

Comparación entre Oracle Exadata e IBM Netezza Data Warehouse Appliance

*por Phil Francisco, Vicepresidente, Product Management y Product Marketing
de IBM Mike Kearney, Senior Director, Product Marketing, IBM*



Contenido

Introducción	2
Proceso de transacciones online (OLTP) y data warehousing	3
Rendimiento de las consultas	5
Simplicidad de operación	10
Valor	14
Conclusión	16

Introducción

Los appliances de data warehouse de IBM Netezza se centran en tecnología diseñada para consultar y analizar gran cantidad de datos. Los appliances de data warehouse de IBM Netezza están alterando el mercado. Deseosos de utilizar datos con el menor coste de operación y propiedad, muchos de nuestros clientes han migrado sus data warehouses desde Oracle. Oracle acaba de sacar al mercado Exadata, una máquina que aparentemente hace todo lo que el appliance de data warehouse IBM Netezza hace, y también procesa transacciones online. Este examen de Exadata e IBM Netezza como plataformas de data warehouse se ha escrito a partir de un punto de vista manifiesto de IBM, aunque con el fin de asegurar la credibilidad nos hemos asesorado con Philip Howard, Director de Investigaciones de Bloor Research y Curt Monash, Presidente de Monash Research.

La innovación nos obliga a pensar y hacer cosas de manera diferente, resolviendo un problema por medio de nuevos enfoques. Los appliances de data warehouse de IBM Netezza ofrecen un excelente rendimiento para las consultas de warehouse de nuestros clientes. Los appliances de data warehouse de IBM Netezza ofrecen simplicidad a los clientes; cualquiera con un conocimiento básico de SQL y Linux tiene las habilidades necesarias para llevar a cabo las pocas tareas administrativas necesarias para

“Netezza formó parte de la inspiración para Exadata. Teradata formó parte de la inspiración para Exadata. Nos gustaría agradecerles que nos obligaran a entrar en el negocio del hardware.”

— Larry Ellison, enero de 2010

mantener niveles de servicio coherentes mediante cargas de trabajo con cambios dinámicos. El rendimiento de los appliances de data warehouse de IBM Netezza con su simplicidad reducen los costes de propiedad y de ejecución de sus data warehouses. Más importante aún, nuestros clientes crean nuevo valor de negocio mediante el despliegue de aplicaciones analíticas que hasta ahora consideraban más allá de su alcance.

“Netezza formó parte de la inspiración para Exadata. Teradata formó parte de la inspiración para Exadata”, reconoció Larry Ellison el 27 de enero de 2010. “Nos gustaría agradecerles que nos obligaran a entrar en el negocio del hardware.”¹ Aunque expresado con el lenguaje peculiar de Larry Ellison, existe un punto

¹ Consulte http://oracle.com.edgesuite.net/ivt/4000/8104/9238/12652/lobby_external_flash_clean_480x360/default.htm

importante en este comentario: sólo los mejores captan la atención de Oracle. Exadata representa una dirección estratégica para Oracle; la adaptación de su sistema de gestión de bases de datos OLTP, en colaboración con un sistema de almacenamiento masivo en paralelo de Sun. Oracle lanzó Exadata V2 con la promesa de un rendimiento extremo para el procesamiento tanto de transacciones online como de consultas analíticas. Por lo tanto, Oracle Exadata V2 es una plataforma de propósito general para la gestión de cargas de trabajo mixtas. Oracle Database se diseñó para OLTP. Pero el data warehousing y la analítica genera demandas más diferentes en su software y hardware que OLTP. Simplificando un poco, algunas cargas de trabajo para data warehouse funcionan mucho mejor y son mucho más asequibles en un sistema diseñado específicamente para la analítica. Las credenciales de data warehouse de Exadata exigen un estudio detallado, particularmente con respecto a la simplicidad y el valor.

Este libro electrónico se inicia con la revisión de las diferencias entre el procesamiento de transacciones online y el procesamiento de consultas y análisis en un data warehouse. A continuación se comenta Exadata y el appliance de data warehouse de IBM Netezza desde las perspectivas del rendimiento de las consultas, la simplicidad de operación y el valor.

Lo único que pedimos a nuestros lectores es que lo hagan como nuestros clientes y socios lo han hecho: dejando a un lado las nociones de cómo debe ser un sistema de gestión de bases de datos, abrirse a nuevas formas de pensar y prepararse para hacer menos, no más, para lograr un mejor resultado.

Una advertencia: el equipo del appliance de data warehouse de IBM Netezza no tiene acceso directo a una máquina Exadata. Nos sentimos afortunados por los comentarios detallados que recibimos de muchas organizaciones que han evaluado ambas tecnologías y han seleccionado los appliances de data warehouse de IBM Netezza. Dado el tamaño de Oracle y su foco en Exadata, la información pública disponible de Exadata es sorprendentemente escasa. Los casos de uso citados por Oracle aportan poco al debate, que por sí mismo preocupa a varios seguidores del sector, p. ej. Information Week.² La información compartida en este artículo se ofrece con el espíritu de la apertura. Todas las imprecisiones son consecuencia de nuestros errores, no un intento de confundir.

² Consulte http://www.informationweek.com/news/business_intelligence/warehouses/showArticle.jhtml?articleID=225702836&cid=RSSfeed_IWK_News

Proceso de transacciones online (OLTP) y data warehousing

Los sistemas OLTP ejecutan muchas transacciones cortas. El alcance de cada transacción es reducido, limitado a uno o a un pequeño número de registros y es tan previsible que muchas veces los datos se guardan en caché. Aunque los sistemas OLTP procesan grandes volúmenes de consultas de base de datos, su objetivo es escribir (UPDATE, INSERT y DELETE) en un conjunto de datos actual. Estos sistemas habitualmente son específicos de un proceso o función de negocio, por ejemplo la gestión del saldo actual de una cuenta de cheques. Sus datos habitualmente se estructuran en forma normal tres (3NF). Los tipos de transacción de los sistemas OLTP son estables y sus requisitos de datos se conocen bien, por lo que estructuras de datos secundarias, como los índices, pueden localizar útilmente registros en el disco, con anterioridad a su transferencia a memoria para su procesamiento.

En comparación, los sistemas de data warehouse se caracterizan por operaciones predominantemente de lectura de base de datos (SELECT) en un conjunto de datos actual e histórico. Mientras que una operación de OLTP accede a un pequeño número de registros, una consulta de data warehouse puede explorar una tabla de miles de millones de filas y unir sus registros

con los de varias otras tablas. Es más, las consultas en un data warehouse normalmente son imprevisibles por naturaleza, por lo que es difícil utilizar estrategias de indexado y caché. Las opciones para la estructuración de datos en el warehouse van desde 3NF hasta modelos dimensionales como los esquemas de estrella y copo de nieve. Los datos de cada sistema que alimenta a un warehouse típico se estructuran para reflejar las necesidades de un proceso de negocio específico. Antes de cargar los datos en el warehouse se limpian, se optimizan y se integran.

Este libro electrónico divide los data warehouses en las categorías de primera o segunda generación. Aunque esta clasificación no permita realizar un escrutinio en gran profundidad, sí refleja el modo en que nuestros clientes hablan de su evolución hacia la generación de más valor a partir de sus datos.

Los data warehouses de primera generación habitualmente se cargan por la noche. Proporcionan información a sus empresas a través de un cuerpo estable de dashboards e informes basados en SQL de evolución lenta. Puesto que estos warehouses sencillos de algún modo se parecen a los sistemas OLTP – sus requisitos de carga de trabajo y de datos son conocidos y estables – las organizaciones suelen adoptar los mismos productos de gestión de base de datos que utilizan para OLTP. Con el producto viene la práctica: los administradores de bases de datos analizan los requisitos de datos de cada informe y generan índices para acelerar la recuperación de datos.

