Josip Stuhli, Sofascore CTO

Scaling: From 0 to 25 million

Sofascore

09:08
94
3 - 3
AP
DETAILS
LINEUPS
STATISTICS
COMMENTARY
Penalty shootout
4
L. Messi, P. Dybala, L. Paredes, G. Montiel
2
K. Mbappé, K. Coman, A. Tchouaméni, R. K. Muani
Number of Ratings: 67k
9.5
Avg.
See all rated matches
9.2
L. Messi
Sofascore ratings
8.5
K. Mbappé
PEN 4 - 2
4 - 2
Gonzalo Montiel
Randal Kolo Muani
3 - 2

09:09
94
Lionel Messi
9.2
Sofascore
statistical rating
View player details
COMPARE
Match Shotmap
Minutes played
120'
Goals
2
Expected Goals (xG)
1.88
Assists
0
Expected Assists (xA)
0.26
Shots on target
4
Shots off target
1
Shots blocked
0
Dribble attempts (succ.)
2 (0)
A. Mac Allister
L. Messi
J. Álvarez

09:10
94
3 - 3
AP
TAILS
LINEUPS
STATISTICS
COMMENTARY
ALL
1ST
2ND
ET1
ET2
Match overview
54%
Ball possession
46%
3.23
Expected Goals (xG)
2.27
5
Big chances
3
20
Total shots
10
2
Goalkeeper saves
7
6
Corner kicks
5
26
Fouls
19
635
Passes
532
23
Tackles
26
18
Free kicks
23
5
Yellow cards
3

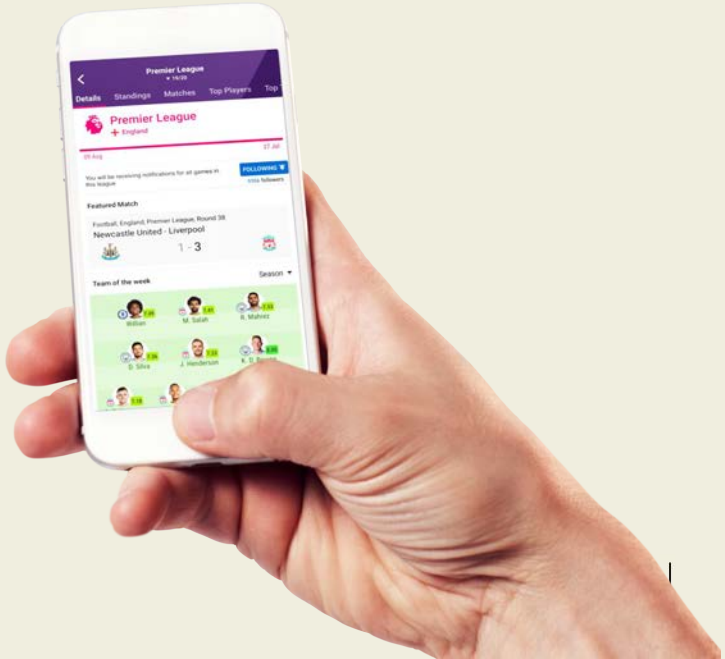
Monthly

28+ million users

1.4 PB+ of traffic

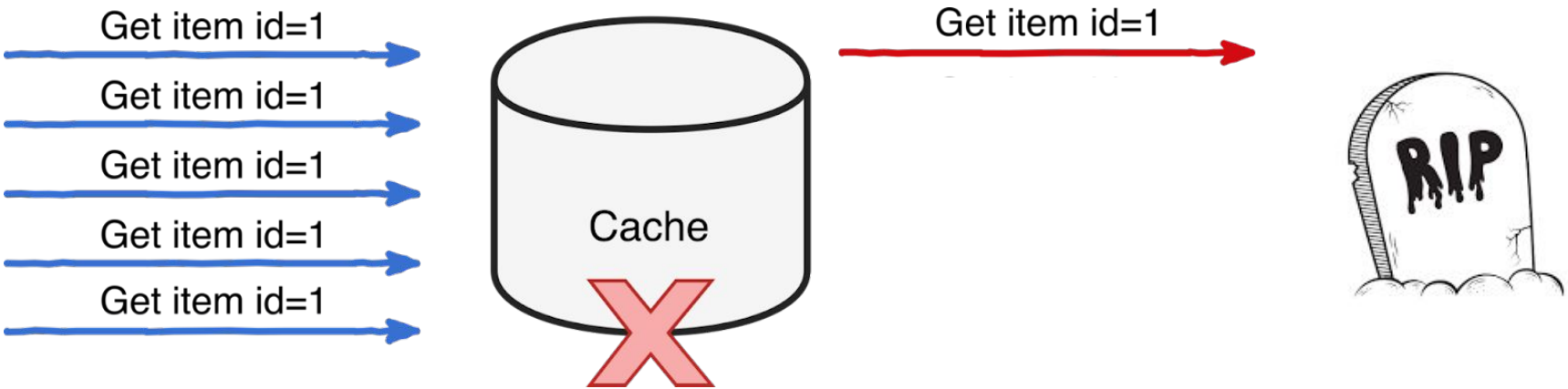
400+ billion requests

1.6M req/s peak



Full page cache

Cache Stampede problem



Proactive caching

+- 2h

MEMCACHED

Full page cache

Know how your users interact with your app (system)

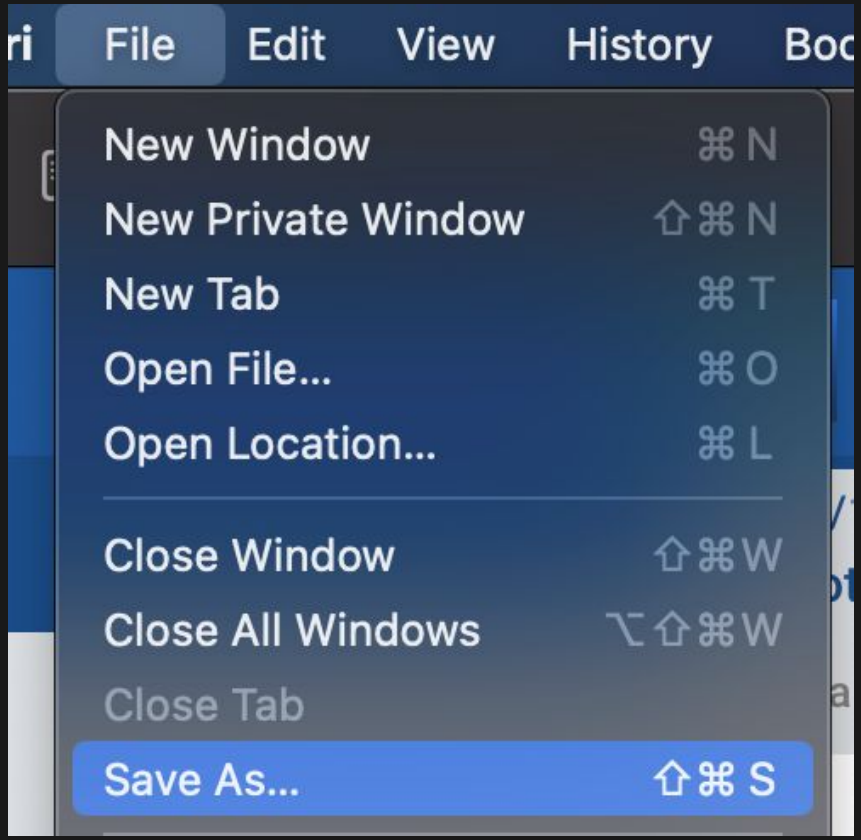
15

SofaScore
EL CLÁSICO
FCB
VS

CPU

Apache
HTTP SERVER

php
5.3


ri
File
Edit
View
History
Boo
New Window ⌘N
New Private Window ⇧⌘N
New Tab ⌘T
Open File... ⌘O
Open Location... ⌘L
Close Window ⇧⌘W
Close All Windows ⌥⇧⌘W
Close Tab
Save As... ⇧⌘S

VERY NICE
GREAT SUCCESS

amazon
web services

Varnish

HTTP accelerator

The cache is often the most critical piece of software for a web-based business

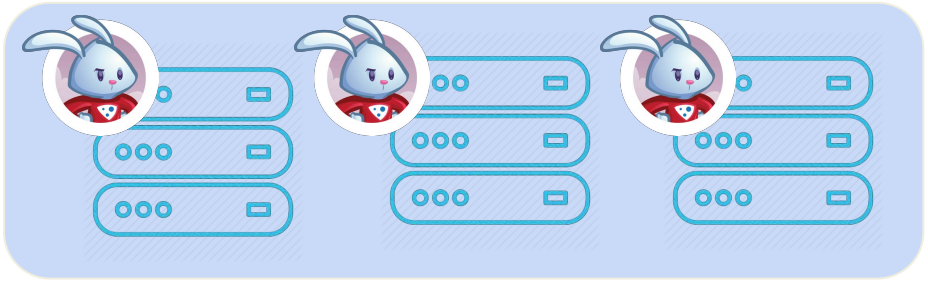
Sofascore

Request coalescing

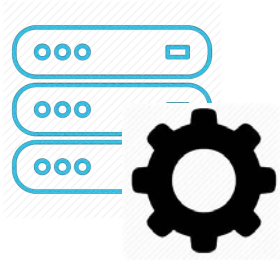
Stateless!

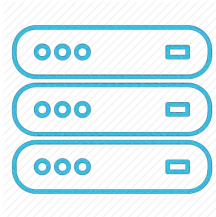
Amazon Machine
Image (AMI)

Elastic
Load Balancer

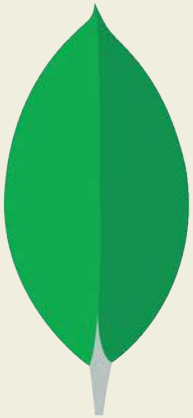


Auto Scaling Group





Be stateless where possible



mongoDB®

webscale

mongoDB®

- Replication issues
- Data type errors
- Difficult to analyse data
- Locking
- Foreign keys

PostgreSQL

Why postgres?

- Open source
- Advanced SQL capabilities
- Minimal locking! (MVCC)
- Great performance!

~ 1 million
queries / s

~ 90k
transactions / s

Pros and Cons

- Simple Replication
- Simple configuration
- Leader election

- Stability
- Foreign keys, consistency
- Analytical power

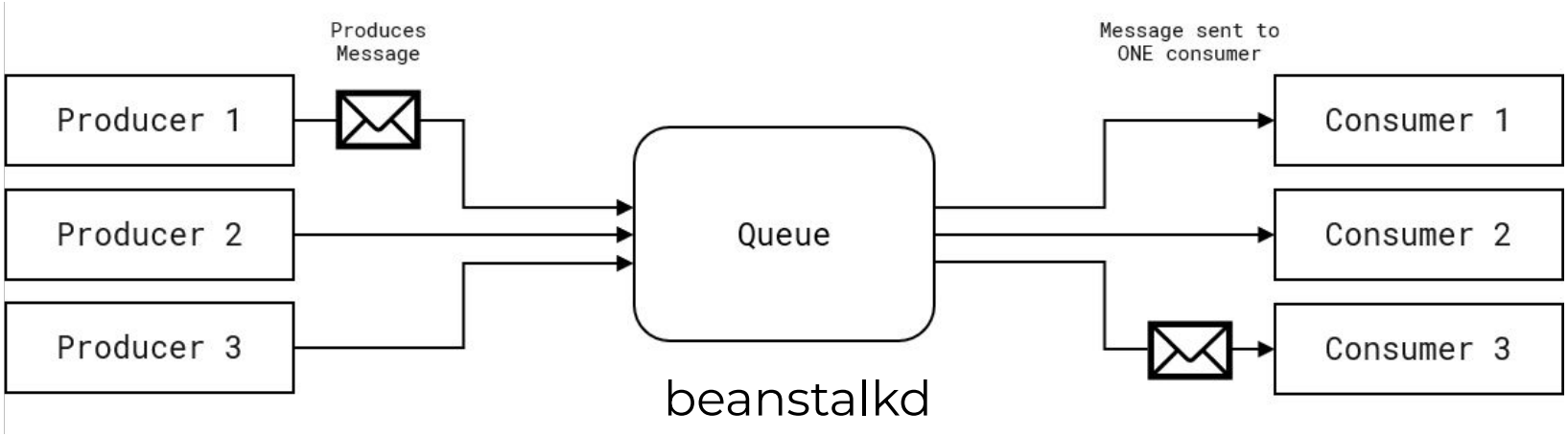
Recognize the right tool for the job and change if necessary

Microservices?

