Cost Optimization at Scale:

Building and Realizing the Economic Case for the AWS Cloud

Shahbaz Alam

AWS Professional Services

September 2016

A Couple Assumptions…

1. You’re using AWS…

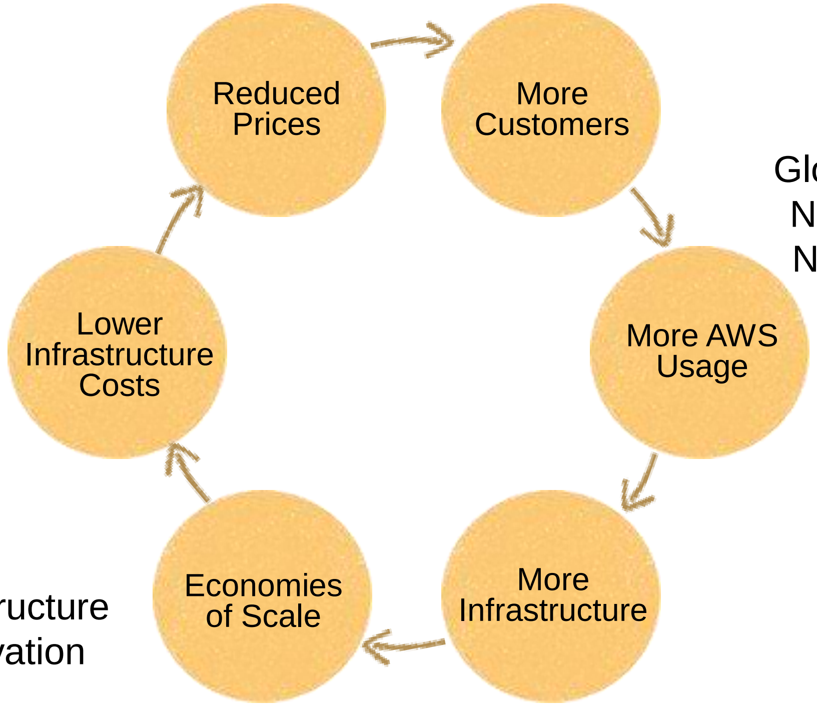
2. You like it!!

But maybe you are spending more than you planned…

Or you’d just like to spend less

What should you do?

AWS Pricing Philosophy



Reduced Prices

More Customers

More AWS Usage

More Infrastructure

Economies of Scale

Lower Infrastructure Costs

Ecosystem
Global Footprint
New Features
New Services

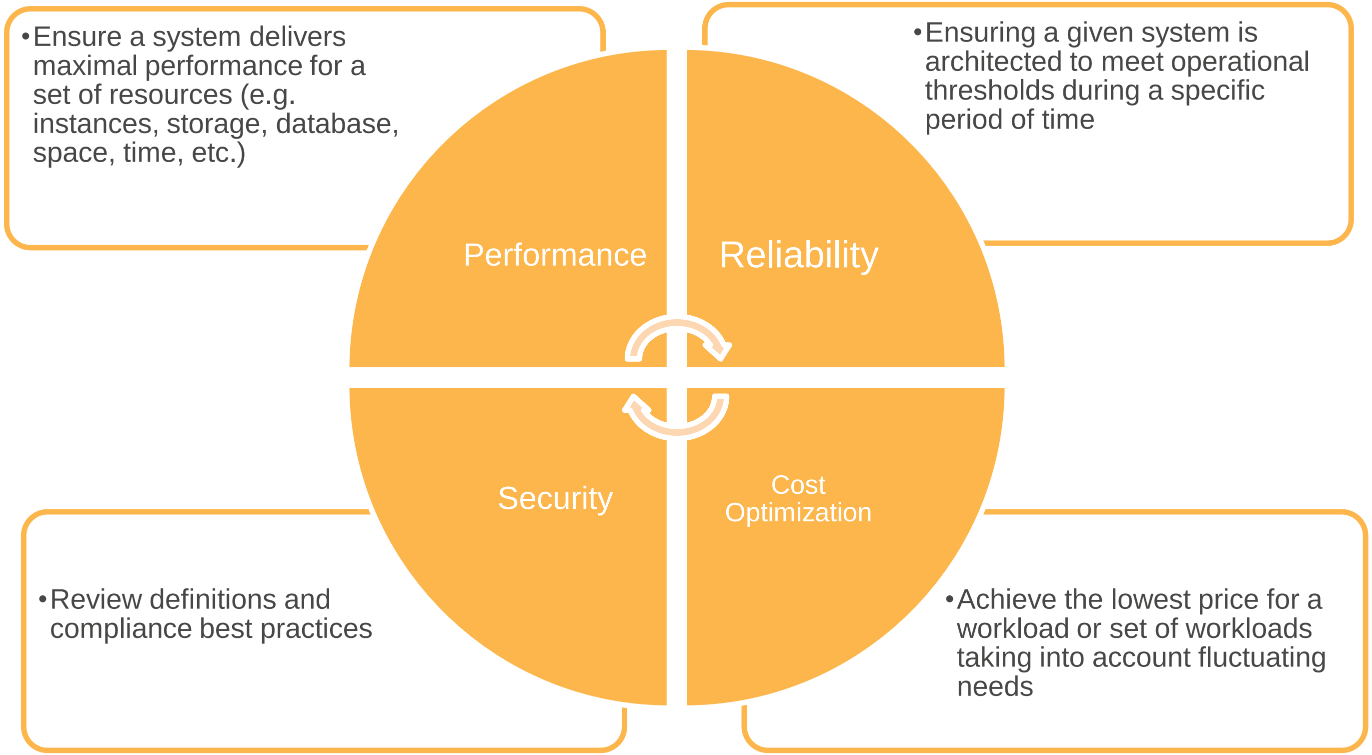
Infrastructure
Innovation

50+

PRICE REDUCTIONS

We pass the savings along to our customers in the form of low prices and continuous reductions

4 Components of AWS Architecture Best Practices



- Ensure a system delivers maximal performance for a set of resources (e.g. instances, storage, database, space, time, etc.)

Performance

- Ensuring a given system is architected to meet operational thresholds during a specific period of time

Reliability

Security

- Review definitions and compliance best practices

Cost Optimization

- Achieve the lowest price for a workload or set of workloads taking into account fluctuating needs

In the beginning . . .

…there was **TCO**

What is TCO?

Definition: ***Comparative*** **total cost of ownership analysis** (acquisition and operating costs) for running an infrastructure environment end-to-end on-premises vs. on AWS.

Used for:

1) Comparing the costs of running an **entire infrastructure environment or specific workload** on-premises or in a co-location facility vs. on AWS

2) Budgeting and **building the business case** for moving to AWS

So how do we do it?

≠

amazon
web services

TCO = Acquisition costs + Operations costs

illustrative

1	**Server costs**	Hardware—server, rack chassis PDUs, Tor switches (+maintenance)	Software—OS, virtualization licenses (+maintenance)	**Facilities cost** Space	Power	Cooling
2	**Storage costs**	Hardware—storage disks, SAN/FC switches	Storage admin costs	**Facilities cost** Space	Power	Cooling
3	**Network costs**	Network hardware—LAN switches, load balancer bandwidth costs	Network admin costs	**Facilities cost** Space	Power	Cooling
4	**IT labor costs**	Server admin virtualization admin				

The diagram doesn’t include every cost item. For example, software costs can include database, management, and middle-tier software costs. Facilities cost can include costs associated with upgrades, maintenance, building security, taxes, and so on. IT labor costs can include security admin and application admin costs.

Questions to explore your existing footprint…

1 **Capacity Planning**

- How do you plan for capacity?
- How many servers have you added in the past year? Anticipating next year?
- Can you switch your hardware on and off and only pay for what is used?

2 **Utilization**

- What is your average server utilization?
- How much do you overprovision for peak load?

3 **Operations**

- Will you run out of data center space some time in the future?
- What was your last year power utility bill for the Data Center(s)?
- Have you budgeted for both average and peak power requirements?

4 **Optimization**

- Are you on AWS today?
- Is your architecture cost-optimized (Auto Scaling, RIs, Spot, Instances turn on/off)?

And, make sure to…

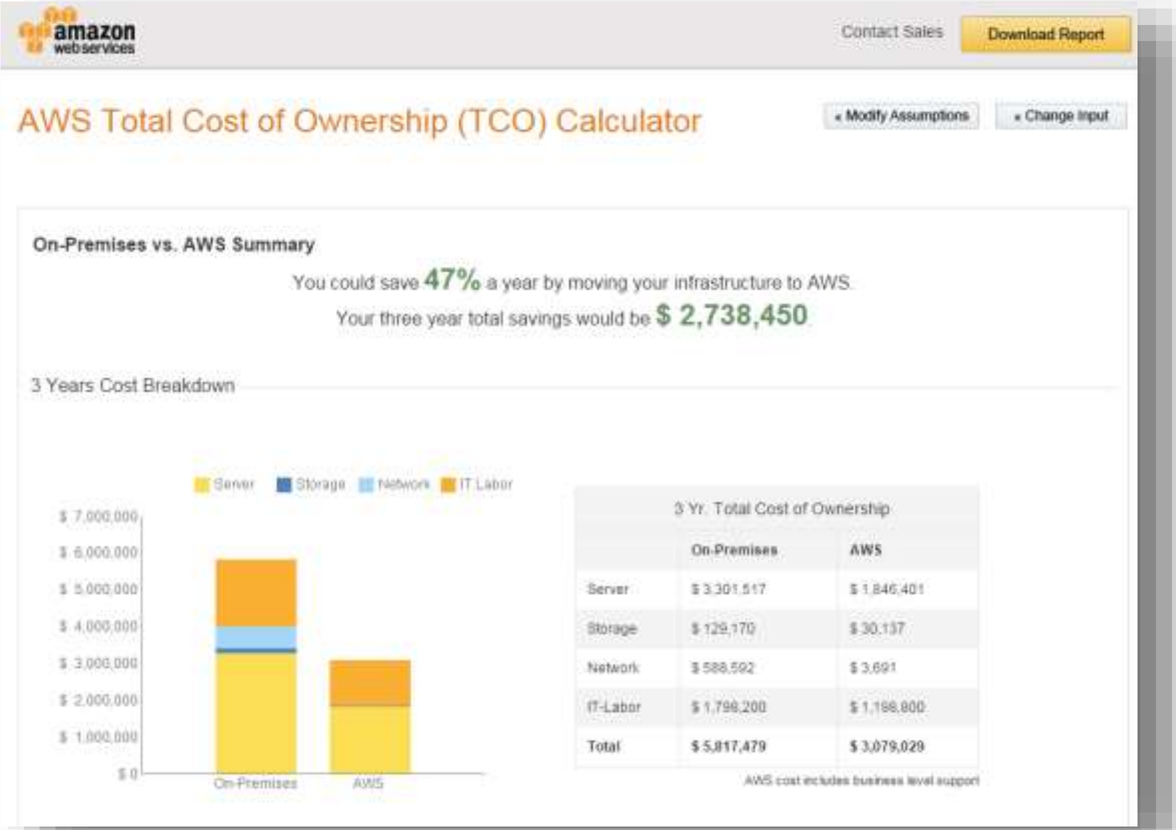
Consider

- Power/Cooling (compute, storage, shared network)
- Data Center Administration (procurement, design, build, operate, network, security personnel)
- Rent/Real Estate (building deprecation, taxes)
- Software (OS, virtualization licensing & maintenance)
- RAW vs. USABLE storage capacity
- Storage Redundancy (RAID penalty, OS penalty)
- Storage Backup costs (tape, backup software)
- Bandwidth, Network Gear & Redundancy (routers, VPN, WAN, etc.)

