

SUMMIT

TORONTO | JUNE 14, 2023

ARC301

Multi-Region or single-Region? Considerations and architectures

John Formento, Jr.
Principal SA
AWS

Tarik Makota
Sr. Principal SA
AWS

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

AWS global infrastructure

Why multi-Region?

Multi-Region fundamentals

Open discussion

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS global infrastructure

© 2023, 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

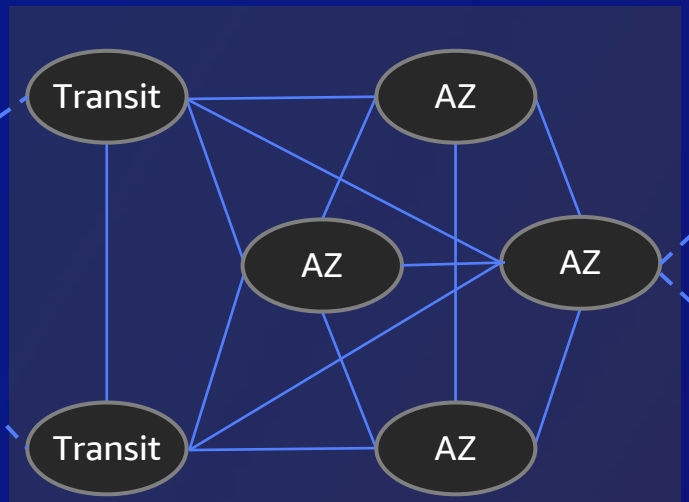
31 AWS Regions worldwide



AWS Regions

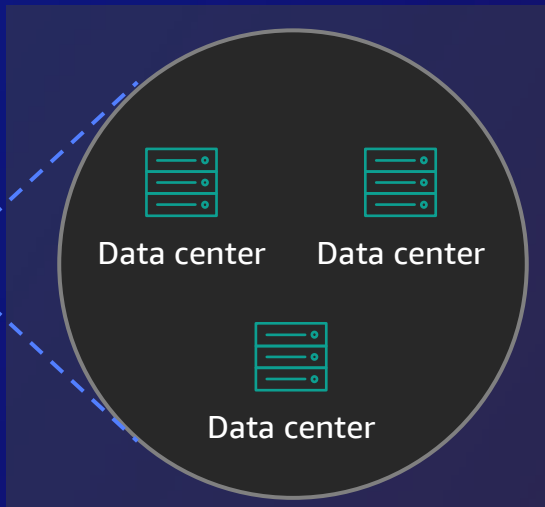
Announced Regions

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

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why multi-Region?

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

“We need to create a multi-Region architecture.”

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

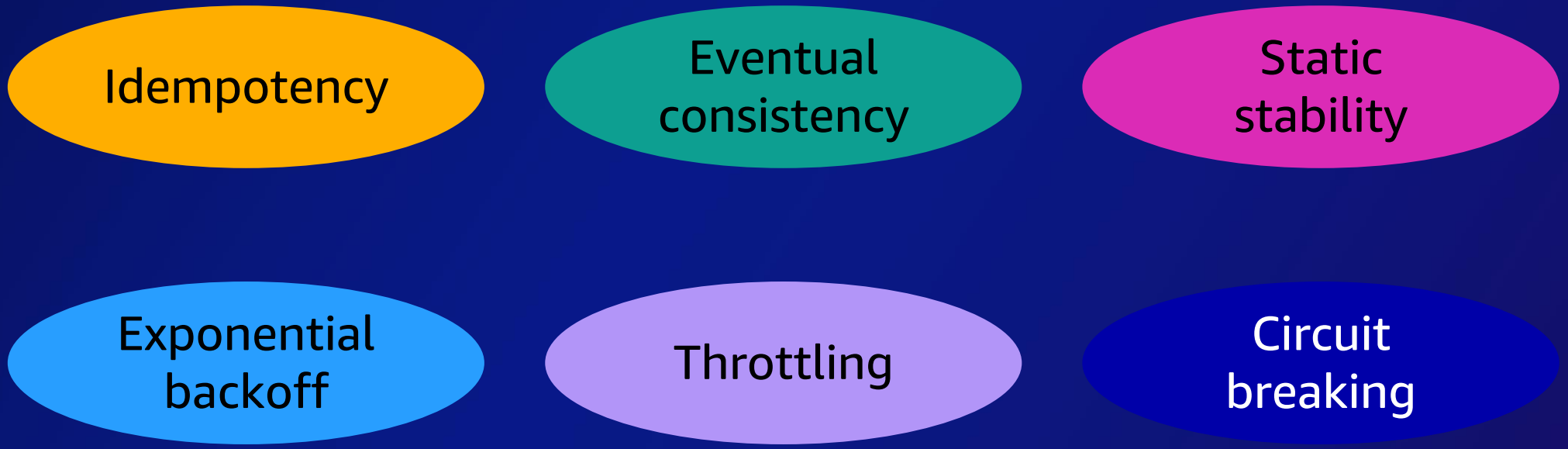
Maybe, but first . . .

Start with improving resilience within Region before multi-Region

- Are you engineering for resilience in Region?
- Are you operating at a high bar within Region?

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Distributed system design best practices



aws
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-Region fundamentals

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Fundamental 1: Understand the requirements

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Are we doing multi-Region for the right reasons?



Increased availability



Regulatory requirements



Improved performance

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What are the business requirements?

- Do we need to meet an extreme resilience requirement?
  - Impact of missing targets?
- Do we need to be able to prove the workload can run from another Region?
  - Regulation guiding approach?

Weigh the requirements against the cost and complexity

Tier	Maximum RTO	Maximum RPO	Criteria	Cost
Platinum	15 minutes	5 minutes	Mission-critical workloads	**$$$**
Gold	15 minutes–6 hours	2 hours	Important, but not mission-critical workloads	**$$**
Silver	**6 hours–few days**	**24 hours**	**Noncritical workloads**	**$**

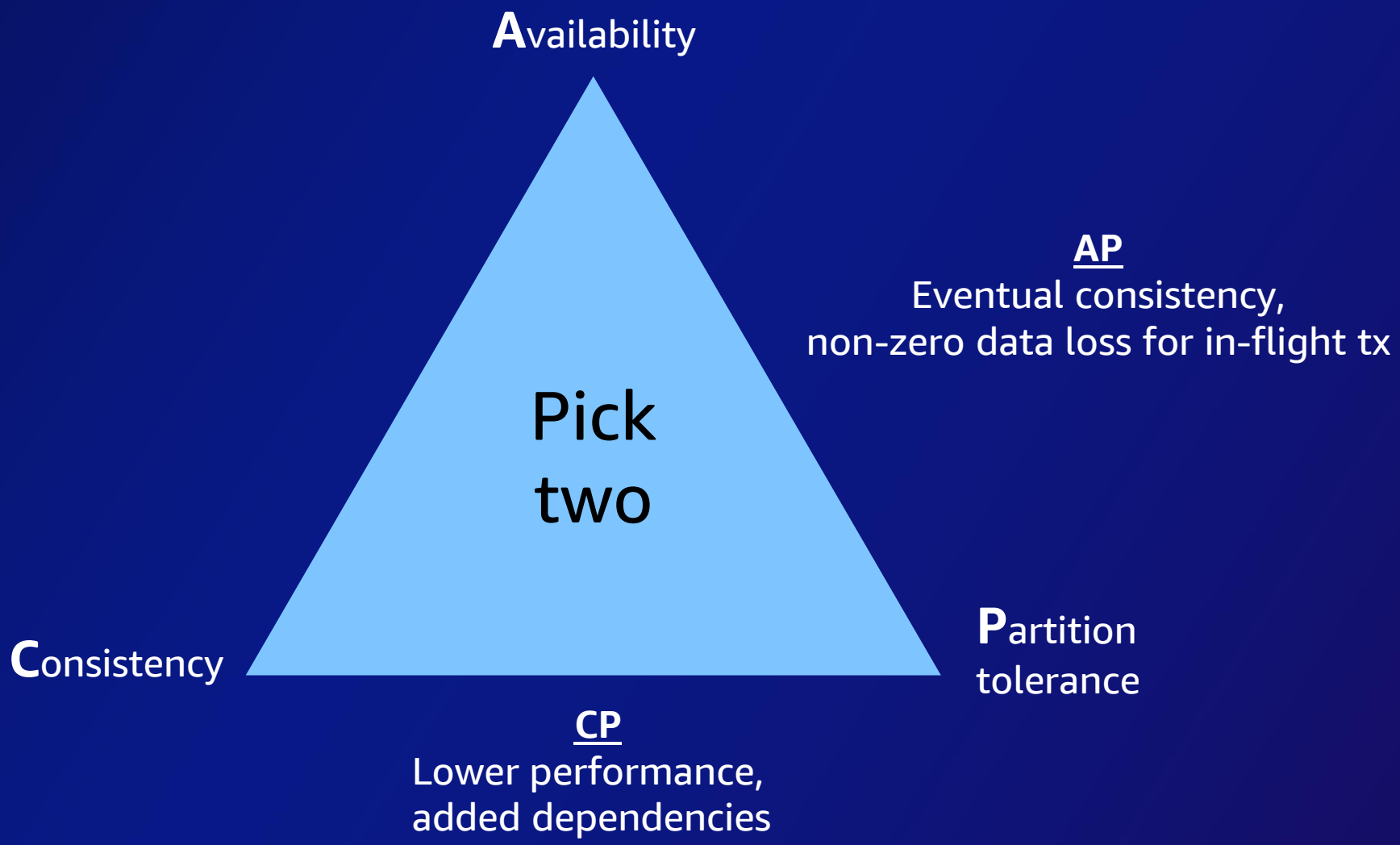
aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Fundamental 2: Understand the data

