AVVS Summits 2014

Dynamic Content Acceleration:
Lightning-Fast Web Apps with
Amazon CloudFront and
Amazon Route 53

Constantin Gonzalez, Solutions Architect Amazon Web Services Germany GmbH





Any Web Application Must Have...

- Tight Security
- High Availability
- High Performance



Why Does Performance Matter?

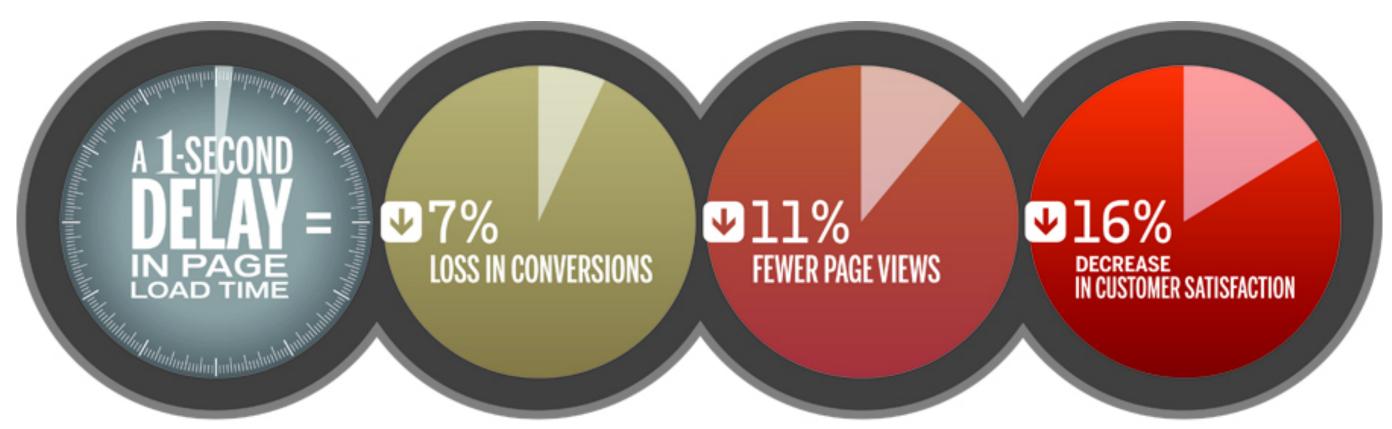
- More Page Views
- Better Customer Experience
- Higher Conversion Rates







How Much Does Performance Matter?





IN DOLLAR TERMS,

this means that if your site typically earns \$100,000 a day, this year

you could lose \$2.5 MILLION in sales.

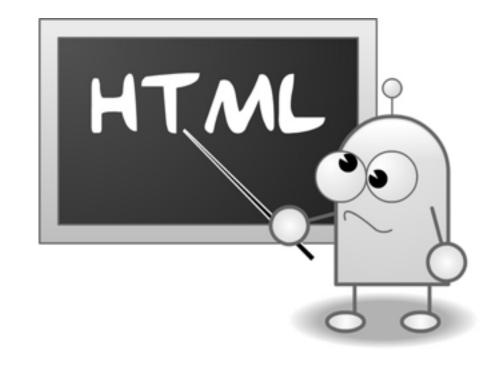






Where Does Performance Matter?

of the end user latency comes from the front-end





How to Improve Performance?

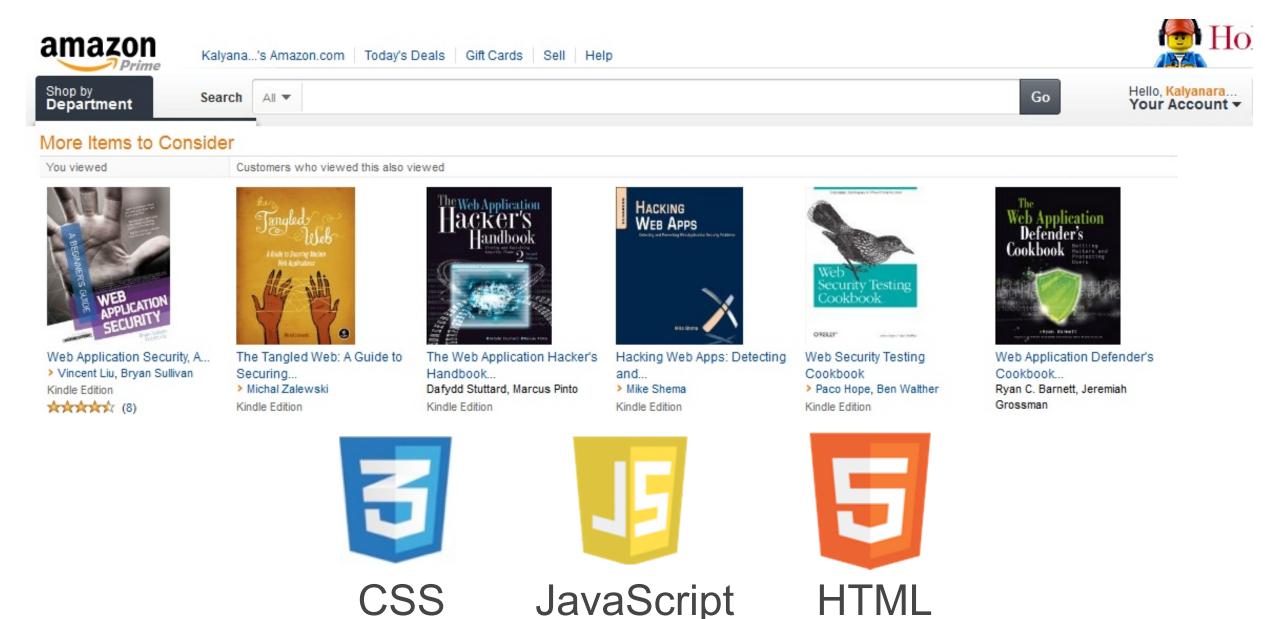
A Typical Web Application Has ...

- Static or Re-Usable Content
- Dynamic or Unique Content



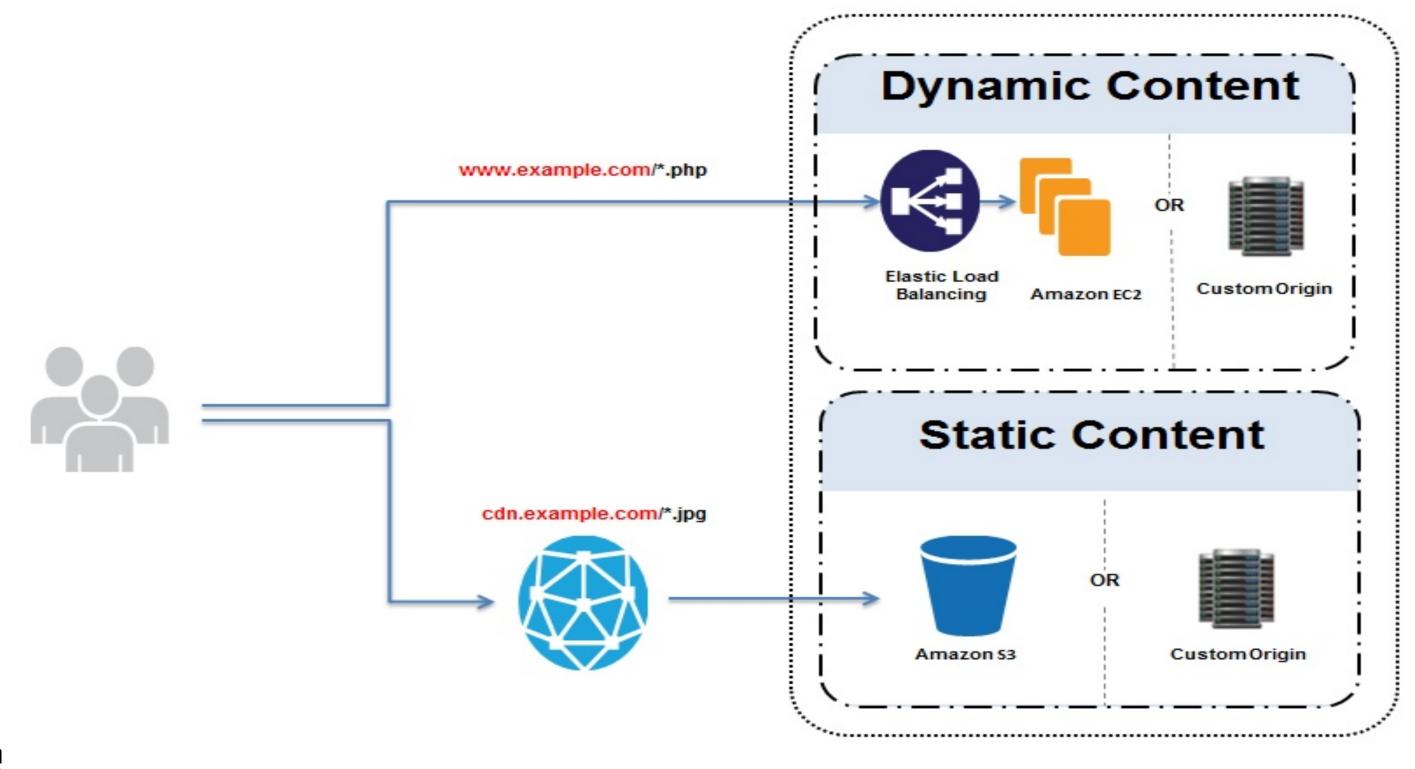
Static or Re-Usable Content

Can be cached (high TTLs or low TTLs)





Typical Architecture





Dynamic or Unique Content

Cannot be cached – but affects 100% of your viewers!



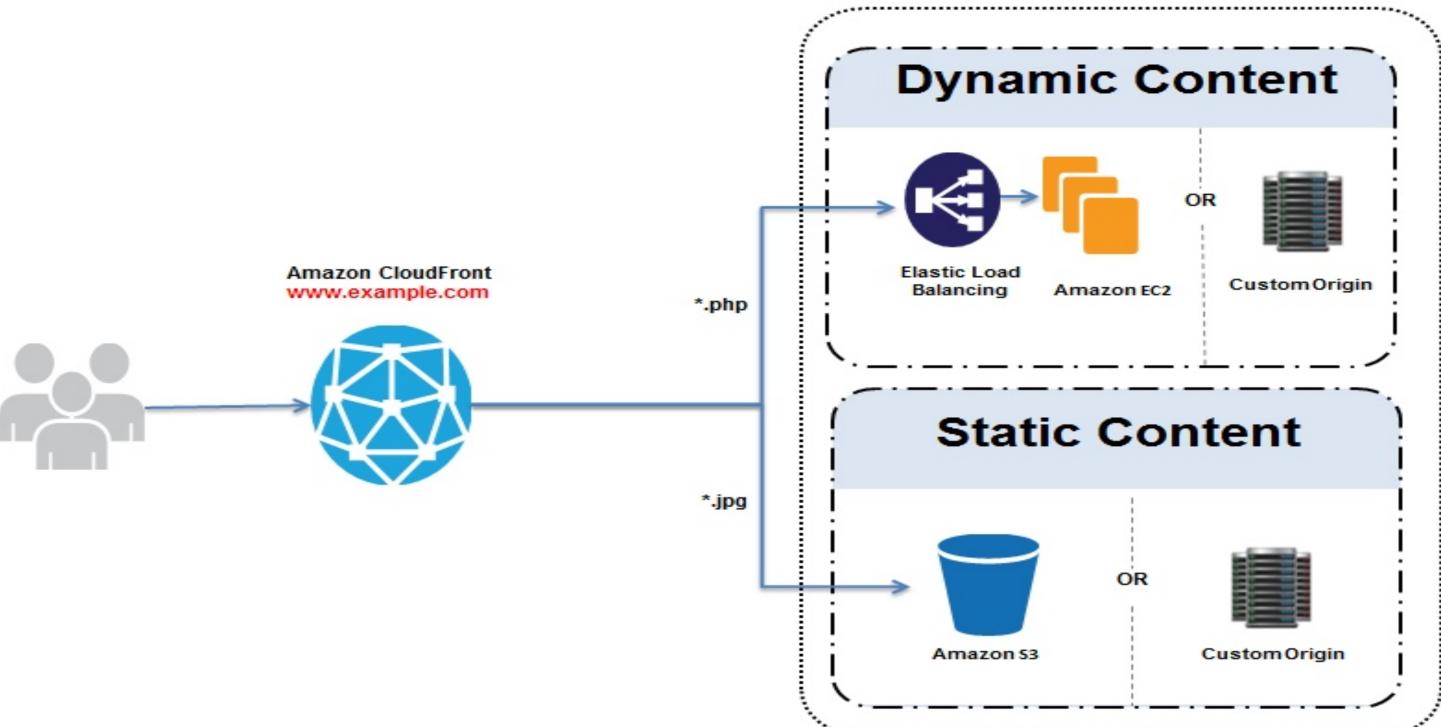








Why Not...?





