

ESTRUTURAS DE INDEXAÇÃO

ÍNDICES MULTINÍVEL

Rodrigo Salvador Monteiro

Slides adaptados do Prof Sean Siqueira (sean@uniriotec.br)

Índices Multinível

- Uma pesquisa binária sobre um índice com b_i blocos
 - ▶ requer aproximadamente $(\log_2 b_i)$ acessos a blocos
 - ▶ cada etapa do algoritmo reduz a parte do arquivo índice que continuamos a pesquisar por um fator de 2.
- índice multinível
 - ▶ Visa reduzir a parte do índice que continuamos a pesquisar através de bfr_i (fator de bloco para o índice) maior do que 2.
- $bfr_i = \text{fan-out (fo)}$ do índice multinível.
- Pesquisar em um índice multinível requer aproximadamente $(\log_{fo} b_i)$ acessos a blocos
 - ▶ número muito menor do que a pesquisa binária se o fan-out for maior do que 2.

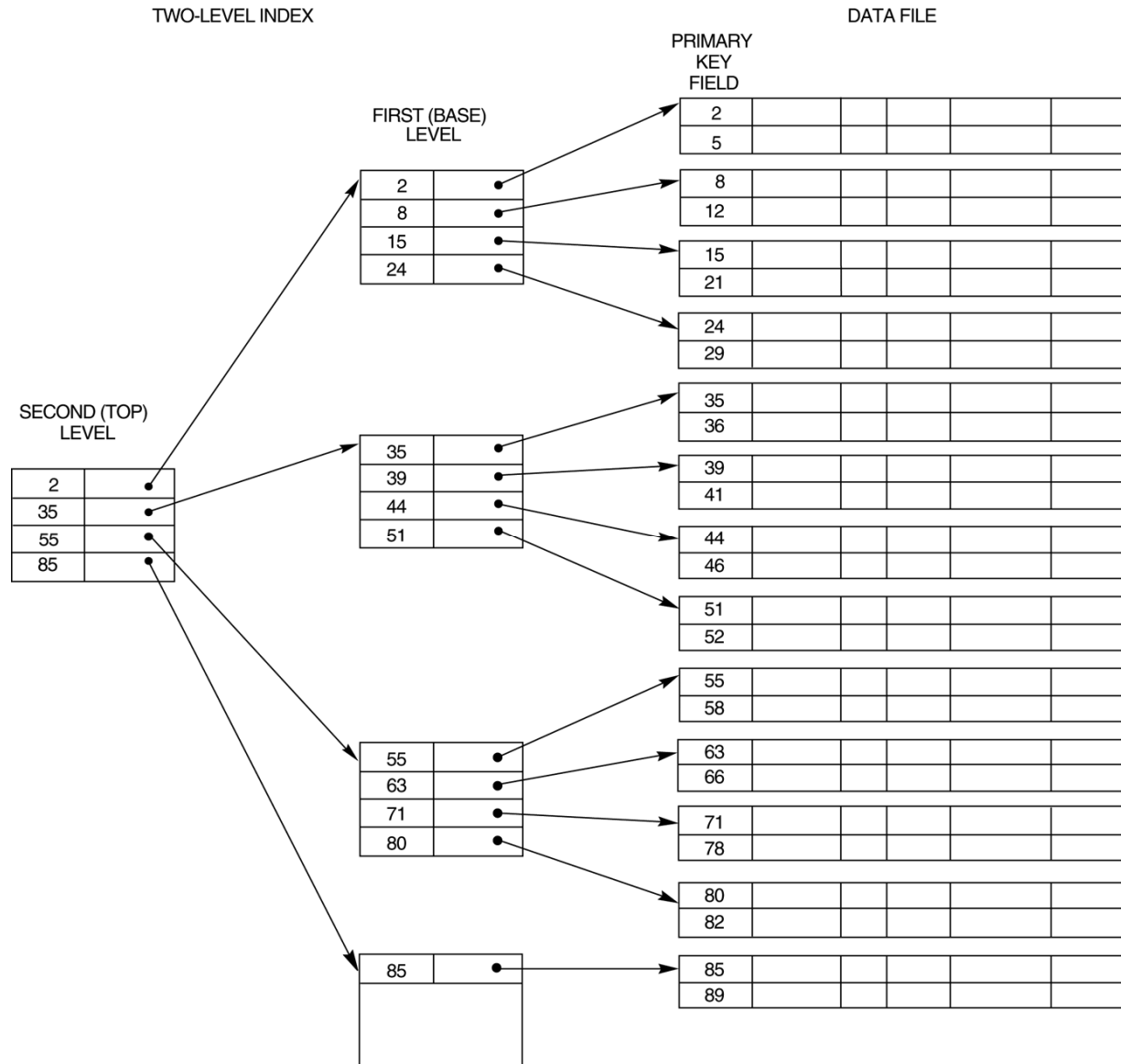
Índices Multinível (cont.)

