http://bit.ly/2ErwJa5

Kombinot on AWS - Running Beyond Cost Effective

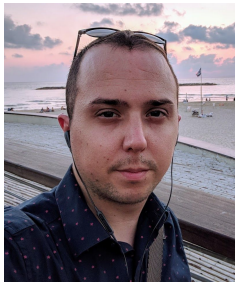
YOT PO.

datree.io

SimilarWeb

ironSource

MEET THE TEAM



Shimon Tolts

Arthur Schmunk

Tal Hibner

Niv Yungelson

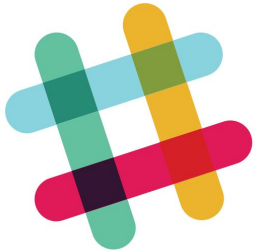
Doron Rogov

Boaz Ziniman

AWS Israel Community

- Founded - Feb **2013**
- **78** meetups with **~6000** Members
- Monthly meetups
- No Marketing, No bullshit
- Monthly Highlight (News)
- All AWS: AI, BigData, Serverless, Containers, etc

Join the Community!



http://bit.ly/2ErwJa5

https://www.meetup.com/AWS-IL/

https://www.meetup.com/AWS-IL/

July meetup - **Big Data on AWS**

Join us on **July 16** in a new AWSome location.

Want to be a speaker? - Come talk to us or E-mail us at:
Niv - lnivy1@gmail.com Or Shimon - shimon@datree.io

AWS News

- You Can Simplify Login with Application Load Balancer Built-in Authentication
- Amazon ECS Adds Options to Speed Up Container Launch Times
- Amazon EKS – Now Generally Available
- Redis 4.0 Compatibility in Amazon ElastiCache

Kombinot on AWS - Running Beyond Cost Effective

- ironSource - ELB hacks - Avi Keinan - Devops
- Yotpo - Cloudfront vs DIY CDN - Andrei Burd - DevOps Team Lead
- SimilarWEB - Migrating Hadoop to AWS - Liran Vaknin - Production Engineer

Questions?