Understand

- Procurement Time, Resource sitting on self
- Cost of Lost Customers
- RTO, RPO

Resources to get you started



AWS Total Cost of Ownership (TCO) Calculator

On-Premises vs. AWS Summary

You could save 47% a year by moving your infrastructure to AWS.

Your three year total savings would be $ 2,738,450

3 Years Cost Breakdown

3 Yr. Total Cost of Ownership

	On-Premises	AWS
Server	$ 3,301,517	$ 1,846,401
Storage	$ 129,170	$ 30,137
Network	$ 588,592	$ 3,691
IT-Labor	$ 1,798,200	$ 1,198,800
Total	**$ 5,817,479**	**$ 3,079,029**

AWS cost includes business level support

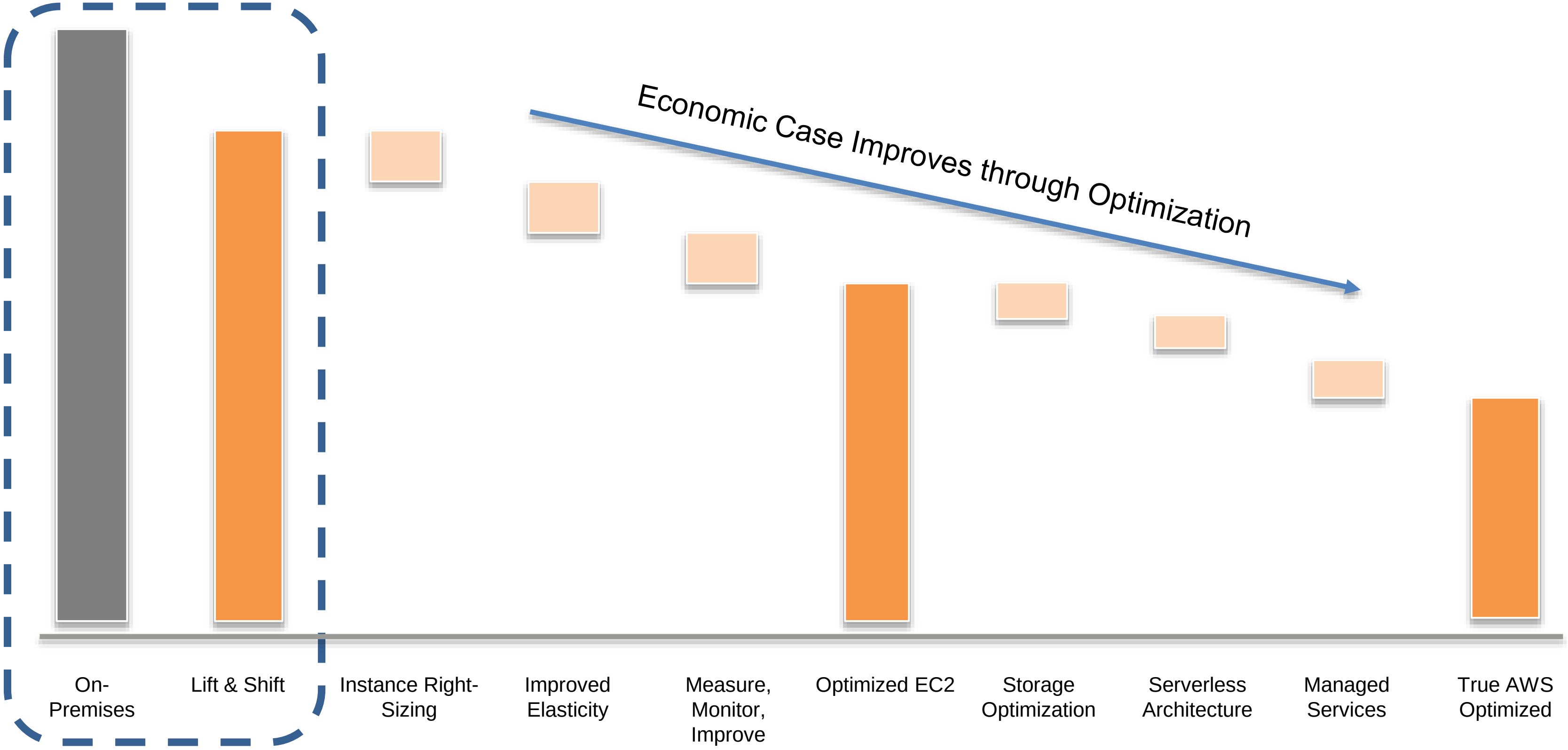
AWS TCO Calculator

https://awstcocalculator.com

Case studies and research

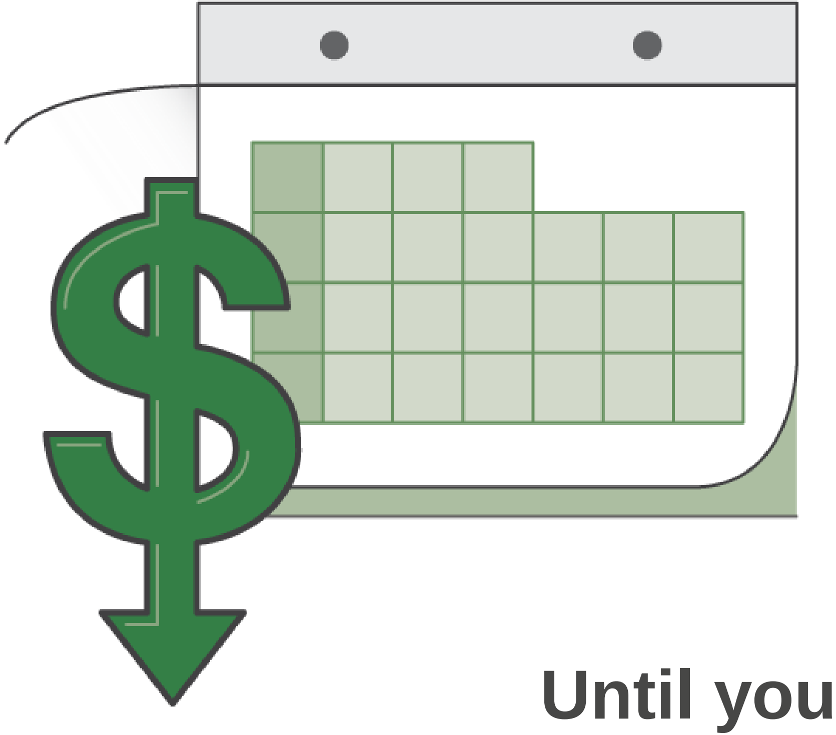
http://aws.amazon.com/economics/

Lowering TCO through cost optimization



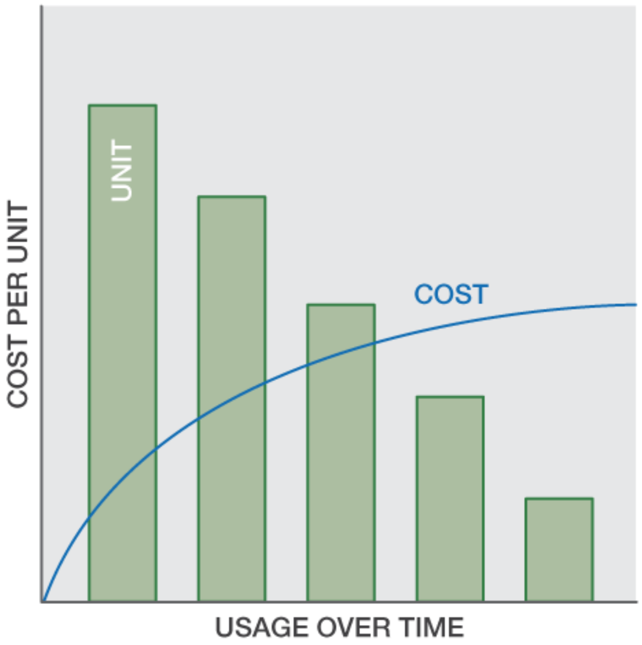
Traditional TCO Comparisons

So you’re feeling pretty good.



Until your CFO shows up with the bill.

