



aws **SUMMIT**

NEW YORK | JULY 12, 2022

DAT201

Modernize your applications with purpose-built AWS databases

Barry Morris

General Manager of Amazon DocumentDB, ElastiCache, Timestream
AWS



Agenda

Key attributes of modern database applications

Modern database architecture

Modernization at Amazon

The pursuit of the “ideal” database

Database modernization playbooks

Typical modernization journey with purpose-built databases

Key attributes of modern applications



Designed for

**Innovation
and
agility**



Without limits on

**Performance
and
scalability**



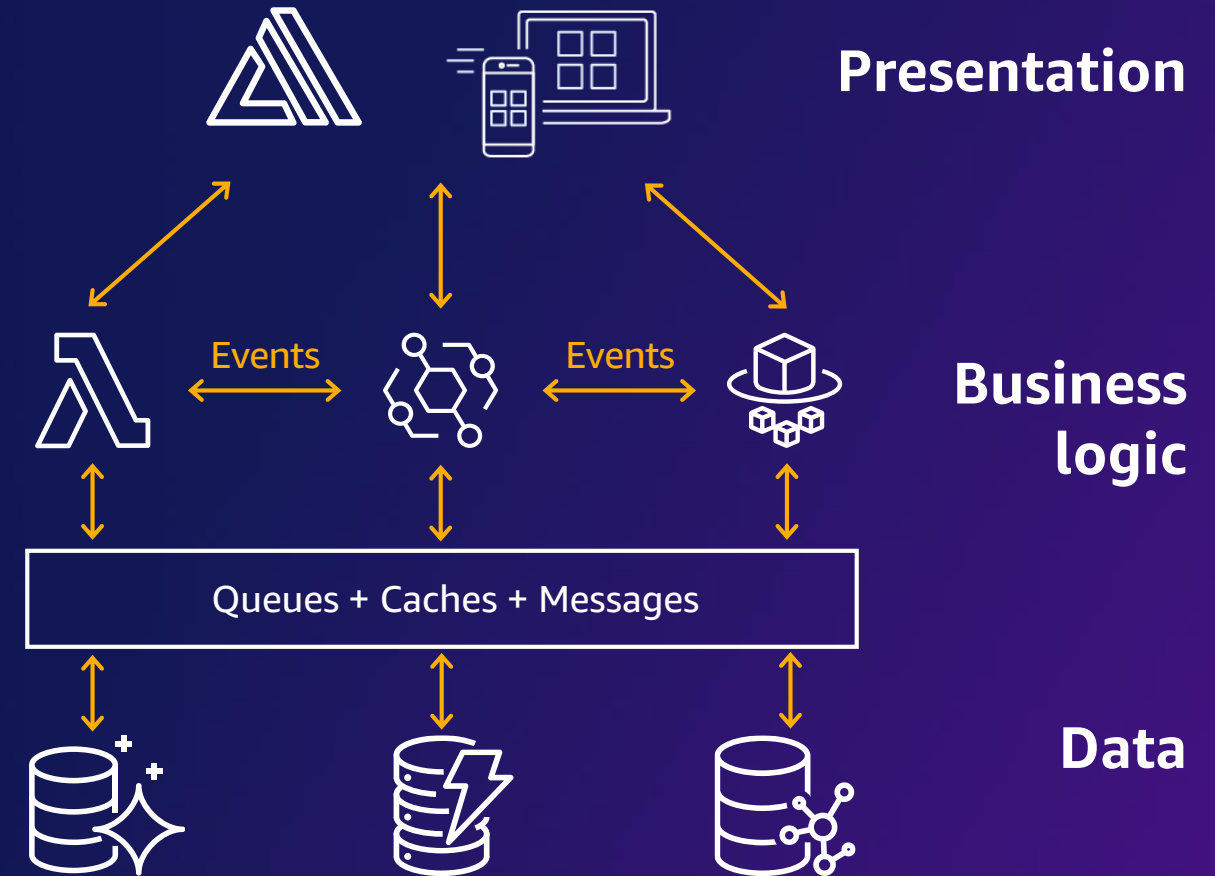
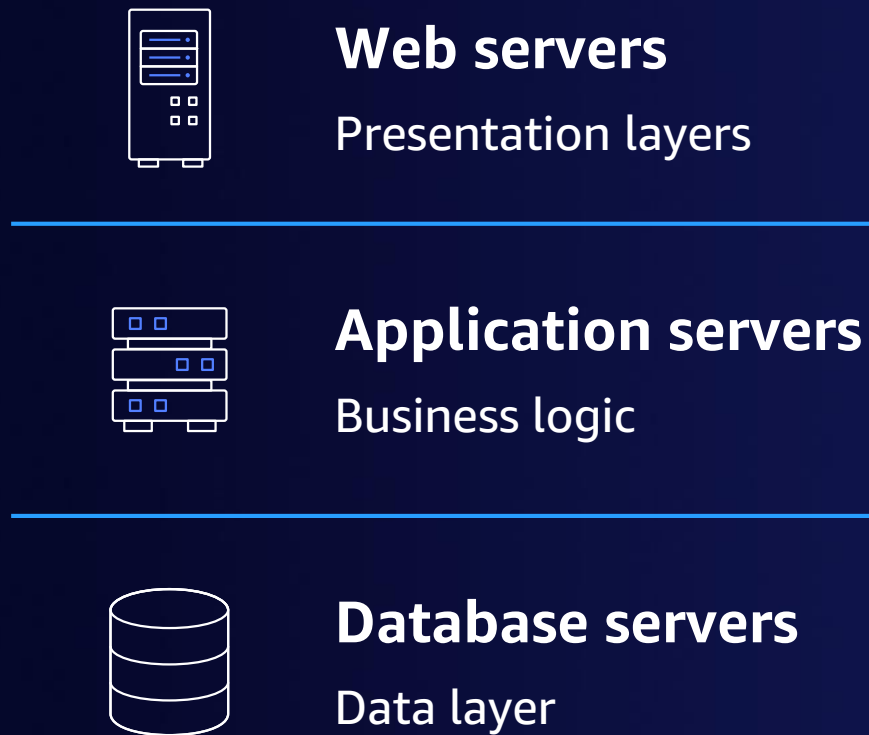
Is

**Highly available,
easily managed,
and
cost-effective**

Modern data architecture

From traditional . . .

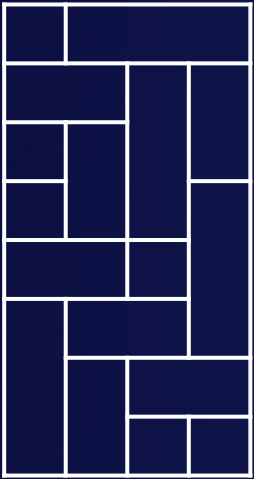
. . . to microservices, decoupled architectures