El avance de las técnicas y de la tecnología OLTP parece tener éxito, hasta que los volúmenes de datos del warehouse dejan atrás a los habitualmente gestionados en los sistemas transaccionales.

En este siglo, las corporaciones y las agencias del sector público aceptan índices de crecimiento de datos de un 30-50% por año como normal. Las tecnologías y las prácticas que tienen éxito en el mundo de OLTP son menos aplicables al data warehousing; el índice como ayuda para la recuperación de datos es un caso paradigmático. Cuando el sistema de base de datos procesa trabajos para cargar datos, también está ocupado actualizando sus múltiples índices. Con grandes volúmenes de datos, esto se convierte en un proceso muy lento, lo que provoca que los trabajos superen sus ventanas de procesamiento asignadas. A pesar de trabajar muchas horas, el equipo técnico no logra ofrecer los niveles de servicio negociados con la empresa. La productividad se ve afectada cuando las unidades de negocio esperan la disponibilidad de datos e informes.

Las tecnologías y las prácticas que tienen éxito en el mundo de OLTP son menos aplicables al data warehousing...

Las organizaciones redefinen la forma en que necesitan y quieren utilizar sus datos; este libro electrónico hace referencia a este desarrollo como el data warehouse de segunda generación. Estos nuevos warehouses, que gestionan enormes conjuntos de datos con facilidad, sirven como memoria corporativa. Cuando se les interroga, recuerdan eventos registrados años antes; estas memorias distantes incrementan la exactitud de las aplicaciones de analítica predictiva. Un goteo constante sustituye a las cargas de proceso por lotes nocturnos, reduciendo la latencia entre el registro de un evento y su análisis. Más allá del SQL sencillo utilizado para llenar informes y dashboards, el warehouse procesa regresiones lineales, Naive Bayes y otros algoritmos matemáticos de analítica avanzada. Al detectar un pico repentino de ventas en un producto de margen alto en tan solo cinco tiendas lleva a un comerciante a saber lo que ha ocurrido y por qué. Este conocimiento aporta información a las estrategias para promocionar actividades de ventas similares en todas las 150 tiendas. El sistema informático que sostiene al warehouse debe ser capaz de gestionar estas puntas repentinas de demanda sin interrumpir los informes y los dashboards normales. Los usuarios de negocio exigen la libertad para utilizar sus datos en el momento y del modo en que prefieran. Su apetito por la inmediatez no deja lugar a las tecnologías cuyo rendimiento depende del trabajo de ajuste de los administradores.

Rendimiento de las consultas

Rendimiento de las consultas con Oracle Exadata

Al adquirir a Sun, Oracle ha llegado a la conclusión a la que el equipo del appliance de data warehouse de IBM Netezza llegó una década antes: los sistemas de data warehouse alcanzan una eficiencia más alta cuando todas las partes, software y hardware, se optimizan para ese objetivo. Exadata se crea a partir de dos subsistemas conectados por una red rápida: un sistema de almacenamiento inteligente que se comunica a través de InfiniBand con una Oracle Database 11g V2 con Real Application Clusters (RAC). Un sistema de bastidor único incluye un nivel de almacenamiento de 14 servidores de almacenamiento, llamados celdas Exadata, en una red masiva de procesamiento en paralelo (MPP), junto con la base de datos Oracle RAC ejecutándose como clúster de discos compartidos de ocho nodos multiproceso simétricos.

Al adquirir a Sun, Oracle ha llegado a la conclusión a la que el equipo del appliance de data warehouse de IBM Netezza llegó una década antes: los sistemas de data warehouse alcanzan una eficiencia más alta cuando todas las partes, software y hardware, se optimizan para ese objetivo.

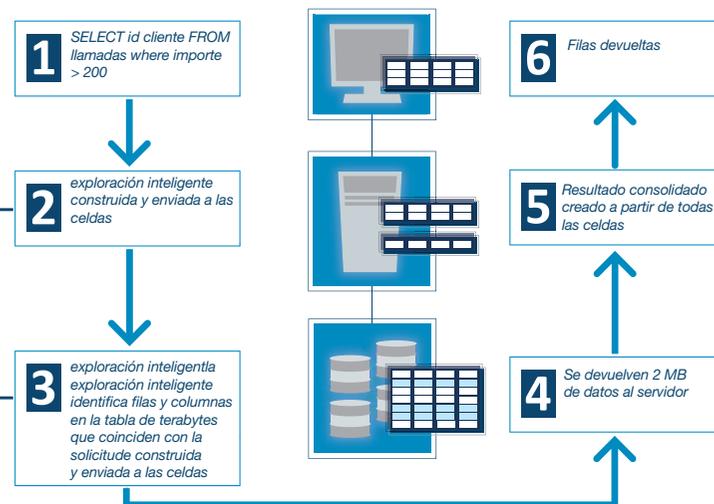
Limitaciones del escaneo inteligente

La exploración inteligente no es exhaustiva; el nivel de almacenamiento MPP de Exadata no puede procesar: tablas organizadas en índices – recomendado por Oracle para texto, imagen, audio y datos espaciales; los bloques de datos que contienen transacciones activas (INSERT, UPDATE, DELETE); uniones SQL entre varias tablas o uniones complejas entre dos tablas; 192 de las 511 funciones de base de datos de Oracle; funciones definidas por el usuario; o agregaciones distintas, habituales en los informes sencillos.

MPP infrutilizado

La ingeniería de Exadata no utiliza totalmente la arquitectura MPP. La gestión de bases de datos no se integra completamente en el nivel de almacenamiento, lo cual implica que se pide pocas cosas al hardware de su red MPP.

PROCESAMIENTO DE EXPLORACIÓN INTELIGENTE DE EXADATA



Oracle llama inteligente al nivel de almacenamiento de Exadata porque procesa proyección SQL, restricción y filtrado de uniones,³ antes de poner el conjunto de datos resultante en la red para su posterior procesamiento por parte de Oracle RAC. Esta técnica recibe el nombre de exploración inteligente. No obstante, el escaneo inteligente no es completo; el nivel de almacenamiento no procesa todas las restricciones. El foro online de Oracle⁴ enumera una serie de operaciones, incluidas las exploraciones de tablas organizadas en índices o tablas en clúster, que no se benefician del escaneo inteligente. Además de éstas, Christian Antognini, autor del libro Resolución de problemas de rendimiento de Oracle, escribe un blog que sugiere que el escaneo inteligente no se utiliza con el tipo de datos TIMESTAMP.⁵ Oracle recomienda implementar tablas de hechos en los data warehouses como tablas organizadas por índices para la ejecución eficiente de consultas en estrella.⁶ El nivel de almacenamiento de Exadata no procesa restricciones en las tablas organizadas por índices, sino que debe pasar todos los registros hasta la base de datos Oracle. El enfoque de Exadata de pasar todos los registros al nivel de base de datos es muy eficaz para OLTP, ya que cada transacción sólo debe recuperar un número reducido de filas. No obstante, un análisis estadístico que requiera el escaneo de una tabla de hechos larga (cientos o miles de millones de filas), y ancha (cientos de columnas) generará una marea de datos que se moverán ineficazmente por la red. Exadata lograría un mejor rendimiento y sería más eficiente si procesara todas las consultas de SQL (cláusulas WHERE) en su nivel de almacenamiento MPP.

Los servidores de almacenamiento de Exadata no pueden comunicarse entre sí; al contrario, toda la comunicación se ve forzada a realizarse a través de la red InfiniBand con Oracle RAC y de nuevo a través de la red hasta el nivel de almacenamiento. Esta arquitectura es beneficiosa para el procesamiento de transacciones online; en el que cada transacción, con un alcance de uno o pocos registros, se puede conseguir moviendo un conjunto reducido de datos desde el almacenamiento hasta la base de datos. Las consultas analíticas, tales como “encontrar todas las cestas de la compra vendidas el mes pasado en el Estado de Washington, Oregón y California que contengan el producto X con el producto Y y con un valor total de más de 35 dólares”, debe recuperar conjuntos de datos mucho más grandes, conjuntos que deben moverse del almacenamiento a la base de datos. Este movimiento ineficiente de datos grandes afecta negativamente al rendimiento de la consulta.

