Fault Tolerant Infrastructure

Building Systems with etcd

@coreoslinux
@brandonphilips

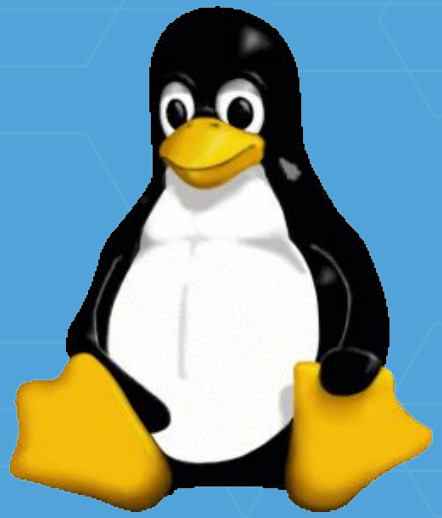
Brandon Philips

CTO, CoreOS

github.com/philips

What is CoreOS?

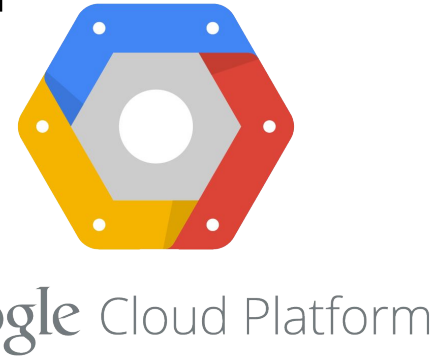
What is CoreOS?




openstack
CLOUD SOFTWARE

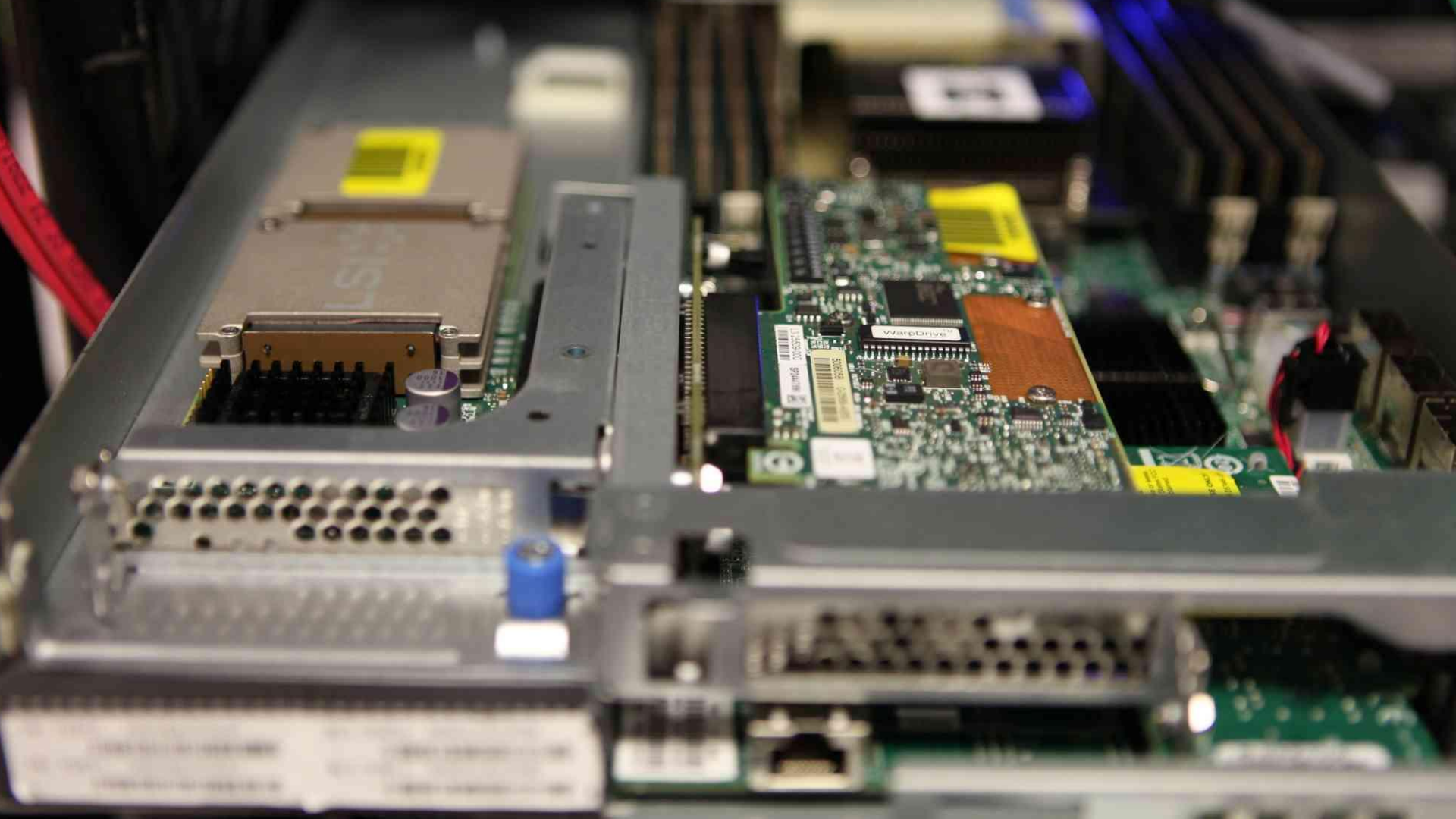

amazon
web services™


DigitalOcean


Google Cloud Platform


rackspace®
the #1 managed cloud company


vmware®


WarpDrive

What is CoreOS?


etcd
4,961
6
38
13
527
Top 100 Community Contributors
CoreOS
16
18
Issues
10k+ total commits
~30 repositories
324 releases
22 CoreOS contributors
605 community contributors
Top 100 Community Contributors

The smartest way to run your container infrastructure.

