DEVNEXUS™

the professional developer conference

Amazon Web Services

Going from development to production

Introduction

Tom Elrod – CTO @ Make & Build

We build cool stuff


amazon
web services™

Compute

Amazon Elastic Compute Cloud (Amazon EC2)

Amazon Elastic MapReduce

Auto Scaling

Storage

Amazon Simple Storage Service (Amazon S3)

Amazon Elastic Block Storage (Amazon EBS)

AWS Import/ Export

AWS Storage Gateway Service

AWS Glacier

Database

Amazon DynamoDB

Amazon Relational Database Service (Amazon RDS)

Amazon ElastiCache

Networking

Amazon Route 53

Amazon Elastic Load Balancing

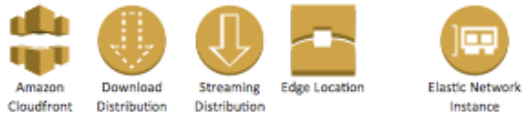
AWS Direct Connect

Amazon Virtual Private Cloud (VPC)

Content Delivery

Amazon Cloudfront

Elastic Network Instance



Application Services

Amazon Simple Queue Service (SQS)

Amazon Cloudsearch

Amazon Simple Email Service (SES)

Amazon Simple Workflow (SWF)

Amazon Simple Notification Service (SNS)

Deployment and Management

Amazon Elastic Beanstalk

AWS Identity and Access Management (IAM)

