AWS re:Invent

ARC219-R

AWS Cost Management tools for cost & usage optimization

Nathan Besh

Cost Lead, Well-Architected

Amazon Web Services

AWS re:Invent

aws

Agenda

Introduction

Governance

Visualization and analysis

Licensing

Right sizing

Pricing models

Bonus: Well-architected

Related breakouts

ARC204 Cost optimizing a workload (200-300)

ENT206 Optimize AWS costs and utilization with AWS management tools (100-200)

CMP410 Save up to 90 percent on CI/CD and test workloads (spot)

ARC209 Running lean architectures: How to be cost-effective on AWS (Presentation)

CMP207 Manage, control, and optimize costs with native AWS products (Chalk talk)

CMP323 Optimize performance and cost for your AWS compute (Presentation)

ENT204 Managing your cloud financials as you scale on AWS (Presentation)

ENT230 Introduction to forecasting costs on AWS (Chalk talk)

Introduction

- Implement fundamental (effective) cost and usage controls
- Analyze and manage your costs and usage
- Dual streams for each topic: 200 and 300 level
  - 200 level: Building understanding an familiarity
  - 300 level: Focusing on scale and efficiency
- Present then implement
  - A few slides on the topic
  - Hands-on labs
- Labs are available online
  - Stay together (don’t worry if you don’t finish)
  - Do them anytime

Governance

Governance

- Who can do what (usage)
- Broad, high-level controls (safety)
- Notifications (learn)
- Implement controls (manage)

Governance: Controls and notifications

- Controls: AWS Identity and Access Management (IAM)
  - Users, groups, policies
  - Roles (not covered today)
  - Limits (not covered today)

- Notifications: AWS Budgets
  - Soft approach (do not enforce)
  - Forecasted and actual
  - Cost, usage, reserved instance and savings plan coverage
  - Refine by accounts, tags, instances, and so on (~10 billing columns)

Lab

https://wellarchitectedlabs.com/Cost/arc219.html

Governance

Visualization and analysis

Visualization

- It can be hard to consume data (tables, CSVs)
- Trend analysis and anomalies
- Dashboards for visibility

- Chose the visualization carefully
  - Bar chart: Quick comparison, first slide, overview of the position

- Line chart (almost always)
  - Cloud: Highly variable usage/costs over time
  - Time on the x-axis with as many data points as permitted

Visualization: Value

- Cost optimization: Business value at the lowest price
- What is your (business) value?
  - How is it measured?

- Process
  - Start high level (total outputs per hour)
  - Slowly build granularity
    - Different types of output
    - Workload components/microservices

Analysis

- What, where, who, why
- Cost and usage report (CUR)

- Ask business questions
  - What happened
  - Trends (positive)
  - Chargeback
  - Large spend/usage (focus for cost optimization opportunities)

Analysis

- Tooling
  - AWS Glue
  - Amazon Athena
  - AWS Lambda (300)

- Uses
  - Basis of reporting within your organization
  - Modeling & forecasting
  - Answer any question on cost and usage

Lab

https://wellarchitectedlabs.com/Cost/arc219.html

Visualization and analysis

Licensing

Licensing

- What is the cost of a software license?
- What is the cost of running licensed software?

M5.xlarge: US-East-1 (N. Virginia)—Shared tenancy

	Linux	RHEL	Windows
Cost/hour	$0.192	$0.252	$0.376

T3.medium: US-East-1 (N. Virginia)—Shared tenancy

	Linux	RHEL	Windows
Cost/hour	$0.0416	$0.1016	$0.06

M5.xlarge: US-East-1 (N. Virginia)—Shared tenancy

- What's different when you launch?
  - Storage
  - Network traffic (updates, chatty OS)
- Console defaults, all free-tier eligible

	Linux + 8 Gb GP2	RHEL + 10 Gb GP2	Windows + 30 Gb GP2
Cost/hour	$0.193096 ($0.192)	$0.253370 ($0.252)	$0.380110 ($0.376)

M5.xlarge: US-East-1 (N. Virginia)—Shared tenancy

- Reserved instances
- 1 year, no up-front, regional/standard

	Linux + 8 Gb GP2	RHEL + 10 Gb GP2	Windows + 30 Gb GP2
Cost/hour	$0.124096 ($0.192)	$0.184370 ($0.252)	$0.311110 ($0.376)

T3.medium: US-East-1 (N. Virginia) —Shared tenancy

- Reserved instances & storage

	Linux	RHEL	Windows
Cost/hour	$0.027096	$0.087370	$0.049110

Licensing

- Understand the total cost of running software
  - Licensing
  - Additional resources (storage, CPU, memory)
  - Operations – advanced features
  - People/skills

- Ensure you get value for spend

Right sizing

Right sizing

- Impactful: One size = half your costs
- Half CPU ~43% reduction of costs (c5.xl -> m5.l)
- Half memory ~10% fewer costs (m5.l -> c5.l)

- Not just compute: database, storage
- Storage is linear: ½ storage = ½ costs
- Storage: PIOPS, type/class

Right sizing

- Factor in total cost (testing, operations)
- Enhance with Amazon CloudWatch (memory)
- Factor in your pricing models
- Become efficient: filter—accounts, tags

- Recommendation process
  - Previous 14 days (seasonal trends)
  - Max CPU <= 1% ➔ terminate
  - 1% <= max CPU <= 40% ➔ modify
  - If future CPU < 80% ➔ recommend

