

K8s is Not Your Platform, It's Just the Foundation



Manuel Pais
co-author of *Team Topologies*

QCon London 2020

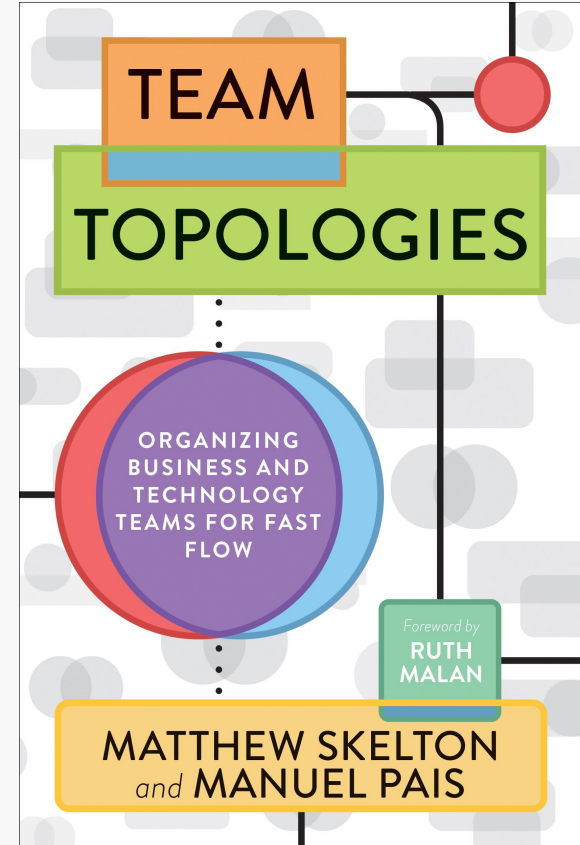
Team Topologies

Organizing business and technology teams for fast flow

Matthew Skelton & Manuel Pais

IT Revolution Press (2019)

<https://teamentopologies.com>





Is Kubernetes a Platform?



Team Cognitive Load

Team Interactions



Getting Started



Is Kubernetes a Platform?

DEVOPS

The next QCon is in [San Francisco, November 11-15, 2019](#). Save an extra \$100 with INFOQ100!

How Airbnb Simplified the Kubernetes Workflow for 1000+ Engineers



LIKE



DISCUSS



BOOKMARKS



MAR 08, 2019 • 3 MIN READ

by

Manuel Pais

FOLLOW

[Melanie Cebula](#), infrastructure engineer at Airbnb, gave a [talk at QCon London](#) [\[slides PDF\]](#) about the internal tooling and strategies Airbnb adopted to support over 1000 engineers concurrently configuring and deploying over 250 critical services to Kubernetes (at a frequency of about 500 deploys per day on average). One key enabler was a layer of abstraction and generation of Kubernetes configuration from higher level primitives using standardized environments and namespaces (and automated validations whenever possible). Also critical was the automation of common workflows for engineers and using the same tools across all environments.

Source: <https://www.infoq.com/news/2019/03/airbnb-kubernetes-workflow>
Melanie's talk: <https://www.infoq.com/presentations/airbnb-kubernetes-services>

DEVOPS

The next QCon is in San Francisco, November 11-15, 2019. Save an extra \$100 with INFOQ100!

How Airbnb Simplified the Kubernetes Workflow for 1000+ Engineers

 LIKE
  DISCUSS
 
 BOOKMARKS
 

MAR 08, 2019 • 3 MIN READ



by

Manuel Pais

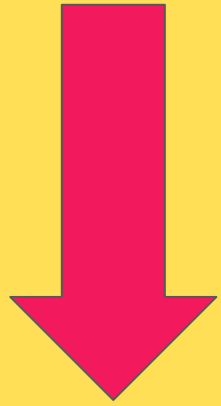
FOLLOW

[Melanie Cebula](#), infrastructure engineer at Airbnb, gave a [talk at QCon London](#) ([slides](#) [PDF](#)) about the internal tooling and strategies Airbnb adopted to support over 1000 engineers concurrently configuring and deploying over 250 critical services to Kubernetes (at a frequency of about 500 deploys per day on average). One key enabler was a layer of abstraction and generation of Kubernetes configuration from higher level primitives using standardized environments and namespaces (and automated validations whenever possible). Also critical was the automation of common workflows for engineers and using the same tools across all environments.

The screenshot shows the Kibana search interface. At the top, the search bar contains the query: `time must field : @timestamp from : "2019-03-07T11:58:48.402Z" to : "2019-03-15T11:59:21.781Z"`. The 'from' and 'to' date filters are circled in yellow. Below the search bar, the results table is displayed with the following columns: Value, Count, Author, Published date, and Action. The first row of results is circled in yellow, showing the title 'How Airbnb Simplified the Kubernetes Workflow for 1000+ Engineers' with a count of 23401, author 'Manuel Pais', and published date '2019/03/08 16:30:00'.

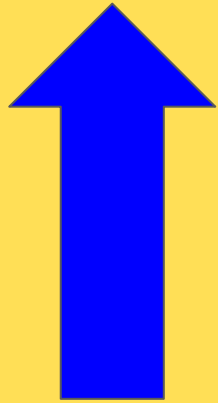
Value	Count	Author	Published date	Action
How Airbnb Simplified the Kubernetes Workflow for 1000+ Engineers	23401	Manuel Pais	2019/03/08 16:30:00	 

Kubernetes “platform”



**microservices
ops complexity**

Kubernetes “platform”



**deploy & run
abstractions**



Still need to...

... sizing hosts

... create/destroy clusters

... update to new K8s versions

... decide on namespaces vs clusters

<insert your fav chore here>

Still need to...

... sizing hosts

... create/destroy clusters

... update to new K8s versions

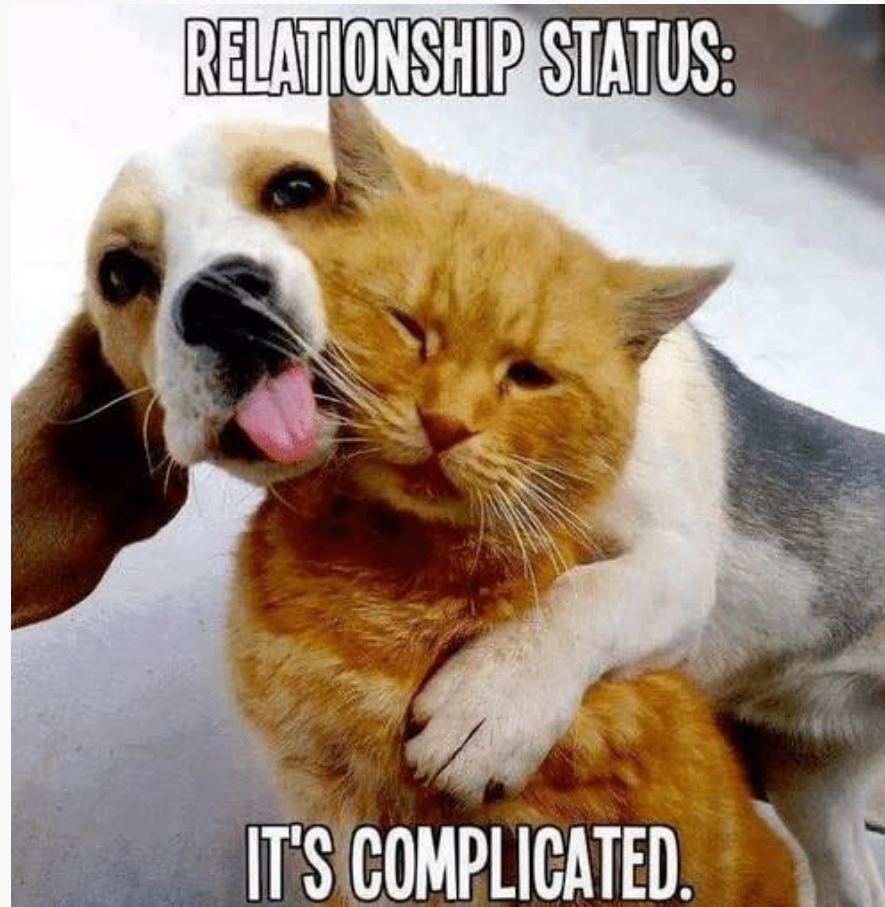
... decide on namespaces vs clusters

worry about security

Who is the provider?

Who is the provider?

Who is the consumer?



“A digital platform is a foundation of self-service APIs, tools, services, knowledge and support which are arranged as a compelling internal product.”

