BOOSTCOM

Kubernetes & Ruby: From Dedicated Servers To Kubernetes

Jakub Kruczek

- General description & ideas (no time to tell about everything, sorry :))
- Join official K8s slack at http://slack.k8s.io/

  channel **#pl-users** & #kubernetes-ruby

Key presentation agenda

1. Past, Now (& Future)
2. Security (remember!)
3. Let’s deploy something
4. Monitoring & Logging
5. Encountered problems

About me

Software Engineer at Boostcom for >3.5y

coding mainly in Ruby, TypeScript and Angular

Started working with Kubernetes >1.5y ago

Deployed, configured and maintain 4 K8s clusters

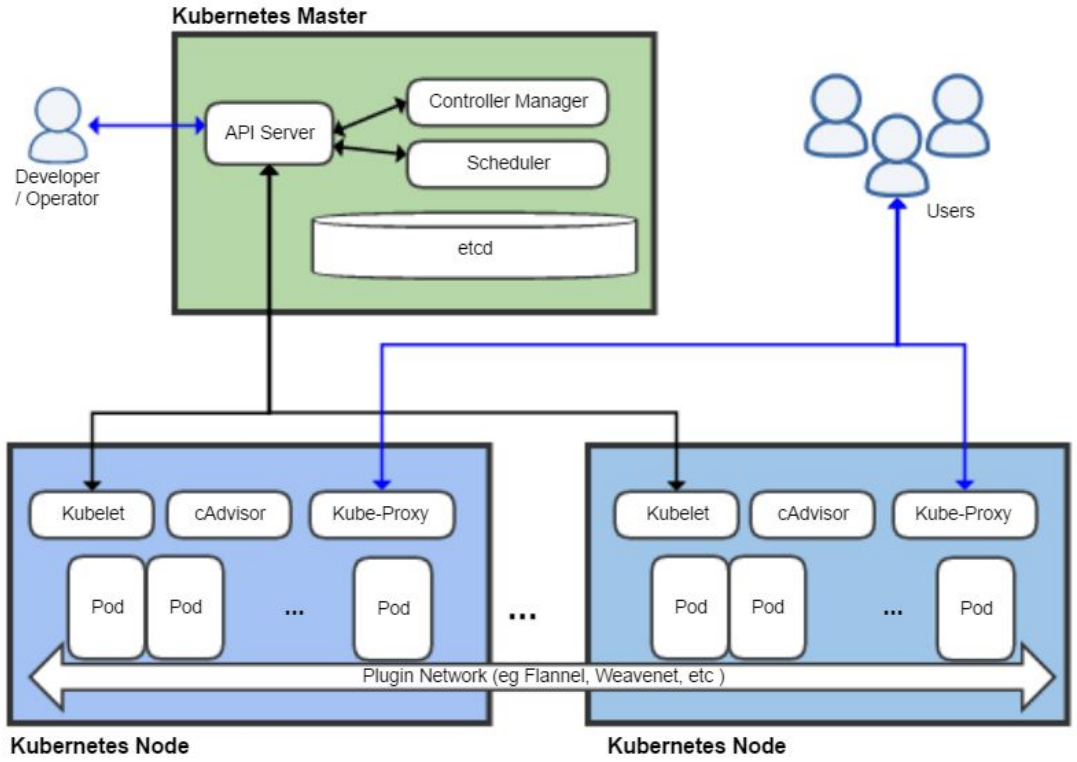
not counting in deleted ones

I like cats :)

BOOSTCOM

What are the Kubernetes?

Production-Grade Container Orchestration

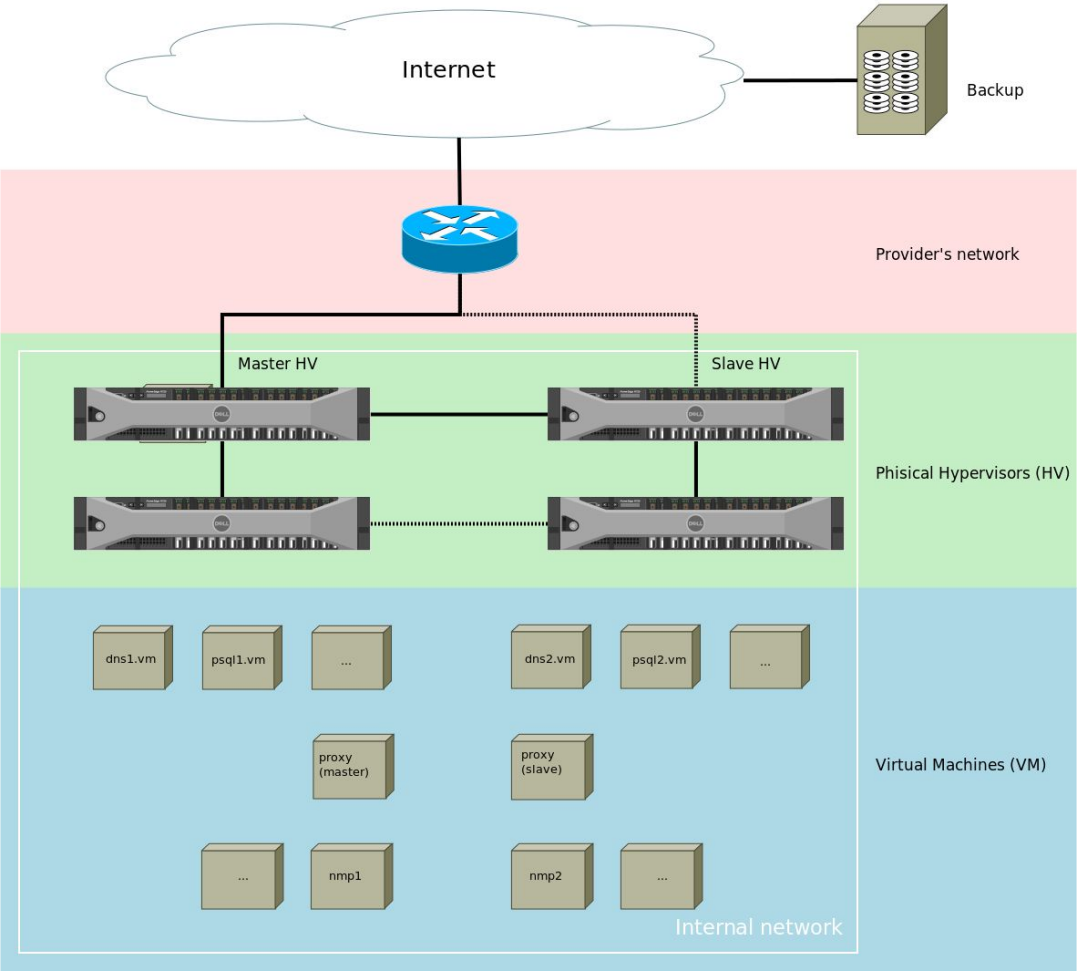


Pod - 1..N container, storage resources, a unique network IP, and additional options