- primeiro nível ou nível base
 - ▶ arquivo ordenado com um valor distinto para cada $K(i)$.
- Índice de segundo nível
 - ▶ índice principal para o primeiro nível
 - ▶ uma entrada para cada bloco do primeiro nível.
 - ▶ bfr_i para o segundo nível – e para todos os níveis subseqüentes
 - mesmo que para o índice do primeiro nível...
- Suponha...
 - ▶ Índice de primeiro nível com r_1 entradas e $bfr_i = fo$
 - ▶ Então... o primeiro nível precisa de $\lceil r_1/fo \rceil$ blocos, que é portanto o número de entradas r_2 necessárias no segundo nível de índice.

Índices Multinível (cont.)

- O terceiro nível
 - ▶ índice principal para o segundo nível
 - ▶ uma entrada para cada bloco do segundo nível
 - número de entradas do terceiro nível seja $r_3 = \lceil r_2 / f_0 \rceil$.
- Quando um segundo nível é necessário?
 - ▶ se o primeiro precisar de mais do que um bloco de armazenamento de disco
 - ▶ terceiro nível somente se o segundo precisar de mais do que um bloco, e assim sucessivamente...
- Até que todas as entradas de algum nível t do índice caibam em um único bloco.
 - ▶ t é chamado de **nível de índice superior**.

Exemplo



Exemplo

- Suponha que o índice secundário denso do exemplo anterior seja convertido em um índice multinível. Calculamos o fator de bloco do índice $bfr_i = 68$ entradas de índice por bloco, que é também o fan-out fo para o índice multinível; o número de blocos de primeiro nível $b_1 = 442$ blocos também foi calculado.
- Qual é o número de blocos de segundo nível?
 $b_2 = \lceil b_1/fo \rceil = \lceil 442/68 \rceil = 7$ blocos
- Qual é o número de blocos de terceiro nível?
 $b_3 = \lceil b_2/fo \rceil = \lceil 7/68 \rceil = 1$ bloco
- Qual é o número de blocos de quarto nível?
Não tem quarto nível
- Qual é o índice superior?
 $t = 3$; o terceiro é o nível superior do índice
- Quantos acessos a blocos são necessários para acessar um registro utilizando este índice multinível?
 $3 + 1 = 4$

