Resilient Routing and Discovery

Simon Eskildsen, Shopify

@Sirupsen

dockercon 15

SF JUNE 22-23

shopify

Shopify

Docker in Production serving the below for 1+ year

165,000+

ACTIVE SHOPIFY MERCHANTS

200+

DEVELOPERS

2

DATACENTERS

3000+

CONTAINERS RUNNING AT ANY TIME

12+

DEPLOYS PER DAY

10,000+

MAX CHECKOUTS PER MINUTE

$8 BILLION+

CUMULATIVE GMV

500+

SERVERS

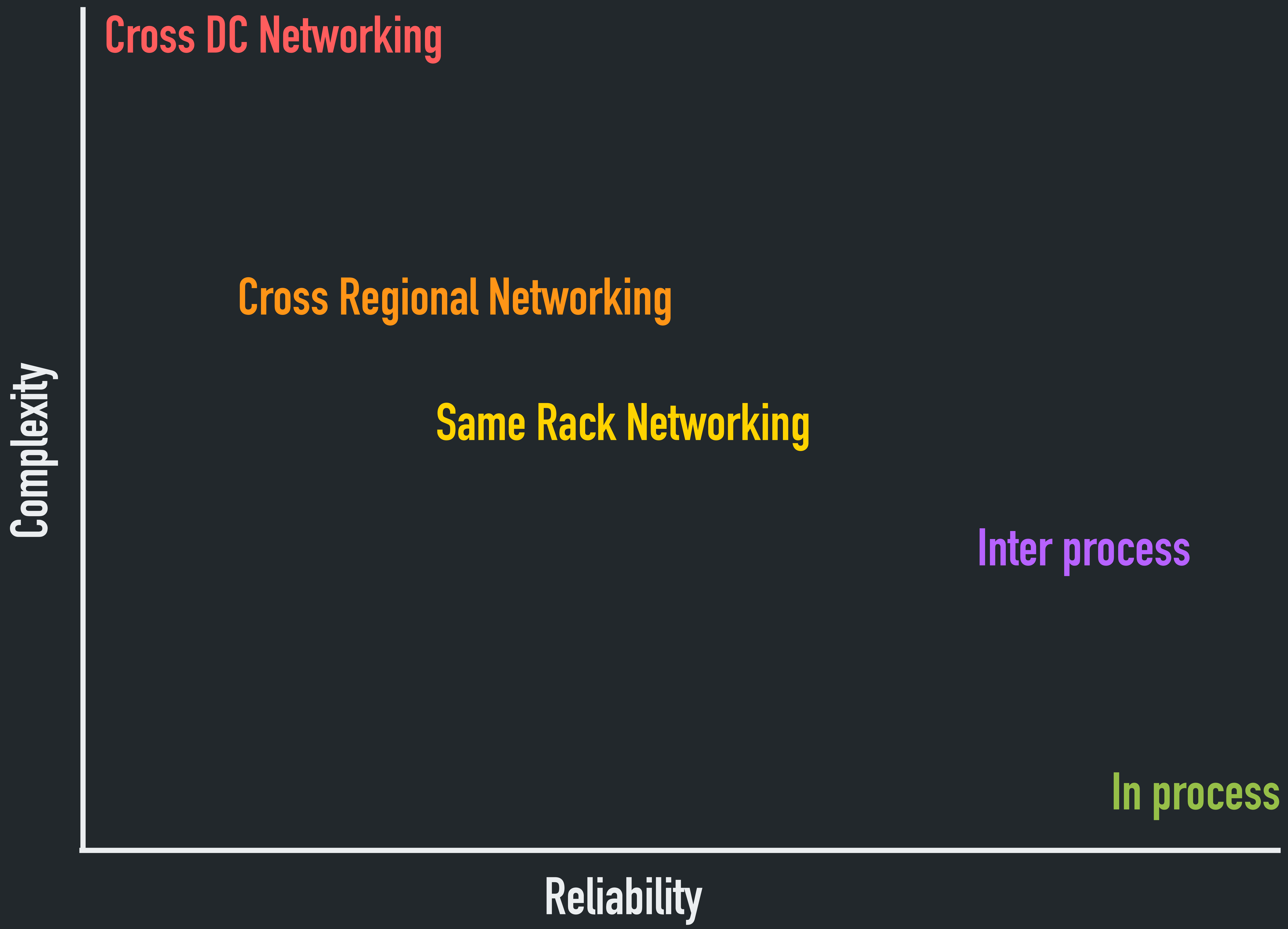
Ruby on Rails

10+ years old

300M unique visits/month

LEAGUE OF APPLE, EBAY AND AMAZON

Building reliable bridges in large distributed systems



Cross DC Networking

Cross Regional Networking

Same Rack Networking

Inter process

In process

Complexity

Reliability

Resiliency

Discovery

Routing

Reliability is your success metric for discovery and routing.

