

**TEMA 56: SISTEMAS DE
ALMACENAMIENTO PARA SISTEMAS
GRANDES Y DEPARTAMENTALES.
DISPOSITIVOS PARA TRATAMIENTO DE
INFORMACIÓN MULTIMEDIA.
VIRTUALIZACIÓN DEL ALMACENAMIENTO.
COPIAS DE SEGURIDAD**

Actualizado a 15/04/2023

1 SOPORTES O MEDIOS DE ALMACENAMIENTO DE DATOS

1.1 DISCOS

1.1.1 DISCOS MAGNÉTICOS (HDD: HARD DISK DRIVE)

Dispositivos de almacenamiento de datos basados en un sistema de grabación magnética. Se caracterizan por su capacidad de almacenamiento (TB), velocidad de rotación (rpm), memoria caché (MB), tasa de transferencia (MB/s) y tiempo medio de acceso (< 10 ms).

Componentes: un *eje* central soporta y permite el giro de los **platos**, los cuales se dividen en **pistas**, y estas a su vez en **sectores** (mínima unidad direccionable), cuyo contenido es leído/escrito por las **cabezas**. La **controladora** gestiona el movimiento de platos y cabezas, la memoria **caché** (que acelera el rendimiento) y el acceso a los datos. La comunicación con el exterior se realiza a través de la **interfaz**.

Interfaz	Bus	Última especificación, velocidad, año
ATA (Advanced Technology Attachment)	Paralelo	ATA8 o Ultra ATA/166 → 166 MB/s (2008)
SCSI (Small Computer System Interface)	Paralelo*	Ultra-640 SCSI → 640 MB/s (2003)
SATA (Serial ATA)	Serie	SATA III → 600 MB/s (2009)
SAS (Serial Attached SCSI)	Serie	SAS-3 → 1.200 MB/s (2013)

*NOTA: SCSI 1, SCSI 2 y SCSI 3.1 (SPI) conectan los dispositivos en paralelo. Mientras que SCSI 3.2 (FireWire), SCSI 3.3 (SSA) y SCSI 3.4 (FC-AL) conectan los dispositivos en serie.

1.1.2 DISCOS DE ESTADO SÓLIDO (SSD: SOLID STATE DISK)

Dispositivos de almacenamiento de datos basados en una memoria no volátil (flash NAND).

- Ventajas: elevada velocidad de lectura/escritura, baja latencia, bajo consumo, menor peso y tamaño.
- Inconvenientes: mayor precio y menor capacidad que discos magnéticos, vida útil limitada.

Pueden utilizar las interfaces descritas de los discos magnéticos, si bien en este caso su rendimiento puede quedar limitado por el propio interfaz. Además pueden utilizar el interfaz **NVMe** (NVM express) sobre el bus **PCIe** (PCI express), diseñado para aprovechar la baja latencia y el paralelismo interno de los dispositivos de almacenamiento de estado sólido y que proporciona un rendimiento mayor.

1.1.3 CABINAS DE DISCOS

Dispositivos de almacenamiento compuestos por un conjunto de discos. Proporcionan rendimiento (mediante caché), integridad de los datos (mediante RAID) y facilitan la gestión de los discos.

Componentes:

- **Front-end:** permite conexión con los servidores mediante un protocolo de acceso por bloques (FC, iSCSI, FCoE) o ficheros (NFS, CIFS...). Caracterizados por: número de puertos y velocidad.

- **Back-end:** ofrece conexión a los distintos tipos de discos internos mediante el protocolo nativo de dichos discos (FC o SAS típicamente).
- **Controladora:** gestiona el funcionamiento de la cabina, conectando front-end y back-end. Proporciona funcionalidades como replicación (RAID), cifrado, deduplicación, compresión, snapshots, copias remotas...

