
AWS
re:Invent
NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

ARC201

Reliable scalability: How Amazon.com scales in the cloud

Seth Eliot
Principal Reliability Solutions Architect, AWS Well-Architected
AWS

Kieran Kavanagh
Principal AI/ML Solutions Architect, Strategic Accounts
AWS

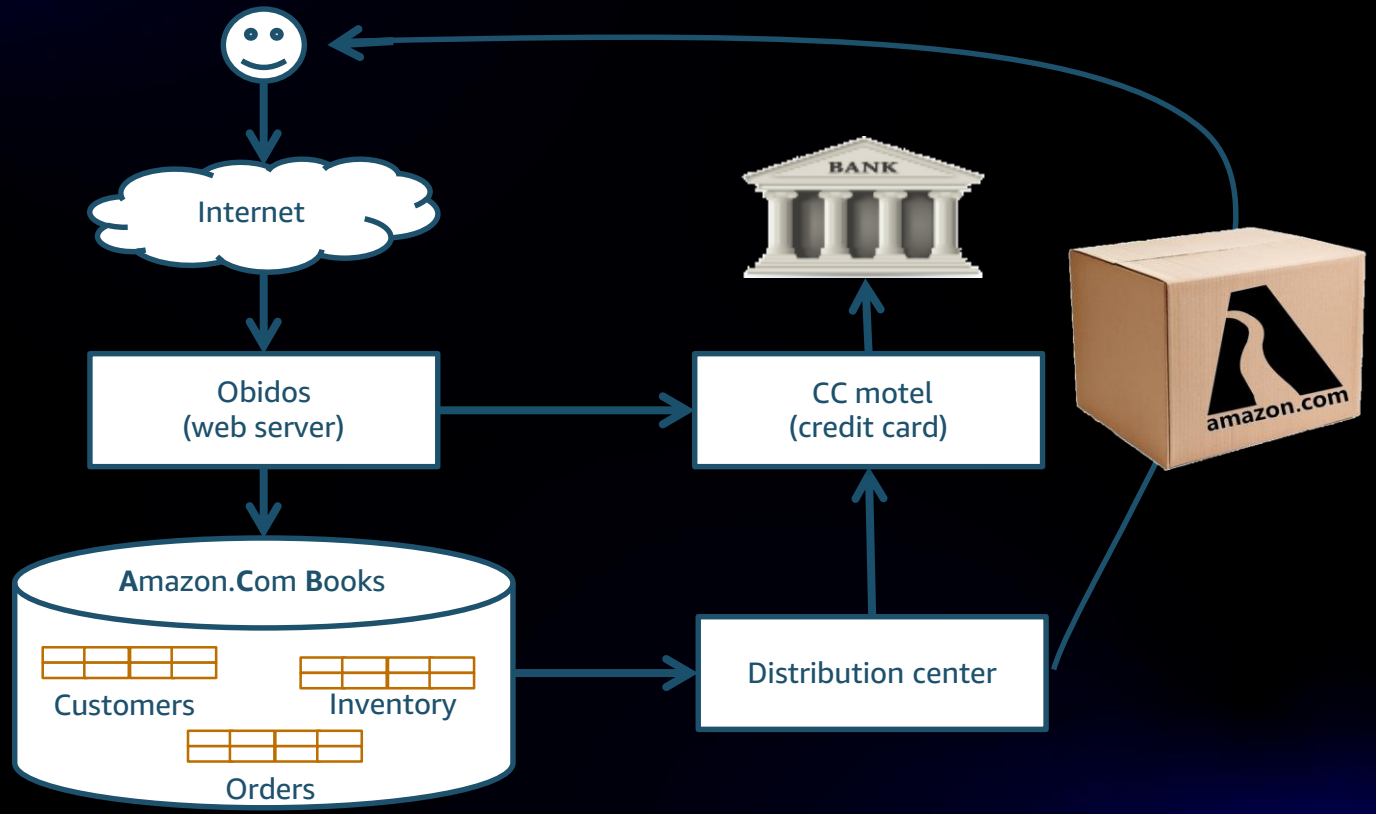
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Welcome to Amazon.com Books! – 1995



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

Architecture 1995



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

Amazon needs scalability



Scalability

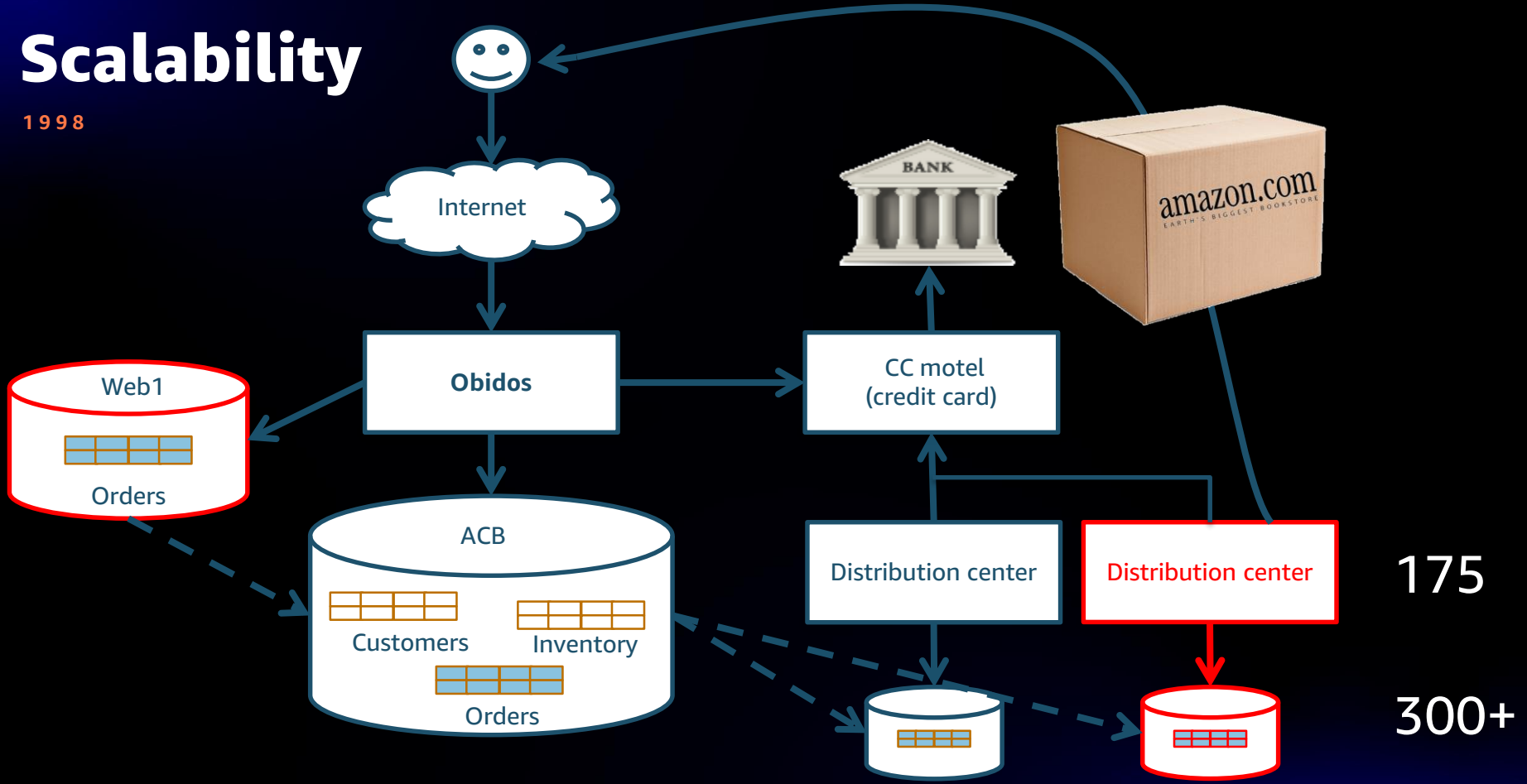
The ability of a workload to perform its agreed function when the load or scope changes

aws

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

Scalability

1998



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

Scalability

SOA – 2000









Internet

Obidos

CC motel

Distribution center

Web1

Orders

Customer service

ACB

Orders

Inventory

CustNA

Customers

aws

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

Amazon got big

AMAZON PRIME DAY 2021 SALES STATISTICS



Number of items purchased by Prime members during Prime Day

175 million

Global Amazon Prime Day sales

$11.2 billion

Source: https://www.statista.com/topics/5385/amazon-prime-day/

aws

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

AWS statistics for Prime Day 2021





89.2 million requests/s

Amazon DynamoDB



47.7 million requests/s

Amazon SQS



11.1 trillion requests/day

Amazon EBS



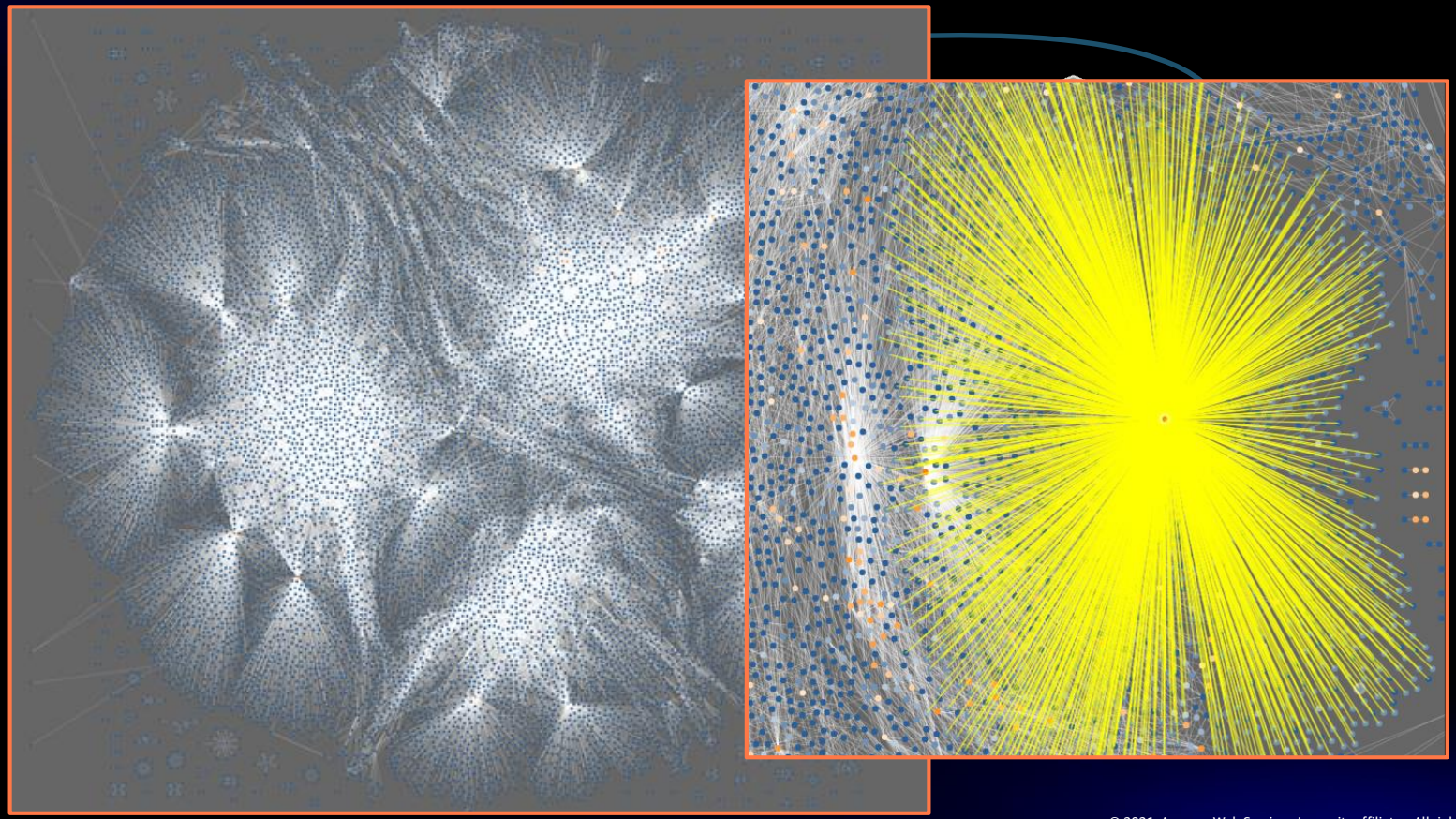
>290 million requests/min
600 billion requests total