Ask in Slack #kombinot_on_aws

http://bit.ly/2ErwJa5

Cut your AWS billing and make your CEO happy

Avi Keinan, DevOps @ ironSource

ironSource

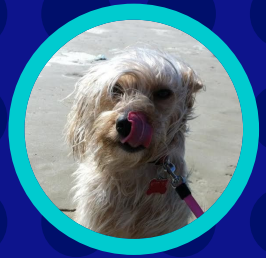
ironSource employees in numbers

Women

36%

Singles

51%



Dogs

60

Couples

8

Tel Avivians

44%

Average Age

31

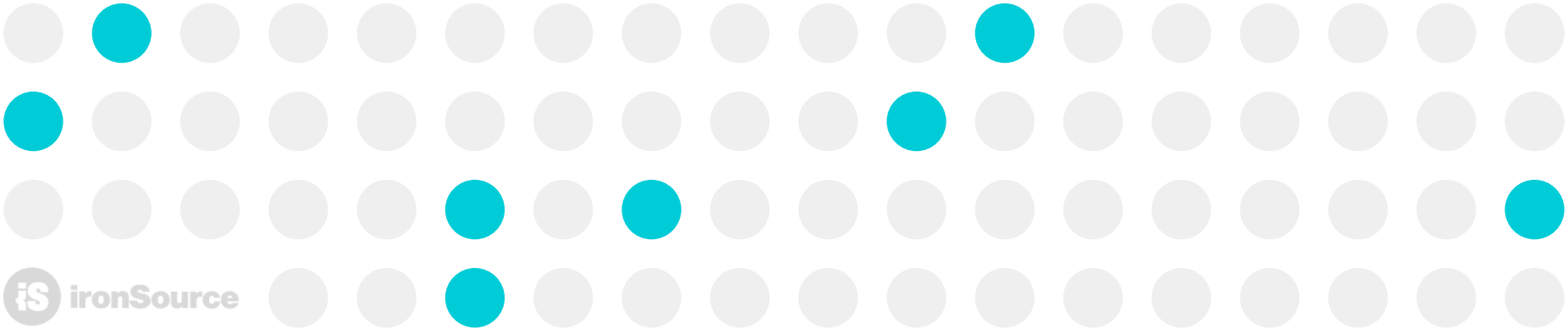
Youngest & Oldest

21/54

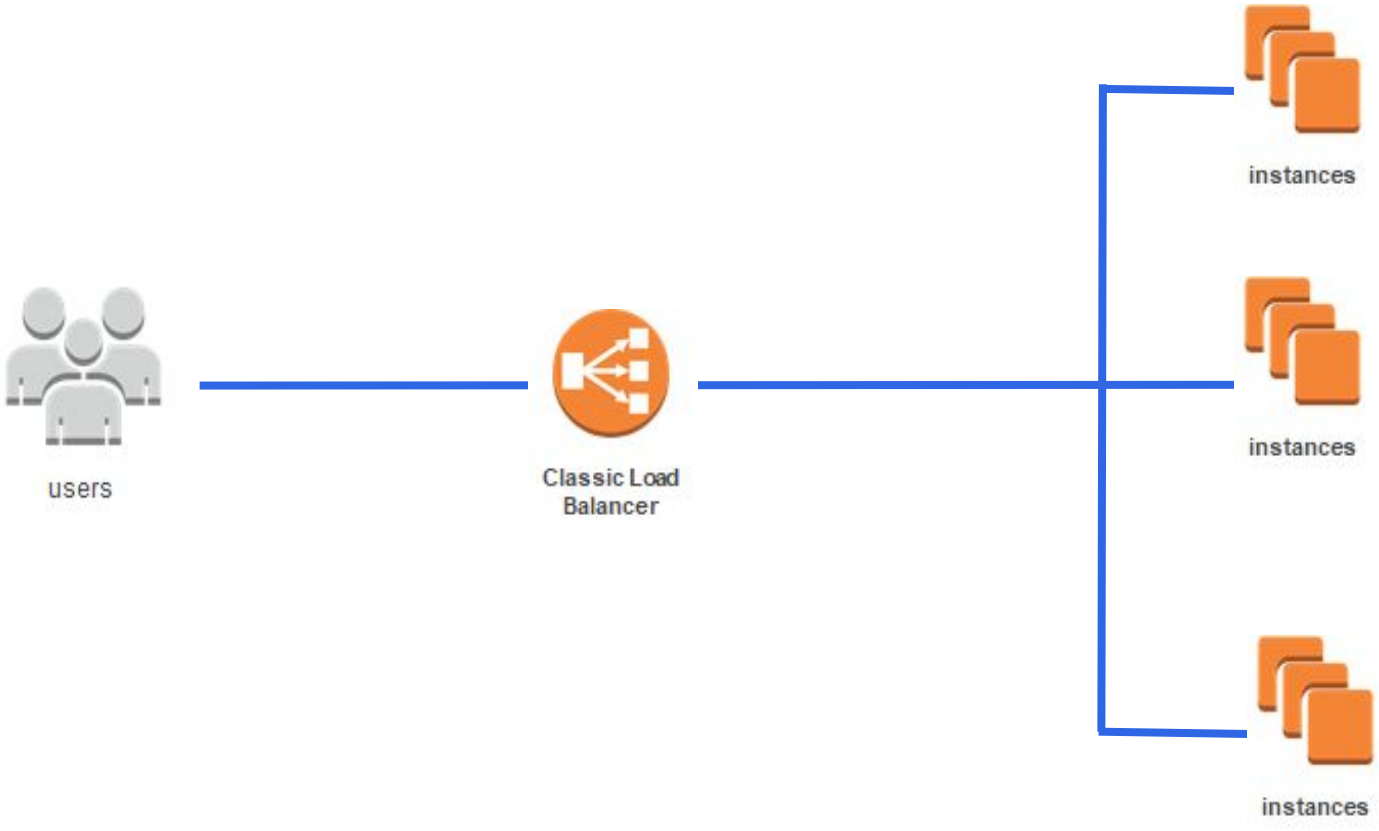
ironSource

Option #1

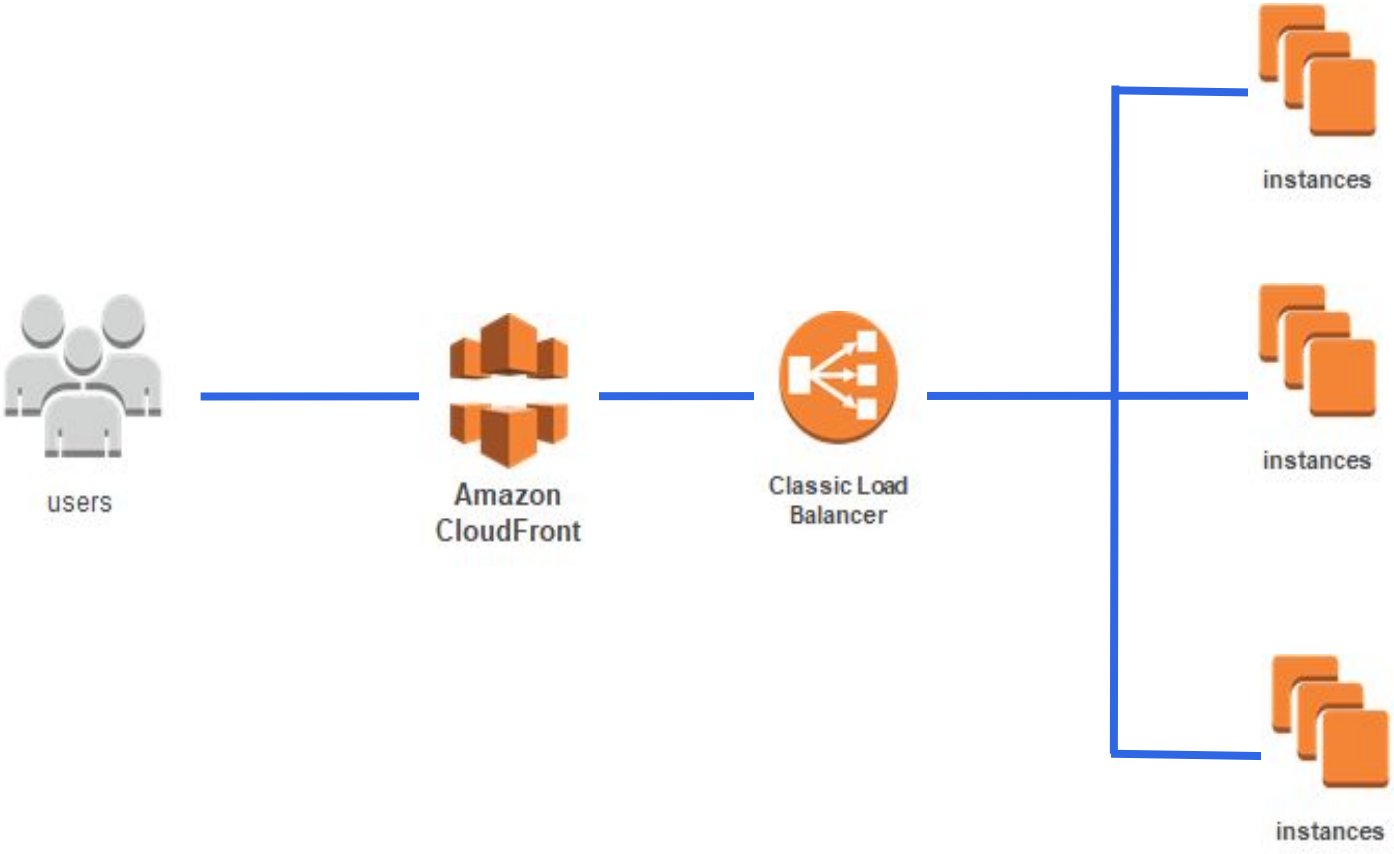
CloudFront in front of your Load Balancer



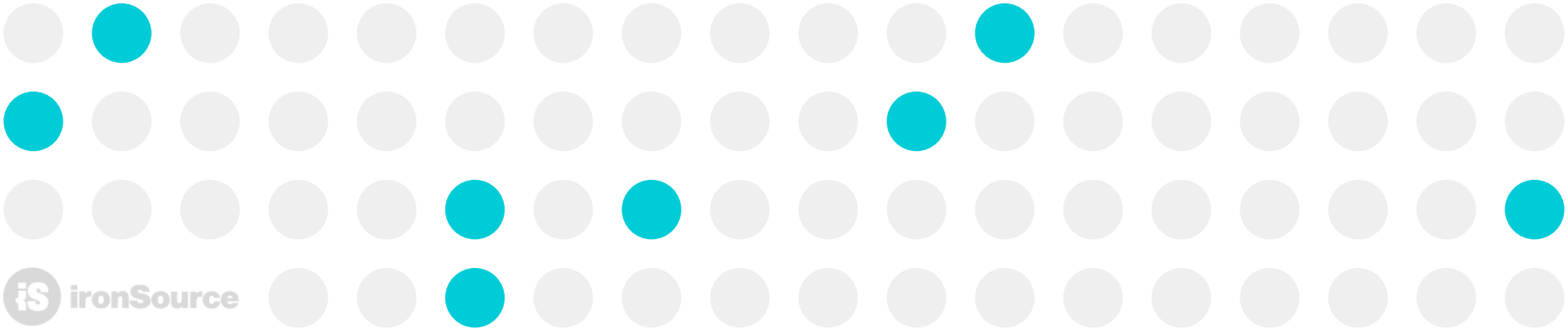
The usual way



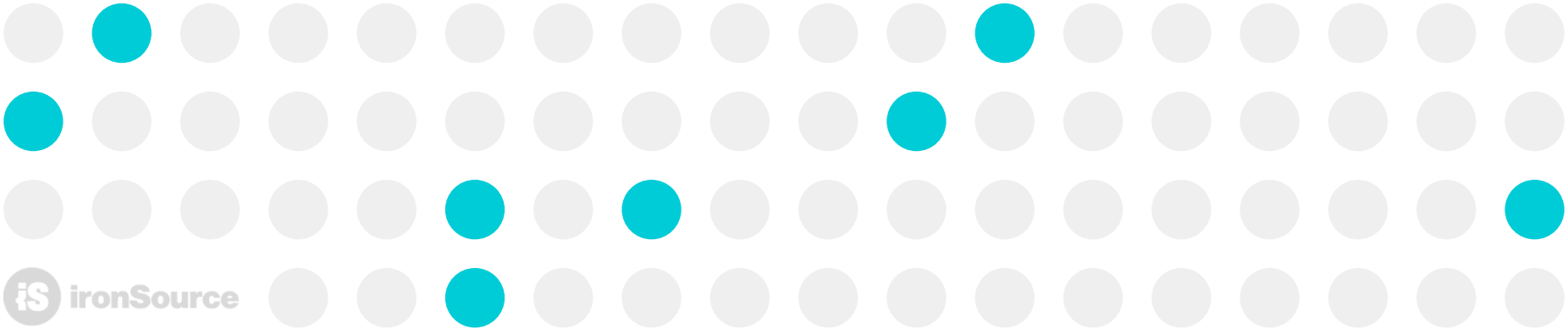
CloudFront in front of your Load Balancer



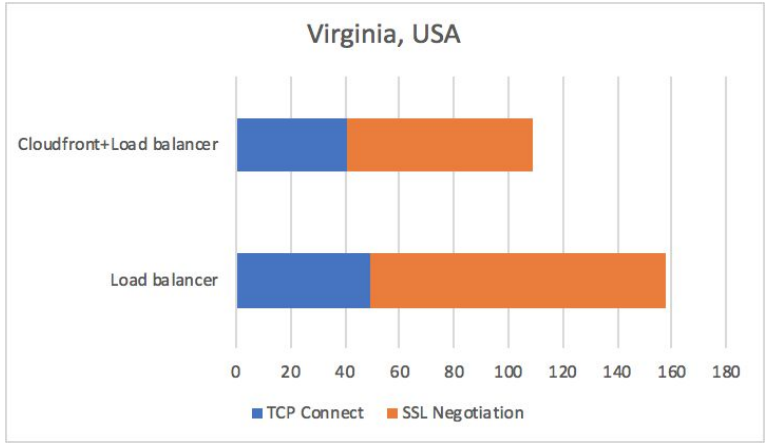
Why?

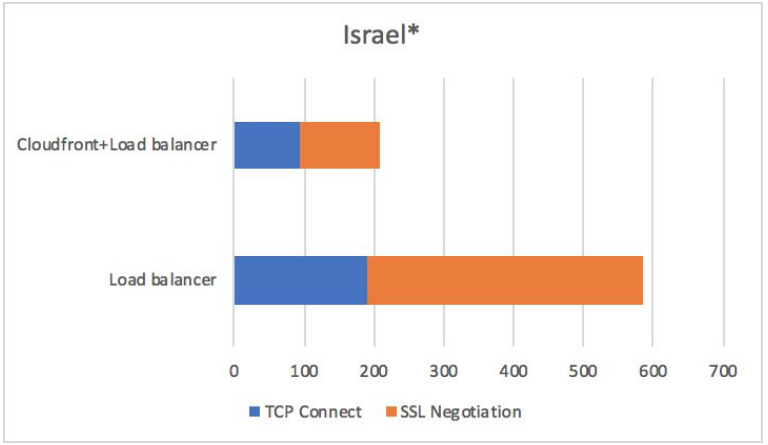


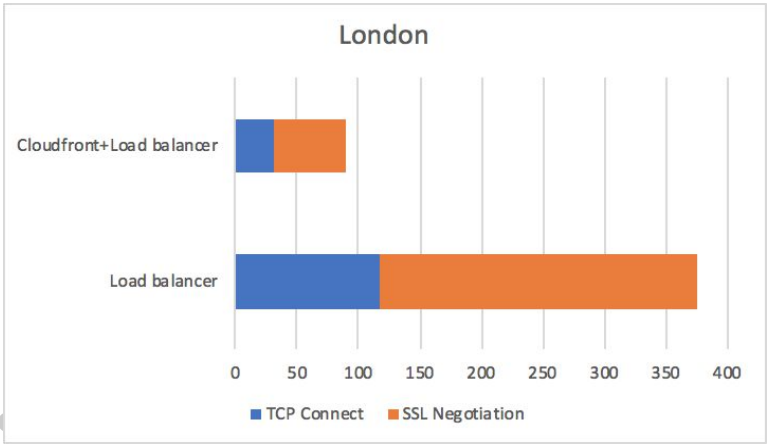
1. Performance

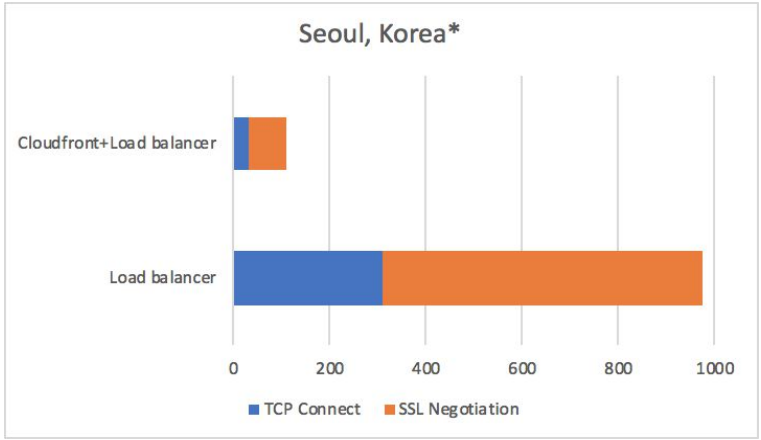


Performance comparison: ELB vs CloudFront&ELB (time in ms)

