PATROCINADO POR

IBM

GUÍA PARA GEEKS

Por qué los desarrolladores de aplicaciones innovadoras adoran los OSDBMS de alta velocidad

LINUX™ JOURNAL

Contenido

Ted Schmidt es un consultor especializado en soluciones de marketing y comercio electrónico para la industria de la fabricación. Trabaja en la gestión de proyectos y productos desde antes del comienzo del movimiento ágil en 2001, y durante más de 20 años, se ha dedicado a gestionar el desarrollo de proyectos y productos para fabricantes de bienes de consumo, dispositivos médicos, electrónica y telecomunicaciones. Cuando no está dedicado al desarrollo de productos, escribe novelas y dirige un pequeño estudio de diseño gráfico en http://FloatingOrange.com. Ted ha dado charlas en conferencias en PMI y escribe en el blog http://FloatingOrangeDesign.Tumblr.com y en su sitio web en http://FloatingOrange.com.

GUÍAS PARA GEEKS:

Información esencial para las personas más técnicas del planeta

Acerca del patrocinador

IBM

IBM es una empresa global de consultoría y tecnología integrada cuya sede central está en Armonk, Nueva York. Con operaciones en más de 170 países, IBM atrae y retiene a algunas de las personas más talentosas del mundo para ayudar a resolver problemas y proporcionar ventajas a las empresas, los gobiernos y las organizaciones sin fines de lucro.

La innovación es el eje de la estrategia de IBM. La empresa se ha reinventado a sí misma a lo largo de numerosas eras tecnológicas y ciclos económicos, y ha creado un valor diferenciador para sus clientes. Hoy en día, a medida que la industria informática cambia radicalmente a un ritmo sin precedentes, IBM es mucho más que una empresa de "*hardware*, *software* y servicios". IBM emerge en la actualidad como una empresa de plataformas de nube y soluciones cognitivas.

Las soluciones cognitivas con tecnología de nube son la clave para la transformación digital de los clientes. Esta transformación exige innovaciones en todos los niveles de la base tecnológica empresarial, desde los procesadores y el diseño de los equipos, hasta el almacenamiento, las redes y la capa de integración. Los sistemas IBM Power Systems, creados con tecnologías abiertas y pensados para las aplicaciones esenciales, ofrecen la infraestructura que está diseñada para las cargas de trabajo cognitivas.

Por qué los desarrolladores de aplicaciones innovadoras adoran los OSDBMS de alta velocidad

TED SCHMIDT

Introducción

Como cualquier desarrollador de aplicaciones sociales, móviles o de Internet de las Cosas (IoT) puede dar fe, los modelos heredados de bases de datos relacionales ya no satisfacen todas nuestras necesidades. No es que haya algún problema con los sistemas tradicionales de gestión de bases de datos, simplemente no fueron diseñaron para abordar la variedad y el volumen de datos que invaden el mundo digital actual; por no mencionar la demanda de velocidad necesaria para procesar tantos datos con el fin de ofrecer la funcionalidad y las características útiles basadas en datos que cada vez más esperamos de prácticamente todo. Por suerte, todo un nuevo mundo de sistemas de gestión de bases de datos de código abierto (OSDBMS) se ha creado precisamente para gestionar la diversidad y la complejidad de los datos actuales, y para poder almacenarlos, analizarlos y trabajar con ellos a la

velocidad necesaria para que sean más valiosos que nunca.

En esta guía, analizaré algunos de los OSDBMS disponibles y las soluciones que ofrecen a los problemas que enfrentamos en el desarrollo de aplicaciones innovadoras para los datos nuevos de fuentes sociales, móviles y de IoT. Si bien el análisis de todos los creadores de OSDBMS no está dentro del alcance de esta guía, incluyo un análisis equilibrado de los más destacados en cada categoría principal, como SQL y NoSQL (incluidos gráficos, documentos y almacenes clave-valor) de código abierto, e incluso los productos en memoria con aceleración por GPU disponibles en la actualidad.

Comenzaré analizando algunos desafíos clave del escenario de las aplicaciones nuevas, lo que incluye los *Big Data* que estas aplicaciones crean y consumen. A continuación, describiré el escenario de los OSDBMS de forma general y luego profundizaré en las principales ofertas de OSDBMS. Finalmente, concluiré con un análisis de la mejor plataforma tecnológica disponible para que estos modernos sistemas de gestión de bases de datos (DBMS) funcionen en el máximo nivel, superando desafíos y batiendo nuevos récords de velocidad, rendimiento y escala para satisfacer las necesidades de las aplicaciones más innovadoras de la actualidad.

Los desafíos de los desarrolladores de aplicaciones innovadoras

Seguimos expectantes ante los desafíos asociados con "*Big Data*" y su evolución. Si bien es posible que *Big Data* tenga diferentes requisitos para diferentes organizaciones e industrias, cuando hablo de *Big Data* aquí, me refiero a ellos específicamente en términos de los problemas al diseñar y desarrollar aplicaciones innovadoras en un mundo que exige acceso instantáneo a un gran volumen de

FIGURA 1. Las cuatro dimensiones de Big Data

datos. Considero las fuentes externas de *Big Data* y los problemas asociados con la recopilación, el almacenamiento, el análisis y la visualización de estos datos. Para un desarrollador de aplicaciones enfocado en proporcionar soluciones innovadoras basadas en datos, las cuatro dimensiones de *Big Data* de interés son: tamaño, velocidad, complejidad y confiabilidad. En esta guía, me centro principalmente en los aspectos de velocidad y complejidad, y en el valor de las soluciones de OSDBMS y NoSQL para abordarlos.