Modernization at Amazon

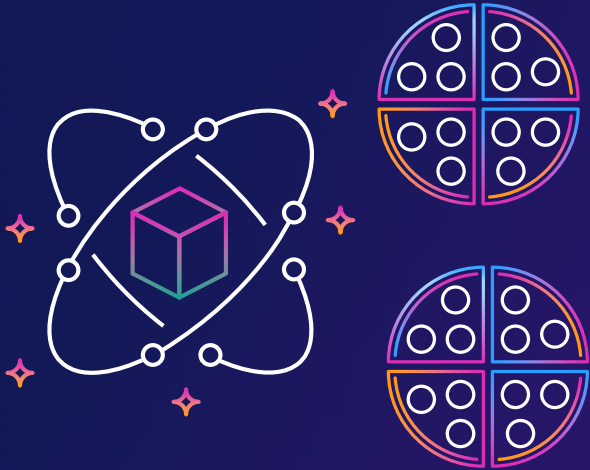
LESSON LEARNED: DECOMPOSE FOR INNOVATION AND AGILITY

Monolithic application



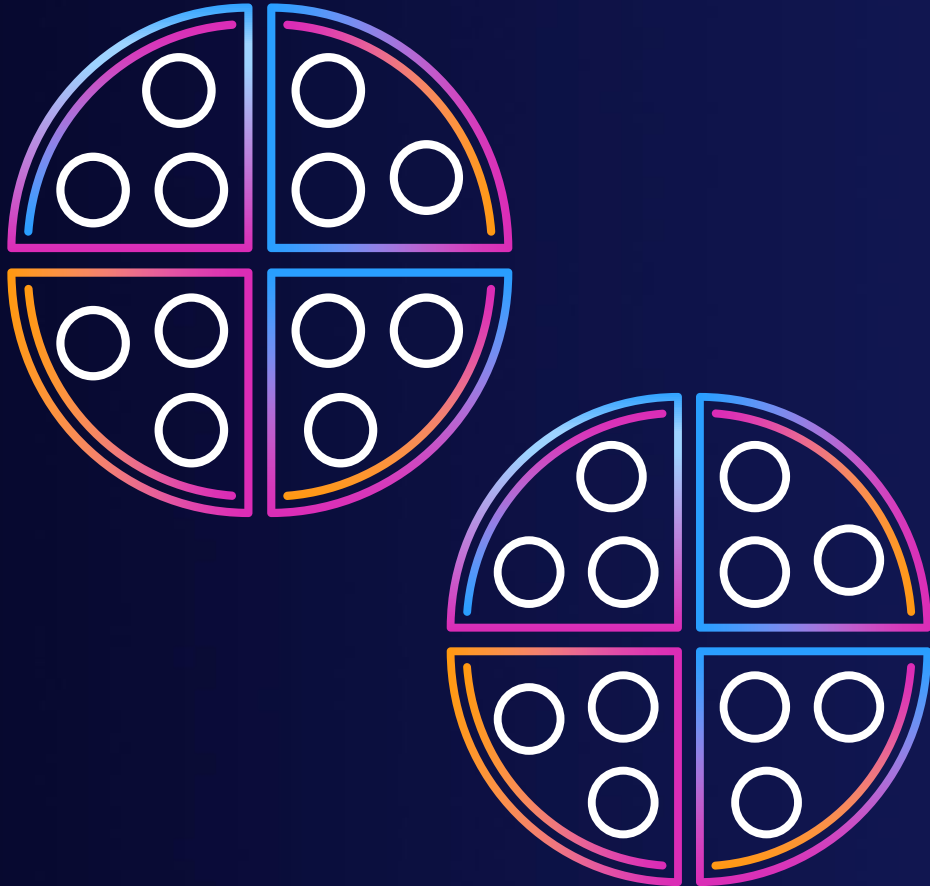
Central database
and dependent teams

Microservices



Independent databases
and two-pizza teams

Two-pizza teams are fast and agile!



Full ownership and autonomy

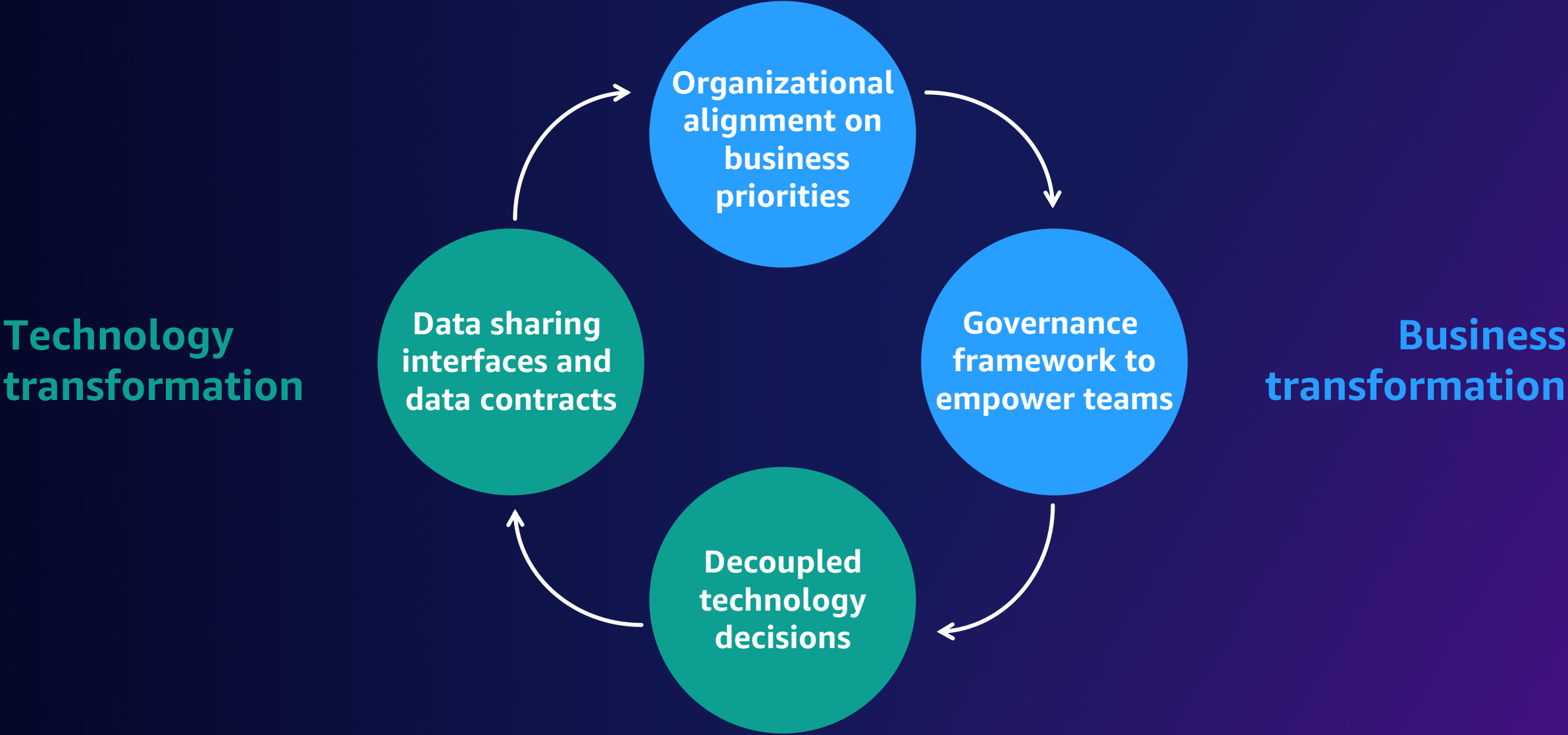
You build it, you run it

DevOps – small, nimble teams

Rapid innovation

Clean interfaces

The modernization virtuous cycle



The pursuit of the “ideal” database

THE IDEAL DATABASE SHOULD DO EVERYTHING WELL

Innovation and agility

Supports **all storage**
(data types) and
all access patterns

Implements **ACID**
transactions and
strong consistency

Performance and scalability

Provides **limitless**
scalability and tolerates
load variability

Is **highly performant**
regardless of query
complexity

Cost-effective and easy to manage

Is **continuously**
available (no downtime)

Is **simple and easy to**
use, tune, and maintain

Is **cost-effective**
and **predictable**

... and more!

The pursuit of the “ideal” database

THE IDEAL DATABASE SHOULD DO EVERYTHING WELL

Innovation and agility

Supports **all storage**
(data types) and
all access patterns

Implements **ACID**
transactions and
strong consistency

Performance and scalability

Provides **horizontal**
scalability and tolerates
partial availability

Is **highly performant**
regardless of query
complexity

Cost-effective and easy to manage

Is **continuously**
available (no downtime)

Is **simple and easy to**
use, tune, and maintain

Is **cost-effective**
and **predictable**

This doesn't exist!

... and more!

Explosion of database solutions

ORACLE®

Microsoft
SQL Server



Amazon DynamoDB



Amazon RDS



Amazon Aurora



Apache
Cassandra

Apache
HBase



Amazon DocumentDB
(with MongoDB
compatibility)



Amazon Keyspaces
(for Apache Cassandra)



Amazon ElastiCache



Amazon MemoryDB
for Redis



CockroachDB



Memcached

... and more!