Shopify started this journey in the fall of 2014

Resiliency

Building a reliable system from unreliable components

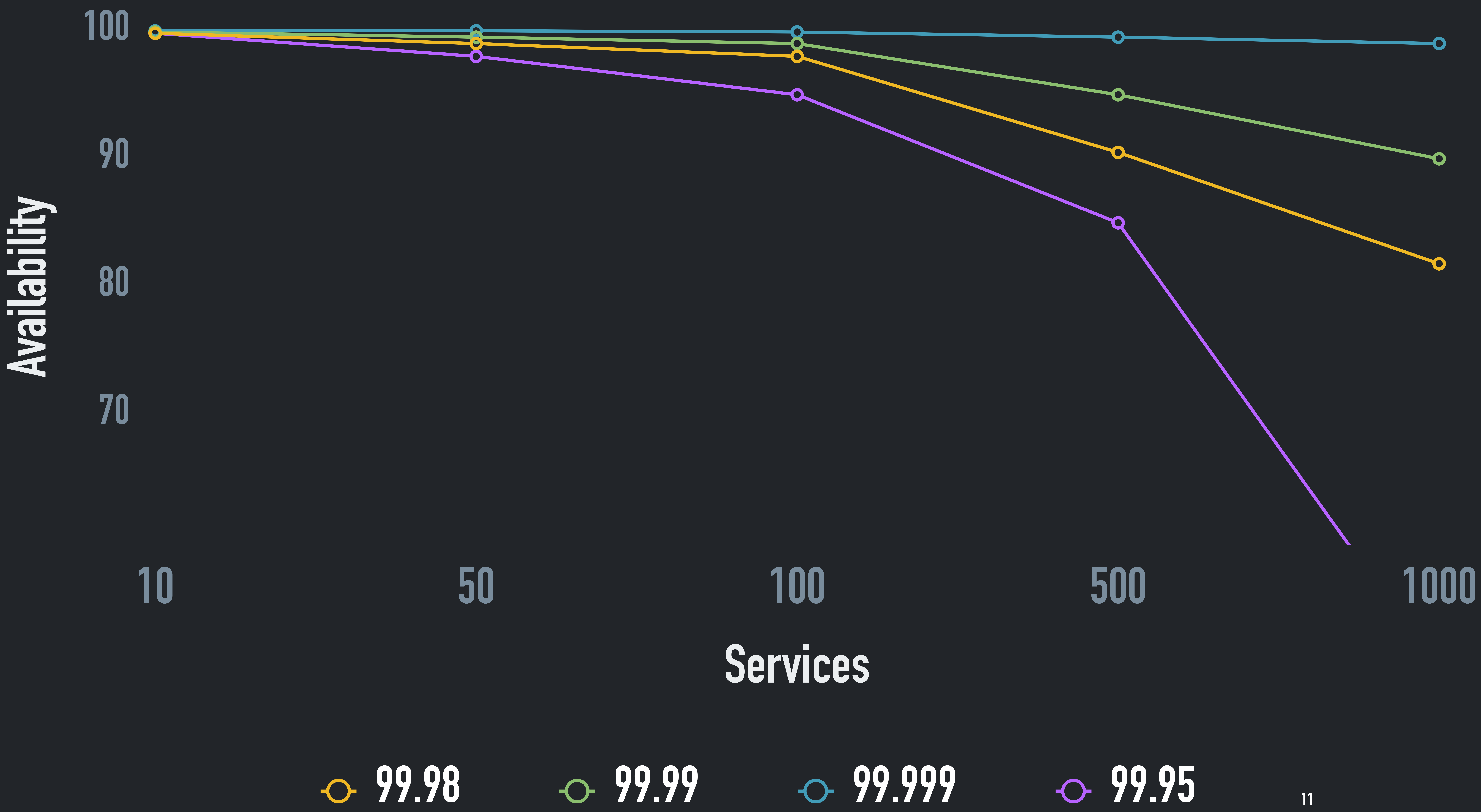
(Micro)service equation

$$\text{Uptime} = A^N$$

Number of services

Availability per service

Total availability


100
90
80
70
Availability
10
50
100
500
1000
Services
99.98
99.99
99.999
99.95

Resiliency Matrix

	Checkout	Admin	Storefront
MySQL Shard	Unavailable	Unavailable	Degraded
MySQL Master	Available	Unavailable	Available
Kafka	Available	Degraded	Available
External HTTP API	Degraded	Available	Unavailable
redis-sessions	Unavailable	Unavailable	Degraded

Objectives for large distributed systems

Building reliable systems from unreliable components

Explore resiliency, service discovery, routing, orchestration and the relationship between them

Recognizing and avoiding premature optimizations and overcompensation

Application should be designed to handle fallbacks

KITH
View Account (Logged in as Simon) | Logout ✕
SHOP
BRANDS
BLOG
LOOKBOOKS
ABOUT
LOCATIONS
SUBSCRIBE
SUPPORT
CART (1)
ALL
SNEAKERS
BOOTS
SHOES
SANDALS
NEWEST TO OLDEST ▼

PUMA X HOUSE OF HACKNEY BECKER - WHITE / GREEN FLORAL

PUMA

$160.00

PUMA X HOUSE OF HACKNEY BASKET - TOTAL ECLIPSE / GREEN PALM

PUMA

$125.00

NEW BALANCE M998 "MONTAUK" - GREY / BLUE / GREEN

NEW BALANCE

$170.00

NEW BALANCE M997 "DISTINCT" - OLIVE / BROWN

NEW BALANCE

$260.00

search

carts

sessions

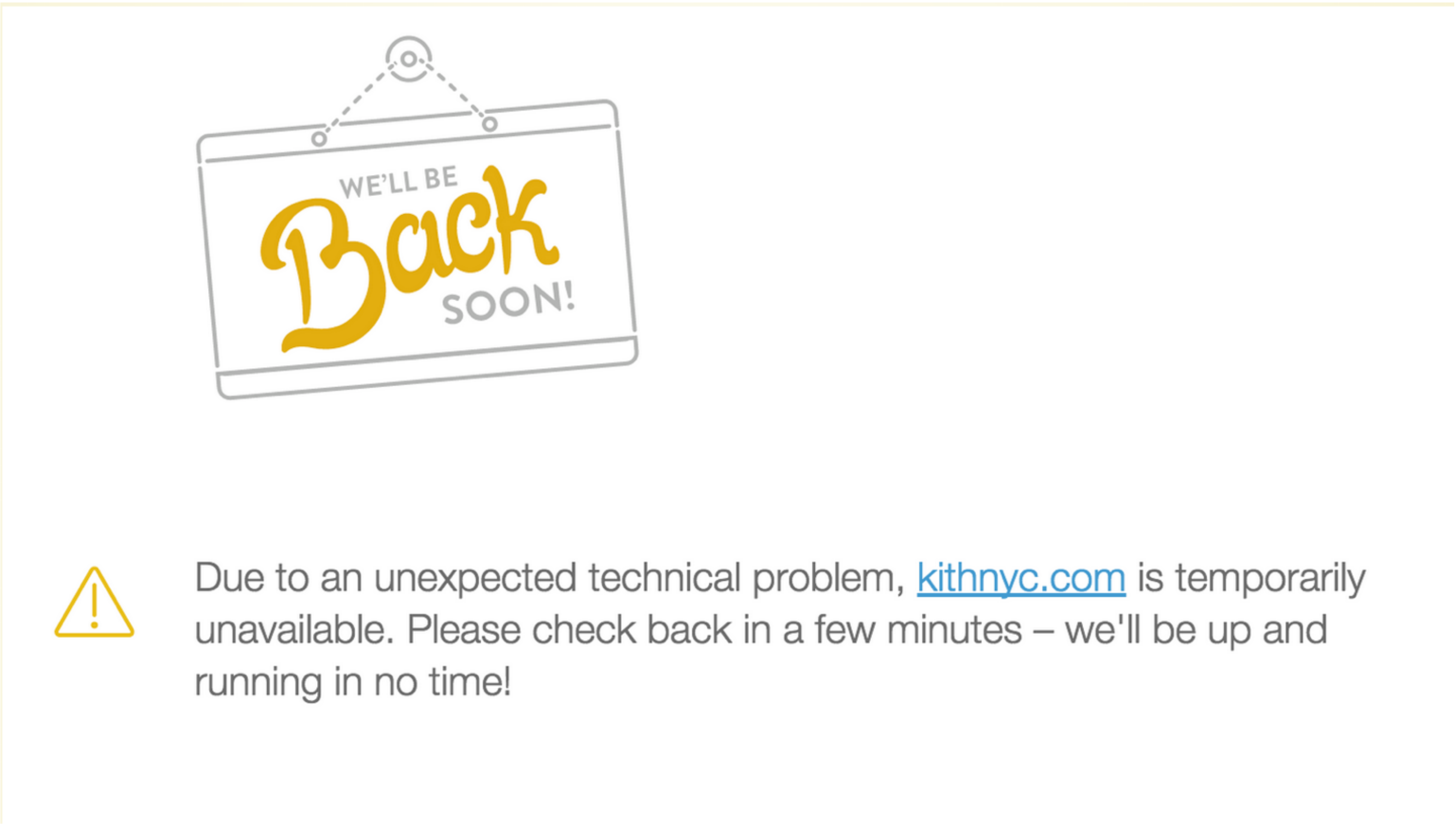
cdn

mysql

KITH
View Account (Logged in as Simon) | Logout ✕
SHOP
BRANDS
BLOG
LOOKBOOKS
ABOUT
LOCATIONS
SUBSCRIBE
SUPPORT
CART (1)
ALL
SNEAKERS
BOOTS
SHOES
SANDALS
NEWEST TO OLDEST
PUMA X HOUSE OF HACKNEY BECKER - WHITE / GREEN FLORAL
PUMA
$1[illegible]
PUMA X HOUSE OF HACKNEY BASKET - TOTAL ECLIPSE / GREEN PALM
PUMA
$125.00
NEW BALANCE M998 "MONTAUK" - GREY / BLUE / GREEN
NEW BALANCE
$170.00
NEW BALANCE M997 "DISTINCT" - OLIVE / BROWN
NEW BALANCE
$260.00

Avoid HTTP 500 for single service failing

.. or suffer the faith of the (micro)service equation



⚠ Due to an unexpected technical problem, kithnyc.com is temporarily unavailable. Please check back in a few minutes – we'll be up and running in no time!

