



Mimmit koodaa - Databases and Analytics overview for the Associate Cert Exam

Veli-Matti Ojala (he/him)

Senior Solutions Architect
AWS for Data Champion

AWS Databases -

Overview for the Associate Cert Exam

AWS Purpose-built databases: the right tool for the job

Purpose-built Databases

Relational

Key-value

Document

In-Memory

Graph

Time-Series

Ledger

Wide Column



Aurora

RDS

DynamoDB

DocumentDB

ElastiCache

MemoryDB

Neptune

Timestream

QLDB

Keyspaces

Amazon RDS for Open Source engines

Set up, operate and scale managed **MySQL**, **PostgreSQL**, and **MariaDB** databases in a few clicks



Fully managed

Deploy using the familiar code, features, and tools of your favorite database

Access the latest major and minor versions



Cost effective

Avoid costly licensing fees associated with commercial databases

Benefit from the latest price-performance capabilities



Highly available

Customize your workload for the availability that you need

Access to Amazon RDS Multi-AZ with one standby or with two readable standbys



Easy Migration

Reduce the complexity of migrating your open source databases

Simple to lift and shift workloads

Your typical and trusted Relational Database Service

Amazon RDS Commercial Engines

Set up, operate and scale managed **Oracle** and **SQL Server** databases in a few clicks



Fully managed

Automate administrative tasks like hardware provisioning, patches, backup and recovery.

Access the latest major and minor versions



Flexible licensing

Choice of 'license included' or 'bring-your-own-license' (BYOL) licensing¹ models

No upfront fees or long-term commitment with pay-as-you-go or reserved pricing options

¹ BYOL only applies to RDS for Oracle



Highly available

Customize your workload for the availability that you need with security and durability

Resilient configuration options including multi-AZs, read replicas, and more



Easy migration

Reduce the complexity of migrating with like-to-like migration

Deploy using the familiar code, features, and tools of your favorite database

No schema conversion, simply migrate to and from

Amazon Aurora

Designed for unparalleled high performance and availability at global scale with full **MySQL** and **PostgreSQL** compatibility at 1/10th the cost of commercial databases



Performance & scalability

- 5x throughput of standard MySQL and 3x of standard PostgreSQL
- Scale out up to 15 read replicas
- Decoupled storage and compute enabling cost optimization
- Fast database cloning
- Distributed, dynamically scaling storage subsystem



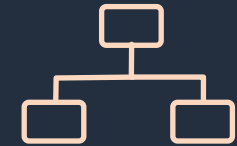
Availability & durability

- 6 copies of data across 3 AZs (customers pays for 1)
- Automatic, continuous, incremental backups with point-in-time recovery (PITR)
- Fault-tolerant, self-healing, auto-scaling storage
- Global Database for disaster recovery



Highly secure

- Network isolation
- Encryption at rest/in transit
- Supports multiple secure authentication mechanisms and audit controls



Fully managed

- Automates time-consuming management of administration tasks like hardware provisioning, database setup, patching, and backups
- Serverless configuration options

