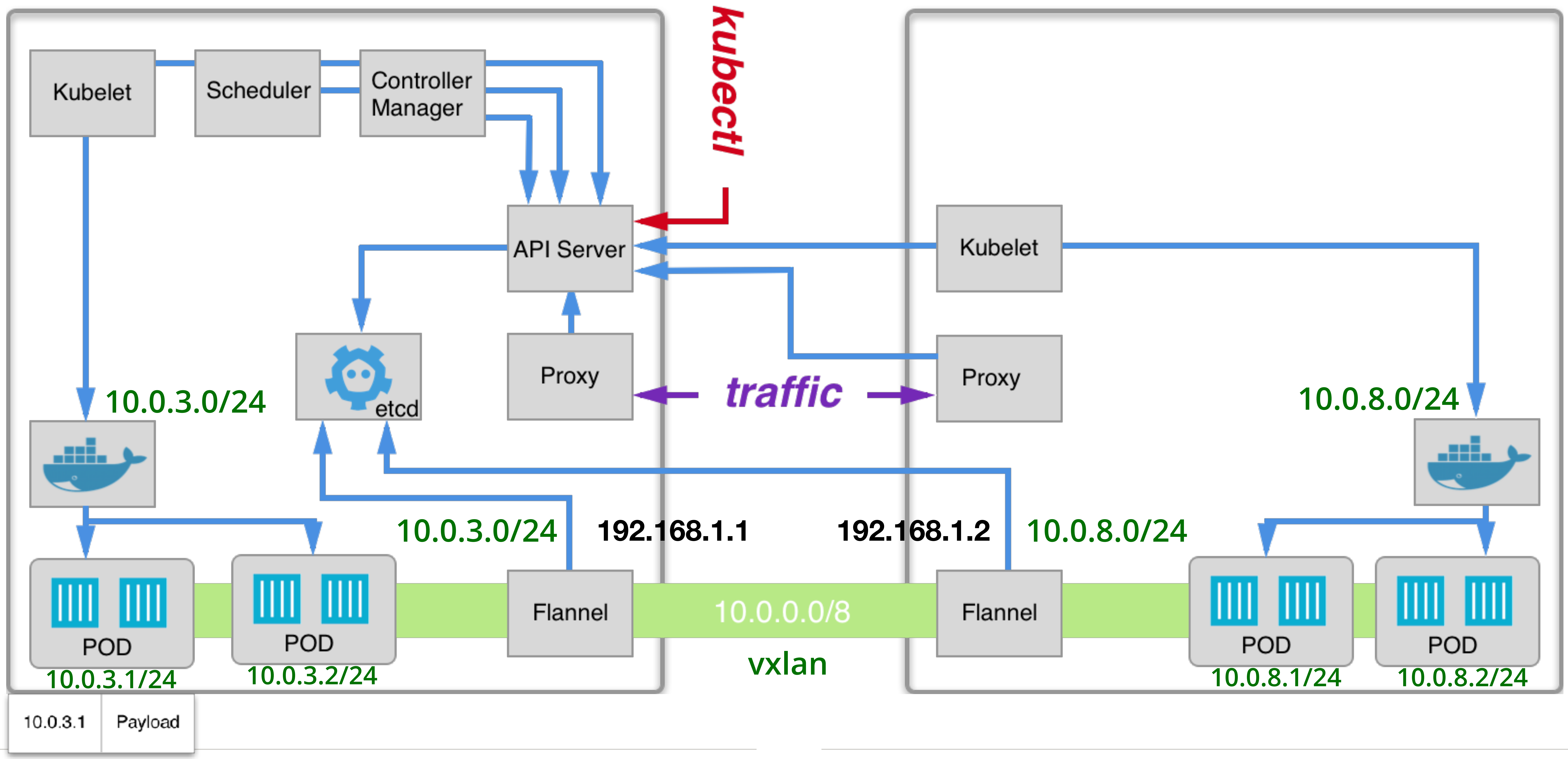
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
10.0.3.0/24
traffic
Kubelet
Proxy
10.0.8.0/24
10.0.3.0/24
192.168.1.1
192.168.1.2
10.0.8.0/24
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD
10.0.3.1/24
10.0.3.2/24
vxlan
10.0.8.1/24
10.0.8.2/24
192.168.1.1
UDP
10.0.3.1
Payload

