infobip

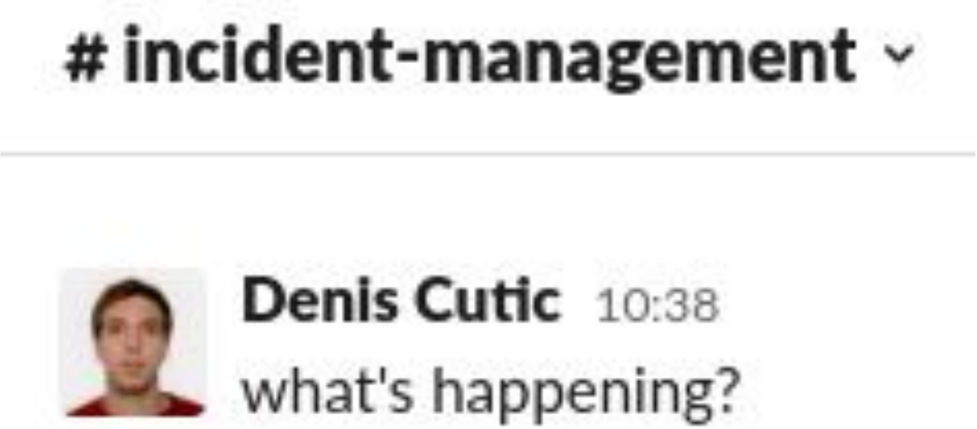
SRE & Happiness

Denis Ćutić
Senior Site Reliability Engineer @ Infobip

JavaCro'21

Let me tell you a story...
Story #1

29.04.2018

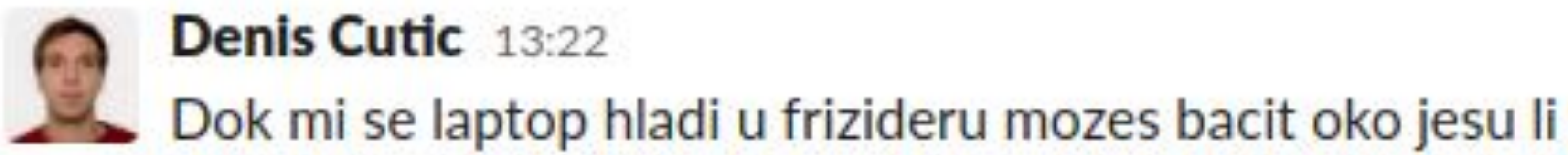


incident-management

Denis Cutic 10:38

what's happening?

What happened?

