

LECCIÓN 2: CONSIDERACIONES DE LA CAPA DE BASE DE DATOS.



CONSIDERACIONES DE LA BASE DE DATOS



Escalabilidad

Requisitos de almacenamiento totales

Tamaño y tipo de objeto

Durabilidad

¿Cuánto rendimiento se necesita?

¿La solución elegida será capaz de escalar en forma ascendente más adelante si resulta necesario?



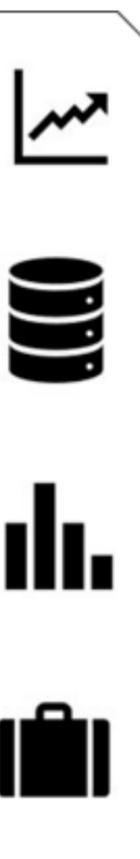
Como arquitecto, enfrentará decisiones sobre qué tipo de base de datos mejor maneja una carga de trabajo específica. Antes de tomar una decisión, hay consideraciones cruciales que deben influir en el proceso de selección.

En primer lugar, la escalabilidad es fundamental. En entornos tradicionales, escalar la capacidad de una base de datos puede ser complejo y consumir mucho tiempo, incluso para expertos. El impacto en el rendimiento durante este proceso puede ser impredecible y, a menudo, requiere tiempo de inactividad. Sin embargo, la importancia de una escalabilidad adecuada no puede ser subestimada. Un sub-aprovisionamiento puede resultar en fallos de aplicaciones, mientras que un sobre aprovisionamiento aumenta los costos al adquirir recursos innecesarios, contradiciendo el principio de optimización de costos de AWS Well-Architected Framework.

Es ideal elegir una base de datos que pueda manejar la carga de trabajo inicial y también puede escalar fácilmente hacia arriba si se requiere un mayor rendimiento en el futuro.

Además, es beneficioso contar con la capacidad de escalar hacia abajo la capacidad de la base de datos si los requisitos de rendimiento o la carga disminuyen con el tiempo. Esto permite reducir costos al instante al ajustar la capacidad. Una solución de escalado automático puede ayudar a optimizar tanto los costos del sistema como los de la mano de obra.

CONSIDERACIONES DE LA BASE DE DATOS: REQUISITOS DE ALMACENAMIENTO



Escalabilidad

Requisitos de almacenamiento totales

Tamaño y tipo de objeto

Durabilidad

¿Qué tan grande debe ser la base de datos?

¿Necesitará almacenar GB,
¿TB o petabytes de



En segundo lugar, al considerar qué base de datos utilizar para gestionar un tipo específico de carga de trabajo, es importante tener en cuenta los requisitos de almacenamiento de esa carga.

¿Cuántos datos se necesitan almacenar en la base de datos? ¿Estamos hablando de gigabytes, terabytes o incluso petabytes de datos?

Las diferentes arquitecturas de bases de datos tienen capacidades máximas de almacenamiento diferentes. Algunas están diseñadas específicamente para aplicaciones tradicionales, mientras que otras son más adecuadas para almacenamiento en caché o gestión de sesiones. Otras aún son ideales para aplicaciones de Internet de las cosas (IoT) o de big data.

Es esencial entender completamente los requisitos totales de almacenamiento para tomar una decisión informada sobre la base de datos adecuada.

CONSIDERACIONES DE LA BASE DE DATOS: TAMAÑO Y TIPO DE OBJETOS



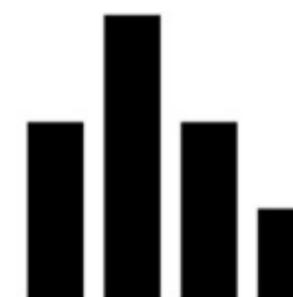
Escalabilidad

Requisitos de almacenamiento totales

Tamaño y tipo de objeto

Durabilidad

¿Es necesario almacenar estructuras de
datos simples, objetos de datos
grandes o ambos?



