



Estrutura de Dados II

Organizações Básicas de Arquivos

Prof Márcio Bueno

ed2tarde@marciobueno.com / ed2noite@marciobueno.com

Introdução

- **Armazenamento de pequeno volume de dados**
 - Distribuição simples dos registros em um arquivo
 - **Armazenamento eficiente**
 - Freqüência de acessos aleatórios não deve ser elevada
- **Armazenamento de grande volume de dados e/ou aumento da complexidade dos acessos**
 - Problemas de eficiência no armazenamento dos arquivos e dos acessos aos registros
 - Solução: **Técnicas sofisticadas de armazenamento e recuperação de dados**

Introdução

- **Estratégias de Organização de Arquivos**

- Arquivo Seqüencial Simples
- Arquivo Seqüencial Ordenado
- Arquivo Seqüencial-Indexado
- Arquivo Indexado
- Arquivo Direto
- Arquivo Invertido

Terminologia

- **Arquivo**
 - Coleção de registros lógicos, cada um representando um objeto ou entidade
- **Registro Lógico (Registro)**
 - Seqüência de itens, sendo cada item **campo** ou **atributo**
 - Cada **campo** representa uma **característica** ou propriedade, possui **nome**, **tipo** e **comprimento**
 - Comprimento pode ser **constante** ou **variável**
- **Registro Físico**
 - Armazenamento do arquivo em **bloco de registros lógicos** chamado registro físico (leitura/gravação)
 - Tamanho do bloco coincide com uma unidade de armazenamento do meio físico (Ex.: setor, trilha)
 - Cada bloco armazena um número inteiro de registros

Terminologia

- **Chave**
 - Seqüência de um ou mais campos de um registro
- **Chave Primária**
 - É uma chave que apresenta um valor diferente para cada registro do arquivo
- **Chave Secundária**
 - Difere de uma chave primária pela possibilidade de ter seu valor repetido em diferentes registros
- **Chave de Acesso**
 - É uma chave utilizada para identificar o(s) registro(s) desejado(s) em uma operação de acesso a um arquivo

Terminologia

- **Argumento de Pesquisa**
 - Valor da chave de acesso em uma operação
- **Chave de um Registro**
 - Valor de uma chave primária em um registro
- **Chave de Ordenação**
 - É a chave primária utilizada para estabelecer a seqüência na qual devem ser dispostos (física ou logicamente) os registros de um arquivo

Arquivo Seqüencial Simples

- **Definição**

- Registros são distribuídos em uma **ordem arbitrária**, um após o outro, dentro da área
- Ordem pode ser a mesma da geração dos registros

- **Vantagem**

- Simplicidade

- **Desvantagem**

- Busca de registro através de acesso seqüencial

Arquivo Seqüencial Ordenado

- **Definição**

- Os registros estão dispostos **ordenadamente**, obedecendo à seqüência determinada por uma chave primária, chamada de **chave de ordenação**.

EMPREGADO

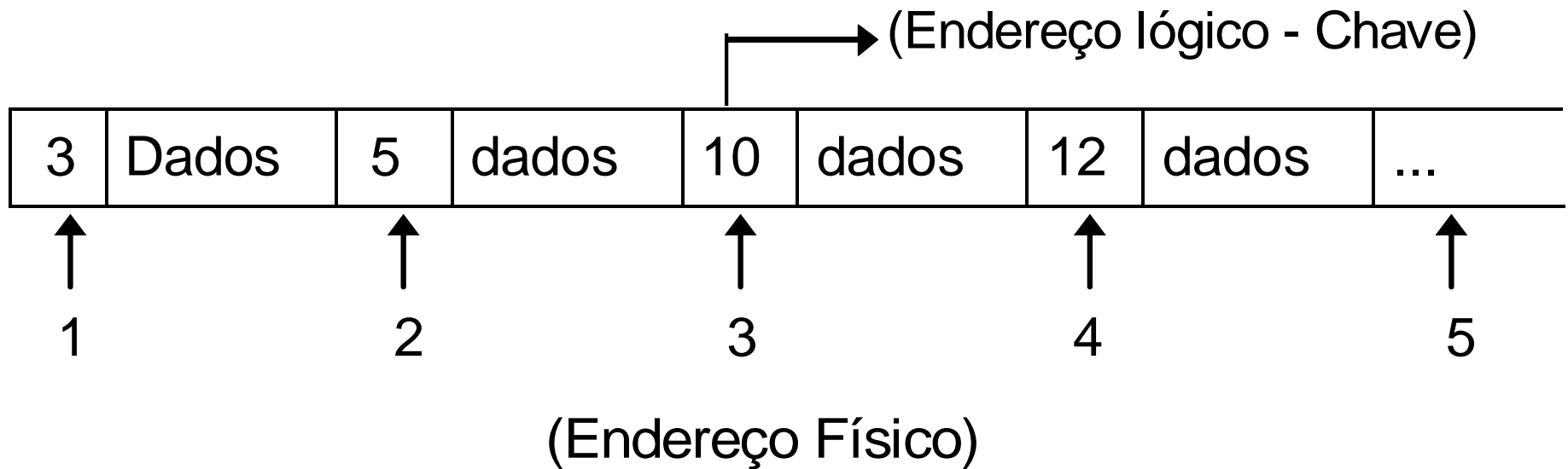
Chave de pesquisa: 1030

| Matricula | Nome | Idade | Salário |
|-----------|----------|-------|---------|
| 1000 | Ademar | 32 | 5000 |
| 1010 | Roberto | 25 | 7500 |
| 1020 | Gerson | 43 | 6000 |
| 1030 | Yeda | 23 | 9000 |
| 1040 | Bernardo | 21 | 4500 |
| 1050 | Ângela | 29 | 5000 |