Sessions data store unavailable

KITH
Search
Account
SHOP
BRANDS
BLOG
LOOKBOOKS
ABOUT
LOCATIONS
SUBSCRIBE
SUPPORT
CART
ALL
SNEAKERS
BOOTS
SHOES
SANDALS
NEWEST TO OLDEST

PUMA X HOUSE OF HACKNEY BECKER - WHITE / GREEN FLORAL

PUMA

$160.00

PUMA X HOUSE OF HACKNEY BASKET - TOTAL ECLIPSE / GREEN PALM

PUMA

$125.00

NEW BALANCE M998 "MONTAUK" - GREY / BLUE / GREEN

NEW BALANCE

$170.00

NEW BALANCE M997 "DISTINCT" - OLIVE / BROWN

NEW BALANCE

$260.00

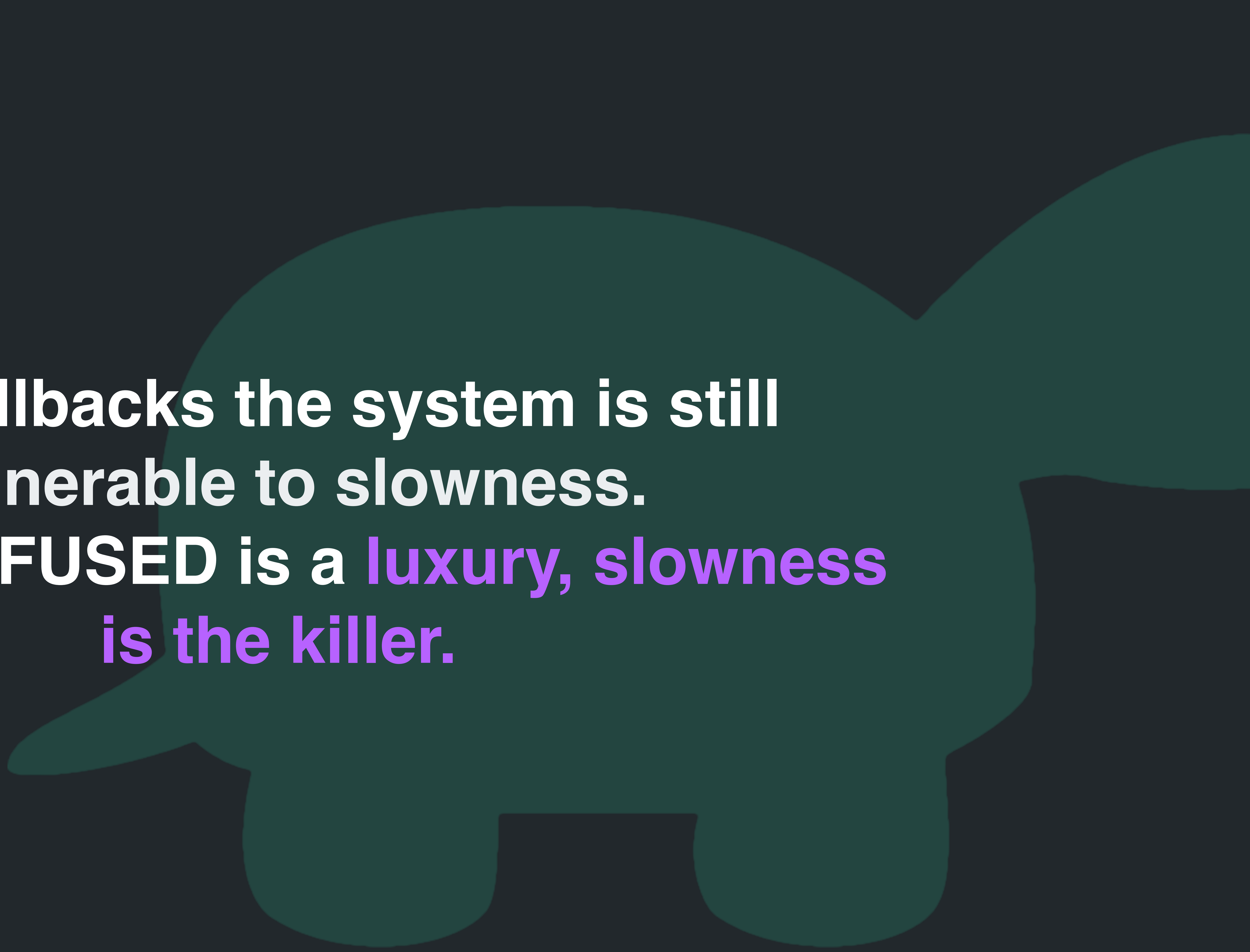
Customer signed out

Simulate TCP conditions with Toxiproxy

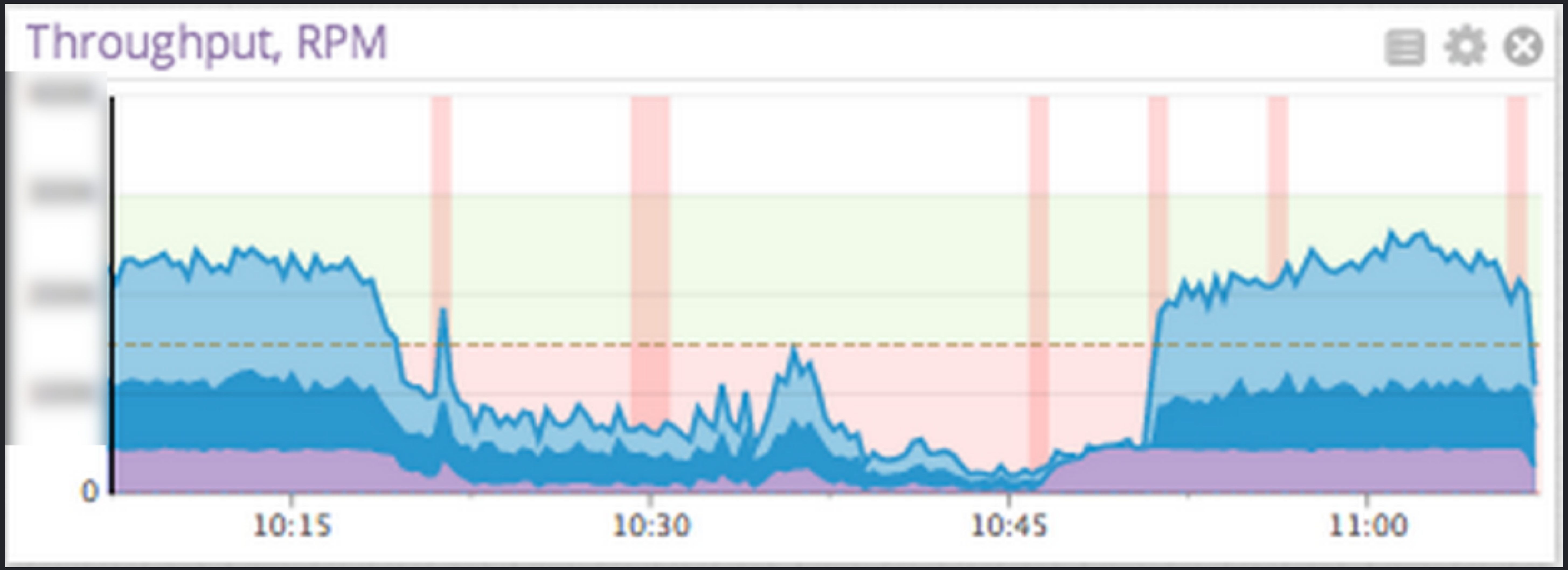
```
curl -i -d '{"enabled":true, "latency":1000}' \
  localhost:8474/proxies/redis/downstream/toxics/latency

curl -i -X DELETE localhost:8474/proxies/redis


Toxiproxy[:mysql_master].downstream(:latency, latency: 1000).apply do
  Shop.first # this takes at least 1s
end

Toxiproxy[/redis/].down do
  session[:user_id] # this will throw an exception
end
```