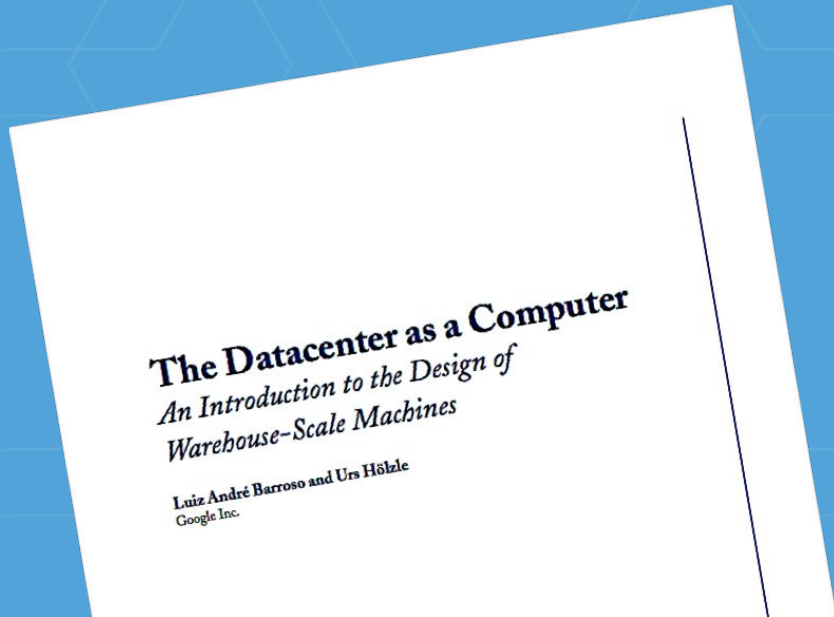
tectonic.com @tectonic

QUAY

Secure hosting for private Docker repositories

quay.io @quayio

Why build CoreOS?


The Datacenter as a Computer
An Introduction to the Design of Warehouse-Scale Machines
Luiz André Barroso and Urs Hölzle
Google Inc.

you

you as a sw engineer

your

```
with Ada.Text_IO;

procedure Hello_World is
 use Ada.Text_IO;
begin
    Put_Line("Hello, world!");
end;
```

```
#include <stdio.h>

int main()
{
    printf("Hello, world!\n");
}
```

```
package main

import "fmt"

func main() {
      fmt.Println("Hello, world!")
}
```

