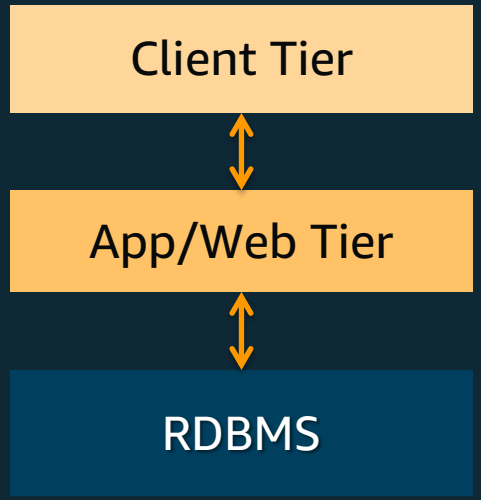
Solutions Architect Professional

Databases on AWS

Traditional Database Architecture

one database for all workloads

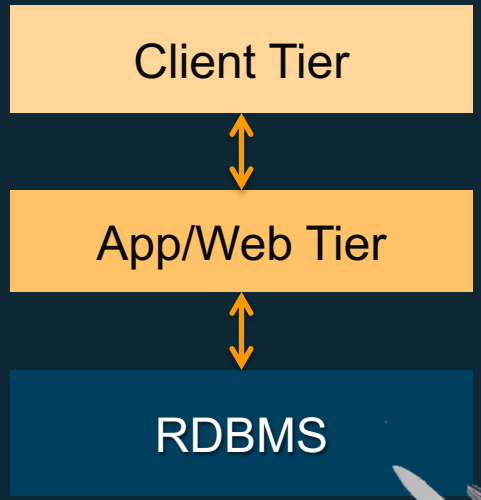


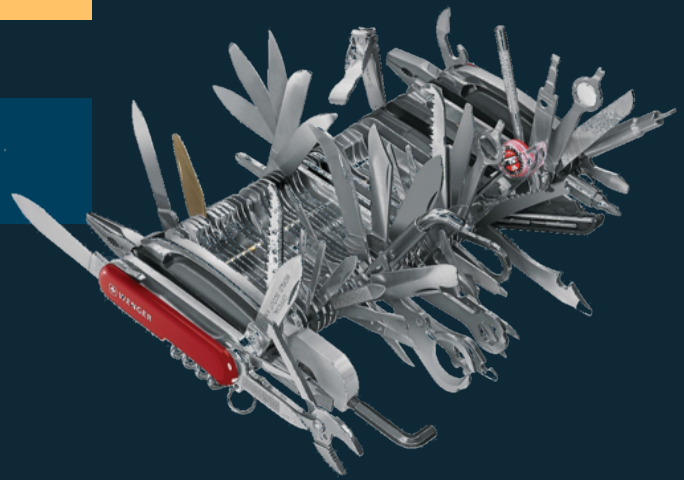
aws

Traditional Database Architecture

Key-value access
Complex queries
OLAP transactions
Analytics

All forced into the relational database

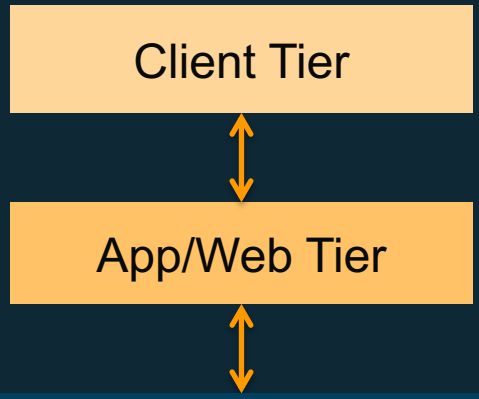




aws

AWS Data Tier Architecture

On AWS choose best database service for each workload

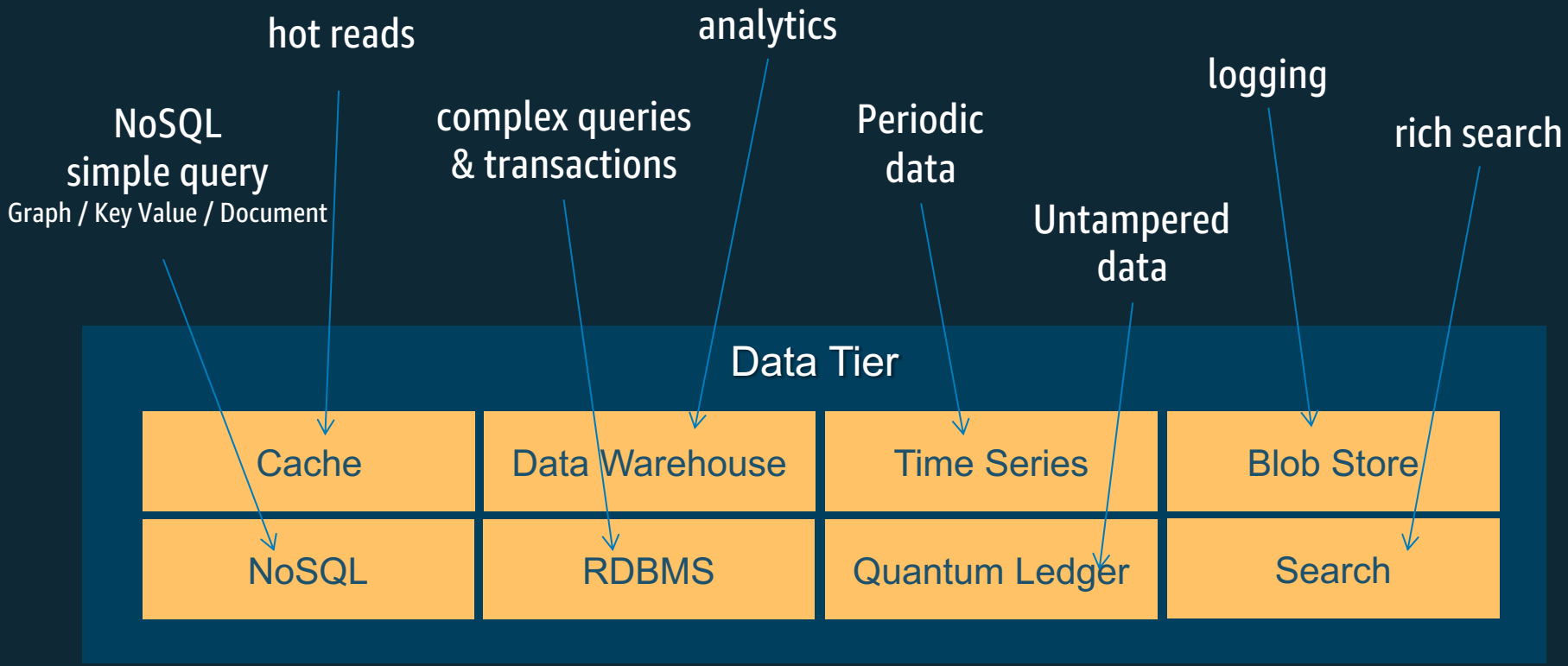


Data Tier

Cache	Data Warehouse	Time Series	Blob Store
NoSQL	RDBMS	Quantum Ledger	Search

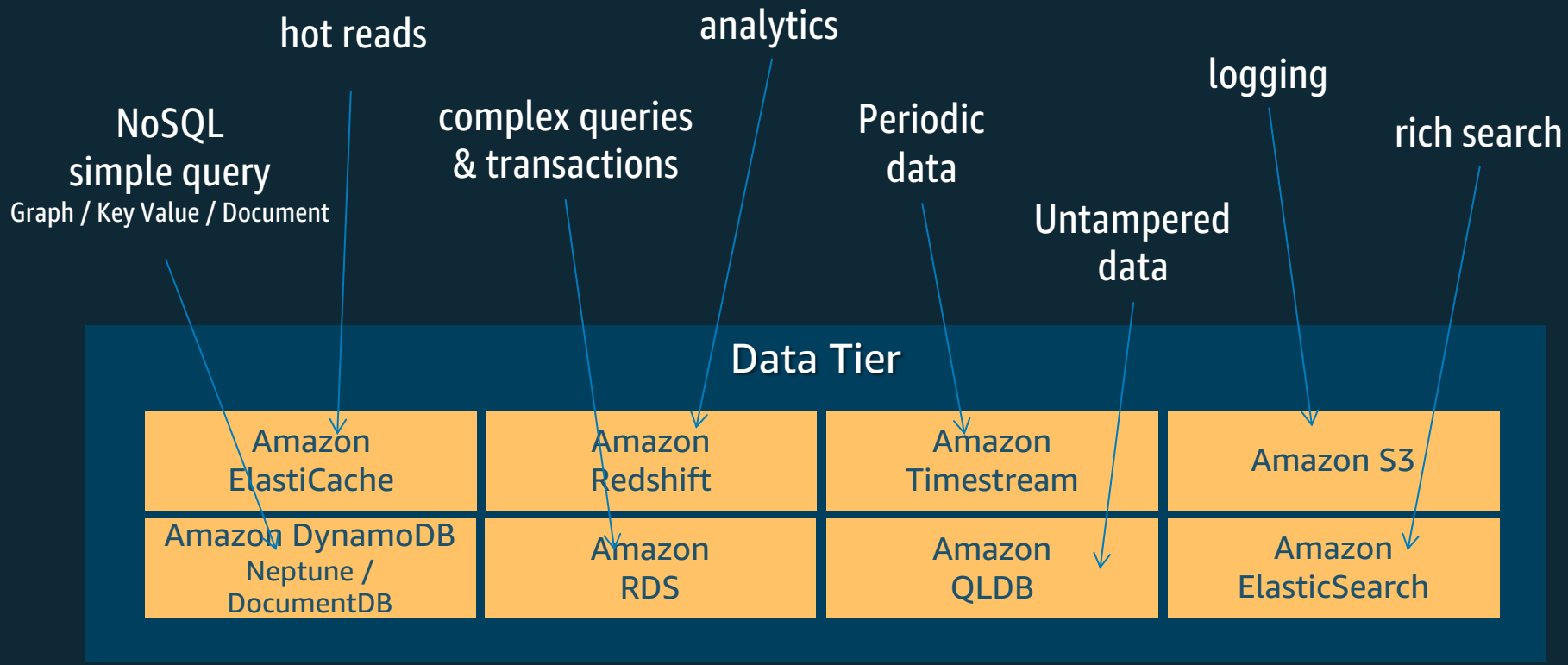
aws

Workload Driven Data Store Selection



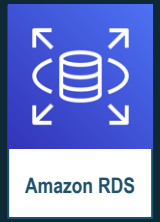
aws

AWS Database Services for the Data Tier

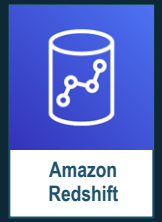


aws

AWS Database Services

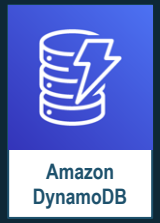


Managed Relational Database Service

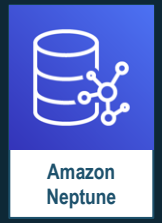


Petabyte-scale Data Warehouse

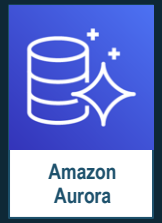
In-Memory Key Value Store



Fully Managed Key-Value and Document Database



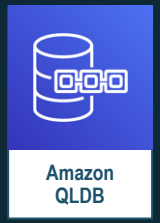
Fully Managed Graph Database



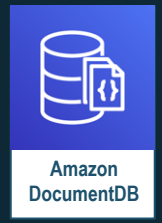
Cloud-Native Relational Database

[Preview]

Fully Managed Time Series Database



Fully Managed Ledger Database



MongoDB Compatible Document Database

aws

Amazon RDS

Managed relational database service with a choice of popular database engines

Amazon Aurora | MySQL | PostgreSQL | MariaDB | Microsoft SQL Server | ORACLE

Easy to administer

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

Performant & scalable

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

Available & durable

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

Secure and compliant