Past, Now (& Future)

Dedicated servers -> Kubernetes

Past - dedicated servers

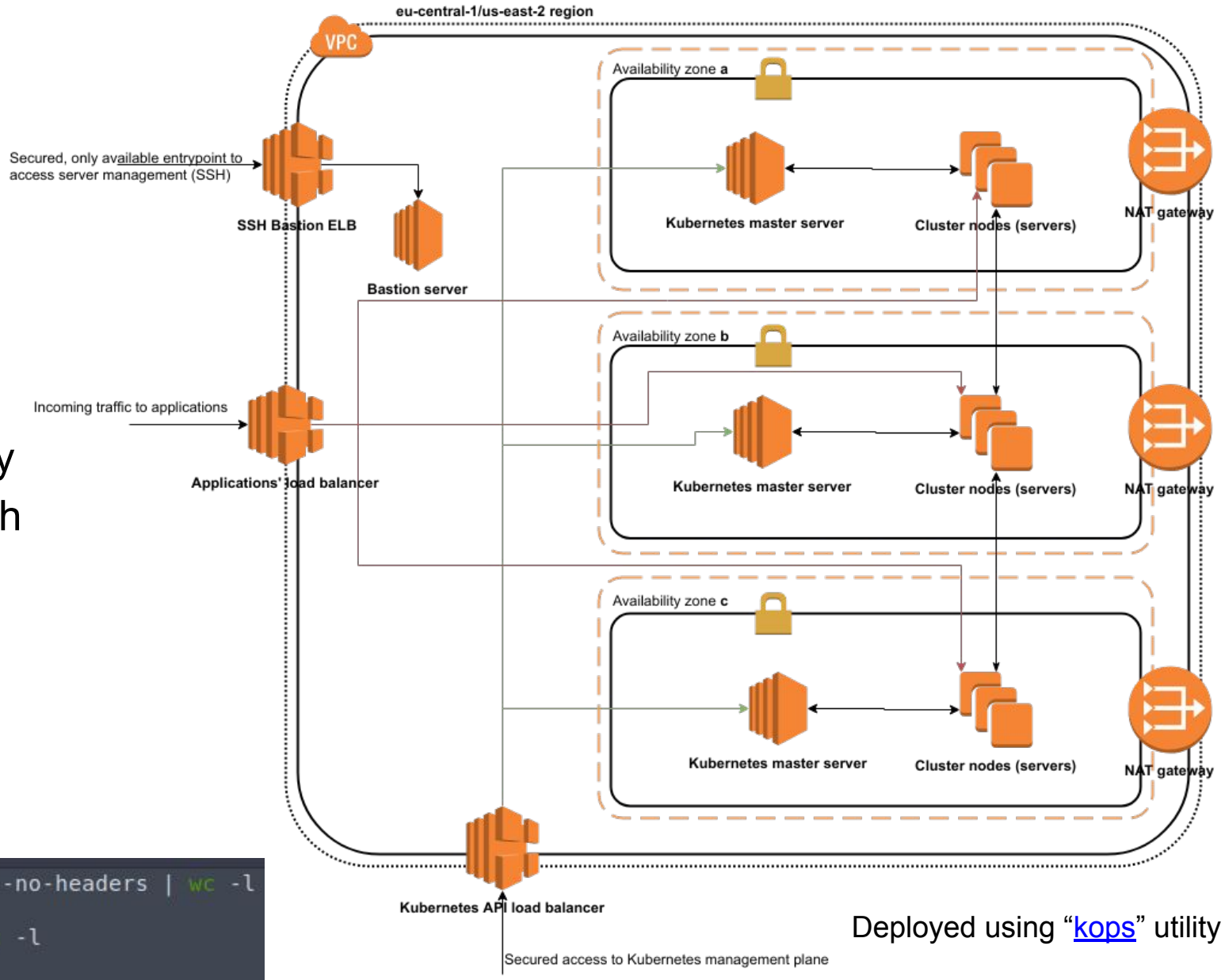


- No real, on-demand scalability
- Single location
- No “Availability Zones”
- Hard to operate and make changes
- GRE tunnels between HV (single tunnel down == everything down...)
- Very custom failover scripts :)
- Very much work needed to make something very simple
- …

~once a month

Now

- Scalable
- Resilient (and fully HA)
- Multiple locations over the world
- Quite easy to operate by developers without much “servers experience”




```
[kruczjak:~] % kprod get pods --all-namespaces --no-headers | wc -l
354
[kruczjak:~] % kprod get nodes --no-headers | wc -l
14
[kruczjak:~] %
```

Deployed using “kops” utility

Ruby App migration

Considerations

- Persistent storage (remember about zones (AWS)!)
- Multiple replicas running & HA
- Shutting down & Starting up (scaling, reboot, etc.)
- CPU & RAM usage - Kubernetes requests and limits

Steps

1. Dockerfile!
2. CI/CD scripts
3. Liveness & Readiness probes
4. Kubernetes configuration (yamls, encrypted secrets)
5. ECR repository (just like Dockerhub)
6. Logging adjustments (STDOUT, format)
7. Monitoring!

Total time to migrate single app: ~2 workdays

Ruby App migration - container image

Dockerfile

```
FROM ruby:2.5.1-alpine

RUN apk --no-cache --update add postgresql-dev tzdata git bind-tools

# Gemfile for caching
ADD Gemfile /app/
ADD Gemfile.lock /app/
WORKDIR /app

RUN apk --no-cache --update add --virtual build-dependencies ruby-dev build-base libxml2-dev && \
    gem install bundler --no-doc && \
    BUNDLE_FORCE_RUBY_PLATFORM=1 bundle install -j 4 --deployment --without development test && \
    apk del build-dependencies

ADD . /app
```

	Alpine	Debian
	398MB	**1.3GB**
500Mb/s	**6.4s**	**20.8s**

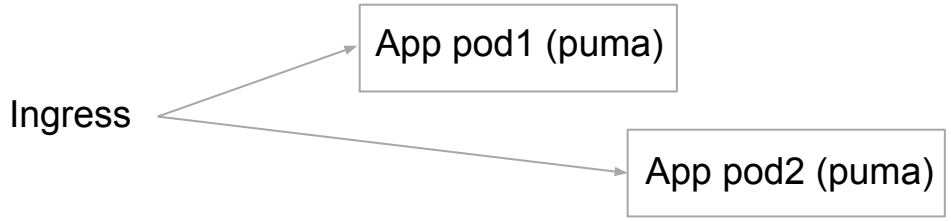
Ruby App migration - resources

https://github.com/linki/chaoskube
“chaoskube periodically kills random pods in your Kubernetes cluster”

