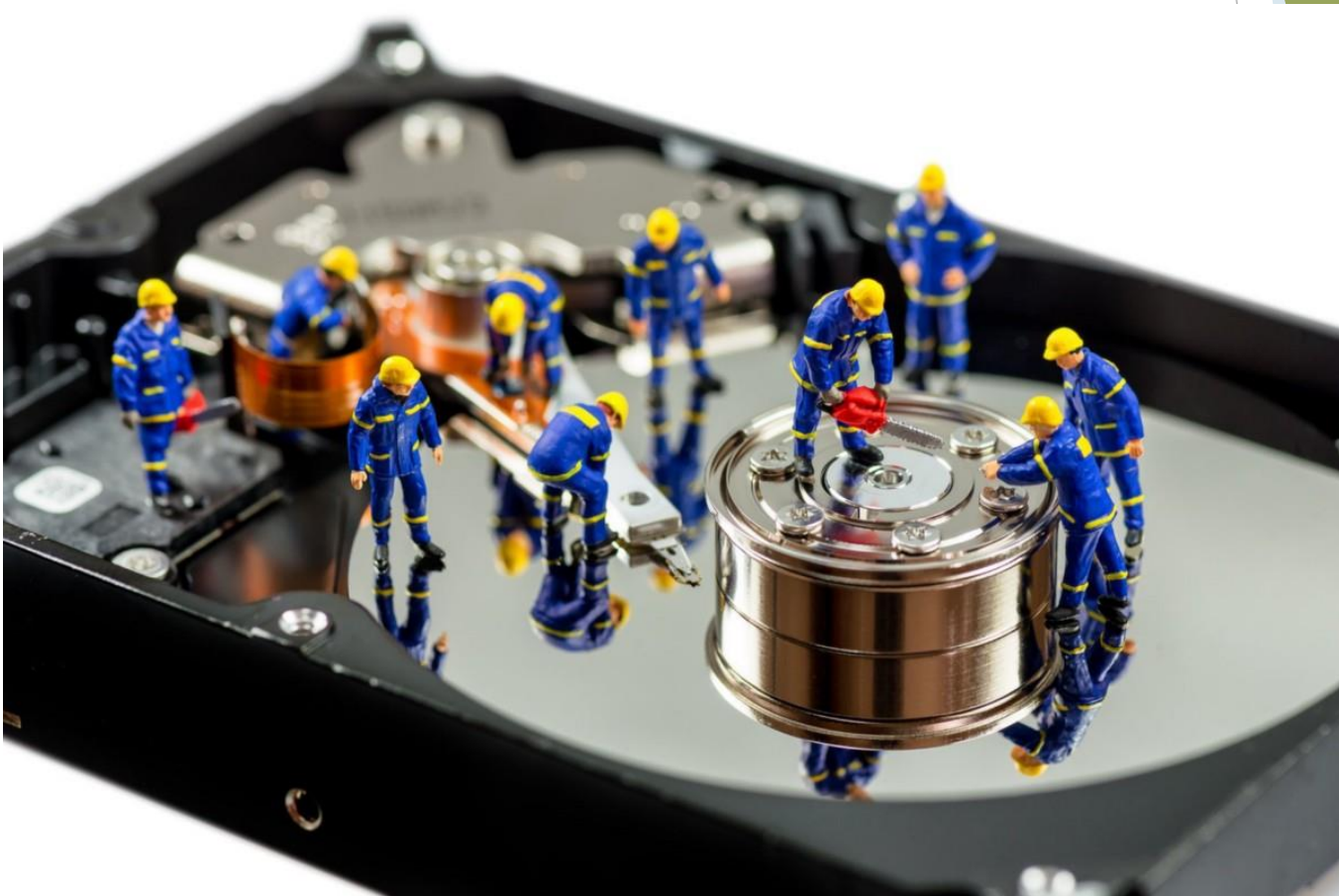


116327 -Organização de Arquivos



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 1

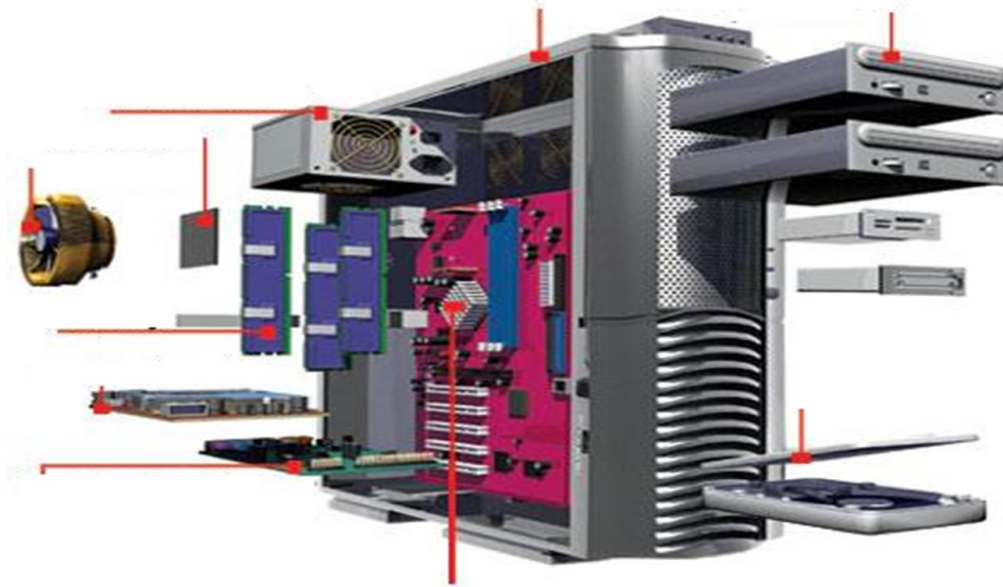
M.Sc. Oscar Gaidos

