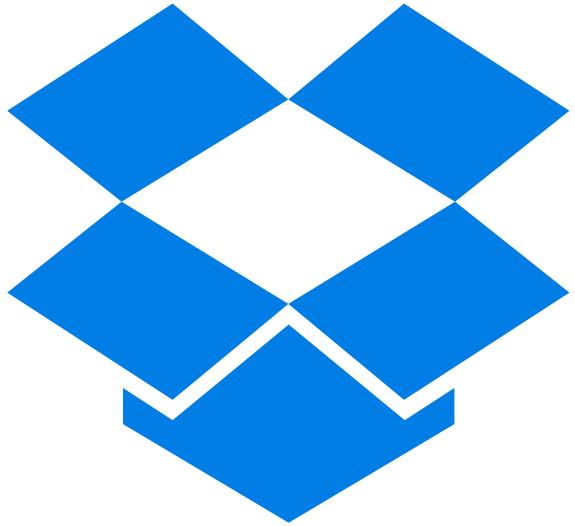
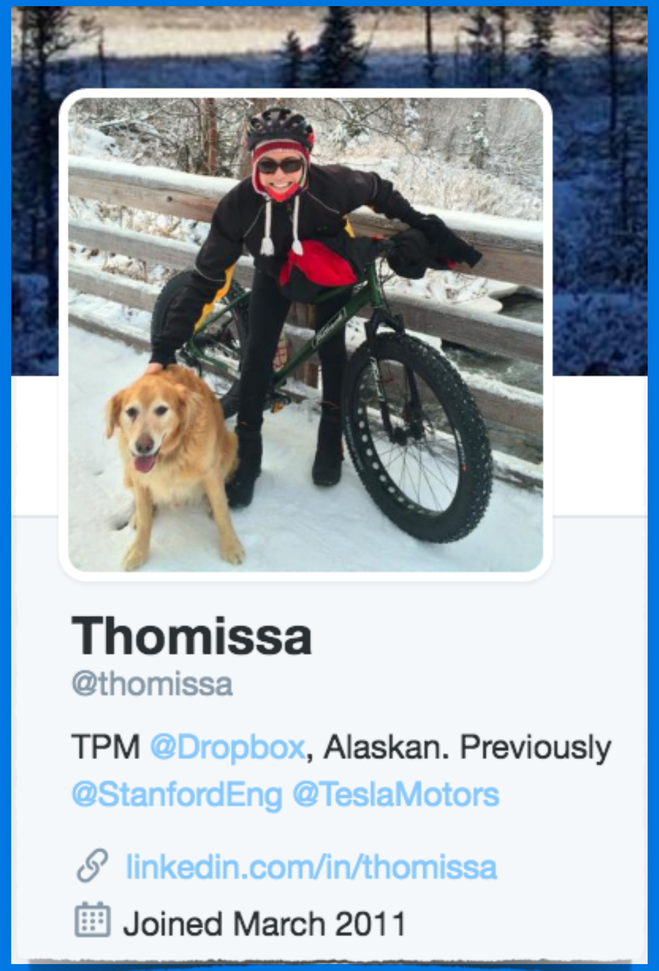


Thomissa Comellas & Tammy Butow

0-100 Days - Running DRTs at Dropbox

Thomissa



Dropbox

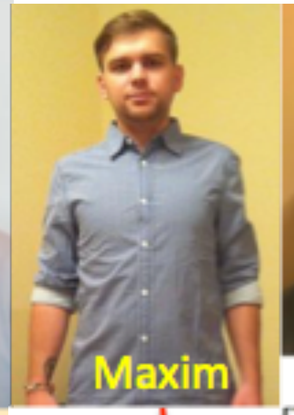
Tammy

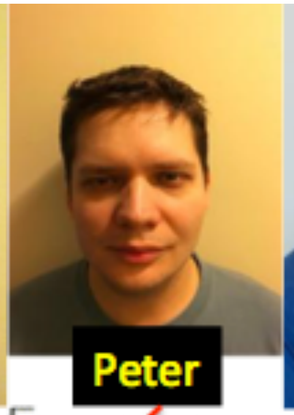
Dropbox

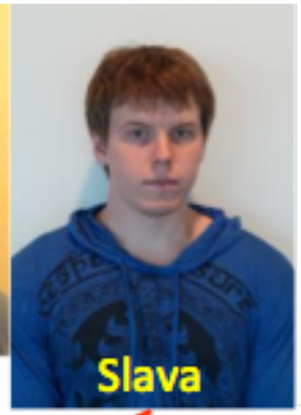

Tammy

Renjish

Dave


Maxim


Peter


Slava


Klamath
Klamath Falls
Eureka
Idaho
Redding
Sacramento
West Sacramento
Nevada
San Francisco
Stockton
Reedley
Lake Mead
Los Angeles
Long Beach
San Diego
Arizona
Colorado


Great Britain
Wales
Scotland
north atlantic ocean
north sea
baltic sea
The Netherlands
Denmark
Ireland
United Kingdom
Germany
Belgium
Luxembourg
Austria
Liechtenstein
Switzerland
France
Andorra
San Marino
Monaco
Italy
adriatic sea
tyrrhenian sea
ionian sea
Gibraltar
Malta

Brian


René

Tool & Maynard James Keenan Are Making New Album Via Dropbox

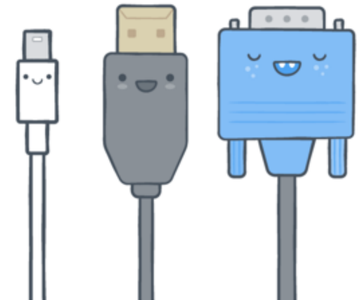
Rock Music | Alternative | Hard Rock | Rock Features | Headlines | Metal

by Brett Buchanan - Nov 23, 2015

Adam Jones discussed Tool's new album in a recent **Rolling Stone** interview.

What will we share today?

- Background on Dropbox
- Thomissa’s 100 days transforming DRTs at Dropbox
- Tammy’s examples of how we currently run DRTs
- What the future looks like for “anti-fragile” at Dropbox



Dropbox

500 million users is the start of something big.
We wouldn't be here without you.

We've scaled to meet your needs...

3,300,000,000

sharing connections have
been created with Dropbox

1,200,000,000

files are saved on Dropbox
every day

DRTs at Dropbox

Disaster Recovery Test

Where do DRTs fit in?

Testing Tools

Chaos Monkey
Greg Orzell
Netflix

Chaos Kong
Luke Koweski
Netflix

DRTD
Tim Doug, David Mah & Brian Cain
Dropbox

Testing Methods

GameDays
Jesse Robbins
Amazon

DiRTs
Kripa Krishnan
Google

DRTs
Thomissa Comellas & Tammy Butow
Dropbox

0–25 Days “Learning and Reflection”

BREAK IT

DRT

Disaster Recovery Test

FIX IT

SEV

High Severity Incident

Sweat The Details

Streamline the SEV reporting process

Derive analytics from failures

Increase view of the details related to SEVs and DRTs

Become more data driven

25–50 Days “Determining the WHY?”

DropSev

All SEVs | SEVs currently open | Report a new SEV

Report a SEV

The incident so far:
Reporter: thomissa
Category: Unavailability

What's the nature of your production emergency?
(Pick the first applicable option)

- **Data Integrity** - user data may be missing or corrupt
- **Security Leak** - user data or Dropbox secrets may be exposed to unauthorized users
- **Product Abuse** - people are doing mean things to our product or other users
- **Other Security Issue** - not covered above
- (•) **External Unavailability** - Dropbox is erroring out or performing badly for some users
- **Internal Tool Unavailability** - something non-user-visible is erroring out or performing badly
- **I'm not sure**

Is the emergency currently happening or are you reporting a resolved issue?

