

Sistemas de Computación

2do Semestre 2010

Gestión de E/S

Discos

Wenceslao Palma M. <wenceslao.palma@ucv.cl>

Conceptos

Considerando el almacenamiento se tiene la siguiente jerarquía: Memoria Caché, Memoria Principal, Discos Magnéticos, Memoria terciaria.

En forma idea siempre es recomendable trabajar con los datos en memoria volátil, debido a la velocidad de acceso, para luego enviarlos a memoria secundaria o terciaria.

Memoria Caché:

- su objetivo es reducir los tiempos de espera.

- es pequeña pero de mucha rapidez.

- es necesario administrarla.

Memoria Principal:

- se utiliza para el procesamiento de los datos.

Conceptos

Disco Magnético:

no es volátil.

dispositivo de acceso aleatorio.

permite almacenar gran cantidad de bytes.

para procesar su contenido es necesario el uso de la memoria principal.

Memoria terciaria:

considera entre otros, CDs, DVDs y Cintas.

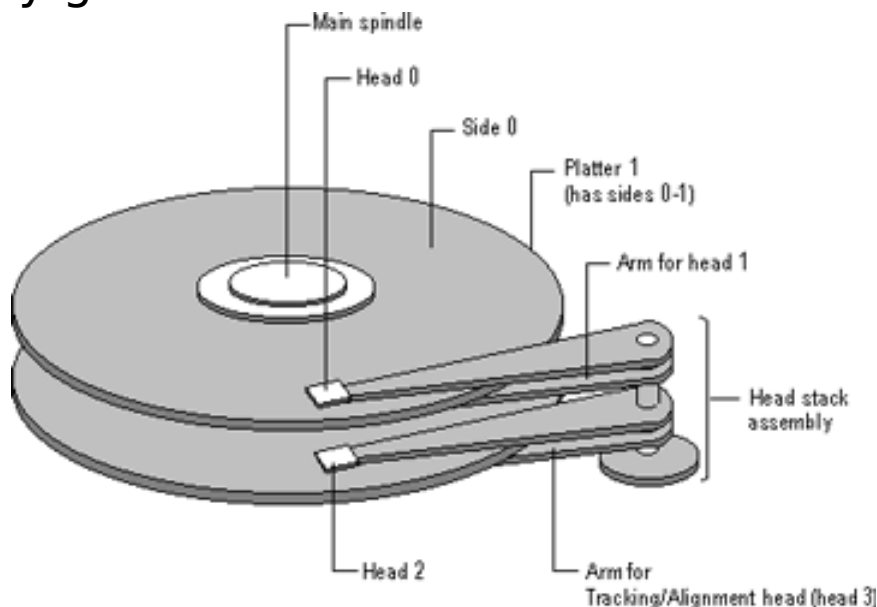
una cinta es un dispositivo de acceso secuencial, es mas lenta que un disco magnético, generalmente se utiliza para respaldos.

La principal desventaja en la utilización de Discos se encuentra en el tiempo de acceso y de recuperación. Sin embargo este se puede disminuir seleccionando una organización adecuada.

Conceptos

Disco Magnético

también conocido como disco duro, se compone de un pack de platos cada uno de los cuales tiene caras, pistas y sectores para almacenar la información. Además posee cabezas, asociadas a un brazo, para leer y grabar la .



Una cabeza se mueve radialmente sobre la superficie de un plato que gira a gran velocidad. Sólo una cabeza puede realizar transferencia de datos en un determinado momento.

Conceptos

El tiempo de acceso a un disco está dado por *tiempo de seek+latencia (rotacional)+ tiempo de transferencia*.

seek: tiempo requerido para transitar entre pistas.

latencia: tiempo requerido para que el disco gire al sector que se desea. Generalmente corresponde a media vuelta.

Transferencia: velocidad con la que se escriben/leen bytes hacia/desde el disco.

Conceptos

Ejercicio: considere el ordenamiento y mezcla de un archivo de 800000 registros. Sólo 10000 se pueden almacenar en RAM.