– Evan Bottcher, 2018



*“A digital platform is a foundation of self-service APIs, tools, services, **knowledge and support** which are arranged as a compelling internal product.”*

– Evan Bottcher, 2018



*“A digital platform is a foundation of self-service APIs, tools, services, knowledge and support which are arranged as a **compelling internal product.**”*

– Evan Bottcher, 2018





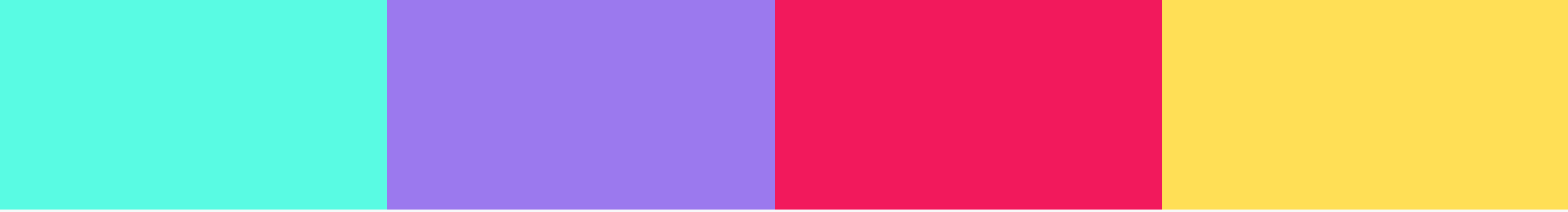
***Kubernetes is not your
platform. It's the foundation.***



“Create a **path of least resistance.**

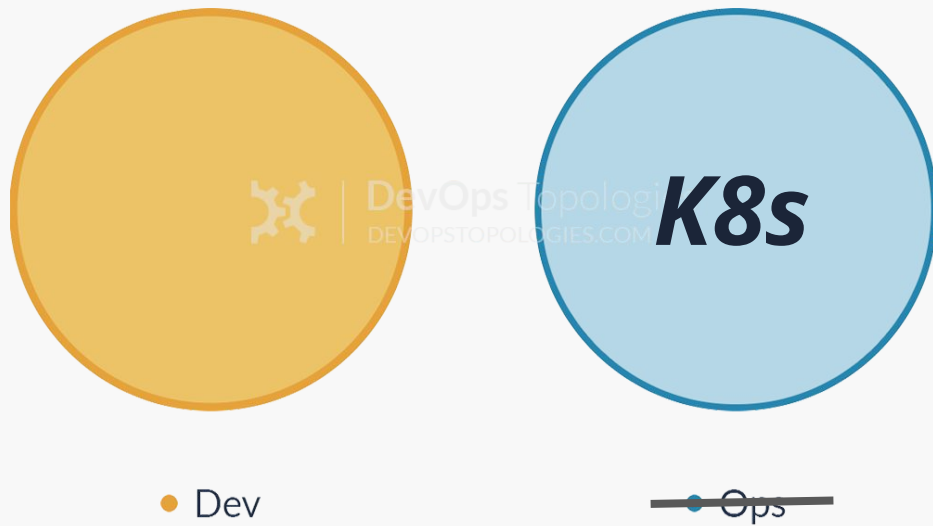
Make the right thing the easiest thing to do.”

– Evan Bottcher, 2018



***The hard thing about
platforms is to constantly
evolve & adapt to
new & old customers.***





Team Cognitive Load

“Cognitive load is the total amount of mental effort being used in the working memory”

- John Sweller

Intrinsic

Extraneous

Germane

“How are
classes
defined in
Java?”

Intrinsic

Extraneous

Germane

“How do I
deploy this
app,
again?”

Intrinsic
Extraneous
Germane

“How do
bank
transfers
work?”

Intrinsic *(skills)*

Extraneous *(mechanics)*

Germane *(domain focus)*



Intrinsic *(skills)*

Extraneous *(mechanics)*

Germane *(domain focus)*

 ***Intrinsic*** *(skills)*

 ***Extraneous*** *(mechanics)*

Germane *(domain focus)*

 ***Intrinsic*** (skills)

 ***Extraneous*** (mechanics)

 ***Germane*** (domain focus)

More: '*Hacking Your Head*'

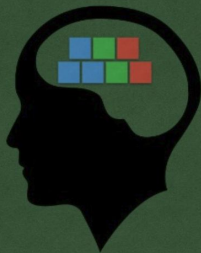
Jo Pearce

(@jdpearce)

Hacking Your Head : Managing Information Overload (extended)

Hacking Your Head : Managing Information Overload

Total cognitive load is comprised of *three* types:

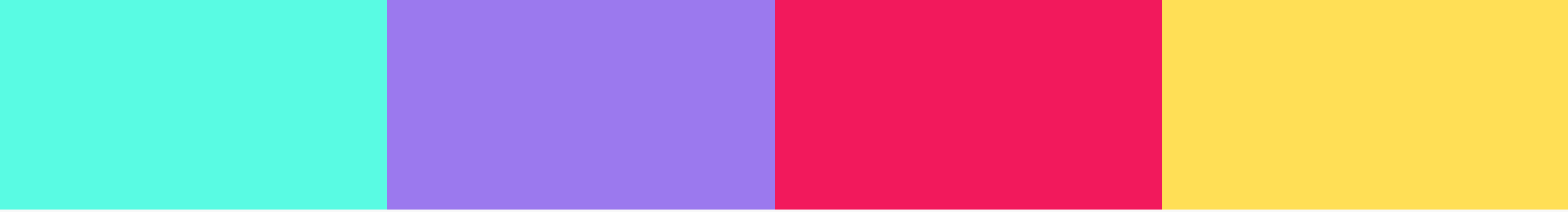


- Intrinsic Load
- Extraneous (Irrelevant) Load
- Germane (Relevant) Load

@jdpearce jopearce.co.uk

11 of 12

<https://www.slideshare.net/JoPearce5/hacking-your-head-managing-information-overload-extended>



***Be mindful of your
platform choices' impact
on teams' cognitive load***



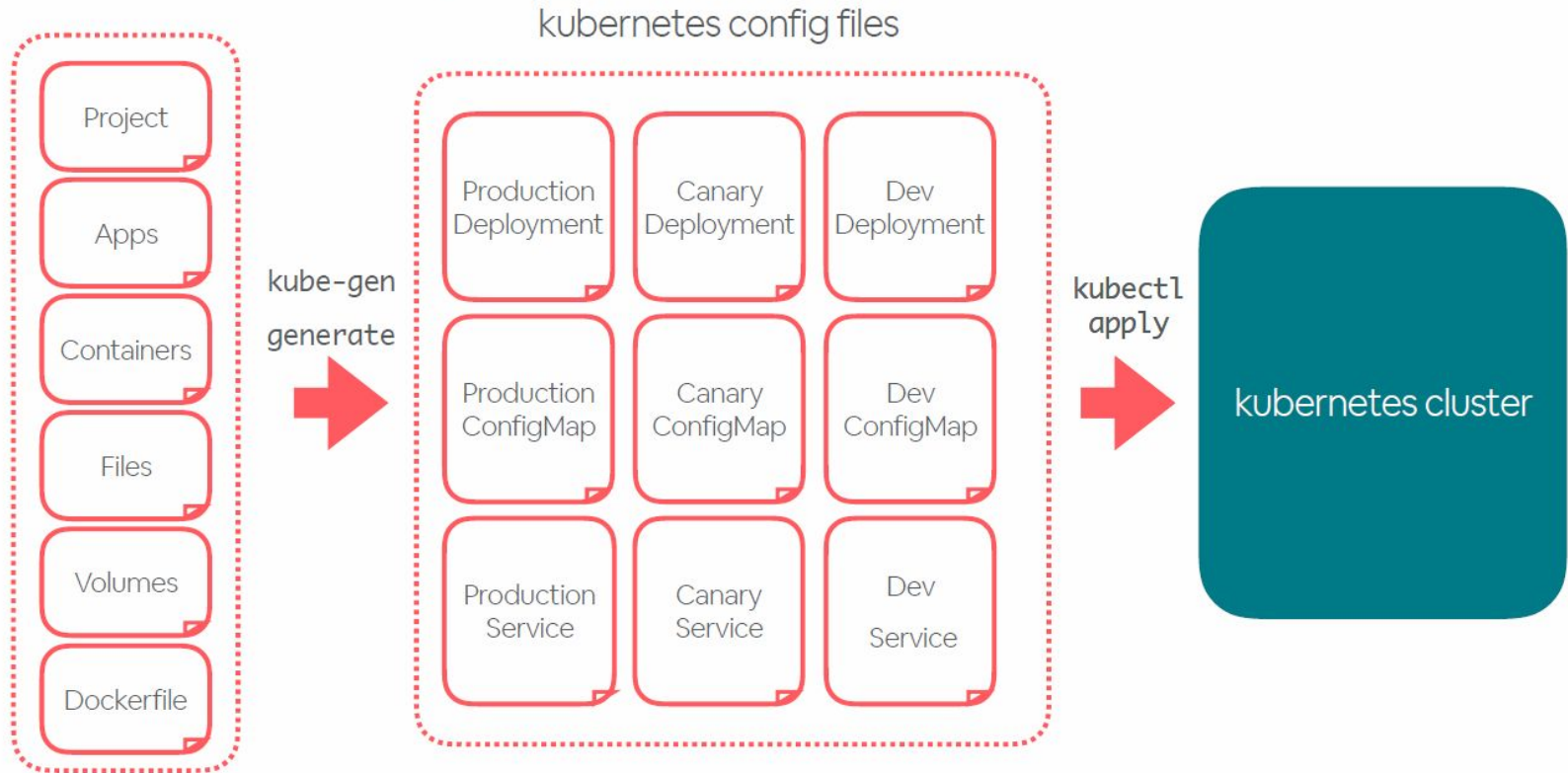
“The best part of my day is when I update 10 different YAML files to deploy a one-line code change.”

