

LECCIÓN 3: AMAZON RDS





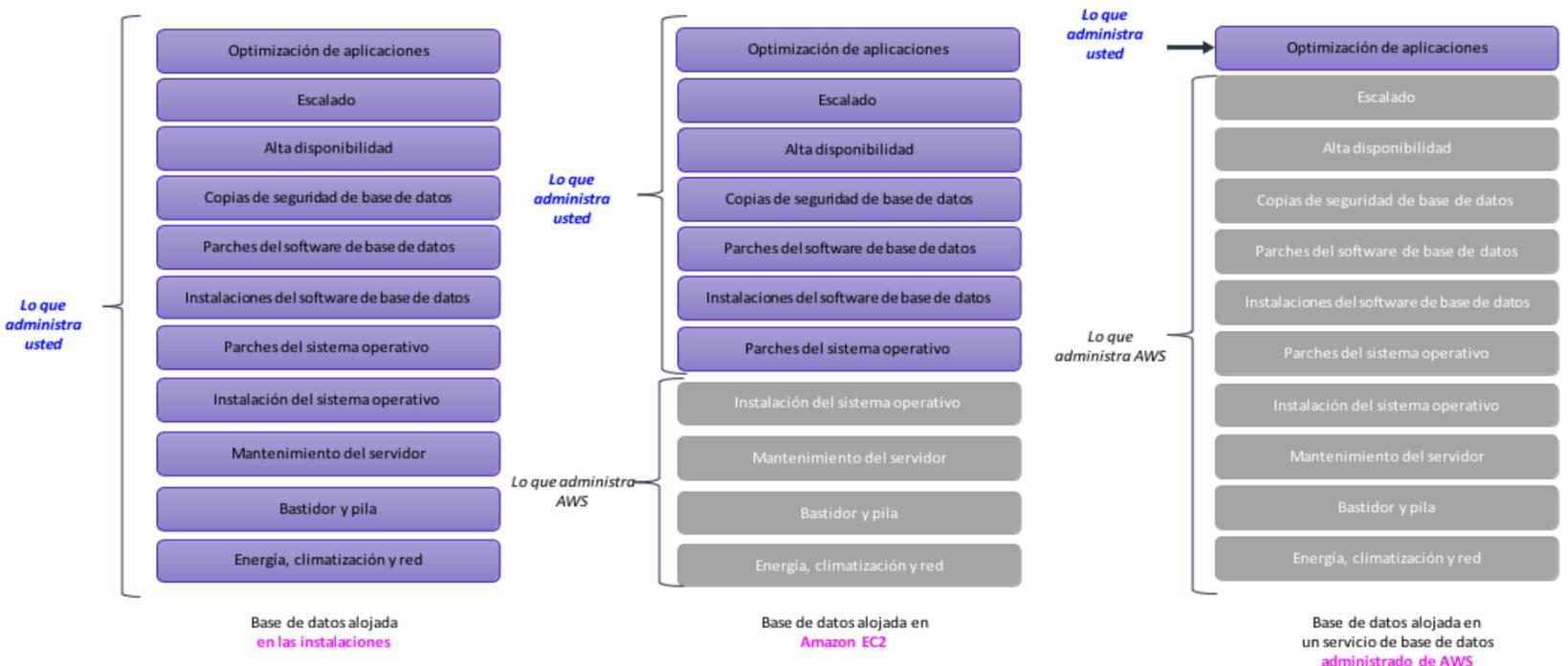
Amazon RDS es un **servicio** de base de datos relacional **completamente administrada**.



Amazon RDS es un servicio en la nube que gestiona completamente bases de datos relacionales, encargándose de su creación y operación. Antes de profundizar en Amazon RDS, exploraremos los beneficios que ofrece como una solución de bases de datos gestionada.



Ventaja de los servicios de bases de datos administradas de AWS



Ahora explicaremos algunas de las ventajas de utilizar los servicios de bases de datos administradas de AWS.

En el primer caso, albergar la base de datos en las instalaciones implica una responsabilidad total.

Esto incluye alimentar los servidores físicos, mantener los sistemas operativos, instalar y aplicar parches de software, realizar copias de seguridad, configurar soluciones de escalado y alta disponibilidad, y optimizar las aplicaciones que utilizan la base de datos.



En el segundo caso, instalar la base de datos en instancias de Amazon EC2 reduce algunas responsabilidades, ya que AWS gestiona el mantenimiento del entorno físico del centro de datos y proporciona instancias EC2 con sistemas operativos preinstalados. Sin embargo, sigue siendo su responsabilidad gestionar todas las configuraciones por encima de la instalación del sistema operativo, lo que conlleva una administración manual de varios recursos.



Por otro lado, la opción de base de datos administrada de AWS ofrece alta disponibilidad, escalabilidad y copias de seguridad integradas que se pueden configurar fácilmente. AWS se encarga de las tareas repetitivas y comunes de administración de la base de datos, mientras que usted solo necesita preocuparse por optimizar la aplicación y garantizar el buen funcionamiento de la capa de base de datos para la aplicación.

Características de Amazon RDS



Patrón de acceso
Transaccional,
análisis ligero



Volumen de datos
Rango de pocos TB



Eficacia
Rendimiento de medio
a alto, baja latencia



Casos de uso empresariales
Transaccional
OLAP

Amazon Relational Database Service (Amazon RDS) es un servicio de gestión de bases de datos relacionales que se encarga completamente de la administración y mantenimiento de la base de datos.

Normalmente, Amazon RDS se utiliza en situaciones donde el acceso a los datos sigue un patrón transaccional o para análisis que no requieren un alto nivel de complejidad.



En términos de capacidad de almacenamiento, Amazon RDS es capaz de manejar conjuntos de datos que van desde tamaños modestos hasta el rango bajo de terabytes. A medida que los requisitos de almacenamiento aumentan, se puede aprovisionar más espacio sin necesidad de tiempo de inactividad. Por ejemplo, el motor Amazon Aurora puede aumentar automáticamente el tamaño del volumen de la base de datos a medida que crecen las necesidades de almacenamiento, escalando hasta un máximo de 64 TB o el límite que se defina. Otros motores como MySQL, MariaDB, Oracle y PostgreSQL permiten escalar hasta 32 TB, mientras que Microsoft SQL Server admite hasta 16 TB. Este escalado de almacenamiento es dinámico y no interrumpe la actividad.



