

AWS Solution Architect Associate Certification Training – Module 16

16. Relational Database Services (RDS)

Introduction to databases

Database refers to a collection of electronic records that could be processed to produce useful information. The data can be accessed, modified, managed, controlled and organized to perform various data-processing operations. The data is typically indexed across rows, columns and tables that make workload processing and data querying efficient.

In enterprise applications, databases involve mission-critical, security-sensitive and compliance-focused record items that have complicated logical relationships with other datasets and grow exponentially over time as the userbase increases. As a result, these organizations require technology solutions to maintain, secure, manage and process the data stored in databases. This is where Database Management System come into play.

Overview of structured and unstructured data

Structured Data: If Only Everything Were This Easy

Structured data is highly organized information that uploads neatly into a relational database (think traditional row database structures), lives in fixed fields, and is easily detectable via search operations or algorithms. Structured data is relatively simple to enter, store, query, and analyze, but it must be strictly defined in terms of field name and type (e.g. alpha, numeric, date, currency), and as a result is often restricted by character numbers or specific terminology. Analysts typically use simple or more complex VLOOKUP queries in Excel spreadsheets or Structured Query Language (SQL) to perform queries on structured data within relational databases.

Structured data leaves out immense amounts of material that do not fit simply into a firm's organization of information. Until recently, structured data was supplemented by this additional information in the form of paper or microfiche. With the improvement of processing by computers, lowered cost of data storage, and the spread of new formats of data, the age of unstructured data began. Now, structured and unstructured data must both be consulted, queried, assimilated and leveraged to make the best business decisions.

Unstructured Data: Everything You Didn't Know You Wanted

Unstructured data may have its own internal structure, but does not conform neatly into a spreadsheet or database. While unruly in nature, it is also incredibly valuable and increasingly available in the form of complex data sources, such as web logs, multimedia content, email, customer service interactions, sales automation, and social media data. Most business interactions, in fact, are unstructured in nature.

The fundamental challenge of unstructured data sources is that they are difficult for nontechnical business users and data analysts alike to unbox, understand, and prepare for analytic use. Beyond issues of structure, is the sheer volume of this type of data. Because of this, current data mining techniques often leave out valuable information and make analyzing unstructured data laborious and expensive.

Introduction to database services of AWS

AWS offers a wide range of database services to fit your application requirements. These database services are fully managed and can be launched in minutes with just a few clicks.

Database services

Relational:

Relational databases store data with pre-defined schema and relationships between them, designed for supporting ACID (Atomicity, Consistency, Isolation, Durability) transactions, maintaining referential integrity, and data consistency.

Used for: Traditional applications, ERP, CRM, and e-commerce.

AWS Offerings

- Amazon Aurora
MySQL, PostgreSQL
- Amazon RDS
MySQL, PostgreSQL, MariaDB, Oracle, SQL Server
- Amazon Redshift

Key-value:

Key-value databases are optimized to store and retrieve key-value pairs in large volumes and in milliseconds, without the performance overhead and scale limitations of relational databases.

Used for: Internet-scale applications, real-time bidding, shopping carts, and customer preferences.

AWS Offerings

- Amazon DynamoDB

Document:

Document databases are designed to store semi-structured data as documents and are intuitive for developers to use because the data is typically represented as a readable document.

Used for: Content management, personalization, and mobile applications.

AWS Offerings

- Amazon DocumentDB (with MongoDB compatibility)

In-memory:

In-memory databases are used for applications that require real time access to data. By storing data directly in memory, these databases provide microsecond latency where millisecond latency is not enough.

Used for: Caching, gaming leaderboards, and real-time analytics.

AWS Offerings:

- Amazon ElastiCache for Redis
- Amazon ElastiCache for Memcached

Graph:

Graph databases are used for applications that need to enable millions of users to query and navigate relationships between highly connected, graph datasets with millisecond latency.

Used for: Fraud detection, social networking, and recommendation engines

AWS Offerings:

- Amazon Neptune

Time Series:

Time series databases are used to efficiently collect, synthesize, and derive insights from enormous amounts of data that changes over time (known as time-series data).

Used for: IoT applications, DevOps, and industrial telemetry.

AWS Offerings:

- Amazon Timestream

Introduction to Amazon RDS

Amazon Relational Database Service (Amazon RDS) makes it easy to set up, operate, and scale a relational database in the cloud. It provides cost-efficient and resizable capacity while automating time-consuming administration tasks such as hardware provisioning, database setup, patching and backups. It frees you to focus on your applications so you can give them the fast performance, high availability, security and compatibility they need.

Benefits

Easy to administer:

Amazon RDS makes it easy to go from project conception to deployment. Use the Amazon RDS Management Console, the AWS RDS Command-Line Interface, or simple API calls to access the capabilities of a production-ready relational database in minutes. No need for infrastructure provisioning, and no need for installing and maintaining database software.