Chave de ordenação

Arquivo Seqüencial Ordenado

- Estrutura Básica



Arquivo Seqüencial Ordenado

• Principais Características

- Neste tipo de arquivo, os registros são gravados em ordem seqüencial por suas respectivas chaves, havendo uma organização perfeita, tanto lógica quanto física.
- Os registros possuem o mesmo formato, assim cada valor de atributo está associado ao nome do atributo pela sua posição relativa no registro. A descrição do formato de registro, denominada *layout* do registro é externa aos dados que ela descreve. Esta descrição vem declarada nos programas através de declarações de tipos e tamanhos.
- Como o formato é único para todas as ocorrências do registro, campos alfanuméricos são dimensionados pelo tamanho máximo ocorrendo, portanto, desperdício de posições de armazenamento.

Arquivo Seqüencial Ordenado

- **Vantagens**

- Acesso seqüencial eficiente

- Operações de acesso a um registro onde a chave de acesso coincide com a chave de ordenação.
- Operação de exibição dos registros do arquivo na seqüência da chave de ordenação

- **Desvantagens**

- Operações de acesso a um registro onde a chave de acesso não coincide com a chave de ordenação.
- Operações de modificação no arquivo: inserção, alteração e remoção de registros.

Arquivo Seqüencial Ordenado

- Operações

- Acesso a um Registro
- Inserção de um Registro
- Exclusão de um Registro
- Alteração de um Registro
- Leitura Exaustiva dos Registros
- Reorganização do Arquivo

Arquivo Seqüencial Ordenado - Operações

- **Acesso a um Registro**

- **Acesso Seqüencial (serial)**

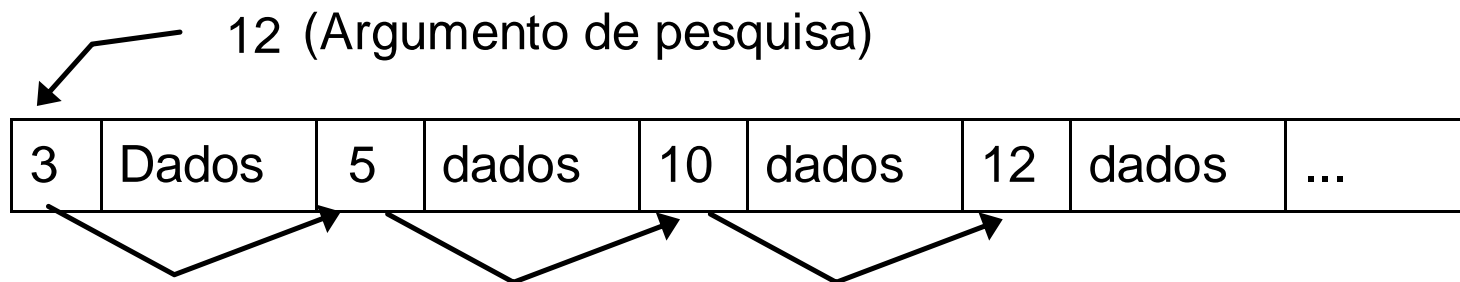
- Consiste na obtenção do registro que segue ao último acessado na seqüência, segundo a chave de ordenação
- **Acesso eficiente**
 - Registros fisicamente armazenados na seqüência de acesso
 - Na maioria dos acessos o registro desejado já estará na memória, por pertencer ao mesmo bloco de seu antecessor

Arquivo Seqüencial Ordenado - Operações

- **Acesso a um Registro**

- **Acesso Seqüencial (serial)**

- *O argumento de pesquisa é comparado com cada registro lido de forma seqüencial.*
- Exemplo: Apresentar os 50 primeiros funcionários pela ordem de matrícula



Arquivo Seqüencial Ordenado - Operações

- Acesso a um Registro

- Acesso Aleatório

- Pesquisa baseada em um argumento de pesquisa
- Seqüência de acesso não mantém necessariamente relação com a ordenação física do arquivo
- Dois tipos:
 - Chave de pesquisa não coincide com chave de ordenação
 - Chave de pesquisa coincide com chave de ordenação

Arquivo Seqüencial Ordenado - Operações

- Chave de pesquisa não coincide com chave de ordenação
 - » Acesso seqüencial
- Chave de pesquisa coincide com chave de ordenação
 - » Em uma mídia de acesso serial, a constatação de registro não encontrado é bem mais rápida
 - » Em uma mídia de acesso direto, pode-se usar pesquisa binária ou por interpolação (mais eficiente)