Amazon
Timestream



Amazon
Neptune

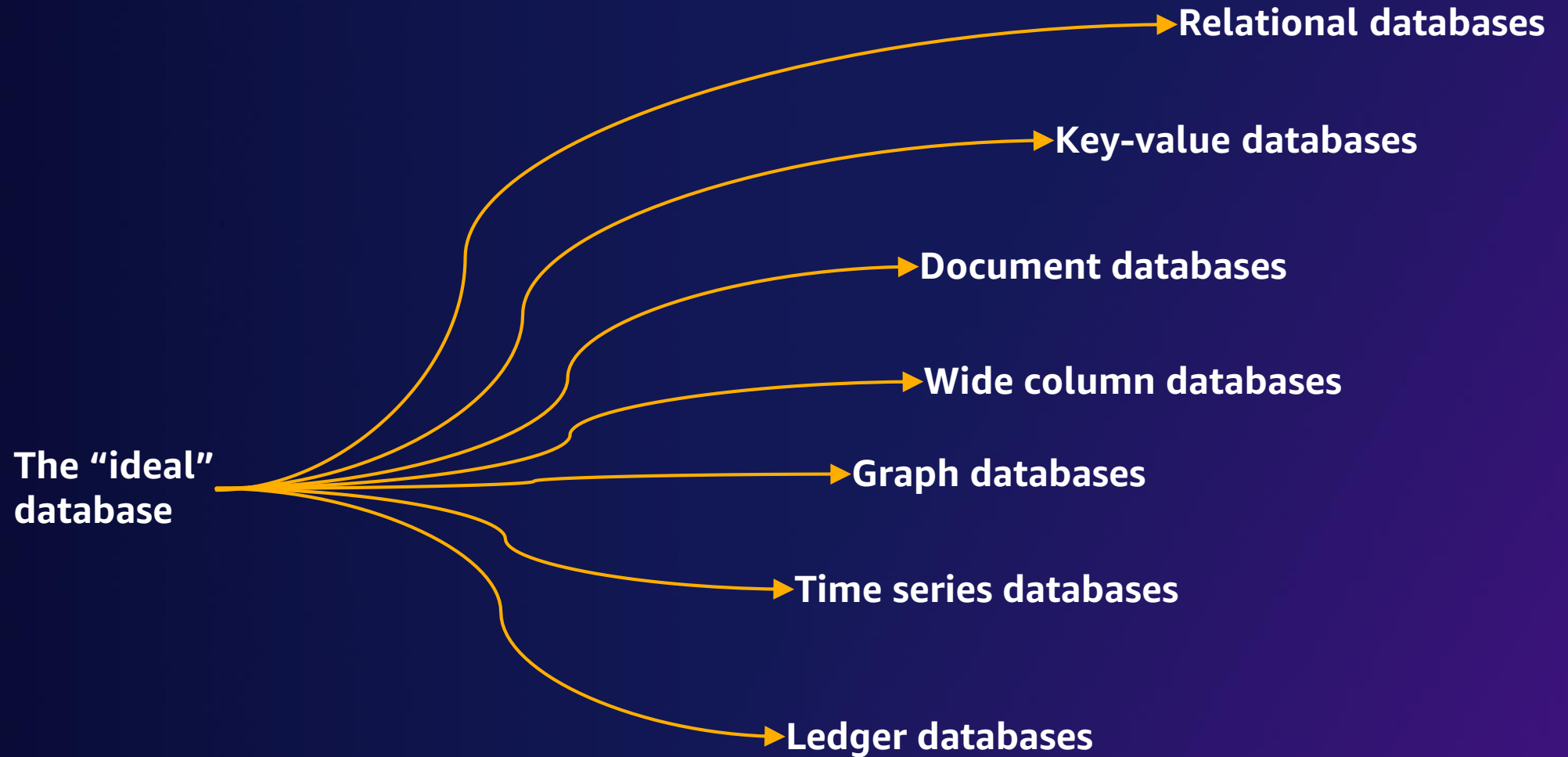


Amazon QLDB



Purpose-built databases

USE THE RIGHT TOOL FOR THE JOB

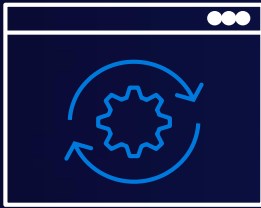


... and more!

Modernization using purpose-built databases

So, where do we start?

Database modernization playbooks



Move to
managed databases

Self-managed → AWS-managed

Oracle	}	Aurora
SQL Server		Amazon RDS
MongoDB		Amazon DocumentDB
Redis		ElastiCache
...		MemoryDB



Break free from
legacy databases

Commercial → Open source

Oracle	}	Aurora
SQL Server		Amazon RDS
IBM DB2		
Sybase		
...		

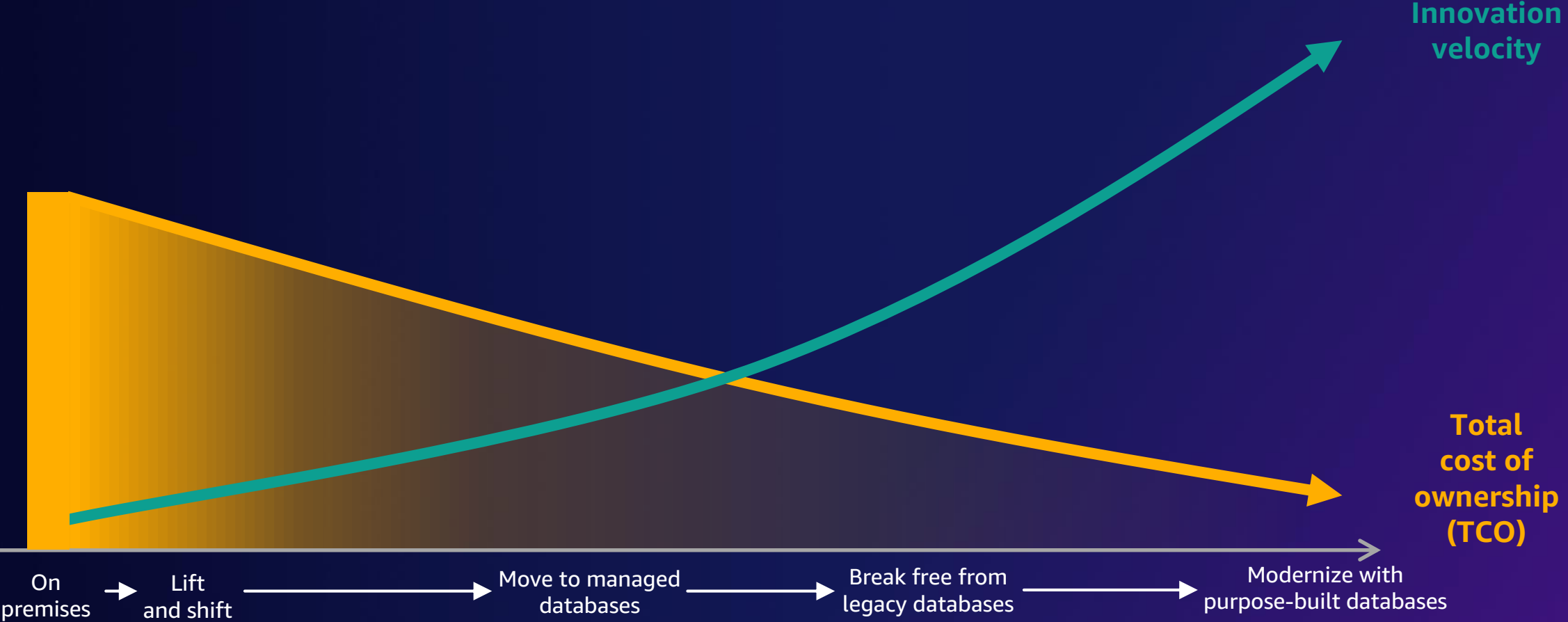


Modernize with
purpose-built databases

Monolithic → Microservices

Oracle	}	DynamoDB
SQL Server		Aurora
MySQL		Neptune
PostgreSQL		Amazon DocumentDB
...		Timestream

Both migration and modernization are important



Modernization using purpose-built databases

It's a journey!

Typical database modernization journeys start here



**Relational
database
application**

Existing business-critical applications

- Modernization of existing applications
- Typically monolithic
- Made to fit relational databases

But also

- Start up new applications
- Rapid prototyping, developed locally
- ORM/ODM heavy for agility

Step 1: Move to a managed database



Relational database application

Adopt a managed database



Managed database

Adopt relational databases built for the cloud
Amazon RDS, Aurora

Agile, easily managed, cost-effective

- Reduce operational burden
- Take advantage of platform features

Amazon RDS

MANAGED RELATIONAL DATABASE SERVICE WITH A CHOICE OF POPULAR DATABASE ENGINES



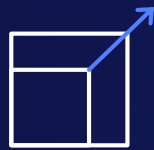
Microsoft
SQL Server

ORACLE



Easy to administer

Easily deploy and maintain hardware, OS, and DB software; built-in monitoring



Performant and scalable

Scale compute and storage with a few clicks; minimal downtime for your application



