

Gentileza de



IBM Platform Computing Edition

Almacenamiento definido por software FOR DUMMIES[®] A Wiley Brand

Aprenda a:

- Controlar los costos de almacenamiento
- Eliminar cuellos de botella
- Usar IBM GPFS para resolver desafíos de gestión de almacenamiento

Lawrence C. Miller
CISSP

Scott Fadden



Baje la copia completa de *Software Defined Storage for Dummies*

<https://ibm.biz/SDSforDummies>



Software Defined Storage For Dummies®, IBM Platform Computing Edition

BgT[USVa bad
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

Copyright © 2014 by John Wiley & Sons, Inc., Hoboken, New Jersey

N[Yg` SbSdVWVWSbgT[USJo` bgVWVWdVhVg[VS.S` SWVSWSWg` efW SSWWVgbWSUo` a
fd`e [fVSWV[Yó` _ aVa a bad` [Yó` _ Wjal VWfcb` [lál _ Wá [lál Xafalab]Sel dV[efcal dS` a Wafch ffbal
VUMfa eYó` `a bW [fW`Se eWUJa` W#) a #`* WWS>WVW6 WUz aeW3 gfdVWae 7eSVaeG` [VaeWV
3_ éclS W#+) (1 e[` WbW [ea bclMa badWclfa WVa 7Vfadalz` aebW[VaeVW6W [ea S`á 7Vfadal WVV
V(dYdVW`6 VdSdS_ Wfa WbW [e ad John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030,
(201) 748-6011, fax (201) 748-6008, o <http://www.wiley.com/go/permissions>.

? SdSeLá WLSW WileY, For Dummies, WYa Dummies Man, The Dummies Way, Dummies.com,
Making Everything Easier k`Se[_ aYWWUa_ WLSWdVWUJa` SVSea` _ SdSeLá WLSW_a_ SdSe
Lá WLSWdVWfcbVSeVWohn Wiley & Sons, Inc. kla ege XISWVW` ae 7eSVaeG` [VaeW3_ éclS kafche
bSeW k` a bgWVW gff[clSdVW` bW [ea WclfazIBM, WYa IBM k GPFS ea` _ SdSeLá WLSW_a_ SdSe
Lá WLSWdVWfcbVSeVWInternational Business Machines Corporation. FalSe`SeVW/ ae_ SdSe
Lá WLSW@a` bchb[VSVWVge dVWf[hae bchb[VScdae John Wiley & Sons, Inc., `a WfaSeaUSVa Ua`
` [Yó` bclVgfa` [bclhWad_ WUJa` Svá WVVWV[clz

>E7:E7 67 D7EBA@E34;>:636 IE@ 93D3@FEB; >3 76:FAD;3> K 7> 3GFAD @A
OTORGA@ 675>3D35;0@ @; 93D3@FEB 3>9G@3 D7EB75FA 67 >3 7J35F;FG6 7
:@F79D;636 67 >AE 5A@F7@;6AE 67 7EF7 FD343-A K7EB75EB 53? 7@F7
DENEGAN FA63 93D3@FEB1 ;@5>GEA1 E@ >.;? :F35;0@1 93D3@FEB 67
3675G35;0@ B3D3 G@ GEA B3DF;5G-3DZ @A E7 5D73DA @; 7JF7@67DA
93D3@FEB 3>9G@3 BAD ? 3F7D;3>7E BDA? A5;A@3>7E A 67 H7@F3ZEL
ASESORAMIENTO Y LAS ESTRATEGIAS DEL PRESENTE PUEDEN NO SER ADECUADOS PARA
TODAS LAS SITUACIONES. ESTE TRABAJO SE VENDE ENTENDIENDO QUE LA EDITORIAL NO
TIENE POR FINALIDAD PRESTAR SERVICIOS LEGALES, CONTABLES U OTRO TIPO DE
SERVICIOS PROFESIONALES. SI SE REQUIERE ASISTENCIA PROFESIONAL, DEBEN BUSCARSE
LOS SERVICIOS DE UN PROFESIONAL COMPETENTE. NI LA EDITORIAL NI EL AUTOR SERÁN
RESPONSABLES DE LOS DAÑOS QUE PUDIERAN SURGIR DEL PRESENTE. EL HECHO DE QUE SE
MENCIONE UNA ORGANIZACIÓN O UN SITIO WEB EN ESTE TRABAJO COMO CITA Y/O
POTENCIAL FUENTE DE INFORMACIÓN ADICIONAL NO SIGNIFICA QUE LA EDITORIAL O EL
AUTOR APRUEBEN LA INFORMACIÓN QUE LA ORGANIZACIÓN O EL SITIO WEB PUDIERA
PROPORCIONAR NI LAS RECOMENDACIONES QUE PUDIERA HACER. ASIMISMO, SE INFORMA A
LOS LECTORES QUE LOS SITIOS WEB INCLUIDOS EN ESTE TRABAJO PUEDEN HABER CAMBIADO O
DESAPARECIDO EN EL PERÍODO TRANSCURRIDO DESDE SU ELABORACIÓN HASTA SU LECTURA.

