

# Organização de Arquivos

Prof<sup>a</sup>. Patricia D M Plentz

UFSC / CTC / INE

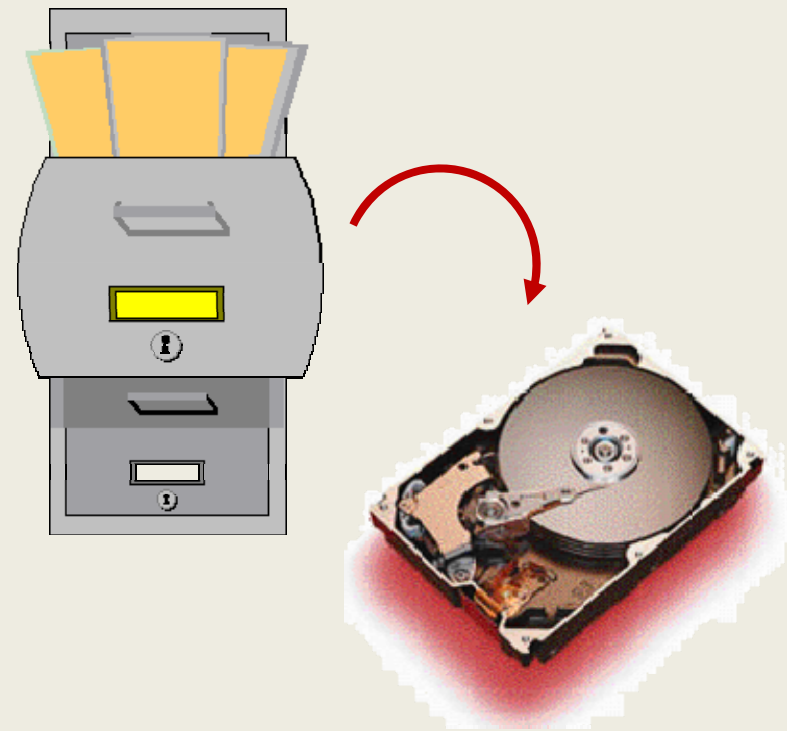
Disciplina INE5609 – Estruturas de Dados

# Objetivos desta Aula

- Estudo de técnicas de armazenamento e recuperação de dados em memória secundária.

- Principais tipos de arquivos:

- Arquivos Seqüenciais;
- Arquivos Indexados;
- Arquivos Invertidos;
- Arquivos de texto.



# Plano de Aula

- Introdução
- Tipos de acesso / dispositivos
- Princípios básicos da Gerência de Arquivos
- Tipos de arquivos
- Principais operações sobre arquivos

# Introdução

- Arquivo
  - Conjunto de dados residente em um  
*Dispositivo de Armazenamento Secundário.*



# Tipos de Acesso / Dispositivos

- Acesso Seqüencial

- Dados podem ser lidos somente na mesma ordem em que foram gravados.
- A gravação dos dados é realizada após a área utilizada.
- Exemplo: Fita Dat

- Acesso Aleatório

- Dados podem ser lidos e gravados\* em qualquer ordem.
- Exemplo: Disco, CD.  
\* à exceção do CD.

# Acesso Seqüencial - Exemplo

- Fita Dat



- Tempo para acesso aos dados é muito grande;
- Utilizada para armazenar dados que serão acessados poucas vezes e não serão alterados.
- Uso típico: Backup.

# Acesso Aleatório - Exemplo

- Disco
  - Tempo para acesso aos dados é menor;
  - Utilizada para armazenar dados que serão freqüentemente acessados e alterados.
  - Uso típico: Banco de dados e sistemas em geral.
- **Base para as técnicas que serão estudadas.**

# Características

## Memória RAM x Memória Secundária

- **Memória RAM**

- **Vantagens:**  
velocidade de acesso aos dados.
- **Desvantagens:**  
volátil.

- **Memória Secundária**

- **Vantagens:** não volátil.
- **Desvantagens:**  
velocidade de acesso aos dados.



# Características

## Memória RAM x Memória Secundária

- A **memória RAM** é utilizada para temporariamente armazenar dados que serão acessados/modificados, isto é, processados;
- A **memória de armazenamento secundário** é utilizada para armazenar dados por um tempo indeterminado, antes e depois de serem processados.

