

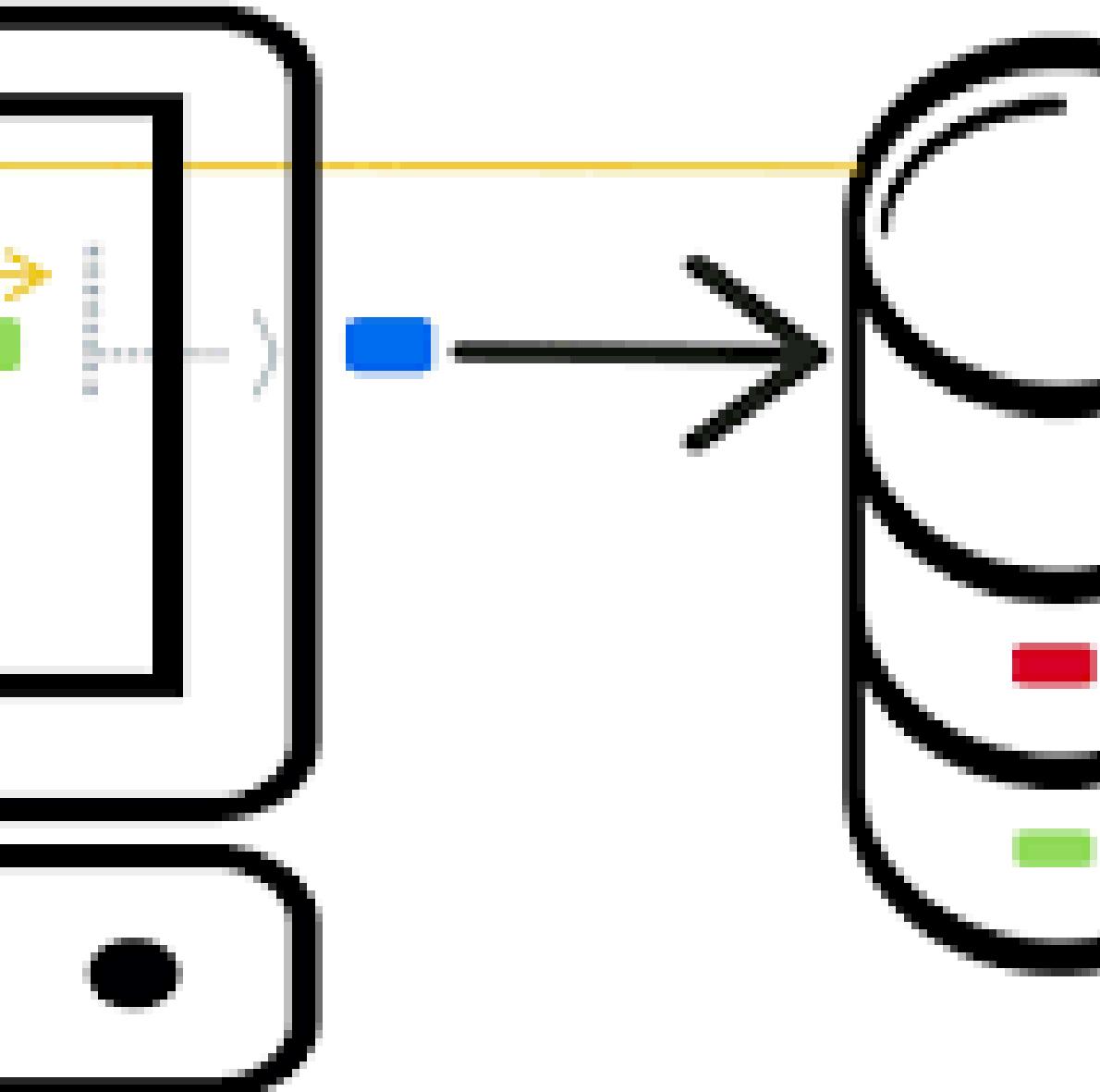
DEDUPLICACIÓN EN SISTEMAS OPERATIVOS MODERNOS

INTEGRANTES:

Jimenez Pérez Patricia Alejandra

Pérez Uribe Jose Alberto





La Deduplicación

es un proceso que elimina las copias excesivas de los datos y reduce significativamente los requisitos de capacidad de almacenamiento.