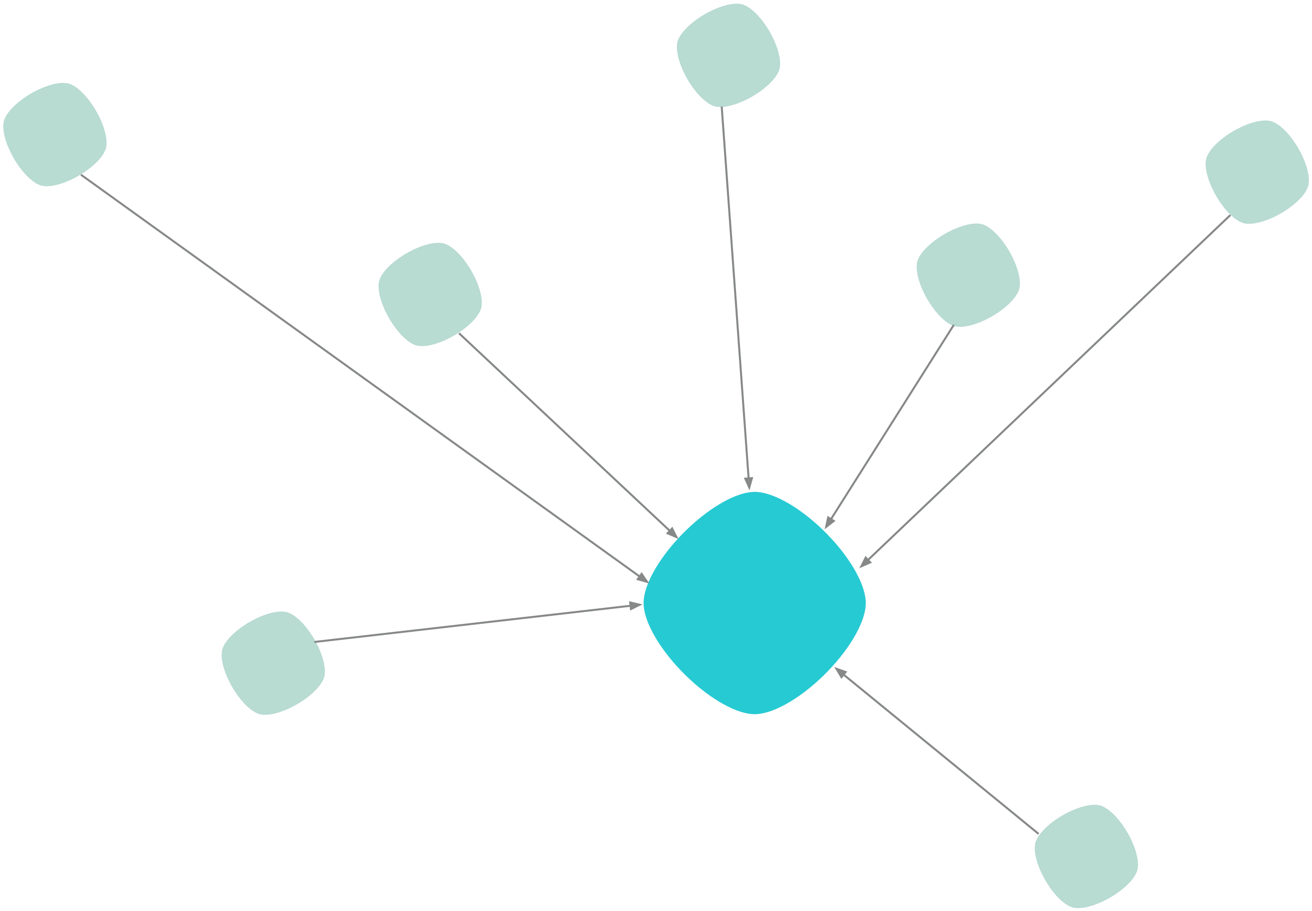


Denis Cutic 13:22

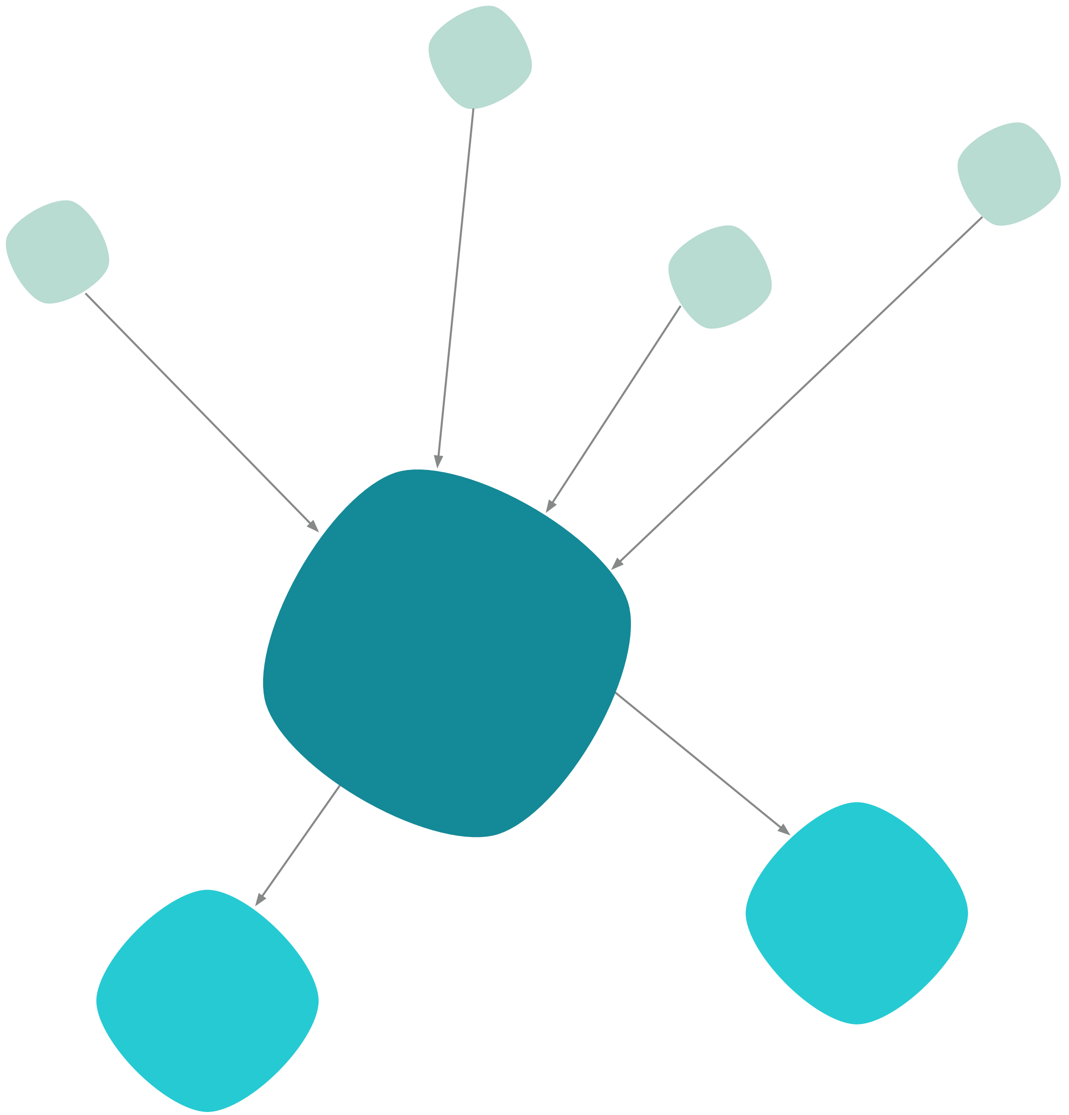
Dok mi se laptop hladi u frizideru mozes bacit oko jesu li hope != strategy

My journey
Story #2

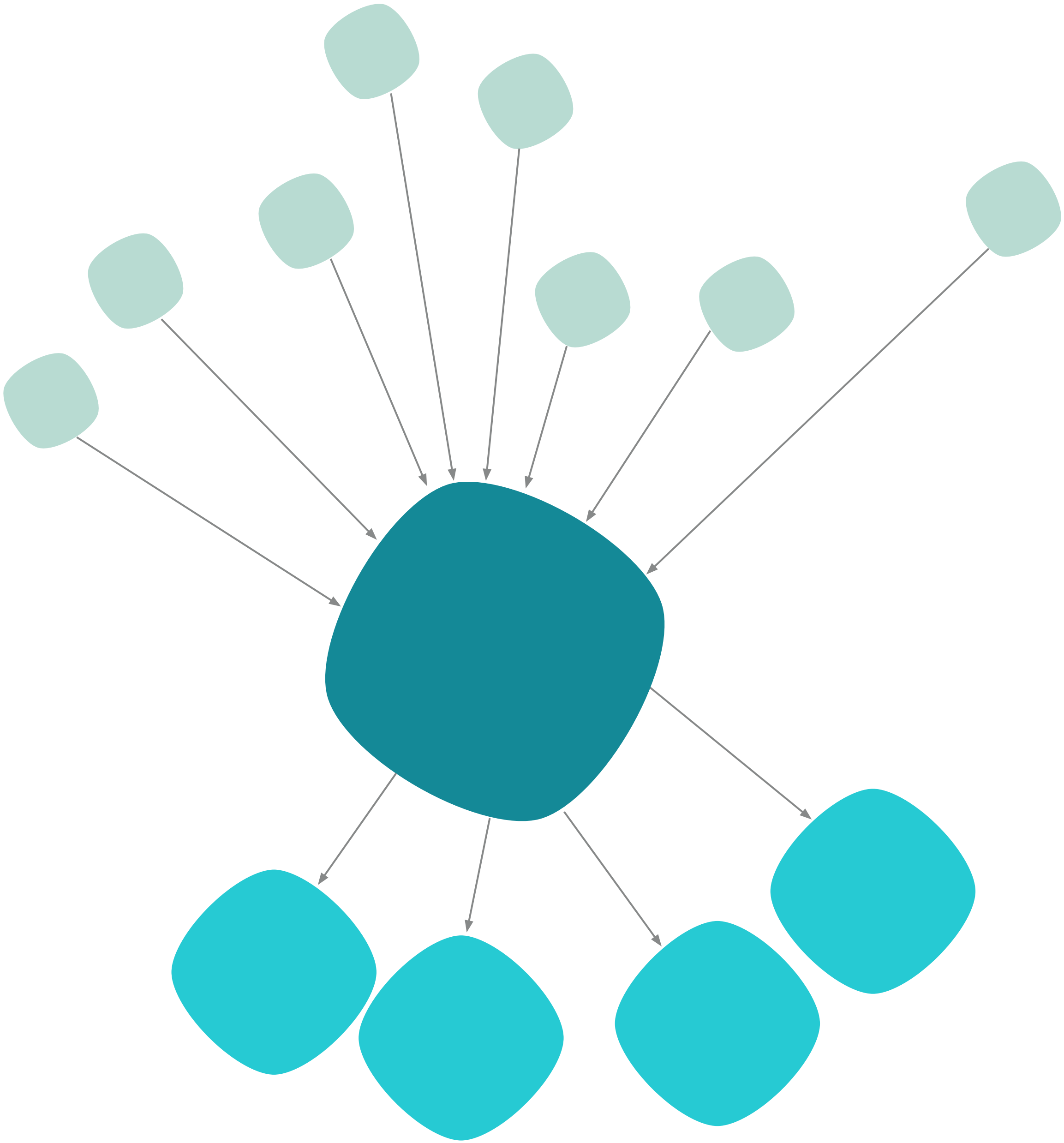
Auth services
Account/User management services