# Características

## Memória RAM x Memória Secundária

- **Solução Intuitiva:**
  - O sistema operacional carrega todos os dados que um programa precisa para a memória RAM.
  - Após o processamento ser realizado, o sistema operacional armazena na memória secundária os dados inseridos/alterados.

- **Problema:**

- Ler e gravar **todos** os dados de um arquivo pode levar muito tempo e talvez eles não caibam na memória;
- Talvez apenas uma parte do arquivo seja necessária.

# Organização de Arquivos

- A organização de arquivos se refere a um conjunto de técnicas para (eficientemente):
  - Classificar arquivos e conteúdos de arquivos;
  - Realizar busca de dados em arquivos;
  - Inserir / Alterar / Eliminar dados em um arquivo;
- Maior restrição: **tempo de acesso aos dados.**

- **Solução** para o problema de leitura / escrita de dados:
  - Trabalhar com dados organizados em arquivos;
  - Trazemos para a memória apenas os dados que queremos alterar;
  - Escrevemos em disco dados que devem ser inseridos no arquivo;
  - Retiramos do disco dados que queremos excluir.

- Maior restrição: **tempo de acesso aos dados.**

### **Solução:**

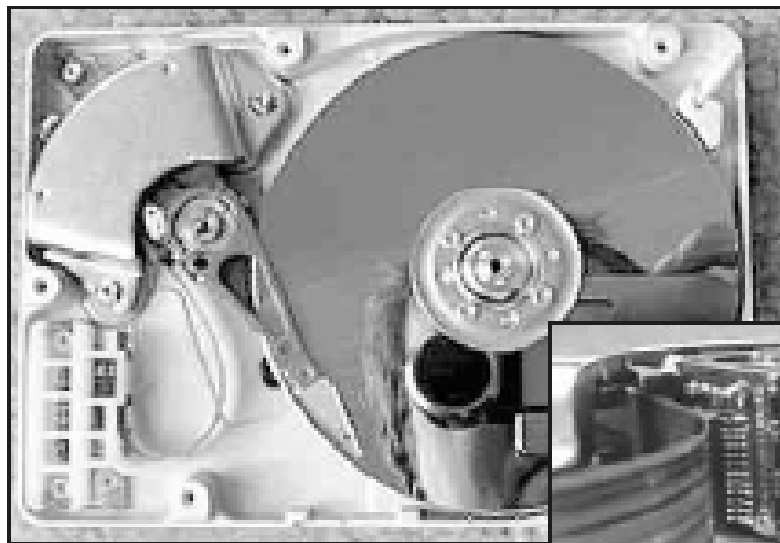
- Organizamos os arquivos de forma a ***minimizar o tempo de acesso*** aos dados;
- ***Eficiência*** na organização tem prioridade maior que economia de espaço (pode ser desperdiçado moderadamente)

# Gerência de Arquivos: Princípio Básico

- Organizar os dados de forma a minimizar o número de acessos à memória secundária para executar uma operação (procura, inserção, alteração, deleção).
- Para isto utiliza-se:
  - Organizações de dados mais ***rígidas*** de que na memória (mesmo que isso signifique desperdício de espaço)
  - técnicas de ***indexação***
  - algoritmos de ***busca***

