ENT202

Driving AWS Cost Efficiency at Your Company

Bill Lynch, VP of Product, Cloudability

December 1, 2016

© 2016, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

amazon web services

What to Expect from the Session

- How to evolve from cloud cost to cloud efficiency
- How to drive cost efficiency at your business
- Useful tips to get started

About me

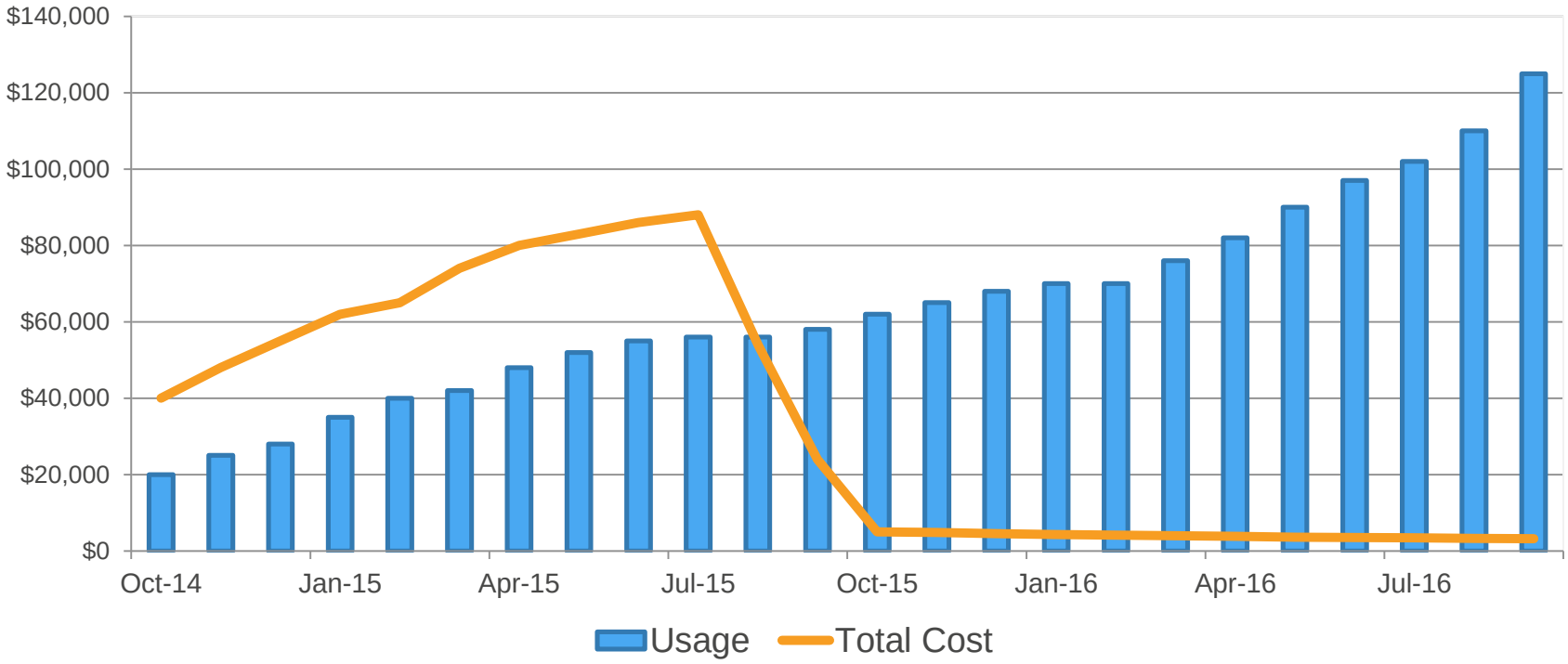
Bill Lynch

VP of Product, Cloudability

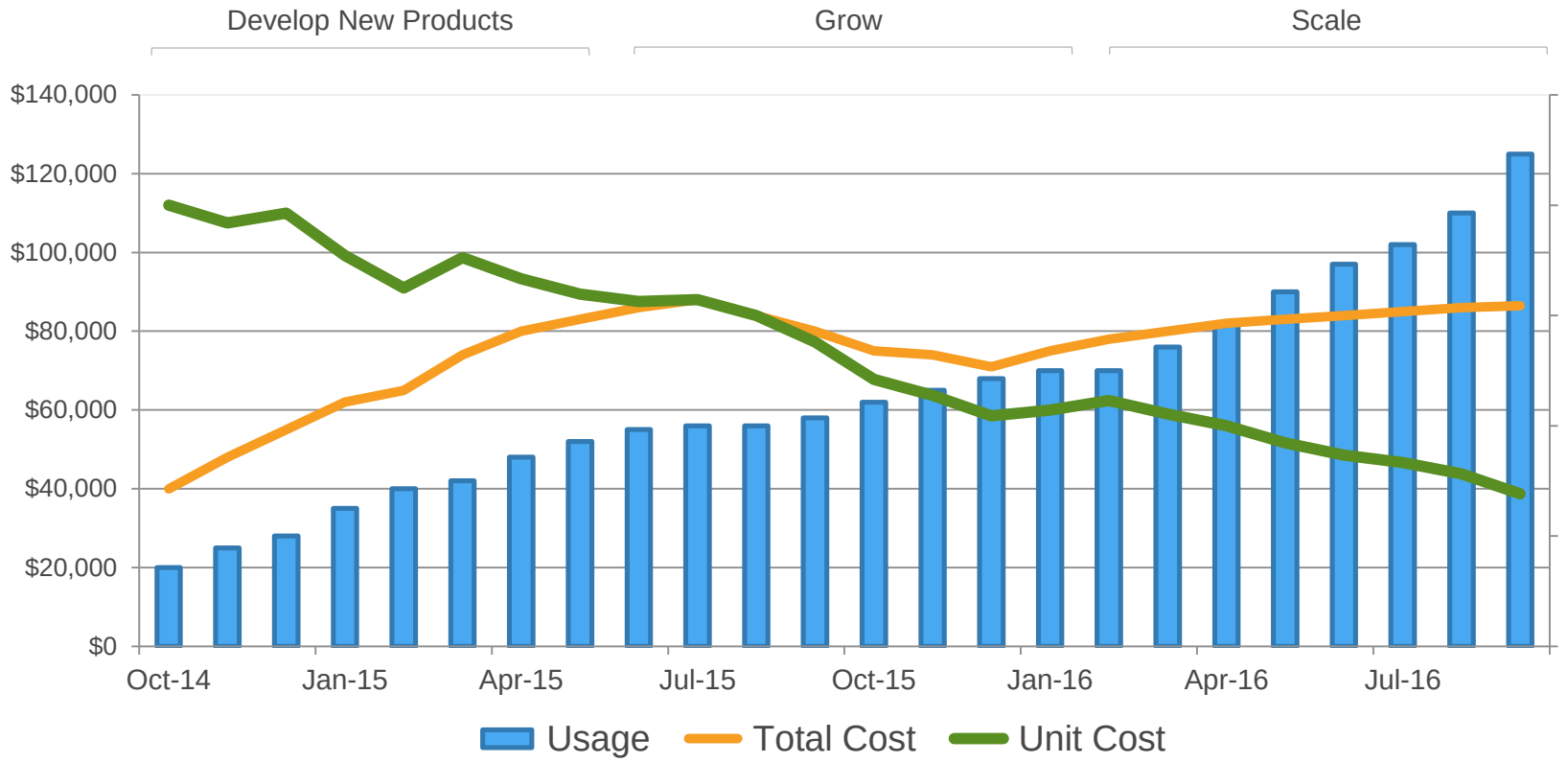
Cloud enthusiast

Examining efficiency

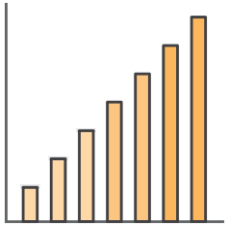
An unrealistic picture of cloud costs



When does efficiency matter?



