

Fundamentos de Sistemas de Operação

Unix Windows NT Netware Mac OS DOS/V/S Vax/VMS
Linux Solaris HP/UX AIX Mach Chorus

Gestão de Ficheiros, Tópicos Avançados
Armazenamento RAID

Estruturas RAID

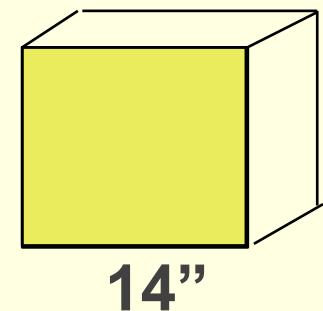
- RAID: Patterson, D. et al. A Case for Redundant Arrays of Inexpensive Disks (RAID). Proceedings of the 1989 ACM-SIGMOD International Conference on the Management of Data, ACM, 1989, pp. 109-116.
- Hoje em dia, a sigla é a mesma, mas diz-se Redundant Array of **Independent** Disks – Armário Redundante de Discos independentes.
- Foram identificados 6 níveis de estruturas RAID (de 0 a 5; recentemente apareceu o RAID-6)
- Os objectivos principais são: elevado desempenho e fiabilidade, obtidos usando múltiplos discos “em paralelo”, tolerando a falha (excepto RAID-0) de um ou mais discos sem perda de dados

Disk Arrays: vantagens

Famílias de Discos

Convencional:

4 tipos de discos

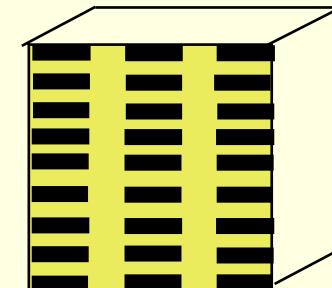
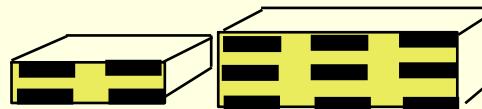


“Low End” → “High End”

Disk Array:

1 só tipo de discos

3.5" →



Estruturas RAID - I

Substituir um disco grande por muitos discos pequenos! (Patterson, 1988)

	IBM 3390 (K)	IBM 3.5" 0061	70 x 3.5"
Capacidade	20 GBytes	320 MBytes	23 GBytes
Volume	2,7 m³	2,8 dm³	2,7 m³
Potência	3 KW	11 W	1 KW
Ritmo de transf	15 MB/s	1.5 MB/s	120 MB/s
Ritmo de ops I/O	600 I/O/s	55 I/O/s	3900 I/O/s
MTTF	250 K Hrs	50 K Hrs	??? Hrs
Custo	\$250K	\$2K	\$150K

Estruturas RAID - II

*Potencial para: grandes taxas de transferência,
de economia de: potência, volume, e custo!!!*

	IBM 3390 (K)	x70
Capacidade	20 GBytes	23 GBytes
Volume	2,7 m ³	2,7 m ³
Potência	3 KW	1 KW
Ritmo de transf	15 MB/s	120 MB/s
Ritmo de ops I/O	600 I/O/s	3900 I/O/s
MTTF	250 Khrs	??? Hrs
Custo	\$250K	\$150K

