



**Perspectiva de EMC sobre
la Deduplicación de Datos
para Procesos de Backup**



Resumen

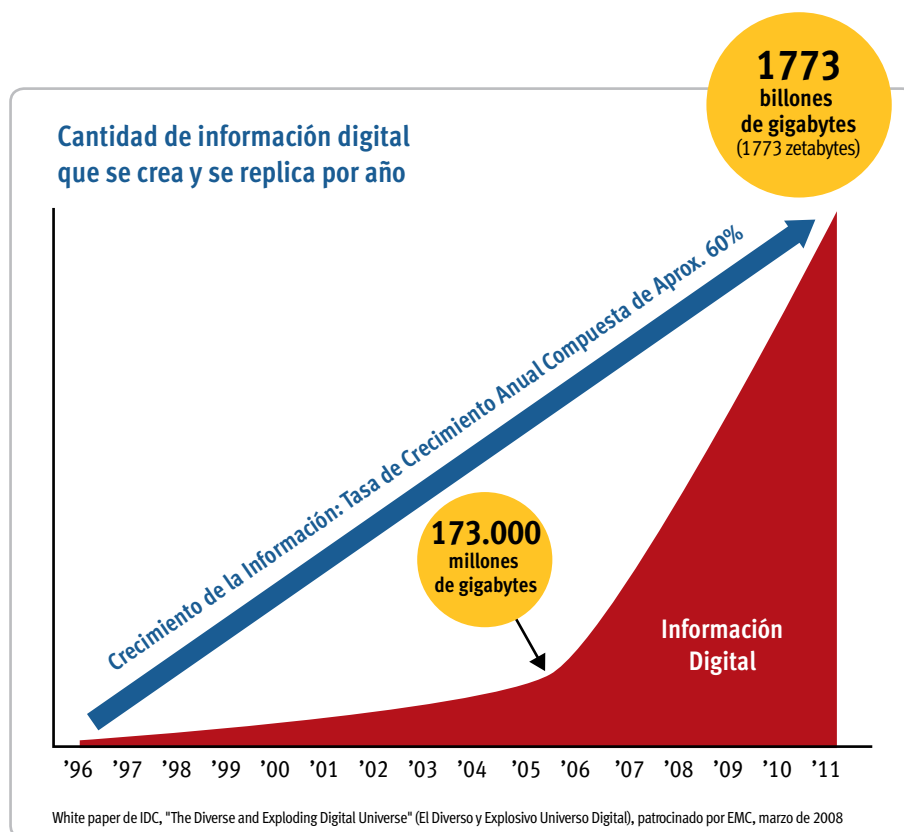
Este documento analiza los factores que impulsan la necesidad de la deduplicación de datos y los beneficios que ésta ofrece como una característica de la estrategia de backup de una empresa. Entre los temas tratados en este documento, se incluyen:

- Aspectos del crecimiento de la información y la presión resultante en los procesos de backup y replicación
- Tecnología de deduplicación de datos actual y su implementación
- Beneficios de la deduplicación de datos y su impacto positivo en los procesos de backup, replicación y recuperación
- Factores que determinan la eficacia de la deduplicación de datos
- Consideración específica de entornos de backup de VMware, oficina remota y backup en disco
- Justificación de una inversión en la deduplicación de datos

Finalmente, el documento describirá brevemente las soluciones de backup con capacidad de deduplicación de EMC que se han optimizado para satisfacer las necesidades de backup, los requerimientos de recuperación y los entornos de las aplicaciones.

La Explosión del Crecimiento de los Datos

IDC predice que el crecimiento del "universo digital", es decir, la información que se crea, captura o replica en formato digital, aumentará diez veces en volumen entre 2006 y 2011 (IDC, The Diverse and Exploding Digital [El Diverso y Explosivo Universo Digital], marzo de 2008). La mayoría de las empresas lidian con el impacto que ocasiona el crecimiento de información en sus procesos de protección de información. Más información, restricciones de costos y niveles de servicios más estrictos están presionando los entornos de backup tradicionales hasta el punto límite. La protección de los datos críticos constituye un desafío para las empresas de todos los tamaños. Los procesos y las herramientas de backup tradicionales no pueden mantener el ritmo, lo que hace que las empresas se esfuercen por administrar las ventanas de backup, el tiempo activo de las aplicaciones y los requerimientos de recuperación de información.



