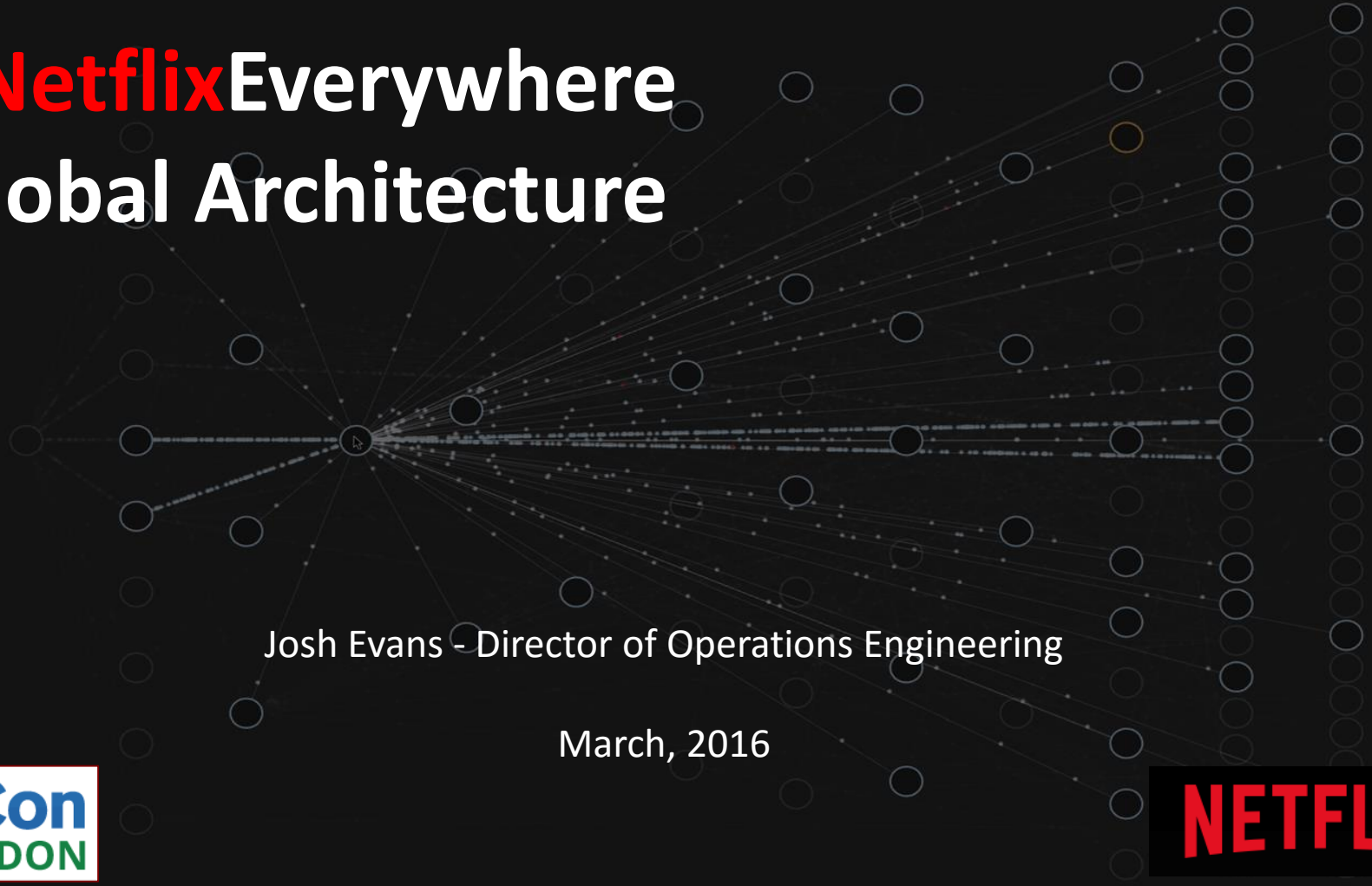
#NetflixEverywhere Global Architecture



Josh Evans - Director of Operations Engineering

March, 2016

QCon LONDON

NETFLIX

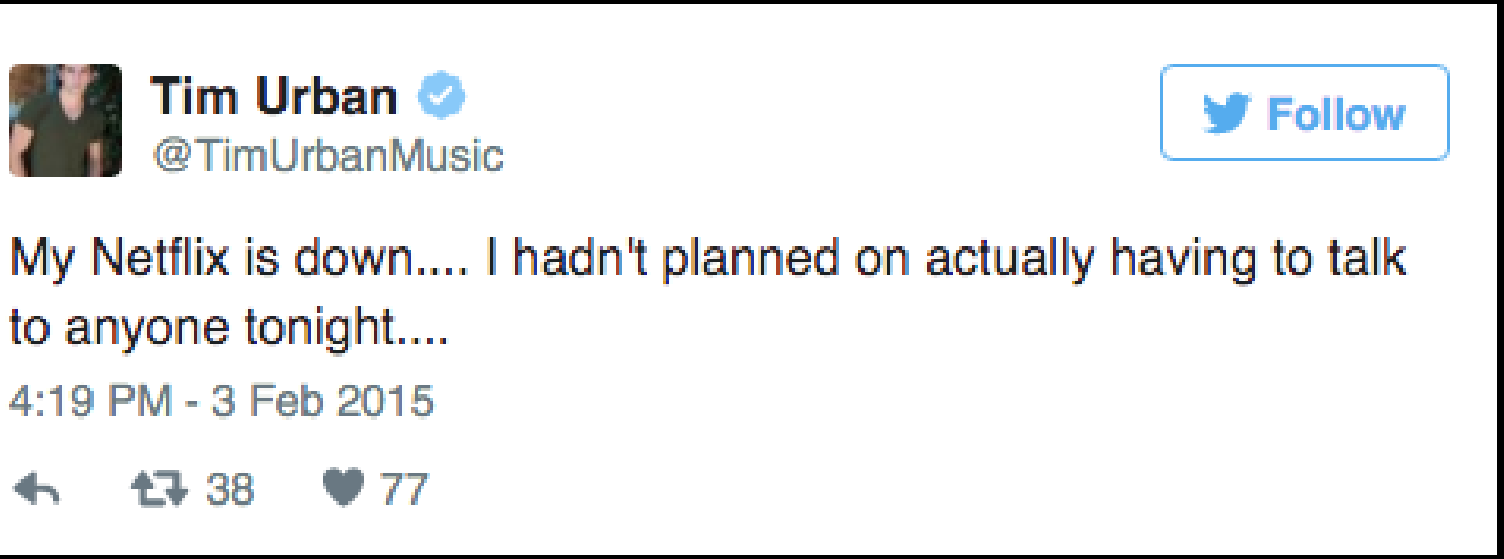
December 24th, 2012

Forbes / Tech

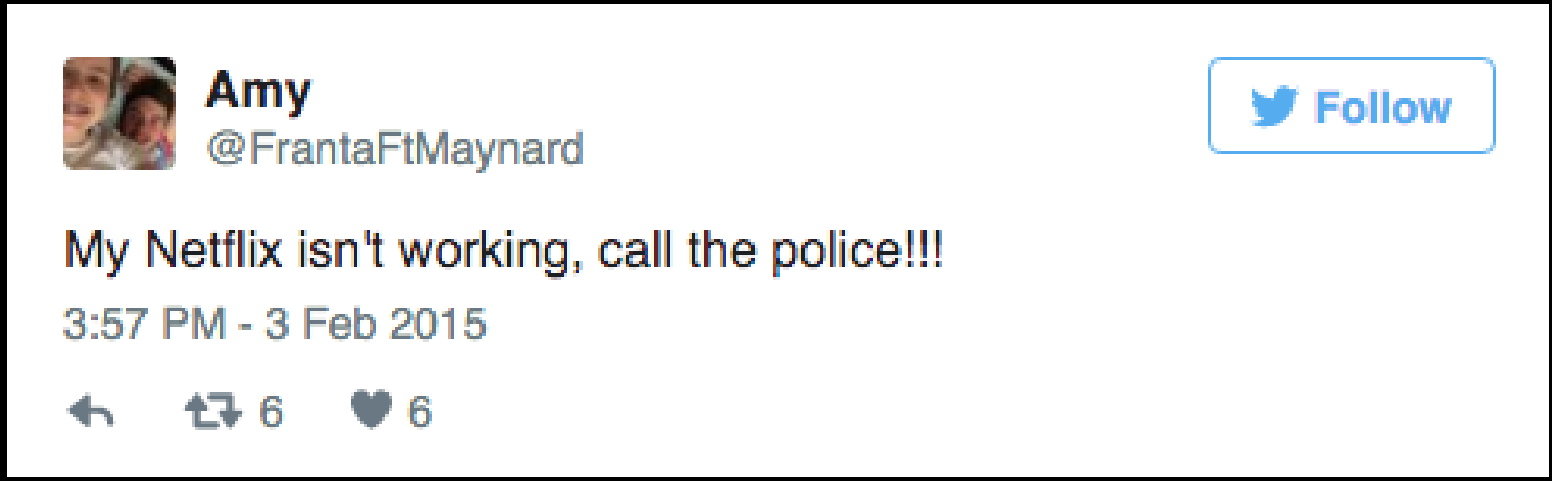
DEC 24, 2012 @ 09:46 PM 105,201 VIEWS

Amazon AWS Takes Down Netflix On Christmas Eve

Disappointment



Outrage


Amy
@FrantaFtMaynard
Follow
My Netflix isn't working, call the police!!!
3:57 PM - 3 Feb 2015
6
6

Withdrawal


The Strumbellas
@thestrumbellas
Follow
I didn't realize my dependency on netflix until it went down and I'm now shaking in the bathtub in a hulk costume quoting 'friends' lines.
4:29 PM - 3 Feb 2015 · Kawartha Lakes, Ontario, Canada
19
47

December 24th, 2012

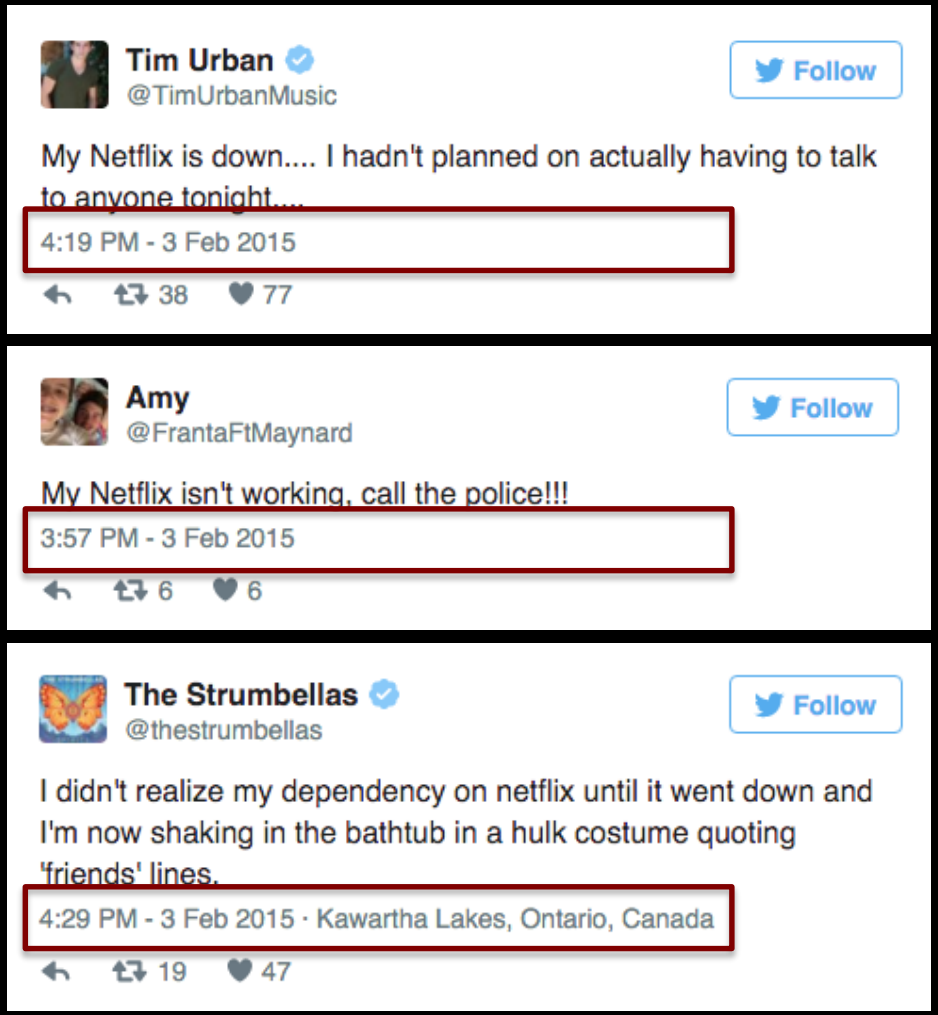
amazon web services

Summary of the December 24, 2012 Amazon ELB Service Event in the US-East Region

We would like to share more details with our customers about the event that occurred with the Amazon Elastic Load Balancing Service ("ELB") earlier this

> The data was deleted by a maintenance process that was inadvertently run against the production ELB state data.

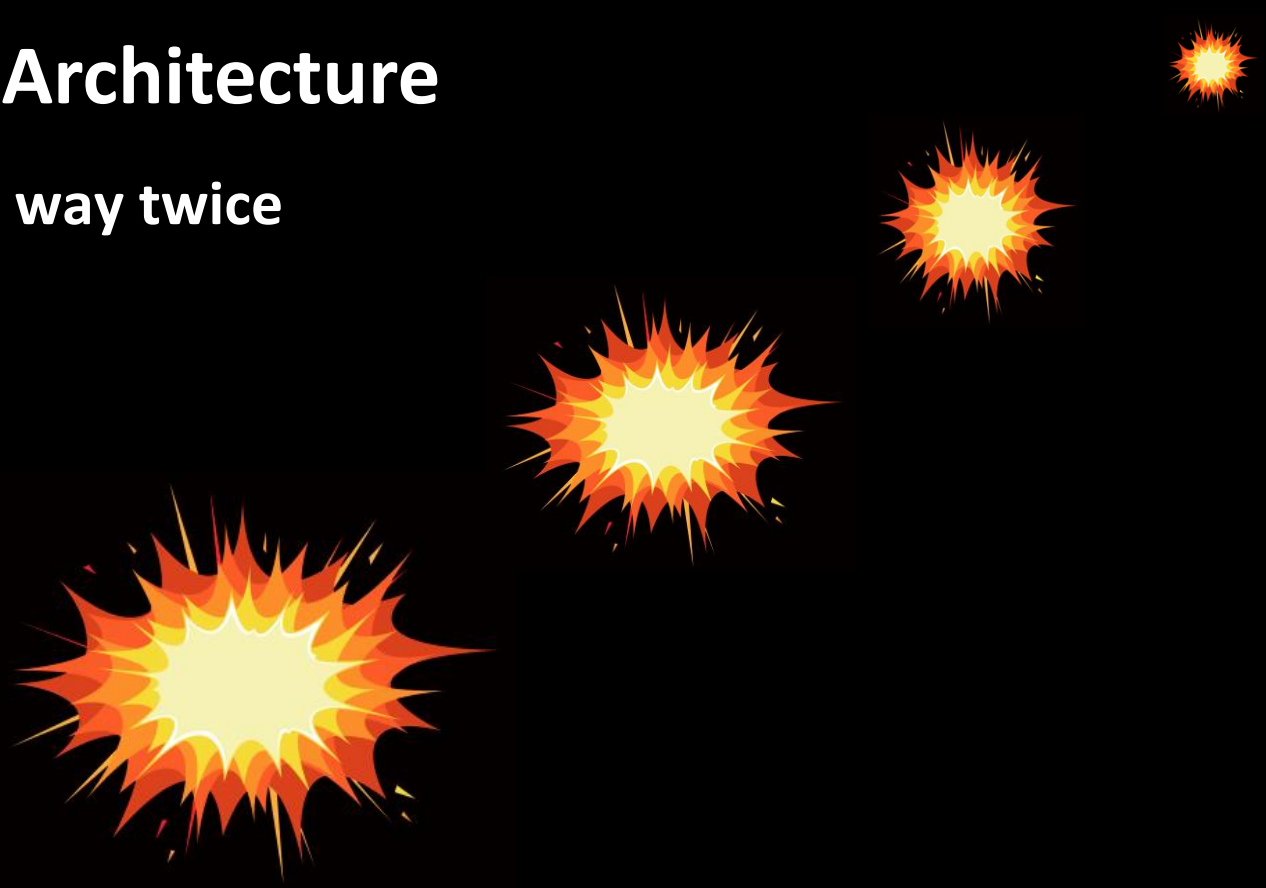
maintained by the ELB control plane to manage the configuration of the ELB load balancers in the region (for example tracking all the backend hosts to which traffic should be routed by each load balancer). The data was deleted by a maintenance process that was inadvertently run against the production ELB state data. This process was run by one of a very small number of developers who have access to this production environment. Unfortunately, the developer did not realize the mistake at the time. After this data was deleted, the ELB control plane began experiencing high latency and error rates for API calls to manage ELB load balancers. In this initial part of the service disruption, there was no impact to the request handling functionality of running ELB load balancers because the missing ELB state data was not integral to the basic operation of running load balancers.


Tim Urban
@TimUrbanMusic
Follow
My Netflix is down.... I hadn't planned on actually having to talk to anyone tonight....
4:19 PM - 3 Feb 2015
38
77
Amy
@FrantaFtMaynard
Follow
My Netflix isn't working, call the police!!!
3:57 PM - 3 Feb 2015
6
6
The Strumbellas
@thestrumbellas
Follow
I didn't realize my dependency on netflix until it went down and I'm now shaking in the bathtub in a hulk costume quoting 'friends' lines.
4:29 PM - 3 Feb 2015 · Kawartha Lakes, Ontario, Canada
19
47

Failure is inevitable

Failure-Driven Architecture

Never fail the same way twice



Failure-Driven Architecture

Never fail the same way twice

#NetflixEverywhere

Our Talk Today

- Introductions
- Failure-Driven Architecture
- Taking It Global

#NetflixEverywhere

Our Talk Today

- **Introductions**
- Failure-Driven Architecture
- Taking It Global

#NetflixEverywhere

Josh Evans – Director of Operations Engineering

1999 – 2009

- Ecommerce (DVD → Streaming)

2009 – 2013

- Playback Services (Activate, Manifests, DRM)

2013 - present

- Operations Engineering
  - CD, RTA, Chaos, Performance

NETFLIX

NETFLIX

Bringing movies & TV shows from all over the world to people all over the world

- Streaming, on demand, subscription
- Global & regional licensing
- Hollywood, independent, international

- Striving for global ubiquity

Device Ubiquity

2007

- Jan – Windows

2008

- May – Roku
- Oct – LG, Samsung Blu-ray
- Oct – Apple Mac
- Nov – XBox 360

2009

- Jun – LG DTV
- Nov –Sony PS3 (disc)
- Nov – Sony Bravia
  - DTV & Blu-ray

Device Ubiquity

2010

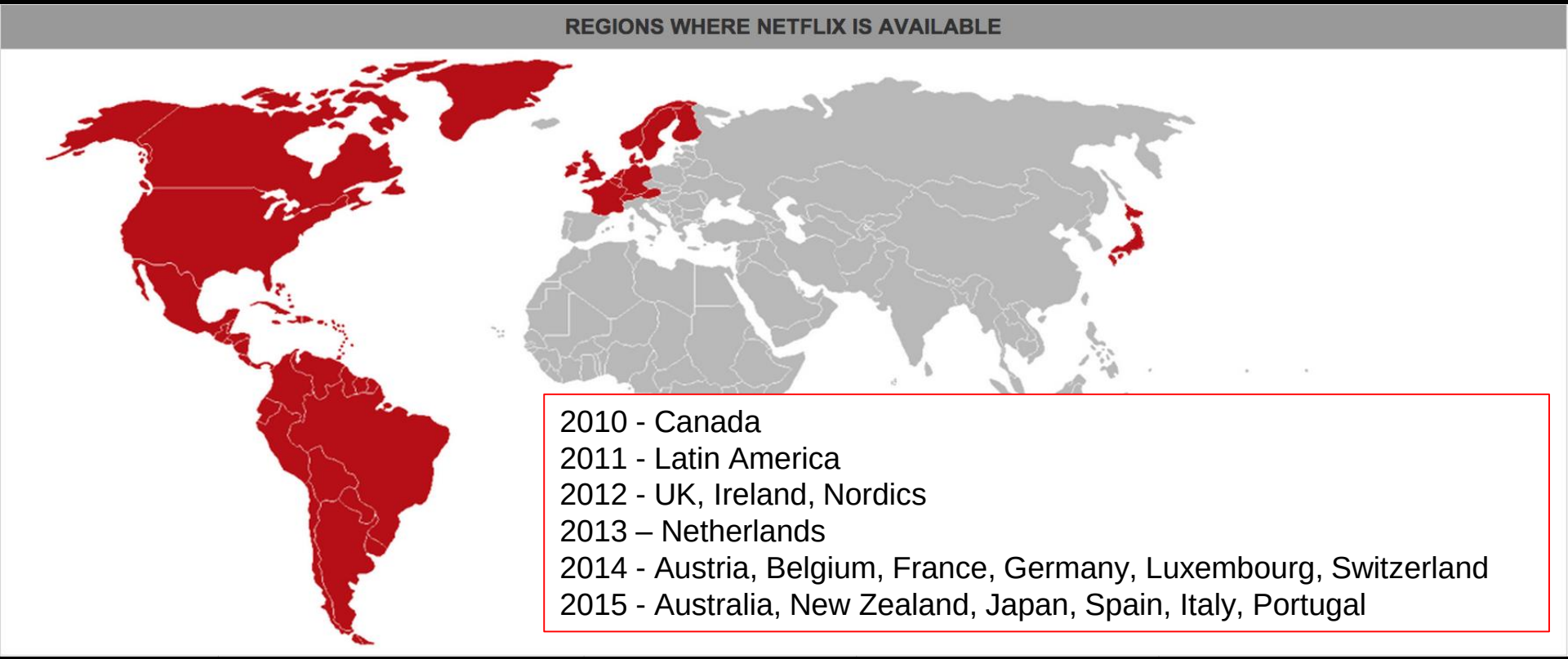
- Mar – Nintendo Wii (disc)
- Apr – Apple iPad
- Aug – Apple iPhone
- Sep – Apple TV
- Oct – Sony PS3 (no disc)
- Oct – Nintendo Wii (no disc)
- Nov – Windows Phone 7

