**UNIVERSIDADE CATÓLICA DE PERNAMBUCO**

**CURSO - CIÊNCIA DA COMPUTAÇÃO**

**DISCIPLINA – ESTRUTURA DE DADOS II**

**ORGANIZAÇÃO DE ARQUIVOS**

**JESIMIEL RUFINO DA SILVA**

**PEDRO HENRIQUE ARAÚJO CALDEIRA**

**RECIFE - PE**

**18/11/2016**

Introdução

A organização de arquivos é o modo como os dados estão organizados internamente e essa estrutura pode variar de acordo com com o tipo de informação que o arquivo contém.

Existem diversas formas de organização de arquivos porém a mais simples é através de uma sequência não estruturada de bytes. Essa maneira intuitiva de armazenar um arquivo consiste na distribuição dos seus registros em uma ordem arbitrária, um após o outro, dentro da área destinada a mantê-los. Esta ordem pode ser, por exemplo, aquela na qual os registros são gerados. Isto causa uma dificuldade na localização dos registros e uma perda de eficiência, porém esta técnica intuitiva é bastante usada

Não há como recuperar partes individuais dos arquivos, como nome de um cliente por exemplo. Os arquivos são lidos/escrito em blocos. Esses blocos são os *registros* que constituem a menor parte que pode ser lida e escrita em um arquivo. Os *registros,* consequentemente, são conjuntos de fields ou campos que por sua vez são agregados de caracteres que constituem a menor unidade lógica de informação em um arquivo.

O armazenamento de pequenos volumes de dados, via de regra, não implica em grandes problemas no que diz respeito à distribuição dos registros dentro de um arquivo, desde que a freqüência de acessos aleatórios a registros não seja muito elevada. Mas à medida que o volume de dados e/ou a freqüência e a complexidade dos acessos aumentam, crescem também os problemas de eficiência do armazenamento dos arquivos e do acesso a seus registros, fazendo da melhoria das técnicas de armazenamento e recuperação de dados uma conseqüência da necessidade de acesso rápido a registros pertencentes a grandes arquivos ou, simplesmente, arquivos muito solicitados.

Serão apresentadas neste trabalho introduções sobre cinco estratégias/técnicas de organização de arquivos voltadas para o acesso por meio de chaves primárias, que são Arquivo sequencial, Arquivo sequencial Indexado, Arquivo Indexado, Arquivo Direto e Arquivo Invertido. Não pretendemos esgotar o assunto Organização de Arquivos neste trabalho, que funcionará assim apenas como uma introdução ao assunto.

Definições

Um **arquivo** é formado por uma coleção de *registros lógicos*, cada um deles representando um objeto ou entidade.

Um **registro lógico**, ou simplesmente *registro*, é formado por uma seqüência de itens, sendo cada item chamado *campo* ou atributo. Cada item corresponde a uma característica ou propriedade do objeto representado.

Cada **campo** possui um *nome*, um *tipo* e um *comprimento*. O comprimento dos valores de um atributo pode ser constante para todos os registros do arquivo, ou variável.

O armazenamento de um arquivo é feito, via de regra, por **blocos**de registros lógicos (um bloco é chamado *registro físico*), sendo, em cada leitura ou gravação, lido ou gravado todo um bloco e não apenas um registro lógico.

Uma **chave** é uma seqüência de um ou mais campos em um arquivo.

Uma **chave primária**é uma chave que apresenta um valor diferente para cada registro do arquivo, de tal forma que, dado um valor da chave primária, é identificado um único registro do arquivo. Usualmente a chave primária é formada por um único campo.

Uma **chave secundária** difere de uma primária pela possibilidade de não possuir um valor diferente para cada registro. Assim, um valor de uma chave secundária identifica um conjunto de registros.

**Chave de acesso** é a chave utilizada para identificar o(s) registro(s) desejado(s) em uma operação de acesso a um arquivo.

**Argumento de pesquisa**é o valor da chave de acesso em uma operação.

**Chave de um registro** é o valor de uma chave primária em um particular registro do arquivo

Chave de ordenação é a chave primária usada para estabelecer a seqüência na qual devem ser dispostos (física ou logicamente) os registros de um arquivo.

