

UD08. BACKUPS

1. BACKUPS

La pérdida de datos en cualquier tipo de negocio es, sin duda, uno de los mayores problemas a los que una empresa puede enfrentarse hoy en día. Por esta razón es muy importante tener una buena política de copias de seguridad que permita la recuperación de los datos en el menor tiempo posible

De forma sencilla podríamos decir que hacer una copia de seguridad es copiar la información en un soporte distinto al habitual, pero en realidad, hacer una copia de seguridad correctamente es algo que debe ser analizado en detalle.

1.1 Política de copias de seguridad

El nombre de la política de copias de seguridad incluye una serie de reglas que deben seguirse para optimizar la creación y recuperación de copias.

Por ejemplo, algunas reglas que pueden incluirse en una política de copias de seguridad pueden ser¹:

- Realizar copias de seguridad diarias de la información que se actualiza con frecuencia, y que son de gran valor para su negocio.
- Realizar copias semanales de información menos sensible, pero que aún contenga un cierto valor para la empresa, además de la información diaria habitual.
- Conserva al menos una semana la copia diaria, y al menos un mes la copia semanal. Puede rotar los medios (cinta de copia de seguridad, DVD regrabable, etc.) pero conservando la información de cierta antigüedad.
- Conserva una copia de cada mes durante al menos un año.
- Mantenga una copia anual de la información para siempre.
- Realiza una simulación de recuperación de copias de seguridad de vez en cuando.
- Se debe intentar que los soportes no se almacenen en el mismo lugar donde se hicieron.

1.2 Tipos de copias

1.2.1 Copias de seguridad completas

El tipo de operación de respaldo más básico y completo es el respaldo completo. Consiste en hacer una copia de todos los datos. El problema es que el tiempo de grabación y el espacio necesario son mayores que en el resto de tipos pero, por otro lado, su tiempo de recuperación es más rápido ya que se accede directamente a los datos sin necesidad de copias previas.

Es habitual que en las operaciones de copia de seguridad se combine la copia de seguridad completa con copias de seguridad incrementales o diferenciales.

Los problemas pueden venir por diferentes razones: deterioro, catástrofes (naturales o causadas por el hombre) y otros problemas.

De forma sencilla podríamos decir que hacer una copia de seguridad es copiar la información en un soporte distinto al habitual, pero en realidad, hacer una copia de seguridad correctamente es algo que debe ser analizado en detalle.

1.1 Política de copias de seguridad

El nombre de la política de copias de seguridad incluye una serie de reglas que deben seguirse para optimizar la creación y recuperación de copias.

Por ejemplo, algunas reglas que pueden incluirse en una política de copias de seguridad pueden ser¹:

- Realizar copias de seguridad diarias de la información que se actualiza con frecuencia, y que son de gran valor para su negocio.
- Realizar copias semanales de información menos sensible, pero que aún contenga un cierto valor para la empresa, además de la información diaria habitual.
- Conserva al menos una semana la copia diaria, y al menos un mes la copia semanal. Puede rotar los medios (cinta de copia de seguridad, DVD regrabable, etc.) pero conservando la información de cierta antigüedad.
- Conserva una copia de cada mes durante al menos un año.
- Mantenga una copia anual de la información para siempre.
- Realiza una simulación de recuperación de copias de seguridad de vez en cuando.
- Se debe intentar que los soportes no se almacenen en el mismo lugar donde se hicieron.

1.2 Tipos de copias

1.2.1 Copias de seguridad completas

El tipo de operación de respaldo más básico y completo es el respaldo completo. Consiste en hacer una copia de todos los datos. El problema es que el tiempo de grabación y el espacio necesario son mayores que en el resto de tipos pero, por otro lado, su tiempo de recuperación es más rápido ya que se accede directamente a los datos sin necesidad de copias previas.

Es habitual que en las operaciones de copia de seguridad se combine la copia de seguridad completa con copias de seguridad incrementales o diferenciales.

1 <https://loogic.com/politicas-de-copias-de-seguridad-no-pierdas-tiempo-ni-dinero/>

1.2.2 Copias de seguridad incrementales

Una operación de copia de seguridad incremental sólo copia los datos que han cambiado desde la última operación de copia de seguridad de cualquier tipo. Se suele utilizar la hora y la fecha del archivo, comparándola con la hora y la fecha de la última copia de seguridad.

Su hora de grabación y el espacio que requieren son los más pequeños, pero por el contrario, la recuperación es la más lenta, ya que para obtener un archivo actual es necesario ir recuperando todas las copias hasta la última completa

1.2.3 Copias de seguridad diferenciales

Es una operación similar a la copia de seguridad incremental, pero en este caso cada vez que se ejecuta, continuará copiando todos los datos que han cambiado desde la anterior copia de seguridad completa.

