Containing Infrastructure
The Internet on Kubernetes

Brandon Philips

CTO & Co-Founder, CoreOS

@BrandonPhilips

3,424,000,000

Internet Users

Source: https://www.infoq.com/news/2014/01/IDC-software-developers

29,000,000

Software Developers and IT Practioners

CoreOS

Source: http://www.internetlivestats.com/internet-users/

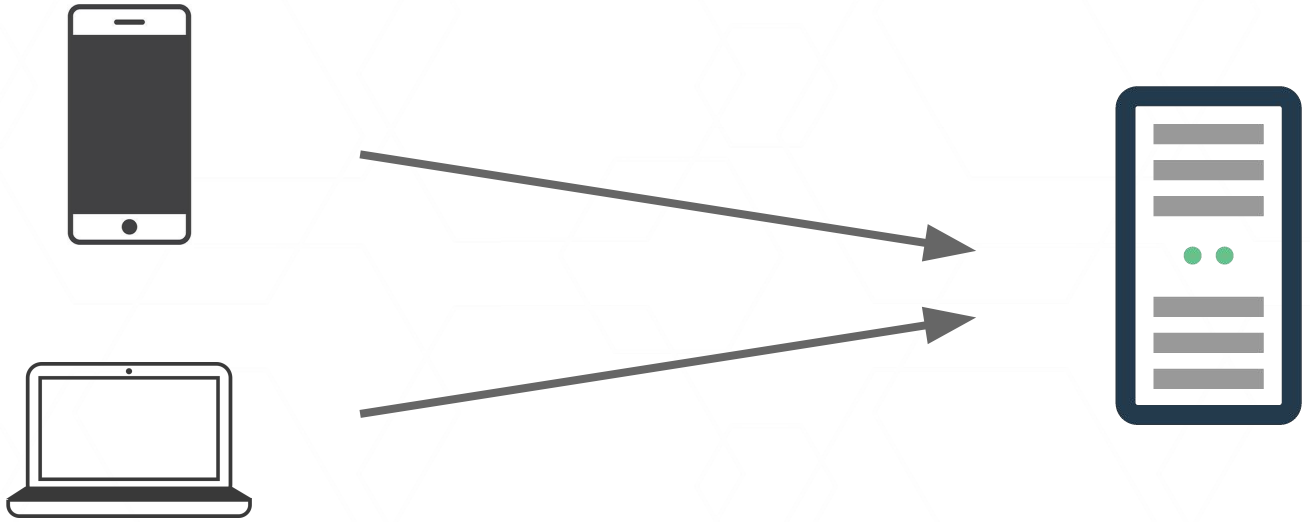
We are Outnumbered!

238,975,082

New Internet Users in 2016

CoreOS

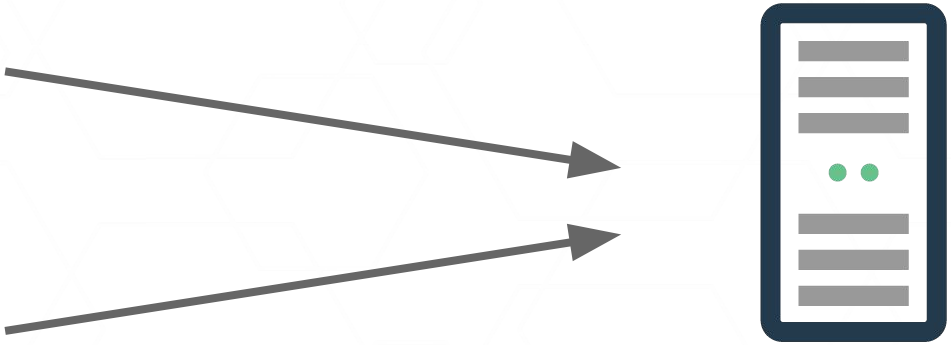
Source: http://www.internetlivestats.com/internet-users/



Documents

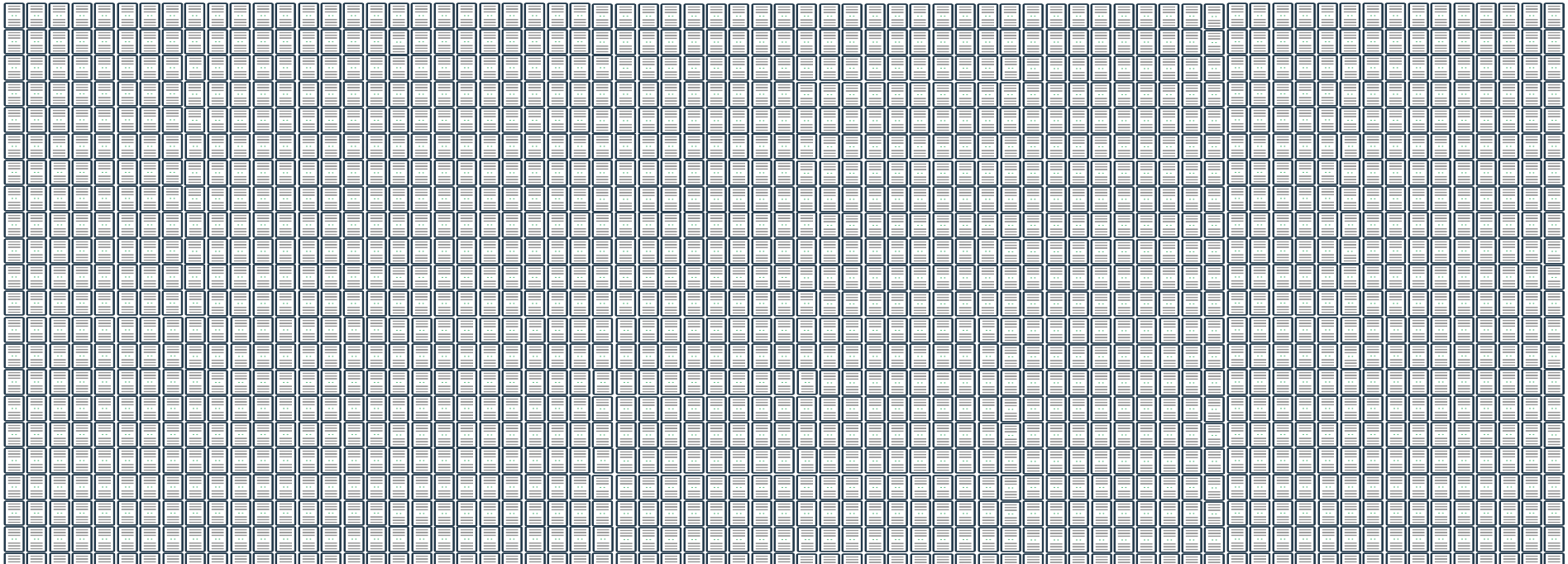
Commerce

Communications



~100,000,000

Servers Worldwide

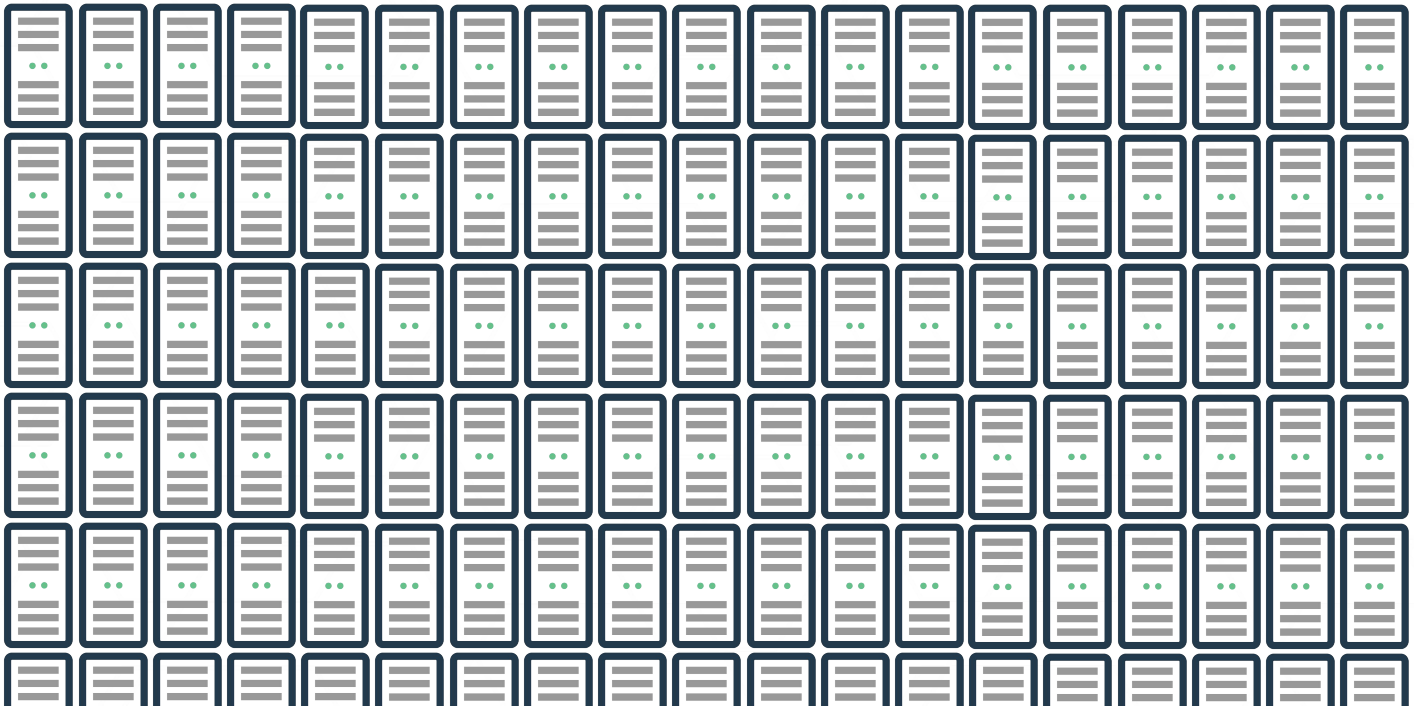


3 Per Person

In the Software and IT Industry

100+ Per Person

At the Internet Giants

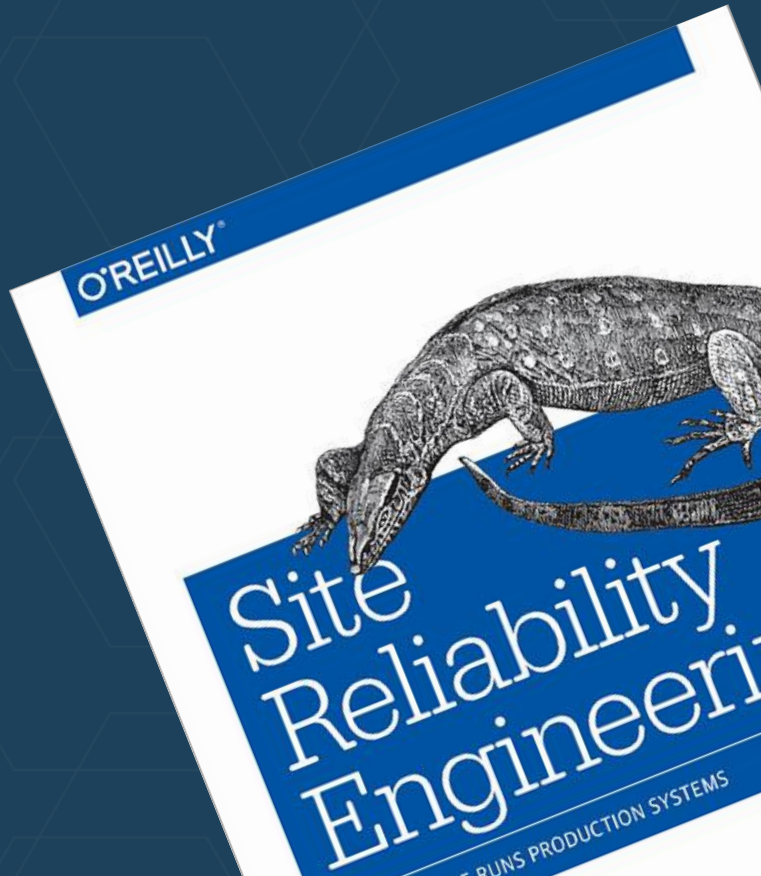


CoreOS

How do they do it?

Software Systems:
Containers, Clustering, Monitoring

Enabling Teams To:
Organize, Specialize, Take Risks



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

your

your

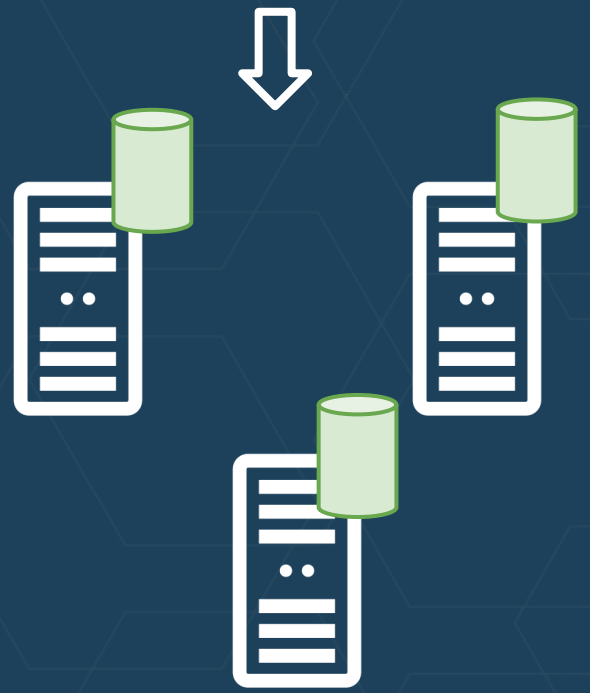
you as an ops engineer

your

quay.io/coreos/etcd
x3

your

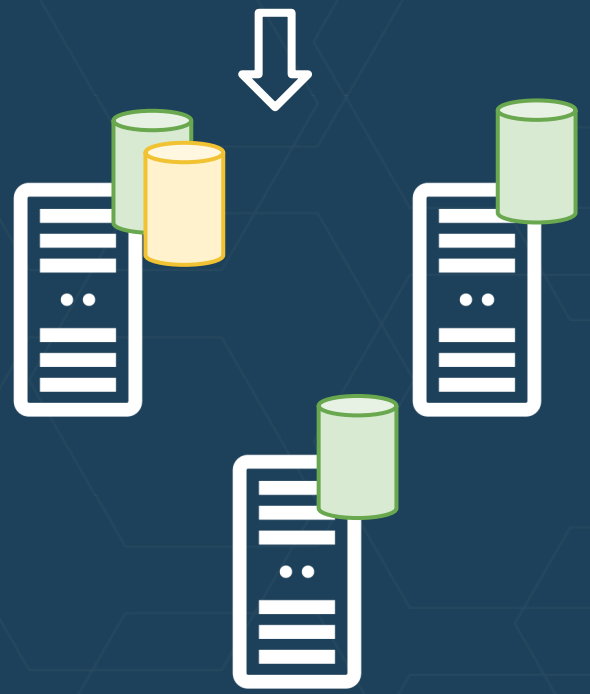
quay.io/coreos/etcd
x3



your

quay.io/coreos/prometheus
x1

your



quay.io/coreos/dex
x2

your

Containers

Application Packaging

```
package main

import "fmt"

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

Building and Hosting

