

Tecnologías de Almacenamiento en Centro de Datos

Salazar, Luis Diego
Universidad Latina de Costa Rica
diego_salazar_f@hotmail.com



Recibido 4/ago/2018
Aprobado 24/oct/2018

Resumen

La cantidad de información digital que es albergada en sistemas de almacenamiento en centros de datos va en crecimiento a un ritmo muy acelerado. Dichos sistemas de almacenamiento pueden llegar a hospedar una cantidad finita de datos, pero nuevos datos son creados día a día y el único límite se encuentra en los dispositivos donde son hospedados físicamente. El siguiente artículo introduce las tecnologías y funcionalidades en sistemas de almacenamiento en centros de datos para aprovechar este recurso al máximo con el objetivo de cumplir con la necesidad mundial de almacenar toda información que es creada bajo demanda a un costo accesible. La conclusión más importante a la que se llega es que los sistemas de almacenamiento ya no son una caja con discos duros, tienen características avanzadas para sacar el mayor provecho a la inversión, así como continuidad e integridad de la información.

Abstract

The amount of digital information kept on storage system located on data centers are growing in accelerated paced. These storage systems can retain a finite amount of data. But new data are created day by day and the only limit reside on the physical storage device where are located the data. This article introduces some technologies and features on storage systems that can maximize this resource with the objective to store all data on demand created for an accessible price. The main and most important conclusion is to remark that storage systems are not only a box with hard drives, they are a completed and integrated systems with advanced features that can maximize all the invest, also allow continuity and integrity of data.

Palabras Claves — Sistemas de Almacenamiento, Información Digital, Crecimiento data, Centros de Datos.

Keywords — Storage System, Digital Information, Growing data, Data Center.

I. INTRODUCCIÓN

Desde el uso masivo del internet a nivel mundial, han pasado más de dos décadas, casi ninguna persona al final del último siglo se imaginaba el impacto a nivel global y cómo la humanidad se apoyaría en esta tecnología, pero para que el internet funcione la información tiene que estar alojada en algún lugar. ¡Es ahí entra en funcionamiento los sistemas de almacenamiento!

Los datos digitales se miden en Bytes y para ilustrar la idea, en la siguiente figura se muestra una visualización de la cantidad de bytes que conforman un Zettabyte (ZB).



Figura 7. A ZETTABYTE SIZE IN NUMBERS! 2018 Recuperado de: <https://image.slidesharecdn.com/must-150404090138-conversion-gate01/95/big-data-analytics-in-markeing-8-638.jpg?cb=1428138172>

Se puede definir un sistema de almacenamiento como un dispositivo físico donde se encuentran distintos elementos que almacenan u brindan la información digital de manera segura.

Toda la información contenida en internet que actualmente es manejada en Zettabyte, es guardada en unidades de almacenamiento en distintos medios tales como discos duros, unidades de cinta o memorias no volátiles, por mencionar algunos ejemplos. Ahora bien, el problema radica en que estos medios o recursos solo pueden contener una cantidad finita de bytes. Al pasar los años, dichos dispositivos de almacenamiento alojados en centros de datos han crecido de manera

vertiginosa en la cantidad de información que pueden albergar, pero no es suficiente. Se ocupan de mecanismos y funcionalidades adicionales para aprovechar al máximo este recurso tan valioso.

Según el estudio de Reinsel, D., Gantz, J., y Rydning, J. (2017) *Date Age 2025: The Evolution of Data to Life-Critical*, el total de la información digital proyectado para el 2018 es de 30 ZB y se estima que para el 2025 el total alcanzará los 163 ZB.

En este artículo se muestran los diferentes tipos de sistemas de almacenamiento en centros de datos y presentará algunas de las características avanzadas y sus mecanismos para aprovechar cada byte de espacio para hospedar toda la información digital que es producida anualmente de manera segura.