Cost optimization is…



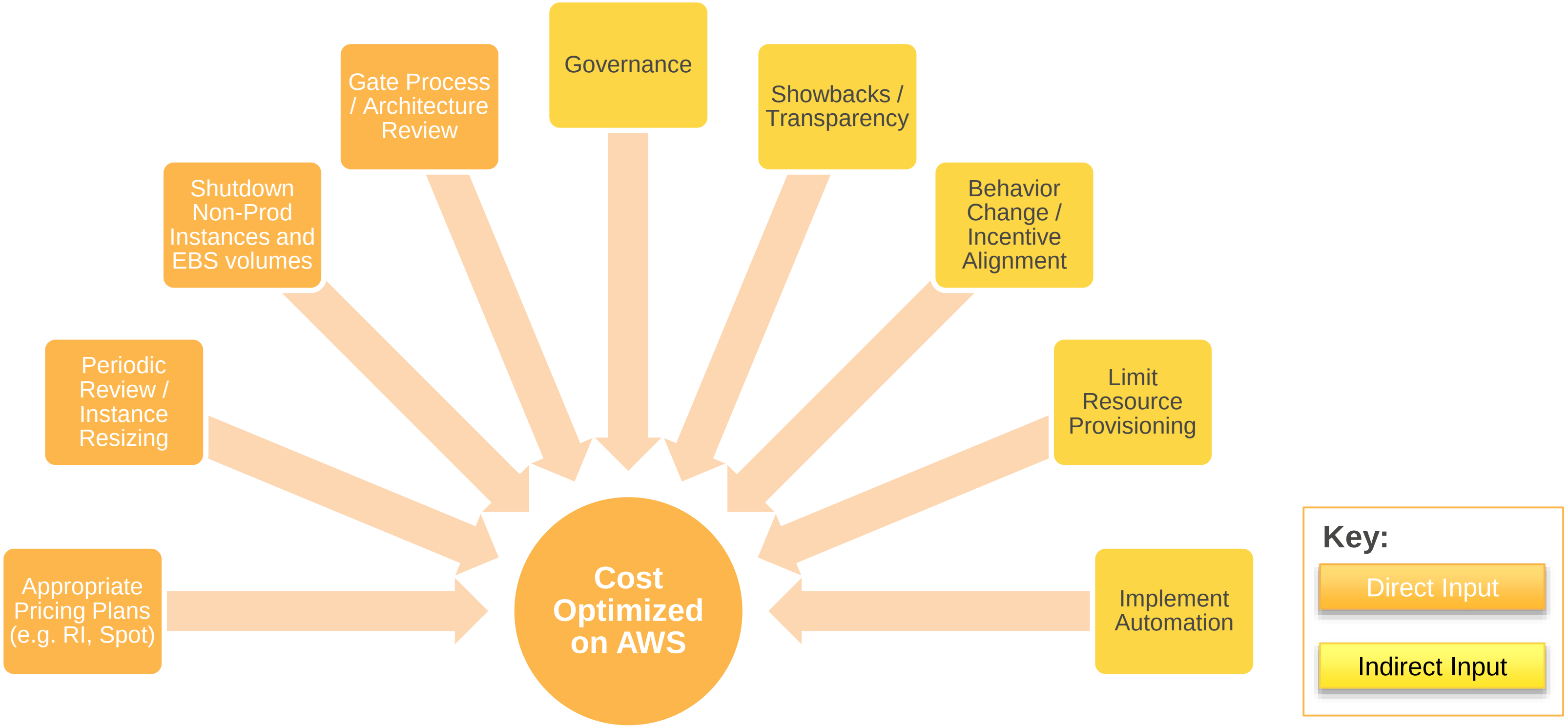
going from…

pay for what you *use*

to…

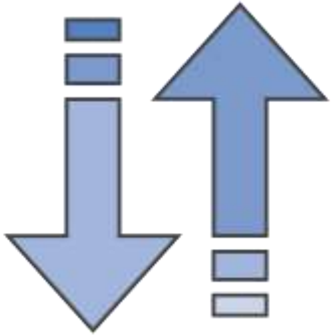
pay for what you *need*

Key inputs to cost optimization on AWS

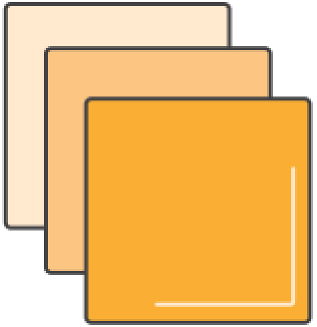

Gate Process / Architecture Review
Governance
Showbacks / Transparency
Shutdown Non-Prod Instances and EBS volumes
Behavior Change / Incentive Alignment
Periodic Review / Instance Resizing
Limit Resource Provisioning
Appropriate Pricing Plans (e.g. RI, Spot)
Cost Optimized on AWS
Implement Automation
Key:
Direct Input
Indirect Input

Where do you start?

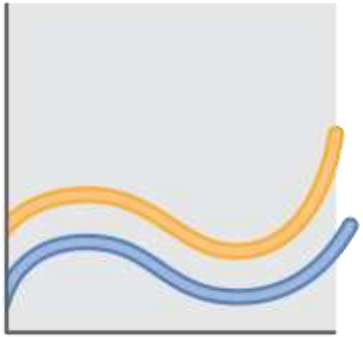
The four pillars of cost optimization



Right-sizing



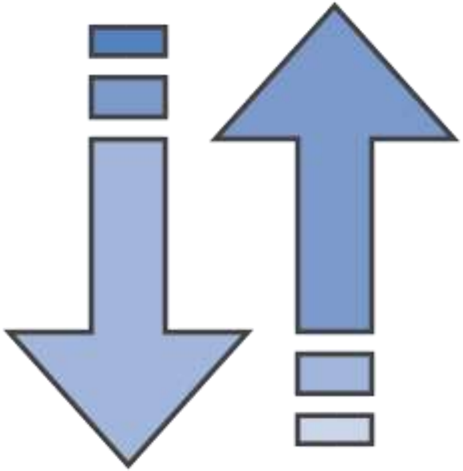
Reserved Instances



Increase elasticity

Measure, monitor, and improve

Right-sizing



Right-sizing

- Selecting the cheapest instance available while meeting performance requirements
- Looking at CPU, RAM, storage, and network utilization to identify potential instances that can be downsized
- Leveraging Amazon CloudWatch metrics and setting up custom RAM metrics

Rule of thumb: Right size, then reserve.

(But if you’re in a pinch, reserve first.)

Reserved Instances

Step 1: RI Coverage

- Cover always-on resources.

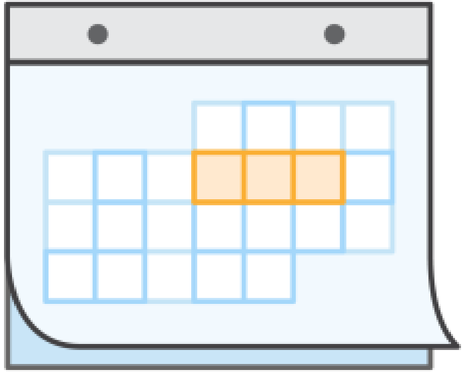
Step 2: RI Utilization

- Leverage RI flexibility to increase utilization.
- Merge and split RIs as needed.

Rule of thumb: Target 70–80% always-on coverage and 95% RI utilization rate.

EC2 Reserved Pricing

Steady State



Reserved Capacity

Upfront payments to
reduce costs

Reserved Instances

Up to 75%+ savings*

(and capacity reservation)

Commitment level

1 year

3 year

AWS services offering RIs

Amazon EC2

Amazon RDS

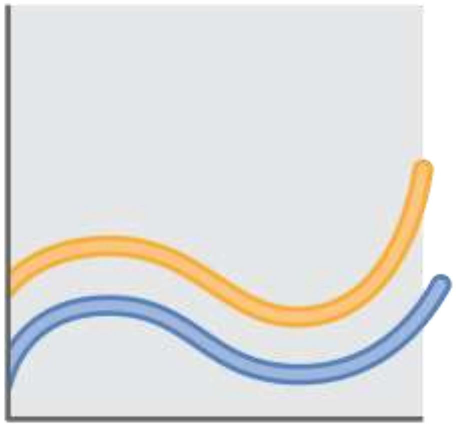
Amazon DynamoDB

Amazon Redshift

Amazon ElastiCache

* Dependent on specific AWS service, size/type, and region

Increase elasticity



Turn off nonproduction instances

- Look for dev/test, nonproduction instances that are running always-on and turn them off.

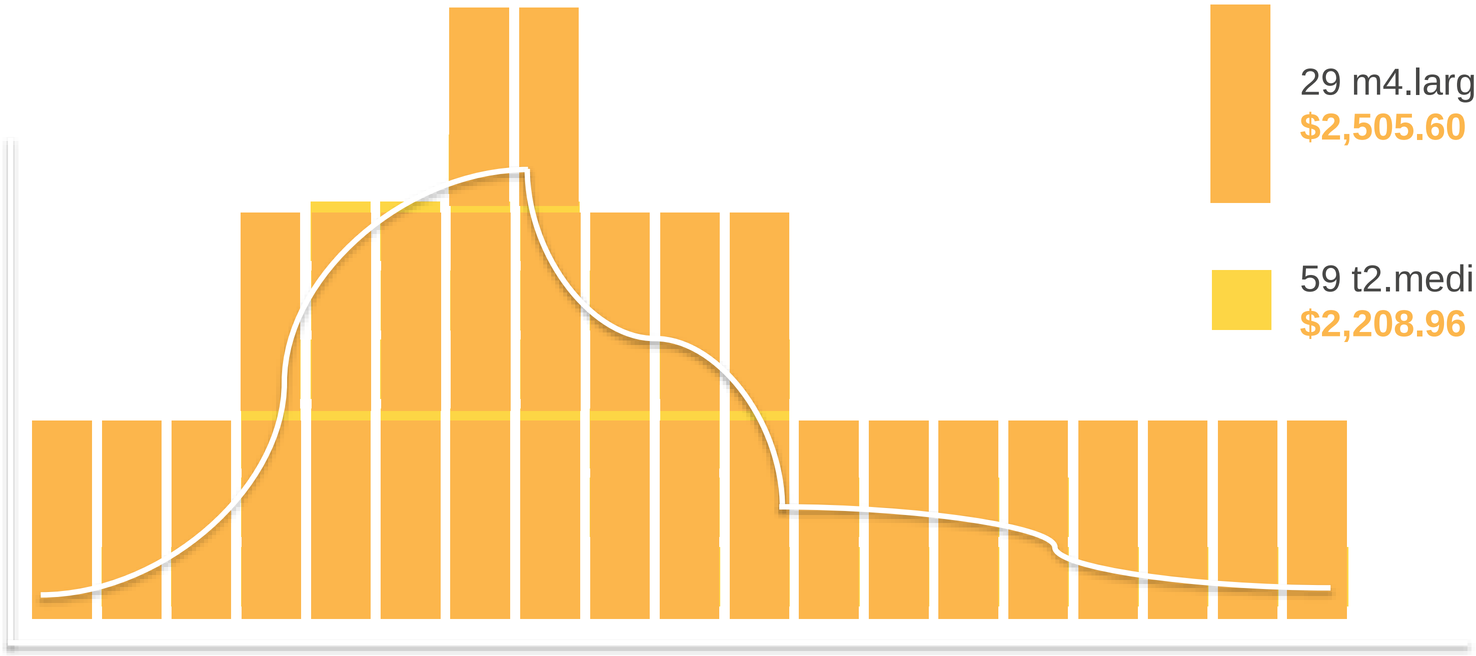
Autoscale production

- Use Auto Scaling to scale up and down based on demand and usage (for example, spikes).

Rule of thumb: Shoot for 20–30% of Amazon EC2 instances running on demand to be able to handle elasticity needs.

