

> Que no le engañen con la deduplicación - Deduplicación de datos moderna con Arcserve UDP

Por Christophe Bertrand, VP de Product Marketing

Demasiados datos, falta de tiempo, falta de espacio de almacenamiento y falta de presupuesto. ¿Le suena? Desde los primeros grandes mainframes, se ha trabajado para optimizar los requisitos de capacidad de almacenamiento y los procesos de protección de los datos.

En el mundo de los sistemas abiertos, los problemas que hemos mencionado anteriormente siguen siendo los mismos que hace años, cuando la primera tecnología de deduplicación se convirtió en la tendencia principal. Las copias de seguridad fallan, ocupan demasiado espacio y son demasiado costosas.

Hoy en día, los volúmenes de datos crecen exponencialmente y las organizaciones de todas las dimensiones luchan por gestionar lo que se ha convertido en un problema muy caro. El almacenamiento más barato ayuda, pero no se trata de una solución eficiente desde el punto de vista operativo para muchas cargas de trabajo. Hay que reducir los datos a niveles más gestionables. Demasiada cantidad de datos provoca verdaderos problemas a las empresas:

- Las empresas aprovisionan en exceso su infraestructura de copia de seguridad para anticipar un futuro rápido crecimiento. Esto es caro.
- Los sistemas heredados no pueden mantener el ritmo y tardan demasiado o están incompletos.
- Las empresas no cumplen los objetivos de punto de recuperación ni los objetivos de tiempo de recuperación.
- Las copias de seguridad sobrecargan la infraestructura y el ancho de banda de la red.
- Las empresas no pueden adoptar nuevas tecnologías, como la copia de seguridad en la nube, porque existen demasiados datos que transferir en redes de grandes áreas.

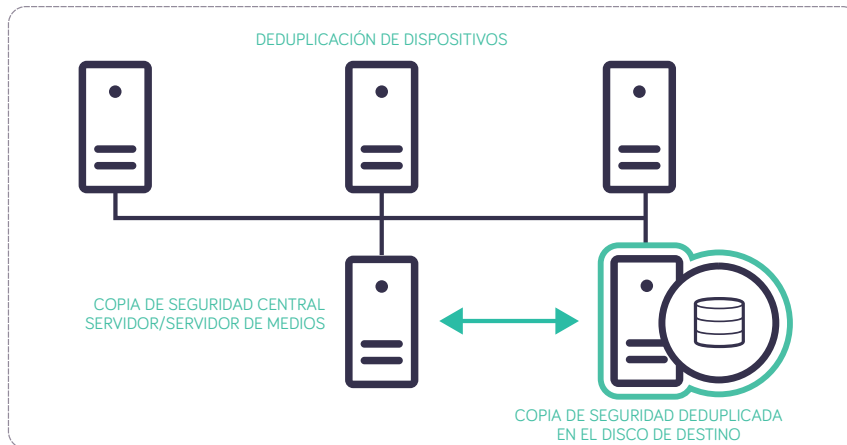
Hoy en día, se necesitan nuevos avances tecnológicos para combatir el crecimiento exponencial e imparable de las máquinas y los datos virtuales. En este artículo, analizaremos el método de Arcserve para la deduplicación de datos.

Deduplicación de destino... De lo más antiguo

Desde hace unos años, las organizaciones se vienen dando cuenta de que no es posible realizar copias de seguridad de Windows con las arquitecturas de copia de seguridad tradicionales. Además de con rendimiento de la copia de seguridad, luchaban contra el coste esencial que supone el almacenamiento de grandes cantidades de datos de copia de seguridad. Los costes de almacenamiento se disparaban a medida que se implementaban los planes de copia de seguridad. Los dispositivos de deduplicación se convirtieron en la solución preferida para abordar el problema. El proceso implicaba la toma de los datos de copia de seguridad, su optimización a través de procesos de deduplicación y su almacenamiento en el disco. "Comprimamos" sus volúmenes de copia de seguridad y ahorremosle dinero. La deduplicación de destino ha funcionado muy bien y sigue en uso hoy en día en muchos entornos.

