



Los hipervínculos son
señalados con esta clave



e-Ciencias de la Información

Tecnologías de almacenamiento de información en el
ambiente digital

► **Salvador E. Vázquez-Moctezuma**

Publicado 01 de julio, 2015 / Revisión bibliográfica 1

Revista electrónica semestral ISSN-1659-4142



Universidad de Costa Rica
Visite el sitio web de e-Ciencias de la Información

➤ Tecnologías de almacenamiento de información en el ambiente digital

Information storage technologies in the digital environment

➤ *Salvador E. Vázquez-Moctezuma*¹

RESUMEN

Las bibliotecas han funcionado por mucho tiempo como depósitos de información, a la cual facilitan el acceso recogiéndola y preservándola. Actualmente el entorno ha cambiado y las bibliotecas se enfrentan al reto de contener un creciente volumen de información digital, de ahí que son necesarias las tecnologías de información masiva. Este trabajo tiene como propósito revisar los métodos de sistemas de almacenamiento de datos masivos, por lo cual se describen y analizan cuatro tipos de tecnologías de almacenamiento en datos: DAS, NAS, SAN y almacenamiento en nube; para ello se utilizó el método documental. Se encontró escasa literatura sobre los sistemas de almacenamiento digital vistos desde la perspectiva de las ciencias de la información, razón por la cual se retoman publicaciones del área de ciencias de la computación e ingeniería en sistemas para comprender el tópico. Se concluye que el almacenamiento de los datos en la era digital se lleva a cabo de forma concentrada y distribuida.

ABSTRACT

Libraries have long functioned as information repositories, providing access to it by collecting and preserving data. In the present, environment has changed, so libraries are challenged to contain an increasing volume of digital information, hence the massive information technologies are needed. The aim of this paper is revising methods of storage systems for massive data; therefore, four types of data storage technologies, which are DAS, NAS, SAN and cloud storage are described and analyzed using the documentary method. It was found scarce literature on digital storage systems viewed from the perspective of information science; for this reason, publications in the field of computer science and engineering systems were used in order to understand the topic. It is concluded that storing data in the digital age is carry out in concentrated and distributed form.

Palabras clave

Sistemas de almacenamiento de datos, alojamiento, sistema de almacenamiento concentrado, almacenamiento en nube.

Keywords

Data storage systems, depot, concentrate storage system, cloud storage.

Recibido: 22 de marzo 2015 | Corregido: 05 de junio 2015 | Aprobado: 22 de junio 2015

¹ Universidad Nacional Autónoma de México. Posgrado en Bibliotecología y Estudios de la Información. MÉXICO. quique.vazquez@hotmail.com

1. Introducción

Con el paso de los siglos el ser humano ha tenido la necesidad de guardar su conocimiento en diferentes soportes desde las antiguas tablas de arcilla hasta los medios digitales. El almacenamiento y recuperación de la información representan uno de los problemas a los que la humanidad se ha tenido que enfrentar desde la invención de la escritura. Con la aparición de la computadora este problema se ha resuelto parcialmente con nuevos dispositivos de almacenamiento, diseños de conexión y estructuras de bases de datos.

Las tecnologías han evolucionado para atender las necesidades de almacenamiento, de ahí que nos encontramos en una etapa compleja donde es más fácil producir datos que guardarlos y administrarlos. Se tiene que cada año en el mundo se produce entre 3 y 5 Exabytes de información (Lyman y Varian, 2003). Aunque en 2007 se almacenaron de manera comprimida 2.9 Zettabytes, se comunicaron cerca de 2 Zettabytes y se ejecutaron 6.4 Exabytes de instrucciones por segundo en computadoras de propósito general (Hilbert y López, 2011), se cree que en el año 2020 se tendrán 35 Zettabytes (Li y Cao, 2014); en consecuencia, las capacidades de almacenamiento han tenido que crecer. Por lo tanto, se intuye que la cantidad de información digital que se produce en el mundo es inmensa; sin embargo, ignoramos su verdadera cantidad, asimismo su enorme dimensión. Existen Exabytes de datos almacenados en servidores de empresas que se han visto en la necesidad de ampliar su capacidad de espacio; en principio pareciera que el tamaño de almacenamiento es un problema, aunque se olvida que la naturaleza de los datos y la administración de la entrada y salida del sistema de información es otra cuestión de suma importancia.

El almacenamiento de datos puede verse desde dos perspectivas. La primera observación se puede hacer desde el punto de las estructuras de sistemas de almacenamiento con opciones como DAS (Direct Attached Storage o Almacenamiento de Conexión Directa), NAS (Network Attached Storage o Almacenamiento Conectado en Red), SAN (Storage Area Network o Red de Área de Almacenamiento) y sistemas de almacenamiento en la nube, que incluye capacidades de espacio en unidades de discos duros tradicionales y sólidos, así como la tecnología de la Memoria de Cambio de Fase (PCM: Phase Change Memory). La segunda visión se enfoca en la naturaleza de los datos en una perspectiva más cercana a la administración de datos; probablemente se pueda tener la capacidad de espacio a través de los sistemas distribuidos de nube, pero surgen inconvenientes relacionados con la consistencia, disponibilidad y tolerancia de partición de los datos; es decir, se trata de una perspectiva más cercana a la administración de datos.

En la actualidad, las estructuras de almacenamiento y operatividad de las computadoras han cambiado. Por esto, el presente artículo tiene como propósito analizar cuatro tipos de sistemas de almacenamiento de datos masivos, los cuales son: DAS, NAS, SAN y almacenamiento en la nube, para ello se empleó una revisión documental.

2. Sistemas de almacenamiento de datos digitales

El ser humano en cada época se ha encontrado en la necesidad de mejorar sus mecanismos de almacenamiento a causa del incremento del número de datos producidos y los retos que surgen por el manejo de información. Desde la antigüedad la información era almacenada, los sumerios utilizaban para ello tablas de arcilla, siendo este un primer sistema de almacenamiento. En la era digital, el almacenamiento de datos surge desde mediados del siglo XX con la aparición de las computadoras y unidades de disco duros (HDD): los datos digitales han cambiado nuestra forma de almacenar, considerando que se necesita mayor capacidad de espacio y velocidad en los dispositivos de entrada y salida para acceder y recuperar información.