Highly performant Relational Database Service



Global Database

Faster disaster recovery and enhanced data locality

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

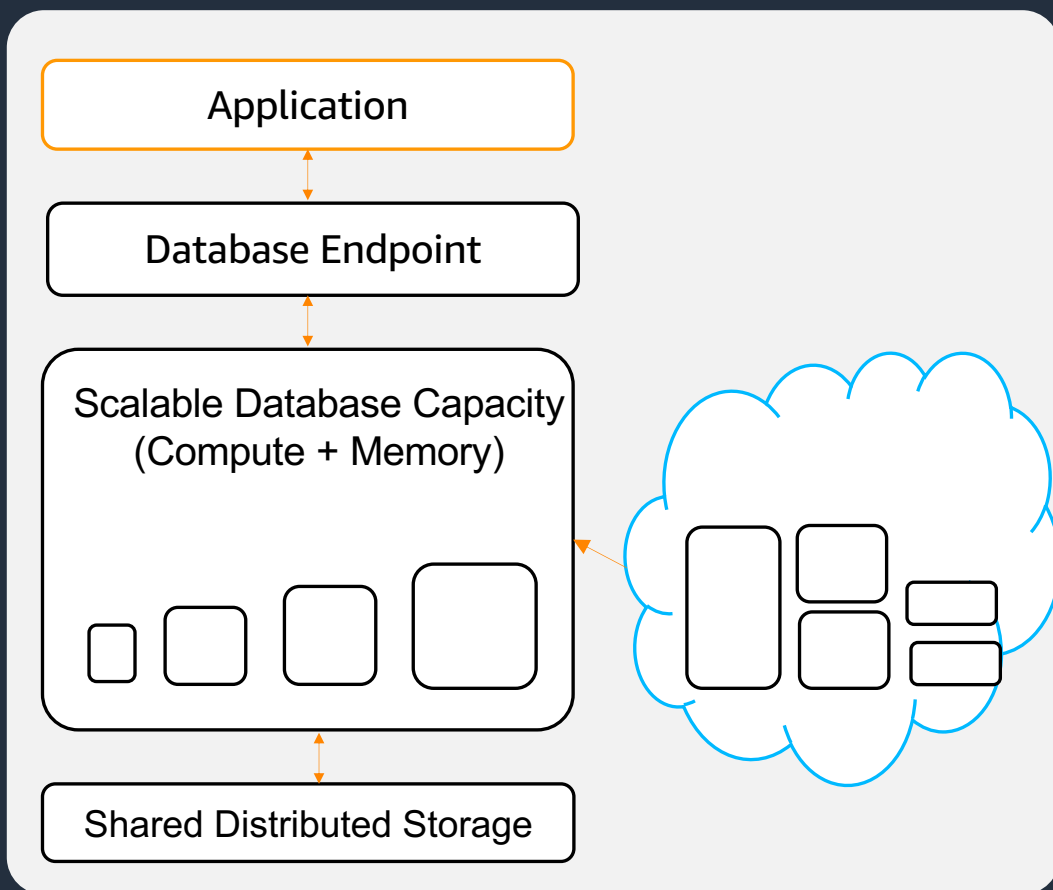
Bring data close to your customer's applications in **different regions**

