



# Estruturas de dados II

## Árvore B+

André Tavares da Silva  
andre.silva@udesc.br

# Conceitos da árvores B+

- É uma variação da estrutura da árvore B
- Características e principais diferenças para as árvores B:
  - Todas as chaves são mantidas em folhas;
  - As chaves são repetidas em nós não folha que formam índices;
  - As folhas são ligadas oferecendo um caminho sequencial para percorrer as chaves;
  - Os nós folhas apontam para os registros em arquivo. Isso permite que os dados e os índices sejam armazenados em arquivos separados.
- Ordem da árvore
  - O número máximo de chaves representa ordem da árvore
  - Todo nó tem no máximo  $M-1$  chaves e no mínimo  $M/2$  (50%) de ocupação
    - Essa característica é válida para  $M > 3$ , já que se com  $M=3$ , quando dividir sempre ficará uma chave única em algum nó (diferente da árvore B que sempre perde um nó ao dividir).

# Árvores B+

- Possui complexidade de tempo logarítmo
  - Acesso aleatório:  $O(\log n)$
  - Inserção:  $O(\log n)$
  - Remoção:  $O(\log n)$
- Aumenta a eficiência da localização do próximo registro
  - Acesso sequencial é de  $O(1)$  em vez de  $O(\log n)$
- Não é necessário manter nenhum ponteiro de registro em nós não folha
  - Somente os nós folha apontam para os registros
  - Nós não folha só guardam o valor da chave de busca

# Utilização da árvores B+

- Banco de Dados
  - Muitos Bancos de Dados são construídos usando o mecanismo de Árvores B+: SQLServer, Oracle, PostgreSQL, MariaDB, SQLite,...
- Sistemas de Arquivos
  - Sistemas de arquivos como ReiserFS, XFS e JFS2 usam Árvores B+ para operações de leitura e escrita em blocos de disco.
- Sistemas de gerência de dados
  - Sistemas como CouchDB, Tokyo Cabinet, Tokyo Tyrant,....

# Operações em árvores B+

- A inserção de uma nova chave em uma árvore B+ é semelhante a inserção em uma árvore B:
  - Inicia localizando o nó folha que deverá ser adicionada a chave
  - Passos
    - Localizar a folha dentro da qual a chave deve ser inserida;
    - Localizar a posição de inserção dentro da folha;
    - Inserir a chave;
    - Se após a inserção a folha estiver completa, realizar a partição (divisão) da página
  - Para a divisão, as  $M-1$  chaves são divididas em dois grupos:
    - As  $M/2$  chaves menores ficam na folha esquerda;
    - As demais chaves ficam na folha da direita;
    - A menor chave da direita é **copiada** para o nó pai.

