

Tema 4

Tecnologías de almacenamiento

Firma

Índice

- **Introducción**
- Organización del almacenamiento en el computador
- Discos duros y dispositivos SSD
- RAID

Almacenamiento permanente

- Concepto

Capacidad de los sistemas informáticos de almacenar información durante un tiempo indefinido, independientemente de que dichos sistemas se encuentren o no conectados a su fuente de alimentación habitual.

- Dispositivos y tecnologías de almacenamiento permanente

Disco duro



Tecnología de almacenamiento

Magnética

CD/DVD



Tecnología de almacenamiento

Óptica

Unidad de estado sólido



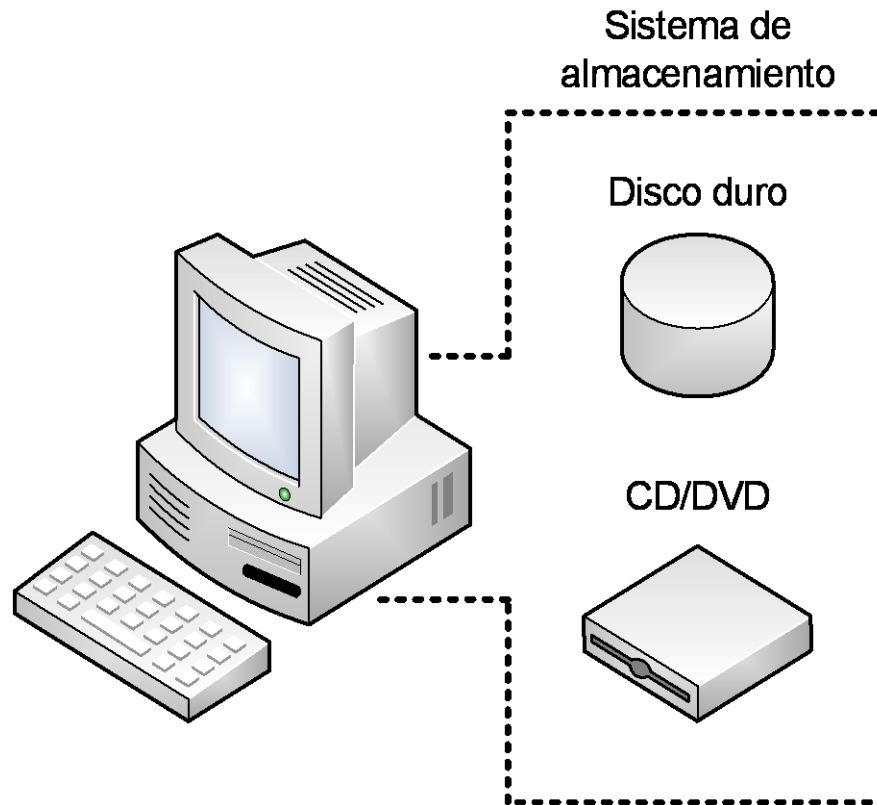
Tecnología de almacenamiento

Memoria flash

Almacenamiento en el ordenador de escritorio

- Característica básica

Simplicidad: el sistema de almacenamiento está habitualmente formado por un disco duro y una unidad óptica.



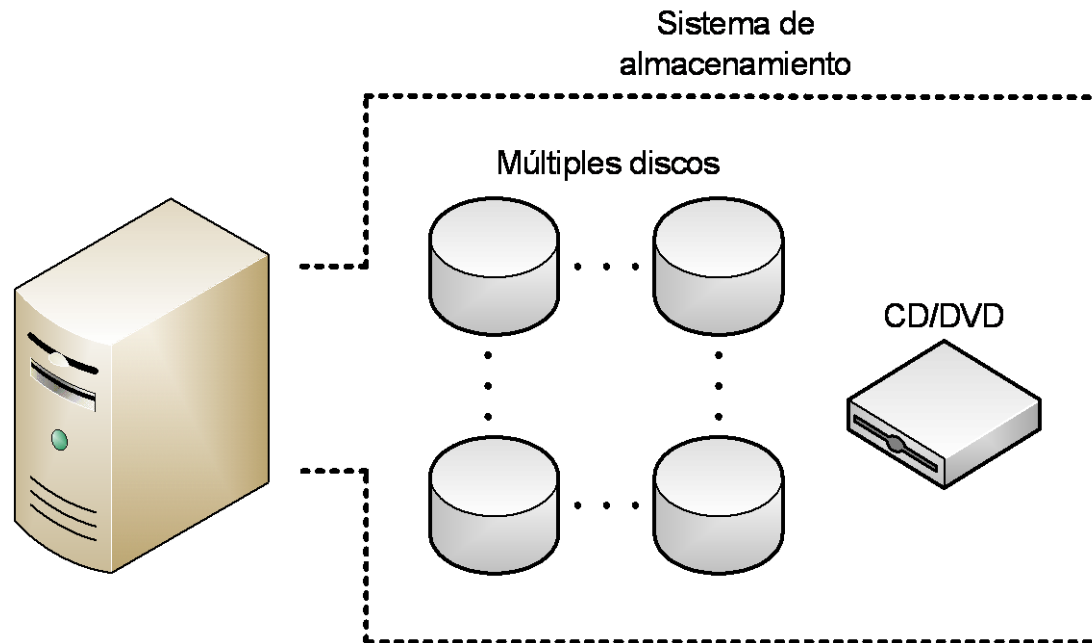
Almacenamiento en el servidor

- Requisitos habitualmente requeridos

- Elevada capacidad de almacenamiento.
- Elevadas prestaciones: bajo tiempo de acceso y elevada tasa de transferencia.
- Tolerancia a fallos y alta disponibilidad.

- Diseño para el cumplimiento de los requisitos

Basado en múltiples discos habitualmente gestionados mediante controladoras RAID.

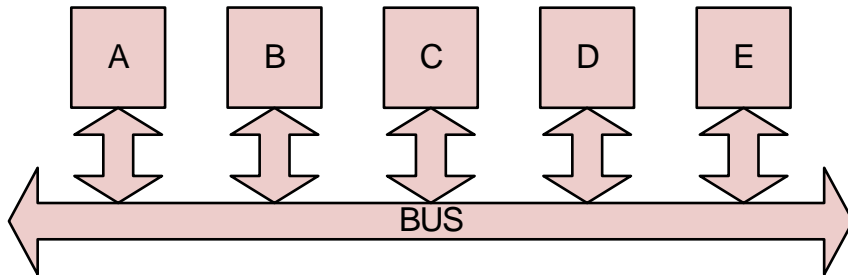


Índice

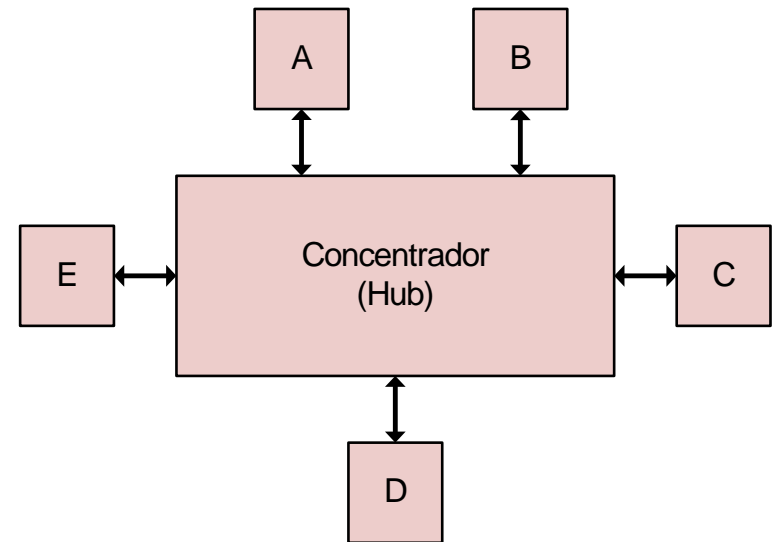
- Introducción
- **Organización del almacenamiento en el computador**
- Discos duros y dispositivos SSD
- RAID

Repaso: Interconexión de dispositivos en el ordenador

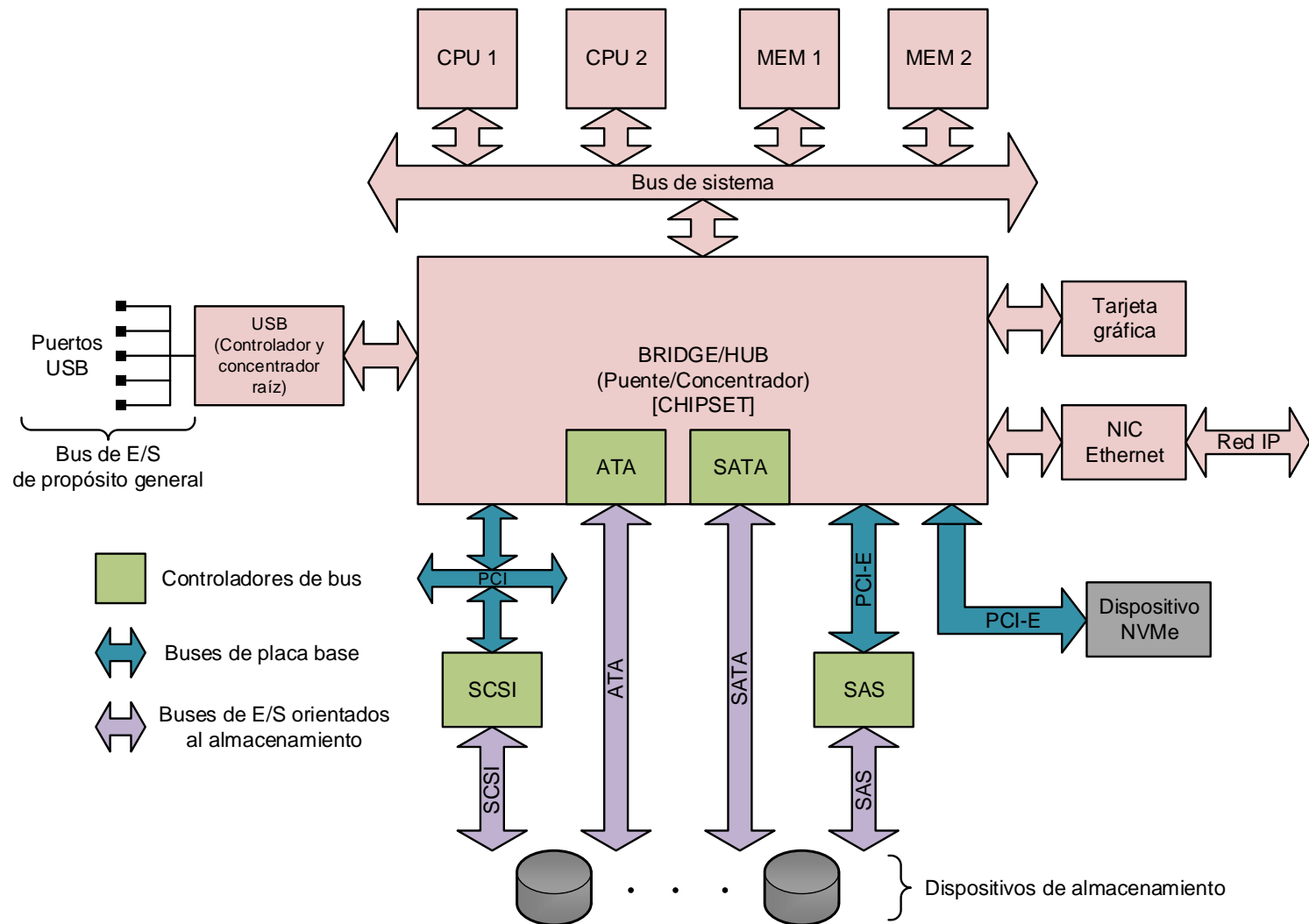
Interconexión en bus



Interconexión mediante concentrador



Jerarquía de conexiones de un ordenador moderno



Elementos de la jerarquía de conexiones de un ordenador moderno

- Bus de sistema

Su objetivo es conectar las CPU y la memoria RAM del ordenador. Se trata de un bus especialmente optimizado para proporcionar un elevado ancho de banda de comunicación entre estos dispositivos.

- Subsistema de concentradores (chipset)

Su objetivo es establecer una vía de comunicación entre los dispositivos conectados al bus de sistema (CPU y memoria) y el resto de dispositivos del computador.