En cuanto al rendimiento, Amazon RDS ofrece dos opciones principales. En primer lugar, proporciona una opción de almacenamiento respaldada por unidades de estado sólido (SSD) de uso general. Esta opción ofrece un rendimiento base consistente de 3 IOPS por GB aprovisionado, con la capacidad de ampliar su uso mediante ráfagas de hasta 3000 IOPS por encima de la base de referencia. Este tipo de almacenamiento es adecuado para una amplia variedad de cargas de trabajo de bases de datos. La segunda opción es el almacenamiento en SSD con IOPS aprovisionadas, que es ideal para cargas de trabajo de bases de datos transaccionales con un alto uso de operaciones de E/S.

Amazon RDS: usos y tipos de base de datos



Amazon
RDS

Funciona bien para aplicaciones que cumplen con lo siguiente:

- Tienen datos más complejos
- Necesitan combinar y unir conjuntos de datos.
- Necesitan reglas de sintaxis aplicadas.

Se admiten seis tipos de bases de datos:

- [Microsoft SQL Server](#)
- [Oracle](#)
- [MySQL](#)
- [PostgreSQL](#)
- [Aurora](#)
- [MariaDB](#)



Como una solución de bases de datos relacionales, Amazon RDS se destaca como una excelente opción para aplicaciones que manejan conjuntos de datos complejos y altamente estructurados.

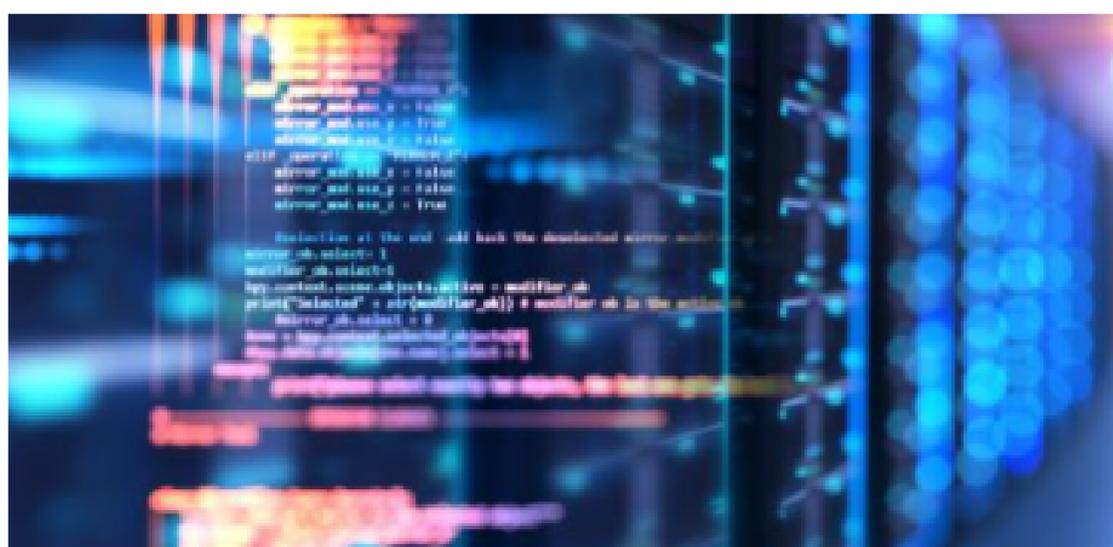
Es especialmente adecuado para situaciones en las que las cargas de trabajo necesitan fusionar y unir datos de manera frecuente, y donde se aplican reglas de sintaxis de forma estricta. Por ejemplo, Amazon RDS es comúnmente utilizado para respaldar aplicaciones tradicionales, sistemas de planificación de recursos empresariales (ERP), sistemas de gestión de relaciones con clientes (CRM) y plataformas de comercio electrónico.

Amazon RDS ofrece una selección de seis motores de bases de datos entre los cuales puede elegir, incluyendo Microsoft SQL Server, Oracle, MySQL, PostgreSQL, Amazon Aurora y MariaDB.

Tamaño de la instancia de base de datos

	Familia T	Familia M	Familia R
Tipo	Instancias ampliables	Instancias de uso general	Instancias con optimización de memoria
Tamaño	1 vCPU/1 GB de RAM a 8 vCPU/32 GB de RAM	2 vCPU/8 GB de RAM a 96 vCPU/384 GB de RAM	2 vCPU/16 GB de RAM a 96 vCPU/768 GB de RAM
Redes	Rendimiento moderado	Alto rendimiento	Alto rendimiento
Carga de trabajo ideal	Menor o variable	Mucho consumo de CPU	Recuentos altos de conexiones con uso intensivo de consultas
Aspectos destacados	T3 puede ampliarse por encima de la referencia por un cargo adicional	M5 ofrece hasta 96 vCPU	R5 ofrece hasta 96 vCPU/768 GiB de RAM