# Exemplo

$M = 4 \rightarrow 25, 11, 3, 8, 14, 40, 20, 9, 2, 6, 28, 10, 1$


Índices


Dados

$M = 4 \rightarrow$  **25**, 11, 3, 8, 14, 40, 20, 9, 2, 6, 28, 10, 1

25			

$M = 4 \rightarrow 25, \mathbf{11}, 3, 8, 14, 40, 20, 9, 2, 6, 28, 10, 1$

<b>11</b>	25		



$M = 4 \rightarrow 25, 11, \mathbf{3}, 8, 14, 40, 20, 9, 2, 6, 28, 10, 1$

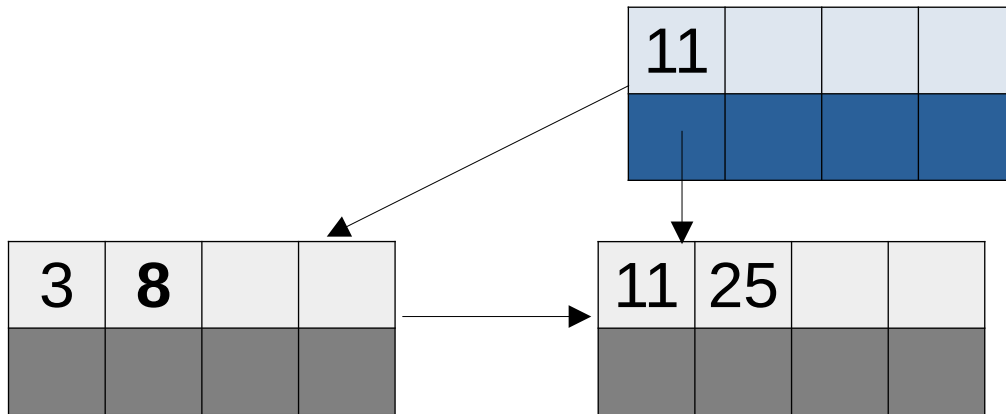
<b>3</b>	11	25	

$M = 4 \rightarrow 25, 11, 3, \mathbf{8}, 14, 40, 20, 9, 2, 6, 28, 10, 1$

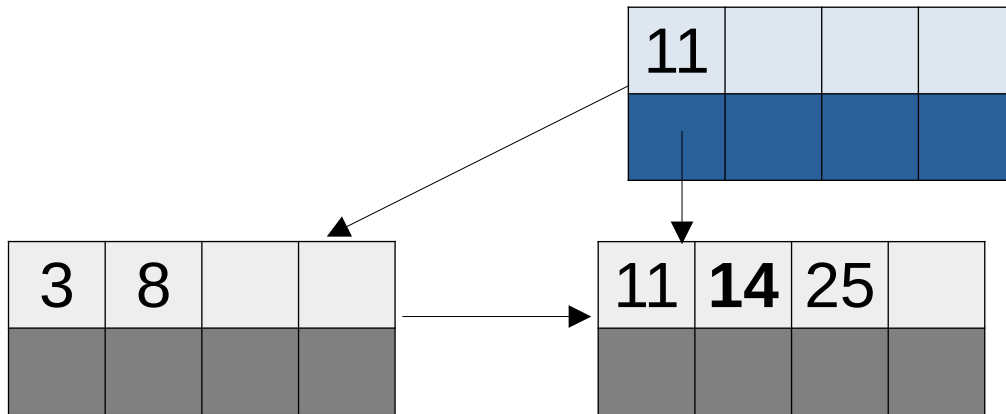
*Split!*

3	<b>8</b>	11	25

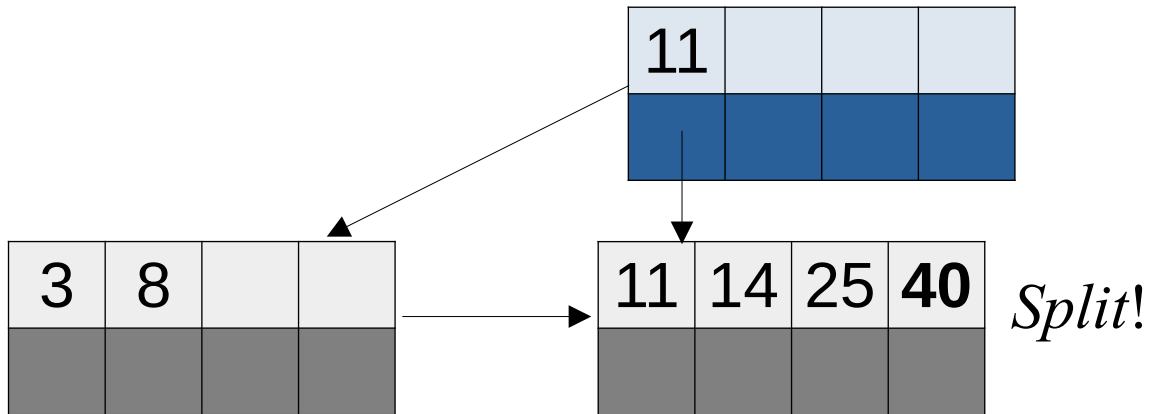
$M = 4 \rightarrow 25, 11, 3, \mathbf{8}, 14, 40, 20, 9, 2, 6, 28, 10, 1$



