

# Don't Panic!

## How to Cope Now You're Responsible for Production

Euan Finlay

@efinlay24 | #DevReach2019







Ahmen Khawaja @AhmenKhawaja · 2 hrs

BREAKING: Queen Elizabeth is being treated at King Edward 7th Hospital in London. Statement due shortly:

[@BBCWorld](#)

← ↻ 7 ★ 1 👤 ⋮



Ahmen Khawaja @AhmenKhawaja · 2 hrs

"Queen Elizabeth has died": [@BBCWorld](#)

← ↻ ★ 👤 ⋮



Ahmen Khawaja @AhmenKhawaja · 2 hrs

False Alarm to Queen's death! She is being treated at King Edward 7th hospital. Statement due shortly

← ↻ 1 ★ 👤 ⋮



Ahmen Khawaja @AhmenKhawaja · 2 hrs

False Alarm: Have deleted previous tweets!!

← ↻ ★ 👤 ⋮

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.



**Financial Times** ✓

@FinancialTimes



Follow

ECB leaves rates unchanged in shock decision [on.ft.com/1Nrekqz](https://on.ft.com/1Nrekqz)

RETWEETS

36

LIKES

6



7:38 AM - 3 Dec 2015





03 Déc 2015  
 Australia 61 2 9777 8600 Brazil 5511 2395 9000 Europe 44 20 7330 7500 Germany 49 69 9204 1210 Hong Kong 852 2977 6000  
 Japan 81 3 3201 8900 Singapore 65 6212 1000 U.S. 1 212 318 2000  
 Copyright 2015 Bloomberg Finance L.P.  
 SN 204778 CET GMT+1:00 H216-2408-2 03-Dec-2015 13:58:36





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

Royal Navy and British Secret Service

secret service agent James Bond  
collaborator Wei Lin of the  
China's External Security Force  
this morning in





**/usr/bin/whoami**

@efinlay24

# `/usr/bin/whodoiworkfor`

No such file or directory.

@efinlay24

Brexit

## Johnson aims to pass Brexit deal this week

Foreign secretary Raab says cross-party coalition of at least 320 MPs are likely to back agreement

● NEW 22 MINUTES AGO

- Johnson can still see the path to Brexit victory
- Johnson seeks Brexit delay from EU after MPs delay key vote
- Boris Johnson weighs next Brexit move after Super Saturday setback



