



Databases on AWS

AWS Certified Solutions Architect – Associate
July, 2021

Karish Chowdhury, Enterprise Solutions Architect
karishch@amazon.com



Agenda

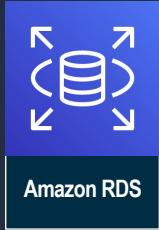
- AWS Database Services Overview
- Traditional vs. AWS Data services model
- Relational Databases on AWS
- Quiz!
- NoSQL Databases on AWS
- Quiz!
- Summary & Sample Exam questions
- Wrap-up
- Additional Topics

AWS Database Services Overview

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



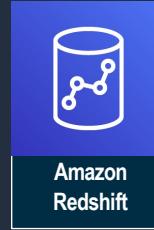
AWS Database Services



Managed Relational
Database Service



Cloud-Native
Relational Database



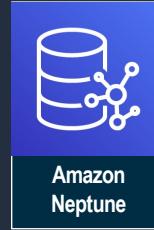
Petabyte-scale Data
Warehouse



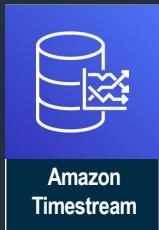
Fully Managed Key-
Value and Document
Database



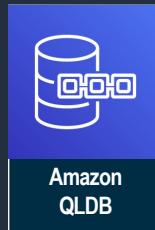
In-Memory Key
Value Store



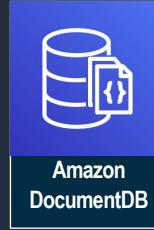
Fully Managed
Graph Database



Fully Managed Time
Series Database



Fully Managed
Ledger Database



MongoDB
Compatible
Document Database

Traditional vs. AWS Data services model

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



Traditional Database Architecture



Traditional Database Architecture

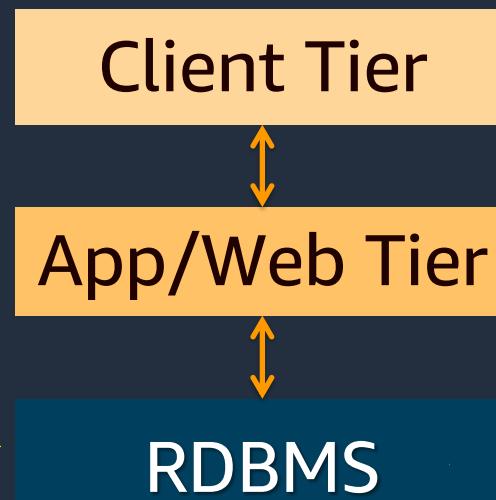
Key-value access

Complex queries

OLAP transactions

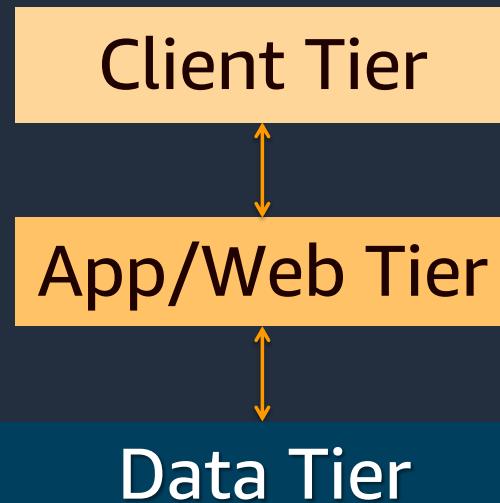
Analytics

*All forced into
the relational
database*



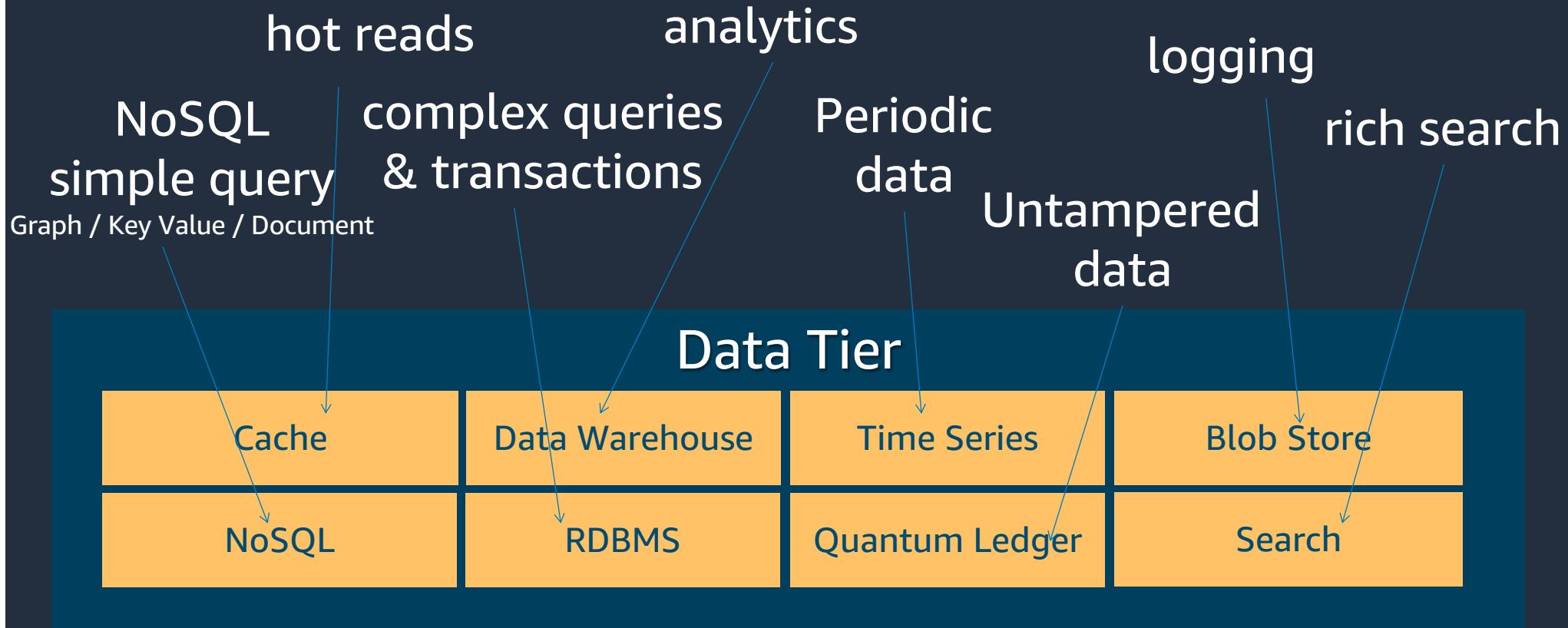
AWS Data Tier Architecture

On AWS choose best database service for each workload

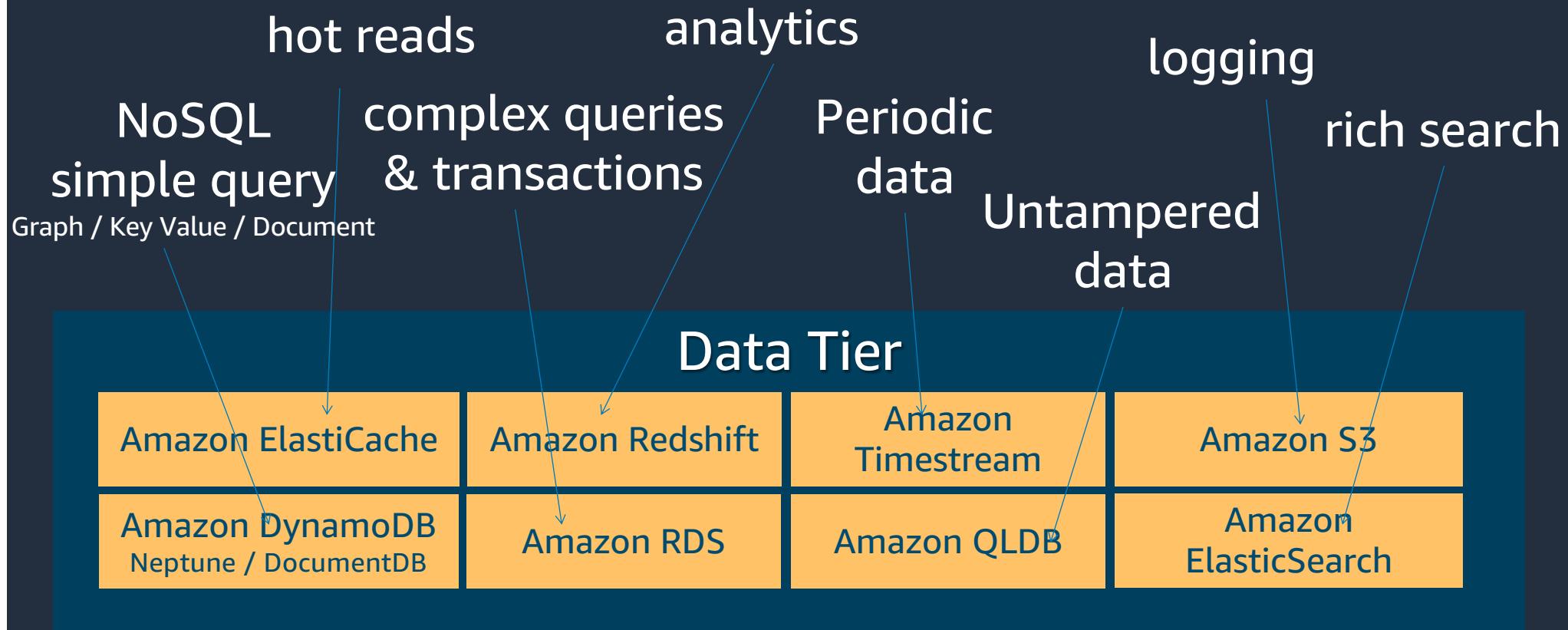


Cache	Data Warehouse	Time Series	Blob Store
NoSQL	RDBMS	Quantum Ledger	Search

Workload Driven Data Store Selection



AWS Database Services for the Data Tier



Relational Databases on AWS

Amazon RDS

Amazon Aurora

Amazon Redshift





Amazon RDS

Amazon Relational Database Service (RDS)