El nivel de almacenamiento de Exadata presenta otras limitaciones. Las celdas de Exadata no pueden procesar agregaciones distintas, que son muy comunes incluso en informes sencillos; no pueden procesar uniones complejas o funciones analíticas utilizadas en las aplicaciones analíticas. Incapaz de resolver estas consultas típicas de data warehousing en su nivel de almacenamiento, Exadata debe enviar conjuntos de datos muy grandes por su red interna hasta Oracle RAC. Este fallo de arquitectura plantea dudas en la idoneidad de Exadata para data warehouses de segunda generación que deben ejecutar consultas analíticas complejas.

Oracle posiciona su uso del switch InfiniBand de 40 Gb/seg como una ventaja frente al appliance de data warehouse de IBM Netezza; en realidad. Exadata necesita esta red costosa debido a la falta de equilibrio y de eficacia del sistema. Los servidores de almacenamiento de Exadata hacen demasiado poco trabajo, por lo que se ponen en la red más datos de los necesarios para moverlos hasta su procesamiento por parte de Oracle RAC, a quien se le pide que haga demasiado trabajo.

En su nivel de base de datos Exadata ejecuta Oracle 11g V2 con Real Application Clusters como arquitectura de discos compartidos en clúster. El uso de esta arquitectura para una plataforma de data warehouse plantea dudas de que la contención del recurso compartido imponga límites en el volumen de datos que la base de datos pueda procesar y el número de consultas que pueda ejecutar simultáneamente. El tiempo y las experiencias de los usuarios dirán si esta preocupación es justificada.

³ Visión general técnica del servidor de almacenamiento y de la máquina de base de datos de Sun Oracle Exadata – Documento técnico de Oracle, octubre de 2009.

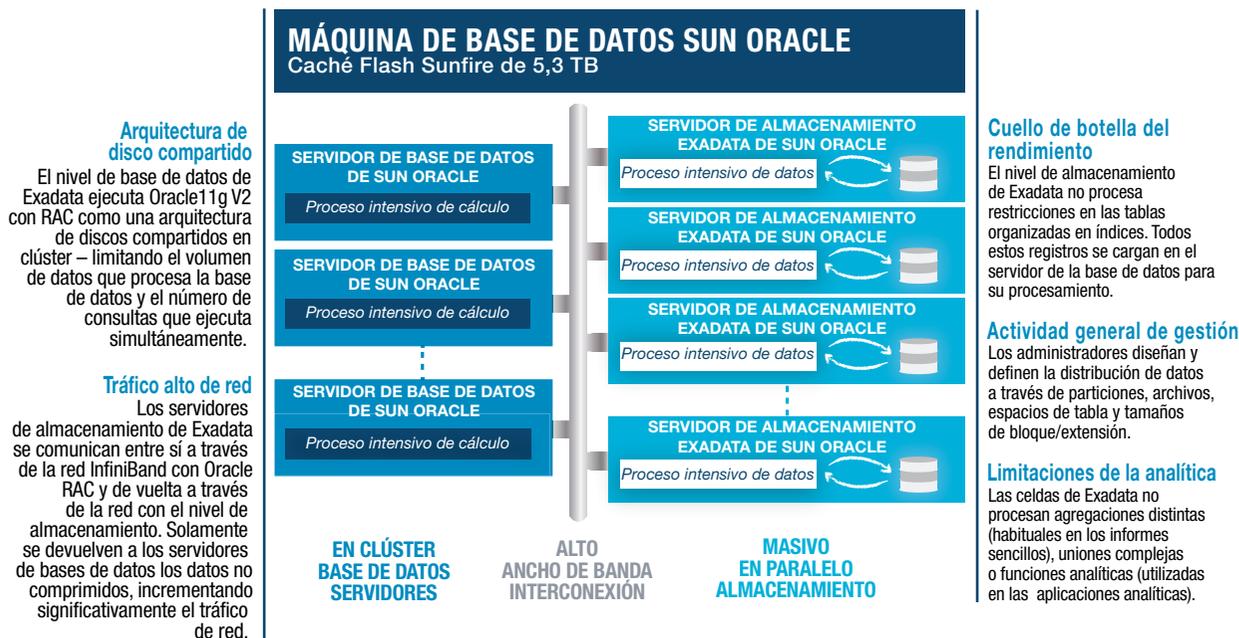
⁴ <http://forums.oracle.com/forums/thread.jspa?threadID=1036774&start=0>
La lista completa es: exploraciones de tablas organizadas en índices o tablas agrupadas en clúster; acceso a un índice comprimido; acceso a un índice de clave inversa; Búsqueda empresarial segura

⁵ Visite el blog de Christian Antognini en <http://antognini.ch/2010/05/exadata-storage-server-and-the-query-optimizer-%E2%80%93-part-2/>

⁶ http://www.oracle.com/technology/products/oracle9i/datasheets/iot/iot_ds.html

Todos los discos del nivel de almacenamiento de Exadata los comparten todos los nodos de la red que ejecuta Oracle RAC. Este almacenamiento comunitario crea el riesgo de que un nodo lea una página mientras otro la está actualizando. Para gestionarlo, Oracle fuerza la coordinación entre nodos. Cada nodo comprueba la actividad de disco de sus iguales para evitar el conflicto. Los técnicos de Oracle llaman a esta actividad “block pinging”. Los ciclos de cálculo consumidos mientras cada nodo comprueba la actividad de disco de sus iguales, o que se pierden mientras un nodo desocupado espera a que otro complete una operación, se pierden. En una arquitectura diseñada específicamente para el data warehousing, estos ciclos se podrían emplear para el procesamiento de consultas, minería de datos o ejecución de análisis.

Excepto en las consultas sencillas, Exadata debe mover grandes conjuntos de datos desde su nivel de almacenamiento hasta el de la base de datos, planteando dudas de su idoneidad como plataforma para un data warehouse moderno.