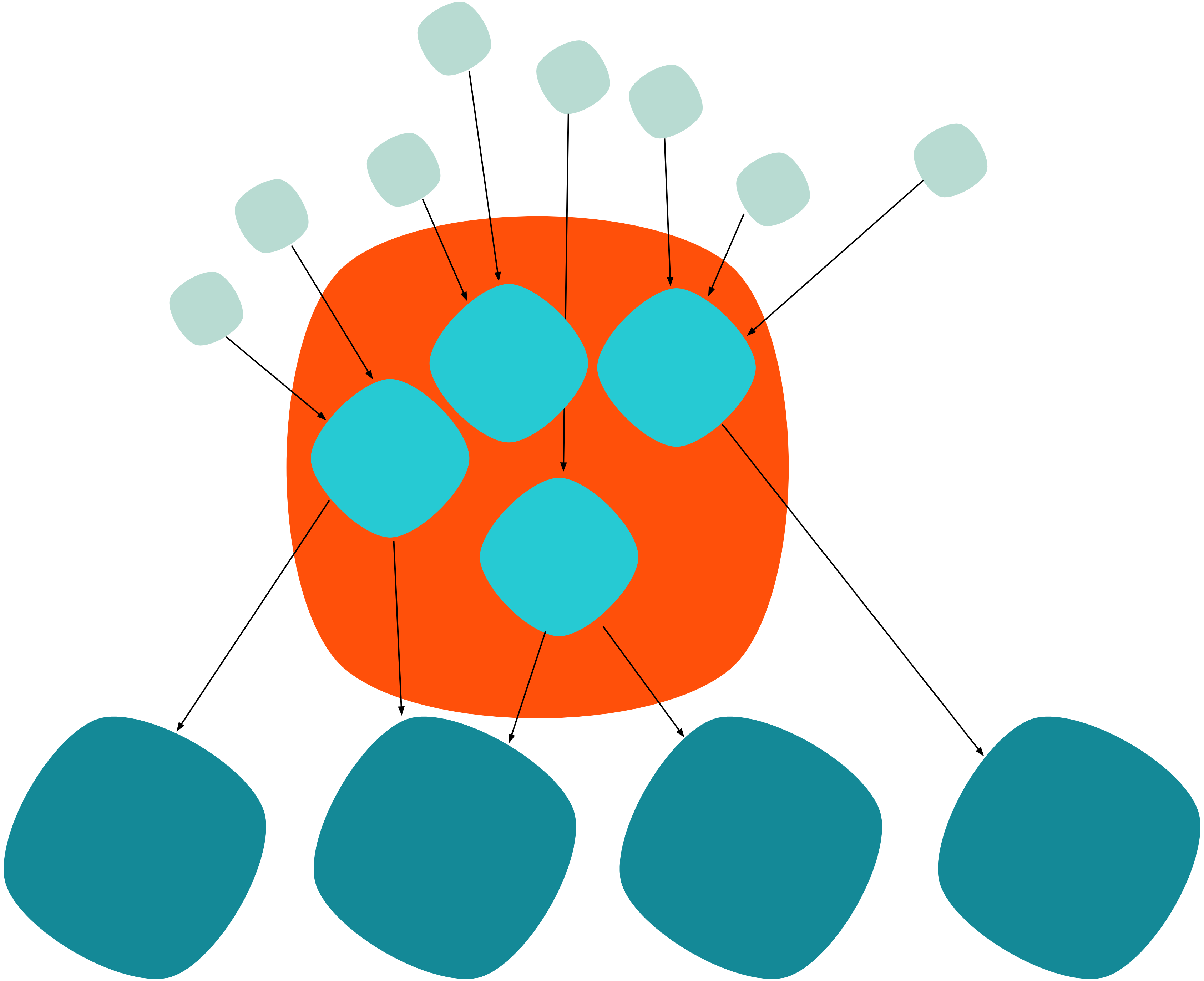
SMS API service



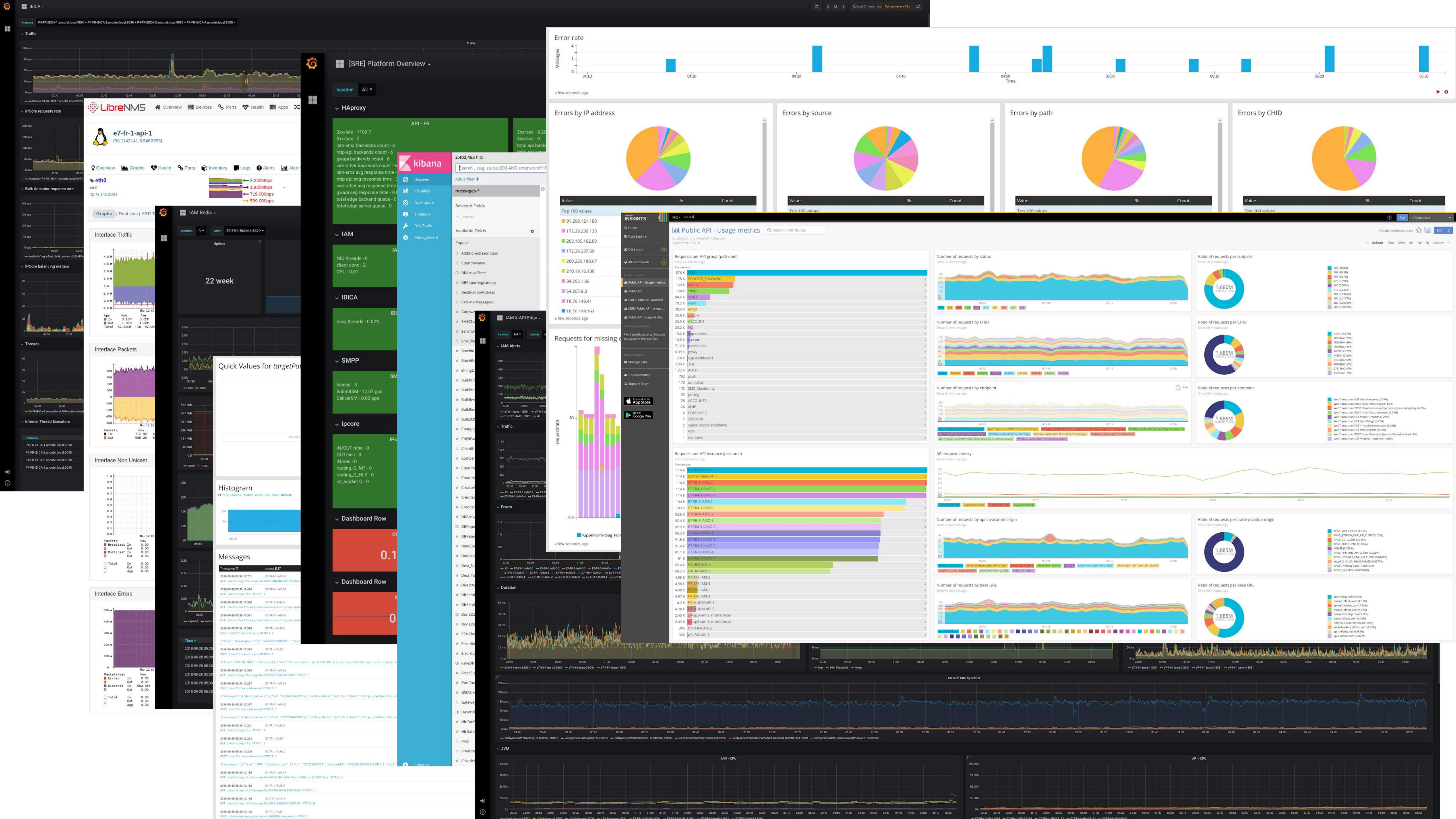
Generic API service



API infrastructure



Survival?



HAproxy

API - FR

2xx/sec - 1109.7
5xx/sec - 0
iam-sms backends count - 6
http-api backends count - 6
gwapi backends count - 6
iam-other backends count - 6
iam-sms avg response time -
http-api avg response time - 0
iam-other avg response time
gwapi avg response time - 0.0
total edge backend queue - 0
total edge server queue - 0

2xx/sec - 8.20
5xx/sec - 0
total api backe

IAM

NIO threads - 0
client conn - 2
CPU - 0.01

IBICA

busy threads - 0.02%

SMPP

binded - 3
SubmitSM - 12.07 pps
DeliverSM - 0.03 pps

ipcore

IN/OUT ratio - 0
OUT/sec - 0
IN/sec - 0
routing_Q_MT - 0
routing_Q_HLR - 0
int_worker_Q - 0

Dashboard Row

0.1

Dashboard Row

0

Errors by IP address

Value
Top 100 values
91.208.121.180
172.29.239.130
203.105.162.80
172.29.237.59
200.220.188.67
210.19.16.130
34.205.1.66
54.207.8.3
10.76.148.41
10.76.148.162