Amazon CloudFront

Source: Jeff Barr, VP and Chief AWS Evangelist

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

Architecture 2021



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

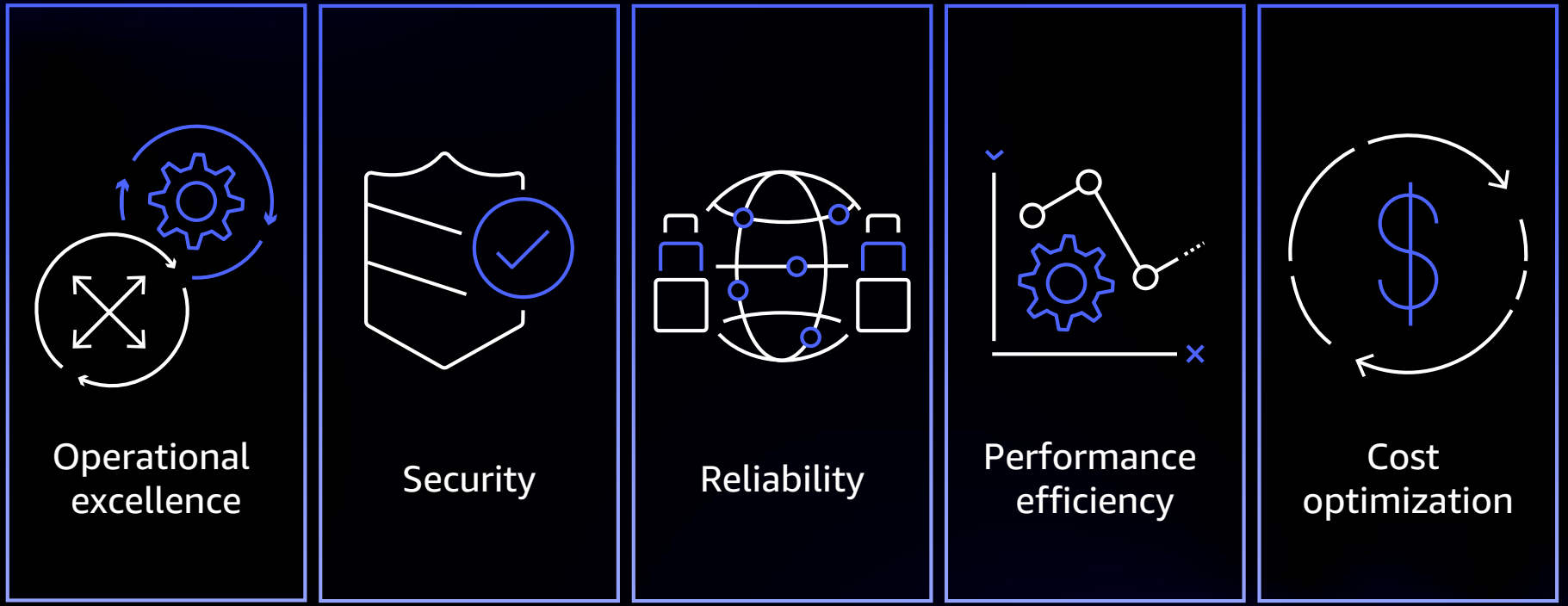
Reliable scalability

Reliability

The ability of a workload to perform its required function correctly and consistently

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

AWS Well-Architected



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

Service-oriented and microservices architectures

Bias for action

Speed matters in business
We value calculated risk taking

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

Architecture evolution

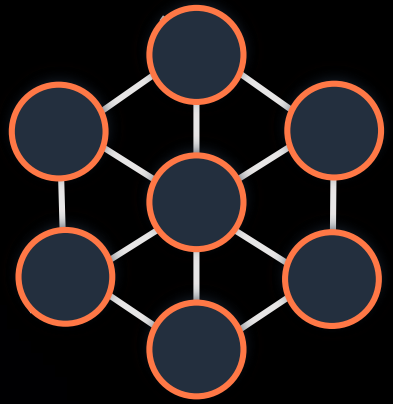
WHEN THE IMPACT OF CHANGE IS SMALL, RELEASE VELOCITY CAN INCREASE



Monolithic application
Does everything
High impact of change



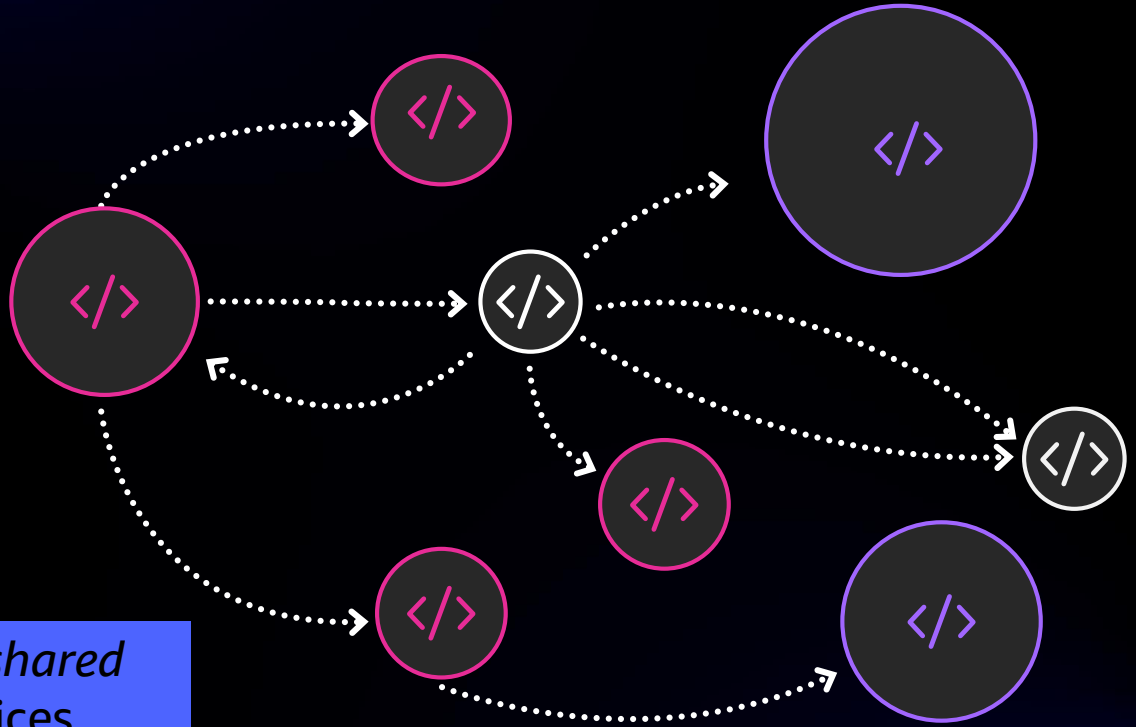
Service-oriented
Does some things
Smaller impact of change



Microservices
Does one thing
Smallest impact of change

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

Microservices deploy and scale independently



Capacity is *shared* among services

aws

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

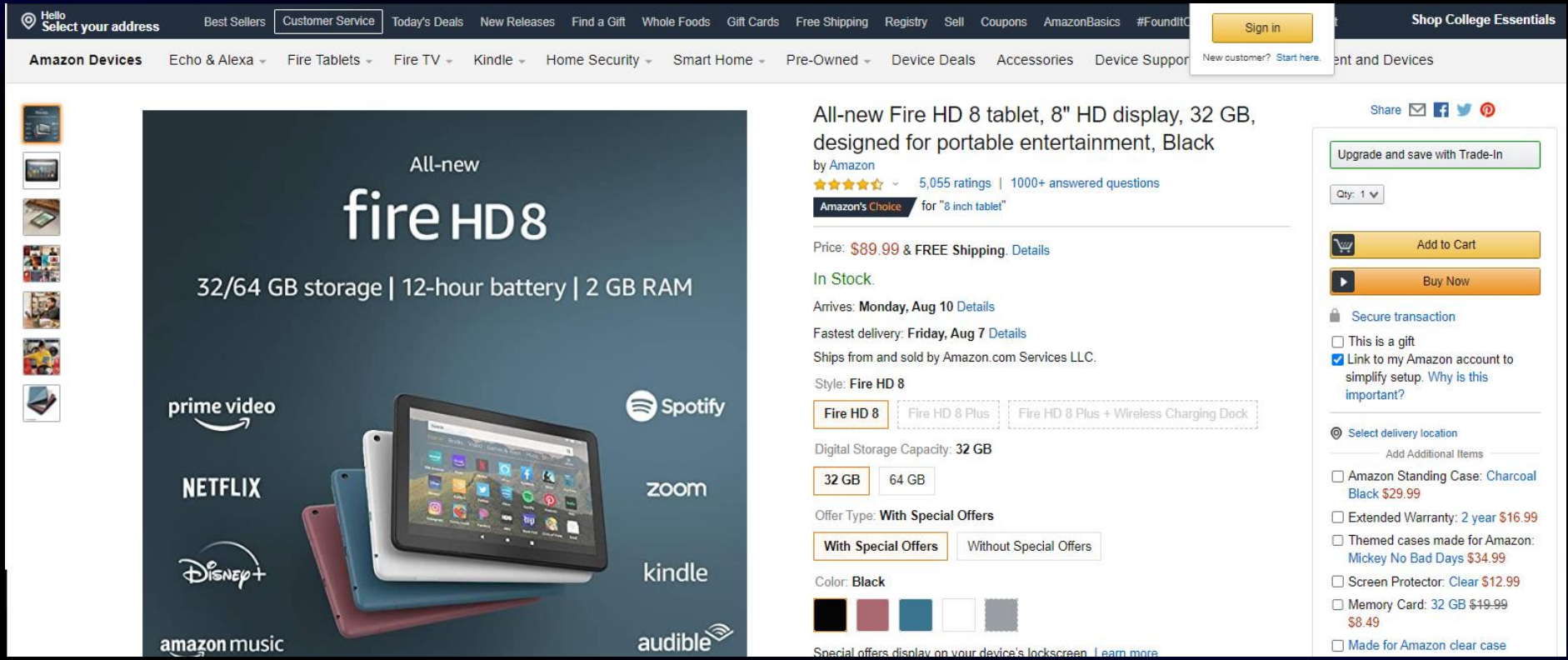
Deployment velocity

150 million deployments per year



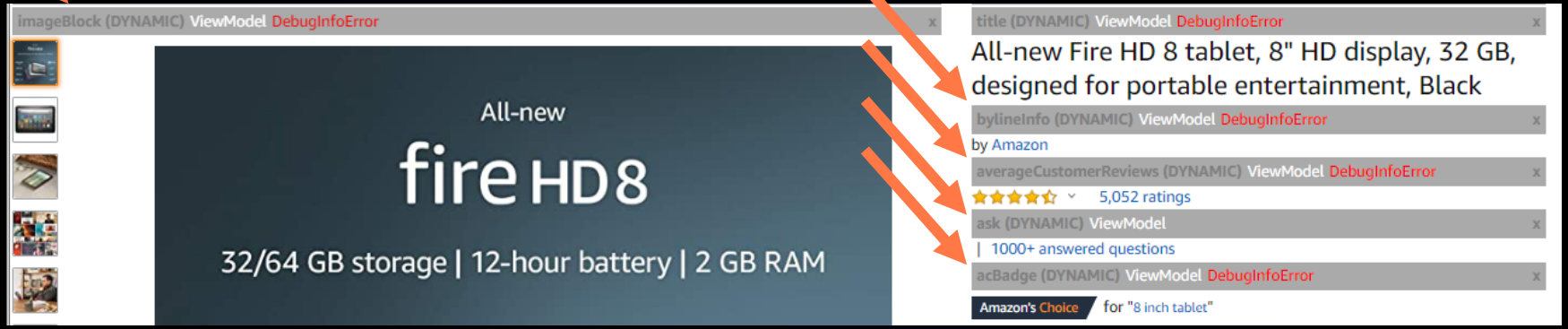
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon “detail page”