2011

- May – Android
- Nov – First e-readers
  - Kindle Fire, Nook


ENTER
EXIT
NETFLIX
FAVORITE CH
FAV APP
1
2
3
A
B

Geographic Ubiquity



2010 - Canada
2011 - Latin America
2012 - UK, Ireland, Nordics
2013 – Netherlands
2014 - Austria, Belgium, France, Germany, Luxembourg, Switzerland
2015 - Australia, New Zealand, Japan, Spain, Italy, Portugal

Language Ubiquity - Subs, Dubs, UI

- English
- Spanish (Latin American)
- Portuguese (Brazilian)
- Dutch
- French
- German
- Japanese
- Spanish (Castilian)
- Italian
- Portuguese (European)

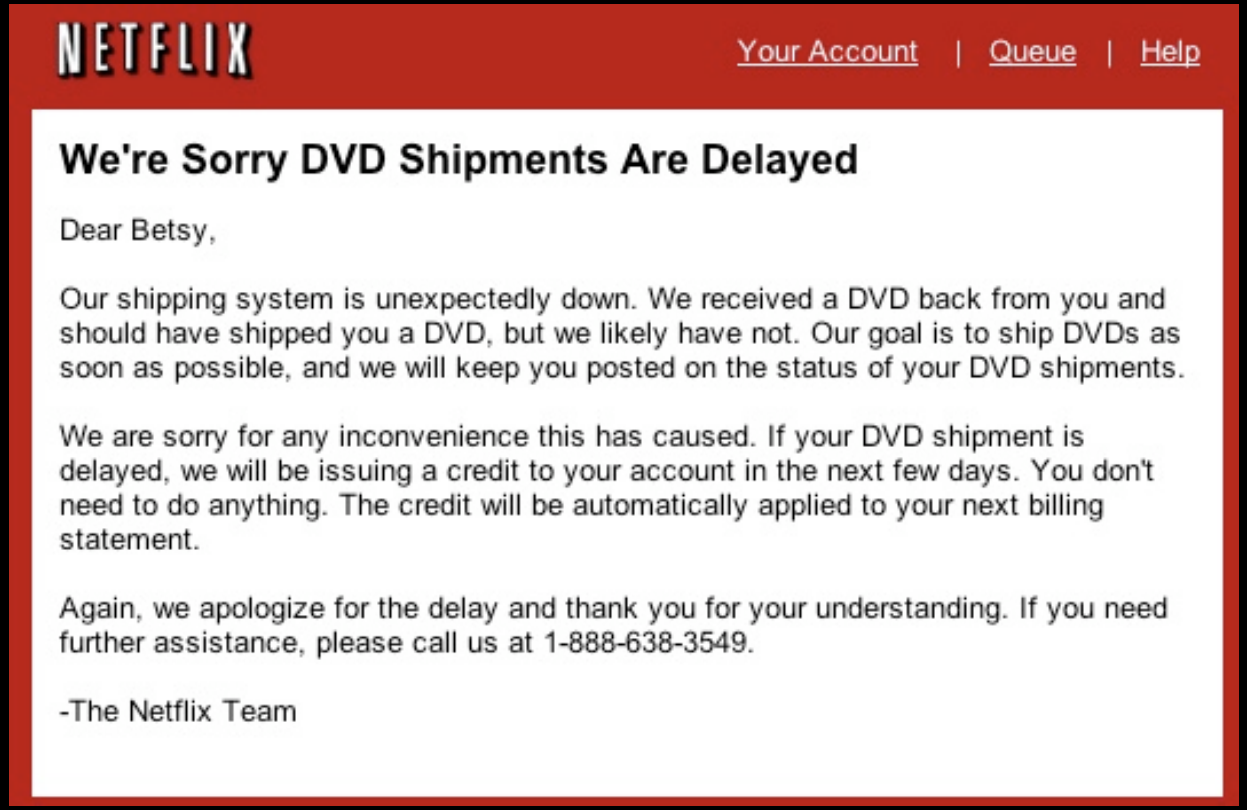
75,000,000

Our Talk Today

- Introductions
- **Failure-Driven Architecture**
- Taking It Global

#NetflixEverywhere

August 2008



NETFLIX

Your Account | Queue | Help

We're Sorry DVD Shipments Are Delayed

Dear Betsy,

Our shipping system is unexpectedly down. We received a DVD back from you and should have shipped you a DVD, but we likely have not. Our goal is to ship DVDs as soon as possible, and we will keep you posted on the status of your DVD shipments.

We are sorry for any inconvenience this has caused. If your DVD shipment is delayed, we will be issuing a credit to your account in the next few days. You don't need to do anything. The credit will be automatically applied to your next billing statement.

Again, we apologize for the delay and thank you for your understanding. If you need further assistance, please call us at 1-888-638-3549.

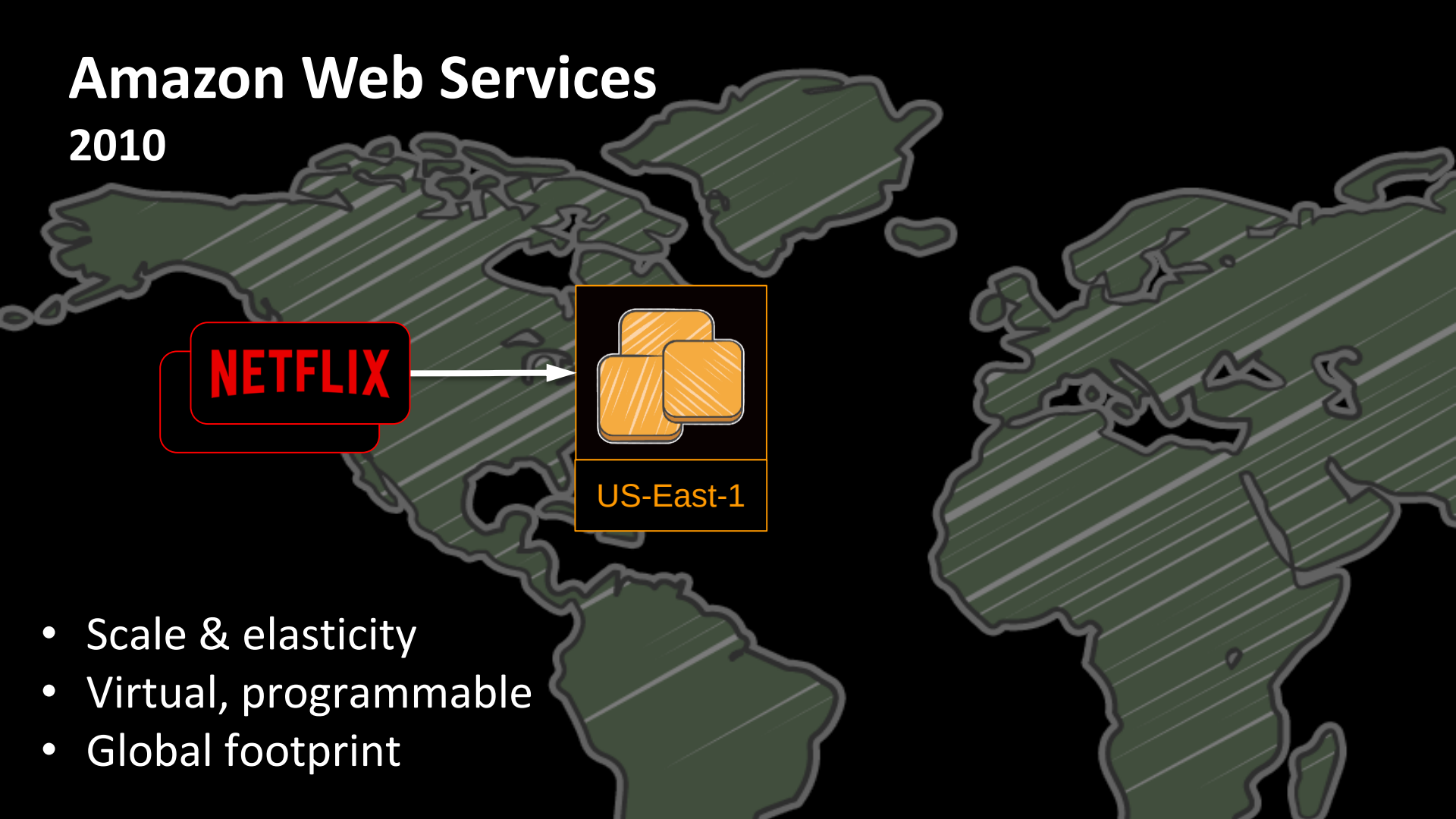
-The Netflix Team

DC2
2009

- No automation, virtualization, standardization
- Manual, error prone, slow
- Big iron & monoliths

Undifferentiated Heavy Lifting

Amazon Web Services
2010



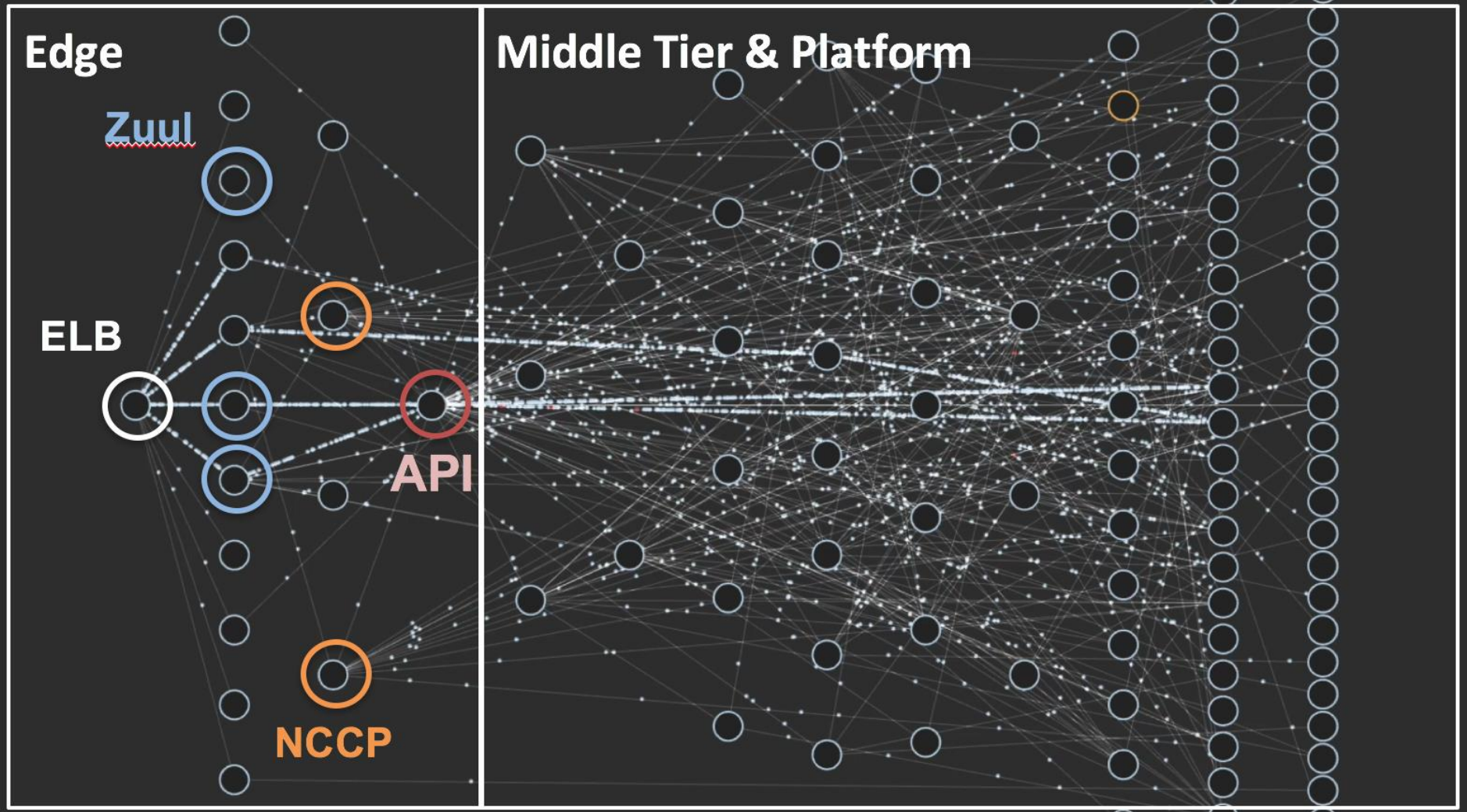
- Scale & elasticity
- Virtual, programmable
- Global footprint

Architectural Pillars

- Micro-services
- Database
- Cache
- Traffic

Architectural Pillars

- Micro-services
- Database
- Cache
- Traffic


Edge
Middle Tier & Platform
Zuul
ELB
API
NCCP

Micro-service Failure



FIT

Fault-Injection
Test Framework

Micro-service Failure



FIT
Fault-Injection
Test Framework

Architectural Pillars

- Micro-services
- **Database**
- Cache
- Traffic

SimpleDB

NoSQL but…

- Not web scale
- Throttling

Modest scale

- 100s of play starts / second
- 10,000s of requests / second
- 10s of billions of records

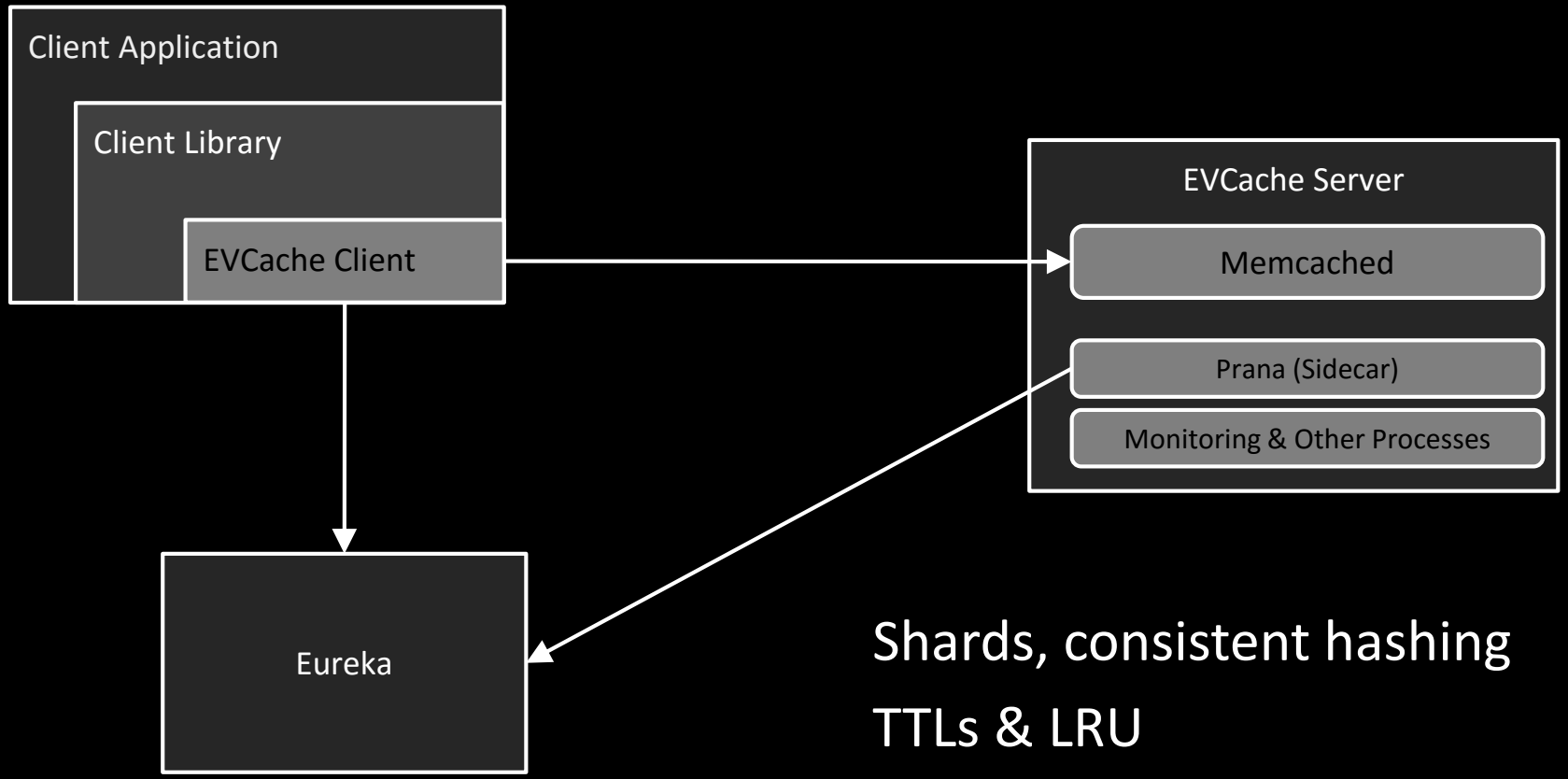
Architectural Pillars

- Micro-services
- Database
- **Cache**
- Traffic

Ephemeral Volatile memCache (EVCache)

