



No deje que el almacenamiento consuma el Retorno de Inversión de su proyecto de nube.

*Autor: George Crump, Analista Senior
Fecha: 17/05/2012*

La inversión que una organización realiza en su iniciativa de virtualización o nube es significativa, pero también lo es el retorno sobre la inversión (return on investment, ROI) que estos proyectos ofrecen. El desafío es que el costo de proporcionar la Infraestructura de Almacenamiento para tales iniciativas puede ser alto y rápidamente consumir todo el ROI que se hubiera obtenido.

Una inversión en la nube no se hace para justificar la compra de un Sistema de Almacenamiento más costoso; se hace para permitir que la tecnología de la información (TI) pueda proporcionarse como un servicio y tenga la agilidad necesaria para satisfacer las demandas del negocio. Se necesitan nuevos enfoques en almacenamiento para evitar que se consuma el ROI de nube.

Para evitar que el almacenamiento no carcoma el ROI de los proyectos de nube, se debe proporcionar el rendimiento requerido por la infraestructura y al mismo tiempo las capacidades requeridas en la forma más eficaz y eficiente en función de costos posible. Los requisitos de rendimiento de una infraestructura de nube son distintos de los requisitos de los servidores individuales y de las aplicaciones con altas exigencias de rendimiento, como bases de datos. No solo debe manejar altibajos en las cargas de entrada/salida que provienen de las aplicaciones de bases de datos virtualizadas, sino que además debe manejar cargas I/O estables pero aleatorias provenientes de la gran cantidad de servidores virtualizados.

La Infraestructura de Almacenamiento debe contar con una serie de atributos de eficiencia de almacenamiento para minimizar la cantidad de capacidad física requerida por el entorno virtual que crece a ritmo acelerado. Estos atributos en general incluyen la asignación dinámica de espacio de volumen (thin provisioning o aprovisionamiento ligero), writable snapshots o instantáneas escribibles para que las máquinas virtuales puedan partir de una fuente común (clonación) y técnicas de eficiencia de espacio que reducen el consumo de capacidad de los conjuntos de datos reales que están siendo implementados (de-duplicación o compresión).

Los atributos agregados al Sistema de Almacenamiento para mejorar la eficiencia no deberían dañar el rendimiento. El Sistema de Almacenamiento debe encontrar el equilibrio justo entre alto rendimiento y uso de la capacidad de almacenamiento de la forma más eficiente posible. Para ello, se requiere un enfoque diferente, en el cual es posible utilizar múltiples tecnologías de almacenamiento.

IBM Storwize V7000 es un ejemplo excelente de una Solución de Almacenamiento apta para esta tarea. Al apalancar virtualización de almacenamiento y aplicar un alto grado de automatización, elimina gran parte de la carga que implica manejar el rendimiento y los costos. El Sistema Storwize V7000 es un componente clave de la Estrategia Smarter Storage de IBM, que a su vez forma parte de la Estrategia Smarter Computing de IBM. El Storwize V7000 ofrece altos niveles de automatización o inteligencia para permitir a los administradores enfocarse en otros desafíos.

Rendimiento con eficiencia de costos

El entorno de nube representa un cambio operativo importante en comparación con las épocas en que se estilaba implementar un servidor por cada aplicación utilizada. Ahora cada servidor aloja múltiples máquinas virtuales, todas con distintos patrones de I/O de almacenamiento. Significa un caudal constante de I/O muy aleatoria, un patrón que presenta grandes desafíos para los medios de disco duro giratorio.

La implementación de Sistemas de Almacenamiento de Alto Rendimiento, a menudo basados en tecnología de estado sólido, promete consumir el ROI que originalmente ofrece un proyecto de nube incluso con mayor rapidez. La situación exige una implementación híbrida de tecnología de estado sólido y disco duro. Sin embargo, si estos Sistemas de Almacenamiento no están correctamente implementados, pueden consumir tiempo adicional de administración e incluso generar la necesidad de un "Especialista del Rendimiento" a tiempo completo, que supervise el entorno de nube y mueva las máquinas virtuales a distintos niveles de almacenamiento en función de los cambios en la demanda de rendimiento.