II. TIPOS SISTEMAS DE ALMACENAMIENTO

Todos los sistemas de almacenamiento para centros de datos son diseñados para cumplir con un tamaño estándar y ser colocado en un gabinete. Usualmente se manejan capacidades entre 5 TB hasta 6 PB, de igual manera cuentan con total redundancia en fuentes de poder, ventiladores, discos duros, controladoras, puertos de comunicación y cableado.

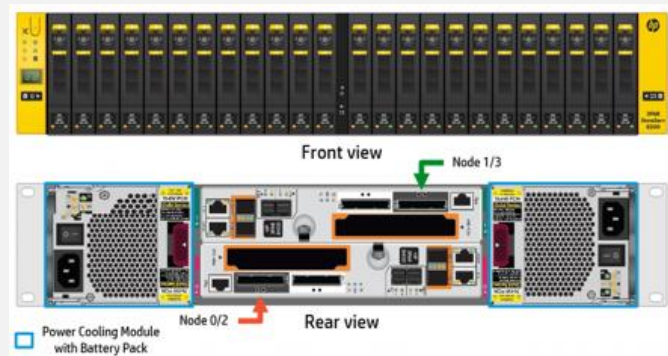


Figura 8. Storage marca 3PAR vista frontal y trasera 2016 Recuperado de: https://hpe.kx.com.tw/file/repository/a00004511enw_omq99_7yi1ipz.pdf

A. Nivel de Bloque

Almacenar datos a nivel de bloque en el

dispositivo significa que al guardar cualquier dato, este se guardará en pequeñas piezas de un tamaño fijo por ejemplo 1024 bytes. La misma tampoco contará con metadata, formato o dueño. La ventaja de almacenar los datos en bloque es el transporte de estos datos que comúnmente es a nivel de una red especializada llamada *Storage Area Network* (SAN) que es mucho más eficiente y rápida. Igualmente, cada bloque puede verse como una unidad de almacenamiento de disco llamada LUN hacia el host y puede ser vista por el sistema operativo en el tipo de archivos que desee formatearlo. Este tipo de almacenamiento es recomendable para almacenar Bases de Datos por su alto rendimiento y alta disponibilidad, como para ser usado en sistemas de archivo de máquinas virtuales, debido a que soporta arranque desde el almacenamiento.

B. Nivel de Archivos

Contrariamente al almacenamiento por bloques donde un dato se descompone en una cantidad fija de bytes que son guardados en el dispositivo, en este tipo, los datos son almacenados en forma de ficheros y de manera estructurada; por ejemplo, una carpeta que contiene una dirección raíz y a su vez contiene subcarpetas y en ella archivos. Este tipo de almacenamiento comúnmente llamado *Network Attached Storage* (NAS) y los archivos pueden ser vistos a través de protocolos como NFS o FTP en el host. La ventaja de usar almacenamiento por archivos es que es sencillo y rápido de implementar. Además, su costo es menor al que se puede usar la red LAN como medio de transporte.

C. Objetos

Este tipo de almacenamiento es el más nuevo debido a las tendencias tecnológicas como *Big Data* y *Cloud*. Es usado para almacenar datos no estructurados como imágenes, videos, audios, textos; en el cual los datos y contenido no pueden ser estructuradas en tablas como en

una base de datos. Su principal ventaja es que es muy flexible y escalable, no tiene un número límite para almacenamiento de objetos, igualmente se pueden alojar objetos mayores a 5 TB.

D. Híbridos

En este tipo de almacenamiento se puede combinar todas las anteriores en un solo dispositivo consolidado. Entonces dependiendo de la necesidad del negocio se utiliza el tipo de almacenamiento más óptimo y eficiente.

