AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

COP207

Inside Amazon operations: How Amazon Alexa gains observability

Justin Brown (he/him)
Senior Solutions Architect
Alexa Excellence

Bobby Hallahan (he/him)
Senior Specialist Solutions Architect
AWS Observability

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

About us



Bobby Hallahan is a Sr. Specialist Solutions Architect on the AWS Observability team

He is passionate about helping customers find innovative solutions to difficult problems



Justin Brown is a Sr. Solutions Architect with the Alexa Excellence team

He has been leading operational cost efficiency, optimization, and observability initiatives for the past 3 years

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Observability overview

Observability with Amazon Alexa Team

Cost attribution

Managing underutilized resources

Functional efficiency

Recent release

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Observability overview

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is observability

A PROPERTY OF A SYSTEM

In control theory, **observability** is a measure of how well internal states of a system can be inferred from knowledge of its external outputs

Source: Wikipedia, “Observability.” https://en.wikipedia.org/wiki/Observability

Observability is a property of a workload, just like the following properties

- Highly available
- Fault-tolerant
- Scalable
- Loosely coupled
- Observable

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is observability

A PROPERTY OF A SYSTEM

In control theory, **observability** is a measure of how well internal states of a system can be inferred from knowledge of its external outputs

Source: Wikipedia, “Observability.” https://en.wikipedia.org/wiki/Observability

- **Observability** is a property of a workload
- **Observable** is an adjective – a system or workload is observable
- **Monitoring** is the act of taking data from an observable system and determining health/performance, and it is a verb – you **monitor** an **observable** system

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Foundation for observability



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Business goals

WHY DO CUSTOMERS NEED OBSERVABILITY



Create new revenue streams



Improve operational and financial efficiency



Lower business risk

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Business goals

WHY DO CUSTOMERS NEED OBSERVABILITY



Create new revenue streams



Improve operational and financial efficiency



Lower business risk

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Observability with Amazon Alexa team

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Operational observability

Where are you on your operational observability journey?

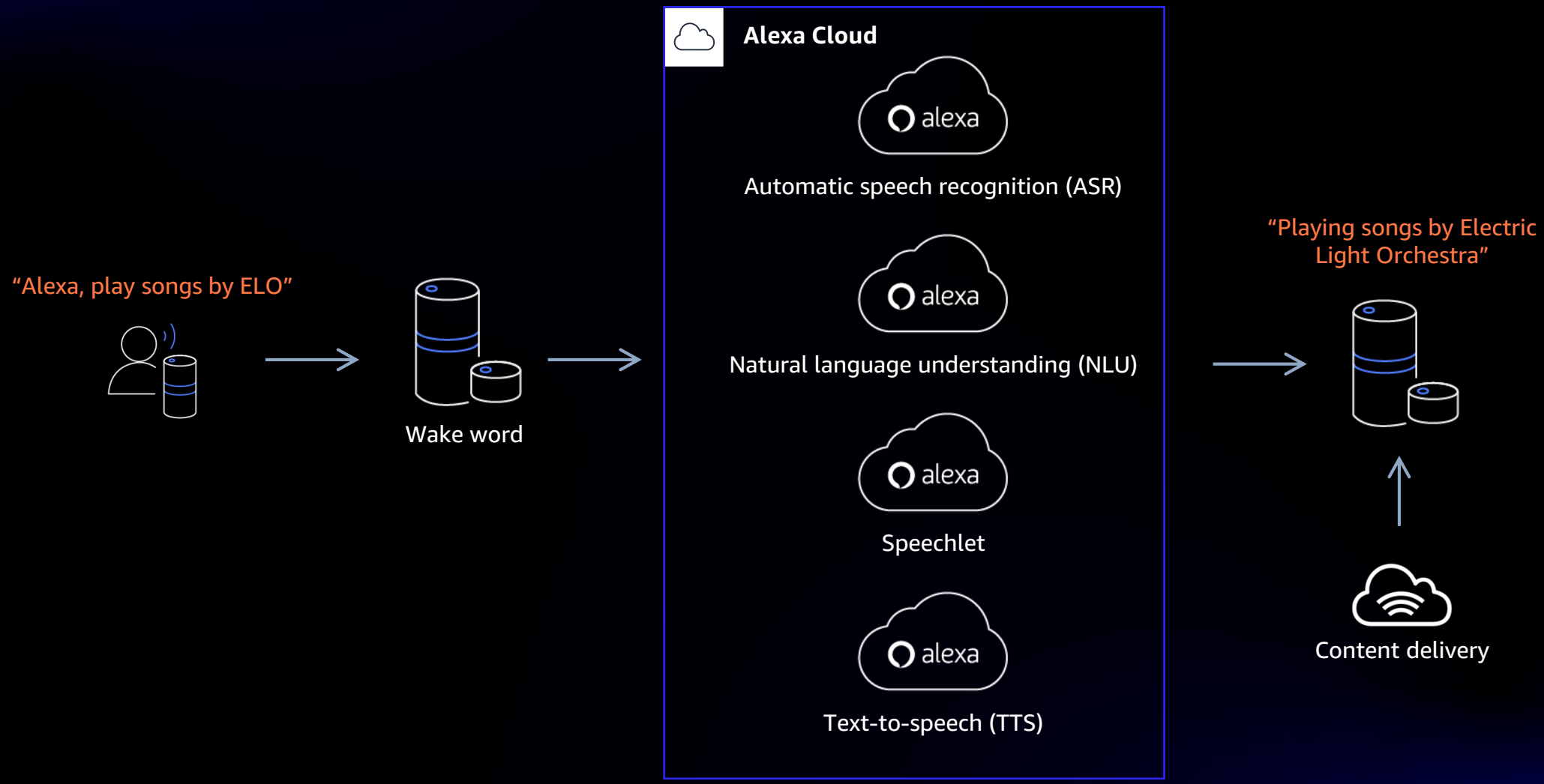
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is Amazon Alexa?