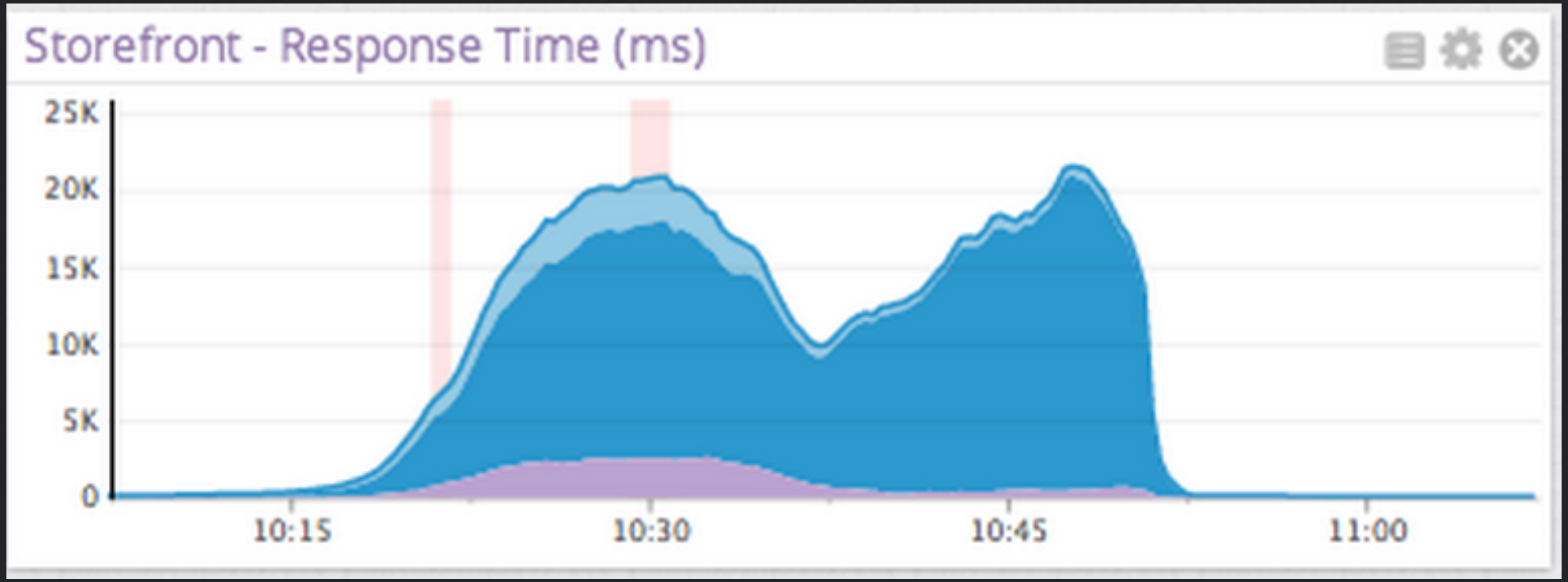
https://github.com/shopify/toxiproxy

With fallbacks the system is still vulnerable to slowness. ECONNREFUSED is a luxury, slowness is the killer.