How Can Amazon CloudFront Help?

- TCP/IP optimizations for the network path
- Keep-Alive connections to reduce RTT
- SSL termination close to viewers
- POST/PUT upload optimizations
- Latency Based Routing
- Low prices, same as static content delivery!







Unique or Reusable Content?





Static or Reusable

Content that does not change for a given period of time





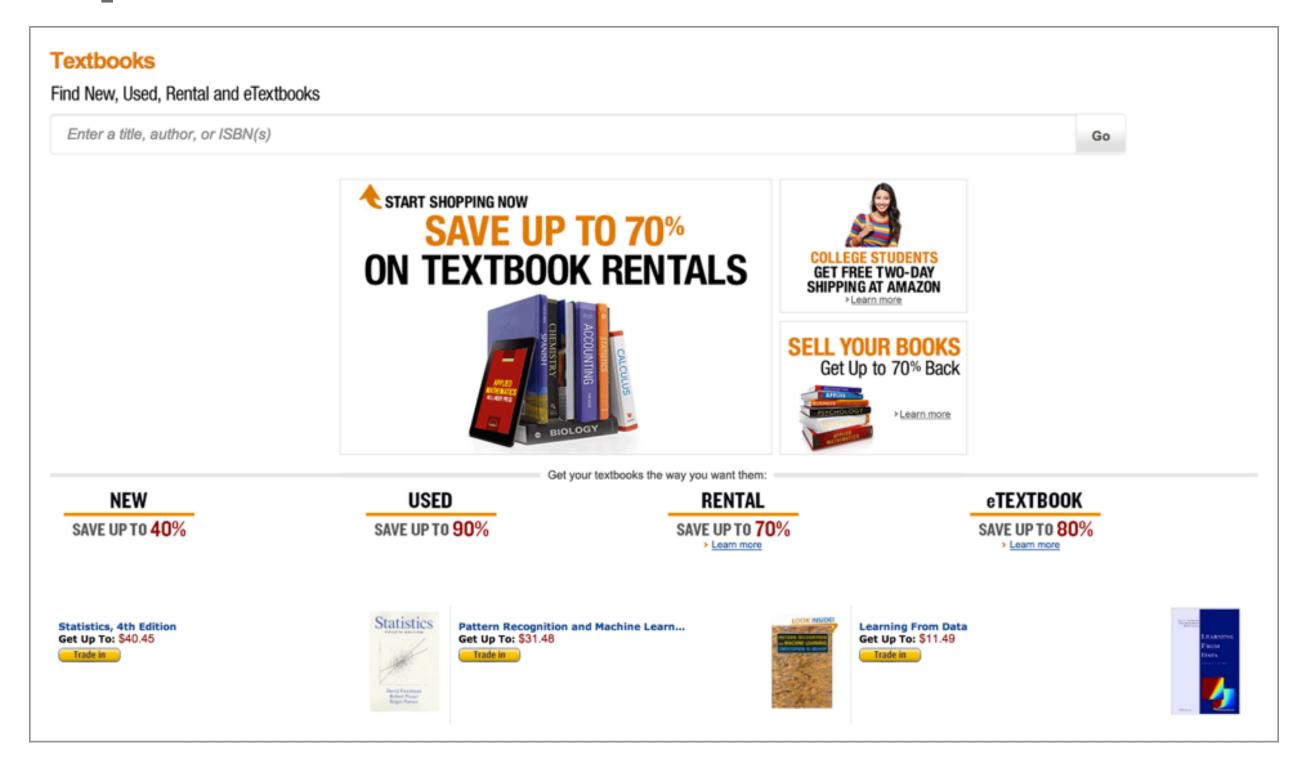
Dynamic or Unique

Content that changes as soon as it is created





Example





Example

Textbooks Index.jsp (dynamic) Find New, Used, Rental and eTextbooks Enter a title, author, or ISBN(s) Go **START SHOPPING NOW SAVE UP TO 70% ON TEXTBOOK RENTALS GET FREE TWO-DAY** SHIPPING AT AMAZON **SELL YOUR BOOKS** Get Up to 70% Back Images (static) >Learn more Get your textbooks the way you want them: NEW USED eTEXTBOOK RENTAL **SAVE UP TO 40% SAVE UP TO 90% SAVE UP TO 70% SAVE UP TO 80%** Learn more Learn more

Statistics, 4th Edition

Get Up To: \$40.45

Pattern Recognition and Machine Learn...

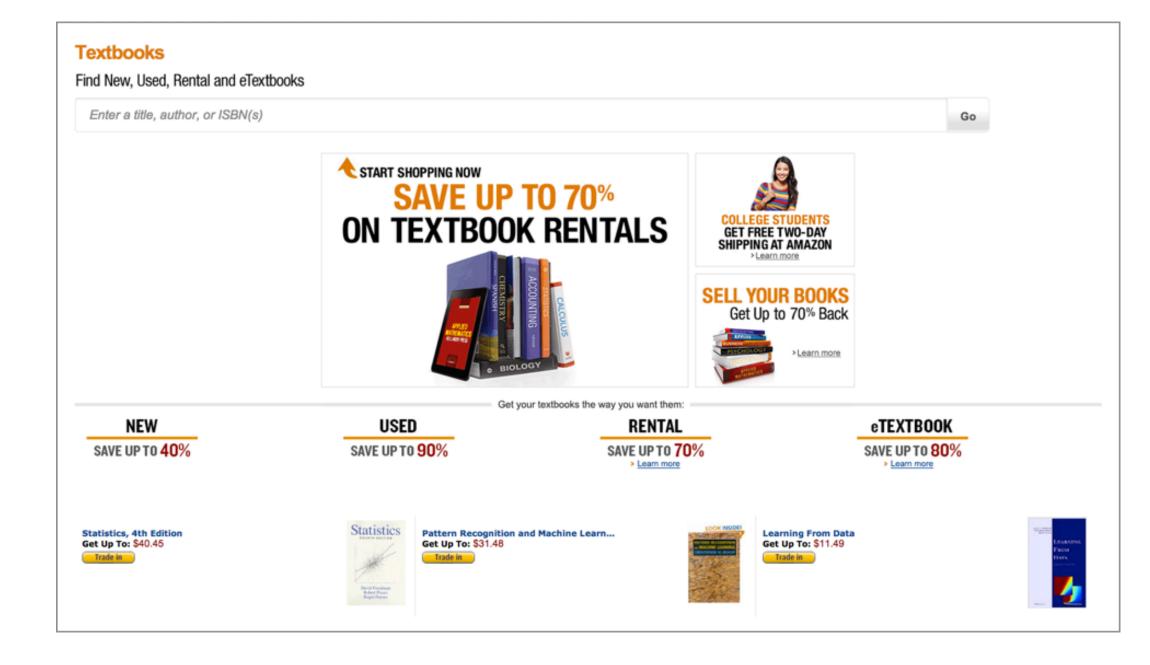
Learning From Data

Get Up To: \$11.49



Example

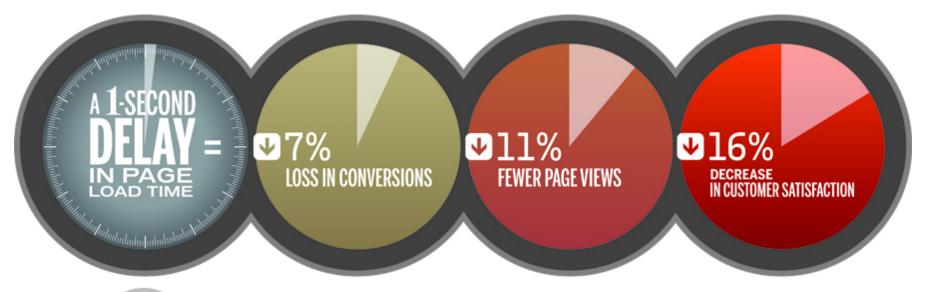






Page Load Time?





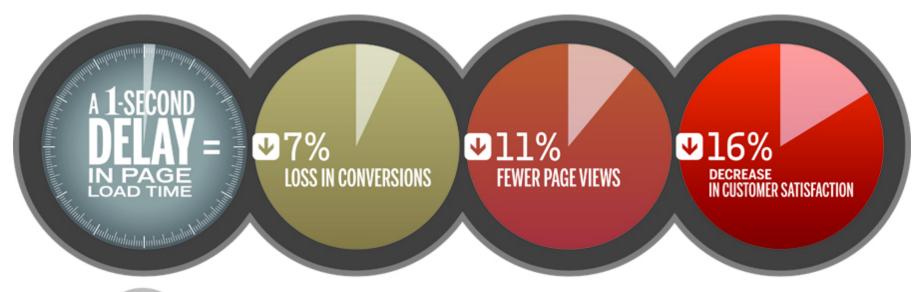






Goal:



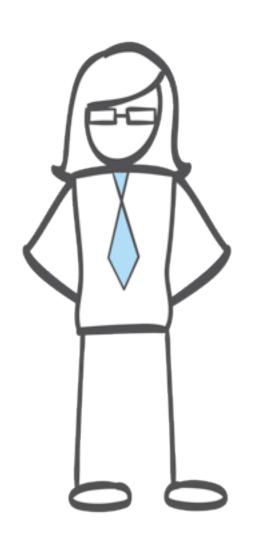




strangeloop www.strangeloopnetworks.com

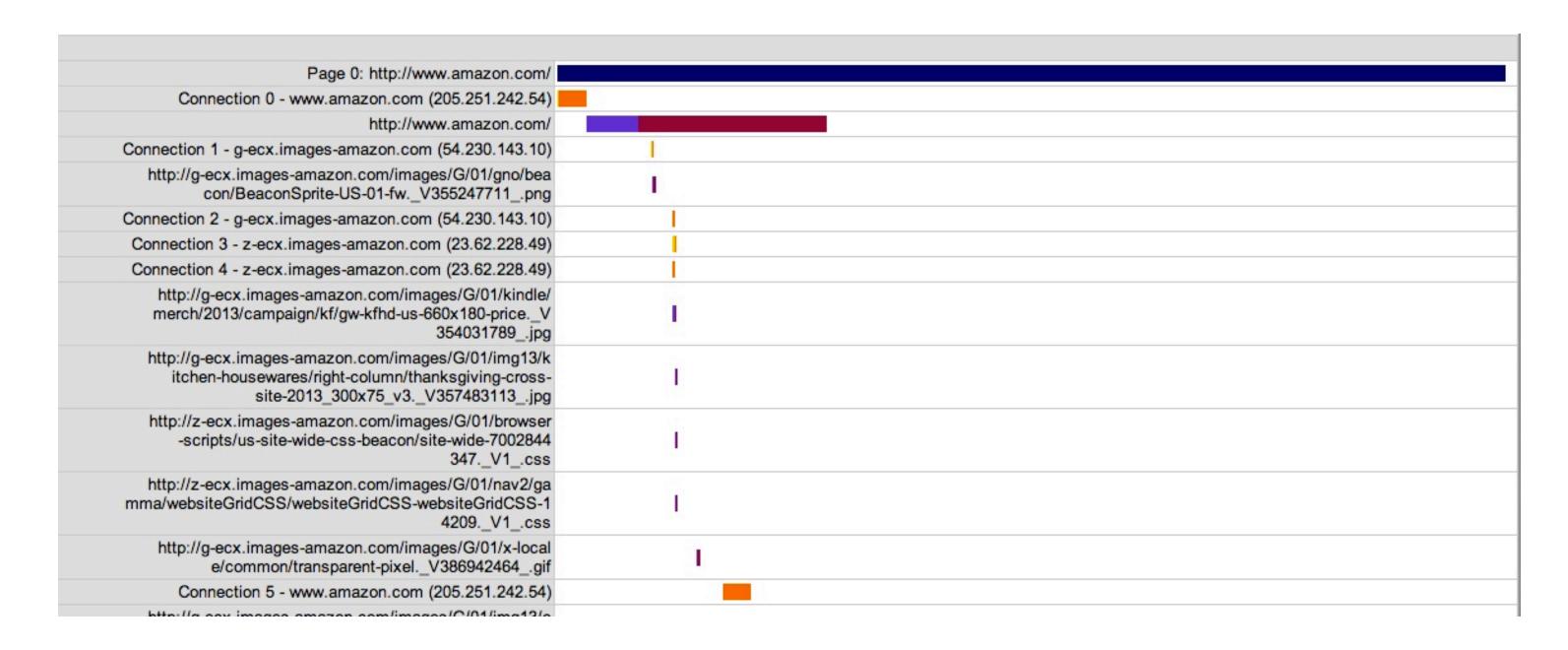


Introduction to Waterfall Graphs





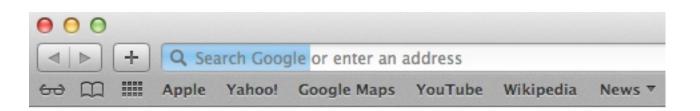
Waterfall Graphs

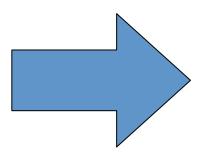




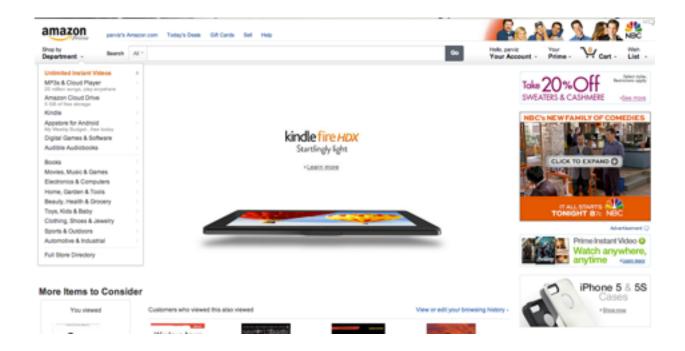
What Happens?