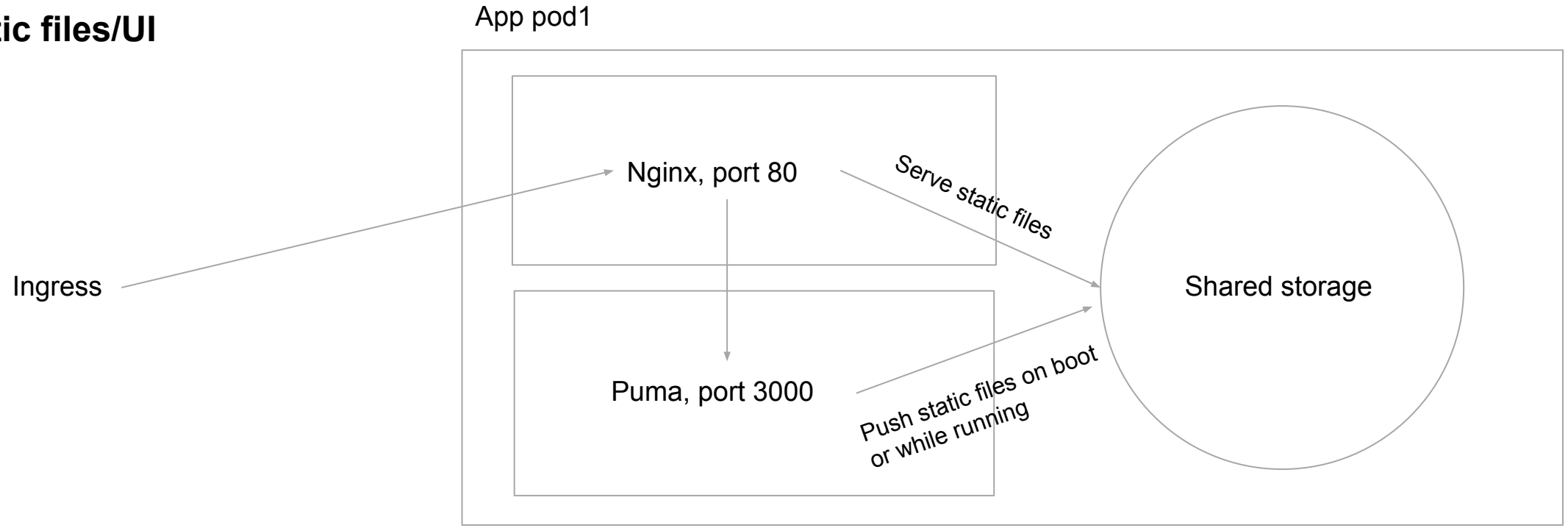
https://kubernetes.io/docs/concepts/workloads/controllers/jobs-run-to-completion/

Ruby App migration - static files/UI

API



Static files/UI



Security (remember!)

K8s without proper configuration isn’t so secure...

I WANT TO EXPLOIT KUBELET!

Inside pod with PHP application

```
curl -Lk https://<nodeip>:10250/runningpods
curl -Lk -X POST https://<nodeip>:10250/run/kube-system/<kube-apiserver-podname>/kube-apiserver -d "cmd=cat
/srv/kubernetes/known_tokens.csv"
```

Solution

- Never allow to connect to node's directly! (hide nodes' IPs)
- Disable "anonymous-auth" on Kubelet
- Enable RBAC authorization on Kubelet

```
/app # curl -k https://172.20.58.165:10250/runningpods/
Unauthorized/app #
```

AWS & EC2 Roles

```
curl -s http://169.254.169.254/latest/user-data | grep -A 1 channels
aws s3 ls s3://<s3_name>/
```

Solution
Use e.g. **kube2iam**

```
/app # curl -s http://169.254.169.254/latest/user-data | grep -A 1 channels
channels:
- s3://[redacted]/addons/bootstrap-channel.yaml
/app # aws s3 ls s3://[redacted]

An error occurred (AccessDenied) when calling the ListObjectsV2 operation: Access Denied
/app #
```

Ways to access K8s API

- Basic Auth
- Using certificates generated using main CA
- (ServiceAccount) Tokens
- **OpenID Connect Tokens**

  - Google
  - GitHub
  - …
  - **KeyCloak (2FA, roles, groups)**

https://github.com/keycloak/keycloak-gatekeeper

OAuth2 (e.g. Grafana, Discourse)

Kibana/ES management dashboard

Prometheus UI (name any app)

Let’s deploy something

Manual deployments are soooo funny xD

Deployment overview

Push to repository

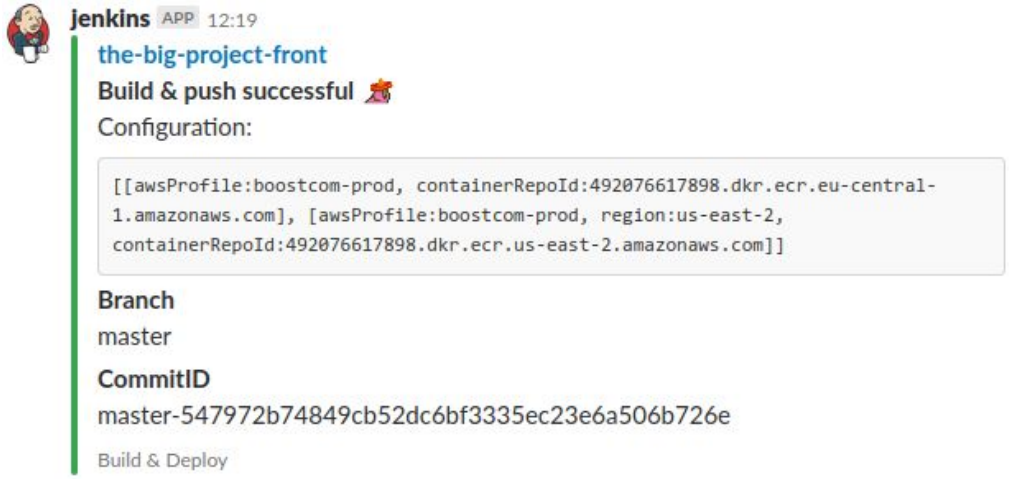
Manual start

HTTPS request

Deployer

Jenkins

- Run tests
- Runs pipeline library:
  - Build Docker image (alpine base)
  - Push image to AWS ECR repository



