

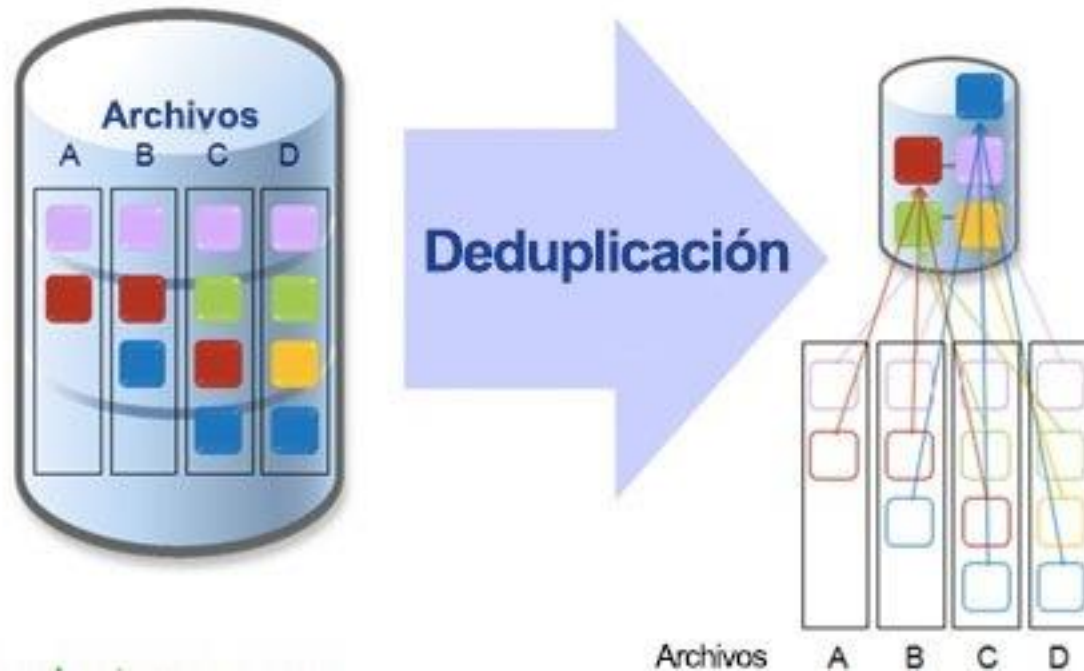
Deduplicación

Felix González Zepeda

Yared Guzmán Marín

¿Qué es?

- Esta técnica de respaldo elimina los datos redundantes almacenados, **guardando una única copia idéntica de los datos**, y reemplazando las copias redundantes por indicadores que apuntan a esa única copia.