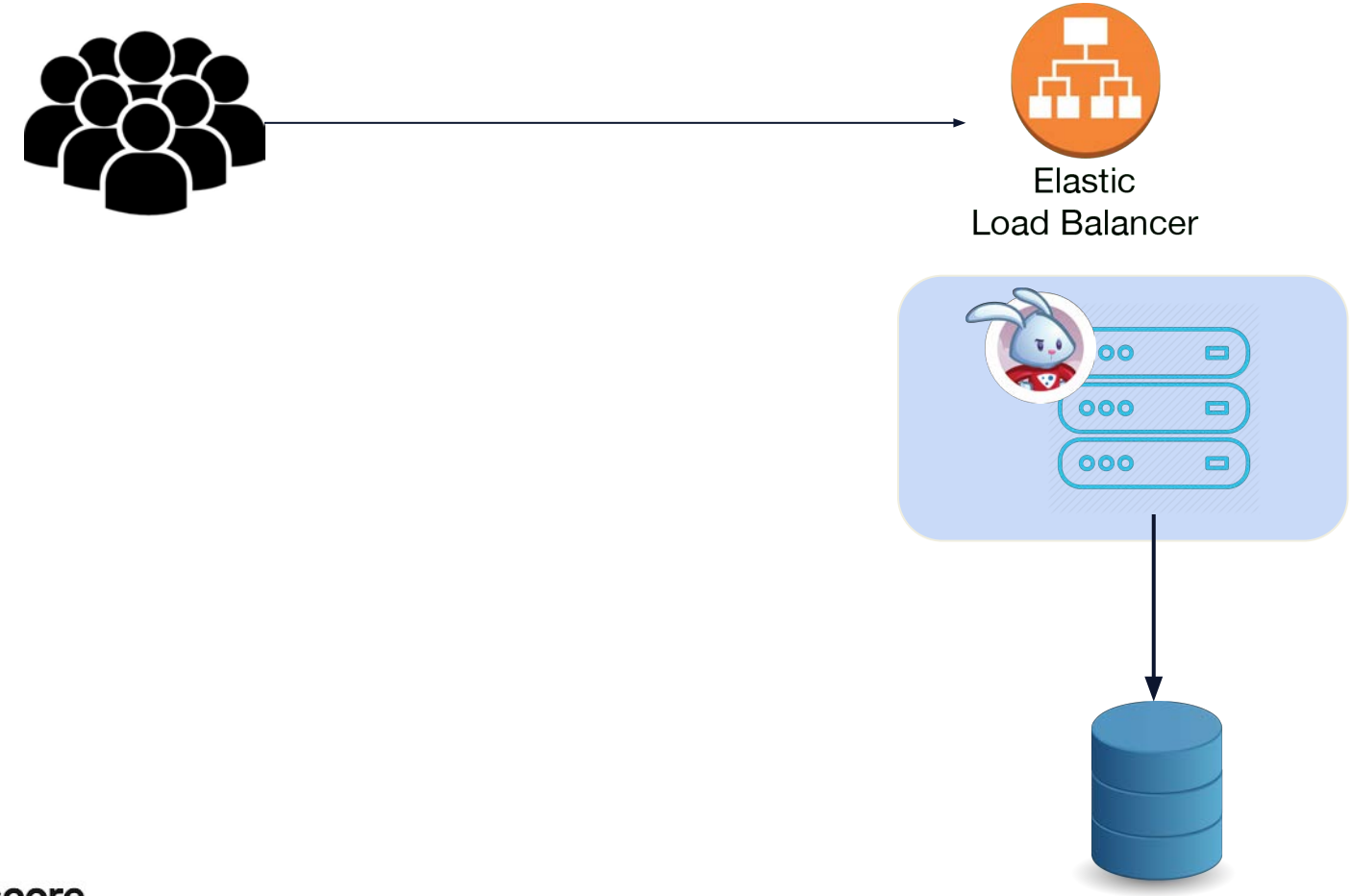
microservices
serverless
Multi AZ
auto scaling
YOUR APP
WITH 0 USERS
COCOA PEBBLES
imgflip.com

