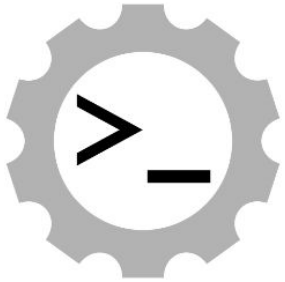
leigh capili

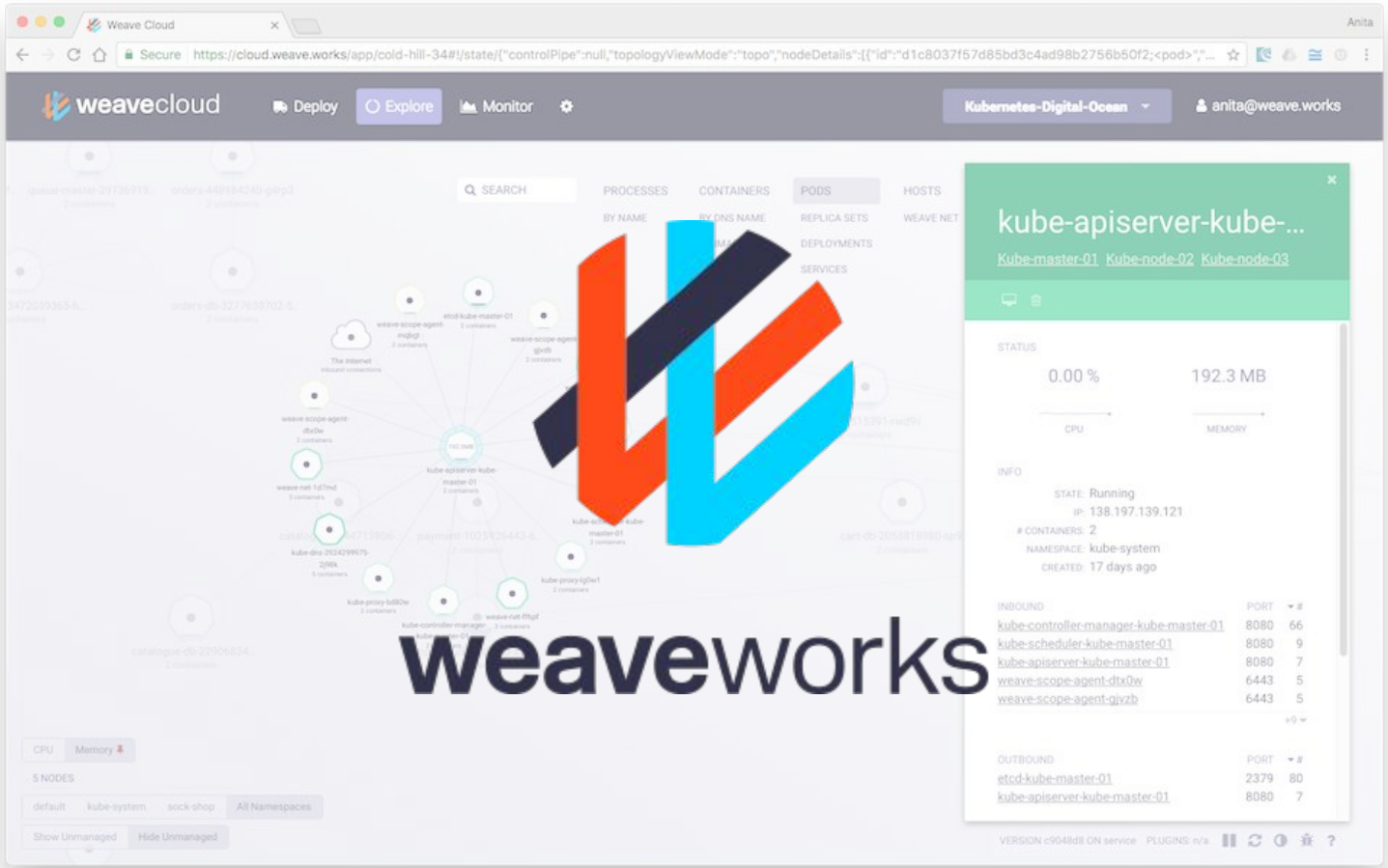


@capileigh

stealthybox

@capileigh

stealthybox



@capileigh

stealthybox

< Performance Art >

 @capileigh  stealthybox

@capileigh stealthybox

503

Service Unavailable

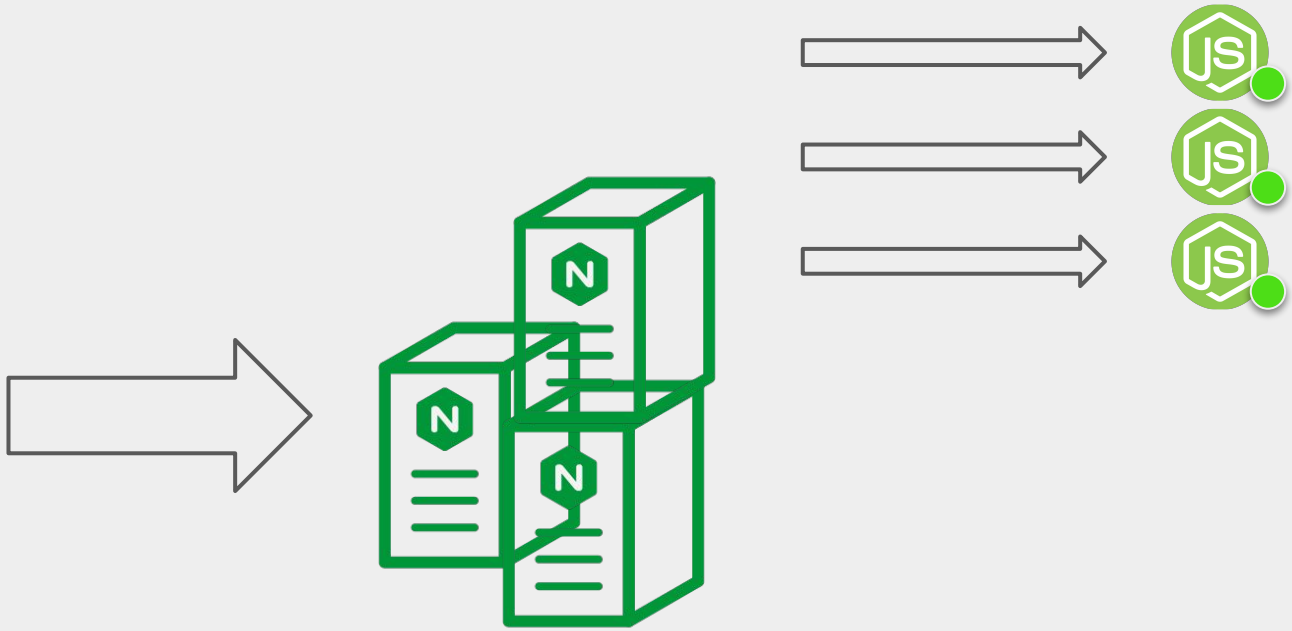
 @capileigh  stealthybox

The Gotchas of Zero-Downtime Traffic

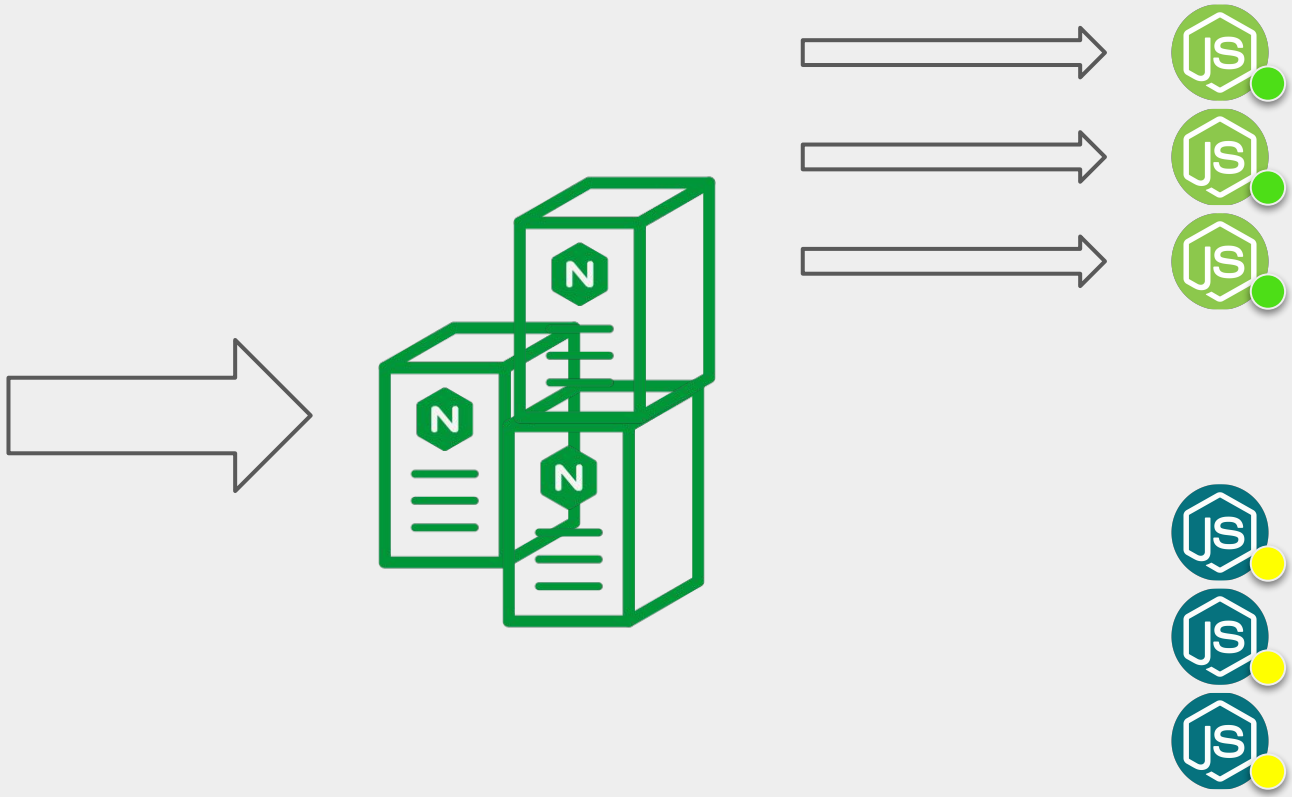
(with k8s!)