Promote to primary 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

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

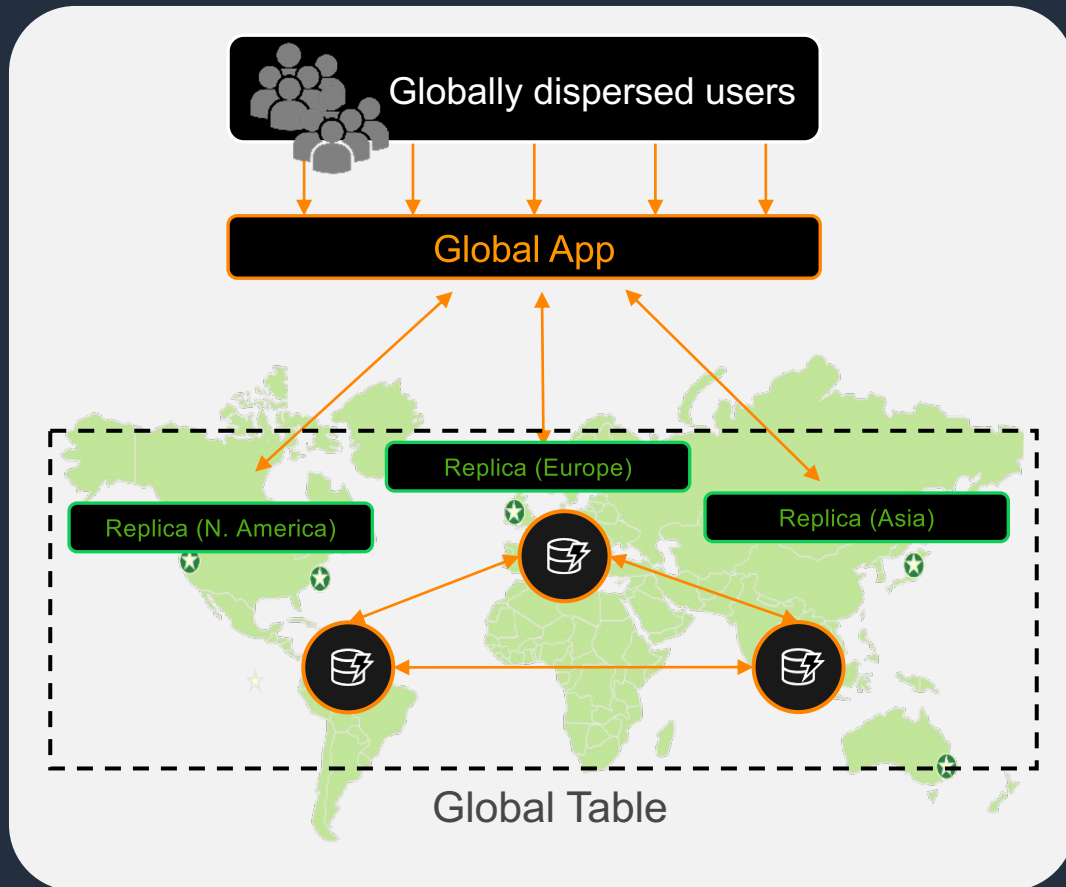
Stores Keys, that can have almost any content.
Mobile, Web, Gaming, Ad tech, IoT, and other applications

Highly Available and Durable



Global Tables

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



Build high performance, globally distributed applications

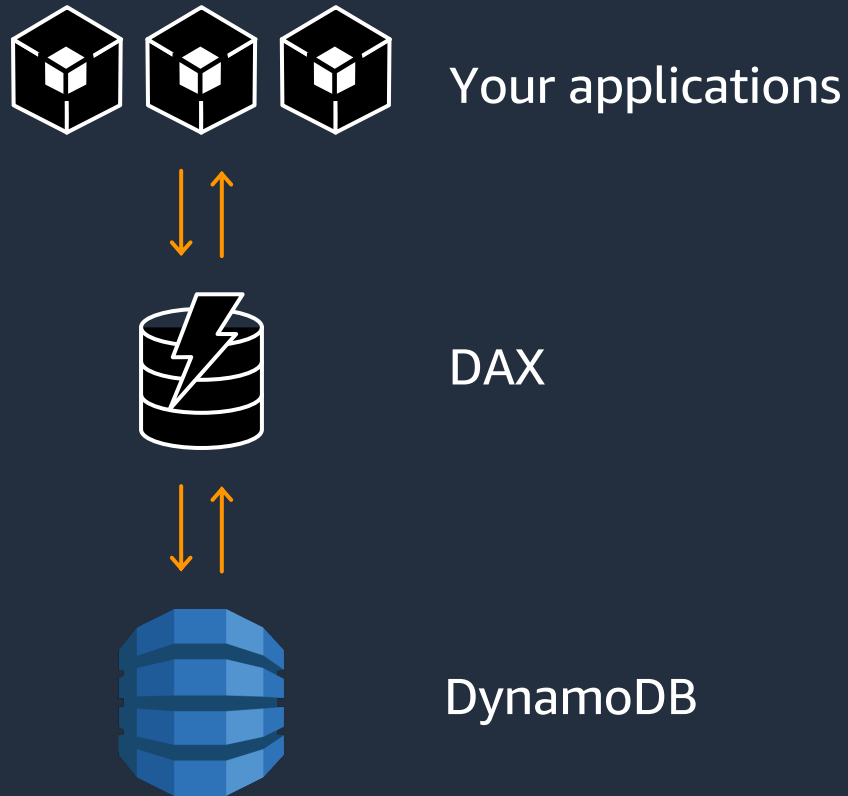
Low latency reads & writes to locally available tables

