Please

Ask questions through the app

Rate Session

Thank you!

zalando

WHY I LOVE KUBERNETES FAILURE STORIES

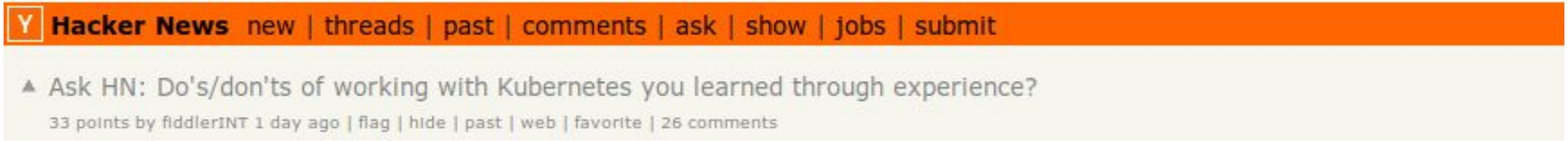
GOTO BERLIN

2019-10-24

HENNING JACOBS

@try_except_

ROLLING OUT KUBERNETES?

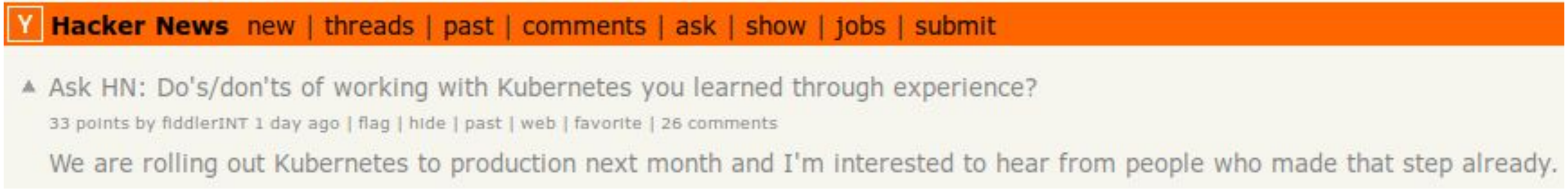

Y
Hacker News new | threads | past | comments | ask | show | jobs | submit

▲ Ask HN: Do's/don'ts of working with Kubernetes you learned through experience?

33 points by fiddlerINT 1 day ago | flag | hide | past | web | favorite | 26 comments

"We are rolling out Kubernetes to production next month and I'm interested to hear from people who made that step already."

DON'T USE IT !!!!!

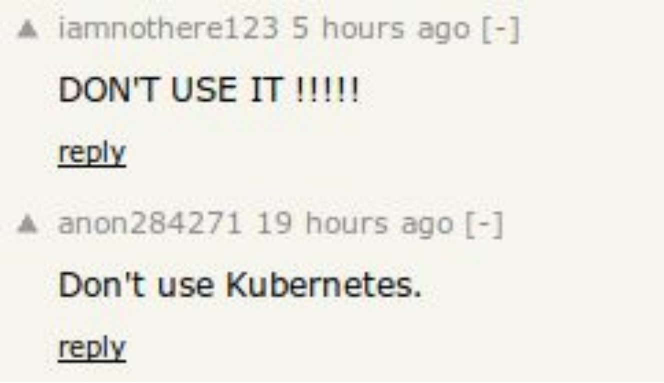


Hacker News new | threads | past | comments | ask | show | jobs | submit

Ask HN: Do's/don'ts of working with Kubernetes you learned through experience?

33 points by fiddlerINT 1 day ago | flag | hide | past | web | favorite | 26 comments

We are rolling out Kubernetes to production next month and I'm interested to hear from people who made that step already.



iamnothere123 5 hours ago [-]

DON'T USE IT !!!!!

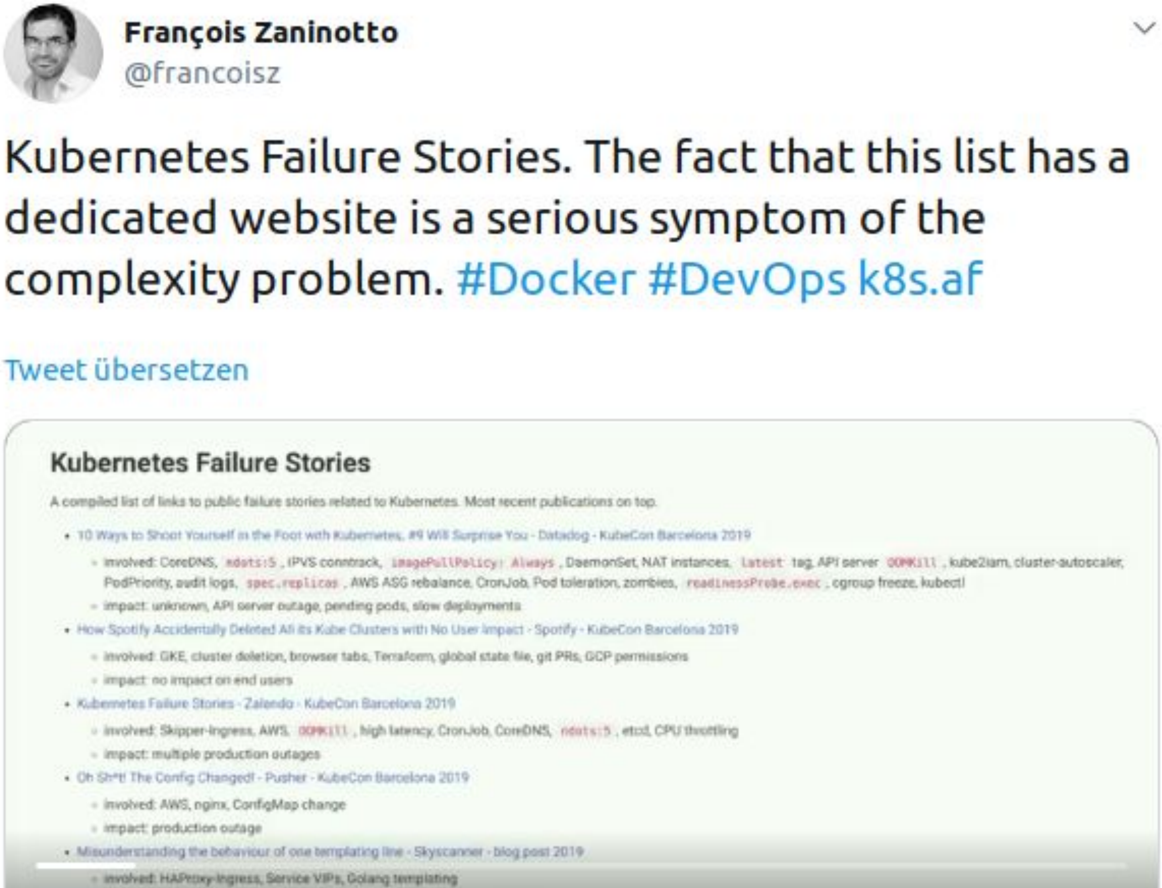
reply

anon284271 19 hours ago [-]

Don't use Kubernetes.

reply

KUBERNETES FAILURE STORIES



François Zaninotto
@francoisz

Kubernetes Failure Stories. The fact that this list has a dedicated website is a serious symptom of the complexity problem. #Docker #DevOps k8s.af

Tweet übersetzen

Kubernetes Failure Stories

A compiled list of links to public failure stories related to Kubernetes. Most recent publications on top.

- 10 Ways to Shoot Yourself in the Foot with Kubernetes, #9 Will Surprise You - Datadog - KubeCon Barcelona 2019
  - involved: CoreDNS, `ndots:5`, IPVS conntrack, `imagePullPolicy: Always`, DaemonSet, NAT instances, `latest` tag, API server `OOMKill`, kube2iam, cluster-autoscaler, PodPriority, audit logs, `spec.replicas`, AWS ASG rebalance, CronJob, Pod toleration, zombies, `readinessProbe.exec`, cgroup freeze, kubectl
  - impact: unknown, API server outage, pending pods, slow deployments
- How Spotify Accidentally Deleted All its Kube Clusters with No User Impact - Spotify - KubeCon Barcelona 2019
  - involved: GKE, cluster deletion, browser tabs, Terraform, global state file, git PRs, GCP permissions
  - impact: no impact on end users
- Kubernetes Failure Stories - Zalando - KubeCon Barcelona 2019
  - involved: Skipper-Ingress, AWS, `OOMKill`, high latency, CronJob, CoreDNS, `ndots:5`, etcd, CPU throttling
  - impact: multiple production outages
- Oh Sh*t! The Config Changed! - Pusher - KubeCon Barcelona 2019
  - involved: AWS, nginx, ConfigMap change
  - impact: production outage
- Misunderstanding the behaviour of one templating line - Skyscanner - blog post 2019
  - involved: HAProxy-Ingress, Service VIPs, Golang templating

ZALANDO AT A GLANCE

~ 5.4 billion EUR
revenue 2018

> 300 million visits per month

> 28 million
active customers

> 400,000
product choices

~ 14,000
employees in Europe

> 80%
of visits via mobile devices

> 2,000
brands

17
countries

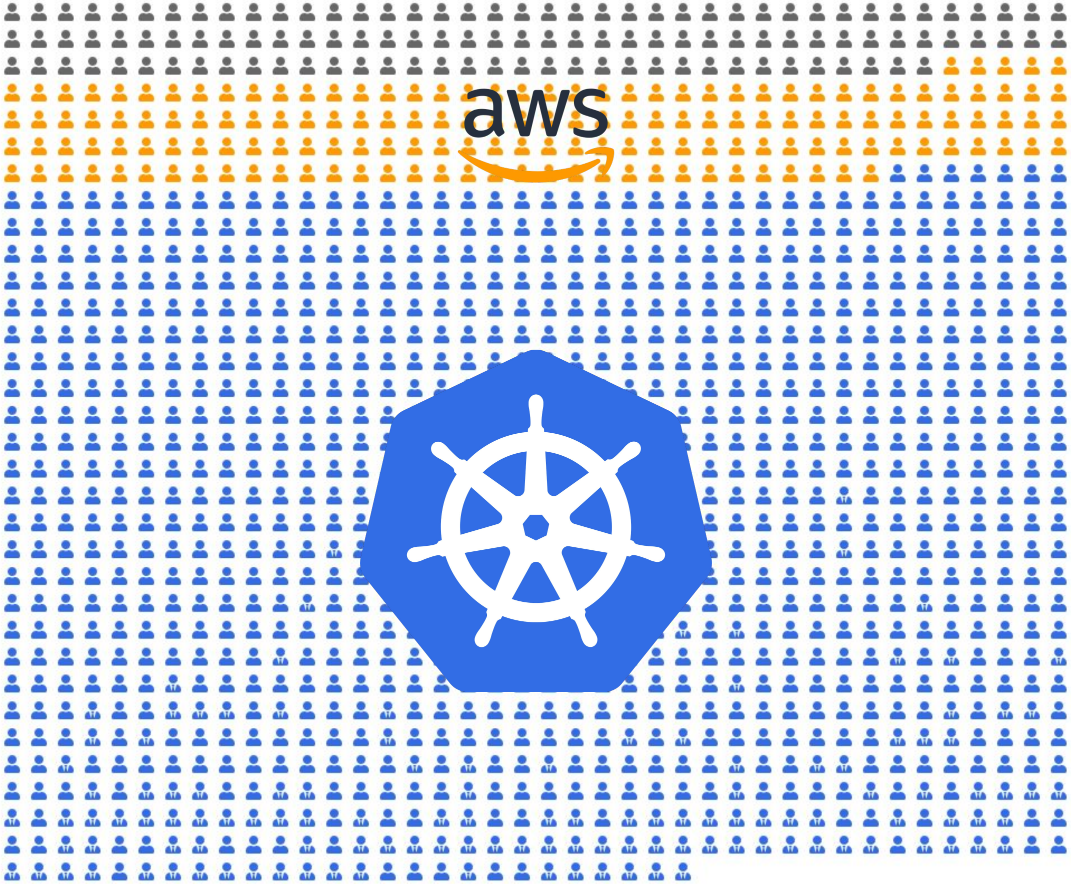
as of June 2019

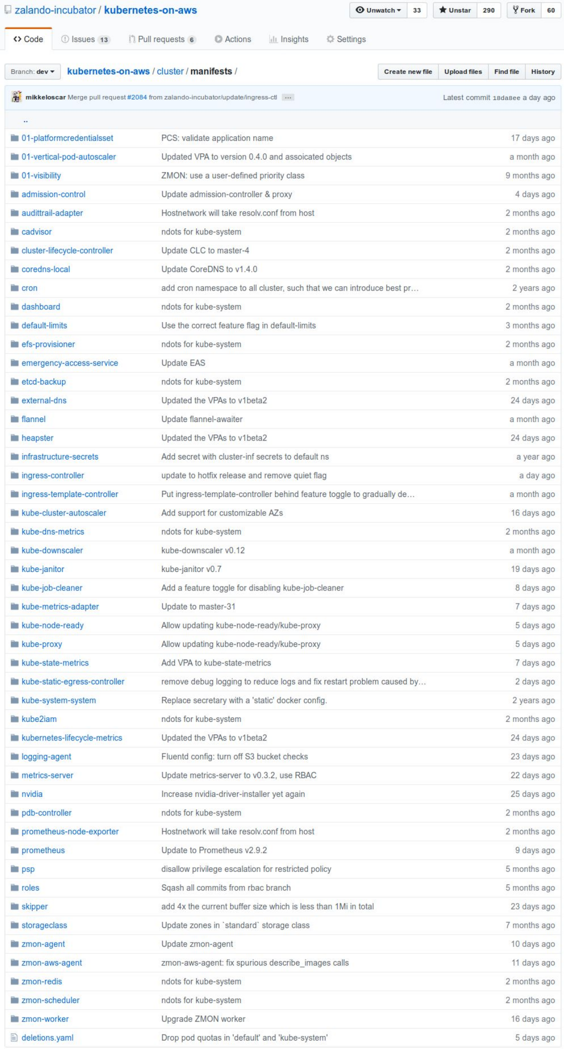
2019: SCALE

396 Accounts

140 Clusters

2019: DEVELOPERS USING KUBERNETES





kubernetes-on-aws / cluster / manifests /

mikkeloscar Merge pull request #2084 from zalando-incubator/update/ingress-ctl … Latest commit 18da8ee a day ago

..		
01-platformcredentialsset	PCS: validate application name	17 days ago
01-vertical-pod-autoscaler	Updated VPA to version 0.4.0 and assoicated objects	a month ago
01-visibility	ZMON: use a user-defined priority class	9 months ago
admission-control	Update admission-controller & proxy	4 days ago
audittrail-adapter	Hostnetwork will take resolv.conf from host	2 months ago
cadvisor	ndots for kube-system	2 months ago
cluster-lifecycle-controller	Update CLC to master-4	2 months ago
coredns-local	Update CoreDNS to v1.4.0	2 months ago
cron	add cron namespace to all cluster, such that we can introduce best pr…	2 years ago
dashboard	ndots for kube-system	2 months ago
default-limits	Use the correct feature flag in default-limits	3 months ago
efs-provisioner	ndots for kube-system	2 months ago
emergency-access-service	Update EAS	a month ago
etcd-backup	ndots for kube-system	2 months ago
external-dns	Updated the VPAs to v1beta2	24 days ago
flannel	Update flannel-awaiter	a month ago
heapster	Updated the VPAs to v1beta2	24 days ago
infrastructure-secrets	Add secret with cluster-inf secrets to default ns	a year ago
ingress-controller	update to hotfix release and remove quiet flag	a day ago
ingress-template-controller	Put ingress-template-controller behind feature toggle to gradually de…	a month ago
kube-cluster-autoscaler	Add support for customizable AZs	16 days ago
kube-dns-metrics	ndots for kube-system	2 months ago
kube-downscaler	kube-downscaler v0.12	a month ago
kube-janitor	kube-janitor v0.7	19 days ago
kube-job-cleaner	Add a feature toggle for disabling kube-job-cleaner	8 days ago
kube-metrics-adapter	Update to master-31	7 days ago
kube-node-ready	Allow updating kube-node-ready/kube-proxy	5 days ago
kube-proxy	Allow updating kube-node-ready/kube-proxy	5 days ago
kube-state-metrics	Add VPA to kube-state-metrics	7 days ago
kube-static-egress-controller	remove debug logging to reduce logs and fix restart problem caused by…	2 days ago
kube-system-system	Replace secretary with a 'static' docker config.	2 years ago
kube2iam	ndots for kube-system	2 months ago
kubernetes-lifecycle-metrics	Updated the VPAs to v1beta2	24 days ago
logging-agent	Fluentd config: turn off S3 bucket checks	23 days ago
metrics-server	Update metrics-server to v0.3.2, use RBAC	22 days ago
nvidia	Increase nvidia-driver-installer yet again	25 days ago
pdb-controller	ndots for kube-system	2 months ago
prometheus-node-exporter	Hostnetwork will take resolv.conf from host	2 months ago
prometheus	Update to Prometheus v2.9.2	9 days ago
psp	disallow privilege escalation for restricted policy	5 months ago
roles	Sqash all commits from rbac branch	5 months ago
skipper	add 4x the current buffer size which is less than 1Mi in total	23 days ago
storageclass	Update zones in `standard` storage class	7 months ago
zmon-agent	Update zmon-agent	10 days ago
zmon-aws-agent	zmon-aws-agent: fix spurious describe_images calls	11 days ago
zmon-redis	ndots for kube-system	2 months ago
zmon-scheduler	ndots for kube-system	2 months ago
zmon-worker	Upgrade ZMON worker	16 days ago
deletions.yaml	Drop pod quotas in 'default' and 'kube-system'	5 days ago

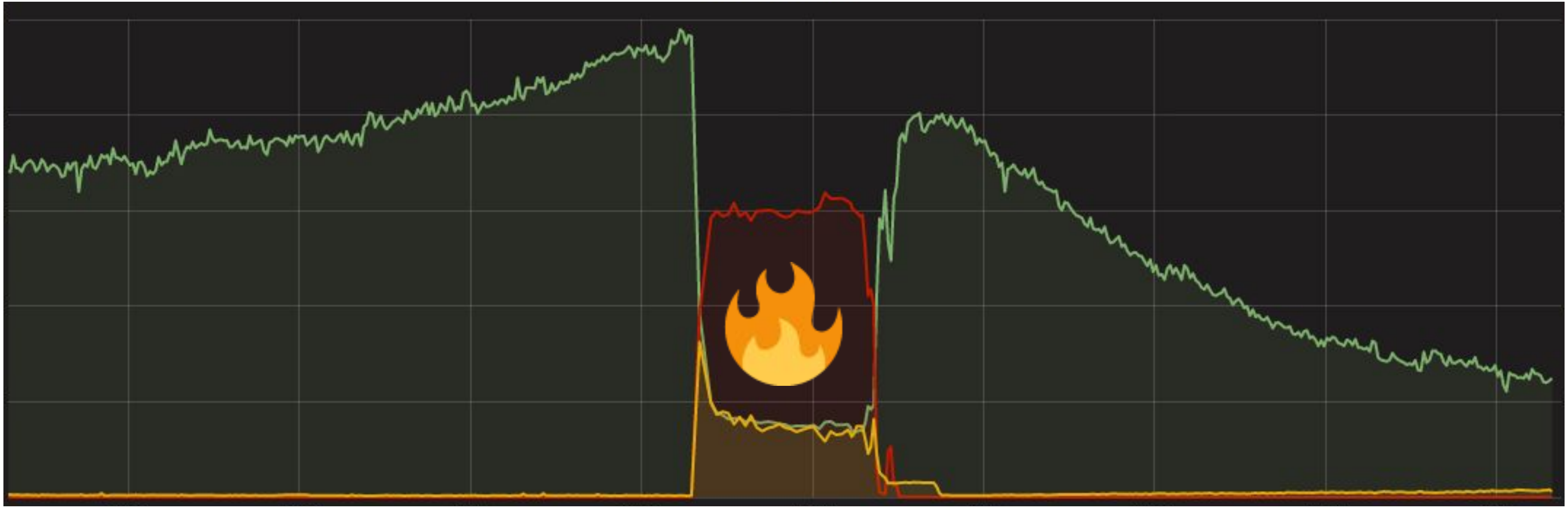
47+ cluster components

INCIDENTS ARE FINE

INCIDENT

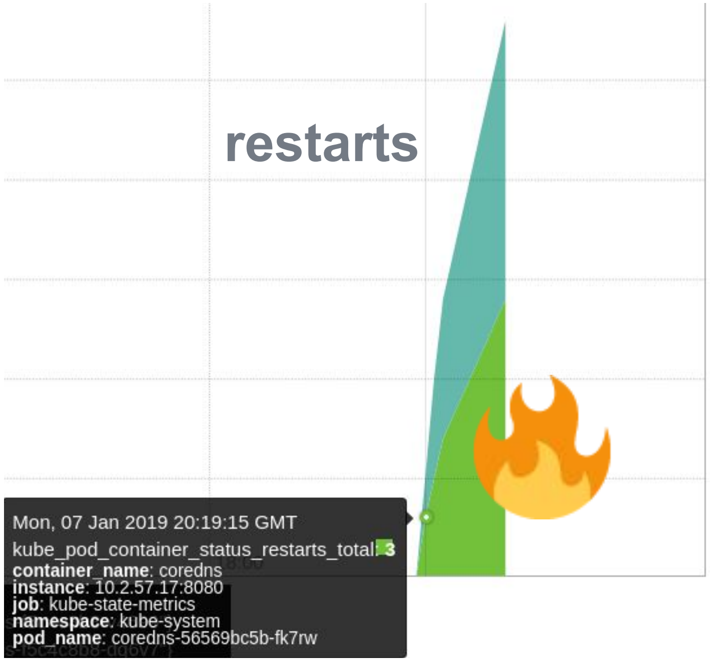
#1

INCIDENT #1: INGRESS ERRORS



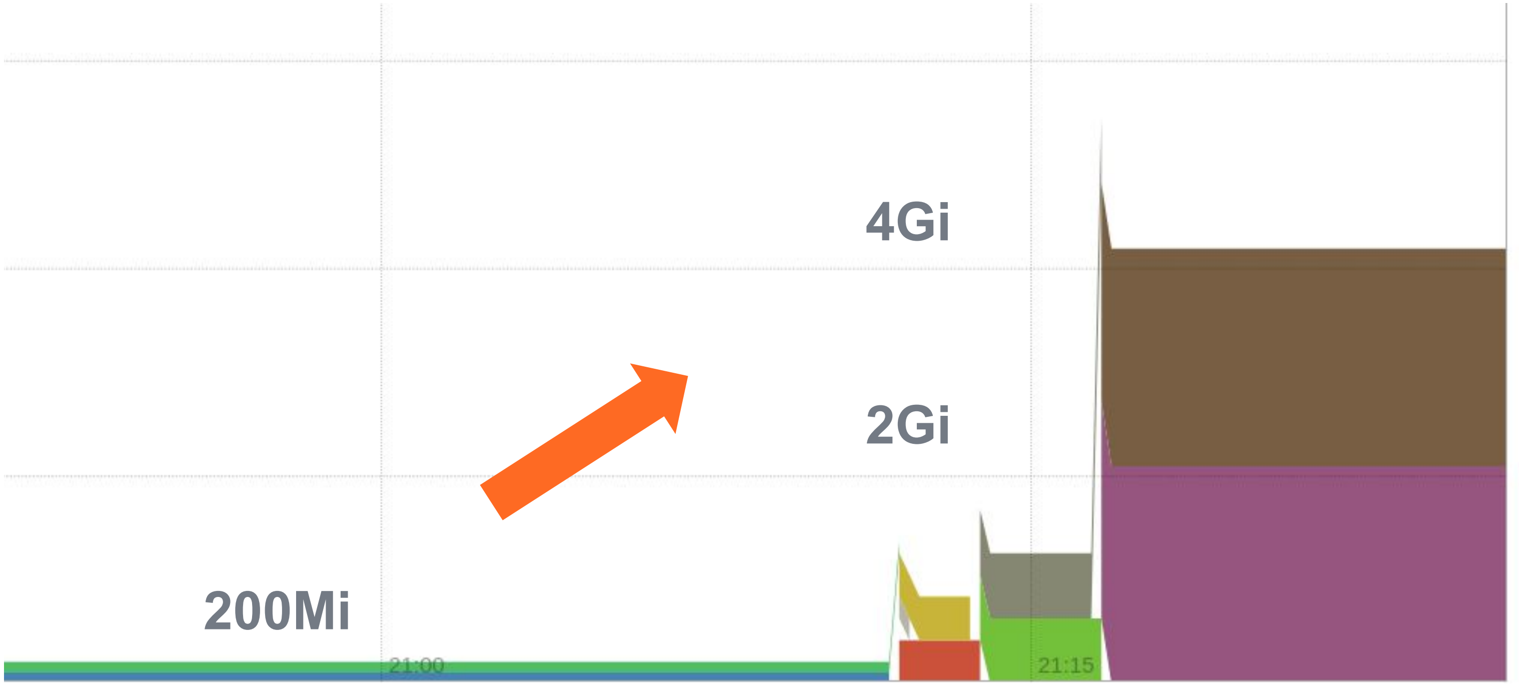
INCIDENT #1: COREDNS OOMKILL

```
coredns invoked oom-killer:
gfp_mask=0x14000c0(GFP_KERNEL),
nodemask=(null), order=0, oom_score_adj=994

Memory cgroup out of memory: Kill process 6428
(coredns) score 2050 or sacrifice child

oom_reaper: reaped process 6428 (coredns),
now anon-rss:0kB, file-rss:0kB, shmem-rss:0kB
```

