Centro de
Investigación en
Computación

Sistemas de Almacenamiento

HDD (Hard Disk Drive)

Discos

Arreglos de discos

Cintas

Etc.

Course

Operating System (with focus on Security)

Instructor

Acosta Bermejo Raúl

Lecture notes

Instituto
Politécnico
Nacional

Table of contents (outline)

Tabla de contenido

1. Introducción
2. Buses de periféricos
    1. ATA (PATA, SATA)
    2. SCSI
    3. SAS
3. Cintas
4. Arreglos de discos (RAID)
5. Almacenamiento en Red
    7. SAN
    8. NAS
6. Protocolos de acceso a discos
7. Temas avanzados

Introducción

Historia

Dispositivos de almacenamiento:

- Discos duros (*hard disks*, HDD *hard disk drives*).
- Floppy drives.
- Optical disc drives
- Cinta (magnética)

Disco duro

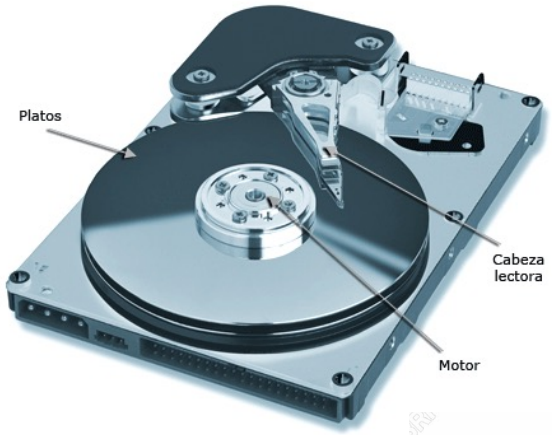
Diseño físico

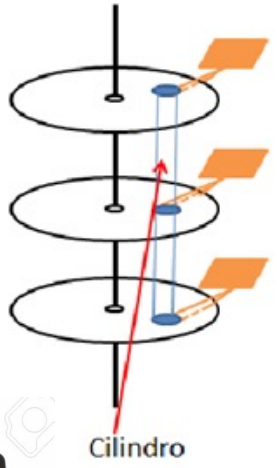
Diseño lógico

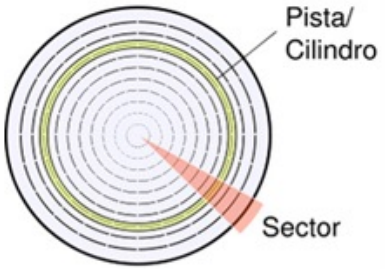
Tabla de particiones

Disco duro

Diseño físico

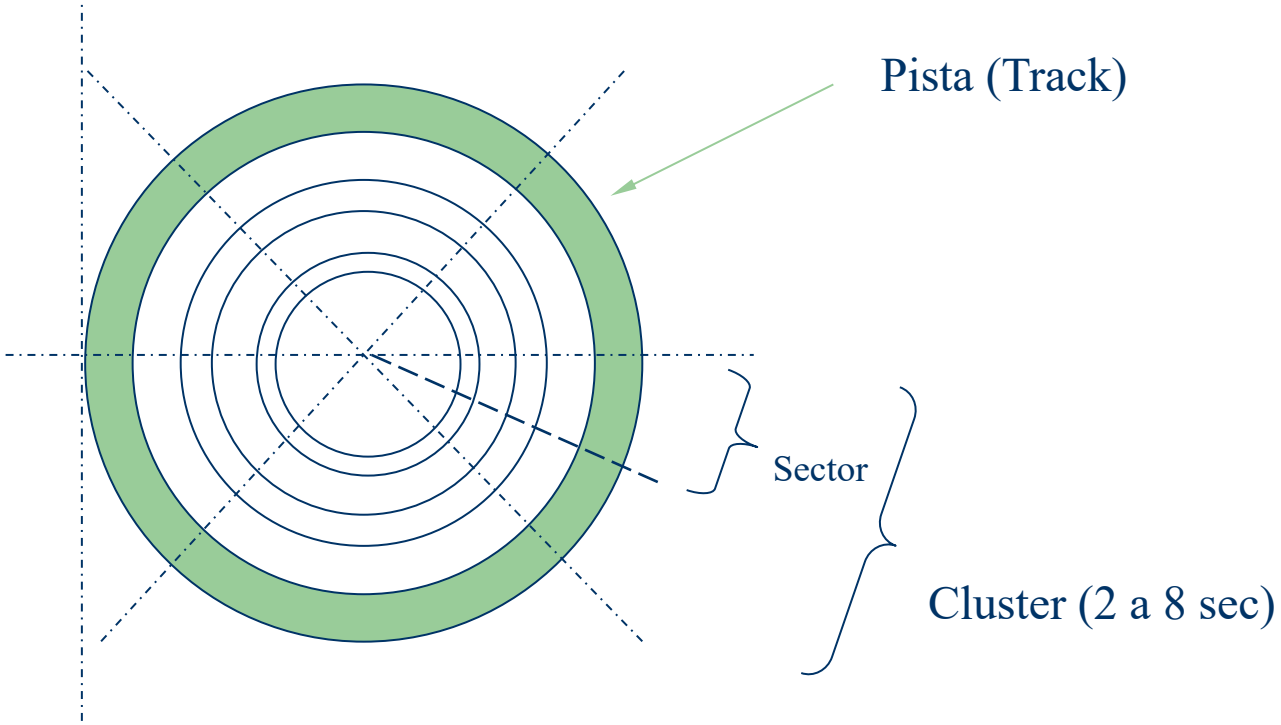






Disco duro

Diseño físico



El dispositivo mecánico (cabeza, head) se mueve por pasos formando cilindros o pistas.

