
AWS
re:Invent

DAT205

Enterprise database transition: License-free cloud databases

Cetin Karakus
Head of Quantitative and Analytical Solutions
bp

Krishna Sikharam
Technology Lead
bp

Paola Lorusso
Solutions Architect
Database Specialist
AWS

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Session Outline

- Enterprise Database Transition - CK
- Market Intelligence Platform - KS
- Migration Process - KS
- Key Tools and Tech - CK
- Best Practices - CK



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Enterprise Database Transition



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is Enterprise Database Transition (EDT) ?



Expensive	→	Cost Effective
Proprietary	→	Open Source
License Restricted	→	License-free

Enterprise Database Transition Drivers

Cost Savings

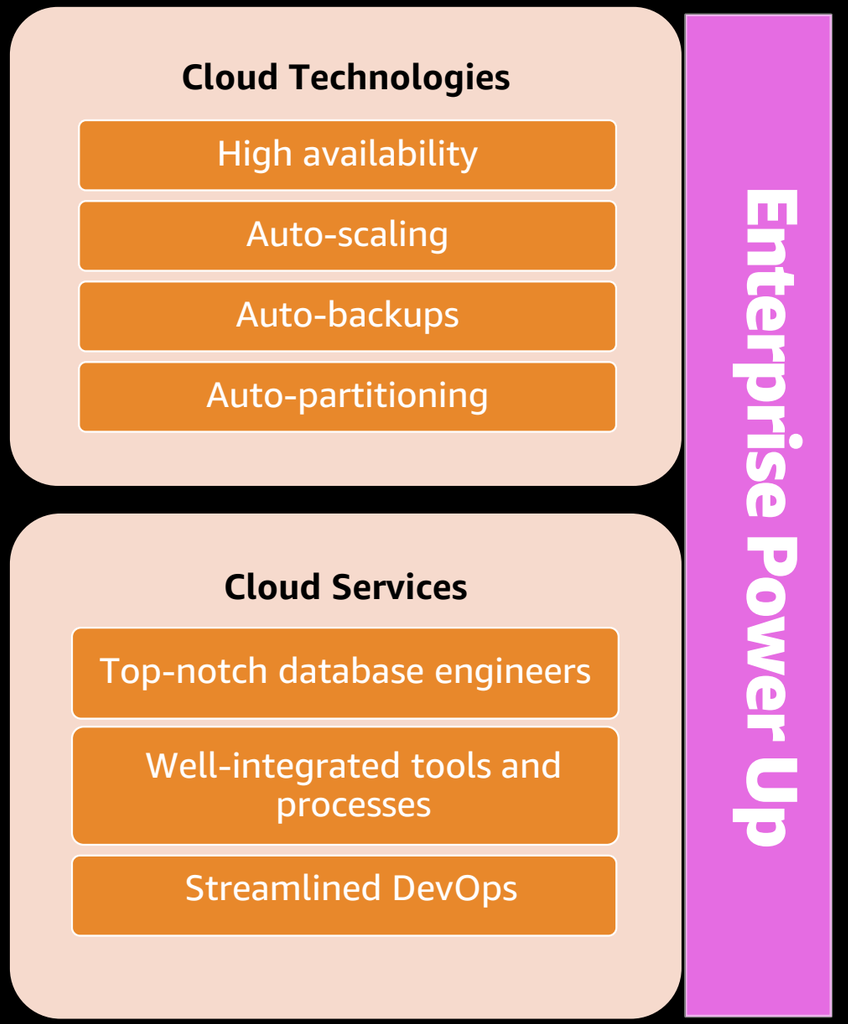
- Potentially huge license cost savings

New Architectural Vistas

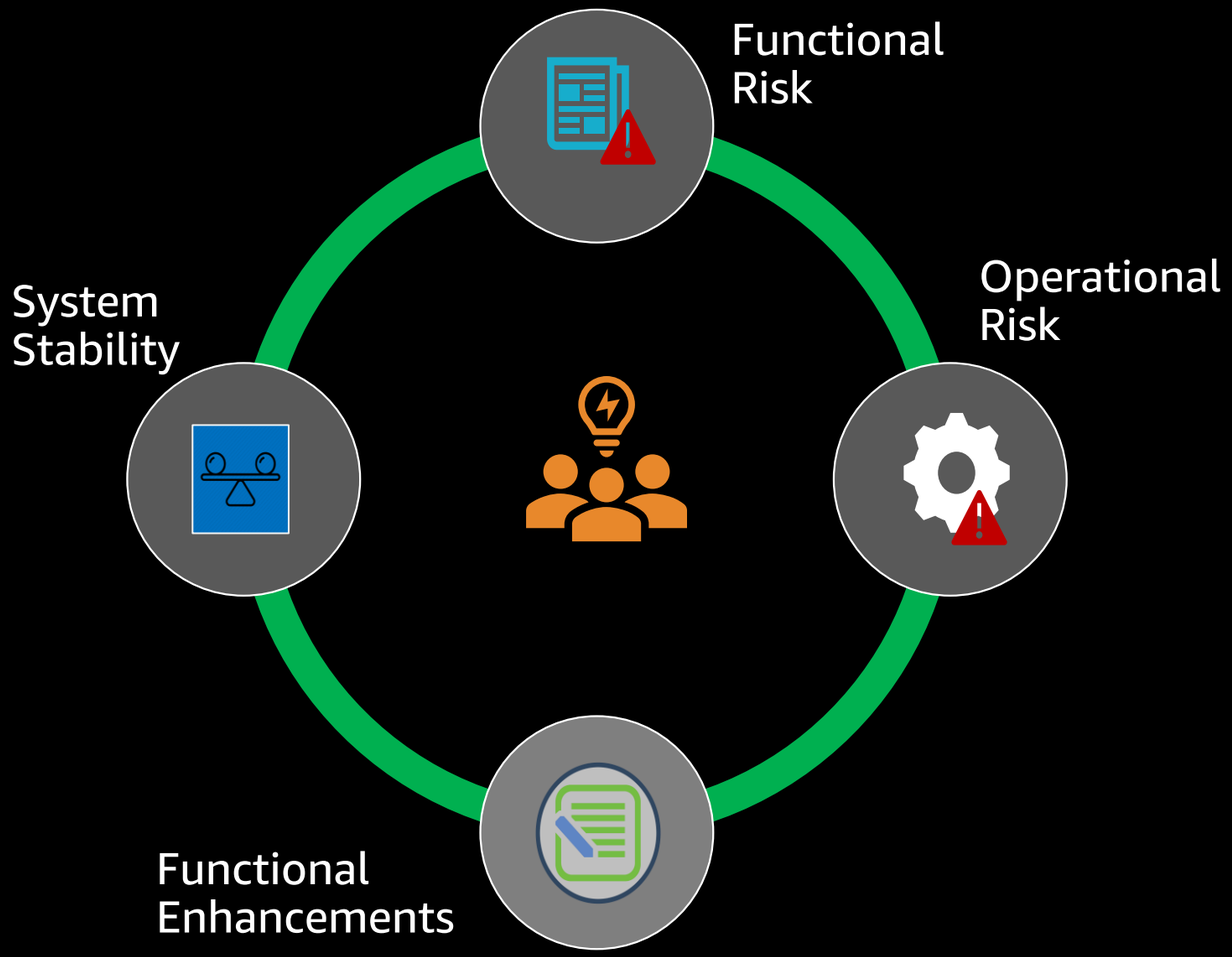
- Without worrying about license-cost explosion

Cloud-native Mindset

- Design solutions based on cloud technologies



Stakeholders' Perspective



Dual Assurance Framework

Functional Assurance

- Iso-Functional
- Iso-Performant
- Iso-Stable

Fail-safety Assurance

- Make-before-Break
- Staged Rollout
- Standby Switch-over

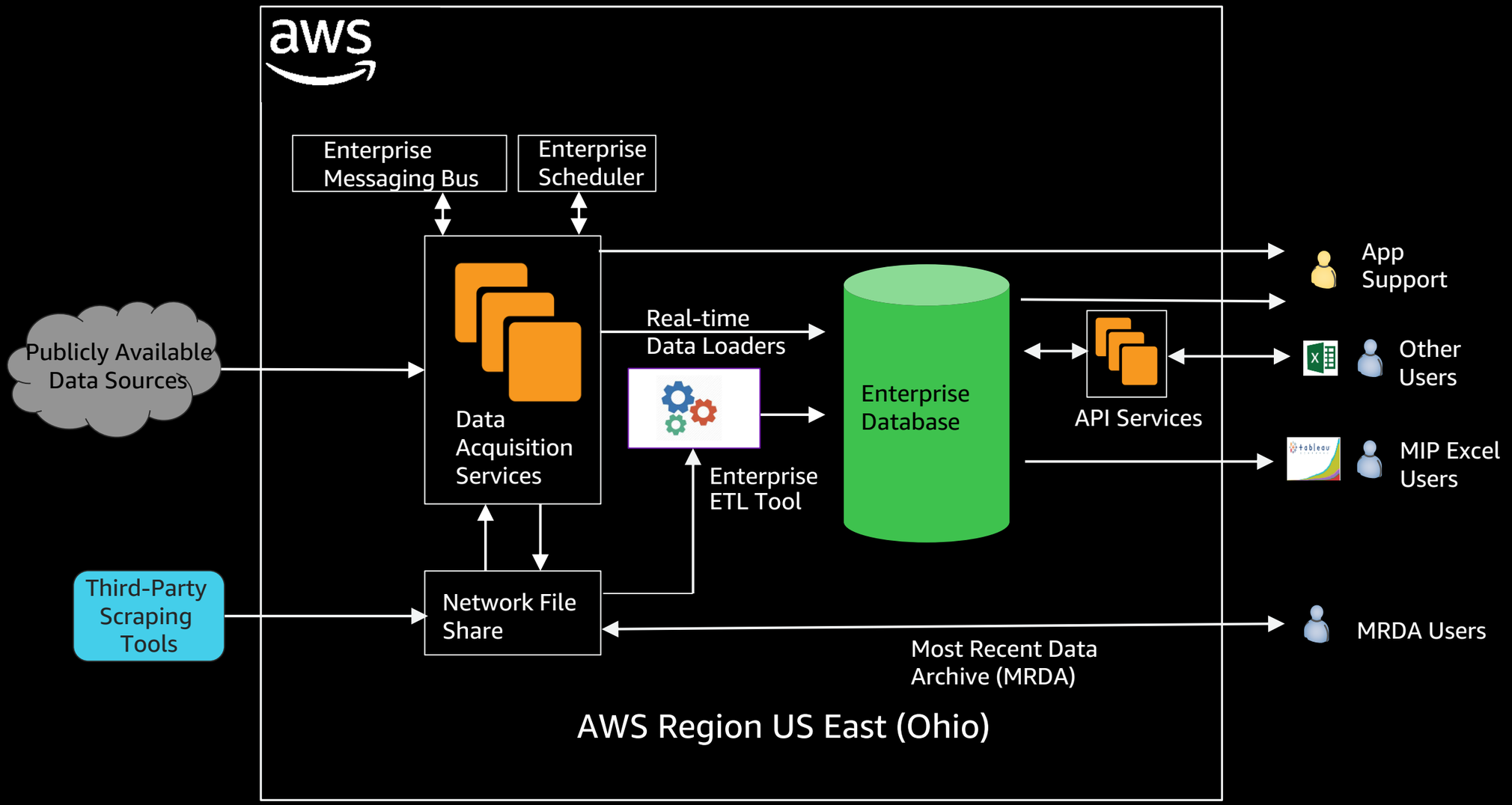
Market Intelligence Platform (MIP)