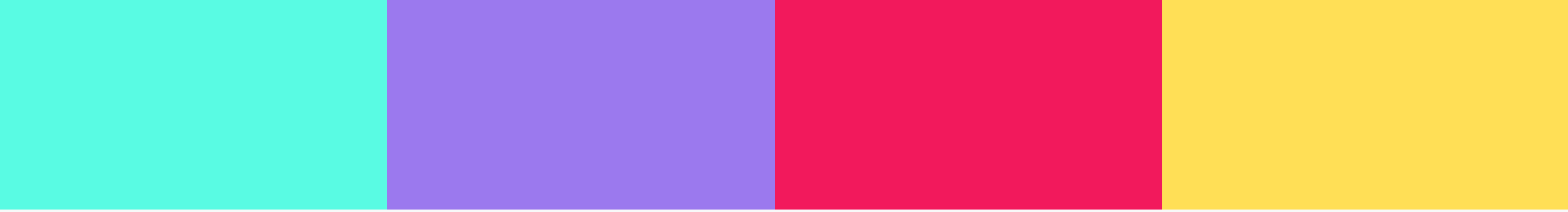
“The best part of my day is when I update 10 different YAML files to deploy a one-line code change.”

– No One, Ever

generating k8s configs

@MELANIECEBULA

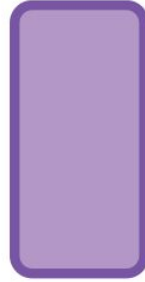




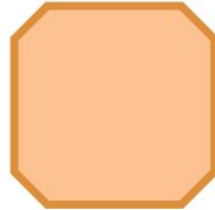
Clarify (platform) service boundaries and provide abstractions to reduce the cognitive load on teams.