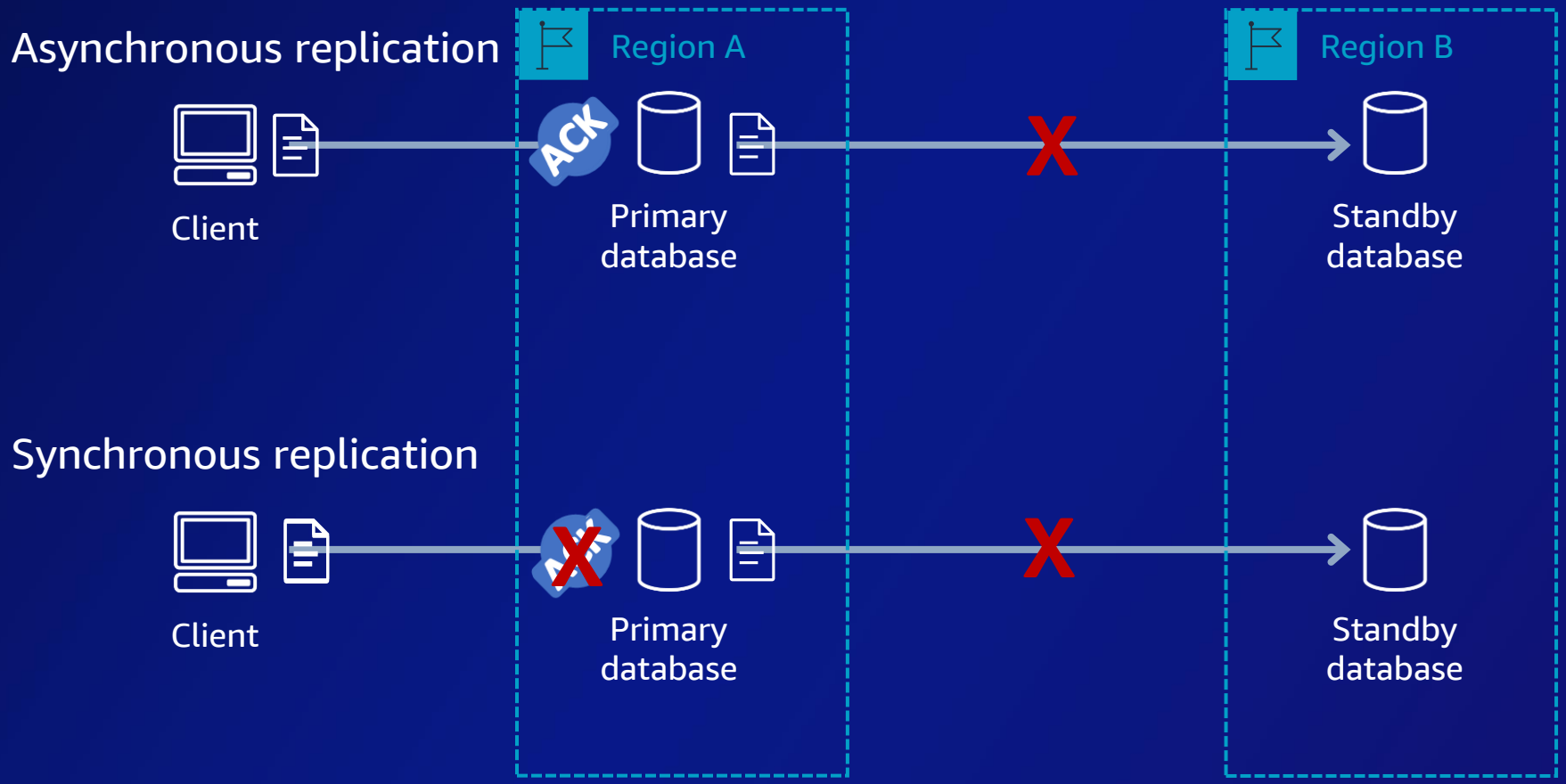
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data consistency requirements



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

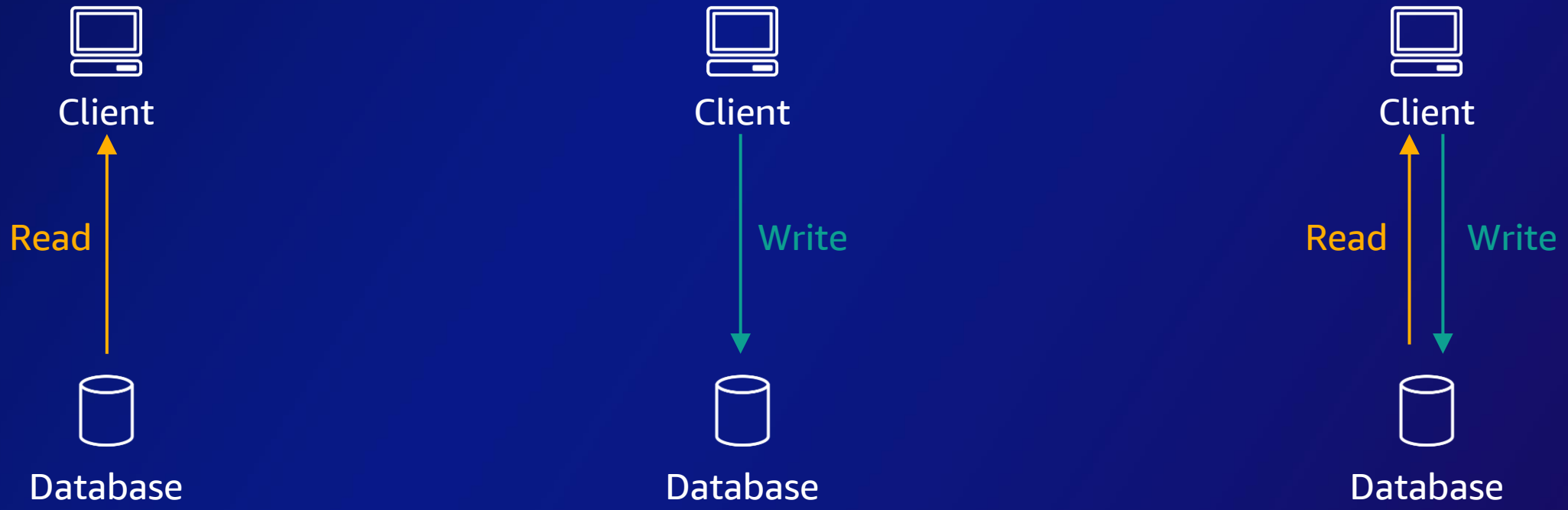
Data consistency requirements



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data access patterns

Understand the access characteristics



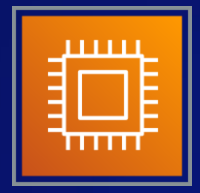
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Fundamental 3:
Understand the dependencies

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS services

- Determine what services are available in desired Regions
- Identify what features will aid in your multi-Region journey


Compute


Storage


Database


Network and content delivery


Analytics


Management and governance


Security

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Other dependencies

- Internal (cloud or on-premises) dependencies
- Third-party dependencies
- Failover mechanism

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Fundamental 4: Operational readiness

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

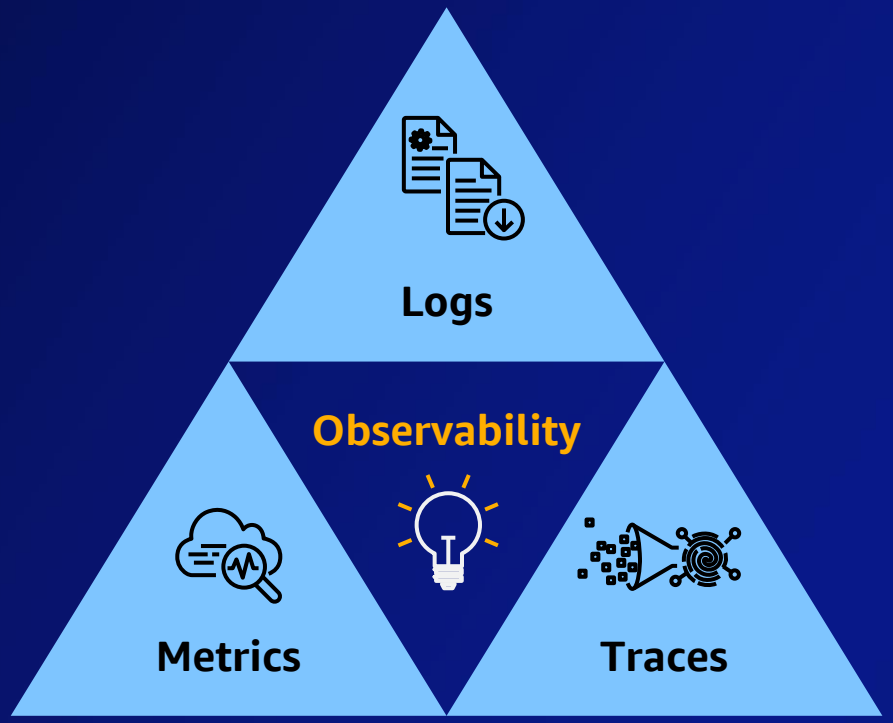
Deploy to one Region at a time

BLUE/GREEN OR CANARY DEPLOYMENTS



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Monitoring: What changes in multi-Region?