Typing the address



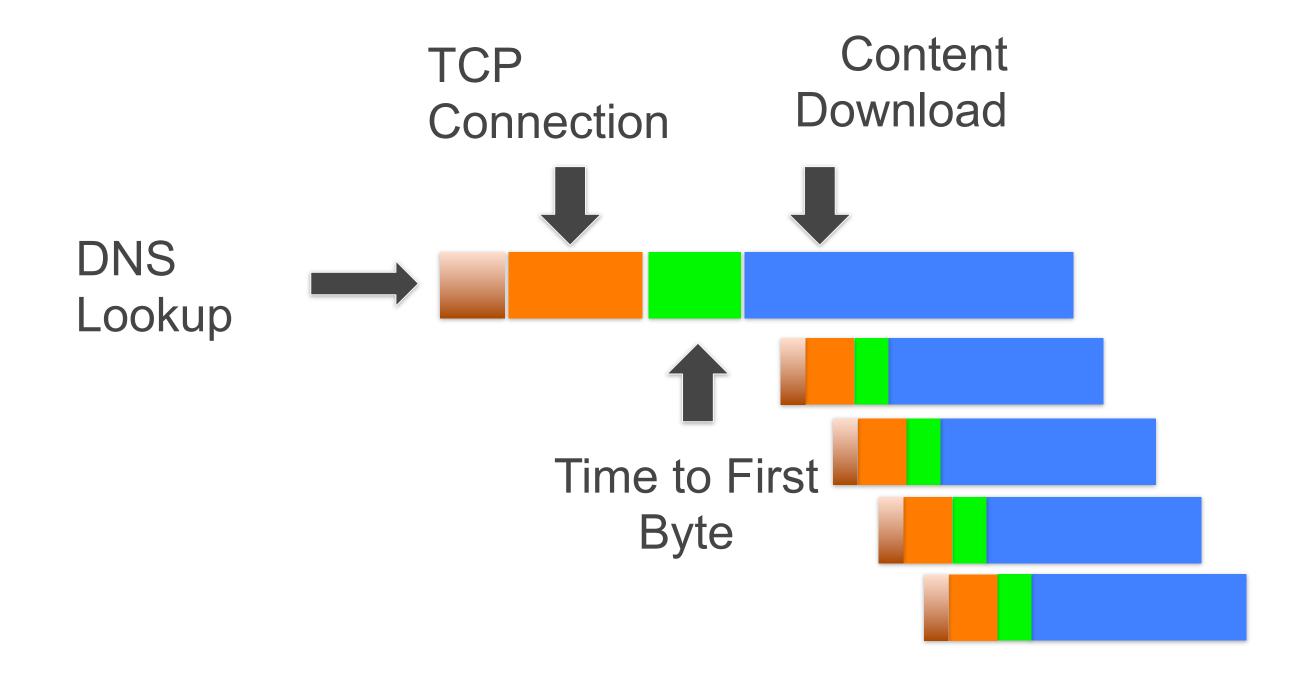


Browser renders



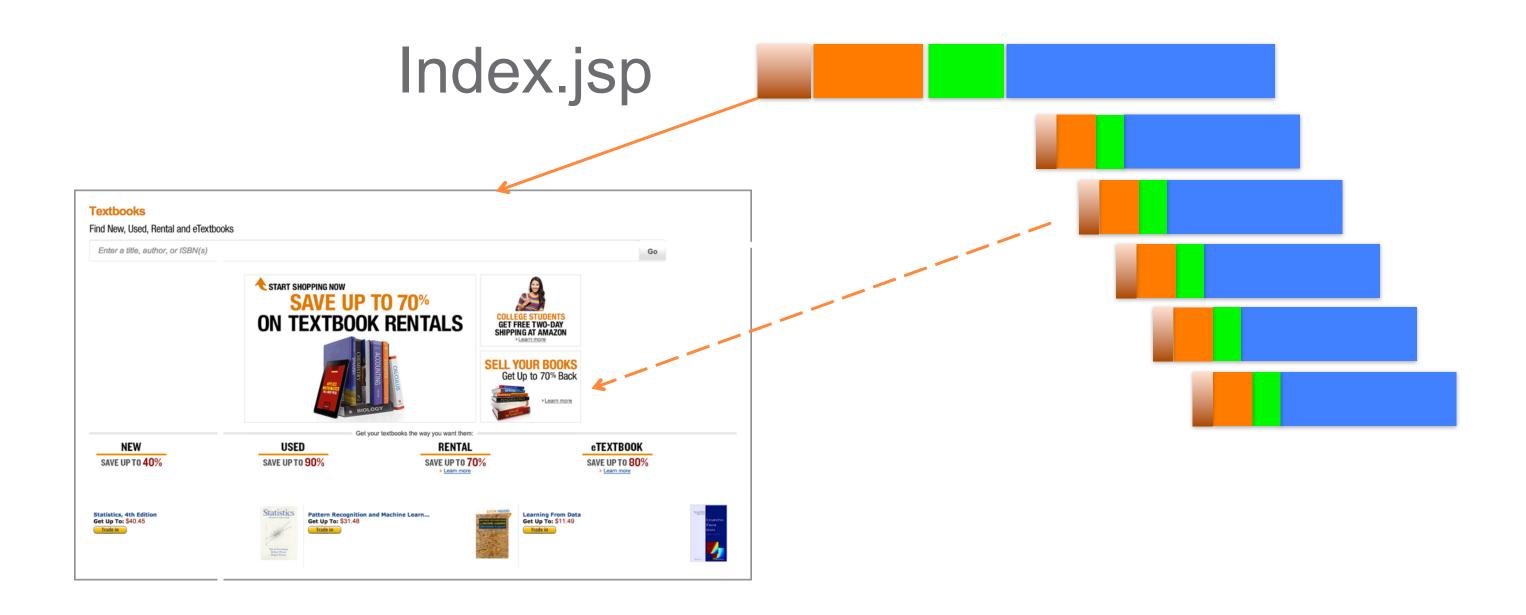


Understanding Waterfall Graphs





Understanding Waterfall Graphs

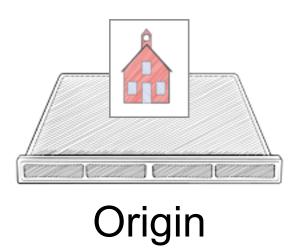




Optimizing Static Content





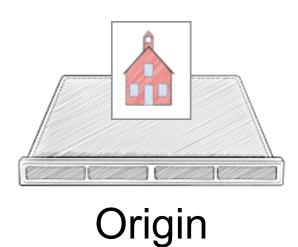






User Request A





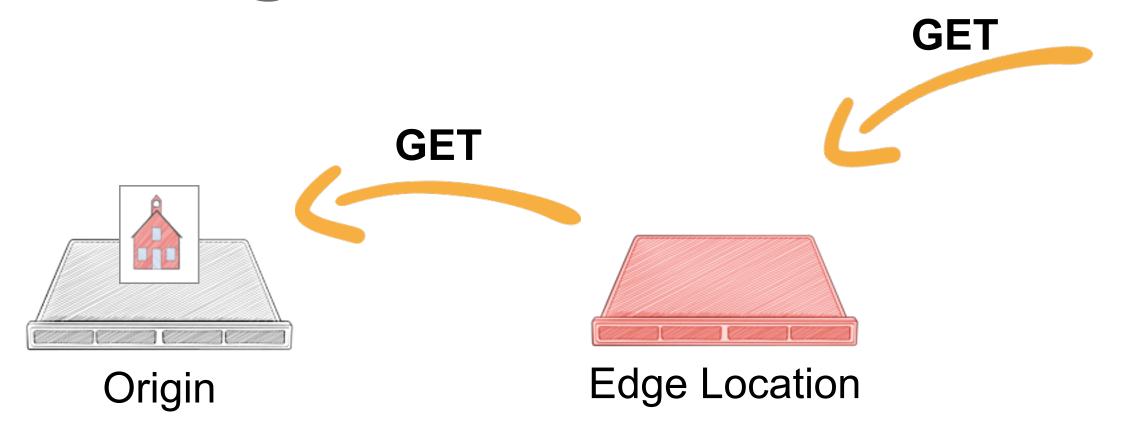






User Request A

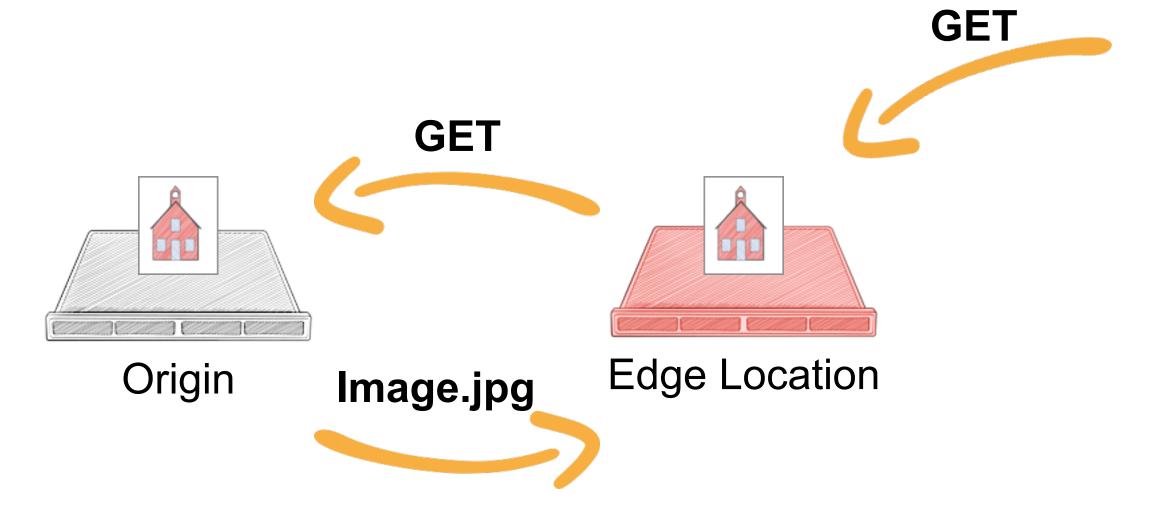






User Request A

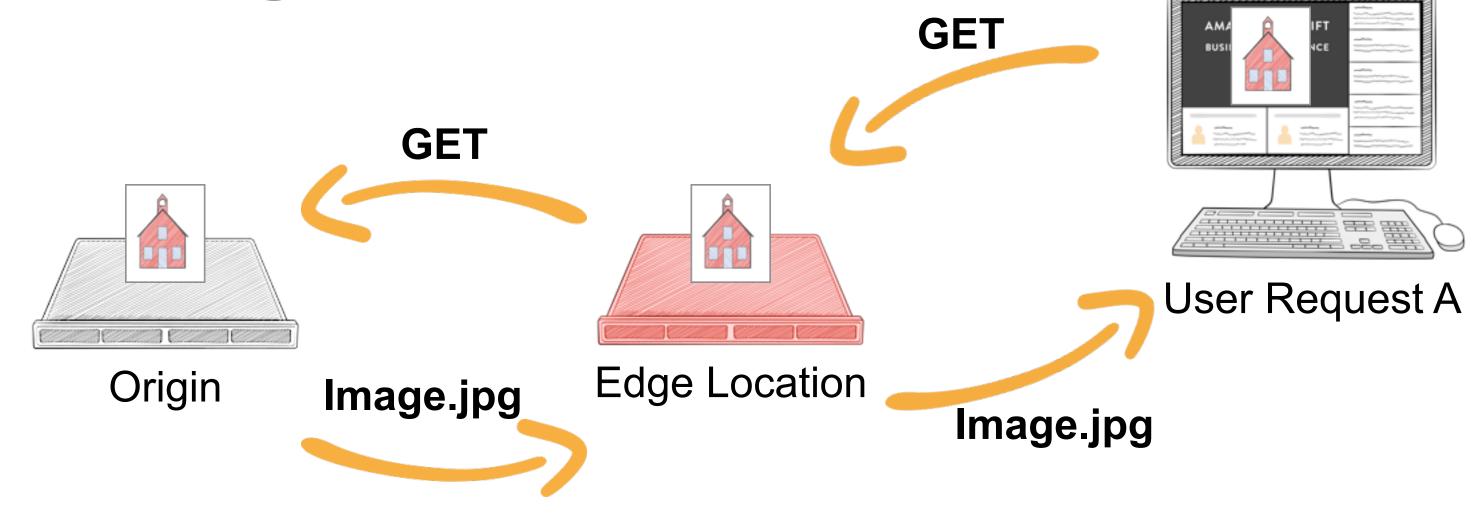




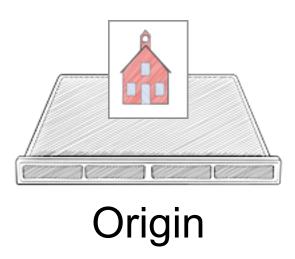


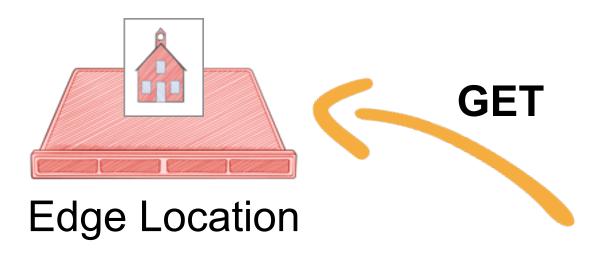
User Request A







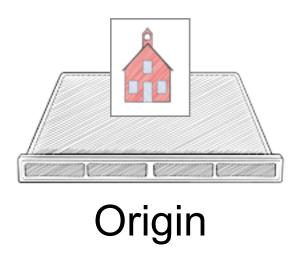


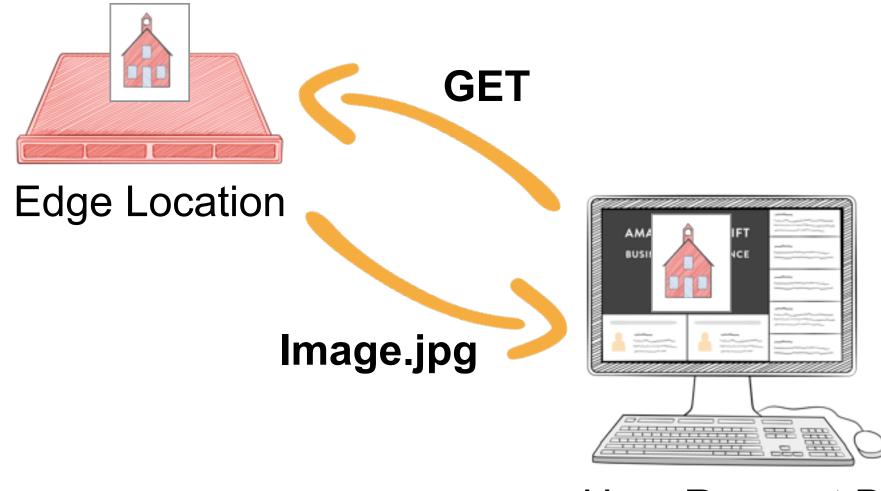




User Request B







User Request B



Optimizing Static Content with Caching

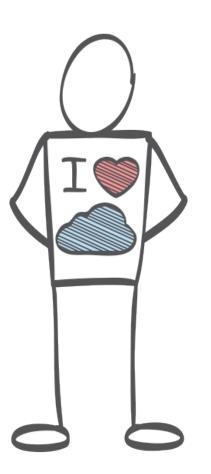
Brings Content Closer to Your Users





Optimizing Static Content with Caching

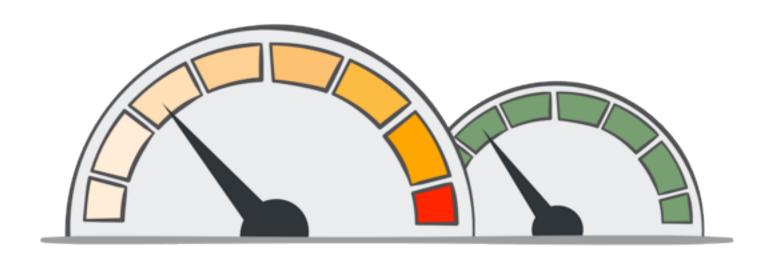
- Brings Content Closer to Your Users
- Improves Experience and Performance





Optimizing Static Content with Caching

- Brings Content Closer to Your Users
- Improves Experience and Performance
- Offloads Your Infrastructure





THE BERLINER PHILHARMONIKER'S DIGITAL CONCERT HALL

Alexander D. McWilliam
Director Software Development
Berlin Phil Media GmbH

















LIVE-STREAMS

CONCERTS

MOVIES

INTERVIEWS

HOW IT WORKS

SIGN UP

The best classical music experience you can have outside a concert hall.

Watch and listen to classical music's finest conductors and soloists performing with the Berliner Philharmoniker live as it happens or on-demand as it suits you.



WATCH FREE CONCERT NOW

With over 40 live concerts performed every season we offer more classical music than you could possibly watch.

soloist, conductor, composer



Q SEARCH

Find out if the Berliner Philharmoniker already played with your favorite artist.





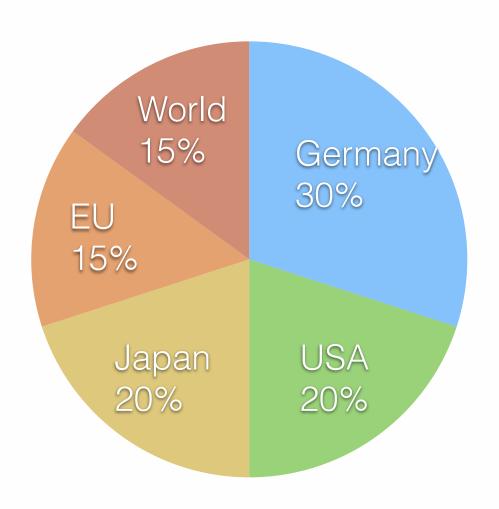






Our Audience

- 400,000 registered users
- from 100 different countries
- aged 18 to 88





Our Content

- 8 bit rates up to 2,500 kbps
- 100 terabytes of video traffic per month
- 2.5 terabytes of video library storage
- Delivery via HLS and progressive download



Other CDNs vs. Cloudfront

- Always 12-month up front commitments
- Insufficient caching performance
- Black box between us and our customers

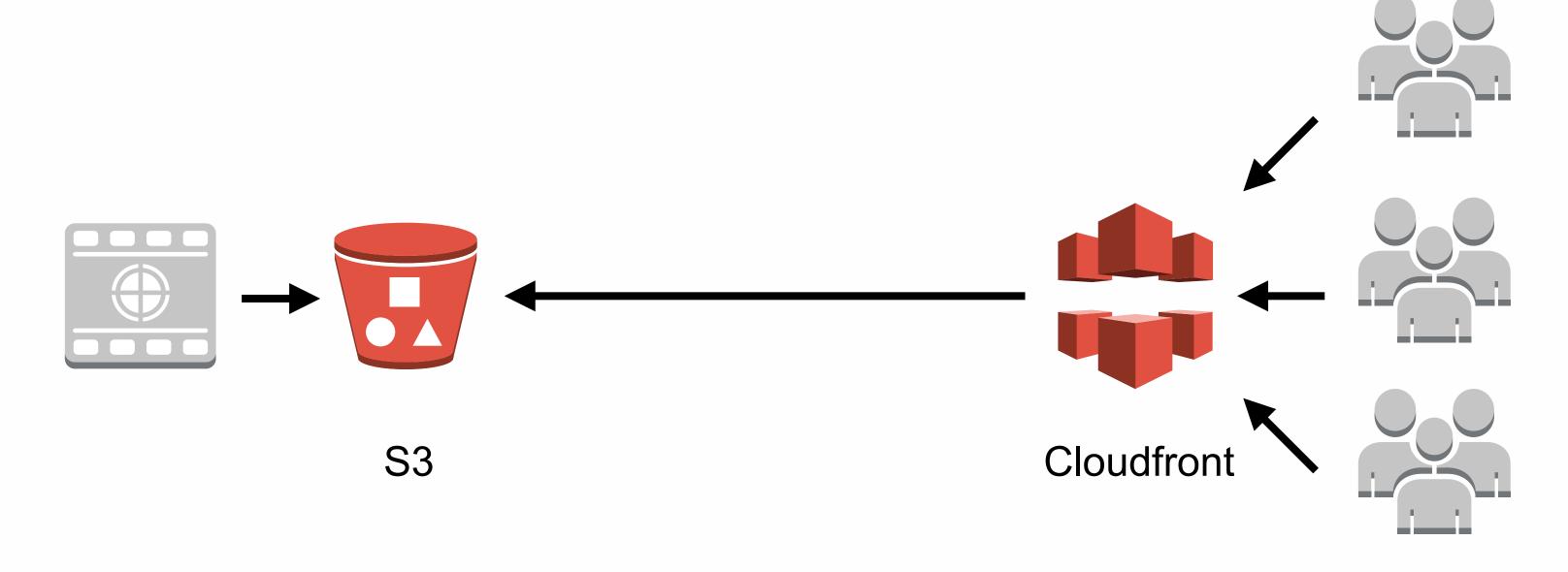
- No commitments
- Hopefully better caching performance
- Extensive documentation