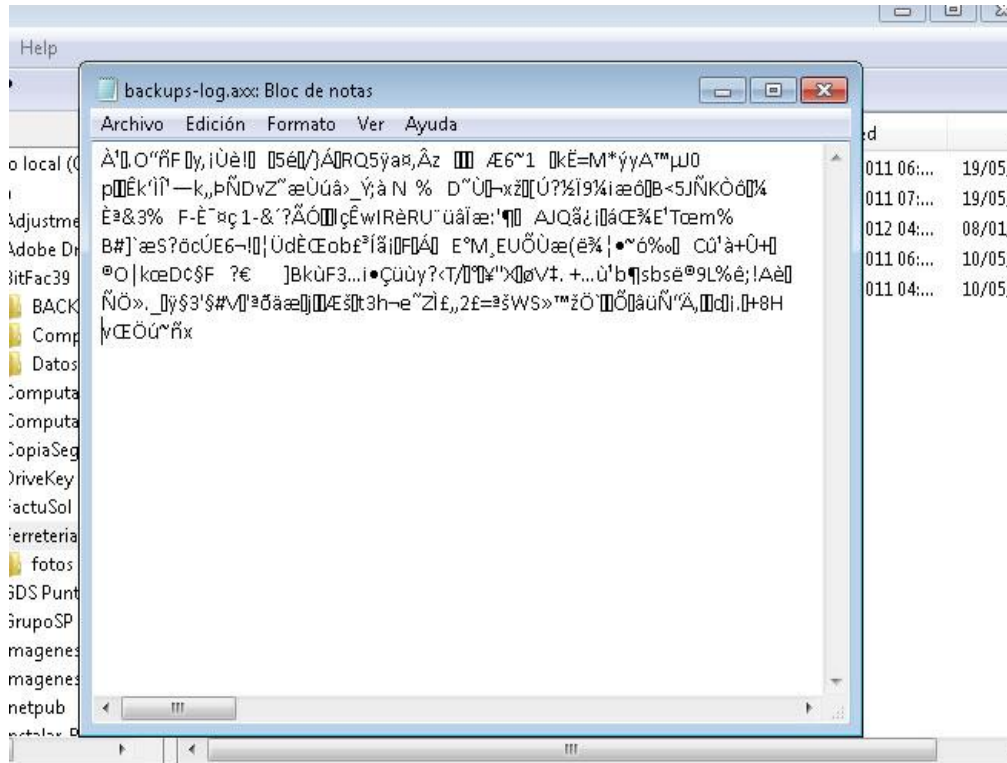


Ventajas

- ▶ **Oficinas remotas:** La deduplicación puede ayudar a responder a una situación corriente en las oficinas remotas que carecen de personal local formado para administrar las *safeguards*.
- ▶ **Recuperación de espacio:** En vista del costo del espacio de los centros de datos, es muy sensato recuperar parte del espacio ocupado por una biblioteca de cintas muy grande y sustituirla por algunas matrices de discos con función de deduplicación, que ocupan mucha menos superficie.
- ▶ Otros:
- ▶ **Deduplicación de datos y archivos duplicados.**
- ▶ **Reducción de la manipulación de soportes.**
- ▶ **Mejora de las cintas.**

Safeguard

- Es una medida tomada para proteger algo o para prevenir algo indeseable.



Cuándo y cómo

- La tarea de deduplicación de datos se puede llevar a cabo en uno o varios lugares entre el origen de los datos y el destino de almacenamiento. La deduplicación que se realiza en el servidor de aplicación o de archivos (antes de la transmisión de los datos de safeguard por la red) se denomina deduplicación del lado cliente (imperativa cuando es importante reducir el consumo de ancho de banda). Alternativamente, la deduplicación de la corriente de safeguard se puede llevar a cabo en el servidor de safeguard, y se denomina deduplicación proxy o deduplicación en el dispositivo de destino, o más en general deduplicación en destino.

Tipos de deduplicacion

- ▶ La deduplicación post-proceso.
- ▶ La deduplicación intermedia.
- ▶ Deduplicación hardware o software



Ratios de deduplicación de datos

- ▶ Ratio será la razón o cociente de dos magnitudes relacionadas.
- ▶ Ejemplo:
- ▶ Una relación de 100: 1 elimina el 99% de los datos. Aumentar la proporción a 500: 1, lo que elimina el 99.8% de los datos, no reduciría muchos más datos ya que la mayor parte de la redundancia ya se ha eliminado.

Desventajas

- ▶ ***Tipo de datos:*** No todos los datos se prestan a la deduplicación: los archivos de imágenes, vídeo y audio, u otros tipos de datos comprimidos menguan poco con la deduplicación.
- ▶ ***Datos efímeros:*** Los datos con parámetros de retención muy bajos normalmente arrojan ratios de deduplicación o reducción muy reducidas.

Fuentes:

- ▶ <https://www.pymesyautonomos.com/tecnologia/que-es-la-deduplicacion-de-datos>
- ▶ <https://searchdatacenter.techtarget.com/es/consejo/Tecnologia-de-deduplicacion-de-datos-casos-que-justifican-la-deduplicacion>
- ▶ <https://searchdatacenter.techtarget.com/es/consejo/Cuando-y-como-utilizar-la-tecnologia-de-deduplicacion-de-datos-en-el-safeguard-de-disco>
- ▶ <https://searchdatabackup.techtarget.com/definition/data-deduplication-ratio>