Arquivo Seqüencial Ordenado - Operações

- **Inserção de um Registro**

- Feita pela técnica de *balance-line*

- Inserir um único registro requer deslocamento dos demais

- *Balance-line*

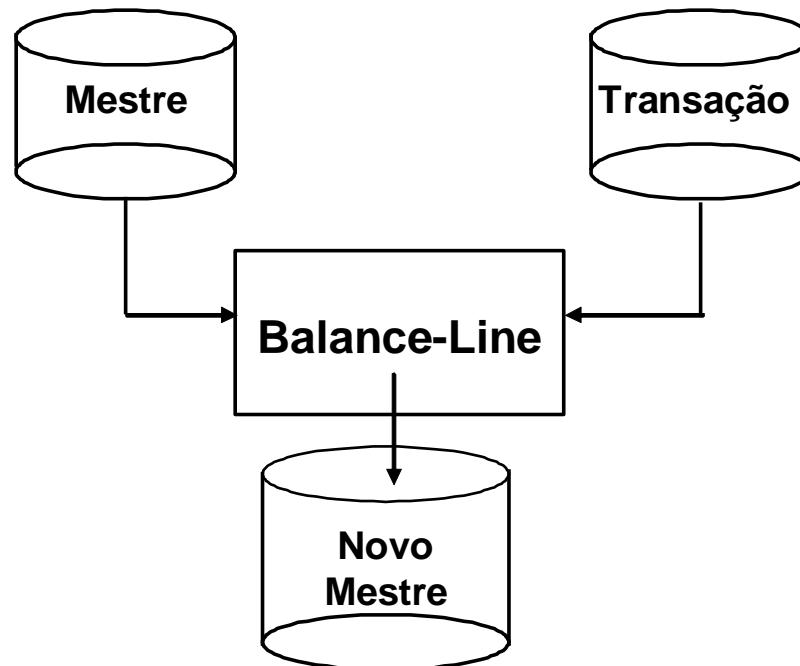
- As inserções e alterações são feitas em um arquivo temporário (são inseridos vários registros e somente depois eles serão inseridos no arquivo original)
- Depois é feita a intercalação do arquivo temporário com o arquivo principal, resultando em um novo arquivo.

Arquivo Seqüencial Ordenado - Operações

• Inserção de um Registro

- Passos:

- Montar um arquivo de transações (ou temporário) com os registros a incluir;
- Ordená-lo da mesma forma que o Arquivo Mestre;
- Intercalar os dois arquivos periodicamente, gerando um Novo Mestre (reorganização do arquivo).



Arquivo Seqüencial Ordenado - Operações

- **Exclusão de um Registro**

- Usa *balance-line*; ou
- Campo adicional, flag, para estado do registro, sendo definido como "excluído"
 - Pesquisa e leitura ignoram tais registros

- **Alteração de um Registro**

- Usa *balance-line*
 - Alteração pode causar aumento do tamanho do registro
 - Alteração pode modificar valor do campo usado como chave de ordenação

Arquivo Seqüencial Ordenado - Operações

- **Leitura Exaustiva dos Registros**
 - Manipula em paralelo os arquivos mestre e transação
- **Reorganização do Arquivo**
 - Operação de intercalação entre os arquivos mestre e transação
 - Operação precede, geralmente, ao uso do arquivo seqüencial para a emissão de relatórios ou efetivação de consultas

Acesso Seqüencial Indexado

- **Motivação**

- Em um arquivo seqüencial, quando o volume de acessos aleatórios torna-se muito grande, é necessário um estrutura de acesso mais eficiente

- **Definição**

- Um arquivo seqüencial-indexado é um arquivo seqüencial acrescido de um índice

Acesso Seqüencial Indexado

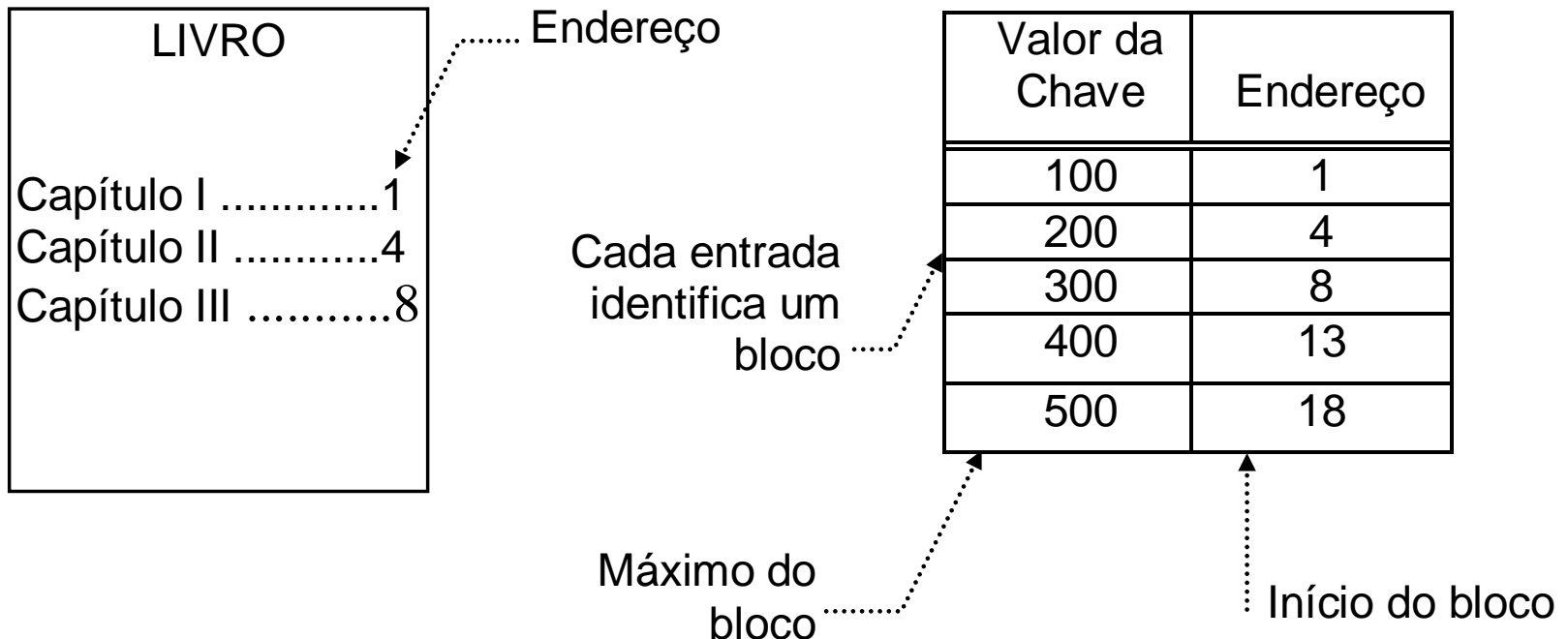
- Um arquivo consta de 3 áreas:
 - Área de índices
 - Arquivo seqüencial criado pelo sistema, no qual cada registro estabelece uma divisão na área primária e contém o endereço do início do segmento e a chave mais alta do mesmo.
 - Desta maneira, o sistema acessa de maneira direta um segmento da área de índices, de forma semelhante a procura de um capítulo de um livro a partir de seu índice.
 - Área primária (principal)
 - Reservada para os registros de dados, classificados em ordem ascendente pelo seu campo chave.
 - Área de excedentes (*overflow*)
 - Reservada para o acréscimo de novos registros que não podem ser colocados na área principal quando se produz uma inserção no arquivo.

