



# Building Your First Modern Apps with Purpose-built Databases

Justin Thomas  
Principal Business Development Manager,  
Amazon Neptune

# Agenda

- New data requirements for modern applications
- Why purpose-built databases?
- Picking the right tool for the job
- Resources
- Q&A

# New data requirements for modern applications

# Rapid expansion of data requirements

## Explosion of data



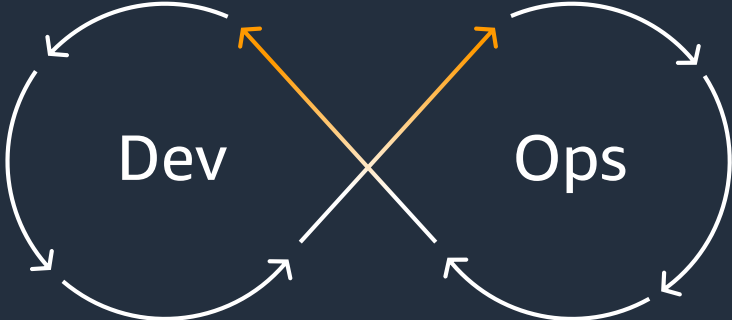
Data grows 10x every 5 years driven by network-connected smart devices

## Microservices change data and analytics requirements



Microservices architecture decreases need for "one size fits all" databases and increases need for real-time monitoring and analytics

## Accelerated rate of change driven by DevOps



Transition from IT to DevOps increases rate of change



Aurora



RDS



DynamoDB



DocumentDB



ElastiCache



Neptune



Timestream



QLDB



Keyspaces

1

**Choice** of all technology

2

**Best of breed** in each category

3

**Designed** for the cloud

The **most complete family** of purpose-built databases

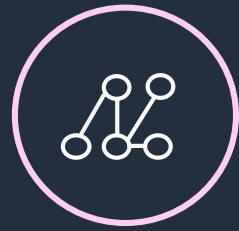
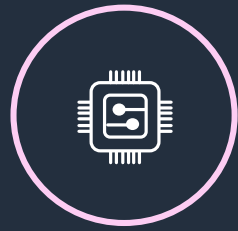
Figure 1: Magic Quadrant for Cloud Database Management Systems



# AWS Recognized as a Cloud Database Leader for 6th Consecutive Year

Gartner, Magic Quadrant for Cloud Database Management Systems, Donald Feinberg, Merv Adrian, Rick Greenwald, Adam Ronthal, Henry Cook, 23 November 2020. This graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from AWS. Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

# Building modern applications with purpose-built databases



## Relational

Referential integrity, ACID transactions, schema-on-write

## Key-value

High throughput, Low latency reads and writes, endless scale

## Document

Store documents and quickly access querying on any attribute

## In-memory

Query by key with microsecond latency

## Graph

Quickly and easily create and navigate relationships between data

## Time-series

Collect, store, and process data sequenced by time

## Ledger

Complete, immutable, and verifiable history of all changes to application data

## Wide Column

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

AWS Service(s)



Lift and shift, ERP, CRM, finance

Real-time bidding, shopping cart, social, product catalog, customer preferences

Content management, personalization, mobile

Leaderboards, real-time analytics, caching

Fraud detection, social networking, recommendation engine

IoT applications, event tracking

Systems of record, supply chain, health care, registrations, financial

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

Common Use Cases

# The Pokémon Company migrates to AWS purpose-built databases

## Challenge:

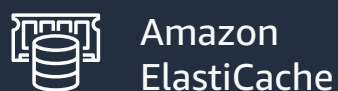
The Pokémon Company International wanted to address the complexity of managing NoSQL and memory-caching systems for a database of more than 300 million users.

## Solution:

The company uses Amazon Aurora as its main user database, Amazon DynamoDB to reduce bot login attempts, and Amazon ElastiCache to enable dynamic caching of user logins.