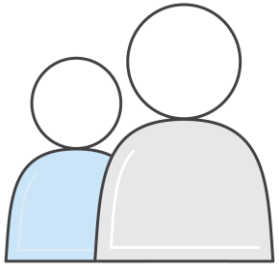
The elements of cost efficiency



Usage

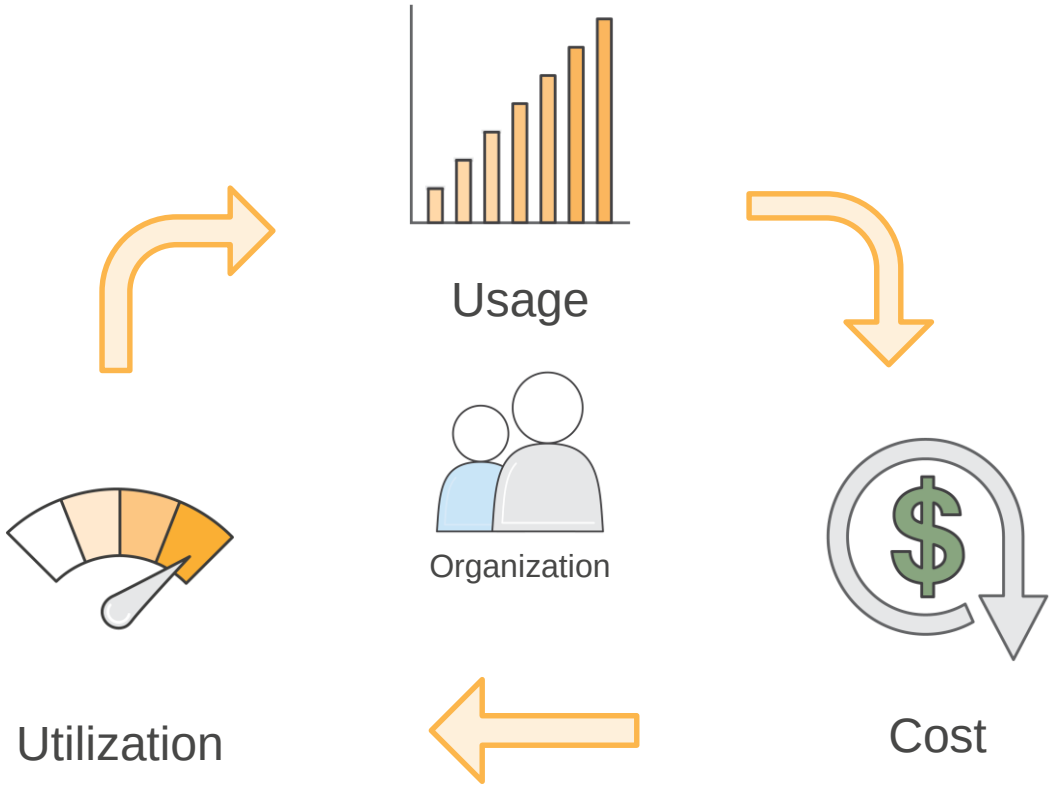
Cost

Utilization

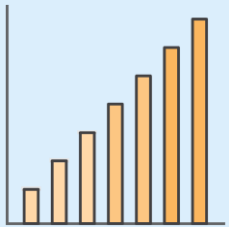


Organization

A virtuous circle



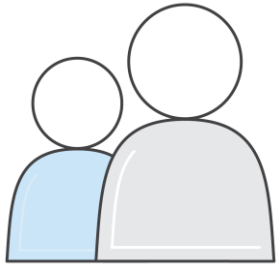
Usage efficiency



Usage

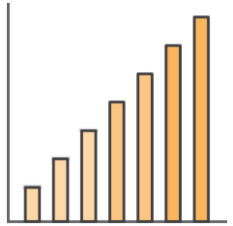
Cost

Utilization



Organization

Examining usage efficiency



Usage efficiency is about determining if you need to use a resource (e.g., an hour of compute).

All other optimizations are downstream.

- How do you measure your usage efficiency?
- What are the ways to affect your efficiency?

