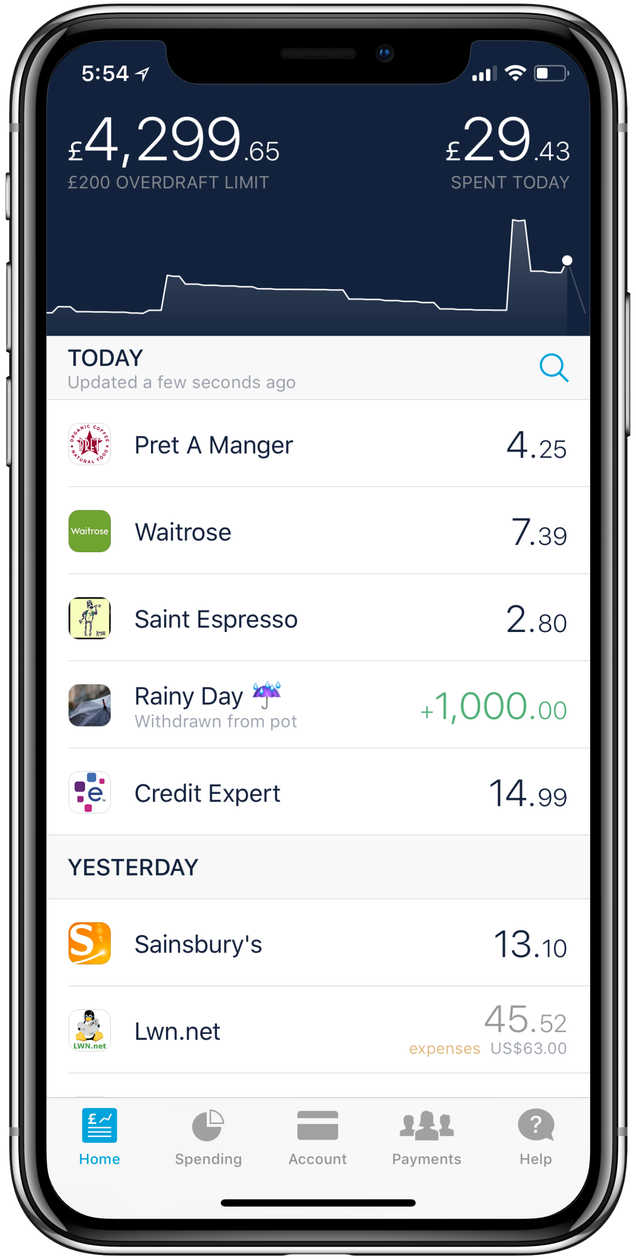
Anatomy of a Production Kubernetes Outage 🔬