Disaster proof with multi-region redundancy

Easy to setup and no application re-writes required

DynamoDB Accelerator (DAX)

High performance



Fully managed, highly available cache for DynamoDB

Even faster—microsecond latency

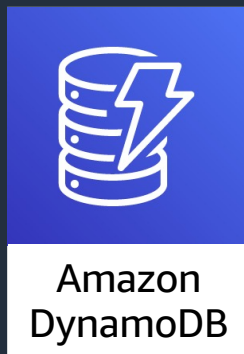
Scales to millions of requests per second

API compatible

NoSQL vs. SQL for a New App

Want simplest possible DB management?

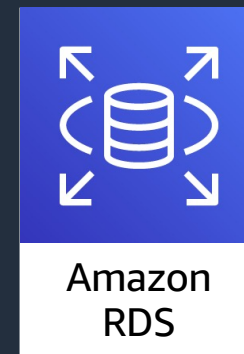
Want app to manage DB integrity?



Need joins, transactions, frequent table scans?

Want DB engine to manage DB integrity?

Team has SQL skills?





Amazon DocumentDB

Fast, scalable, highly available **MongoDB-compatible database** service



Millions of requests per second,
millisecond latency



Same code, drivers, and tools
you use with MongoDB



Simple and
fully managed



Secure and
compliant



2x throughput of
managed MongoDB services



Deeply integrated
with AWS services

Stores JSON-like files - you can query their content



Amazon MemoryDB for Redis

Redis-compatible, durable, in-memory database service



Ultra-fast performance

Microsecond read and single-digit millisecond write latencies with millions of transactions per second



Redis compatibility

Flexible and friendly data structures and APIs



Durability and high availability

Multi-AZ transactional for durability and high availability



Security

Amazon VPC, encryption at rest and in transit, access control list (ACL)



Fully managed

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



High scalability

More than 100 TB of storage per cluster (with 1 replica per shard)

This is your Cache - Also **Amazon ElastiCache, Redis, Memcached**



Amazon Neptune

Fast, reliable graph database built for the cloud

Open



Supports Apache TinkerPop and W3C RDF graph models

Fast



Query billions of relationships with millisecond latency

Reliable



Six replicas of data across three Availability Zones with full backup and restore

Easy



Build powerful queries easily with Gremlin and SPARQL

Graph database - Think social media connections



Amazon Timestream

Fast, scalable, fully managed time-series database

1,000x faster and 1/10th
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

When Timestamp is the key - IoT Data, Log data...

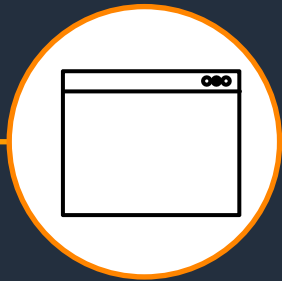
Amazon Quantum Ledger Database



Fully managed ledger database

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

Immutable and transparent



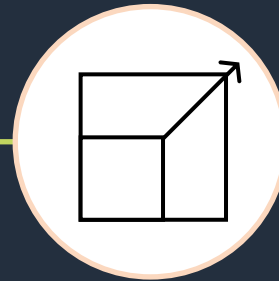
Append-only, immutable journal tracks history of all changes that cannot be deleted or modified; get full visibility into entire data lineage

Cryptographically verifiable



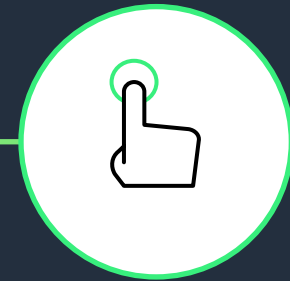
All changes are cryptographically chained and verifiable

