

IFSULDEMINAS, *campus* Muzambinho
Curso de Ciência da Computação

Formas Básicas de Organização de Arquivos

Prof. Ricardo José Martins
ricardo.martins@muz.ifsuldeminas.edu.br



Formas Básicas de Organização

Tipos de Arquivos

- Arquivo Seqüencial
- Seqüencial Indexado
- Indexado
- Direto



Formas Básicas de Organização

■ Arquivo Seqüencial

- Em um arquivo seqüencial os registros estão dispostos ordenadamente, obedecendo à seqüência determinada pela chave primária, chamada chave de ordenação.
- O acesso a um registro, dado o argumento de pesquisa, é facilitado se a chave de acesso coincide com a chave de ordenação, pois nos demais casos não há vantagens.
- **Acesso a um registro**
 - **acesso serial** : consiste na obtenção do registro que segue o último acessado, na seqüência ditada pela chave de ordenação.
 - **Acesso aleatório** : a indicação do registro a ser acessado é feita pela especificação de um argumento de pesquisa

Se a chave de acesso coincide com a chave de ordenação, torna-se viável a utilização de um método de pesquisa mais eficiente que a pesquisa seqüencial. Um método bastante conhecido é a **pesquisa binária**, no qual o primeiro registro a ser acessado é aquele que ocupa a posição média do arquivo.



Formas Básicas de Organização

■ Arquivo Seqüencial

□ Pesquisa binária

Se a chave do registro for igual ao argumento de pesquisa, esta termina com sucesso; caso contrário, ocorre uma das duas seguintes situações:

- a chave do registro é **maior** do que o argumento de pesquisa e o processo de pesquisa é repetido para a **metade inferior** do arquivo.
- a chave do registro é **menor** do que o argumento de pesquisa e o processo de pesquisa é repetido para a **metade superior** do arquivo.

A busca é encerrada sem sucesso quando a área de pesquisa assumir comprimento zero. O número máximo de comparações será: **$\log_2 N + 1$**



Formas Básicas de Organização

■ Arquivo Seqüencial

Chave de Ordenação				Atributos ou campos	
	Nome	Idade	Salário		
Ademar		34	1,500	registros	
Francisco		45	2,500		
Paulo		22	1,000		
Pedro		30	2,900		
Sônia		26	2,300		

Arquivo Seqüencial



Formas Básicas de Organização

■ Arquivo Seqüencial

□ Inserção de um registro

- A maneira usual de processar inserções é por meio de um arquivo de transações contendo os registros a serem inseridos, ordenados pela mesma chave do arquivo original.
- Isto se deve porque a inserção de um registro isolado apresenta um custo proibitivo, pois todos aqueles registros com chaves superiores ao inserido seriam necessariamente deslocados
- O arquivo de transações pode ser usado como uma extensão do arquivo original, até assumir um tamanho que justifique a operação de **reorganização** (intercalação).

□ Exclusão de um registro

- É usualmente implementada como a inserção, sendo as indicações de inclusão coletadas em um **arquivo de transações**.
- Pode ainda ser implementada uma remoção lógica do registro, como o uso de um campo adicional que indique o estado de cada registro.



Formas Básicas de Organização

■ Arquivo Seqüencial

□ Alteração de um registro

- Consiste na modificação de um ou mais atributos de um registro. O registro deve ser localizado e seus campos alterados, sendo, após, gravado novamente.
- Se alteração modifica o valor do campo (chave de ordenação) que determina a seqüência dos registros no arquivo, implica em uma remoção do registro antigo e inserção de um novo registro.

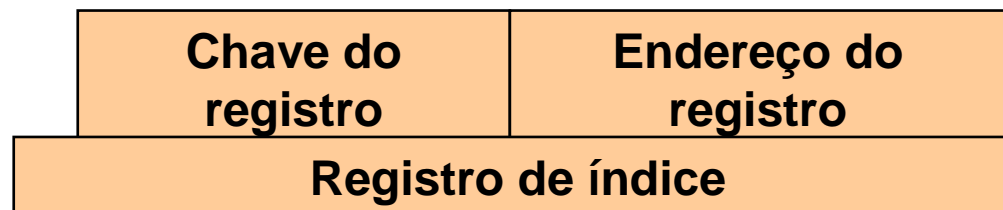
□ Reorganização do arquivo

- Consiste na geração de um arquivo seqüencial A , obtido da intercalação do arquivo original (S) com o arquivo de transações (T). É feita uma leitura exaustiva de S , juntamente com T , e os registros transferidos para o arquivo A ($A = S + T$)

