

# Armazenamento de Dados

Allan Gonçalves de Oliveira

IC/UFMT

# Roteiro

- 1 Hierarquia da memória
- 2 Discos
- 3 Uso eficiente do espaço de armazenamento secundário
- 4 Otimização do tempo de acesso ao espaço de armazenamento secundário

# Hierarquia da memória

- Cache

## Definição

É parte integrante do processador, capaz de armazenar dados ou instruções de máquina.

- Memória Principal

## Definição

Característica principal é o acesso aleatório.

# Hierarquia da memória

- Cache

## Definição

É parte integrante do processador, capaz de armazenar dados ou instruções de máquina.

- Memória Principal

## Definição

Característica principal é o acesso aleatório.

# Hierarquia da memória

- Cache

## Definição

É parte integrante do processador, capaz de armazenar dados ou instruções de máquina.

- Memória Principal

## Definição

Característica principal é o acesso aleatório.

# Hierarquia da memória

- Cache

## Definição

É parte integrante do processador, capaz de armazenar dados ou instruções de máquina.

- Memória Principal

## Definição

Característica principal é o acesso aleatório.

# Hierarquia da memória

- Memória Virtual

## Definição

É onde as variáveis de programas, arquivos lidos ficam armazenadas. Geralmente é o próprio HD.

- Armazenamento Secundário

## Definição

É a forma de armazenamento mais lenta e mais espaçosa, tem acesso aleatório mas com um tempo de latência intrínseco.

# Hierarquia da memória

- Memória Virtual

## Definição

É onde as variáveis de programas, arquivos lidos ficam armazenadas. Geralmente é o próprio HD.

- Armazenamento Secundário

## Definição

É a forma de armazenamento mais lenta e mais espaçosa, tem acesso aleatório mas com um tempo de latência intrínseco.



# Hierarquia da memória

- Memória Virtual

## Definição

É onde as variáveis de programas, arquivos lidos ficam armazenadas. Geralmente é o próprio HD.

- Armazenamento Secundário

## Definição

É a forma de armazenamento mais lenta e mais espaçosa, tem acesso aleatório mas com um tempo de latência intrínseco.

# Hierarquia da memória

- Memória Virtual

## Definição

É onde as variáveis de programas, arquivos lidos ficam armazenadas. Geralmente é o próprio HD.

- Armazenamento Secundário

## Definição

É a forma de armazenamento mais lenta e mais espaçosa, tem acesso aleatório mas com um tempo de latência intrínseco.

# Hierarquia da memória

- Armazenamento terciário

## Definição

São componente para armazenar uma grande quantidade de dados, contudo operações de leitura/escrita são significativamente mais lentas.

Ex: Fita magnética.

- Armazenamento volátil e não volátil

## Definição de Volátil

"Esquece" o que está armazenado quando a energia é desligada.

## Definição de Não Volátil

Mantem o que está armazenado quando a energia é desligada.

# Hierarquia da memória

- Armazenamento terciário

## Definição

São componente para armazenar uma grande quantidade de dados, contudo operações de leitura/escrita são significativamente mais lentas.

Ex: Fita magnética.

- Armazenamento volátil e não volátil

## Definição de Volátil

"Esquece" o que está armazenado quando a energia é desligada.

## Definição de Não Volátil

Mantém o que está armazenado quando a energia é desligada.

# Hierarquia da memória

- Armazenamento terciário

## Definição

São componente para armazenar uma grande quantidade de dados, contudo operações de leitura/escrita são significativamente mais lentas.

Ex: Fita magnética.

- Armazenamento volátil e não volátil

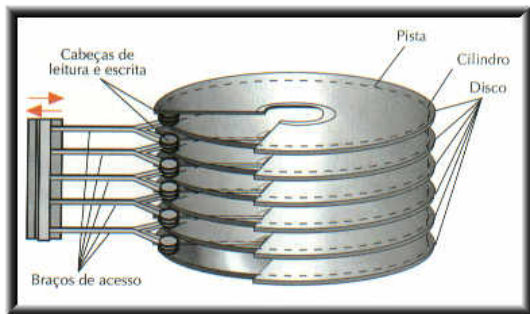
## Definição de Volátil

"Esquece" o que está armazenado quando a energia é desligada.