```
$ git clone git://github.com/coreos/etcd

$ docker build . -t quay.io/coreos/etcd:v3.1.0

$ docker push quay.io/coreos/etcd:v3.1.0
```

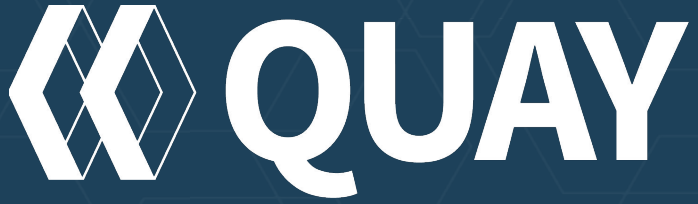
Dockerfile Build System

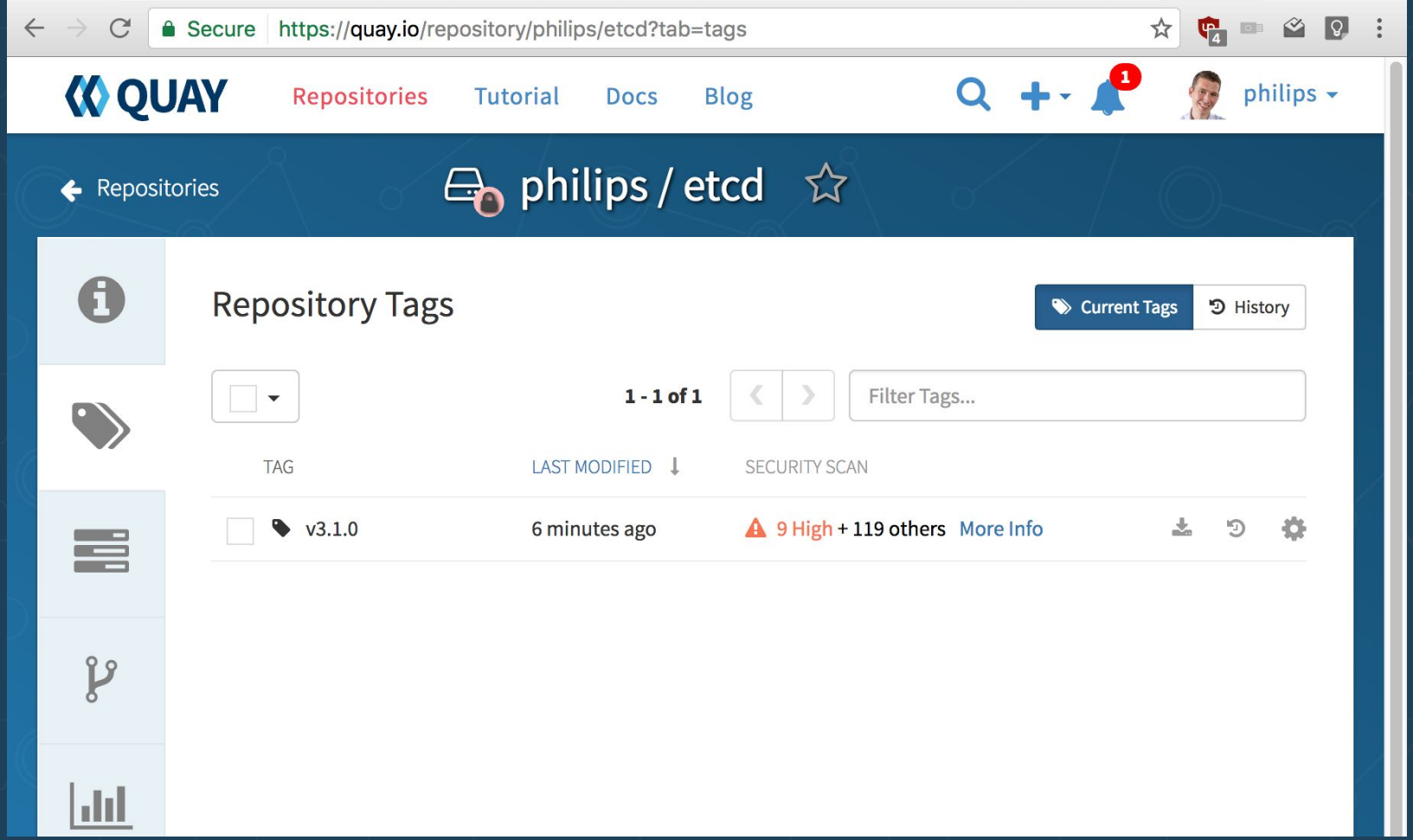
```
$ cat Dockerfile
FROM golang
ADD . /go/src/github.com/coreos/etcd
ADD cmd/vendor /go/src/github.com/coreos/e...
RUN go install github.com/coreos/etcd
EXPOSE 2379 2380
ENTRYPOINT ["etcd"]
```


container
image


/bin/etcd
/lib/libc


quay.io/philips/etcd


QUAY


Secure
https://quay.io/repository/philips/etcd?tab=tags
QUAY
Repositories
Tutorial
Docs
Blog
1
philips
Repositories
philips / etcd
Repository Tags
Current Tags
History
1 - 1 of 1
Filter Tags...
TAG
LAST MODIFIED
SECURITY SCAN
v3.1.0
6 minutes ago
9 High + 119 others
More Info

Running Containers

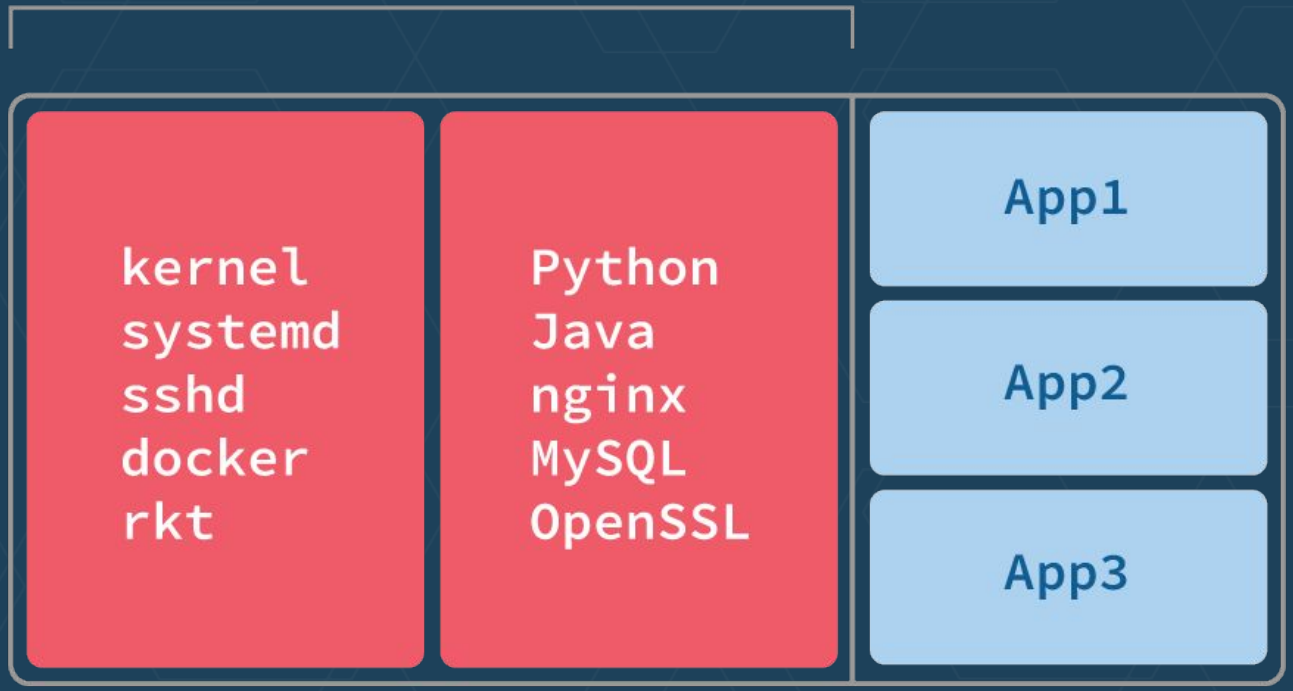
Application Packaging

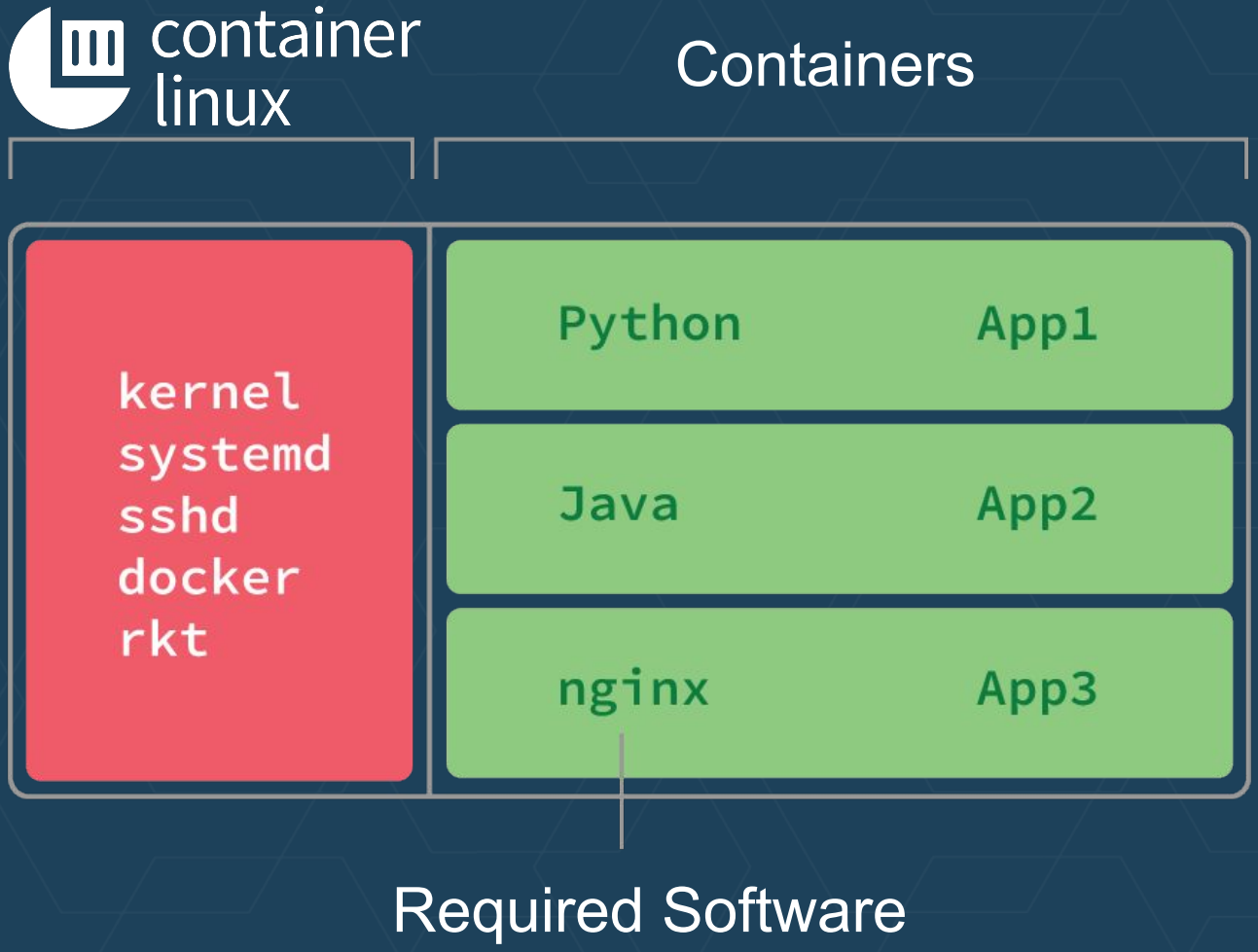
Normal Linux Processes!

```
$ docker run -d quay.io/philips/etcd:v3.1.0 -d
$ pidof etcd
24725
$ cat /proc/24725/mounts | grep overlay
/ overlay workdir=/var/lib/docker/overlay/7134
```

Abstract away app from the OS

Traditional Linux




container linux
Containers
kernel
systemd
sshd
docker
rkt
Python
App1
Java
App2
nginx
App3
Required Software

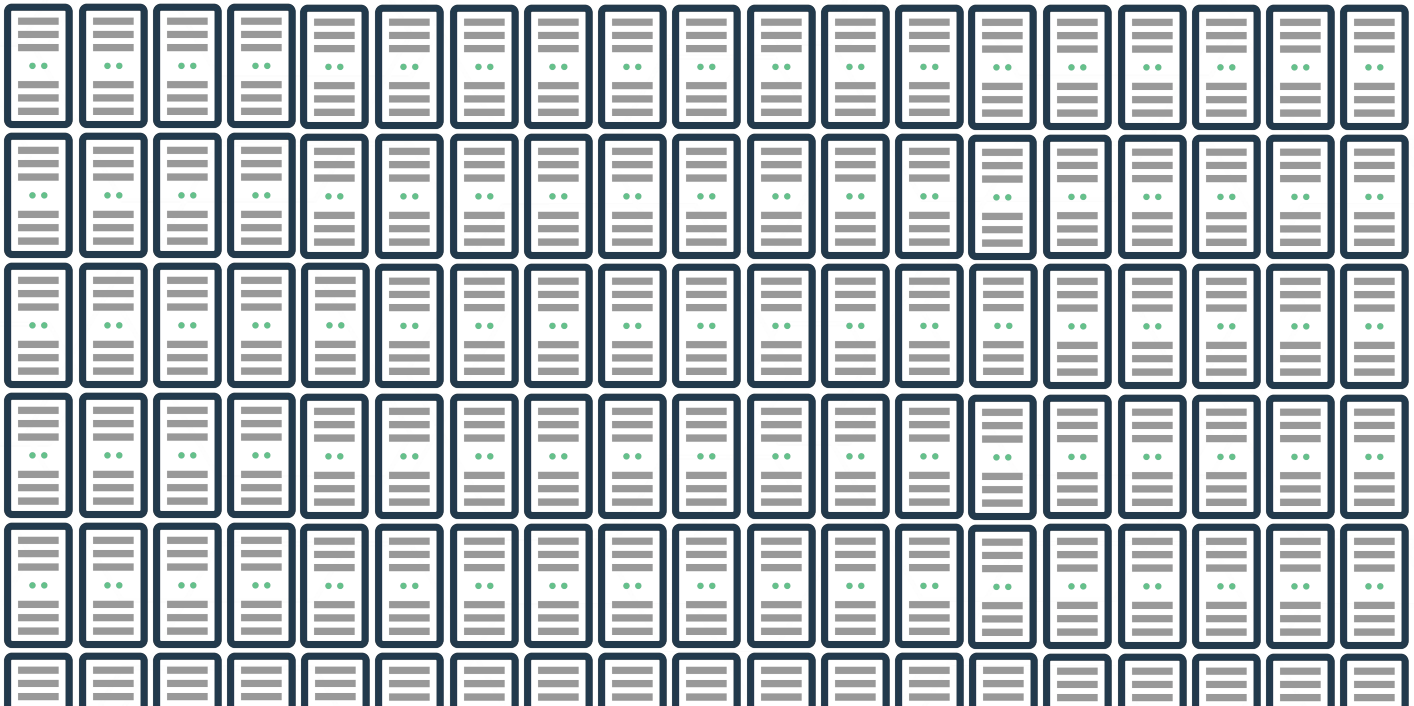
Clustering

Clustering

Server Botnets!

100+ Per Person

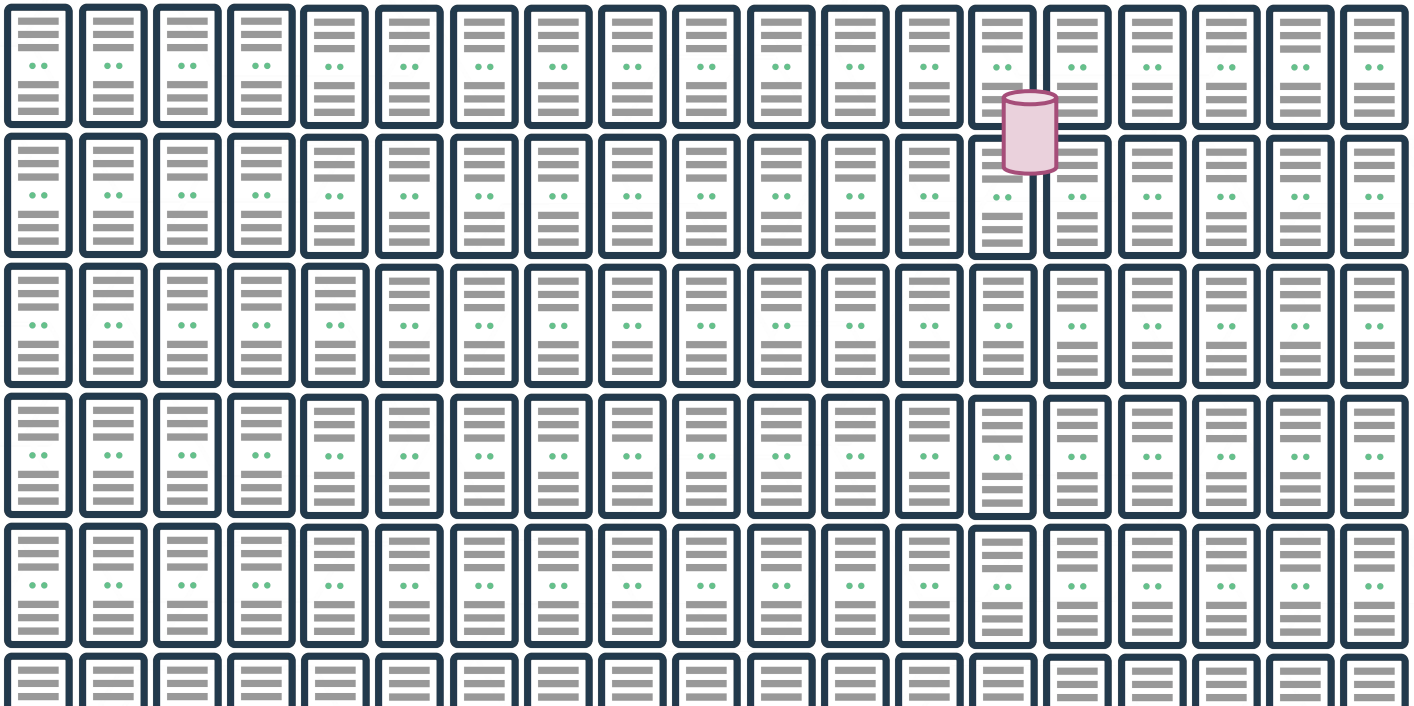
At the Internet Giants



CoreOS

100+ Per Person

Too many for manual placement



CoreOS

100+ Per Person

Too many for manual placement

CoreOS

100+ Per Person

Too many for manual placement

CoreOS

```
$ while read host; ssh $host ... < hosts
```

???

```
$ while read host; ssh $host … < hosts
```

???

