

Informe técnico

# El valor económico de Data Domain

Validación de la rentabilidad de las soluciones Data Domain

Por Vinny Choinski, analista ejecutivo de ESG Lab;  
y Jason Buffington, analista principal

Mayo de 2017

Este informe técnico de ESG fue encargado por Dell EMC y se distribuye con licencia de ESG.

## Tabla de contenido

Tabla de contenido.....	2
Introducción.....	3
Beneficios de la arquitectura de Data Domain .....	4
Beneficios económicos de Data Domain .....	6
La eficiencia de la deduplicación es importante.....	7
Velocidad y escalabilidad.....	8
Consideraciones relativas al ahorro de costos.....	9
La gran verdad.....	11

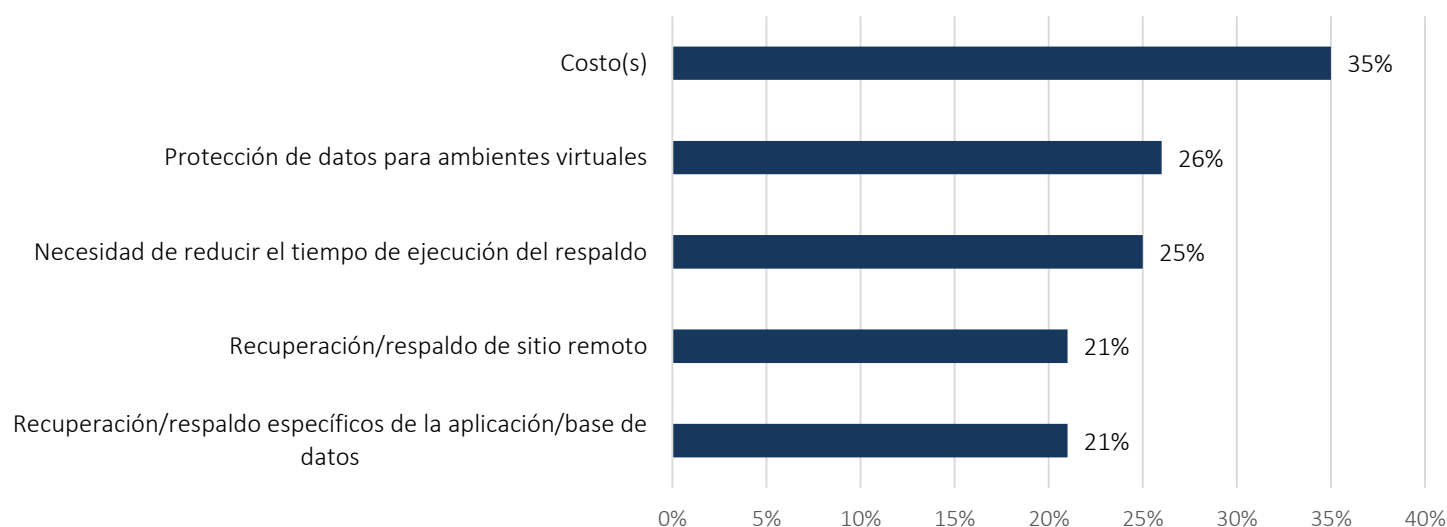
## Introducción

A fin de superar a la competencia en innovación y avances, las organizaciones deben seguir una ruta coherente para mantener su infraestructura moderna. La TI se encuentra bajo constante presión para ofrecer una infraestructura optimizada para las nuevas iniciativas de negocio y las aplicaciones de soporte, y, a la vez, intentar contener o incluso reducir los costos. De hecho, los encuestados de la investigación continua de ESG señalan, de manera unánime, que la reducción de los costos es el principal generador de negocios que afecta los gastos de TI. Cuando se les preguntó en una encuesta reciente acerca de cómo las organizaciones planeaban contener los costos el siguiente año, el 27 % de los encuestados respondió que lo haría mediante la compra de nuevas tecnologías con mejor ROI.<sup>1</sup>

Para desglosar un grupo específico, el 35 % de los administradores de TI encargados de la implementación de tecnologías y procesos de protección de datos hace referencia, de manera habitual, a los costos como uno de sus retos principales (consulte la figura 1). Este grupo parece considerar especialmente problemáticos los problemas centrados en las cargas de trabajo, sobre todo los retos relacionados con la protección de ambientes virtualizados y oficinas remotas. Sin embargo, cuando examinamos la clasificación general de los retos de protección de datos, surge un asunto más grande. Los costos y la virtualización son las principales preocupaciones, pero las siguientes tres consideraciones que se mencionan son de carácter operacional y funcional. Se trata de retos relacionados con el rendimiento, el respaldo y la recuperación de arquitectura distribuida, y la protección de aplicaciones y bases de datos.<sup>2</sup>

**Figura 1. Retos tecnológicos y procesos de protección de datos actuales**

**¿Cuáles de las siguientes opciones definiría como retos relacionados con las tecnologías y los procesos de protección de datos actuales de su organización? (Porcentaje de encuestados, N=387)**