Formas Básicas de Organização

■ Arquivo Seqüencial Indexado

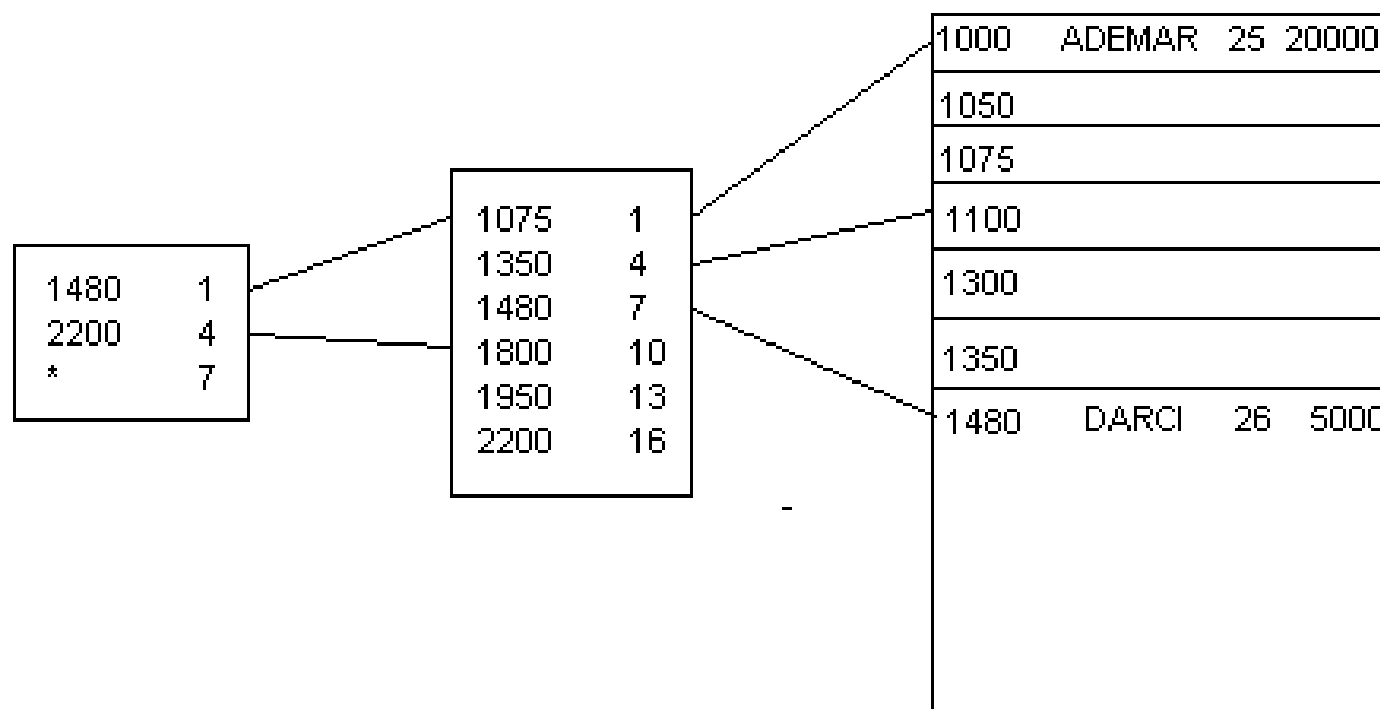
- Quando em um arquivo seqüencial o volume de acesso torna-se muito grande, configura-se a necessidade de utilização de uma estrutura de acesso, associada ao arquivo, que ofereça maior eficiência na localização do registro desejado.
- Um arquivo seqüencial acrescido de um índice (estrutura de acesso) constitui um **Arquivo Seqüencial Indexado**.
- Um **índice** é formado por uma coleção de pares, cada um deles associando um valor da chave de acesso a um endereço no arquivo.
- Um índice ocupa um espaço bem menor que o registro de dados correspondente, com isto, a pesquisa pode ser feita com maior rapidez que se fosse feita diretamente sobre o arquivo de dados.