```
$ while read host; ssh $host ... < hosts
```

Problems: No monitoring, no state to recover

```
$ kubectl run --replicas=3 quay.io/coreos/dex
```

CoreOS

```
$ kubectl run --replicas=3 quay.io/coreos/dex
```

Solution: Monitoring, and state on computers

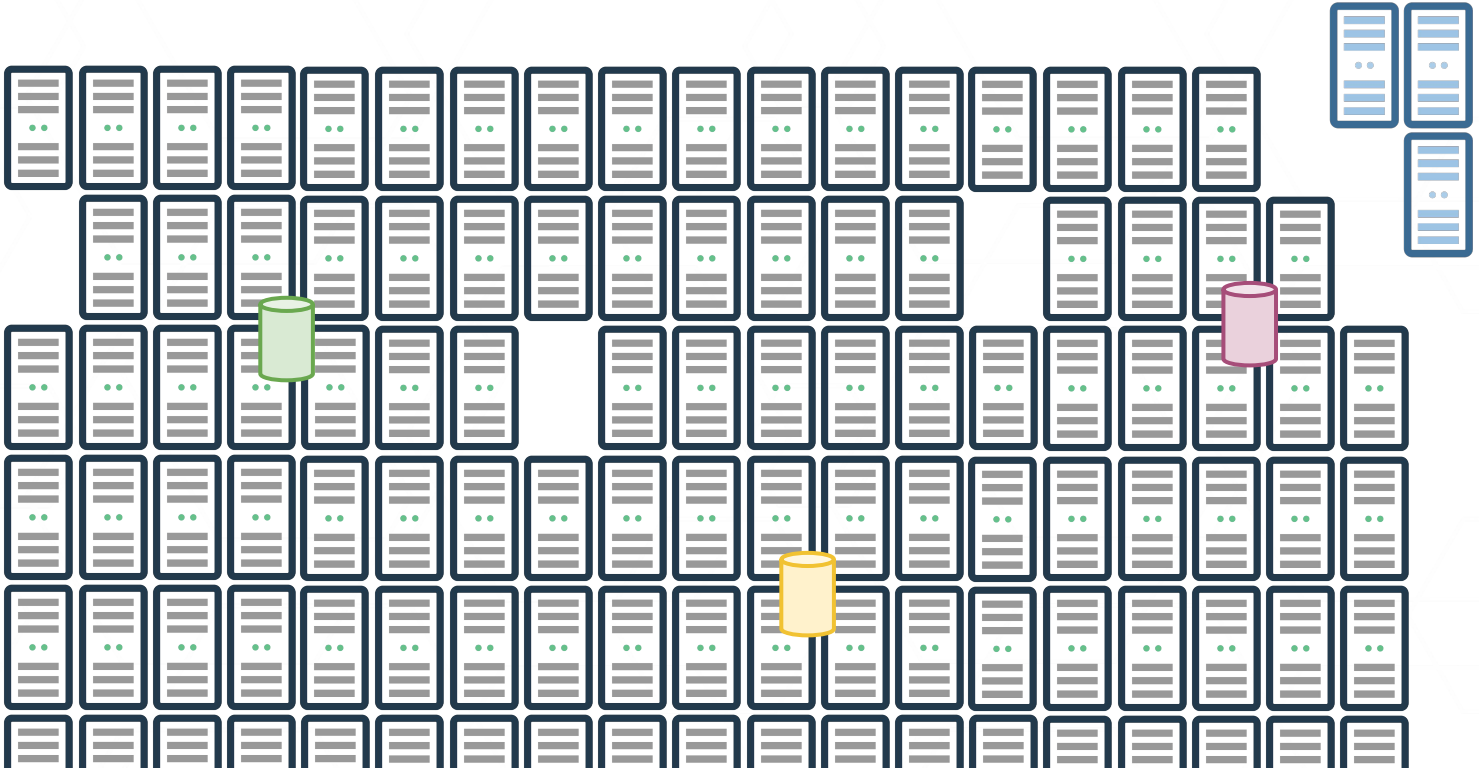
CoreOS

```
$ kubectl run --replicas=3 quay.io/coreos/dex
```

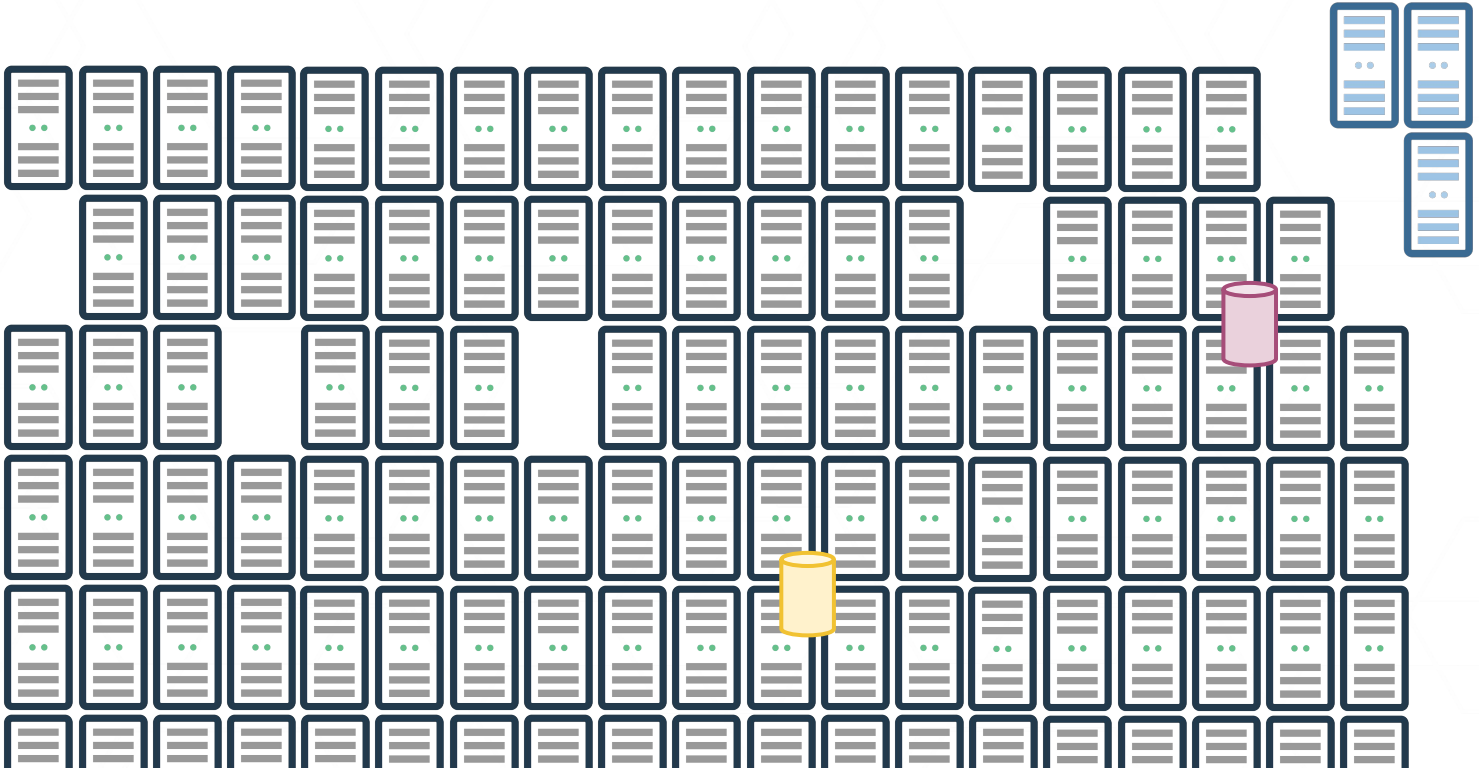
```
$ kubectl run --replicas=3 quay.io/coreos/dex
```

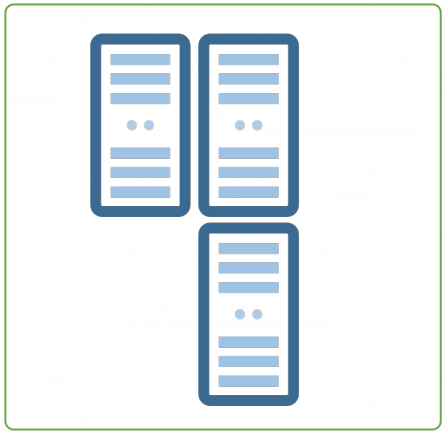
CoreOS