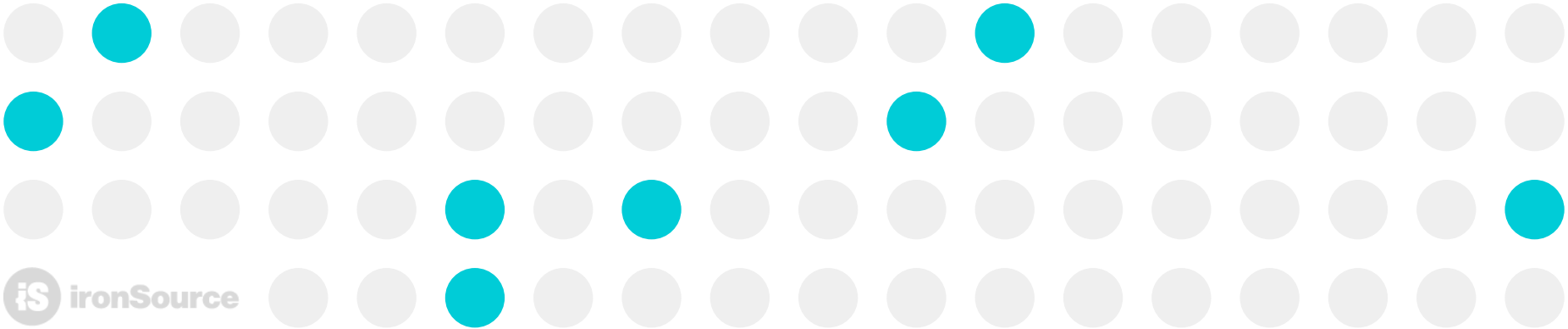








2. Price

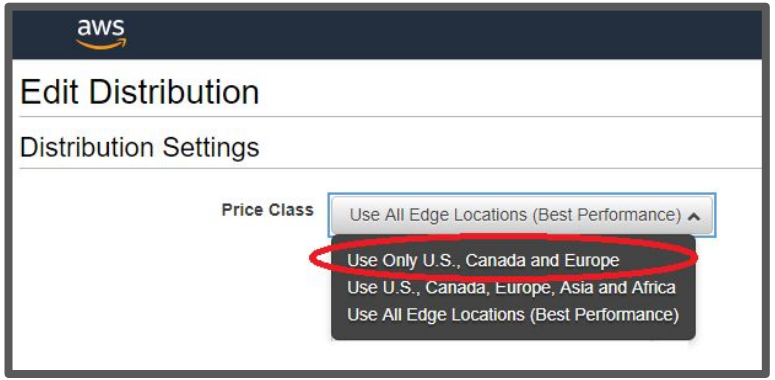


Cloudfront pricing

Reduce your data transfer costs by up to 16%

Depends on:

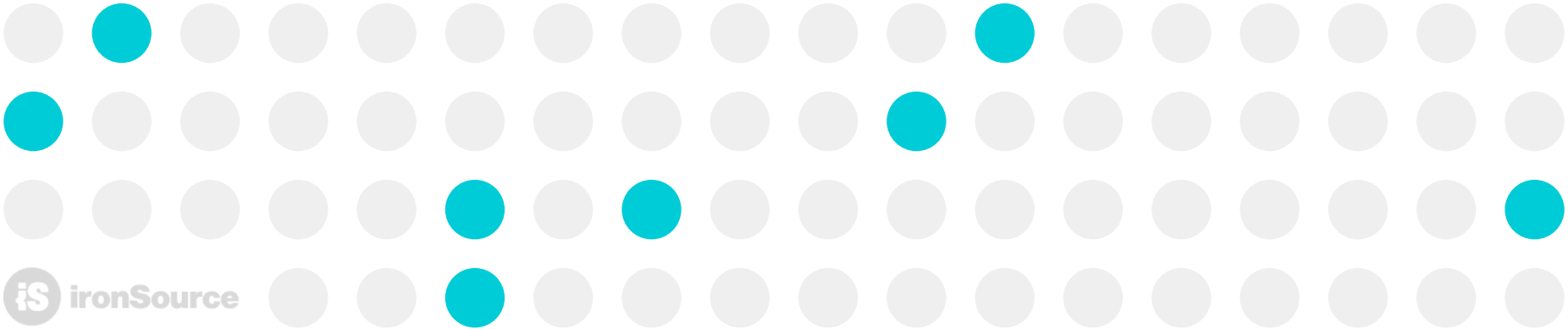
1. **Amount of requests: 1M HTTPS requests = 1$.** (Average object size).
2. **CloudFront Price Class (Cheapest - US/EU/CA):**



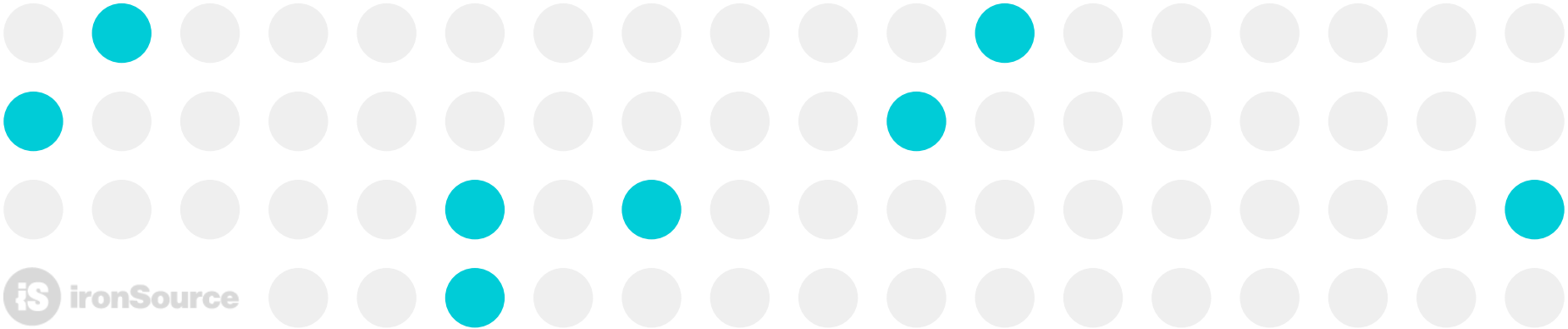
3. **Reserved capacity - Contact AWS Sales.**

Option #2

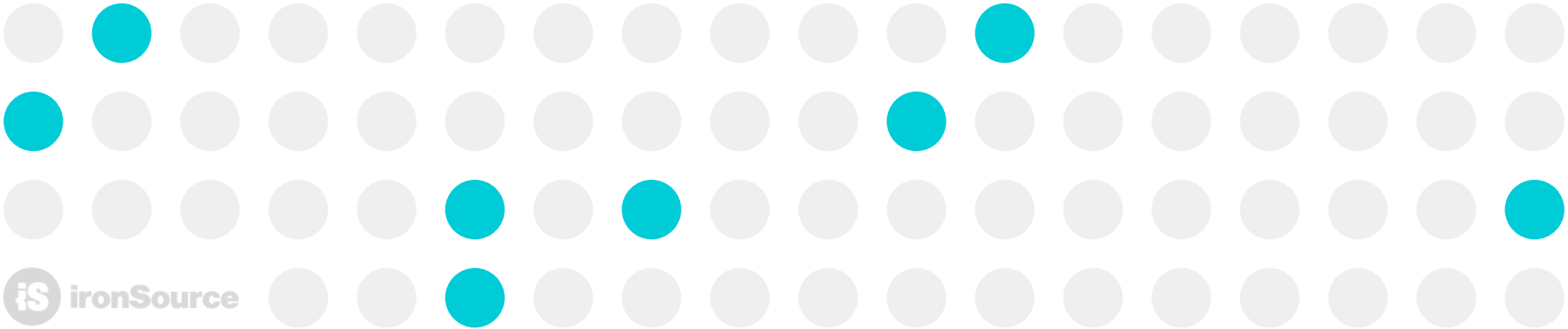
New load balancer methods: RLB & NLB



Why?



