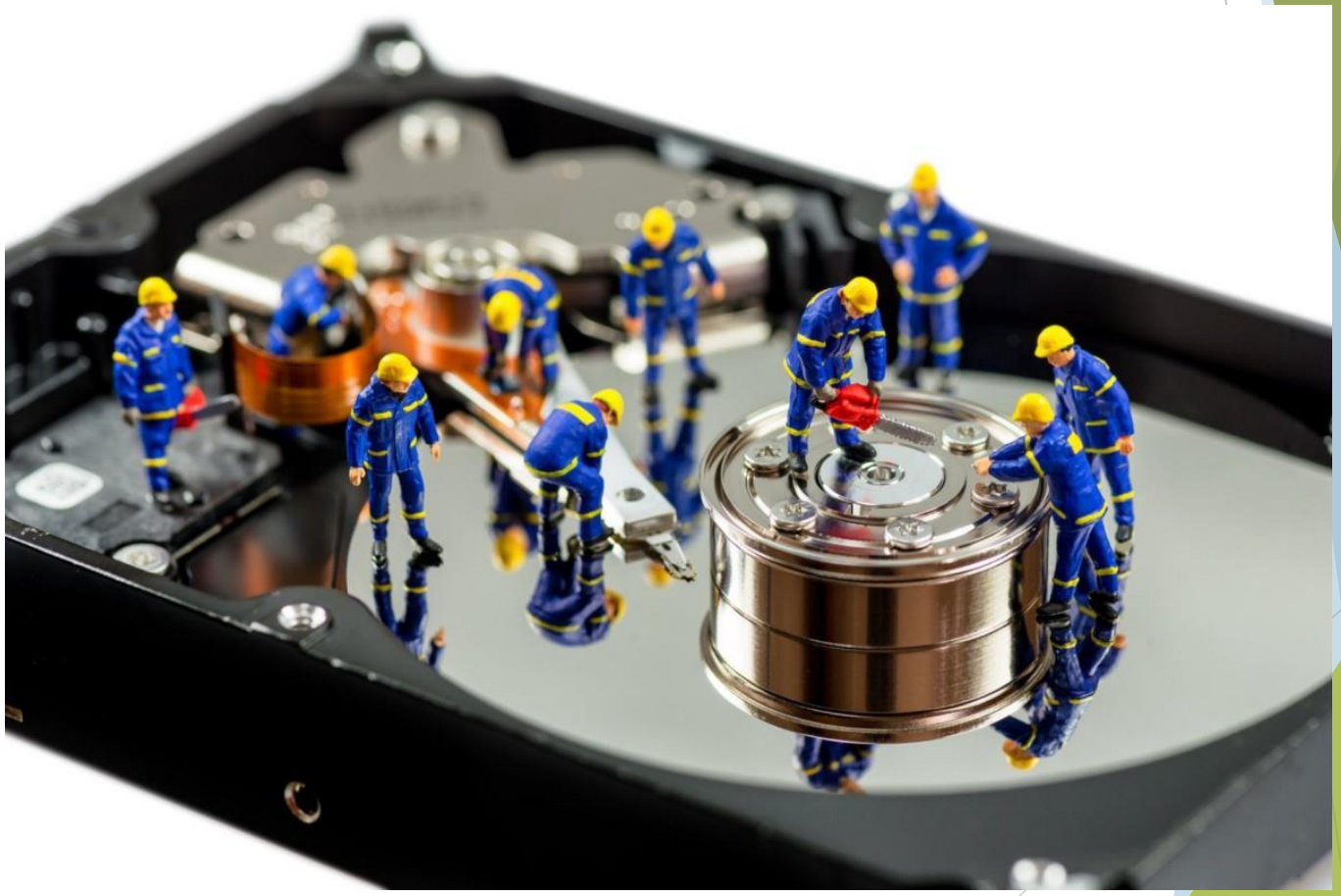
116327 -Organização de Arquivos



Organização de Arquivos

Disciplina: 116327

Prof. Oscar Fernando Gaidos Rosero

Universidade de Brasília – UnB
Instituto de Ciências Exatas – IE
Departamento de Ciência da Computação – CIC

Organização de Arquivos
Aula 5

M.Sc. Oscar Gaidos

Universidade de Brasília (UnB)

Sumário

Armazenagem Secundária

1. Organização dos discos magnéticos
2. Tempo de acesso a disco
3. Fitas magnéticas (aplicações, tamanho do bloco, velocidade de transmissão)
4. Outros dispositivos
5. Transferência dos dados entre o programa e dispositivo de armazenagem secundária
6. Conceitos e Técnicas de gerenciamento de buffers

UnB

Formatação em Discos Setorizados

Formatação Física (comando via BIOS)

- Divisão dos discos em trilhas, cilindros e setores.
- É chamada formatação de baixo nível.
- Coloca as marcas no início e final de cada setor e os intervalos entre trilhas e setores.
- Setores danificados são marcados.

Formatação lógica (comando via S.O.)

- Consiste em criar a estrutura de controle para o sistema de arquivo* utilizado pelo sistema operacional (FAT16, FAT32, NTFS, EXT2, EXT3, etc).
- Grava informações no disco de forma que arquivos possam ser lidos, escritos e localizados.

---

*Um sistema de arquivos é um conjunto de estruturas lógicas e de rotinas que permitem ao SO controlar o acesso ao disco rígido. Diferentes sistemas operacionais usam diferentes sistemas de arquivos.

UnB

Custo de acesso a Disco

O custo em termos de tempo, de um acesso a disco é uma combinação dos tempos de posicionamento (seek) e rotação do disco, e do tempo de transferência dos dados.

Tempo de Seek

- Tempo necessário para mover o braço de acesso para o cilindro correto. O tempo depende da distancia a ser percorrida pelo braço.

Rotacional Delay (latência)

- Tempo que leva para o disco rotacionar tal que o setor desejado esteja sob a cabeça de leitura/escrita.