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

Amazon “detail page” has 100s of service calls



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

SUBSCRIPTION BOXES | BEAUTY | FOOD & DRINK | LIFESTYLE | PETS | TOYS & KIDS | T-SHIRT CLUBS | MANAGE | SELL

Amazon Subscription Boxes: Toys & Kids

Shop the top toy subscription boxes for kids and adults









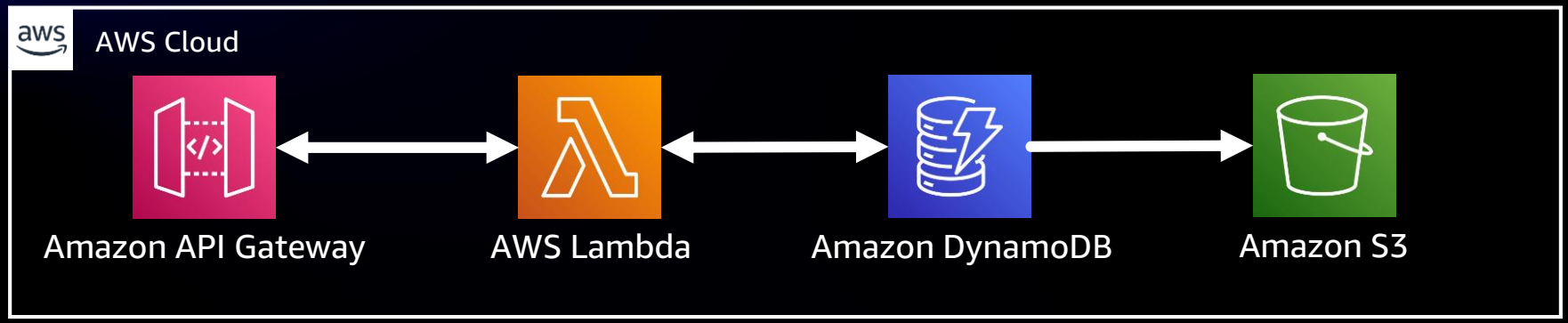




aws

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

Amazon subscription boxes



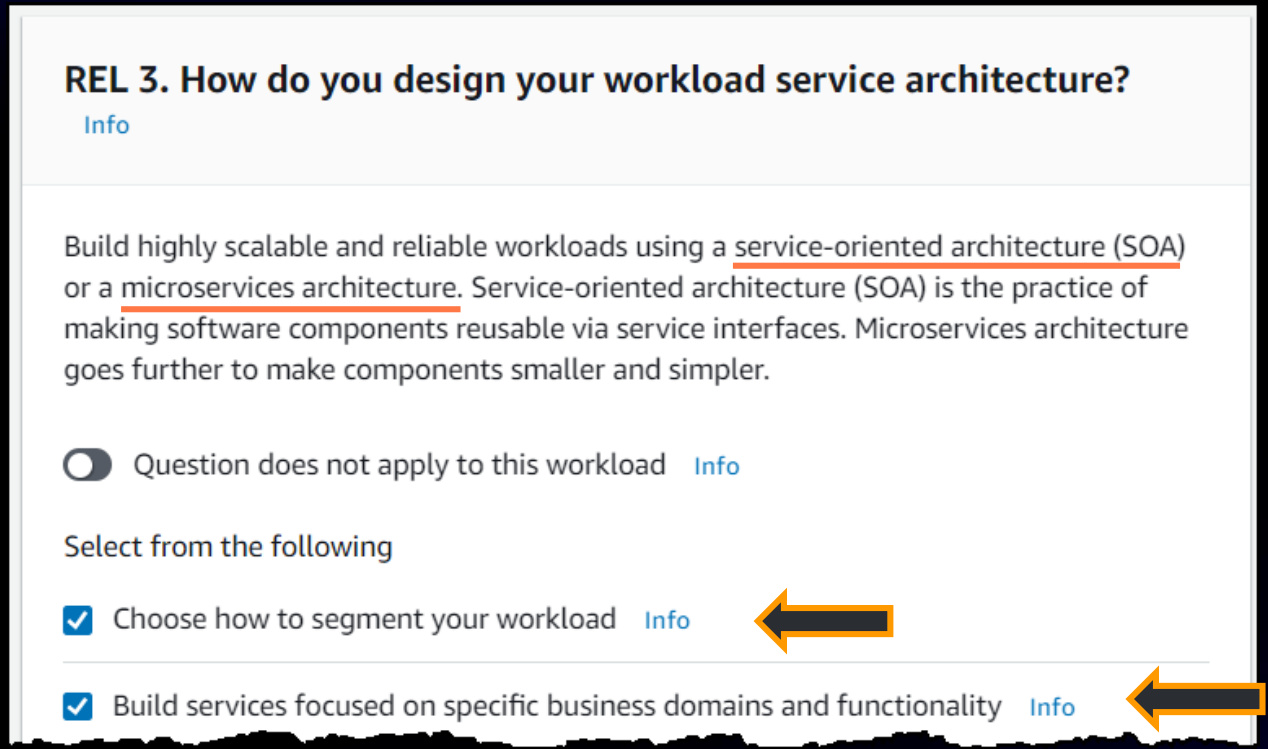
Chooses items in the subscription box for customer

"Nanoservice"

- Built and maintained by 3 engineers
- Handles millions of events/day
- Delegates to other systems: Fulfillment system

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

Reliability pillar – Service architecture



AWS Well-Architected

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

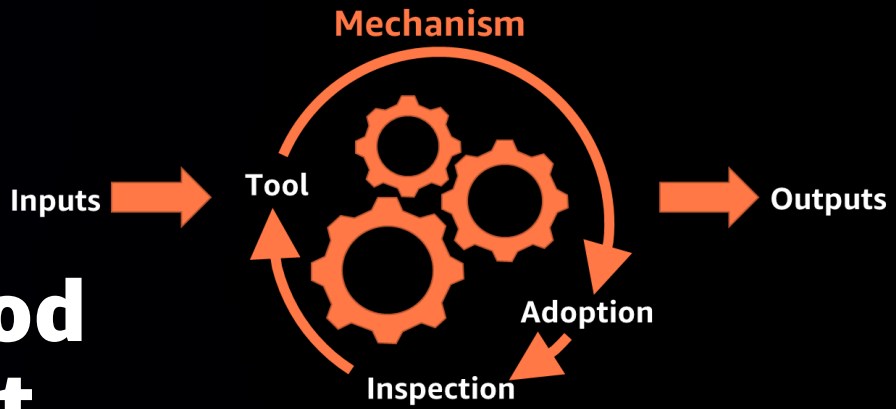
Review and test

Insist on the highest standards

Leaders are continually raising the bar and driving their teams to deliver high-quality products

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

Mechanisms



"When you ask for good intentions, you're not asking for a change, because people already had good intentions."

Jeff Bezos

Founder and Executive Chair, Amazon.com

aws

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

Code review

- All code and configuration destined for production
- If no consensus, then need a second reviewer
- It’s iterative – CR changes must also be reviewed



“By approving a change, the code reviewer is putting their stamp of approval on the code, and also shares in the accountability of the quality of the change.”

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

Principal engineers

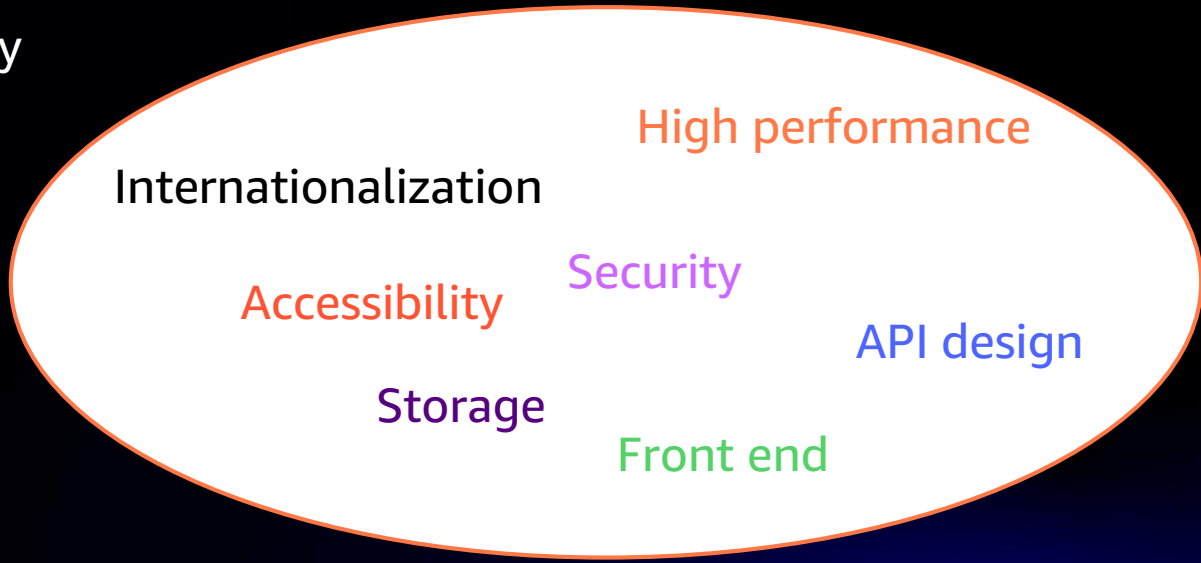
Principal, senior principal, and distinguished engineers across Amazon

- Exemplary practitioner
- Learn, educate, and clarify

Principal consult

Principal design review



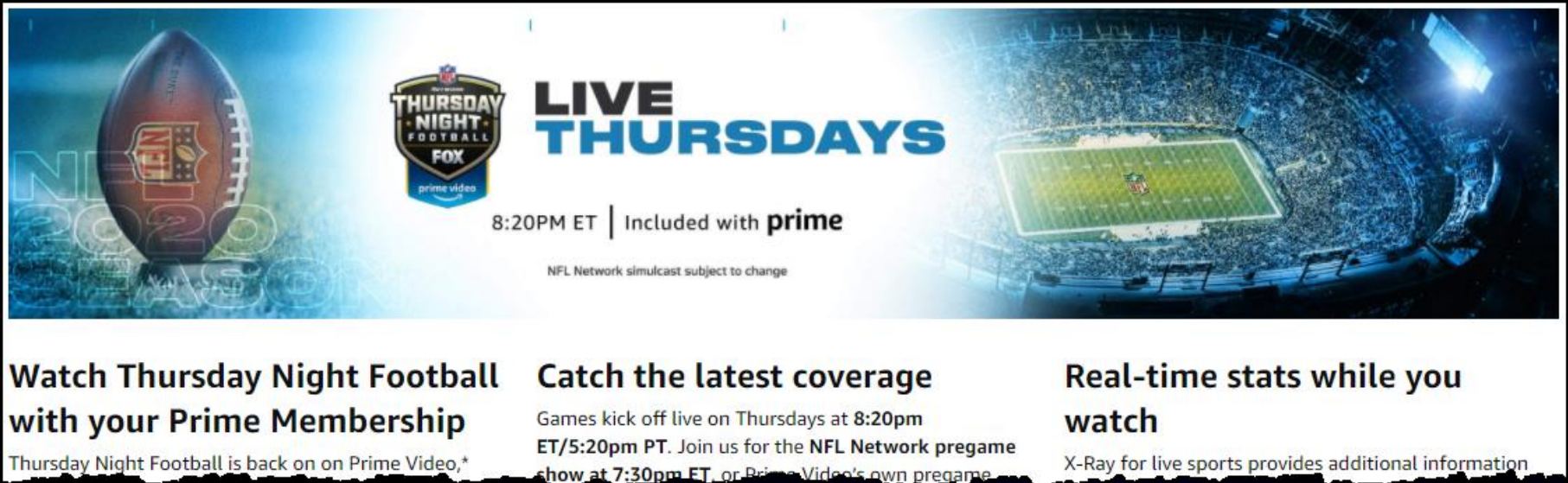


aws

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

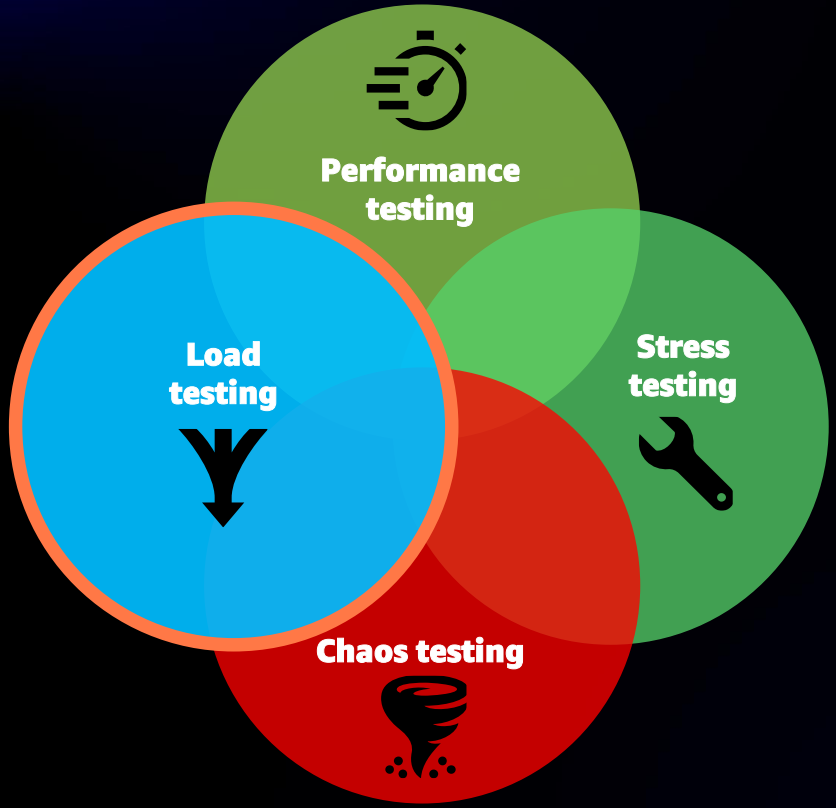
Prime Video game day testing

Prepare for Thursday Night Football . . . and beyond



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

Types of tests performed



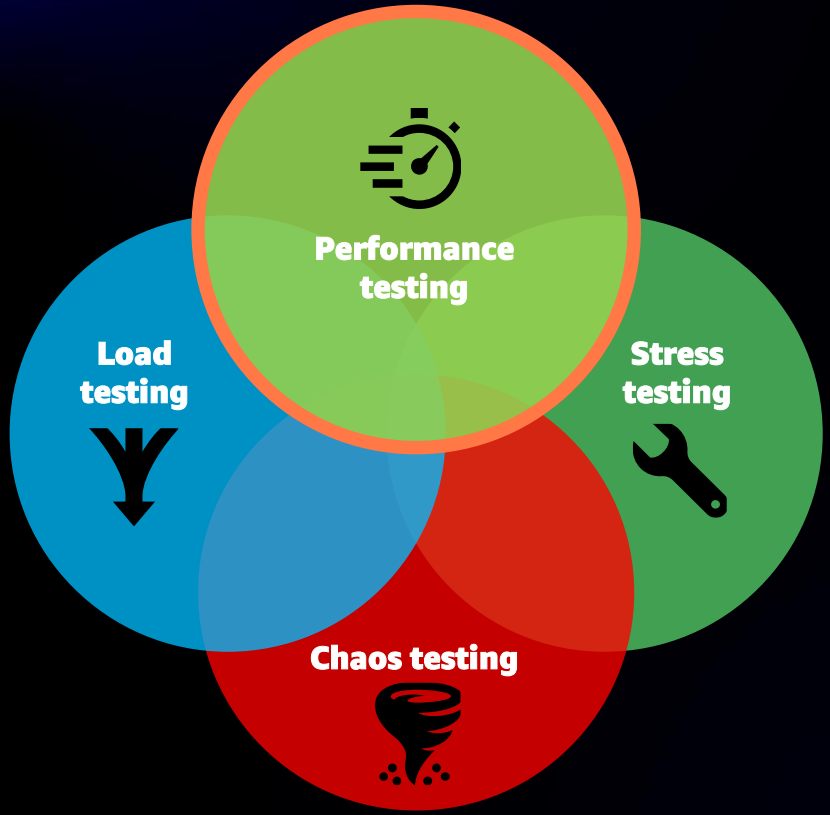
Can our services handle an expected realistic **load**?