Analysis **The Big Read**

### Fidelity's search for the technology of tomorrow



**Donald Trump**

Trump drops plans to host G7 at his Miami resort



**Goldman Sachs Group**

Goldman Sachs banker charged with insider trading



**Reliance Industries Ltd**

India's Reliance eyes sports streaming rights



**Syrian crisis**

Ankara counts itself a victor after Syria incursion

**1211**

**Production Systems**

**FT**

FINANCIAL  
TIMES

# 245

## Platinum Systems

~150

Daily Releases

**~150**

**(including Fridays)**

**FT**

FINANCIAL  
TIMES

**60+**

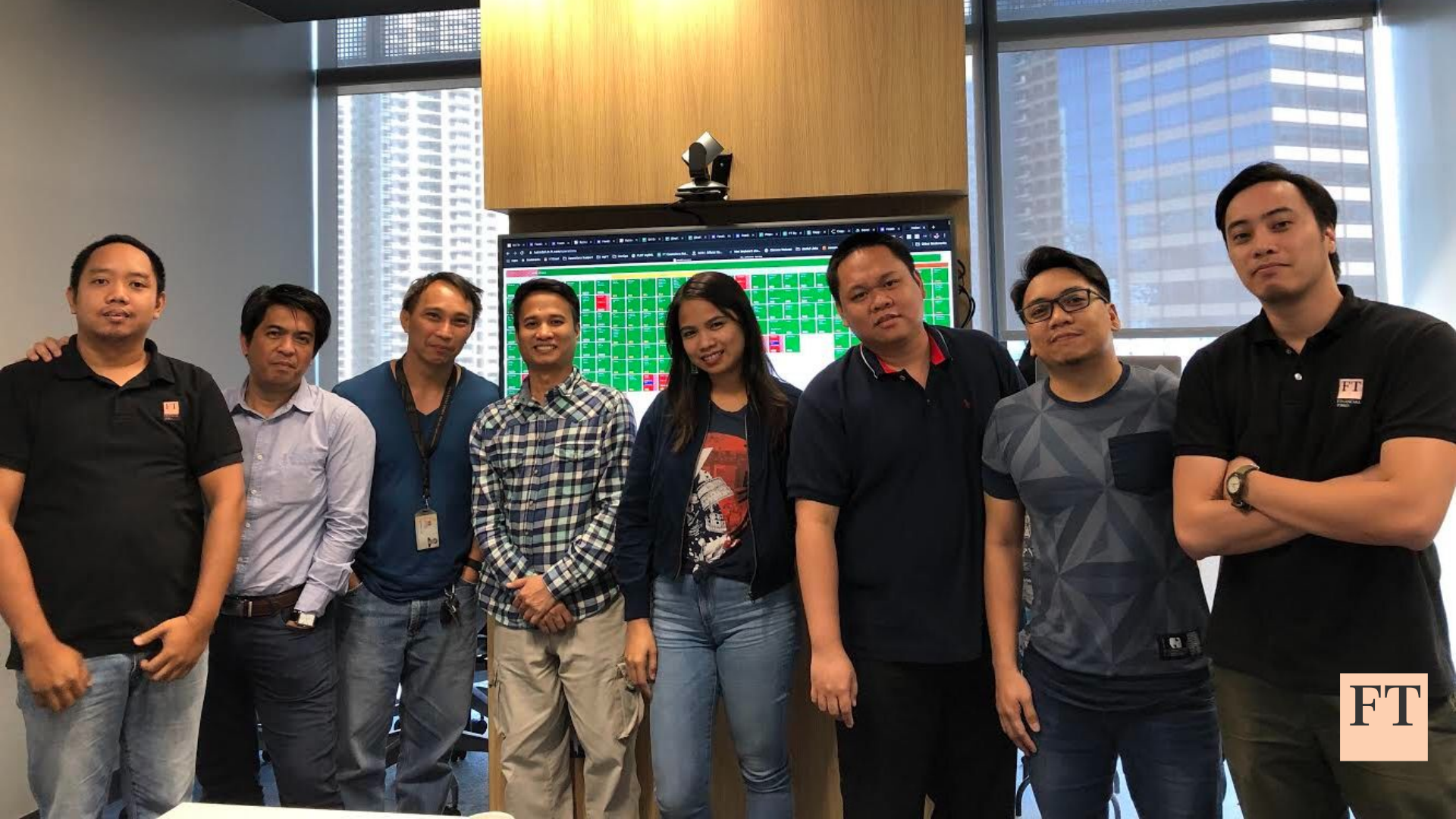
**Third-Party Providers**











# Your team is now on call.

And you're mildly terrified.

@efinlay24

# Obligatory audience interaction.

@efinlay24

**Everyone feels the same  
when they start out.**

I still do today.

@efinlay24

**How do we get comfortable  
with supporting production?**

@efinlay24





# The Ghosts of Incidents...

> Future

# The Ghosts of Incidents...

Future

> Present

# The Ghosts of Incidents...

Future  
Present  
> Past



# The Ghosts of Incidents...

> Future

Present

Past

**Handling incidents is the  
same as any other skill.**

@efinlay24

**Get comfortable  
with your alerts.**

@efinlay24



**Delete the alerts  
you don't care about.**

@efinlay24

**Have a plan for  
when things break.**

@efinlay24

**Keep your documentation  
up to date.**

@efinlay24

## System: FT.com Article Page

[General information](#)[Ownership & knowledge](#)[Technical overview](#)[Data governance](#)[Related resources](#)[Failover](#)[Data recovery](#)[Release](#)[Key Management](#)[Monitoring](#)[Troubleshooting](#)[More Information](#)[Service Operability Review](#)[Miscellaneous](#)

## System: FT.com Article Page

Show the definition of "System" ▼

[View runbook](#)[View Heimdall dashboard](#)[View SOS rating](#)

### General information ✎

[Delete](#)[Edit](#)

**Code** ⓘ

**Name** ⓘ

**Description** ⓘ

**Primary URL** ⓘ

**Service tier** ⓘ

**Lifecycle stage** ⓘ

### Ownership & knowledge ✎

**Delivered by team** ⓘ

**Supported by team** ⓘ

# Biz Ops

System: FT.com Article Page

General information

Ownership & knowledge

Configuration

Data Governance

Related resources

Failover

Data recovery

Release

Key Management

Monitoring

Troubleshooting

More Information

Service Operability Review

Miscellaneous

System: FT.com Article Page

Show the definition of "System" ▾

View runbook

View Heimdall dashboard

View SOS rating

General information

Code

Name

Description

Primary url

Service tier

Lifecycle stage

Ownership & knowledge

Delivered by team

Supported by team

FT.com Article Page

FT.com Article Page

A service that provides the user-facing layer of articles for FT.com

https://www.ft.com/content/01b2e311-e2b0-11e8-a6e5-792428919cee

Premium

Production

Next

Next

Delete

Edit

The central place to find info on all of the FT's systems, products and teams.

## FT.com Article Page

Basic Information

Contact Information

Related systems/products

Technical overview

Monitoring

**Troubleshooting**

Failover

Data recovery

Release

Key Management

More Information

# Troubleshooting ✍

## First line troubleshooting

Below are a list of possible issues and troubleshooting steps to mitigate them:

## Error pages being returned on some or all article pages

Multiple article pages are returning a 404, 503, or other [FT.com](#) error page

- Are there an unusually large number of requests to the site? this could be a DDOS attack ([Grafana](#) logs will tell you this). If this is the case see the Customer Products [cyber attack panic guide](#)
- Check whether ([Elasticsearch](#) or [UPP](#)) are experiencing issues - if so, follow troubleshooting steps for those systems
- If Elasticsearch and UPP are fine the problem is probably with [FT.com](#). Is there a stable region that you can [failover](#) to? If so, do that, and then notify Second Line support
- Has there been a new release of the `next-article` app in the last hour? If so, try [rolling back the release](#). \*NB If it is within working hours it is preferable to contact Second Line to do this
- If none of the above have solved the issue then contact Second Line support

# Contact Information

## Support and delivery team

For support in and out of office hours. This team builds and supports the system

<b>Team name</b>	Next
<b>Slack</b>	ft-next-support
<b>Email</b>	[REDACTED]
<b>Phone</b>	See support rota linked below
<b>Support rota</b>	[REDACTED]
<b>Preference</b>	[REDACTED]
<b>Tech leads</b>	<ul style="list-style-type: none"><li>• [REDACTED] <a href="#">more details</a></li><li>• [REDACTED] <a href="#">more details</a></li><li>• [REDACTED] <a href="#">more details</a></li></ul>





runbook.md ×



runbook.md > abc # FT.com Article Page > abc ## Known About By

```
1 # FT.com Article Page
2
3 A service that provides the user facing layer of articles for FT.com.
4
5 ## Primary URL
6
7 https://www.ft.com/content/00b2e3c8-e2b0-11e8-a6e5-792428919cee
8
9 ## Service Tier
10
11 Platinum
12
13 ## Lifecycle Stage
14
15 Production
16
17 ## Delivered By
18
19 next
20
21 ## Supported By
22
23 next
24
```