Available and durable

Automatic multi-AZ data replication; automated backup, snapshots, and failover



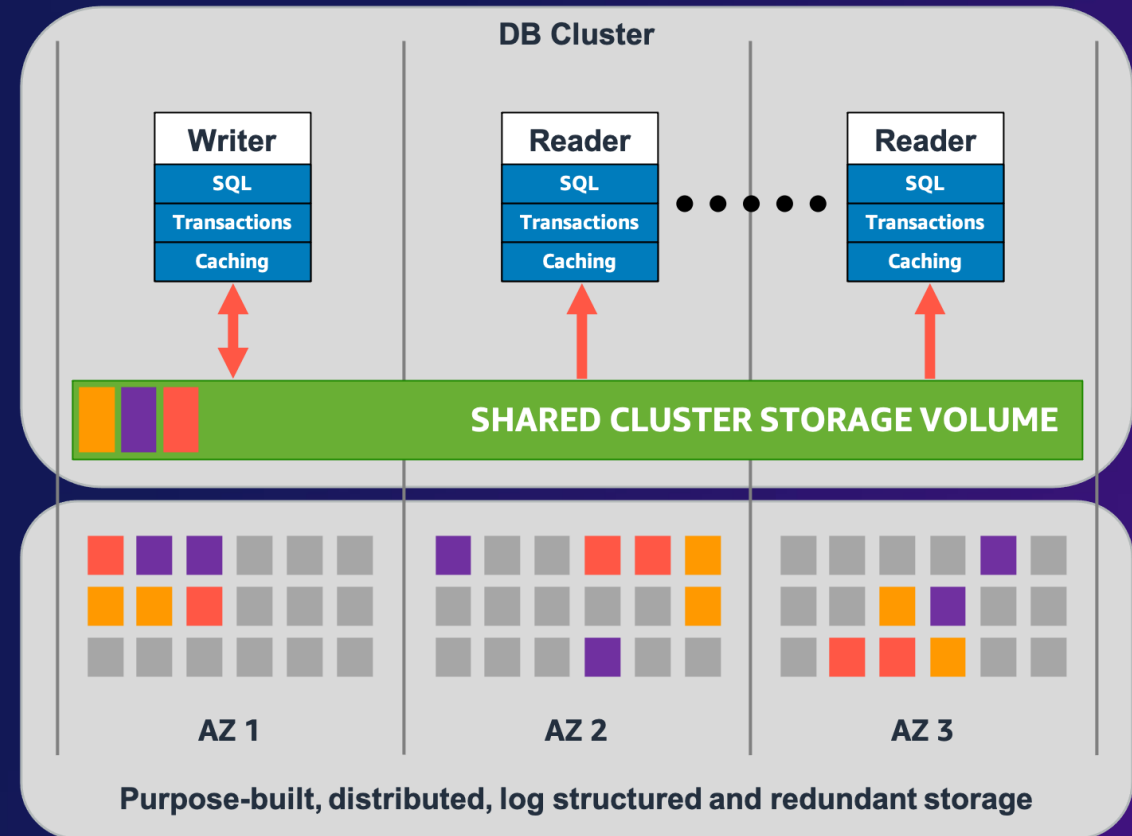
Secure and compliant

Data encryption at rest and in transit; helps customers achieve their compliance goals through tooling

Aurora

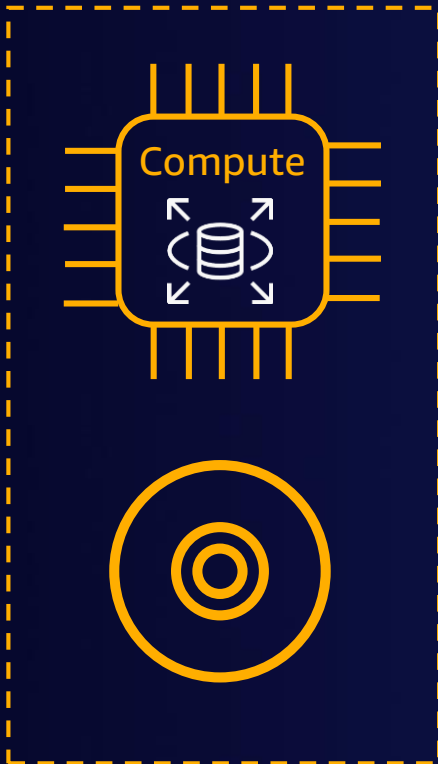
ENTERPRISE DATABASE AT OPEN-SOURCE PRICE, DELIVERED AS A MANAGED SERVICE

- Drop-in compatibility with MySQL and PostgreSQL
- Compatible with T-SQL (Microsoft SQL Server) using BabelFish for Aurora PostgreSQL
- Simplicity and cost-effectiveness of open-source databases
- Throughput and availability of commercial databases
- Simple pay-as-you-go pricing
- Global database with deployment in multiple Regions
- Serverless DB clusters that are instantly scalable

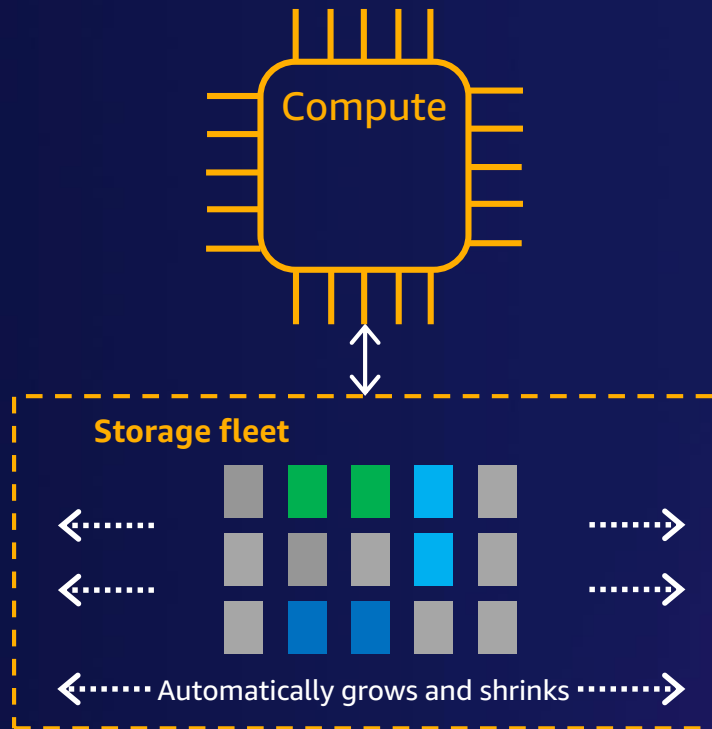


Evolution of AWS-managed relational databases

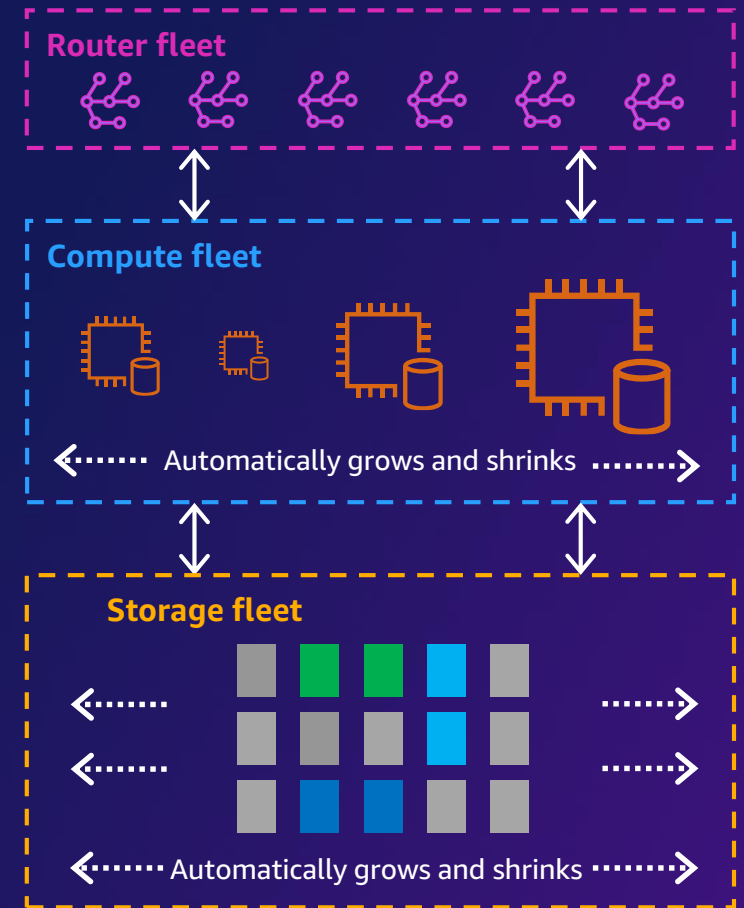
Amazon RDS DB instance



Aurora-provisioned DB cluster



Aurora serverless



What's new in Aurora



AVAILABLE IN GA

Aurora Serverless v2

- Scale to hundreds of thousands of transactions in a fraction of a second
- Scale capacity up and down in fine-grained increments
- Save up to 90% vs. provisioning for peak load
- Multi-AZ, global database, read replicas support

AVAILABLE IN GA

Babelfish for Aurora PostgreSQL

- Translation layer allows Aurora to understand Microsoft SQL Server's T-SQL
- Lower risk and complete migrations faster
- Run T-SQL code alongside new open-source functionality and continue developing with familiar tools

AVAILABLE IN GA

DevOps Guru for RDS

- Detects and diagnoses performance bottlenecks and operational issues
- Delivers recommendations to help resolve issues in minutes
- No ML experience required
- Continuously monitors workloads at scale



Reddit was founded in 2005 and ranks as the 17th most visited website in the world. The social media company was experiencing hyperscale growth, with its monthly active users surging 30% year over year.

Challenge

A substantial operational burden associated with self-managing its relational database was constraining Reddit's team and putting quality service delivery to its users at risk.

Solution

Migrating from self-managed PostgreSQL on Amazon EC2 to Aurora was a key component of Reddit's success. It has resulted in not only higher database reliability, faster backup restoration and point-in-time recovery, and fast automated failovers but also improved productivity for engineers.

“Using Aurora, we could build a longer-term strategy for how we were dealing with data and do more strategic thinking for long-term projects.”