Durante la década de 1950 surgieron las computadoras electrónicas, grandes y costosas máquinas denominadas Mainframes, que funcionaban como centros de procesamiento de grandes organizaciones, caracterizándose por su capacidad de interrelacionar enormes bases de datos y soportar diversos dispositivos periféricos, asimismo atender una gran cantidad de usuarios. Es bien sabido que en los años 60 aparecieron las terminales que se conectaban a una unidad central de proceso permitiendo una conexión directa, rápida y eficiente; sin embargo, al incrementarse el número de terminales y dispositivos periféricos decayó la velocidad de comunicación. Más tarde, en la década de 1970 aparecieron las minicomputadoras que brindaron en ese momento una reducción en el congestionamiento a causa de que el usuario no tenía que conectarse a una unidad central de proceso. En este contexto se produce un retroceso en la forma de procesar información, debido a que se pierde la centralización de la información, al acarrearla en disquetes.

A mediados de la década de 1970, se dio la comercialización de dispositivos de almacenamiento de acceso directo o discos duros de IBM (International Business Machines Corp.), que implicaban una capacidad aproximada de 30 Megabytes; que si bien no era demasiado, sí era muy alto para aquel tiempo (Glossbrenner, 1993). Entonces, el alto costo y la necesidad de compartir y comunicar información entre usuarios obligaron a que los fabricantes y desarrolladores concibieran las redes locales.

Las redes de microcomputadoras, en principio, se formaban por simples conexiones de punto a punto con otra microcomputadora, permitiendo a los usuarios consultar recursos almacenados en los disco duro; no obstante, existía el inconveniente de que la conexión de los equipos permitía también un acceso total a los discos duros, por lo cual había problemas de seguridad e integridad en los datos. Debido a esto, las redes se empezaron a consolidar y surgieron los *file server* que permitían a los usuarios de la red acceder a la misma información y compartir archivos con niveles de seguridad para que los usuarios no pudieran ingresar de forma indiscriminada a toda la información o a los recursos comunes de la red. Para lograr esto, se tenía una computadora que fungía como servidor y se encargaba de administrar los recursos conectados permitiendo un uso eficiente e integridad en la información. A medida que las redes de área local fueron creciendo en tamaño y complejidad surgió la necesidad de comunicarlás entre sí, formando redes más amplias; y con la llegada de internet se maximizó la información, así como el espacio de almacenamiento.

En cuanto a los sistemas de almacenamiento y servidores tienen entre sus componentes unidades de disco duro tradicionales o sólidos. Los datos almacenados a través de unidades de disco duro tradicionales tienen amplias capacidades de almacenamiento, aunque con poco rendimiento en el acceso, tomando en cuenta su naturaleza de funcionamiento, pues la velocidad del giro de los platos magnéticos implica un funcionamiento lento, razón por la cual son remplazados por discos de estado sólido (SSD) con chips que permiten una mayor velocidad en el acceso a la información guardada. Estos últimos tienen la desventaja de ser más caros que los disco duros tradicionales; no obstante, el costo del SSD sería menor al de las nuevas memorias emergentes (Kim, Seshadri, Dickey, y Chiu, 2014).

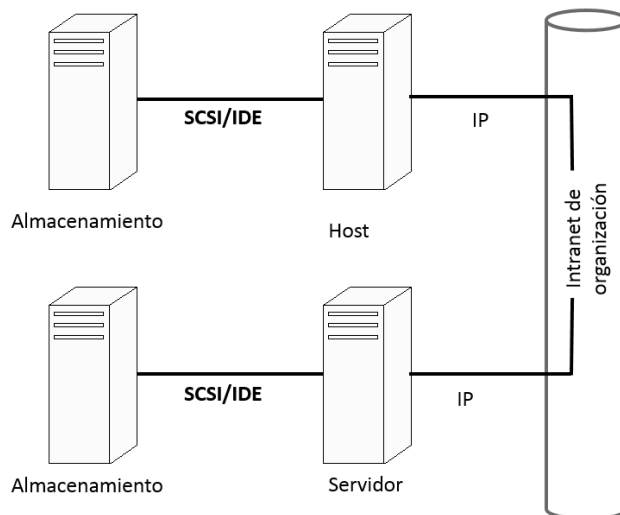
El desarrollo de la tecnología en almacenamiento sigue presentando nuevas formas como la PCM, que de acuerdo con Gopalakrishnan et al. (2010), tiene la bondad de facilidad de integración, escalabilidad, velocidad y resistencia; esta podría ser una opción en un sistema de almacenamiento de grandes capacidades de datos. Además, sus bondades prometen ser memorias no volátiles de próxima generación (Yoon, Bian y Kim, 2014; Wang, Jiang, Zhang y Yang, 2015), considerando que muchas aplicaciones modernas exigen cada vez más para el manejo de grandes cantidades de datos.

Nos encontramos ante una etapa de transición, de modo que las unidades de disco duro tradicional, sólido y memorias de PCM en este tiempo nos ayudarán a resolver nuestros problemas de almacenamiento de grandes datos. Este desarrollo tecnológico ha permitido que los centros de datos dedicados al almacenamiento estén estructurados de una o varias combinaciones de las siguientes cuatro formas: DAS, SAN, NAS y almacenamiento en la nube. A continuación describimos cada uno de ellos.

2.1. Almacenamiento de Conexión Directa (DAS)

Direct Attached Storage o DAS es una de las formas más sencillas y tradicionales del almacenamiento de conexión directa, donde las unidades de disco se encuentran conectadas directamente con los servidores o host a través de una interfaz de datos SCSI o IDE (Figura 1). De acuerdo con Zhao (2006) las conexiones en DAS tienen muchas ventajas, tales como: su instalación es fácil; el software es poco complejo; el costo en mantenimiento es bajo; la tecnología presenta madurez técnica, buena compatibilidad y, relativamente, es de menor gasto. Sin embargo, su deficiencia aparece en cuatro aspectos: (1) la capacidad de almacenamiento está limitada por el servidor; (2) su rendimiento de almacenamiento es directamente afectado por el servidor; (3) los servidores dispersos geográficamente se limitan al intercambio de información y gestión cuando se tiene un servidor aislado; (4) la carga de almacenamiento de datos y el acceso en el servidor hará en general tener un pobre rendimiento.

FIGURA 1
Almacenamiento de Conexión Directa (DAS)



Fuente: Elaboración propia con base en Sadlier (2003).

