Don't Panic!

How to Cope Now You're Responsible for Production

Euan Finlay
@efinlay24

FT
FINANCIAL TIMES

https://www.theguardian.com/uk-news/2015/jun/03/queens-health-bbc-tweet-global-news-alert/

FT
FINANCIAL TIMES

December 3, 2015 12:36 pm

ECB leaves rates unchanged in surprising decision

Claire Jones, Frankfurt

Share Author alerts Print Clip Comments

The European Central Bank has left interest rates unchanged, dashing expectations of a cut to its deposit rate.

https://www.businessinsider.com/ft-publishes-incorrect-ecb-announcement-early-2015-12/

FT FINANCIAL TIMES

https://www.businessinsider.com/ft-publishes-incorrect-ecb-announcement-early-2015-12/

FINANCIAL TIMES

https://twitter.com/Ludo_Dufour/status/672401158653218816/

FT
FINANCIAL TIMES

DECEMBER 3, 2015

On Thursday we published an incorrect story on FT.com that stated the European Central Bank had confounded expectations by deciding to hold interest rates rather than cut them. The story was published a few minutes before the decision to cut rates was announced.

The story was wrong and should not have been published. The article was one of two pre-written stories — covering different possible decisions — which had been prepared in advance of the announcement. Due to an editing error it was published when it should not have been. Automated feeds meant that the initial error was compounded by being simultaneously published on Twitter.

The FT deeply regrets this serious mistake and will immediately be reviewing its publication and workflow processes to ensure such an error cannot happen again. We apologise to all our readers.


OBITUARIES
Commander James Bond

/usr/bin/whoami

/usr/bin/whodoiworkfor

No such file or directory.

fastFT

- Honda set to close Swindon plant in fresh blow to UK manufacturing 10M AGO
- Seven MPs resign from Labour party in challenge to Jeremy Corbyn 2H AGO
- German regulator bans shorting of Wirecard shares 3H AGO
- China leads Asia equities rally ahead of trade talks

Protectionism

EU threatens retaliation if US imposes punitive car tariffs

European Commission to 'react in a swift and adequate manner' to any levies

3 HOURS AGO

- Donald Trump likely to take his time regarding auto tariffs
- Donald Trump's ill-timed rift with Europe
- Trump administration delays decision on car tariffs

Facebook Inc

Facebook joins Amazon and Google in AI chip race

NEW 30 MINUTES AGO

Citigroup Inc

Citi CEO says machines may cut thousands of call centre jobs

Labour Party UK

Seven MPs resign from Labour in challenge to Corbyn

Split by pro-EU moderates comes

Federal Reserve

Fed nears decisions on its asset portfolio

Some officials think balance sheet

Automobiles

Honda set to close Swindon plant in blow to UK manufacturing

https://www.ft.com/

FT FINANCIAL TIMES

Your team is now on call.

And you're mildly terrified.

Obligatory audience interaction.

Everyone feels the same when they start out.

I still do today.

How do you get comfortable with supporting production?

The Ghosts of Incidents...

> Future

FT
FINANCIAL
TIMES

The Ghosts of Incidents...

Future

> Present

FT
FINANCIAL TIMES

The Ghosts of Incidents...

Future

Present

> Past

FT

FINANCIAL TIMES


The Ghost of Incidents Future

Handling incidents is the same as any other skill.

Get comfortable with your alerts.