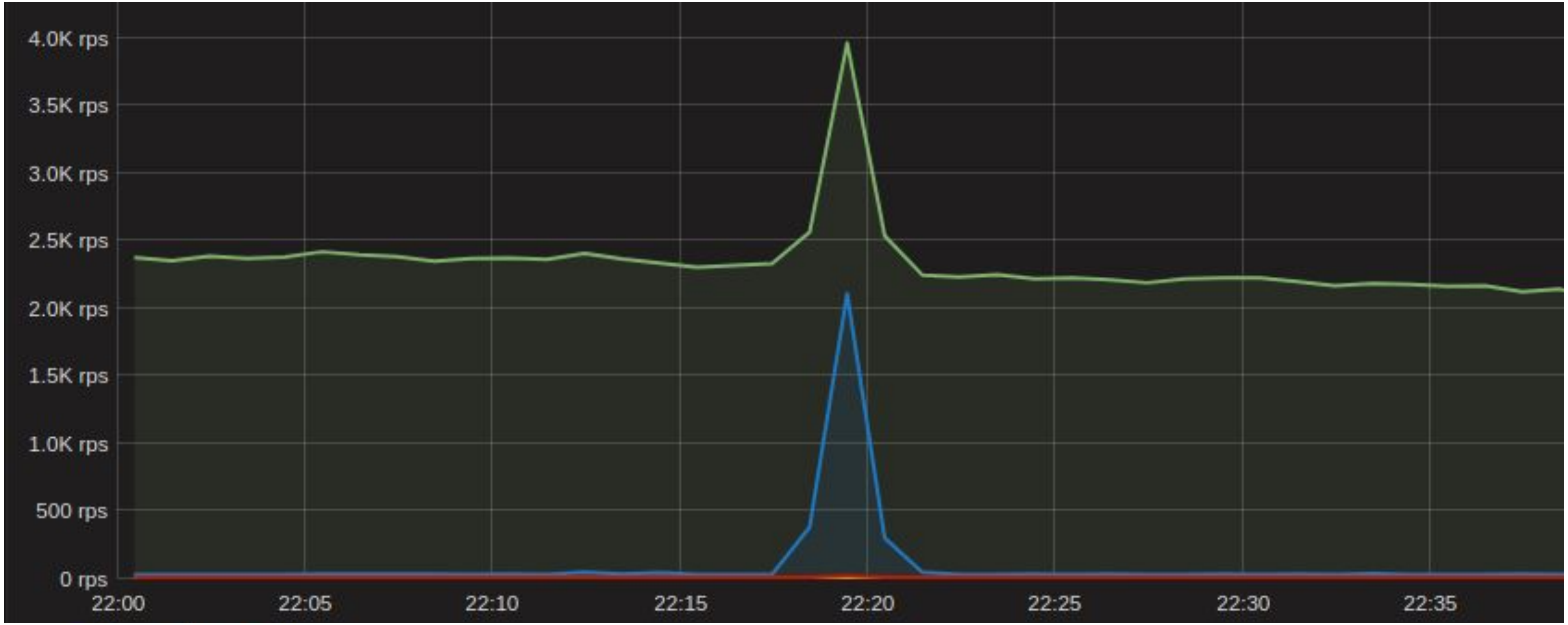


Mon, 07 Jan 2019 20:19:15 GMT
kube_pod_container_status_restarts_total: 3
container_name: coredns
instance: 10.2.57.17:8080
job: kube-state-metrics
namespace: kube-system
pod_name: coredns-56569bc5b-fk7rw

STOP THE BLEEDING: INCREASE MEMORY LIMIT

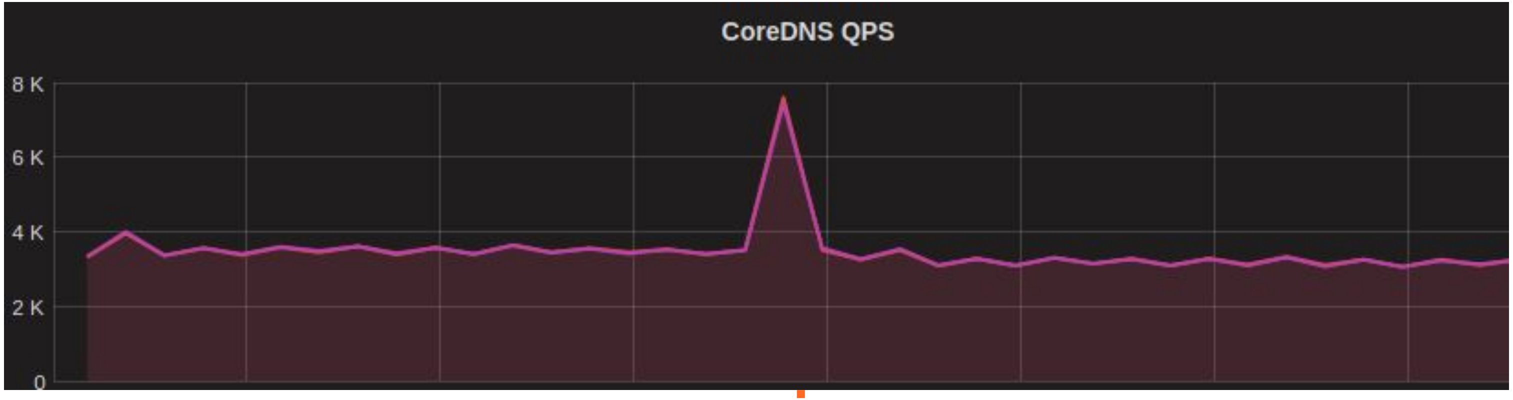


SPIKE IN HTTP REQUESTS

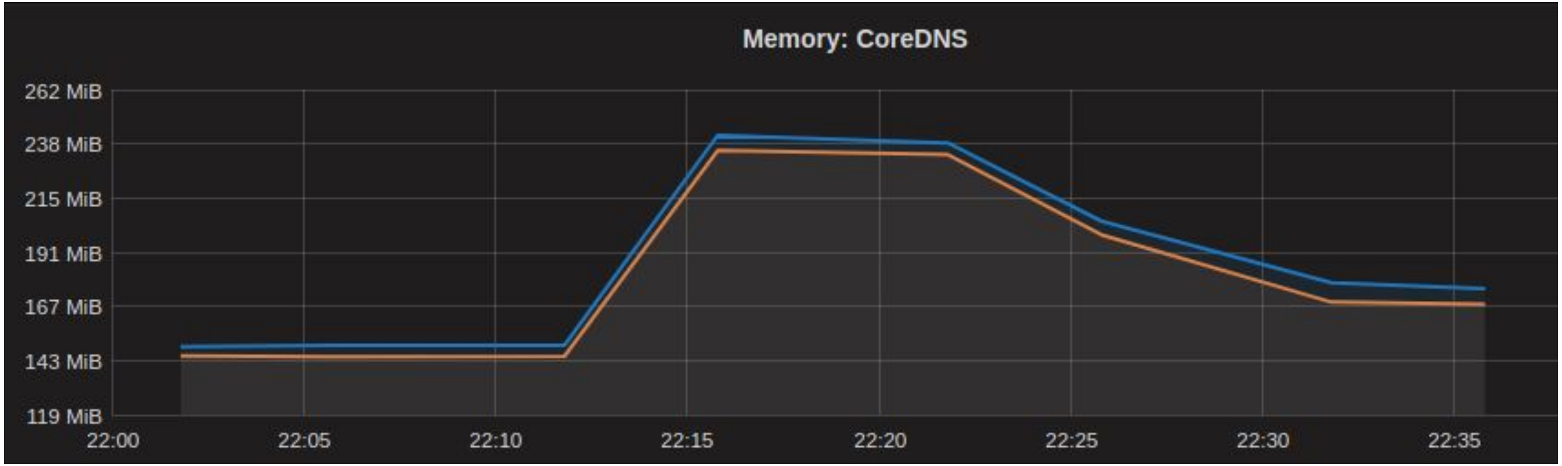


zalando

SPIKE IN DNS QUERIES



INCREASE IN MEMORY USAGE



INCIDENT #1: CONTRIBUTING FACTORS

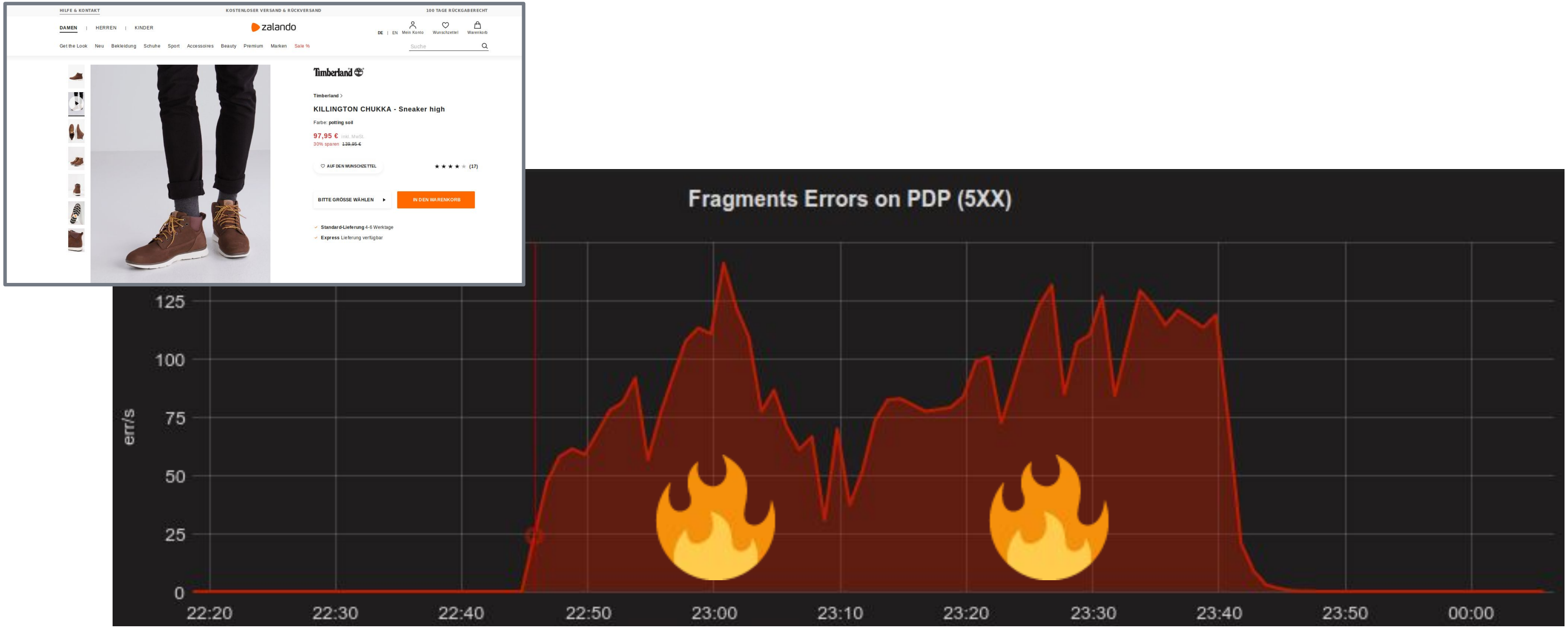
- HTTP retries
- No DNS caching
- Kubernetes ndots:5 problem
- Short maximum lifetime of HTTP connections
- Fixed memory limit for CoreDNS
- Monitoring affected by DNS outage

github.com/zalando-incubator/kubernetes-on-aws/blob/dev/docs/postmortems/jan-2019-dns-outage.md

INCIDENT

#2

INCIDENT #2: CUSTOMER IMPACT



HILFE & KONTAKT KOSTENLOSER VERSAND & RÜCKVERSAND 100 TAGE RÜCKGABERECHT

DAMEN | HERREN | KINDER zalando DE | EN Mein Konto Wunschzettel Warenkorb

Get the Look Neu Bekleidung Schuhe Sport Accessoires Beauty Premium Marken Sale % Suche

Timberland

Timberland >

KILLINGTON CHUKKA - Sneaker high

Farbe: **potting soil**

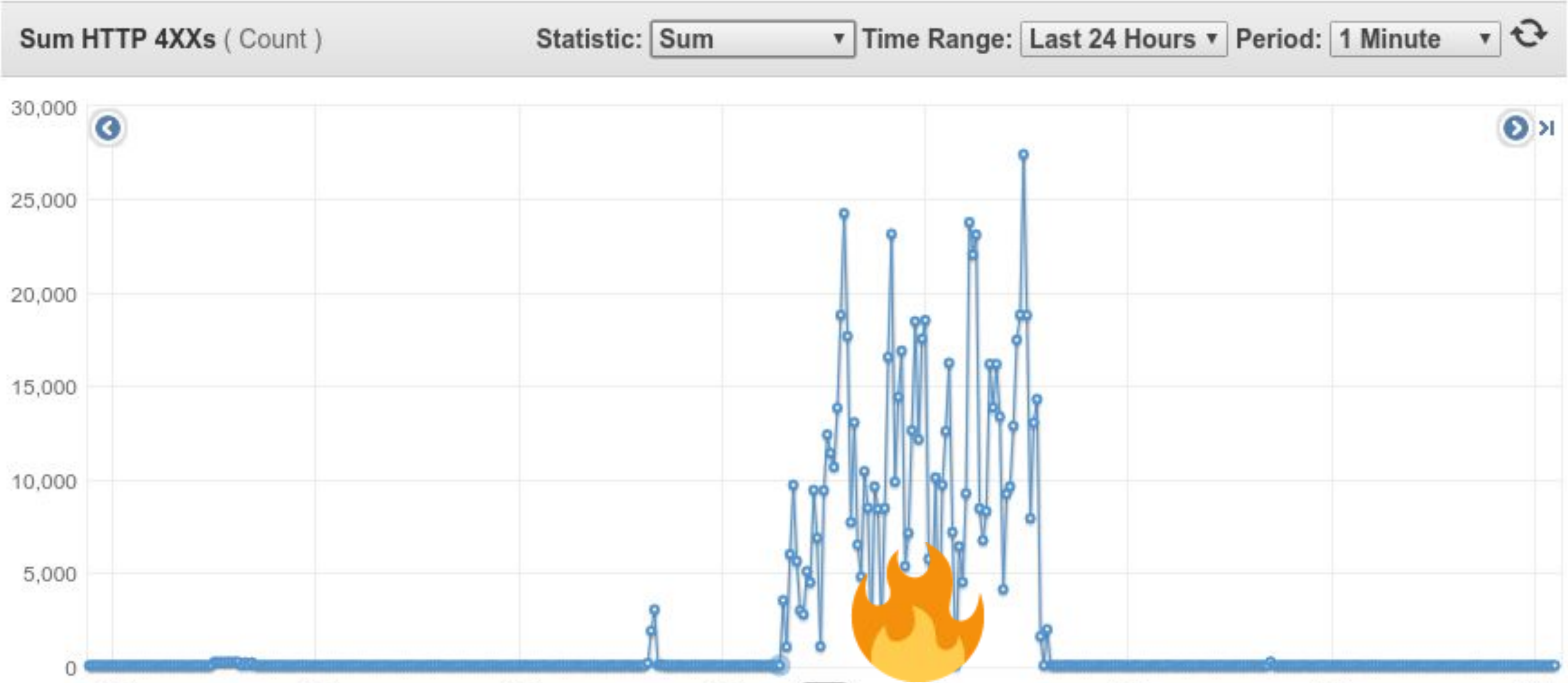
97,95 € inkl. MwSt.
30% sparen ~~139,95 €~~

AUF DEN WUNSCHZETTEL ★★★★☆ (17)

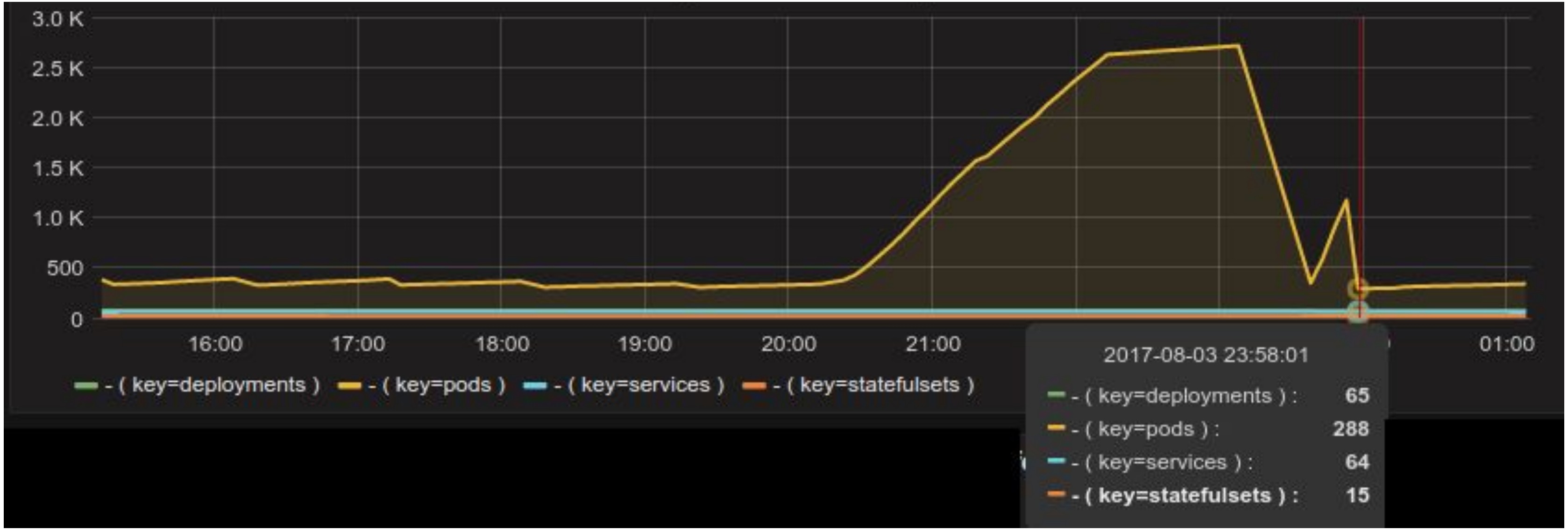
BITTE GRÖSSE WÄHLEN ▸ IN DEN WARENKORB

✓ **Standard-Lieferung** 4-6 Werktage
✓ **Express** Lieferung verfügbar

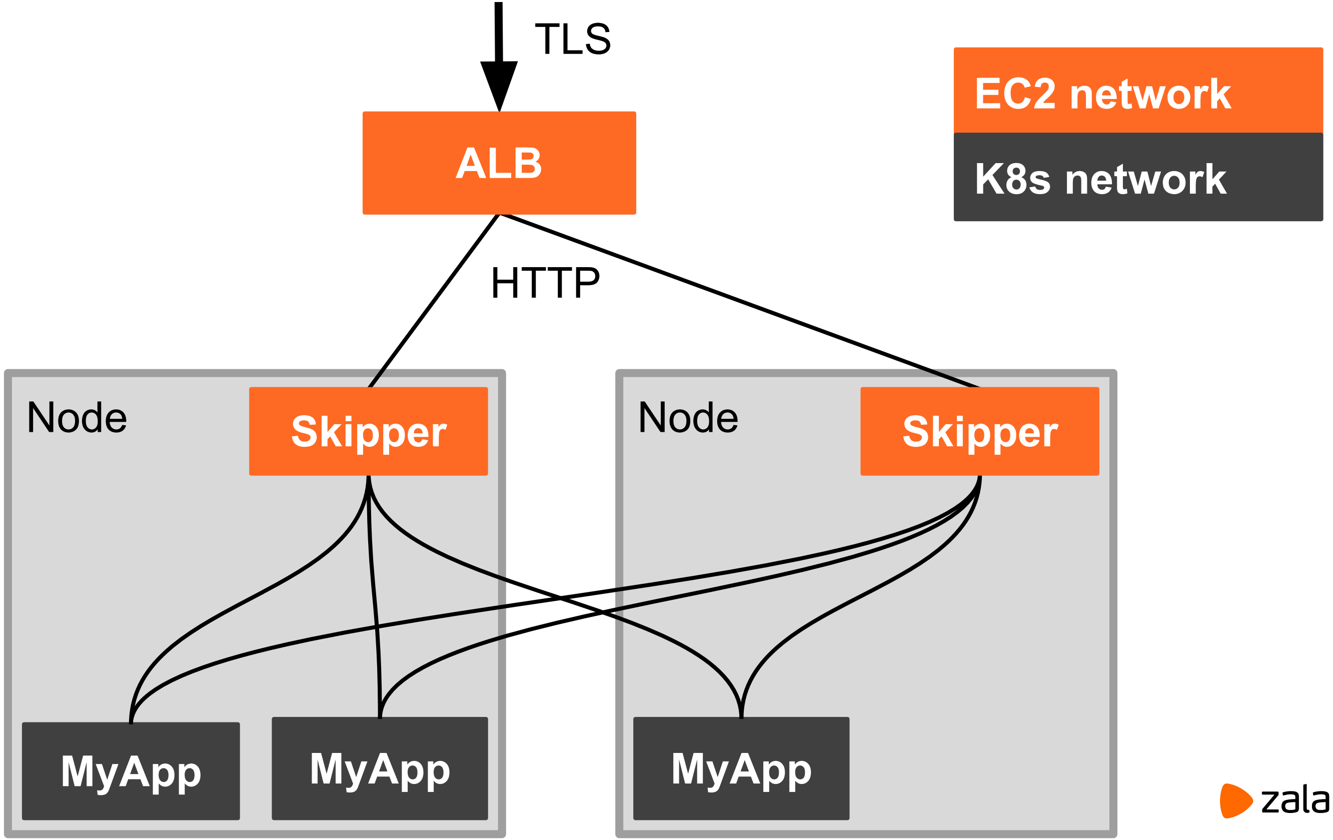
INCIDENT #2: IAM RETURNING 404



INCIDENT #2: NUMBER OF PODS



LIFE OF A REQUEST (INGRESS)



TLS

ALB

HTTP

EC2 network

K8s network

Node

Skipper

MyApp

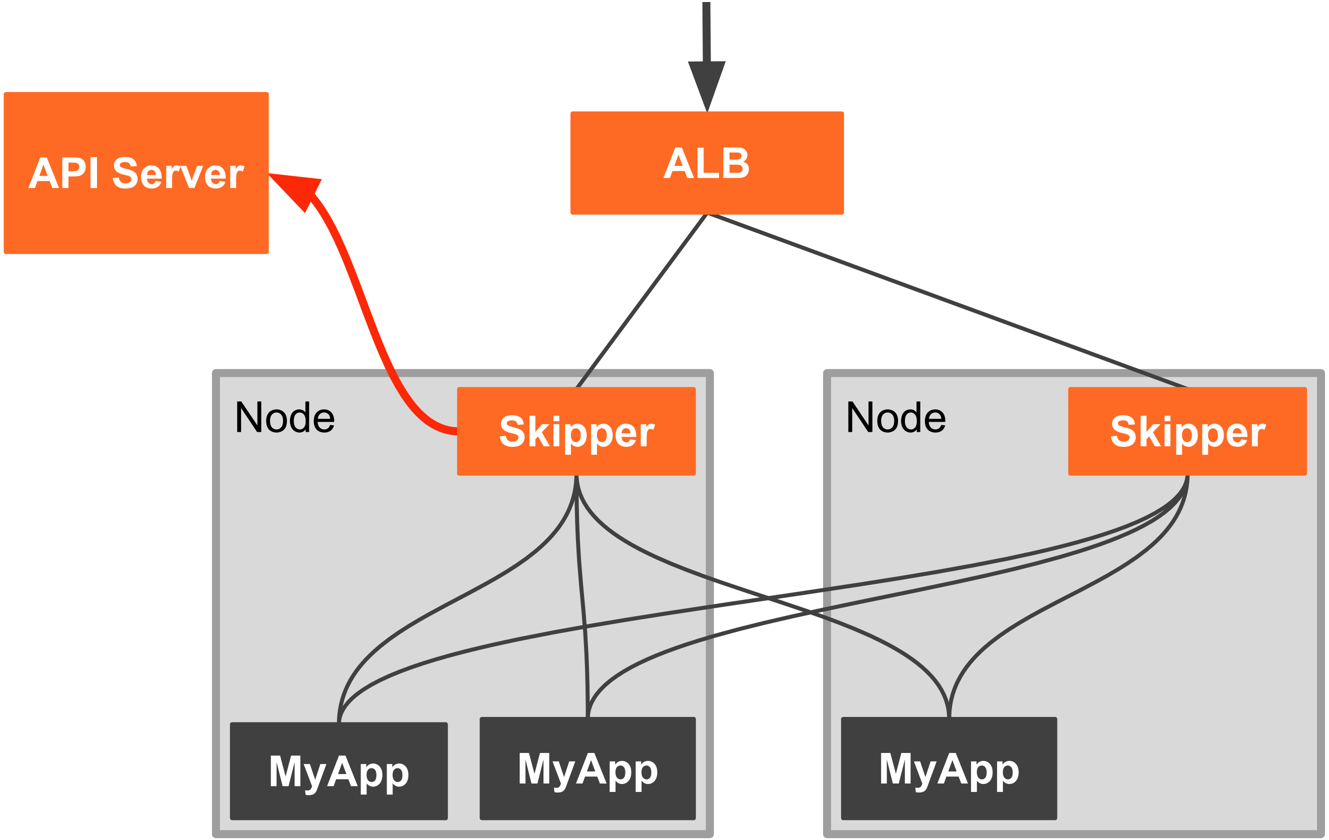
MyApp

Node

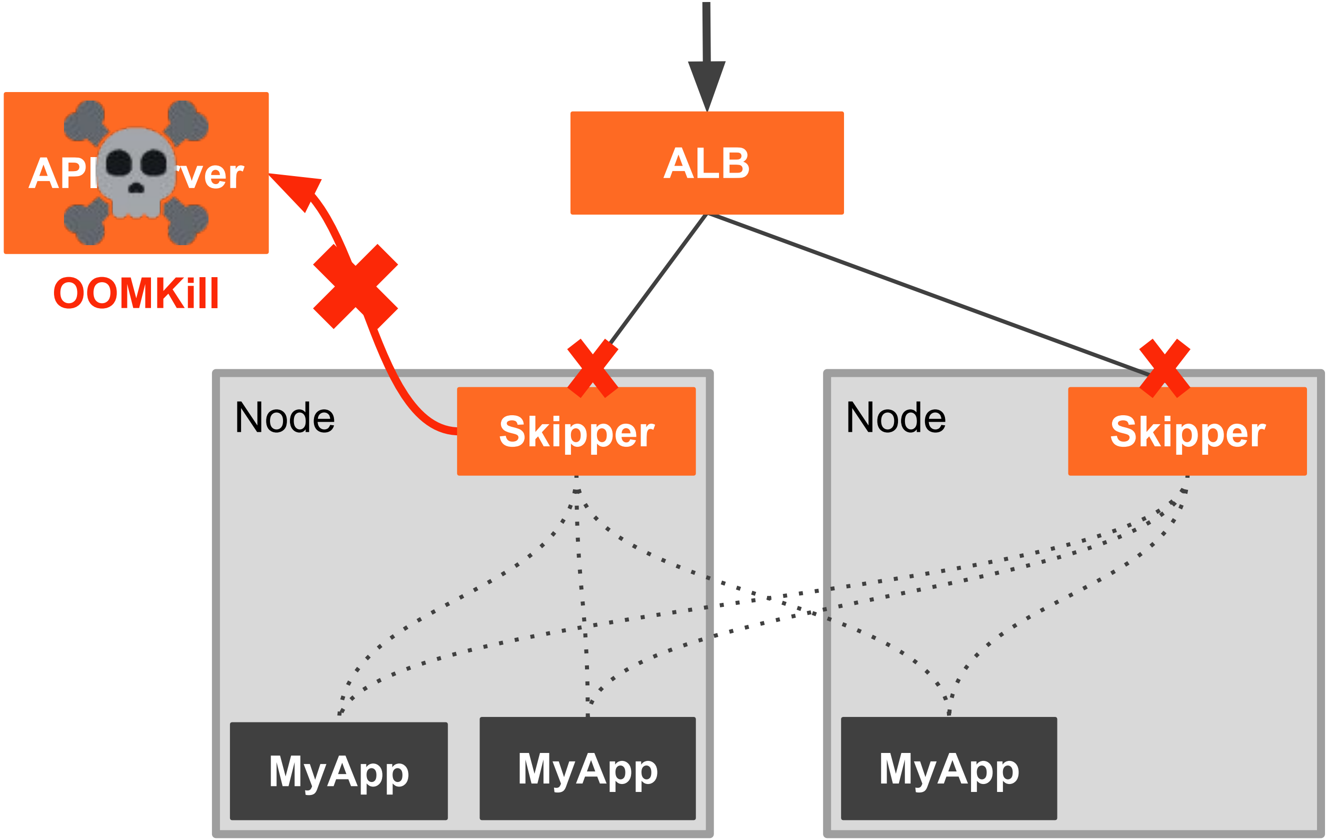
Skipper

MyApp

ROUTES FROM API SERVER



API SERVER DOWN



INCIDENT #2: INNOCENT MANIFEST

```
apiVersion: batch/v2alpha1
kind: CronJob
metadata:
  name: "foobar"
spec:
  schedule: "*/15 9-19 * * Mon-Fri"
  jobTemplate:
    spec:
      template:
        spec:
        restartPolicy: Never
        concurrencyPolicy: Forbid
        successfulJobsHistoryLimit: 1
        failedJobsHistoryLimit: 1
        containers:
            ...
```

