

BASES DE DATOS NOSQL EN AWS

CONTENIDO

BASES DE DATOS NOSQL EN AWS	2
INTRODUCCIÓN	2
¿QUÉ ES UNA BASE DE DATOS NOSQL?	2
Ventajas de las bases de datos NoSQL	3
Casos de uso de las bases de datos NoSQL	4
TIPOS DE BASES DE DATOS NOSQL.....	5
Base de datos de clave-valor	5
Bases de datos de documentos	5
Bases de datos de grafos	5
Bases de datos en memoria.....	5
Bases de datos de motores de búsqueda.....	6
¿CUÁNDO ELEGIR BASES DE DATOS NOSQL EN LUGAR DE BASES DE DATOS SQL?	8
AMAZON DYNAMODB	9
CARACTERÍSTICAS DE AMAZON DYNAMODB.....	10
VENTAJAS DE USAR AMAZON DYNAMODB.....	11
COMPARACIÓN ENTRE AMAZON RDS Y AMAZON DYNAMODB	12
AMAZON DOCUMENTDB.....	13
CARACTERÍSTICAS DE AMAZON DOCUMENTDB.....	14
COMPARACIÓN ENTRE AMAZON DYNAMODB Y AMAZON DOCUMENTDB	17
AMAZON NEPTUNE	18
CARACTERÍSTICAS DE AMAZON NEPTUNE	19

BASES DE DATOS NOSQL EN AWS

INTRODUCCIÓN

Las bases de datos NoSQL están diseñadas para modelos de datos específicos y almacenan los datos en esquemas flexibles que se escalan con facilidad para aplicaciones modernas. Las bases de datos NoSQL son ampliamente reconocidas porque son fáciles de desarrollar, por su funcionalidad y el rendimiento a escala.

A continuación, vamos a explicar algunos conceptos para comprender mejor las bases de datos NoSQL.

¿QUÉ ES UNA BASE DE DATOS NOSQL?

Una base de datos NoSQL es un enfoque de diseño de bases de datos que permite almacenar y consultar datos fuera de las estructuras tradicionales presentes en las bases de datos relacionales. A diferencia de las bases de datos SQL, que utilizan tablas y relaciones, las bases de datos NoSQL adoptan diferentes modelos de datos, como documentos, grafos o columnas.

Estas bases de datos son muy útiles para organizar y gestionar información no estructurada o semiestructurada, y están diseñadas para soportar grandes volúmenes de datos. A diferencia de las bases de datos relacionales, no utilizan el lenguaje SQL para consultas, aunque pueden usarlo como herramienta de apoyo.

Las bases de datos NoSQL son bases de datos no relacionales, son más flexibles a diferencia de las bases de datos SQL, que utilizan tablas con columnas fijas y filas. No siguen un esquema rígido y pueden adaptarse a diferentes formatos de datos, como documentos, grafos o claves. Algunos ejemplos populares de bases de datos NoSQL incluyen MongoDB, Cassandra y Redis.

Ejemplo de datos en una base de datos no relacional:

Clave	Valor
1	Nombre: John Doe Dirección: Cualquier calle 123 Bebida favorita: café con leche mediano
2	Nombre: Mary Major Dirección: Calle principal 100 Fecha de nacimiento: 5 de julio de 1994

Ventajas de las bases de datos NoSQL

Las aplicaciones modernas se enfrentan a varios desafíos que las bases de datos NoSQL pueden resolver. Por ejemplo, las aplicaciones procesan un gran volumen de datos de fuentes dispares, como las redes sociales, los sensores inteligentes y las bases de datos de terceros. Todos estos datos dispares no encajan perfectamente en el modelo relacional. La aplicación de estructuras tabulares puede provocar redundancia, duplicación de datos y problemas de rendimiento a escala.

Las bases de datos NoSQL están diseñadas específicamente para modelos de datos no relacionales y tienen esquemas flexibles para crear aplicaciones modernas. Son ampliamente reconocidas por su facilidad de desarrollo, su funcionalidad y el rendimiento a escala. Algunos de los beneficios de las bases de datos NoSQL son:

Flexibilidad: Las bases de datos NoSQL generalmente ofrecen esquemas flexibles que permiten un desarrollo más rápido y más iterativo. El modelo de datos flexible hace que las bases de datos NoSQL sean ideales para datos semiestructurados y no estructurados.

Escalabilidad: Las bases de datos NoSQL generalmente están diseñadas para escalar horizontalmente usando clústeres distribuidos de hardware, en lugar de escalar añadiendo servidores caros y sólidos. Algunos proveedores de la nube manejan estas operaciones en segundo plano, como un servicio completamente administrado.

Alto rendimiento: Las bases de datos NoSQL están optimizadas para modelos de datos y patrones de acceso específicos. Esto permite un mayor rendimiento que si intentara lograr una funcionalidad similar con bases de datos relacionales.

Altamente funcional: Las bases de datos NoSQL proporcionan API (Interfaz de programación de aplicaciones: Pieza de código que permite a diferentes aplicaciones comunicarse entre sí y compartir información y funcionalidades) altamente funcionales y tipos de datos que están diseñados específicamente para cada uno de sus respectivos modelos de datos.

Casos de uso de las bases de datos NoSQL

Puede usar bases de datos NoSQL para crear una amplia variedad de aplicaciones móviles, de Internet de las cosas (IoT), de juegos y web de alto rendimiento que proporcionan excelentes experiencias de usuario a escala. La gama de bases de datos NoSQL y sus respectivos casos de uso son muy variados. Si bien es difícil presentar un conjunto representativo de casos de uso, a continuación, proporcionamos algunos ejemplos ilustrativos:

Administración de datos en tiempo en tiempo real: Puede ofrecer recomendaciones en tiempo real, personalización y experiencias de usuario mejoradas con bases de datos NoSQL. Por ejemplo, Disney+ ofrece su amplia biblioteca de contenido digital a más de 150 millones de suscriptores mediante la tecnología de base de datos NoSQL. Puede escalar y ofrecer funciones populares, como Seguir viendo, Mi lista y Recomendaciones personalizadas con Amazon DynamoDB.