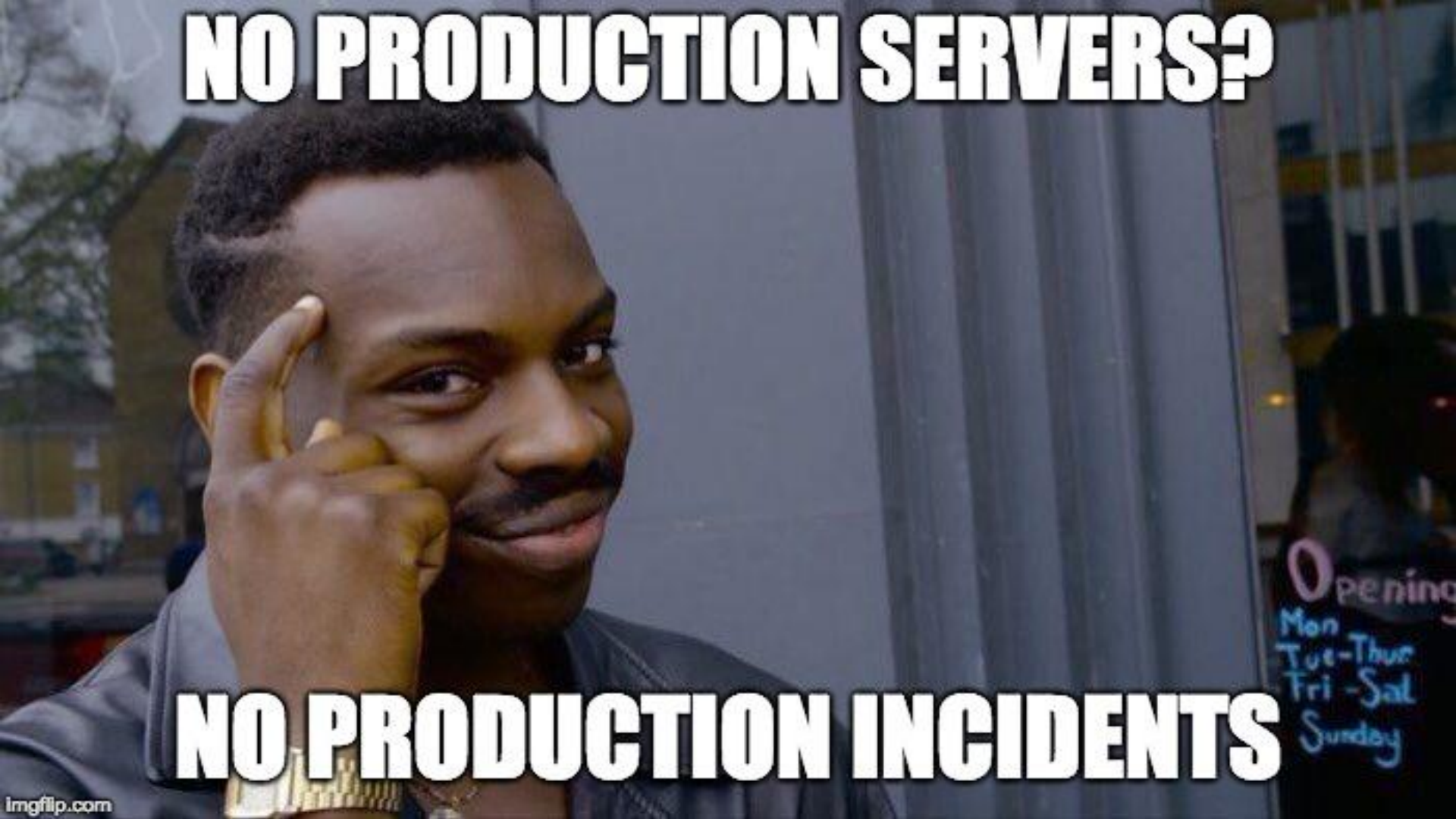
Delete alerts you don't care about.

Have a plan for when things break.

Keep your documentation up to date.

Practice regularly.