Tempo de transferência

- Tempo necessário para um byte ser lido da superfície do disco e transferido para o buffer interno do controlador

UnB

Custo de acesso a disco

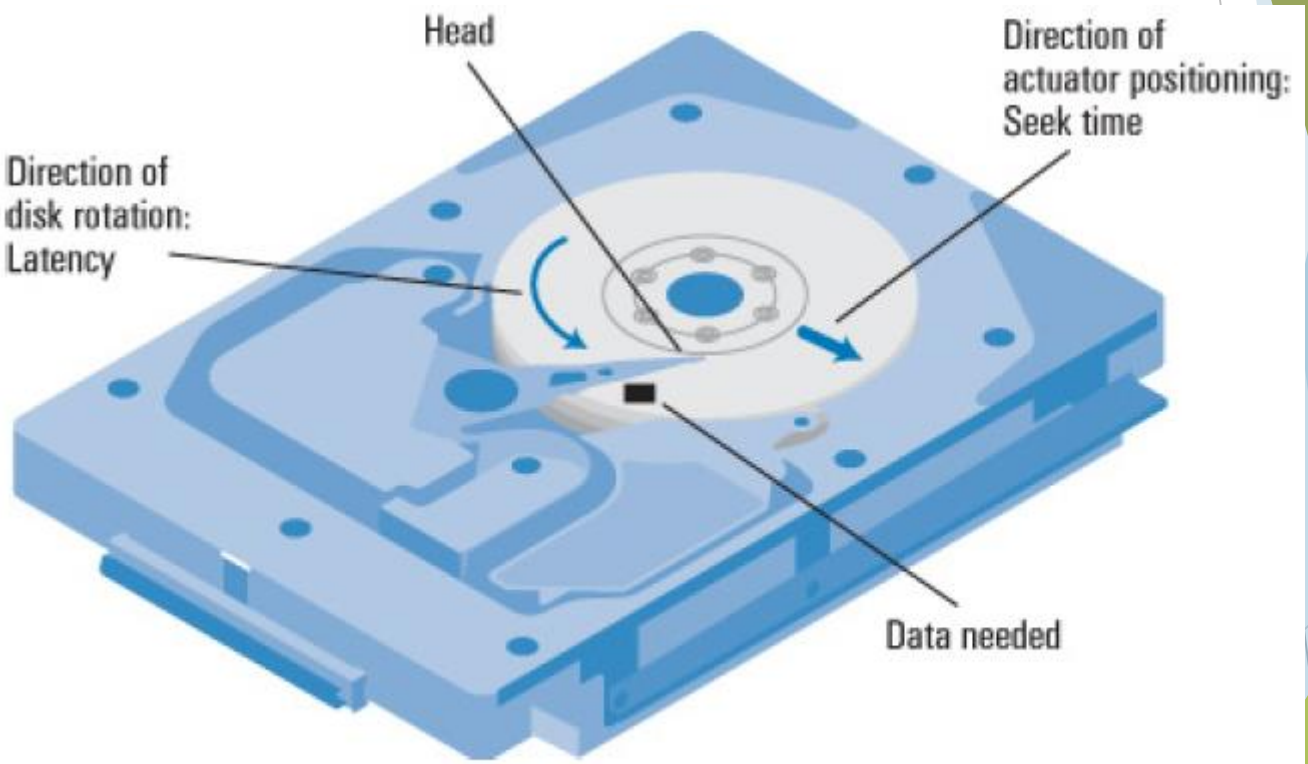


Figura: Custos de acesso ao disco

UnB

Custo de acesso a disco: Seeking

- O seeking pode ser muito caro (custo de acesso elevado!).
- Geralmente, é impossível saber exatamente quantas trilhas serão percorridas em cada busca e o que se faz é tentar determinar o tempo médio de busca necessário para uma certa operação em arquivo, assumindo que as posições inicias e finais para cada acesso são aleatórias.
- Foi descoberto, através de experiências empíricas, que uma busca percorre, em média, 1/3 do total de trilhas, sendo que o tempo gasto para percorrer esse numero de trilhas tem sido geralmente usado pelos fabricantes como indicador do tempo meio de busca*.

---

*Em 1991 o tempo médio de busca fornecido pelos fabricantes variava entre 10 e 40 milissegundos. Hoje esse tempo está entre 5 e 13 ms.

UnB

Custo de acesso a disco: Seeking

- O que ocorre quando um arquivo é acessado sequencialmente, e o arquivo está armazenado em vários cilindros consecutivos?

- O que ocorre quando dois arquivos, localizados em extremos opostos do disco (um no cilindro mais externo, e outro no cilindro mais interno), são acessados alternadamente?

UnB

Custo de acesso a disco: Seeking

- O que ocorre quando um arquivo é acessado sequencialmente, e o arquivo está armazenado em vários cilindros consecutivos?

  O seek time é reduzido!

- O que ocorre quando dois arquivos, localizados em extremos opostos do disco (um no cilindro mais externo, e outro no cilindro mais interno), são acessados alternadamente?

UnB

Custo de acesso a disco: Seeking

- O que ocorre quando um arquivo é acessado sequencialmente, e o arquivo está armazenado em vários cilindros consecutivos?

  O seek time é reduzido!

- O que ocorre quando dois arquivos, localizados em extremos opostos do disco (um no cilindro mais externo, e outro no cilindro mais interno), são acessados alternadamente?

  O seek time é elevado!

  (Isso se torna mais crítico ainda em ambientes multiusuário)

UnB

Custo de acesso a disco: Rotacional Delay do Disco (Latência)

Latência

Atraso necessário para o disco rotar tal que o setor desejado esteja sob a cabeça de leitura/escrita.

- Como na busca, em geral considera-se a média
- Esse tempo é estimado como sendo metade do tempo gasto em uma rotação inteira do disco*.

Exemplo

Um arquivo ocupa duas ou mais trilhas, existem muitas trilhas disponíveis em um cilindro e o arquivo é escrito sequencialmente, com um único comando write.

Quando a primeira trilha esta completa, a escrita na segunda trilha pode ter início imediato, sem nenhum rotational delay. (A latência é virtualmente nula)

---

*Na prática, o atraso devido ao movimento de rotação, esse tempo deve ser menor que a média.

UnB

Custo de acesso a Disco: Tempo de transferência

Uma vez que o dado desejado esta sob a cabeça de leitura, ele pode ser transferido, e o tempo gasto nesta operação é dado pela fórmula:

$$\text{Tempo de Transferência} = \left(\frac{\text{Número de bytes transferidos}}{\text{Número de bytes por trilha}} \cdot \text{Tempo de rotação} \right)$$

Se o drive é setorizado, o tempo de transferência de um setor depende do número de setores por trilha.