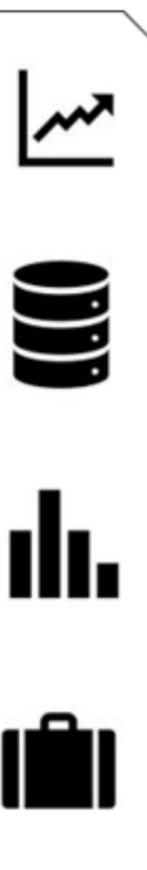


En tercer lugar, al seleccionar una base de datos para sus cargas de trabajo particulares, es importante considerar el tamaño y el tipo de objetos que necesitará almacenar.

¿Está buscando almacenar datos simples con estructuras básicas o necesita manejar objetos de datos más grandes y complejos? ¿Quizás requiere almacenar ambos tipos de objetos en su base de datos?



CONSIDERACIONES DE LA BASE DE DATOS: DURABILIDAD



Escalabilidad

Requisitos de almacenamiento totales

Tamaño y tipo de objeto

Durabilidad

¿Qué nivel de **durabilidad, disponibilidad y capacidad de recuperación de datos** se requiere?

¿Se aplican obligaciones normativas?



Finalmente, como último aspecto para tener en cuenta al seleccionar su base de datos, es importante considerar los requisitos de durabilidad de los datos que va a almacenar.

La durabilidad de los datos se refiere a la garantía de que sus datos permanezcan seguros y no se pierdan, mientras que la disponibilidad de los datos se refiere a su capacidad para acceder a ellos cuando lo necesite.

¿Qué nivel de durabilidad y disponibilidad de datos necesita para su negocio? Si los datos que está almacenando son críticos para su empresa, es crucial elegir una solución de base de datos que garantice la redundancia de datos almacenados en múltiples ubicaciones físicas separadas geográficamente. Sin embargo, es importante tener en cuenta que esta solución puede implicar costos adicionales, por lo que debe encontrar un equilibrio entre sus necesidades empresariales y las consideraciones de costos.

Otro factor importante a considerar es si existen restricciones de residencia de datos o requisitos normativos aplicables a sus datos. Por ejemplo, ¿hay regulaciones de privacidad de datos a nivel regional que deba cumplir? En tal caso, es recomendable elegir una solución de base de datos que garantice el cumplimiento de estas regulaciones.



Ahora que revisó las consideraciones clave, considere las dos categorías de opciones de base de datos disponibles:

Relacional

Ejemplos tradicionales

Microsoft SQL Server
Base de datos de Oracle
MySQL



No relacional

Ejemplos tradicionales

MongoDB
Cassandra
Redis

Existen numerosos tipos de bases de datos entre los cuales puede elegir, y muchas de estas opciones están altamente especializadas. Sin embargo, en términos generales, las bases de datos suelen clasificarse en dos categorías principales: relacionales y no relacionales.

Las bases de datos relacionales son las más comunes y conocidas para la mayoría de las personas. Ejemplos típicos incluyen Microsoft SQL Server, Oracle Database y MySQL.

Por otro lado, las bases de datos no relacionales, aunque más recientes en su desarrollo, han sido parte del panorama informático durante varias décadas y desempeñan un papel crucial en el mundo moderno de la tecnología. Algunos ejemplos notables incluyen MongoDB, Cassandra y Redis.