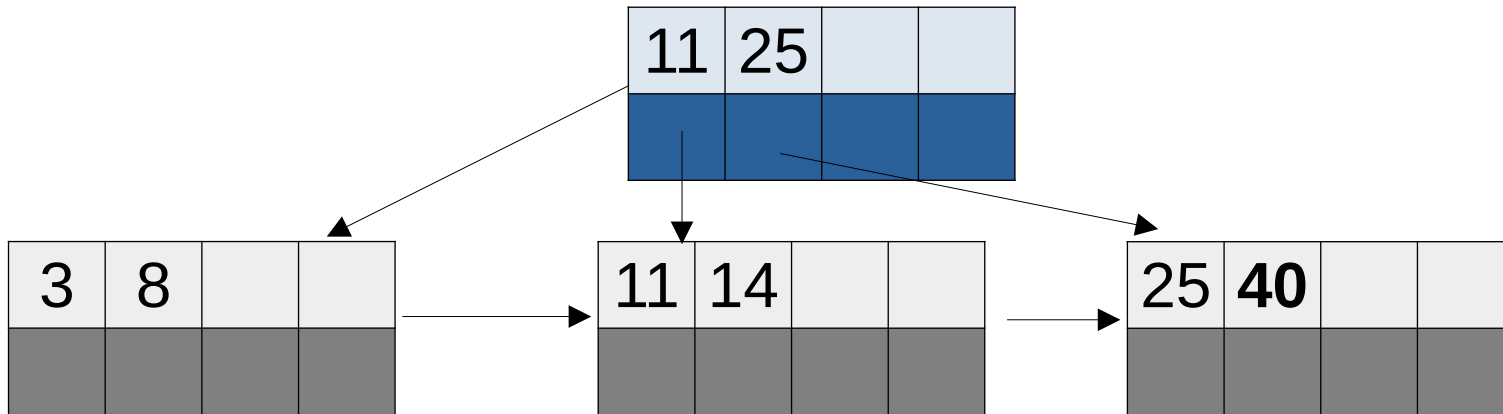
$M = 4 \rightarrow 25, 11, 3, 8, \mathbf{14}, 40, 20, 9, 2, 6, 28, 10, 1$



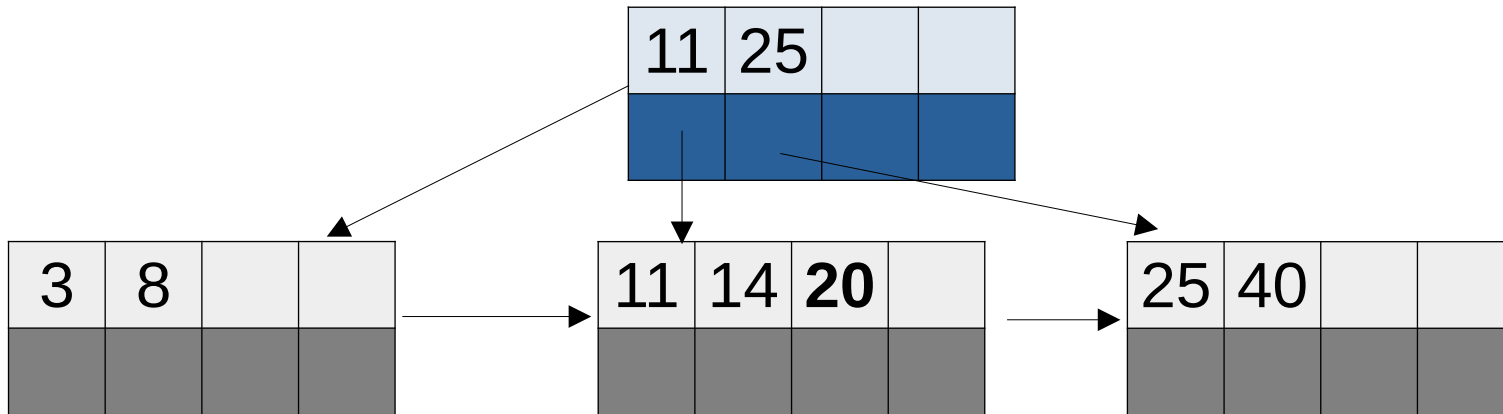
$M = 4 \rightarrow 25, 11, 3, 8, 14, \mathbf{40}, 20, 9, 2, 6, 28, 10, 1$