La dimensión de velocidad, o "celeridad", se refiere a la rapidez con que los datos se crean, almacenan, procesan, recuperan y demás, y en última instancia, se utilizan para análisis. Cada vez más, las organizaciones demandan que los datos se procesen en tiempo real y se transmitan directamente a los procesos de toma de decisiones de una empresa. Consideremos los sistemas de control de tráfico. Si bien las demandas de velocidad continuarán creciendo, nuestra capacidad para seguir el ritmo de esa demanda se encuentra obstaculizada por algunas cuestiones.

Obviamente, la latencia (la diferencia entre cuándo se recopilan los datos y cuándo están disponibles para su uso) va a impactar en la velocidad. Además de la latencia, está el fantasma amenazante de la ley de Moore. La ley de Moore establece que el poder de procesamiento informático total se debería duplicar cada dos años. Si bien ha funcionado durante un tiempo, el sentido común ahora indica que existe un límite físico respecto de lo que los chips de silicona pueden ofrecer. Es una cuestión de física e impacta directamente en la capacidad de las aplicaciones de continuar ganando la velocidad necesaria para procesar la creciente cantidad de datos cada vez más complejos en tiempo real.

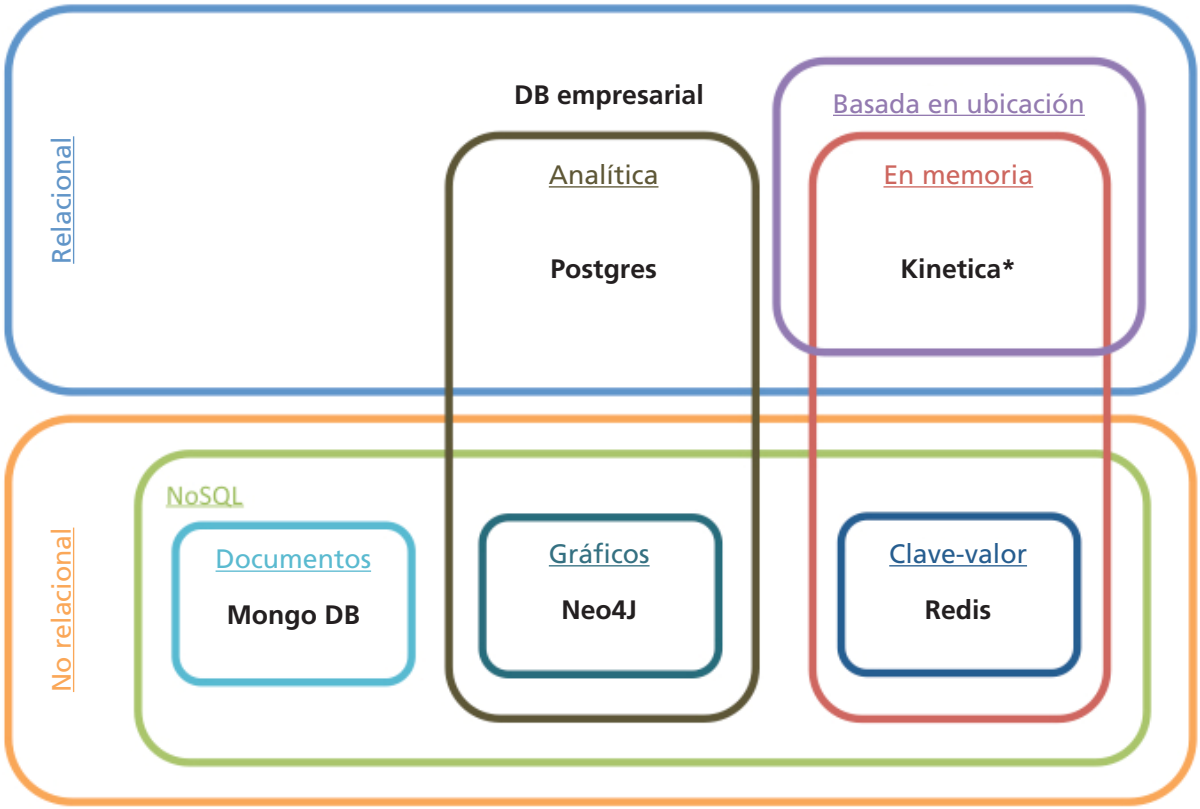
Todo esto significa que, como desarrolladores, hoy enfrentamos un desafío mayor, ya que no podemos simplemente confiar en la ley de Moore como la respuesta a las demandas de más velocidad. Debemos ser más innovadores, y eso significa considerar las ventajas de las soluciones de código abierto, en vez de seguir confiando en las bases de datos tradicionales, para gestionar los grandes volúmenes de datos con los que ahora trabajamos y a partir de los cuales intentamos ofrecer características, funcionalidad y conocimientos útiles lo antes posible.

Qué aporta el código abierto

Existen docenas de soluciones de DBMS de código abierto. Aquí, analizaré cinco que ofrecen ventajas específicas cuando se trata del desarrollo de nuevas aplicaciones innovadoras: MongoDB, Redis, Neo4j, PostgreSQL y Kinetica. (Revelación absoluta: si bien Kinetica no es una solución de código abierto, forma parte de un ecosistema de tecnología abierta a través de su membresía y participación en la Fundación OpenPOWER. Como tal, y gracias a sus ventajas particulares en el procesamiento en tiempo real de grandes

conjuntos de datos de *streaming*, la incluyo en este análisis. La Fundación OpenPOWER, a la cual me referiré más adelante en esta guía, es una organización de membresía técnica abierta que fundamentalmente busca responder a los límites de la ley de Moore permitiendo que las compañías miembros personalicen sus CPU POWER de formas nuevas e innovadoras para aprovechar al máximo toda la velocidad y potencia posibles. Recomiendo enfáticamente leer más información sobre la Fundación OpenPOWER en **https://openpowerfoundation.org**, pero basta señalar aquí que es donde se está produciendo la innovación).