Looking at usage vs. usefulness

Amazon EC2

Generalized Compute

AWS Lambda

Functions, no servers

Attributes of usage and usefulness

Amazon EC2

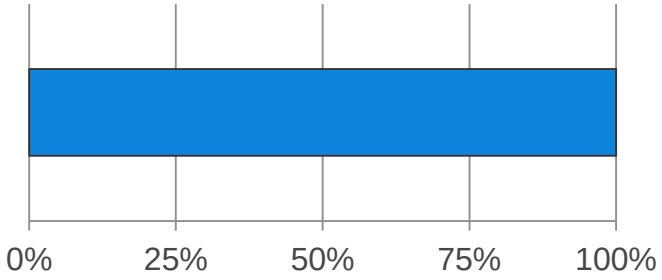
Generalized Compute

AWS Lambda

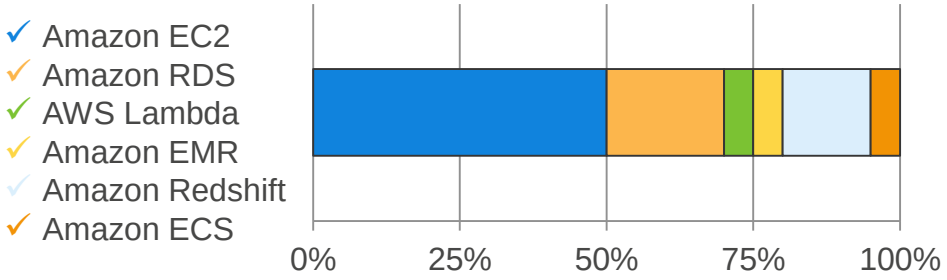
Functions, no servers

Waste Potential	High	Low
Constraints	Low	High
Payment	By Hour	By Use

Service usage efficiency



100% EC2 Usage
Likely usage inefficient



50% EC2, plus other services
Likely usage efficient

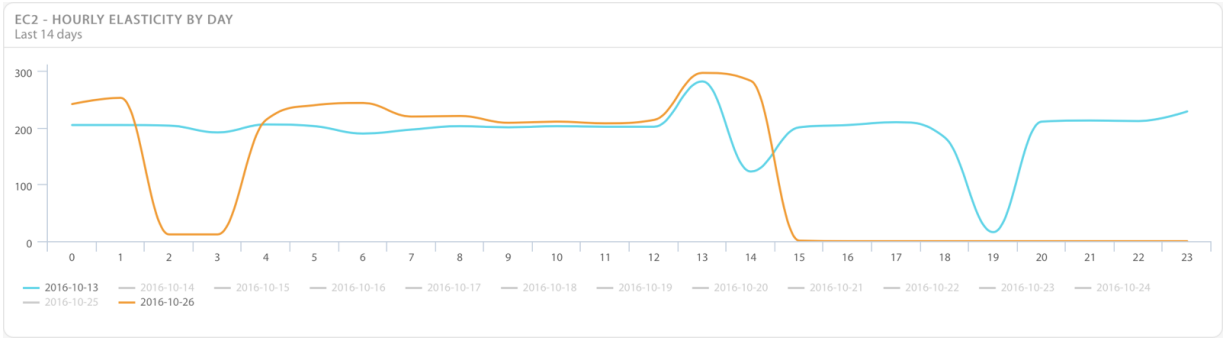
Cost efficiency strategies per service

Service	Waste Potential	Constraints	Strategy
EC2	High	Low	Optimize cost, utilization
ECS	Medium	High	Optimize cluster utilization
Lambda	Low	High	Optimize usage
RDS	Low	Medium	Optimize cost
Redshift	Low	High	Optimize cost

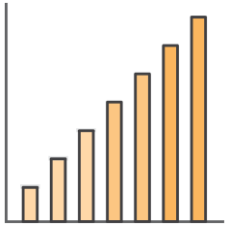
Tip: always look at per-service optimizations. The strategies you use will vary.