INCIDENT #2: FIXED CRON JOB

```
apiVersion: batch/v2alpha1
kind: CronJob
metadata:
  name: "foobar"
spec:
  schedule: "7 8-18 * * Mon-Fri"
  concurrencyPolicy: Forbid
  successfulJobsHistoryLimit: 1
  failedJobsHistoryLimit: 1
  jobTemplate:
    spec:
      activeDeadlineSeconds: 120
      template:
        spec:
          restartPolicy: Never
          containers:
```

INCIDENT #2: LESSONS LEARNED

- Fix Ingress to stay “healthy” during API server problems
- Fix Ingress to retain last known set of routes
- Use quota for number of pods

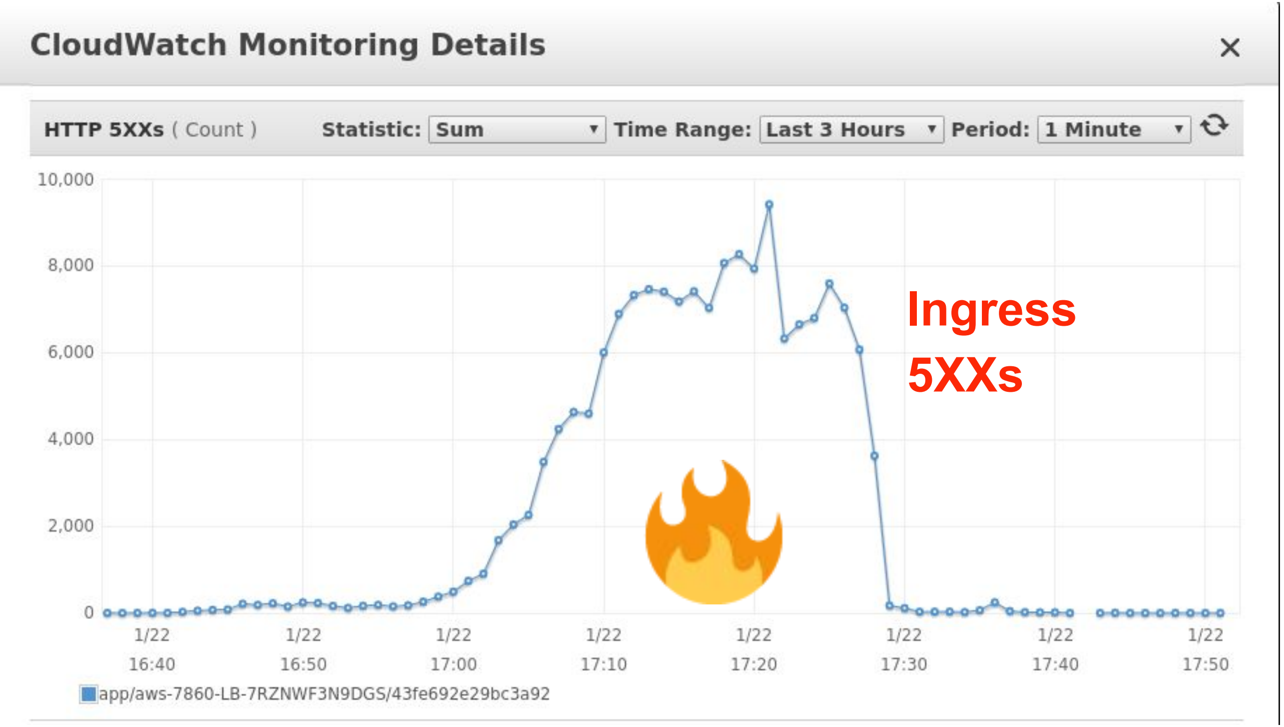
```
apiVersion: v1
kind: ResourceQuota
metadata:
  name: compute-resources
spec:
  hard:
    pods: "1500"
```

NOTE: we dropped quotas recently
github.com/zalando-incubator/kubernetes-on-aws/pull/2059

INCIDENT

#3

INCIDENT #3: IMPACT

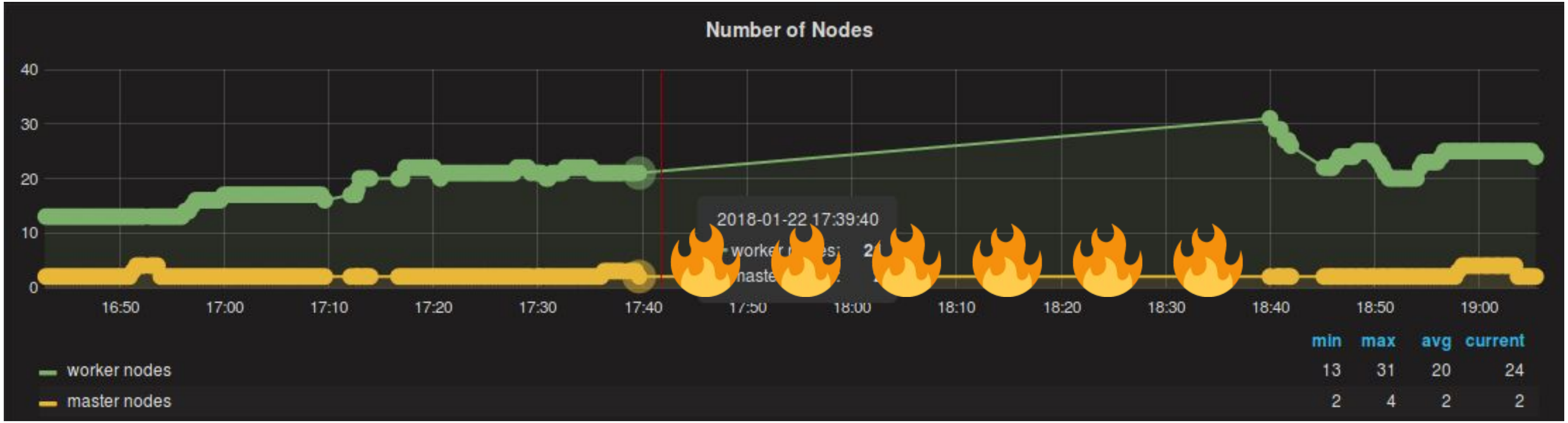


CloudWatch Monitoring Details

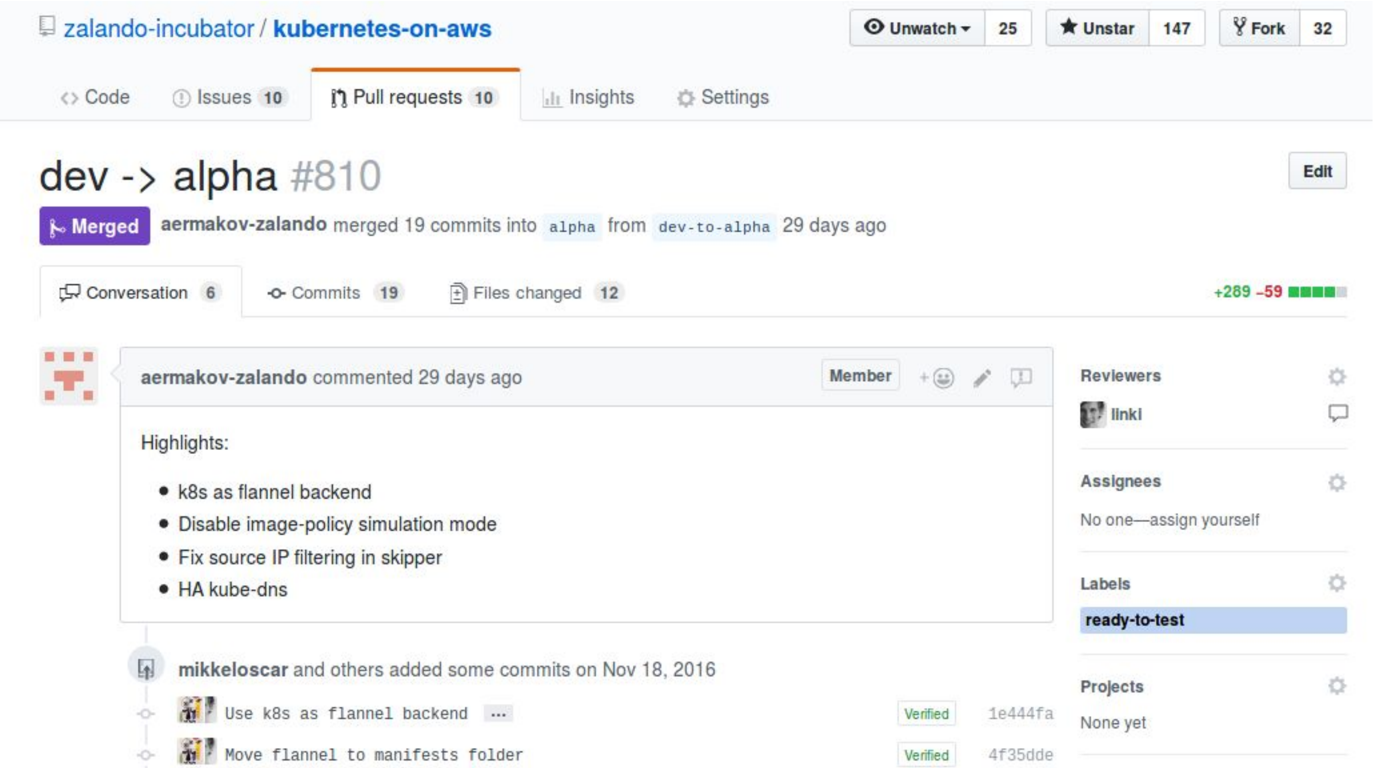
HTTP 5XXs (Count) Statistic: Sum Time Range: Last 3 Hours Period: 1 Minute

Ingress
5XXs

INCIDENT #3: CLUSTER DOWN?



INCIDENT #3: THE TRIGGER



dev -> alpha #810

Merged **aermakov-zalando** merged 19 commits into `alpha` from `dev-to-alpha` 29 days ago

aermakov-zalando commented 29 days ago

Highlights:

- k8s as flannel backend
- Disable image-policy simulation mode
- Fix source IP filtering in skipper
- HA kube-dns

mikkeloscar and others added some commits on Nov 18, 2016

```
Use k8s as flannel backend      Verified  1e444fa
Move flannel to manifests folder  Verified  4f35dde
```

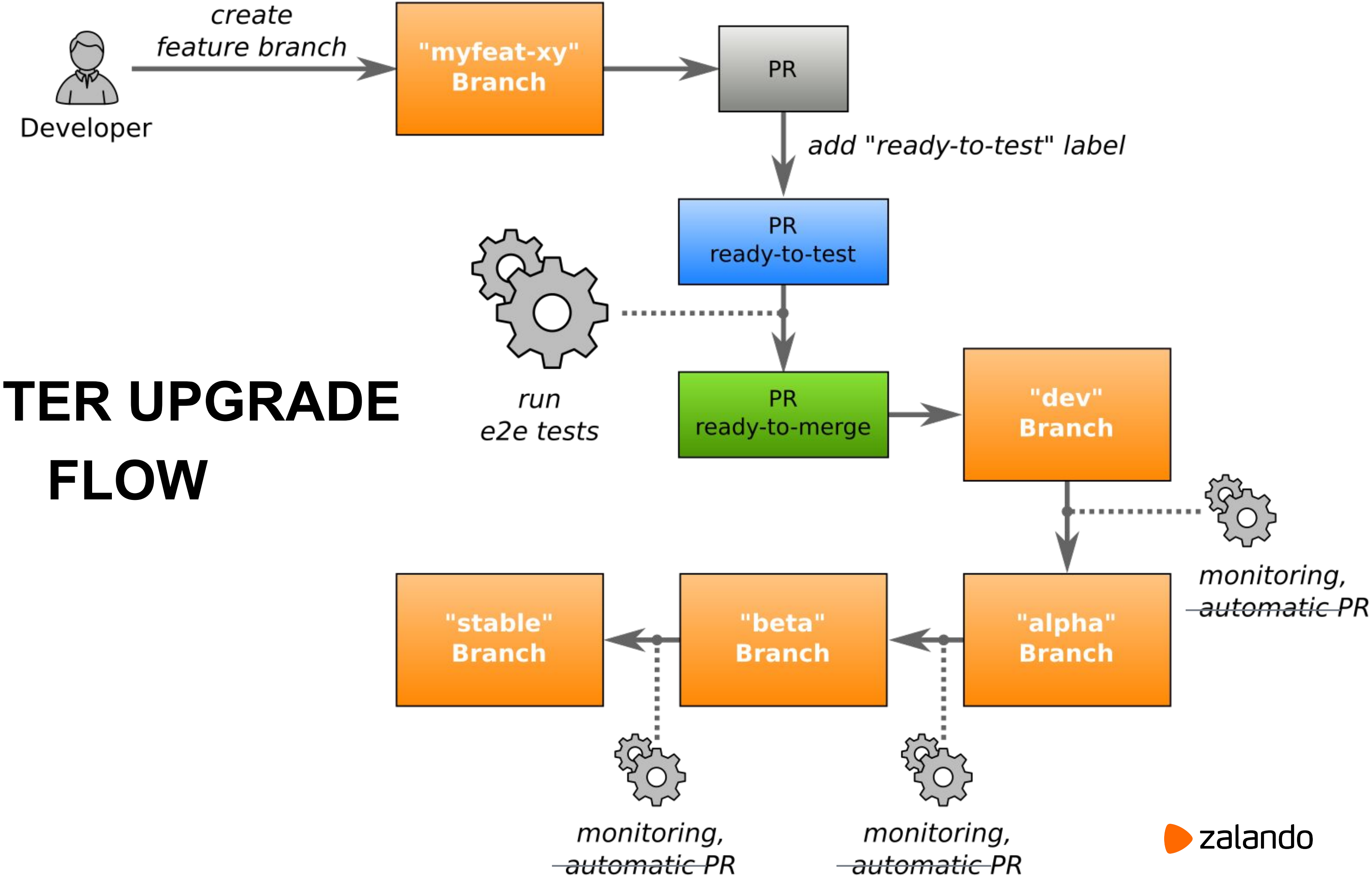
What We Believe

VOL. 1 ISSUE 6

Human Error is NEVER the Root Cause

https://www.outcome-eng.com/human-error-never-root-cause/

CLUSTER UPGRADE FLOW



Developer

create feature branch

"myfeat-xy" Branch

PR

add "ready-to-test" label

PR ready-to-test

run e2e tests

PR ready-to-merge

"dev" Branch

monitoring,
~~automatic PR~~

"alpha" Branch

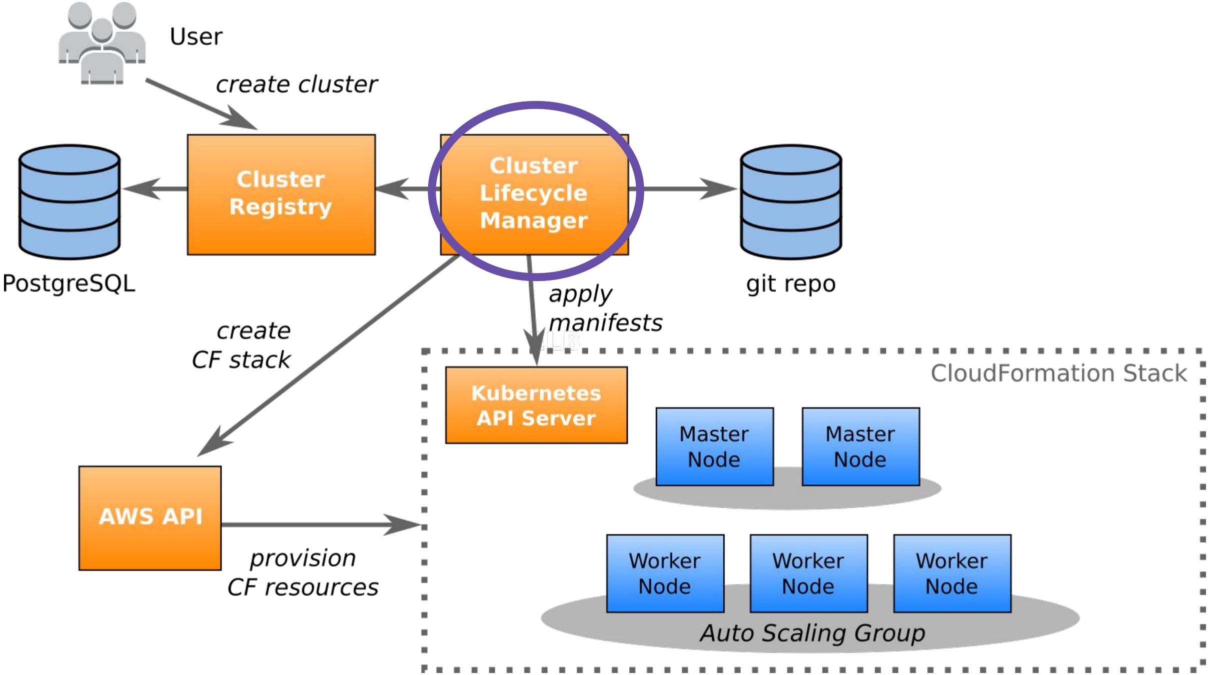
monitoring,
~~automatic PR~~

"beta" Branch

monitoring,
~~automatic PR~~

"stable" Branch

CLUSTER LIFECYCLE MANAGER (CLM)



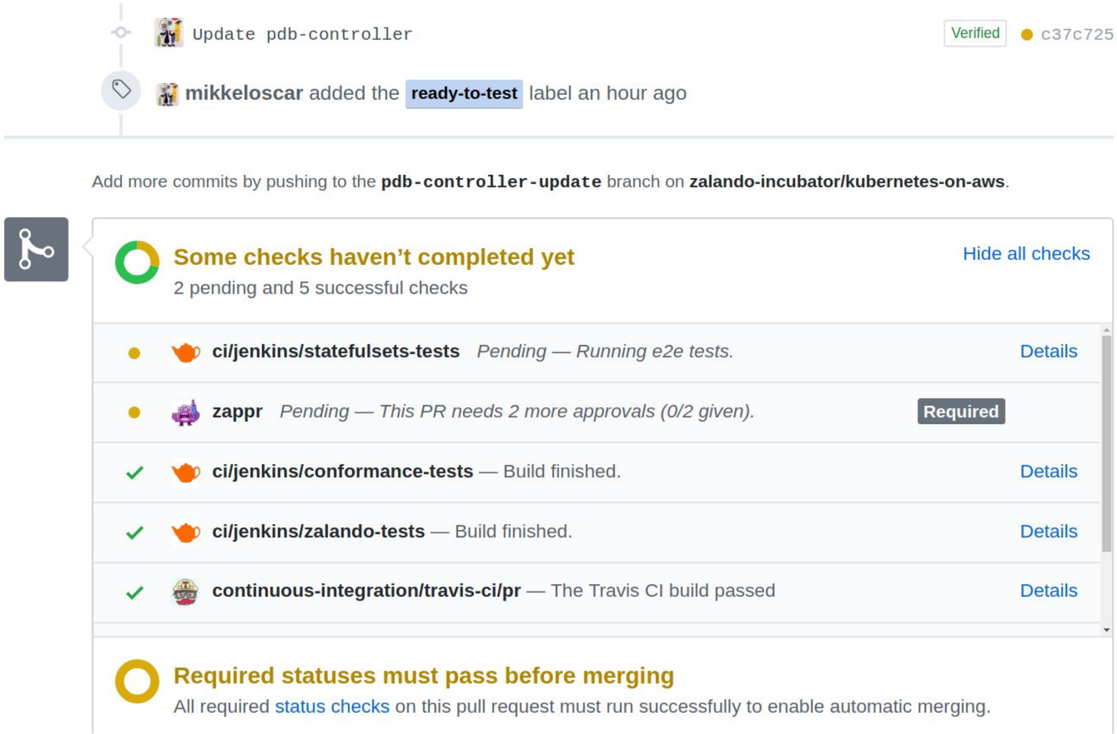
github.com/zalando-incubator/cluster-lifecycle-manager

CLUSTER CHANNELS

Channel	**Description**	**Clusters**
`dev`	Development and playground clusters.	**3**
`alpha`	Main infrastructure clusters (**important to us**).	**2**
`beta`	Product clusters for the rest of the organization (non-prod).	**60+**
`stable`	Product clusters for the rest of the organization (prod).	**60+**

github.com/zalando-incubator/kubernetes-on-aws

E2E TESTS ON EVERY PR



Update pdb-controller

mikkeloscar added the **ready-to-test** label an hour ago

Add more commits by pushing to the **`pdb-controller-update`** branch on **zalando-incubator/kubernetes-on-aws**.