Data encryption at rest and in transit; industry compliance and assurance programs

aws

If you host your databases on-premises…

- App optimization
- Scaling
- High availability
- Database backups
- DB s/w patches
- DB s/w installs
- OS patches
- OS installation
- Server maintenance
- Rack & stack
- Power, HVAC, net

aws

If you host your databases in Amazon EC2…

- App optimization
- Scaling
- High availability
- Database backups
- DB s/w patches
- DB s/w installs
- OS patches
- OS installation
- Server maintenance
- Rack & stack
- Power, HVAC, net

- OS installation
- Server maintenance
- Rack & stack
- Power, HVAC, net

aws

aws

If you choose Amazon RDS…

- App optimization
- Scaling
- High availability
- Database backups
- DB s/w patches
- DB s/w installs
- OS patches
- OS installation
- Server maintenance
- Rack & stack
- Power, HVAC, net

- Scaling
- High availability
- Database backups
- DB s/w patches
- DB s/w installs
- OS patches
- OS installation
- Server maintenance
- Rack & stack
- Power, HVAC, net

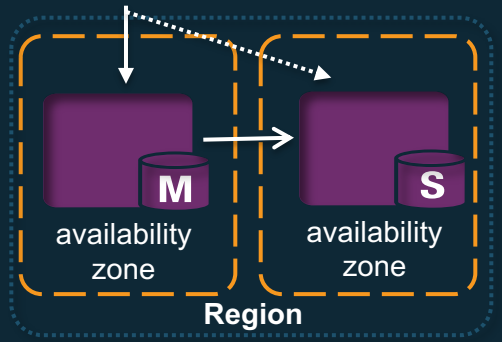
aws

aws

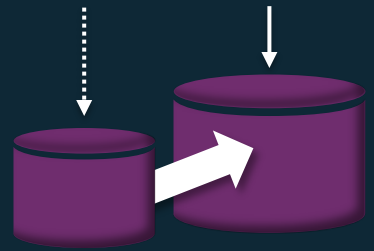
Key Amazon RDS Features

Amazon RDS Configuration	Improve Availability	Increase Throughput	Reduce Latency
Push-Button Scaling		✓	
Multi AZ	✓		
Read Replicas		✓	
Provisioned IOPS		✓	✓

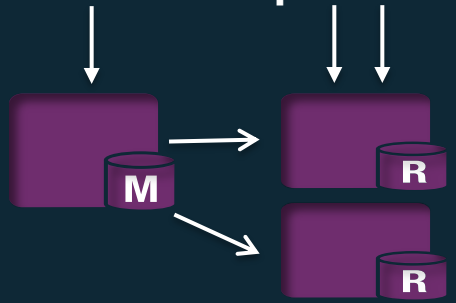
Multi-AZ



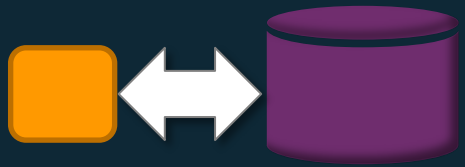
Push-Button Scaling



Read Replicas



Provisioned IOPS



aws

Amazon Aurora

MySQL and PostgreSQL compatible relational database built for the cloud

Performance and availability of commercial-grade databases at 1/10th the cost

Performance & scalability

5x throughput of standard MySQL and 3x of standard PostgreSQL; scale-out up to15 read replicas

Availability & durability

Fault-tolerant, self-healing storage; six copies of data across three AZs; continuous backup to S3