Problemas com Índices Multinível

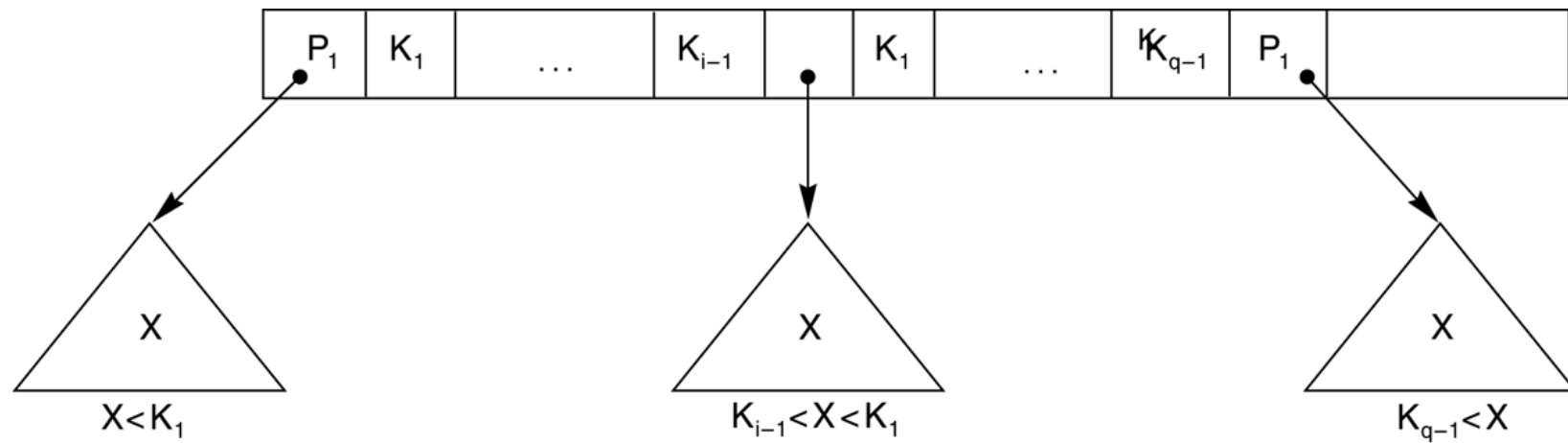
- Inclusões e exclusões

- ▶ Todos os níveis de índices são arquivos fisicamente ordenados

- ▶ Solução:

- Adotar um índice multinível que deixa algum espaço em cada um de seus blocos para inserir novas entradas.
 - Índice multinível dinâmico

Árvores de busca...



Revisão de estrutura de árvore

- Uma árvore é formada por nós;
- Cada nó tem um pai (exceto o nó raiz) e vários nós filhos (zero ou mais)
- Nó que não possui filhos chama-se nó folha;
- Nó que não é nó folha é chamado nó interno;
- O nível de um nó é sempre uma unidade maior que seu pai, com o nível do nó raiz sendo zero;
- Uma subárvore de um nó consiste daquele nó e todos os seus descendentes

Árvores-B

- Árvore sempre balanceada
- Espaço desperdiçado através da exclusão, caso haja algum, nunca se torne excessivo
- Algoritmos para inclusão e exclusão mais complexos