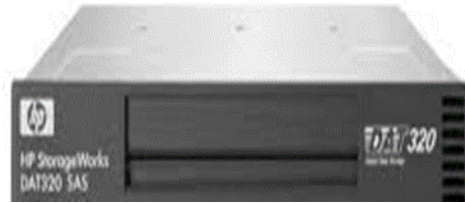
Es una solución intermedia. Almacena más datos que una copia de seguridad incremental, aunque normalmente mucho menos que una copia de seguridad completa. Además, la realización de copias de seguridad diferenciales requiere más espacio y tiempo que las copias de seguridad incrementales, pero menos que las copias de seguridad completas.



2. HARDWARE PARA COPIAS DE SEGURIDAD

Además de los discos duros y los medios ópticos (CD, DVD, Blu-Ray), existen otros sistemas de almacenamiento más enfocados a la creación y restauración de copias de seguridad de gran volumen o al aumento de la fiabilidad de los datos. Dentro de estos sistemas tenemos principalmente:

- Cintas magnéticas. Similares a las viejas cintas de cassette. Son lentas (el tiempo de acceso es de unos segundos) y de acceso secuencial, pero de gran capacidad de almacenamiento (en muchos casos, para aumentar la capacidad de las cintas se comprimen los datos). Su principal tarea es el almacenamiento de grandes cantidades de datos
- Discos duros montados en RAID. Su objetivo es aumentar la fiabilidad de los datos almacenados.



2.1 Sistemas RAID

Los sistemas RAID son uno de los métodos más comunes y estandarizados para aumentar el rendimiento y/o la fiabilidad de los sistemas de almacenamiento en disco. Sus siglas significan Redundant Array of Inexpensive Disks.

La mejor manera de entenderlo es analizando sus propias siglas:

- Matriz de discos: Es una estructura compuesta de varios discos (al menos dos). Cuantos más discos tenga, más fiabilidad.
- Redundante: la información se almacena repetida, ya sea en su totalidad o en parte.
- Económico: El rendimiento de un RAID implementado con discos de gama media puede ser mayor que el de un disco de gama alta.

Los discos utilizados pueden ser de cualquier tipo de interfaz, IDE/PATA, SATA o SCSI, aunque históricamente ha sido el último el que se ha implementado con más asiduidad. Hoy en día, con la mejora en el rendimiento y el precio de los discos duros SATA, se está haciendo muy común ver esta interfaz.

Los sistemas RAID se pueden desplegar de dos maneras:

- A través de software. Es un sistema más lento, que también implica la carga de trabajo del procesador. La gran ventaja es que puede ser implementado en cualquier sistema que contenga al menos dos conectores para discos duros, y por lo tanto es más barato.
- A través del hardware. Mucho más optimizado que el método anterior, ya que todo el trabajo se realiza por un elemento especializado llamado controlador RAID.

Hasta hace poco, las controladoras RAID costaban mucho dinero, pero actualmente la mayoría de las controladoras de discos SATA integran alguna placa RAID (nivel) para que el usuario sólo necesite los discos y la configuración.

Los sistemas RAID se clasifican en los llamados niveles RAID. Cada uno de estos niveles requiere un número de recursos (discos duros) y ofrece una serie de mejoras de fiabilidad y/o rendimiento.

2.2 JBOD (Just a Bunch Of Disks)

Aunque este nivel no es en sí mismo un nivel RAID (no aumentan ni la fiabilidad ni el rendimiento) muchos controladores RAID lo implementan porque es muy simple.

Consiste simplemente en la unión lógica de varios discos duros, de manera que cuando uno se llena, el sistema comienza a utilizar el siguiente, pero sin que el usuario se dé cuenta de nada.

Es útil cuando hay pequeños discos duros que por sí solos no son muy útiles pero que pueden resolver problemas de espacio (por ejemplo si estamos creando video)

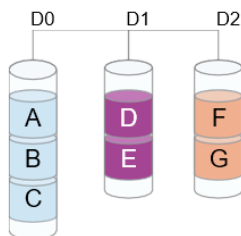


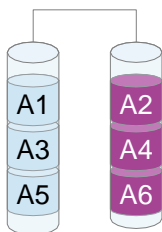
Figure 3. JBOD schema

En la figura se muestra un ejemplo de funcionamiento. Cuando el disco D0 está lleno, el siguiente archivo (D) se guarda automáticamente en D1 sin que el usuario se dé cuenta de nada. Lo mismo ocurre cuando se llena D1; el archivo F se registra automáticamente en D2.

2.3 RAID 0

El propósito del RAID 0 es aumentar el rendimiento porque no tiene redundancia. De hecho, un fallo de uno de los discos del RAID 0 implica la pérdida de toda la información. Para implementar este nivel se necesitan al menos dos discos duros. El sistema divide la información por bloques, distribuyendo los archivos por los N discos duros configurados al mismo tiempo.