@capileigh

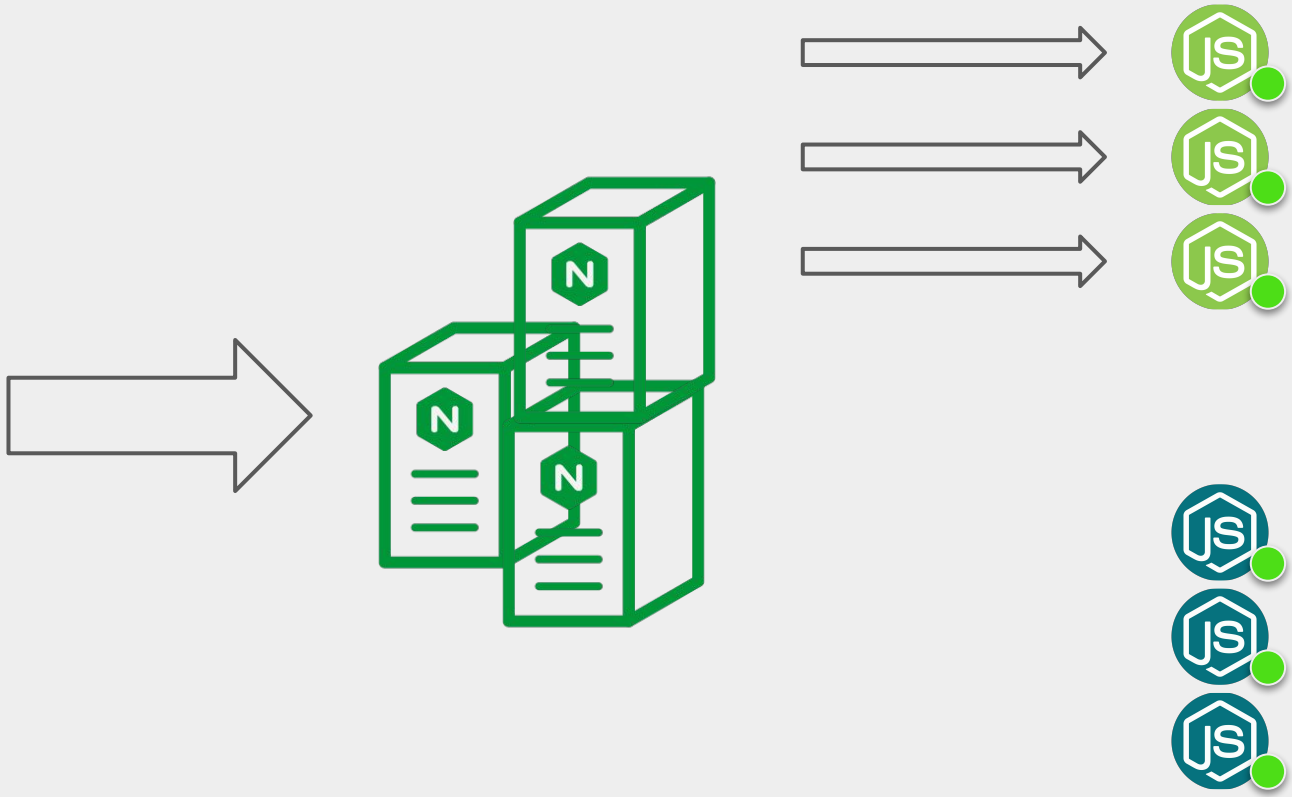
stealthybox



@capileigh stealthybox



@capileigh stealthybox



@capileigh stealthybox

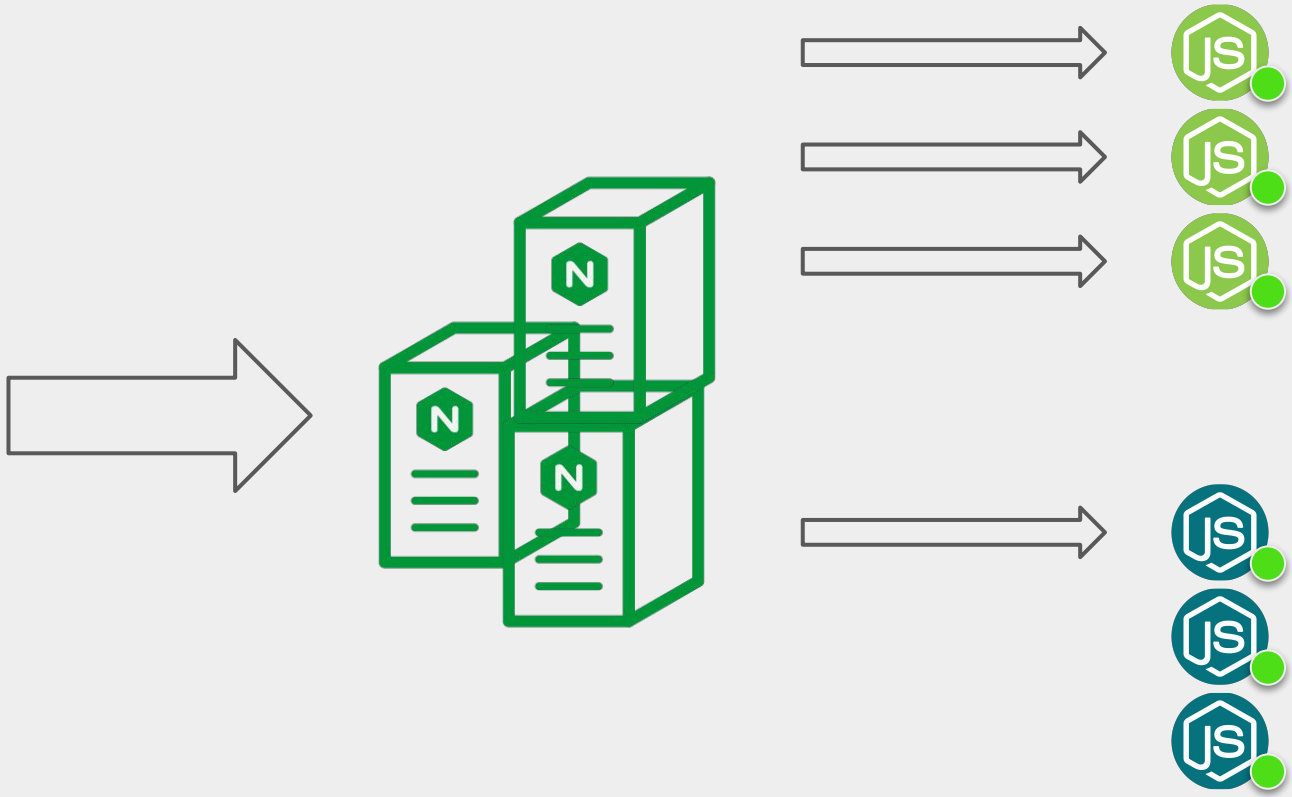
```
185.93.230.3 - - [20/Apr/2017:08:30:29 -0300] "GET /feeds/posts/default HTTP/1.1" 404 6919 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10
_12_0) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/54.0.2840.470 Safari/537.36"
192.88.135.6 - - [20/Apr/2017:08:34:03 -0300] "GET / HTTP/1.1" 200 4772 "http://www.phoenixhollow.com" "Mozilla/5.0+(compatible; UptimeRo
bot/2.0; http://www.uptimerobot.com/)"
185.93.229.3 - - [20/Apr/2017:08:36:32 -0300] "GET /feeds/posts/default HTTP/1.1" 301 5 "http://www.slumpedoverkeyboarddead.com/feeds/pos
ts/default" "Go 1.1 package http"
185.93.229.3 - - [20/Apr/2017:08:36:33 -0300] "GET /feeds/posts/default HTTP/1.1" 404 6919 "https://www.slumpedoverkeyboarddead.com/feeds
/posts/default" "Go 1.1 package http"
192.88.134.3 - - [20/Apr/2017:08:43:37 -0300] "GET /my-website-is-down-now-what-part-4/ HTTP/1.1" 301 5 "-" "msnbot-media/1.1 (+http://se
arch.msn.com/msnbot.htm)"
192.88.134.13 - - [20/Apr/2017:08:49:38 -0300] "GET /robots.txt HTTP/1.1" 200 1320 "-" "Mozilla/5.0 (compatible; bingbot/2.0; +http://www
.bing.com/bingbot.htm)"
192.88.134.13 - - [20/Apr/2017:08:49:48 -0300] "GET /contact.php HTTP/1.1" 200 2223 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 7_0 like Mac
OS X) AppleWebKit/537.51.1 (KHTML, like Gecko) Version/7.0 Mobile/11A465 Safari/9537.53 (compatible; bingbot/2.0; +http://www.bing.com/bi
ngbot.htm)"
185.93.231.3 - - [20/Apr/2017:08:57:47 -0300] "GET /search/label/til?m=0 HTTP/1.1" 200 25347 "-" "Mozilla/5.0 (compatible; Baiduspider/2.
0; +http://www.baidu.com/search/spider.html)"
185.93.229.6 - - [20/Apr/2017:08:59:15 -0300] "GET /robots.txt HTTP/1.1" 200 114 "-" "Mozilla/5.0 (compatible; bingbot/2.0; +http://www.b
ing.com/bingbot.htm)"
192.88.134.13 - - [20/Apr/2017:09:02:09 -0300] "GET /robots.txt HTTP/1.1" 200 1320 "-" "Mozilla/5.0 (compatible; bingbot/2.0; +http://www
.bing.com/bingbot.htm)"
192.88.134.13 - - [20/Apr/2017:09:02:15 -0300] "GET /code/jquery.js HTTP/1.1" 200 57254 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 7_0 like
Mac OS X) AppleWebKit/537.51.1 (KHTML, like Gecko) Version/7.0 Mobile/11A465 Safari/9537.53 (compatible; bingbot/2.0; +http://www.bing.co
m/bingbot.htm)"
192.88.134.13 - - [20/Apr/2017:09:02:23 -0300] "GET /menu.css HTTP/1.1" 200 3727 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 7_0 like Mac OS
X) AppleWebKit/537.51.1 (KHTML, like Gecko) Version/7.0 Mobile/11A465 Safari/9537.53 (compatible; bingbot/2.0; +http://www.bing.com/bingb
ot.htm)"
192.88.134.13 - - [20/Apr/2017:09:02:29 -0300] "GET /styles/styles.css HTTP/1.1" 200 2910 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 7_0 lik
e Mac OS X) AppleWebKit/537.51.1 (KHTML, like Gecko) Version/7.0 Mobile/11A465 Safari/9537.53 (compatible; bingbot/2.0; +http://www.bing.
com/bingbot.htm)"
185.93.230.3 - - [20/Apr/2017:09:08:02 -0300] "GET /feed HTTP/1.1" 301 5 "http://www.slumpedoverkeyboarddead.com/" "Mozilla/5.0 (compatib
le; inoreader.com-like FeedFetcher-Google)"
185.93.230.3 - - [20/Apr/2017:09:08:03 -0300] "GET /feed/ HTTP/1.1" 304 0 "http://www.slumpedoverkeyboarddead.com/" "Mozilla/5.0 (compati
ble; inoreader.com-like FeedFetcher-Google)"
185.93.228.3 - - [20/Apr/2017:09:16:24 -0300] "GET /jon-watson/ HTTP/1.1" 200 32007 "-" "Mozilla/5.0 (Windows NT 6.1; WOW64; rv:40.0) Gec
ko/20100101 Firefox/40.1"
192.88.134.3 - - [20/Apr/2017:09:18:21 -0300] "GET /2009/10/29/ubuntu-9-10-screen-shots/ HTTP/1.1" 404 6919 "-" "Mozilla/5.0 (compatible;
 bingbot/2.0; +http://www.bing.com/bingbot.htm)"
185.93.231.3 - - [20/Apr/2017:09:19:14 -0300] "GET /tag/fidonet/feed/ HTTP/1.1" 200 3969 "-" "Mozilla/5.0 (compatible; Baiduspider/2.0; +
http://www.baidu.com/search/spider.html)"
185.93.231.18 - - [20/Apr/2017:09:20:22 -0300] "GET /feed/ HTTP/1.1" 200 11559 "-" "Mozilla/5.0 (compatible; Kraken/0.1; http://linkfluen
ce.net/; bot@linkfluence.net)"
```