Exemplo

Por exemplo, se existem 32 setores por trilha, o tempo necessário para transferir um setor seria:

$$\frac{1}{32} \cdot \text{Tempo de rotação}$$

UnB

Custo de acesso a disco

- Os diferentes modos de acesso a arquivo podem afetar os tempos de acesso.
- Vamos comparar os tempos de acesso a um dado arquivo considerando acesso sequencial e acesso aleatório.
- No primeiro caso, o máximo do arquivo é processado a cada acesso; no segundo caso, apenas um registro é acessado por vez. Em ambos os casos desejamos acessar todos os registros do arquivo.

UnB

Exemplo de especificações de HDs

Characteristic	Seagate Cheetah 9	Western Digital Caviar AC22100	Western Digital Caviar AC2850
Capacity	$9000MB$	$2100MB$	$850MB$
Seek time	$0.78ms$	$1ms$	$1ms$
Average seek time	$8ms$	$12ms$	$10ms$
Maximum seek time	$19ms$	$22ms$	$22ms$
Spindle speed	$10000rpm$	$5200rpm$	$5200rpm$
Average rotational delay	$3ms$	$6ms$	$6.6ms$
Maximum transfer rate	$6ms/track$	$12ms/track$	$13.3ms/track$
Bytes per sector	512	512	512
Sectors per track	170	63	63
Tracks per cylinder	16	16	16
Cylinders	526	4092	1654

Tabela: Especificações de HD's

UnB

Tempo de acesso

Os diferentes modos de acesso a arquivo podem afetar os tempos de acesso.

Vamos comparar os tempos de acesso a um dado arquivo considerando acesso sequencial e acesso aleatório.

- No primeiro caso, o máximo do arquivo é processado a cada acesso.
- No segundo caso, apenas um registro é acessado por vez. Em ambos os casos desejamos acessar todos os registros do arquivo.

UnB

Exemplo: Característica do disco

Características do disco

- Tempo mínimo de busca: $6ms$
- Tempo médio de busca: $13ms$
- Latência: $8.3ms$
- Taxa máxima de transferência: $16.7mseg/trilha$ ou $1229bytes/msec$
- Bytes por setor: 512
- Setores por trilha: 100
- Trilhas por cilindro: 12
- Trilhas por superfície: 1748 (igual ao número de cilindros)
- Fator de Interleave 1
- Tamanho do cluster: 10 setores
- Tamanho do extent: máximo de 10 clusters

Como o HD usa um cluster de 10 setores, e o extent é de 10 clusters

(100 setores = 1 trilha), o espaço é alocado para os arquivos em unidades de uma trilha ou seja: Alocação por trilhas: 10 clusters, com 10 setores cada.

UnB

Exemplo

Tarefa

Determinar o tempo necessário para ler um arquivo com 40000 registros de 256 bytes cada.

Primeiro precisamos saber como esta distribuído o arquivo no disco:

- Cada cluster (5120 bytes) pode armazenar 20 (256*20=5120) registros e, portanto, o arquivo ocupará uma sequência de 2000 clusters. Como o menor extent é 10 clusters, 200 extents são necessários para alocar o arquivo, que vai ocupar 200 trilhas.
- Vamos assumir que as 200 trilhas estão espalhadas aleatoriamente no disco (pior caso).

UnB

Exemplo

Podemos então calcular o tempo para leitura do arquivo, setor por setor, **em sequência.**

Para cada trilha (sequência de setores consecutivos), o processo de leitura envolve as seguintes operações.

Especificação

- Tempo médio de busca: 13 ms
- Latência (rotational delay): 8.3 ms
- Leitura de uma trilha: 16.7 ms (tempo de transferência)

Total:= 13 + 8.3 + 16.7 = 38 ms/estent

Para 200 trilhas := 200*38ms = 7600 ms = 7.6s

UnB

Exemplo

Agora vamos calcular o tempo de leitura para este mesmo arquivo, usando acesso **aleatório**. Isso significa que em vez de ler um setor depois do outro, **o acesso é feito por registros**, o que envolve mudar de trilha a cada vez que um novo registro é lido.

Para cada registro, essa operação envolve:

- Tempo médio de busca: 13 ms
- Latência (rotational delay): 8,3 ms
- Ler um cluster: 1,67 ms ((1/10) * 16,7) (tempo de transferência)

Total:= 13 + 8,3 + 1,67 = 22,97 ms

Para 40000 registros, temos:

40000*22,97 ms = 918800 ms = 15,4 min

UnB

Custo de acesso a disco:

- A diferença no desempenho entre ambos os tipos de acesso é muitíssimo importante.
- Ler o máximo de informação a cada posicionamento no disco, se possível, é muito melhor do que ficar deslocando a cabeça de R/W a cada novo registro a ser lido, fazendo uma nova busca para cada registro
- O tempo de busca é muito caro e deve ser minimizado sempre que possível.

UnB

Disco com um Gargalo

Disco com um Gargalo

- Os discos estão ficando cada vez mais rápidos, mas são muito mais lentos que as redes.
- Disco é o fator limitante de um processo (gargalo).
- CPU e LAN esperam tempos enormes pela transmissão de dados do disco.
- Isso normalmente significa que um processo é disk-bound (ou Disco-limitado), ou seja, CPU e rede têm que esperar pelos dados sendo transmitidos pelo disco.

Técnicas para minimizar o gargalo

- Multiprogramação
- Striping
- RAM disk
- Disk Cache

UnB

Disco como um Gargalo: Multiprogramação

Multriprogramação

- Permite que a CPU faça outras coisas enquanto espera pelos dados.
- Enquanto espera, o processo é bloqueado.
- O tempo da CPU é cotizado entre outros processos.
- Processo bloqueado acorda ao receber sinal de interrupção.
  - Volta para a fila dos processos prontos (ready) a espera da CPU.

UnB

Disco com um Gargalo: Striping

Striping

- Divisão do arquivo em partes, alocando cada parte em um disco diferente, e fazendo com que cada disco transmita partes do arquivo simultaneamente (sistema caro, pois requer múltiplos drivers de disco!).
- Esta técnica é um tipo de pseudo-paralelismo!

Princípio de Projeto

Onde quer que você encontre um gargalo no sistema, considere replicar a origem do gargalo e configurar o sistema de modo que essas sejam processadas em paralelo.

UnB

Disco com um Gargalo: RAM Disk

RAM Disk

Como a RAM tem ficado cada vez mais barata, uma outra técnica para resolver o gargalo imposto pelo disco seria ter um RAM disk.