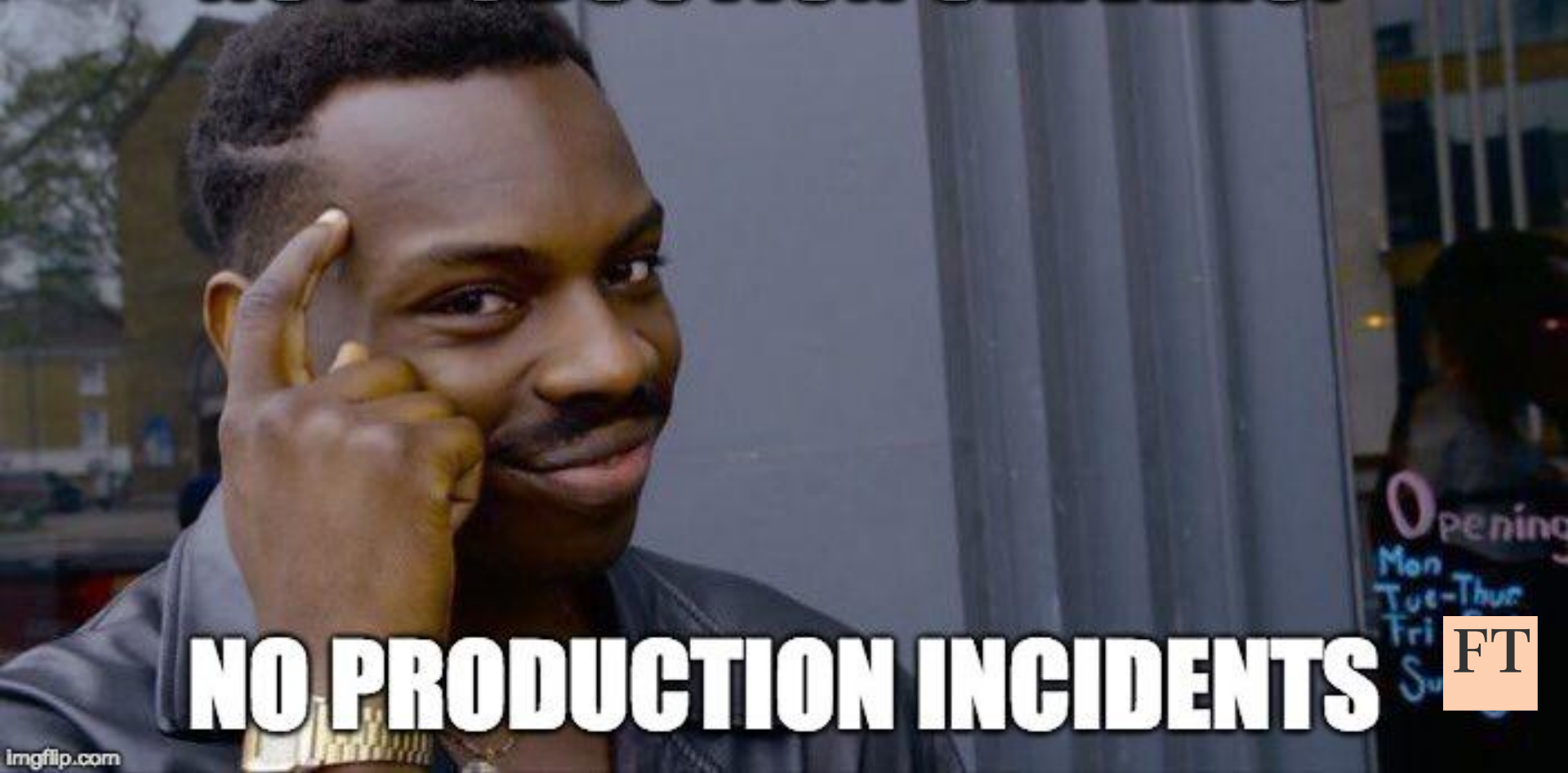


**Practice regularly.**

@efinlay24

***"The Gang Deletes  
Production"***

**NO PRODUCTION SERVERS?**



**NO PRODUCTION INCIDENTS**

**FT**

# Break things, and see what happens.

Did your systems do what you expected?

@efinlay24



# The Planned Datacenter Disconnect

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

@efinlay24





[REDACTED] 28 Jul 2018 at 10:13

I'm sorry, I'm on my phone, in the car - not much more investigation I can do on this!





Twitter feed content:

- Twitter Post 1: [Faded text]
- Twitter Post 2: [Faded text]
- Twitter Post 3: [Faded text]
- Twitter Post 4: [Faded text]
- Twitter Post 5: [Faded text]
- Twitter Post 6: [Faded text]
- Twitter Post 7: [Faded text]
- Twitter Post 8: [Faded text]
- Twitter Post 9: [Faded text]
- Twitter Post 10: [Faded text]

**Have a central place for  
reporting changes and problems.**

@efinlay24



19:25  
Seeing aws dx link issues again-checking

Pasted image at 2018-07-27, 5:28 PM ▾



19:39  
Methode alerts are firing

intermittent



19:40  
yep we have network issues again at PR (edited)



Looks like the MPLS Verizon cct is down

So far no impact reported...

monitoring for now



19:44  
We have reports of publishing not working, and problems with Methode portalpub connecting to UPP again



19:45  
thanks

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



19:46  
switching portalpub off in PR

**We're not perfect.**

But we always try to improve.

@efinlay24

## How we respond to incidents



# Response

As is the case in almost any industry, we have incidents at Monzo. While this sounds a little scary, the term "incident" just means something going wrong or not working as expected.