Acesso Seqüencial Indexado

- Estrutura

- Índice

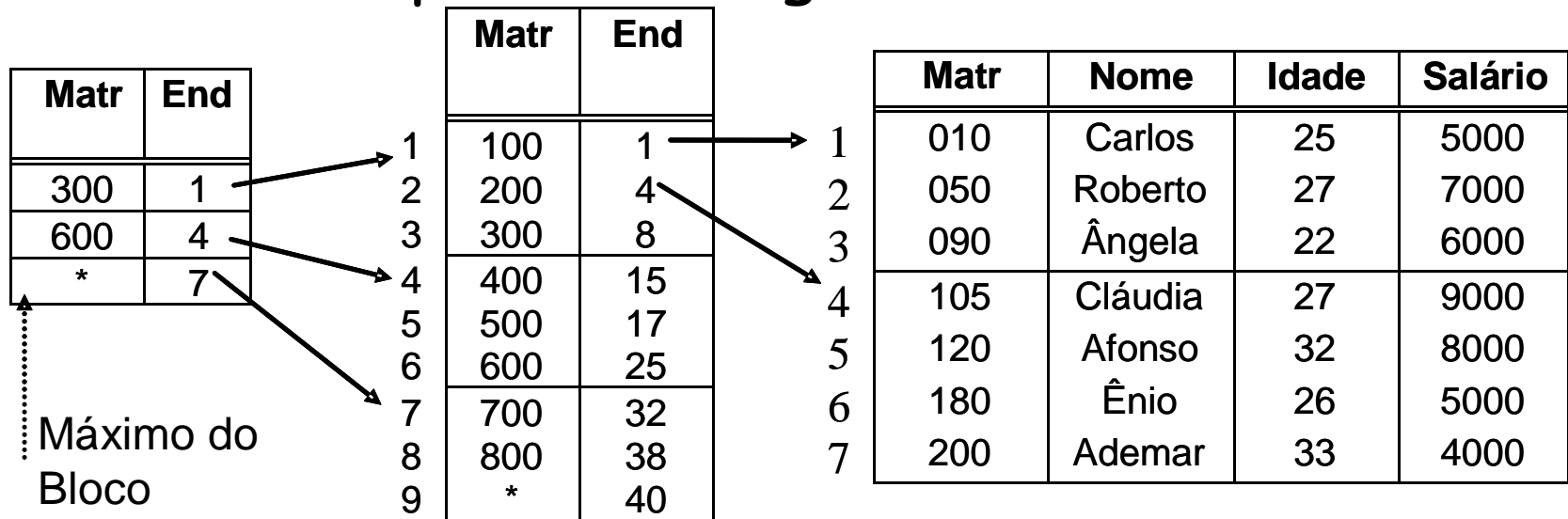
- Um índice é formado por uma coleção de pares, cada um deles associando um valor de chave a um endereço do arquivo.
 - Deve ser especificado um índice para cada chave de acesso.



Acesso Seqüencial Indexado

- Índice

- A finalidade de um índice é permitir a rápida determinação do endereço de um registro, dado um argumento de pesquisa.
- Um índice pode ser organizado em níveis:



Índice em 2 níveis

Acesso Seqüencial Indexado

- Área de Extensão

- Contém os registros inseridos após a criação do arquivo principal.
- Extensão da área principal de dados do arquivo
- São necessárias porque não é viável a implementação da operação de inserção de registros do mesmo modo que nos arquivos seqüenciais
 - Registros mudam de endereço obrigando a uma completa alteração das entradas dos índices

Acesso Seqüencial Indexado

- Área de Extensão

- Pode ser implementado de duas formas:
 - Modelo 1: Cada registro da área de extensão possui um campo elo indicando o seu antecessor (ou sucessor);
 - Modelo 2: Usar um campo de elo para cada bloco de registro contendo a lista de extensões do bloco
- Podem existir várias áreas de extensão
 - Uma para cada bloco ou grupo de blocos adjacentes;
 - Uma ou mais áreas adicionais usadas sempre que ocorre uma inserção em um bloco cuja respectiva área de extensão já está cheia.