Jason Harvey
Principal Engineer
Reddit

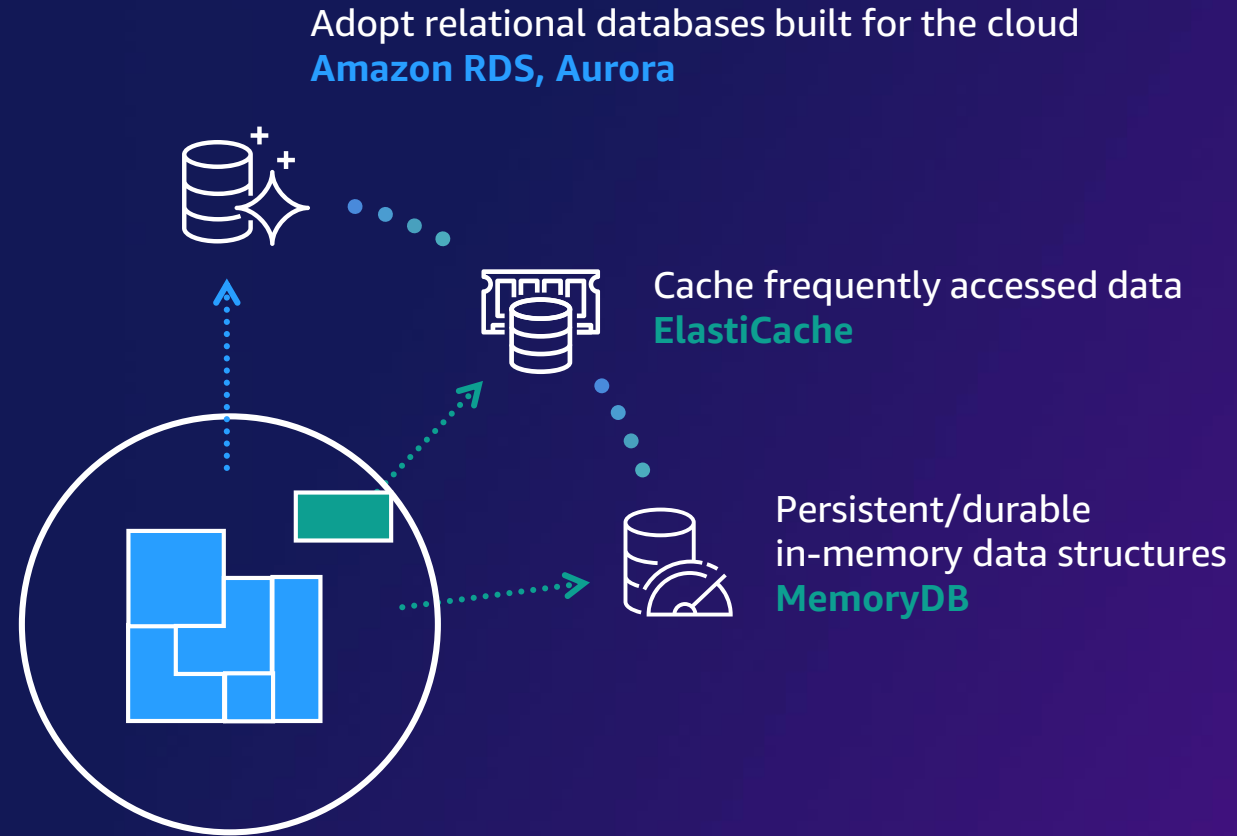
Step 2: Cache and persistent in-memory data

Address scaling pain

- Low-hanging fruit, easy to adopt
- Offload repetitive read operations
- Offload session state
- Caching

Increase performance

- Frequent counters
- Fast-changing rankings
- Submillisecond data access





PayPay Corporation is one of Japan's leading mobile payment applications, with more than 30 million users completing millions of cashless transactions every day. PayPay has been using ElastiCache for Redis both as a distributed cache and a low-latency persistent data store.

Challenge

The team was asked to build the QR Code Payment Service in a short period: 3 months.

Solution

PayPay adopted a microservices architecture on AWS designed to support the growth of the business. Each microservice team can choose the right database solution according to the application requirements.

“ We chose ElastiCache because we needed a fully managed and highly scalable distributed Redis cluster for numerous microservices. For example, our authentication service is based on access tokens stored in ElastiCache for Redis. Because every API call requires authentication, this service handles a huge amount of traffic per second and requires submillisecond performance. ElastiCache for Redis makes this possible. ”

Shilei Long
Senior Software Architect
PayPay Corporation



MemoryDB

REDIS COMPATIBILITY WITH MULTI-AZ DURABILITY

Ultra-fast performance

Lowest-latency database offered by AWS

Fully managed

AWS-managed hardware and software setup, configuration, monitoring, and snapshots

Durability and high availability

Multi-AZ transactional log for durability and replicas for high availability

Redis compatibility

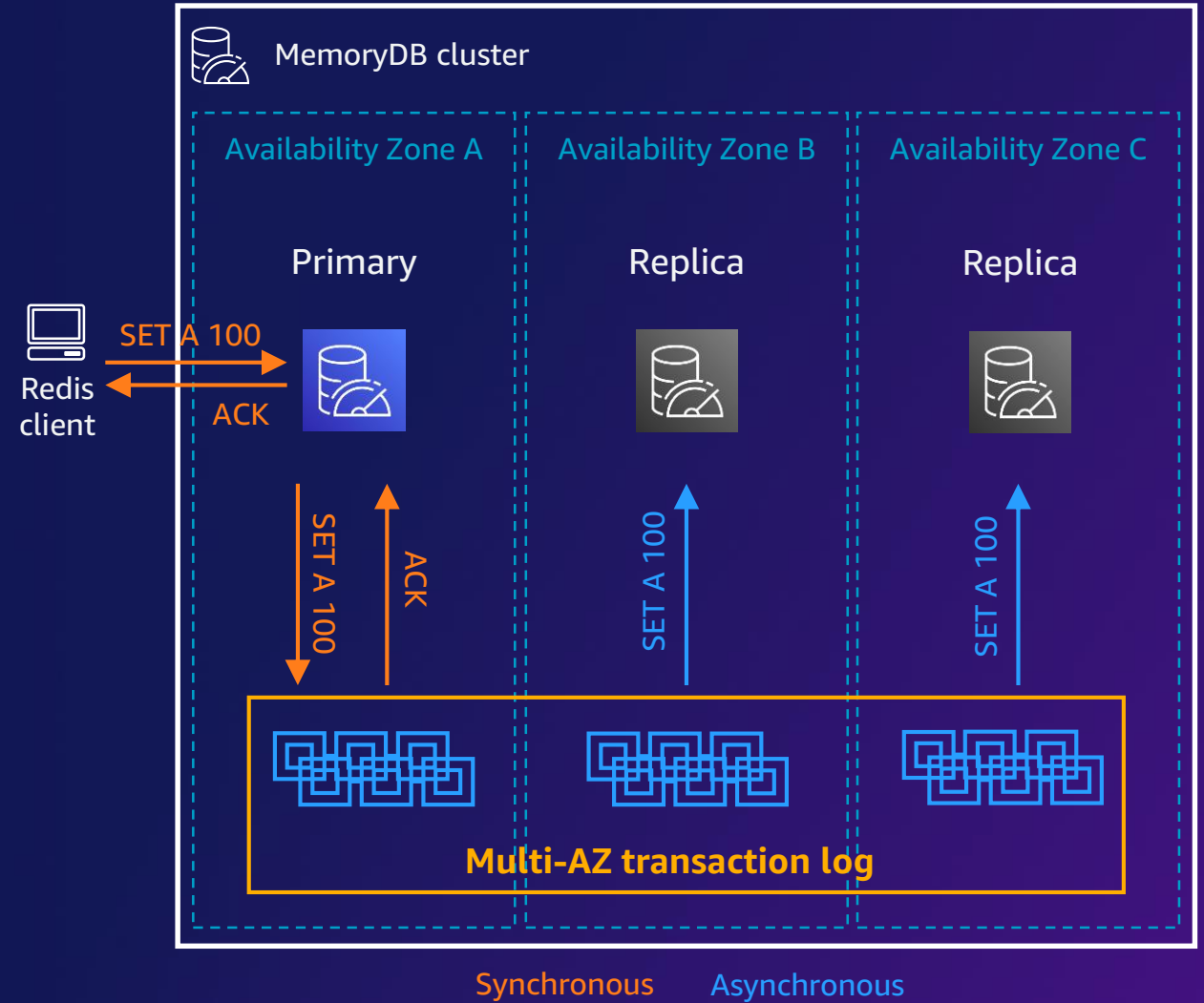
Flexible and friendly Redis APIs and data structures

High scalability

Up to 500 nodes and 128 TiB of in-memory storage per cluster (with 1 replica per shard)

Security

Amazon VPC, encryption at rest and in transit, access control lists (ACLs)



What's new in ElastiCache and MemoryDB



AVAILABLE IN GA

New JSON support for MemoryDB and ElastiCache

- Store, fetch, and update JavaScript Object Notation (JSON) data inside Redis without managing custom code
- Efficiently retrieve data to improve performance, reduce cost
- For Redis version 6.2+ in available regions, no additional cost