Trasladarse a un OSDBMS tiene algunas ventajas tangibles, y a medida que los conjuntos de herramientas de gestión evolucionan hacia un nivel empresarial, las limitaciones anteriores desaparecen. La ventaja más evidente del código abierto en general es el bajo costo asociado con la inexistencia de *software* registrado, e incluso *hardware* registrado. Pero, como verá cuando analice los casos de uso específicos a los que cada una de las soluciones de OSDBMS se adapta mejor, la velocidad, la flexibilidad y la toma de decisiones inteligente son fundamentales. Eso no pretende desestimar la ventaja que aporta el código abierto a la velocidad de la evolución y la innovación. Las soluciones registradas por naturaleza evolucionan con mayor lentitud ante las cambiantes demandas del entorno. Sin embargo, las soluciones de código abierto aprovechan los aportes de una gran comunidad de muchas voces abocadas a resolver los problemas del mundo real. Así, las soluciones de código abierto evolucionan e innovan con mayor rapidez de lo que es capaz cualquier solución registrada.



*Si bien Kinetica no es código abierto, es parte de un ecosistema de tecnología abierta

FIGURA 2. El escenario de **OSDBMS**

Consideremos las dos categorías principales en el escenario de OSDBMS: no relacionales (que incluye MongoDB, Redis y Neo4j) y relacionales (que incluye EDB Postgres y Kinetica).

En el lado no relacional del escenario está Redis, una base de datos clave-valor NoSQL especialmente útil en la industria de los juegos (entre muchas otras), donde los datos de alta velocidad, simples pero sofisticados, son esenciales. Luego, Neo4j es una base de datos de gráficos NoSQL especialmente efectiva para almacenar las relaciones entre puntos de datos. Finalmente, MongoDB es una base de datos de documentos NoSQL, excelente como base de datos de uso general pero especialmente útil por no tener un esquema, lo que significa que puede almacenar todo tipo de datos diferentes con una gran flexibilidad.

En cuanto a la categoría relacional, si bien Kinetica no es una solución de código abierto, merece reconocimiento por la absoluta velocidad que aporta. Además de ser una base de datos relacional basada en la ubicación, Kinetica es una base de datos en memoria con aceleración por GPU. La clave de su velocidad sin precedentes para procesar grandes volúmenes de datos de *streaming* se encuentra en esa última parte sobre la aceleración por GPU (luego desarrollaré este tema). Por último, pero no por eso menos importante, la otra base de datos relacional que analizo aquí, EDB Postgres Advanced Server, es realmente la presentación de nivel empresarial de PostgreSQL de EDB, un producto de desarrollo de código abierto de una comunidad en la que participa activamente y que apoya religiosamente. También es excelente para la analítica.

En el lado no relacional del escenario se encuentra otra base de datos en memoria: Redis. Redis es una base de datos clave-valor NoSQL especialmente útil en la industria de los juegos. Neo4j es otra base datos NoSQL que, como base de datos de gráficos, es particularmente efectiva para almacenar las relaciones entre puntos de datos. Finalmente, MongoDB es también una base de datos NoSQL, pero de documentos, lo que significa que si bien es una buena base de datos de uso general, su real ventaja surge de la ausencia de un esquema. Sin un esquema, es posible almacenar toda clase de datos diferentes, como aumentos de temperatura o rotaciones por segundo.

Recuerde que aunque las ventajas de cada OSDBMS, que con frecuencia coinciden, en definitiva, se basan en el caso de uso específico, todas estas bases de datos comparten una ventaja similar: forman parte de un modelo de desarrollo de tecnología abierta. No solo eso, sino que en la infraestructura correcta, cada OSDBMS ofrece la tan preciada ventaja de la velocidad sin precedentes para una innovación real en el desarrollo de

aplicaciones de uso intensivo de datos.

DBMS de almacenamiento de documentos de MongoDB

MongoDB es una base de datos de código abierto con un modelo de datos orientado a los documentos. Almacena datos en el formato binario JSON (BSON), que amplía el formato JSON para incluir otros tipos de datos, como int, long, fecha, punto flotante y otros. Estos documentos BSON permiten que sea mucho más rápido y sencillo modelar los datos de la aplicación a los de la base de datos, porque los documentos BSON están alineados con la estructura de los objetos en el lenguaje de programación. Cada documento contiene múltiples campos, y cada campo contiene un valor de un tipo de dato específico, como subdocumentos o matrices. Los documentos que son similares en estructura se agrupan entonces en colecciones. En un sistema de gestión de bases de datos relacionales (RDBMS), las colecciones serían tablas, los documentos, filas, y los campos, columnas.

Un modelo de datos orientado a los documentos no tiene esquemas. Por lo tanto, a diferencia de un RDBMS, que almacena valores NULOS en los campos vacíos, en MongoDB, si no hay datos, no hay campo para almacenarlos. Esto significa que no necesita preocuparse por los cambios en un esquema existente cuanto está desarrollando, lo que le permite mayor agilidad para cumplir con los requisitos empresariales en constante evolución. Esto abre la puerta a la innovación, ya que facilita la evolución de las aplicaciones.

El modelo orientado a los documentos de MongoDB también disminuye la necesidad de crear combinaciones, porque no divide documentos normalizados en tablas más pequeñas. Con menos combinaciones, mejora considerablemente la escalabilidad y la velocidad. Lo bueno de

MongoDB es una solución especialmente buena cuando necesita desplegar con rapidez aplicaciones web desarrolladas en JavaScript, utilizar muchos contadores en tiempo real o almacenar una gran cantidad de imágenes.