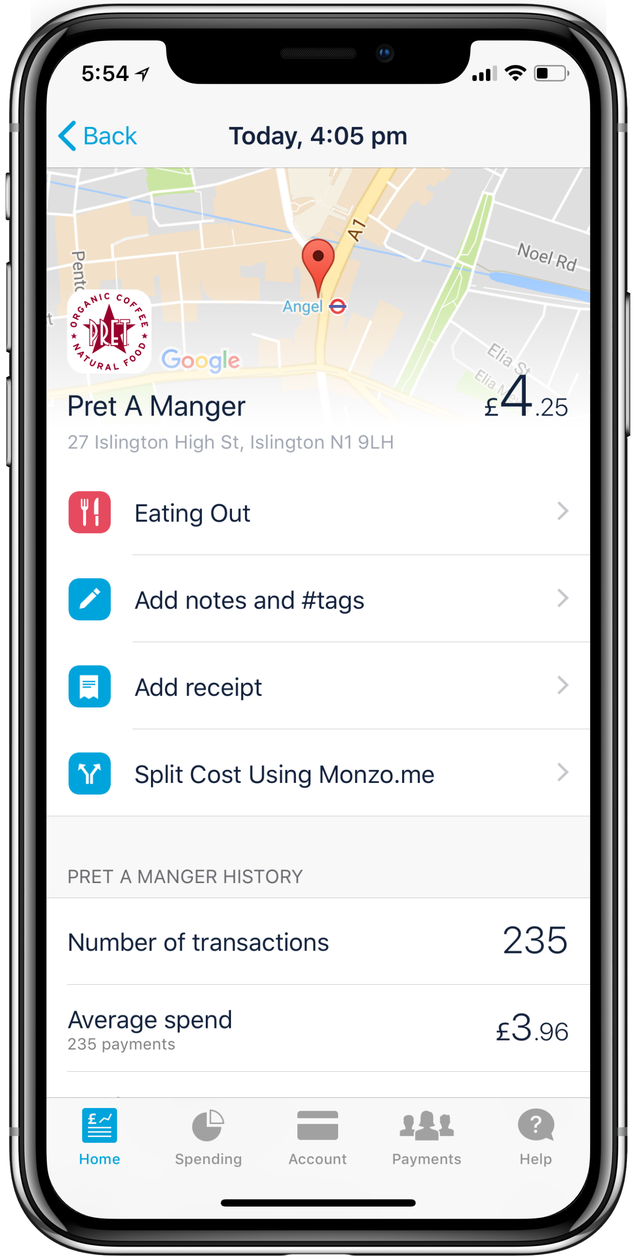
Oliver Beattie

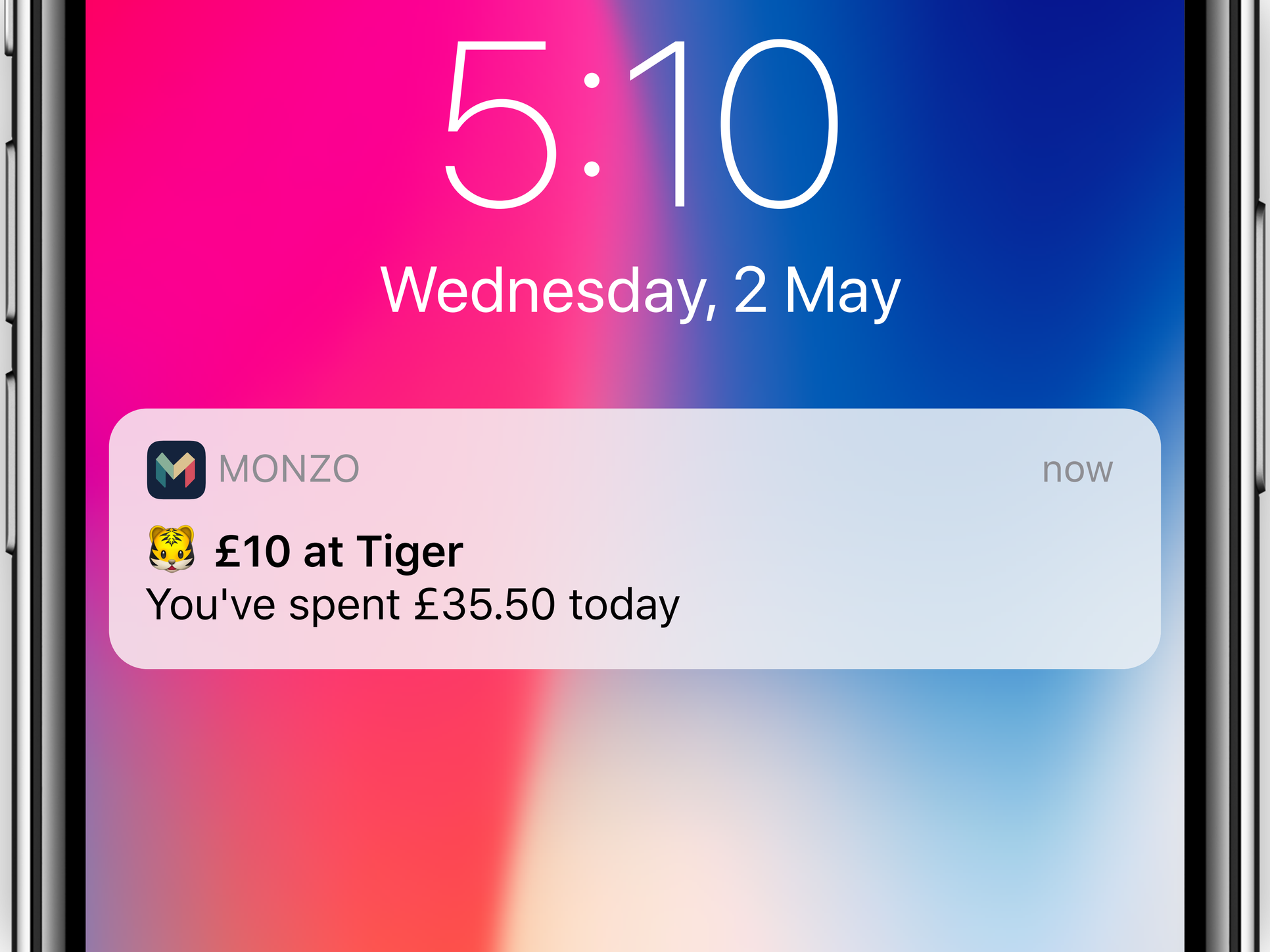
Head of Engineering, Monzo Bank


monzo
5354 2000 0000 0000
EXP
END 08/1
469-624-882


£1,064.25
CARD BALANCE
9:30
£51.60
SPENT TODAY
TODAY
Sainsbury's
James Nicholson
Thanks for dinner!
YESTERDAY
Tommi's Burger Joint
40.00
Hailo
11.60
Top Up
MONDAY, 15 MAY
19.49
13.45
expenses
+1,000.00
Home
Spending
Card
Contacts
Help
monzo
.000 0000 0000
EXP END 08/19
882
mastercard


5:54
£4,299.65
£200 OVERDRAFT LIMIT
£29.43
SPENT TODAY
TODAY
Updated a few seconds ago
Pret A Manger
4.25
Waitrose
7.39
Saint Espresso
2.80
Rainy Day
Withdrawn from pot
+1,000.00
Credit Expert
14.99
YESTERDAY
Sainsbury's
13.10
Lwn.net
45.52
expenses US$63.00
Home
Spending
Account
Payments
Help


5:54
Back
Today, 4:05 pm
A1
Noel Rd
Angel
Google
Pret A Manger
£4.25
27 Islington High St, Islington N1 9LH
Eating Out
Add notes and #tags
Add receipt
Split Cost Using Monzo.me
PRET A MANGER HISTORY
Number of transactions
235
Average spend
235 payments
£3.96
Home
Spending
Account
Payments
Help


