



Área do Conhecimento de Ciências Exatas e Engenharias Bacharelado em Ciência da Computação

Organização de Arquivos Sequencial e Sequencial Indexado

Consultas em Banco de Dados

Prof. Daniel Luis Notari

Agosto - 2017





Problematização

“Qual é a maneira mais apropriada para organizar o registro do arquivo de dados para a organização sequencial?”





Sinopse

- A organização de arquivos sequencial está baseada na ordenação dos registros baseadas em uma chave primária sequencial e crescente.
- Os campos do registro são definidos usando os tipos primários de dados (inteiro, real, caractere, boolean).
- O tamanho do registro é definido contando-se o número de bytes dos tipos de dados do registro, podendo ser de tamanho fixo ou variável.
- A forma de pesquisa no arquivo de dados faz-se através de uma pesquisa exaustiva ou através da pesquisa binária.





Sinopse

- A organização sequencial-indexada difere-se da sequencial pelo uso de um arquivo de índice e pela área de extensão.
- O índice tem o objetivo de diminuir o tempo gasto com a pesquisa exaustiva, onde a pesquisa primeiro é feita no arquivo de índice depois acessa-se diretamente o arquivo de dados.
- A área de extensão é utilizada como uma área temporária para a inserção de registros sem alterar o arquivo de dados.
- Por fim, a operação de reorganização de arquivos é utilizada para inserir os dados da área de extensão quando esta esgota o seu tamanho.





Momento 1: Organização de Arquivo Sequencial





Características

- Os registros são dispostos ordenadamente
- Obedecendo à sequência determinada por uma chave primária, chamada chave de ordenação.
- A ordem lógica e física dos registros armazenados é a mesma.
- Já que os registros em arquivos sequenciais são armazenados em sucessão contínua, acessar o registro n do arquivo (começando no início do arquivo) requer que os registros $n-1$ também sejam lidos.





Exemplo

NÚMERO	NOME	IDADE	FUNÇÃO
1000	ADEMAR	25	PROGRAMADOR I
1050	CLAUDIA	30	ANALISTA II
2010	IARA	27	PROGRAMADOR I
2046	MIGUEL	28	ANALISTA I





Características

- O principal uso é o processamento em série ou sequencial de registros.
- A vantagem de poder acessar rapidamente os registros de forma contínua, torna-se uma desvantagem quando o arquivo é utilizado para acessar um registro diferente que não seja o “seguinte”.
- Em média metade do arquivo deve ser lido para se ter acesso a um determinado registro.
- Os arquivos sequenciais são úteis para classificação e acesso a grande volume de dados.





Características

- Os arquivos sequenciais podem ou não terem chave.
- Cada registro lógico num arquivo com chave tem um item de dado chamado chave e pode ser utilizado para ordenar os registros fisicamente.
- Os registros sem chave estão ordenados em série, cada novo registro é incluído no final do arquivo.
- Em alguns casos, é usual a adição de um campo numérico para identificar cada registro em substituição a uma chave primária que não exista.





Características

- O acesso a um registro, dado um argumento de pesquisa, é facilitado se a chave de acesso coincide com a chave de ordenação, pois nos demais casos não há vantagem na sequencialidade do arquivo.





Abertura de um arquivo

- Um arquivo sequencial
 - pode ser aberto tanto nos modos de leitura e escrita.
- Ao abrir-se um arquivo sequencial,
 - o ponteiro de leitura sempre aponta para o primeiro registro do arquivo.





Consulta a Registros

- O acesso aleatório a registros
 - se caracteriza pela identificação do registro através da especificação de um argumento de pesquisa,
 - tem como resultado registros que não estão armazenados de forma contínua dentro de um arquivo sequencial.
- Pode-se considerar dois tipos de acessos:
 - aqueles cujas chave de acesso coincide com a chave de ordenação (chave primária) e
 - aqueles onde a chave de acesso envolve qualquer atributo.





