

# SDS: Camino a una estrategia de datos de nube híbrida

## Contenido

Contenido.....	1
Introducción.....	1
¿Qué es SDS?.....	1
¿Dónde se aplica SDS?.....	2
Indicadores clave (KPIs) de una implementación SDS exitosa.....	3
El marco SDS.....	3
SDS es el trampolín para la Nube Híbrida .....	4
IBM Spectrum Storage Suite.....	4
Ejemplo: Aprovechar el almacenamiento de nube como un nivel .....	5
Ejemplo: Proteger los datos entre múltiples medios....	5
Ejemplo: Construir la nube de almacenamiento propia, ser el Proveedor de Servicio Propio (SP).....	5
Recomendación.....	5

## Introducción

En esta edición de *Insight*, Neuralytix analiza las características del almacenamiento definido por software (*software-defined storage*, SDS) y su impacto en la economía impulsada por datos. En particular, en este documento se presenta una descripción completa de la solución SDS *IBM Spectrum Software Suite* y se analiza de qué modo aborda las necesidades de negocio de la empresa moderna.

Neuralytix estima que hacia 2020, el 30% de los nuevos proyectos informáticos netos se construirán en algún tipo de infraestructura convergente con software SDS. Nuestra posición sobre el SDS es firme: sencillamente creemos que todas las organizaciones deberían planear incluirlo en su estrategia de TI.

## ¿Qué es SDS?

Un error común con respecto al SDS es creer que es un *producto* singular: el software que se instala para obtener las características y funciones esperadas del SDS. En cierto sentido, esto es correcto. El SDS debe liberar a la empresa de tener que usar múltiples soluciones puntuales que a menudo vienen de múltiples proveedores.

Según investigaciones de Neuralytix, cerca del 100% de todas las organizaciones tienen algún tipo de almacenamiento formal (o bien “TI de las sombras”) dentro de la nube.

También estimamos que aprox. el 95% de todas las organizaciones grandes procuran tener una estrategia de almacenamiento híbrido que incluye almacenamiento basado sobre la nube, mientras que el 80% de los clientes del mercado mediano está analizando cómo integrar el almacenamiento en la nube (*in-cloud*) con el almacenamiento físico en instalaciones de la empresa (*on-premise*).

En todo caso, según nuestras estimaciones, aprox. el 55% de las organizaciones busca activamente lanzar una estrategia SDS dentro de los próximos 12 meses, y alrededor del 25% de las organizaciones desea comenzar a implementar una estrategia SDS dentro de los próximos 36 meses.

Conforme las organizaciones recurren a la nube y procuran reducir la dependencia de almacenamiento físico dedicado, el SDS proporciona flexibilidad y coherencia para cubrir distintos modelos de implementación con una experiencia operativa uniforme. Sin embargo, el SDS es, en realidad, un marco. Dentro de ese marco hay componentes móviles e interdependencias.

Ese marco incluye varias funciones:

- 🌐 Virtualización de almacenamiento *on-premise* e *in-cloud*;
- 🌐 Orquestación de todos los recursos existentes, incluso los que están basados sobre la nube;
- 🌐 Sistema de archivos escalable que puede extenderse a la nube;
- 🌐 Servicios de datos, como la presentación del almacenamiento en diversos protocolos. Por ejemplo, iSCSI, S3, HDFS, NFS, SMB, etc.;
- 🌐 Niveles (“tiers”) de datos dentro y entre almacenamiento *on-premise* e *in-cloud*;
- 🌐 Integración de datos *on-premise* e *in-cloud*; y
- 🌐 Calidad de servicio.

Para unir todas estas funciones con el fin de crear un “único software” cohesivo y percibido como tal, se necesita un proveedor con experiencia en todos los aspectos del marco SDS.

Una forma simple de considerar lo que hace el SDS es concebirlo como *virtualización de datos*. Esencialmente, el SDS “libera” los datos de la infraestructura de hardware. Al hacerlo, el negocio puede enfocarse en el valor de la información que puede extraerse de los datos.