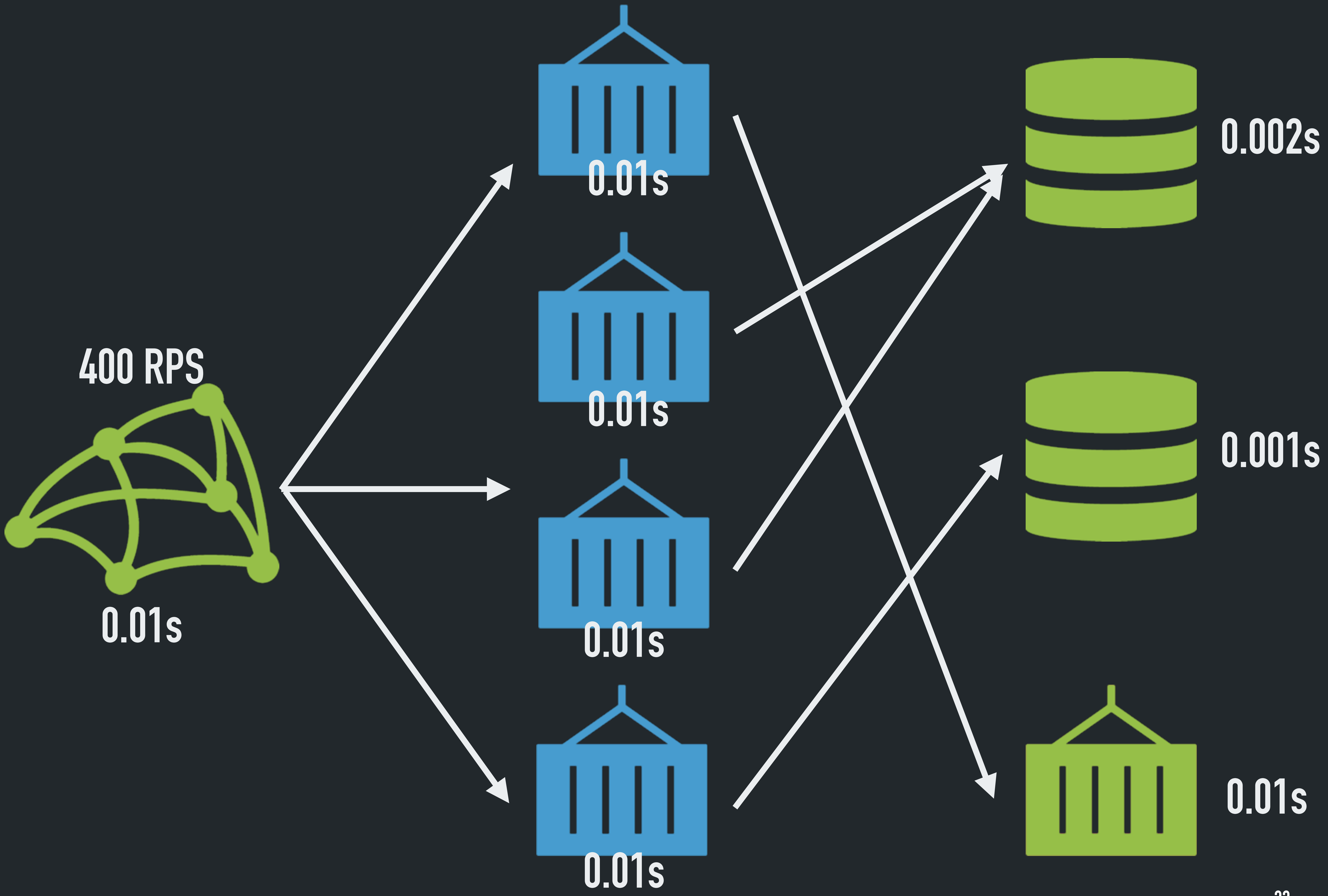


Little’s law

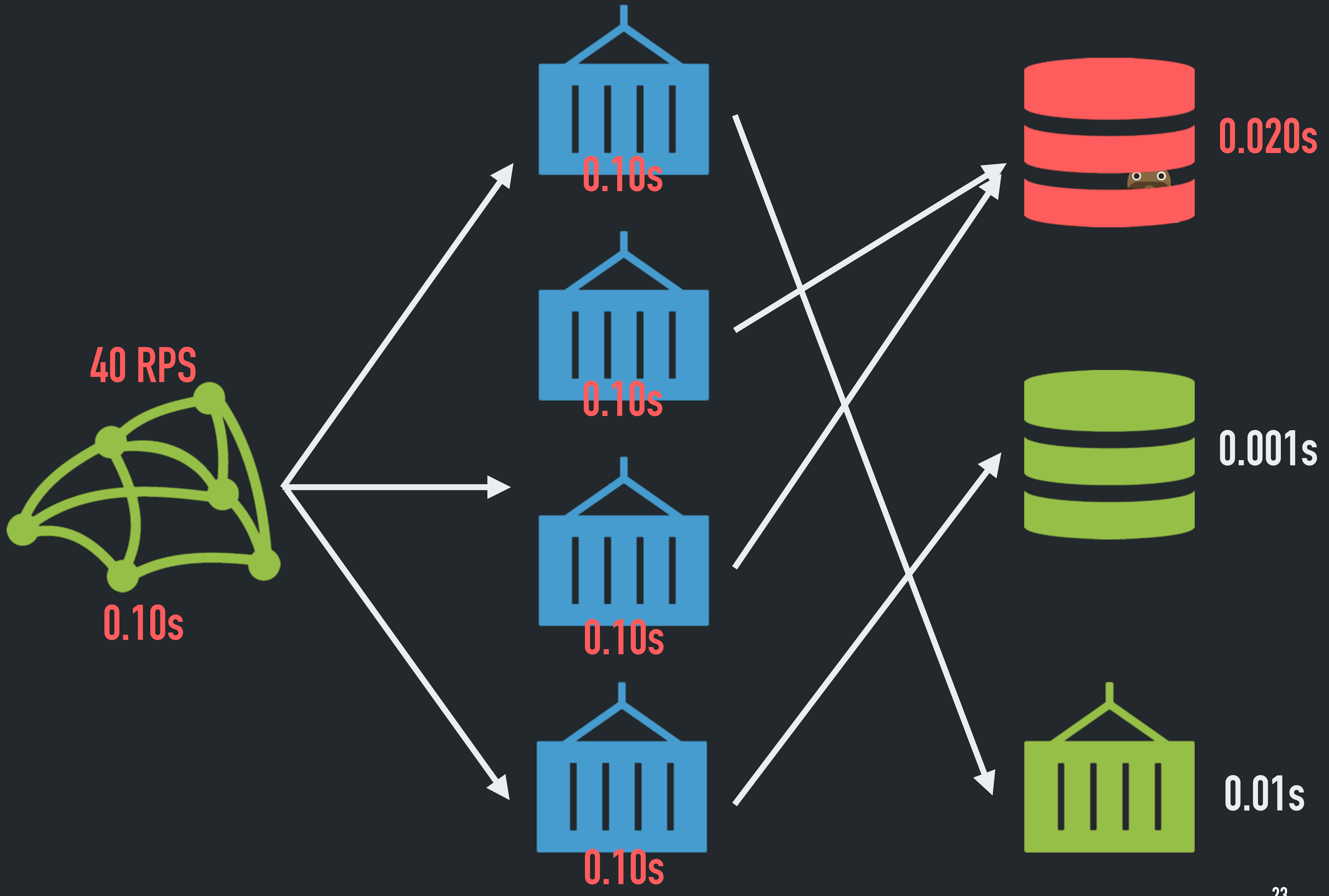

Throughput, RPM
0
10:15
10:30
10:45
11:00


Storefront - Response Time (ms)
25K
20K
15K
10K
5K
0
10:15
10:30
10:45
11:00

Infrastructure operating normally



Database latency increases by 10x, throughput drops 10x



Beating Little’s law is your first priority as you add services

Resiliency Toolkits

Bulk Heads, Circuit Breakers, ..

The Pragmatic Programmers

Release It!

Design and Deploy Production-Ready Software

Michael T. Nygard

Release It book

finagle

twitter/finagle

HYSTRIX

DEFEND YOUR APP

netflix/hystrix

shopify/semian

NETFLIX

Resiliency Maturity Pyramid

Region Gorilla

Latency Monkey

Kill Nodes (Chaos Monkey)

Production Practise Days (Games)

Resiliency Patterns

Application-Specific Fallbacks

Toxiproxy tests and matrix

Testing with mocks

No resiliency effort

Discovery

Infrastructure source of truth

Services

Instances of services

Metadata

Deployed revision, leader, ..