OBJETIVOS

- **Ahorro de Espacio:** La deduplicación busca eliminar copias redundantes de datos almacenados en el sistema de archivos, permitiendo un uso más eficiente del espacio en disco.
- **Mejora del Rendimiento:** Al reducir la cantidad de datos duplicados, se aceleran las operaciones de lectura y escritura en el sistema de archivos, lo que mejora el rendimiento general del sistema operativo.
- **Gestión Eficiente de Recursos:** Al minimizar la duplicación de datos, se optimiza el uso de recursos como la memoria y el procesador, lo que contribuye a un funcionamiento más eficiente del sistema.
- **Respaldos más Rápidos y Eficientes:** La deduplicación facilita la realización de copias de seguridad al reducir la cantidad de datos que se deben respaldar, lo que agiliza el proceso de respaldo y ahorra espacio en los medios de almacenamiento de respaldo.



Tipos y Técnicas de Desduplicación

Deduplicación a nivel de archivo

Deduplicación a nivel de bloque

Deduplicación en línea

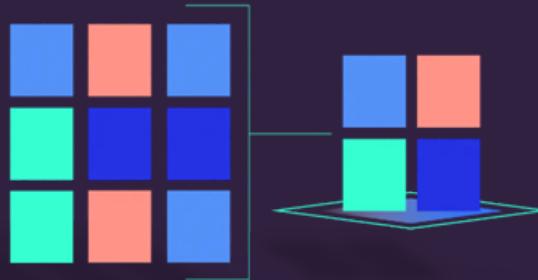
Deduplicacion fuera de línea



Principales implementaciones

Implementación a nivel de bloque

Se divide los datos en bloques más pequeños y se calcula un hash criptográfico para cada bloque.



Implementación a nivel de archivo

Identifica archivos duplicados y guarda solo una copia, ahorrando espacio. Se calcula un hash para todo el archivo y se compara con una base de datos de hashes de archivos anteriores.



Consideraciones

Estos enfoques son fundamentales y proporcionan una comprensión de cómo se puede implementar. Sin embargo, es importante destacar que existen mas opciones que se adaptan a diferentes aplicaciones, requisitos de rendimiento y del sistema.



Desafíos y Recomendaciones

1 Manejo de Colisiones de Hash

Considerar mecanismos para manejar colisiones de hash y mantener la integridad de los datos.

2 Planificación y Recursos

Evaluar los requisitos de recursos de almacenamiento y computación necesarios para la deduplicación.

3 Seguridad y Confidencialidad

Implementar medidas adecuadas para preservar la privacidad y confidencialidad de los datos desduplicados.

Complicaciones y Riesgos

La deduplicación de datos, a pesar de sus beneficios evidentes en términos de ahorro de espacio de almacenamiento, no está exenta de complicaciones y riesgos.

1 Pérdida accidental de datos

Si no se implementa correctamente, se puede perder información crítica.

2 Problemas de rendimiento

El cálculo de hashes y la gestión de grandes conjuntos de datos pueden ralentizar el sistema.

3 Complejidad en datos cambiantes

Mantener la integridad de los datos puede ser un desafío con datos en constante cambio.

4 Posibles colisiones de hash

Existe la posibilidad de que dos bloques diferentes generen el mismo hash.

5 Seguridad y Privacidad

La privacidad y la confidencialidad pueden ser preocupaciones en entornos sensibles.

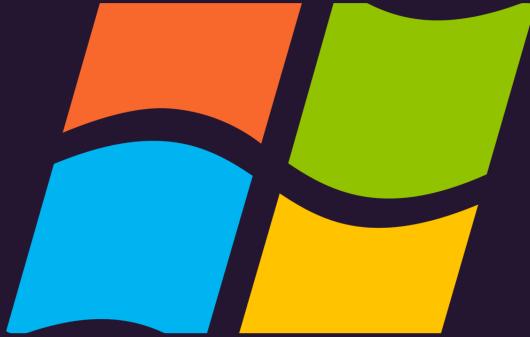
6 Complejidad en sistemas distribuidos

En entornos distribuidos, la deduplicación puede volverse más compleja.

7 Requisitos de recursos

Los sistemas deben ser adecuados en cuanto a capacidad ya que se requiere un uso significativo de recursos.