Universidade de Brasília (UnB)

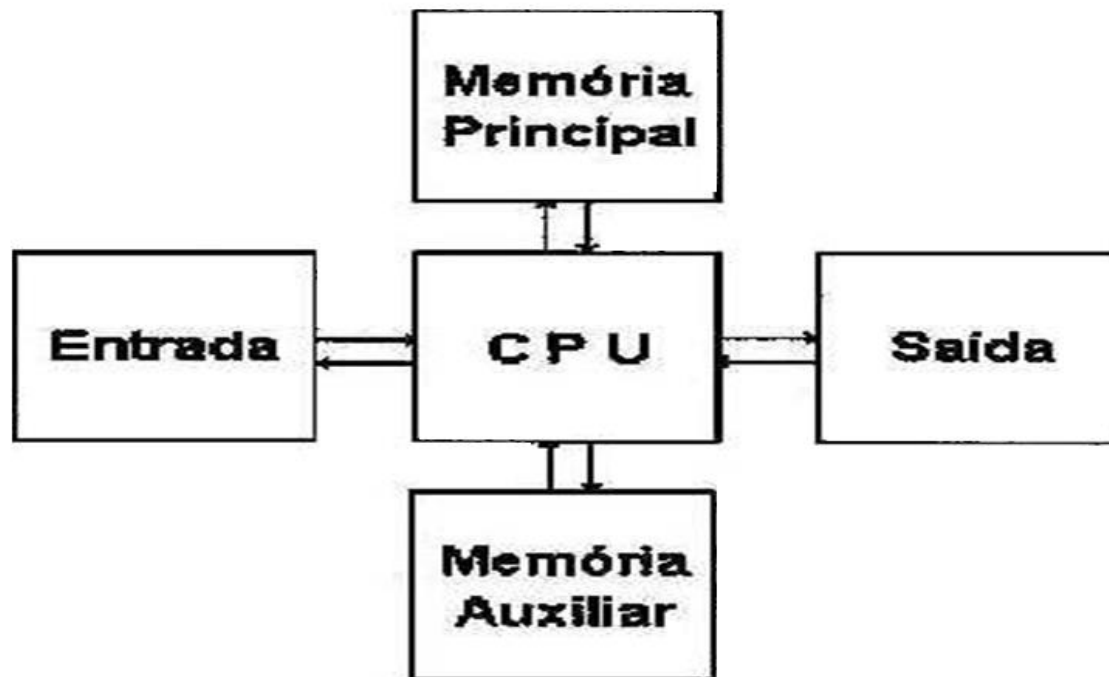
Sumário

- ▶ Conceitos Iniciais
- ▶ Histórico
- ▶ Arquivos
- ▶ Unix e DOS

Quais são as partes de um computador?