Deployer

```
[kruczjak:~/git/boost/nmp_dashboard/.deploy] develop+* ± tree
.
├── config.yaml
├── dev
│   ├── 10-configuration
│   │   ├── a59582db0df375a88d2998291d64fd9efd7335a4464cab58c12993ba2508d85d
│   │   ├── nmp-dashboard-config.configmap.yaml.erb
│   │   ├── nmp-dashboard-nginx-config.configmap.yaml.erb
│   │   ├── nmp-dashboard-secrets.secret.yaml.erb
│   │   └── secrets.ejson
│   ├── 20-predeployment
│   │   └── nmp-dashboard-db-migrate.job.yaml.erb
│   ├── 30-deployment
│   │   ├── nmp-dashboard-puma.ingress.yaml
│   │   ├── nmp-dashboard-puma.web-server-v1.tmpl.yaml.erb
│   │   └── nmp-dashboard-sidekiq.sidekiq-v1.tmpl.yaml.erb
│   └── 40-postdeployment
│       └── nmp-dashboard-seeds.job.yaml.erb
└── prod
    ├── 10-configuration
    │   ├── nmp-dashboard-config.configmap.yaml.erb
    │   ├── nmp-dashboard-nginx-config.configmap.yaml.erb
    │   ├── nmp-dashboard-secrets.secret.yaml.erb
    │   └── secrets.ejson
    ├── 20-predeployment
    │   └── nmp-dashboard-db-migrate.job.yaml.erb
```

shared configuration

deployment step

environment

"template"

encrypted secrets in app repository

Deployer APP 15:36

Boostcom Jenkins

reporting_integrations

⚠️ Starting deployment

environment

stg

Image

master-c3a82c782fd4ed53187e9a949acfd0263ed69e10

Boostcom Deployer

Boostcom Jenkins

reporting_integrations

Error encountered: `Psych::SyntaxError`

```
/usr/local/lib/ruby/2.4.0/psych.rb:377:in `parse': (<unknown>): could not find
expected ':' while scanning a simple key at line 27 column 3
```

environment

stg

Image

master-c3a82c782fd4ed53187e9a949acfd0263ed69e10

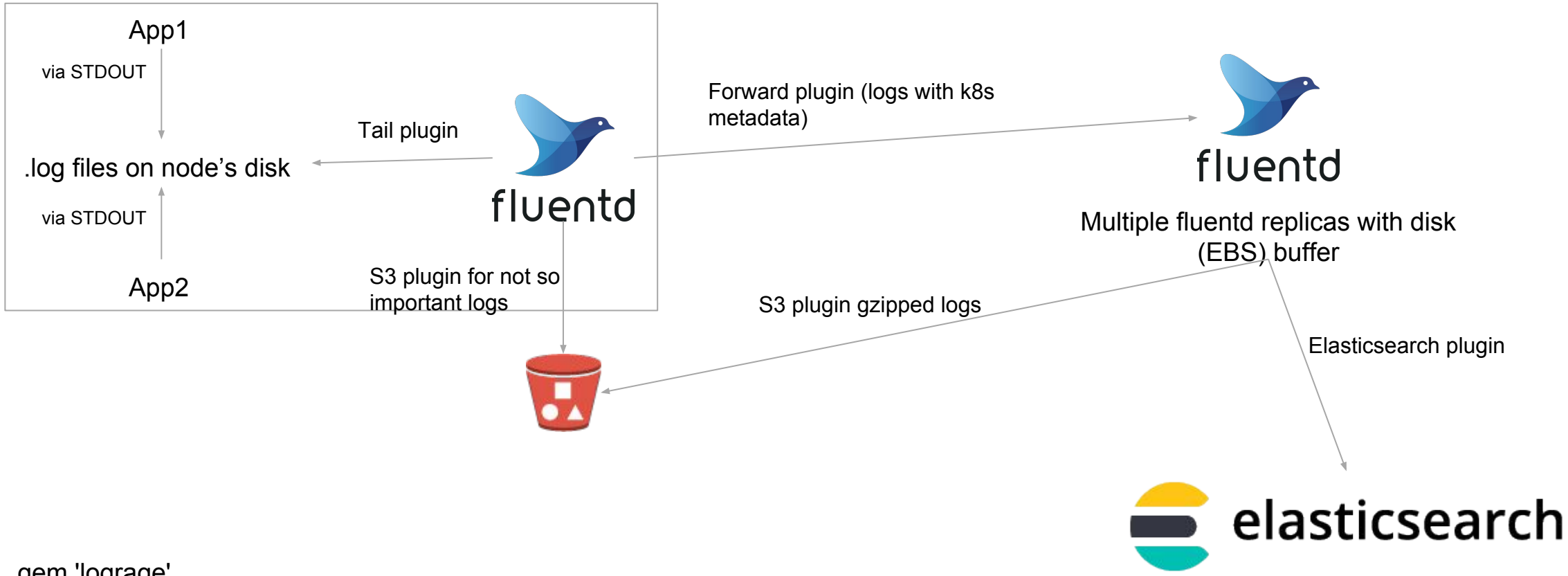
Boostcom Deployer

Monitoring & Logging

Is one thing enough?

Logging - EFK stack on k8s

Single node




```
gem 'lograge'
gem 'logstash-event'
gem 'logstash-logger'
gem 'grape_logging' # for grape
```

Logging - EFK stack on k8s

November 18th 2018, 01:32:49	/api/infinity-mall/settings/webform_options	200	infinity-mall	[200] GET /api/infinity-mall/settings/webform_options (Api::SettingsController#webform_options)

Table JSON

View surrounding documents | View single document

⏲	@timestamp	November 18th 2018, 01:32:49.356
t	@version	1
t	_id	5b48JGcBHM8OAx0yIypo
t	_index	containers_bl_webforms-api-puma_2018.11.18
#	_score	-
t	_type	fluentd
t	action	webform_options
t	controller	Api::SettingsController
t	docker.container_id	3f96bcb19c58041e5428782a6eaf216c5620d823fa80cf3a8731002cc4a18577
#	duration	35.75
t	format	json
t	host	webforms-api-puma-7bddcd5bbd-gqmzp
t	kubernetes.container_name	webforms-api-puma
t	kubernetes.host	ip-172-20-105-25.eu-central-1.compute.internal
t	kubernetes.labels.pod-template-hash	3688781668
t	kubernetes.labels.run	webforms-api-puma
t	kubernetes.master_url	https://100.64.0.1:443/api
t	kubernetes.namespace_id	fe80faff-0669-11e8-85ca-0a82f5c8558a
t	kubernetes.namespace_name	bl
t	kubernetes.pod_id	2995c0e6-e865-11e8-9809-0a9195bbab9a
t	kubernetes.pod_name	webforms-api-puma-7bddcd5bbd-gqmzp
t	log	{"method":"GET","path":"/api/infinity-mall/settings/webform_options","format":"json","controller":"Api::SettingsController","action":"webform_options","status":200,"duration":35.75,"view":0.9,"params":{"source":"webforms","slug":"infinity-mall"},"@timestamp":"2018-11-18T00:32:49.356+00:00","@version":"1","message":"[200] GET /api/infinity-mall/settings/webform_options (Api::SettingsController#webform_options)","severity":"INFO","host":"webforms-api-puma-7bddcd5bbd-gqmzp","tags":["149.156.124.9"]}
t	message	[200] GET /api/infinity-mall/settings/webform_options (Api::SettingsController#webform_options)
t	method	GET
t	params.slug	infinity-mall
t	params.source	webforms
t	path	/api/infinity-mall/settings/webform_options

