

# Humane on Call: 24x7 support without losing it

*Mario Fernandez, Cecilia Hwang, Shahid  
Aslam, Donald Dragoti*

**ThoughtWorks®**

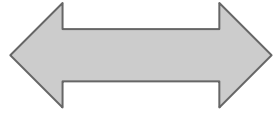
# Introduction

On Call?

Somebody is **always** available for support

24x7

TWer



TW



Client

# Flow

*An on-call rotation that is fit for human consumption*

# What's a rotation?



- One developer at-a-time
- Weekly rotations
- 24-hour availability



# What happens during a rotation?



- Receive alerts
- Respond to alerts within 30 minutes
- Work on incidents
- Close alerts

# What is an incident?



CloudWatch alerts:

- Elasticsearch free storage space low

Prometheus alerts:

- Kube pod crash looping
- Kube node unreachable
- Target down
- Kube daemonset rollout stuck
- ...

User or 2nd Level support created incidents

# Tooling

*The right tool for the job*

# PagerDuty

**PagerDuty**btcmvc.pagerduty.com

Hello Shahid Aslam, you have one open incident assigned to you:

INCIDENT #26781

**[FIRING:1] (ServiceNotAccessible)**

[View Incident](#)

DETAILS

resolved:

num\_resolved: 0

num\_firing: 1

firing: Labels:

- alertname = ServiceNotAccessible

- instance =

- job = blackbox-exporter

- prometheus = monitoring/k8s

- severity = warning

- type =

Annotations:

- summary = Endpoint not accessible for last 90s

Source: [http://prometheus-k8s-0.308u/graph/gu.expr=probe\\_success+%3D%3D+0&g0.tab=1](http://prometheus-k8s-0.308u/graph/gu.expr=probe_success+%3D%3D+0&g0.tab=1)

STATUS

Triggered

URGENCY

↑ High

ASSIGNED TO

[Shahid Aslam](#)

OPENED ON

Apr 04, 2020 at 8:16 PM (Berlin)

SERVICE

[Daytona \[Platform\] Endpoints](#)

Managed by Terraform

ESCALATION POLICY

[Daytona \[Platform\] Escalation Policy](#)

You can disable these messages (or turn off HTML emails) by adjusting [your notification rules](#).

# OpsGenie

Opsgenie Alerts Incidents Who is on-call Teams Services Analytics Settings


   CH

 Your Opsgenie admin has started connecting your account to an Atlassian account. [Finish setting up your Atlassian account.](#)

## Alerts

Create alert ...

{q} ? Search Save ...

☐ Select < Feb 13, 2021 9:08 PM - Mar 15, 2021 9:08 PM > Last Month 

### Saved searches

PREDEFINED

All

Open





Closed

Un'Acked

Not seen

Assigned to me

SHARED WITH ME

<input type="checkbox"/>	#5216	P1	ALARM: "elasticsearch-storage-for-ddc-prod" in EU (Frankfurt)	Shahid Aslam	ACK'ED
	x1		Orbit	Close	Assign ...
				Mar 15, 2021 7:47 PM (GMT+01:00)	
<input type="checkbox"/>	#5178	P1	[FIRING:1] kube-state-metrics critical (KubePodCrashLooping aws-cluster-autoscaler-chart http prod 10.20.33.66:8080 daytona aut...		CLOSED
	x1		Orbit		
				Mar 15, 2021 6:00 PM (GMT+01:00)	
<input type="checkbox"/>	#5170	P1	[FIRING:1] kube-state-metrics critical (KubeDeploymentReplicasMismatch autoscaler-aws-cluster-autoscaler-chart http prod 10.20....		CLOSED
	x1		Orbit		
				Mar 15, 2021 5:46 PM (GMT+01:00)	
<input type="checkbox"/>	#4855	P3	[FIRING:1] fluentd warning (FluentdOutputError metrics test 10.6.32.86:24231 daytona out_cloudwatch_logs		CLOSED
	x1		Orbit		



# Features that makes us happy

# Auto-closing



E.g., pod not healthy alert  
Then pod becomes ready  
And alert closes itself

# Grouping



Different grouping options makes life easier.

- OpsGenie
- AlertManager



# Priorities

P1 = production  
P3 = non-production



# Notification escalation policies



A background image of a man with a beard and glasses, wearing a ThoughtWorks lanyard, smiling and interacting with a woman at a conference. The image is overlaid with a semi-transparent purple filter.

# Substandard tools

# Internal Tool Sadist Made



- No auto-closing
- No grouping
- No notification escalation policies

# Friction

*Every little problem looks way worse at 3am*

Tell me what is wrong!



#1855



[FIRING:1] kube-state-metrics warning (KubeNodeUnreachable NoSchedule http int 10.5.41.114:8080

Mario Fernandez

CLOSED

x1

node.kubernetes.io/unreachable ...