Ejemplos en Sistemas Operativos Modernos



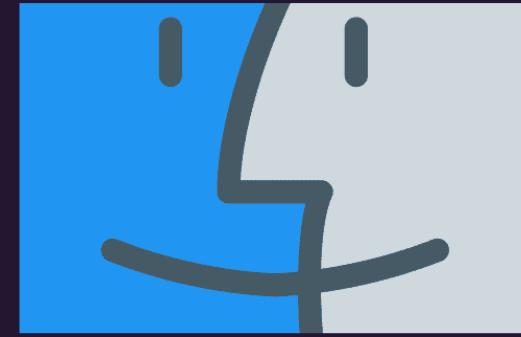
Windows

La deduplicación se realiza a nivel de volumen, identificando bloques duplicados y optimizando el espacio de almacenamiento.



Linux

Btrfs ofrece soporte nativo para deduplicación a nivel de bloques. Herramientas como rsync pueden realizar deduplicación a nivel de archivos.



macOS

Time Machine utiliza técnicas de deduplicación a nivel de bloques para optimizar el espacio de almacenamiento.

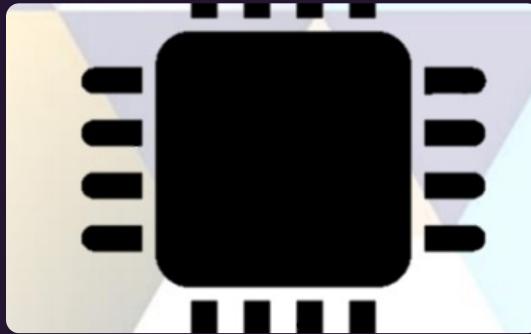
Para sistemas operativos menos comunes o de menor uso, la implementación de la deduplicación puede variar y dependerá en gran medida de la disponibilidad de herramientas y tecnologías específicas.

Ejemplos en Sistemas Operativos de menor uso



UNIX

La deduplicación se realiza a nivel de volumen, identificando bloques duplicados y optimizando el espacio de almacenamiento.



EMBEBIDOS

Btrfs ofrece soporte nativo para deduplicación a nivel de bloques. Herramientas como rsync pueden realizar deduplicación a nivel de archivos.



Sistemas propietarios especializados

Time Machine utiliza técnicas de deduplicación a nivel de bloques para optimizar el espacio de almacenamiento.

Cada sistema operativo tiene sus propias herramientas y métodos para implementar la deduplicación, adaptándose a sus características y funcionalidades particulares.



Herramientas y Métodos

Windows:

- **Herramienta Integrada:** Windows Server Data Deduplication es una herramienta integrada en algunos sistemas Windows Server que permite la deduplicación de datos en volúmenes específicos.

macOS:

- **Time Machine:** Aunque no es una herramienta de deduplicación pura, Time Machine en macOS realiza optimizaciones para ahorrar espacio al eliminar duplicados entre copias de seguridad.
- **APFS:** El sistema de archivos APFS, utilizado en macOS High Sierra y versiones posteriores, incluye características que pueden ayudar a reducir la duplicación de datos.

Unix:

- **ZFS:** ZFS, que es compatible con varios sistemas operativos Unix (como Solaris, FreeBSD, y algunas distribuciones de Linux), ofrece capacidades de deduplicación.

Linux:

- **Btrfs:** El sistema de archivos Btrfs, que es compatible con Linux, ofrece funcionalidades de deduplicación a nivel de bloques.

Sistemas Embebidos:

- La deduplicación en sistemas embebidos puede depender del sistema operativo y las herramientas específicas utilizadas en el entorno embebido. Algunos sistemas embebidos pueden no contar con funciones de deduplicación, mientras que otros pueden depender de soluciones de almacenamiento personalizadas.

Sistemas Propietarios Especializados:

- En sistemas propietarios especializados, la deduplicación puede depender de las soluciones de almacenamiento y gestión de datos proporcionadas por el fabricante del sistema. Estas soluciones a menudo están diseñadas para adaptarse a las necesidades específicas del entorno.

Conclusiones

La desduplicación de archivos en sistemas operativos modernos ofrece beneficios significativos en términos de espacio de almacenamiento, rendimiento del sistema y gestión de datos. Sin embargo, también presenta desafíos y riesgos que deben abordarse cuidadosamente. Con herramientas y métodos adecuados, es posible aprovechar al máximo esta técnica para optimizar el uso del espacio y mejorar la eficiencia operativa.



GRACIAS POR SU ATENCIÓN