Highly scalable:

You can scale your database's compute and storage resources with only a few mouse clicks or an API call, often with no downtime. Many Amazon RDS engine types allow you to launch one or more Read Replicas to offload read traffic from your primary database instance.

Available and durable:

Amazon RDS runs on the same highly reliable infrastructure used by other Amazon Web Services. When you provision a Multi-AZ DB Instance, Amazon RDS synchronously replicates the data to a standby instance in a different Availability Zone (AZ). Amazon RDS has many other features that enhance reliability for critical production databases, including automated backups, database snapshots, and automatic host replacement.

Fast:

Amazon RDS supports the most demanding database applications. You can choose between two SSD-backed storage options: one optimized for high-performance OLTP applications, and the other for cost-effective general-purpose use. In addition, Amazon Aurora provides performance on par with commercial databases at 1/10th the cost.

Secure:

Amazon RDS makes it easy to control network access to your database. Amazon RDS also lets you run your database instances in Amazon Virtual Private Cloud (Amazon VPC), which enables you to isolate your database instances and to connect to your existing IT infrastructure through an industry-standard encrypted IPsec VPN. Many Amazon RDS engine types offer encryption at rest and encryption in transit.

Inexpensive:

You pay very low rates and only for the resources you actually consume. In addition, you benefit from the option of On-Demand pricing with no up-front or long-term commitments, or even lower hourly rates via our Reserved Instance pricing.

Concepts of Amazon RDS services

Why do you want a managed relational database service? Because Amazon RDS takes over many of the difficult or tedious management tasks of a relational database:

- When you buy a server, you get CPU, memory, storage, and IOPS, all bundled together. With Amazon RDS, these are split apart so that you can scale them independently. If you need more CPU, less IOPS, or more storage, you can easily allocate them.
- Amazon RDS manages backups, software patching, automatic failure detection, and recovery.
- To deliver a managed service experience, Amazon RDS doesn't provide shell access to DB instances, and it restricts access to certain system procedures and tables that require advanced privileges.
- You can have automated backups performed when you need them, or manually create your own backup snapshot. You can use these backups to restore a database. The Amazon RDS restore process works reliably and efficiently.

- You can get high availability with a primary instance and a synchronous secondary instance that you can fail over to when problems occur. You can also use MySQL, MariaDB, or PostgreSQL Read Replicas to increase read scaling.
- You can use the database products you are already familiar with: MySQL, MariaDB, PostgreSQL, Oracle, Microsoft SQL Server.
- In addition to the security in your database package, you can help control who can access your RDS databases by using AWS Identity and Access Management (IAM) to define users and permissions. You can also help protect your databases by putting them in a virtual private cloud.

Different types of database engines supported in AWS

AWS database services include Amazon Relational Database Service (Amazon RDS), with support for six commonly used database engines, Amazon Aurora, a MySQL-compatible relational database with five times the performance, Amazon DynamoDB, a fast and flexible NoSQL database service, Amazon Redshift, a petabyte-scale data warehouse service, and Amazon ElastiCache, an in-memory cache service with support for Memcached and Redis. AWS also provides the AWS Database Migration Service, a service which makes it easy and inexpensive to migrate your databases to AWS cloud.

Database machine types, Storage, security groups

Amazon RDS Instance/machine Types:

Amazon RDS provides a selection of instance types optimized to fit different relational database use cases. Instance types comprise varying combinations of CPU, memory, storage, and networking capacity and give you the flexibility to choose the appropriate mix of resources for your database. Each instance type includes several instance sizes, allowing you to scale your database to the requirements of your target workload.

Not every instance type is supported for every database engine, version, edition or region.

General purpose

T3 instances are the next generation burstable general-purpose instance type that provide a baseline level of CPU performance with the ability to burst CPU usage at any time for as long as required. T3 instances offer a balance of compute, memory, and network resources and are ideal for database workloads with moderate CPU usage that experience temporary spikes in use.

T3

T3 instances accumulate CPU credits when a workload is operating below baseline threshold. Each earned CPU credit provides the T3 instance the opportunity to burst with the performance of a full CPU core for one minute when needed. Amazon RDS T3 instances are configured for Unlimited mode, which means they can burst beyond the baseline over a 24-hour window for an additional charge.

Features:

- Burstable CPU, governed by CPU Credits, and consistent baseline performance
- Unlimited mode to ensure performance during peak periods
- Powered by the AWS Nitro System, a combination of dedicated hardware and lightweight hypervisor
- AWS Nitro System and high frequency Intel Xeon Scalable processors result in better price performance than T2 instances

T2

T2 instances are burstable general-purpose performance instances that provide a baseline level of CPU performance with the ability to burst above the baseline. T2 instances are a good choice for a variety of database workloads including micro-services and test and staging databases.