Consulta a Registros

- No primeiro caso,
 - onde a chave de acesso é a chave de ordenação,
 - O método mais conhecido é a ***Pesquisa Binária***.
- No segundo caso,
 - onde a chave acesso de não coincide com a chave de ordenação,
 - Deve-se fazer uma *pesquisa sequencial do registro(s) especificado(s)*.
 - Esta pesquisa consiste no exame de cada registro do arquivo, a partir do primeiro, até ser localizado aquele que possui, para a chave de acesso, um valor igual ao argumento de pesquisa
 - ou então, ser atingido o final do arquivo, o que significa que nenhum registro do arquivo satisfaz ao argumento de pesquisa.





Pesquisa Binária

- Neste método
 - o primeiro registro a ser consultado é aquele que ocupa a posição média do arquivo
 - e particiona o arquivo iterativamente de acordo com a comparação entre o valor da chave e o argumento de pesquisa.
 - Se a chave do registro for igual ao argumento de pesquisa, a pesquisa termina com sucesso, caso contrário,
 - ocorre uma das duas seguintes situações:
 - a) A chave do registro é maior do que o argumento de pesquisa e o processo de busca é repetido para a metade inferior do arquivo.
 - b) A chave do registro é menor do que o argumento de pesquisa e o processo de busca é repetido para a metade superior do arquivo.





Pesquisa Binária

- A busca é encerrada,
 - com sucesso, quando algum registro possuir a chave igual ao argumento de pesquisa,
 - se sem sucesso, quando a área de pesquisa, que a cada comparação é reduzida pela metade, assumir o comprimento zero(0).





```
procedure pesq_bin(chave, inicio, fim, pos: integer; achou: boolean);
```

```
var
```

```
    meio: integer;
```

```
begin
```

```
    achou:= falso;
```

```
    if (fim-inicio) > 0
```

```
    then begin
```

```
        meio:= int((inicio+fim)/2)+1;
```

```
        seek(arq, meio);
```

```
        read(arq, registro);
```

```
        if registro.chave = chave
```

```
        then begin
```

```
            achou:= true;
```

```
            pos:= meio;
```

```
        end
```

```
    else if registro.chave > chave
```

```
        then pesq_bin(chave, inicio, meio, pos, achou)
```

```
        else pesq_bin(chave, meio, fim, pos, achou);
```

```
    end
```

```
end;
```





Pesquisa Binária

- Com a pesquisa binária,
 - o número máximo de comparações necessário para localizar um registro ou determinar que ele não está presente no arquivo é igual a: $\lceil \log_2 N \rceil + 1$,
 - sendo o número de registros do arquivo.
 - Este desempenho é bem melhor do que o da pesquisa sequencial,
 - para a qual o número médio de comparações é dado por: $(N+1)/2$.





Numero	Nome	Idade	Salário
1000	Ademar	25	500
1050	Afonso	27	700
1075	Angela	22	600
1100	Antônio	28	850
1300	Carlos	23	750
1350	Cesar	55	900
1400	Claudia	25	800
1440	Cristiano	30	1000
1480	Darci	20	750
1600	Diogo	26	600
1700	Edson	35	500
1800	Eder	26	550
1850	Elias	32	650
1900	Flavio	28	780
1950	Gerson	39	700
1975	Geraldo	34	2500
2000	Helena	42	500

Pesquisa Binária

Demonstre o seu
funcionamento





Inserção de Registros

- As inserções de registros
 - possuem um custo muito alto, pois estas consistem em:
 - a) Verificar-se qual a posição adequada para o novo registro (ordem de chave primária)
 - b) Deslocar-se todos os registros que possuírem chave maior do que aquele que está sendo incluído, para uma posição adiante
 - c) Incluir novo registro
- Este tipo de procedimento pode acabar tendo um custo proibitivo, dependendo do tamanho do arquivo. Pode-se usar um arquivo auxiliar.





Alteração de Registros

- Essa operação
 1. consistirá na identificação do registro a ser alterado
 2. através de uma consulta (leitura exaustiva ou pesquisa binária)
 3. modificação dos valores de alguns atributos
 4. regravação do registro.





Exclusão de Registros

- A exclusão física
 - assim como a inserção, pode possuir um custo proibitivo
 - Precisa-se deslocar todos os registros que são posteriores ao registro excluído.
- A exclusão lógica
 - incluir um campo ao registro do arquivo que conterà um indicador de exclusão
 - Para excluir-se um registro é suficiente alterar o valor deste campo e regravar-se no arquivo.