**Arquivos Sequenciais**

Em um arquivo sequencial, os registro são dispostos ordenadamente, obedecendo a sequência determinada por uma chave primária, chamada chave de ordenação. Na figura abaixo, é apresentado um exemplo de arquivo sequencial, no qual é usado como chave de ordenação o atributo NÚMERO.

Esta organização, que representa um aperfeiçoamento em relação àquela na qual os registros são dispostos aleatoriamente, representa, também, uma perda de flexibilidade por não acomodar com simplicidade as operações de modificação do arquivo.

O acesso a uma registro, dado um argumento de pesquisa, é facilitado se a chave de acesso coincide com a chave de ordenação (ou com sua parte inicial), pois, nos demais casos, não há vantagem na sequencialidade do arquivo.

**Arquivo Sequencial:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **NÚMERO** | **NOME** | **SALÁRIO** |
| 1 | 100 | PEDRO | 3000 |
| 2 | 150 | JOÃO | 1500 |
| 3 | 200 | MARIA | 2500 |
| 4 | 250 | CARLA | 3000 |
| 5 | 300 | MAX | 2000 |

ÁREA DE DADOS NO DISCO

**Arquivos Sequenciais Indexados**

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, que ofereça maior eficiência na localização de um registro identificado por um argumento de pesquisa do que os métodos vistos para arquivos seqeenciais.

Um arquivo sequencial, acrescido em um índice (estrutura de acesso) constitui um a*rquivo sequencial 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. Assim, um índice é sempre específico para uma chave de acesso. Além do arquivo sequencial e do índice, um arquivo sequencial indexado possui áreas de extensão que são utilizadas para a implementação da operação de inserção de registros.

A finalidade de um índice é permitir rápida determinação do endereço de um registro do arquivo, dado um argumento de pesquisa. O endereço identifica a posição onde está armazenado o registro, na memória secundária.

Usualmente, cada entrada do índice, formada por um par (chave do registro, endereço do registro), ocupa um espaço bem menor do que o registro de dados correspondente, o que faz com que a área ocupada pelo índice seja menor do que aquela ocupada pelos dados. Com isto a pesquisa sobre o índice pode ser feito com maior rapidez do que se fosse feita diretamente sobre o arquivo de dados correspondente. Este fato constitui a justificativa maior para a utilização dos índices.

**Arquivo Sequencial Indexado:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **NÚMERO** | **ENDEREÇO** | | 100 | 1 | | 150 | 2 | | 200 | 3 | | 250 | 4 | | 300 | 5 |   ÍNDICE | |  |  |  |  | | --- | --- | --- | --- | |  | **NÚMERO** | **NOME** | **SALÁRIO** | | 1 | 100 | PEDRO | 3000 | | 2 | 150 | JOÃO | 1500 | | 3 | 200 | MARIA | 2500 | | 4 | 250 | CARLA | 3000 | | 5 | 300 | MAX | 2000 |     ÁREA DE DADOS NO DISCO |

Áreas de Extensão

A área de extensão (também chamada área de overflow) destina-se a conter os registros inseridos, em um arquivo sequencial indexado, após a criação do arquivo. Ela constitui uma extensão da área principal de dados do arquivo.

Áreas de extensão são necessárias em arquivos seqüenciais indexados, porque neles não é viável a implementação da operação de inserção de registros do mesmo que nos arquivos seqüenciais. Naquele processo, a maioria dos registros muda de endereço, o que obrigaria uma completa alteração nas entradas do índice, a cada atualização do arquivo.

Uma possível implementação de áreas de extensão em um arquivo sequencial indexado consiste em destinar um em cada registro da área principal um campo de elo para conter o endereço da lista encadeada de seus sucessores ou antecessore alocados na área de extensão.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **NÚMERO** | **ENDEREÇO** | | 100 | 1 | | 150 | 2 | | 175 | 2 | | 200 | 3 | | 250 | 4 | | 300 | 5 |   ÍNDICE | |  |  |  |  | | --- | --- | --- | --- | |  | **NÚMERO** | **NOME** | **ELO** | | 1 | 100 | PEDRO | - | | 2 | 150 | JOÃO | 10 | | 3 | 200 | MARIA | - | | 4 | 250 | CARLA | 20 | | 5 | 300 | MAX | - |   ÁREA DE DADOS NO DISCO |
|  | |  |  |  |  | | --- | --- | --- | --- | |  | **NÚMERO** | **NOME** | **ELO** | | 10 | 175 | BILL | - | | 50 |  |  | - |   ÁREA DE EXTENSÃO |

