

AWS re:Invent

NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

How Capital One Accelerates Innovation with AWS Databases

Ryan Thurston

Principal Go-To-Market
Specialist NoSQL Databases
Amazon Web Services

Geetha Gopal

VP of Software Engineering,
Head of Cloud & Productivity Engineering
US Card, Capital One

Preston Burroughs

Sr. Director of Database and
Software Engineering
Enterprise Data, Capital One

Meet our speakers



Ryan Thurston

Principal Go-To-Market Specialist
NoSQL Databases
Amazon Web Services



Geetha Gopal

VP of Software Engineering,
Head of Cloud & Productivity Engineering
US Card, Capital One



Preston Burroughs

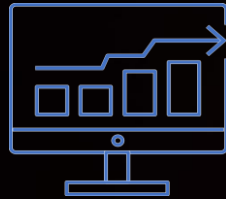
Sr. Director of Database and
Software Engineering
Enterprise Data, Capital One

Macro trends in industry fueling cloud adoption



Retail

Moving to online sales channels, accelerating retail adoption



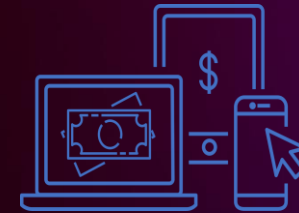
Financial Services

Closing physical stores and going digital



Insurance

Digital channels are now the norm, and usage-based models proliferate



Mobile

Rise in contactless payments; buy now, pay later; and instant payments

Benefits of modernization

Business benefits

Improve customer experience



Increase the efficiency of developers



Increase agility



Improve ROI and reduce TCO



Modern app characteristics

Scales to millions of users

Global availability

Responds in milliseconds

Handles petabytes of data

Modern data architecture

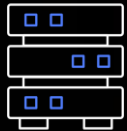
From traditional . . . tightly coupled

. . . to microservices, decoupled architectures



Web servers

Presentation layers
(Web)