your

your

your

your

d474e8c57737625c

your

Signed By: Alice

d474e8c57737625c

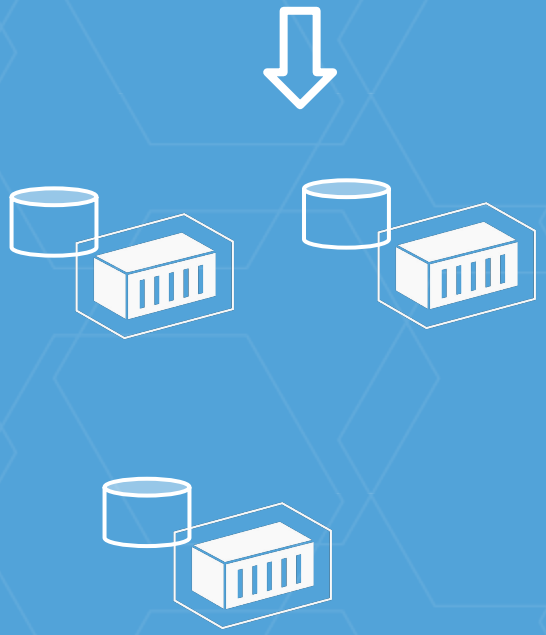
you as an ops engineer

your

com.example.webapp
x3

your

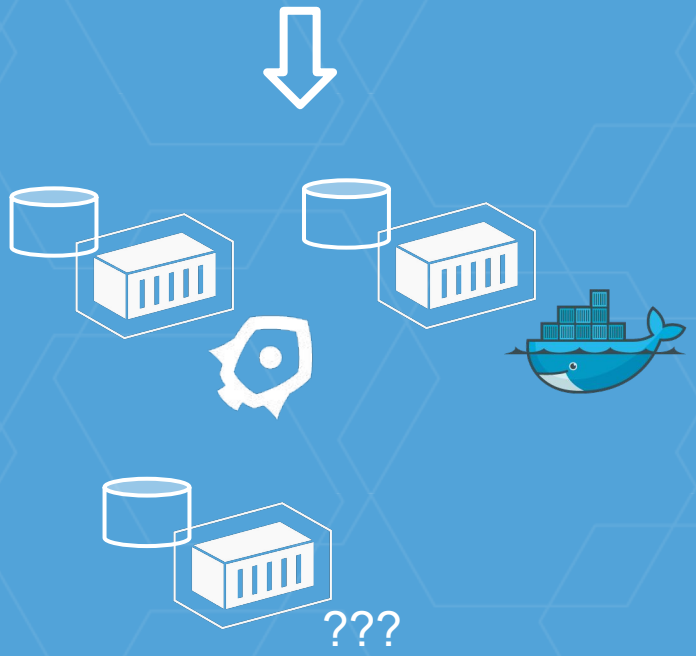
com.example.webapp
x3



your

your

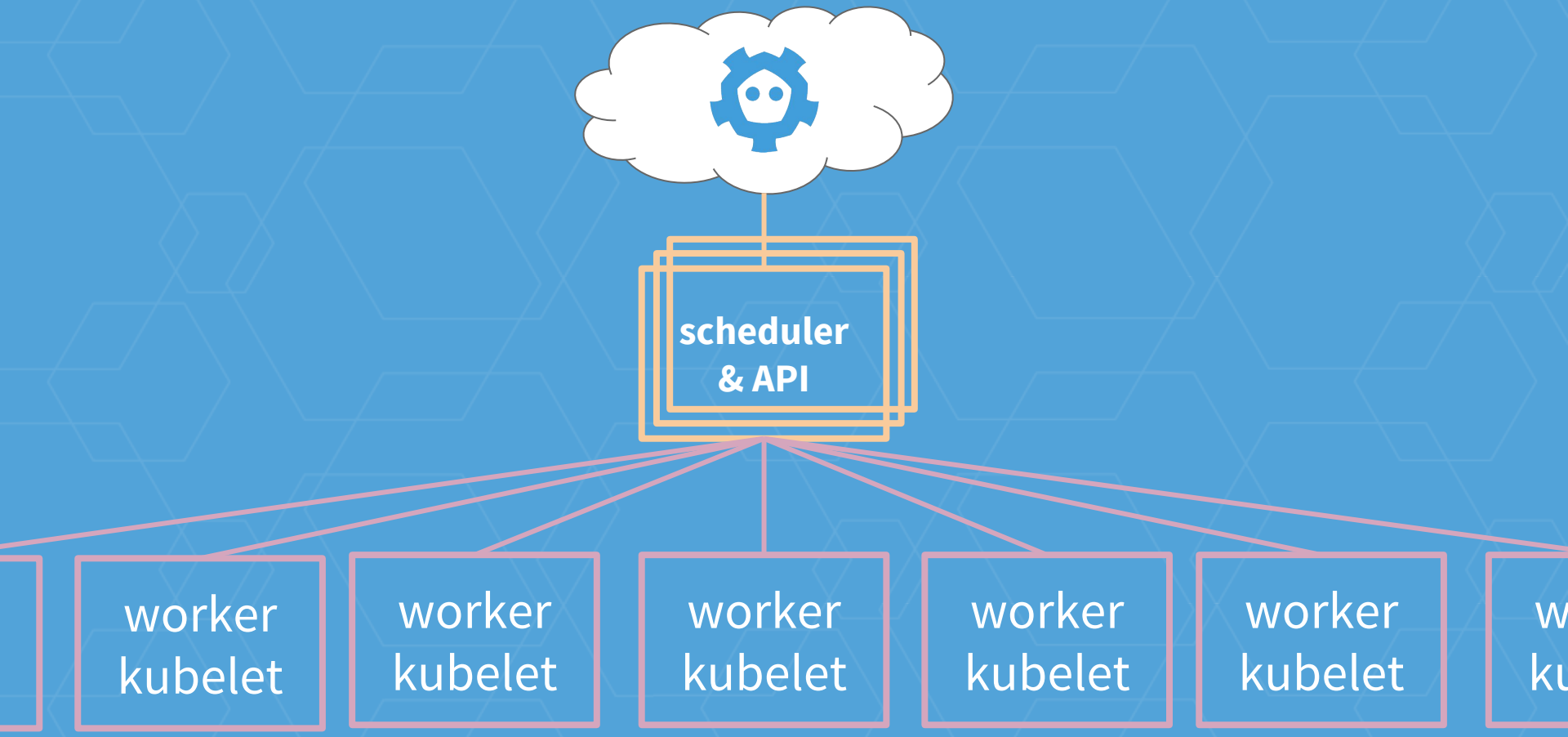
com.example.webapp
x3



How do we do it?

cluster operations

architecture in practice


scheduler
& API
worker
kubelet
worker
kubelet
worker
kubelet
worker
kubelet
worker
kubelet