Seguridad en la nube: Puede usar bases de datos de grafos para descubrir rápidamente relaciones complejas en sus datos. Por ejemplo, Wiz reimaginó la seguridad en la nube como un gráfico con Amazon Neptune. Wiz ayuda a sus clientes a mejorar su postura de seguridad al identificar y corregir rápidamente los riesgos más críticos. Utilizan un modelo gráfico almacenado en Amazon Neptune para descubrir la combinación tóxica de factores de riesgo que representan riesgos críticos. Los motores de riesgo de Wiz recorren el gráfico y, en cuestión de segundos, entrelazan una serie de factores de riesgo interconectados en un gráfico de seguridad.

Aplicaciones de alta disponibilidad: Las bases de datos NoSQL distribuidas son excelentes para crear aplicaciones de alta disponibilidad y baja latencia para mensajería, redes sociales, uso compartido de archivos y más. Por ejemplo, Snapchat tiene más de 290 millones de usuarios que envían miles de millones de imágenes y mensajes de video a diario. Utiliza sistemas de bases de datos NoSQL para reducir la latencia media del envío de mensajes en un 20%.

TIPOS DE BASES DE DATOS NOSQL

Existen varios sistemas de bases de datos NoSQL diferentes debido a las variaciones en la forma en que administran y almacenan los datos sin esquemas. A continuación, explicamos algunos de los tipos más comunes:

Base de datos de clave-valor: Las bases de datos de clave-valor son altamente particionables y permiten un escalado horizontal en niveles que otros tipos de bases de datos NoSQL no pueden lograr. Una base de datos de clave-valor almacena datos como un conjunto de pares clave-valor en los que una clave sirve como un identificador único. Las claves y los valores pueden ser cualquier cosa, desde objetos simples hasta objetos compuestos complejos. Los casos de uso como juegos, tecnología publicitaria e IoT se prestan particularmente bien con el diseño de datos de almacén clave-valor. Amazon DynamoDB está diseñado para proporcionar un rendimiento constante con una latencia de milisegundos de un solo dígito para cualquier escala de cargas de trabajo.

Bases de datos de documentos: Las bases de datos de documentos tienen el mismo formato de modelo de documento que los desarrolladores utilizan en el código de sus aplicaciones. Almacenan los datos como objetos JSON que son flexibles, semiestructurados y de naturaleza jerárquica. La naturaleza flexible, semiestructurada y jerárquica de los documentos y las bases de datos de documentos permite que evolucionen según las necesidades de las aplicaciones. El modelo de base de datos de documentos funciona bien con catálogos, perfiles de usuario y sistemas de administración de contenido en los que cada documento es único y evoluciona con el tiempo. Amazon DocumentDB (con compatibilidad para MongoDB) y MongoDB son bases de datos de documentos conocidas que proporcionan unas API poderosas e intuitivas para un desarrollo flexible e iterativo.

Bases de datos de grafos: El objetivo de las bases de datos de grafos es facilitar la creación y ejecución de aplicaciones que funcionen con conjuntos de datos con un nivel alto de conexión. Usan nodos para almacenar entidades de datos y periféricas para almacenar relaciones entre entidades. Un borde siempre tiene un nodo inicial, un nodo final, un tipo y una dirección. Puede describir las relaciones entre elementos principales y secundarios, las acciones, la propiedad y similares. No hay límite para la cantidad y el tipo de relaciones que un nodo puede tener. Los casos de uso típicos para una base de datos de grafos incluyen redes sociales, motores de recomendaciones, detección de fraude y gráficos de conocimiento. Amazon Neptune es un servicio de base de datos de grafos totalmente administrado.

Bases de datos en memoria: Mientras que otras bases de datos no relacionales almacenan datos en discos o SSD, los almacenes de datos en memoria están diseñados para eliminar la

necesidad de acceder a los discos. Son ideales para aplicaciones que requieren tiempos de respuesta de microsegundos o que tienen grandes picos de tráfico. Puede usarlas en aplicaciones de juegos y tecnología publicitaria para funciones como tablas de clasificación, tiendas de sesiones y análisis en tiempo real. Amazon MemoryDB para Redis es un servicio de base de datos en memoria duradero y compatible con Redis que ofrece latencia de lectura de microsegundos, latencia de escritura de milisegundos de un solo dígito y durabilidad Multi-AZ. Amazon ElastiCache es un servicio de almacenamiento en caché en memoria completamente administrado compatible con Redis y Memcached, para atender cargas de trabajo de baja latencia y alto rendimiento. El Acelerador de Amazon DynamoDB (DAX) es otro ejemplo de un almacén de datos diseñado específicamente que hace que las lecturas de DynamoDB sean mucho más rápidas.

Bases de datos de motores de búsqueda: Una base de datos de motores de búsqueda es un tipo de base de datos no relacional que se dedica a la búsqueda de contenido de datos, como los registros de salida de las aplicaciones que utilizan los desarrolladores para solucionar problemas. Utilizan índices para categorizar características similares entre los datos y facilitar la capacidad de búsqueda. Las bases de datos de los motores de búsqueda están optimizadas para clasificar datos no estructurados, como imágenes y videos. Amazon OpenSearch Service está diseñado para proporcionar visualizaciones en tiempo real y análisis de datos generados por máquinas al indexar, agregar y buscar registros y métricas semiestructuradas.