Application servers

Business logic



Database servers

Data layer



Presentation

Web, Mobile, Desktop



Events



Events



Business logic

AWS Lambda, EKS, ECS

Queues + Caches + Messages



Aurora



DyanmoDB

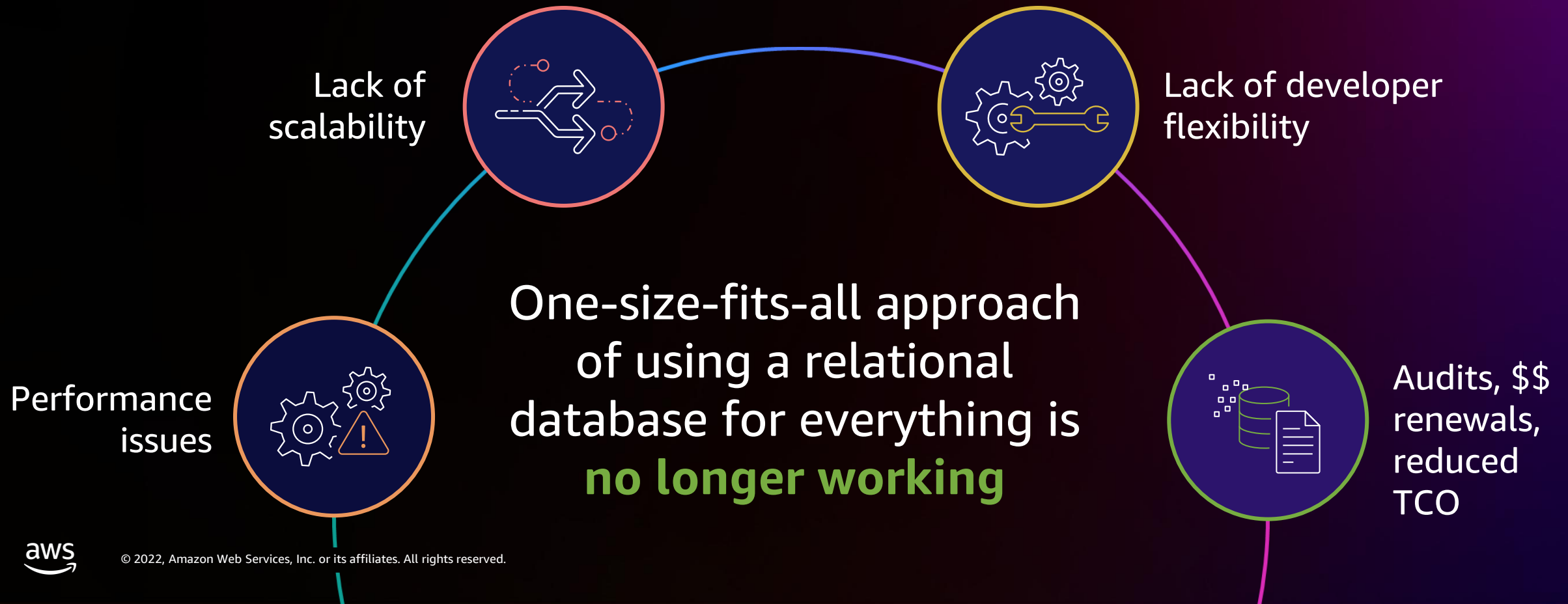


Graph

Data layer

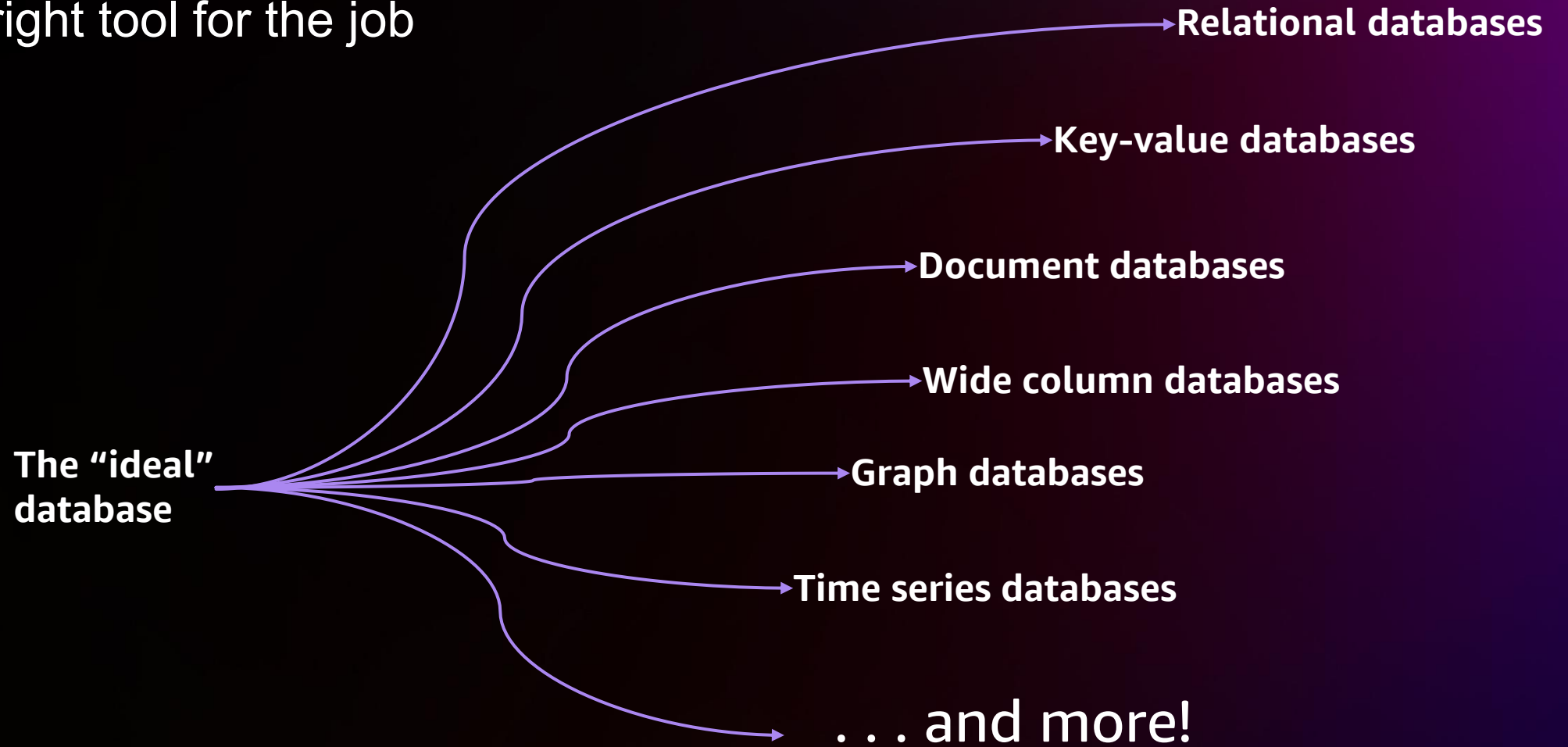
Purpose Built

Developers want the **right database** to meet their application's unique requirements



Purpose-built databases

Use the right tool for the job



AWS delivers the right database for the right job



Relational

Referential integrity, ACID transactions, schema-on-write



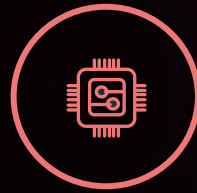
Key-value

High throughput, Low latency reads and writes, endless scale



Document

MongoDB compatible, flexible schema, ad-hoc queries & indexing



In-memory

100% open source Redis compatible, microsecond latency



Graph

Quickly and easily create and navigate relationships between data



Time-series

Collect, store, and process data sequenced by time



Blockchain

Execute transactions without the need for trusted, central authority



Wide Column

Scalable, highly available, and managed Apache Cassandra-compatible service

AWS
Service(s)



DynamoDB



DocumentDB



ElastiCache



Neptune



Timestream



AMB



Keyspaces
Managed Cassandra

Common
Use Cases

Oracle, SQL Server,
PostgreSQL,
MySQL, MariaDB

Single
millisecond
performance at
any scale, query
patterns known
up front

Content
management,
personalization,
mobile

Caching,
Leaderboards,
real-time
analytics

Fraud detection,
social
networking,
recommendation
engine

IoT, DevOps,
Clickstream

Supply chain,
Immutable
records

Build low-latency
applications,
leverage open
source, migrate
Cassandra to the
cloud