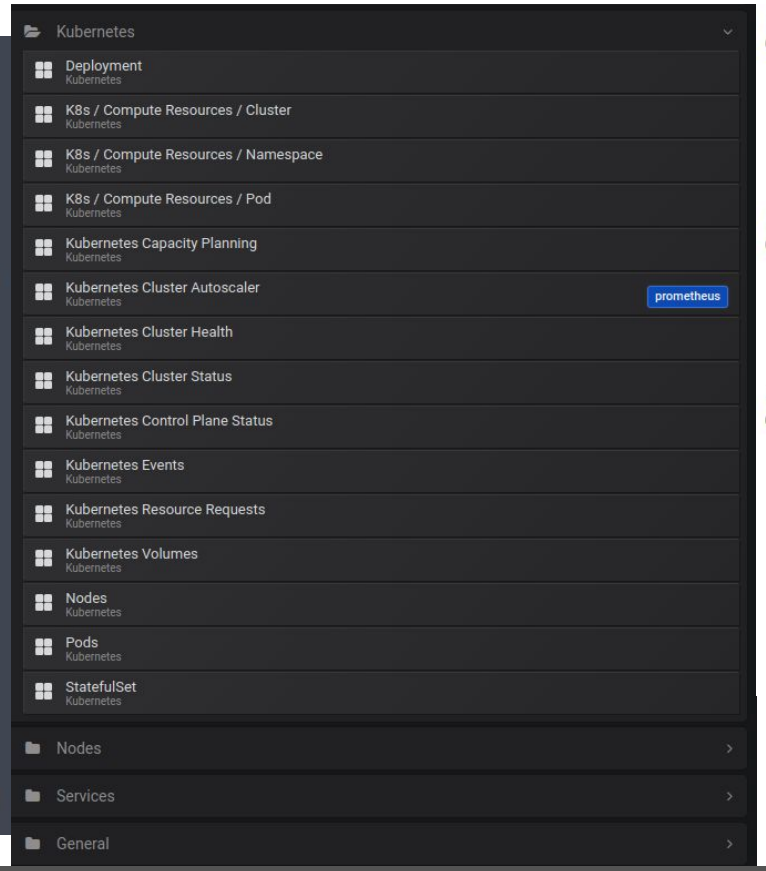
Monitoring - Prometheus & Grafana

https://github.com/coreos/prometheus-operator - service discovery, very easy to configure new metrics from pods and alerts

Libs

- https://github.com/discourse/prometheus_exporter
- https://github.com/yabeda-rb

```
[kruczjak:~] 130 % kprod get pods -n monitoring
NAME
alertmanager-prometheus-operator-alertmanager-0
alertmanager-prometheus-operator-alertmanager-1
alertmanager-prometheus-operator-alertmanager-2
alerts-keycloak-proxy-7664786d76-5jd49
grafana-6f98fc7c5d-67wq7
kube-slack-7784cdf55c-4x6cz
monitoring-grafana-5bc84c849b-fddjd
postgresql-cluster-production-prometheus-prometheus-postgrc27vb
prometheus-keycloak-proxy-7dbf58645c-4bjq8
prometheus-operator-kube-state-metrics-69995f9794-q4xqr
prometheus-operator-operator-6dc4454b7b-c9n8x
prometheus-operator-prometheus-node-exporter-2p9vz
prometheus-operator-prometheus-node-exporter-775n8
prometheus-operator-prometheus-node-exporter-9dtnw
prometheus-operator-prometheus-node-exporter-crdqq
prometheus-operator-prometheus-node-exporter-czbxp
prometheus-operator-prometheus-node-exporter-dk2dw
prometheus-operator-prometheus-node-exporter-f8z8v
prometheus-operator-prometheus-node-exporter-gthg2
prometheus-operator-prometheus-node-exporter-k8dkv
prometheus-operator-prometheus-node-exporter-kx97x
prometheus-operator-prometheus-node-exporter-pmqxt
prometheus-operator-prometheus-node-exporter-sf5sx
prometheus-operator-prometheus-node-exporter-wtvwc
prometheus-operator-prometheus-node-exporter-zw9v2
prometheus-prometheus-operator-prometheus-0
prometheus-prometheus-operator-prometheus-1
redis-cluster-production-2-prometheus-prometheus-redis-expvs6t6
redis-cluster-production-prometheus-prometheus-redis-expor4nl7l
```



K8s PROD AlertManager APP 22:56

[FIRING:1] sidekiq-dmp-audience (SidekiqTooManyEnqueuedJobs sidekiq 9292 100.115.240.8:9292 utilities sidekiq-console-web-654b7678-cbttg monitoring/prometheus-operator-prometheus sidekiq-console-metrics-dmp-audience critical)

Sidekiq sidekiq-dmp-audience has queue higher than 300 for 7 minutes. Current: 38376

K8s PROD AlertManager APP 23:31

[RESOLVED] sidekiq-dmp-audience (SidekiqTooManyEnqueuedJobs sidekiq 9292 100.115.240.8:9292 utilities sidekiq-console-web-654b7678-cbttg monitoring/prometheus-operator-prometheus sidekiq-console-metrics-dmp-audience critical)

Sidekiq sidekiq-dmp-audience has queue higher than 300 for 7 minutes. Current:

K8s PROD AlertManager APP 12:29

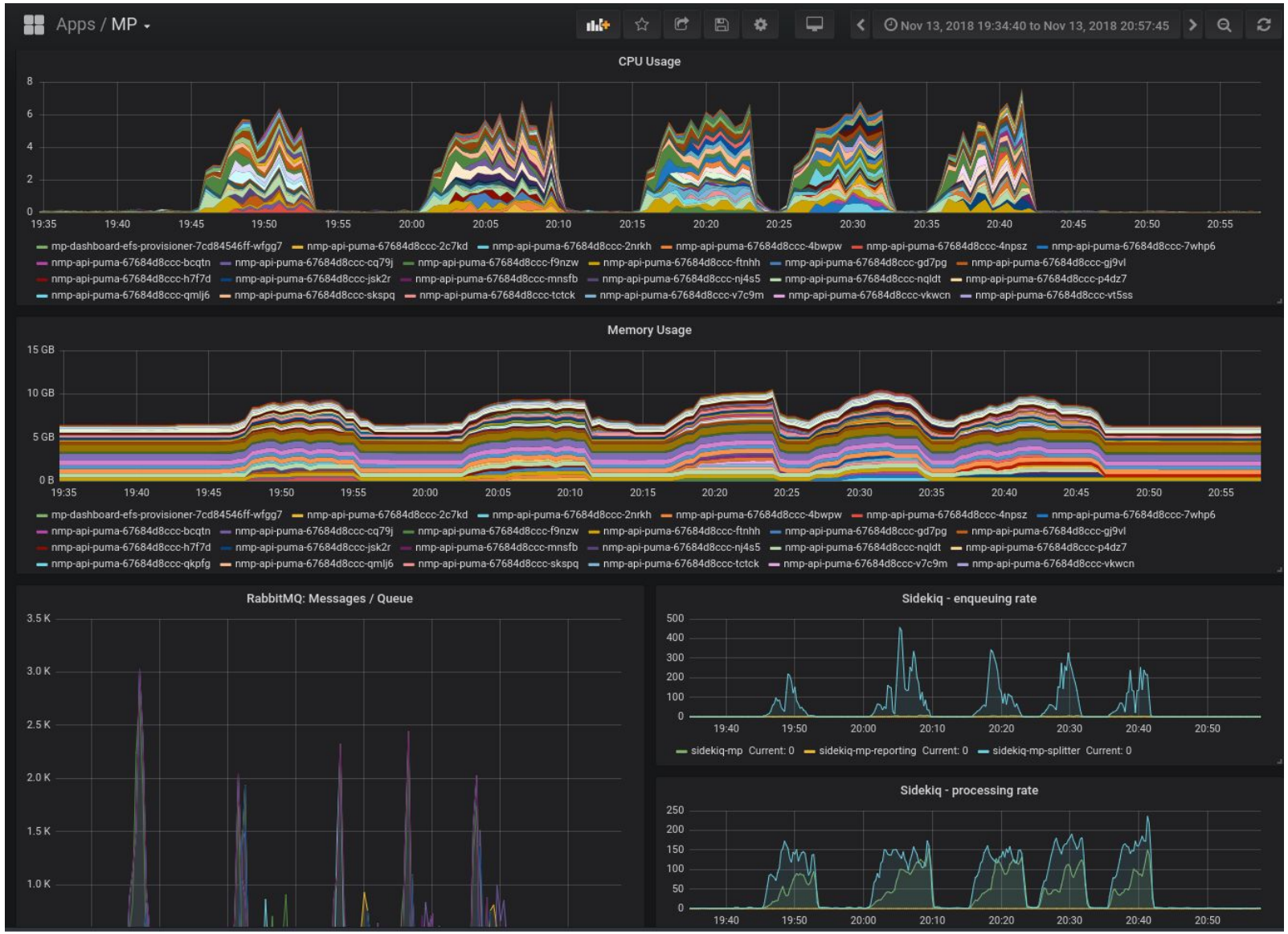
[FIRING:1] node-exporter (metrics monitoring monitoring/prometheus-operator-prometheus prometheus-operator-prometheus-node-exporter)

Load5m on node 172.20.59.52:9100 is high! Currently: 6.4
Load15m on node 172.20.95.178:9100 is high! Currently: 5.02
Load5m on node 172.20.95.178:9100 is high! Currently: 6.14

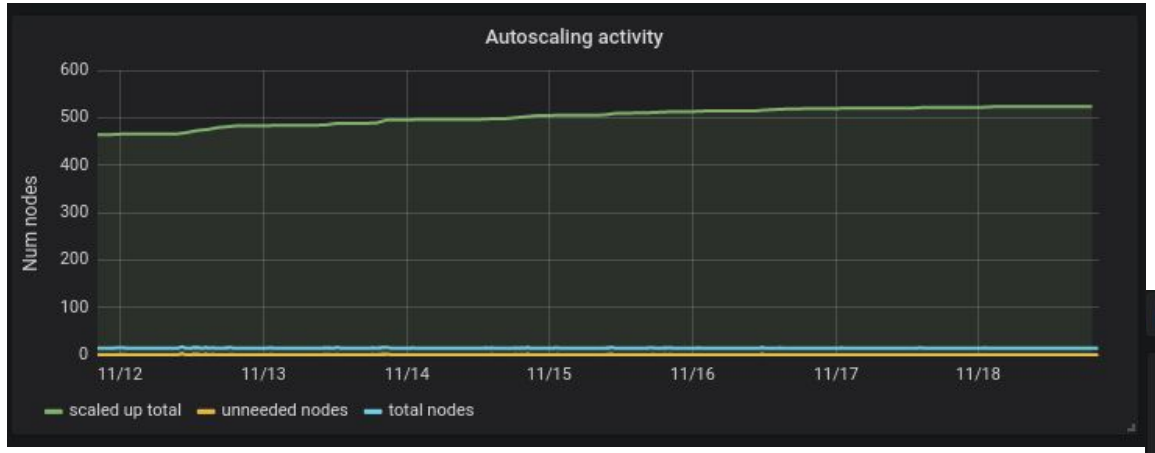
[RESOLVED] node-exporter (metrics monitoring monitoring/prometheus-operator-prometheus prometheus-operator-prometheus-node-exporter)

