



UNIVERSIDAD
DE
CÓRDOBA

Cloud Computing Architecting con Amazon AWS

Análisis, Diseño y Procesamiento de
Datos Aplicados a las Ciencias y a las
Tecnologías

Carlos Checa Moreno



ÍNDICE DE CONTENIDOS

EJERCICIO 1	2
EJERCICIO 2	2
EJERCICIO 3	3
EJERCICIO 4	3
EJERCICIO 5	4
EJERCICIO 6	5
EJERCICIO 7	6
EJERCICIO 8	8
REFERENCIAS.....	9

EJERCICIO SOBRE AWS RDS

Ejercicio 1

¿Qué son 'The five Vs of data'? Enumera y describe brevemente cada una de ellas.

“The five Vs of data” hace referencia a las 5 características primordiales e inherentes del *Big Data*, estas son: velocidad, volumen, valor, variedad y veracidad [1]:

- **Velocidad:** La rapidez con la que se genera y se transfiere el dato para análisis inmediato.
- **Volumen:** La cantidad total de datos total.
- **Valor:** El beneficio que se puede extraer de los datos para la toma de decisiones.
- **Variedad:** La diversidad en los tipos de datos, como datos estructurados y no estructurados.
- **Veracidad:** La calidad y precisión del dato, asegurando su confiabilidad para el análisis.

Ejercicio 2

Describe brevemente qué es AWS RDS

Amazon RDS significa Amazon Relational Database Service, es decir, es un servicio de Amazon que nos ofrece la posibilidad de administrar bases de datos relacionales. Los usuarios de este servicio podrán crear este tipo de bases de datos, configurarlas, operar con ellas o incluso escalarlas si fuera necesario [2].

Ejercicio 3

Explica las ventajas de una instancia AWS RDS con respecto a usar una base de datos on premises o una base de datos ejecutándose en una instancia EC2.

En la siguiente tabla (Tabla 1) que nos ofrece Amazon [3] podemos observar como la principal ventaja que nos ofrece hospedar una base de datos en un Managed AWS Database Service, como RDS, radica en la gran cantidad de responsabilidades que nos ahorramos.

Podemos observar como una On Premise es la que mayor esfuerzo requiere pues debemos tener un control de instalaciones de parches, mantenimiento del servidor, redes, etc. Mientras que, si hospedamos una base de datos en Amazon EC2, tendríamos menos responsabilidades, pero seguiría siendo necesario, por ejemplo, ocuparnos de las backups, parches e instalaciones de las bases de datos. Por otra parte, si eligiésemos hacer una instancia AWS RDS tendríamos que ocuparnos exclusivamente de la optimización de nuestra aplicación.

Tasks	Host Database on Premises	Host Database in Amazon EC2	Host Database in a Managed AWS Database Service
Power, HVAC, and Network	x		
Rack and Stack	x		
Server Maintenance	x		
OS Installation	x		
OS Patches	x		
Database Installation	x	x	
Database Patches	x	x	
Database Backups	x	x	
High Availability	x	x	
Scaling	x	x	
Application Optimization	x	x	x

Tabla 1. Less responsibility with managed AWS database services

Ejercicio 4

Enumera al menos 3 database engines que pueden usarse con AWS RDS.

Con AWS RDS pueden usarse las siguientes database engines: MySQL, MariaDB, PostgreSQL, Oracle y SQL Server [3].

Ejercicio 5

Sobre availability and durability. Explica las diferencias entre

1. **Multi-AZ DB Cluster**
2. **Multi-AZ DB instance**
3. **Single DB instance**
4. <https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/multi-az-db-clusters-concepts.html>

En cuanto a disponibilidad y durabilidad, las diferentes configuraciones de bases de datos presentan las siguientes diferencias [4]:

- **Multi-AZ DB Cluster:** Este despliegue ofrece alta disponibilidad con una configuración semisíncrona. Tiene una instancia principal de escritura y dos réplicas de lectura ubicadas en tres zonas de disponibilidad diferentes dentro de la misma región de AWS. Esto permite mayor capacidad para cargas de lectura, menor latencia en las escrituras, y una recuperación rápida en caso de fallos en una zona de disponibilidad. Es una opción ideal para entornos que necesitan alta disponibilidad y bajo tiempo de inactividad.
 - **Multi-AZ DB Instance:** En esta configuración, Amazon RDS crea una réplica en una segunda zona de disponibilidad que se mantiene en sincronía con la instancia principal. Sin embargo, solo la instancia principal es accesible para operaciones de lectura y escritura. Esta opción asegura disponibilidad en caso de fallos, pero no proporciona la misma capacidad de lectura o latencia de escritura que un Multi-AZ DB Cluster.
 - **Single DB Instance:** Es una configuración de bajo costo con una sola instancia en una sola zona de disponibilidad. No ofrece redundancia o conmutación por error, por lo que si falla la instancia o su zona de disponibilidad, la base de datos estará fuera de
-

servicio hasta que se recupere. Esta es una opción más simple la cual es adecuada para entornos no críticos donde la alta disponibilidad no es esencial.