Highly secure

Network isolation, encryption at rest/transit

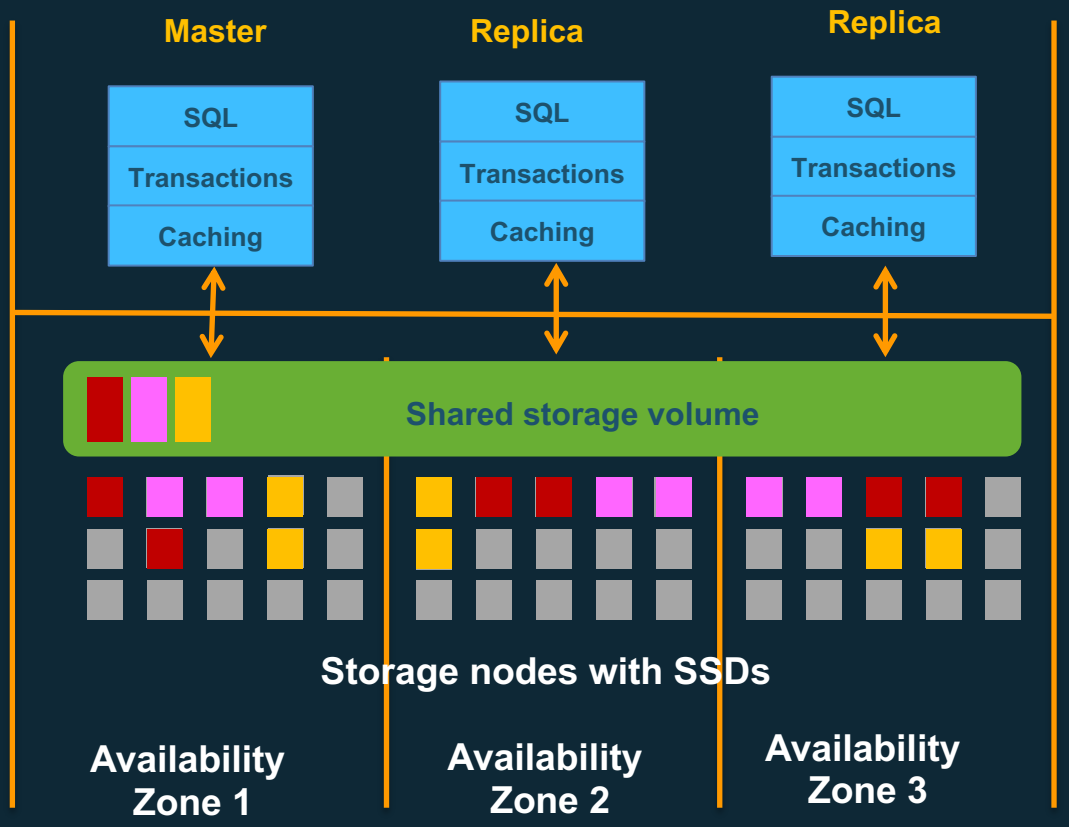
Fully managed

Managed by RDS: no hardware provisioning, software patching, setup, configuration, or backups

aws

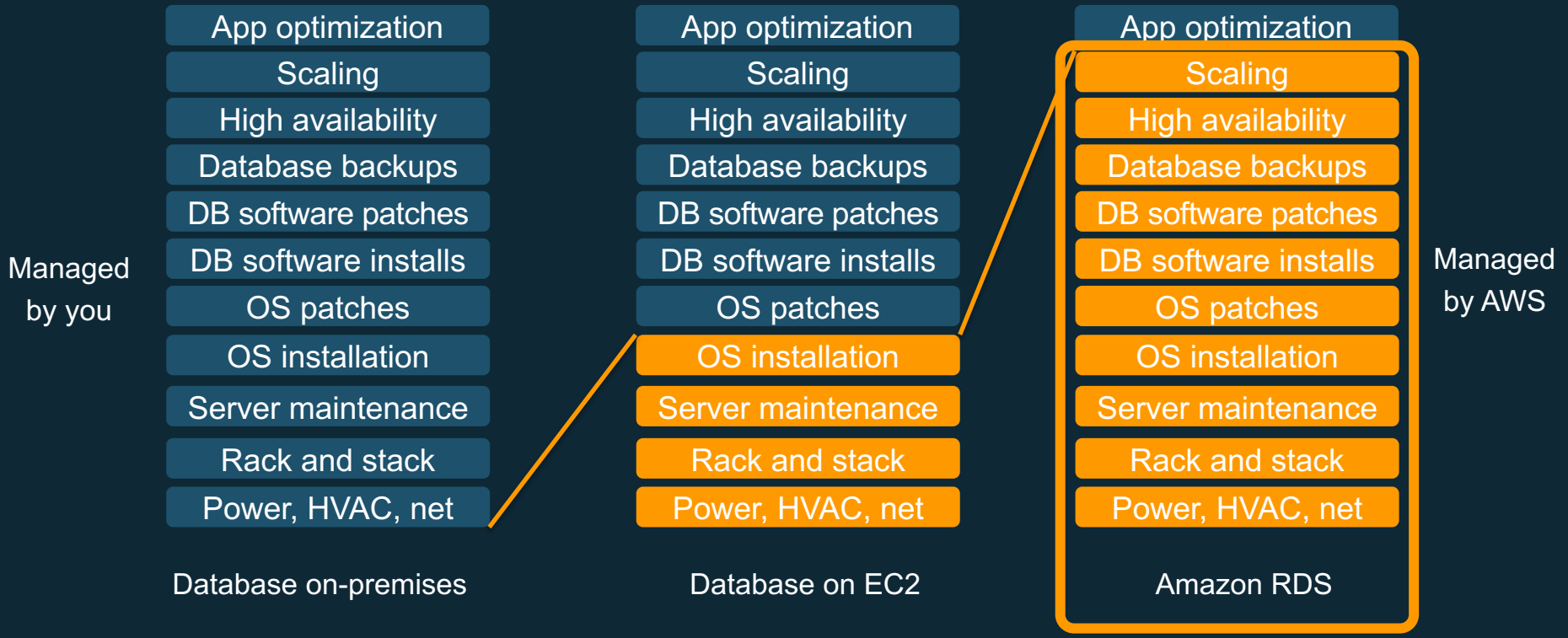
Scale-out, distributed, multi-tenant architecture

- Purpose-built log-structured distributed storage system designed for databases
  - Storage volume is striped across hundreds of storage nodes distributed over 3 different Availability Zones
  - Six copies of data, two copies in each Availability Zone to protect against AZ+1 failures
  - Master and replicas all point to the same storage



aws

Everything you get from Amazon RDS…



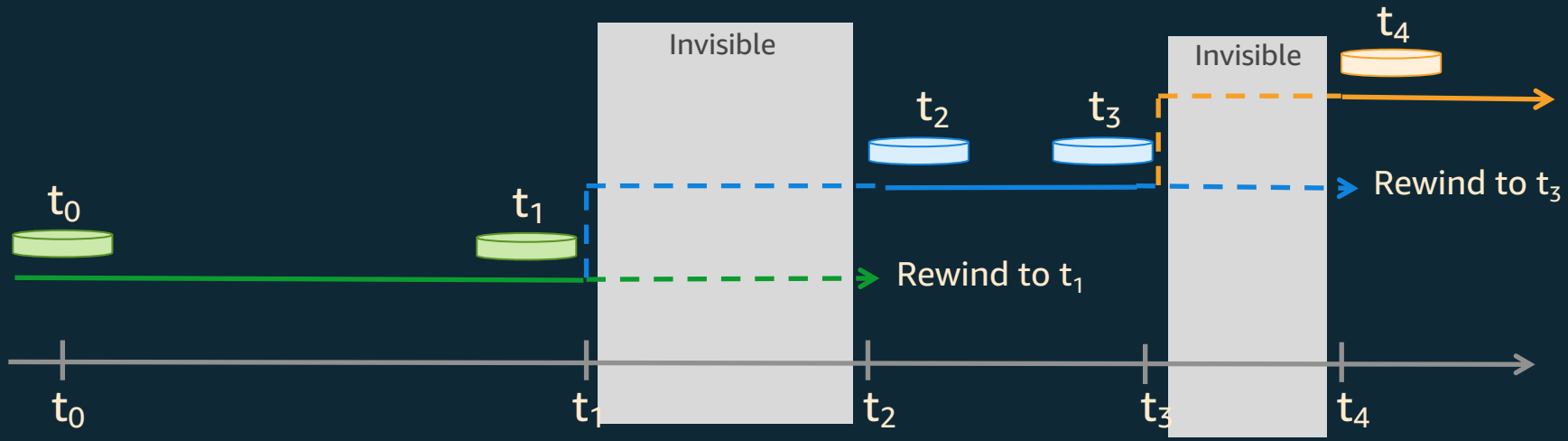
…and more

Up to 64TB of storage – auto-incremented in 10GB units

- Automatic storage scaling up to 64 TB—no performance impact
- Continuous, incremental backups to Amazon S3
- Instantly create user snapshots—no performance impact
- Automatic restriping, mirror repair, hot spot management, encryption