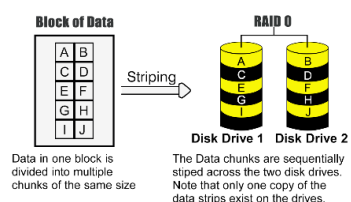
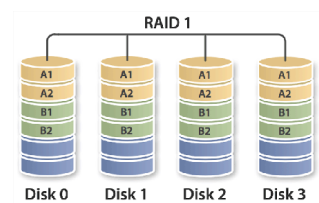
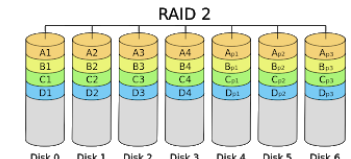
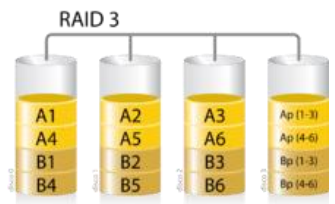
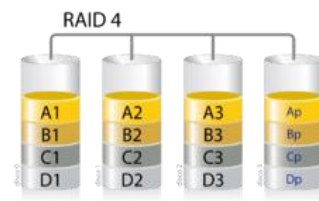
Conceptos:

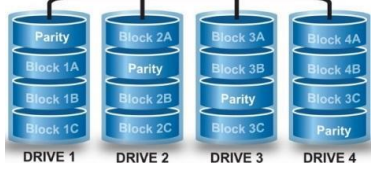
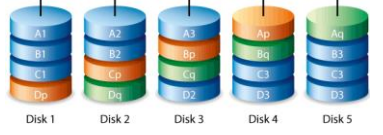
- **Thick provisioning:** los volúmenes se crean con la totalidad del espacio requerido.
- **Thin provisioning:** tecnología empleada para optimizar el espacio mediante el aprovechamiento de los bloques físicos no escritos realmente. Los volúmenes se crean con un tamaño mínimo y crecen según sea necesario.
- **Autotiering:** tecnología sw empleada para optimizar el acceso por niveles a datos almacenados en discos de distinto rendimiento (y coste) según la frecuencia de acceso a los mismos.
- **Deduplicación:** tecnología que optimiza el espacio ocupado al eliminar las copias duplicadas de datos, usando referencias a los datos ya almacenados. Ideal para backups por la habitual reiteración de datos entre copias.

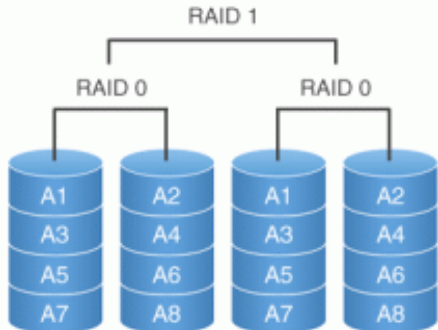
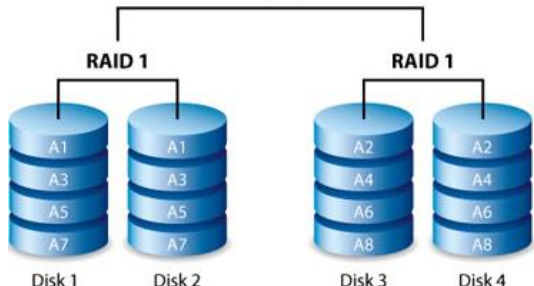
1.1.4 RAID (REDUNDANT ARRAY OF INEXPENSIVE/INDEPENDENT DISKS)

Tecnología basada en la combinación de varios discos duros para formar una única unidad lógica donde se distribuyen o replican los datos. Los niveles RAID utilizan estos mecanismos:

- **Particionado (stripping):** Cada lectura/escritura se divide y las distintas partes son atendidas por discos diferentes. Ventajas: mayor tasa de transferencia y capacidad.
- **Duplicación:** Se replican los datos o se almacena información adicional que permita recuperarlos. Ventajas: mayor integridad y tolerancia a fallos.
 - **Mirroring:** Cada lectura/escritura se envía simultáneamente a varios discos.
 - **Paridad:** Se obtiene una función resumen de los datos a almacenar, permitiendo reconstruir los datos en caso de pérdida. Ventaja: utiliza menos espacio que mirroring.