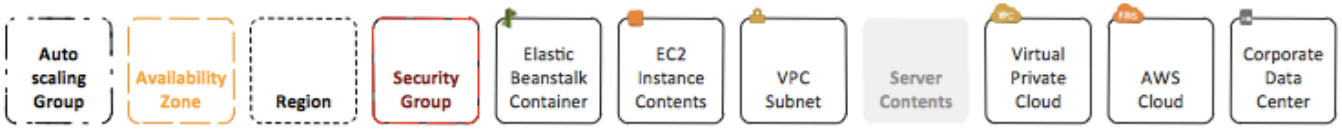
AWS CloudFormation

Monitoring

Amazon CloudWatch

Non-Service Specific

Groups



Narrowing the field

EC2

RDS

Elastic Load Balancer

S3 Bucket

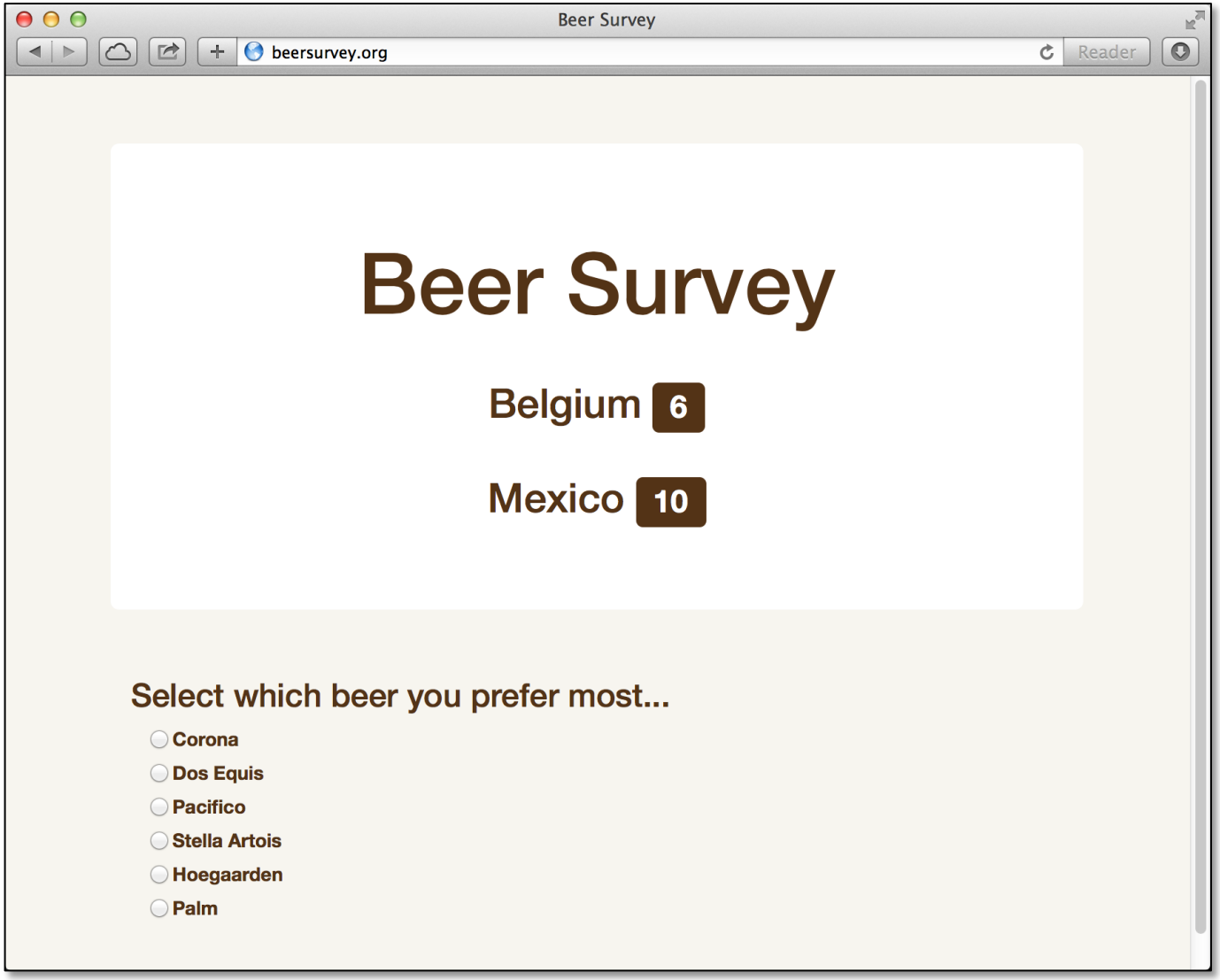
Route 53

CloudFront

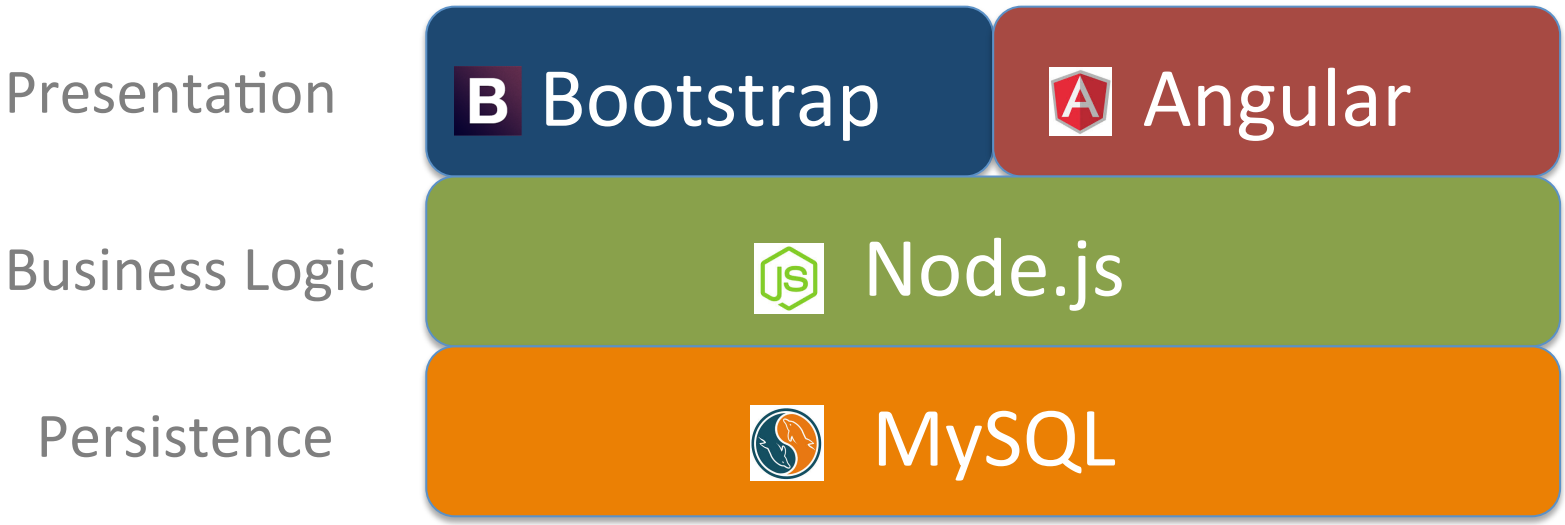
CloudFormation

Beer Survey

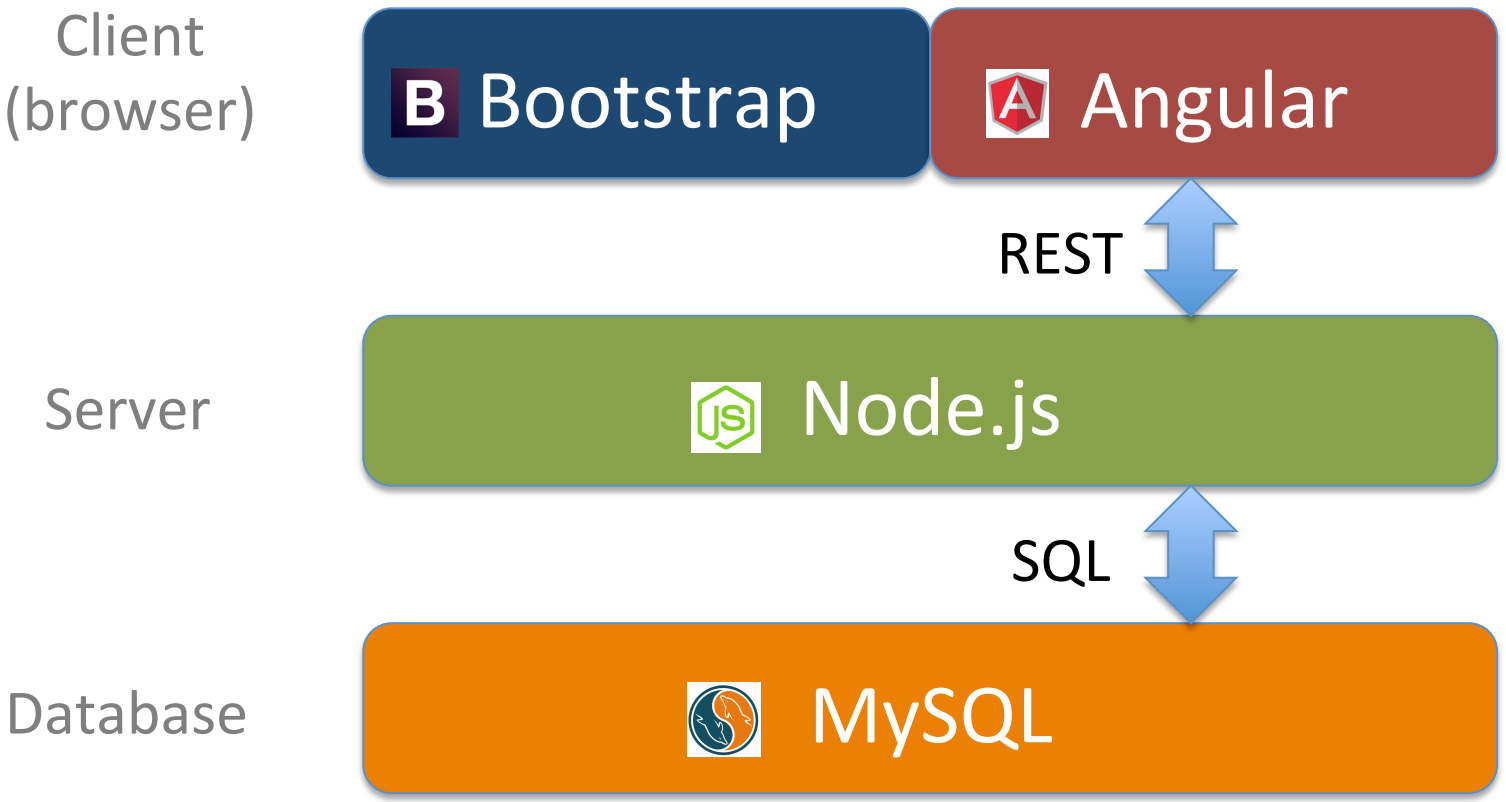
Web Application



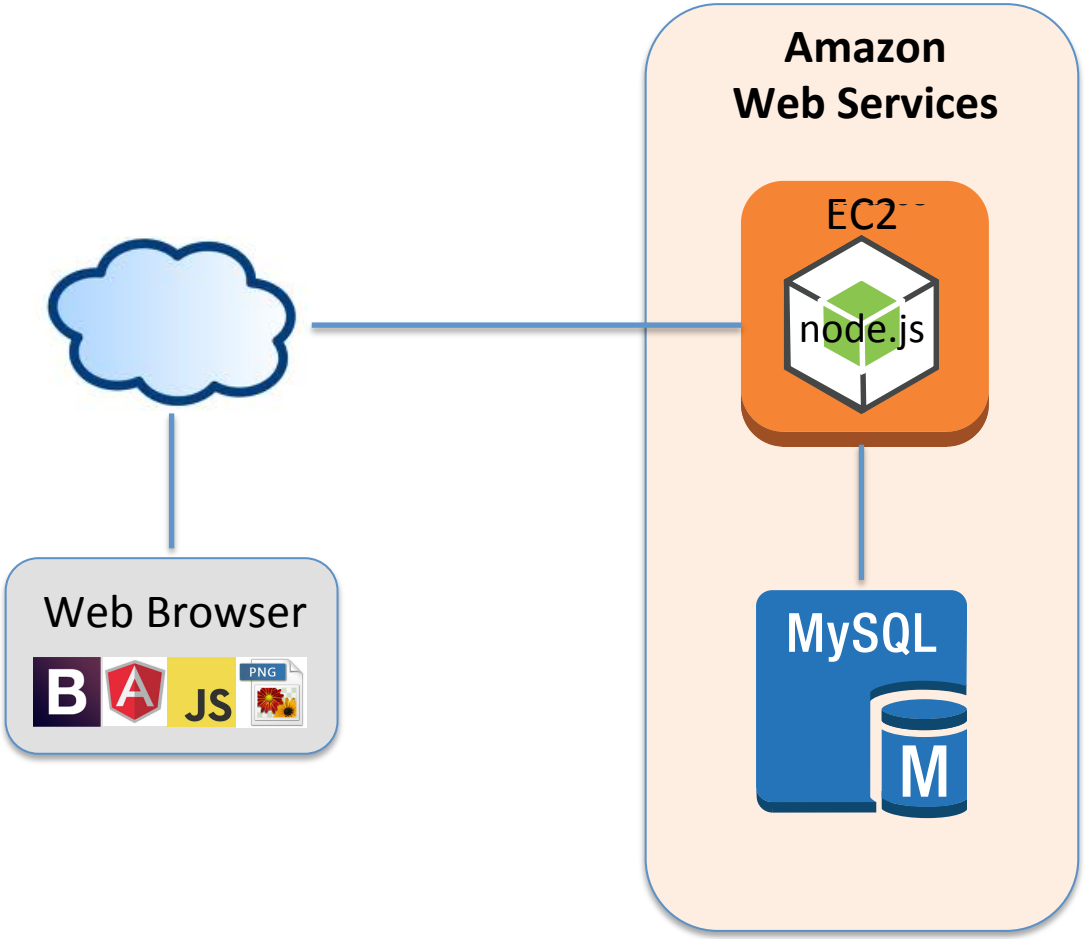
Beer Survey – Tech Stack



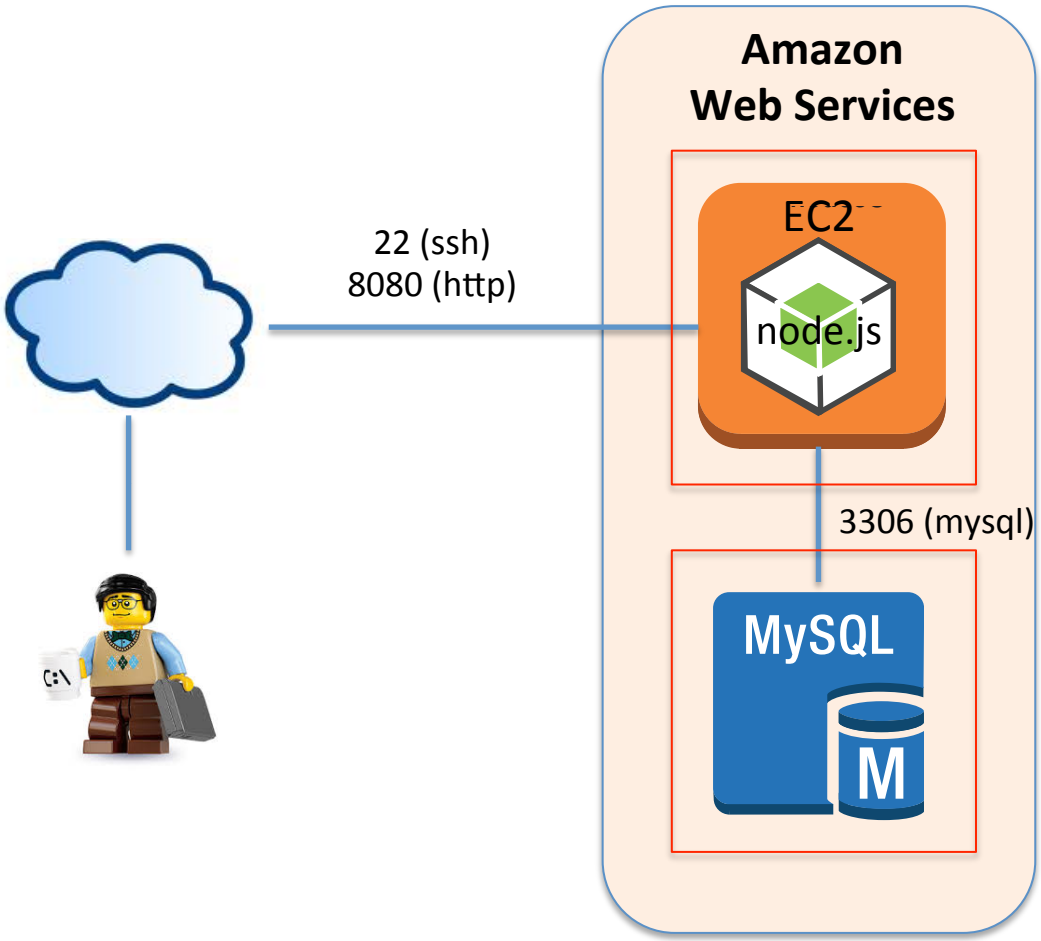