aws

Database backtrack

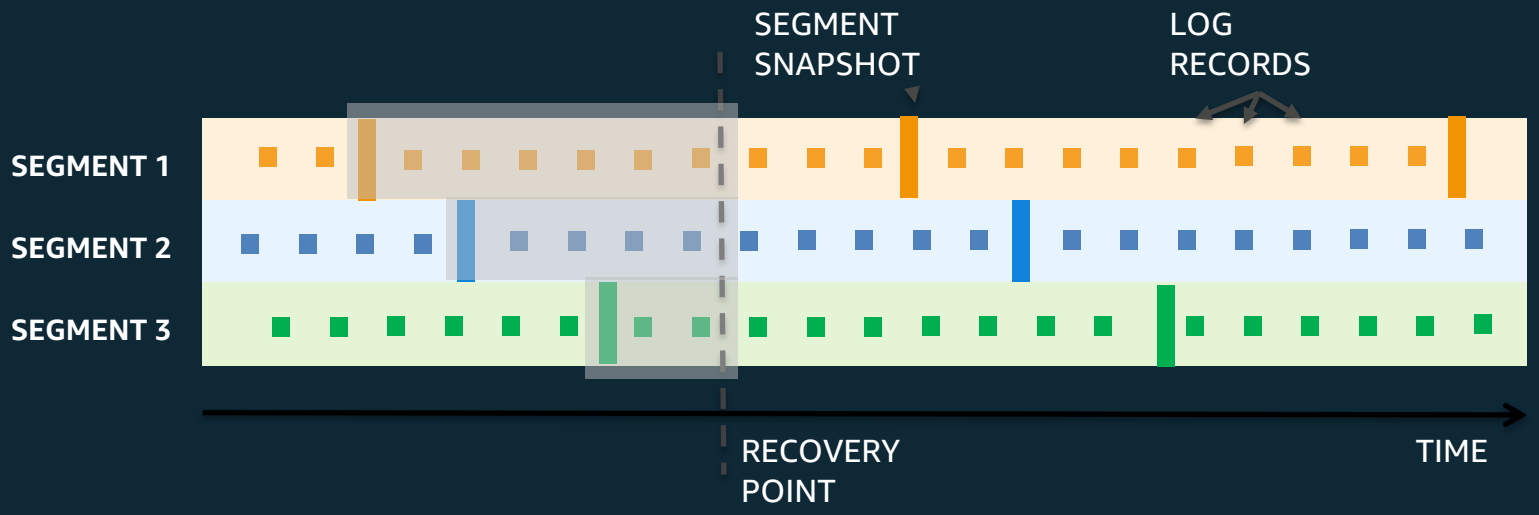


Backtrack brings the database to a point in time without requiring restore from backups

- Backtracking from an unintentional DML or DDL operation
- Backtrack is not destructive. You can backtrack multiple times to find the right point in time

aws

How does backtrack work?



We keep periodic snapshot of each segment; we also preserve the redo logs

For backtrack, we identify the appropriate segment snapshots

Apply log streams to segment snapshots in parallel and asynchronously

aws

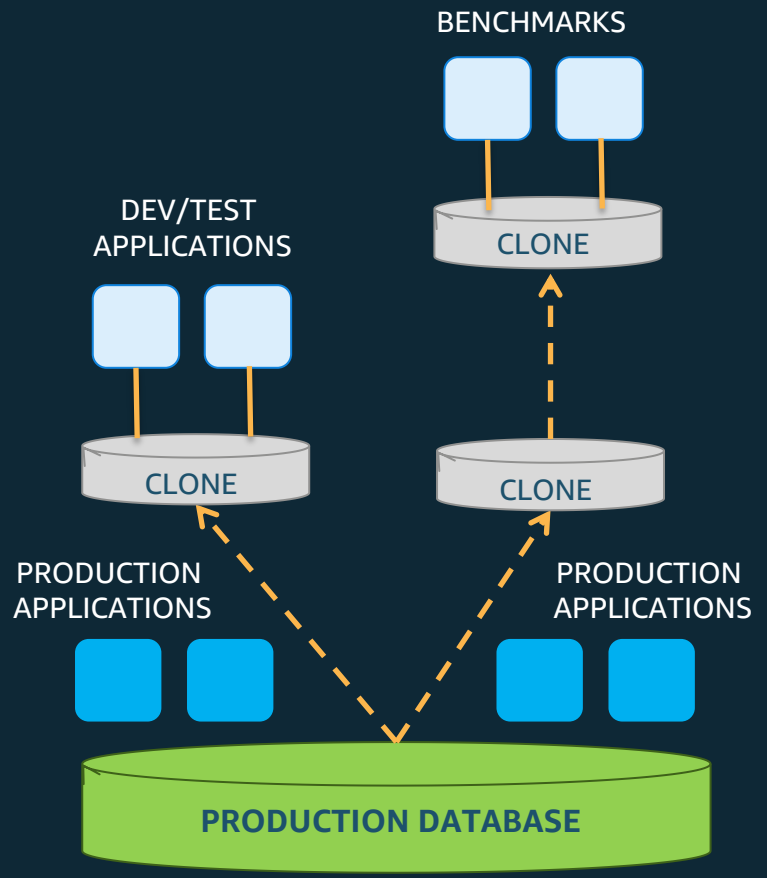
Fast database cloning

Clone database without copying data

- Creation of a clone is nearly instantaneous
- Data copy happens only on write – when original and cloned volume data differ

Example use cases

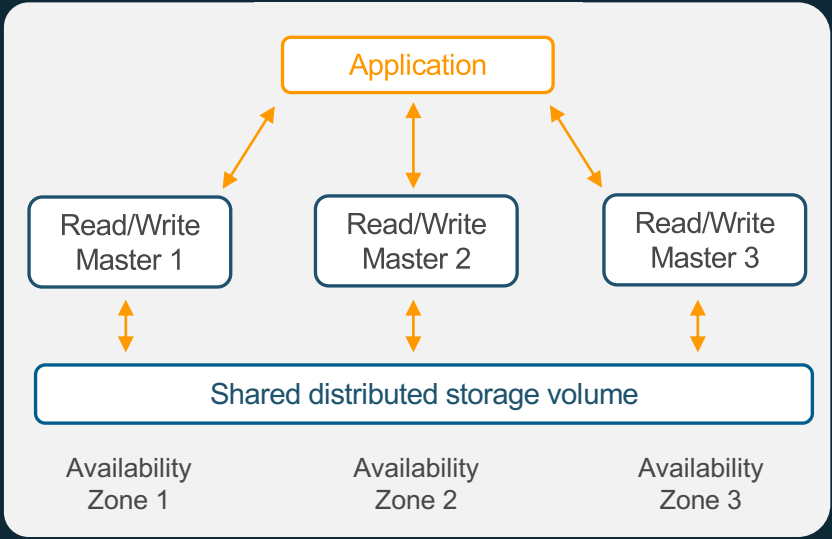
- Clone a production DB to run tests
- Reorganize a database
- Save a point in time snapshot for analysis without impacting production system.