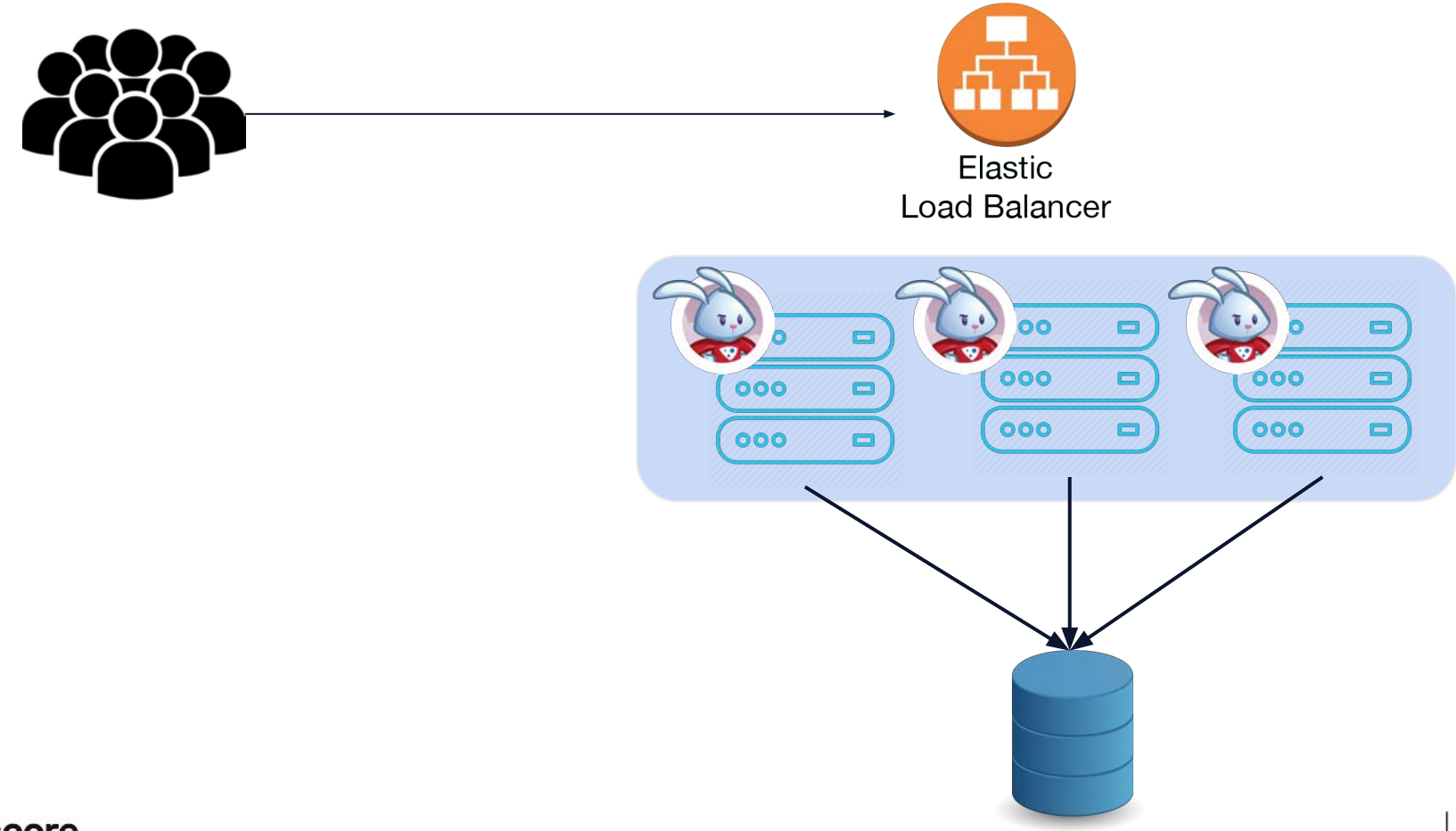
Async communication

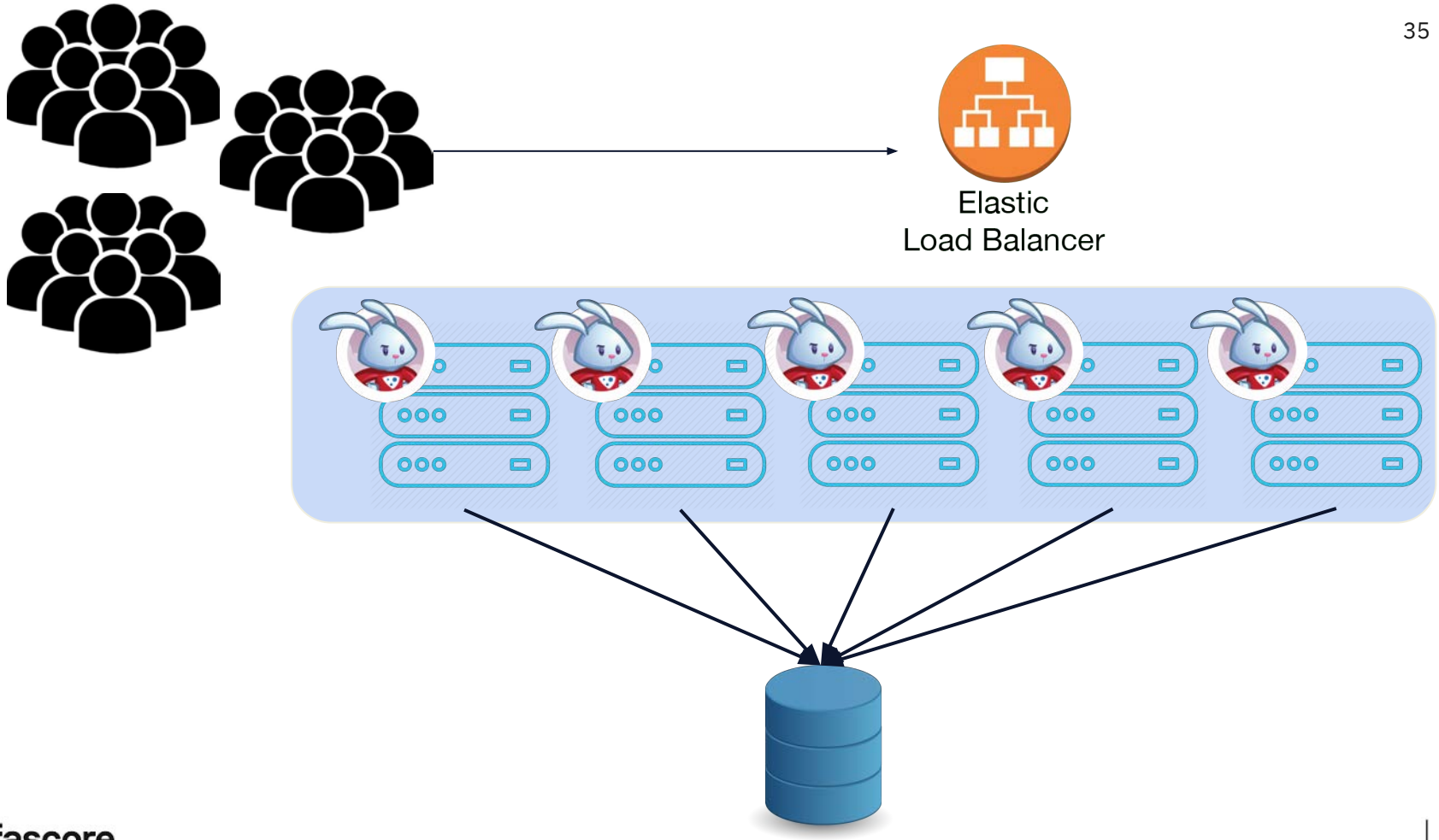
Queues

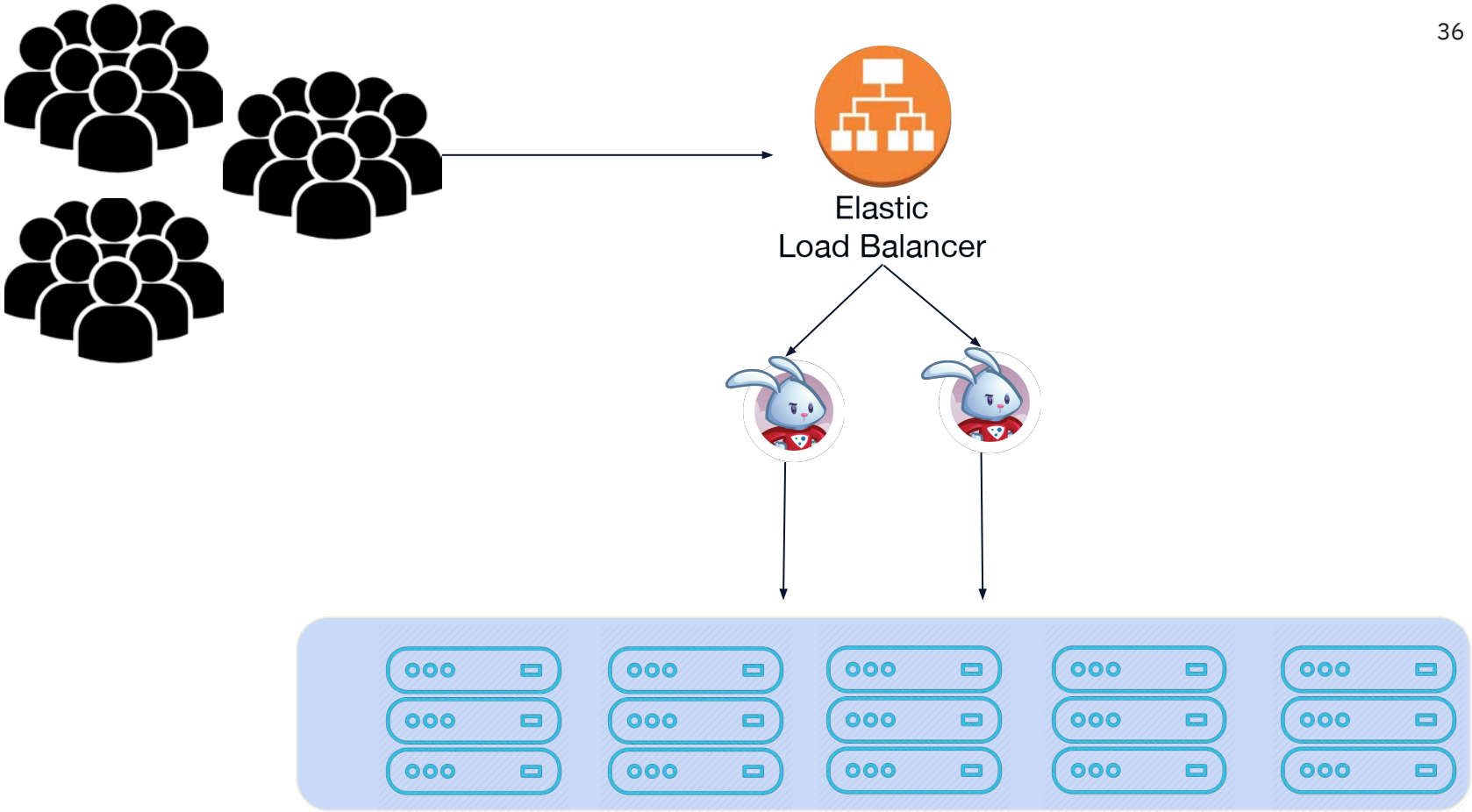

Produces
Message
Message sent to
ONE consumer
Producer 1
Producer 2
Producer 3
Queue
Consumer 1
Consumer 2
Consumer 3
beanstalkd