## Results:

- Cuts monthly costs by tens of thousands of dollars
- Reduces number of nodes from 300 to 30
- Experiences zero hours of downtime or performance degradation after migration





# FINRA uses different AWS databases for different jobs



## Market data search:

- Massive data volume
- Need quick lookups for searches



Amazon  
DynamoDB

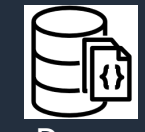
## Session state:

- In-memory store for sub-millisecond site rendering



Amazon  
ElastiCache

## Data collection



Amazon DocumentDB

## Fraud detection



Amazon Neptune

## Relational data:

- Referential integrity
- Primary transactional database



Amazon Aurora



Amazon RDS

# Picking the right tool for the job

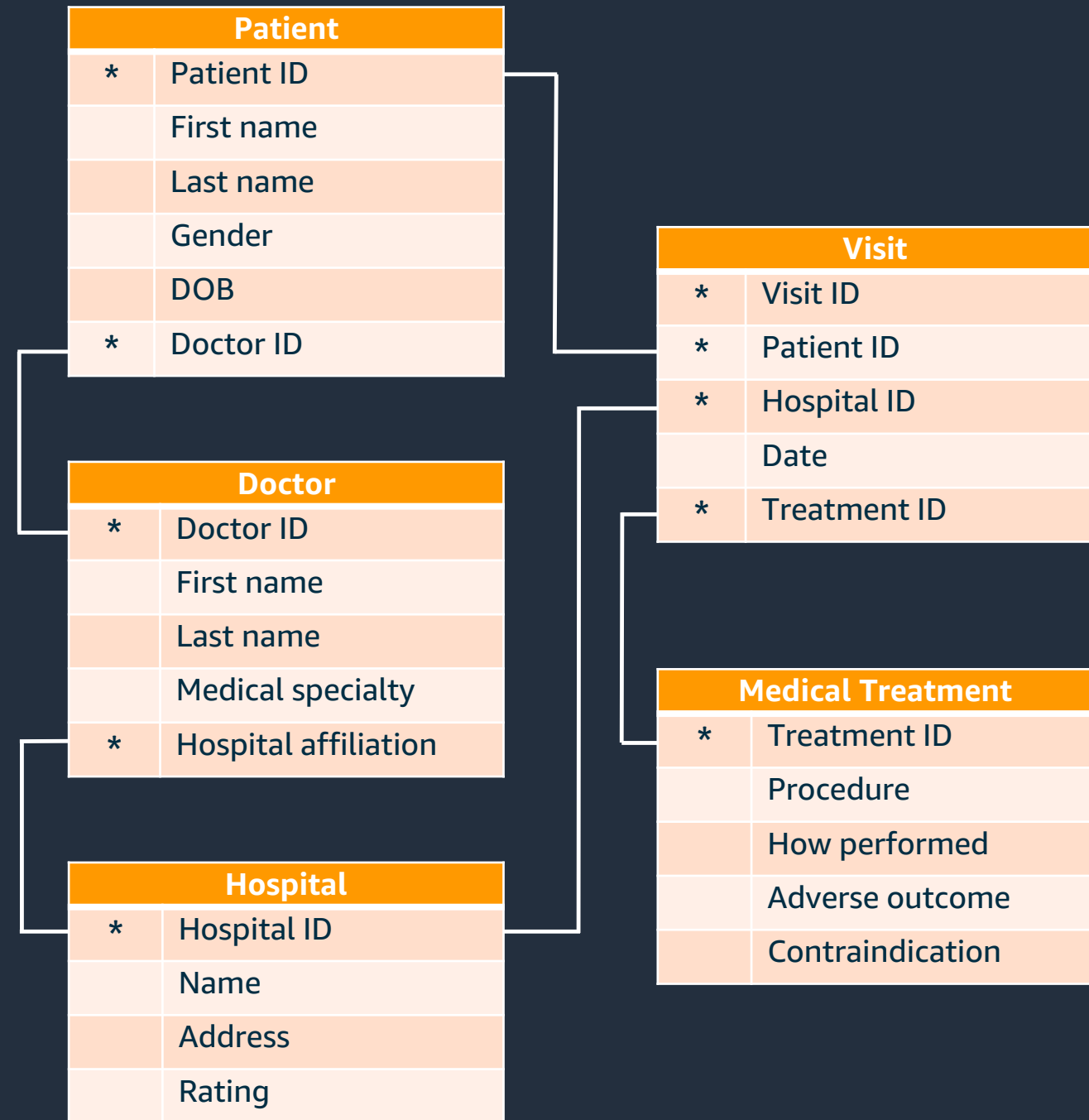


# Relational data

- Divide data among tables
- Highly structured
- Relationships established via keys enforced by the system
- Data accuracy and consistency