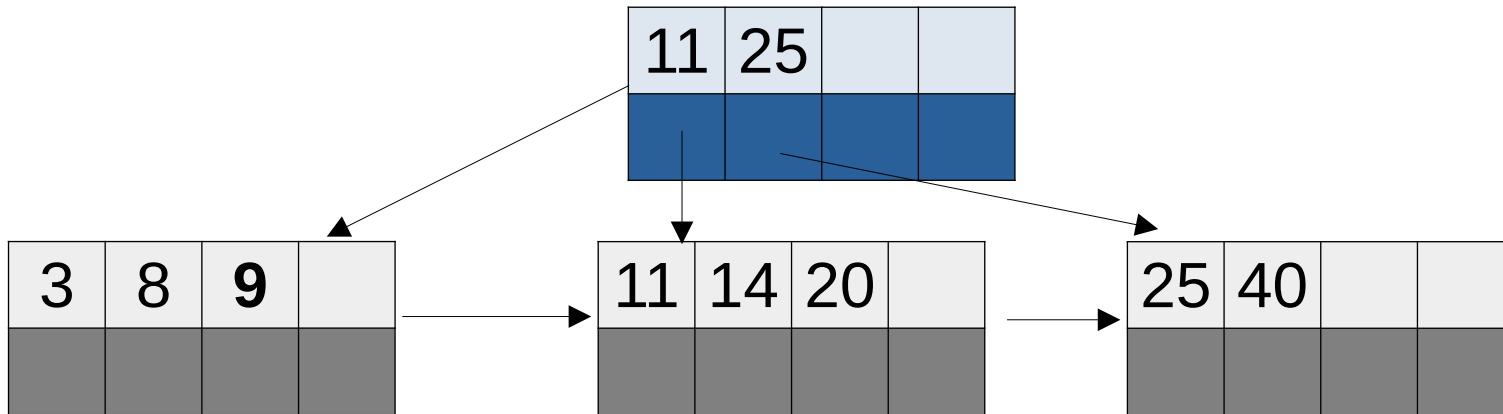
$M = 4 \rightarrow 25, 11, 3, 8, 14, \mathbf{40}, 20, 9, 2, 6, 28, 10, 1$