Highly scalable



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

Easy to use



Flexible document model, query with familiar SQL-like interface

Crypto database for immutable needs

Amazon Keyspaces (for Apache Cassandra)



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

Apache Cassandra-compatible



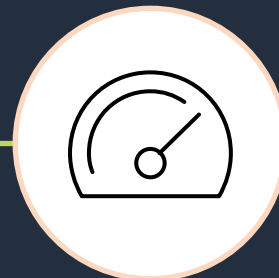
Use the same Cassandra drivers and tools

No servers to manage



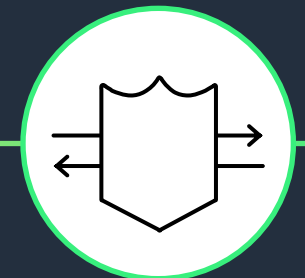
No need to provision, configure, and operate large Cassandra clusters

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

Wide column database - hybrid between key-value and relational

AWS Data and Analytics - Overview for the Associate Cert Exam

Modern data strategy on AWS



AWS Analytics Pillars



Scalable data lakes



Purpose-built for performance and cost



Serverless and easy to use



Unified data access, security, and governance



Built-in machine learning



Scalable data lakes

Broadest portfolio
of analytics tools

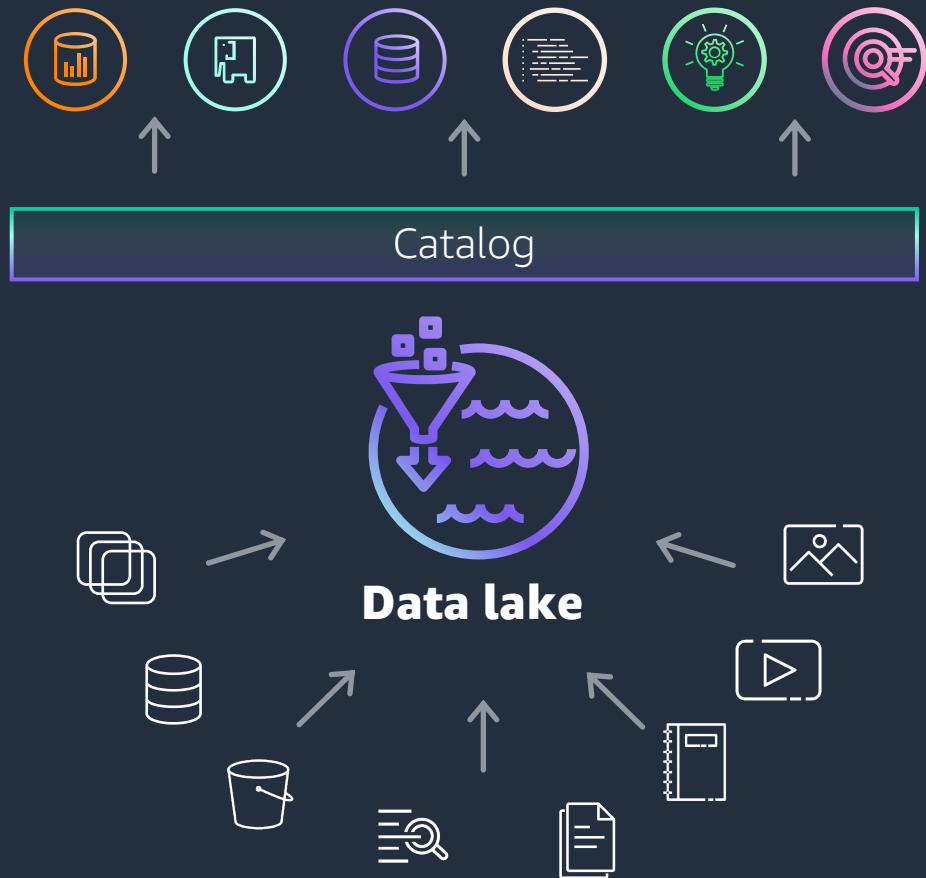


**Amazon
S3**

Unmatched durability,
availability, and scalability

Built to store and retrieve
any amount of data

The benefits of data lakes



Store all your data in open formats

Cost-effectively scale storage to exabytes

Decouple storage from compute

Choice of analytical and ML engines

Process data in place

AWS Lake Formation

BUILD A SECURE
DATA LAKE IN DAYS



Build data lakes quickly



Simplify security management



Provide self-service access to data

AWS Glue

SIMPLE, SCALABLE,
AND SERVERLESS



Integrate data faster



Automate at scale



No servers to manage

AWS Glue: Key Capabilities

SERVERLESS DATA INTEGRATION SERVICE

Scalable Data Integration Engine



Built-in data transforms



Execution engine



Monitor

Centralized and Unified Data Governance



Glue data catalog



Glue crawlers



Lake formation

Connect and Ingest Data



Glue connectors



Glue connector marketplace



Variety of interfaces

User Productivity and Data Ops



Persona specific tools



Productivity tools



Data ops tools



AMAZON
REDSHIFT

Data
warehousing



AMAZON
ATHENA

Query all your
