

~~Why look for something else when Kubernetes does exist?~~

Nomad as container orchestrator - lessons learned

Anton Vorobiev
Infrastructure tech lead @Showmax
Swiss army knife @vorobiev.cloud

What is Showmax?

A world-scalable platform, delivering the best SVOD experience on the African continent and beyond.

- Built on microservice architecture (we love containers)
- We are distributed (EU + Africa)
- Engineering in Prague and Beroun



<https://kinkyrhino.co.za/product/showmax-and-chill>

I'm Anton.

- In Showmax:
 - Infrastructure tech lead
(assisting operation team)
- In vorobiev.cloud:
 - Swiss-army knife (helping companies to adopt k8s and automate stuff)

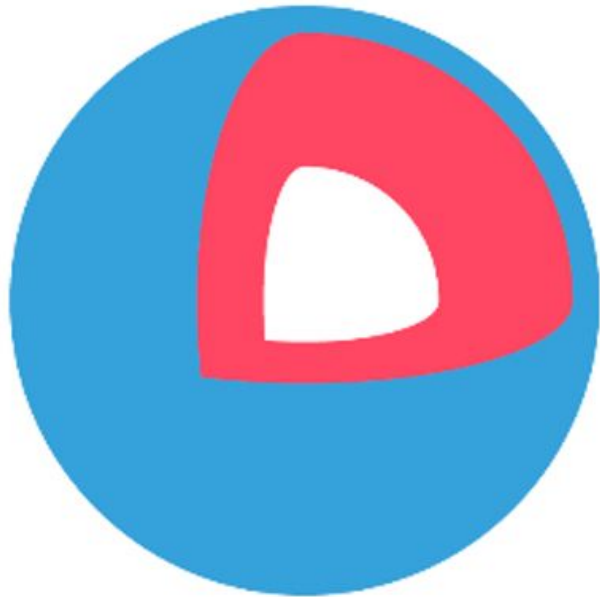


I'm Anton.



Past

- CoreOS fleet
 - a distributed init system
 - systemd-like unitfiles
 - **custom** tooling
 - **rollout** integration
- <https://12factor.net/>
- <https://content.pivotal.io/blog/beyond-the-twelve-factor-app>



Present



Deprecation warning



fleet is no longer developed or maintained by CoreOS. After February 1, 2018, it will no longer be available from the CoreOS Quay registry, but will not be shipped as part of Container Linux.

<https://github.com/coreos/fleet>

Future ???



Almost any k8s howto *

How to ~~draw an owl~~

set up a k8s cluster

1.



2.



set up some etcds

1. ~~Draw some circles~~

do

kubernetes

2. ~~Draw~~ the rest of the ~~drawing~~

<https://knowyourmeme.com/photos/572078-how-to-draw-an-owl>

* “k8s the hard way” is just an exception that proves the rule

Almost any k8s howto

How to ~~draw an owl~~

set up a k8s cluster

1.



2.