5:10
Wednesday, 2 May
MONZO
now
£10 at Tiger
You've spent £35.50 today

> 500 micro services

Built on open source software

Story of an outage

CAST OF CHARACTERS 🎭😈

Kubernetes

etcd

Linkerd

Humans

etcd upgrade

⏰ 2 WEEKS BEFORE THE OUTAGE

Deployment of faulty service

Scaled to zero replicas

Ledger change deployed

⏰ START OF PARTIAL OUTAGE

Ledger change rolled back

⏰ 2 MINS INTO THE OUTAGE

Linkerd identified as unhealthy

⏰ 6 MINS INTO THE OUTAGE

Begin restarting Linkerd pods

⏰ **16 MINS** INTO THE OUTAGE

New Linkerd pods cannot start

Kubernetes apiserver restarted

⏰ 27 MINS INTO THE OUTAGE

Finish restarting Linkerd pods

Matt Heath 2:38 PM

shit

PagerDuty APP 2:39 PM

Triggered #243: DOWN alert: Monzo platform healthchecks

Assigned: Priyesh Patel

Service: Platform health

Integration: Pingdom

⏰ **ESCALATED TO TOTAL OUTAGE** **1 HR 3 MINS** INTO THE OUTAGE

Linkerd `NullPointerException` observed on start up

⏰ **1 HR 17 MINS** INTO THE OUTAGE

Linkerd/k8s incompatibility found

Empty services deleted

⏰ END OF OUTAGE 1 HR 21 MINS

IMPACT 🔥

1 hour, 21 mins of cluster downtime

Vast majority of payments succeeded throughout

ROOT CAUSES 🔬

Bug in gRPC client library affecting etcd

Incompatibility between Kubernetes + Linkerd

```
"endpoints": []
```

K8S < 1.6

```
"endpoints": []
```

K8S < 1.6

vs.

K8S 1.6+

```
"endpoints": null
```

ROOT CAUSES 🔬

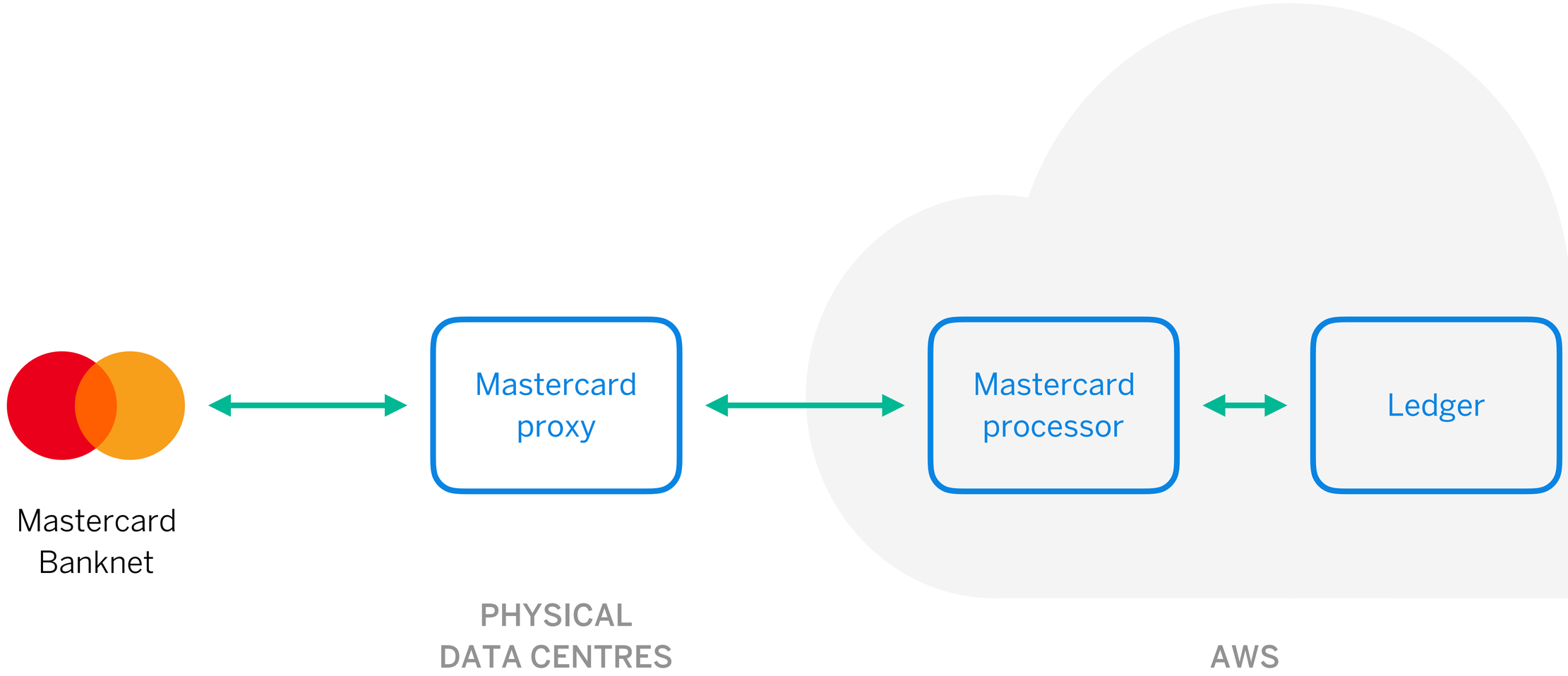
Bug in gRPC client library affecting etcd