```
$ kubectl run --replicas=3 quay.io/coreos/dex
```



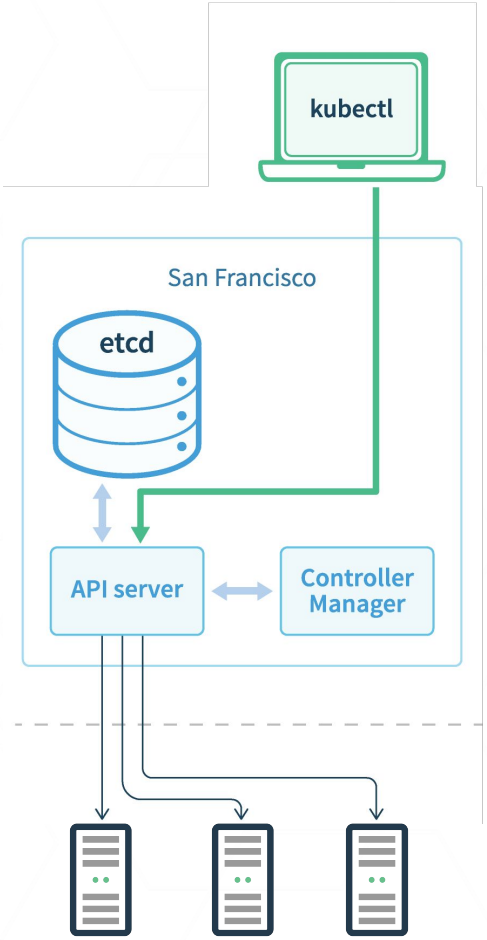
CoreOS





???

???



Simple cluster operations

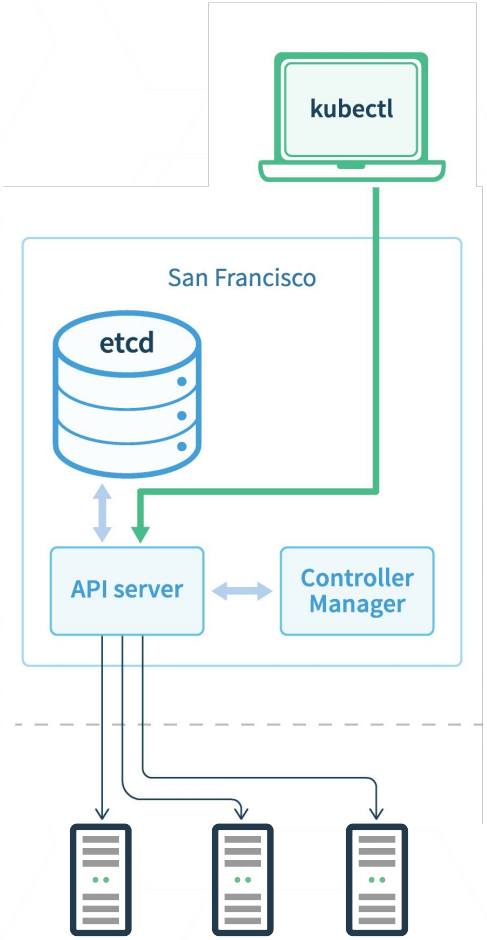
Secure and Simple API

Friendly operational tools

CoreOS

Clustering

Facing Failure

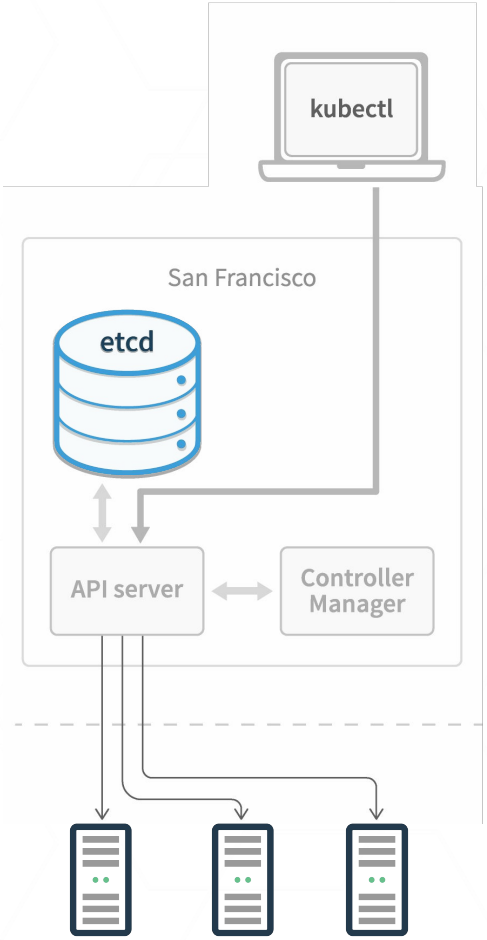


Simple cluster operations

Secure and Simple API

Friendly operational tools

CoreOS



etcd Overview

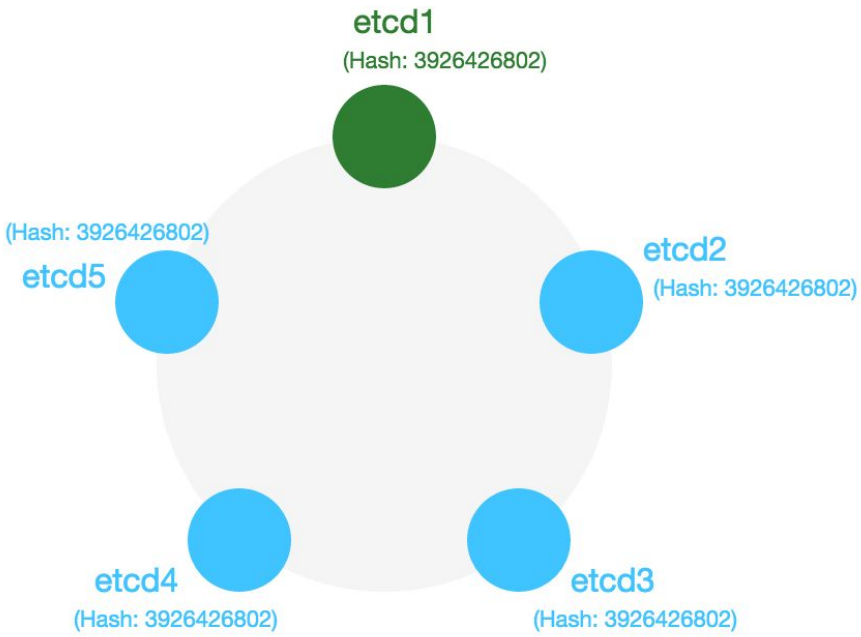
Introduced in 2013 by CoreOS

Persistent database of Kubernetes

Auto-leader election for availability

CoreOS

Play etcd Home Code Issues



etcd1 etcd2 etcd3 etcd4 etcd5

PUT GET DELETE Submit

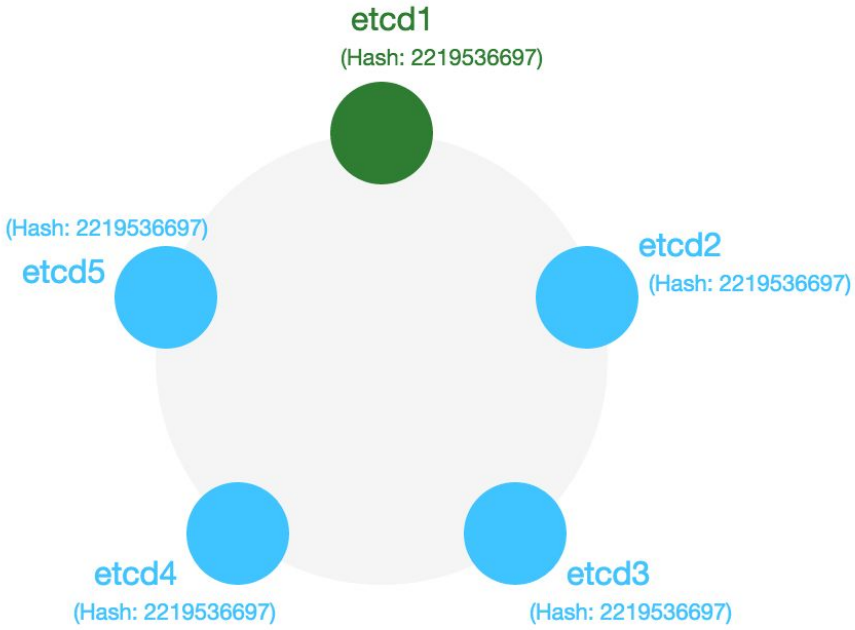
foo

BAZ

[PUT] success! "foo" : "BAZ" (at 2016-03-16 05:36:42 PST)

Log

Play etcd Home Code Issues



etcd1 etcd2 etcd3 etcd4 etcd5

PUT GET DELETE Submit

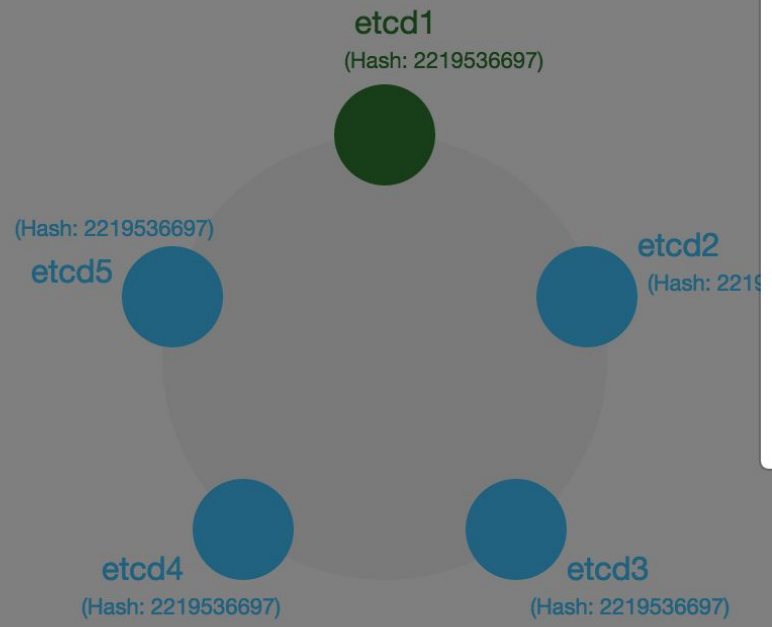
foo

bar

[PUT] success! "foo" : "bar" (at 2016-03-16 05:38:17 PST)

Log

Play etcd Home Code Issues



etcd1

Kill Restart

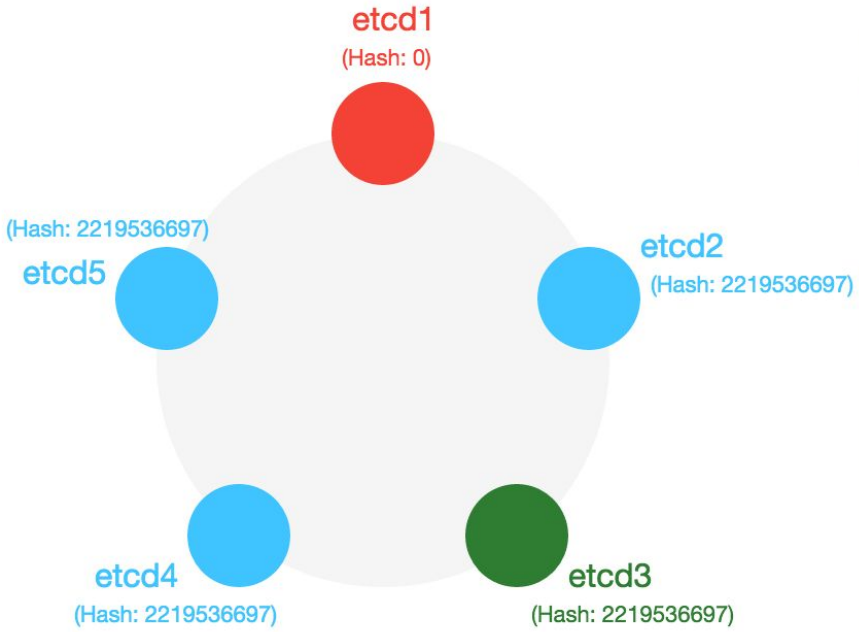
ID: b414e782d52bc45a
Endpoint: 10.128.0.2:1278
State: Leader
Number of Keys: 5
Hash: 2219536697

etcd5

[PUT] success! "foo" : "bar" (at 2016-03-16 05:38:17 PST)

Log

Play etcd Home Code Issues



etcd1 etcd2 etcd3 etcd4 etcd5

PUT GET DELETE Submit

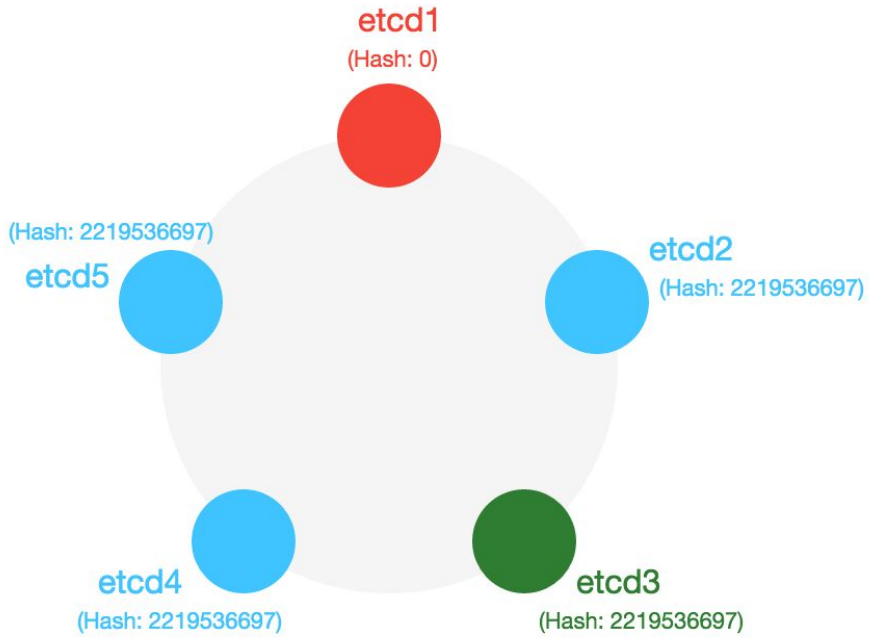
foo

bar

[PUT] success! "foo" : "bar" (at 2016-03-16 05:38:17 PST)

Log

Play etcd Home Code Issues



etcd1 etcd2 etcd3 etcd4 etcd5

PUT GET DELETE Submit

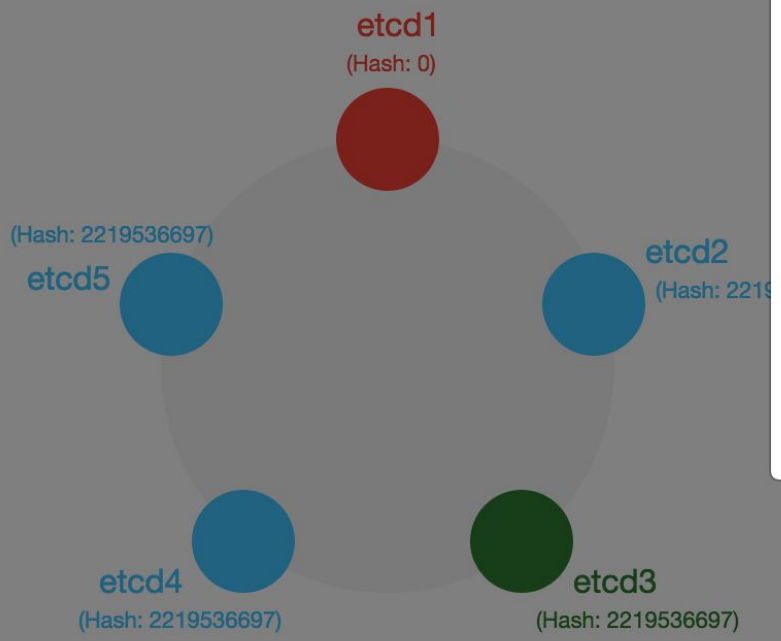
foo

BAZ

[PUT] success! "foo" : "bar" (at 2016-03-16 05:38:17 PST)

Log

Play etcd Home Code Issues



etcd1

Kill Restart

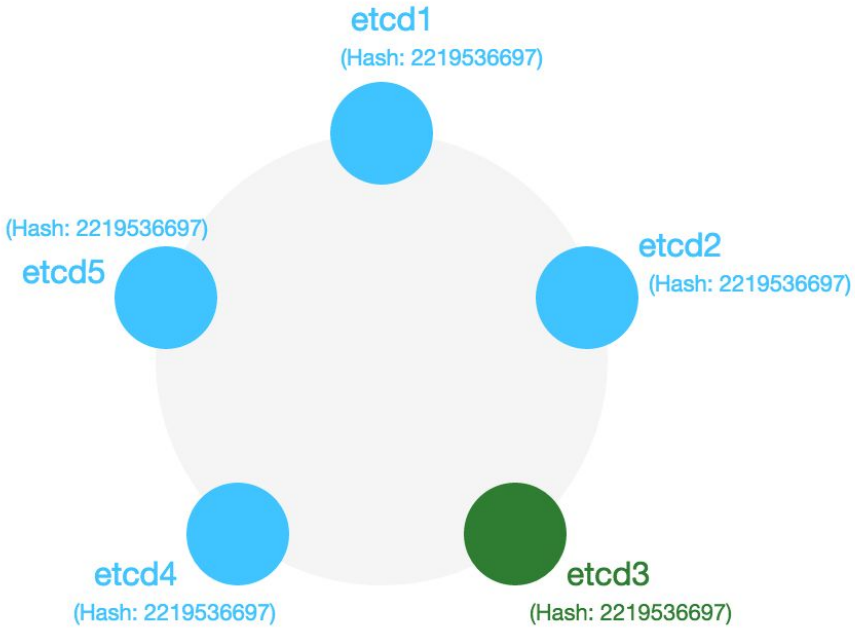
ID: unknown
Endpoint: 10.128.0.2:1278
State: unreachable
Number of Keys: 0
Hash: 0

etcd5

[PUT] error grpc: timed out trying to connect (key: "foo")

Log

Play etcd Home Code Issues