Tamaño del registro 100 bytes

Se utiliza un disco de 3600 RPM que posee seek de 18 ms y transferencia de 1229 bytes/ms

Cuanto es el tiempo requerido para realizar el ordenamiento y mezcla?

Conceptos

Para mejorar el desempeño:

minimizar movimientos del brazo.

lo ideal para la latencia es que la cabeza se encuentre justo en el sector que contiene el dato que se requiere. Sectores contiguos v/s sectores intercalados.

transferencia en paralelo desde el pack de discos, para esto se requieren varios brazos. Ante esto nace el concepto de RAID (Redundant Array of Independent Disks)

RAID (Redundant Array of Independent Disks)

El array es administrado por una controladora que contiene RAID firmware.

RAID administrado por hardware es más rápido que el administrado por software.

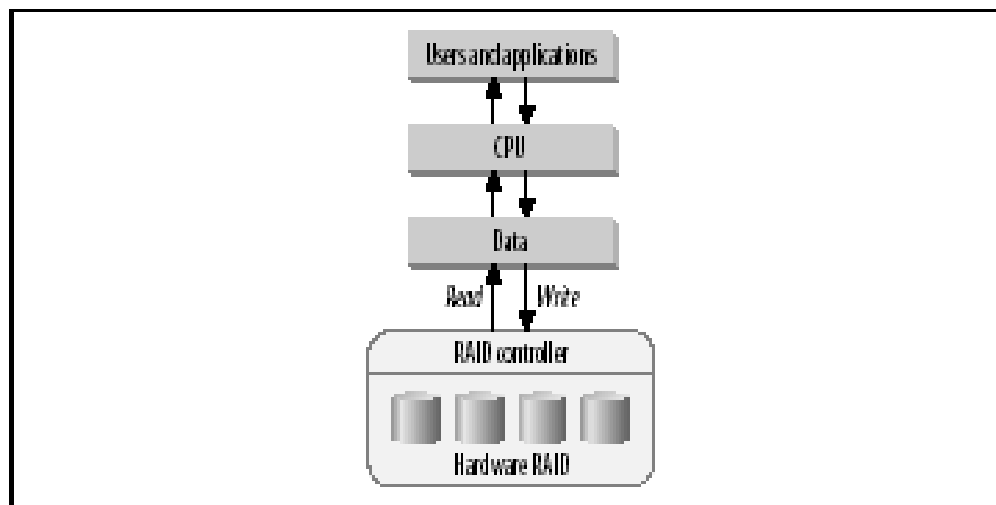


Figure 2-2. Disk controllers shift the array functions off the CPU, yielding an increase in performance.

La controladora posee una BIOS que proporciona las herramientas de administración para la configuración y mantención.

El sistema operativo ve el arreglo como un “gran disco duro”.

RAID (Redundant Array of Independent Disks)

También existen soluciones externas, las cuales son conectadas mediante SCSI, Ethernet o Fibra.

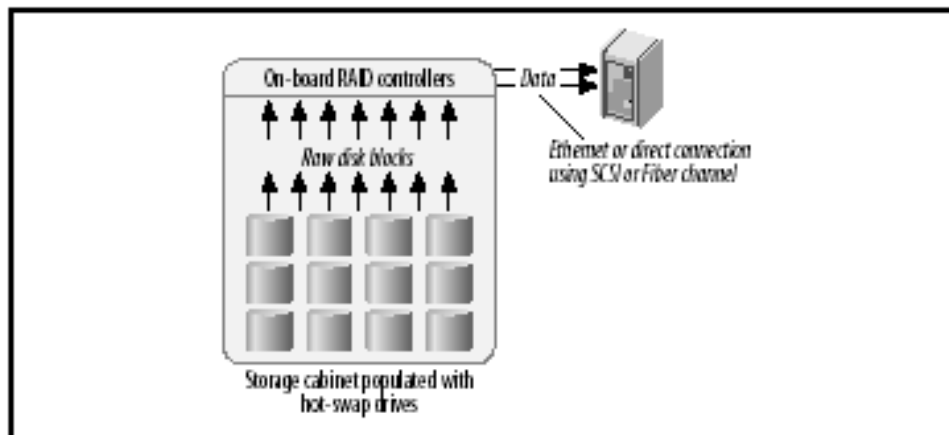


Figure 2-3. Outboard RAID systems are internally managed and connected to a system to which they appear as a single hard disk.