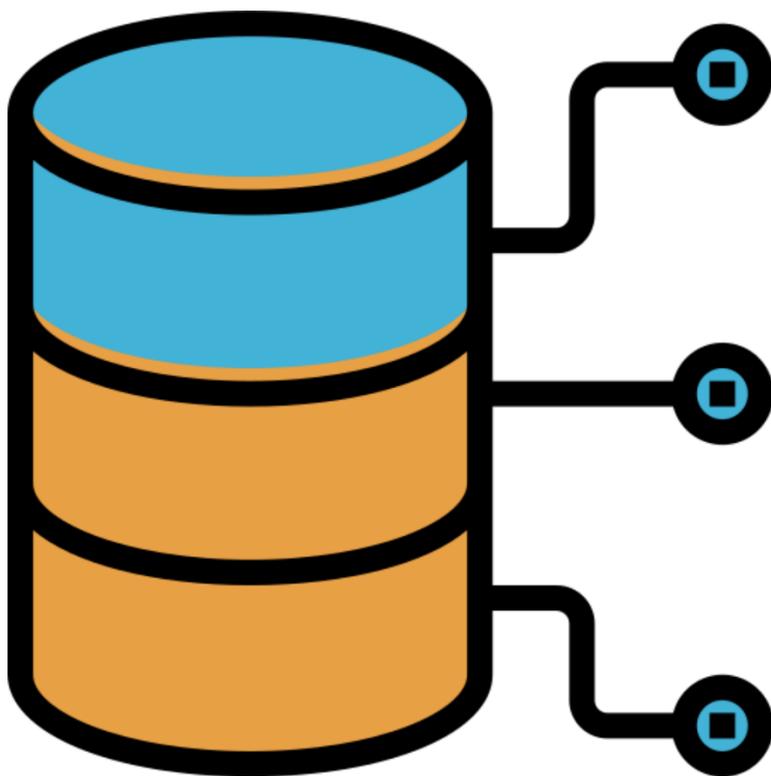
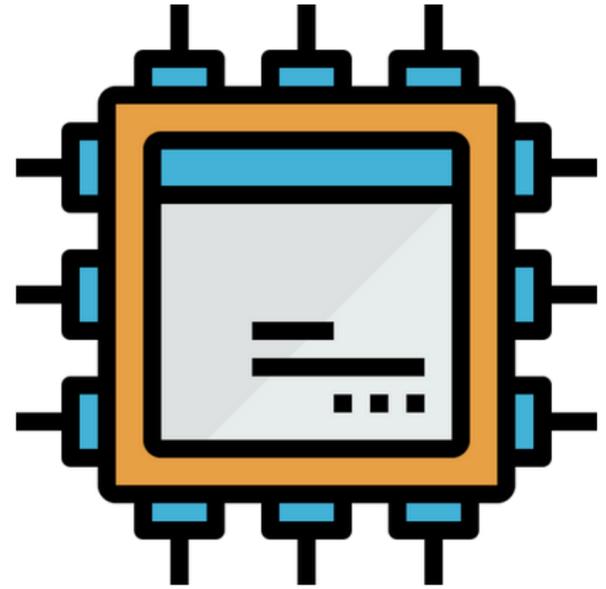
Cada tipo de base de datos en Amazon RDS se ejecuta en un servidor, a excepción de Aurora, que ofrece una opción sin servidor. Además, Amazon RDS ofrece una variedad de tipos de instancias de bases de datos, cada uno optimizado para diferentes tipos de cargas de trabajo.



La familia T de tipos de instancia proporciona un nivel básico de rendimiento de la CPU con la capacidad de aumentar su uso mediante ráfagas según sea necesario. Por ejemplo, las instancias T3 ofrecen un equilibrio entre recursos informáticos, de memoria y de red, diseñadas para cargas de trabajo de bases de datos que tienen un uso moderado de la CPU y pueden experimentar picos temporales de uso.

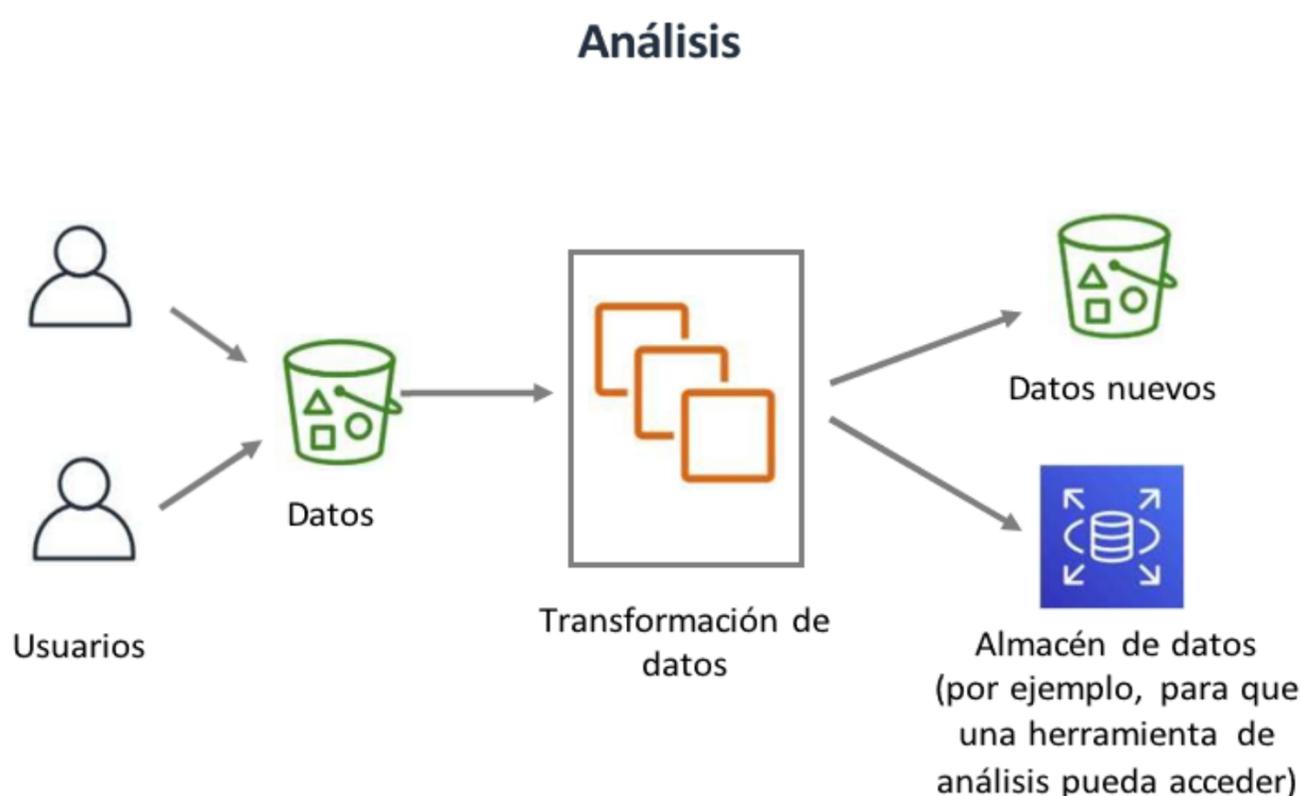
Otra opción es la familia M de instancias, que también es de uso general, pero ofrece más opciones para cargas de trabajo que requieren un uso intensivo de la CPU. Estas instancias son ideales para bases de datos pequeñas y medianas utilizadas en aplicaciones empresariales o de código abierto.