Orchestration

Aid to make things happen across components

Location

Global

Geo-replicated discovery

Regional

Single datacenter

Discovery Backbone Properties

No single point of failure

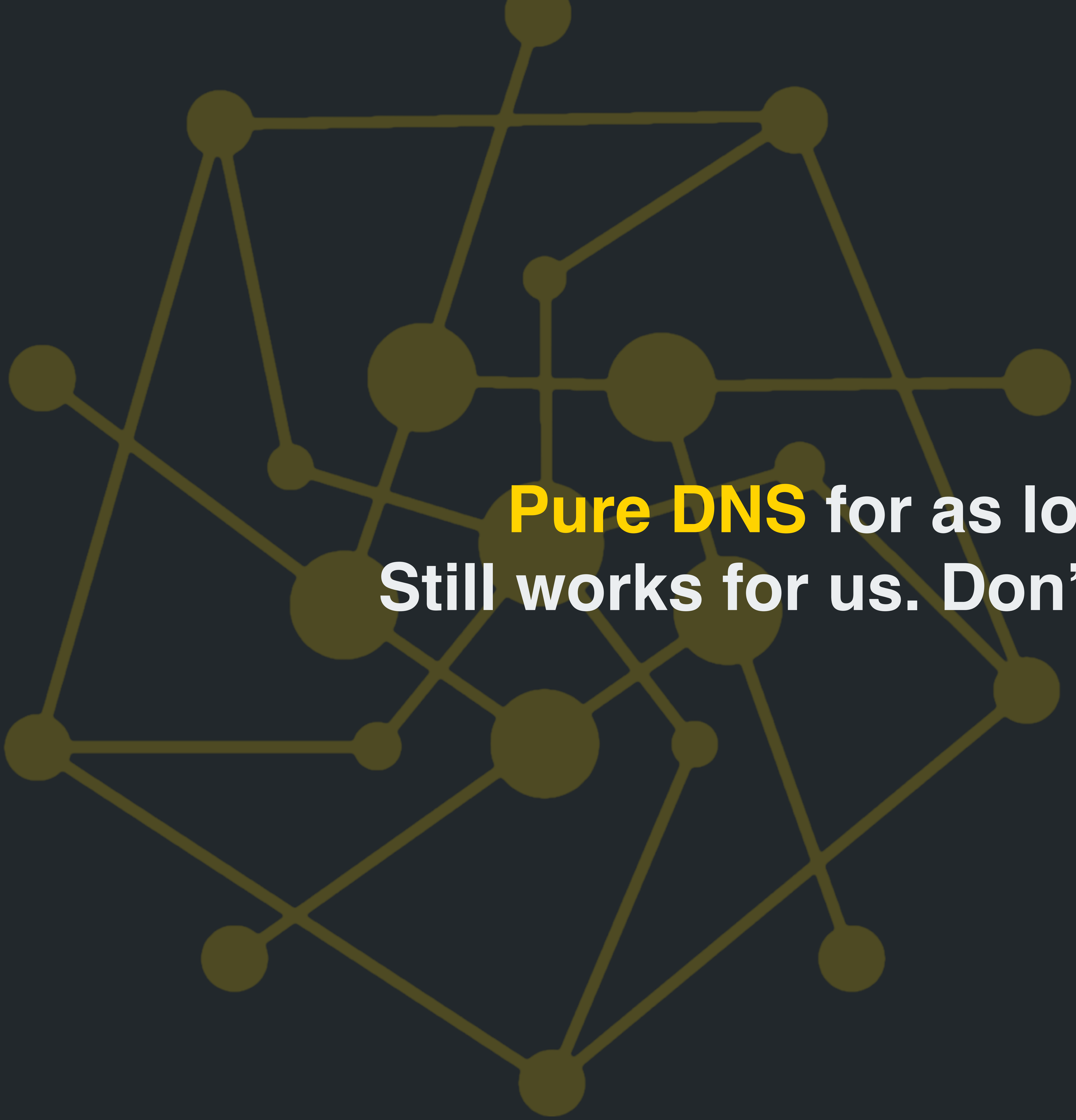
Stale reads better than no reads: A > C

Reads order of magnitude larger than writes

Fast convergence

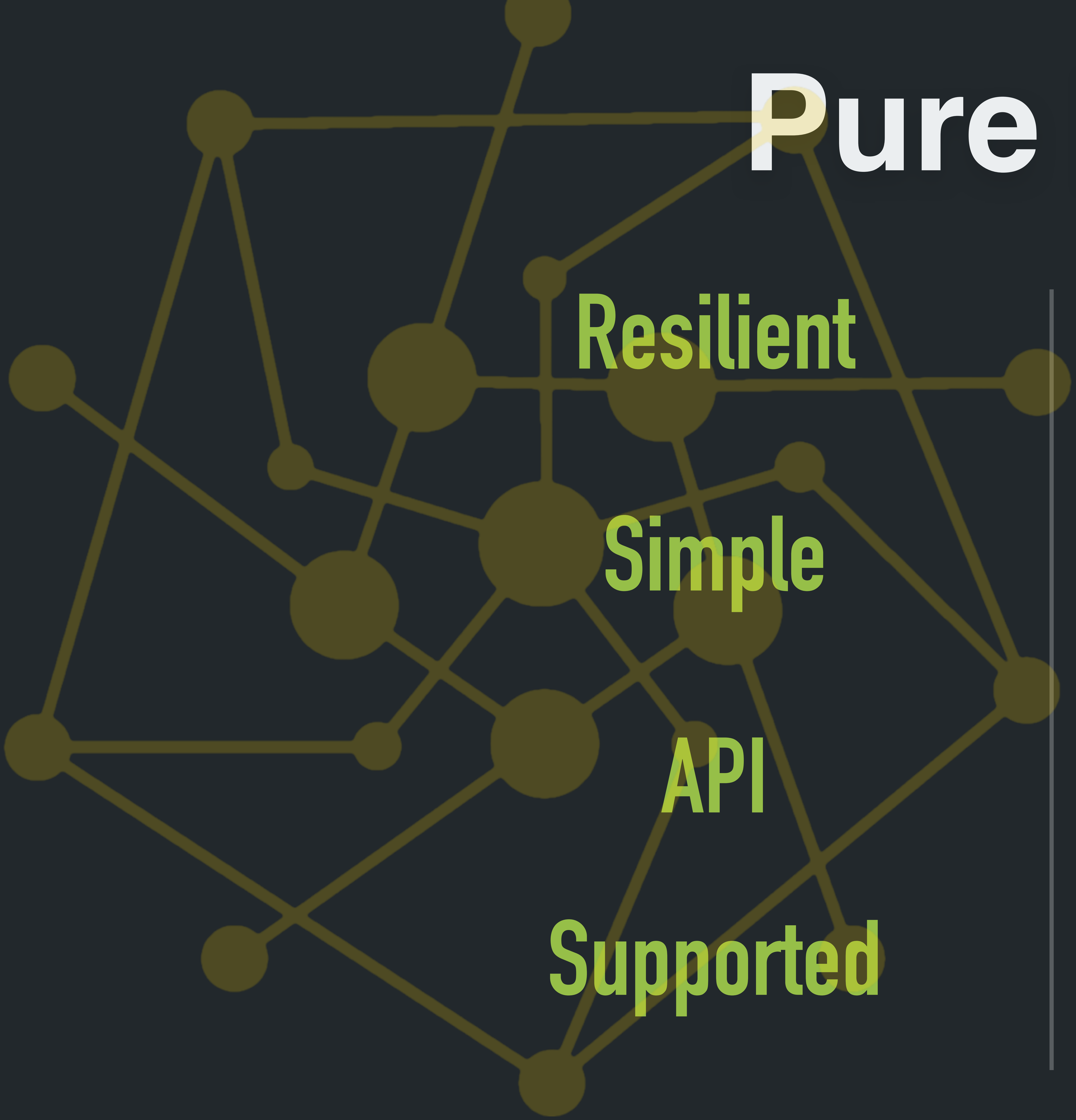
New and Old School

Consul	DNS
Zookeeper	Chef, Puppet, ..
Eureka	Network
Etcd	Hardcoded values



Pure DNS for as long as you can.
Still works for us. Don't overcompensate.