CHS = Cilindre, Head, Sector

Disco duro

Diseño físico

Ejemplos

3 1/2 (720K) 9 sec/pis
5 1/4 (1.2M) 15 sec/pis
Double Side (DS)
Single Side (SS)

Drive

Disco duro

Diseño lógico

Tabla de Particiones	*Partición 1*	*Partición 2*	.....	*Partición N*

1. Normalmente sólo puede haber 4 particiones físicas o 3 físicas y una extendida.
2. Dentro de la partición extendida se pueden crear particiones lógicas.
3. Cada SO puede manejar su propio concepto de partición lógica, por ejemplo Linux maneja ahora VL (Volúmenes Lógicos).
4. Sólo una de las particiones es “booteable” (edo. en la tabla de particiones, se cambia).
5. Antes de usar una partición esta se tiene que formatear para crear el *Sistema de archivos*.

Disco duro

Tabla de particiones

Contiene las características del DD:

Campo	***Tamaño*** (bytes)
Cabeza de inicio	1
Sector de inicio	1
Cilindro de inicio	1
Id del sistema	1
Cabeza final	1
Sector final	1
Cilindro final	1
1er Sector de la Part.	4
Sectores en la Part.	4

00h Desconocido, 01h DOS 12bits FAT,
04h DOS 16bits 05h DOS Ext disk 16 bits FAT

Hay varias herramientas que pueden manipular la tabla de particiones

Fdisk, Norton Utilities

Disco duro

Tabla de particiones

MBR (*Master Boot Record*)

- Es el primer sector (sector 0) de un dispositivo de almacenamiento.
- Se usa para:
  - El arranque del SO con bootstrap, o
  - Almacena la tabla de particiones.
  - Usa el modelo Cilindro-Cabeza-Sector (CHS).

GPT (Tabla de partición GUID, *Global Unique Identifier*)

- Es un estándar para la colocación de la tabla de particiones en un disco duro físico.
- Es parte del estándar EFI propuesto por Intel para remplazar el BIOS de IBM .
- Sustituye la MBR usado con el BIOS.
- Usa un moderno modo de direccionamiento lógico (LBA) en lugar del CHS.

Disco duro

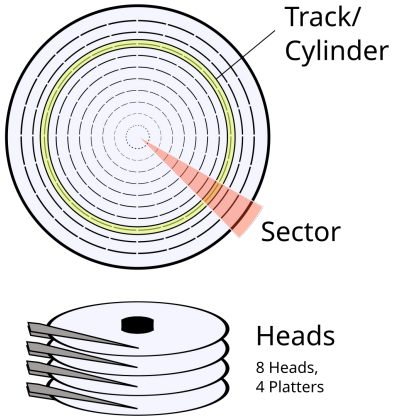
Modo de direccionamiento lógico

LBA (*Logical Block Addressing*, Modo de Direccionamiento Lógico).

- It is a common scheme used for specifying the location of blocks of data stored on computer storage devices, generally secondary storage systems such as hard disk drives.
- LBA is a particularly simple linear addressing scheme; blocks are located by an integer index, with the first block being LBA 0, the second LBA 1, and so on.
- The IDE standard included 22-bit LBA as an option, which was further extended to 28-bit with the release of ATA-1 (1994) and to 48-bit with the release of ATA-6 (2003), whereas the size of entries in on-disk and in-memory data structures holding the address is typically 32 or 64 bits.
- Most hard disk drives released after 1996 implement logical block addressing.

Disco duro

Equivalencias

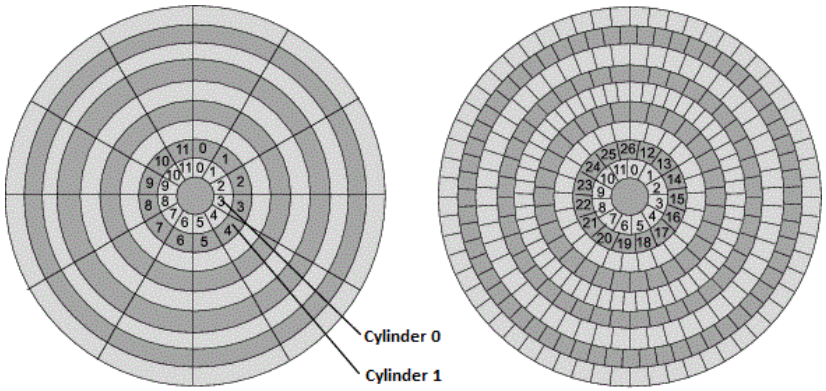


CHS vs LBA

LBA = ((C x HPC) + H) x SPT + S – 1
where:
CH S are the **C**ylinder number, the **H**ead number, and the **S**ector number
LBA is the logical block address
HPC is the number of heads per cylinder
SPT is the number of sectors per track

LBA	C	H	S
0	0	0	0
1	0	0	1
2	0	0	2
3	0	0	3
4	0	0	4
5	0	0	5
6	0	0	6
7	0	0	7
8	0	0	8
9	0	0	9
10	0	1	0
11	0	1	1
12	0	1	2
13	0	1	3
14	0	1	4
15	0	1	5
16	0	1	6
17	0	1	7
18	0	1	8
19	0	1	9
Cylinder 0			