DIFERENCIAS ENTRE LAS BASES DE DATOS NOSQL y SQL

Durante décadas, el modelo de datos predominante en el desarrollo de aplicaciones fue el modelo de datos relacional que almacenaba los datos en tablas compuestas por filas y columnas. Se utilizó lenguaje de consulta estructurado (SQL) para crear y editar estas tablas relacionales. Las bases de datos SQL modelan las relaciones de datos como tablas. Las filas de la tabla representan una recopilación de valores relacionados de un objeto o una entidad. Cada columna de la tabla representa un atributo de los datos, y un campo (o celda de la tabla) almacena el valor real del atributo. Puede usar un sistema de administración de bases de datos relacionales (RDBMS) para acceder a los datos de muchas maneras diferentes sin tener que reorganizar las propias tablas de la base de datos.

No fue sino hasta mediados y finales de la década del 2000 que otros modelos de datos flexibles comenzaron a adoptarse y aumentó su uso significativamente. Para diferenciar y categorizar estas nuevas clases de bases de datos y modelos de datos, se adoptó el término NoSQL, que significa no solo SQL o sin SQL. Con frecuencia, los términos NoSQL y no relacional se usan indistintamente. Las diferencias clave entre las bases de datos relacionales y no relacionales se muestran en la siguiente tabla:

	BASES DE DATOS RELACIONALES	BASES DE DATOS NoSQL
Cargas de trabajo óptimas	Las BD relacionales están diseñadas para aplicaciones de OLTP (procesamiento de transacciones en línea) altamente coherentes y transaccionales. También son buenas para el OLAP (procesamiento analítico en línea)	Las bases de datos NoSQL están diseñadas para varios patrones de acceso a datos que incluyen aplicaciones de baja latencia. Las bases de datos de búsqueda NoSQL están diseñadas para hacer análisis sobre datos semiestructurados
Modelo de datos	El modelo relacional normaliza los datos en tablas conformadas por filas y columnas. Un esquema define estrictamente las tablas, las filas, las columnas, los índices, las relaciones entre las tablas y otros elementos de las bases de datos. La base de datos impone la integridad referencial en las relaciones entre tablas	Las bases de datos NoSQL proporcionan una variedad de modelos de datos, como clave-valor, documentos, gráficos y columnas, que están optimizados para el rendimiento y la escala
Propiedades ACID (Atomicidad, coherencia, aislamiento y durabilidad)	Las bases de datos relacionales ofrecen propiedades ACID: <ul style="list-style-type: none"> - La atomicidad requiere que una transacción se ejecute por completo o no se ejecute en absoluto. - La consistencia requiere que los datos deban acoplarse al esquema de la base de datos una vez confirmada una transacción. - El aislamiento requiere que las transacciones simultáneas se ejecuten por separado. - La durabilidad requiere la capacidad de recuperarse de un error inesperado del sistema o de un corte de 	La mayoría de bases de datos NoSQL hace concesiones al flexibilizar algunas de las propiedades ACID de las bases de datos relacionales a favor de un modelo de datos más flexible que puede escalar horizontalmente. Esto hace que las bases de datos NoSQL sean una excelente opción para casos de uso de baja latencia y alto rendimiento que necesitan escalar horizontalmente más allá de las limitaciones de una sola instancia

	energía y volver al último estado conocido.	
Rendimiento	Normalmente, el rendimiento depende del subsistema de disco. Se necesita la optimización de consultas, índices y estructura de tabla para lograr el máximo rendimiento	El rendimiento, por lo general, depende del tamaño del clúster de hardware subyacente, la latencia de red y la aplicación que efectúa la llamada
Escalado	Las bases de datos relacionales generalmente escalan verticalmente las capacidades de computación del hardware o la ampliación mediante la adición de réplicas para cargas de trabajo de solo lectura	Las bases de datos NoSQL suelen ser particionables. Esto se debe a que los patrones de acceso pueden escalar horizontalmente mediante el uso de arquitectura distribuida para aumentar el rendimiento que proporciona un rendimiento constante a una escala casi ilimitada
API	Solicita almacenar y recuperar datos que están comunicados mediante consultas que se ajustan a un lenguaje de consulta estructurado (SQL). Estas consultas son analizadas y ejecutadas por la base de datos relacional	Las API basadas en objetos permiten a los desarrolladores almacenar y recuperar fácilmente estructuras de datos. Las claves de partición permiten que las aplicaciones busquen pares de clave-valor, conjuntos de columnas o documentos semiestructurados que contengan atributos y objetos de aplicación serializados

¿CUÁNDO ELEGIR BASES DE DATOS NOSQL EN LUGAR DE BASES DE DATOS SQL?

Una base de datos NoSQL es la mejor opción para gestionar datos indeterminados, no relacionados o que cambian rápidamente. Su uso es intuitivo para los desarrolladores cuando la aplicación dicta el esquema de la base de datos. Puede usarlas para aplicaciones que cumplan las siguientes condiciones:

- Necesidad de esquemas flexibles que permitan un desarrollo más rápido e iterativo.
- Prioridad en el rendimiento por encima de una sólida coherencia de los datos y para mantener las relaciones entre las tablas de datos (integridad referencial).
- Exigencia de un escalado horizontal mediante la fragmentación de los servidores.
- Admisión de datos semiestructurados y sin estructurar.

No siempre tiene que elegir entre un esquema de base de datos relacional y no relacional. Puede emplear una combinación de bases de datos SQL y NoSQL en sus aplicaciones. Este enfoque híbrido es bastante común y garantiza que cada carga de trabajo se asigne a la base de datos correcta para obtener una relación precio-rendimiento óptima.