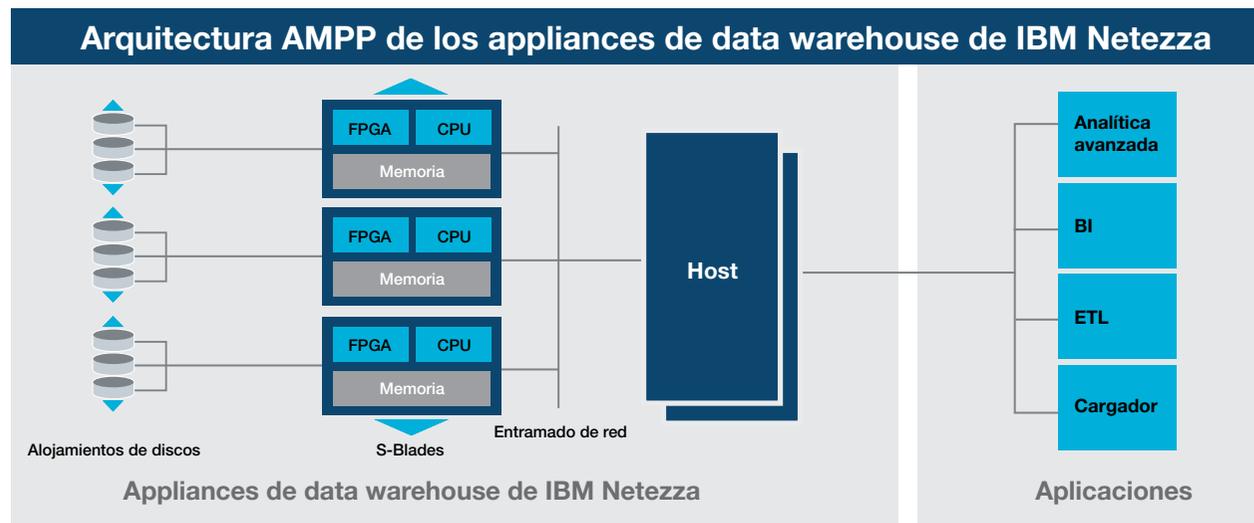


Casando su tecnología de base de datos existente con un nuevo nivel de almacenamiento “inteligente”, Exadata elimina el atasco de rendimiento de disco que sufre Oracle cuando se comunica con almacenamiento convencional. Puede parecer que Exadata presenta una interesante oportunidad para los CIOs que buscan consolidar varios sistemas OLTP en una única plataforma. No obstante, tras un examen más detallado, se hace patente que el volumen de almacenamiento y el software de almacenamiento incluidos en Exadata se ven superados por la gran mayoría de entornos OLTP. Excepto en las consultas sencillas, Exadata debe mover grandes conjuntos de datos desde su nivel de almacenamiento hasta el de la base de datos, planteando dudas de su idoneidad como plataforma para un data warehouse moderno.

Rendimiento de las consultas con el appliance de data warehouse de IBM Netezza

El appliance de data warehouse de IBM Netezza se ha diseñado desde el principio como plataforma de data warehousing. IBM Netezza emplea una arquitectura Asymmetric Massively Parallel Processing (AMPP). Un host de multiproceso simétrico⁷ ocupa el frente de una red de nodos de proceso masivo en paralelo (MPP). El appliance de data warehouse de IBM Netezza utiliza esta red MPP para procesar la carga pesada del warehousing y del análisis de datos.

⁷ Los appliances de data warehouse de IBM Netezza tienen dos hosts SMP para redundancia pero sólo uno de ellos está activo a la vez.



Un nodo de la red de appliances de data warehouse de IBM Netezza recibe el nombre de S-Blade (Snippet Blade), un servidor independiente que contiene unidades de procesamiento central (CPUs) de varios núcleos. Cada CPU forma equipo con un Field Programmable Gate Array (FPGA) multimotor y gigabytes de memoria de acceso aleatorio. Puesto que las CPUs tienen su propia memoria, se dedican exclusivamente al análisis de los datos y nunca se ven distraídas por el seguimiento del block ping o de la actividad de congelación de clúster en otros nodos, como ocurre en los sistemas de bases de datos con memoria compartida.

Un FPGA es un chip semiconductor equipado con un gran número de puertas internas programables para implementar casi cualquier función lógica, y es particularmente eficaz en la gestión del streaming de tareas de procesamiento. Fuera del appliance de data warehouse de IBM Netezza, los FPGAs se utilizan en aplicaciones tales como procesos de señales digitales, imagen médica y reconocimiento de la voz. El equipo de ingeniería del appliance de data warehouse de IBM ha creado máquinas de software en los FPGAs de nuestros appliances para acelerar el proceso de los datos antes de que llegue a la CPU. En el interior de cada bastidor de Exadata, Oracle dedica 14 servidores de almacenamiento de ocho vías para conseguir menos de lo que logra el appliance de data warehouse de IBM Netezza con 48 FPGAs incorporados en nuestros

servidores blade. Cada FPGA – al igual que un cuadrado “1x1” de sílice – realiza su trabajo con una gran eficacia, consumiendo poca potencia y generando poco calor.

IBM no tomó un sistema antiguo con limitaciones conocidas y lo equilibró con un nuevo nivel de almacenamiento más inteligente; los appliances de data warehouse de IBM Netezza se han diseñado como plataformas optimizadas para data warehousing.

La comunicación entre nodos en la red MPP de IBM Netezza se produce en un entramado de red que ejecuta un protocolo personalizado basado en IP que utiliza completamente todo el ancho de banda multiseccional y elimina las congestiones incluso con un tráfico de red con ráfagas sostenidas. La red se optimiza para poder escalarse hasta más de mil nodos, permitiendo que cada nodo inicie grandes transferencias de datos simultáneamente a todos los demás nodos. Estas transferencias aportan una gran eficiencia a las tareas de procesamiento típicas del data warehousing y de la analítica avanzada. Al igual que una sentencia

SQL se beneficia de su procesamiento en la arquitectura MPP del appliance de data warehouse de IBM Netezza, también lo hacen los algoritmos computacionales complejos en el núcleo de la analítica avanzada. Las generaciones anteriores de tecnología separan físicamente el procesamiento de aplicación del procesamiento de base de datos, introduciendo ineficiencias y restricciones cuando los conjuntos de datos grandes salen del warehouse para dirigirse a las plataformas de procesamiento analítico y viceversa. El appliance de data warehouse de IBM Netezza lleva el cálculo pesado de la analítica avanzada hasta su red MPP, ejecutando los algoritmos en cada una de las CPUs físicamente cercanas a los datos, haciendo que el movimiento de datos sea redundante y aumentando el rendimiento. Los algoritmos se benefician de la ejecución en los numerosos nodos de la red MPP del appliance de data warehouse de IBM Netezza, liberada de las limitaciones impuestas en sistemas en clúster menos escalables.

A diferencia de la competencia, IBM no tomó un sistema antiguo con limitaciones conocidas y lo equilibró con un nuevo nivel de almacenamiento más inteligente; los appliances de data warehouse de IBM Netezza se han diseñado como plataformas optimizadas para data warehousing. Los appliances de data warehouse de IBM Netezza ofrecen gran rendimiento, facilitando la vida de programadores, administradores y usuarios.

Un cliente de un appliance de data warehouse de IBM Netezza del sector de los servicios financieros utilizaba el enfoque Lean para analizar los gastos de recursos necesarios para gestionar su data warehouse de Oracle. En la construcción y mantenimiento de índices, agregaciones, vistas materializadas y data marts aprendieron que más del 90% del trabajo de su equipo TI se perdía necesariamente o era procesamiento sin valor añadido.

Simplicidad de operación

Simplicidad de operación con Oracle Exadata

Antes de que el warehouse pueda ejecutar consultas debe cargarse con datos. El nivel de almacenamiento de Exadata es una red MPP. Los sistemas MPP tienen un buen rendimiento y se escalan bien cuando todos los nodos participan por igual en la tarea de cálculo que se está realizando. Los datos deben distribuirse por igual, con el mismo volumen de datos relevantes en cada nodo para cada consulta, tanto como sea posible. Para distribuir por igual los datos en la red de servidores de almacenamiento de Exadata, los administradores tienen que estar formados y tener experiencia en el diseño, la gestión y el mantenimiento de particiones complejas, archivos, espacios de tabla, índices, tablas y tamaños de bloques y extensión. “Un sistema sería mucho mejor si no se apoyara mucho en el particionamiento complejo para lograr un buen rendimiento.”⁸

Un cliente de IBM del sector de los servicios financieros utilizaba el enfoque Lean⁹ para analizar los gastos de recursos necesarios para gestionar su data warehouse de Oracle. En la construcción y mantenimiento de índices, agregaciones, vistas materializadas y data marts aprendieron que más del 90% del trabajo de su equipo TI se perdía innecesariamente o era procesado sin valor añadido. El coste de esta pérdida de tiempo se traduce en costes innecesarios de licencias de hardware y software, terabytes de almacenamiento perdido,

desarrollo más largo y ciclos de carga de datos, largos períodos de no disponibilidad de los datos, datos obsoletos, cargas con un rendimiento bajo y excesivos costes administrativos.