LBA	C	H	S
20	1	0	0
21	1	0	1
22	1	0	2
23	1	0	3
24	1	0	4
25	1	0	5
26	1	0	6
27	1	0	7
28	1	0	8
29	1	0	9
30	1	1	0
31	1	1	1
32	1	1	2
33	1	1	3
34	1	1	4
35	1	1	5
36	1	1	6
37	1	1	7
38	1	1	8
39	1	1	9
Cylinder 1			



CHS Adressing **LBA Adressing**

Formatear
Crea la estructura de
Cilindros y

Disco duro

Equivalencias

Herramientas

Linux

- Fdisk. La herramienta más antigua y aun disponible. Solo modo txt.
- Cada distribución tiene su herramienta
  - GNU. Gparted. Incluso puede estar instalado en una Pendrive.
    - https://gparted.org/live**cd**.php
  - KDE. KDE Partition Manager
    - .
- Herramientas con otras funcionalidades
  - https://clonezilla.org/ Free
  - https://www.acronis.com/ Comercial

Windows

- Bootear de una USB un Linux
  - https://pendrivelinux.com/

ATA

Advanced Technology Attachment

History
PATA
SATA

ATA

History

Resumen de historia y versiones

1. ATA-1
    1. Conocido como ATA/ATAPI actualmente. Originalmente conocido como IDE. Desarrolado por Western Digital (WD).
    2. Apareció en 1986 en las Compaq.
    3. El original era un bus ISA de 16 bits.
2. ATA-2
    1. Enhanced IDE (EIDE) propuesto por WD en 1994 (estandar en el 96).
    2. Otros nombres de otros fabricantes: Fast ATA, Fast ATA-2.
3. ATA-4
    1. Conodido originalmente como Ultra DMA (UDMA).
    2. Velocidades de 16 MB/s a 33 MB/s.

El término Integrated Drive Electronics (IDE), Enhanced IDE (EIDE) son sinónimos de ATA (actualmente Parallel ATA, o PATA).

PATA

Parallel ATA

Es una interfaz para dispositivos de almacenamiento. Es el resultado de una larga historia de desarrollos técnicos incrementales:

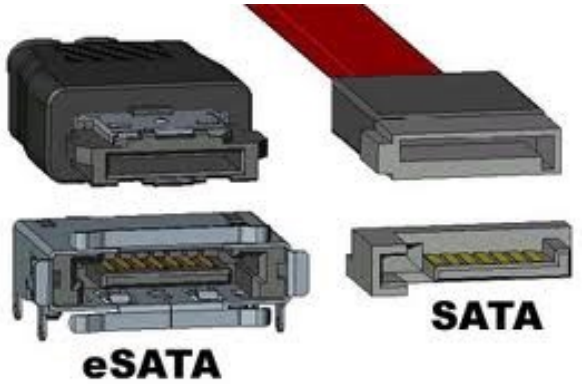
1. Originalmente se le llamó AT Attachment (ATA).
2. Otros nombres adoptados: ATA/ATAPI.
3. Evolucionó del IDE de Western Digital.
4. Al surgir el SATA en 2003 se le renombró a PATA.

SATA

Serial ATA

Publicado en el 2003:

1. Es una arquitectura punto a punto.
    1. Cada dispositivo se conecta directamente a un controlador SATA.
    2. No como en el PATA en una interface segmentada en maestras y esclavas.
2. Mismo conector para equipos de escritorio y servidores.
    1. El ATA utiliza conectores diferentes: servidores (3.5"), cliente (2.5").
3. **eSATA**.- Estándar para unidades externas.
4. Velocidades:
    1. SATA I.- 150 MB/s.
    2. SATA II.- 300MB/s.
    3. SATA III.- 600 MB/s.
    4. eSATA.- 115 MB/s.



SCSI

Small Computer System Interface

History

SAS

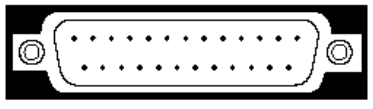
SCSI

Historia

- Es una interfaz estandarizada para la transmisión de datos entre distintos dispositivos del bus de la computadora.
- Fue creado por Larry Boucher y Dal Allan, y normalizado en 1986.
- Normalmente es más caro que los discos ATA y por eso sólo se extendió su uso en servidores y equipos de video/audio.
- Se usó en la Commodore Amiga y equipos de Apple y Sun.
- Recientemente Apple ya lo substituyo por completo por IDE y Sun por SATA.
- Utiliza CCS (*Command Common Set*), un conjunto de comandos para acceder a los dispositivos que los hacen más o menos compatibles.

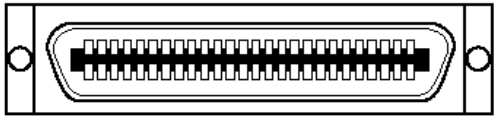
SCSI

Conectores

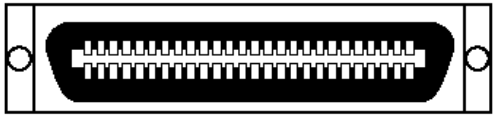


DB-25, Male External

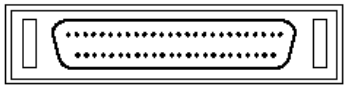
DB-25, Female External



Low-Density, 50-pin, Male External



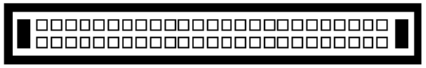
Low-Density, 50-pin, Female External



High-Density, 50-pin, Male External

High-Density, 50-pin, Female External

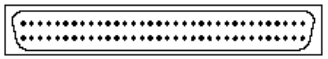
Low-Density, 50-pin, Male Internal



Low-Density, 50-pin, Female Internal

High-Density, 68-pin, Male External

High-Density, 68-pin, Female External



High-Density, 68-pin, Male Internal

High-Density, 68-pin, Female Internal

SCSI

Tipos

SCSI 1

Bus de 8 bits, velocidad de 5MBps, 50 pins, 6mts. de long. max del cable, hasta 7 dispositivos conectados.