La Figura 1 ilustra las interacciones básicas entre el SDS y la infraestructura de hardware que lo rodea:

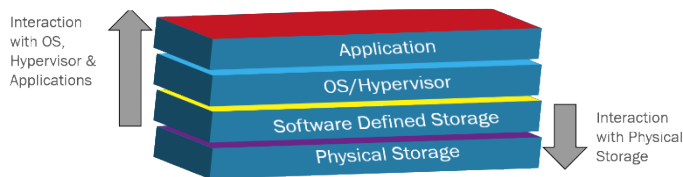


Figura 1: Diagrama de bloques de interacciones SDS (Neuralytix 2016) (Interacción con OS, hipervisor y aplicaciones / Interacción con almacenamiento físico)

Un elemento clave que es común a todo el software SDS es que puede, a través de un único panel, administrar todas las interacciones entre las aplicaciones y el sistema operativo o hipervisor, y la capacidad de almacenamiento físico.<sup>1</sup>

Como resultado, la gestión de datos es mucho más simple. El administrador ya no tiene que asociar aplicaciones específicas con sistemas de almacenamiento específicos que llevan a la formación de silos de datos y a una utilización subóptima de la capacidad global de almacenamiento. Además, en software SDS más avanzado, la analítica puede contribuir a asegurar que la colocación de datos sea tal que los acuerdos de nivel de servicio (SLA), medidos por ejemplo en IOPS o redundancia de datos (como código de eliminación) protejan los datos.

En otras palabras, en lugar de que la TI sea tecnología de *infraestructura*, verdaderamente puede ser tecnología de la *información*. Los resultados de negocio deben ser simples: crecimiento y una mayor ventaja competitiva.

## ¿Dónde se aplica SDS?

En el documento de Neuralytix titulado *A Buyer's Guide to Software-Defined Storage (SDS): Establishing a Sustainable Data Strategy* (mayo de 2016), Neuralytix recomendó a los clientes:

*“El SDS debe ser parte de la estrategia de toda organización de IT.”*

Reafirmamos también aquí esta recomendación.

Aparte de lo mencionado sobre “liberar” datos, el SDS tiene otras funciones igualmente críticas. Al ser definido por software (es decir, impulsado por software), el SDS puede permitir el aprovisionamiento y movimiento rápido de máquinas virtuales y hardware bare metal.

La capacidad de dar soporte a múltiples protocolos de acceso de datos permite el movimiento de datos entre métodos de acceso tradicionales, como SMB y NFS, y métodos de próxima generación, como HDFS y S3. Los últimos son críticos para aplicaciones que requieren escala masiva, como Hadoop y otras aplicaciones de Big Data.

El movimiento de datos no se limita a la transformación de *on-premise* a *on-premise*, sino que con protocolos como S3, permite que los datos fluyan fácilmente entre las modalidades *on-premise* e *in-cloud* para la capacidad de almacenamiento.

¡La liberación de datos (que es el resultado de poder tener un flujo libre de datos) puede acelerar el tiempo de falla! La frase “fracasa seguido, fracasa rápido” es válida aquí también. La idea es que a través de la experimentación, que




muchas veces no da el resultado deseado, la empresa pueda probar la mayor cantidad de situaciones posibles y llegar al estado deseado.

Además, a través de la experimentación, las empresas han encontrado nuevas oportunidades que antes no sospechaban que existían, y pueden abrir el camino a la innovación, que puede incluir productos o soluciones nuevas o mejoradas, un mejor servicio al cliente, mayores niveles de precios y mayor productividad.




Los datos liberados, cuando se usan adecuadamente, permiten a la empresa acelerar el tiempo de obtención de conocimientos y el tiempo de acción. Esto se traduce directamente en el logro de objetivos, ya se trate de crecimiento del negocio o de ventaja competitiva. El objetivo último es mejorar la salud general del negocio.