Formas Básicas de Organização

I



Arquivo seqüencial indexado estruturado em dois níveis



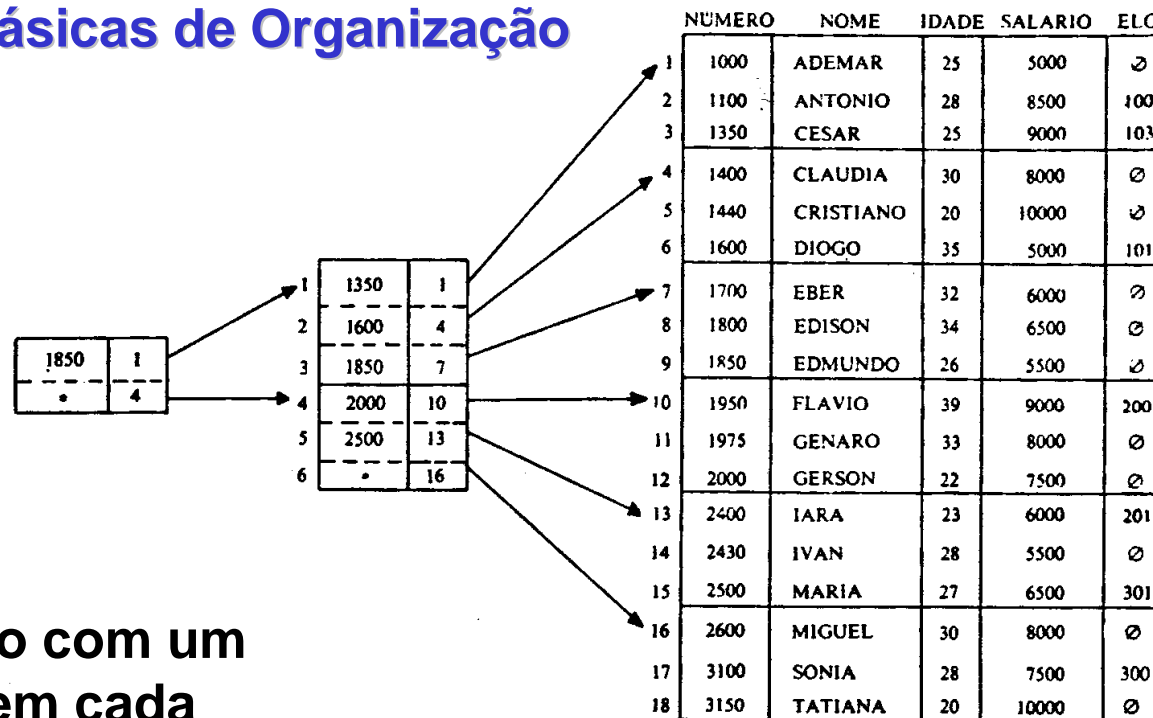
Formas Básicas de Organização

■ Arquivo Seqüencial Indexado

■ Área de extensão

- Destina-se a conter os registros inseridos, em um arquivo seqüencial indexado. Ela constitui uma extensão da área principal de dados do arquivo de dados.
- Elas são necessárias porque não é viável a implementação da operação de inserção de registro do mesmo modo que nos arquivos seqüenciais.
- 2 alternativas para implementação da área de extensão :
 - Adicionar um campo de elo em cada registro para conter o endereço da lista encadeada de seus antecessores na área de extensão.
 - Adicionar um campo de elo para cada bloco de registros, destinado a conter o endereço da lista de extensões do bloco.