- Managed Relational Database

Amazon RDS

Managed relational database service with a choice of popular database engines

Amazon
Aurora



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

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



you

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



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



you

OS installation

Server maintenance

Rack & stack

Power, HVAC, net



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

you



Scaling

High availability

Database backups

DB s/w patches

DB s/w installs

OS patches

OS installation

Server maintenance

Rack & stack

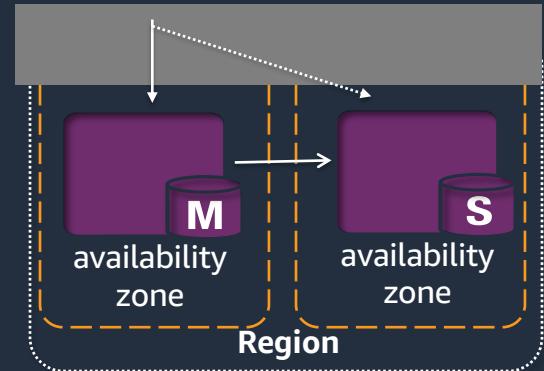
Power, HVAC, net



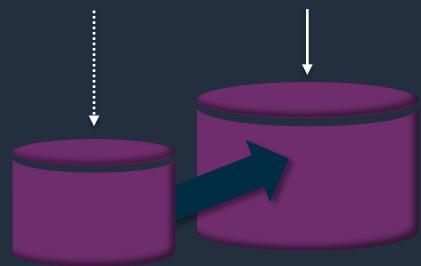
Key Amazon RDS Features

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

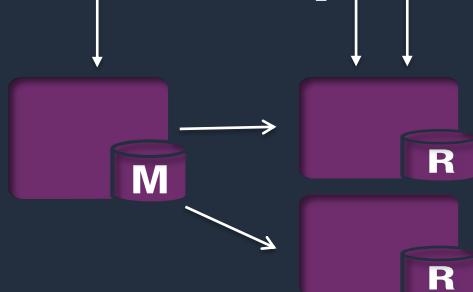
Multi-AZ



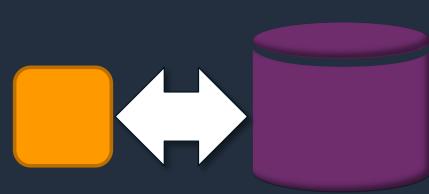
Push-Button Scaling

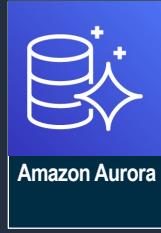


Read Replicas



Provisioned IOPS





Cloud-Native Relational Database

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 to 15 read replicas



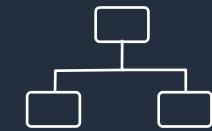
Availability & durability

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



Highly secure

Network isolation, encryption at rest/transit

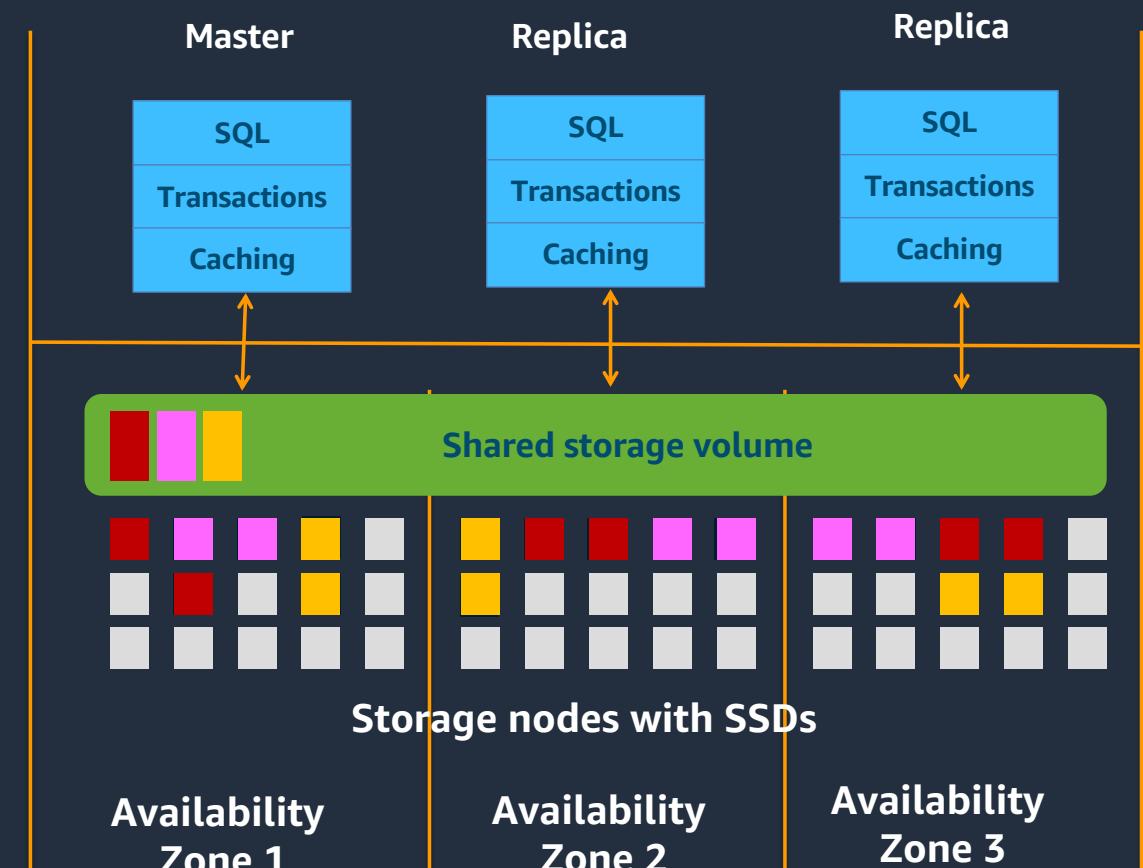


Fully managed

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

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



Everything you get from Amazon RDS...

Managed
by you

- App optimization
- Scaling
- High availability
- Database backups
- DB software patches
- DB software installs
- OS patches
- OS installation
- Server maintenance
- Rack and stack
- Power, HVAC, net

Database on-premises

- App optimization
- Scaling
- High availability
- Database backups
- DB software patches
- DB software installs
- OS patches
- OS installation
- Server maintenance
- Rack and stack
- Power, HVAC, net

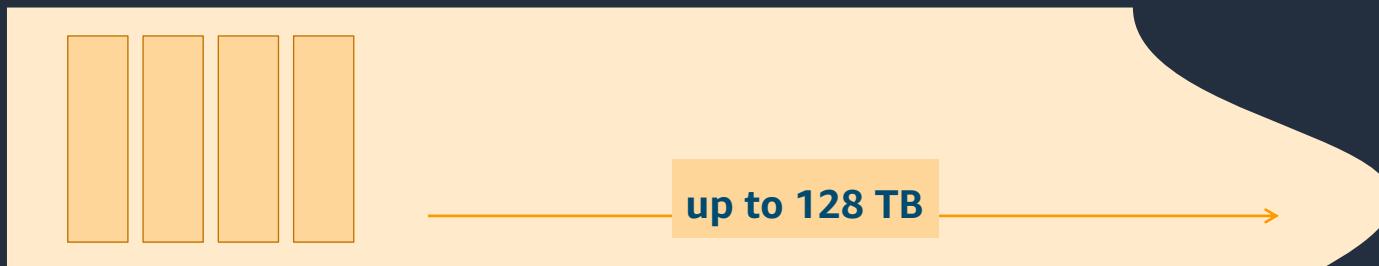
Database on EC2

- App optimization
- Scaling
- High availability
- Database backups
- DB software patches
- DB software installs
- OS patches
- OS installation
- Server maintenance
- Rack and stack
- Power, HVAC, net

Amazon RDS

Managed
by AWS

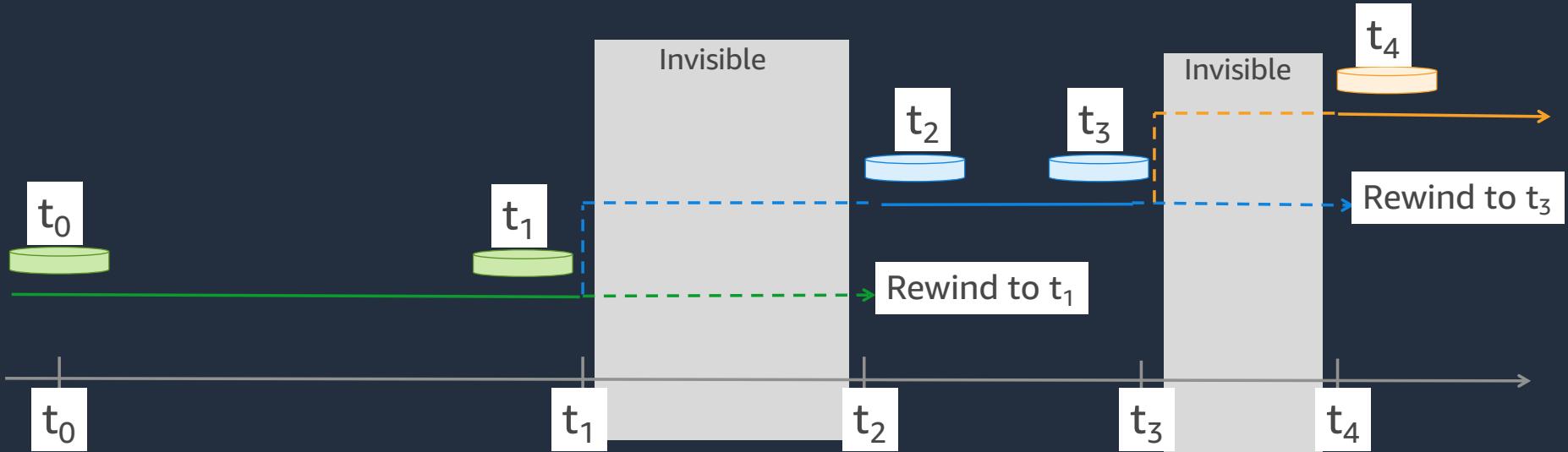
...and more



Up to 128 TB of storage – auto-incremented in 10GB units

- Automatic storage scaling up to 128 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