## Indicadores clave (KPIs) de una implementación SDS exitosa

Neuralytix cree que en toda tecnología de la información moderna, el éxito debe medirse en referencia a tres indicadores o KPIs:

-  Previsibilidad;
-  Repetibilidad y
-  Escalabilidad.

El departamento de TI debe enfocarse en la creación de valor a través de la Información. Como tal, la infraestructura debe exhibir atributos que demuestren su aptitud para la creación de valor, la ventaja competitiva y el crecimiento.

-  La infraestructura óptima es aquella en la que el resultado y el desempeño siempre son **predecibles**, sin importar que se agreguen o eliminen aplicaciones, nodos o capacidad de cómputo o almacenamiento.
-  Los procesos involucrados en la escala deben ser **repetibles**. Como se mencionó, cualquiera sea el cambio que se haga a la infraestructura, debe seguir un proceso repetible (preferentemente, automatizado), que arroje resultados también repetibles.
-  Idealmente, la infraestructura debería ser infinitamente **escalable**, o por lo menos previsible para los próximos cinco a diez años en cada dimensión: aplicación, cómputo, almacenamiento, etc.

Neuralytix cree que IBM Spectrum Storage Suite aborda los tres KPIs, para ayudar a los clientes a asegurar un SDS exitoso, e impulsar el crecimiento de negocio, la innovación y la ventaja competitiva.

## El marco SDS

Como se señaló, el SDS es un marco, no un único producto de software. Este marco tiene en su interior un conjunto de soluciones puntuales interdependientes que, cuando se reúnen, hacen que el SDS sea lo que es.

En el nivel más básico, el SDS debe tener virtualización de almacenamiento y un *pool* de almacenamiento escalable. Para que el SDS aporte valor a un centro de cómputo, es necesario que esté en condiciones de virtualizar todos los recursos de almacenamiento en el centro de cómputo. Pero proporcionar simplemente virtualización agregaría una capa innecesaria en una pila de almacenamiento ya compleja. Por lo tanto, para aportar valor a la pila de almacenamiento, el SDS *debe* ser capaz de presentar un sistema de archivos que aumente la escala a todos los recursos de almacenamiento disponibles. Idealmente, esto se hace sin movimiento de datos. El desafío para el SDS es si el sistema de almacenamiento *scale-out* es maduro, confiable y con características empresariales.

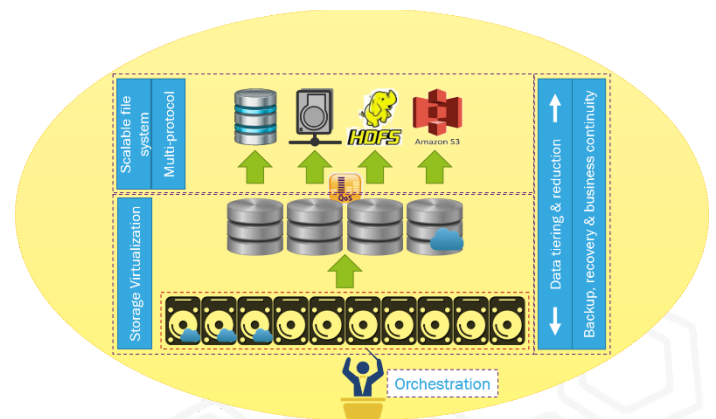


Figura 2: El Marco SDS (Neuralytix 2016) (Virtualización de almacenamiento, Sistema de archivos escalable, Niveles y reducción de datos, Resguardo, recuperación y continuidad de negocio, Orquestación)

Además de la virtualización de almacenamiento y el sistema de archivo *scale-out*, el SDS debe poder presentar los datos en la misma manera en que las aplicaciones los requieren. Cuando mayor es la diversidad de aplicaciones, mayor debe ser la cantidad de protocolos a los que el SDS da soporte.

Por lo tanto, si los datos virtualizados son en bloque, debe poder presentarlos como bloque, y lo mismo se aplica a los datos de archivo, datos de objeto, etc. Los servicios de datos que el software SDS debe proporcionar al nivel más básico incluyen la presentación de datos en bloque, archivo, S3 y protocolos HDF5.