Niveles simples		
Nivel	Descripción	N.º mínimo discos
RAID 0	<p>Stripping a nivel de bloque. Información dividida en bandas/stripes distribuidas entre los discos, sin redundancia ni paridad.</p> <ul style="list-style-type: none"> Capacidad total del disco virtual, la del disco físico más pequeño x nº discos. Incrementa tasa de transferencia y escritura n veces (n nº discos). No redundancia datos: si falla 1 bloque de 1 stripe de un disco, se pierde todo el stripe. Y si falla 1 disco, se pierde el RAID. Se usa cuando es necesario rendimiento pero no fiabilidad, por ejemplo carpetas de trabajo para edición de video. 	<p>2</p> 
RAID 1	<p>Mirroring a nivel de bloque.</p> <ul style="list-style-type: none"> Capacidad total del disco virtual, la del disco físico más pequeño, aunque lo normal es que se realicen grupos con todos los discos iguales, con lo que sería igual a la capacidad de uno de los 2 discos. Si el grupo de RAID es de más de 2 discos, en cantidad siempre par, la capacidad sería la mitad del total del grupo. Rendimiento: el más lento de los dos. Tolera fallos simples de 1 disco. Si fallan los 2 discos, no. Si hay más de 2 discos, tolera el fallo de la mitad de los discos del grupo de RAID. 	<p>2</p> 
RAID 2	<p>Stripping a nivel de bit con disco de paridad dedicado.</p> <ul style="list-style-type: none"> Todos los discos giran a la misma velocidad. No se usa prácticamente. 	<p>3</p> 
RAID 3	<p>Stripping a nivel de byte con disco de paridad dedicado (Mínimo 3 discos)</p>	<p>3</p> 
RAID 4	<p>Stripping a nivel de bloque con disco de paridad dedicado (Mínimo 3 discos)</p>	<p>3</p> 

RAID 5	<p>Stripping a nivel de bloque con paridad distribuida (1 bloque de paridad por disco).</p> <p>Eficiencia de uso = $1 - 1/n$</p> <p>Capacidad total del disco virtual, si todos de igual C, $C_{RAID5} = C \times (n-1)$</p> <p>Se pierde la capacidad correspondiente a un disco por cada grupo de RAID para paridad (Mínimo 3 discos)</p> <p>Puede fallar un disco sin sufrir pérdida de información</p>	<p>3</p> <p>RAID 5</p> 
RAID 6	<p>Stripping a nivel de bloque con paridad distribuida (2 bloques de paridad por disco).</p> <p>Eficiencia = $1 - 2/n$</p> <p>Capacidad total del disco virtual, si todos de igual C, $C_{RAID6} = C \times (n-2)$</p> <p>Se pierde la capacidad correspondiente a dos discos por cada grupo de RAID para paridad (Mínimo 4 discos)</p> <p>Pueden fallar dos discos sin sufrir pérdida de información</p>	<p>4</p> <p>RAID 6</p> 

Niveles compuestos		
Nivel	Descripción	N.º mínimo discos
RAID 0+1	Mirroring dos arrays <div style="text-align: center;">  </div>	4
RAID 1+0	Striping (RAID 0) de dos arrays RAID 1 <div style="text-align: center;">  </div>	4
RAID 0+3 (o RAID 53)	Striping a nivel de byte (RAID 3) de arrays RAID 0	6
RAID 3+0	Striping a nivel de bloque (RAID 0) de arrays RAID 3	6
RAID 0+5	Striping a nivel de bloque (RAID 5) de arrays RAID 0	6
RAID 5+0	Striping a nivel de bloque (RAID 0) de arrays RAID 5	6
RAID 1+5	Striping a nivel de bloque (RAID 5) de arrays RAID 1	6
RAID 5+1	Mirroring a nivel de bloque (RAID 1) de arrays RAID 5	6

Notas:

- RAID 2, 3, 4 y 5 permiten recuperar la pérdida de 1 disco.
- RAID 6 permite recuperar la pérdida de 2 discos.
- RAID 0+1 tiene la misma tolerancia a fallos que RAID 5.
- RAID 1+0 tiene la misma tolerancia a fallos que RAID 1.
- Se denomina RAID 5E y RAID 6E a las variantes de RAID 5 y RAID 6 que incluyen discos de reserva, los cuales pueden estar conectados y preparados (hot spare) o en espera (standby spare). Permiten minimizar el tiempo de reconstrucción en caso de fallo.

1.2 CINTAS MAGNÉTICAS

Soporte de almacenamiento de datos donde se graba en pistas sobre una banda magnética. Proporcionan gran capacidad a bajo coste, pero con menor rendimiento que los discos y mayor degradación con el uso. El acceso a los datos es secuencial. Usos: respaldo de ficheros.