Lab

https://wellarchitectedlabs.com/Cost/arc219.html

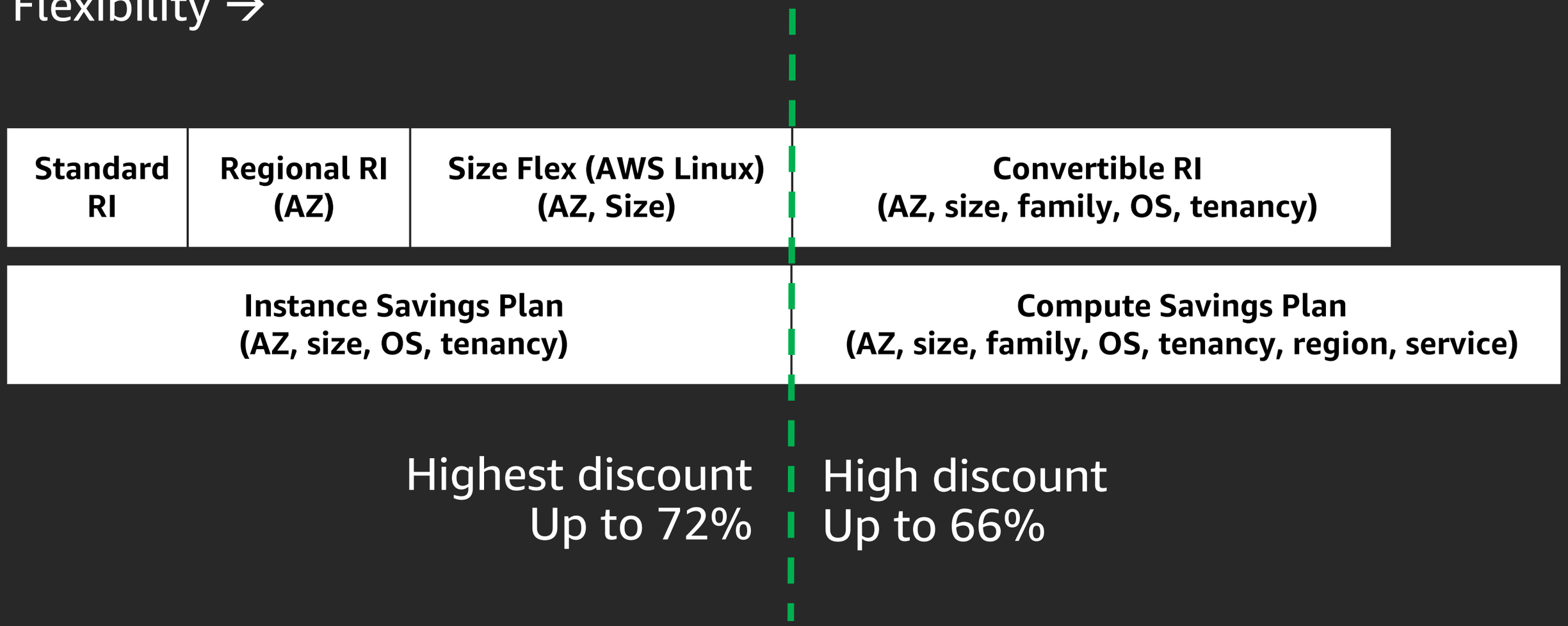
Right sizing

Pricing models

Pricing models

- On Demand (OD)
- Spot: instances, defined duration, fleet
- Reserved Instances (RI)
- Savings Plan (SP)

Pricing models – flexibility & discounts



SP order of operation

- Start in the account that purchased the SP
- Apply SP commitment from highest discount to lowest discount
- Exhaust all savings in this account
- Look at usage across all other accounts
- Apply SP commitment from highest discount to lowest discount

Coverage vs. utilization (RI)

- 100 instances of Linux, m5.large, 1-year commit, no upfront
- On Demand: $84,096
- RI discount = 36%

- 50% coverage (50 RI/50 OD): $68,766
- 25% coverage (25 RI/75 OD): $76,431

- Purchase 128 RI (need 100): $68,398
- Purchase 143 RI (need 100): $76,413

Coverage vs. utilization (SP)

- 100 instances of Linux, m5.large, 1-year commit, no upfront
- On Demand: $84,096
- 100% coverage instance SP @ 36%: $53,436
- 100% coverage compute SP @ 26%: $62,196

74 instances running: $62,231 On Demand

64 instances running: $53,821 On Demand	
Instance SP: $385 (savings)	Compute SP: $8,375 (loss)

Pricing models = commitment = investment

- What is the risk?
  - Risk = investment over the full commitment
  - How quickly can I pay off the full commitment with savings?
  - Formula = (monthly discounted amount x 12) / (monthly On Demand)
  - Break even = wrong
- What is the return?
  - Savings from discount
  - Variable
- What if it is unused?
  - Don’t worry

Lab

https://wellarchitectedlabs.com/Cost/arc219.html

Pricing models

Bonus: Well-architected tool

Cost optimization: Where should I focus?

- Well-architected tool
- Nine questions
  - Different focus areas of cost optimization
  - Best practices in each area
- Open discussion within your organization
  - No right or wrong
  - How you are achieving the outcomes
- Free learning tool for your organization
- Track your progress and improvement

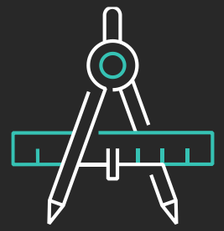
Lab

https://wellarchitectedlabs.com/Cost/arc219.html

Well-architected tool

Learn to architect with AWS Training and Certification

Resources created by the experts at AWS to propel your organization and career forward



Free foundational to advanced digital courses cover AWS services and teach architecting best practices

Classroom offerings, including Architecting on AWS, feature AWS expert instructors and hands-on labs

Validate expertise with the **AWS Certified Solutions Architect - Associate** or **AWS Certification Solutions Architect - Professional** exams

Visit aws.amazon.com/training/path-architecting/

aws training and certification

Thank you!

Nathan Besh

natbesh@amazon.com

Please complete the session survey in the mobile app.