Beer Survey – Architecture



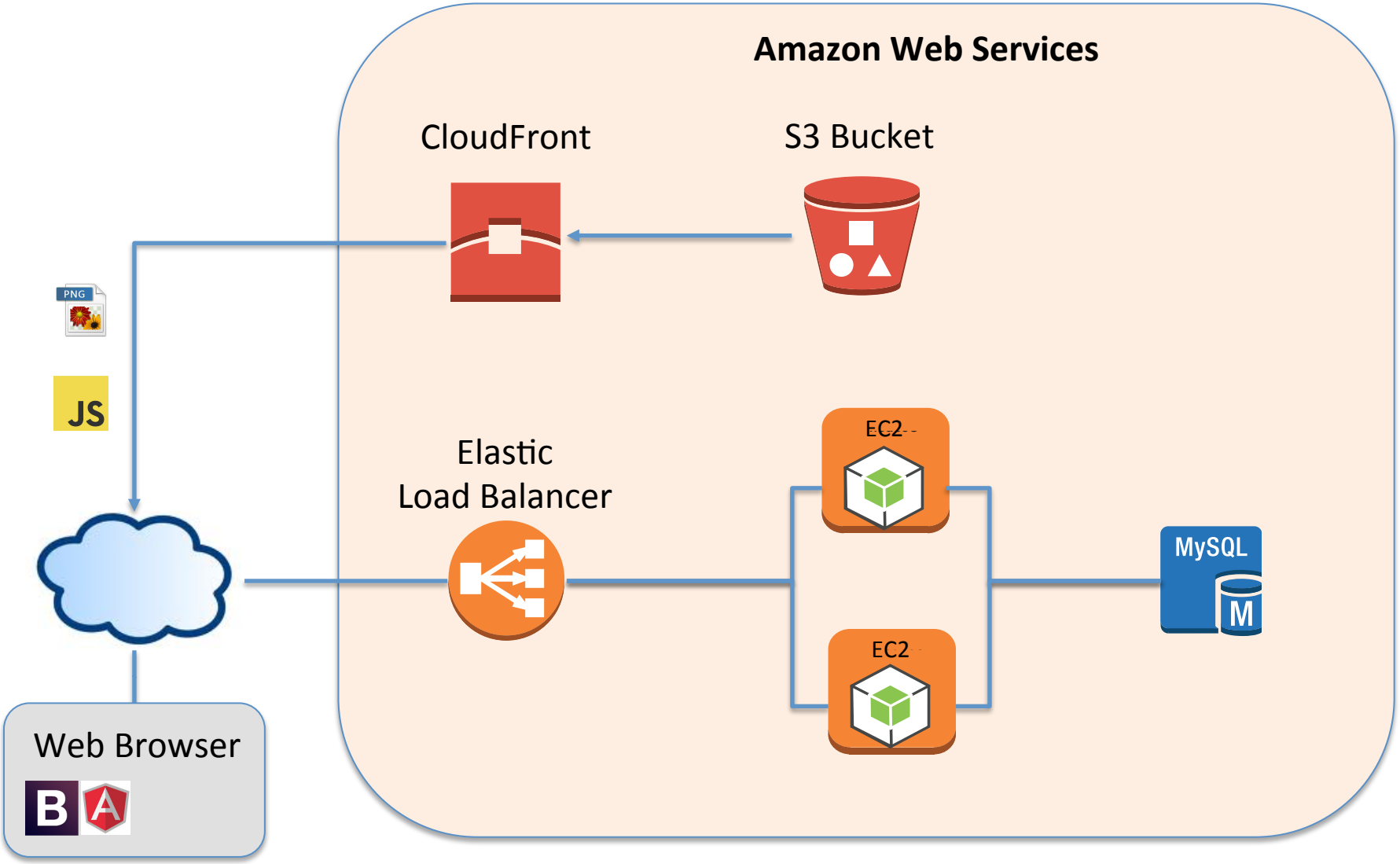
Development Environment



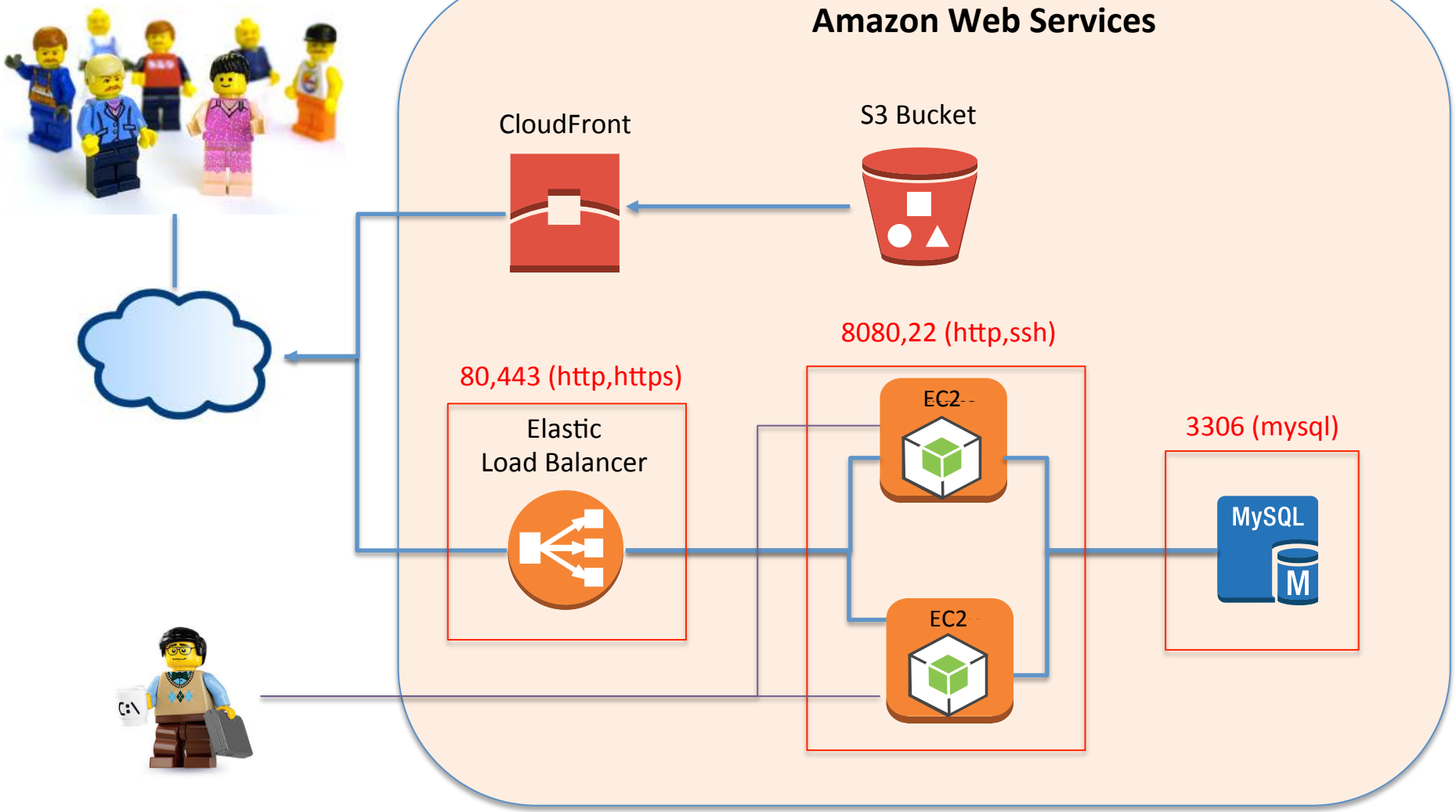
Development Environment



Production Environment



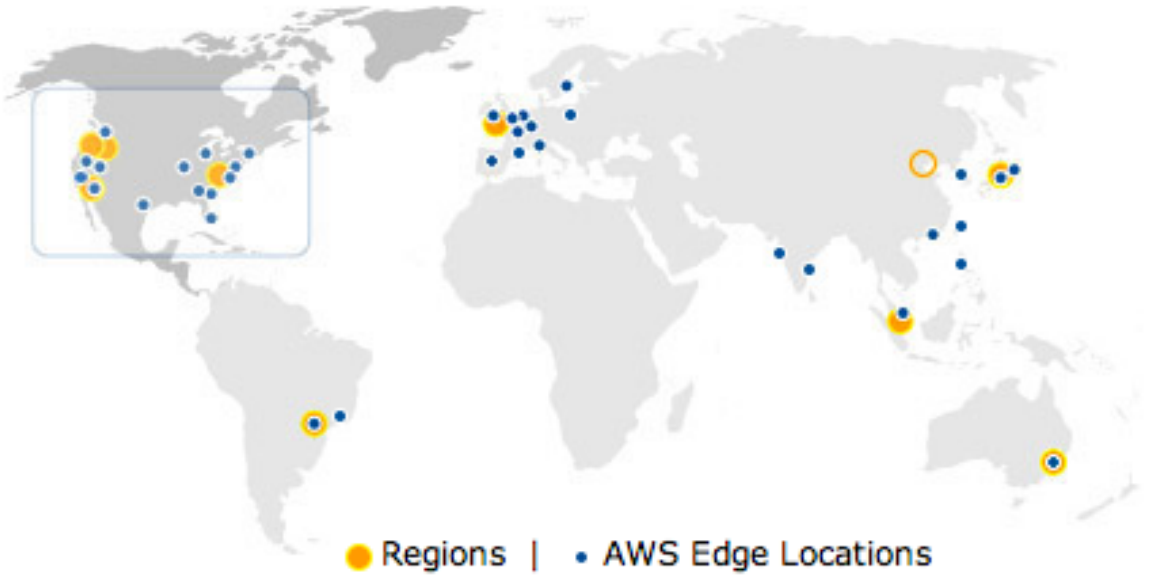
Production Environment



EC2

- Virtual servers that can run applications
- Provision new server instances in minutes
- Pay only for capacity actually used
- Up to 20 instances per account (apply for more)
- Tip: Do not use for sending email

