

Banco de dados II

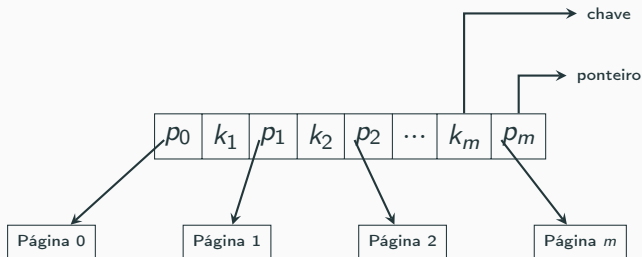
03 - Indexação Baseada em Árvore

Marcos Roberto Ribeiro



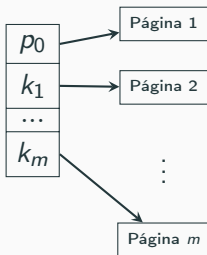
Instituto Federal Minas Gerais - Campus Bambuí

2018



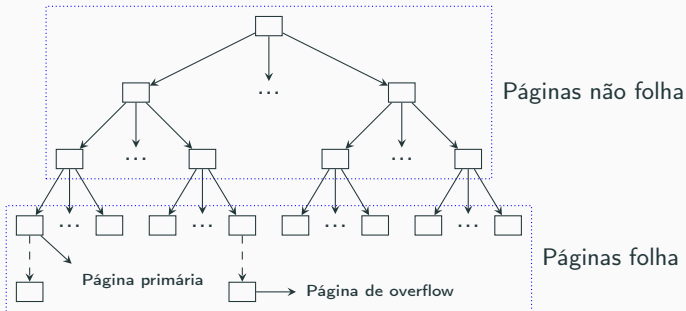
- k : chave de pesquisa
- k^* : entrada de dados correspondente à chave de pesquisa k
- k^* na página $0 < k \leq k^*$ na página 1
- Entradas de índice: $(k, \text{ponteiro})$

Índice Primitivo



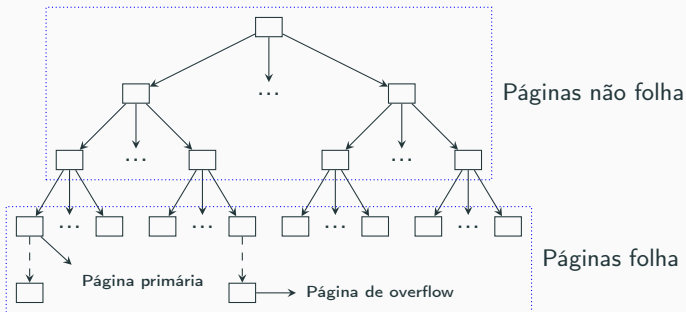
- Pesquisa binária eficiente
- Tamanho menor do que arquivo de dados
- Inserções e exclusões precisam mover entradas

Árvores ISAM I



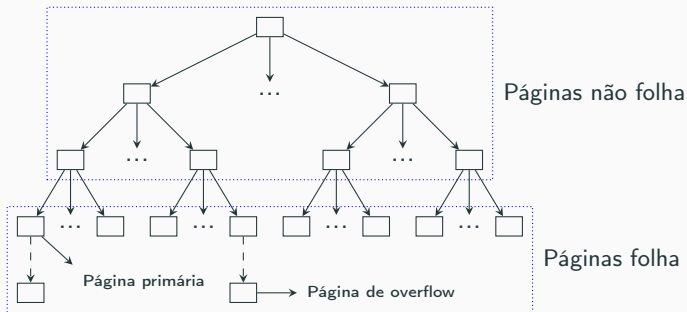
- *Indexed Sequential Access Method (ISAM)*
- Quando o índice é criado todas as páginas folhas primárias são alocadas sequencialmente ordenadas pela chave
- Depois disto elas não sofrem alterações (a ISAM é uma estrutura estática)
- Se uma página folha enche, cria-se uma página de *overflow*

Árvores ISAM II



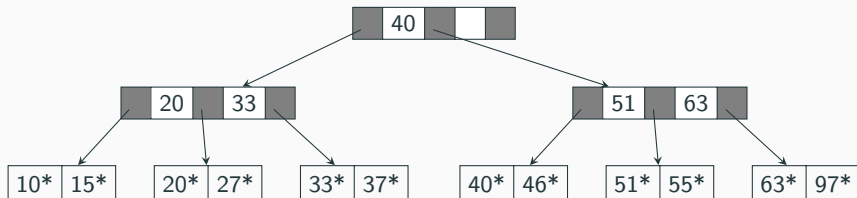
- A árvore ISAM suporta pesquisas por igualdade e por intervalo
- As operações de inserção e remoção são feitas apenas nas folhas
- As páginas de overflow podem ser excluídas, mas as páginas primárias devem permanecer (mesmo que estejam vazias)