SCSI 2

Existes 2:

Fast.- Bus de 8 bits, vel. 10MBps, 50 pins, 3mts, 7 disp.

Wide.- Bus de 16 bits, 68 pins, 3 mts., 16 disp.

SCSI 3

3.1 SPI Existen: Ultra,, Ultra Wide, Ultra 2 (**80MBps**).

3.2 FireWire (IEEE 1394)

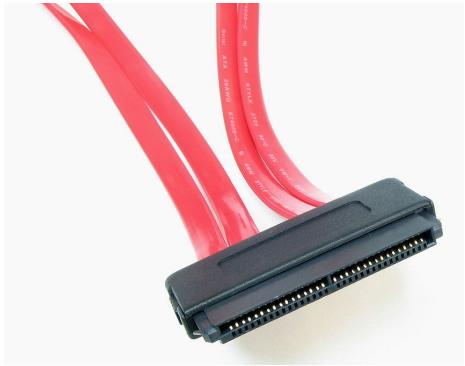
3.3 Serial Storage Architecture

3.4 Fibre Channel Arbitrated Loop (fibra **10km** o coaxial **24mts**)

SAS

Serial Atached SCSI

- It is a point-to-point serial protocol that moves data to and from computer storage devices such as hard drives and tape drives.
- SAS replaces the older Parallel SCSI (Small Computer System Interface, usually pronounced "scuzzy") bus technology that first appeared in the mid-1980s.
- SAS, like its predecessor, uses the standard SCSI command set. SAS offers backward compatibility with SATA, versions 2 and later.
- This allows for SATA drives to be connected to SAS backplanes. The reverse, connecting SAS drives to SATA backplanes, is not possible.



Tape

Cintas magnéticas

Historia

Formatos

Cintas magnéticas

Tape

- Es un tipo de medio o soporte de almacenamiento de datos que se graba en pistas sobre una banda plástica con un material magnetizado, generalmente óxido de hierro o algún cromato.
- Hay diferentes tipos de cintas, tanto en sus medidas físicas como en su constitución química, así como diferentes formatos de grabación, especializados en el tipo de información que se quiere grabar.
- En 1949 Edvac fue la primera computadora que empleó la cinta magnética como medio de almacenamiento de datos, fue de las primeras computadoras que procesaba con sistema binario en lugar de decimal y un lector grabador de cinta magnética.
- Univac en 1955 fue de las primeras computadoras que solucionó la necesidad de convertir grandes cantidades de información previamente almacenada en tarjetas.

Cintas magnéticas

Tape

- Linear Tape-Open (**LTO**) es una tecnología de cinta magnética de almacenamiento de datos, desarrollada originalmente a finales de 1990 como alternativa de estándares abiertos a los formatos de cinta magnética patentada que estaban disponibles en ese momento.
- Hewlett-Packard, IBM y Seagate iniciaron el Consorcio LTO.

Cintas magnéticas

Tape

Formatos LTO

- LTO-1: 200 GB
- LTO-2: 400 GB
- LTO-3: 800 GB
- LTO-4: 1,6 TB
- LTO-5: 3 TB Velocidad de transferencia de hasta 1TB/hora

Existen equipos que virtualizan cintas.

Algunas marcas son Ultrium.

RAID

Redundant Array of Independet Disks

Arreglos de discos

Historia

Formatos

Arreglos de discos

RAID (Redundant Array of Independet Disks)

Es una tecnología de almacenamiento que combina multiples discos en una sólo unidad lógica.

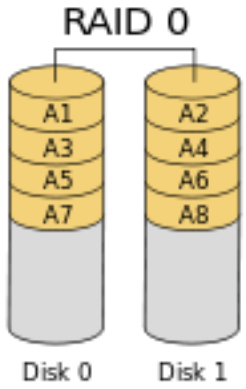
1. Creado en 1978 por D. Patterson, G.A. Gibson, y R. Katz de la Univ. de California en Berkeley. Publicado por ACM en SIGMOD (Management Of Data).
2. Los datos son distribuidos en varios discos de varias formas que reciben el nombre de RAID número .
3. El número refleja el nivel de redundancia (replicación) y de rendimiento.
4. Los niveles de RAID y los formatos de datos son estandarizados por la SNIA (Storage Network Industry Association).

Arreglos de discos

Niveles de RAID

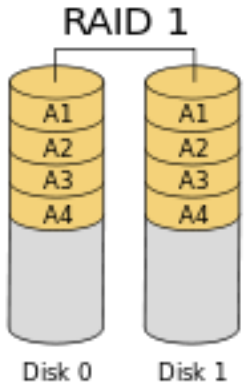
RAID 0

Se usan 2 o más discos que se agrupan para dar la apariencia de que son uno sólo. Los datos se almacenan de manera distribuida (**striping**) en los discos.



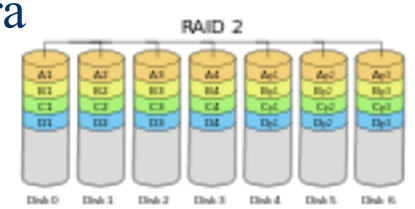
RAID 1

Discos espejo.- Los datos son escritos de manera identica en los discos. Lectura y escritura requiere el doble de tiempo.