Por último, las instancias de la familia R están diseñadas para manejar cargas de trabajo de bases de datos que requieren un uso intensivo de la memoria.



Algunos motores de base de datos admiten clases de instancias de bases de datos adicionales. Para obtener más información sobre las opciones disponibles, consulte la documentación de AWS sobre las instancias de base de datos de Amazon RDS y, específicamente, la página que detalla cómo elegir la clase de instancias de bases de datos para obtener más detalles sobre las diferentes opciones disponibles

Amazon RDS: ejemplo de caso de uso



Imaginemos un escenario empresarial donde Amazon RDS forma parte de una solución implementada para abordar un desafío específico.

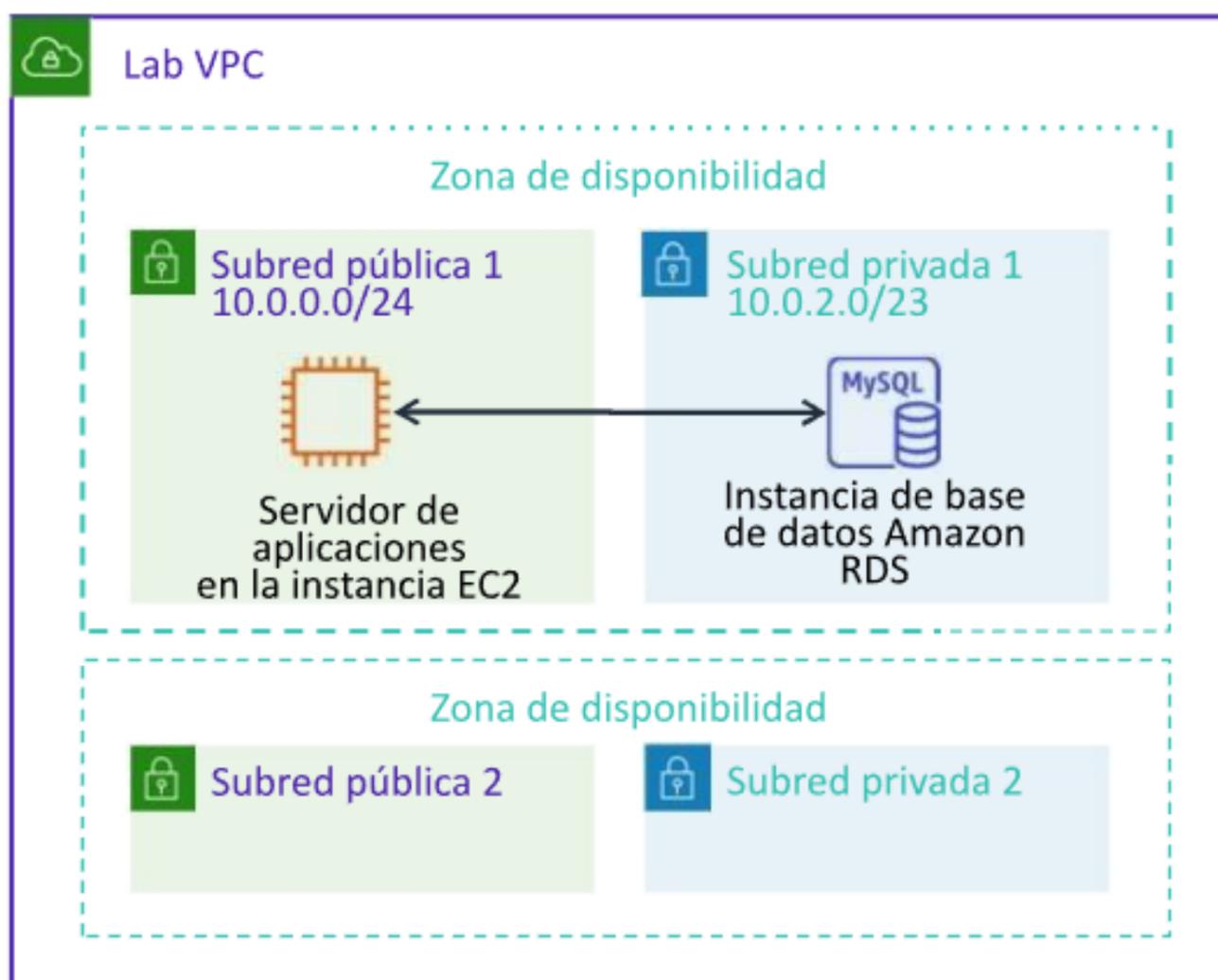
En este ejemplo, una empresa busca obtener información valiosa de sus datos comerciales. Los usuarios cargan los datos en un depósito de Amazon Simple Storage Service (Amazon S3). Luego, un grupo de instancias de Amazon EC2 se utiliza para procesar y transformar los datos. Este proceso puede realizarse de manera programada y por lotes a intervalos regulares. Posteriormente, los datos transformados se almacenan en otro depósito de S3. Sin embargo, algunos de estos datos también se insertan en una instancia de RDS designada como el "Almacén de datos". Esta instancia de RDS se convierte en una fuente accesible para consultas, ya sea mediante software de análisis externo o utilizando un servicio de análisis de AWS capaz de ejecutar consultas SQL para extraer información valiosa con fines comerciales.



AHORA COMPLETARÁ LA UNIDAD 4 - LABORATORIO GUIADO: CREACIÓN DE UNA BASE DE DATOS DE AMAZON RDS.

Creación de una base de datos de Amazon RDS.