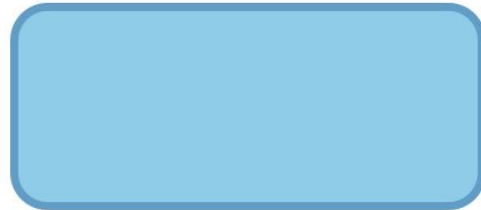
Stream-aligned
team



Enabling team



Complicated
subsystem team



Platform team



Stream-aligned
team



Platform team

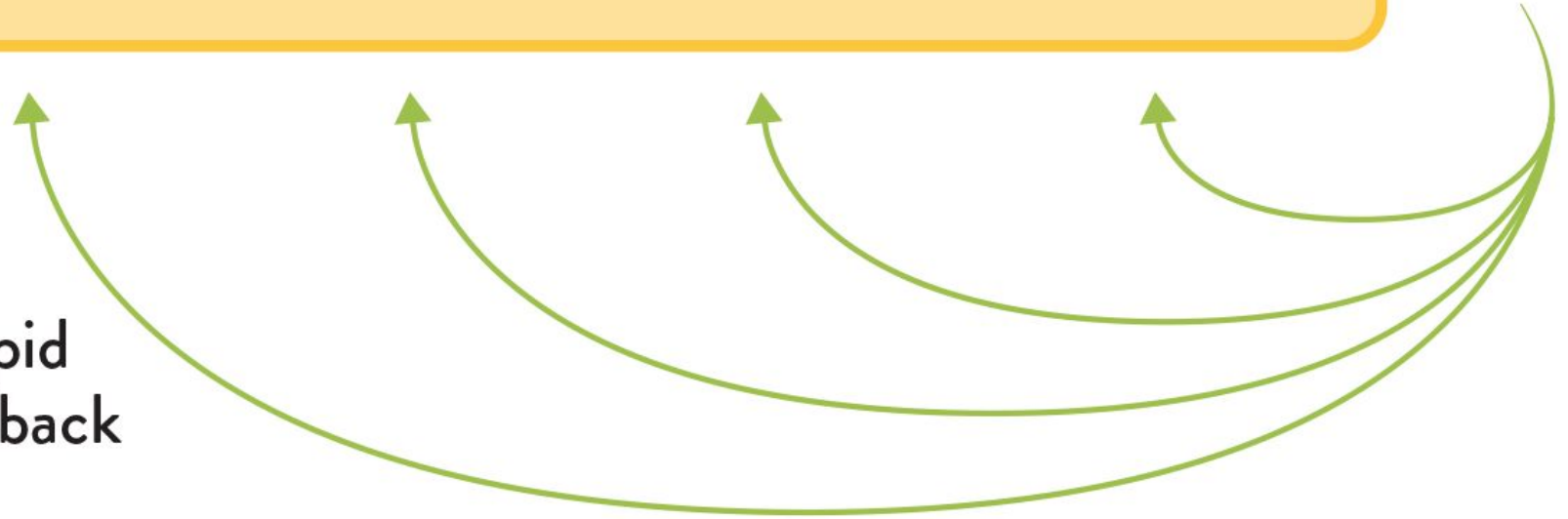




Product Team

LIVE

**Rapid
Feedback**

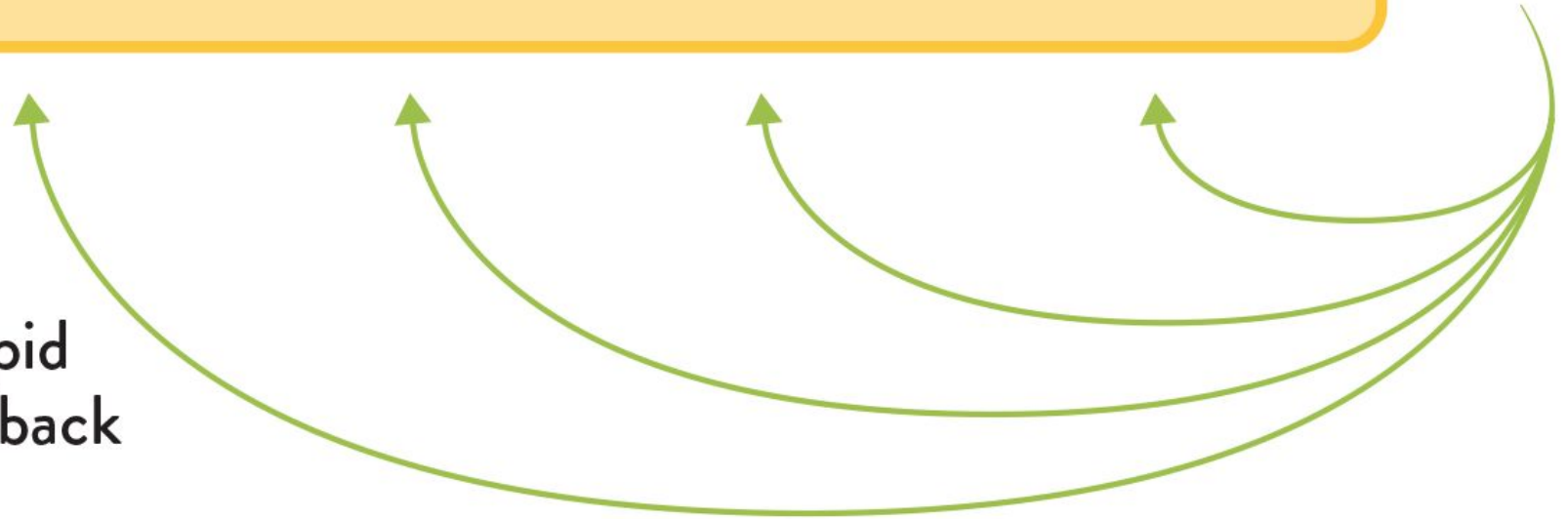




Stream-Aligned Team

LIVE

**Rapid
Feedback**





U switch

RV

Convergence to Kubernetes

Standardisation to Scale

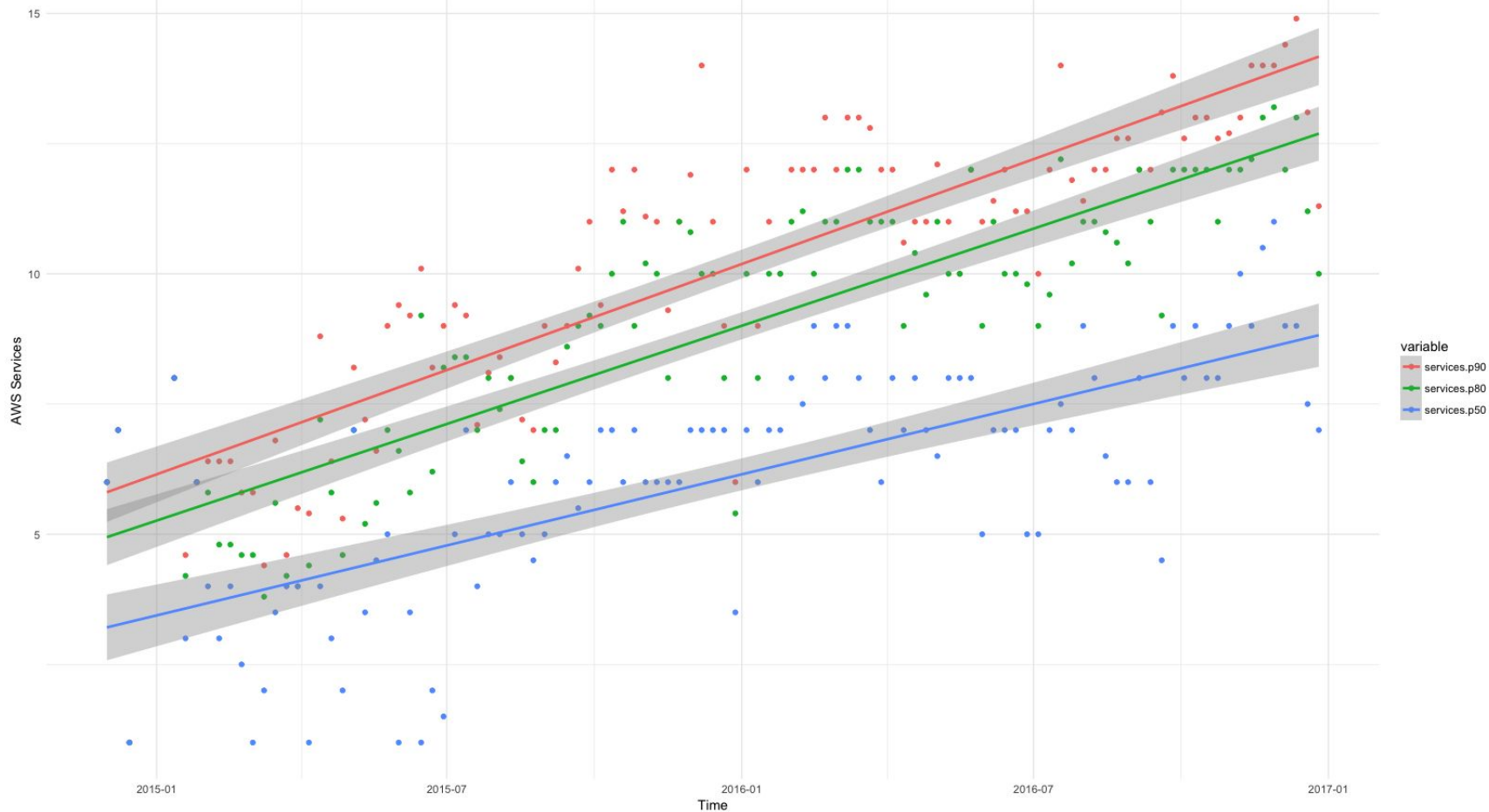


Paul Ingles

Follow

Jun 18, 2018 · 13 min read

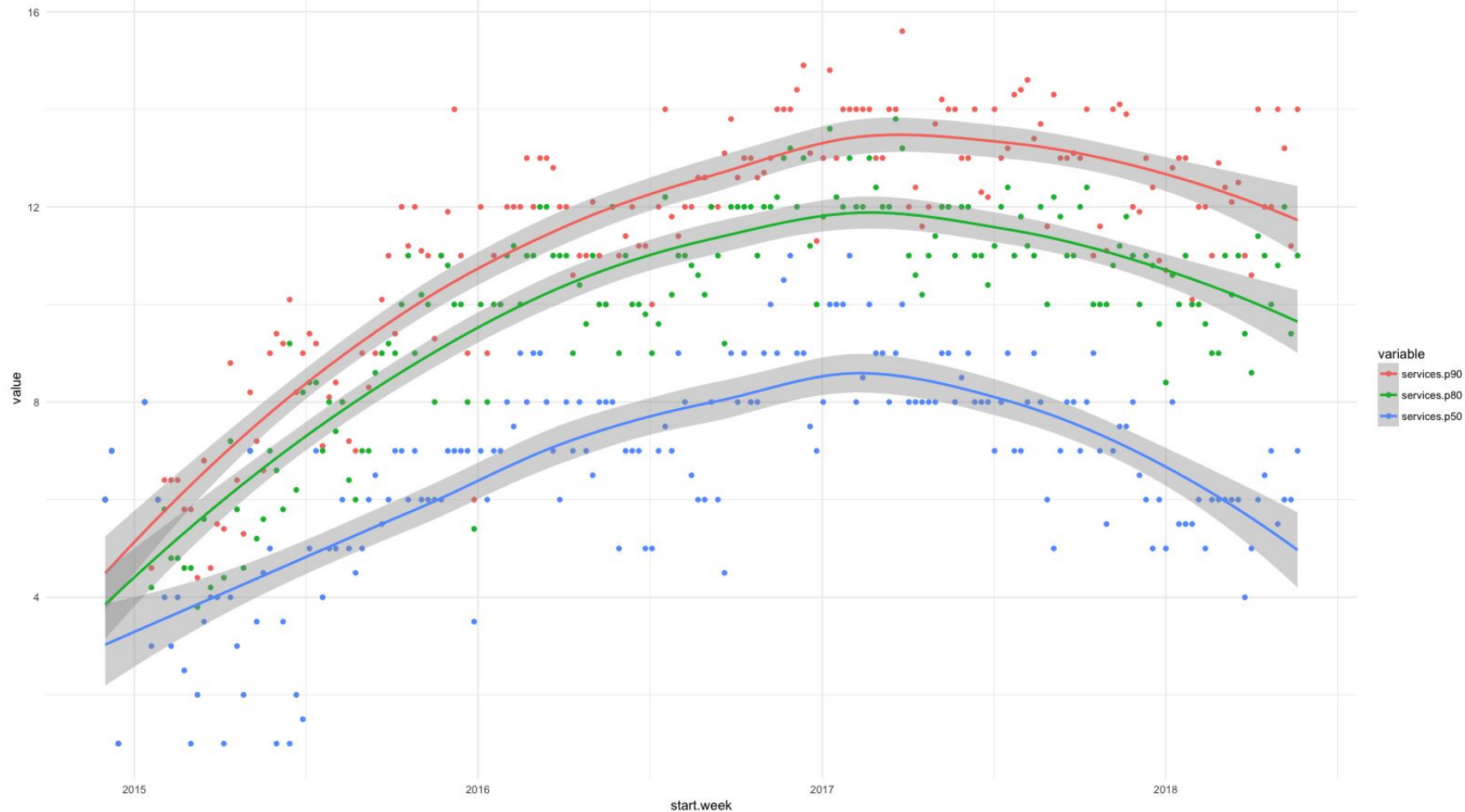
Source: <https://medium.com/@pingles/convergence-to-kubernetes-137ffa7ea2bc>



Low-level AWS service calls (EC2, IAM, STS, Autoscaling, etc.) from January 2015 to January 2017

*“We didn’t change our organization because we wanted to use Kubernetes, **we used Kubernetes because we wanted to change our organization.**”*

- Paul Ingles



Low-level AWS service calls since Kubernetes adoption in January 2017

Platform Purpose



**enable stream-aligned
teams to deliver work
autonomously with
self-service capabilities ...**

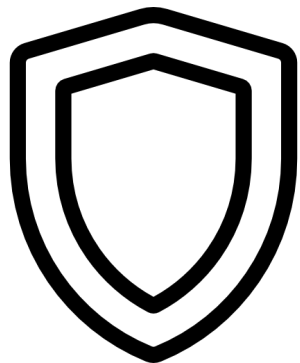
Platform Purpose



... in order to reduce
extraneous cognitive load
on stream-aligned teams

*“We wanted to scale our teams but maintain the principles of what helped us move fast: **autonomy**, work with **minimal coordination**, **self-service** infrastructure.”*

- Paul Ingles



Treat the platform as a product



Reliable Fit for Purpose Focused on DevEx

Reliable Platform



on-call support
service status pages
suitable comms channels
response time for incidents
downtime planned & announced

Fit for Purpose Platform



prototyping
fast, regular feedback
agile, iterative practices
few(er) services, high(er) quality
skilled product management

#DevEx Focused Platform



Speak the same language

Right level of abstractions for
your engineering teams today

“Kubernetes helps us in a few ways:

- ***Application-focused abstractions***
- *Operate and configure clusters to*
minimise coordination ”
- Paul Ingles



Dynamic Database Credentials

Multi-Cluster Load Balancing

Alerts + SLOs

Source (Joseph Irving): <https://t.co/99gwRH7dU2>



▼ availability



SLI

Proportion of valid requests that are successful. HTTP statuses 5XX are considered a failure.

Objective

99.50%

SLI

Total Error B...

55689

Remaining E...

46212

99.92%

ErrorBudget



Heimdall APP 20:46

firing: [blue]

[Affiliate products API] capone healthy traffic dropped below threshold



Heimdall APP 20:53

resolved: [blue]

[Affiliate products API] capone healthy traffic dropped below threshold

```
---
apiVersion: vaultwebhook.uswitch.com/v1alpha1
kind: DatabaseCredentialBinding
metadata:
  name: mybinding
  namespace: mynamespace
spec:
  serviceAccount: my_service_account
  database: mydb
  role: readonly
```

```
apiVersion: networking.k8s.io/v1beta1
kind: Ingress
metadata:
  name: my-ingress
spec:
  rules:
  - host: host.usw.co
    paths:
    - path: /
      backend:
        serviceName: my-service
        servicePort: 80
status:
  loadBalancer:
    ingress:
    - hostname: lb.eu-west-1.elb.amazonaws.com
```

```
apiVersion: vedfolnir.uswitch.com/v1alpha1
kind: ServiceLevelObjective
metadata:
  annotations:
    name: an-slo-name
    namespace: a-namespace
spec:
  slo:
  - name: availability
    description: Proportion of valid requests that are successful
    objective: "0.97"
    period: 28d
    counter:
      name: nginx_ingress_controller_requests
      prometheus: ingress
      badSelector:
        matchExpressions:
        - status=~"(4|5..)"
      validSelector:
        matchExpressions:
        - ingress="my-app-production"
        - path!~"/(status|metrics)"
```

2018

Infra platform
started with few
services

First customer
(centralized
logging, metrics,
auto scaling)



2018

Infra platform
started with few
services

First customer
(centralized
logging, metrics,
auto scaling)



2019

Started using SLAs
and SLOs, clarifying
reliability/latency/etc

**Growing traffic in
platform vs AWS**

2018

Infra platform
started with few
services

First customer
(centralized
logging, metrics,
auto scaling)



2019

Started using SLAs
and SLOs, clarifying
reliability/latency/etc

**Growing traffic in
platform vs AWS**

...