Using right-sizing and elasticity to lower cost

More smaller instances vs. fewer larger instances

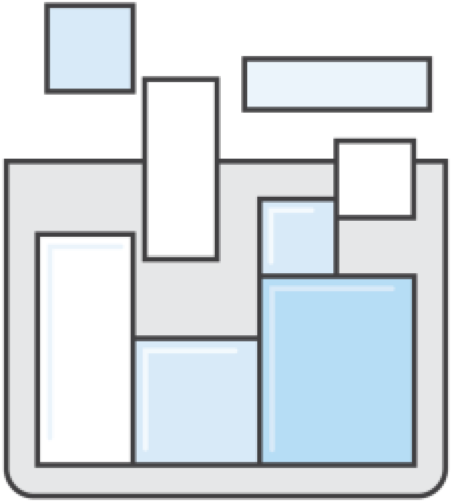


🟧 29 m4.large @ $0.12 /hr
$2,505.60 / mo*

🟨 59 t2.medium @ $0.052/hr
$2,208.96 / mo*

*Assumes Amazon Linux instances in the US-East (N. Virginia) Region at 720 hours per month

EC2 Spot Pricing



Time or instance flexible

Experiment and/or build cost sensitive businesses

Users with urgent computing needs or large amounts of additional capacity

Consider Spot for Elastic Workloads

90%

Savings!*

Options

- Spot Fleet to maintain instance availability
- Spot Block durations (1-6 hours) for workloads that must run continuously

Commitment level

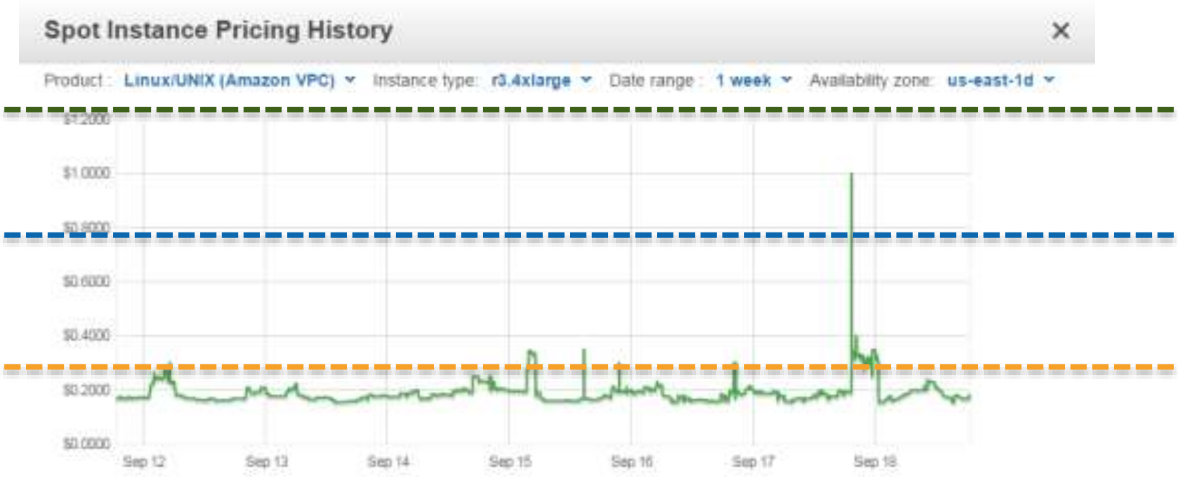
- None

* Compared to On Demand price based on specific EC2 instance type, region, and Availability Zone

Spot Rules

Markets where the price of compute changes based on supply and demand

You'll never pay more than your bid.



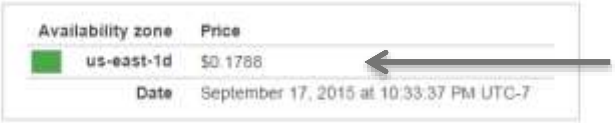
Spot Instance Pricing History

Product : Linux/UNIX (Amazon VPC)   Instance type: r3.4xlarge   Date range : 1 week   Availability zone: us-east-1d

75% of OD

50% of OD

25% of OD

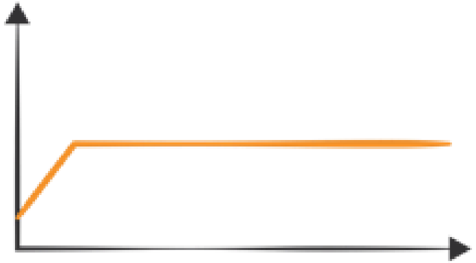


Availability zone	Price
🟩 us-east-1d	$0.1788
Date	September 17, 2015 at 10:33:37 PM UTC-7

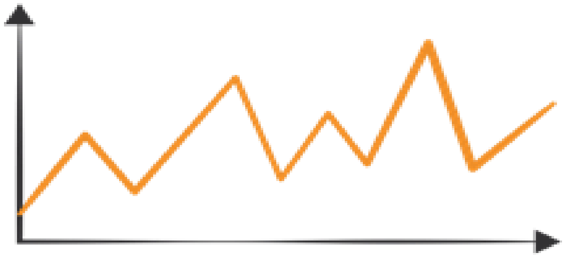
You pay the market price **87% discount!**

Strike a Balance

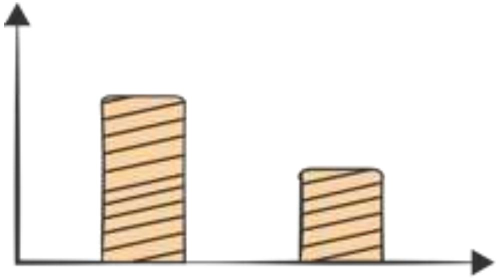
Reserved Instances

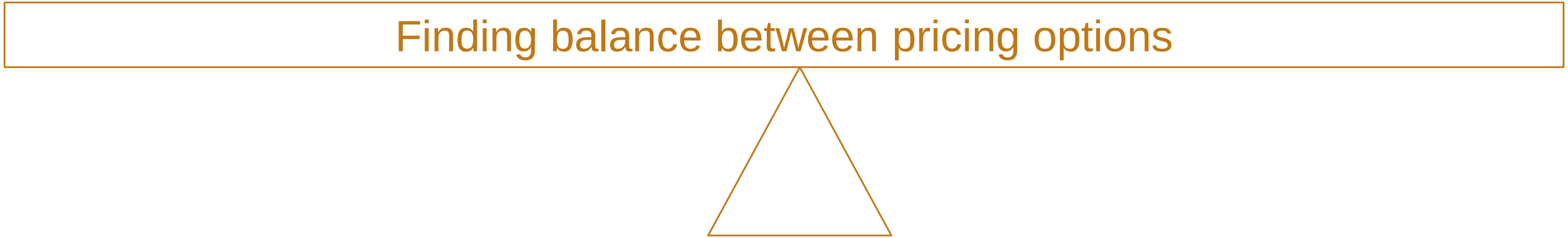


On Demand



Spot

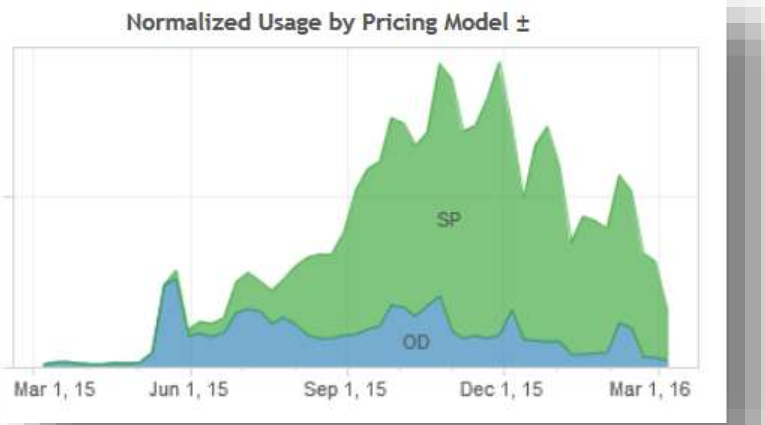




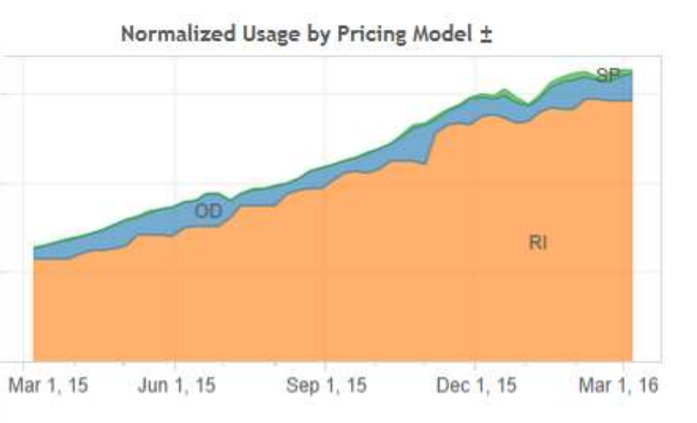
Finding balance between pricing options

Consumption model by industry

Web Scale (e.g. Adtech) Company

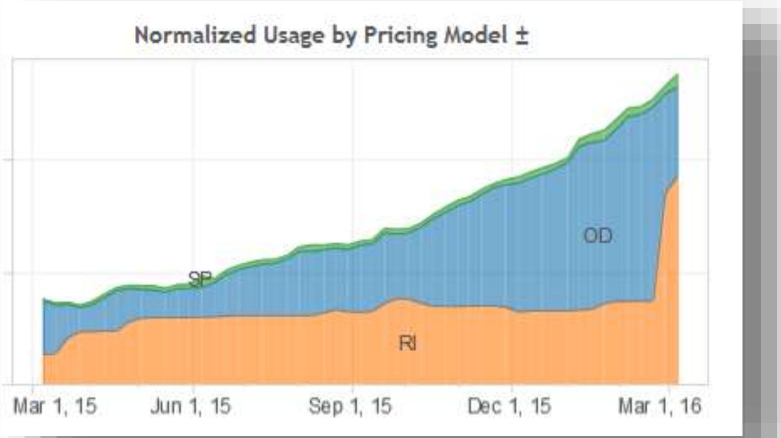


Enterprise SaaS Company

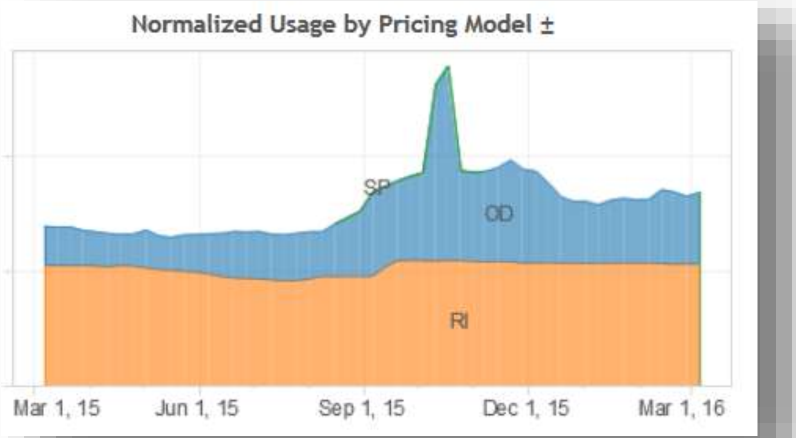


Consumption model by industry (cont…)