Los Procesos de Backup se Enfrentan a la Multiplicación del Crecimiento de Datos

La mayoría de las soluciones de backup tradicionales protegen los mismos datos de manera reiterada, lo que incrementa el almacenamiento de información total administrado en cinco a treinta o más veces. Considere el efecto de un backup de file server completo semanal. Con una política de retención de cinco semanas, la gran mayoría del contenido se protege cinco veces. Ahora, considere un servidor de correo electrónico protegido con un backup completo cada noche; a la mayor parte del contenido de ese servidor de correo se le realiza backup 35 veces durante esas cinco semanas.

Las empresas necesitan soluciones que las ayuden a administrar esta explosión de información. Además, las regulaciones gubernamentales y las solicitudes de presentación de información en procedimientos legales pueden sobrecargar aún más los recursos y las capacidades de las soluciones de protección de datos tradicionales. La falla en el cumplimiento de normas o presentación de información de manera oportuna puede generar importantes costos y multas. La legislación reciente también ha expuesto el riesgo de enviar cartuchos de cinta físicos como uno de los mayores problemas de seguridad de la infraestructura de TI actual.

Como se mencionó anteriormente, las soluciones de backup tradicionales requieren una programación rotatoria de procesos de backup completos e incrementales, los cuales mueven una enorme cantidad de datos redundantes semana tras semana. Y la mayoría de las empresas crean una segunda copia de esta información para enviarla a un sitio secundario con fines de recuperación ante desastres. Cuando se consideran en conjunto, los costos de backup tradicional en lo que respecta a ancho de banda, infraestructura de almacenamiento de información y tiempo (sin mencionar los costos de capital) constituyen una carga significativa para el departamento de TI.

El backup de archivos y datos redundantes es uno de los motivos principales por los cuales las ventanas de backup implican horas de producción, utilizan recursos de la red en forma excesiva y requieren mucha capacidad de almacenamiento de información para mantener datos de backup innecesarios. Este desperdicio es aún mayor ya que la replicación de archivos duplicados y la repetición de los procesos de backup de datos redundantes no aportan ningún valor adicional al negocio.

¿Qué es la Deduplicación de Datos?

Si fuese posible, las empresas de TI sólo protegerían los datos únicos de sus backups. En vez de guardar todos los datos de manera reiterada, el escenario ideal es aquel donde se guarda únicamente el contenido nuevo o único. La deduplicación de datos ofrece esta capacidad básica. Ofrece la capacidad de detectar y eliminar los datos redundantes dentro de un dataset. Un dataset puede abarcar una sola aplicación o toda una empresa. Los elementos de datos redundantes pueden ser archivos enteros o segmentos de datos dentro un archivo. En todos los casos, el objetivo del proceso de deduplicación consiste en almacenar elementos de datos únicos sólo una vez y, al mismo tiempo, poder reconstituir todo el contenido en su formato original según demanda, con un 100% de confiabilidad a velocidades de disco.

La deduplicación de datos es fundamental para mejorar la protección de la información, optimizar las operaciones de backup, reducir la infraestructura de backup, disminuir las ventanas de backup y eliminar la carga de las redes de información.

Beneficios de la Deduplicación de Datos para Backup

Los productos de deduplicación de datos eficaces permiten a las empresas lidiar con la realización de backup del almacenamiento de información. La deduplicación de datos elimina los datos redundantes para economizar los requerimientos de almacenamiento y recuperación ante desastres para los datos. Como se observó anteriormente, existe una tremenda redundancia de datos en los entornos de backup. La implementación eficaz de la deduplicación de datos permite a las empresas proteger y hacer restore de la información en una fracción del espacio físico y del costo operativo de su infraestructura de almacenamiento de información para backup actual.

Una inversión en deduplicación de datos aporta enormes beneficios para el negocio. Entre los beneficios que brinda al negocio la deduplicación de datos, se incluyen:

Menores costos de infraestructura. Al eliminar los datos redundantes del backup, se necesita mucha menos infraestructura para mantener las imágenes de backup. La deduplicación de datos se traduce directamente en menos capacidades de almacenamiento de información para mantener las imágenes de backup. Menores requerimientos de capacidad implican costos de adquisición más bajos y costos de energía y enfriamiento reducidos.