Aurora Multi-Master

First relational database service with scale-out reads and writes across multiple data centers

Scale out both reads **and writes**



Zero application downtime from ANY instance failure

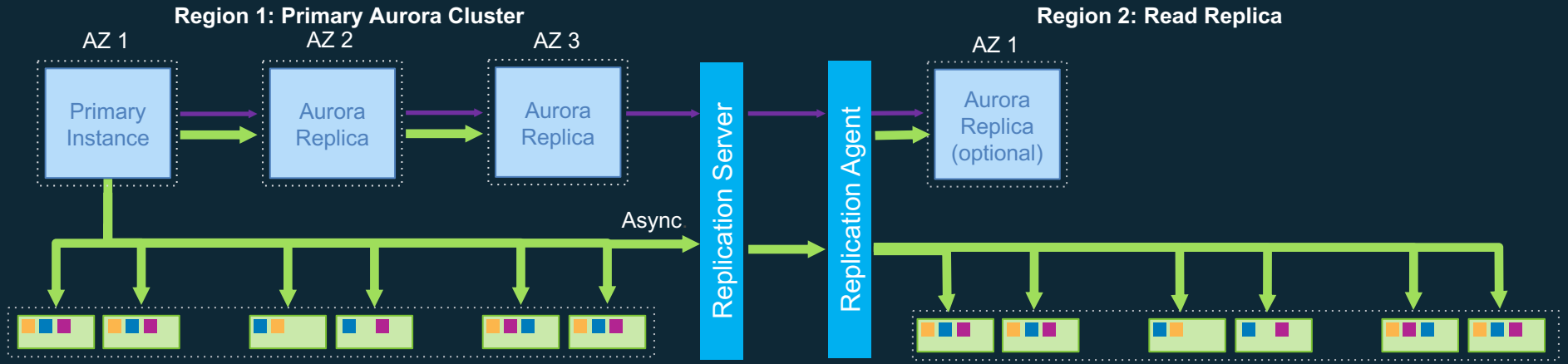
Zero application downtime from ANY AZ failure

Faster write performance and higher scale

Sign up for single-region multi-master preview today; multi-region multi-master **coming in 2019**

aws

Global database – physical replication



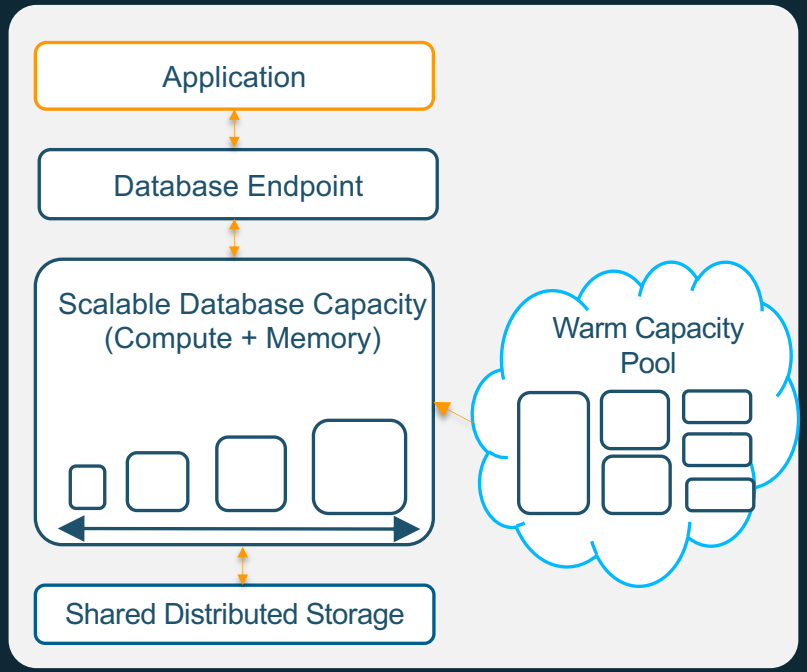
Consistently fast, low-lag, high-performance replication for global relational databases

- Global-scale replication in seconds or less
- Dedicated replication infrastructure ensures unconstrained performance
- Local reads, faster recovery, tighter DR objectives, and seamless cross-region migration

aws

Aurora Serverless

On-demand, auto-scaling database for applications with variable workloads

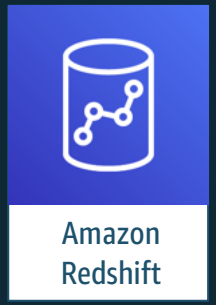


Starts up on demand, shuts down when not in use

Automatically scales with no instances to manage

Pay per second for the database capacity you use

aws



for as low as
$934/TB per year

Petabyte scale

Massively parallel

Columnar Store

Relational data warehouse

Fully managed = no admin

aws

Redshift cluster architecture

Massively parallel, shared nothing architecture

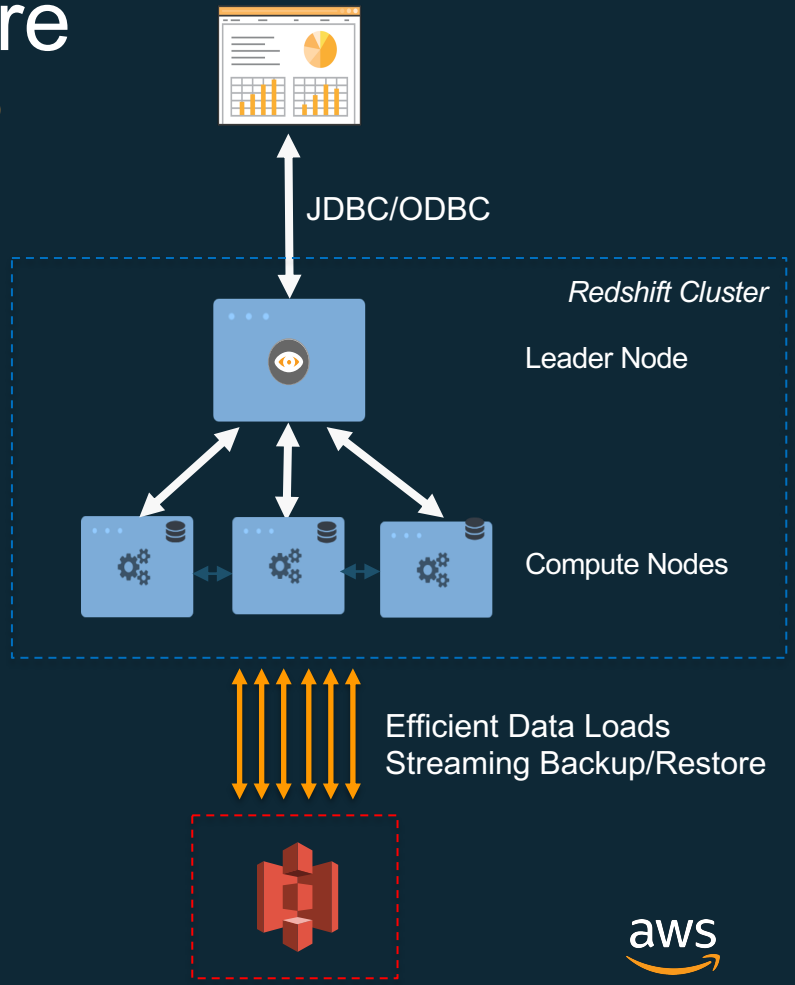
Streaming Backup/Restore from S3

Leader node

- SQL endpoint
- Stores metadata
- Coordinates parallel SQL processing

Compute nodes

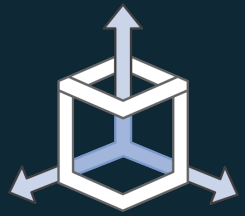
- Local, columnar storage
- Executes queries in parallel
- Load, backup, restore
- 2, 16, or 32 slices



Redshift Spectrum

Run SQL queries directly against data in S3 using thousands of nodes

High concurrency: Multiple clusters access same data



No ETL: Query data in-place using open file formats

Full Amazon Redshift SQL support

Fast at exabyte scale

Elastic and highly available

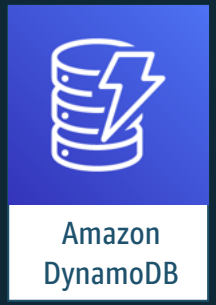
On-demand, pay-per-query

aws

Highlights

- Redshift is a managed data warehouse intended for analytics workloads
- Patching, backup/restore, and resize are fully managed by the service
- It uses a distributed, massively parallel architecture that scales horizontally to meet throughput requirements
- Redshift uses a c-store architecture, but still supports ANSI SQL including Transactions and Foreign Keys
- You can implement any type of data model on Redshift, but some types of data models scale better than others
- Redshift is extremely cost effective, and can offer similar performance for 1/10th the cost of Oracle, Teradata, or Netezza (as low as $1000/TB)

aws



NoSQL database

Seamless scalability

Zero admin

Single-digit millisecond latency

Multi-Master

Multi-Region

aws

Highly available and durable

Built for high durability

Designed to support
99.99%
of availability

WRITES

3-way replication
Persisted to disk
(Custom SSD)