Fuente: Enterprise Strategy Group, 2017

Estos retos son coherentes, además, con los mandatos principales de los líderes de TI, los cuales están relacionados con mejores niveles de velocidad, agilidad, confiabilidad y control de costos, todos aspectos que se informaron, de manera coherente, en informes anteriores de ESG sobre la modernización de la protección de datos. El análisis de ESG confirma que las incoherencias entre aquello con lo que lidiaban los implementadores y aquello que los líderes ordenaban solía dar

<sup>1</sup> Fuente: Informe de investigación de ESG, [Encuesta sobre las intenciones de gasto en TI de 2017](#), marzo de 2017.

<sup>2</sup> Fuente: Informe de investigación de ESG, [Modernización de la protección de datos](#), que se publicará próximamente.

lugar a una escasa transformación o modernización real. Los implementadores no eran capaces de actuar sobre los mandatos de los líderes debido a impedimentos técnicos en sus ambientes de protección de datos.<sup>3</sup>

Este informe analiza cómo los sistemas Data Domain, y las soluciones integradas incorporadas en la arquitectura de Data Domain, como el dispositivo de protección de datos integrado de Dell EMC, permiten brindar la agilidad que los implementadores de TI necesitan para transformar sus infraestructuras a fin de cumplir con los mandatos de los líderes de TI, solucionar los retos de protección de datos actuales y, lo más importante, reducir los costos del almacenamiento de datos de protección. Los hallazgos de ESG se basan en una auditoría y en un análisis de los indicadores clave del rendimiento (KPI): datos reales de los sistemas implementados, incluidos el precio de compra original, los costos del ambiente, la capacidad/utilización y el rendimiento.

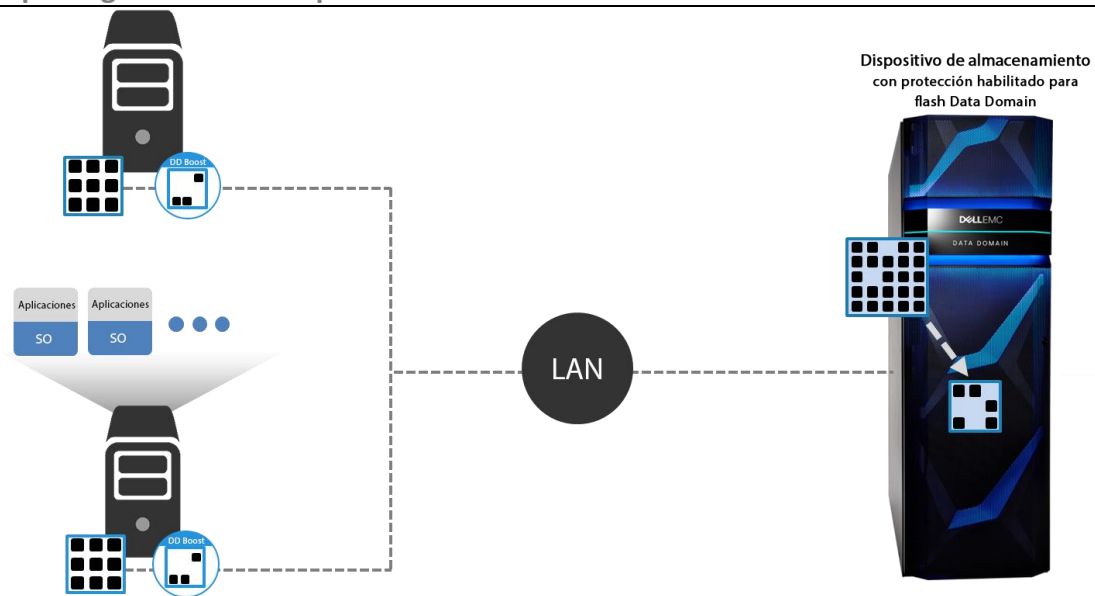
## Beneficios de la arquitectura de Data Domain

En 2006, EMC decidió no agregar el código de deduplicación de datos a las soluciones de protección de datos existentes, lo que provocó un cambio en su estrategia de protección de datos. Este cambio comenzó con el aprovechamiento de la tecnología de dos adquisiciones clave. La primera compra proporcionó una tecnología de deduplicación basada en el origen, y la segunda fue Data Domain, cuya tecnología de deduplicación está basada en el destino. Las tecnologías de estas dos adquisiciones se convirtieron en componentes fundamentales de las soluciones de protección de datos de EMC. Actualmente, bajo la marca Dell EMC, los sistemas Data Domain, incluido el dispositivo de protección de datos integrado, son dispositivos de protección de datos especialmente diseñados, completamente integrados y habilitados para flash que fueron diseñados para reducir la cantidad de almacenamiento en disco necesaria para conservar y proteger los datos. Gracias a la integración nativa en la arquitectura de ambos tipos de deduplicación de datos, basada en el origen y basada en el destino, estos sistemas hacen lo posible por completar más respaldos en menos tiempo, proporcionar restauraciones más rápidas y confiables, y reducir la cantidad de capacidad de almacenamiento necesaria para la protección de datos. La generación más reciente de sistemas Data Domain introdujo la inclusión de la tecnología flash para metadatos, lo que permite un rendimiento a escala, además de una restauración y un acceso instantáneo 20 veces más rápidos a las máquinas virtuales directamente desde Data Domain, en comparación con las generaciones anteriores.