Configuración de la comunicación de aplicaciones web con una instancia de base de datos



En este laboratorio guiado, realizará las siguientes tareas:

1. Creación de una base de datos de Amazon RDS
2. Configuración de la comunicación de aplicaciones web con una instancia de base de datos

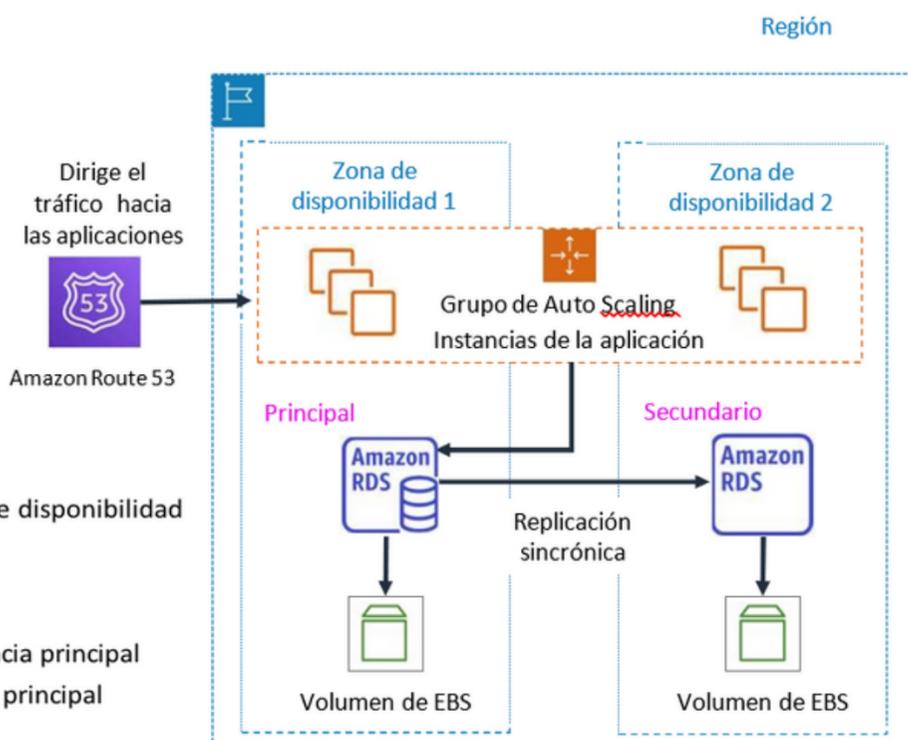
Implementación Multi-AZ para alta disponibilidad

Beneficios

- Durabilidad mejorada
- Mayor disponibilidad
- La conmutación por error al modo de espera se produce automáticamente

Condiciones de la conmutación por error automatizada

- Pérdida de disponibilidad en la zona de disponibilidad principal
- Pérdida de conectividad de red con la instancia principal
- Error de unidad informática en la instancia principal
- Error de almacenamiento en instancia principal



Amazon RDS garantiza la alta disponibilidad mediante la implementación de un enfoque Multi-AZ. Esto implica que Amazon RDS automáticamente crea y mantiene una instancia de base de datos en espera, sincrónica, en una zona de disponibilidad diferente. La base de datos principal se replica de manera sincrónica en esta instancia en espera, ubicada en distintas zonas de disponibilidad, lo que proporciona redundancia de datos. Esta configuración ayuda a evitar interrupciones en las operaciones de entrada/salida y minimiza los tiempos de latencia durante las copias de seguridad del sistema.

La ejecución de una instancia de base de datos con alta disponibilidad puede mejorar la continuidad del servicio durante el mantenimiento planificado del sistema. Además, protege las bases de datos contra errores en la instancia principal y contra interrupciones en la zona de disponibilidad. Esta capacidad de alta disponibilidad y durabilidad hace que las instancias de base de datos sean ideales para cargas de trabajo críticas en producción.



En caso de una falla de infraestructura, Amazon RDS realiza automáticamente una conmutación por error a la instancia en espera. Una vez completada la conmutación por error, las operaciones de la base de datos pueden reanudarse. Importante hay que destacar que el punto de conexión de la instancia de base de datos permanece inalterado después de la conmutación por error, lo que permite que las aplicaciones continúen operando sin necesidad de intervención manual por parte del administrador.

Las condiciones que pueden desencadenar una conmutación por error incluyen la pérdida de disponibilidad en la zona de disponibilidad principal, la desconexión de red con la base de datos principal, fallas en la unidad informática de la instancia principal o problemas de almacenamiento en la instancia principal.

Réplicas de lectura para el rendimiento

Beneficios

- Rendimiento mejorado
- Mayor disponibilidad
- Diseñado para garantizar la seguridad

Compatible con

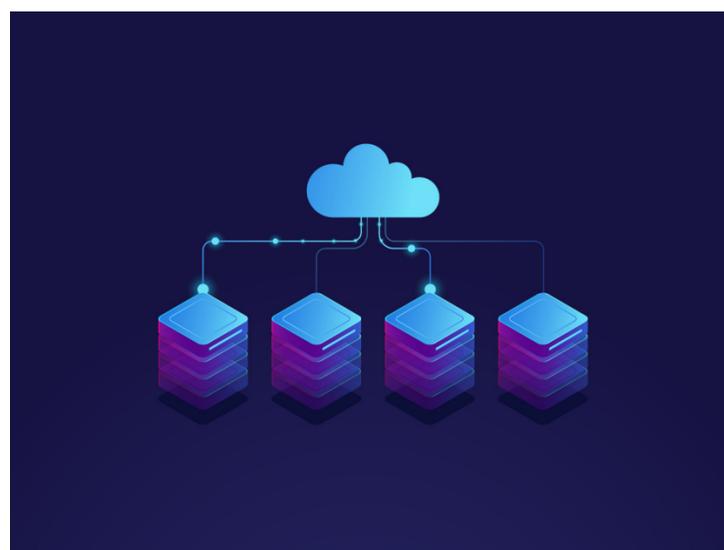
- MySQL
- MariaDB
- PostgreSQL
- Oracle

Límites

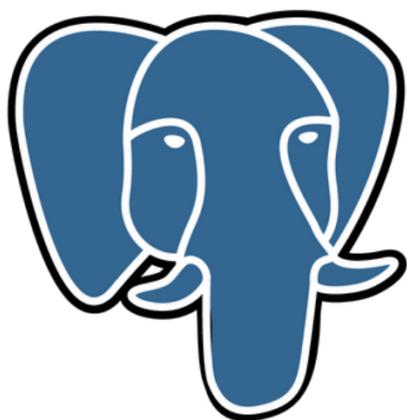
- Cinco réplicas de lectura por instancia principal
- Para obtener una consistencia estricta de lectura después de escritura, lea desde la instancia principal



Amazon RDS ofrece la capacidad de crear réplicas de lectura utilizando una instancia de base de datos principal como punto de partida. Estas réplicas de lectura son copias especiales de la instancia de origen, actualizadas de forma asíncrona conforme se realizan cambios en la instancia principal.