etcd1 | etcd2 | etcd3 | etcd4 | etcd5

PUT | GET | DELETE | Submit

foo

BAZ

[PUT] error grpc: timed out trying to connect (key: "foo")

Log


etcd


etcd

Kubernetes Everywhere

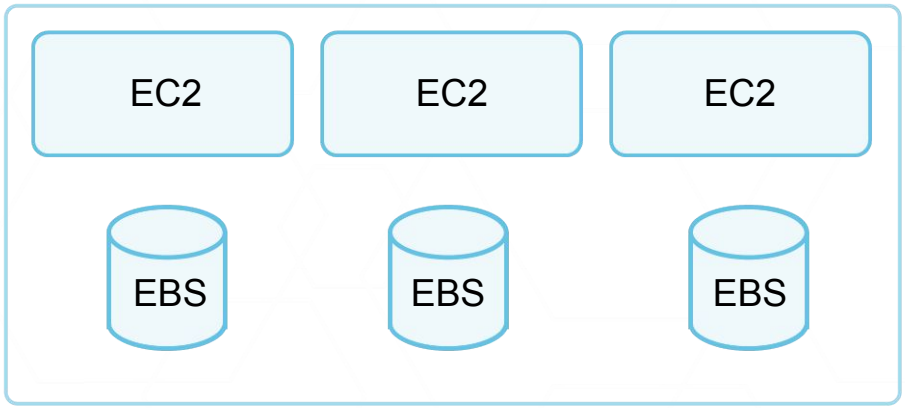
Consistent Infrastructure Everywhere

kubectl

k8s API

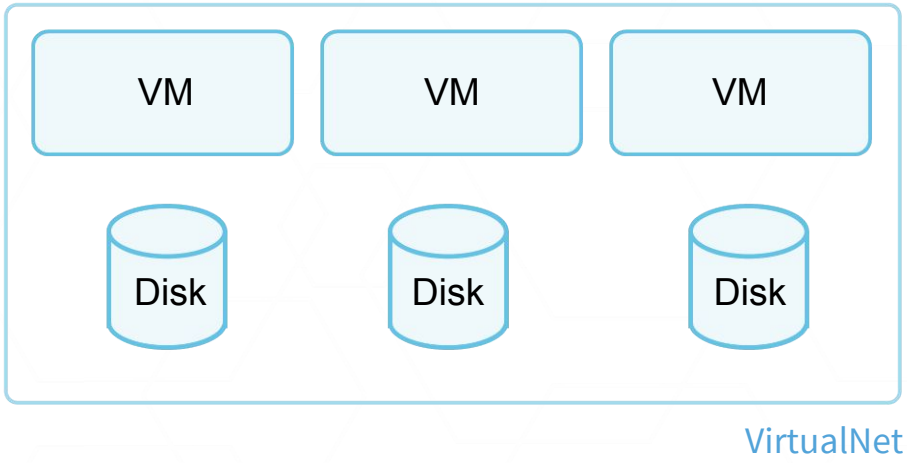
AWS APIs

kubernetes v1.5.2



kubectl

kubernetes v1.5.2



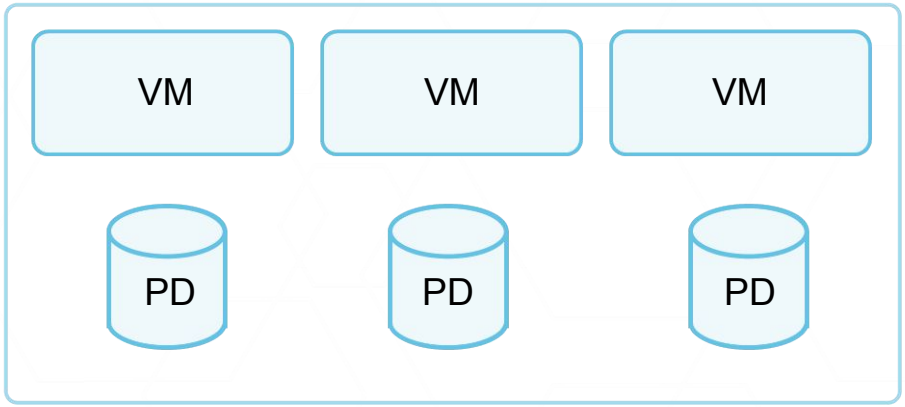
kubectl

k8s API

Google APIs

kubernetes

v1.5.2



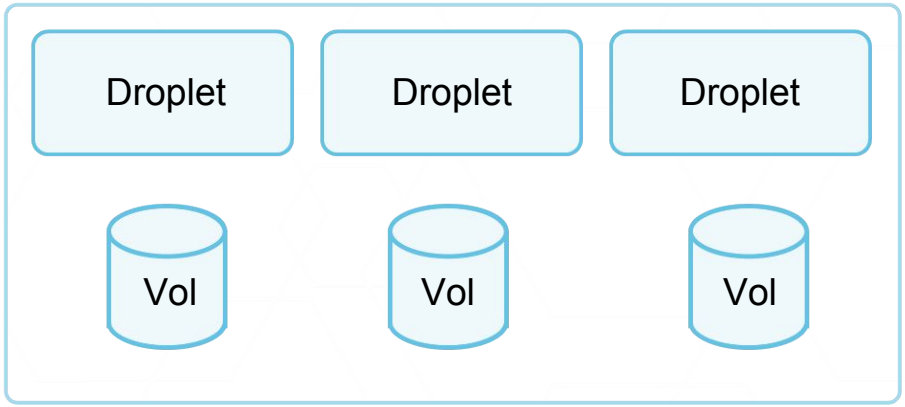
VirtualNet

kubectl

k8s API

Digitalocean APIs

kubernetes v1.5.2



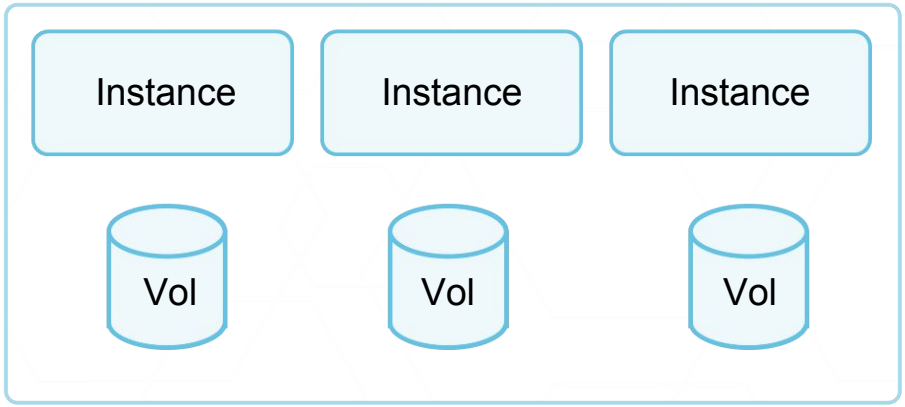
kubectl

k8s API

OpenStack APIs

kubernetes

v1.5.2



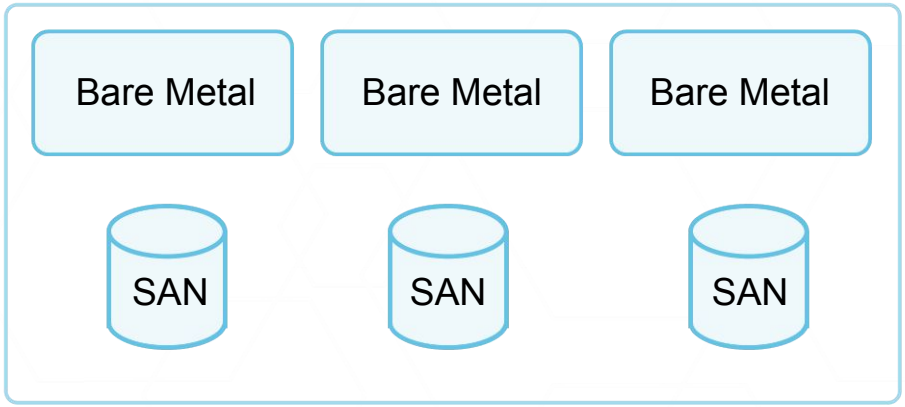
VirtualNet

kubectl

k8s API

kubernetes

v1.5.2



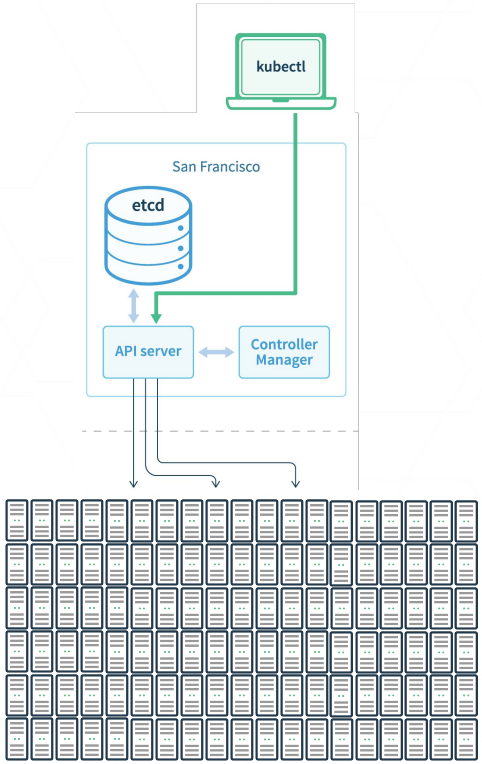
Ethernet

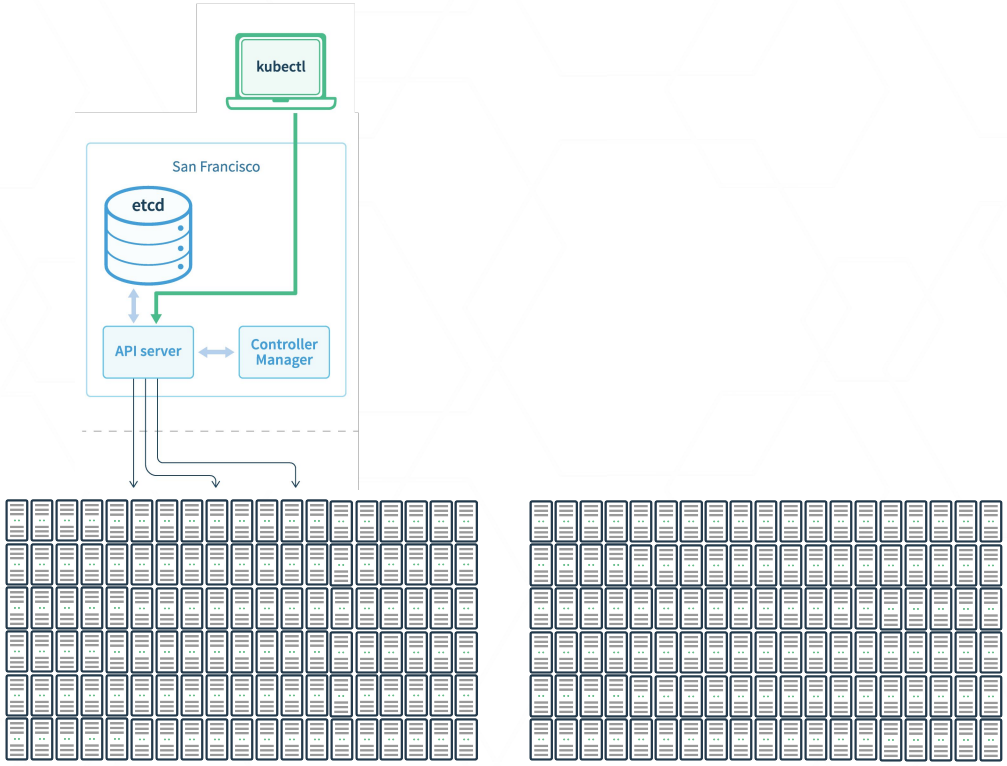
Consistency on all major components

- **Compute:** Physical, Virtual Machine, Cloud
- **Networking:** VXLAN, BGP, IPIP, forwarding, etc
- **Storage:** EBS, NFS, GlusterFS, Cinder, etc
- **Load Balancing:** ELB, nginx, Cloud LB, etc

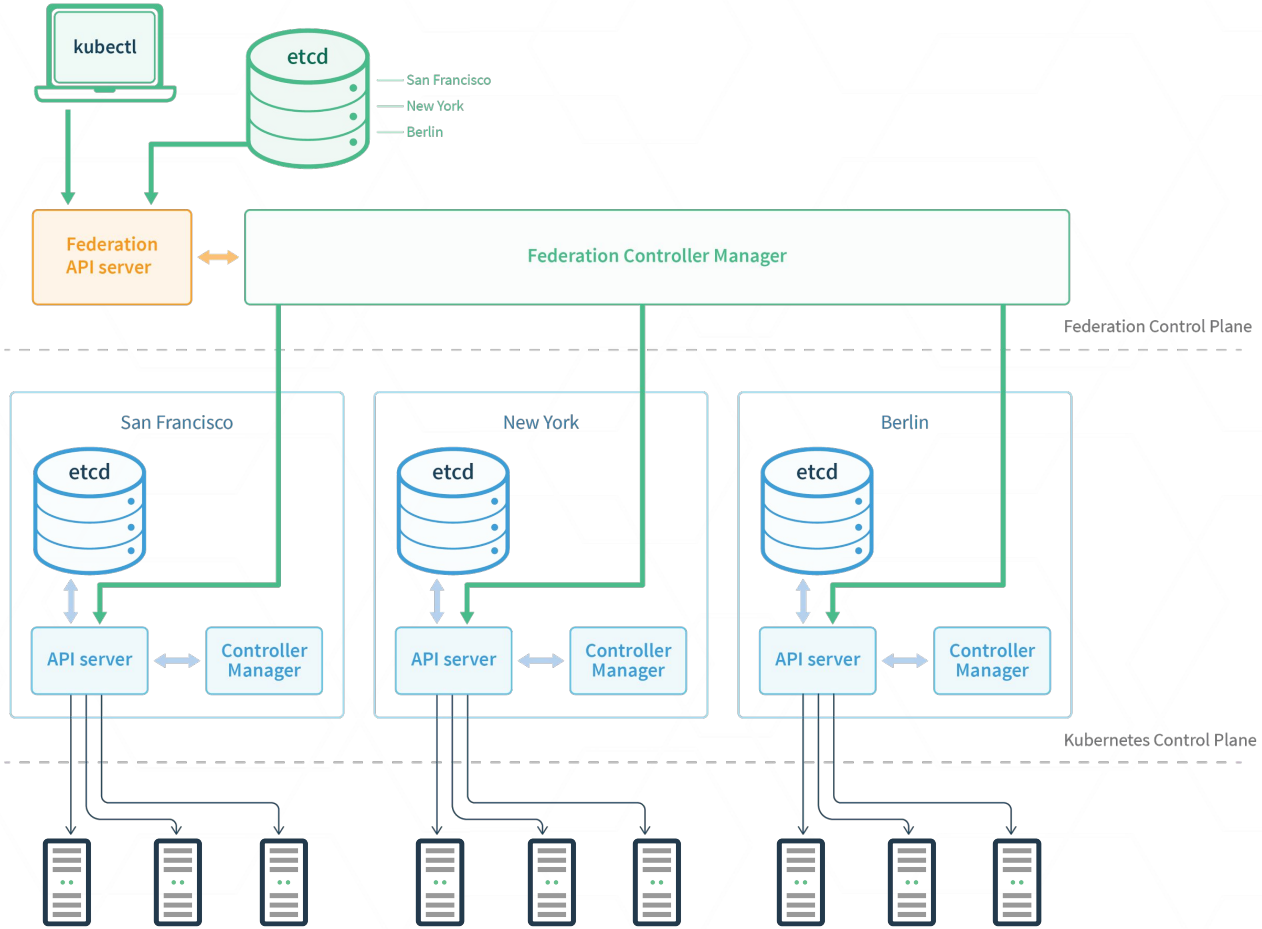
Federation

Handling Too Much Success

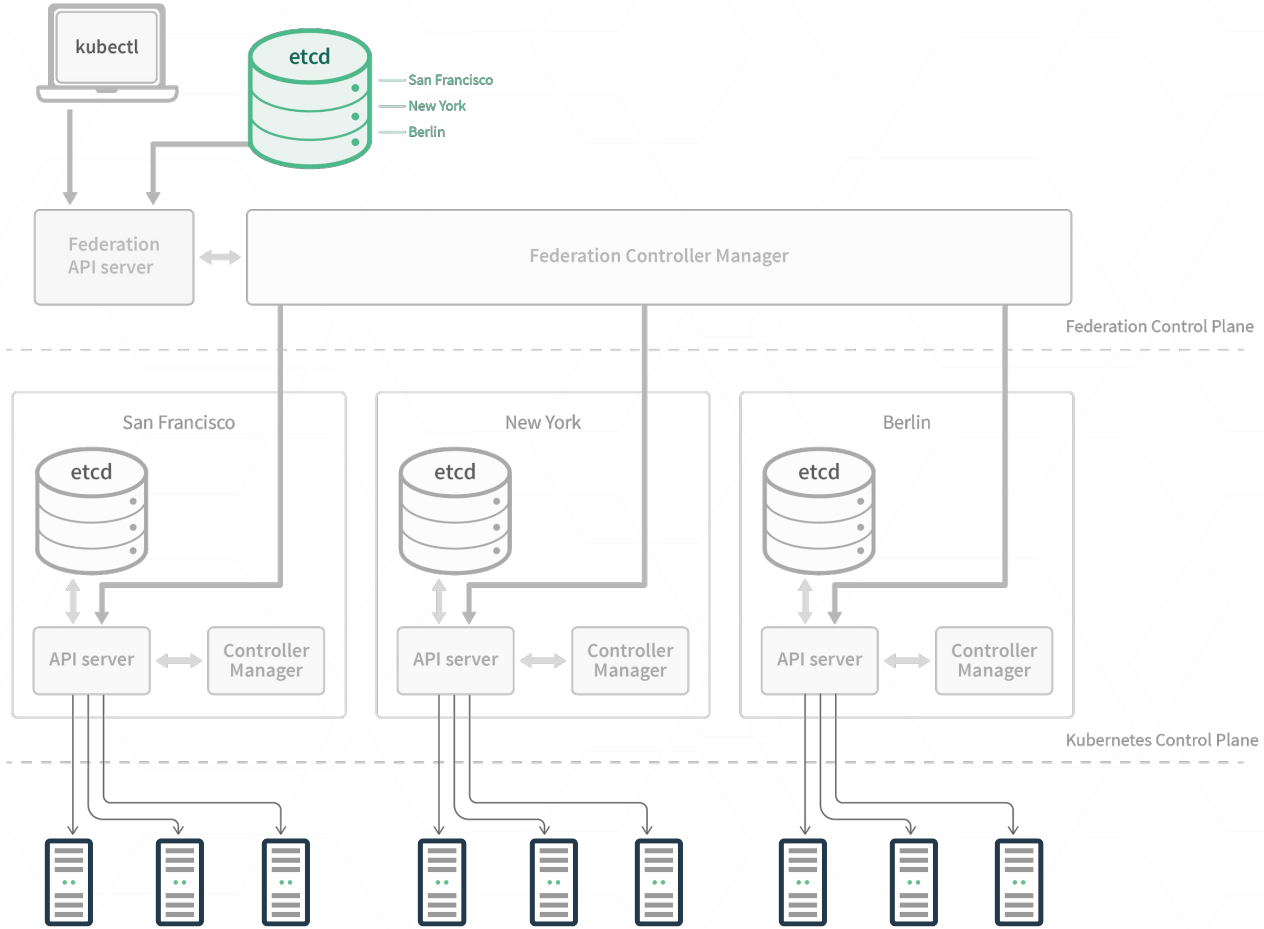

kubectl
San Francisco
etcd
API server
Controller Manager


kubectl
San Francisco
etcd
API server
Controller Manager

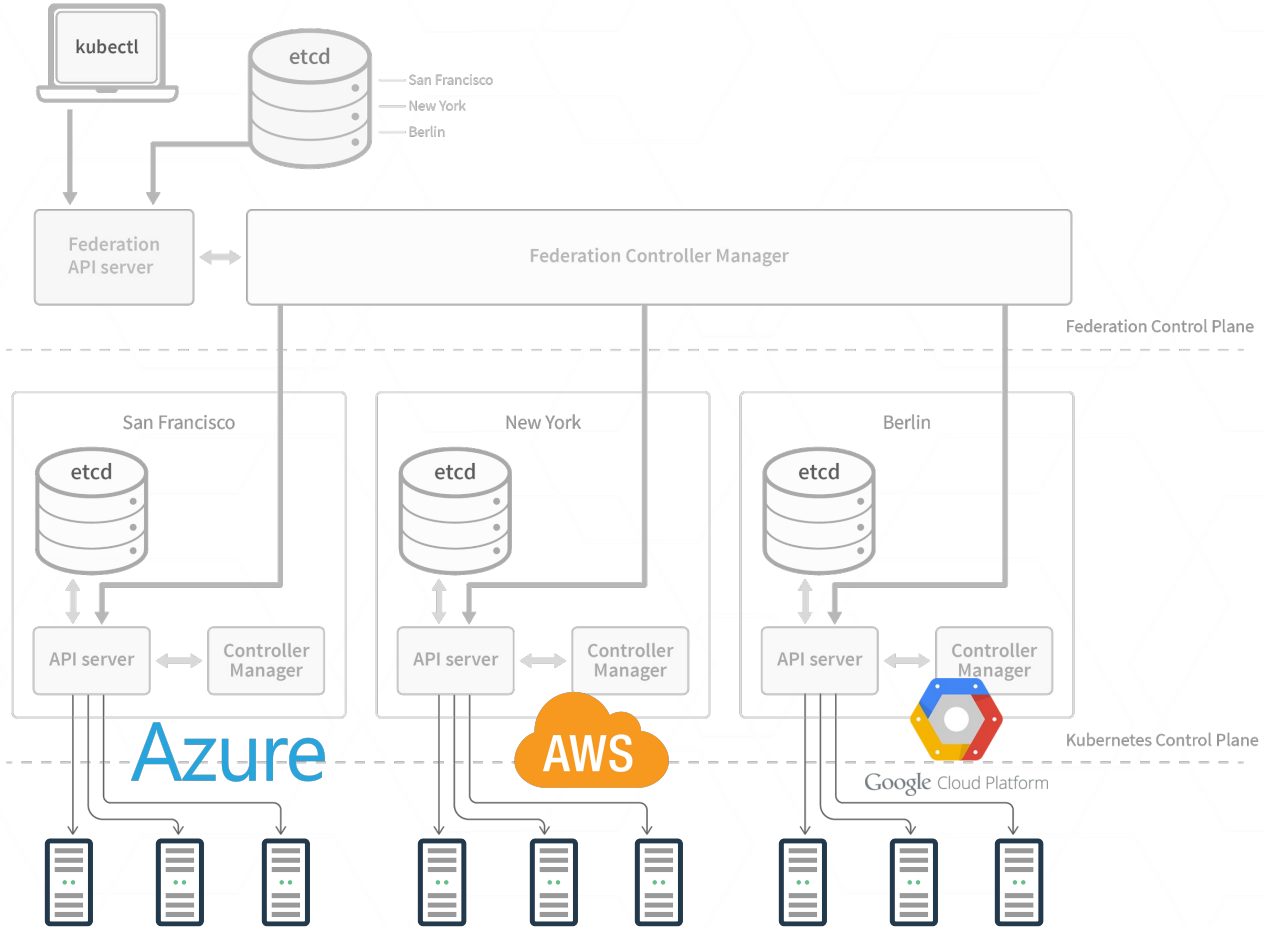

kubectl
San Francisco
etcd
API server
Controller Manager

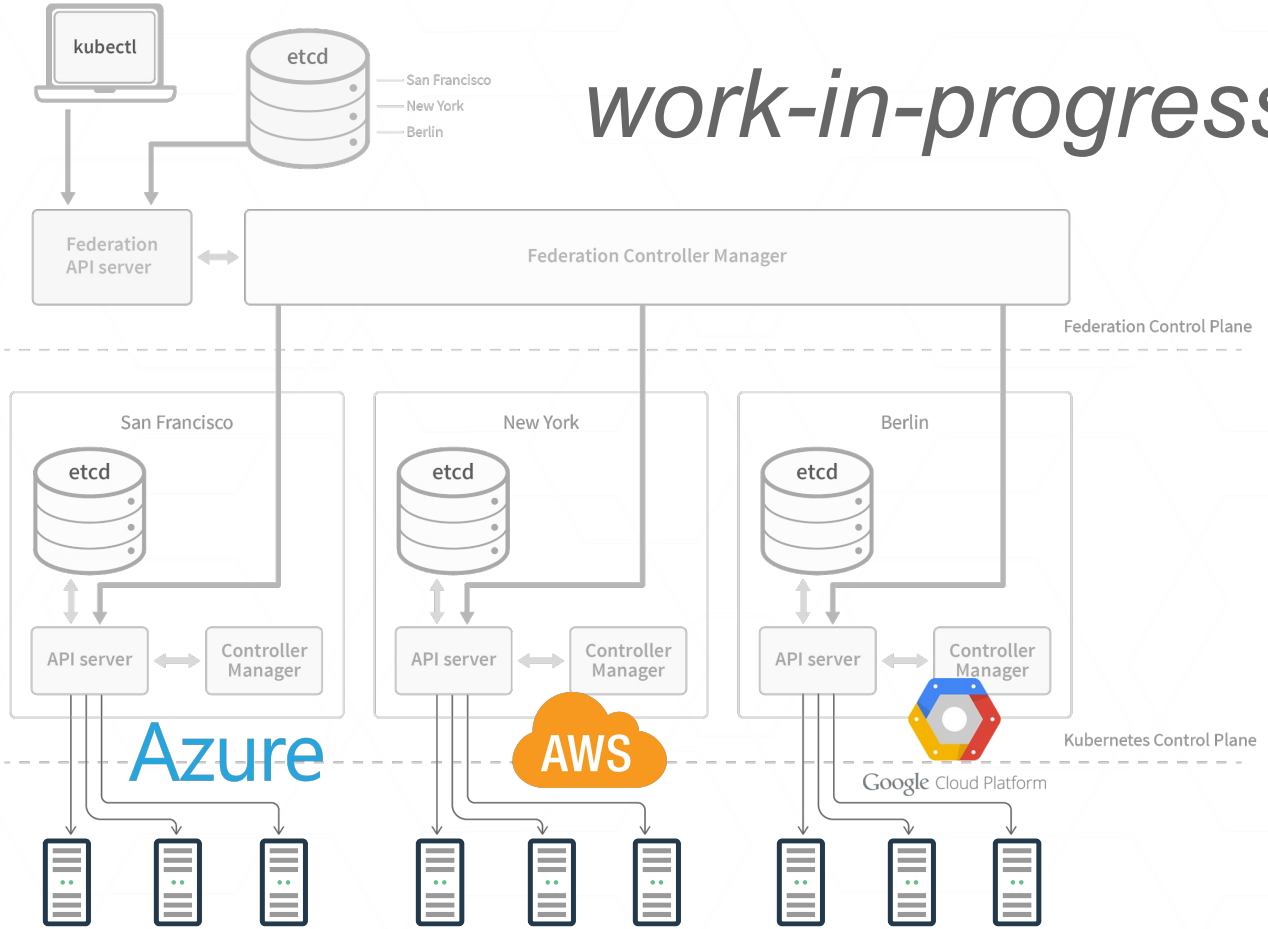

kubectl
etcd
San Francisco
New York
Berlin
Federation API server
Federation Controller Manager
Federation Control Plane
San Francisco
New York
Berlin
etcd
etcd
etcd
API server
Controller Manager
API server
Controller Manager
API server
Controller Manager
Kubernetes Control Plane

CoreOS


kubectl
etcd
San Francisco
New York
Berlin
Federation API server
Federation Controller Manager
Federation Control Plane
San Francisco
New York
Berlin
etcd
etcd
etcd