"The Gang Deletes Production"

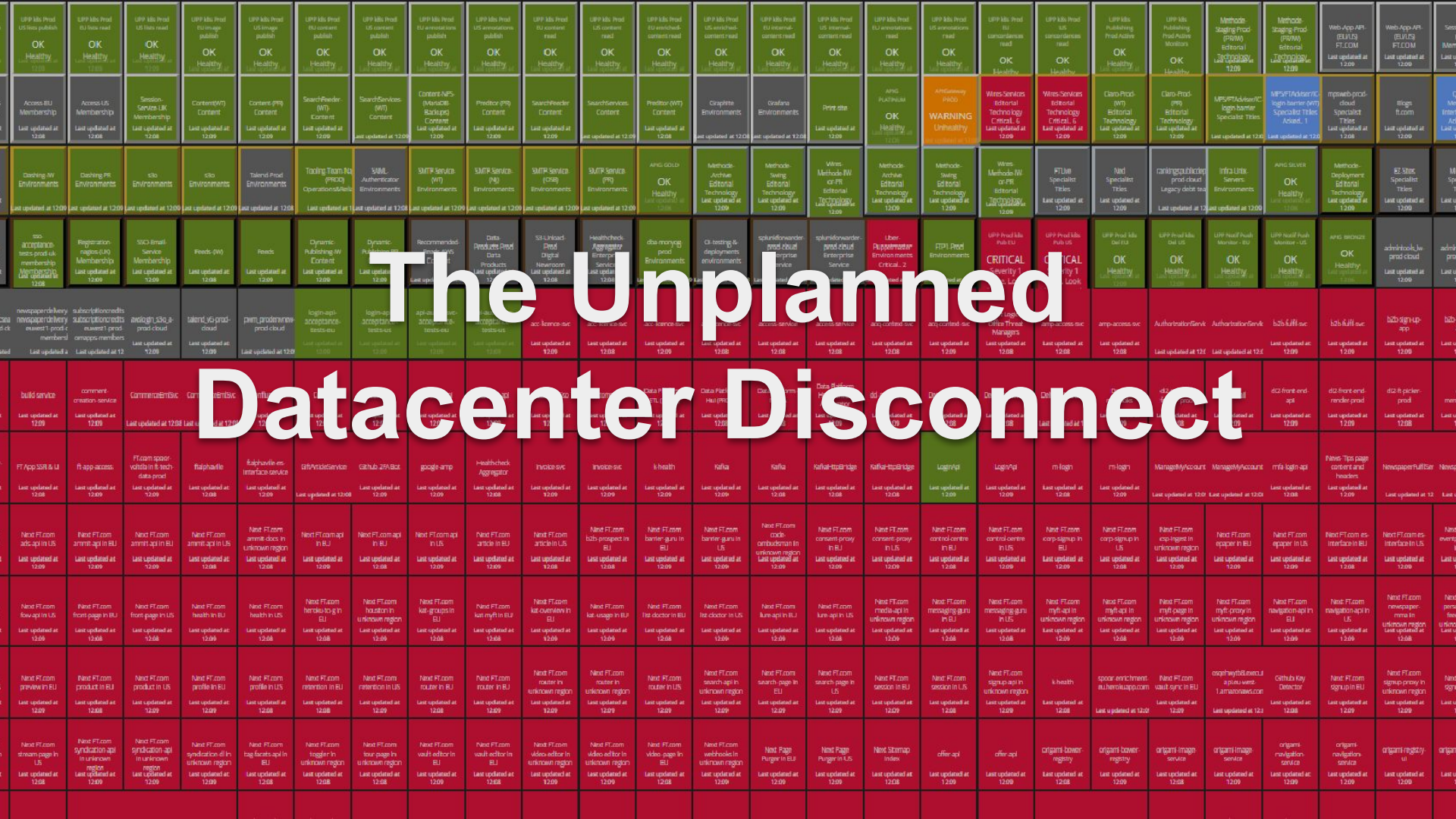

NO PRODUCTION SERVERS?
NO PRODUCTION INCIDENTS
Opening
Mon
Tue-Thur
Fri -Sat
Sunday
imgflip.com

Break things, and see what happens.

Did your systems do what you expected?


The Planned
Datacenter Disconnect

We got complacent, and stopped running datacenter failure tests...


The Unplanned
Datacenter Disconnect

Have a central place for reporting changes and problems.