- Replicate only high-level metrics to the other Region
  - User experience
  - User count
  - Transaction count
  - Replication status
- Monitor the health of the application in Region A from tooling in Region B
- Use telemetry from Region B for determining when to fail over from application in Region A

aws
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cost and complexity

- Engineering effort
  - Is the application(s) ready for multi-Region?
- Operational overhead
  - People and process
  - Testing
  - Enhanced deployment processes
- Cost
  - Isolation of application stacks between Regions can result in 2x cost
  - Increased costs from the operational efforts and engineering efforts are on top of infrastructure costs

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

skillbuilder.aws

Your time is now

Build in-demand cloud skills your way

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!



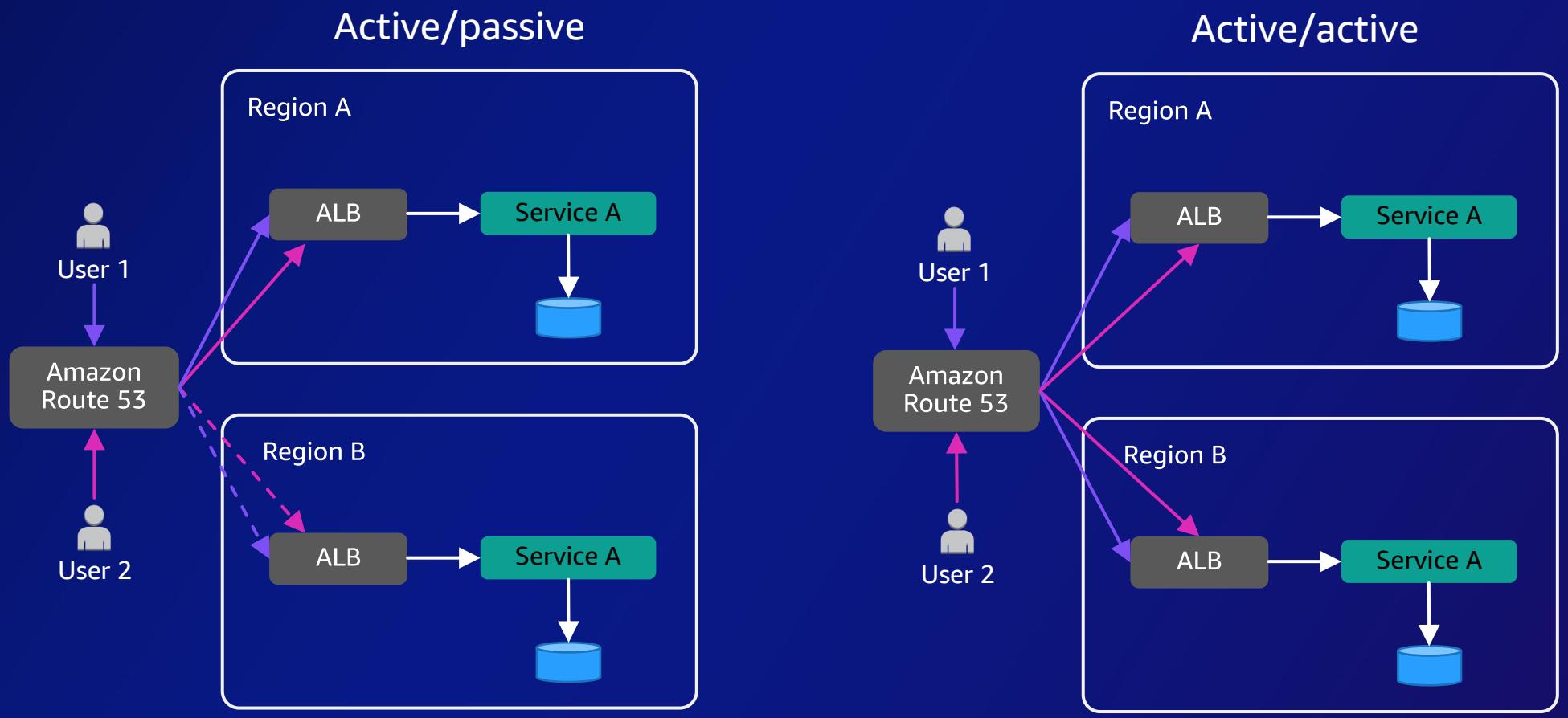
Please complete the session survey in the mobile app