Capital One at a glance

Fortune 100 company, founded in 1994

Top 10 US bank, serving over
100M customers

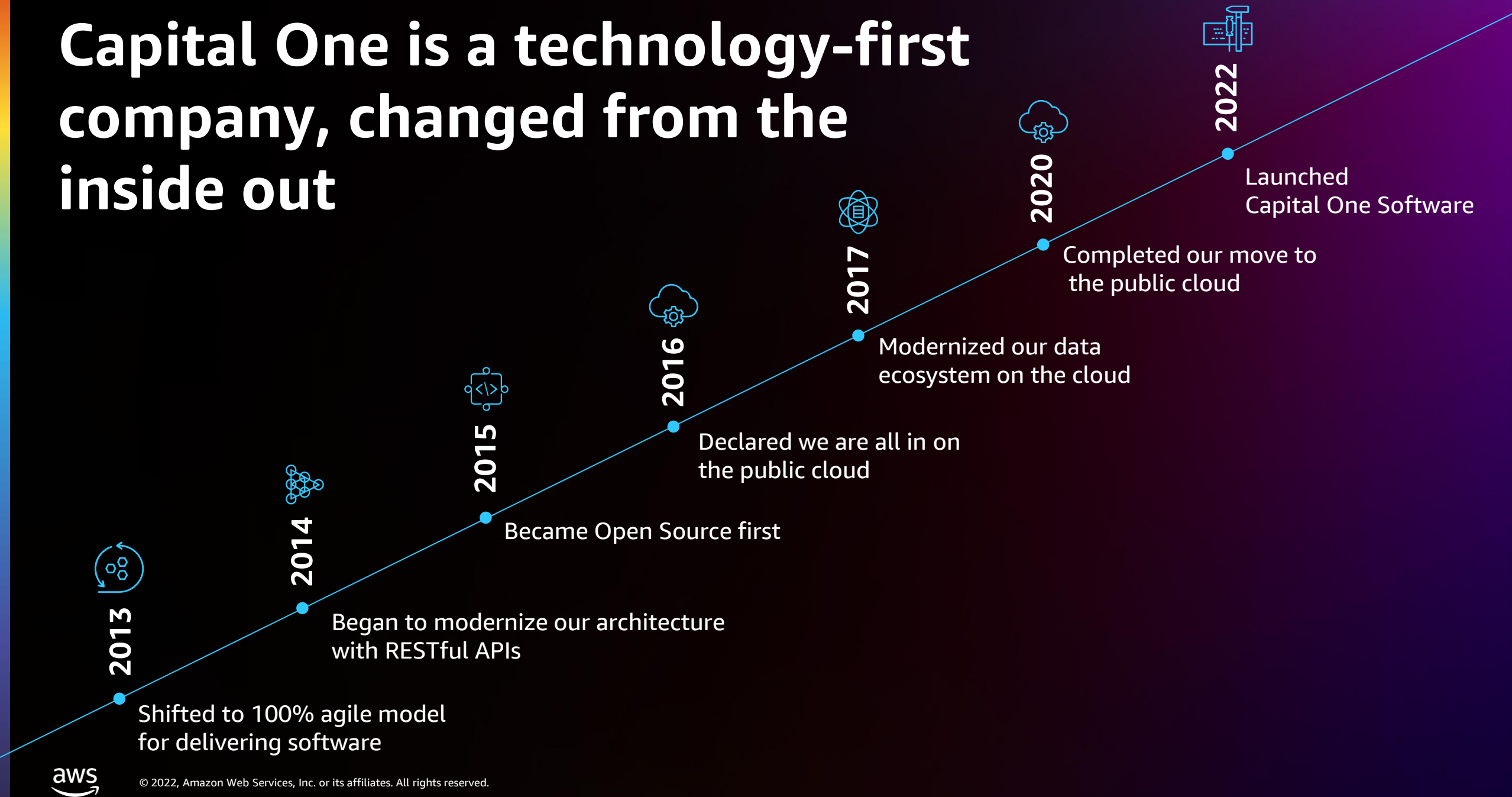
Data and technology pioneer in the
financial services industry



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



Capital One is a technology-first company, changed from the inside out



Agenda

Why we adopted purpose-built databases

Key steps we took and decisions we made to accelerate adoption

Use cases - Document DB, Aurora Global, Dynamodb

Migration guidance

Conclusion



Thriving on the cloud

We modernized and migrated thousands of workloads

We use modern development practices

We harness innovative cloud services while adhering to regulatory requirements

Our journey is not over – cloud-native, real-time, intelligent

Purpose-built databases have been a big part of our innovation journey



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



Database management is time consuming

Database management requires specialized skill and knowledge across the DB lifecycle

Database infrastructure

Database maintenance

Database monitoring

Reduce
variability

Standardize
patterns

Automate
lifecycle tasks

We have created paved paths for purpose-built databases that meet majority of our business needs

Database Type	Paved Paths
Relational	Aurora Global, RDS Aurora
Key Value	DynamoDB
Document	DocumentDB
Graph	Neptune
Wide Column	Key Spaces*
Time Series	Timestream*

* Evaluating



Our Database Selection Tool guides teams with picking the right database and pattern for their use case

Key considerations

Data model

Data access patterns

Query type

Resiliency requirements

Scalability requirements

Cost

Automate Lifecycle tasks

Cross-Region auto-failover

- Automate regional failure detection
- Trigger alert to engineering teams with one-click auto-failover option

Automated governance

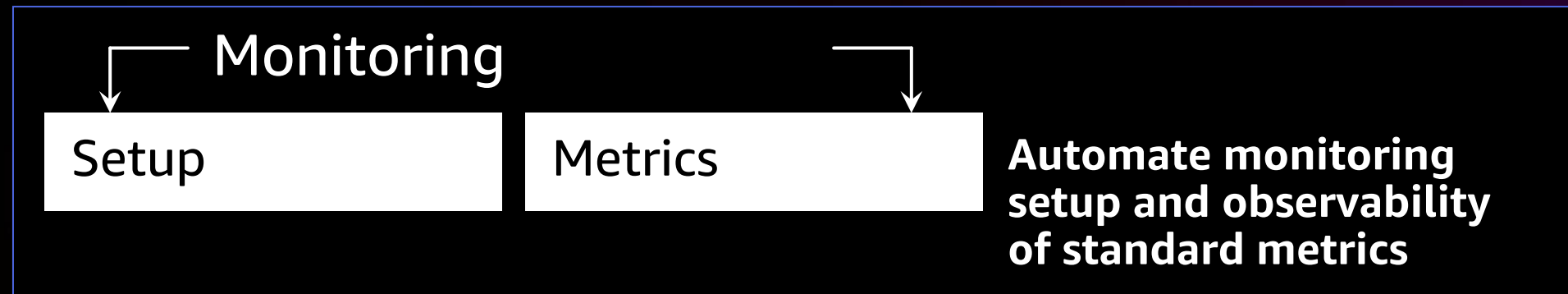
- Access provisioning
- Security and compliance checks/remediation

Version upgrades

- Automate near-zero downtime version upgrades

Automate resiliency testing

- Hands-free resiliency testing



DocumentDB



Business use case - Credit decisioning

- Need for near real-time credit product decisioning
- Data from several external data sources form user profiles
- The data structure of these user profiles vary and can result in large documents
- User profiles stored, converted to features, fed to decisioning process
- It is critical to achieve ms read and write latencies to make quick decisions