Estas características son las más básicas que deberían esperarse del software SDS. Una forma de considerar este conjunto de características es en conjunto, para formar la primera generación de software SDS.

La maduración del software SDS trae una serie de características nuevas y más capaces.

Otros servicios de datos que el SDS idealmente debe proporcionar incluyen *data tiering*, deduplicación de datos, compresión de datos, resguardo y recuperación, *archiving*, y continuidad de negocio.

Todos los servicios de datos deben presentarse con un nivel de calidad compatible con las aplicaciones que tienen acceso a ellos.

La capa de orquestación y gestión del software SDS es un buen indicador de si el software SDS considerado es verdaderamente una solución SDS de próxima generación.

La orquestación y gestión de datos y almacenamiento no deben confinarse a soluciones *on-premise*. Deben abordar la tendencia hacia construir nubes híbridas. En otras palabras, la capa de orquestación y gestión SDS debe poder integrar, controlar y administrar en las modalidades *on-premise* e *in-cloud* (sea en nube pública o privada) a través de una interfaz singular, y tratar el almacenamiento basado sobre la nube como un nivel dentro de la estrategia de almacenamiento general.

La gestión idealmente debe ser accesible desde la nube, para administrar tanto el almacenamiento *on-premise* como *in-cloud*. Los datos deben poder viajar libremente y sin límite entre los destinos de almacenamiento fijos y de nube.

## SDS es el trampolín para la Nube Híbrida

Una vez resueltas estas cuestiones -todas las capas de abstracción, gestión, orquestación, etc.- el objetivo final es abstraer los datos de la infraestructura.

A menudo es lo que se conoce como movilidad de datos. La movilidad de datos permite que los datos viajen en forma dinámica del usuario final a los sistemas clave, desde los dispositivos móviles a sistemas de reporte y analítica. También permite que los datos viajen a y de la nube según sea necesario o deseable.


Conforme aumenta el procesamiento que tiene lugar en el entorno de nube y crece el volumen de datos que se ingiere de puntos de recolección de datos basados en la nube, las soluciones SDS necesitan incorporar movilidad de datos no solo dentro del centro de cómputo, sino también entre distintos centros de cómputo, apalancando las nubes públicas o privadas, así como entre destinos de almacenamiento *on-premise* e *in-cloud*.


Aparte del procesamiento de datos, la movilidad de datos también sirve para permitir a las organizaciones aprovechar la economía de los destinos de almacenamiento de datos *on-premise* e *in-cloud*. Por ejemplo, la mejor opción para almacenar datos temporales puede ser la nube, donde es más barato por un tiempo corto, en lugar de archivar a largo plazo, para el cual puede ser mejor el almacenamiento en cinta lineal, fuera de línea o en un lugar remoto.

## IBM Spectrum Storage Suite

IBM Spectrum Storage Suite es una oferta SDS singular. Mientras que la mayoría de las ofertas SDS comerciales se enfocan en dar soluciones puntuales centradas en bloque, archivo, objeto o máquina virtual, IBM Spectrum Storage Suite es un conjunto de soluciones que ofrece flexibilidad y agilidad al cliente, permitiéndole adquirir e implementar solo los componentes relevantes para sus necesidades, y reunir todo el soporte para múltiples aplicaciones, protocolos, cargas de trabajo y ubicaciones (*on-premise* o en la empresa e *in-cloud* o en la nube) que las organizaciones requieren de su software SDS.

IBM Spectrum Storage Software Suite tiene seis componentes:

 **Spectrum Accelerate:** ofrece un desempeño uniforme de datos y una red expansible de servidores, ideal para implementaciones rápidas de bases de datos que crecen, soporte de máquinas virtuales y cargas transaccionales.

 **Spectrum Archive:** permite gestionar los datos almacenados en cinta lineal de la misma manera que los datos almacenados en unidades magnéticas o de disco flash.