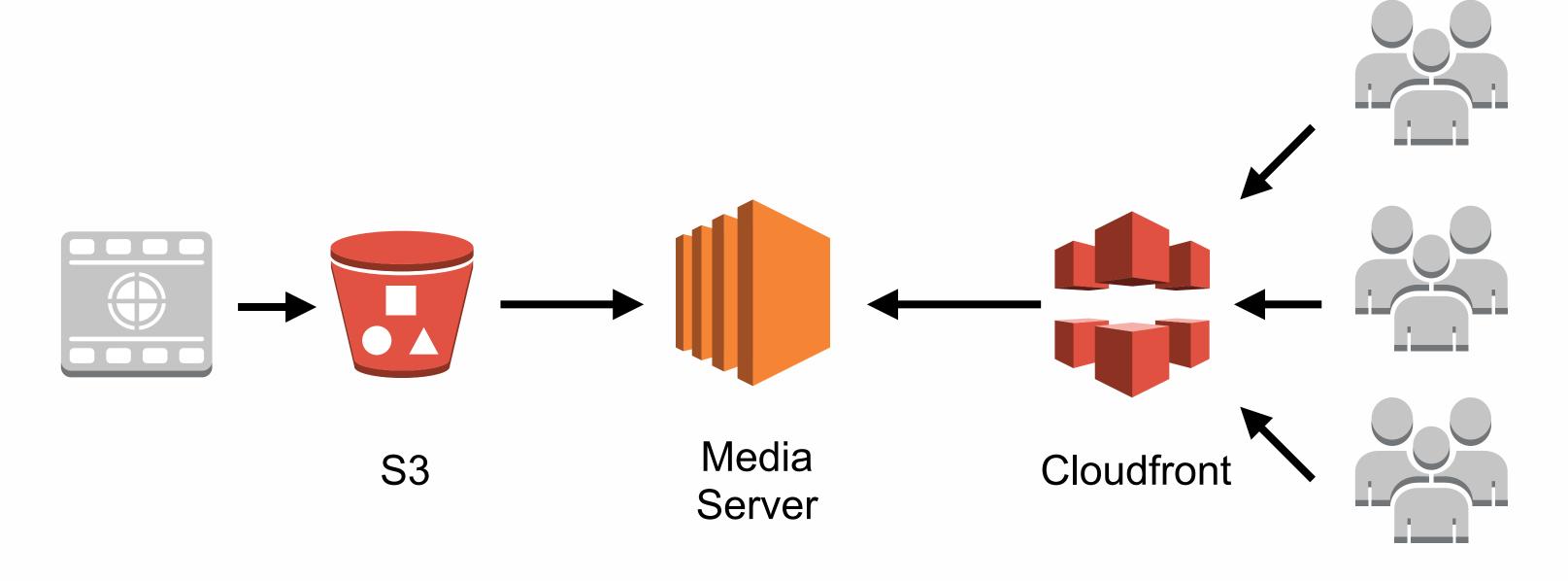


Our First Cloud (progressive download only)



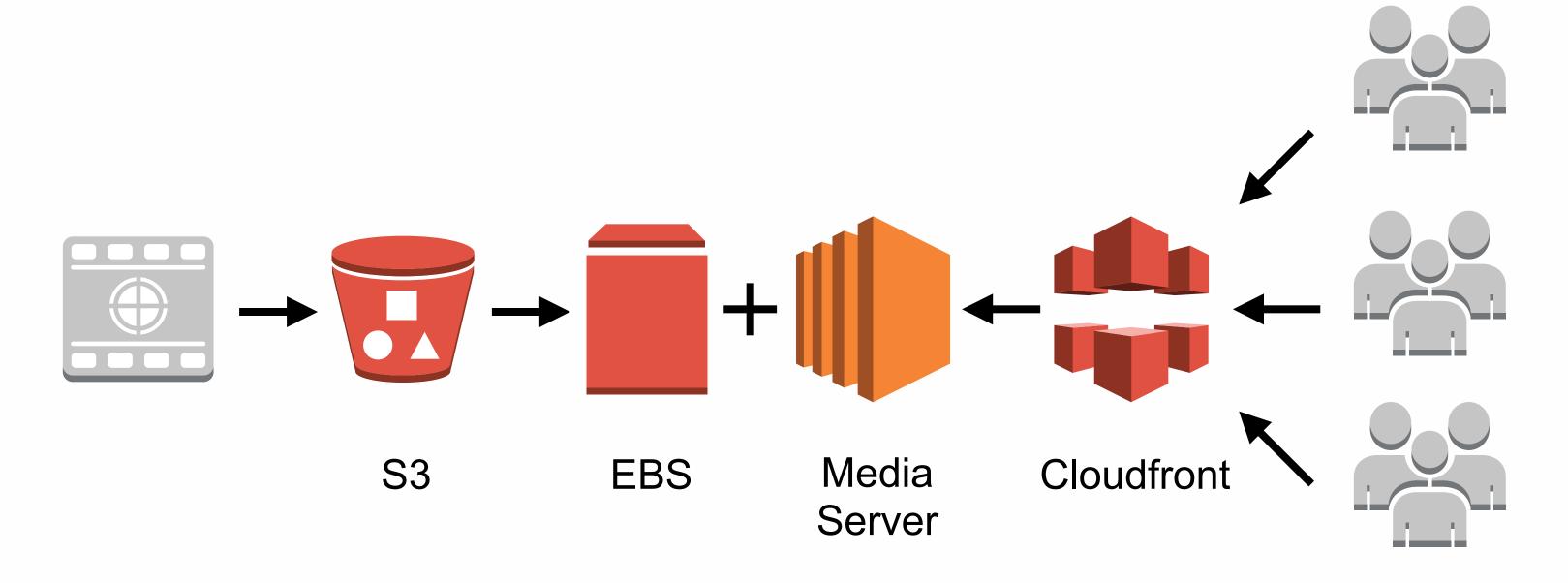


Our Second Cloud (progressive download + streaming)





Our Second Cloud (progressive download + streaming)





Verdict

- Cheaper
- More fun implementing
- Slightly better performance



Why did it not work?

- We have too many objects and too large objects*
- We have too few users
- dispersed too far around the planet

...resulting in too many CACHE MISSES.



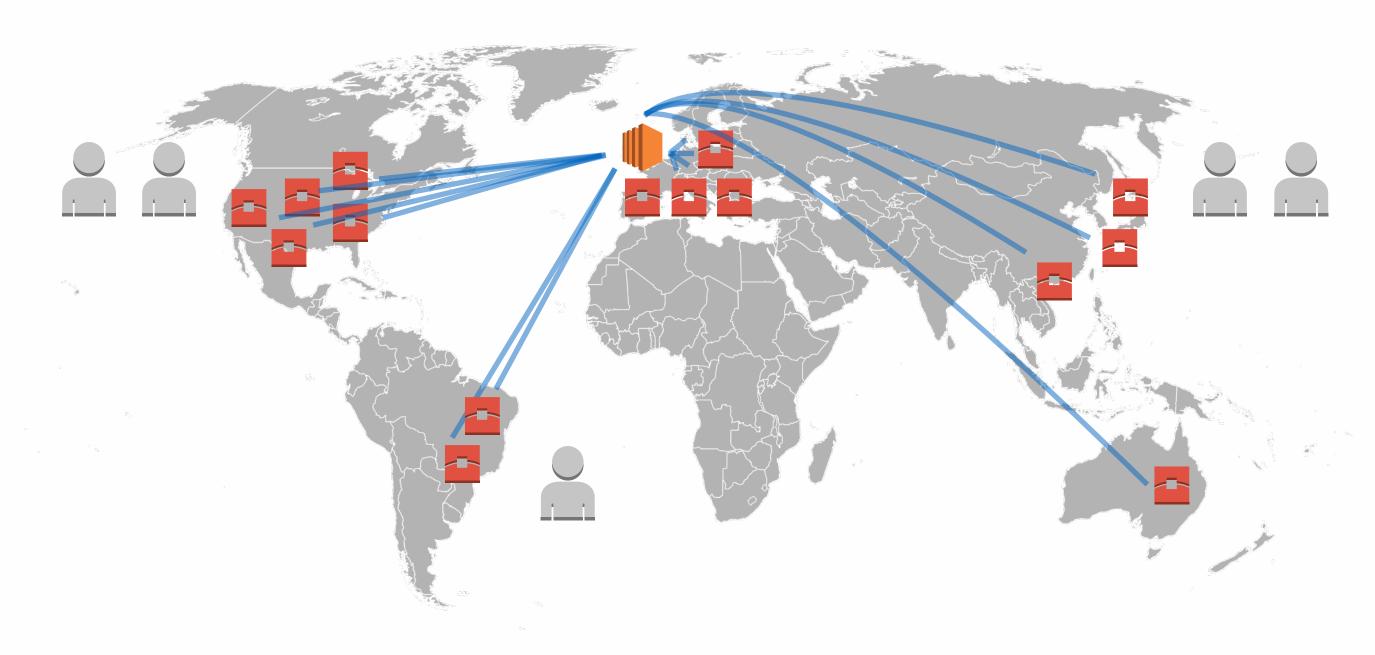
All CDNs assume that a sufficiently large number of users are requesting a sufficiently small number of objects.



The first user is the pawn.



One origin is not enough



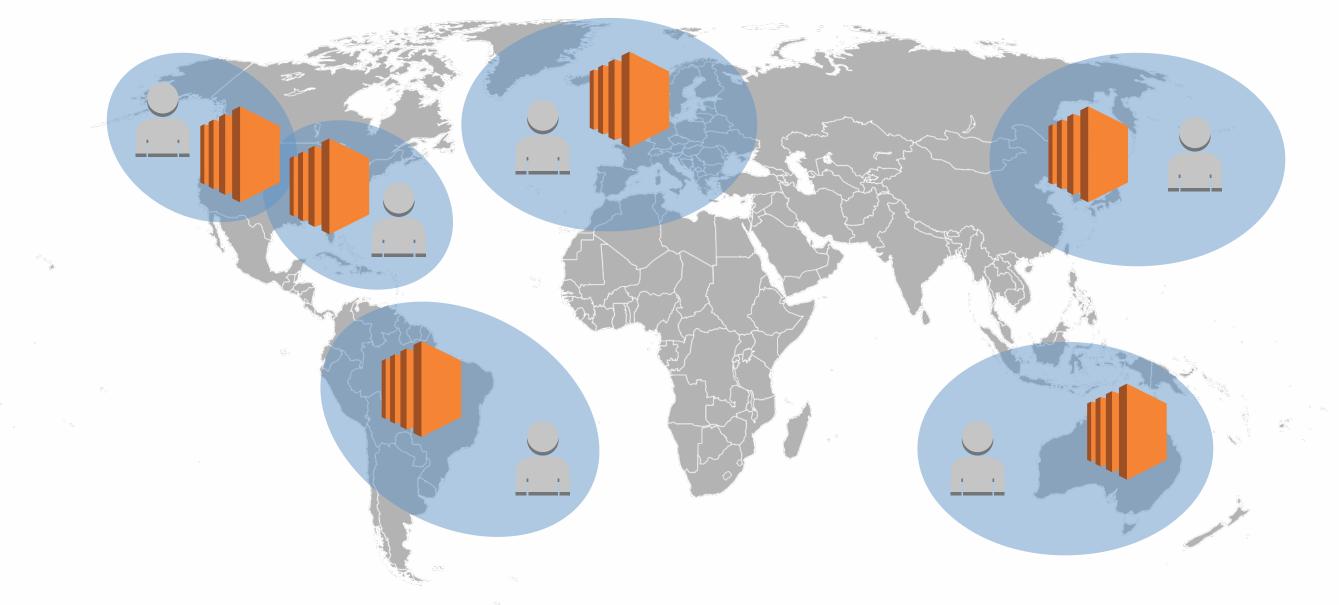


If the user can't come to us we must go to the user.



Our Third Cloud – with Route53







Challenges

- We had to do our own URL signing with mod_auth_token
- We had to copy our library across the world 6x
- We have to keep all regions in sync with S3/Ireland



New verdict

- Not cheaper
- Maximum control
- Awesome performance



What about live streaming?





Live streaming with Cloudfront

- We have very few and very small objects*
- We have a lot of users
- (still dispersed around the planet)

...resulting in many CACHE HITS.