AMAZON DYNAMODB

Amazon DynamoDB es un servicio de base de datos NoSQL sin servidor que le permite desarrollar aplicaciones modernas a cualquier escala. Al ser una base de datos sin servidor, solo paga por lo que usa; no tiene arranques en frío (proceso de iniciar la base de datos desde un estado completamente apagado), actualizaciones de versión, periodos de mantenimiento, parches, ni tiempo de inactividad debido a tareas de mantenimiento. DynamoDB ofrece un amplio conjunto de controles de seguridad y estándares de cumplimiento.

Para las aplicaciones distribuidas a nivel mundial, las tablas globales de DynamoDB son una base de datos multirregional y multiactiva con un SLA de disponibilidad del 99,999 % y una mayor resiliencia. La fiabilidad de DynamoDB es compatible con la realización de copias de seguridad administradas, recuperación a un momento dado y más.

Amazon DynamoDB es una base de datos NoSQL, de clave-valor, totalmente administrada y sin servidor, con un rendimiento de milisegundos de un solo dígito a cualquier escala. Al ser sin servidor o serverless, no es necesario administrar las instancias subyacentes ni la infraestructura que la alimenta.

Con DynamoDB, se crean tablas. Una tabla de DynamoDB es solo un lugar donde puede almacenar y consultar datos. Los datos se organizan en elementos y los elementos tienen atributos. Los atributos son solo funciones diferentes de los datos. Si tiene un elemento o 2 millones de elementos en la tabla, DynamoDB administra el almacenamiento subyacente por usted y no debe preocuparse por el escalado del sistema, ya sea vertical u horizontal.

DynamoDB almacena los datos de forma redundante en las zonas de disponibilidad y refleja los datos en múltiples unidades para usted. Esto hace que la carga de operar una base de datos de alta disponibilidad sea mucho menor.

DynamoDB, además de ser escalable de forma masiva, también tiene un alto rendimiento y un tiempo de respuesta de milisegundos. Cuando tiene aplicaciones con potencialmente millones de usuarios, es de suma importancia contar con escalabilidad y tiempos de respuesta fiables y ultrarrápidos.

DynamoDB es una base de datos no relacional, por lo tanto, tiene esquemas flexibles simples y no esquemas rígidos complejos, que presentan varias tablas que se relacionan entre sí. Con DynamoDB puede agregar y quitar atributos de los elementos de la tabla, y en cualquier momento, no todo el elemento de la tabla tiene que tener los mismos atributos. Esto es ideal para conjuntos de datos que tienen alguna variación de un elemento a otro.

Debido a esta flexibilidad, no se pueden ejecutar consultas SQL complejas en él, en su lugar, escribiría consultas basadas en un pequeño conjunto de atributos que se designan como claves. Por lo tanto, las consultas que se ejecutan tienden a ser más sencillas y se centran en una colección de elementos de una tabla y no en consultas que abarcan varias tablas. Este patrón de consulta, junto con otros factores, incluyendo la forma que se diseña el sistema subyacente, le permite a DynamoDB ser muy rápido en el tiempo de respuesta y altamente escalable.

CARACTERÍSTICAS DE AMAZON DYNAMODB

Amazon DynamoDB es un servicio de base de datos NoSQL sin servidor que admite modelos de datos de clave-valor y de documentos. Los desarrolladores pueden usar Amazon DynamoDB para crear aplicaciones modernas y sin servidor que pueden ejecutarse con poco y escalar globalmente. Amazon DynamoDB se escala para admitir tablas de prácticamente cualquier tamaño con escalado horizontal automatizado.

La disponibilidad, durabilidad y tolerancia a errores están integradas y no se pueden deshabilitar, por lo que no es necesario estructurar sus aplicaciones para estas capacidades.

Amazon DynamoDB se ha diseñado para ejecutar aplicaciones de alto rendimiento a escala de Internet que sobrecargarían las bases de datos relacionales tradicionales. Con más de diez años de inversión pionera en innovación, Amazon DynamoDB ofrece una escalabilidad ilimitada con un rendimiento uniforme de milisegundos de un solo dígito y una disponibilidad de hasta el 99,999 %.

A continuación, veamos con más detalle las características de Amazon DynamoDB:

Sin servidor: Con Amazon DynamoDB no hay servidores que aprovisionar, parchar o administrar, y no hay software que instalar, mantener o utilizar. Amazon DynamoDB no tiene versiones (principal, secundaria ni parche), no hay ventanas de mantenimiento y ofrece un mantenimiento sin tiempo de inactividad. Los precios bajo demanda de Amazon DynamoDB ofrecen precios de pago por uso, se escalan a cero y escalan automáticamente las tablas para ajustar la capacidad y mantener el rendimiento sin necesidad de administración.

Escalado automático: A medida que disminuye o aumenta el tamaño de la base de datos, DynamoDB escala de forma automática para ajustarse a los cambios en la capacidad mientras mantiene un rendimiento uniforme.

Esto lo convierte en una opción adecuada para los casos prácticos que requieren alto rendimiento durante el escalado.

Seguridad y fiabilidad: Amazon DynamoDB ayuda a proteger los datos mediante el cifrado y realiza copias de seguridad continuas de los datos para protegerlos.

Ofrece cifrado en reposo y en tránsito, así como copias de seguridad continuas para proteger los datos. Además, es altamente disponible y duradero.

Rentabilidad: DynamoDB ofrece precios de pago por uso y se escala automáticamente según la demanda. Puedes optimizar los costos seleccionando el modo de capacidad adecuado y la clase de tabla correcta. Es una solución rentable para empresas de todos los tamaños.

VENTAJAS DE USAR AMAZON DYNAMODB

Algunas de las ventajas de usar Amazon DynamoDB son:

Rendimiento a escala: Ofrezca aplicaciones con un rendimiento y almacenamiento coherentes, de alto rendimiento y prácticamente ilimitados.

SLA del 99.999%: Obtenga un rendimiento rápido de lecturas y escrituras locales con la base de datos de tablas globales en varias regiones.

