Tema: 5 Copias de seguridad.

1. Introducción.

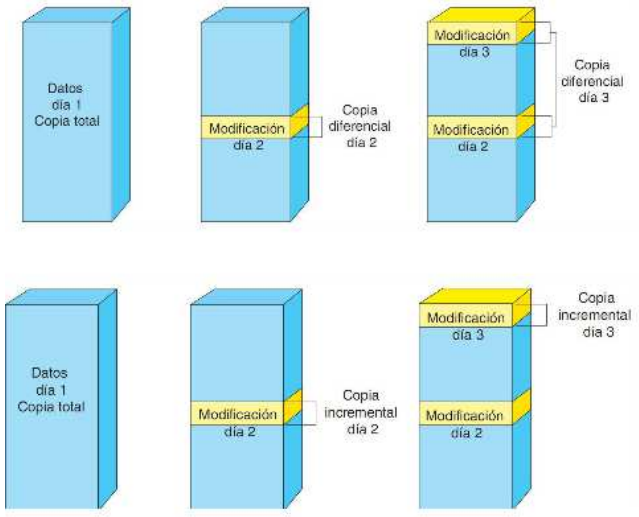
Una copia de seguridad, respaldo, copia de respaldo, copia de reserva (del inglés backup) en ciencias de la información e informática es una copia de los datos originales fuera de la infraestructura que se realiza con el fin de disponer de un medio para recuperarlos en caso de su pérdida. Las copias de seguridad son útiles ante distintos eventos y usos: recuperar los sistemas informáticos y los datos de una catástrofe informática, natural o ataque; restaurar una pequeña cantidad de archivos que pueden haberse eliminado accidentalmente, corrompido, infectado por un virus informático u otras causas; guardar información histórica de forma más económica que los discos duros y además permitiendo el traslado a ubicaciones distintas de la de los datos originales; etc.

El proceso de copia de seguridad se complementa con otro conocido como restauración de los datos, que es la acción de leer y grabar en la ubicación original u otra alternativa los datos requeridos. La pérdida de datos es muy común, el 66 % de los usuarios de Internet han sufrido una seria pérdida de datos en algún momento.

2. Tipos de copias de seguridad.

2.1. Según la cantidad de datos.

En general, existen 3 tipos distintos de copias de seguridad o backups:



- Copia de seguridad completa: como sugiere su nombre, este tipo de backup hace un respaldo completo de todos archivos del nuestro equipo. El backup abarca el 100% de las informaciones, por lo que normalmente requiere más tiempo en realizarse y ocupa más espacio. Si uno está

seguro que quiere protegerlo todo por igual es la mejor solución.

- Copia de seguridad diferencial: únicamente contiene los archivos que han cambiado desde la última vez que se hizo la copia completa. Por lo tanto, se incluyen sólo los archivos nuevos y/o modificados.

- Copia de seguridad incremental: se hace una copia de todos los archivos que han sido modificados desde que fue ejecutado el último backup completo, diferencial o incremental. Es el método más rápido para realizar copias de seguridad.

2.2. Según el tipo de datos.

Otra clasificación de los tipos de copias de seguridad es dependiendo del tipo de información:

- Copias de sistema: Utilizadas para una recuperación rápida del sistema, son menos importantes, ya que se pueden volver a reinstalar todas las aplicaciones que teníamos instalado. Para este tipo de copias existen multitud de aplicaciones de terceros como acronis, clonezilla, Norton ghost, etc. Además, el propio sistema puede incluir utilidades de autorecuperación, como en Windows los puntos de restauración.
- Copias de datos: Son las más importantes, en ellas se copian las ubicaciones deseadas, y desde esa copia se puede restaurar la información, al estado en que se realizó la copia.

Los comandos usuales para realizar copias de seguridad, son en Windows copy, xcopy y robocopy. El Linux el comando es cp. Lo ideal es además de copiar, comprimir y empaquetar. Para ello el comando es tar.

Por último, destacar que desde una copia de seguridad se puede divulgar la información, por lo que las copias de seguridad se deben custodiar como el resto del sistema.

3. Manipulación de los datos de la copia de seguridad

Es una práctica habitual el manipular los datos guardados en las copias de seguridad para optimizar tanto los procesos de copia como el almacenamiento.

- Compresión.

La compresión es el mejor método para disminuir el espacio de almacenamiento necesario y de ese modo reducir el costo.

- Redundancia.

Cuando varios sistemas guardan sus copias de seguridad en el mismo sistema de almacenamiento, existe la posibilidad de redundancia en los datos copiados. Si tenemos estaciones con el mismo sistema operativo compartiendo el mismo almacén de datos, existe la posibilidad de que la mayoría de los archivos del sistema sean comunes. El almacén de datos realmente sólo necesita almacenar una copia de esos ficheros para luego ser utilizada por cualquiera de las estaciones. Esta técnica puede ser aplicada al nivel de ficheros o incluso al nivel de bloques de datos, reduciendo el espacio utilizado para almacenar.

- Deduplicación.

Es una técnica especializada de compresión de datos para eliminar copias duplicadas de datos repetidos. Un término relacionado con la deduplicación de datos es la compresión inteligente de datos. Esta técnica se usa para optimizar el almacenamiento de datos en disco y también para reducir la cantidad de información que debe enviarse de un dispositivo a otro a través de redes de comunicación.:

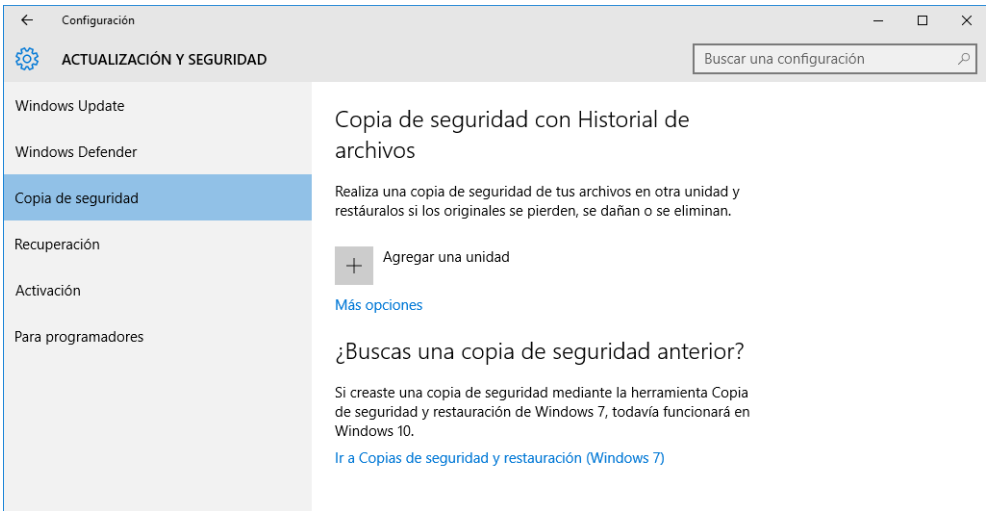
Los procesos de deduplicación a nivel de archivo examinan los ficheros en su totalidad para determinar si están duplicados, lo que se conoce como almacenamiento de instancia única, que es idéntico a un backup incremental clásico. Sin embargo, otros procesos dividen los datos en bloques y tratan de encontrar duplicados en ellos (duplicación al nivel de los bloques). La deduplicación a nivel de bloques produce más granularidad y una reducción mayor del espacio de almacenamiento que la de nivel de archivo. Pero la verdadera potencia está en la deduplicación a nivel de bytes, al realizar una comparación byte a byte de las

corrientes de datos se consigue un mayor nivel de precisión garantizando la eliminación de datos redundantes.

- Cifrado.

La alta capacidad de los soportes de almacenamiento desmontables implica un riesgo de perderse o ser robados. Si se cifra la información de estos soportes se puede mitigar el problema, aunque esto presenta nuevos inconvenientes. Primero, cifrar es un proceso que consume mucho tiempo de CPU y puede bajar la velocidad de copiado. En segundo lugar, una vez cifrados los datos, la compresión es menos eficaz.