# Estrutura do Disco

## Estrutura do disco



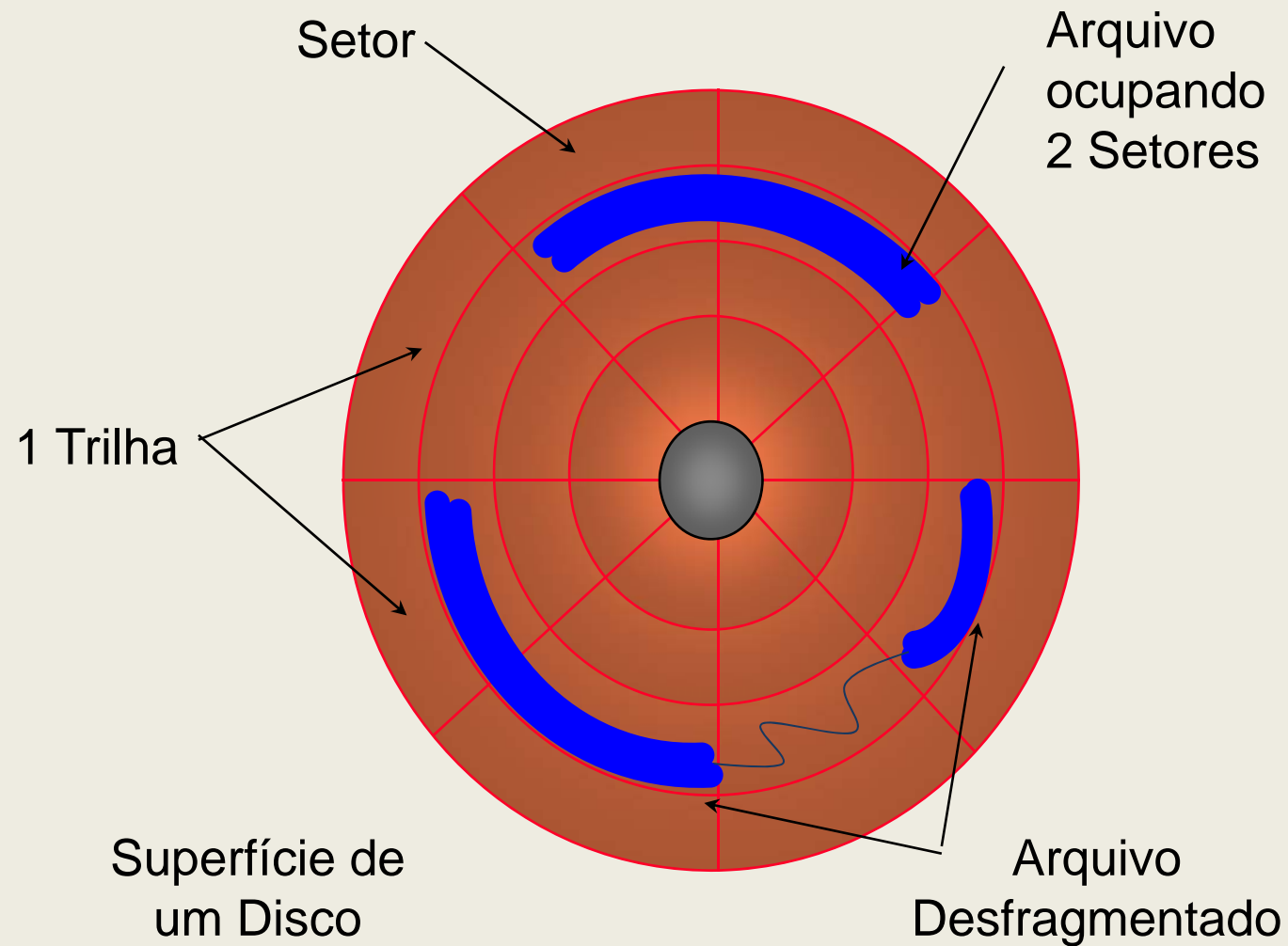


# Estrutura do Disco

- Disco - Exemplo típico para uma memória secundária de acesso aleatório.
- Compreender a forma de armazenamento é importante.
- Modelo de estruturação do disco serve para todos os outros meios secundários.

# Armazenamento em Disco





- **Modelo Simplificado:**

- Dados são organizados em superfícies, trilhas e setores, blocos.
- Um arquivo pode ser imaginado como sendo constituído por uma seqüência de dados no disco.
- Acesso é feito através do posicionamento de um cabeçotes de r/w em qualquer ponto.
- Uma unidade de alocação do disco (um bloco ou um setor) possui um endereço físico no disco.
- Podemos endereçar estes da mesma forma que fazemos com a RAM

- Modelo mais realista:

- O disco pode apresentar **fragmentação**: isto é, os dados não estão exatamente em espaços contíguos, mas em unidades de alocação longe umas das outras, encadeadas como uma lista.
- O encadeamento é tarefa do **sistema operacional** e do **hardware** (controladora) do disco.