La deduplicación de destino es atractiva porque no requiere que el usuario cambie drásticamente sus configuraciones o políticas de software de copia de seguridad y, en su lugar, solo tienen que cambiar el destino de los flujos de copia de seguridad.

La deduplicación de destino se produce sobre la marcha o como un proceso posterior (escritura completa en el disco y, a continuación, optimización para acelerar las copias de seguridad).



Deduplicación basada en el destino¹

En una de sus sesiones educativas de 2008, SNIA proporciona una descripción de los casos de deduplicación de destino. En el caso de la deduplicación de destino, se puede ver que el software de copia de seguridad actúa como el instigador de los datos y envía todos los flujos de datos de copia de seguridad no deduplicados al disco de destino o al dispositivo VTL.

Independientemente de que el proceso tenga lugar en la memoria sobre la marcha o a posteriori, ¿por qué habría que pagar otro nivel de tecnología? Se trata de un método y de una tecnología del pasado. Hoy en día existe una forma mejor.

Deduplicación distribuida... ¡Carísima!

Algunos fabricantes introducen costes adicionales cobrando básicamente cada agente de deduplicación distribuida de cada sistema que necesite protección. Aunque esto puede conllevar algunos beneficios o mejoras desde el punto de vista de la recuperación tras desastres, la multiplicación de sistemas caros, licencias de software y los requisitos de ancho de banda asociados son un suma y sigue. Además, aunque puedan existir herramientas, este tipo de arquitectura puede ser costoso desde el punto de vista administrativo.

Deduplicación en la fuente... ¡Tan de nueva generación!

¿Por qué deduplicar datos a posteriori si se puede realizar la copia de seguridad únicamente de los datos nuevos y únicos que necesite en la fuente? Siempre y cuando no afecte al cliente, puede ahorrarse ese ancho de banda que utiliza para enviar flujos de copia de seguridad al destino.

¿Por qué no compartir toda la inteligencia deduplicada con todos sus clientes? De eso trata la deduplicación en la fuente global y ahí es donde debe llegar el sector. La deduplicación en la fuente global es hacia donde muchos usuarios finales han reconocido que la tecnología debe avanzar. El reto de la deduplicación en la fuente global es garantizar que el sistema fuente (cliente) no se vea estancado por el software de deduplicación.

¹Fuente: http://www.snia.org/sites/default/education/tutorials/2008/spring/data-management/Hamilton-D_Deduplication_Methods_Data_Efficiency.pdf



Arcserve Unified Data Protection

Con Arcserve Unified Data Protection (UDP), los administradores de copias de seguridad pueden optimizar los requisitos de almacenamiento y del ancho de banda, así como acelerar la protección y la recuperación en varios lugares. Además, la solución permite la rehidratación de los datos in situ para conseguir una restauración granular rápida, incluido el sistema en cinta.

Arcserve UDP realiza una duplicación en la fuente global en los nodos que se deba proteger. Solo transmite datos desde el nodo fuente único para el servidor de punto de recuperación UDP, que sirve de almacén de datos central. Después de tomar una instantánea completa inicial de la copia de seguridad, el resto de copias de seguridad capturan solo los cambios incrementales en los datos. Puesto que se trata de un proceso muy eficiente para el cliente:

- Se reduce la cantidad de datos de los que se realiza la copia de seguridad;
- Se reduce la cantidad de datos transferidos por la red;
- Se reduce la carga en el servidor de producción y la frecuencia de las copias de seguridad puede aumentar para proporcionar excelentes objetivos de punto de recuperación y proteger el sistema, las aplicaciones y los datos.

Con la deduplicación en la fuente global de Arcserve UDP, nunca realizará copias de seguridad de datos dos veces, incluidos los SO, hipervisores, etc. La deduplicación en la fuente global de Arcserve UDP reduce drásticamente la cantidad de datos transferidos durante los ciclos de copia de seguridad. Esto se aplica al software de UDP y a los dispositivos de UDP que ejecutan el mismo software.