Para ver información general sobre nuestros productos y servicios, o información sobre cómo crear su propia edición personalizada "for Dummies" para su negocio u organización, contacte a nuestro Departamento de Desarrollo de Negocios en EE.UU. al 877-409-4177, o escriba a: info@dummies.biz, o visite: www.wiley.com/go/custompub. Para ver información sobre licencias de la marca For Dummies para productos o servicios, contacte a BrandedRights&Licenses@Wiley.com.

ISBN: 978-1-118-84178-5 (pbk); ISBN: 978-1-118-84314-7 (ebk)

Fabricado en EE.UU.

10 9 8 7 6 5 4 3 2 1

Contenido

Introducción	1
Capítulo 1: Almacenamiento 101	3
Acceso a datos y desafíos de gestión	4
Tres funciones importantes del almacenamiento	6
Definiendo tipos de almacenamiento	7
Tecnologías de disco duro y SSD	13
Cluster File Systems	15
Capítulo 2: ¿Qué es almacenamiento definido por software?	17
Definiendo almacenamiento definido por software.....	17
Beneficios clave del almacenamiento definido por software.....	18
Soporte para archivos, bloques y objetos.....	21
Capítulo 3: Profundizando en IBM GPFS	23
Explicación de conceptos GPFS	23
Eliminación de los cuellos de botella de datos.....	24
Simplificación de la gestión de datos.....	26
Configuraciones de clusters	28
Cómo compartir datos entre Clusters GPFS	31
Gestión del ciclo de vida completo a menor costo.....	35
Soluciones empresariales basadas en GPFS	36
Capítulo 4: Conociendo IBM GPFS Storage Server	39
Entrega de una solución principio a fin	40
Cómo llevar el rendimiento paralelo hacia la tolerancia a fallas.....	42
Integración de los conceptos.....	44
Capítulo 5: Almacenamiento definido por software en el mundo real	45
Educación superior	45
Energía.....	47
Análisis de ingeniería.....	50
Ciencias de vida.....	53
Medios y entretenimientos.....	55
Investigación y desarrollo.....	57
Capítulo 6: Diez formas de utilizar el almacenamiento definido por software	59

Capítulo 2

¿Qué es el almacenamiento definido por software?

En este capítulo

- ▶ Cómo armar las piezas del almacenamiento definido por software
 - ▶ Reconociendo los beneficios del almacenamiento definido por software
 - ▶ Introducción del almacenamiento definido por software GPFS
-

El almacenamiento definido por software es un concepto relativamente nuevo en la industria de computación y el almacenamiento, y puede referirse a distintas tecnologías e implementaciones. Forma parte de una tendencia de industria más amplia que incluye las redes definidas por software (*software defined networking / SDN*) y los centros de datos definidos por software (*software defined data centers / SDDC*). Este capítulo explica exactamente de qué se trata el almacenamiento definido por software.