- Bus de placa base

Su objetivo es proporcionar un mecanismo de integración de interfaces en el computador mediante el mecanismo de las ranuras de expansión. Los controladores de buses de E/S pueden conectarse a este bus.

- Controladores de bus

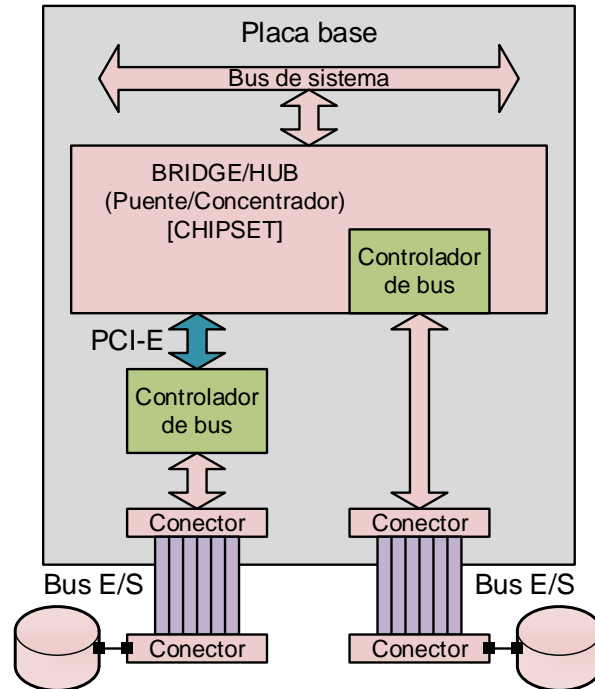
El objetivo de estos dispositivos es controlar buses de E/S. Para ello, gestionan el direccionamiento de los dispositivos en el bus y proporcionan un puente de comunicación entre el bus y otros dispositivos del sistema de interconexiones del computador.

- Buses de E/S

El objetivo de estos buses es proporcionar un mecanismo de conexión de periféricos al computador.

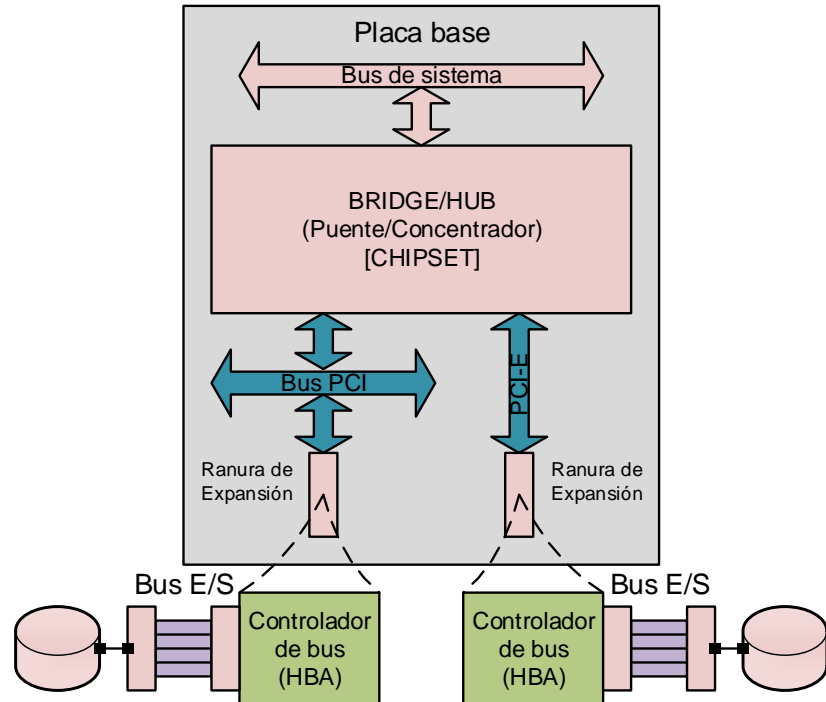
Alternativas de implementación de los controladores de bus

Controlador de bus en placa base



La electrónica del controlador está integrada en la placa base: esta debe proporcionar un conector para conectar el cable que actuará como bus de E/S.

Controlador de bus en HBA



La electrónica del controlador se implementa mediante un HBA (Host Bus Adapter), que se integra en la placa base mediante una ranura de expansión perteneciente a un bus de placa base. El HBA proporciona un conector para conectar el cable que actuará de bus E/S.

Buses de E/S

- El beneficio de la estandarización

Posibilitar la integración entre periféricos y ordenadores de múltiples fabricantes.

- Definición y mantenimiento de los estándares

Los estándares son gestionados, o bien por consorcios de fabricantes del sector, o bien por comités pertenecientes a institutos de estandarización, como el INCIS.

- Aspectos definidos por un estándar de bus

- Características físicas

Define aspectos tales como formatos de cables y conectores, especificaciones eléctricas, frecuencias de trabajo, etc.

- Arquitectura lógica

Define fundamentalmente las líneas que forman el bus así como su funcionalidad.

- Interfaz de programación

Define el conjunto de comandos disponibles para el manejo del bus.

PREGUNTA TÍPICA

Buses de E/S estándares

- Buses de E/S de propósito general

- USB (Universal Serial Bus)

- ✓ Objetivo

Proporcionar un mecanismo simple para la conexión de periféricos externos a un ordenador.

- ✓ Características más destacadas

- Plug & Play.
 - Capacidad de alimentar a los dispositivos conectados.

- Buses de E/S orientados al almacenamiento

- ATA (Advanced Technology Attachment)

- SCSI (Small Computer System Interface)

- SATA (Serial ATA)

- SAS (Serial Attached SCSI)

ATA: Aspectos generales

- Nomenclatura alternativa

IDE (Integrated Drive Electronics)

- Arquitectura

- Bus paralelo de ancho 16
- Capacidad de direccionamiento de un controlador: dos dispositivos de almacenamiento.

- Ámbito de uso

Ordenadores personales

- Evoluciones del estándar

7 evoluciones: ATA1 - ATA7

- Prestaciones

Hasta 133MB/s en ATA7.

ATA: Implementación física

- Controlador

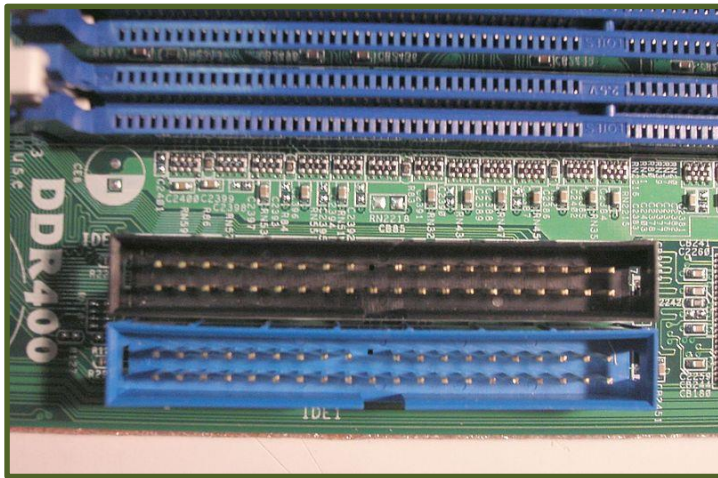
En los últimos años, los controladores ATA se han implementado de forma estándar en placa base, formando parte del chipset.

- Bus

Se implementa mediante un cable de 40 u 80 conductores. En el cable de 80, solo 40 transportan señales, los otros 40 son de aislamiento.

- Conectores

Cada controlador ATA tiene asignado un conector ATA en la placa base, que se utiliza para conectar el cable al controlador.



Conectores ATA



Cable ATA

SCSI: Aspectos generales

- Arquitectura

- Bus paralelo diseñado con dos tipos de ancho: 8/16
- Capacidad de almacenamiento: 8 dispositivos para ancho 8, 16 para ancho 16.

- Ámbito de uso

Servidores y sistemas de almacenamiento externos (uds. de cinta)

- Evoluciones del estándar

9 evoluciones, desde la SCSI-1 hasta la ultra-320 SCSI

- Prestaciones

320MB/s en la ultra-320 SCSI

SCSI: Implementación física

- Controlador

Se implementa con mayor frecuencia mediante HBA

Conector para cable interno

Conector para cable externo



HBA SCSI

- Bus

Se implementa mediante un cable de 68 conductores para los buses de ancho 16



Cable para conexión de dispositivos externos



Cable para conexión de dispositivos internos

El cambio hacia el bus serie

- Prestaciones (ancho de banda)
 - Factores determinantes en las prestaciones de un bus
 - ✓ Acho (tamaño de los datos transmitidos)
 - ✓ Frecuencia de trabajo
 - Comparativa en prestaciones bus paralelo / bus serie respecto a ancho y frecuencia
 - ✓ Ancho: Bus paralelo ventajoso
 - ✓ Frecuencia: Bus paralelo desventajoso

- Existen un conjunto de problemas de diseño electrónico en los buses paralelos que limitan su frecuencia de trabajo, agravándose dicha limitación en la medida que aumenta la longitud del bus.
- La longitud requerida para un bus de E/S es significativa.

✓ Resultado combinado (Acho + frecuencia): Bus serie ventajoso

- Espacio ocupado (*footprint*) por los elementos del bus

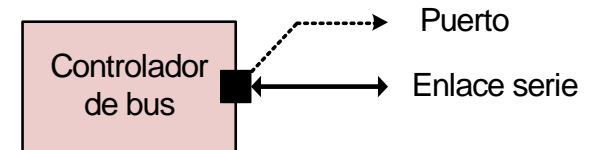
Los cables serie son más estrechos ocupando menos espacio. Asimismo, debido a que tienen muchos menos conductores, sus conectores tendrán muchos menos pines, resultando más pequeños.

- Espacio ocupado: Bus serie ventajoso

Interconexión de dispositivos mediante bus serie

- Concepto de puerto

Punto de conexión proporcionado por un controlador de bus para conectar un enlace serie.