RAID 2

No se usa.- Los datos son escritos en varios discos (striping) pero a nivel de **Bits**. Además utiliza información de **paridad** para detectar errores.



Arreglos de discos

Niveles de RAID

RAID 3

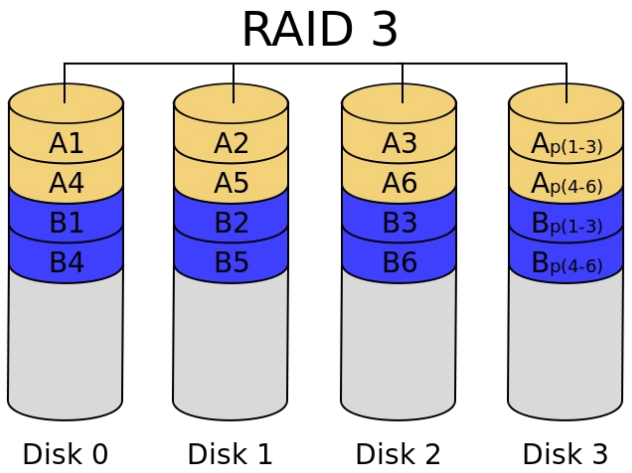
No se usa. Hace stripping a nivel de **Bytes**.
Calcula paridad colocandola exclusivamente en un disco.

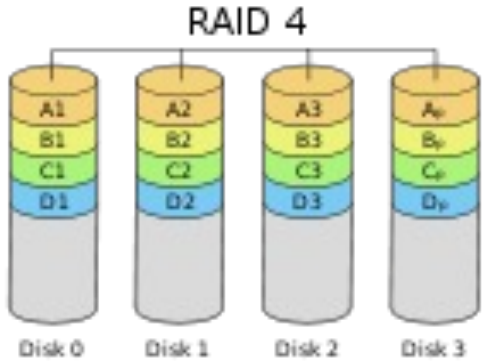
RAID 4

Casi no se usa. Hace stripping a nivel de **Bloques**.
Calcula paridad colocada exclusivamente en un disco.

RAID 5

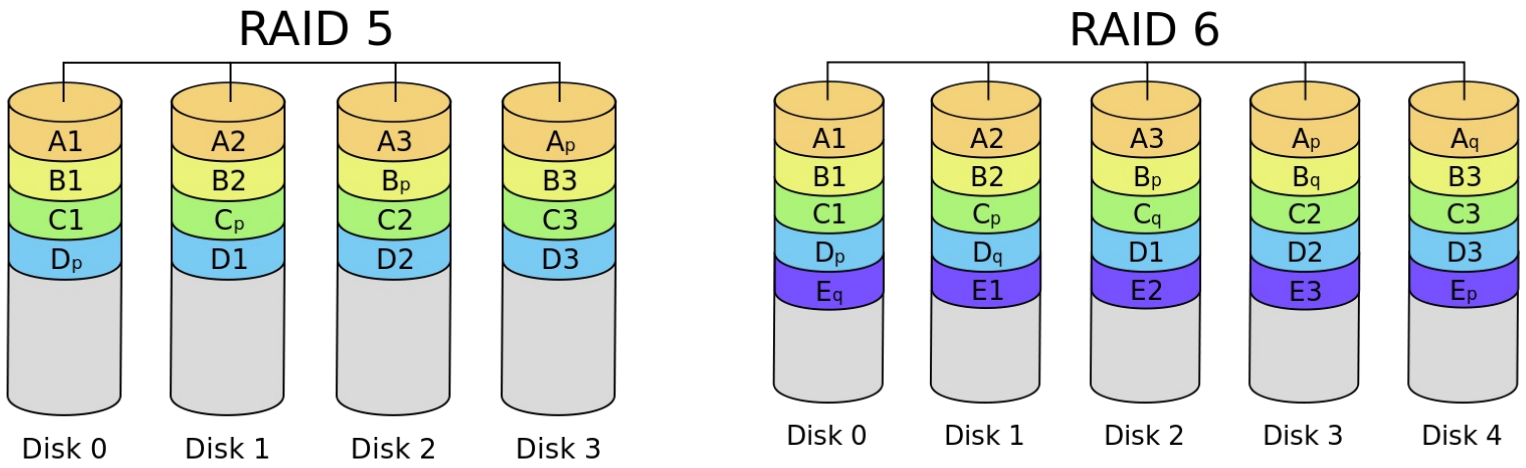
Más usado. Hace striping de bloques y de paridad.
Requiere como mínimo 3 discos y puede operar con 2.





Arreglos de discos

Niveles de RAID



RAID 6

Hace stripping de datos (bloques) y de paridad.

Calcula doble nivel de paridad.

