

Resumo AED II

Arquivo

Unidade lógica de informação produzida por processos, pode ser visto como um espaço de endereçamento (em disco). São características desejáveis aos arquivos:

- **Persistência:** manter seus dados armazenados após o computador haver sido desligado;
- **Concorrência:** estar disponível para ser acessado por múltiplos processos simultaneamente.
- **Maior capacidade de armazenamento**

Estrutura interna de arquivos de dados

Muitos arquivos tratam-se de uma lista de dados, ou *registros*, organizados em **campos**.

Métodos para organização em campos

Comprimento fixo

Cada campo ocupa um dado tamanho fixo em memória. Este tamanho fixo, sendo conhecido, permite recuperar o conteúdo respectivo a cada campo.

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

Desvantagens

Dado um conteúdo de tamanho variável, o espaço alocado e não utilizado para o armazenamento deste conteúdo faz com que o arquivo seja desnecessariamente maior.

Campos com indicador de comprimento

O tamanho de cada campo é especificado logo antes deste.

05	Maria	05	Rua 1	03	123	10	São Carlos
04	João	05	Rua A	03	255	09	Rio Claro
05	Pedro	06	Rua 10	02	56	10	Rib. Preto

Desvantagens

Há um consumo de memória adicional para armazenar os indicadores de comprimento de cada campo, e não mais (comparado ao exemplo anterior) é possível acessar os campos diretamente, apenas sequencialmente.

Campos separados por delimitadores

Seleciona-se um caractere especial para designar o fim de cada campo. Em comparação ao exemplo anterior, há um consumo menor de memória para com o uso de delimitadores.

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

Desvantagens

Percorrer os campos agora exige percorrer o arquivo seus caracteres um a um.

Tags

O campo oferece uma dada informação semântica sobre si.

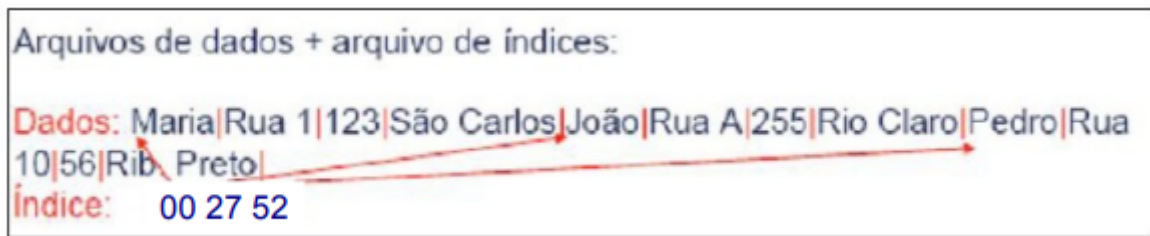
Nome=	Maria		Endereço=	Rua 1		Número=	123		Cidade=	São Carlos	
Nome=	João		Endereço=	Rua A		Número=	255		Cidade=	Rio Claro	
Nome=	Pedro		Endereço=	Rua 10		Número=	56		Cidade=	Rib. Preto	

Desvantagem

O acréscimo das palavra-chave pode ocupar um espaço significativo na memória.

Índices

Um conjunto de índices armazenados externamente ao arquivo, ou em um cabeçalho deste, indica o deslocamento dos campos com relação ao início do arquivo.



Cabeçalho, ou descritor

Conjunto de informações de interesse sobre um dado arquivo, que o acompanha (*metadata*). Informações tais como a descrição dos campos de um registro, data de criação e atualização, adesão um padrão de arquivo (como jpeg, tiff, pdf);

Armazenamento não volátil de arquivos (memória secundária)

Disco: conjunto de "pratos" empilhados nos quais dados são gravados nas superfícies;

Superfícies: são organizadas em trilhas;

Trilhas: são organizadas em setores;

Cilindro: conjunto de trilhas na mesma posição.

Um **setor** é a menor porção endereçável do disco. A divisão de uma trilha em setores é definida pelo disco, e não pode ser alterada.

Seek

Denominação dada à movimentação mecânica do disco necessária à posicionar a cabeça de leitura sobre uma dada trilha qualquer, e tê-la percorrida. São características do seek:

- O conteúdo de todo um cilindro pode ser lido em um único seek.
- É o movimento mais lento da operação de leitura ou escrita e, por isso, deve ser evitado ao máximo. Uma aproximação do tempo de acesso total pode ser calculado por: tempo de acesso total = número de acessos (seeks) × tempo de acesso.

Paginação

Processo em que um conjunto de setores adjacentes entre si é transposto em sua inteiridade para a memória principal, de forma a agilizar na memória principal o acesso de

informações dispostas sequencialmente na memória secundária. Por "página" refere-se ao conteúdo transposto na paginação, e por "moldura de página" refere-se a porção da memória principal destinado a armazenar uma página.

Memória virtual

Uma técnica que usa a **memória secundária** como uma **cache** para armazenamento secundário. Houve duas motivações principais: permitir o compartilhamento seguro e eficiente da memória entre vários programas e remover os transtornos de programação de uma quantidade pequena e limitada na memória principal.

Políticas de reposição de páginas

- Menos Recentemente Utilizada (LRU);
- Menos Frequentemente Utilizada (LFU);
- Ordem de chegada FIFO.

Organização de arquivos

Bloco: unidade de transferência de dados entre a memória principal e a memória secundária.

Alocação sequencial

Blocos são alocados sequencialmente no disco (pelos cilindros)

Vantagens e desvantagens

- Fácil leitura, não é necessário buscar blocos quando estes estão dispostos desta forma.
- Expansão complicada: se não houver espaço disponível até o próximo arquivo, todos os blocos necessitam ser realocados para outro local em que possam ser dispostos desta forma.
- Provoca maior fragmentação externa (buracos na memória entre arquivos).

Métodos de ajuste sequencial

Para encaixar blocos contínuos de arquivos na memória, um dos seguintes ajustes podem ser utilizados:

- **Primeiro ajuste:** Seleciona o primeiro trecho encontrado (a partir do início da lista) grande o suficiente.

- **PRóximo ajuste:** Selecciona o próximo trecho grande o suficiente

Acesso sequencial x acesso direto (aleatório ou randômico)

Apesar do custo de um seek, o acesso é direto, não é necessário ler dados anteriores ou posteriores na mesma trilha para se obter acesso a um dado.

Árvores B

Inserção

Ocorre sempre nas folhas, no caminho da raiz até a folha, nós cheios são subdivididos (a chave do meio na lista de chaves é enviada para o nó atual, um novo nó à direita do nó filho carrega a metade das chaves de maior valor).