



**Flash ya no es exclusivamente para
bases de datos**

*Elaborado por: George Crump, Lead
Analyst Prepared: Agosto del 2016*

Un sistema de almacenamiento basado en memoria, o lo que hoy denominamos un sistema de almacenamiento totalmente en flash, es una herramienta que TI ha empleado en el pasado para acelerar los entornos de bases de datos. Estos entornos están hechos a la medida para los atributos de flash por su naturaleza de alta transacción y pequeña E/S. Sin embargo, la considerable disminución del precio de los sistemas de almacenamiento all-flash y la aceptación de las tecnologías de eficiencia de datos han hecho que las cargas de trabajo de nivel medio, como la virtualización, sean ideales para los sistemas de almacenamiento all-flash. Pero, recientemente, gracias a la constante disminución de los costos en flash y a su creciente densidad de soporte físico, la memoria flash también ofrece un valor real para cargas de trabajo basadas en archivos o en objetos.

El estado de los datos basados en archivos

Los sistemas de archivos, especialmente, el almacenamiento de objetos, se consideran a menudo como el vertedero digital del centro de datos. Pero el caso de uso de los sistemas de almacenamiento basados en archivos ha cambiado drásticamente durante los últimos años gracias al fenómeno de big data. Muchas organizaciones tienen en marcha proyectos como Hadoop, SAS, Spark o Splunk para analizar los datos que los dispositivos de la Internet de las cosas (IoT) crean y dan como retroalimentación a la organización. Además, algunas organizaciones tienen la necesidad de proporcionar transmisión de medios a sus empleados o clientes. Por último, dado que flash comienza a dominar el conjunto de aplicaciones de producción tradicionales, estas organizaciones están observando la necesidad de proporcionar una estrategia de protección de datos que sea más rápida que las configuraciones actuales basadas en disco duro o cinta.

Tradicionalmente, estos sistemas de almacenamiento secundario se han basado en un disco duro, pero ha llegado el momento de considerar la posibilidad de trasladarlos hacia el almacenamiento all-flash.

El caso de almacenamiento secundario all-flash

Los sistemas de almacenamiento secundario están diseñados para almacenar, principalmente, cargas de trabajo basadas en archivos, representadas por iniciativas de big data, proyectos de DevOps y almacenamiento tradicional de archivos. Es importante comprender que el almacenamiento secundario no es sinónimo de "no importante". El almacenamiento secundario es vital para la mayoría de las organizaciones, pero no necesita el mismo nivel de rendimiento y disponibilidad que requieren las cargas de trabajo de almacenamiento primario. Pero, en la medida en que los casos de uso para almacenamiento basado en archivos siguen aumentando, también lo hace la demanda por cada vez más rendimiento.

Existen dos motivos principales para considerar flash para estas cargas de trabajo, cuyos datos tradicionalmente residían en el almacenamiento secundario. La primera es que ahora estas cargas de trabajo tienen una necesidad legítima de rendimiento inherente a un sistema de almacenamiento basado en flash.

El beneficio del alto rendimiento es inmediatamente evidente en entornos como Hadoop, Spark y Splunk, donde los análisis de datos más rápidos y más frecuentes generan una mejor toma de decisiones. Otro impulsor clave para el almacenamiento secundario de alto rendimiento es la personalización. Las organizaciones están aprendiendo que si la aplicación que utilizan sus clientes personaliza la experiencia, para presentar al usuario el contenido que más le interesa, existe un aumento en el compromiso del cliente.

Así, aunque la mayor parte de los beneficios de un sistema basado en flash está relacionada con su reducción en la latencia e IOPs extremadamente altos, un sistema flash que sirva al mercado de almacenamiento secundario también debe ofrecer un ancho de banda increíble, que es valioso en situaciones de carga a granel o para protección de datos.

La segunda razón para considerar flash es la densidad. El almacenamiento basado en flash con el diseño adecuado, incluso sin eficiencia de datos, puede ofrecer mucha más capacidad por unidad de bastidor que el sistema de almacenamiento típico basado en disco duro. Mientras el costo de ese sistema puede ser mayor, al considerar el costo de espacio de piso del centro de datos, los menores requisitos de energía, los ahorros en enfriamiento y la reducción de la mano de obra operacional y los gastos operativos, el sistema de almacenamiento flash (especialmente, si es particularmente denso) es menos costoso que un sistema enfocado en big data que se base en disco duro.