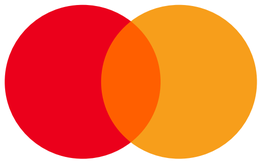
Incompatibility between Kubernetes + Linkerd

Human error

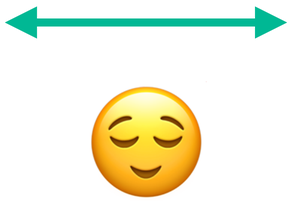
LESSONS 🎓

Defence in depth


Mastercard
Banknet
Mastercard
proxy
Mastercard
processor
Ledger
PHYSICAL
DATA CENTRES
AWS



Mastercard
Banknet



Mastercard
proxy

PHYSICAL
DATA CENTRES

Mastercard
processor

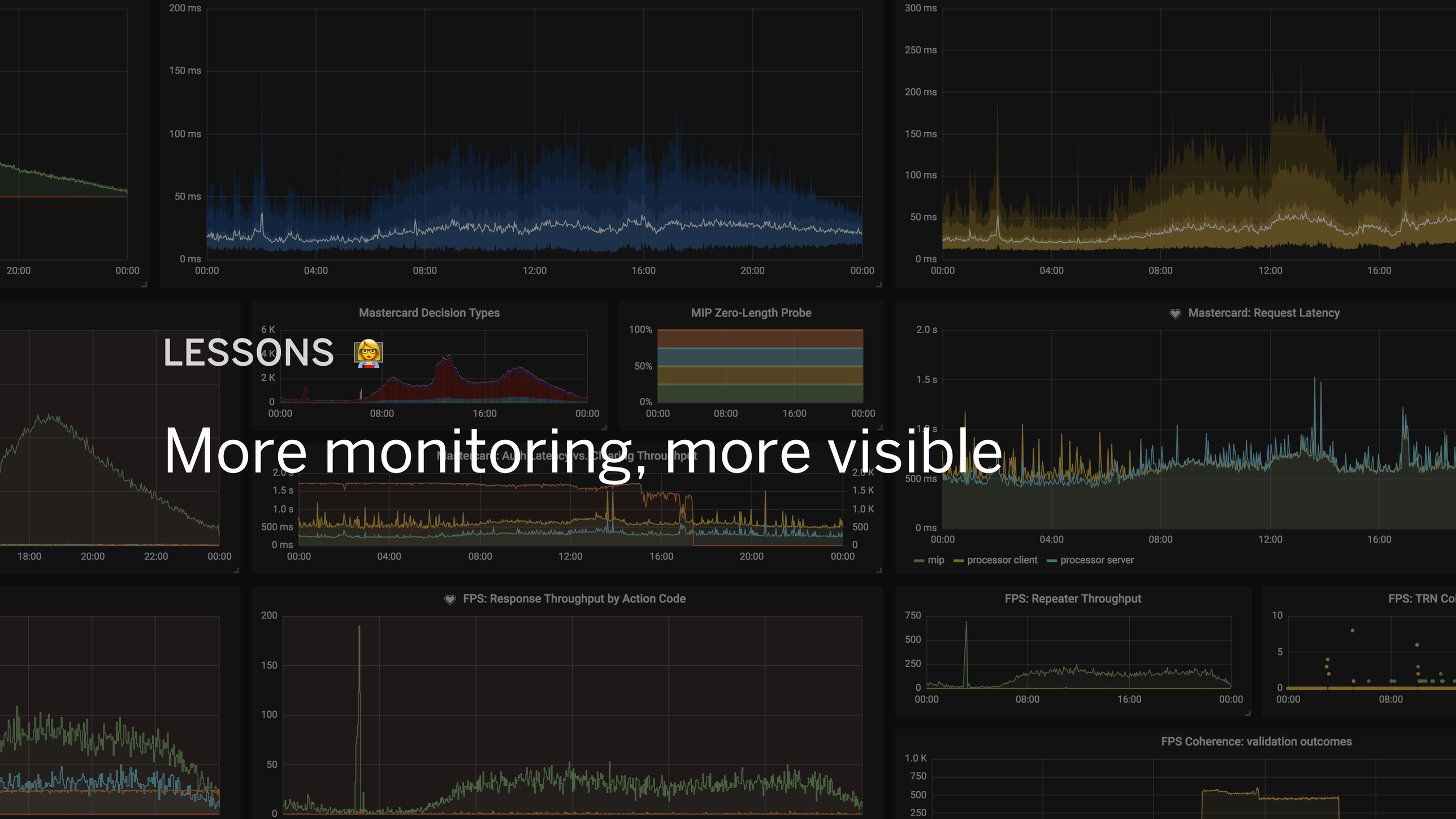
Ledger

AWS

LESSONS 🎓

Chaos engineering

“Chaos Engineering is the discipline of experimenting on a distributed system in order to build confidence in the system’s capability to withstand turbulent conditions in production.”



LESSONS 👩‍🏫

More monitoring, more visible

LESSONS 🎓

Be transparent; embrace the community

oliver Oliver Beattie Monzo

oliver Oct '17

Hi everyone 👋 I'm Monzo's Head of Engineering, and as I promised on Friday I'd like to share some more information about what happened during this outage. Because the nature of the issue was technical, this post is also quite technical. 🤓

It's important to note that we had two major incidents last week that many of you will have experienced (sorry again.) The first incident lasted most of the week and affected only our prepaid product – ie. Monzo Alpha and Beta cards. The second outage affected both the prepaid product and our new current account for a period of around 1½ hours on Friday afternoon. This post is about the latter.

You can learn more about our overall backend architecture in this blog post 753 I published last year, but it's important to understand the role of a few components in our stack at a high level to understand this issue:

- Kubernetes 102 is a system which deploys and manages all of our infrastructure. Monzo's backend is written as several hundred microservices, packaged into Docker containers. Kubernetes manages these Docker containers and ensures they are running properly across our fleet of AWS nodes.