Our requirements

- Support large documents and support growth in document size
- Millisecond read and write latencies < 100 ms reads, < 20 ms writes
- Fully managed in Capital One VPC, application to DB TLS connection
- Cross-region auto-failover capability with less than 2 Mins RTO and RPO
- Minimal to no application code change (Mongo compatible)
- No downtime migration
- Functional capability - Text search, native s3-archival
- Storage growth to 128 TB

Amazon DocumentDB met most of our needs



Fully
managed

- Built-in high availability
- Backups enabled by default
- Durable by default
- Security best practices by default
- Automatic patching
- Monitoring and alerting
- Global Replication



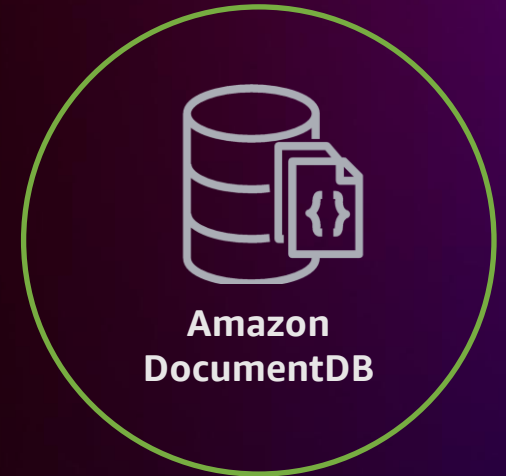
Scalable

- Scale compute in minutes
- Storage and IO autoscaling
- Storage scales
- Scale out replicas for millions of reads

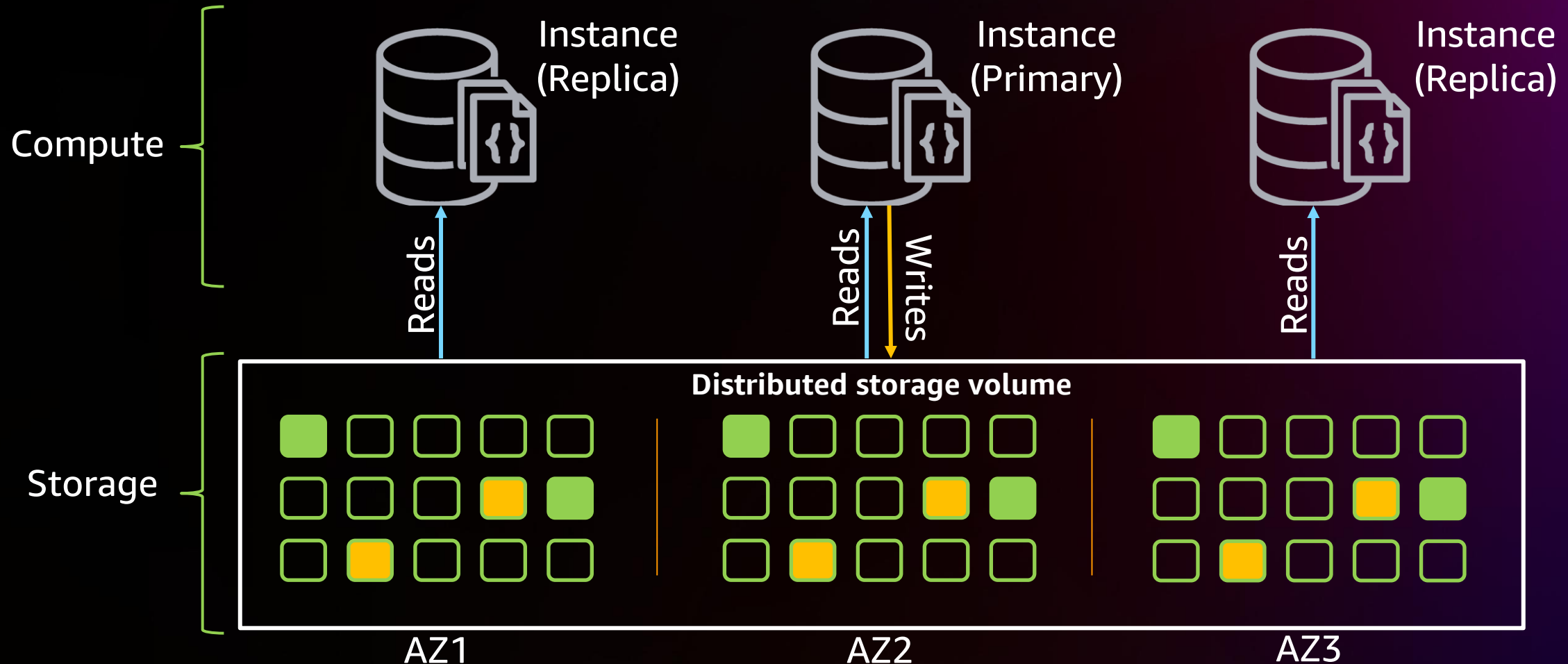


MongoDB API
compatible

- Applications, drivers, and tools can be used with Amazon DocumentDB with little or no change
- Supports hundreds of APIs, operators, and stages
- Continually working backward from customers to deliver the capabilities they need