MongoDB, frente a otras bases de datos NoSQL, es que aún puede utilizar combinaciones si quiere combinar datos de múltiples colecciones. MongoDB también ofrece capacidades automáticas de particionamiento y compatibilidad con aplicaciones geoespaciales, lo que la hace ideal para el tipo de aplicaciones que analizo aquí. En comparación con el particionamiento de los RDBMS, que resulta complicado por las múltiples tablas y combinaciones, el particionamiento con MongoDB se realiza particionando el espacio de claves, porque la clave es la identificación del documento y el documento es el valor en los almacenes de documentos clave-valor.

En realidad, todas las bases de datos NoSQL tienen alguna forma de particionamiento que reduce la latencia y mejora la escalabilidad. Gracias al uso del particionamiento automático y los documentos BSON, MongoDB ofrece una base de datos flexible y de alta velocidad que permite un enfoque más ágil y con mayor capacidad de respuesta a los requisitos empresariales en constante evolución. MongoDB es una solución especialmente buena cuando necesita desplegar con rapidez aplicaciones web desarrolladas en JavaScript, utilizar muchos contadores en tiempo real o almacenar una gran cantidad de imágenes. Es sorprendentemente rápida cuando se la utiliza para satisfacer las necesidades de elaboración de informes y consultas de IoT en tiempo real. Y con la compatibilidad geoespacial, es ideal para usarla en

aplicaciones cuando es importante saber dónde está el usuario o mostrarle adónde ir.

Los desarrolladores también apreciarán la capacitación *on-demand* que ofrece MongoDB. Asimismo, proporciona asistencia para el desarrollo basada en proyectos, frente a la asistencia basada en el servidor, lo que lo convierte en una gran ayuda tanto para desarrolladores como administradores de operaciones.

EDB Postgres Advanced Server

EDB Postgres es en realidad una versión ampliada de la base de datos relacional PostgreSQL de código abierto, distribuida por EnterpriseDB. Si bien EDB Postgres solía ser una versión detrás de PostgreSQL, realmente se esfuerza por ser una parte activa e integral de la comunidad de PostgreSQL.

La velocidad y la escalabilidad son las ventajas clave que aporta PostgreSQL, y EDB lo complementa con un amplio conjunto de herramientas de nivel empresarial. Tener acceso al ecosistema de la comunidad PostgreSQL es una clara ventaja, pero otras surgen también de esta base de datos escalable y de alto rendimiento.

PostgreSQL admite múltiples tipos de datos, incluidos los definidos por el usuario (como XML), colecciones de tablas y *varrays*. Admite datos de textos, indexación y búsqueda, y permite realizar operaciones de lectura/escritura para una ejecución sin bloqueos empleando el control de concurrencia multiversión. Finalmente, los procedimientos almacenados se pueden escribir en lenguajes como C/C++, Java, JavaScript, Python, Perl y Ruby, lo que brinda gran libertad y flexibilidad. Además, EDB Postgres ofrece características de seguridad de nivel empresarial, incluidos los perfiles de contraseña ampliada. A través del complemento PostGIS de código abierto, EDB Postges se puede

implementar como una base de datos espacial *back-end* para sistemas de información geográfica (GISes). También incluye una herramienta de GUI para crear y depurar desencadenantes y procedimientos almacenados.

Además, EDB Postgres ofrece optimización del escalado vertical y mejoras de la escalabilidad ampliada para bloquear subsistemas, lo que mejora el rendimiento. Hace posible la integración en bases de datos diferentes (admitiendo bases de datos relacionales, de documentos y clave-valor), lo que permite combinar datos no estructurados, estructurados y transaccionales. También puede usarla para aplicaciones de solo lectura donde la alta velocidad es esencial. Permite que los administradores de bases de datos (DBA) prioricen de forma selectiva en los procesos tanto el consumo de E/S como de la CPU, y es compatible con Oracle. EDB ofrece una suscripción para desarrolladores de Postgres que proporciona acceso directo a la pericia de Postgres, videos técnicos, gran cantidad de documentación y una comunidad increíblemente sólida. Asimismo, proporciona una suite de herramientas de nivel empresarial para aliviar las cargas habituales asociadas con la migración, la integración y la gestión.

Cabe destacar que EDB Postgres se vende como un DBMS con suscripción, que incluye todas las actualizaciones, el mantenimiento y la asistencia, además del software.

El resultado es que EDB Postgres Advanced Server es una excelente solución de base de datos relacional moderna porque es segura, escalable, flexible y rápida, especialmente cuando se ejecuta en una arquitectura de servidor optimizada (luego desarrollaré este tema). Ofrece todo lo que usted espera de un RDBMS de nivel empresarial, sin las elevadas tasas de licencia asociadas con otros proveedores, pero con toda la innovación que busca en una solución de código abierto.

Base de datos de gráficos Neo4j

Si bien puede parecerlo al principio, realmente no hay nada mágico con las bases de datos de gráficos. Un gráfico está compuesto de dos elementos básicos: nodos y relaciones. Cada nodo representa un dato: una cosa, una entidad. Cada relación representa cómo se relacionan dos nodos. Los sitios de redes sociales donde los usuarios se siguen unos a otros, como Facebook o Tumblr, son típicos ejemplos de esta idea. Los usuarios son los nodos y el "seguimiento" es la relación entre los nodos.