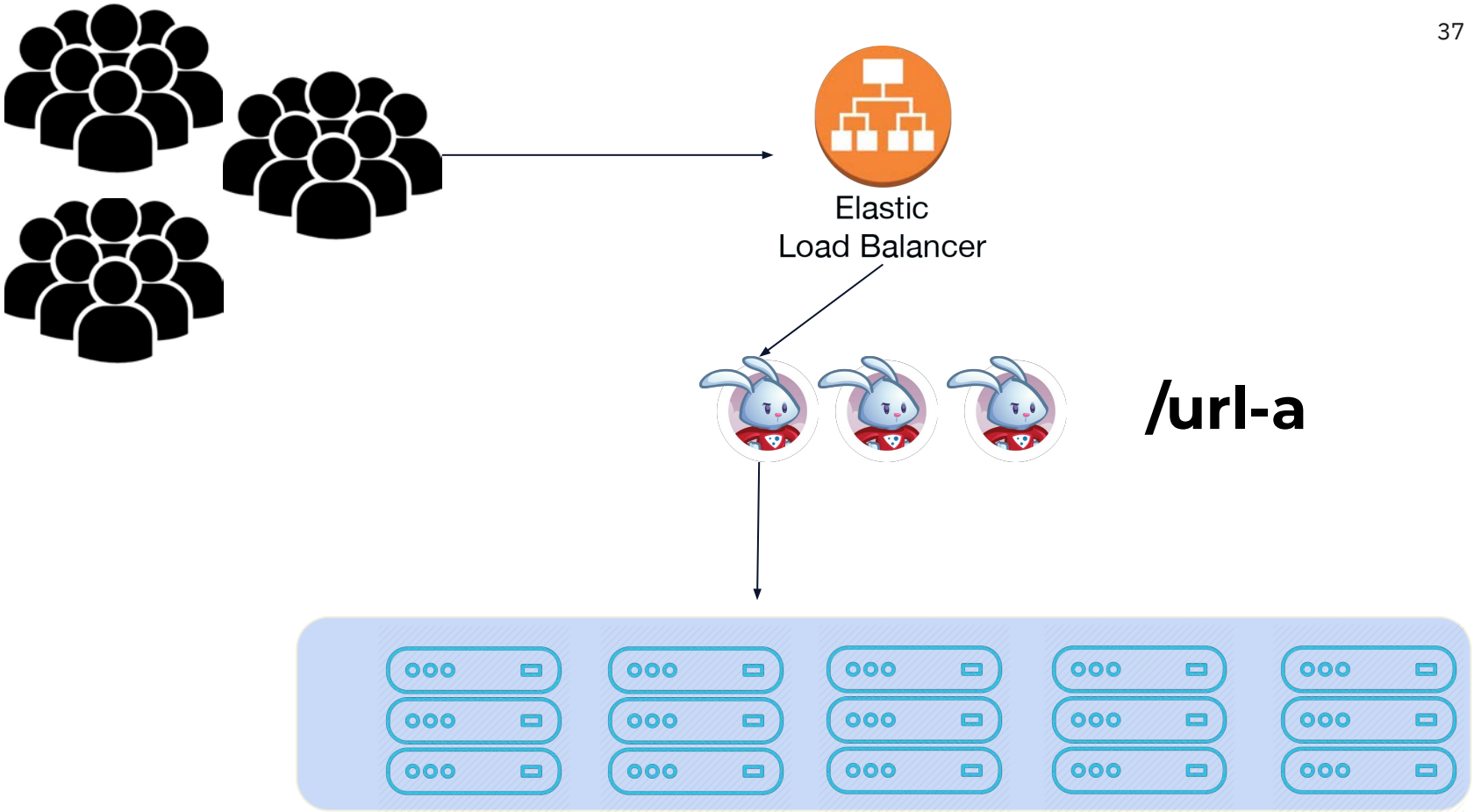
Queue everything that is slow

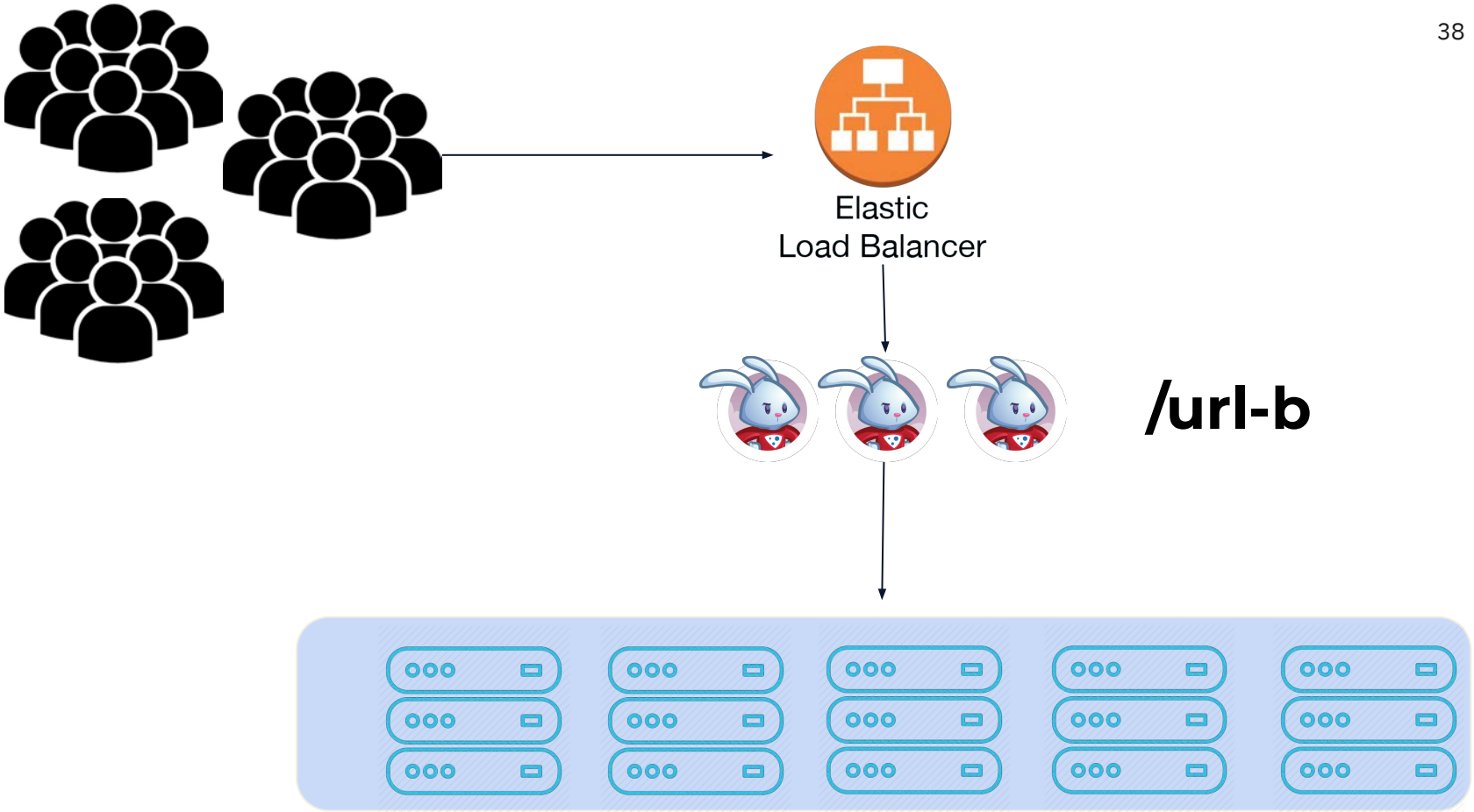

Elastic
Load Balancer

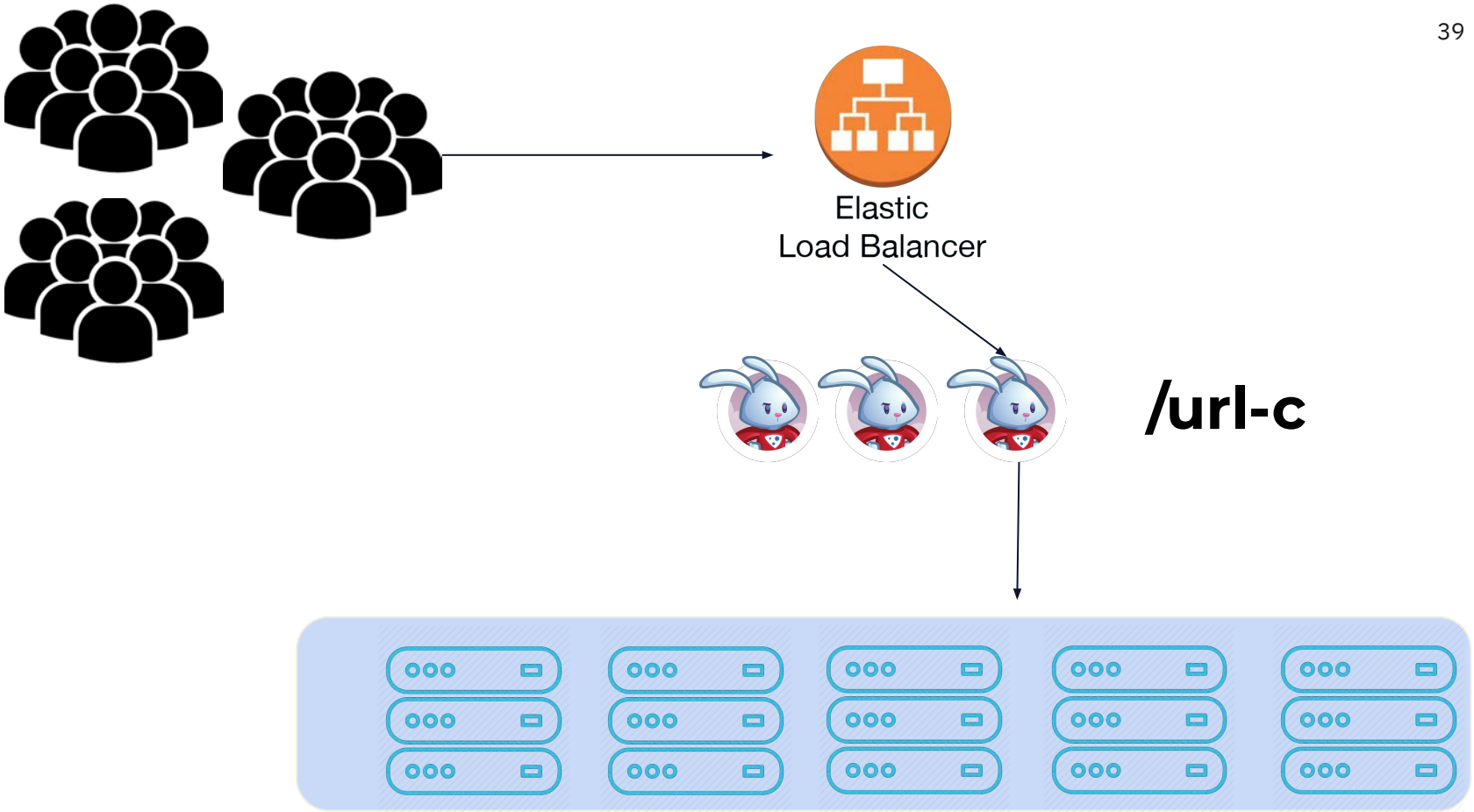

Elastic
Load Balancer

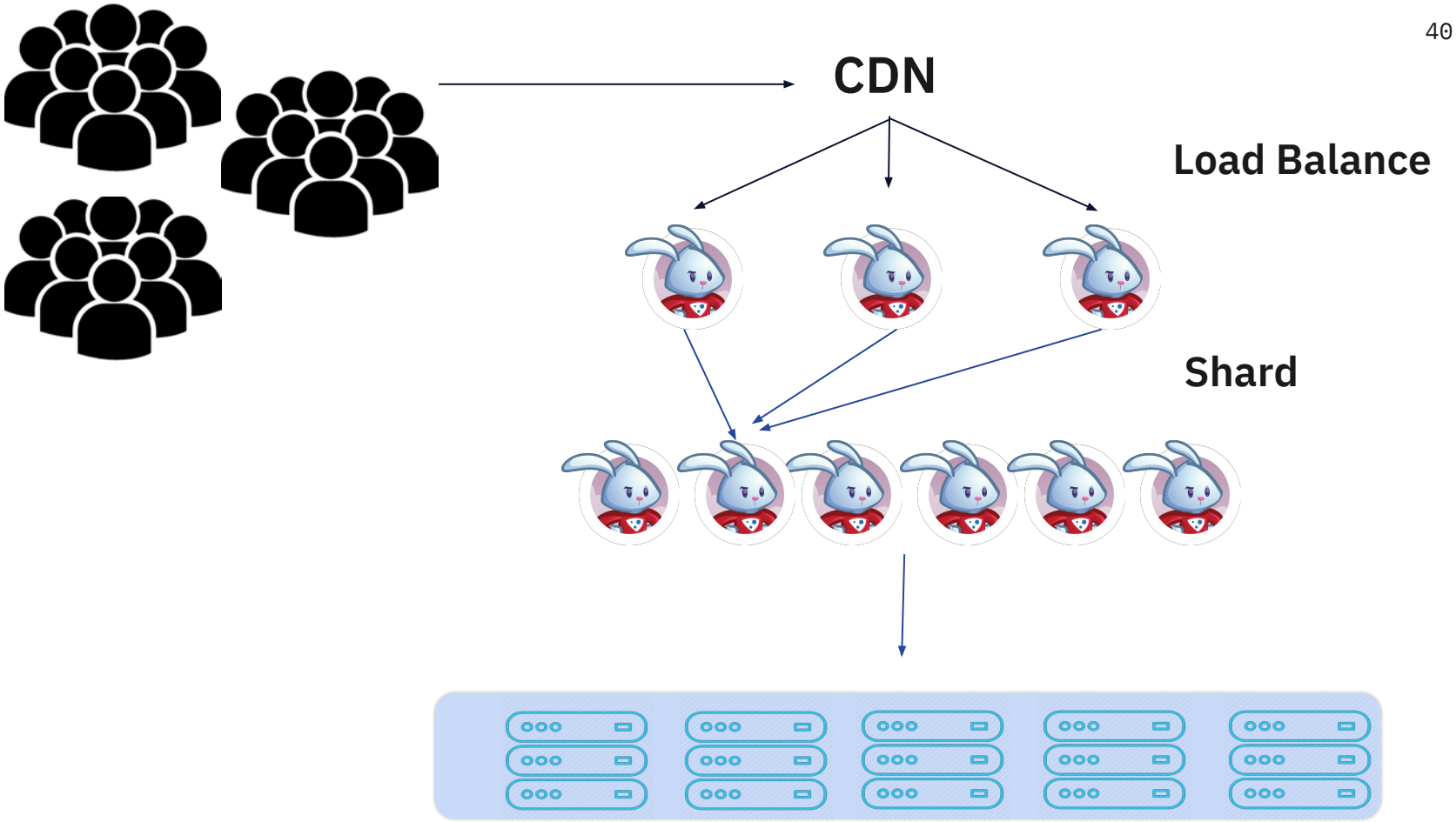

Elastic
Load Balancer

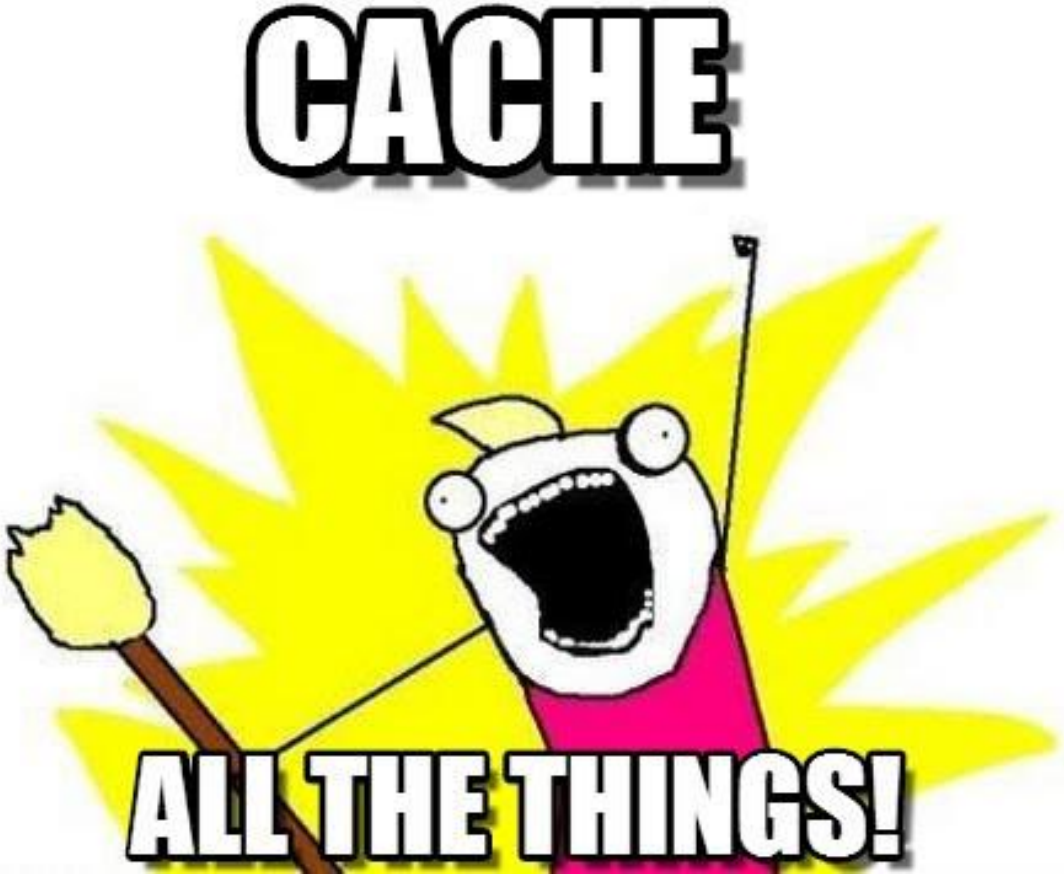

Elastic
Load Balancer


Elastic
Load Balancer
/url-a


Elastic
Load Balancer
/url-b


Elastic
Load Balancer
/url-c


CDN
Load Balance
Shard


CACHE
ALL THE THINGS!