Some checks haven't completed yet

2 pending and 5 successful checks

- ● **ci/jenkins/statefulsets-tests** *Pending — Running e2e tests.* Details
- ● **zappr** *Pending — This PR needs 2 more approvals (0/2 given).* **Required**
- ✓ **ci/jenkins/conformance-tests** — Build finished. Details
- ✓ **ci/jenkins/zalando-tests** — Build finished. Details
- ✓ **continuous-integration/travis-ci/pr** — The Travis CI build passed Details

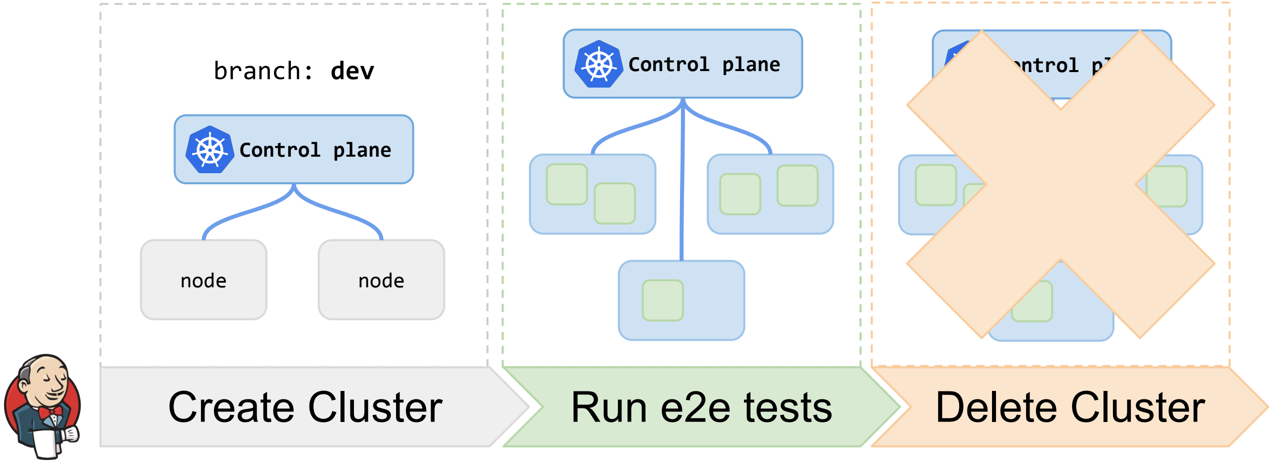
Required statuses must pass before merging

All required status checks on this pull request must run successfully to enable automatic merging.

github.com/zalando-incubator/kubernetes-on-aws

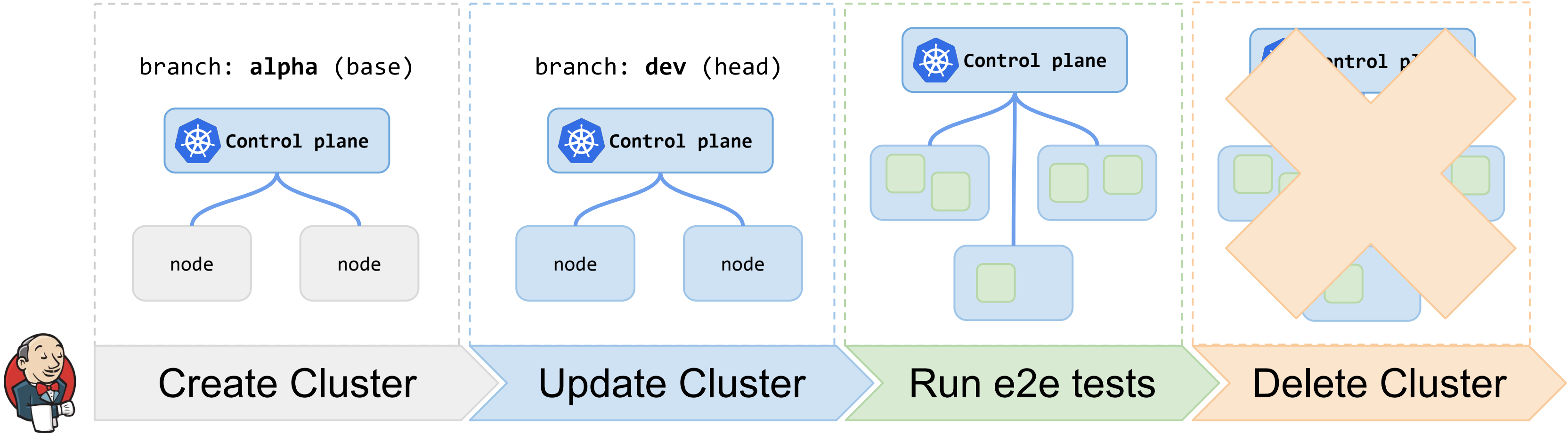
RUNNING E2E TESTS (BEFORE)

Testing **dev** to **alpha** upgrade



RUNNING E2E TESTS (NOW)

`Testing` **`dev`** `to` **`alpha`** `upgrade`



zalando

INCIDENT #3: LESSONS LEARNED

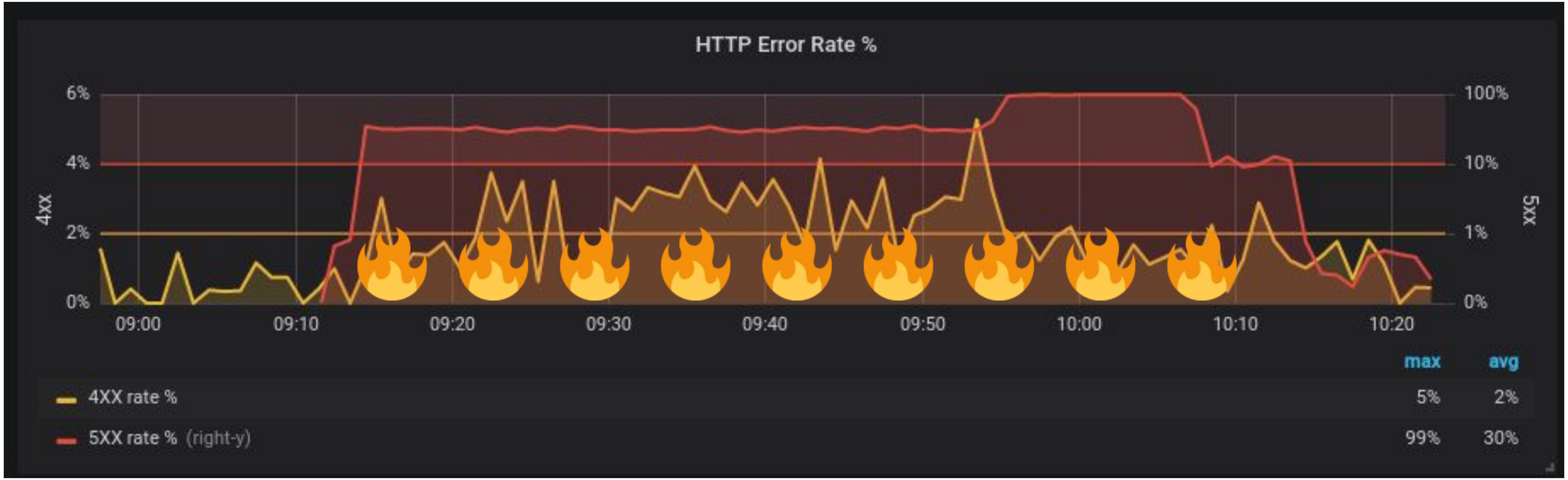
- Automated e2e tests are pretty good, but not enough
- Test the diff/migration automatically
  - Bootstrap new cluster with previous configuration
  - Apply new configuration
  - Run end-to-end & conformance tests

srcco.de/posts/how-zalando-manages-140-kubernetes-clusters.html

INCIDENT

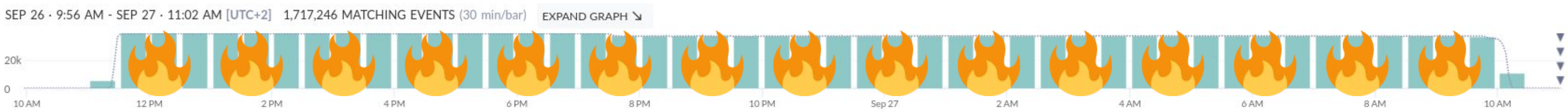
#4

INCIDENT #4: IMPACT



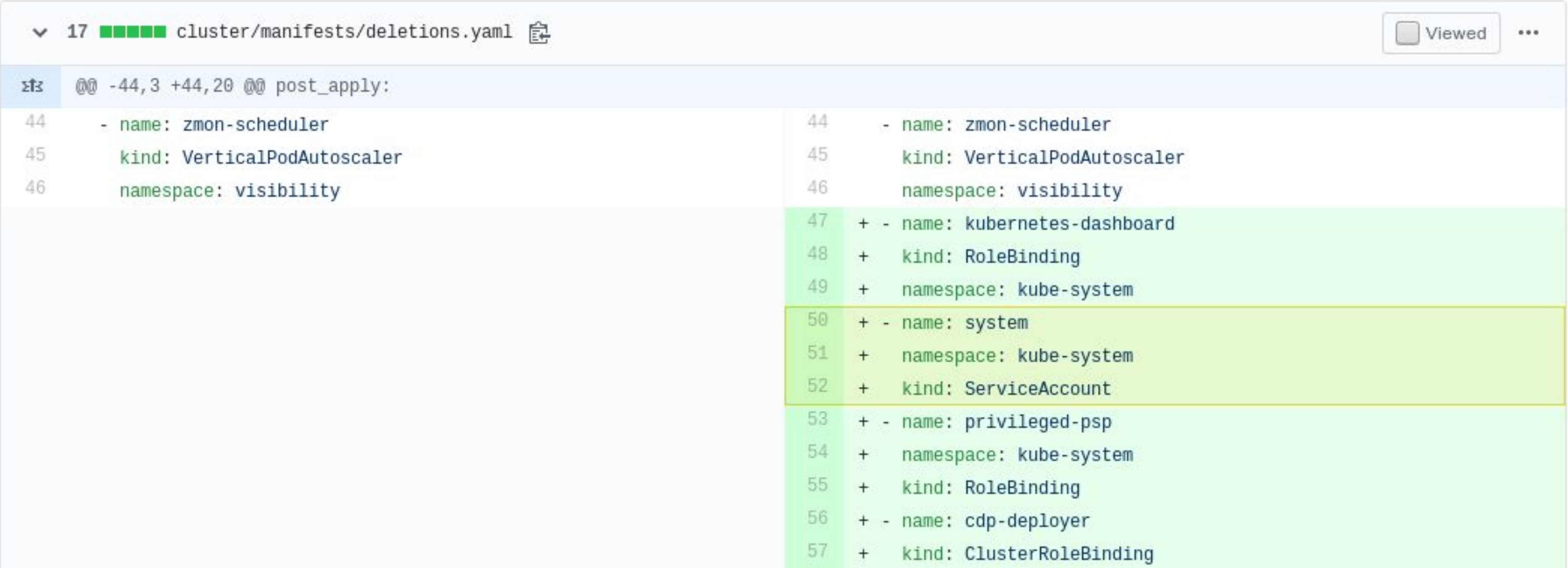
INCIDENT #4: FLANNEL ERRORS

Failed to list *v1.Node: Unauthorized



SEP 26 · 9:56 AM - SEP 27 · 11:02 AM [UTC+2] 1,717,246 MATCHING EVENTS (30 min/bar)

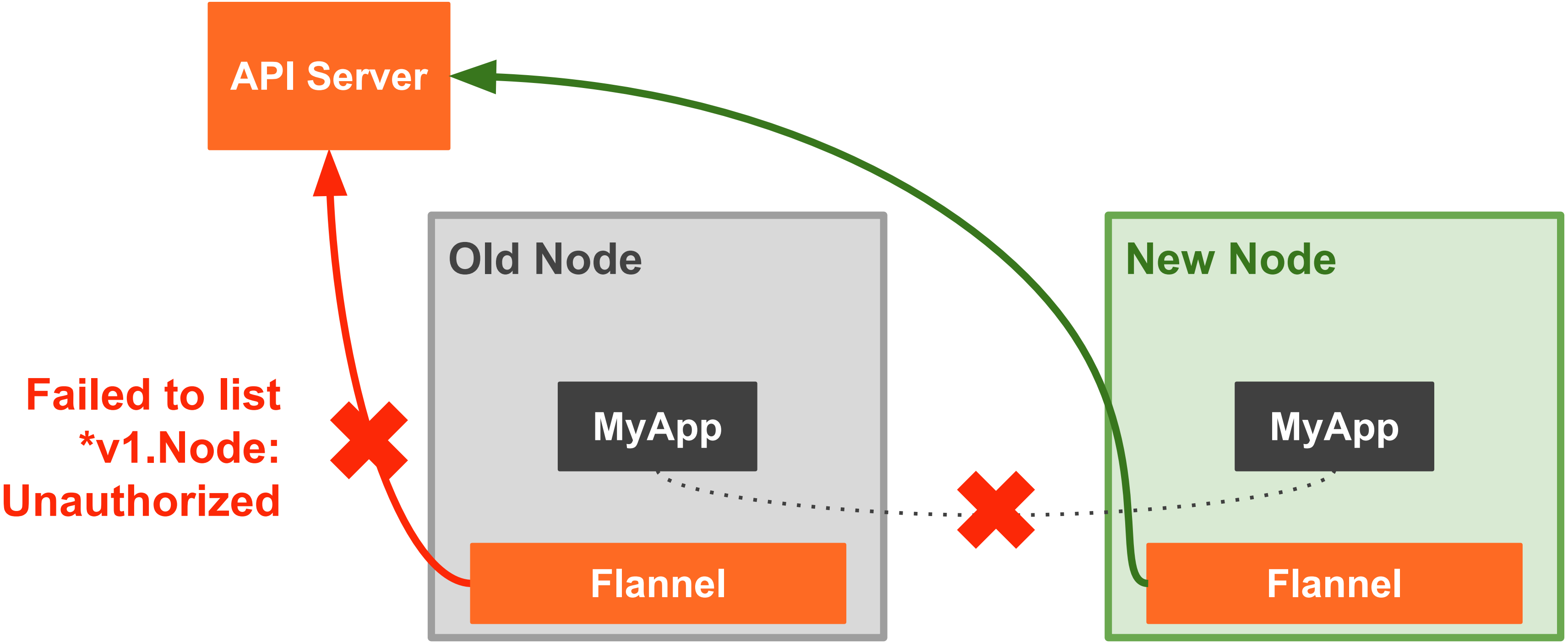
INCIDENT #4: RBAC CHANGES



```
@@ -44,3 +44,20 @@ post_apply:
  - name: zmon-scheduler
    kind: VerticalPodAutoscaler
    namespace: visibility
+ - name: kubernetes-dashboard
+   kind: RoleBinding
+   namespace: kube-system
+ - name: system
+   namespace: kube-system
+   kind: ServiceAccount
+ - name: privileged-psp
+   namespace: kube-system
+   kind: RoleBinding
+ - name: cdp-deployer
+   kind: ClusterRoleBinding
```

github.com/zalando-incubator/kubernetes-on-aws/pull/2510

INCIDENT #4: NETWORK SPLIT



Failed to list *v1.Node: Unauthorized

INCIDENT

#5

INCIDENT #5: IMPACT

Error during Pod creation:

```
MountVolume.SetUp failed for volume
 "outfit-delivery-api-credentials" :
 secrets "outfit-delivery-api-credentials" not found
```

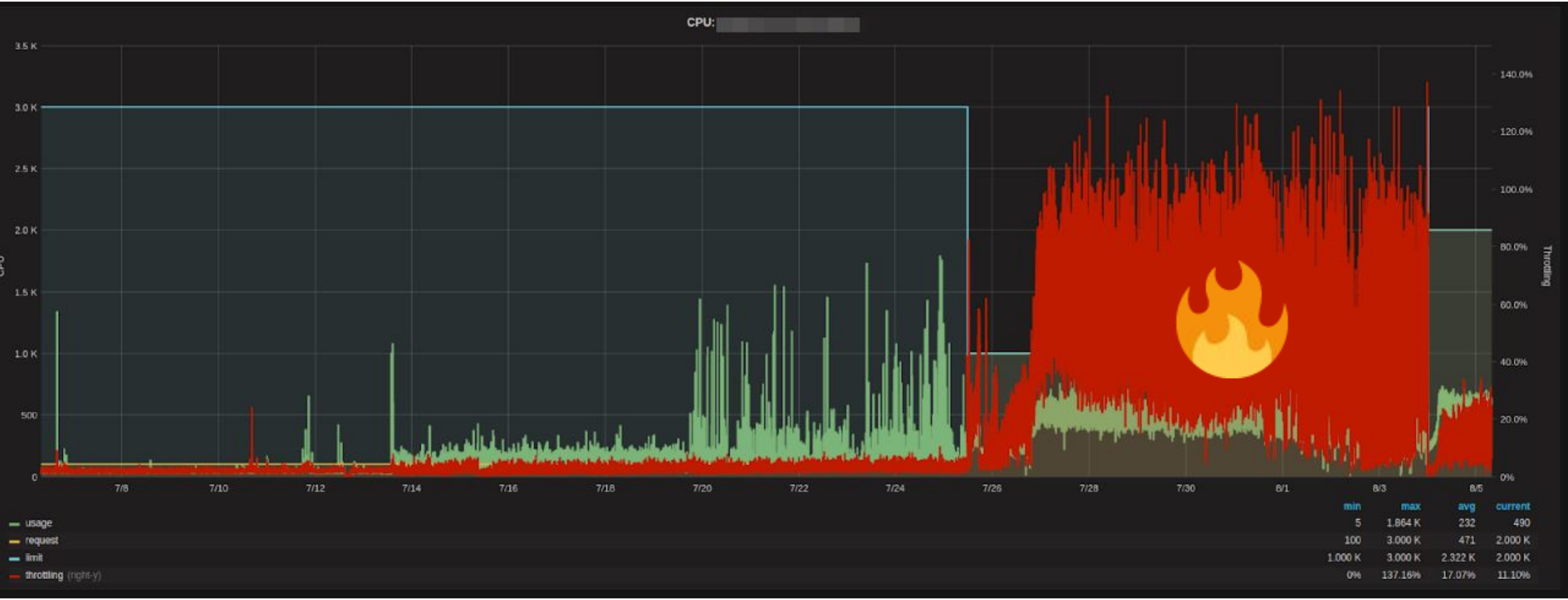
⇒ All new Kubernetes deployments fail 🔥

INCIDENT #5: CREDENTIALS QUEUE

```
17:30:07 | [pool-6-thread-1    ] | Current queue size: 7115, current number of active workers: 20
17:31:07 | [pool-6-thread-1    ] | Current queue size: 7505, current number of active workers: 20
17:32:07 | [pool-6-thread-1    ] | Current queue size: 7886, current number of active workers: 20
..
17:37:07 | [pool-6-thread-1    ] | Current queue size: 9686, current number of active workers: 20
..
17:44:07 | [pool-6-thread-1    ] | Current queue size: 11976, current number of active workers: 20
..
19:16:07 | [pool-6-thread-1    ] | Current queue size: 58381, current number of active workers: 20
```

🔥

INCIDENT #5: CPU THROTTLING



	min	max	avg	current
usage	5	1.864 K	232	490
request	100	3.000 K	471	2.000 K
limit	1.000 K	3.000 K	2.322 K	2.000 K
throttling (right-y)	0%	137.16%	17.07%	11.10%

INCIDENT #5: WHAT HAPPENED

Scaled down IAM provider
to reduce **Slack**

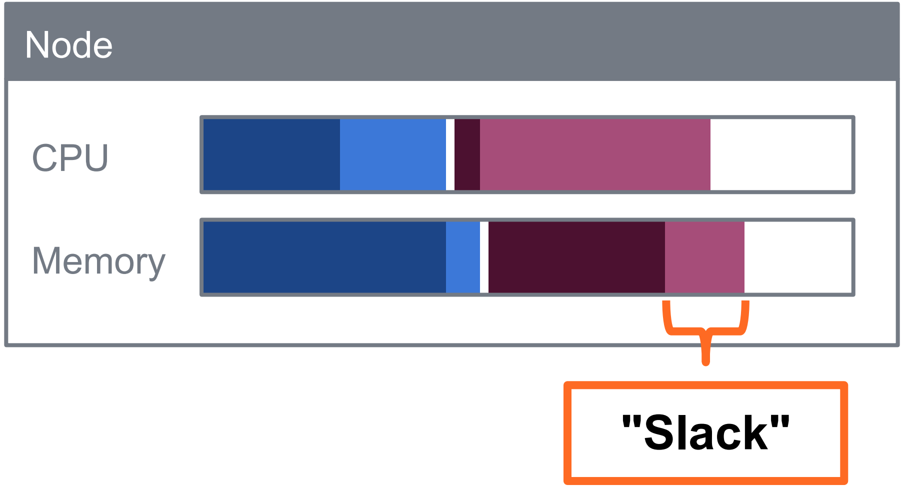
+ Number of deployments increased

⇒ Process could not process credentials fast enough

SLACK

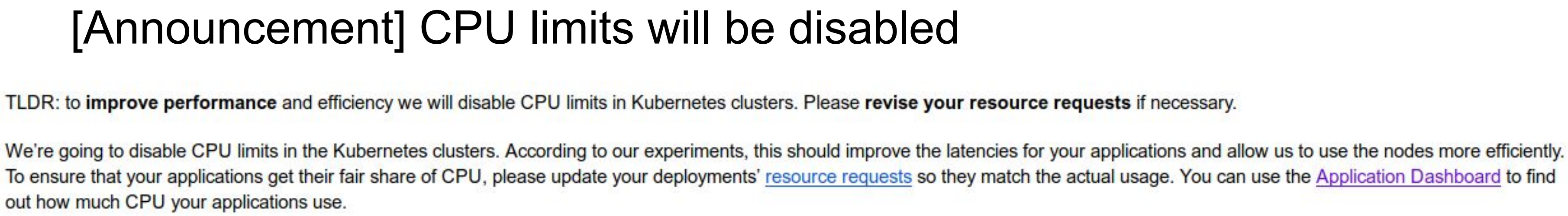
CPU/memory requests "block" resources on nodes.

Difference between actual usage and requests → **Slack**



DISABLING CPU THROTTLING

```
kubelet … --cpu-cfs-quota=false
```

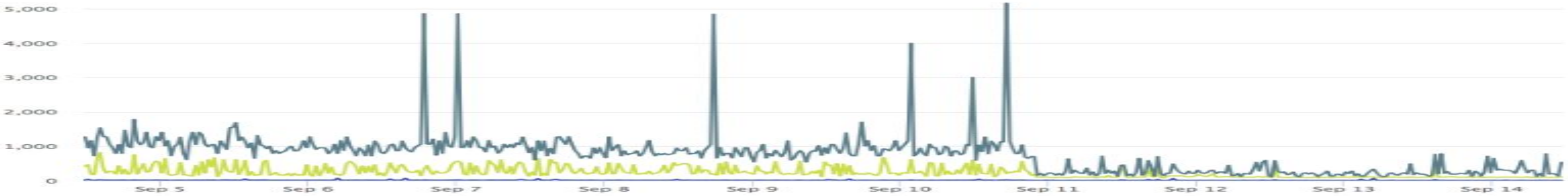


[Announcement] CPU limits will be disabled

TLDR: to **improve performance** and efficiency we will disable CPU limits in Kubernetes clusters. Please **revise your resource requests** if necessary.

We're going to disable CPU limits in the Kubernetes clusters. According to our experiments, this should improve the latencies for your applications and allow us to use the nodes more efficiently. To ensure that your applications get their fair share of CPU, please update your deployments' resource requests so they match the actual usage. You can use the Application Dashboard to find out how much CPU your applications use.