Public API - Usage metrics

Created by Josip.Antolis@infobip.com
Last edited 7/24/18

Requests per API group (pick one!)
Since 60 minutes ago

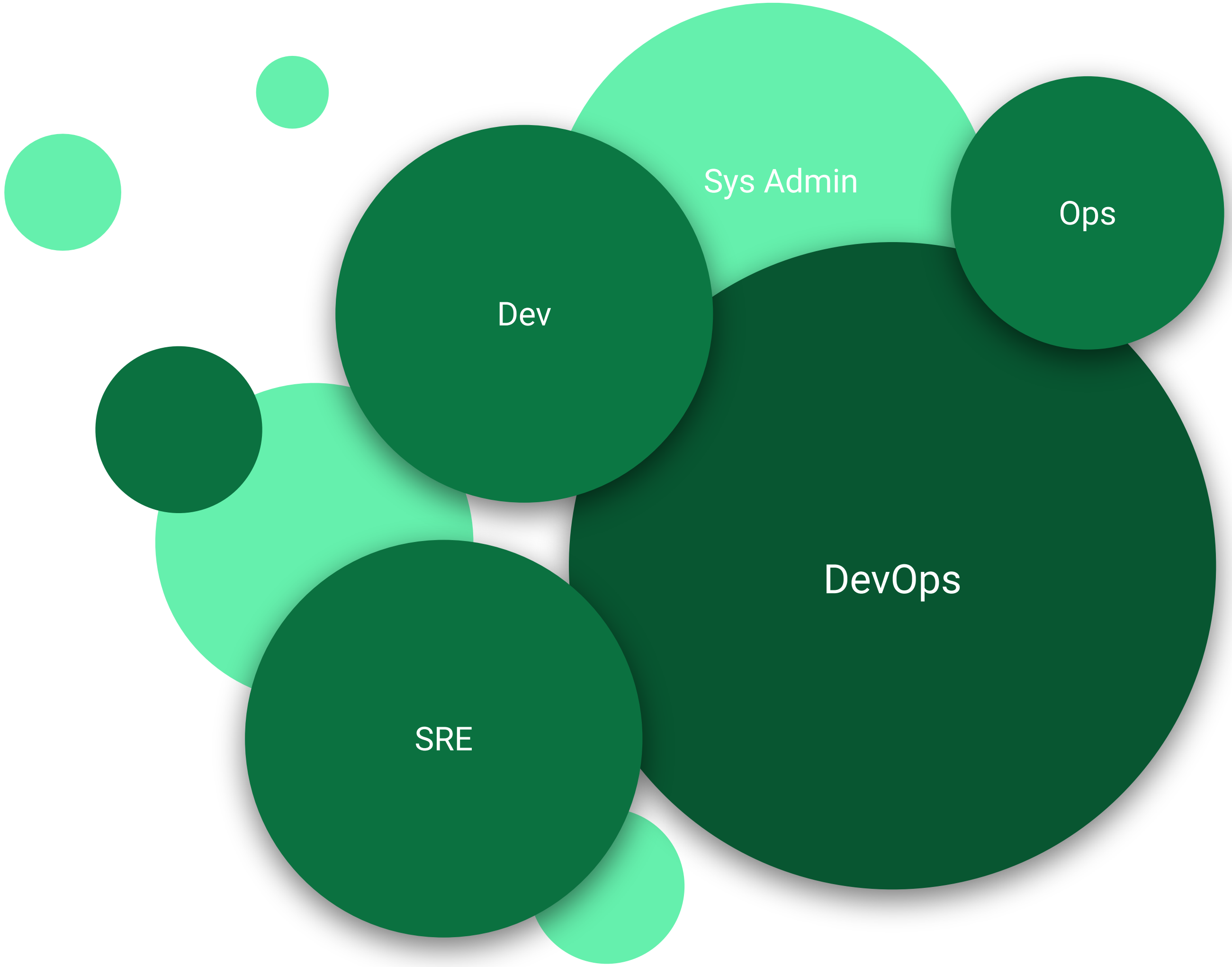
Transactions	
876 K	SMS
175 K	PROCESS_TRACKING
169 K	Mobile
154 K	OMNI
84.6 K	VOICE
70.2 K	viber
38.4 K	email
26.8 K	people
23.5 K	ACCOUNT
23.2 K	NC
13.2 K	cup-reports
10.8 K	custom
7.12 K	people-dev
6.39 K	proxy
2.8 K	cup-dashboard
2.04 K	2FA
1.22 K	AUTH
792	push
173	omnichat
172	SMS_Monitoring
32	pricing
28	ACCOUNTS
26	MNP
6	CUSTOMER
6	ZENDESK
2	superuserapi-salesforce
1	VoIP
1	numbers

Requests per API instance (pick one!)
Since 60 minutes ago

Transactions	
114 K	E7-FR-1-IAMS-3
114 K	e7-fr4-1-iams-4
114 K	E7-FR-1-IAMS-2
114 K	E7-FR4-1-IAMS-2
114 K	E7-FR4-1-IAMS-3
104 K	E7-FR-1-IAMS-1
103 K	E7-FR4-1-IAMS-1
93.6 K	E7-FR-1-IAMO-2
92.4 K	E7-FR-1-IAMO-3
92.2 K	E7-FR4-1-IAMO-3
92.2 K	E7-FR-1-IAMO-1
91.8 K	E7-FR4-1-IAMO-1
91.4 K	E7-FR4-1-IAMO-2
91.1 K	E7-FR-1-IAMO-4
91 K	E7-FR4-1-IAMO-4
69.4 K	P4-HK2-IAM-1
68.2 K	P4-HK2-IAM-2
4.98 K	P7-AFR-IAM-2
4.98 K	P7-AFR-IAM-4
4.98 K	P7-AFR-IAM-1
4.97 K	P7-AFR-IAM-3
4.3 K	P4-MUNM-API-1
4.28 K	P4-MUNM-API-2
2.43 K	p4-sp3-iam-2.ancotel.local
2.43 K	p4-sp3-iam-1.ancotel.local
305	P7-TFR4-IAM-2
300	p7-tfr4-iam-1

SRE

Infobip’s journey Story #3 - Prequel



Sys Admin

Ops

Dev

DevOps

SRE

DevOps VS SRE

Infobip's journey
Story #3

~~DevOps~~
Startup

Reliability

Beginner level

- Reactive
- Hopefully someone will catch failures
- Hopefully it will be someone inside the company
- Hopefully someone will notice some patterns
- Hopefully we will meet the agreed SLA

Sys Admins + DevOps + Ops

Reliability

Advanced level

- Reactive
- Observability tools in place
- Service-level monitoring from client PoV
- Promoted through company culture
- Support process in place
- Incident management process
- Starting with post-incident reviews

Sys Admins + DevOps + Ops +
QA + SecOps + SRE

Reliability

Shipping stuff to space level

Not there yet...

Reliability

Professional level

- Dedicated team
- Proactive
- Collecting and analysing incident data
- Identifying and escalating issues on organisational level
- Unifying and improving processes
- Transparency

SRE @ Infobip

Numbers

Company

5 Business Areas

26 Requirement Areas

100+ teams

~900 engineers

~3000 employees total

Products / Platform

23 products and channels

39 DCs

50+ locations

3 clouds (on prem + 2 public)

Rate of change (monthly)

~30k deployments

~30B client interactions

~30k active web users

50+ maintenances

1 release (of all products)

SRE

5 team members

30+ years of IB experience

10+ different IB job titles

50+ years of IT experience

Platform monitoring

Platform, high-level alerts

Open channels to support and teams

Request teams to expose relevant metrics

Driving SLO adoption

Incident Management

Owners of the IM process

Helping/Handling incidents

Incident commanders for complex incidents

Collection and analysis of incident meta-data

Monthly, quarterly, yearly reports

Post-incident reviews

Tooling

Automating operational tasks

Automating processes

Reviewing usage of observability, alerting and escalation tools

Educating how to best use the above

Coordination

High-impact and critical maintenances

Handling reliability-related inquiries by clients

Client integration when high levels of reliability is required

Mediators between stakeholders

Culture

Blameless incident culture

Data-driven decision making

Contextualizing SRE practices for our way of work

Defining best practices: HA, monitoring, availability, etc.

Promoting a client-centric view of problems on our platform

Culture

... if you do [have problems]

we will look for you

we will find you

and we will help you

Product

Reliability review

Driving reliability improvements

1 godina SREće (1 year of happiness)

- 0 to 5 SREs
- Incident number increase
  - Improved detection
  - Improved reports
- On average, incident duration is halved when SRE member involved
  - No better metric ATM :(
- Reporting speed-up
  - Monthly: from weeks to < 1 day
  - Self-service dashboards for managers
- High, cross-company, visibility

Road to SRE

WHY SRE?