Retención por períodos más largos. Dado que la deduplicación de datos reduce la cantidad de contenido del backup diario, los usuarios pueden extender las políticas de retención. Esto puede aportar un beneficio significativo a los usuarios que actualmente requieren retención por períodos más extensos, pero se ven limitados por los procesos y políticas actuales.

Mejor protección de datos. La deduplicación de datos permite a muchas empresas crear imágenes de backup completos diarios. Muchas de estas empresas se habían visto obligadas a realizar backups completos semanales e incrementales diarios debido a las limitaciones de las ventanas de backup. La deduplicación reduce los requerimientos de capacidad de almacenamiento de información, lo que permite políticas de backup más exigentes con mejores tiempos de restore.

Menor cantidad, mejor performance. Al reducir la cantidad total del tamaño de las imágenes de backup, las empresas aumentan la posibilidad de afrontar el reemplazo del almacenamiento en cinta tradicional por el disco para backup. El backup en disco permite imágenes de backup muy confiables y a alta velocidad, lo que satisface la necesidad de ventanas de backup más cortas y tiempos de recuperación más rápidos.

Grandes reducciones en ancho de banda para backup. La deduplicación de datos reduce la cantidad de contenido en una imagen de backup, con lo cual se reduce el gasto de replicación remota de ese contenido y se permiten backups replicados en forma remota. Además, al utilizar la deduplicación de datos en el cliente (basada en origen), los datos redundantes se extraen del proceso de backup antes de mover datos durante el proceso. Esto significa que los procesos de backup se completan moviendo sólo los elementos de datos dentro de los archivos que no se encuentran incluidos en la base de datos de backup. Los procesos de backup deduplicados basados en origen sólo requieren una parte del ancho de banda de red que utilizan los procesos de backup tradicionales.

Mayor seguridad y administración optimizada. Al aprovechar el almacenamiento de información basado en disco para las imágenes de backup, las empresas pueden mejorar la seguridad de la información ya que se eliminan los riesgos, los retrasos y los costos asociados con el manejo y el bunkering de cintas físicas. En lugar de los métodos tradicionales para enviar cintas entre las instalaciones, ahora el contenido puede trasladarse a través de la red.

Es importante comprender que la deduplicación de datos no es una solución única. Existen muchas opciones en cuanto a cómo, cuándo y dónde realizar la deduplicación. Existen impactos en el performance y la eficacia que deben tenerse en cuenta antes de finalizar la estrategia de deduplicación de datos.

Un Análisis Más Detallado de la Deduplicación de Datos

Existen muchas formas de deduplicación de datos. Por lo general, no hay una manera ideal de implementar la deduplicación de datos en toda una empresa. En cambio, las empresas pueden implementar más de una estrategia de deduplicación a fin de maximizar los beneficios. En el momento de elegir una solución de deduplicación, resulta importante comprender claramente el entorno y los desafíos de backup que representa.

Existen tres formas básicas de deduplicación de datos. Si bien las definiciones varían, algunas formas de deduplicación de datos, como la **compresión**, están presentes desde hace décadas. En los últimos tiempos, el **almacenamiento de información de instancia única** permitió la eliminación de los archivos redundantes de entornos de almacenamiento de información como el archiving. Más recientemente, observamos la incorporación de la **deduplicación dentro de los archivos**. Estos tres tipos de deduplicación de datos se describen más adelante.

Compresión de Datos

La compresión de datos es un método para reducir el tamaño de los archivos. La compresión de datos funciona dentro de un archivo para identificar y eliminar el espacio vacío que aparece como patrones repetitivos. Esta forma de deduplicación de datos es local para el archivo y no toma en cuenta otros archivos ni segmentos de datos dentro de esos archivos. La compresión de datos está disponible desde hace muchos años, pero al limitarse a cada archivo en particular, los beneficios son limitados si se compara la compresión de datos con otras formas de deduplicación. Por ejemplo, la compresión de datos no será eficaz para reconocer y eliminar archivos duplicados, pero comprimirá cada uno de los archivos de manera independiente.

Almacenamiento de Información de Instancia Única

Una forma de deduplicación es la eliminación de varias copias de un archivo. Los entornos de almacenamiento de información de instancia única (SIS, Single-Instance Storage) tienen la capacidad de detectar y eliminar copias redundantes de archivos idénticos. Una vez almacenado un archivo en un sistema de almacenamiento de información de instancia única, todas las demás referencias al mismo archivo se referirán a la única copia original. Los sistemas de almacenamiento de información de instancia única comparan el contenido de los archivos para determinar si el archivo entrante es idéntico a un archivo existente en el sistema de almacenamiento de información. El almacenamiento de información accedido por contenido generalmente está equipado con la funcionalidad de almacenamiento de información de instancia única.