En resumen, Multi-AZ DB Cluster ofrece la mayor disponibilidad y capacidad de lectura, seguido de Multi-AZ DB Instance, y Single DB Instance como opción básica sin redundancia.

Ejercicio 6

Sobre DB instance class types en instance configuration. Explica las diferencias entre:

1. **General-purpose**
2. **Memory-optimized**
3. **Compute-optimized**
4. **Burstable-performance**
5. **Optimized Reads**
6. <https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Concepts.DBInstanceClass.Types.html>

Un DB instance class type determina la capacidad de cómputo y memoria de una instancia de base de datos. Amazon RDS admite los siguientes tipos de clase de instancia, donde el asterisco (*) representa la generación, el atributo opcional y el tamaño [5]:

- General purpose – db.m*
- Memory optimized – db.z*, db.x*, db.r*
- Compute optimized – db.c*
- Burstable performance – db.t*

Podemos ver una especificación del hardware de cada instancia ofrecida por Amazon en la siguiente web: [Hardware specifications for DB instance classes](#) [6].

Cada uno de estos tipos se puede aprovechar para un objetivo en concreto:

- **General-purpose:** Estas instancias ofrecen un balance entre cómputo, memoria y rendimiento de red, lo que las hace adecuadas para una amplia variedad de cargas de trabajo de base de datos.
- **Memory-optimized:** Estas instancias están diseñadas para aplicaciones que necesitan un alto rendimiento de memoria. Son adecuadas para bases de datos con grandes volúmenes de datos.
- **Compute-optimized:** Estas instancias están diseñadas para aplicaciones que requieren un alto rendimiento de cómputo, proporcionando más potencia de procesamiento en relación con la memoria.
- **Burstable-performance:** Estas instancias permiten alcanzar niveles de rendimiento superiores en períodos de alta demanda, con un costo generalmente menor. Son útiles para aplicaciones con patrones de uso variables, como cargas de trabajo de bases de datos de baja a media actividad que pueden experimentar picos ocasionales de procesamiento.

Por último, Amazon también nos ofrece **Optimized Reads** [7]: esta opción mejora el rendimiento de las consultas en bases de datos MySQL o en Multi-AZ clusters mediante el uso de un *instance store*, que es un almacenamiento temporal a nivel de instancia. Este almacenamiento se basa en unidades NVMe SSD, que están físicamente conectadas al servidor.

Ejercicio 7

Storage type. AWS gestiona las instancias RDS mediante instancias EC2 que usan volúmenes de disco AWS EBS volumes. Explica las diferencias entre:

1. **Provisioned IOPS SSD**
 2. **General Purpose SSD**
 3. **Magnetic**
 4. https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/CHAP_Storage.html#Concepts.Storage
-

Amazon nos ofrece la siguiente tabla comparativa [8]:

	Volúmenes SSD de uso general de Amazon EBS		Volúmenes de SSD de IOPS aprovisionadas para Amazon EBS		Volúmenes de HDD con rendimiento optimizado	Volúmenes de HDD en frío
Tipo de volumen	gp3	gp2	io2 Block Express 3	io1	st1	sc1
Durabilidad	99,8 % - 99,9 % de durabilidad (0,1 % - 0,2 % tasa anual de errores)		99,999 % de durabilidad (0,001 % tasa anual de errores)	99,8 % - 99,9 % de durabilidad (0,1 % - 0,2 % tasa anual de errores)	99,8 % - 99,9 % de durabilidad (0,1 % - 0,2 % tasa anual de errores)	
Casos de uso	<ul style="list-style-type: none"> • Cargas de trabajo transaccionales • Entornos de desarrollo y pruebas • Volúmenes de arranque • Aplicaciones interactivas de baja latencia • Bases de datos de tamaño mediano y una sola instancia • Escritorios virtuales 		Cargas de trabajo que requieren lo siguiente: <ul style="list-style-type: none"> • Latencia inferior a milisegundos • Rendimiento de IOPS sostenido • Más de 64 000 IOPS o 1000 MB/s de rendimiento 	<ul style="list-style-type: none"> • Cargas de trabajo que requieren un rendimiento sostenido de IOPS o más de 16,000 IOPS • Cargas de trabajo de bases de datos con uso intensivo de operaciones de E/S 	<ul style="list-style-type: none"> • Big data • Procesamiento de registros • Data warehouses 	<ul style="list-style-type: none"> • Almacenamiento orientado al rendimiento para datos a los que se accede con poca frecuencia • Escenarios en los que es importante el costo de almacenamiento más bajo
Tamaño del volumen	1 GiB - 16 TiB		4 GiB - 64 TiB 4	4 GiB - 16 TiB	125 GiB - 16 TiB	
Máximo de IOPS por volumen	16 000 (E/S de 64 KiB)	16 000 (E/S de 16 KiB)	256 000 (E/S de 16 KiB) 5	64 000 (E/S de 16 KiB)	500	250
Rendimiento máximo por volumen	1000 MB/s	250 MB/s 1	4000 MB/s	1000 MB/s 2	500 MB/s	250 MB/s
Amazon EBS Multi-attach	No compatible		Compatible		No compatible	
Reservas de NVMe	No compatible		Compatible	No compatible	No compatible	
Volumen de arranque	Compatible				No compatible	