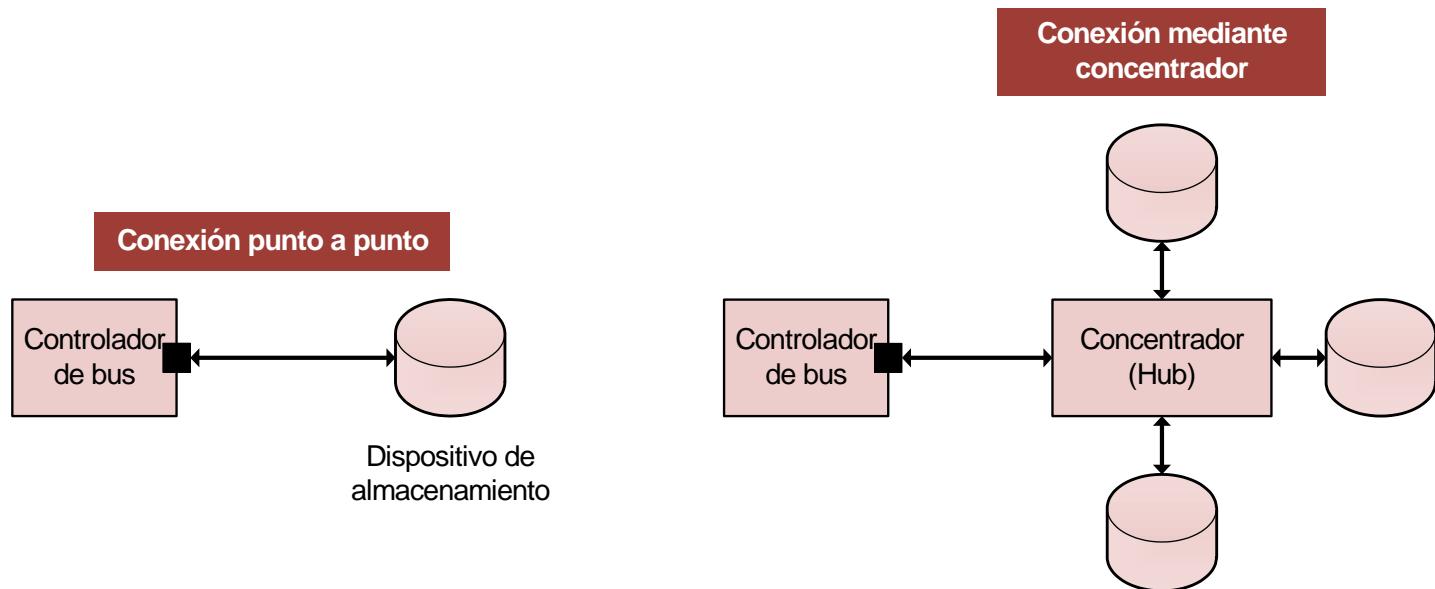
- Modelos de conexión

- Punto a punto

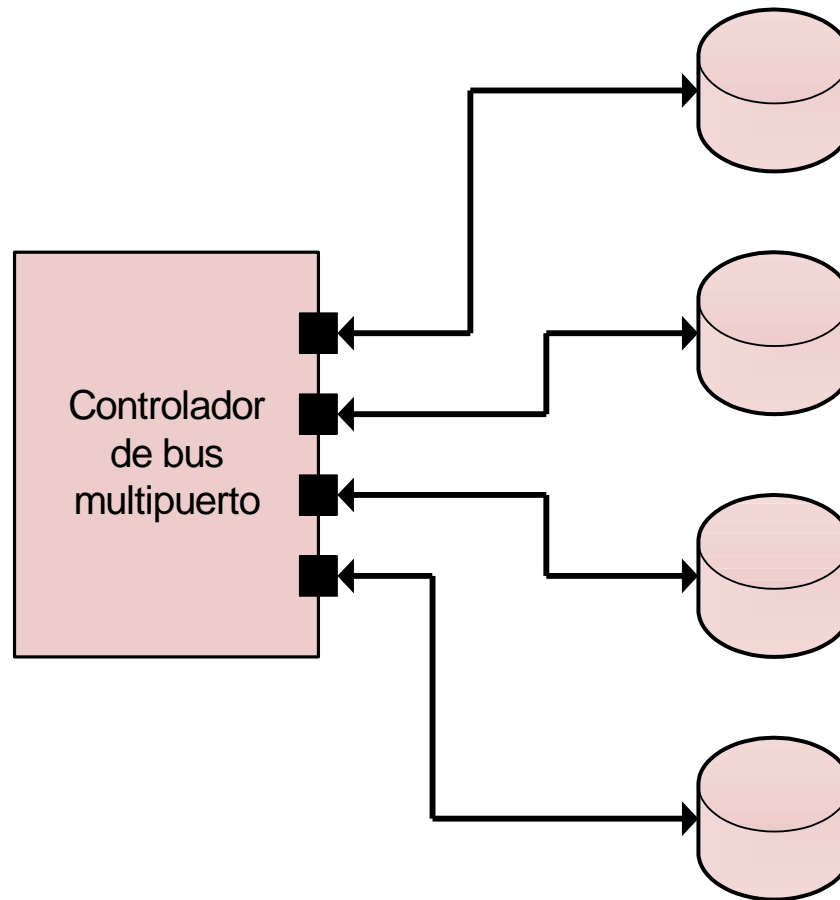
Un puerto, un dispositivo

- Mediante concentrador

Múltiples dispositivos se conectan a un puerto mediante un concentrador



Controladores multipuerto



Proporcionar un mecanismo de conexión al computador para múltiples dispositivos de almacenamiento, utilizando conexiones punto a punto.

SATA: Aspectos generales

- Nomenclatura

SATA significa Serial ATA.

- Compatibilidad con el estándar ATA

- 100% compatible en la interfaz de programación.
- No compatible a nivel físico (cables y conectores)

- Arquitectura

- Canal de comunicación

Serie bidireccional (dos canales de transmisión)

- Controlador de bus

Habitualmente multipuerto.

- Capacidad de direccionamiento

Cada puerto puede direccionar hasta quince dispositivos utilizando un concentrador. Habitualmente, un puerto se conectará a un único dispositivo.

- Ámbito de uso

- Ordenadores personales
- Servidores de gama baja
- Conexión de dispositivos ópticos en todo tipo de equipos.

SATA: Estándar

- Evolución del estándar y sus prestaciones

Revisión	Nombre comercial	Frecuencia (GHz)	Ancho de banda (MB/s)
1.0	SATA 1.5 Gbit/s	1,5	150
2.0	SATA 3 Gbit/s	3	300
3.0	SATA 6 Gbit/s	6	600

- Cálculo del ancho de banda (SATA 1.5 Gbit/s)

$$AB(10) = 1.5 \times 10 \text{ ciclos/seg} \cdot \text{bit/ciclo} \cdot \text{byte}/10\text{bits} = 1.5 \times 10 \text{ bytes/s} = 150\text{MB/s}$$

Nota: Los bytes se codifican con 10 bits con objeto de incluir información de sincronización

- Comparación de prestaciones ATA/SATA
 - ATA-7: 133 MB/s
 - SATA 6 Gbit/s: 600 MB/s

SATA: Implementación física

- Controlador

Se implementan de forma estándar en placa base, formando parte del chipset.

- Enlace serie

Se implementa mediante un cable de 7 conductores: 4 para los dos canales de transmisión + 3 masas

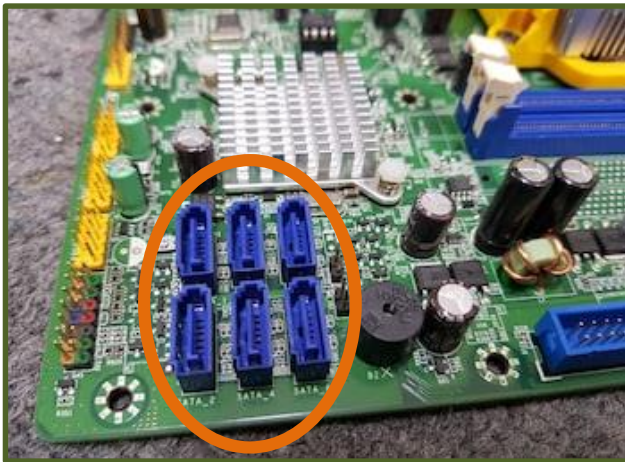
- Cable SATA



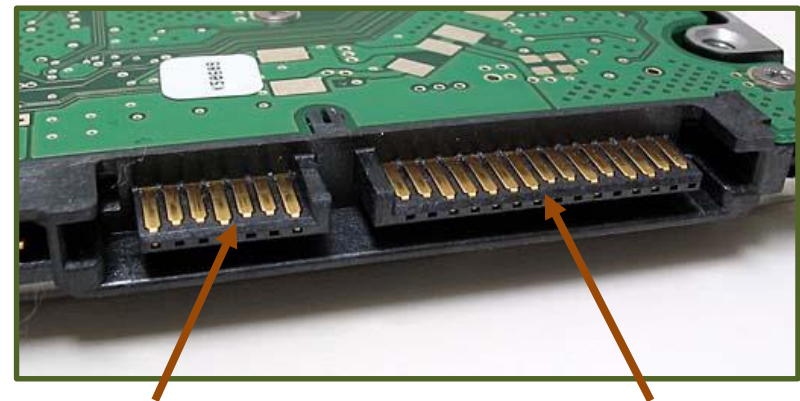
El cable SATA ocupa mucho menos espacio que el cable ATA

- Conectores

Conectores en la placa base



Conectores en el disco



Conector de datos

Conector de alimentación

SAS: Aspectos generales

- SAS significa Serial Attached SCSI
- Compatibilidad con el estándar SCSI
 - 100% en la interfaz de programación
 - No compatible a nivel físico (cables y conectores)
- Ámbito de uso

- Servidores de gama media y alta.
- Cabinas de almacenamiento.

SAS: Arquitectura (I)

- Canal de comunicación

Serie bidireccional idéntico al canal SATA.

- Concepto ampliado de puerto

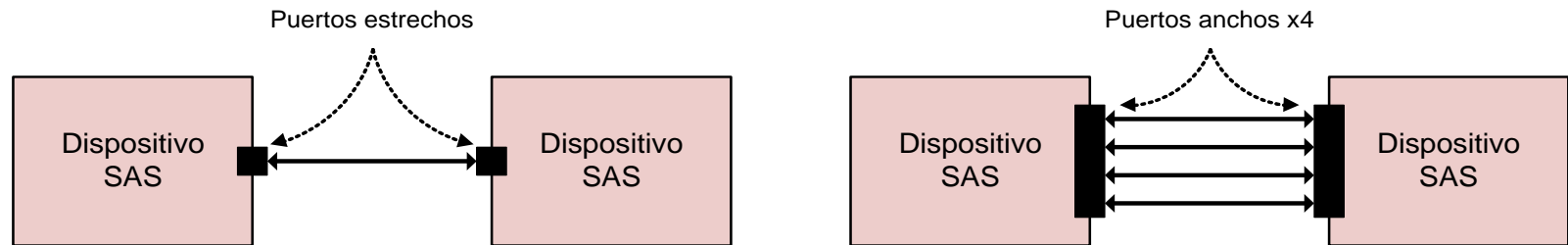
En SAS, los canales serie pueden trabajar en solitario o en agregación. Un puerto es un punto de conexión para un canal o para una agregación de canales.

- Puerto estrecho (*narrow port*)

Incluye un solo canal.

- Puerto ancho (*wide port*)

Incluye dos, cuatro u ocho canales.



- Interconexión de dispositivos SAS

- Punto a punto

- A través de concentradores

- Los concentradores SAS se denominan “expanders”
 - Se pueden interconectar hasta dos niveles de “expanders”

SAS: Arquitectura (II)

- Direccionamiento de dispositivos

- Dominio SAS

Conjunto de dispositivos gestionados por un puerto, o bien directamente, o mediante concentradores.

- Roles de dispositivo en un dominio

- ✓ *Initiator* (habitualmente, el puerto de un controlador)

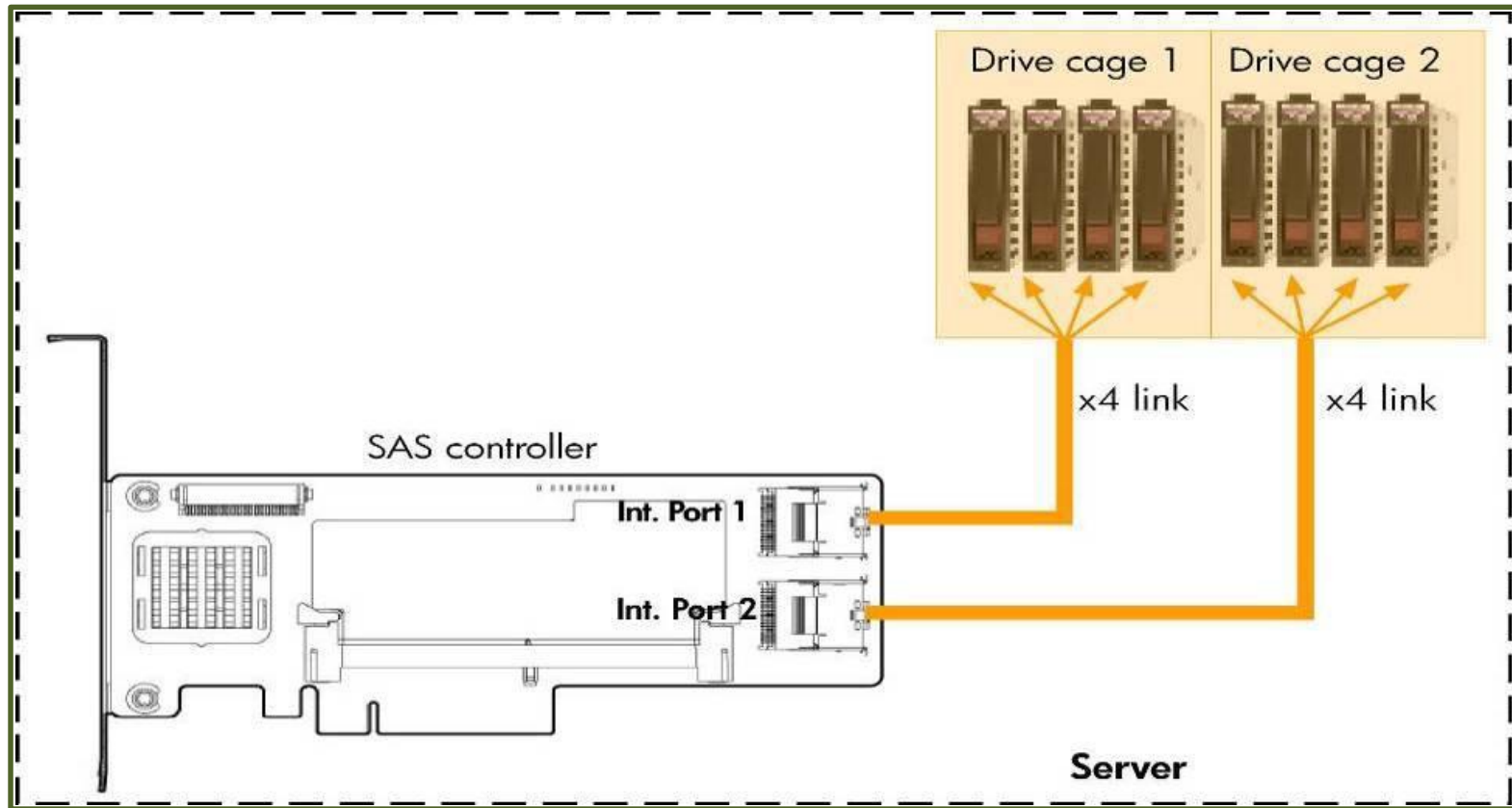
- ✓ *Target* (habitualmente, un dispositivo de almacenamiento)

- Capacidad de direccionamiento

- Hasta ocho dispositivos sin el uso de concentradores (puerto x8)
 - Hasta 128 dispositivos en un único concentrador.
 - Hasta 128x128 dispositivos utilizando dos niveles de concentradores.