- Uma parte da memória configurada para simular o comportamento de um disco, **exceto a velocidade e não-volatilidade.**
- Nos anos 80, quando disco rígidos eram caros, e floppy drives eram lentos, esse recurso foi muito popular para simular um disco flexível.

UnB

Disco como um Gargalo: RAM Disk

RAMDisk is a software driver that emulates as fully as possible the low-level functionality of a hard disk with system RAM. RAMDisk speeds up applications because RAM is much faster than mechanical hard disks for storing and retrieving data. Applications that do a lot of reading and writing to storage, like database queries, will show the most improvement with RAMDisk.

A RAM-Disk, Ramdisk or Ramdrive *is a virtual solid state disk that uses a segment of active computer memory, RAM, as secondary storage, a role typically filled by hard drives. Access times are greatly improved, because RAM is approximately a hundred times faster than hard drives. However, the volatility of RAM means that data will be lost if power is lost, e.g. when the computer is turned off. RAM disks can be used to store temporary data or hold uncompressed programs for short periods.*

RAM disks were popular as boot media in the 1980s, when hard drives were expensive, floppy drives were slow, and a few systems, such as the Amiga series and the Apple IIgs, supported booting from a RAM disk. At the cost of some main memory, the system could be soft-rebooted and be back in the operating system in mere seconds instead of minutes. Some systems had battery-backed RAM disks so their contents could persist when the system is shut down.

UnB

Disco como um Gargalo: RAM Disk

A proper disk cache in the operating system will usually obviate the performance motivation for a RAM disk; a disk cache fulfills a similar role (fast access to data that is notionally stored on a disk) without the various penalties (data loss in the event of power loss, static partitioning, etc.). RAM disks are, however, indispensable in situations in which a physical disk is not available, or where access to, or changing a physical disk is not desirable (such as in the case of Live CDs). They can also be used in a kiosk-style device where any changes made to a system are not committed and the original configuration is to be loaded each time the computer is turned on.

The advent of SATA has meant that RAM disks can be interfaced as a normal hard drive, although with extremely high transfer speeds.

Another way to use RAM to store files is the temporary filesystem. The difference between temporary filesystem and a RAM disk is that the RAM disk (/dev/ram0 etc.) is fixed-sized and acts like a disk partition, whereas the temporary filesystem (/dev/shm; in Source Mage GNU/Linux also /tmp) grows and shrinks to fit the files put on it.

Ramdisks have the advantage of being much faster than hard drives and only require special software (and of course the computer's RAM). Their disadvantage is that they are limited to main memory and data is lost on loss of power unless other measures (such as battery backup) are used.

UnB

Disco como um Gargalo: Disk Cache

Disk Cache

- Pode ser uma porção da memória RAM. É usada para acelerar o acesso aos dados no disco.
- Esta porção é configurada para conter páginas (um ou mais blocos contíguos) de dados do disco.
- Quando o dado é requisitado, o conteúdo do cache é verificado para ver se já não contém a informação desejada (parte do princípio de que, muitas vezes, certos conjuntos de dados são acessados seguidamente).
- Caso o dado não esteja na memória, a página contendo a informação é buscada no disco, e colocada no lugar de uma página corrente em memória.
- Atualmente os discos possuem um bloco de memória interna que funciona como o Disk Cache, de forma que não seja utilizada parte da RAM.

UnB

Disco como um Gargalo

As duas últimas técnicas (RAM Disk e Disk Cache) são exemplos de **buffering**.

UnB

RAID – Redundant Arrays of Inexpensive Disks

Idéia:

- Combinar vários discos rígidos de modo a criar um sistema virtual maior.
- Em vez de operar cada disco de forma independente, pode-se obter velocidades maiores, maior resistência a erros, e mais confiabilidade ligando as unidades pelo hardware, formando um array de unidades ou RAID.
- Sistema de tolerância a falhas que utiliza um conjunto de vários discos rígidos.
- Conceito original: uma configuração que agrupasse discos baratos e de pouca capacidade, chamados inexpensive capaz de substituir os discos de grande capacidade dos mainframes.

UnB

Sistemas RAID

Sistemas Raid (Redundant Arrays of Inexpensive Disks)

- São conjuntos de discos rígidos, normalmente de 3,5”, dispostos de tal modo que se consiga um dos seguintes objetivos: maior capacidade de armazenamento, redundância (tolerância a falhas) dos elementos que possibilitem a restauração dos dados, em caso de falha de um dos discos, maior velocidade de acesso ou vários dos anteriores.
- O modo de designar estes conjunto de discos, segundo a sua configuração, é Raid-0 à Raid-6.
- Essa tecnologia foi desenvolvida em 1987/1988 na Universidade de Califórnia em Berkeley.

UnB

Sistemas RAID

- A distribuição de dados em vários discos pode ser gerenciada por hardware ou por software.

- Software: o sistema operacional gerencia os discos do array.
- Hardware: através da instalação de uma placa controladora de RAID do tipo PCI (Peripheral Component Interconnect) Express num PC servidor, por exemplo.

UnB

Sistemas RAID

O sistema RAID consiste em um conjunto de dois ou mais discos rígidos com dois objetivos básicos:

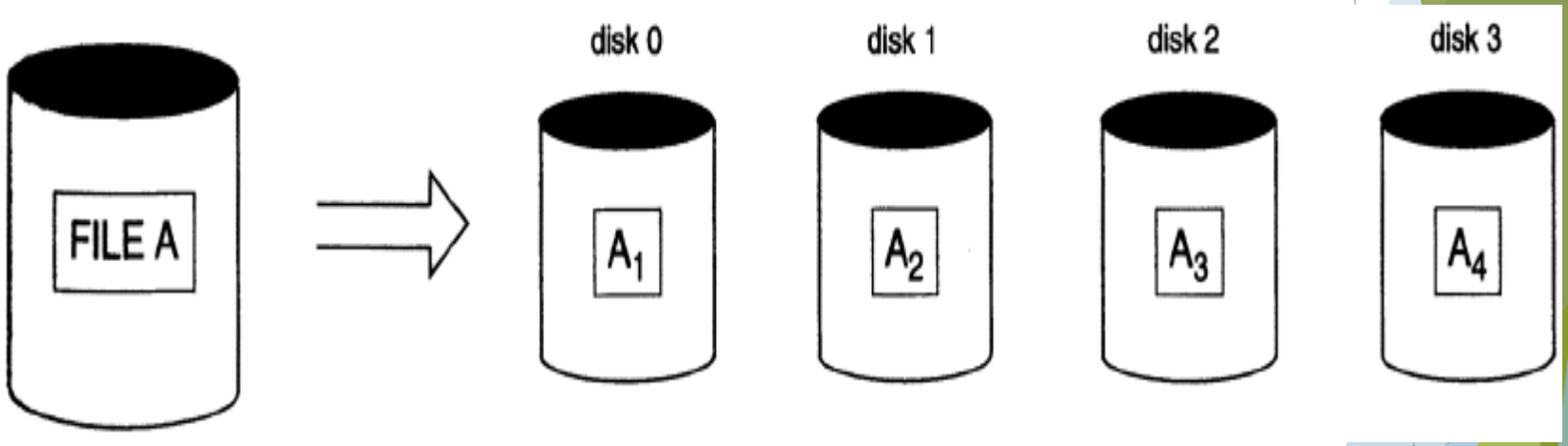
- Tornar o sistema de disco mais rápido (isto é, reduzir tempo de transferência de dados para a CPU). Já que vários discos acessados em paralelo fornecerão uma taxa de I/O superior a de um único disco.
- Aumentar confiabilidade (tornar o sistema de disco mais seguro) através da armazenagem de informação redundante e vários discos e assim permitir a recuperação dos dados em caso de falhas no disco.