AVAILABLE IN PREVIEW

AWS Controllers for Kubernetes (ACK) for MemoryDB

- Define and use MemoryDB resources using ACK from your Kubernetes cluster
- Use MemoryDB to support Kubernetes applications without needing to define resources outside of the cluster

AVAILABLE IN GA

ElastiCache for Memcached supports encryption of data in transit

- Support for encryption in transit using TLS 1.2 and 1.3 for Memcached clusters
- All network traffic between clients and clusters is encrypted
- Available in all public AWS Regions and AWS GovCloud (US), no additional cost

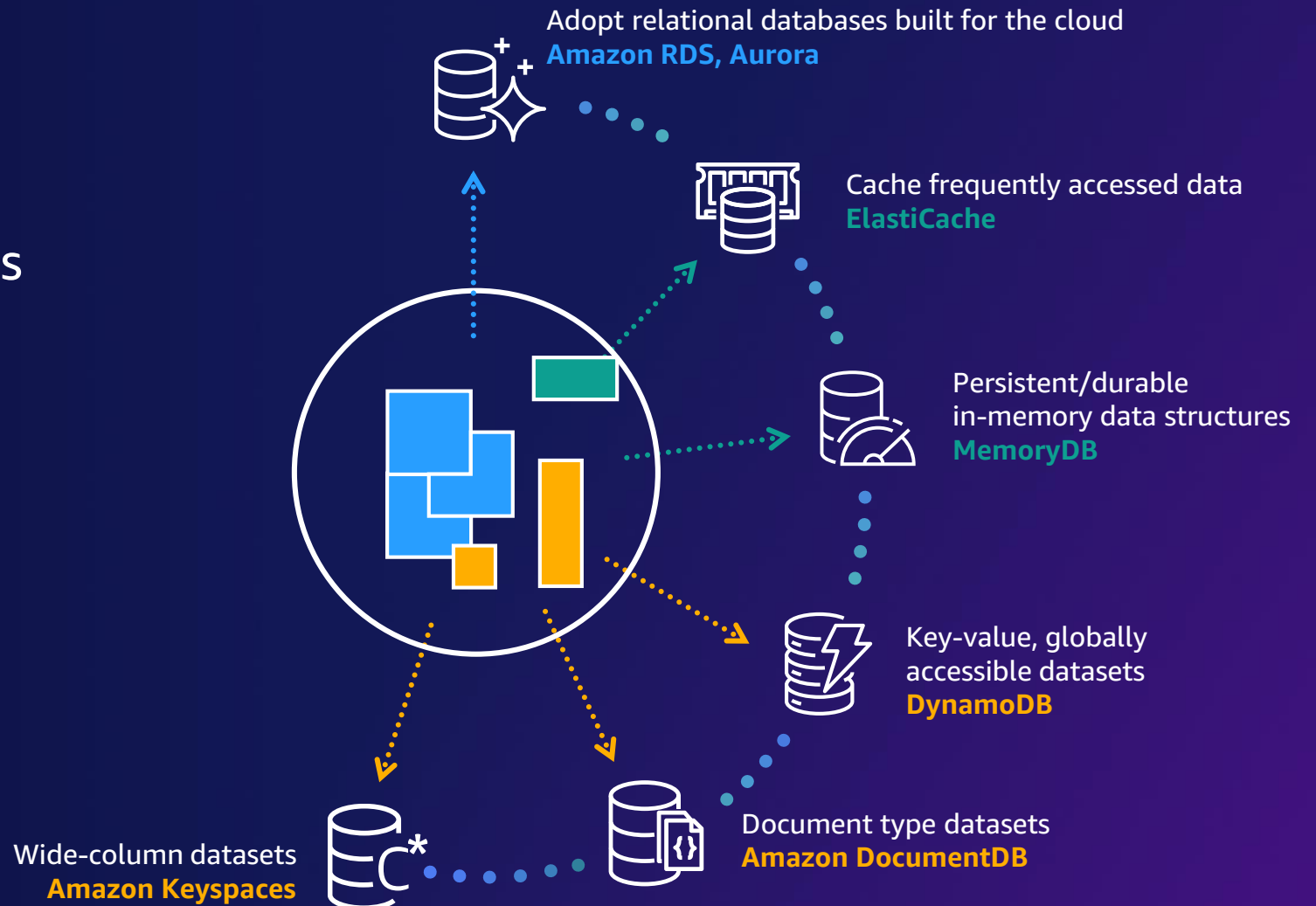
Step 3: Nonrelational query patterns

Breaking down the monolith

- Right tool for the job
- Microservice access pattern drives data store decision

Benefits

- Agility
- Scalability
- Performance



DynamoDB

FAST AND FLEXIBLE NOSQL DATABASE AT ANY SCALE



Performance at scale

- Delivers consistent, single-digit millisecond latency
- Handles millions of requests per second



No servers to manage

- Maintenance free
- Auto-scaling
- On-demand capacity mode
- Up to 99.999% SLA



Enterprise-ready

- ACID transactions
- Encryption at rest
- Continuous backups, on-demand backup and restore
- Integration with other AWS services



Global tables

- Build global applications
- Get fast access to local data
- Automated global replication

DynamoDB is ideal for

Serverless and event-driven architectures • Globally resilient services • High-throughput workloads



What's new in DynamoDB



AVAILABLE IN GA

NoSQL Workbench for DynamoDB added support

- Now supports table and global secondary index (GSI) control plane operations such as CreateTable, UpdateTable, and DeleteTable
- Perform all data plane and control plane operations from one convenient application

AVAILABLE IN GA

DynamoDB support for limit request option for PartiQL operations

- Supports limiting the number of items processed in PartiQL operations as an optional parameter on each request
- Can help reduce the cost and duration of each request when you know you need less than a certain number of items



Zoom Video Communications, Inc. provides videotelephony and online chat services through a cloud-based peer-to-peer software platform and is used for teleconferencing, telecommuting, distance education, and social relations.

Challenge

Beginning in early 2020, Zoom's software usage saw a significant global increase after quarantine measures were adopted in response to the COVID-19 pandemic.

Solution

DynamoDB global tables using on-demand capacity allowed Zoom to scale and meet that global increase in demand for its services.

“ On the backend, we were able to manage this surge with Amazon DynamoDB for Zoom Meetings. Using DynamoDB global tables in conjunction with on-demand mode enabled us to scale nearly infinitely with no performance issues, even with our sudden spike in usage. ”

Yasin Mohammed

Engineering Manager, Cloud Operations
Zoom Video Communications, Inc.



Amazon DocumentDB

FAST, SCALABLE, AND FULLY MANAGED MONGODB-COMPATIBLE DATABASE SERVICE



Fast and scalable

- Scale compute in minutes
- Storage and I/O auto scaling
- Storage scales to 64 TB without sharding
- Scale out to 15 replicas for millions of reads
- Globally distributed



Enterprise-ready

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



MongoDB-compatible

- Applications, drivers, and tools can be used 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 is ideal for

Content management • **Real-time big data** • **User or product profiles**

What's new in Amazon DocumentDB



**AVAILABLE IN
PREVIEW**

Amazon DocumentDB Performance Insights

- Tune and monitor performance
- Quickly assess database load
- Determine when and where to take action
- Visualization on one easy-to-understand dashboard

AVAILABLE IN GA

New Amazon DocumentDB Region availability

- Now available in 18 Regions
- Launched May 2022: China (Beijing) Region, operated by Sinnet
- Launched March 2022: Europe (Milan) Region

AVAILABLE IN GA

Amazon DocumentDB Free Trial

- 750 hours – db.t3.medium instance (2 vCPUs, 4GB RAM)
- 5 GB storage
- 30 million I/Os
- 5GB backup



Rappi is Colombia's first tech startup to be valued at more than \$3.5 billion. It offers on-demand, last-mile delivery for almost anything – from food and clothing to insurance and financial services – in 9 Latin American countries.

Challenge

Needed to transition from a monolithic infrastructure to a microservices-driven approach in order to keep pace of innovation required to compete in the on-demand delivery business.

Solution

Amazon DocumentDB allowed Rappi to increase stability and performance, lower maintenance needs, and free up its data platform team to focus on value-adding activities that help move the app forward.

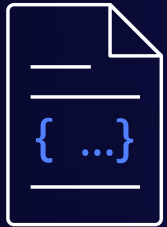
“With all the changes we had to do everywhere to cope with the new world reality, I don't remember touching anything on the Amazon DocumentDB side. It really coped with the change to the load without any problem, requiring almost no work from our side.”