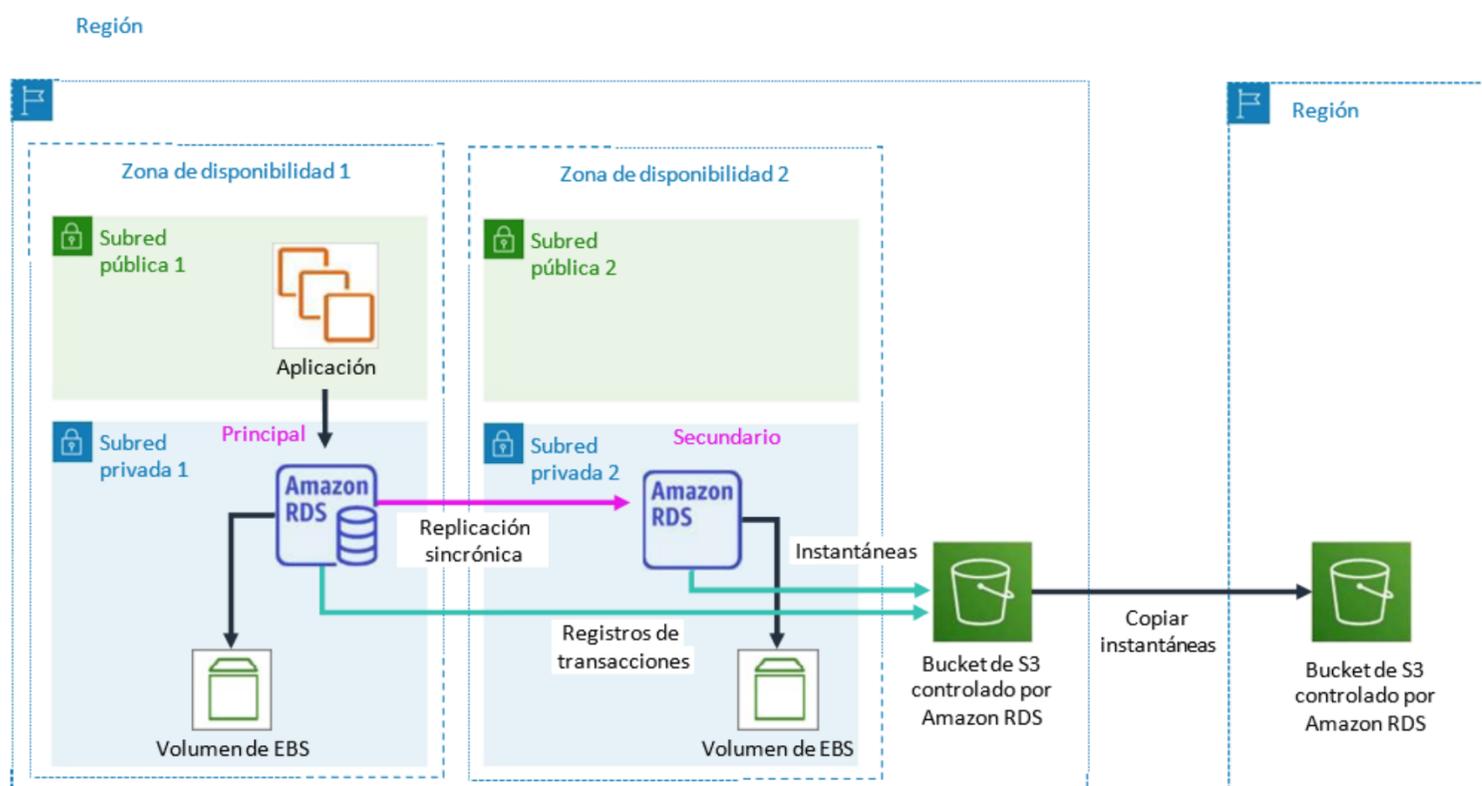


El uso de réplicas de lectura proporciona mejoras significativas en el rendimiento. Por ejemplo, puede aliviar la carga de la instancia de base de datos principal redirigiendo consultas de lectura hacia las réplicas de lectura. Además, las réplicas de lectura mejoran la disponibilidad del sistema. Para cargas de trabajo que requieren un alto número de lecturas, es posible escalar más allá de los límites de capacidad de una sola instancia de base de datos mediante el uso de réplicas de lectura. Estas réplicas también pueden ser promovidas manualmente para convertirse en instancias de base de datos independientes según sea necesario.



Las réplicas de lectura están disponibles en Amazon RDS para MySQL, MariaDB, PostgreSQL y Oracle. Cada uno de estos motores de base de datos admite hasta un máximo de cinco réplicas de lectura por instancia principal. Es importante tener en cuenta que si se necesita una consistencia estricta entre las operaciones de lectura y escritura (es decir, que lo que se lee siempre sea lo que se acaba de escribir, incluso si se lee de inmediato), entonces se debe acceder a la instancia de base de datos principal. De lo contrario, para distribuir la carga y mejorar el rendimiento, se puede leer desde una réplica de lectura.

Réplicas de lectura para el rendimiento



Consideremos un ejemplo de arquitectura para una solución de copia de seguridad utilizando Amazon RDS.