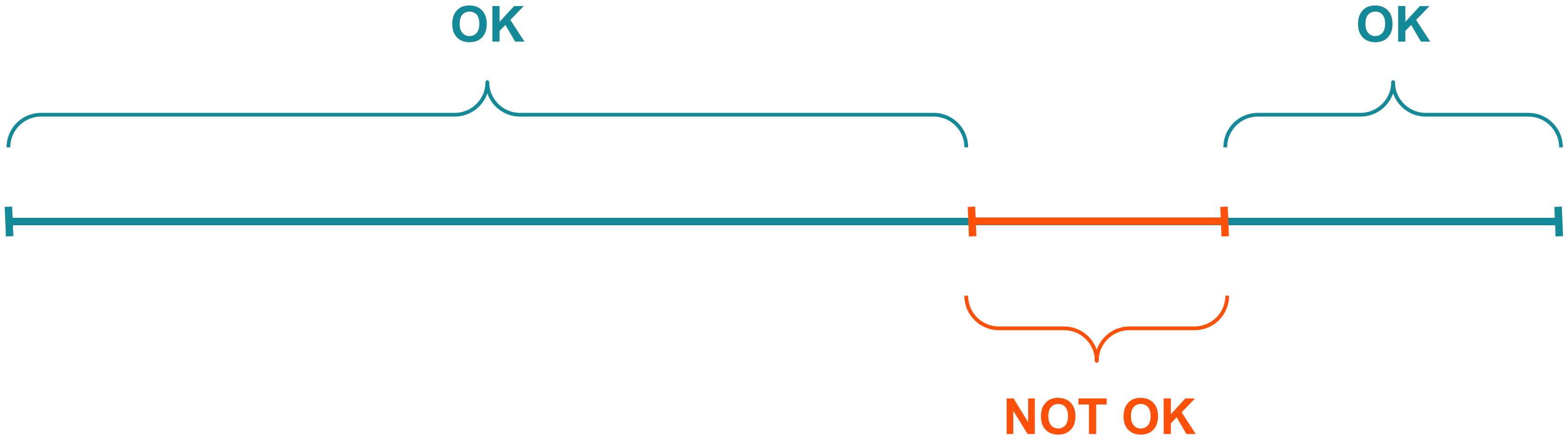
Fulfill contractual obligations

Proactively build and maintain reliable services

WHY SRE?

hope != strategy

WHAT IS NEEDED?



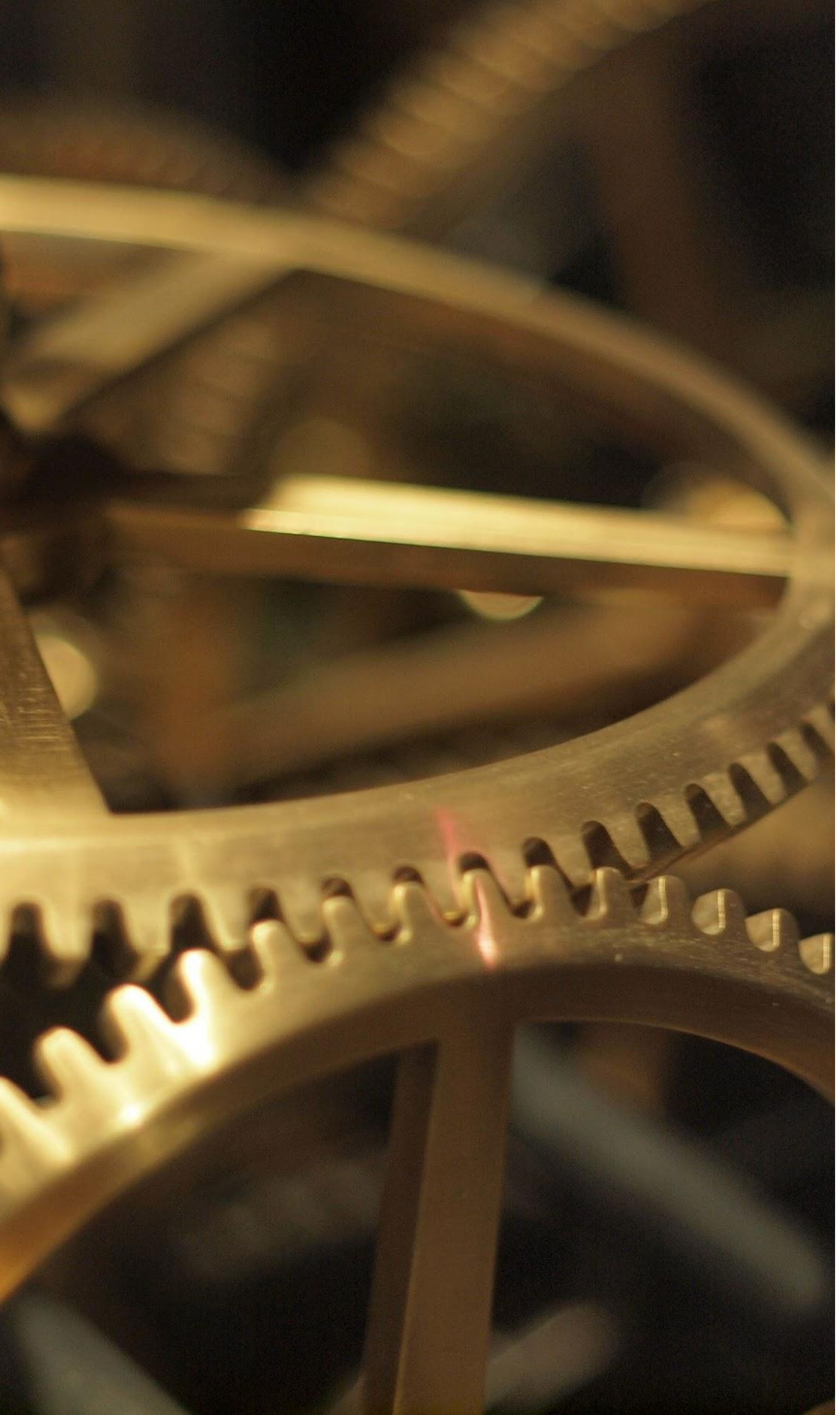
Reliability measures the functioning of a service over a period of time, under specified conditions.

WHAT

Measure

- Monitoring
- Observability is the base
- It is not trivial to have good measures
- If you don't have it, start today
- Once you have it, improve it constantly

Progressive improvement beats delayed perfection.
Mark Twain



WHAT

Function

- Categorization ok / not ok
- Define it carefully
- Make it observable
- Define it from the client PoV
  - Whoever or whatever the client in your context is

If the client receives a 200 OK response, because the request was put into a queue, are they getting the service they paid for?

WHAT

Incident

- Categorize incident / not incident
- When should people report and escalate problems?
- What are the thresholds?
- How to define them?
- They will happen
- Be ready
- SLOs make it easier to answer these

Success is not final, failure is not fatal:
it is the courage to continue that counts.
Winston Churchill

WHAT

Incident management

- Are all incidents equal?
- Are they equally severe?
- Do they have the same priority?
- How does one respond to an incident?
- Define the incident response process
  - Teach it
  - Exercise it
  - Improve it
- Guides
  - ITIL, ITSM, OODA
- Incident reports
  - For transparency
  - For improvements rather focus on specific incidents

WHAT

Disaster scenarios

- Disaster will happen too
- Are you ready?
    - Facebook was
- What is the cost of the service being down
    - Day?
    - Week?
    - Month?

> *I think the environmental impact of this disaster is likely to have been very, very modest.*
> *Tony Hayward, BP CEO*

WHAT

On-call

- Organizing incident response
- Protecting people
  - Their well-being
  - Their work-life balance
  - Ther happiness
- Organizing rotations
- Clear responsibilities
- Clear expectations

Have you tried turning it off and on again?
IT Crowd

WHAT

Culture

> *Culture eats strategy for breakfast.*
>
> *Peter Drucker*

HOW TO SRE?