Computador Simplificado



Definição (Memória Primária)

- ▶ A memória primária, conhecida como memória interna ou simplesmente memória, é a única que o CPU acessa diretamente. A CPU continuamente executa instruções que se encontram na memória principal.
- ▶ A memória principal pode ser dividida em:
- ▶ Registradores: Lugar para armazenar dados dentro do processador. Normalmente um registro armazena de 32 a 64 bits.
- ▶ Cache: É uma memória intermediária entre os ultra-rápidos registros e uma memória mais lenta (RAM). A função do cache é de aumentar a performance do processador.
- ▶ RAM: Memória Principal de Acesso Randômico. A RAM é conectada ao CPU mediante de um bus (memory bus). A tecnologia utilizada para a RAM normalmente é do tipo volátil.
- ▶ ROM: Memória somente de leitura não volátil. Normalmente utilizada para armazenar a BIOS, e normalmente é mais lenta que a RAM.

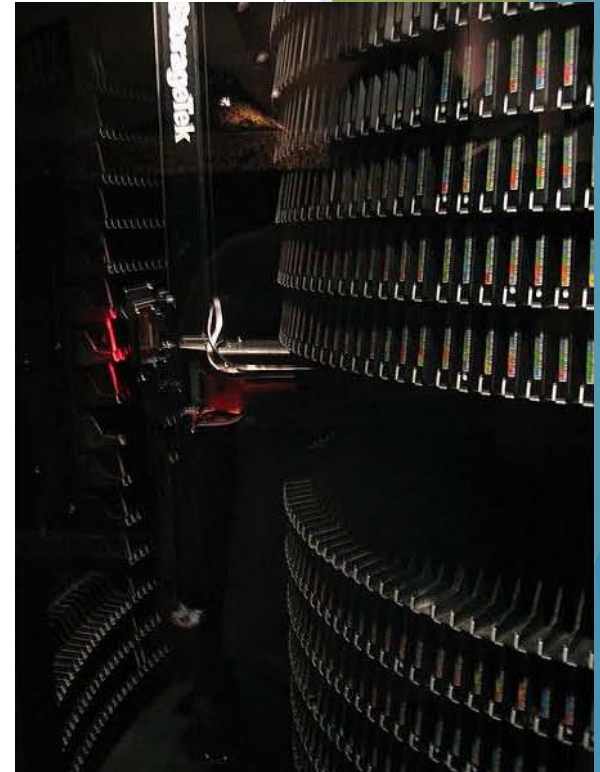
Definição (Memória Secundária)

- ▶ A memória secundária não é acessada diretamente pelo CPU, são utilizados canais de acesso de E/S para transferir dados para a memória principal.
- ▶ A memória secundária NÃO é volátil e permite o armazenamento massivo de dados.
- ▶ Normalmente a memória secundária é implementada por meio de Discos Rígidos.
- ▶ Outros dispositivos são: CDs, DVDs, flash drivers (pen drives), floppy disks, fita magnética, cartões perfurados, etc.