Definiendo el almacenamiento definido por software

En su nivel más básico, *software defined storage* o almacenamiento definido por software es almacenamiento de clase empresarial que utiliza hardware estándar, con todas las funciones importantes de almacenamiento y gestión realizadas dentro de un software inteligente. El almacenamiento definido por software ofrece servicios de almacenamiento automatizados, basados en políticas y con conciencia de aplicaciones, a través de la orquestación de la infraestructura de almacenamiento subyacente, con soporte para un entorno general definido por software. El hardware estándar incluye:

- ✓ **Almacenamiento en disco**, como SAN, NAS y disk arrays o JBOs (conjunto de discos, o "just a bunch of disks")
- ✓ **Dispositivos de red**, como switches e interfaces de red
- ✓ **Servidores** para procesamiento, administración y gestión del almacenamiento

Características adicionales que el almacenamiento definido por software puede incluir:

- ✓ Administración automatizada, basada en políticas para funciones de gestión, tales como gestión del ciclo de vida de la información (ILM) y aprovisionamiento
- ✓ Virtualización del almacenamiento
- ✓ Planos separados de control y datos para administrar la infraestructura de almacenamiento y los datos en la infraestructura de almacenamiento, respectivamente
- ✓ Arquitectura con escalabilidad scale-out masiva

Estas características se diferencian de los sistemas de almacenamiento tradicionales que dependen fuertemente de controladores basados en hardware personalizados para realizar funciones de almacenamiento. Los sistemas NAS, DAS y SAN (que se comentan en el Capítulo 1) son ejemplos de sistemas de almacenamiento típicos basados en hardware que utilizan controladores RAID especiales y firmware personalizado no portable para desempeñar sus funciones de almacenamiento.

Beneficios clave del almacenamiento definido por software

Las empresas hoy reconocen muchos beneficios significativos en el almacenamiento definido por software en sus centros de datos. Entre otros: mayor flexibilidad, gestión automatizada, eficiencia de costos y escalabilidad sin límites.

Mayor flexibilidad y agilidad

Las plataformas de almacenamiento empresarial tradicionales, tales como SAN y NAS (comentadas en el Capítulo 1) en general se basan sobre sistemas propietarios y vienen con un alto costo total de propiedad (TCO). Las soluciones SAN suelen requerir el uso de switches SAN costosos y complejos, conjuntos de almacenamiento y otros componentes propietarios.

Los dispositivos NAS son relativamente baratos pero tienen una escalabilidad limitada. Cuando uno se queda sin espacio en un dispositivo NAS, simplemente comienza a agregar más. Sin embargo, no se trata de una verdadera capacidad scale-out porque cada NAS individual se presenta como un almacenamiento separado e individual que se maneja aparte.

Una solución de almacenamiento definido por software aumenta la flexibilidad, al permitir a las organizaciones utilizar hardware estándar propietario y, en muchos casos, aprovechar la infraestructura de almacenamiento existente como parte de su solución de almacenamiento empresarial. Además, las organizaciones pueden lograr una escala masiva con una solución de almacenamiento definido por software, sumando componentes de hardware heterogéneos e individuales según sea necesario, para aumentar la capacidad, y mejorar el rendimiento de la solución.

Uso inteligente de recursos y gestión automatizada

La gestión automatizada y sobre la base de políticas que ofrecen las soluciones de almacenamiento definidas por software ayudan a impulsar las eficiencias operativas y de costos. Por ejemplo, el almacenamiento definido por software maneja funciones de almacenamiento importantes, como ILM, caché de disco, snapshots, replicación, striping y clustering. En suma, estas capacidades del almacenamiento definido por software le permiten poner los datos correctos en el lugar correcto, en el momento correcto y con el rendimiento y el costo correcto, automáticamente.

Eficiencia de costos

En lugar de usar un costoso hardware propietario, el almacenamiento definido por software utiliza hardware estándar para reducir sustancialmente tanto los costos de adquisición como el costo total de propiedad (TCO) para una solución de almacenamiento de clase empresarial. El software en una solución de almacenamiento definida por software se basa sobre estándares y administra la infraestructura de almacenamiento, así como los datos dentro del sistema de almacenamiento.