En esta configuración, tanto las instantáneas como los registros de transacciones de la base de datos de RDS se guardan en un depósito de S3 administrado por Amazon RDS. Amazon RDS ofrece la opción de configurar copias de seguridad automáticas o manuales de la base de datos en una segunda región de AWS. Además, es posible copiar las instantáneas en cuentas de AWS separadas.

Para obtener más detalles sobre cómo configurar este proceso, se puede consultar la documentación de AWS sobre cómo Copiar una instantánea.

Ahora, el instructor puede elegir demostrar la configuración de copias de seguridad automatizadas y réplicas de lectura en Amazon RDS mediante la consola de administración de AWS.

Amazon Aurora

Amazon Aurora es un motor de base de datos relacional completamente administrada compatible con MySQL y PostgreSQL.

Se usa para el procesamiento de transacciones en línea (OLTP).

Ofrece hasta cinco veces más rendimiento que MySQL

Ofrece hasta tres veces más rendimiento que PostgreSQL

Replica datos de seis maneras en tres zonas de disponibilidad.

Requiere pocos cambios en su aplicación existente.

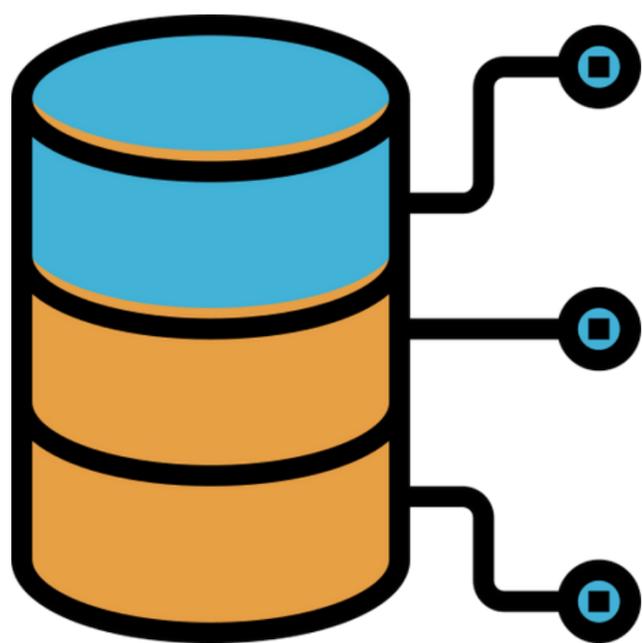


Amazon
Aurora

Los detalles del punto de referencia están disponibles para MySQL y PostgreSQL

Amazon Aurora se presenta como una opción de motor de base de datos dentro de Amazon RDS.

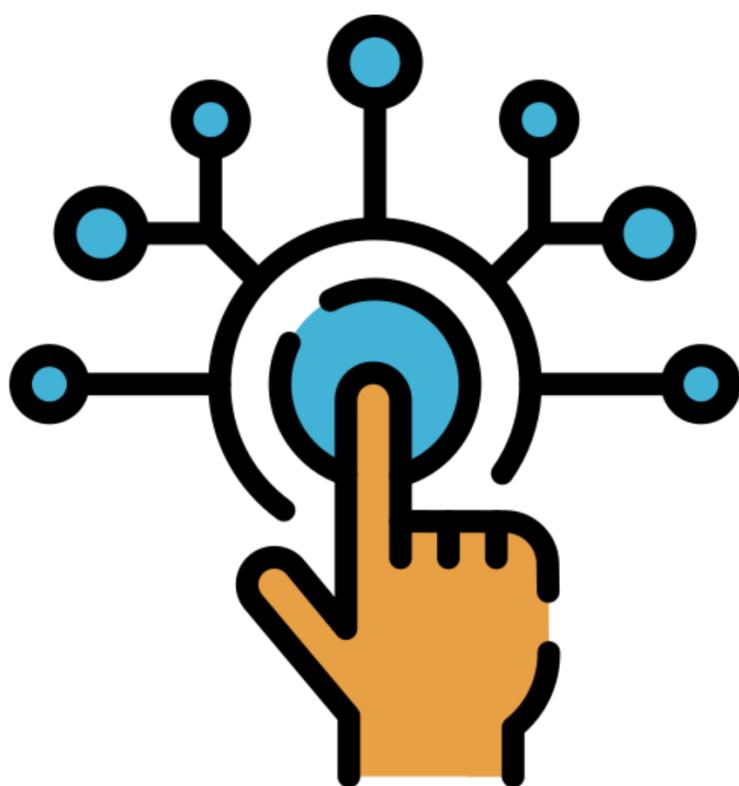
Aurora es una solución de base de datos relacional, compatible con MySQL y PostgreSQL, diseñada específicamente para la nube. Integra el rendimiento y la disponibilidad típicos de las bases de datos empresariales convencionales con la simplicidad y la economía propias de las bases de datos de código abierto.



Comparada con las bases de datos MySQL y PostgreSQL estándar, Aurora puede ser hasta cinco veces más veloz en el caso de MySQL y tres veces más rápido en el de PostgreSQL. Ofrece los niveles de seguridad, disponibilidad y confiabilidad propios de las bases de datos empresariales a una fracción del costo.

La documentación detallada sobre estas capacidades y cómo aprovecharlas se encuentra disponible en las guías de referencia de rendimiento de Amazon Aurora MySQL y Amazon Aurora PostgreSQL.

Aurora proporciona un sistema de almacenamiento distribuido, robusto y con recuperación automática que puede ajustar su escala automáticamente hasta alcanzar los 64 TB por instancia de base de datos. Además, ofrece un alto rendimiento y disponibilidad, incluyendo hasta 15 réplicas de lectura de baja latencia, recuperación a un momento dado, copias de seguridad continuas en Amazon S3 y replicación en tres zonas de disponibilidad.



Se utiliza comúnmente para el procesamiento de transacciones en línea (OLTP), que implica la capacidad de manejar grandes volúmenes de usuarios simultáneos y ejecutar operaciones de inserción y actualización. Un ejemplo típico de OLTP es un sistema de entrada de pedidos. Es importante distinguir el OLTP del procesamiento de análisis en línea (OLAP), que se caracteriza por un menor volumen de consultas más complejas.