Incidents can happen anywhere, and cover everything from office building issues, to technology outages that impact our customers. We can't stop incidents from happening, but we can make sure we're ready to deal with them. And we can use them as a way to learn more about how things really work.



<https://monzo.com/blog/2019/07/08/h>

In previous posts we've shared [How we monitor Monzo](#) so we know what's going on with our systems, and now we structure our [on-call](#)

## How we respond to incidents

# An easy way to report technology problems that could affect the business.

As the digital world of any industry, we have incidents at Monzo. While this sounds a little scary, the term "incident" just means something going wrong or not working as expected.

Incidents can come anywhere, and come in everything from office building issues, to technology outages that impact our customers.

We can't stop incidents from happening, but we can make sure we're ready to deal with them. And we can use them as a way to learn more about how things really work.

In previous posts we've shared [How we monitor Monzo](#) so we know what's going on with our systems, and now we structure our [on-call](#)



<https://monzo.com/blog/2019/07/08/h>



Jump to...

Threads

Shared channels

# slack-corp-shared

Channels

# announcements

# anyone-for-pintz

# bracken-bakes

# brackenhous

# btlg

# engineering

# ft-boardroom

# ft-snowmates

# ft-sofia

# ft-tech-incidents

# ft-test-incidents

# ftlearning

# general

# inc-archimedes-files-mis...

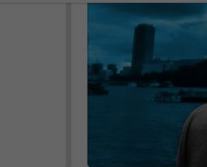
# inc-money-cover-page-t...

# inc-oracle-ipm-invoice-c...

# incident-slackbot-poc

# integration-guild

# ops-and-rel



Completed CRO

Fyi... Live blog is

Thanks

International Ed

UK Edition clear



### Report an Incident



#### Report

What's the problem?

See the User Guide for more info: <https://bit.ly/ft-response-user-guide>

#### Summary (optional)

Can you share any more details?

#### Impact (optional)

Who or what might be affected?

Think about affected people, systems, and processes

#### Incident Lead (optional)

Choose an option...

The person currently handling incident coordination & communication

#### Severity (optional)

Choose an option...

again for all the help! Really appreciate it"

**Write code  
*you can fix at 3am***





# The Ghosts of Incidents...

Future

> Present

Past

**Calm down,  
and take a deep breath.**

It's probably ok.

@efinlay24

# Don't dive straight in.

Go back to first principles.

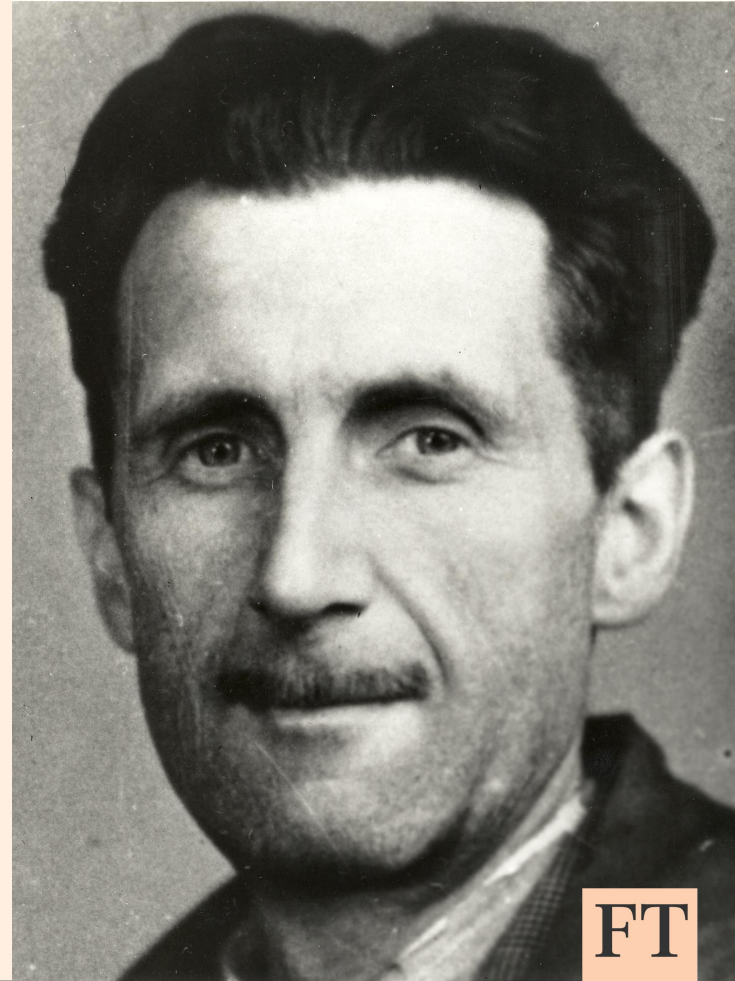
@efinlay24

# What's the actual impact?

@efinlay24

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

George Orwell, probably.