 Amazon Aurora

 Amazon RDS





A leader in 3D design, engineering, and entertainment software, Autodesk makes software for people who make things

### Challenge:

Their ACM (Access Control Management) app connected to a single RDS MySQL database instance, limiting the available scaling options

### Solution:

They chose **Amazon RDS** for MySQL first, and migrated to **Aurora** for improved scalability, availability, and performance

### Result:

- Aurora provides up to five times the throughput of standard MySQL databases
- Aurora scales up to 64 TB, with no need to scale manually
- Up to 15 low-latency Aurora Replicas improve availability





# Key-value data

- Simple key-value pairs
- Partitioned by keys
- Resilient to failure
- High-throughput, low-latency reads and writes
- Consistent performance at scale

```
PUT {
  TableName: "Gamers",
  Item: {
    "GamerTag": "Hammer57",
    "Level": 21,
    "Points": 4050,
    "Score": 483610,
    "Plays": 1722
  }
}
```

```
GET {
  TableName: "Gamers",
  Key: {
    "GamerTag": "Hammer57",
    "ProjectionExpression": "Points"
  }
}
```

Gamers				
Primary key	Attributes			
Gamer tag	Level	Points	High score	Plays
Hammer57	21	4050	483610	1722
FluffyDuffy	5	1123	10863	43
Lol777313	14	3075	380500	1307
Jam22Jam	20	3986	478658	1694
ButterZZ_55	7	1530	12547	66
...	...	...	...	...



# Amazon DynamoDB

Fast and flexible key value database service for any scale



## Performance at scale

Consistent, single-digit millisecond response times at any scale; build applications with virtually unlimited throughput



## Serverless architecture

No hardware provisioning, software patching, or upgrades; scales up or down automatically; continuously backs up your data



## Enterprise security

Encrypts all data by default and fully integrates with AWS Identity and Access Management for robust security



## Global replication

Build global applications with fast access to local data by easily replicating tables across multiple AWS Regions

# Dropbox saves millions by building scalable metadata store on AWS

## Challenge:

Dropbox had a capacity crunch in its on-premises metadata store due to fast data growth in some partitions. It had three options: double its on-premises storage, delete swaths of metadata, or find a new solution.

## Solution:

Dropbox turned to Amazon DynamoDB and Amazon Simple Storage Service for its new managed storage system, which saved the company millions of dollars and has room for virtually unlimited user metadata.

## Results:

- Launched metadata storage system on AWS in 1 year
- Cut cost per user gigabyte by a factor of 5.5
- Ingests data at 4,000–6,000 queries per second



# Amazon Keyspaces (for Apache Cassandra)

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



Apache Cassandra-compatible



Use the same Cassandra drivers and tools

No servers to manage



No need to provision, configure, and operate large Cassandra clusters

Single-digit millisecond performance at scale



Scale tables up and down automatically  
Virtually unlimited throughput and storage

Highly available and secure



99.99% availability SLA within an AWS Region  
Data encrypted at rest; integrated with IAM

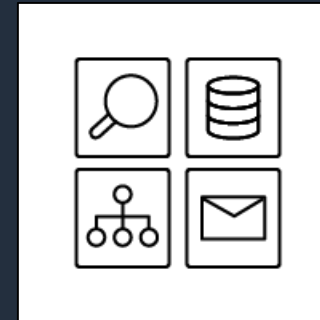




# Why use a document database?

The JSON document model maps naturally to application data

---



Each document can have a different data structure and is independent of other documents

---



Index on any key in a document, and run ad hoc and aggregation queries across your data set