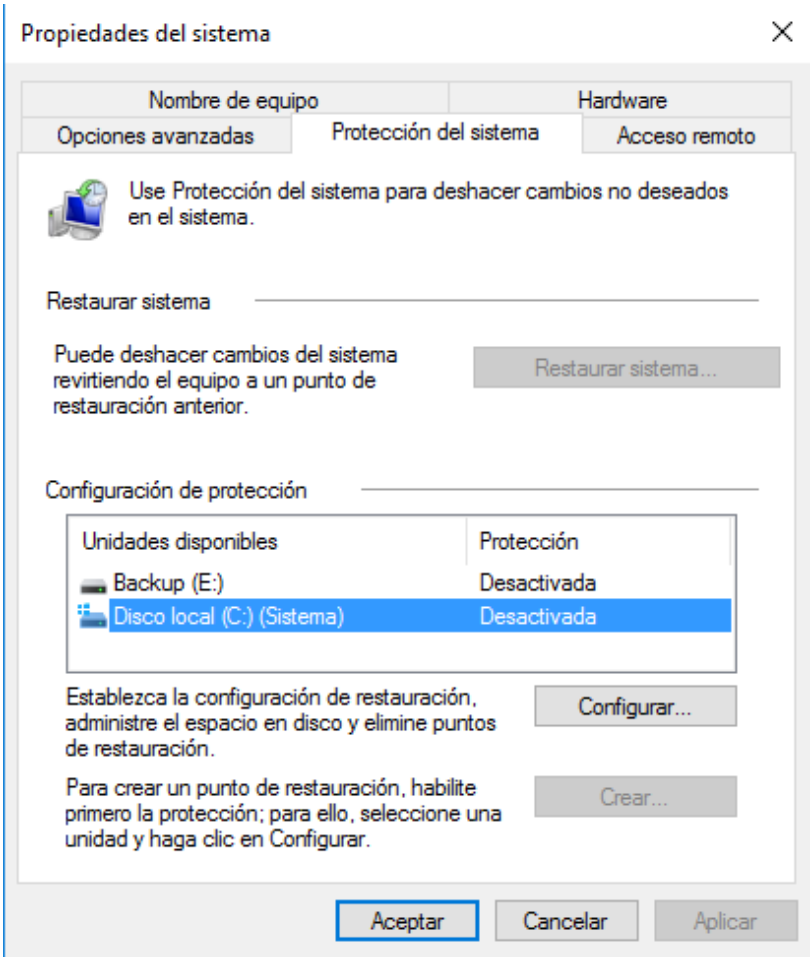
4. Copias en Windows 10.



Cuando vamos a realizar un backup, lo último que queremos es que se cuelen virus, malware, ficheros temporales y otros datos inútiles o peligrosos. Por eso, antes de hacer una copia de seguridad, sigue estos pasos:

- Realiza una limpieza del sistema con un programa como CCleaner Free o equivalente, para eliminar archivos temporales o cachés inservibles.
- Desinstala los programas que realmente ya no utilizas.
- Realiza un chequeo del sistema con un antivirus y un software antimalware del tipo Malwarebytes Antimalware.

Las copias de seguridad se deben realizar en un sistema de ficheros distinto del que se desea respaldar. Para realizar copias de seguridad en Windows 10, icono de Configuración, y luego entra en Actualización y Seguridad y Copias de Seguridad. Verás una opción que se llama Copia de Seguridad con Historial de Archivos. Si la activas y pinchas en Más Opciones, podrás elegir los plazos entre copias o la unidad de destino, pero ahora puedes añadir al backup las carpetas que tú quieres con el botón Agregar una Carpeta:



4.1. Puntos de restauración.

Windows 10 realizará la creación de un punto de restauración de manera automática en el momento en el que se producen cambios importantes en el sistema operativo como por ejemplo una actualización de Microsoft. Sin embargo, esto puede que no sea suficiente para ti y en algún momento decidas crear un punto de restauración. Esto es totalmente posible y sobre todo recomendable en aquellos casos en los que vallamos a realizar cambios en el sistema como por ejemplo la edición del registro del sistema.

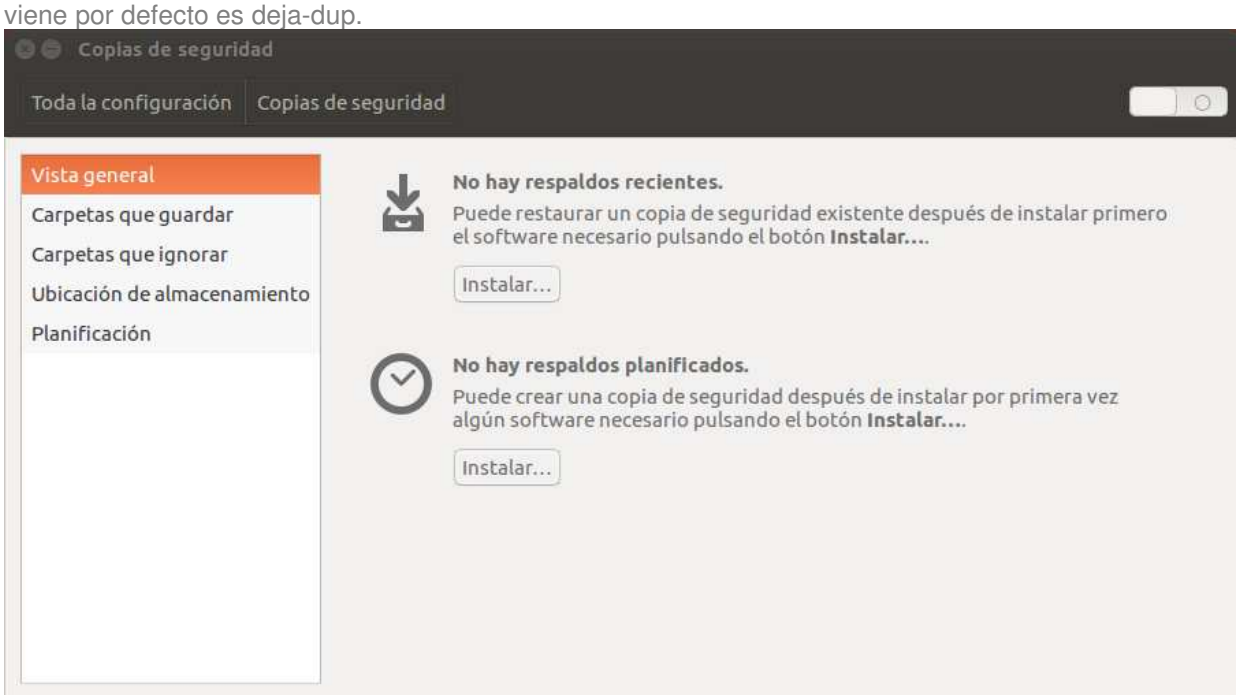
Para la creación de un punto de restauración, lo primero que tendremos que hacer será realizar la búsqueda de la herramienta: Crear un punto de restauración. Para ello podrás búscala desde el cuadro de búsqueda que se encuentra en la barra de tareas de Windows 10. Una vez que la herramienta se muestre en los resultados de búsqueda, solo tendrás que hacer clic sobre ella para abrirla.

Cuando la ventana se halla abierto, lo que haremos primero será configurar el espacio que ocupará nuestro punto de restauración. Para ello tendremos que hacer clic sobre el botón Configurar junto a la descripción: Establezca la configuración de restauración, administre el espacio en disco y elimine puntos de restauración. Esto provocará la apertura de una nueva ventana en la que podrás configurar el uso del espacio que este punto de restauración ocupará en tu disco duro. Recuerda que la opción que debe estar marcada en esta ventana es la de: Activar protección del sistema.

Para restaurar desde un punto de restauración, se accede igual y se selecciona restaurar sistema.