Regions



US East (N. Virginia) $

US West (Oregon) $

US West (N. California) $$

South America (Sao Paulo) $$$

Asia Pacific (Singapore) $$$

Asia Pacific (Tokyo) $$$$

Asia Pacific (Sydney) $$$

EU (Ireland) $$

Instant Type Families

Family	Types	Popular Use cases
General Purpose	M1, M3	Small and mid-size databases, data processing, encoding, and caching.
Compute-optimized	C1, CC2, C3	High-traffic web applications, ad serving, batch processing, video encoding, and distributed analytics.
GPU	G2, CG1	Game streaming, 3D application streaming, and other server-side graphics workloads (G2). Computational chemistry, rendering, financial modeling, and engineering design (CG1)
Memory-optimized	M2, CR1	High performance databases, distributed memory caches, and in-memory analytics.
Storage-optimized	HS1, I2, HI1	NoSQL databases like Cassandra and MongoDB, and scale out transactional databases (I2, HI1). Data warehousing, Hadoop, and cluster file systems (HS1).

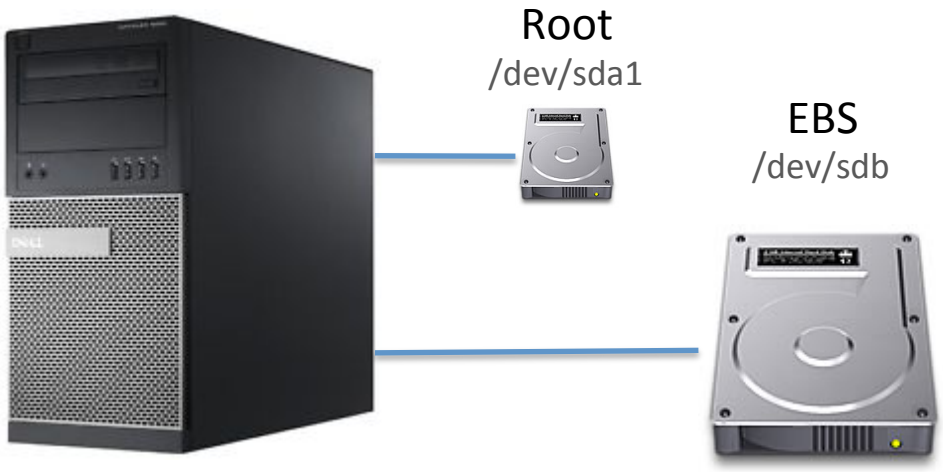
CPU capacity - ECUs & vCPUs

Type	CPU
m1.small	Opteron 2218
c1.medium	Xeon E5410
m1.large	Xeon E5430
m2.xlarge	Xeon X5550
c1.xlarge	Xeon E5410

- **ECU** – One EC2 Compute Unit provides the equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor
- **vCPU** – The number of virtual CPUs per instance

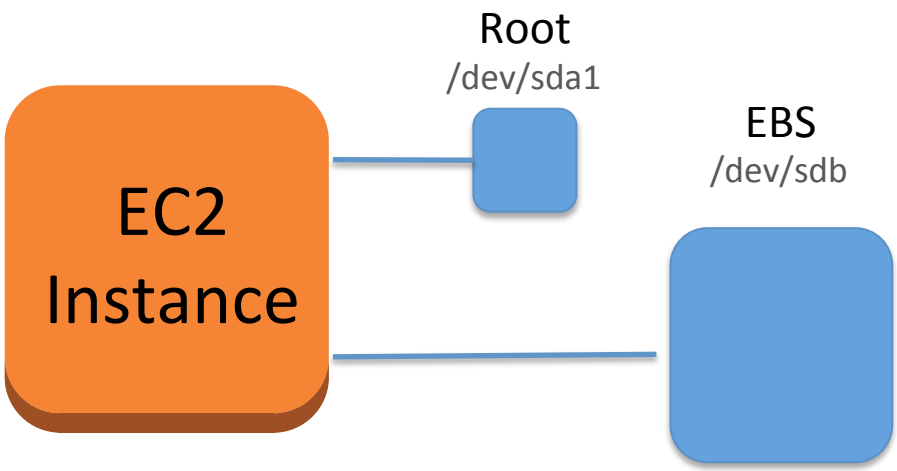
EBS – Elastic Block Store

"Amazon Elastic Block Store (Amazon EBS) provides persistent block level storage volumes for use with Amazon EC2 instances in the AWS Cloud."



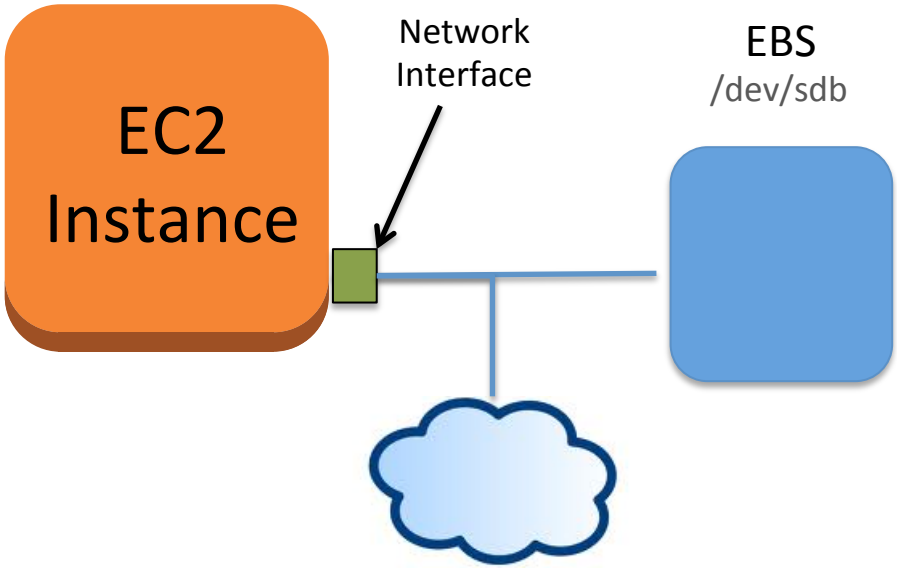
EBS – Elastic Block Store

"Amazon Elastic Block Store (Amazon EBS) provides persistent block level storage volumes for use with Amazon EC2 instances in the AWS Cloud."



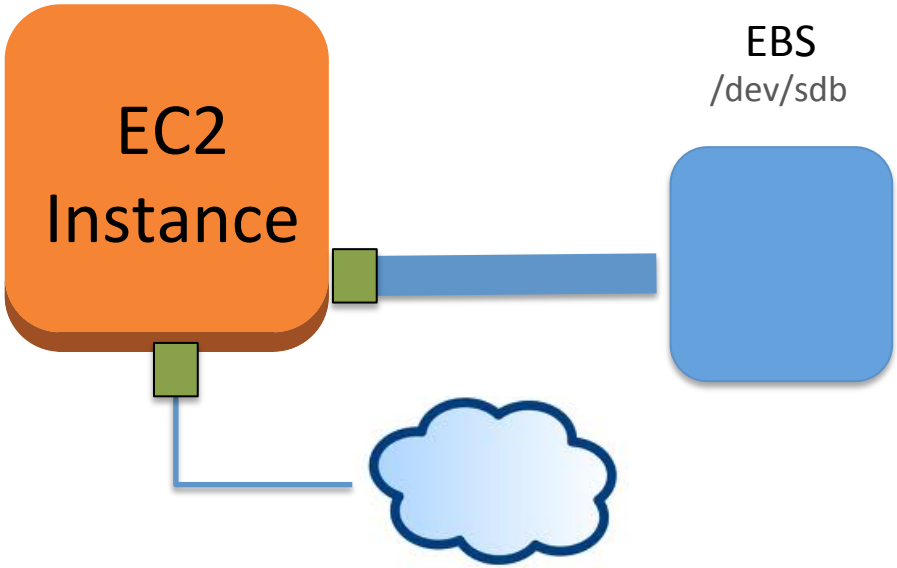
EBS – Standard

EC2 instance normally communicate with EBS volumes using its network interface, routing Internet traffic and I/O operations over the single network interface.



EBS – Optimized

Instances with *EBS Optimized* badge receive another network interface, dedicated only to EBS' traffic, with a fixed throughput of 500Mbps for **m1.large** types and 1Gbps for the others.