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Reference architectures

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Two macro patterns

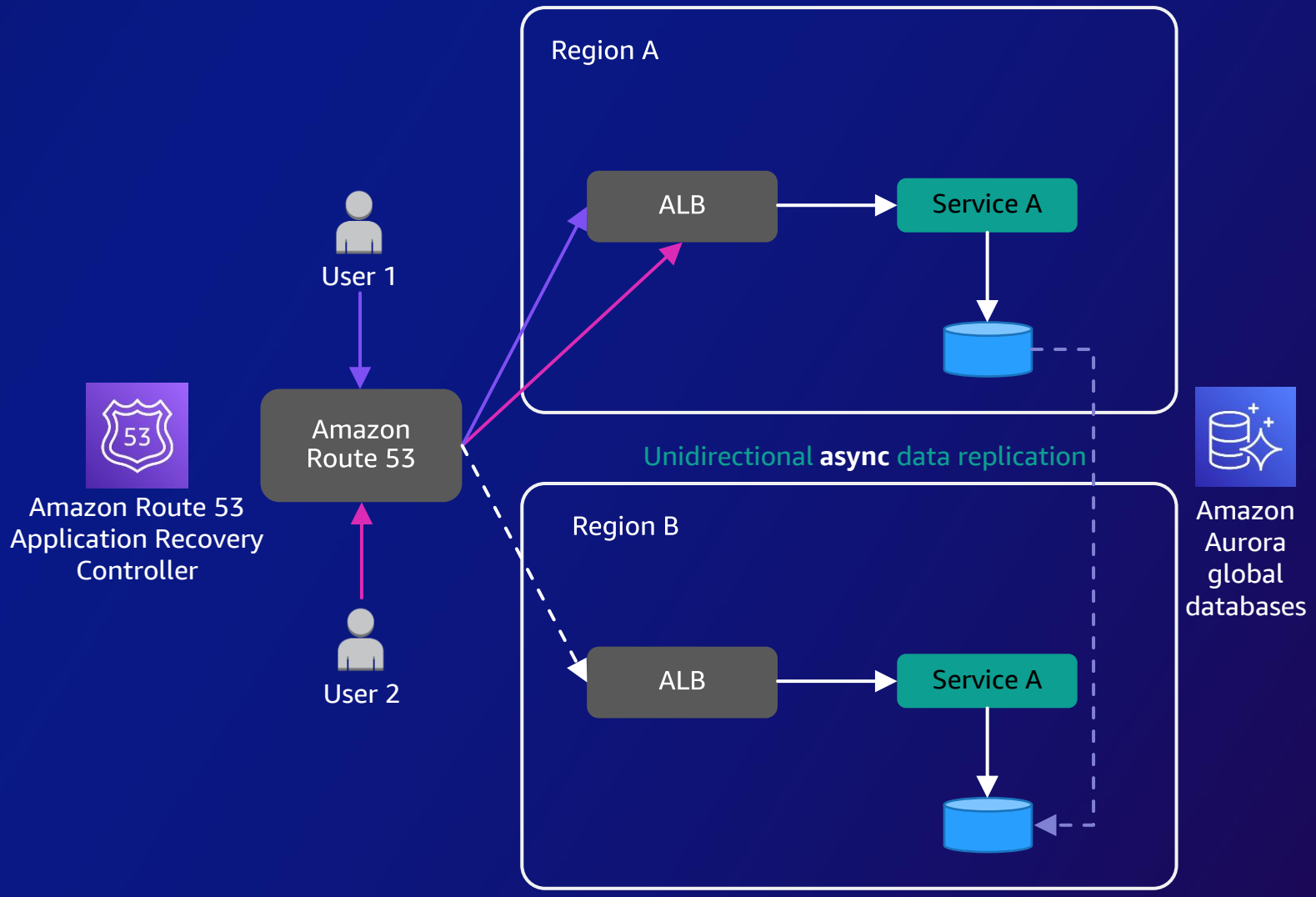


© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pattern 1: Active/passive

PILOT LIGHT, WARM STANDBY, ACTIVE/HOT STANDBY

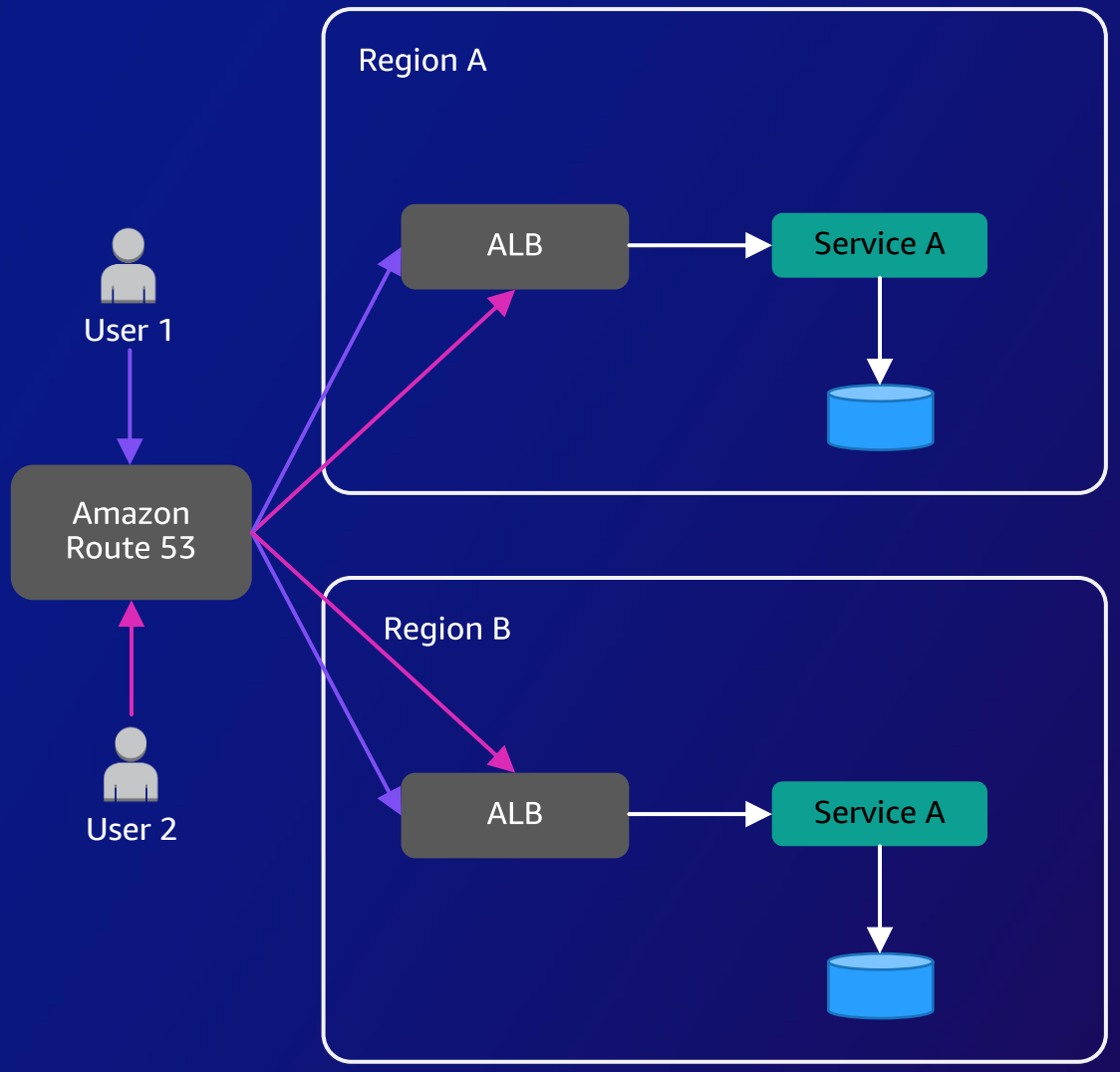
- Use cases
  - Regional failover/failback for DR/ operational continuity
  - Regulatory requirements
- Key design considerations
  - Observability (health checks)
  - CI/CD, code and configuration drifts
  - Routing strategy
  - Failover/failback procedures



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pattern 2: Active/active

- Use cases
  - Extreme high availability
  - Performance – latency sensitivity
  - Low RTO/RPO

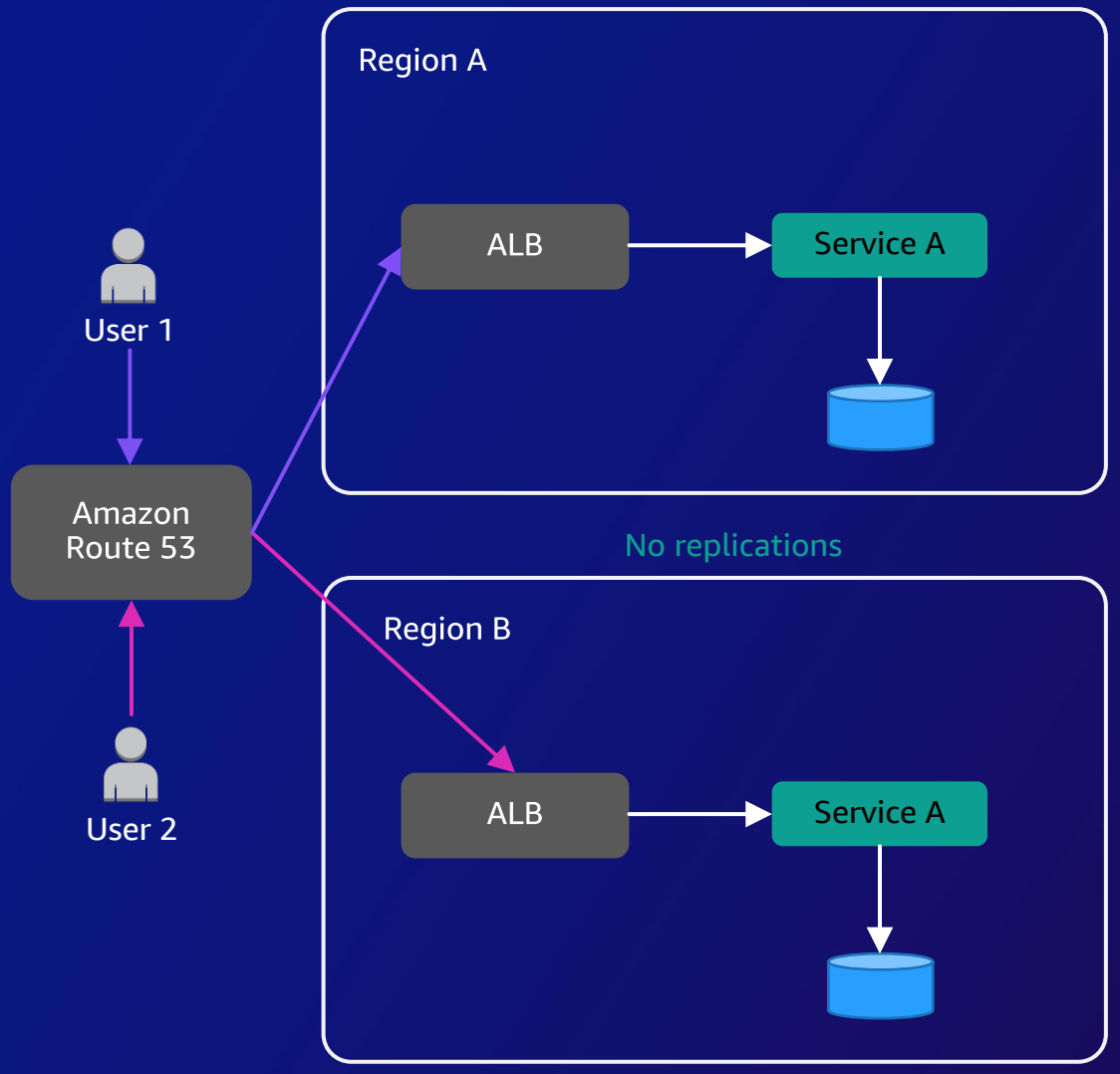


© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pattern 2.1: Active/active

REGIONAL SHARDING

- Use cases
  - Global expansion
  - Data locality, regulatory
  - Performance – latency-sensitive
- Key design considerations
  - Observability (health checks)
  - CI/CD, code and configuration drifts
  - Routing strategy