Amazon Redshift

Amazon Redshift es un servicio de almacenamiento de datos.

Se utiliza para el procesamiento de análisis en línea (OLAP).

Almacena conjuntos muy grandes de datos.

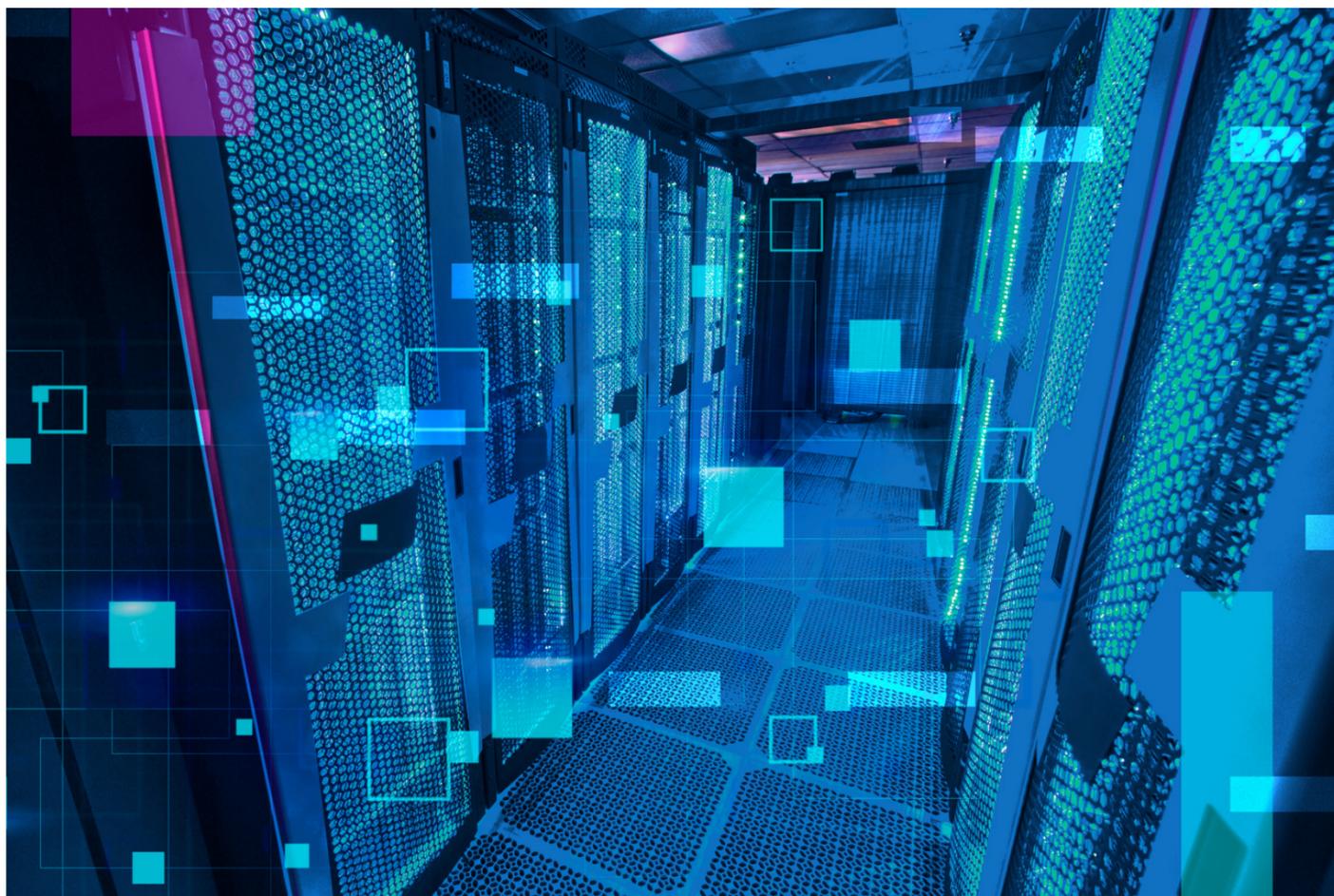
- Almacena datos muy estructurados y de acceso frecuente en Amazon Redshift.
- También puede almacenar exabytes de datos estructurados, semiestructurados y no estructurados en Amazon S3.



Amazon Redshift es una alternativa de base de datos relacional proporcionada por AWS, pero a diferencia de Amazon RDS, opera de manera independiente.

Este servicio está diseñado para gestionar almacenamientos de datos a gran escala, llegando incluso a petabytes, y permite análisis de lago de datos. Usualmente, Amazon Redshift se emplea para almacenar datos altamente estructurados que se acceden con regularidad. Además, tiene la capacidad de acceder directamente a los datos almacenados en Amazon S3, lo que significa que puede administrar cantidades masivas de datos estructurados, semiestructurados y no estructurados en Amazon S3.

Amazon Redshift garantiza un rendimiento rápido y constante, incluso bajo cargas de trabajo intensivas con miles de consultas simultáneas. Esto se mantiene consistente tanto si se accede a los datos almacenados en el propio servicio de almacenamiento de datos de Amazon Redshift, como si se accede al lago de datos de Amazon S3. Además, es compatible con una variedad de formatos de archivo abiertos como Parquet, JSON, Avro y CSV, y permite realizar consultas directamente a Amazon S3 mediante SQL.



Estos son algunos de los aprendizajes clave de esta sección del unidad:

- Los servicios de bases de datos administradas de AWS manejan las tareas de administración para que pueda centrarse en sus aplicaciones
- Amazon RDS admite Microsoft SQL Server, Oracle, MySQL, PostgreSQL, Aurora y MariaDB
- Las implementaciones Multi-AZ de Amazon RDS ofrecen alta disponibilidad con conmutación por error automática
- Puede tener hasta cinco réplicas de lectura por base de datos principal para mejorar el rendimiento
- Amazon Aurora es un motor de base de datos relacional completamente administrada compatible con MySQL y PostgreSQL
- Amazon Redshift es una oferta de base de datos relacional de almacenamiento de datos