Amazon DocumentDB: Built-for-the-cloud architecture



Amazon DocumentDB – Global clusters

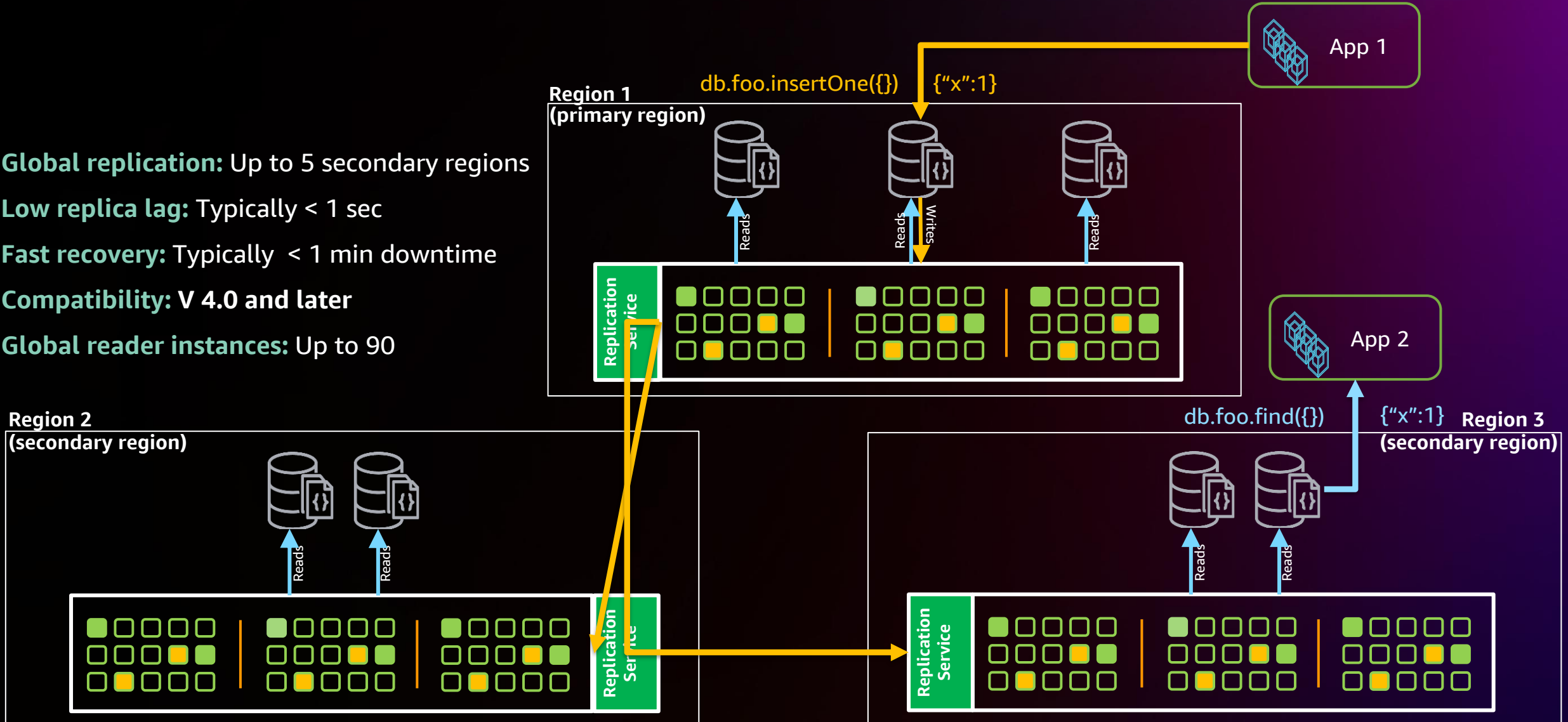
Global replication: Up to 5 secondary regions

Low replica lag: Typically < 1 sec

Fast recovery: Typically < 1 min downtime

Compatibility: V 4.0 and later

Global reader instances: Up to 90



What we did

- We had a nearly 2-month intense DocumentDB engagement starting with immersion day
- Two extensive Proof of Concepts (POCs) with the real-world use case
 - Read/write standard size documents yielding 5x throughput
 - Read/write large documents yielding 5x throughput
- The engagement ended with decisions on what capabilities Capital One would build versus AWS roadmap

Role	Prepare		Execute POC1		Execute POC2	
	September		October			
Week Ending	28-Oct	4-Nov	11-Nov	18-Nov	25-Nov	2-Dec
Dev	Compatibility & Sizing Assessment					
		Application changes & query workarounds, as needed	App to cluster connection established			
	DocumentDB Immersion Day					
			Establish connection from Application to DocumentDB	Wrap up POC1		
			Execute load testing script			
			Validate against defined success criteria			
					Insert large documents to two identified collections	Wrap up POC2
			Cluster Provisioned		Validate against defined success criteria	
DBA	Create Cluster and add users per RBAC					
	Architecture and Security Group Approval					
			Set up Monitoring & Alarms / configure CloudWatch dashboard			
					POC Support	

Amazon DocumentDB – Concerns & mitigation

Criteria	Result
Lack of storage compression Use case requires large amount of storage per year	Short term: non-native archival Long term: Sharding
Downtime with maintenance operations Can take up to 30 seconds	Short term: Non-critical - subscribe to maintenance events Critical - DMS for upgrades Long term: Proxy solution will minimize downtime (Roadmap)
Failover time for regional failovers. DB failover is not fully automated and can take several minutes	Short term: Automate global failover process using AWS CLI / SDK Long term: Managed failover (Roadmap)