**Arquivos Indexados**

Nos arquivos seqüenciais indexados, o compromisso de manter os registros fisicamente ordenados pelo valor da chave de ordenação, com o objetivo de prover um acesso serial eficiente, acarreta uma série de problemas, principalmente no que diz respeito à operação de inserção de um registro, conduzindo à necessidade de utilização de áreas de extensão e efetivação de reorganizações periódicas.

À medida que decresce a frequência de acessos seriais, relativamente à frequência de acessos aleatórios, a manutenção da sequencialidade física do arquivo encontra uma compensação cada vez menor em termos de eficiência de acesso, até tornar-se antieconômica. A partir deste ponto, torna-se mais conveniente o uso de um *arquivo indexado*, no qual os registros são acessados sempre através de um ou mais índices, não havendo qualquer compromisso com a ordem física de instalação dos registros.

A liberdade na escolha do endereço no qual um registro é armazenado representa um ganho de flexibilidade que permite maior eficiência, principalmente na operação de inserção de um registro, conduzindo, também, a uma simplificação da estrutura geral do arquivo, sendo dispensados os mecanismos complexos de administração de áreas de extensão.

Arquivo Indexado:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **NÚMERO** | **ENDEREÇO** | | 100 | 4 | | 150 | 3 | | 200 | 1 | | 250 | 5 | | 300 | 2 |   ÍNDICE | |  |  |  |  | | --- | --- | --- | --- | |  | **NÚMERO** | **NOME** | **SALÁRIO** | | 1 | 200 | PAULO | 3100 | | 2 | 300 | JOSÉ | 4500 | | 3 | 150 | MARIA | 2500 | | 4 | 100 | MARISA | 5000 | | 5 | 250 | FABIO | 2500 |   ÁREA DE DADOS NO DISCO |

Índices

Em um arquivo indexado, podem existir tantos índices quantas forem as chaves de acesso aos registros. Um índice consiste de uma entrada para cada registro considerado relevante com relação à chave de acesso associada ao índice. As entradas do índice são ordenadas pelo valor da chave de acesso, sendo cada uma delas constituída por um par (chave do registro, endereço do registro). A sequencialidade física das entradas no índice visa a tornar mais eficiente o processo de busca e permitir o acesso serial ao arquivo.

Um índice é dito exaustivo quando possui uma entrada para cada registro do arquivo e seletivo quando possui entradas apenas para um subconjunto de registros. O subconjunto é definido por uma condição relativa à chave de acesso e/ou a outros atributos do arquivo.Um exemplo de índice seletivo seria o índice dos funcionários estáveis (há mais de 10 anos na empresa) sobre o cadastro geral de funcionários de uma empresa.

O maior problema relacionado com a utilização de arquivos indexados diz respeito à necessidade de atualização de todos os índices, quando um registro é inserido no arquivo. Atualizações nos índices também são necessárias quando a alteração de um registro envolve atributos associados a índices. Nos arquivos seqüenciais indexados, a necessidade de alteração dos índices é eliminada pelo uso de áreas de extensão e encadeamento na implementação de inserções; no entanto, esta estratégia não é condizente com a idéia de arquivos indexados, nos quais a manutenção constante dos índices é necessária.

**Arquivos Diretos**

A idéia básica de um arquivo direto consiste na instalação dos registros em endereços determinados com base no valor de uma chave primária, de modo que se tenha acesso rápido aos registros especificados por argumentos de pesquisa, sem que haja necessidade de percorrer uma estrutura auxiliar (índice).

Um arquivo direto é semelhante a um arquivo indexado, no sentido de que, nos dois casos, o objetivo principal é a obtenção de acesso aleatório eficiente. Em um arquivo direto, aos invés do índice é usada uma função que calcula o endereço do registro a partir do argumento de pesquisa.