---

<sup>3</sup> Fuente: Informe de investigación de ESG, [Tendencias en la modernización de la protección de datos de 2015](#), septiembre de 2015.

Figura 2. Descripción general de la arquitectura de Data Domain



Fuente: Enterprise Strategy Group, 2017

Entre las principales características arquitectónicas de Data Domain, se incluyen las siguientes:

- **Stream-Informed Segment Layout (SISL):** SISL permite que los sistemas Data Domain lleven a cabo el 99 % de los procesos de deduplicación en la CPU y la memoria RAM. Esto significa que los sistemas no dependen de la cantidad de discos para aumentar el rendimiento.
- **Segmentación de longitud variable:** a fin de obtener tasas óptimas de deduplicación, los sistemas Data Domain aprovechan la segmentación de longitud variable para dividir los flujos en función de la estructura natural de los datos. Esto permite al sistema determinar si el segmento es único antes de comprimirlo y almacenarlo.
- **Deduplicación en línea:** los sistemas Data Domain realizan la deduplicación en la CPU y en la memoria a medida que el sistema recibe el flujo de respaldo. Esto significa que solo los datos únicos se envían y se almacenan en el disco. De esta manera, se descarta la necesidad de utilizar un área de almacenamiento provisional en disco y recursos de procesamiento para la deduplicación posterior al proceso.
- **Software Data Domain Boost:** el software Data Domain Boost distribuye las partes del proceso de deduplicación a los clientes de aplicaciones o al servidor de respaldo. Con DD Boost, solamente los datos únicos recorren la conexión entre el servidor de respaldo o los clientes y el sistema Data Domain. Esto ayuda a liberar recursos en la red del cliente y en el sistema Data Domain para obtener un rendimiento mejorado de la deduplicación en el destino.
- **Arquitectura de invulnerabilidad de datos de Data Domain:** Data Domain está diseñado para garantizar que los datos puedan recuperarse de manera confiable. La arquitectura de invulnerabilidad de datos proporciona la verificación de lecturas y escrituras en línea, lo cual brinda protección contra problemas de integridad de los datos durante la recopilación y la recuperación de datos, y permite la recuperación automática. La detección continua de fallas y la autorreparación garantizan que los datos sigan siendo recuperables durante todo su ciclo de vida en el sistema Data Domain.