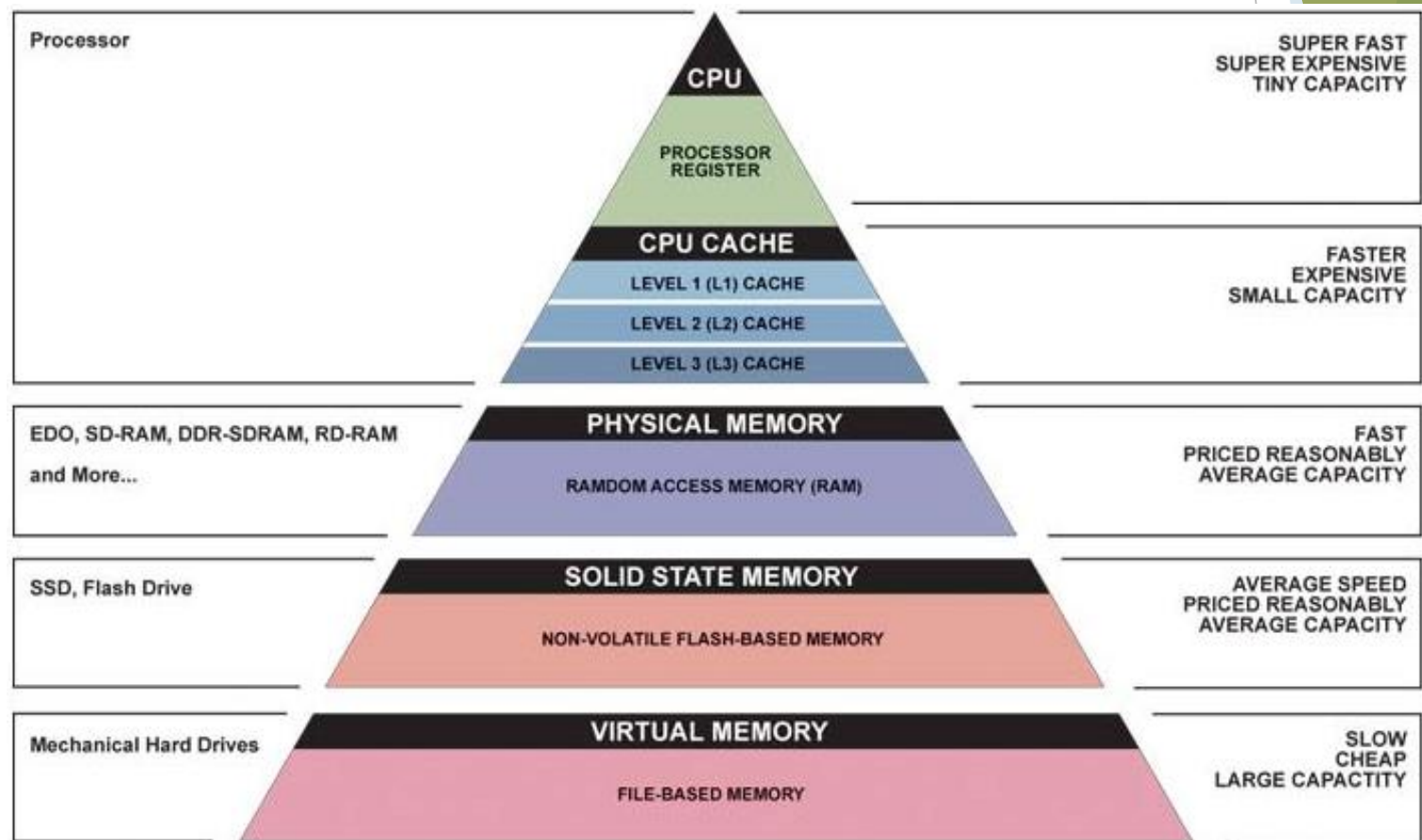


Definição (Memória Terciária)

- ▶ A memória terciária proporciona um terceiro nível de armazenamento, porém, na maioria de sistemas, ela não esta presente.
- ▶ A memória terciária representa unidades de armazenamento removíveis as quais são montadas por algum mecanismo robótico (automático).
- ▶ Normalmente, os dados são transferidos para a memória secundária. A memória terciária é feita para armazenar dados que são raramente acessados