1. Price



AWS ELB price:

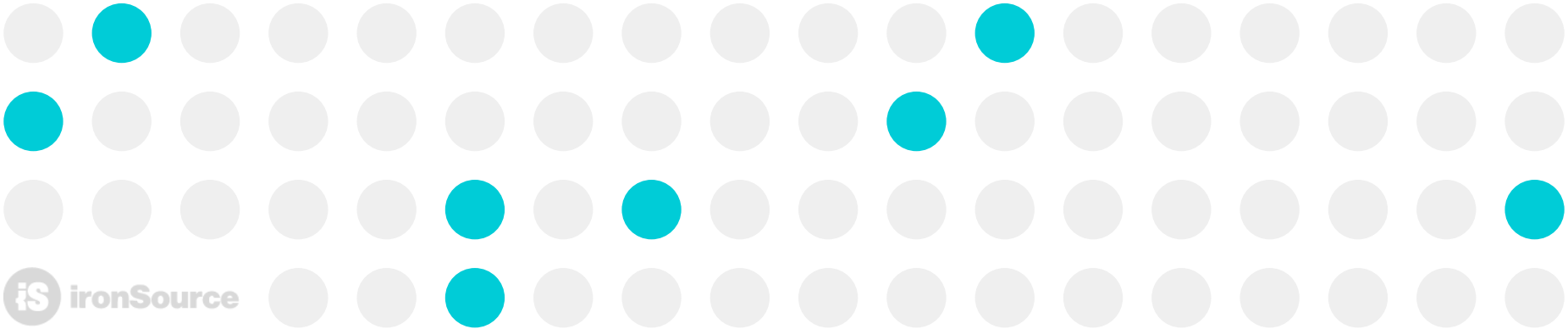
Classic Load Balancer Pricing

$0.008 per GB of data processed by a Classic Load Balancer

Huge amount of traffic = Tens of thousands of dollars

RLB?

Roman's Load Balancer



RLB (Proof of concept)

1. All instances are spots and have elastic ips.
2. Users go directly to the instances.
3. When AWS reclaims a spot instance, its elastic ip is reattached to on demand machine with multiple nics.
4. Because we don’t use any load balancers we end up saving a lot of money.

Spot Instance

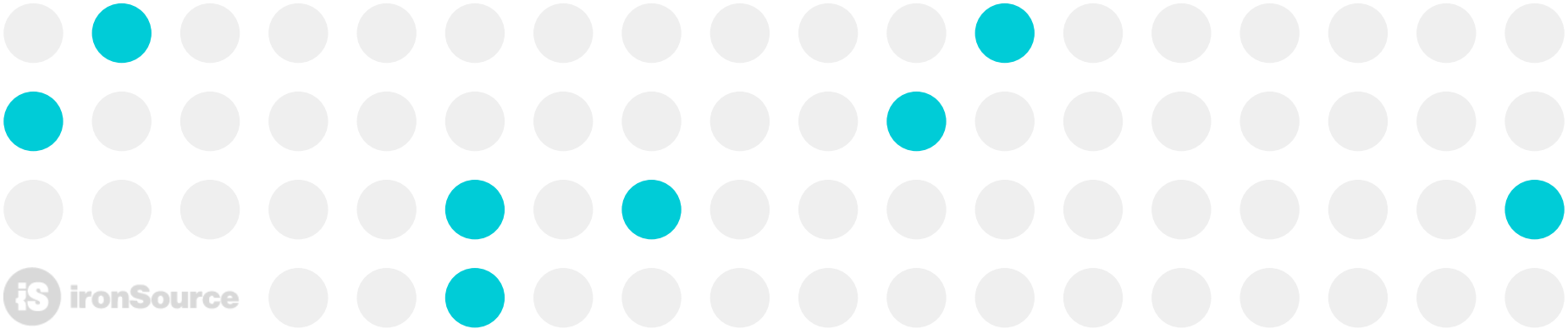
optimized instance

Amazon Route 53

NLB?

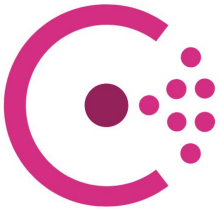
Noy's Load Balancer

(not Network load balancer)



NLB (Production)

1. Used in one of our event logging systems.
2. HAProxy spot instances in autoscale environment.
3. Machine ips are registered in Route 53 using HashiCorp Consul.
4. When AWS reclaims a spot instance, our Client SDK will keep trying to send the event until it's acknowledged.



Q&A

Andrei Burd
DevOps Team Lead @

YOTPO.

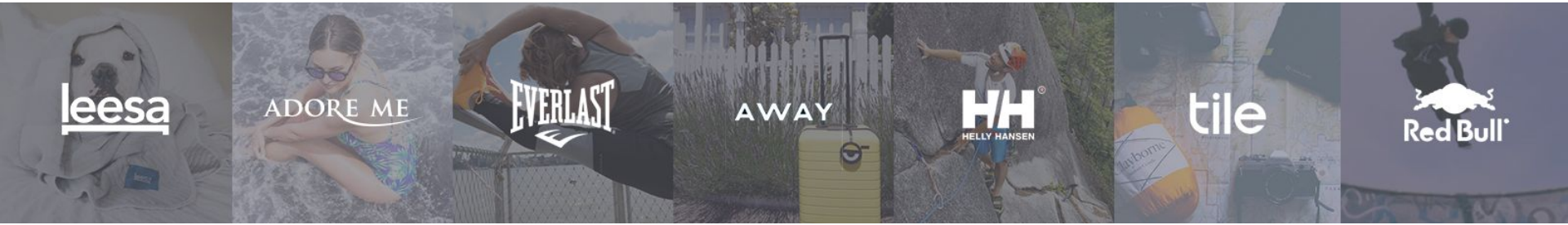
bandrei@yotpo.com

burdandrei

burdandrei


YOT
PO.

Yotpo in a nutshell



100%
YOY growth

300K
Clients
(30K+ Paying)

200M
Shoppers
(350M During Nov Sales)

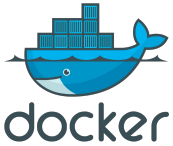
290+
People in TLV, NYC
& London

$101M+ Raised

MARKER LLC · INNOVATION ENDEAVORS · BLUMBERG CAPITAL · BESSEMER VENTURE PARTNERS · Vintage Investment Partners

YOT PO.

Stack



AWS

- Networking
- Compute
- Storage
- Monitoring
- Security

HashiCorp

- Consul
- Packer
- Terraform
- Vault
- Nomad

Docker

- ECR

CI/CD

- Jenkins
- Git
- VCS
- jFrog
- Travis

Monitoring

- Elasticsearch
- Logstash
- Kibana
- Grafana
- Telegraf
- InfluxDB

YOTPO.

Free Cheese

Google Analytics

Google Analytics Collection Limits and Quotas

Google Analytics

=

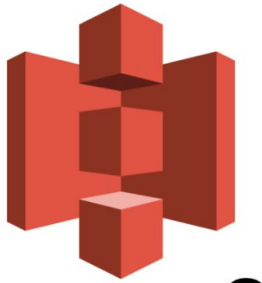
$

https://developers.google.com/analytics/devguides/collection/analyticsjs/limits-quotas

YOT PO.

Open Source

Open Source



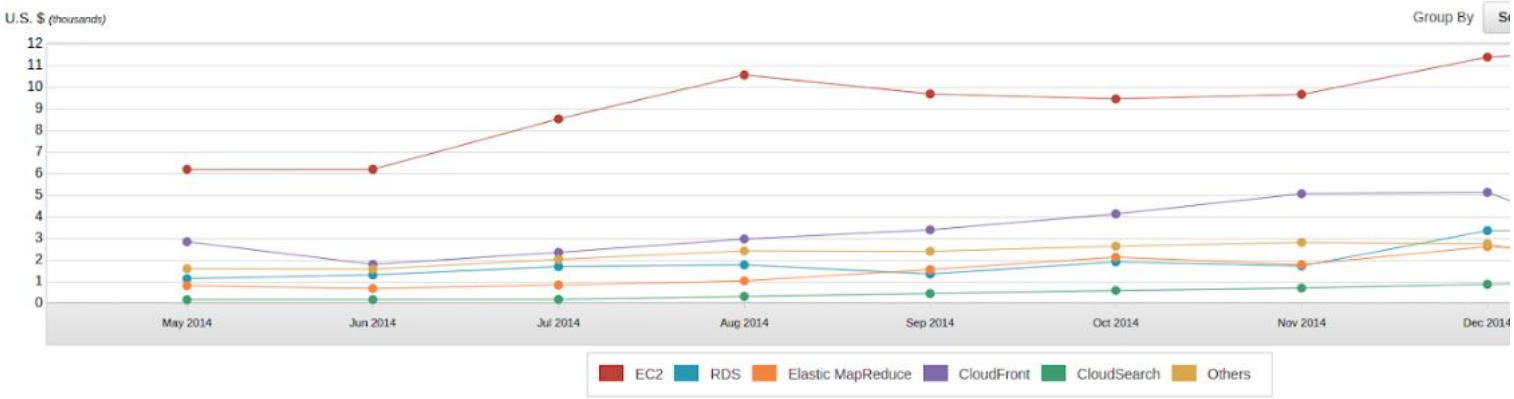
Amazon S3

YOT PO.

Payday

SNOWPLOW

amazon cloudfront



YOT PO.

Fair price?