En resumidas cuentas, cualquier equipo, máquina virtual o servidor del que se realice una copia de seguridad se comunica con el servidor de punto de recuperación (RPS) de Arcserve UDP que gestiona un índice de bases de datos global de los archivos de todas las máquinas en cualquier parte. El servidor RPS hace el trabajo de averiguar aquellos datos de los que hay que realizar la copia de seguridad e introduce los datos nuevos según sea necesario a la vez que elimina las copias duplicadas. En combinación con la tecnología de compresión, el RPS puede comprimir datos en el disco casi a más del 92%.

El índice de bases de datos de deduplicación RPS se almacena en un SSD de altas prestaciones para mejorar aún más el rendimiento y la eficiencia y reducir los costes en comparación con el enfoque basado en la memoria simplemente.

Destinations: Recovery Point Server

Actions - | Add a Recovery Point Server

Name	Plan Count	Data Protected	Deduplication	Compression	Overall Data Reduction	Space Occupied
✓ [blurred]	6	15.57 TB	61%	34%	74%	4.02 TB
✓ [blurred]	5	14.41 TB	54%	39%	72%	4.06 TB
✓ [blurred]	1	6.24 TB	28%	40%	56%	2.72 TB
✓ [blurred]	6	11.78 TB	88%	37%	92%	936.78 GB
✓ [blurred]	1	94.36 GB	66%	37%	78%	20.41 GB
! [blurred]	1	N/A	N/A	N/A	N/A	N/A
✓ [blurred]	4	12.05 TB	64%	33%	76%	2.04 TB
✓ [blurred]	3	5.59 TB	76%	33%	84%	915.21 GB
✓ [blurred]	1	6.25 TB	28%	40%	56%	2.72 TB
✓ [blurred]	4	7.61 TB	68%	39%	81%	1.47 TB
✓ [blurred]	5	9.69 TB	89%	29%	92%	757.4 GB

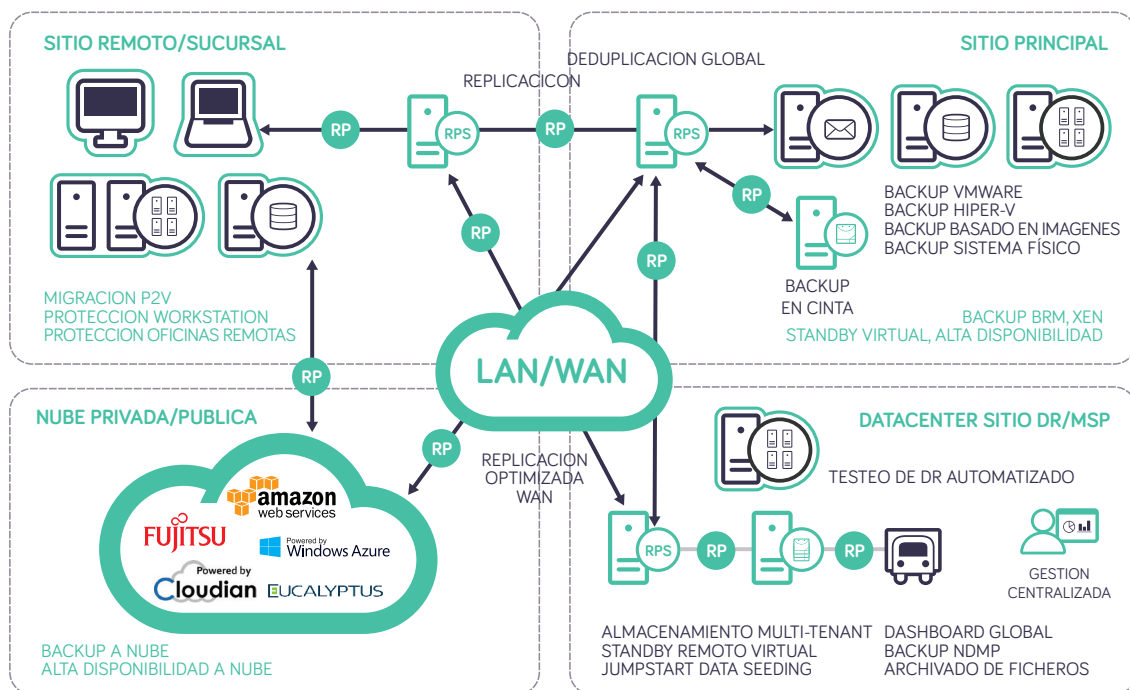
En esta captura de pantalla del cliente de la vida real, puede observar la eficiencia de la deduplicación combinada con las capacidades de compresión que aplicamos a la fuente de datos, lo que da como resultado unos niveles generales de reducción de datos muy altos.