set up some etcds

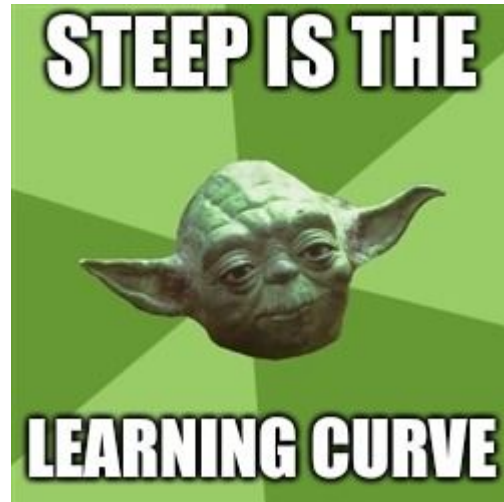
1. ~~Draw some circles~~

do

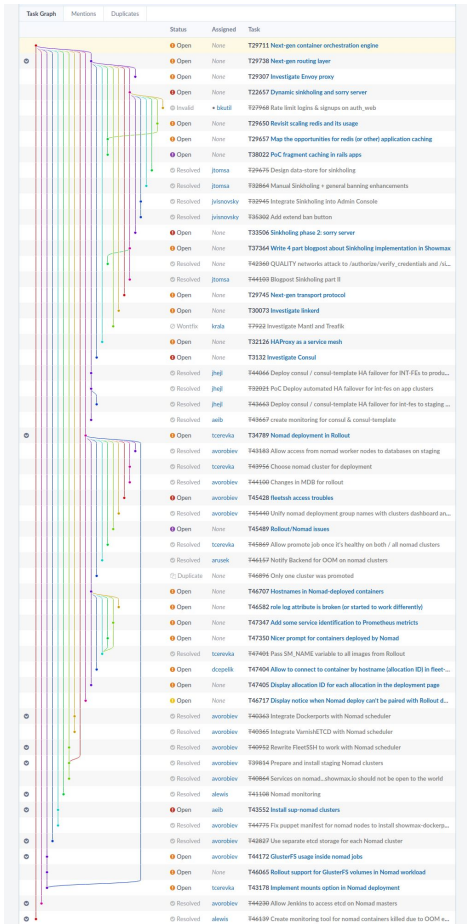
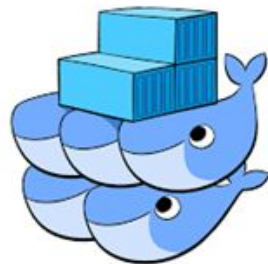
kubernetes

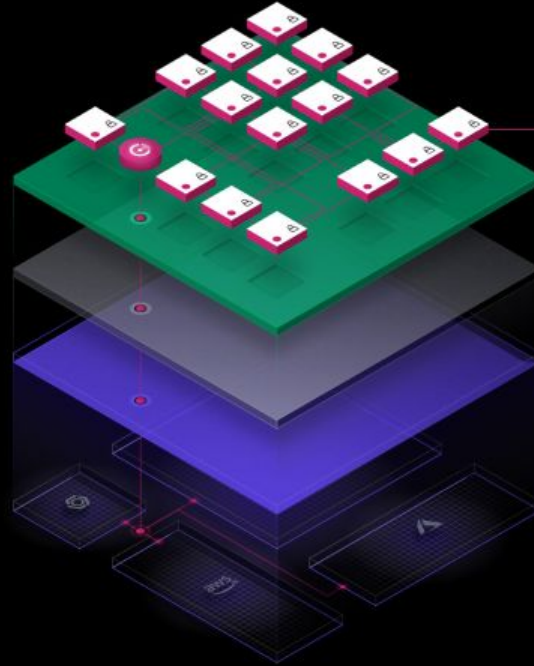
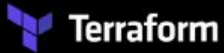
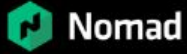
2. ~~Draw~~ the rest of the ~~drawing~~

<https://knowyourmeme.com/photos/572078-how-to-draw-an-owl>



Future





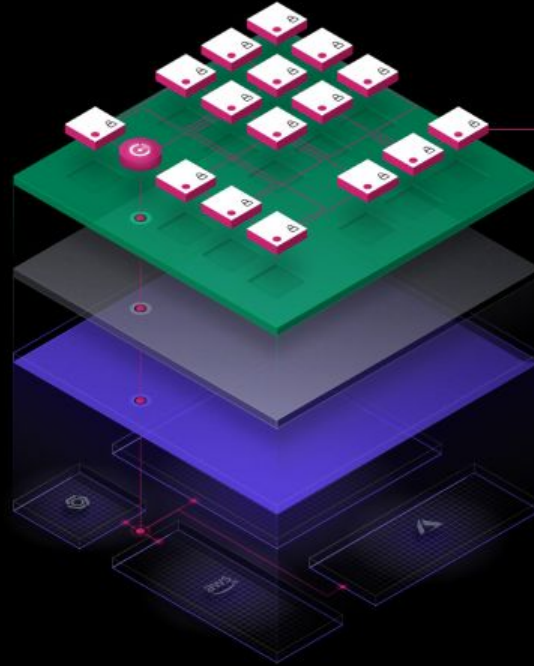
Networking
Connect infrastructure
and applications