Alexa, what are you?

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How does Alexa work?



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Our operational challenges

How might your environment experience similar challenges?

- Large number of distributed resources and accounts
- Fast-paced experimentation and innovation
- Highly variable infrastructure use cases and architecture patterns
- Globally distributed collaboration
- Complex multi-component orchestration
- High standards for end-to-end SLAs and SLAs between components
- Multiple entry points to the Alexa ecosystem

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Operational observability

What happens if we don’t invest in operational observability?

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Operational observability

How efficiently am I investing in infrastructure that delivers real value to my customers?

- Cost attribution
- Managing underutilized resources
- Functional efficiency



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Guiding principles

Where do I start and what should I consider?

- Shared core operational framework
- Focus on iterations over mature-state vision
- Take stock of what is available
- Consider automation of patterns and best practices
- Reevaluate operational trends often
- Doesn’t have to be complex to be effective



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The Alexa journey

Operational observability with Amazon EC2

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cost attribution

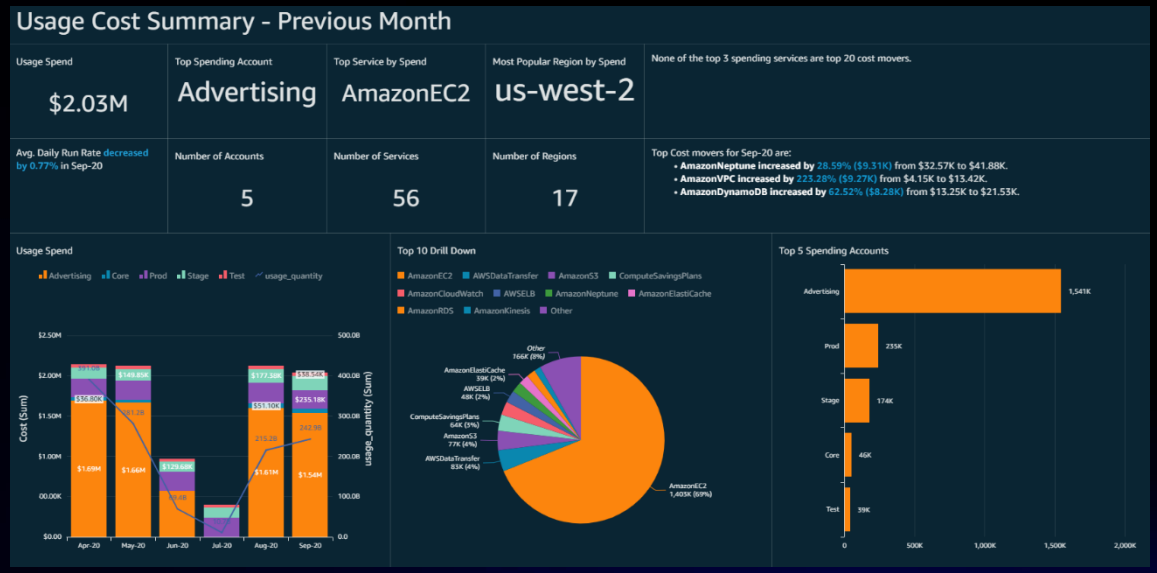
Where are my infrastructure investments going?