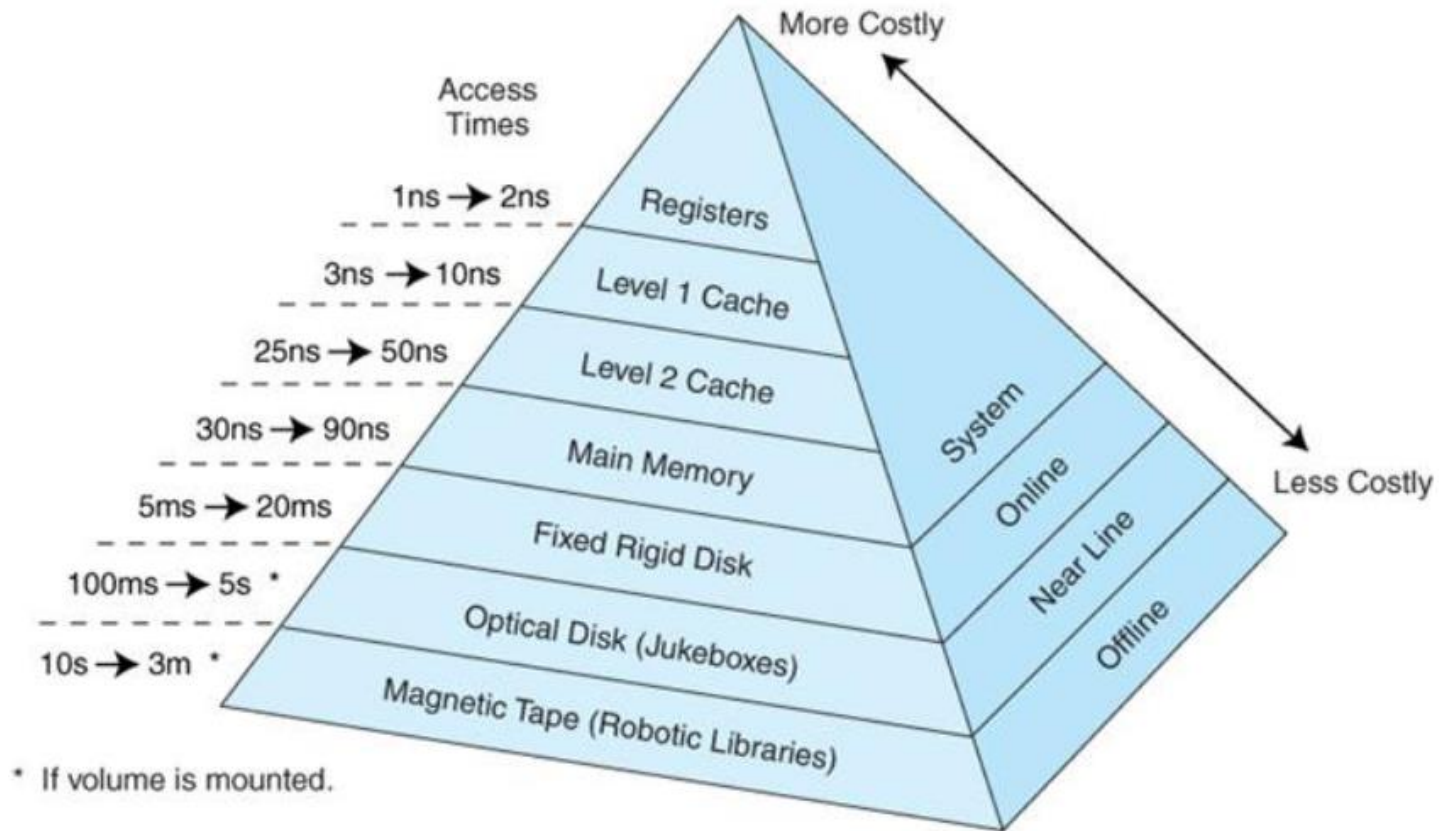


Motivação



▲ Simplified Computer Memory Hierarchy
Illustration: Ryan J. Leng

Motivação



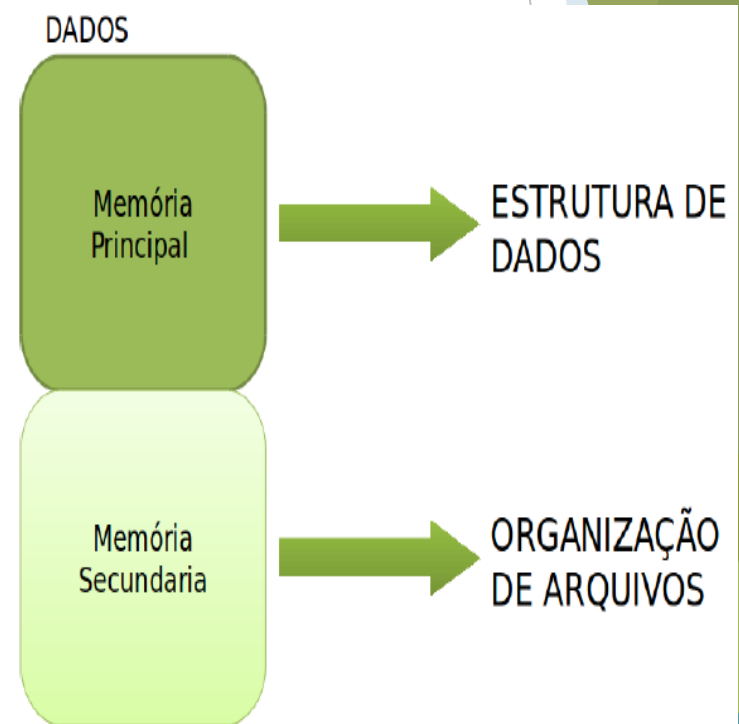
Dispositivos diferentes requerem abordagens diferentes!

Organização de Arquivos

- ▶ Definição (Organização de Arquivos)
- ▶ A organização de arquivos é uma disciplina que estuda tanto a representação de dados em arquivos através de estruturas como as operações necessárias para acessar estes dados na **memória secundária**.
- ▶ As estruturas de dados de arquivos permitem:
- ▶ Realização de operações de leitura, escrita e alteração de dados.
- ▶ Suportam a busca de dados que seguem um determinado critério de busca, e/ou leitura de dados em uma determinada ordem.