Reorganização do Arquivo

- A reorganização é necessária para:
 - Eliminar registros excluídos logicamente
 - Intercalação do arquivo auxiliar utilizado para as inclusões
 - Processamento do arquivo em lote das alterações (ou outra operação) que não puderam ser realizadas em tempo real.





Momento 2: Organização de Arquivo Sequencial Indexado





Características

- Quando em um arquivo sequencial
 - o volume de acessos aleatórios torna-se muito grande,
 - configura-se a necessidade de utilização de uma estrutura de acesso associada ao arquivo,
 - a qual ofereça maior eficiência na localização de um registro,
 - identificado por um argumento de pesquisa, do que os métodos vistos para arquivos sequenciais.





Características

- Um arquivo sequencial
 - acrescido de um índice (estrutura de acesso) constitui um Arquivo Sequencial Indexado.
- Um arquivo de índices
 - é um arquivo formado por uma coleção de duplas,
 - cada uma delas associando um valor de chave de acesso a um endereço do arquivo de dados.
 - Um índice é sempre específico para uma chave de acesso.
 - Nos casos em que é desejado acesso aleatório, por meio de várias chaves de acesso, deve ser construído um índice para cada uma delas.





Exemplo

ÁREA DE DADOS NO DISCO

	NÚMERO	NOME	ELO
1	100	PEDRO	-
2	150	JOÃO	10
3	200	MARIA	-
4	250	CARLA	20
5	300	MAX	-

ÍNDICE

	NÚMERO	ENDEREÇO
1	100	1
2	150	2
3	175	2
4	200	3
5	250	4
6	275	4
7	300	5

Porque o índice possui mais
Valores do que o arquivo
de dados?





Características

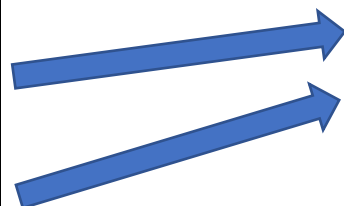
- Um arquivo sequencial indexado
 - possui áreas de extensão que são utilizadas para a implementação da operação de inclusão de registros.
 - Um arquivo sequencial indexado mantém as mesmas características de um arquivo sequencial (agilidade no acesso serial a registros) e
 - amplia a versatilidade no acesso aleatório a registros.
- Entretanto, a implementação deste tipo de organização pode ser um pouco mais complexa devido as suas características.





ÁREA DE DADOS NO DISCO

	NÚMERO	NOME	ELO
1	100	PEDRO	-
2	150	JOÃO	10
3	200	MARIA	-
4	250	CARLA	20
5	300	MAX	-



	NÚMERO	NOME	ELO
10	175	BILL	-
20	275	NARA	-
30	-		
40	-		
50	-		

ÁREA DE EXTENSÃO

ÍNDICE

Endereço
lógico

	NÚMERO	ENDEREÇO
1	100	1
2	150	2
3	175	2
4	200	3
5	250	4
6	275	4
7	300	5

Endereço
lógico





Índices

- A finalidade de um índice
 - é permitir a rápida determinação do endereço de um registro do arquivo,
 - dado um argumento de pesquisa.
 - O endereço determina a posição onde está armazenado o registro no dispositivo de armazenamento.





Índices

- O arquivo de índices
 - ocupa uma área muito menor que a ocupada pelo arquivo de dados.
 - Desta forma, a pesquisa sobre um arquivo de índices pode ser feita com maior rapidez do que se fosse feita diretamente sobre o arquivo de dados correspondente.
- Com o objetivo de tornar mais eficiente o processo de pesquisa,
 - os arquivos de índices podem ser estruturados em níveis,
 - sendo o número de níveis proporcional ao número de entradas (números de índices).





Índices

- O índice associado a chave de ordenação
 - é chamado ***índice primário***,
 - e os demais são chamados de ***índices secundários***.
- Em ambos os casos,
 - as entradas do arquivo de índice são ordenadas pelo valor da chave de acesso a ele associada,
 - com o propósito de permitir maior eficiência no processo de pesquisa.
- O fato de que os registros se apresentam fisicamente ordenados pelo valor da chave de acesso,
 - é aproveitado na organização de índices primários para a obtenção de flexibilidade nas operações de inserção de registros e economia de espaço no armazenamento do índice.