# Armazenamento em Disco

- Custo de acesso ao disco:
  - **Tempo de busca (*seek time*):** tempo de deslocamento do mecanismo de acesso de uma trilha para outra;
  - **Latência rotacional:** tempo para que o início do bloco que contenha o registro a ser lido passe pelo cabeçote de leitura/gravação;
  - **Tempo de transferência:** quantidade de tempo necessário para que um bloco (ou setor) seja transferido para um *buffer* de memória.

# Tipos de Arquivos

- Sob o ponto de vista de armazenamento de informações, existe basicamente dois tipos de arquivos:
  - Arquivos de formato variável
  - Arquivos de registro.

# Aplicação Típica


- **Acesso ao arquivo inteiro de formato variável:**
  - Computação gráfica
  - Engenharia de software
  - CAD
  - Documentos
- **Acesso a registros**
  - Banco de dados
  - Processamento transacional



# Arquivos de Formato Variável

- A única unidade de endereçamento possível é o byte.
  - Os arquivos não são organizados em estruturas.
  - Exemplo: arquivo texto.
  - São tratados como arquivos de acesso seqüencial.

# Arquivos de Registros

- São organizados em unidades, chamadas **registros**.
  - São conjuntos de dados de tamanho fixo, mesmo que uma parte não seja aproveitada.
- As técnicas de Organização de Arquivos são voltadas para arquivos do tipo registro.

# Terminologia Empregada

- **Arquivo**
  - Coleção de registros lógicos;
- **Registro lógico**
  - Seqüência de campos ou atributos da entidade ou do objeto sendo modelado;
- **Campos**
  - Corresponde a cada uma das informações que se deseja modelar a respeito da entidade ou objeto considerado. Ex.: Nome, salário, cargo, etc.

# Terminologia Empregada

- Exemplo:

- Entidade: Carro;
- Atributos: cor, marca, modelo, placa.

Quais são os possíveis atributos da entidade Livro?

## Arquivo

| Marca     | Modelo | Cor      | Placa    |
|-----------|--------|----------|----------|
| Ford      | Fiesta | Preta    | CVC 0665 |
| Citroen   | C3     | Vermelho | BGH 9815 |
| Chevrolet | Corsa  | Branco   | SKL 6632 |

Diagram illustrating the structure of a data file (Arquivo) for cars. The table shows columns for Marca, Modelo, Cor, and Placa. Annotations include:

- Orange arrows pointing from 'Fiesta' and 'Vermelho' to the label 'Atributos'.
- A red arrow pointing from the entire table structure to the label 'Registro'.