Development
Run applications

Security
Secure infrastructure
and applications

Operations
Provision infrastructure





Networking
Connect infrastructure
and applications

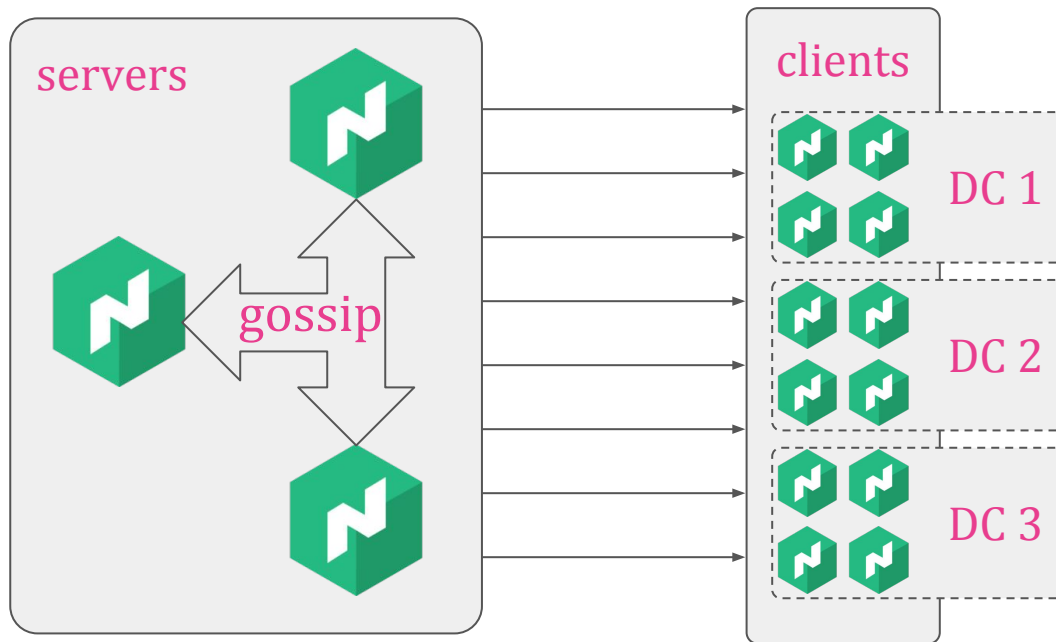
Development
Run applications

Security
Secure infrastructure
and applications

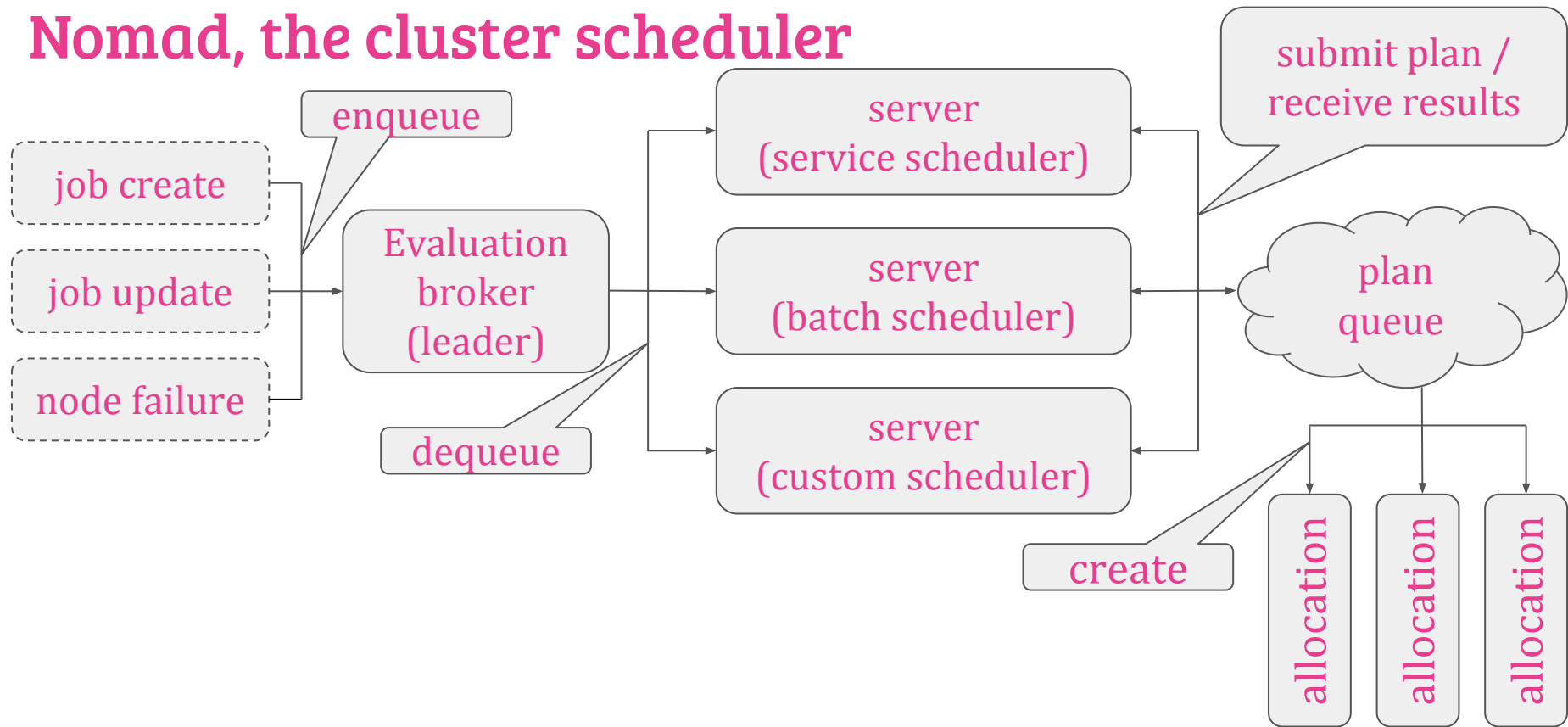
Operations
Provision infrastructure



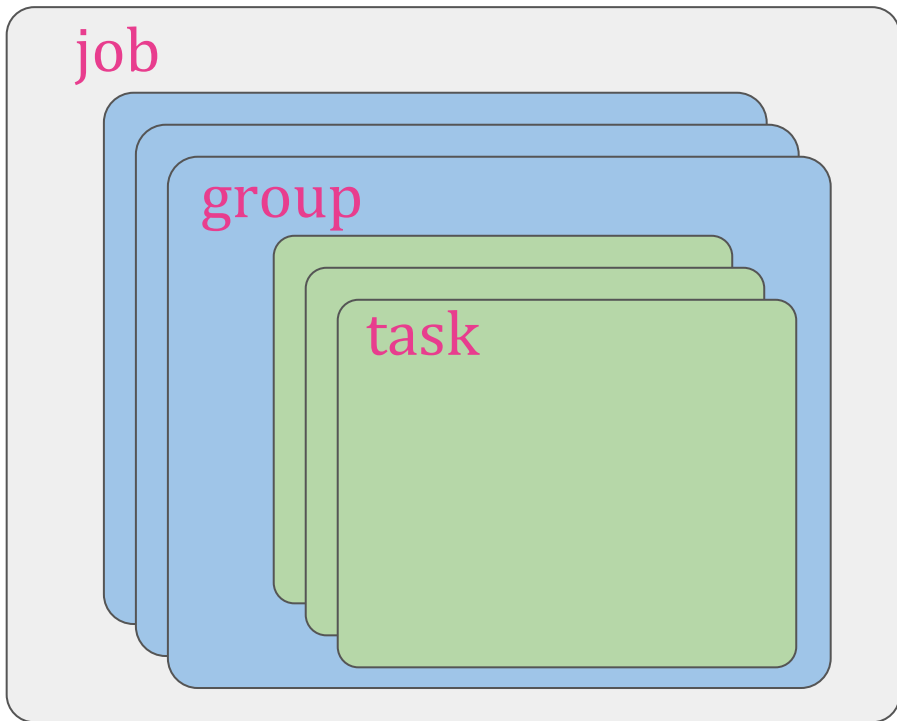
Nomad, the cluster scheduler



Nomad, the cluster scheduler



Nomad, the cluster scheduler



```
job "docs" {  
  datacenters = ["dc1"]  
  
  group "example" {  
    task "server" {  
      driver = "docker"  
  
      config {  
        image = "hashicorp/http-echo"  
        args = [  
          "-listen", ":5678",  
          "-text", "hello world",  
        ]  
      }  
  
      resources {  
        network {  
          mbits = 10  
          port "http" {  
            static = "5678"  
          }  
        }  
      }  
    }  
  }  
}
```


Nomad, the cluster scheduler

- Flexible (enough) deployments
 - service
 - batch
 - system
- Schedules apps that can't easily be containerized
- Device plugins
- Resource limits
- Scheduling (anti-)affinities
- Scheduling constraints

1 million containers challenge

- Cluster of 5 Nomad servers
- 5000 Nomad clients
- 1 000 000 containers under 5 minutes
- rate of 3,750 containers per second