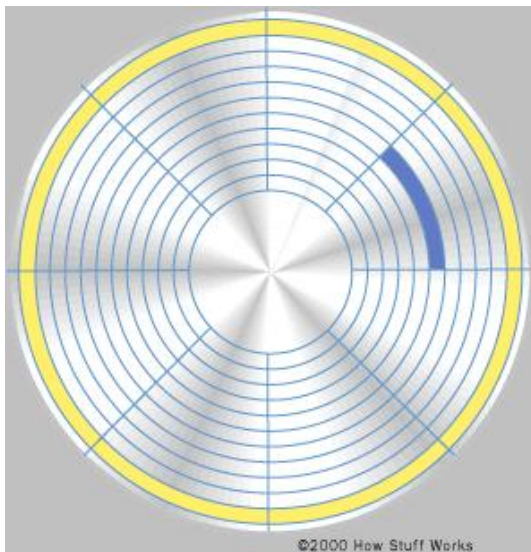
## Definição de Não Volátil

Mantem o que está armazenado quando a energia é desligada.

# Discos



# Discos - Setor



# Discos - Setor X Blocos

- Qual a diferença entre setor e bloco?
  - Setor é a unidade física do disco
  - Bloco é a unidade lógica do disco



# Discos - Setor X Blocos

- Qual a diferença entre setor e bloco?
- Setor é a unidade física do disco
- Bloco é a unidade lógica do disco

# Discos - Setor X Blocos

- Qual a diferença entre setor e bloco?
- Setor é a unidade física do disco
- Bloco é a unidade lógica do disco

# Latência

## Definição

É o tempo gasto entre o comando de leitura e o momento em que o conteúdo do bloco requisitado aparece na memória.

Pode ser dividido em:

- Tempo ocupado pelo processador para processar a solicitação
- Tempo de busca: é o tempo que leva para posicionar o conjunto de cabeças no cilindro certo.
- Latência rotacional: é o tempo para o disco girar até o setor certo.
- Tempo de transferência

# Latência

## Definição

É o tempo gasto entre o comando de leitura e o momento em que o conteúdo do bloco requisitado aparece na memória.

Pode ser dividido em:

- Tempo ocupado pelo processador para processar a solicitação
- Tempo de busca: é o tempo que leva para posicionar o conjunto de cabeças no cilindro certo.
- Latência rotacional: é o tempo para o disco girar até o setor certo.
- Tempo de transferência

# Latência

## Definição

É o tempo gasto entre o comando de leitura e o momento em que o conteúdo do bloco requisitado aparece na memória.

Pode ser dividido em:

- Tempo ocupado pelo processador para processar a solicitação
- Tempo de busca: é o tempo que leva para posicionar o conjunto de cabeças no cilindro certo.
- Latência rotacional: é o tempo para o disco girar até o setor certo.
- Tempo de transferência

# Latência

## Definição

É o tempo gasto entre o comando de leitura e o momento em que o conteúdo do bloco requisitado aparece na memória.

Pode ser dividido em:

- Tempo ocupado pelo processador para processar a solicitação
- Tempo de busca: é o tempo que leva para posicionar o conjunto de cabeças no cilindro certo.
- Latência rotacional: é o tempo para o disco girar até o setor certo.
- Tempo de transferência

# Latência

## Definição

É o tempo gasto entre o comando de leitura e o momento em que o conteúdo do bloco requisitado aparece na memória.

Pode ser dividido em:

- Tempo ocupado pelo processador para processar a solicitação
- Tempo de busca: é o tempo que leva para posicionar o conjunto de cabeças no cilindro certo.
- Latência rotacional: é o tempo para o disco girar até o setor certo.
- Tempo de transferência

# Latência

## Definição

É o tempo gasto entre o comando de leitura e o momento em que o conteúdo do bloco requisitado aparece na memória.

Pode ser dividido em:

- Tempo ocupado pelo processador para processar a solicitação
- Tempo de busca: é o tempo que leva para posicionar o conjunto de cabeças no cilindro certo.
- Latência rotacional: é o tempo para o disco girar até o setor certo.
- Tempo de transferência



# Latência

## Definição

É o tempo gasto entre o comando de leitura e o momento em que o conteúdo do bloco requisitado aparece na memória.