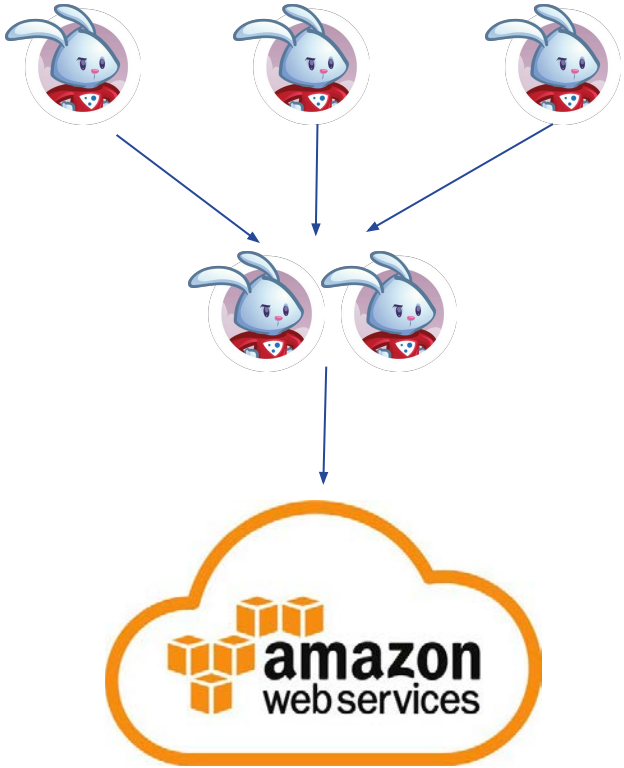
But note: cache invalidation is hard

Cache-Control header

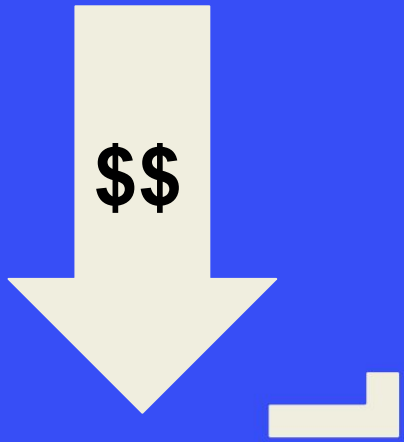
max-age: 60
s-maxage: 300
stale-while-revalidate: 3600
stale-if-error: 86400

amazon
web services

MB = $$$


amazon
web services

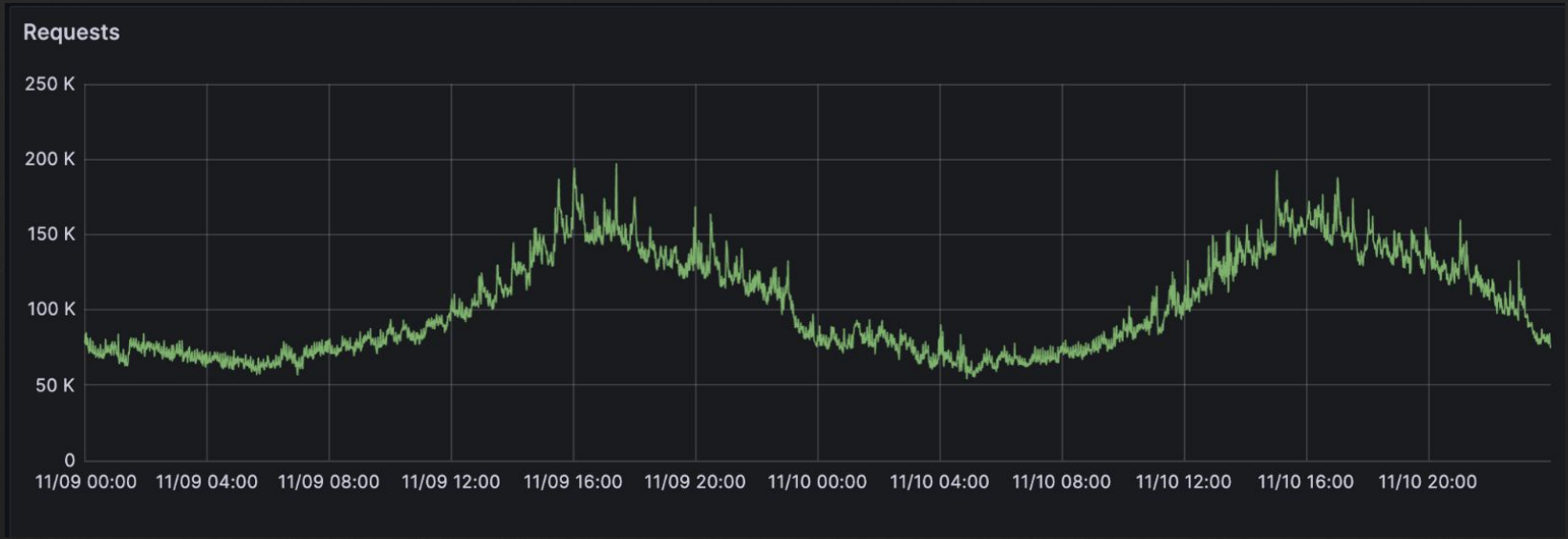
10x

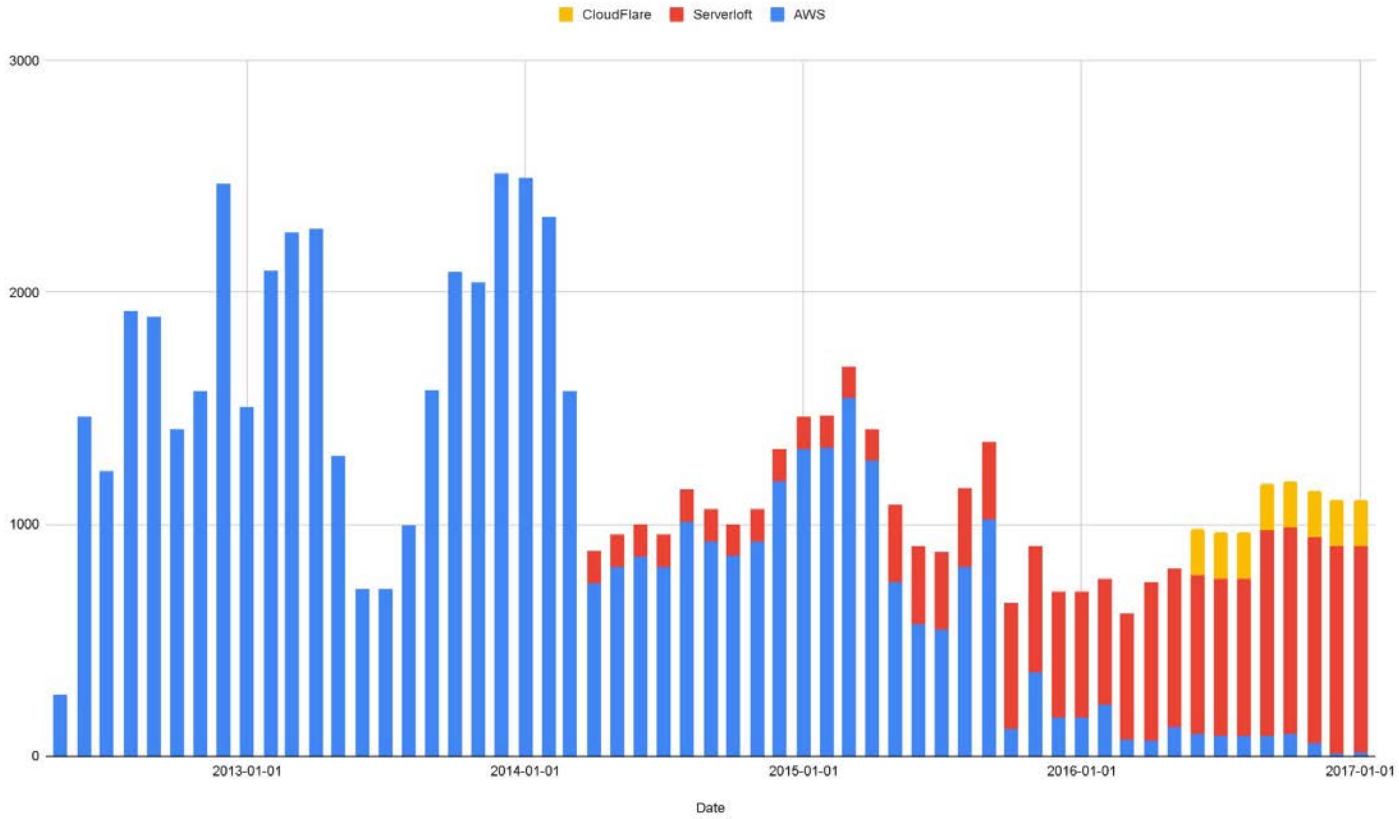


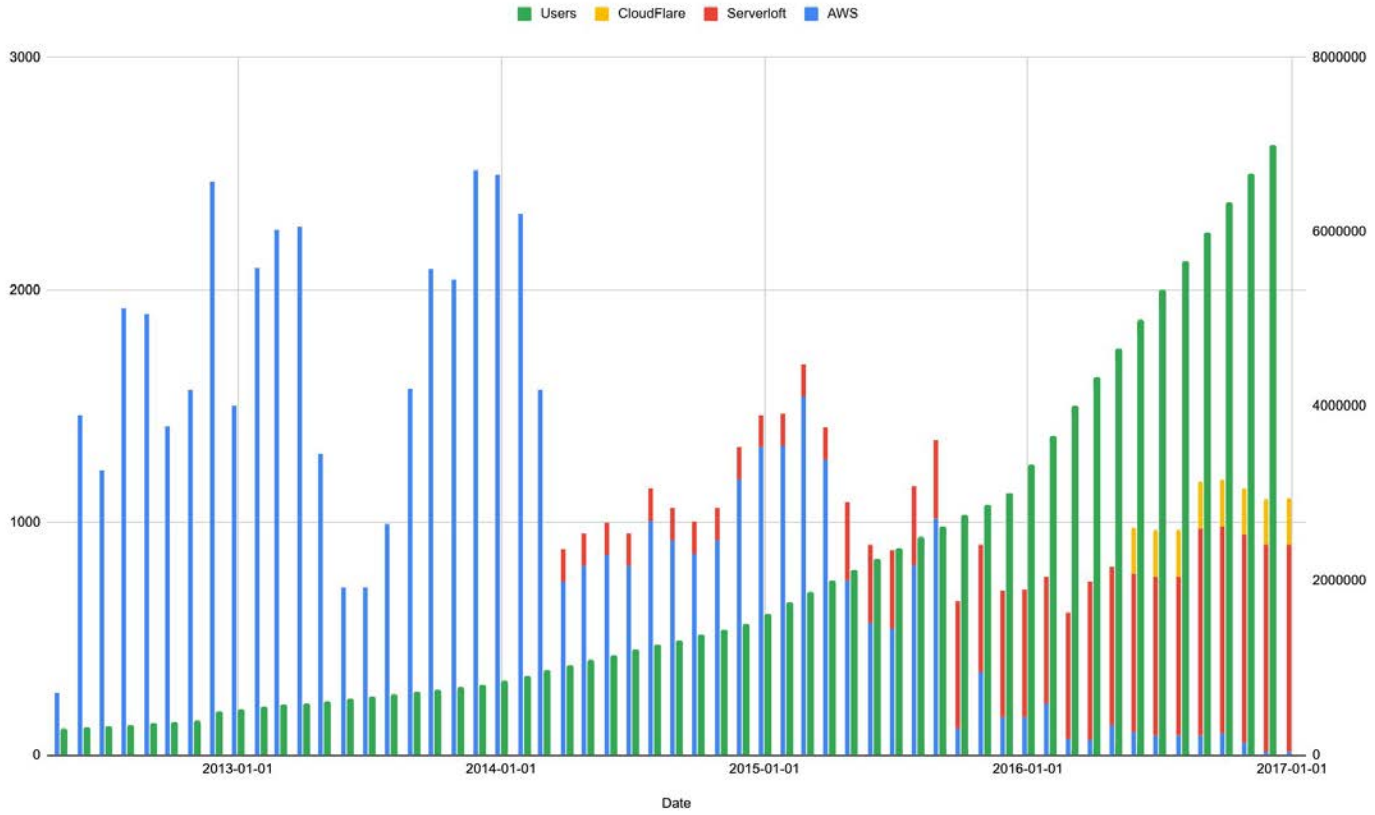
Sofascore

Migrating everything from cloud to DC

Sofascore


Requests
250 K
200 K
150 K
100 K
50 K
0
11/09 00:00
11/09 04:00
11/09 08:00
11/09 12:00
11/09 16:00
11/09 20:00
11/10 00:00
11/10 04:00
11/10 08:00
11/10 12:00
11/10 16:00
11/10 20:00


