

Introdução

Projeto de Banco de Dados

Organização de Arquivos e Índices

- O entendimento de como os arquivos de um BD e seus índices são organizados passa pelo conhecimento de como os dados são acessados.

Localizar os Dados

- Localizar e apresentar os dados ao usuário requer várias camadas de software:
 - 1.O SGBD determina que registro é necessário e pede ao gerenciador de arquivos para buscar o registro.
 - 2.O gerenciador de arquivo solicita ao gerenciador de disco a página que contém o registro desejado.
 - 3.Finalmente, o gerenciador de discos determina o local físico da página desejada no disco e emite a requisição E/S de disco necessária.

Gerenciador de Discos

- Faz parte do SO.
- É de sua responsabilidade as operações de E/S físicas.
- Ele conhece os endereços de disco físico.

Gerenciador de Discos

- Gerenciador de arquivos:
 - Usa os recursos disponibilizados pelo gerenciador de discos.
 - Permite ao SGBD ver o disco como uma coleção de arquivos.
 - O gerenciador de disco permite que desconheça os detalhes de E/S do disco físico e veja ele de uma forma lógica como E/S de página.
 - Essa função é chamada de gerenciamento de página.
 - O gerenciador de página é responsável por gerenciar os conjuntos de páginas.

Gerenciador de Discos

- Gerenciador de página:
 - Quando um novo conjunto de página é criado, uma quantidade “X” de páginas faz parte desse conjunto.
 - De uma maneira simples, uma página tem, além do dado armazenado, um número único dentro do conjunto e um ponteiro apontando para a próxima página.
 - A primeira página de cada conjunto de páginas – página zero – contém uma lista de conjuntos de páginas existente atualmente no disco, juntamente com um ponteiro para a primeira página de cada um desses conjuntos de páginas.

Gerenciador de Discos

- Conjuntos de Páginas:
 - Como as páginas estão organizadas impactam diretamente sobre a velocidade de recuperação dos dados.
 - Essa velocidade de recuperação nem sempre tem o ganho esperado.
 - Índices são necessários para melhorar a velocidade de recuperação.

Índices

- São usados para aumentar a velocidade da recuperação de registros nas respostas a certas condições de busca.
- É um tipo especial de arquivo que armazena duas informações por entrada:
 - Um valor de dados.
 - Um ponteiro.

Índices

- Valor de dados:
 - É um valor para algum campo do arquivo indexado.
- Ponteiro:
 - Identifica um registro do arquivo que tenha o valor para o campo desejado.
 - Esse campo é chamado de campo indexado ou chave de índice.
 - Qualquer campo do arquivo pode ser usado para criar um índice.

Índices

- Exemplo:
 - Tendo como base um arquivo de alunos, onde a busca pelo nome do aluno seja uma constante.
 - Levando em consideração que o nome do aluno não é um campo chave, para localizar um determinado aluno pelo nome é necessário fazer uma busca linear.
 - Caso exista um arquivo indexado pelo nome do aluno, pode ser efetuada uma busca binária sobre esse arquivo.
 - Ao encontrar o nome do aluno desejado, basta obter o ponteiro que identifica o registro do arquivo de alunos.

Índices

- Ponto Forte:
 - Ganho de velocidade nas operações de leitura.
- Ponto Fraco:
 - Overhead causado nas operações de mudança de status do banco de dados.
 - Exemplo:
 - A inclusão de um novo registro no arquivo indexado, obrigatoriamente, também deve ser feita no índice de uma maneira que o índice continue consistente.

Índices

- Os tipos de índices em nível único:
 - Índices primários.
 - Índices clustering.
 - Índices secundários.

Índices

- Índice primário:
 - O valor armazenado é do mesmo tipo do campo chave do arquivo de dados.
 - Esse tipo de índice também é utilizado para manter a unicidade do campo chave.
 - O arquivo só pode ter um índice primário.

Índices

- Índice clustering:
 - Utilizado por arquivos que mantêm os seus registros ordenados fisicamente por um campo não chave.
 - Não é necessário manter a unicidade do campo indexado.
 - O arquivo só pode ter um índice clustering.

Índices

- Índice secundário:
 - Permite que o arquivo seja indexado por um campo que seja...
 - Chave candidata .
 - Não chave.
 - Um arquivo pode ter vários índices secundários.

Índices

- Índices Densos:
 - Tem uma entrada de índice para cada registro do arquivo de dados.
 - Índice secundário.
- Índices esparsos ou não densos:
 - Não tem uma entrada de índice para cada registro do arquivo de dados.
 - Índice primário.
 - Índice clustering.

Índices

- Árvore-B:

- É “um tipo particularmente comum e importante de índice”. [DATE]
- Possui “condições adicionais que garantem que uma árvore esteja sempre balanceada”. [ELMASRI]

Índices

- Árvore-B:

- É um índice multinível.

Divide o índice em dois conjuntos

- inter-relacionados denominados:
 - Conjunto de sequência.
 - Conjunto de índices.

Índices

- Árvore-B:

- Conjunto de sequência:
 - Índice normal.
 - Geralmente denso.
 - Entradas do índice são agrupadas por páginas.
 - As páginas são encadeadas.
 - As páginas fornecem um acesso sequencial aos dados indexados.

Índices

- Árvore-B:

- Conjunto de índice:

- Índice esparso.
- As entradas apontam para um conjunto de sequência.
- O nível superior consiste num nó apontando para muitos conjuntos de sequência.
- Definição:
<P1 <K1, Pr1>, P2, <K2, Pr2>, ..., <Kq-1, Prq-1>, Pq>
 - Pi é um ponteiro de árvore (ponteiro para um conjunto de sequência).
 - Pri é um ponteiro de dados.
 - Ki é um valor chave do índice.