# What's already been tried?

@efinlay24





**Is there definitely a  
problem?**

@efinlay24

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

LAST UPDATED ON 10/21

LAST UPDATED ON 10/21

LAST UPDATED ON 10/21

LAST UPDATED ON 10/21

LAST UPDATED ON 10/21

OS/CO Prod US

OK  
Healthy

MASHERV2-TESTS-PROD-UK



LAST UPDATED ON 10/21

MASHERV2-TESTS-PROD-US



LAST UPDATED ON 10/21

DON'T  
WORRY

EVERYTHING  
IS  
FINE

PROBABLY  
OK

SEMANTIC Test Park Royal  
Acimed... 0

SEMANTIC Test Watford

NOT  
CRITICAL

PAT - TEST



DYNPUB / EXTRAPUB Int Park Royal  
Content

SEMANTIC Int Park Royal

LAST UPDATED ON 10/21

LAST UPDATED ON 10/21

LAST UPDATED ON 11/22

LAST UPDATED ON 11/22

LAST UPDATED ON 11/21

PAT - TEST



LAST UPDATED ON 11/22



# What's the Minimum Viable Solution?

@efinlay24

**Get it running  
before you get it fixed.**

@efinlay24

**Go back to basics.**

@efinlay24

*It's not DNS*

*There's no way it's DNS*

*It was DNS*

*-SSBroski*



FT



**Don't be afraid to call  
for help.**

@efinlay24



## TEAMS

## Guide: Understand team effectiveness

## ● Introduction

● Define what makes a team effective

● Define "high performing teams"

● Collect data and measure effectiveness

● Identify dynamic factors of effective teams

🔍 Tool: Help teams determine their own needs

🔍 Tool: Foster psychological safety

● Help teams take action

## Introduction

Much of the work done at Google, and in many organizations, is done collaboratively by teams. The team is the molecular unit where real production happens, where innovative ideas are conceived and executed. Good teams are empowered, experienced, composed of like-minded people, but it's also where interpersonal issues, ill-suited skill sets, and unclear group goals can hinder productivity and cause friction.

Following the success of [Google's Project Oxygen research](#) where the People Analytics team studied [what makes Google managers effective](#), Google researchers applied a similar method to discover the secrets of effective teams at Google. Code-named [Project Aristotle](#) - a tribute to Aristotle's quote, "the whole is greater than the sum of its parts" (as the Google researchers believed employees can do more working together than alone) - the goal was to answer the question: "What makes a team effective at Google?"

Read about the researchers behind the work in [The New York Times: What Google Learned From Its Quest to Build the Perfect Team](#)

# What makes an effective team at Google?

1

## Psychological Safety

Team members feel safe to take risks and be vulnerable in front of each other.

2

## Dependability

Team members get things done on time and meet Google's high bar for excellence.

3

## Structure & Clarity

Team members have clear roles, plans, and goals.

4

## Meaning

Work is personally important to team members.

5

## Impact

Team members think their work matters and creates change.

re:Work

FT

FINANCIAL  
TIMES




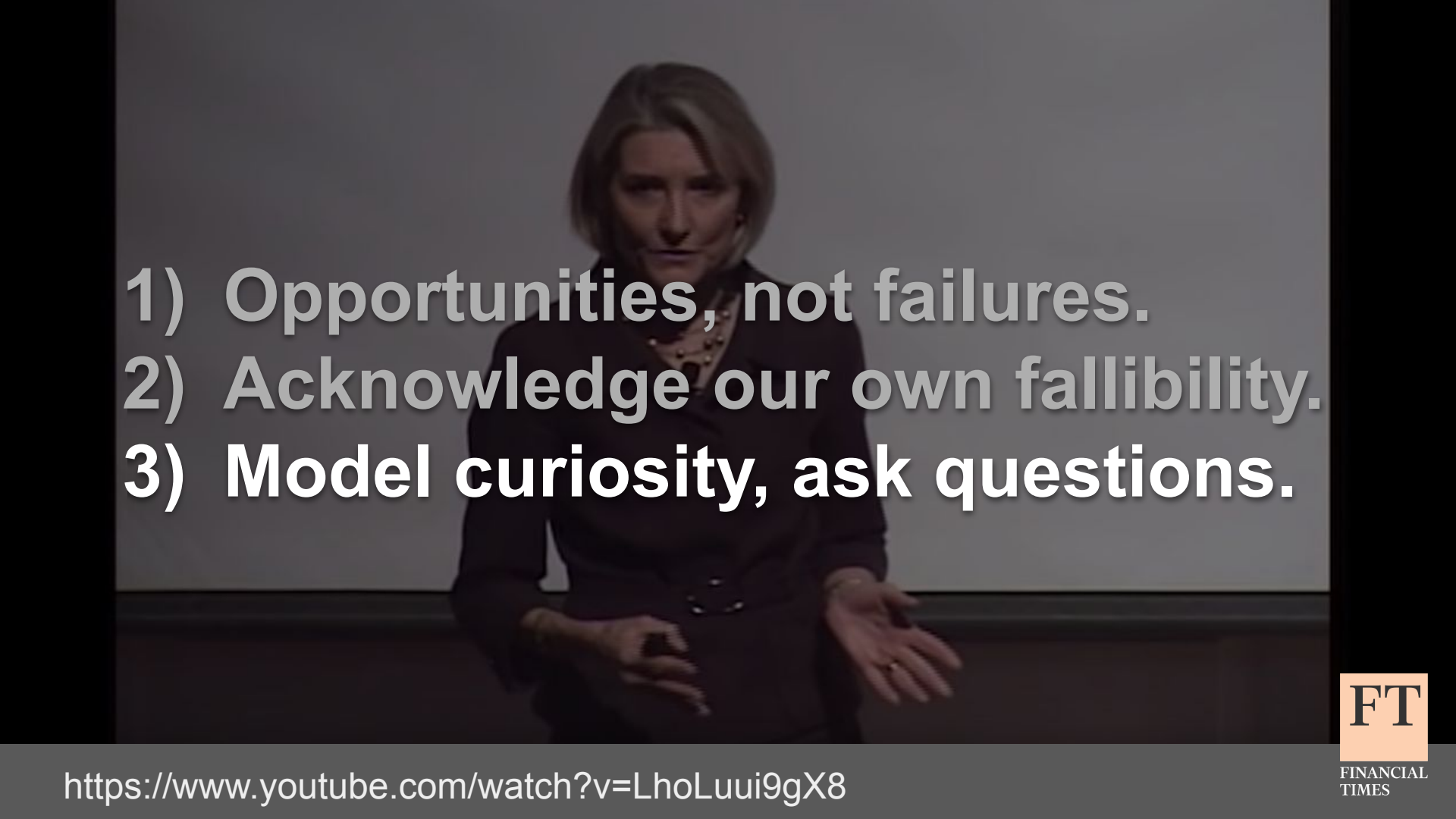
<https://www.youtube.com/watch?v=LhoLuui9gX8>

A woman with short blonde hair, wearing a dark suit and a necklace, is speaking on a stage. She has her hands open in a gesture. The background is a plain, light-colored wall. A large white text overlay is centered on the image.