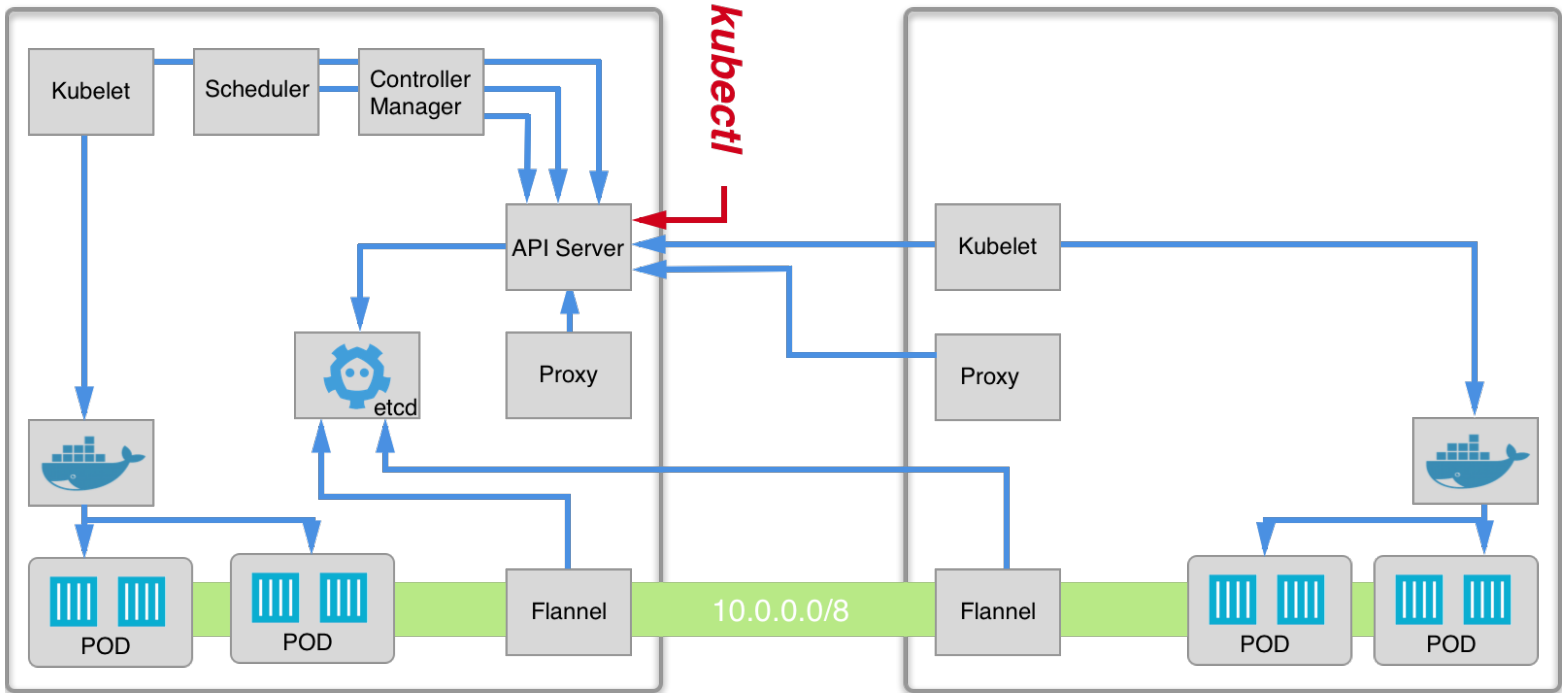
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
Kubelet
Proxy
10.0.3.0/24
10.0.8.0/24
10.0.3.0/24
192.168.1.1
192.168.1.2
10.0.8.0/24
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD
vxlan
10.0.3.1/24
10.0.3.2/24
10.0.8.1/24
10.0.8.2/24
10.0.3.1
Payload

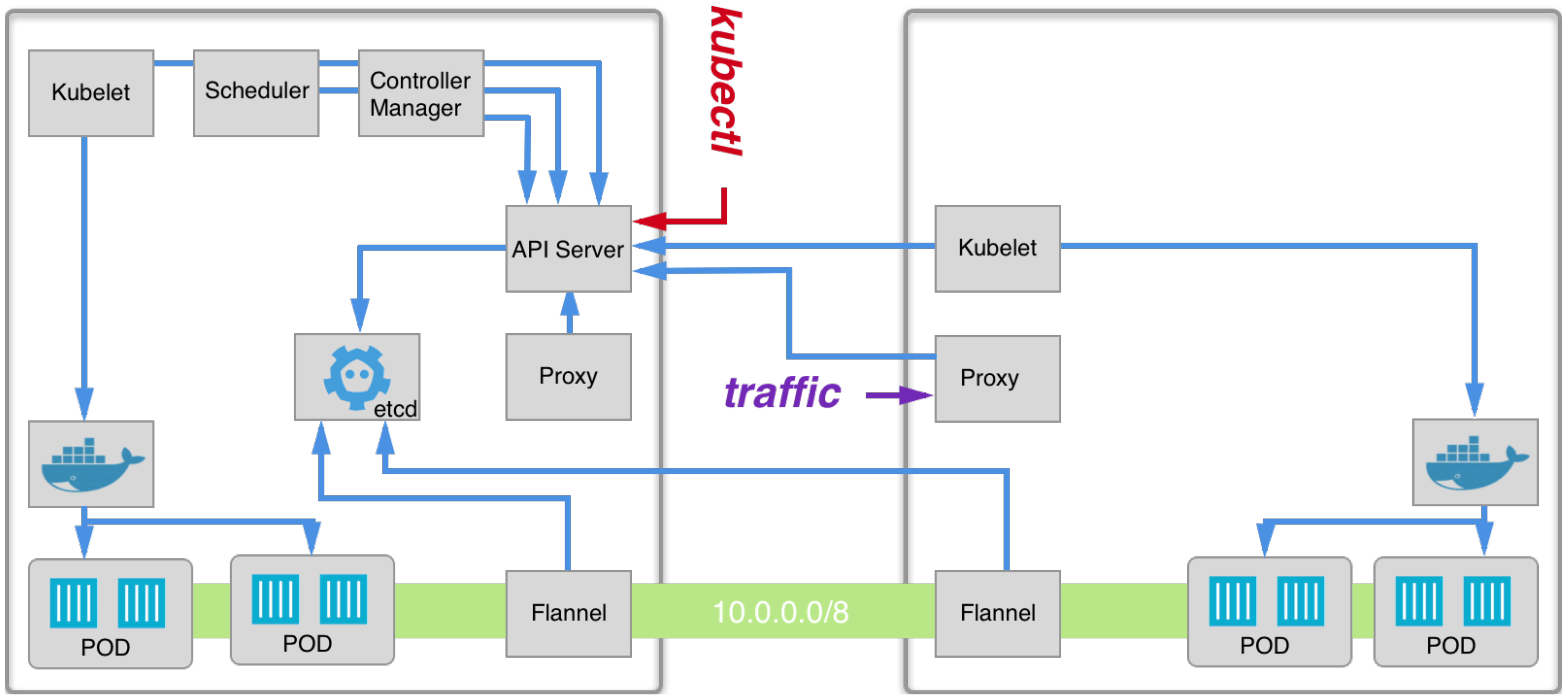
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
Kubelet
etcd
Proxy
Proxy
traffic
10.0.3.0/24
10.0.8.0/24
10.0.3.0/24
192.168.1.1
192.168.1.2
10.0.8.0/24
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD
10.0.3.1/24
10.0.3.2/24
vxlan
10.0.8.1/24
10.0.8.2/24
10.0.3.1
Payload

