

Classificação e Pesquisa de Dados

Aula 22

Organização de Arquivos: Arquivos Sequenciais e Arquivos Sequenciais Indexados

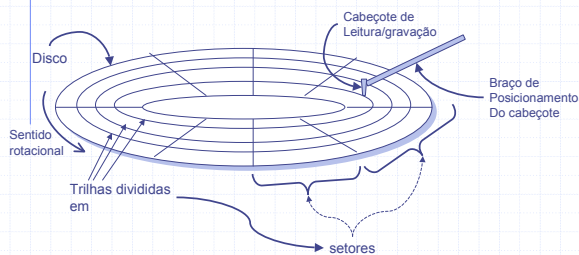
UFRGS

INF01124

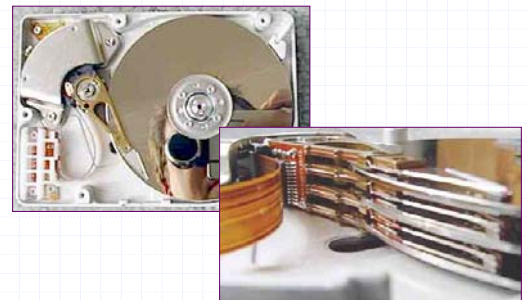
Organização de Arquivos

- ◆ Propósito
 - Estudo de técnicas de armazenamento e recuperação de dados em memória secundária
- ◆ Principais técnicas
 - Arquivos sequenciais
 - Arquivos indexados
 - Arquivos invertidos
 - Arquivos de texto

Armazenamento em disco



Estrutura do disco



Armazenamento em disco: características

- ◆ Várias superfícies de gravação/leitura
- ◆ Cilindro: definido como sendo um conjunto de trilhas verticalmente alinhadas e com mesmo diâmetro
 - Registros que pertencem a um mesmo cilindro não requerem deslocamento do mecanismo de acesso

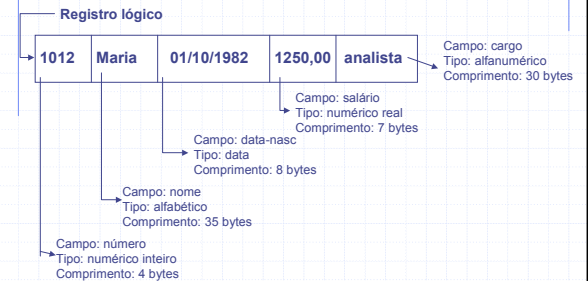
Armazenamento em disco: características

- ◆ 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

Terminologia empregada

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

Arquivo de funcionário



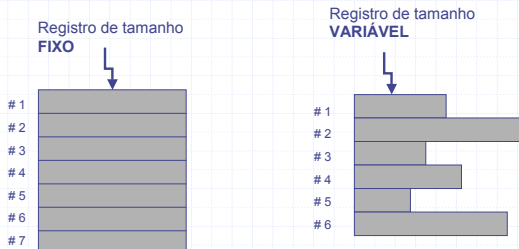
Terminologia empregada

- ◆ 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

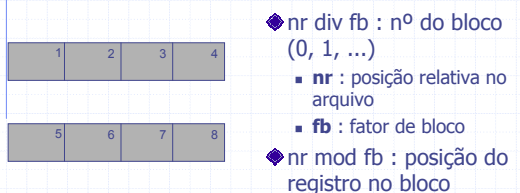
Terminologia empregada

- 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

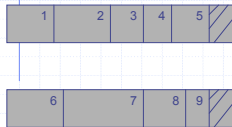
Tipos de registros lógicos



Registros de tamanho fixo



Registros de tamanho variável



- ◆ Problemas de acesso, exceto o serial
- ◆ Problemas de atualização, podendo haver aumento de tamanho do registro

Arquivo seqüencial - AS

◆ Caracterização

- Registros geralmente encontram-se fisicamente ordenados por uma chave primária ou chave de ordenação

Número	Nome	Idade	Salário
1000	Ademar	25	900
1050	Afonso	27	500
1075	Carlos	22	1200
1100	Darci	25	1500
1300	Eber	39	500
1350	Genaro	19	650
1400	Helena	19	420
1440	Maria	21	900
1480	Ramon	22	1340
...