- Tecnología de grabación:
 - Helicoidal: graban diagonalmente, permitiendo mayor capacidad.
 - Longitudinal: graban en varias pistas. Mayor velocidad y menor capacidad que las de grabación helicoidal.
- Tipos de cintas (datos de últimas versiones comerciales):

TIPO	CAPACIDAD NATIVA	VELOCIDAD
DAT (Digital Audio Tape)	160 GB	24 MB/s
TRAVAN	20 GB	8 MB/s
DLT (Digital Linear Tape)	800 GB	60 MB/s
LTO (Linear Tape-Open) La última generación es LTO-9	18 TB	360 MB/s

Las cintas se ubican en librerías, que actualmente se encuentran automatizadas, pudiendo ser bien cargadores automáticos (con un pequeño número de cartuchos) o bien robots (con miles de ellos).

2 SISTEMAS DE ALMACENAMIENTO DE DATOS

2.1 MODELOS DE CONECTIVIDAD DEL ALMACENAMIENTO

Los modelos más simples se basan bien en el uso de discos internos en los servidores o bien en la conexión directa de una cabina de discos a un servidor (DAS o Direct Attached Storage). Este último proporciona mayor rendimiento y funcionalidades adicionales (integridad, alta disponibilidad, replicación, posibilidad de ampliación...), si bien en ambos casos no se comparte el almacenamiento entre servidores, lo que produce una infrutilización de la capacidad de almacenamiento y no permite la creación de clústeres de servidores.

Los modelos más usados (SAN y NAS) se basan en dos modos de acceso a los datos:

- **Acceso a bloques:** el acceso es a bajo nivel (bloques de datos).
- **Acceso a ficheros:** el acceso es a alto nivel (ficheros de datos).

Diferencias entre SAN y NAS:

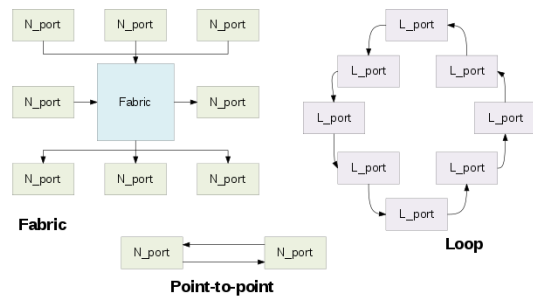
- **SAN (Storage Area Network):**
 - Red independiente de la red de datos y dedicada exclusivamente al almacenamiento, de altas prestaciones, normalmente basada en fibra óptica y con estructura conmutada.
 - Se compone de: dispositivos de almacenamiento, conmutadores y fibra óptica.
 - El dispositivo de almacenamiento (cabina) presenta LUNs (Logical Unit Number) o discos virtuales a los servidores, los cuales gestionan el sistema de ficheros.
 - Opera a nivel de bloques de datos.
- **NAS (Network attached storage):**
 - Dispositivo de almacenamiento conectado a una red de datos (típicamente Ethernet).
 - La NAS crea y gestiona el sistema de ficheros internamente y lo presenta a los servidores a través de una red de datos mediante protocolos como NFS o CIFS.
 - Opera a nivel de ficheros de datos.

2.1.1 CARACTERÍSTICAS DE UNA RED SAN

- **Capas de una red SAN:**
 - Capa cliente: servidores provistos de una tarjeta HBA (Host Bus Adapter), equivalente a las tarjetas NIC (Network Interface Card) de una red Ethernet.
 - Capa de red: cables de fibra óptica, SAN hubs y SAN switches.
 - Capa de dispositivos: cabina de discos y librerías de cintas.

- **Topologías:**

- Conmutada (Switched Fabric)
- Bucle arbitrado (Arbitrated Loop)
- Punto a punto (Point-to-Point)