El entorno de uso de este tipo de arquitectura de almacenamiento es ideal para el intercambio de archivos localizados en ambientes con un único servidor o unos cuantos servidores, por ejemplo, una pequeña biblioteca que no necesita compartir información a través de largas distancias. Naturalmente se convierte en la elección previa para las pequeñas unidades de información, debido a que cuentan con menos recursos financieros. Principalmente es utilizado en las computadoras personales y pequeños servidores, que soportan solo aplicaciones que requieren capacidades bajas de almacenamiento y no admite directamente equipos múltiples de almacenamiento compartido.

2.2. Almacenamiento Conectado en Red (NAS)

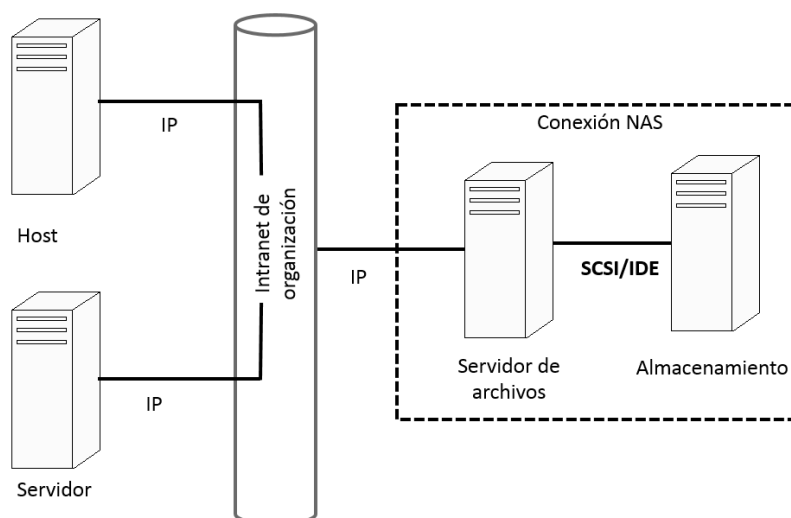
Para proveer el almacenamiento en trama es necesaria una LAN o WAN, además de un dispositivo de almacenamiento dedicado y diseñado para esta infraestructura; su propósito es proporcionar a los usuarios un sistema de servicio de acceso e intercambio de información. El almacenamiento en red se caracteriza por el depósito masivo de datos, lo que incluye intercambio de datos limitados, fiabilidad y seguridad en los datos, y así como el simplificado y unificado en la gestión de datos. Aunque su principal bondad es la capacidad de expansión, donde se proporcionan tasas de transmisión de la información de acuerdo al volumen de datos. Las conexiones SAN y NAS son ejemplos claros del almacenamiento en red; las analizamos a continuación.

El Almacenamiento Conectado en Red o *NAS* (del acrónimo *inglés Network Attached Storage*) es un dispositivo que se conecta a la red y provee un almacén de datos que permite a varios hosts acceder al mismo lugar de almacenamiento a través de una red IP. El espacio de almacenamiento se presenta en la red con un nodo dedicado a través de un servidor de archivos, aunque en sistemas recientes este dispositivo puede ser un dispositivo inmerso en la

red (Figura 2). NAS y LAN están en la misma red física; por lo tanto, NAS depende de ciertas características de LAN. Para ello necesita un gran ancho de banda en red y de muy alta potencia de procesamiento del CPU: cuando no se cumplen estas condiciones, la red se congestiona y su rendimiento se reduce (Liu y Yi, 2012).

Con el servidor de archivos se gestiona la entrada y salida de datos en el disco duro; además, se regula el acceso entre varios clientes de red. Para Edelson (2004), el almacenamiento en NAS tiene dos características. En primer lugar, es la conexión física, puesto que se conecta el servidor de archivos directamente al equipo de almacenamiento y otro punto a la red, evitando así la carga de entrada y salida de datos en el servidor; en segundo lugar, técnicamente, se reducen los movimientos del brazo de la unidad de disco duro y, por lo tanto, se reduce el desgaste. Sin embargo, en esencia la estructura de este tipo de almacenamiento muestra que todavía es un equipo de servidor tradicional.

FIGURA 2
Almacenamiento Conectado en Red (NAS)



Fuente: Elaboración propia con base en Sadlier (2003).

Los principales beneficios de NAS son la facilidad de comunicación entre una computadora y el sistema de almacenamiento en comparación con una conexión de computadora a computadora. El intercambio y recuperación de datos mediante una sola fuente de almacenamiento genera menos errores, menos trabajo al tratar de mantener copias de seguridad, y mayor precisión en la búsqueda de información. Estos sistemas son más seguros, porque en lugar de almacenar los datos en un solo disco duro distribuyen copias de los datos entre distintos discos duros que actúan como uno solo. Cuando un disco duro falla, se alerta al administrador de redes, y la información continúa estando disponible para todos los usuarios (Edelson, 2004).

Todos sabemos que un disco duro dañado puede ser reemplazado por uno nuevo sin necesidad de que los usuarios lo perciban, ya que desde sus computadores continuarán trabajando normalmente: continuarán teniendo acceso a la red y, por lo tanto, a la información que

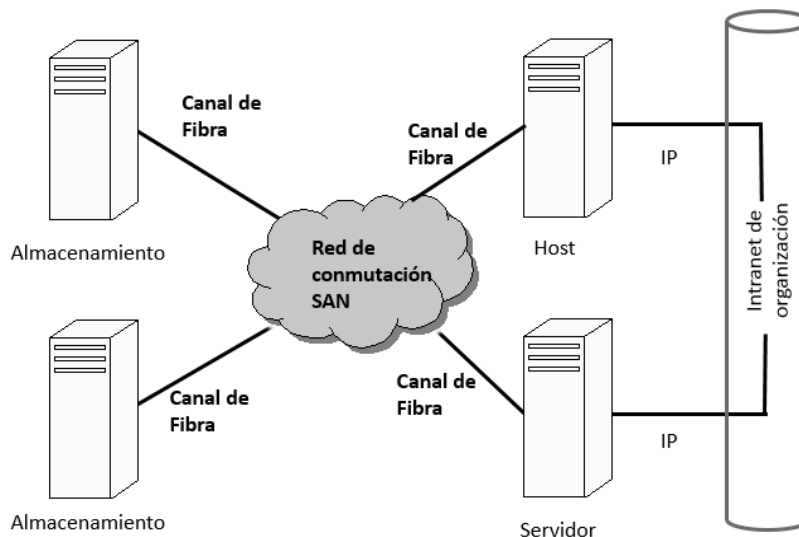
necesitan. En esta misma línea un disco de gran tamaño puede ser más barato que varios discos de menor capacidad de espacio. El sistema NAS, según Cunhe (2002), tiene ventajas tales como facilidad en la instalación, complementos o extensiones (*plugs*), precio, flexibilidad de conexión, fácil mantenimiento, seguridad de autenticación, administración de espacio en disco y escalabilidad. Así las cosas, NAS es una opción ideal para organizaciones pequeñas y medianas que buscan, de una manera simple y rentable, lograr el acceso de datos rápido en nivel de archivo para varios clientes.