Each company does it differently

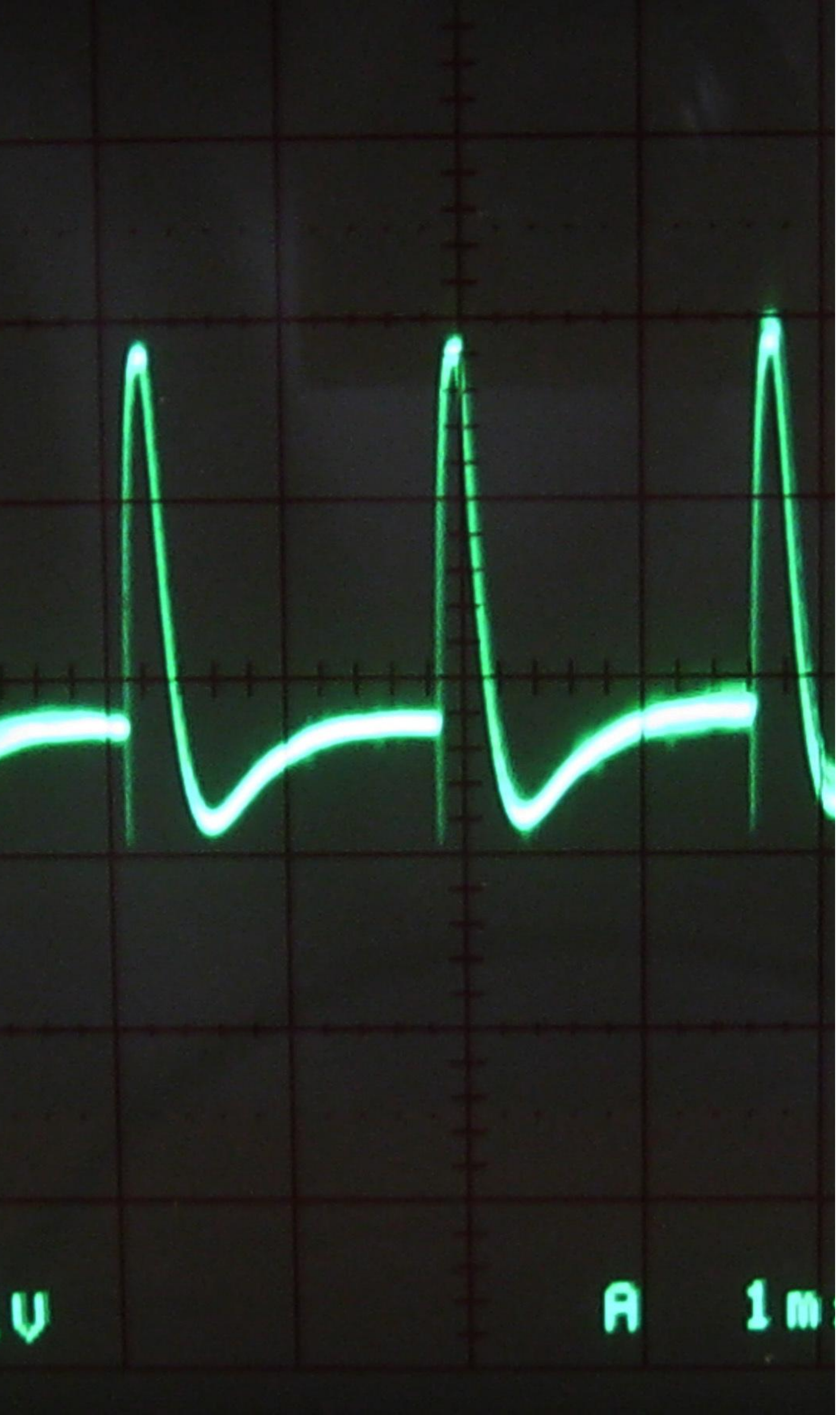
Needs to be aligned with the company culture

Start by adopting practices, one by one

HOW

Post-mortems / Post-incident reviews

- Incidents are complex
- Incidents are unique
- Focus on finding all the contributing causes
  - There's rarely a single, root cause
- Define planned actions
  - Make sure they are executed



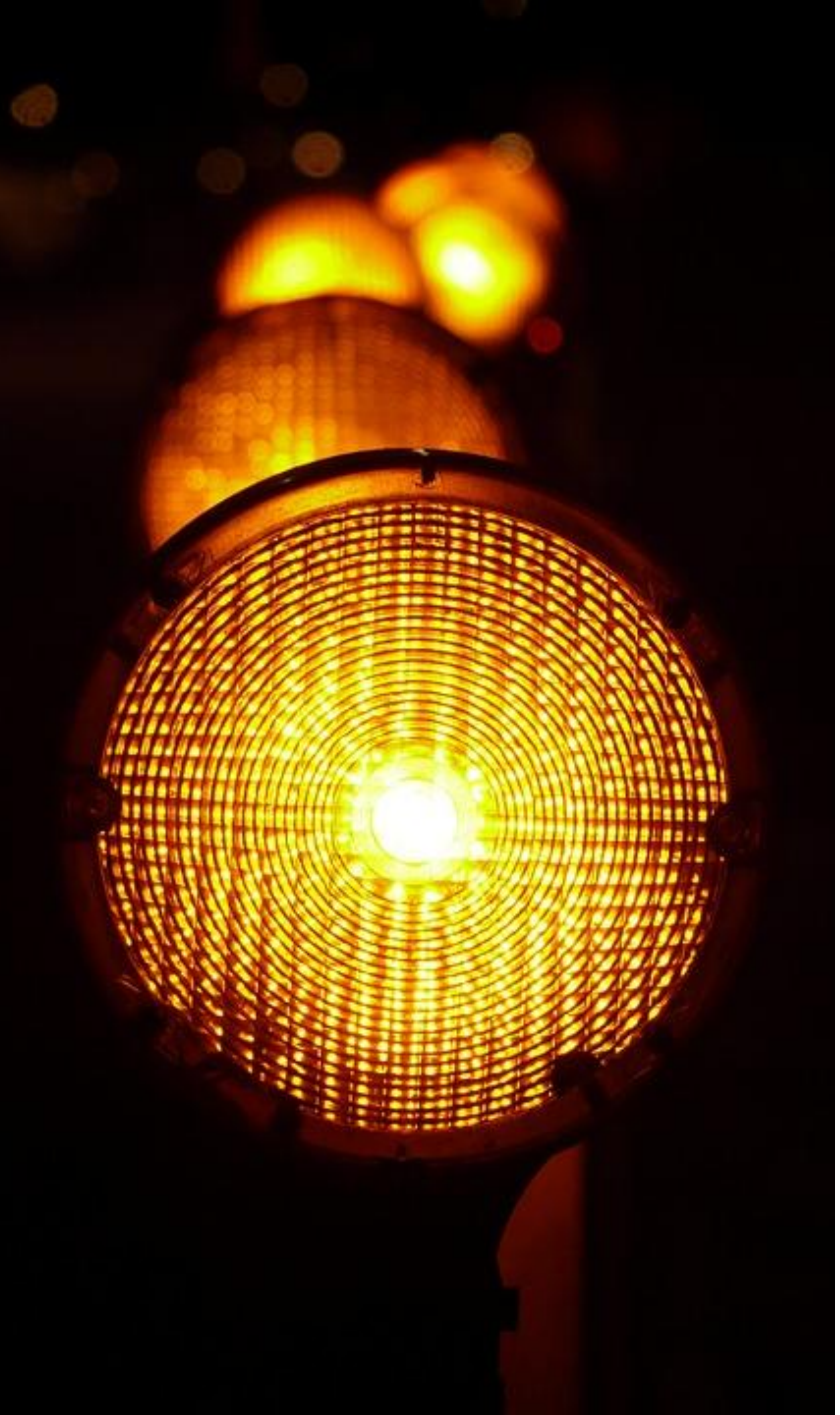
HOW

SLO - Service Level Objective

- SLIs are not just another metric
  - All stakeholders agree on the importance
  - Relates to business value
- Does not have to be 100% precise
  - Constantly improve
- If objective is not met, actions are taken
- Set realistic targets
  - No point in failing, constantly
  - SLO is always stricter than SLA
- Do not make it a KPI / OKR

When a measure becomes a target,
it ceases to be a good measure.