Aurora Global Database

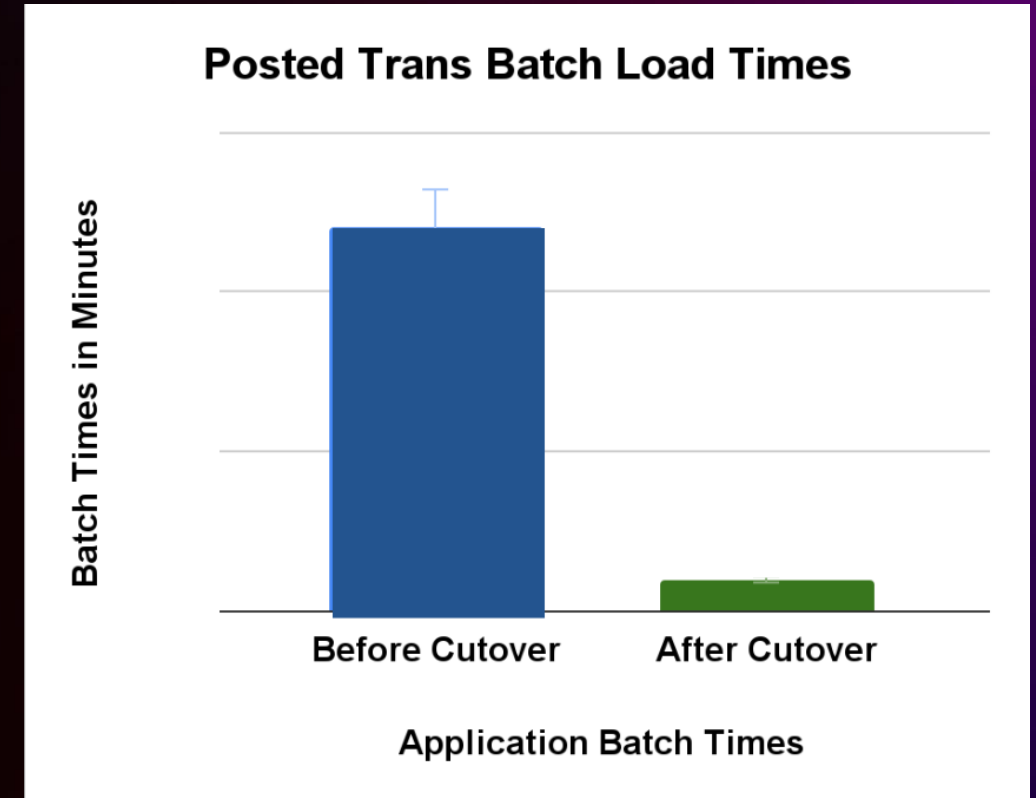


Business use case - Customer transactions

- Tracking customer transactions - purchases, reversal, rewards, payments, are critical to providing great spend experience
- Relational, batch data that tracks several parameters for a transaction such as customer, merchant, payment network and so on
- Database is a third party postgres database, provisioned on EC2 instances
- Team spent several hours in database maintenance

We migrated customer transactions database to Aurora Global

- DMS was used to migrate from third party database to Aurora Global
- We developed best practices for relational databases along the way
- We automated life-cycle tasks including developed regional auto-failover capabilities



10x improvement in Batch load times - Database optimization, Aurora performance

Business use case - Payments

- Our payment orchestration service interacts externally with vendor services
- Due to rigid vendor timeout configuration, incomplete transactions can result in inefficient post-remediation processes
- Need for the payment service orchestration system to be highly resilient and configured to near zero data loss to enable transaction retries
- Need a transactional data store to achieve 4 seconds SLA during system issues

We fronted Aurora Global with a Cache layer to meet our business needs

- The solution establishes a new pattern for high resiliency databases within Capital One and creates an “always on” data store by adding a Redis caching layer in front of the data layer
- In the event of a database failure or AWS regional outage, the data layer will perform automatic regional failover while the cache layer serves traffic
- Once the failover is complete, changes made in the cache is flushed back out to the database