**Spectrum Control:** proporciona gestión unificada de SDS así como almacenamiento físico tradicional con tecnologías de *tiering* de datos y reducción de datos que reducen los costos de almacenamiento hasta 73%<sup>ii</sup>.

**Spectrum Protect:** proporciona capacidades de backup, snapshot y recuperación, con rapidez y escalable a través de toda la organización y la nube híbrida.

**Spectrum Scale:** trae un sistema de archivos maduro, de alto desempeño y alta escalabilidad, con soporte para objetos y HDFS.

**Spectrum Virtualize:** permite virtualizar cualquier recurso de almacenamiento. Protege y permite que las inversiones existentes se integren al marco SDS.

La investigación realizada por Neuralytix indica que cerca del 100% de las organizaciones tienen algún tipo de almacenamiento *in-cloud* ya sea formal o de TI en las sombras. En realidad, todas las organizaciones están tratando de encontrar el equilibrio entre almacenamiento *on-premise* e *in-cloud* de sus datos, y llegar a una estrategia de nube híbrida.

Sin IBM Spectrum Software, la mayoría de las organizaciones se encontrarán en una situación en la que podrían estar migrando a una solución SDS *on-premise* y a una solución *in-cloud* totalmente disociadas, lo cual dista de ser ideal.

En cuanto a la idea de movilidad de datos, los datos deben fluir libremente y con mínima fricción entre destinos *on-premise* e *in-cloud*. Para ello, se requiere una solución capaz de abstraer destinos de almacenamiento fijo y en la nube con una sola plataforma de gestión y orquestación. IBM Spectrum Storage Software Suite cumple este objetivo.

## Ejemplo: Aprovechar el almacenamiento de nube como un nivel

Para tener un nivel (*“tier”*) de nube en cualquier estrategia de almacenamiento, se debe contar con un sistema de archivo escalable compatible que se extienda del almacenamiento *on-premise* al almacenamiento *in-cloud*. IBM se encuentra exclusivamente capacitada para esto. IBM Spectrum Scale, uno de los componentes del marco SDS de IBM, es un sistema de archivos de alto desempeño y alta escalabilidad para implementaciones SDS físicas o en la nube. Esto permite a un cliente definir el

almacenamiento de nube como un nivel dentro de la estrategia general de niveles de datos.

Si los clientes prefieren una nube pública, el mismo software Spectrum Scale también puede aprovechar los protocolos S3 o Swift para pasar los datos a nubes públicas como Amazon Web Services (AWS) S3. De hecho, si se quiere, el almacenamiento basado en archivos *on-premise* puede convertirse en almacenamiento de nube basado sobre objeto (S3) al instante para permitir que los datos que antes se dirigían al centro de cómputo se aprovechen en la nube.

De eso se trata la verdadera movilidad de datos.

## Ejemplo: Proteger los datos en múltiples medios

Tradicionalmente, la protección de datos se logra mediante copias de resguardo (*backup*) a un destino local para una rápida recuperación y, en última instancia, a cinta lineal. Como IBM ha podido apalancar el almacenamiento de nube como nivel o *“tier”*, utilizando Spectrum Protect, los datos pueden resguardarse en múltiples destinos, que incluyen discos locales, la nube y cinta lineal. Ello da a los clientes flexibilidad y agilidad, al permitirles aprovechar el medio de resguardo disponible que se sea más efectivo en relación costo-beneficio, en vista de sus necesidades.

## Ejemplo: Construir la nube de almacenamiento propia, ser su propio Proveedor de Servicio (SP)

El objetivo de muchas organizaciones de TI es convertirse en el proveedor de servicio de sus usuarios (clientes). Proporcionar capacidad para almacenar datos valiosos que los usuarios puedan aprovechar es una capacidad crítica.

Como operación crítica, la provisión de almacenamiento debe cumplir el más alto nivel de KPIs descritos en términos de previsibilidad y repetibilidad, dando por sentada la escalabilidad.