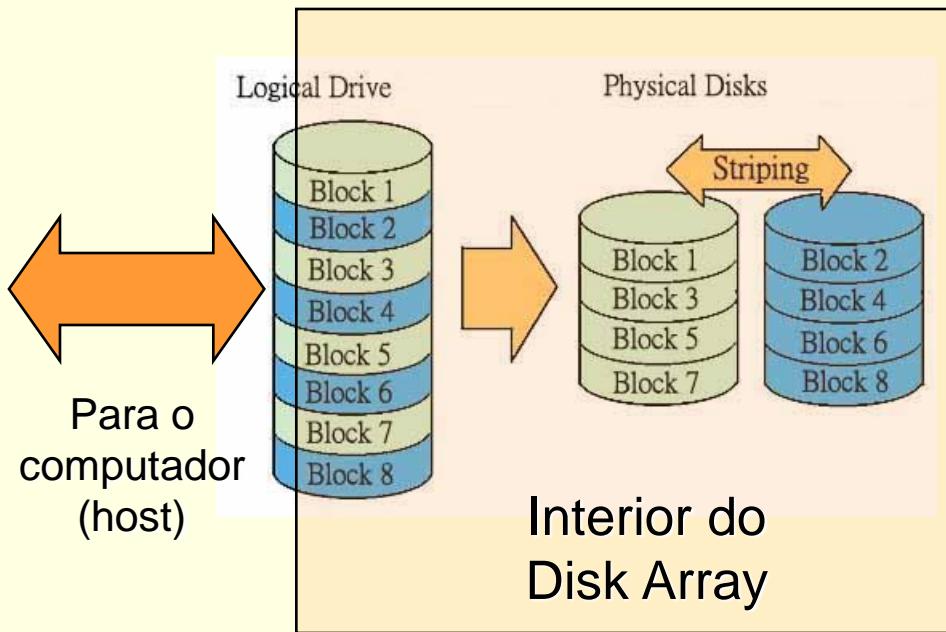
Um disco moderno



- Simplesmente... um disco
 - Seagate ST3146855FC ($\approx \$250$)
15K rpm, 16MB cache
Random Seek (R/W): 3.5 / 4 ms
Avg. Track-to-Track (R/W): 0.2 / 0.4 ms
Transfer Rate (Inter., min/max): 96/160 MB/s
Sust. transf Rate (min/max): 73/125 MB/s
Transfer Rate (External): 800 MB/s

RAID-0

□ RAID-0 com 2 discos (descrição simplificada)



Quando o host transfere (lê ou escreve) um “registo” os 2 discos são acedidos.

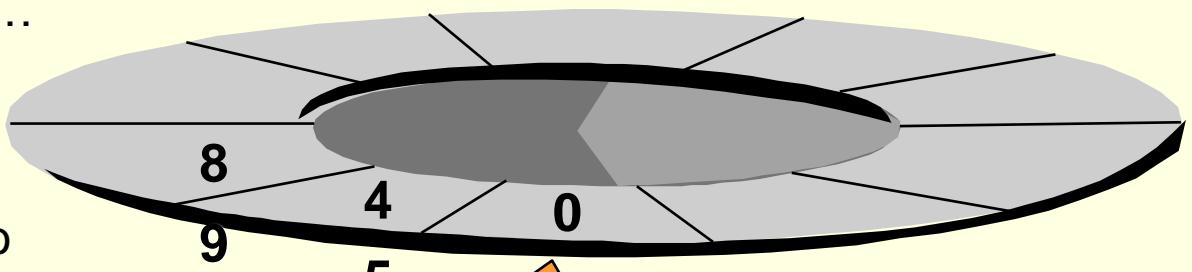
A taxa de transferência é 2 vezes superior à do disco individual.

Em caso de falha de um dos discos o “disco lógico” já “não funciona”, perdem-se os dados.