En contraste, la escalabilidad se presenta como desventaja a causa de que la capacidad se limita por los equipos y dispositivos conectados; asimismo, NAS no podrá ser integrado cuando no esté bien configurado, y gracias a esto el sistema de archivos no podrá formarse. También existe un inconveniente con las copias de seguridad: si se hacen en horas de mayor tráfico de datos, es seguro que el consumo de ancho de banda y rendimiento será limitado. Por esto, cuando el número de usuarios simultáneos no es muy grande, NAS sería una económica y racional elección, pero no es adecuado con aplicaciones de grandes bases de datos (Tian, 2006). En realidad, NAS tiene que ser visto como un equipo de almacenamiento auxiliar en una red, el cual está directamente conectado a una red usando un hub o switch, y comunicándose por medio del protocolo TCP / IP; sin duda, NAS está orientado al paso de mensajes y archivos, formato en el cual transmite los datos.

2.3. Red de Área de Almacenamiento (SAN)

Existe la red de área de almacenamiento o *Storage Area Network*, por sus siglas en inglés SAN. Se centra en el almacenamiento de datos utilizando una topología de red flexible, además, con conexiones de fibra óptica que permiten alta velocidad en la transferencia de datos; ofrece la conmutación entre múltiples nodos (Figura 3). Sin duda, SAN es otro enfoque de almacenamiento compartido que a menudo se usa en la nube. En SAN, la gestión del almacenamiento de datos se encuentra relativamente independiente a la red de área local, con el fin de lograr el máximo grado de intercambio de datos, así como la extensión del sistema (Sadlier, 2003).

FIGURA 3
Red de Área de Almacenamiento (SAN)



Fuente: Elaboración propia con base en Sadlier (2003).

La tecnología SAN, para Xu y Zhu (2006), se orienta a la alta velocidad de procesamiento de datos masivos, lo que incluye alta velocidad en el acceso, almacenamiento seguro, intercambio de datos, respaldo de datos, migración de los datos, entre otras ventajas de los sistemas distribuidos. Se debe considerar que muchas organizaciones usan conexiones SAN con cable UTP otras con fibra óptica; esta última se caracteriza por la alta velocidad de transmisión de información.

El canal de fibra o FC es de gran fiabilidad, a causa de la tecnología de interconexión en gigabytes que provee la comunicación simultánea entre distintas estaciones de trabajo, mainframes, servidores, sistemas de almacenamiento y otros periféricos de entrada o salida. De ahí que el FC es ideal para mover grandes volúmenes de datos a través de largas distancias rápidamente y de forma fiable. La velocidad del puerto de fibra ahora ha alcanzado 4 GB para la transferencia de datos; aunque en ocasiones las limitaciones de SAN se deben al uso de cable UTP y del protocolo IP, esto se ha superado con el FC que da un mayor alcance y con un funcionamiento estable entre los dispositivos (Cunhe, 2002). El rasgo característico de esta arquitectura es el costo que sigue siendo demasiado alto para el uso general (Yang, Shih, Huang, Jiang y Chu, 2014).

Por su parte, el rendimiento es mayor con el uso del canal de fibra en comparación con las conexiones de cable par trenzado. En contraste, cuando se necesita gestionar el sistema de almacenamiento SAN será complejo por la gran cantidad de información, además del alto costo de la infraestructura para el uso de fibra óptica. Así SAN FC es adecuado para grandes unidades de información que tienen mayores presupuestos y requerimientos altos de transferencia y transmisión de datos.

2.4. Comparación DAS y NAS

Una larga historia tiene DAS. Se trata de un sistema de conexión de almacenamiento ampliamente utilizado, pero que a medida que se lleva a cabo el avance tecnológico ha sido poco a poco desplazado por NAS y SAN. En la Tabla 1 se distingue que NAS en comparación con DAS es superior en indicadores y desempeño. En el contexto de las unidades de información la conexión NAS es la elección apropiada para centros con exigencias en capacidad de almacenamiento, acceso además por el precio y ventajas en rendimiento. Una comparación de NAS y SAN no solo incluye un contraste técnico sino también un análisis de precio, efectividad y tendencias.

TABLA 1
Comparación entre DAS y NAS

ELEMENTOS DE COMPARACIÓN	DAS	NAS
Estructura de almacenamiento	Centralizado	Distribuido/ Centralizado
Conexión	SCSI	IP
Inversión inicial de equipo	Bajo	Bajo
Costo de mantenimiento	Alto	Bajo
Costo de expansión	Alto	Bajo
Sistemas operativos interdependientes	No	Sí
Expansibilidad	Bajo	Alto
Administración y mantenimiento	Complejo	Simple
Distancia de transmisión	50 metros	Ilimitado
Utilización de unidades de disco duro	Bajo	Alto
Rendimiento general	Alto	Bajo
Entornos de aplicación	Pequeñas y medianas unidades de información	Medianas unidades de información

Fuente: Elaboración propia.

2.5. Comparación de NAS y SAN

Los sistemas de almacenamiento NAS y SAN son soluciones completamente diferentes en su arquitectura. SAN se apoya en la transferencia de grandes volúmenes de datos mientras que NAS utiliza el modo de transferencia de archivos. Una comparación completa de NAS y SAN involucra un contraste técnico y también un análisis de costo-efectividad y usabilidad. En la Tabla 2 se compara NAS y SAN; este último sistema es dividido por la conexión IP (SAN IP) y el canal de fibra (SAN FC).