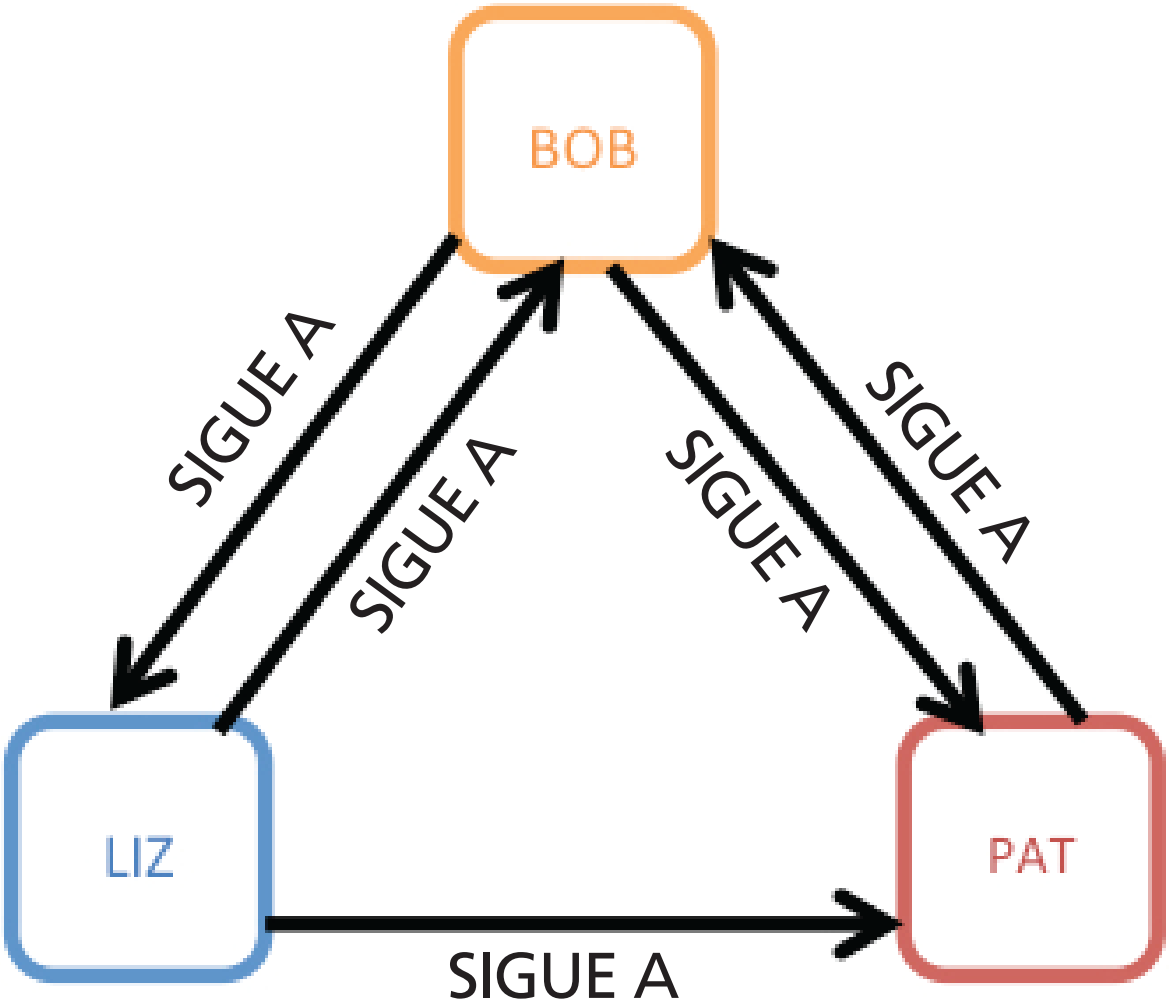


FIGURA 3. Nodos y relaciones

En una base de datos de gráficos, las relaciones tienen prioridad, lo que significa que los modelos de datos son más simples y más expresivos, y no tiene que preocuparse sobre cuestiones como las claves externas. Además, significa que nunca puede tener una relación sin dos nodos y que no puede eliminar un nodo sin eliminar también sus relaciones.

Las bases de datos de gráficos tienen dos características clave que deben considerarse para comprender la ventaja que una base de datos de gráficos le ofrece al desarrollo de aplicaciones. La primera es la diferencia entre el almacenamiento de gráficos nativo y no nativo. Las bases de datos de gráficos nativas, como Neo4j, están especialmente diseñadas para almacenar y gestionar gráficos, frente a las bases de datos orientadas a los objetos o relacionales que se adaptan para ser similares a las de gráficos. Las bases de datos de gráficos no nativas utilizan bases de datos orientadas a los objetos o relacionales para almacenar datos. Por lo tanto, cuando el volumen de los datos y la complejidad de las consultas aumentan, una base de datos de gráficos no nativa termina siendo mucho más latente.

La segunda característica favorable de Neo4j es su motor de procesamiento de gráficos. En el procesamiento de gráficos nativo, los nodos conectados se señalan directamente entre sí. Esto se llama adyacencia sin índice y es el método más eficiente para procesar datos en una base de datos de gráficos. El motor de procesamiento de gráficos nativo de Neo4j proporciona rendimiento constante en tiempo real porque evita las costosas búsquedas de índice que las bases de datos no nativas deben realizar.

Estas características son excelentes para las aplicaciones de gestión de acceso e identidades, donde necesita rastrear usuarios y autorizaciones a gran velocidad, o para motores de recomendación en tiempo real, que impulsan muchas aplicaciones de personalización y productos de comercio electrónico. Y, como sugerí antes, esas características son indispensables cuando necesita realizar un análisis en tiempo real de los datos de

aplicaciones sociales.

Neo4j es particularmente ágil gracias a su modelo de datos adaptable, usted puede responder a las necesidades empresariales emergentes con cambios en los modelos de datos sin preocuparse por el impacto en la funcionalidad actual. Asimismo, está desarrollada para la velocidad. Gracias a las relaciones explícitas entre los nodos, Neo4j evita la ineludible lentitud posterior cuando los conjuntos de datos aumentan.