Clustered memcached optimized for Netflix use cases

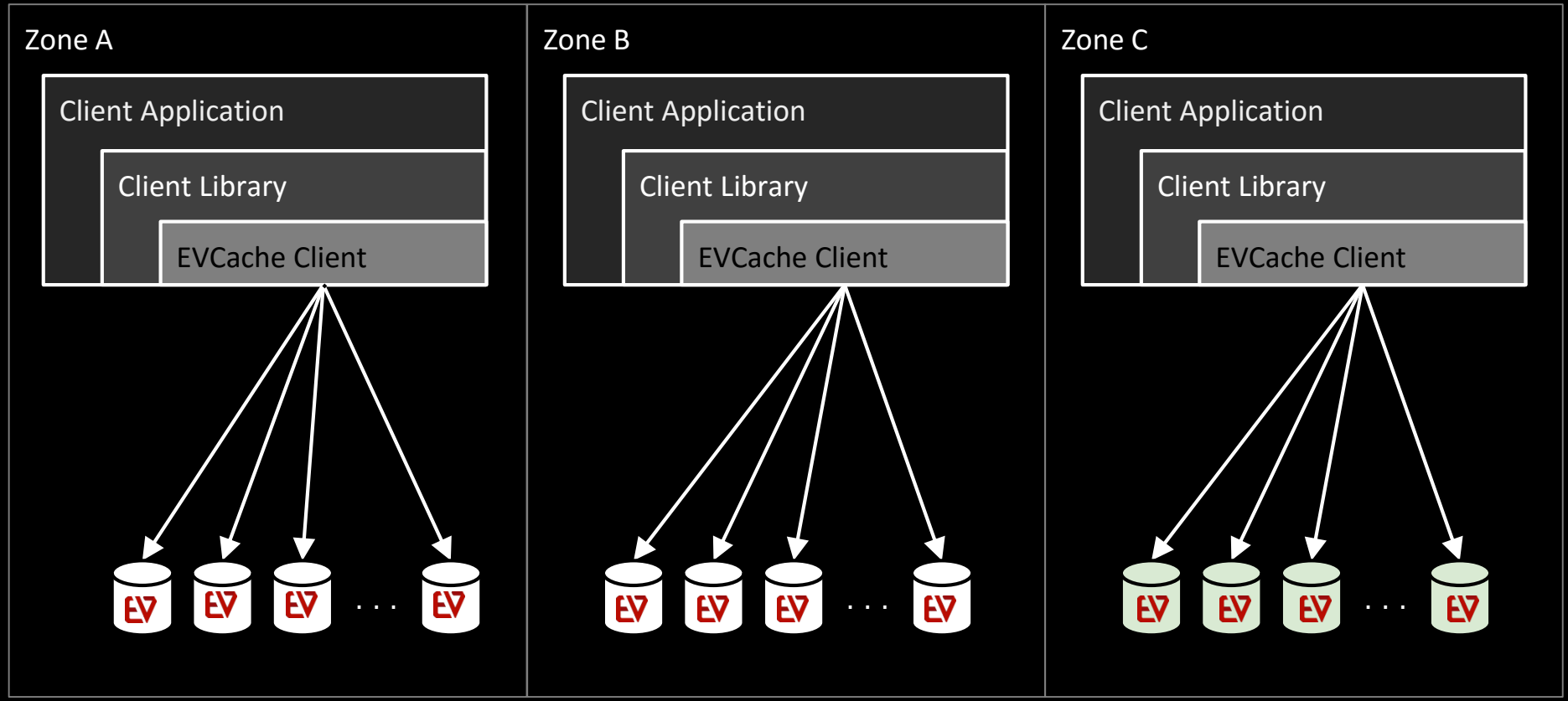
EVCache Architecture



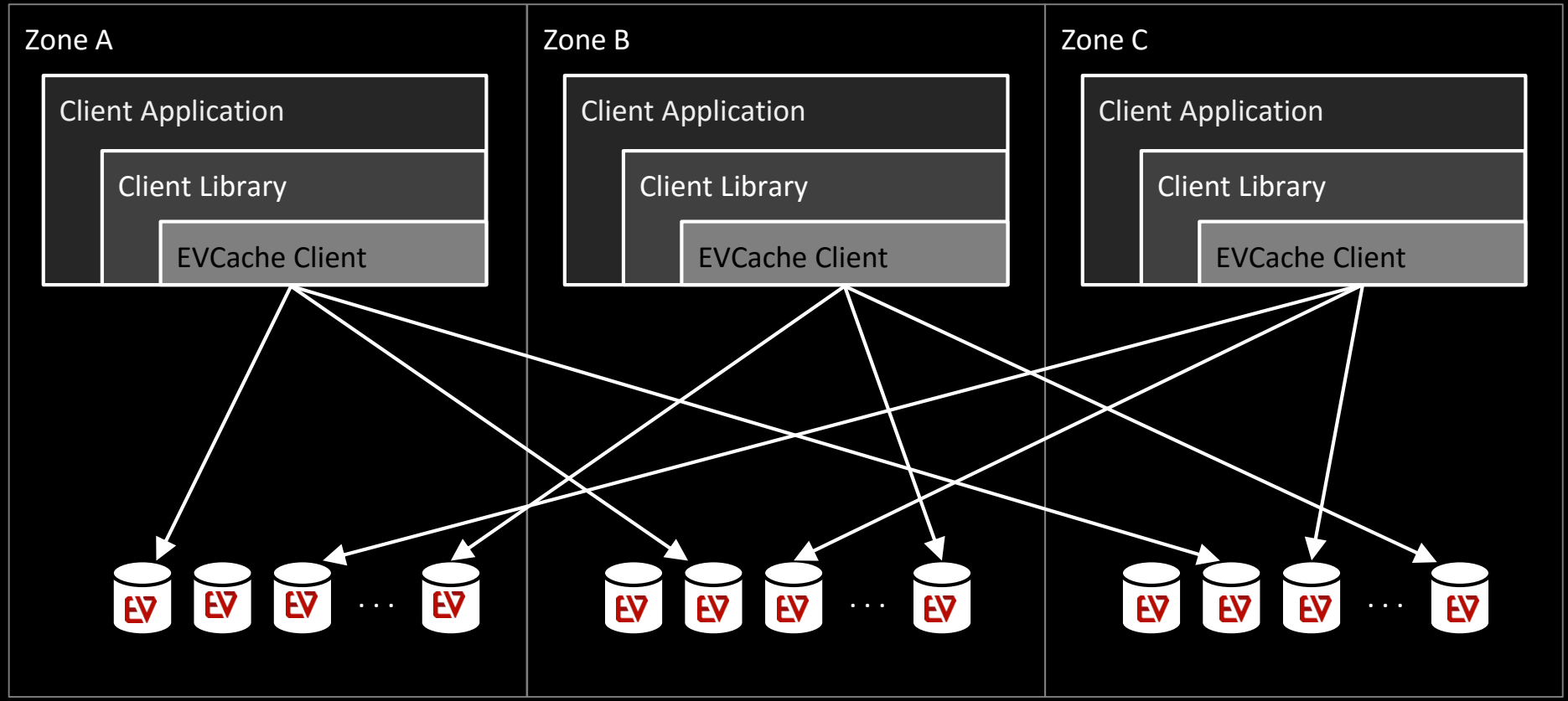
Shards, consistent hashing

TTLs & LRU

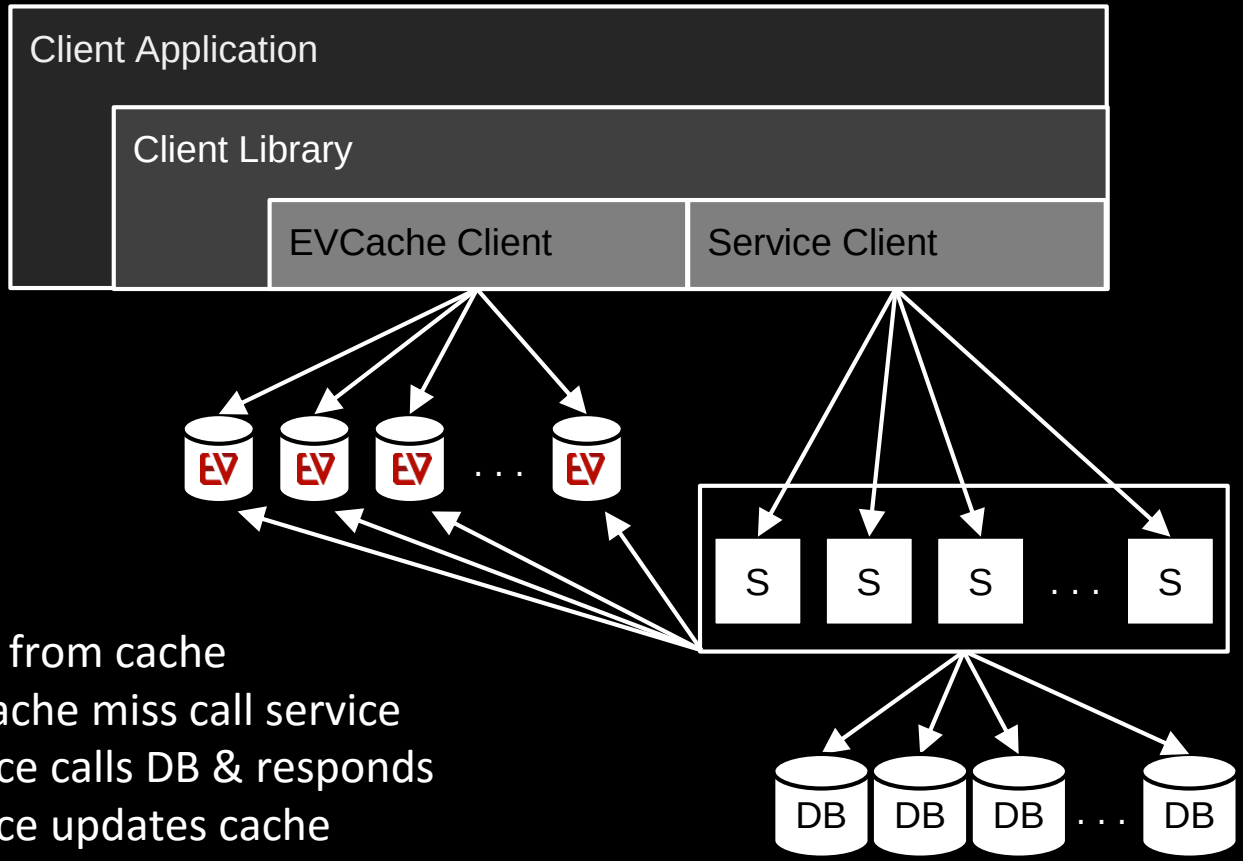
Reads


Zone A
Client Application
Client Library
EVCache Client
Zone B
Client Application
Client Library
EVCache Client
Zone C
Client Application
Client Library
EVCache Client

Writes


Zone A
Client Application
Client Library
EVCache Client
Zone B
Client Application
Client Library
EVCache Client
Zone C
Client Application
Client Library
EVCache Client

Fronting Micro-services



1. Read from cache
2. On cache miss call service
3. Service calls DB & responds
4. Service updates cache

Linear Scaling

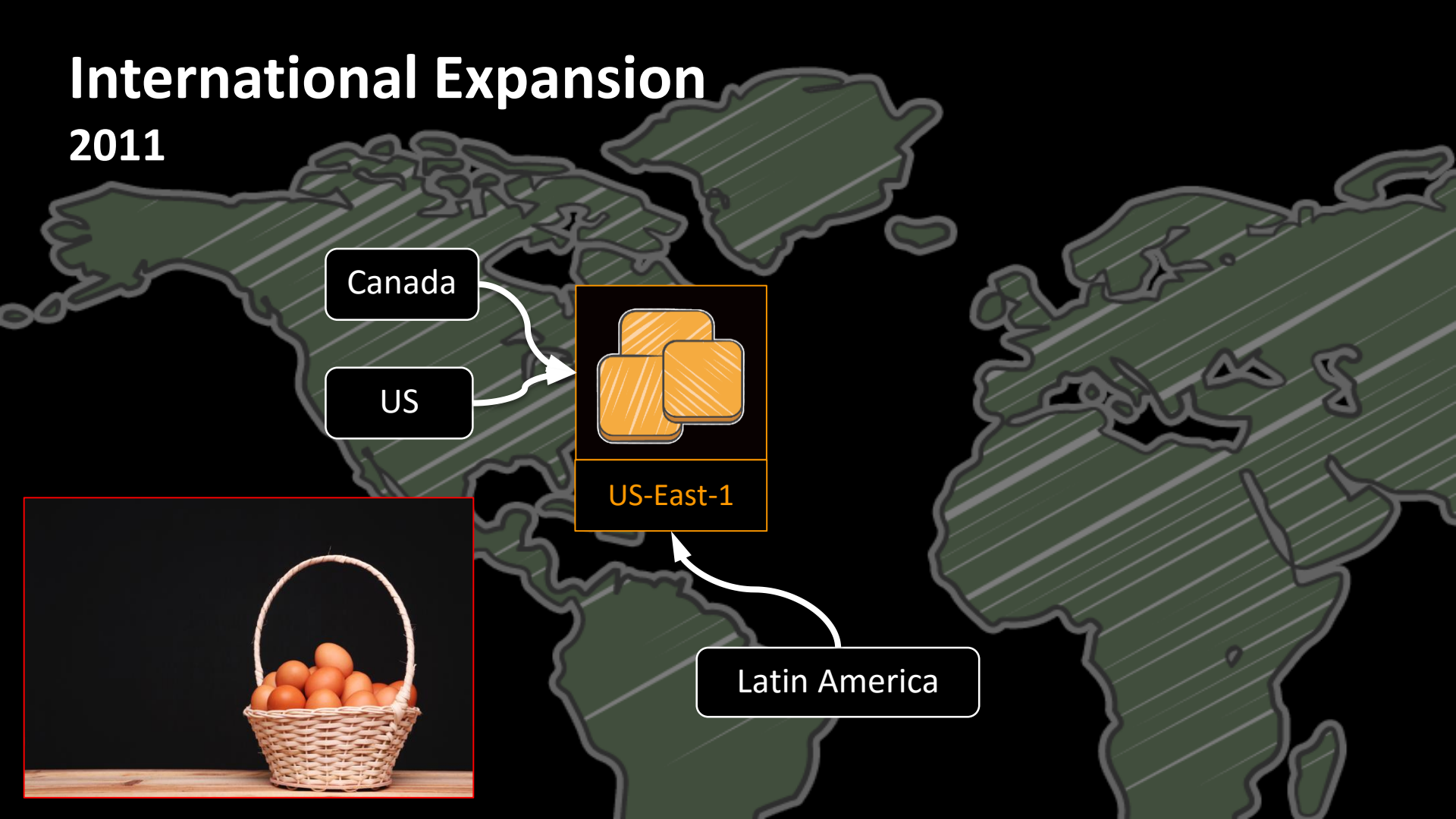
- 30 million requests/sec
- 2 trillion requests per day globally

- Hundreds of billions of objects
- Tens of thousands of memcached instances

- Milliseconds of latency per request

International Expansion

2011



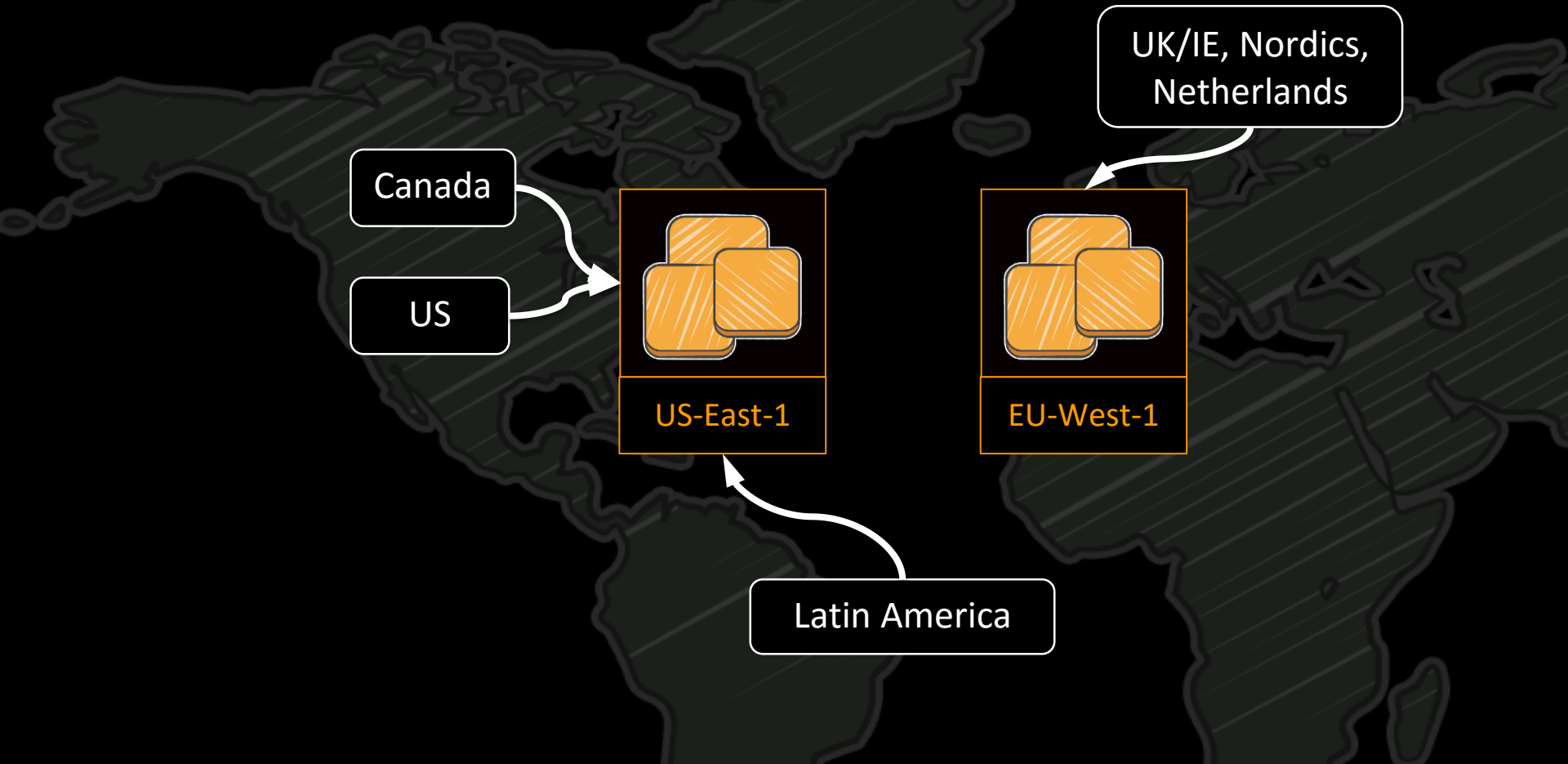
Cloud Islands

2012

Architectural Pillars

- Micro-services
- Database
- Cache
- Traffic

DNS Geo Mapping



Architectural Pillars

- Micro-services
- **Database!**
- Caching
- Traffic

Scalable, Durable, Global

Why Cassandra?

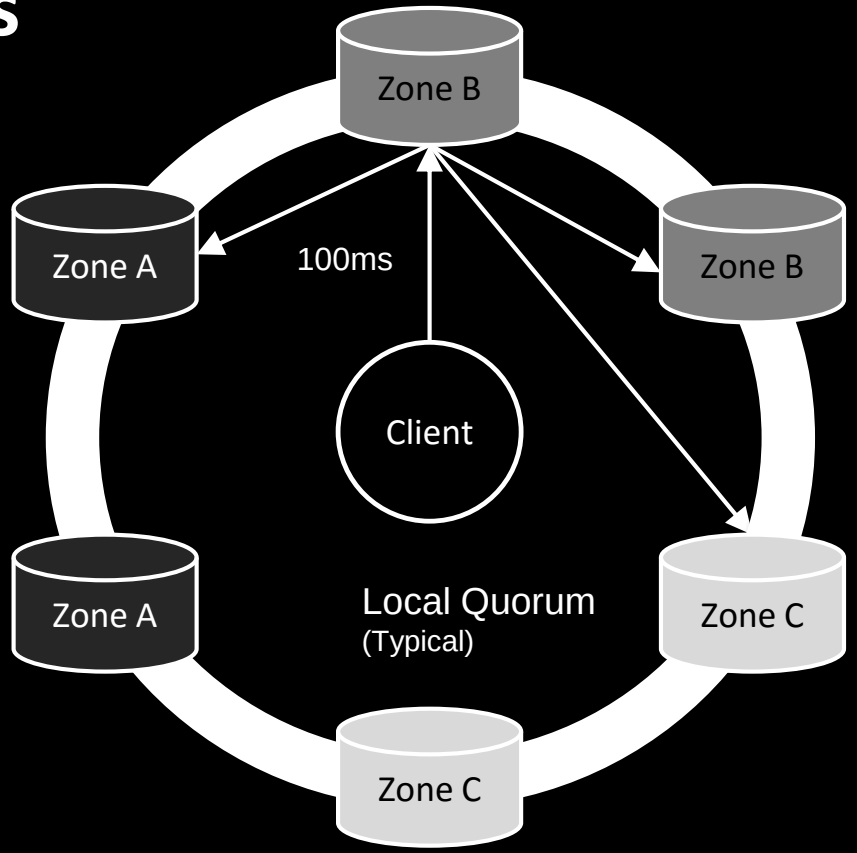
- NoSQL at scale
- Open source
- Multi-region
- Multi-directional

- CAP Choices
  - Availability
  - Partition tolerance
  - Eventual consistency*

Single Region, Multiple AZs

1. Client writes to any node
2. Coordinator replicates to nodes
3. Nodes ack to coordinator
4. Coordinator acks to client
5. Write to commit log

- Hinted handoff to offline nodes



Not quite fast enough

December 24th, 2012

amazon web services

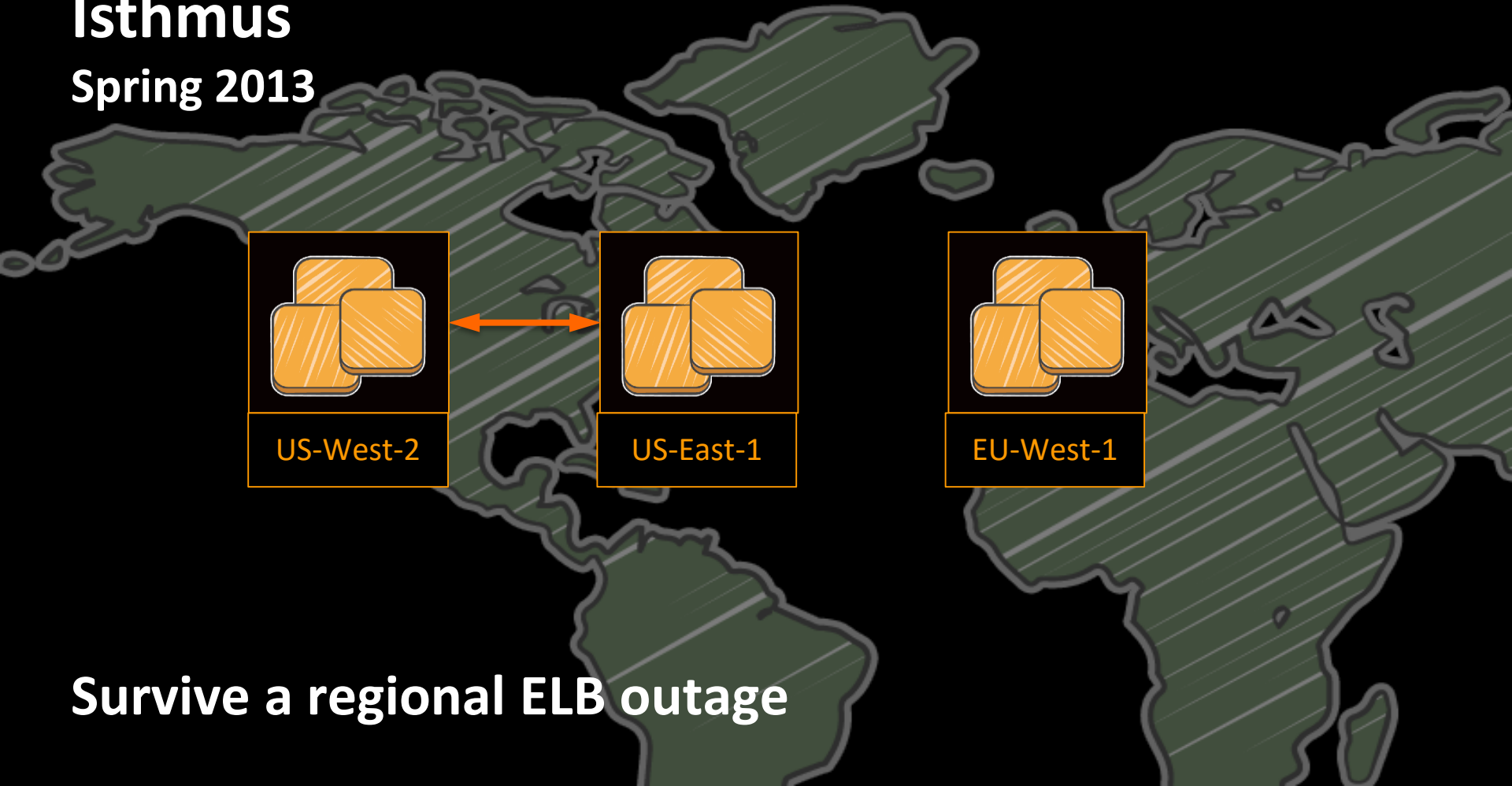
Summary of the December 24, 2012 Amazon ELB Service Event in the US-East Region

We would like to share more details with our customers about the event that occurred with the Amazon Elastic Load Balancing Service ("ELB") earlier this week in the US-East Region. While the service disruption only affected applications using the ELB service (and only a fraction of the ELB load balancers were affected), the impacted load balancers saw significant impact for a prolonged period of time.

The service disruption began at 12:24 PM PST on December 24th when a portion of the ELB state data was logically deleted. This data is used and maintained by the ELB control plane to manage the configuration of the ELB load balancers in the region (for example tracking all the backend hosts to which traffic should be routed by each load balancer). The data was deleted by a maintenance process that was inadvertently run against the production ELB state data. This process was run by one of a very small number of developers who have access to this production environment. Unfortunately, the developer did not realize the mistake at the time. After this data was deleted, the ELB control plane began experiencing high latency and error rates for API calls to manage ELB load balancers. In this initial part of the service disruption, there was no impact to the request handling functionality of running ELB load balancers because the missing ELB state data was not integral to the basic operation of running load balancers.