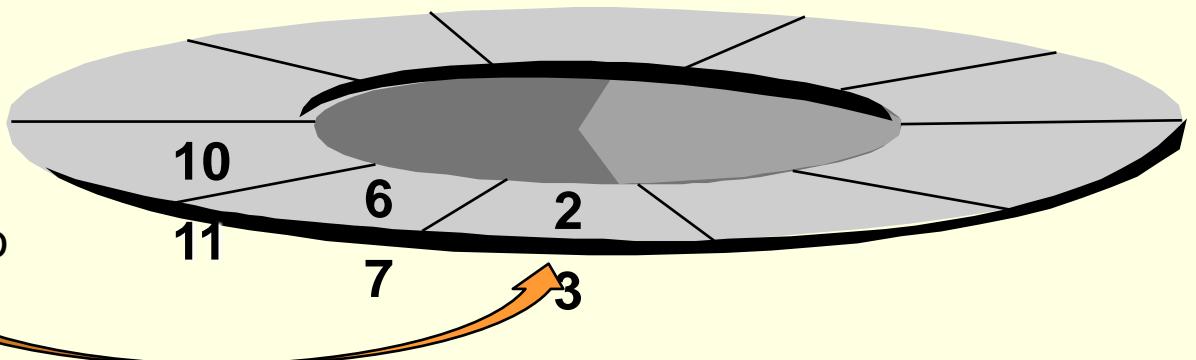
O espaço útil é a soma do espaço dos discos físicos.

Disco individual: numeração dos blocos

- Num disco individual como este, com 4 superfícies, os blocos são numerados assim, de forma que dumas só vez se conseguem transferir 4 blocos...



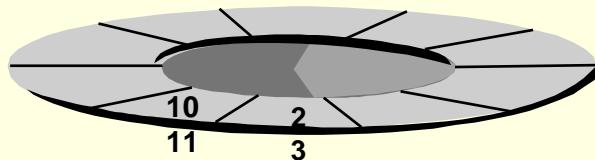
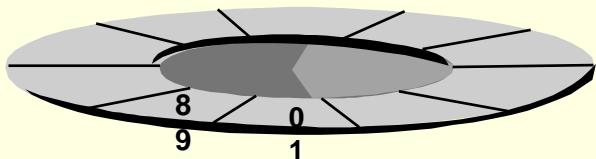
Na “parte” de baixo
fica o bloco 1



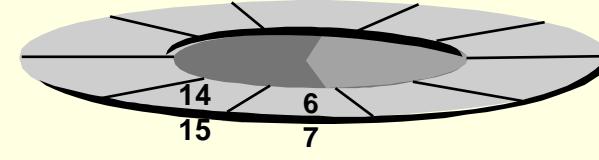
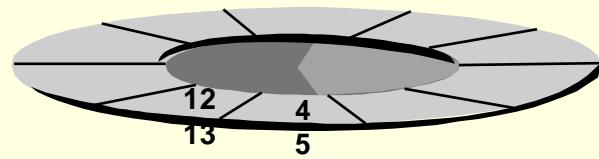
Na “parte” de baixo
fica o bloco 3

RAID-0: renumeração dos blocos

- Num RAID-0 com 2 dos mesmos discos, os blocos são renumerados assim, de forma que dumas só vez se conseguem transferir 8 blocos: do disco 1 os blocos 0..3 e do disco 2 os blocos 4..7