Addressed critical
cross-functional
needs (GDPR,
security, alerts +
SLOs as a service)

**Adoption by HMMT
(Highest Money
Making Team)**

2018

Infra platform
started with few
services

First customer

(centralized
logging, metrics,
auto scaling)



RV

2019

Started using SLAs
and SLOs for clarifying
reliability/latency/etc

**Growing traffic in
platform vs AWS**



...

Addressed critical
cross-functional
needs (GDPR,
security, alerts +
SLOs as a service)

**Adoption by HMMT
(Highest Money
Making Team)**



Platform Metrics



product metrics

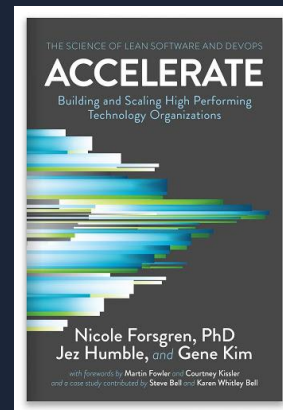
4 key metrics: *'Accelerate'*

lead time

deployment frequency

mean time to restore (MTTR)

change fail percentage



Platform Metrics



product metrics

user satisfaction metrics



AGREE OR DISAGREE

BUILD

I can effectively build my software.

I have the tools to validate my software.

DELIVER

I can reliably deliver my software to dev, stage, and prod.

I can efficiently manage my cloud infrastructure.

RUN

I can measure the operational metrics of my services.

I can understand the cost of running my service.

COMPELLING

I feel platform tools are consistently improving.

I can voice problems that result in improvements.

My tools are best in class.

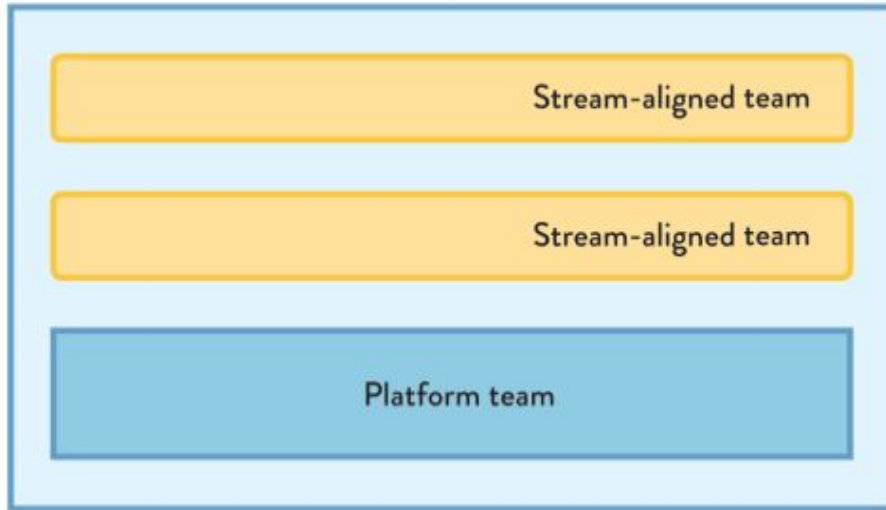
Platform Metrics



product metrics

user satisfaction metrics

adoption & engagement metrics



Platform Metrics



product metrics

user satisfaction metrics

adoption & engagement metrics

reliability metrics



▼ availability



SLI

Proportion of valid requests that are successful. HTTP statuses 5XX are considered a failure.

Objective

99.50%

SLI

Total Error B...

55689

Remaining E...

46212

99.92%

ErrorBudget



Platform Metrics



product metrics

(Accelerate metrics for platform services)

user satisfaction metrics

(Accelerate metrics for business services, NPS, etc)

adoption & engagement metrics

(% teams onboard, per platform and per service)

reliability metrics

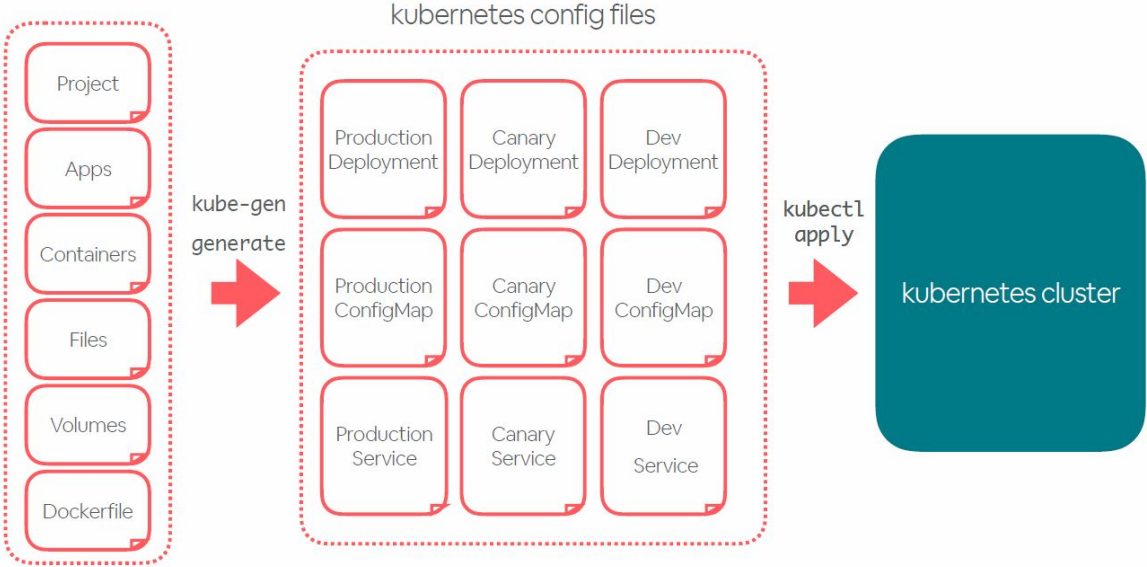
(SLOs, latency, #Incidents, etc)



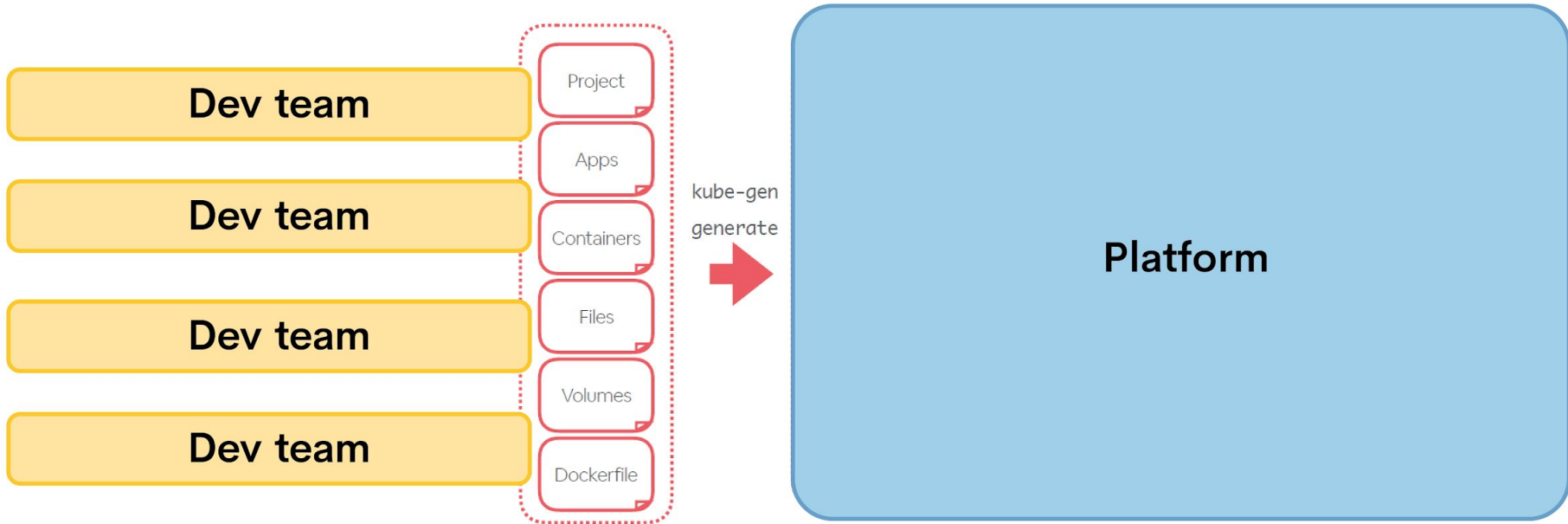
***The success of platform
teams is the success of
stream-aligned teams***