SAS: Ejemplo de topología de interconexión (I)

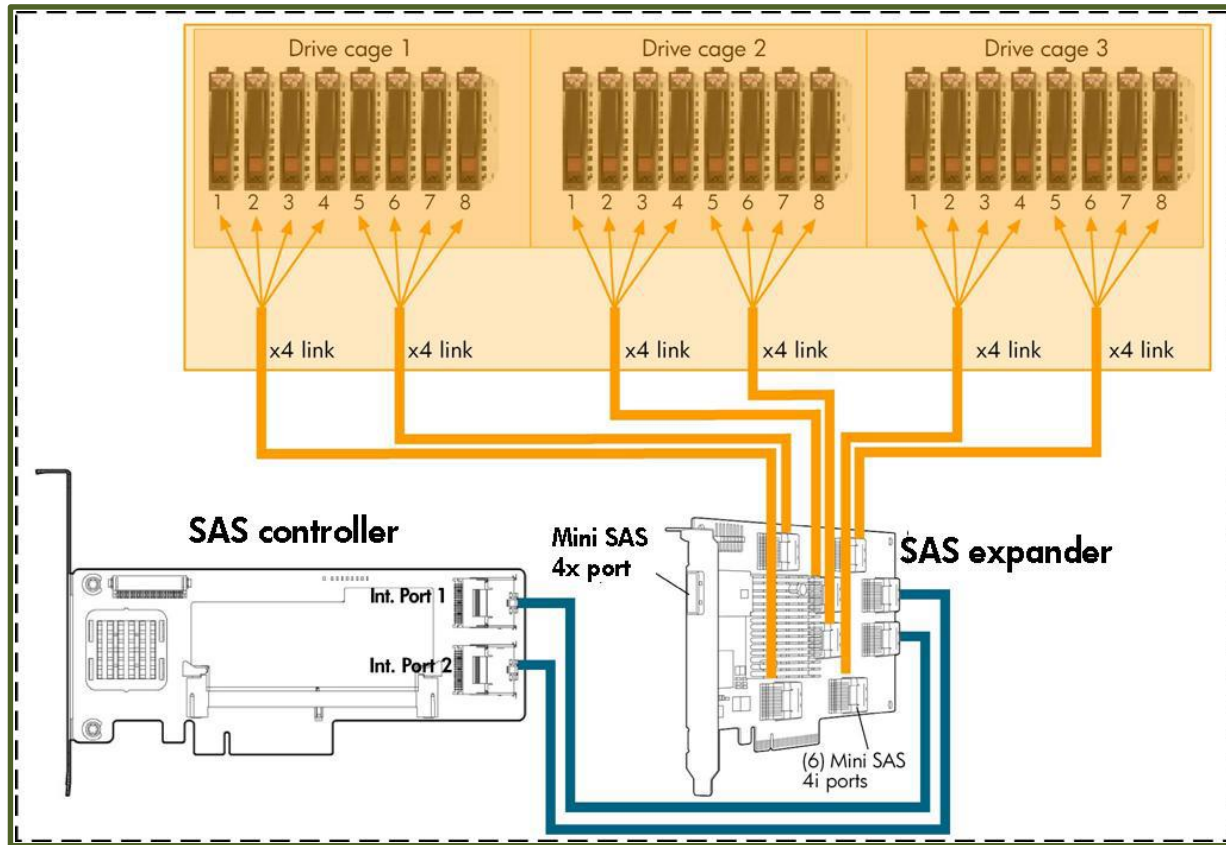
Topología basada en conexiones punto a punto



- Topología SAS típica para la organización del almacenamiento interno de un servidor
- Dos dominios SAS con cinco dispositivos cada uno (un initiator y cuatro targets)

SAS: Ejemplo de topología de interconexión (II)

Topología basada en concentradores



- Topología SAS para la organización del almacenamiento interno de un servidor de alta capacidad
- Dos dominios SAS con trece dispositivos cada uno (un initiator, doce targets)

SAS: Estándar

- Evolución del estándar y sus prestaciones

Versión	Frecuencia (GHz)	Ancho de banda (MB/s)
SAS-1	3	300
SAS-2	6	600
SAS-3	12	1.200

- Cálculo del ancho de banda (SAS-2)

$$AB(SAS-2) = 6 \cdot 10 \text{ ciclos/seg} \cdot \text{bit/ciclo} \cdot \text{byte}/10\text{bits} = 6 \cdot 10 \text{ bytes/seg} = 600\text{MBps}$$

Nota: como en el caso de SATA, en SAS los bytes se codifican con 10 bits.

- Ancho de banda de un puerto SAS-3 x4

$$4 \cdot 1200 = 4800\text{MBps}$$

- Comparación de prestaciones SCSI/SAS
 - Ultra-320 SCSI: 320 MB/s
 - SAS-3: 1200 MB/s

SAS: Implementación física (I)

- Controlador

Se implementa tanto en placa base como mediante HBA



HBA SAS

- Enlace serie

- Se implementa mediante un cable de 7 conductores: 4 para los dos canales de transmisión + 3 masas
- La implementación del enlace SAS es idéntica a la del enlace SATA

SAS: Implementación física (II)

- SAS Backplane

- Concepto

Placa de circuito impreso que proporciona un conjunto de conectores SAS.

- Objetivo

Proporcionar una infraestructura de conexión para dispositivos de almacenamiento con capacidad de conexión en caliente.

- Uso

En la inmensa mayoría de los servidores de gama media y alta, los dispositivos de almacenamiento SAS se conectan mediante un SAS backplane.

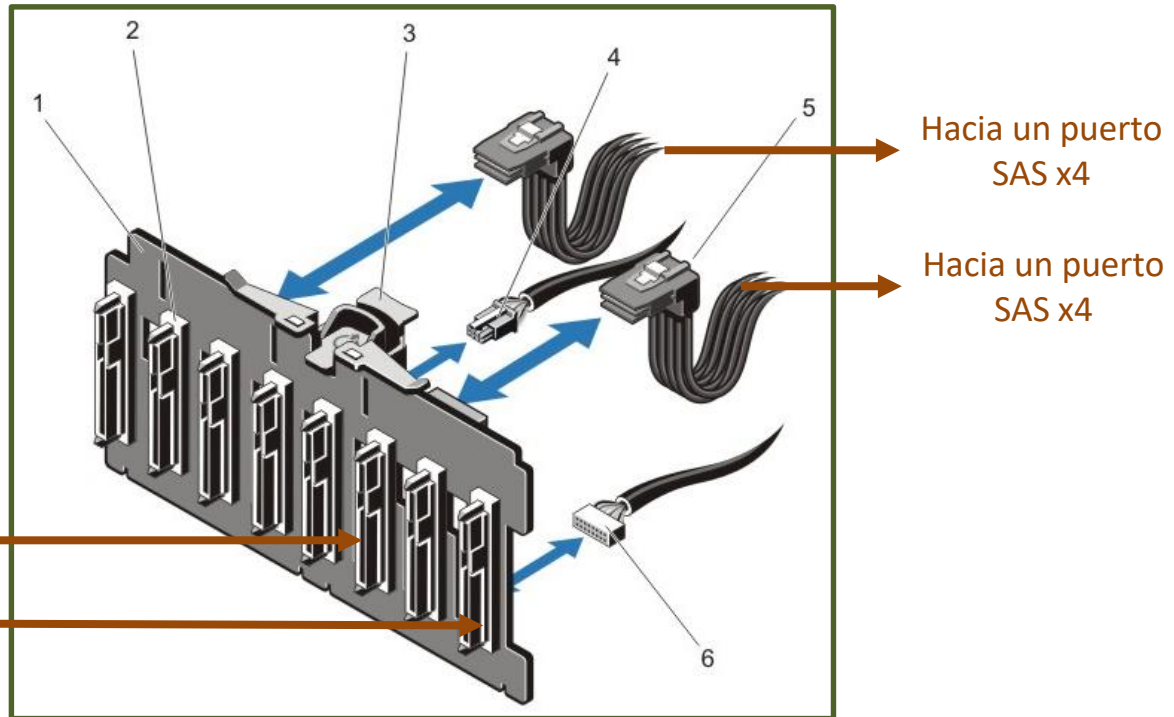
2) Conector SAS para disco

4) Cable de alimentación

5) Cable SAS

Disco SAS

Disco SAS



SAS: Tecnología de discos con doble puerto

- Concepto

Discos que tienen dos puertos SAS de conexión

- Objetivos

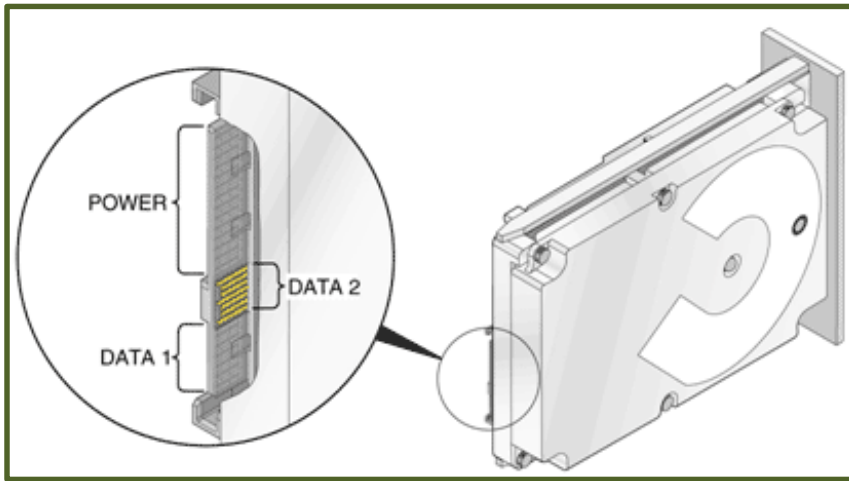
- Permitir la implementación de topologías de conexión de discos con bus redundante

Aplicable tanto a dispositivos HDD como SSD.

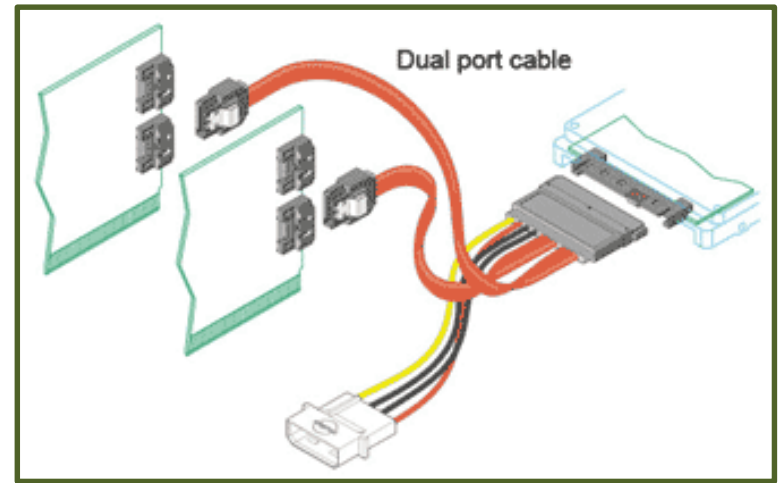
- Duplicar la tasa de transferencia de datos entre el controlador y el disco

Aplicable solo a dispositivos SSD.