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Show me the money

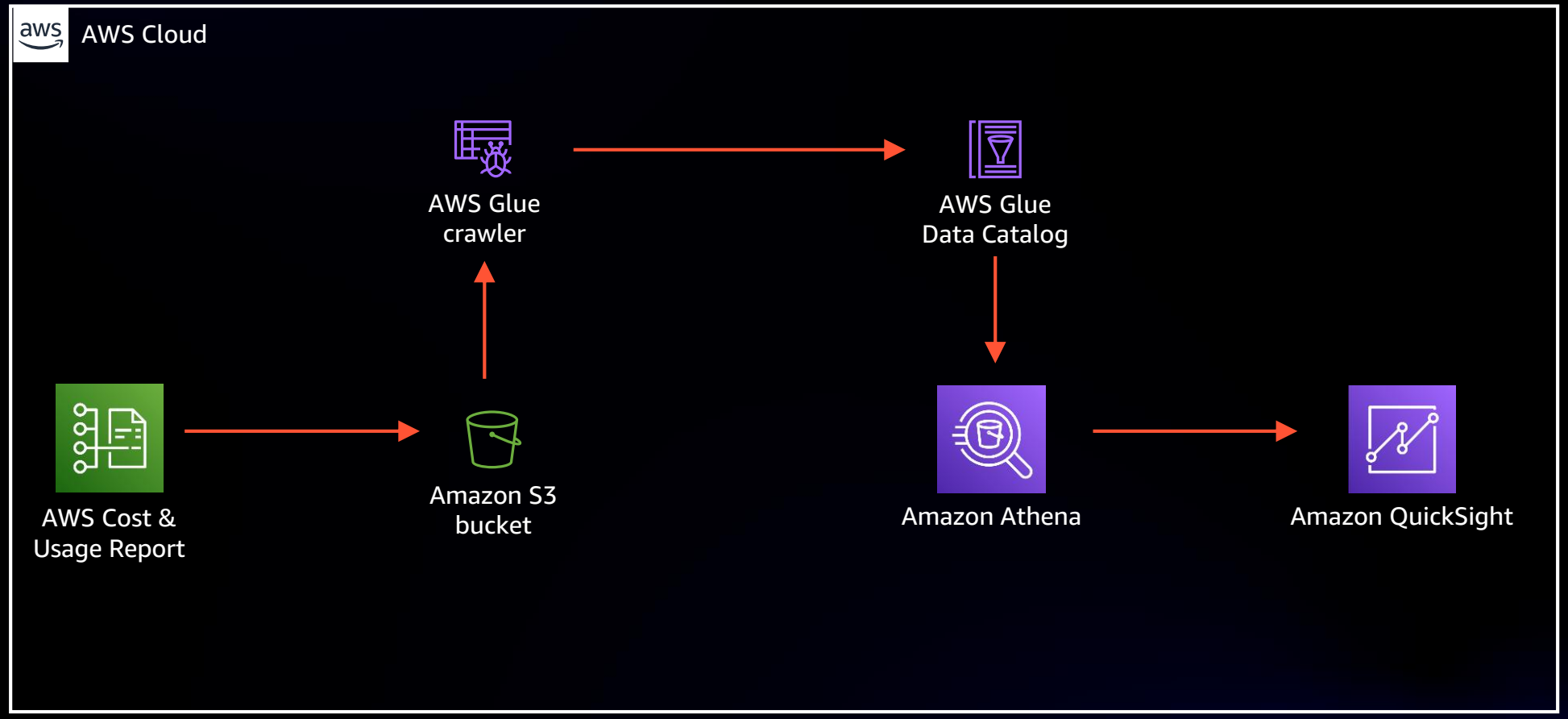
Can I trace my customer priorities to my infrastructure investment?

- Distributed teams make decisions to fit their use case
- Provisioning options and billing models
- Resource cost allocation tagging
- Trends and forecasting



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Cost & Usage Report with Amazon QuickSight



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Cost & Usage Report with Amazon QuickSight



https://www.wellarchitectedlabs.com/cost/200_labs/200_enterprise_dashboards/

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Managing underutilized resources

Baseline infrastructure optimization

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Infrastructure investment value