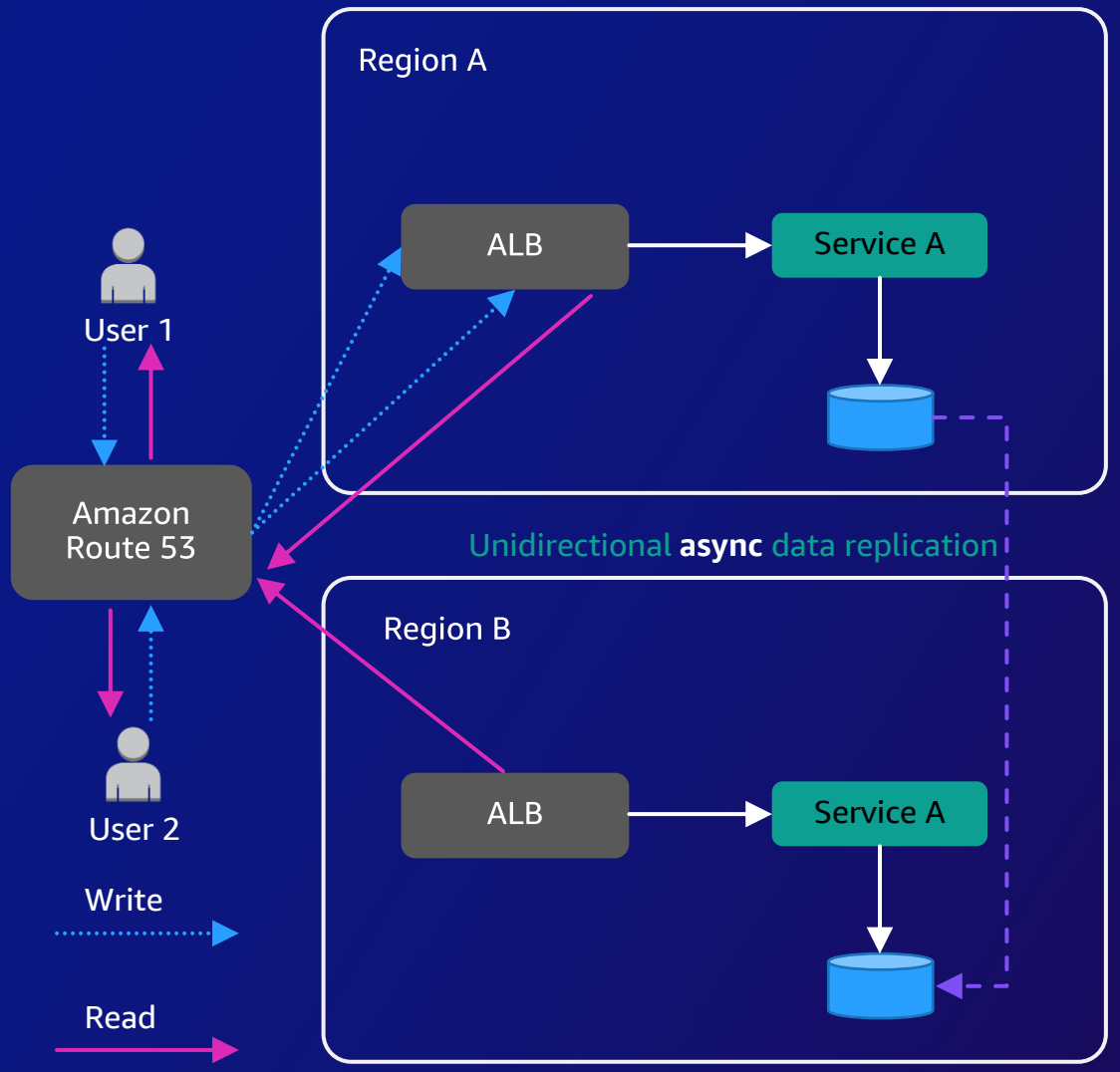


© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pattern 2.2: Active/active

SINGLE WRITER; MULTIPLE READERS

- Use cases
  - High availability
  - Latency sensitivity
  - Eventual consistency – async replication
  - Read-heavy workloads
- Key design considerations
  - Observability (health checks)
  - CI/CD, code and configuration drifts
  - Routing strategy
  - **Failover/failback procedures (writer)**
  - **Read after write consistency**

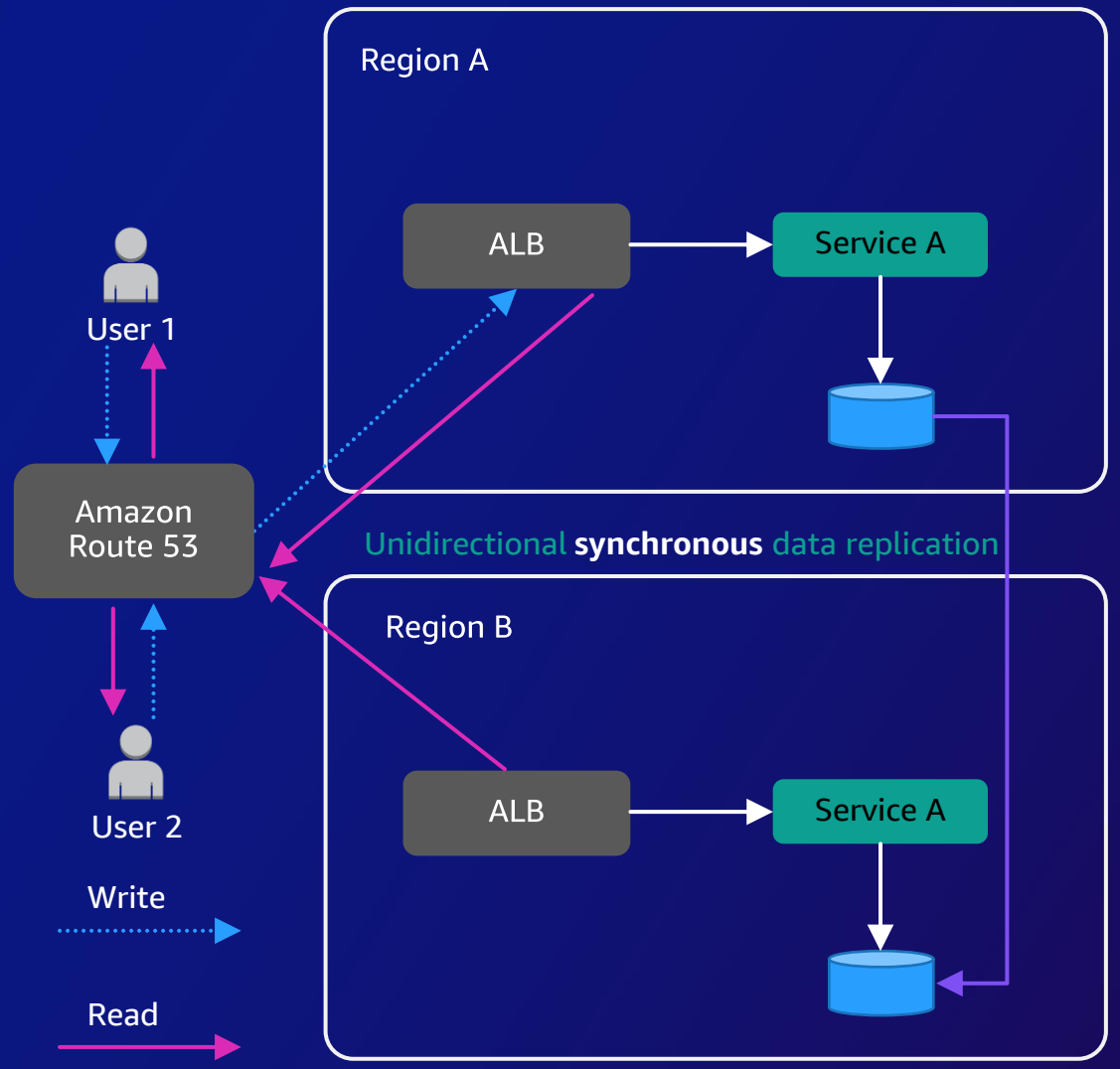


© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pattern 2.3: Active/active

SINGLE WRITER; MULTIPLE READERS (STRONG CONSISTENCY)

- Use cases
  - Read-heavy workloads
  - Strong read after write consistency
  - Regional failover/failback for DR/operational continuity
  - RPO = 0; regulatory requirements
- Key design considerations
  - Observability (health checks)
  - CI/CD, code and configuration drifts
  - Routing strategy
  - Failover/failback procedures (writer)
  - **Increased latency – performance**