The baseline performance and ability to burst are governed by CPU Credits. T2 instances receive CPU Credits continuously at a set rate depending on the instance size, accumulating CPU Credits when they are idle, and consuming CPU credits when they are active.

Features:

- High frequency Intel Xeon processors
- Burstable CPU, governed by CPU Credits, and consistent baseline performance
- Free Tier eligible
- Balance of compute, memory, and network resources

M5

M5 instances are the latest generation of General Purpose Instances and provide improved performance over M4. This family provides a balance of compute, memory, and network resources, and is a good choice for many database workloads.

Features:

- 2.5 GHz Intel Xeon® Platinum 8175 processors with new Intel Advanced Vector Extension (AVX-512) instruction set
- New larger instance size, m5.24xlarge, offering 96 vCPUs and 384 GiB of memory
- Up to 25 Gbps network bandwidth using Enhanced Networking
- Powered by the AWS Nitro System, a combination of dedicated hardware and lightweight hypervisor

M4

M4 instances provide a balance of compute, memory, and network resources, and are a good choice for many database workloads including small and mid-size databases for open source or enterprise applications.

Features:

- 2.3 GHz Intel Xeon® E5-2686 v4 (Broadwell) processors or 2.4 GHz Intel Xeon® E5-2676 v3 (Haswell) processors
- EBS-optimized by default at no additional cost
- Support for Enhanced Networking
- Balance of compute, memory, and network resources

Memory Optimized

R5

R5 instances are the latest generation of memory optimized instances that deliver 5% additional memory per vCPU than R4 with the largest size providing 768 GiB of memory. In addition, R5 instances deliver a 10% price per GiB improvement and a ~20% increased CPU performance over R4.

Features:

- Up to 768 GiB of memory per instance
- Intel Xeon Platinum 8000 series (Skylake-SP) processors with a sustained all core Turbo CPU clock speed of up to 3.1 GHz
- Powered by the AWS Nitro System, a combination of dedicated hardware and lightweight hypervisor

R4

R4 instances are optimized for memory-intensive database workloads and offer better price per GiB of RAM than R3.

Features:

- High Frequency Intel Xeon E5-2686 v4 (Broadwell) processors
- DDR4 Memory
- Support for Enhanced Networking

X1e

X1e instances are optimized for high-performance databases. X1e instances offer one of the lowest price per GiB of RAM among Amazon RDS instance types.

Features:

- High frequency Intel Xeon E7-8880 v3 (Haswell) processors
- One of the lowest prices per GiB of RAM
- Up to 3,904 GiB of DRAM-based instance memory

X1

X1 instances are optimized for large-scale, enterprise-class and in-memory applications, and offer one of the lowest price per GiB of RAM among Amazon RDS instance types.

Features:

- High frequency Intel Xeon E7-8880 v3 (Haswell) processors
- One of the lowest prices per GiB of RAM
- Up to 1,952 GiB of DRAM-based instance memory

Instance Features

1. Burstable Performance Instances

Amazon RDS allows you to choose between Fixed Performance Instances (e.g. M5 and R5) and Burstable Performance Instances (e.g. T3). Burstable Performance Instances provide a baseline level of CPU performance with the ability to burst above the baseline.

2. Database Storage Options

Storage for Amazon RDS for MySQL, MariaDB, PostgreSQL, Oracle, and SQL Server is built on Amazon EBS, a durable, block-level storage service. Amazon RDS provides three volume types to best meet the needs of your database workloads: General Purpose (SSD), Provisioned IOPS (SSD), and Magnetic. General Purpose (SSD) is an SSD-backed, general purpose volume type that we recommend as the default choice for a broad range of database workloads. Provisioned IOPS (SSD) volumes offer storage with consistent and low-latency performance, and are designed for I/O intensive database workloads. Magnetic volumes provide a low cost per gigabyte and are provided for backwards compatibility.

Amazon Aurora features a distributed, fault-tolerant, self-healing storage system that auto-scales up to 64TB per database instance. It delivers high performance and availability with up to 15 low-latency read replicas, point-in-time recovery, continuous backup to Amazon S3, and replication across three Availability Zones (AZs).

3. EBS-optimized Instances

EBS-optimized instances enable Amazon RDS to fully use the IOPS provisioned on an EBS volume. EBS-optimized instances deliver dedicated throughput between Amazon RDS and Amazon EBS, with options between 500 and 4,000 Megabits per second (Mbps) depending on the instance type used. The dedicated throughput minimizes contention between Amazon EBS I/O and other traffic from your RDS instance, providing the best performance for your EBS volumes. EBS-optimized instances are designed for use with

both Standard and Provisioned IOPS Amazon EBS volumes. When attached to EBS-optimized instances, Provisioned IOPS volumes can achieve single digit millisecond latencies and are designed to deliver within 10% of the provisioned IOPS performance 99.9% of the time. AWS recommend using Provisioned IOPS volumes with EBS-optimized instances or instances that support cluster networking for applications with high storage I/O requirements.