- (•) The emergency is ongoing
- The emergency has been resolved

Does the emergency last more than 20 minutes?

- (•) YES!
- No/Don't know.

Are any of these bonus catastrophes currently occurring? If multiple are occurring, pick the first one on the list. X

- ( ) Users can't login
- ( ) Network capacity is <50% of normal
- ( ) >25% of billing requests are failing
- ( ) This is affecting third-party sites via the API
- ( ) None of the above

Romantic Hedgehog

Create issue

Configure Fields

Issue Type* SEV

Some issue types are unavailable due to incompatible field configuration and/or workflow associations.

Summary*

SEV Level None

IMOC

Start typing to get a list of possible matches.

TLOC

Start typing to get a list of possible matches.

Affected System(s) (Infra Only)

Contbin/Replication
Databases
Delphi
DNS
Edgestore
Emailservice

Start Time (UTC)

Date and time when the SEV began

SEVs & DRTs at Dropbox

DropSev: SEV filed	SEV auto-named	JIRA ticket created w/ initial data	Ticket complete w/ AIs, DRTs, etc assigned	Weekly Reliability Review [SEVs]

DRTs

Reliability Working Group: Every Two Weeks

DRTs as a Product

How do teams run DRTs?

How have DRTs changed over time?

What do teams like / dislike about DRTs?

How do teams prioritize DRTs?

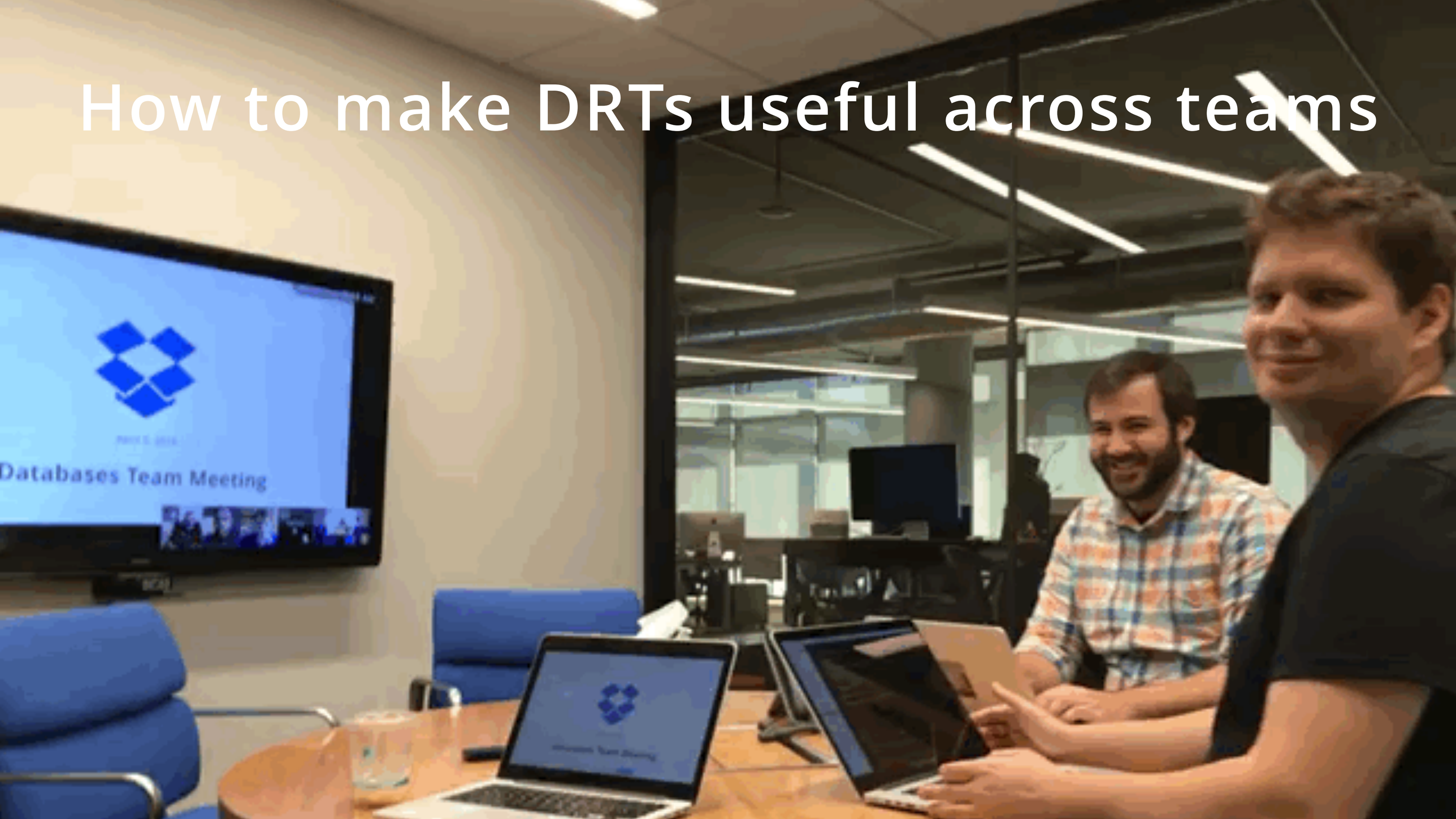
Try my MVP?

	Known	Unknown
Known	Compliance (K/K)	Change (K/U)
Unknown	Checking (U/K)	Control (U/U)

50–75 Days “Determining the HOW”

Dropbox

How to make DRTs useful across teams



Questions To Guide Teams

What are the weaknesses

How does code behavior differ from expected

How can you improve visibility

How confident are you in regards to failure modes

How well do you know your system inter-dependencies

75–100 Days “Determining the WHAT”

Dropbox

What do you do now?

Maturity level 2

Start with past SEVs impacting the system

Test the most general hardware and software failure modes

And do the tests you've meaning to run, you know, for a while.