También proporciona excelente asistencia en línea para desarrolladores, lo que incluye una formidable biblioteca de documentos, base de conocimientos, acceso a entornos aislados y, al igual que los otros OSDBMS que describí, un sólido sistema de asistencia de la comunidad.

Sistemas de bases de datos en memoria

Los sistemas de bases de datos en memoria (IMDBS), como Redis y Kinetica, almacenan datos en la memoria principal, a diferencia de los sistemas de bases de datos tradicionales que están diseñados para almacenar datos en medios permanentes. Si bien técnicamente podría almacenar una base de datos tradicional en la memoria RAM, aún se vería agobiado con la sobrecarga de un sistema diseñado para el almacenamiento en disco. Precisamente porque un IMDBS almacena datos en la memoria, evitando así la sobrecarga de las operaciones de E/S y caché, es progresivamente más veloz que un DBMS tradicional. A su vez, debido a este diseño simple, los IMDBS tienen requisitos mucho más bajos de CPU y memoria.

Las grandes candidatas para los IMDBS son las aplicaciones que necesitan acceso veloz a los datos, y también su rápida manipulación. Las aplicaciones sociales, de comercio electrónico y de mercado financiero, y los sistemas integrados en tiempo real son excelentes opciones para los IMDBS, ya que pueden obtener ventajas reales de su velocidad. Y la rapidez no es la única característica de los IMDBS, también tienen muy buena escalabilidad. No es

No es inusual que un IMDBS crezca más allá de su rango de tamaño de terabytes a la vez que mantiene todas las ventajas de rendimiento sobre las soluciones de DBMS tradicionales.

inusual que un IMDBS crezca más allá de su rango de tamaño de terabytes a la vez que mantiene todas las ventajas de rendimiento sobre las soluciones de DBMS tradicionales.

Redis Una base de datos clave-valor, o almacén, es aquella diseñada para almacenar, recuperar y gestionar matrices asociativas. Una matriz asociativa es un modelo de datos simple donde cada clave se asocia con un único valor en una colección, una relación que se conoce como par clave-valor. Una cadena arbitraria, como un nombre de archivo, *hash* o URI, representa la clave en cada par clave-valor. El valor, que se almacena como un *blob*, puede ser cualquier tipo de datos, como una imagen o un documento. Como el valor se almacena como un *blob*, no requiere definición de esquema ni modelado de datos con anticipación. Esto también elimina la necesidad de indexar los datos para mejorar el rendimiento. No puede simplemente filtrar ni controlar lo que devuelve una solicitud en base al valor, porque este es opaco.

Los almacenes clave-valor utilizan los comandos obtener, poner y eliminar en lugar de un lenguaje de consulta, lo que significa que la ruta para recuperar datos es una solicitud directa al objeto en la memoria. en lugar de un lenguaje de consulta, lo que significa que la ruta para recuperar datos es una solicitud directa al objeto en memoria. La relación entre los datos no se calcula, por lo que no hay sobrecarga de la optimización. No necesita preocuparse por dónde almacenar índices, por la velocidad de la red ni por el equilibrio en un sistema distribuido. Gracias a esta simplicidad, un

almacén clave-valor es muy rápido y flexible, simple de usar, altamente escalable y portátil. Redis es un almacén de estructura de datos clave-valor en memoria de código abierto (con licencia BSD), que se utiliza como una base de datos, caché y agente de mensajes.

Una de las ventajas de usar Redis surge del uso de sus comandos primitivos, como LPUSH, LTRIM y LREM. Esto permite que las tareas lentas y complejas con los almacenes de datos tradicionales se realicen de una forma mucho más simple. Por ejemplo, en una aplicación web, los artículos borrados se pueden eliminar de la caché usando LREM, o puede usar LPUSH para insertar una identificación de contenido en el encabezado de la lista almacenada en una clave para mostrar los listados de elementos más recientes en una página de inicio, y puede usar LTRIM para limitar ese número de elementos de la lista. Con estos tres simples comandos primitivos, Redis facilita en gran medida el trabajo de un desarrollador.

Gracias a su simplicidad, velocidad y baja latencia, Redis también es una excelente solución para el desarrollo de aplicaciones de comercio electrónico si busca almacenar de manera eficiente los perfiles y las preferencias de los usuarios, para recomendar productos en función de lo que ven los usuarios, o para presentar cupones y anuncios publicitarios en tiempo real personalizados según los hábitos de consumo de un cliente. Dado que todos los datos están en la memoria, se eliminan las demoras para encontrar esos datos, lo que tiene como resultado un rendimiento de máxima velocidad. Al usar Redis como caché frente a otra DB, por ejemplo, se obtienen grandes ventajas en cuanto a velocidad.

Personalmente, también valoro la asistencia en línea disponible para Redis. Incluye una lista completa de comandos como parte de una guía de programación detallada, junto con múltiples tutoriales, guías administrativas y otros recursos para desarrolladores. Para obtener información completa y detallada sobre las numerosas ventajas que Redis ofrece, visite

https://redis.io.