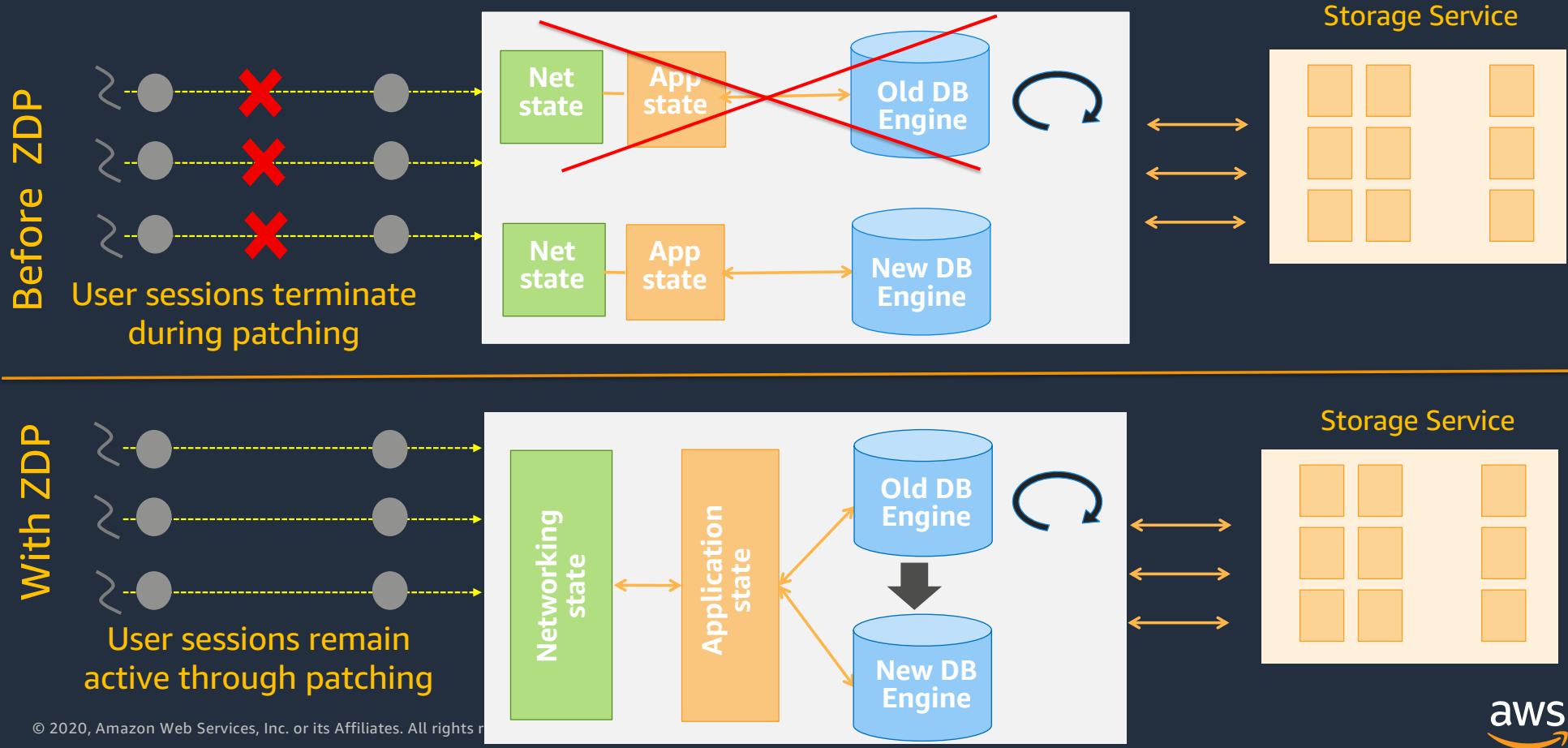
Database backtrack



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

Zero downtime patching



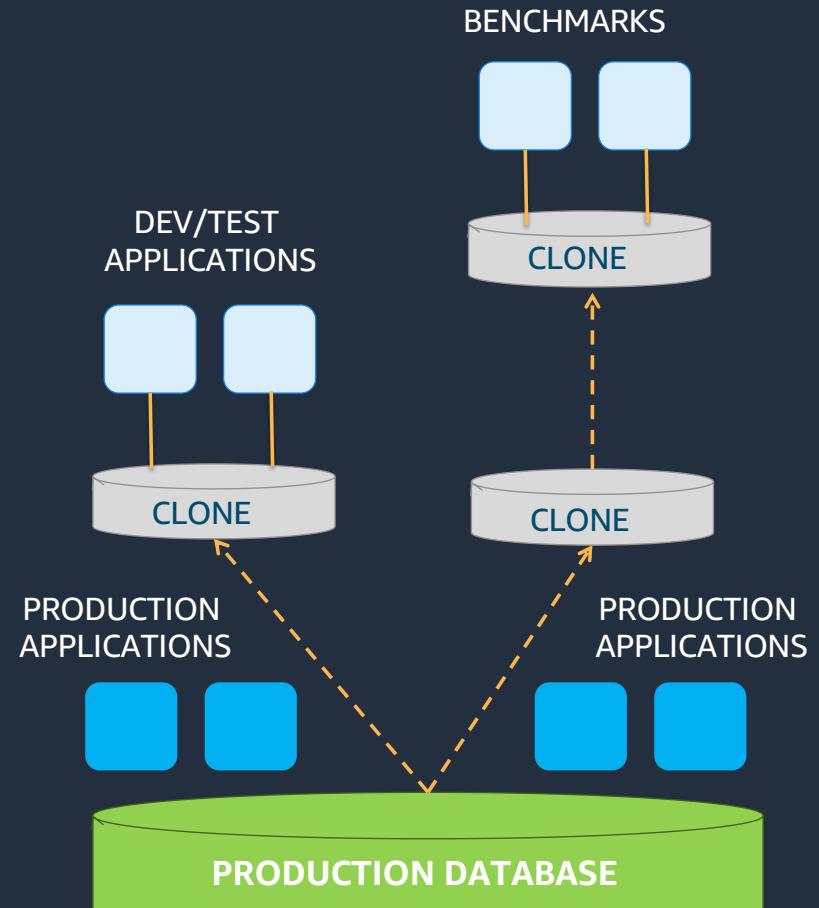
Fast database cloning

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

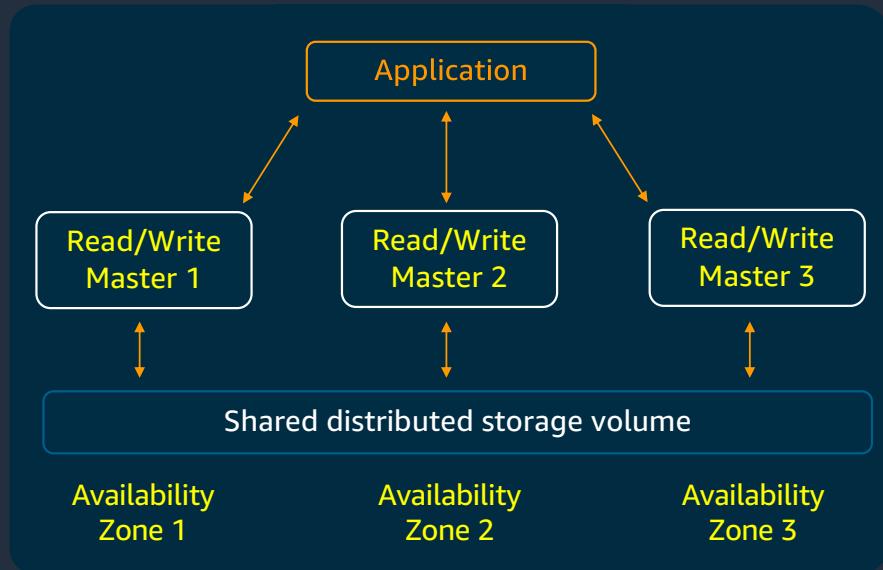
- Clone a production DB to run tests
- Reorganize a database
- Save a point in time snapshot for analysis without impacting production system.



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

Global database

Faster disaster recovery and enhanced data locality

Promote read-replica to a master
for faster recovery **in the event
of disaster**

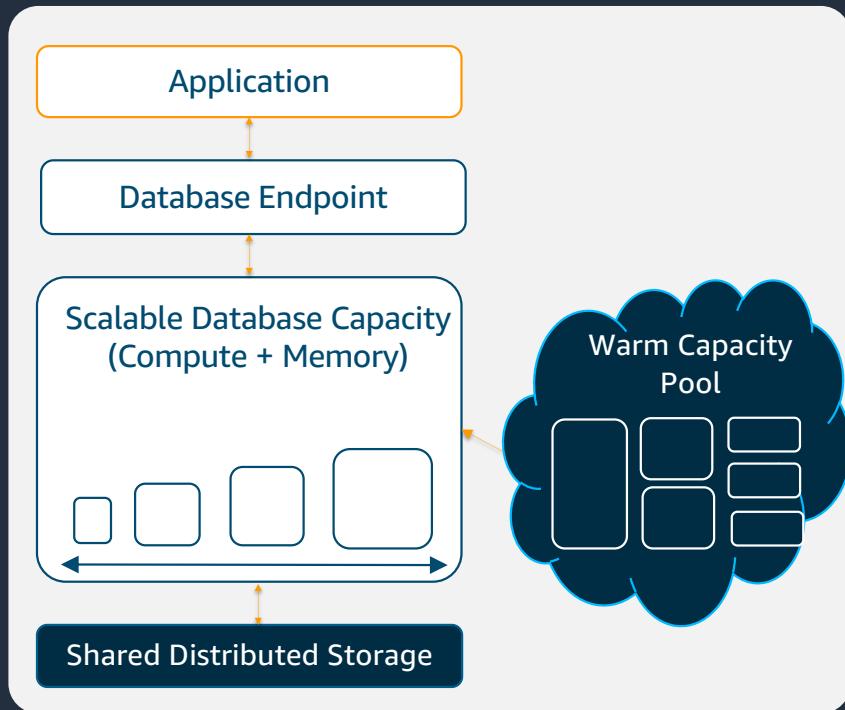
Bring data close to your
customer's applications in
different regions

Promote to a master for **easy
migration**



Aurora Serverless

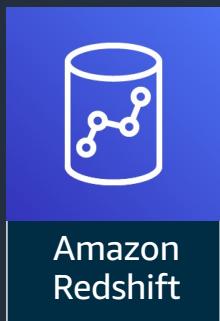
On-demand, auto-scaling database for applications with variable workloads



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



Petabyte scale

Massively parallel

Columnar Store

Relational data warehouse

Fully managed = no admin

*for as low
as
\$934/TB
per year*

A green curved arrow points from the word "parallel" towards the italicized price information.