API server
Controller Manager
API server
Controller Manager
API server
Controller Manager
Kubernetes Control Plane

CoreOS


kubectl
etcd
San Francisco
New York
Berlin
Federation API server
Federation Controller Manager
Federation Control Plane
San Francisco
New York
Berlin
etcd
etcd
etcd
API server
Controller Manager
API server
Controller Manager
API server
Controller Manager
Azure
AWS
Google Cloud Platform
Kubernetes Control Plane


kubectl
etcd
San Francisco
New York
Berlin
work-in-progress
Federation API server
Federation Controller Manager
Federation Control Plane
San Francisco
New York
Berlin
etcd
etcd
etcd
API server
Controller Manager
API server
Controller Manager
API server
Controller Manager
Azure
AWS
Google Cloud Platform
Kubernetes Control Plane

CoreOS

Service Discovery

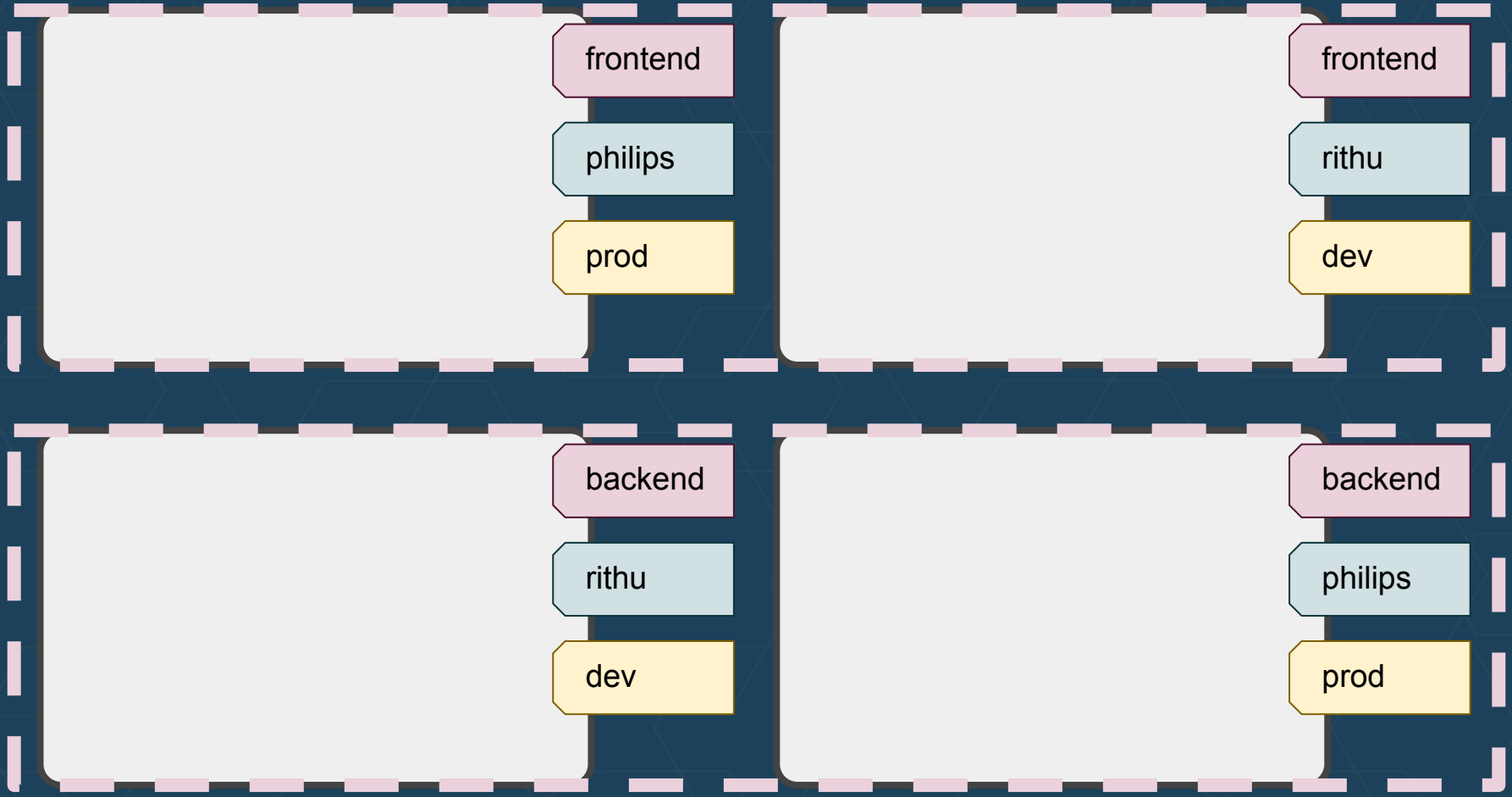
Labels, The Kubernetes Way

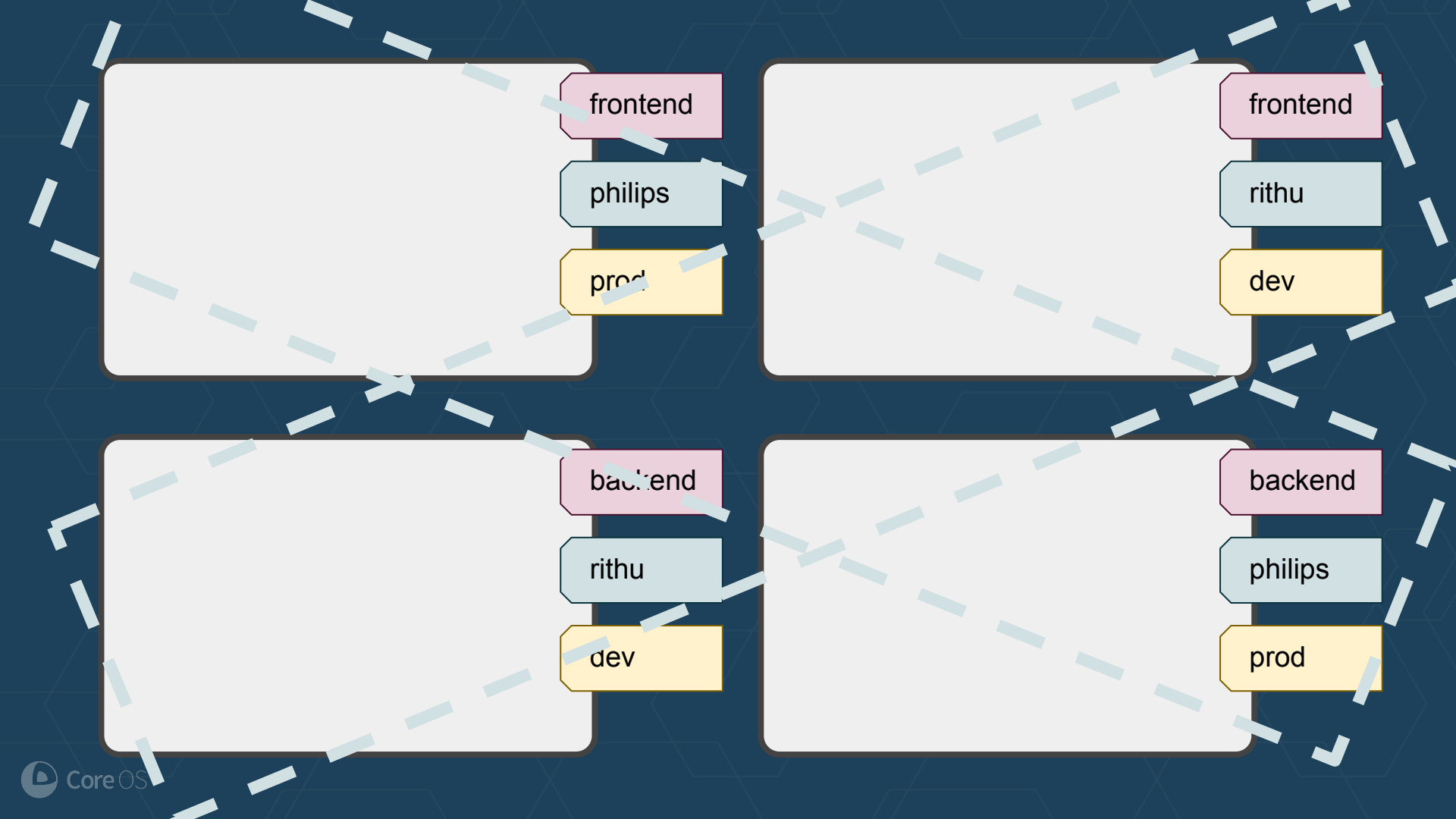

frontend
philips
prod

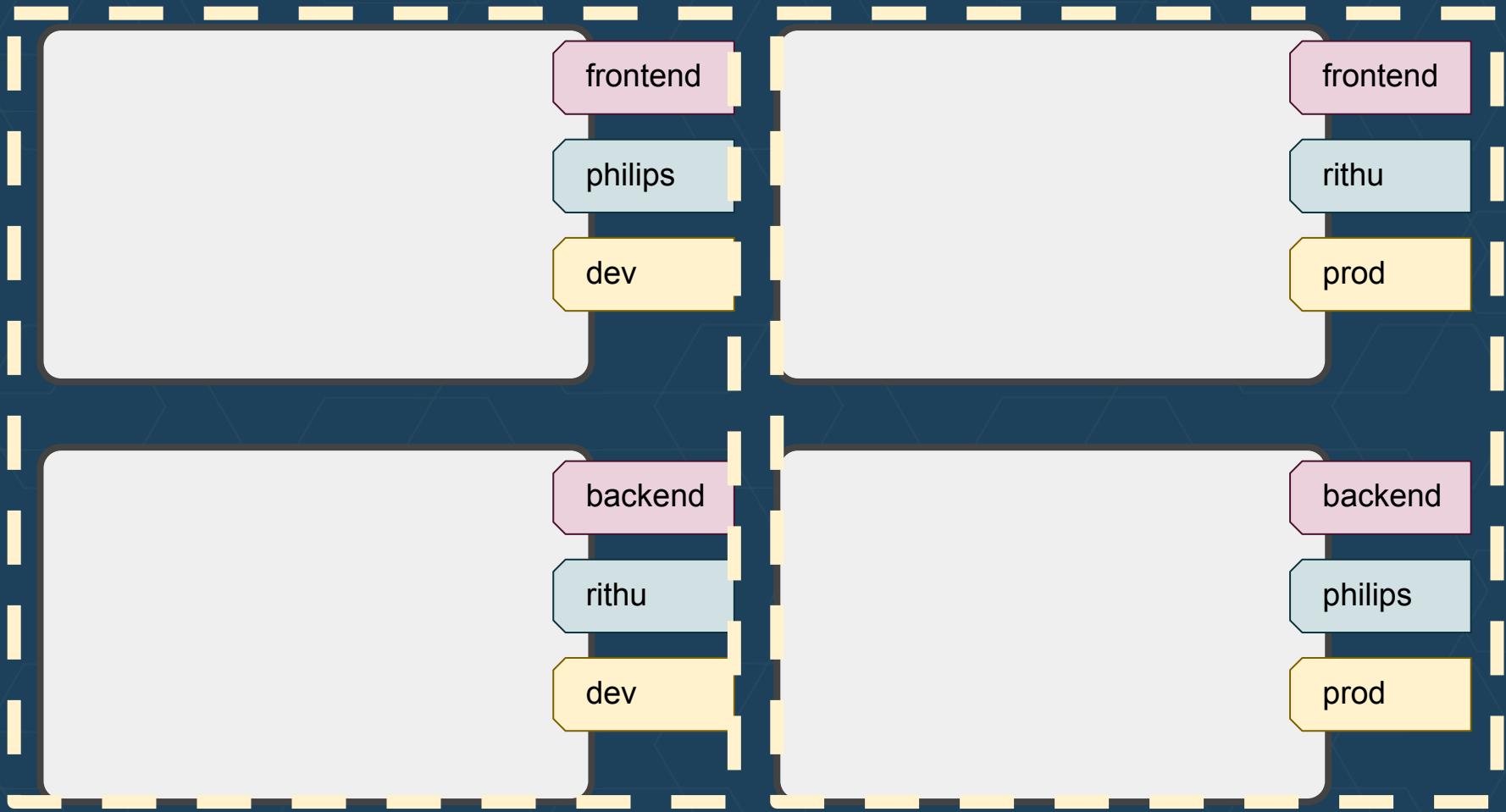

frontend
rithu
dev


backend
rithu
dev


backend
philips
prod


frontend
philips
prod
frontend
rithu
dev
backend
rithu
dev
backend
philips
prod

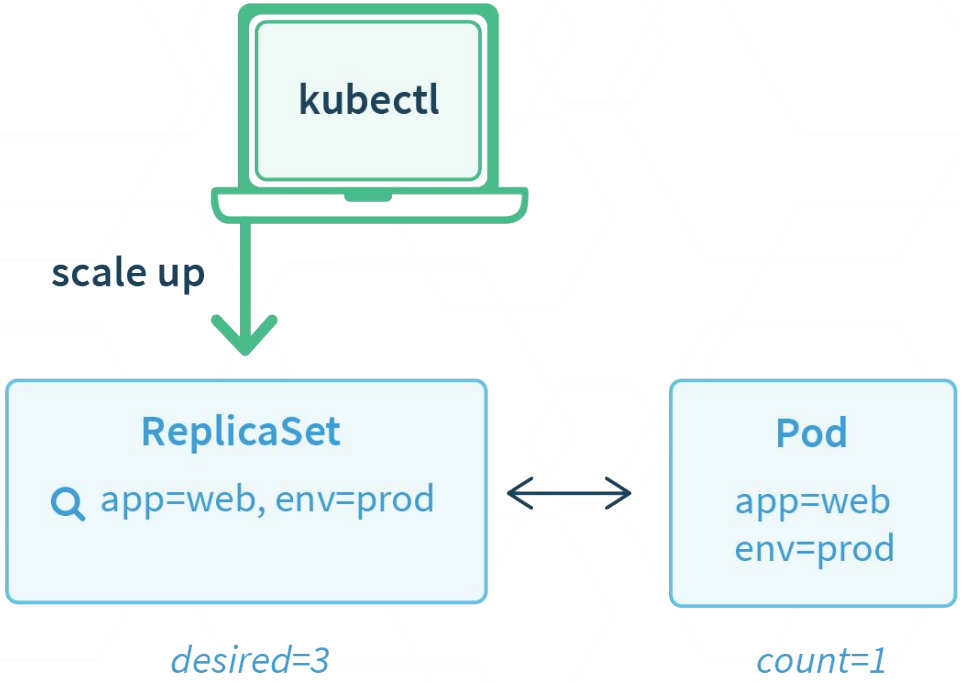

frontend
philips
prod
frontend
rithu
dev
backend
rithu
dev
backend
philips
prod


frontend
philips
dev
frontend
rithu
prod
backend
rithu
dev
backend
philips
prod

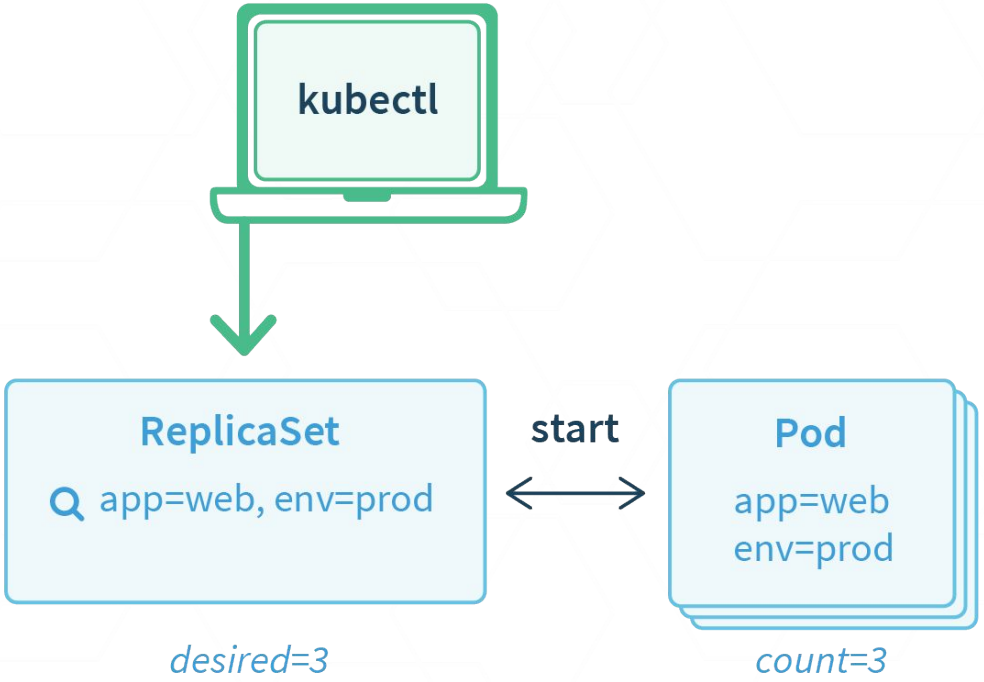
Scaling Applications

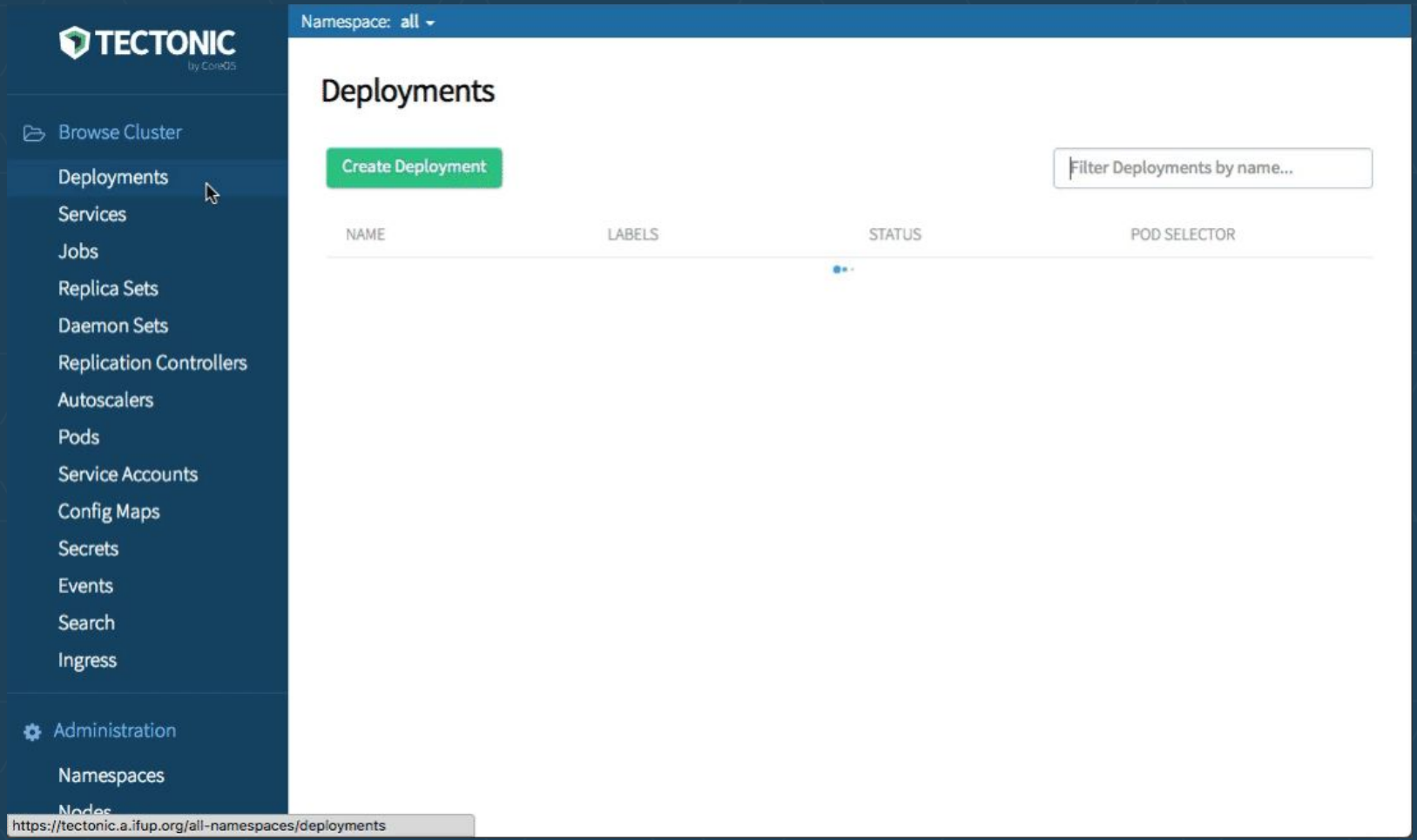
Reacting to Demand

ReplicaSet



ReplicaSet




TECTONIC
by CoreOS
Namespace: all
Browse Cluster
Deployments
Services
Jobs
Replica Sets
Daemon Sets
Replication Controllers
Autoscalers
Pods
Service Accounts
Config Maps
Secrets
Events
Search
Ingress
Administration
Namespaces
Nodes
Deployments
Create Deployment
Filter Deployments by name...
NAME
LABELS
STATUS
POD SELECTOR
https://tectonic.a.ifup.org/all-namespaces/deployments

Overview

Scaling Complex Apps

Creating a Database is Easy on Kubernetes