@capileigh

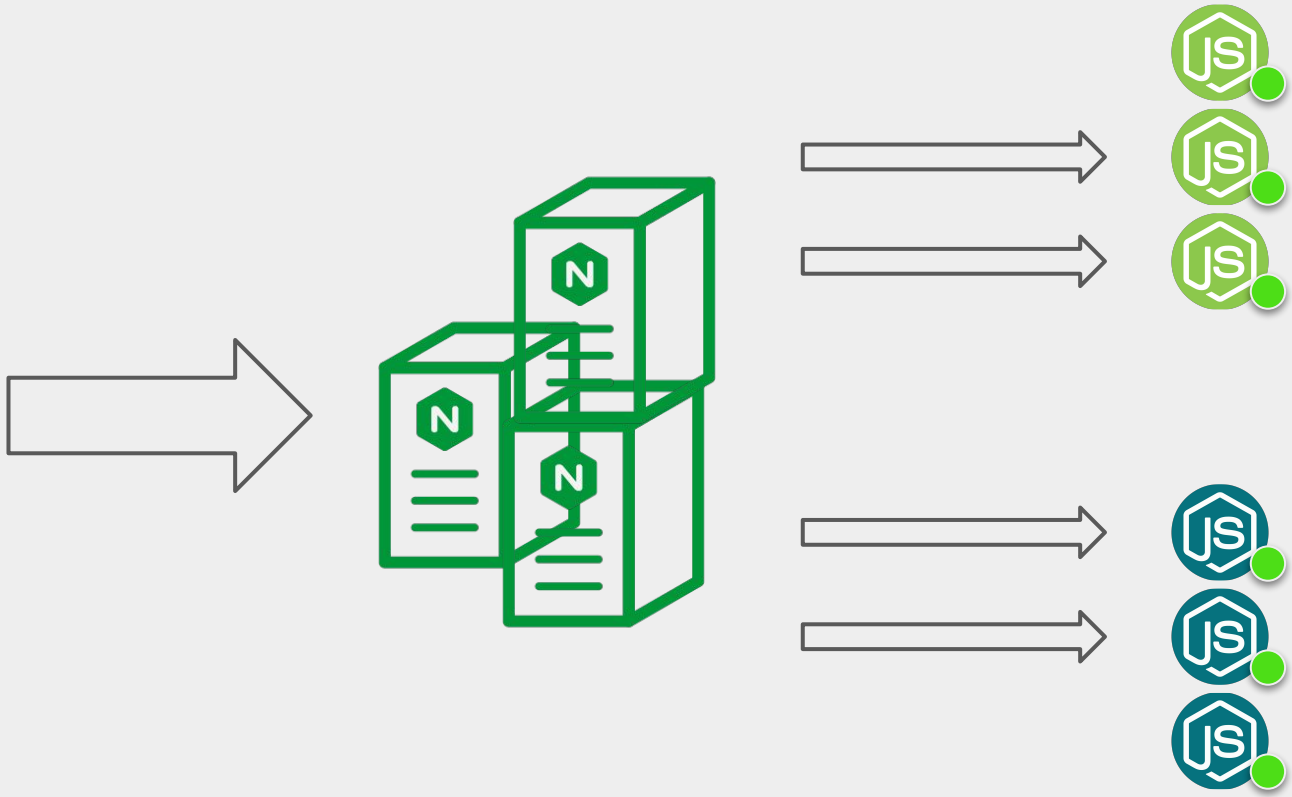
stealthybox

@capileigh

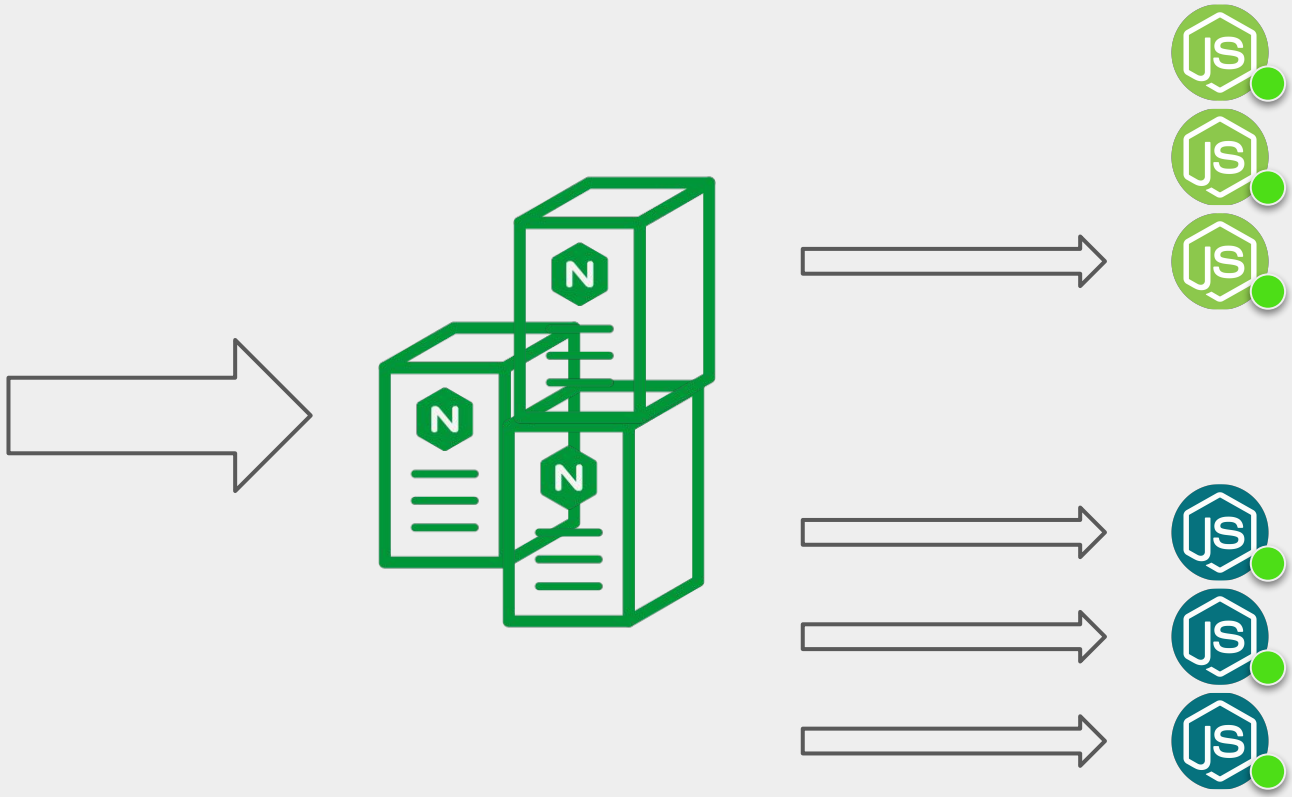
stealthybox



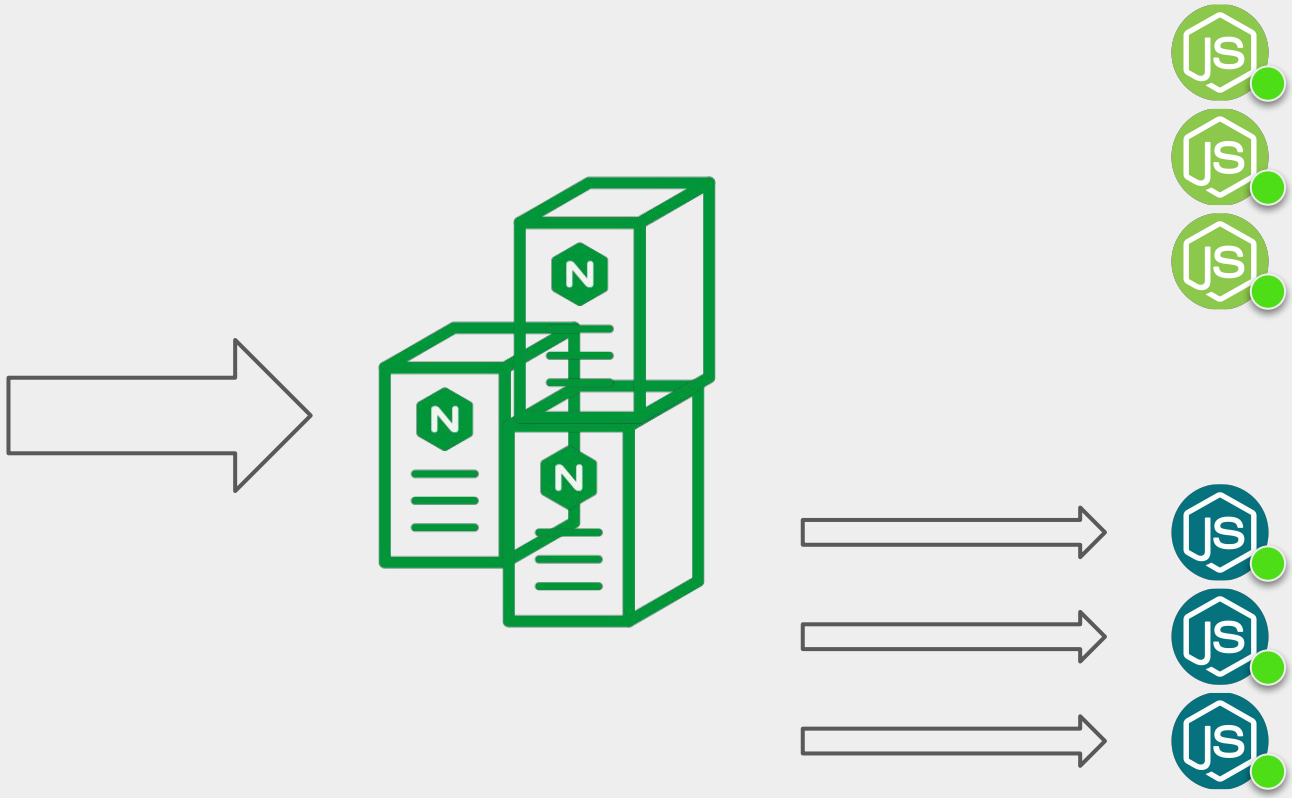
@capileigh stealthybox



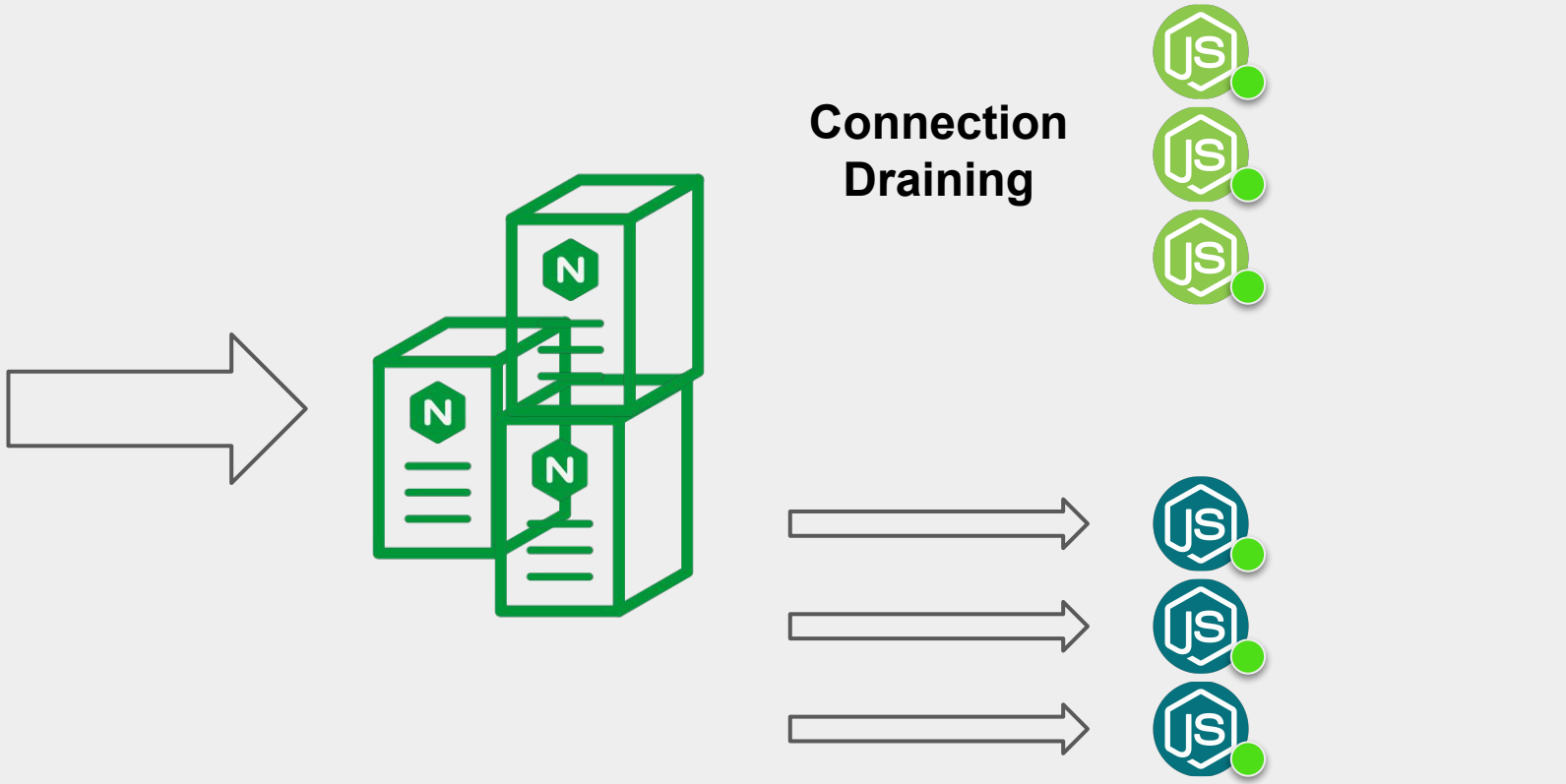