Script for Magic Pocket DRT: mp_shed_block_load_via_dns.py

Living Runbook

https://paper.dropbox.com/doc/Cluster·

Living document (e.g. paper doc) describing current state of DRT runbook

Cadence

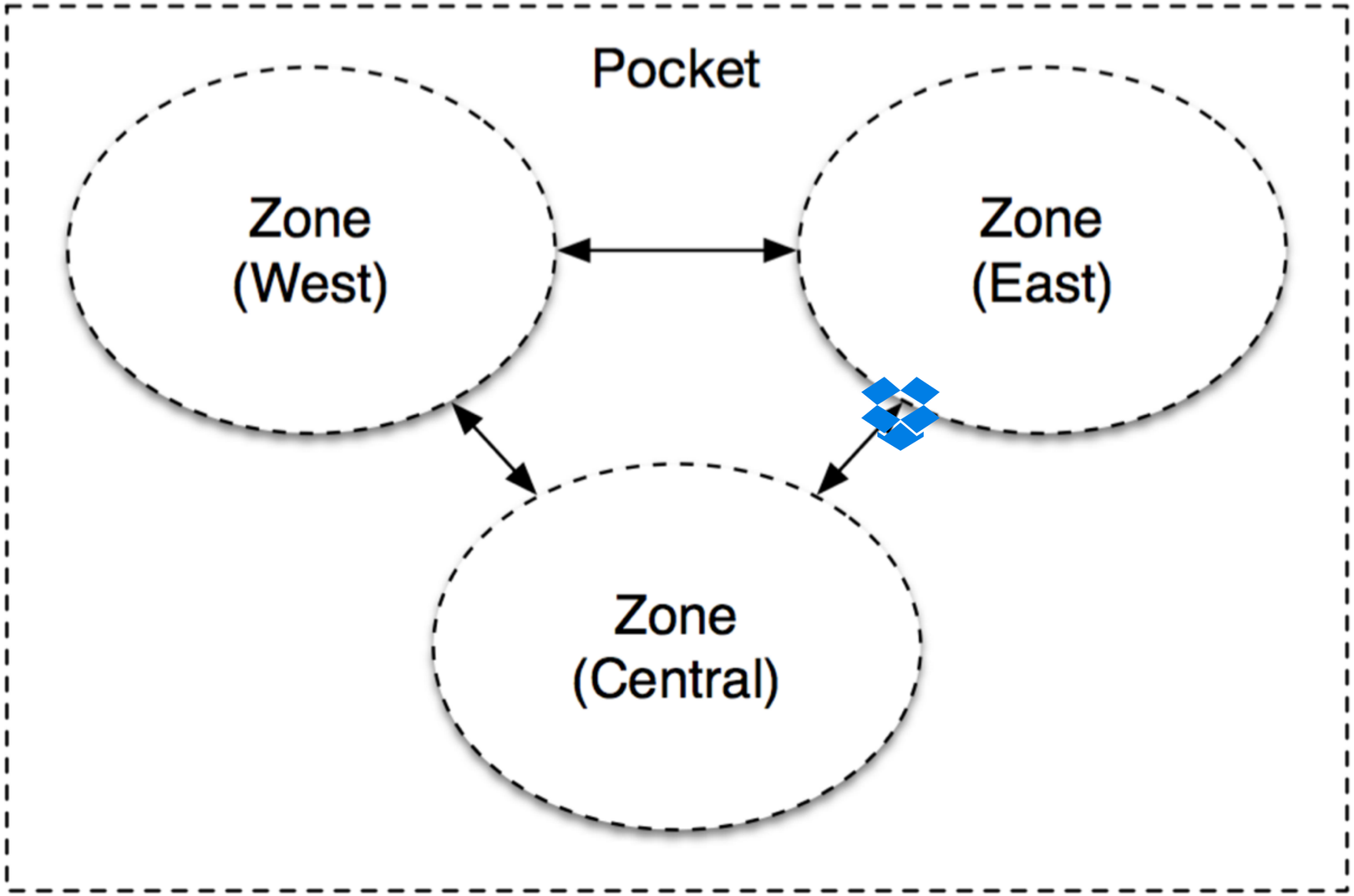
None
Every two weeks
Every four weeks
Every six weeks
Every quarter

How often DRT should be repeated until retired

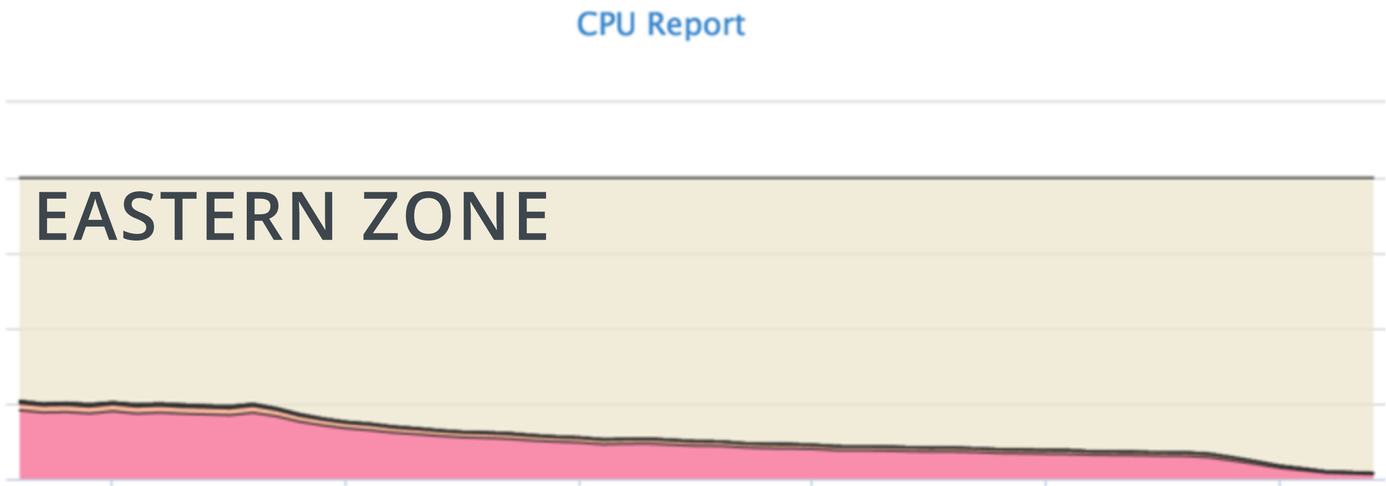
Affected System(s) (Infra Only)