READS

Strongly or eventually consistent
No latency trade-off

Data is always replicated to
three Availability Zones

aws

Backup and restore

The only cloud database to provide on-demand and continuous backups

On-demand backups for long-term data archival and compliance

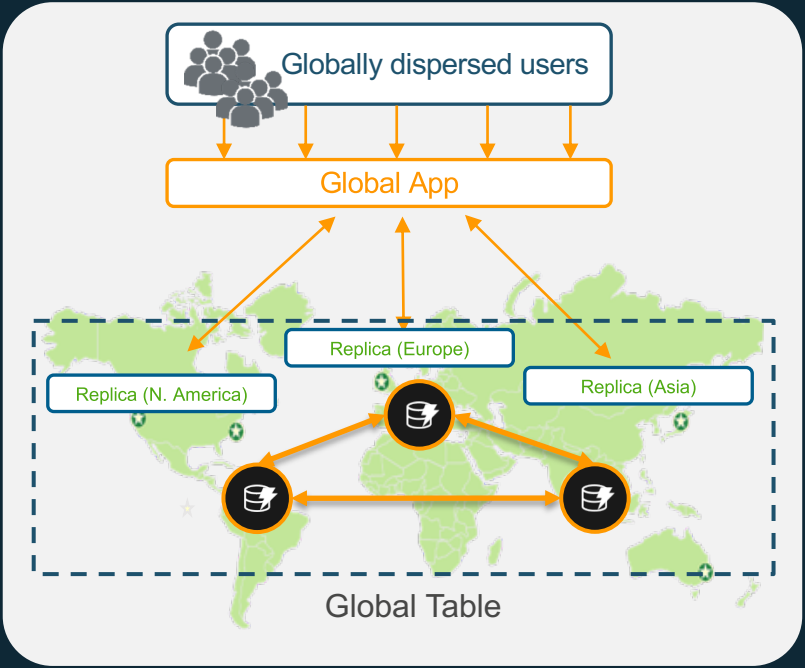
Point in time restore for short term retention and data corruption protection (35 days)

Point in time recovery with restore times in a few hours depending on table size

aws

Global Tables

The first fully-managed, multi-master, multi-region database

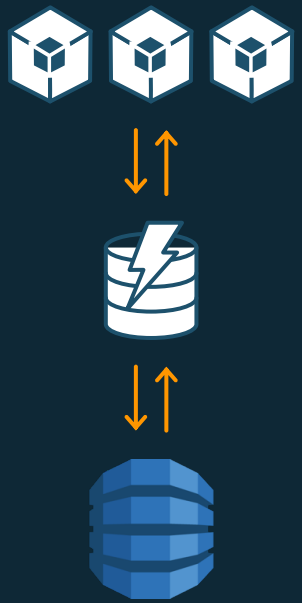


- Build high performance, globally distributed applications
- Low latency reads & writes to locally available tables
- Disaster proof with multi-region redundancy
- Easy to setup and no application re-writes required

aws

DynamoDB Accelerator (DAX)

High performance



Fully managed, highly available cache for DynamoDB

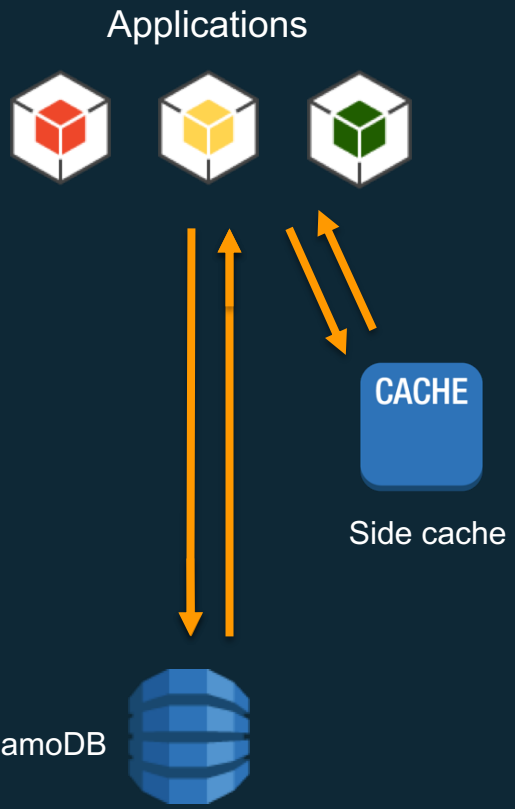
Even faster—microsecond latency

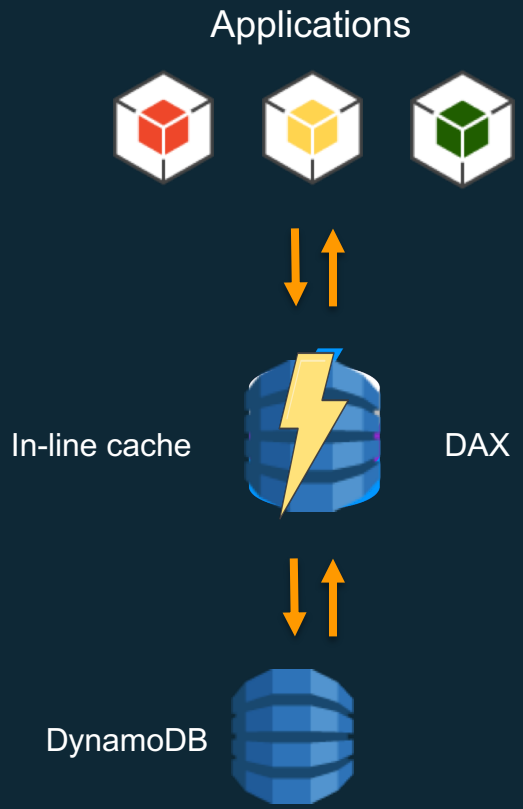
Scales to millions of requests per second

API compatible

aws

Traditional cache vs. DAX

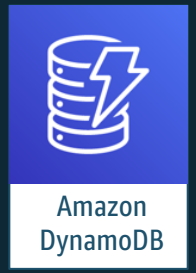




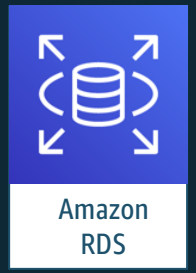
aws

NoSQL vs. SQL for a new app: how to choose?

- Want simplest possible DB management?
- Want app to manage DB integrity?
- Focus on performance and availability at any scale



- Need joins, transactions, frequent table scans?
- Want DB engine to manage DB integrity?
- Team has SQL skills?



aws

µs is the new *ms*

aws

Amazon ElastiCache

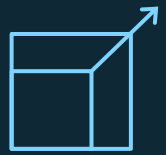
Fully-managed, Redis or Memcached compatible, low-latency, in-memory data store

Extreme Performance

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

Fully Managed

AWS manages all hardware and software setup, configuration, monitoring



Easily Scalable

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

aws

ElastiCache Redis

#1 Key-Value Store*

Fast in-memory data store in the cloud. Use as a database, cache, message broker, queue

Highly Available & Reliable

Read replicas, multiple primaries, multi-AZ with automatic failover

Fully Managed & Hardened

AWS manages hardware, software, setup, configuration, monitoring, failure recovery, and backups

Easily Scalable

Cluster with up to 6.1 TiB of in-memory data

Read scaling with replicas

Write and memory scaling with sharding

Scale out or in

Secure & Compliant

VPC for cluster isolation, encryption at rest/transit, HIPAA compliance

*: https://db-engines.com/en/ranking

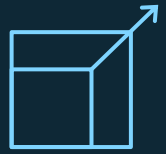
ElastiCache Memcached

Fully Managed Memcached

Fast in-memory data store in the cloud. Use as a cache to reduce latency and improve throughput

Secure & Hardened

VPC for cluster isolation



Easily Scalable

Sharding to scale in-memory cache with up to 20 nodes and 8.14 TiB per cluster

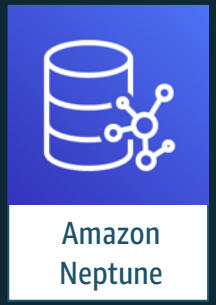
aws

Choosing between Redis and Memcached

	Memcached	Redis
Sub-millisecond latency	Yes	Yes
Developer ease of use	Yes	Yes
Data partitioning	Yes	Yes
Support for a broad set of programming languages	Yes	Yes
Advanced data structures	-	Yes
Multithreaded architecture	Yes	-
Snapshots	-	Yes
Replication	-	Yes
Transactions	-	Yes
Pub/Sub	-	Yes
Lua scripting	-	Yes
Geospatial support	-	Yes

When to use ElastiCache vs DAX?