Goodhart's Law

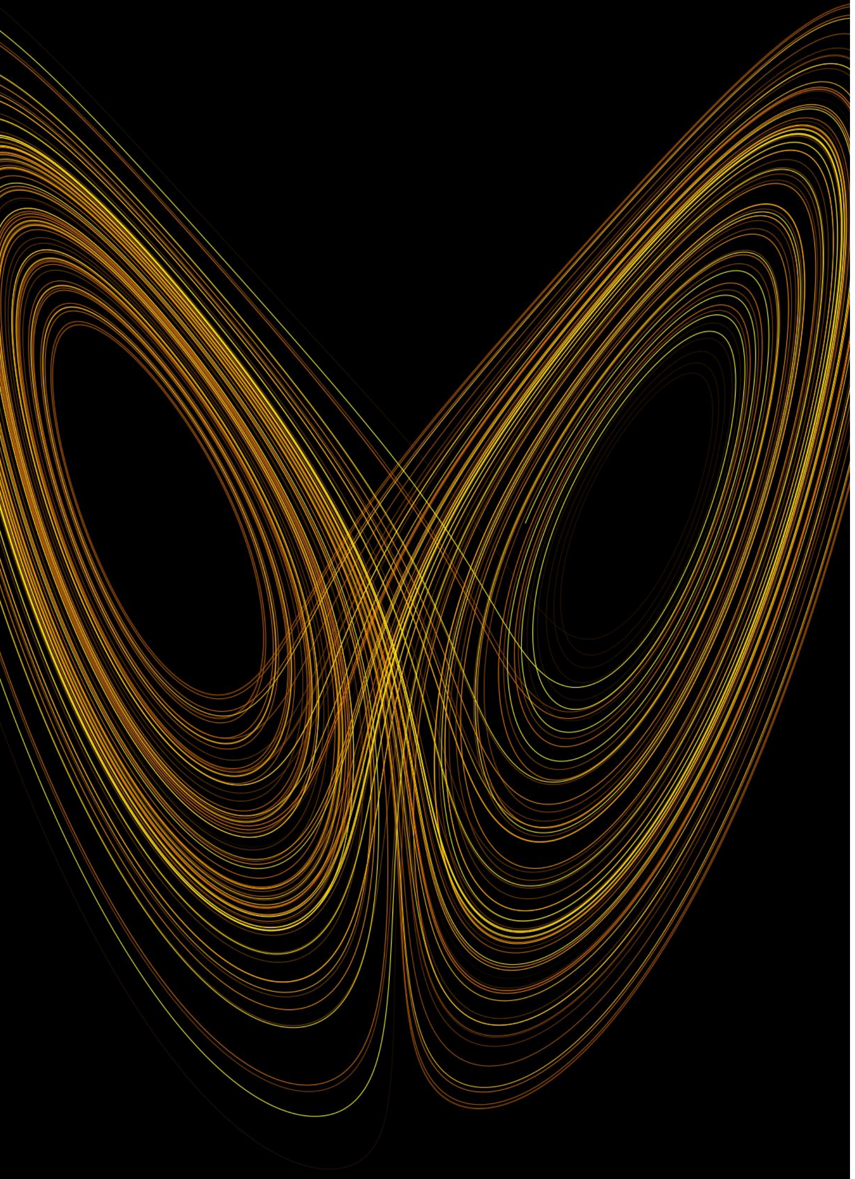


HOW

Automated alerting

- Good alerting is hard to set up
- Requires maintenance
- Requires constant improvements
- To properly scale, requires a strategy
- Differentiate between
  - Alerts - as few as possible
  - Troubleshooting data - as much as CEO is willing to pay
  - Notifications - calling you in the middle of the night
- Not all metrics need to be alerted on
- Not all alerts need to trigger notifications
- Not all alerts need to trigger end-of-the-world notifications

Be alert… the world needs more lerts.
Woody Allen



HOW

Chaos engineering

- Handy practice
- Uncovers some types of problems
- Can be used to improve reliability
- Start practicing when you think you are reliable enough

HOW

Other practices

- Fire drills, game days, disaster recovery tests
- MTTx metrics
  - Collection and analysis
  - Beware of averages
  - TTx histograms
- Service and organisation registry
  - Keep the two in sync
- Eliminating toil
  - Meta-practice, should be included in everything
- Data analysis, statistics
- Many more

> *Average: a random number that falls somewhere between the maximum and 1/2 the median. Most often used to ignore reality.*
>
> *Gil Tene*

k8s?

The Google Model
We Are Now SRE
SRE Center of Practice / Excellence
Embedded SRE
Github repo: How They SRE...

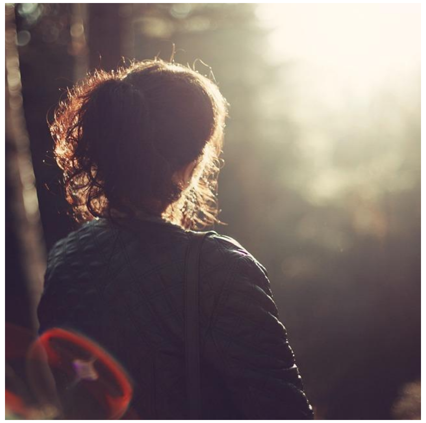
Epilogue

Start small

- Start with anything
- Start with what you have
- Practice makes perfect
- Improve constantly

Start smart

- Do not reinvent the wheel
- There’s tons of resources
- Do not ignore the history of how practices evolved and why



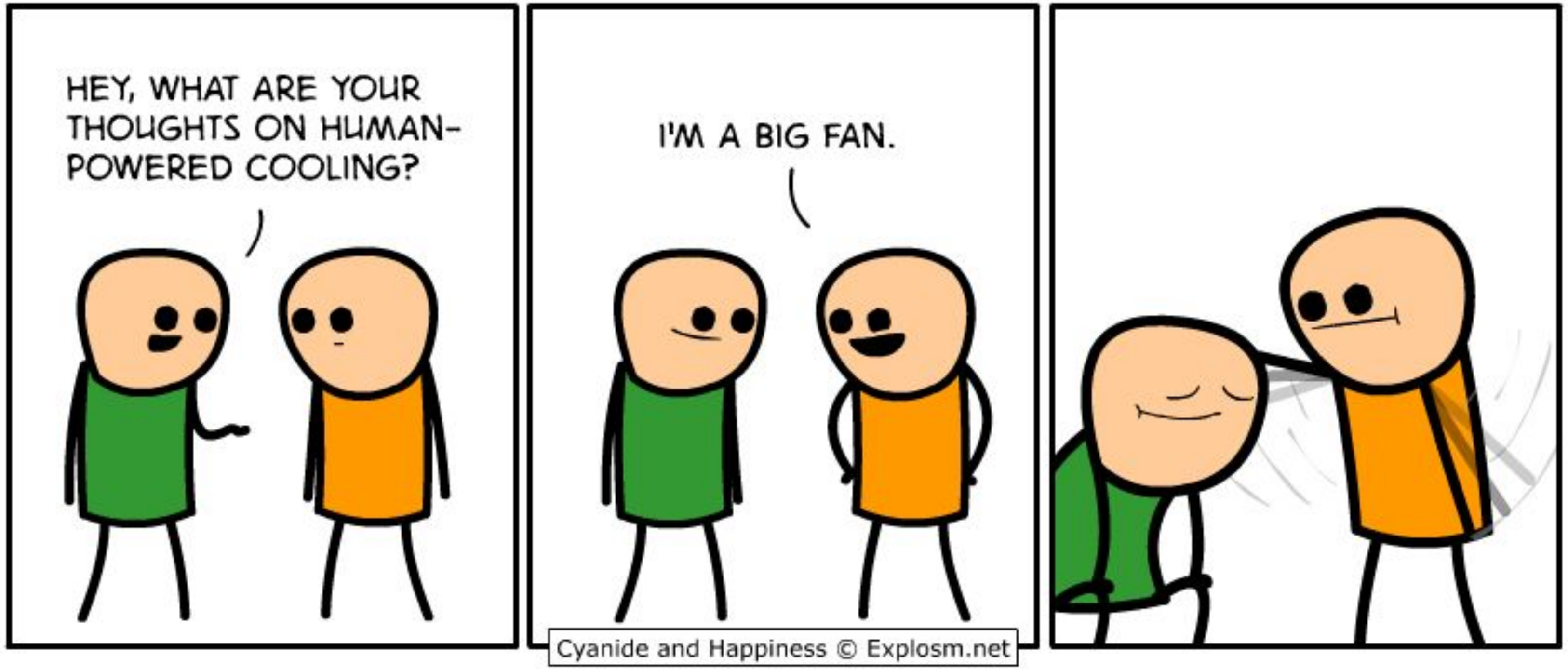
Be kind to yourself

- SRE, not a role, a condition
- Handling production is challenging
- Lots of context switching
- Lots of quick decisions
- Can be highly stressful

Be kind to others

- Communication is essential
- Lots of stakeholders and their specific dialects
- Lots of different cultures
- Make people responsible for their actions and services

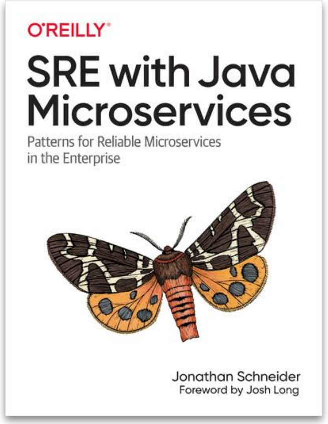
Strategy for engaging humans doing ops with something worthy of their mental capacity