Team Interactions

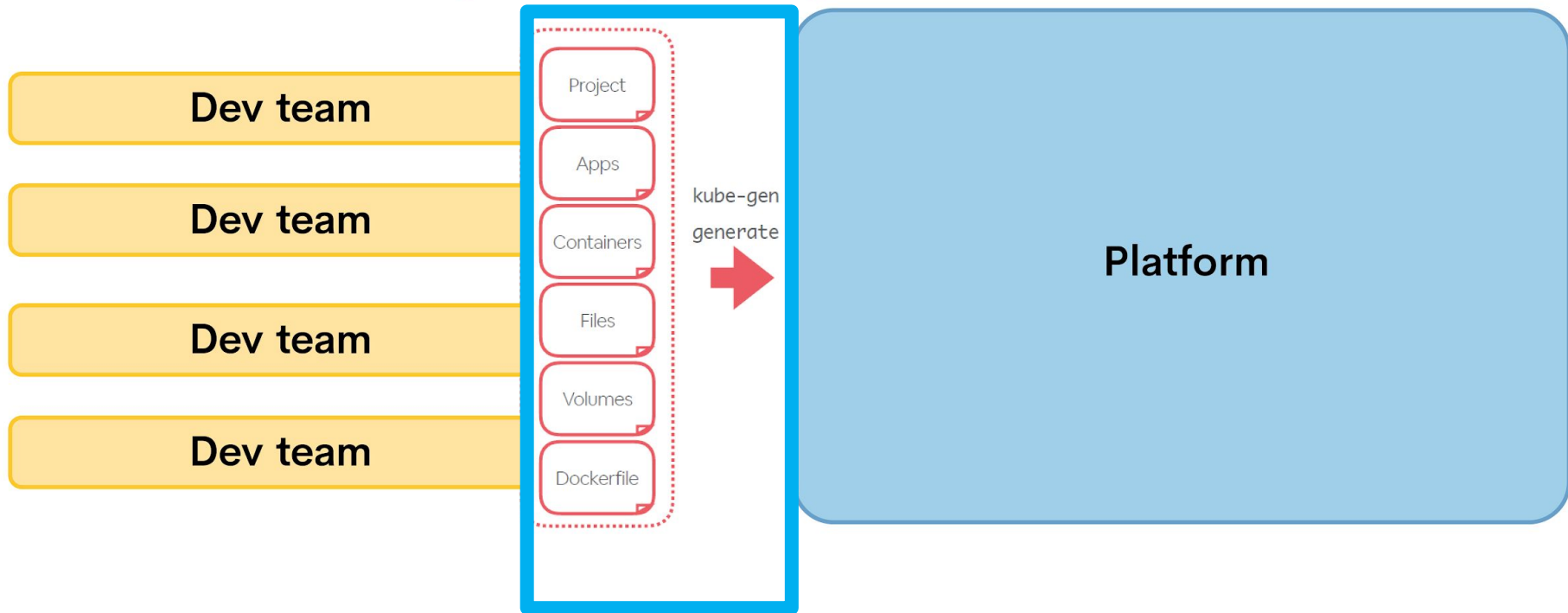
generating k8s configs



generating k8s configs

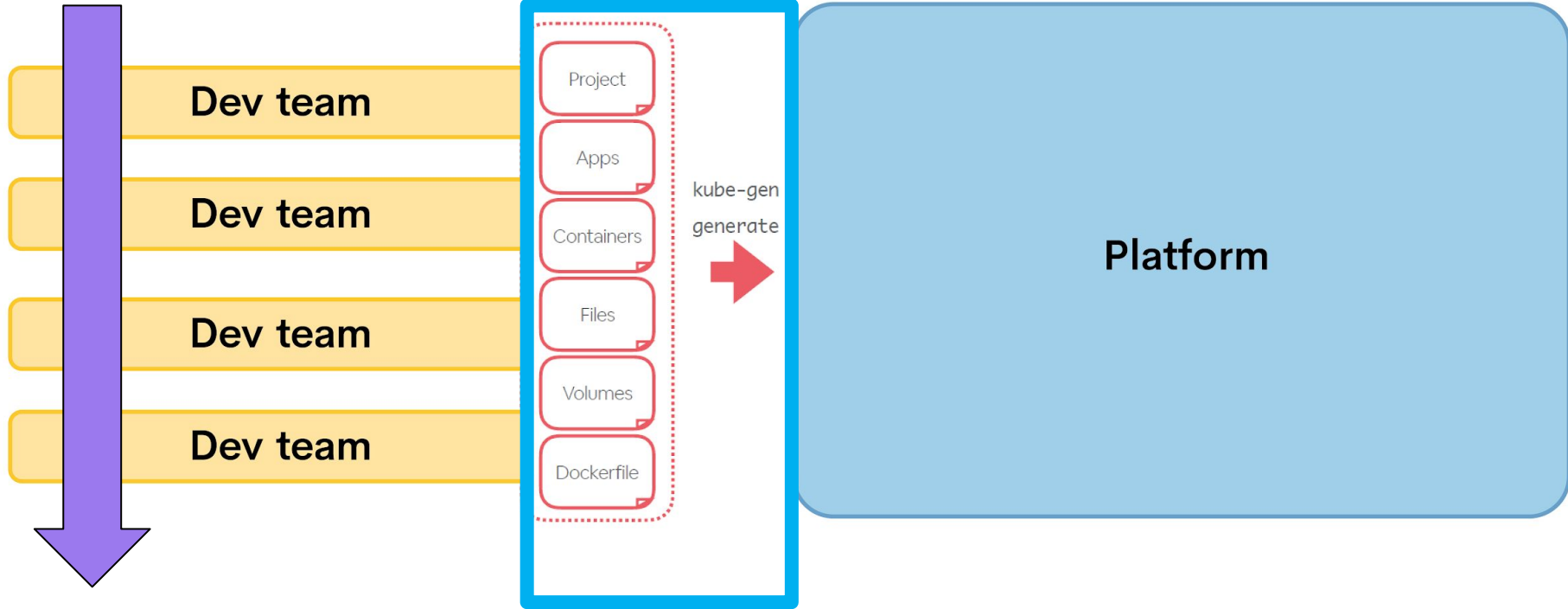


generating k8s configs



service boundary

generating k8s configs



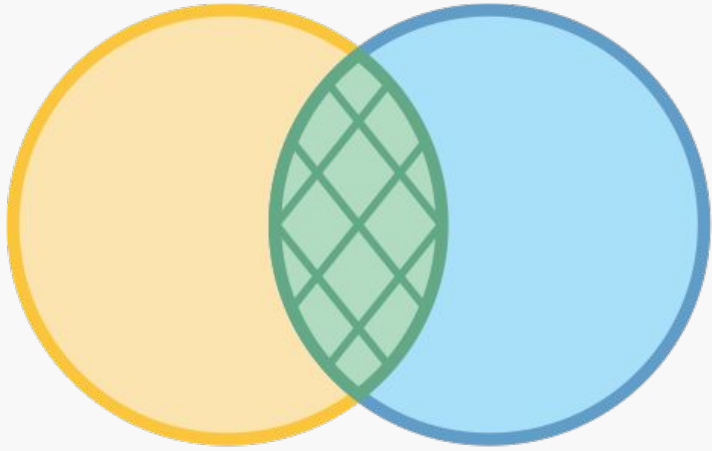
cognitive load

service boundary

Platform Behaviors



**strong collaboration with
stream-aligned teams for
any new service or evolution**

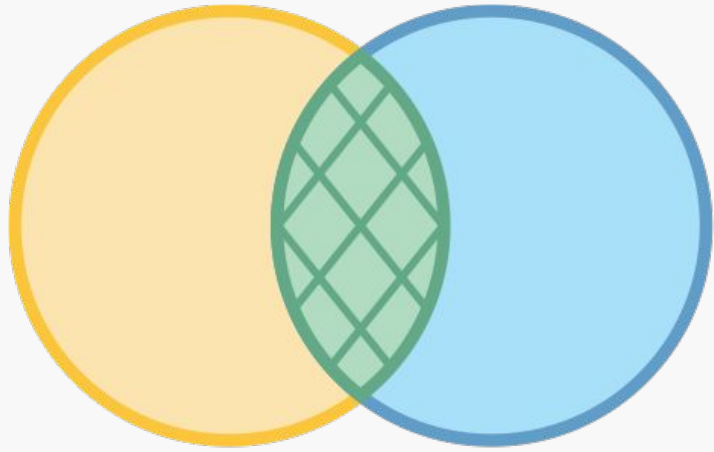


Collaboration