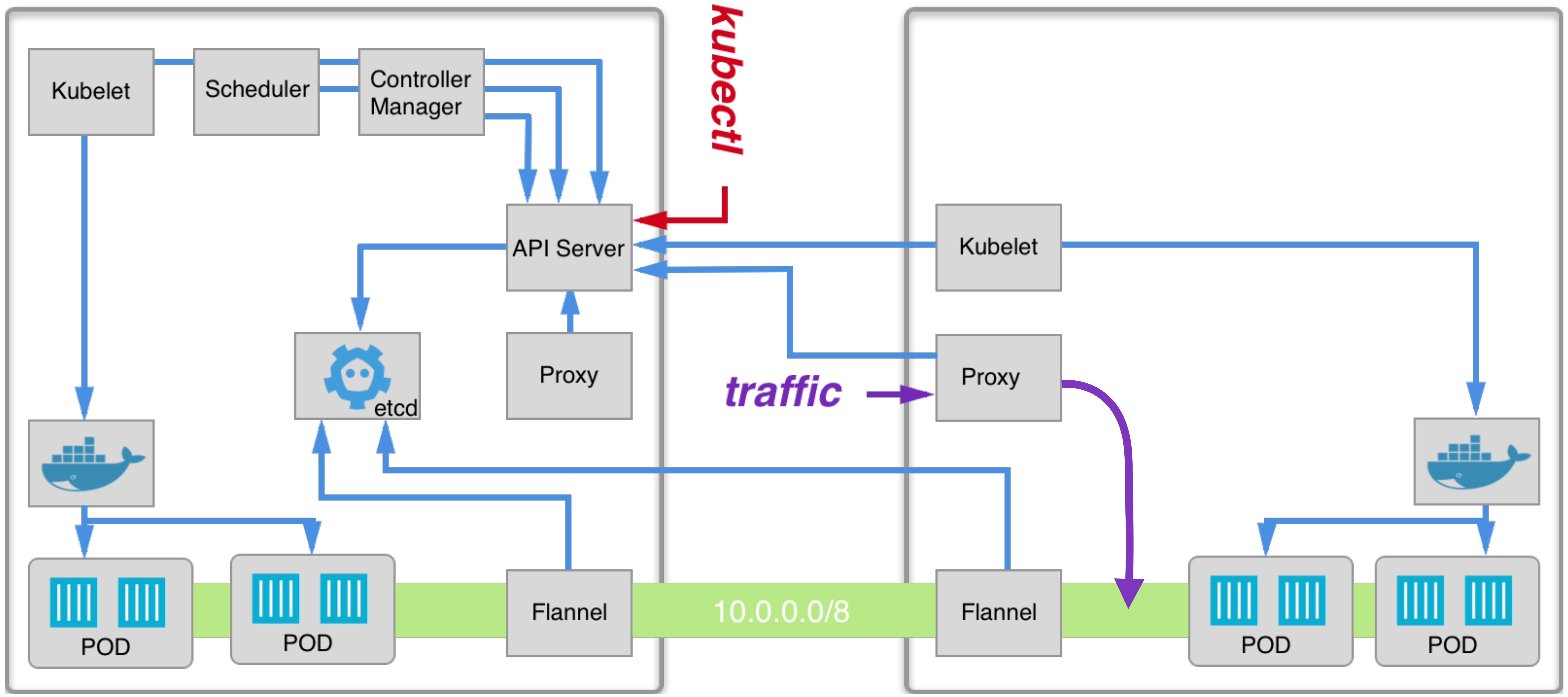
NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
Kubelet
etcd
Proxy
Proxy
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD

NETWORKING

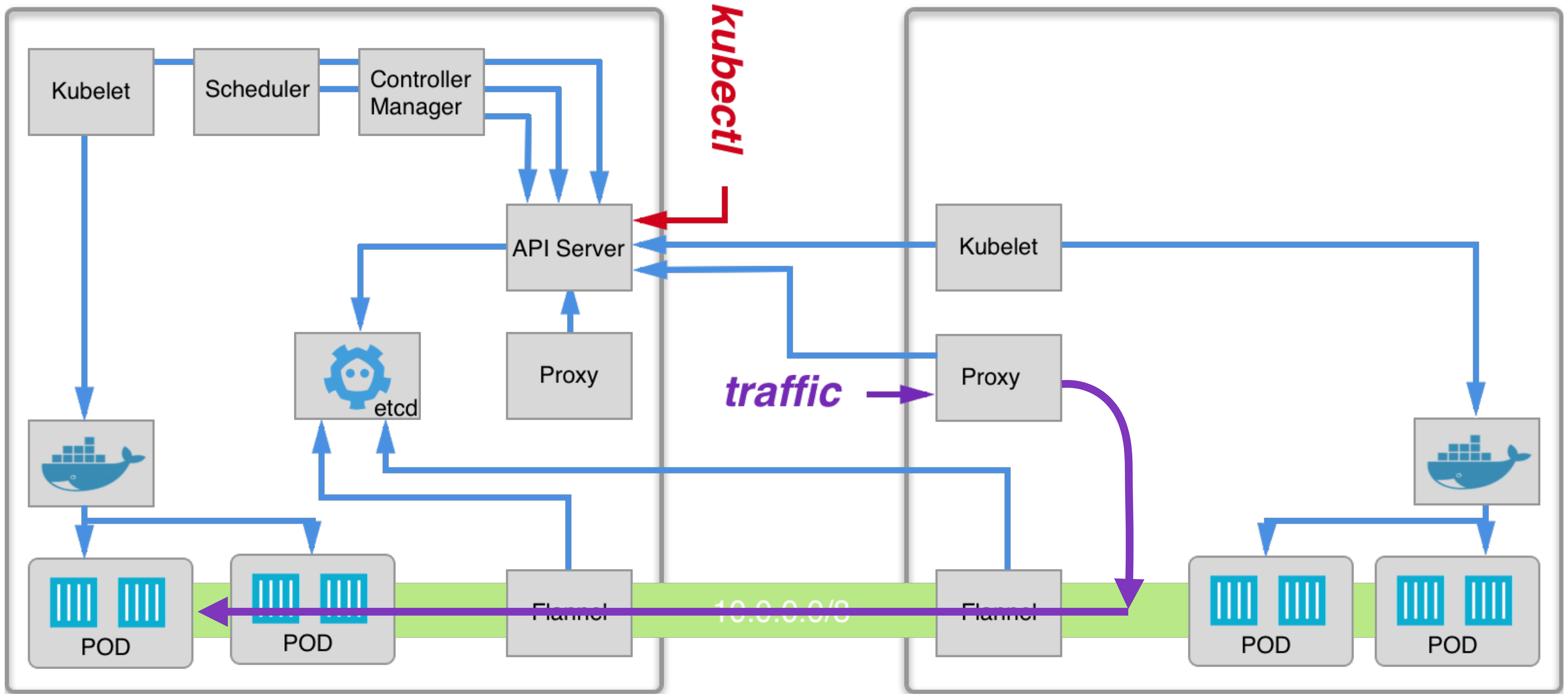

kubectl
Kubelet
Scheduler
Controller Manager
API Server
etcd
Proxy
POD
POD
Flannel
10.0.0.0/8
traffic
Kubelet
Proxy
Flannel
POD
POD

NETWORKING

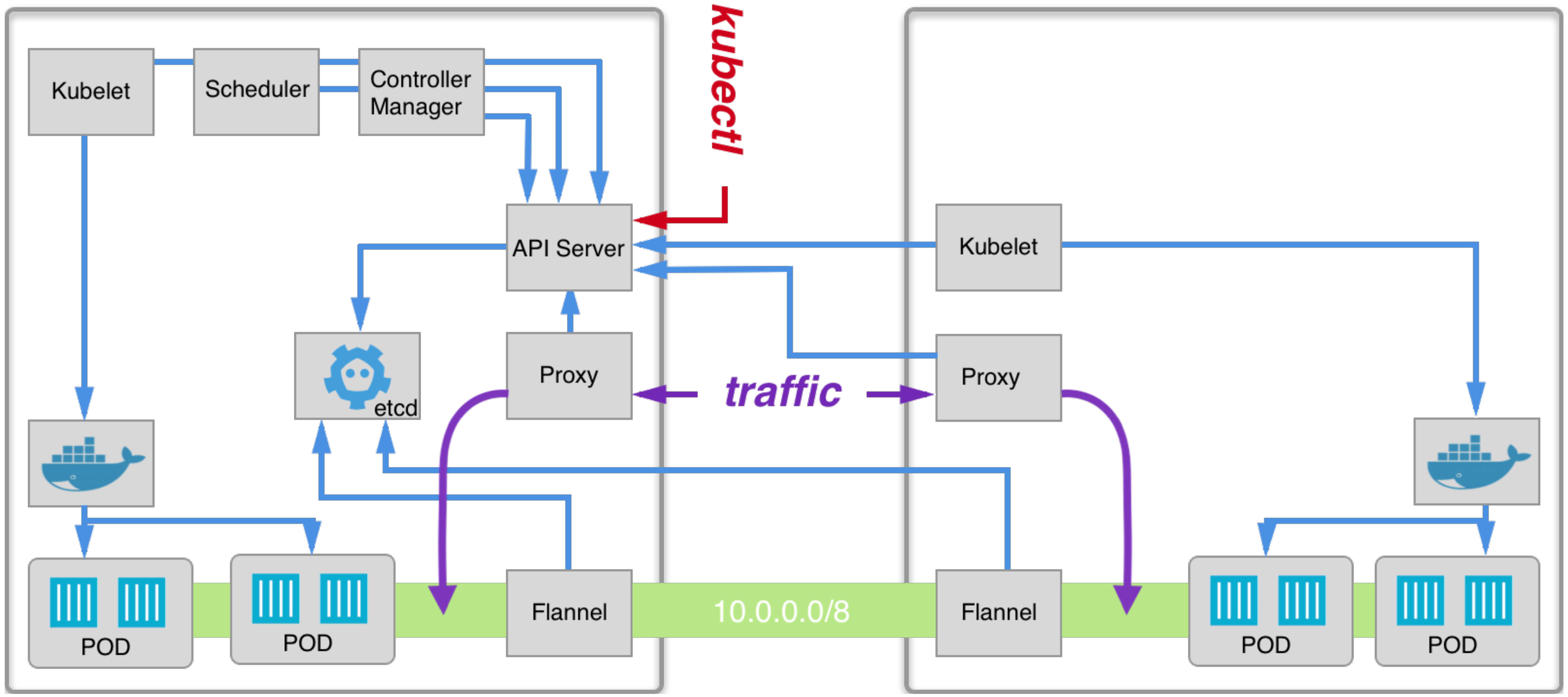


kubectl
Kubelet
Scheduler
Controller Manager
API Server
etcd
Proxy
POD
POD
Flannel
10.0.0.0/8
traffic
Kubelet
Proxy
Flannel
POD
POD