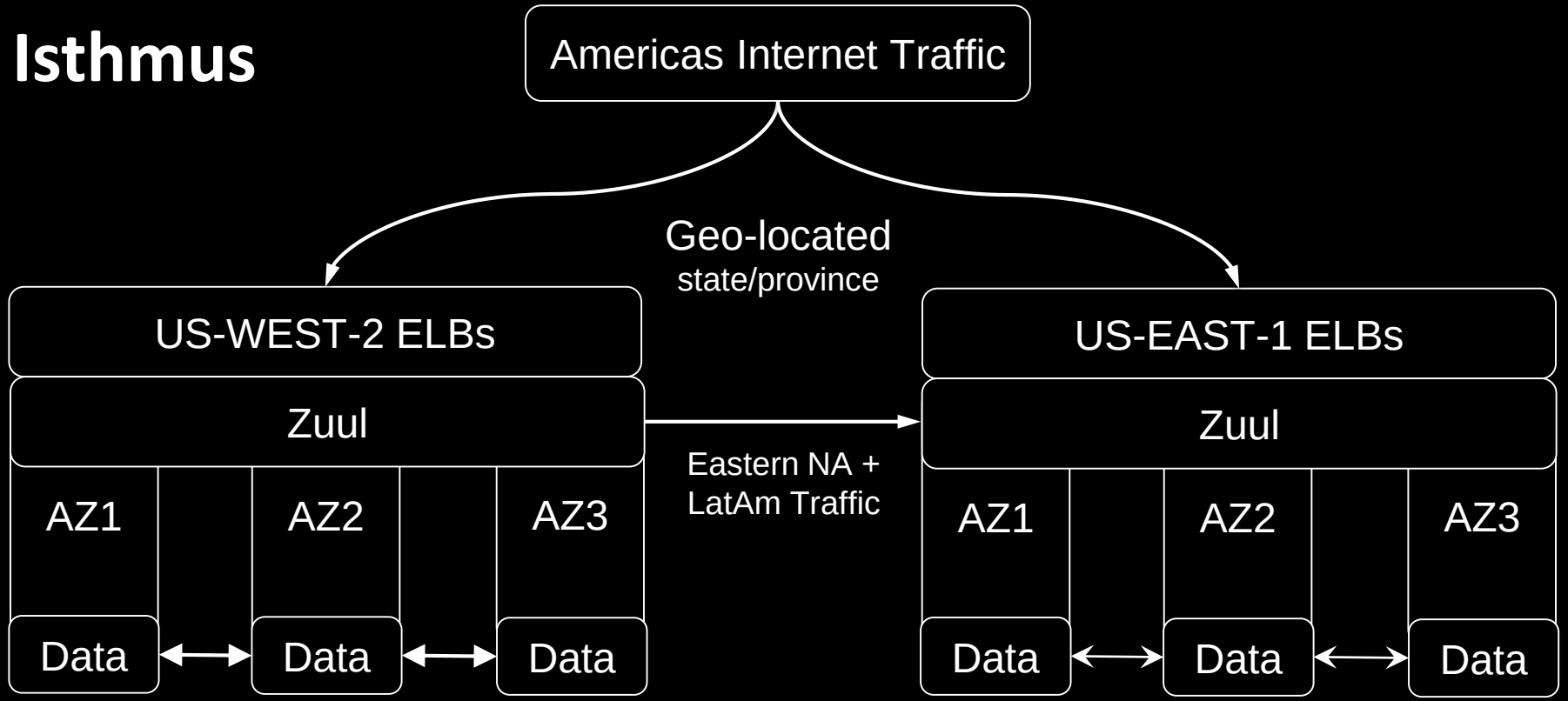
Isthmus

Spring 2013



Survive a regional ELB outage

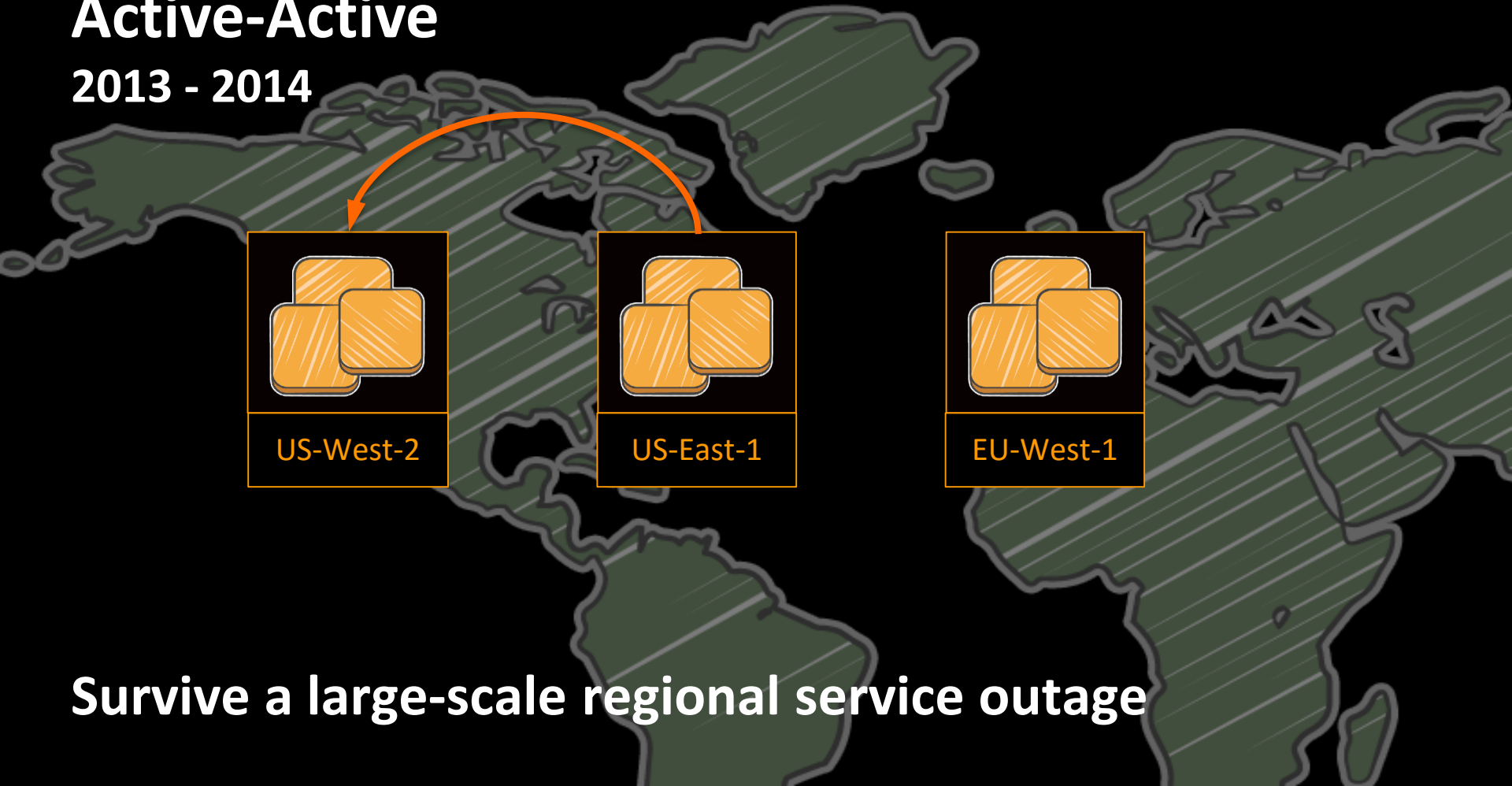
Isthmus



- Zuul routes locally or remotely
- Eureka - multi-region aware

Active-Active

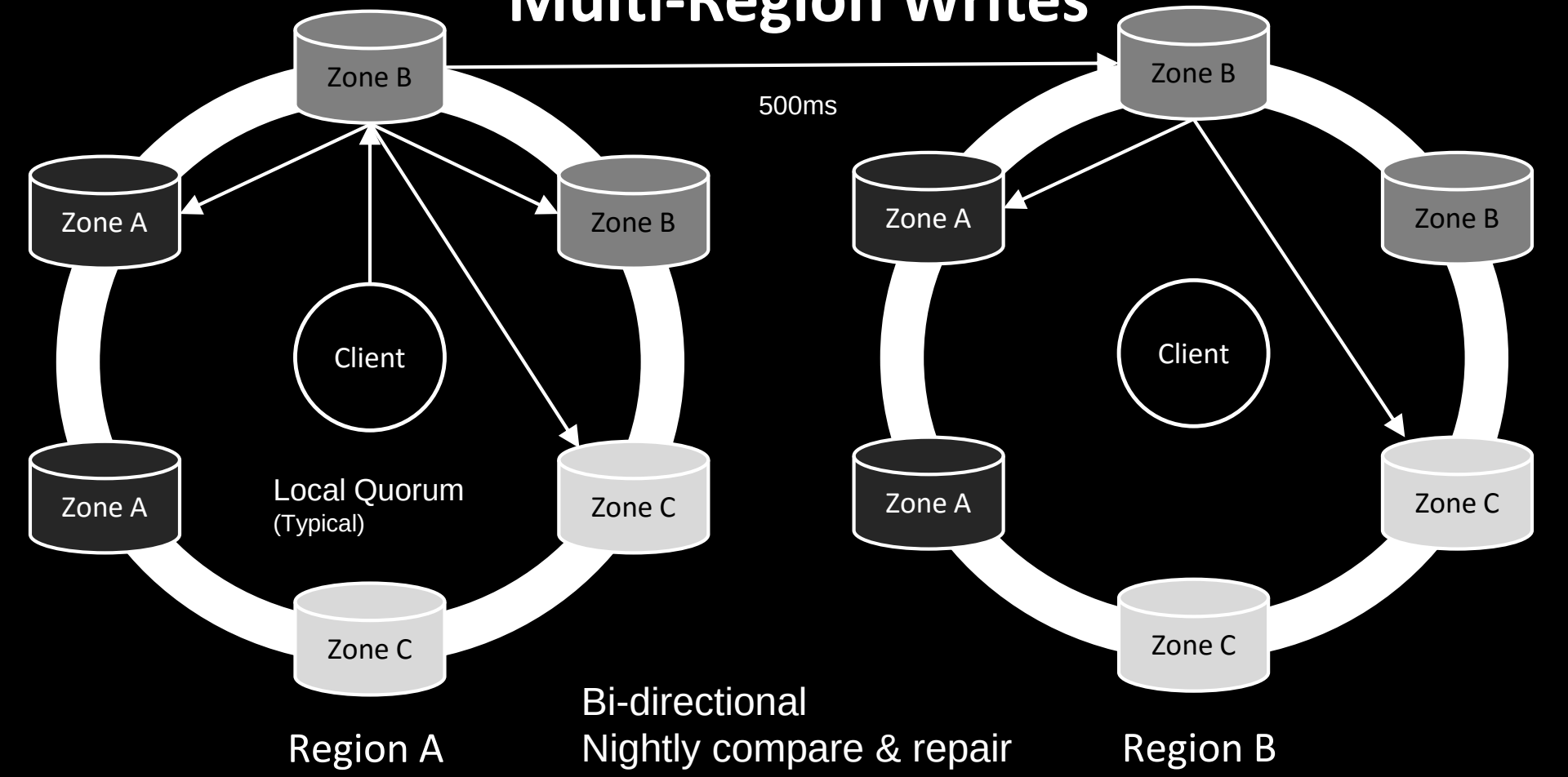
2013 - 2014



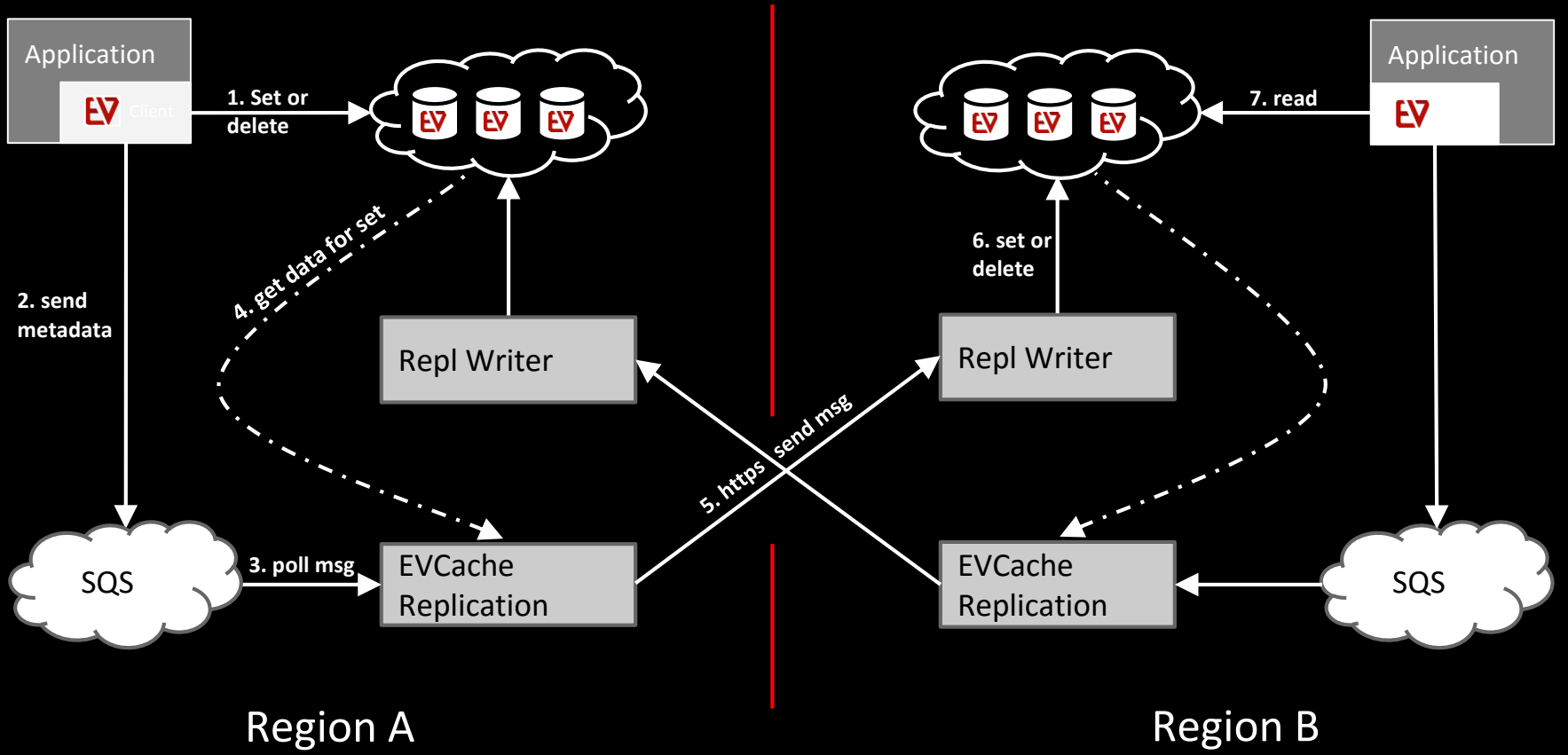
Survive a large-scale regional service outage

Active-Active Data Replication

Multi-Region Writes

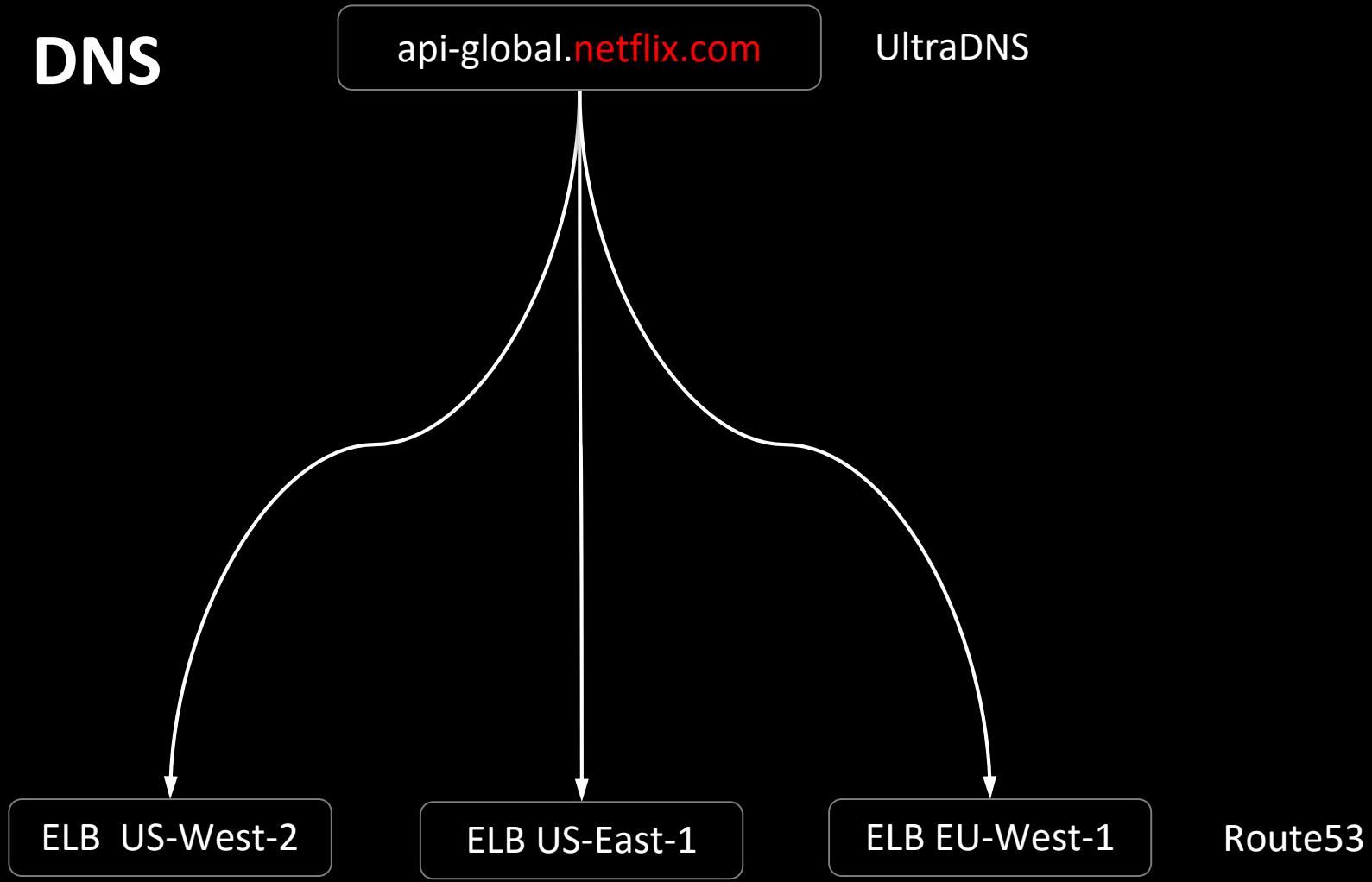


EVCache Cross-Region Replication

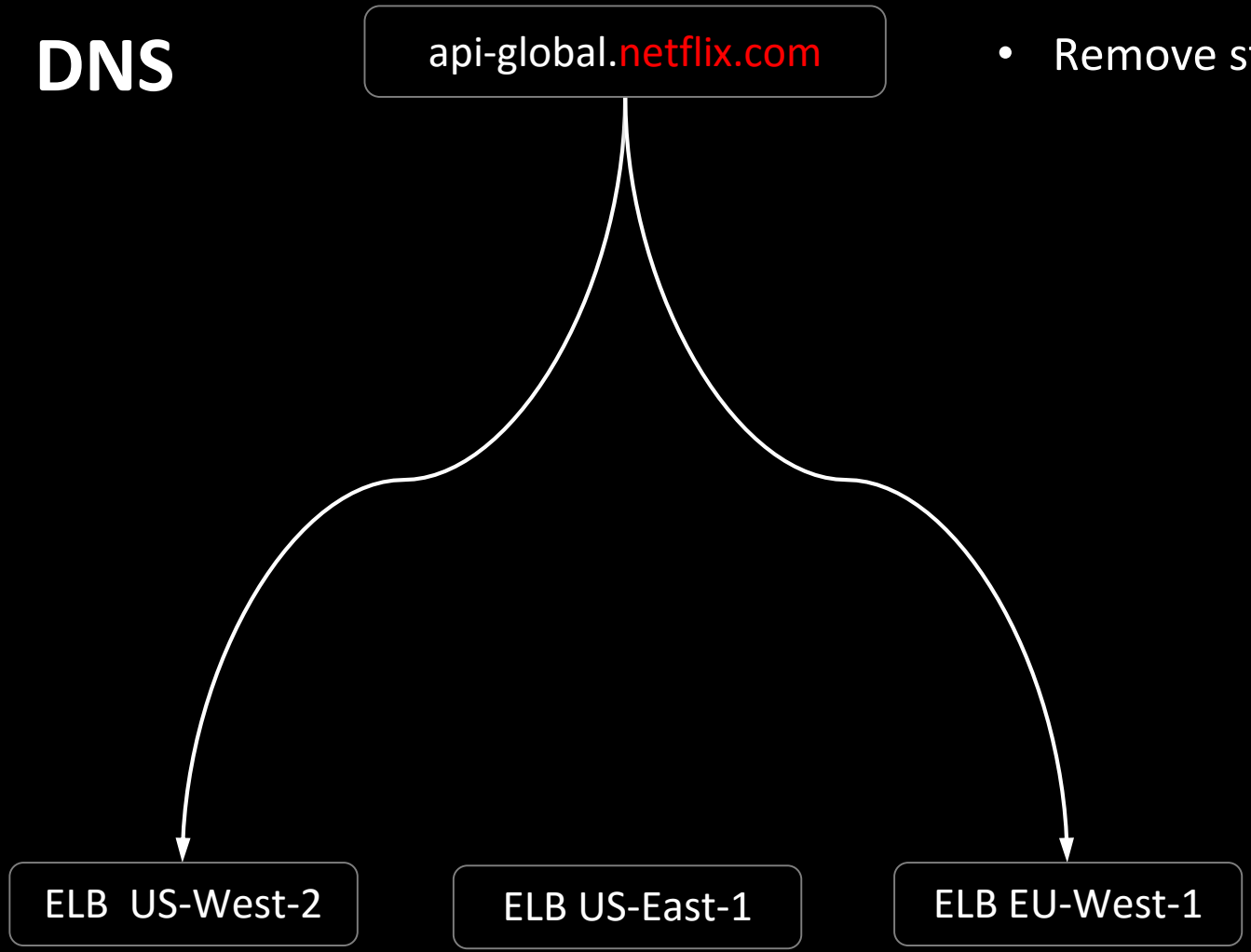


Active-Active Traffic Management

DNS

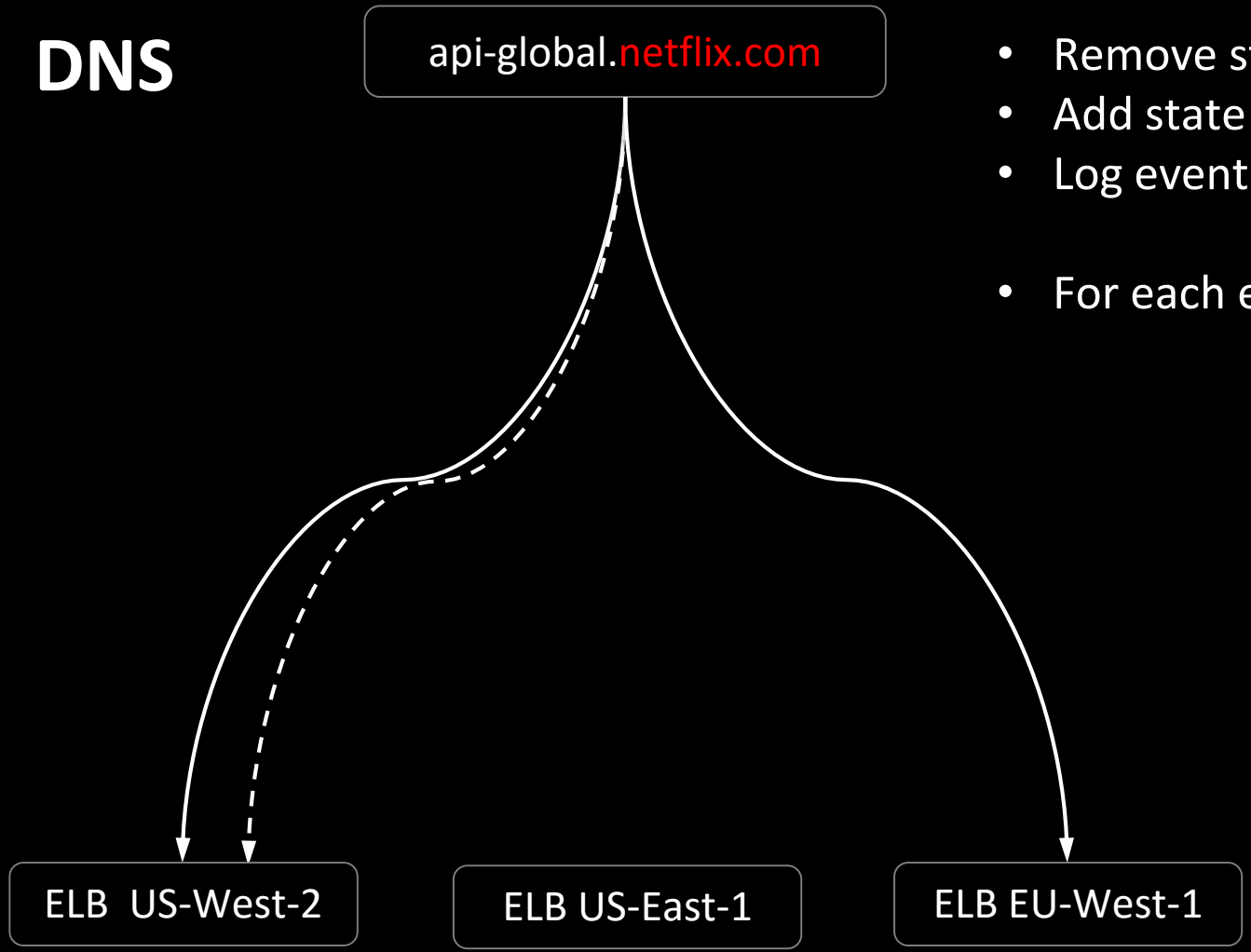

api-global.netflix.com
UltraDNS
ELB US-West-2
ELB US-East-1
ELB EU-West-1
Route53