Detalle del conector de un disco con doble puerto



Conexión de disco con doble puerto a una topología con bus SAS redundante



SAS: Compatibilidad con SATA

- El bus SAS admite dispositivos SATA

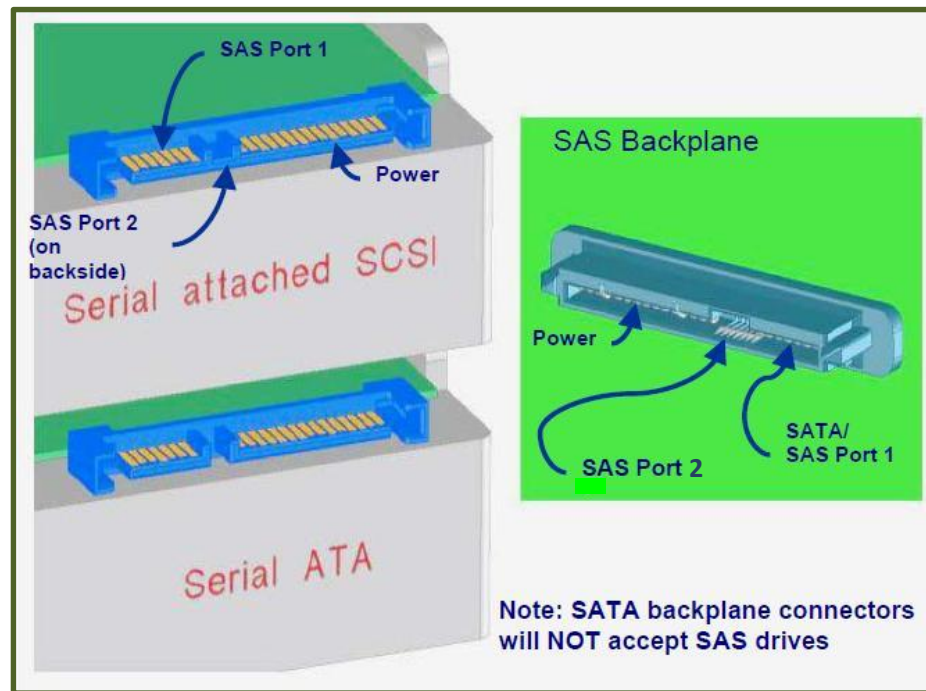
- Implicaciones relativas al controlador

Los controladores SAS implementan el protocolo SATA además del SAS.

- Implicaciones relativas a los conectores

Los conectores SAS deben ser compatibles con los dispositivos SATA.

Comparación conectores SAS/SATA



NVMe: Aspectos generales

- NVMe significa Non-Volatile Memory Express

- Concepto

Especificación de interfaz de dispositivo lógico para el acceso a dispositivos de memoria no volátil a través del bus PCIe.

- Productos de la implementación del estándar NVMe

Drivers de dispositivos

- Rediseño del modelo de driver desde cero

- Objetivo

Explotar al máximo las características de los dispositivos de memoria no volátil.

- ¿Qué hace posible el rediseño desde cero?

Se diseñan solo para dispositivos de memoria no volátil, liberándose de la gestión de discos HDD.

- Sistemas operativos con soporte nativo para NVMe

Todos los Windows modernos (10 y las versiones de servidor desde 2012 R2) y todas las distribuciones Linux con la versión del kernel 3.3 o superior.

- Diferencia esencial de NVMe respecto a SAS y SATA

SAS y SATA especifican buses físicos (basados en enlaces serie), además de una interfaz de dispositivo lógico. Por el contrario, NVMe solo especifica una interfaz de dispositivo lógico, ya que el bus físico utilizado por los dispositivos basados en esta especificación es el PCIe.

- Ámbito de uso

Todo tipos de sistemas, incluyendo ordenadores de escritorio, estaciones de trabajo, portátiles y servidores.

Mejoras de NVMe frente a SAS y SATA

- Manejo de múltiples colas de comandos de E/S de mayor profundidad

- Colas de comandos en SATA

Una, de profundidad 32.

- Colas de comandos en SAS

Una, de profundidad 256.

- Colas de comandos en NVMe

Hasta 64 K, de profundidad 64 K.

- Beneficio

Posibilitar el paralelismo en el acceso al dispositivo de almacenamiento.

- Racionalización y simplificación del código de E/S (usa menos de la mitad de instrucciones)

- Beneficio

Reduce el tiempo de ejecución de las rutinas de acceso a los dispositivos en el SO, disminuyendo la latencia de las operaciones y aumentando la productividad de las mismas.

Factores de forma de dispositivos NVMe y formatos de conexión al bus PCIe

NOTA sobre PCIe: Las interfaces PCIe gestionan enlaces serie (*lanes*). El número de enlaces en una interfaz puede ser x1, x2, x4, x8, x16 y x32

- 1) Tarjeta de expansión PCIe estándar: x4, x8
- 2) Tarjeta M.2: x2, x4
- 3) Dispositivo 2.5" (small form factor): x4 (con. U.2)

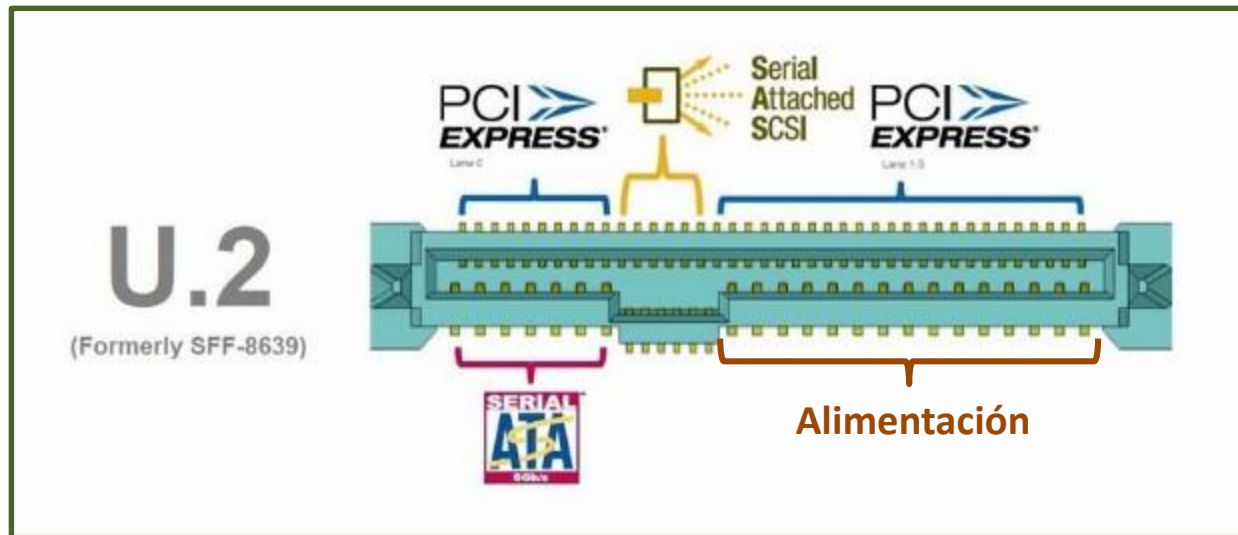
Anchos de banda PCIe 3.0

x1	984.6 MB/s
x2	1.97 GB/s
x4	3.94 GB/s
x8	7.88 GB/s



Conector U.2 para backplane

- Diseñado para la conexión en caliente de dispositivos NVMe
- Conexión PCIe x4
- Compatible con dispositivos SAS y SATA



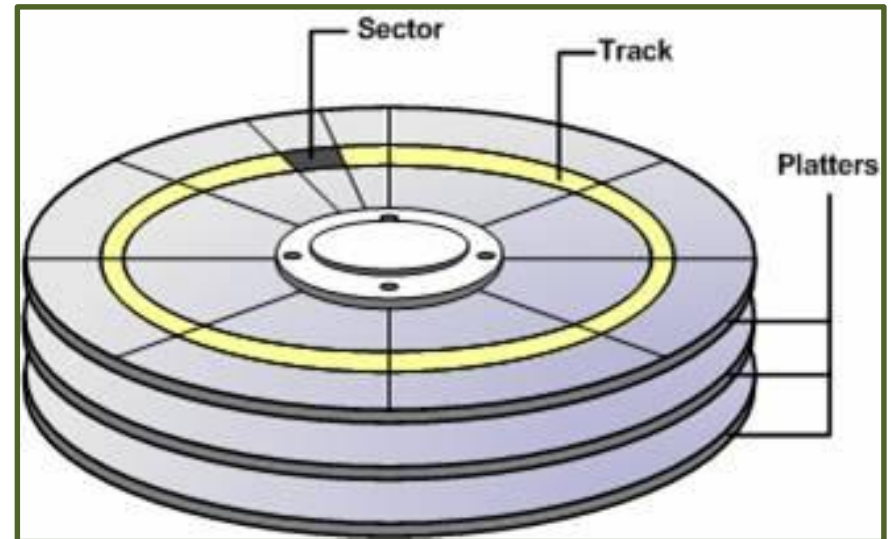
Índice

- Introducción
- Organización del almacenamiento en el computador
- **Discos duros y dispositivos SSD**
- RAID

Estructura de un disco duro (HDD)



Organización de la información
en un disco duro



En un disco duro la información se organiza en superficies, pistas y sectores.

Características de un disco duro

- Capacidad

- Concepto

Número de bytes almacenados por el disco.

- Expresión

- Se expresa habitualmente en GB o TB.-

- Consumo

Número de vatios consumidos durante su funcionamiento.

- Factor de forma

- Concepto

Hace referencia al tamaño del disco.

- Tipos

✓ LFF (Large Form Factor): 3,5"

✓ SFF (Small Form Factor): 2,5"

- Comparación LFF/SFF

Menor footprint	Mayor capacidad	Menor consumo
SFF	LFF	SFF

Características de un disco duro

- Velocidad de rotación

Valores habituales: 5400, 7200, 10000, 15000

- Cache

Memoria que almacena los sectores más frecuentemente accedidos, con objeto de incrementar las prestaciones. Rango de tamaños: [16MB, 256MB]

- Interfaz

SAS, SATA

- Número de puertos de datos

Modelos SATA: 2
Modelos SAS: 1

- Fiabilidad

- Forma de expresión habitual: AFR (Annualized Failure Rate)

Indica la probabilidad de fallo durante un año ininterrumpido de funcionamiento.

- Valores típicos del AFR en discos de clase empresarial

Alrededor del 0,5%.

Características de un disco duro / prestaciones

- Aspectos de las prestaciones
 - Aspecto 1: Velocidad de posicionamiento de las cabezas

Se expresa mediante el tiempo medio de acceso

Tiempo medio utilizado por las cabezas para posicionarse en el sector requerido por una operación de r/w.

Componentes del tiempo medio de acceso

- ✓ Tiempo de búsqueda medio (*average seek time*)

Tiempo medio que tardan las cabezas en moverse desde una pista a otra aleatoriamente elegida.

- ✓ Latencia (*latency*)

tiempo medio requerido para acceder a un sector, una vez que la cabeza se ha posicionado en la pista que contiene dicho sector. Depende de la velocidad de rotación del disco.

- Aspecto 2: Velocidad de transferencia cabeza/disco

Se expresa como la velocidad de transferencia sostenida (*sustained transfer rate*)

Número de bytes por unidad de tiempo que pueden ser movidos entre el disco y la interfaz que lo controla. Valores máximos actuales: alrededor de 250 MB/seg

Tiempo medio de acceso = tiempo de búsqueda medio + latencia = 3ms + 3ms = 6ms

latencia = $0,5 \text{ vueltas} / (10000 \text{ vueltas/min} \cdot \text{min}/60 \text{ seg}) = 0,003 \text{ seg} = 3 \text{ ms}$

Tipos de acceso a disco y su relación con las prestaciones

- Acceso aleatorio

- Concepto

Acceso a grupos de datos de pequeño tamaño y dispersos

- Ejemplo de escenario de acceso aleatorio

Bases de datos relacionales

- Factor determinante en la velocidad del acceso aleatorio

Tiempo medio de acceso

- Concepto de IOPS

Número de operaciones de E/S aleatorias por segundo. Es la inversa del tiempo medio de acceso.

- Acceso secuencial

- Concepto

Acceso a bloques de datos de gran tamaño ubicados en el disco de forma contigua

- Ejemplo de escenario de acceso secuencial

Streaming de vídeo

- Factor determinante en la velocidad de acceso secuencial

Velocidad de transferencia sostenida.