Mientras la deduplicación a nivel de archivo evita el almacenamiento de archivos que son un duplicado de otro archivo, muchos archivos que se consideran únicos según la medición del almacenamiento de información de instancia única pueden tener una gran cantidad de redundancia dentro de los archivos o entre ellos. Por ejemplo, sólo sería necesario un elemento pequeño (por ejemplo, un nuevo dato insertado en la diapositiva del título de una presentación) para que el almacenamiento de información de instancia única considere que dos archivos de gran tamaño son distintos y que deben almacenarse sin deduplicación.

Deduplicación Dentro de los Archivos

La deduplicación dentro de los archivos detecta los datos redundantes dentro de los archivos y entre ellos, en lugar de buscar archivos idénticos como en las implementaciones de almacenamiento de información de instancia única (SIS). Mediante la deduplicación dentro de los archivos, se detectan y eliminan las copias redundantes de datos, incluso cuando los datos duplicados se encuentran dentro de archivos separados. Esta forma de deduplicación descubre los elementos de datos únicos dentro de una empresa y detecta cuándo estos elementos se utilizan dentro de otros archivos. De esta manera, la deduplicación dentro de los archivos elimina el almacenamiento de datos duplicados en una empresa. La deduplicación de datos dentro de los archivos ofrece enormes beneficios incluso cuando los archivos no son idénticos, pero tienen elementos de datos que ya se han reconocido en algún lugar de la empresa.

Existen dos formas de implementación de deduplicación dentro de los archivos. La deduplicación dentro de los archivos de **longitud fija** utiliza una longitud de datos fija arbitraria para buscar datos duplicados dentro de los archivos. Si bien su diseño es simple, los segmentos de longitud fija pierden muchas oportunidades de detectar datos redundantes dentro de los archivos. (Considere el caso en el que se agrega el nombre de una persona a la página del título de un documento; todo el contenido del documento cambiará, lo que provocará la falla de la herramienta de deduplicación para detectar equivalencias). Las implementaciones de **longitud variable** no están confinadas a ninguna longitud de segmentos arbitraria. Las implementaciones de longitud variable comparan los tamaños de segmentos de datos con la duplicación que ocurre naturalmente dentro de los archivos, lo que aumenta significativamente los índices de deduplicación total (en el ejemplo anterior, la deduplicación de longitud variable detectará todos los segmentos duplicados, independientemente del lugar donde se produzcan los cambios).

¿Dónde se Lleva a Cabo la Deduplicación de Datos?

Se deben tomar en cuenta algunas cuestiones relacionadas con la tecnología en el momento de determinar la mejor solución de deduplicación para la empresa. Por ejemplo, se debe considerar si la deduplicación se produce en el origen de la información o en el destino del backup. Además, se debe tener en cuenta la adecuación de las arquitecturas de deduplicación inmediata o deduplicación programada para el entorno de backup. Estas características se describen más adelante en esta sección.

Deduplicación Basada en Origen

La deduplicación de datos basada en origen permite la eliminación de datos redundantes en el origen. Esto significa que la deduplicación de datos se lleva a cabo al comienzo del proceso de backup, antes de que la información se transmita al entorno de backup. La deduplicación basada en origen puede reducir radicalmente la cantidad de datos de backup enviados a través de las redes durante los procesos de backup. Esto es de gran importancia si existen cuellos de botella en el proceso de backup relacionados con las redes, los recursos compartidos (como en VMware®) o las ventanas de backup. Asimismo, también se produce una reducción importante en los requerimientos de capacidad necesarios para almacenar las imágenes de backup.

Deduplicación Basada en Destino

La deduplicación basada en destino constituye una alternativa para la deduplicación basada en origen. La deduplicación basada en destino se lleva a cabo en el dispositivo de almacenamiento de información de backup. Esta forma de deduplicación no precisa que los usuarios cambien el software de backup existente. La deduplicación basada en destino precisa que todas las imágenes de backup se copien al dispositivo para backup; por lo tanto, el proceso de backup basado en destino no es una solución que permite reducir los requerimientos de ancho de banda para backup de cliente a destino.