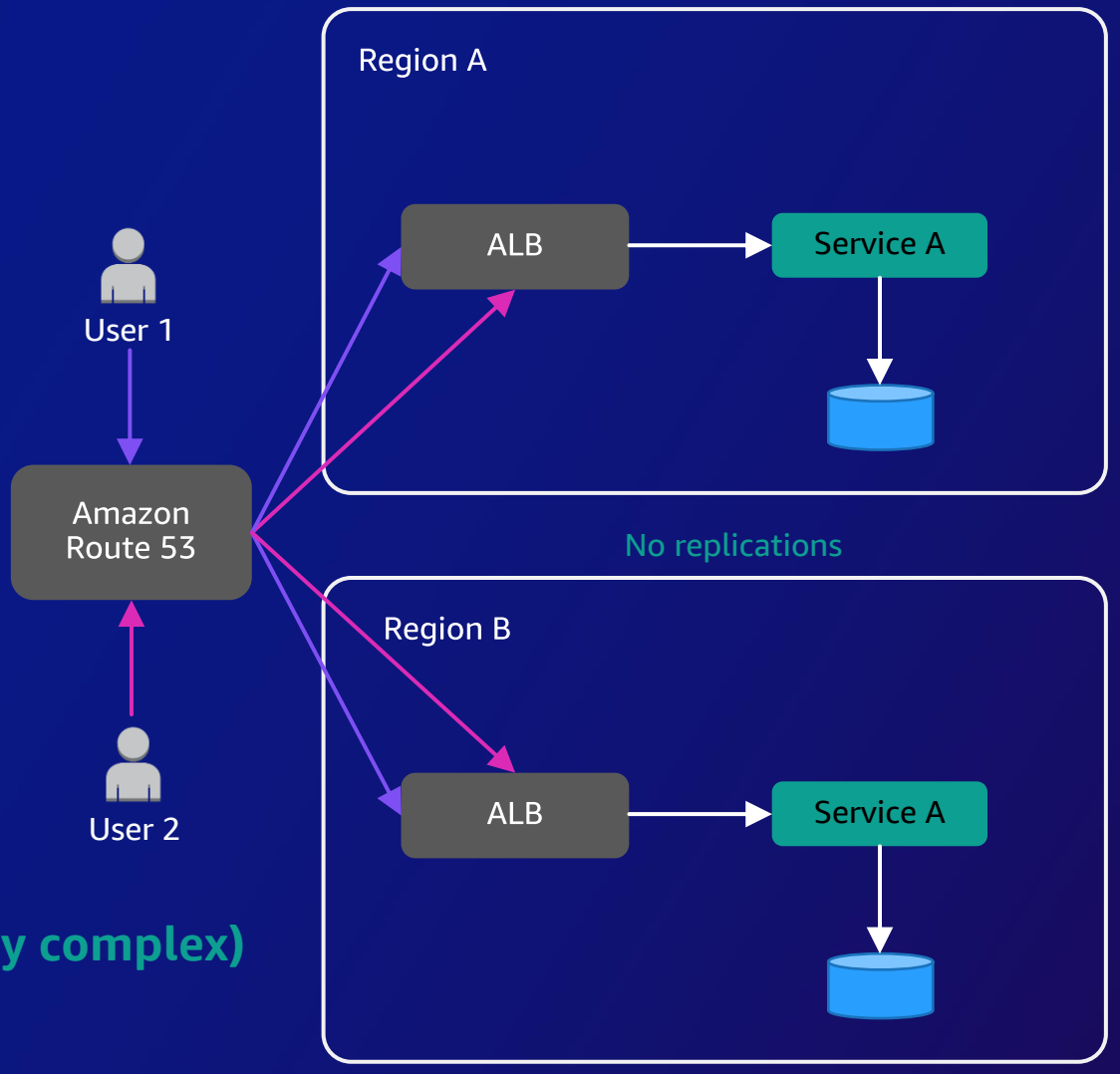


© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pattern 2.4: Active/active

DUAL WRITES

- Use cases
  - Read-heavy workloads
  - Strong read after write consistency
  - Regional failover/failback for DR/operational continuity
  - RPO = 0; regulatory requirements
- Key design considerations
  - Observability (health checks)
  - CI/CD, code and configuration drifts
  - Routing strategy
  - **Idempotency**
  - **Managing data atomicity at the client side (highly complex)**