Pure DNS



- Resilient
- Simple
- API
- Supported

- Failovers?
- Slow convergence
- Not a data store
- Not for orchestration

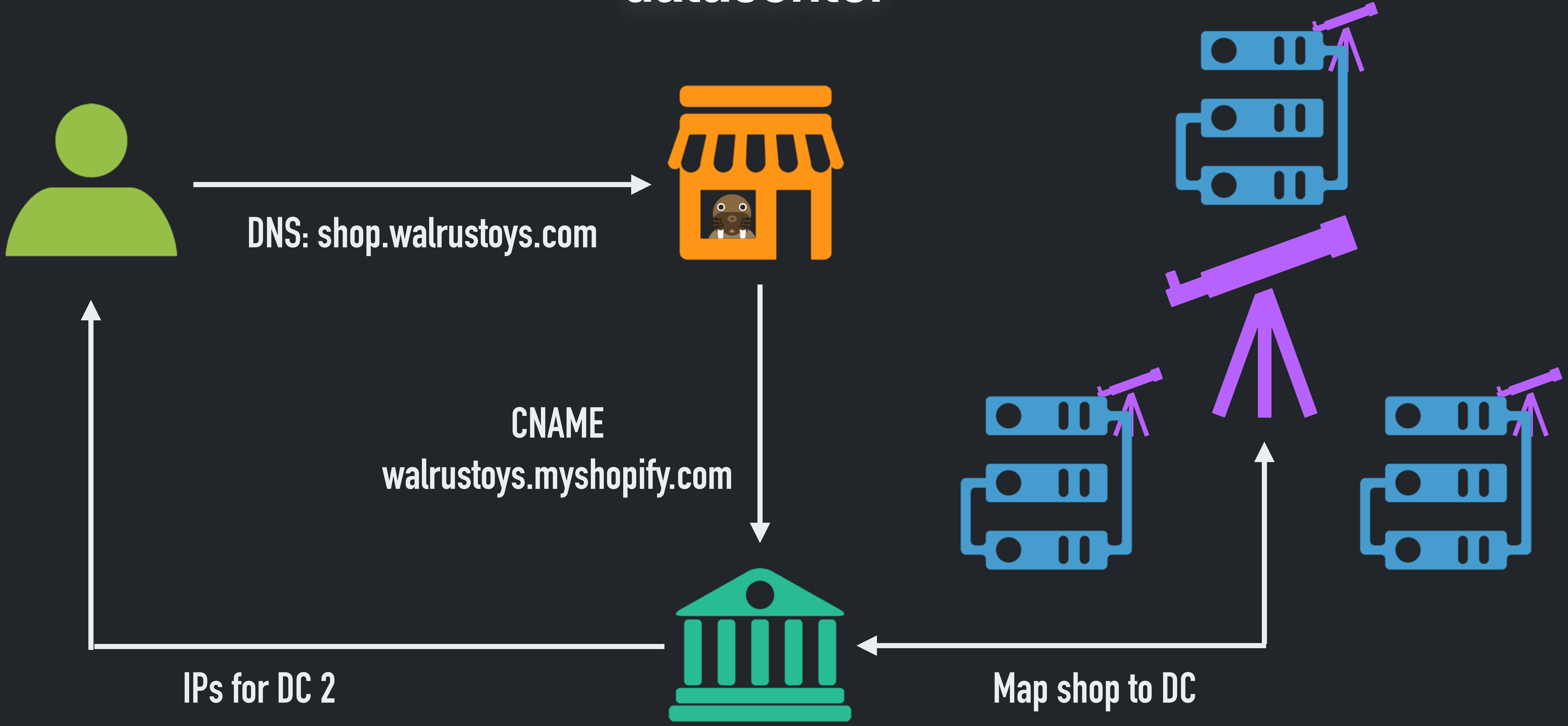
Global discovery and orchestration most pressing issue for Shopify

Orchestration of datacenter failovers

Too many Sources of Truth

Component	Source of Truth
Network	NetEng?
MySQL	DBAs?
Application	Cookbooks
Redis	Cookbooks
Load Balancers	Hardcode value in config file

Routing shops to the right datacenter



DNS: shop.walrustoys.com

CNAME walrustoys.myshopify.com

IPs for DC 2

Map shop to DC

DNS problematic when..

Multiple owners of data

Fast converge

Lots of change in instances

Zookeeper

- Scalable stale reads
- Consistent
- Orchestration
- Trusted

- Not complete discovery
- Complex clients
- Operational burden
- Shoehorn

Complex client problem

- Connecting directly risky
- Proxy pattern
- Dumping to files
- Stale reads

Routing

Routing responsibilities

Load balance

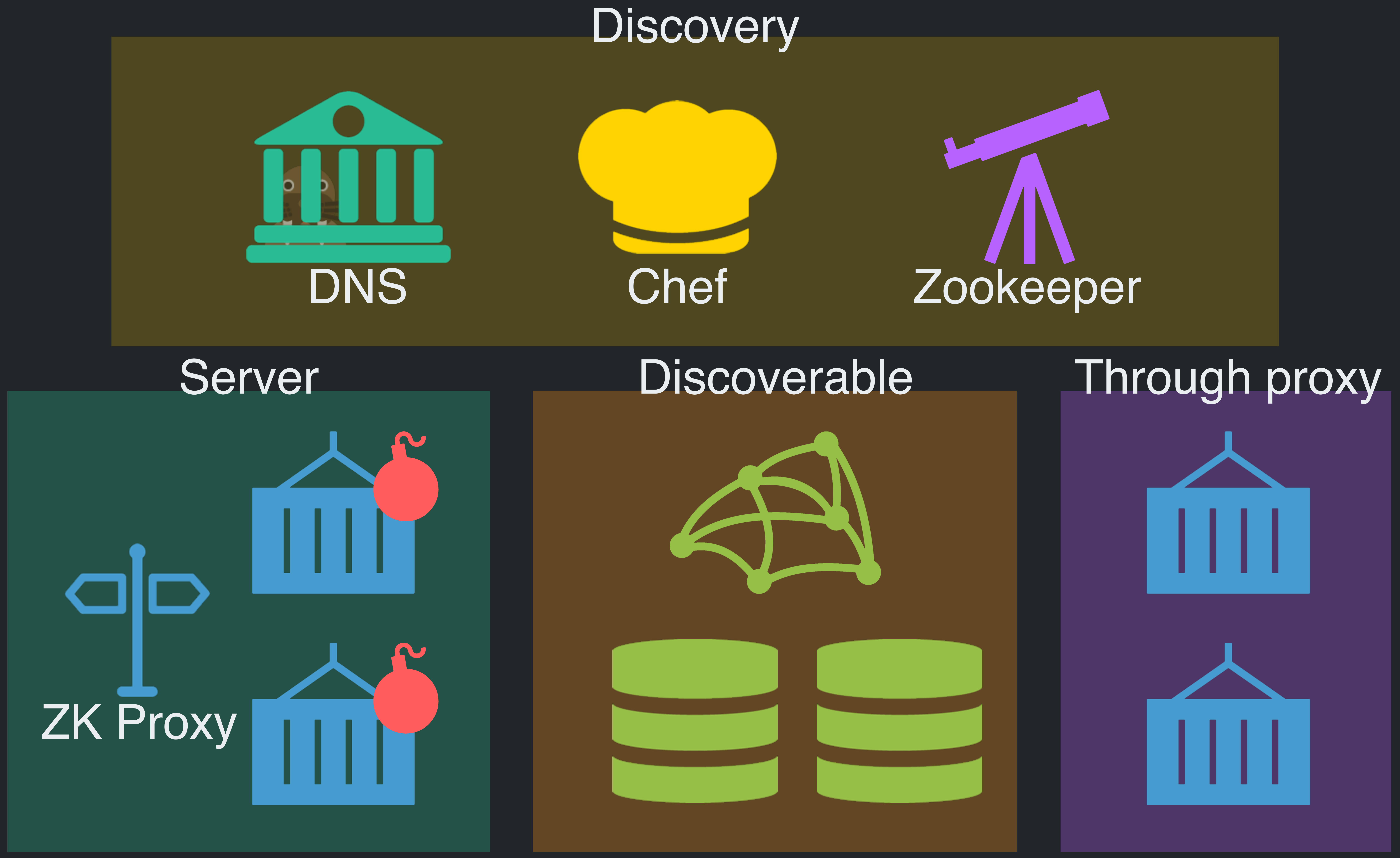
Protect applications against unhealthy resources: circuit breaker, bulk heads, rate limiting, …