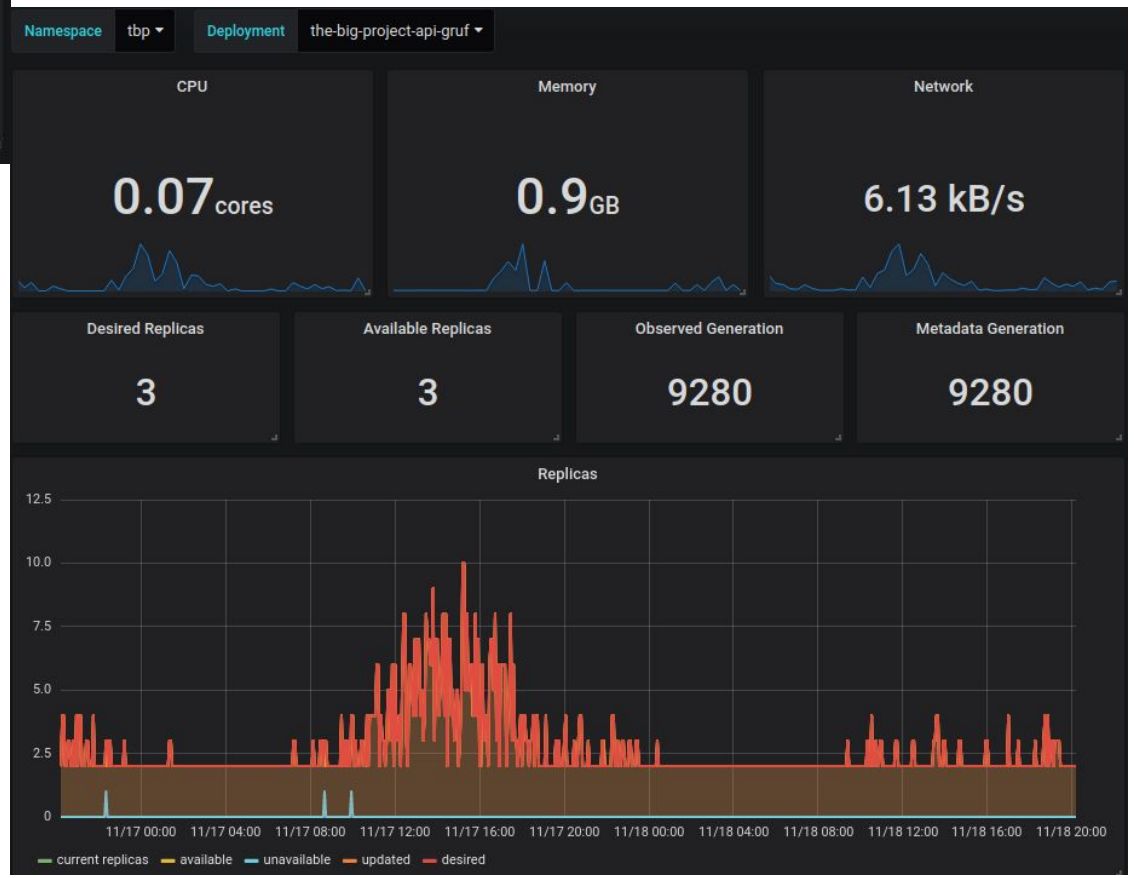
Load5m on node 172.20.59.52:9100 is high! Currently: 6.4
Load15m on node 172.20.95.178:9100 is high! Currently: 5.02
Load5m on node 172.20.95.178:9100 is high! Currently: 6.14

Monitoring - Prometheus & Grafana



Monitoring - Prometheus & Grafana





Distributed tracing - jaeger

https://github.com/salemove/jaeger-client-ruby - example ruby client

Jaeger UI | Lookup by Trace ID... | Search | Compare | Dependencies | About Jaeger

BL API (Production): api_v3/member_passwords#create_and_send_otp_by_msisdn

Search... | View Options

Trace Start: **November 18, 2018 1:37 AM** | Duration: **1.48s** | Services: **6** | Depth: **16** | Total Spans: **26**

0ms | 370.63ms | 741.27ms | 1.11s | 1.48s

Service & Operation	0ms – 1.48s
BL API (Production) api_v3/member_passwords#create_and_send_otp_by_msisdn	91.43ms
BL API (Production) /grpc_token_service.TokensService/Fetch	10.32ms
Token Service (Production) grpc_token_service.tokens_service.fetch	6.42ms
BL API (Production) /grpc_crm_service.MembersService/GetByMsisdn	13.92ms
CRM Data Service (Production) grpc_crm_service.members_service.get_by_msisdn	10.12ms
BL API (Production) OneTimePasswords::CreatorAndSenderWorker	0.09ms
BL API (Production) OneTimePasswords::CreatorAndSenderWorker	286.53ms
BL API (Production) /grpc_crm_service.MembersService/Get	18.82ms
CRM Data Service (Production) grpc_crm_service.members_service.get	5.75ms
BL API (Production) /grpc_crm_service.MembersService/Update	17.52ms
CRM Data Service (Production) grpc_crm_service.members_service.update	13.34ms
BL API (Production) /grpc_boostcom_loyalty.Messages.MessagesService/SendPasswordSms	105.82ms
TBP (Production) grpc_boostcom_loyalty.messages.messages_service.send_password_sms	101.01ms
TBP (Production) /grpc_crm_service.MembersService/Get	9.11ms
CRM Data Service (Production) grpc_crm_service.members_service.get	5.01ms
TBP (Production) Sms::ServiceSenderWorker	0.18ms
TBP (Production) Sms::ServiceSenderWorker	119ms
TBP (Production) /v2/messages	108.2ms
NMP Api (Production) messages#create	101.6ms
NMP Api (Production) MtsPersistWorker	2.47ms
NMP Api (Production) MtsPersistWorker	28.1ms
NMP Api (Production) RoutingWorker	1.12ms
NMP Gateways (Production) RoutingWorker	39.84ms
NMP Gateways (Production) ClxIntegratedSenderWorker	0.27ms
NMP Gateways (Production) ClxIntegratedSenderWorker	269.3ms
NMP Gateways (Production) /sendsms	241.76ms

Encountered problems

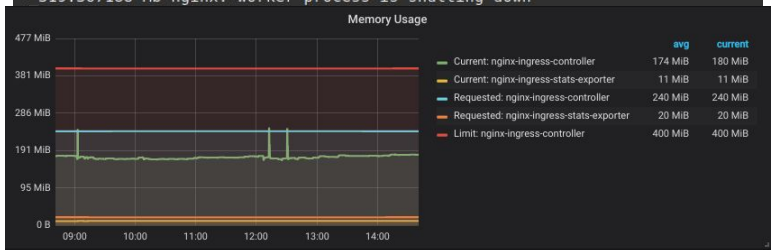
Some things not so great...