OS operations
machine configuration

cluster operations

distributed configuration

github.com/philips/hacks/tree/master/etcd-demos

etcd

/etc
distributed

open source software

failure tolerant

durable

watchable

exposed via HTTP

runtime reconfigurable

Data Store API

-X GET

- Get Wait

-X PUT

- Put Create CAS

-X DELETE

- Delete CAD

etcd basics
clusters

Typical Cluster

Leader

Follower

etcd basics

API

etcd basics

fault tolerance

Available

Leader

Follower

Available

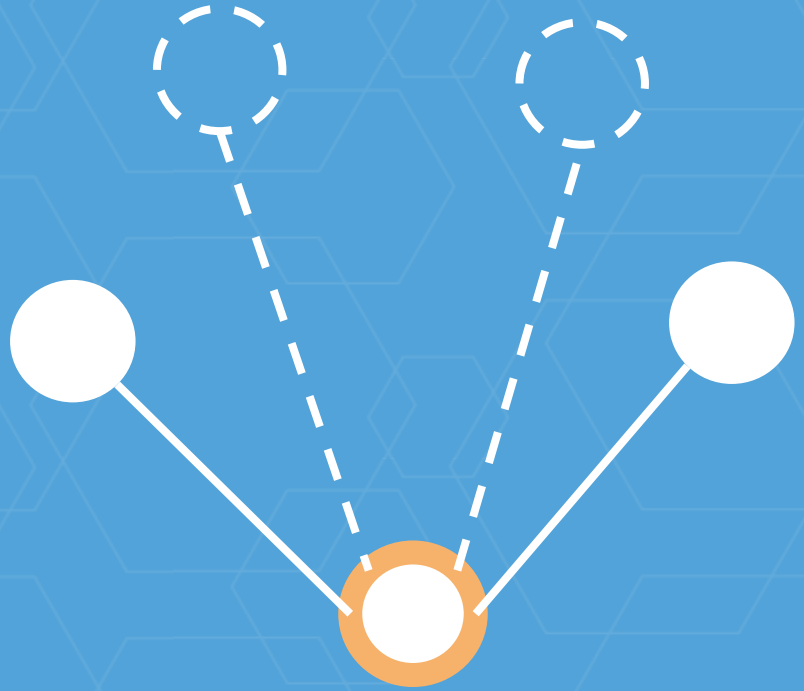