# Amazon DocumentDB

Fast, scalable, and fully managed MongoDB-compatible database service

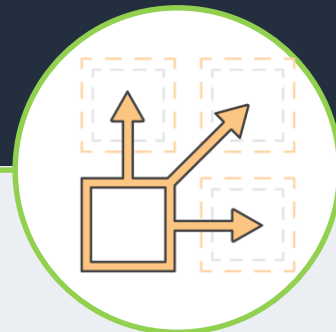


## Fast



Millions of requests per second with millisecond latency

## Scalable



Separation of compute and storage scales both independently; scale out to 15 read replicas in minutes

## Fully managed



Managed by AWS: no hardware provisioning; auto patching, quick setup, secure, and automatic backups

## MongoDB compatible



Compatible with MongoDB 3.6 and 4.0; use the same SDKs, tools, and applications with Amazon DocumentDB. Migrate workloads with AWS DMS.

Purpose-built document database engineered for the cloud

# FINRA better manages millions of documents with DocumentDB



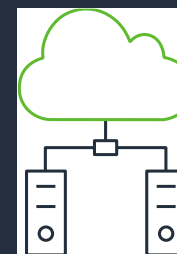
“ One of the reasons that we picked a managed service is we are big in requirements around security, audit, versioning, records management. All of these are very important to us, so by going to a managed service like using DocumentDB...we get all of these things free for us so that we do not have to worry about building a lot of this by ourselves.

- **Ranga Rajagopal**, Senior Director, Enterprise Data Platforms, FINRA

”



Amazon DocumentDB

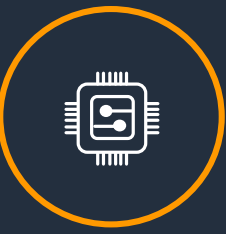


FINRA processes 4 million filings and filing attachments per year. The regulator's existing solution to store millions of documents as XML in a relational database was consuming too much storage, required custom tooling, and was difficult to manage.

FINRA chose Amazon **DocumentDB** (with MongoDB compatibility) as a managed JSON document store, making it simpler to query and index regulatory documents, reduce development cycles, and extend usability of data.

Unlike in a relational DB, the storage and compute are decoupled in DocumentDB, which enables FINRA's databases to scale automatically, independently and quickly, regardless of the size of their data.

# In-memory databases: usage patterns



Caching



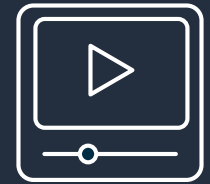
Real-time analytics store



Gaming leaderboards



Geospatial



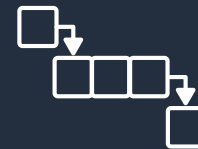
Media streaming store



Session store



Chat apps pub/sub



Job queue



Machine learning real-time model scoring



# Amazon ElastiCache

# Amazon ElastiCache

Managed, Redis or Memcached-compatible in-memory data store



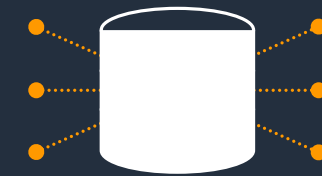
## Unlimited scale

Read scaling with replicas. Write and memory scaling with sharding. Non disruptive scaling