Árvores ISAM III



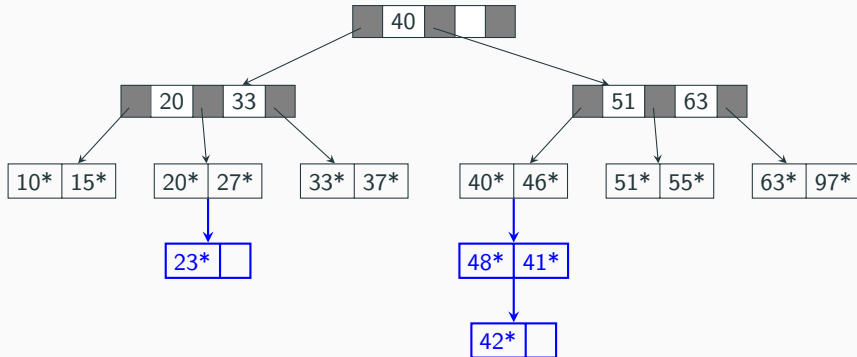
- Pelo fato das alterações ocorrerem apenas nas folhas, podem surgir longas cadeias de overflow que podem prejudicar o desempenho
- Por outro lado, o acesso concorrente não precisa bloquear as páginas não folha

Exercício

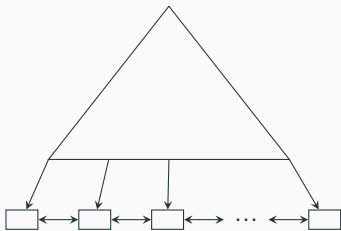


Inserir na árvore os registros com chave igual a 23, 48, 41, 42

Exercício - Resposta

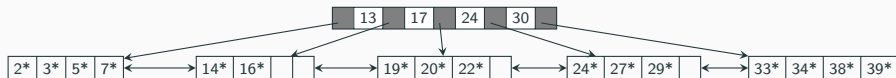


Árvores B+

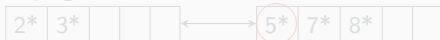


- Estrutura de índice mais usada
- As alterações mantêm a árvore balanceada
- Ocupação mínima de 50% das páginas (exceto raiz)
- As páginas folhas são ligadas
- Cada página contém m entradas, onde $d \leq m \leq 2d$
- d é chamado de ordem da árvore
- A raiz tem $1 \leq m \leq 2d$ registros

Inserção de Entradas



- Inserir o registro 8*
- Deve ser inserido na primeira folha, mas ela está cheia
- Precisamos dividir a página

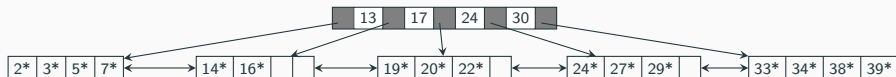


- A primeira entrada da segunda folha é **copiada** para o nó pai
- O nó pai também está cheio, temos que dividir novamente



- A primeira chave da segunda página (nó não folha) é **movida** para o nó pai
- Como não existe um nó pai, criamos uma nova raiz

Inserção de Entradas



- Inserir o registro 8*
- Deve ser inserido na primeira folha, mas ela está cheia
- Precisamos dividir a página

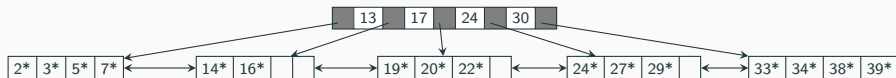


- A primeira entrada da segunda folha é **copiada** para o nó pai
- O nó pai também está cheio, temos que dividir novamente



- A primeira chave da segunda página (nó não folha) é **movida** para o nó pai
- Como não existe um nó pai, criamos uma nova raiz

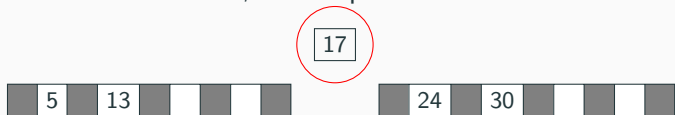
Inserção de Entradas



- Inserir o registro 8*
- Deve ser inserido na primeira folha, mas ela está cheia
- Precisamos dividir a página