$M = 4 \rightarrow 25, 11, 3, 8, 14, 40, \mathbf{20}, 9, 2, 6, 28, 10, 1$

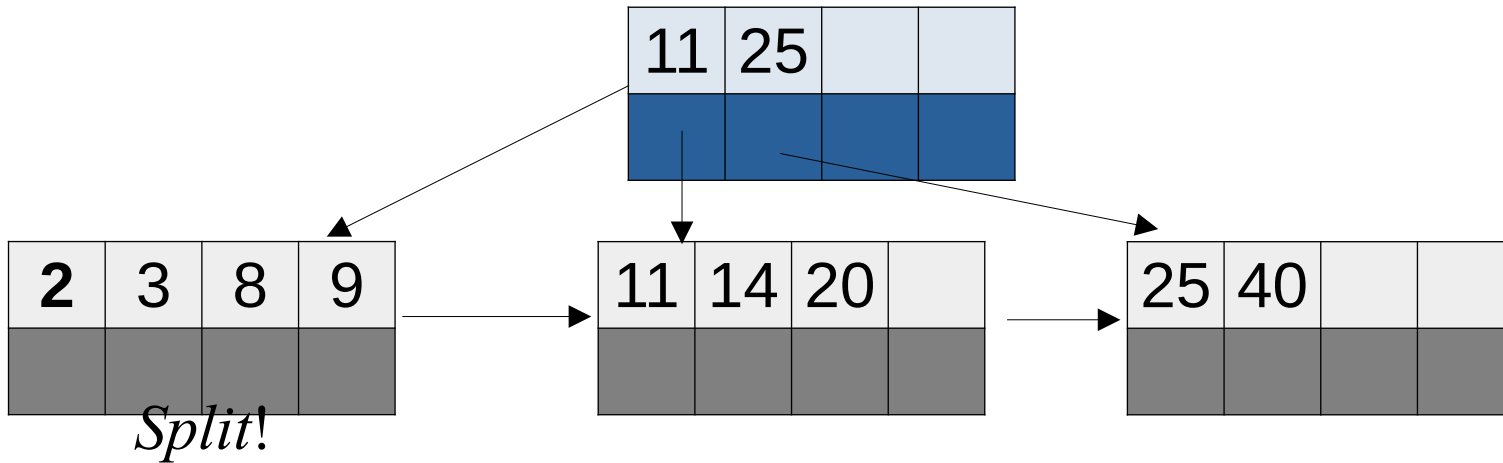


$M = 4 \rightarrow 25, 11, 3, 8, 14, 40, 20, \mathbf{9}, 2, 6, 28, 10, 1$

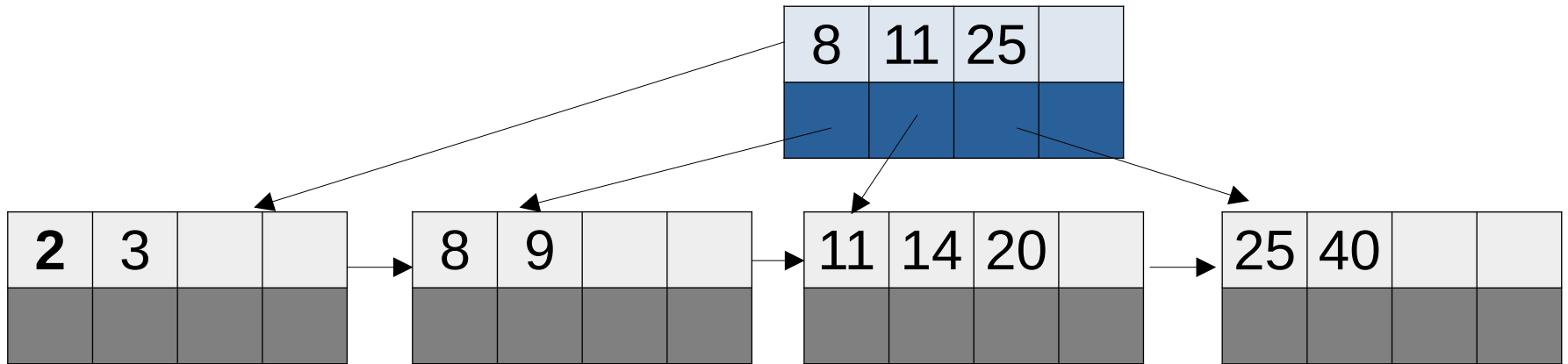




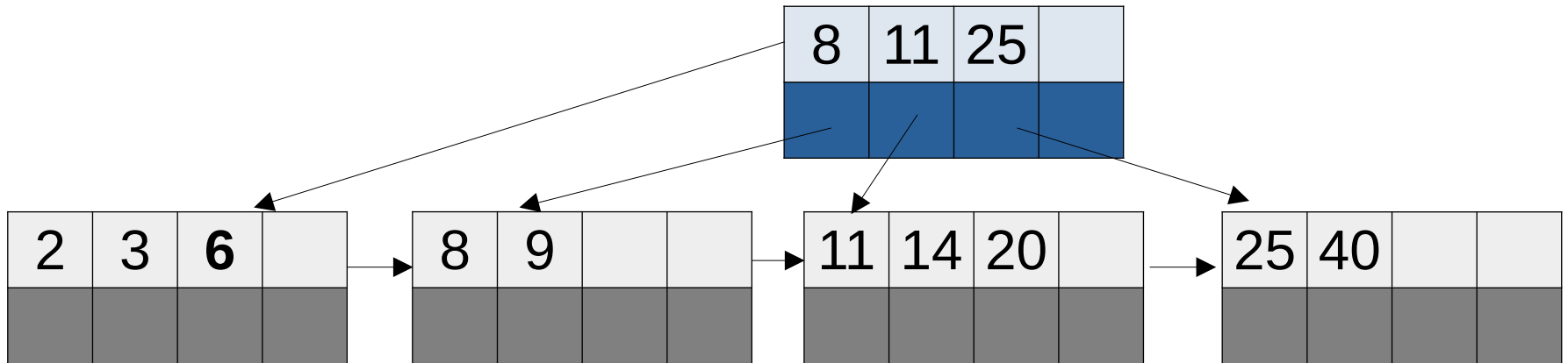
$M = 4 \rightarrow 25, 11, 3, 8, 14, 40, 20, 9, \mathbf{2}, 6, 28, 10, 1$



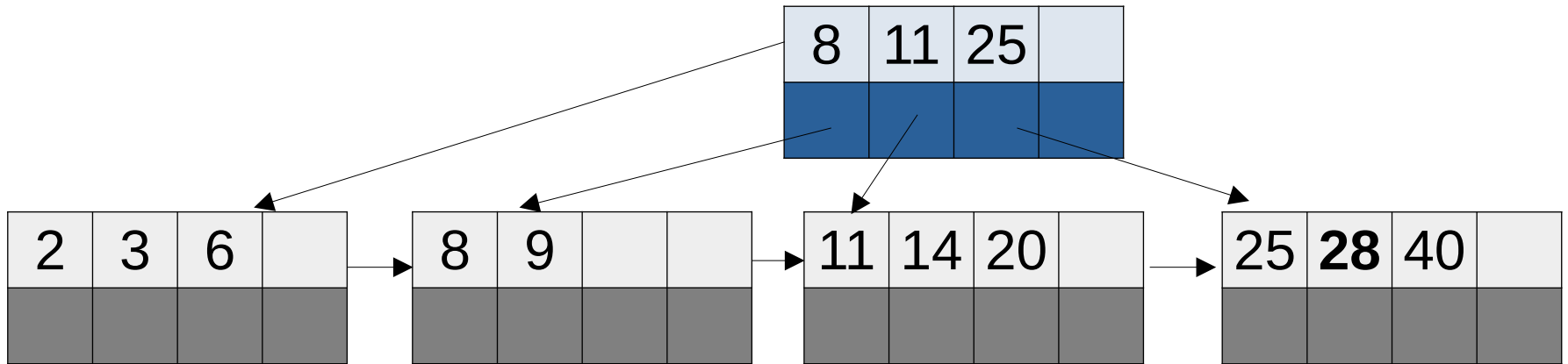
$M = 4 \rightarrow 25, 11, 3, 8, 14, 40, 20, 9, \mathbf{2}, 6, 28, 10, 1$