Further employments of Cloudfront

- DONE: static website assets (GFX/CSS/JS)
- DONE: static app content via JSON API
- NEXT: full site acceleration incl. dynamic content



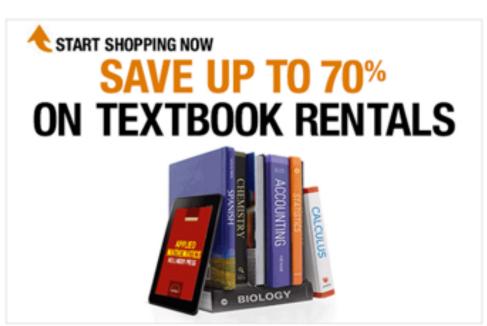


Textbooks

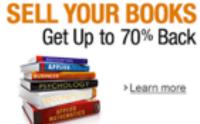
Find New, Used, Rental and eTextbooks

Enter a title, author, or ISBN(s)

Go







Get your textbooks the way you want them:

NEW

SAVE UP TO 40%

USED

SAVE UP TO 90%

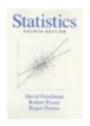
RENTAL

SAVE UP TO 70%

eTEXTB00K

SAVE UP TO 80%

Statistics, 4th Edition Get Up To: \$40.45



Pattern Recognition and Machine Learn... Get Up To: \$31.48



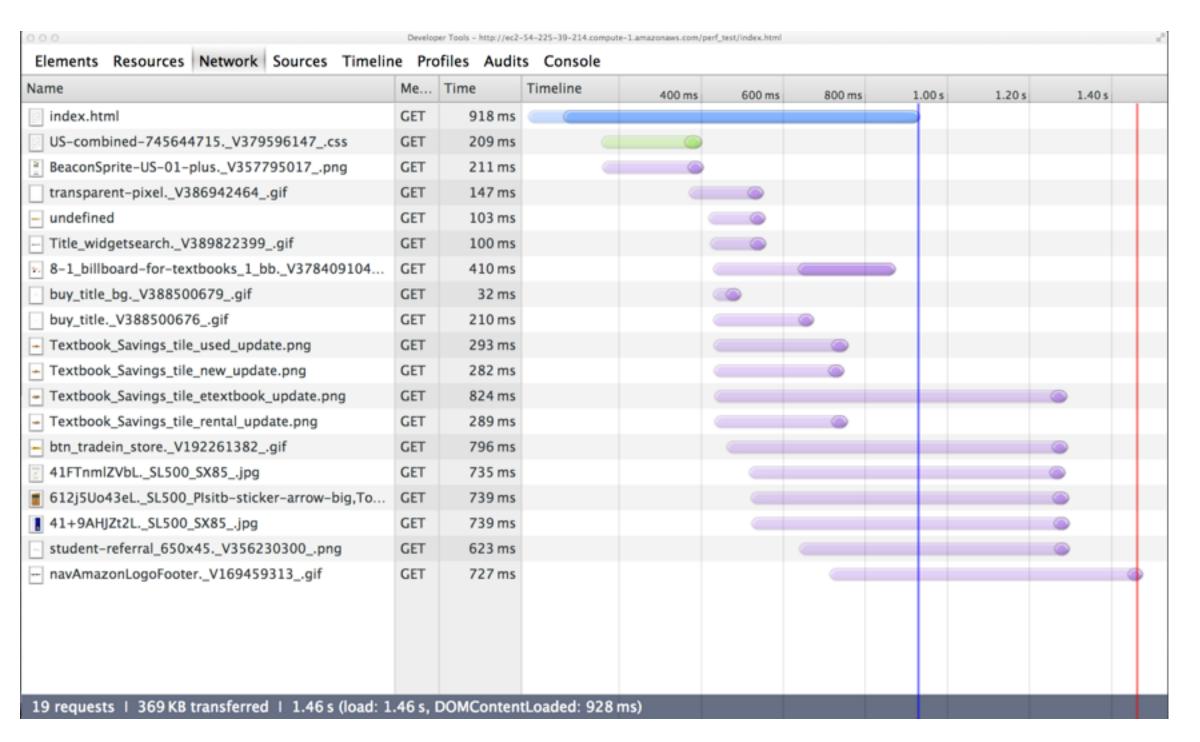


Learning From Data Get Up To: \$11.49



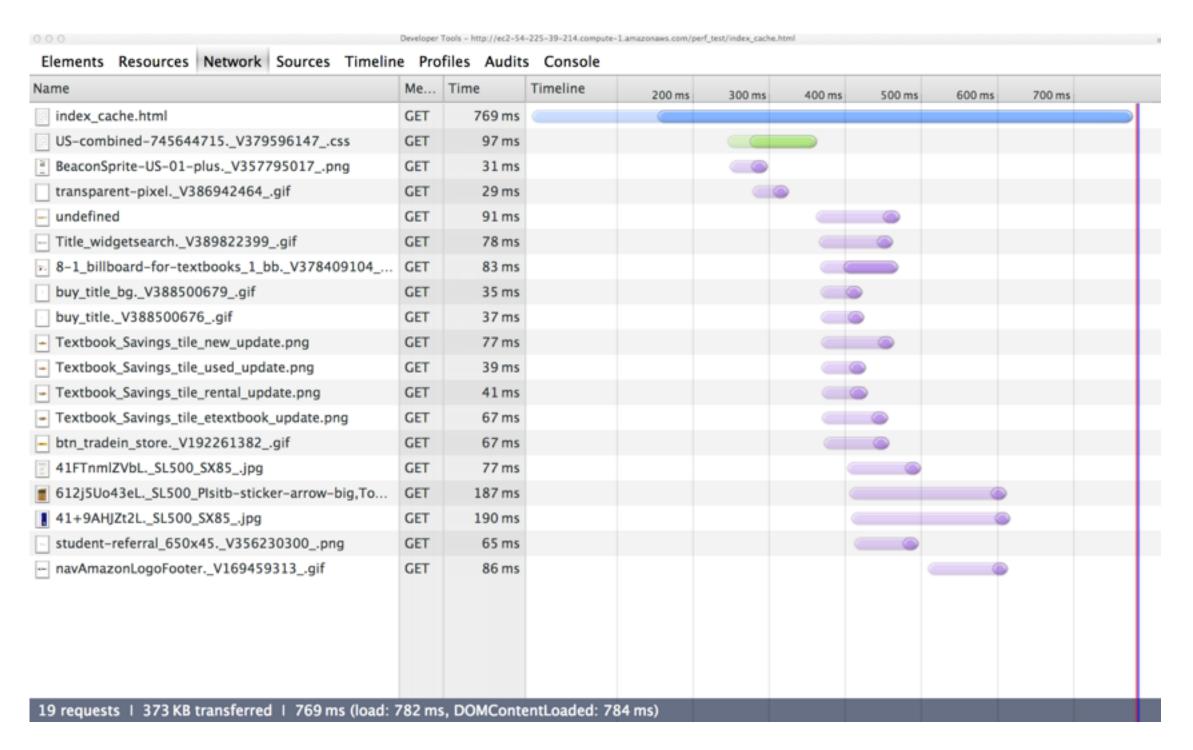


Before Caching: 1460 ms



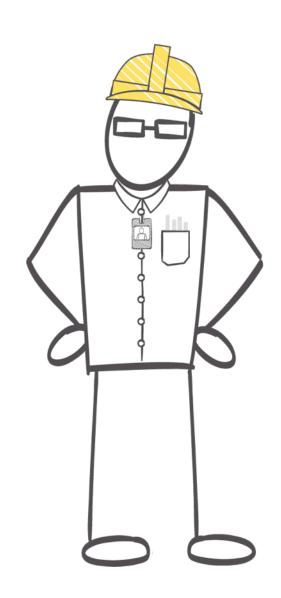


After Caching: 770 ms





Goal:







Cache as Much as You Can!





Cache as Much as You Can!

Collect Logs From Your Web Tier





Cache as Much as You Can!

- Collect Logs From Your Web Tier
- Run a Report on Your Logs (EMR, RDS, Redshift)
- Identify Top N URLs





- 220 /index.jsp
- 200 /images/book1.gif
- 120 /css/style.css
- 119 /js/script1.js
 - 110 /factory/create_image?name=book1&size=10x10
 - 100 /api/GetBooks?category=math
 - 90 /api/GetBooks?category=math&lang=spanish
 - 80 /api/GetBooks?top=10



Static or Reusable

Content that does not change for a given period of time





Static or Reusable

Content that does not change for a given period of time

CloudFront caches content for any period of time:

Hours, Minutes, Seconds





Content with Query Strings

110 /factor/create_image?name=book1&size=10x10

Reusable?



Content with Query Strings

110 /factor/create_image?name=book1&size=10x10

Yes!

- CloudFront can cache content with query strings
- Every unique query string combination is a new object in CloudFront's cache



- 220 /index.jsp
- 200 /images/book1.gif
- 120 /css/style.css
- 119 /js/script1.js
- 110 /factory/create_image?name=book1&size=10x10
- 100 /api/GetBooks?category=math
- 80 /api/GetBooks?top=10



Why Caching for Smaller Time Units?

1000 /api/GetBooks?top=10

- Example: Read heavy API with 1000 requests per second
- Offload your web-tier from handling 1000 RPS
- Offload your load balancer
- Provision less capacity and reduce cost



How About the Base Page?

220 /index.jsp

Reusable?



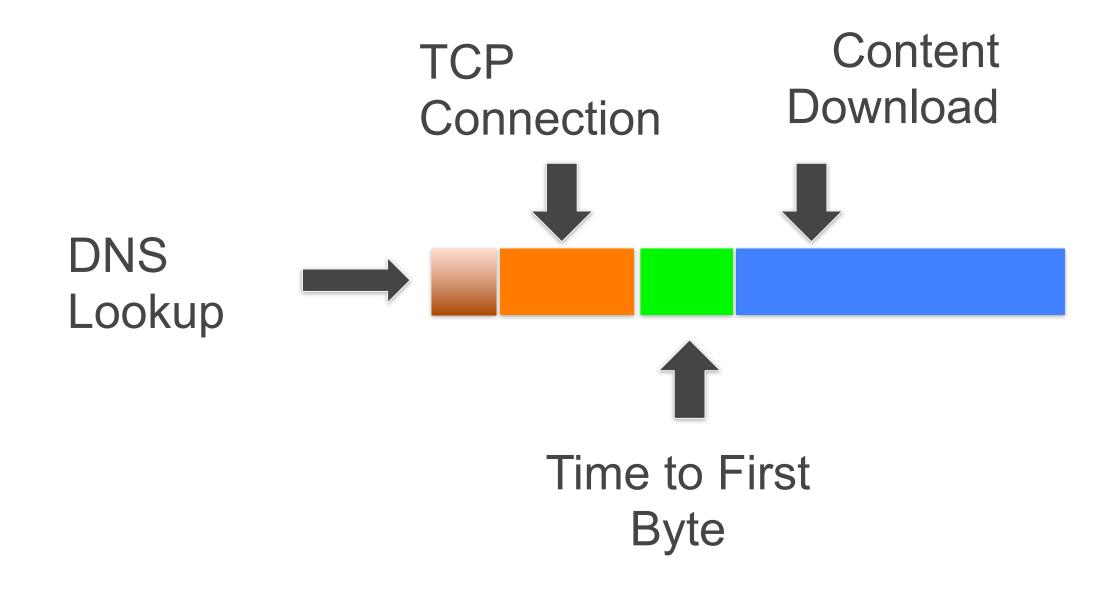
Optimizing Dynamic Content





Response Time

DNS + Connection + First Byte + Content Download





DNS



DNS



Amazon Route 53



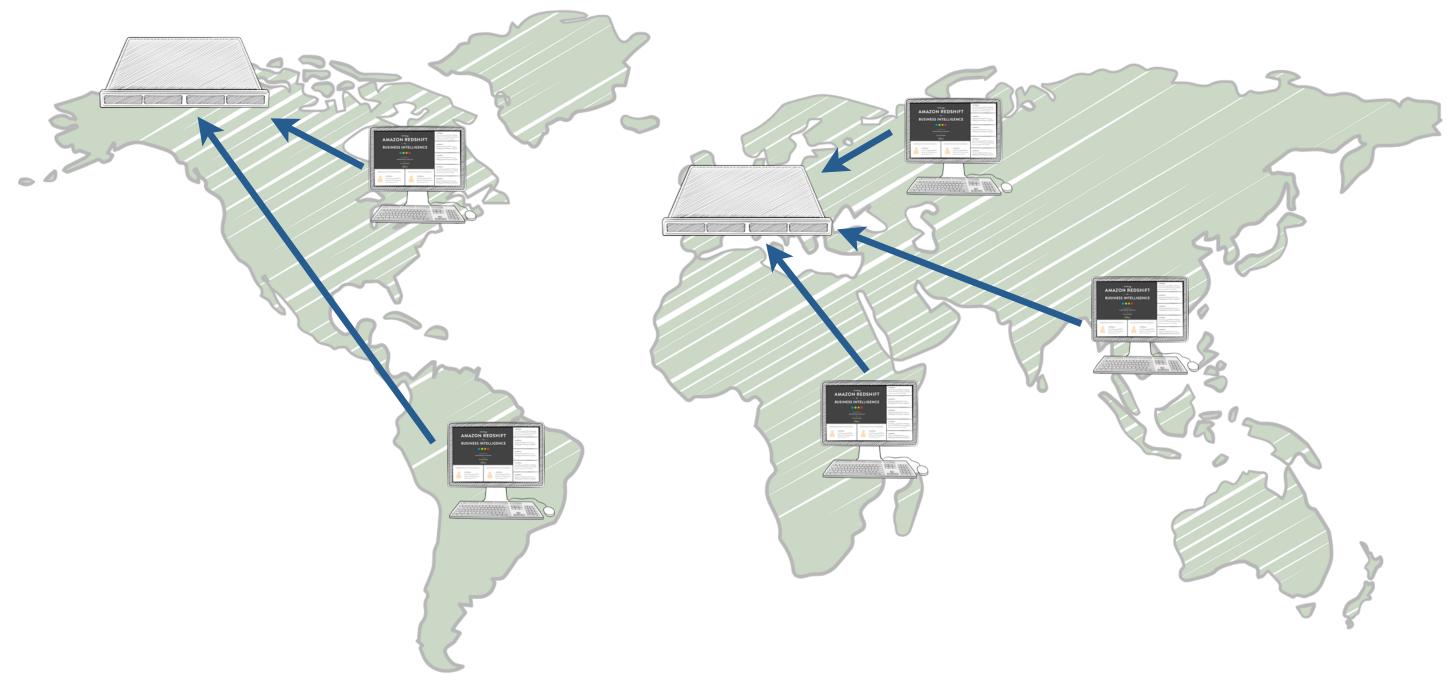
- Managed DNS
- Fast
- Low latency
- Global network
- Queries routed to the nearest DNS server