En vista del costo asociado al Almacenamiento de Unidad de Estado Sólido (Solid State Drive), la inversión en el dispositivo flash debe utilizarse al máximo aprovechando toda la capacidad disponible. No puede utilizarse a medias como el almacenamiento en disco duro tradicional. Además, no es posible almacenar cualquier tipo de datos en el dispositivo. Es necesario almacenar la cantidad máxima de los tipos de datos más apropiados para asegurar que la utilización llegue al máximo de su capacidad en la forma más eficiente en función de costos posible.

Es posible mover manualmente los datos a un área de Almacenamiento Flash y algunos entornos de hipervisor como VMware que permiten la migración en vivo de imágenes de servidores virtualizados entre distintos tipos de almacenamiento en disco. Pero este proceso a menudo no es práctico porque los conjuntos de datos son demasiado dinámicos, lo cual aumentaría las probabilidades de que los datos incorrectos se encuentren en el lugar incorrecto en el momento incorrecto.

También puede considerarse el costo de administrar un sistema separado que no esté integrado con el Sistema de Almacenamiento original. Muchos clientes preferirán comprar un sistema individual de un proveedor alternativo para satisfacer la demanda de rendimiento. Estos sistemas no pueden integrarse entre sí y nuevamente, mientras que algunos hipervisores como VMware permiten migrar máquinas virtuales entre sistemas disímiles, sigue habiendo desafíos en la administración proactiva de estos dispositivos independientes.

IBM Storwize V7000 aborda ambos problemas con la tecnología IBM Easy Tier y la virtualización del almacenamiento. Esta característica, también disponible en el IBM DS8000 y en el SAN Volume Controller, monitorea automáticamente los datos almacenados por el hipervisor. Conforme la frecuencia de acceso de ciertos conjuntos de datos aumenta, es posible transferir esos datos automáticamente a la capa de almacenamiento flash. Easy Tier mantendrá la capa flash llena con los datos intensivos en I/O para que sean servidos desde la capa que ofrece el mayor rendimiento.

La inversión en SSD se maximiza porque la capa flash ahora está llena de los datos más apropiados posibles. Easy Tier seguirá analizando dinámicamente todos los datos, para que, conforme los datos en la capa del disco duro adquieran una intensidad de I/O mayor a la de los datos en el nivel flash, puedan sustituirse. Con este proceso, se asegura que el almacenamiento premium basado en flash quede reservado para los datos que aprovechan su alto rendimiento y que la inversión premium se utilice al máximo.

Puede haber una combinación de dispositivos de almacenamiento que integren un mismo volumen, pero tanto la infraestructura de nube como los administradores de almacenamiento manejan una sola entidad lógica. El Sistema Storwize V7000 integra en forma transparente los distintos tipos de medios detrás de escena y coloca los datos en el medio más adecuado para satisfacer sus requisitos de rendimiento.

El proceso Easy Tier está diseñado para funcionar automáticamente en el fondo. Esto permite al administrador de almacenamiento concentrarse en otras tareas, en lugar de tener que cuidar que se cumplan los requisitos de rendimiento de las máquinas virtuales de alta exigencia. Una tecnología por capas como esta preserva el ROI de la implementación del servidor virtual, ya que libera capital humano y no requiere un "especialista en rendimiento", pues el Storwize V7000 cumple la función del especialista. El retorno sobre la inversión de la implementación del proyecto de nube se preserva aún más, minimizando la cantidad de almacenamiento de estado sólido de alto precio requerido para satisfacer las exigencias de rendimiento.

Esta automatización no solo preserva el ROI de los proyectos de nube, sino que en realidad puede extenderlo. Una tecnología automatizada de colocación de datos también permite implementaciones incluso más densas de máquinas virtuales por sistema host, dado que el rendimiento del Sistema de Almacenamiento puede reaccionar en forma automática a las cargas de trabajo adicionales. Si se necesita mayor rendimiento, la tecnología Easy Tier brinda información acerca de cuánto almacenamiento de estado sólido debe implementarse, lo cual evita gastos excesivos.