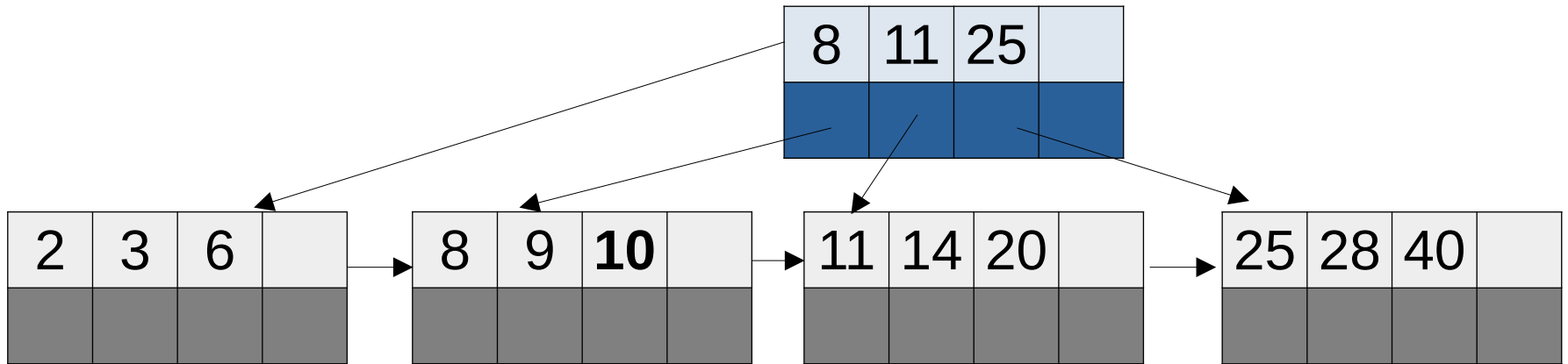
$M = 4 \rightarrow 25, 11, 3, 8, 14, 40, 20, 9, 2, \mathbf{6}, 28, 10, 1$



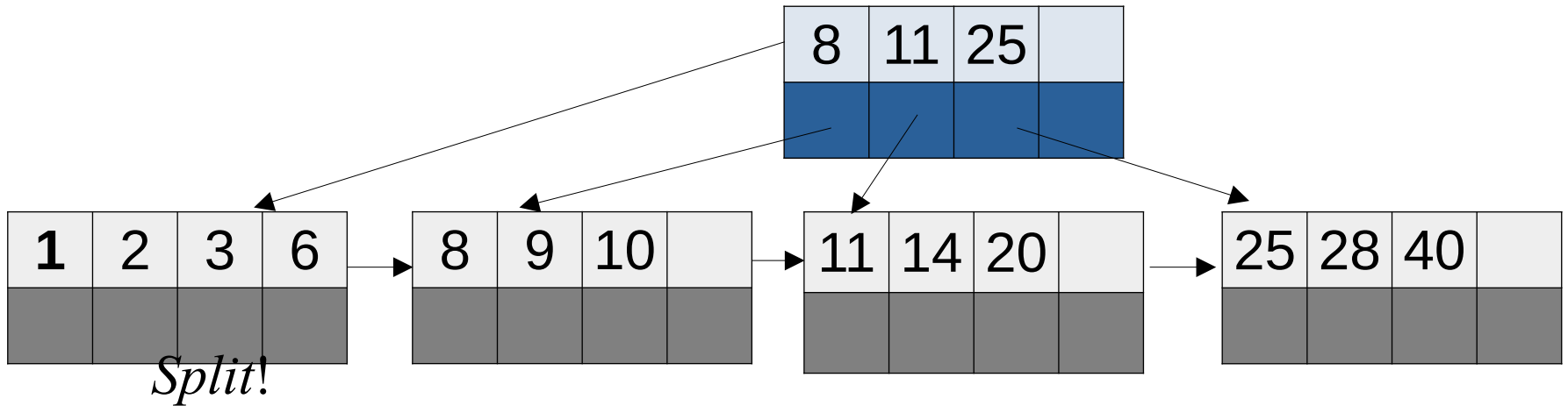
$M = 4 \rightarrow 25, 11, 3, 8, 14, 40, 20, 9, 2, 6, \mathbf{28}, 10, 1$