¿Cuándo se Lleva a Cabo la Deduplicación de Datos de Destino?

Deduplicación Inmediata Basada en Destino

Existen dos clasificaciones del lugar en el que se lleva a cabo la deduplicación. La deduplicación inmediata se produce a medida que llegan datos nuevos al destino y el dispositivo reconoce los datos duplicados antes de escribir los datos deduplicados en el sistema de almacenamiento de información de destino. Toda la deduplicación genera overhead en cuanto al tiempo necesario para identificar y eliminar la duplicación. Debido al hecho de que las implementaciones de deduplicación inmediata incurren en una penalización por tiempo, los usuarios deben considerar el efecto de agregar latencia al path de datos. Esto puede representar un desafío para los entornos en los que la reducción de las ventanas de backup es fundamental.

Deduplicación Programada Basada en Destino

La deduplicación programada constituye una alternativa para la deduplicación inmediata. Los productos de deduplicación que se clasifican como deduplicación programada efectúan la deduplicación fuera del path de datos. Con el fin de eliminar el overhead del path de datos, las implementaciones de deduplicación programada almacenan primero los datos desde el origen y después realizan la deduplicación posterior al proceso. Las implementaciones de deduplicación programada tienen la capacidad de almacenamiento de información para mantener los datos intermedios antes de la deduplicación. El principal beneficio de la deduplicación programada es que permite al path de datos del origen operar con la mínima latencia, lo que agiliza el movimiento de datos. Además, la deduplicación programada permite que la imagen de backup nativa completa esté disponible, lo que brinda un restore rápido directamente desde el destino de disco. Los usuarios deben considerar que la deduplicación programada requiere capacidad de almacenamiento de información adicional para almacenar las imágenes de backup nativas antes de su deduplicación.

Factores que Determinan la Eficacia de la Deduplicación

La primera decisión que se debe tomar es determinar si la deduplicación se justifica en un entorno en particular. Muchos entornos de aplicaciones se adaptan bien a la deduplicación de datos y se adaptan de manera óptima a una implementación de deduplicación específica. Algunos entornos no aprovecharán al máximo los beneficios de la deduplicación y diversas soluciones de deduplicación proporcionan diferentes índices de deduplicación. El índice de deduplicación es el tamaño de los datos a los que se les hace backup en comparación con el tamaño de las imágenes de backup en disco. Existen diversos factores que contribuyen al índice de deduplicación final que se obtendrá. Algunos de estos factores son intrínsecos a la tecnología de deduplicación y otros constituyen factores relacionados con el entorno. A continuación se presenta un resumen de diversos factores comunes que afectan la eficacia de la deduplicación.

- **Período de retención.** Cuanto más extenso sea el período de retención de datos en backup, mayor será la frecuencia de datos duplicados ya existentes en el almacenamiento de imágenes de backup.
- **Relación entre backups completos y backups incrementales.** Cuanto mayor sea la frecuencia con la que se realizan backups completos, mayor será la ventaja de la deduplicación para reducir el tamaño del almacenamiento de backup.
- **Tasa de cambio.** Cuanto menos cambios se produzcan en el contenido entre backups, mayor será la eficacia de la deduplicación.
- **Tipo de datos.** Los datos de orígenes naturales (audio, imágenes, escaneos) son sumamente únicos en comparación con los datos generados por aplicaciones (documentos, correo electrónico, presentaciones). Cuanto más únicos sean los datos, existirá menos duplicación intrínseca. La deduplicación a nivel de archivo puede aun ser posible, pero existe menos probabilidad de que se produzca la deduplicación dentro de los archivos.
- **Método de deduplicación.** La deduplicación dentro de los archivos de longitud de segmentos variable detectará la mayor cantidad de deduplicación en una empresa.

Consideraciones Especiales en Cuanto al Ancho de Banda y al Tipo de Archivo

Algunos archivos tienen bajos niveles de redundancia dentro de su contenido. Entre los ejemplos de estos archivos con bajo contenido de redundancia se incluyen archivos de video, audio, comprimidos y cifrados. Estos archivos no tendrán una gran frecuencia de coincidencias de duplicación y, por lo tanto, no obtendrán un beneficio significativo con la deduplicación. Los entornos que se componen principalmente de estos archivos con baja redundancia no lograrán alcanzar altos índices de deduplicación. La aplicación de deduplicación a los entornos con poca redundancia resolverá estos archivos en almacenamiento de información de instancia única en las imágenes de backup.

