

Organização de Arquivos

Arquivos

Arquivo é um conjunto de **registros** que possuem a mesma estrutura.

Registro: conjunto de pares atributo-valor que identifica um indivíduo de uma entidade

Exemplo:

Funcionário:

<nome, João>, <CPF, 123456789-00>, <data-nascimento, 28/06/1980>, <salário, 3000>

Arquivos

- **Registro Lógico (Registro)**
 - **Registro Físico**
 - Armazenamento do arquivo em 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**
 - Sequência 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

Operações Básicas

- **Criação:** alocação e inicialização da área de dados, assim como de seus descritores
- **Destruição:** liberação da área de dados e descritores usados na representação da tabela
- **Inserção:** inclusão de novo registro na tabela
- **Exclusão:** remoção de um registro da tabela
- **Alteração:** modificação dos valores de um ou mais atributos/campos da tabela
- **Consulta:** obtenção dos valores de todos os campos de um registro, dada uma chave de entrada

Tipos de Arquivos

- Arquivo Sequencial
- Arquivo Sequencial Ordenado
- Arquivo Seqüencial-Indexado
- Arquivo Indexado
- Arquivo Direto
- Arquivo Invertido

Arquivo Sequencial

Definição

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

• **Vantagem**

- Simplicidade

• **Desvantagem**

- Busca de registro através de acesso seqüencial

Arquivo Sequencial

Nome	Idade	Salário
Ademar	32	5000
Roberto	25	7500
Gerson	43	6000
Yeda	23	9000
Bernardo	21	4500
Ângela	29	5000

Arquivo Sequencial Ordenado

- As tabelas são mantidas ordenadas pela chave primária.
- Não é necessário ir até o final do arquivo para ver que uma chave não está no arquivo
- Chave > chave busca
 - sabe-se que a chave procurada não está lá

Arquivo Sequencial Ordenado

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 Sequencial Ordenado

- Vantagens
 - Acesso seqüencial eficiente (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 Sequencial Ordenado


Como atualizar o arquivo e mantê-lo ordenado?

- Executar uma ordenação a cada atualização?

Solução: Arquivo de Transações


- Chave primária
- Identificador do tipo de transação(" I ", "E ", " M ")
- PARAMODIFICAÇÕES:
 - Nome do atributo a modificar
 - Novo valor do atributo
- PARAINCLUSÕES
 - Valores de todos os atributos

Arquivo Sequencial Ordenado



02	João	21/04/1990
11	Maria	02/07/1986
13	Ana	05/07/1978
17	Mario	04/02/1976
20	Rosa	03/01/1980
21	Ricardo	24/10/1985
56	Marcelo	30/10/1992

Arquivo Mestre



11	E		
17	M	Nome	Maria
18	I	Andrea	02/04/1990
26	I	Jonas	03/07/2000

Arquivo de Transações

Arquivo Sequencial Ordenado

Reorganização do Arquivo

- Operação de intercalação entre os arquivos mestre e transação

Arquivo Sequencial Indexado

- Arquivo Seqüencial + índice

- Área de índices
 - Arquivo seqüencial criado pelo sistema
 - contém o endereço do início do segmento e a chave mais alta
- Área primária (principal)
 - Reservada para os registros de dados, classificados em ordem ascendente pelo seu campo chave.
- Área de excedentes (overflow)
 - Novos registros que não podem ser colocados na área principal quando se produz uma inserção no arquivo.

Arquivo Sequencial Indexado

Valor da Chave	Endereço
550	1
700	4

↖ Máximo do bloco

↑ Início do bloco

1

4

Matrícula	Nome	Idade
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

Arquivo Sequencial 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

Extensão - Modelo 1

Índice

600	01
700	04
920	07
*	10

1

4

7

ÁREA DE EXTENSÃO

200				
201				
202				
203				

Funcionários

Matrícula	Nome	Idade	Elo
450	ANA	25	
500	ZÉLIA	21	
550	ÉLIDA	37	201
630	ADRIANO	32	
680	CARLOS	34	200
700	WAGNER	30	
780	ROBERTO	25	
900	ALBERTO	40	
920	PAULO	27	
1000	JUCA	31	
1050	ALVES	30	202
2000	YEDA	32	

690 JOÃO 23
600 PEDRO 25
1500 ANTÔNIO 30
820 MARIA 29

Extensão - Modelo 2

600	01
750	04
950	07
*	10

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

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	

ÁREA DE EXTENSÃO

690	JOÃO	23	200
600	PEDRO	25	201
1500	ANTÔNIO	30	202
820	MARIA	29	203
			204
			205

Arquivo Sequencial Indexado

- Tipos de Acesso
 - Acesso serial
 - Direto sobre a área de dados + extensão sem usar o índice
- Acesso aleatório
 - Usando o índice
 - Obter o endereço do próprio registro
 - Obter endereço do bloco: realizar a busca dentro do bloco e incluir mais acessos à área de extensão
- Leitura exaustiva (todos os registros)
 - Igual ao acesso serial!

Arquivo Sequencial Indexado

- 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

Arquivo Sequencial Indexado

Reorganização

Desempenho da operações está relacionado ao número de inclusões e exclusões

Periodicamente, excluir fisicamente os logicamente excluídos e realizar limpeza na área de extensão;

Um novo índice deve ser gerado.

Arquivo Sequencial Indexado

Principais características

- Permite acesso aleatório satisfatório;
- Permite acesso seqüencial eficiente, pela chave primária
- Permite, com relativa facilidade, as inserções e exclusões, através do uso de áreas de extensão

Arquivo Indexado

- 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

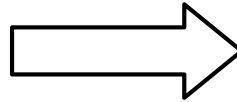
- Índice
 - Classifica-se em:
 - Exaustivo: Quando possui uma entrada para cada registro do arquivo
 - Seletivo: Uma entrada para cada subconjunto de registros

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



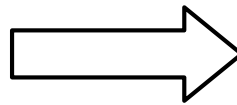
Endereço do reg.	NroFunc	Depto	Salário	Sexo
301	1000	A	1.000	M
302	1066	B	2.000	F
303	1140	A	1.500	F
304	1162	C	1.000	M
305	1234	C	2.000	M
306	1305	C	2.500	F
307	1312	A	3.000	F

Arquivo Indexado

- Índice Seletivo

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

Depto	Nro serial do índice primário
A	1, 3, 7
B	2

*Índice seletivo
(depto)*

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

*Índice seletivo
(sexo)*

Arquivo Indexado

- 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

- Uso de uma função (*hashing*) que calcula o endereço do registro a partir do valor da chave do registro

Arquivo Direto

Argumento de Pesquisa

$$E = F(c)$$

C = 3150

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

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 de Endereço:
 - Funções Determinísticas e Probabilísticas
 - Determinísticas: gera um único endereço
 - 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

- 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

A cada um dos valores da chave de acesso é associada uma lista de identificação de registros, chamada lista invertida.

Exemplo: arquivo com uma inversão associada a IDADE

Matr	Nome	Idade	Salário
1000	ANTÔNIO	25	5000
1050	AFONSO	20	7000
2400	CRISTIAN	25	5500
1850	EDSON	21	5500
1440	YEDA	22	5000
3150	SANDRA	22	7000
2000	FLÁVIA	20	5500
1900	ROBERTO	20	10000

Inversão

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

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