AO vs ED

- ▶ Como visto, o estudo de Organização de arquivos consiste na aplicação de algoritmos e estruturas de dados (ED) para a resolução de problemas de armazenamento, processamento e acesso de dados da memória secundária.
- ▶ Logo, tanto ED como AO estudam:
- ▶ Como representar dados
- ▶ Operações associadas aos dados e às estruturas que representam esses dados



Representação de dados

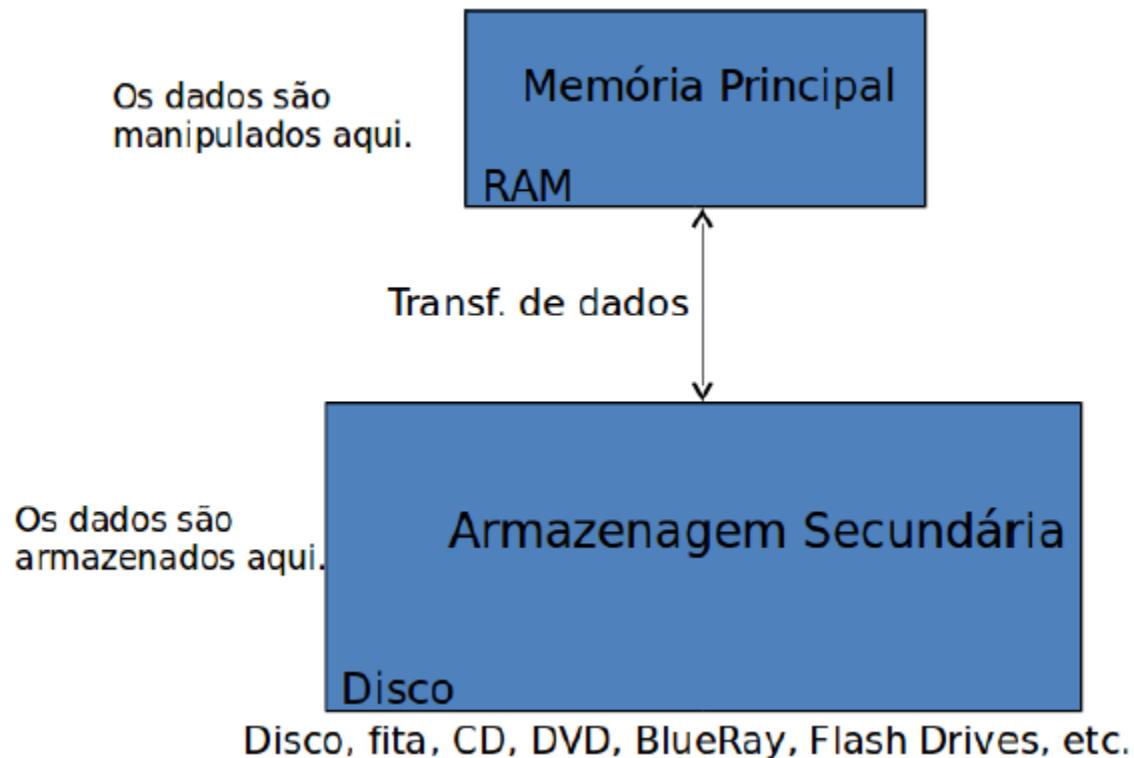
- ▶ Porque é importante estudar formas de representar dados?