Marcelo Soto
Head of Data Platform
Rappi



Amazon Keyspaces

FULLY MANAGED APACHE CASSANDRA-COMPATIBLE DATABASE SERVICE



Compatible with Apache Cassandra

- Existing applications
- New applications
- Developer skill sets



Performance at any scale

- Handles millions of requests per second
- Delivers single-digit-millisecond latency



Serverless

- No compaction
- Fully managed
- Auto scaling



Enterprise-ready

- Point-in-time recovery backups
- Encryption at rest
- IAM integration

Amazon Keyspaces is ideal for workloads with

Schema validation • **Multiple independent sorts** • **Rows up to 1 MiB in size**

What's new in Amazon Keyspaces



AVAILABLE FOR GA

Amazon Keyspaces Spark Cassandra Connector support

- Easily read and write data between Amazon Keyspaces and Spark
- Run managed, serverless Cassandra workloads in Spark-based analytics pipelines
- Available in AWS Regions where Amazon Keyspaces is offered

AVAILABLE FOR GA

Amazon Keyspaces support in AWS SDK

- Automate resource management
- Use third-party automation tools that depend on the AWS SDK to manage resources
- No additional cost

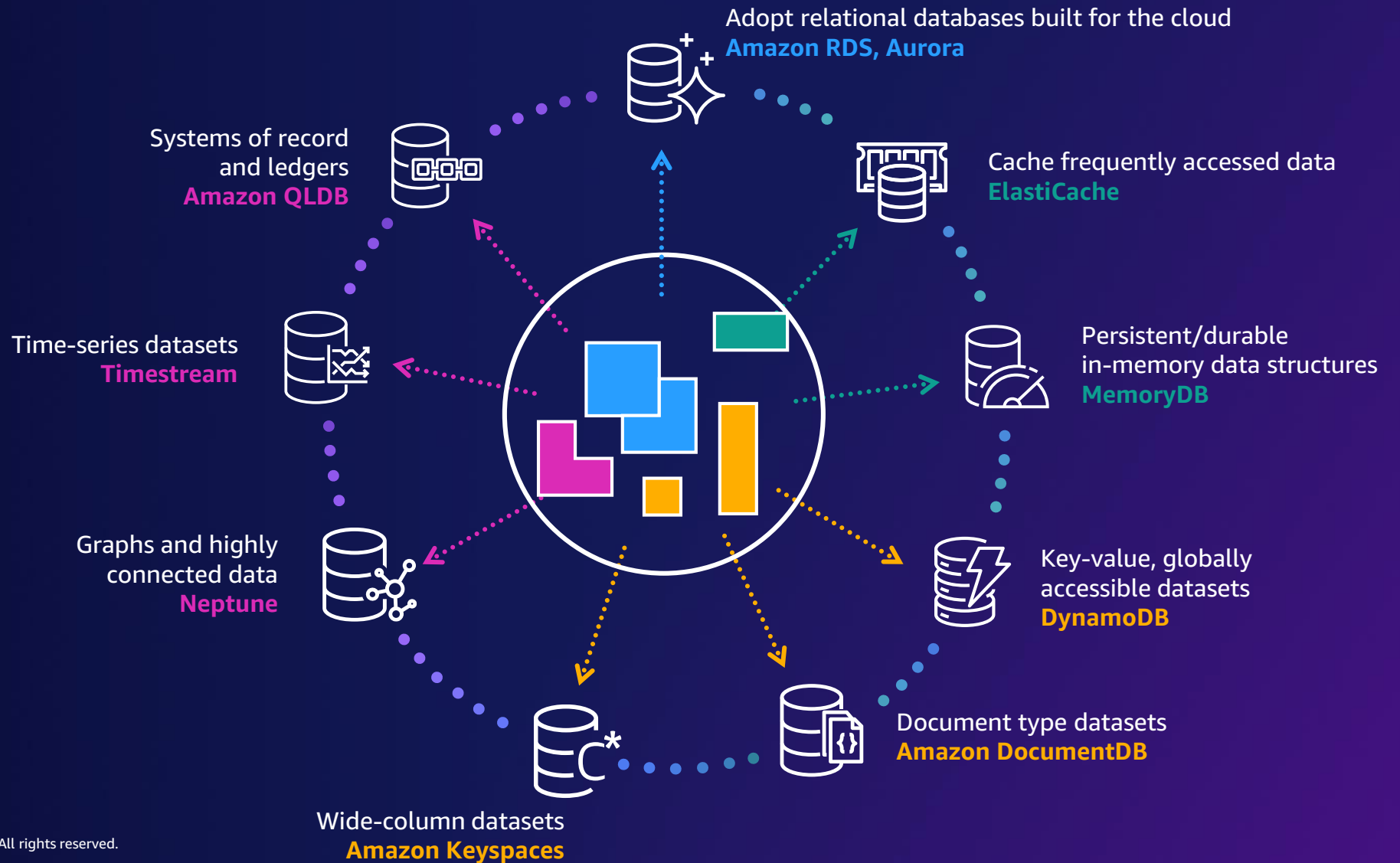
Step 4: Specialized data interactions

Specialized datasets

- Social graphs
- Recommendation engines
- Measurements
- Systems of record
- Digital records

Benefits

- Agility
- Scalability
- Performance



Neptune

FAST, RELIABLE GRAPH DATABASE BUILT FOR THE CLOUD

Open



Supports property graphs and W3C RDF graph models

Fast



Query billions of relationships with millisecond latency

Reliable



6 replicas of your data across 3 AZs with full backup and restore

Easy



Build powerful queries easily with Gremlin and SPARQL

Useful when

Relationships matter as much as the data
Results depend on the **strength, weight, or quality** of relationships

Neptune is ideal for

Social networking • **Recommendations** • **Fraud detection** • **Life sciences**

What's new in Neptune



AVAILABLE FOR GA

openCypher for Neptune

- Draw on SQL knowledge with the SQL-inspired syntax
- There are no additional charges for using openCypher
- Compatible with customers' existing property graphs

AVAILABLE FOR GA

Neptune free trial

- 750 hours of db.t3.medium instance
- 10 million I/Os
- 1 GB storage
- 1 GB backup storage

Dream11 fantasy sports Social Network Service

Dream11, the flagship brand of Dream Sports, is the world's largest fantasy sports platform, with more than 100 million users. Its Social Network Service helped it evolve as a social sports platform.

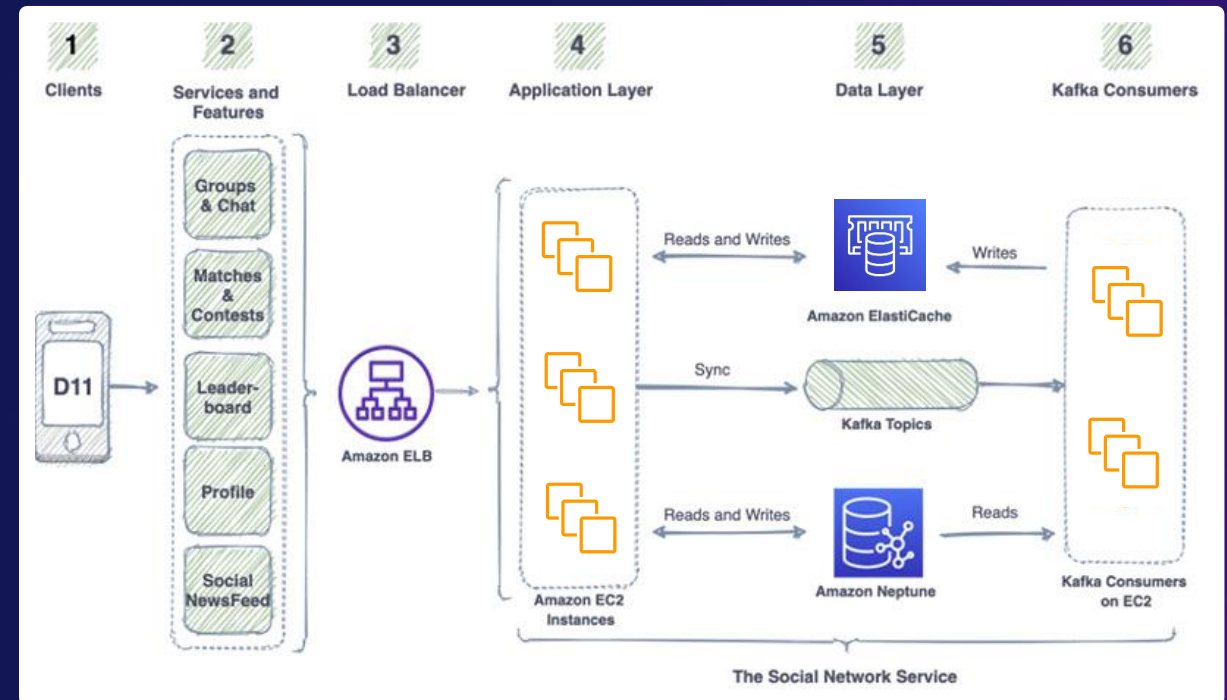
Challenges

As the Social Network Service gained popularity, it needed to scale from thousands to millions of followers per user and withstand spiky behavior around popular match times.

Solution

Neptune provides persistent graph storage with proven high availability, scalability, and performance supporting social graph traversals such as friends in common and clusters of friends.