Aceleración por GPU con Kinetica. Nuevamente, Kinetica no es una base de datos de código abierto, sino parte del ecosistema de *hardware/software* abierto de la Fundación OpenPOWER. Kinetica es una base de datos en memoria y distribuida que es acelerada por las unidades de procesamiento gráfico (GPU). Una GPU es simplemente un circuito diseñado para acelerar la creación de imágenes para mostrar alterando y manipulando la memoria con rapidez. Mientras una CPU tiene muchos núcleos y una gran memoria caché, una GPU tiene miles de núcleos, lo que da lugar a aumentos de velocidad más de 100 veces superior a la de una CPU en algunos casos. Como son excelentes para aceptar grandes cantidades de datos y realizar la misma operación una y otra vez, las GPU originalmente se concibieron para el renderizado de juegos en 3D. Recientemente, las GPU se han empezado a usar acelerando las cargas de trabajo computacionales en funciones como modelado financiero, investigación, exploración de energía e inteligencia artificial. De hecho, gracias a la arquitectura de procesamiento paralelo que produce velocidades de procesamiento hasta 100 veces superior, Kinetica es ideal para analizar grandes volúmenes de datos de *streaming*. Es perfecta para el análisis predictivo que define las cargas de trabajo de inteligencia artificial (AI).

Kinetica aprovecha la potencia de procesamiento de la GPU para gestionar grandes conjuntos de datos, especialmente los datos de *streaming*, en una fracción del tiempo y en un espacio de *hardware* mucho menor que las bases de datos tradicionales. Esto es especialmente útil para las aplicaciones de IoT y la visualización geoespacial. Incluye herramientas de visualización que pueden renderizar grandes cantidades de datos, y no hay necesidad de preparar el esquema antes de analizar los datos. Kinetica es una herramienta aliada excelente para los sistemas transaccionales, los almacenes y los lagos de datos, y como es totalmente compatible con SQL, es fácil consultar.

También admite REST, JSON, Java, JavaScript, C++ y Python, entre otros, por lo que resulta ideal para los desarrolladores. Esto significa que puede explorar grandes conjuntos de datos más rápido que nunca sin tener que aprender nuevos lenguajes de programación o consulta, ni desarrollar nuevos modelos de datos.

Hasta aquí, he analizado varias ofertas de bases de datos del escenario de los OSDBMS. En la categoría no relacional están MongoDB, con su modelo de datos orientado a los documentos; Neo4j, con su modelo de gráficos, y Redis, que ofrece un enfoque de clave-valor. En el campo relacional están EDB Postgres, una excelente base de datos analítica, y Kinetica, una opción en memoria. La ventaja que comparten todos estos sistemas de gestión de bases de datos es la velocidad, esencial al hablar del desarrollo de aplicaciones analíticas de *Big Data*.

Ahora vayamos, finalmente, a una plataforma que resulta ideal para alojar estos OSDBMS y las aplicaciones desarrolladas a partir de estos: los servidores OpenPOWER LC diseñados desde cero para *Big Data* por IBM y sus socios de la Fundación OpenPOWER.

Por qué la implementación de OSDBMS en los sistemas OpenPOWER de IBM es el enfoque correcto

Aprovechar las capacidades de los OSDBMS para crear soluciones realmente innovadoras requiere no solo velocidad de procesamiento, sino también colaboración. Antes en este *ebook*, mencioné a la Fundación OpenPOWER, un consorcio con más de 250 miembros que incluye a algunos de los nombres más importantes en el mundo de la tecnología: IBM, Google, Mellanox Technologies, Tyan, Xilinx y Canonical. Hace años que la Fundación OpenPOWER colabora en el diseño de sistemas basados en la arquitectura de procesadores POWER de IBM. Los servidores OpenPOWER

FIGURA 4. Por qué IBM POWER8 es ideal para OSDBMS

LC de IBM son una de las más recientes manifestaciones de esa colaboración comercialmente disponible.

Los servidores OpenPOWER LC de IBM con tecnología de procesadores POWER8 fueron diseñados para cargas de trabajo de *Big Data*, incluidas las soluciones de OSDBMS que he analizado aquí. IBM POWER8 ofrece el cuádruple de caché de procesador, subprocesos y ancho de banda de memoria que las plataformas básicas.

IBM POWER8 ejecuta Linux estándar de la industria de Red Hat, SUSE y Canonical. Esto hace que trasladar las aplicaciones de Linux x86 a Power sea más atractivo y sencillo que nunca. Linux en Power brinda la plataforma innovadora que los desarrolladores realmente necesitan para aprovechar la potencia y la escala de los OSDBMS para aplicaciones de uso intensivo de datos.

El diseño del POWER8 combina potencia informática, ancho de banda de la memoria y rendimiento de E/S para producir la velocidad necesaria para las cargas de trabajo analíticas y de *Big Data*. El procesador POWER8 está diseñado para ofrecer el cuádruple de subprocesos por núcleo y ancho de banda de la memoria frente a la infraestructura básica, como así también mayor capacidad de la memoria, con sistemas de escalado externo *(scale-out)* que entregan hasta dos terabytes en un servidor de dos *sockets* hasta alcanzar 16 terabytes para servidores de escalado empresariales. Además, POWER8 proporciona una caché por procesador cuatro veces superior a una latencia más baja, lo que permite procesar más datos con mayor rapidez.

MongoDB Para MongoDB, esto es fantástico, porque le permite obtener una plataforma que ofrece una visualización integrada en tiempo real de todos sus datos. Según IBM, MongoDB en POWER8 proporciona un rendimiento por servidor un 40 % superior al de Intel Xeon. Es una excelente solución para la expansión del servidor del centro de datos, y si también se consideran los costos de despliegue, MongoDB en POWER8 ofrece el doble de rendimiento por dólar en comparación con los sistemas basados en x86. Es una excelente noticia si quiere ahorrar el dinero de la empresa para la futura innovación.