Amazon CloudFront Pricing

Request Pricing for All HTTP Methods (per 10,000)

$0.0075 - $0.0220

File:Transparent.gif

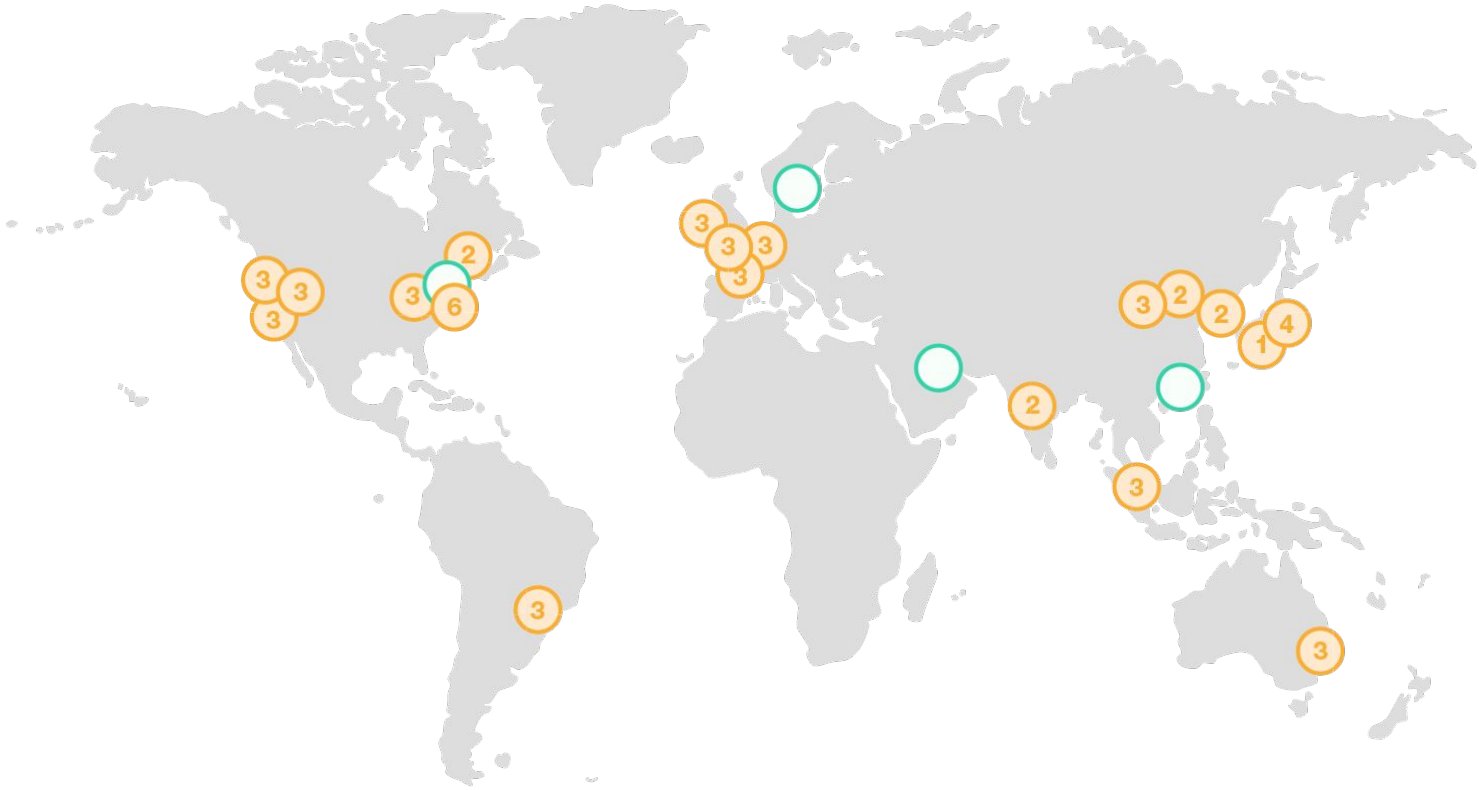
From Wikimedia Commons, the free media repository

No higher resolution available.

Transparent.gif (1 × 1 pixels, file size: 42 bytes, MIME type: image/gif)

YOTPO.

Options



Analytics

The `ngx_http_empty_gif_module` module emits single-pixel transparent GIF.

Example Configuration

```
location = /_.gif {
    empty_gif;
}
```

Analytics

7 Bytes/request = $

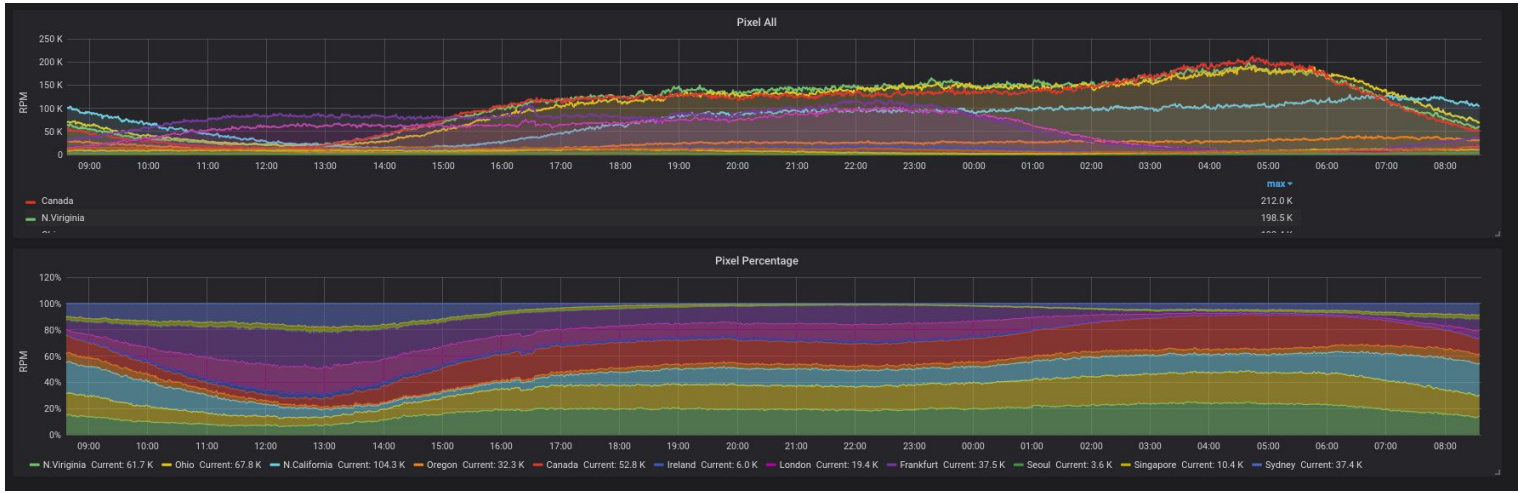
```
bandrei@Bandrei-T460[15:33:44][last: 0s][~] ~$ curl -i http://p.yotpo.com/aws-user-group-israel/combinot
HTTP/1.1 200 OK
Content-Type: image/gif
Date: Sun, 03 Jun 2018 12:34:29 GMT
Server: nginx
Set-Cookie: pixel=dd89f87b-c2f1-4a39-5d54-d4b277a657ce; Path=/; Domain=yotpo.com; Max-Age=31536000; HttpOnly
Content-Length: 35
Connection: keep-alive
```

Route53

p.yotpo.com.	A	ALIAS pixel-c5-us-west-2-1568874472.us-west-2.elt	Yes	dcc2a236-2914-4419-888c-347650c59ece	us-west-2
p.yotpo.com.	A	ALIAS pixel-c5-us-west-1-967756041.us-west-1.elb.	Yes	44e494ef-a302-40e8-862b-94776aefaaae	us-west-1
p.yotpo.com.	A	ALIAS pixel-c5-us-east-2-641342301.us-east-2.elb.a	Yes	9ce67d96-58f8-4db0-9d4e-3448e31c6d6a	us-east-2
p.yotpo.com.	A	ALIAS pixel-c5-us-east-1-1970857725.us-east-1.elb	Yes	fe0b9256-54d7-4390-8a83-d9e6e6402f52	us-east-1
p.yotpo.com.	A	ALIAS pixel-c5-sa-east-1-588023405.sa-east-1.elb.a	Yes	5b1b4b12-3ad5-4af1-babf-012077024d5d	sa-east-1
p.yotpo.com.	A	ALIAS pixel-c5-eu-west-2-568835368.eu-west-2.elb.	Yes	b634bb50-007c-4970-9c0b-b8cfca22ee4c	eu-west-2
p.yotpo.com.	A	ALIAS pixel-c5-eu-west-1-1230377036.eu-west-1.ell	Yes	740242a4-8f61-4307-aaa9-a16bfd1471de	eu-west-1
p.yotpo.com.	A	ALIAS pixel-c5-eu-central-1-937032432.eu-central-1	Yes	bff78aee-ce26-41f6-a1ce-5b9a30a41fc8	eu-central-1
p.yotpo.com.	A	ALIAS pixel-c5-ca-central-1-862040907.ca-central-1	Yes	4ee7d278-7584-42df-9dd1-a82ab016a29c	ca-central-1
p.yotpo.com.	A	ALIAS pixel-c5-ap-southeast-2-1785855677.ap-sout	Yes	b5fb5067-4697-4964-91f8-74c05b3c6bab	ap-southeast-2
p.yotpo.com.	A	ALIAS pixel-c5-ap-southeast-1-37982051.ap-southe	Yes	f200e7bf-66c5-462b-a132-2fee5d9feadf	ap-southeast-1
p.yotpo.com.	A	ALIAS pixel-c5-ap-south-1-448955295.ap-south-1.el	Yes	2104e2af-17de-4533-b938-05ca150f0048	ap-south-1
p.yotpo.com.	A	ALIAS pixel-c5-ap-northeast-2-1538821530.ap-nortl	Yes	44eb8d5c-1a5f-441f-95e8-00f89e9b7665	ap-northeast-2
p.yotpo.com.	A	ALIAS pixel-ap-northeast-1-1956858513.ap-northea	Yes	a17de7ee-9093-4067-b769-a517ec1c63af	ap-northeast-1

YOT PO.

