

Árvores-B: Definição

Profa. Dra. Cristina Dutra de Aguiar Ciferri

Árvore-B

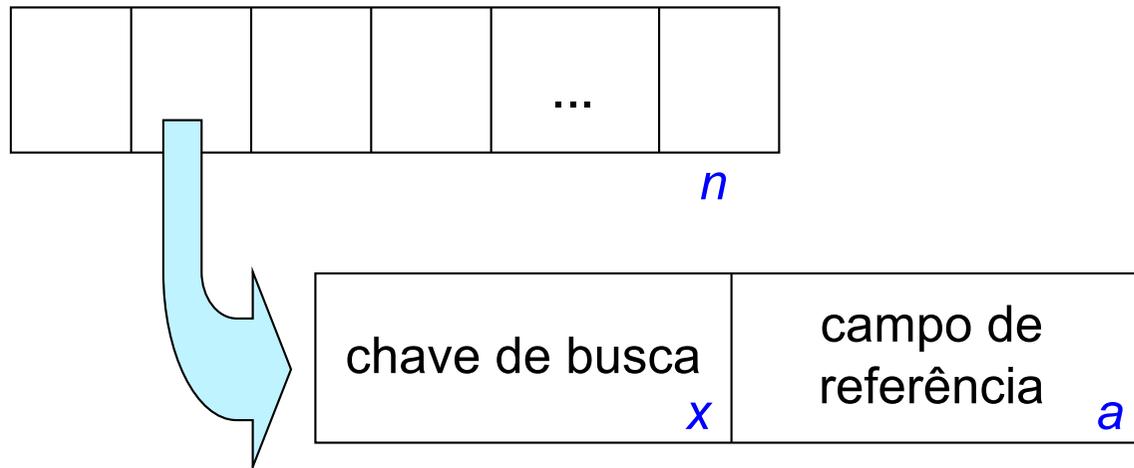
- Método genérico para o armazenamento e a recuperação de dados
 - voltado para arquivos volumosos
 - proporciona rápido acesso aos dados
 - possui custo mínimo de *overhead*
 - Referência
 - Bayer, R.; McCreight, E. *Organization and Maintenance of Large Ordered Indexes*.
 - Boing Corporation, 1972.
-

Sistemas de Banco de Dados

- 1979
 - árvore-B é, de fato, a organização padrão para indexação
 - Atualmente
 - amplamente utilizada
 - existem variantes
 - árvore-B⁺
 - árvore-B^{*}
 - árvore-B paralela
-

Características Gerais

- Organizar e manter um índice para um arquivo de acesso aleatório altamente dinâmico
- Índice
 - n elementos (x,a) de tamanho fixo



Propósito Geral

- Índice
 - extremamente volumoso
 - *Buffer-pool* pequeno
 - apenas uma parcela do índice pode ser carregada em memória principal
 - operações baseadas em disco
-

Características

- Balanceada
 - *Bottom-up* para a criação (em disco)
 - nós folhas → nó raiz
 - as chaves na raiz da árvore emergem naturalmente
 - boas chaves separadoras
-