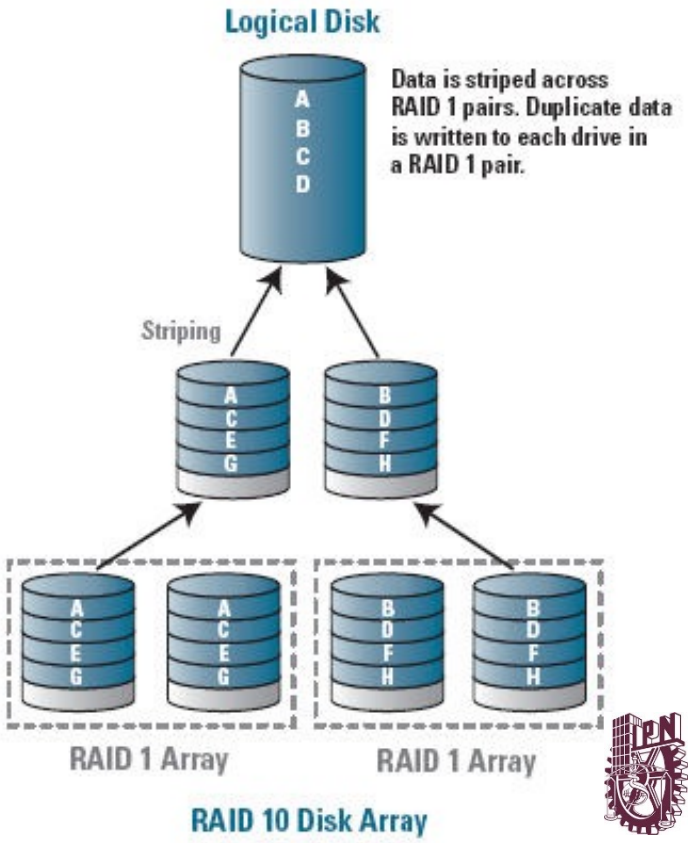
Nota importante.- El manejo del striping y de la paridad se hace por hardware (lo usal y más rápido, controladora) pero también se puede con software.

Arreglos de discos

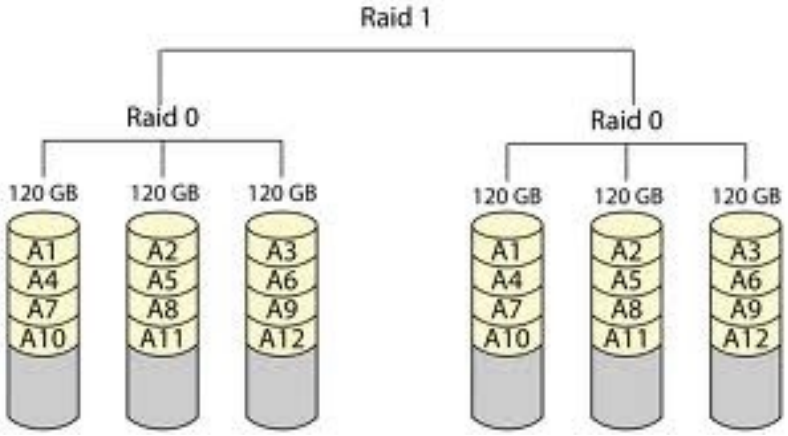
RAIDs anidados o hibridos

RAID 1+0 (o 10)

- Se construye como:
  - El primer nivel es RAID 1.
  - El segundo nivel es RAID 0.
- Significa que se tienen:
  - 2 RAID 1 combinados en un RAID 0.
  - Al menos 4 discos.



RAID 01



Arreglos de discos

RAIDs anidados o hibridos

Linux

- En Linux se pueden crear RAIDs de manera virtual utilizando el comando mdadm.
- El comando mdadm permite crear RAIDs 0, 1, 4, 5, 6 entre otras funcionalidades.
- Antes del comando mdadm se creo el paquete raidtools. El cual es más básico pero aun tiene comandos útiles.
- Como es habitual en Linux, la información de los RAIDs se puede consultar a través de la información colocada en /proc/mdstat.

NAS/SAN

Arreglos en red

Historia

Formatos

NAS

Network Attached Storage

Marcas conocidas QNAP

Almacenamiento conectado en red, **Network Attached Storage** (NAS)

- Es el nombre dado a una tecnología de almacenamiento dedicada a compartir la capacidad de almacenamiento de un computador (servidor) con computadoras personales o servidores clientes.
- Se hace a través de una red (normalmente TCP/IP), haciendo uso de un sistema operativo optimizado
- Se da acceso con los protocolos CIFS, NFS, FTP o TFTP.
- Los protocolos de comunicaciones NAS están basados en archivos por lo que el cliente solicita el archivo completo al servidor y lo maneja localmente.
- Están orientados a manipular una gran cantidad de pequeños archivos.
- Los protocolos usados son protocolos de compartición de archivos como Network File System (**NFS**) o Microsoft Common Internet File System (**CIFS**).

SAN

Storage Area Network

Una **red de área de almacenamiento**, en inglés *Storage Area Network* (**SAN**), es una red de almacenamiento integral. Se trata de una arquitectura completa que agrupa los siguientes elementos:

- Una red de alta velocidad de canal de fibra o iSCSI.
- Un equipo de interconexión dedicado (conmutadores, puentes, etc).
- Elementos de almacenamiento de red (discos duros).

Una SAN es una red dedicada al almacenamiento que está conectada a las redes de comunicación de una compañía.