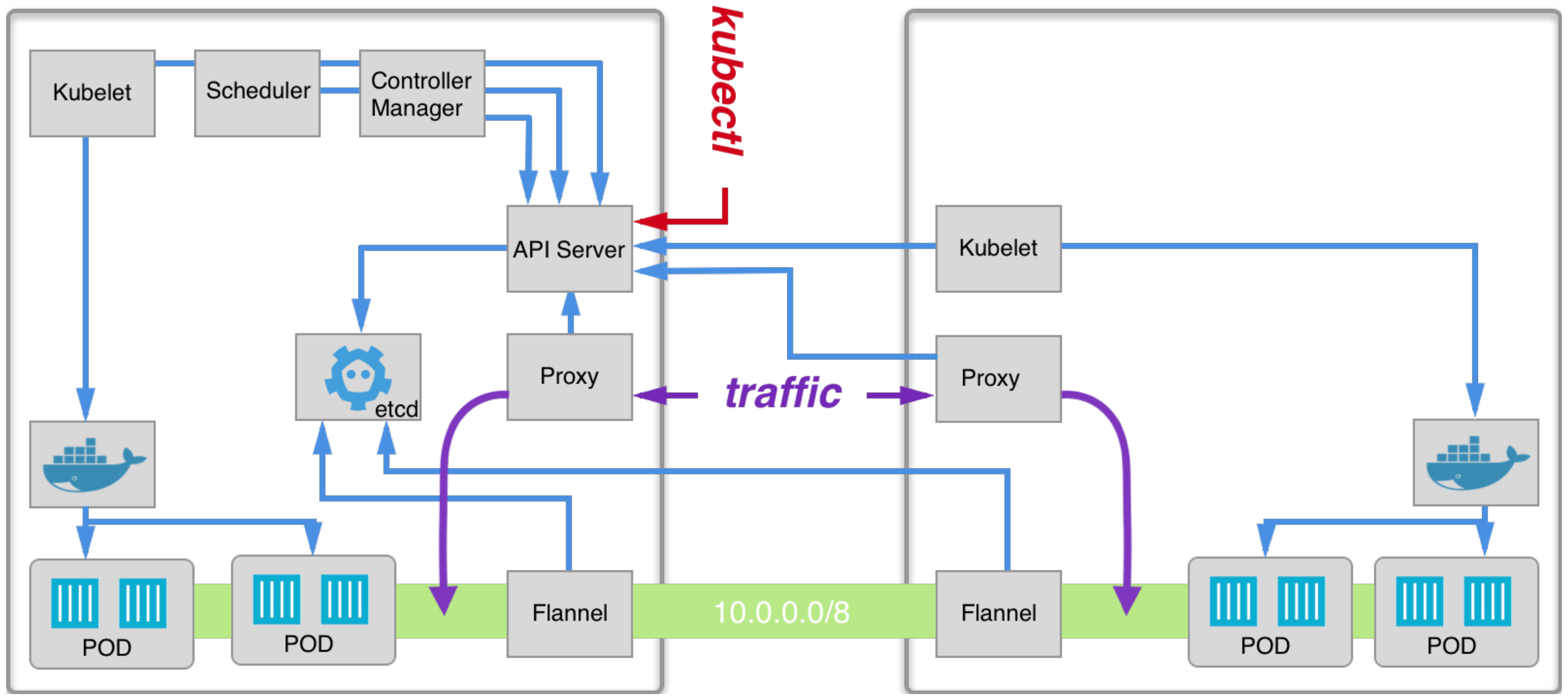
NETWORKING


kubectl
Kubelet
Scheduler
Controller Manager
API Server
etcd
Proxy
traffic
Kubelet
Proxy
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD

NETWORKING

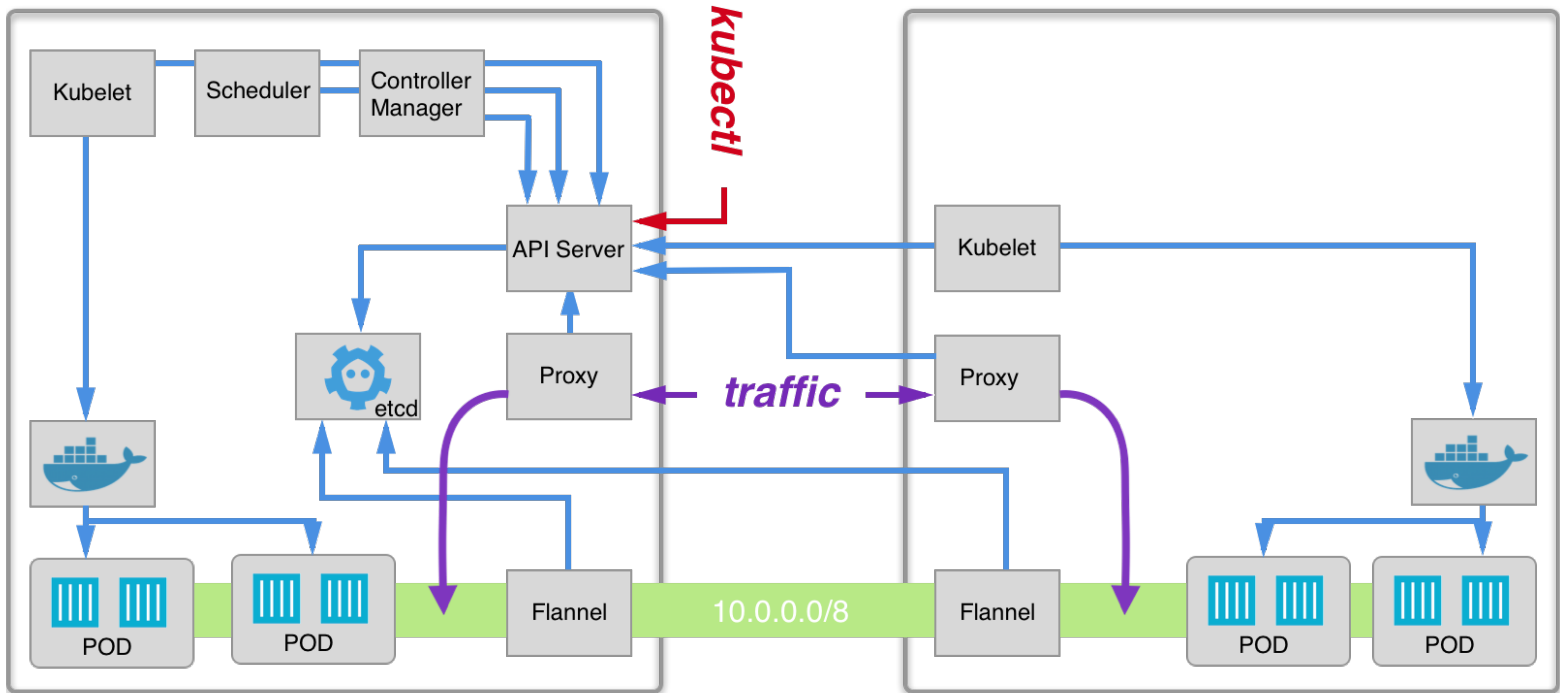

Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
POD
POD
Flannel
10.0.0.0/8
Kubelet
Proxy
Flannel
POD
POD

NETWORKING


Kubelet
Scheduler
Controller Manager
kubectl
API Server
etcd
Proxy
traffic
Kubelet
Proxy
POD
POD
Flannel
10.0.0.0/8
Flannel
POD
POD

NETWORKING

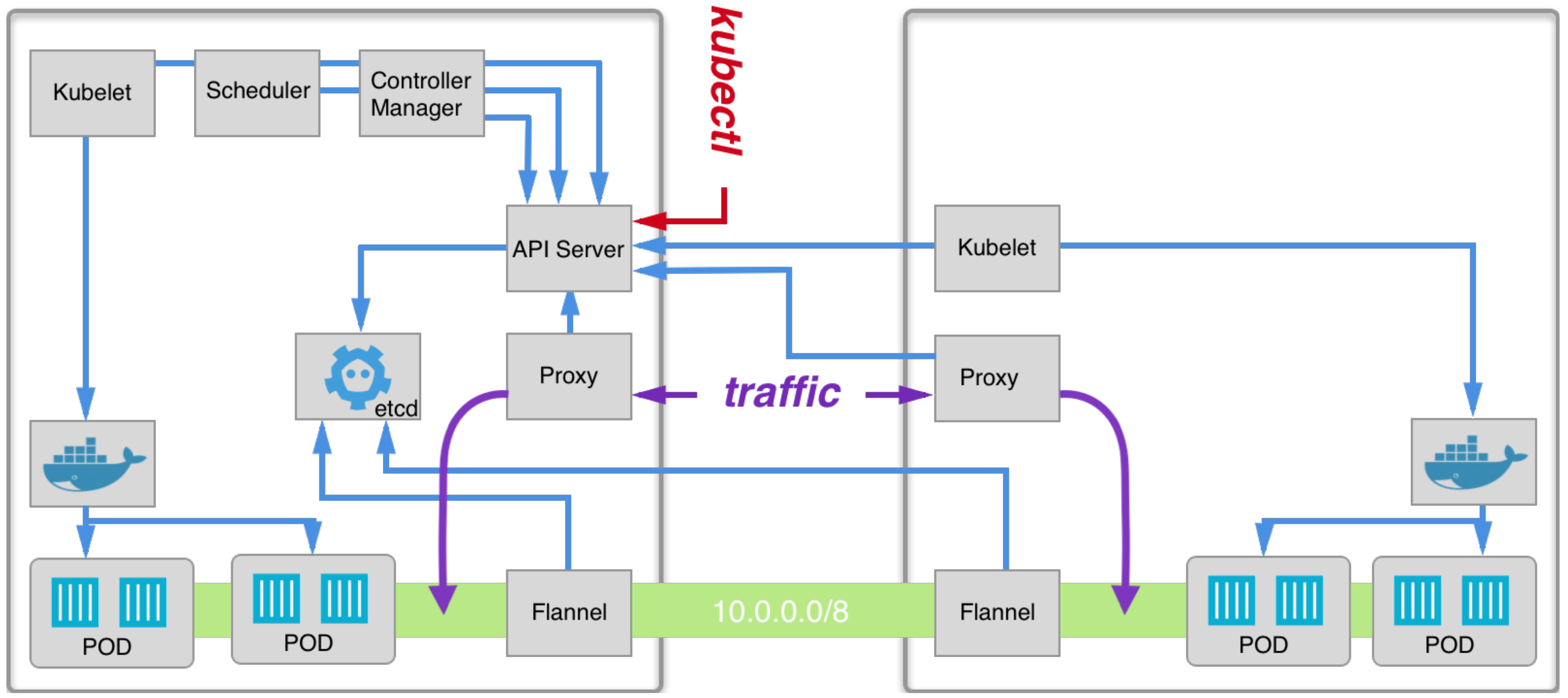
Flannel:



NETWORKING

Flannel:

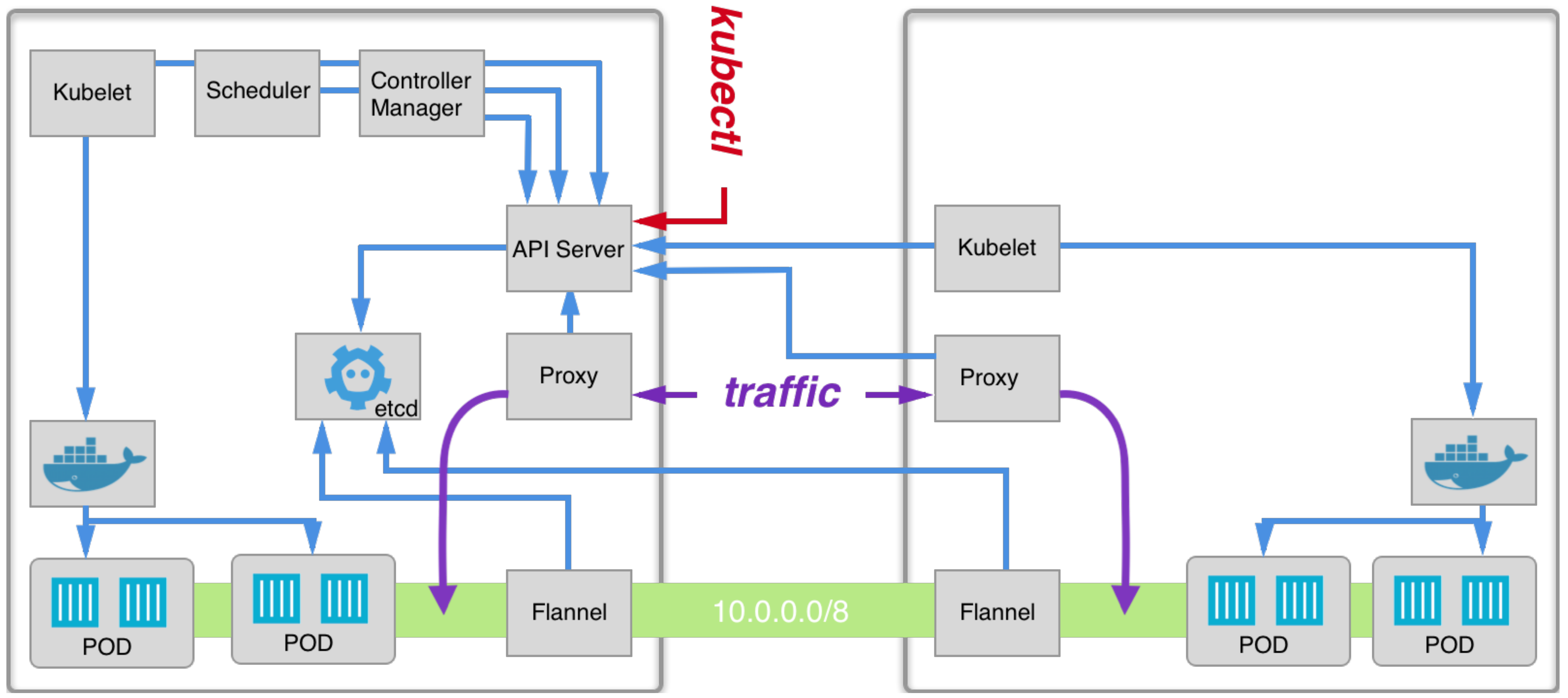
- uses etcd as a database



NETWORKING

Flannel:

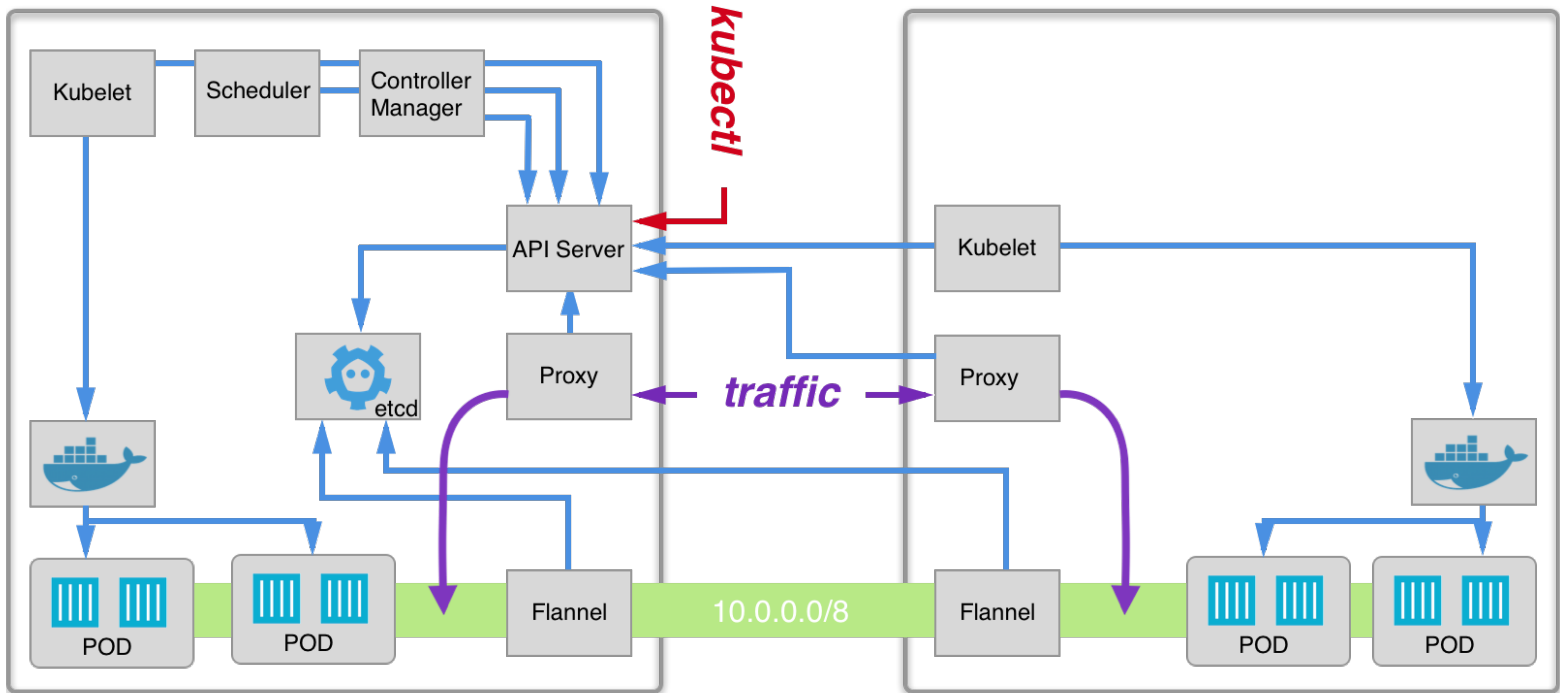
- uses etcd as a database
- agent allocates a subnet for a node out of larger address space