CloudFlare
Serverloft
AWS
3000
2000
1000
0
2013-01-01
2014-01-01
2015-01-01
2016-01-01
2017-01-01
Date


Users
CloudFlare
Serverloft
AWS
3000
2000
1000
0
8000000
6000000
4000000
2000000
0
2013-01-01
2014-01-01
2015-01-01
2016-01-01
2017-01-01
Date

Cost and usage Info

Current month costs

$1,684,974.20

Forecasted month end costs

$2,002,283.67 Up 446% over last month

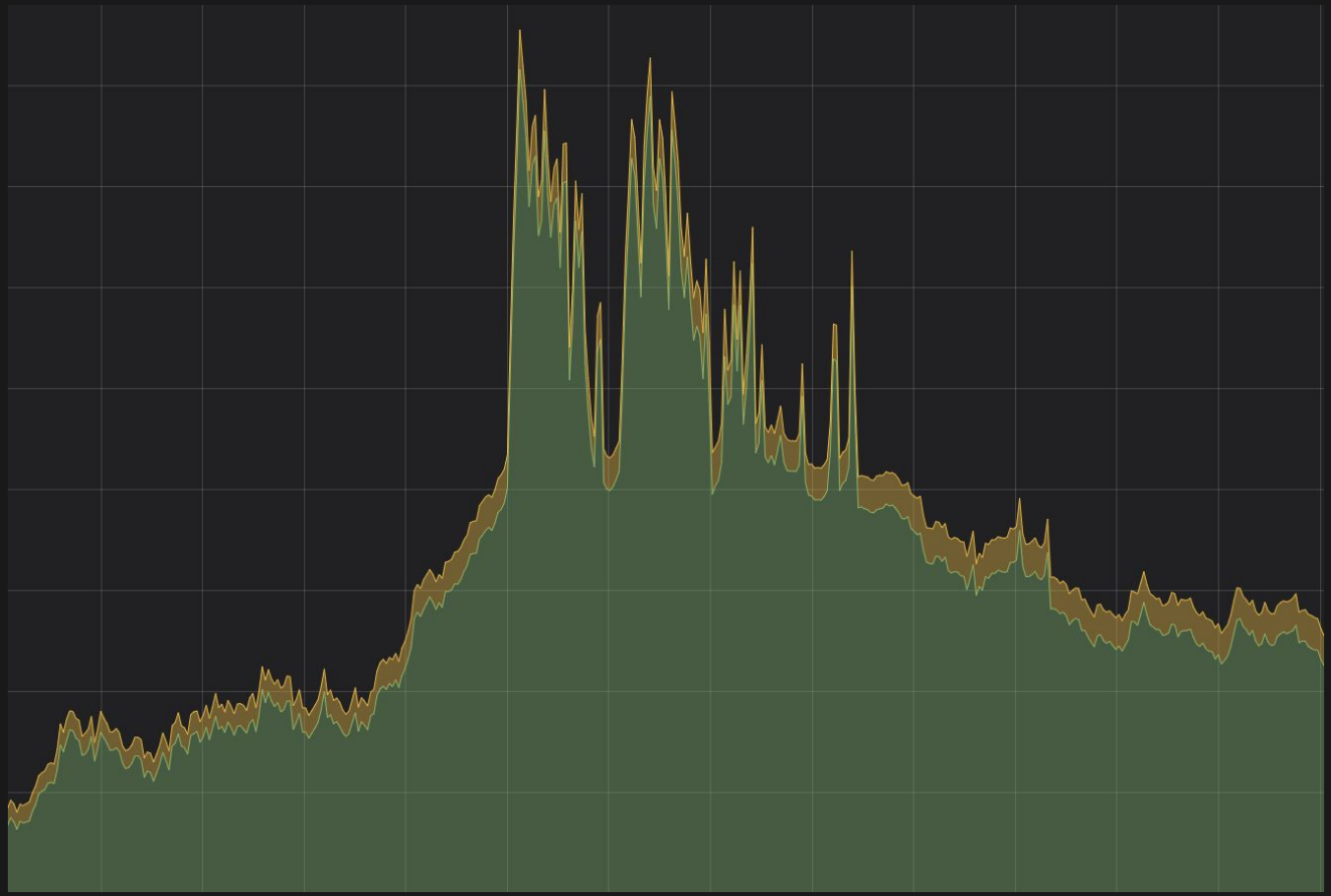
Last month costs

$366,573.53

Lot’s of spare RAM

Cache!

Cost of people?



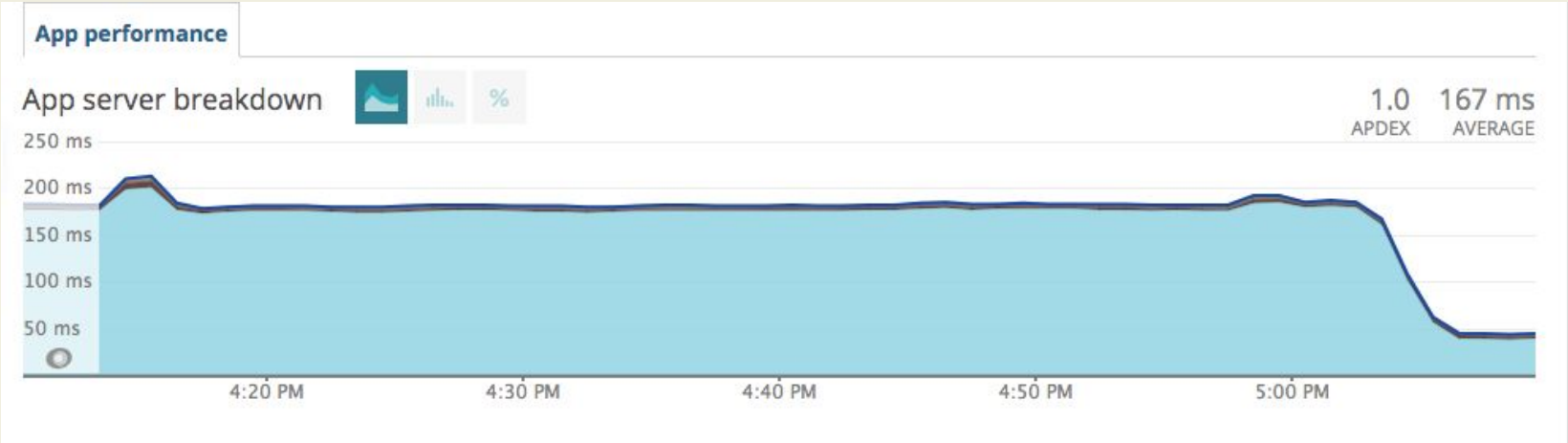
APM

Type Web

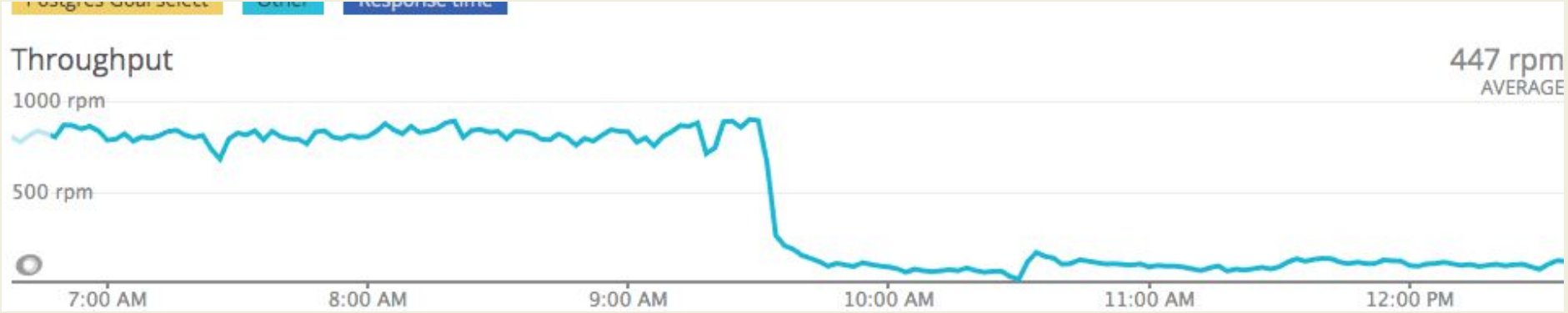
Most time consuming

/sofa_web_event_list_json	5.9%
/sofa_web_event_details_json	5.64%
/app_api_v1_user_events	4.65%
/app_api_v1_event_votes	3.84%
/sofascore_mobile_mobile4_event_head2head	3.79%
/sofa_web_team	3.48%
/app_api_v1_player_events	3.05%
/sofa_web_event_details	2.84%
/sofa_web_event_details_lineups_json	2.62%
/sofascore_mobile_mobile4_event_vote	2.55%
/sofa_web_event_details_amp	2.45%
/sofascore_mobile_mobile4_team_events	1.92%
/sofascore_mobile_mobile4_event_tweets	1.57%