[redacted] 19:25

Seeing aws dx link issues again-checking

Pasted image at 2018-07-27, 5:28 PM

[redacted] 👀 19:39

Methode alerts are firing

intermittent

[redacted] 19:40

yep we have network issues again at PR (edited)

😢 1

Looks like the MPLS Verizon cct is down

So far no impact reported...

monitoring for now

[redacted] 🇦🇺 19:44

We have reports of publishing not working, and problems with Methode portalpub connecting to UPP again

[redacted] 19:45

thanks [redacted]

^^Verizon are saying PR site is affected by an issue affecting multiple locations

[redacted] 👀 19:46

switching portalpub off in PR

We're not perfect.

But we always try to improve.

The Ghosts of Incidents...

Future

> Present

Past

FT
FINANCIAL TIMES


The Ghost of Incidents Present

Calm down, and take a deep breath.

It's probably ok.

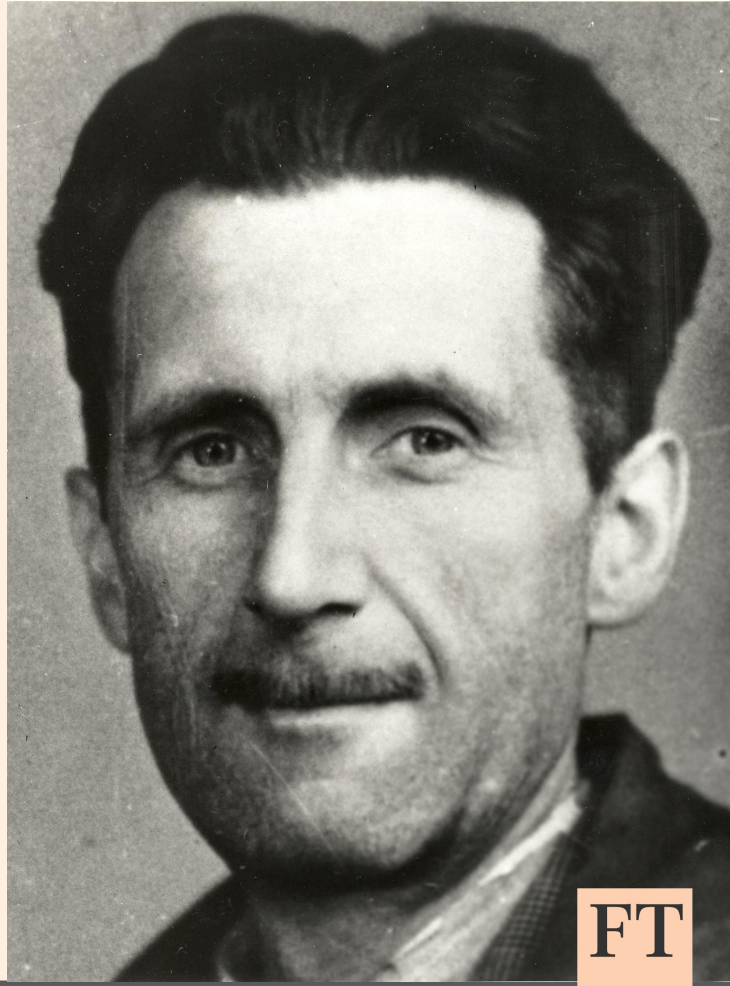
Don't dive straight in.

Go back to first principles.

What's the actual impact?

"All incidents are **equal**, but some incidents are **more equal** than others."

George Orwell, probably.



What's already been tried?


ENERGY

Is there definitely a problem?

DYNPUB / EXTRAPUB Prod Park Royal Content

DYNPUB / EXTRAPUB Prod Watford Content

SEMANTIC Prod Park Royal Content

SEMANTIC Prod Watford Content

SEMANTIC Prod AWS Content Downtime: 5

DON'T PANIC

CD-CD Prod US

OK

Healthy

MASHERY-V2-TESTS-PROD-UK

MASHERY-V2-TESTS-PROD-US

DON'T WORRY

EVERYTHING IS FINE

PROBABLY OK

SEMANTIC Test Park Royal Acted...k

SEMANTIC Test Watford

NOT CRTICAL

PAT_TEST

0

DYNPUB / EXTRAPUB Int Park Royal Content

SEMANTIC Int Park Royal

PAT_TEST

0

What's the minimum viable solution?