DNS



- Remove state from geo bucket

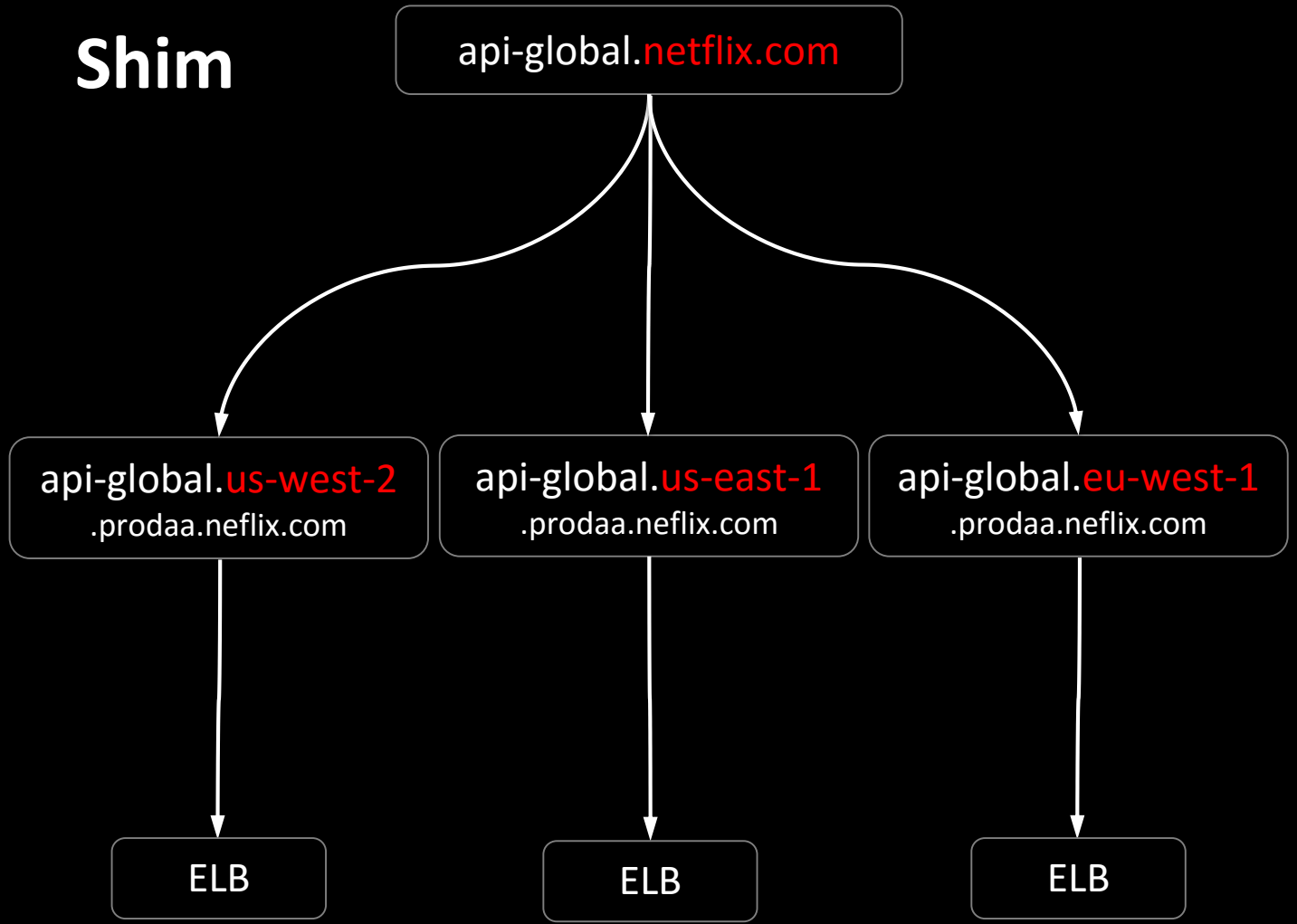
DNS



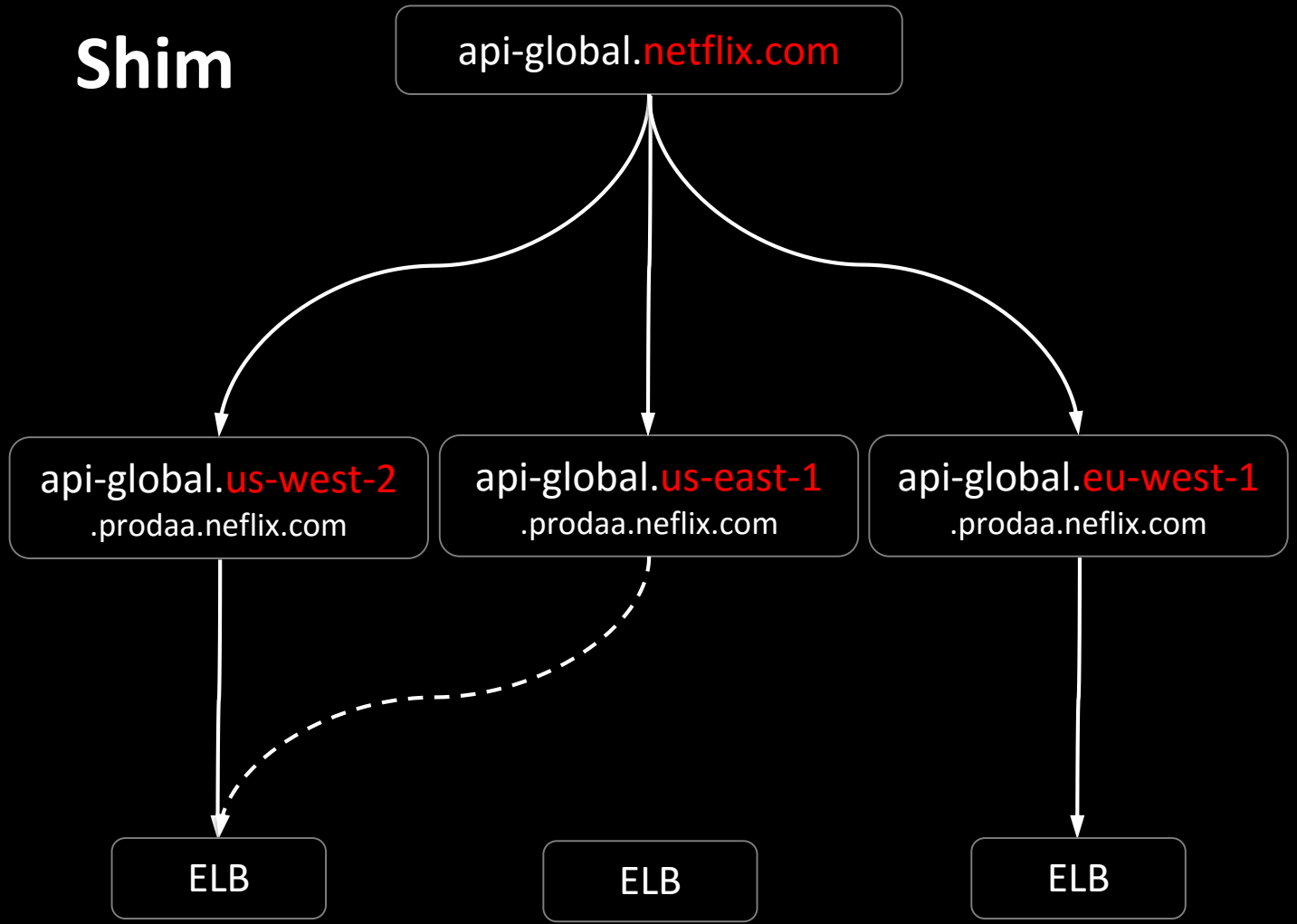
- Remove state from geo bucket
- Add state to geo bucket
- Log event

- For each end point

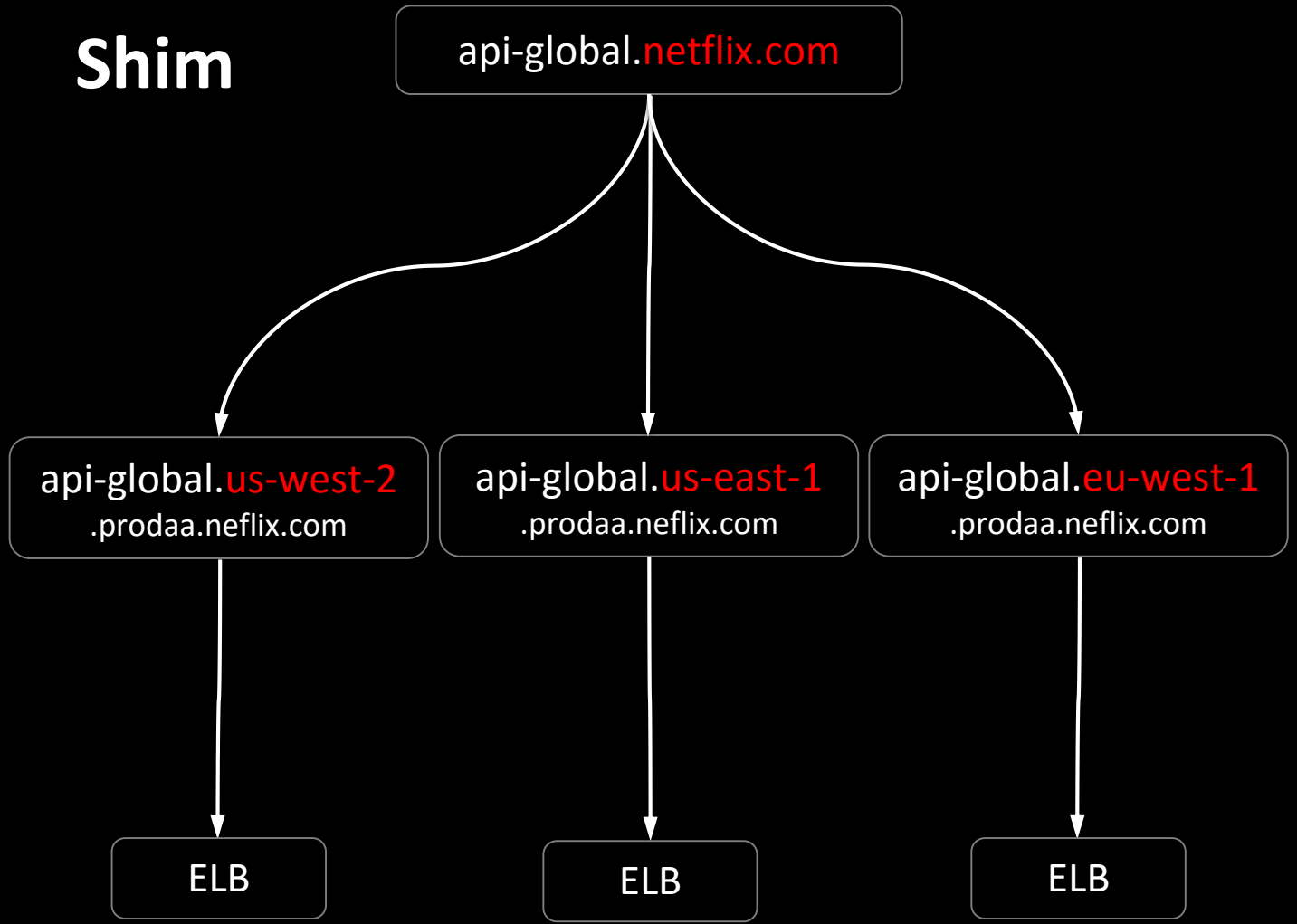
Shim


api-global.netflix.com
api-global.us-west-2
.prodaa.neflix.com
api-global.us-east-1
.prodaa.neflix.com
api-global.eu-west-1
.prodaa.neflix.com
ELB
ELB
ELB

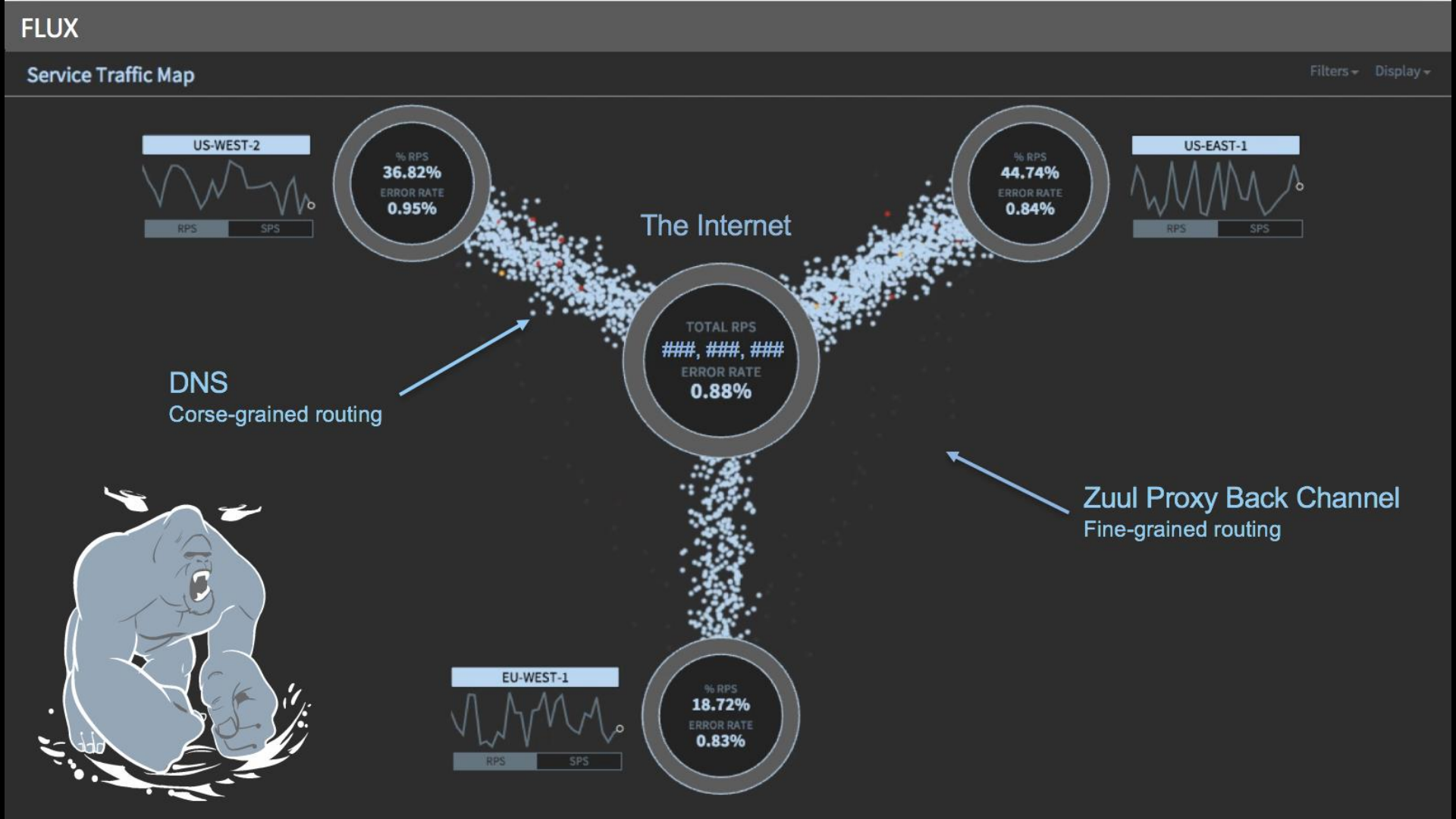
Shim


api-global.netflix.com
api-global.us-west-2
.prodaa.neflix.com
api-global.us-east-1
.prodaa.neflix.com
api-global.eu-west-1
.prodaa.neflix.com
ELB
ELB
ELB

Shim


api-global.netflix.com
api-global.us-west-2
.prodaa.neflix.com
api-global.us-east-1
.prodaa.neflix.com
api-global.eu-west-1
.prodaa.neflix.com
ELB
ELB
ELB

Active-Active Failover


FLUX
Service Traffic Map
Filters
Display
US-WEST-2
RPS
SPS
% RPS
36.82%
ERROR RATE
0.95%
The Internet
% RPS
44.74%
ERROR RATE
0.84%
US-EAST-1
RPS
SPS
TOTAL RPS
###, ###, ###
ERROR RATE
0.88%
DNS
Corse-grained routing
Zuul Proxy Back Channel
Fine-grained routing
EU-WEST-1
RPS
SPS
% RPS
18.72%
ERROR RATE
0.83%

Our Talk Today

- Introductions
- Failure-Driven Architecture
- **Taking It Global**

#NetflixEverywhere

January 6th, 2016

Geographic Ubiquity

REGIONS WHERE NETFLIX IS AVAILABLE

Language Ubiquity

Before Global

- English
- Spanish (Latin American)
- Portuguese (Brazilian)
- Dutch
- French
- German
- Japanese
- Spanish (Castilian)
- Italian
- Portuguese (European)

Global

- Chinese
- Korean
- Arabic

ステップ1/4: アカウントを作成

盛りだくさんの映画やドラマを視聴でき、**最初の1ヵ月は無料**です。

NETFLIX

名を入力してください

たなか　田中　タナカ　田仲

あ か さ た な は ま や ら わ
い き し ち に ひ み ゆ り を
う く す つ ぬ ふ む よ る ん
え け せ て ね へ め ゛ れ ー
お こ そ と の ほ も ゜ ろ 小

abc123　␣　⌫

戻る　次へ

田中

姓

メールアドレス

パスワード (4〜50文字)

プライバシーポリシー

□Netflixサービスについてご連絡を差し上げる場合があります。

ご質問ですか? 1-866-579-7172□24時間対応


ءادوسلا ةمئاقلا
2013 ةويزفلتلا-14 1 مسوملا ULTRA HD 4K 5.1
تاقيقحتلا بتكمل هسفن لوحت هنأ .محري ال امامت ،باذج ،يكذ
.طورشلا ضعب عم .يلاردفلا
نلعأ يس يب امارد "ءادوسلا ةمئاقلا" حبصأ ،لوألا همسوم يف
.فينصت
UltraHD ك 4
Breaking Bad
TOM CRUISE
Jerry Maguire
THE SMURFS 2
Will Smith
HITCH
THE BLACKLIST
NETFLIX
سكايلفتن يف ةيبعش
NETFLIX
ARCHER
FX
the office
Parks and Recreation
BRAIN GAMES