Tackling elasticity

- Very good proxy for usage efficiency
- Tag everything, and determine usage and spend by team or service
- Architect for and leverage Auto Scaling and/or the Spot market
- Measure and set goals (hours, cost)



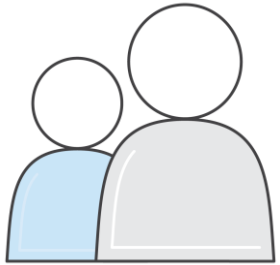
Cost efficiency



Usage

Cost

Utilization



Organization

Examining cost efficiency

Cost efficiency means you're paying the *best* rate for service usage. You know there is a theoretical minimum to pay, so how?

- How do you measure cost efficiency?
- What are the ways to achieve the best rate?

Reserved Instances

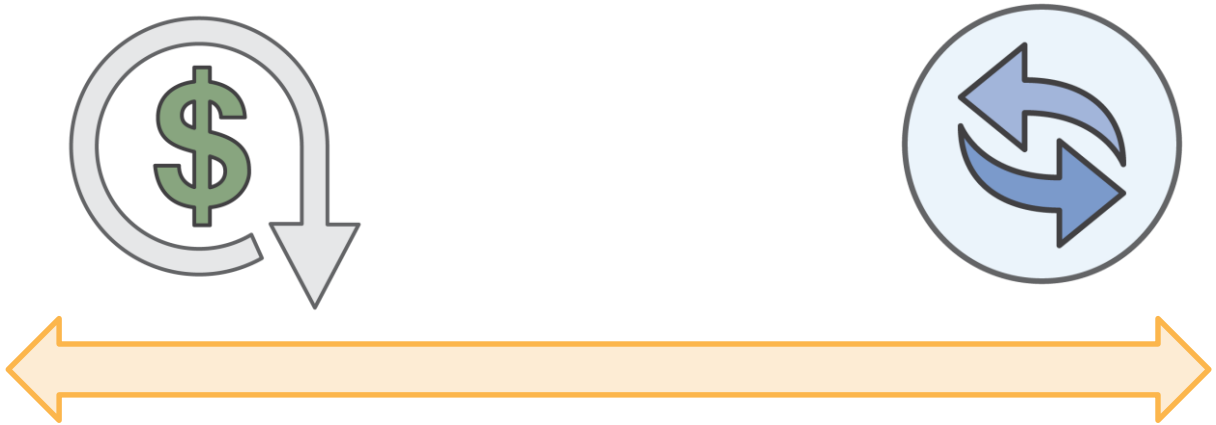
- Pricing discounts in exchange for a longer-term commitment
- Not a physical instance, more like a book of coupons
- Modify over time to better fit your compute

See also:

Re:Invent 2016 Session **ENT307**

“Saving at Scale with Reserved Instances”

Standard vs. convertible RIs



Standard RI
Max Savings

Convertible RI
Max Flexibility

Regional Benefit
Capacity vs. Utilization

Standard vs. convertible RIs

m4.xlarge on-demand pricing: $0.239/hr, $2,093.64/yr

	1 Year, Standard	3 Year, Convertible		3 Year, Standard
No Upfront	**31%** $0.164	**38%** $0.148		N/A
Partial Upfront	**41%** $0.140	**47%** $0.126	← 24% → $0.030	**60%** $0.096
All Upfront	**42%** $0.138	**48%** $0.124	← 28% → $0.035	**63%** $0.089

Price of flexibility
$306.60/yr

RI tips

- Determine business priorities: savings, flexibility, capacity
- Use regional benefit to minimize waste
- Buy centrally
- Optimal portfolio will be a mix of standard and convertible
- For all and partial upfront RIs you will need to amortize costs

Serverless cost efficiency

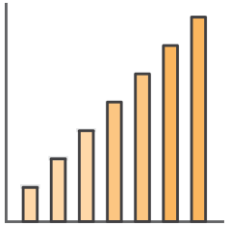
EC2 m4.xlarge: $0.239 / hr

Lambda 512 MB: $0.000000834* / invocation

- Optimize for usage
- Look at usage per account
- Track functional usage; group with other costs