RAID (Redundant Array of Independent Disks)

Niveles RAID

Existen distintas alternativas de configurar un arreglo de discos.

Generalmente un nivel se define en base a los requerimientos de aplicaciones que lo utilizan.

Básicamente constituyen un compromiso entre redundancia y rendimiento.

RAID 0(striping)

En rigor no corresponde a un nivel RAID, debido a que no presenta redundancia.

Conviene utilizarlo cuando los datos no son críticos y el rendimiento es importante.

RAID (Redundant Array of Independent Disks)

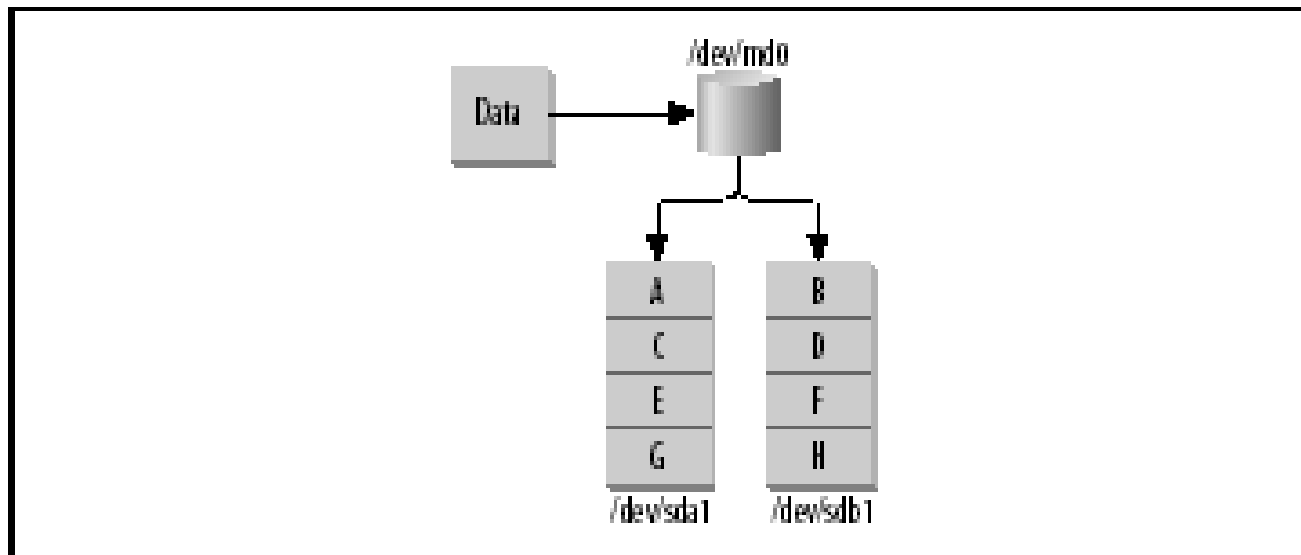


Figure 2-5. RAID-0 (striping) writes data consecutively across multiple drives.

RAID (Redundant Array of Independent Disks)

RAID 1(mirroring)

Proporciona la forma más completa en cuanto redundancia.

Puede soportar muchas fallas sin la necesidad de un algoritmo de recuperación.