@capileigh stealthybox



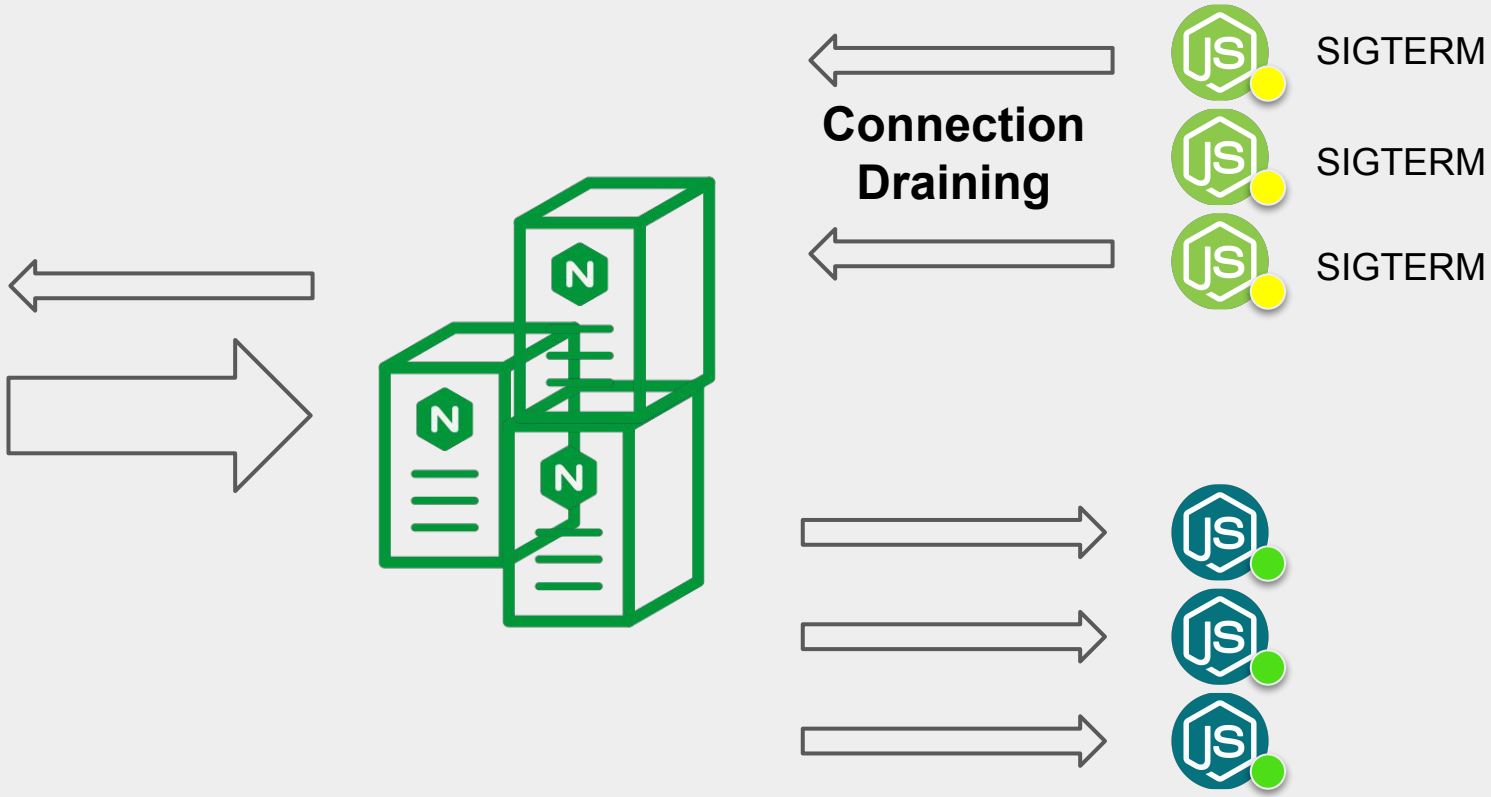
@capileigh stealthybox



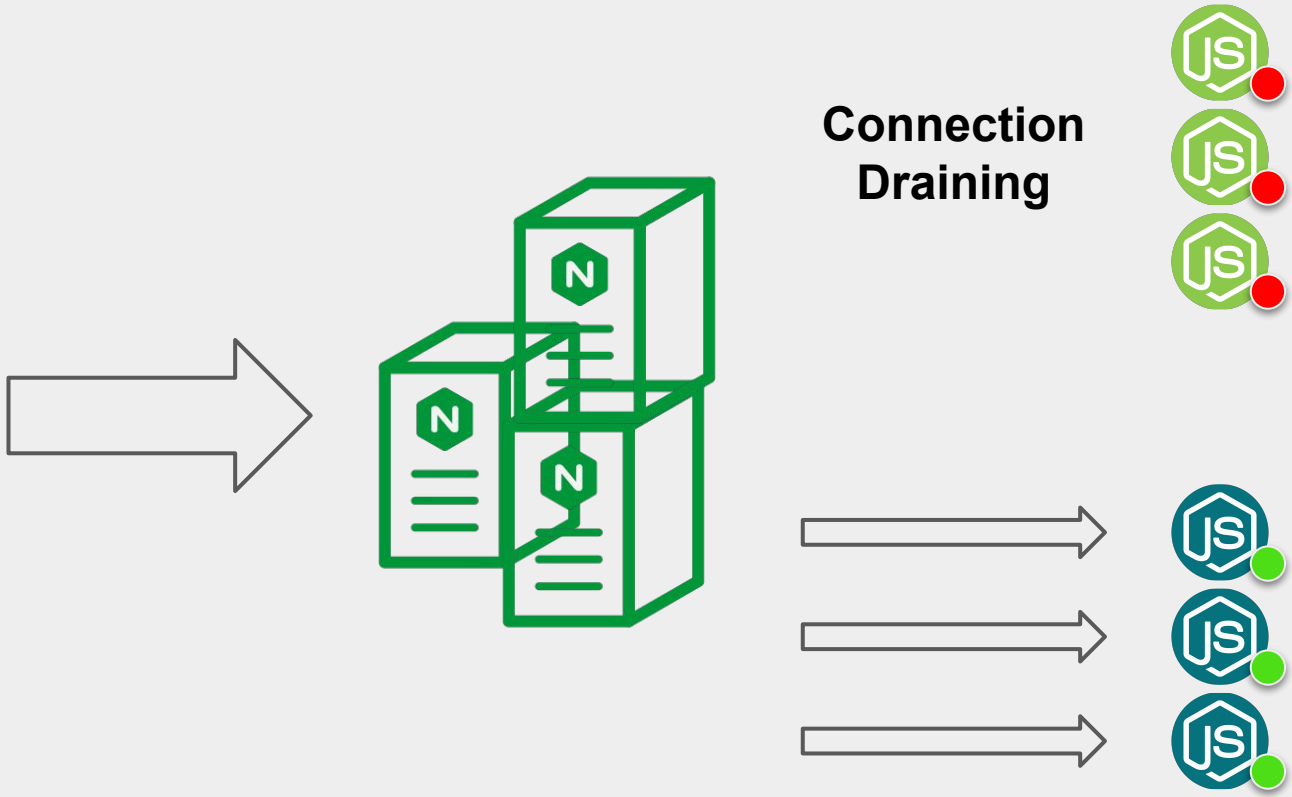
@capileigh stealthybox



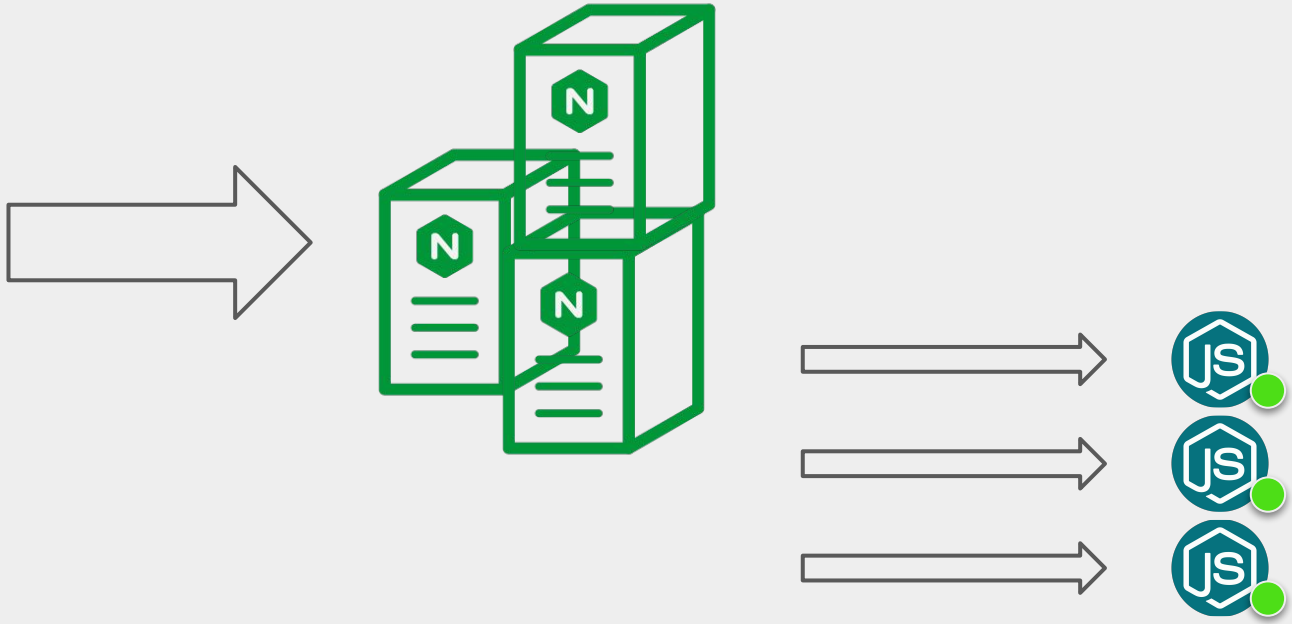
@capileigh stealthybox