* In English, 8.34 ten millionths of a dollar

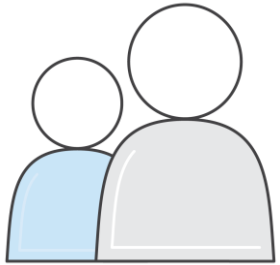
Utilization efficiency



Usage

Cost

Utilization



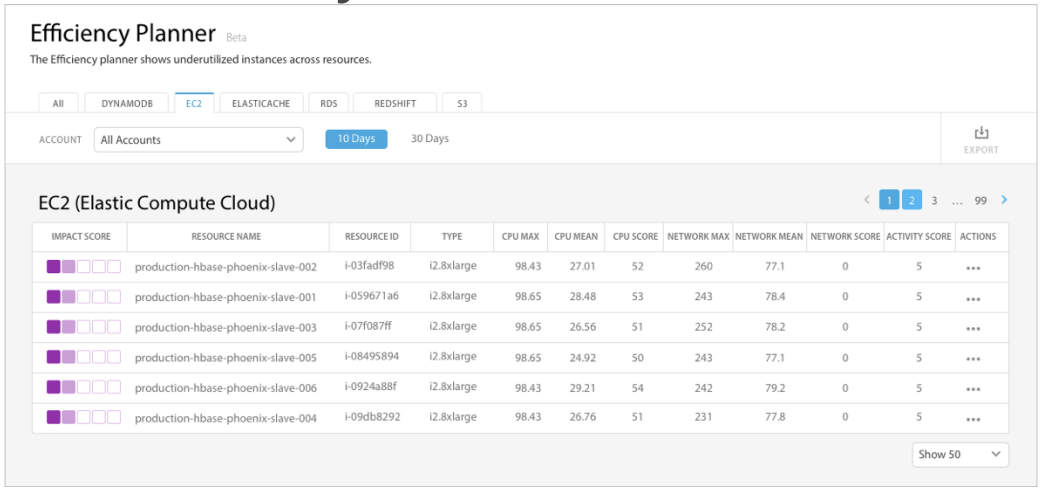
Organization

Examining utilization

- Assuming you needed the hour (usage), and you paid the lowest amount for it (cost), then did you maximize its value while using it?
- “Maximizing its value” is a relative term and entirely depends on what you value
- Measurement is critical

Measures of utilization

- Out-of-the-box metrics from AWS; you'll need to correlate with costs
- Custom metrics will improve understanding of utilization
- Know where to focus your efforts



Container efficiency

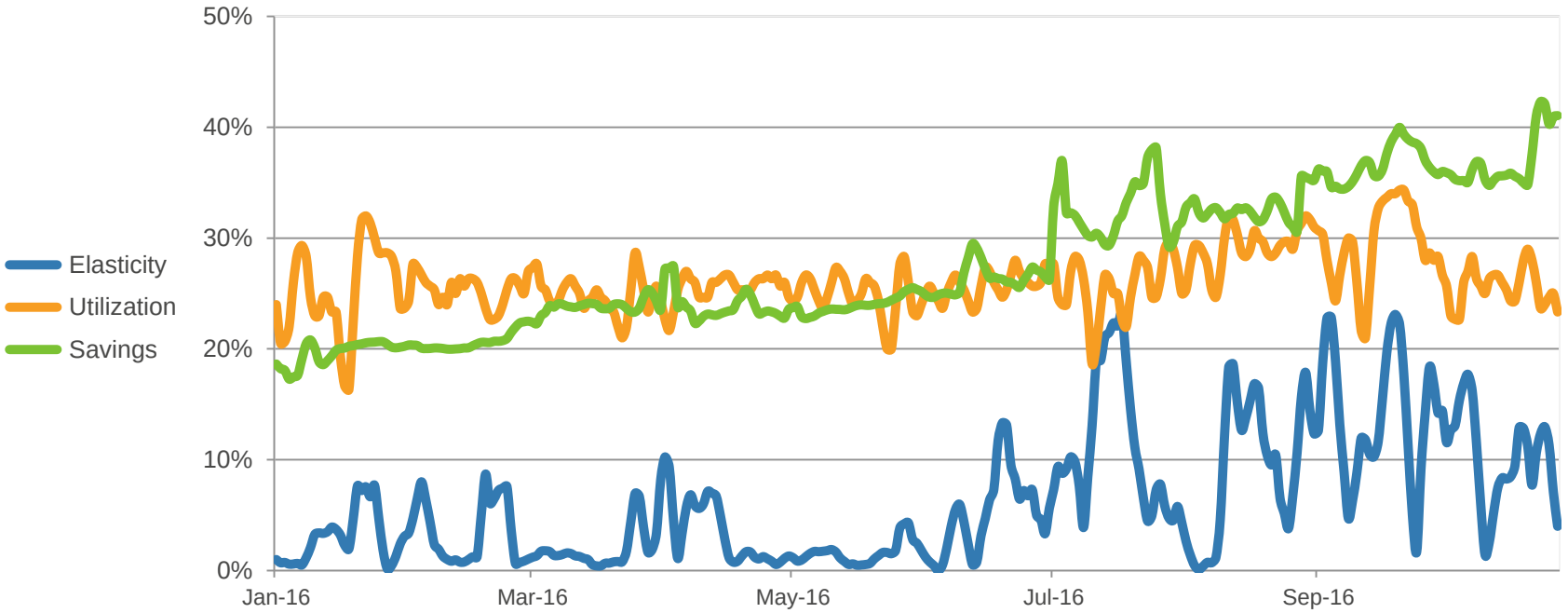
- Path to expert cost allocation is elusive, and it’s early
- For utilization, examine a cluster in the same way as an EC2 instance