Verdadera deduplicación global. ¡Increíble!

La deduplicación "global" en todos los clientes de la infraestructura es vital para limitar el almacenamiento y transferencia innecesarios de los datos de copia de seguridad duplicados. Los datos se deduplican en todos los nodos y tareas y en todos los lugares. La deduplicación global en Arcserve UDP va más allá de los límites de algunos proveedores cuya deduplicación solo se aplica a la caché de réplica de WAN y no a lo que se almacena realmente en el disco, con lo que se reducen los posibles beneficios generales en términos de ahorro de ancho de banda y almacenamiento.

Para la protección de los datos en sitios locales y remotos (nube privada), el índice de bases de datos de deduplicación global de Arcserve UDP se replica y distribuye para que todos los datos en la fuente y en el destino se deduplican en todos los servidores RPS. Puesto que los datos de copia de seguridad se deduplican globalmente antes de transferirse al RPS de destino, solo se envían los cambios a través de la red, lo que mejora el rendimiento y reduce el uso del ancho de banda. Todo el proceso se protege mediante cifrado en el nivel del almacén de datos y contraseñas por sesión.



Arquitectura de deduplicación global de Arcserve UDP



La "red neta"

La deduplicación de Arcserve UDP no es una tecnología mejor simplemente, sino que también aporta varios beneficios empresariales:

- **Realización de copias de seguridad más rápidas.** Con menos datos que transmitir y almacenar, las copias de seguridad son más rápidas. Esto es importante en situaciones en que el volumen total de los datos amenace con llevar tanto tiempo que no finalice antes de que deba empezar el siguiente.
- **Mejora del rendimiento del cliente.** El agente Arcserve tiene un impacto mínimo en el rendimiento del cliente dado que la mayoría del procesamiento se efectúa en el servidor de punto de recuperación. En el caso de entornos virtualizados sin agentes, la copia de seguridad de VMware y Hyper-V reduce el riesgo de la formación de cuellos de botella y problemas de rendimiento en el nivel del hardware. Dicho de otro modo, garantiza que las copias de seguridad no estancuen los servidores virtualizados.
- **Simplificación de la infraestructura de copia de seguridad.** Enrutando las copias de seguridad a través del servidor de punto de recuperación, es más fácil dirigir los datos de copias de seguridad a distintos lugares, por ejemplo, a la nube por medio de Azure o Amazon Web Services, o una cinta local en una nube privada exterior.
- **Mejora de la resistencia y de la disponibilidad.** Dado que los datos deduplicados se almacenan centralmente en el servidor de punto de recuperación, es fácil proteger la infraestructura de copia de seguridad. Por ejemplo, se puede replicar toda la información del almacén de datos en la nube, en otro servidor del mismo centro de datos o en el exterior.
- **Reducción del ancho de banda requerido para las copias de seguridad.** El servidor de punto de recuperación solo introduce datos nuevos o cambiados de un cliente con altos niveles de granularidad, incluso en fragmentos de tan solo 4 KB. Esto dota a las copias de seguridad de una eficiencia extrema.
- **Cumplimiento de los acuerdos de nivel de servicio de punto de recuperación y de tiempo de recuperación.** Gracias a la deduplicación en la fuente global de Arcserve, es más fácil cumplir los objetivos de tiempo y de punto de recuperación. Con menos datos que deben transferirse, las copias de seguridad pueden ser más frecuentes y las restauraciones más rápidas.

Compruébelo usted mismo y pruebe Arcserve UDP gratis durante 30 días. Simplemente, acceda a arcserve.com/free-backup-software-trials

Consulte nuestra calculadora de deduplicación UDP en: arcserve.com/calculator



Para obtener más información sobre UDP de Arcserve, visite arcserve.com