Leader

Follower

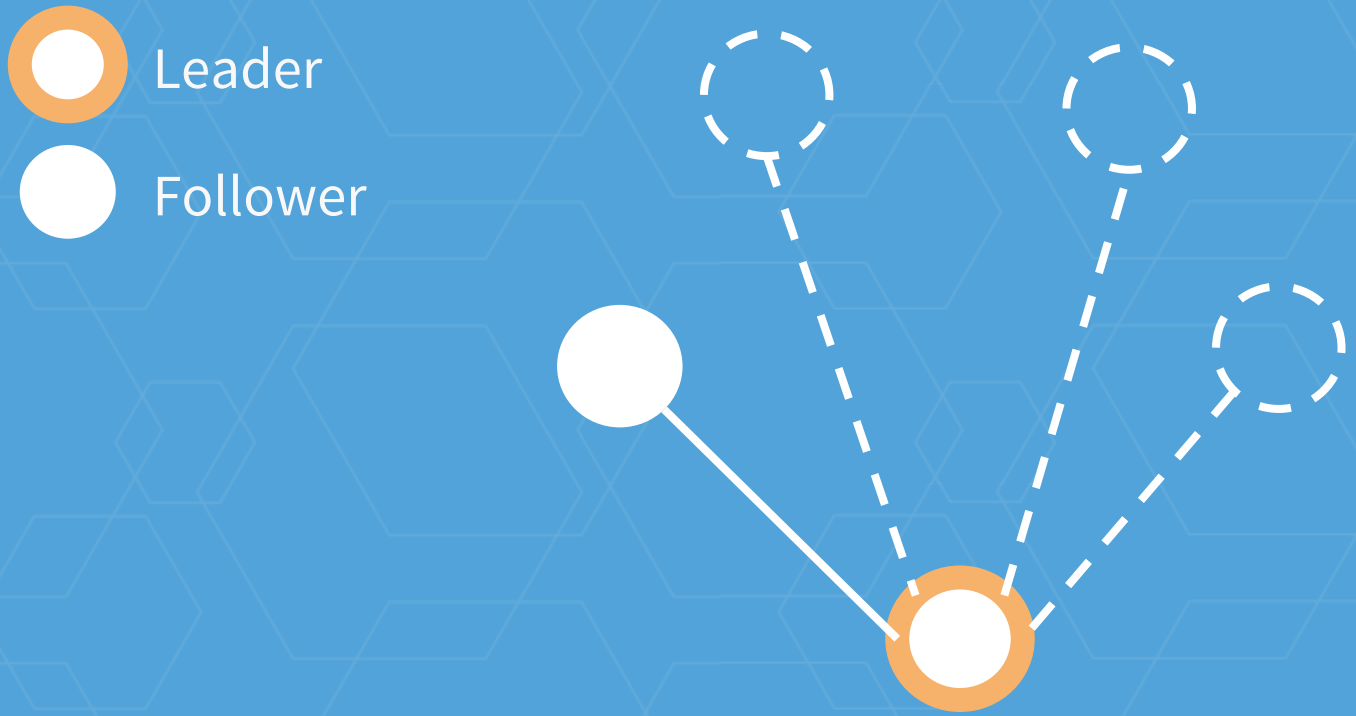
Available

Leader

Follower



Unavailable



etcd basics

leader fault tolerance

Available

Leader

Follower

Available

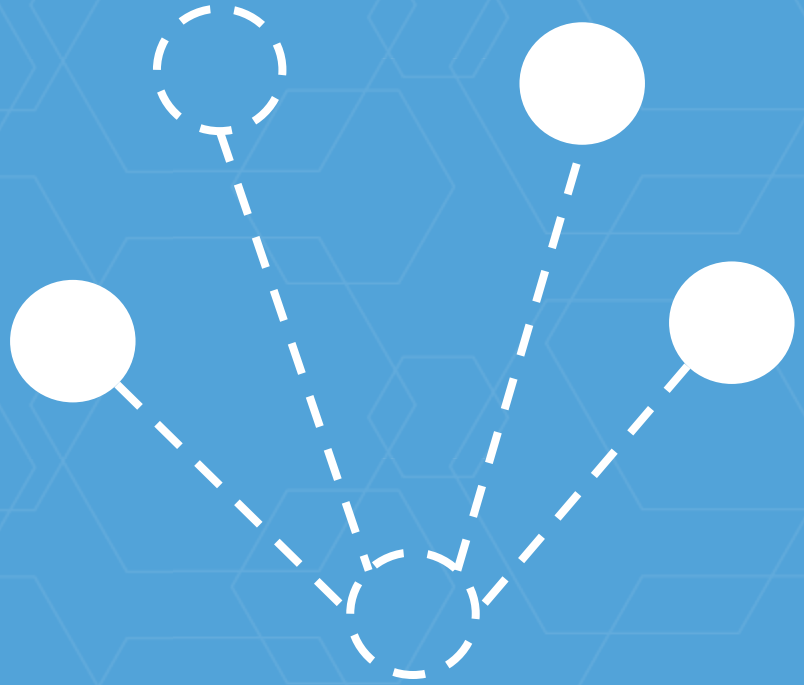
Leader

Follower

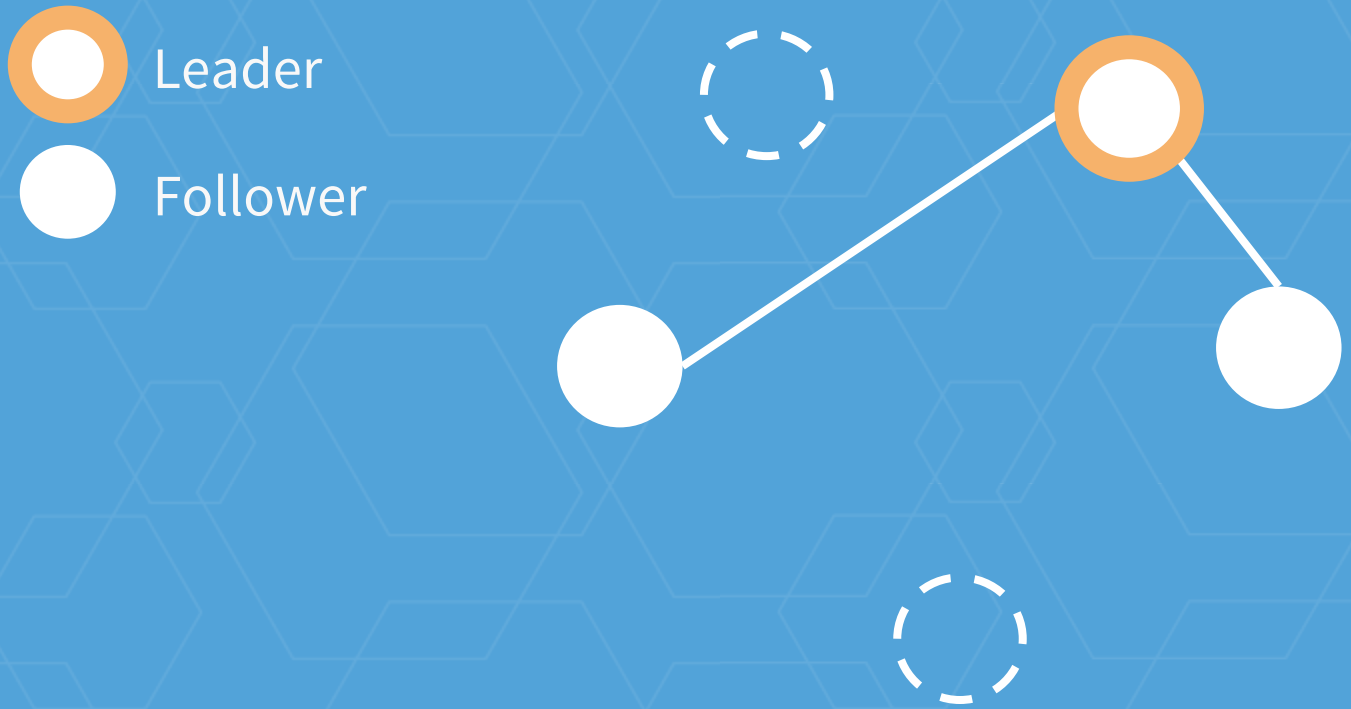
Temporarily Unavailable

Leader

Follower



Available



Unavailable

etcd durability

wal, snapshots, backups

etcd bootstrap

discovery, static

```
$ curl discovery.etcd.io/new?size=5
discovery.etcd.io/6eadeac2
```