ElastiCache	DAX
Redis/Memcached Open Source	Easy Cache compatible with DynamoDB API
Cache Compatible with all Databases	DynamoDB only
6+ TiB (6000+ GiB)	488 Gib (vertically scaled)
200+ Commands	Get, Put, Update, Query, Scan
Data Structures (String, Lists, Sets, Sorted Sets, Hashes, Bit Arrays, HyperLogLogs)	DynamoDB Data Types
Advanced Eviction Policies No Eviction, allkeys-lru, volatile-lru, allkeys-random, volatile-random, volatile-ttl	Time-to-Live Cache (TTL), Least Recently Used (LRU), Write-Through Eviction
More Control on Cache Content	Plug & Play Cache
Durability (RoadMap 2018)	Durability with DynamoDB
Cost for Cache (cheaper)	Cost for Cache (more expensive)



Fully managed graph database

Supports open graph APIs

Scalable

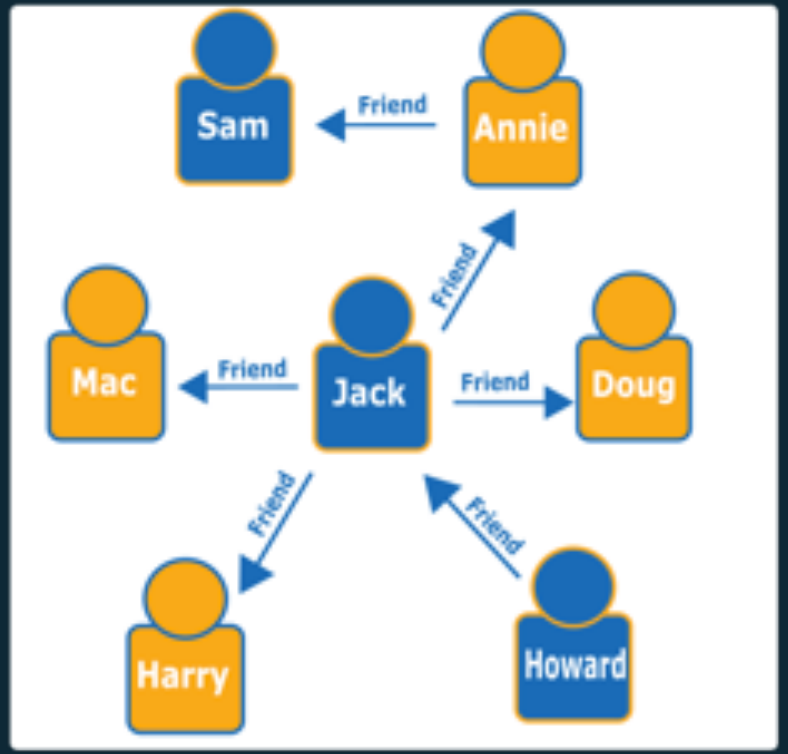
ACID compliant

Multi-AZ

aws

Use cases for highly connected data

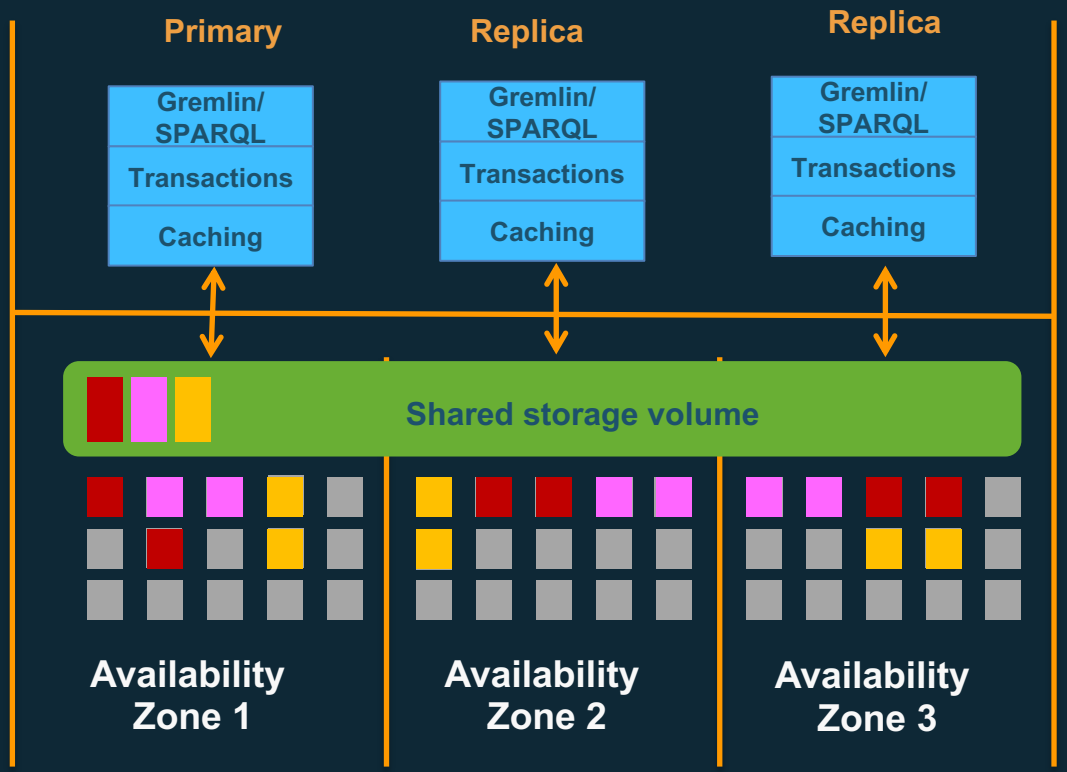
- Social networking
- Recommendations
- Knowledge graphs
- Fraud detection
- Life sciences
- Network and IT operations



aws

Distributed storage architecture

- Performance, availability, durability
- Scale-out replica architecture
- Shared storage volume with 10 GB segments striped across hundreds of nodes
- Data is replicated 6 times across 3 AZs
- Hotspot rebalance, fast database recovery
- Log applicator embedded in storage layer
  - Ship only the log
  - Less work on engine
  - Minimizes network traffic

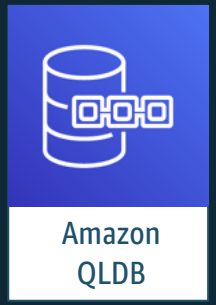


Delivered as a managed service

aws

Fully managed service

- Easily configurable via the console
- Multi-AZ high availability
- Up to 15 read replicas
- Supports encryption at rest
- Supports encryption in transit (TLS)
- Backup and restore, point-in-time recovery



Fully managed ledger database

Immutable and transparent

Cryptographically verifiable

Scalable

Serverless

aws

Common use cases

Banking & Finance
Keeping track of transactions, trades and accounts

E-Commerce
Where's my stuff?