• Área de Extensão - Modelo 1

Índice

| | |
|-----|----|
| 600 | 01 |
| 750 | 04 |
| 950 | 07 |
| * | 10 |

Funcionários

| Matrícula | Nome | Idade | Elo |
|-----------|---------|-------|-----|
| 450 | ANA | 25 | |
| 500 | ZÉLIA | 21 | |
| 550 | ÉLIDA | 37 | |
| 630 | ADRIANO | 32 | |
| 680 | CARLOS | 34 | |
| 700 | WAGNER | 30 | |
| 780 | ROBERTO | 25 | |
| 900 | ALBERTO | 40 | |
| 920 | PAULO | 27 | |
| 1000 | JUCA | 31 | |
| 1050 | ALVES | 30 | |
| 2000 | YEDA | 32 | |

| | | |
|------|---------|----|
| 690 | JOÃO | 23 |
| 600 | PEDRO | 25 |
| 1500 | ANTÔNIO | 30 |
| 820 | MARIA | 29 |
| 695 | JOSÉ | 40 |
| 790 | RAMOS | 50 |
| 698 | RONALDO | 31 |
| 440 | JÚLIO | 27 |

1
2
3
4
5
6
7
8
9
10
11
12

ÁREA DE EXTENSÃO

| | | | |
|-----|--|--|--|
| 200 | | | |
| 201 | | | |
| 202 | | | |
| 203 | | | |
| 204 | | | |
| 205 | | | |
| 206 | | | |
| 207 | | | |

Área de Extensão - Modelo 2

Índice

| | |
|-----|----|
| 600 | 01 |
| 750 | 04 |
| 950 | 07 |
| * | 10 |

Funcionários

| | Matrícula | Nome | Idade | Elo |
|----|-----------|---------|-------|-----|
| 1 | 450 | ANA | 25 | |
| 2 | 500 | ZÉLIA | 21 | |
| 3 | 550 | ÉLIDA | 37 | |
| 4 | 630 | ADRIANO | 32 | |
| 5 | 680 | CARLOS | 34 | |
| 6 | 700 | WAGNER | 30 | |
| 7 | 780 | ROBERTO | 25 | |
| 8 | 900 | ALBERTO | 40 | |
| 9 | 920 | PAULO | 27 | |
| 10 | 1000 | JUCA | 31 | |
| 11 | 1050 | ALVES | 30 | |
| 12 | 2000 | YEDA | 32 | |

| | | |
|------|---------|----|
| 690 | JOÃO | 23 |
| 600 | PEDRO | 25 |
| 1500 | ANTÔNIO | 30 |
| 820 | MARIA | 29 |
| 695 | JOSÉ | 40 |
| 790 | RAMOS | 50 |
| 698 | RONALDO | 31 |
| 440 | JÚLIO | 27 |

ÁREA DE EXTENSÃO

| | | | | |
|-----|--|--|--|--|
| 200 | | | | |
| 201 | | | | |
| 202 | | | | |
| 203 | | | | |
| 204 | | | | |
| 205 | | | | |
| 206 | | | | |
| 207 | | | | |

Acesso Seqüencial Indexado - Operações

- **Acesso**

- **Acesso serial**

- Direto sobre a área de dados + extensão sem usar o índice

- **Acesso aleatório**

- Usando o índice
- Pode obter o endereço do próprio registro ou de seu bloco
- No segundo caso, precisa-se de uma busca dentro do bloco e incluir mais acessos à área de extensão

- **Leitura exaustiva** (todos os registros)

- **Igual ao acesso serial!**

Acesso Seqüencial Indexado - Operações

- **Inclusão**
 - Usa as áreas de extensão
 - Dois modos: vide explicação na lista de exemplos
- **Exclusão**
 - Pode ser colocada um marca de excluído no campo "status" do registro
- **Alteração**
 - Pesquisa-se o registro no arquivo;
 - Se a alteração não envolver a chave de ordenação, o registro é gravado na mesma posição;
 - Se envolver a chave de ordenação, usa-se **Exclusão + Inclusão**

Acesso Seqüencial Indexado - Operações

• Reorganização

- Desempenho da operações é degradado à medida que ocorrem um grande número de inclusões e exclusões
- Periodicamente, excluir fisicamente os logicamente excluídos e realizar limpeza na área de extensão;
- Um novo índice deve ser gerado
- O ponto de reorganização deve ser estabelecido, por exemplo, 75% de utilização da área de extensão e em horário que o arquivo não esteja sendo utilizado

Acesso Seqüencial Indexado

- Principais características

- Permite acesso aleatório satisfatório;
- Permite acesso seqüencial eficiente, pela chave primária
 - Exemplo: impressão de relatório de todo estoque
- Permite, com relativa facilidade, as inserções e exclusões, através do uso de áreas de extensão.