Receive upstreams from discovery layer

	Trusted	Scriptable	Resiliency	Dynamic upstreams	Discovery built in	TCP	Library/Proxy
yours	Don’t do this	Of course	It’s perfect	I got it	Easy	Obviously, it’s Go	
OS nginx	YES	3rd party (ngx-lua). Not complete (no TCP support).	Possible for HTTP via ngx-lua. No TCP yet	Sidekick for new upstreams. Manipulate existing via ngx-lua	No, try via sidekick/ ngx-lua	Landed in 1.9.0, stabilized in nginx+	Proxy
haproxy	YES	Lua support in master	Not scriptable, only rate limiting built-in	Sidekick and reloads (with iptables wizardry), manipulate existing admin socket	No, try via sidekick	Built as L4	Proxy
vulcand	Maybe?	middlewares, requires forking	SOME, only circuit breaker	Beautiful HTTP API	etcd support	No, only supports HTTP currently (not in ROADMAP.md)	Proxy
finagle	YES	YES, completely centered around plugins	YES, sophisticated FailFast module	YES	Zookeeper support	Application-level	Library, requires JVM
smartstack	Somewhat	However much HAProxy is, adapters	NO, same as HAProxy	YES	Zookeeper support	Yes, uses HAProxy	Proxy + discovery

With a polyglot stack, we just use simple proxies and DNS

Current Stack



Discovery

DNS

Chef

Zookeeper

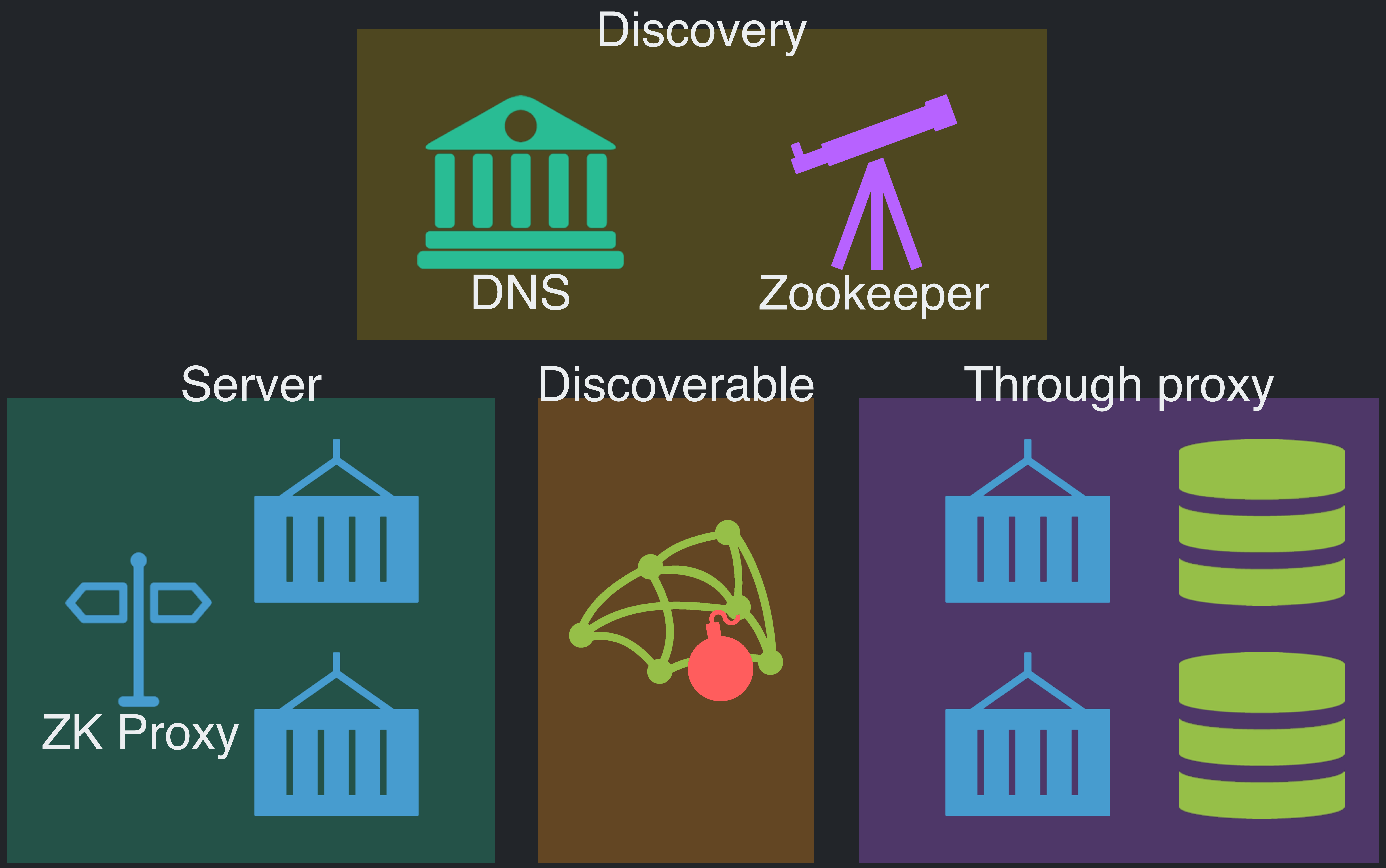
Server

ZK Proxy

Discoverable

Through proxy

Future Stack



Discovery

DNS

Zookeeper

Server

ZK Proxy

Discoverable

Through proxy

Docker's future role in discovery, routing and resiliency

Final remarks

Figure out service discovery value for your company, don’t overcompensate—your metric is reliability

Infrastructure teams own integration points, don’t leave it up to everyone to jump in

Build resiliency into the system, don’t make it opt in, be able to reason about entire system’s state and test

Thank you

Simon Eskildsen, Shopify

@Sirupsen

Server by Konstantin Velichko from the Noun Project

basket by Ben Rex Furneaux from the Noun Project

container by Creative Stall from the Noun Project

people by Wilson Joseph from the Noun Project

mesh network by Lance Weisser from the Noun Project

Conductor by By Luis Prado from the Noun Project

Jar by Yazmin Alanix from the Noun Project

Broken Chain by Simon Martin from the Noun Project

Book by Ben Rex Furneaux from the Noun Project

network by Jessica Coccimiglio from the Noun Project

server by Creative Stall from the Noun Project

components by icons.design from the Noun Project

switch button by Marco Olgio from the Noun Project

Pile of leaves (autumn) by Aarthi Ramamurthy

Bridge by Toreham Sharman from the Noun Project

collaboration by Alex Kwa from the Noun Project

converge by Creative Stall from the Noun Project

change by Jorge Mateo from the Noun Project

tag by Rohith M S from the Noun Project

whale by Christopher T. Howlett from the Noun Project

file by Marlou Latourre from the Noun Project

Signpost by Dmitry Mirolyubov from the Noun Project

Arrow by Zlatko Najdenovski from the Noun Project

Chef by Ross Sokolovski from the Noun Project