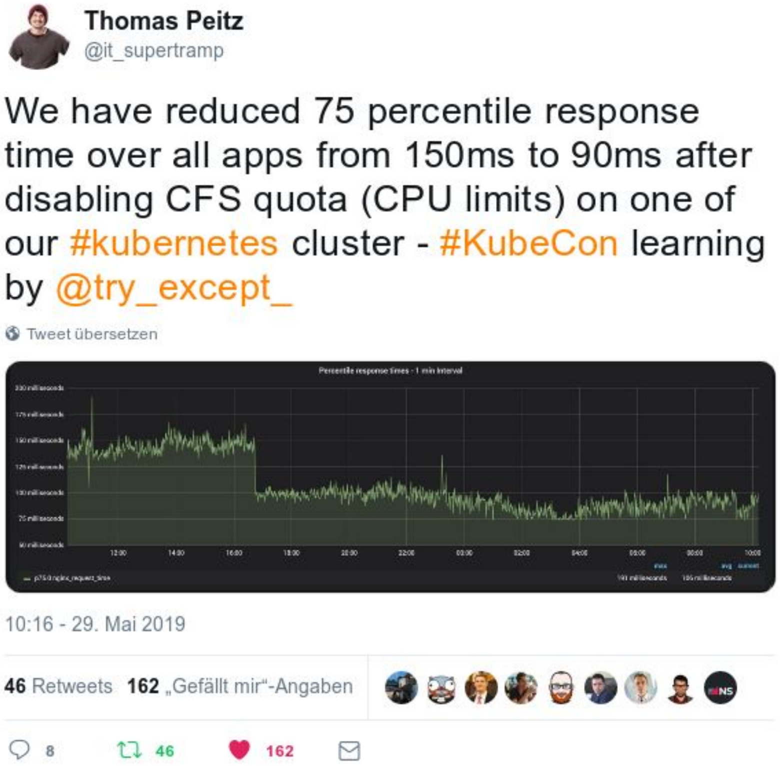
⇒ Ingress Latency Improvements



https://www.youtube.com/watch?v=4QyecOoPsGU

zalando

DISABLING CPU THROTTLING



Thomas Peitz
@it_supertramp

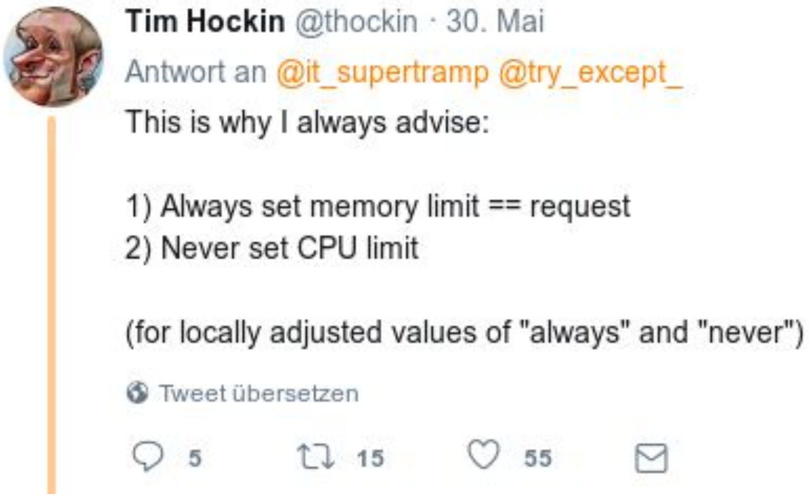
We have reduced 75 percentile response time over all apps from 150ms to 90ms after disabling CFS quota (CPU limits) on one of our #kubernetes cluster - #KubeCon learning by @try_except_

Tweet übersetzen

10:16 - 29. Mai 2019

46 Retweets **162** „Gefällt mir"-Angaben

8 46 162



Tim Hockin @thockin · 30. Mai

Antwort an @it_supertramp @try_except_

This is why I always advise:

1) Always set memory limit == request
2) Never set CPU limit

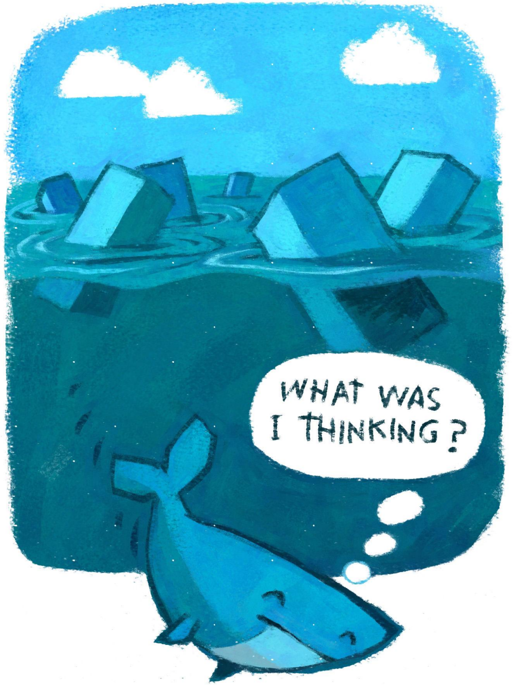
(for locally adjusted values of "always" and "never")

Tweet übersetzen

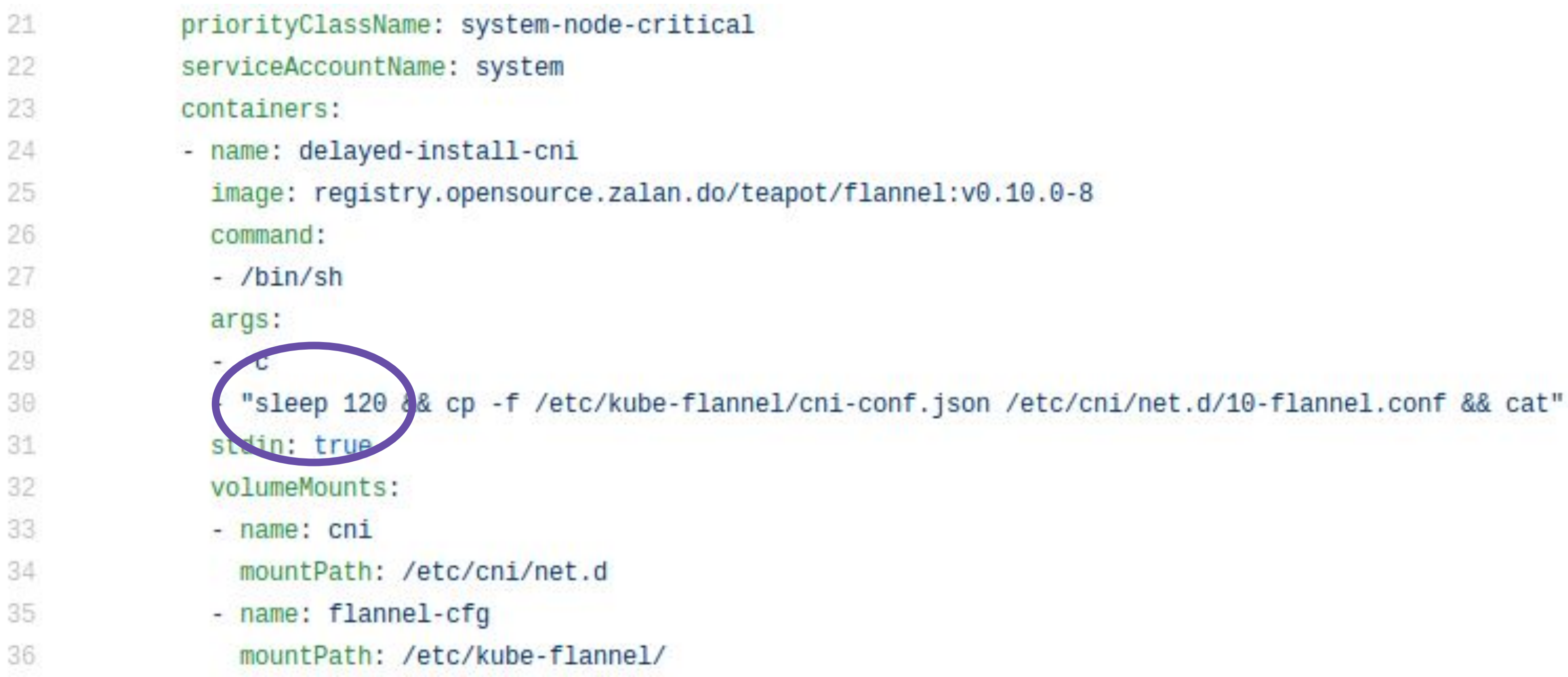
5 15 55

MORE TOPICS

- 🔥 **Graceful Pod shutdown** and race conditions (endpoints, Ingress)
- 🔥 **Incompatible Kubernetes changes**
- 🔥 CoreOS **ContainerLinux** "stable" won't boot
- 🔥 Kubernetes **EBS volume handling**
- 🔥 **Docker**



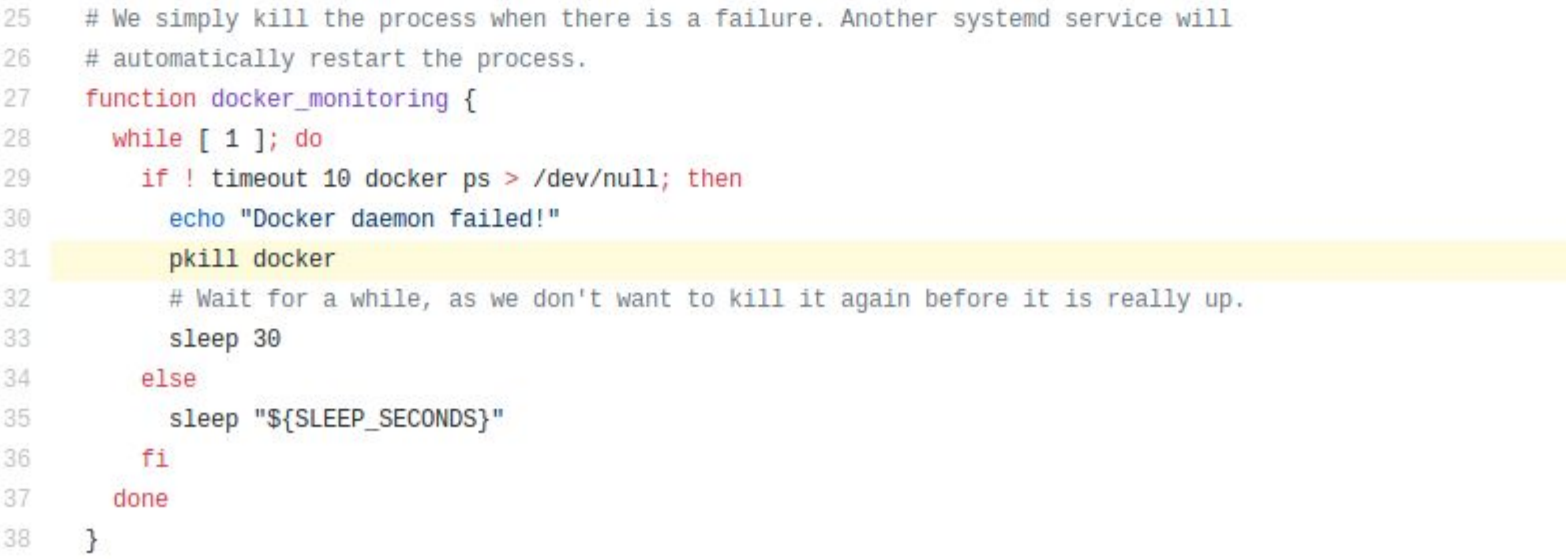
RACE CONDITIONS..



```
priorityClassName: system-node-critical
serviceAccountName: system
containers:
- name: delayed-install-cni
  image: registry.opensource.zalan.do/teapot/flannel:v0.10.0-8
  command:
  - /bin/sh
  args:
  - -c
  - "sleep 120 && cp -f /etc/kube-flannel/cni-conf.json /etc/cni/net.d/10-flannel.conf && cat"
  stdin: true
  volumeMounts:
  - name: cni
    mountPath: /etc/cni/net.d
  - name: flannel-cfg
    mountPath: /etc/kube-flannel/
```

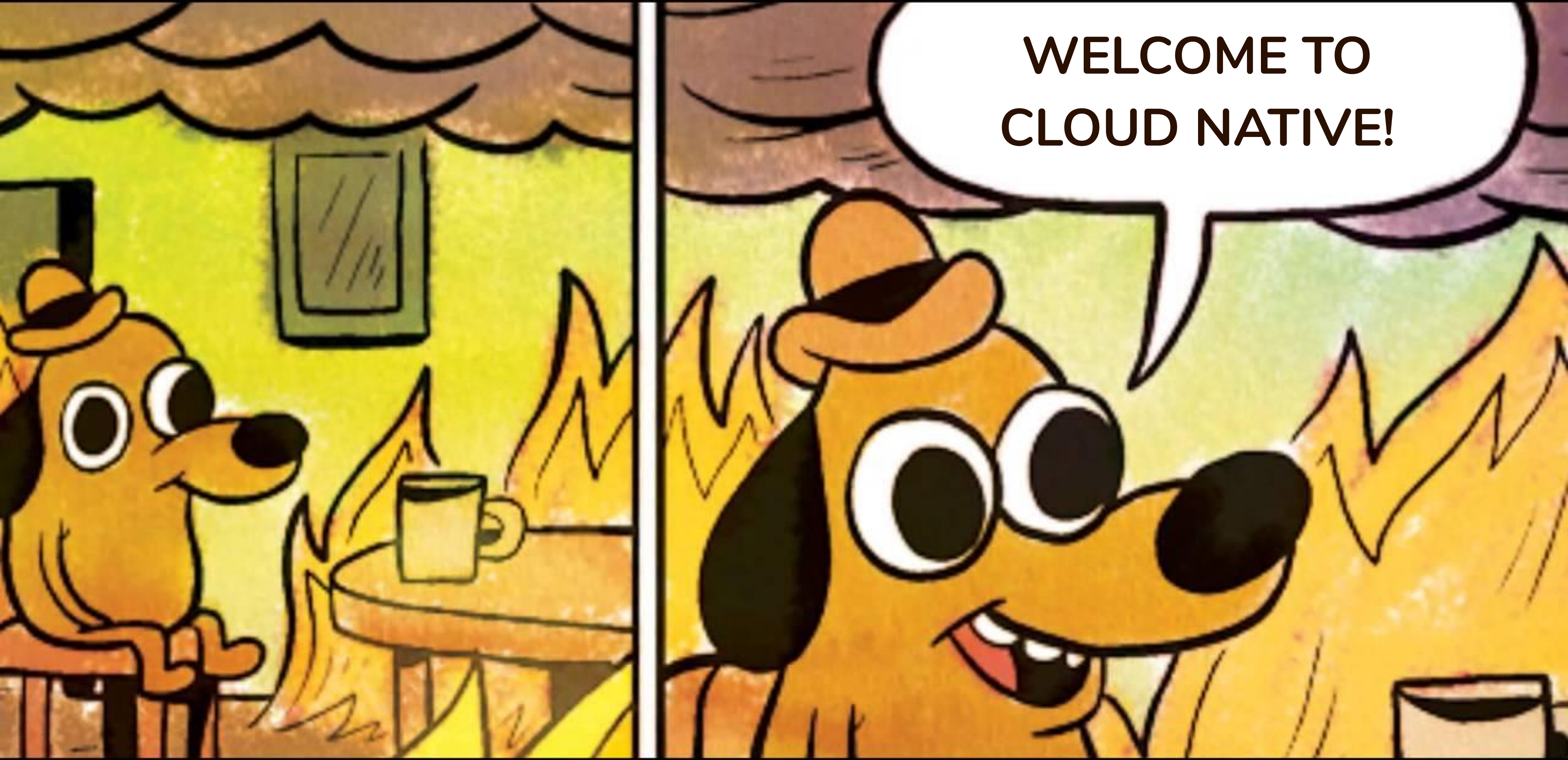
github.com/zalando-incubator/kubernetes-on-aws

DOCKER.. (ON GKE)



```
# We simply kill the process when there is a failure. Another systemd service will
# automatically restart the process.
function docker_monitoring {
  while [ 1 ]; do
    if ! timeout 10 docker ps > /dev/null; then
      echo "Docker daemon failed!"
      pkill docker
      # Wait for a while, as we don't want to kill it again before it is really up.
      sleep 30
    else
      sleep "${SLEEP_SECONDS}"
    fi
  done
}
```

https://github.com/kubernetes/kubernetes/blob/8fd414537b5143ab039cb910590237cabf4af783/cluster/gce/gci/health-monitor.sh#L29

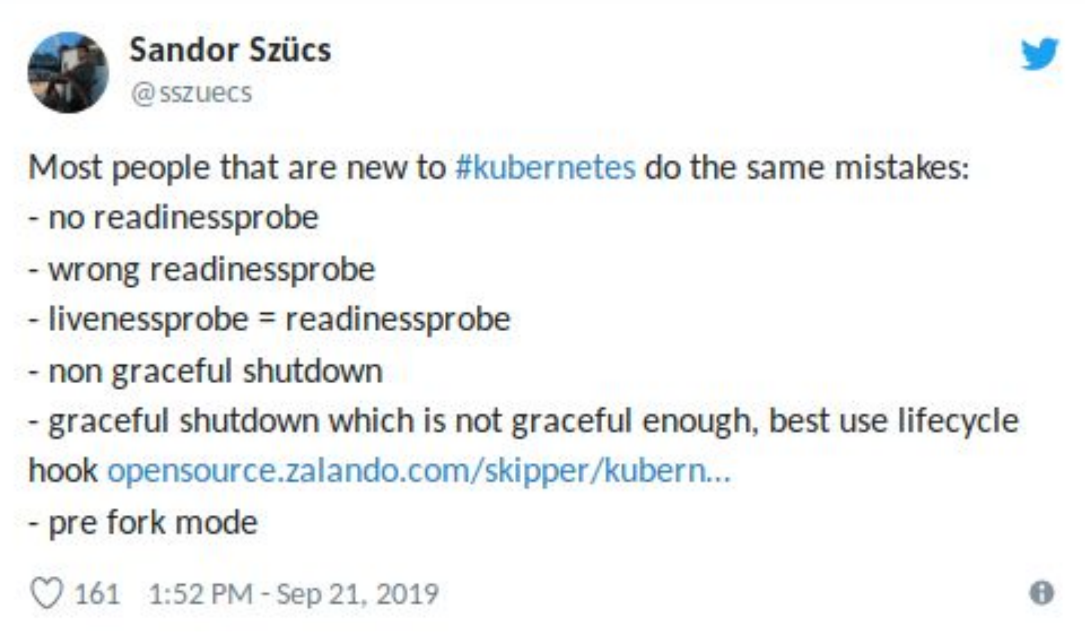

WELCOME TO
CLOUD NATIVE!

COMMON PITFALLS

COMMON PITFALLS

- Insufficient e2e tests
- Readiness & Liveness Probes
- Resource Requests & Limits
- DNS

READINESS & LIVENESS PROBES



Sandor Szücs
@sszuecs

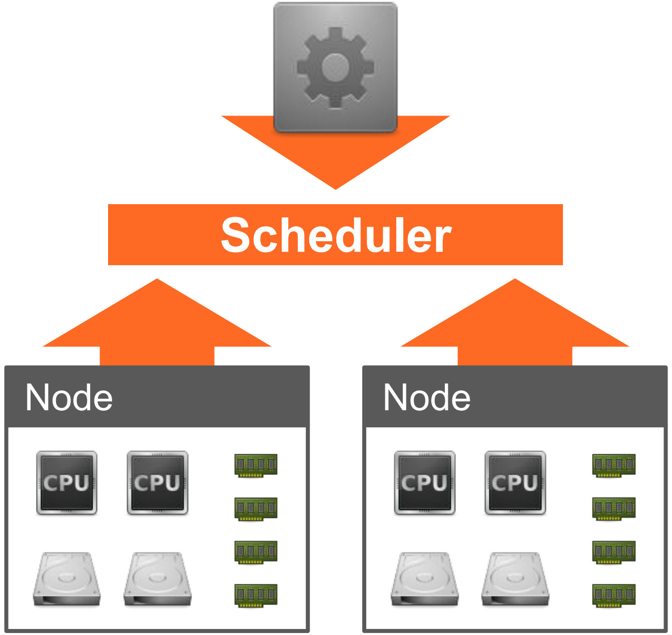
Most people that are new to #kubernetes do the same mistakes:
- no readinessprobe
- wrong readinessprobe
- livenessprobe = readinessprobe
- non graceful shutdown
- graceful shutdown which is not graceful enough, best use lifecycle hook opensource.zalando.com/skipper/kubern...
- pre fork mode

161 1:52 PM - Sep 21, 2019

srcco.de/posts/kubernetes-liveness-probes-are-dangerous.html

RESOURCE REQUESTS & LIMITS

- No resource requests / limits
- QoS
- OOM
- Overcommit
- CPU throttling


Scheduler

Node

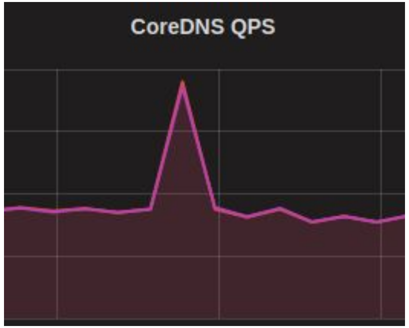
Node

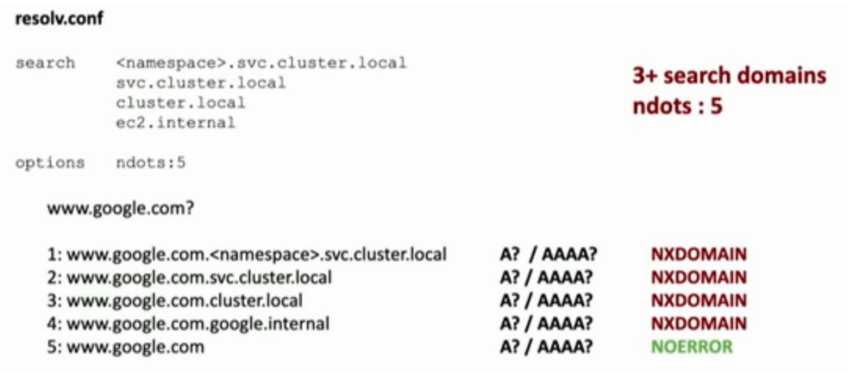
youtube.com/watch?v=4QyecOoPsGU

DNS

- ndots: 5
- musl, conntrack, UDP
- overload

It's ~~never~~ always DNS.





resolv.conf

```
search     <namespace>.svc.cluster.local
           svc.cluster.local
           cluster.local
           ec2.internal

options    ndots:5
```

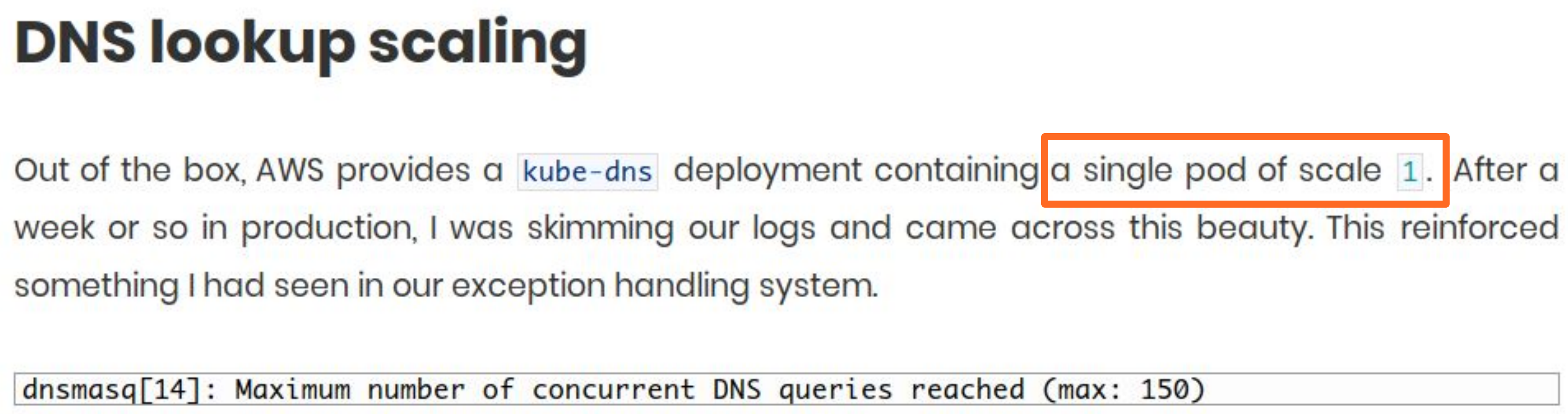
3+ search domains
ndots : 5

www.google.com?

1: www.google.com.<namespace>.svc.cluster.local	A? / AAAA?	NXDOMAIN
2: www.google.com.svc.cluster.local	A? / AAAA?	NXDOMAIN
3: www.google.com.cluster.local	A? / AAAA?	NXDOMAIN
4: www.google.com.google.internal	A? / AAAA?	NXDOMAIN
5: www.google.com	A? / AAAA?	NOERROR

youtube.com/watch?v=QKI-JRs2RIE

AWS EKS IN PRODUCTION

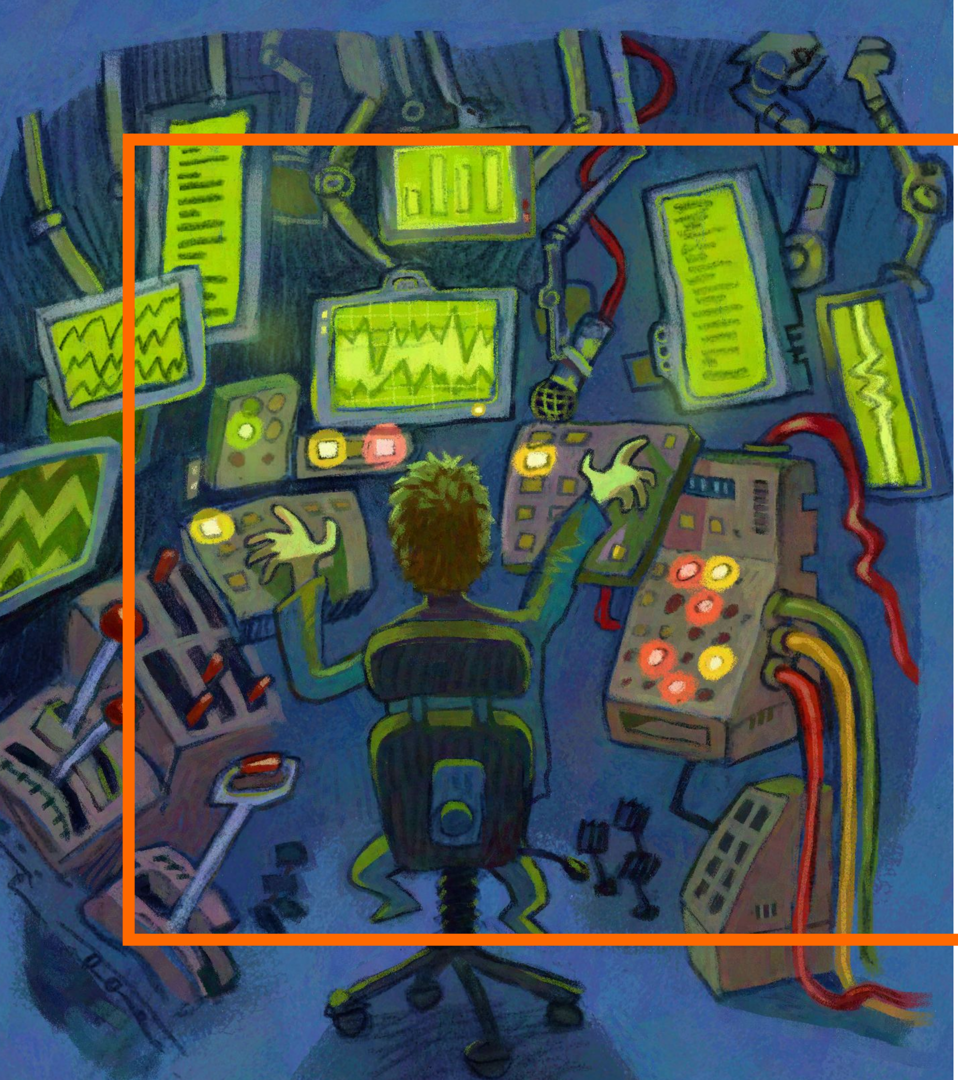


DNS lookup scaling

Out of the box, AWS provides a `kube-dns` deployment containing a single pod of scale `1`. After a week or so in production, I was skimming our logs and came across this beauty. This reinforced something I had seen in our exception handling system.

```
dnsmasq[14]: Maximum number of concurrent DNS queries reached (max: 150)
```

kubedex.com/90-days-of-aws-eks-in-production/

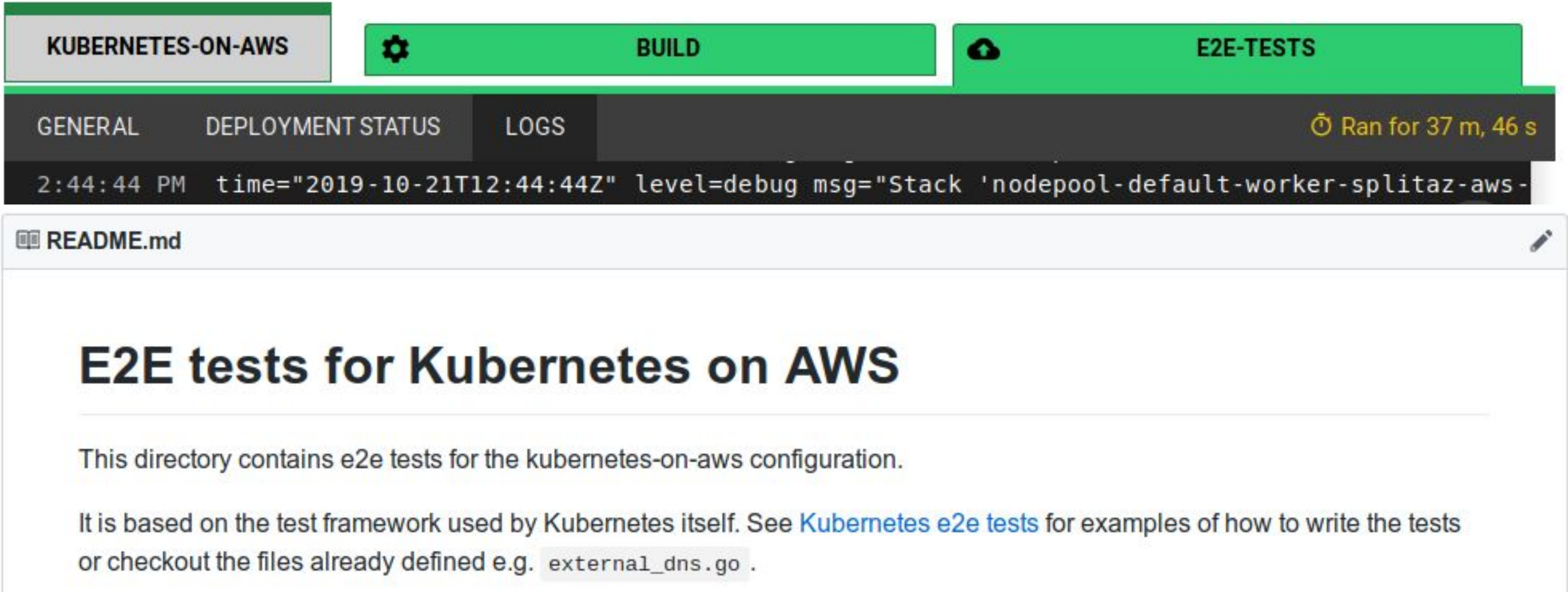


TOOLS & PRACTICES TO HELP YOU

HELPFUL

- Automated e2e tests
- Monitoring
- OpenTracing
- Kubernetes Web View
- Emergency access
- Kubernetes Failure Stories