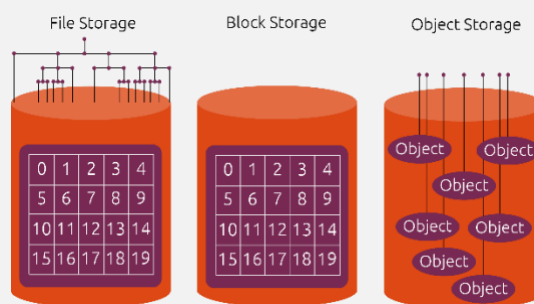


Figura 9. Tipos de Almacenamiento. 2018

Recuperado de:

<https://blog.ubuntu.com/2015/05/18/what-are-the-different-types-of-storage-block-object-and-file>

III. CARACTERÍSTICAS AVANZADAS

La información contenida en un sistema de almacenamiento es invaluable para el dueño, pues un solo documento de unos pocos bytes puede representar cientos de horas de trabajo. Incluso los documentos pueden clasificarse y mantenerse en sitios altamente confidenciales. Es por eso que estos dispositivos tienen que contar con elementos que no permitan perder información y contar con la seguridad apropiada para que solo los usuarios autorizados puedan manipularla. Así mismo tienen que manejar mecanismos que puedan optimizar el espacio y la eficiencia energética sin perder rendimiento o velocidad.

A. Replicación

Un dispositivo de almacenamiento puede ser replicado parcial o totalmente sin importar la distancia física entre ellos. Pero hay que planear de la mejor forma y tomar en cuenta, cual debe ser su mejor ubicación, con el fin de evitar el desabastecimiento de energía, en caso de algún tipo de desastre natural o provocado, como es el caso de los incendios. Después de los atentados a las torres gemelas en el año 2001, se ha tomado mucha importancia a esta tecnología, en ese trágico evento donde se perdieron cientos de vidas, también se perdió información digital con valores superiores a los millones de dólares. La razón principal fue porque las empresas tenían su replicación del centro de datos en la torre contigua y viceversa. Por lo tanto, al caer las dos torres, la Información se perdió. En la siguiente figura se puede observar la mejor práctica para implementar un proceso de replicación. Entre el sitio A y B se puede tener una replicación en tiempo real con una distancia menor a 50 Kilómetros. Por otra parte, existe una replicación entre el sitio A y C simultáneamente, que puede tener una distancia de miles de kilómetros entre sí, que, aunque no es a tiempo real por las velocidades de conexión, representa un punto a favor en caso de que haya un desastre entre los sitios A y B que pueda tener mayor probabilidad de que suceda debido a la cercanía entre ellos, se podría pensar el caso de un terremoto o huracán donde muchas estructuras se ven comprometidas al existir distancias mayores.

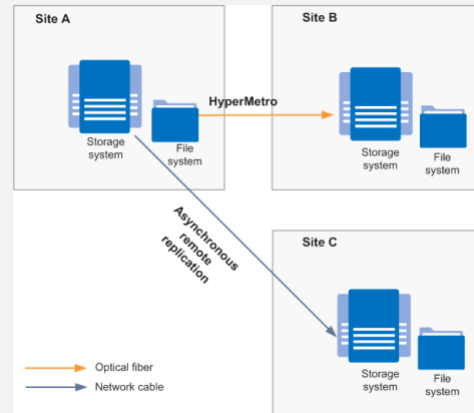


Figura 10. Mejores Prácticas Replicación 2018 Recuperado de:

<http://support.huawei.com/enterprise/product/images/31f3969a83314c8688e7ad00cf08c607>

B. Volúmenes delgados

Este es un mecanismo que utiliza el sistema de almacenamiento para optimizar el aprovisionamiento del espacio al usuario. En la antigua manera si un usuario ocupaba 1 TB de espacio, ese era la cantidad de que se reservaba y brindaba de manera íntegra, aunque nunca llegara a necesitarlo. Ahora con volúmenes delgados si el mismo usuario le solicita 1 TB, si verá que le están suministrando ese espacio, pero en el equipo de almacenamiento solo se verá consumido el espacio real utilizado y crecerá bajo demanda hasta alcanzar el TB de espacio que originalmente se le brindó. Se podrán imaginar cuanto espacio es ahorrado en el almacenamiento cuando son provisionados miles de volúmenes que van creciendo bajo demanda. Permite usar de la manera más eficiente el espacio y ahorrar dinero en recursos que no se está utilizando.