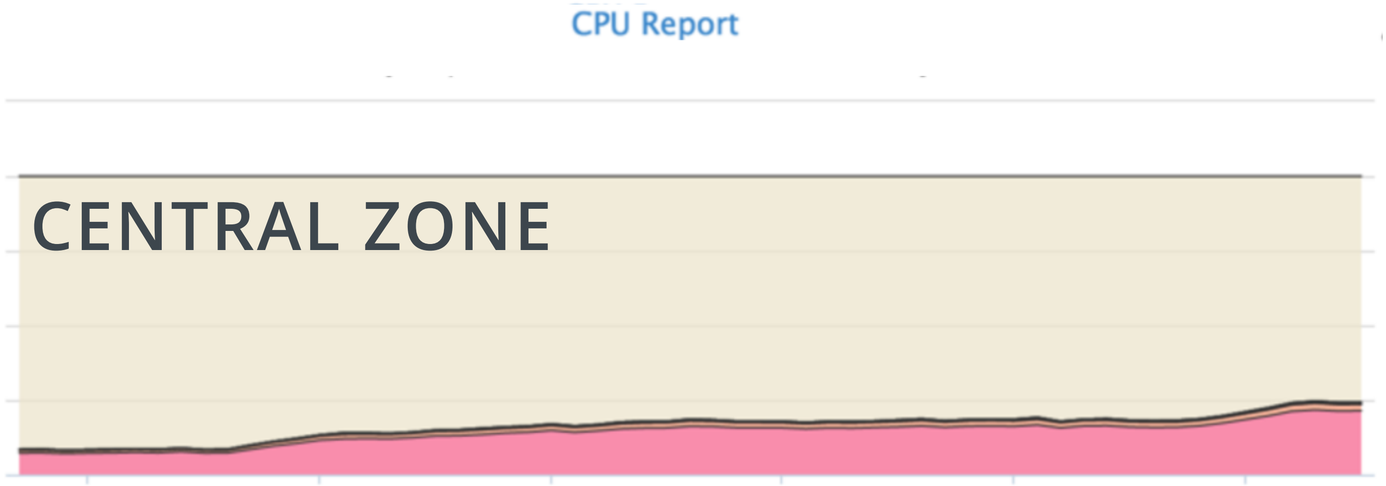
Hedwig
IPVS
Kafka
Livefill
Magic Pocket

MP Failover DRT

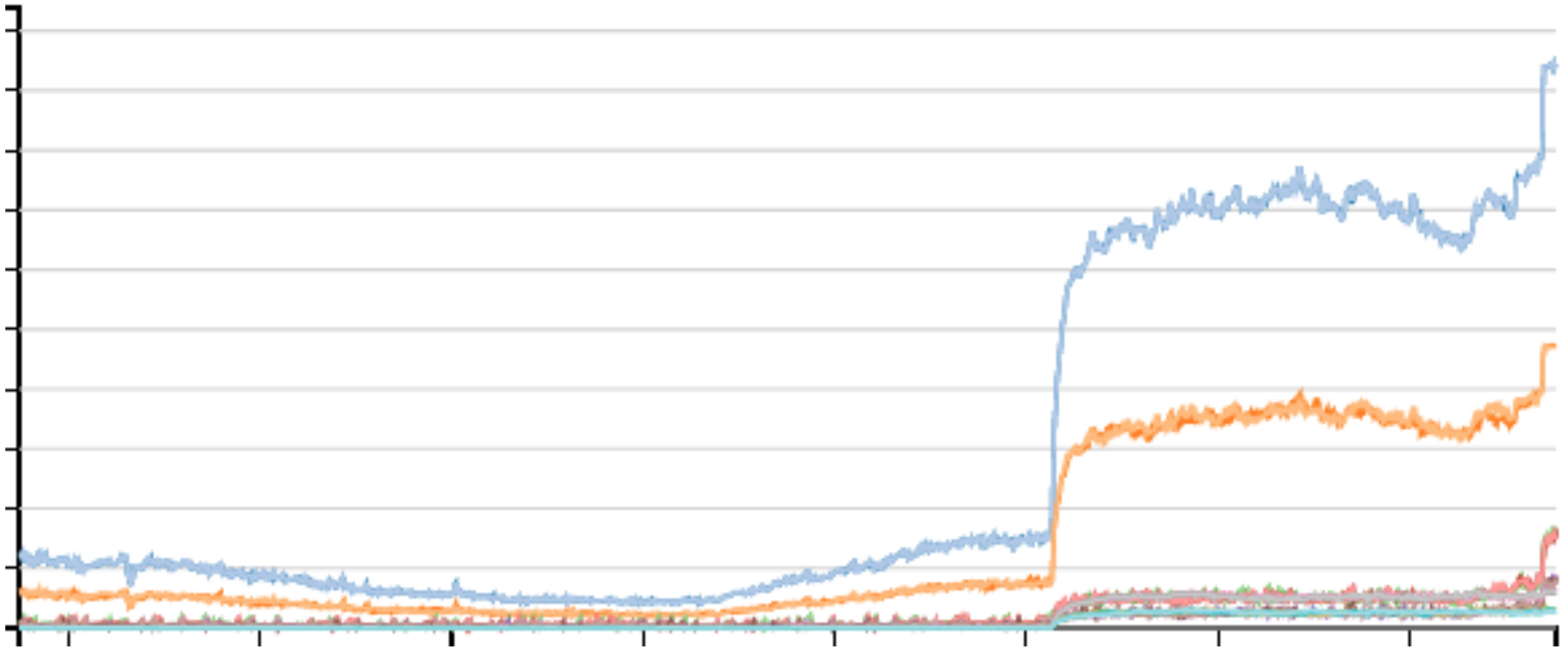


Failover - block-misc





Networking



Learning from your DRTs

Engage Networking team for DRTs

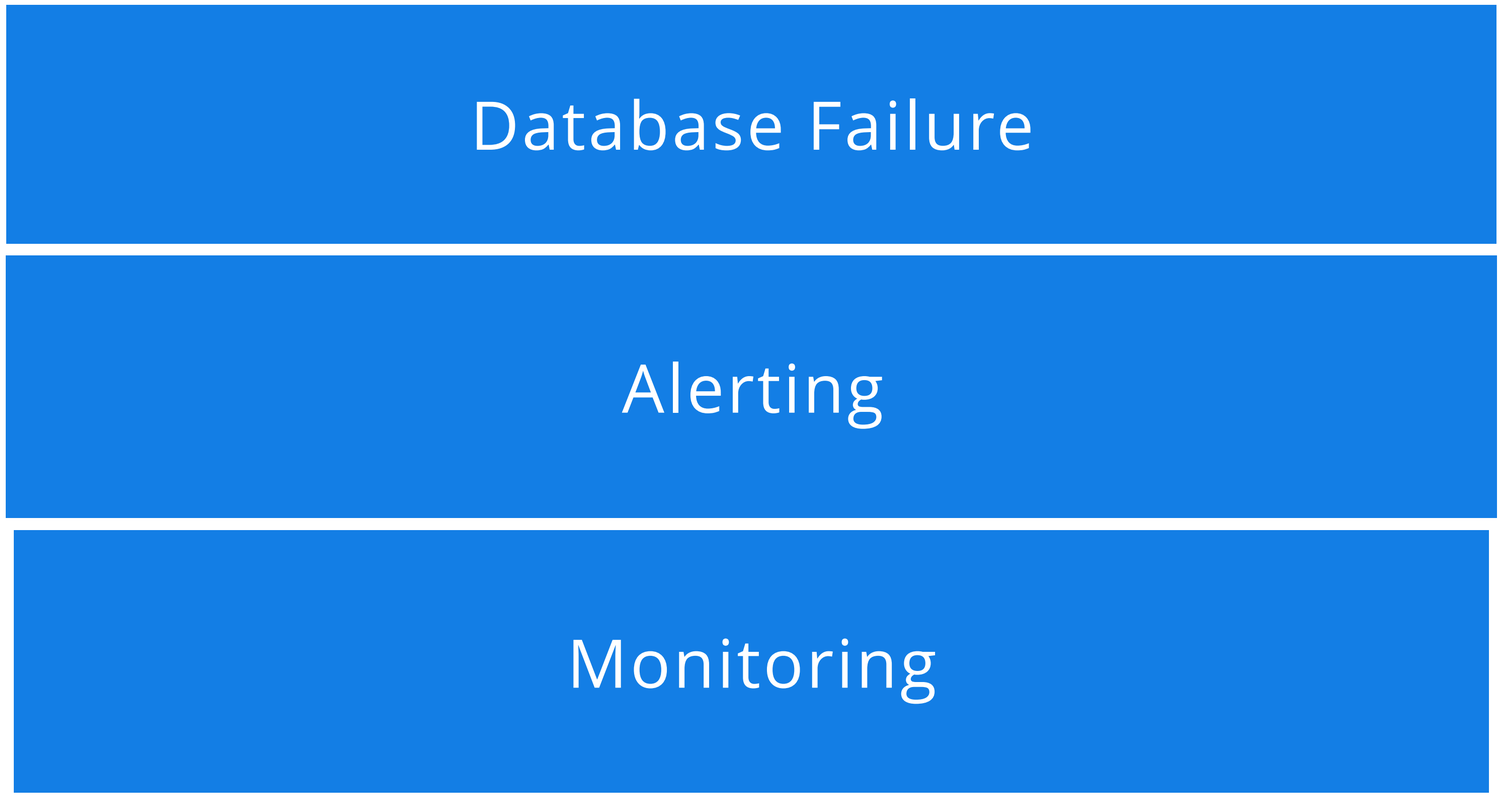
Communicate in advance of DRTs

DRT automated calendar

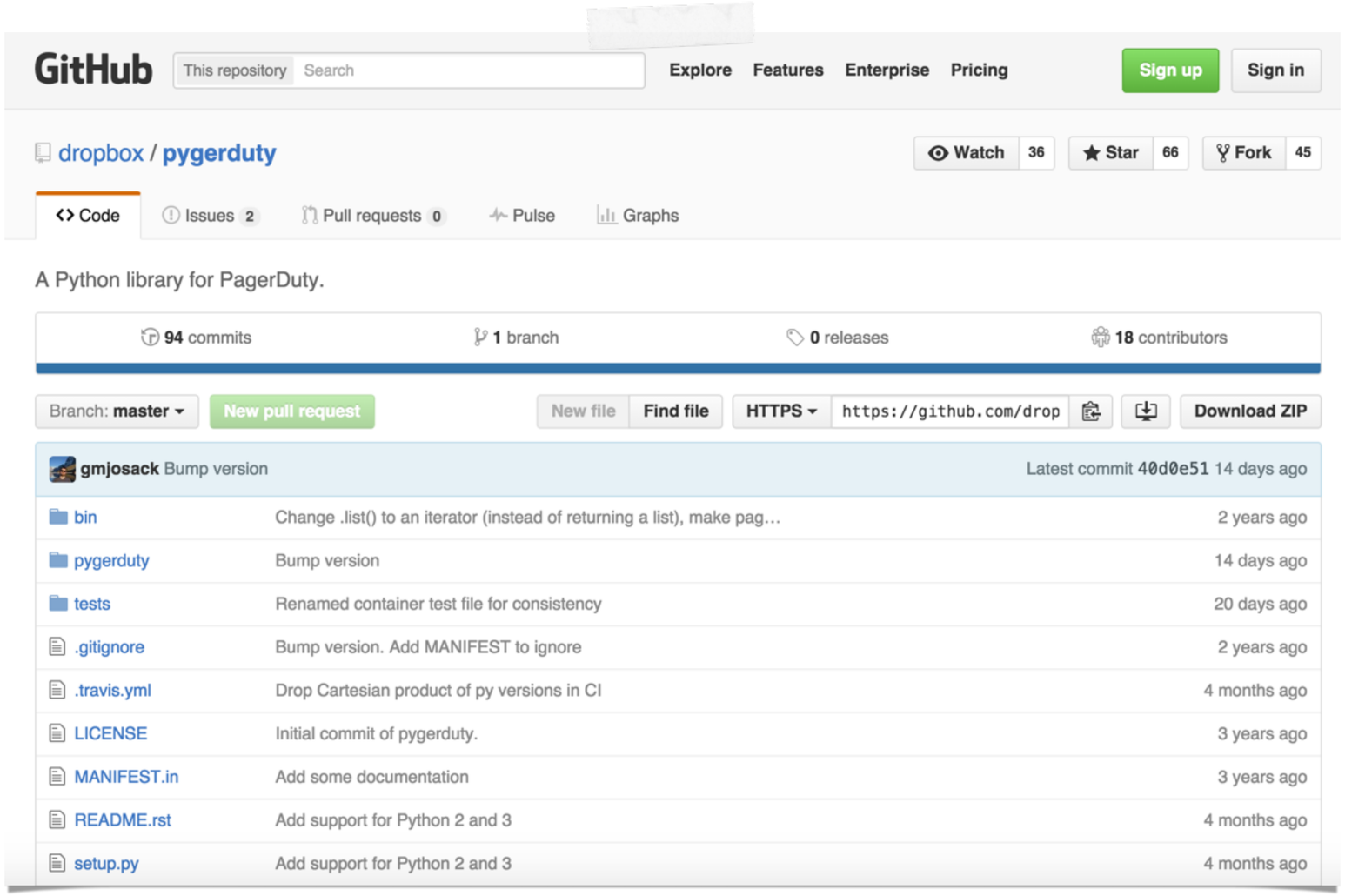
DRT dashboards for teams

Databases DRTs at Dropbox

Databases DRTS at Dropbox

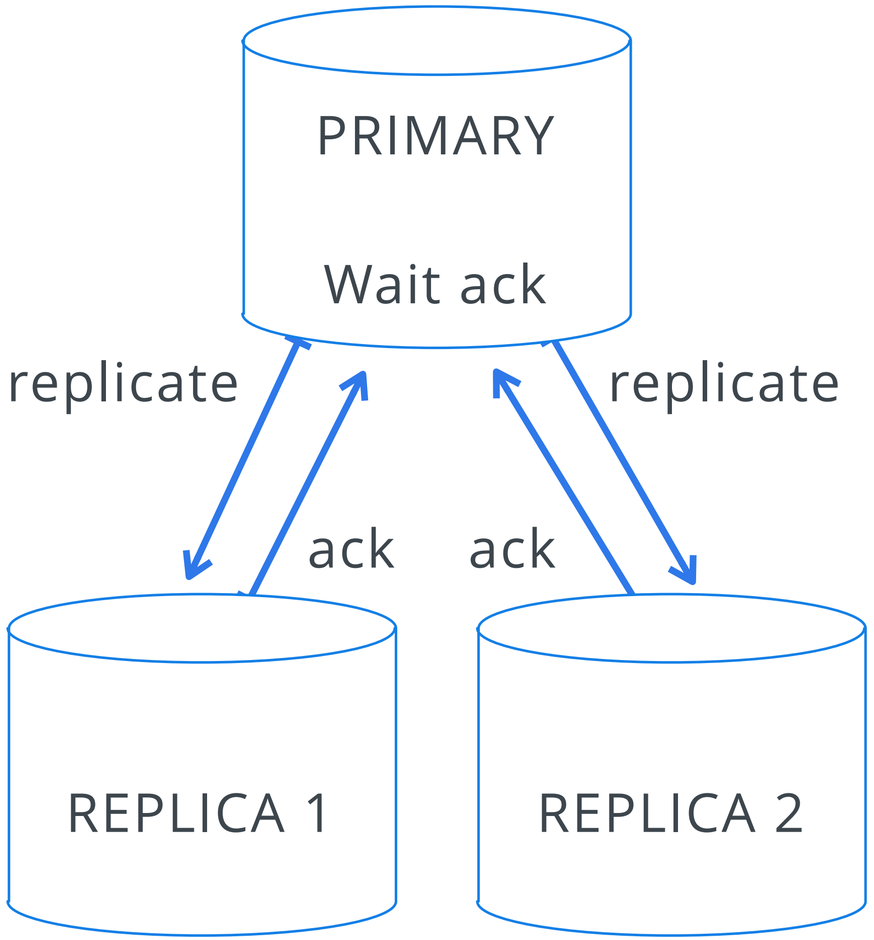


dropbox/Pygerduty




YouTube
Search
Example #2: server_file_journal
CREATE TABLE `server_file_journal` (
`id` int(10) unsigned,
`filename` varchar(255),
`latest` tinyint(1),
`ns_id` int(10) unsigned,
`prev_rev` int(10) unsigned,
[...]
PRIMARY KEY (`ns_id`, `latest`, `id`)
) ENGINE=InnoDB;
52:05 / 1:08:16
How We've Scaled Dropbox
Stanford
Subscribe
535,006
61,935 views
Add to
Share
More
509
15
Published on Sep 10, 2012
(Feburary 22, 2012) Kevin Modzelewski talks about Dropbox and its History. He describes the technological issues faced by Dropbox and the actions they have to take in order to continuously improve it.