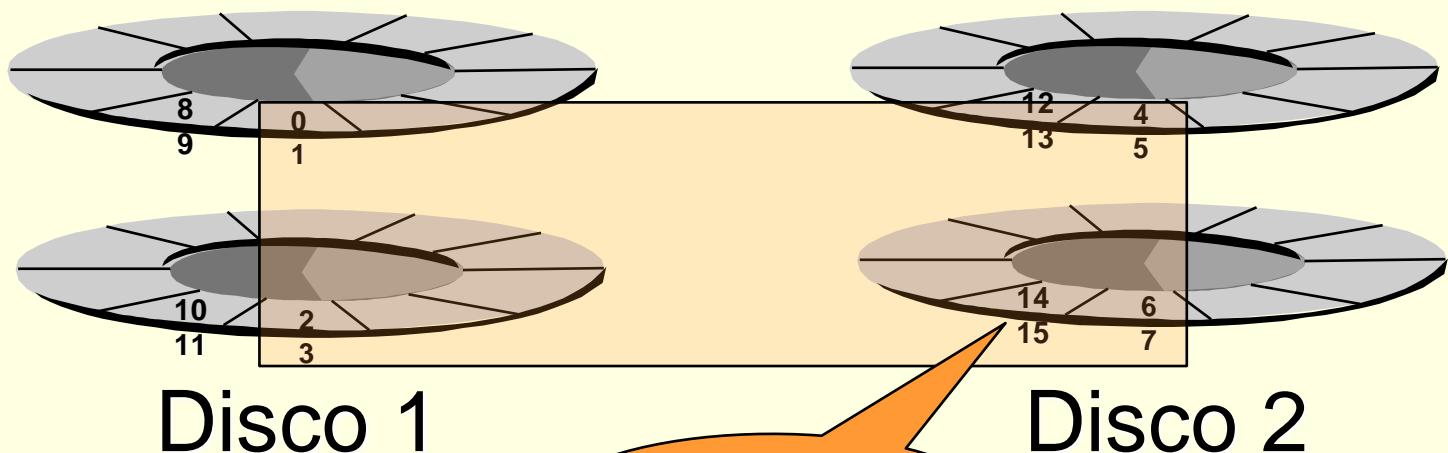
Disco 1



Disco 2

RAID-0: stripes

- A unidade de transferência é a faixa (*stripe*), que neste exemplo tem 8 blocos – tem uma profundidade (*depth*) 4 (são 4 blocos em cada disco) e uma largura (*width*) 2 (são 2 discos)
 - **Stripe size = Stripe Depth x Stripe Width**

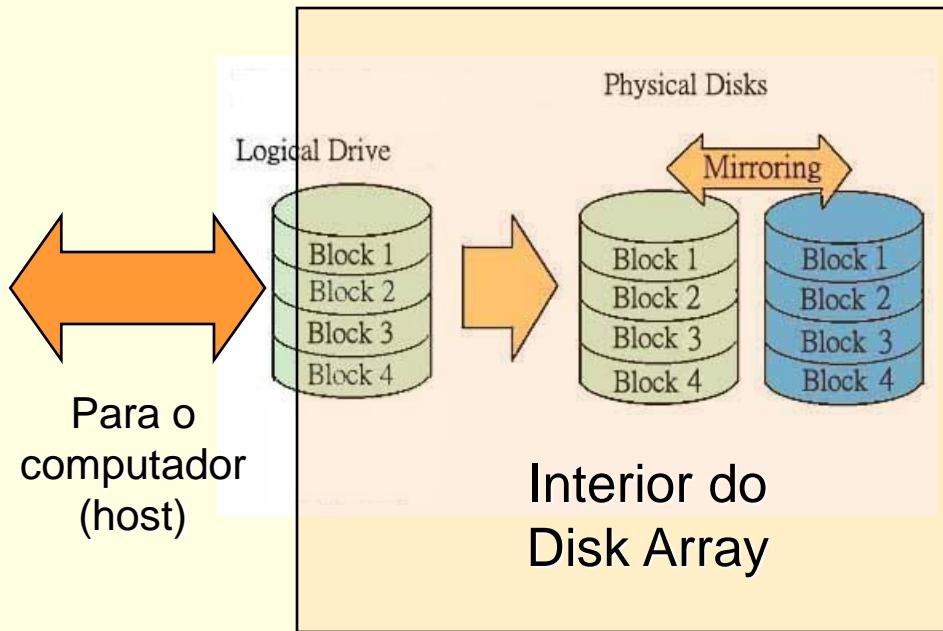


RAID-0

- Os “caixotes de discos” (*disk arrays*) empresariais são construídos para ter desempenho muito elevado,
 - Os discos giram sincronizados,
 - No RAID-0, as cabeças dos diferentes discos movem-se alinhadas
- Consequências, num RAID-0 com N discos:
 - A largura de banda das leituras e escritas feitas a um volume RAID-0 é aprox. igual a $N \times LB$ de um disco individual
 - A latência de um volume RAID-0 é aprox. a mesma de um só disco

RAID-1 (mirror)

□ RAID-1 com 2 discos (descrição simplificada)



Quando o host lê um “registro”, os 2 discos podem ser acedidos; quando escreve, os mesmos dados são escritos nos 2 discos.