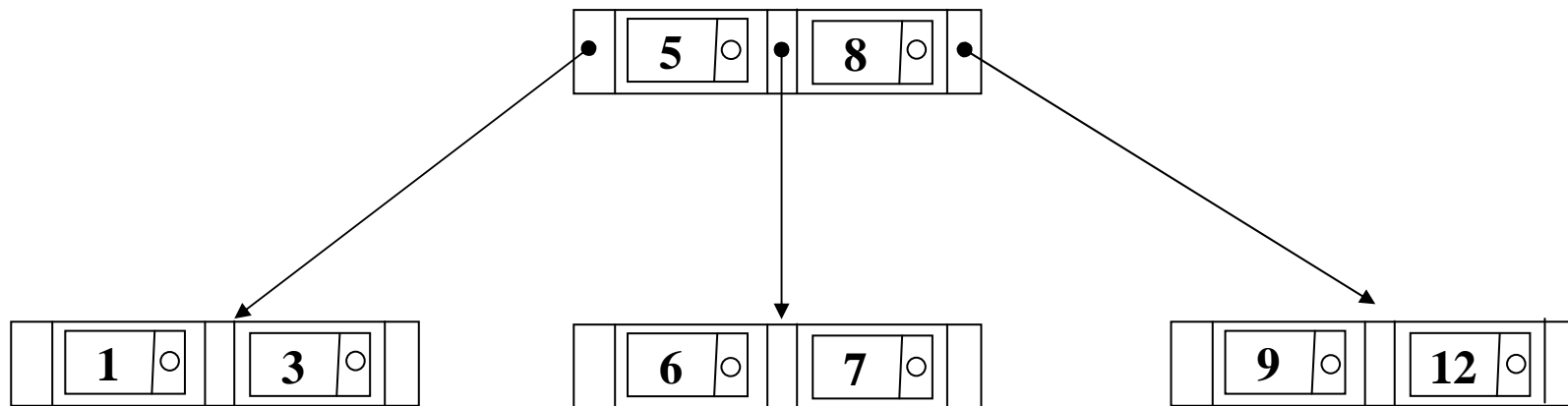
Árvores-B

- Uma árvore-B de ordem p quando utilizada como uma estrutura de acesso em um campo chave para pesquisar registros em um arquivo de dados pode ser definida da seguinte maneira:
 - ▶ Cada nó interno da árvore-b é da forma $\langle P_1, \langle K_1, Pr_1 \rangle, P_2, \langle K_2, Pr_2 \rangle, \dots, \langle K_{q-1}, Pr_{q-1} \rangle, P_q \rangle$, onde $q \leq p$.
 - ▶ Dentro de cada nó, $K_1 < K_2 < \dots < K_{q-1}$
 - ▶ Para todos os valores X do campo chave de pesquisa na subárvore referenciada por P_i , temos: $K_{i-1} < X < K_i$ para $1 < i < q$; $X < K_i$ para $i=1$; e $K_{i-1} < X$ para $i = q$
 - ▶ Cada nó possui no máximo p ponteiros de árvore

Árvores-B (cont.)

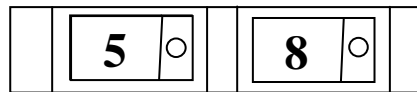
- ▶ Cada nó, exceto os nós raiz e as folhas, possui pelo menos $\lceil p/2 \rceil$ = ponteiros de árvore. O nó raiz possui pelo menos dois ponteiros de árvore a não ser que ele seja o único nó na árvore.
- ▶ Um nó com q ponteiros de árvore, $q \leq p$, possui $q-1$ valores de campo chave de pesquisa (e portanto possui $q-1$ ponteiros de dados).
- ▶ Todos os nós folhas encontram-se no mesmo nível.
- ▶ Os nós folhas possuem a mesma estrutura que os nós internos, exceto pelo fato de que todos os seus ponteiros de árvore P_i são nulos.