Ejemplos de discos gama empresarial

Seagate

Gama Exos orientada a prestaciones

Interfaz	SAS a 12 Gb/s
Factor de forma	SFF
Capacidades (GB)	300 - 2400
Cache (MB)	256
r.p.m.	10.000, 15.000
Velocidad de transferencia sostenida máxima (MB/s)	315
IOP	-
AFR	0,44%

Gama Exos orientada a capacidad

Interfaz	SAS a 12 Gb/s SATA 6 Gb/s
Factor de forma	LFF
Capacidades (TB)	1 - 14
Cache (MB)	256
r.p.m.	7.200
Velocidad de transferencia sostenida máxima (MB/s)	261
IOP	-
AFR	0,35%

Dispositivos SSD

- SSD significa *Solid State Drive*
- Concepto

Dispositivo de almacenamiento permanente basado en memoria flash y que se conecta al sistema de conexiones del computador mediante una interfaz SATA, SAS o PCIe.

Ejemplo de dispositivo SSD con Interfaz SATA



Factores de forma de dispositivos SSD y sus interfaces de conexión

- 2,5" SFF
 - Interfaces de conexión
 - ✓ SATA
 - ✓ SAS
 - ✓ PCIe x4 (mediante conector U.2)
- M.2
 - Interfaces de conexión
 - ✓ SATA
 - ✓ PCIe x2 o x4
- Tarjeta de expansión PCIe estándar
 - Interfaces de conexión
 - ✓ PCIe x4 o x8

Factores de forma de dispositivos SSD

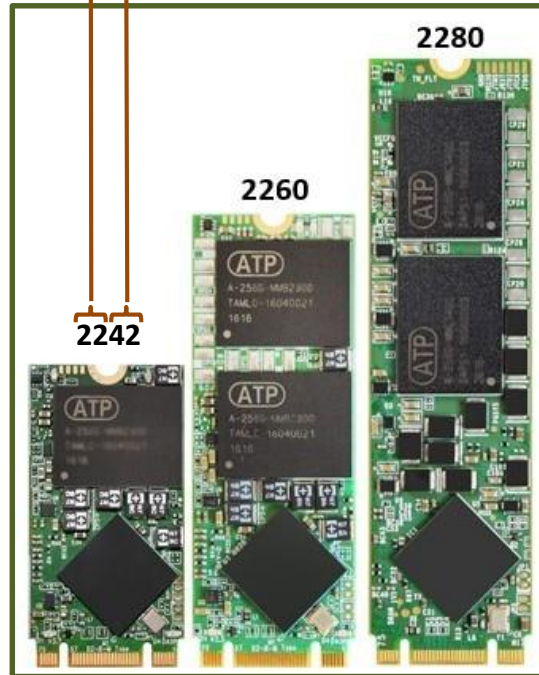
2,5" SFF



M.2

→ Ancho (en milímetros)

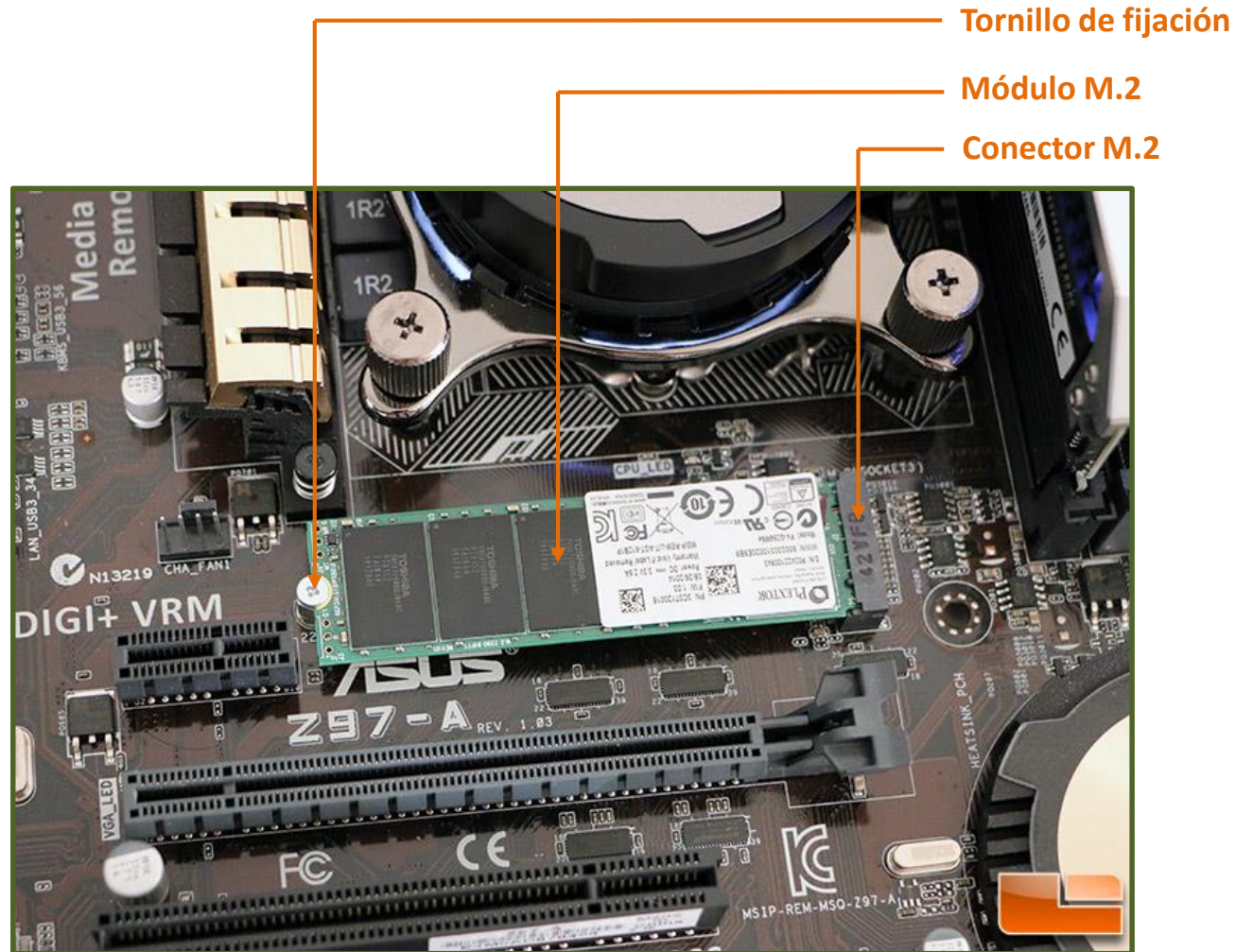
→ Largo (en milímetros)



Tarjeta de expansión PCIe



Instalación de un dispositivo M.2



Dispositivos SSD de gama empresarial (Seagate)

Gama Nytro	SAS a 12 Gb/s	SATA a 6 Gb/s
Factor de forma	SFF	SFF
Capacidades	15 TB – 400 GB	3,8 TB – 240 GB
Lectura secuencial sostenida (MB/seg) (1 puerto)	Hasta 1.100	Hasta 564
Lectura secuencial sostenida (MB/seg) (2 puertos)	Hasta 2.100	-
Escritura secuencial sostenida (MB/seg) (1 puerto)	Hasta 970	Hasta 536
Escritura secuencial sostenida (MB/seg) (2 puertos)	Hasta 1.400	-
Lectura aleatoria (IOPS) sostenida (1 puerto)	Hasta 200.000	Hasta 94.000
Lectura aleatoria (IOPS) sostenida (2 puertos)	Hasta 240.000	-
Escritura aleatoria (IOPS) sostenida (1 puerto)	Hasta 80.000	Hasta 61.000
Escritura aleatoria (IOPS) sostenida (2 puertos)	Hasta 80.000	-
AFR	0,35 %	0,5 %

Comparativa SSD / HDD

Característica	Comparativa
Capacidad máxima	Similares
Prestaciones acceso aleatorio	SSD muy superior (varios órdenes de magnitud)
Prestaciones acceso secuencial	SSD superior
Fiabilidad	Similares
Coste / GB	SSD más caro (x4, aproximadamente)

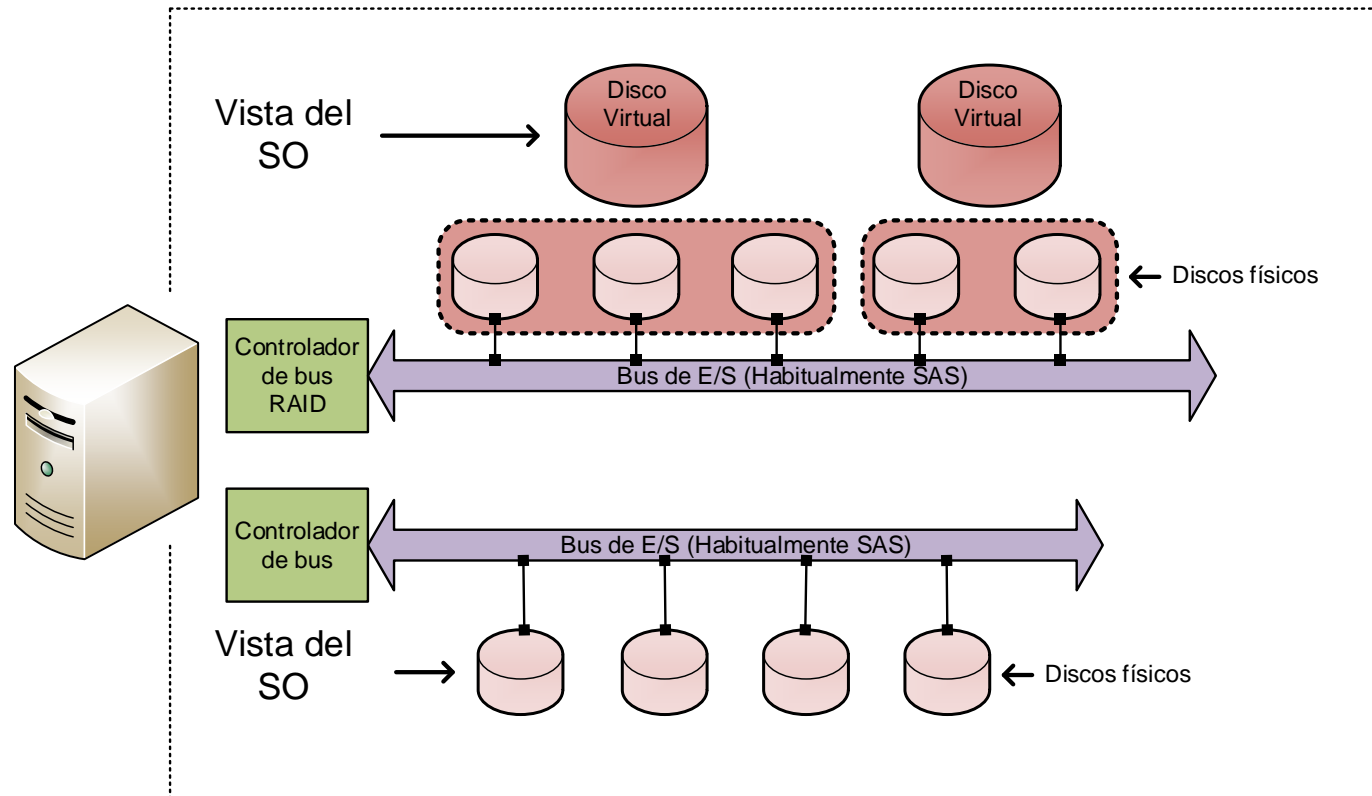
Índice

- Introducción
- Organización del almacenamiento en el computador
- Discos duros y dispositivos SSD
- **RAID**

Concepto

- RAID significa Redundant Array of Independent Disks
- Concepto

Capacidad añadida de algunos controladores de bus que permite gestionar los discos físicos en agrupaciones, que los hace funcionar en colaboración, de modo que son presentados al SO como si fueran un disco único, al que se hace referencia habitualmente como "disco virtual".



Aspectos generales

- Objetivos

- Incremento de prestaciones

Se consigue mediante fragmentación de datos (*data striping*). _____

Los datos se fragmentan entre múltiples unidades físicas, habilitándose de esta forma el paralelismo en el acceso a la información, lo que incrementa las prestaciones.

- Tolerancia a fallos

Se consigue mediante redundancia en el almacenamiento de la información. _____

La información se redunda entre los diversos discos del RAID, de forma que el sistema pueda continuar funcionando aunque se produzca el fallo en alguno de los discos físicos del RAID.