Optimizations



Longer cache

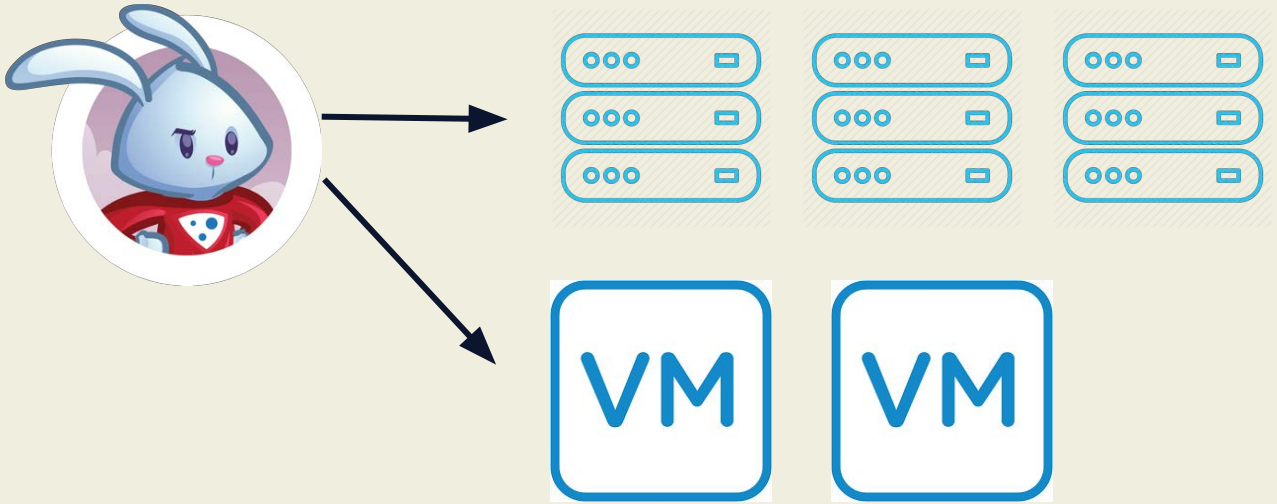


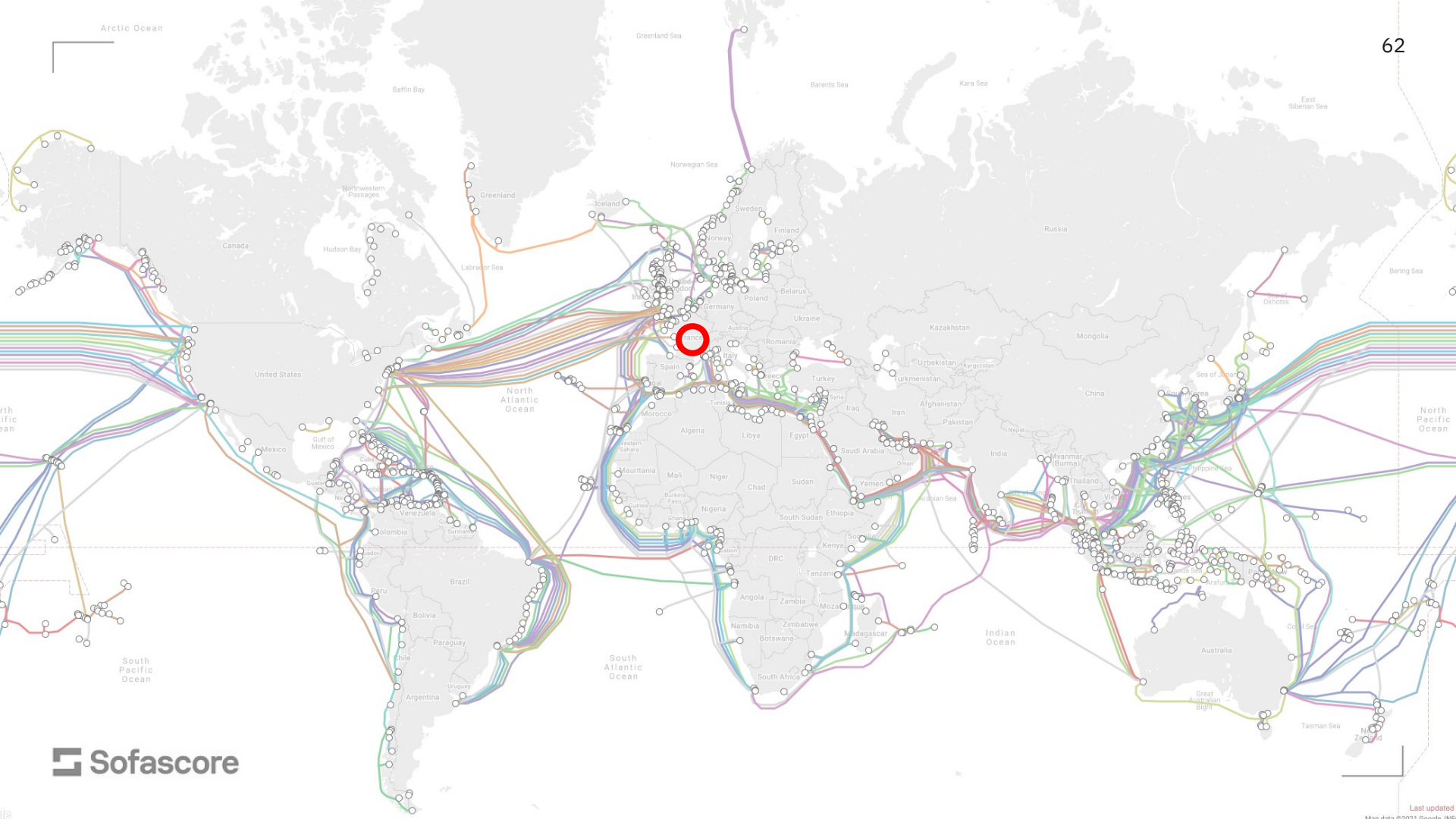
Monitor everything optimize something

Sofascore

What if the number of users exceeds current capacity?

Dedicated + VPS





Sofascore

Last updated

Josip Stuhli Nov 10, 2020, 10:25 PM

to [redacted]

Hi,

I hope this email finds you well.

I have been informed by other mods that you are from New Zealand. I'm interested in seeing what kind of user experience mobile users from countries outside Europe get using SofaScore.
Could you record your screen while using SofaScore app so I can see what kind of latency are you getting? You can open the app, go through live matches, open details, maybe some tabs, such as lineups and statistics. Maybe open up match list for Saturday/Sunday, team/player info and such.

I'm hoping to use that info to speed up the loading of the data in the app itself.

Thanks,
Josip Stuhli
SofaScore CTO

C [redacted]

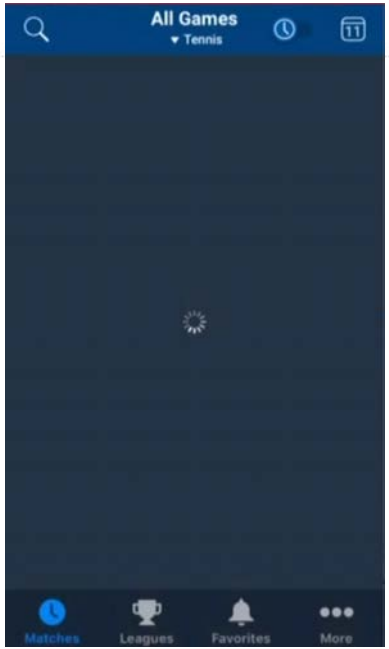
to me

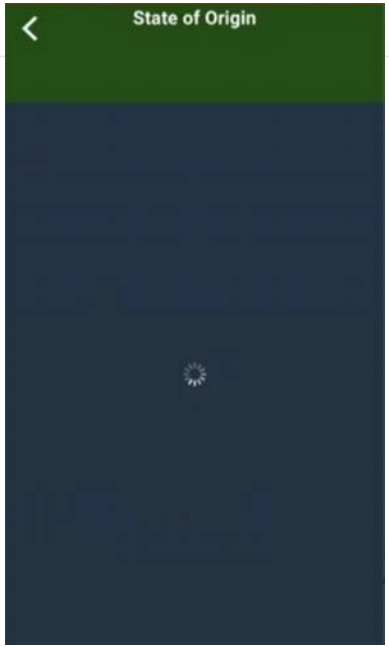
Yes mate I'm from New Zealand but living in Australia and everything seems to be working fine