<https://www.hashicorp.com/c1m>

Traffic balancer

fabio

Overrides 1.5.9 Github

Routing Table

type to filter routes

#	Service	Source	Dest	Options	Weight
1	apache-webserver	/	http://172.31.55.41:20119/		33.33%
2	apache-webserver	/	http://172.31.55.174:27708/		33.33%
3	apache-webserver	/	http://172.31.53.180:21188/		33.33%



pic: <https://www.nomadproject.io/guides/load-balancing/fabio.html>

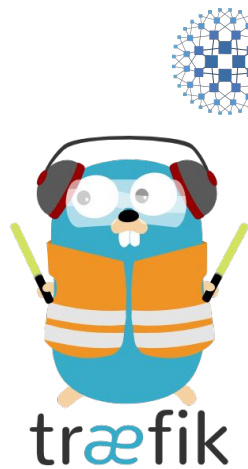
<https://fabio.net>

```
job "fabio" {
  datacenters = ["dc1"]
  type = "system"

  group "fabio" {
    task "fabio" {
      driver = "docker"
      config {
        image = "fabiolb/fabio"
        network_mode = "host"
      }

      resources {
        cpu = 200
        memory = 128
        network {
          mbits = 20
          port "lb" {
            static = 9999
          }
          port "ui" {
            static = 9998
          }
        }
      }
    }
  }
}
```

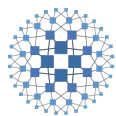
Traffic balancer



traefik



VARNISH
SOFTWARE



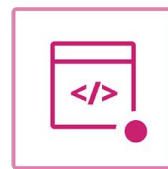
HAProxy



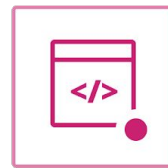
INTERNET



LOAD BALANCER



APP



APP



APP



CONSUL
CLUSTER

CONSUL
TEMPLATE



pic: <https://learn.hashicorp.com/consul/integrations/nginx-consul-template>

Demo (hopefully)

Things that hurt

- Small community - often times no choice of a tool with single available option (e.g. Fabio)
- Small community - low update/maintenance velocity (e.g. Spark on Nomad)
- `${NOMAD_ALLOC_INDEX}` / `${NOMAD_ALLOC_ID}` - worked around
- `localhost bind()` - WIP
(<https://github.com/hashicorp/nomad/issues/2615>)

Some state of the art WTF stuff

- Inconsistent cli output

```
Placement Metrics
Node                job-anti-affinity  node-reschedule-penalty  node-affinity  binpack  final score
6fc7950f-6af1-001f-dc85-76690174e6a2  0                0                        0              0.192    0.192
16a5491e-a3ed-8f4e-af18-e29e05f45b52  0                0                        0              0.192    0.192
d711dc5e-20f5-fb6f-f227-2c91081672d2  -1               0                        0              0.342    -0.329
root@master:~#
```

```
Placement Metrics
Node                node-affinity  binpack  job-anti-affinity  node-reschedule-penalty  final score
6fc7950f-6af1-001f-dc85-76690174e6a2  0             0.192    0                  0                        0.192
16a5491e-a3ed-8f4e-af18-e29e05f45b52  0             0.192    0                  0                        0.192
d711dc5e-20f5-fb6f-f227-2c91081672d2  0             0.342    -1                 0                        -0.329
root@master:~#
```


Some state of the art WTF stuff

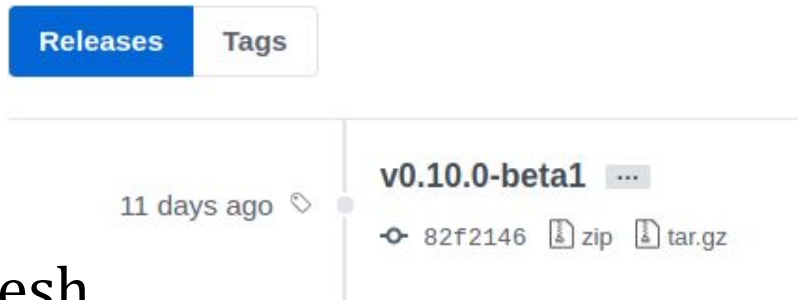
- Inconsistent



<https://imgur.com/gallery/JmYCgIq>

(Near) future

- Consul Connect based service mesh
- Shared network namespace within task group
- Host Volumes



<https://www.hashicorp.com/blog/hashicorp-nomad-0-10-beta>

The point is



<https://me.me/i/you-must-choose-but-choose-wisely-meme-com-3791503>

THANK YOU!

Get in touch!

Anton Vorobiev

 [@vorobiev cloud](https://twitter.com/vorobiev_cloud)
anton@vorobiev.cloud