Escalable a cero sin servidores: Optimice los costos con una base de datos sin servidor totalmente administrada que se escala automáticamente para adaptarse a sus necesidades.

Integración sencilla con AWS: Aproveche más sus datos con herramientas integradas para realizar análisis, extraer información y mucho más.

En síntesis, Amazon DynamoDB es un servicio completamente administrado de base de datos NoSQL sin servidor con tiempos de respuesta de milisegundos de un solo dígito a cualquier escala, lo que le permite desarrollar y ejecutar aplicaciones modernas pagando solo por lo que utiliza.

COMPARACIÓN ENTRE AMAZON RDS Y AMAZON DYNAMODB

En la búsqueda base de datos, tenemos en el lado relacional a Amazon RDS diseñado para eliminar el trabajo pesado indiferenciado de los administradores de bases de datos, con alta disponibilidad y recuperación automática proporcionadas, donde usted controla los datos, el esquema y la red.

En el lado no relacional está Amazon DynamoDB, usando un par de valor de clave que no requiere ningún esquema avanzado, capaz de funcionar una base de datos global en solo pulsar un botón, con un rendimiento masivo. Tiene potencial de escala de petabytes y acceso API granular.

Cada clase de base de datos está diseñada para mejorar los casos de uso que usted se puede imaginar. ¿Qué base de datos es mejor? Depende de su caso práctico.

Aquí tenemos una comparación entre ambos:

Tipo de Base de Datos: Amazon RDS: Diseñado para bases de datos relacionales (SQL).
DynamoDB: Orientado a bases de datos NoSQL.

Escalabilidad: Amazon RDS: Escalable, pero la capacidad de almacenamiento varía según el motor de base de datos utilizado (por ejemplo, Aurora, MySQL, PostgreSQL).
DynamoDB: Altamente escalable, admite tablas de cualquier tamaño.

Rendimiento: Amazon RDS: Ofrece opciones de almacenamiento SSD con diferentes niveles de rendimiento (hasta 40,000 IOPS por instancia de base de datos).
DynamoDB: Alto rendimiento, con baja latencia y respuesta rápida debido al uso de SSD y caché en memoria (DynamoDB Accelerator).

Costo: Amazon RDS: Más caro que DynamoDB, pero ofrece más funciones y flexibilidad.
DynamoDB: Costo efectivo y se factura según la capacidad de lectura/escritura y el almacenamiento.

En conclusión, si necesitamos una base de datos fácil de configurar, operar y escalar, que admita consultas complejas y transacciones, Amazon RDS es la mejor opción. Si buscamos alto rendimiento, escalabilidad sin problemas y un enfoque costo-efectivo, Amazon DynamoDB es la elección adecuada.

AMAZON DOCUMENTDB

Amazon DocumentDB (con compatibilidad con MongoDB) es una base de datos de documentos creada específicamente para la administración de datos JSON a escala, completamente administrada que hace que sea fácil y rentable operar con cargas de trabajo de documentos clave, sin tener que administrar infraestructura. Amazon DocumentDB simplifica su arquitectura al brindar prácticas recomendadas de seguridad integradas, copias de seguridad continuas e integraciones nativas con otros servicios de AWS. Puede mejorar sus aplicaciones con capacidades de inteligencia artificial (IA) generativa y machine learning (ML) mediante la búsqueda vectorial para Amazon DocumentDB y la integración con Amazon SageMaker Canvas.

En Amazon DocumentDB, el almacenamiento escala automáticamente hasta los 128 TiB en clústeres basados en instancias y 4 PiB en clústeres elásticos de Amazon DocumentDB, con un impacto escaso o inexistente en su aplicación. Amazon DocumentDB admite millones de solicitudes por segundo con hasta 15 réplicas de lectura de baja latencia en cuestión de minutos, sin ningún tiempo de inactividad de la aplicación, con independencia del tamaño de los datos.

Amazon DocumentDB ofrece un SLA del 99,9 % y hace que sus datos sean duraderos en tres zonas de disponibilidad (AZ) de una región al replicar las nuevas escrituras de seis formas para garantizar que los datos sigan siendo legibles en el caso poco común de que se produzca un error total de AZ más un error simultáneo adicional en un nodo de almacenamiento en una AZ diferente. Al replicar las nuevas escrituras de seis maneras, Amazon DocumentDB es resiliente a los errores y garantiza que las conmutaciones por error no tengan pérdida de datos dentro de una región. Los clientes solo pagan por una copia del almacenamiento.

Los clientes pueden usar AWS Database Migration Service (DMS) para migrar con facilidad sus bases de datos de MongoDB autoadministradas a Amazon DocumentDB prácticamente sin tiempo de inactividad.

CARACTERÍSTICAS DE AMAZON DOCUMENTDB

Rendimiento a escala: Clústeres elásticos de Amazon DocumentDB permite a los clientes administrar millones de escrituras y lecturas por segundo, de modo que pueden escalar sus bases de datos de documentos en minutos con un periodo de inactividad mínimo o nulo y sin perjudicar el rendimiento.

Amazon DocumentDB utiliza una arquitectura optimizada en memoria y de escala ajustable que permite la evaluación rápida de consultas en conjuntos de documentos de gran tamaño.

Con unos pocos clics en la consola de administración de AWS, los clientes pueden escalar o reducir los recursos de computación y de memoria mediante la creación de instancias de réplicas nuevas del tamaño que desee o a través de la eliminación de instancias.

Amazon DocumentDB aumentará automáticamente el tamaño del volumen de almacenamiento a medida que se incrementen las necesidades del clúster en relación con el almacenamiento. No es necesario que los clientes aprovisionen un almacenamiento excesivo para la base de datos de documentos con el objetivo de afrontar el crecimiento futuro.