Amazon Redshift – Data Warehousing

*Fast, powerful, and simple data warehousing at 1/10 the cost
Massively parallel, petabyte scale*

Fast



Columnar storage technology to improve I/O efficiency and parallelize queries. Data load scales linearly.

Inexpensive



As low as \$1000 per terabyte per year, 1/10th the cost of traditional data warehouse solutions

Scalable



Resize your cluster up and down as your performance and capacity needs change

Secure



Data encrypted at rest and transit. Isolate clusters with VPC. Manage your own keys with KMS

Redshift cluster architecture

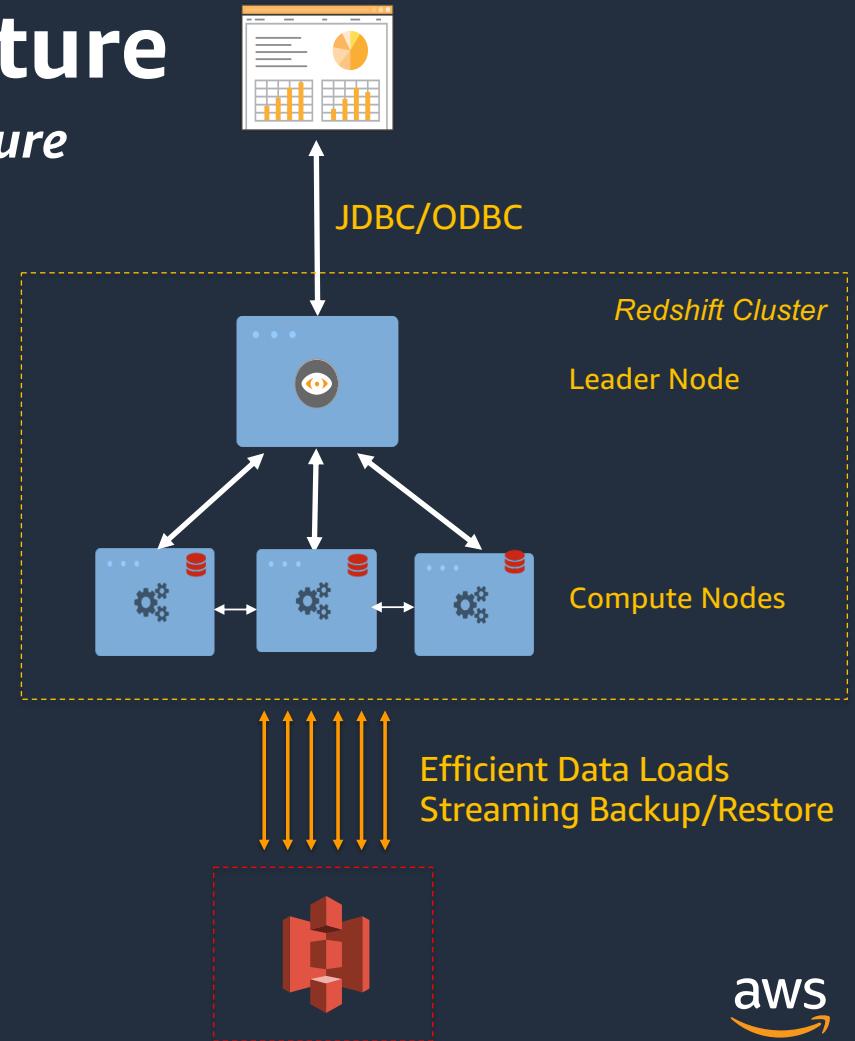
*Massively parallel, shared nothing architecture
Streaming Backup/Restore from S3*

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

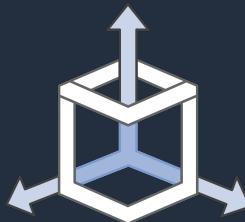


Redshift Spectrum

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



High concurrency: Multiple clusters access same data



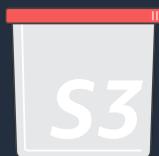
No ETL: Query data in-place using open file formats



Full Amazon Redshift SQL support



Fast at exabyte scale



Elastic and highly available



On-demand, pay-per-query

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)

Knowledge Check

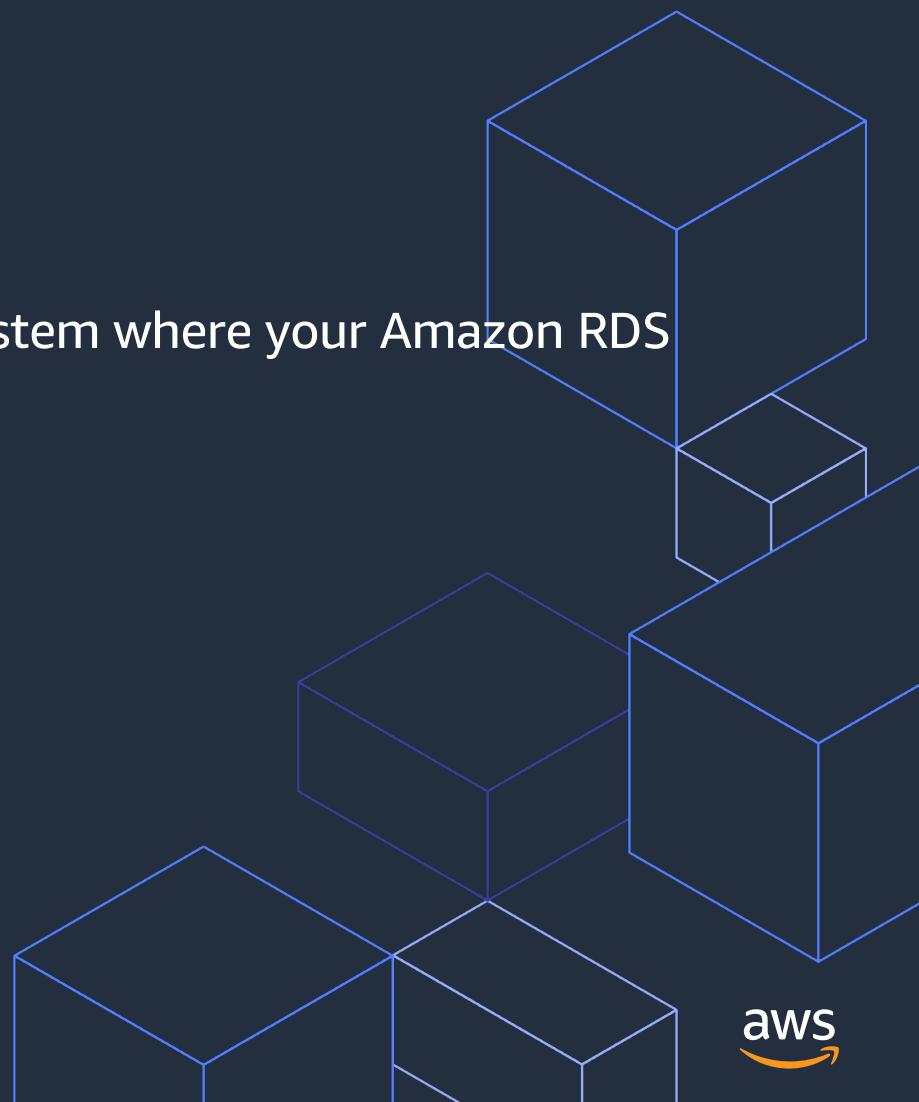
© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Poll #1

You can SSH into and control the operating system where your Amazon RDS MySQL instance is running.

- True
- False



Poll #2

Under what circumstances would I choose provisioned IOPS over standard storage when creating an RDS instance?

- A. For Test Database Environment
- B. If you have workloads that are not sensitive to latency/lag
- C. If you need to run an I/O-Intensive relational database for mission critical application in Production
- D. If your Business was trying to save money

Poll #3

Which AWS service is ideal for Business Intelligence Tools/Data Warehousing?

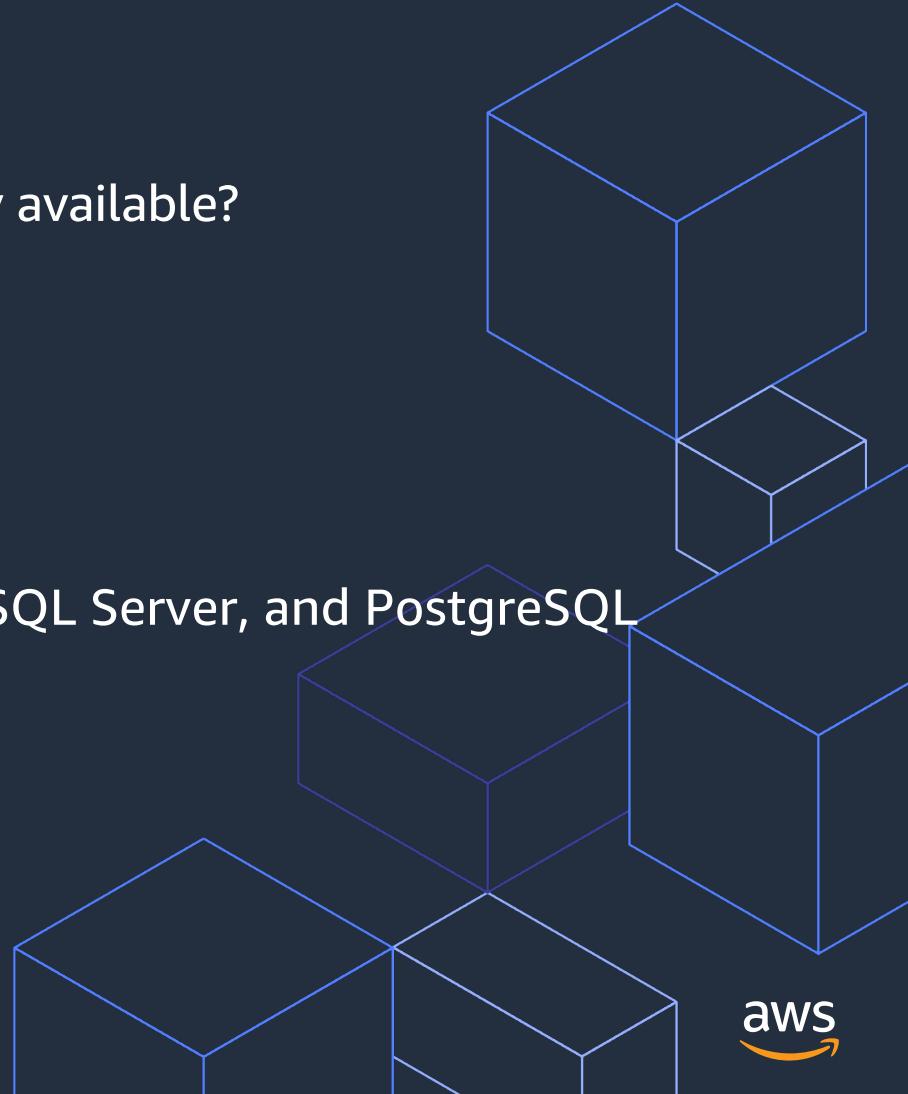
- A. DynamoDB
- B. ElastiCache
- C. Redshift
- D. Elastic Beanstalk



Poll #4

Which set of RDS database engines is currently available?