```
$ kubectl run db --image=quay.io/my/db
```

Managing a Distributed Database is Harder

- **Resize/Upgrade** - coordination for availability
- **Reconfigure** - tedious generation / templating
- **Backup** - requires coordination on instances
- **Healing** - observe and act for recovery

Introducing

Operators

The Dream

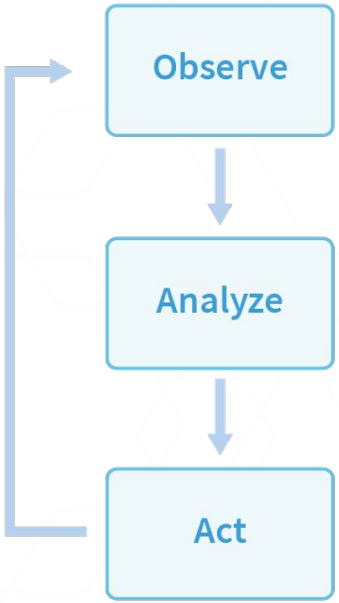
```
$ cat my-db-cluster.yaml
spec:
  clusterSize: 3
  readReplicas: 2
  version: v4.0.1
```

An Operator represents human operational knowledge in software, to reliably manage an application.

etcd Operator Resource

```
$ cat etcd-cluster.yaml
spec:
  clusterSize: 3
  version: v3.1.0
```

etcd Operator



Cluster “A” has 2 running pods:
- name: A-000, version 3.0.9
- name: A-001, version 3.1.0

Differences from desired config:
- should be version 3.1.0
- should have 3 members

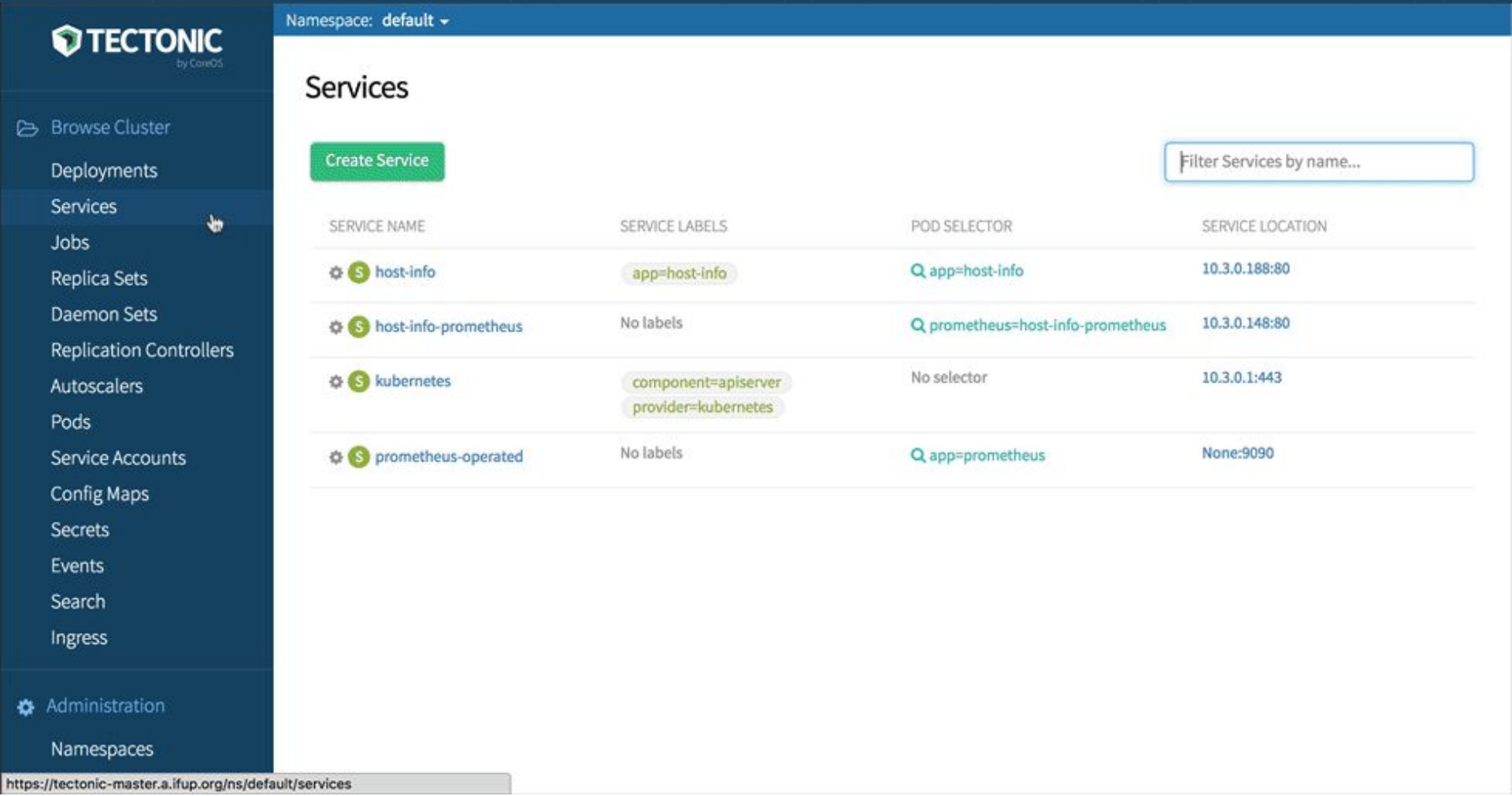
How to get to desired config:
- Recover 1 member
- Back up cluster
- Upgrade to 3.1.0

Monitoring

Foundation of Production

Prometheus Operator

- Operates Prometheus on k8s
- Handles common tasks:
  - Create/Destroy
  - Monitor Configuration
  - **Services Targets via Labels**
- Configured by resources


TECTONIC
by CoreOS
Namespace: default
Browse Cluster
Deployments
Services
Jobs
Replica Sets
Daemon Sets
Replication Controllers
Autoscalers
Pods
Service Accounts
Config Maps
Secrets
Events
Search
Ingress
Administration
Namespaces
Services
Create Service
Filter Services by name...
SERVICE NAME
SERVICE LABELS
POD SELECTOR
SERVICE LOCATION
host-info
app=host-info
app=host-info
10.3.0.188:80
host-info-prometheus
No labels
prometheus=host-info-prometheus
10.3.0.148:80
kubernetes
component=apiserver
provider=kubernetes
No selector
10.3.0.1:443
prometheus-operated
No labels
app=prometheus
None:9090
https://tectonic-master.a.ifup.org/ns/default/services

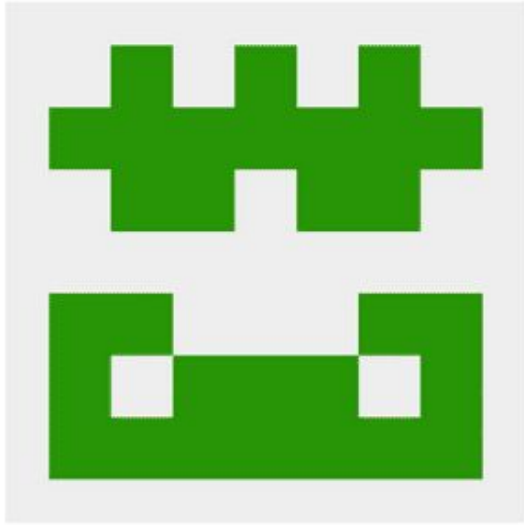
host.ifup.org

Prometheus Time Series Colle

CoreOS

host.ifup.org

You are visitor: 3919



host-info-2478745576-m6pff

Unknown Zone

Networking

lo

- 127.0.0.1/8
- ::1/128

eth0

What's Next?

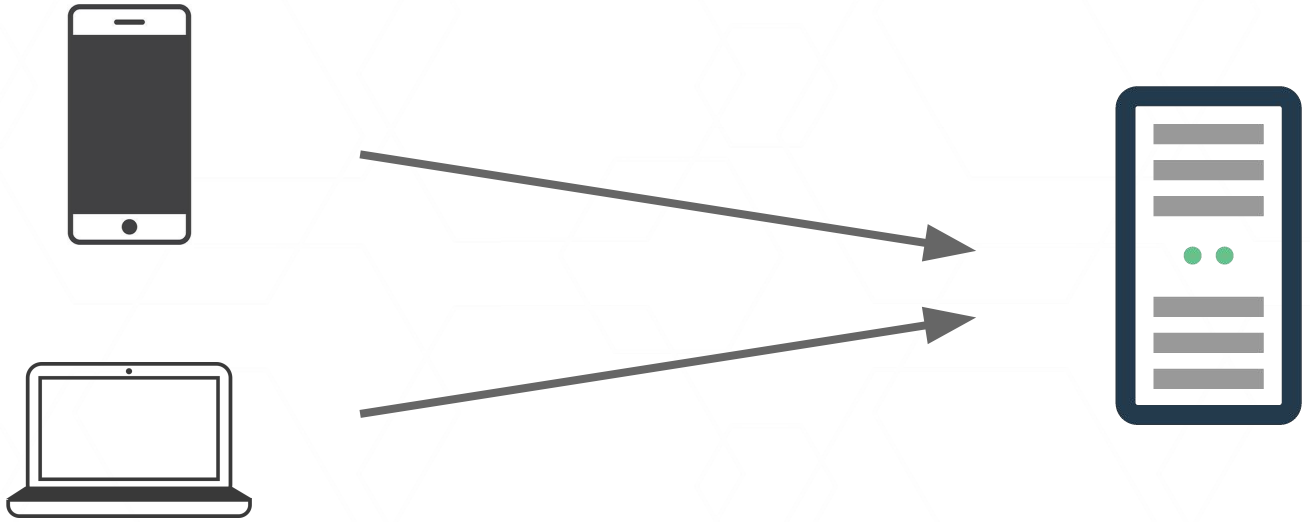
Kubernetes Next Steps

Kubernetes Next Steps

- Healthy growth of the open source community
- Better metrics and monitoring
- Ever improving security defaults
- Support for more cloud platforms
- More pre-packaged applications
- https://github.com/kubernetes/features

CoreOS Mission

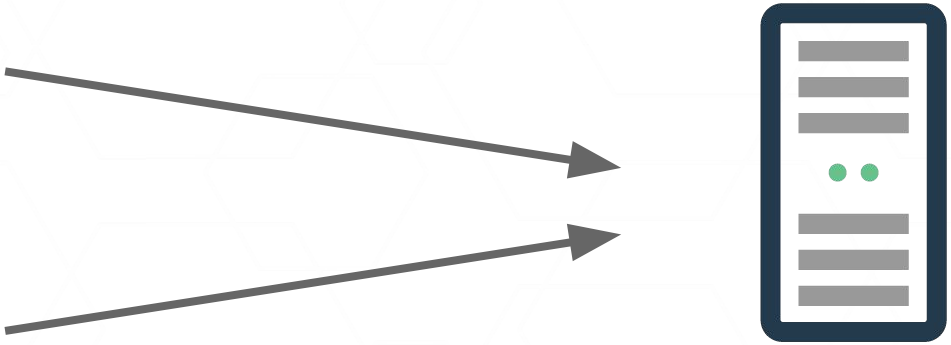
Secure the Internet



Documents

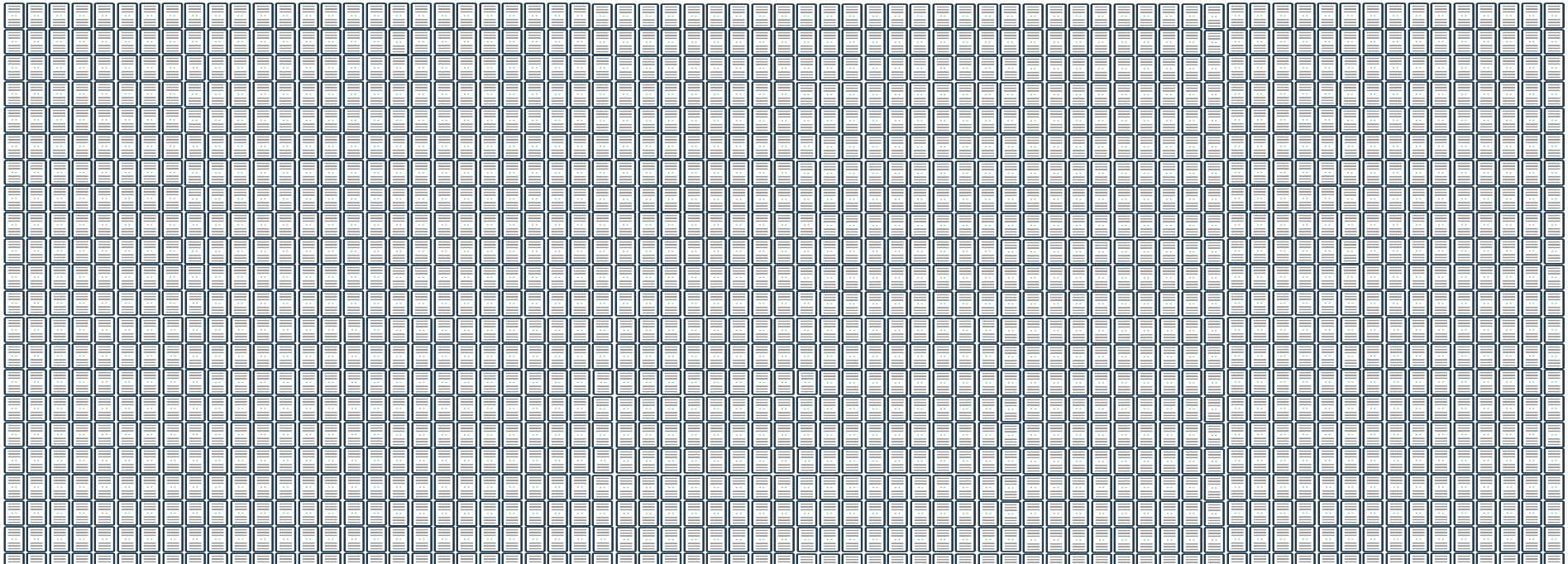
Commerce

Communications



~100,000,000

Servers Worldwide



CoreOS

Self Driving Infrastructure

What is self-driving infrastructure

Applying updates to our infrastructure software much like the way updates are applied to our phones

Why self driving infrastructure

Patching Software is important

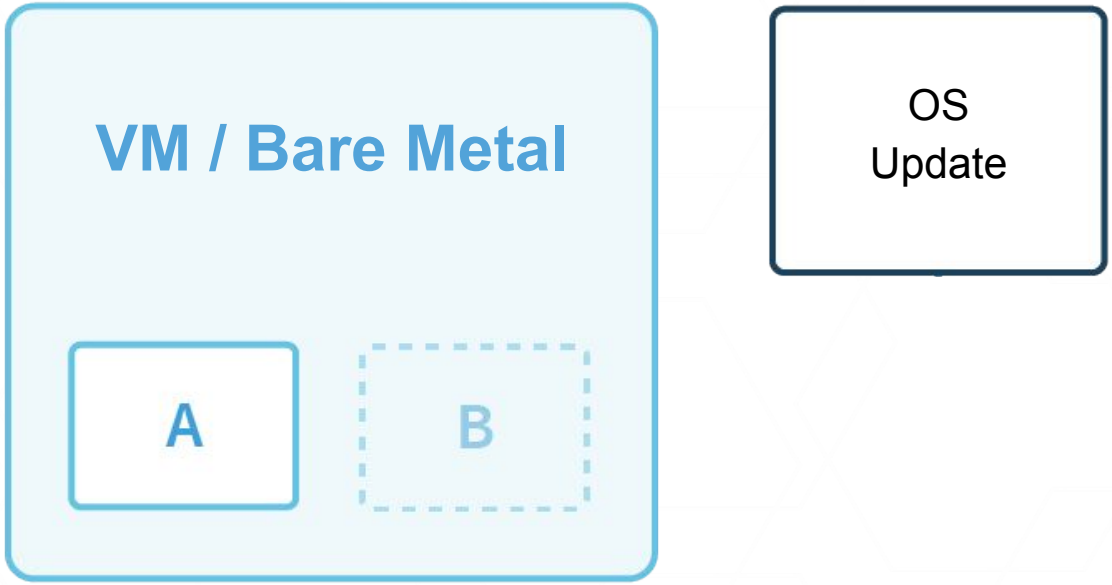
#	CVE ID	CWE ID	# of Exploits	Vulnerability Type(s)	Publish Date	Update Date	Score
1	CVE-2016-9083	190		DoS Overflow Mem. Corr. Bypass	2016-11-27	2016-11-29	7.2
drivers/vfio/pci/vfio_pci.c in the Linux kernel through 4.8.11 allows local users to bypass integer overflow ch impact, by leveraging access to a vfio PCI device file for a VFIO_DEVICE_SET_IRQS ioctl call, aka a "state m							
2	CVE-2016-8666	400		DoS	2016-10-16	2016-11-28	7.8
The IP stack in the Linux kernel before 4.6 allows remote attackers to cause a denial of service (stack consu GRO path for packets with tunnel stacking, as demonstrated by interleaved IPv4 headers and GRE headers,							
3	CVE-2016-8632	119		DoS Overflow +Priv	2016-11-27	2016-11-28	7.2
The tipc_msg_build function in net/tipc/msg.c in the Linux kernel through 4.8.11 does not validate the relat allows local users to gain privileges or cause a denial of service (heap-based buffer overflow) by leveraging							
4	CVE-2016-7914	125		DoS +Info	2016-11-16	2016-12-02	7.1
The assoc_array_insert_into_terminal_node function in lib/assoc_array.c in the Linux kernel before 4.5.3 do information from kernel memory or cause a denial of service (invalid pointer dereference and out-of-bounds demonstrated by the keyutils test suite.							
5	CVE-2016-7425	119		DoS Overflow +Priv	2016-10-16	2016-11-28	7.2
The arcmsr_iop_message_xfer function in drivers/scsi/arcmsr/arcmsr_hba.c in the Linux kernel through 4.8 or cause a denial of service (heap-based buffer overflow) via an ARCMSR_MESSAGE_WRITE_WQBUFFER co							
6	CVE-2016-7039	399		DoS	2016-10-16	2016-11-28	7.8
The IP stack in the Linux kernel through 4.8.2 allows remote attackers to cause a denial of service (stack co the GRO path for large crafted packets, as demonstrated by packets that contain only VLAN headers, a relat							
7	CVE-2016-6187	119		Overflow +Priv	2016-08-06	2016-11-28	7.2