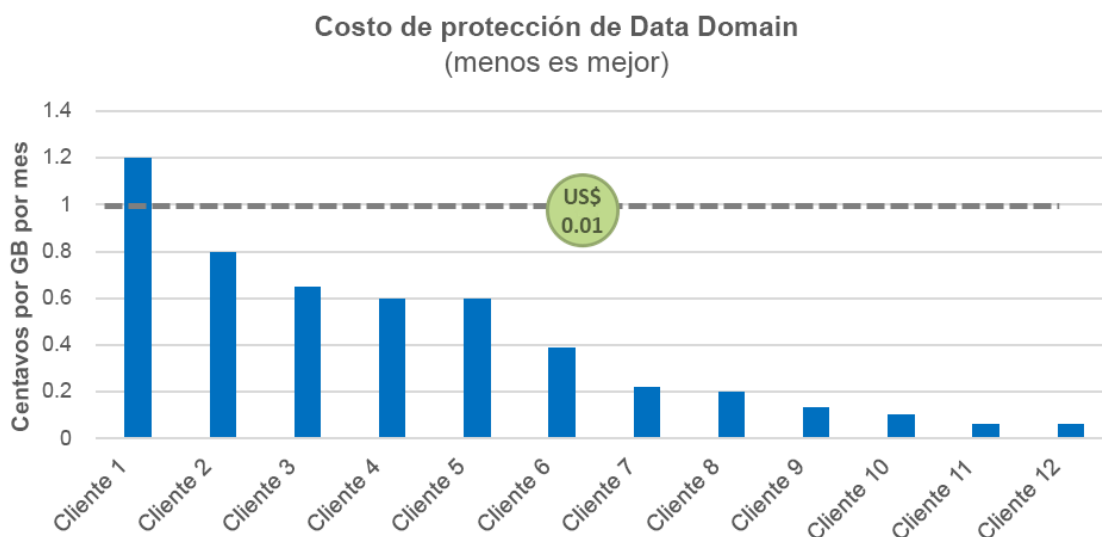
## Beneficios económicos de Data Domain

Analicemos detenidamente cómo la arquitectura de Data Domain produce beneficios económicos para las partes interesadas del negocio. La deduplicación de datos permite a los clientes almacenar más datos en la misma cantidad de espacio del disco físico. Esto reduce los requisitos de capacidad de almacenamiento y los costos. La deduplicación de datos implementada en el cliente o en el origen también permite ahorrar capacidad, pero cuenta con el beneficio adicional de mejorar el rendimiento del respaldo. Con la deduplicación en el origen, solo los bloques de datos únicos se envían desde el origen hasta el destino durante la operación de respaldo, lo cual reduce significativamente el tráfico de red. Esta eficiencia de la red mejorada permite el crecimiento de los datos de respaldo mediante el uso de una infraestructura de red existente y, posiblemente, elimina o posterga la necesidad de realizar actualizaciones de red costosas. Obviamente, cuantos menos datos deban transferirse, más rápido será el rendimiento del respaldo. Las duraciones de respaldo más cortas también permiten a los clientes aumentar la frecuencia de los respaldos, lo que reduce el riesgo de pérdida de datos, que puede ser extremadamente costoso para una organización. Con el software DD Boost, los sistemas Data Domain admiten la deduplicación tanto en el origen como en el destino, lo que brinda a los clientes la flexibilidad de implementar la deduplicación donde sea más significativo para sus ambientes.

Tal como se muestra en la figura 3, el análisis de los datos reales de ESG (incluidos el hardware, el software, la potencia, el enfriamiento y la deduplicación) demuestra que los sistemas Data Domain son capaces de atender fácilmente el almacenamiento en los ambientes de protección de datos por fracciones de un centavo por GB por mes.

Observe cómo el costo de protección varía entre 0.062 y 1.2 centavos por gigabyte para los doce clientes analizados por ESG. Esta variedad relativamente amplia de ahorros se debe principalmente a la cantidad de ahorros en capacidad de deduplicación que han logrado los clientes de Dell EMC en los ambientes de producción. El balance de este informe analiza detenidamente los costos de protección con un enfoque centrado en cómo los ahorros de deduplicación pueden incrementarse mediante una combinación de punto a punto del software Dell EMC Data Protection Suite y del hardware Data Domain.

**Figura 3. El valor económico de Data Domain**

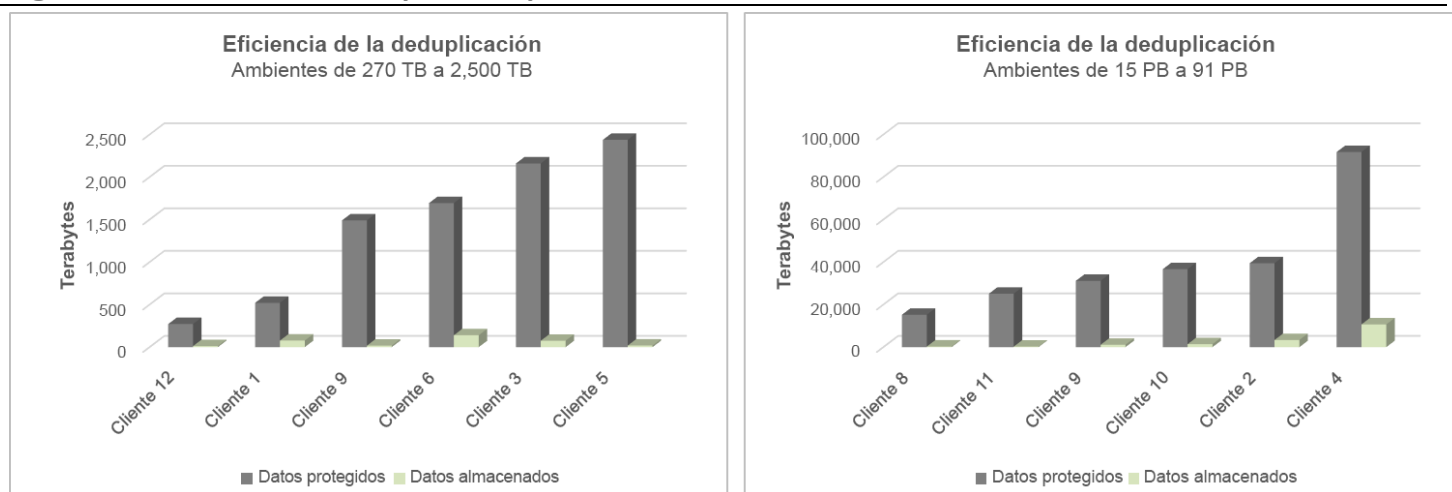


Fuente: Enterprise Strategy Group, 2017

## La eficiencia de la deduplicación es importante

ESG comenzó su estudio sobre el valor económico de los ahorros en capacidad de deduplicación de Data Domain auditando y analizando los datos de soporte del servicio Call Home de 12 clientes activos de Dell EMC. Tal como se muestra en la figura 4, la cantidad de datos que se debía proteger en los ambientes de los clientes variaba, aproximadamente, de 270 TB a más de 90 PB. Los clientes abarcaban múltiples sectores, incluidos los sectores de tecnología, fabricación, seguros y atención médica. La selección de clientes en los diferentes sectores se diseñó con el objetivo de capturar resultados de deduplicación en los distintos tipos de conjuntos de datos. En la figura 4, las barras grises muestran la cantidad de datos protegidos por cada ambiente. Las barras verdes indican la cantidad de datos almacenados luego de la deduplicación.