@capileigh stealthybox

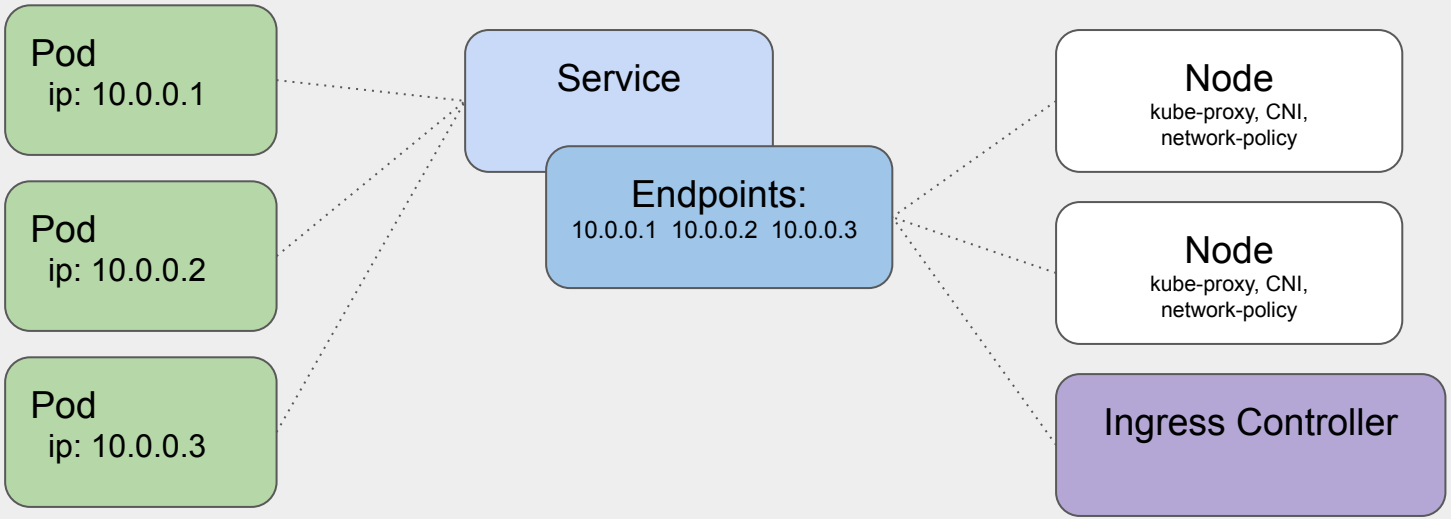


@capileigh stealthybox

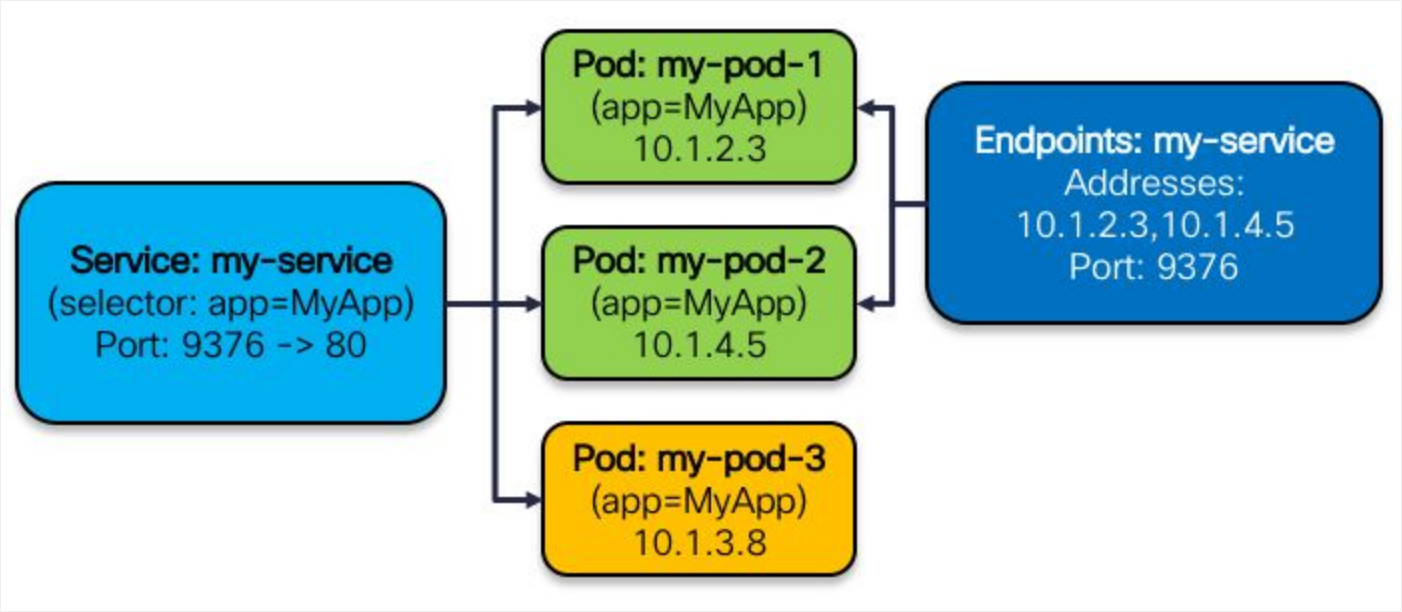


@capileigh stealthybox

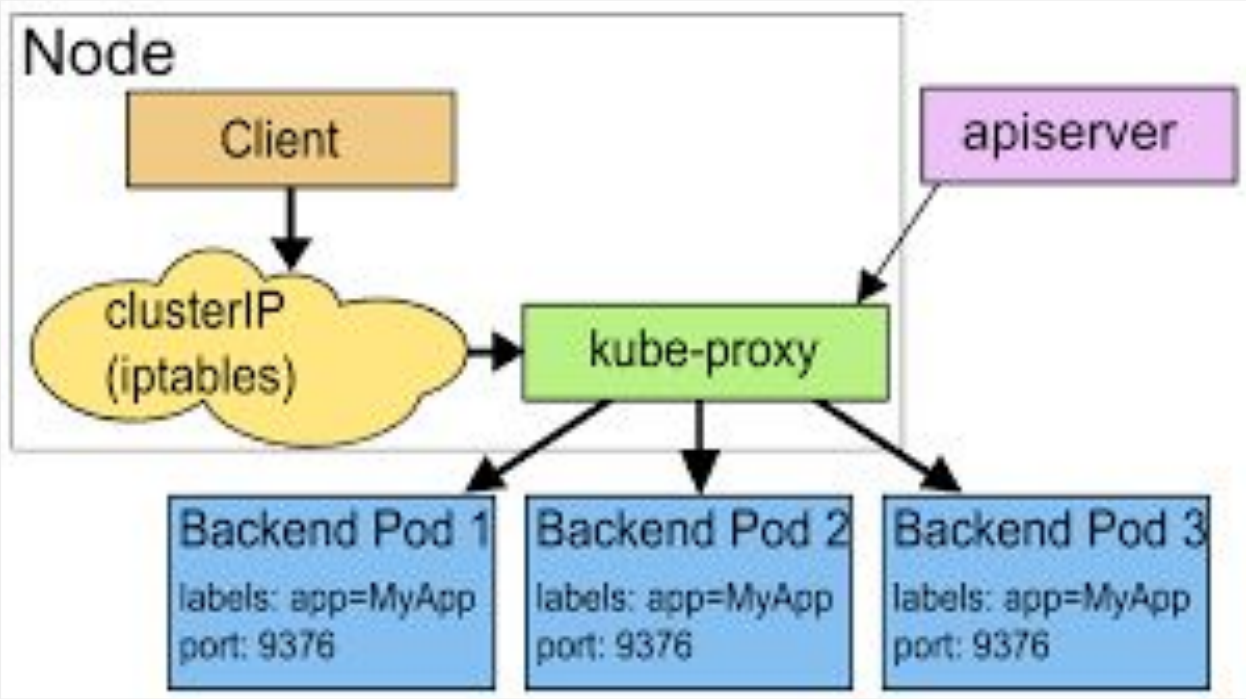
 @capileigh  stealthybox



@capileigh stealthybox



@capileigh stealthybox


Node
Client
apiserver
clusterIP
(iptables)
kube-proxy
Backend Pod 1
labels: app=MyApp
port: 9376
Backend Pod 2
labels: app=MyApp
port: 9376
Backend Pod 3
labels: app=MyApp
port: 9376