Tabla 2. Comparativa storage type [8]

Algunas diferencias destacables entre los tipos de almacenamiento que podemos observar son:

- **Provisioned IOPS SSD:** Este tipo de almacenamiento está diseñado específicamente para aplicaciones que requieren un rendimiento de entrada/salida (IOPS) constante y extremadamente alto, pudiendo alcanza hasta 256,000 IOPS y 4,000 MB/s de rendimiento. Además, es el único tipo de almacenamiento compatibles con Multi-Attach, lo que permite que varias instancias EC2 accedan simultáneamente al mismo volumen, una característica especialmente útil en aplicaciones que requieren alta disponibilidad.
- **General Purpose SSD:** Este almacenamiento ofrece un balance entre costo y rendimiento, lo que lo convierte en una opción versátil para la mayoría de las aplicaciones de propósito general. General Purpose SSD puede alcanzar hasta 16,000

IOPS y 1,000 MB/s de rendimiento, lo cual es adecuado para cargas de trabajo de transacciones moderadas. Su ventaja principal es el equilibrio que ofrece entre rendimiento adecuado y un costo accesible, por lo que es la opción preferida para la mayoría de las aplicaciones de mediana escala.

- **Magnetic (HDD):** Este almacenamiento se enfoca en ofrecer un bajo costo para cargas de trabajo que no requieren un alto rendimiento, como el almacenamiento de grandes volúmenes de datos de acceso infrecuente. Los volúmenes HDD están disponibles en dos variantes: st1 (rendimiento optimizado) y sc1 (almacenamiento en frío). El rendimiento es notablemente más bajo, con un máximo de 500 IOPS en st1 y 250 IOPS en sc1.

Ejercicio 8

Describe brevemente con tus palabras la característica de RDS: Managing capacity automatically with Amazon RDS storage autoscaling

1. https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_PIOPS.Autoscaling.htm

Esta característica se ocupa de aumentar automáticamente el volumen de nuestra base de datos cuando no estemos quedando sin espacio dada una carga de trabajo poco predecible [9]. Por ejemplo, si lanzamos una nueva aplicación móvil y repentinamente empieza a tener más descargas de las esperadas, es probable que se acabara excediendo el almacenamiento establecido al inicio. Para este tipo de situaciones, Amazon RDS storage autoscaling se encargaría de aumentar automáticamente el tamaño de nuestra base de datos, de tal forma que no perdiésemos usuarios.

REFERENCIAS

- [1] Robinson, S. (2023). 5V's of big data. TechTarget.
<https://www.techtarget.com/searchdatamanagement/definition/5-Vs-of-big-data>
- [2] Amazon. (s.f.). Amazon Relational Database Service. Recuperado el 14 de noviembre de 2024 de <https://aws.amazon.com/es/rds/>
- [3] AWS Academy (2024). AWS Academy Cloud Architecting Module 06 Student Guide Version 3.0.0. 16. <https://awsacademy.instructure.com/courses/94098/modules/items/8678030>
- [4] Amazon. (s.f.). Multi-AZ DB cluster deployments for Amazon RDS. Recuperado el 14 de noviembre de 2024 de <https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/multi-az-db-clusters-concepts.html>
- [5] Amazon. (s.f.). What is Amazon Relational Database Service (Amazon RDS)? Recuperado el 14 de noviembre de 2024 de <https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html#Welcome.Concepts.DBInstance.instance-class>
- [6] Amazon. (s.f.). Hardware specifications for DB instance classes. Recuperado el 14 de noviembre de 2024 de <https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Concepts.DBInstanceClass.Summary.html>
- [7] Amazon. (s.f.). Improving query performance for RDS for MySQL with Amazon RDS Optimized Reads. Recuperado el 14 de noviembre de 2024 de <https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/rds-optimized-reads.html#rds-optimized-reads-overview>
- [8] Amazon. (s.f.). Tipos de volúmenes de Amazon EBS. Recuperado el 14 de noviembre de 2024 de https://docs.aws.amazon.com/es_es/ebs/latest/userguide/ebs-volume-types.html

- [9] Amazon. (s.f.). Managing capacity automatically with Amazon RDS storage autoscaling. Recuperado el 14 de noviembre de 2024 de https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_PIOPS.Autoscaling.html
-