Get it running *before* you get it fixed.

Go back to basics.

It's not DNS

There's no way it's DNS

It was DNS

-SSBroski


THIS IS FINE.

Don't be afraid to call for help.


The One Where a Director
Falls Through the Ceiling

Communication is key.

Especially to our customers.

Put someone in charge.

Pingdom APP 14:42

k8s - UPP Prod Delivery US: Annotations Read Aggregate Healthcheck is down (Incident #4073269)
upp-prod-delivery-us.ft.com • View details

k8s - UPP Prod Delivery UK: Annotations Read Aggregate Healthcheck is down (Incident #4073041)
upp-prod-delivery-eu.ft.com • View details

k8s - UPP Prod Delivery UK: Content Read Aggregate Healthcheck is down (Incident #4073077)
upp-prod-delivery-eu.ft.com • View details

k8s - UPP Prod Delivery US: Content Publish Aggregate Healthcheck is down (Incident #4073290)
upp-prod-delivery-us.ft.com • View details

k8s - UPP Prod Delivery US: Annotations Read Aggregate Healthcheck is up (Incident #4073269)
upp-prod-delivery-us.ft.com • View details

k8s - UPP Prod Delivery US: Image Publish Aggregate Healthcheck is down (Incident #4073407)
upp-prod-delivery-us.ft.com • View details

k8s - UPP Prod Delivery UK: Image Publish Aggregate Healthcheck is down (Incident #4073095)
upp-prod-delivery-eu.ft.com • View details

Software can be chaotic, but we make it work

Expert

Trying Stuff Until it Works

O RLY?

The Practical Developer
@ThePracticalDev

Create a temporary incident channel.

[redacted] 11:26

Same for `EA460F68-495E-D18F-3130-C220FA23EF8E` in prod-us at 04:01:57

[redacted] 11:26

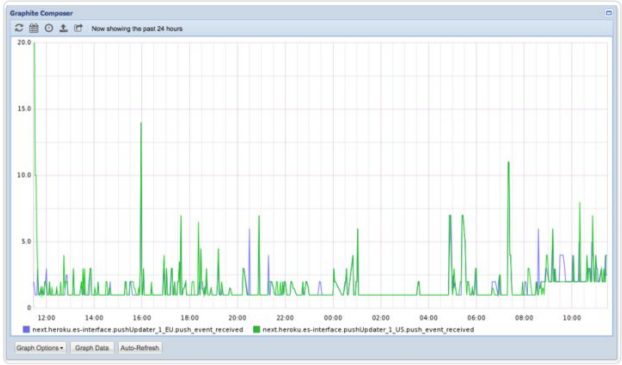
So [redacted] says that in the last 24 hours, Next have got 20% more push event notifications in the US than in the EU...

[redacted] 11:26

Both request times line up exactly with the notification timestamps

[redacted] 11:27

uploaded and commented on this image: **Push notifications received**



> green is US region

[redacted] 11:28

`suggestions-rw-neo4j` (that's v2 annotations as I understand it) wrote annotations at 04:01:36 (uk) and 04:01:51 (us)

So they would have been present before the notifications were sent out

[redacted] 11:29

[redacted] and I have confirmed that no annotations will still allow the article to show in both enrichedcontent and internalcontent