NETWORKING

Flannel:

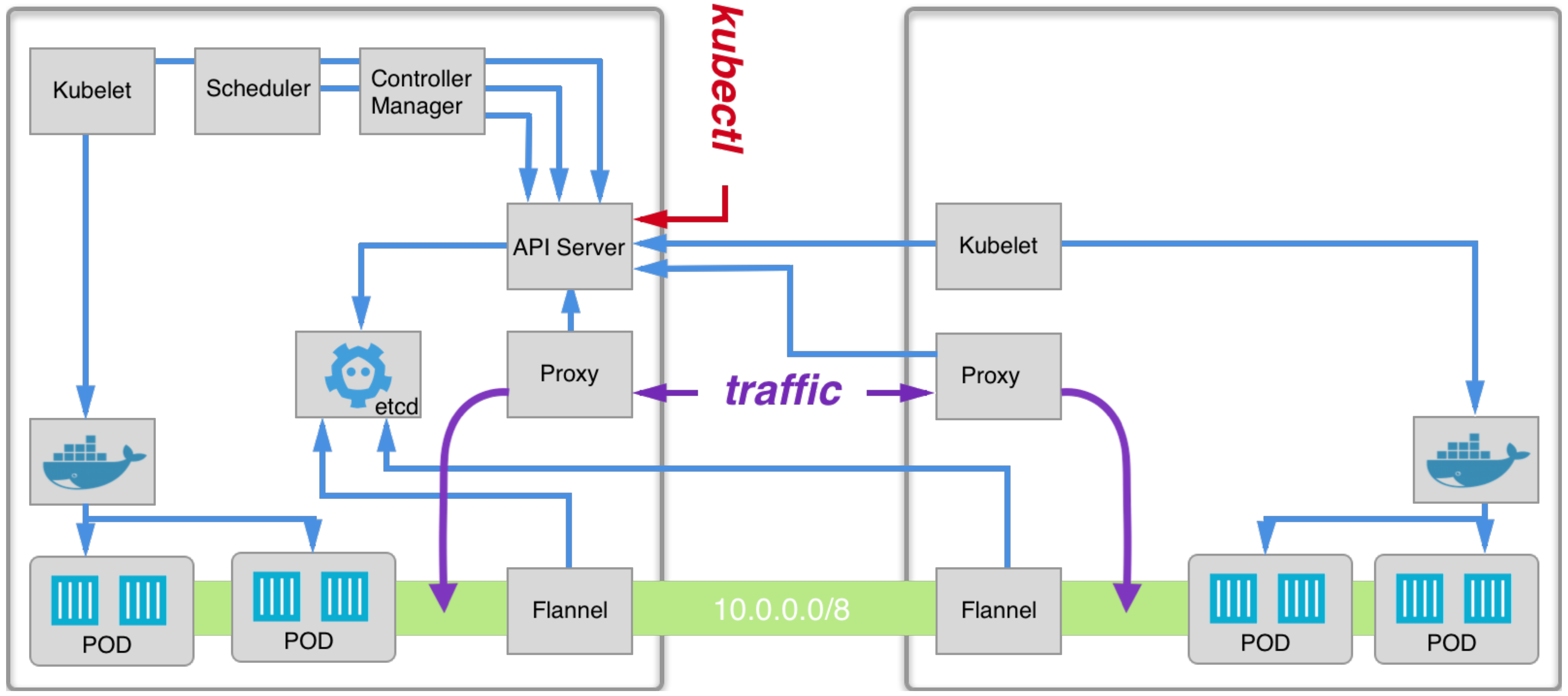
- uses etcd as a database
- agent allocates a subnet for a node out of larger address space
- tells docker to use that subnet



NETWORKING

Flannel:

- uses etcd as a database
- agent allocates a subnet for a node out of larger address space
- tells docker to use that subnet
- uses vxlan to encapsulate and foward packages between nodes



DEMO

Q&A

THANK YOU