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

MIP Architecture



MIP Database Overview

- ✓ Multiple schemas
- ✓ Complex partitioning
- ✓ Terabytes of data
- ✓ Star schema
- ✓ Thousands of user queries
- ✓ Low latency
- ✓ Business critical
- ✓ HA requirement
- ✓ Complex ingestion

EnterpriseDB Features	
Enterprise Edition	Y
Partitioning	Y
Application Features	
Sequences	Y
Stored Procedures	Y
Triggers	Y
Misc. Features	
Star Schema	Y
Enterprise User Security	Y

Factor	Size
Schema	8
Tables	1914
Table Partitions	56682
Table Sub Partitions	19394
Index	2586
Triggers	25
Total Data Volumes	4.5 TB
View	5918
Procedure	206
Function	137

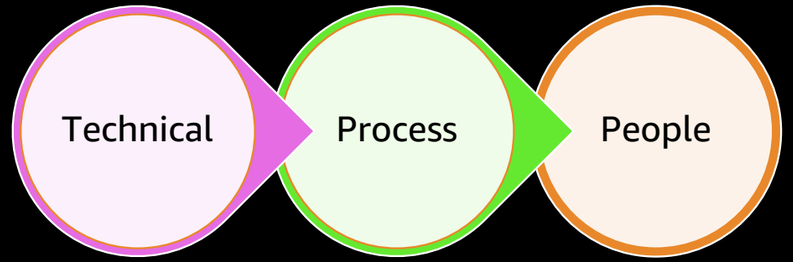
- 150 to 300 active connections to database with a peak of 650
- Peak Read IOPS is up to 16000
- Peak Write IOPS is up to 15000
- 192 MIP API users (last 90 days)
- 390k API calls (last 90 days)
- 95 Discovery users
- 1.2 million Discovery Calls (last 90 days)
- 120+ data sources
- ~ 17.9 billion rows

