



**ESCUELA POLITÉCNICA
SUPERIOR DE CÓRDOBA**
Universidad de Córdoba



Data-Driven Organizations
Apartado II

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

Máster Universitario en Inteligencia Computacional e Internet de las Cosas

Autor: Marcos Rivera Gavilán

Email: i92rigam@uco.es

Fecha: 10 Enero 2025



UNIVERSIDAD DE CÓRDOBA



Índice

Introducción	3
Ejercicio 1	3
Ejercicio 2	5
Ejercicio 3	6
Ejercicio 4	6
Ejercicio 5	7
Ejercicio 6	8
Ejercicio 7	9
Ejercicio 8	9
Bibliografía	10

Introducción

Después de realizar el laboratorio ACAv3 Module 6: Guided lab: Creating an Amazon RDS Database, contesta brevemente a las siguientes preguntas.

Ejercicio 1

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

Se trata de las cinco dimensiones que ayudan a entender los desafíos y las oportunidades en el análisis y la gestión de grandes volúmenes de datos:

Volumen: Se refiere a la cantidad de datos que se generan y almacenan. En el contexto de Big Data, el volumen puede ser enorme, proveniente de diversas fuentes como redes sociales, sensores, dispositivos IoT, transacciones financieras, y más. Este alto volumen requiere capacidades de almacenamiento y procesamiento especializadas para su manejo.

Velocidad: Describe la rapidez con la que se generan, transmiten y procesan los datos. En muchos casos, los datos deben ser analizados en tiempo real o casi en tiempo real, como en aplicaciones de monitoreo en línea o en sistemas de recomendación. La alta velocidad presenta desafíos de infraestructura y de procesamiento.

Variedad: Representa la diversidad de tipos y fuentes de datos, que pueden incluir datos estructurados (como tablas en bases de datos), semi-estructurados (como archivos XML o JSON), y no estructurados (como videos, imágenes o texto). La variedad exige técnicas diferentes para integrar y analizar datos que no están en un formato uniforme.

Veracidad: Se refiere a la calidad y confiabilidad de los datos. No todos los datos que se recolectan son precisos o relevantes, y puede haber ruido, duplicados o errores. La veracidad implica trabajar en la limpieza y verificación de datos para asegurar que los análisis realizados sean confiables y útiles.

Valor: Esta dimensión enfatiza la importancia de que los datos tengan un propósito y aporten información valiosa para la organización. A pesar de que se disponga de grandes volúmenes de datos, su verdadero valor radica en la capacidad de extraer información útil que apoye en la toma de decisiones y la generación de ventajas competitivas.

Ejercicio 2

Describe brevemente qué es AWS RDS

AWS RDS (Amazon Relational Database Service) es un servicio de Amazon Web Services que facilita la configuración, operación y escalado de bases de datos relacionales en la nube. RDS permite a los usuarios administrar bases de datos sin tener que preocuparse por la infraestructura subyacente, ya que AWS se encarga de tareas como actualizaciones de software, respaldos automáticos, recuperación ante fallos y escalabilidad.

RDS es compatible con varios motores de bases de datos populares, como Amazon Aurora, MySQL, PostgreSQL, MariaDB, Oracle y Microsoft SQL Server. Esto permite a las empresas elegir la tecnología de base de datos que mejor se adapte a sus necesidades y migrar o integrar sus sistemas existentes con facilidad.

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.

AWS RDS facilita la gestión de bases de datos al automatizar tareas como configuración, backups, actualizaciones y recuperación ante fallos, en comparación con una base de datos on-premises o en EC2, donde estas tareas son manuales. Sus principales características son:

- **Gestión Automática:** RDS se despliega automáticamente.
- **Escalabilidad:** RDS aumenta la capacidad automáticamente.
- **Alta Disponibilidad:** Permite crear distintas réplicas de la instancia.
- **Seguridad Integrada:** Ofrece cifrado y control de acceso.
- **Backups Automáticos:** Permite restaurar en puntos específicos.
- **Optimización de Costes:** Permite el pago por uso.