- **Tipos de red de almacenamiento:**

- **SAN FC (SAN Fiber Channel):**
 - Se componen de un conjunto de switches que conmutan las tramas FC, permitiendo la conectividad de servidores y dispositivos de almacenamiento.
 - Alto rendimiento (la última especificación, con 128 Gbps, alcanza 25.600 MB/s).
 - WWN (World Wide Name): dirección de 64 bits asignada unívocamente a cada dispositivo.
 - FCID: identificador de 24 bits asociado a cada dispositivo de la red SAN FC. Se usa en lugar del WWN para optimizar la conmutación de tramas.
 - DomainID: identificador de los switches FC.
 - ISL (Inter Switch Links): interconexión entre switches FC.
 - Fabric: interfaz que permite a cualquier nodo asociado conectar con cualquier otro sin necesidad de conocer el esquema de interconexión entre nodos.
- **iSCSI (Internet SCSI):**
 - Se basa en el intercambio de tramas sobre TCP/IP en lugar de sobre FC.
 - Menor coste económico que una SAN FC pero menor rendimiento a igual ancho de banda.
- **FCoE (Fiber Channel over Ethernet):**
 - Permite unificar las redes de datos (Ethernet) y la red de almacenamiento (SAN) en una única red Ethernet.
 - Los servidores requieren tarjetas específicas CNA (Converged Network Adapter) que combinan las funcionalidades de una NIC y una HBA.
- **iFCP (Internet Fiber Channel Protocol):** permite tráfico Fiber Channel sobre la red TCP/IP. Actuando como un gateway, permite conectar sistemas Fiber Channel con redes IP.
- **FCIP (Fiber Channel over IP):** interconecta sistemas de SAN FC sobre redes IP. Permite la transmisión de tramas Fiber Channel (FC) a través de túneles IP.
- SAN para mainframe IBM: ESCON y FICON. Tecnología propietaria.

- **Seguridad en una SAN:**

- **Zoning:** segmentación de una red SAN en zonas en base al WWN de los dispositivos.
- **LUN Masking:** presentación de una LUN únicamente a los nodos autorizados.

- **Port Binding:** limitación de la conectividad de un nodo a un puerto del switch.
- **Mecanismos de replicación:**
 - **Copia local:** replicación de los datos en el mismo dispositivo de almacenamiento. Protege frente a errores humanos o de software, pero no errores hardware.
 - **Copia remota:** replicación de los datos en dos o más CPD de forma simultánea. Protege frente a errores hardware, garantizando la alta disponibilidad.
 - **Copia síncrona:** los datos escritos se replican en el sistema remoto y sólo cuando se ha escrito en ambos se devuelve al servidor la confirmación. Garantiza que los datos remotos son en todo momento copia de los locales.
 - **Copia asíncrona:** el servidor recibe la confirmación de escritura cuando se ha realizado en el sistema local y posteriormente se envían los datos al centro remoto, por lo que existe un desfase entre las copias.

2.2 ARQUITECTURAS DE BACKUP

A diferencia de los mecanismos de replicación en tiempo real, los sistemas de backup realizan copias de los datos con una periodicidad definida, y permiten recuperar una versión de los datos en un momento anterior en el tiempo.

Tipos de backup:

- **Completo:** copia de todos los archivos y directorios seleccionados. Se borra el bit de modificado de cada archivo, lo que permite que futuros backups solo copien los archivos modificados.
- **Incremental:** se examina el bit de modificado, copiando solo los archivos que han cambiado desde el último backup incremental o completo, y se borra este bit en los archivos respaldados.
- **Diferencial:** como el incremental, salvo que no se elimina el bit de modificación. Requiere más espacio y tiempo que el incremental pero las restauraciones solo requieren la copia completa y la diferencial más reciente.
- **Intermedio:** como el completo, salvo que no se borra el bit de modificado.

Un esquema común de respaldo es el método abuelo-padre-hijo, que utiliza 3 generaciones de cintas (en total $12+4+7=23$ cintas):

- Se realiza un backup completo cada mes y se guarda durante un año (abuelo).
- Se realiza un backup completo cada semana y se guarda durante un mes (padre).
- Se realiza un backup diario (completo, incremental o diferencial) y se guarda durante una semana (hijo).

3 VIRTUALIZACIÓN DEL ALMACENAMIENTO Y BACKUP

3.1 VIRTUALIZACIÓN DEL ALMACENAMIENTO

La virtualización de almacenamiento busca presentar el almacenamiento como una vista lógica a uno o varios hosts, abstrayendo los detalles de los dispositivos físicos de almacenamiento. La capa de virtualización muestra a los hosts un espacio de almacenamiento uniforme, agregando y gestionando el espacio de múltiples dispositivos de almacenamiento potencialmente de distintos fabricantes y tipos.