Para hacer eso, IBM Spectrum Accelerate puede permitir a las organizaciones de TI tomar almacenamiento y servidores bare metal y presentar capacidades de almacenamiento en cuestión de minutos.

Spectrum Accelerate también puede operar *in-cloud* en Softlayer, la nube IaaS de IBM. Esto significa que el almacenamiento puede aprovisionarse en minutos para operaciones *on-premise* e *in-cloud* para aplicaciones basadas sobre la nube, todo con igual interfaz, proceso, orquestación y gestión.

Los Proveedores de Servicio y el área de TI también pueden apalancar Spectrum Accelerate para crear destinos de *backup* para clientes externos, porque Spectrum Accelerate fue diseñado con *multi-tenancy* en mente.

## Recomendación

Como lo indican los ejemplos, IBM Spectrum Software puede implementarse en una diversidad de casos de uso. Con frecuencia, es posible combinar múltiples casos de uso con los componentes apropiados de IBM Spectrum Software para lograr un entorno SDS de gran agilidad y alto desempeño.

Neuralytix cree que IBM Spectrum Software Suite se diferencia de otras ofertas SDS por ser una solución completa y ágil.

IBM tiene en cuenta todos los elementos necesarios que se esperan en cualquier solución de almacenamiento para una empresa ágil y escalable.

El conjunto de soluciones IBM Spectrum Software Suite aborda las necesidades multiprotocolo, multiaplicación, multi-ubicación y multicarga de trabajo de la empresa moderna. Reúne soluciones maduras en un formato que

permite agilidad de desarrollo y capacidades completas.

La capacidad de abordar almacenamiento en las dos modalidades, *on-premise* e *in-cloud*, es un atributo cada vez más solicitado por empresas que entienden la economía y los beneficios de la capacidad de almacenamiento en la nube.

El SDS no debería ser una solución puntual. Debe ofrecer un conjunto abarcador de servicios de datos y almacenamiento. IBM lo hace, y muy bien.

Neuralytix recomienda a sus clientes IBM Spectrum Suite como una opción adecuada a la hora de definir su estrategia de almacenamiento.

### Notas

<sup>i</sup> Extraído de [A Buyer's Guide to Software-Defined Storage \(SDS\): Establishing a Sustainable Data Strategy](#) (Neuralytix, Mayo de 2016)

<sup>ii</sup> Basado sobre cálculos de IBM

#### Sobre Neuralytix

Neuralytix es la firma de asesoramiento, consultoría e investigación de mercado de TI líder a nivel global. Neuralytix se basa en tres principios

- Los análisis cuantitativos de distintas tecnologías no representan la visión contemporánea o futura del mercado;
- Debemos considerar dominios de tecnología en lugar de tecnologías aisladas; en otras palabras, un enfoque holístico del problema de negocio que se intenta resolver, y
- Nuestro rol primario es ayudar a nuestros clientes a alcanzar sus resultados de negocio.

Entre nuestros clientes están los proveedores de TI más reconocidos, así como proveedores emergentes y privados. Nuestra sección de servicios a usuarios finales sigue creciendo agresivamente, junto con nuestra oferta de Inversores Neuralytix, destinada a influenciadores financieros. Nuestros clientes proveedores que cotizan en bolsa tienen una capitalización de mercado combinada de aprox. US\$4 billones; con influencia de aprox. US\$1,5 billones en gasto anual de TI.

Nuestros clientes valoran la relación con Neuralytix porque les da la ventaja de ser los primeros, y les ayuda a alcanzar los resultados de negocio deseados, a través del principio central de todo lo que hacemos, que es entregar Estrategias para el mañana.

#### Contacto

Neuralytix®  
755 Sansome Street  
Suite 360  
San Francisco, CA 94111

- Tel: (415) 919-1188
- Email: [info@neuralytix.com](mailto:info@neuralytix.com)
- Twitter @Neuralytix

#### Oficinas Globales:

- Australia
- Brasil
- Hong Kong, SAR
- Alemania
- Reino Unido