North America

US East (Northern Virginia) Region
EC2 Availability Zones: 5* Launched 2006

US West (Northern California) Region
EC2 Availability Zones: 3* Launched 2009

US West (Oregon) Region
EC2 Availability Zones: 3 Launched 2011

AWS GovCloud (US) Region
EC2 Availability Zones: 2 Launched 2011

Asia Pacific

Asia Pacific (Singapore) Region
EC2 Availability Zones: 2 Launched 2010

Asia Pacific (Tokyo) Region
EC2 Availability Zones: 3 Launched 2011

Asia Pacific (Sydney) Region
EC2 Availability Zones: 2 Launched 2012

China (Beijing) Region
EC2 Availability Zones: 1 Coming Soon

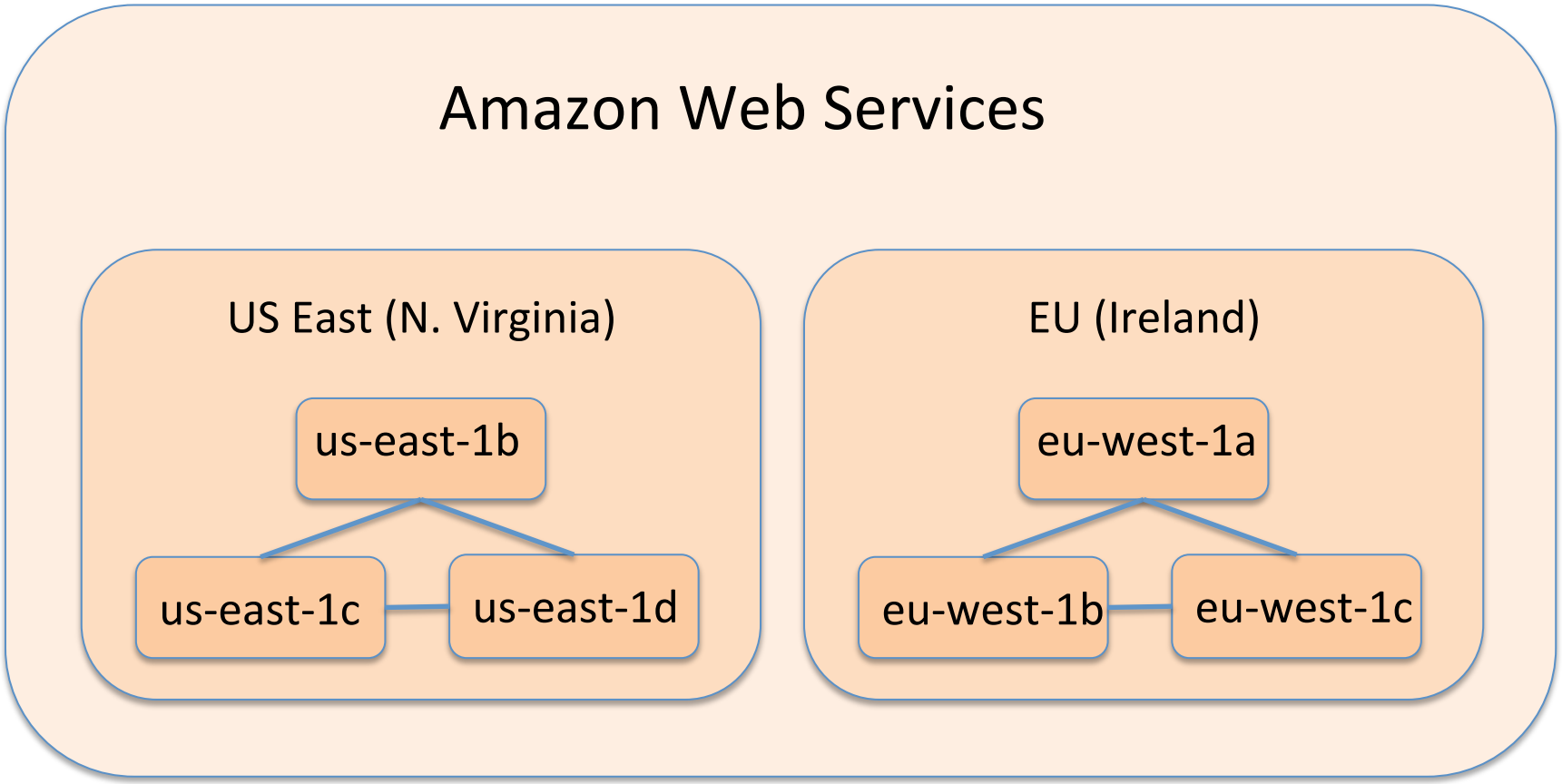
Europe / Middle East / Africa

EU (Ireland) Region
EC2 Availability Zones: 3 Launched 2007

South America

São Paulo Region
EC2 Availability Zones: 2 Launched 2011

Availability Zones



IOPS
(input/output operations per second)

- Provisioned IOPS volumes can achieve single digit millisecond latencies and are designed to deliver within 10% of the provisioned IOPS performance 99.9% of the time.
- Provisioned IOPS volumes support up to 30 IOPS per GB.
- Enables provisioning 4000 IOPS on a volume as small as 134 GB up to 1 TB.
- As a point of reference, a standard EBS volume will generally provide about 100 IOPS on average.

Elastic IP Addresses

- Map static IP address to EC2 instance
- Limited to 5 per account (can apply for more)
- Charged for unallocated IP addresses ($.01/hr)
- Reverse DNS records can be configured (requires filling out a form)

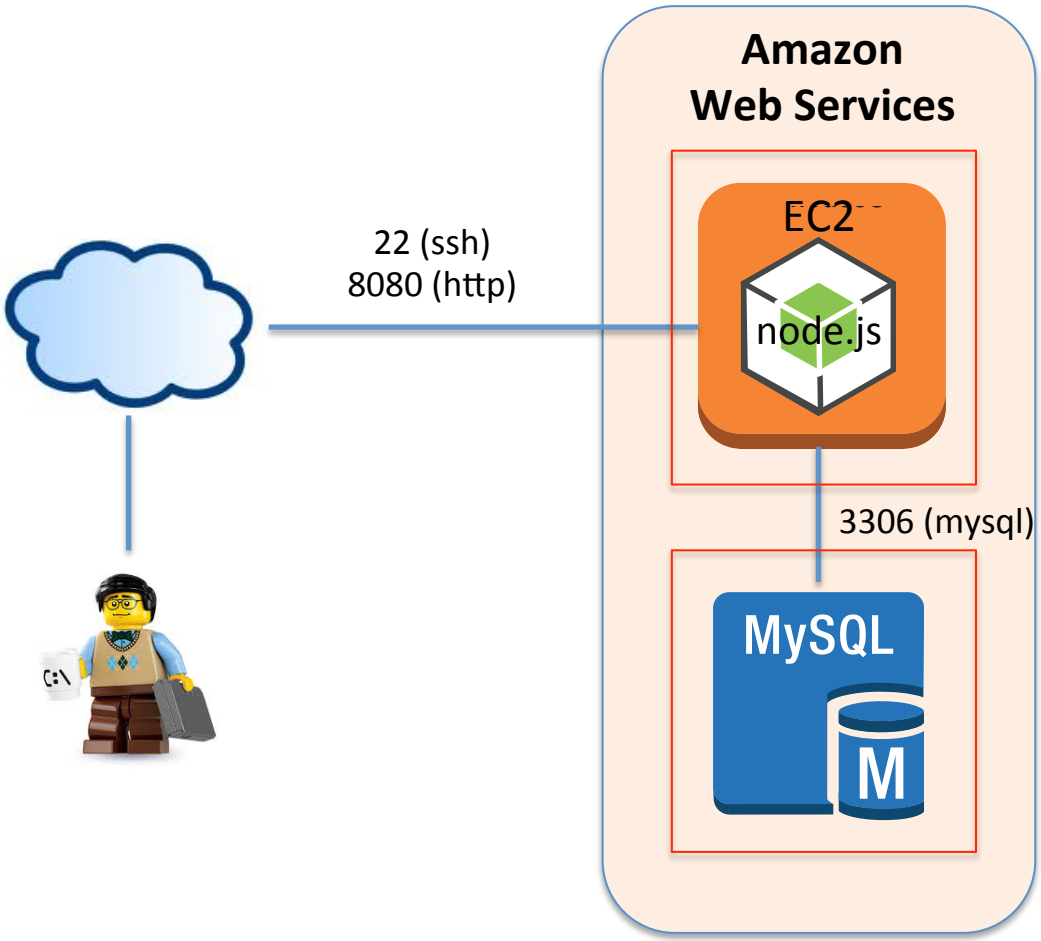
Security Groups

- Firewall for EC2 instances
- Up to 5 security groups per network interface
- Up to 50 rules to a security group
- iptables setup on linux boxes with just port 22 open by default.

