
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

CMP319

Optimize for cost and availability with capacity management

Brian Phillippi

Principal Product Manager – Tech, Amazon EC2
Amazon Web Services

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- Function of capacity management
- Capacity availability
- Capacity management lifecycle
- Optimize across cost, performance, and availability

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Function of capacity management

HIGHEST PERFORMANCE AT THE LOWEST COST



Performance

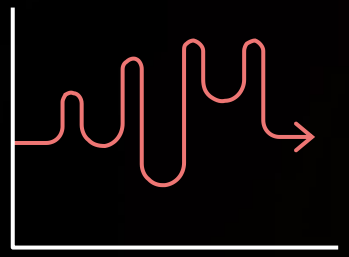


Cost

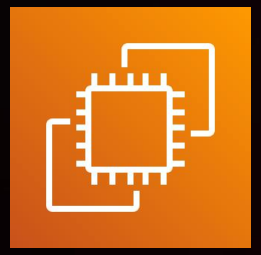
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EC2 On-Demand



Pay for
what you use



500+
instance types



84 Availability
Zones

aws

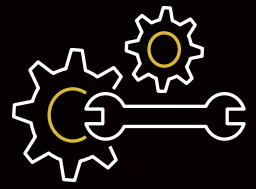
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Situations needing higher levels of availability

FOR BUSINESS CRITICAL WORKLOADS



Disaster recovery



Maintenance



Regulatory requirements



Peak events

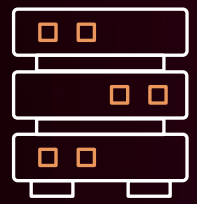
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Increasing availability



Flexibility



Capacity Reservations

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Becoming Flexible



Compute Optimizer



Amazon EC2
Auto Scaling



Amazon EC2 Spot



AWS Support

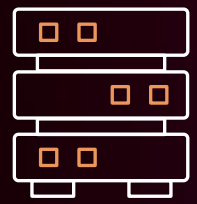
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Increasing availability



Flexibility



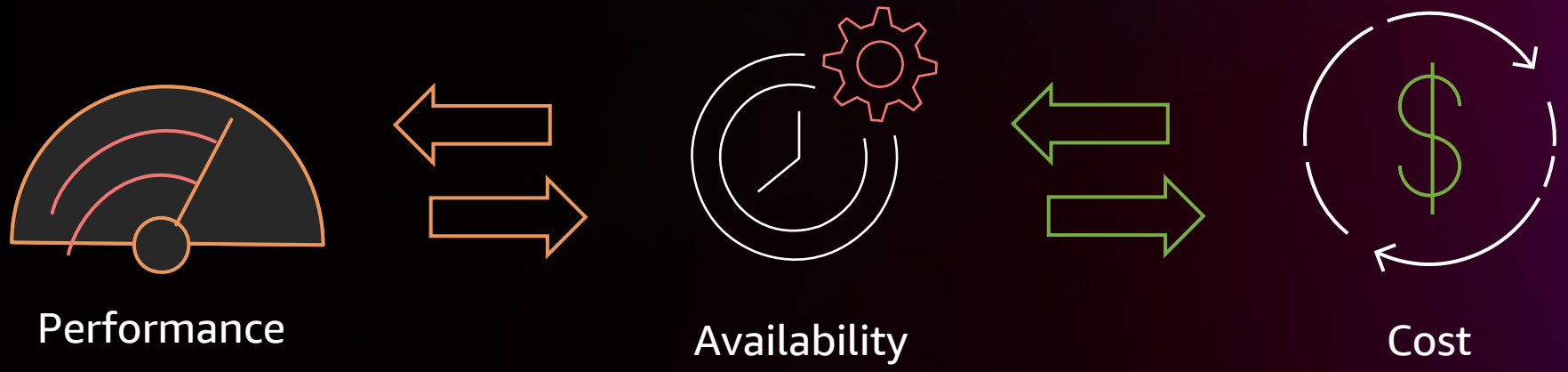
Capacity Reservations

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Capacity management success



The key to successful capacity management is finding the right balance between performance, availability, and cost for your workloads

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Capacity management lifecycle