discovery

discovery

discovery

discovery

Leader

Follower

etcd reconfig

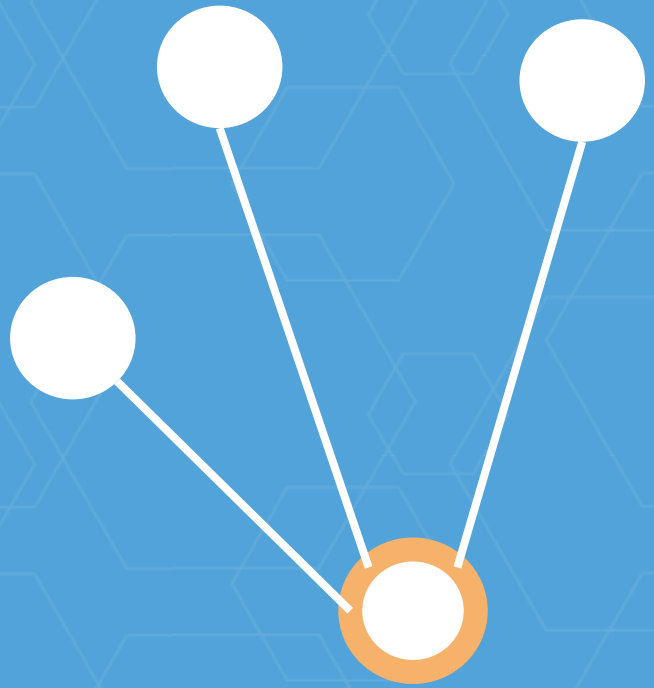
live addition and removal

Leader

Follower

Leader

Follower



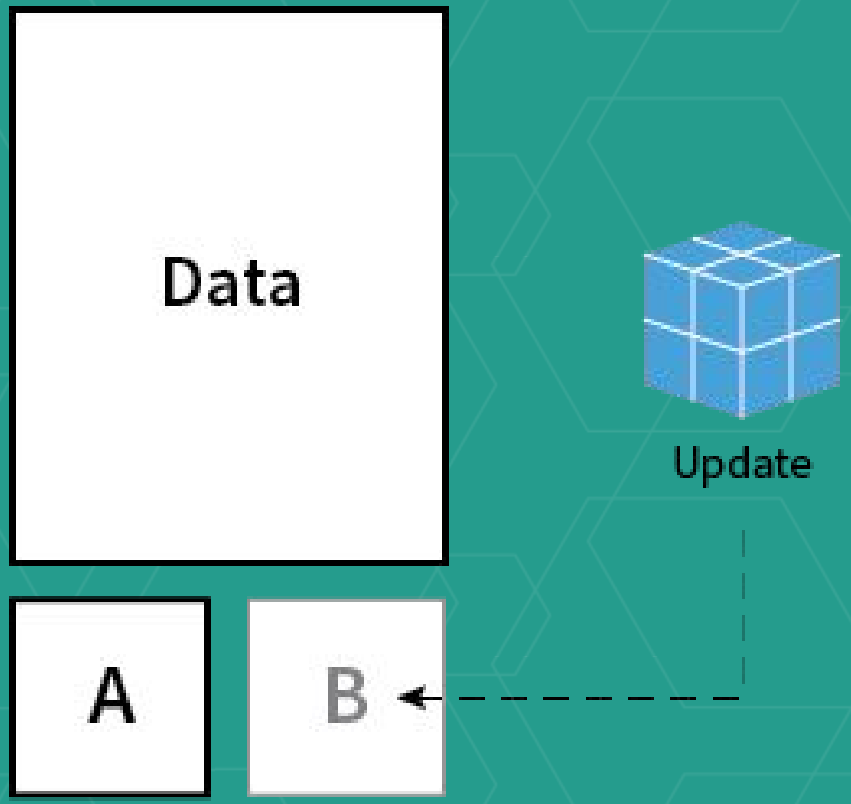
Leader

Follower

etcd apps

etcd apps

reboot locksmith


Data
Update
A
B


Data
A
B

Cluster Wide Reboot Lock

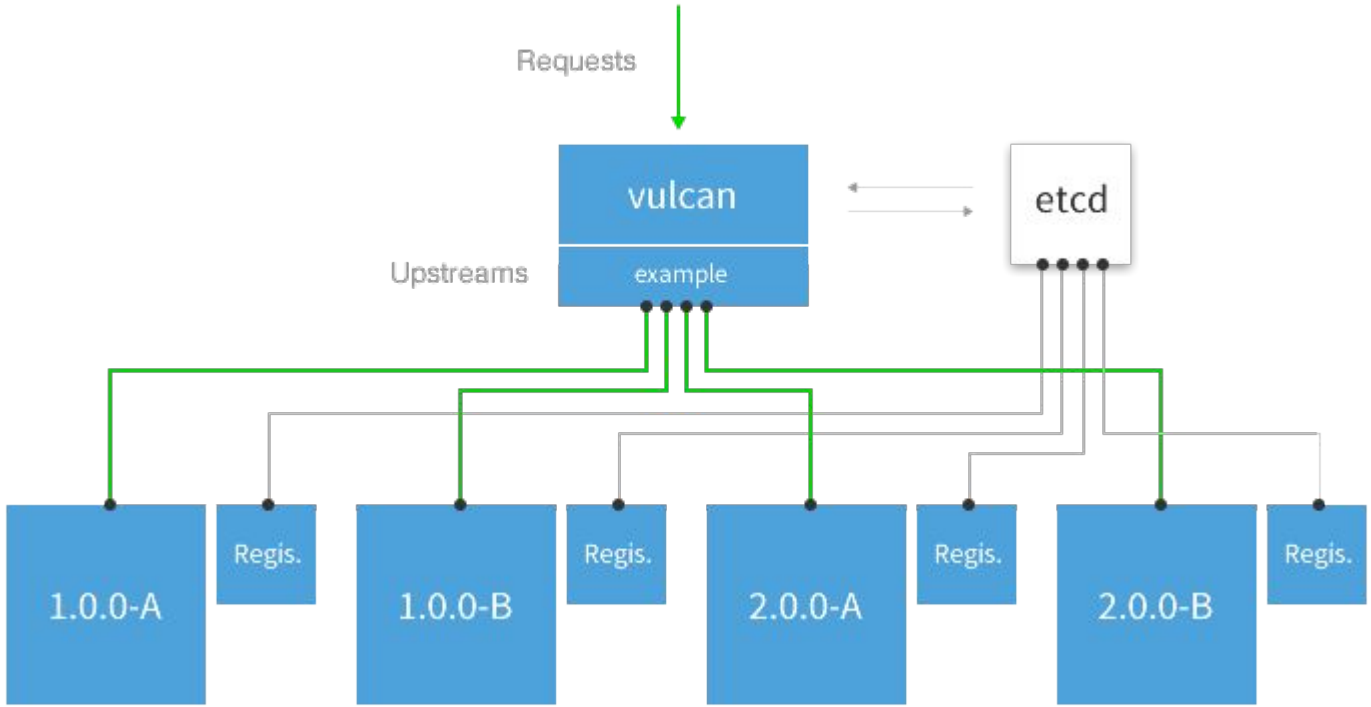
- Need to reboot? Decrement the semaphore key atomically with etcd.
- manager.Reboot() and wait...
- After reboot increment the semaphore key in etcd atomically.

etcd apps

skydns

etcd apps

vulcand


Requests
vulcan
etcd
Upstreams
example
1.0.0-A
Regis.
1.0.0-B
Regis.
2.0.0-A
Regis.
2.0.0-B
Regis.

etcd apps

confd

kubernetes
pulling it together

scheduling

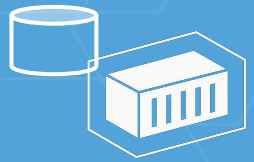
k8s/mesos/etc scheduler

scheduling

getting work to servers

```
$ scp app host:/opt
$ ssh host systemd-run /opt/app
```

```
$ scp app host:/opt
$ ssh host systemd-run /opt/app
```