Without Route 53





With Route 53





DNS



Amazon Route 53



DNS



Amazon Route 53

Connection

First Byte



DNS

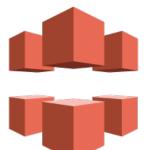
Connection

First Byte









Amazon Route 53

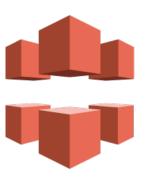
Amazon CloudFront Keep Alive

Amazon CloudFront Keep Alive



TCP/IP Hand Shake

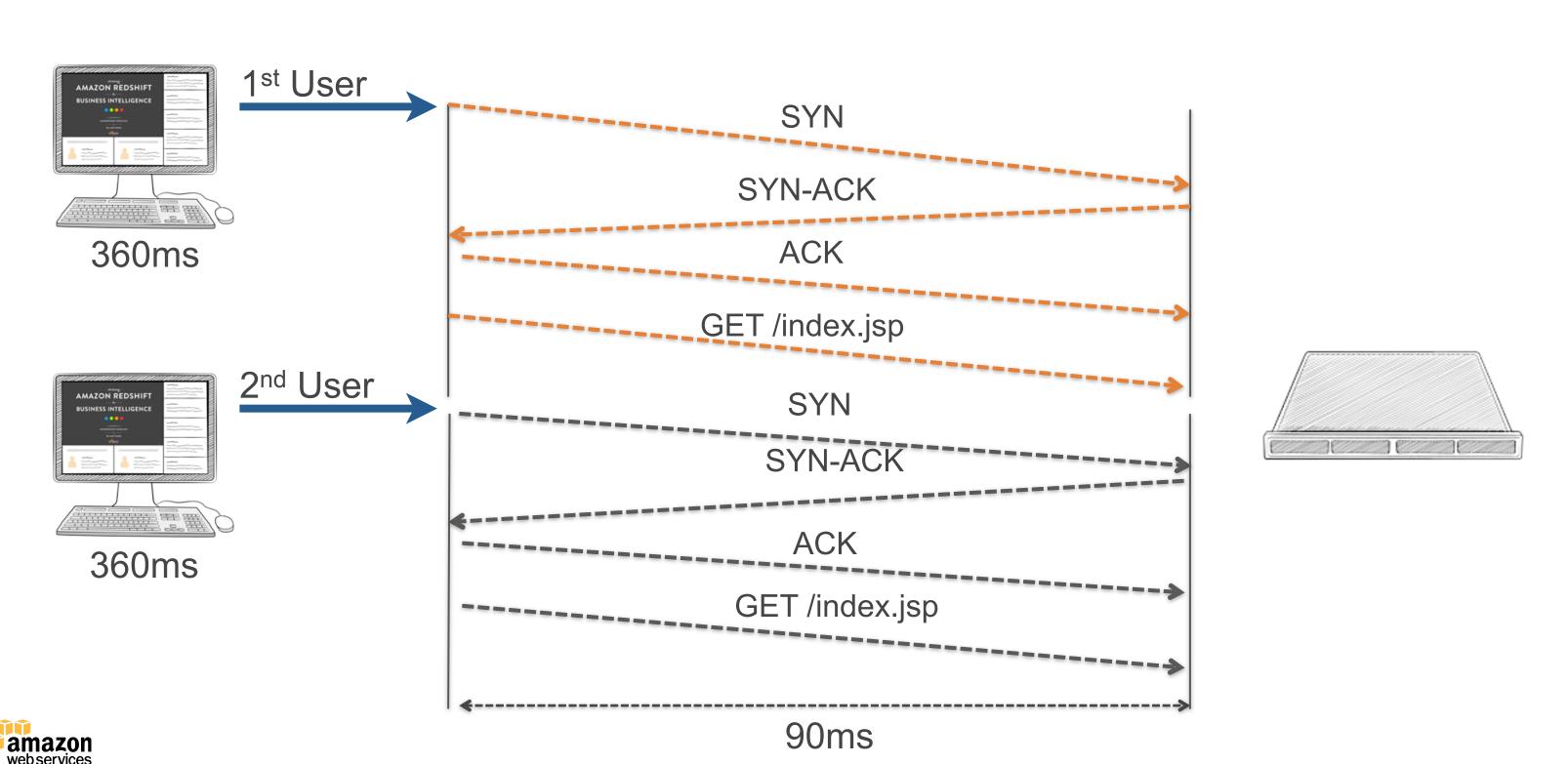
- HTTP Runs on TCP/IP
- TCP uses TCP handshake
- TCP handshake costs time
- Every HTTP Connection needs to complete TCP Handshake



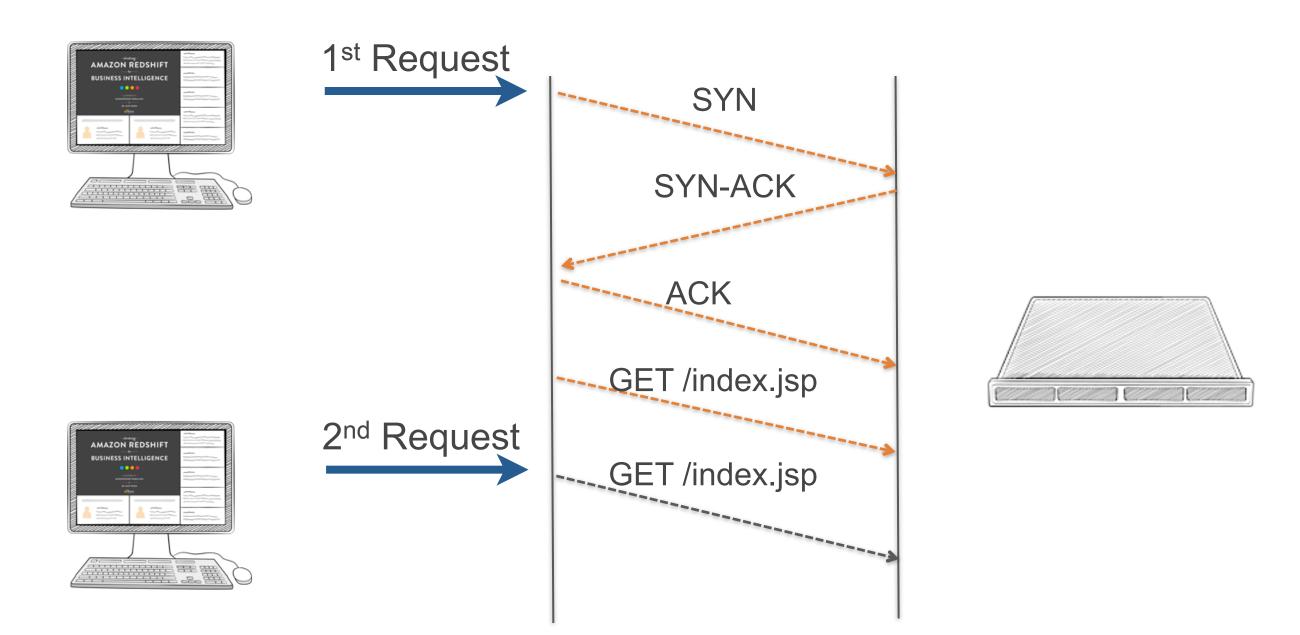
Amazon CloudFront



Two Users Without CloudFront

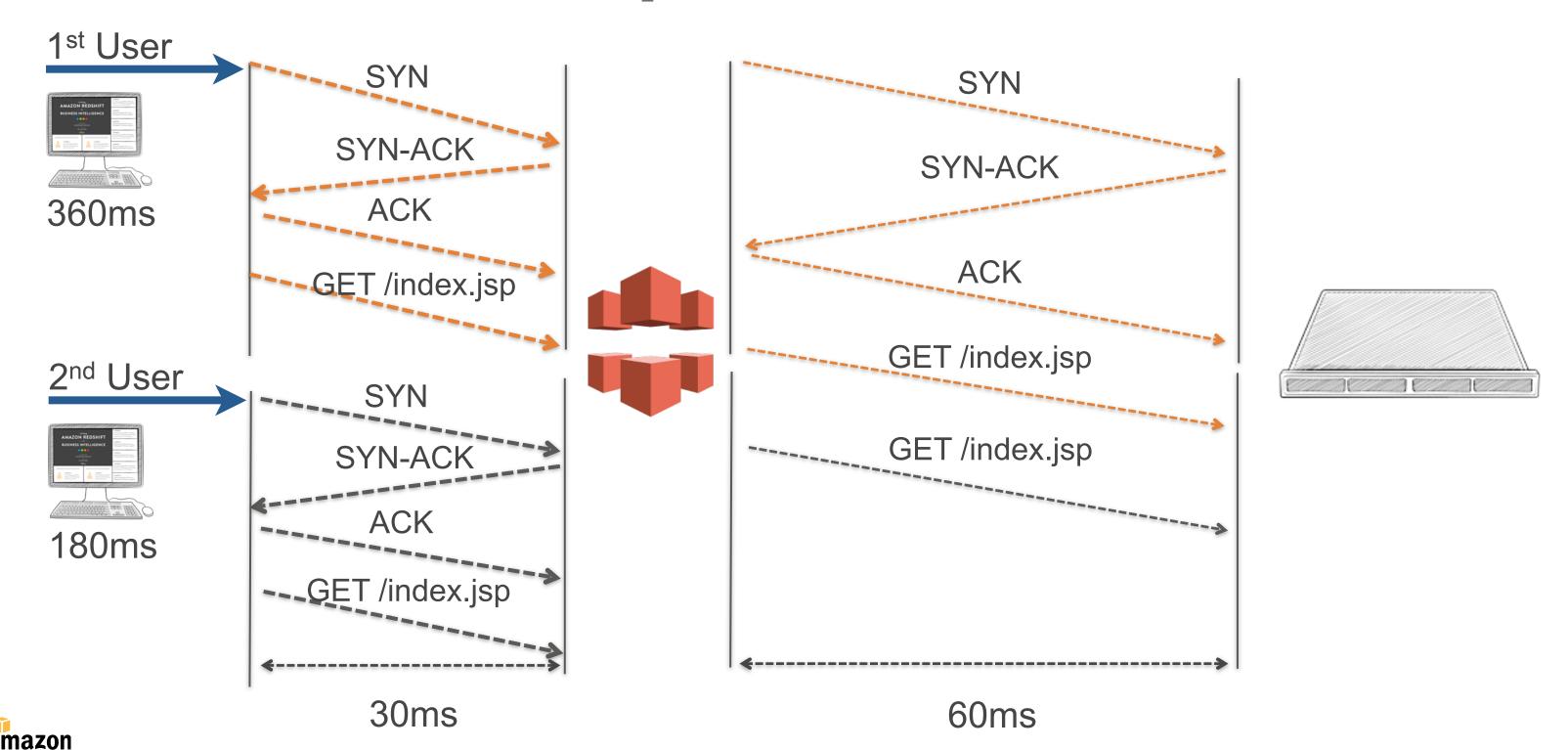


Keep Alive





CloudFront Keep Alive



CloudFront Keep Alive

- More Users ≠ More Connections
- Offloads Origin's CPU/Memory
- Improves Response Time