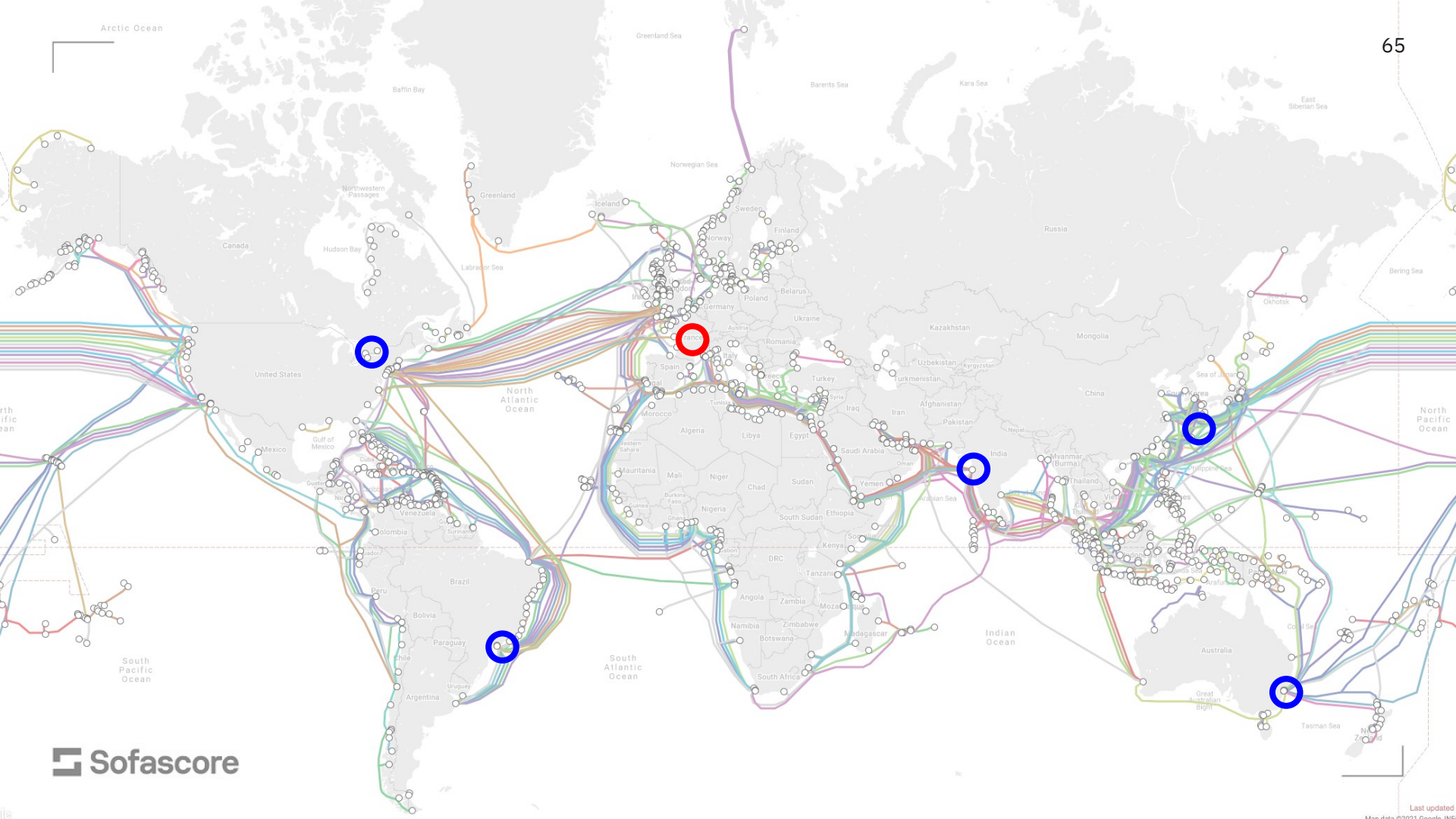
•••

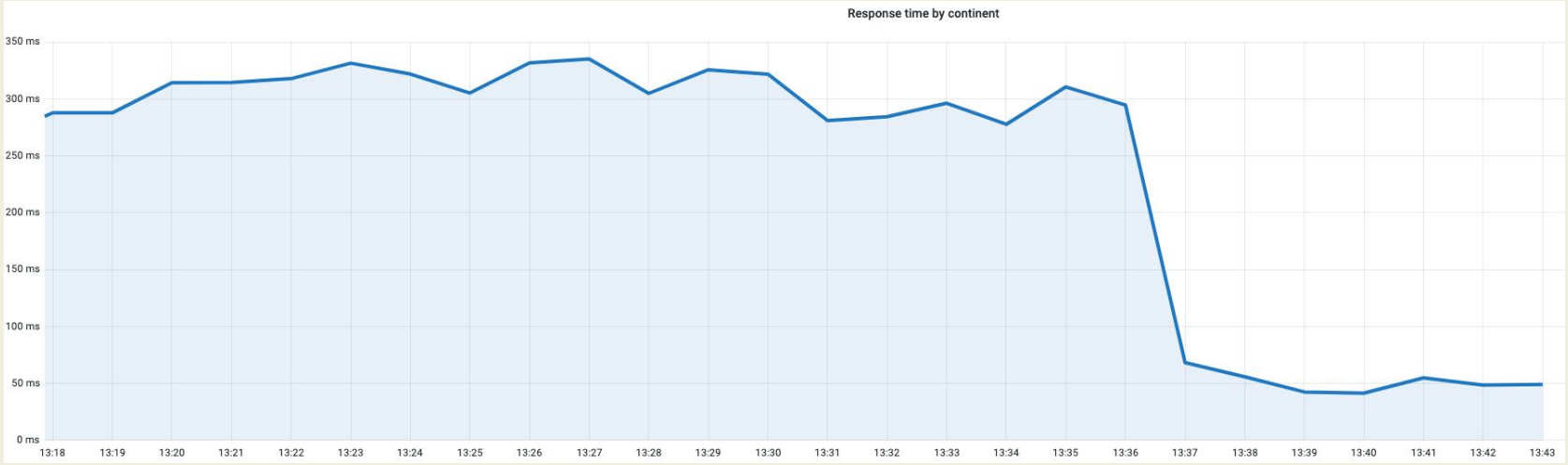
RPReplay_Final1605088002.mp4

14 MB





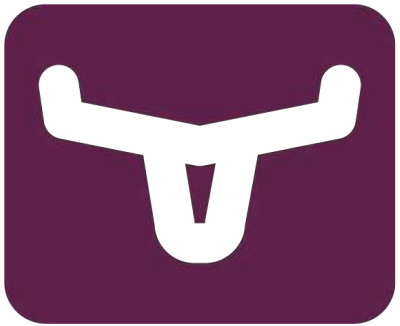

Arctic Ocean
Greenland Sea
Barents Sea
Kara Sea
East Siberian Sea
Baffin Bay
Norwegian Sea
Northwestern Passages
Greenland
Iceland
Sweden
Finland
Norway
Russia
Canada
Hudson Bay
Labrador Sea
Bering Sea
Belarus
Poland
Germany
Ukraine
Kazakhstan
Mongolia
Austria
Romania
Italy
Spain
Uzbekistan
Kyrgyzstan
Turkmenistan
Turkey
United States
North Atlantic Ocean
Morocco
Tunisia
Syria
Iraq
Iran
Afghanistan
Pakistan
China
Sea of Japan
South Korea
North Pacific Ocean
Algeria
Libya
Egypt
Saudi Arabia
Nepal
India
Mexico
Gulf of Mexico
Western Sahara
Mauritania
Mali
Niger
Chad
Sudan
Oman
Yemen
Myanmar (Burma)
Thailand
Philippine Sea
Burkina Faso
Nigeria
Ethiopia
South Sudan
Arabian Sea
Venezuela
Colombia
Kenya
DRC
Tanzania
Brazil
Peru
Bolivia
Paraguay
Chile
Uruguay
Argentina
Angola
Zambia
Namibia
Zimbabwe
Botswana
Madagascar
South Africa
Indian Ocean
South Pacific Ocean
South Atlantic Ocean
Australia
Great Australian Bight
Tasman Sea
Sofascore
Last updated


Response time by continent
350 ms
300 ms
250 ms
200 ms
150 ms
100 ms
50 ms
0 ms
13:18
13:19
13:20
13:21
13:22
13:23
13:24
13:25
13:26
13:27
13:28
13:29
13:30
13:31
13:32
13:33
13:34
13:35
13:36
13:37
13:38
13:39
13:40
13:41
13:42
13:43

Australia

Shouldn’t CDN do that?

kubernetes



LONGHORN

Real time updates

Sofascore

NATS Messaging

NATS is an open-source messaging system. The NATS server is written in the Go programming language. Client libraries to interface with the server are available for dozens of major programming languages. The core design principles of NATS are performance, scalability, and ease of use. Wikipedia

Developer(s): Synadia Communications, Inc.

Initial release: 2011

License: Apache 2.0

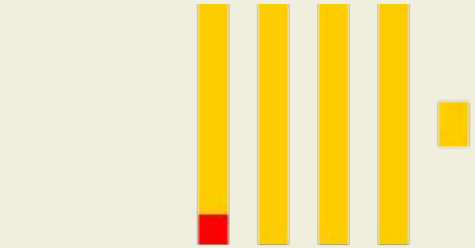
Stable release: 2.8.2 / 4 May 2022; 6 days ago

Written in: Go

500k+ msgs/sec

What to do with 2 PB of data?

Sofascore



ClickHouse

+

1.75 PB -> 46.8TB
1.4 Trillion rows

Insert:

rows_per_sec	2.15 Mil

ClientIP	count()	uniq_methods	uniq_user_agent	uniq_routes	min(EdgeStartTimes	max(EdgeStartTimes
203.161.33.185	9382	1	1	14	2024-11-19 17:31:13	2024-11-19 18:30:12
185.51.61.123	5151	1	1	3	2024-11-19 18:10:17	2024-11-19 18:27:37
192.64.113.199	3706	1	1	3	2024-11-19 17:31:00	2024-11-19 18:29:19
223.19.209.2	2725	1	2	31	2024-11-19 17:31:08	2024-11-19 18:29:32
46.134.139.172	2643	1	1	21	2024-11-19 17:31:00	2024-11-19 17:42:41
2a02:8070:6482:7...	2335	2	3	6	2024-11-19 18:26:16	2024-11-19 18:29:33
41.113.128.118	2221	1	1	40	2024-11-19 17:32:57	2024-11-19 18:29:29
203.175.9.145	1840	1	234	10	2024-11-19 17:31:11	2024-11-19 18:29:25
36.134.84.234	1591	1	1	6	2024-11-19 17:31:03	2024-11-19 18:29:04
98.201.93.87	1494	2	1	43	2024-11-19 17:38:28	2024-11-19 18:29:16

Clickhouse + Superset

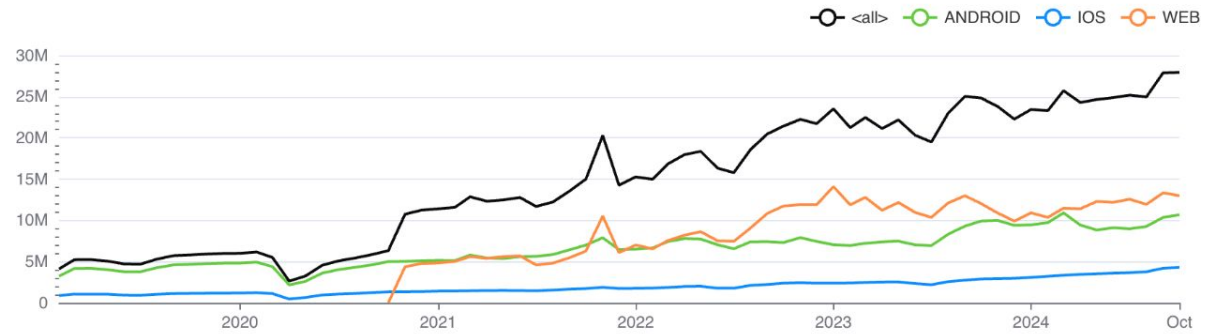
MAU YoY

October 2024

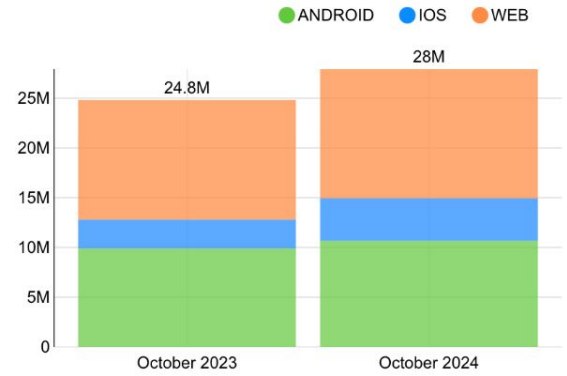
28M

+12.5% compared to previous year

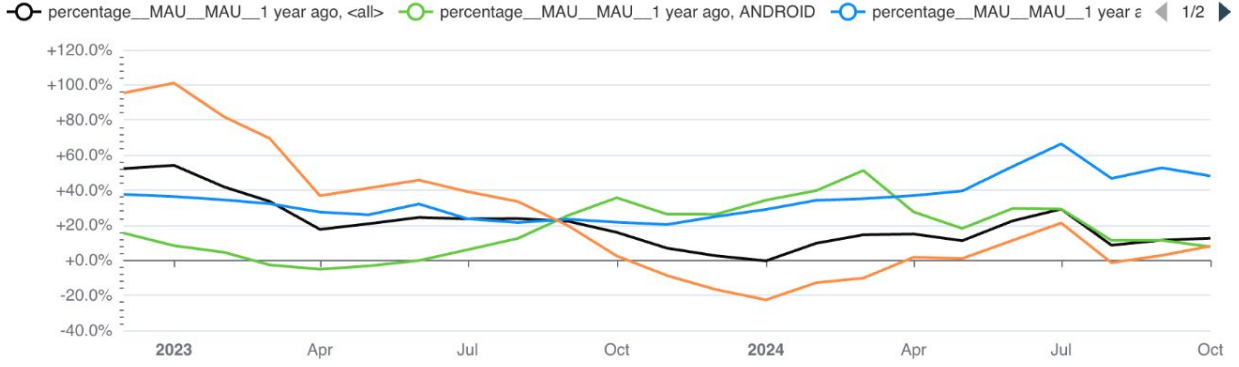
MAU Over Time



MAU YoY Composition



MAU YoY Change Over Time



Active users by First user source

#1 (direct)
119,891
66.38%

FIRST USER SOURCE	ACTIVE USERS
(direct)	120K
google	37K
google-play	17K
Apple	5K

Active users by Audience

#1 All Users
15,382
11.44%

AUDIENCE	ACTIVE USERS
All Users	15K
Growth - Android use...	15K
ads_click custom les...	14K
Open_event (Do) GA4	11K

Views by Page title and screen name

#1 Football - scor...ds | Sofascore
10,080
18%

PAGE TITLE AND S...	VIEWS
Football - scores, sch...	10K
Crystal Palace vs Ma...	2.6K
Calcio - risultati, cale...	2.1K
Jogos de Futebol de...	1.9K

Event count by Event name

#1 screen_view_custom
41,157,034
32.72%

EVENT NAME	EVENT COUNT
screen_view_custom	41M
ads_impression_cust...	15M
ad_impression	12M
ads_value	12M

Infra cost is 0.6% of revenue

Take home

- Cache all the things!
- Be stateless where possible
- Know how users interact with the app
- Recognize the right tool for the job and change if necessary
- Cloud should be the default, but keep your options open
- Queue everything that is slow
- Monitor everything - optimize something
- Keep it simple for as long as you can

Thank you!

https://www.linkedin.com/in/jstuhli/

Sofascore