- Niveles RAID

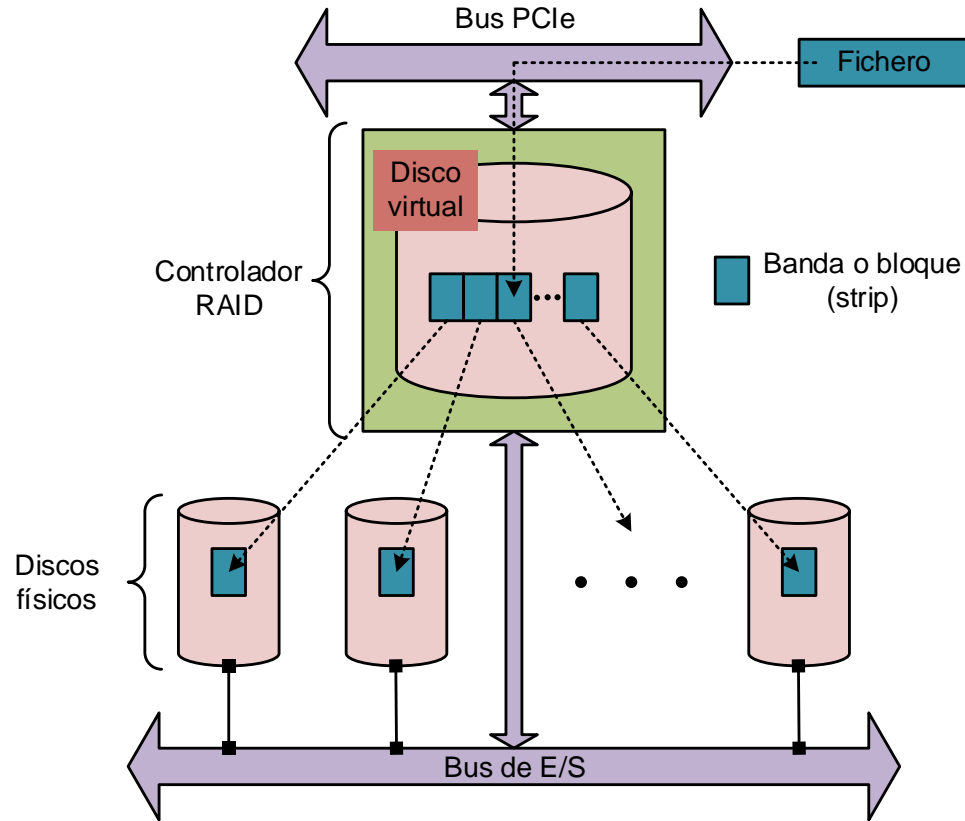
- Concepto

Representan las diferentes formas de distribuir la información entre los discos del RAID.

- Niveles RAID más comunes

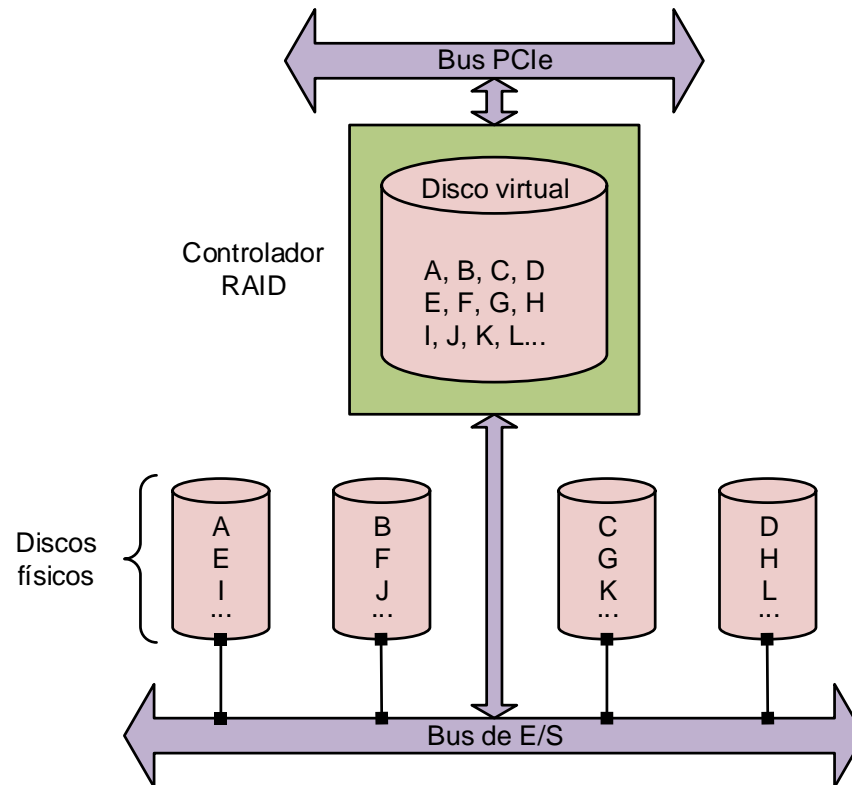
0, 1, 10, 5 y 6.

Concepto de banda o bloque (*strip*)



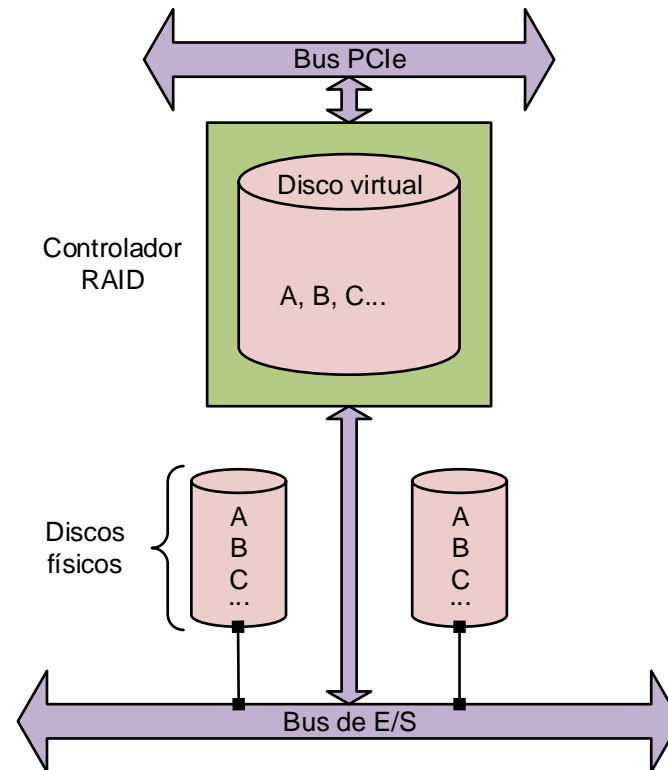
- **Definición:** Unidad de fragmentación de la información a escribir en el sistema RAID.
- **Tamaño:** Mínimo alrededor de 2KB, máximo alrededor de 2MB, y típico alrededor de 64KB.
- **Aplicación:** Se aplica a nivel de fichero.

RAID 0: Fragmentación (*striping*)



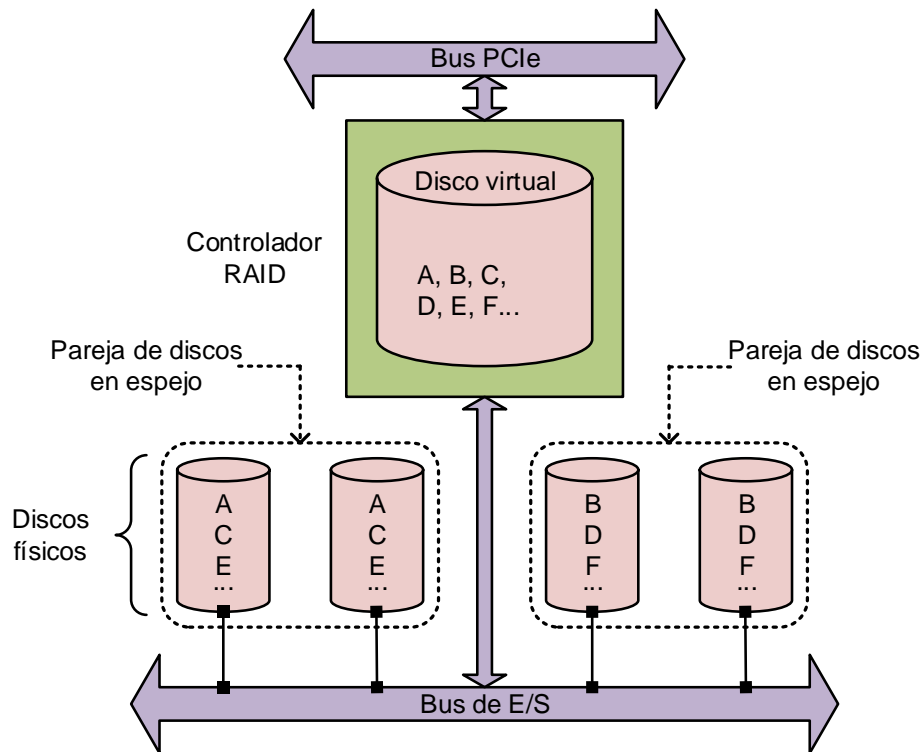
- Funcionamiento: Los bloques a escribir se almacenan en los discos siguiendo una distribución secuencial
- Redundancia: No
- Prestaciones en operaciones de lectura: $N \times P$ (siendo N el número de discos y P las prestaciones de un disco)
- Prestaciones en operaciones de escritura: $N \times P$

RAID 1: Espejo (*mirroring*)



- **Funcionamiento:** Los bloques a escribir se copian en todos los discos del RAID. Este tipo de RAID se configura con dos discos.
- **Redundancia:** Sí
- **Eficiencia en el almacenamiento:** 1/2
- **Prestaciones en operaciones de lectura:** 2P
- **Prestaciones en operaciones de escritura:** P

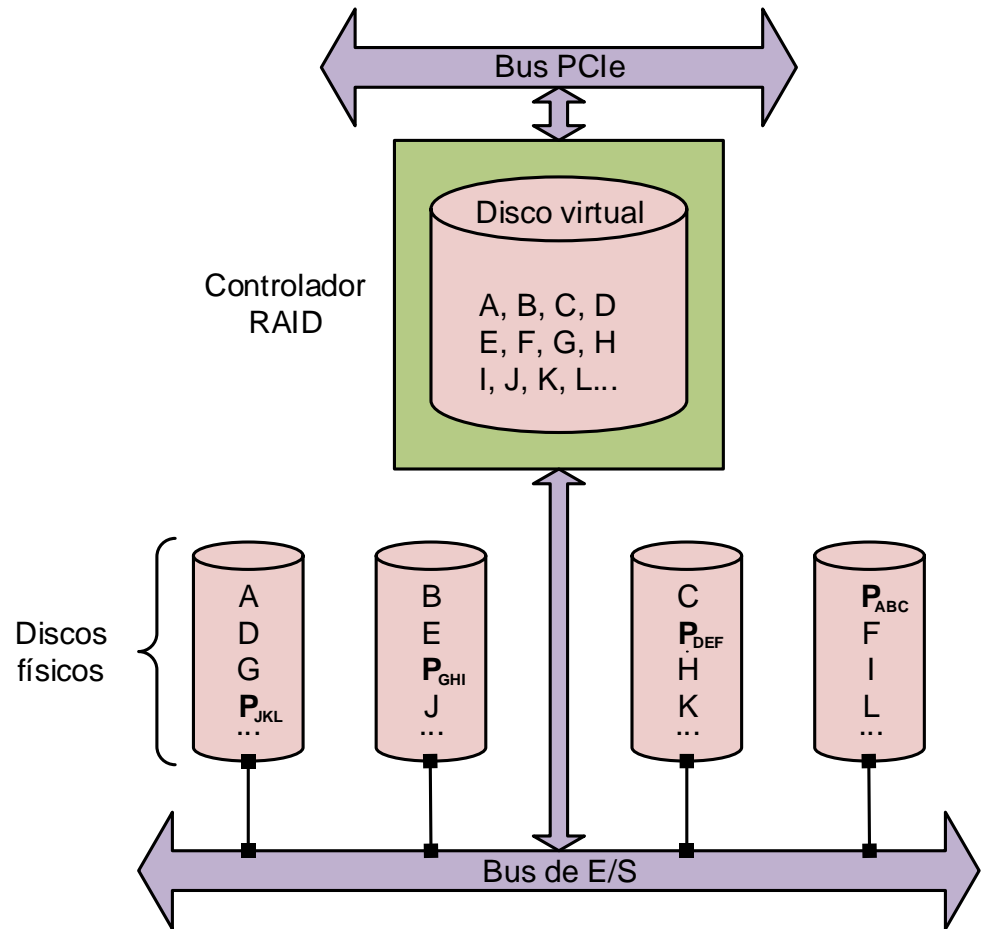
RAID 10: Espejo (*mirroring*)