Ejercicio 4

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

Tres motores de bases de datos que pueden usarse con AWS RDS son:

- MySQL
- PostgreSQL
- Oracle

Ejercicio 5

Sobre availability and durability. Explica las diferencias entre:

Multi-AZ DB Cluster, Multi-AZ DB instance, Single DB instance

En AWS RDS, las opciones de Multi-AZ DB Cluster, Multi-AZ DB Instance, y Single DB Instance ofrecen diferentes niveles de disponibilidad y durabilidad para las bases de datos.

Multi-AZ DB Cluster: Configuración con una instancia principal y dos réplicas en diferentes zonas de disponibilidad (AZs). Permite failover rápido y lecturas desde réplicas. Ideal para aplicaciones críticas.

Multi-AZ DB Instance: Una instancia principal y una réplica en standby en otra AZ. Solo la instancia principal maneja lecturas/escrituras; la réplica en standby es para failover en caso de fallo. Buena para alta disponibilidad sin lecturas en réplicas.

Single DB Instance: Instancia única en una sola AZ, sin réplicas ni failover automático. Adecuada para entornos de desarrollo o aplicaciones sin requisitos de alta disponibilidad.

Ejercicio 6

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

General-purpose, Memory-optimized, Compute-optimized, Burstable-performance, Optimized Reads

DB Instance Class Types en AWS RDS:

General-purpose: Balance entre cómputo, memoria y almacenamiento. Para cargas moderadas (ej. db.m5).

Memory-optimized: Alta memoria para manejar grandes volúmenes de datos en memoria. Ideal para análisis en tiempo real (ej. db.r5).

Compute-optimized: Mayor capacidad de procesamiento. Adecuado para tareas de cómputo intensivo (ej. db.c5).

Burstable-performance: Rendimiento variable con acumulación de créditos de CPU. Para cargas intermitentes (ej. db.t3).

Optimized Reads: Configuración para lecturas rápidas. Usado en aplicaciones de análisis y reporting (ej. db.r5d).

Ejercicio 7

Storage type. AWS gestiona las instancias RDS mediante instancias EC2 que usan volúmenes de disco AWS EBS volumes. Explica las diferencias entre: Provisioned IOPS SSD, General Purpose SSD, Magnetic

Tipos de almacenamiento en AWS RDS:

Provisioned IOPS SSD (io1): Almacenamiento SSD de alto rendimiento. Ideal para aplicaciones que requieren latencia baja.

General Purpose SSD (gp2/gp3): Almacenamiento SSD, adecuado para la mayoría de las aplicaciones. Ajusta IOPS según el volumen.

Magnetic (estándar): Almacenamiento HDD de bajo costo y rendimiento limitado. Usado para aplicaciones de baja prioridad.

Ejercicio 8

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

Amazon RDS Storage Autoscaling es una característica que permite ajustar automáticamente la capacidad de almacenamiento de una base de datos en RDS según las necesidades del usuario. Esto significa que, a medida que crece la base de datos y se requiere más espacio, RDS puede incrementar el almacenamiento sin necesidad de intervención manual.

Bibliografía

ResearchGate. (n.d.). *Representation of the five Vs of big data*. Retrieved March 17, 2025, from https://www.researchgate.net/figure/Representation-of-the-five-Vs-of-big-data_fig2_338166638

Segment. (n.d.). *Big data characteristics*. Segment Data Hub. Retrieved March 17, 2025, from <https://segment.com/data-hub/big-data/characteristics/>

Coursera. (n.d.). *5 Vs of big data*. Coursera Articles. Retrieved March 17, 2025, from <https://www.coursera.org/articles/5-vs-of-big-data>

Amazon Web Services. (n.d.). *Amazon RDS documentation*. Retrieved March 17, 2025, from <https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html>

GeeksforGeeks. (n.d.). *Amazon RDS – Introduction to Amazon Relational Database System*. Retrieved March 17, 2025, from <https://www.geeksforgeeks.org/amazon-rds-introduction-to-amazon-relational-database-system/>

Wikipedia. (n.d.). *Servicio de Bases de Datos Relacionales de Amazon*. Retrieved March 17, 2025, from https://es.wikipedia.org/wiki/Servicio_de_Bases_de_Datos_Relaciones_de_Amazon