Content Ubiquity

March 18th, 2016

Daredevil Season 2

All episodes, all devices, all countries

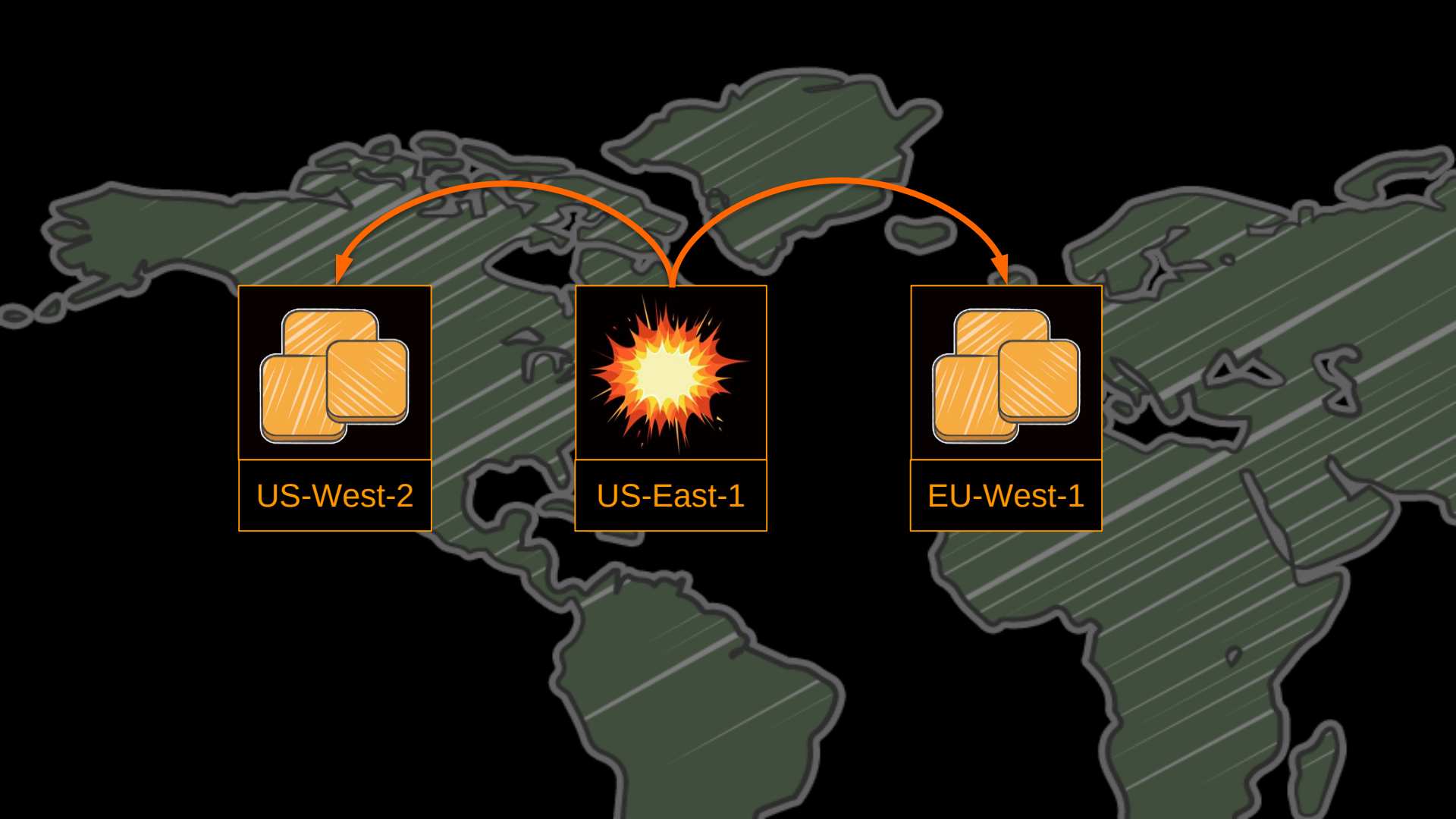
Simultaneously

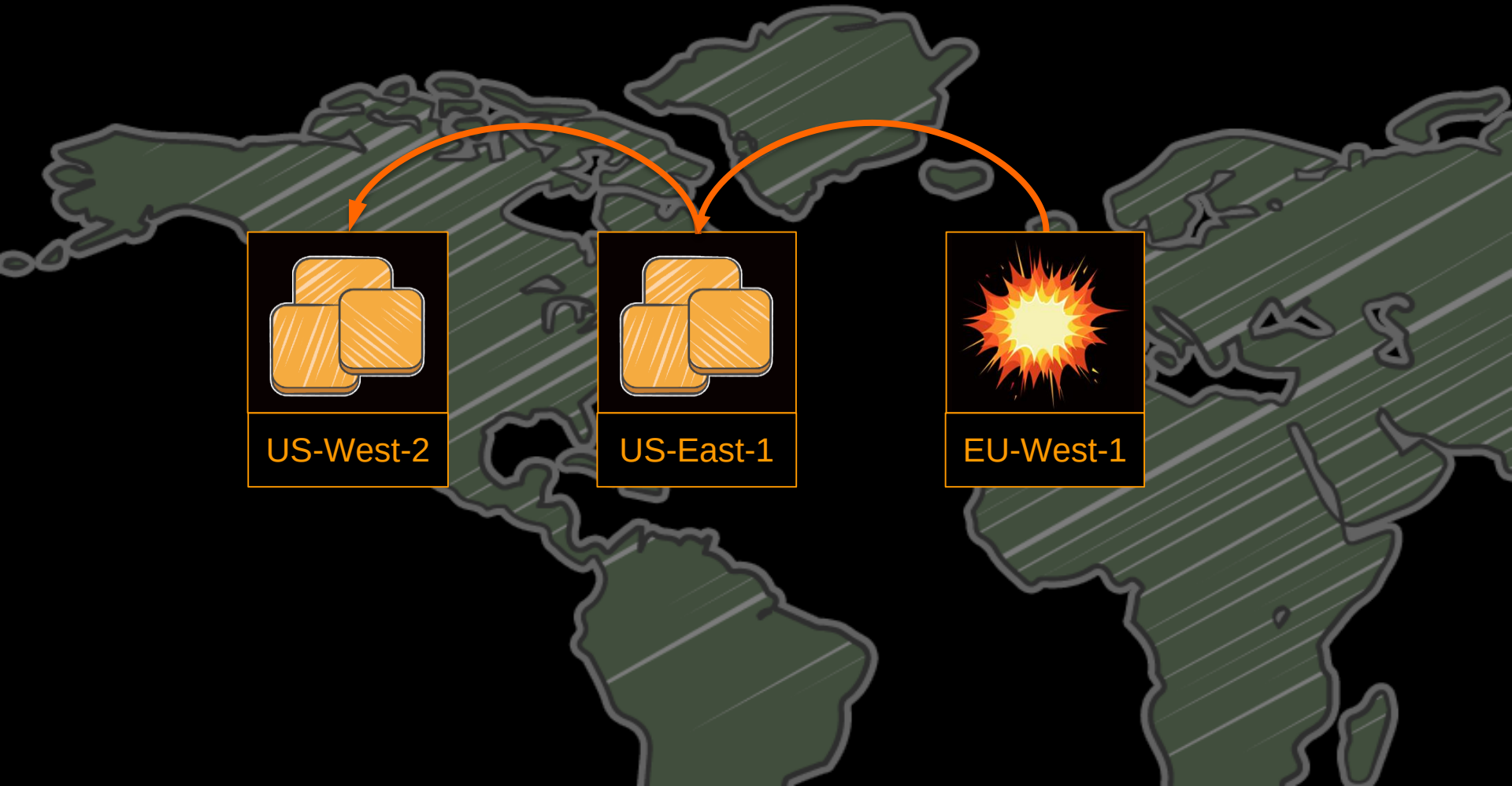
Ubiquitous, Resilient Architecture

Netflix Global

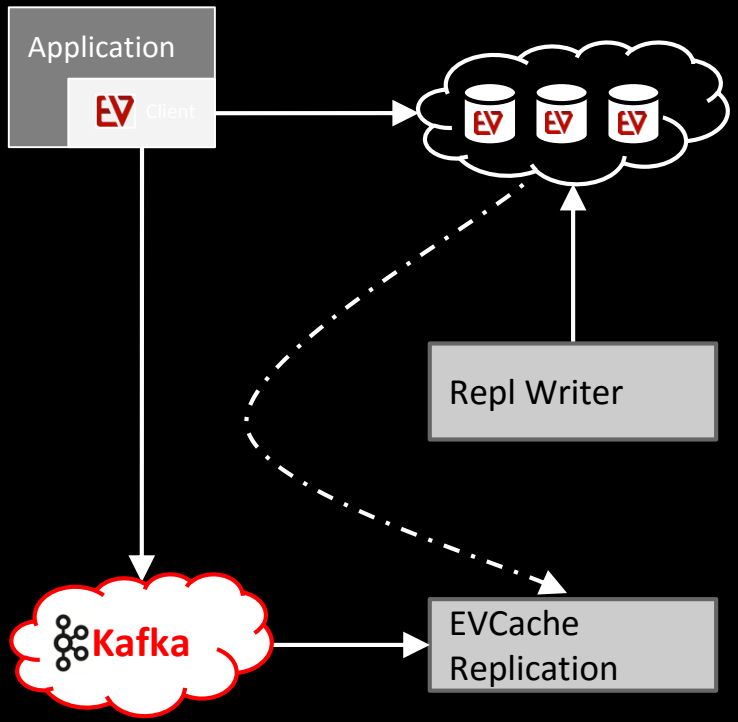
2015

Reliably and efficiently serve any customer from any region


US-West-2
US-East-1
EU-West-1


US-West-2
US-East-1
EU-West-1

Ubiquitous Data

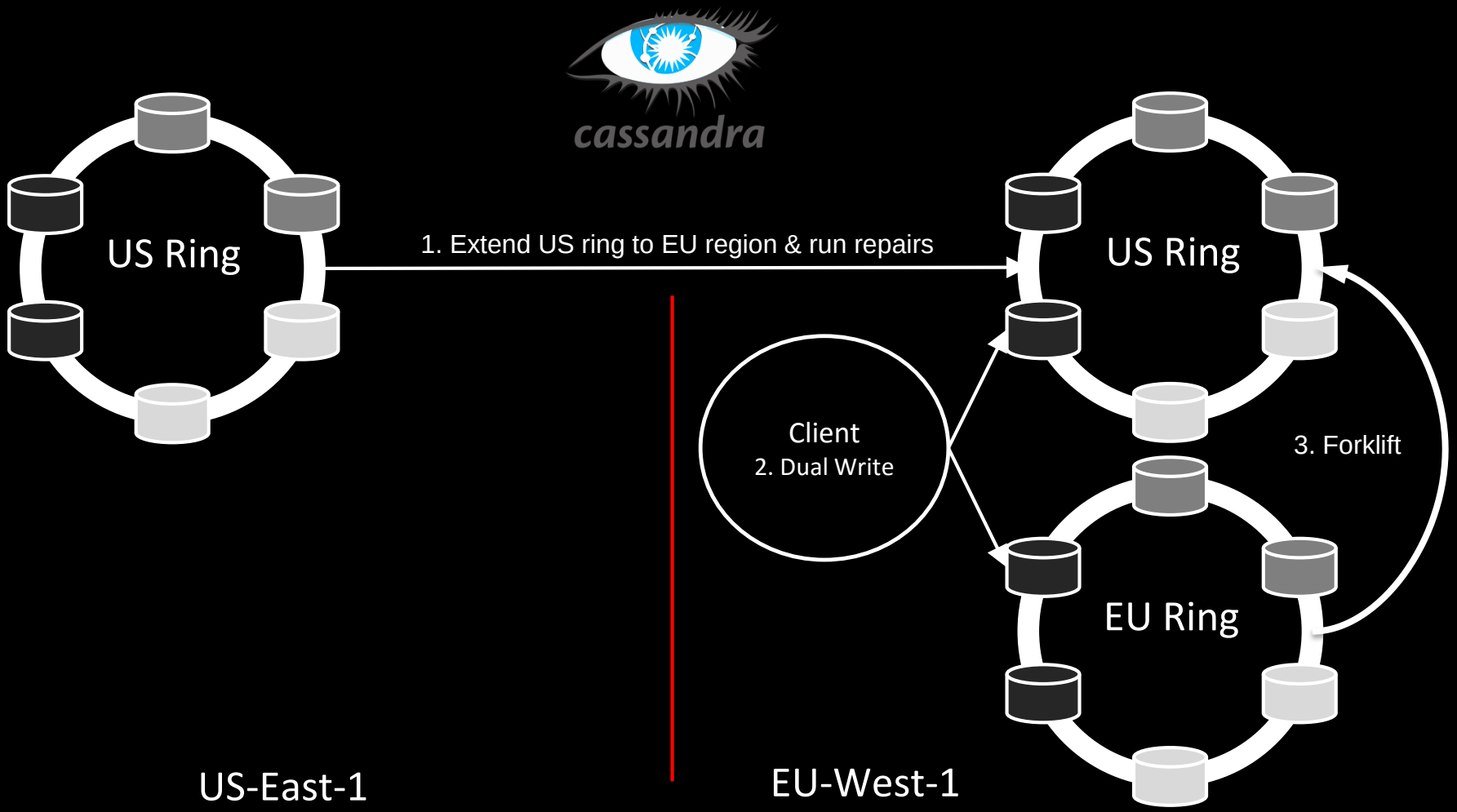


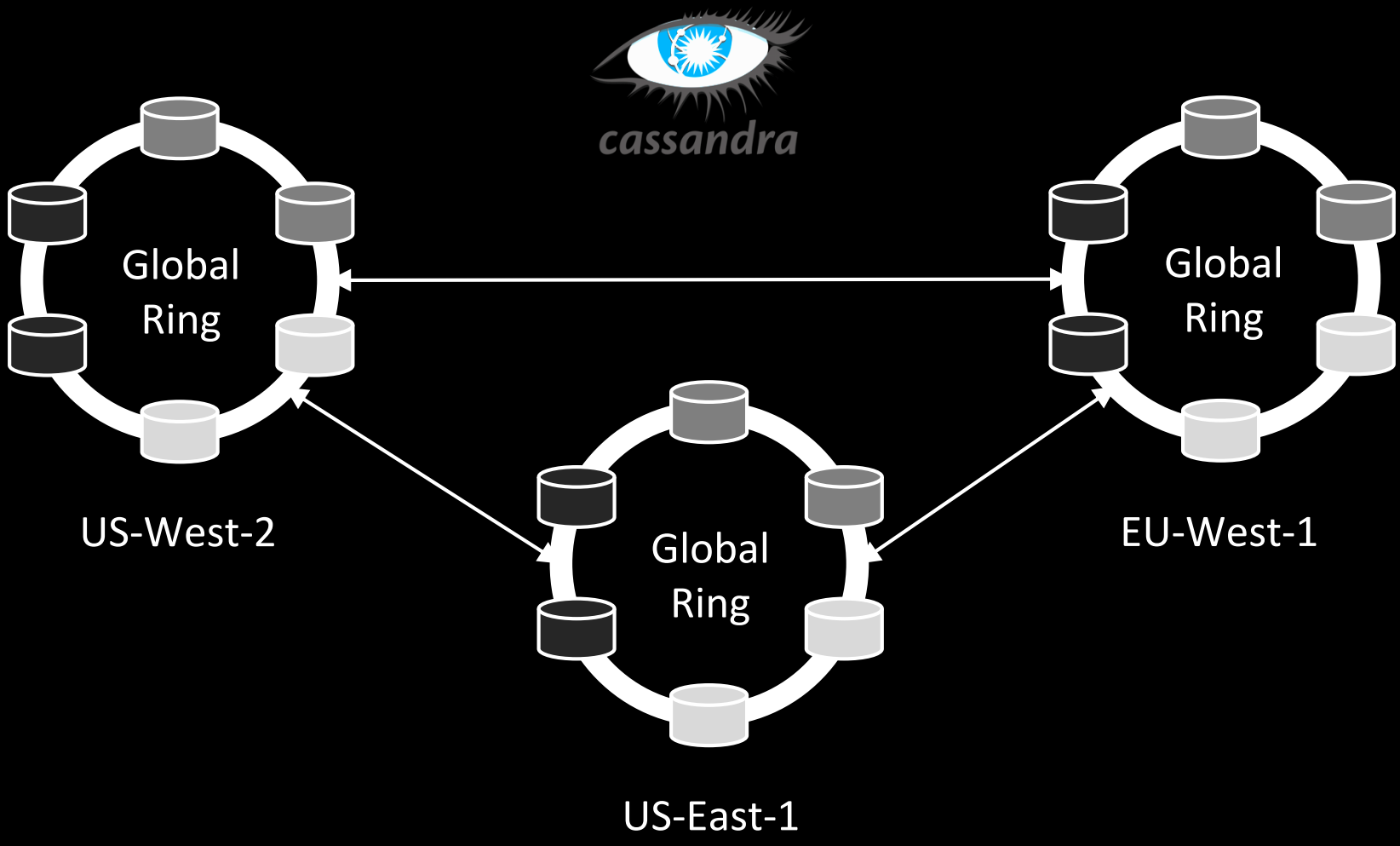
SQS

- High latency
- Read once

Kafka

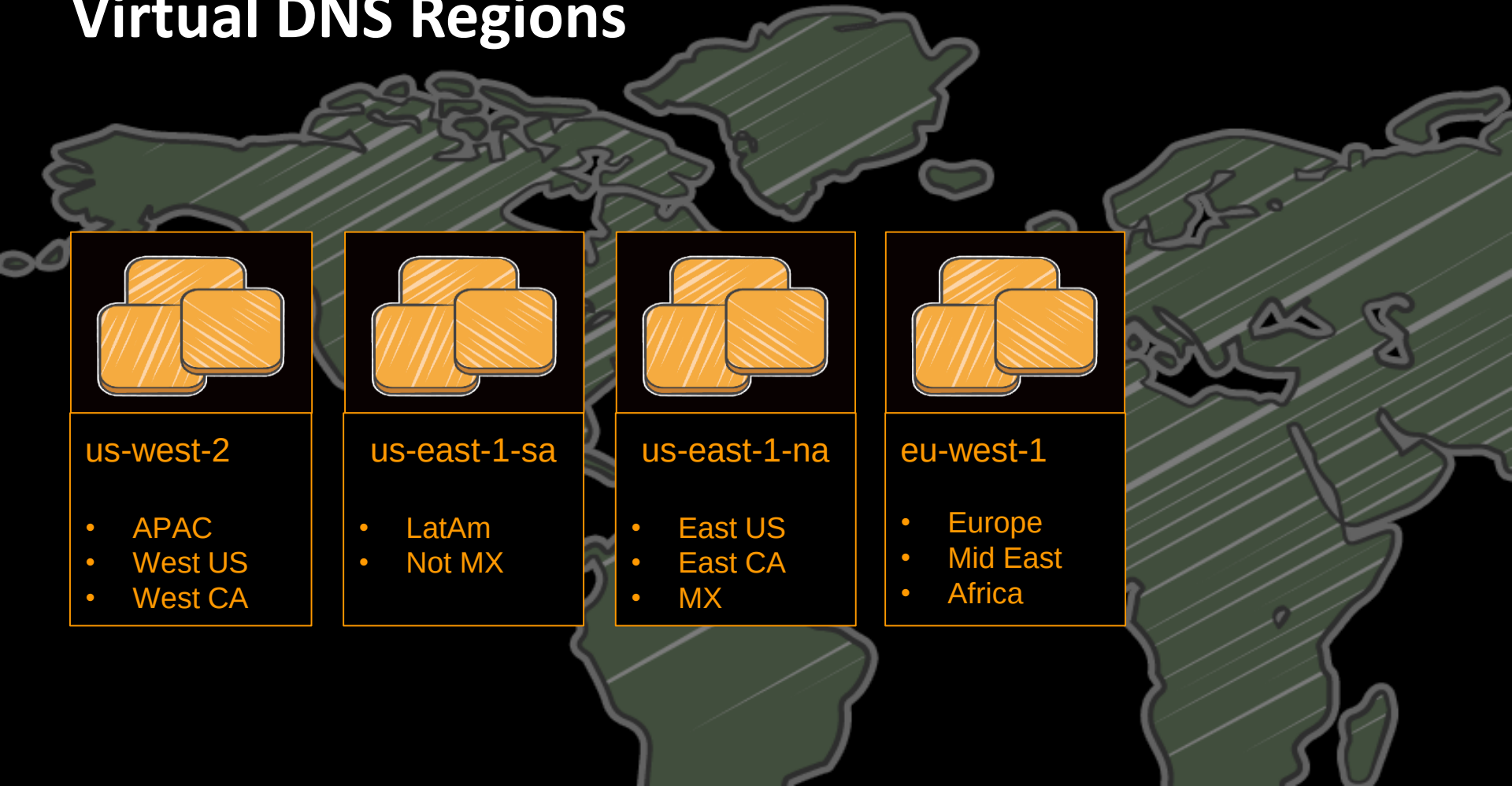
- Low latency
- Multiple readers
- > 1M replications/sec


cassandra
US Ring
1. Extend US ring to EU region & run repairs
US Ring
Client
2. Dual Write
3. Forklift
EU Ring
US-East-1
EU-West-1