**1) Opportunities, not failures.**

<https://www.youtube.com/watch?v=LhoLuui9gX8>

- 
- A woman with short blonde hair, wearing a dark blazer and a necklace, is speaking. She has her hands slightly raised in a gesturing motion. The background is a plain, light-colored wall.
- 1) Opportunities, not failures.
  - 2) Acknowledge our own fallibility.

- 
- 1) Opportunities, not failures.
  - 2) Acknowledge our own fallibility.
  - 3) Model curiosity, ask questions.





# The One Where a Director Falls Through the Ceiling

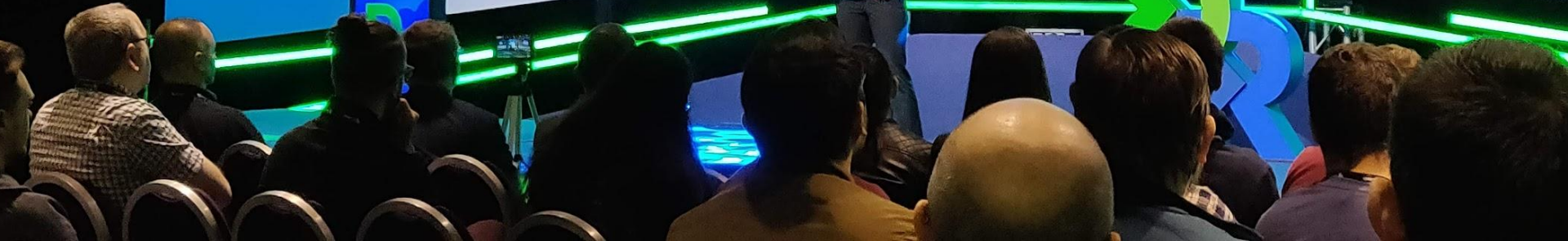




# Communication is key.

Especially to our customers.

@efinlay24





# Designate an incident lead.

@efinlay24

# On-Call and Incident Response: Lessons for Success, the New Relic Way



By Beth Adele Long • Oct. 24th, 2018 • Software Engineering

 [DevOps](#), [incident response](#), [on-call](#), [SRE](#)



*(Editor's note: This post is adapted from a pair of posts originally published on February 13, 2018.)*

Far too many companies continue to use on-call rotations and incident response processes that leave team members feeling stressed out, anxious, and generally miserable. Notably, plenty of good engineers are turning down jobs specifically for that reason.

It doesn't have to be this way. At New Relic, our DevOps practice has allowed us to create on-call



k8s - UPP Prod Delivery US: Annotations Read Aggregate Healthcheck is down  
(Incident #4073269)

[upp-prod-delivery-us.ft.com](https://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](https://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](https://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](https://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](https://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](https://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](https://upp-prod-delivery-eu.ft.com) • [View details](#)

*Software can be chaotic, but we make it work*



*Expert*

# Trying Stuff Until it Works

○ RLY?

*The Practical Developer*  
*@ThePracticalDev*

FT



# Create a temporary incident channel.

@efinlay24

# #inc-liveblogs-sorry-page

👤 0 | 🗨️ 11 | Sorry Page Not Found on Live Blog - [REDACTED]

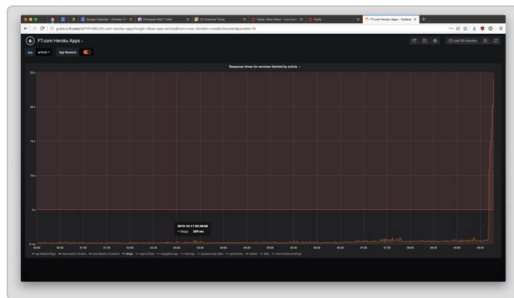
Thursday, 17 October



[REDACTED]

🕒 12:48

image.png ▾



1 reply 4 days ago



**Euan Finlay** 🇮🇪 12:48

possible, looking now



[REDACTED]

🕒 12:48

not that high traffic according to elb



[REDACTED]

set the channel description: Live Blogs responding slowly / Sorry Page Not Found



[REDACTED]

joined #inc-liveblogs-sorry-page.



[REDACTED]

🕒 12:49

Previous incident, [REDACTED]

Is is traffic from the app killing blogs?

And we're seeing the effects of that on [www.ft.com](http://www.ft.com)



[REDACTED]

That is about the time the deal was announced (1040 ish)



[REDACTED]

🕒 12:51

Memory, CPU and response codes looks ok



# Incident 131

**RESOLVED** - MAJOR SEVERITY

## Summary

We are noticing intermittent 'Sorry Page Not Found' when loading the live blog on the homepage

- **Reporter:** [REDACTED]
- **Lead:** [REDACTED]
- **Start Time:** Oct. 17, 2019, 9:46 a.m.
- **Report Time:** Oct. 17, 2019, 9:46 a.m.
- **End Time:** Oct. 21, 2019, 9:14 a.m.
- **Duration:** 95 hrs 27 mins
- **Participants:**
  - [REDACTED] (41 messages)
  - [REDACTED] (7 messages)
  - euan.finlay (37 messages)
  - [REDACTED] (35 messages)
  - [REDACTED] (34 messages)