Exadata no hace gran cosa para simplificar la gestión de un data warehouse de Oracle. Los administradores deben gestionar varias capas de servidores, cada uno de ellos con imágenes de sistema operativo, firmware, sistemas de archivos y software que deben mantenerse. Oracle sugiere que los DBAs deben prever dedicar 26% menos de tiempo a gestionar 11g, la versión de la base de datos de Exadata, que el que dedican en despliegues anteriores a 10g. Si esto se confirma en la práctica y Exadata reduce en una cuarta parte el tiempo que los clientes pierden en una administración sin valor, Oracle ha dado un paso en la buena dirección. Los appliances de IBM Netezza se han diseñado para no perder el tiempo de los clientes. “El equipo de DBA solamente hace la copia de seguridad del entorno y gestiona

⁸ Curt Monash en <http://www.dbms2.com/2009/09/21/notes-on-the-oracle-database-11g-release-2-white-paper/>

⁹ Con sus raíces en la fabricación, “Lean” es una práctica que utiliza herramientas y técnicas de Six Sigma para analizar el gasto inútil de recursos, así como identificar las actividades que no generan valor en el producto o servicio para su eliminación.

¹⁰ Cliente utilizando Oracle para OLTP y Netezza para data warehousing citado en el foro de Exadata Vs Netezza de LinkedIn en http://www.linkedin.com/groupAnswers?viewQuestionAndAnswers=&gid=2602952&discussionID=11385070&ik=1275353329699&trk=ug_qa_q&goback=.ana_2602952_1275353329699_3_1

el modelo de seguridad de alto nivel para el appliance y esto es todo. No necesitan hacer nada más (por ejemplo, el concepto de indexado es foráneo para ellos en temas relacionados con IBM).¹⁰

Los usuarios de negocio no sólo exigen que las consultas se completen rápidamente, sino que también esperan tener un rendimiento coherente; un informe que ayer se completaba en cinco segundos y hoy en tres minutos seguramente creará una incidencia que requerirá una respuesta del personal del servicio de soporte TI. Los warehouses están inevitablemente sujetos a las demandas de cargas de trabajo dinámicas y variadas. Se cargan los datos procedentes de sistemas OLTP a través de trabajos de proceso por lotes o del goteo de canales de información, las tareas administrativas tales como la copia de seguridad, restauración y limpieza se ejecutan en segundo plano fuera de la vista del negocio y se actualizan constantemente los dashboards. Al mismo tiempo, las aplicaciones con un uso intensivo de cálculo – como las que predicen las reclamaciones que pueden ser fraudulentas o irregulares – crean cargas pesadas y repentinas en la infraestructura del warehouse. Un rendimiento coherente en la empresa crea dos requisitos del warehouse: rendimiento coherente de las consultas y gestión eficaz de la carga de trabajo. Esto simplifica la asignación de potencia de cálculo disponible a todos los trabajos que requieren servicio, habitualmente basada en las prioridades acordadas con la empresa.

La filosofía de la gestión de la carga de trabajo de Oracle es ofrecer a los administradores varios parámetros de ajuste. Los parámetros de Oracle tienen un alto grado de dependencia entre sí, y en Exadata deben definirse algunos de ellos con el mismo valor en todos los procesadores de la red. Esta complejidad obliga a los administradores a cambiar experimentalmente los valores de los parámetros, ajustando sobre la marcha las demandas imprevistas en el warehouse. El logro y el mantenimiento de un rendimiento coherente para grandes comunidades de usuarios, con diferentes requisitos de datos y aplicación, mediante el aumento y disminución de cargas, es una tarea compleja que requiere un alto grado de experiencia en Oracle de los administradores del warehouse.

En los sistemas OLTP con una población estable y bien conocida de transacciones, la empresa puede protegerse de esta complejidad. Los administradores de bases de datos tienen una amplia oportunidad durante la fase de desarrollo de una aplicación, de analizar los requisitos de datos de cada operación y tener el tiempo de diseñar, probar y ajustar la base de datos. Los data warehouses son distintos. Un evento en el mundo exterior crea la necesidad de analizar los datos como no se había intentado hasta ahora. La necesidad de información inmediata no deja tiempo a los administradores para analizar cada consulta y optimizar su recuperación de datos. Un warehouse incapaz de procesar inmediatamente las peticiones, tal como se formulan, niega a la empresa las oportunidades de pasar a la acción.

“La forma en que hicimos una prueba de concepto con ellos [appliance de data warehouse de IBM Netezza] fue: nos enviaron una caja, la colocamos en nuestro centro de datos y la conectamos a nuestra red. Antes de 24 horas ya estaba en pleno funcionamiento. No estoy exagerando, fue así de sencillo.”

Simplicidad de operación con un appliance de data warehouse de IBM Netezza

Los clientes de IBM Netezza confirman que nuestros appliances son sencillos de instalar y utilizar. “La forma en que hicimos una prueba de concepto con ellos [IBM Netezza] fue: nos enviaron una caja, la colocamos en nuestro centro de datos y la conectamos a nuestra red”, dijo. “Antes de 24 horas ya estaba en pleno funcionamiento. No estoy exagerando, fue así de sencillo.”¹¹

Este comentario es del vicepresidente de tecnología de una importante empresa de redes sociales que ya está utilizando bases de datos de Oracle y software RAC.

¹¹ http://www.computerworld.com/s/article/9126930/EHarmony_finds_data_warehouse_match_with_Netezza?source=rss_news

Hay que dejar claras algunas cosas para tener un enfoque sencillo

- SIN monitorización/ajuste (GES y GCS) de interconexión de clúster
 - SIN conocimiento/ajuste específicos de RAC (los DBAs con experiencia en RAC no son muy habituales)
 - SIN configuración y dimensionamiento de dbspace/ tablespace
 - SIN configuración y dimensionamiento de registros rehacer/ físico
 - SIN configuración y dimensionamiento de registros de diario/ lógico
 - SIN configuración y dimensionamiento de página/bloque para tablas
 - SIN configuración y dimensionamiento de extensión para tablas
 - SIN asignación y monitorización de espacio temporal
 - SIN integración de recomendaciones de kernel de sistema operativo
 - SIN mantenimiento de niveles de parche recomendados de sistema operativo
 - SIN sesiones JAD para configurar host/red/almacenamiento
 - SIN sugerencias de consulta (p. ej. first_rows) y optimizador (p. ej., optimizer_index_cost_adj)
 - SIN statspack (estadísticas, aciertos de caché, monitorización de eventos de espera)
 - SIN ajuste de memoria (SGA, búffers de bloques, etc.)
 - SIN planificación/creación/mantenimiento de índices
 - Estrategias de particionamiento sencillo: HASH o ROUND ROBIN
-

La reducción del tiempo para ser productivo es un buen punto de partida; la filosofía de IBM es la de llevar simplicidad a todas las fases del data warehousing. La primera tarea a la que se enfrenta un cliente es la carga de los datos. Un appliance de data warehouse de IBM Netezza automatiza la distribución de datos. La experiencia de los proyectos de prueba de concepto es que los clientes cargan sus datos en un appliance de data warehouse de IBM Netezza utilizando la distribución automática, ejecutan consultas y comparan los resultados con los de los entornos de Oracle altamente ajustados. Para todas las consultas excepto las más sencillas, la distribución automática es suficiente para que un appliance de data warehouse de IBM Netezza supere a Oracle. Posteriormente los clientes pueden analizar todas sus consultas para identificar las que se pueden acelerar mediante la redistribución de los datos en diferentes claves. El appliance de data warehouse de IBM Netezza facilita esta tarea.

Todas las consultas remitidas al appliance de data warehouse de IBM Netezza se procesan automáticamente en su red masiva en paralelo sin la intervención de los administradores de bases de datos. Las consultas y los análisis entran en el appliance de data warehouse de IBM Netezza a través de la máquina host en la que el optimizador, el compilador y el planificador las descomponen en muchas partes o fragmentos distintos, y distribuyen estas instrucciones

a la red MPP de nodos de procesamiento, o S-Blades, y todos ellos pasan a procesar simultáneamente sus cargas de trabajo en la porción de datos gestionados localmente.

