

Desduplicación

Lugo Gracia Julio César

Universidad Nacional Autónoma de México

Facultad de Ingeniería

17/Octubre/2019

Esquema Expositivo

- 1 ¿Qué es desduplicación?
- 2 Modo de operación del mecanismo
- 3 Desduplicación en línea y fuera de línea
- 4 Referencias

- Es una técnica especializada de compresión de datos para eliminar copias duplicadas de datos repetidos.
- Los chunks (o pedazos de datos) son identificados y guardados durante el proceso de análisis.

Funcionamiento

- Para identificar qué datos están duplicados, el mecanismo más utilizado es calcular el hash criptográfico de los datos.
- Este mecanismo permite una búsqueda rápida y confiable de coincidencias, ya sea por archivo o por bloque.
- Este método de deduplicación basado en bloques se subdivide en dos variantes: con longitud de bloques fija y longitud de bloques variables.
- Puede realizarse en línea o fuera de línea.

Hash Criptográfico

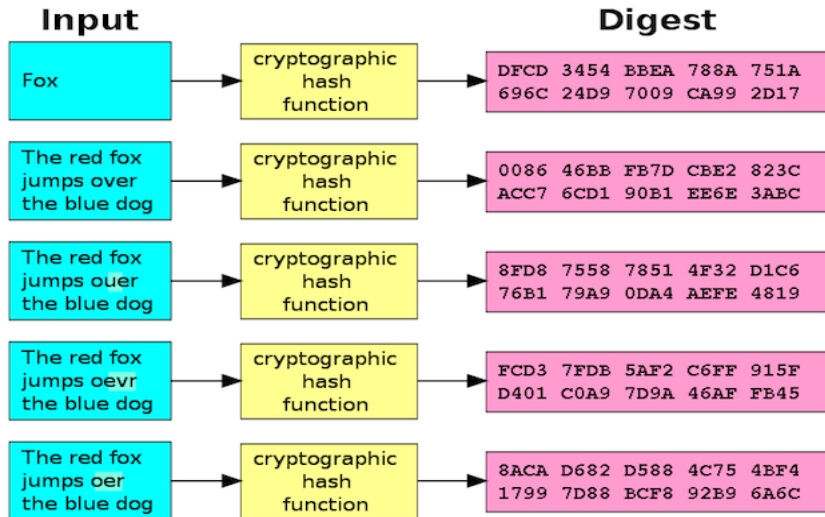


Figura: Hash Criptográfico

Desduplicación en línea

En línea

- Es la segunda fase de reducción de datos.
- Los datos los procesa y consolida antes de grabarlos en disco.
- La comparación hash y el hashing se realizan al momento.
- Duplica bloques que nunca se han grabado en el sistema de disco de destino.

En este proceso:

- Los hash se calculan sólo una vez
- Los hash se almacenan en metadatos
- Los datos se comparan, comparando el hash
- El tamaño del bloque es de 8 KB

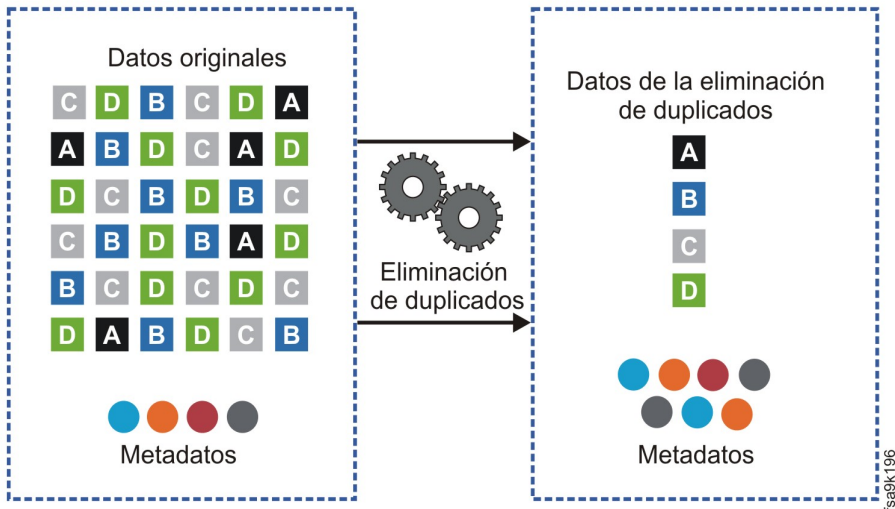


Figura: Desduplicación en línea

Offline o post process

- Permite la reducción de datos más eficiente.
 - Los datos se depositan primero en el disco y allí se deduplican.
 - Los datos no están disponibles de forma inmediata para eliminar redundancias.
 - Los datos esperan a que finalice el proceso de backup.
-
- Los datos son redireccionados al disco duro tres veces antes de que se puedan replicar o se pueda acceder a ellos.
 - El sistema de backup prepara dos zonas de almacenamiento separadas.
 - Se requiere más espacio físico en el disco que en el caso inline.

Post-Processing vs. Inline Deduplication

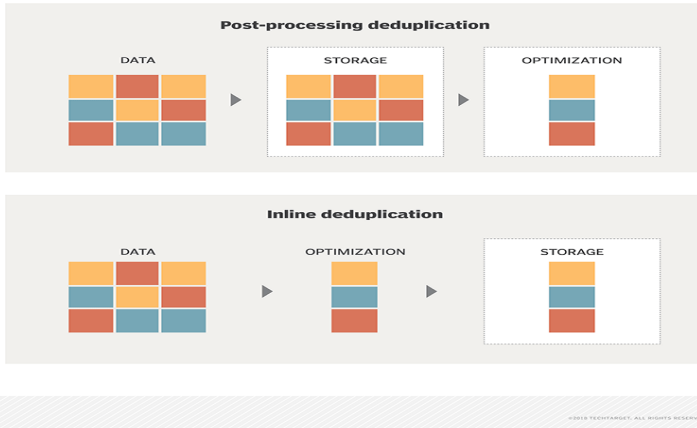


Figura: Desduplicación en fuera de línea



Hash criptográfico

<https://latam.kaspersky.com/blog/que-es-un-hash-y-como-funciona/2806/>



Desduplicación en línea

<https://www.ibm.com>



Desduplicación en fuera de línea

<https://searchdatabackup.techtarget.com>



Fundamentos de Sistemas Operativos (2015)

pp 275 - 276