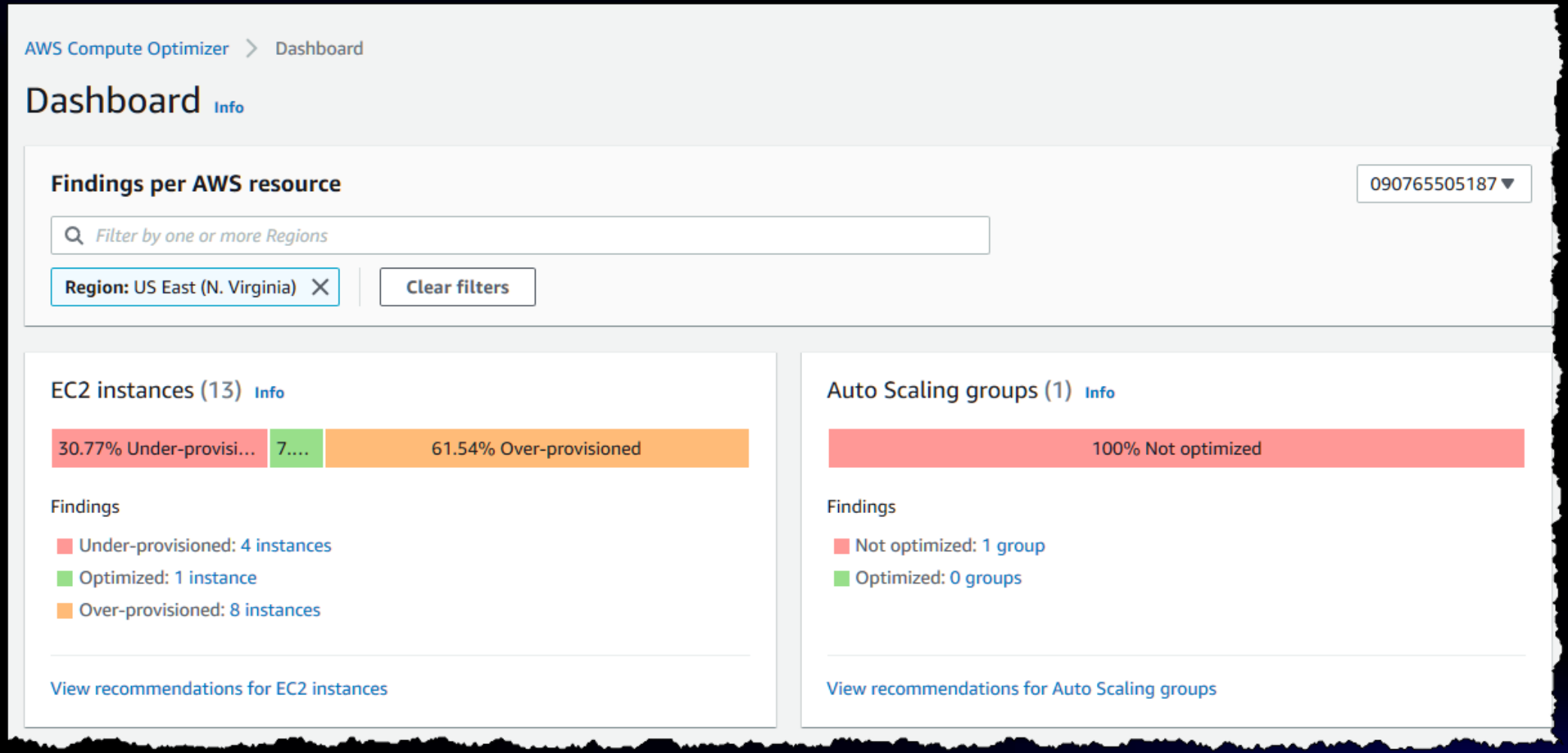
Where can I optimize resources when they are not delivering value?

- AWS Compute Optimizer and AWS Trusted Advisor
- Data transformation and visualization
- Automated guardrails and cleanup
- CloudWatch metrics aggregation