SAN

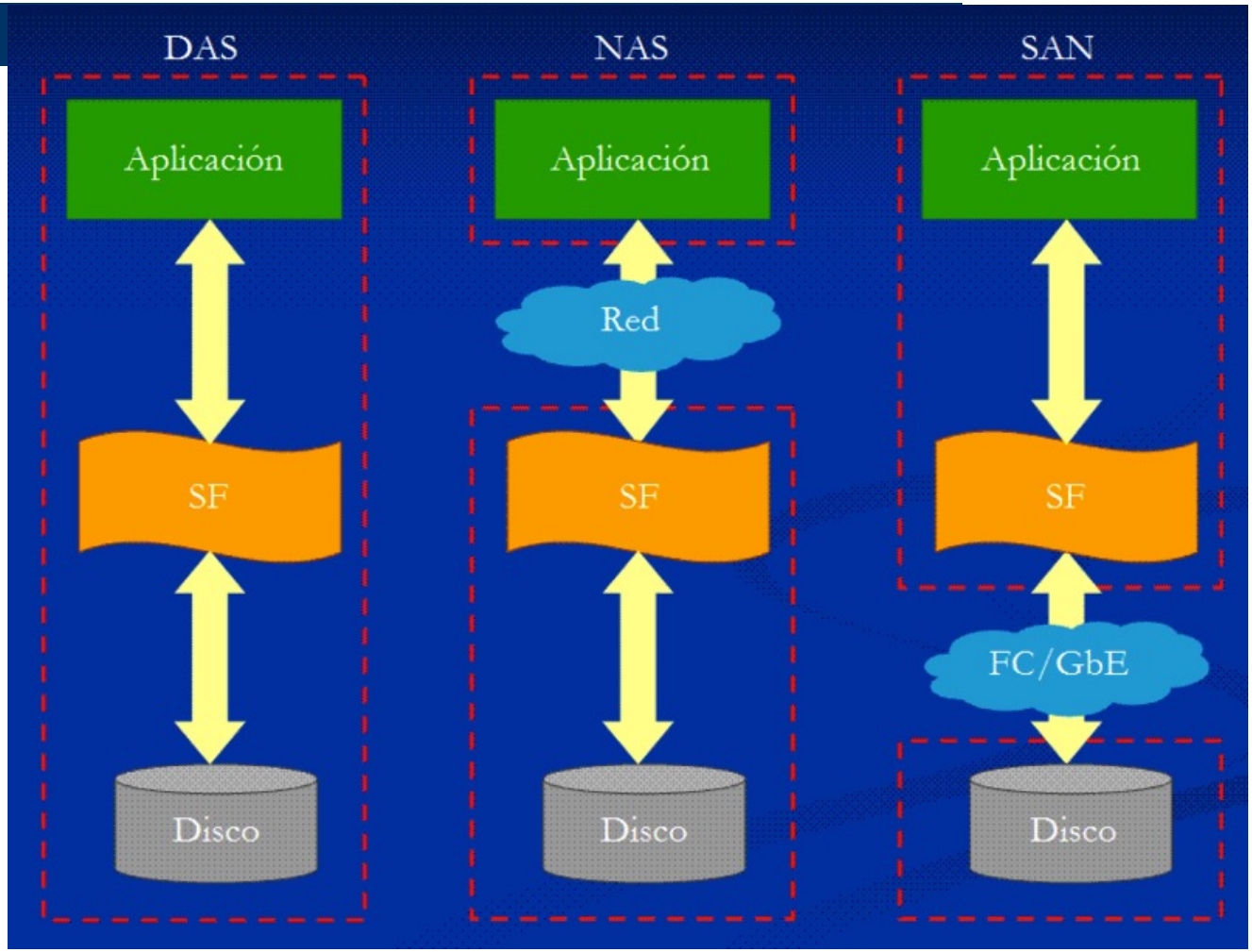
Storage Area Network

- El rendimiento de la SAN está directamente relacionado con el tipo de red que se utiliza.
- En el caso de una red de canal de fibra, el ancho de banda es de aproximadamente 100 megabytes/segundo (1.000 megabits/segundo) y se puede extender aumentando la cantidad de conexiones de acceso.
- La capacidad de una SAN se puede extender de manera casi ilimitada y puede alcanzar cientos y hasta miles de terabytes.
- Principalmente, está basada en tecnología fibre channel y más recientemente en iSCSI.
- Existen protocolos básicos usados en una red de área de almacenamiento:
  - FC-AL
  - FC-SW
  - SCSI
  - FCoE