- **Funcionamiento:** Los discos se organizan en parejas que funcionan en espejo. Los bloques se almacenan en las parejas de discos siguiendo una distribución secuencial.
- **Redundancia:** Sí
- **Eficiencia en el almacenamiento:** 1/2
- **Prestaciones en operaciones de lectura:** NP
- **Prestaciones en operaciones de escritura:** NP/2

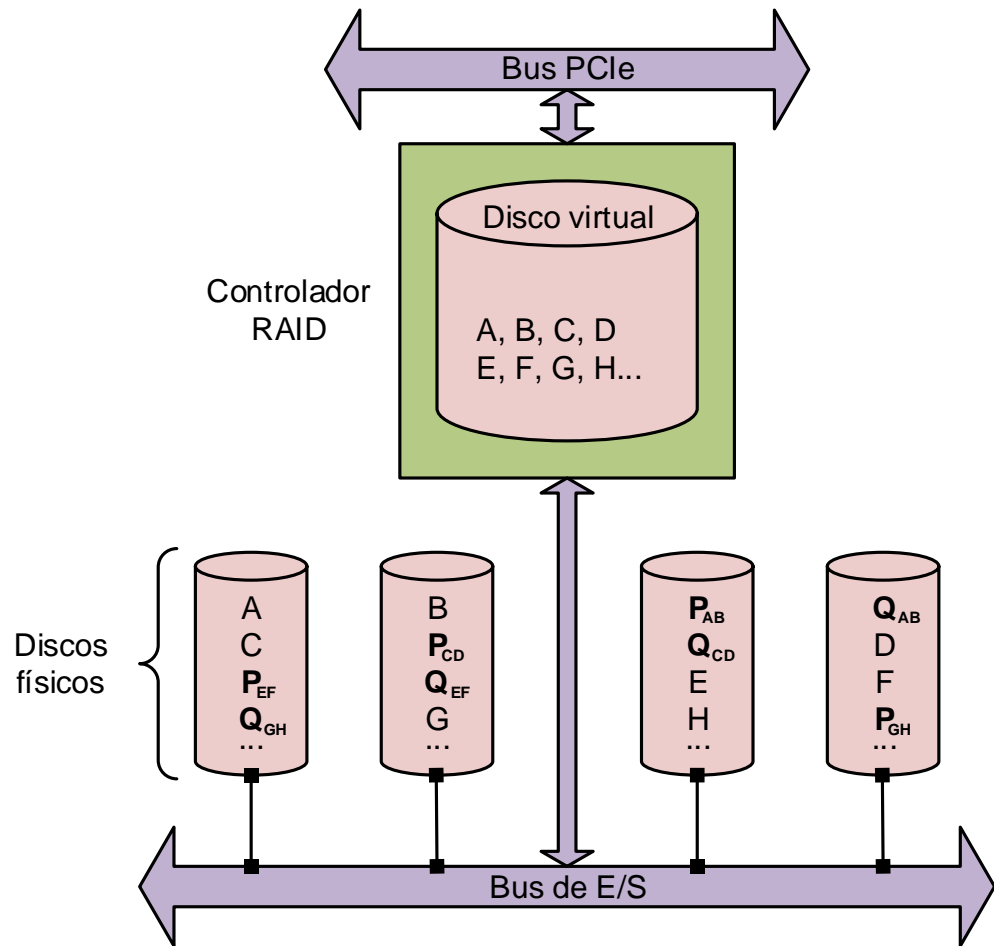
RAID 5: Paridad (*parity*)

- **Funcionamiento:** Dado un RAID de N discos, para cada (N-1) bloques de datos a escribir se genera un bloque de paridad.
- El bloque de paridad se calcula realizando la operación XOR sobre los bloques de datos correspondientes.
- Los bloques de datos se distribuyen en los discos de manera secuencial.
- Los bloques de paridad se intercalan entre los bloques de datos, de manera que queden uniformemente distribuidos entre todos los discos.
- Ante la pérdida de cualquier bloque de datos (debido al fallo de un disco), la información de dicho bloque puede calcularse a partir de los bloques de datos restantes del grupo, junto con el bloque de paridad.
- **Redundancia:** Sí
- **Eficiencia en el almacenamiento:** $(N-1)/N$
- **Prestaciones en op. de lectura:** $(N-1)P$
- **Prestaciones en op. de escritura:** $NP-C$



RAID 6: Paridad doble (*parity*)

- **Funcionamiento:** Dado un RAID de N discos, para cada (N-2) bloques de datos a escribir se generan dos bloques de paridad.
- Los bloques de paridad se calculan realizando la operación XOR sobre los bloques de datos correspondientes.
- Los bloques de datos se distribuyen en los discos de manera secuencial.
- Los bloques de paridad se intercalan entre los bloques de datos, de manera que queden uniformemente distribuidos entre todos los discos.
- El RAID es tolerante al fallo de dos discos.
- **Redundancia:** Sí
- **Eficiencia en el almacenamiento:** N-2
- **Prestaciones en op. de lectura:** (N-2)P
- **Prestaciones en op. de escritura:** NP-2C



Cache de escritura

- ¿Qué algoritmo de escritura se utiliza?

Funciona con algoritmo "write-back".

- ¿Cómo funciona?

Todas las escrituras se hacen en la caché. Los datos se mueven a los discos en el momento más conveniente de modo que se maximicen las prestaciones globales del sistema de almacenamiento.

- ¿Qué ocurre si se interrumpe la energía en el servidor?

La caché es alimentada por un sistema de condensadores durante el tiempo necesario para volcar su contenido a una memoria flash. Cuando el sistema vuelve a arrancar, tras recuperarse la energía, el contenido de la flash se vuelca a los discos.

- Denominación de la tecnología utilizada

Flash-backed write cache (FBWC)

Conveniencia de la cache en función del tipo de dispositivos del RAID

- RAID de dispositivos HDD

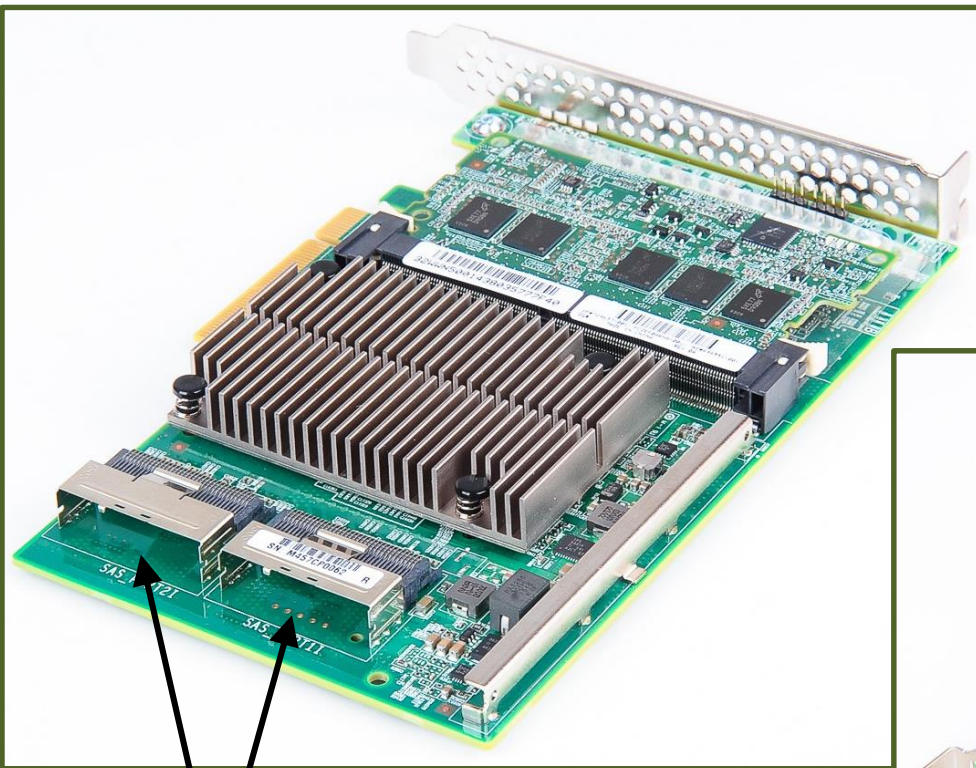
La cache es esencial para obtener unas prestaciones razonables, especialmente si el porcentaje de escrituras en la carga de trabajo es significativo.

- RAID de dispositivos SSD

La cache no es necesaria

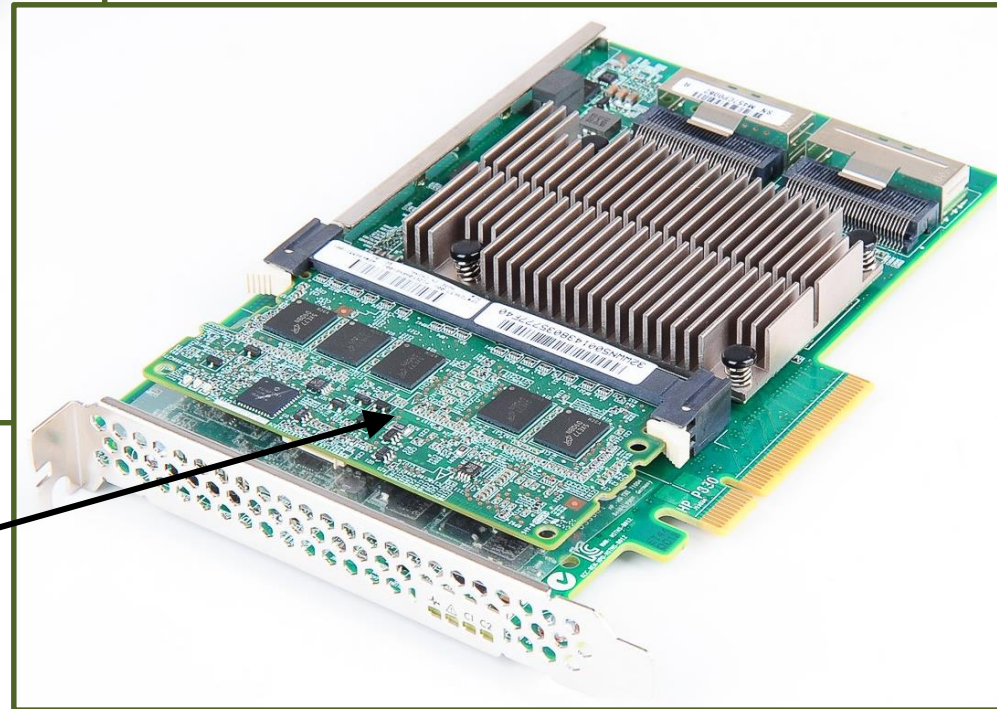
Ejemplo de controladora RAID de clase empresarial

HPE Smart Array P830



Puertos SAS x8

Condensador



FBWC

HPE Smart Array P830

Propiedades esenciales	
Interfaz de placa base	PCIe
Interfaz de disco	SAS 12 Gb/s
Puertos físicos	2 x 8
Niveles de RAID soportados	0, 1, 10, 5, 50, 6, 60, 10 ADM
Cache	4 GB

Ejercicio: cálculo del tiempo medio de acceso de un disco 10K, sabiendo que el tiempo de búsqueda medio es de 3ms

Ejercicio: ejemplo de cálculo de paridad en RAID 5

0	1	0	1
1	1	0	0
0	0	0	0
1	1	1	1

$$\begin{aligned}0 \text{ xor } 1 \text{ xor } 0 &= 1 \\1 \text{ xor } 1 \text{ xor } 0 &= 0 \\0 \text{ xor } 0 \text{ xor } 0 &= 0 \\1 \text{ xor } 1 \text{ xor } 1 &= 1\end{aligned}$$

0	1	0	1
1	1	0	0
0	0	0	0
1	1	1	1

$$\left. \begin{aligned}0 \text{ xor } 1 \text{ xor } 0 &= 1 \\1 \text{ xor } 1 \text{ xor } 0 &= 0 \\0 \text{ xor } 0 \text{ xor } 0 &= 0 \\1 \text{ xor } 1 \text{ xor } 1 &= 1\end{aligned} \right\} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \checkmark$$