TIPO DE BASE DE DATOS RELACIONAL



Beneficios:

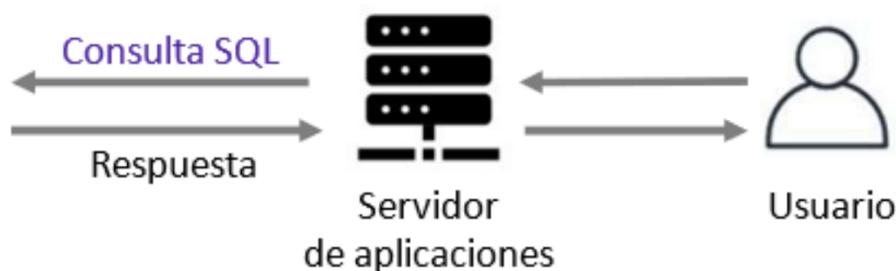
- Facilidad de uso
- Integridad de los datos
- Almacenamiento de datos reducido
- Lenguaje común (lenguaje de consulta estructurada o SQL)

La base de datos relacional es ideal cuando sucede lo siguiente:

- Necesita reglas estrictas de esquema, conformidad con ACID y cumplimiento de la calidad de los datos.
- No necesita una capacidad extrema de lectura o escritura.
- No necesita un rendimiento extremo.
 - Un RDBMS puede ser la mejor solución de menor esfuerzo



Sistema de administración de bases de datos relacionales (RDBMS)



Las bases de datos relacionales, a menudo llamadas sistemas de gestión de bases de datos relacionales (RDBMS), continúan siendo la categoría más popular y ampliamente utilizada en el mundo de la tecnología. De hecho, según un estudio de clasificación de DB-Engines en 2019, representan más del 75% de la popularidad total en todas las categorías. Los líderes en este aspecto fueron Oracle, MySQL, Microsoft SQL Server y PostgreSQL, todos ellos sistemas relacionales.

La preferencia por las bases de datos relacionales se sostiene por diversas razones. Estas incluyen su facilidad de uso, la garantía de integridad de los datos, su eficiencia en el almacenamiento y su integración con el lenguaje de consulta estructurada (SQL), el cual es ampliamente aceptado para interactuar con datos almacenados en RDBMS. SQL facilita la consulta y manipulación de datos estructurados de forma consistente y es compatible con la mayoría de los proveedores.

Si su caso de uso se ajusta a estructuras de datos rígidas, donde el esquema está bien definido y no cambia con frecuencia, una base de datos relacional sería una elección apropiada. Las bases de datos relacionales también respaldan transacciones ACID, lo que garantiza la integridad de los datos mediante transacciones atómicas, consistentes, aisladas y duraderas. Sin embargo, si su aplicación requiere una alta demanda de lectura o escritura, es posible que una base de datos relacional no sea la mejor opción.

Aunque las bases de datos relacionales ofrecen buen rendimiento en general, pueden no ser óptimas para escenarios que requieren una capacidad extrema de procesamiento. A pesar de esto, en muchos casos, una base de datos relacional sigue siendo la solución más sencilla y eficaz.

TIPO DE BASE DE DATOS NO RELACIONAL



Beneficios

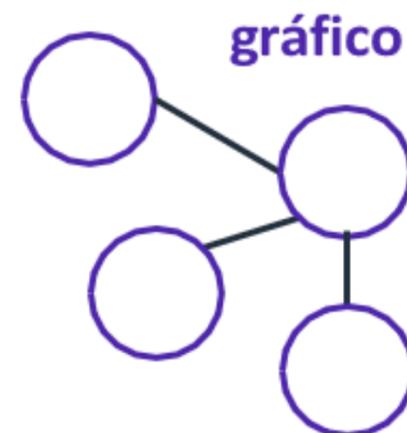
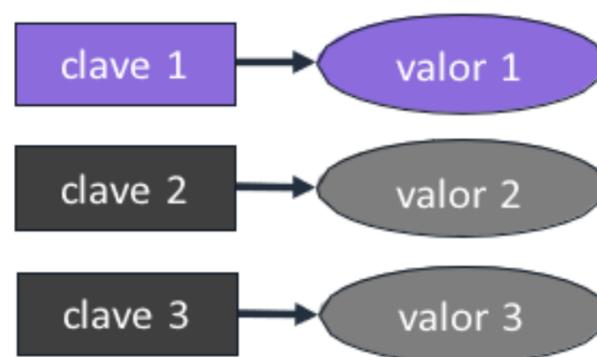
- Flexibilidad
- Escalabilidad
- Alto rendimiento
- API altamente funcionales

La base de datos no relacional es ideal cuando sucede lo siguiente:

- La base de datos debe escalarse **de forma horizontal** para manejar un volumen de datos masivo.
- Los datos no se ajustan bien a los esquemas tradicionales.
- Las tasas de lectura o escritura superan las que se pueden sustentar económicamente por medio de un RDBMS tradicional.