Testing Server File Journal



MySQL 5.6
semi-sync

Running a database primary DRT

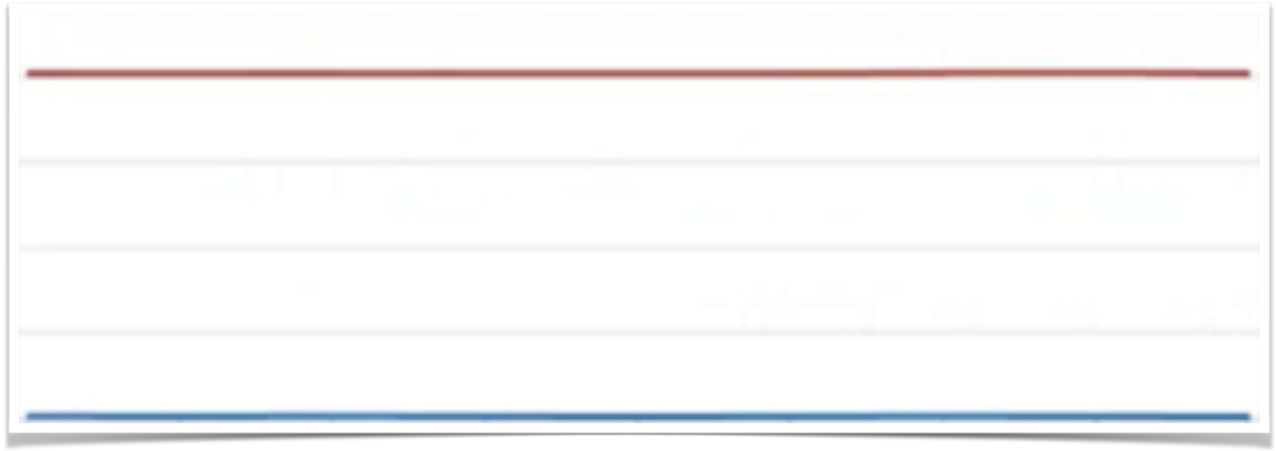
Item	Response	Reasoning
Due Date [Date Scheduled]	5/11/16	Surfaces potential clash with other teams Allows for sprint planning (and entry on calendar)
Assignee [Person executing]	Brian C	Fame
Assignee Team	Databases, Filesystems	Tag needed for board / tracking
Summary [Title / Description]	SFJ Primary Failure	Readability / Reference
Affected Systems	Server File Journal (SFJ)	Tracking dependencies, educating for additional DRTs, future scheduling
**Living runbook link (e.g. paper doc)	+Filesystems DRT with Databases	Visibility into historical vs. current (could be a web link field)
**Cadence	Every quarter	[every six weeks, every two weeks, every quarter] determine effective scheduling / KTLO budgeting

We check that the daily push has happened or will happen later

“At x:xx I will be performing a controlled primary failure for SFJ. This will affect xxx, xxx and xxx, impact should be minimal. I will be monitoring graphs. In case of issues please alert me in #databases or #serving-team.”

[0:00]

Choose host to perform DRT on with Filesystems team
Check threads running for the primary host are low and steady



[0:20] Run command:

$ select hostname, global.rpl_semi_sync_slave_enabled

Expected outcome:

host1 1

host2 1

host3 1

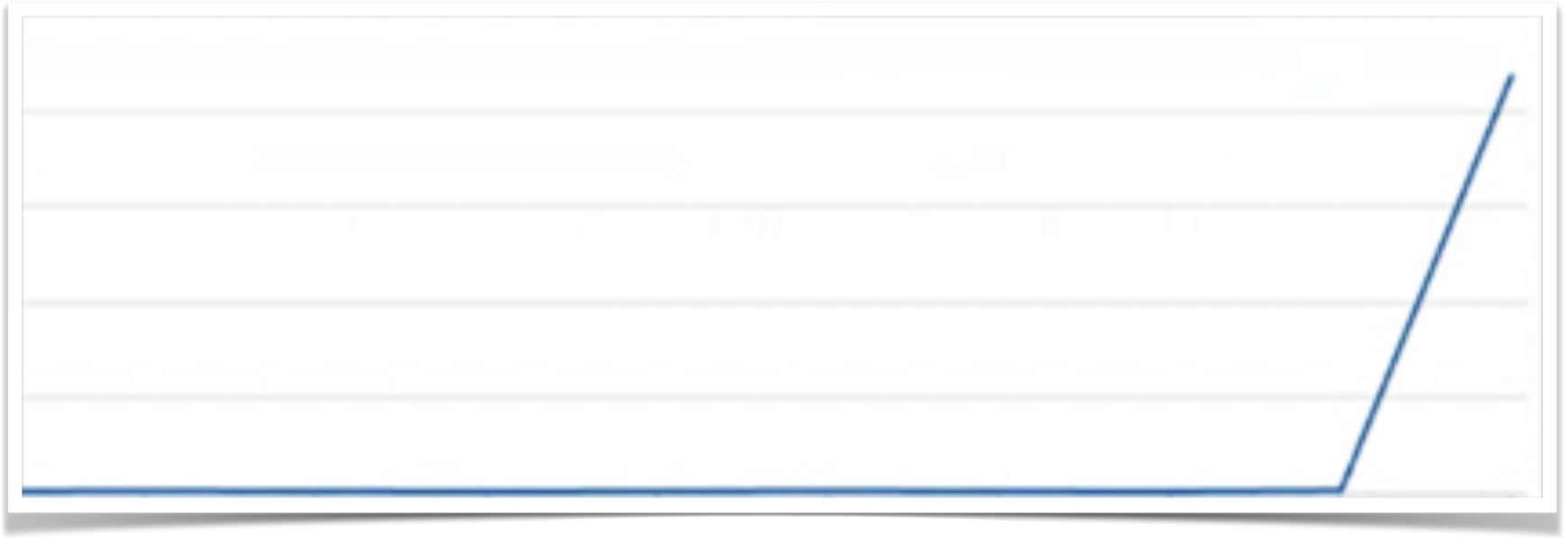
[0:22] Run command:

$ stop slave; set global rpl_semi_sync_slave_enabled=0; start slave;

Expected outcome:

“Waiting for semi-sync ACK from slave”

[0:26] Check if threads running has spiked on mysql perf dashboard



[0:27] Likely to see a spike in lock failures and commit errors going up on SFJ dashboards:

[0:30] Threads running will continue to rise

[0:32] You should receive a PagerDuty alert for “threads_running_sfj”

[0:34] End the DRT by enabling semi-sync, you will start to see it drop back down, it will take a few minutes

Run command:

$ stop slave; set global rpl_semi_sync_slave_enabled=1; start slave

[0:39] Expect to see a large drop in threads running

[0:40] DRT Complete, resolve PagerDuty alert

Running a SQL Proxy DRT

Assignee [Person executing]	Brian C	Fame
Assignee Team	Databases	Tag needed for board / tracking
Summary [Title / Description]	SQL Proxy Failure	Readability / Reference
Affected Systems	SQL Proxy	Tracking dependencies, educating for additional DRTs, future scheduling
**Cadence	Every quarter	[every six weeks, every two weeks, every quarter] determine effective scheduling / KTLO budgeting

[0:00]
Log into to any sql-proxy host

[0:20] Run command:
$ status sqlproxy

Expected outcome:
sqlproxy_0 SubTaskRunning started:2016-5-31T04:55:23Z uptime:

[0:22] Run command:

$ stop sqlproxy_global

[0:26] Wait for a few mins and monitor the availability graphs

[0:27] Now start sql proxy again

$ start sqlproxy_global

Expected outcome:

start successful: sqlproxy

[0:27] DRT passed!