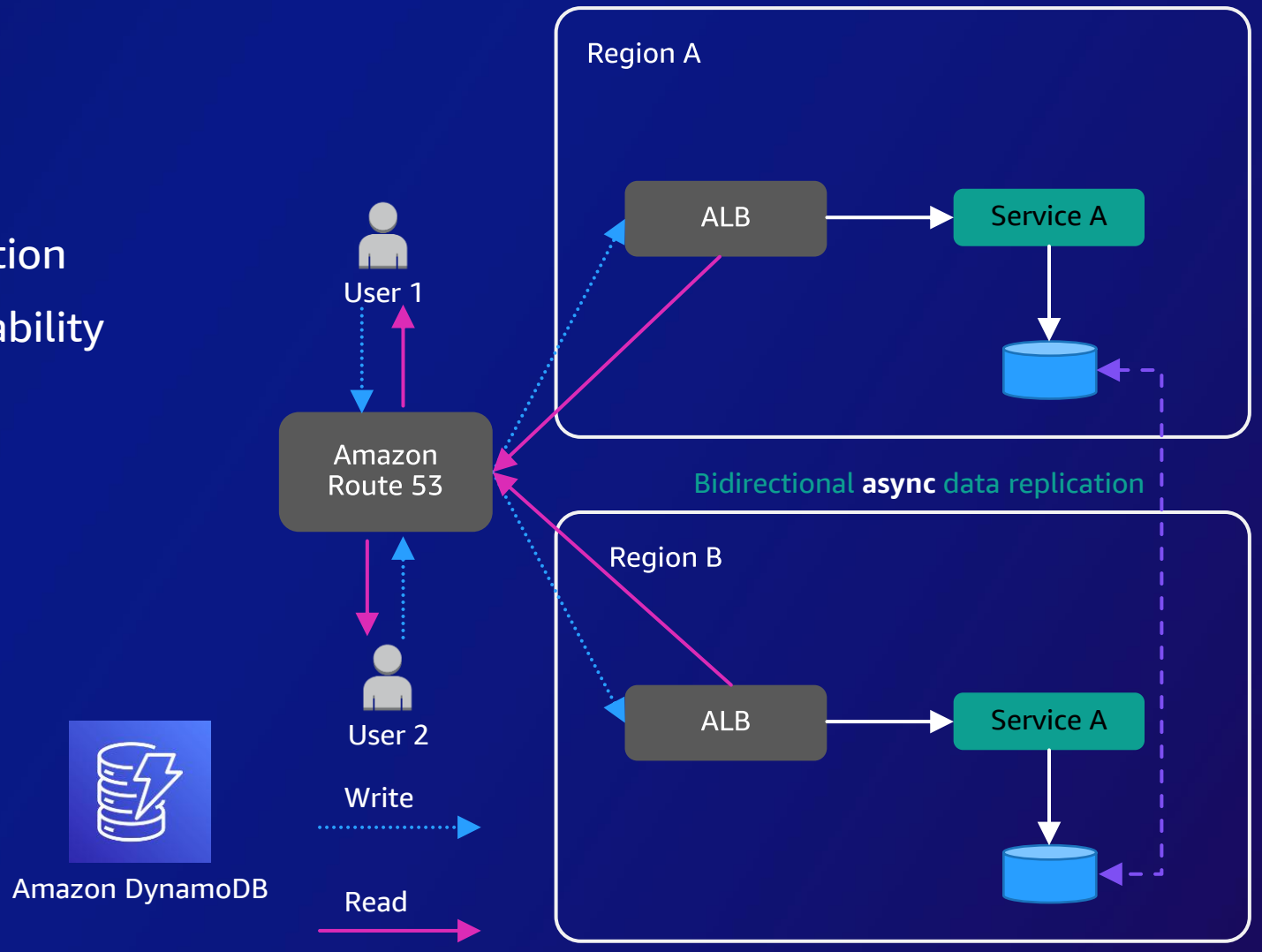


© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pattern 2.5: Active/active

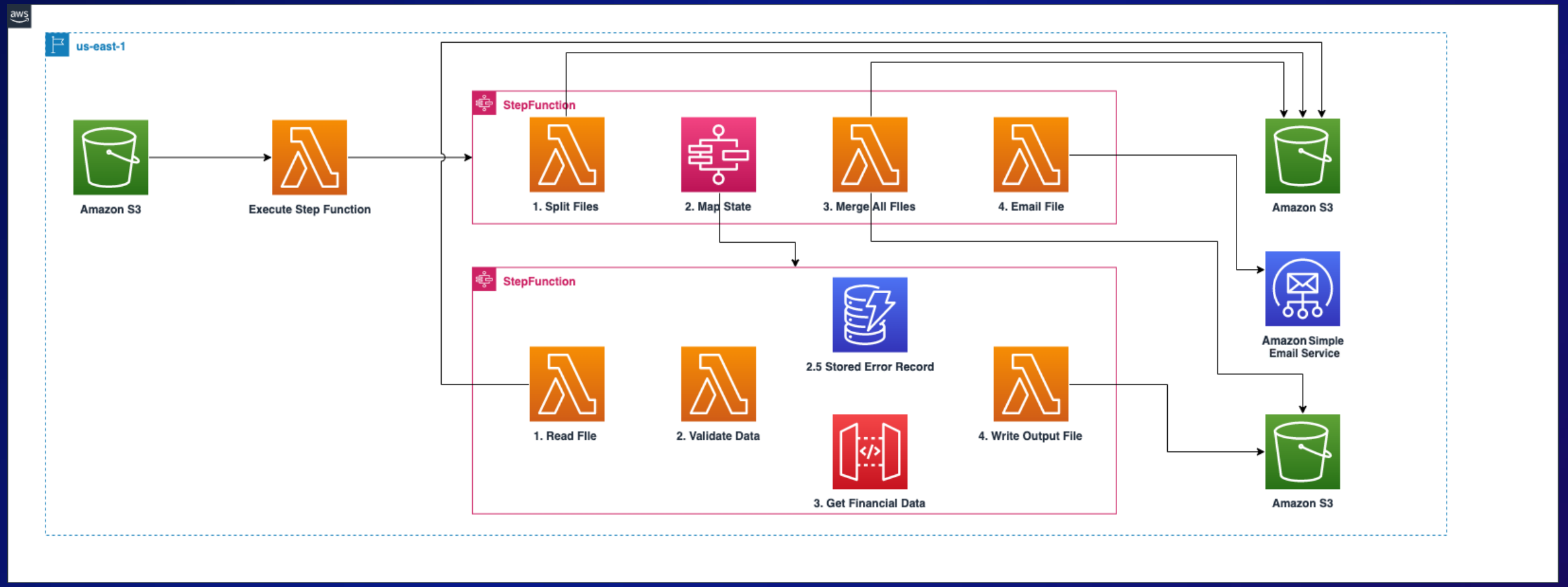
MULTIPLE WRITERS; MULTIPLE READERS

- Use cases
  - High availability
  - Latency sensitivity
  - Eventual consistency – async replication
  - Read-/write-heavy workloads – scalability
- Key design considerations
  - Observability (health checks)
  - CI/CD, code and configuration drifts
  - Routing strategy
  - **Conflict resolution**
  - **Read after write consistency**



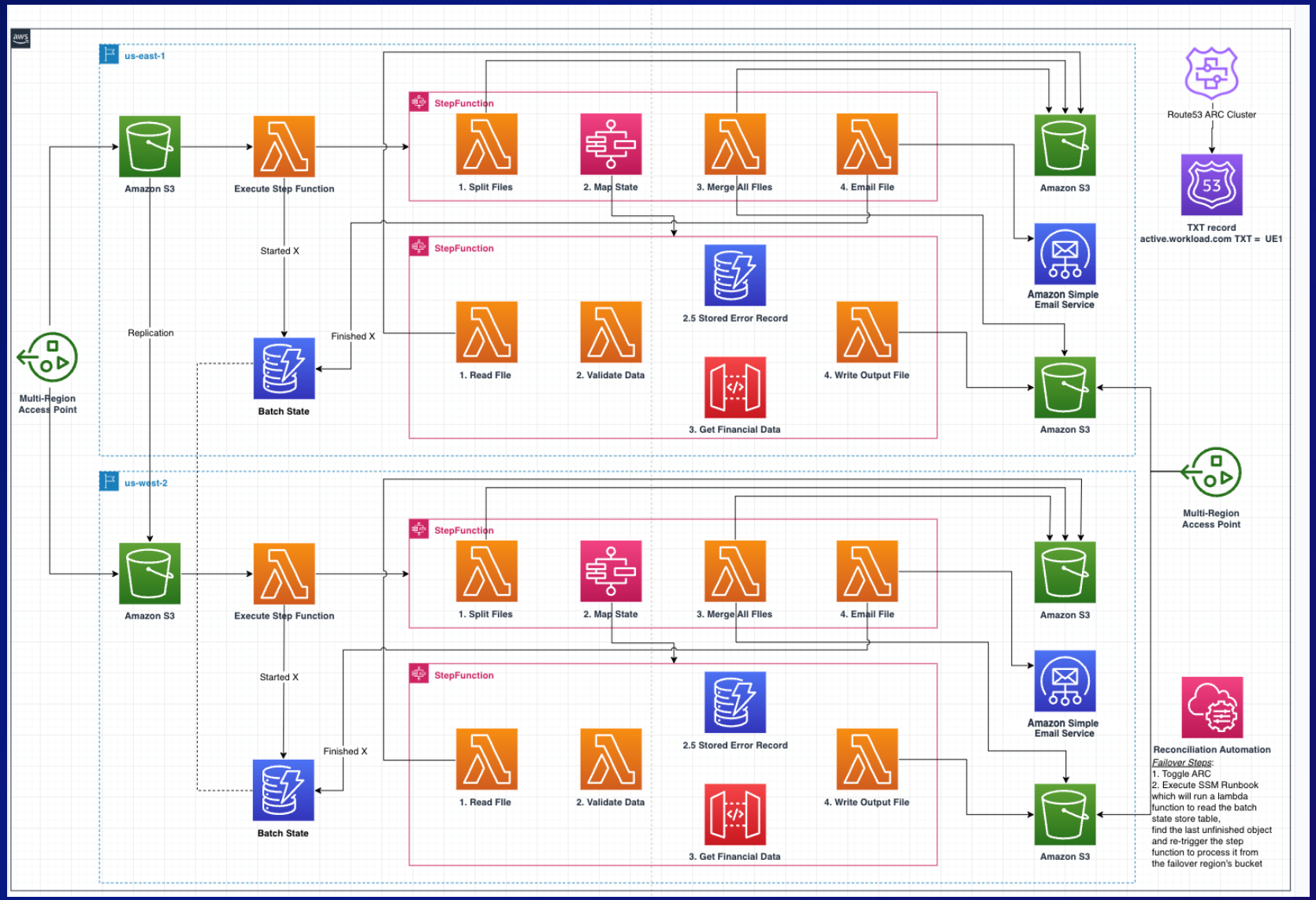
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-Region batch



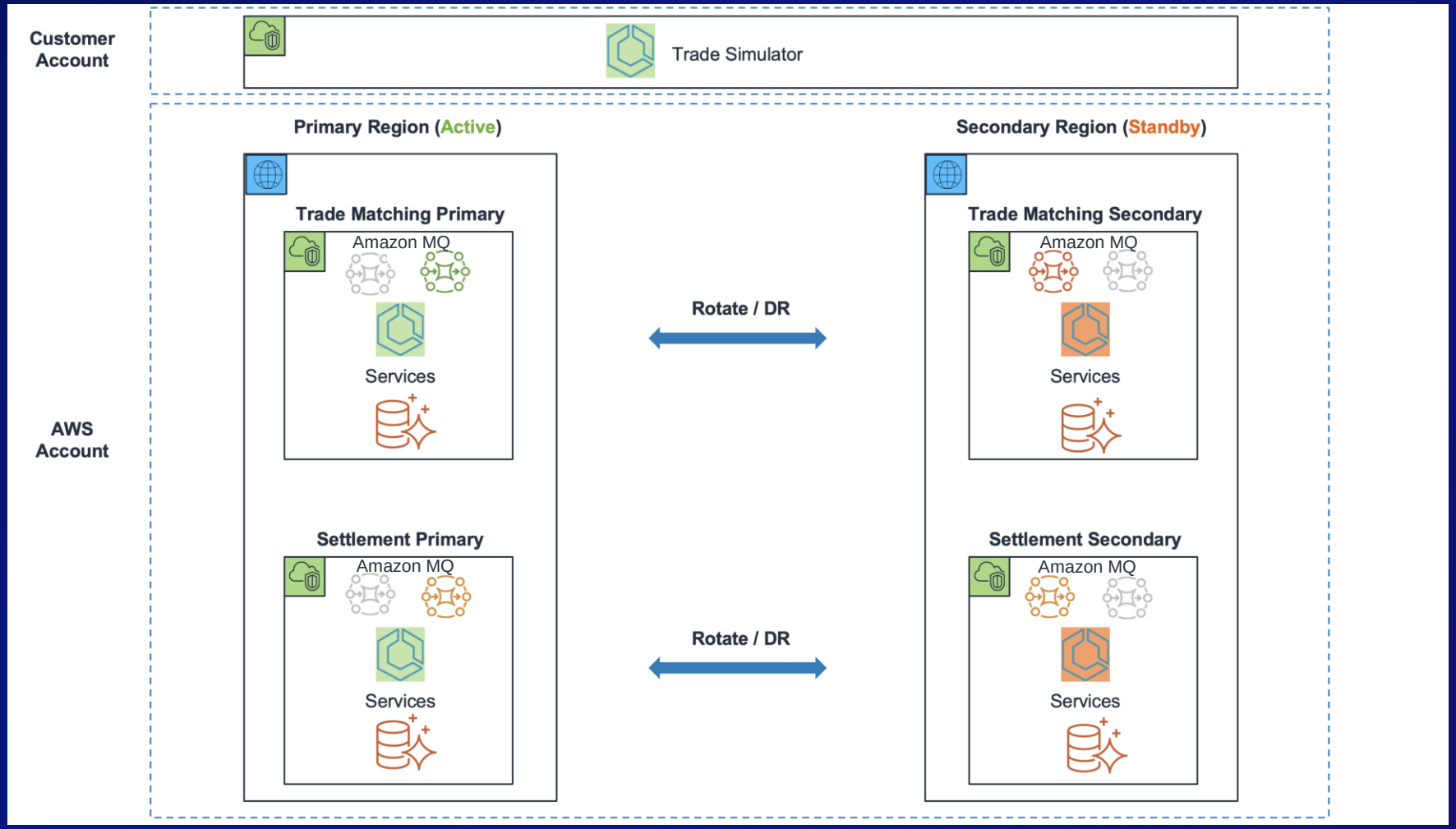
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-Region batch