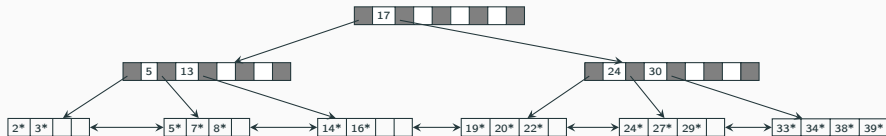


- A primeira entrada da segunda folha é **copiada** para o nó pai
- O nó pai também está cheio, temos que dividir novamente



- A primeira chave da segunda página (nó não folha) é **movida** para o nó pai
- Como não existe um nó pai, criamos uma nova raiz

Árvore após Inserção

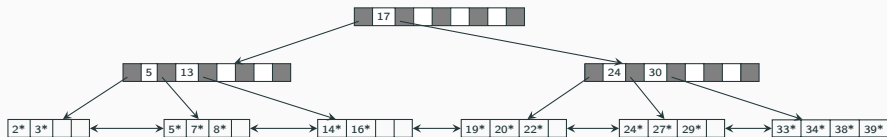


Inserção com Redistribuição



- A redistribuição não é muito vantajosa, em geral, ela apenas atrasa a divisão

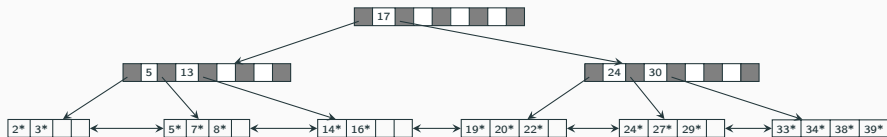
Exclusão de Entradas



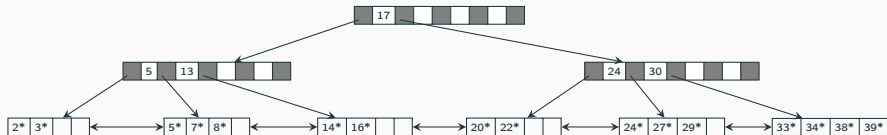
- Excluir a entrada 19*
- Podemos apenas remover da página (ela permanece com ocupação mínima)



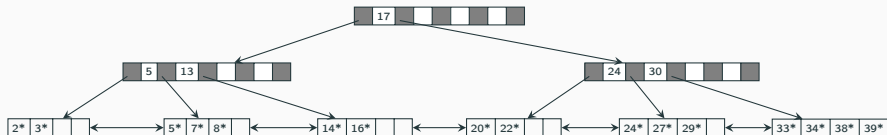
Exclusão de Entradas



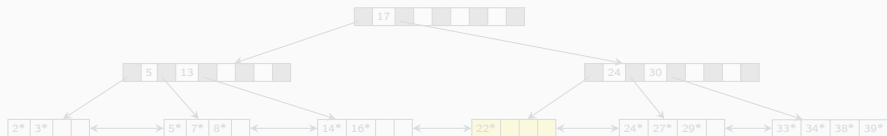
- Excluir a entrada 19*
- Podemos apenas remover da página (ela permanece com ocupação mínima)



Exclusão com Redistribuição I

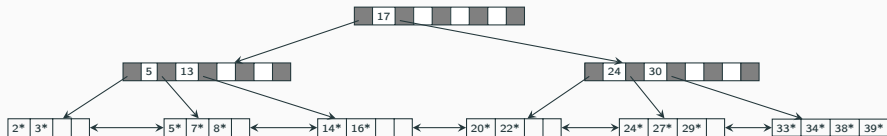


- Suponha agora a exclusão da entrada 20* (depois de excluirmos 19*)

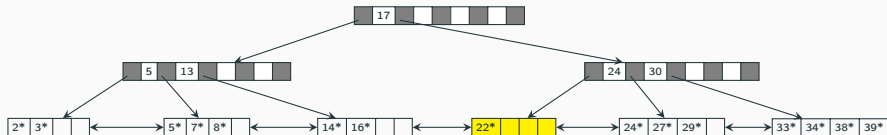


- A página fica com ocupação a baixo do mínimo

Exclusão com Redistribuição I



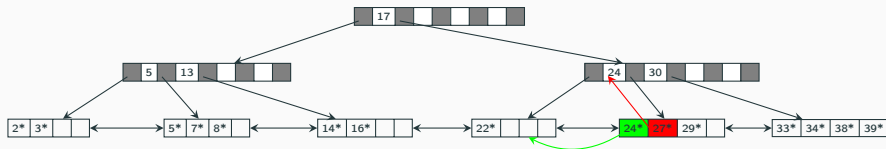
- Suponha agora a exclusão da entrada 20* (depois de excluirmos 19*)