ElastiCache for
Redis



Amazon Aurora

Dynamo DB

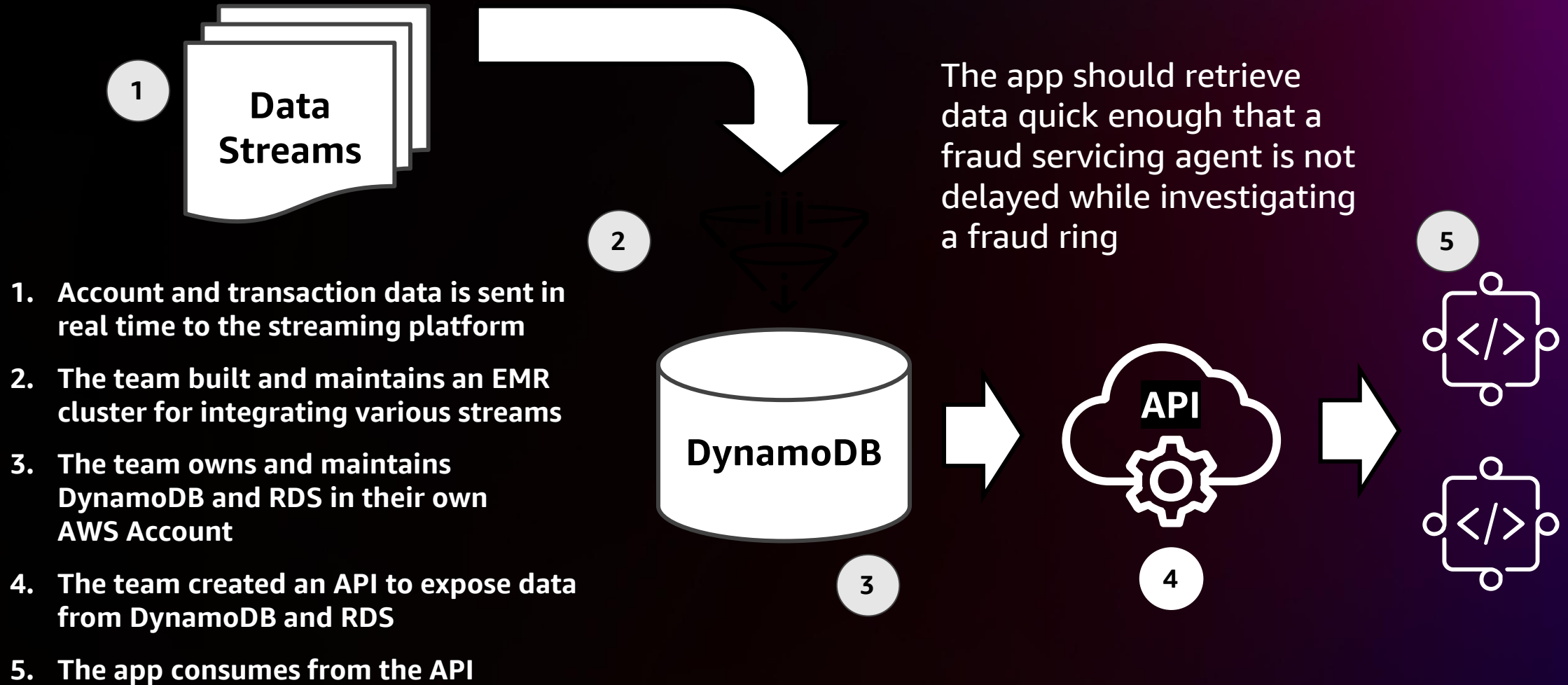


Use case

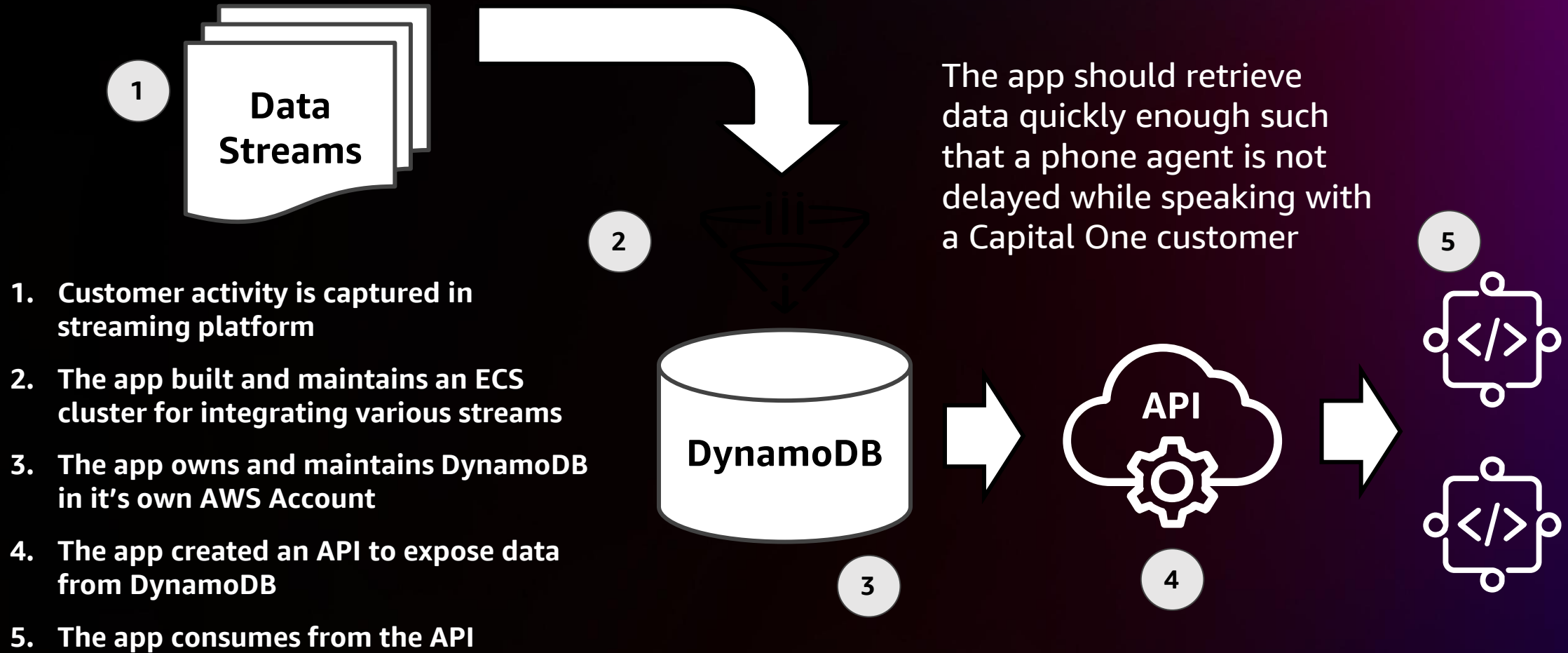
Functional need - Low Latency database used to enable real-time experiences.

- Used for “fast” writes and reads of data
 - **Time window:** Streaming
 - **Consistency:** Real time - Eventual
 - **Read Perf:** ~3-20ms
- Technologies that are highly optimized for specific query types