As principais técnicas para atingir esses objetivos são:

- Stripping de dados (divisão de dados) com ou sem informação de paridade.
- Espelhamento (Mirroring).

UnB

Stripping (divisão de dados)



RAID-0

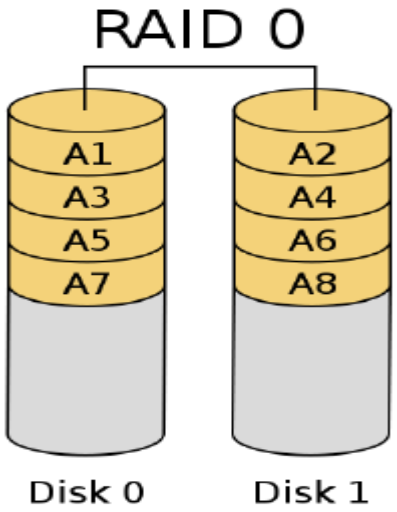
A técnica utilizada no RAID-0 é o Stripping sem informação de paridade dos dados.

Objetivo:

- Melhorar o desempenho no acesso

Contras

- Não oferece redundância => Não é tolerante à falhas



RAID 0 (bloco de dados)

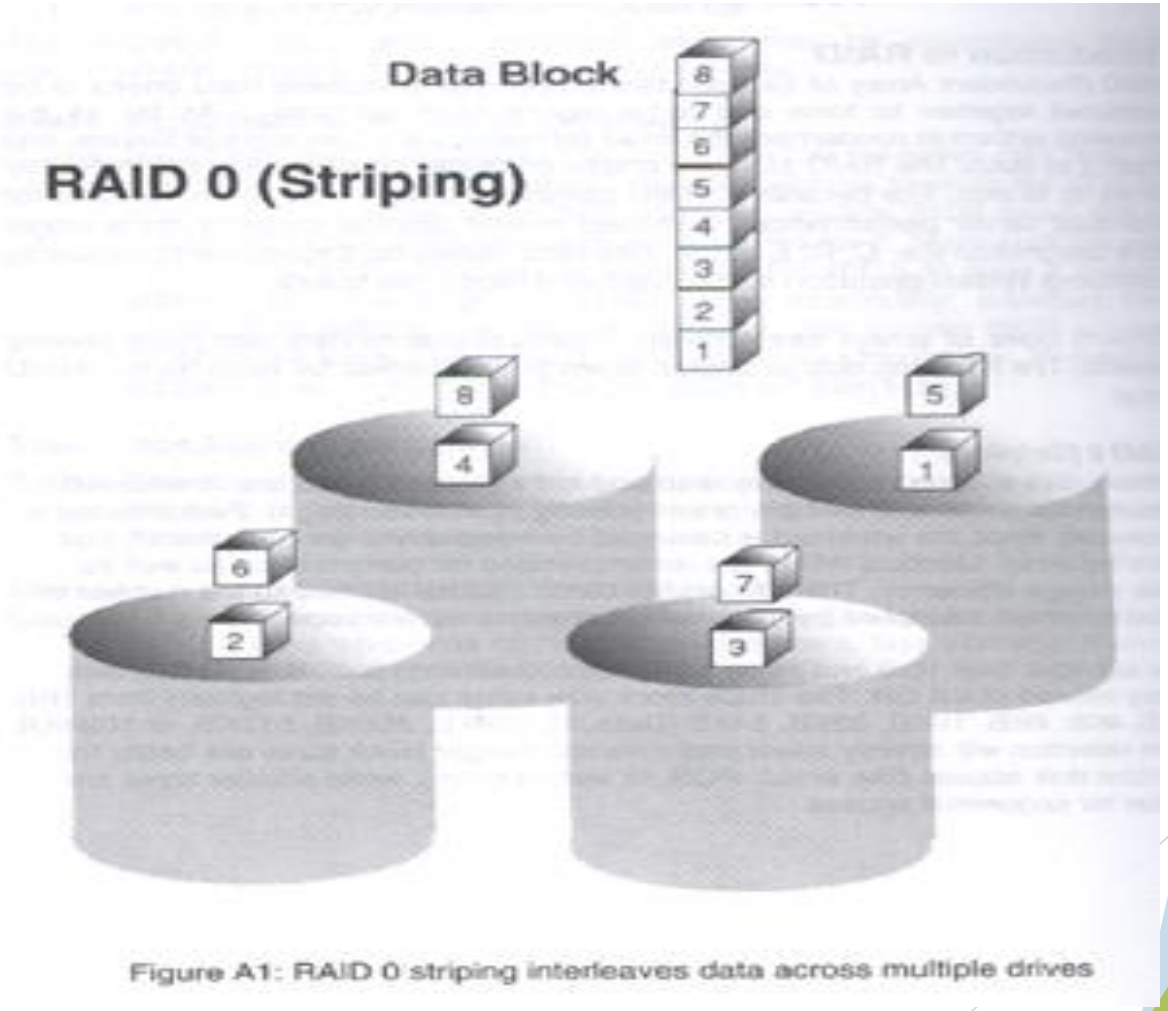


Figure A1: RAID 0 striping interleaves data across multiple drives

UnB

RAID-1

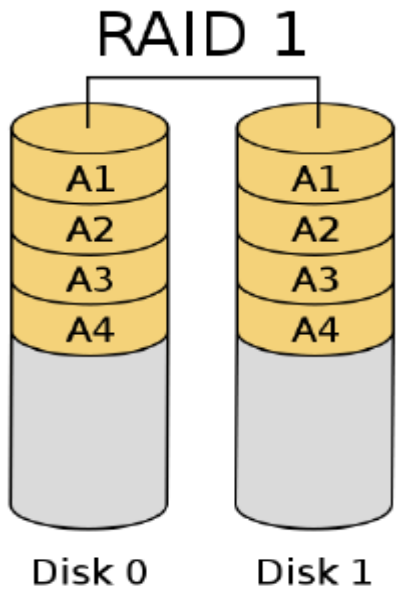
A técnica utilizada no RAID-1 é o Mirroring (Espelhamento)