Onboarding Enterprise



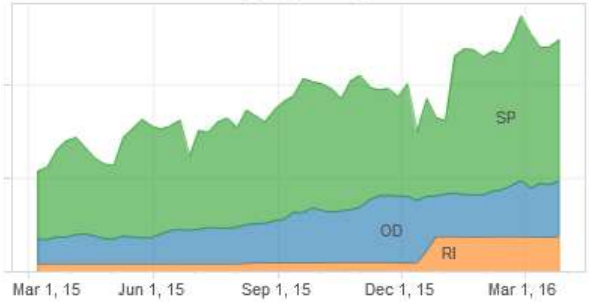
Gaming Company



Consumption model workload…

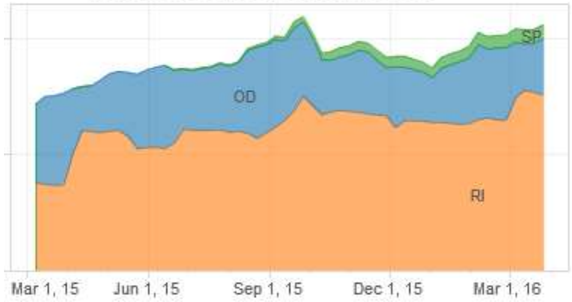
Dev Test

Normalized Usage by Pricing Model ±



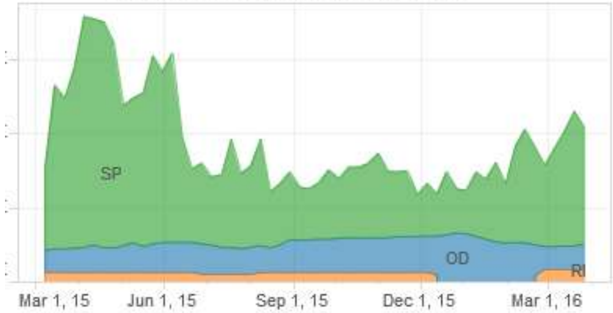
Enterprise Applications

Normalized Usage by Pricing Model ±



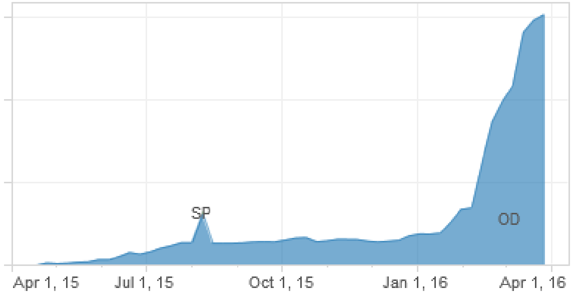
Data Science

Normalized Usage by Pricing Model ±



New app development

Normalized Usage by Pricing Model ±



EC2 cost optimization options

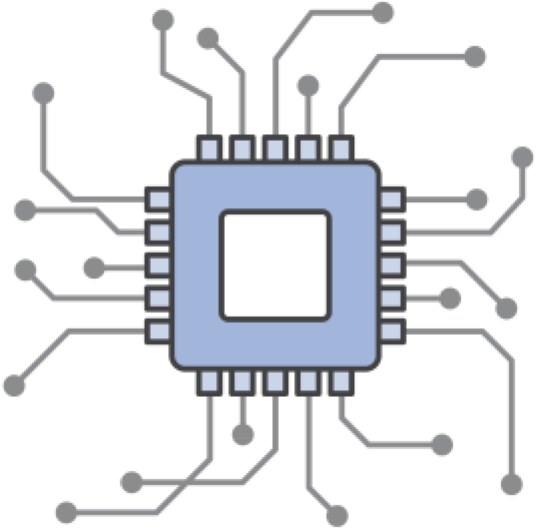
Cost Savings	EC2	Benefit
Base Price	On Demand	• No Commitment • Pay only what you use • No capacity reservation • No interruption
< 10%	Scheduled Reserved Instances	• Commitment of 1,200 hours for one year • Specified schedule • Capacity reservation; no interruption
30% – 75%	Standard Reserved Instances	• Commitment of one year or three years • Capacity reservation; no interruption
40% – 60%	Spot Blocks	• Bid for 2-6 hours blocks of time • No long term commitment • No interruption
Up to 85%	Spot	• Bid for instances • Interrupted if market price higher than your bid price • 2 minute advanced notice

Putting it all together: case study

Challenge:

Minimizing *unit costs* during a period of massive growth.

Elastic compute unit (ECU)



