

# ACH2024

## Aula 10

# Organização de arquivos, e acesso em memória secundária

Prof Helton Hideraldo Bís caro

# Arquivos

- O que é um arquivo?

# Arquivos

- O que é um arquivo?
  - Unidade lógica de informação produzida por processos
  - Pode ser visto como um espaço de endereçamento (em disco)

# Arquivos

- O que é um arquivo?
  - Unidade lógica de informação produzida por processos
  - Pode ser visto como um espaço de endereçamento (em disco)
- Quais os tipos de arquivos?

# Arquivos

- O que é um arquivo?
  - Unidade lógica de informação produzida por processos
  - Pode ser visto como um espaço de endereçamento (em disco)
- Quais os tipos de arquivos?
  - Texto (.txt, .doc, ...)
  - Imagem (.tiff, .jpeg, .gif,...)
  - Planilha (.xls, .csv, .txt)

# Arquivos

- Quais são algumas características desejáveis de arquivos (em geral)?

# Arquivos

- Quais são algumas características desejáveis de arquivos (em geral)? Algumas das principais:
  - Persistência: conteúdo precisa ser mantido mesmo quando o computador é desligado, isto é, preciso ser armazenado em um dispositivo de memória não volátil (memória secundária), normalmente o disco
  - Concorrência: múltiplos processos devem poder acessar o mesmo arquivo simultaneamente
  - Capacidade de armazenamento: quero armazenar uma grande quantidade de informação, maior do que a capacidade da memória principal (às vezes até mesmo para um ÚNICO arquivo!)

# Arquivos

- Quais são algumas características desejáveis de arquivos (em geral)? Algumas das principais:
  - Persistência: conteúdo precisa ser mantido mesmo quando o computador é desligado, isto é, preciso ser armazenado em um dispositivo de memória não volátil (memória secundária), normalmente o disco
  - Concorrência: múltiplos processos devem poder acessar o mesmo arquivo simultaneamente
  - Capacidade de armazenamento: quero armazenar uma grande quantidade de informação, maior do que a capacidade da memória principal (às vezes até mesmo para um ÚNICO arquivo!)
    - Quais tipos de arquivos costumam ser grandes assim?



# Arquivos

- Quais são algumas características desejáveis de arquivos (em geral)? Algumas das principais:
  - Persistência: conteúdo precisa ser mantido mesmo quando o computador é desligado, isto é, preciso ser armazenado em um dispositivo de memória não volátil (memória secundária), normalmente o disco
  - Concorrência: múltiplos processos devem poder acessar o mesmo arquivo simultaneamente
  - Capacidade de armazenamento: quero armazenar uma grande quantidade de informação, maior do que a capacidade da memória principal (às vezes até mesmo para um ÚNICO arquivo!)
    - Quais tipos de arquivos costumam ser grandes assim? **Arquivos de DADOS, como planilhas (conjunto de registros – linhas, cada registro com seus campos).**

# Arquivos

- Nas próximas aulas veremos como esses arquivos (conjunto de registros) são organizados no disco, pensando na viabilização de:

- Leitura sequencial e aleatória
- Alteração de registros
- Inserção de registros
- Exclusão de registros

Tudo isso lembrando que o arquivo tem que passar pela memória principal (ou seja, sua organização em disco tem que ser compatível com a estrutura da memória principal).

- Ligação direta com as disciplinas de Sistemas Operacionais e Banco de Dados

# Estrutura interna de arquivos de dados

- Muitos arquivos de dados são uma lista de dados, cada um podendo conter vários campos
  - Ex: linhas de uma tabela, seja de uma planilha, de um .csv, de uma tabela de banco de dados, ...
  - Cada dado normalmente tem uma chave, pela qual é possível realizar operações de busca e ordenação
- Chamaremos cada dado (contendo a chave e demais campos) de **registro**
- Como separar os campos? Como separar os registros?



## Métodos para organização em campos

---

- Comprimento fixo
- Indicador de comprimento
- Delimitadores
- Uso de *tags* (etiquetas)



## Campos com tamanho fixo

- Cada campo ocupa no arquivo um **tamanho fixo**, pré-determinado
- O fato do tamanho ser conhecido garante que **é possível recuperar cada campo**
  - Como?

Maria	Rua 1	123	São Carlos
João	Rua A	255	Rio Claro
Pedro	Rua 10	56	Rib. Preto



## Campos com tamanho fixo

---

```
struct {  
    char last[10];  
    char first[10];  
    char city[15];  
    char state[2];  
    char zip[9];  
} set_of_fields;
```



## Campos com tamanho fixo

---

- Quais as **desvantagens** desta abordagem?