Objetivo:

- Trazer redundância de dados => Maior tolerância à falhas

Contras:

- Espaço em disco destinado à redundâncias.



RAID 1 (Bloco de Dados)

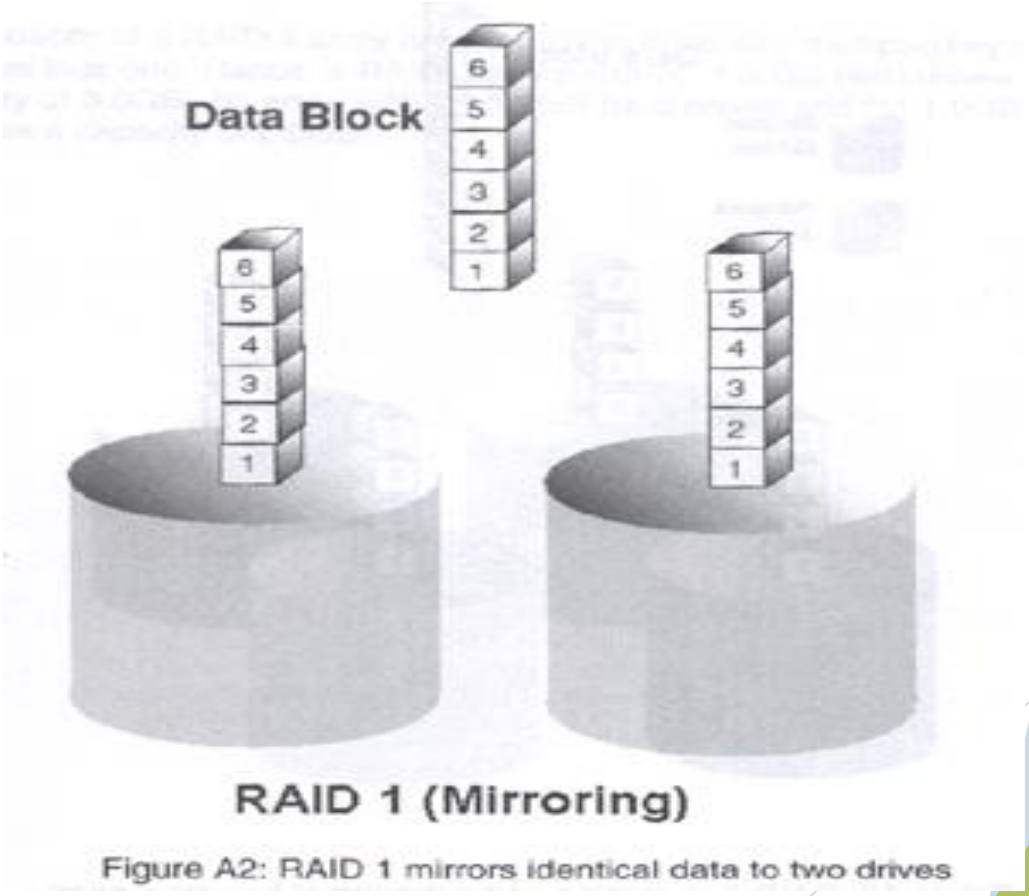


Figure A2: RAID 1 mirrors identical data to two drives

RAID-2

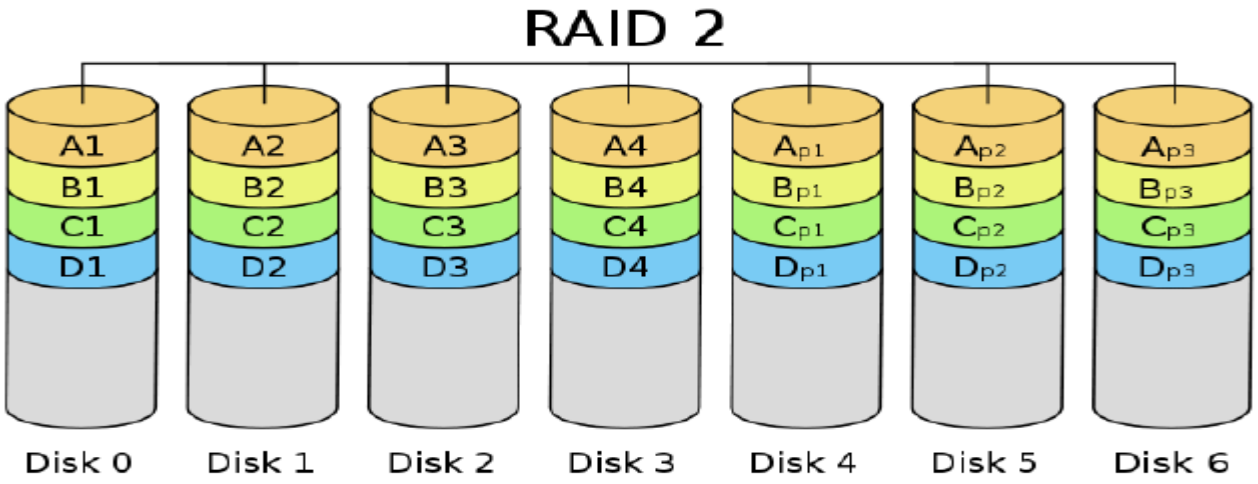
Cada bit do dado é gravado em um disco diferente, bem como cada código de Hamming do dado para uma possível recuperação de informação.