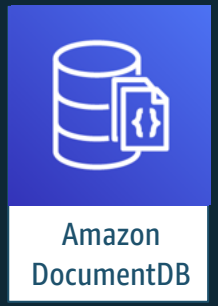
Transport & Logistics
Tracking transportation of goods

HR & Payroll
Tracking changes to an individual's profile

Manufacturing
Recording components used in manufacturing

Government
Tracking vehicle title history

aws



Fully managed document database

MongoDB-compatible

Multi-AZ

Performance at scale

aws

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

aws

Use cases for document databases

Content Management

Mobile

Personalization

Catalog

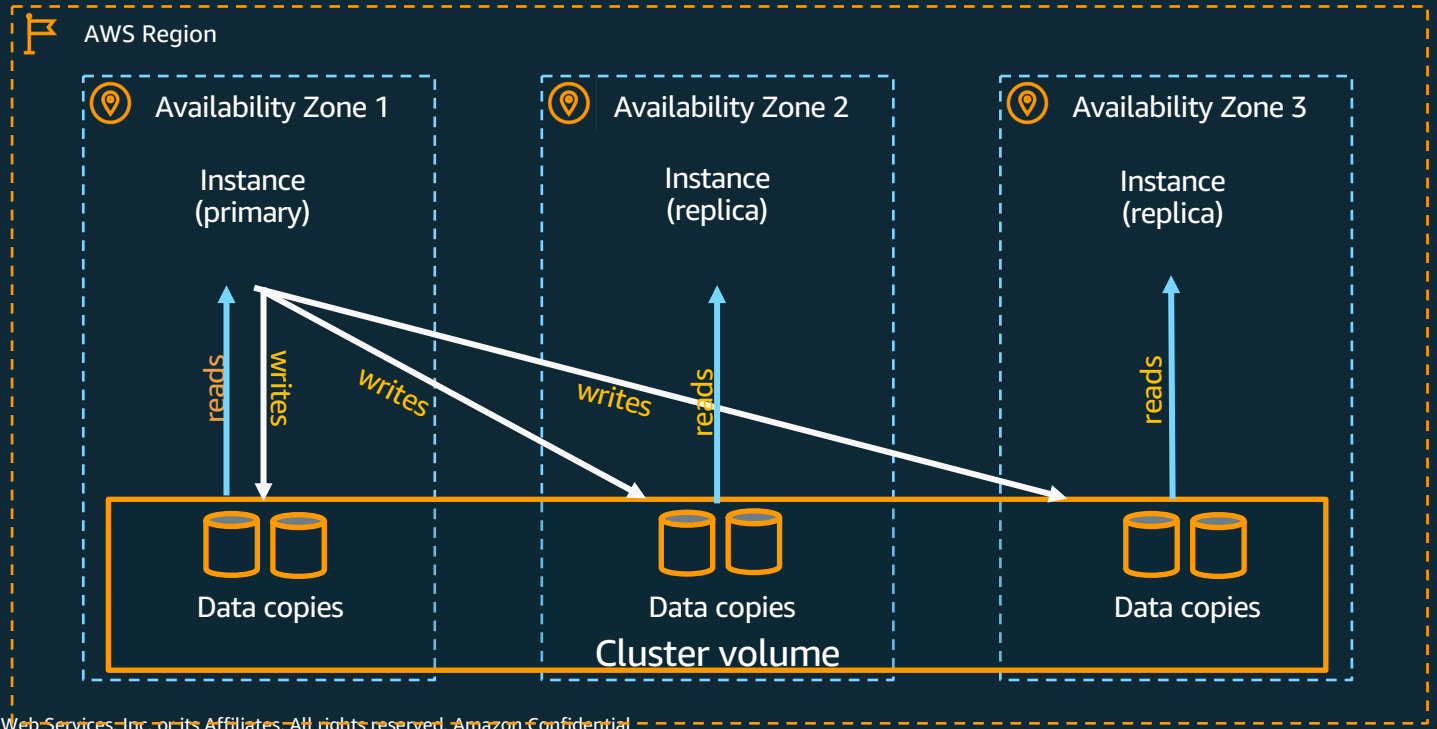
Retail and Marketing

User profiles

aws

DocumentDB Architecture

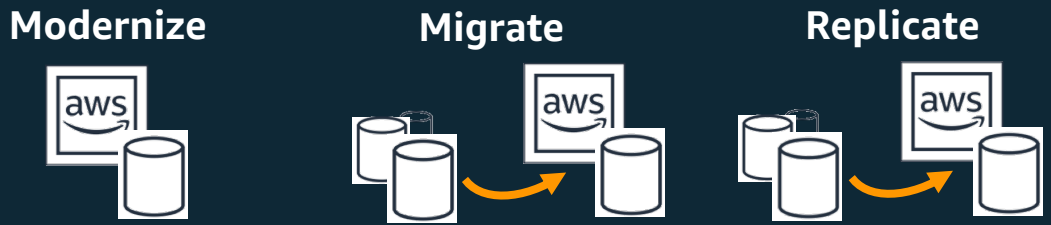
Separate compute and storage provide 2x throughput of current MongoDB managed services



Database Migration Service & Schema Conversion Tool

AWS Database Migration Service (DMS) easily and securely migrates and/or replicate your databases *and* data warehouses to AWS

AWS Schema Conversion Tool (SCT) converts your commercial database and data warehouse schemas to open-source engines or AWS-native services, such as Amazon Aurora and Redshift

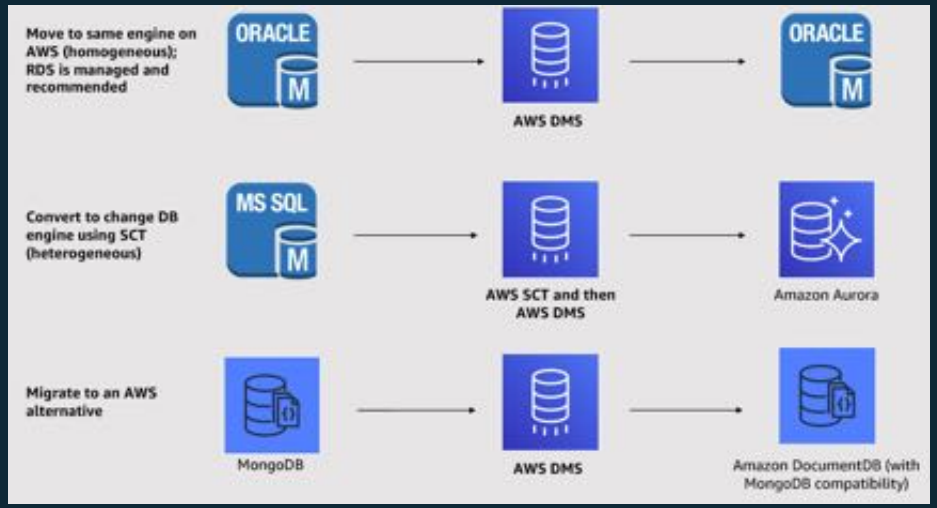


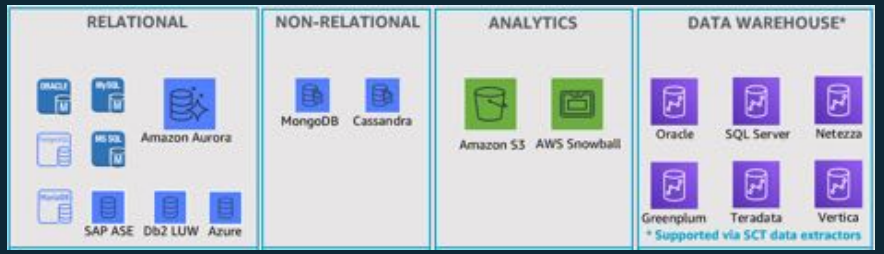
aws

Flexible, powerful migration tooling

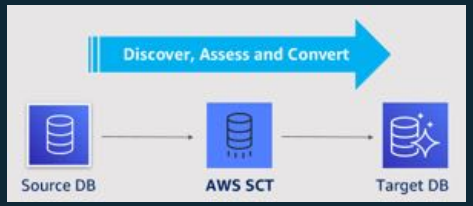
Most sources and targets, higher conversion automation

Database Migration Service (DMS)





Schema Conversion Tool (SCT)



For heterogeneous migrations (e.g., Oracle to Aurora PostgreSQL or Aurora MySQL, and Cassandra to Amazon DynamoDB)

aws

AWS database services

Purpose-built databases, the right tool for the right job

Relational

- Aurora
  - MySQL
  - PostgreSQL
- RDS
  - MySQL
  - PostgreSQL
  - MariaDB
  - ORACLE
  - Microsoft SQL Server

Non-relational

- DynamoDB — Key value
- ElastiCache — In-memory
- Neptune — Graph
- DocumentDB — Document
- Timestream — Time series
- QLDB — Ledger

Database Migration Service