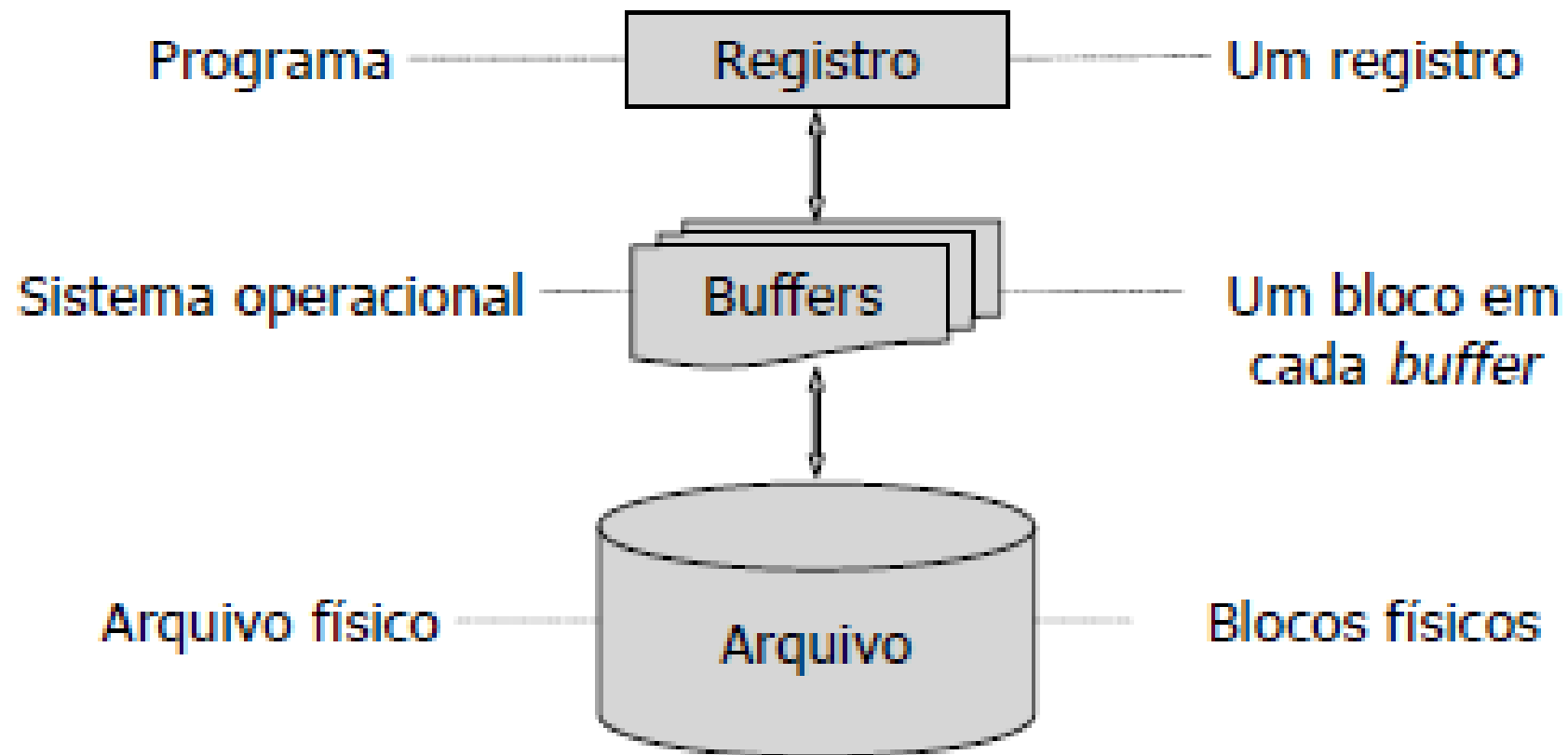
# Terminologia Empregada

- Registro físico
  - O armazenamento de um arquivo é feito, via de regra, por blocos de registros lógicos;
  - Um bloco corresponde a quantidade de dados transferidos em um acesso simples;
  - Um bloco de registros lógicos corresponde a um registro físico.