Los requisitos de un sistema de almacenamiento secundario all-flash de densidad extrema

El requisito principal de una matriz de almacenamiento secundario all-flash es que necesita ser extremadamente denso. Los profesionales de TI deben esperar 1/2 petabyte de capacidad en 3 U. Esto es porque los módulos flash pueden estar estrechamente empacados en un gabinete de almacenamiento, a diferencia de los sistemas de disco duro, donde hay que preocuparse por el calor y la vibración causada por decenas de unidades que están girando. Además, el factor de forma de flash es más pequeño y no necesita estar rodeado por la mecánica de una unidad de disco duro. Las densidades de estas proporciones también deberían bajar los costos de energía y enfriamiento del nivel de almacenamiento secundario y, normalmente, repercutir positivamente sobre el costo de la granja de servidores asociada.

Confiable y protegido

Un sistema all-flash secundario no necesita cumplir los mismos requisitos de alta disponibilidad y rendimiento de una matriz all-flash. Pero sí necesita proporcionar un mejor rendimiento que un sistema de almacenamiento secundario basado en unidad de disco duro tradicional. Al mismo tiempo, si bien no necesita cumplir con la meta de 24x7x365 de una matriz de almacenamiento all-flash primario, necesita ser muy confiable y ofrecer una alta integridad de datos.

Software neutral

A diferencia de las cargas de trabajo de almacenamiento primario, donde el tipo general de aplicación es conocido, los sistemas de almacenamiento secundario necesitan dar soporte a diversas cargas de trabajo y aplicaciones. La mayoría de las aplicaciones comunes que tendrán acceso a los datos en el sistema de almacenamiento secundario, como Spark, Splunk, Hadoop, Couchbase y Cassandra, ni siquiera existían hace unos cinco años. Es razonable esperar más innovación en estas plataformas y esperar más aplicaciones nuevas durante los próximos años.

Como resultado, el hardware de almacenamiento debe ser neutral respecto al software, lo que significa que gran variedad de software de almacenamiento y sistemas de archivos pueden tomar ventaja de él. El software de almacenamiento debe adaptarse de modo que pueda presentar ambos protocolos heredados, como NFS, y protocolos emergentes, como Hadoop File System (HDFS) y Object Storage.

El software también debe descargar parte de las responsabilidades de protección de datos e integridad de datos del sistema de almacenamiento, al ser capaz de asegurar que las copias de datos sobrevivan a un fallo de un componente y asegurar que los datos escritos hoy se puedan potencialmente leer de igual manera años más tarde.

Rentable

Por último, la realidad es que este nivel de almacenamiento secundario, aunque el rendimiento se ha vuelto cada vez más importante para este nivel en los últimos años, se está comparando con el almacenamiento basado en disco duro. Gracias a los ahorros en costos de energía y densidad, así como al significativo aumento del rendimiento, estos sistemas no necesitan mejorar el precio del disco. Pero necesitan acercarse. Los profesionales de TI deben esperar comprar estos sistemas por menos de \$ 1 por GB.

Es posible obtener mayores ahorros al no sobrecargar los sistemas con software "incluido". Recuerde: el mercado de datos secundarios es tan inestable, que cualquier software incluido puede ser inservible en un par de años. Los proveedores de almacenamiento deben enfocarse en proporcionar sistemas de almacenamiento de alta calidad y alta densidad, diseñados para complementar el software, no para competir con él.

Conclusión

Ha llegado el momento en que los sistemas de almacenamiento secundario se trasladen hacia all-flash. Los sistemas de almacenamiento secundario se benefician del rendimiento de flash tanto como las aplicaciones tradicionales de producción. La capacidad de analizar instantáneamente los datos proporciona a las organizaciones grandes ventajas. Pero esta nueva plataforma de flash tiene que ser diferente. Tiene que ser más densa que sus homólogos de almacenamiento primario. No solo reduce el espacio de piso del centro de datos y los costos de energía, sino que también proporciona un análisis mucho más enriquecido de los datos, ya que una sección transversal más grande puede ser analizada en menos tiempo que si estuviera en almacenamiento de disco duro. Por último, el nivel secundario todavía es una entidad en evolución. Debe permitir el envío de datos de origen de seres humanos y máquinas. Un enfoque impulsado por software es ideal para esta plataforma.

Este documento fue desarrollado con financiamiento de [IBM](#). Aunque el documento puede utilizar material públicamente disponible de varios proveedores, incluido IBM, no refleja necesariamente las posiciones de los proveedores sobre los temas tratados en este documento.