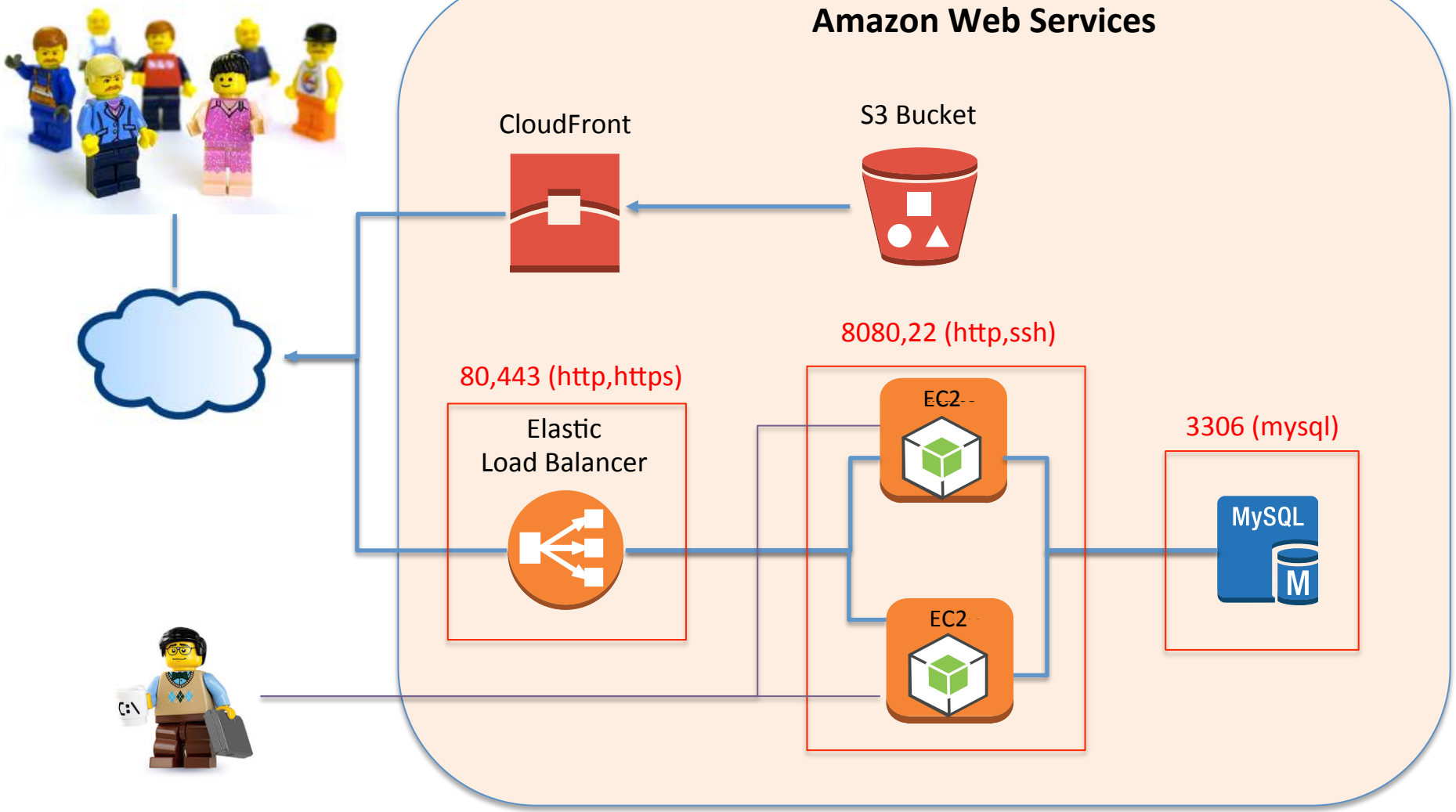
RDS

- Relational Database Service (RDS)
- Provision additional storage on-the-fly with zero downtime
- Automatic patches, backs up and replication
- Multi-AZ – synchronous replication and automatic failover to standby instance
- Read Replicas for read-heavy workloads
- Store up to 3TB and IOPS up to 30,000

Development Environment



Production Environment



Elastic Load Balancer

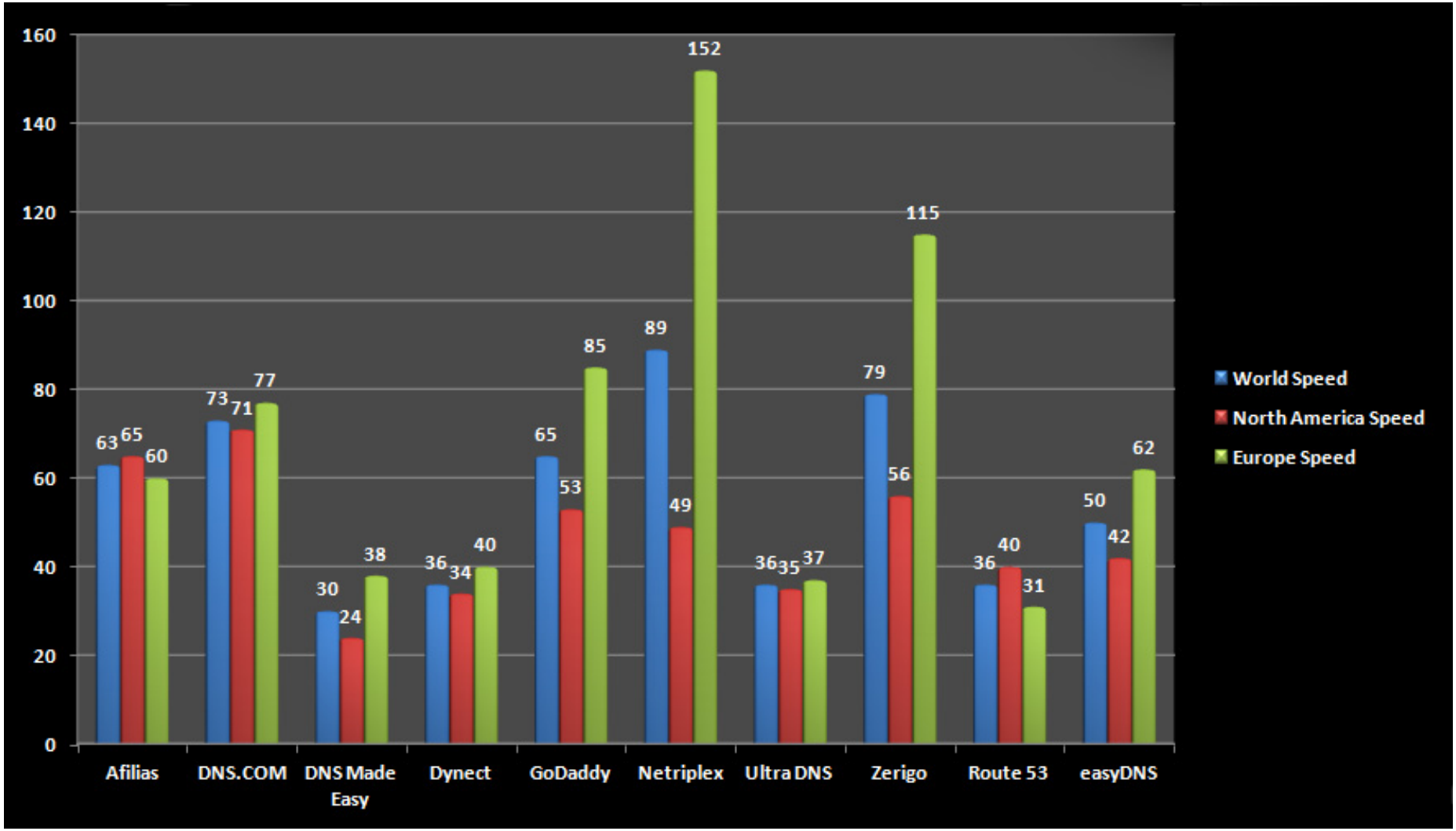
- Load balances requests across multiple EC2 instance and multiple Availability Zones
- Support SSL termination
- Supports sticky sessions
- Allows for Auto Scaling
- Up to 10 per region

Route 53

DNS service

- Latency Based Routing
- Weighted Round Robin
- DNS Failover
- AWS integration (ELB, CloudFront, S3)
- 100% SLA
- Servers in United States(14), Europe(10), Asia(9), Australia(1), South America(2)

Route 53


160
140
120
100
80
60
40
20
0
63
65
60
73
71
77
30
24
38
36
34
40
65
53
85
89
49
152
36
35
37
79
56
115
36
40
31
50
42
62
Afilias
DNS.COM
DNS Made Easy
Dynect
GoDaddy
Netriplex
Ultra DNS
Zerigo
Route 53
easyDNS
World Speed
North America Speed
Europe Speed

S3

Simple Storage Service

- 99.999999999% durability, with 99.99% availability
- File size can range between 1 byte to 5 terabytes, number of files is unlimited
- HTTP & BitTorrent protocols, REST & SOAP API