Un fragmento que llega a cada uno de los S-Blades del appliance de data warehouse de IBM Netezza inicia la lectura de datos comprimidos del disco a memoria. A continuación, el FPGA lee los datos de los búffers de memoria y utiliza su motor de compresión para descomprimirlos, transformando instantáneamente cada bloque del disco en el equivalente de 4-8 bloques de datos en el FPGA. La ingeniería que se esconde detrás del appliance de data warehouse de IBM Netezza acelera el componente más lento de cualquier data warehouse: el disco. A continuación, en los flujos de datos de FPGA que entran en el motor de proyecto, el cual filtra las columnas en base a los parámetros especificados en la cláusula SELECT de la consulta SQL que se está procesando. Sólo los registros que cumplen la cláusula SELECT se pasan al motor de restricciones, en el que se bloquea mediante puertas el paso de las filas no necesarias para procesar la consulta, en base a las restricciones especificadas en la cláusula WHERE. El motor de visibilidad mantiene la compatibilidad ACID (Atomicity (atomicidad), Consistency (coherencia), Isolation (aislamiento) y Durability (durabilidad)) con velocidades de streaming. Todo este trabajo, la depuración constante de columnas y filas no necesarias, se lleva a cabo en un FPGA con eficiencia energética que mide tan sólo 6 cm cuadrados. Si la ingeniería en la que se basa el appliance de data warehouse de IBM Netezza no necesita mover datos, no lo hace.

Una vez se completa el preprocesamiento del FPGA, éste devuelve sólo el conjunto de registros recortado resultante a la memoria del S-Blade en la que la CPU realiza operaciones de base de datos de nivel superior como las clasificaciones, joins y agregaciones, haciéndolo en paralelo con las otras CPUs de la red MPP. La CPU también puede aplicar algoritmos complejos incorporados en el código del fragmento para su procesamiento con analítica avanzada. Por último, la CPU ensambla todos los resultados intermedios del flujo de datos y produce un resultado para el fragmento, que se envía a través del entramado de red hasta otros S-Blades o el host, según indique el código del fragmento. Cuando los datos necesarios para un JOIN no son contiguos en un nodo, el entramado de red internodal del appliance de data warehouse de IBM Netezza redistribuye con eficacia y sencillez en la parte final del ciclo de procesamiento, después de que la base de datos haya completado las restricciones y las proyecciones. Algunos algoritmos muy complejos requieren la comunicación entre nodos para calcular la respuesta. La ingeniería del appliance de data warehouse de IBM Netezza utiliza una interfaz de paso de mensajes para comunicar los resultados provisionales y producir el resultado final.

Y, puesto que los bloques de datos comprimidos originales siguen estando en memoria, se pueden reutilizar automáticamente en consultas posteriores que requieran datos similares, a través del caché de tabla del

Tan sólo tres meses después de pasar a un appliance de data warehouse de IBM Netezza, un cliente explica que su equipo ofreció más aplicaciones analíticas que las que podían suministrar en los tres años anteriores con Oracle.

appliance de data warehouse de IBM Netezza – un mecanismo automatizado que no requiere formación o implicación del DBA.

Puesto que el appliance de data warehouse de IBM Netezza aplica un paralelismo completo en todas las tareas, su sistema de gestión de cargas de trabajo juega un papel crítico en el control del número de recursos de cálculo del appliance que se pone a disposición de cada uno de los trabajos. En la arquitectura del appliance de IBM, un componente de software controla todos los recursos del sistema: procesadores; discos; memoria; red. Esta elegancia es la base del sistema de gestión de cargas de trabajo del appliance de data warehouse de IBM Netezza. El sistema de gestión de cargas de trabajo del appliance de data warehouse de IBM Netezza facilita a los administradores la asignación de recursos de cálculo a los usuarios y los

grupos, en base a prioridades acordadas con la empresa y mantienen tiempos de respuesta coherentes para varias comunidades.

Los appliances de data warehouse de IBM Netezza eliminan el trabajo perdido del ajuste de bases de datos. Equipados para tomar sus propias decisiones inteligentes, los appliances de IBM Netezza no requieren ajustes y poca administración del sistema. Las pocas tareas administrativas necesarias para mantener un rendimiento coherente mediante cargas de trabajo dinámicas y cambiantes están al alcance de cualquiera con experiencia en Linux y SQL. Todo lo que el administrador debe hacer es asignar los recursos del appliance de data warehouse de IBM Netezza a los grupos de la comunidad de usuarios y ceder el control al sistema de gestión de cargas de trabajo. Liberado de los constantes ciclos de administración de bases de datos, el personal técnico participa con la empresa para investigar nuevas formas de creación de valor en la utilización de los datos. Tan sólo tres meses después de pasar a un appliance de data warehouse de IBM Netezza, un cliente explica que su equipo ofreció más aplicaciones analíticas que las que pudieron suministrar en los tres años anteriores con Oracle. El procesamiento de aplicaciones analíticas cercanas a donde se gestionan los datos, utilizando la misma plataforma MPP que la utilizada para procesar SQL, representa una oportunidad real para que las organizaciones incrementen radicalmente el valor que generan a partir de los datos.

Valor

Valor con Exadata

Como pone de manifiesto el análisis de tiempo perdido llevado a cabo por el cliente de servicios financieros tanto del appliance de data warehouse de IBM Netezza como de Oracle, el uso de Oracle para el data warehousing consume recursos laborales. Oracle sugiere que la última versión de su sistema de gestión de bases de datos puede reducir esta pérdida en un 26%.¹² Los clientes del appliance de data warehouse de IBM Netezza atestiguan que estas tareas de administración de bajo nivel y técnicamente exigentes son sencillamente innecesarias; en este sentido no es defendible que la operación de una base de datos de Oracle exija a los administradores dedicar la mayor parte de su tiempo en el cuidado de la tecnología subyacente, mientras que los clientes del appliance de data warehouse de IBM Netezza dedican dicho tiempo a crear valor utilizando los datos.

El nuevo nivel de almacenamiento de Exadata añade otra capa de complejidad que los administradores deben ajustar y gestionar. Puesto que Exadata es muy nuevo y muy pocos data warehouses que utilizan esta tecnología están en producción, las proyecciones de sus costes de propiedad son prematuras. No obstante, los clientes pueden esperar que para lograr un alto rendimiento coherente de Exadata incurrirán en costes substanciales en el diseño y la administración de las bases de datos.

Aunque la incorporación de un nuevo nivel de almacenamiento elimina el cuello de botella del rendimiento del disco en la base de datos de Oracle, la ingeniería de Exadata está más adaptada al procesamiento masivo en paralelo que la plena

utilización de la arquitectura. La incapacidad de Oracle de integrar la gestión de datos totalmente en el nivel de almacenamiento de Exadata implica que se pide poco al hardware en su red MPP. Esto hace aumentar el coste de adquisición de Exadata; los clientes pagan por hardware que nunca utilizarán completamente y pagan por una nueva capa adicional de software de almacenamiento con capacidades limitadas. Estos costes aumentan durante la vida útil del warehouse. Los clientes pagan por un espacio infrutilizado en sus centros de datos que retornaría un mayor valor si se utilizara para alojar un sistema informático más eficiente.