Aumente el rendimiento de lectura para admitir solicitudes de aplicaciones de volumen alto mediante la creación de hasta 15 réplicas de lectura de bases de datos.

Con compatibilidad con MongoDB: Amazon DocumentDB es compatible con las herramientas y los controladores de MongoDB 3.6, 4.0 y 5.0. En Amazon DocumentDB, los clientes pueden usar la gran mayoría de las aplicaciones, controladores y herramientas que utilizan con sus bases de datos no relacionales de MongoDB de código abierto. En Amazon DocumentDB se simulan las respuestas que espera un cliente de un servidor de MongoDB mediante la implementación de las API de MongoDB 3.6, 4.0 y 5.0 de código abierto de Apache 2.0 en un sistema de almacenamiento personalizado, distribuido, tolerante a errores y con recuperación automática, por medio del cual se proporciona a los clientes el nivel de rendimiento, escalabilidad y disponibilidad necesario para utilizar cargas de trabajo de MongoDB esenciales a escala.

Aquí, vamos a hacer un paréntesis para hablar de MongoDB:

(

MongoDB es una base de datos de documentos que ofrece gran escalabilidad y flexibilidad. A diferencia de las bases de datos relacionales tradicionales, MongoDB utiliza documentos flexibles en lugar de tablas y filas para procesar y almacenar datos. Algunas características clave de MongoDB son:

Modelo de documentos: MongoDB almacena datos en documentos similares a JSON, lo que

permite que los campos varíen entre documentos y que la estructura de datos cambie con el tiempo. Esto facilita el trabajo con datos en aplicaciones.

Consultas e indexación avanzada: MongoDB ofrece consultas ad hoc, indexación y agregación en tiempo real, lo que permite un acceso potente a los datos y su análisis.

Escalabilidad y alta disponibilidad: MongoDB es una base de datos distribuida en su núcleo, lo que significa que la escalabilidad horizontal, la alta disponibilidad y la distribución geográfica están integradas y son fáciles de usar.

Fácil desarrollo: El modelo de documentos de MongoDB es fácil de aprender y usar, y proporciona a los desarrolladores todas las funcionalidades necesarias para satisfacer requisitos complejos a cualquier escala.

En resumen, MongoDB es una herramienta poderosa para almacenar y gestionar grandes cantidades de datos estructurados y no estructurados. Es ampliamente utilizado en una variedad de aplicaciones.

)

Rentabilidad: No hay ningún compromiso inicial con Amazon DocumentDB, usted paga un cargo por hora por cada instancia que lanza y, cuando termine de usar una instancia de Amazon DocumentDB, puede eliminarla o pausarla. No es necesario sobre aprovisionar el almacenamiento como margen de seguridad y solo se pagará por el almacenamiento que realmente consuma.

Amazon DocumentDB ofrece una configuración de almacenamiento optimizada para E/S. Amazon DocumentDB optimizado para E/S ofrece un ahorro de costos de hasta un 40 % para aplicaciones con uso intensivo de E/S en las que los cargos relacionados superan el 25% del gasto total en bases de datos de Amazon DocumentDB.

Amazon DocumentDB le permite elegir entre configuraciones de almacenamiento estándar y optimizadas para E/S para su clúster de base de datos. La flexibilidad le permite maximizar la relación precio-rendimiento al elegir la configuración adecuada en función de sus necesidades

Completamente administrado: Comenzar a usar Amazon DocumentDB es sencillo, tan solo debe lanzar un clúster de Amazon DocumentDB con la consola de administración de AWS. Las instancias de Amazon DocumentDB están preconfiguradas con los parámetros y configuraciones correspondientes a la clase de instancia que seleccionó. Los clientes pueden iniciar el clúster y conectar la aplicación en cuestión de minutos sin realizar configuraciones adicionales.

Con Amazon DocumentDB, se mantendrá actualizada la base de datos de los clientes con los parches más recientes. Los clientes pueden decidir si quieren implementar revisiones en el clúster y cuándo pueden hacerlo a través de la administración de versiones del motor de la base de datos.

Altos niveles de seguridad y cumplimiento: Amazon DocumentDB se ejecuta en la nube virtual privada de Amazon (VPC), lo que permite a los clientes aislar el clúster en la red virtual y conectarse a la infraestructura de TI local con redes privadas virtuales (VPN) con IPsec cifradas estándares de la industria. Además, con la configuración de la VPC de Amazon DocumentDB, los clientes pueden configurar los parámetros del firewall y controlar el acceso de red al clúster.

Amazon DocumentDB es compatible con el control del acceso basado en roles (RBAC) que incluye roles integrados y roles definidos. El RBAC permite a los clientes aplicar privilegios mínimos como práctica recomendada, ya que restringe las acciones que los usuarios están autorizados a llevar a cabo. Amazon DocumentDB se integra con AWS Identity and Access Management (IAM) y ofrece a los clientes la posibilidad de controlar las acciones que los usuarios y grupos de AWS IAM pueden llevar a cabo en recursos específicos de Amazon DocumentDB, incluidos clústeres, instancias, instantáneas y grupos de parámetro.

Amazon DocumentDB permite a los clientes cifrar sus bases de datos mediante las claves que cree y controle a través de AWS Key Management Service (KMS). En un clúster que se ejecuta con el cifrado de Amazon DocumentDB, los datos almacenados en reposo en el almacenamiento subyacente están cifrados, al igual que las copias de seguridad, las instantáneas y las réplicas automatizadas que se encuentran en el mismo clúster. De manera predeterminada, las conexiones entre un cliente y Amazon DocumentDB están cifradas en tránsito con TLS.

Amazon DocumentDB se diseñó para cumplir con los estándares de seguridad más exigentes y para facilitar a los clientes la verificación de la seguridad y el cumplimiento de sus propias obligaciones normativas y de cumplimiento.