Figura 4. Eficiencia de la deduplicación por cliente



Fuente: Enterprise Strategy Group, 2017

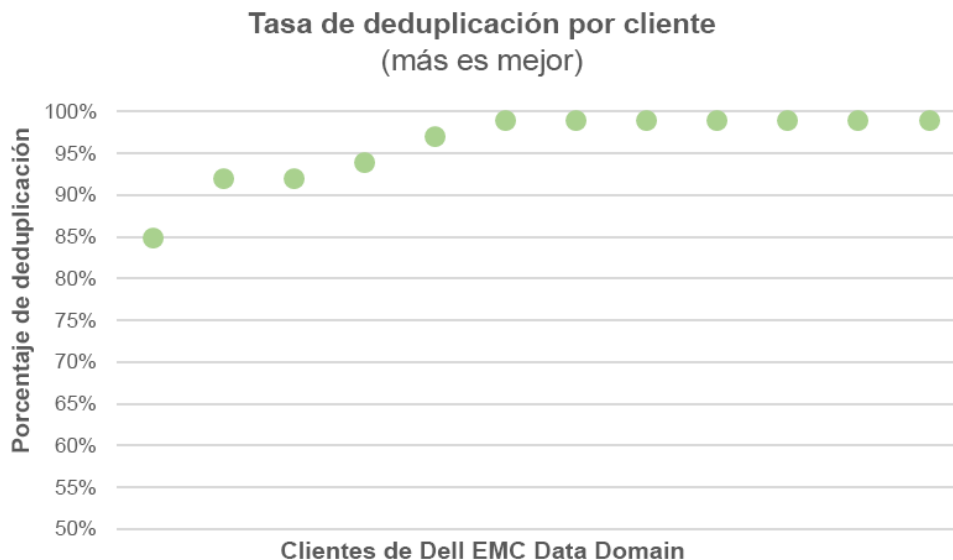
- **Deduplicación:** la mayor tasa de deduplicación que se observó en los datos analizados fue de 126:1. Este resultado provenía de un cliente que hacía uso del software de respaldo y recuperación Data Protection Suite con un dispositivo Data Domain, y se identificó como el primer punto de datos en el lado izquierdo de la figura 4. Para todos los clientes analizados, se observó una tasa de deduplicación de 41:1 en promedio.
- **Protección:** incluso el cliente con la tasa de deduplicación más pequeña que pudo observarse, de aproximadamente 9:1, pudo proteger casi 92 PB de datos utilizando solo 10 PB de capacidad. Este cliente se identifica en el último punto de datos en el lado derecho de la figura 4. Dicho cliente es parte del sector de fabricación y cuenta con un ambiente altamente distribuido.

La figura 5 brinda otra visión sobre el efecto de la deduplicación de Data Domain. Esta figura muestra el porcentaje de deduplicación logrado para cada uno de los mismos 12 ambientes del cliente. Las tasas de deduplicación oscilan entre un 85 % y un 99 %. Por lo general, mientras más tiempo haya estado recibiendo datos la solución Data Domain en el ambiente, mayor será la tasa de deduplicación. Esto se debe a que es más probable que vea los mismos patrones de datos con el paso del tiempo y que solamente necesite almacenar datos únicos.

La funcionalidad de segmentación de longitud variable de la arquitectura de Data Domain es un componente clave para lograr este nivel de deduplicación. Gracias a la segmentación de longitud variable, la solución Data Domain puede alinearse más fácilmente con los patrones naturales, incluidos los marcadores de registro de fecha y hora de la base de datos, en las estructuras de datos que se envían al dispositivo para protección. Esta segmentación de longitud variable produce niveles de deduplicación significativamente más altos que las arquitecturas de longitud fija. Además, permite una escalabilidad

más efectiva dentro de un solo pool de almacenamiento. Esto también da como resultado menos dispositivos para administrar, lo que permite ahorrar en los costos operativos.

**Figura 5. Porcentaje de deduplicación por cliente**



Fuente: Enterprise Strategy Group, 2017

Otro de los beneficios de la deduplicación de segmentación de longitud variable alta surge cuando se replican datos entre dispositivos Data Domain. Una menor cantidad de datos significa menos datos que replicar. Esto no solo significa un mejor rendimiento de replicación, sino también requisitos más bajos de ancho de banda de WAN para el proceso de replicación.