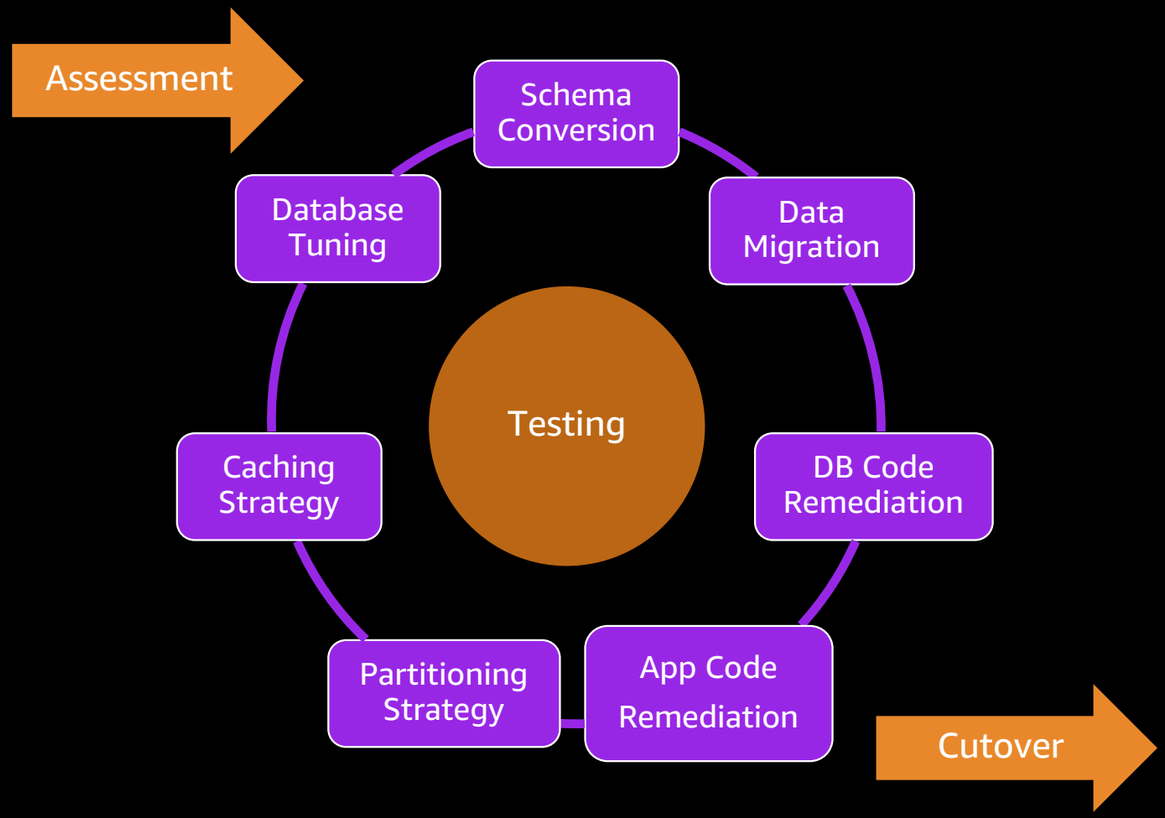
Migration Process

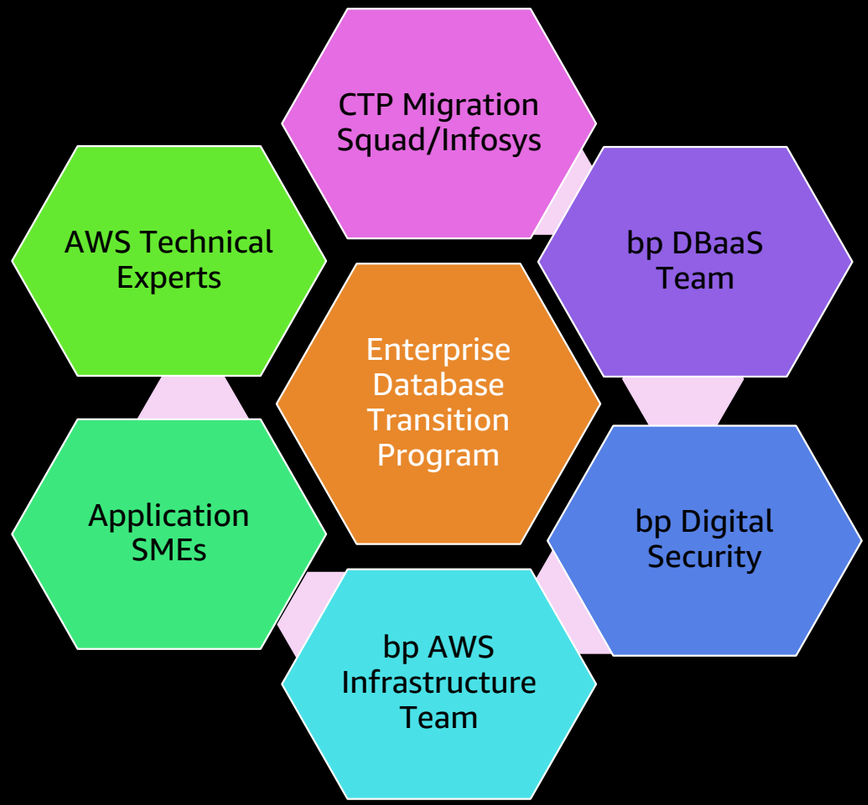


© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

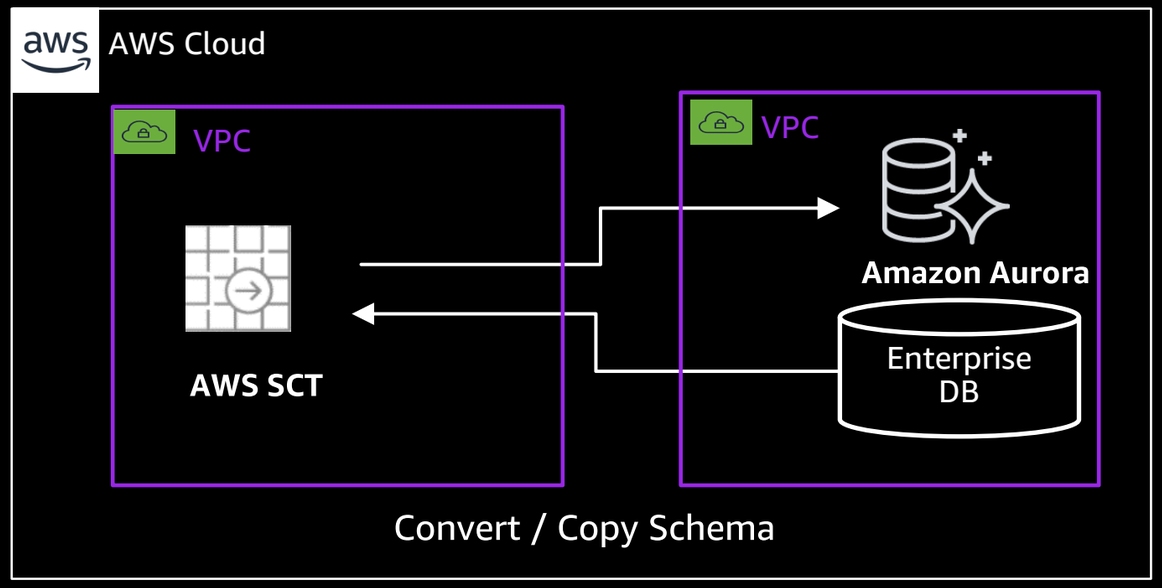
Migration Approach







Schema Conversion



Setup and Configure
AWS Schema Conversion Tool (AWS SCT)