24 hour



Kung Foo

74 lines (60 sloc) | 1.19 KB

```
1  module "pixel-us-east-1" {
2    source = "pixel"
3    region = "us-east-1"
4  }
5
6  module "pixel-us-east-2" {
7    source = "pixel"
8    region = "us-east-2"
9  }
10
11 module "pixel-us-west-1" {
12   source = "pixel"
13   region = "us-west-1"
14 }
```

https://github.com/kung-foo/multiregion-terraform

YOT PO.

Today

10K RPS

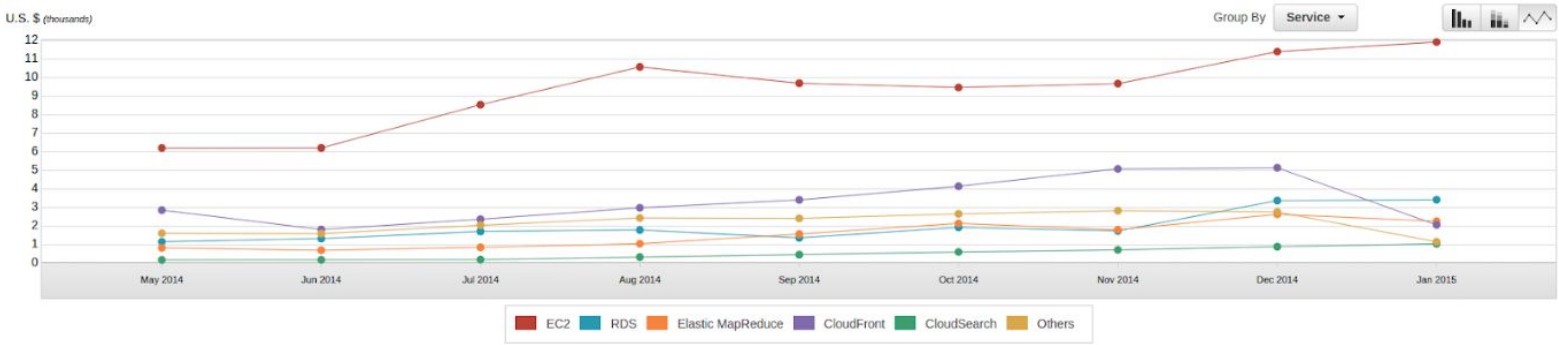
$25K

$1100

14 regions * 2 c5l instances

YOT PO.

Happy CFO

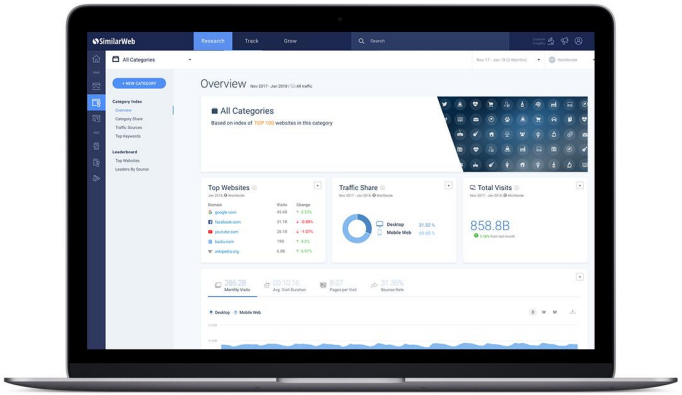


HashiCorp User Group
ISRAEL

Migrating Hadoop to AWS Without Busting the Bank

June, 2018 | Liran Vaknin, Production Engineer

SimilarWeb



POWERING BETTER DECISIONS

SimilarWeb empowers businesses to make better decisions by equipping them with the insights they only dreamed were possible.

About Us

2013

Launched

~400

Employees

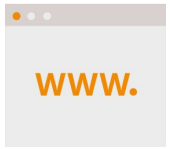
+$112M

Funding

8

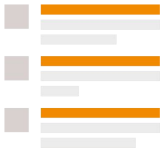
Offices

What We Do



80M
websites

4M
apps



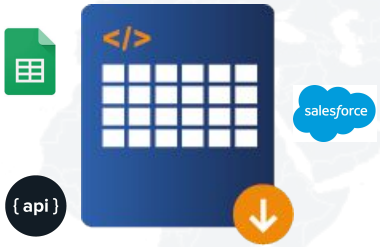
250
categories

190+
countries

Delivering Our Insights

SimilarWeb

PLATFORM



API &
INTEGRATIONS

SERVICES &
CONSULTING

Serving layer already migrated to AWS

Data layer is next!