- **Tasas de deduplicación:** las tasas de deduplicación auditadas y analizadas oscilaban entre un 85 % y un 99 %, con un promedio del 96 %. Esto quiere decir que solo era necesario almacenar una fracción (el 1 % o menos, en el mejor de los casos) de los datos de producción en el sistema Data Domain para protección.
- **Por debajo del 90 %:** únicamente un cliente de los datos analizados presentó una tasa de deduplicación inferior al 90 %. Una vez más, este cliente representa el ambiente altamente distribuido con pools de protección distribuidos.
- **90 % y por encima de este porcentaje:** el resto de los clientes analizados lograron tasas de deduplicación por encima del 90 %, y la mayoría de los ambientes de clientes ejecutaban una tasa de deduplicación del 99 %.

### Velocidad y escalabilidad

El aumento de la capacidad de almacenamiento a través de la deduplicación no siempre va acompañado de un alto rendimiento. Data Domain enfrenta este desafío mediante el aprovechamiento de dos elementos clave de rendimiento: el software DD Boost y Stream-informed Segment Layout (SISL).

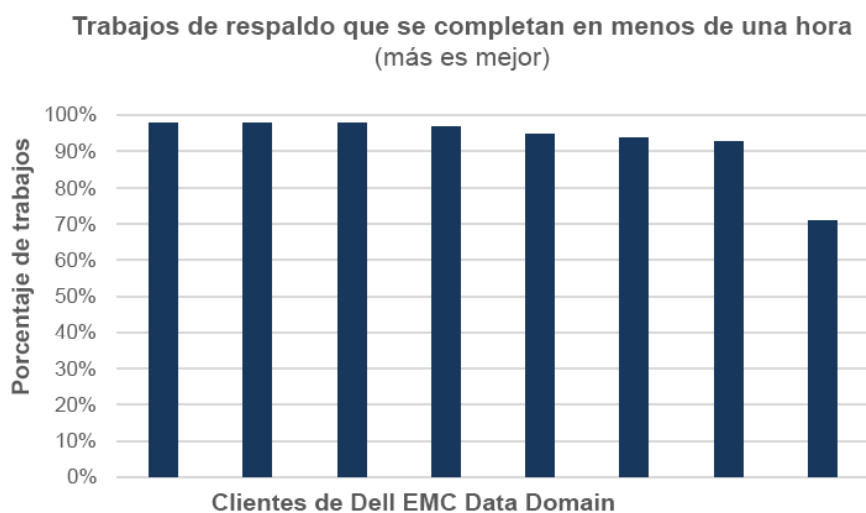
SISL es un elemento arquitectónico de la plataforma Data Domain que permite que el 99 % del proceso de deduplicación se lleve a cabo en la CPU y la memoria RAM. Esto significa que la solución no depende de un conteo alto de ejes de discos para el rendimiento. Como consecuencia, las soluciones con un espacio físico más pequeño en la misma controladora de Data Domain pueden lograr el mismo rendimiento que las configuraciones de mayor capacidad. Además, dado que las



controladoras de Data Domain utilizan las CPU Intel, estas obtienen los beneficios de las mejoras relacionadas con el rendimiento de cada versión nueva de los procesadores Intel.

El software DD Boost también ayuda a brindar rendimiento mediante la distribución de partes del proceso de deduplicación a los clientes de aplicaciones o al servidor de respaldo. Con DD Boost, solamente es necesario transferir los datos únicos desde los clientes o el servidor de respaldo al sistema Data Domain. DD Boost reduce la cantidad de datos transferidos hasta en un 99 %. ESG auditó el beneficio neto de finalizar los trabajos de respaldo de manera más rápida. Tal como se muestra en la figura 6, la mayoría de los clientes normalmente completan los trabajos de respaldo en una hora o menos.