Automate Migration Task

AWS SCT Migration Assessment Report

Schemas

Privileges

Stored Procedures

Tables

Functions

Triggers

Packages

Sequences

Partitions

Materialized Views

User Defined Types

Indexes

Assessment and Remediation

Key Assessment Findings

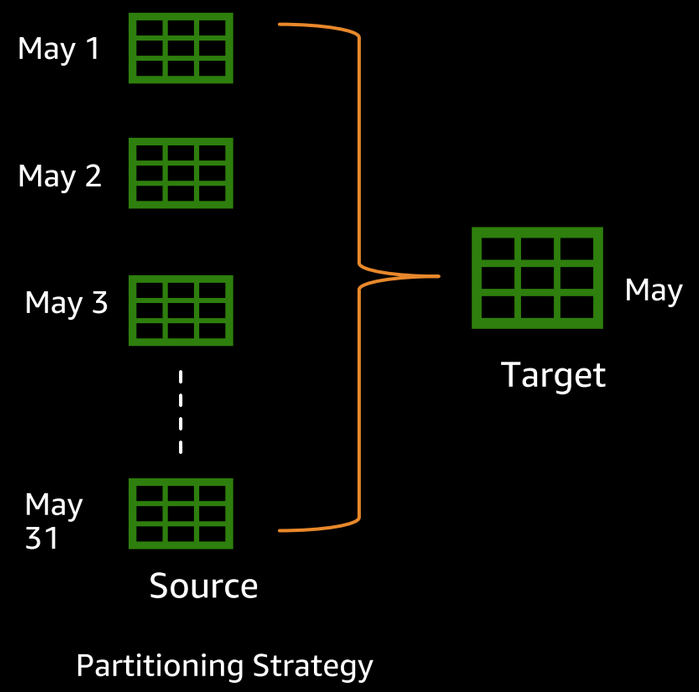
- ✓ Bitmap indexes
- ✓ Global temporary tables
- ✓ Partitions
- ✓ Dynamic SQL
- ✓ Materialized views
- ✓ Packages and procedures
- ✓ System functions

Database Code Remediation

- ✓ Revise partitioning strategy
- ✓ Bitmap indexes to B-Tree indexes
- ✓ Refresh strategy for materialized views
- ✓ Convert packages to procedures
- ✓ Auto vacuum
- ✓ Redefine security model

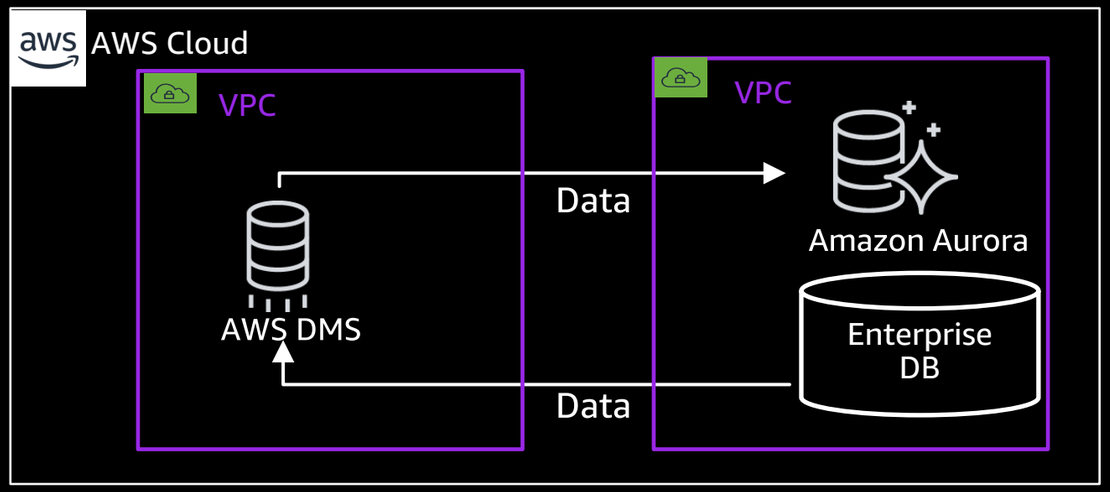
Application Code Remediation

- ✓ ETL driver
- ✓ Merge statement for Upsert
- ✓ Convert stored procedure to functions
- ✓ Dynamic queries



Data Migration

- ✓ Setup and Configure AWS Database Migration Service (AWS DMS)
- ✓ Full Load Transfer
- ✓ Change Data Capture
- ✓ Steady State

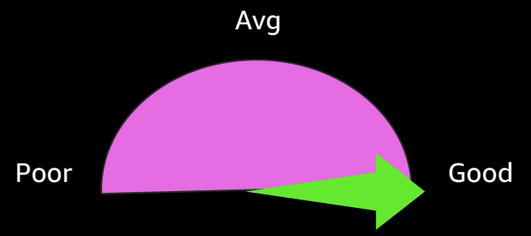


mipdw-monthly-forward-series-2005-12-31	Load complete	100%	Full load	mipspostgres	mipspostgres	mipsreplication	9/11/2020, 11:06:08 PM GMT-0500	11 h 58 m
mipdw-monthly-forward-series-2006-01-01-2006-12-31	Load complete	100%	Full load	mipspostgres	mipspostgres	mipsreplication	9/11/2020, 11:06:08 PM GMT-0500	6 h 34 m
mipdw-monthly-forward-series-2007-01-01-2007-12-31	Load complete	100%	Full load	mipspostgres	mipspostgres	mipsreplication	9/11/2020, 11:06:09 PM GMT-0500	8 h 21 m
mipdw-monthly-forward-series-2008-01-01-2008-12-31	Load complete	100%	Full load	mipspostgres	mipspostgres	mipsreplication	9/11/2020, 11:06:09 PM GMT-0500	16 h 11 m
mipdw-monthly-forward-series-2009-01-01-2009-12-31	Load complete	100%	Full load	mipspostgres	mipspostgres	mipsreplication	9/11/2020, 11:06:09 PM GMT-0500	17 h 5 m
mipdw-monthly-forward-series-2010-01-01-2010-12-31	Load complete	100%	Full load	mipspostgres	mipspostgres	mipsreplication	9/11/2020, 11:06:09 PM GMT-0500	19 h 21 m
mipdw-monthly-forward-series-2011-01-01-2011-12-31	Load complete	100%	Full load	mipspostgres	mipspostgres	mipsreplication	9/11/2020, 11:06:09 PM GMT-0500	19 h 49 m
mipdw-monthly-forward-series-2012-01-01-2012-06-30	Load complete	100%	Full load	mipspostgres	mipspostgres	mipsreplication	9/11/2020, 11:06:09 PM GMT-0500	12 h 11 m
mipdw-monthly-forward-series-2012-07-01-2012-12-31	Load complete	100%	Full load	mipspostgres	mipspostgres	mipsreplication	9/11/2020, 11:06:10 PM GMT-0500	13 h 30 m