TABLA 2
Comparación entre NAS y SAN

ELEMENTOS DE COMPARACIÓN	NAS	SAN FC	SAN IP
Costo	Bajo	Alto	Medio
Expansibilidad	Medio	Alto	Alto
Administración y mantenimiento	Simple	Complejo	Simple
Recursos compartidos	Datos de archivos	Recursos de almacenamiento	Recursos de almacenamiento
Conexión	IP	Canal de fibra	IP
Modo de acceso	Archivo de datos	Bloques de datos	Bloques de datos
Medio de transmisión	<i>Cable de par trenzado</i>	Canal de fibra	<i>Cable de par trenzado</i>
Distancia de transmisión	Ilimitado	100 metros	Ilimitado
Efecto electromagnético	Sí	No	Sí
Medida de transmisión	Rápido/ Gigabytes	4 GB	Rápido/ Gigabytes
Rendimiento general	Medio	Alto	Medio
Entornos de aplicación	Medianas unidades de información	Grandes unidades de información	Medianas unidades de información

Fuente: Elaboración propia.

En la Tabla 2, se observan varias diferencias entre SAN IP y SAN FC en materia de transmisión. Sin embargo, el SAN FC no es adecuado para pequeñas bibliotecas debido a su costo, gestión y su complejo mantenimiento. De los tres tipos de tecnologías de almacenamiento, el más adecuado para las unidades de información es la conexión de almacenamiento SAN IP, debido a que es equitativo su costo por el rendimiento que ofrece.

Por último, el almacenamiento de conexión directa (DAS), almacenamiento conectado a red (NAS), y la red de área de almacenamiento (SAN) son sistemas de almacenamiento usados con frecuencia. No obstante, estas arquitecturas de almacenamiento tienen inconvenientes graves y limitaciones cuando se trata de sistemas distribuidos a gran escala debido a la cantidad de espacio y administración de los datos.

3. Almacenamiento basado en la nube

El desarrollo del almacenamiento de datos en la nube, mejor conocido como *cloud computing*, se da gracias al uso de equipos virtuales (Furht y Escalante, 2011); implica una infraestructura informática invisible para el usuario, pero al utilizarla parece que se tuviera un equipo físico real, permitiendo la gran ventaja de determinar el número de procesamiento, el sistema operativo, el tamaño de memoria RAM y de disco de almacenamiento. Esta elasticidad en la infraestructura es una de las técnicas usables en *Big data* (Sakr, Batista y Alomari, 2011; Schadt, Linderman, Soreson, Lee y Nolan, 2010): las tecnologías de virtualización han hecho que la computación sea accesible, asequible y rentable.

El nombre de *cloud computing* proviene de la utilización del símbolo con forma de nube o *cloud*, que es el diagrama usado en sistemas como una abstracción para determinar internet, mientras que *computing* implica la informática. Una buena definición de esto es la que ha formulado el NIST (Mell y Grance, 2011; Joyanes Aguilar, 2012), que la considera como

un modelo que permite el acceso bajo demanda a través de la red a un conjunto compartido de recursos de computación configurables (como por ejemplo red, servidores, almacenamiento, aplicaciones y servicios) que pueden ser rápidamente aprovisionados con el mínimo esfuerzo de gestión o interacción del proveedor del servicio (Mell y Grance, 2009, citados por Joyanes Aguilar, 2012, p. 91).

Una bondad de los entornos de la nube es que, sin duda, proporciona una posible herramienta para el almacenamiento de grandes volúmenes de datos. El almacenamiento en la nube o *cloud storage* es el espacio para acopiar datos, información, objetos digitales, y otros, que se acceden por internet a través de un servicio *web*, mediante un navegador como Explorer, Firefox, Chrome o Safari. Además de un aprovisionamiento de recursos informáticos bajo demanda, con control variable para el usuario y neutrales ante sistemas operativos (Sosinsky, 2011), estas características hacen único al almacenamiento en la nube. Hay que tener en cuenta que el almacenamiento puede ser brindado por un proveedor de servicios (nube pública) o una versión privada (nube privada); esta última es creada por una organización particular para su uso interno, con un completo control de los recursos en tecnologías de información.

El servicio de almacenamiento en la nube significa que un proveedor renta espacio en su centro de almacenamiento a usuarios finales que carecen de almacenamiento propio o no desean adquirirlo. También, se usa cuando no se dispone de personal técnico especializado en la administración de sistemas informáticos, o cuando se adolece de conocimiento para implementar y mantener infraestructura en almacenamiento.

Las tecnologías del *cloud computing* ofrecen principalmente tres modelos de servicio, de acuerdo al NIST (Mell y Grance, 2011): el primero es la infraestructura como servicio (IaaS), el segundo es la plataforma como servicio (PaaS) y, por último, el software como servicio (SaaS). No obstante, para fines de almacenamiento en la nube el más adecuado es la modalidad IaaS, a causa de que el proveedor ofrece al usuario recursos como capacidad de procesamiento, de

almacenamiento, o comunicaciones, que el usuario puede utilizar para ejecutar cualquier tipo de software, desde sistemas operativos hasta aplicaciones.

Una de las grandes ventajas del almacenamiento en la nube es el ahorro de recursos económicos. El almacenamiento se alquila a un proveedor utilizando el modelo de pago por gigabyte almacenado o pago por unidades de datos transferidos. Pues el usuario únicamente paga por la cantidad de datos que transfiere y aloja en los servidores del proveedor, además se tiene que el almacenamiento en la nube permite buena extensibilidad y escalabilidad en el almacenamiento de la información, necesario cuando se manejan grandes cantidades de datos (Joyanes Aguilar, 2012). La nube no tiene delimitaciones geográficas como los países, implicando que nuestra información puede acabar siendo deslocalizada en una o varias regiones del mundo, en uno de los centros de procesos de datos del proveedor de servicios.

3.1. Proveedores de servicios de almacenamiento en la nube

A continuación se identifican y describen los principales proveedores de almacenamiento en la nube.