Representação de dados

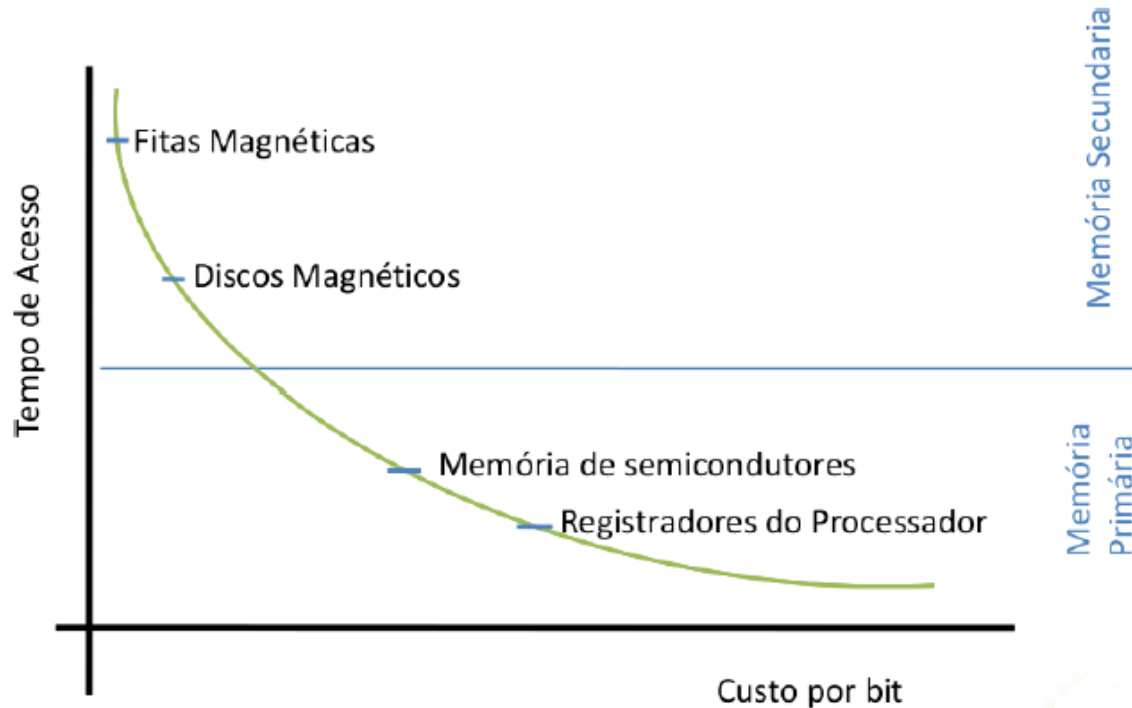
- ▶ Porque é importante estudar formas de representar dados?
- ▶ Dentro do âmbito da ciência e Engenharia da Computação, a representação de dados permite:
- ▶ Armazenar
- ▶ Acessar
- ▶ Processar e Organizar de maneira adequada e eficiente

Organização de Arquivos

► Porque organizar arquivos?



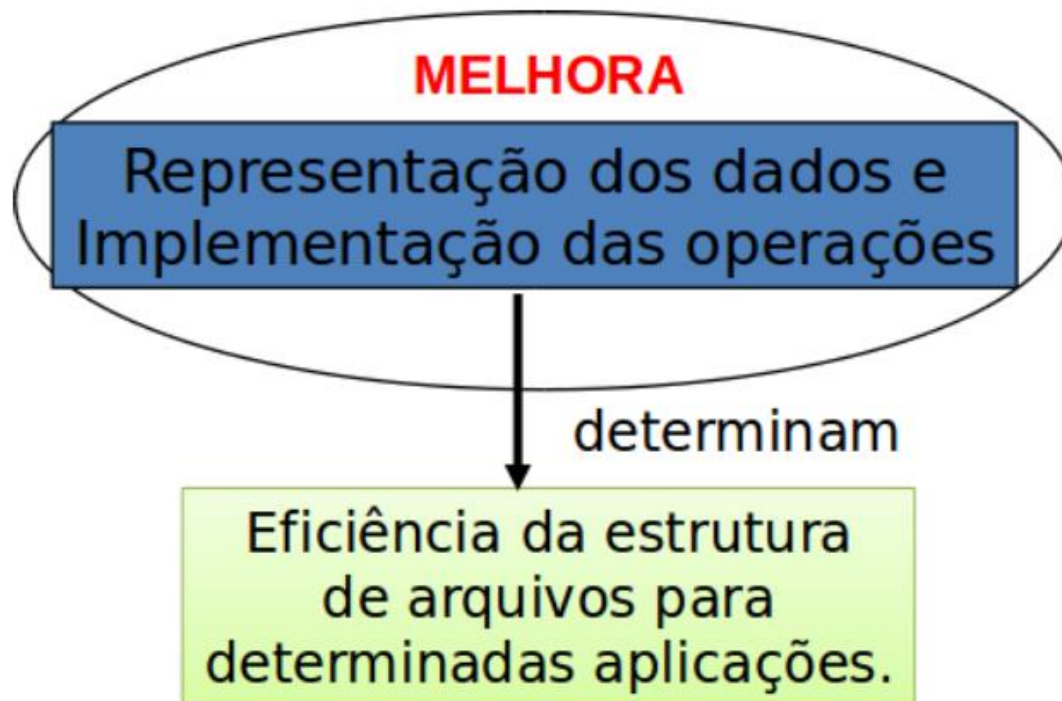
Custo vs Espaço



- ▶ Tempo (típico) para obter uma informação da(o):
- ▶ Memória RAM: 60 nanosegundos (60 bilhonésimos de segundo).
- ▶ Disco: 15 milisegundos (15 milésimos de segundo).