- A. Aurora, MySQL, SQL Server, Cassandra
- B. PostgreSQL, MariaDB, MongoDB, Aurora
- C. Amazon Aurora, MySQL, MariaDB, Oracle, SQL Server, and PostgreSQL
- D. MariaDB, SQL Server, MySQL, Cassandra

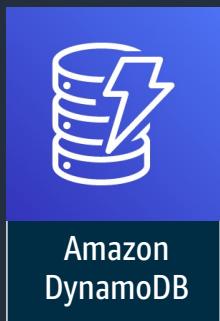


NoSQL Databases on AWS

Amazon DynamoDB
Amazon ElastiCache
Amazon Neptune
Amazon Timestream
Amazon QLDB
Amazon DocumentDB

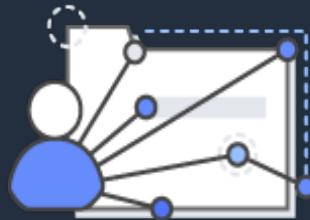
© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.





NoSQL database
Seamless scalability
Zero admin
Single-digit millisecond latency
Multi-Master
Multi-Region

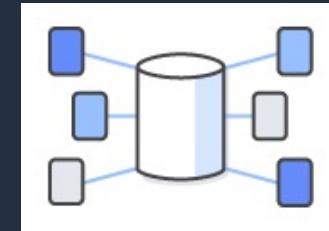
Amazon DynamoDB



Fully managed



Consistently fast at any scale



Highly available
and durable



Secure



Integrates with AWS Lambda,
Amazon Redshift, and more



Cost-effective

Highly available and durable



Backup and Restore

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



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



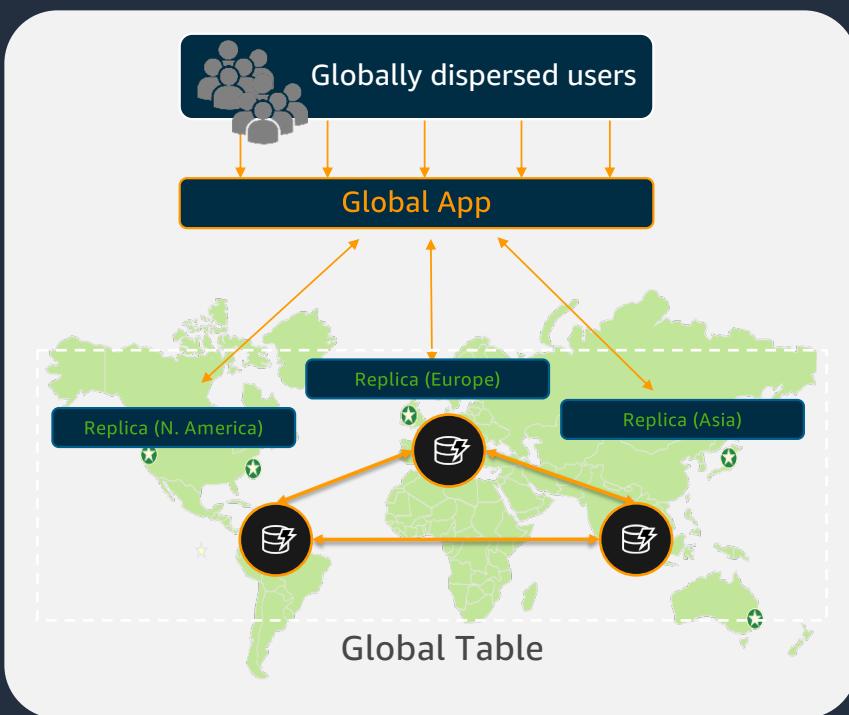
Point in time restore
for short term
retention and data
corruption protection
(35 days)



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

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

Capacity managed for you

Provisioned

Govern max consumption

Auto Scaling

Set a minimum

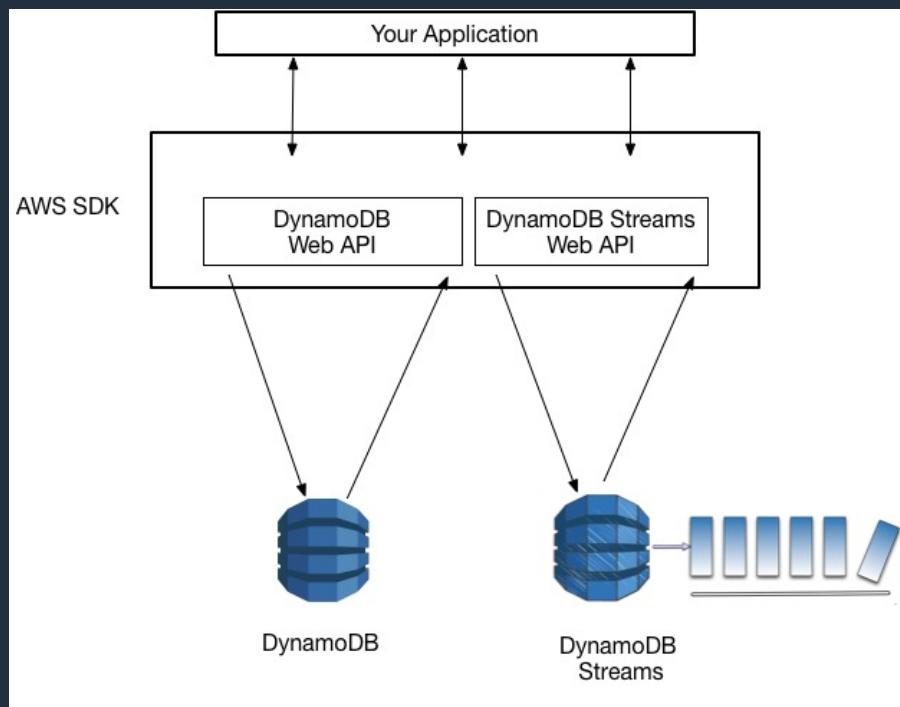
On-Demand

No limit

Start at zero

DynamoDB Streams

Capture changes to items as they occur



Time-ordered sequence of item-level modifications

For inserts, updates & deletes

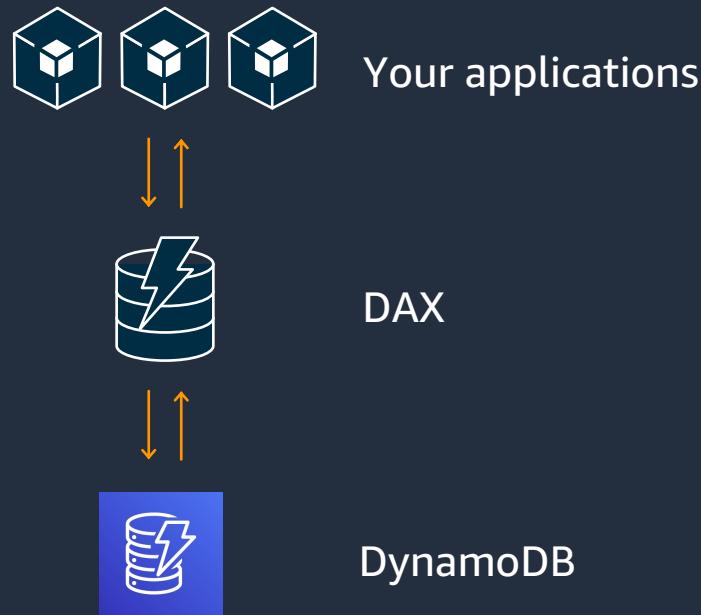
Stored for up to 24 hours

Enable when creating the table or later

No performance impacts from enabling stream on a table

DynamoDB Accelerator (DAX)

High performance at scale



Fully managed, highly available cache for DynamoDB

Even faster — microsecond latency

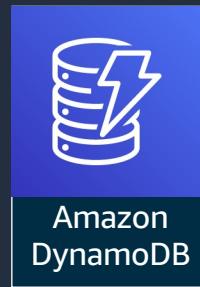
Scales to millions of requests per second

API compatible

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

Want simplest possible DB management?

Want app to manage DB integrity?

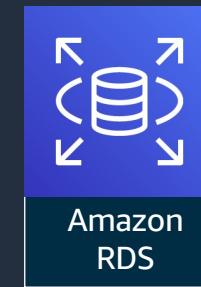


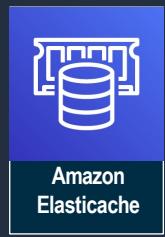
© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Need joins, transactions, frequent table scans?

Want DB engine to manage DB integrity?

Team has SQL skills?