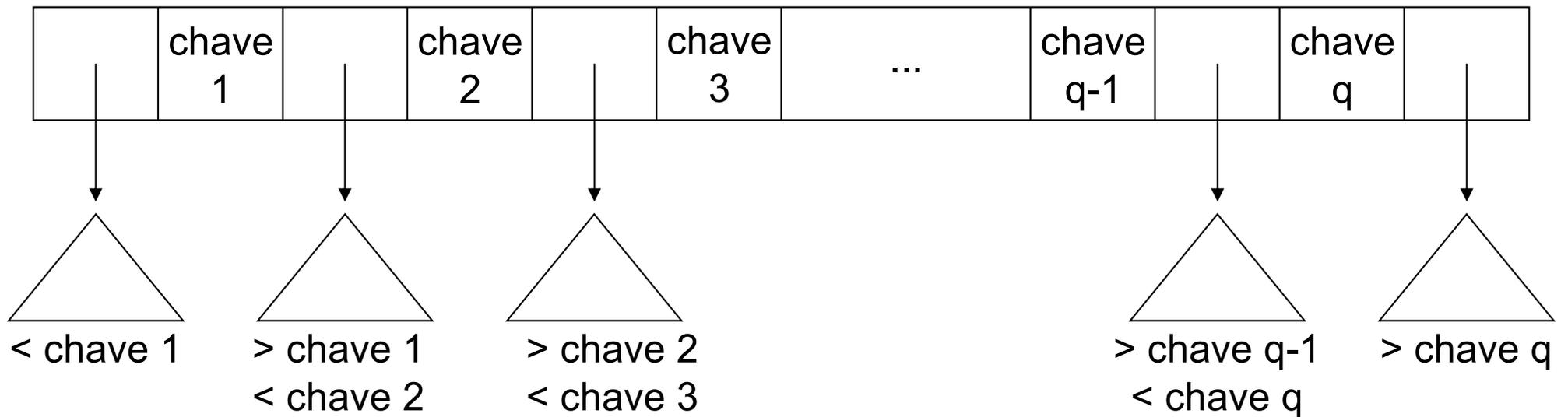
Características

- Nó (= página de disco)
 - sequência ordenada de chaves
 - conjunto de ponteiros
 - número de ponteiros = número de chaves + 1
 - não há uma árvore explícita dentro de uma página (ou nó da árvore)
-