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 do que uma chave
- Quando exige-se respostas em tempo real
- Aplicações com inserções/exclusões arbitrárias

Acesso a um registro

Arquivo Seqüencial

- ◆ Serial ou seqüencial
- ◆ Aleatório

Arquivo seqüencial indexado

◆ Indicação de uso

- Quando o volume de acessos aleatórios torna-se significativo, surge a necessidade de uma estrutura de acesso eficiente para a localização de um registro dado o argumento de pesquisa

Arquivo seqüencial indexado

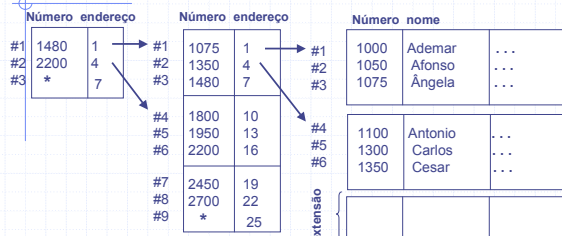
◆ Caracterização

- Esta organização consiste essencialmente de um arquivo seqüencial acrescido de um índice, oferecendo acesso serial ordenado e aleatório eficientes
- Além do arquivo seqüencial e do índice, esta organização ainda prevê uma área de extensão (ou área de *overflow*), utilizada para a implementação da operação de inserção de registros

O que é um índice

- ◆ Estrutura de acesso formada por uma coleção de pares do tipo <chave, endereço>, cada um deles associando o valor de uma chave de acesso a um endereço do arquivo na *memória secundária*
- ◆ Principais vantagens
 - Ocupa menos espaço → maior rapidez de busca
 - Permite acesso ordenado sem rearranjar o arquivo
 - Permite múltiplos caminhos de acesso
 - Oferece maior facilidade de acesso a registros de tamanho variável

Arquivo seqüencial indexado



Para tornar a pesquisa mais eficiente, o índice normalmente é estruturado em vários níveis (árvore B+, por exemplo).

Tipos de Índices

- ◆ Índice primário
 - Índice associado a chave de ordenação
- ◆ Índices secundários
 - Não estão associados a chave de ordenação

Área de extensão

- ◆ Alternativas de implementação
 1. Em cada registro da área principal, associar um campo de elo para conter o endereço da lista encadeada de seus antecessores (ou sucessores), alocados na área de extensão
 2. Em cada bloco de registros, associar um campo de elo, destinado a conter o endereço da lista de extensões do bloco
 - Manutenção da seqüencialidade física dos registros dentro de cada bloco da área principal
 - Todos os registros da lista de extensão de um bloco possuem ordem maior do que todos os da área principal

Acesso a um registro

Arquivo Seqüencial Indexado

- ◆ Serial : diretamente sobre a área de dados
- ◆ Aleatório: é feito via índice

Inserção de um registro

- ◆ Determinação do local onde deve ocorrer a inserção → busca no arquivo via índice
- ◆ Com área de extensão
 - Inserção do registro na lista de extensão do seu sucessor na área principal ou no bloco selecionado
- ◆ Sem área de extensão
 - As inserções são feitas em endereços liberados por exclusões ou reservados para este fim quando da geração do arquivo

Exclusão de um registro

- ◆ Determinação do local onde deve ocorrer a exclusão → busca no arquivo via índice
- ◆ Colocação da marca de excluído

Alteração de um registro

- ◆ Determinação do local onde deve ocorrer a alteração → busca no arquivo via índice
- ◆ Procedimento normal (lê → altera → grava)
 - Quando a alteração não envolve a chave de ordenação e não aumenta o comprimento do registro
- ◆ Procedimento diferenciado
 - Lê → exclui → re-insere registro modificado

Requisitos para reorganização

- ◆ Acessos freqüentes às áreas de extensão
- ◆ Necessidade de desconsideração dos registros excluídos
- ◆ Ponto de reorganização

Procedimento de reorganização

- ◆ Leitura exaustiva e transferência de todos os registros para uma nova área
- ◆ Todos os registros são colocados na área principal, ficando a área de extensão toda livre
- ◆ Registros excluídos desde a última reorganização são retirados fisicamente do arquivo
- ◆ Geração de um novo índice