Orbit

Feb 17, 2021 5:59 PM (GMT+01:00)

# Spurious Alerts



P3

Apr 24, 2020 12:10 PM (GMT+01:00) - Mario Fernandez

x1

#4556

[+ Add tag](#)**[FIRING:3] node-exporter warning (NodeClockSkewDetected) metrics prod daytona-metrics space-lab daytona-metrics/prometheus-opera...**

CLOSED

Details

Activity log

Responder states

Source

<http://prometheus-operator-alertmanager.daytona-metrics:9093/#/alerts?receiver=daytona-platform-opsg>

ELAPSED TIME

Integration

Orbit API space-lab prod eu-central-1 (API)

0h 15m 39s

Responders

[Orbit](#)

Owner Team

[Orbit](#)

Alias

7f9aefc8ab57616e3be9b63902d8b4dbc2c408f9154d98d9272b905f4c7a4415

Last Updated At

Apr 24, 2020 12:26 PM (GMT+01:00)

Description

<https://github.com/kubernetes-monitoring/kubernetes-mixin/tree/master/runbook.md#alert-name-nodclockskewdetected> Clock skew detected.

Alerts Firing:

Labels:

- alertname = **NodeClockSkewDetected**
- endpoint = metrics
- environment = prod
- instance = 10.0.36.183:9100
- job = node-exporter
- namespace = daytona-metrics
- pod = prometheus-operator-prometheus-node-exporter-wh7kk
- product = space-lab
- prometheus = daytona-metrics/prometheus-operator-prometheus
- region = eu-central-1
- service = prometheus-operator-prometheus-node-exporter
- severity = warning

Annotations:

- message = Clock on 10.0.36.183:9100 is out of sync by more than 300s. Ensure NTP is configured correctly on this host.
- runbook\_url = <https://github.com/kubernetes-monitoring/kubernetes-mixin/tree/master/runbook.md#alert-name-nodclockskewdetected>
- summary = Clock skew detected.

# Continuous Tuning

## Graph

[View in metrics](#)

## KibanaHealthyNodes

OK

KibanaHealthyNodes &lt;= 1 for 6 datapoints within 18 minutes

Count



KibanaHealthyNodes

## Details

Name

kibana-health-check-space-lab-test

State

OK

Namespace

AWS/ES

Datapoints to alarm

6 out of 6

Type

Metric alarm

Threshold

KibanaHealthyNodes &lt;= 1 for 6 datapoints within 18 minutes

Metric name

KibanaHealthyNodes

Missing data treatment

Treat missing data as bad (breaching threshold)

Description

This metric is a health check for Kibana nodes for space-lab-test domain

Last change

2021-02-16 13:31:01

DomainName

space-lab-test

ClientId

[REDACTED]

Statistic

Sum

Period

180

Percentiles with low samples

evaluate

ARN

[REDACTED]

# Understanding

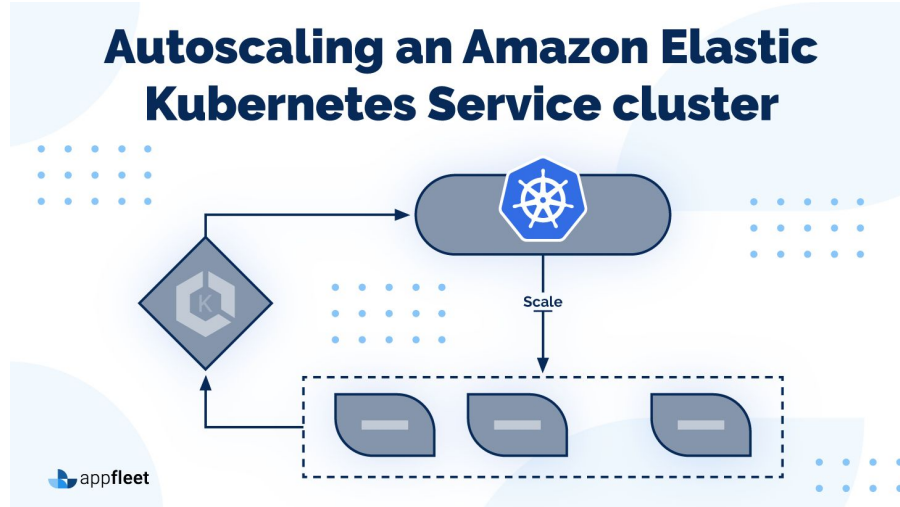
*Silos will literally keep you awake*

The one thing you don't really understand



Complexity hinders our ability to reason  
about systems

# Autoscaling an Amazon Elastic Kubernetes Service cluster





# You build it you run it

*Or, you find somebody to blame instead*

Story from the trenches

Delegating responsibility

Efficiently assigning blame

**24x7 is a  
state of mind**