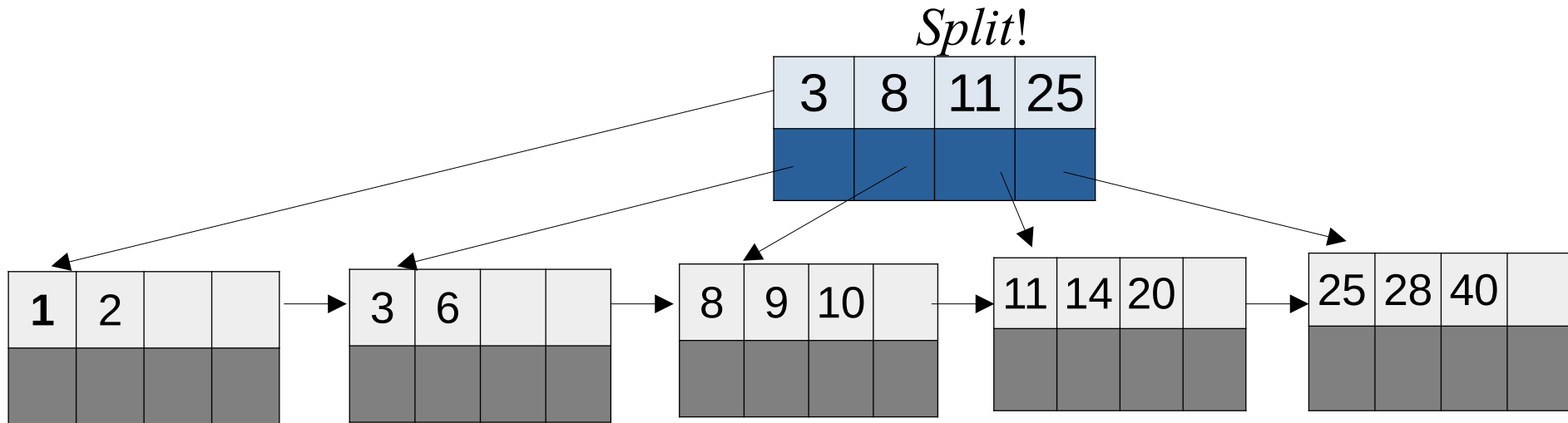
$M = 4 \rightarrow 25, 11, 3, 8, 14, 40, 20, 9, 2, 6, 28, \mathbf{10}, 1$



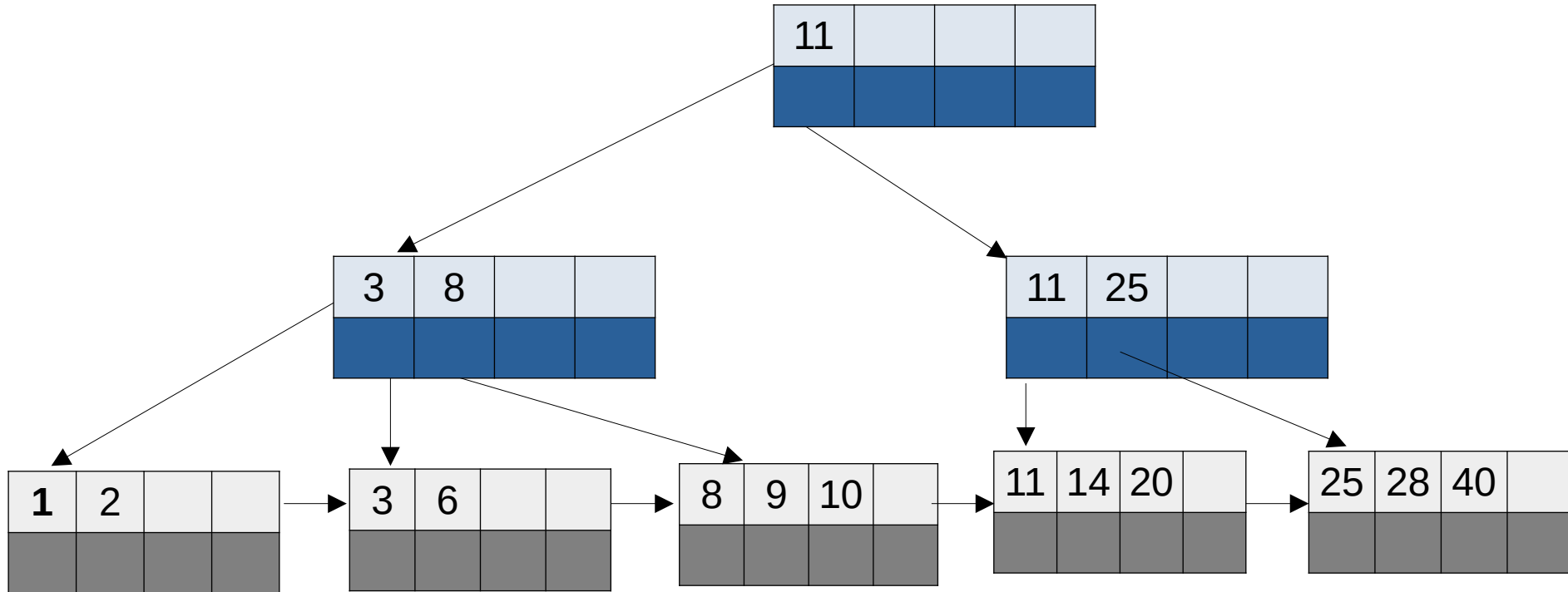
$M = 4 \rightarrow 25, 11, 3, 8, 14, 40, 20, 9, 2, 6, 28, 10, 1$