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

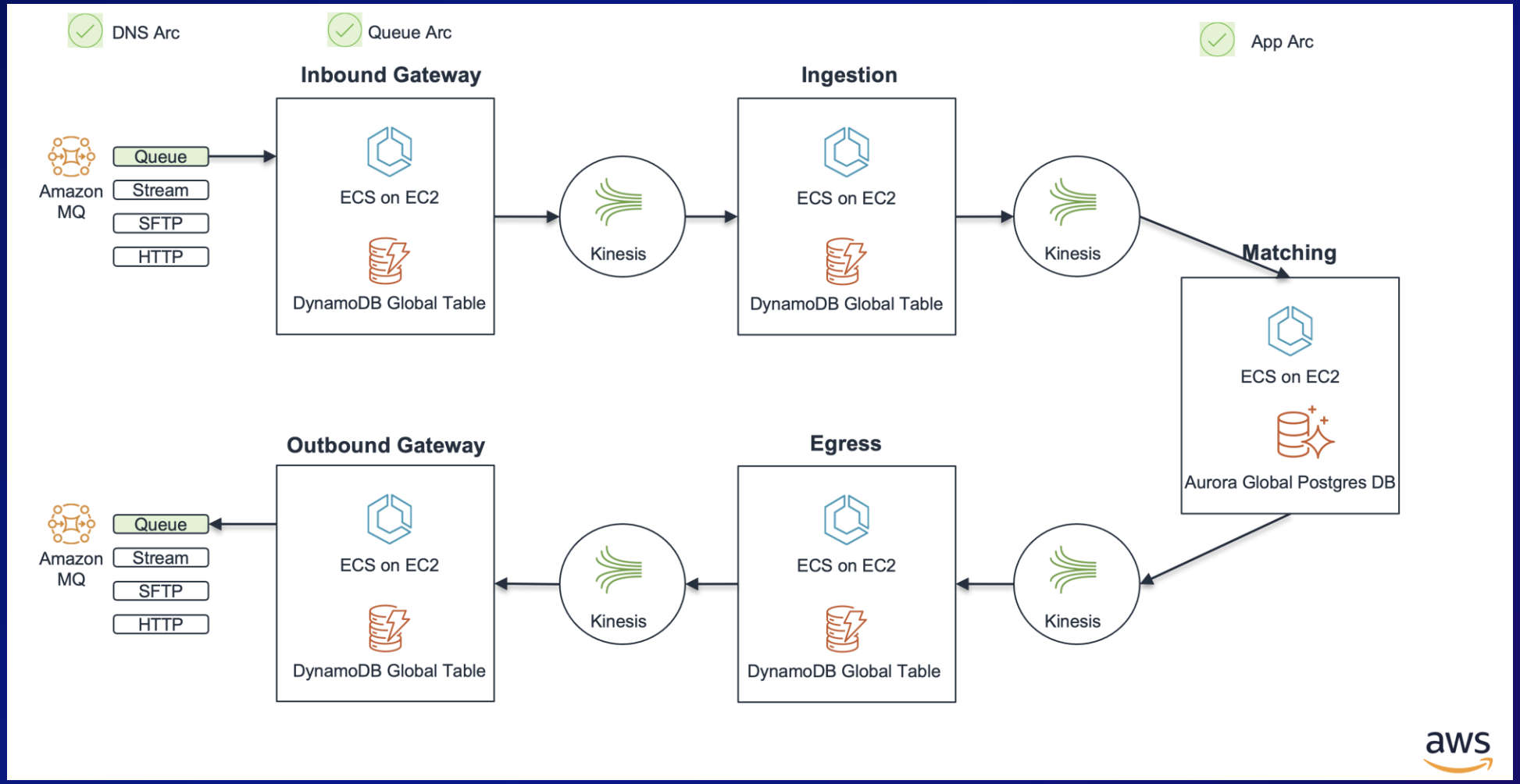
Multi-Region trade and settlement application



https://github.com/aws-samples/multi-region-resiliency-reference-implementation

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

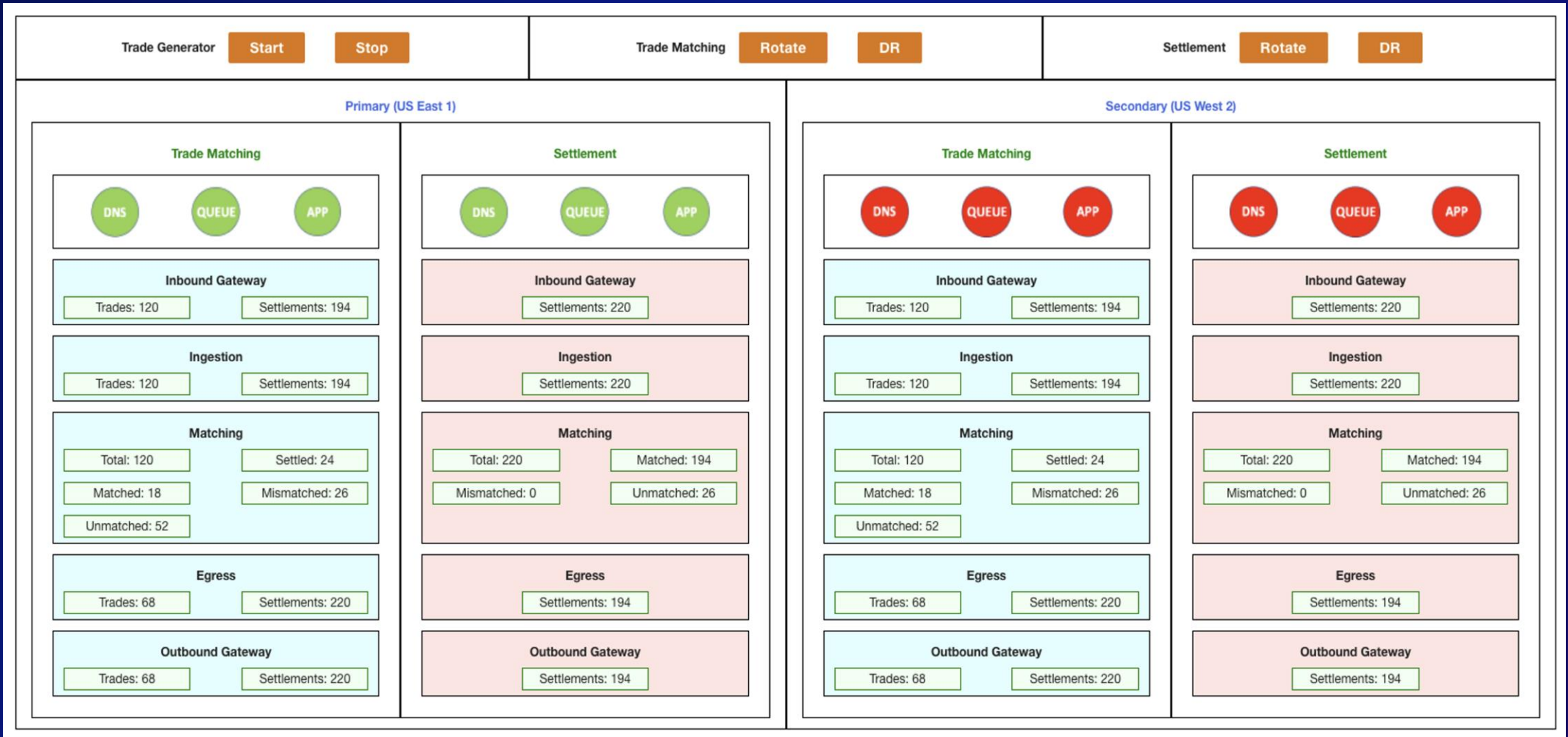
Multi-Region trade and settlement application



https://github.com/aws-samples/multi-region-resiliency-reference-implementation

aws
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-Region trade and settlement application



https://github.com/aws-samples/multi-region-resiliency-reference-implementation

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.