Objetivos:

- Proporcionar velocidade e tolerâncias à falhas. Além disso, é possível recuperar o dado olhando para o código de Hamming dele, caso haja falha em um dos discos.

Contras

- Complexidade e custo de armazenamento.



Código de Hamming: permite a transferência e armazenamento de dados de forma segura e eficiente

RAID-3

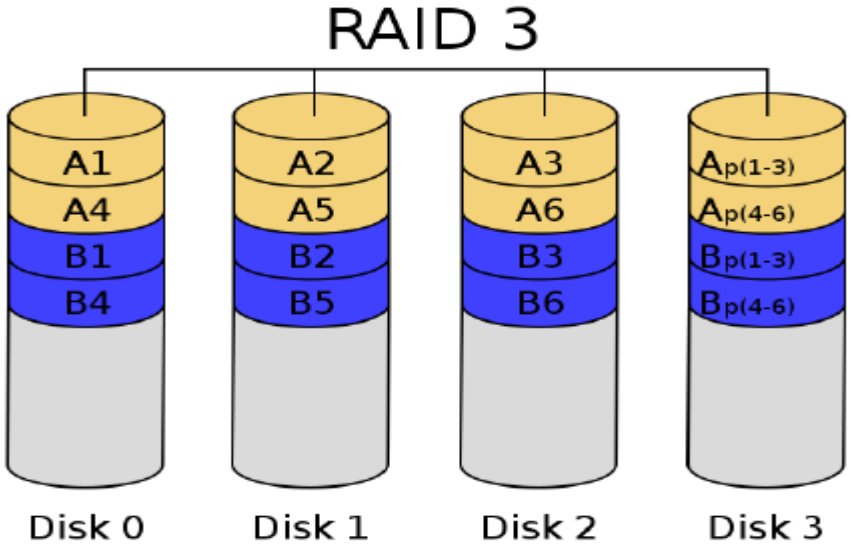
Cada byte do dado é gravado em um disco diferente, bem como o código de paridade dos bytes.

Objetivos:

- Proporcionar velocidade e detecção de erros.

Contras:

- Complexidade e custo de armazenamento.



Info de Paridade: Bits contendo informações extras são adicionadas aos dados. No caso de uma falha no disco, a paridade combinada com os dados remanescentes no disco pode ser utilizada para recriar os dados perdidos.

RAID-4

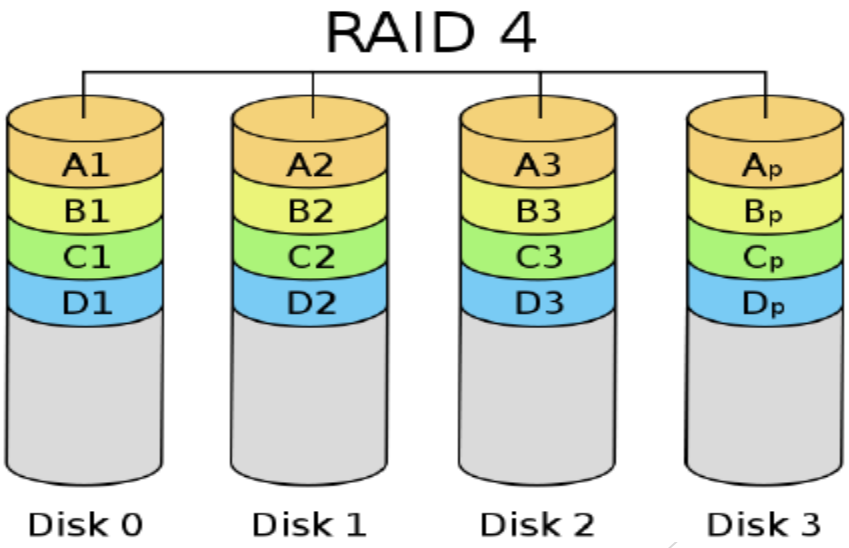
O dado é dividido em blocos e cada um destes é gravado em um disco diferente. O código de paridade associado aos blocos é gravado em um disco dedicado.

Objetivos:

- Proporcionar velocidade e detecção de erros

Contras:

- Complexidade e custo de armazenamento



RAID-5

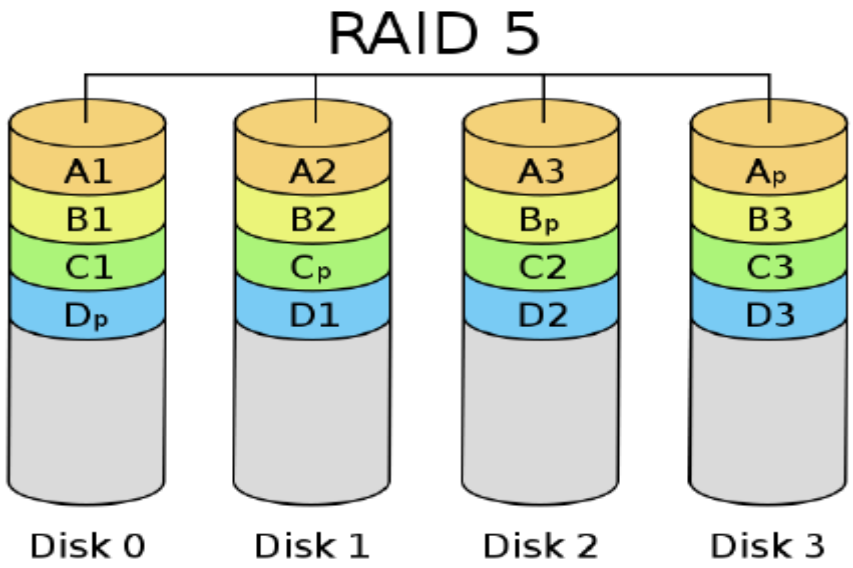
O dado é dividido em blocos e cada um destes é gravado em um disco diferente. É semelhante ao RAID-4, senão pelo fato do código de paridade ser gravado em discos diferentes.