NAS vs SAN

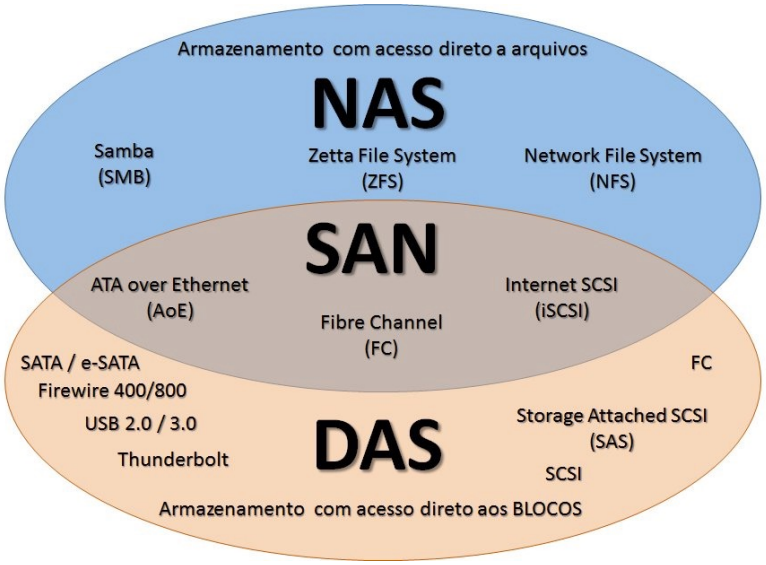
Cada uno en su ambito

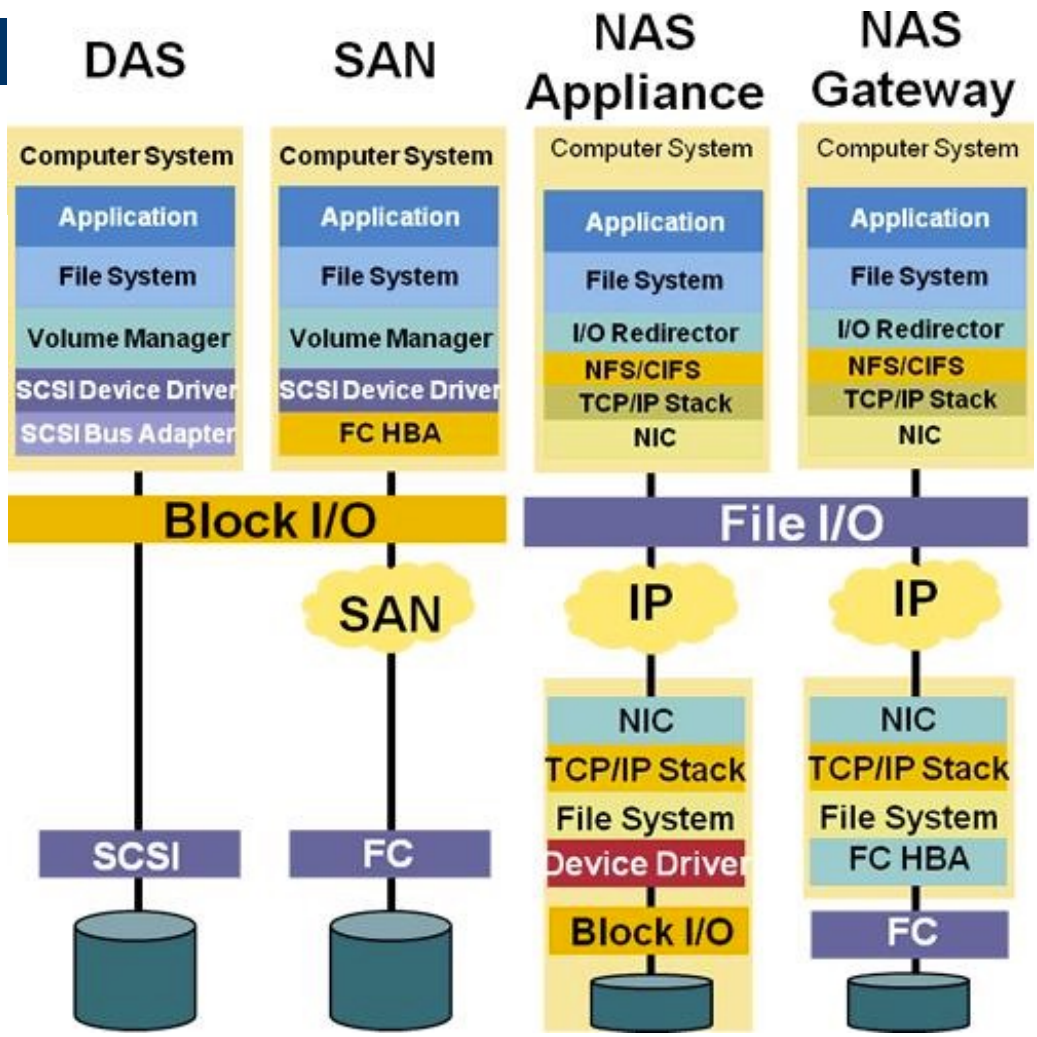
El DAS es el acceso tradicional que se hace directamente (local) al sistema de archivos y al disco.



NAS vs SAN

Cada uno en su ámbito



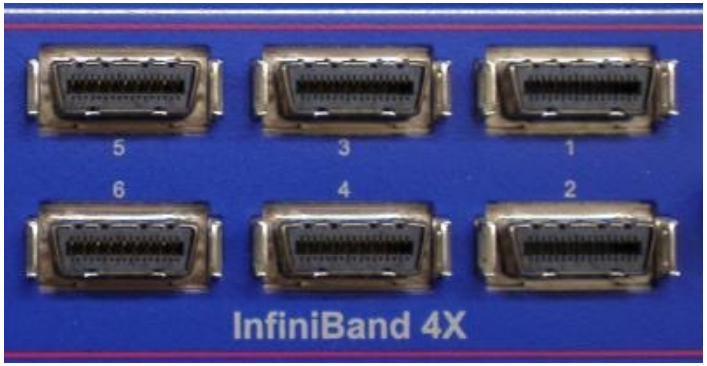


Protocolos de acceso a discos

Introducción

Lista de tecnologías

1. ATA over Etherenet (AoE)
2. Fibre Channel Protocol (FCP) SCSI over Fibre Channel.
3. Fibre Channel over Ethernet (FCoE)
4. HyperSCSI, SCSI over Ethernet
5. iSCSI, SCSI over TCP/IP
6. iSCSI Extension for RDMA (iSER), iSCSI over InfiniBand (switch).



Temas avanzados

Introducción

Investigar IOPS (Input/Ouput Per Second)

- IOPS es una medida común de rendimiento usada en los benchmarks de dispositivos de almacenamiento.
- Como cualquier benchmark, el número de IOPS publicado por los fabricantes no garantiza un rendimiento de la aplicación real ya que se hacen las pruebas en condiciones de laboratorio.

Temas avanzados

Benchmarks tools

IOmeter

IOzone

HD Tune

SQLIO

FileBench

Bonnie++

The end

Contacto

Raúl Acosta Bermejo

http:www.cic.ipn.mx
http://www.ciseg.cic.ipn.mx/

racostab@ipn.mx
racosta@cic.ipn.mx

57-29-60-00
Ext. 56652