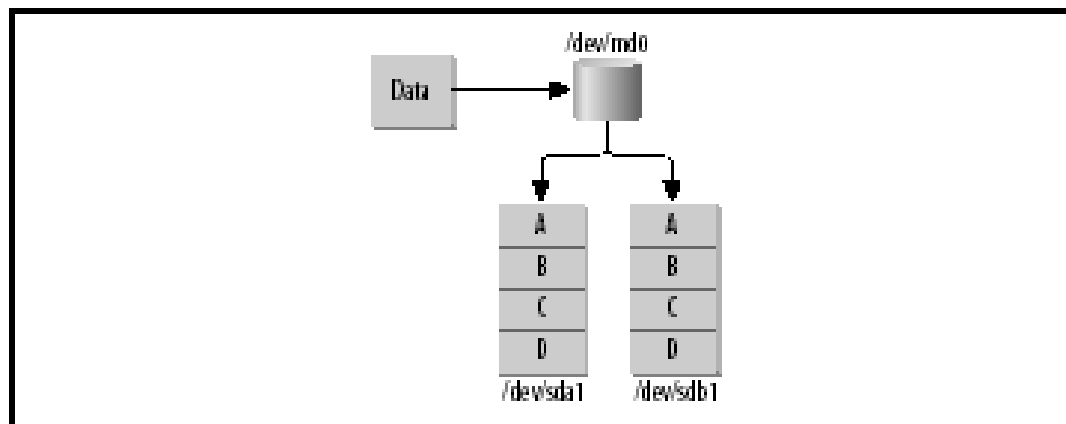


Figure 2-7. Fully redundant RAID-1.

Para un array de N discos, es posible soportar la falla de N-1 de ellos sin pérdida de datos.

RAID (Redundant Array of Independent Disks)

Sin embargo, el rendimiento de las operaciones de escritura es inversamente proporcional al aumento de discos en el array.

Por otro es posible realizar lecturas en forma concurrente.

IMPORTANTE: el costo de implementación es como mínimo el doble del requerimiento de almacenamiento.

RAID (Redundant Array of Independent Disks)

RAID 4

Utiliza un disco para almacenar información de paridad, la cual se utiliza ante un eventual desastre.

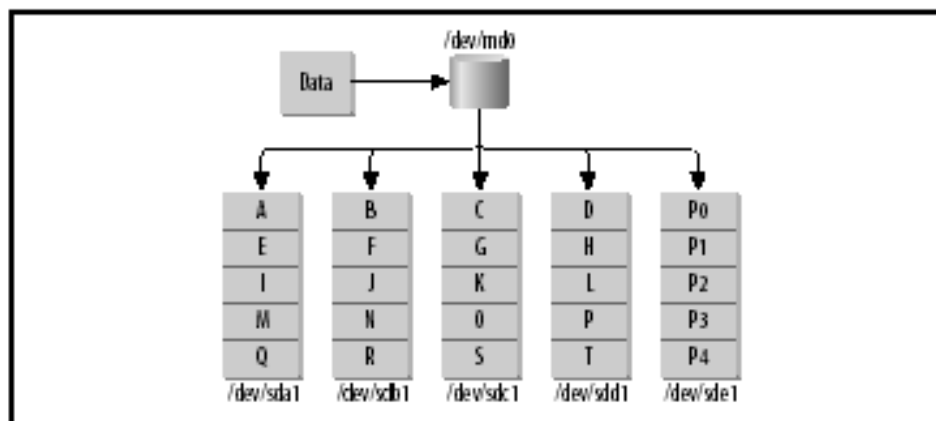


Figure 2-8. RAID-4 stripes data to all disks except a dedicated parity drive.