Pod Shutdown

kube-apiserver receives delete
Pod marked as **Terminating**
+ **aysnc** consequence:
Service controller removes Endpoint

PreStop hooks run
PID 1 of all containers receive **SIGTERM**
Termination Grace Period Seconds
PID 1 of all containers receive **SIGKILL**

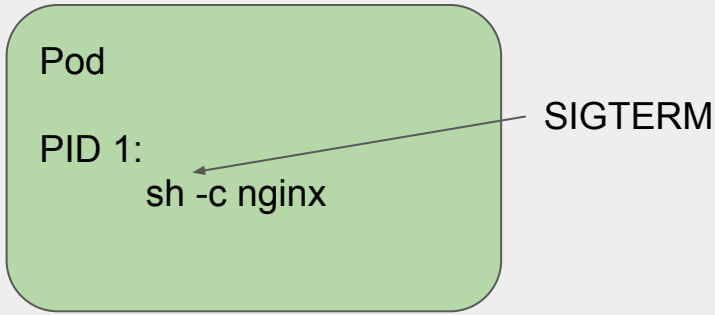

@capileigh

stealthybox

Gotcha #1

Dockerfile:

CMD nginx
vs.
ENTRYPOINT [“nginx”]



Gotcha #2

STOPSIGNAL

```
### Shutdown Behavior
Normal shutdown behavior of kubernetes Pods and docker containers is:
1. send the process a SIGTERM
2. if the process has not exited after terminationGracePeriodSeconds, send SIGKILL

for the docker runtime, you can rewrite SIGTERM to a different signal using a
`STOPSIGNAL` layer/directive
(It's not clear whether this directive is formally supported by CRI)
```

@capileigh

stealthybox

Gotcha #3

Readiness / Liveness Probes

Kubernetes can't watch your logs...
but it can watch these probes:

- `Liveness used to check if Process is OK`
- `Readiness used to check if Pod should receive traffic`

Be intentional with timeouts and periods


@capileigh

stealthybox

Gotcha #4

PreStop lifecycle hook or.... in-app integration

- Important because the definition of Graceful shutdown for most programs is dissonant with Kubernetes' expectations:

Endpoints update async, independent of Pod Lifecycle.

kube-proxy and ingress-controllers depend on Endpoints.

When preStop is running or SIGTERM is sent,

your app will likely still be receiving connections.

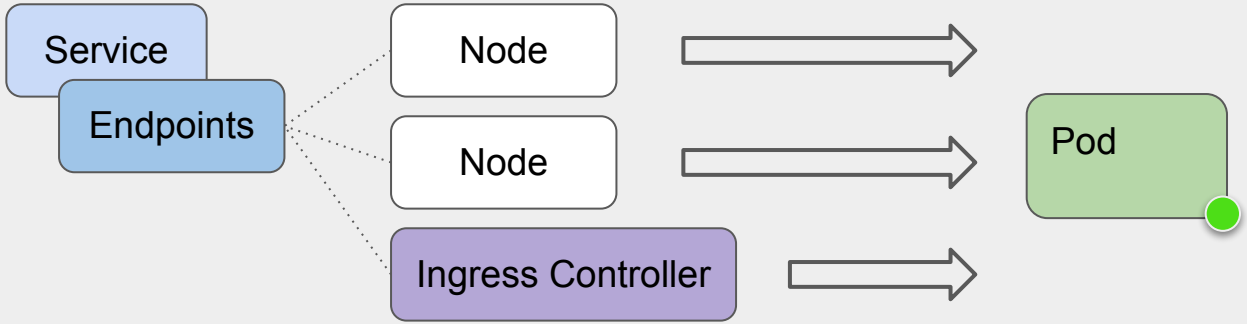
@capileigh

stealthybox

Gotcha #4

PreStop lifecycle hook or.... in-app integration

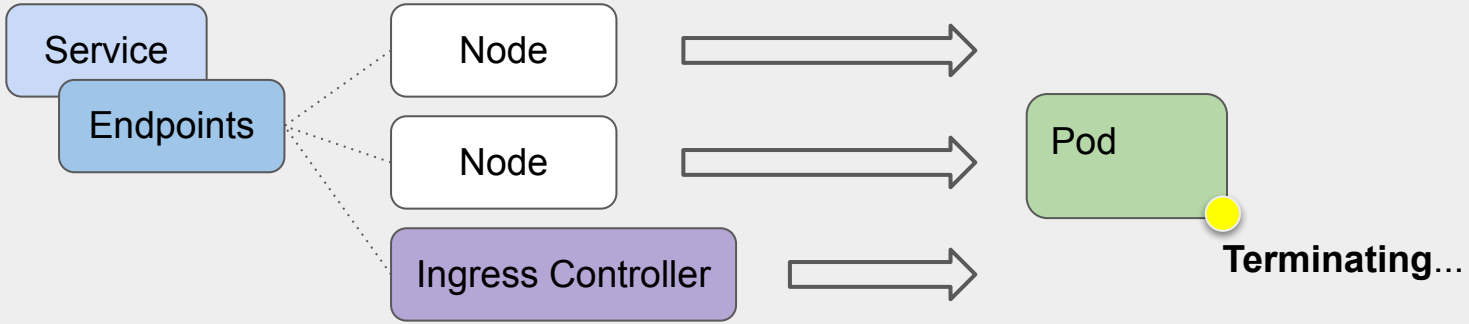
- Important because the definition of Graceful shutdown for most programs is dissonant with Kubernetes' expectations:



Gotcha #4

PreStop lifecycle hook or.... in-app integration

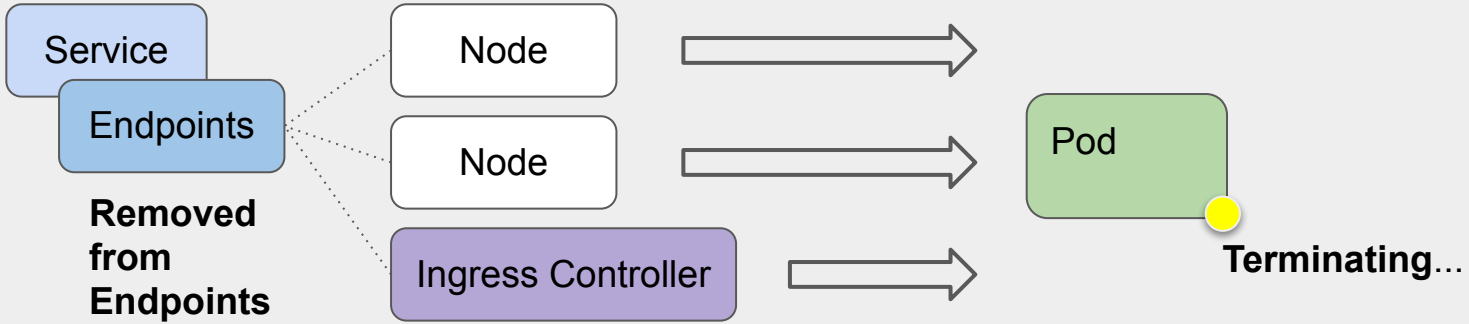
- Important because the definition of Graceful shutdown for most programs is dissonant with Kubernetes' expectations:



Gotcha #4

PreStop lifecycle hook or.... in-app integration

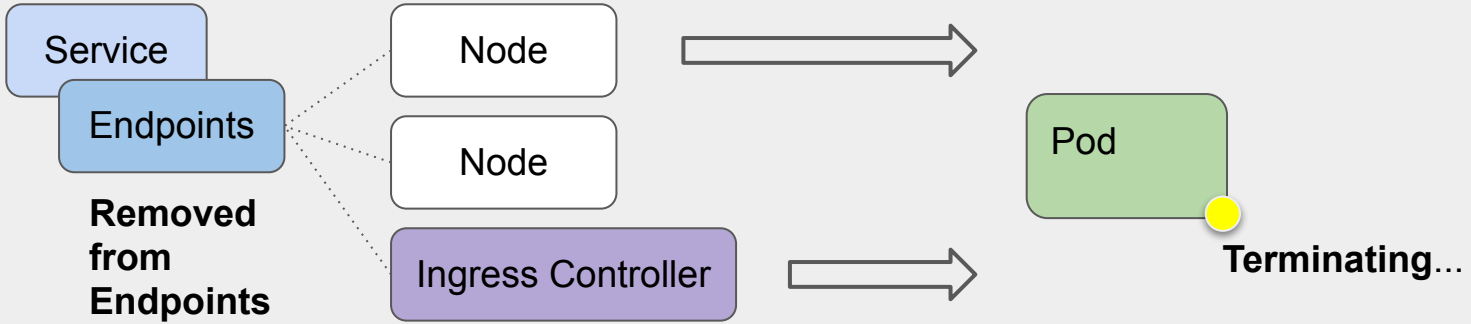
- Important because the definition of Graceful shutdown for most programs is dissonant with Kubernetes' expectations:



Gotcha #4

PreStop lifecycle hook or.... in-app integration

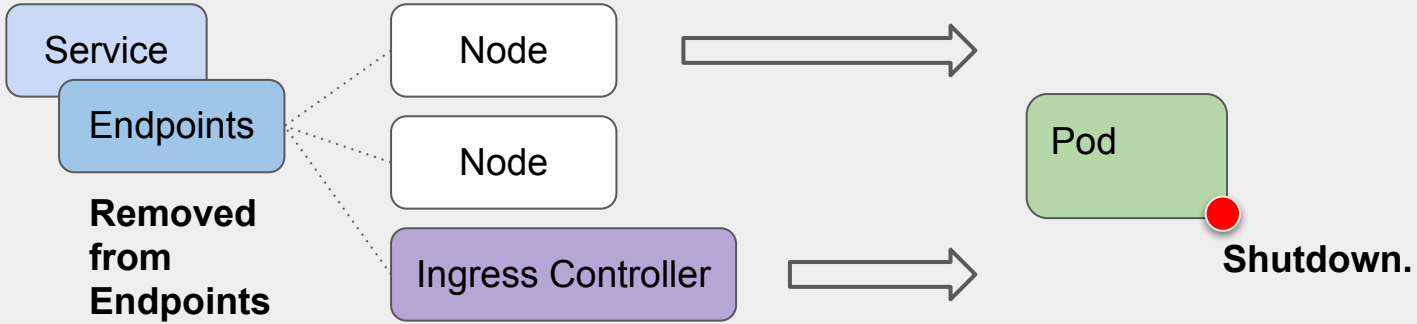
- Important because the definition of Graceful shutdown for most programs is dissonant with Kubernetes' expectations:



Gotcha #4

PreStop lifecycle hook or.... in-app integration

- Important because the definition of Graceful shutdown for most programs is dissonant with Kubernetes' expectations:



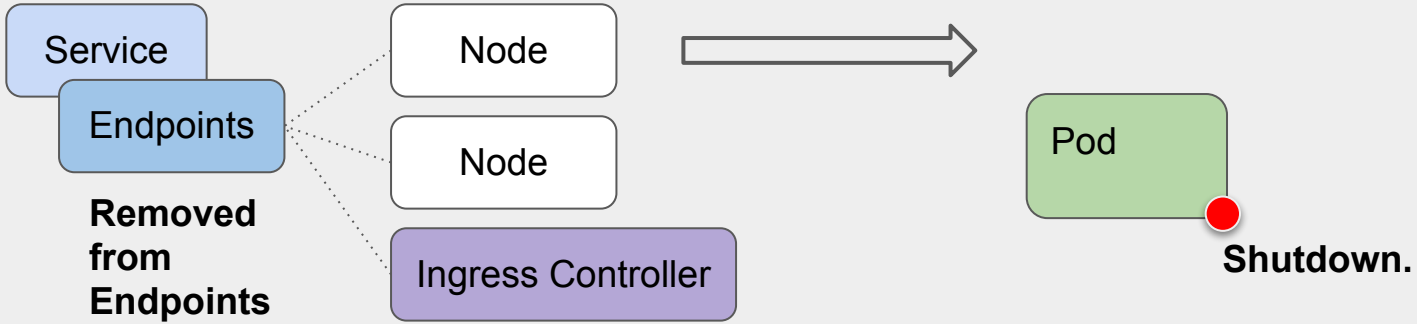

@capileigh

stealthybox

Gotcha #4

PreStop lifecycle hook or.... in-app integration

- Important because the definition of Graceful shutdown for most programs is dissonant with Kubernetes' expectations:

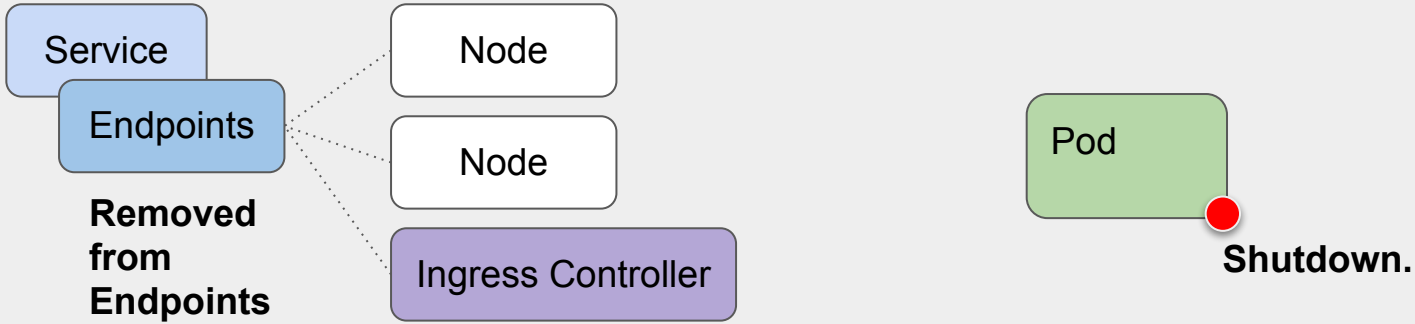


 @capileigh  stealthybox

Gotcha #4

PreStop lifecycle hook or.... in-app integration

- Important because the definition of Graceful shutdown for most programs is dissonant with Kubernetes' expectations:



Gotcha #4

PreStop lifecycle hook or.... in-app integration

- Important because the definition of Graceful shutdown for most programs is dissonant with Kubernetes' expectations:

"**Stop receiving** connections"
vs.
"**Start draining** connections"


@capileigh

stealthybox

Gotcha #4

PreStop lifecycle hook or.... in-app integration

- Important because the definition of Graceful shutdown for most programs is dissonant with Kubernetes' expectations:

Uwsgi: override the SIGTERM handler with an internal function...

https://github.com/unbit/uwsgi/issues/849#issuecomment-118869386

NGINX: need to build your own image with STOPSIGNAL SIGHUP

Was added in the upstream image and then removed

Both LB's also need "/bin/sleep" preStop hooks

@capileigh

stealthybox

Gotcha #4

PreStop lifecycle hook or.... in-app integration

- Important because the definition of Graceful shutdown for most programs is dissonant with Kubernetes' expectations:

- In-app integration leaks platform abstractions into your code
- Perhaps need to propose a new `sleep:` lifecycle hook for supporting "FROM scratch" images. (nothing to `exec:`)

@capileigh

stealthybox

https://github.com/stealthybox/zero-downtime

 @capileigh  stealthybox

Gotcha #5

Deployments:

.spec.strategy.rollingUpdate.**maxUnavailable**
Use percentage or 0 when replica count == 1

This used to default to 1 -- Fixed in apps/v1beta1:
https://github.com/kubernetes/kubernetes/pull/39683
Make sure you're using >= **apps/v1** API


@capileigh

stealthybox

Gotcha #6

Deployments:

Make sure that your app can stay warm according to these periods:

.spec.strategy.**minReadySeconds**

.spec.strategy.**progressDeadlineSeconds**

Also take care that this does not exceed capacity:

.spec.strategy.**rollingUpdate.maxSurge**

@capileigh

stealthybox

Gotcha #7

Mismatched signal lifecycle with side-cars:

Example:

If you're using **cloudsql-proxy** to connect your app to your db, your preStop hooks and graceful shutdown periods should be either **synchronized** or **scheduled** so that they do not effectively race.

If your app is in graceful shutdown and the proxy is not sleeping, it will exit and drop your db connections.

 @capileigh  stealthybox

Rules of Uptime

1. **entrypoint** should handle or pass **signals**
2. **STOPSIGNAL** may need to be changed
3. Use diff. periods for **Liveness/Readiness** Probes
4. Sleep in **preStop** hooks to drain connections
5. Use the newer **apps/v1** Deployment
6. Keep your app **warm** during a RollingUpdate
7. Synchronize shutdown of **side-cars**

 @capileigh  stealthybox

@capileigh

stealthybox