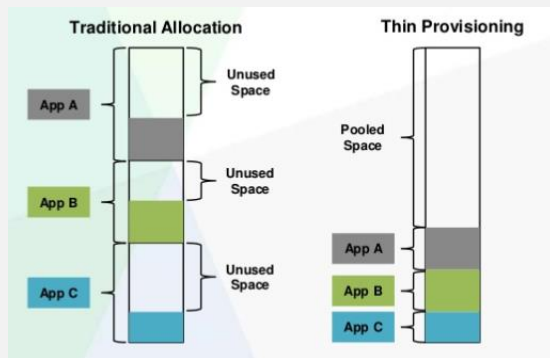


Figura 5. Aprovechamiento del Espacio 2018
Recuperado de:
<https://image.slidesharecdn.com/theskinnyonthinprovisioninginvirtualenvironments-140805100811-phpapp02/95/the-skinny-on-thin-provisioning-in-virtual-environments-4-638.jpg?cb=1428575476>

C. Deduplicación

Con esta tecnología podemos salvaguardar el espacio y ahorrar recursos al máximo, el sistema de almacenamiento realiza un barrido de la información almacenada y busca redundancias en las secuencias de bytes, esto con el fin de solo almacenar una copia y borrar el resto haciendo una referencia a la primera. En la siguiente figura se visualiza a alto nivel lo que esta tecnología realiza. Se puede imaginar una empresa que envía un correo a toda la organización, en la manera tradicional, cada copia de correo es almacenada por cada usuario, entonces un correo electrónico que pesa 1MB para el sistema de almacenamiento puede convertirse en 1GB muy rápido. Con Deduplicación solo se almacena una copia y el resto queda solo como una referencia. Y eso es solo un correo electrónico, se puede figurar la cantidad de información repetida (fotos, audios, textos, entre otros) que solo genera pérdida de espacio.



Figura 6. Información después de Deduplicar 2018
Recuperado de:
<https://grumpy.tech/content/images/2017/11/image.jpg>

D. Destrucción segura de datos

Este elemento cobra más importancia al pasar los años, generalmente un sistema de almacenamiento es usado entre 5 y 10 años por una compañía y cuando deja de ser utilizado se desecha o reutiliza. Pero los datos almacenados son importantes y no basta con formatear las unidades provisionadas. Esta característica permite destruir los datos en un disco duro y no existe la posibilidad de recuperar la Información. Para ello se sobrescriben los datos haciendo un barrido bit por bit del disco duro durante múltiples secuencias de escritura. Es usado de manera frecuente en compañías financieras y de gobierno.

E. Auto nivelamiento

En un sistema de almacenamiento en donde se pueden mezclar diferentes medios físicos que almacenan los datos, estos medios difieren en la capacidad de almacenamiento, velocidad y ancho de banda. Los tres más comunes son:

- Nivel 1: SSD
- Nivel 2: SAS
- Nivel 3: NL-SAS

El almacenamiento de Nivel 1 se realiza en discos de estado sólido (SSD) proporciona un almacenamiento dedicado a suplir necesidades prioritarias de lectura y escritura, por ejemplo, una base de datos. Los discos duros SAS

brindan un punto intermedio entre precio y rendimiento, son usados en ambientes donde lo almacenado es importante y accedido con frecuencia, como un sistema operativo. Por último, en nivel 3, los discos NL-SAS son los que presentan menor rendimiento, pero son más baratos y presentan una gran densidad de espacio, se suelen usar para almacenar respaldos.

En la siguiente figura se muestra la mejor práctica para poblar un sistema de almacenamiento con respecto a su almacenamiento. Se observa que tan solo un 3% se debe dedicar a SSD, ya que solo se utiliza en información que tiene que ser accesada de manera instantánea. Por el contrario, la mayoría de espacio se espera que se encuentre en discos NL-SAS debido a que no son accedidos de manera frecuente y es mucha data.