5. Copias en Linux.

En el modo grafico existen varias herramientas de copias de seguridad, en Ubuntu la que viene por defecto es deja-dup.



Como todo sistema UNIX, Linux provee herramientas estándar para realizar las copias de seguridad de los discos.

La utilidad tar es una herramienta disponible en todas las versiones de Unix/Linux, que permite volcar ficheros individuales o directorios en un fichero único. Por si solo el comando tar es un empaquetador, es decir, solo unifica en un único fichero varios de ellos. El comando tar es un comando muy complejo, que ofrece una gran cantidad de opciones. A continuación, algunas de las combinaciones más usadas.

-f : Permite especificar el nombre del archivo.
-c : Crea un nuevo archivo tar.
-r : Agrega archivos a un paquete existente.
-t : Muestra el contenido de un paquete.
-u : Agrega archivos al paquete, pero sólo si estos son más recientes que los existentes.
-x : Desempaqueta archivos de un paquete (extrae).
-z : Comprime con gzip el paquete generado.
-j : Comprime con bzip2 el paquete generado.
-v : Da los nombres de los archivos procesados.
-p : Conserva los permisos de los ficheros.

Los archivos creados mediante tar terminan en .tar. Cuando el archivo tar ha sido además comprimido con gzip, la terminación será .tar.gz, con bzip2, será .tar.bz2.

Las opciones más comunes son:

- Empaquetar: tar -cvf archivo.tar /directorio/
- Desempaquetar: tar -xvf archivo.tar

Si se usa la opción z, es equivalente a realizarlo en 2 pasos, uno para empaquetar y otro para comprimir con el comando gzip. Para comprimir se utiliza:

gzip -9 archivo.tar

El resultado será archivo.tar.gz.

Para descomprimir, se puede poner gzip -d y el nombre del archivo gz a descomprimir.

gzip -d archivo.tar.gz

En esta ocasión archivo.tar.gz será descomprimido y reemplazado por archivo.tar. Un equivalente a gzip -d es el comando gunzip:

gunzip archivo.tar.gz

La cantidad de compresión obtenida depende de varios factores, típicamente, texto o código fuente se reduce en un porcentaje del 60 al 70%. Un fichero ya comprimido, como la mayoría de archivos gráficos (gif, jpg), difícilmente reducirán su tamaño.
Desde modo el menú contextual también se puede empaquetar y comprimir seleccionando comprimir.

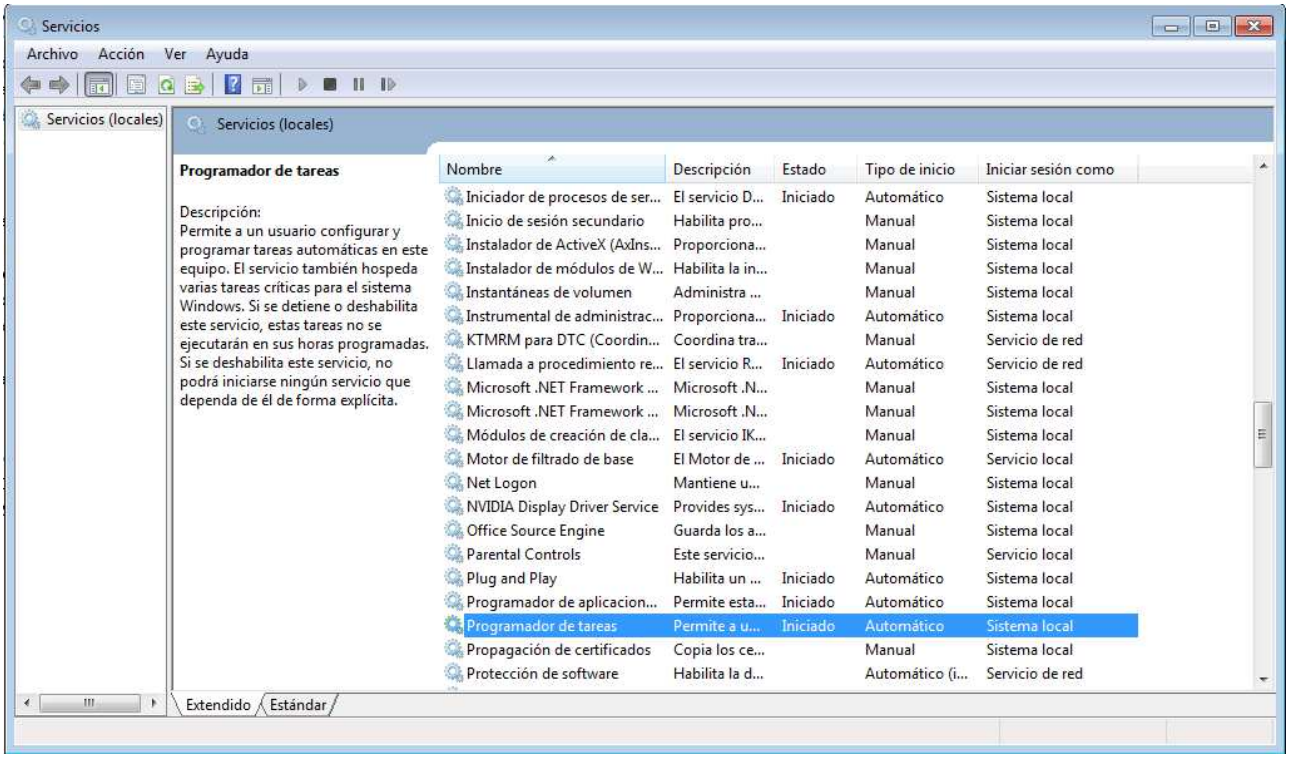
En Linux existen varias herramientas de compresión, que generan distintos tipos de archivos. Así las siguientes extensiones de archivo están asociadas a distintas herramientas, Zip, .Rar, .Gz, .bz2, .Bzip2, .Lha, .Arj, .Zoo.

6. Tareas programadas en Windows.

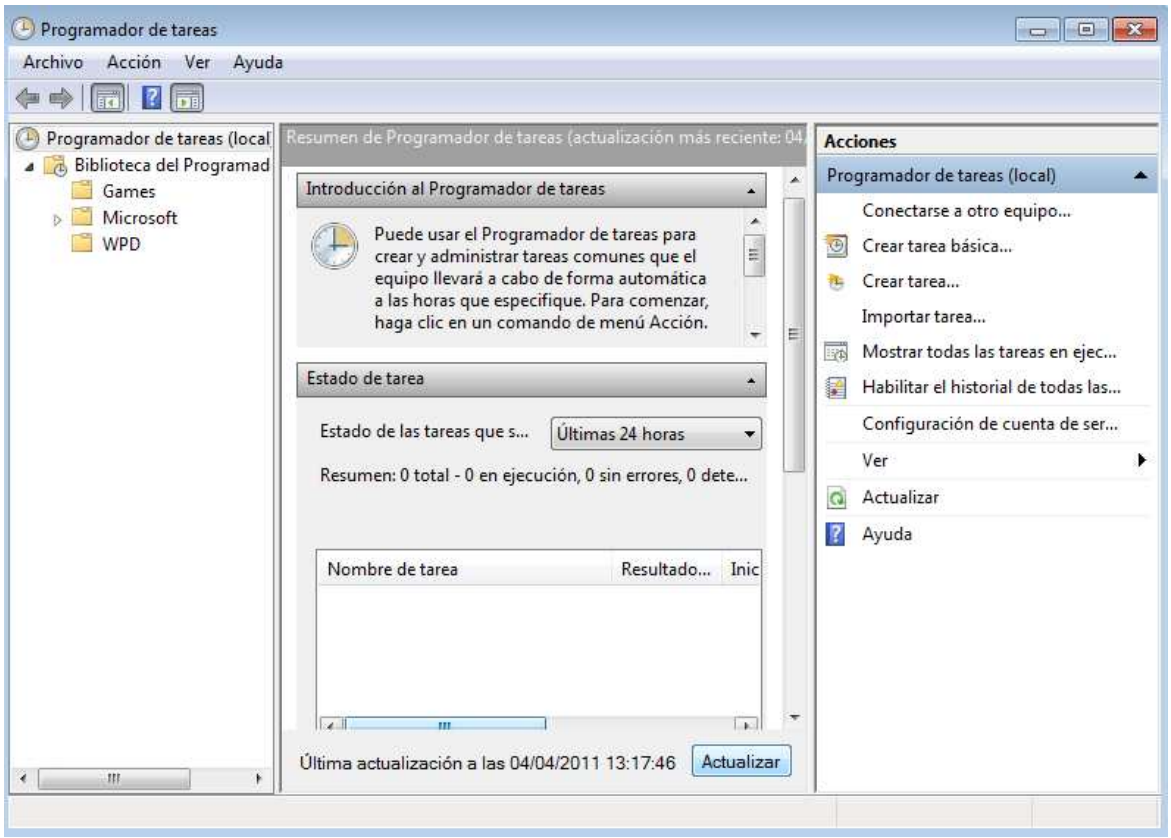
Windows cuenta con un programador de tareas que nos permite:

- Programar una tarea para que se ejecute diaria, semanal o mensualmente, o determinadas horas (por ejemplo, al iniciarse el sistema).
- Cambiar la programación de una tarea.
- Detener una tarea programada.
- Personalizar la forma en que se ejecutará una tarea a la hora programada.

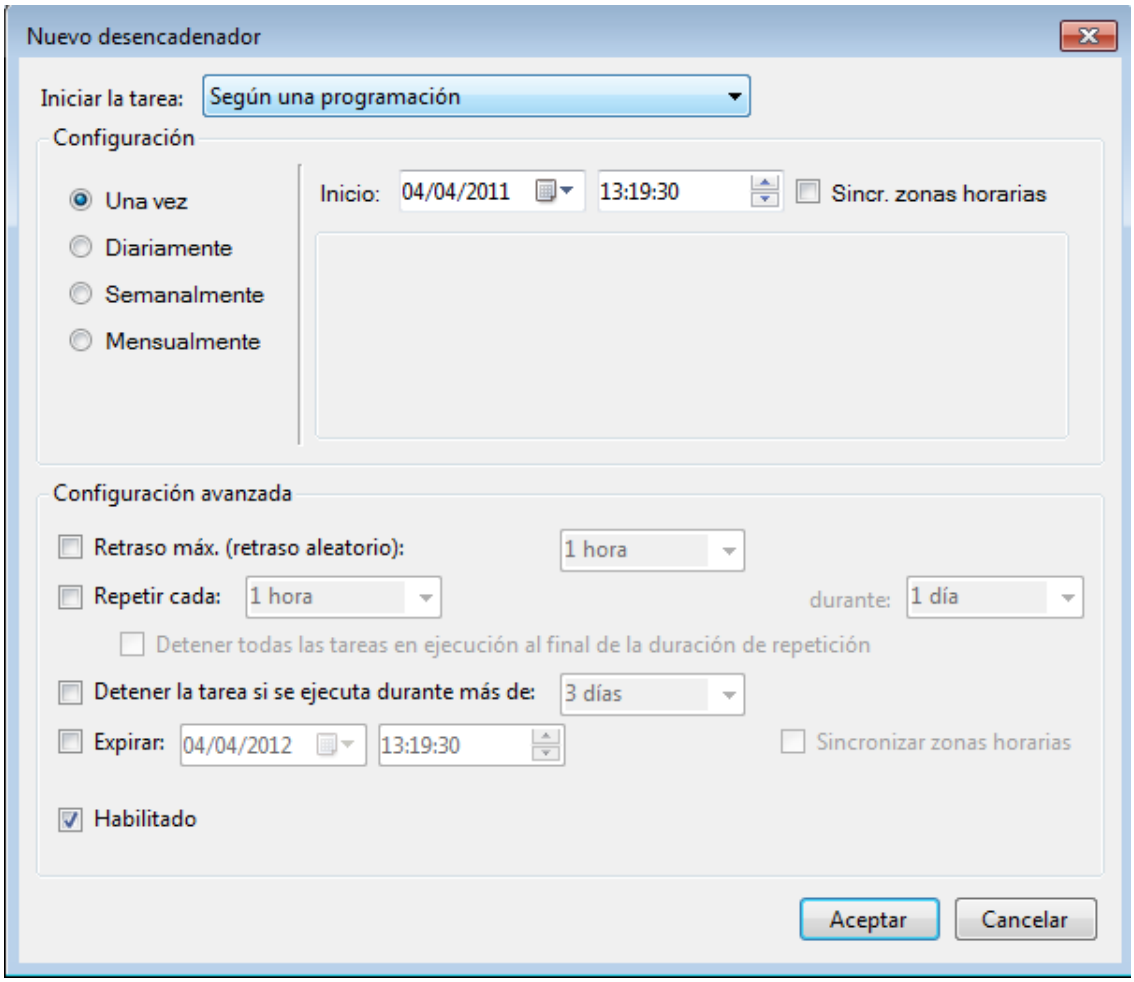
Para ello, se tiene que poner en marcha el servicio correspondiente, el servicio de programador de tareas:



Desde Windows 7, se accede por panel de control ⇒ Sistema y seguridad ⇒ Herramientas administrativas y, a continuación, haga doble clic en Programador de tareas.



Si pulsamos en crear tareas.



También se pueden programar tareas desde línea de comandos. La utilidad se llama schtasks (de scheduler – planificador- y tasks – tareas).

Los parámetros que incluyen son:

/Create	Crea una nueva tarea programada.
/Delete	Borra las tareas programadas.
/Query	Muestra todas las tareas programadas.
/Change	Cambia las propiedades de la tarea programada.
/Run	Ejecuta la tarea programada inmediatamente.
/End	Detiene la tarea programada que se está ejecutando actualmente
/?	Muestra esta ayuda/uso.

7. Tareas programas en Linux.

Existen en Linux, varios programas que realizarían la programación de tareas. El más común para gnome es gnome-schedule, aunque existen otros como kalarm o kcron. Para ello sería necesario instalar la herramienta.

En Linux existen comandos para realizar la programación de tareas. En particular existen 3 que suelen ser complementarias:

- cron: para planificar tareas periódicas, especificando el momento.
- anacron: es un programa libre que ejecuta asíncronamente tareas programadas en sistemas UNIX de manera periódica.
- at: es una herramienta que permite programar la ejecución de uno o varios programas en un momento futuro.

De todos el más importante y utilizado es cron. Anacron es útil en equipos que no funcionan todo el día, mientras que en servidores cuya ejecución no para, se utiliza

cron.cron es un administrador regular de procesos en segundo plano (*demonio*) que ejecuta procesos o guiones a intervalos regulares (por ejemplo, cada minuto, día, semana o mes). Los procesos que deben ejecutarse y la hora en la que deben hacerlo se especifican en el fichero crontab. El nombre *cron* viene del griego *chronos* que significa "tiempo".

Para que se ejecute las tareas es necesario que el servicio cron este corriendo. Anteriormente este servicio se llamaba crond.

Para utilizar el programador de tareas existen 3 opciones:

crontab –e : para añadir una nueva tarea.

crontab –l :para listar las tareas programadas.

crontab –r : para eliminar todas las tareas.