Changing business needs

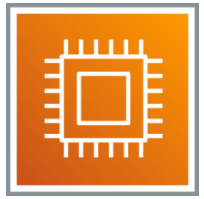
Plan



Gather information

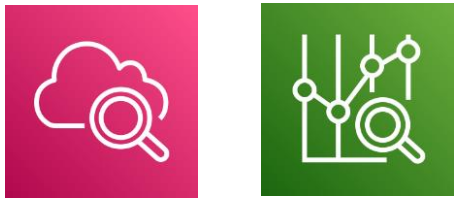
Pre-optimize

Acquire



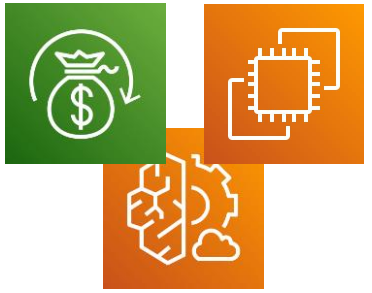
Launch or procure instances

Monitor



Discover the insights you need to take action

Optimize



Optimize for cost, availability and performance

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Plan – Considerations

GATHER INFORMATION, PRE-OPTIMIZE

- What instance type/capabilities best meet my needs?
- How flexible can I be?
- Can my workload run on AMD or Arm?
- Am I scaling up gradually or as a step function?
- What does my workload look like on a time series?
- Do I need to have a specific type of capacity?
- What will I do if I can’t get that capacity?
- How long will I need to run this workload?

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

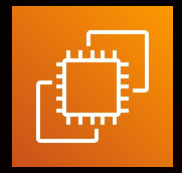
Acquire - Considerations

LAUNCH OR PROCURE INSTANCES

- Does the tool allow me to express flexibility?
- What will I do if I can't get the exact capacity that I need?
- How do I get capacity reservations for an upcoming event?



Auto Scaling groups



Run instances



Amazon EMR



Capacity Reservations

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Monitor

DISCOVER THE INSIGHTS YOU NEED TO TAKE ACTION

- Are my capacity reservations being fully utilized?
- Are there other accounts I can share my reservations with?
- What should I cover with a capacity reservation?
- Where am I underutilizing my savings plans?
- Do I want to be notified of an anomaly?



CloudWatch



AWS Cost Explorer



AWS Cost and Usage Report

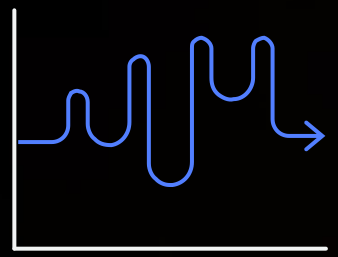
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS has multiple pricing models to optimize your costs

On-Demand

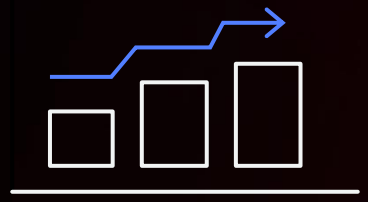
Pay for what you use with no long-term commitments