A taxa de transferência pode de ser 2 vezes superior à do disco individual em leitura, mas é igual em escrita.

Em caso de falha de um dos discos o “disco lógico” “continua a funcionar”.

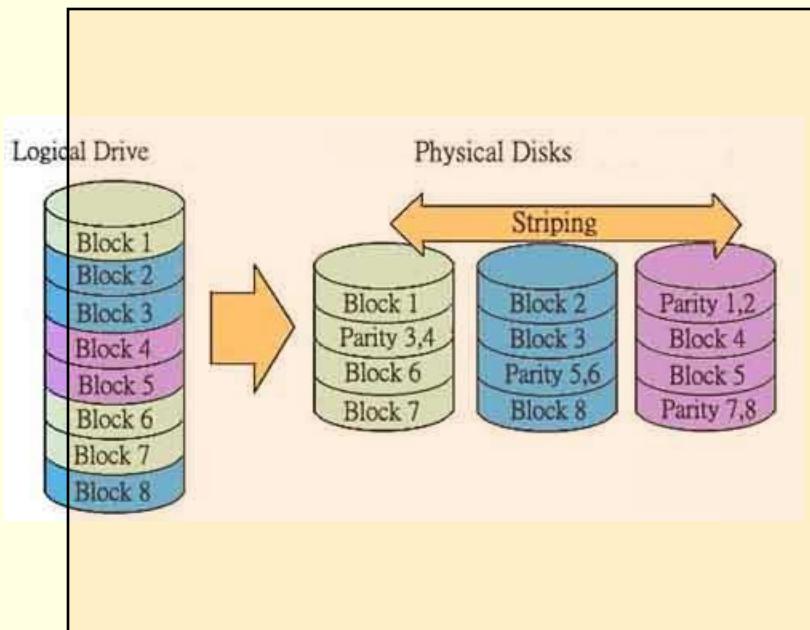
O “espaço útil” é metade do espaço físico...

RAID-1

- Num volume RAID-1 (em *arrays* empresariais),
 - Os discos giram sincronizados,
 - As cabeças dos diferentes discos podem por vezes mover-se sem estar alinhadas, nas leituras
- Consequências, num RAID-1 :
 - A largura de banda das leituras feitas a um volume RAID-1 é aprox. igual a 2 x LB de um disco individual
 - A LB nas escritas é aprox. a mesma de um só disco
 - A latência de um volume RAID-1 é aprox. a mesma de um só disco

RAID-5

□ RAID-5 com 3 discos (descrição simplificada)



Quando o host lê um “registo”, os 3 discos podem ser acedidos; quando escreve, lê-se os blocos que estão na mesma stripe, juntase o(s) novo(s) bloco(s) e recalcula-se a paridade.

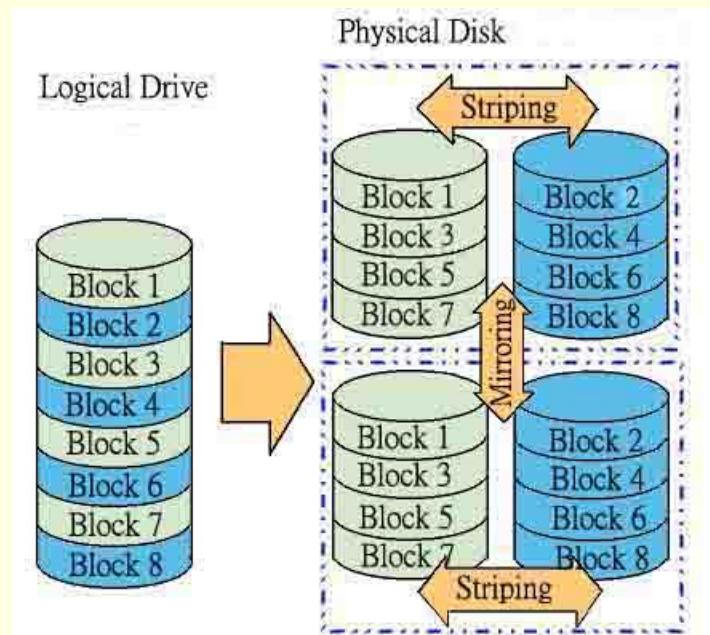
A taxa de transferência pode ser 2 vezes superior à do disco individual em leitura, mas é bastante pior em escrita.