AUTOMATED E2E TESTS



```
2:44:44 PM  time="2019-10-21T12:44:44Z" level=debug msg="Stack 'nodepool-default-worker-splitaz-aws-
```

E2E tests for Kubernetes on AWS

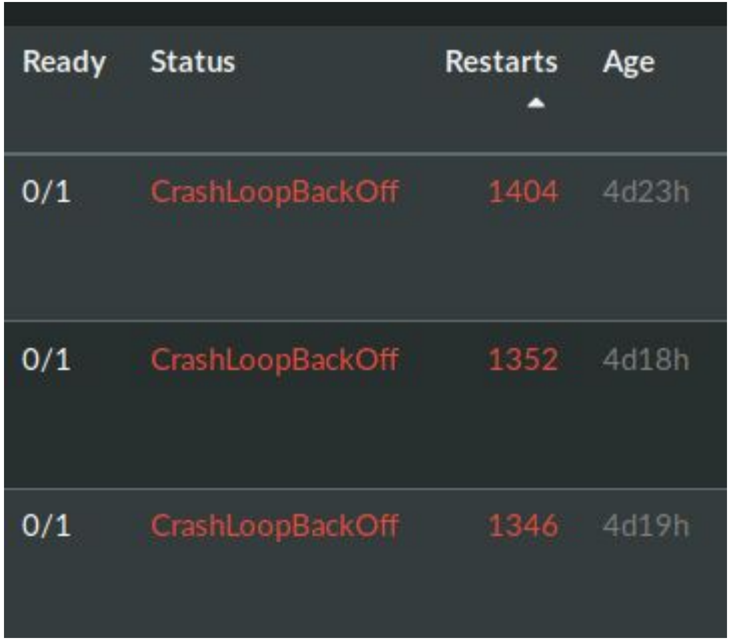
This directory contains e2e tests for the kubernetes-on-aws configuration.

It is based on the test framework used by Kubernetes itself. See Kubernetes e2e tests for examples of how to write the tests or checkout the files already defined e.g. `external_dns.go` .

github.com/zalando-incubator/kubernetes-on-aws/tree/dev/test/e2e

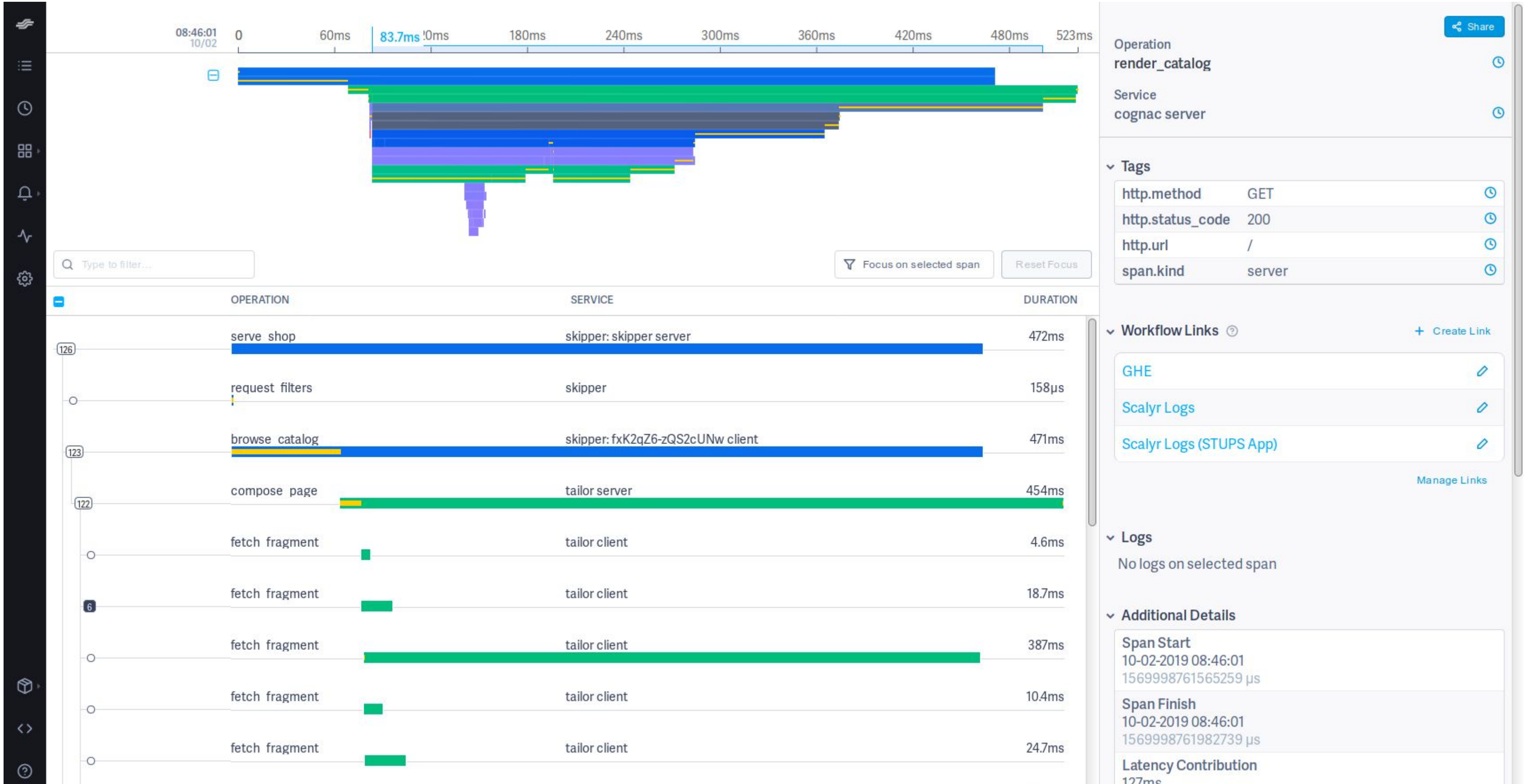
MONITORING

- Alert on symptoms, not potential causes
- What do you need to monitor to ensure cluster availability?
- SLOs



Ready	Status	Restarts	Age
0/1	CrashLoopBackOff	1404	4d23h
0/1	CrashLoopBackOff	1352	4d18h
0/1	CrashLoopBackOff	1346	4d19h

OPENTRACING



OPERATION	SERVICE	DURATION
serve_shop	skipper: skipper server	472ms
request_filters	skipper	158µs
browse_catalog	skipper: fxK2qZ6-zQS2cUNw client	471ms
compose_page	tailor server	454ms
fetch_fragment	tailor client	4.6ms
fetch_fragment	tailor client	18.7ms
fetch_fragment	tailor client	387ms
fetch_fragment	tailor client	10.4ms
fetch_fragment	tailor client	24.7ms

Operation
render_catalog

Service
cognac server

Tags

http.method	GET
http.status_code	200
http.url	/
span.kind	server

Workflow Links

GHE
Scalyr Logs
Scalyr Logs (STUPS App)

Logs

No logs on selected span

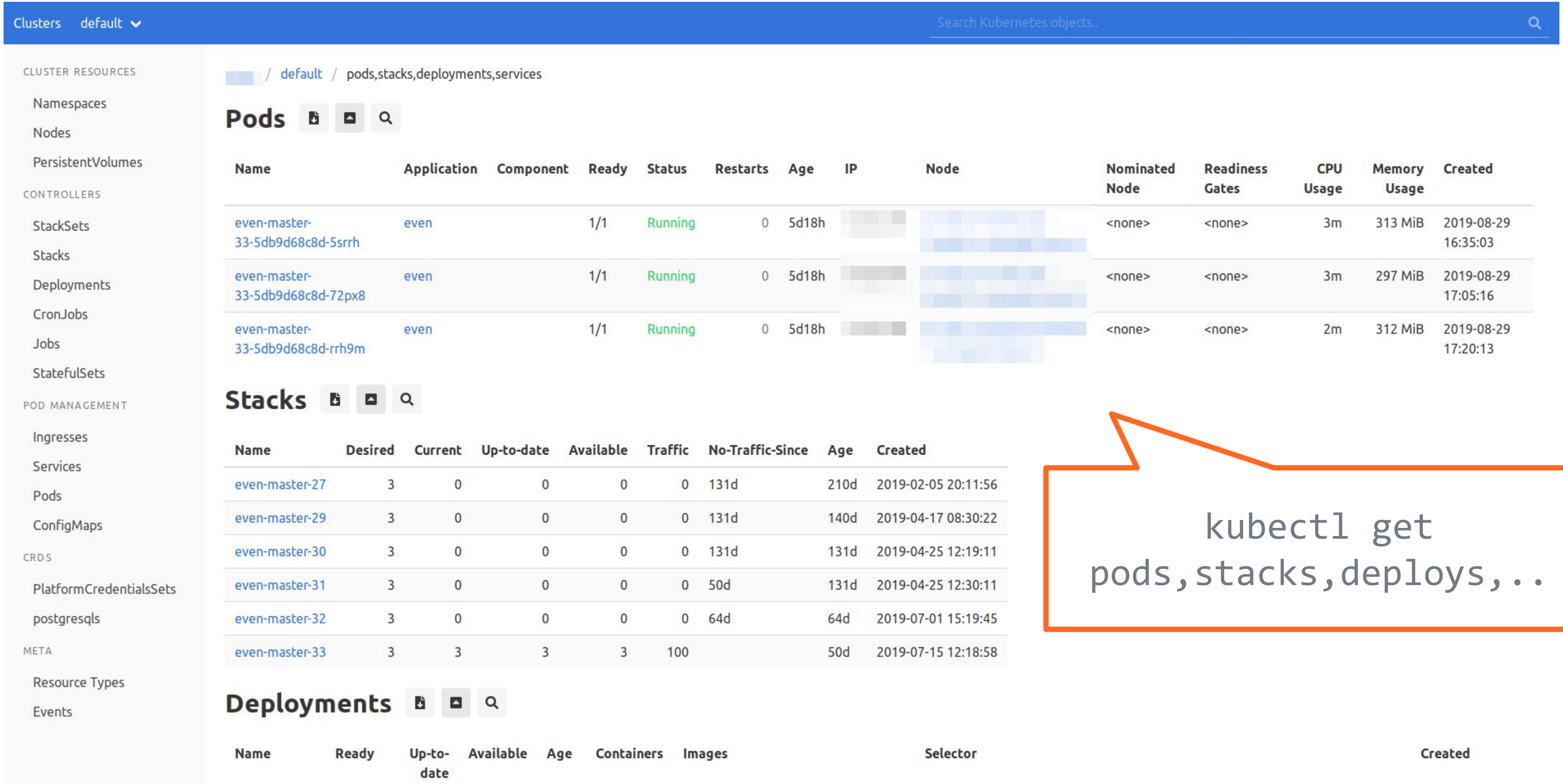
Additional Details

Span Start
10-02-2019 08:46:01
1569998761565259 µs

Span Finish
10-02-2019 08:46:01
1569998761982739 µs

Latency Contribution
127ms

KUBERNETES WEB VIEW



CLUSTER RESOURCES
- Namespaces
- Nodes
- PersistentVolumes

CONTROLLERS
- StackSets
- Stacks
- Deployments
- CronJobs
- Jobs
- StatefulSets

POD MANAGEMENT
- Ingresses
- Services
- Pods
- ConfigMaps

CRDS
- PlatformCredentialsSets
- postgresqls

META
- Resource Types
- Events

[redacted] / default / pods,stacks,deployments,services

Pods

Name	Application	Component	Ready	Status	Restarts	Age	IP	Node	Nominated Node	Readiness Gates	CPU Usage	Memory Usage	Created
even-master-33-5db9d68c8d-5srrh	even		1/1	Running	0	5d18h	[redacted]	[redacted]	<none>	<none>	3m	313 MiB	2019-08-29 16:35:03
even-master-33-5db9d68c8d-72px8	even		1/1	Running	0	5d18h	[redacted]	[redacted]	<none>	<none>	3m	297 MiB	2019-08-29 17:05:16
even-master-33-5db9d68c8d-rrh9m	even		1/1	Running	0	5d18h	[redacted]	[redacted]	<none>	<none>	2m	312 MiB	2019-08-29 17:20:13

Stacks

Name	Desired	Current	Up-to-date	Available	Traffic	No-Traffic-Since	Age	Created
even-master-27	3	0	0	0	0	131d	210d	2019-02-05 20:11:56
even-master-29	3	0	0	0	0	131d	140d	2019-04-17 08:30:22
even-master-30	3	0	0	0	0	131d	131d	2019-04-25 12:19:11
even-master-31	3	0	0	0	0	50d	131d	2019-04-25 12:30:11
even-master-32	3	0	0	0	0	64d	64d	2019-07-01 15:19:45
even-master-33	3	3	3	3	100		50d	2019-07-15 12:18:58

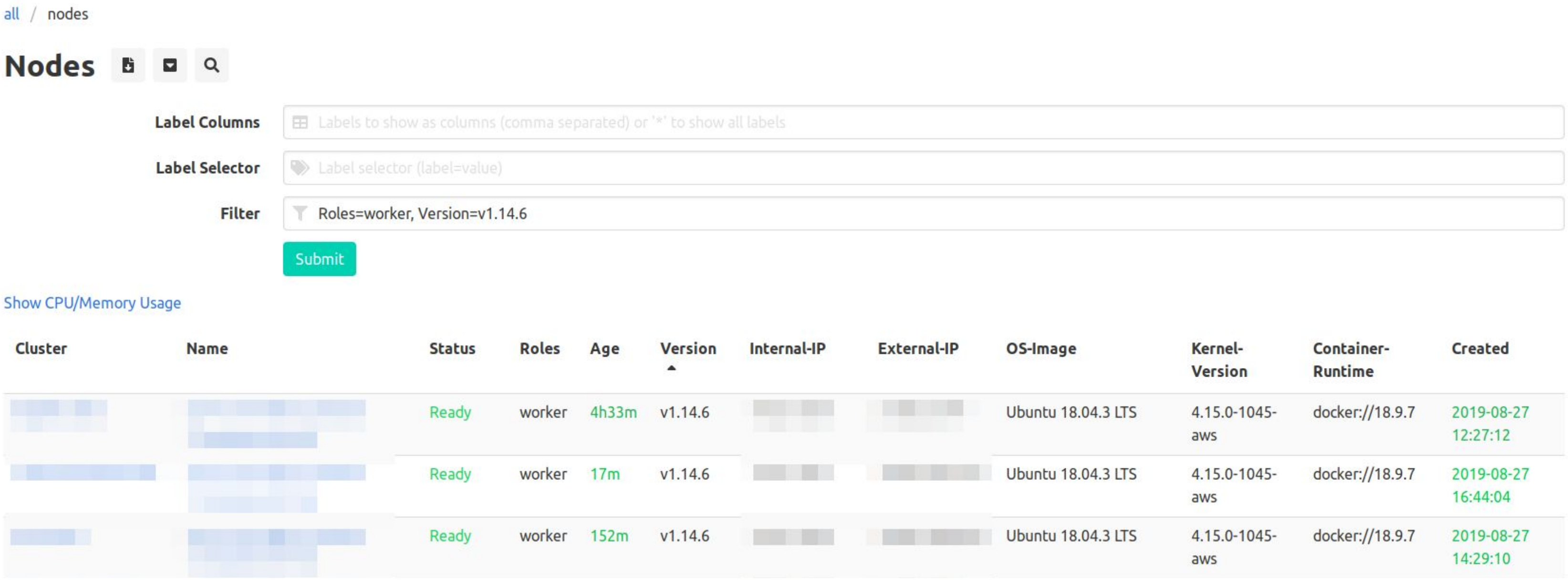
Deployments

Name	Ready	Up-to-date	Available	Age	Containers	Images	Selector	Created

```
kubectl get
pods,stacks,deploys,..
```

UPGRADE TO KUBERNETES 1.14

"Found 1223 rows for 1 resource type in 148 clusters in 3.301 seconds."



all / nodes

Nodes

Label Columns: Labels to show as columns (comma separated) or '*' to show all labels

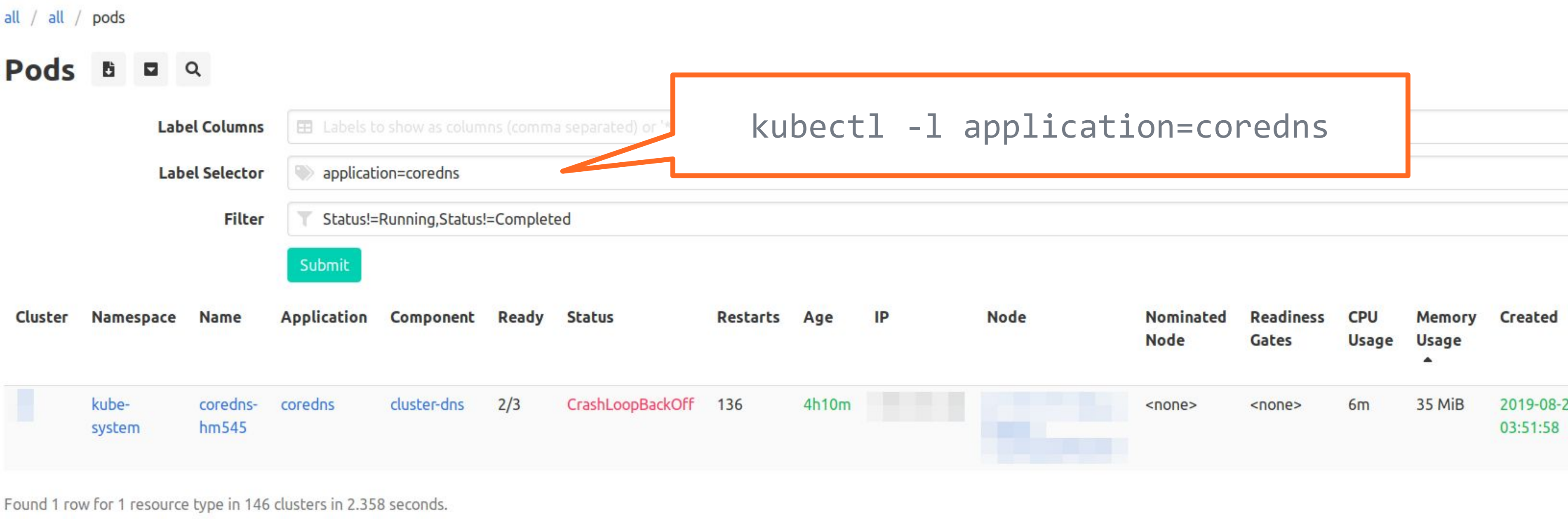
Label Selector: Label selector (label=value)

Filter: Roles=worker, Version=v1.14.6

Show CPU/Memory Usage

Cluster	Name	Status	Roles	Age	Version	Internal-IP	External-IP	OS-Image	Kernel-Version	Container-Runtime	Created
[redacted]	[redacted]	Ready	worker	4h33m	v1.14.6	[redacted]	[redacted]	Ubuntu 18.04.3 LTS	4.15.0-1045-aws	docker://18.9.7	2019-08-27 12:27:12
[redacted]	[redacted]	Ready	worker	17m	v1.14.6	[redacted]	[redacted]	Ubuntu 18.04.3 LTS	4.15.0-1045-aws	docker://18.9.7	2019-08-27 16:44:04
[redacted]	[redacted]	Ready	worker	152m	v1.14.6	[redacted]	[redacted]	Ubuntu 18.04.3 LTS	4.15.0-1045-aws	docker://18.9.7	2019-08-27 14:29:10

MULTIPLE CONDITIONS



```
kubectl -l application=coredns
```

all / all / pods

Pods

Label Columns: Labels to show as columns (comma separated) or '*

Label Selector: application=coredns

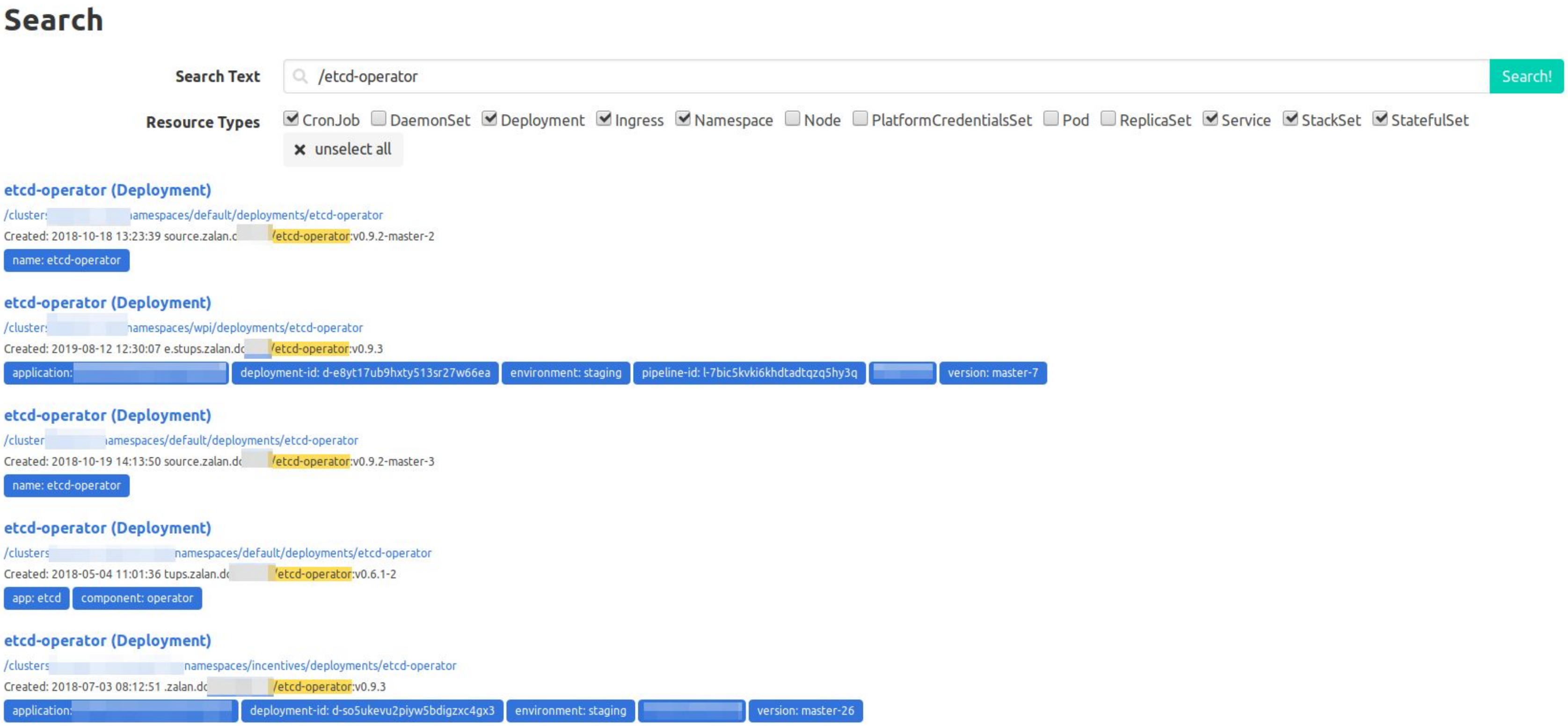
Filter: Status!=Running,Status!=Completed

Submit

Cluster	Namespace	Name	Application	Component	Ready	Status	Restarts	Age	IP	Node	Nominated Node	Readiness Gates	CPU Usage	Memory Usage	Created
[redacted]	kube-system	coredns-hm545	coredns	cluster-dns	2/3	CrashLoopBackOff	136	4h10m	[redacted]	[redacted]	<none>	<none>	6m	35 MiB	2019-08-2 03:51:58

Found 1 row for 1 resource type in 146 clusters in 2.358 seconds.