Alta disponibilidad: Los clústeres globales de Amazon DocumentDB permiten la recuperación de desastres en caso de interrupciones de servicio en toda una región y posibilitan las lecturas globales de baja latencia.

El estado de las instancias y el clúster de su Amazon DocumentDB se supervisa continuamente, si ocurre un error en la instancia que contiene la base de datos, la instancia y los procesos asociados se reinician automáticamente. La recuperación de Amazon DocumentDB no necesita el potencialmente prolongado reinicio de los registros de rehacer de la base de datos, por lo que los plazos de reinicio de la instancia son, normalmente, de 30 segundos o menos.

En caso de que ocurra un error en una instancia, Amazon DocumentDB automatiza la conmutación por error a una de las hasta 15 réplicas de Amazon DocumentDB que los clientes hayan creado en cualquiera de las tres zonas de disponibilidad. Si no se aprovisionó ninguna réplica de Amazon DocumentDB y se produce un error, Amazon DocumentDB intentará crear automáticamente una nueva instancia para los clientes.

Cada sección de 10 GB del volumen de almacenamiento se replica de seis formas en tres zonas de disponibilidad (AZ). Amazon DocumentDB utiliza almacenamiento tolerante a errores que administra de manera transparente la pérdida de hasta dos copias de datos sin que ello afecte la disponibilidad de escritura de la base de datos y hasta tres copias sin que incida en la disponibilidad de lectura.

IA generativa y machine learning: Amazon DocumentDB ofrece funciones que permiten que los modelos de machine learning (ML) e inteligencia artificial (IA) generativa funcionen con los datos almacenados en Amazon DocumentDB en tiempo real. Los clientes ya no tendrán que perder tiempo administrando una infraestructura independiente, escribiendo código para conectarse con otro servicio y duplicando datos de su base de datos principal. Con la búsqueda vectorial para Amazon DocumentDB, puede almacenar, indexar y buscar millones de vectores con tiempos de respuesta de milisegundos (un vector es una representación numérica que representa el significado semántico de datos no estructurados, como texto, imágenes y video).

En resumen, Amazon DocumentDB es un servicio de base de datos de documentos completamente administrado creado por AWS. Está diseñado especialmente para trabajar con aplicaciones basadas en documentos JSON. En esencia, Amazon DocumentDB es compatible con MongoDB, lo que significa que puede operar con cargas de trabajo MongoDB sin tener que administrar la infraestructura subyacente. Es una excelente opción para almacenar, consultar e indexar datos en formato JSON flexible, y escalar a millones de solicitudes de lectura y escritura de documentos por segundo con petabytes de almacenamiento.

COMPARACIÓN ENTRE AMAZON DYNAMODB Y AMAZON DOCUMENTDB

Amazon DynamoDB y Amazon DocumentDB son dos populares soluciones de bases de datos NoSQL basadas en la nube de Amazon. A continuación, haremos una comparación entre ambas:

1. Modelo de Datos:

DynamoDB: Admite modelos de datos clave-valor y documentos. Es completamente administrado y puede escalar horizontalmente, lo que lo hace altamente escalable.

DocumentDB: Diseñado para manejar modelos de datos JSON. Es compatible con MongoDB y ofrece una replicación automática de datos para alta disponibilidad.

2. Latencia y Escalabilidad:

DynamoDB: Ofrece latencia de milisegundos de un solo dígito y puede escalar para manejar más de 20 millones de solicitudes por segundo.

DocumentDB: También ofrece baja latencia y alta escalabilidad, pero su límite de almacenamiento es menor que el de DynamoDB.

3. Uso Ideal:

DynamoDB: Perfecto para aplicaciones con alta demanda de rendimiento y baja latencia, como juegos, tecnología publicitaria y aplicaciones IoT.

DocumentDB: Adecuado para cargas de trabajo críticas de MongoDB que requieren consultas complejas y análisis.

Es decir, si necesitamos un almacén de claves-valor con alta escalabilidad, DynamoDB es la elección. Si trabajamos con datos JSON y necesitamos una base de datos compatible con MongoDB, DocumentDB es la opción adecuada.

AMAZON NEPTUNE

Amazon Neptune es un servicio de bases de datos de gráficos sin servidor, rápido, fiable y completamente administrado por Amazon Web Services (AWS). Está diseñado para almacenar y consultar grandes conjuntos de datos relacionales o de gráficos altamente conectados. Neptune permite crear y ejecutar fácilmente aplicaciones gráficas interactivas que trabajan con datos altamente conectados, como motores de recomendaciones, detección de fraudes, gráficos de conocimiento, descubrimiento de fármacos y seguridad de red.

El componente principal de Neptune es un motor de base de datos de gráficos de alto rendimiento y personalizado. Está diseñado para una escalabilidad y disponibilidad superiores. Proporciona seguridad integrada, copias de seguridad continuas e integraciones con otros servicios de AWS. Además, Neptune ofrece replicación de datos entre regiones para lecturas y escrituras de baja latencia, recuperación de desastres y escalabilidad para aplicaciones distribuidas a nivel mundial.

Las consultas SQL para datos fuertemente conectados son complejas y difíciles de ajustar para lograr un buen nivel de rendimiento. En cambio, Amazon Neptune permite utilizar los lenguajes populares de consulta de grafos Apache TinkerPop Gremlin, SPARQL y openCypher de W3C para ejecutar consultas eficientes que son fáciles de escribir y tienen un buen rendimiento en datos conectados. Esta característica reduce de manera significativa la complejidad del código y permite crear aplicaciones que procesen relaciones con mayor rapidez. Podemos iniciar rápidamente una instancia de base de datos de Neptune con unos pocos pasos en la consola de Neptune.