Performance Measurement

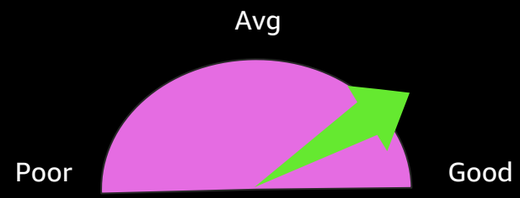
Ingestion Performance

Factors (Rows/Second)	Decision Criteria			Enterprise DB (Large Load)	PostgreSQL (Large Load)
	Negative	Neutral	Positive		
Read Throughput			✓	108	73
Write Throughput			✓	115	81
Update Throughput		✓		20	20
Delete Throughput			✓	-	-
WF Execution Time			✓	12 mins 9 secs	11 mins 44 secs



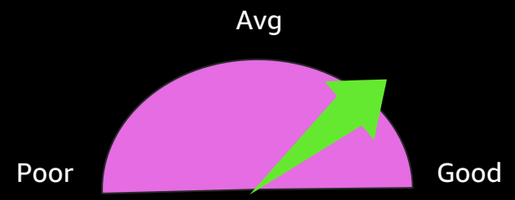
Query Performance

Factors (Query exec in Seconds)	Decision Criteria			PostgreSQL (Large RS) in %
	Negative	Neutral	Positive	
Operation Data			✓	35
Partitioned Data			✓	25
Non-Partitioned Data			✓	0
Aggregations		✓		200

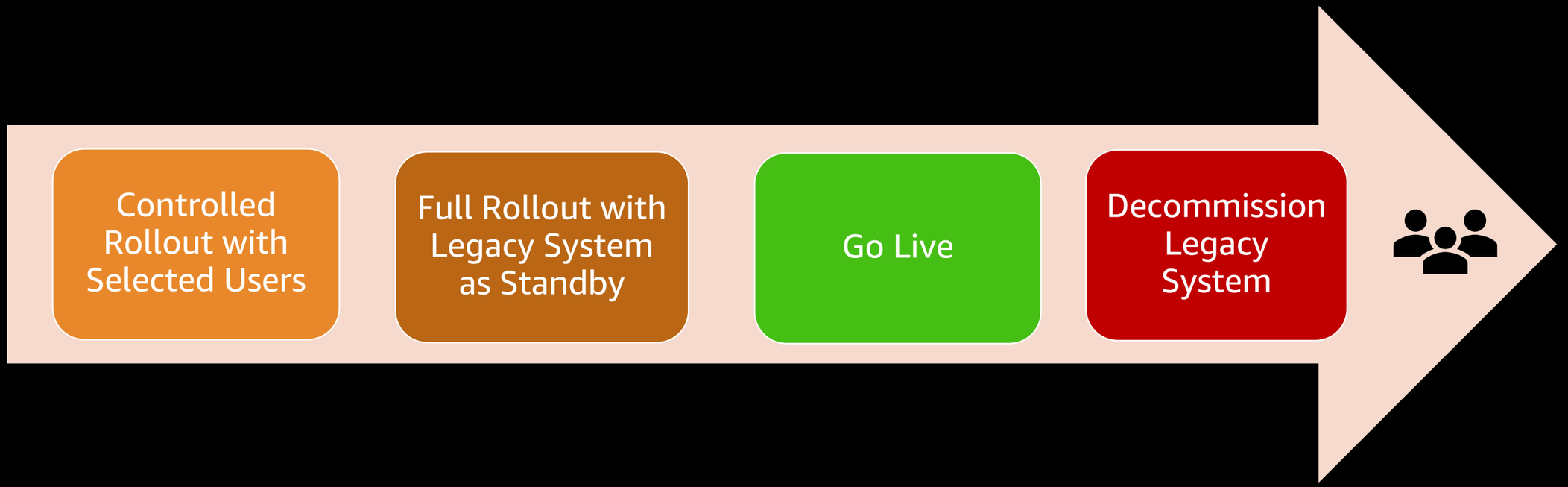


Application Performance

Factors (Response time in Seconds for 80% of Queries)	Decision Criteria			PostgreSQL (Large RS) in %
	Negative	Neutral	Positive	
Discovery			✓	0
MIP API			✓	20