The apparmor_setprocattr function in security/apparmor/lsm.c in the Linux kernel before 4.6.5 does not val AppArmor setprocattr hook.							
8	CVE-2016-5829	119		DoS Overflow	2016-06-27	2016-11-28	7.2
Multiple heap-based buffer overflows in the hiddev_ioctl_usage function in drivers/hid/usbhid/hiddev.c in th have unspecified other impact via a crafted (1) HIDIOCGUSAGES or (2) HIDIOCSUSAGES ioctl call.							
9	CVE-2016-5828	20		DoS	2016-06-27	2016-11-28	7.2
The start_thread function in arch/powerpc/kernel/process.c in the Linux kernel through 4.6.3 on powerpc pl service (invalid process state or TM Bad Thing exception, and system crash) or possibly have unspecified oth							
10	CVE-2016-5343	119		DoS Overflow Mem. Corr.	2016-10-10	2016-11-28	7.5
drivers/soc/qcom/qdsp6v2/voice_svc.c in the QDSP6v2 Voice Service driver for the Linux kernel 3.x, as use other products, allows attackers to cause a denial of service (memory corruption) or possibly have unspecifi buffer overflow.							
11	CVE-2016-5342	119		DoS Overflow	2016-08-30	2016-11-28	7.2
Heap-based buffer overflow in the wcnss_wlan_write function in drivers/net/wireless/wcnss/wcnss_wlan.c ir Innovation Center (QUIC) Android contributions for MSM devices and other products, allows attackers to ca /dev/wcnss_wlan with an unexpected amount of data.							
12	CVE-2016-5340	20		Bypass	2016-08-07	2016-11-28	7.2
The is_ashmem_file function in drivers/staging/android/ashmem.c in a certain Qualcomm Innovation Center the KGSL Linux Graphics Module, which allows attackers to bypass intended access restrictions by using the							
13	CVE-2016-5195	362		+Priv	2016-11-10	2016-12-02	7.2
Race condition in mm/gup.c in the Linux kernel 2.x through 4.x before 4.8.3 allows local users to gain privile read-only memory mapping, as exploited in the wild in October 2016, aka "Dirty COW."							

HEWLETT·PACKARD
hp 15C
ON
f
g
SST
BST
LBL
GTO
HYP
HYP-1
R/S
P/R
PSE
GSB
RTN
Σ
SIN
SIN-1
DIM
COS
COS-1
(i)
TAN
TAN-1
I
PRGM
CLEAR
REG
PREFIX
RND
R↑
STO
INT
FRAC
RCL
MEM
USER
CLx
ENTER
LSTx
RAN #
EEX
π
RESULT
CHS
ABS
MATRIX
1/x
Δ%
E
yx
%
D
10x
LOG
C
ex
LN
B
x2
A
7
DEG
FIX
8
RAD
SCI
9
GRD
ENG
÷
x≤y
SOLVE
4
SF
5
CF
DSE
6
F?
ISG
×
x=0
1
→P
→R
2
→H
→H.MS
3
→DEG
→RAD
−
TEST
Re≷Im
0
x̄
x!
·
s
ŷ,r
Σ+
Σ-
L.R.


ENCRYPTION
LAST WEEK TONIGHT
0:19 / 18:00
CC
HD

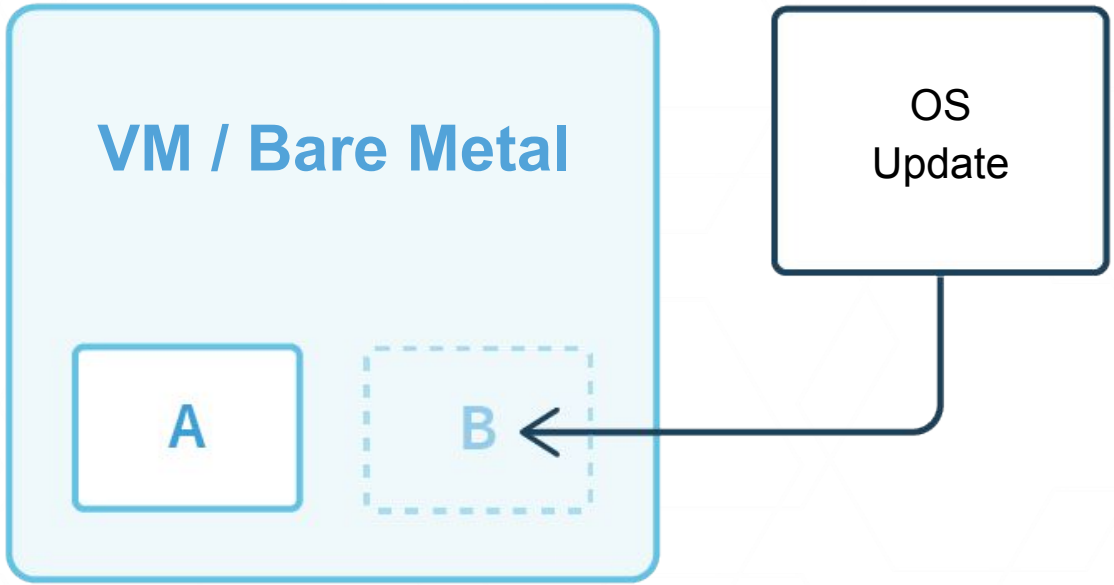
Updating Container Linux

container linux



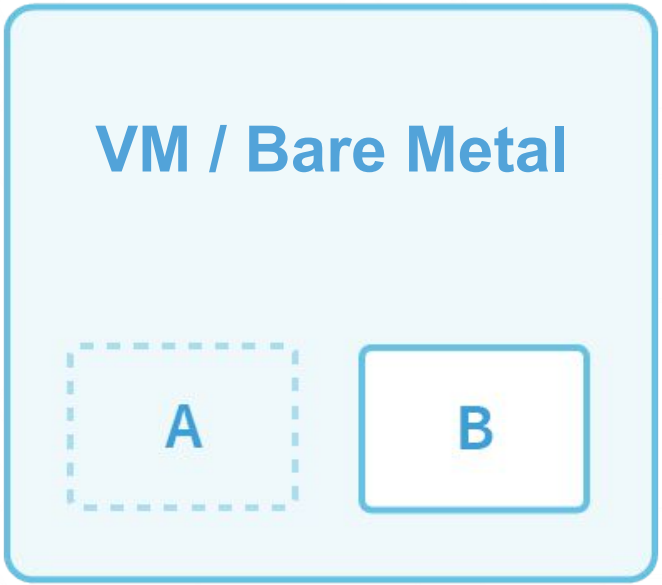
Updating Container Linux

container linux



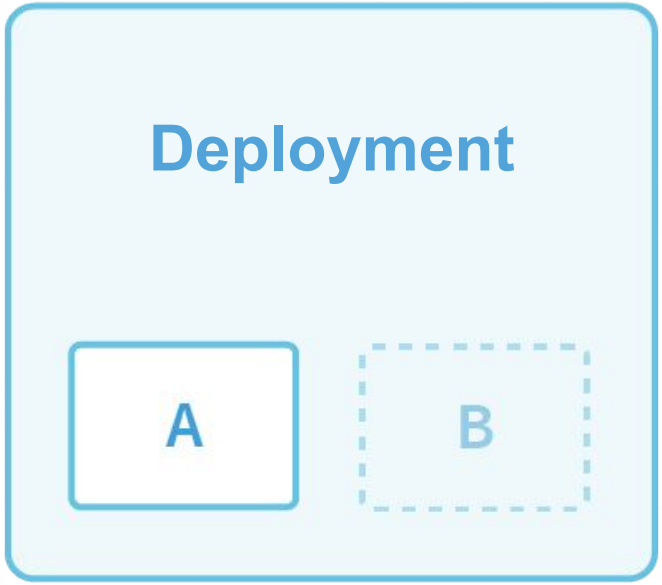
Updating Container Linux

container linux



Updating Kubernetes

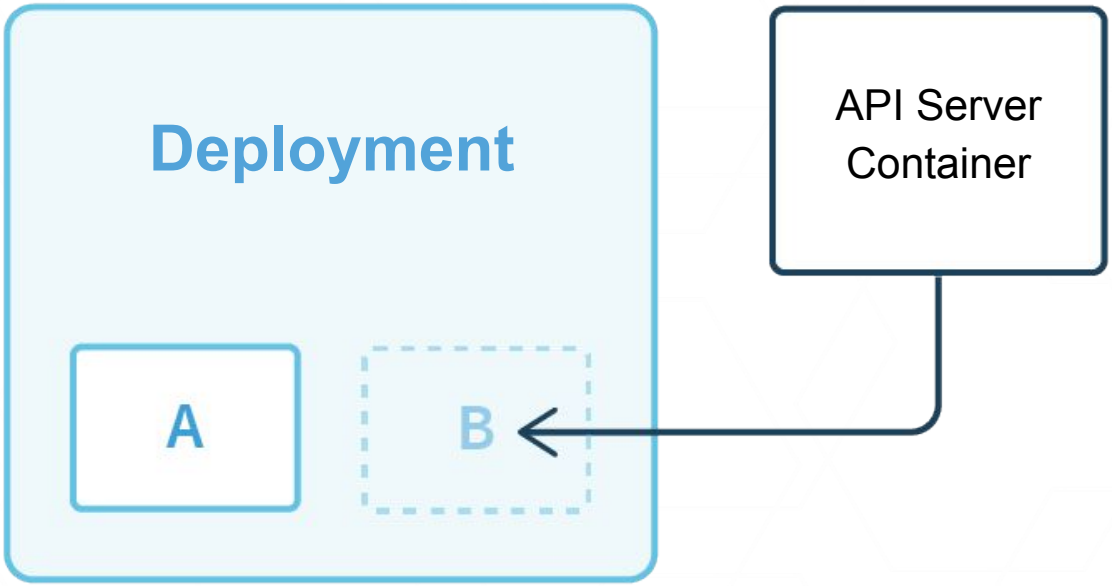
Kubernetes



API Server
Container

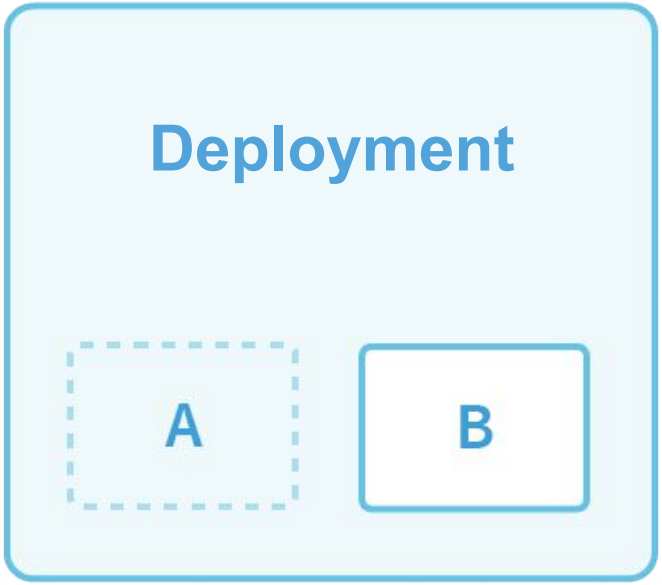
Updating Kubernetes

Kubernetes



Updating Kubernetes

Kubernetes



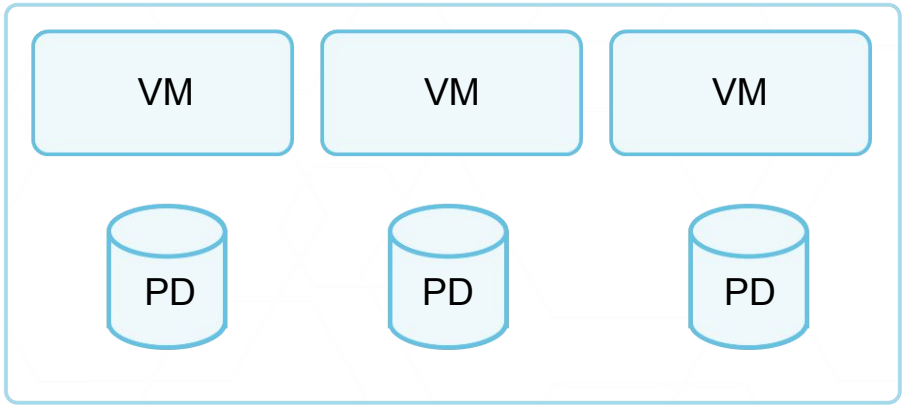
kubectl

k8s API

Azure APIs

kubernetes

v1.5.1



VirtualNet

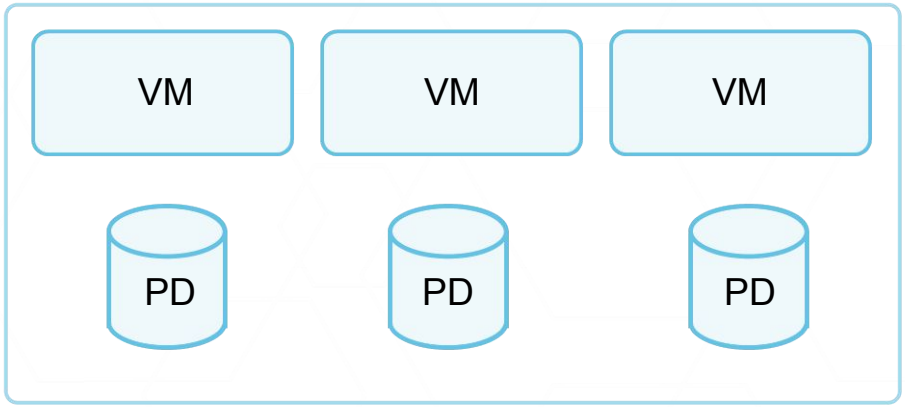
kubectl

k8s API

Azure APIs

kubernetes

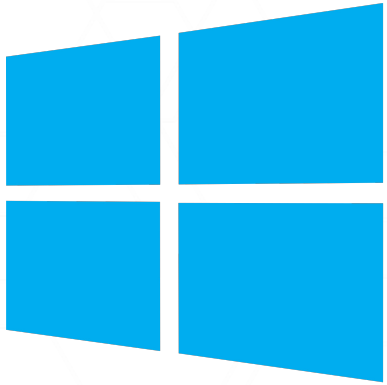
v1.5.2

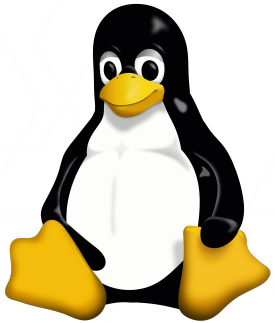


Free Stuff

Get Kubernetes For Yourself

Minikube - Kubernetes on your laptop





macOS

All-in-one Virtual Machine

Single file download

Latest Kubernetes

github.com/kubernetes/minikube

Tectonic Free Tier

Pure upstream Kubernetes

Production configuration

Bare metal and AWS installer

(more coming soon)

coreos.com/tectonic

Join Us

Build great code

github.com/coreos

Hack on Operating Systems

Build distributed systems

Geek out on internet identity protocols

Container Focused Operating System

Clustered Database

Identity & Federation

10+ other major projects

github.com/kubernetes

Most active project on GitHub

Dozens of special interest groups

Not just code! Docs, PM, and more

kubernetes

Work WIth Us

Community is taking off

kubernetes

What container and PaaS tools are used to manage OpenStack applications?

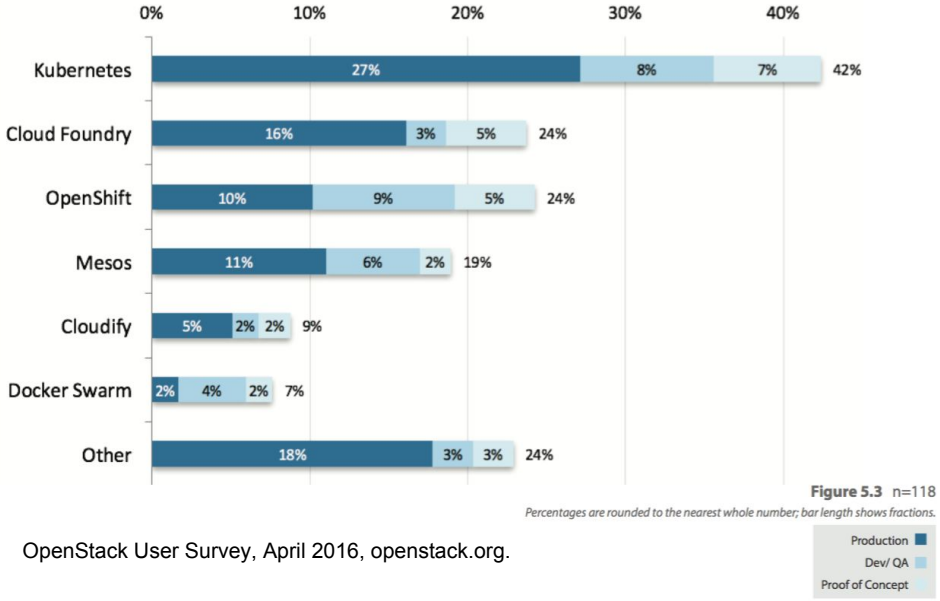
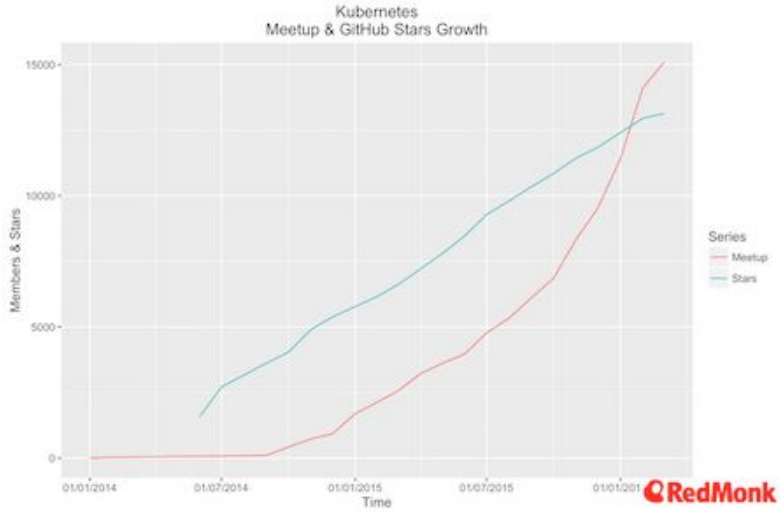


Figure 5.3 n=118

Percentages are rounded to the nearest whole number; bar length shows fractions.

OpenStack User Survey, April 2016, openstack.org.

Developer engagement



"The Further Evolution of Kubernetes," March 2016.

CoreOS

CoreOS is running the world's containers

We're hiring: careers@coreos.com

OPEN SOURCE	ENTERPRISE
90+ Projects on GitHub, 1,000+ Contributors	Support plans, training and more
container linux dex etcd	TECTONIC QUAY
coreos.com	sales@coreos.com

May 31 - June 1, 2017
San Francisco

coreos.com/fest

@coreosfest

Thanks!

QUESTIONS?

brandon.philips@coreos.com

@BrandonPhilips

linkedin.com/in/brandonphilips

LONGER CHAT?

Let’s talk!

IRC

More events: coreos.com/community

We’re hiring: coreos.com/careers

Preview

Self-Driving Kubernetes