EC2

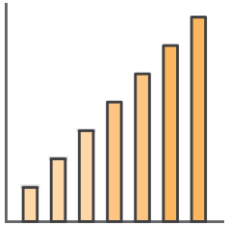
ECS

Metrics	Host	Cluster
Rightsize Action	Auto Scaling group, others	Cluster capacity

Customer efficiency example



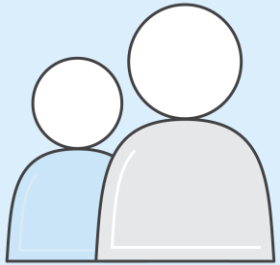
Organizational efficiency



Usage

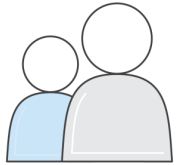
Cost

Utilization



Organization

Organizational efficiency



Cloud is different than traditional IT; therefore, the behavior to manage it needs to evolve. Start preparing now rather than waiting for unpleasant surprises.

- How quickly can you make decisions?
- And how fast can your company move while also managing spend?

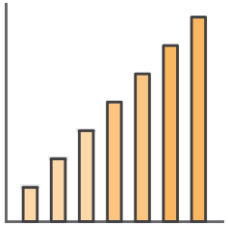
Roles involved in cost efficiency

- Ops and Finance, aka “fin-ops”
- Centralized RI buying is optimal; run through finance is even better, aka “RI Czar”
- Cost visibility at every level; awareness of impact of change

Tips for interfacing with your business

- Speak in terms of efficiency, particularly unit costs
- Tie to other business metrics like customer-facing KPIs
- Train finance on things they can control
- Make sure the right visibility in to usage and utilization efficiency exists for the rest of the business

Summary



Usage

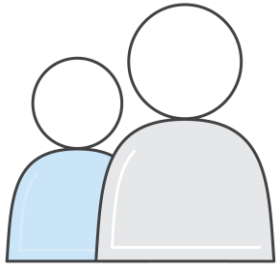
- Leverage usage-efficient services
- Invest in elasticity
- Optimize cost and utilization after usage

Cost

- Leverage RIs for more immediate savings; make sure to know business tradeoffs
- Architect for Spot and Auto Scaling

Utilization

- Right-size your resources to maximize value
- Determine what you don’t need



Organization

- Drive visibility and information sharing
- Assemble your team, especially an RI Czar
- Establish regular RI-buying meetings

Thank you!

amazon
web services

Remember to complete your evaluations!