Production Cutover





Engage with Stakeholders



Identify Support Staff for Cutover



Prepare Migration Checklist and Acceptance Criteria



Create RFC for Migration and Cutover



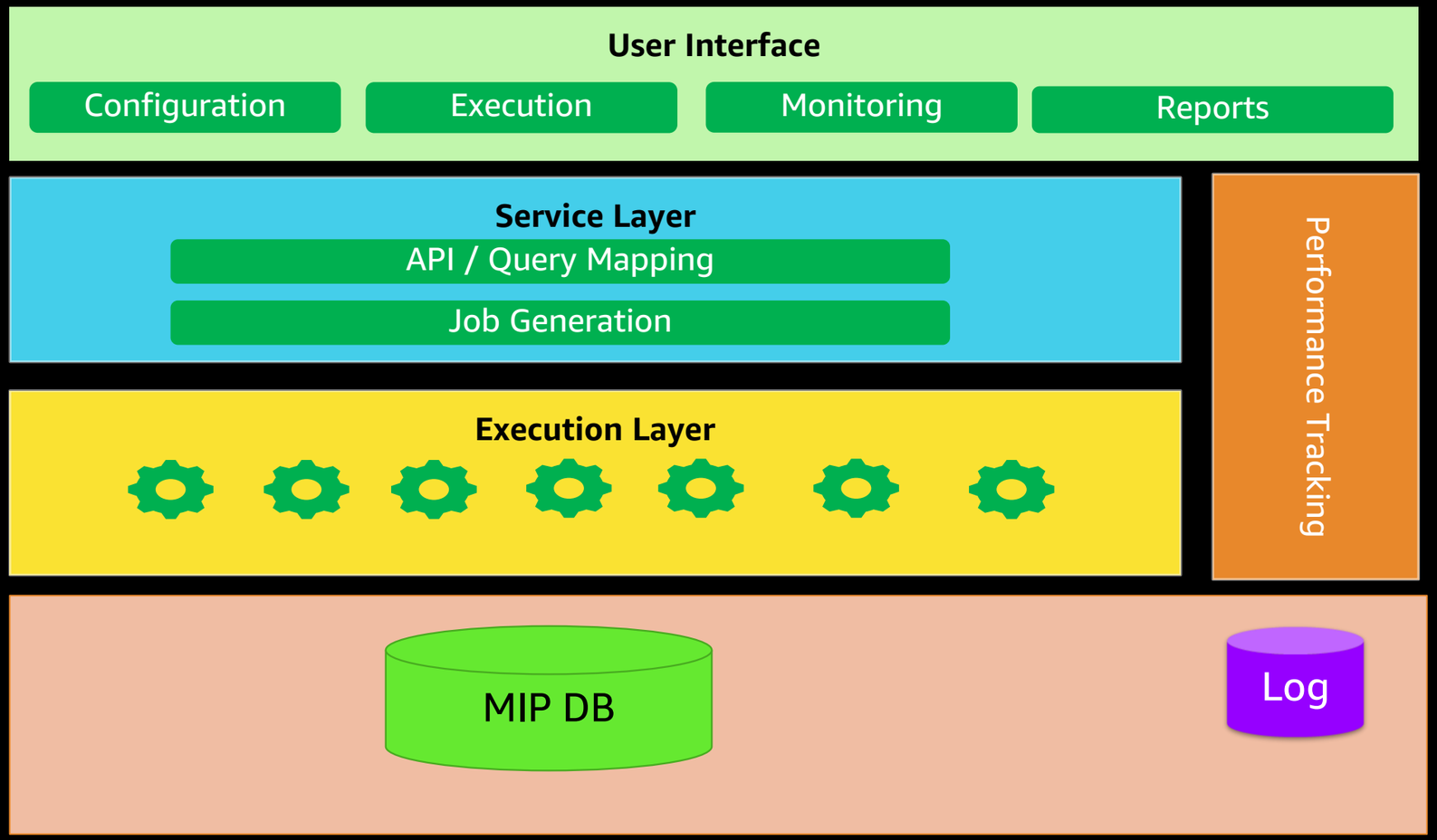
Submit Decommissioning Request

Key Tools and Tech



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

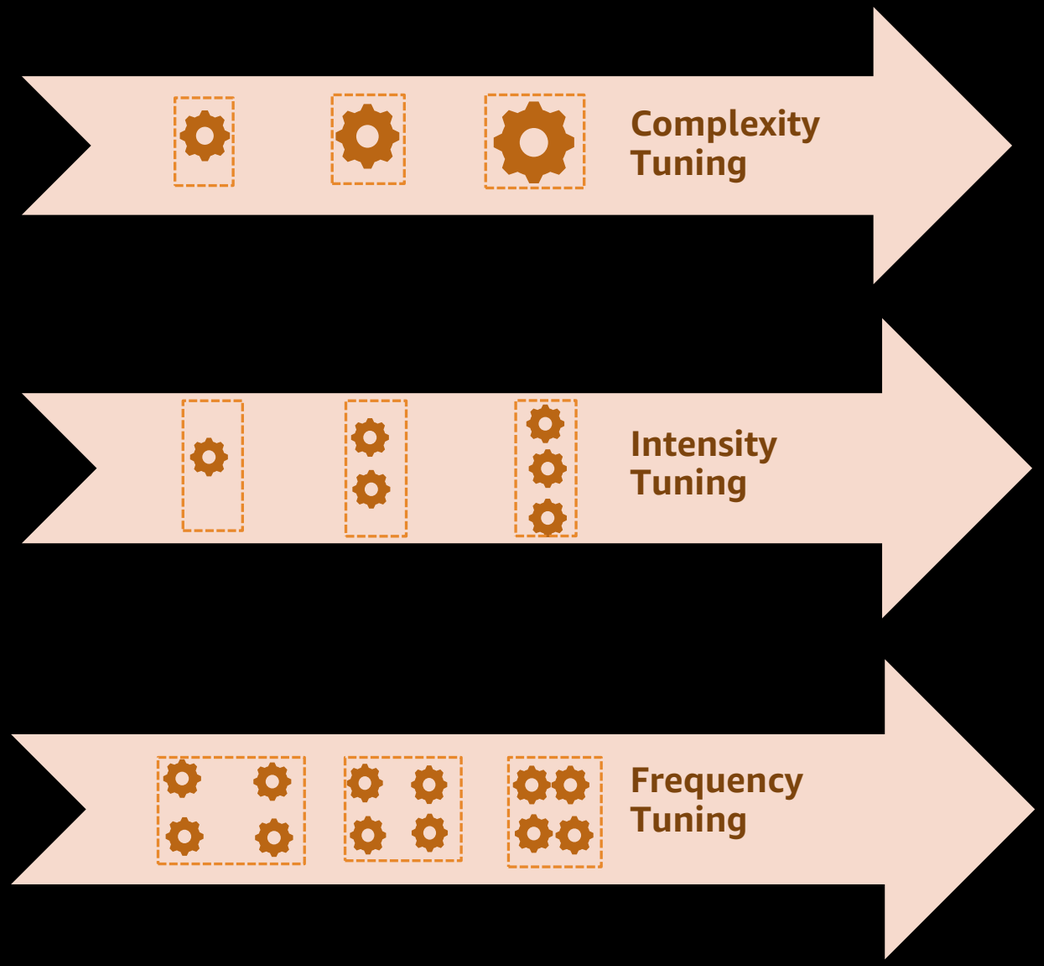
Automatic Load Simulator (ALS)