Uma árvore-B de ordem $p=3$



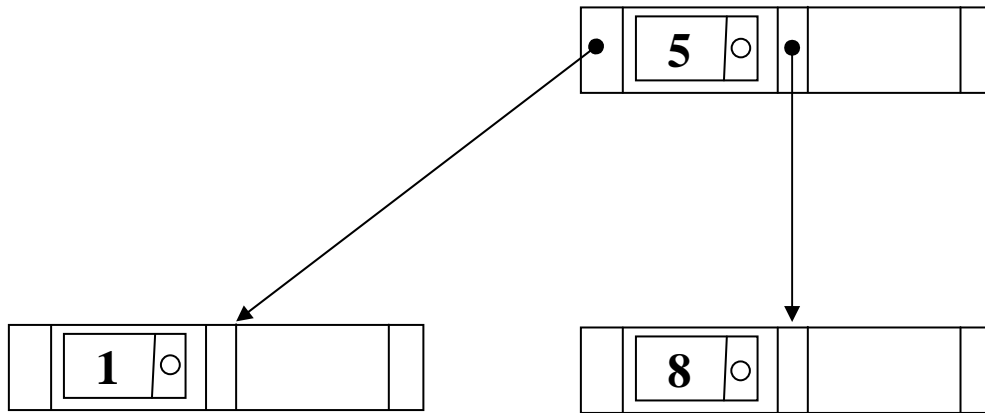
Uma árvore-b de ordem $p=3$

Inserção dos valores 8, 5, 1



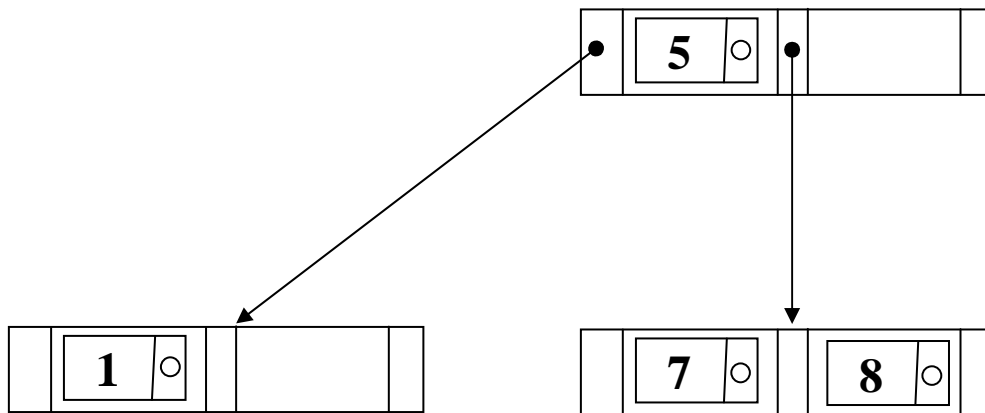
Uma árvore-b de ordem $p=3$

Inserção dos valores 8, 5, 1, 7



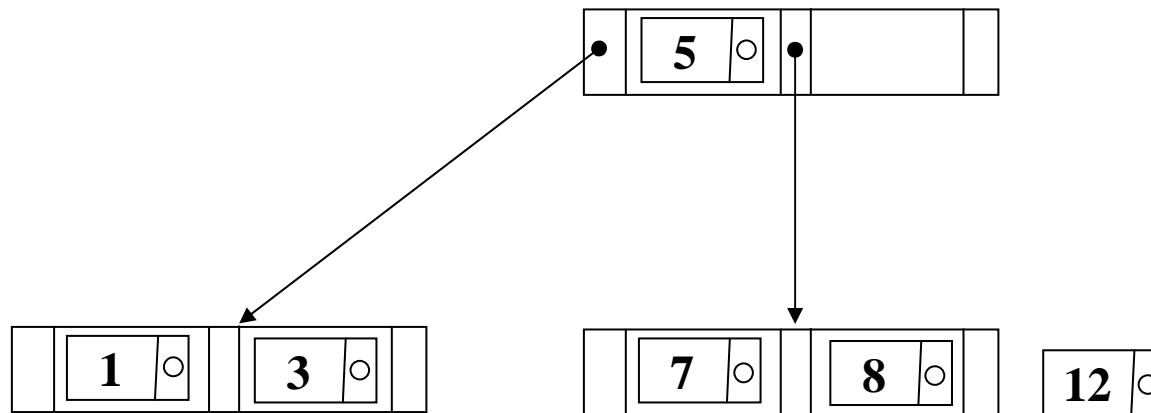
Uma árvore-b de ordem $p=3$

Inserção dos valores 8, 5, 1, 7, 3



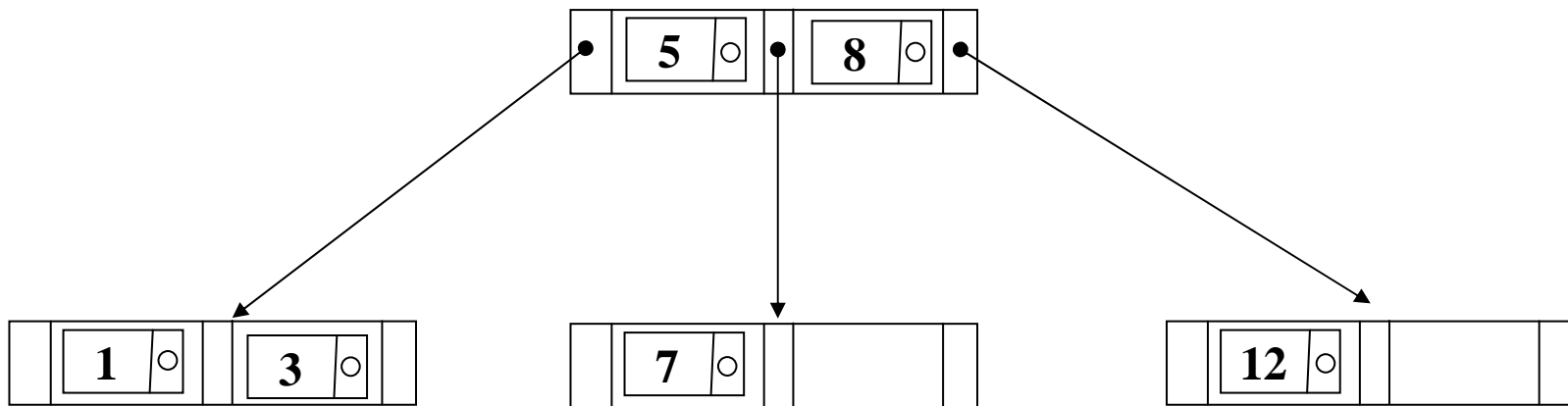
Uma árvore-b de ordem $p=3$

Inserção dos valores 8, 5, 1, 7, 3, 12



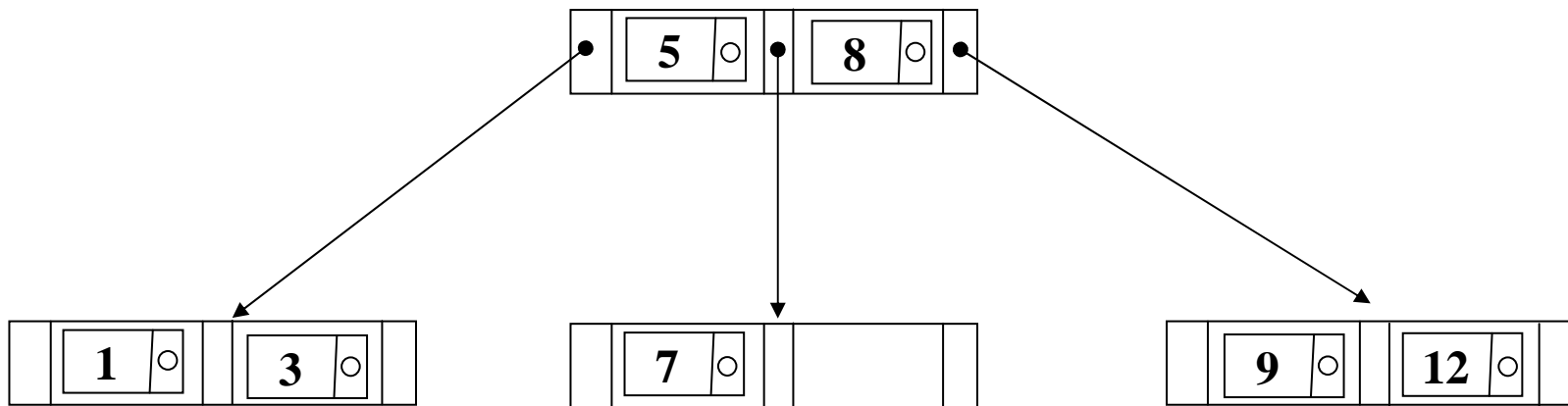
Uma árvore-b de ordem $p=3$

Inserção dos valores 8, 5, 1, 7, 3, 12, 9



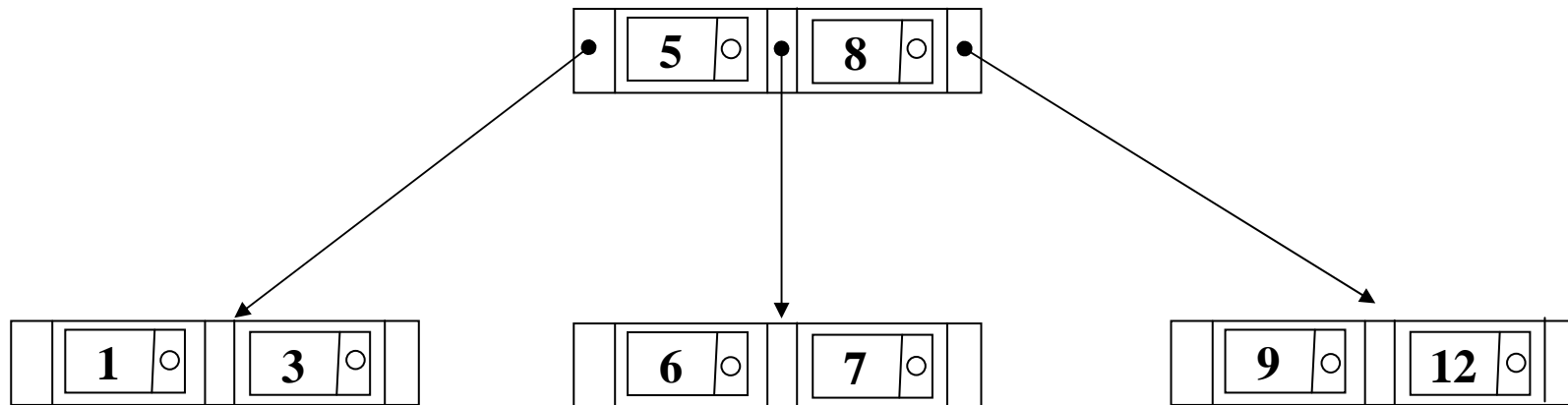
Uma árvore-b de ordem $p=3$