$M = 4 \rightarrow 25, 11, 3, 8, 14, 40, 20, 9, 2, 6, 28, 10, 1$



$M = 4 \rightarrow 25, 11, 3, 8, 14, 40, 20, 9, 2, 6, 28, 10, 1$





# Considerações sobre implementação de árvores B+

- Pode-se usar três arquivos:
  - Um para armazenar os metadados
    - Ponteiro para raiz da árvore
    - *Flag* indicando se a raiz é folha
  - Um arquivo para armazenar os índices (nós internos)
  - Um arquivo para armazenar os dados (folhas)

# Considerações sobre implementação de árvores B+

- O **arquivo de índice** é estruturado em nós (blocos/páginas);
- Contém os nós internos da árvore;
- Cada nó possui:
  - Um inteiro  $m$  representando o número de chaves no nó;
  - Flag indicando se aponta para nó folha (dados) ou não (índices);
  - Ponteiro para nó pai (para facilitar implementação);
  - $P_0, (s_1, p_1), (s_2, p_2), \dots, (s_m, p_m)$ , onde
    - $s_x$  é uma chave
    - $p_x$  é um ponteiro para uma página dentro do arquivo

# Considerações sobre implementação de árvores B+

- O **arquivo de dados** é estruturado em nós (blocos/páginas);
- Contém os nós folhas da árvore;
- Cada nó possui:
  - Um inteiro  $m$  representando o número de registros no nó;
  - Ponteiro para nó pai (para facilitar implementação);
  - Ponteiro para a próxima página;
  - **$m$**  registros.

# Considerações sobre implementação de árvores B+

- Se um sistema de arquivo tem tamanho de bloco de  $B$  bytes e as chaves a serem armazenadas têm tamanho de  $k$  bytes, a árvore B+ mais eficiente é com ordem mínima (50% de  $M$ ) de  $d = (B / k) - 1$
- Exemplo prático:
  - Tamanho do bloco  $B$  em disco de 4096 bytes;
  - Tamanho da chave  $k$  de 4 bytes (inteiro, por exemplo);
  - $d = (4096 / 4) - 1 = 2023$
- Qual o tamanho da ordem máxima nessa situação?
  - $M = 2 * d = 2046$

# Considerações sobre implementação de árvores B+

- As funções de inclusão e remoção devem garantir que as ordens máximas ( $M-1$ ) e mínimas ( $M/2$ ) da árvore sejam respeitadas (para  $M > 3$ );
- Em algumas implementações os nós folhas podem possuir  $M$  elementos (*overflow* só ocorre com valor maior que  $M$ ). Alguns autores consideram que os nós folha (dados) possam ter tamanho distinto dos nós de índice;
- Nós de índice também podem ser representados como nos slides sobre árvores B (diferente dos nós folha):





# Estruturas de dados II

## Árvore B+

André Tavares da Silva  
andre.silva@udesc.br