Arquivo Indexado

- **Motivação**

- Nos arquivos seqüenciais ordenados manter os **registros fisicamente ordenados**, com o objetivo de prover acesso serial eficiente, dificulta a inserção de um registro exigindo a utilização de áreas de extensão e da efetivação de reorganizações periódicas
- À medida que decresce a freqüência de acessos seriais e cresce os acessos aleatórios, a manutenção da seqüencialidade é inviável

- **Definição**

- Existência de um ou mais índices para acesso aos registros, **não havendo qualquer compromisso com a ordem física de instalação dos registros**
- Considera a possibilidade de acesso por **qualquer campo do registro!**

Arquivo Indexado

- Índice

- Podem existir tantos índices quantas forem as chaves de acesso aos registros.
- Um índice contém um conjunto de entradas ordenadas pela chaves de acesso
 - Permite uma busca mais eficiente e o acesso serial ao arquivo
- Cada entrada contém o valor do atributo e um ponteiro ao endereço do registro que o contém

Arquivo Indexado

- Índice

- Classifica-se em:

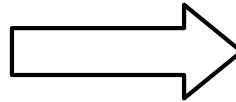
- **Exaustivo**: Quando possui uma entrada para cada registro do arquivo
- **Seletivo**: Uma entrada para cada subconjunto de registros

Arquivo Indexado

- Índice Exaustivo

| NroSerial | NroFunc | Endereço do registro |
|-----------|---------|----------------------|
| 1 | 1000 | 301 |
| 2 | 1066 | 302 |
| 3 | 1140 | 303 |
| 4 | 1162 | 304 |
| 5 | 1234 | 305 |
| 6 | 1305 | 306 |
| 7 | 1312 | 307 |

Método de acesso do SO



| Endereço do reg. | NroFunc | Depto | Salário | Sexo |
|------------------|---------|-------|---------|------|
| 301 | 1000 | A | 1.000 | M |
| 302 | 1066 | B | 2.000 | F |
| 303 | 1140 | A | 1.500 | F |
| 304 | 1162 | C | 1.000 | M |
| 305 | 1234 | C | 2.000 | M |
| 306 | 1305 | C | 2.500 | F |
| 307 | 1312 | A | 3.000 | F |

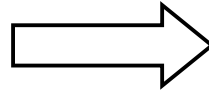
Índice Seletivo

Índice exaustivo (primário):

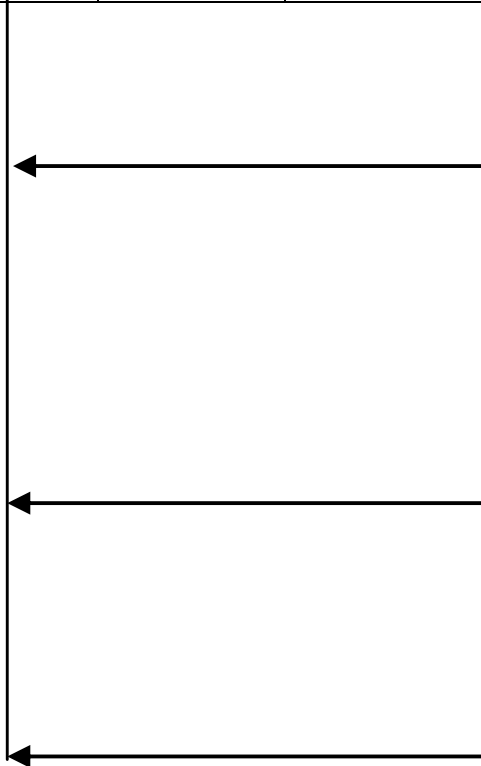
Área de dados

| NroSerial | NroFunc | Endereço do registro |
|-----------|---------|----------------------|
| 1 | 1000 | 301 |
| 2 | 1066 | 302 |
| 3 | 1140 | 303 |
| 4 | 1162 | 304 |
| 5 | 1234 | 305 |
| 6 | 1305 | 306 |
| 7 | 1312 | 307 |

Método de acesso do SO



| Endereço do reg. | NroFunc | Depto | Salário | Sexo |
|------------------|---------|-------|---------|------|
| 301 | 1000 | A | 1.000 | M |
| 302 | 1066 | B | 2.000 | F |
| 303 | 1140 | A | 1.500 | F |
| 304 | 1162 | C | 1.000 | M |
| 305 | 1234 | C | 2.000 | M |
| 306 | 1305 | C | 2.500 | F |
| 307 | 1312 | A | 3.000 | F |



| Salário | Nro serial do índice primário |
|---------|-------------------------------|
| 1.000 | 1, 4 |
| 1.500 | 3 |
| 2.000 | 2, 5 |
| 2.500 | 6 |
| 3.000 | 7 |

Índice seletivo (salário)

| Depto | Nro serial do índice primário |
|-------|-------------------------------|
| A | 1, 3, 7 |
| B | 2 |
| C | 4, 5, 6 |

Índice seletivo (depto)

| Salário | Nro serial do índice primário |
|---------|-------------------------------|
| M | 1, 4, 5 |
| F | 2, 3, 6, 7 |

Índice seletivo (sexo)

Acesso Indexado - Operações

- **Acesso**

- **Acesso serial**

- Utilizando-se o **índice apropriado**, onde a identificação é simplificada, pois as entradas dos índices são ordenadas.
- Neste caso, a memória guarda um bloco do índice, reduzindo o número de leituras a disco.
- Os índices são guardados em um único cilindro, mesmo que o endereço encontrado esteja em um cilindro diferente.

- **Acesso aleatório**

- Requer uma busca sobre o índice.

- **Leitura exaustiva (todos os registros)**

- Para uma leitura exaustiva, são realizados sucessivos acessos seriais sobre índices exaustivos.

Acesso Indexado - Operações

- **Inclusão**

- O registro é armazenado em qualquer endereço vago;
- Os seus pares são inseridos nos índices correspondentes;
- Para o tratamento dos índices é utilizada a árvore-B

- **Exclusão**

- É liberada a área de dados ocupada e são removidas as entradas correspondentes a este registro

- **Alteração**

- Primeiro busca-se o registro
- Em seguida os campos são alterados e gravados na mesma posição

Arquivo Indexado

- **Vantagens**

- Operação de inserção mais eficiente.
- Possibilidade de acessos aleatórios via índices.