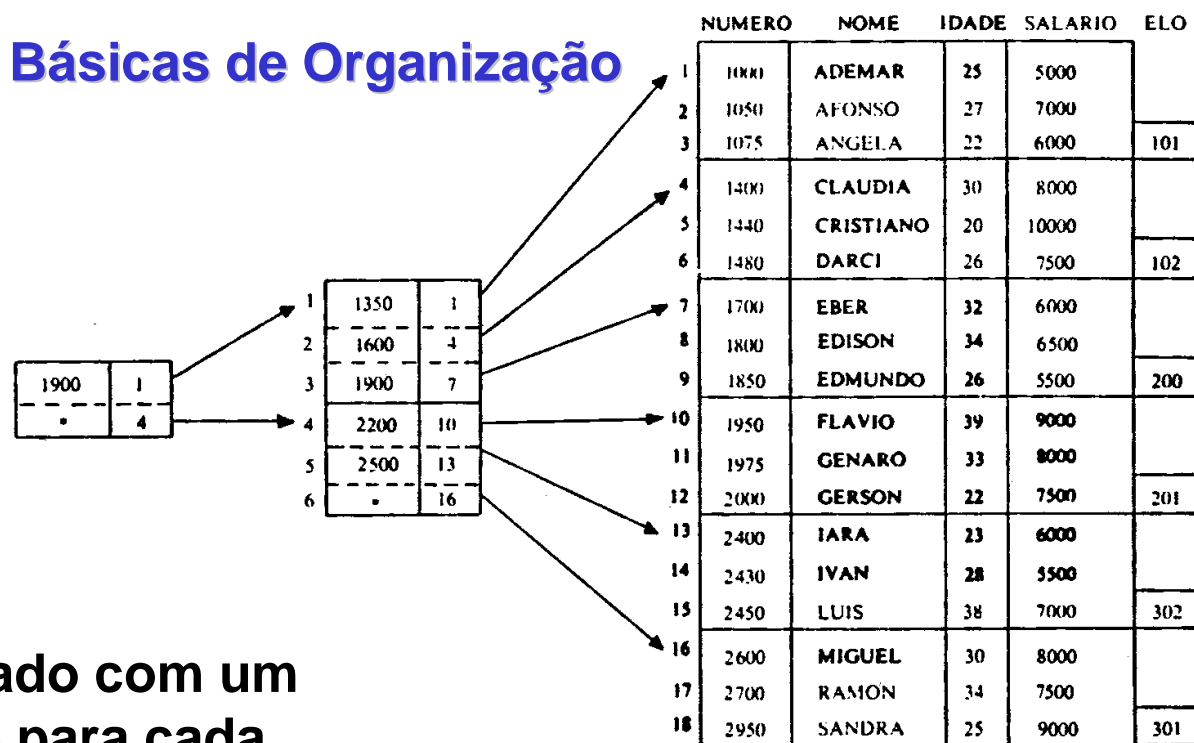
Formas Básicas de Organização



Arquivo Indexado com um campo de elo em cada registro

ÁREAS DE EXTENSÃO	100	1050	AFONSO	27	7000	102-
	101	1480	DARCI	26	7500	∅
	102	1075	ANGELA	22	6000	∅
	103	1300	CARLOS	23	7500	∅
	200	1900	ENIO	28	7800	∅
	201	2200	HELENA	25	8500	∅
	202					
	203					
	300	2700	RAMON	34	7500	302
	301	2450	LUIS	38	7000	∅
	302	2950	SANDRA	25	9000	∅
	303					

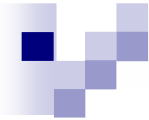
Formas Básicas de Organização



Arquivo Indexado com um campo de elo para cada bloco de registros.

ÁREAS DE EXTENSÃO

100	1350	CESAR	25	9000	Ø
101	1100	ANTONIO	28	8500	103
102	1600	DIOGO	35	5000	Ø
103	1300	CARLOS	23	7500	100
200	1900	ENIO	28	7800	Ø
201	2200	HELENA	25	8500	Ø
202					
203					
300	3150	TATIANA	20	10000	Ø
301	3100	SONIA	28	7500	300
302	2500	MARIA	27	6500	Ø
303					



Formas Básicas de Organização

■ Arquivo Seqüencial Indexado

□ Acesso a um registro

- **acesso serial** : pode ser feito diretamente sobre a área de dados, havendo apenas neste caso cuidados adicionais pela existência de áreas de extensão.
- **Acesso aleatório** : é feito com a utilização do índice. O argumento de pesquisa define o caminhamento sobre o índice, que conduz ao endereço do registro desejado. O endereço obtido pode ser o do próprio registro ou o endereço do bloco que o contém.

□ Inserção de um registro

- Requer a efetivação de uma busca no índice para determinação do local a ser inserido o novo registro, e por fim sua inserção na área de extensão.
- Pode requerer deslocamento de registros dentro do bloco.