Ventajas:

- Permite migraciones de datos entre dispositivos de forma transparente.
- Centraliza la gestión del espacio de diferentes dispositivos en una herramienta única.
- Permite mejorar la escalabilidad del sistema de almacenamiento al no depender de un dispositivo en concreto, permitiendo habitualmente el crecimiento o disminución de espacio asignado en caliente.
- Puede añadir características o inteligencia a los dispositivos existentes, como sistemas de replicación, caché, tiering, etc.

La virtualización puede implementarse:

- A nivel de host: El sistema operativo virtualiza el almacenamiento, agrupando todos los recursos disponibles y permitiendo administrarlo de forma uniforme. Un caso de uso habitual es el de los hipervisores de virtualización, que pueden agrupar el almacenamiento de varios orígenes y presentarlo particionado a las máquinas virtuales.
- A nivel de red: Más común en sistemas SAN, la inteligencia reside en un dispositivo de la red, que agrupa los accesos de los dispositivos finales de almacenamiento y los presenta de forma unificada a los hosts.
- Basada en cabinas: Posibilidad en algunos modelos de cabinas de virtualizar los recursos físicos de que dispone. Por ejemplo, algunos modelos disponen de discos de distintos tipos o clases de rendimiento, realizando tiering de los datos en función de su frecuencia de utilización. De esta forma, se ubican los datos más accedidos en los discos de más rendimiento, mientras que los datos no usados se almacenan en los dispositivos más lentos y baratos. Los sistemas RAID pueden considerarse un caso de este tipo de virtualización, si bien muy simple.

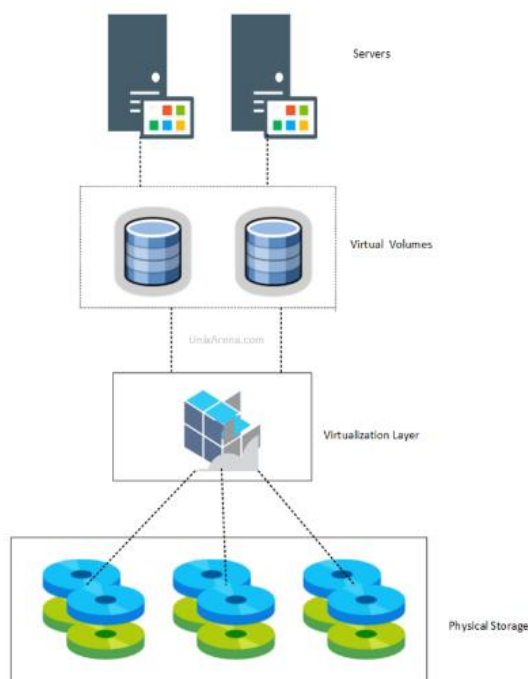


Figura 1: obtenido de <https://www.unixarena.com/2019/08/virtualization-hypervisor-basic-interview-questions.html/>

3.2 VIRTUALIZACIÓN DE BACKUP

Los sistemas **VTL** (Virtual Tape Library) permiten presentar un dispositivo de almacenamiento (normalmente una cabina de discos) como si fuera una librería de cintas para su uso con un software de backup preexistente. Proporcionan mayor rendimiento tanto en backup como en la recuperación. Comúnmente se implementa mediante un "appliance".

4 NUEVAS TENDENCIAS EN ALMACENAMIENTO

Polarización en el almacenamiento, ubicando el almacenamiento primario en configuraciones de alto rendimiento y el almacenamiento secundario en soluciones de menor coste. Actualmente, los sistemas híbridos (basados en el uso conjunto de discos magnéticos y tecnología flash para cachear la información) son una buena opción para lograr un alto rendimiento a la vez que mantienen un coste por GB óptimo.

Acercamiento del almacenamiento al cómputo y uso de hardware de propósito general. Este concepto se conoce también como **Server SAN**, basado en el uso de dispositivos de almacenamiento DAS (Direct Attached Storage) conectados directamente a los servidores, que permite combinar almacenamiento y cómputo en el mismo hardware.

Software Defined Storage (SDS): arquitectura de almacenamiento que separa el almacenamiento subyacente mediante una capa de datos virtual, permitiendo una mayor flexibilidad y menores costes. Forma parte de la tendencia que incluye las redes definidas por software (SDN) y los centros de datos definidos por software.