Un cron job del usuario consta de seis campos y luce como la siguiente línea:

1 2 3 4 5 /path/to/command arg1 arg2

Donde:

1: Minuto (0-59)

2: Horas (0-23)

3: Día (0-31)

4: Mes (0-12 [12 == diciembre])

5: Día de la semana (0-7 [7 or 0 == domingo])

```
minuto (0-59),
|  hora (0-23),
|  |  día del mes (1-31),
|  |  |  mes (1-12
|  |  |  |  día de la semana (0-6 donde 0=Domingo)
|  |  |  |  |   Script o comando.
15 02 *  *  * /path/to/command arg1 arg2
```

Ejemplo de tareas programadas seria:

Por ejemplo:

*30 10 * * 1 root /usr/bin/who >> /home/quien.tex*

Ejecuta la orden **who** todos los lunes a las **10:30** y guarda la salida en el fichero *quien.tex*

Para especificar dos o más valores en cada variable, estas deben estar separadas por comas, siguiendo con el ejemplo anterior:

*0,30 * * * 1 root /usr/bin/who >> /home/quien.tex*

Ejecuta la orden who todos los lunes cada media hora y guarda la salida en el fichero *quien.tex*

Si queremos que se ejecute cada 15 minutos sería:

*0,15,30,45 * * * * root /usr/bin/who >> /home/quien.tex*

o

**/15 * * * * root /usr/bin/who >> /home/quien.tex*

En este ejemplo veremos como pasarle más de un comando al cron y de paso como puede programarse una descarga:

*30 21 * * * root cd /media/sda7/dexter/distributions/isos;wget* http://example.com/fichero_a_descargar.loquesea

Este otro es para programar el apagado del PC. En este caso todos los sábados a las 21.30

*30 21 * * 6 root /sbin/shutdown -h now*

Hay varios valores predefinidos que se pueden utilizar para sustituir la expresión CRON.

Entrada	Descripción	Equivale A
@yearly	Se ejecuta una vez al año	0 0 1 1 *
@annually	(igual que @yearly)	0 0 1 1 *
@monthly	Se ejecuta una vez al mes	0 0 1 * *
@weekly	Se ejecuta una vez a la semana	0 0 * * 0
@daily	Se ejecuta una vez al día	0 0 * * *

@midnight	(igual que @daily)	0 0 * * *
@hourly	Se ejecuta una vez cada hora	0 * * * *

También está disponible @reboot, que permite a un trabajo ejecutarse una vez cada vez que el demonio cron se inicie, que eso típicamente coincidirá con el arranque del servidor.

Así se puede saber cada hora que usuarios están en el sistema:

@hourly root /usr/bin/who >> /home/quien.tex

8. Asegurar la información.

8.1. Particiones.

Una partición de disco, es el nombre genérico que recibe cada división presente en una sola unidad física de almacenamiento de datos. Toda partición tiene su propio sistema de archivos (formato); generalmente, casi cualquier sistema operativo interpreta, utiliza y manipula cada partición como un disco físico independiente, a pesar de que dichas particiones estén en un solo disco físico.

Existen distintos esquemas de particiones para la distribución de particiones en un disco. Los más conocidos y difundidos son MBR (Master Boot Record) y GPT (GUID Partition Table).

En Windows, las particiones reconocidas son identificadas con una letra seguida por dos puntos (por ejemplo, C:), aunque también pueden ser montadas en directorios (por ejemplo C:\Users). Prácticamente todo tipo de discos magnéticos y memorias flash (como pendrives) pueden particionarse. En sistemas UNIX y UNIX-like, las particiones de datos son montadas en un mismo y único árbol jerárquico, en el cual se montan a través de una carpeta, proceso que sólo el superusuario (root) puede realizar.

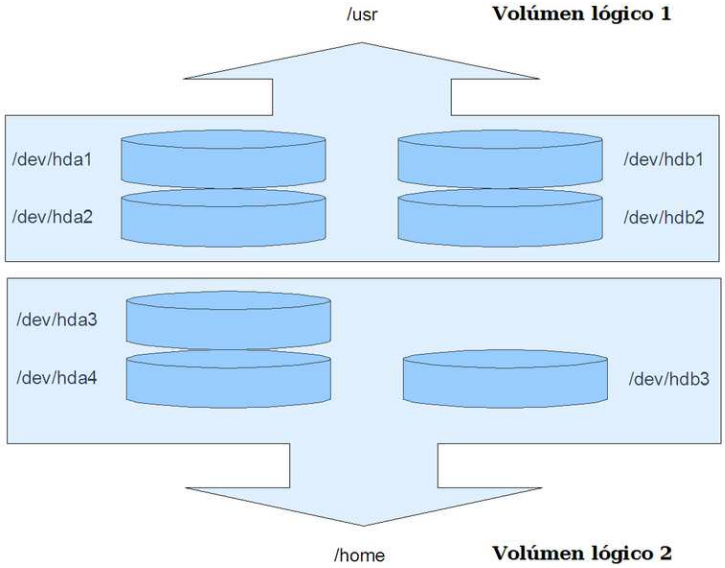
Es común que en los sistemas basados o similares a UNIX generalmente se usan hasta con 3 particiones: la principal, montada en el directorio raíz (/); una segunda que se usa para montar el directorio /home, el cual contiene las configuraciones de los usuarios; y finalmente, una tercera llamada swap, que se usa para la memoria virtual temporal. Sin embargo, 2 particiones (/, y swap) es el mínimo suficiente en estos sistemas operativos. A las particiones de intercambio (swap) no se les asigna un directorio; este tipo de particiones se usa para guardar ciertas réplicas de la memoria RAM, para que de esta forma la RAM tenga más espacio para las tareas en primer plano, guardando las tareas en segundo plano dentro de la partición de intercambio. Algunos sistemas tipo UNIX están diseñados para funcionar con una sola partición, sin embargo, estos diseños no son muy comunes.

8.2. Volúmenes.

En sistemas operativos, el término volumen hace referencia a una única área accesible de almacenamiento, como ser una partición de un disco duro.

Un volumen posee una letra de unidad (C:, D:, E:, etc.), un tipo de sistema de archivo, y un medio de almacenamiento correspondiente (disco duro, CD-ROM, disquetera, etc.).

Por ejemplo, si un disco duro está particionado en tres, significa que posee tres volúmenes, pero esto no implica que volumen sea sinónimo de partición (pero volumen sí es sinónimo de unidad lógica). Cada volumen puede tener un sistema de archivo distinto asociado.



También un sistema operativo puede reconocer una partición sin haber reconocido un volumen asociado a esta, y esto puede suceder porque el sistema operativo no reconoce el sistema de archivos que utiliza esa partición.

En definitiva, los volúmenes existen a nivel lógico en el sistema operativo, en cambio, las particiones existen a nivel físico en un medio de almacenamiento. Por lo general hay una correspondencia uno a uno, pero no siempre.

En Windows la gestión de volúmenes se realiza desde el comando diskpart.

En Linux se instala la herramienta lvm2.

8.3. Discos básicos y dinámicos.

Se distinguen dos tipos de discos: discos dinámicos o discos básicos. Los discos básicos se utilizan de la forma tradicional, mientras que los discos duros dinámicos permiten la utilización de volúmenes.

Desde Windows para convertir el tipo de discos, se utiliza la herramienta diskpart, select disk <disco>, convert <dinamic-basic>

8.4. RAID.

RAID (del inglés redundant array of independent disks), hace referencia a un sistema de almacenamiento de datos que utiliza múltiples unidades de almacenamiento de datos (discos duros o SSD) entre los que se distribuyen o replican los datos.

Dependiendo de su configuración (a la que suele llamarse nivel), los beneficios de un RAID respecto a un único disco son uno o varios de los siguientes: mayor integridad, tolerancia frente a fallos, tasa de transferencia y capacidad. En sus implementaciones originales, su ventaja clave era la habilidad de combinar varios dispositivos de bajo coste y tecnología más vieja en un conjunto que ofrecía mayor capacidad, fiabilidad, velocidad o una combinación de éstas que un solo dispositivo de última generación y coste más alto.