Eficiencia de almacenamiento con eficacia de costos

El análisis de la mayoría de los entornos de nube muestra que la mayor parte de los conjuntos de datos se sirven mejor desde una capa de disco mecánica porque se accede a ellos con muy poca frecuencia. De hecho, hasta 90% de los datos no están activos a diario. Aunque la tecnología de disco duro ya es de bajo costo, debería buscarse una mayor optimización de esa inversión para maximizarla. La mayor eficiencia del almacenamiento no solo ahorra gastos de capital sino también espacio en el centro de cómputo y en los bastidores, en comparación con el gran consumo de energía que supone tener un desfile interminable de nuevas unidades de disco. Aunque los discos fueran gratuitos, el costo de alojar, alimentar y refrigerar esos discos sería significativo.

Sin embargo, no es sólo la capacidad del disco duro la que puede aprovechar esta eficiencia. Si puede ser optimizada, la capa de estado sólido ofrece un retorno sobre la inversión aún mayor, ya que la tasa de costo por GB es mayor. Si la inversión para comprar suficiente tecnología de estado sólido para almacenar 10% de los datos puede estirarse hasta alojar efectivamente 20% de los datos, esto se traducirá en una importante suba del rendimiento a un conjunto de datos más amplio, eliminando virtualmente el acceso a los datos que no se encuentren en la capa de estado sólido; un suceso denominado "tier miss".

El proceso de eficiencia

Hay un proceso de tres pasos que los Sistemas de Almacenamiento como el Storwize V7000 pueden ofrecer en un proyecto de nube. En primer lugar, los volúmenes pueden aprovisionarse en forma ligera, luego pueden clonarse y finalmente, el conjunto de datos restante debe optimizarse. Idealmente, el sistema de almacenamiento debe estar en condiciones de ofrecer una pila completa de estas funciones, pero la optimización de la capacidad restante es potencialmente el paso más importante de todos.

Aprovisionamiento ligero

Aunque este término se ha convertido en un concepto universal, es necesario entender todos sus matices. El aprovisionamiento ligero es el proceso de permitirle a un sistema host físico o a una máquina virtual que "piense" que tiene la capacidad que requiere pero solo asignar esa capacidad a medida que la necesita. En suma, el espacio en disco se asigna en forma dinámica al instante, a medida que el sistema host o la máquina virtual lo necesitan. Esta operación de escritura dinámica requiere que el Sistema de Almacenamiento tenga suficiente desempeño para asignar y escribir los datos entrantes con un impacto mínimo en el rendimiento de la aplicación host. Storwize V7000 es un ejemplo de un sistema que posee esa capacidad.

Otra opción es utilizar las capacidades del hipervisor para hacer su propio aprovisionamiento ligero. El desafío es la actividad de escritura extra que una función de asignación dinámica requerirá y la carga que colocará en el hipervisor de la nube y su sistema de archivos. Gran parte de este problema de rendimiento puede resolverse utilizando una capa de almacenamiento en estado sólido y nuevamente, un producto como IBM Easy Tier reduciría gran parte del problema de rendimiento.

Clonación / Instantáneas escribibles

El segundo paso consiste en permitir el uso de la clonación o imágenes instantáneas escribibles. Como el aprovisionamiento ligero, esto puede hacerse a través del Sistema de Almacenamiento o el hipervisor de nube, atribuyendo la debida importancia a las mismas preocupaciones sobre rendimiento. Esta técnica genera una imagen maestra de un servidor virtual y luego la utiliza repetidas veces. Por ejemplo, una colección de servidores de base de datos puede tener el mismo sistema operativo y software de base de datos. Puede crearse una imagen maestra y luego usar una instantánea para crear cada servidor subsiguiente con detalles específicos para cada máquina virtual que se mantiene en la porción escribible de cada instantánea.