- **Desvantagens**

- Acesso serial ineficiente.
- Necessidade de manutenção de índice (inserções ou alterações envolvendo atributos associados aos índices)

Arquivo Direto

- **Motivação**

- Acesso rápido aos registros especificados por argumentos de pesquisa , sem percorrer uma estrutura auxiliar (índice)

- **Definição**

- Em um arquivo direto, ao invés de um índice é utilizada uma função (*hashing*) que calcula o endereço do registro a partir do valor da chave do registro.

Arquivo Direto

EMPREGADO

| | Matr | Nome | Idade | Salário |
|---|-------------|-------------|--------------|----------------|
| 1 | 1100 | ANTÔNIO | 28 | 8500 |
| 2 | | | | |
| 3 | 1440 | CRISTIAN | 30 | 9500 |
| 4 | 1800 | EDSON | 34 | 6500 |
| 5 | | | | |
| 6 | 2950 | SANDRA | 25 | 9000 |
| 7 | 1950 | FLÁVIA | 39 | 9000 |
| 8 | | | | |
| 9 | 3150 | TATIANA | 20 | 10000 |

Argumento de Pesquisa

C = 3150

$E = F(c)$

onde

E = Endereço

F = Função matemática

C = Chave primária

Arquivo Direto

- **Semelhança com o Arquivo Indexado**
 - O acesso aleatório é eficiente
- **Diferenças para o Arquivo Indexado**
 - No indexado, o endereço é independente do valor da chave
 - Não são previstos, no arquivo direto, acessos seriais.

Arquivo Direto

- Cálculo do Endereço - Determinação da função F
 - **Funções Determinísticas**
 - Dada uma chave de acesso, gera um único endereço.
 - Em termos práticos não despertam maiores interesses.
 - **Funções Probabilísticas**
 - Para cada valor da chave de acesso, gera um endereço tão único quanto possível.
 - Quando houver coincidência (duas chaves gerando o mesmo endereço) este fato é denominado de COLISÃO.
 - Objetivo das funções probabilísticas
 - Preservar a ordem dos registros
 - Aumentar o grau de unicidade (uniformidade) dos registros sobre o arquivo.

Arquivo Direto

• Cálculo do Endereço - Determinação da função F

- Exemplo 1

- Considere que o número das matrículas dos funcionários esteja entre 900 e 3150 e os endereços disponíveis estejam entre 1 e 37. A função escolhida para gerar estes endereços foi:

- **Função:** $E(x) = \frac{\text{Chave} - \text{Min}}{(\text{Max} - \text{Min}) / 37} + 1$

$(\text{Max} - \text{Min}) / 37 \longrightarrow$ *Refere-se ao maior endereço disponível.*

- Considerando o exemplo, a função seria $E(x) = \frac{x-900}{61} + 1$ \longrightarrow *Valor arredondado*

- Se as chaves de acesso forem:

1000

1400

1600

- Teremos os endereços:

2

9

12

Arquivo Direto

- Cálculo do Endereço - Determinação da função F

- Exemplo 1 - continuação

- $E(1000) = \frac{(1000-900)}{(3150 - 900) / 37} + 1 = 2$

- $E(1400) = \frac{(1400-900)}{(3150 - 900) / 37} + 1 = 9$

- $E(1600) = \frac{(1600-900)}{(3150 - 900) / 37} + 1 = 12$

Arquivo Direto

- Cálculo do Endereço - Determinação da função F

- Exemplo 1 - continuação

EMPREGADO

| | Matrícula | Nome | Idade | Salário |
|-----------|-------------|----------|-------|---------|
| 1 | 900 | ANTÔNIO | 28 | 8500 |
| 2 | 1000 | ROBERTO | 36 | 9500 |
| 3 | 1010 | CRISTIAN | 30 | 9500 |
| 4 | 1100 | EDSON | 34 | 6500 |
| 5 | | | | |
| 6 | 1200 | SANDRA | 25 | 9000 |
| 7 | 1300 | FLÁVIA | 39 | 9000 |
| 8 | | | | |
| 9 | 1400 | TATIANA | 20 | 10000 |
| 10 | 1480 | MARIA | 25 | 8000 |
| 11 | | | | |
| 12 | 1600 | DIOGO | 23 | 5000 |

Arquivo Direto

• Cálculo do Endereço - Determinação da função F

- Exemplo 2

- Posso ter funções que não preservam a ordem dos registros.
- São as chamadas *funções de aleatorização*.

$$E(x) = (x \text{ módulo } 31) + 1$$

| C r e s c e n t e | O r d e m | ↓ | Argumento de pesquisa | Endereço encontrado |) Endereços em ordem aleatória |
|---|-----------------------|---|-----------------------|---------------------|--------------------------------|
| | | | 1000 → | 09 | |
| | | | 1050 → | 25 | |
| | | | 1075 → | 22 | |
| | | | 1100 → | 16 | |
| | | | 1300 → | 30 | |

Arquivo Direto

• Tratamentos de Colisão

- Tratamento por Endereçamento Aberto

- O endereço colidido é guardado no primeiro endereço livre.
- Exemplo: $E(1440) = 1440 - 900 / 61 + 1 = 9$