## Campos com tamanho fixo

---

- O espaço alocado (e não usado) aumenta desnecessariamente o tamanho do arquivo (**desperdício**)
  - Solução inapropriada quando se tem uma grande quantidade de dados com tamanho variável
  - Razoável apenas se o comprimento dos campos é realmente fixo ou apresenta pouca variação





## Campos com indicador de comprimento

- O tamanho de cada campo é armazenado imediatamente antes do dado
- Desvantagens desta abordagem?

05	Maria	05	Rua 103	123	10	São Carlos
04	João	05	Rua A	03255	09	Rio Claro
05	Pedro	06	Rua 100	256	10	Rib. Preto



## Campos com indicador de comprimento

- O **tamanho de cada campo** é armazenado imediatamente antes do dado
  - Se o tamanho do campo é inferior a 256 bytes, o espaço necessário para armazenar a informação de comprimento é um único byte
- **Desvantagens** desta abordagem?

05	Maria	05	Rua 103	123	10	São Carlos
04	João	05	Rua A03	255	09	Rio Claro
05	Pedro	06	Rua 100	256	10	Rib. Preto

Gasto de espaço para cada campo (vale a pena ou não dependendo da variabilidade do tamanho dos campos)

Complica o cálculo da posição de um dado campo



## Campos separados por delimitadores

- **Caractere(s) especial(ais)** (que não fazem parte do dado) são escolhido(s) para ser(em) inserido(s) ao final de cada campo
  - Ex.: para o campo *nome* pode-se utilizar /, tab, #, etc...
  - Espaços em branco não servem na maioria dos casos

Maria	Rua 1	123	São Carlos
João	Rua A	255	Rio Claro
Pedro	Rua 10	56	Rib. Preto



## Uso de uma *tag* do tipo "keyword=value"

- Vantagem: o campo fornece **informação semântica** sobre si próprio
  - Fica mais fácil identificar o conteúdo do arquivo
- **Desvantagem:** as *keywords* podem ocupar uma porção significativa do arquivo

```
Nome=Maria|Endereço=Rua 1|Número=123|Cidade=São Carlos|
Nome=João|Endereço=Rua A|Número=255|Cidade=Rio Claro|
Nome=Pedro|Endereço=Rua 10|Número=56|Cidade=Rib. Preto|
```

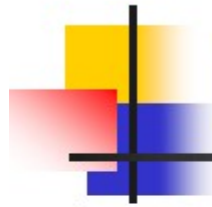


## Métodos para organização em registros

---

- Tamanho fixo
- Número fixo de campos
- Indicador de tamanho
- Uso de índice
- Utilizar delimitadores





## Registros de tamanho fixo

- Analogamente ao conceito de **campos de tamanho fixo**, assume que todos os registros têm o **mesmo tamanho**, com **campos de tamanho fixo ou não**
  - Um dos métodos mais comuns de organização de arquivos

Registro de tamanho fixo e campos de tamanho fixo:

Maria	Rua 1	123	São Carlos
João	Rua A	255	Rio Claro
Pedro	Rua 10	56	Rib. Preto

Registro de tamanho fixo e campos de tamanho variável:

Maria	Rua 1	123	São Carlos	← Espaço vazio →
João	Rua A	255	Rio Claro	← Espaço vazio →
Pedro	Rua 10	56	Rib. Preto	← Espaço vazio →



## Registros com número fixo de campos

- Ao invés de especificar que cada registro contém um tamanho fixo, podemos especificar um **número fixo de campos**
  - O tamanho do registro é **variável**
  - Neste caso, os campos seriam separados por delimitadores

Registro com número fixo de campos:

Maria|Rua 1|123|São Carlos|João|Rua A|255|Rio Claro|Pedro|Rua  
10|56|Rib. Preto|



## Indicador de tamanho para registros

- O indicador que precede o registro fornece o seu **tamanho total**
  - Os campos são separados internamente por **delimitadores**
  - Boa solução para registros de tamanho variável

Registro iniciados por indicador de tamanho:

27Maria|Rua 1|123|São Carlos|25João|Rua A|255|Rio Claro|27Pedro|Rua  
10|56|Rib. Preto|





## Utilizar um índice

- Um **índice externo** poderia indicar o deslocamento de cada registro relativo ao início do arquivo
  - Pode ser utilizado também para calcular o **tamanho dos registros**
  - Os campos seriam separados por **delimitadores**

Arquivos de dados + arquivo de índices:

**Dados:** Maria|Rua 1|123|São Carlos|João|Rua A|255|Rio Claro|Pedro|Rua  
10|56|Rib. Preto|  
**Índice:**    00 27 52



## Utilizar delimitadores

- Separar os registros com **delimitadores** análogos aos de fim de campo
  - O delimitador de campos é mantido, sendo que o método combina os dois delimitadores
  - Note que delimitar fim de campo é diferente de delimitar fim de registro

Registro delimitado por marcador (#):

Maria|Rua 1|123|São Carlos|#João|Rua A|255|Rio Claro|#Pedro|Rua 10|56|Rib.  
Preto|