Optimización de la capacidad

El aprovisionamiento ligero libera la capacidad que fue asignada a un servidor o máquina virtual en particular pero que no se encuentra en uso. Es importante porque se trata de espacio libre que no puede ser operado por ninguna otra técnica de optimización de la capacidad. La clonación y las instantáneas escribibles eliminan gran parte de la redundancia que se ve en una infraestructura en nube. Se supone que la optimización de la capacidad baja el impacto en el conjunto de datos restante.

La compresión y la de-duplicación son dos técnicas comunes utilizadas para optimizar la capacidad. La compresión elimina información redundante desde adentro de un archivo, mientras que la de-duplicación elimina información redundante entre todos los archivos. La compresión puede ofrecer eficiencia en cada archivo, mientras que la de-duplicación solo puede ofrecer optimización si existen correlaciones de identidad entre archivos.

La de-duplicación tiene un papel importante en las técnicas de resguardo de la empresa en donde el nivel de redundancia de datos es intencionadamente alto. Pareciera que tiene el mismo valor en el entorno virtual excepto cuando se utiliza un Sistema de Almacenamiento robusto como Storwize V7000. Estos sistemas pueden eliminar una gran parte de los datos redundantes a través del uso de clonación e instantáneas. La mayoría de los datos que permanecen son únicos. Como resultado, las tecnologías de compresión como IBM Real-Time Compression, pueden proporcionar un retorno sobre la inversión mayor porque funcionan en todos los archivos, no solo en los redundantes.

La compresión en tiempo real es ideal para la infraestructura de nube

El IBM Storwize V7000 proporciona una versión totalmente integrada de la tecnología Real-Time Compression de IBM. Esta tecnología está diseñada específicamente para cargas de trabajo de almacenamiento y permite la optimización de cargas activas primarias de alto rendimiento. En la infraestructura de nube, un gran conjunto de datos activos es la norma. Como las máquinas virtuales y los sistemas que las alojan son muy activos, toda estrategia de optimización idealmente debería trabajar sobre los datos activos y proporcionar optimización, sin sacrificar rendimiento.

En general, la tecnología Real-Time Compression de IBM conservadoramente ofrecerá una reducción del 50% en los requisitos de almacenamiento de datos. Pruebas recientes en entornos de nube demostraron específicamente que las razones de compresión de los sistemas operativos virtuales Linux pueden llegar al 70%, en tanto que los sistemas operativos virtuales Windows pueden llegar al 50%. Incluso los Documentos de Microsoft Office que son pre-comprimidos muestran una ganancia adicional del 20% en la eficiencia de almacenamiento después de Real-Time Compression. La mayoría de las otras tecnologías de eficiencia del almacenamiento tienen un impacto en el rendimiento que los usuarios deben aceptar o tratar de resolver. Real-Time Compression permite a los usuarios reducir los costos de almacenamiento, bajar el espacio en rack y minimizar los requisitos de refrigeración, sin afectar el nivel de rendimiento.

La compresión, cuando se realiza en tiempo real, hace que todo sea más eficaz. La capa SSD descrita puede almacenar el doble de datos o puede ser reducida en tamaño, para ahorrar presupuesto en la capa más cara. Incluso el ancho de banda se torna más eficiente, porque es posible transferir el doble de datos a través del mismo segmento.

La ventaja de comprimir en tiempo real

Cuando la eficiencia del almacenamiento, como la compresión, se realiza en tiempo real, la gestión de datos se simplifica. El administrador de almacenamiento no tiene que diferenciar entre áreas de almacenamiento optimizadas y no optimizadas. Tampoco tiene que configurar un horario especial para ejecutar la optimización del almacenamiento. Finalmente, la capacidad neta disponible siempre se refleja con precisión y al instante. No hay que hacer "matemáticas en una servilleta" para ver cuánta capacidad "real" estará disponible después de que se dispare el proceso de compresión.