Para muchos entornos, el ancho de banda para backup también constituye una consideración especial. Para entornos que tienen limitaciones de recursos o de ancho de banda (por ejemplo, oficinas remotas, VMware), la deduplicación basada en origen ofrece las ventajas más sustanciales. Dado que la deduplicación basada en origen identifica los datos redundantes antes de copiar los datos para backup, sólo una pequeña parte del contenido deberá moverse a través de la red, con lo cual se reducen los requerimientos de ancho de banda para backup.

Deduplicación por Tipo de Entorno

Ninguna oferta o estrategia de deduplicación resulta ideal para todos los entornos de aplicaciones. Los diversos entornos de aplicaciones se adaptan mejor a la deduplicación basada en origen o bien a la deduplicación basada en destino. En algunos entornos, se prefiere la deduplicación inmediata en lugar de la deduplicación programada. En esta sección, analizaremos diversos entornos de aplicaciones y demostraremos cómo se realiza la implementación ideal de la deduplicación de datos.

Deduplicación de Datos en Entornos LAN y SAN

La deduplicación de datos puede aprovecharse en diferentes formatos; desde infraestructuras simples de backup en disco basado en LAN hasta infraestructuras de bibliotecas de cintas virtuales. Cada una de estas soluciones ofrece sus propias ventajas y las soluciones de deduplicación de datos aprovechan estas arquitecturas para ampliar los beneficios para un entorno en particular.

El backup en disco basado en LAN permite a los usuarios crear fácilmente una solución mediante el uso de la infraestructura LAN existente. Para las empresas que aún no implementaron SAN Fibre Channel, el backup en disco basado en LAN suele ser la solución preferida. Para los usuarios que ya cuentan o planean implementar SAN Fibre Channel, las bibliotecas de cintas virtuales (VTLs) constituyen una firme consideración. Las VTLs son sistemas de almacenamiento de información basados en disco especializados que se presentan como una biblioteca de cintas a la aplicación de backup, mientras brindan velocidad y confiabilidad mayores. Cualquiera de estas soluciones son fáciles de implementar dentro de entornos de aplicaciones de backup tradicionales.

Deduplicación de Datos en Entornos VMware

VMware es la solución de virtualización más implementada en la industria y se utiliza cada vez más para entornos de producción de mayor misión crítica. Por ello, la ampliación de la protección de datos de máquinas virtuales constituye una función importante. VMware consolida varios servidores virtuales en un único servidor físico. En un proceso de backup tradicional, cada uno de esos servidores virtuales tiene un agente de backup que, al ejecutarse, consume muchos recursos de memoria, CPU, NIC y memoria caché. Cuando se inician varios agentes en una máquina VMware, el performance del sistema general puede verse afectado. Al implementar la deduplicación en la infraestructura VMware, se pueden generar grandes ahorros de recursos, lo que permite un mayor grado de consolidación y menor presión sobre la plataforma de servidores.

Deduplicación de Datos en Backup de Oficina Remota

El backup de oficina remota suele ser un desafío difícil. Muchas empresas tienen personal limitado con habilidades insuficientes para ofrecer niveles de servicios adecuados. Los desafíos clave de la protección de datos de oficina remota incluyen no sólo personal capacitado sino también ancho de banda WAN limitado, alto costo de ancho de banda adicional, equipos propensos a fallas, procesos manuales, falta de administración centralizada y altos índices de crecimiento de datos. El riesgo que genera la exposición o la pérdida de datos de sitios remotos puede ser realmente muy alto. Como resultado, no es de sorprender que la protección de datos de oficinas remotas sea una cuestión primordial para muchos ejecutivos de TI.