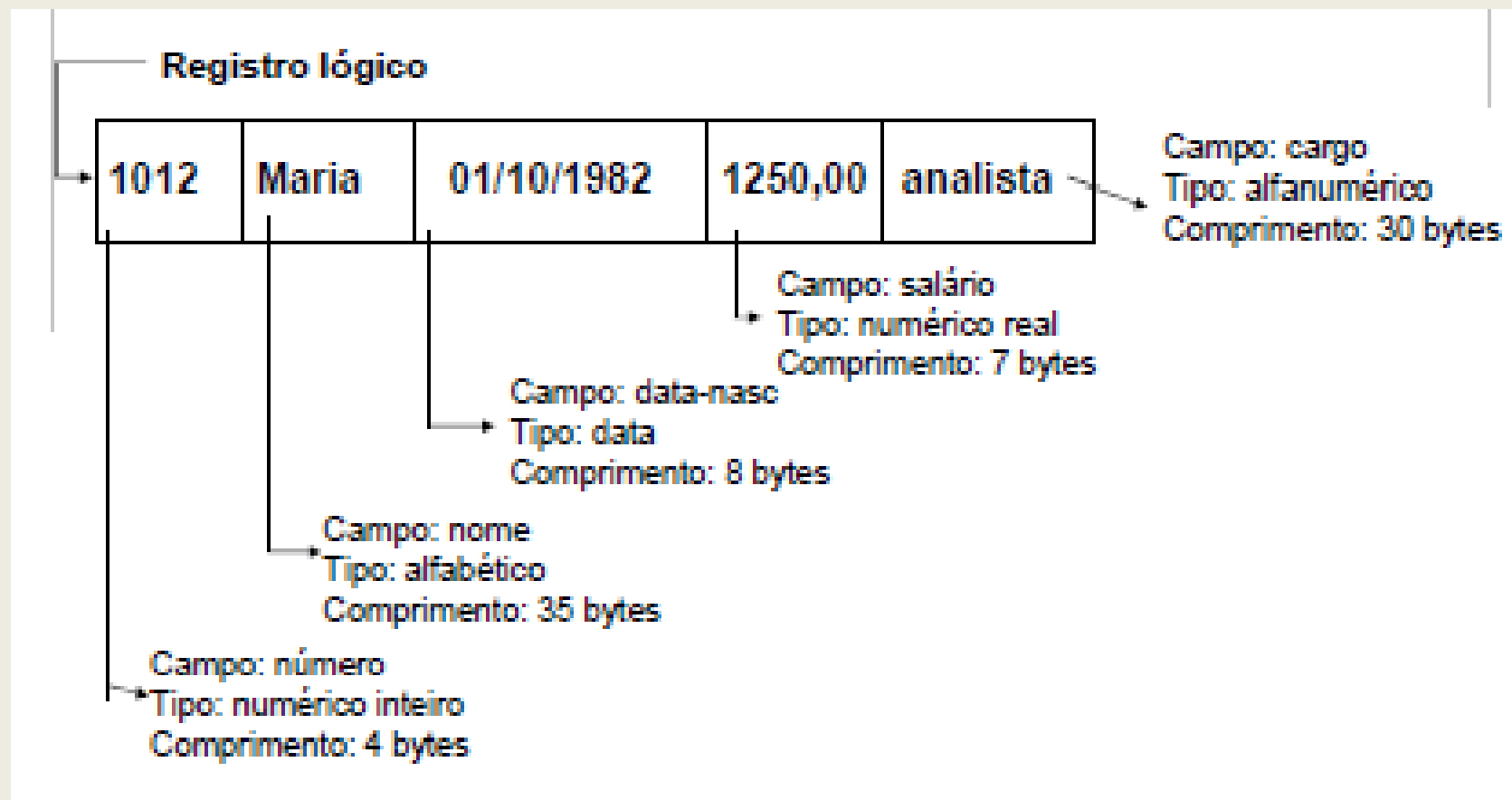
# Terminologia Empregada

- Em cada leitura ou gravação é lido ou gravado um bloco e não apenas um registro lógico
  - Usualmente, o tamanho dos blocos é constante e coincidente com uma unidade de armazenamento do meio físico utilizado (ex.: setores, trilhas em disco magnético).
  - **Fator de bloco:** número (inteiro) de registros lógicos por bloco.