Real-Time Compression no solo es más eficiente que la de-duplicación en un entorno de nube sino que también es más eficiente que los algoritmos de compresión tradicionales que se basan en segmentos de tamaño fijo que se comprimen para producir una salida de tamaño variable. El tamaño del segmento tendrá un impacto en el rendimiento o en los índices de compresión. Por ejemplo, si se usan grandes segmentos, hay una mejor probabilidad de encontrar una oportunidad de compresión, pero el rendimiento I/O se verá afectado, porque debe procesarse un segmento entero incluso para un nivel bajo de I/O. Si los datos se descomponen en segmentos más pequeños, el rendimiento I/O tiene un menor impacto, pero el índice de compresión se reduce.

IBM Real-Time Compression, por el contrario, utiliza una entrada de tamaño variable, lo cual le permite adaptarse a las cargas de trabajo, comprimiendo solo los datos que se escriben, lo cual ayuda a mejorar el desempeño para cargas de trabajo activas que a menudo incluyen pequeñas operaciones de escritura. También agrupa varias I/O, para que los datos a los que se accede en forma conjunta se compriman juntos. Este enfoque permite niveles de índices de compresión mayores, y al mismo tiempo minimiza la I/O en disco y permite el acceso aleatorio a los datos para aplicaciones como bases de datos y correo electrónico.

Integración Storwize V7000

La integración de Real-Time Compression al IBM Storwize V7000 hace que la compresión sea una opción fácil de habilitar. Cuando se crea un volumen, hay una opción de seleccionar que el volumen se comprima. Una vez seleccionado, la tarea del administrador del almacenamiento ha terminado. No hay ajustes ni programación de tareas de optimización. Esto tiene su atractivo para el administrador del almacenamiento, quien a menudo tiene muchas tareas que cumplir. La conversión de volúmenes no comprimidos a volúmenes comprimidos es tan simple como generar el volumen comprimido y luego hacer una copia en espejo del volumen anterior y romper el espejo con el nuevo volumen comprimido que se convierte en el objetivo de almacenamiento primario.

Finalmente, como Storwize V7000 puede virtualizar tanto el almacenamiento interno que trae como el almacenamiento externo de terceros que se conecta, existe la capacidad de extender la vida útil de conjuntos más antiguos utilizando la compresión. Storwize V7000 no solo simplificará su gestión, sino que además puede duplicar su capacidad.

Conclusión

La Infraestructura de Nube implica una carga exclusiva para el almacenamiento que la sustenta, que exige un rendimiento I/O muy alto y una utilización muy eficiente. La mayoría de los Sistemas de Almacenamiento tienden a enfocarse en uno de estos atributos (rendimiento o eficiencia) pero no en los dos al mismo tiempo. Éste debe poder ofrecer el rendimiento requerido y la eficiencia necesaria para que el ROI pueda mantenerse y con el sistema correcto, mejorarse.

IBM Storwize V7000 – gracias a sus niveles de almacenamiento automatizado, que se suman a sus capacidades de eficiencia – ofrece la respuesta combinada justa que se requiere del Sistema de Almacenamiento, en tanto que mantiene el sistema en un nivel fácil de administrar. Easy Tier constituye una forma automatizada y simple de aprovechar al máximo una inversión en SSD y su exclusiva prestación Real-Time Compression simplifica el proceso de eficiencia del almacenamiento. La capacidad del Sistema Storwize V7000 de ofrecer rendimiento y eficiencia en forma automatizada y simplificada le permiten expandir el ROI que ofrece la Infraestructura de Nube.

A medida que se expande la cantidad de instancias virtuales dentro de la Infraestructura de Nube, también se expanden las complejidades del almacenamiento para hacer frente a esos desafíos. Los Sistemas de Almacenamiento deben ofrecer altos niveles de automatización y volverse más inteligentes, para preservar el ROI de los proyectos de nube. La Estrategia Smarter Computing de IBM – y dentro de ella Smarter Storage – son ejemplos excelentes de cómo la tecnología puede potenciar esta expansión.

Este documento se elaboró con financiación de IBM. Si bien el documento puede utilizar materiales disponibles al público de diversos proveedores, incluso IBM, no necesariamente refleja las posiciones de dichos proveedores sobre las cuestiones que trata.