Índices

- A localização de um registro
 - por meio de um argumento de pesquisa é realizada em duas etapas:
 - a) consulta-se o arquivo de índice e determina-se em qual endereço está o registro;
 - b) pesquisa-se o endereço selecionado (por pesquisa sequencial ou binária) para localizar-se o registro desejado.
 - c) O segundo passo pode determinar o acesso a área de extensão através da consulta dos elos.



Índices

Índice Secundário

	Num	End
1	1480	1
2	2200	4
3	*	7

Índice Primário

	Num	End
1	1075	1
2	1350	4
3	1480	7
4	1800	10
5	1950	13
6	2200	16
7	2450	19
8	2700	22
9	*	25

Arquivo de Dados (Funcionários)

	Num	Nome	I	Salário
1	1000	Ademar	25	500
2	1050	Afonso	27	700
3	1075	Angela	22	600
4	1100	Antônio	28	850
5	1300	Carlos	23	750
6	1350	Cesar	25	900
7	1400	Claudia	30	800
8	1440	Cristiano	20	1000
9	1480	Darci	26	750
1	1600	Diogo	35	500
1	1700	Edar	32	600
1	1800	Edson	34	650
1	1850	Elias	26	550
1	1900	Flavio	28	780
1	1950	Geraldo	39	900
1	1975	Gerson	33	800
1	2000	Helena	22	750
1	2200	Iara	23	850
1	2400	Ivan	28	600
2	2430	Luiz	38	550
2	2450	Maria	27	7

CS
ADE
SUL

50
ANOS



Área de Extensão

- Destina-se a conter os registros inseridos
 - após a criação do arquivo.
 - constitui uma extensão da área principal de dados do arquivo.
- As áreas de extensão
 - são necessárias porque a inclusão de um novo registro provocaria um deslocamento dos registros para manter a sequencialidade
 - acarretaria uma reestruturação dos arquivos de índices, pode-se perceber o quanto seria demorada esta operação, o que torna praticamente inviável.





Área de Extensão

- A implementação de áreas de extensão não desperdiça espaço apesar de possuir uma implementação um pouco mais complexa.
- Duas são as alternativas principais para a implementação de áreas de extensão em um arquivo sequencial-indexado:
 - i. Incluir em cada registro da área principal de dados um campo de elo para conter o endereço da lista encadeada de seus antecessores (ou sucessores), alocados na área de extensão;
 - ii. Incluir-se o campo de elo em cada bloco, destinado a conter o endereço da lista de extensões do bloco, sendo mantida a sequencialidade física dentro de cada bloco da área principal, e sendo que todos os registros da lista de extensão de um bloco possuem ordem maior do que todos os da área principal do mesmo bloco.





Área de Extensão

- A segunda alternativa
 - implica na mudança de endereços de registros dentro de um bloco,
 - pode ser utilizada somente naqueles casos em que o índice possui uma entrada para cada bloco.
 - A existência de índices secundários, sobre o mesmo arquivo, torna inviável esta alternativa.





Consulta a Registros

- O acesso aleatório a registros é realizado com a utilização do índice.
- O argumento de pesquisa define o caminharmento sobre o índice,
 - que conduz ao endereçamento do registro desejado.
 - O endereço obtido do índice pode ser o próprio endereço do registro ou o endereço do bloco que o contém.
 - Neste último caso, é necessária ainda a efetivação de uma busca no bloco para a localização do registro, a qual pode requerer acessos à área de extensão.





Inclusão de Registros

1. Efetivar uma busca através da estrutura de índices para se verificar em qual bloco de dados o registro deve ser incluído.
2. Uma vez determinada a sua posição, o registro é incluído na área de extensão do seu sucessor na área principal e atualizam-se os elos (método *i*, acima).





Inclusão de Registros

3. Já pelo método *ii*, o registro é inserido no bloco selecionado.

- o registro será inserido na área de extensão se sua chave for maior do que a do último registro da área principal;
- caso contrário o último registro da área principal é deslocado para a área de extensão e o novo registro é inserido em sua posição correta dentro do bloco principal,
- podendo requerer deslocamentos dentro do bloco.