Amazon
ElastiCache

In-Memory Key Value Store

μs is the new ms



In-memory databases
and data grids



Specialized hardware
such
as multi-core
processors,
GPUs, accelerators



Data reduction
approaches
such as sampling,
aggregation

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

Internet-scale apps need low latency and high concurrency



Users	1M+
Data volume	TB-PB-EB
Locality	Global
Performance	Milliseconds to microseconds
Request Rate	Millions
Access	Mobile, IoT, Devices
Scale	Up-Out-In
Economics	Pay as you go
Developer access	Instant API access

ElastiCache Redis

#1 Key-Value Store*

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

Fully Managed & Hardened

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

Secure & Compliant

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

Highly Available & Reliable

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

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

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

Amazon ElastiCache

- In-memory cache in the cloud
- Improve latency and throughput for read-heavy workloads
- Supports open-source caching engines
 - Memcached
 - Redis
- Fully managed
- Multi-AZ



Examples

- Caching of MySQL database query results
- Caching of post-processing results
- Caching of user session and frequently accessed data



Knowledge Check

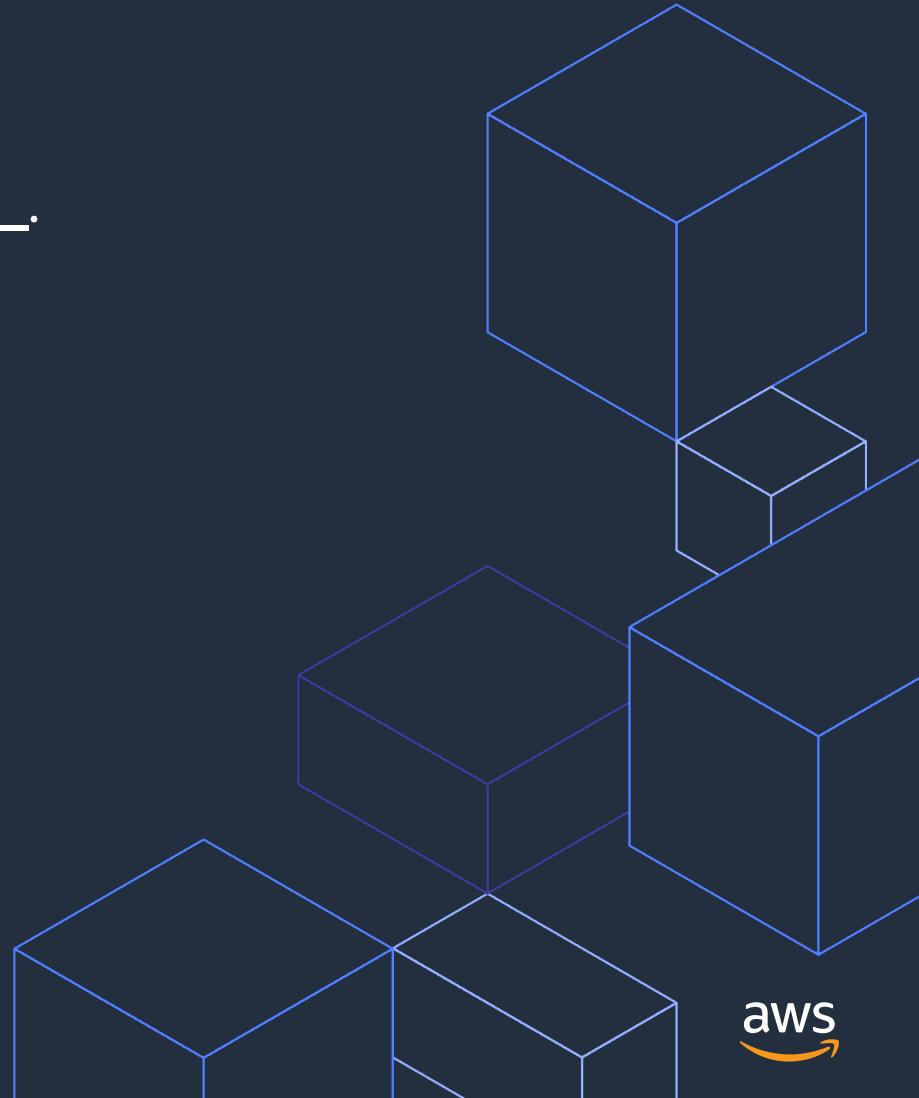
© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Poll #5

AWS's NoSQL product offering is known as ____.

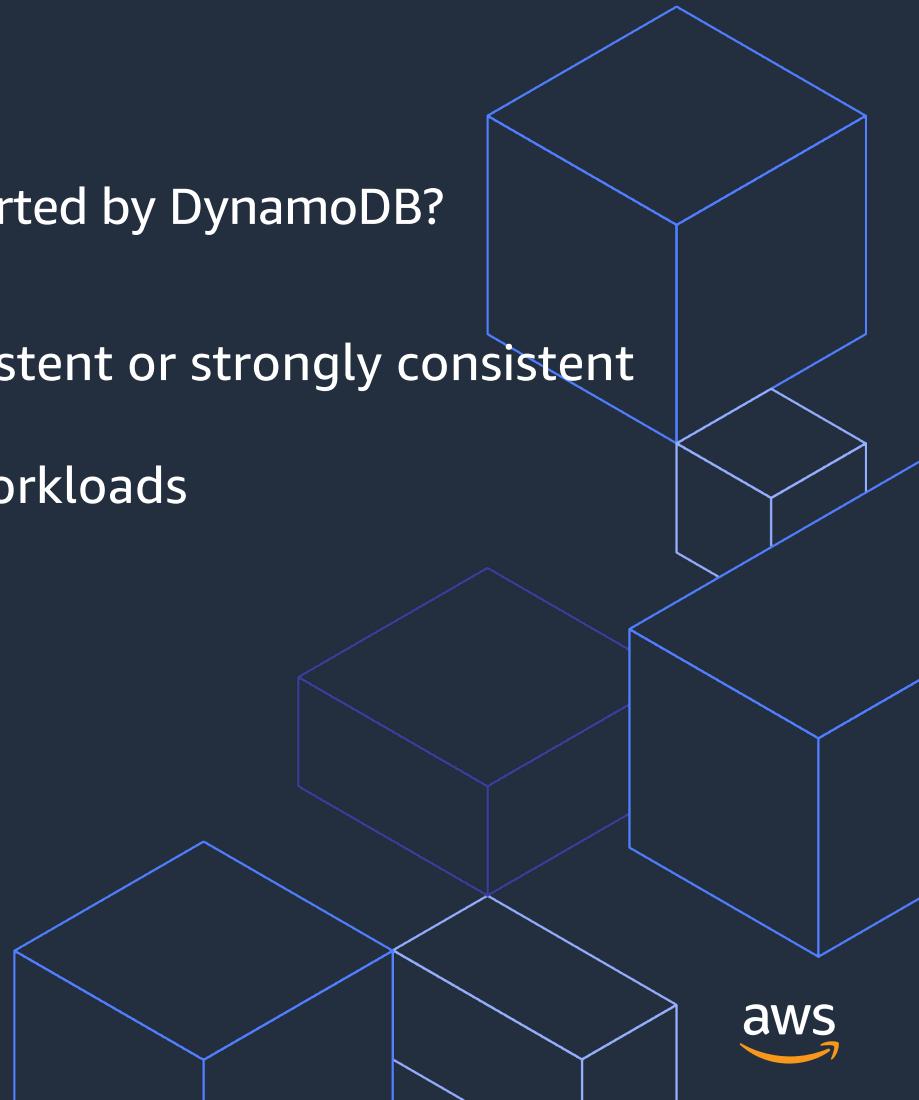
- A. DynamoDB
- B. RDS
- C. MongoDB
- D. MySQL



Poll #6

Which of the following is NOT a feature supported by DynamoDB?

- A. Data reads that are either eventually consistent or strongly consistent
- B. Amazon DynamoDB supports MongoDB workloads



Migrating Databases to AWS - Challenges

- How will my on-premises data migrate to the cloud?
- How can I make it transparent to my users?
- Afterwards, how will on-premises and cloud data interact?
- How can I integrate my data assets within AWS?
- Can I get help moving off of commercial databases?
- How can I move data to my data lake?

What are DMS and SCT?

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



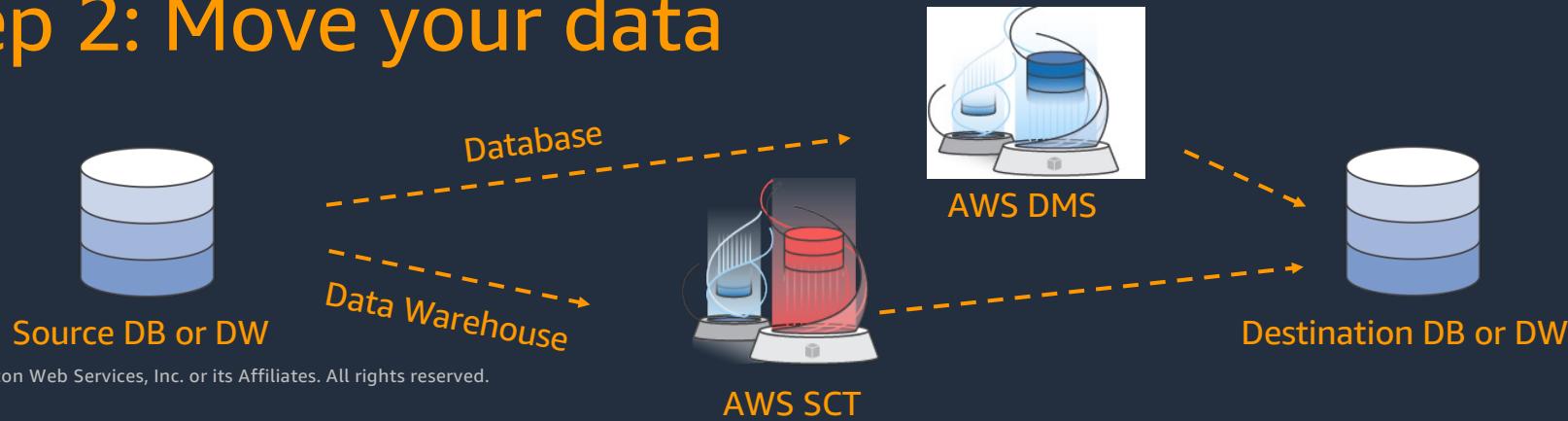
AWS Schema Conversion Tool (SCT) converts your commercial database and data warehouse schemas to open-source engines or AWS-native services, such as Amazon Aurora and Redshift