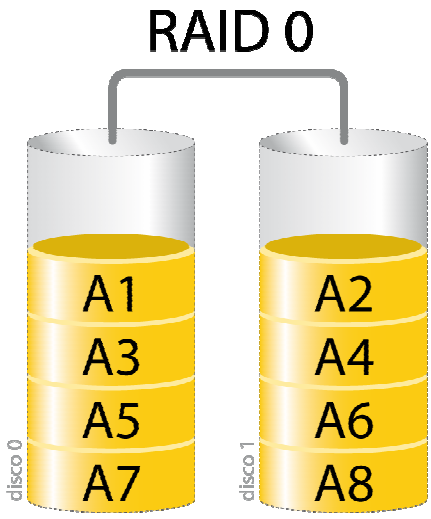
En el nivel más simple, un RAID combina varios discos duros en una sola unidad lógica. Así, en lugar de ver varios discos duros diferentes, el sistema operativo ve uno solo. Los RAID suelen usarse en servidores y normalmente (aunque no es necesario) se implementan con unidades de disco de la misma capacidad. Debido al descenso en el precio de los discos duros y la mayor disponibilidad de las opciones RAID incluidas en los chipsets de las placas base, los RAID se encuentran también como opción en las computadoras personales más avanzadas. Esto es especialmente frecuente en las computadoras dedicadas a tareas intensivas y que requiera asegurar la integridad de los datos en caso de fallo del sistema. Esta característica está disponible en los sistemas RAID por hardware (dependiendo de qué estructura elijamos). Por el contrario, los sistemas basados en software son mucho más flexibles y los basados en hardware añaden un punto de fallo más al sistema (la controladora RAID).

Todas las implementaciones pueden soportar el uso de uno o más discos de reserva (hot spare), unidades preinstaladas que pueden usarse inmediatamente (y casi siempre automáticamente) tras el fallo de un disco del RAID. Esto reduce el tiempo del período de reparación al acortar el tiempo de reconstrucción del RAID.

Los niveles RAID más comúnmente usados son:

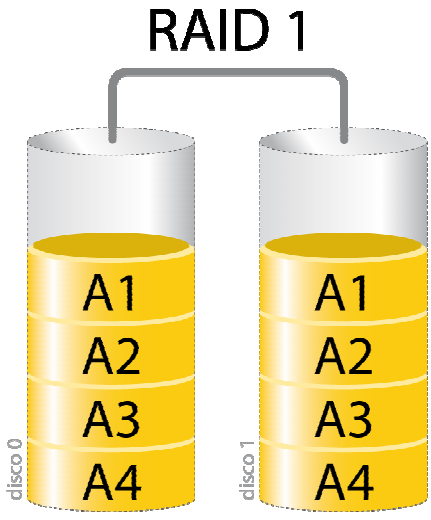
- RAID 0: Conjunto dividido
- RAID 1: Conjunto en espejo
- RAID 5: Conjunto dividido con paridad distribuida.

8.5. RAID 0 (Data Striping, Striped Volume)



Un RAID 0 (también llamado conjunto dividido, volumen dividido, volumen seccionado) distribuye los datos equitativamente entre dos o más discos (usualmente se ocupa el mismo espacio en dos o más discos) sin información de paridad que proporcione redundancia. Es importante señalar que el RAID 0 no era uno de los niveles RAID originales y que no es redundante. El RAID 0 se usa normalmente para proporcionar un alto rendimiento de escritura ya que los datos se escriben en dos o más discos de forma paralela, aunque un mismo fichero solo está presente una vez en el conjunto. RAID 0 también puede utilizarse como forma de crear un pequeño número de grandes discos virtuales a partir de un gran número de pequeños discos físicos. Un RAID 0 puede ser creado con discos de diferentes tamaños, pero el espacio de almacenamiento añadido al conjunto estará limitado por el tamaño del disco más pequeño (por ejemplo, si un disco de 450 GB se divide con uno de 100 GB, el tamaño del conjunto resultante será sólo de 200 GB, ya que cada disco aporta 100 GB). Una buena implementación de un RAID 0 dividirá las operaciones de lectura y escritura en bloques de igual tamaño, por lo que distribuirá la información equitativamente entre los dos discos. También es posible crear un RAID 0 con más de dos discos, si bien, la fiabilidad del conjunto será igual a la fiabilidad media de cada disco entre el número de discos del conjunto; es decir, la fiabilidad total —medida como MTTF o MTBF— es (aproximadamente) inversamente proporcional al número de discos del conjunto (pues para que el conjunto falle es suficiente con que lo haga cualquiera de sus discos). No debe confundirse RAID 0 con un Volumen Distribuido (Spanned Volume) en el cual se agregan múltiples espacios no usados de varios discos para formar un único disco virtual. Puede que en un Volumen Distribuido el fichero a recuperar esté presente en un solo disco del conjunto debido a que aquí no hay una distribución equitativa de los datos (como dijimos para RAID 0), por lo tanto, en ese caso no sería posible la recuperación paralela de datos y no tendríamos mejora del rendimiento de lectura.

8.6. RAID 1 (espejo)



Un RAID 1 crea una copia exacta (o espejo) de un conjunto de datos en dos o más discos. Esto resulta útil cuando queremos tener más seguridad desaprovechando capacidad, ya que, si perdemos un disco, tenemos el otro con la misma información. Un conjunto RAID 1 sólo puede ser tan grande como el más pequeño de sus discos. Un RAID 1 clásico consiste en dos discos en espejo, lo que incrementa exponencialmente la fiabilidad respecto a un solo disco; es decir, la probabilidad de fallo del conjunto es igual al producto de las probabilidades de fallo de cada uno de los discos (pues para que el conjunto falle es necesario que lo hagan todos sus discos).

Adicionalmente, dado que todos los datos están en dos o más discos, con hardware habitualmente independiente, el rendimiento de lectura se incrementa aproximadamente como múltiplo lineal del número de copias; es decir, un RAID 1 puede estar leyendo simultáneamente dos datos diferentes en dos discos diferentes, por lo que su rendimiento se duplica. Para maximizar los beneficios sobre el rendimiento del RAID 1 se recomienda el uso de controladoras de disco independientes, una para cada disco (práctica que algunos denominan splitting o duplexing).

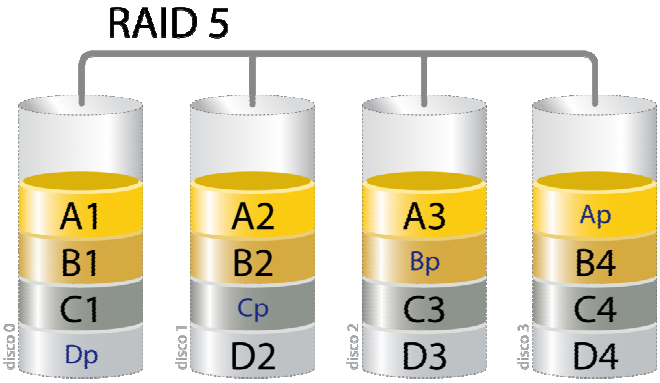
Como en el RAID 0, el tiempo medio de lectura se reduce, ya que los sectores a buscar pueden dividirse entre los discos, bajando el tiempo de búsqueda y subiendo la tasa de transferencia, con el único límite de la velocidad soportada por la controladora RAID. Sin embargo, muchas tarjetas RAID 1 IDE antiguas leen sólo de un disco de la pareja, por lo que su rendimiento es igual al de un único disco. Algunas implementaciones RAID 1 antiguas también leen de ambos discos simultáneamente y comparan los datos para detectar errores.

Al escribir, el conjunto se comporta como un único disco, dado que los datos deben ser escritos en todos los discos del RAID 1. Por tanto, el rendimiento de escritura no mejora.

El RAID 1 tiene muchas ventajas de administración. Por ejemplo, en algunos entornos 24/7, es posible «dividir el espejo»: marcar un disco como inactivo, hacer una copia de seguridad de dicho disco y luego «reconstruir» el espejo. Esto requiere que la aplicación de gestión del conjunto soporte la recuperación de los datos del disco en el momento de la división. Este procedimiento es menos crítico que la presencia de una característica de snapshot en algunos sistemas de archivos, en la que se reserva algún espacio para los cambios, presentando una vista estática en un punto temporal dado del sistema de archivos. Alternativamente, un conjunto de discos puede ser almacenado de forma parecida a como se hace con las tradicionales cintas.