CloudFront Keep Alive Test

Test	CPU %
Without CloudFront	20 %
With CloudFront	6 %





DNS

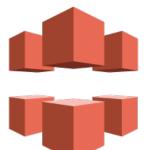
Connection

First Byte









Amazon Route 53

Amazon CloudFront Keep Alive

Amazon CloudFront Keep Alive



DNS

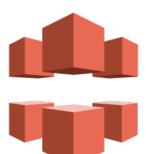
Connection

First Byte









Amazon Route 53

Amazon CloudFront Keep Alive & SSL Termination

Amazon CloudFront Keep Alive



CloudFront SSL Termination

- CloudFront Supports SSL
- Terminate SSL at the Edge (Half-Bridge)
- Or Terminate SSL at the Edge and
 Use SSL to the Origin as Well (Full-Bridge)
- Use CloudFront SSL Certificate
- Or Bring Your Own





DNS

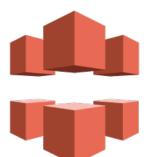
Connection

First Byte









Amazon Route 53

Amazon CloudFront Keep Alive & SSL Termination

Amazon CloudFront Keep Alive



DNS

Connection

First Byte

Content Download

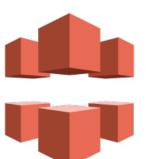














Amazon CloudFront Keep Alive & SSL Termination

Amazon CloudFront Keep Alive

Amazon CloudFront TCP/IP Optimizations



CloudFront TCP/IP Optimizations

TCP Slow Start Optimizations





TCP Slow Start Optimization Performance Results

Test	# Of Packets	Response Time Per Request	Response Time For 200 Requests
Without CloudFront	2605	170 ms	33.876 ms
With CloudFront	896	96 ms	19.24 ms



CloudFront TCP/IP Optimizations

- TCP Slow Start Optimizations
- HTTP PUT/POST Optimizations





DNS

Connection

First Byte

Content Download

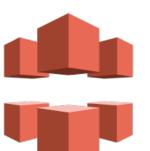












Amazon Route 53

Amazon CloudFront Keep Alive & SSL Termination

Amazon CloudFront Keep Alive

Amazon CloudFront TCP/IP Optimizations



But You Can Do More





Route 53 Latency Based Routing



Route 53



Route 53 Latency Based Routing

- Create multiple stacks in different EC2 regions
- Create LBR records with geo information tags
- Route 53 routes users to the lowest latency endpoint
- Better performance and availability
- Easy to use, low cost
- With or without CloudFront











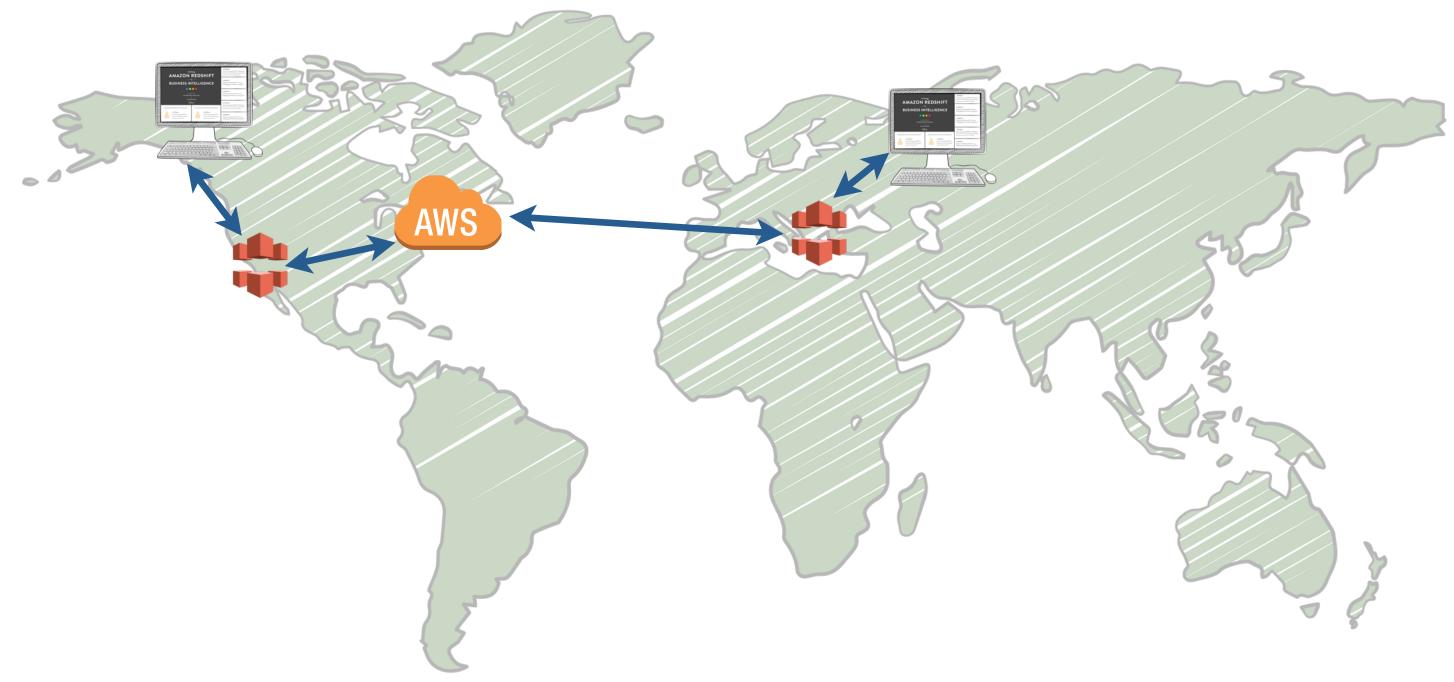






































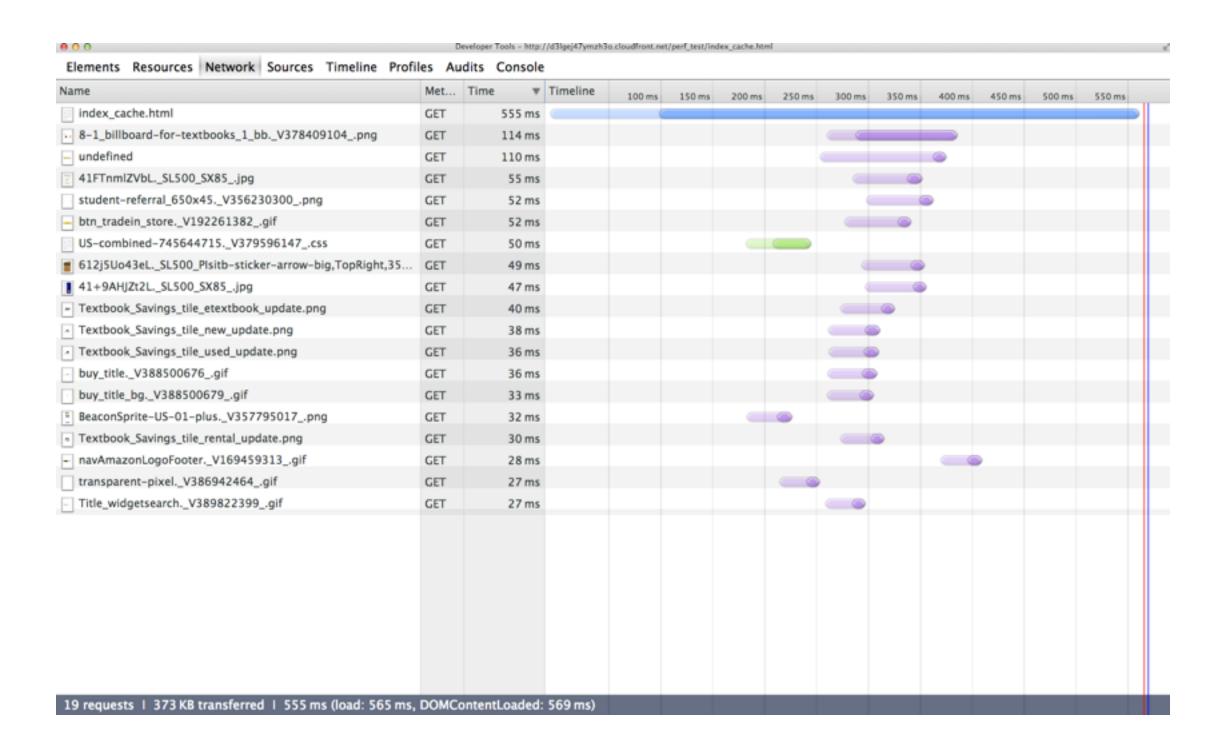






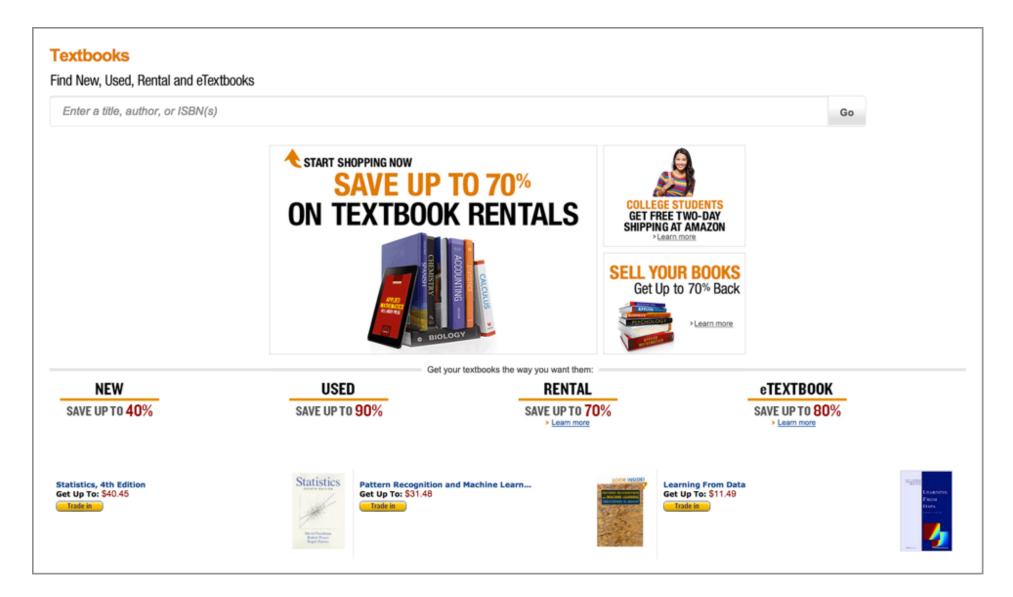


After CloudFront Dynamic Content Optimization: 555 ms

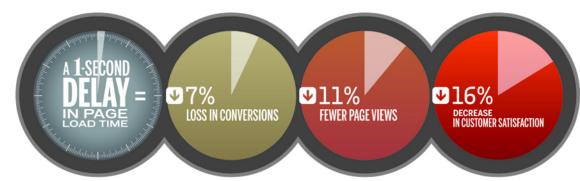




Goal Reached!







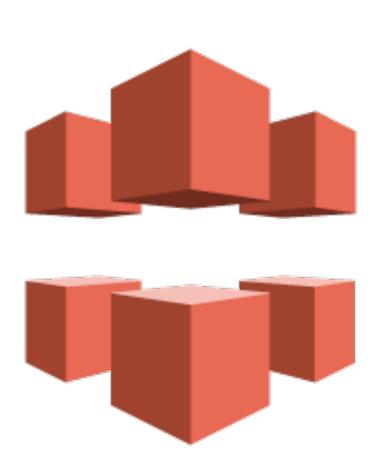






Bonus: Fault Tolerance

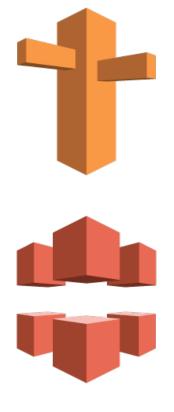






Bonus: Fault Tolerance

- Route 53 Health Checks
- CloudFront Reduces Load on Your Origins
- CloudFront Works Together With Route 53
 Latency Baced Routing
- CloudFront Fails Over to Cached Content
- CloudFront Customized Error Pages





Summary

- Accelerate all your content with CloudFront
- Use CloudFront with Route 53 latency-based routing to improve your performance
- Design for failure with CloudFront and Amazon Route 53



AVVS Summits 2014

Thank You!

Constantin Gonzalez
Alexander D. MacWilliam