Pode ser dividido em:

- Tempo ocupado pelo processador para processar a solicitação
- Tempo de busca: é o tempo que leva para posicionar o conjunto de cabeças no cilindro certo.
- Latência rotacional: é o tempo para o disco girar até o setor certo.
- Tempo de transferência

- Há dez superfícies, com 10.000 trilhas cada uma
- As trilhas têm em média 1.000 setores de 512 bytes cada
- 20% de cada trilha são usados nos intervalos

Pergunta: Qual é a capacidade do disco?

- Há dez superfícies, com 10.000 trilhas cada uma
- As trilhas têm em média 1.000 setores de 512 bytes cada
- 20% de cada trilha são usados nos intervalos

Pergunta: Qual é a capacidade do disco?

- Há dez superfícies, com 10.000 trilhas cada uma
- As trilhas têm em média 1.000 setores de 512 bytes cada
- 20% de cada trilha são usados nos intervalos

Pergunta: Qual é a capacidade do disco?

- Há dez superfícies, com 10.000 trilhas cada uma
- As trilhas têm em média 1.000 setores de 512 bytes cada
- 20% de cada trilha são usados nos intervalos

Pergunta: Qual é a capacidade do disco?

# Uso eficiente do espaço de armazenamento secundário

- Um SGBD deve trabalhar com uma quantidade de dados que não cabem na memória.
- Com isso é necessário uma medida para estimar o tempo de processamento.
- Qual?

# Uso eficiente do espaço de armazenamento secundário

- Um SGBD deve trabalhar com uma quantidade de dados que não cabem na memória.
- Com isso é necessário uma medida para estimar o tempo de processamento.
- Qual?

# Uso eficiente do espaço de armazenamento secundário

- Um SGBD deve trabalhar com uma quantidade de dados que não cabem na memória.
- Com isso é necessário uma medida para estimar o tempo de processamento.
- Qual?



# Uso eficiente do espaço de armazenamento secundário

## Custo de E/S

O número de acessos a blocos (leitura e escritas) é uma boa aproximação do tempo exigido para algum algoritmo classificar os dados.

# Exemplo

Vamos supor que temos:

- 10.000.000 de tuplas
  - Temos cada bloco com 4096 ou  $2^{12}$  bytes
  - Se cada bloco armazena 40 tuplas de 100 bytes cada
  - Teremos 250.000 blocos
  - Se tivermos somente 50 Mbytes de memória
  - Teremos  $\frac{50 \times 2^{20}}{2^{12}} = 12.800$  blocos em memória
  - Como ordenar esse conjunto de dados?

# Exemplo

Vamos supor que temos:

- 10.000.000 de tuplas
- Temos cada bloco com 4096 ou  $2^{12}$  bytes
- Se cada bloco armazena 40 tuplas de 100 bytes cada
- Teremos 250.000 blocos
- Se tivermos somente 50 Mbytes de memória
- Teremos  $\frac{50 \times 2^{20}}{2^{12}} = 12.800$  blocos em memória
- Como ordenar esse conjunto de dados?

# Exemplo

Vamos supor que temos:

- 10.000.000 de tuplas
- Temos cada bloco com 4096 ou  $2^{12}$  bytes
- Se cada bloco armazena 40 tuplas de 100 bytes cada
- Teremos 250.000 blocos
- Se tivermos somente 50 Mbytes de memória
- Teremos  $\frac{50 \times 2^{20}}{2^{12}} = 12.800$  blocos em memória
- Como ordenar esse conjunto de dados?

# Exemplo

Vamos supor que temos:

- 10.000.000 de tuplas
- Temos cada bloco com 4096 ou  $2^{12}$  bytes
- Se cada bloco armazena 40 tuplas de 100 bytes cada
- Teremos 250.000 blocos
- Se tivermos somente 50 Mbytes de memória
- Teremos  $\frac{50 \times 2^{20}}{2^{12}} = 12.800$  blocos em memória
- Como ordenar esse conjunto de dados?

# Exemplo

Vamos supor que temos:

- 10.000.000 de tuplas
- Temos cada bloco com 4096 ou  $2^{12}$  bytes
- Se cada bloco armazena 40 tuplas de 100 bytes cada
- Teremos 250.000 blocos
- Se tivermos somente 50 Mbytes de memória
- Teremos  $\frac{50 \times 2^{20}}{2^{12}} = 12.800$  blocos em memória
- Como ordenar esse conjunto de dados?