Organização de Arquivos

- Uma eficiente representação dos dados, e organização dos mesmos melhora o tempo de acesso.



Melhora no tempo de recuperação de dados da memória secundária

Organização de arquivos

- ▶ Apesar de mais barata que a memória primária, o acesso à memória secundária é **caro** em termos de tempo.
- ▶ Organização de arquivos
- ▶ Necessitamos de representar de maneira adequada (através de estruturas de dados) os dados em memória secundária. Através de **algoritmos** eficientes nestas estruturas, procuramos recuperar os dados da memória secundária visando diminuir o **tempo de acesso** nesta memória.

Conceitos Iniciais

- ▶ Definição (Entidade)
- ▶ Objeto sobre o qual se armazenam informações. Ex: empregados, alunos, contas correntes, centros de custo, etc.
- ▶ Definição (Classe de Entidades)
- ▶ Classe de Entidade: Um conjunto de entidades que representam objetos da **mesma natureza**. Ex: empregados de uma empresa, alunos de uma escola, itens de um almoxarifado, etc.
- ▶ Definição (Atributos)
- ▶ Atributo: Uma determinada **característica** de uma entidade. Por exemplo, a classe de empregados de uma empresa pode ter os seguintes atributos:
 - ▶ Nome
 - ▶ Matrícula
 - ▶ Cargo
 - ▶ Salário

Conceitos Iniciais

- Um conjunto de pares <atributo, valor do atributo>, que identifica uma entidade é chamado de registro de entidade.

Atributo	Valor do Atributo
Nome	Petronilo Pancrácio II
Matrícula	2.0004-11
Cargo	Contador

- Se um conjunto de registros que caracteriza uma classe de entidades é mantido na memória principal é comumente chamado de **tabela**. Se é mantido na memória secundária é normalmente chamado de **arquivo**.

Conceitos iniciais

- ▶ Na organização de arquivos existe um conceito abstrato de dado, chamado campo.
- ▶ Definição (Campo)
- ▶ O campo pode ser definido como sendo a menor unidade lógica de informação. Logo, um registro pode ser definido como um conjunto de campos. Relacionando com o conceito de entidade, podemos dizer que cada campo representa um determinado atributo de uma determinada classe de entidade.

Conceitos iniciais

- ▶ Definição (arquivos)
- ▶ Considerando determinado dispositivo de armazenamento e do ponto de vista da organização dos dados em registros, um arquivo pode ser definido como uma coleção do mesmo tipo de estruturas de dados, com informações sobre uma determinada classe de entidade que está armazenando em tal dispositivo. Essa estrutura pode ser um agregado de valores heterogêneos e homogêneos.
- ▶ O conteúdo (informação) está organizada em arrays de bytes. Porém, os dados que compõem a informação podem estar organizados sob alguma estrutura.

Definição (banco de dados)

- ▶ Um banco de dados é uma coleção de dados. Ele é composto de uma coleção de arquivos com algumas propriedades características, tais como redundância, isolamento, etc.
- ▶ Para um banco de dados existir, é essencial a presença de arquivos físicos, um programa gerenciador, uma autoridade ou administrador.

Histórico I

1. Arquivamento em fitas magnéticas => acesso sequencial
2. Criação de discos magnéticos
 - ▶ Possibilidade de acesso direto
 - ▶ Índices: Acessos feitos em arquivos pequenos consultando o valor da chave, que possui um ponteiro associado referente à informação à esta. Para arquivos grandes, essa metodologia se torna inviável.
3. Surgimento de estruturas em árvore: Indexam um arquivo de tamanho grande de maneira que a busca se torne viável.
 - ▶ Árbores Binárias
 - ▶ Árbores AVL
 - ▶ Árbores Paginadas
 - ▶ Árbores B
 - ▶ Tempo de acesso em $O(\log_k N)$, onde k é o número de entradas indexadas em um único bloco de uma árvore B.
 - ▶ Possibilidade de encontrar um dado específico entre milhões de outros com cerca de 3 ou 4 acessos a disco
 - ▶ Árbores B+

Histórico II

4. Hashing: Metodologia que possibilita a recuperação do dado em tempo $O(1)$, dado que o tamanho do arquivo não se altere muito com o passar do tempo.

5. Hashing extensível: Permite a recuperação do dado em $O(1)$, independente do crescimento do arquivo com relação ao tempo.

Próxima Aula

- ▶ Próxima Aula
- ▶ Introdução (continuação).