A consistent measure of CPU processing power

The growth challenge

2x

YoY Compute Growth

1,192 ECU

August 2015

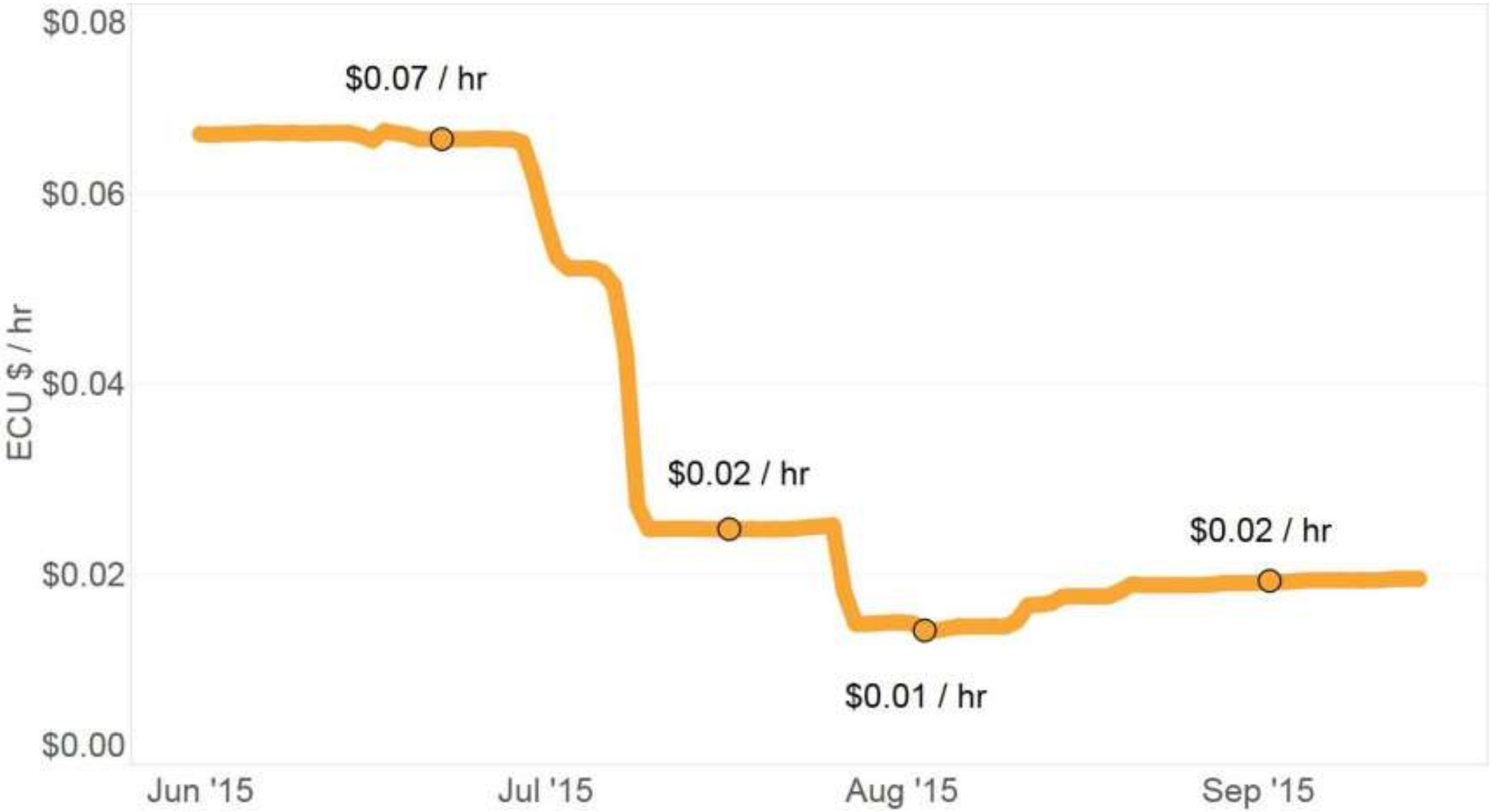
584 ECU

August 2014

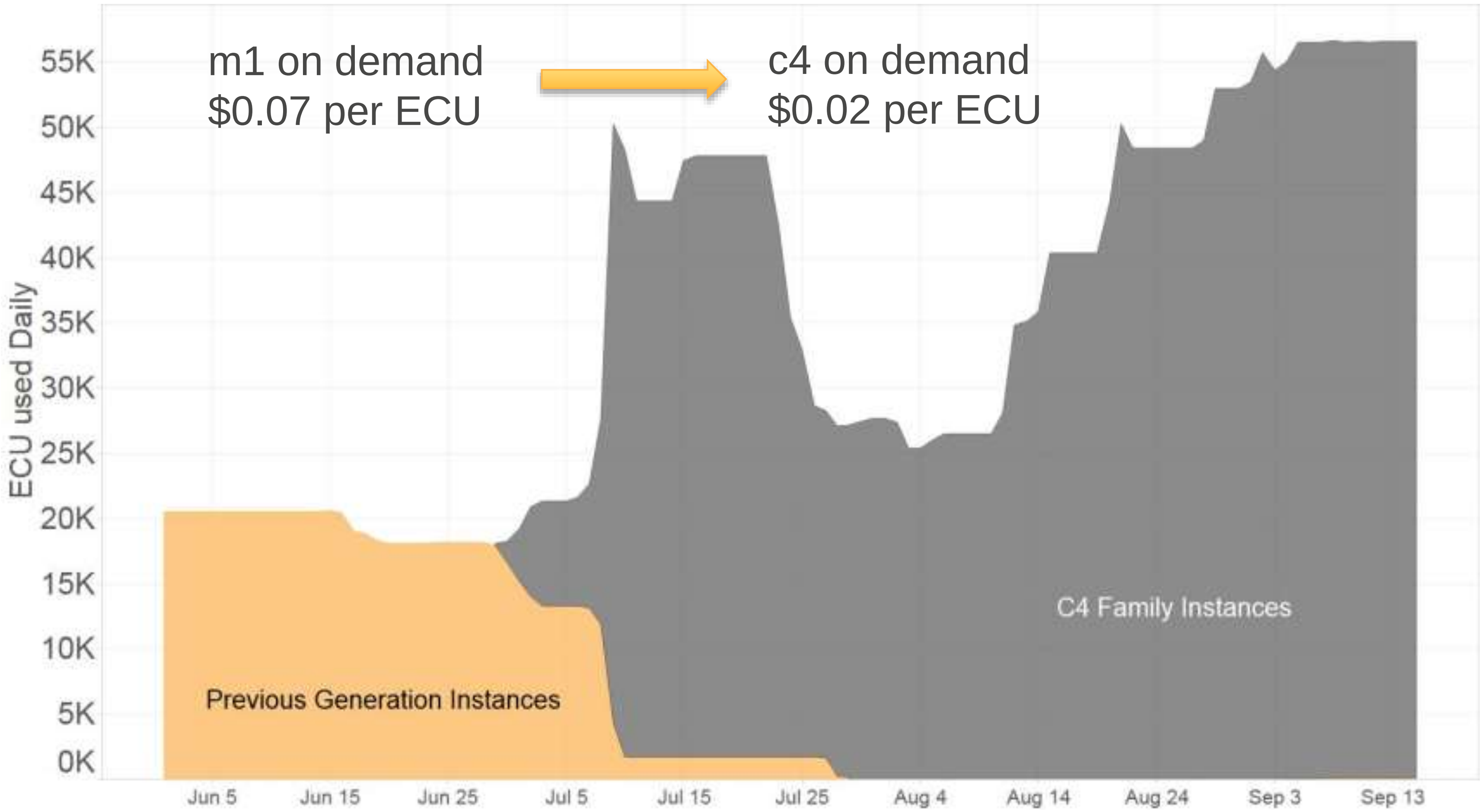
33%

decrease in monthly EC2 costs!