Stateful spiky workloads

Savings Plans

Up to 72% off On-Demand prices for 1- or 3-year hourly usage commitments



Committed and steady-state usage

Spot

Spare capacity at up to 90% off On-Demand prices



Fault-tolerant, flexible, stateless workloads

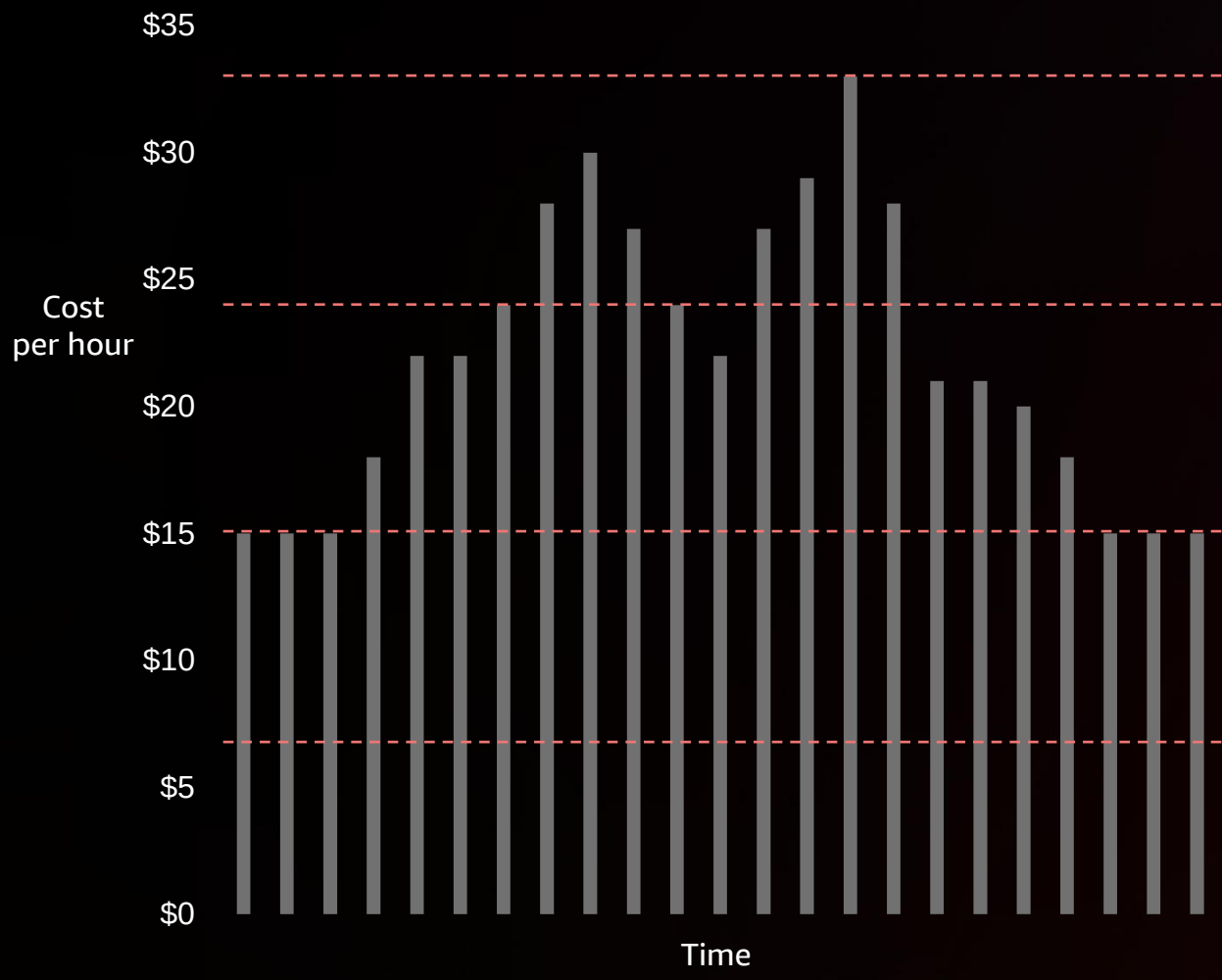
The best practice is to combine all purchase options

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Optimize cost

COMBINE PURCHASE MODELS TO ACHIEVE A COST OPTIMIZED ENVIRONMENT



Spiky, unpredictable EC2 usage that is not suitable for Spot

On-Demand
List price

Workloads that are fault-tolerant, flexible and stateless

Spot instances
Discounts up to 90%

Steady state usage where the flexibility to change is important

Compute Savings Plans
Discounts up to 66%

Steady EC2 usage unlikely to change during the commitment time frame

EC2 Instance Savings Plans
Discounts up to 72%

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Airbnb optimizes usage and costs by using Savings Plans and actionable cost data on AWS

Challenge

Airbnb wanted to expand its technical capabilities so that it could grow sustainably

It sought to improve operational efficiency by reducing costs and delivering actionable business metrics

Solution

The company relies on multiple AWS solutions, including Amazon Cost and Usage Reports, to manage capacity, optimize storage and compute capacity, and deliver insights that inform business and technical decision-making

Benefits

- Built a custom cost and usage data pipeline
- Reduced storage costs by about 27%
- Optimized compute spending using Savings Plans

- Company: Airbnb
- Industry: Travel
- Country: United States
- Website: airbnb.com



"Using Savings Plans has been a significant improvement to Airbnb's cloud management process. It's helped reduce our operational workload while also driving meaningful cost savings for our business."