- ✓ Mostly positive testing
- ✓ Emphasis on prod behavior

aws

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

Types of tests performed



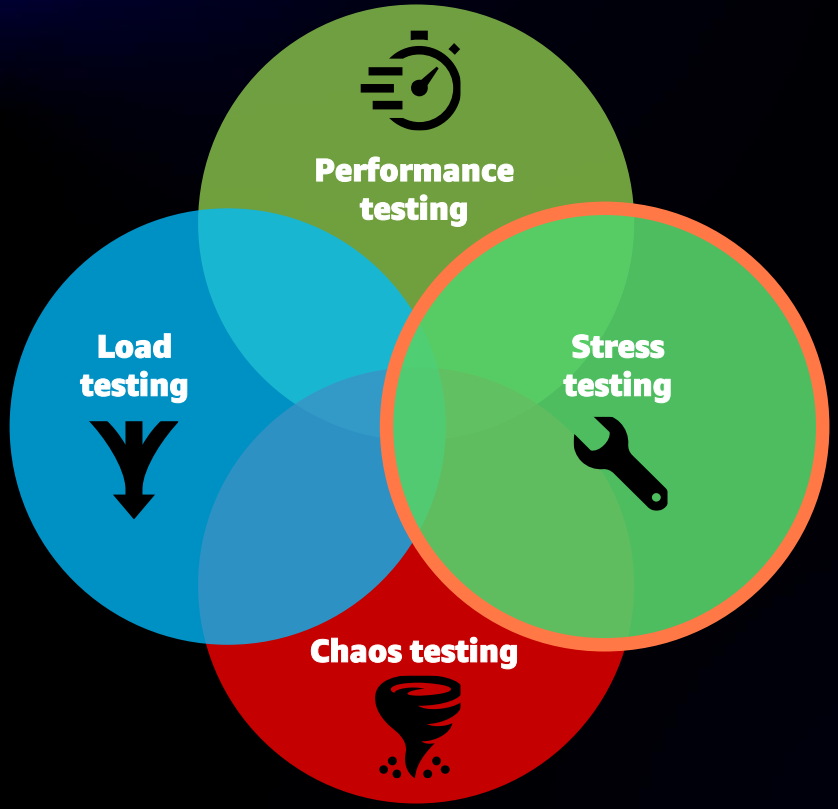
How does load change **performance**?

- ✓Latency
- ✓Errors
- ✓Throughput

aws

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

Types of tests performed



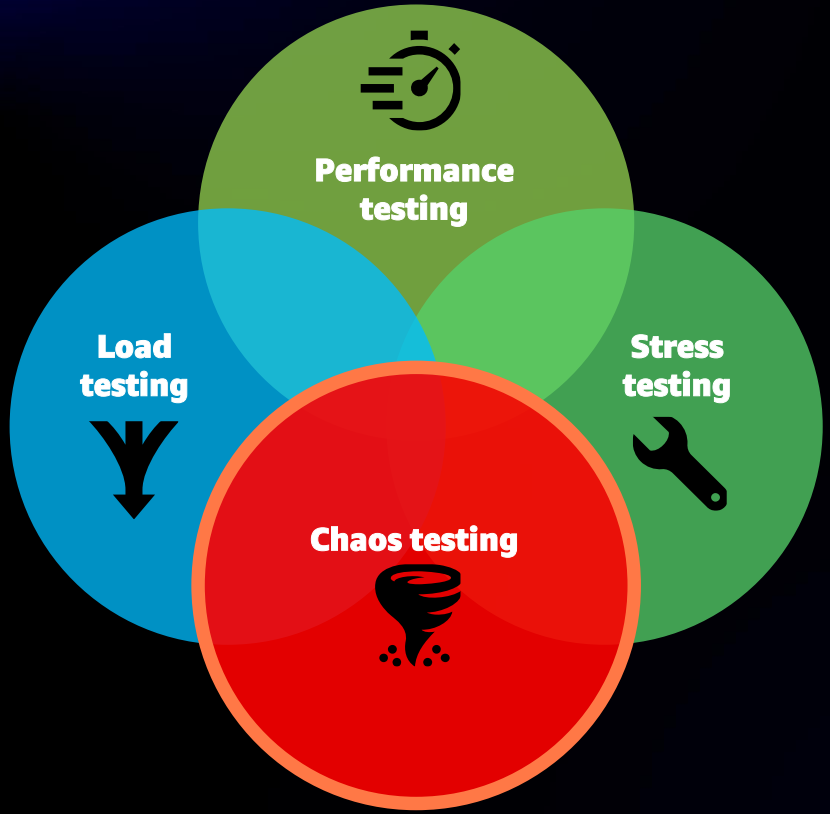
When and where is a **breaking point visible**?

- ✓ Emphasis on breaking the system
- ✓ Not on prod-like behavior

aws

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

Types of tests performed



How will services handle underlying **failures**?

- ✓ Network latency
- ✓ CPU or memory maxed out
- ✓ Dependencies failing

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

Reliability pillar – Test reliability

REL 12. How do you test reliability? Info

After you have designed your workload to be resilient to the stresses of production, testing is the only way to ensure that it will operate as designed, and deliver the resiliency you expect.

Question does not apply to this workload Info

Select from the following

- [x] Use playbooks to investigate failures Info
- [x] Perform post-incident analysis Info
- [x] Test functional requirements Info
- [x] Test scaling and performance requirements Info
- [x] Test resiliency using chaos engineering Info
- [x] Conduct game days regularly Info

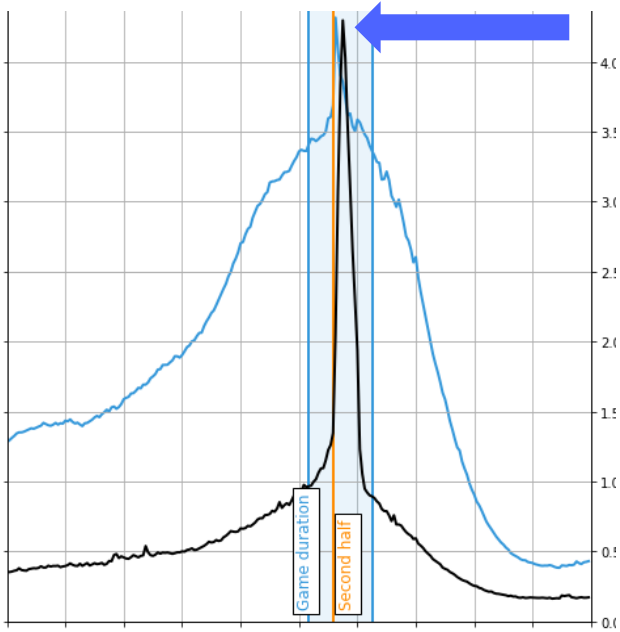


AWS
Well-Architected

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

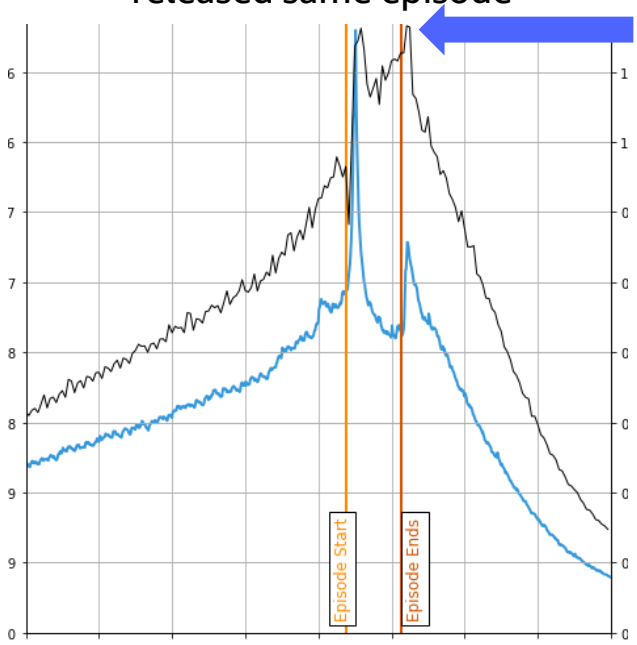
Real users can be unpredictable

The match gets interesting . . . subscribing in the second half



Stream starts
Sign-ups

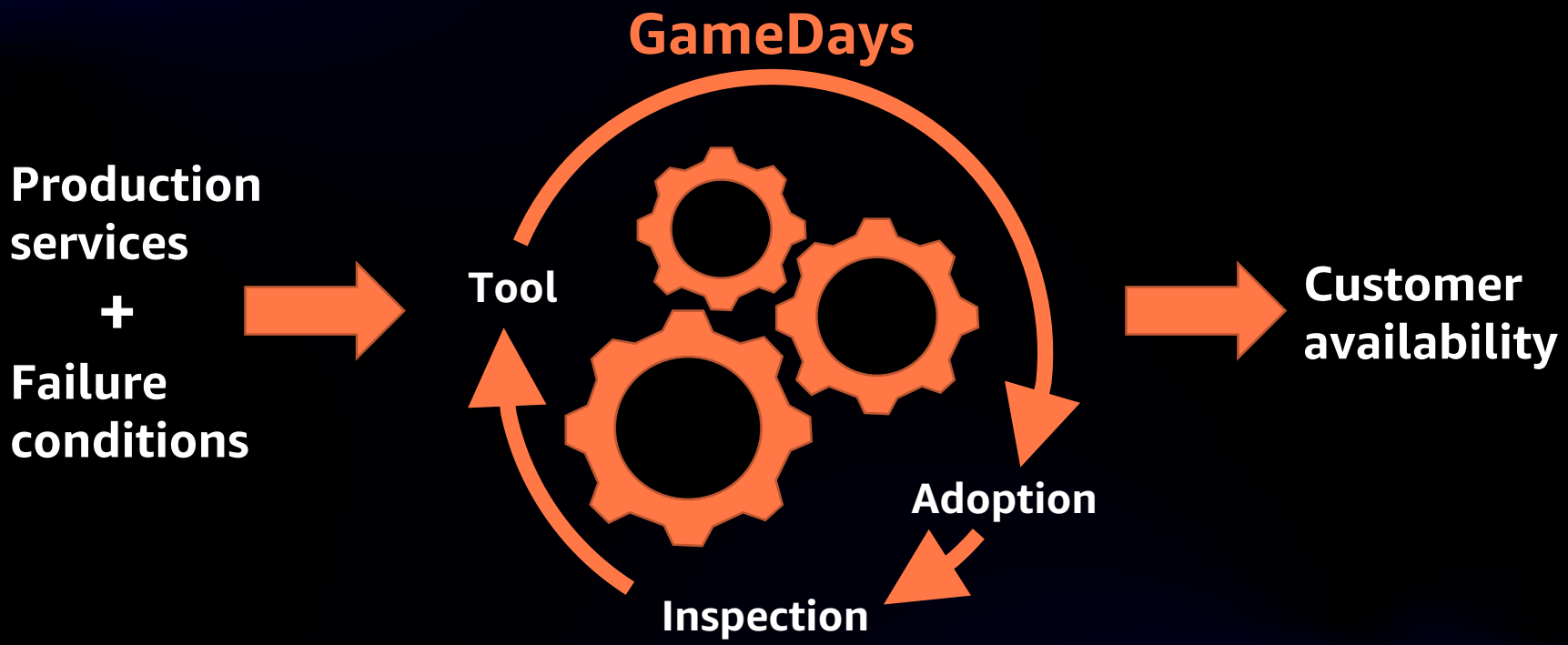
Customers rewatching a newly released same episode