Exclusão de Registros

- A exclusão
 - é realizada logicamente através de um campo que indica que o registro está excluído ou não.
 - Estes registros serão desconsiderados pelas consultas e poderão ser utilizados, como dito acima, pelas inclusões.





Alteração de Registros

1. Localiza-se o registro como na consulta (através da estrutura de índices) caso ele seja identificado por um argumento de pesquisa que envolva a chave de acesso.
2. Verifica-se se o registro está no arquivo de dados ou na área de extensão.
3. A seguir alteram-se os campos desejados
4. Regrava-se o registro.





Reorganização do Arquivo

- A reorganização de um arquivo sequencial indexado será necessária sempre que:
 - a área de extensão tornar-se muito grande, diminuindo a performance nos acessos ao arquivo;
 - for necessária a eliminação dos registros excluídos logicamente (se estas áreas não forem aproveitadas pelo algoritmo de inclusão de registros);
 - se precisar reestruturar os arquivos de índices para melhorar o acesso aos registros.





Momento 3: Exercícios





Exercício

Numero	Nome	Idade	Salário
1000	Ademar	25	500
1050	Afonso	27	700
1075	Angela	22	600
1100	Antônio	28	850
1300	Carlos	23	750
1350	Cesaar	55	900
1400	Claudia	25	800
1440	Cristiano	30	1000
1480	Darci	20	750
1600	Diogo	26	600
1700	Edson	35	500
1800	Eder	26	550
1850	Elias	32	650
1900	Flavio	28	780
1950	Gerson	39	700
1975	Geraldo	34	2500
2000	Helena	42	500

1. Crie o registro de dados para o arquivo ao lado usando registro de tamanho fixo com o caractere `\n` de final de linha e sem separador entre os campos.
2. Implemente um procedimento para inserir os dados da tabela ao lado e outro procedimento para mostrar os dados.
3. Implemente a pesquisa binária.
4. Implemente um arquivo de índice para o campo número
5. Use a pesquisa binária para pesquisar no arquivo de índice e, depois o comando seek para pesquisar no arquivo de dados.





Exercício

6. Dado o arquivo **sequencial indexado** ao lado, *com um campo de elo de extensão para cada registro* e com uma área de extensão única para todos os blocos, preencha o campo elo de todos os registros e a área de extensão, após a inclusão dos seguintes funcionários:

- 1450 CRISTINA 23
- 1440 ROBERTO 29
- 1435 MARCELO 30
- 2450 PAULO 32





Exercício

Índice		Funcionário			
		Matrícula	Nome	Idade	Elo
1430	01	1230	Adriana	18	
2640	04	1330	Almir	19	
3990	07	1430	Carlos	20	
4965	11	2400	Cláudio	19	
		2600	Cristiane	18	
		2640	Fernando	18	
		3270	<u>Cida</u>	18	
		3370	Marcelo	19	
		3850	Sérgio	18	
		3990	Zoe	19	
		4670	Paulo	20	
		4760	Renato	19	
		4860	Henrique	18	
		4965	Buarque	19	
		Área de extensão			
		201			
		202			
		203			
		204			
		205			
		206			
		207			





Exercício

7. Dado o arquivo **sequencial indexado** abaixo, *com um campo de elo de extensão para cada bloco de registro* e com uma área de extensão única para todos os blocos, preencha o campo elo de todos os registros e a área de extensão, após a inclusão dos seguintes funcionários:

- 1450 CRISTINA 23
- 1460 ROBERTO 29
- 1650 MARCELO 30
- 1340 PAULO 32



Exercício

		Funcionário				
Índice			Matrícula	Nome	Idade	Elo
1430	01	1	1230	Adriana	18	
2640	04	2	1330	Almir	19	
3990	07	3	1430	Carlos	20	
4965	11	4	2400	Cláudio	19	
		5	2600	Cristiane	18	
		6	2640	Fernando	18	
		7	3270	Cida	18	
		8	3370	Marcelo	19	
		9	3850	Sérgio	18	
		10	3990	Zoe	19	
		11	4670	Paulo	20	
		12	4760	Renato	19	
		13	4860	Henrique	18	
		14	4965	Buarque	19	
Área de extensão						
		201				
		202				
		203				
		204				
		205				
		206				
		207				