1º endereço livre

| | | | | |
|----|-------------|---------|----|-------|
| 7 | 1300 | FLÁVIA | 39 | 9000 |
| 8 | | | | |
| 9 | 1400 | TATIANA | 20 | 10000 |
| 10 | 1480 | MARIA | 25 | 8000 |
| 11 | | | | |
| 12 | 1600 | DIOGO | 23 | 5000 |



| | | | | |
|----|-------------|--------------|-----------|-------------|
| 7 | 1300 | FLÁVIA | 39 | 9000 |
| 8 | | | | |
| 9 | 1400 | TATIANA | 20 | 10000 |
| 10 | 1480 | MARIA | 25 | 8000 |
| 11 | <i>1440</i> | <i>PAULO</i> | <i>23</i> | <i>6000</i> |
| 12 | 1600 | DIOGO | 23 | 5000 |

Arquivo Direto

- **Tratamentos de Colisão**

- **Tratamento por Encadeamento**

- Procura-se um endereço, colocando-se um elo para encadeá-lo ao anterior. Neste caso, duas alternativas podem ser adotadas:

- **Encadeamento puro**: os que colidem formam uma lista encadeada na área de dados
- **Utilizando áreas de extensão**: semelhante ao usado no arquivo seqüencial indexado

Arquivo Direto - Operações

- **Acesso serial**
 - Acesso serial: só é possível se for utilizada uma função que preserve a ordem dos registros. Neste caso, para o acesso serial, basta ler a área de dados.
 - A leitura exaustiva segue o mesmo requisito.
- **Acesso aleatório**
 - Aplicar a função de cálculo
- **Inserção**
 - Empregar a função de cálculo
- **Exclusão**
 - É acessado o registro e colocada a marca de excluído
- **Alteração**
 - Se não envolver a chave de acesso: localizar e alterar.
 - Caso contrário, excluir e inserir

Arquivo Invertido

- **Motivação**

- As organizações de arquivos vistas até aqui foram voltadas para a chave primária.
- Existem, entretanto, organização de arquivos voltada para chaves secundárias que visam resolver o problema de um conjunto de registros.
- A cada um dos valores da chave de acesso é associada uma lista de identificação de registros, chamada lista invertida.

- **Estrutura de um arquivo invertido**

- **Inversão**: é o conjunto de listas invertidas associadas a uma chave de acesso.
- Um arquivo pode ter uma ou mais inversões.

Arquivo Invertido

- **Estrutura de um arquivo invertido**

- Exemplo: arquivo com uma **inversão associada a IDADE**

EMPREGADO

| | Matr | Nome | Idade | Salário |
|---|-------------|-------------|--------------|----------------|
| 1 | 1000 | ANTÔNIO | 25 | 5000 |
| 2 | 1050 | AFONSO | 20 | 7000 |
| 3 | 2400 | CRISTIAN | 25 | 5500 |
| 4 | 1850 | EDSON | 21 | 5500 |
| 5 | 1440 | YEDA | 22 | 5000 |
| 6 | 3150 | SANDRA | 22 | 7000 |
| 7 | 2000 | FLÁVIA | 20 | 5500 |
| 8 | 1900 | ROBERTO | 20 | 10000 |

Inversão

| IDADE | ENDEREÇO |
|--------------|-----------------|
| 20 | 2 7 8 |
| 21 | 4 |
| 22 | 5 6 |
| 25 | 1 3 |

Arquivo Invertido

- **Estrutura de um arquivo invertido**
 - **Vantagem**
 - Permite o acesso direto a um conjunto de registros
 - **Desvantagem**
 - As listas só são válidas para aquela disposição física
 - Se o arquivo sofrer uma reorganização, as inversões terão que ser regeradas
 - Para superar esta desvantagem, **pode-se implementar as listas por chaves primárias.**
 - Tem-se uma perda de eficiência.

Arquivo Invertido

- **Estrutura de um arquivo invertido**
 - Exemplo 2: Inversão associada a SALÁRIO, por chave primária da entidade Empregado

| SALÁRIO | NÚMEROS | | |
|----------------|----------------|------|------|
| 5000 | 1000 | 1440 | |
| 5500 | 1850 | 2000 | 2400 |
| 7000 | 1050 | 3150 | |
| 10000 | 1900 | | |

Arquivo Invertido

• Estrutura de um arquivo invertido

- Mapa de bits

- Outra opção para a implementação dos arquivos
- Exemplo 3: **Inversão associada ao SALÁRIO**

Os que recebem 5000

| 5000 | 5500 | 7000 | 10000 | | Matr | Nome | Idade | Salário |
|------|------|------|-------|---|------|----------|-------|---------|
| 1 | 0 | 0 | 0 | 1 | 1000 | ANTÔNIO | 25 | 5000 |
| 0 | 0 | 1 | 0 | 2 | 1050 | AFONSO | 20 | 7000 |
| 0 | 1 | 0 | 0 | 3 | 2400 | CRISTIAN | 25 | 5500 |
| 0 | 1 | 0 | 0 | 4 | 1850 | EDSON | 21 | 5500 |
| 1 | 0 | 0 | 0 | 5 | 1440 | YEDA | 22 | 5000 |
| 0 | 0 | 1 | 0 | 6 | 3150 | SANDRA | 22 | 7000 |
| 0 | 1 | 0 | 0 | 7 | 2000 | FLÁVIA | 20 | 5500 |
| 0 | 0 | 0 | 1 | 8 | 1900 | ROBERTO | 20 | 10000 |

Bibliografia

- FURTADO, Antonio Luz; SANTOS, C. S. Dos. **Organização de bancos de dados**. 6. ed. Rio de Janeiro: Campus, 1986.