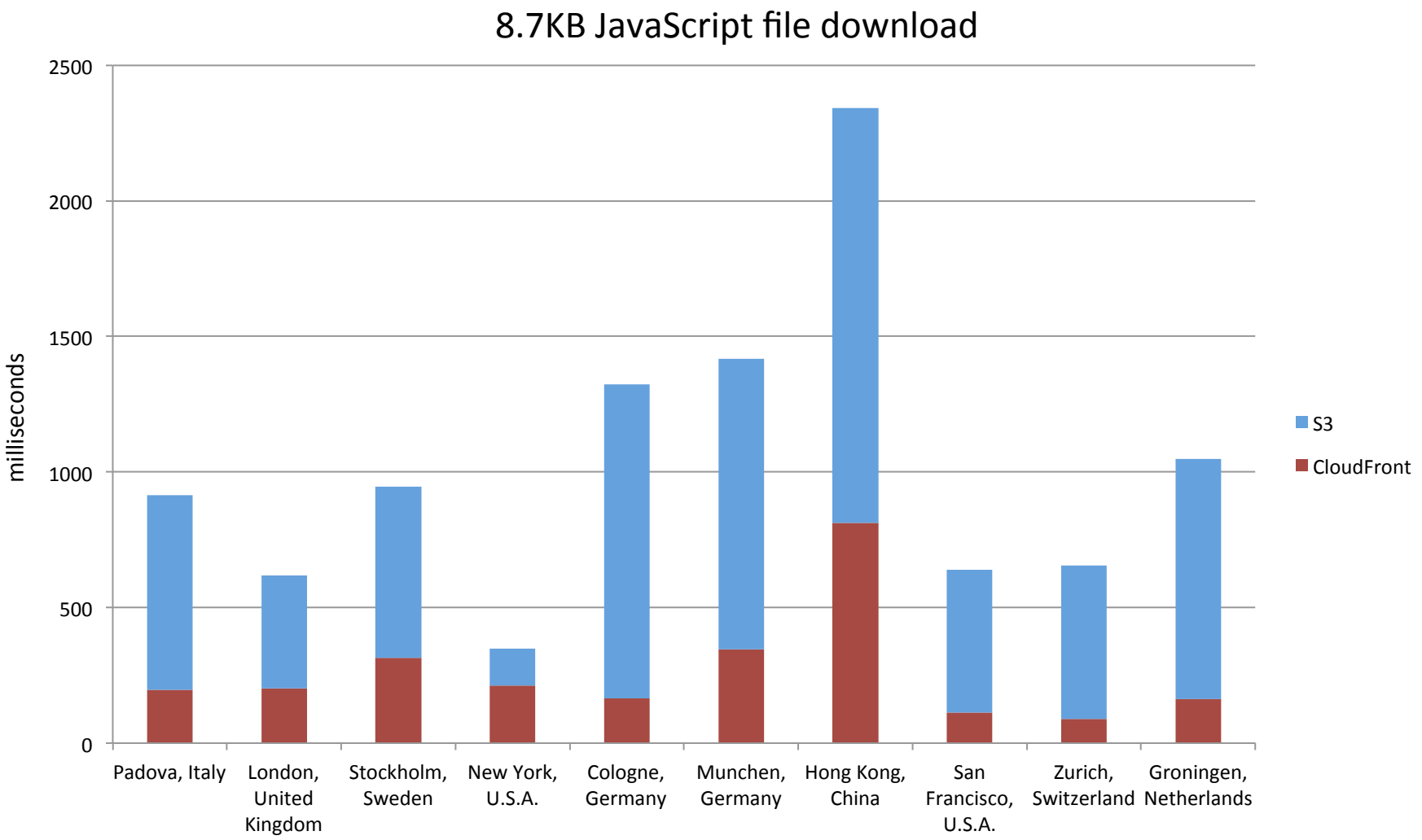
Reduced Redundancy Storage (RRS) & Amazon Glacier

CloudFront

Content Delivery Network (CDN)

- Custom SSL support
- Private Content
- Geo Restrictions
- Custom Error Responses
- Live Streaming

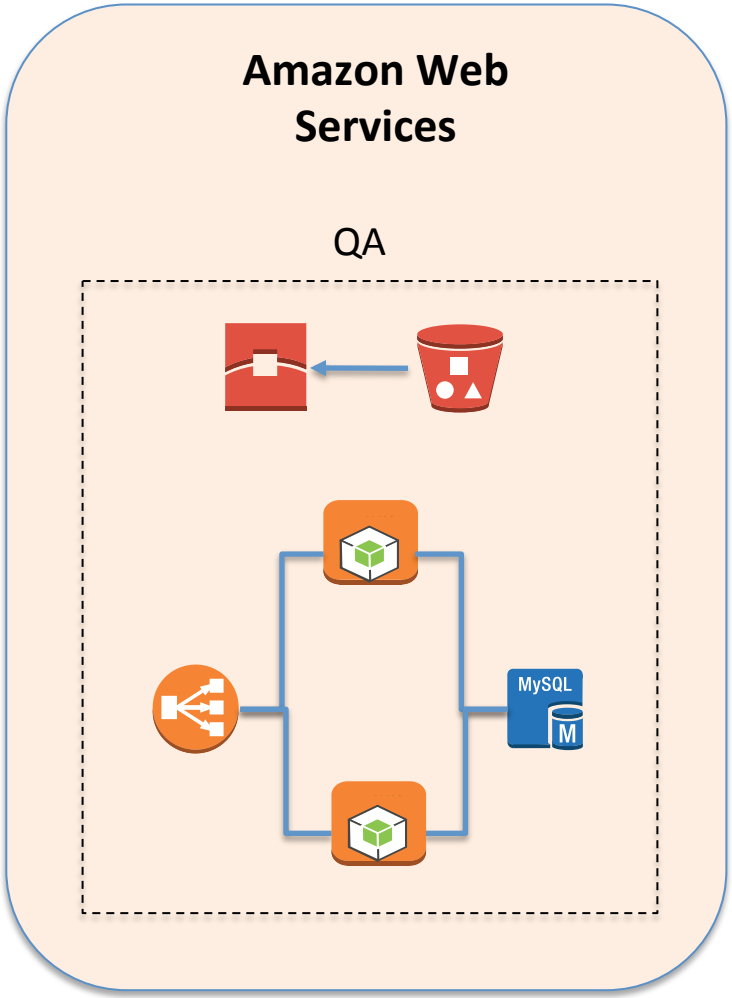
CloudFront


8.7KB JavaScript file download
milliseconds
2500
2000
1500
1000
500
0
Padova, Italy
London, United Kingdom
Stockholm, Sweden
New York, U.S.A.
Cologne, Germany
Munchen, Germany
Hong Kong, China
San Francisco, U.S.A.
Zurich, Switzerland
Groningen, Netherlands
S3
CloudFront

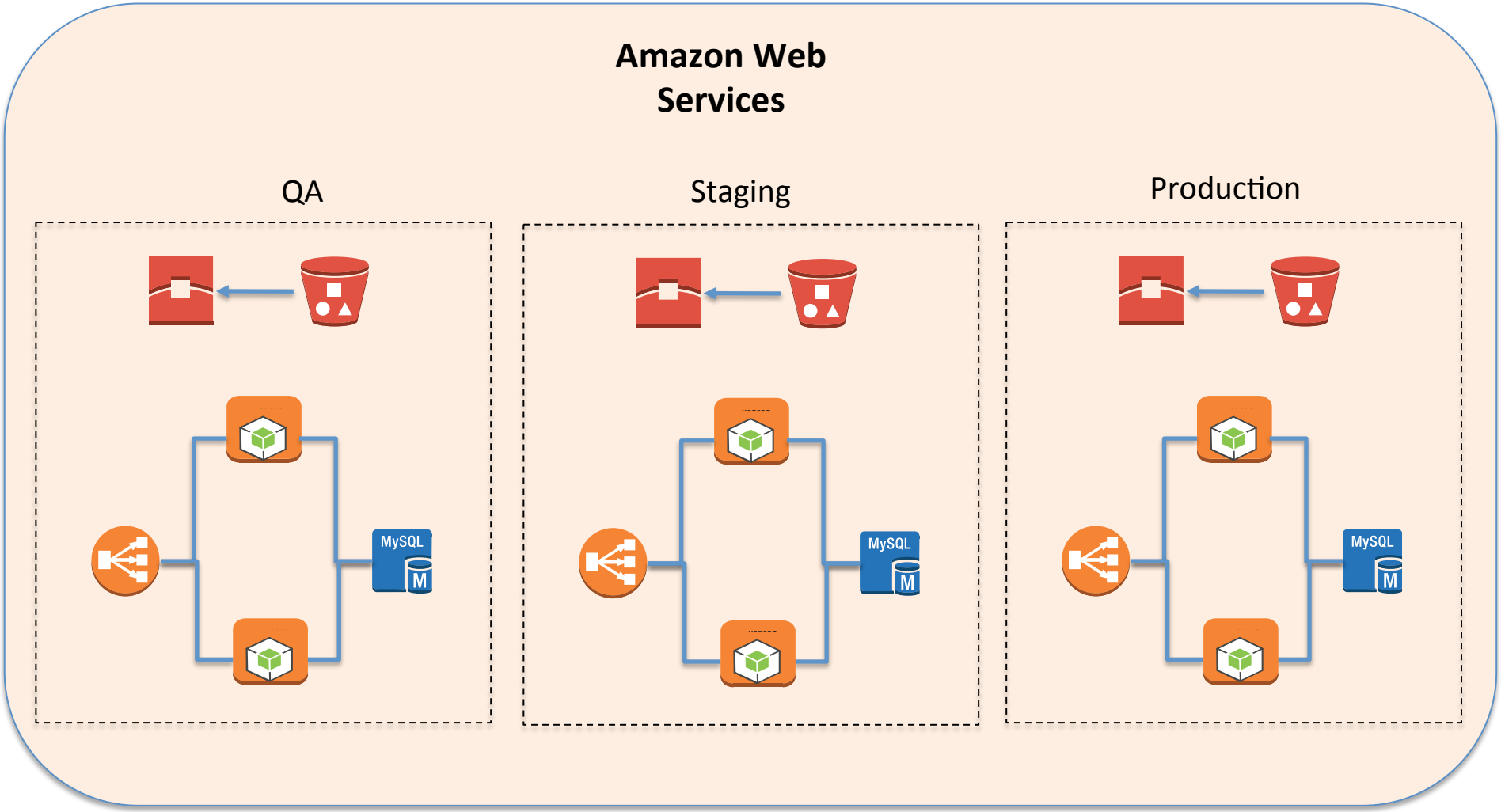
CloudFormation

- Template for creating a Stack that includes AWS resources
- Ability to update or delete entire Stack
- Automatic rollback on failures
- CloudFormer for creating Stack from existing resources

CloudFormation



CloudFormation



What is all this going to cost?