- etcd 122 is a distributed database used by Kubernetes to store information about which services are deployed, where they are running, and what state they're in. Kubernetes requires a stable connection to etcd in order to work properly, although if etcd does go down all of our services do continue running – they just can't be upgraded, or scaled up or down.
- linkerd 458 is a piece of software that we use to manage the communication between all of the services in our backend. In a system like ours, thousands of network calls are happening every second, and linkerd does the job of routing and load balancing all of these calls. In order to know where to route these calls, it relies on being able to receive updates about where services are located from Kubernetes.

Timeline

- **Two weeks before:** The Platform team makes some changes to our etcd cluster to upgrade it to a new version, and also to increase the size of the cluster. Previously, this cluster consisted of three nodes (one in each of our three zones 126); we raise this to nine (three in each zone.) Because etcd relies on being able to achieve a quorum 118 to make progress, this means that in this setup we can tolerate the simultaneous loss of an entire zone and a single node in another zone.

Oct 2017
95 / 185
Oct 2017
Dec 2017

obeattie commented on 29 Oct 2017 • edited

Not to add to the noise, but we've encountered this issue in production, and it ended up leading to a complete cluster outage (through a very unfortunate series of events.)

I have gathered all the relevant logs from our 3 k8s master and 9 etcd nodes. There may not be anything of additional interest there, but if you would like to see them please let me know and I can share them privately.

❤️ 10

timothysc commented on 31 Oct 2017 • edited

Owner

@obeattie I'm sooo sorry. I'll update the client tomorrow, and going to poke folks about getting the next rev in line for release.
/cc **@luxas @roberthbailey @jbeda**

❤️ 1


monzo.com/careers

@obeattie