$ fab deploy:app

$ fab deploy:app

$ fab deploy:app

$ fab deploy:collector-app

$ fab deploy:collector-app

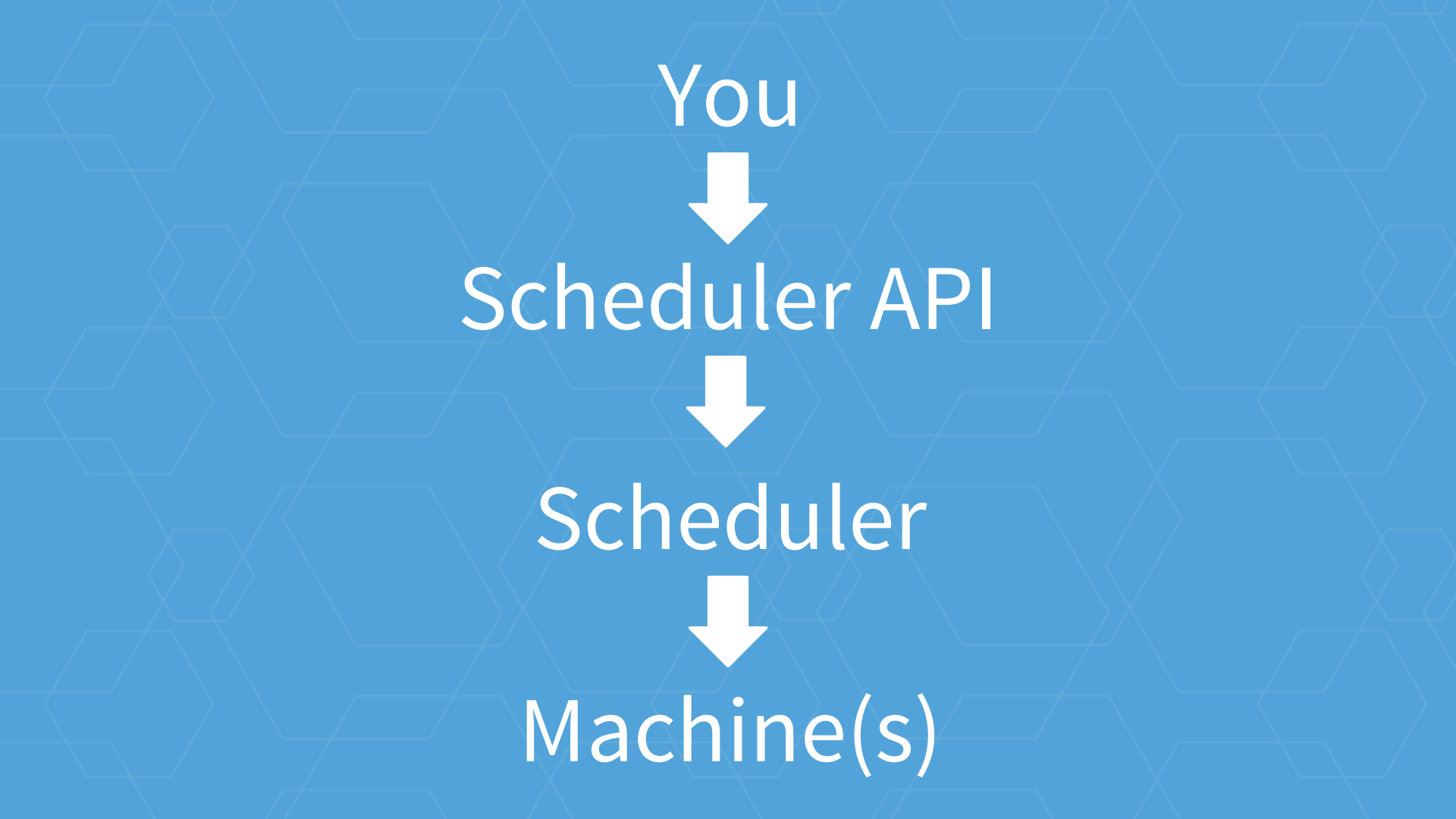
$ fab deploy:collector-app

$ fab deploy deploy:collector-app

$ fab lowest-loadaverage

$ fab lowest-loadaverage host1

```
$ fab lowest-loadaverage
host1
$ fab -H host1 deploy:job
```


You
Scheduler API
Scheduler
Machine(s)

```
while true {
  todo = diff(desState, curState)
  schedule(todo)
}
```

```
while true {
  todo = diff(desState, curState)
  schedule(todo)
}
```

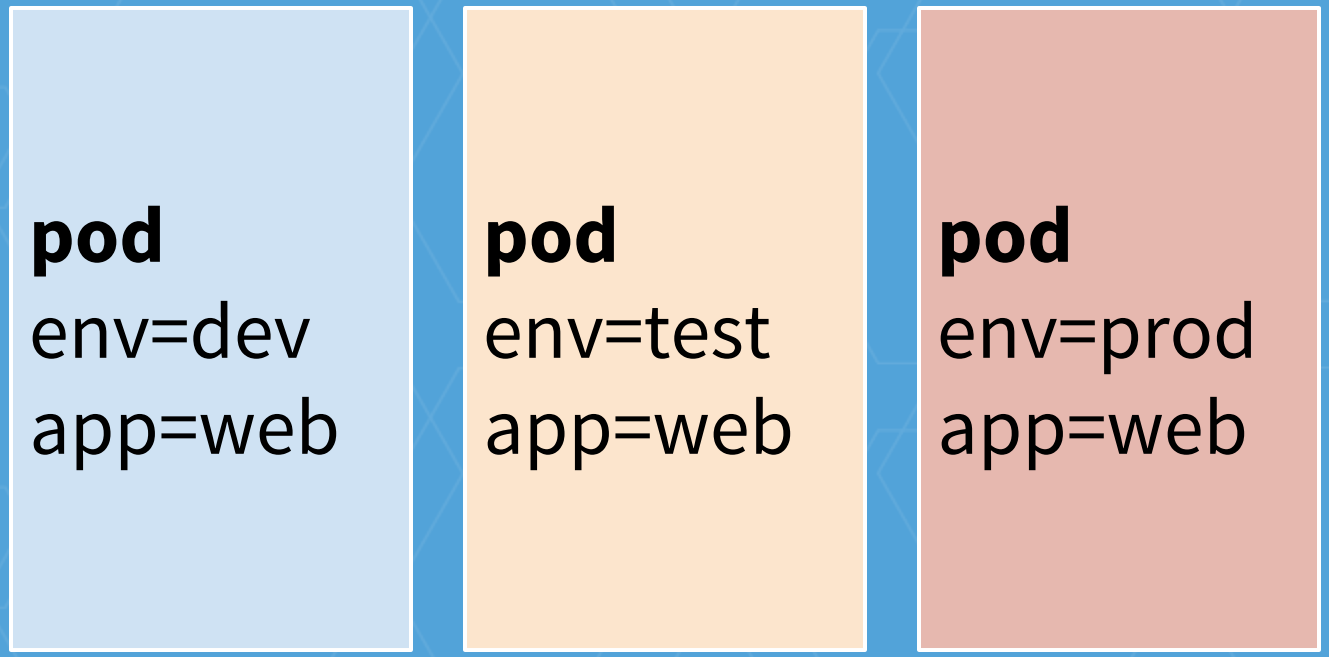
```
while true {
  todo = diff(desState, curState)
  schedule(todo)
}
```