## Timeline

09:53:33

[https://meet.google.com/\[REDACTED\]](https://meet.google.com/[REDACTED])

10:03:04

from looking at Blogs Web backend servers: - not seeing Out Of Memory errors like last incident - not seeing any unusual CPU load, CPU credits are fine

10:07:38

releasing router to prod



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

@efinlay24

**Tired people don't think  
good.**

@efinlay24

**Sometimes we have to  
leave things broken.**

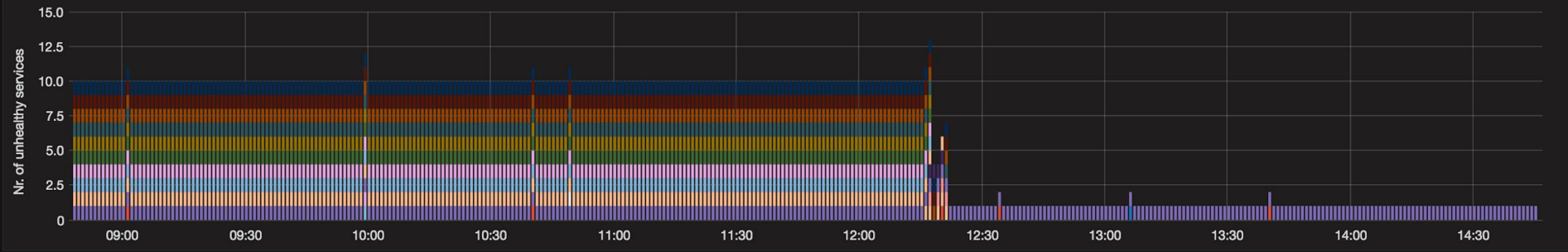
@efinlay24



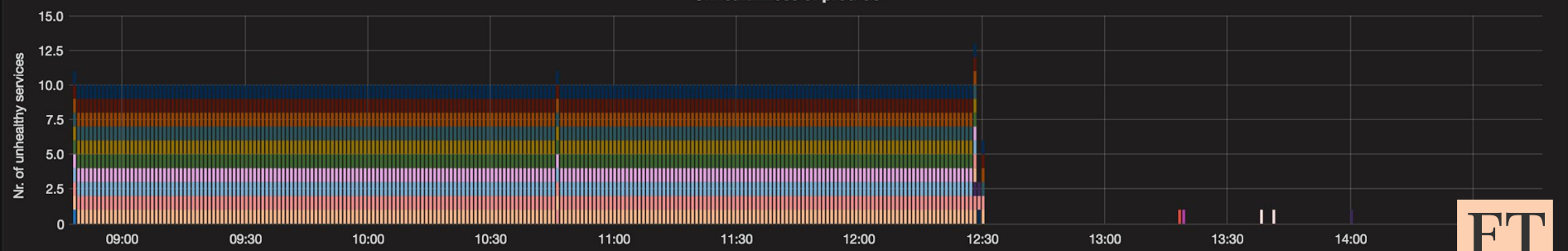
environment: prod-uk + prod-us + pub-prod-uk