Stream starts
Concurrent streams

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

Resilience validation mechanism

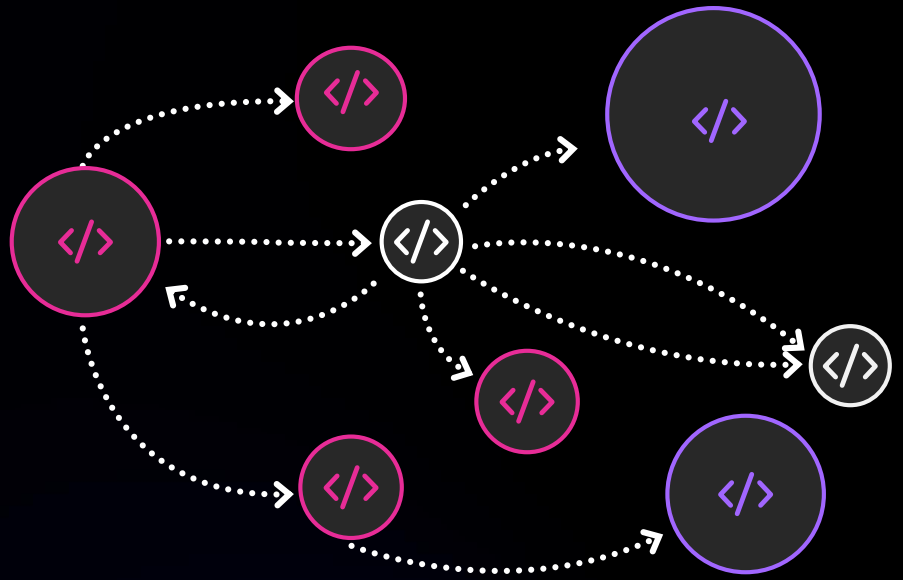


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Dependencies

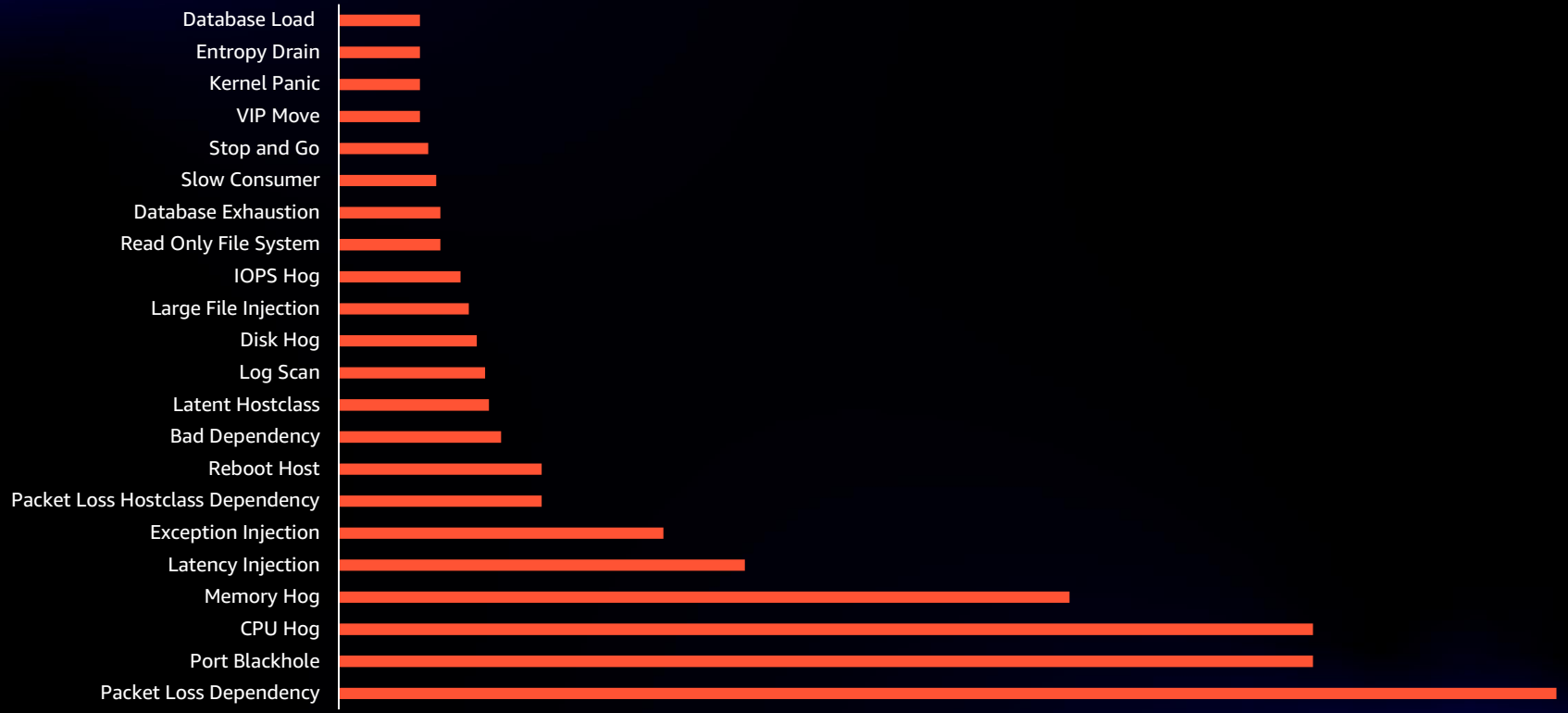
Microservices architecture

- What happens when your dependencies fail?
- How do you know?



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

Failure injection tests used by Prime Video

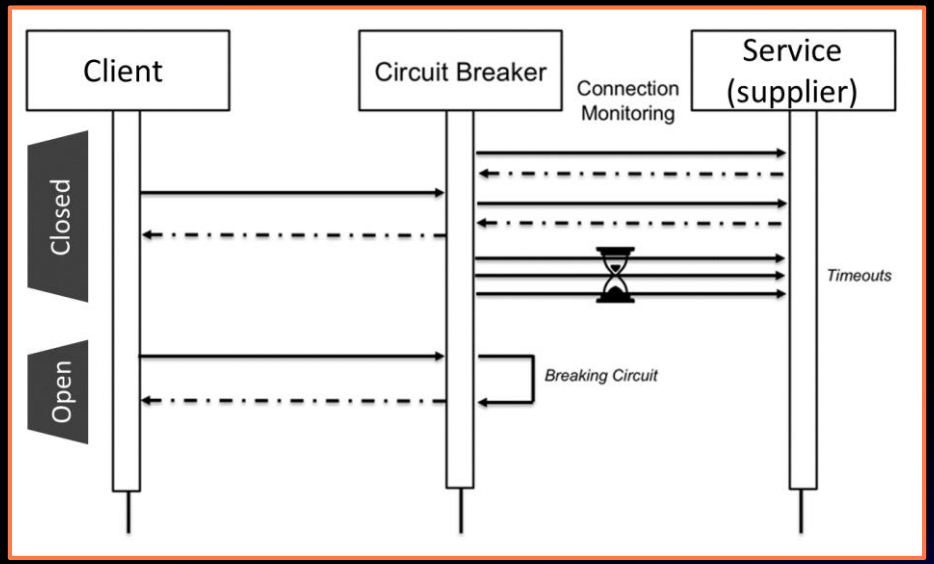


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Failure injection test (FIT) use case

Problem: Downstream dependency throttling caused availability drop

Solution: Implemented fast-fail circuit-breaker



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

High-velocity events

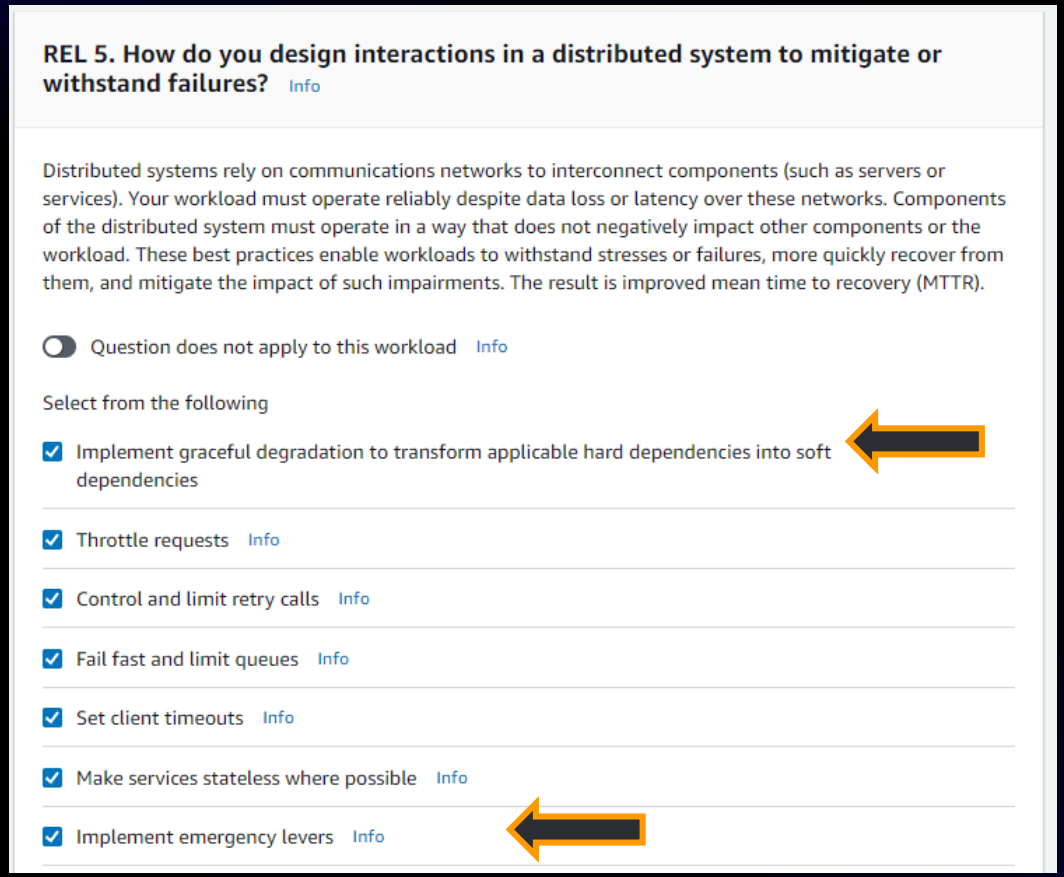
Protective actions

- Contingencies and mitigation levers
- Prioritization of critical workloads
- Capabilities to shield downstreams
- Control by title or customer group



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

Reliability pillar – Mitigate failures



REL 5. How do you design interactions in a distributed system to mitigate or withstand failures? Info

Distributed systems rely on communications networks to interconnect components (such as servers or services). Your workload must operate reliably despite data loss or latency over these networks. Components of the distributed system must operate in a way that does not negatively impact other components or the workload. These best practices enable workloads to withstand stresses or failures, more quickly recover from them, and mitigate the impact of such impairments. The result is improved mean time to recovery (MTTR).

Question does not apply to this workload Info

Select from the following

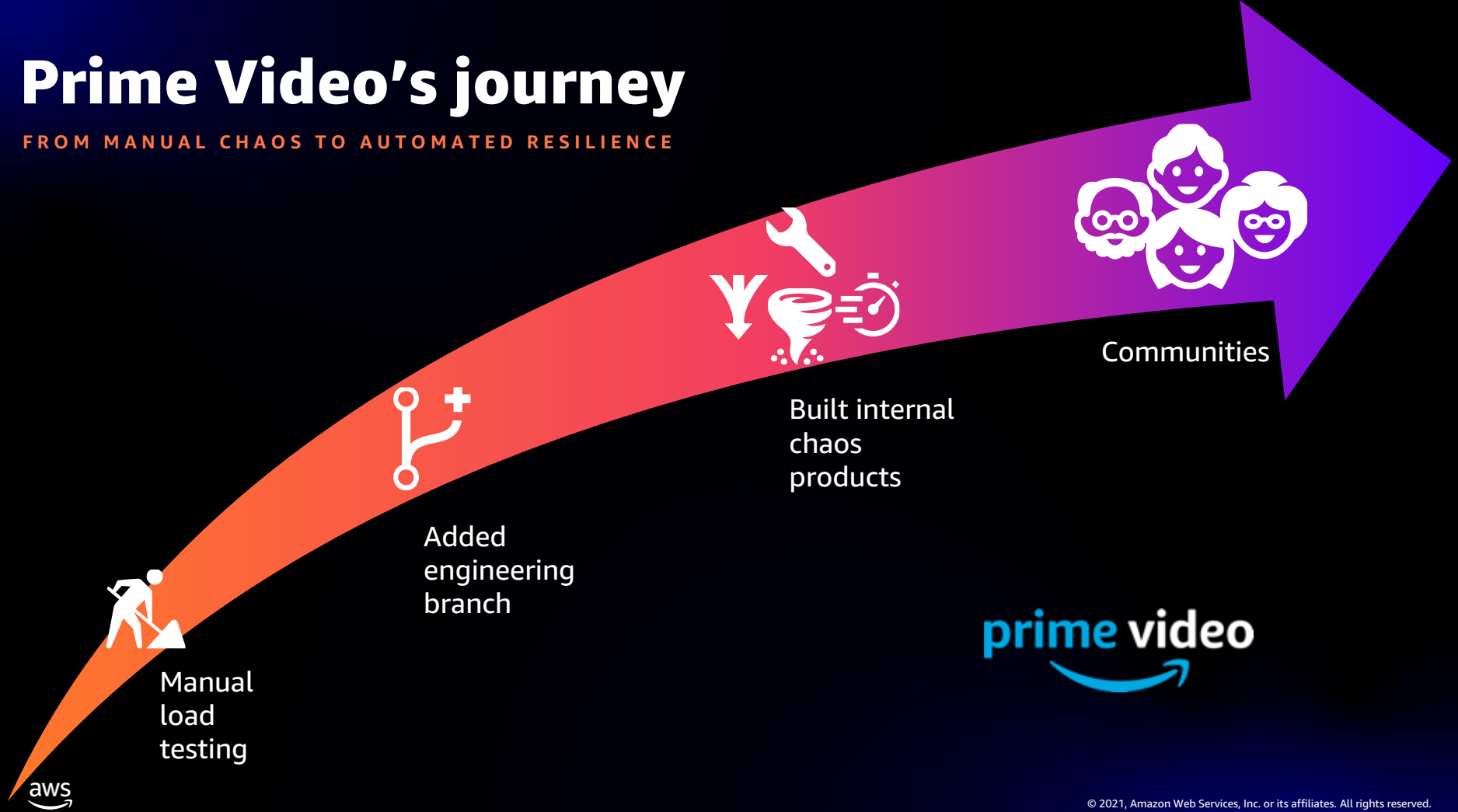
- [x] Implement graceful degradation to transform applicable hard dependencies into soft dependencies
- [x] Throttle requests Info
- [x] Control and limit retry calls Info
- [x] Fail fast and limit queues Info
- [x] Set client timeouts Info
- [x] Make services stateless where possible Info
- [x] Implement emergency levers Info