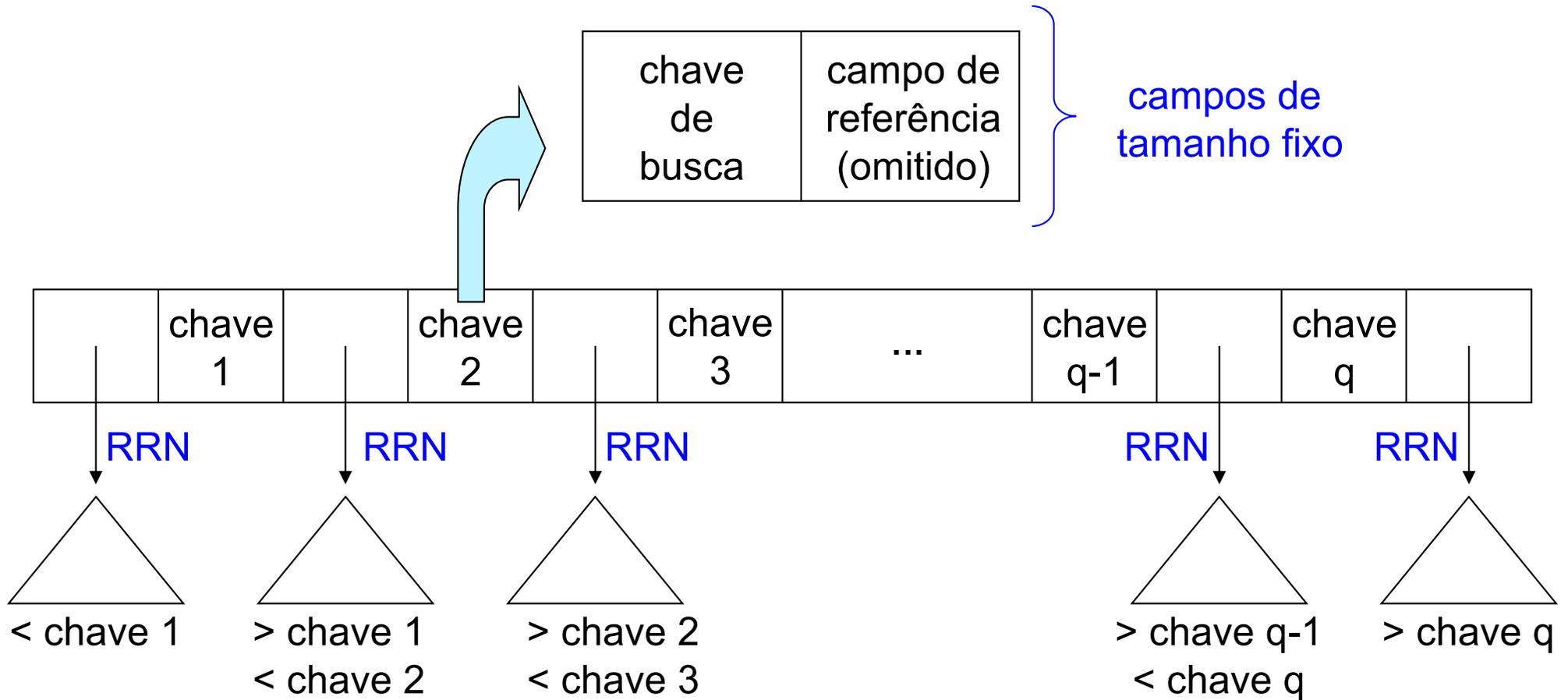
Características

- Ordem (Knuth 1973)
 - número máximo de ponteiros que pode ser armazenado em um nó
 - exemplo: árvore-B de ordem 8
 - máximo de 7 chaves e 8 ponteiros
 - Observações
 - número máximo de ponteiros é igual ao número máximo de descendentes de um nó
 - nós folhas (nível mais baixo das chaves) não possuem filhos, e seus ponteiros são nulos
-