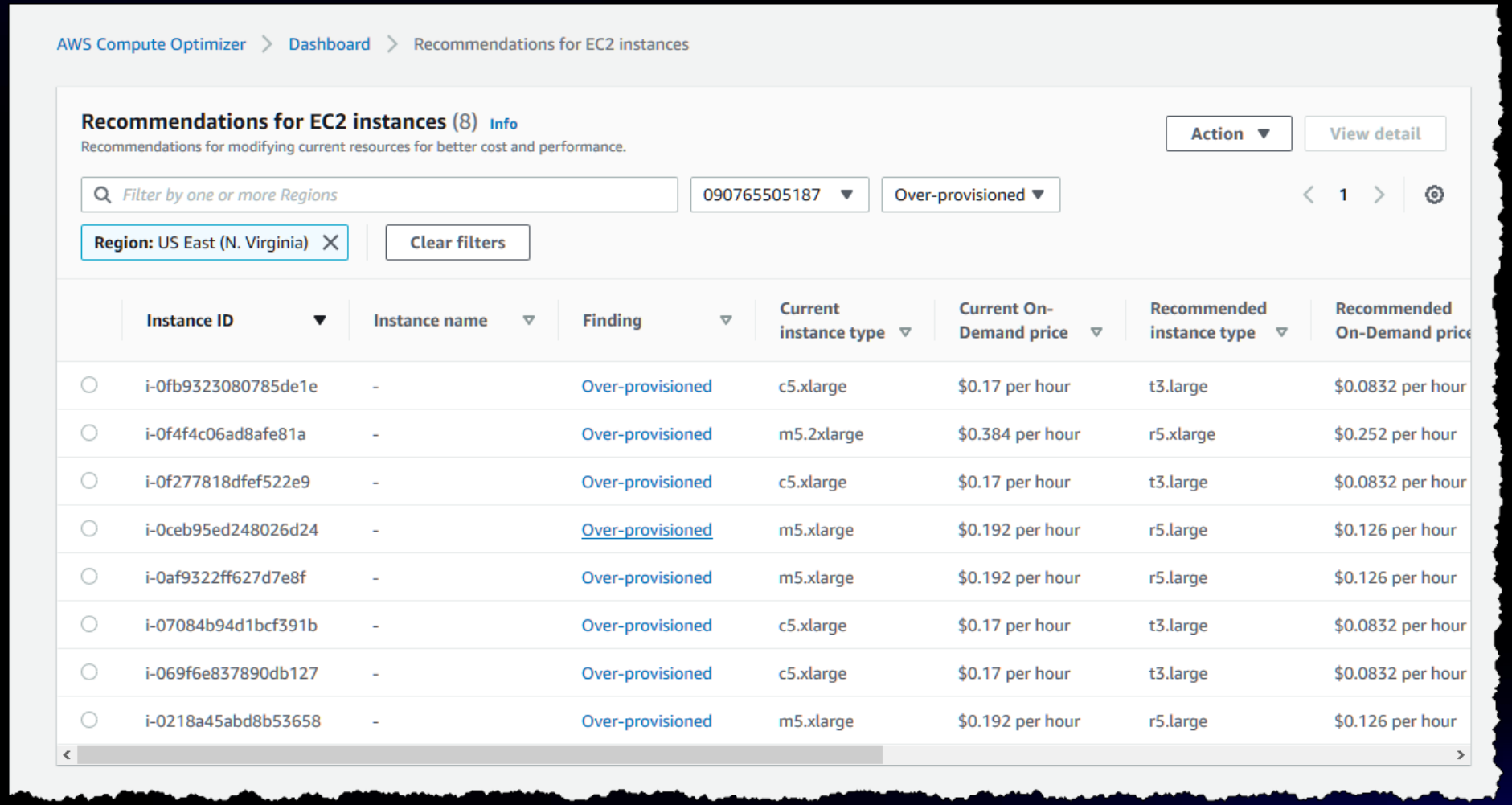
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Compute Optimizer



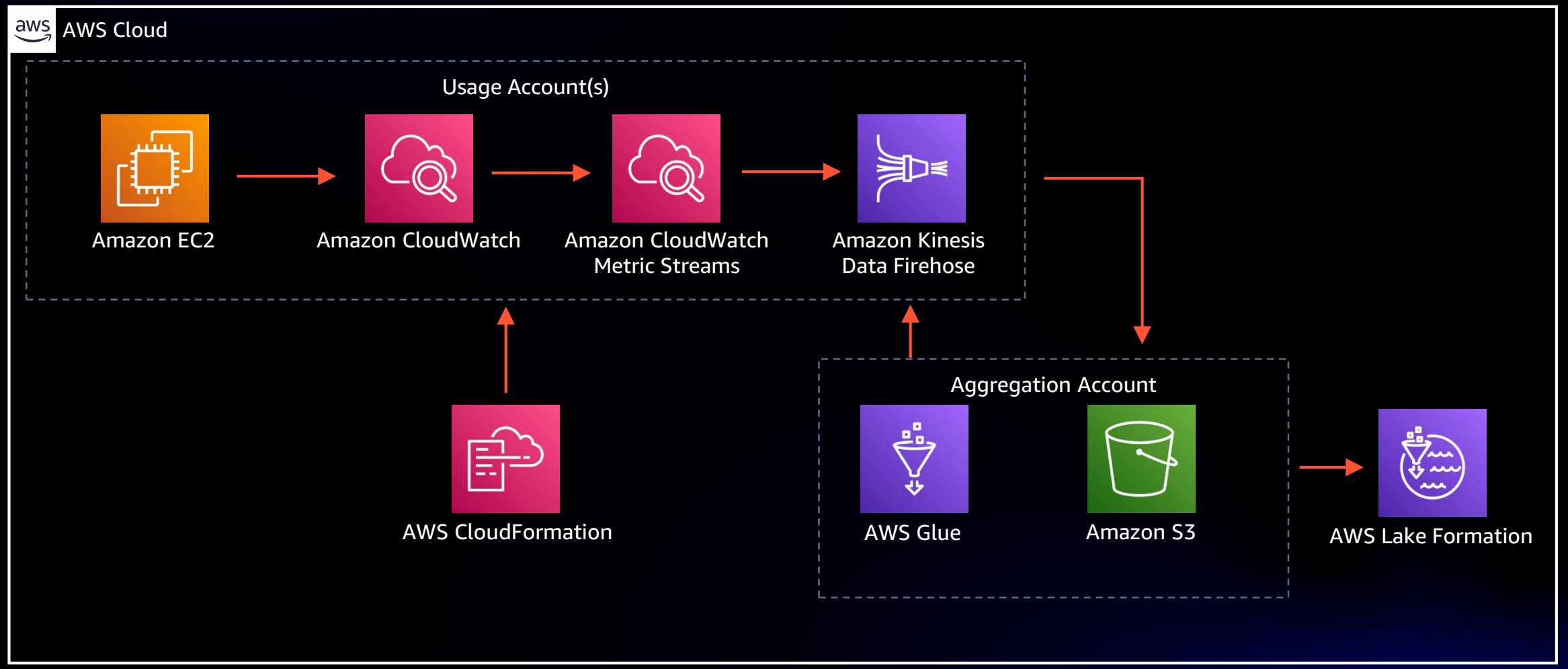
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Compute Optimizer