AWS Well-Architected

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

Prime Video's journey

FROM MANUAL CHAOS TO AUTOMATED RESILIENCE



prime video

aws

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

Understanding services in production

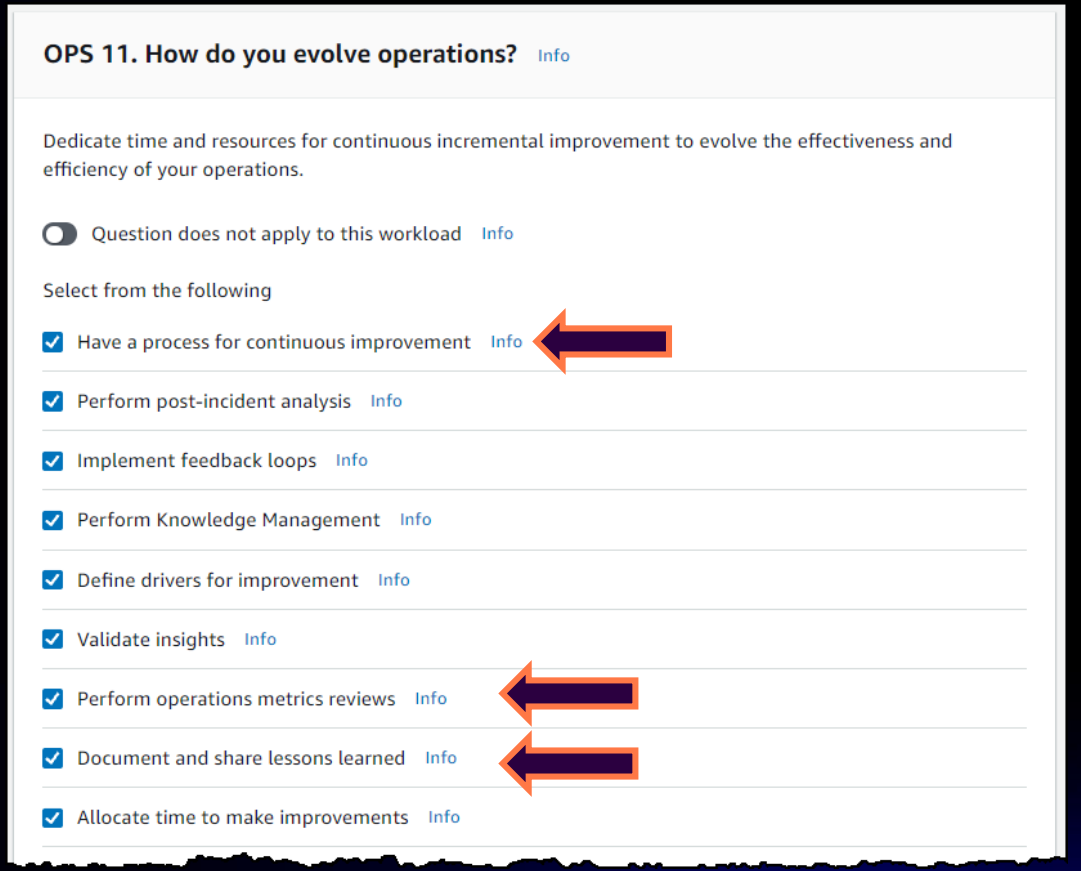
Dive deep

Leaders operate at all levels, stay connected to the details, and audit frequently

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

Weekly ops review

- Success stories
- Metrics
- What can improve?



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

The wheel

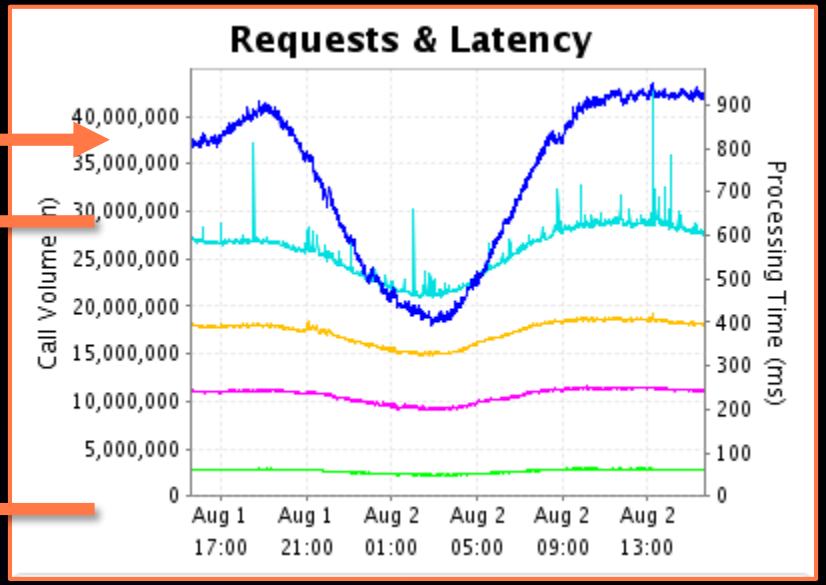
"Winner" gets to share metrics, success stories, and challenges with senior leadership



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

Every service has a monitoring dashboard

- Number of requests
- Latency – p50, p90, p99, and p99.9
- Availability (9s)
- And more



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

Monitoring for Amazon

UNDERSTANDING IMPACT ON CUSTOMERS

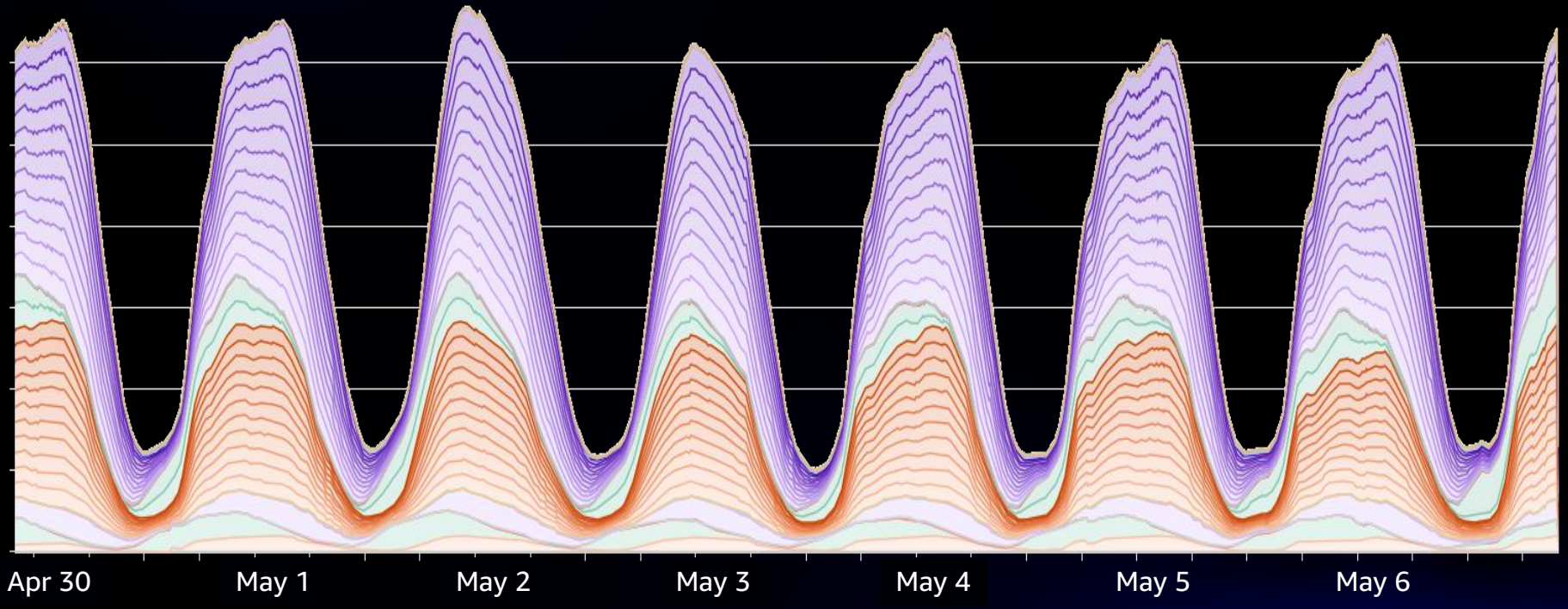
Alarm if orders drop versus predicted



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

Ring weekly load pattern

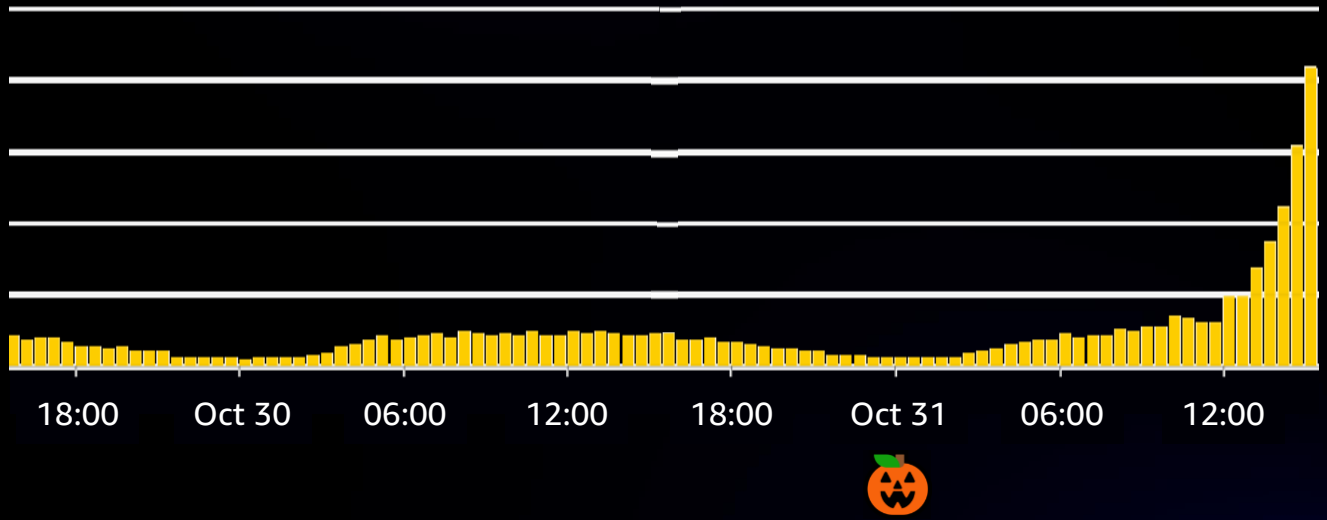
VIDEO CAPTURE AND REQUEST FOR PROCESSING



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

Ring's "Prime Day": Halloween!