Send an email that it is finished

Running a databases replica DRT

Dropbox

Assignee [Person executing]	Peter B	Fame
Assignee Team	Databases Filesystems	Tag needed for board / tracking
Summary [Title / Description]	Server File Journal Replica Failure	Readability / Reference
Affected Systems	Server File Journal (SFJ)	Tracking dependencies, educating for additional DRTs, future scheduling
**Living runbook link (e.g. paper doc)		Visibility into historical vs. current (could be a web link field)
**Cadence	Every quarter	[every six weeks, every two weeks, every quarter] determine effective scheduling / KTLO budgeting

We check that the daily push has happened or will happen later

We send an email to serving-announce mailing list

“At x:xx I will be performing a controlled replica failure for SFJ. This will affect xxx, xxx and xxx, impact should be minimal. I will be monitoring graphs. In case of issues please alert me in #databases or #serving-team.”

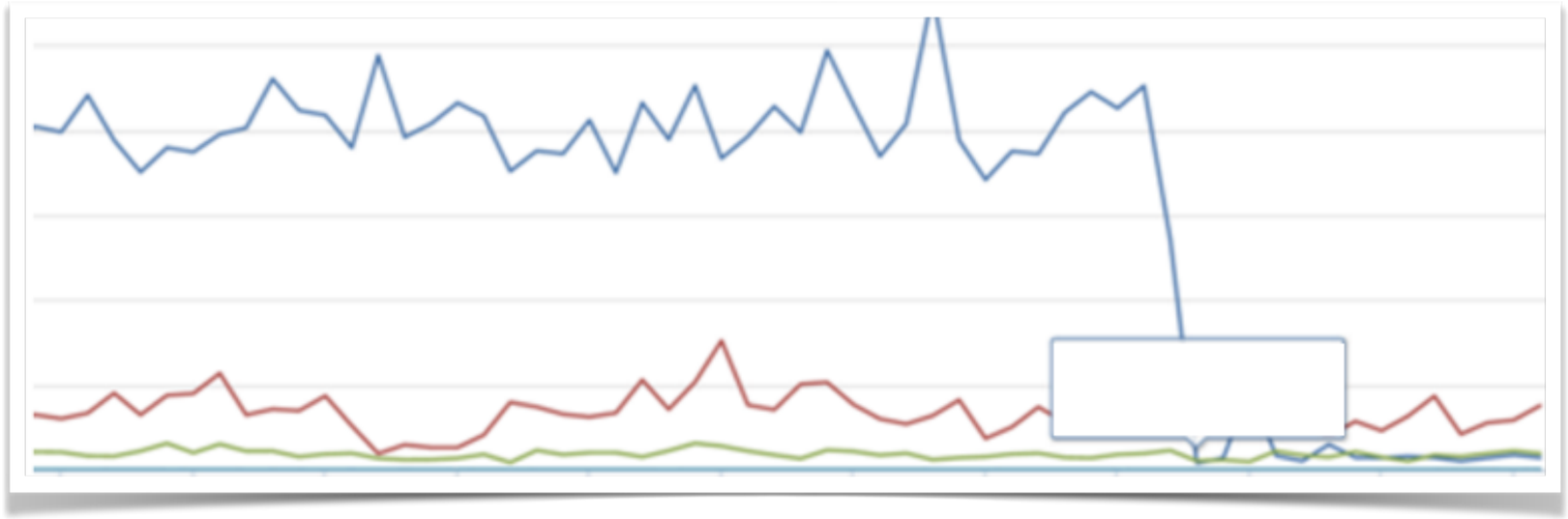
[0:00] Choose host to perform DRT on with Filesystems team

[0:10] Fail one replica from production by killing mysqld on host

[0:13] Filesystems team will let you know that the build is back to green

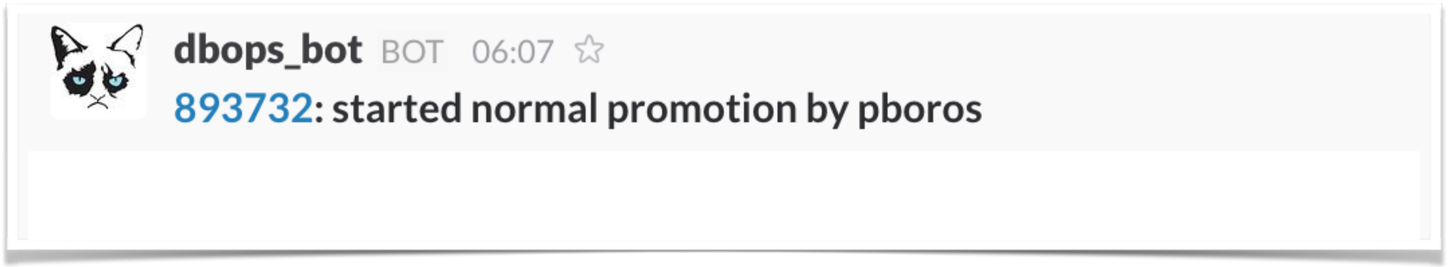
[0:15] InnoDB reads will drop to almost 0

[0:15] auto_replace script will kick in and cloning will commence



DRT complete and passed:

- dbops bot will post in slack
- Clone will complete successfully



Running an Edgestore stage DRT

Dropbox

[0:20] Run command:

$ select hostname, global.rpl_semi_sync_slave_enabled

Expected outcome:

host1 1

host2 1

host3 1

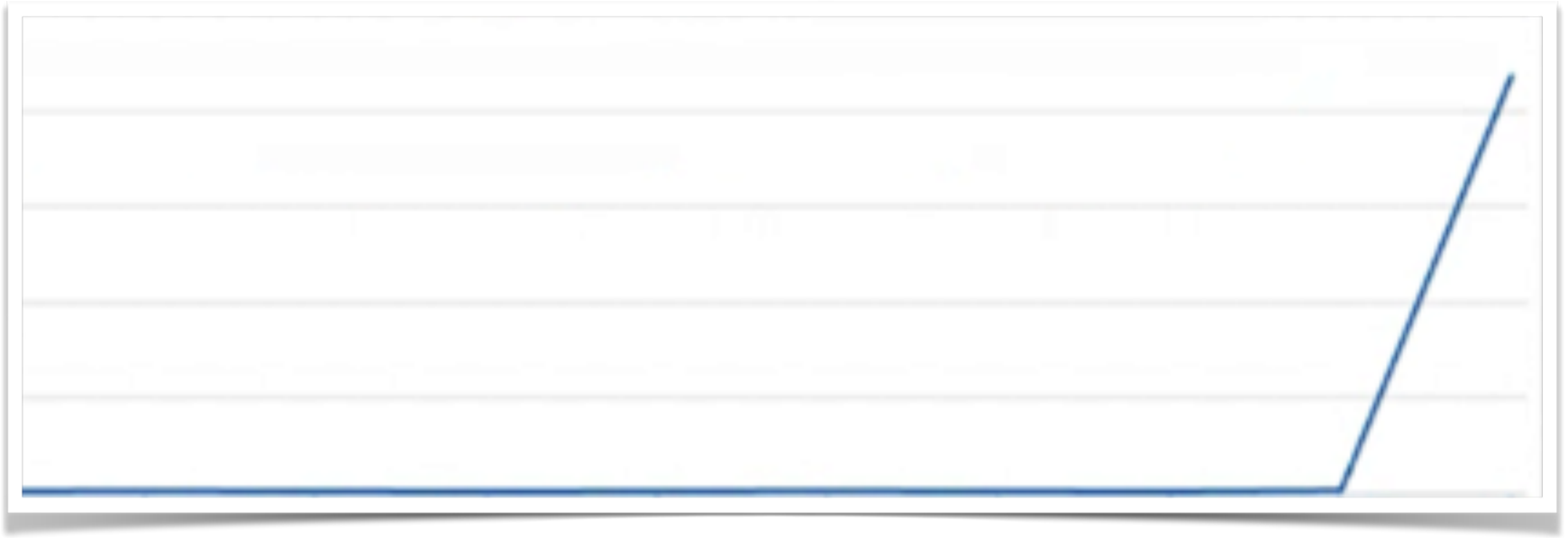
[0:22] Run command:

$ stop slave; set global rpl_semi_sync_slave_enabled=0; start slave;

Expected outcome:

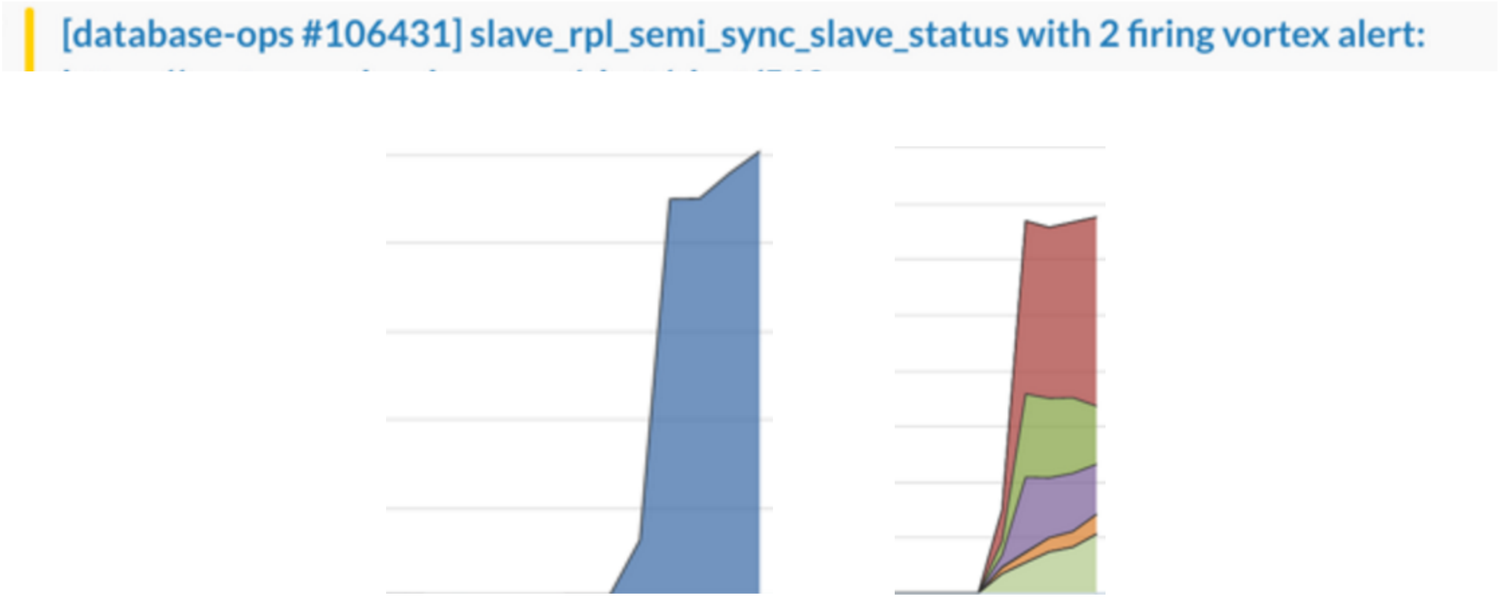
“Waiting for semi-sync ACK from slave”

[0:26] Check if threads running has spiked on mysql perf dashboard



[0:27] Likely to see a spike in errors on the edgestore stage dashboard

You will see alerts in Slack “slave_rpl_semi_sync_slave_status”



[0:30] Threads running and threads connected will continue to rise

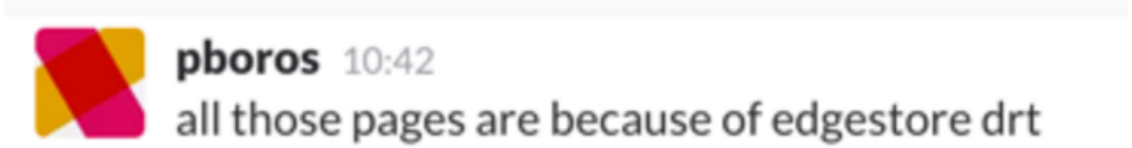
[0:32] You should receive a PagerDuty alert

[0:34] End the DRT by enabling semi-sync, you will start to see it drop back down, it will take a few minutes

Run command:

$ stop slave; set global rpl_semi_sync_slave_enabled=1; start slave

DRT is finished!



Learning from databases DRTs

AIs

DRIs

Edit

Recommended by Vaibhav Bhembre, Richard Murby, and 1 other

Tammy Butow

SRE Manager @Dropbox ☁️❤ I like databases, storage, automation, performance, go, Linux & death ...

Apr 16 · 6 min read

Better Sprints with Team Traditions

3

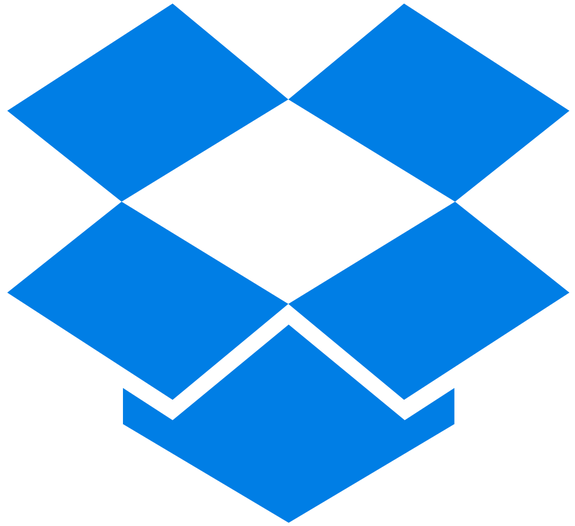
Tammy Butow

https://medium.com/@tammybutow/better-sprints-with-team-traditions

Improvements For Distributed Teams

The Future of anti-fragile at Dropbox

Dropbox



0-100 Days - Running DRTs at Dropbox


Thomissa
@thomissa
TPM @Dropbox, Alaskan. Previously @StanfordEng @TeslaMotors
linkedin.com/in/thomissa
Joined March 2011

Thank you!

Q & A


Tammy Bütow
@tammybutow
SRE Manager @Dropbox Metadata Storage. I like team traditions, databases, automation, Go, Linux & death metal. \m/ Previously @DigitalOcean @NAB @QUT.
San Francisco, CA
tammybutow.com
Joined June 2009