RAID (Redundant Array of Independent Disks)

RAID 5

aquí se elimina el uso de un disco exclusivo para el tema de la paridad. Se utiliza un bloque por cada disco

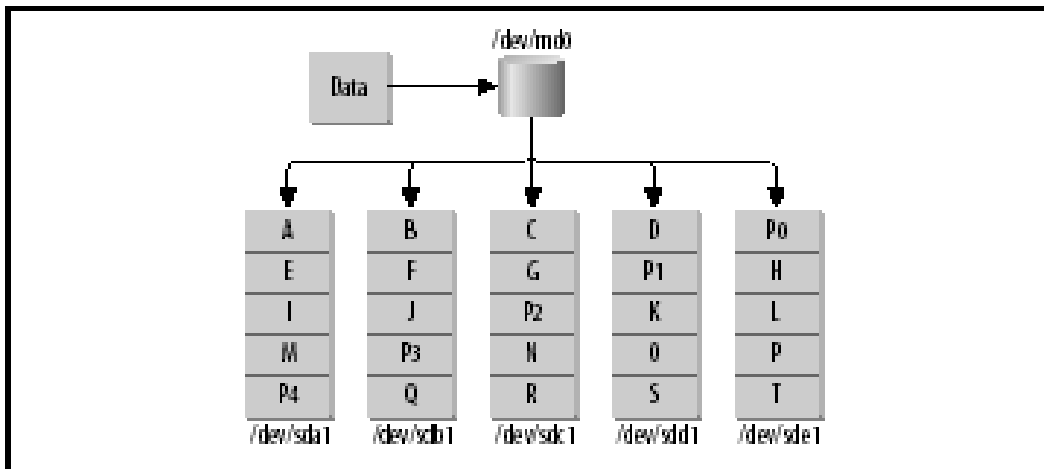


Figure 2-9. RAID-5 eliminates the dedicated parity disk by distributing parity across all drives.

RAID (Redundant Array of Independent Disks)

Hibridos

Combinan diferentes niveles RAID para incrementar rendimiento y confiabilidad.

RAID 10(striping mirror)

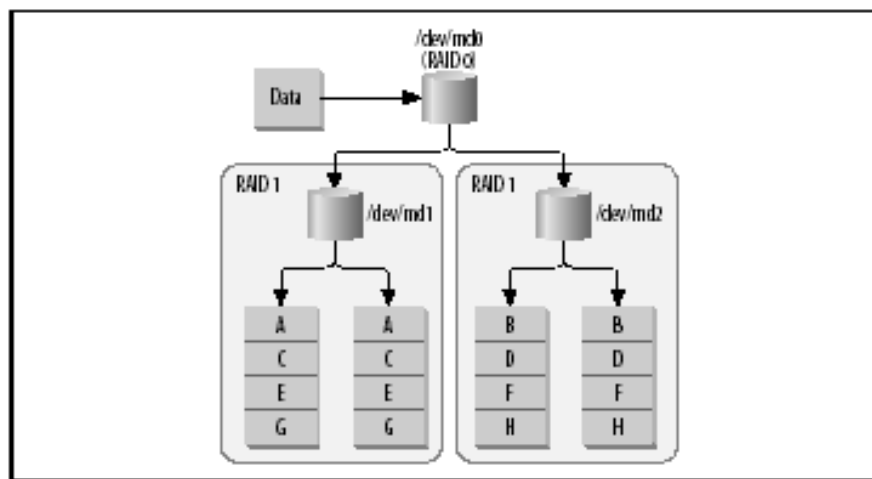


Figure 2-10. A hybrid array formed by combining two mirrors, which are then combined into a stripe.

RAID (Redundant Array of Independent Disks)

RAID 50 (striping parity)