As duas organizações possuem diferenças importantes, além do modo pelo qual é feito o acesso. uma delas é o fato de que nos arquivos indexados, ao contrário dos diretos, o endereço onde um registro é armazenado independe do valor de sua chave, e uma outra, muito importante, diz respeito a acessos seriais, que nos arquivos indexados são providos por meio de índices e nos arquivos diretos não são previstos, de acordo com a idéia básica.

Arquivo Direto:

Chave: 150; Pos = Função(Chave); Pos = 3;

|  |  |  |  |
| --- | --- | --- | --- |
|  | **NÚMERO** | **NOME** | **SALÁRIO** |
| 1 | 200 | PAULO | 3100 |
| 2 |  |  |  |
| **3** | **150** | **MARIA** | **2500** |
| 4 |  |  |  |
| 5 | 250 | FABIO | 2500 |

ÁREA DE DADOS NO DISCO

Informações sobre Arquvos Indexados

Arquivos Indexados

Manutenção da ordem dá um pouco de trabalho como você deve ter visto pelo nosso último algoritmo de intercalação balanceada existe uma alternativa para eu manter o registro na ordem em que eles estiver e mesmo assim ter acesso a qualquer um deles numa velocidade muito azuado uma boa velocidade essa estrutura é a de arquivos indexados os arquivos indexados criam para a gente se esqueçam estruturas que nos facilitam essa busca seja por nome seja por idade seja código né como é que estão Mais esses essas estruturas que você tem que permitem acesso aleatório é um registro dada uma determinada chave ou seja vamos as palavras chaves aqui primeira que me interessa acesso aleatório o arquivo sequencial ele é feito para acesso sequencial tô pegando e gesso depois o próximo depois eu próximo depois o próximo depois o próximo pois pronto depois passa até chegar no fim quando eu falo de acesso aleatório no meio qual é a chave a chave que vai ser usada na organização de serviços por escolher a minha chave é o nome então todos os meus arquivos estarão organizados em torno desse nome se eu quero Roberto procurar alguém chamar Roberto eu vou saber que esse cara tá depois do Paulo e antes do Túlio certo então Faber tu tá ali no meio aí eu consigo fazer essa busca mais rápido agora se eu faço Como que eu vou saber vou ter que percorrer um por um para ir testando porque aquela organização tá torrent no ar Então eu preciso definir qual é a sua chave eu vou criar uma estrutura paralela chamada ainsi que trará para mim todos todo morganization específicas e gesso em toda aquela chatice não consigo fazer acesso aleatório complicado mas não é um exemplo eu tenho aqui alguns dados de livros alguns poucos sua quase do tamanho da tela Eu tenho 7 livros com 7 autores e com seus preços aquele o arquivo de dados é um arquivo que a gente poderia pensar que estaria organizado por ordem de chegada tá tão os registos vão sendo cadastrado sempre no fim do arquivo de repente se eu excluo um registro do arquivo só abre um buraco pode ser que eu Aproveite esse buraco para inserir um novo registro então é essa peração Tiraria qualquer ordem lógica de qualquer campo do meu registro Inclusive a ordem de código tá bom se eu seu aproveita o buraco no meu arquivo deixado por registros excluídos então eu não posso dizer que há uma organização qualquer nem a de chegar não é lembrado que ela nem chegado no aumente a ordem do código né quando não vou repetir se lembra só que código número sequencial não significativo fica mais ou menos estabelece a ordem de chegada de serviços muito bem então nem assino a senha para você pode vir aqui o código 8 tá lá no fim do arquivo e ele não é o maior código eu não tem ordem nenhuma nesse meu arquivo de dados mas eu quero fazer busca por código na hora que você chegar lá no sistema passei aí eu preciso acessar o livro tall Tee Light que tinha do livro que diz qual que é o código dele você coloca lá 13 muito bem se não achar que não está ordenado por ódigo nem por ti vem pro alto aí preço como é que eu acho esse 13 teria que percorrer um por um aí eu vou lá a minha sutura de apoio que é o índice me disse nada mais é do que um vetor vetor com dois Campos o primeiro campo é a chave nesse caso o meu arquivo trabalha com uma ordenação