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Collecting Amazon CloudWatch metrics – Amazon EC2



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Operationalizing CloudWatch data

How can I use CloudWatch data to identify quick wins?

- **Example**: Derive an underutilized compute ratio (UCR) for EC2
- Set a CPU utilization threshold for underutilized classification
- Create a ratio of underutilized / total
- Group by metadata, tags, and allocation details
- Identify top areas to change consumption patterns

$$UCR = \frac{SUM(\text{data points under threshold})}{SUM(\text{total data points})}$$

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Functional efficiency

Fine-tuning your operational environment

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Functional efficiency

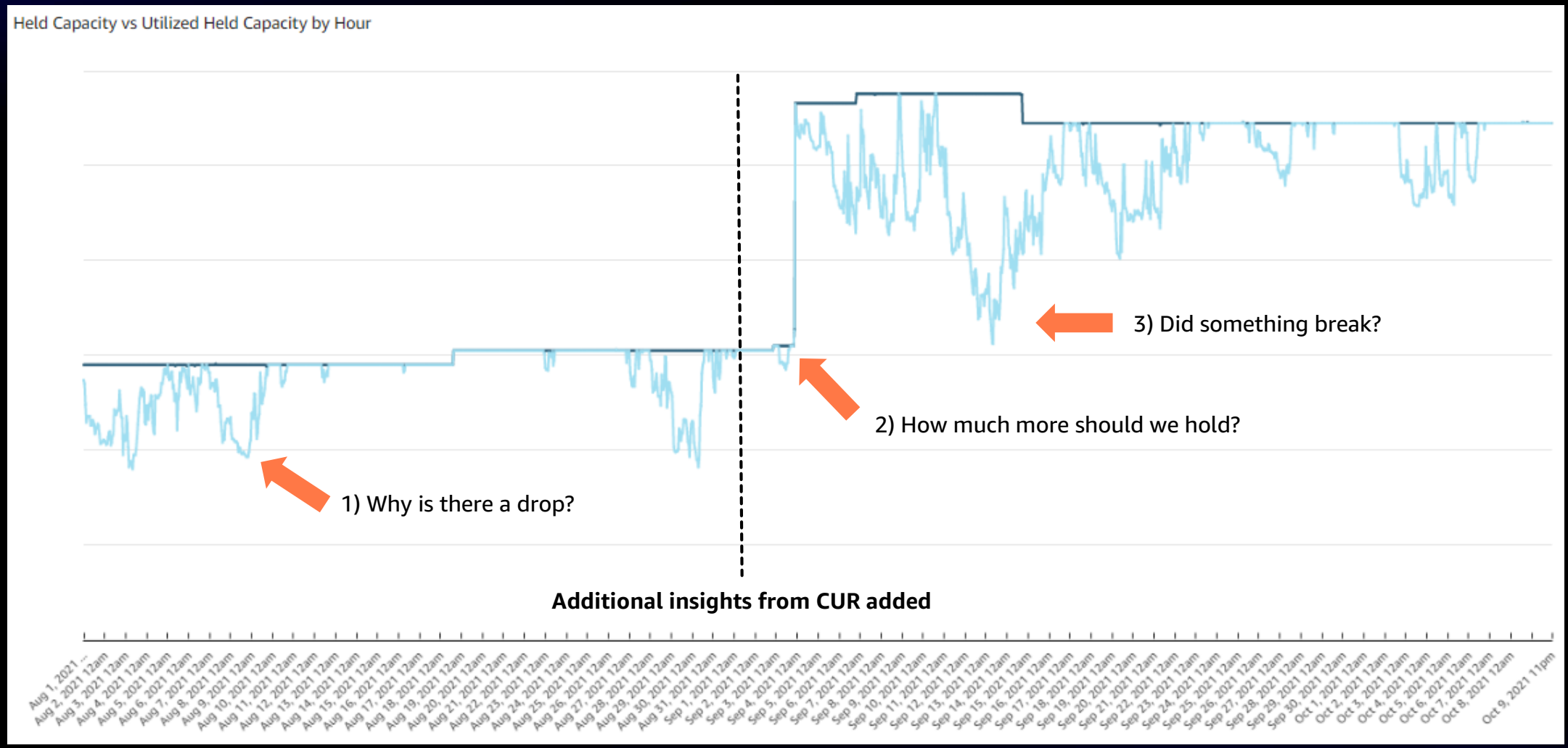
How efficient am I? Is it good enough? What does good mean?