EDB Postgres Advanced Server EDB Postgres Advanced Server también se ejecuta en Linux little-endian Linux en POWER8, lo que elimina las barreras de portabilidad. Ejecutar EDB Postgres Advanced Server en los servidores OpenPOWER LC de IBM ofrece subprocesos de alto rendimiento, más caché, mayor ancho de banda para los datos y el doble de mejora en la relación precio-rendimiento en comparación con los sistemas basados en x86. Los *benchmarks* de IBM han demostrado que los servidores OpenPOWER LC tienen un rendimiento por núcleo un 60 % superior frente a Intel Xeon. Una vez más, esta es la oportunidad perfecta para que su

A través de esta solución, que funciona con cualquier cliente de Redis sin cambios en la API estándar de Redis, un solo servidor POWER8 con aceleración CAPI Flash puede procesar más de 200.000 operaciones por segundo con una latencia de submilisegundos; eso es rápido.

empresa despliegue más cargas de trabajo en menos tiempo y gaste menos en despliegues de infraestructura para sus aplicaciones de *Big Data*, dejando más recursos disponibles para la innovación.

Redis También miembro de la Fundación OpenPOWER, Redis Labs e IBM Power Systems colaboran estrechamente para brindar una solución Redis optimizada para POWER8 y su Coherent Accelerator Processor Interface (CAPI) como parte de la compatibilidad de Redis con IBM Data Engine for NoSQL, que ejecuta Redis en IBM FlashSystem 840, la tarjeta CAPI Flash de IBM y Redis Labs Enterprise Cluster (RLEC) para *software* Flash como reemplazo de la memoria RAM. A través de esta solución, que funciona con cualquier cliente de Redis sin cambios en la API estándar de Redis, un solo servidor POWER8 con aceleración CAPI Flash puede procesar más de 200.000 operaciones por segundo con una latencia de submilisegundos; eso es rápido. A su vez, puede almacenar el 90 % de un conjunto de datos de multiterabytes en Flash y solo el 10 % en la memoria RAM. En comparación con una solución Redis totalmente basada en RAM, esto reduce los costos de despliegue en más de un 70 %. Las pruebas de Redis han demostrado que los servidores OpenPOWER LC de IBM tienen un rendimiento un 67 % superior frente a los x86.

Neo4j Neo4j en servidores OpenPOWER LC de IBM ofrece la plataforma para bases de datos de gráficos más escalable del mundo, capaz de almacenar y procesar gráficos asombrosamente grandes. Uno de los principales desafíos en el procesamiento de gráficos a escala es cómo manejar el tamaño del conjunto de datos sin comprometer las capacidades en tiempo real. Con los 56 TB disponibles de memoria ampliada en el servidor LC gracias a la tecnología CAPI y Flash que describí antes, el tamaño de las consultas en tiempo real aumenta, porque el tamaño de los gráficos que se pueden almacenar en memoria se ha incrementado. Sin embargo, la línea de servidores LC con POWER8 se equilibra también. Cada núcleo puede manejar ocho subprocesos de *hardware* a la vez, para un total de 96 subprocesos concurrentes en un chip de 12 núcleos. Los controladores de memoria en chip permiten un gran ancho de banda en la E/S del sistema y la memoria. Con la aceleración CAPI habilitada en un servidor OpenPOWER LC de IBM, el rendimiento es casi del doble que el de Intel Xeon.

Kinetica Obtener el mayor rendimiento de Kinetica realmente depende de la velocidad con que los datos se trasladan entre la CPU y la GPU, porque Kinetica se diseñó para aprovechar la memoria del sistema. La nueva tecnología NVIDIA NVLink e IBM POWER8 proporcionan el enfoque más avanzado y más asequible para ofrecer analítica de alto rendimiento con Kinetica. Las interconexiones, como NVIDIA NVLink, abren una ruta más amplia entre la CPU y la GPU, lo que permite que Kinetica aproveche al máximo la memoria del sistema. El procesamiento de la GPU ya no se limita a la velocidad con que los datos se pueden mover a través del subsistema de E/S, permitiendo que Kinetica admita conjuntos de datos más grandes. IBM ha demostrado que Kinetica produce un increíble rendimiento 2,5 veces superior cuando se ejecuta en un servidor OpenPOWER LC de IBM con NVLink en comparación con un sistema x86 similar.

Conclusión

Hay todo un mundo complejo generando un gran volumen de datos también complejos. Las bases de datos, los enfoques de desarrollo y las plataformas de procesamiento tradicionales no van a proporcionar el rendimiento necesario para satisfacer las demandas actuales de capacidades analíticas y de datos en tiempo real.

Los OSDBMS les ofrecen a los desarrolladores innovadores varias ventajas, como rendimiento y flexibilidad para la avalancha de datos qe se generan y mueven en la actualidad, y quizá algo todavía más importante, las ventajas de modelos de desarrollo abiertos, incluido el acceso a una comunidad dinámica que se adapta con rapidez y eficacia, al cambio y a los problemas del mundo real. En los OSDBMS, no solo encontramos funciones aceleradas y con mayor capacidad de respuesta, sino que las herramientas de nivel empresarial y la ausencia de las restricciones de las soluciones registradas hacen que los OSDBMS sean más útiles y accesibles.

A medida que las demandas exigidas a los desarrolladores de aplicaciones sigan evolucionando, las propias aplicaciones deben continuar innovando en respuesta a esas demandas. El ecosistema abierto de los OSDBMS ofrece el entorno necesario para que las ideas y la innovación fluyan en soluciones reales y útiles. A través de una estrecha colaboración, IBM y los líderes del ámbito de los OSDBMS ofrecen plataformas líderes de la industria para aprovechar al máximo estas nuevas tecnologías de bases de datos.

Descubra usted mismo cómo puede aprovechar las ventajas del desarrollo de aplicaciones con los OSDBMS en POWER haciendo clic en este enlace:

https://www-01.ibm.com/marketing/iwm/dre/signup?source=mrs-form-12148&S_PKG=ov53321.n