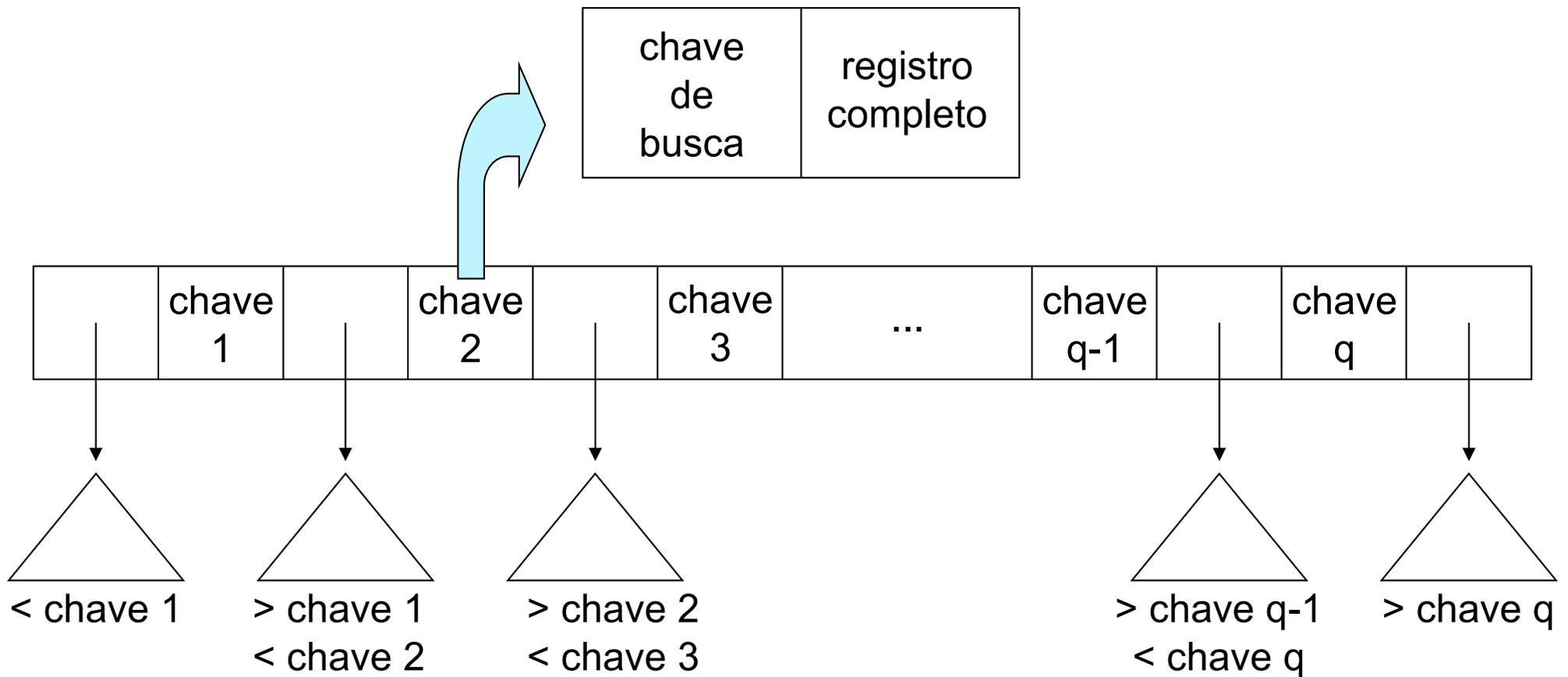
Estrutura Lógica de um Nó



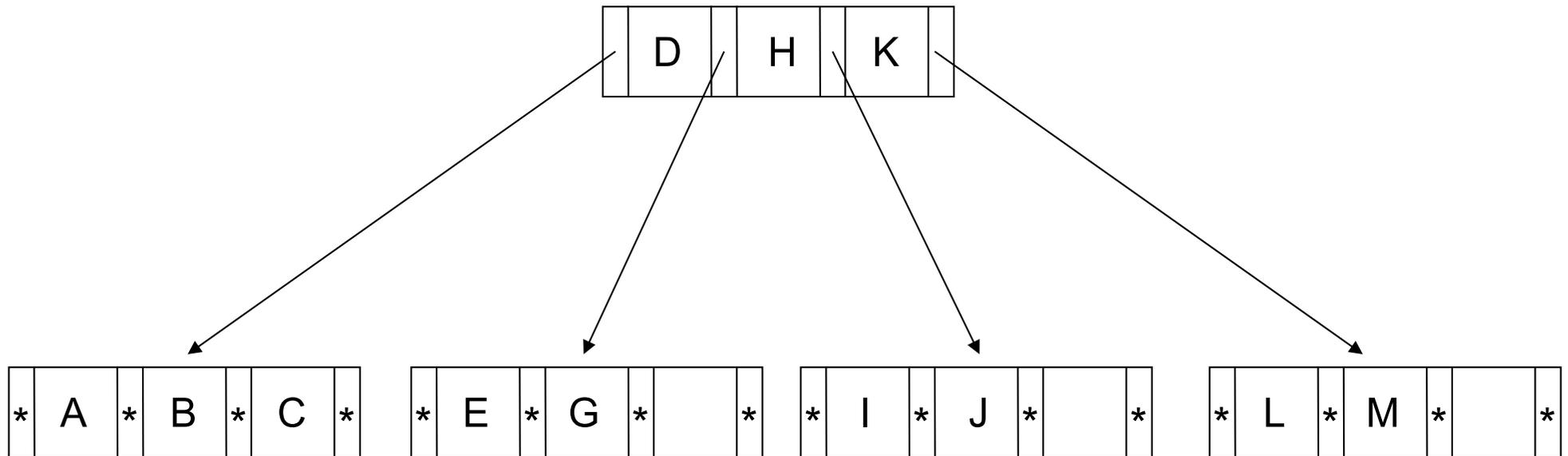
Estrutura Lógica de um Nó



Estrutura Lógica de um Nó



Exemplo



Definição Formal

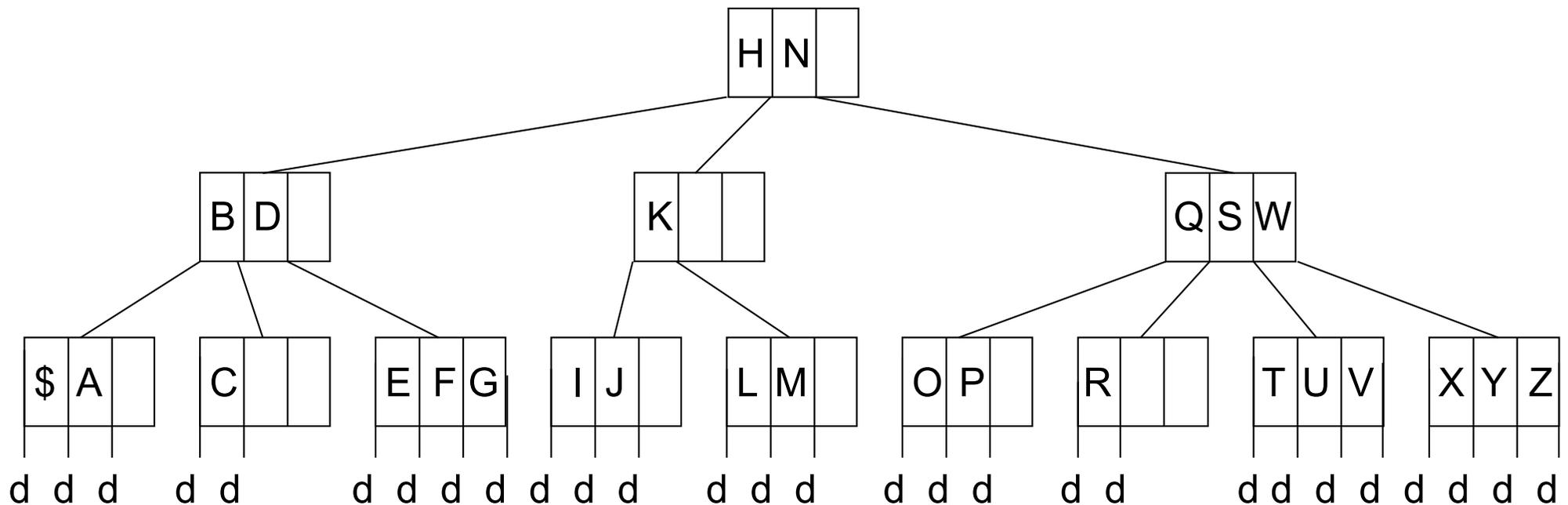
- **Árvore-B com ordem m**
 - cada página possui um máximo de m descendentes
 - cada página, exceto a raiz e as folhas, possui no mínimo $\lceil m/2 \rceil$ descendentes → taxa de ocupação
 - a raiz possui pelo menos 2 descendentes, a menos que seja um nó folha
 - todas as folhas aparecem no mesmo nível
 - uma página interna com k descendentes contém $k-1$ chaves
 - uma folha possui no mínimo $\lceil m/2 \rceil - 1$ chaves e no máximo $m - 1$ chaves → taxa de ocupação
-

Complexidade (Pior Caso)

- Profundidade do caminho de busca
 - número máximo de acessos a disco
 - Exemplo
 - tamanho da página de disco
 - ex: árvore-B de ordem 512 → 511 chaves/página
 - número de chaves
 - ex: 1.000.000 de chaves
 - ⇒ número de níveis que pode ser atingido?
-