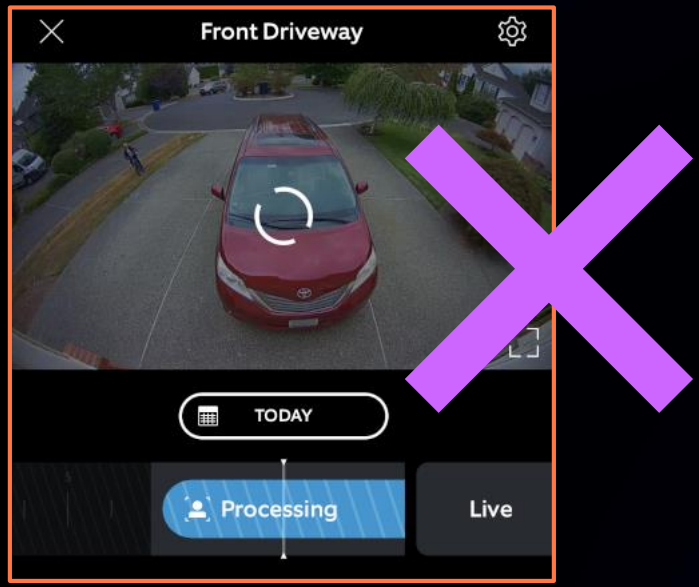
PREPARING FOR HIGH DEMAND



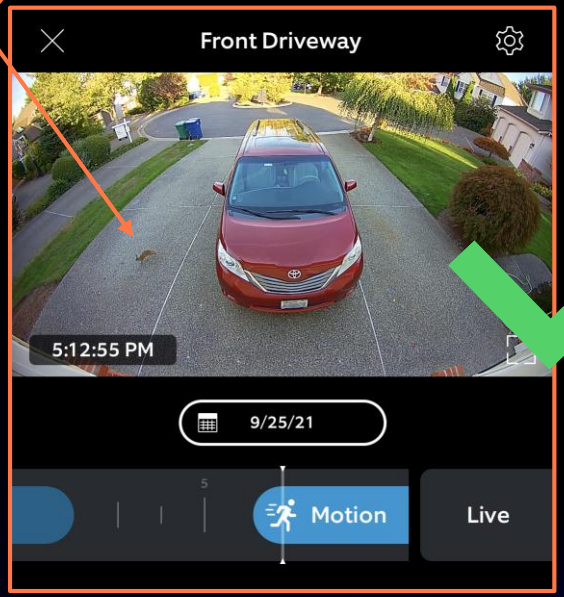


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Nearly instant video processing

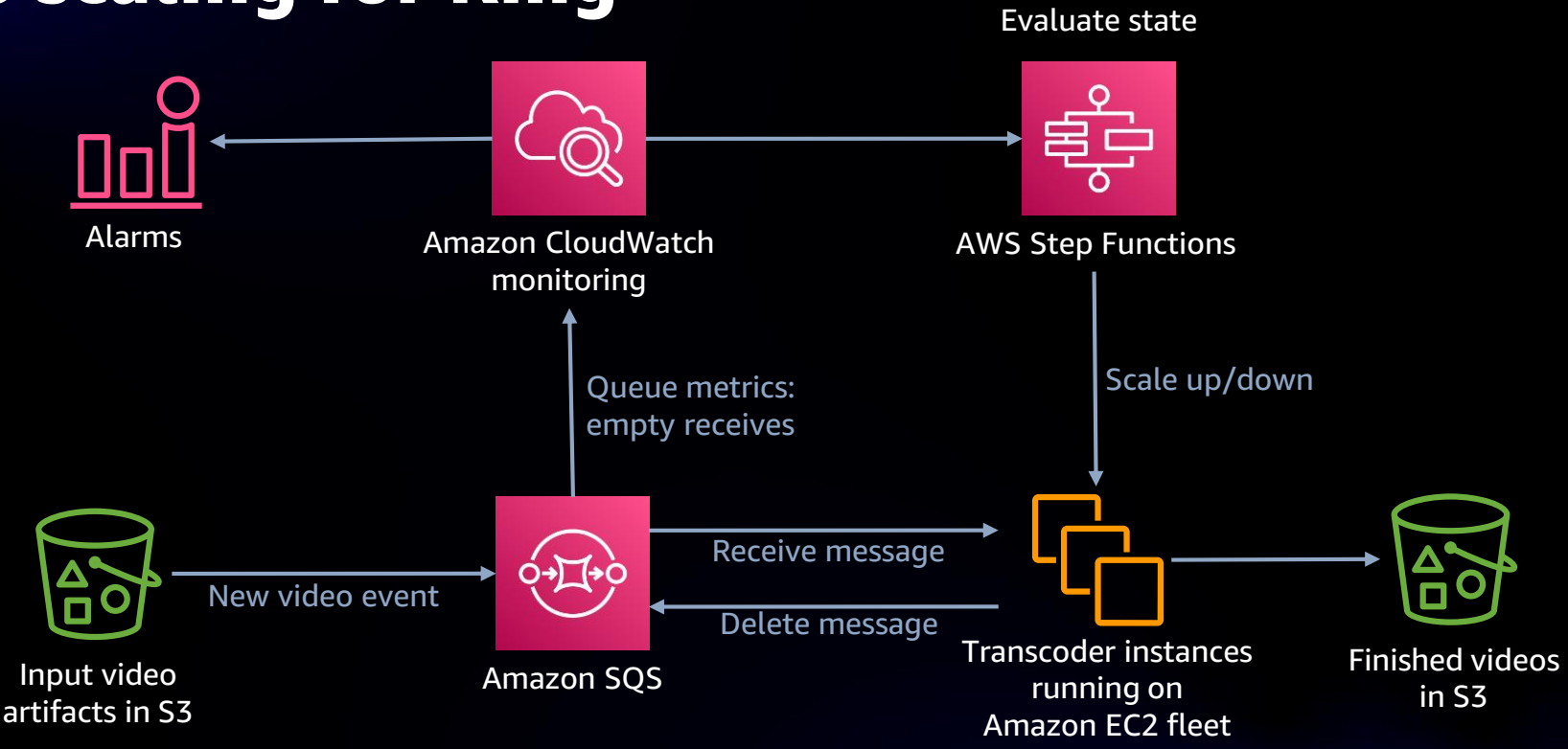






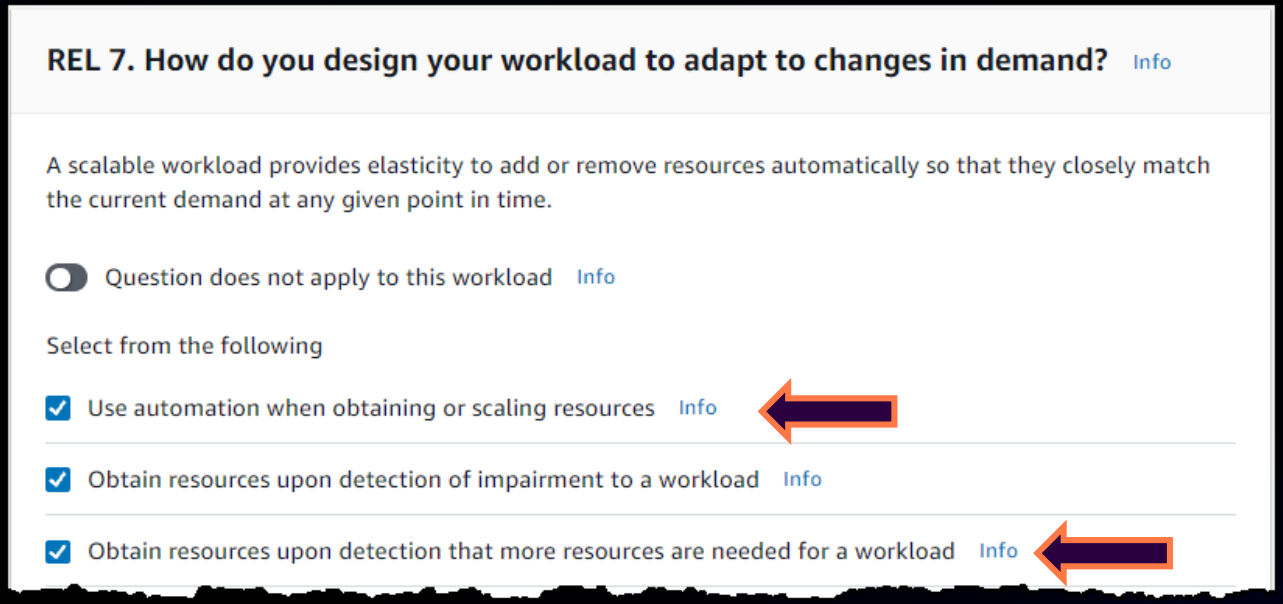
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Auto scaling for Ring



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

Reliability pillar – Adapt to changes in demand



AWS Well-Architected

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

Going global

Customer obsession

Leaders start with the customer and work backwards. They work vigorously to earn and keep customer trust.

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

AWS Regions and Availability Zones (AZs)

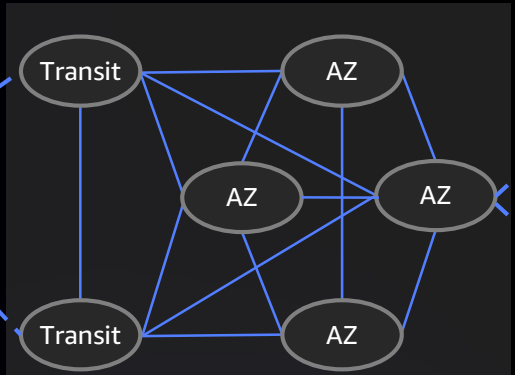
AWS REGIONS ARE PHYSICAL LOCATIONS AROUND THE WORLD WHERE WE CLUSTER DATA CENTERS

25 AWS Regions worldwide



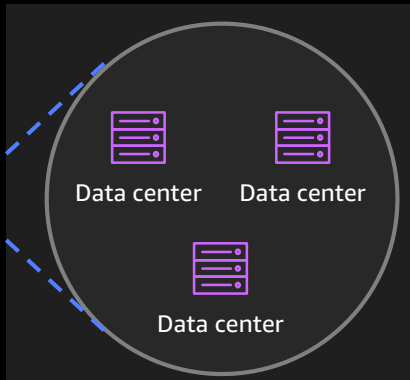
- AWS Regions
- Announced Regions

Each AWS Region has multiple AZs



A Region is a physical location in the world

Each AZ is one or more discrete data centers



Data centers, each with redundant power, networking, and connectivity, are housed in separate facilities

aws

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

Globalization

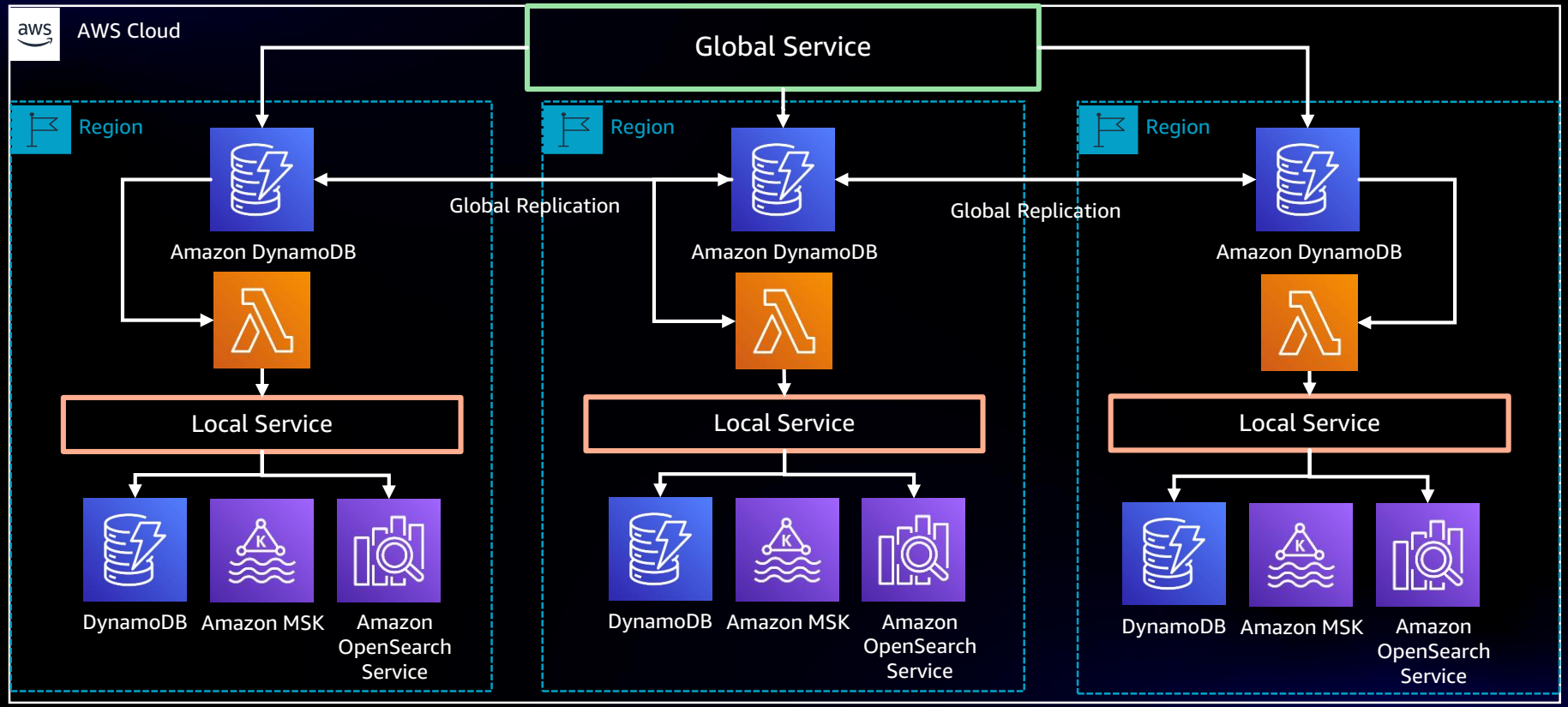
AMAZON ADVERTISING

- Entities created in one Region to be available globally
- Challenge – Our stacks were regional; true globalization would have involved 100s of teams to change their architecture
- DynamoDB Global Tables to rescue!