Ari Siegel, Senior Finance Manager, Airbnb

About Airbnb

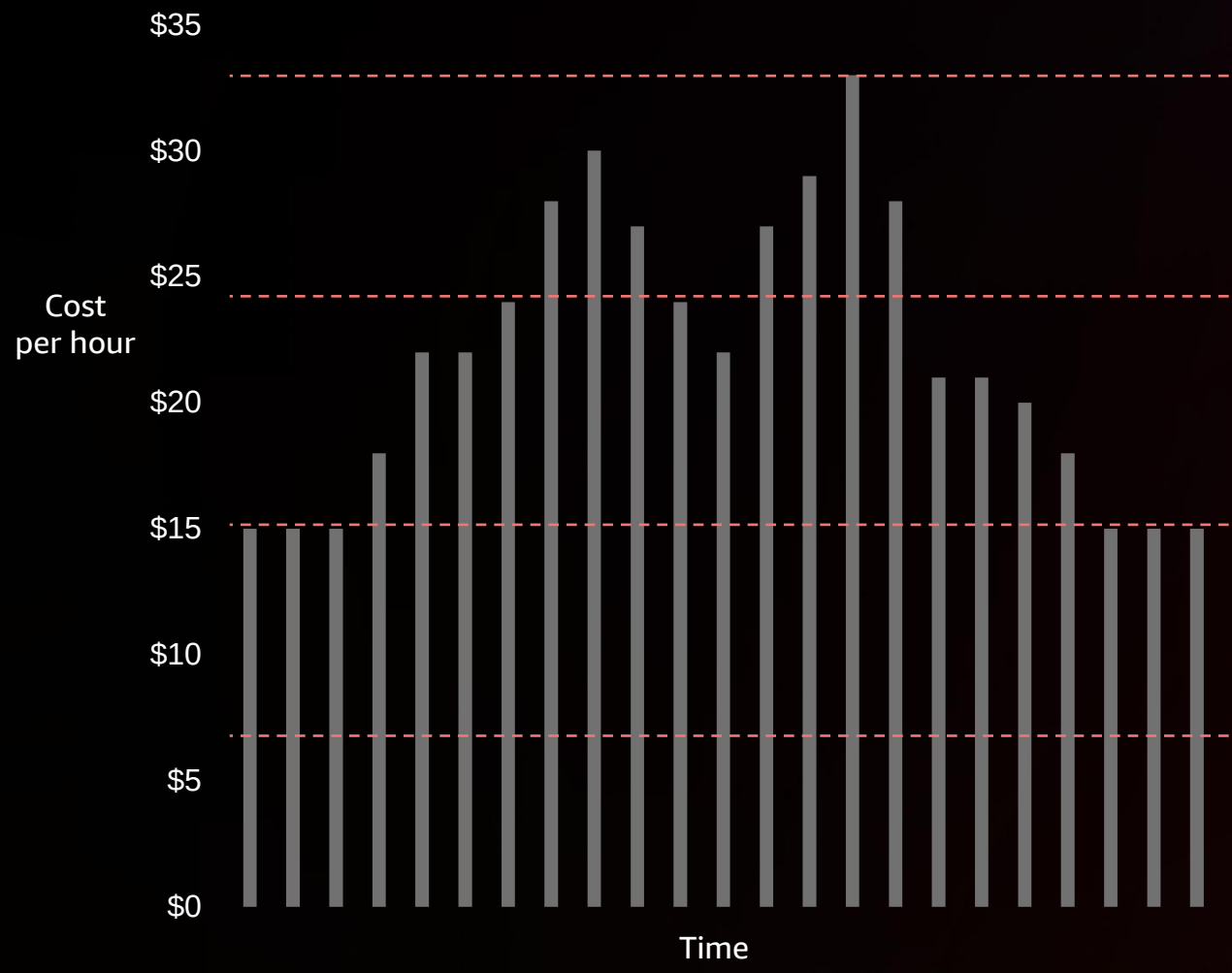
Airbnb was born in 2007 when two hosts welcomed three guests to their San Francisco home

It has since grown to four million hosts who have welcomed more than one billion guest arrivals across over 220 countries and regions

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Optimize availability

COMBINE AVAILABILITY MODELS ACHIEVE A COST-OPTIMIZED ENVIRONMENT



Spiky, unpredictable EC2 usage that is not suitable for Spot

On-Demand
Instance type/ AZ flexibility

Workloads that are fault-tolerant, flexible and stateless

Spot Instances
Instance type/ AZ Flexibility

Steady state usage where the flexibility to change is important

Capacity Reservations
and
On-Demand

Steady EC2 usage unlikely to change during the commitment time frame

Capacity Reservations
Hold for as long as you want

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Brian Phillippi
phbrian@amazon.com

LinkedIn :





Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.