- Amazon- Es un sitio web de comercio electrónico; no solo es la librería virtual más grande del mundo, sino que es una de los principales proveedores de servicios en la nube y, además de su eficiencia, presenta mejor garantía de servicio. El servicio de *Amazon Simple Storage Service* (Amazon S3) es una interfaz de servicios *web* que puede utilizarse para almacenar y recuperar grandes cantidades de datos, en cualquier momento y desde cualquier parte de la web. Es una infraestructura de almacenamiento de datos altamente escalable, fiable y de rápida transferencia. También, se distingue por el pago por uso y por consiguiente un ahorro a las organizaciones, con un precio por \$0.0300 por GB alojado al mes y con capacidad de renta hasta de 5000 TB (Amazon, 2015).
- IBM- Los servicios de *cloud computing* de IBM incluyen también el almacenamiento con *IBM Smart Business Storage Cloud*, el cual surgió por el crecimiento en los volúmenes de datos y la diversidad de formatos de archivo (IBM, 2015). Así esta solución permite que los usuarios tengan un acceso eficiente, rentable y experimenten una disminución del rendimiento e interrupciones. En este sentido, el proveedor brinda soporte para combatir las deficiencias en la gestión del almacenamiento, ayudando a validar requisitos de la red de área de almacenamiento SAN.
- Google- El gigante de la industria de la informática ofrece servicios a través de *Google App Engine*, que es una plataforma que ofrece construcción y alojamiento de aplicaciones *web* con la infraestructura de Google, donde solo se paga lo que se utiliza. Permite que los recursos informáticos sean fáciles de construir, mantener y escalar a medida que crecen las necesidades de almacenamiento y tráfico de web. La fijación de precios *Google Cloud Storage* se basa en una tarifa plana para su almacenamiento y una tasa de uso de la red. El uso del almacenamiento de proyectos y uso de banda ancha se calculan en gigabytes (GB), permitiendo alcanzar hasta más de 90TB de uso (Google, 2015).
- Windows- El principal monopolio de ofimática también nos ofrece capacidad de almacenamiento por medio de *Microsoft Azure*, que proporciona la:

flexibilidad de almacenar y recuperar grandes cantidades de datos sin estructurar, como documentos y archivos multimedia del tipo Blobs de Azure, datos estructurados basados en Nosql con Tablas de Azure, mensajes confiables con Colas de Azure; y utilizar Archivos de Azure basados en SMB para migrar aplicaciones locales a la nube... [Ofrece] hasta 500 TB de almacenamiento total por cuenta. (Microsoft, párr. 1 y 3, 2015).

Además, el Almacenamiento de Azure “replica automáticamente los datos para ayudar a protegerse frente a errores de hardware inesperados y tener la garantía de que estarán disponibles cuando se necesite” (Microsoft, párr. 4, 2015).

- Sun Microsystems (Oracle):

The Sun Cloud Storage Service provides Internet-based access to scalable, on-demand, pay-per-use file storage capabilities. When you register with the Sun Cloud, you are given an account which you can use to store files. Your account storage space grows and shrinks as you add and remove content. That is, you only use the space that you need. You can access the files in your storage account from any computer at any time. [El servicio de Sun Cloud Storage Services proporciona acceso basado en Internet para capacidades de almacenamiento de archivos escalables bajo demanda y pago por uso. El espacio de almacenamiento crece y se contrae al añadir y eliminar contenido. Es decir, sólo utiliza el espacio que el usuario necesita. Se puede acceder a los archivos desde cualquier computadora conectada a internet, en cualquier momento²]. (Oracle, párr. 1, 2015)

There are two ways to store your files in your account, file-based and object-based. The file-based storage enables you to create storage drives, called volumes, in which you can create a hierarchy of folders and files. The object-based storage enables you to create buckets in which you can place objects. [Hay dos formas de almacenar los archivos en la cuenta del cliente, basado en archivos y basado en objetos. El almacenamiento basado en archivos le permite crear unidades de almacenamiento, llamados volúmenes, en las que se puede crear una jerarquía de carpetas y archivos. Mientras en el almacenamiento basado en objetos permite crear cubos en el que se pueden colocar objetos³]. (Oracle, párr. 2, 2015)

- NetApp- Es una compañía donde “se desarrollan productos innovadores de software y sistemas de almacenamiento que ayudan a clientes de todo el mundo a almacenar, administrar, proteger y retener uno de los activos más preciados de la empresa: los datos” (NetApp, párr. 1, 2015). Para ello ofrecen un alojamiento de información con capacidad de 200 TB por cliente, además de ofertar una amplia gama de soluciones que va de mover datos a las nubes y entre ellas hasta proteger y recuperar datos (NetApp, 2015).

Además de los anteriores, existe un amplio número de proveedores de almacenamiento más pequeños, pero que de igual forman ofrecen espacio de almacenamiento. Entre estos podemos citar: *Openstack, Zip Cloud, Sugar Sync, ATMOS, GoGird, Rackspace, Arsys, Strato, IDrive, Open Drive, Mozy, Dropbox, Box, JustCloud y ADrive*. Todos ellos ofrecen precios competitivos y el pago por uso.

² Traducción propia.

³ Traducción propia.

Las nuevas formas de almacenamiento -en la nube- nos generan dudas sobre seguridad, privacidad e integridad de los datos gracias a que la información está deslocalizada y formando con frecuencia parte de las preocupaciones que se tienen sobre este modelo de almacenamiento. Si se pretende usar el almacenamiento en la nube hay que considerar los aspectos mencionados anteriormente y poner en balanza si aquella flexibilidad de almacenamiento vale la pena a pesar de los riesgos de seguridad que se corren.

4. Los sistemas de almacenamiento en las bibliotecas

Constantemente las unidades de información están usando las nuevas tecnologías de almacenamiento, desde la llegada de los archivos electrónicos se utilizó el microfilm, cintas magnéticas, discos ópticos, entre otros dispositivos diseñados para albergar información (Moorthy y Karisiddappa, 2000). Sin embargo, en las últimas tres décadas se han utilizado los discos duros tradicionales para contener el volumen de información digital, aunque poco a poco han sido sustituidos por memorias no volátiles y arquitecturas en sistemas de almacenamiento.

Por ejemplo, se sabe que los catálogos del sistema bibliotecario de la Universidad Nacional Autónoma de México (UNAM) tienen una arquitectura de almacenamiento del tipo SAN con fibra óptica, de ahí que cuenta con un espacio de 6TB para la colección de fondo antiguo, mientras que el catálogo de Scielo México se conforma por 600GB. En contraste, la colección de CLASE y Periódica tienen 1 TB disponible de almacenamiento, así como la que dispone SERIUNAM. Otro dato interesante es la cantidad de espacio destinada para la colección de tesis, que se contempla en 700GB.⁴ Por otra parte, el portal de revistas científicas y arbitradas de la UNAM rebasa 700 GB, considerando su base de datos y archivos, el cual sigue creciendo y pertenece a la infraestructura denominada "UNAM Cloud"⁵.