- Custom tags and metadata
- Custom metrics and alarms
- Establish functionally specific baselines
- Evaluate capacity impacts on efficiency
- Benchmark, iterate, improve



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

CloudWatch ODCR Utilization

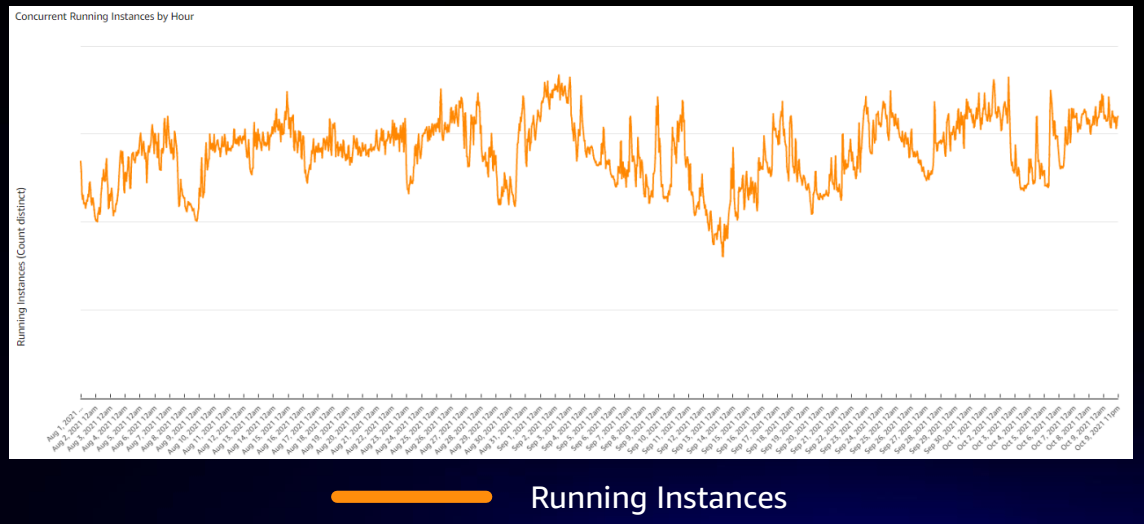


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Extending the Cost & Usage Report (CUR)