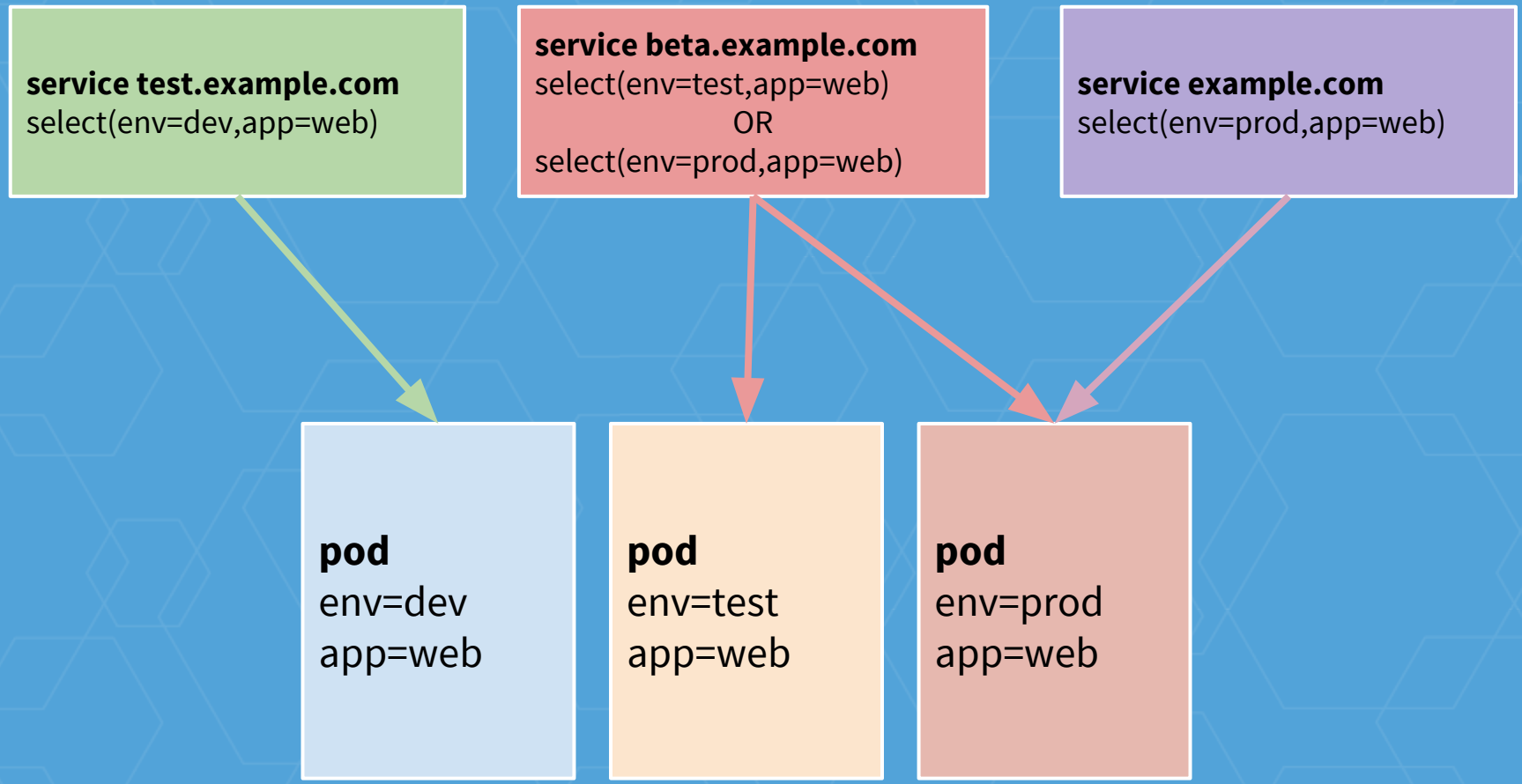
```
while true {
  todo = diff(desState, curState)
  schedule(todo)
}
```

services

dns, LBs, k8s labels

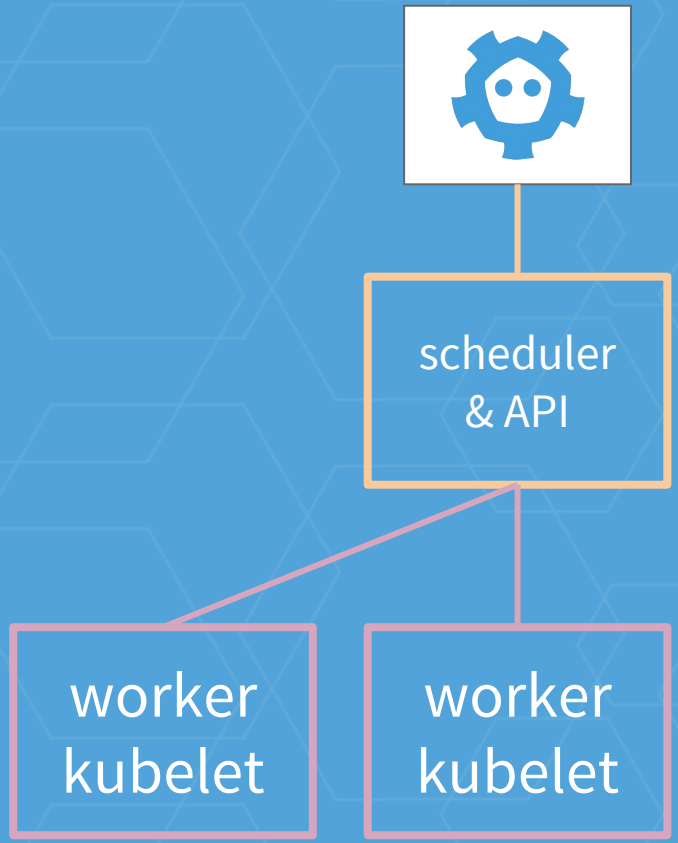
k8s labels
flexible service discovery

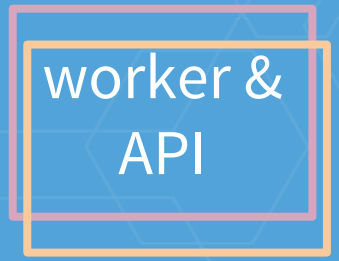

pod
env=dev
app=web
pod
env=test
app=web
pod
env=prod
app=web


service test.example.com
select(env=dev,app=web)
service beta.example.com
select(env=test,app=web)
OR
select(env=prod,app=web)
service example.com
select(env=prod,app=web)
pod
env=dev
app=web
pod
env=test
app=web
pod
env=prod
app=web

github.com/coreos/coreos-kubernetes

etcd.ngrok.io


scheduler
& API
worker
kubelet
worker
kubelet



works on 1 node too

work with us

coreos.com/careers

thank you

@coreoslinux
@tectonicstack
@brandonphilips