😞 1

Hmm, that's a big gap between 1 and 5

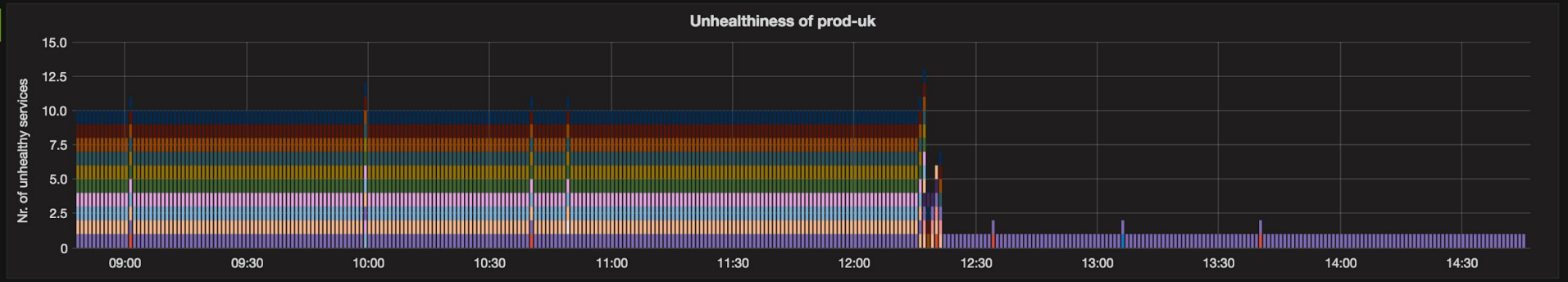
If you think you're
over-communicating,
it's probably just the right amount.

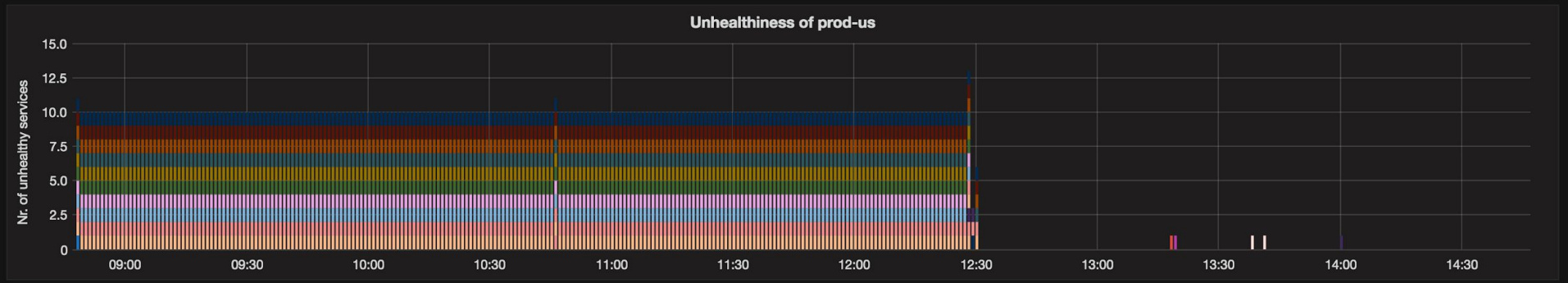
Tired people don't think good.

CoCo Health Timeline

Zoom Out | Last 6 hours | Refresh every 1m

environment: prod-uk + prod-us + pub-prod-uk | services: All





"The Gang Serves Traffic From Staging"