En muchos casos, las organizaciones pueden aprovechar sus inversiones actuales en infraestructura de almacenamiento, redes y servidores para implementar una solución de almacenamiento definida por software que es sumamente eficiente en función de costos.

DATOS TÉCNICOS: En un informe de julio 2012, Gartner, Inc. reveló que el costo de adquisición promedio por gigabyte de los sistemas de almacenamiento tradicionales multi-tier estaba en el rango de \$0,9/GB a \$ 5/GB. En comparación, las soluciones de almacenamiento definido por software promedian \$0,4/GB.



Escala de datos elástica sin límites

A diferencia de los sistemas tradicionales de almacenamiento, como SAN y NAS, el almacenamiento definido por software le permite aumentar la escala (scale out) con hardware estándar relativamente barato, en tanto que puede seguir administrando el almacenamiento como un único sistema de clase empresarial. Al aumentar la escala de la infraestructura de almacenamiento, el rendimiento y la rentabilidad siguen mejorando. Por ejemplo, IBM General Parallel File System (GPFS) ofrece mejoras de órdenes de magnitud en rendimiento I/O a medida que se agrega hardware, en comparación con NAS convencional (Figura 2-1). GPFS es el tema de un análisis más detallado en el Capítulo 3.

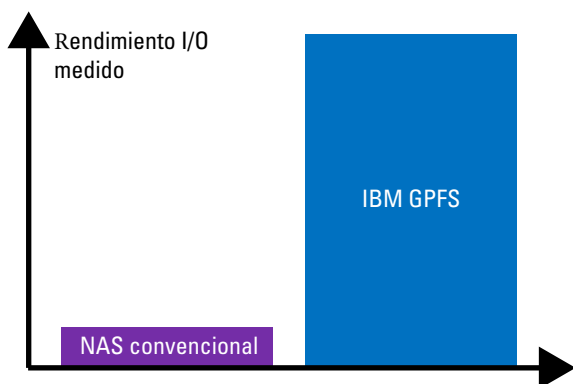


Figura 2-1: IBM GPFS ofrece rendimiento I/O extremo.

El almacenamiento definido por software ofrece escalabilidad masiva y casi ilimitada. Por ejemplo, IBM GPFS da soporte a:

- ✓ tamaño de sistema de archivos máximo: un millón de yottabytes
- ✓ 2^{63} (o aprox. 9 trillones) de archivos por sistema de archivos
- ✓ IPv6
- ✓ 1 a 16.384 nodos en un cluster



Un yottabyte equivale a un billón de terabytes.

Soporte para archivos, bloques y objetos

El almacenamiento definido por software en general se refiere a software que maneja la creación, colocación, protección y recuperación de datos. IBM GPFS es una solución de almacenamiento definido por software de clase empresarial que se ejecuta sobre plataformas IBM y de terceros y que está disponible como software que puede implementarse sobre una variedad de sistemas de hardware commodity y como parte de un "appliance" integrado, el GPFS Storage Server (GSS), que se describe en el Capítulo 4.

Los desarrolladores de grandes aplicaciones a escala Cloud han expresado un interés particular en el almacenamiento definido por software. Para ellos, las pesadas soluciones existentes controladas por hardware simplemente no ofrecen escalabilidad, tienen costos prohibitivos y son demasiado inflexibles para aumentar la capacidad en forma dinámica para los datos de aplicaciones necesarias para responder a sus necesidades de negocio en el futuro. Muchos de ellos han enfocado su desarrollo en OpenStack, la plataforma de computación de nube de fuente abierta para nubes públicas y privadas. En el recuadro "Foco en OpenStack" se presenta información más detallada.

RECUERDE: Para desarrolladores OpenStack, GPFS puede unificar el almacenamiento con una forma común de almacenar imágenes VM, dispositivos de bloque, objetos y archivos. El almacenamiento definido por software permite este tipo de integración. GPFS funciona como GPFS Native RAID (GNR) y la colocación de datos sobre la base de políticas le ofrece la flexibilidad de poner los datos en la mejor ubicación en el mejor nivel (tier) (en cuanto a rendimiento y costo), en el momento correcto. El almacenamiento definido por software permite la implementación sobre hardware estándar de industria heterogéneo, como se muestra en la Figura 2-2.