Em caso de falha de um dos discos o “disco lógico” “continua a funcionar”.

“Perde-se” o espaço de um disco...

RAID-0/1

- *RAID-0/1 com 4 discos*



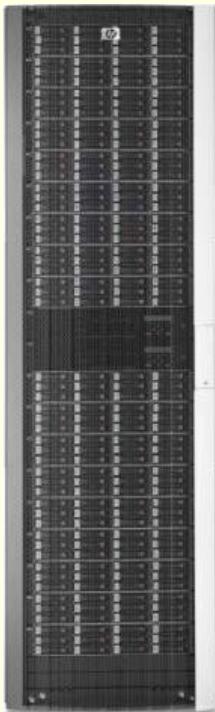
Fundamentos de Sistemas de Operação

Unix Windows NT Netware Mac OS DOS/V/VS Vax/VMS
Linux Solaris HP/UX AIX Mach Chorus

Alguma informação adicional
(só para conhecimento, não para avaliação)

Disk Arrays modernos (1)

HP EVA 4400 – gama média (modelo mais baixo)



Número max. de discos: 96

Número max.@velocidade de portas: 4@4Gbs

Cache máx.: 4 GB

Tecnologia/velocidade/capacidade dos discos:

- SSD, 72 GB
- FC 15k e 10krpm, 600 GB => 58 TB
- FATA 7.2k rpm, 1TB

RAID: 0, 1, 0+1, 5, 0+5, 6

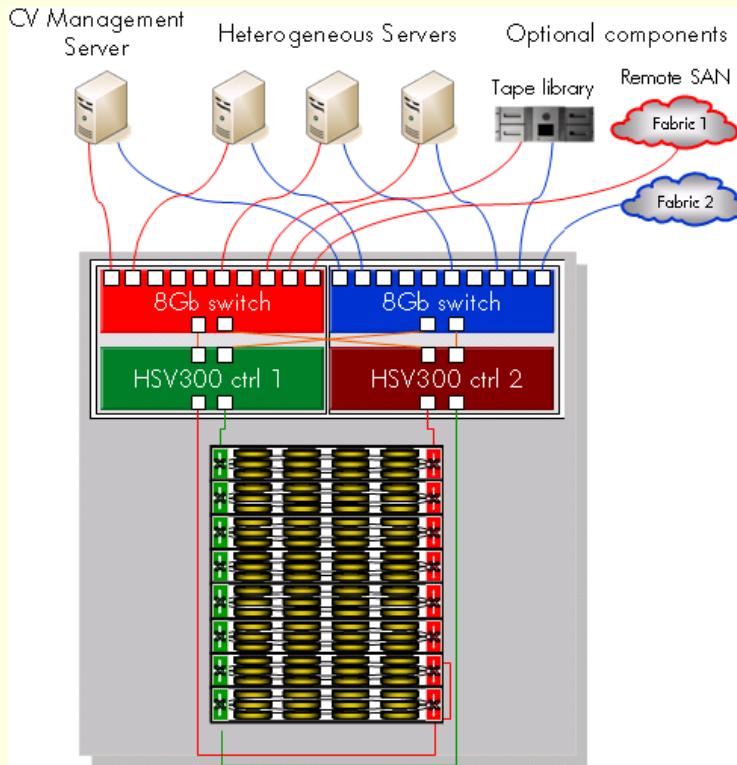
IOPS: cache: 141k, disco: 26k (R), 7.2k (W)