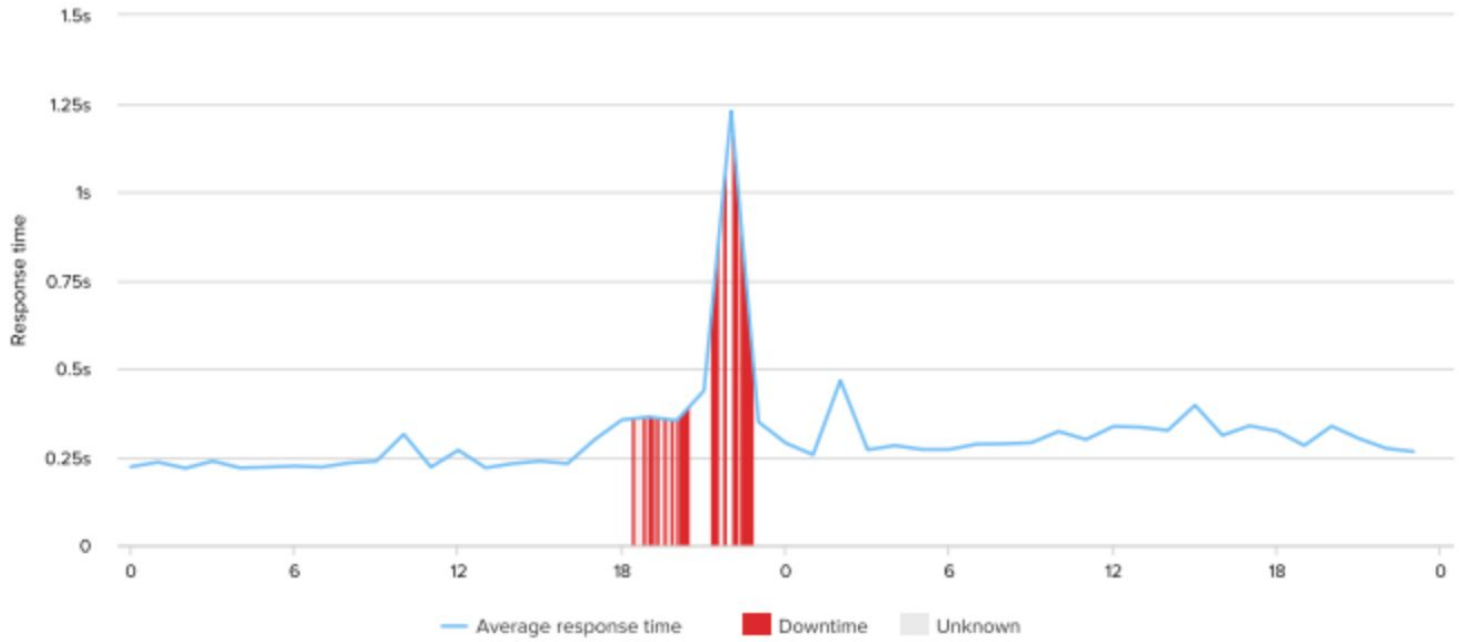
Uptime report

Report dates: 26/07/2016 to 27/07/2016 (Europe/Lo

Content API V2 Lists API [redacted] Type: HTTP

Response time

Downtime	Outages	Uptime	Max resp. time	Min resp. time	Avg resp. time
1h51m	21	96.15%	1.23s	218ms	302ms



It wasn't great,
but it wasn't the end of the world.

The Ghosts of Incidents...

Future

Present

> Past

FT

FINANCIAL TIMES


The Ghost of Incidents Past

Congratulations!
You survived.

It probably wasn't **that** bad, was it?

Run a learning review with everyone involved.

Incident reports are important.

NEVER HAVE I FELT SO CLOSE TO ANOTHER SOUL

AND YET SO HELPLESSLY ALONE

AS WHEN I GOOGLE AN ERROR

AND THERE'S ONE RESULT

A THREAD BY SOMEONE WITH THE SAME PROBLEM

AND NO ANSWER

LAST POSTED TO IN 2003

Feb 10, 2017 - GitLab

Postmortem of database outage of January 31

Postmortem on the database outage of January 31 2017 with the lessons we learned.

← Back to company

On January 31st 2017, we experienced a major service outage for one of our products, the online service GitLab.com. The outage was caused by an accidental removal of data from our primary database server.

This incident caused the GitLab.com service to be unavailable for many hours. We also lost some production data that we were eventually unable to recover. Specifically, we lost modifications to database data such as projects, comments, user accounts, issues and snippets, that took place between 17:20 and 00:00 UTC on January 31. Our best estimate is that it affected roughly 5,000 projects, 5,000 comments and 700 new user accounts. Code repositories or wikis hosted on GitLab.com were unavailable during the outage, but were not affected by the data loss. GitLab Enterprise customers, GitHost customers, and self-hosted GitLab CE users were not affected by the outage, or the data loss.

https://about.gitlab.com/2017/02/10/postmortem-of-database-outage-of-january-31/

FT FINANCIAL TIMES

"Until a restore is attempted, a backup is both **successful** and **unsuccessful**."