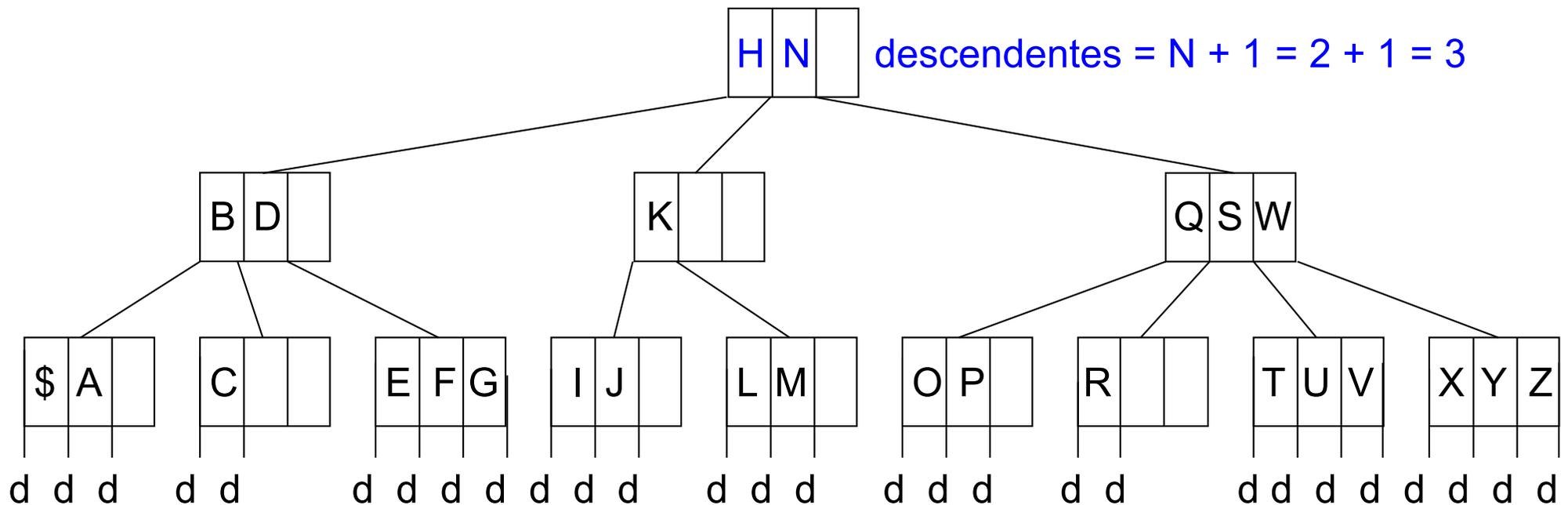
Característica

número de descendentes de um nível da árvore-B = número de chaves contidas no nível em questão e em todos os níveis acima (N) + 1



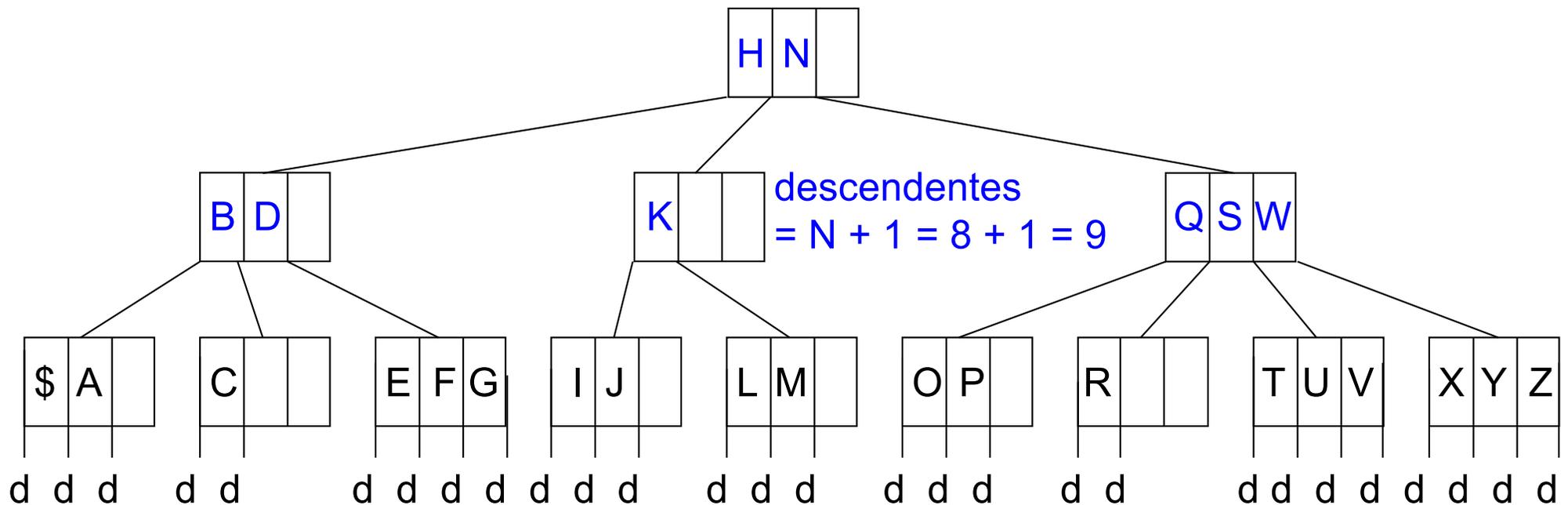
Característica

número de descendentes de um nível da árvore-B = número de chaves contidas no nível em questão e em todos os níveis acima (N) + 1