Solving the growth challenge



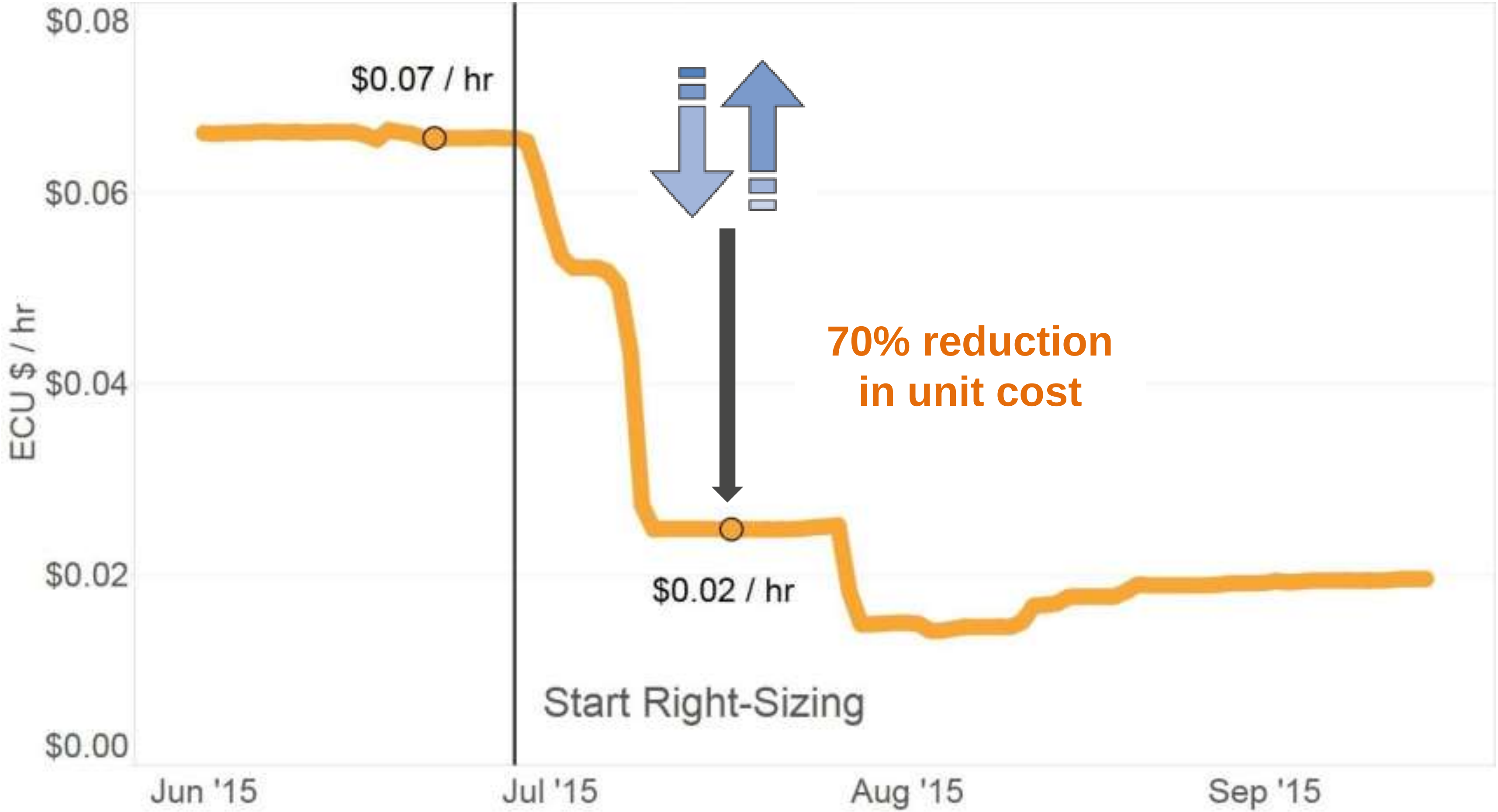
Step 1: Right-size and update instances



m1 on demand
$0.07 per ECU

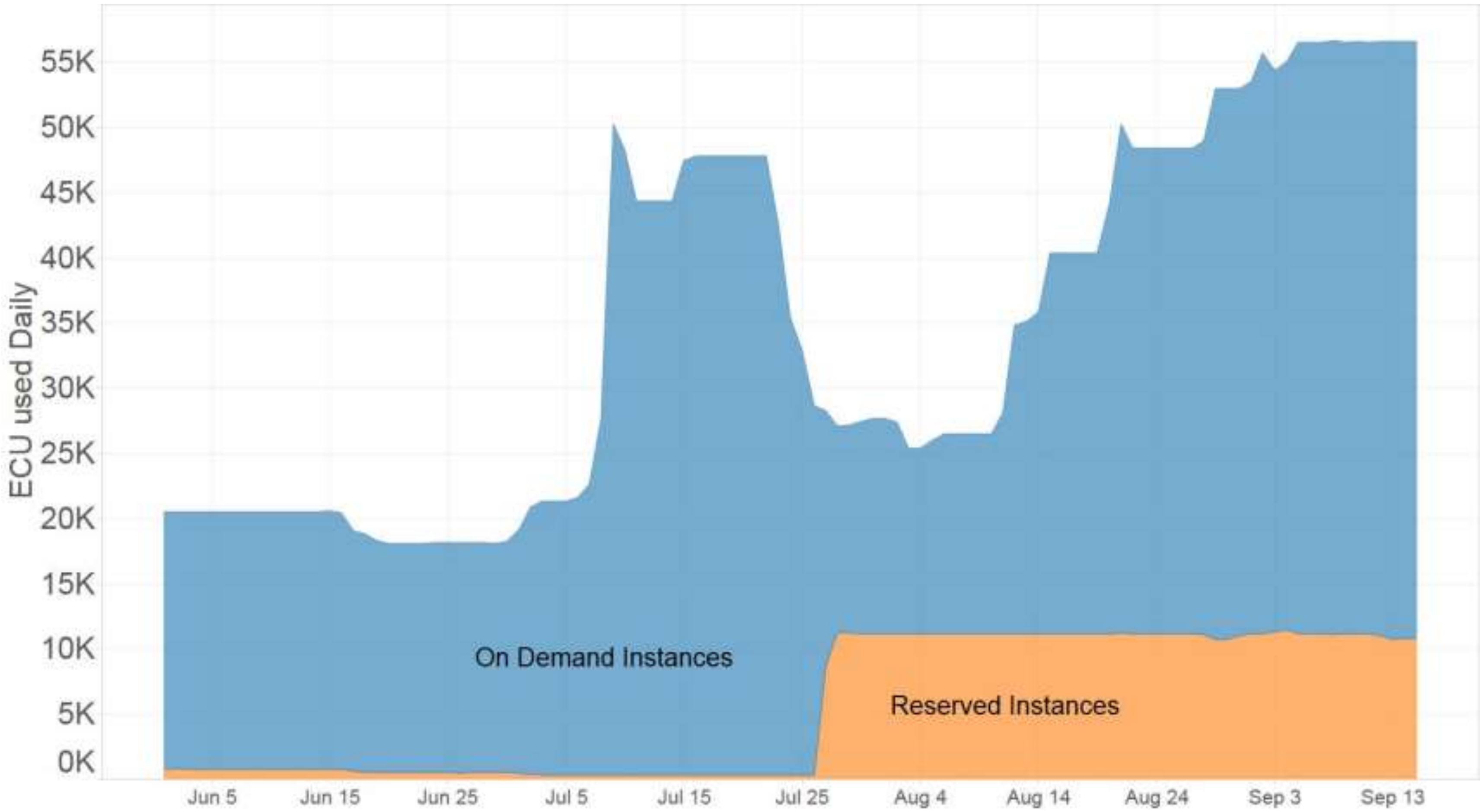
c4 on demand
$0.02 per ECU

The impact of right-sizing

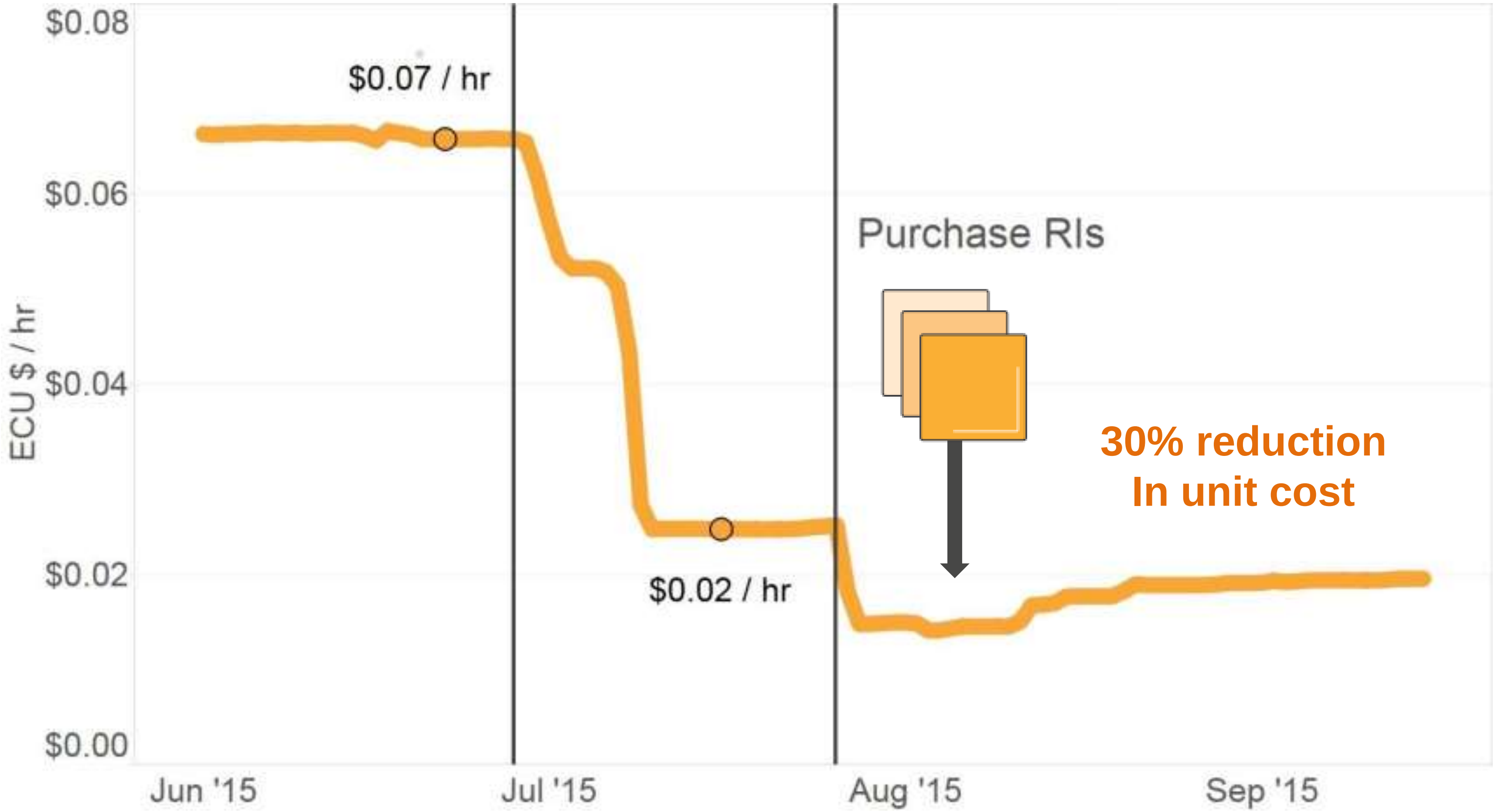


70% reduction in unit cost

Step 2: Reserve

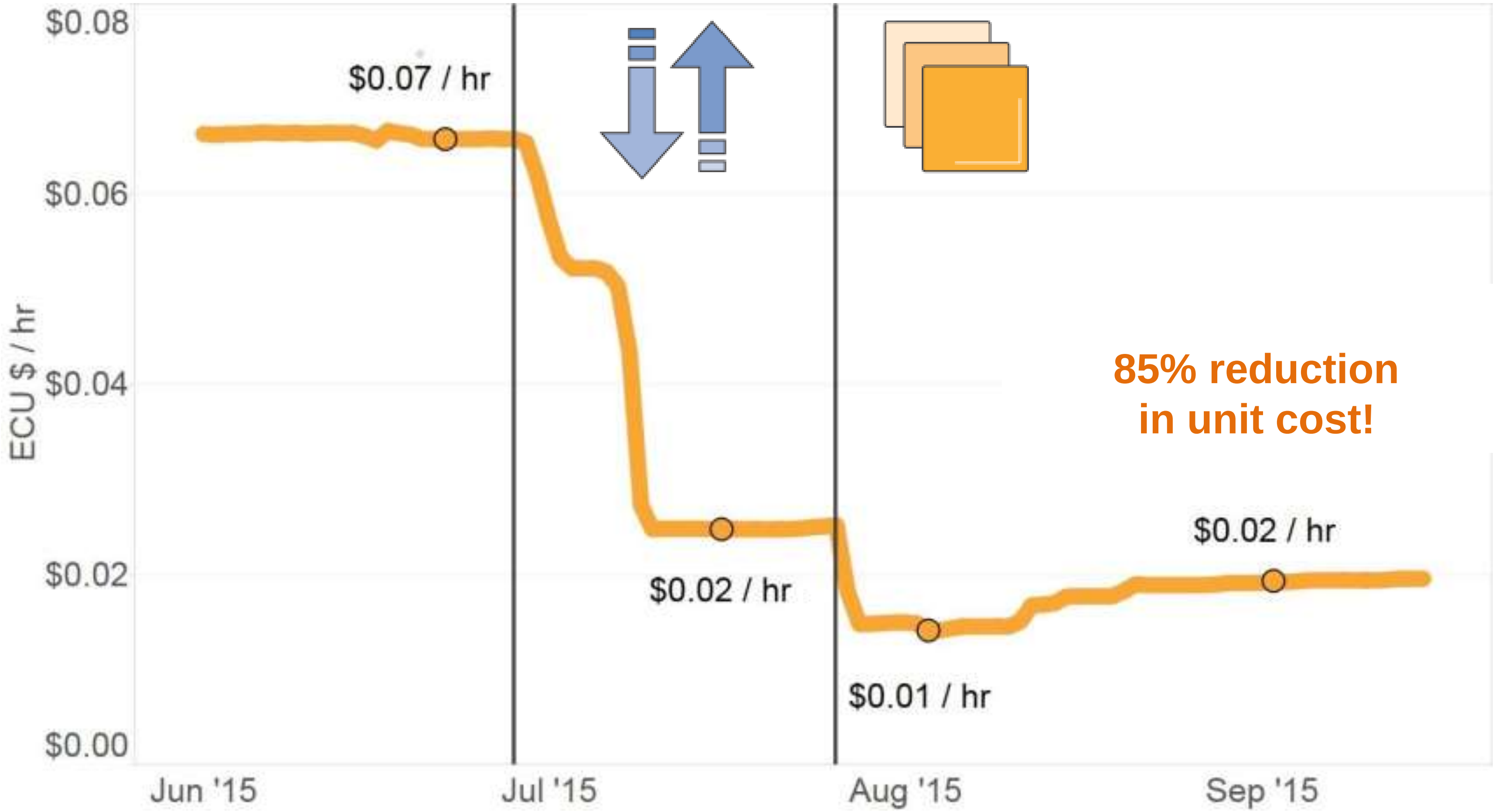


The impact of reservations



30% reduction
In unit cost

Putting it together



85% reduction in unit cost!

Sounds pretty easy, right?

Not really.

In reality, it is very complex.

- Scale
- Behavioral change
- Visibility
- Ownership

Cost optimization governance (Remember the fourth pillar?)

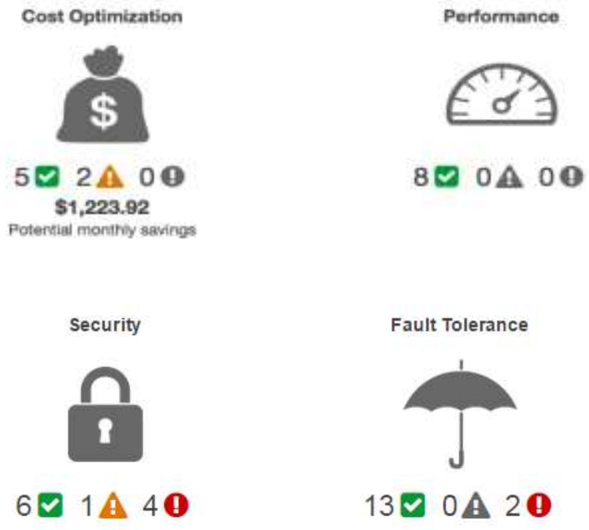
Uncovering the cost optimization opportunities

1. Auto-tag resources.
2. Identify always-on nonprod.
3. Identify instances to downsize.
4. Recommend RIs to purchase.
5. Dashboard our status.
6. Report on savings.

AWS options

AWS Trusted Advisor

Cost Explorer



Cost Optimization

5 ✓ 2 ⚠ 0 ❗

$1,223.92

Potential monthly savings

Performance

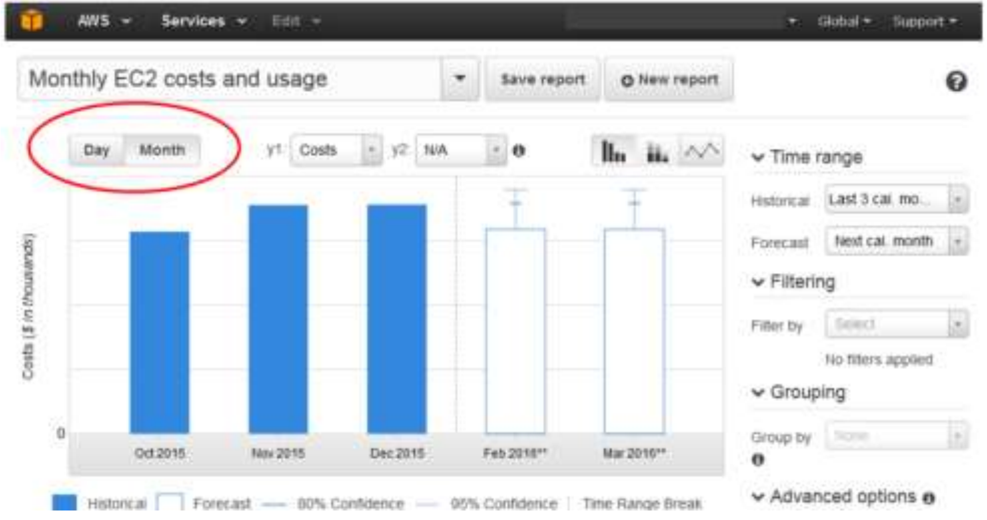
8 ✓ 0 ⚠ 0 ❗

Security

6 ✓ 1 ⚠ 4 ❗

Fault Tolerance

13 ✓ 0 ⚠ 2 ❗



Monthly EC2 costs and usage

y1: Costs y2: N/A

Historical: Last 3 cal. mo...