SEARCHING ACROSS 140+ CLUSTERS



Search

Search Text /etcd-operator

Resource Types ☑ CronJob ☐ DaemonSet ☑ Deployment ☑ Ingress ☑ Namespace ☐ Node ☐ PlatformCredentialsSet ☐ Pod ☐ ReplicaSet ☑ Service ☑ StackSet ☑ StatefulSet

etcd-operator (Deployment)
/cluster[redacted]amespaces/default/deployments/etcd-operator
Created: 2018-10-18 13:23:39 source.zalan.c[redacted]/etcd-operator:v0.9.2-master-2
name: etcd-operator

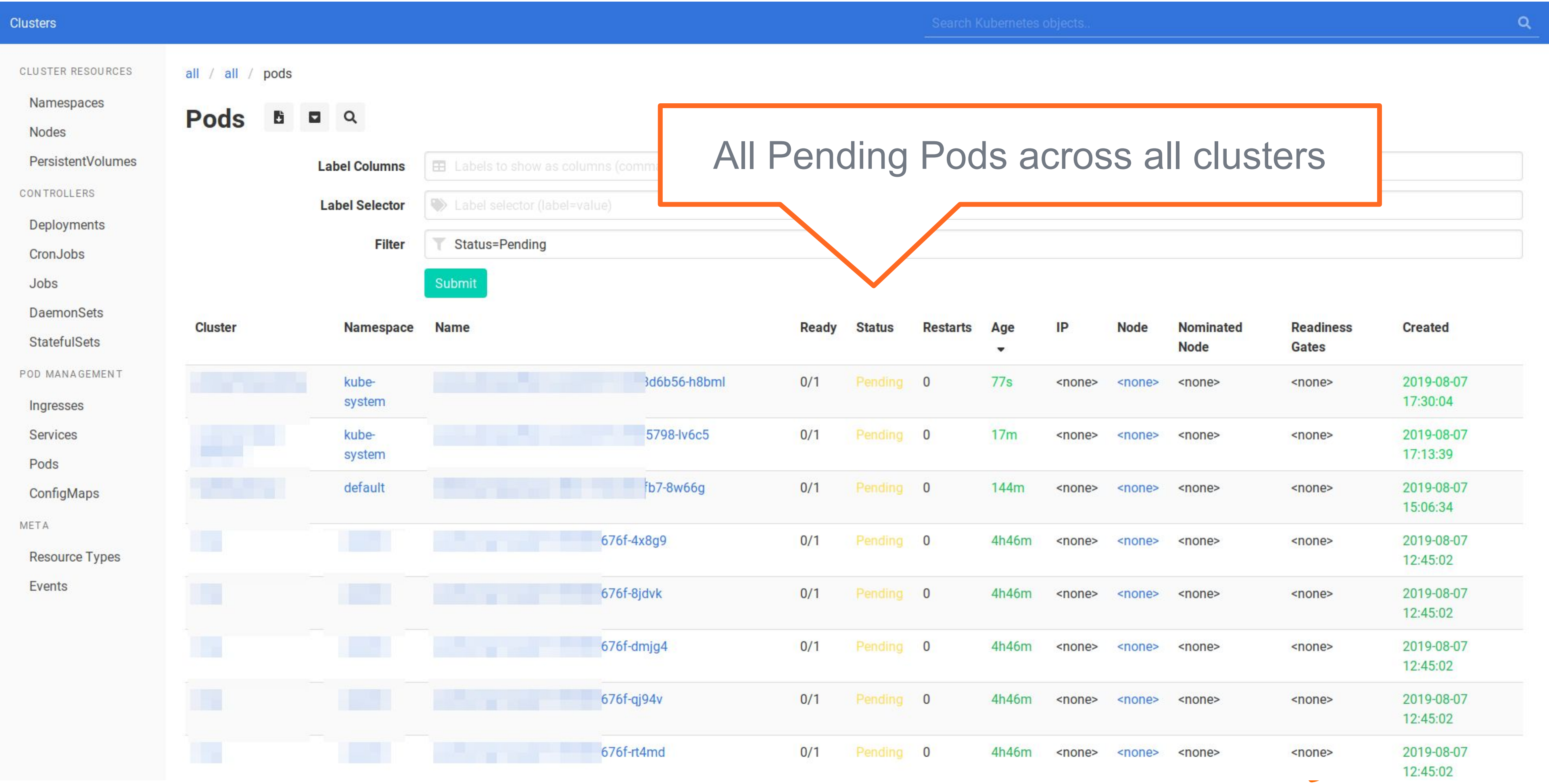
etcd-operator (Deployment)
/cluster[redacted]amespaces/wpi/deployments/etcd-operator
Created: 2019-08-12 12:30:07 e.stups.zalan.dc[redacted]/etcd-operator:v0.9.3
application:[redacted] deployment-id: d-e8yt17ub9hxty513sr27w66ea environment: staging pipeline-id: l-7bic5kvki6khdtadtqzq5hy3q [redacted] version: master-7

etcd-operator (Deployment)
/cluster[redacted]amespaces/default/deployments/etcd-operator
Created: 2018-10-19 14:13:50 source.zalan.dc[redacted]/etcd-operator:v0.9.2-master-3
name: etcd-operator

etcd-operator (Deployment)
/clusters[redacted]namespaces/default/deployments/etcd-operator
Created: 2018-05-04 11:01:36 tups.zalan.dc[redacted]'etcd-operator:v0.6.1-2
app: etcd component: operator

etcd-operator (Deployment)
/clusters[redacted]namespaces/incentives/deployments/etcd-operator
Created: 2018-07-03 08:12:51 .zalan.dc[redacted]/etcd-operator:v0.9.3
application:[redacted] deployment-id: d-so5ukevu2piyw5bdigzxc4gx3 environment: staging [redacted] version: master-26

codeberg.org/hjacobs/kube-web-view



Pods

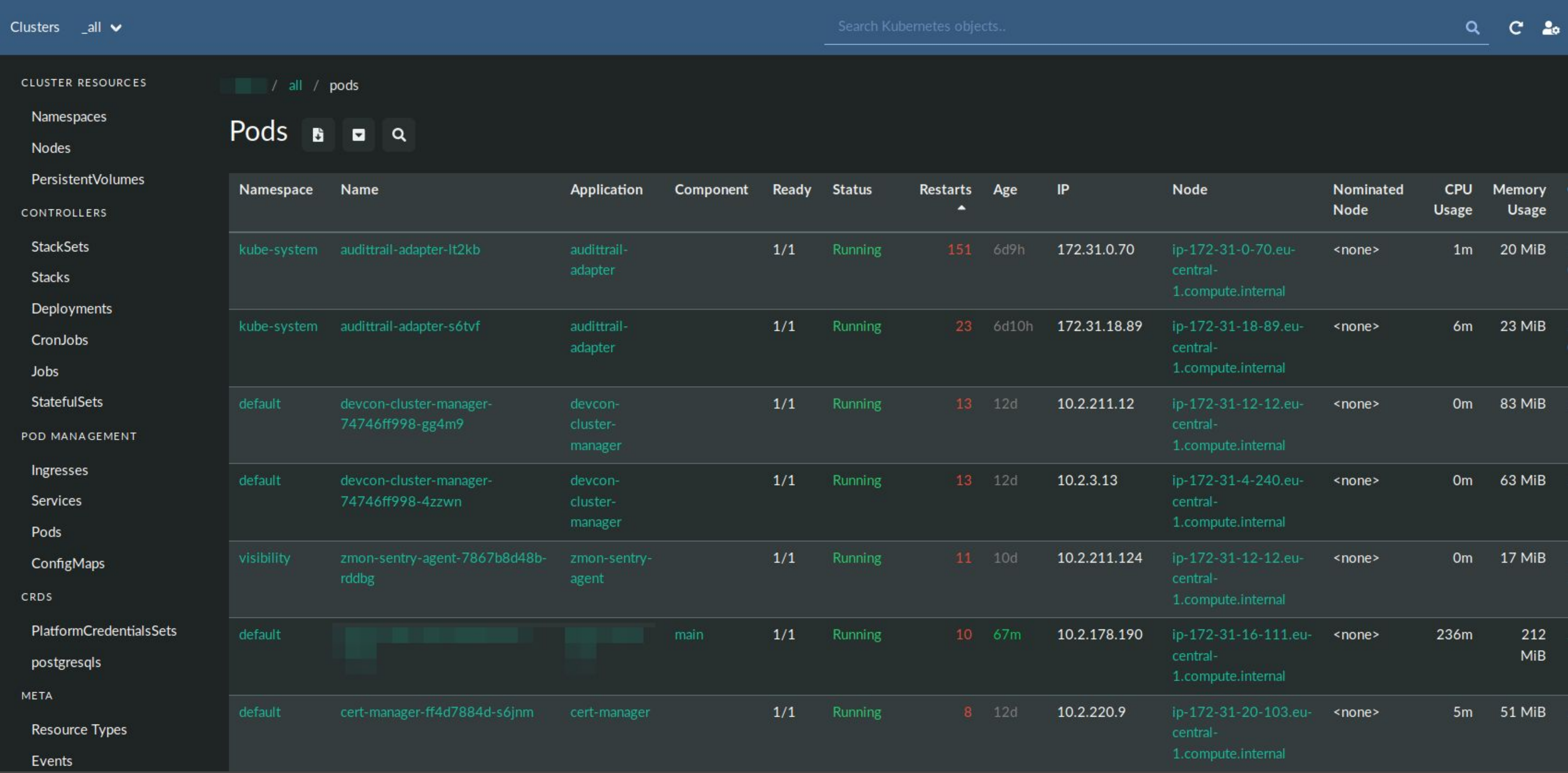
Label Columns: Labels to show as columns (comma

Label Selector: Label selector (label=value)

Filter: Status=Pending

All Pending Pods across all clusters

Cluster	Namespace	Name	Ready	Status	Restarts	Age	IP	Node	Nominated Node	Readiness Gates	Created
[redacted]	kube-system	[redacted]3d6b56-h8bml	0/1	Pending	0	77s	<none>	<none>	<none>	<none>	2019-08-07 17:30:04
[redacted]	kube-system	[redacted]5798-lv6c5	0/1	Pending	0	17m	<none>	<none>	<none>	<none>	2019-08-07 17:13:39
[redacted]	default	[redacted]fb7-8w66g	0/1	Pending	0	144m	<none>	<none>	<none>	<none>	2019-08-07 15:06:34
[redacted]	[redacted]	[redacted]676f-4x8g9	0/1	Pending	0	4h46m	<none>	<none>	<none>	<none>	2019-08-07 12:45:02
[redacted]	[redacted]	[redacted]676f-8jdvk	0/1	Pending	0	4h46m	<none>	<none>	<none>	<none>	2019-08-07 12:45:02
[redacted]	[redacted]	[redacted]676f-dmjg4	0/1	Pending	0	4h46m	<none>	<none>	<none>	<none>	2019-08-07 12:45:02
[redacted]	[redacted]	[redacted]676f-qj94v	0/1	Pending	0	4h46m	<none>	<none>	<none>	<none>	2019-08-07 12:45:02
[redacted]	[redacted]	[redacted]676f-rt4md	0/1	Pending	0	4h46m	<none>	<none>	<none>	<none>	2019-08-07 12:45:02



[redacted] / all / pods

Pods

Namespace	Name	Application	Component	Ready	Status	Restarts ▲	Age	IP	Node	Nominated Node	CPU Usage	Memory Usage
kube-system	audittrail-adapter-lt2kb	audittrail-adapter		1/1	Running	151	6d9h	172.31.0.70	ip-172-31-0-70.eu-central-1.compute.internal	<none>	1m	20 MiB
kube-system	audittrail-adapter-s6tvf	audittrail-adapter		1/1	Running	23	6d10h	172.31.18.89	ip-172-31-18-89.eu-central-1.compute.internal	<none>	6m	23 MiB
default	devcon-cluster-manager-74746ff998-gg4m9	devcon-cluster-manager		1/1	Running	13	12d	10.2.211.12	ip-172-31-12-12.eu-central-1.compute.internal	<none>	0m	83 MiB
default	devcon-cluster-manager-74746ff998-4zzwn	devcon-cluster-manager		1/1	Running	13	12d	10.2.3.13	ip-172-31-4-240.eu-central-1.compute.internal	<none>	0m	63 MiB
visibility	zmon-sentry-agent-7867b8d48b-rddbg	zmon-sentry-agent		1/1	Running	11	10d	10.2.211.124	ip-172-31-12-12.eu-central-1.compute.internal	<none>	0m	17 MiB
default	[redacted]	[redacted]	main	1/1	Running	10	67m	10.2.178.190	ip-172-31-16-111.eu-central-1.compute.internal	<none>	236m	212 MiB
default	cert-manager-ff4d7884d-s6jnm	cert-manager		1/1	Running	8	12d	10.2.220.9	ip-172-31-20-103.eu-central-1.compute.internal	<none>	5m	51 MiB

codeberg.org/hjacobs/kube-web-view

EMERGENCY ACCESS SERVICE

Emergency access by referencing Incident

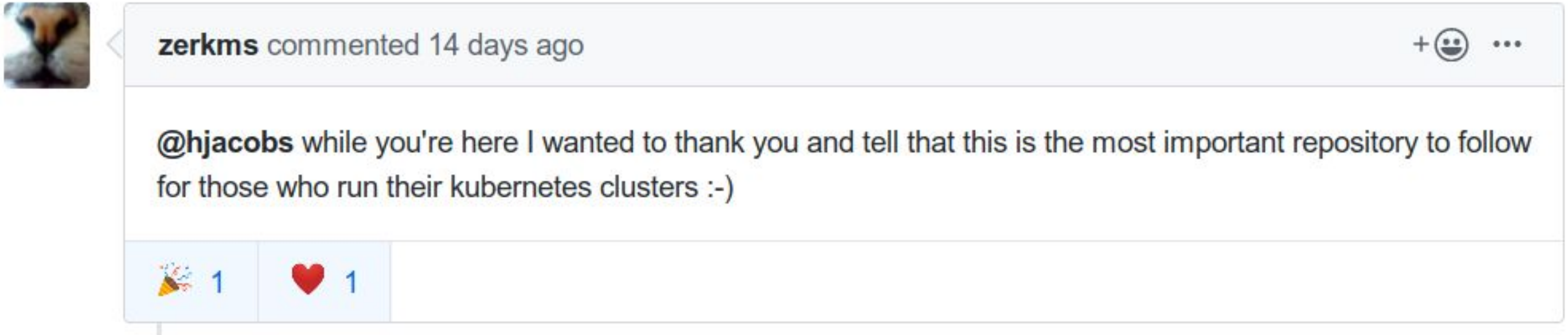
```
zkubectl cluster-access request \
    --emergency -i INC REASON
```

Privileged production access via 4-eyes

```
zkubectl cluster-access request REASON
zkubectl cluster-access approve USERNAME
```

KUBERNETES FAILURE STORIES

Learning about production pitfalls!



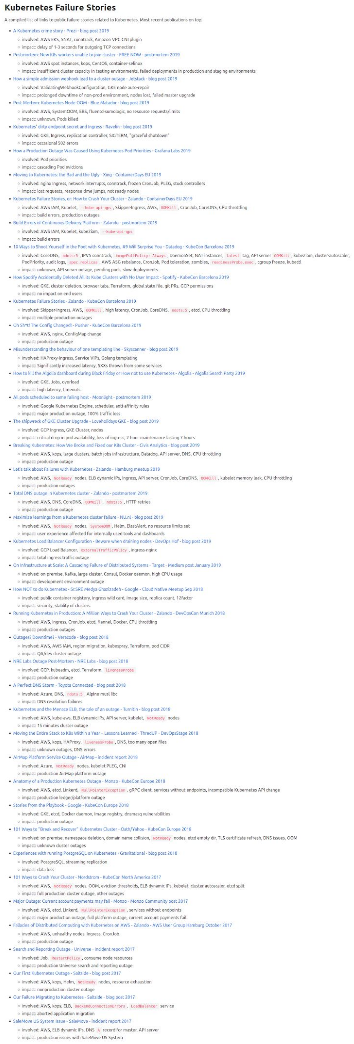
zerkms commented 14 days ago

@hjacobs while you're here I wanted to thank you and tell that this is the most important repository to follow for those who run their kubernetes clusters :-)

🎉 1 ❤️ 1

https://k8s.af

https://k8s.af



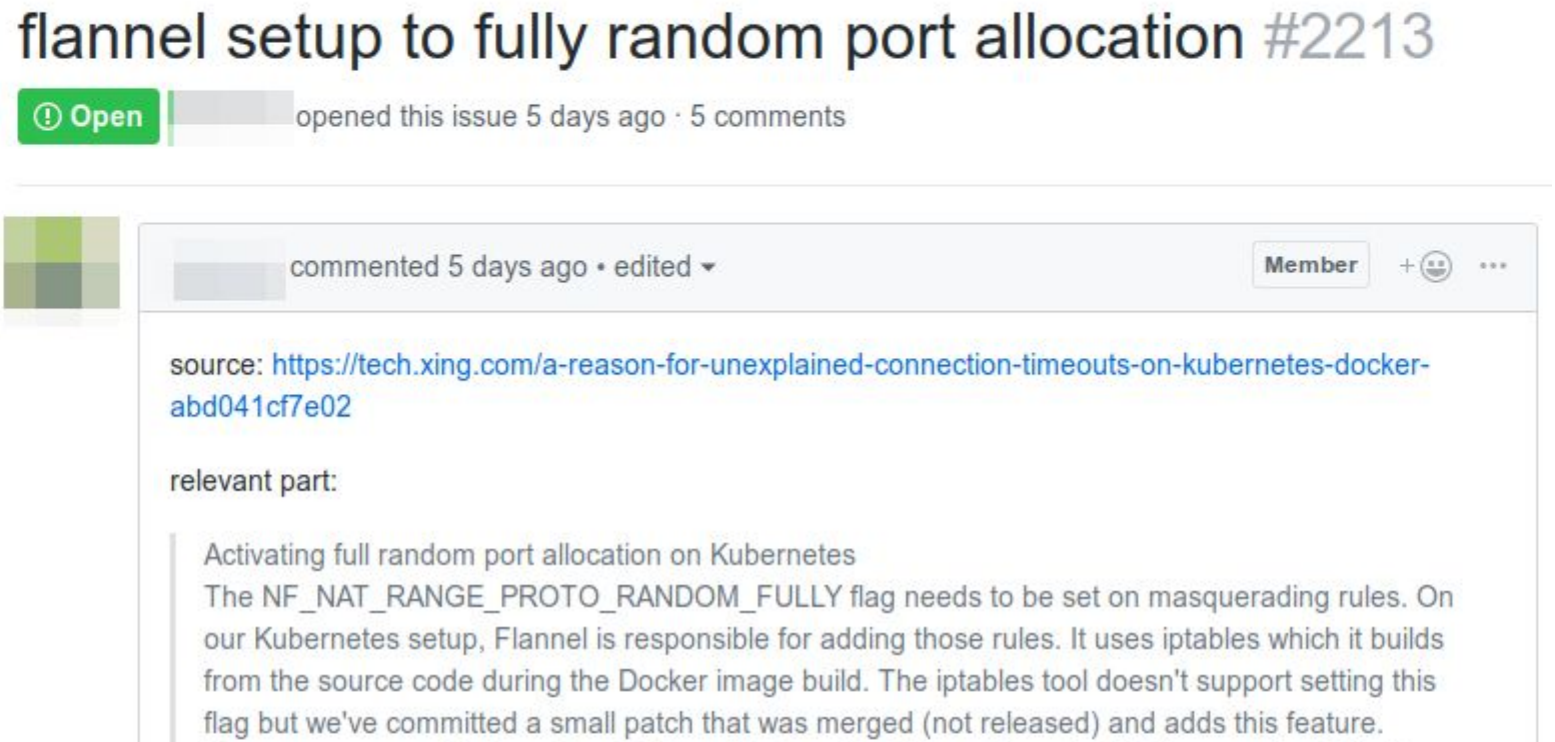
Kubernetes Failure Stories

A compiled list of links to public failure stories related to Kubernetes. Most recent publications on top.

- A Kubernetes crime story - Prezi - blog post 2019
  - involved: AWS EKS, SNAT, conntrack, Amazon VPC CNI plugin
  - impact: delay of 1-3 seconds for outgoing TCP connections
- Postmortem: New K8s workers unable to join cluster - FREE NOW - postmortem 2019
  - involved: AWS spot instances, kops, CentOS, container-selinux
  - impact: insufficient cluster capacity in testing environments, failed deployments in production and staging environments
- How a simple admission webhook lead to a cluster outage - Jetstack - blog post 2019
  - involved: ValidatingWebhookConfiguration, GKE node auto-repair
  - impact: prolonged downtime of non-prod environment, nodes lost, failed master upgrade
- Post Mortem: Kubernetes Node OOM - Blue Matador - blog post 2019
  - involved: AWS, SystemOOM, EBS, fluentd-sumologic, no resource requests/limits
  - impact: unknown, Pods killed
- Kubernetes' dirty endpoint secret and Ingress - Ravelin - blog post 2019
  - involved: GKE, Ingress, replication controller, SIGTERM, "graceful shutdown"
  - impact: occasional 502 errors
- How a Production Outage Was Caused Using Kubernetes Pod Priorities - Grafana Labs 2019
  - involved: Pod priorities
  - impact: cascading Pod evictions
- Moving to Kubernetes: the Bad and the Ugly - Xing - ContainerDays EU 2019
  - involved: nginx Ingress, network interrupts, conntrack, frozen CronJob, PLEG, stuck controllers
  - impact: lost requests, response time jumps, not ready nodes
- Kubernetes Failure Stories, or: How to Crash Your Cluster - Zalando - ContainerDays EU 2019
  - involved: AWS IAM, Kubelet, `--kube-api-qps` , Skipper-Ingress, AWS, `OOMKill` , CronJob, CoreDNS, CPU throttling
  - impact: build errors, production outages
- Build Errors of Continuous Delivery Platform - Zalando - postmortem 2019
  - involved: AWS IAM, Kubelet, kube2iam, `--kube-api-qps`
  - impact: build errors
- 10 Ways to Shoot Yourself in the Foot with Kubernetes, #9 Will Surprise You - Datadog - KubeCon Barcelona 2019
  - involved: CoreDNS, `ndots:5` , IPVS conntrack, `imagePullPolicy: Always` , DaemonSet, NAT instances, `latest` tag, API server `OOMKill` , kube2iam, cluster-autoscaler, PodPriority, audit logs, `spec.replicas` , AWS ASG rebalance, CronJob, Pod toleration, zombies, `readinessProbe.exec` , cgroup freeze, kubectl
  - impact: unknown, API server outage, pending pods, slow deployments
- How Spotify Accidentally Deleted All its Kube Clusters with No User Impact - Spotify - KubeCon Barcelona 2019
  - involved: GKE, cluster deletion, browser tabs, Terraform, global state file, git PRs, GCP permissions
  - impact: no impact on end users
- Kubernetes Failure Stories - Zalando - KubeCon Barcelona 2019
  - involved: Skipper-Ingress, AWS, `OOMKill` , high latency, CronJob, CoreDNS, `ndots:5` , etcd, CPU throttling
  - impact: multiple production outages
- Oh Sh*t! The Config Changed! - Pusher - KubeCon Barcelona 2019
  - involved: AWS, nginx, ConfigMap change
  - impact: production outage
- Misunderstanding the behaviour of one templating line - Skyscanner - blog post 2019
  - involved: HAProxy-Ingress, Service VIPs, Golang templating
  - impact: Significantly increased latency, 5XXs thrown from some services
- How to kill the Algolia dashboard during Black Friday or How not to use Kubernetes - Algolia - Algolia Search Party 2019
  - involved: GKE, Jobs, overload
  - impact: high latency, timeouts
- All pods scheduled to same failing host - Moonlight - postmortem 2019
  - involved: Google Kubernetes Engine, scheduler, anti-affinity rules
  - impact: major production outage, 100% traffic loss
- The shipwreck of GKE Cluster Upgrade - Loveholidays GKE - blog post 2019
  - involved: GCP Ingress, GKE Cluster, nodes
  - impact: critical drop in pod availability, loss of ingress, 2 hour maintenance lasting 7 hours
- Breaking Kubernetes: How We Broke and Fixed our K8s Cluster - Civis Analytics - blog post 2019
  - involved: AWS, kops, large clusters, batch jobs infrastructure, Datadog, API server, DNS, CPU throttling
  - impact: production outage
- Let's talk about Failures with Kubernetes - Zalando - Hamburg meetup 2019
  - involved: AWS, `NotReady` nodes, ELB dynamic IPs, Ingress, API server, CronJob, CoreDNS, `OOMKill` , kubelet memory leak, CPU throttling
  - impact: production outages
- Total DNS outage in Kubernetes cluster - Zalando - postmortem 2019
  - involved: AWS, DNS, CoreDNS, `OOMKill` , `ndots:5` , HTTP retries
  - impact: production outage
- Maximize learnings from a Kubernetes cluster failure - NU.nl - blog post 2019
  - involved: AWS, `NotReady` nodes, `SystemOOM` , Helm, ElastAlert, no resource limits set
  - impact: user experience affected for internally used tools and dashboards
- Kubernetes Load Balancer Configuration - Beware when draining nodes - DevOps Hof - blog post 2019
  - involved: GCP Load Balancer, `externalTrafficPolicy` , ingress-nginx
  - impact: total ingress traffic outage
- On Infrastructure at Scale: A Cascading Failure of Distributed Systems - Target - Medium post January 2019
  - involved: on-premise, Kafka, large cluster, Consul, Docker daemon, high CPU usage
  - impact: development environment outage
- How NOT to do Kubernetes - Sr.SRE Medya Ghazizadeh - Google - Cloud Native Meetup Sep 2018
  - involved: public container registery, ingress wild card, image size, replica count, 12factor
  - impact: security, stablity of clusters.
- Running Kubernetes in Production: A Million Ways to Crash Your Cluster - Zalando - DevOpsCon Munich 2018
  - involved: AWS, Ingress, CronJob, etcd, flannel, Docker, CPU throttling
  - impact: production outages
- Outages? Downtime? - Veracode - blog post 2018
  - involved: AWS, AWS IAM, region migration, kubespray, Terraform, pod CIDR
  - impact: QA/dev cluster outage
- NRE Labs Outage Post-Mortem - NRE Labs - blog post 2018
  - involved: GCP, kubeadm, etcd, Terraform, `livenessProbe`
  - impact: production outage
- A Perfect DNS Storm - Toyota Connected - blog post 2018
  - involved: Azure, DNS, `ndots:5` , Alpine musl libc
  - impact: DNS resolution failures
- Kubernetes and the Menace ELB, the tale of an outage - Turnitin - blog post 2018
  - involved: AWS, kube-aws, ELB dynamic IPs, API server, kubelet, `NotReady` nodes
  - impact: 15 minutes cluster outage
- Moving the Entire Stack to K8s Within a Year – Lessons Learned - ThredUP - DevOpsStage 2018
  - involved: AWS, kops, HAProxy, `livenessProbe` , DNS, too many open files
  - impact: unknown outages, DNS errors
- AirMap Platform Service Outage - AirMap - incident report 2018
  - involved: Azure, `NotReady` nodes, kubelet PLEG, CNI
  - impact: production AirMap platform outage
- Anatomy of a Production Kubernetes Outage - Monzo - KubeCon Europe 2018
  - involved: AWS, etcd, Linkerd, `NullPointerException` , gRPC client, services without endpoints, incompatible Kubernetes API change
  - impact: production ledger/platform outage
- Stories from the Playbook - Google - KubeCon Europe 2018
  - involved: GKE, etcd, Docker daemon, Image registry, dnsmasq vulnerabilities
  - impact: production outage
- 101 Ways to "Break and Recover" Kubernetes Cluster - Oath/Yahoo - KubeCon Europe 2018
  - involved: on-premise, namespace deletion, domain name collision, `NotReady` nodes, etcd empty dir, TLS certificate refresh, DNS issues, OOM
  - impact: unknown cluster outages
- Experiences with running PostgreSQL on Kubernetes - Gravitational - blog post 2018
  - involved: PostgreSQL, streaming replication
  - impact: data loss
- 101 Ways to Crash Your Cluster - Nordstrom - KubeCon North America 2017
  - involved: AWS, `NotReady` nodes, OOM, eviction thresholds, ELB dynamic IPs, kubelet, cluster autoscaler, etcd split
  - impact: full production cluster outage, other outages
- Major Outage: Current account payments may fail - Monzo - Monzo Community post 2017
  - involved: AWS, etcd, Linkerd, `NullPointerException` , services without endpoints
  - impact: major production outage, full platform outage, current account payments fail
- Fallacies of Distributed Computing with Kubernetes on AWS - Zalando - AWS User Group Hamburg October 2017
  - involved: AWS, unhealthy nodes, Ingress, CronJob
  - impact: production outage
- Search and Reporting Outage - Universe - incident report 2017
  - involved: Job, `RestartPolicy` , consume node resources
  - impact: production Universe search and reporting outage
- Our First Kubernetes Outage - Saltside - blog post 2017
  - involved: AWS, kops, Helm, `NotReady` nodes, resource exhaustion
  - impact: nonproduction cluster outage
- Our Failure Migrating to Kubernetes - Saltside - blog post 2017
  - involved: AWS, kops, ELB, `BackendConnectionErrors` , `LoadBalancer` service
  - impact: aborted application migration
- SaleMove US System Issue - SaleMove - incident report 2017
  - involved: AWS, ELB dynamic IPs, DNS `A` record for master, API server
  - impact: production issues with SaleMove US System