4. Enhanced Networking

Enhanced Networking enables you to get significantly higher packet per second (PPS) performance, lower network jitter and lower latencies. This feature uses a new network virtualization stack that provides higher I/O performance and lower CPU utilization compared to traditional implementations. Amazon RDS automatically enables Enhanced Networking for supported database instance types.

Security Groups

Security groups control the access that traffic has in and out of a DB instance. Three types of security groups are used with Amazon RDS: DB security groups, VPC security groups, and Amazon EC2 security groups. In simple terms, these work as follows:

- A DB security group controls access to EC2-Classic DB instances that are not in a VPC.
- A VPC security group controls access to DB instances and EC2 instances inside a VPC.
- An EC2 security group controls access to an EC2 instance.

Database replica's

Amazon RDS Read Replicas provide enhanced performance and durability for database (DB) instances. This feature makes it easy to elastically scale out beyond the capacity constraints of a single DB instance for read-heavy database workloads. You can create one or more replicas of a given source DB Instance and serve high-volume application read traffic from multiple copies of your data, thereby increasing aggregate read throughput. Read replicas can also be promoted when needed to become standalone DB instances. Read replicas are available in Amazon RDS for MySQL, MariaDB, PostgreSQL and Oracle as well as Amazon Aurora.

Database Encryption

You can encrypt your Amazon RDS DB instances and snapshots at rest by enabling the encryption option for your Amazon RDS DB instances. Data that is encrypted at rest includes the underlying storage for a DB instances, its automated backups, Read Replicas, and snapshots.

Amazon RDS encrypted DB instances use the industry standard AES-256 encryption algorithm to encrypt your data on the server that hosts your Amazon RDS DB instances. Once your data is encrypted, Amazon RDS handles authentication of access and decryption of your data transparently with a minimal impact on performance. You don't need to modify your database client applications to use encryption.

Overview of Encrypting Amazon RDS Resources

Amazon RDS encrypted DB instances provide an additional layer of data protection by securing your data from unauthorized access to the underlying storage. You can use Amazon RDS encryption to increase data protection of your applications deployed in the cloud, and to fulfill compliance requirements for data-at-rest encryption.

Amazon RDS also supports encrypting an Oracle or SQL Server DB instance with Transparent Data Encryption (TDE). TDE can be used with encryption at rest, although using TDE and encryption at rest simultaneously might slightly affect the performance of your database. You must manage different keys for each encryption method

To manage the keys used for encrypting and decrypting your Amazon RDS resources, you use the AWS Key Management Service (AWS KMS). AWS KMS combines secure, highly available hardware and software to provide a key management system scaled for the cloud. Using AWS KMS, you can create encryption keys and define the policies that control how these keys can be used. AWS KMS supports CloudTrail, so you can audit key usage to verify that keys are being used appropriately. Your AWS KMS keys can be used in combination with Amazon RDS and supported AWS services such as Amazon Simple Storage Service (Amazon S3), Amazon Elastic Block Store (Amazon EBS), and Amazon Redshift.

For an Amazon RDS encrypted DB instance, all logs, backups, and snapshots are encrypted. A Read Replica of an Amazon RDS encrypted instance is also encrypted using the same key as the master instance when both are in the same region. If the master and Read Replica are in different regions, you encrypt using the encryption key for that region.

Snapshot backups – automatic and manual snapshots

Amazon RDS creates and saves automated backups of your DB instance. Amazon RDS creates a storage volume snapshot of your DB instance, backing up the entire DB instance and not just individual databases.

Amazon RDS creates automated backups of your DB instance during the backup window of your DB instance. Amazon RDS saves the automated backups of your DB instance according to the backup retention period that you specify. If necessary, you can recover your database to any point in time during the backup retention period.

Automated backups follow these rules:

- Your DB instance must be in the ACTIVE state for automated backups to occur. Automated backups don't occur while your DB instance is in a state other than ACTIVE, for example STORAGE_FULL.
- Automated backups and automated snapshots don't occur while a copy is executing in the same region for the same DB instance.

You can also back up your DB instance manually, by manually creating a DB snapshot

Maintaining a DB Instance

Periodically, Amazon RDS performs maintenance on Amazon RDS resources. Maintenance most often involves updates to the DB instance's underlying hardware, underlying operating system (OS), or database

engine version. Updates to the operating system most often occur for security issues and should be done as soon as possible.

Some maintenance items require that Amazon RDS take your DB instance offline for a short time. Maintenance items that require a resource to be offline include required operating system or database patching. Required patching is automatically scheduled only for patches that are related to security and instance reliability. Such patching occurs infrequently (typically once every few months) and seldom requires more than a fraction of your maintenance window.