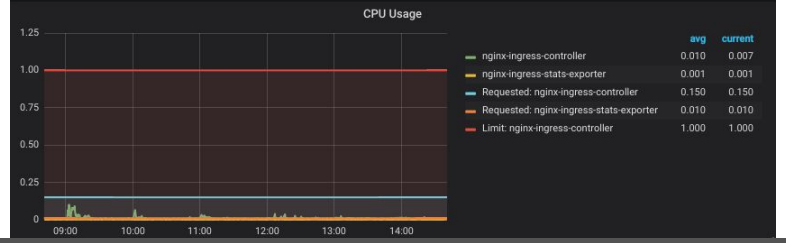
Choose “ingress” wisely

nginx-ingress

- C and Golang
- Nginx is not really cloud-first
- Nginx packed inside Golang “configuration wrapper”
- Reconfiguration == nginx reload
- Advanced stuff == manual nginx configuration snippets

```
# ps -eo size,pid,user,command | awk '{ hr=$1/1024 ; printf("%13.6f Mb ",hr) } {
     0.000000 Mb COMMAND
     0.191406 Mb /usr/bin/dumb-init /nginx-ingress-controller --default-backend-
em/nginx-ingress-controller
     0.214844 Mb awk { hr=$1/1024 ; printf("%13.6f Mb ",hr) } { for ( x=4 ; x<=N
     0.328125 Mb /bin/sh
     0.328125 Mb /bin/sh
     1.074219 Mb ps -eo size,pid,user,command
    42.843750 Mb /nginx-ingress-controller --default-backend-service=kube-syster
ntroller
    55.367188 Mb nginx: master process /usr/sbin/nginx -c /etc/nginx/nginx.conf
   318.488281 Mb nginx: worker process is shutting down
   318.488281 Mb nginx: worker process is shutting down
   318.867188 Mb nginx: worker process is shutting down
   319.367188 Mb nginx: worker process
   319.367188 Mb nginx: worker process
   319.367188 Mb nginx: worker process
   319.367188 Mb nginx: worker process is shutting down
   319.367188 Mb nginx: worker process is shutting down
```





traefik

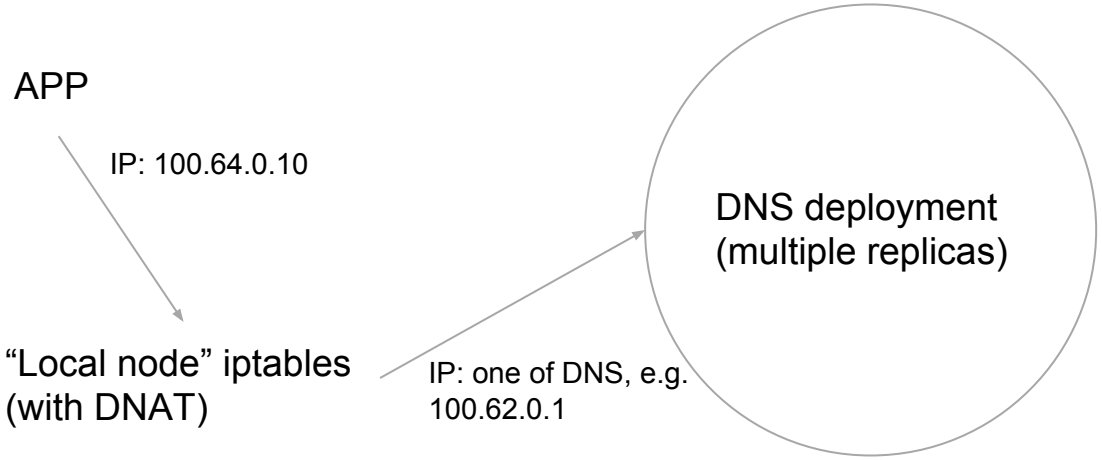
- Golang
- Designed for cloud
- Kubernetes native backend implementation
- Advanced Kubernetes annotations
- Nice UI
- Really easy

The “DNS” bug

Actually - Linux kernel’s conntrack bug, while using DNAT and two UDP packets at the same time

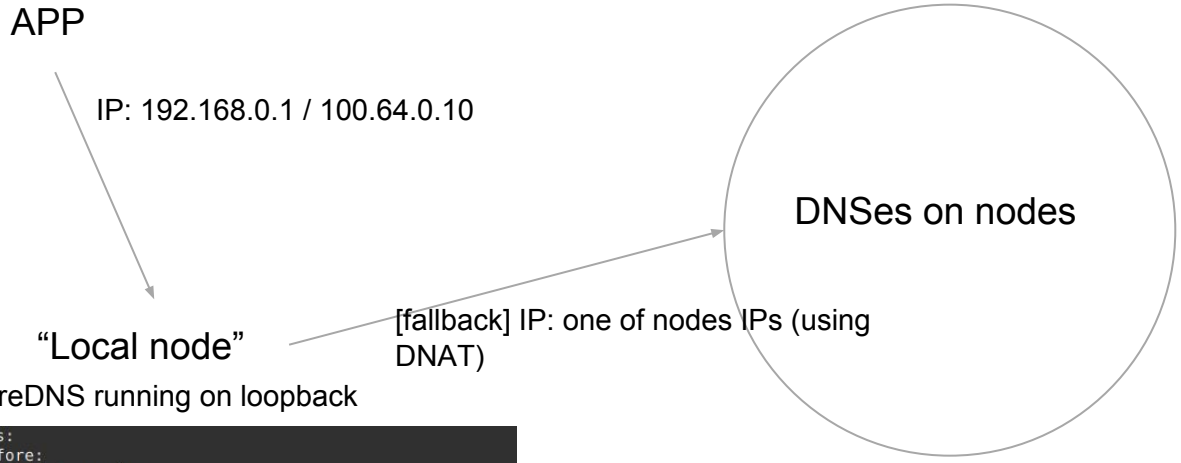
```
irb(main):002:0> 1000000.times { time = Time.now; Resolv::DNS.new.getaddress("mp.dev.boost.no"); stoptime = Time.now; p "#{stoptime-time}" if stoptime - time > 0.02 }
"0.055170046"
"5.003717779"
"5.002724427"
"0.069260178"
"0.020280993"
"5.005113935"
"0.052574319"
"0.027052438"
"0.036984"
"5.004995591"
"5.00289057"
"5.006228433"
```

Standard deployment




```
/usr/share/nginx/html # cat /etc/resolv.conf
nameserver 100.64.0.10
search tbp.svc.cluster.local svc.cluster.local cluster.local us-east-2.compute.internal
options ndots:5
```

Workaround deployment



CoreDNS running on loopback

```
hooks:
- before:
  - kubelet.service
  manifest: |
    Type=oneshot
    ExecStart=/bin/bash -c "ifconfig lo:0 192.168.0.1 up"
  roles:
  - Node
  - Master
```

```
/usr/share/nginx/html # cat /etc/resolv.conf
nameserver 192.168.0.1
nameserver 100.64.0.10
search tbp.svc.cluster.local svc.cluster.local cluster.local eu-central-1.compute.internal
options ndots:5
```

Summary

Summary

Pros

- Simple to operate and configure
- Super scalable
- Server downtimes reduced almost to 0
- Very popular with big community and mass of dedicated services
- Versatile
- Simple app deployments (compared to dedicated servers)

Cons

- You have to learn basics first (which can be a bit hard) [developers]
- You have to know something about networking. [maintaining/operating]
- Not for single app (overhead)
- Not so easy to deploy on barebone servers (with good configuration)

We’re hiring!

Know Ruby/PHP/Angular?

Send your CV to techjobs@boostcom.no

BOOSTCOM

+47 73 60 60 23

contact@boostcom.no

boostcom.no