Modelos de ejemplo

Clave-valor



Las bases de datos no relacionales, a menudo conocidas como bases de datos NoSQL, se distinguen de las bases de datos relacionales por estar diseñadas específicamente para modelos de datos particulares, con esquemas flexibles que se adaptan a las necesidades de las aplicaciones modernas. Son muy valoradas por su flexibilidad, capacidad de escalabilidad, alto rendimiento y API altamente funcional.

Utilizan una variedad de modelos de datos, como clave-valor, gráficos, documentos, almacenamiento en memoria y búsqueda. Los esquemas en las bases de datos NoSQL son dinámicos, lo que significa que no necesariamente requieren una estructura fija para cada objeto almacenado. Además, estas bases de datos pueden escalar horizontalmente aumentando simplemente el número de servidores en los que operan.

Las bases de datos NoSQL ofrecen esquemas flexibles, lo que permite que cada objeto tenga una estructura diferente. Pueden almacenar datos estructurados, como lo haría una base de datos relacional; datos semi estructurados, como documentos JSON; y datos no estructurados, como archivos de imágenes o mensajes de correo electrónico. Son especialmente efectivas cuando los datos pueden presentar inconsistencias estructurales.

Por último, estas bases de datos están optimizadas para diferentes modelos de datos, adaptándose a las necesidades específicas de cada aplicación.

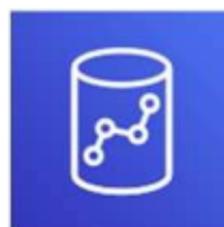
OPCIONES DE BASE DE DATOS DE AMAZON

Existen más opciones de bases de datos; estas opciones son ejemplos comunes

Bases de datos relacionales



Amazon
RDS



Amazon
Redshift



Amazon
Aurora

Bases de datos no relacionales



Amazon
DynamoDB



Amazon
ElastiCache



Amazon
Neptune

Céntrese en este módulo

Ahora que ha comprendido las diferencias entre las bases de datos relacionales y no relacionales, demos un vistazo a algunos ejemplos de ofertas de bases de datos de Amazon que se ajustan a estas dos categorías.

Entre las opciones más utilizadas de bases de datos relacionales de AWS se encuentran Amazon RDS, Amazon Redshift y Amazon Aurora.

Del mismo modo, algunas de las bases de datos no relacionales más populares son Amazon DynamoDB, Amazon ElastiCache y Amazon Neptune.

En esta unidad, nos adentraremos en el estudio de Amazon RDS y Amazon DynamoDB con mayor profundidad. Para obtener una lista completa de los servicios de bases de datos disponibles en AWS y sus casos de uso más comunes, puede consultar la página de información general de las bases de datos de AWS.

ESTOS SON ALGUNOS DE LOS APRENDIZAJES CLAVE DE ESTA LECCIÓN DE LA UNIDAD 4:

Cuando elija una base de datos, tenga en cuenta la escalabilidad, los requisitos de almacenamiento, el tipo y tamaño de los objetos que se almacenarán y los requisitos de durabilidad.

Las bases de datos relacionales tienen reglas estrictas de esquema, proporcionan integridad de los datos y admiten SQL.

Las bases de datos no relacionales se escalan de forma horizontal, proporcionan mayor escalabilidad y flexibilidad, y funcionan bien para datos semiestructurados y no estructurados.