BW acesso sequencial:

- cache: 1380 MB/s
- disco: 775 MB/s (R), 550 MB/s (W)

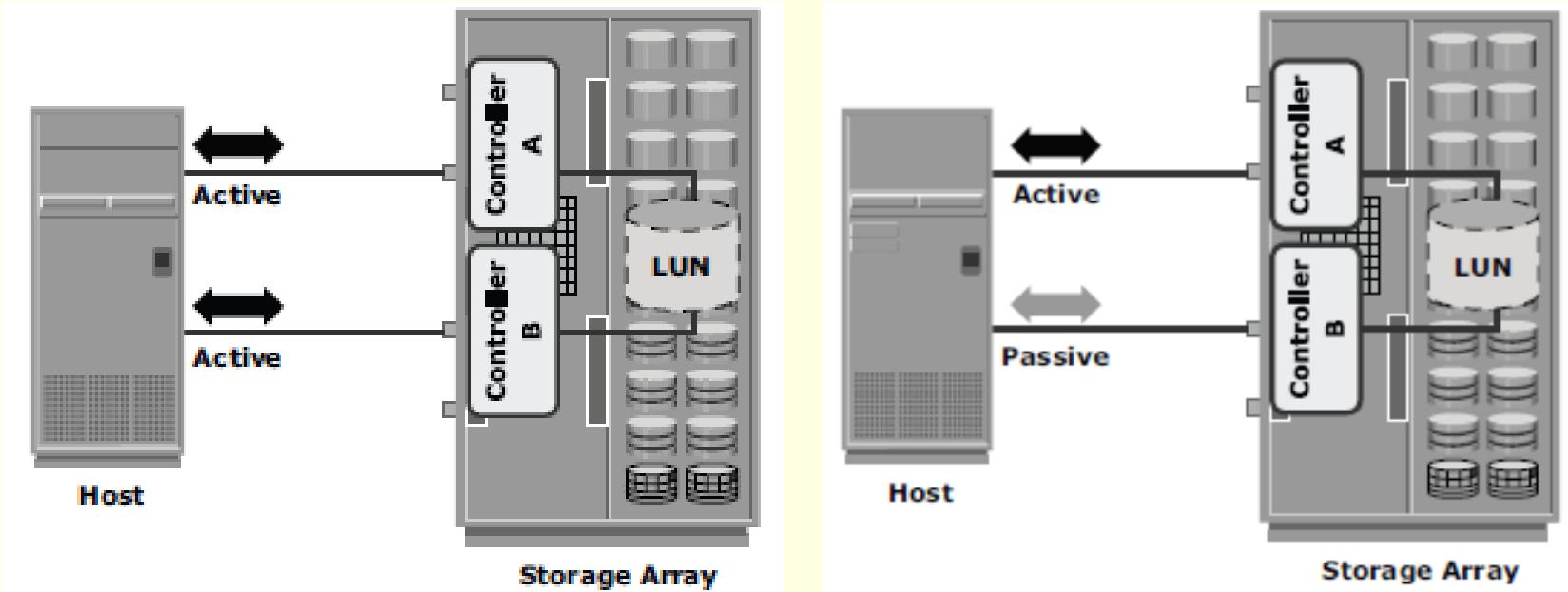
Disk Arrays modernos (2)

HP EVA 4400 – arquitectura interna



4 portas FC-SW
para o exterior
4 portas FC-AL
para o interior

Alta disponibilidade (HA)



SAN – Storage Area Network

Unix Windows NT Netware Macos DOS/VMS Vax/VMS
Linux Solaris HP/UX AIX Mach Chorus