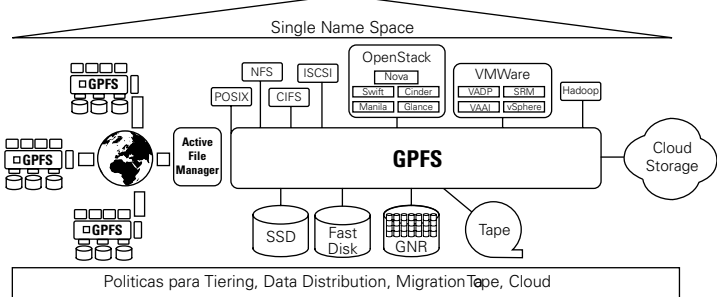


Figura 2-2: GPFS ofrece un plano de almacenamiento común.

Foco en OpenStack

OpenStack tiene una arquitectura modular con diversos componentes, entre otros:

✔ **OpenStack Compute (Nova):**

Un controlador de cloud computing fabric

✔ **Block Storage (Cinder):** Ofrece dispositivos de almacenamiento de nivel de bloque persistentes

✔ **Object Storage (Swift):** Sistema de almacenamiento redundante escalable

Con OpenStack, usted puede controlar pools de recursos de procesamiento, almacenamiento y redes en todo un centro de datos. Y aunque OpenStack ofrece versiones de fuente abierta de almacenamiento de bloques y objetos, muchos desarrolladores OpenStack identificaron la necesidad de almacenamiento más robusto para respaldar aplicaciones a escala Cloud. Si bien muchos desarrolladores OpenStack sienten que pueden diseñar la arquitectura para evitar las limitaciones de capacidades de cómputo y robustez de OpenStack, el almacenamiento tiene un "estándar mucho más alto" en cuanto a resiliencia y confiabilidad.

Respondiendo a la necesidad de un robusto almacenamiento definido por software, el release OpenStack "Havana" incluye un driver OpenStack Block Storage Cinder para IBM GPFS, que proporciona a los arquitectos que construyen nubes públicas, privadas e híbridas acceso a las características y capacidades del sistema de almacenamiento definido por software líder de la industria para el segmento empresarial. Y Cinder es tan sólo el comienzo.

La visión que tiene IBM para GPFS y

OpenStack es crear un único plano scale-out para todo el centro de datos o para múltiples centros de datos conectados en el mundo.

GPFS unifica imágenes, dispositivos de bloque, objetos y archivos OpenStack VM con soporte para Nova, Cinder, Swift y Glance, junto con interfaces POSIX como NFS y CIFS para integrar aplicaciones legacy. La capacidad de usar un único sistema de archivos GPFS para manejar volúmenes (Cinder), imágenes (Glance), sistemas de archivos compartidos (Manila), y usar clones de archivos para compartir datos con eficiencia y rapidez dentro y entre componentes será una gran ventaja para desarrolladores de aplicaciones a escala Cloud que utilizan OpenStack.

La robustez y las características de GPFS combinadas con extensiones de objetos OpenStack Swift pueden proporcionar un almacenamiento extenso de objetos de grado empresarial con alta eficiencia de almacenamiento, integración de cinta, replicación de área amplia, tiering transparente, checksums, snapshots y ACLs: capacidades que la mayoría de las ofertas de almacenamiento basado en objetos hoy no pueden igualar. OpenStack sobre GPFS ofrece eficiencias atractivas en una única solución de almacenamiento unificada que puede dar soporte a objetos y acceso a archivos a los mismos datos con protección de datos GPFS Native RAID. El almacenamiento de objetivos OpenStack Swift en GPFS puede reducir la cantidad de almacenamiento bruto que se necesita usar, en comparación con sistemas de almacenamiento de objetos que dependen estrictamente de la replicación.