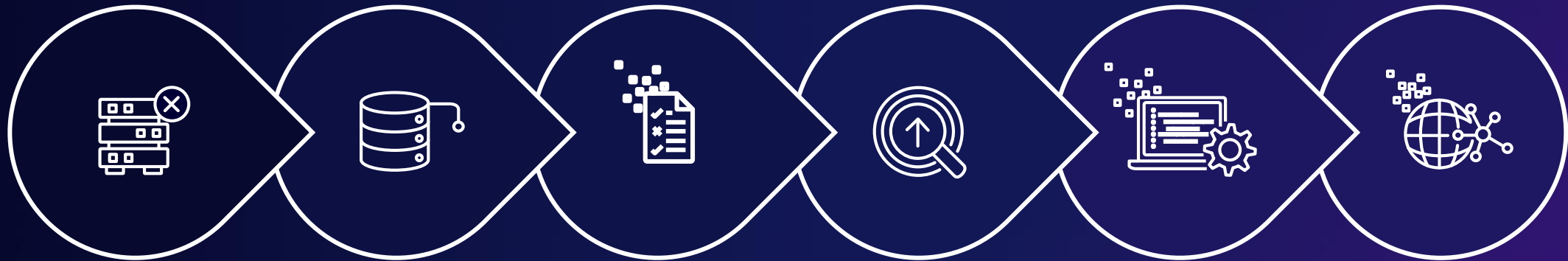
ElastiCache for Redis enables extremely high throughput API calls such as counts and sorted lists of followers.



Timestream

SERVERLESS, FULLY MANAGED TIME SERIES DATABASE

Collect, store, and process data sequenced by time **at any scale**



Serverless auto scaling
for ingesting, storing,
and querying

Data lifecycle
management

No predefined
schema

Access data across
storage tiers with
adaptive query engine

Built-in
time series
analytics in SQL

Integrates with
ML, analytics,
data collectors

Timestream is ideal for

Real-time performance tracking • Forecasting • Trend analysis • Event tracking

What's new in Timestream



AVAILABLE IN GA

Multi-measure records

- New data modeling capability for cost-effective data storage, performant data access, and ease of use
- Store multiple measurements in a single table row, instead of storing one per table row

AVAILABLE IN GA

Scheduled queries

- Orders of magnitude faster queries for operational dashboards and business reports, at a fraction of the cost
- Pre-compute and store frequently accessed aggregates and computations, power your dashboard using these pre-computed results

AVAILABLE IN GA

Magnetic store writes

- Optimize data storage spend for applications generating late arrival data
- Enable writes directly to magnetic store, avoid longer memory store retention to account for late data arrival

Amazon QLDB

FULLY MANAGED LEDGER DATABASE SERVICE

Immutable



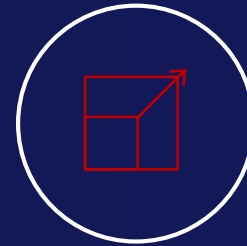
Append-only, immutable journal
tracks history of all changes

Verifiable



All changes are **cryptographically chained** and **verifiable**

Serverless



Scalable and highly available,
no need to provision storage
and I/O in advance

Easy to use



PartiQL offers familiar SQL
operators, ION data model is
rich and expressive

Amazon QLDB is ideal for

Audit databases • **State change modeling** • **Systems of record** • **Blockchain alternatives**

What's new in Amazon QLDB



AVAILABLE IN GA

New Amazon QLDB shell version

- Download the QLDB shell tailored to your favorite operating system and begin querying a QLDB ledger without any other installation steps
- Provides improved query runtime statistics

AVAILABLE IN GA

JSON output support

- Receive journal data in newline-delimited JSON format
- Enables direct integration with common software libraries, tools, and popular Amazon services such as AWS Glue and Athena

AVAILABLE IN GA

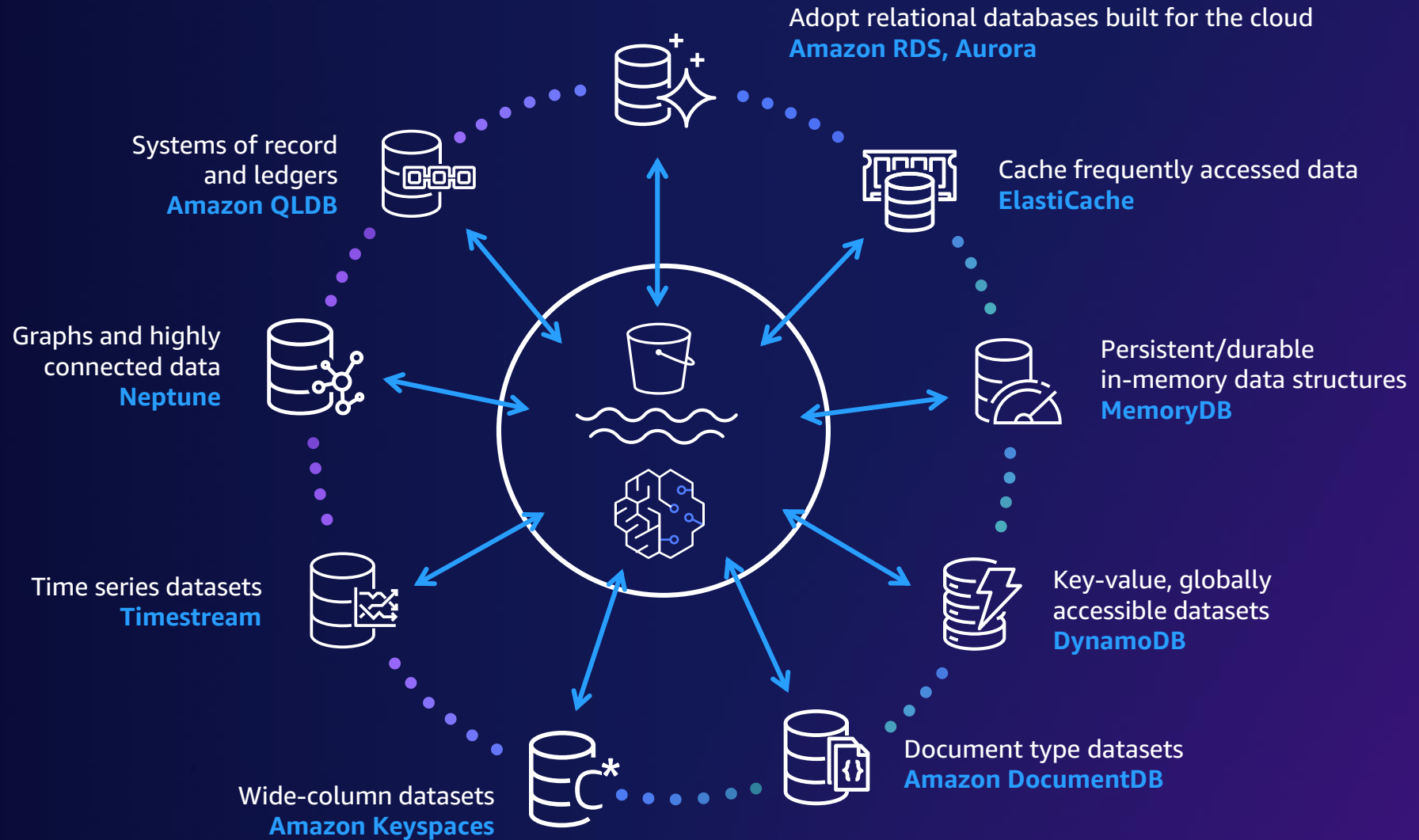
Console query editor

- Improved interface for authoring queries, debugging transactions, and exploring results
- Filter your results in your preferred format

Next steps: Unify and innovate



Unify using data lakes and innovate using ML



To sum it all up



Purpose-built modernization is a journey

BOTH ORGANIZATIONALLY AND TECHNOLOGICALLY

- 1** **Move relational workloads to managed DB services**
Aurora, Amazon RDS
Improve: agility, ease of management, cost-efficiency
- 2** **Implement caching and offload in-memory workloads**
ElastiCache, MemoryDB
Improve: scalability and performance
- 3** **Move nonrelational query patterns to NoSQL**
DynamoDB, Amazon DocumentDB, Amazon Keyspaces
Improve: agility, innovation, scalability, performance, ease of management, cost-efficiency
- 4** **Move specialized datasets and query patterns to specialized databases**
Neptune, Timestream, Amazon QLDB
Improve: agility, innovation, scalability, performance, ease of management, cost-efficiency

Learn in-demand AWS Cloud skills



AWS Skill Builder

Access **500+ free** digital courses and Learning Plans

Explore resources with a variety of skill levels and **16+** languages to meet your learning needs

Deepen your skills with digital learning on demand



Train now



AWS Certifications

Earn an industry-recognized credential

Receive Foundational, Associate, Professional, and Specialty certifications

Join the **AWS Certified community** and get exclusive benefits



Access **new** exam guides

Thank you!

Barry Morris

bamorri@amazon.com