Objetivos:

- Proporcionar velocidade e detecção de erros

Contras

- Complexidade e custo de armazenamento



RAID 5

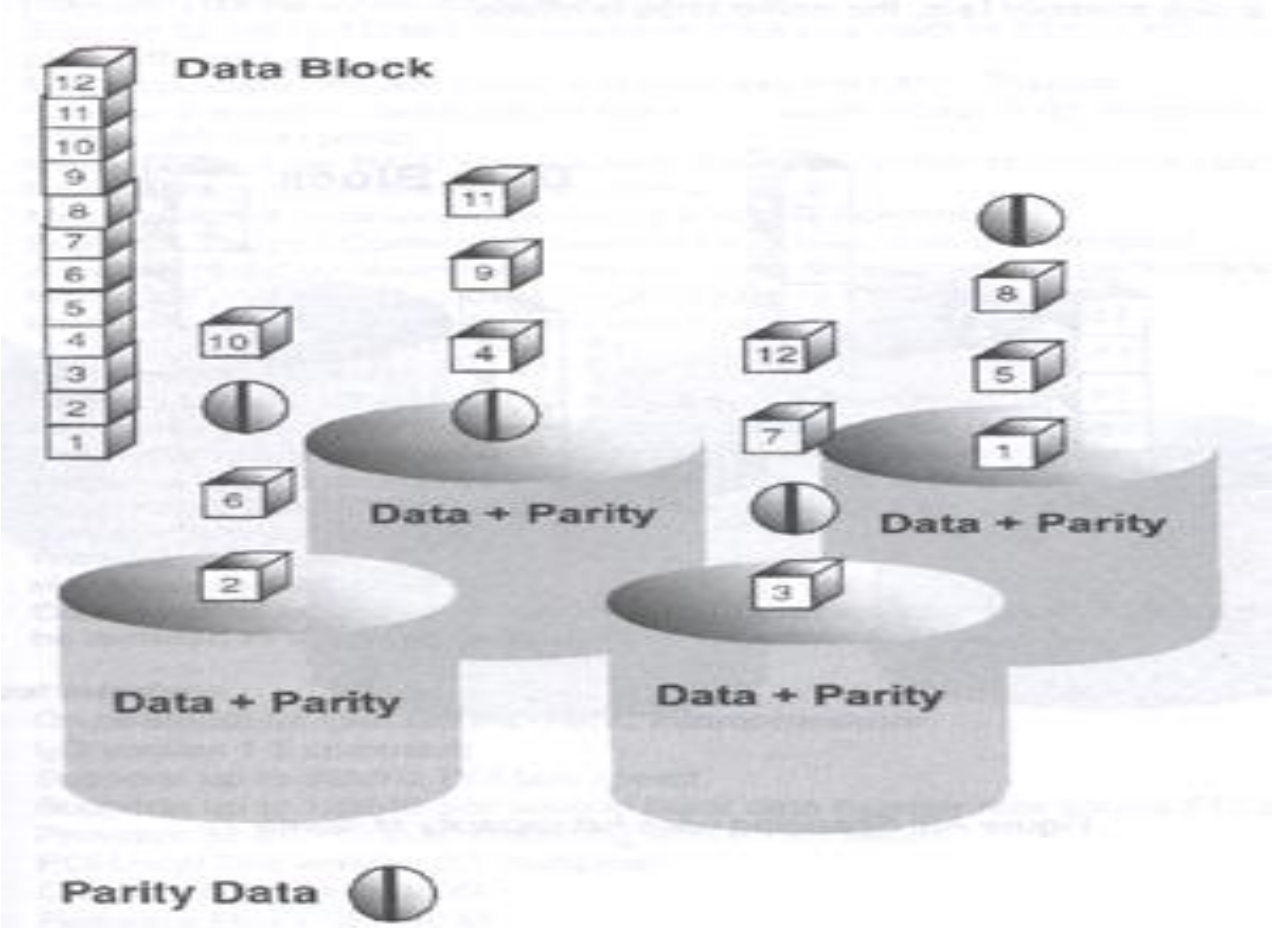


Figure A5: RAID 5 stripes all drives with data and parity info

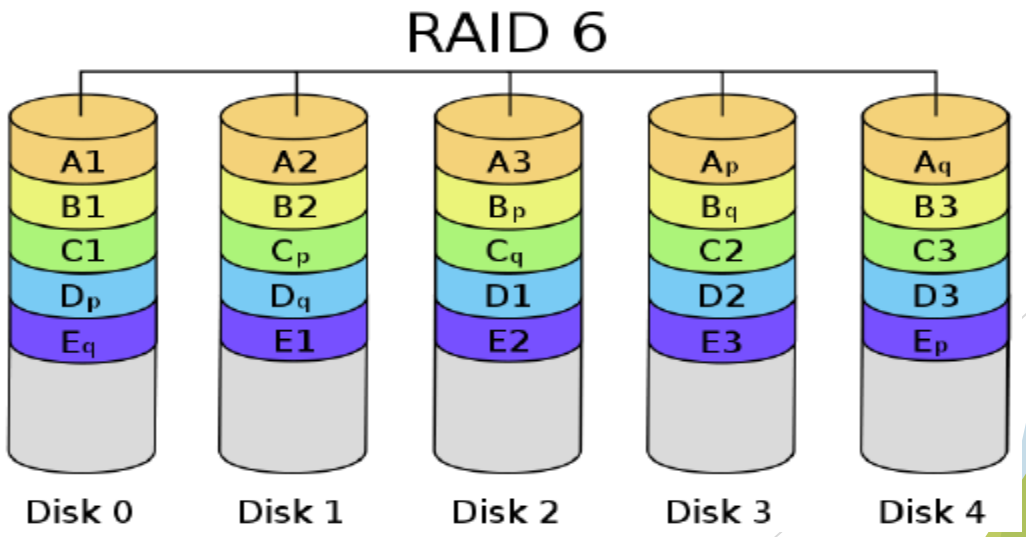
RAID-6

Idêntico ao RAID-5, senão por um fator. Um código de paridade extra de um bloco é gravado em um disco diferente do código de paridade original do bloco. Objetivos:

- Proporcionar velocidade e detecção de erros

Contras:

- Complexidade e custo de armazenamento



Sistemas RAID

Observação

Todos os sistemas RAID vistos anteriormente que usam códigos de paridade, possuem certa tolerância à falha. Caso haja falha em um disco, ao combinarmos o código de paridade com os dados remanescentes do disco, é possível recuperar a informação.

Sistemas RAID

As designações numéricas dos RAIDs são arbitrárias!

- Não significa que RAID 1 é melhor ou pior que RAID 5.
- Os números só funcionam como meio de identificar cada tecnologia.
- Depende da funcionalidade desejada:

1. Utilização eficiente da capacidade da unidade
2. Menor número de unidades
3. Maior confiabilidade
4. Melhor desempenho

UnB

Próxima Aula

Armazenagem Secundária (continuação).