Aunque los costes socavan el valor, una pregunta fundamental es si Exadata ayuda a los clientes a crear valor. Los data warehouses de primera generación juegan un papel importante en mantener informada a la organización del pasado reciente, aunque los datos liberan un mayor potencial mediante la analítica avanzada y otras capacidades de los data warehouses de segunda generación, como se ha comentado antes en el artículo. Oracle RAC junto con el almacenamiento tradicional ha tenido un éxito técnico limitado en esta área y aún no ha demostrado tener éxito en este rol hasta la fecha. El nivel de almacenamiento de Exadata no puede procesar uniones complejas, agregaciones distintas y funciones analíticas. Es difícil imaginarse cómo dos tecnologías, individualmente mal equipadas para analizar en profundidad conjuntos de datos muy grandes con un alto rendimiento, conseguirán esta hazaña cuando se conecten mediante una red rápida y se alojen en el mismo bastidor.

Valor con el appliance de data warehouse de IBM Netezza

Los ingenieros de los appliances de data warehouse de IBM Netezza integran la gestión de los datos y el análisis en el interior de redes masivas en paralelo que no comparten nada. Un resultado que tenemos previsto lograr de esta innovación es la simplicidad para nuestros clientes, que se traduce directamente en unos costes radicalmente más bajos de posesión y operación de data warehouses que los posibles con productos de bases de datos tradicionales, como por ejemplo la de Oracle.

Las demandas en los data warehouses se han desplazado más allá del procesamiento de SQL sencillo; para utilizar totalmente los datos el warehouse debe ser capaz de ejecutar modelos predictivos, gráficos de investigación y otras aplicaciones analíticas. Como ilustración, una empresa de servicios financieros – sabiendo que la siguiente compra más probable por parte de una familia que recientemente ha solicitado una hipoteca y anteriormente ha adquirido productos de inversión, productos de préstamo y una cuenta corriente pero que nunca ha adquirido pólizas de seguros, es un producto de inversión seguido de otra hipoteca – puede crear campañas de marketing de valor para el cliente y con una alta probabilidad de éxito.¹³

¹² <http://www.dbms2.com/2009/09/21/notes-on-the-oracle-database-11g-release-2-white-paper/>

¹³ Consulte las redes bayesianas dinámicas para la adquisición de patrones de análisis: aplicación de venta cruzada de servicios financieros por Anita Prinzie, Marketing Group, Manchester Business School y Dirk Van den Poel, Department of Marketing, Ghent University.

Evaluación de los sistemas		IBM Netezza	Oracle
Elemento	IBM Netezza 1000	Exadata v2 (SAS)	
Rendimiento y arquitectura	MPP	<ul style="list-style-type: none"> MPP Real Optimizado para data warehousing y analítica 	<ul style="list-style-type: none"> Híbrido – nodos de almacenamiento en paralelo y nodo de cabecera de clúster SMP Arquitectura generalizada
	Arquitectura de hardware	<ul style="list-style-type: none"> S-Blades de procesamiento completo (1 núcleo CPU + 1 núcleo FPGA / 1 unidad de disco) Nodo host SMP utilizado principalmente para interfaz de usuario/ aplicaciones Redistribución blade-blade independiente 	<ul style="list-style-type: none"> Almacenamiento inteligente (1 núcleo CPU / 1,5 unidades de disco) Nodos de clúster SMP ejecutando Oracle 11g RAC InfiniBand (nodos Exadata a clúster SMP) Participación de nodo de cabecera en todas las redistribuciones de datos
	Streaming de datos	<ul style="list-style-type: none"> El rendimiento de FPGA ayuda a S-Blade – descompresión, filtrado de predicados, aplicación de seguridad a nivel de fila >9% de trabajo realizado en los S-Blades 	<ul style="list-style-type: none"> Nodos Exadata utilizados principalmente para descompresión y filtrado de predicados Gran parte del trabajo de DW y analítica realizado en el nodo de cabecera de SMP
	Analítica en base de datos	<ul style="list-style-type: none"> Plataforma MPP totalmente dedicada a la analítica Funciones definidas por el usuario, agregaciones y tablas Soporte de lenguaje: C/C++, Java, Python, R, Fortran Soporte de paradigma: SQL, Matrix, Grid, Hadoop Conjunto incorporado de >50 analíticas clave (totalmente paralelizadas) Entorno de desarrollo integrado: Eclipse y R GUI con asistentes 	<ul style="list-style-type: none"> Procesamiento analítico limitado sólo al clúster de nodo de cabecera Funciones definidas por el usuario y agregaciones Soporte de lenguaje: C/C++, Java Soporte de paradigma: SQL, Matrix (menor) Funciones analíticas básicas
	Escala	<ul style="list-style-type: none"> Rendimiento lineal y escalabilidad de tamaño de datos Gestión de cargas de trabajo empresarial y totalmente funcional y otras funciones 	<ul style="list-style-type: none"> Rendimiento no lineal y escalado del tamaño de datos – rendimiento y cuello de botella e/s en el clúster de nodos de cabecera
Simplicidad	Gestión e integración de sistemas de appliance	<ul style="list-style-type: none"> Sin ajustes, sin indexado, sin particiones Sistema equilibrado desarrollado para ofrecer el mejor precio-rendimiento 	<ul style="list-style-type: none"> Gran dependencia del rendimiento al ajuste El rendimiento depende de las habilidades en el diseño de bases de datos físicas, incluidos los índices y las particiones

Conclusión

Este análisis, que queda fuera de las capacidades de SQL, requiere una técnica llamada Redes bayesianas dinámicas. No obstante, el análisis utiliza los mismos datos procesados por SQL para crear informes y dashboards, lo que sugiere un rol ampliado del warehouse.

Los appliances de data warehouse de IBM Netezza se han diseñado desde el principio para procesar tanto SQL como las aplicaciones de la analítica avanzada. Los appliances de data warehouse de IBM Netezza liberan a los clientes de los lenguajes propietarios. Los clientes pueden portar aplicaciones existentes a los appliances de data warehouse de IBM Netezza o también pueden optar por desarrollar nuevas aplicaciones analíticas en el lenguaje que prefieran, incluidos R, C/C++, Java, Python y Fortran. Los clientes pueden utilizar una biblioteca incorporada de algoritmos paralelizados en base de datos, incluida la preparación de datos, la minería de datos, la analítica predictiva, geoespacial y la álgebra de matrices. Adicionalmente, los clientes pueden elegir trabajar con Hadoop / MapReduce como, por ejemplo, un mecanismo de ingestión altamente escalable para preprocesar grandes conjuntos de datos generados por aplicaciones web de interfaz pública y registros web, antes de cargarlos en los appliances de data warehouse de IBM Netezza para el análisis bajo demanda.

Los appliances de data warehouse de IBM Netezza surgen como principal alternativa de Oracle para el data warehousing. El paso de data warehouses y marts de Oracle a un data warehouse de IBM Netezza crea nuevas oportunidades, no riesgos. La mayoría de clientes de IBM ya han recorrido este camino, muchos de ellos asociándose con empresas de integración de sistemas con un sólido historial de migraciones con éxito. Exadata es una evolución de la plataforma OLTP de Oracle y se posiciona como plataforma de propósito general tanto para OLTP como para la analítica. El sistema de gestión de bases de datos de Oracle se ha diseñado para OLTP, en el que los volúmenes de datos son relativamente modestos en comparación con los data warehouses. La actividad de base de datos de un sistema OLTP se puede evaluar antes de ponerlo en producción; los administradores tienen tiempo para diseñar, probar y optimizar la recuperación de datos de cada transacción. Los data warehouses deben procesar inmediatamente cualquier consulta que la empresa deba formular a sus datos; las tecnologías que requieren la mediación de un administrador no son adecuadas para esta tarea. El constreñimiento de esta tecnología a un rol distinto al del procesamiento de transacciones produce una gran tensión en las personas y en los procesos encargados de gestionar y operar un data warehouse.

“Este [appliance de data warehouse de IBM Netezza] es el primer producto de base de datos con una hoja de ruta de producto a largo plazo que se alinea perfectamente con nuestra hoja de ruta. A esto lo llamamos nuestra base de datos bajo demanda.”