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

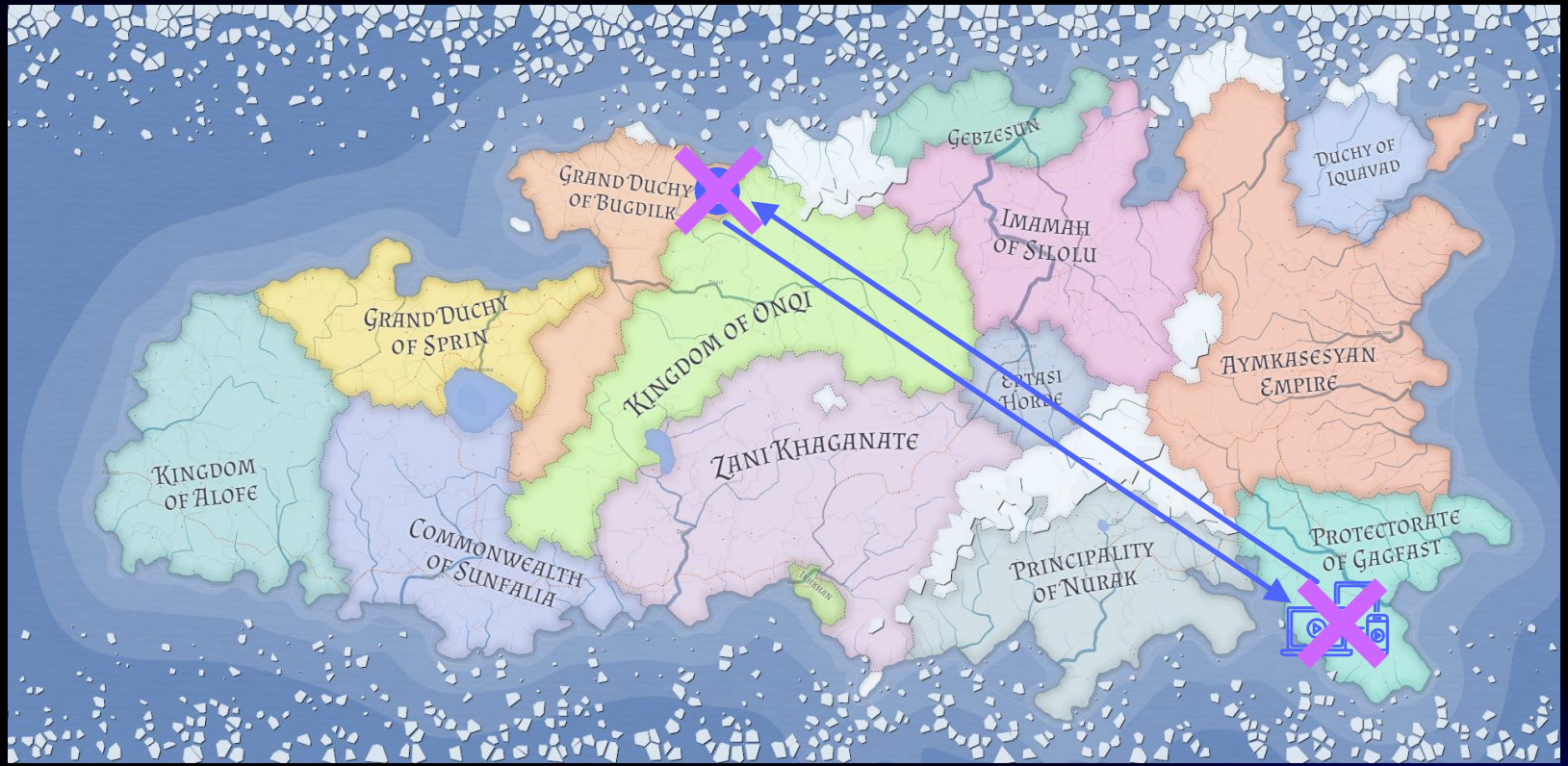
Amazon Ads service globalization



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

Prime Video multi-Region live playback

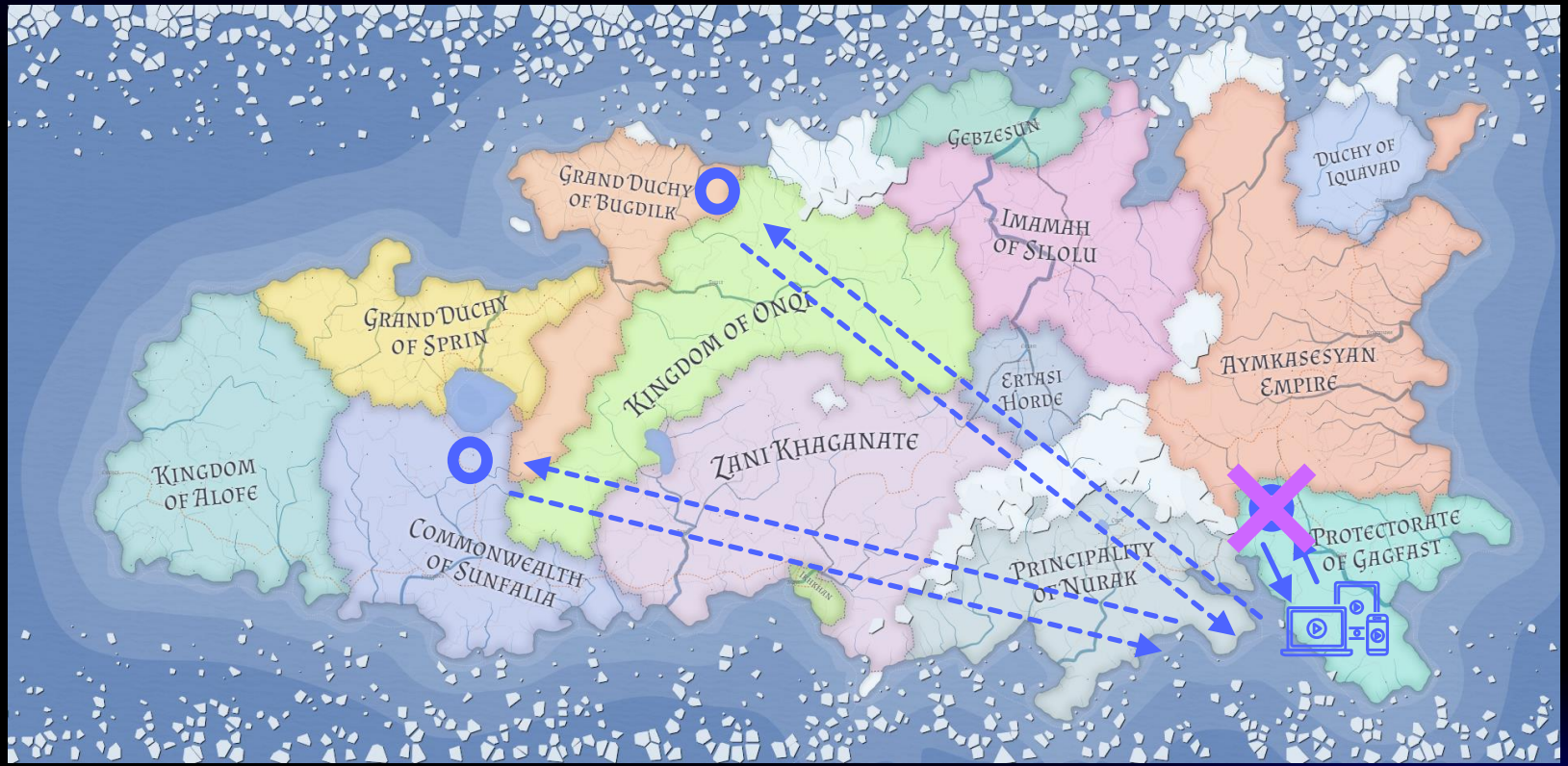
GET CAPACITY CLOSE TO CUSTOMERS AND MITIGATE FAILURES



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

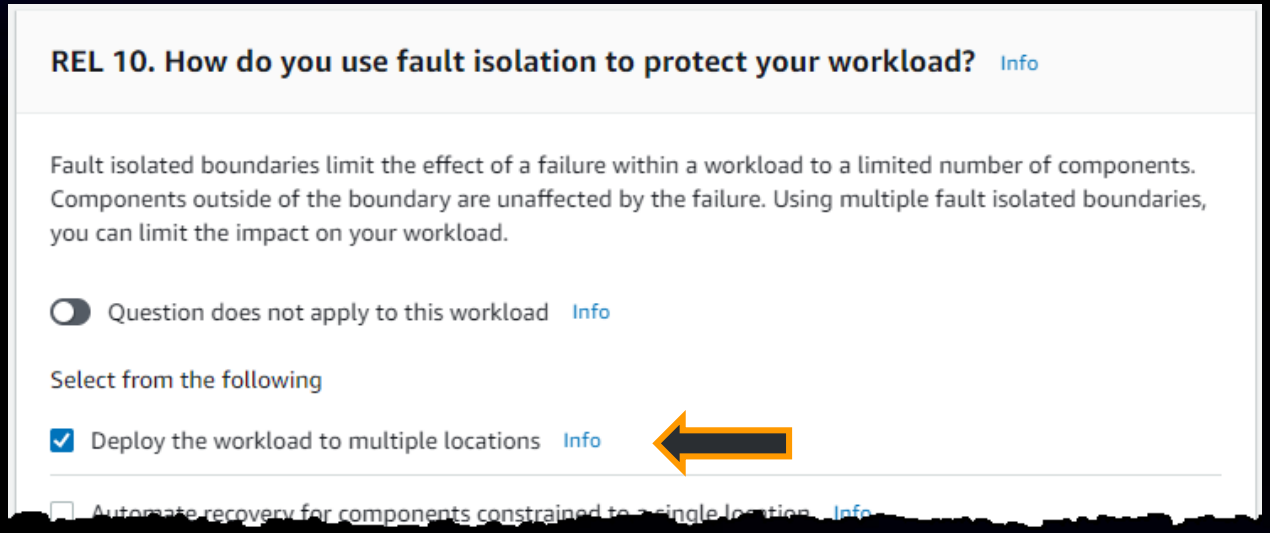
Prime Video multi-Region live playback

GET CAPACITY CLOSE TO CUSTOMERS AND MITIGATE FAILURES



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

Reliability pillar – Use fault isolation



AWS Well-Architected

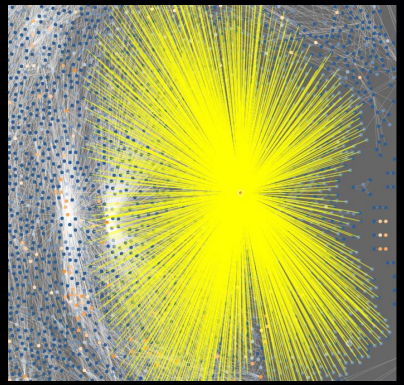
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Summary

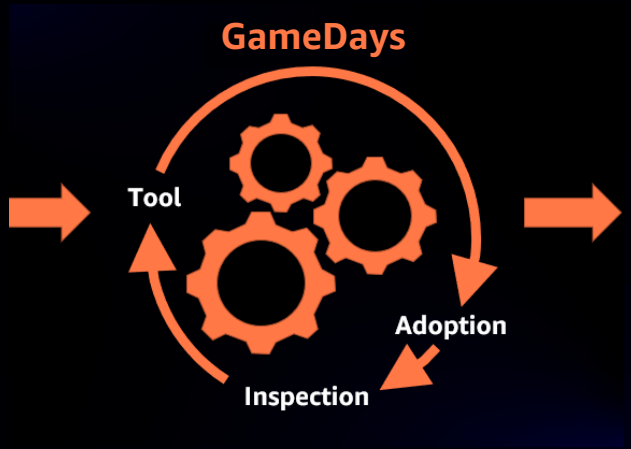
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Reliable scalability for Amazon.com

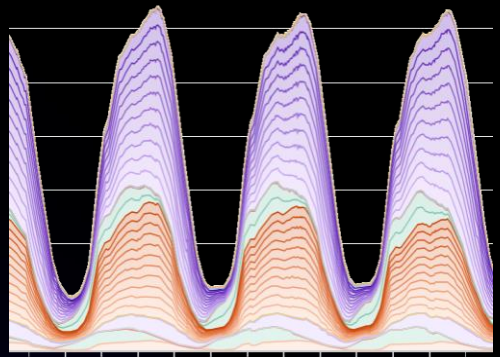
Service-oriented and microservices architectures



Review and test



Understanding services in production



aws

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

Resources

AWS Well-Architected Framework: aws.com/well-architected

AWS Well-Architected Labs: wellarchitectedlabs.com

The Amazon Builders' Library: aws.com/builders-library

AWS Well-Architected Framework whitepaper: bit.ly/waframework



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

Related sessions

ADM301 [Breakout]: Under the hood at Amazon Ads

AMZ201 [Breakout]: Lumos: Amazon's internal, large-scale serverless payments system

AMZ202 [Breakout]: Driving sustainable operations at Amazon and Climate Pledge Arena

AMZ301 [Breakout]: How Amazon migrated a large ecommerce platform to AWS Graviton

AMZ302 [Breakout]: MLOps at Amazon: How to productionize ML workloads at scale

AMZ303 [Chalk Talk]: Amazon's approach to DevSecOps practices

INO201 [Breakout]: Amazon's culture of innovation

INO202 [Breakout]: How Amazon.com transforms customer experiences through AI/ML

ROB201 [Breakout]: Optimizing robotics and autonomous systems with AWS and Amazon

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

Thank you!

Seth Eliot

Twitter: @setheliot
linkedin.com/in/setheliot/

Kieran Kavanagh

linkedin.com/in/kierankavanagh/

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