Our Self-Hosted Cluster

Capacity

- 200 machines
- 13TB of RAM
- 4800 CPU cores
- 5PB of storage

Cost

- 100,000$ per month
- Additional operational cost

Utilization And Limitations

- Paid monthly
- Not very flexible
- Human API
- Manual maintenance

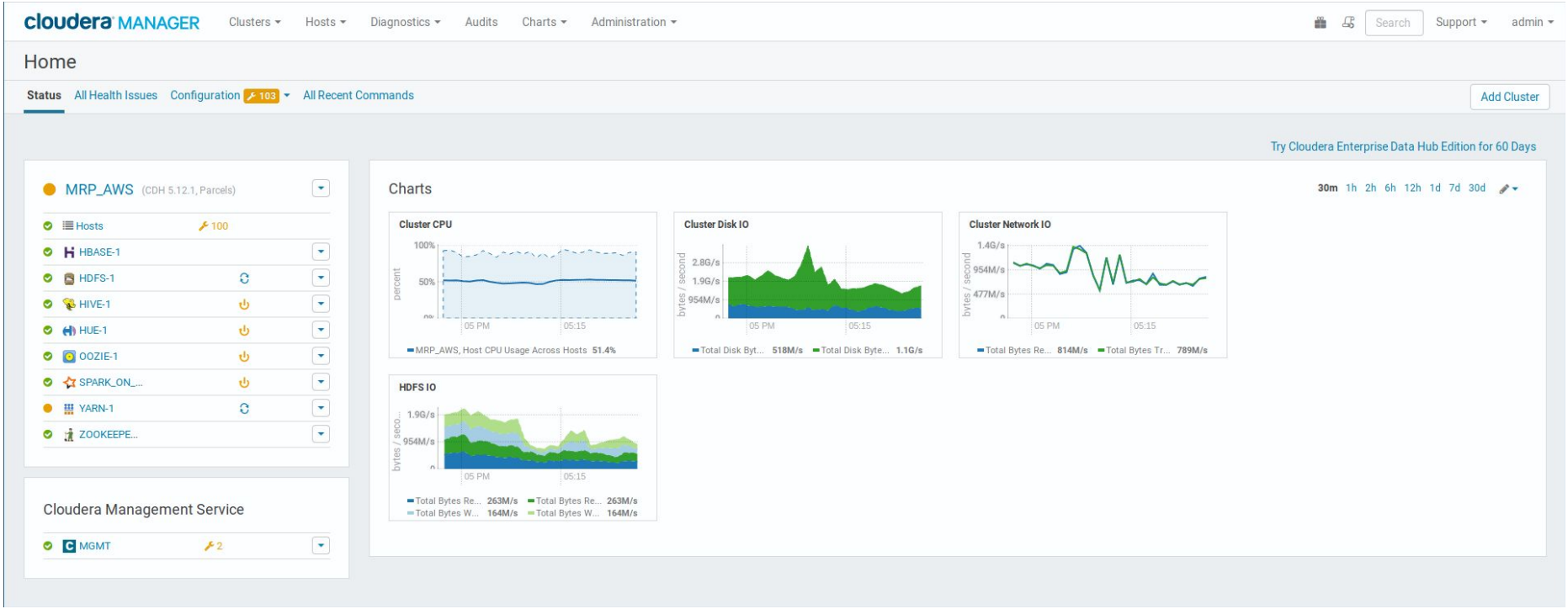
Self-Hosted Cluster Utilization

Apache Hadoop

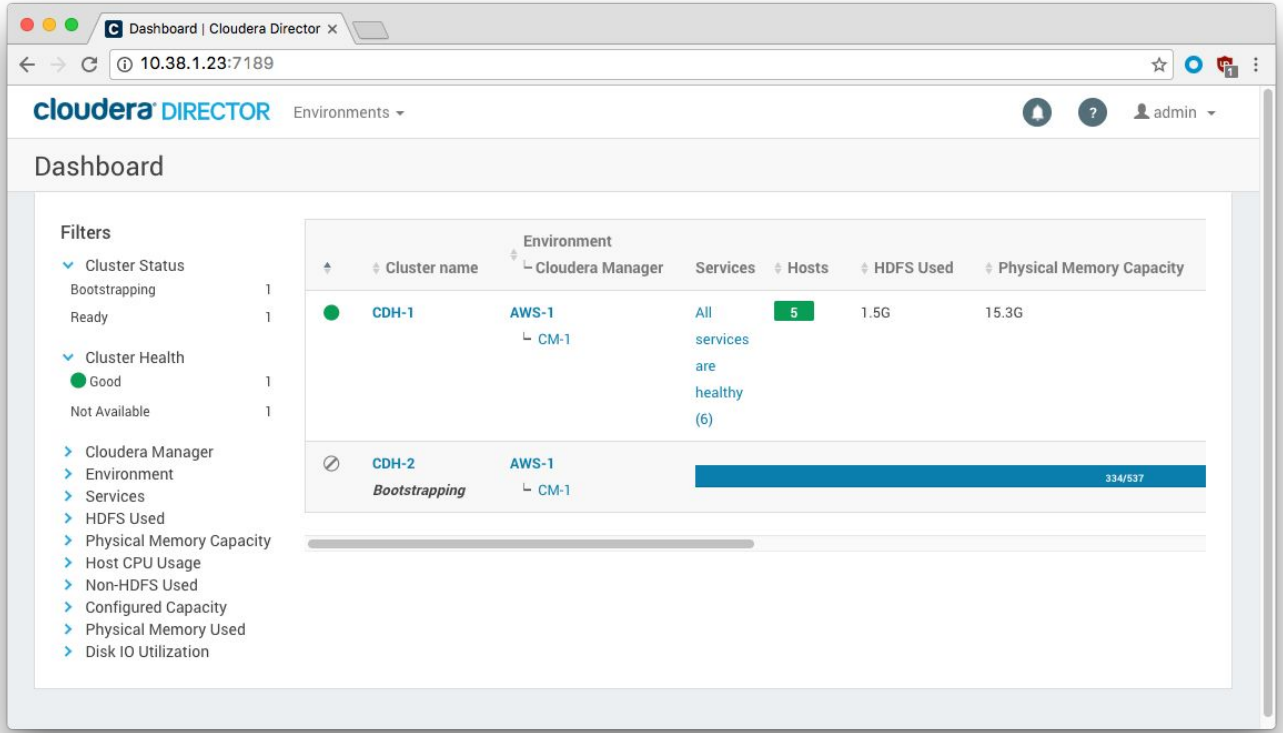
*"**Apache Hadoop** is a collection of open-source software utilities that facilitate using a **network of many computers** to solve problems involving **massive amounts of data** and computation. It provides a software framework for **distributed storage and processing** of big data using the MapReduce programming model."*

~ https://en.wikipedia.org/wiki/Apache_Hadoop

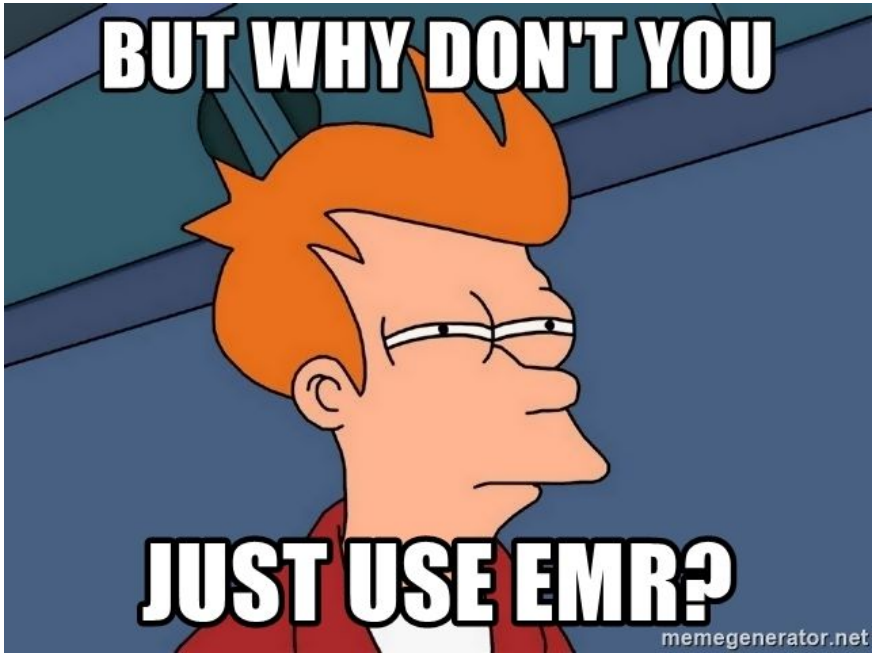
Cloudera Manager



Cloudera Director



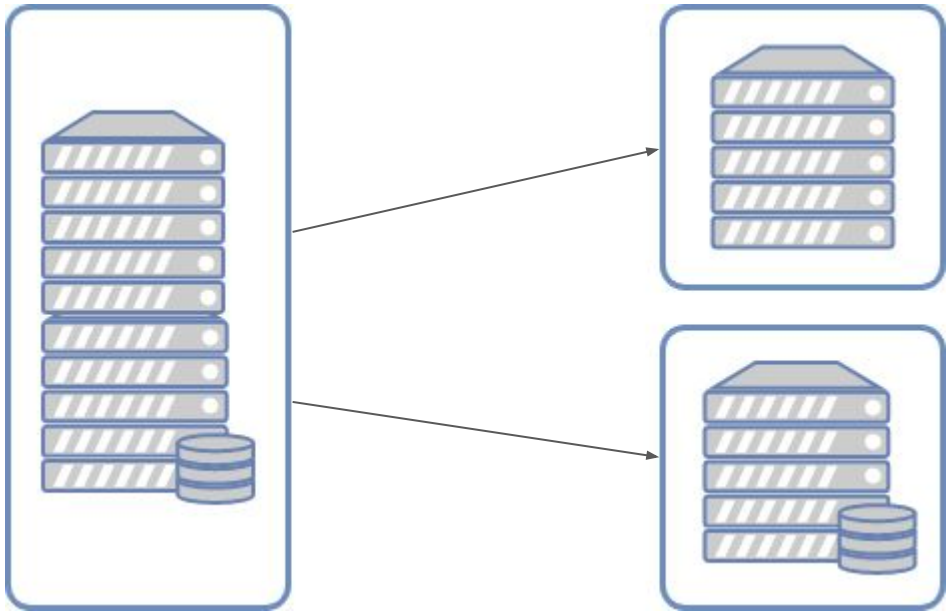
Wait a second..



Separate Compute and Storage

What we've gained?

- Storage is predictable, we can reserve instances (in our case, up to **40%** discount).
- Compute power varies, therefore it scales.



Choosing Instance Type (Datanodes)

EC2Instances.info Easy Amazon EC2 Instance Comparison

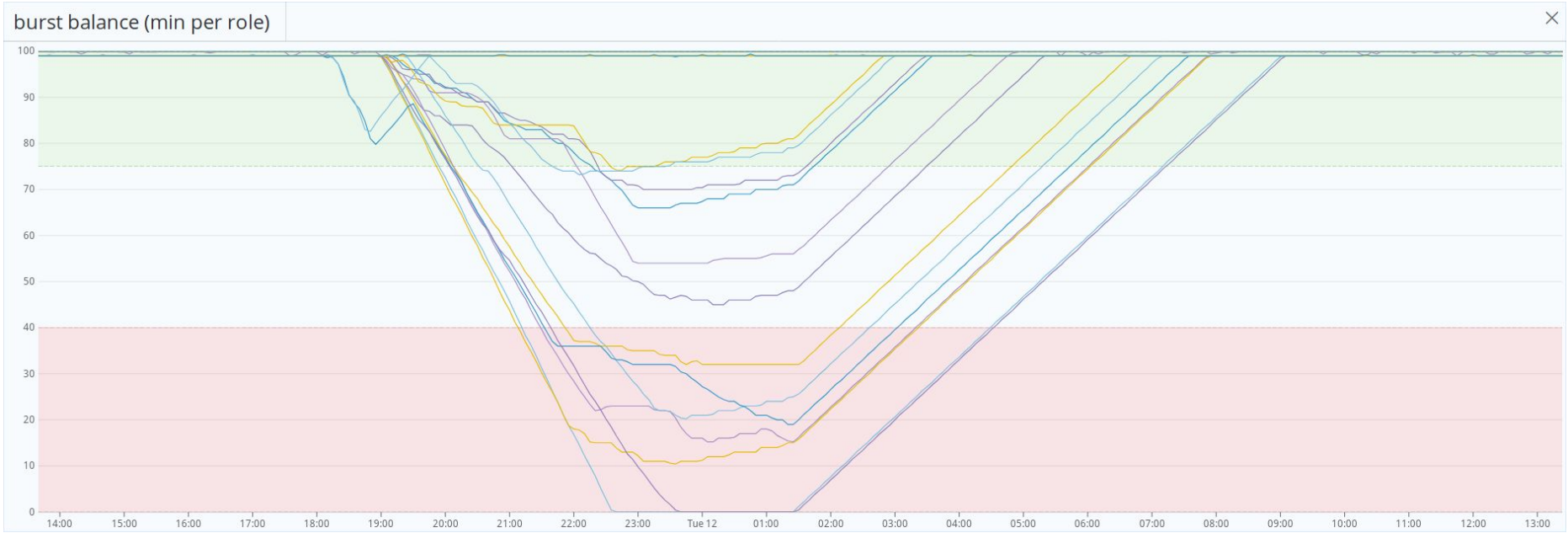
EC2 | RDS

Region: US East (N. Virginia) | Cost: Hourly | Reserved: 1-year - No Upfront | Columns | Compare Selected | Clear Filters | CSV