## Consistent high performance

In-memory data store and cache for sub-millisecond response times

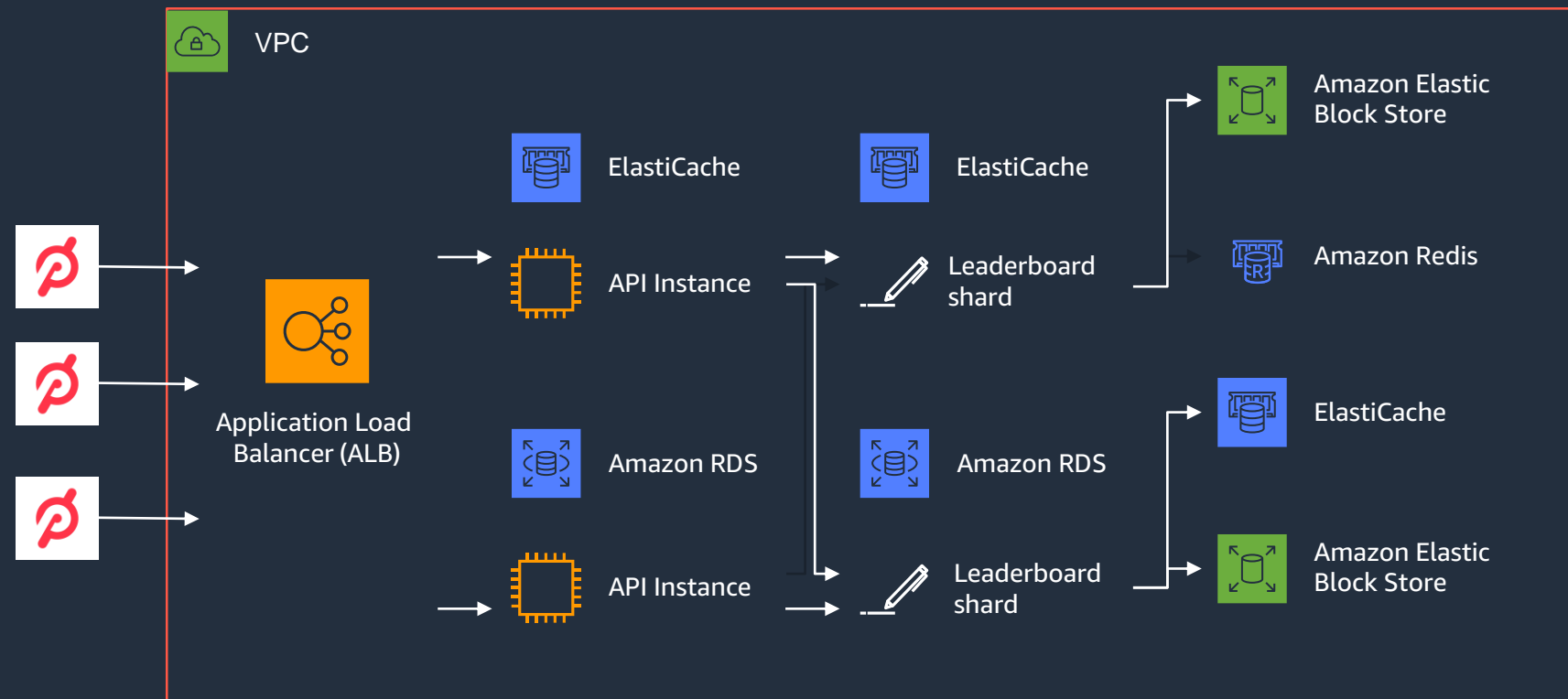


## Fully managed

AWS manages all hardware and software setup, configuration, monitoring



## Powering the leaderboard with AWS



- Required millisecond response time for real-time leaderboards with thousands of users and up to 20K requests/sec
- Use ElastiCache Redis as a in-memory data store for real-time data computation
- Use RDS MySQL as the persistent data store

# When to use graph?



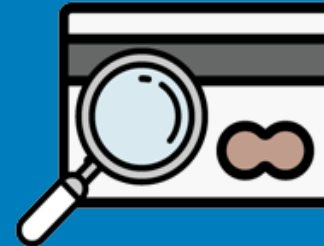
Social  
networking



Recommendations



Knowledge  
graphs



Fraud  
detection



Life  
Sciences



Network & IT  
operations

## With connected datasets

- Where relationships between data matter just as much as the data itself
- When results are dependent on the basis of the *strength*, *weight*, or *quality* of relationships

# Amazon Neptune

Fast, reliable graph database built for the cloud



## OPEN



Supports Apache TinkerPop & 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



# SIEMENS

*Ingenuity for life*

Siemens is a global powerhouse focusing on the areas of electrification, automation, and digitalization

## Challenge:

They were faced with isolated data silos from different departments that resulted in data inaccessibility, inefficient workflows, and low data quality

## Solution:

- Industrial knowledge graphs for capturing Siemens Domain Knowledge
- Providing knowledge graphs as a service



# Amazon Neptune ML



*Easy, fast, and accurate predictions on graphs with graph neural networks (GNNs), powered by the Deep Graph Library (DGL) and Amazon SageMaker*