services: All

### Unhealthiness of prod-uk



### Unhealthiness of prod-us



*"The Gang Serves Traffic  
From Staging"*





## Response time

Downtime

1h51m

Outages

21

Uptime

96.15%

Max resp. time

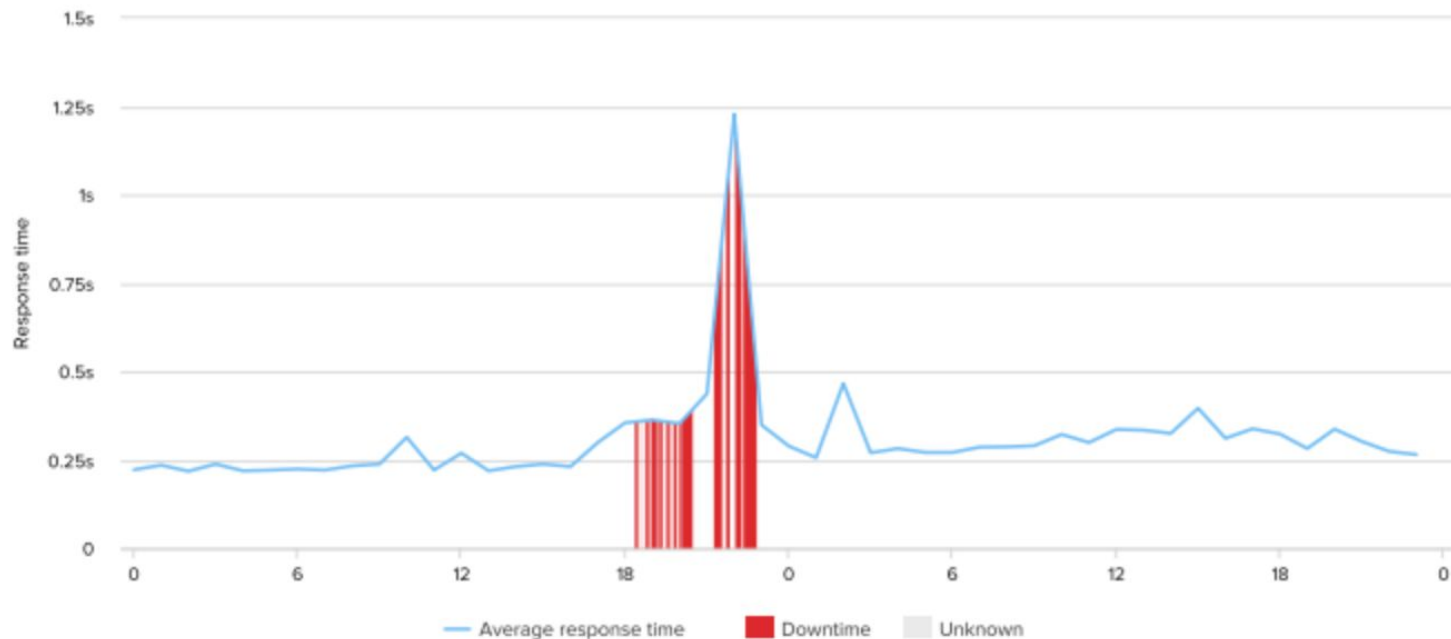
1.23s

Min resp. time

218ms

Avg resp. time

302ms



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

@efinlay24



# The Ghosts of Incidents...

Future

Present

> Past

# Congratulations! You survived.

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

@efinlay24

# Run a incident review with everyone involved.

Nobody died, so it's not a post-mortem.

@efinlay24

FT

FINANCIAL  
TIMES

**Incident reports  
are important.**

@efinlay24

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







Search or jump to...



Pull requests Issues Marketplace Explore



Financial-Times [redacted] Private

Watch 52 Star 4 Fork 0

Code Issues 59 Pull requests 0 Actions Security Insights Settings

Filters is:issue is:open

Labels 14 Milestones 0

New issue

<input type="checkbox"/>	<span>59 Open</span> <span>95 Closed</span>	Author	Labels	Projects	Milestones	Assignee	Sort
<input type="checkbox"/>	<span>!</span> <b>The one where a Brexit twist blew up live blogs (again)</b> #249 opened 3 days ago by [redacted] 4 of 9	[redacted]				[redacted]	1
<input type="checkbox"/>	<span>!</span> <b>The one where The Rachman Review podcasts were missing from ft.com/app</b> #248 opened 12 days ago by [redacted] 0 of 4	[redacted]					1
<input type="checkbox"/>	<span>!</span> <b>The one where links broke on the homepage</b> #247 opened 18 days ago by [redacted]	[redacted]					1
<input type="checkbox"/>	<span>!</span> <b>2019-10-01 - The one where ft.com was down because of a bad alias on Elasticsearch US.md</b> <b>type-incident-report</b> #246 opened 20 days ago by [redacted]	[redacted]					6
<input type="checkbox"/>	<span>!</span> <b># The one where we could not purge old content</b> <b>type-incident-report</b> #245 opened 27 days ago by [redacted]	[redacted]					
<input type="checkbox"/>	<span>!</span> <b>The one where a Boris blunder broke live blogs</b> <b>system-[redacted]</b> <b>type-incident-report</b> #242 opened 27 days ago by [redacted] 7 of 9	[redacted]					19
<input type="checkbox"/>	<span>!</span> <b>The one where editors couldn't edit the frozen edition in the app</b> <b>system-[redacted]</b> <b>type-incident-report</b> #241 opened 27 days ago by [redacted] 5 of 7	[redacted]					
<input type="checkbox"/>	<span>!</span> <b>The one where an empty Brexit list broke the Front Page</b> <b>system-[redacted]</b> <b>type-incident-report</b>	[redacted]				[redacted]	10



# 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, GitHub customers, and self-hosted GitLab CE users were not affected by the outage, or the data loss.



"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 GitHub instances were not affected.

**Identify what can be  
improved for next time.**

@efinlay24

# 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

# We had problems with bank transfers on 30th May. Here's what happened and how we're fixing it for the future.



On the 30th of May 2019 between 09:54 and 19:20, around a quarter of bank transfers into Monzo accounts were failing or delayed by several hours. And bank transfers from Monzo accounts were delayed by a few minutes.

During this time, you might've had trouble getting payments from other banks, had payments into your account take a while to arrive, or seen bank transfers arrive in your Monzo account then get reversed later.

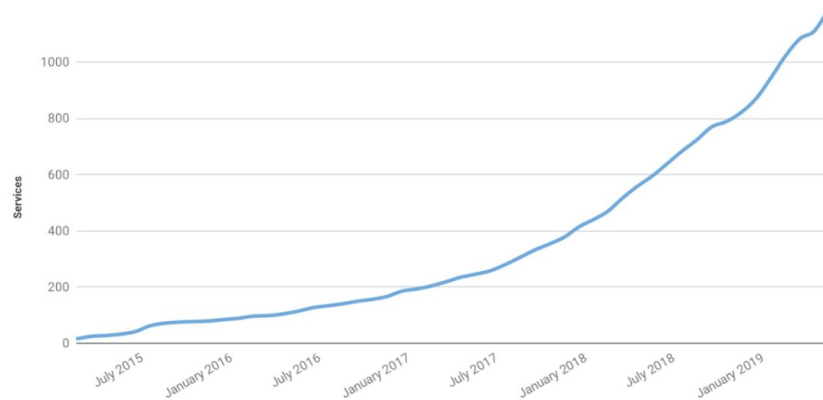




## We were scaling up Cassandra to keep apps and card payments working smoothly

As more and more people start using Monzo, we have to scale up Cassandra so it can store all the data and serve it quickly and smoothly. We last scaled up Cassandra in October 2018 and projected that our current capacity would tide us over for about a year.

But during this time, lots more people started using Monzo, and we increased the number of microservices we run to support all the new features in the Monzo app.



# Nearly the end.

Don't clap yet.

@efinlay24

# Feedback is welcome.



<https://www.teierik.com/aevreach/day>

# Failure is inevitable.

And that's ok.

@efinlay24

# The end.

"Please clap."  
Jeb Bush, 2016

@efinlay24



# We're hiring in Sofia!

<https://ft.com/dev/null/>

@efinlay24  
euan.finlay@ft.com

FT