La única ventaja de RAID 0 es una mejora en el rendimiento ya que puede ser leído o escrito a la vez en varios discos, pudiendo, teóricamente, conseguir hasta el doble de velocidad con dos discos duros idénticos o incluso más si el número de discos es mayor, siempre que cuando se conecten para que el acceso pueda ser simultáneo



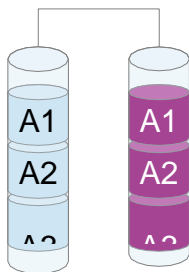
En la figura se muestra un ejemplo de funcionamiento. Un archivo A se divide en bloques que se escriben en diferentes discos duros. Obviamente, si un disco falla, los bloques de ese disco son irrecuperables y por lo tanto también lo es el archivo.

Este sistema sólo debe utilizarse cuando la integridad de la información no sea crítica.

2.4 RAID 1

La redada 1, también, llamada espejamiento. Es un sistema en el que hay una total redundancia ya que se hace una copia en espejo de toda la información almacenada. Su objetivo fundamental es aumentar la fiabilidad, pero también el rendimiento en la lectura, ya que es posible hacerlo leyendo dos discos al mismo tiempo (no el de escritura, que permanece igual).

En caso de error el tiempo de recuperación es 0, ya que mientras se desconecta uno de los discos y se conecta otro es posible utilizar la copia existente en el "espejo" (aunque es aconsejable no hacerlo por seguridad). Una vez conectado el nuevo disco hay un tiempo de copia en el que los datos del espejo se copian en el nuevo disco.

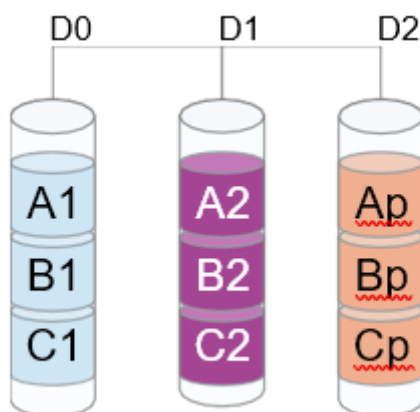


En la figura se muestra un ejemplo de funcionamiento. Un archivo copia sus bloques tanto en el disco 0 como en el disco 1.

Si tienes datos mínimamente importantes es muy interesante, y podríamos decir casi obligatorio, instalarlo.

2.5 RAID 4

El sistema RAID 4 divide los archivos en bloques del tamaño de los sectores del disco duro (normalmente 512 bytes). Esto ya aporta ventajas de rendimiento (y cuanto más discos duros mejor), pero también incluye un sistema de corrección de errores (un ECC, por ejemplo) que ocupa uno de los discos. Por lo tanto, se necesitan al menos 3 discos.



En la figura se muestra un ejemplo de funcionamiento. El archivo A distribuye sus bloques entre los discos D0 y D1, y en el disco D2 almacena la paridad generada del archivo. En caso de fallo de uno de los discos, el sistema es capaz de "sobrevivir", pero no así al fallo de dos de ellos.

2.6 RAID 5

Es el más utilizado en los sistemas profesionales. Es muy similar al RAID 4, pero con la información de paridad distribuida por cada uno de los bloques de datos, siempre en un disco diferente a los que el conjunto de datos se almacenan. Esto significa un rendimiento ligeramente mejor que el de RAID 4 en escritura

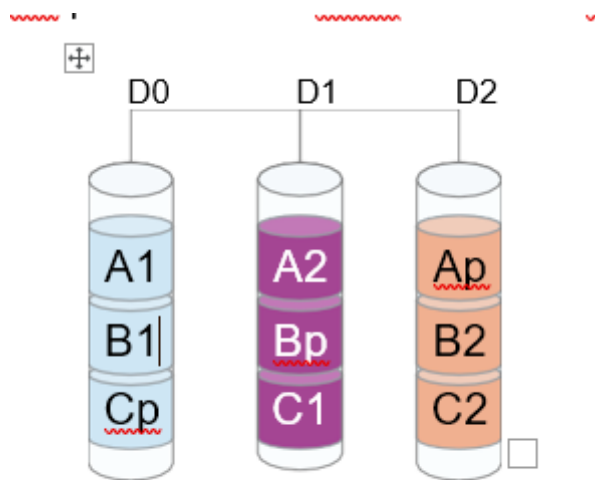


Figure 7. RAID 5 schema

Un sistema RAID 5 sobrevive al fallo de uno de los discos, pero no al fallo de dos, ya que el bloque de paridad puede combinarse matemáticamente y obtener uno de los bloques.