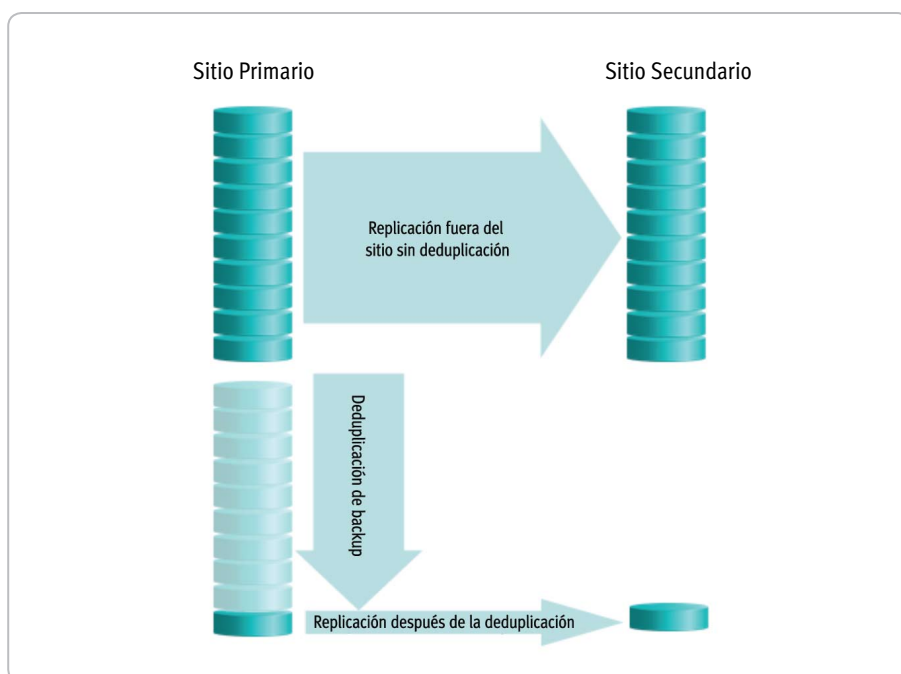
Al implementar la deduplicación de datos y la reducción consecuente en el costo de ancho de banda, los usuarios pueden almacenar de manera centralizada todos los backups de sitios remotos. La centralización de backups reduce los gastos operativos y disminuye los riesgos asociados con las actividades de backup remotas independientes. Al llevar a cabo la deduplicación en el origen, antes del envío de datos, la empresa se beneficia con la gran reducción de los requerimientos de ancho de banda para backup mientras se permite la administración centralizada de backup.

La Deduplicación Permite que la Replicación Fuera del Sitio sea Asegurable y Confiable

La deduplicación de backup brinda beneficios sustanciales al considerar la recuperación ante desastres. A fin de contar con una eficaz recuperación ante desastres basada en backup, las imágenes de backup deben conservarse fuera del sitio a una distancia suficiente del data center para adaptarse a varias situaciones de desastre en el sitio. Desafortunadamente, la replicación de backup (copia remota) por el cable duplica los requerimientos de almacenamiento de información de backup, lo que implica un aumento en infraestructura, tiempo, overhead operativo y costos de ancho de banda.

Para garantizar la protección ante desastres en el sitio, es necesario copiar la imagen de backup en el sitio secundario. Tradicionalmente, las empresas trasladaban las cintas de backup a sitios seguros y recuperaban esas cintas para facilitar la recuperación en el sitio. Los requerimientos de nivel de servicio de recuperación cada vez más exigentes obligaron a muchas empresas a agilizar el proceso de recuperación. Como resultado, la replicación de backup mediante redes se ha convertido en el método predilecto para la protección ante desastres en el sitio.

La deduplicación de datos y la replicación de esos datos deduplicados forman una sólida estrategia para lograr un proceso de recuperación ante desastres más económico. La menor cantidad de datos de imagen de backup originadas por la deduplicación reduce directamente la cantidad de contenido necesario en el sitio secundario. La deduplicación de backup reduce directamente la infraestructura, el tiempo, el overhead operacional y los costos de ancho de banda necesarios para crear imágenes de backup en el sitio de recuperación ante desastres. Además, la replicación de backup de disco a disco elimina los riesgos asociados con el traslado de cartuchos de cinta físicos.