- Many factors go into cost…
  - **Region** (Virginia, California, Ireland,…)
  - **AMI License** (Linux, RHEL, SUSE,Ubuntu,…)
  - **Type** (m1.small, m3.medium,c3.large,…)
  - **EBS** (size(GB), Standard/Optimized, IOPS,…)
  - **Monitoring** (CloudWatch)
  - **Tenancy** (Shared/Dedicated)
  - **Data Transfer** (IN/OUT)

AWS Costs Cheat Sheet

Storage

Instance Storage	N/A
EBS Standard	10c /gb-month (provisioned)
EBS PIOPS	12.5c /gb-month (provisioned)
EBS Requests	10c per million
EBS Provisioned IOPS	10c /month (2.628m requests)
EBS Snapshot	12.5c /gb-month (stored)
S3 Standard	12.5c /gb-month (stored)
S3 Reduced Redundancy	9.3c /gb-month (stored)
Glacier	1c /gb-month (stored)

Regional Cost Differences

Region	Variation
US East (Virginia)	BASE
US West (Oregon)	0%
US West (California)	12.5%
EU (Ireland)	8%
Asia (Singapore)	8%
Asia (Toyko)	15%
South America (Sao Paulo)	45%

Instance Sizes

API name	Compute	Memory	Hourly Cost
t1.micro	< 2	0.613	0.02
m1.small	1	1.7	0.08
m1.medium	2	3.75	0.16
m1.large	4	7.5	0.32
m1.xlarge	8	15	0.64
m2.xlarge	6.5	17.1	0.45
m2.2xlarge	13	34.2	0.9
m2.4xlarge	26	68.4	1.8
c1.medium	5	1.7	0.165
c1.xlarge	20	7	0.66
cc1.4xlarge	33.5	23	1.3
cc2.8xlarge	88	60.5	2.4
cg1.4xlarge	33.5	22	2.1
hi1.4xlarge	35	60.5	3.1

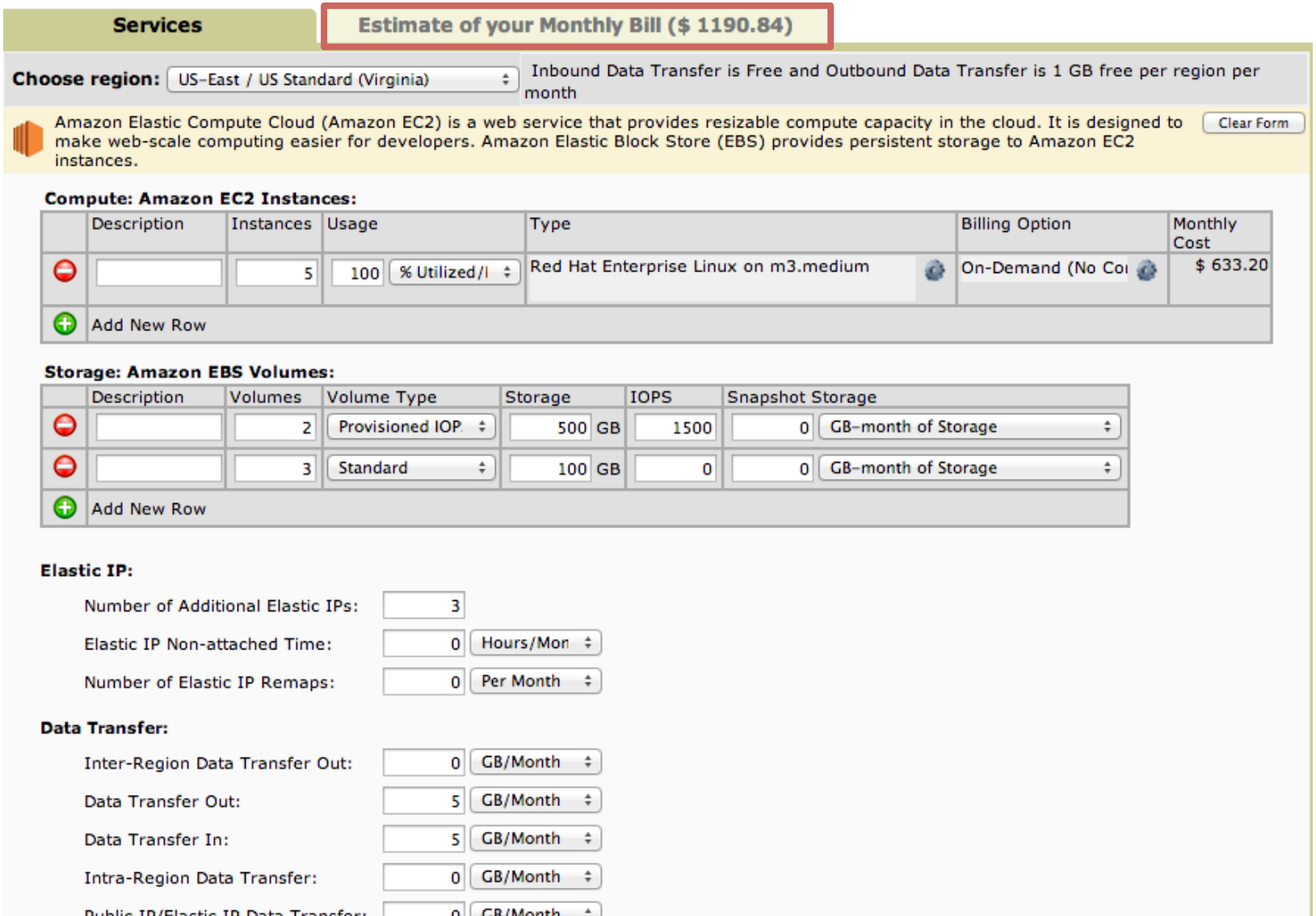
Network

Data IN	FREE
Data OUT	12c /gb
Data within AZ	FREE
Data within Region	1c /gb
Managed Data (EIP/ELB)	1c /gb
Load Balancer	2.5c /hour
Load Balanced Traffic	.8c /gb
IP Address Per Instance	FREE
Extra or Unused IP Address	.5c /hour

Time Key

Day = 24 Hours
Week = 168 Hours
Month = 730 Hours
Year = 8,760 Hours
Month = 2,628,000 Seconds

AWS Simple Monthly Calculator



EC2 Purchase Options

- On-Demand
  - Pay as you go at hourly rate
  - Nothing up front, no commitment
- Reserved
  - 1-3 year commitment
  - One-time upfront payment, reduced hourly rate
- Spot
  - Bid for hourly rate as becomes available

Reserved Instances - Utilization

3-Year RI Percentage Savings Over On-Demand Comparison*

Annual Utilization	Light Utilization RI	Medium Utilization RI	Heavy Utilization RI
20%	**25%**	-7%	-77%
40%	**40%**	33%	11%
60%	45%	**46%**	41%
80%	48%	52%	**56%**
100%	49%	59%	**65%**

* Rates are compared for an m1.xlarge instance 3-year Reserved Instance, % savings on effective hourly rates are roughly the same for all instance types

Heavy Utilization Pricing

	1-Year Term		3-Year Term	
	Upfront	**Hourly**	**Upfront**	**Hourly**
General Purpose - Current Generation				
m3.medium	$317	$0.026 per Hour	$481	$0.022 per Hour
m3.large	$633	$0.053 per Hour	$961	$0.043 per Hour
m3.xlarge	$1266	$0.105 per Hour	$1922	$0.086 per Hour
m3.2xlarge	$2531	$0.209 per Hour	$3844	$0.172 per Hour

DIY vs On-Demand vs Reserved

Annual Cost Comparison (100% utilization)

	Do-It-Yourself	EC2 On-Demand	EC2 Reserved (1 Year Term)	EC2 Reserved (3 Year Term)
Usage Costs	-	$ 157,680	$ 75,411	$ 48,123
Server Hardware	$ 20,129	-	-	-
Network Hardware	$ 4,026	-	-	-
Hardware Maintenance	$ 28,986	-	-	-
Operating System	$ –	-	-	-
Facility Expense	$ 131,382	-	-	-
Remote Hands Support	$ 1,014	-	-	-
Data Transfer Costs	$ 10,071	$ 6,138	$ 6,138	$ 6,138
TOTAL COST	**$ 195,608.00**	**$ 163,818.00**	**$ 81,550.00**	**$ 54,263.00**

References & Resources

- Entire AWS Web Site - http://aws.amazon.com/
- EC2 and EBS Performance - http://www.slideshare.net/AmazonWebServices/stg302-28617072
- ECU CPU Benchmarking - http://blog.cloudharmony.com/2010/05/what-is-ecu-cpu-benchmarking-in-cloud.html
- Route 53 speed comparison - http://www.dnscomparison.com/route53.html
- CloudVertical cost cheat sheet - https://blog.copper.io/aws-cost-cheat-sheet-2/
- Simple Monthly Calculator - http://calculator.s3.amazonaws.com/index.html
- How AWS Pricing Works - http://d36cz9buwru1tt.cloudfront.net/AWS_Pricing_Overiew.pdf
- Beer Survey App - http://beersurvey.org

Q&A

or

Go drinks some beer?