8.7. RAID 5.

Un RAID 5 (también llamado distribuido con paridad) es una división de datos a nivel de bloques que distribuye la información de paridad entre todos los discos miembros del conjunto. El RAID 5 ha logrado popularidad gracias a su bajo coste de redundancia. Generalmente, el RAID 5 se implementa con soporte hardware para el cálculo de la paridad. RAID 5 necesitará un mínimo de 3 discos para ser implementado.



En el gráfico de ejemplo anterior, una petición de lectura del bloque «A1» sería servida por el disco 0. Una petición de lectura simultánea del bloque «B1» tendría que esperar, pero una petición de lectura de «B2» podría atenderse concurrentemente ya que sería servida por el disco 1.

Cada vez que un bloque de datos se escribe en un RAID 5, se genera un bloque de paridad dentro de la misma división (stripe). Un bloque se compone a menudo de muchos sectores consecutivos de disco. Una serie de bloques (un bloque de cada uno de los discos del conjunto) recibe el nombre colectivo de división (stripe). Si otro bloque, o alguna porción de un bloque, es escrita en esa misma división, el bloque de paridad (o una parte del mismo) es recalculada y vuelta a escribir. El disco utilizado por el bloque de paridad está escalonado de una división a la siguiente, de ahí el término «bloques de paridad distribuidos». Las escrituras en un RAID 5 son costosas en términos de operaciones de disco y tráfico entre los discos y la controladora.

Los bloques de paridad no se leen en las operaciones de lectura de datos, ya que esto sería una sobrecarga innecesaria y disminuiría el rendimiento. Sin embargo, los bloques de paridad se leen cuando la lectura de un sector de datos provoca un error de CRC. En este caso, el sector en la misma posición relativa dentro de cada uno de los bloques de datos restantes en la división y dentro del bloque de paridad en la división se utilizan para reconstruir el sector erróneo. El error CRC se oculta así al resto del sistema. De la misma forma, si falla un disco del conjunto, los bloques de paridad de los restantes discos son combinados matemáticamente con los bloques de datos de los restantes discos para reconstruir los datos del disco que ha fallado «al vuelo».

Lo anterior se denomina a veces Modo Interino de Recuperación de Datos (Interim Data Recovery Mode). El sistema sabe que un disco ha fallado, pero sólo con el fin de que el sistema operativo pueda notificar al administrador que una unidad necesita ser reemplazada: las aplicaciones en ejecución siguen funcionando ajenas al fallo. Las lecturas y escrituras continúan normalmente en el conjunto de discos, aunque con alguna degradación de rendimiento. La diferencia entre el RAID 4 y el RAID 5 es que, en el Modo Interno de Recuperación de Datos, el RAID 5 puede ser ligeramente más rápido, debido a que, cuando el CRC y la paridad están en el disco que falló, los cálculos no tienen que realizarse, mientras que en el RAID 4, si uno de los discos de datos falla, los cálculos tienen que ser realizados en cada acceso.

El fallo de un segundo disco provoca la pérdida completa de los datos.

El número máximo de discos en un grupo de redundancia RAID 5 es teóricamente ilimitado, pero en la práctica es común limitar el número de unidades. Los inconvenientes de usar grupos de redundancia mayores son una mayor probabilidad de fallo simultáneo de dos discos, un mayor tiempo de reconstrucción y una mayor probabilidad de hallar un sector irrecuperable durante una reconstrucción. A medida que el número de discos en un conjunto RAID 5 crece, el MTBF (tiempo medio entre fallos) puede ser más bajo que el de un único disco. Esto sucede cuando la probabilidad de que falle un segundo disco en los N-1 discos restantes de un conjunto en el que ha fallado un disco en el tiempo necesario para detectar, reemplazar y recrear dicho disco es mayor que la probabilidad de fallo de un único disco. Una alternativa que proporciona una protección de paridad dual, permitiendo así mayor número de discos por grupo, es el RAID 6.

Algunos vendedores RAID evitan montar discos de los mismos lotes en un grupo de redundancia para minimizar la probabilidad de fallos simultáneos al principio y el final de su vida útil.

Las implementaciones RAID 5 presentan un rendimiento malo cuando se someten a cargas de trabajo que incluyen muchas escrituras más pequeñas que el tamaño de una división (stripe). Esto se debe a que la paridad debe ser actualizada para cada escritura, lo que exige realizar secuencias de lectura, modificación y escritura tanto para el bloque de datos como para el de paridad. Implementaciones más complejas incluyen a menudo cachés de escritura no volátiles para reducir este problema de rendimiento.

En el caso de un fallo del sistema cuando hay escrituras activas, la paridad de una división (stripe) puede quedar en un estado inconsistente con los datos. Si esto no se detecta y repara antes de que un disco o bloque falle, pueden perderse datos debido a que se usará una paridad incorrecta para reconstruir el bloque perdido en dicha división. Esta potencial vulnerabilidad se conoce a veces como «agujero de escritura». Son comunes el uso de caché no volátiles y otras técnicas para reducir la probabilidad de ocurrencia de esta vulnerabilidad.

Muchas controladoras permiten anidar niveles RAID, es decir, que un RAID pueda usarse como elemento básico de otro en lugar de discos físicos. Resulta instructivo pensar en estos conjuntos como capas dispuestas unas sobre otras, con los discos físicos en la inferior.

Los RAID anidados se indican normalmente uniendo en un solo número los correspondientes a los niveles RAID usados, añadiendo a veces un «+» entre ellos. Por ejemplo, el RAID 10 (o RAID 1+0) consiste conceptualmente en múltiples conjuntos de nivel 1 almacenados en discos físicos con un nivel 0 encima, agrupando los anteriores niveles 1. En el caso del RAID 0+1 se usa más esta forma que RAID 01 para evitar la confusión con el RAID 1. Sin embargo, cuando el conjunto de más alto nivel es un RAID 0 (como en el RAID 10 y en el RAID 50), la mayoría de los vendedores eligen omitir el «+», a pesar de que RAID 5+0 sea más informativo.

Lo que RAID puede hacer.

- RAID puede mejorar el uptime. Los niveles RAID 1, 0+1 o 10 o 5 (sus variantes, como el 50) permiten que un disco falle mecánicamente y que aun así los datos del conjunto sigan siendo accesibles para los usuarios. En lugar de exigir que se

realice una restauración costosa en tiempo desde una cinta, DVD o algún otro medio de respaldo lento, un RAID permite que los datos se recuperen en un disco de reemplazo a partir de los restantes discos del conjunto, mientras al mismo tiempo permanece disponible para los usuarios en un modo degradado. Esto es muy valorado por las empresas, ya que el tiempo de no disponibilidad suele tener graves repercusiones. Para usuarios domésticos, puede permitir el ahorro del tiempo de restauración de volúmenes grandes, que requerirían varios DVD o cintas para las copias de seguridad.

- RAID puede mejorar el rendimiento de ciertas aplicaciones. Los niveles RAID 0 y 5 usan variantes de división (stripping) de datos, lo que permite que varios discos atiendan simultáneamente las operaciones de lectura lineales, aumentando la tasa de transferencia sostenida. Las aplicaciones de escritorio que trabajan con archivos grandes, como la edición de vídeo e imágenes, se benefician de esta mejora. También es útil para las operaciones de copia de respaldo de disco a disco. Además, si se usa un RAID 1 o un RAID basado en división con un tamaño de bloque lo suficientemente grande se logran mejoras de rendimiento para patrones de acceso que implique múltiples lecturas simultáneas (por ejemplo, bases de datos multiusuario).

Lo que RAID no puede hacer.