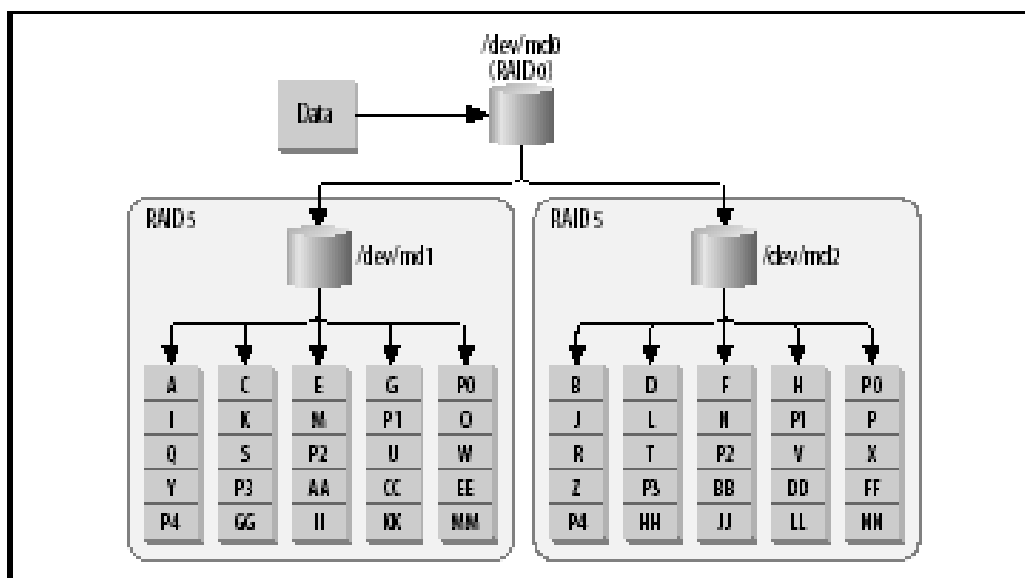


Figure 2-12. A hybrid array formed by combining RAID-5 arrays into a striped array.

RAID (Redundant Array of Independent Disks)

Comparaciones en el rendimiento

	RAID 1	RAID 0	RAID 4	RAID 5
Operación de escritura	Lenta, mas aun si se agregan discos.	Mucho mejor que un único disco.	Comparable a raid 0 con un disco menos.	Comparable a raid 0 con un disco menos

	RAID 1	RAID 0	RAID 4	RAID 5
Operación de lectura	rápida, mas aun si se agregan discos.	El mejor	Comparable a raid 0 con un disco menos.	Comparable a raid 0 con un disco menos

RAID (Redundant Array of Independent Disks)

Fallas y tipos de aplicaciones

	RAID 1	RAID 0	RAID 4	RAID 5
Nro de fallas	N-1	0	1	1

	RAID 1	RAID 0	RAID 4	RAID 5
Aplicaciones	Servidor de imágenes. En gnral, sistemas con poca volatilidad .		Equivalente a raid 5	Servidor de archivos, bases de datos.

RAID (Redundant Array of Independent Disks)

Fallas en los discos

un beneficio de RAID es la capacidad de afrontar fallas sin la intervención del usuario.

Hot-spares : reemplazo de un disco sin intervención del usuario.

Hot-swap : reemplazo del un disco con intervención del usuario.

RAID (Redundant Array of Independent Disks)

Consideraciones de hardware

No debemos olvidar que un array es solo un componente mas de un sistema computacional.

Muchos factores afectan el rendimiento y la capacidad de expansión de un arreglo de discos:

- Throughput del bus.

- Canales de I/O.

- Throughput del protocolo de acceso al disco.

- Velocidad del disco.

- CPU y memoria.

Considerando estos factores la idea es poner atención a los posibles cuellos de botella que se puedan generar.

RAID (Redundant Array of Independent Disks)