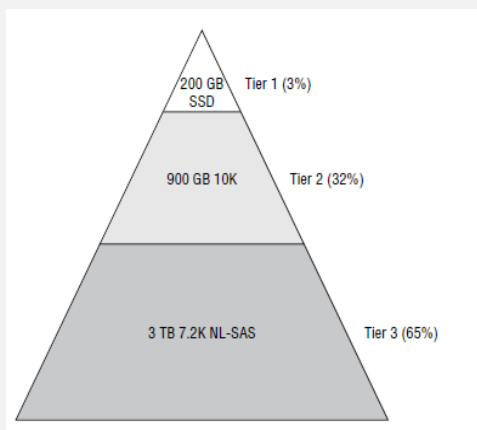


Figura 7. Niveles de Almacenamiento 2016. Recuperado de: https://www.acnc.com/media/images/site_library/1061_ART_-_Tiered_Storage.png

El Auto Nivelamiento es una característica avanzada de un sistema de almacenamiento que permite mover esta información de manera inteligente entre los niveles, esto produce que el dispositivo funcione siempre a su mejor rendimiento entonces sin tener que hacer una gran inversión económica. Esta tecnología escanea toda la información guardada y la

comparara con la frecuencia de uso, horas del día donde es más accesada y tipo de información. Cuando el sistema mapea esta información y el dispositivo es menos usado, mueve la información entre las unidades de almacenamiento sin ninguna afectación.

Como se observa en la siguiente figura, después del proceso de Auto Nivelamiento la información que es almacenada y es usada con más frecuencia, se guarda en discos SSD o también llamados Flash, por el contrario, los de menor uso se guardan en disco NL-SAS. Esto permite tener un balance entre rendimiento y costo.

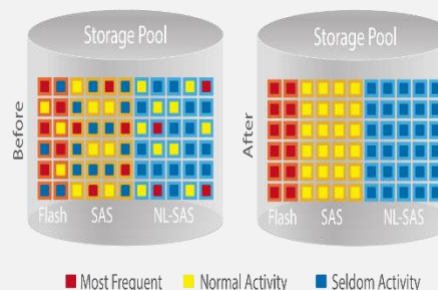


Figura 8. Antes y Después de Auto Nivelamiento 2018 Recuperado de: https://www.qnap.com/solution/qtier-auto-tiering/_images/Qtire-12G-SAS.png

IV. CONCLUSIÓN

- Hay 4 tipos de sistema de almacenamiento y dependiendo de la información que se quiera o considere necesario guardar, se elige el más adecuado.
- La replicación permite un aseguramiento de la información en caso de un evento de recuperación de desastres. Por ello se sugiere usar la técnica de replicar en dos sitios alternos a diferentes distancias.
- Al utilizar las características de Volúmenes delgados junto con Deduplicación y Auto Nivelamiento se obtiene un aprovechamiento de espacio muy superior al que se ha venido usando tradicionalmente sin que se

afecte el rendimiento o sea necesario realizar inversiones económicas.

- Es importante para la protección de la información, que cuando la misma ya no se utilice o se vuelva obsoleta, se realice el mecanismo de destrucción segura de los datos.

REFERENCIAS

File vs. Block vs. Object Storage. (2018).

Recuperado de SNIA:

<https://www.snia.org/sites/default/files/ESF/SNIA-Block-File-Object-Storage-Webcast-Final.pdf>

Deduplicación de Datos. (2018). Recuperado de DELL EMC:

<https://mexico.emc.com/corporate/glossary/data-deduplication.htm>

Data Age 2025: The Evolution of Data to Life-Critical. (2017). Recuperado de Storagenewsletter.:

<https://www.storagenewsletter.com/2017/04/05/total-ww-data-to-reach-163-zettabytes-by-2025-idc/>

Automatic Storage Tiering. (2018).

Recuperado de Gartner:

<https://www.gartner.com/it-glossary/automatic-storage-tiering-auto-tiering>