HEY, WHAT ARE YOUR THOUGHTS ON HUMAN-POWERED COOLING?
I'M A BIG FAN.
Cyanide and Happiness © Explosm.net

THANK YOU!

infobip

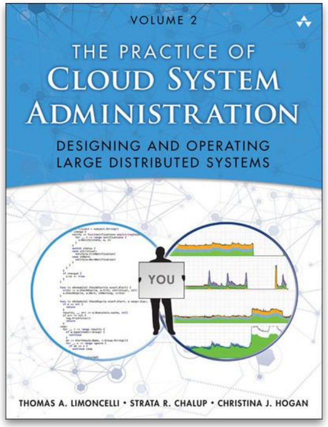
References


O'REILLY
SRE with Java Microservices
Patterns for Reliable Microservices in the Enterprise
Jonathan Schneider
Foreword by Josh Long

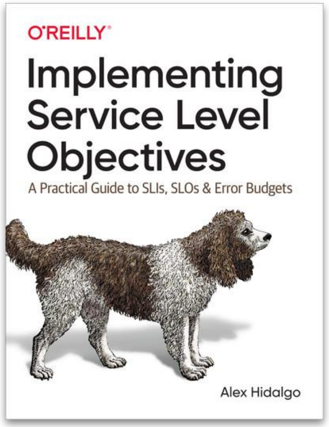
- Code-level / microservice architecture reliability
- Observability: deep-dive
- Recommended for all SW engineers

The Pragmatic Programmers
Release It!
Second Edition
Design and Deploy Production-Ready Software
Michael T. Nygard
Edited by Katharine Dvorak

- Stability patterns
- Examples of real-life failures and how to mitigate them
- Recommended for SW engineers working with distributed systems

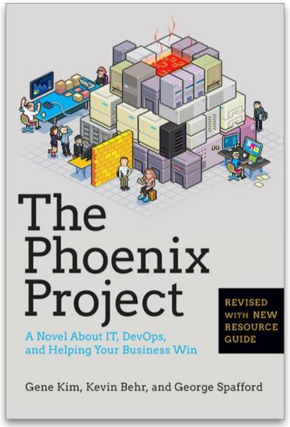

VOLUME 2
THE PRACTICE OF CLOUD SYSTEM ADMINISTRATION
DESIGNING AND OPERATING LARGE DISTRIBUTED SYSTEMS
YOU
THOMAS A. LIMONCELLI • STRATA R. CHALUP • CHRISTINA J. HOGAN

- Overview of ops required in the cloud
- Design, operate, assess, improve
- Recommended for tech-savvy managers, new / evolving Sys Amins, SREs, Devs doing Ops


O'REILLY
Implementing Service Level Objectives
A Practical Guide to SLIs, SLOs & Error Budgets
Alex Hidalgo

- All you need to know about SLOs
- Deep-dive into the subject
- Recommended for managers, architects and senior SW engineers

Newsletter: https://sreweekly.com/

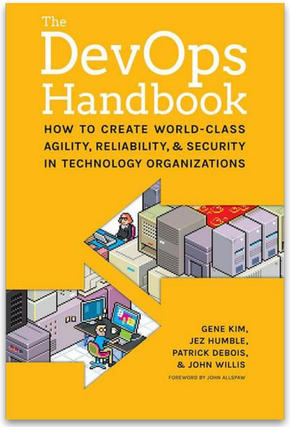


The Phoenix Project

REVISED WITH NEW RESOURCE GUIDE

A Novel About IT, DevOps, and Helping Your Business Win

Gene Kim, Kevin Behr, and George Spafford



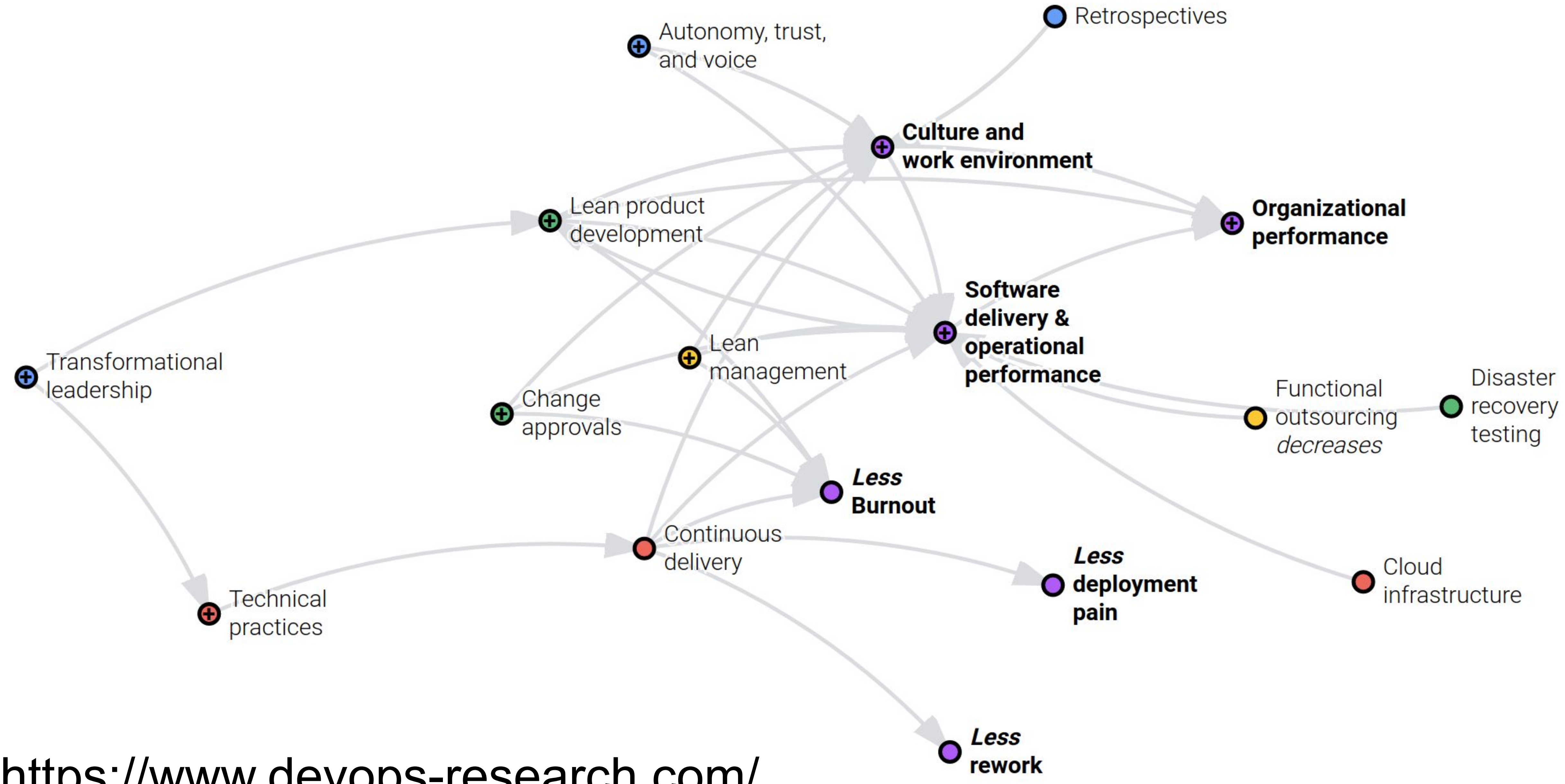
The DevOps Handbook

HOW TO CREATE WORLD-CLASS AGILITY, RELIABILITY, & SECURITY IN TECHNOLOGY ORGANIZATIONS

GENE KIM, JEZ HUMBLE, PATRICK DEBOIS, & JOHN WILLIS

FOREWORD BY JOHN ALLSPAW

- Phoenix project: DevOps explained as a fictional story
- Handbook: Why, what and how to DevOps
- Recommended to all thinking they need to hire a DevOp


Retrospectives
Autonomy, trust, and voice
Culture and work environment
Organizational performance
Lean product development
Software delivery & operational performance
Lean management
Transformational leadership
Change approvals
Functional outsourcing *decreases*
Disaster recovery testing
Less Burnout
Continuous delivery
Less deployment pain
Cloud infrastructure
Technical practices
Less rework

https://www.devops-research.com/