Use case 1 : Fraud



Use case 2 : Bank App for Associates



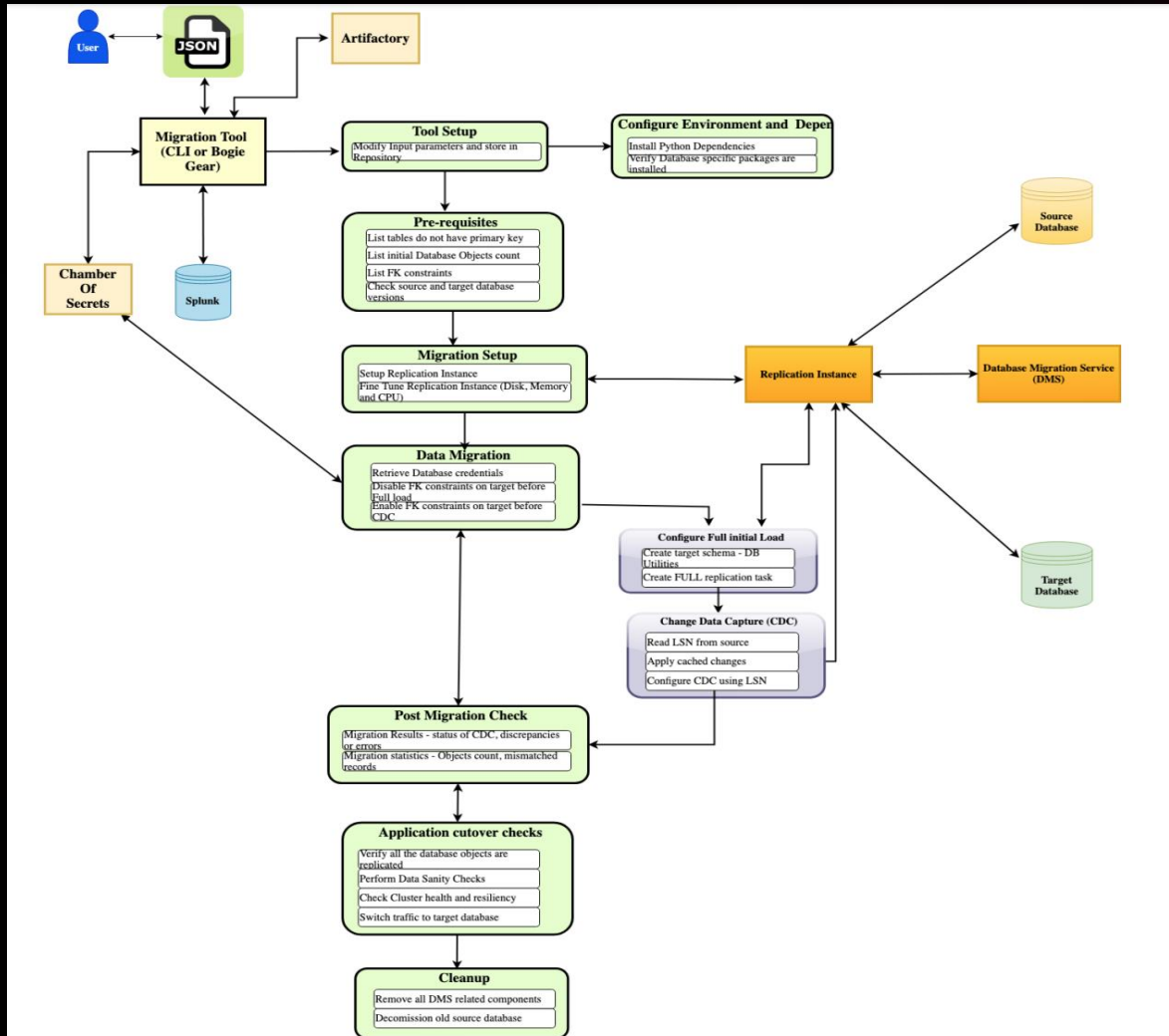
Data migrations



Summary

- There are multiple use cases for massive data movement within Capital One
- AWS DMS has limitations and pre-requisites that need to be met
- We developed an internal Database Migrator tool based on AWS DMS
- The Migrator tool is integrated into capital one tooling and will evolve as we work with AWS to influence their DMS roadmap

DB Migrator analyzes databases for migration readiness and helps migrate with near-zero downtime



- DB Migrator targets database migrations across a wide spectrum of scenarios with near-zero downtime requirements
- The tool guides users at every stage of migration - migration readiness, migration, validation and clean-up
- Creates DMS resources and applies config defaults based on automated analysis of database configuration and usage, thereby taking guesswork out of the picture

DB Migrator bridges several limitations in DMS and makes it easy to use DMS safely and efficiently

Capability	Description
Schema Objects	Schema objects are migrated as part of the process
Logging	Additional logging to enable end-to-end view of migration process
Constraints	Updates instance settings to allow constraints during replication
Replication Instance Sizing	Automatically sized based on the Database volume and usage
DB Utilities	Database version is checked and appropriate DB utilities are used
Data Validations	Data validation can be performed any time during the upgrade process
DMS Resources	Tool launches appropriate AWS resources to handle each stage of the migration process
Database Credentials	Database credentials are retrieved from internal vault during setup process
Resource Cleanup	DB Migrator performs cleanup after successful cutover

Our partnerships with AWS and Engineering Teams are the key to our success

Our database strategies had to evolve – purpose-built databases power our innovative experiences

We built software that helped us accelerate adoption

AWS prioritized features we needed to adopt



NoSQL programs/investments

Program Name	Description	Duration	Cost	Outcome
Cost Analysis	NoSQL sizing questionnaire using your workload metrics to generate a cluster sizing estimate	Self-serve	None	sizing and pricing estimate
Compatibility Assessment	The NoSQL compatibility tool will examine log files from MongoDB to determine if your applications use operators that are not supported in DocumentDB. This tool will report use of unsupported APIs, and output the unsupported log lines into a file for future use.	Self-serve	None	MongoDB -> DocumentDB compatibility report
Immersion Day	NoSQL Immersion Days provide customers with modular content and hands-on labs to learn about use cases, architecture, best practices, migrations, security, monitoring, and more.	1-2 days	None	Enable customers to build NoSQL POC, or implement/migrate workload
Well-Architected Lens	WAL assesses customer workloads with a focus on optimizing performance, reliability, security, cost and operational excellence.	Half day	None	Well Architected review report with recommendations to optimize workload.
Springboard	Customized migration game plan that begins with decision makers (sponsor) and diverse stakeholders. Includes 1. current state operations & business commitments 2. specific architectural recommendations on AWS 3. proposed milestones 4. follow up action items	Half day	None	Engagement readout that serves as execution plan.
Data Labs	Data Labs provide technical resources to help customer build tangible deliverables that accelerate data modernization initiatives. Customers who've built NoSQL POCs are given prescriptive architectural guidance, best practices, and technical roadblock removal by AWS Data Lab engineers and NoSQL experts. Customers leave the lab with a well-architected prototype, a path to production, and greater knowledge of NoSQL technologies.	4 days	None	Well architected prototype with path to production
Professional Services	AWS Pro Serve provides hands-on support to review customers' database environments and challenges, requirements, desired outcomes, and success metrics for migrating workloads to NoSQL databases. Suitable use cases are identified, and customer gets experience implementing a demo of their solution using sample data. A documented roadmap proposal with next steps to implement migrations to NoSQL is provided by end of the engagement.	Varies	SOW based	Hands-on support to guide customer from POC->assessment->migration

Thank you!

Visit Capital One at booth #2004



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



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