Database migration process

Step 1: Convert or Copy your Schema



Step 2: Move your data

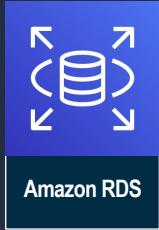


Summary

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



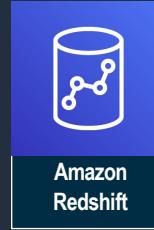
AWS Database Services



Managed Relational
Database Service



Cloud-Native
Relational Database



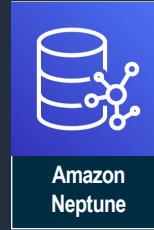
Petabyte-scale Data
Warehouse



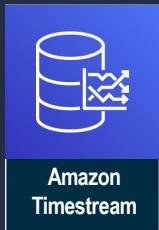
Fully Managed Key-
Value and Document
Database



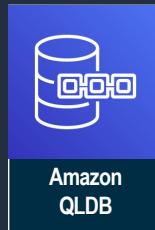
In-Memory Key
Value Store



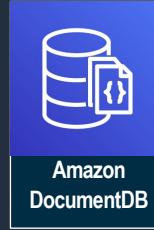
Fully Managed
Graph Database



Fully Managed Time
Series Database



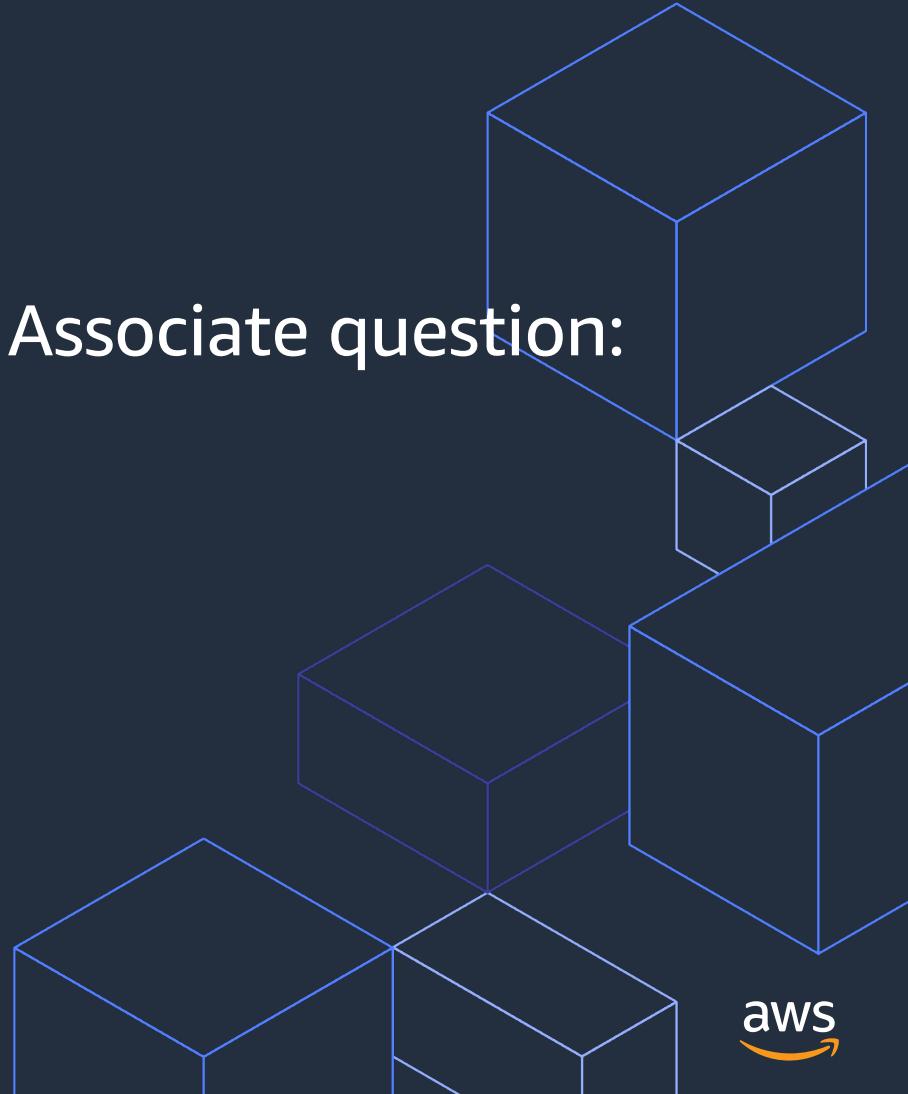
Fully Managed
Ledger Database



MongoDB
Compatible
Document Database

Sample AWS Solutions Architect Associate question:

Database



Poll #7

You work for an advertising company that has a real-time bidding application. You are also using CloudFront on the front end to accommodate a worldwide user base. Your users begin complaining about response times and pauses in real-time bidding. What is the best service that can be used to reduce DynamoDB response times by an order of magnitude (milliseconds to microseconds)?

- A. CloudFront Edge Caches
- B. ElastiCache
- C. DAX
- D. DynamoDB Auto Scaling

Poll #8

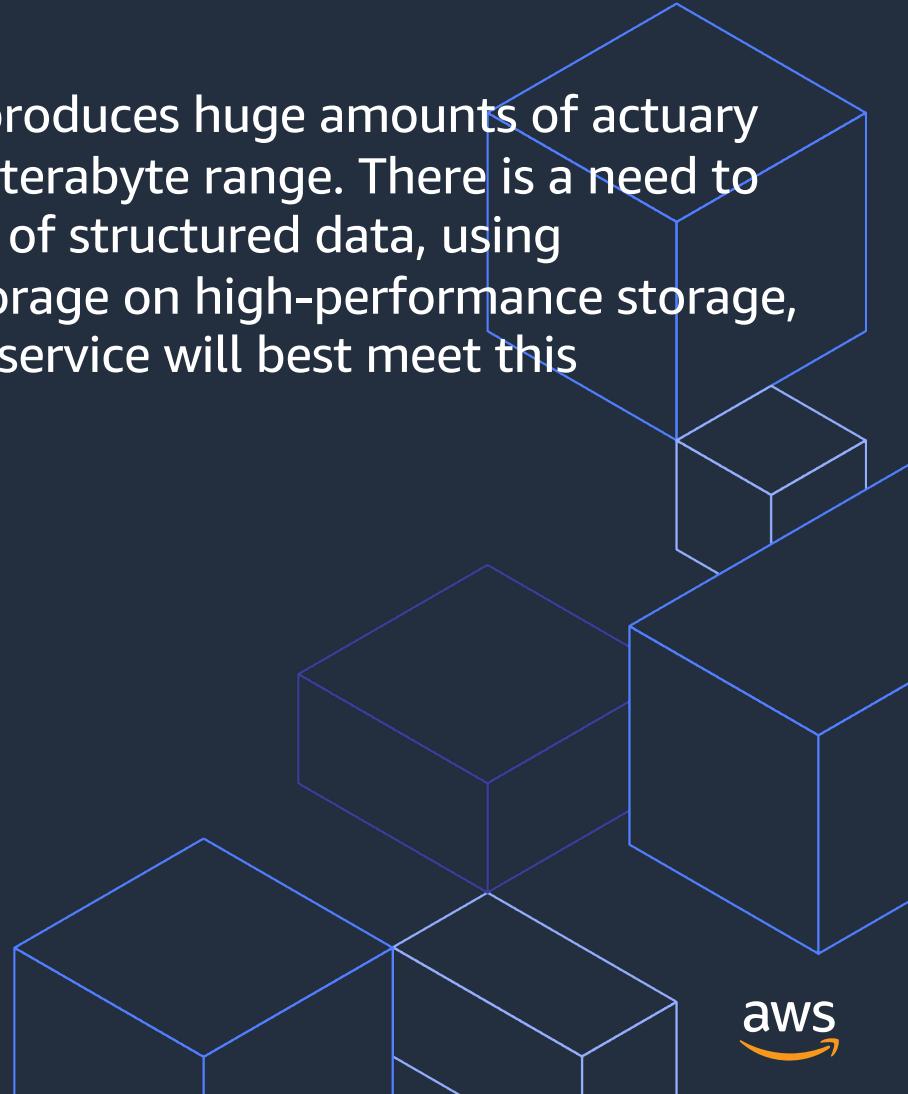
A Solutions Architect has been assigned the task of helping the company development optimize the performance of their web application. End users have been complaining about slow response times. The Solutions Architect has determined that improvements can be realized by adding ElastiCache to the solution. What can ElastiCache do to improve performance?

- A. Queue up requests and allow the processor time to catch up
- B. Offload some of the write traffic to the database
- C. Delivers up to 10x performance improvement from milliseconds to microseconds or even at millions of requests per second.
- D. Cache frequently accessed data in-memory.

Poll #9

A financial institution has an application that produces huge amounts of actuary data, which is ultimately expected to be in the terabyte range. There is a need to run complex analytic queries against terabytes of structured data, using sophisticated query optimization, columnar storage on high-performance storage, and massively parallel query execution. Which service will best meet this requirement?

- A. Elasticache
- B. RDS
- C. DynamoDB
- D. Redshift



Poll #10

Your company has performed a Disaster Recovery drill which failed to meet the Recovery Time Objective (RTO) desired by executive management. The failure was due in large part by the amount of time taken to restore proper functioning on the database side. You have given management a recommendation of implementing synchronous data replication for the RDS database to help meet the RTO. Which of these options can perform synchronous data replication in RDS?