2.7 RAID 0+1 o RAID 01

Es un sistema RAID anidado, por lo que cada sistema RAID es visto por otro sistema RAID como un único disco duro. En este caso, el sistema de discos en RAID 0 unido por RAID 1.

Lo que se consigue es un rendimiento tan alto como el de RAID 0, pero con la tolerancia a fallos que ofrece RAID

1. Se requieren al menos 4 discos y es soportado por casi todos los controladores.

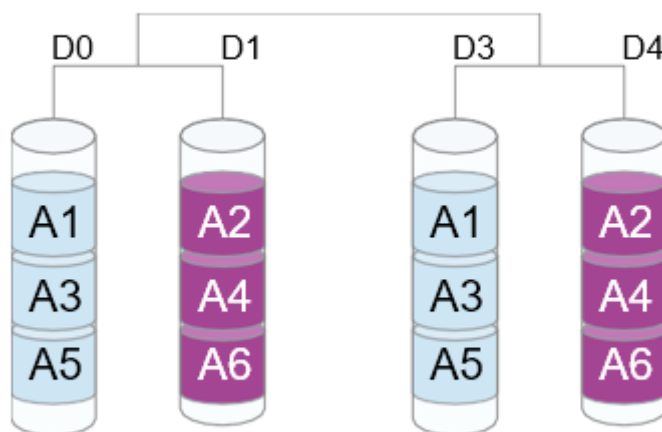


Figure 8. RAID 01 schema

Si un disco de un RAID 0 falla no pasa nada, pero los datos se perderán si falla alguno de los otros discos del RAID 0.

2.8 RAID 10

No lo confundas con el RAID 01. Se utiliza más que el anterior. Es un conjunto de varios discos RAID 1 unidos a través de un RAID 0. Es un poco más tolerante a los fallos, ya que admite el fallo de varios discos siempre que se produzcan en diferentes RAID 1

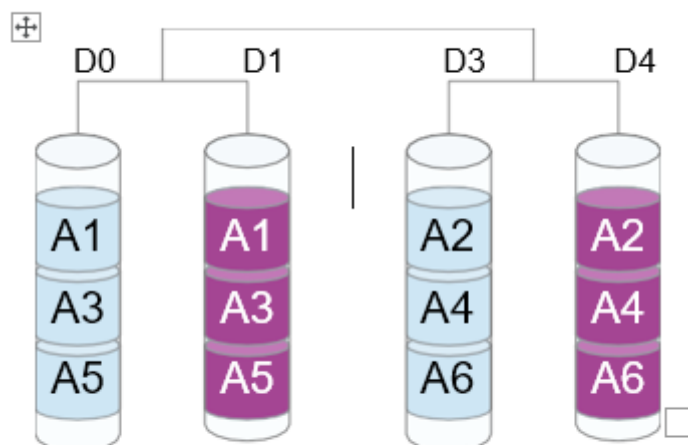


Figure 9. RAID 10 schema

2.9 RAID 50

Es muy similar al RAID 10, pero uniendo varios RAID 5 en lugar del RAID 1. Al utilizar la paridad en lugar de una copia en espejo, se gana más espacio y con un número suficiente de discos duros, la tolerancia a los fallos puede ser mejor.

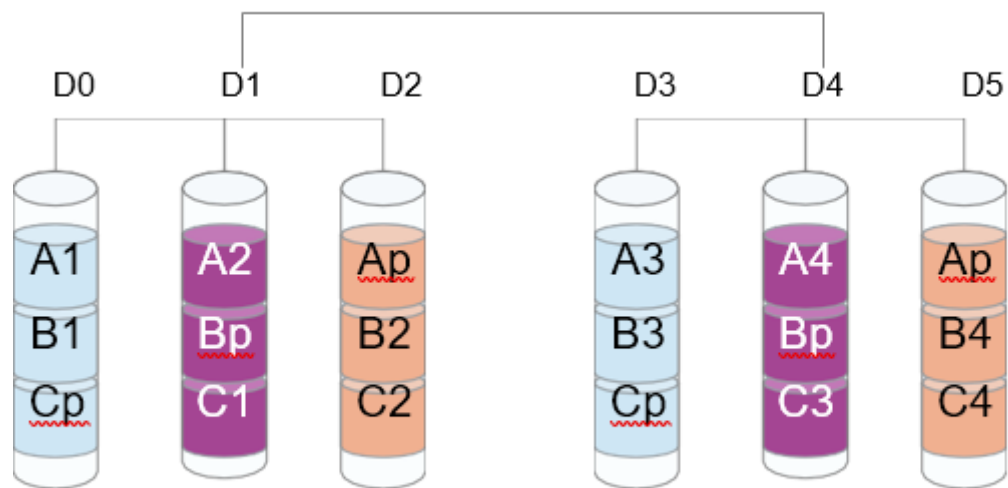


Figure 10. RAID 50 schema