Forecast: Next cal. month

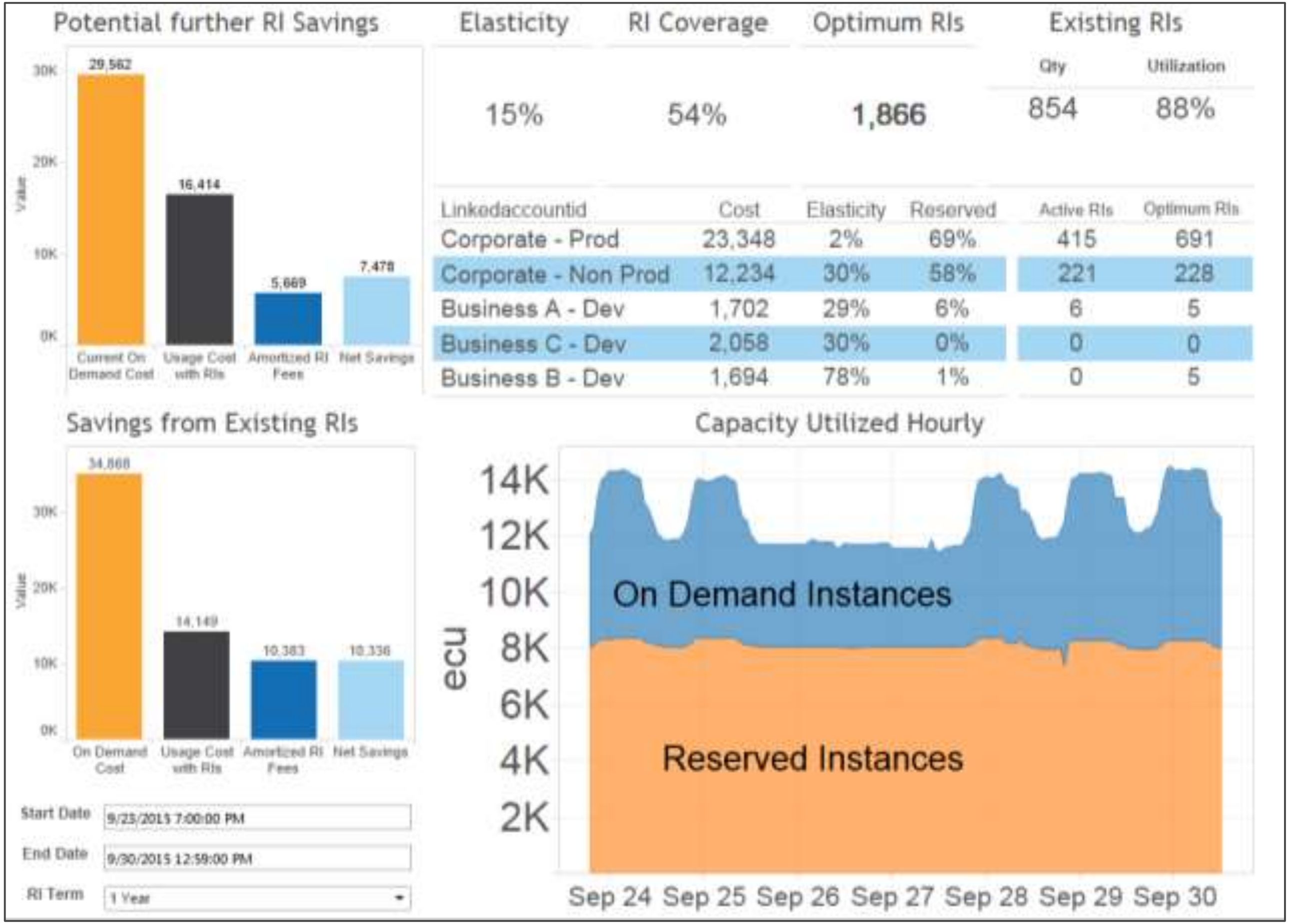
Filter by: Select

No filters applied

Group by: None

Reserved Instances and right-sizing options

Example: reasonable optimization dashboard



Elasticity	RI Coverage	Optimum RIs	Existing RIs	
			Qty	Utilization
15%	54%	1,866	854	88%

Linkedaccountid	Cost	Elasticity	Reserved	Active RIs	Optimum RIs
Corporate - Prod	23,348	2%	69%	415	691
Corporate - Non Prod	12,234	30%	58%	221	228
Business A - Dev	1,702	29%	6%	6	5
Business C - Dev	2,058	30%	0%	0	0
Business B - Dev	1,694	78%	1%	0	5

Start Date: 9/23/2015 7:00:00 PM

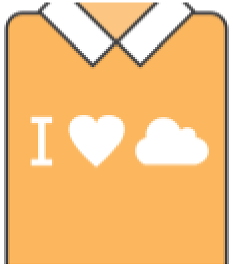
End Date: 9/30/2015 12:59:00 PM

RI Term: 1 Year

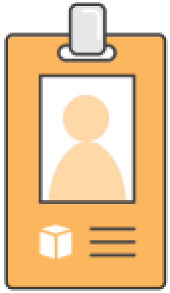
Creating a culture of cost transparency

Targets and metrics

Cloud Competency Center



AWS Enterprise Support



Cost Metrics

A company’s overall AWS cost should be evaluated as a unit cost ratio with respect to another defined metric:

$$Unit\ Cost = \frac{Total\ Cost}{Individual\ or\ Business\ Metric}$$

Examples

- Unit cost per revenue generated
- Unit cost per product or business unit
- Unit cost per internal user
- Unit cost per customer or subscriber

Putting it all together

Where to start

Set up a Cloud Competency Center

Bring in the right tools

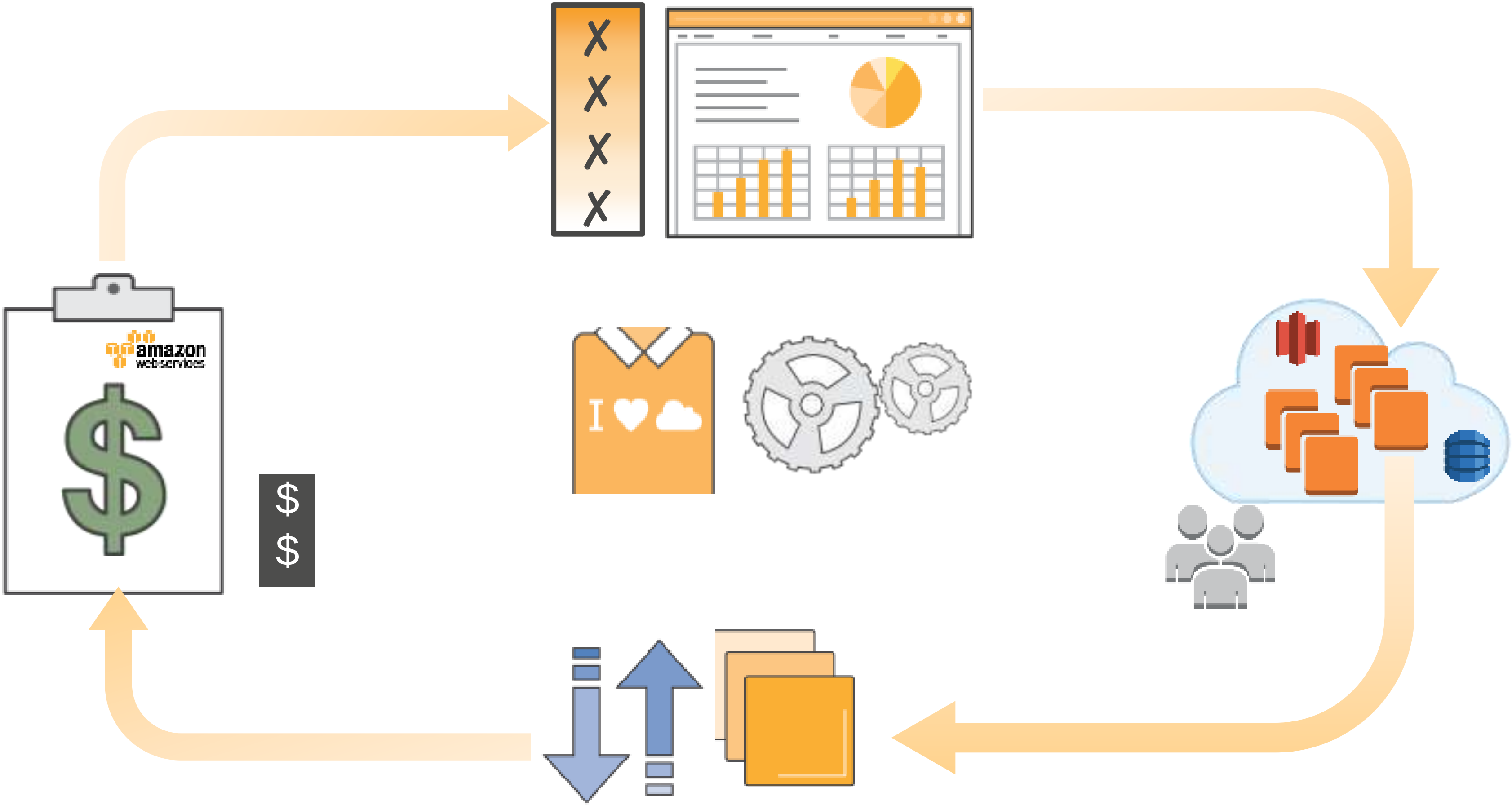
Use metrics to reinforce behavior

amazon web services | Partner Network

PREMIER CONSULTING PARTNER

Use partners to accelerate!

Cycle of cost optimization



Remember to complete your evaluations!

AWS SUMMIT

Thank You!

amazon web services