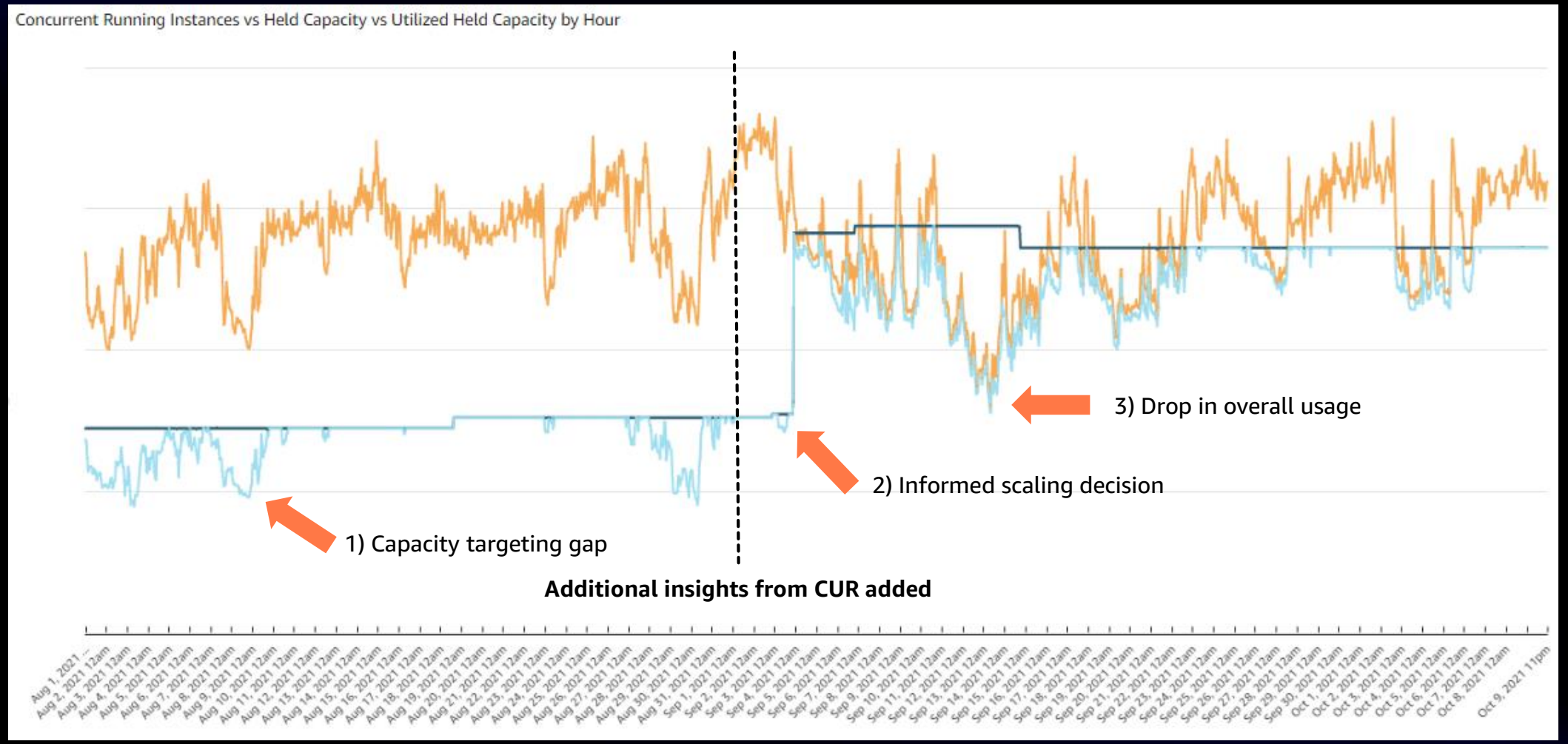
Objective: Hourly time series visibility into concurrent EC2 footprint

- For each resource key use MIN(start_date) and MAX(end_date)
- Group by CUR metadata
- Cross-join to all dates/hours in date range (instance lifecycle)
- Note edge cases such as hibernation
- Sum distinct resource keys per hour



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Extended CUR and CloudWatch



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Putting it all together

Infrastructure ROI observability

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Putting it all together

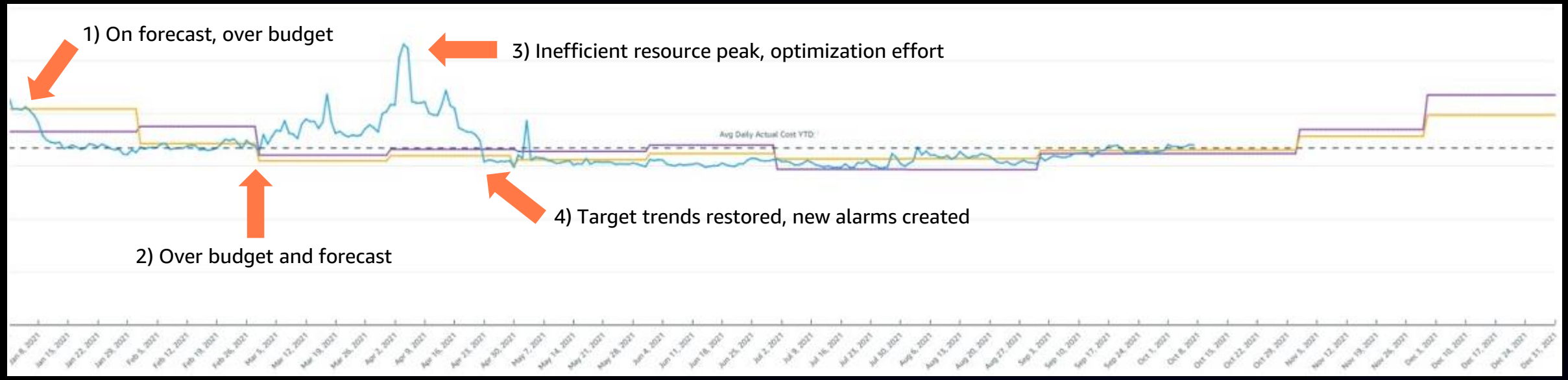
Is cost attribution, managing underutilized resources, and functional efficiency delivering the overall intended value to my customers?

- Combining CUR, CloudWatch, and financials
- Budget and SLA performance
- Business health metrics, trends, and threshold alarms
- Reassess and raise the bar

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Putting it all together

Example: Combining trends for actual usage/cost, budget, forecast and then use CloudWatch derived insights to drive efficiency initiatives.



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Recent release

Custom Dashboard Widgets

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Recap and Call to Action

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.