- RAID no protege los datos. Un conjunto RAID tiene un sistema de archivos, lo que supone un punto único de fallo al ser vulnerable a una amplia variedad de riesgos aparte del fallo físico de disco, por lo que RAID no evita la pérdida de datos por estas causas. RAID no impedirá que un virus destruya los datos, que éstos se corrompan, que sufran la modificación o borrado accidental por parte del usuario ni que un fallo físico en otro componente del sistema afecte a los datos.
- RAID no simplifica la recuperación de un desastre. Cuando se trabaja con un solo disco, éste es accesible normalmente mediante un controlador ATA o SCSI incluido en la mayoría de los sistemas operativos. Sin embargo, las controladoras RAID necesitan controladores software específico. Las herramientas de recuperación que trabajan con discos simples en controladoras genéricas necesitarán controladores especiales para acceder a los datos de los conjuntos RAID. Si estas herramientas no los soportan, los datos serán inaccesibles para ellas.
- RAID no mejora el rendimiento de todas las aplicaciones. Esto resulta especialmente cierto en las configuraciones típicas de escritorio. La mayoría de aplicaciones de escritorio y videojuegos hacen énfasis en la estrategia de buffering y los tiempos de búsqueda de los discos. Una mayor tasa de transferencia sostenida supone poco beneficio para los usuarios de estas aplicaciones, al ser la mayoría de los archivos a los que se accede muy pequeños. La división de discos de un RAID 0 mejora el rendimiento de transferencia lineal pero no lo demás, lo que hace que la mayoría de las aplicaciones de escritorio y juegos no muestren mejora alguna, salvo excepciones. Para estos usos, lo mejor es comprar un disco más grande y rápido, en lugar de dos discos más lentos y pequeños en una configuración RAID 0.
- RAID no facilita el traslado a un sistema nuevo. Cuando se usa un solo disco, es relativamente fácil trasladar el disco a un sistema nuevo: basta con conectarlo, si cuenta con la misma interfaz. Con un RAID no es tan sencillo: la BIOS RAID debe ser capaz de leer los metadatos de los miembros del conjunto para reconocerlo adecuadamente y hacerlo disponible al sistema operativo. Dado que los distintos fabricantes de controladoras RAID usan diferentes formatos de metadatos (incluso controladoras de un mismo fabricante son incompatibles si corresponden a series diferentes) es virtualmente imposible mover un conjunto RAID a una controladora diferente, por lo que suele ser necesario mover también la controladora. Esto resulta imposible en aquellos sistemas donde está integrada en la placa base. Esta limitación puede obviarse con el uso de RAID por software, que a su vez añaden otras diferentes (especialmente relacionadas con el rendimiento).

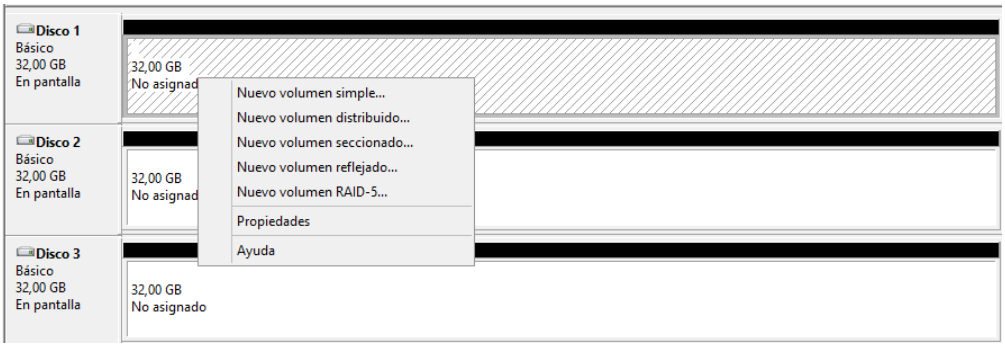
9. Administración de discos.

En los sistemas Windows servidor, se emplea desde el administrador de discos, en Windows clientes, (como Windows10), se utiliza una herramienta llamada espacio de almacenamiento. En ambos se puede usar la herramienta diskpart y desde Linux para trabajar con sistemas raid se utiliza la herramienta mdadm.

En Windows existen 2 formas de gestionar los discos, el modo gráfico y desde la consola de administración de discos.

Los tipos de volúmenes son los siguientes:

- Volumen simple: Un volumen simple se compone por espacio libre de un único disco dinámico (disco normal).
- Volumen distribuido: Un volumen distribuido se forma a partir de la capacidad de varios discos dinámicos.
- Volumen seccionado (RAID 0): Un volumen seccionado almacena en bandas de dos o más discos dinámicos. Un volumen seccionado proporciona un acceso más rápido a sus datos que un volumen simple y distribuido. Las operaciones de lectura y escritura se realizan en paralelo en los discos que componen la banda.
- Volumen reflejado (RAID 1): Un volumen reflejado o disco espejo duplica sus datos en 2 discos dinámicos. De esta forma si se rompe un volumen se conservan los datos en la otra unidad.
- RAID 5: Necesita un mínimo de 3 discos, y aporta tolerancia a fallos y rendimiento.



Desde modo consola la herramienta es diskpart. Esta es una herramienta muy completa, así para crear un sistema raid 5 los comandos serian:

list disk => Para listar los discos físicos.

select disk 1 => Para seleccionar el disco1

online disk => Para poner el disco en línea

convert dynamic => Crear un disco dinámico

Realizamos lo mismo con el resto de discos

create volume raid disk=1,2,3 => crea el sistema raid 5

list volume => Listamos los volúmenes

assign letter = x => Para asignarle una letra al volumen

exit => Salir de la herramienta diskpart

format x: /q /FS:NTFS => Para formatear el volumen
Desconectamos uno de los discos.

```
DISKPART> select disk 3

El disco 3 es ahora el disco seleccionado.

DISKPART> offline disk

DiskPart desconectó correctamente el disco seleccionado.
```

Para recuperar un sistema donde falla un disco (por ejemplo el disco 3), y colocando otro disco de iguales características, llamado disco 4.

```
DISKPART> list disk

  Núm Disco  Estado      Tamaño   Disp     Din  Gpt
  ---------- ----------  -------  -------  ---  ---
  Disco 0    En pantall    15 GB      0 B
  Disco 1    En pantall    20 GB      0 B   *
  Disco 2    En pantall    20 GB      0 B   *
* Disco 3    Sin conexi    20 GB      0 B   *
  Disco 4    Sin conexi    20 GB    20 GB
  Disco M0   Falta          0 B      0 B   *
```

Ahora al disco nuevo (disco 4), lo convertimos a dinámico y borramos sus atributos ya que nos aparecerá como solo lectura y no nos dejaría convertirlo en dinámico:

```
DISKPART> select disk 4

El disco 4 es ahora el disco seleccionado.

DISKPART> attribute disk clear readonly

Atributos de disco borrados correctamente.

DISKPART> convert dynamic

DiskPart convirtió correctamente el disco seleccionado en el formato dinámico.

DISKPART>
```

Nos queda reparar, para ello seleccionamos el volumen que queremos añadir nuestro disco

list volume => listamos los volumenes

select volume 0 => seleccionamos el volumen sobre el que queremos trabajar

repair disk=4 => Añadimos el disco 4 para reparar el sistema

Ahora eliminamos el disco que a fallado:

```
DISKPART> select disk M0

El disco M0 es ahora el disco seleccionado.

DISKPART> delete disk

DiskPart eliminó correctamente el disco que faltaba.
```

Y ya tendríamos recuperado nuestro RAID 5.

En Linux la herramienta se llama mdadm, que se puede instalar en cualquier momento, pero para disco distinto al de sistema. Si se desea utilizar con particiones de sistema, es necesario instalar la versión server.

Para crear un sistema raid 0 se puede utilizar el siguiente formato

```
# mdadm --create --verbose /dev/md0 --level=0 --raid-devices=2  /dev/sdb1 /dev/sdc1
```

En este ejemplo se crea un RAID-5 con cinco discos y un tamaño de chunk de 128KB:

```
# mdadm -Cv /dev/md0 -l5 -n5 -c128 /dev/sd{a,b,c,d,e}1
```