En esta misma línea, existen bibliotecas que usan soluciones en la nube. Por ejemplo, la Biblioteca Nacional de Australia (NLA) (Goldner, 2010), que ha utilizado la web para llevar a cabo dos tareas. La primera es la combinación de las colecciones de las bibliotecas australianas con otras colecciones australianas e internacionales importantes y fuentes de información de la Wikipedia; la segunda es abrir la mayor parte de este contenido para que el público pueda etiquetarlo, editarlo, recogerlo y revisarlo.

La explosión de los proyectos de digitalización en la última década ha impulsado esta reunión de información en nuevas direcciones de almacenamiento. Por ejemplo, el proyecto Hathi se caracteriza por ser una construcción de un depósito de libros digitalizados y revistas de las principales bibliotecas de investigación en los Estados Unidos. Otro proyecto que usa la tecnología del Cloud es OAISTER; dicho servicio, iniciado por la Universidad de Michigan y ahora gestionado por OCLC, busca cosechar todos los principales repositorios digitales de todo el mundo (Goldner, 2010).

⁴ Información proporcionada por el Ing. Mauricio Sánchez Reséndiz de la Secretaría Técnica de Informática y Telecomunicaciones, Dirección General de Bibliotecas, UNAM.

⁵ Información proporcionada por el Ing. Miguel Ángel Ortiz Camilo de la Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM.

También, la British Library a través de S3 y Nexsan han implementado soluciones en la nube con un espacio de 300 TB de almacenamiento para albergar materia digital de origen del Reino Unido y la digitalización de partes de sus propias colecciones (Solid State Solutions, 2015). En contraste, la Biblioteca del Congreso lanzó un programa piloto que junto con DuraCloud usarán la nube para probar el acceso perpetuo a los contenidos digitales. La misión del Programa de Preservación es el desarrollo de una estrategia nacional para recopilar y preservar los contenidos digitales disponibles, especialmente los materiales que se crean solo en formatos digitales, para las generaciones actuales y futuras (Allen y Morris 2014).

De forma general, las tecnologías de almacenamiento de información van cambiando constantemente de modo que el profesional de la información debe conocer la variedad que existe para seleccionar la solución de acuerdo a sus necesidades y entorno. Entre más información se tenga, será necesario crear diseños de sistemas de almacenamiento más complejos.

Por otra parte, el almacenamiento de conexión directa (DAS) y el almacenamiento conectado en red (NAS) nos ofrecerán un almacenamiento concentrado en un lugar geográfico para cantidades de información pequeñas y medianas. En cambio, la red de área de almacenamiento (SAN) y el almacenamiento en la nube proveerán un mayor espacio de almacenamiento, así como su acceso desde cualquier punto geográfico por medio de internet, aunque esto traiga consigo otros inconvenientes relacionados principalmente con la seguridad de la información, ya que al estar la información en varios servidores dispersos geográficamente y conectados al internet los hacen vulnerables cuando se carecen de sistemas de seguridad.

5. Conclusiones

De acuerdo a lo encontrado sobre las tecnologías de almacenamiento de información en el ambiente digital podemos concluir lo siguiente:

- Con el paso del tiempo, el ser humano ha creado distintas tecnologías de almacenamiento para guardar y recuperar su información. Desde mediados del siglo XX, la producción de información electrónica provocó la generación de distintos dispositivos de almacenamiento como las unidades de disco duro.
- Se tiene el registro que se producen en el mundo entre 3 y 5 *Zettabytes* de datos; por lo tanto, se intuye que la cantidad de información digital es inmensa; sin embargo, ignoramos su verdadera y enorme dimensión. Hay que tener en cuenta que los datos producidos se actualizan y expiran provocando que se carezca de una cifra exacta sobre la cantidad de datos generados, por lo cual el espacio de almacenamiento es variante.
- La tecnología del almacenamiento, por una parte, tiene que ser vista primero desde la perspectiva física de componentes (capacidad de espacio de unidades de discos duros tradicionales y sólidos, memorias PCM, cableado UTP y Fibra óptica, etc.); por otra parte, debe considerarse su diseño o arreglo del sistema de almacenamiento.
- Las tecnologías para almacenar datos de forma concentrada son: el Almacenamiento de Conexión Directa (DAS) y Almacenamiento Conectado en Red (NAS). Mientras que los sistemas de Red de Área de Almacenamiento (SAN) y el almacenamiento en la nube son

sistemas de almacenamiento distribuido, que ofrecen escalabilidad, amplia capacidad de almacenamiento y administración de grandes cantidades de datos.

- Con respecto al almacenamiento en la nube, se tienen inconvenientes de seguridad causados por la deslocalización de la información. El uso de un proveedor de servicios, integridad de los datos y fiabilidad del servicio de almacenamiento en la nube son elementos que deben ser evaluados por el profesional de la información para determinar si se usa un sistema distribuido o en la nube para almacenar grandes cantidades de datos.
- Los sistemas de almacenamiento deberán ser escalables, es decir, no estar limitados a una capacidad de espacio, debido a que la información crece con el tiempo.
- La tecnología de almacenamiento se encuentra en un constante estado de cambio; es fácil caer en la trampa de elegir la tecnología emergente de almacenamiento, pero la mejor opción será aquella que satisfaga nuestras necesidades además de que sea rentable con posibilidades de expansión y escalabilidad.

6. Referencias

Allen, E. y Morris, C. (2014). *Library of Congress and DuraCloud launch pilot program using cloud technologies to test perpetual access to digital content. News from the Library of Congress* [página web]. Recuperado de <http://www.loc.gov/today/pr/2009/09-140.html>

Amazon. (2015). *Amazon Simple Storage Service* [página web]. Recuperado de <http://aws.amazon.com/es/s3/>

Microsoft. (2015). *Microsoft Azure* [página web]. Recuperado de <http://azure.microsoft.com/es-es/>

Cunhe, L. (2002). Approach to the Information Storage Technologies under Network Environment. *Journal of the China Society for Scientific and Technical Information*, 1, 48-51.

Edelson, E. (2004). Security in network attached storage (NAS) for workgroups. *Network Security*, 4, 8-12.

Furht, B. y Escalante, A. (2011). *Handbook of Cloud Computing*. New York: Springer.

Glossbrenner, A. (1993) *File and disk management: From chaos to control*. México: Mcgraw-Hill.