**Figura 6. Rendimiento del trabajo de respaldo**



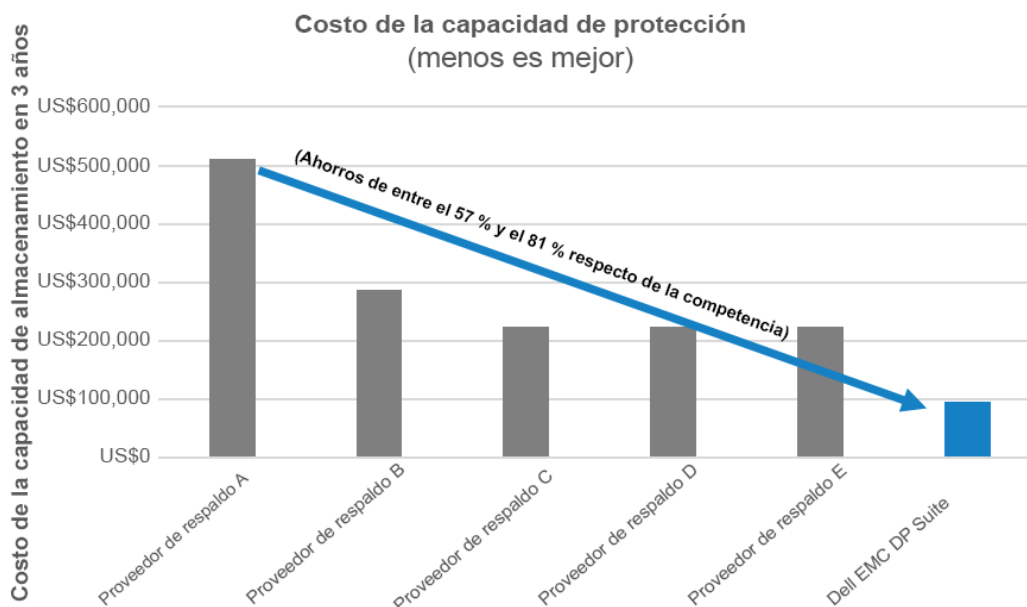
Fuente: Enterprise Strategy Group, 2017

- **Menos de 15 minutos:** para un cliente del sector tecnológico, se necesitó menos de 15 minutos para completar el 76 % de todos los trabajos de respaldo.
- **Menos de 1 hora:** para ocho de los doce clientes analizados, se necesitó menos de una hora para completar entre el 70 % y el 98 % de todos los trabajos de respaldo (tal como se muestra en la Figura 6).
- **Menos de 4 horas:** para un cliente en el sector de seguros, se necesitó menos de cuatro horas para completar todos los trabajos de respaldo.
- **Más de 4 horas:** para dos de los ambientes más grandes (en cuanto a la cantidad de aplicaciones y al tamaño de estas), se necesitó más de cuatro horas para completar todos los trabajos de respaldo.

### Consideraciones relativas al ahorro de costos

ESG completó el estudio sobre el valor económico de los ahorros en capacidad de deduplicación de Data Domain auditando y analizando los datos de soporte del servicio Call Home de más de 15,000 sistemas Data Domain implementados en todo el mundo. Esta muestra se obtuvo de los sistemas Data Domain que registran el software de respaldo en uso y las tasas de deduplicación logradas. Los datos muestran que la eficiencia de deduplicación de Data Domain se puede aprovechar aún más cuando se implementa una solución de protección de datos de Dell EMC de punto a punto. Como se muestra en la figura 7, la combinación de Data Domain con Data Protection Suite incrementó los ahorros de deduplicación y redujo el costo de capacidad de protección de un 57 % a un 81 %.

Figura 7. Incremento en los ahorros con Dell EMC Data Domain y Data Protection Suite



Fuente: Enterprise Strategy Group, 2017

La solución Data Domain brinda otros beneficios económicos, además del costo de ahorro de capacidad que se muestra en la figura 7. El análisis de ESG de los datos de clientes reales de Data Domain revelaron también los siguientes beneficios:

- Reducción del ancho de banda de red:** al igual que la mayoría de las aplicaciones empresariales, las soluciones de respaldo y recuperación dependen de los recursos de red para funcionar. Esto hace referencia tanto a los recursos de LAN para la restauración y el respaldo locales, como a los recursos de WAN para la DR y la continuidad del negocio. Las soluciones ineficientes de protección de datos pueden consumir rápidamente los recursos de red valiosos, e incluso afectar la productividad y la conectividad del usuario. Mediante el aprovechamiento de la deduplicación de Data Domain, un fabricante multinacional pudo reducir la utilización de su ancho de banda de protección de datos local y remota en un 98 %.
- Rendimiento:** la deduplicación de datos eficiente puede tener un gran impacto en la mejora del rendimiento general de la protección de datos. Mediante la implementación de DD Boost, que reduce la cantidad de datos que se deben transferir entre el cliente y el sistema Data Domain, un fabricante de equipos pesados pudo mejorar sus velocidades de restauración y respaldo en un 50 %. El mismo cliente mejoró su preparación para la DR en un 90 %.
- Reducción del espacio físico del centro de datos:** las instalaciones, especialmente la cantidad necesaria para admitir los requisitos de espacio, potencia y enfriamiento de un centro de datos moderno, no son económicas. Puede resultar difícil, y en algunas ocasiones imposible, expandir físicamente un centro de datos sin realizar una reubicación. Con Data Domain, una cadena nacional de tiendas departamentales eliminó las cintas físicas en su ambiente de protección de datos y liberó tres filas completas de espacio de rack muy valioso en el centro de datos.

## La gran verdad

Los principales mandatos de protección de datos de los líderes de TI se centran en mejorar la agilidad y confiabilidad fundamentales de las soluciones en uso. El mandato que le sigue inmediatamente es la reducción de costos, la cual se considera también una prioridad entre los implementadores de la protección de datos. Estos retos no deben considerarse contradictorios ni mutuamente excluyentes; de hecho, todos pueden abordarse con el uso de soluciones de protección de datos mejoradas, las cuales están diseñadas tanto para la eficiencia como para la confiabilidad y la funcionalidad.