42 stories so far

INTERNAL TICKETS BASED ON FAILURE STORIES



flannel setup to fully random port allocation #2213

[redacted] opened this issue 5 days ago · 5 comments

[redacted] commented 5 days ago • edited

source: https://tech.xing.com/a-reason-for-unexplained-connection-timeouts-on-kubernetes-docker-abd041cf7e02

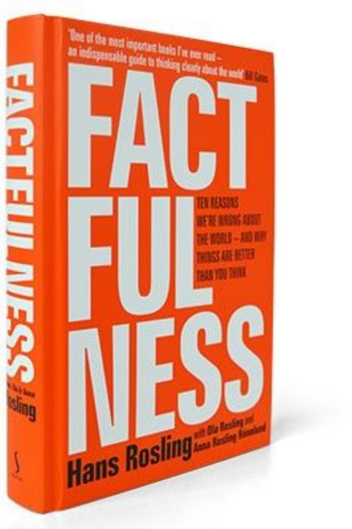
relevant part:

> Activating full random port allocation on Kubernetes
> The NF_NAT_RANGE_PROTO_RANDOM_FULLY flag needs to be set on masquerading rules. On our Kubernetes setup, Flannel is responsible for adding those rules. It uses iptables which it builds from the source code during the Docker image build. The iptables tool doesn't support setting this flag but we've committed a small patch that was merged (not released) and adds this feature.

FACTFULNESS

Things can be both better and bad!

How would failure stories for your non-K8s infra look like?


'One of the most important books I've ever read – an indispensable guide to thinking clearly about the world' Bill Gates
FACT FUL NESS
TEN REASONS WE'RE WRONG ABOUT THE WORLD – AND WHY THINGS ARE BETTER THAN YOU THINK
Hans Rosling
with Ola Rosling and Anna Rosling Rönnlund
FACTFULNESS

https://k8s.af

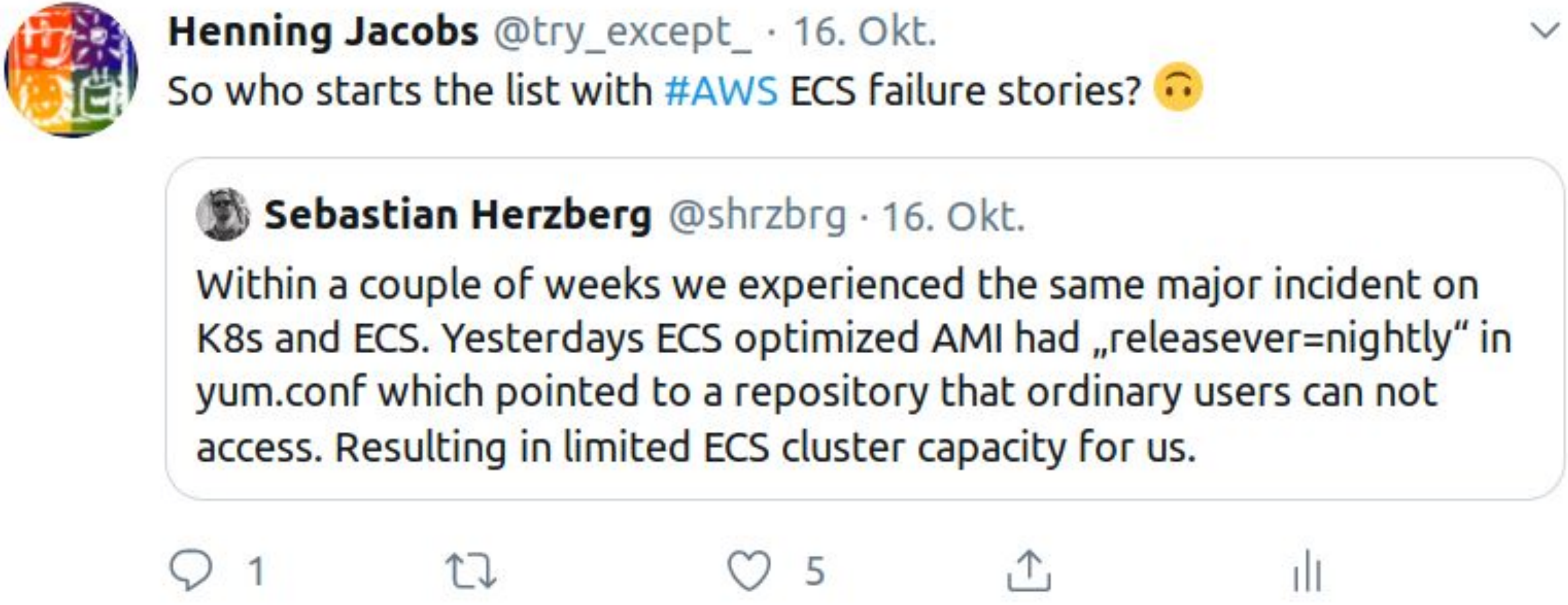
"I'M TORN ON THIS LIST"

It is important to learn from the mistakes from others, so I like it [...], BUT it feels like folks are **using these examples as reasons to stay away from Kubernetes**.

There could be a [...] **larger list of system failures where K8s is not involved**. Similarly there could [...] be a list of "Encryption Failure Stories" but that doesn't mean we shouldn't encrypt things. [...]

https://news.ycombinator.com/item?id=20168521

FAILURE STORIES CAN BE FOUND EVERYWHERE

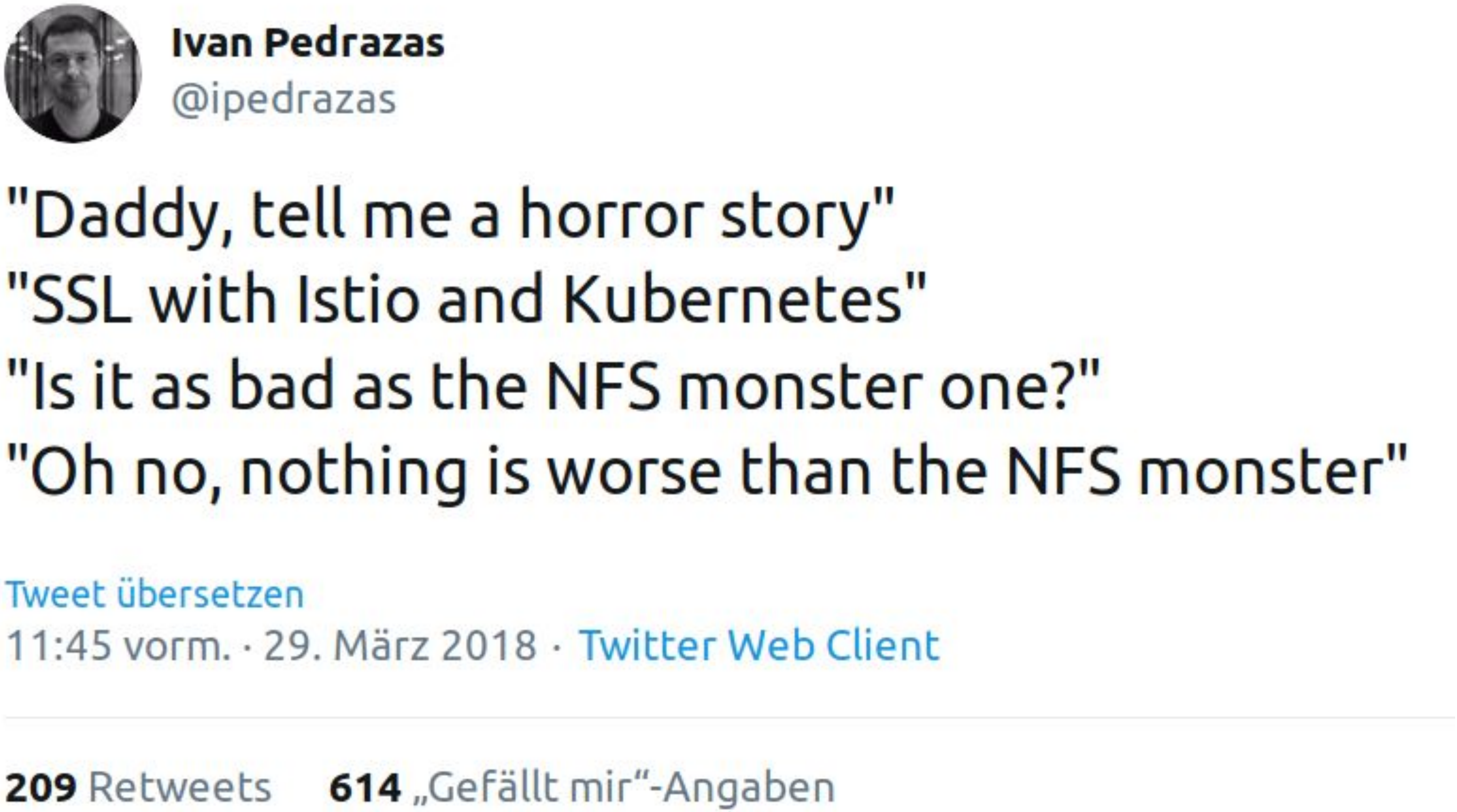


Henning Jacobs @try_except_ · 16. Okt.

So who starts the list with #AWS ECS failure stories? 🙃

> **Sebastian Herzberg** @shrzbrg · 16. Okt.
>
> Within a couple of weeks we experienced the same major incident on K8s and ECS. Yesterdays ECS optimized AMI had „releasever=nightly“ in yum.conf which pointed to a repository that ordinary users can not access. Resulting in limited ECS cluster capacity for us.

zalando



Ivan Pedrazas
@ipedrazas

"Daddy, tell me a horror story"
"SSL with Istio and Kubernetes"
"Is it as bad as the NFS monster one?"
"Oh no, nothing is worse than the NFS monster"

11:45 vorm. · 29. März 2018 · Twitter Web Client

209 Retweets **614** „Gefällt mir“-Angaben

WHY KUBERNETES?

WHY KUBERNETES?

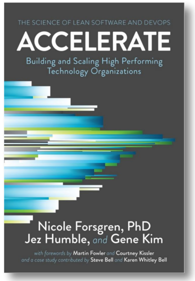
- provides enough **abstractions** (StatefulSet, CronJob, ..)
- provides **consistency** (API spec/status)
- is **extensible** (annotations, CRDs, API aggreg.)
- certain **compatibility** guarantee (versioning)
- widely **adopted** (all cloud providers)
- works across environments and implementations

srcco.de/posts/why-kubernetes.html

WHY KUBERNETES?

(for Zalando)

- Efficiency
- Common Operational Model
- Developer Experience
- Cloud Provider Independent
- Compliance and Security
- Talent



THE SCIENCE OF LEAN SOFTWARE AND DEVOPS

ACCELERATE

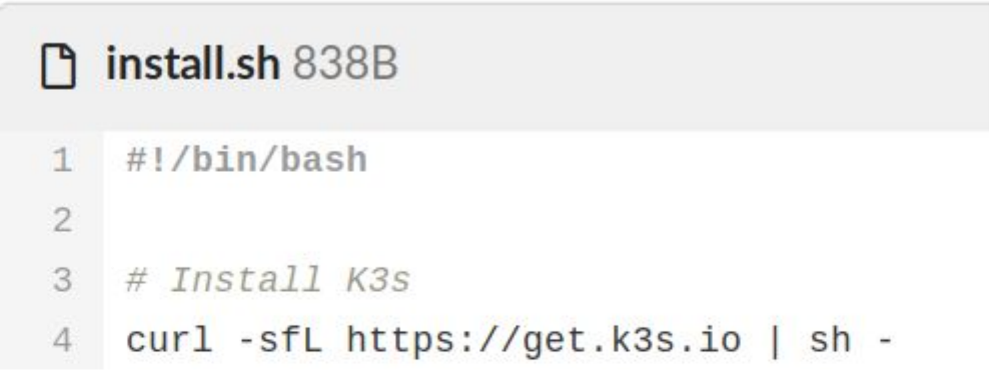
Building and Scaling High Performing Technology Organizations

Nicole Forsgren, PhD
Jez Humble, and Gene Kim

with forewords by Martin Fowler and Courtney Kissler
and a case study contributed by Steve Bell and Karen Whitley Bell

COMPLEXITY FOR GOOGLE-SCALE INFRA?

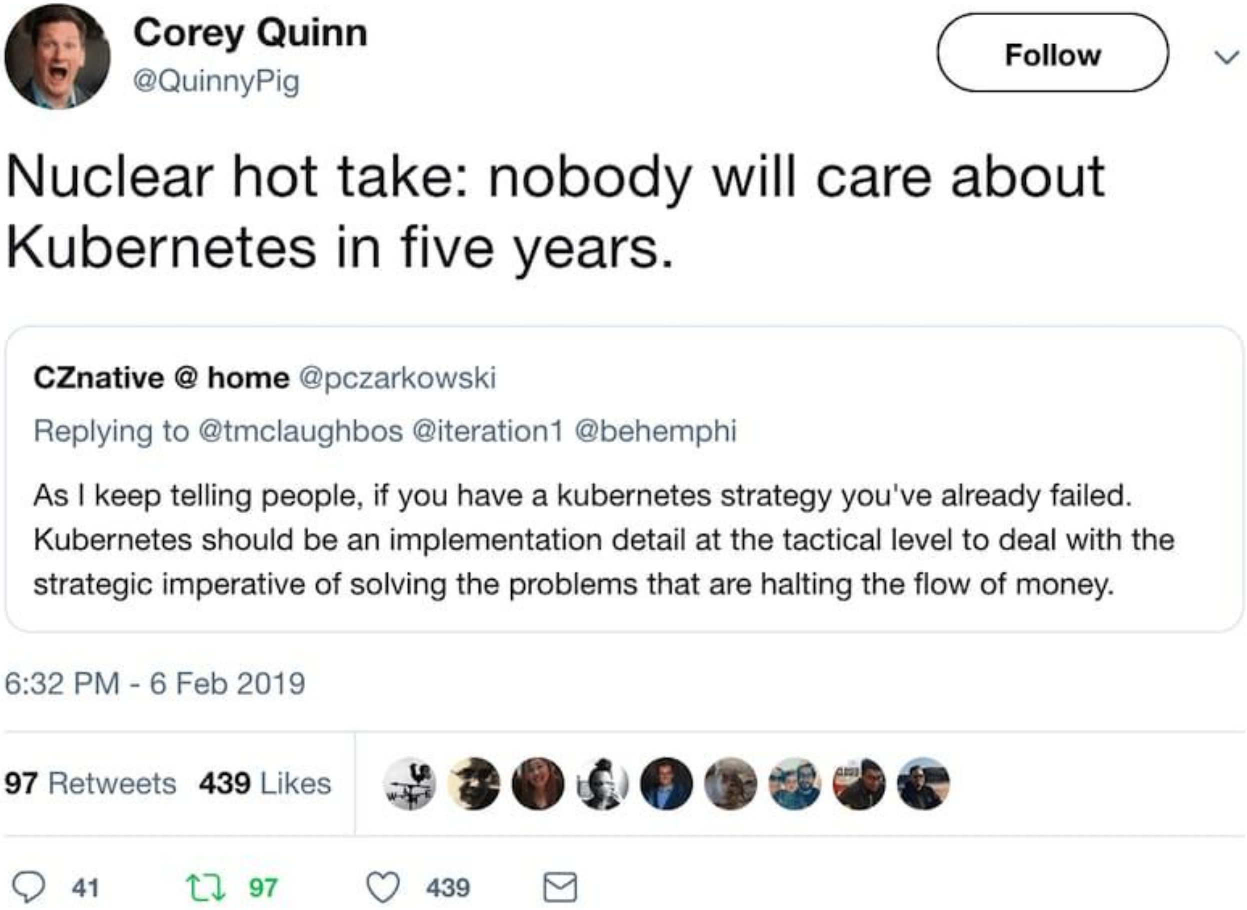
- Managed DO cluster: 4 minutes
- K3s single node: 2 minutes



install.sh 838B

```
#!/bin/bash

# Install K3s
curl -sfL https://get.k3s.io | sh -
```

demo.j-serv.de



Corey Quinn
@QuinnyPig

Nuclear hot take: nobody will care about Kubernetes in five years.

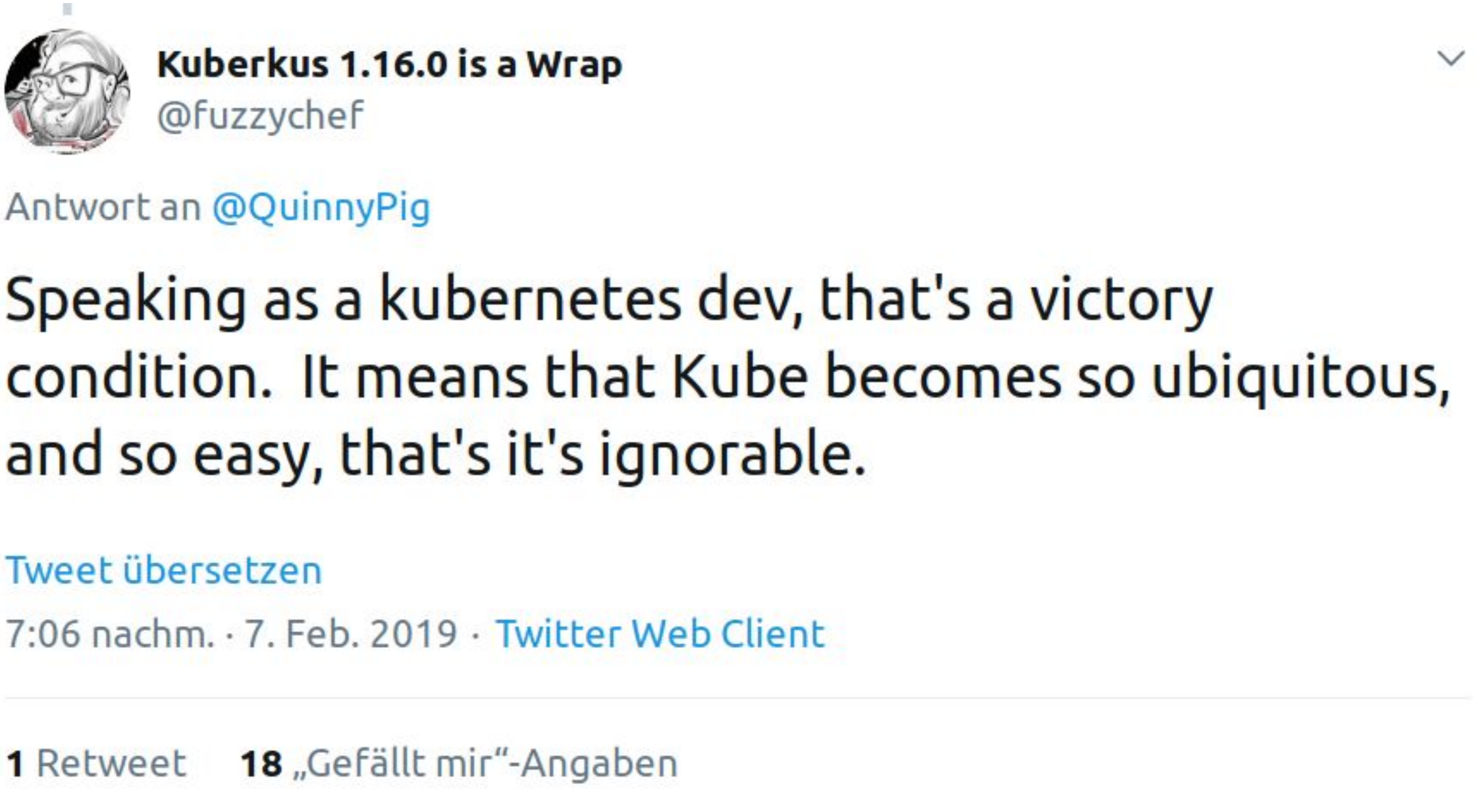
CZnative @ home @pczarkowski
Replying to @tmclaughbos @iteration1 @behemphi

As I keep telling people, if you have a kubernetes strategy you've already failed. Kubernetes should be an implementation detail at the tactical level to deal with the strategic imperative of solving the problems that are halting the flow of money.

6:32 PM - 6 Feb 2019

97 Retweets **439** Likes

MAYBE THAT'S GOOD?



Kuberkus 1.16.0 is a Wrap
@fuzzychef

Antwort an @QuinnyPig

Speaking as a kubernetes dev, that's a victory condition. It means that Kube becomes so ubiquitous, and so easy, that's it's ignorable.

Tweet übersetzen

7:06 nachm. · 7. Feb. 2019 · Twitter Web Client

1 Retweet **18** „Gefällt mir"-Angaben

OPEN SOURCE & MORE

Kubernetes Web View
codeberg.org/hjacobs/kube-web-view

Skipper HTTP Router & Ingress controller
github.com/zalando/skipper

Kubernetes Janitor
github.com/hjacobs/kube-janitor

Postgres Operator
github.com/zalando-incubator/postgres-operator

More Zalando Tech Talks
github.com/zalando/public-presentations

QUESTIONS?

HENNING JACOBS

SENIOR PRINCIPAL

henning@zalando.de

@try_except_

Illustrations by @01k

Please

Remember to rate this session

Thank you!