cassandra
Global Ring
Global Ring
Global Ring
US-West-2
EU-West-1
US-East-1

Ubiquitous Traffic Management

Virtual DNS Regions



us-west-2

- APAC
- West US
- West CA

us-east-1-sa

- LatAm
- Not MX

us-east-1-na

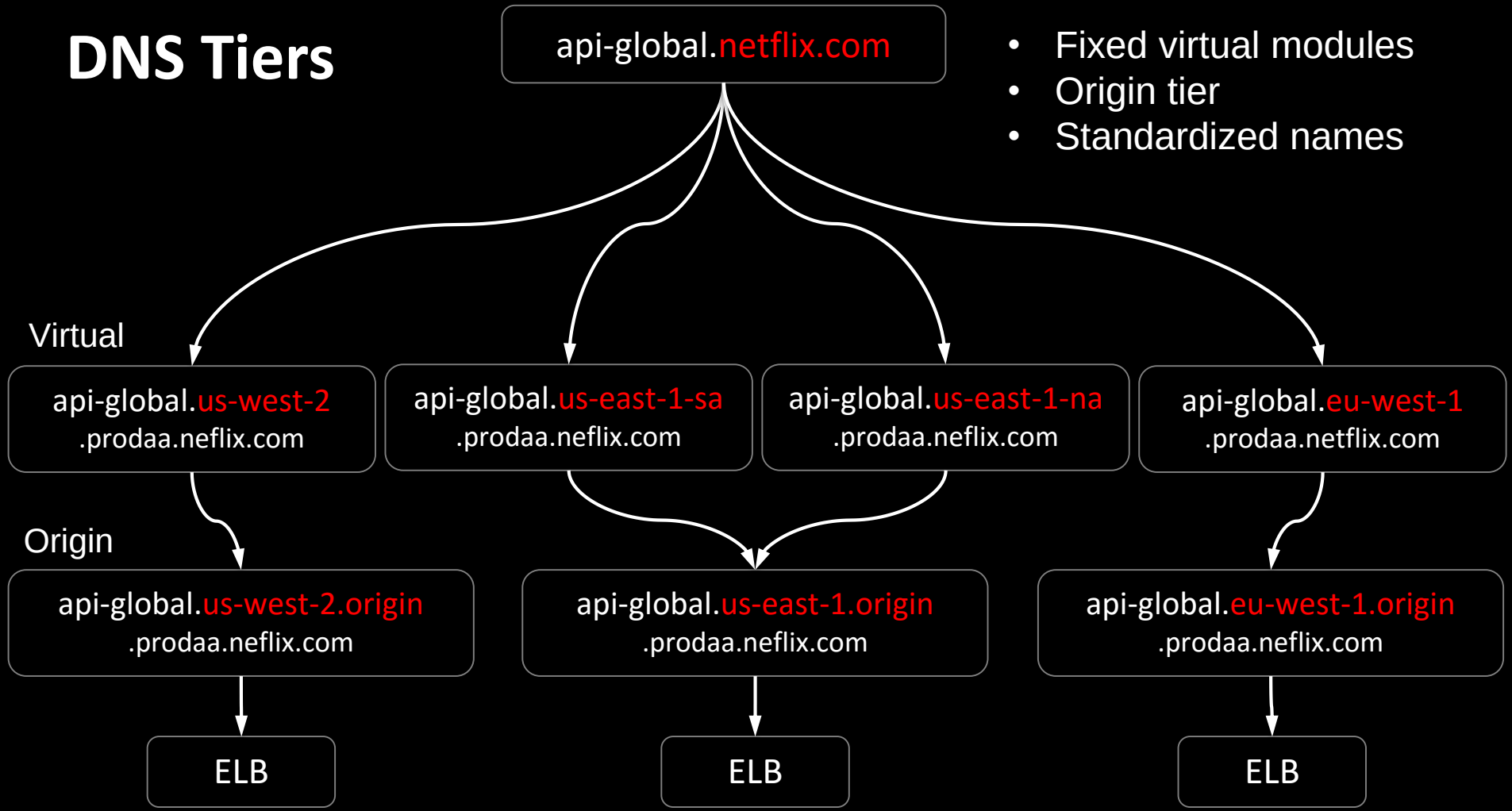
- East US
- East CA
- MX

eu-west-1

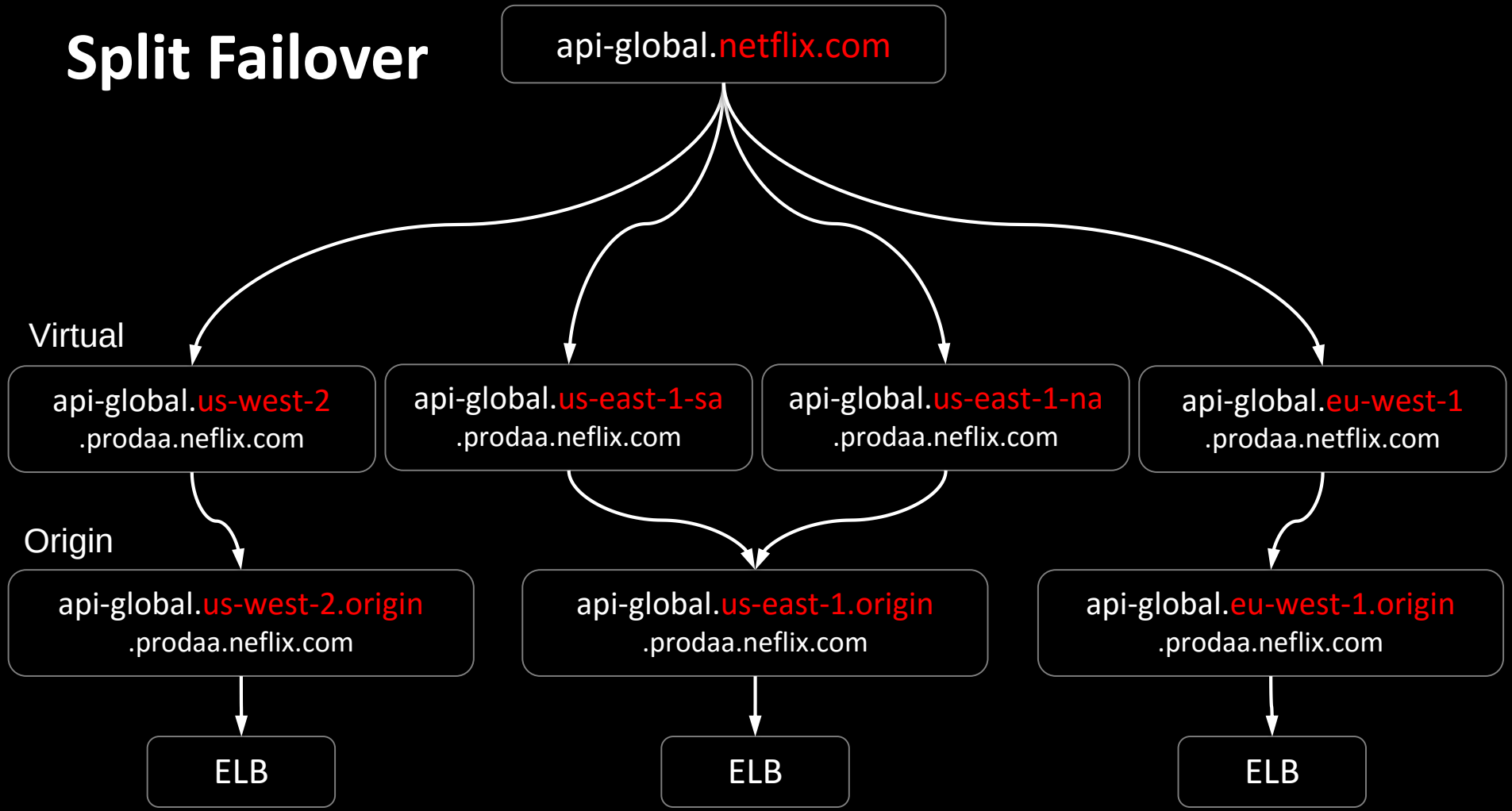
- Europe
- Mid East
- Africa

DNS Tiers

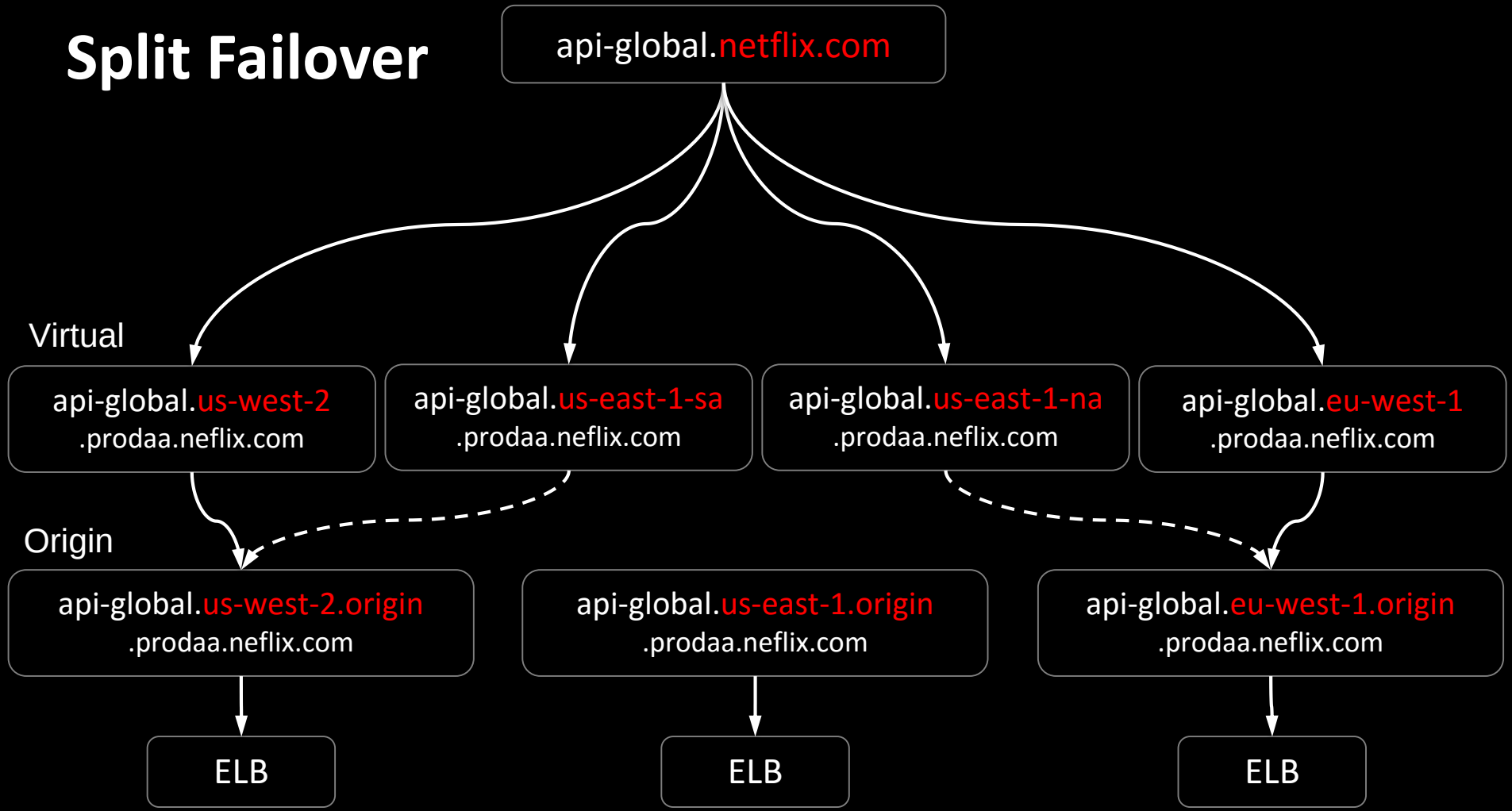
- Fixed virtual modules
- Origin tier
- Standardized names