Goldner, M. R. (2010). Winds of change: libraries and cloud computing. *BIBLIOTHEK Forschung und Praxis*, 34(3), 270-275.

Google. (2015). *Google App Engine* [página web]. Recuperado de <https://cloud.google.com/storage/>

Gopalakrishnan, K., Shenoy, R.S., Rettner, C.T., Virwani, K., Bethune, D.S., Shelby, R.M., Kurdi, B.N. (2010). Highly-scalable novel access device based on mixed ionic electronic conduction (MIEC) materials for high density phase change memory (PCM) Arrays. En *Symposium on VLSI Technology* (pp. 205-206). Hawaii, United State.

- Hilbert, M. y López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, 332(60), 60-65.
- International Business Machines Corp [IBM]. (2015). *IBM Smart Business Storage Cloud* [página web]. Recuperado de <http://www-935.ibm.com/services/us/en/it-services/storage-services/smart-business-storage-cloud/>
- Joyanes Aguilar, L. (2012). Computación en la nube: Notas para una estrategia española en Cloud Computing. *Revista del Instituto Español de Estudios Estratégicos*, (0), 87-110. Recuperado de http://publicaciones.defensa.gob.es/pprevistas/46f2896b-fb63-65ab-9bdd-ff0000451707/pubData/source/R_IEEE_0.pdf
- Kim, H., Seshadri, S., Dickey, C. L., y Chiu, L. (2014). Evaluating phase change memory for enterprise storage systems: A study of caching and tiering approaches. *ACM Transactions on Storage (TOS)*, 10(4), 15.
- Li, M. y Cao, S. (2014). A serie method of massive information storage, retrieval and sharing. En *Mechatronics and Automation (ICMA), 2014 IEEE International Conference on* (pp. 1171-1175). IEEE.
- Liu, S. J. y Yi, Z. J. (2012). Research of network mass storage mode based on san. En D. Jin y S. Kin (eds.), *Advances in Electronic Commerce, Web Application and Communication* (pp. 279-284). Springer Berlin Heidelberg.
- Lyman, P. y Varian, H. (2003). *How much information*. Estados Unidos: Universidad de California. Recuperado de <http://www2.sims.berkeley.edu/research/projects/howmuchinfo2003>
- Mell, P. y Grance, T. (2011). *The NIST definition of cloud computing: Recommendations of the National Institute of Standards and Technology*. Estados Unidos: U. S. Department of Commerce.
- Moorthy, A. L., y Karisiddappa, C. R. (2000). Mass Storage Technologies for Libraries & information Centres. *DESIDOC Journal of Library & Information Technology*, 20(5), 3-20
- NetApp. (2015). *NetApp - México* [página web]. Recuperado de <http://www.ats.avnet.com/lac/es-es/regions/Mexico/suppliers/Pages/Mexico-NetApp.aspx>
- Oracle. (2015). *Sun Cloud Storage Service User Quick Start Tutorial* [página web]. Recuperado de <http://www.oracle.com/technetwork/systems/usr-quickstart-138183.html#about>
- Sadlier, G. (2003). Storage Area Networks: An Information Security Perspective. *Information Systems Security*, 12(5), 29-41.
- Sakr, S., Liu, A., Batista, D.M. y Alomari, M. (2011). A survey of large scale data management approaches in cloud environments. *IEEE Communications Surveys and Tutorials*, 13(3), 311-336.
- Schadt, E. E., Linderman, M. D., Sorenson, J., Lee, L. y Nolan, G. P. (2010). Computational solutions to large-scale data management and analysis. *Nature Reviews Genetics*, 11(9), 647-657.

- Solid State Solutions. (2015). *Case study: British Library*. [página web]. Recuperado de <http://www.s3.co.uk/customers/british-library>
- Sosinsky, B. (2011). *Cloud Computing Bible*. New York: Wiley.
- Tian, L. (2006). The Selection and Application of Library Storage Technology Programs. *Library Work and Study*, 5, 28-30.
- Wang, R., Jiang, L., Zhang, Y. y Yang, J. (2015). SD-PCM: Constructing Reliable Super Dense Phase Change Memory under Write Disturbance. En *Proceedings of the Twentieth International Conference on Architectural Support for Programming Languages and Operating Systems* (pp. 19-31). Nueva York: ACM.
- Xu, J. y Zhu, S. (2006). Research on the Application of SAN in Digital Library. *Library and Information Service of Zhejiang Universities and Colleges*, 3, 20-21.
- Yang, C. T., Shih, W. C., Huang, C. L., Jiang, F. C. y Chu, W. C. C. (2014). On construction of a distributed data storage system in cloud. *Computing*, 1-26.
- Yoon, S. K., Bian, M. Y., y Kim, S. D. (2014). An integrated memory-disk system with buffering adapter and non-volatile memory. *Design Automation for Embedded Systems*, 1-18.
- Zhao, G. (2006). Research on Digital Library Data Storage Program. *Shanxi Library Journal*, 3, 31-34.



e-Ciencias de la Información

¿Desea publicar su trabajo?

[Ingrese aquí](#) O escríbanos a la siguiente dirección:
revista.ebci@ucr.ac.cr**2011****Origen: respuesta a una necesidad**

En el año 2011, la Escuela de Bibliotecología y Ciencias de la Información (EBCI) de la Universidad de Costa Rica (UCR) reconoció la importancia de crear **nuevas y mejores alternativas** para difusión de la investigación. **e-Ciencias de la Información** es la respuesta a un contexto actual marcado por una mayor **apertura, flexibilidad, y rigurosidad** en la verificación de los datos y su procesamiento.

2013**Revista de la UCR**

e-Ciencias de la Información es una **revista científica** que aborda las nuevas temáticas de desarrollo e investigación en las Ciencias de la Información, en el **ámbito nacional e internacional**. Así, colabora significativamente en el progreso de esta disciplina. Por sus parámetros de calidad, pertenece al grupo de **las revistas más importantes de la UCR** y se encuentra ampliamente indizada en los importantes catálogos.

En la actualidad**Posicionamiento internacional**

La revista admite trabajos en las siguientes áreas, siempre que se relacionen **directamente** con las Ciencias de la Información:

Bibliotecología, Documentación, Tecnologías de la Información y la Comunicación, Investigación, Análisis Estadísticos y Bibliometría, Archivística, Sistemas de Información, Informática, Comunicación y Bibliotecas Escolares.