□ Exclusão de um registro

- É implementada pela colocação de uma marca de excluído em um campo adicional do registro que indica seu estado.

□ Alteração de um registro

- É feita uma pesquisa por meio do índice para a localização do registro. Se a alteração não envolve a chave de ordenação, este é lido, seus campos alterados, e novamente gravado.



Formas Básicas de Organização

■ Arquivo Indexado

- Os registros são acessados sempre através de um ou mais índices, não havendo qualquer compromisso com a ordem física de instalação dos registros

NRO	END			NRO	NOME	IDADE
1075	2		0	1480	IARA	23
1350	1		1	1350	EDMUNDO	25
1480	0		2	1075	CRISTIANO	21
1800	5		3	1950	TATIANA	56
1950	3		4	2200	GERSON	34
2200	4		5	1800	ANGELA	21

Arquivo Indexado



Formas Básicas de Organização

■ Arquivo Indexado

□ Índices

Em um arquivo indexado pode haver tantos índices quanto forem as chaves de acesso aos registros. As entradas são ordenadas pelo valor da chave de acesso, sendo cada uma delas constituída por um par (chave, endereço do registro).

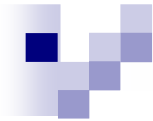
Problema : necessidade de atualização dos índices quando um registro é inserido.

□ Acesso a um registro

- **acesso serial** : é feito com a utilização de um dos índices, escolhido de acordo com a seqüência desejada.
- **Acesso aleatório** : requer a efetivação de uma busca sobre o índice correspondente à chave de acesso utilizada.

□ Inserção de um registro

O registro é armazenado em qualquer endereço vago dentro da área alocada para o arquivo. A seguir, seus pares relativos às chaves de acesso para as quais existem índices são inseridos não índices correspondentes.



Formas Básicas de Organização

■ Arquivo Indexado

□ Exclusão de um registro

- **Física** : é liberada a área de dados ocupada pelo registro excluído e são removidas dos índices as entradas a ele correspondentes.
- **Lógica** : Marcar a entrada correspondente ao registro excluído, pela substituição do endereço do registro por um endereço inválido.

□ Alteração de um registro

- O mesmo que no arquivo seqüencial indexado.



Formas Básicas de Organização

■ Arquivo Direto

- A idéia básica de um arquivo direto consiste na instalação dos registros em endereços determinados. Acesso a um registro com base no valor de uma chave primária. Ao invés de um índice é usada uma função de avaliação que calcula o endereço do registro a partir do argumento de pesquisa.
- **Cálculo do endereço**
 - O primeiro problema associado com os arquivos diretos, é o da determinação de uma função F , que transforme o valor C da chave de um registro no endereço E que lhe corresponde no arquivo.
 - Funções **determinísticas** : associam um único valor da chave de acesso a cada endereço. Na prática é impossível sua implementação.
 - Funções **probabilísticas** : geram para cada valor da chave um endereço “tão único quanto possível” , podendo gerar, para valores distintos da chave, o mesmo endereço, fato que, este é denominado **colisão**.



Formas Básicas de Organização

■ Arquivo Direto

□ Tratamento das colisões

Dizemos que há uma **colisão** quando dois valores diferentes da chave de acesso é atribuído o mesmo endereço.

□ Tratamento de colisões

□ Tratamento por endereçamento aberto

Ao ocorrer encadeamento em uma operação de inserção, é feita uma busca sobre o arquivo para localização de um endereço livre.

■ Pesquisa seqüencial

$E = F(C) \rightarrow$ procura em $E+1, E+2, E+3, \dots M, 1, \dots, E-1$

■ Realeatorização

$E = F(C) \rightarrow$ calcula outro endereço $E = F'(C)$, repete-se até encontrar endereço livre.



Formas Básicas de Organização

■ Arquivo Direto

☐ Tratamento por encadeamento

Registros que colidem são inseridos em uma lista encadeada.

☐ Utilização de áreas de extensão

Cada bloco possui um *header* o qual armazena o endereço da lista encadeada, a ele correspondente na área de extensão.

☐ Acesso a um registro

Acesso serial : somente é provido por um arquivo direto por meio de uma função que preserve a ordem dos registros.

☐ Exclusão de um registro

Uso de um campo para exclusão lógica.