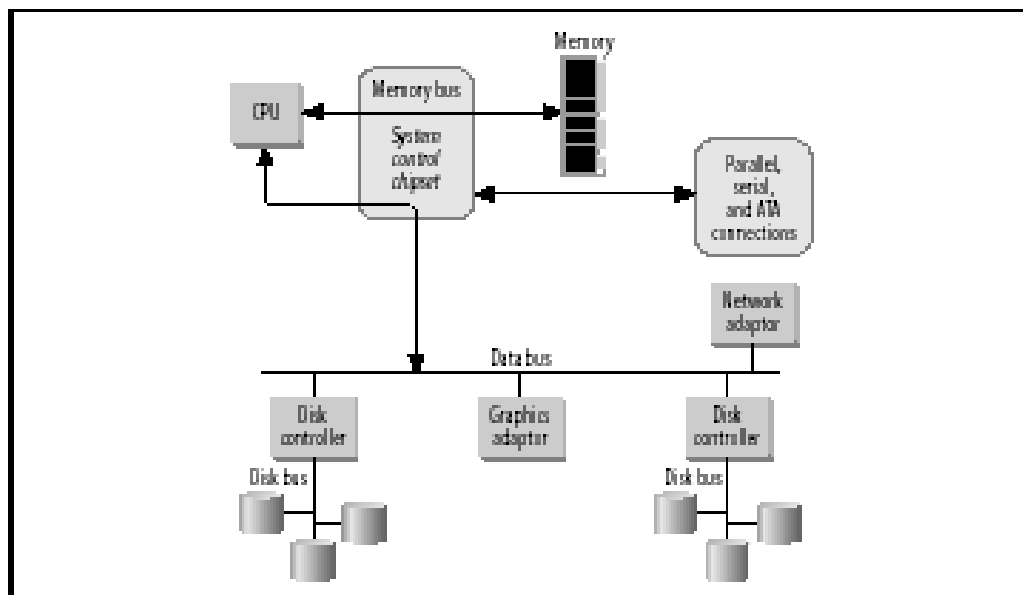


Figure 2-15. When disk I/O occurs, data travels over both the data bus and the disk bus, each a potential bottleneck.

La velocidad de los buses de datos y de disco tiene un impacto directo en el desempeño del sistema. Es sencillo agregar más controladoras de disco aumentando el throughput pero el bus de datos es uno solo.

RAID (Redundant Array of Independent Disks)

Bus PCI

la velocidad es determinada por dos factores: el ancho del bus (bus-width) y la velocidad del bus (bus-speed)

bus-width: indica cuantos bytes de datos se pueden enviar a la vez.

bus-speed: especifica cuantas veces por segundo es posible transferir datos por el bus.

Tipo de Bus	Width (bits)	Ciclos (MHz)	Throughput (MB/s)
ISA (xt)	8	8.33	8.33
ISA (at)	16	8.33	16.66
PCI	32-64	33.33-66.66	133.33-533.33

RAID (Redundant Array of Independent Disks)

Si se construye un RAID considerando 3 tarjetas SCSI cada una de las cuales tienen una velocidad de 80MB/s, es posible generar un cuello de botella si se cuenta con una tarjeta madre que posee un bus PCI de 32 bits corriendo a 33MHz ($80\text{MB/s} \times 3 \text{ v/s } 133.33 \text{ MB/s}$)

RAID (Redundant Array of Independent Disks)

Protocolos de acceso a disco

tienen un impacto en el rendimiento y escalabilidad del array.

Cada protocolo tiene bien definido su throughput y el número de dispositivos que se pueden conectar.

En las soluciones basadas en arrays, el más usado es SCSI.

El bus-width SCSI se conoce como narrow (8 bits) o wide (16 bits). Determinan la cantidad de discos que se pueden conectar a un bus SCSI.

narrow : 8, sin embargo la cantidad efectiva es 7.

wide : 16, la cantidad efectiva es 15.

En ambos casos se debe a que se considera la tarjeta controladora.

RAID (Redundant Array of Independent Disks)

Nombre	width	Signal rate (MHz)	Troughput (MB/s)
SCSI-1, SCSI,Narrow SCSI	8	5	5
Fast SCSI	8	10	10
Fast Wide SCSI	16	10	20
Ultra SCSI	8	20	20
Ultra wide SCSI	16	20	40
Ultra2 SCSI, Ultra2 Narrow	8	40	40
Ultra2 wide SCSI	16	40	80
Ultra3 SCSI, Ultra 160 SCI	16	80	160

RAID (Redundant Array of Independent Disks)

Referencias

Managing RAID on Linux. Derek Vadala. Ed. O'Reilly.