# Exemplo

Vamos supor que temos:

- 10.000.000 de tuplas
- Temos cada bloco com 4096 ou  $2^{12}$  bytes
- Se cada bloco armazena 40 tuplas de 100 bytes cada
- Teremos 250.000 blocos
- Se tivermos somente 50 Mbytes de memória
- Teremos  $\frac{50 \times 2^{20}}{2^{12}} = 12.800$  blocos em memória
- Como ordenar esse conjunto de dados?

# Exemplo

Vamos supor que temos:

- 10.000.000 de tuplas
- Temos cada bloco com 4096 ou  $2^{12}$  bytes
- Se cada bloco armazena 40 tuplas de 100 bytes cada
- Teremos 250.000 blocos
- Se tivermos somente 50 Mbytes de memória
- Teremos  $\frac{50 \times 2^{20}}{2^{12}} = 12.800$  blocos em memória
- Como ordenar esse conjunto de dados?



# Classificação por intercalação de vários caminhos e duas fases

## Definição

- Fase 1: Classificar itens de dados com um tamanho que caiba na memória principal. Assim teremos um conjunto de sublistas classificadas.
- Fase 2: Mesclar todas as sublistas em uma única lista.

# Classificação por intercalação de vários caminhos e duas fases

## Definição

- Fase 1: Classificar itens de dados com um tamanho que caiba na memória principal. Assim teremos um conjunto de sublistas classificadas.
- Fase 2: Mesclar todas as sublistas em uma única lista.

# Classificação por intercalação de vários caminhos e duas fases

## Definição

- Fase 1: Classificar itens de dados com um tamanho que caiba na memória principal. Assim teremos um conjunto de sublistas classificadas.
- Fase 2: Mesclar todas as sublistas em uma única lista.

# Exemplo

Com o algoritmos anterior teremos:

- Com 12.800 blocos por vez na memória teremos certa de 20 sublistas
- Quanto tempo demora a primeira fase?
- Quanto tempo demora a segunda fase?
- Se cada operação de E/S de um bloco leva 15 milissegundos, quanto tempo demora o algoritmo?

# Exemplo

Com o algoritmos anterior teremos:

- Com 12.800 blocos por vez na memória teremos certa de 20 sublistas
- Quanto tempo demora a primeira fase?
- Quanto tempo demora a segunda fase?
- Se cada operação de E/S de um bloco leva 15 milissegundos, quanto tempo demora o algoritmo?

# Exemplo

Com o algoritmos anterior teremos:

- Com 12.800 blocos por vez na memória teremos certa de 20 sublistas
- Quanto tempo demora a primeira fase?
- Quanto tempo demora a segunda fase?
- Se cada operação de E/S de um bloco leva 15 milissegundos, quanto tempo demora o algoritmo?

# Exemplo

Com o algoritmos anterior teremos:

- Com 12.800 blocos por vez na memória teremos certa de 20 sublistas
- Quanto tempo demora a primeira fase?
- Quanto tempo demora a segunda fase?
- Se cada operação de E/S de um bloco leva 15 milissegundos, quanto tempo demora o algoritmo?

# Otimização do tempo de acesso ao espaço de armazenamento secundário

- Colocar juntos blocos que serão acessados no mesmo cilindro
- Dividir os dados em discos pequenos
- "Espelhar" o disco
- Realizar a transferência prévia de blocos



# Otimização do tempo de acesso ao espaço de armazenamento secundário

- Colocar juntos blocos que serão acessados no mesmo cilindro
- Dividir os dados em discos pequenos
- "Espelhar" o disco
- Realizar a transferência prévia de blocos

# Otimização do tempo de acesso ao espaço de armazenamento secundário

- Colocar juntos blocos que serão acessados no mesmo cilindro
- Dividir os dados em discos pequenos
- "Espelhar" o disco
- Realizar a transferência prévia de blocos

# Otimização do tempo de acesso ao espaço de armazenamento secundário

- Colocar juntos blocos que serão acessados no mesmo cilindro
- Dividir os dados em discos pequenos
- "Espelhar" o disco
- Realizar a transferência prévia de blocos

# Exercício

Implementar um algoritmo que dado uma lista de nomes, ordene-os segundo o algoritmo estudado, supondo que na memória cabem no máximo 10 nomes.