— Director de Datos, Large Equities Exchange Group

Oracle avisa a los clientes que Exadata tiene una arquitectura similar a la de los appliances de data warehouse de IBM Netezza pero mejor, porque los appliances de data warehouse de IBM Netezza no admiten todos los tipos de datos o estándares SQL, y porque no soporta la minería de datos o la alta concurrencia. Los clientes de los appliances de data warehouse de IBM Netezza discrepan: “Este [appliance de data warehouse de IBM Netezza] es el primer producto de base de datos con una hoja de ruta de producto a largo plazo que se alinea perfectamente con nuestra hoja de ruta. A esto lo llamamos nuestra base de datos bajo demanda”¹⁵, dijo el director de datos de un gran grupo de intercambio de capitales.

¹⁵ www.netezza.com/customers/nyse-euronext-video.aspx

dadas las características diferentes de sus cargas de trabajo, pocos clientes intentan ejecutar sistemas OLTP y data warehouse en la misma infraestructura; para ello se necesita un constante ajuste y optimización. Los técnicos están en una posición difícil: o aceptan un rendimiento comprometido tanto para OLTP como para el data warehouse, o reconfiguran incesantemente la base de datos en un intento vano de satisfacer las demandas conflictivas de las diferentes cargas de trabajo.

Como se ha mencionado anteriormente en este libro electrónico, las organizaciones que evolucionan hacia los data warehouses de segunda generación ejecutan sus sistemas OLTP y warehouse en distintas plataformas, cada una de ellas configurada específicamente para las necesidades de sus cargas de trabajo.

El único data warehouse que realmente importa es su data warehouse – las aplicaciones que se ejecutan en sus datos en su centro de datos. Una prueba de concepto (PoC) in situ crea la oportunidad de que un departamento TI investigue totalmente una tecnología, aprendiendo a utilizar los appliances de data warehouse de IBM Netezza de modo que ayuden a sus compañeros de empresa a extraer más valor de los datos. Para aprovechar al máximo esta oportunidad es necesario gestionar la PoC con la misma disciplina dedicada a otros proyectos.

Curt Monash ofrece un sabio consejo en su blog “Best practices for analytic DBMS POCs”¹⁶, que incluye la implicación de un consultor independiente para conducir el proyecto hasta un resultado exitoso. Para las organizaciones cualificadas que deseen conocer el rendimiento de sus warehouses en un appliance de data warehouse de IBM Netezza, sin coste ni riesgo alguno, IBM ofrece su TestDrive. Para reservar uno, visite:

ibm.com/software/data/netezza/

Appliances de data warehouse de IBM Netezza: utilizarlos es disfrutarlos.

¹⁶ www.dbms2.com/2010/06/14/best-practices-analytic-database-poc/ffmore-2297

Transmítalo

Comparta este documento con sus amigos y compañeros de trabajo.

Démos su opinión

¿Qué piensa de las ideas y los argumentos de este libro electrónico? Nos gustaría saber qué le ha gustado, qué no le ha gustado o qué querría comentar más.

- Converse con la Comunidad de data warehouses de IBM Netezza:

www.enzeecommunity.com/groups/

Póngase en contacto

- Visite los blogs de IBM Netezza:

<http://thinking.netezza.com>

- Visite la página web de IBM:

ibm.com/software/data/netezza/

- Visite la página web de la Comunidad de data warehouses de IBM Netezza:

www.enzeecommunity.com

Acerca de los autores

Phil Francisco, Vicepresidente, Product Management y Product Marketing de IBM

Phil Francisco cuenta con más de 20 años de experiencia en desarrollo de tecnología y marketing global de tecnología. Como vicepresidente de Product Management y Product Marketing en IBM, acoge nuevas estrategias de producto y de negocio, dirige la cartera de productos e impulsa programas de marketing de producto. Antes de IBM, Francisco era Vicepresidente de Marketing en PhotonEx, líder en el desarrollo de sistemas de transporte óptico de 40 Gb/s para proveedores de redes de telecomunicaciones básicas. Antes de PhotonEx, Francisco ocupó el cargo de Vicepresidente de Product Marketing del Optical Networking Group de Lucent Technologies, donde trabajó con algunas de las operadoras de telecomunicaciones más importantes del mundo en la planificación e implementación de soluciones de redes ópticas. El Sr. Francisco es propietario de una patente en arquitecturas de redes ópticas avanzadas. Se licenció en Ingeniería Eléctrica y Ciencias Informáticas con el grado de magna cum laude, en la Escuela Moore de Ingeniería Eléctrica en la Universidad de Pennsylvania. Obtuvo el grado de Máster en Ingeniería Eléctrica en la Universidad de Stanford y realizó el Programa de Administración Avanzada en la Escuela de Negocios de Fuqua en la Universidad de Duke. Lea el blog de Phil: http://thinking.netezza.com/blogs_by/phil-francisco

Mike Kearney, Senior Director, Product Marketing de IBM

Mike ha trabajado en las tecnologías de la información desde hace más de 25 años. En IBM se ha especializado en la comunicación de valor del rendimiento, el valor y la simplicidad del appliance de Netezza, a las organizaciones que buscan generar un mayor retorno de sus datos. Mike ha trabajado anteriormente en Vignette, BMC Software y Oracle Corporation y en empresas de los sectores de los servicios financieros, telecomunicaciones y energía. Mike es licenciado en ciencias por la Universidad de Londres y doctor en ciencias por la Universidad de Coventry.

Acerca de Netezza, empresa de IBM

Netezza, empresa de IBM, es un líder global en appliances de data warehouse, analíticos y de monitorización, que simplifican radicalmente la analítica de alto rendimiento en toda una empresa. La tecnología de Netezza permite a las organizaciones procesar un gran volumen de datos capturados con una velocidad excepcional, proporcionando una ventaja competitiva y operacional significativa en los sectores actuales que hacen un uso intensivo de datos, como los medios de comunicación digitales, energía, servicios financieros, gobierno, sanidad y ciencias de la vida, comercio y telecomunicaciones.

Acerca de IBM Data Warehousing y Soluciones de Analítica

IBM proporciona el portafolio más amplio y completo de software, hardware y soluciones de data warehousing, gestión de la información y analítica de negocio para ayudar a los clientes a maximizar el valor de sus activos de información y descubrir nuevos conocimientos para tomar decisiones mejores y más rápidas, así como optimizar el resultado de la empresa.

Para más información

Si desea obtener más información acerca de IBM Data Warehousing y Soluciones de Analítica, póngase en contacto con su representante de ventas de IBM o visite: ibm.com/software/data/netezza/

IBM España, S.A.
Tel.: 900 100 400
c/Sta. Hortensia, 26-28
28002 Madrid
España

La página de presentación de IBM puede encontrarse en
ibm.com/es

IBM, el logotipo de IBM, ibm.com y Netezza son marcas registradas de International Business Machines Corporation en Estados Unidos y/o en otros países. Si éstas y otras marcas registradas de IBM son resaltadas en su primera aparición en esta información con un símbolo de marca registrada (® o ™), estos símbolos indican marcas registradas en EE.UU. o marcas registradas según el derecho constitucional de propiedad de IBM en el momento que se publicó esta información. Estas marcas registradas también pueden ser marcas registradas o marcas registradas según el derecho consuetudinario en otros países. Encontrará una lista actualizada de las marcas registradas de IBM en la Web en “Información de copyright y marcas registradas” en:

ibm.com/legal/copytrade.shtml

Microsoft y SQL Server son marcas registradas de Microsoft Corporation en Estados Unidos y/o en otros países.

Oracle y Exadata son marcas registradas de Oracle y/o sus filiales.

Los nombres de otras empresas, productos y servicios pueden ser marcas comerciales o marcas de servicio de terceros.

© Copyright IBM Corporation 2012



Por favor, recicle