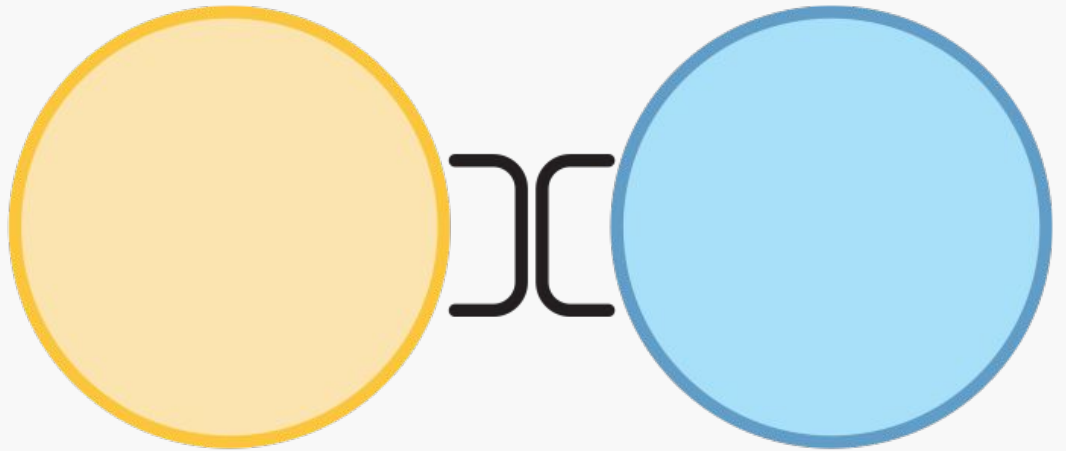
Platform Behaviors



**provide support and great
documentation for stable
services**

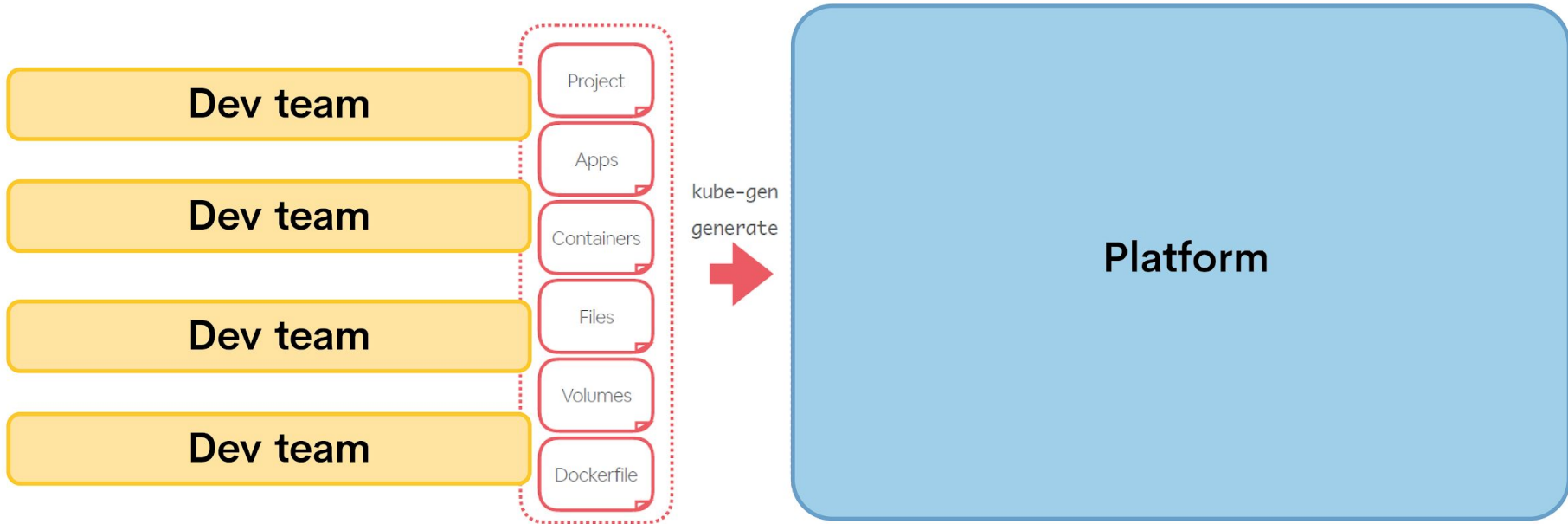


Collaboration

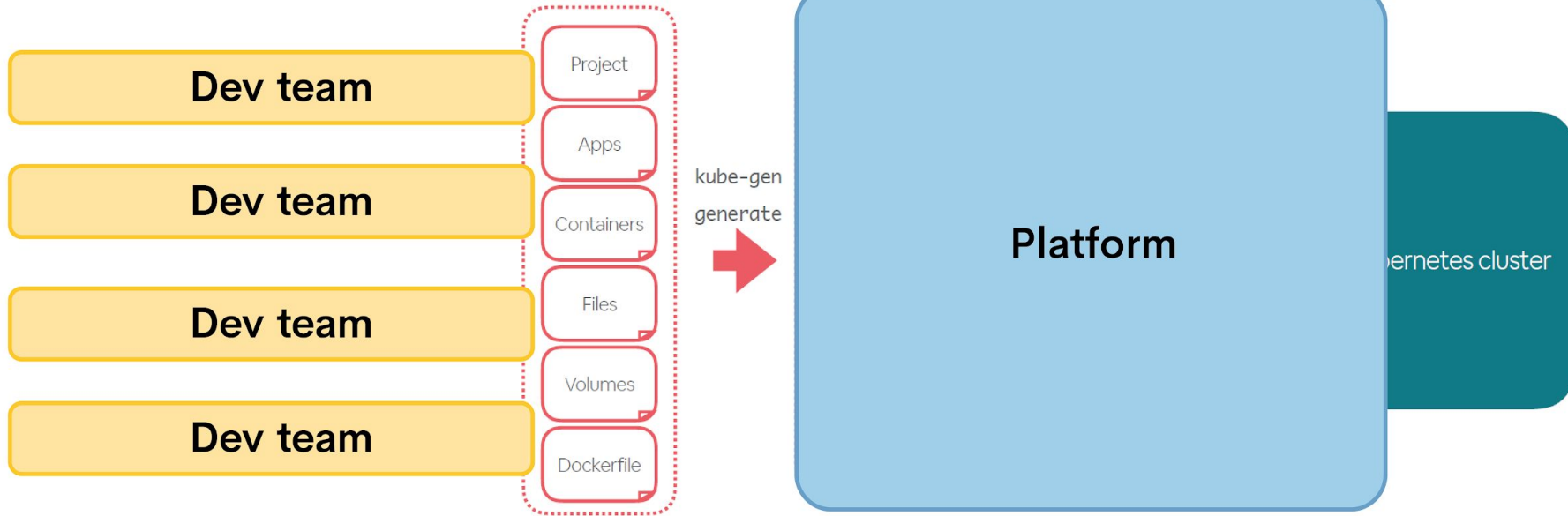


X-as-a-service

generating k8s configs



generating k8s configs



Dev team

k gen

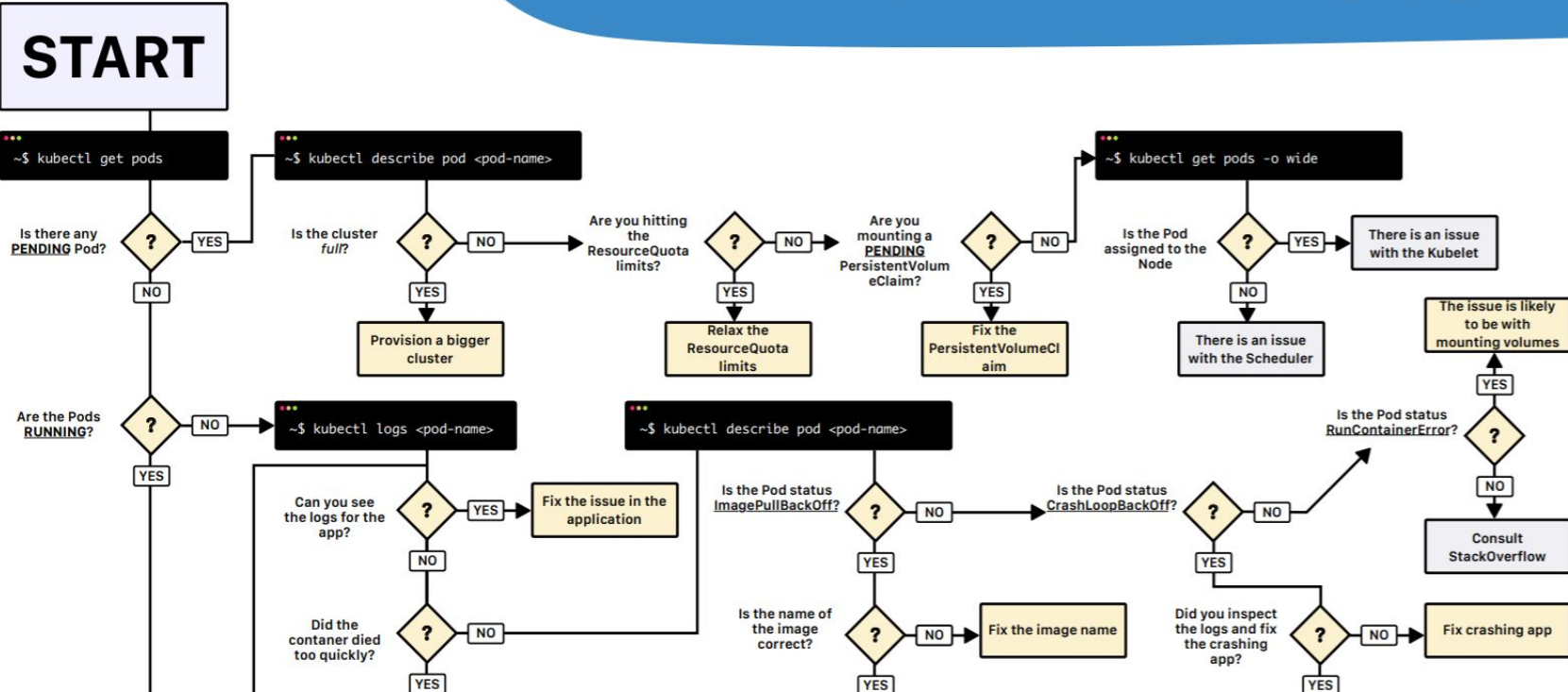
k deploy

Infrastructure team

Troubleshooting Kubernetes deployments
Read the blog article at
<https://learnk8s.io/troubleshooting-deployments>