CARACTERÍSTICAS DE AMAZON NEPTUNE

Algunas características clave de Amazon Neptune son:

Alto rendimiento y escalabilidad: Neptune es una base de datos de gráficos de alto rendimiento. Almacena y examina los datos de gráficos de manera eficiente, y usa una arquitectura optimizada en memoria y de escala ajustable que permite la evaluación rápida de consultas de gran tamaño.

Con unos pocos pasos en la consola de administración de AWS, puede aumentar o reducir los recursos informáticos y de memoria del clúster. Con Neptune puede escalar creando nuevas instancias de réplica del tamaño deseado o eliminar instancias. Las operaciones de escalamiento de computación normalmente se llevan a cabo en cuestión de minutos.

Neptune usa una arquitectura de almacenamiento distribuido y compartido que crecerá automáticamente a medida que aumenten las necesidades de almacenamiento de datos.

Con Neptune puede aumentar el rendimiento de lectura para admitir solicitudes de aplicaciones de gran volumen mediante la creación de hasta 15 réplicas de lectura de bases de datos.

Alta disponibilidad y durabilidad: La salud de su base de datos de Neptune y su instancia de EC2 subyacente se supervisa continuamente. Si se produce un error en la instancia que respalda a la base de datos, se reinicia la base de datos, así como los procesos asociados de forma automática.

En caso de error de instancia, Neptune automatiza la conmutación por error a una de hasta 15 réplicas de Neptune que haya creado en cualquiera de las tres zonas de disponibilidad. Si aún no se aprovisiona ninguna réplica de Neptune y se produce un error, Neptune intentará crear automáticamente una nueva instancia de base de datos.

Cada porción de 10 GiB del volumen de la base de datos es duradera en tres zonas de disponibilidad. Neptune utiliza un almacenamiento tolerante a errores que gestiona de forma transparente la pérdida de hasta dos copias de datos sin afectar a la disponibilidad de escritura de la base de datos y hasta tres copias sin afectar a la disponibilidad de lectura.

Alto nivel de seguridad: Neptune se ejecuta en Amazon Virtual Private Cloud (Amazon VPC), lo que le permite aislar la base de datos en su propia red virtual y conectarse a la infraestructura de TI local en las instalaciones con VPN con IPsec cifradas estándar de la industria. Además, al usar la configuración de Neptune VPC, puede configurar los ajustes del firewall y controlar el acceso de red a las instancias de la base de datos.

Está integrado con AWS Identity and Access Management (IAM) y le permite controlar las acciones que los usuarios y grupos de IAM pueden realizar en recursos específicos de Neptune, como las instancias de bases de datos, las instantáneas de bases de datos, los grupos de parámetros de bases de datos, las suscripciones a eventos de bases de datos y los grupos de opciones de bases de datos.

Neptune admite el cifrado en tránsito con TLS versión 1.2. Permite cifrar las bases de datos mediante las claves que crea y controla a través de AWS Key Management Service (AWS KMS). En una instancia de base de datos que se ejecuta con el cifrado de Neptune, los datos almacenados en reposo en el almacenamiento subyacente están cifrados, al igual que las copias de seguridad, las instantáneas y las réplicas automatizadas que se encuentran en el mismo clúster.

Administración total: Puede comenzar a utilizar Neptune lanzando una nueva instancia de base de datos de Neptune mediante la consola de administración de AWS. Las instancias de base de datos de Neptune vienen configuradas previamente con parámetros y ajustes adecuados para la clase de instancia de base de datos que ha seleccionado. Puede lanzar una instancia de base de datos y conectar su aplicación en cuestión de minutos sin configuraciones adicionales.

Neptune mantendrá su base de datos actualizada con los parches más recientes.

Puede clonar una base de datos de Neptune con solo unos pasos en la consola de administración de AWS, sin afectar al entorno de producción. El clon se distribuye y replica en tres zonas de disponibilidad.

ML e IA generativa: Amazon Neptune ML funciona con SageMaker, que utiliza GNN (redes neuronales gráficas), una técnica de ML diseñada específicamente para grafos, para realizar predicciones rápidas y más precisas con datos grafos. Con Neptune ML, puede mejorar la precisión de la mayoría de las predicciones de grafos en más de un 50 % en comparación con las predicciones hechas con métodos no grafos.

La búsqueda vectorial en lugar de datos gráficos le brinda un costo total de propiedad más bajo en general y una sobrecarga de administración más simple, ya que no necesita administrar almacenes de datos separados, crear canalizaciones ni preocuparse por mantener los almacenes de datos sincronizados.

Cumplimiento y conformidad: Neptune está dentro del alcance de más de 20 estándares internacionales de conformidad que van desde FedRAMP (Moderado y Alto) hasta SOC (1,2,3), y también es elegible para HIPAA.

Rentabilidad: Con Neptune no hay ningún compromiso por adelantado, se paga una tarifa por hora por cada instancia que se lanza o por los recursos de la base de datos que se consumen para el uso sin servidor. Cuando termine con una instancia de base de datos de Neptune, puede eliminarla. No es necesario sobre aprovisionar almacenamiento como margen de seguridad y solo se paga por el almacenamiento que realmente se consume.

En síntesis, Neptune utiliza un motor de base de datos de gráficos de alto rendimiento optimizado para almacenar miles de millones de relaciones y consultar el gráfico con baja latencia. Es compatible con lenguajes de consulta de gráficos como Apache TinkerPop, Gremlin, openCypher y SPARQL.

Además, ofrece alta disponibilidad con réplicas de lectura, recuperación a un momento dado y copias de seguridad continuas en Amazon S3. Además, admite cifrado en reposo y en tránsito.

Amazon Neptune es una solución poderosa para trabajar con datos de gráficos altamente conectados y ofrece una experiencia completamente administrada para los desarrolladores y administradores de bases de datos.