ALS Architecture

- An in-house tool
- Generates realistic loads on MIP
- Ensures basic functional assurance
- Ensures iso-performance characteristics of new system

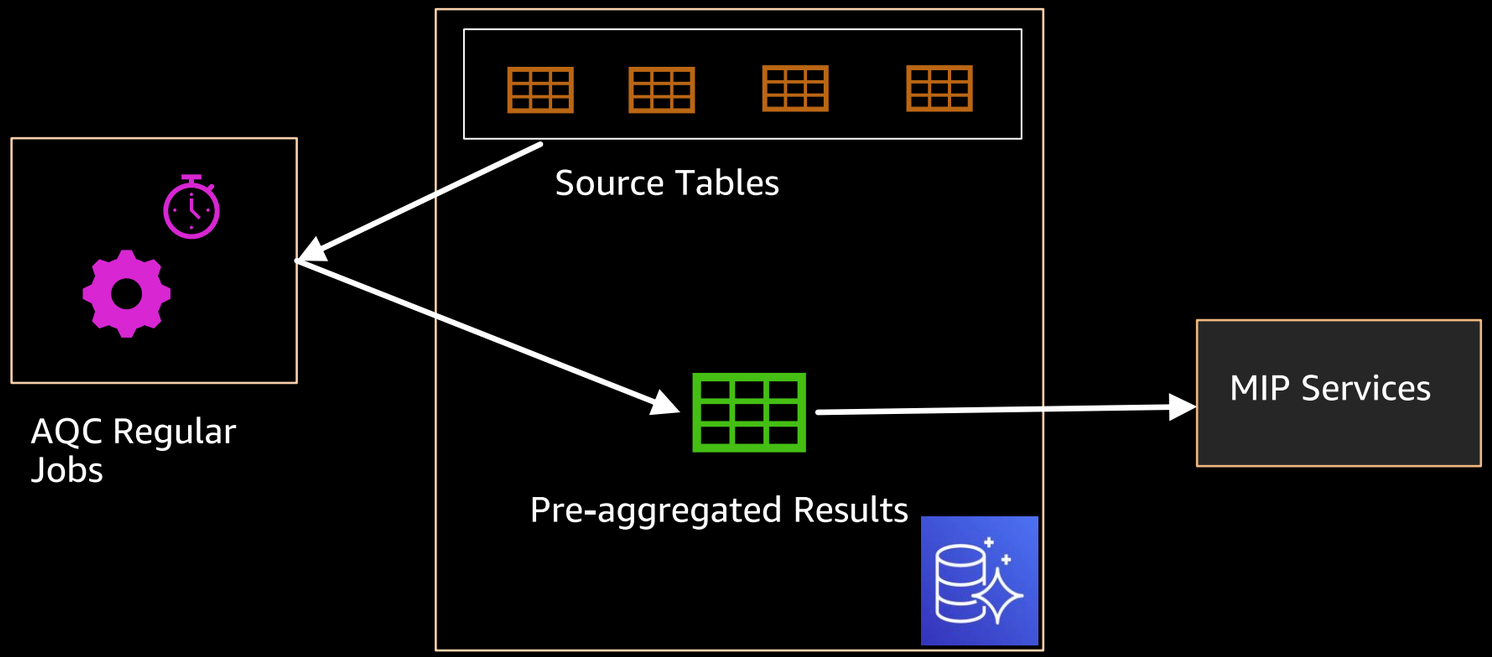
Load Simulation Approaches



Key Simulation Parameters

- System Load Factor (SLF)
  - Isoload: SLF = 100%
  - Overload: SLF > 100%
  - Underload: SLF < 100%
- Intensity Bundling Window (IBW)
  - Low value → High resolution, intensity focus
  - High value → Low resolution, frequency focus

Aggregate Query Caching (AQC)



- Workaround for PostgreSQL performance shortcomings over large aggregate queries
- Pre-aggregated results stored in separate tables
- Fast read access for aggregated results
- Tunable regular jobs to keep source tables and results in sync

Best Practices



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Key Practices for Success



Firm-wide
EDT Governance



Dual Assurance
Framework



Most Challenging
Project First



Decentralized
Empowering and
Accountability



Centralized
Knowledge Base

Coordinated
Collaboration with
AWS Team

Thank you!



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.