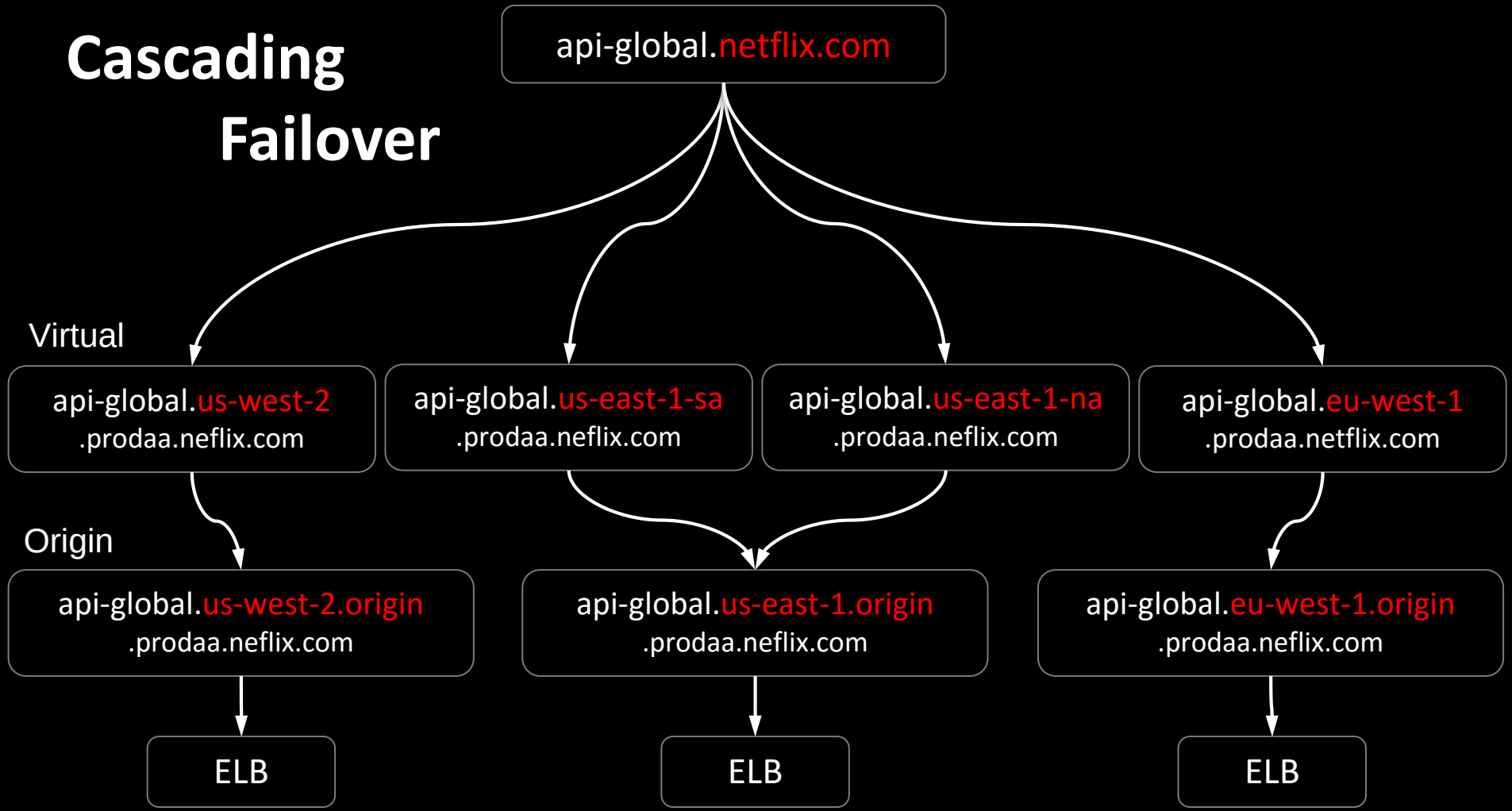
Split Failover



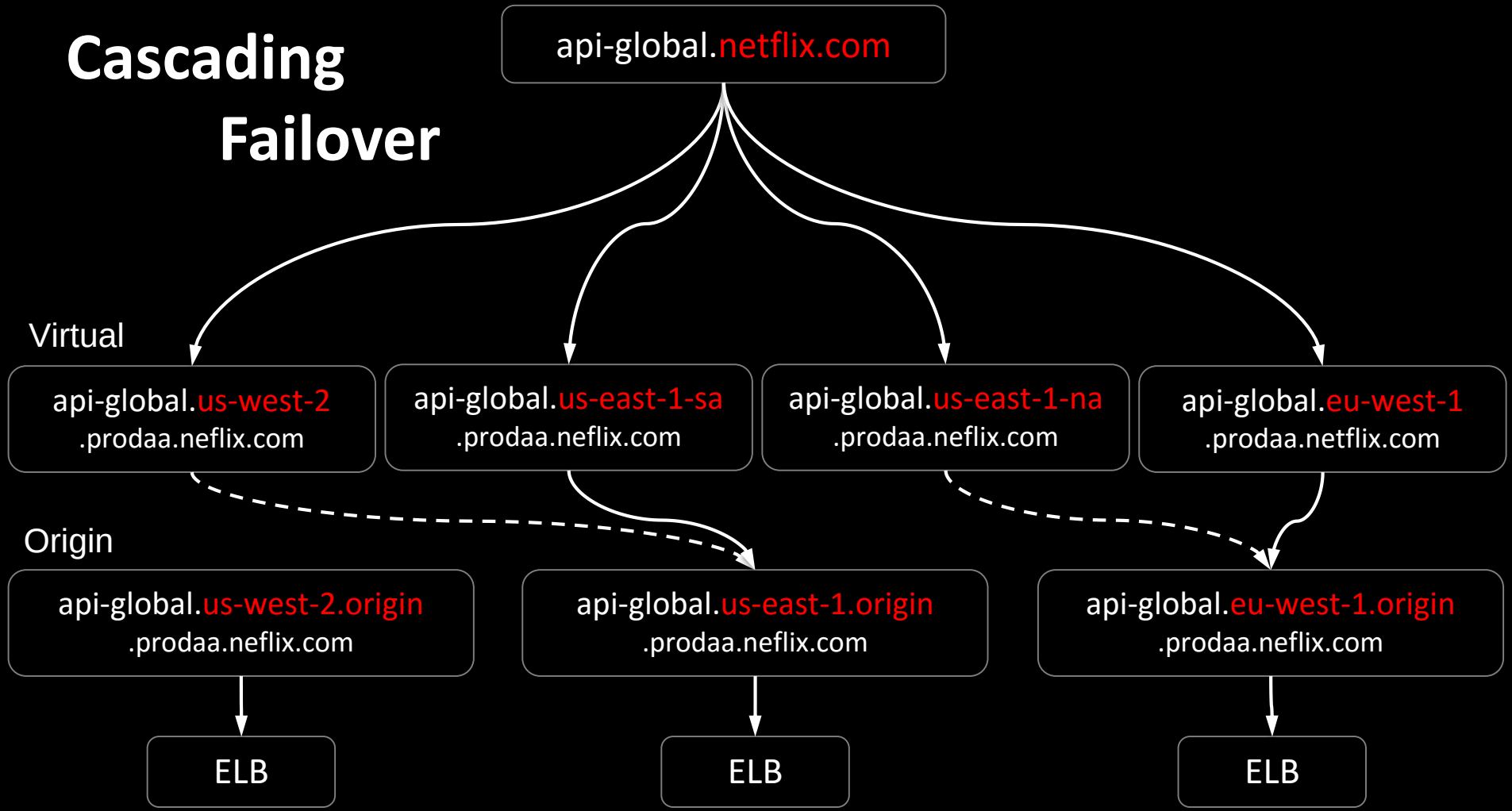
Split Failover



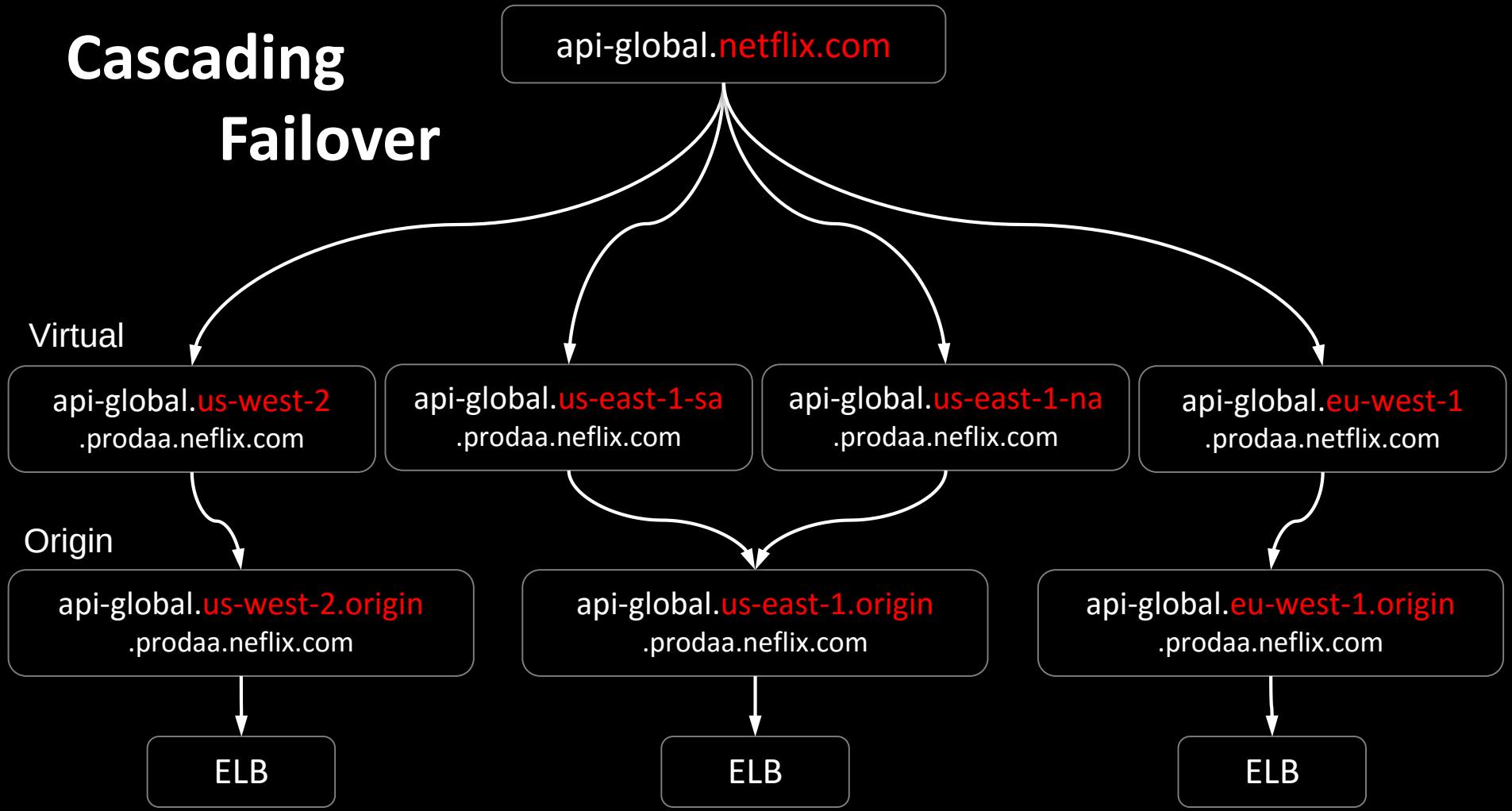
Cascading Failover

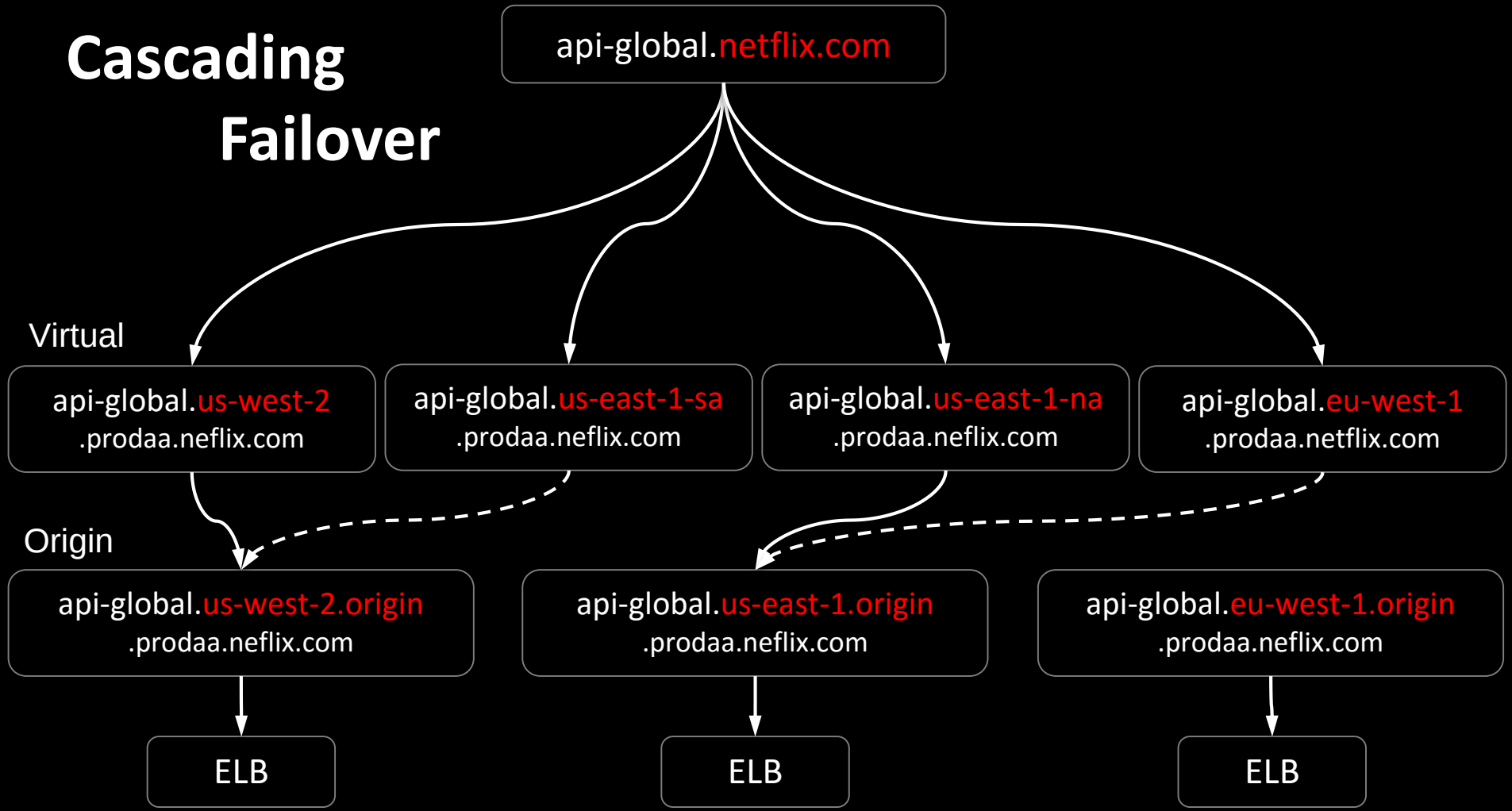


Cascading Failover



Cascading Failover

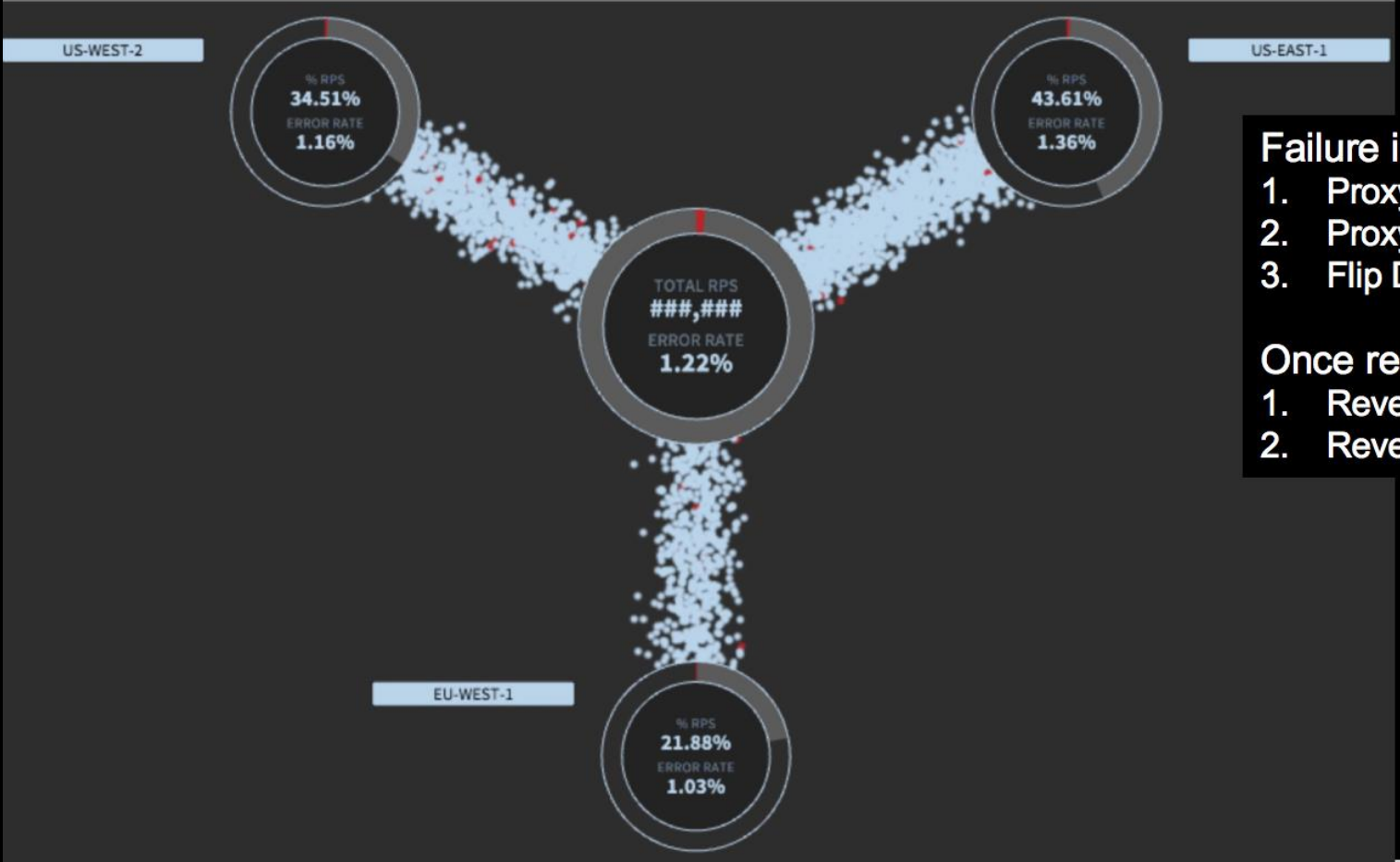



Cascading Failover
api-global.netflix.com
Virtual
api-global.us-west-2 .prodaa.neflix.com
api-global.us-east-1-sa .prodaa.neflix.com
api-global.us-east-1-na .prodaa.neflix.com
api-global.eu-west-1 .prodaa.netflix.com
Origin
api-global.us-west-2.origin .prodaa.neflix.com
api-global.us-east-1.origin .prodaa.neflix.com
api-global.eu-west-1.origin .prodaa.neflix.com
ELB
ELB
ELB

FLUX

Service Traffic Map

Filters Display



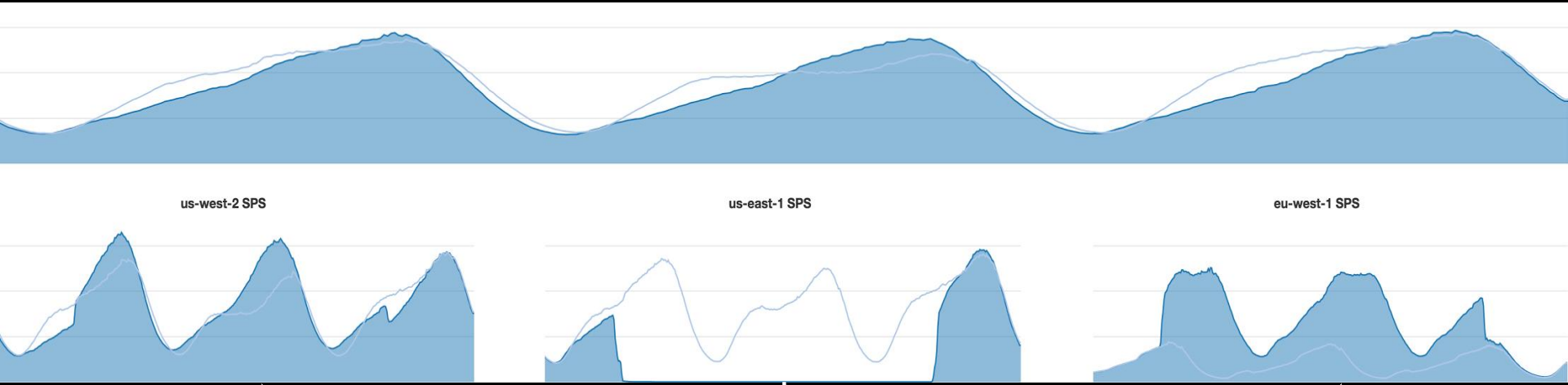
Failure in US-East-1

1. Proxy to EU-West-1 and
2. Proxy to US-West-2
3. Flip DNS to savior regions

Once recovered

1. Revert DNS
2. Revert Zuul

Multi-region Failover




January 6th, 2016

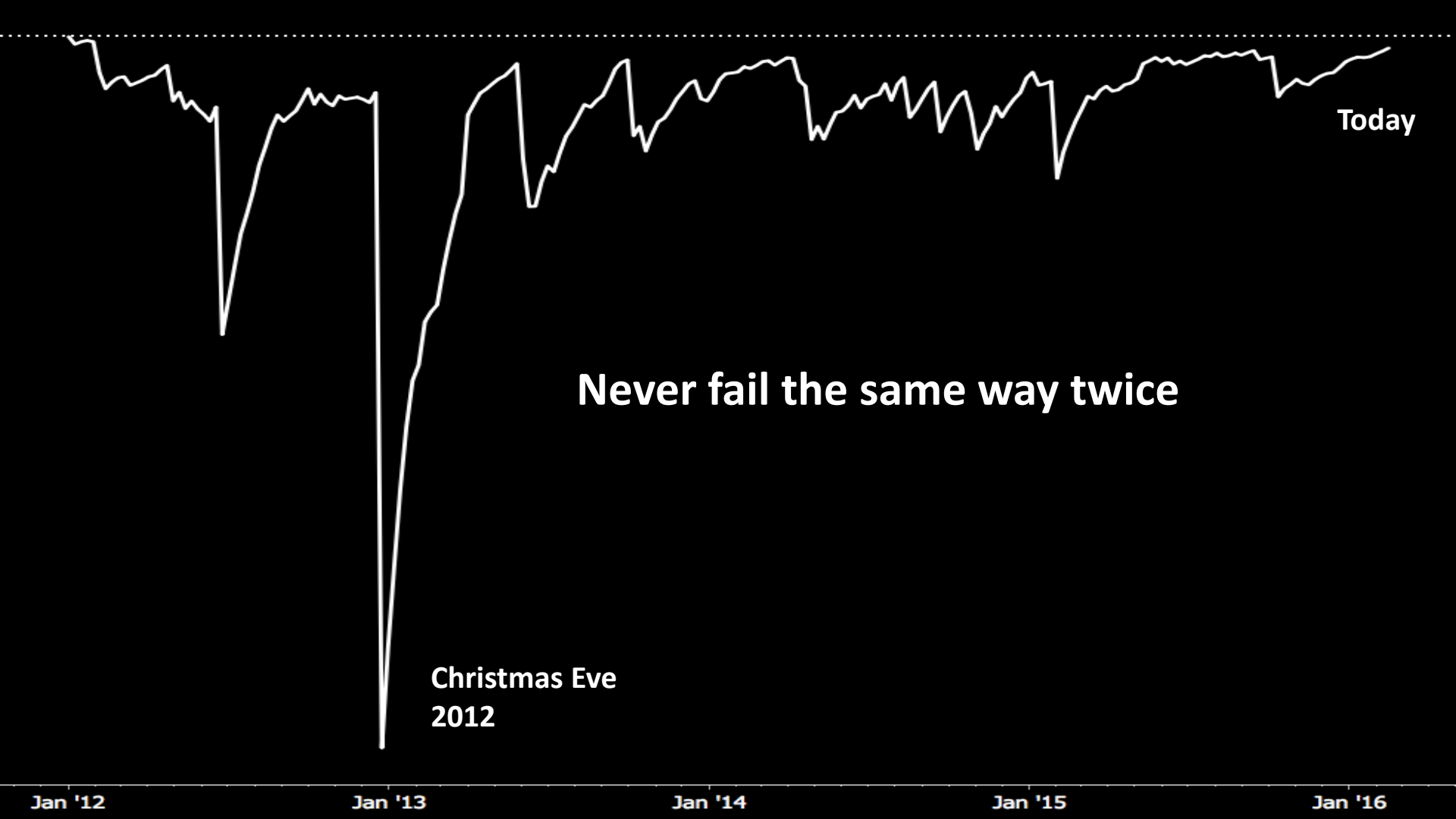

“Going global is just like having a baby.”
- Reed Hastings, Netflix CEO

What's next?

- Global latency
- Edge computing
- ML-based monitoring
- Self-healing systems
- Capacity utilization
- Fast, autonomous traffic
- Integrate DB & caching

#NetflixEverywhere

Takeaways


Today
Never fail the same way twice
Christmas Eve
2012
Jan '12
Jan '13
Jan '14
Jan '15
Jan '16

Know your resiliency patterns

Pattern	Properties
DC	SPoF, infrastructure heavy lifting
Cloud (one region)	Multiple DCs, one control plane
Islands	Regional containment
Isthmus	Regional ELB bypass
Active-active	Regional failover
Global	Ubiquity, resiliency, efficiency

Invest in architectural pillars

- Micro-services
- Database
- Caching
- Traffic

Think globally, act locally

#NetflixEverywhere

NETFLIX | OSS

Data Persistence

Storing and Serving data in the Cloud.

Handling over a trillion data operations per day requires an interesting mix of “off the shelf OSS” and in house projects. No single data technology can meet every use case or satisfy every latency requirement. Our needs range from non-durable in-memory stores like Memcached and Redis, to searchable datastores such as Elastic and durable must-never-go-down datastores like Cassandra and MySQL.

Our Cloud usage and the scale at which we consume these technologies, has required us to build tools and services that enhance the datastores we use. We’ve created the sidecars Raigad and Priam to help with the deployment, management and backup/recovery of our hundreds of Elastic and Cassandra clusters. We’ve created EVCache and Dynomite to use Memcached and Redis at scale. We’ve even developed the Dyno client library to better consume Dynomite in the Cloud.

netflix.github.io

NETFLIX | OSS

Common Runtime Services & Libraries

Runtime containers, libraries and services that power microservices

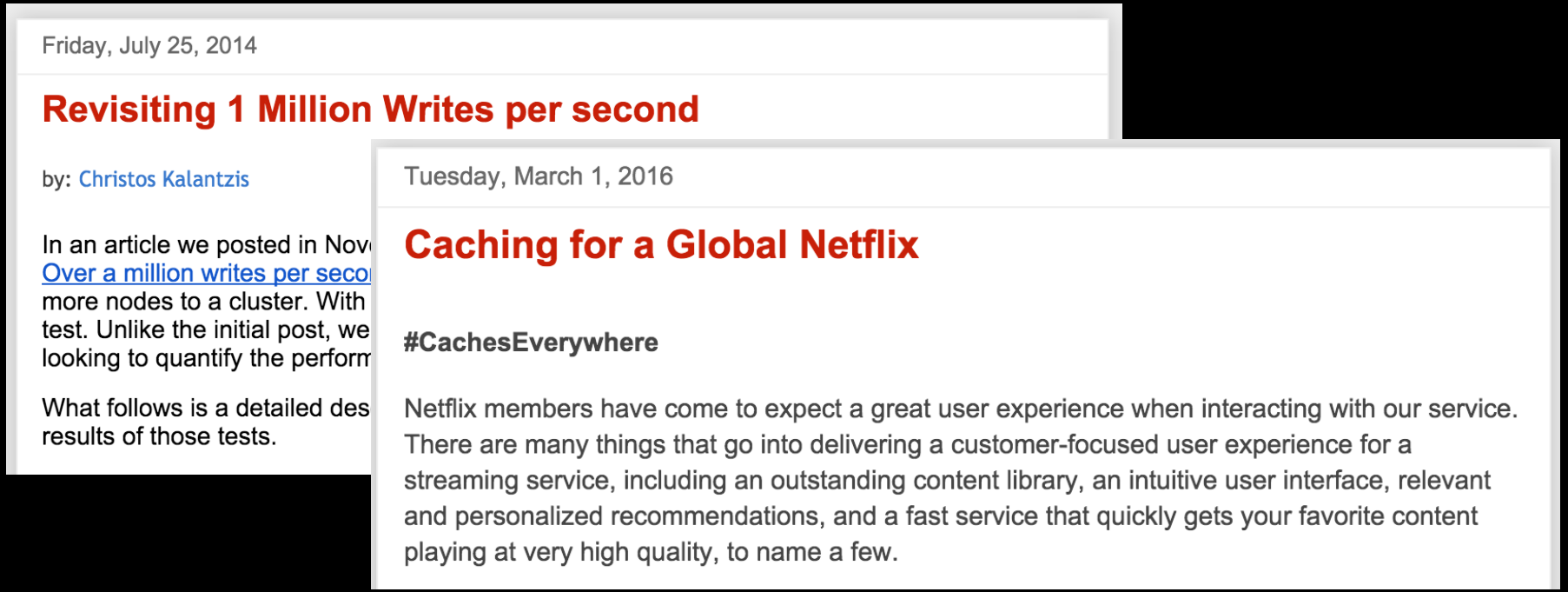
The cloud platform is the foundation and technology stack for the majority of the services within Netflix. The cloud platform consists of cloud services, application libraries and application containers. Specifically, the platform provides service discovery through Eureka, distributed configuration through Archaius, resilent and intelligent inter-process and service communication through Ribbon. To provide reliability beyond single service calls, Hystrix is provided to isolate latency and fault tolerance at runtime. The previous libraries and services can be used with any JVM based container.

The platform provides JVM container services through Karyon and Governator and support for non-JVM runtimes via the Prana sidecar. While Prana provides proxy capabilities within an instance, Zuul (which integrates Hystrix, Eureka, and Ribbon as part of its IPC capabilities) provides dyamically scriptable proxying at the edge of the cloud deployment.

The platform works well within the EC2 cloud utilizing the Amazon autoscaler. For container applications and batch jobs running on Apache Mesos, Fenzo is a scheduler that provides advanced scheduling and resource management for cloud native frameworks. Fenzo provides plugin implementations for bin packing, cluster autoscaling, and custom scheduling optimizations can be implemented through user-defined plugins.

netflix.github.io

Netflix Tech Blog



Friday, July 25, 2014

Revisiting 1 Million Writes per second

by: Christos Kalantzis

In an article we posted in Nov
Over a million writes per seco
more nodes to a cluster. With
test. Unlike the initial post, we
looking to quantify the perform

What follows is a detailed des
results of those tests.

Tuesday, March 1, 2016

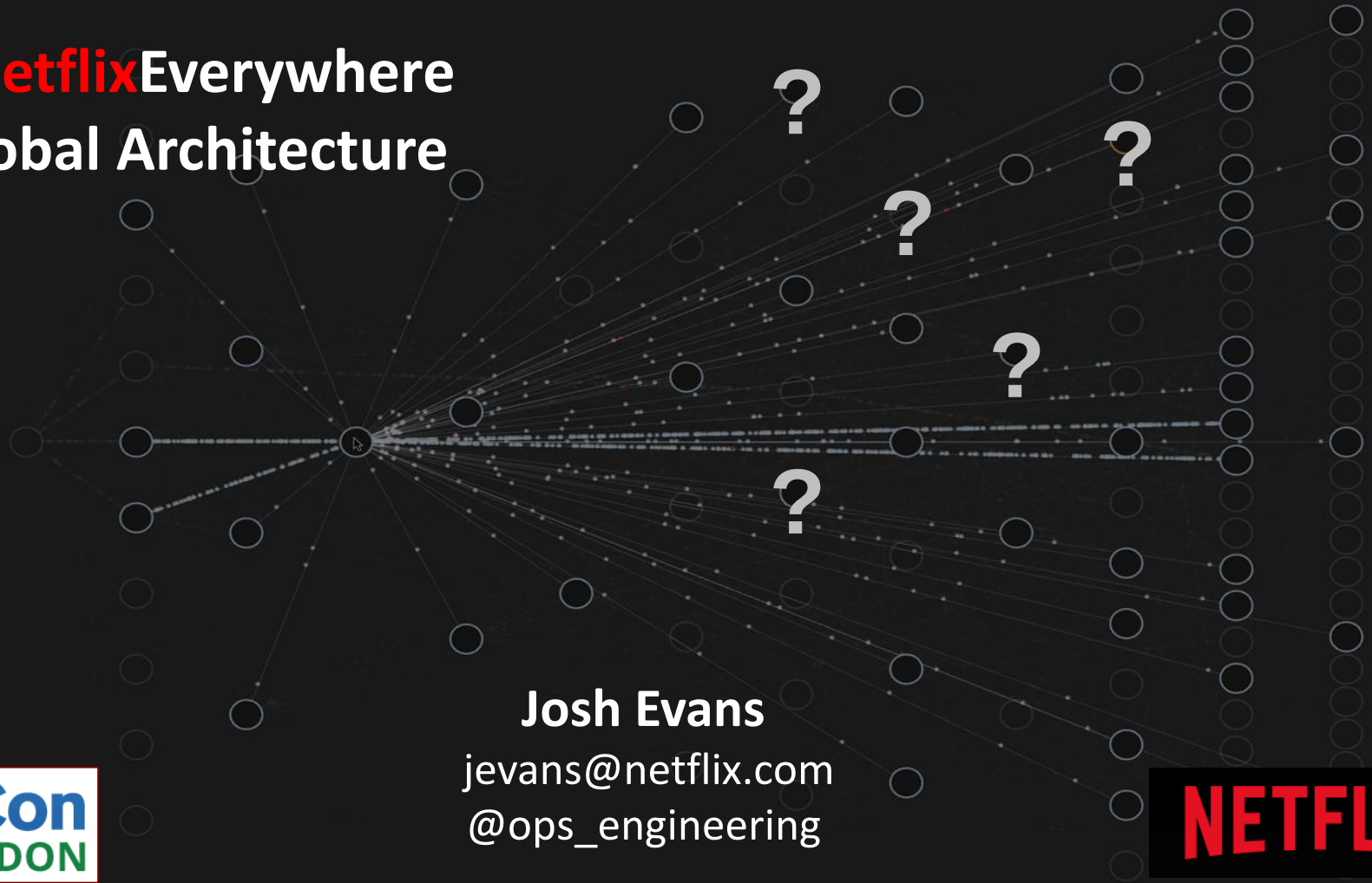
Caching for a Global Netflix

#CachesEverywhere

Netflix members have come to expect a great user experience when interacting with our service. There are many things that go into delivering a customer-focused user experience for a streaming service, including an outstanding content library, an intuitive user interface, relevant and personalized recommendations, and a fast service that quickly gets your favorite content playing at very high quality, to name a few.

techblog.netflix.com

#NetflixEverywhere Global Architecture



Josh Evans
jevans@netflix.com
@ops_engineering

QCon LONDON

NETFLIX