Integrations  
and ML

## Make predictions on graph data without ML expertise

Automatically choose and train the best ML model for your workload, enabling you to make ML-based predictions on graph data in hours instead of weeks

## Use state-of-the-art graph machine learning

Up to 50% more accurate, GNNs are state-of-the-art ML purpose-built to use the relationships in graphs based on research from Stanford University

## Scale to large datasets

For graph applications with billions of relationships in knowledge graphs, fraud detection, or product recommendation

# Ledgers and audit tables have been around for a while



## Banking and finance

Keeping track of transactions, trades, and accounts



## Manufacturing

Recording components used in manufacturing



## Ownership

Maintaining records of asset ownership or provenance



# Amazon Quantum Ledger Database

Fully managed ledger database

Track and verify history of all changes made to your application's data

Immutable and transparent



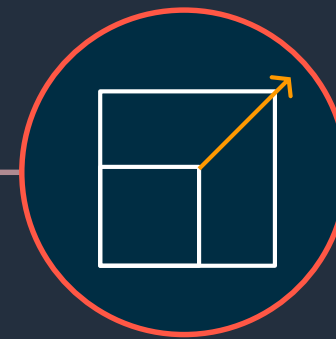
Append-only, immutable journal tracks history of all changes which cannot be deleted or modified. Get full visibility into entire data lineage

Cryptographically verifiable



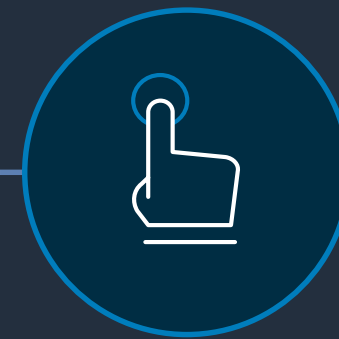
All changes are cryptographically chained and verifiable

Highly scalable



Executes 2 – 3X as many transactions than ledgers in common blockchain frameworks

Easy to use



Flexible document model, query with familiar SQL-like interface

# BMW

## Digital Vehicle Passport

### Challenge

BMW needs to track trusted, verifiable automotive data in order to get full transparency on transactions across multiple entities.

### Solution

Started building a BMW Digital Vehicle Passport App that provides a transparent and complete history of vehicle data such as fueling, inspection, oil changes, diagnostics, repairs, tire changes, and sales across multiple partners. Amazon QLDB is at the core of this solution and provides BMW the verified data with a centralized trust



# BMW GROUP

# Time series data

**Time-series data is a sequence of data points recorded over a time interval for measuring events that change over time**





# Amazon Timestream

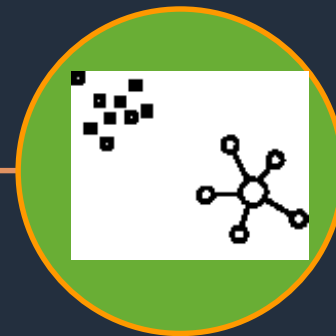
## Fast, scalable, fully managed time series database

1,000x faster and 1/10<sup>th</sup> the cost of relational databases



Collect data at the rate of millions of inserts per second (10M/second)