Erwin Schrödinger

Timeline

On January 31st an engineer started setting up multiple PostgreSQL servers in our staging environment. The plan was to try out pgpool-II to see if it would reduce the load on our database by load balancing queries between the available hosts. Here is the issue for that plan: infrastructure#259.

± 17:20 UTC: prior to starting this work, our engineer took an LVM snapshot of the production database and loaded this into the staging environment. This was necessary to ensure the staging database was up to date, allowing for more accurate load testing. This procedure normally happens automatically once every 24 hours (at 01:00 UTC), but they wanted a more up to date copy of the database.

± 19:00 UTC: GitLab.com starts experiencing an increase in database load due to what we suspect was spam. In the week leading up to this event GitLab.com had been experiencing similar problems, but not this severe. One of the problems this load caused was that many users were not able to post comments on issues and merge requests. Getting the load under control took several hours.

We would later find out that part of the load was caused by a background job trying to remove a GitLab employee and their associated data. This was the result of their account being flagged for abuse and accidentally scheduled for removal. More information regarding this particular problem can be found in the issue "Removal of users by spam should not hard delete".

Publication of the outage

In the spirit of transparency we kept track of progress and notes in a publicly visible Google document. We also streamed the recovery procedure on YouTube, with a peak viewer count of around 5000 (resulting in the stream being the #2 live stream on YouTube for several hours). The stream was used to give our users live updates about the recovery procedure. Finally we used Twitter (https://twitter.com/gitlabstatus) to inform those that might not be watching the stream.

The document in question was initially private to GitLab employees and contained name of the engineer who accidentally removed the data. While the name was added by the engineer themselves (and they had no problem with this being public), we will redact names in future cases as other engineers may not be comfortable with their name being published.

Data loss impact

Database data such as projects, issues, snippets, etc. created between January 31st 17:20 UTC and 23:30 UTC has been lost. Git repositories and Wikis were not removed as they are stored separately.

It's hard to estimate how much data has been lost exactly, but we estimate we have lost at least 5000 projects, 5000 comments, and roughly 700 users. This only affected users of GitLab.com, self-hosted instances or GitHost instances were not affected.

Identify what can be improved for next time.

Improving recovery procedures

We are currently working on fixing and improving our various recovery procedures. Work is split across the following issues:

1. Overview of status of all issues listed in this blog post (#1684)
2. Update PS1 across all hosts to more clearly differentiate between hosts and environments (#1094)
3. Prometheus monitoring for backups (#1095)
4. Set PostgreSQL's max_connections to a sane value (#1096)
5. Investigate Point in time recovery & continuous archiving for PostgreSQL (#1097)
6. Hourly LVM snapshots of the production databases (#1098)
7. Azure disk snapshots of production databases (#1099)
8. Move staging to the ARM environment (#1100)
9. Recover production replica(s) (#1101)
10. Automated testing of recovering PostgreSQL database backups (#1102)
11. Improve PostgreSQL replication documentation/runbooks (#1103)
12. Investigate pgbarman for creating PostgreSQL backups (#1105)
13. Investigate using WAL-E as a means of Database Backup and Realtime Replication (#494)
14. Build Streaming Database Restore
15. Assign an owner for data durability

https://about.gitlab.com/2017/02/10/postmortem-of-database-outage-of-january-31/

Nearly the end.

Don't clap yet.

Failure is inevitable.

And that's ok.

The end.

"Please clap."
Jeb Bush, 2016

@efinlay24

euan.finlay@ft.com

We're hiring!
https://ft.com/dev/null/

FT
FINANCIAL TIMES