- A página fica com ocupação a baixo do mínimo

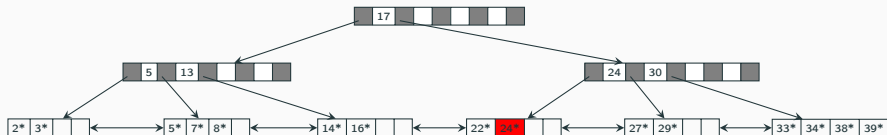
Exclusão com Redistribuição II

- Uma solução é puxar entradas de páginas irmãs para termos a ocupação mínima
- Quando isto acontece temos que atualizar o nó pai

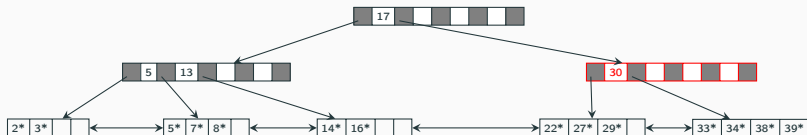


- Poderíamos optar por não redistribuir e juntar a página com uma irmã
- Porém, é melhor fazer a redistribuição primeiro porque podem haver inserções em seguida

Exclusão com Junção de Páginas I

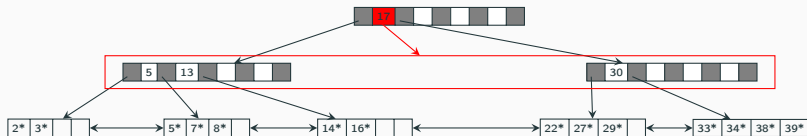


- Suponha agora se seja preciso remover a entrada 24*
- A página irmã não possui entradas sobrando para redistribuirmos
- Neste caso, precisamos juntar as páginas

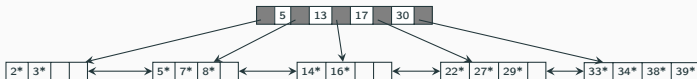


- Após a junção, o pai ficou com ocupação abaixo do mínimo
- Seu irmão também não possui entradas sobrando para redistribuirmos
- É necessário fazer outra junção

Exclusão com Junção de Páginas II



- Durante a junção, puxamos a entrada do pai que separa os dois nós



- Após a junção podemos remover a raiz original
- O nó resultante da junção passa a ser a nova raiz

Lidando com Duplicatas em Chaves de Pesquisa

- Quando o índice não agrupado deve permitir chaves de pesquisas duplicadas é necessário fazer alterações nas folhas da árvore
- Cada entrada nas folhas fica ligada a uma lista de **rid**

Compressão de Chaves

- A técnica de compressão de chaves é muito usada para chaves que contenham campos textuais
- Chaves de pesquisa maiores ocupam mais espaço nas páginas e prejudicam o espalhamento da árvore
- Os valores de chave do índice servem apenas para direcionar o tráfego, não precisamos do valor completo
- Exemplo:



- É preciso verificar o maior valor do filho a esquerda e o menor valor do filho a direita

Carregamento em Massa

- O carregamento em massa (ou *bulk-loading*) ocorre quando precisamos de carregar todos os dados de uma tabela (por exemplo, na restauração de um backup)
- Neste caso, inserir registro por registro e atualizar o índice pode tomar muito tempo
- Se houver mais de um índice, podemos fazer o carregamento em massa considerando o índice agrupado
- Primeiro ordenamos os dados pela chave e depois criamos o índice

O Conceito de Ordem

- Na prática a ordem da árvore B+ não é o número de entradas, mas sim a quantidade de espaço ocupado
- Os principais motivos para isto são:
 - Nós folhas e nós não folhas podem guardar números diferentes de entradas
 - Podem existir registros de tamanho variável

Referências I



Date, C. J. (2004).

Introdução a sistemas de bancos de dados.

Elsevier, Rio de Janeiro.



Elmasri, R. and Navathe, S. B. (2011).

Sistemas de banco de dados.

Pearson Addison Wesley, São Paulo, 6 edition.



Ramakrishnan, R. and Gehrke, J. (2008).

Sistemas de gerenciamento de banco de dados.

McGrawHill, São Paulo, 3 edition.



Silberschatz, A., Korth, H. F., and Sudarshan, S. (2007).

Sistema de bancos de dados.