Trillions of daily events



Adaptive query processing engine maintains steady, predictable performance

Time-series analytics



Built-in functions for interpolation, smoothing, and approximation

Serverless



Automated setup, configuration, server provisioning, software patching

# Customer use case

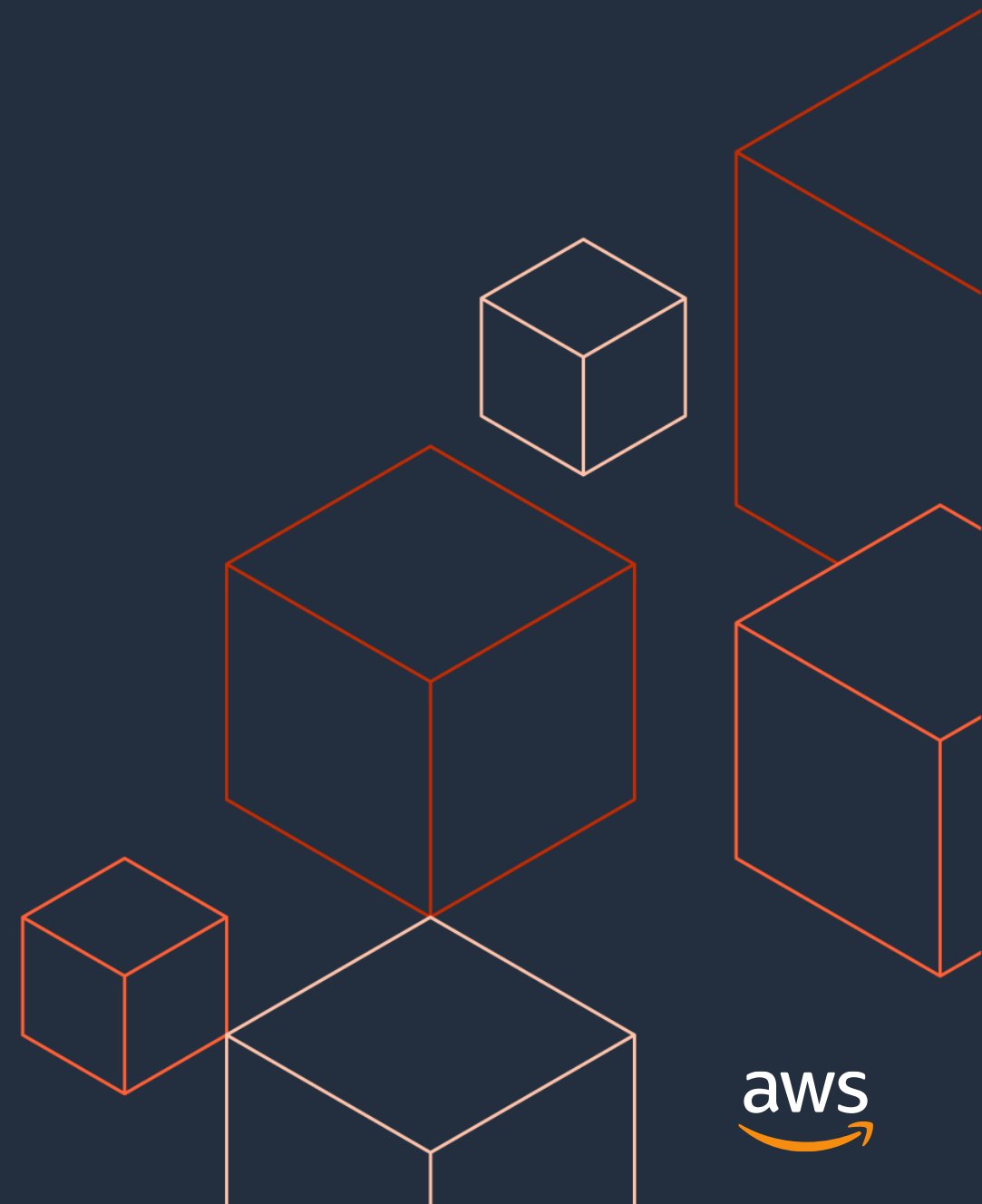
**Trimble Inc., is a leading technology provider of productivity solutions for the construction, agriculture, geospatial, and transportation industries.**

“Whenever possible, the Trimble Cloud Platform team tries to leverage AWS’ managed service offerings. We are excited to now use Amazon Timestream as a serverless time series database supporting our IoT monitoring solution. Timestream is purpose-built for our IoT-generated time series data, and will allow us to reduce management overhead, improve performance, and reduce cost for our existing monitoring system.” **David Kohler, Engineering Director – Trimble**





# Conclusion



# Our approach



Architect services ground-up for the cloud and for the explosion of data



Offer a portfolio of purpose-built services, optimized for your workloads



Help you innovate faster through managed services



Provide services that help you migrate existing apps and databases to the cloud

# Get started

See more information at:  
[aws.amazon.com/databases](https://aws.amazon.com/databases)

Contact us at:  
<https://aws.amazon.com/contact-us/>

# Q&A