Filter: Min Memory (GiB): 60 Min vCPUs: 16 Min Storage (GiB): 0

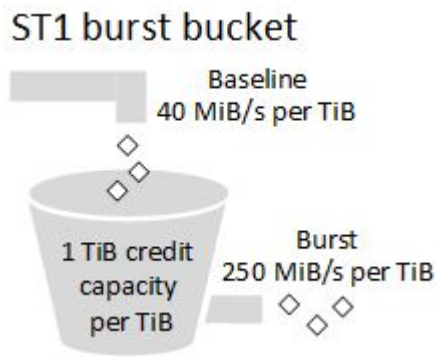
Name	API Name	Memory	vCPUs	Instance Storage	Network Performance	Linux On Demand cost	Linux Reserved cost
M5 General Purpose Quadruple Extra Large	m5.4xlarge	64.0 GiB	16 vCPUs	EBS only	High	$0.768000 hourly	$0.491000 hourly
M4 General Purpose Quadruple Extra Large	m4.4xlarge	64.0 GiB	16 vCPUs	EBS only	High	$0.800000 hourly	$0.496000 hourly
M5 General Purpose Quadruple Extra Large	m5d.4xlarge	64.0 GiB	16 vCPUs	600 GiB (2 * 300 GiB NVMe SSD)	High	$0.904000 hourly	$0.576000 hourly
H1 Quadruple Extra Large	h1.4xlarge	64.0 GiB	16 vCPUs	4000 GiB (2 * 2000 GiB HDD)	Up to 10 Gigabit	$0.936000 hourly	$0.636000 hourly
R4 High-Memory Quadruple Extra Large	r4.4xlarge	122.0 GiB	16 vCPUs	EBS only	Up to 10 Gigabit	$1.064000 hourly	$0.672000 hourly
G3 Quadruple Extra Large	g3.4xlarge	122.0 GiB	16 vCPUs	EBS only	Up to 10 Gigabit	$1.140000 hourly	$0.778000 hourly
I3 High I/O Quadruple Extra Large	i3.4xlarge	122.0 GiB	16 vCPUs	3800 GiB (2 * 1900 GiB NVMe SSD)	Up to 10 Gigabit	$1.248000 hourly	$0.857000 hourly
R3 High-Memory Quadruple Extra Large	r3.4xlarge	122.0 GiB	16 vCPUs	320 GiB SSD	High	$1.330000 hourly	$0.836000 hourly
C5 High-CPU 9xlarge	c5.9xlarge	72.0 GiB	36 vCPUs	EBS only	10 Gigabit	$1.530000 hourly	$0.964000 hourly
C4 High-CPU Eight Extra Large	c4.8xlarge	60.0 GiB	36 vCPUs	EBS only	10 Gigabit	$1.591000 hourly	$1.008000 hourly
C3 High-CPU Eight Extra Large	c3.8xlarge	60.0 GiB	32 vCPUs	640 GiB (2 * 320 GiB SSD)	10 Gigabit	$1.680000 hourly	$1.168000 hourly
C5 High-CPU 9xlarge	c5d.9xlarge	72.0 GiB	36 vCPUs	900 GiB NVMe SSD	10 Gigabit	$1.728000 hourly	$1.090000 hourly

Burst Balance Issues

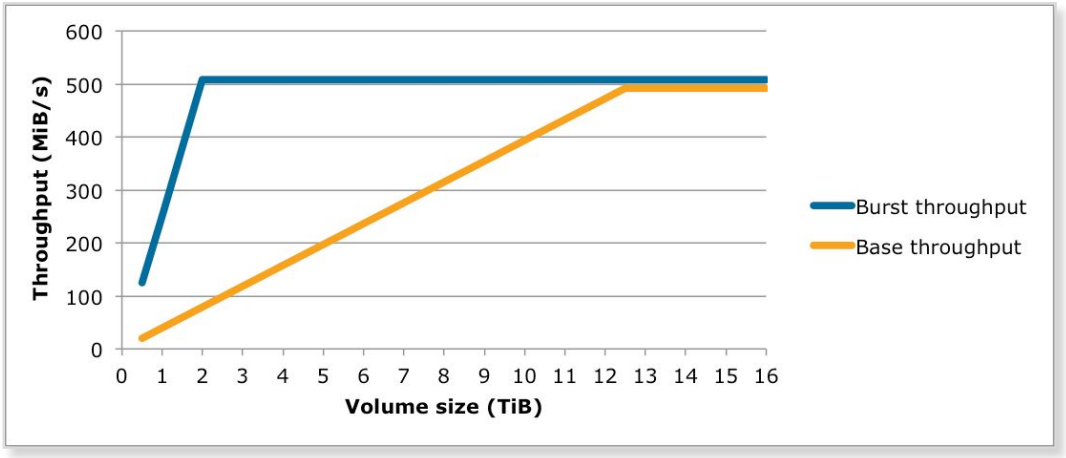


Choosing the Right EBS

Calculating Baseline / Burst Throughput



* Bucket and Credit Model



St1 Volumes, https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html

Choosing Instance Type (Workers)

EC2Instances.info Easy Amazon EC2 Instance Comparison

EC2 | RDS

Region: US East (N. Virginia) | Cost: Hourly | Reserved: 1-year - No Upfront | Columns | Compare Selected | Clear Filters | CSV

Filter: Min Memory (GiB): 32 Min vCPUs: 32 Min Storage (GiB): 500

Name	API Name	Memory	vCPUs	Instance Storage	Network Performance	Linux On Demand cost	Linux Reserved cost
C3 High-CPU Eight Extra Large	c3.8xlarge	60.0 GiB	32 vCPUs	640 GiB (2 * 320 GiB SSD)	10 Gigabit	$1.680000 hourly	$1.168000 hourly
C5 High-CPU 9xlarge	c5d.9xlarge	72.0 GiB	36 vCPUs	900 GiB NVMe SSD	10 Gigabit	$1.728000 hourly	$1.090000 hourly
H1 Eight Extra Large	h1.8xlarge	128.0 GiB	32 vCPUs	8000 GiB (4 * 2000 GiB HDD)	10 Gigabit	$1.872000 hourly	$1.272000 hourly
Cluster Compute Eight Extra Large	cc2.8xlarge	60.5 GiB	32 vCPUs	3360 GiB (4 * 840 GiB HDD)	10 Gigabit	$2.000000 hourly	$1.090000 hourly
I3 High I/O Eight Extra Large	i3.8xlarge	244.0 GiB	32 vCPUs	7600 GiB (4 * 1900 GiB NVMe SSD)	10 Gigabit	$2.496000 hourly	$1.714000 hourly
R3 High-Memory Eight Extra Large	r3.8xlarge	244.0 GiB	32 vCPUs	640 GiB (2 * 320 GiB SSD)	10 Gigabit	$2.660000 hourly	$1.672000 hourly
M5 General Purpose 12xlarge	m5d.12xlarge	192.0 GiB	48 vCPUs	1800 GiB (2 * 900 GiB NVMe SSD)	10 Gigabit	$2.712000 hourly	$1.727000 hourly
C5 High-CPU 18xlarge	c5d.18xlarge	144.0 GiB	72 vCPUs	1800 GiB (2 * 900 GiB NVMe SSD)	25 Gigabit	$3.456000 hourly	$2.180000 hourly
H1 16xlarge	h1.16xlarge	256.0 GiB	64 vCPUs	16000 GiB (8 * 2000 GiB HDD)	25 Gigabit	$3.744000 hourly	$2.545000 hourly
I3 High I/O 16xlarge	i3.16xlarge	488.0 GiB	64 vCPUs	15200 GiB (8 * 1900 GiB NVMe SSD)	25 Gigabit	$4.992000 hourly	$3.427000 hourly
I3 High I/O Metal	i3.metal	512.0 GiB	N/A vCPUs	15200 GiB (8 * 1900 GiB NVMe SSD)	25 Gigabit	$4.992000 hourly	$3.427000 hourly
M5 General Purpose 24xlarge	m5d.24xlarge	384.0 GiB	96 vCPUs	3600 GiB (4 * 900 GiB NVMe SSD)	25 Gigabit	$5.424000 hourly	$3.454000 hourly

Now we can scale!

Self-Hosted Cluster Utilization

There is more that can be done!

Utilizing Spot Instances for Workers

on-demand hourly cost: **1.68$**

Spot hourly price cost (varies): **0.4367$**

That's **74%** discount!

Which saves us **1.09M $**/year! (Over ~100 nodes)

* pricing for c3.8xlarge

All of these things adds up..

Takeaways

1. Cost == Monthly bill + Operational overhead + Maintenance + Reaction time
2. Dare to think about solutions outside of the AWS ecosystem.
3. Full utilization of paid resources with the ability to respond fast is vital for a growing business.
4. Using large EBS volumes will give you more headroom in terms of burst balance.
5. TAG EVERYTHING!

Q & A

THANK YOU