Por qué EMC para la Deduplicación de Backup

En la actualidad, EMC cuenta con la cartera de ofertas de deduplicación más amplia de la industria. Las empresas de todos los tipos confían en EMC por su experiencia y sus consejos en el momento de seleccionar las soluciones de backup en disco óptimas con capacidades avanzadas de deduplicación de datos. La estrategia de deduplicación de EMC consiste en ofrecer soluciones que ofrecen:

- Deduplicar al nivel más alto de abstracción. EMC ayuda a las empresas a preservar el valor mediante la deduplicación de contenido en lugar de los datos y permitir la reutilización generalizada durante el ciclo de vida.
- Deduplicar lo más cerca al origen como resulte práctico. EMC maximiza los ahorros durante todo el ciclo de vida de la información, ya que maximiza los beneficios que ofrece la deduplicación a la vez que se minimizan los requerimientos de almacenamiento de información y el overhead de los recursos.
- Encontrar el enfoque más interoperable en varios casos de uso. La tecnología y las mejores prácticas de EMC permiten combinaciones de almacenamiento de información y movimiento desde el sitio primario al secundario y al terciario con fines de recuperación y retención.

Los productos de backup de EMC que cuentan con deduplicación incluyen:

EMC® Avamar®. Avamar ofrece deduplicación global de datos basada en origen mediante el uso de una longitud de segmentos variable para lograr la mayor eficacia de la deduplicación. Esta tecnología líder en la industria resulta ideal para entornos remotos, implementaciones de VMware y entornos con limitaciones en ventanas de backup y ancho de banda en data centers.

EMC Disk Library 3D 1500 y 3D 3000. Plataformas de backup en disco basado en LAN muy escalables con deduplicación de datos que ofrecen almacenamiento de información rentable, retención más prolongada en el sitio y menores costos de replicación. Estas plataformas asequibles incluyen deduplicación de datos dentro de los archivos basada en políticas para adaptarse mejor a los diversos entornos, así como también compresión de datos de hardware y replicación IP de flujos de datos deduplicados para la protección fuera del sitio.

EMC Disk Library DL4000. Esta plataforma de backup en disco basada en SAN escalable a nivel empresarial soporta la deduplicación de datos nativos. La plataforma de arquitectura en red EMC Disk Library 4000 incluye deduplicación de datos dentro de los archivos basada en políticas así como motores de replicación, deduplicación y backup independientes para garantizar la escalabilidad. Se obtienen ahorros adicionales por medio de la tecnología de rotación del disco y unidades SATA de 1TB de bajo consumo.

EMC NetWorker®. NetWorker ahora está disponible con capacidades de deduplicación de datos. NetWorker ofrece al usuario una opción de nivel de servicio, es decir, la oportunidad de realizar procesos de backup tradicionales así como procesos de backup no tradicionales, para optimizar aún más las operaciones de recuperación y backup en toda la empresa desde una consola central.

EMC Services. Estos servicios prestados por asociados de negocios y por EMC incluyen evaluaciones de backup, cálculo de deduplicación así como servicios de implementación y diseño para identificar claramente los beneficios, crear una solución y agilizar la implementación de backup optimizado que aprovecha la deduplicación de datos.

En resumen: Justificación de la Deduplicación

Muchas empresas están implementando backup en disco para perfeccionar o incluso eliminar su infraestructura de backup y recuperación basada en cinta. ¿Por qué la deduplicación del contenido de backup en disco es tan atractiva? La deduplicación reduce los costos de backup y aporta beneficios incrementales en las áreas siguientes:

- Reduce los costos de infraestructura de cintas
- Reduce los requerimientos de capacidad de disco
- Reduce las ventanas de backup
- Acelera la recuperación de datos en comparación con la cinta
- Reduce los costos de replicación fuera del sitio
- Elimina los riesgos asociados con el manejo de cintas
- Disminuye la dependencia en bibliotecas de cintas para backup

Permita que EMC y los asociados de negocios de EMC en todo el mundo ayuden a su empresa con ofertas de deduplicación de datos y backup en disco de nivel internacional. La innovadora cartera de productos de EMC, así como su gran experiencia en administración de información y sus mejores prácticas comprobadas, le garantizan que obtendrá el mayor valor de la infraestructura de la información con las mejores eficacias y los costos más bajos.



EMC South Cone
Torre Alem Plaza
Av. Leandro N. Alem 855 Piso 17
C1001AAD
Buenos Aires, Argentina
Teléfono: +54 11 4021-3622
Fax: +54 11 4021-3643

EMC Mexico
Torre Mayor
Paseo de la Reforma # 505 Piso 39
Col. Cuauhtémoc, Del. Cuauhtémoc
C.P. 06500
México D.F.
Teléfono: +52-55-50 80 37 00
Fax: +52-55-52 11 6442 y 52 11 6070
www.EMC.com
Visite el sitio Web de EMC correspondiente a su región.