# Estrutura de Acesso



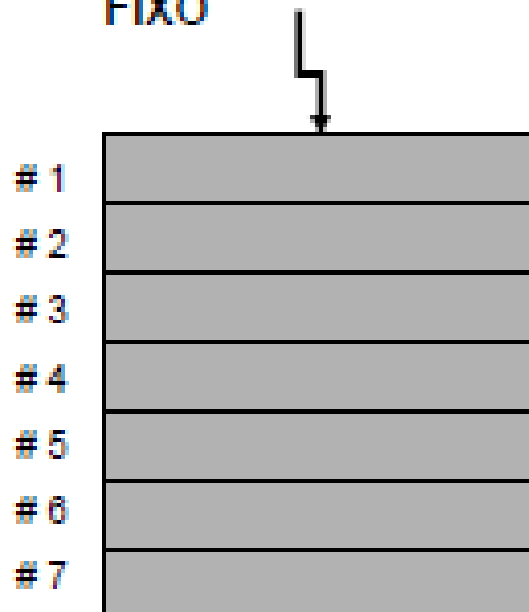
# Exemplo: Arquivo de Funcionário



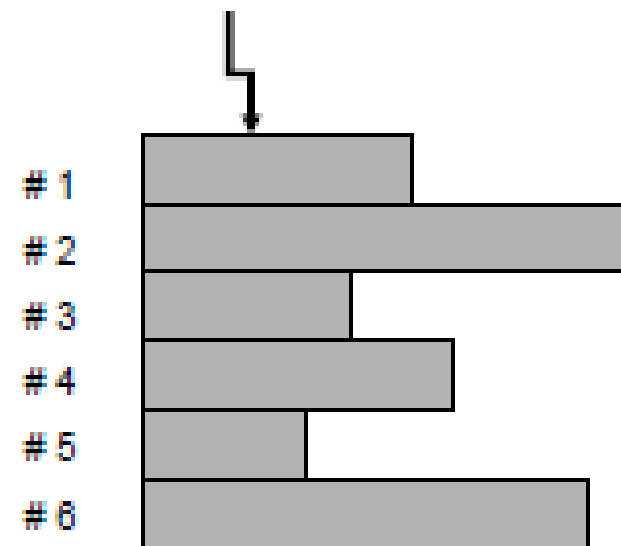


# Tipos de Registros Lógicos

Registro de tamanho  
**FIXO**



Registro de tamanho  
**VARIÁVEL**



# Exemplo de Registros com Campo Variável

|   |   |   |   |   |     |     |     |
|---|---|---|---|---|-----|-----|-----|
| A | B | C | D | E | F01 | F02 | F03 |
|---|---|---|---|---|-----|-----|-----|

A – Código

B – Nome

C – Tamanho do campo D

D – Endereço

E – Ocorrências de F (1 a 10)

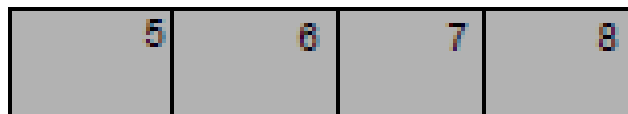
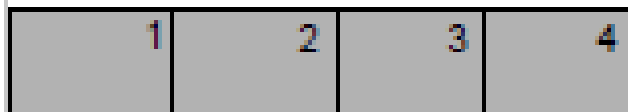
F – Disciplinas

01 Pascal

02 Inglês

03 Organização  
de Dados

# Registros de Tamanho Fixo



- ✳  $nr \div fb$  : bloco onde está o registro
- ✳  $nr \bmod fb$  : posição do registro no bloco
- $nr$  : posição relativa no arquivo
- $fb$  : fator de bloco

# Registros de Tamanho Variável



- Problemas de acesso, exceto o serial;
- Problemas de atualização, podendo haver aumento de tamanho do registro;
- As vezes é conveniente deixar alguns itens em um tamanho padrão, por serem usados com grande frequência. Assim, evita-se a expansão à cada vez que são manipulados.

# Arquivos Seqüenciais

# Arquivo Seqüencial

- Para acessar um dado registro em um arquivo seqüencial é necessário percorrer todos os que o antecedem.
  - É a forma mais simples de implementar operações de inserção, remoção e consulta.
- Os registros podem estar:
  - Desordenados;
  - Ordenados (chaves): fisicamente ou logicamente


# Arquivo Seqüencial

- Desordenado:

|   | PACIENTE | SANGUE | RH | DOAÇÃO   | CONTATO       | PESO |
|---|----------|--------|----|----------|---------------|------|
| 1 | Zé       | A      | +  | 12/12/03 | (79)223.1542  | 80   |
| 2 | Beto     | B      | -  | 12/10/02 | (79)8801.2017 | 70   |
| 3 | Clô      | O      | +  | 11/09/04 | (79)9941.1388 | 72   |
| 4 | Gil      | AB     | -  | 02/02/05 | (81)2212.6711 | 95   |
| 5 | Sara     | A      | -  | 10/01/04 | (11)2213.0909 | 73   |

# Arquivo Seqüencial

- Ordenados Fisicamente:

|   | PACIENTE<br> | SANGUE | RH | DOAÇÃO   | CONTATO       | PESO |
|---|---|--------|----|----------|---------------|------|
| 1 | Beto  | B      | -  | 12/10/02 | (79)8801.2017 | 70   |
| 2 | Clô   | O      | +  | 11/09/04 | (79)9941.1388 | 72   |
| 3 | Gil   | AB     | -  | 02/02/05 | (81)2212.6711 | 95   |
| 4 | Sara  | A      | -  | 10/01/04 | (11)2213.0909 | 73   |
| 5 | Zé  | A      | +  | 12/12/03 | (79)223.1542  | 80   |