Si bien la eficiencia se da de muchas maneras, por lo general se basa en el costo de hacer negocios y en la interoperabilidad:

- La capacidad de brindar almacenamiento con protección de manera económica mediante la reducción del costo de servicio de las aplicaciones de protección de datos.
- La interoperabilidad entre las soluciones de software de respaldo, dado que solo una de cada cuatro organizaciones cuenta con una sola aplicación de respaldo.<sup>4</sup>
- La eficacia de la deduplicación y de la escalabilidad del sistema de almacenamiento con protección, que puede utilizarse para reducir la cantidad, y el costo, de los sistemas que deben implementarse y administrarse.
- La capacidad de aprovechar el mismo pool de almacenamiento con protección optimizado con respecto a la capacidad para los datos de respaldo y archiving.
- La interoperabilidad (e integración) entre el almacenamiento con protección y la infinidad de plataformas, de manera tal que los propietarios de aplicaciones (p. ej.: administradores de bases de datos y administradores virtuales) puedan utilizar sus propias herramientas para una recuperación/protección adicional, y continuar siendo ciudadanos de TI responsables y aprovechando el almacenamiento con protección eficiente y centralizado.

Proporcionando lo que muchos en el sector consideran sinónimo de “almacenamiento con protección” o “deduplicación”, Dell EMC ha ampliado la plataforma Data Domain de un simple componente de la infraestructura de protección de datos a la base de un ecosistema que incluye almacenamiento de producción, proveedores de respaldo, proveedores de archiving y aceleradores de plug-in para las aplicaciones (es decir, DD Boost), entre otros elementos. El dispositivo integrado de protección de datos de Dell EMC es una extensión de este ecosistema con un software de protección y una analítica preconfigurados. Por lo tanto, existe un impulso significativo para seguir innovando frente a las demandas actuales del mercado, lo que genera mayores ganancias de eficiencia dentro de los mecanismos de deduplicación de Data Domain, mejoras con sus integraciones y expansiones de sus funcionalidades más allá de la deduplicación en las instalaciones, tales como la organización de la nube en niveles, la replicación hacia/desde los dispositivos alojados en la nube, los dispositivos virtuales, la recuperación ante desastres en la nube, etc.

Uno de los elementos reconocibles más importantes de cualquier plataforma de deduplicación es el rigor de garantizar la integridad de los datos dentro del repositorio, ya que mucho se basa en la integridad de cada bloque único dentro del sistema. Por ello, es fundamental que las organizaciones investiguen de manera diligente con qué fervor y frecuencia el sistema verifica el almacenamiento subyacente para garantizar la integridad. El enfoque de Dell EMC con respecto a esto ha sido un sello distintivo de la plataforma Data Domain desde su creación a través de lo que se denomina arquitectura de invulnerabilidad de datos (DIA).

<sup>4</sup> Fuente: Informe de investigación de ESG, *Modernización de la protección de datos*, próximo a ser publicado.

En función de una auditoría de los datos de campo recopilados en más de 15,000 ambientes de producción, ESG ha confirmado que la combinación del hardware Dell EMC Data Domain y del software Data Protection Suite reduce entre un 57 % y un 81 % el costo de capacidad de protección, en comparación con los ambientes Data Domain que utilizan un software de respaldo de la competencia. Sería bueno que las organizaciones que dudan en invertir en Dell EMC Data Domain debido al “precio” vuelvan a considerar los beneficios económicos a lo largo del tiempo, lo que incluye el costo de reducir la pérdida de datos y el tiempo fuera, y de mejorar el rendimiento y, lo más importante, incluye un costo de protección menor a un centavo por GB por mes.

Todos los nombres de marcas comerciales son propiedad de sus respectivas empresas. La información incluida en esta publicación se obtuvo por medio de fuentes que The Enterprise Strategy Group (ESG) considera confiables, pero no está garantizada por ESG. Esta publicación puede contener opiniones de ESG que están sujetas a cambios periódicos. Los derechos de esta publicación pertenecen a The Enterprise Strategy Group, Inc. Cualquier reproducción o redistribución de esta publicación, en su totalidad o en parte, ya sea en formato impreso, electrónico o de otro tipo, a personas no autorizadas para recibirla sin el consentimiento expreso de The Enterprise Strategy Group, Inc., constituye una violación de las leyes de derechos de autor de los Estados Unidos y estará sujeta a una acción por daños civiles y, en caso de ser pertinente, a un juicio penal. Si tiene consultas, comuníquese con ESG Client Relations llamando al 508-482-0188.



**Enterprise Strategy Group** es una empresa integrada de análisis, investigación y estrategia de TI reconocida en todo el mundo por brindar inteligencia e información valiosa útiles a la comunidad mundial de TI.

© 2017, The Enterprise Strategy Group, Inc. Todos los derechos reservados.