data using SQL
or Python



AMAZON
EMR

Big data
processing



AMAZON
OPENSEARCH
SERVICE

Log and search
analytics



AMAZON
KINESIS & MSK

Real-time
analytics

Amazon Athena

QUERY ALL YOUR
DATA USING
SQL OR PYTHON



Simple, instant start



Interactive, advanced analytics



Open and flexible



Cost-effective



SIMPLE, INSTANT START

Serverless, no setup

Instant start,
optimized runtimes
for fast results

Point to S3 and
start querying



INTERACTIVE, ADVANCED ANALYTICS

Federated queries
across **35+** data stores

Use PySpark ecosystem

Simplified notebooks on
console for PySpark



OPEN AND FLEXIBLE

ANSI SQL, Apache Spark

Multiple formats,
compression types, and
complex joins
and data types



COST EFFECTIVE

Pay only for what you use

SQL: Save **30%–90%** on
per-query costs
through compression

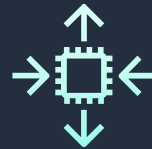
Python: minimize idle
compute charges

Amazon Redshift

THE BEST PRICE-PERFORMANCE
FOR CLOUD DATA WAREHOUSING



Analyze all your data
Datawarehouse for massive
amounts of data



Price-performance at any scale
Integrates nicely with S3



Easy, secure, and reliable
Massively Parallel Processing - MPP
(think columnar)

Amazon EMR

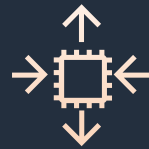
RUN BIG DATA APPLICATIONS
IN THE CLOUD



Fully managed and customizable



Latest open-source releases



Multiple deployment models



Automatically scale up and down

Amazon S3 data lake integration

Amazon OpenSearch Service

SUCCESSOR TO
AMAZON ELASTICSEARCH SERVICE



Fully managed

Can compliment your databases



Log and search analytics

Integrates nicely with S3,
Firehose and other AWS Services



Cost effective

Amazon Kinesis

COLLECT, PROCESS, AND
ANALYZE VIDEO AND DATA
STREAMS IN REAL TIME



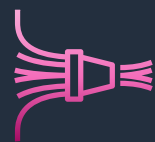
Kinesis Data Streams



Kinesis Data Analytics



Kinesis Video Streams



Kinesis Data Firehose

Amazon QuickSight

CLOUD-NATIVE BI SOLUTION
FOR ILLUMINATING
ORGANIZATIONAL INSIGHTS



Auto-scaling and serverless



Internal and/or external users



Deeply integrated with AWS services



Augmented insights on-demand



Thank you!

**Best of luck to the
Cert Exam!**