- A. RDS Multi-AZ
- B. DAX
- C. Read Replicas
- D. AWS Database Migration

Additional Resources

- Further Deep Dive on Specific Services and Database Use Cases
 - AWS FAQ: <https://aws.amazon.com/faqs/>
 - AWS Database Blog: <https://aws.amazon.com/blogs/database/>
 - AWS Online Tech Talks: <https://aws.amazon.com/events/online-tech-talks/>
 - Note: Filter for “Databases” Category
- Labs and Tutorials
 - Event Engine Hash: <https://dashboard.eventengine.run/login?hash=7fdf-0b7f6c2f54-04>
 - Hands-on Tutorials: <https://aws.amazon.com/getting-started/hands-on/>
 - Note: Filter for “Databases” Category
 - RDS Labs from ACloudGuru Certified SA Associate course
 - Create an RDS instance
 - RDS backups, multi-AZ and Read Replicas
 - Deploying an Amazon RDS Multi-AZ and Read Replica in AWS

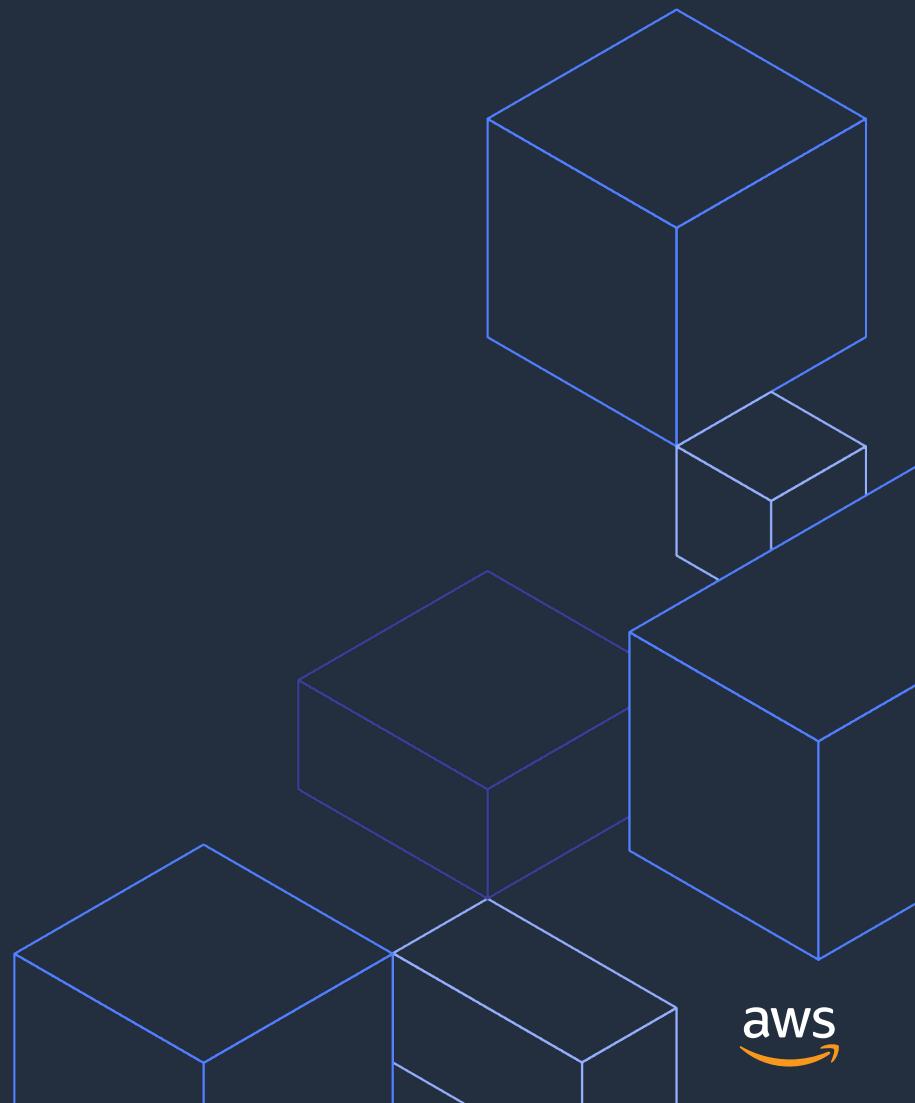
Thank you!

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

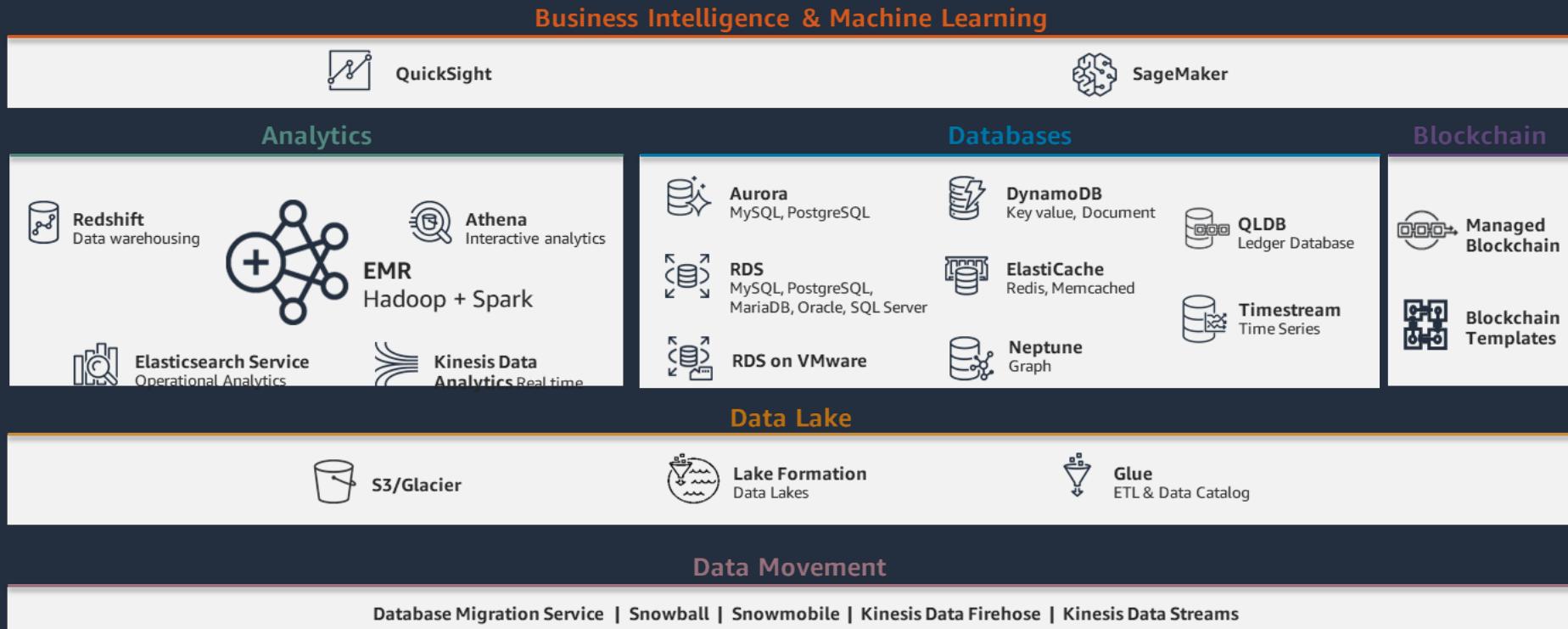


Additional Topics

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



AWS Analytics Stack



Amazon EMR

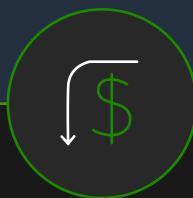
Easily Run Spark, Hadoop, Hive, Presto, HBase, and more big data apps on AWS

Latest versions



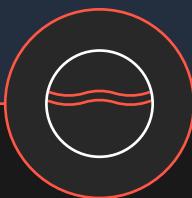
Updated with latest open source frameworks within 30 days

Low cost



50–80% reduction in costs with EC2 Spot and Reserved Instances
Per-second billing for flexibility

Use S3 storage



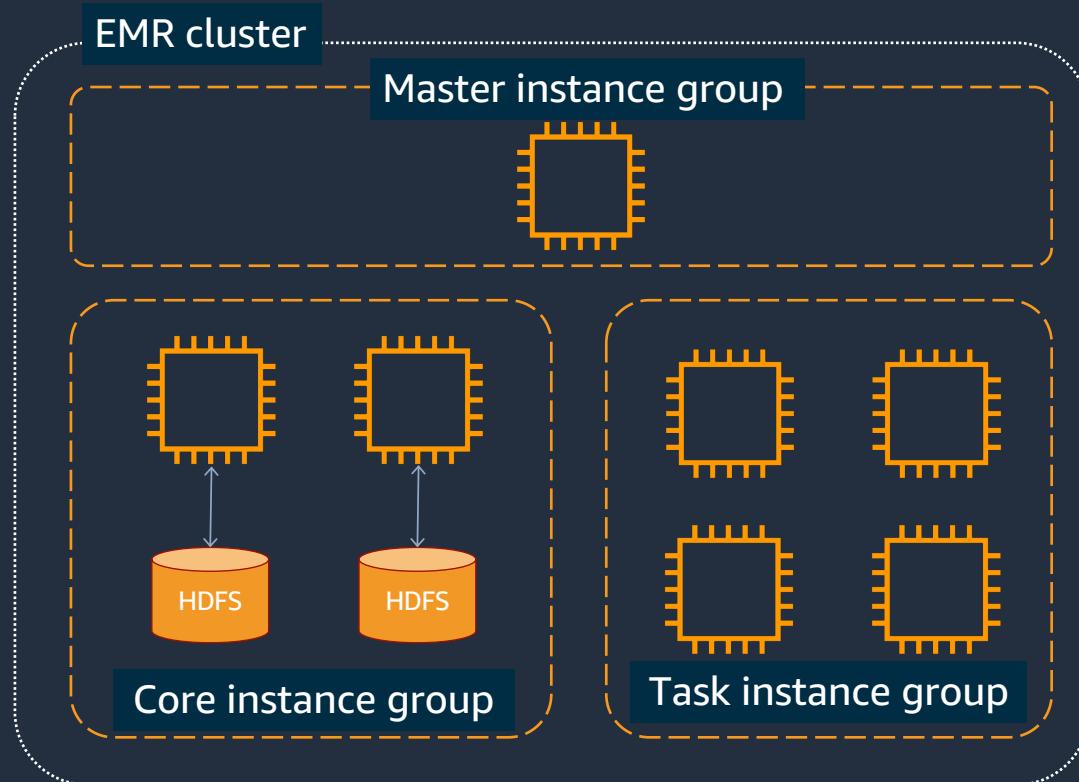
Process data in S3 securely with high performance using the EMRFS connector

Easy



Fully managed no cluster setup, node provisioning, cluster tuning

EMR Node Types



Master Node must keep running

Core nodes can be added and removed gracefully

Cluster can tolerate loss of task nodes.