Read the blog article at

<https://learnk8s.io/troubleshooting-deployments>





Dev team

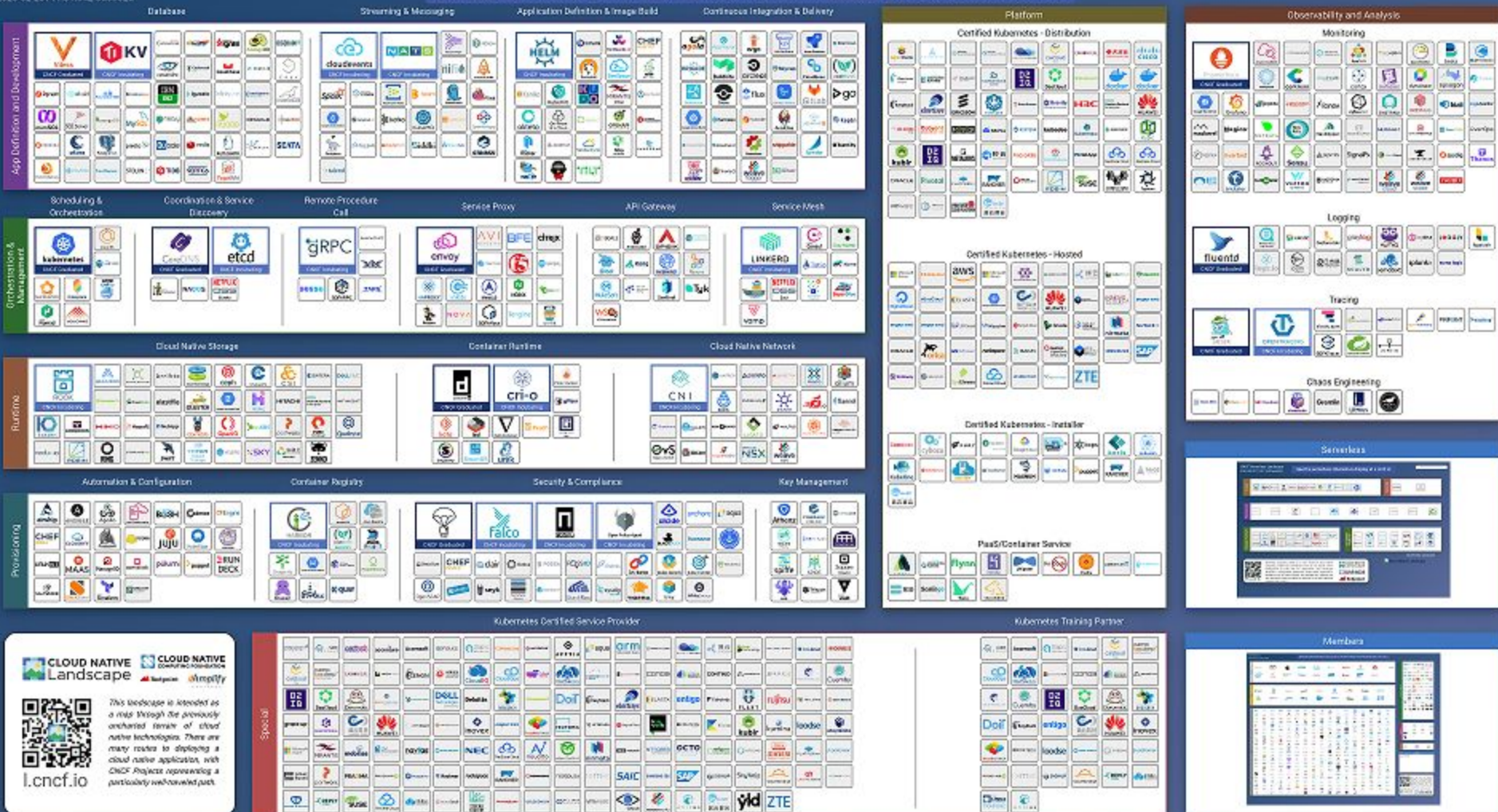
k gen

k deploy

k diagnose

...

Infrastructure team



uswitch / heimdall

👁 Watch 8

★ Star 16

🍴 Fork 6

Code

Issues 0

Pull requests 0

Actions

Projects 0

Wiki

Security

Insights

Generate PrometheusRule CRDs from Ingress annotations and Go templates

🔄 69 commits

🌿 1 branch

📦 0 packages

🏷 8 releases

👤 8 contributors

📄 Apache-2.0

Branch: master

New pull request

Create new file

Upload files

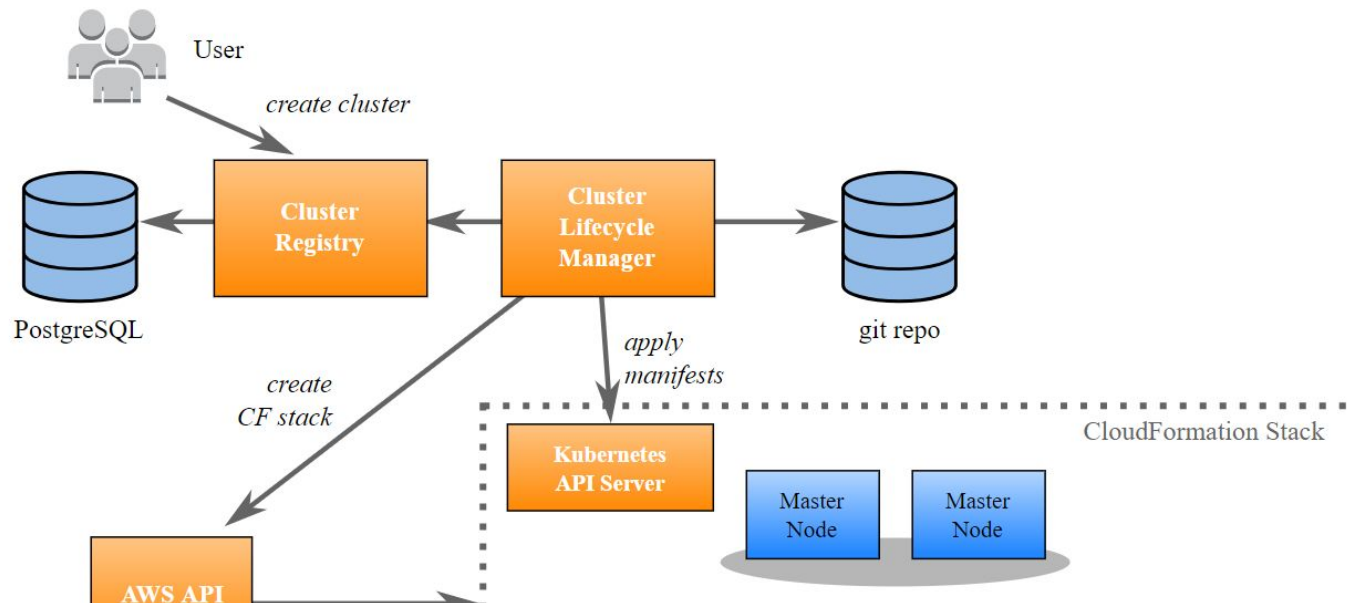
Find file

Clone or download

Cluster Lifecycle Manager (CLM)

build passing coverage 52% go report A+

The Cluster Lifecycle Manager (CLM) is a component responsible for operating (create, update, delete) Kubernetes clusters. It interacts with a Cluster Registry and a configuration source from which it reads information about the clusters and keep them up to date with the latest configuration.



Getting Started with team-centric Kubernetes adoption

1 - Assess cognitive load

How well can the team understand the platform/Kubernetes abstractions they need to use on a regular basis?

2 - Define your platform

What's the gap between your Kubernetes implementation and an internal digital platform?

3 - Team Interactions

Who is responsible for what? Who is impacted? How do you collaborate on new platform internal services?

Collaboration vs X-as-a-Service

More platform examples



Zalando Kubernetes at Zalando



Mercedes DevOps Adoption at Mercedes-Benz.io

Twilio Platforms at Twilio: Unlocking Developer Effectiveness

Adidas Where Cloud Native Meets the Sporting Goods Industry

ITV ITV's Common Platform v2 Better, Faster, Cheaper, Happier

MAN Truck & Bus How to Manage Cloud Infrastructure at MAN Truck & Bus

Farfetch UX | DevOps - The Trojan Horse for Implementing a DevOps Culture



Why teams fail with Kubernetes—and what to do about it

There are major implications to how teams must interact when you're using Kubernetes—especially as you scale. Here are key approaches to ...

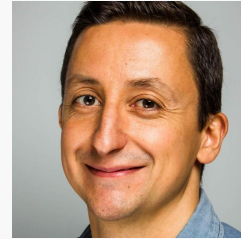
techbeacon.com

Thank you!

teamtologies.com



Matthew Skelton, Conflux
@matthewpskelton



Manuel Pais, Independent
@manupaisable



Copyright © [Conflux Digital Ltd](#) 2018-2020. All rights reserved.
Registered in England and Wales, number 10890964

Icons made by [Freepik](#) from [www.flaticon.com](#) - used under license