Inserção dos valores 8, 5, 1, 7, 3, 12, 9, 6



Uma árvore-b de ordem $p=3$

Inserção dos valores 8, 5, 1, 7, 3, 12, 9, 6



Exemplo

- Suponha que:
 - ▶ O campo de pesquisa seja $V = 9$ bytes de tamanho
 - ▶ O tamanho do bloco seja $B = 512$ bytes
 - ▶ Um ponteiro de registro (dados) seja $P_r = 7$ bytes
 - ▶ Um ponteiro de bloco seja $P = 6$ bytes
- Quantos ponteiros de árvore (ou ordem p) poderemos ter em cada nó de modo que cada nó da árvore-B corresponda a um bloco de disco?

$$(p * P) + ((p-1) * (P_r + V)) \leq B$$

$$(p * 6) + ((p-1) * (7 + 9)) \leq 512$$

$$(22 * p) \leq 528$$

$$p \leq 24$$

$$p = 23$$

Exemplo

- Suponha que:
 - ▶ o campo de pesquisa seja um campo chave sem ordenação
 - ▶ construímos uma árvore-b nesse campo
 - ▶ cada nó da árvore-B esteja 69% cheio
- Quantos ponteiros e valores teremos em média em cada nó?
cada nó, em média, terá $p * 0,69 = 23 * 0,69 = 15,87$ ou
aproximadamente 16 ponteiros e portanto 15 valores de campos chave
- Qual é o fan-out médio?
fan-out médio (fo) = 15
- Supondo uma árvore-B até o nível 3, quantos nós, entradas e ponteiros temos em cada nível?

Raiz	1 nó	15 entradas	16 ponteiros
Nível 1	16 nós	240 entradas	256 ponteiros
Nível 2	256 nós	3840 entradas	4096 ponteiros
Nível 3	4096 nós	61.440 entradas	