# Arquivo Seqüencial

- Ordenado logicamente (por links)

|   | PACIENTE | SANGUE | RH | DOAÇÃO   | CONTATO       | PESO | LINK |
|---|----------|--------|----|----------|---------------|------|------|
| 0 |          |        |    |          |               |      | 2    |
| 1 | Zé       | A      | +  | 12/12/03 | (79)223.1542  | 80   | -1   |
| 2 | Beto     | B      | -  | 12/10/02 | (79)8801.2017 | 70   | 3    |
| 3 | Clô      | O      | +  | 11/09/04 | (79)9941.1388 | 72   | 4    |
| 4 | Gil      | AB     | -  | 02/02/05 | (81)2212.6711 | 95   | 5    |
| 5 | Sara     | A      | -  | 10/01/04 | (11)2213.0909 | 73   | 1    |

# Arquivo Seqüencial - Chave Primária

- Uma chave é um campo usado para identificar um registro como único.
  - Índice é uma estrutura de dados que mantém o conjunto de chaves de um arquivo e o endereço respectivo do registro de cada chave.

# Arquivo Seqüencial

- Registros fisicamente ordenados por uma chave primária ou chave de ordenação.

| Número | Nome   | Idade | Salário |
|--------|--------|-------|---------|
| 1000   | Ademar | 25    | 900     |
| 1050   | Afonso | 34    | 1200    |
| 1075   | Carlos | 51    | 1500    |
| 1100   | Darci  | 42    | 970     |
| 1300   | Elber  | 38    | 875     |
| 1350   | Genaro | 29    | 1340    |
| 1400   | Helena | 20    | 1190    |
| ...    | ...    | ...   | ...     |

# Arquivo Seqüencial

- Indicação de Uso
  - Memória de acesso seqüencial;
  - Indicado para arquivos que sofrem recuperações / atualizações por lotes (em *batch*);
- Contra-indicação
  - Quando há mais de uma chave;
  - Quando exige-se respostas em tempo real;
  - Aplicações com inserções / exclusões arbitrárias.

# Arquivo Seqüencial

## Exercício:

- Há vantagens em manter arquivos seqüenciais ordenados por links?
  - Vantagens em relação a que outra organização?
  - Em relação a que critérios?
    - ✓ Espaço de armazenamento?
    - ✓ Tempo de processamento das operações básicas?
  - Há desvantagens?

# Arquivo Seqüencial Desordenado

## Aplicação de Arquivos Desordenados:

- Na implementação de aplicações cujo tempo destinado ao desenvolvimento é curto.
- Para manter pouco dados, mas de forma permanente.
- Quando os dados são ordenados naturalmente, conforme a entrada destes.

# Arquivo Seqüencial Desordenado


- Operações que podem ser realizadas:
  - Inserção
  - Exclusão
  - Consulta
  - Alteração
  - Classificação
  - Recuperação



## • Inserção

- Os registros são inseridos no final do arquivo.

|   | Paciente | Sangue | RH  | Doação | Peso |
|---|----------|--------|-----|--------|------|
| 0 | Zé       | ...    | ... | ...    | ...  |
| 1 | Maria    | ...    | ... | ...    | ...  |
| 2 | Ana      | ...    | ... | ...    | ...  |
| 3 | João     | ...    | ... | ...    | ...  |
| 4 |          |        |     |        |      |

 Inserção  
ocorre  
aqui.

\* *Somente se não houver preocupação com repetição de chaves.*

## • Exclusão

- Requer do programador cuidado com os espaços livres.
- Os espaços livres podem ser invalidados (exclusão lógica) para posterior reutilização ou compactação (exclusão física).
- A exclusão pode ser efetuada registro a registro ou em lotes.

- **Exclusão - Exercício:**

- Sendo objetivo do programador evitar desperdício de memória, a implementação de exclusões lógicas devem provocar alteração da operação de inclusão em arquivo seqüencial desordenado? Justifique:

- **Busca (Pesquisa):**

- Consiste em exibir os campos:

- ✓ De um registro referente a uma dada chave – busca específica.

- ✓ De todos os registros do arquivo – busca total.

- Depende a organização do arquivo.

- Trata-se de uma operação fim da computação merecendo atenção redobrada do programador.

- **Alteração:**

Para execução desta é necessário:

- Fazer a busca do registro a ser alterado;
- Ler os novos dados;
- Gravar estes (novos) sobre os já existentes.
- É conveniente alterar chaves de ordenação e busca?

# Referências Bibliográficas

- Material Prof. Aldo von Wangenheim (INE/UFSC);
- Material Prof<sup>a</sup>. Kenia Kodel (UFS);