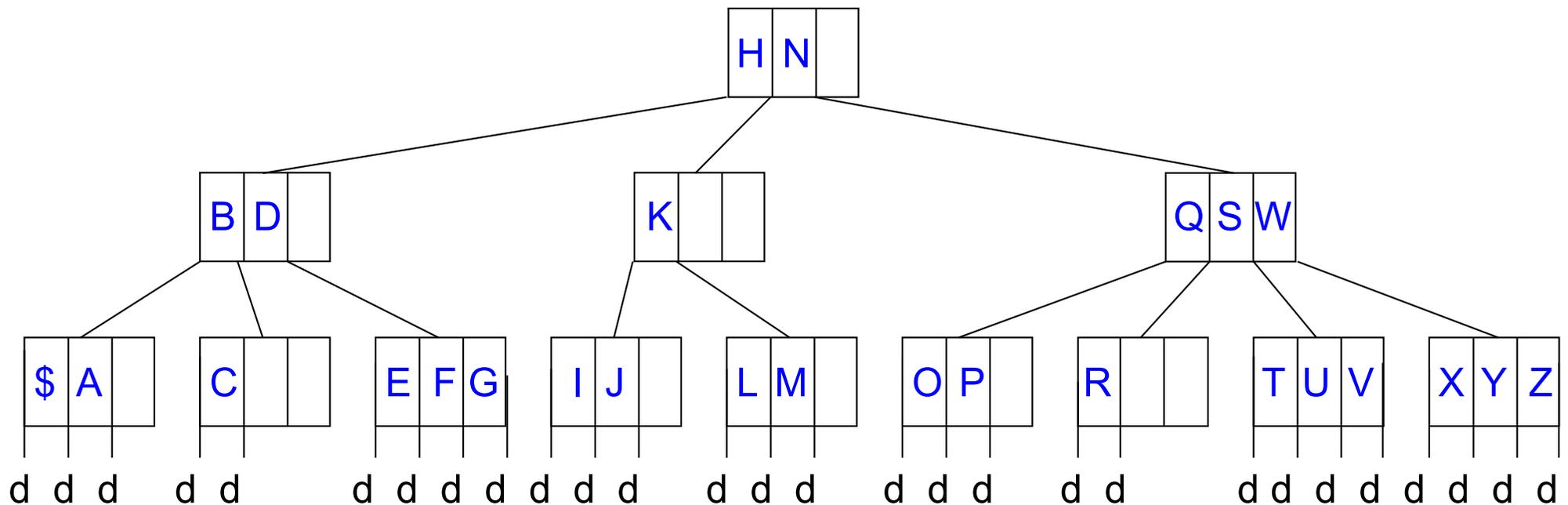
Característica

número de descendentes de um nível da árvore-B = número de chaves contidas no nível em questão e em todos os níveis acima (N) + 1



Característica

número de descendentes de um nível da árvore-B = número de chaves contidas no nível em questão e em todos os níveis acima (N) + 1



$$\text{descendentes} = N + 1 = 27 + 1 = 28$$

Característica

- Propriedades da árvore-B de ordem m
 - cálculo do número mínimo de descendentes de um nível (**análise do pior caso**)

nível	número mínimo de descendentes
1	2
2	$2 \times \lceil m/2 \rceil$
3	$2 \times \lceil m/2 \rceil \times \lceil m/2 \rceil = 2 \times \lceil m/2 \rceil^2$
4	$2 \times \lceil m/2 \rceil \times \lceil m/2 \rceil \times \lceil m/2 \rceil = 2 \times \lceil m/2 \rceil^3$
...	...
d	$2 \times \lceil m/2 \rceil^{d-1}$

nível 1 (raiz):
tem no mínimo
2
descendentes

Observação 2

- Propriedades da árvore-B de ordem m
 - cálculo do número mínimo de descendentes de um nível (*análise do pior caso*)

nível	número mínimo de descendentes
1	2
2	$2 \times \lceil m/2 \rceil$ ←
3	$2 \times \lceil m/2 \rceil \times \lceil m/2 \rceil = 2 \times \lceil m/2 \rceil^2$
4	$2 \times \lceil m/2 \rceil \times \lceil m/2 \rceil \times \lceil m/2 \rceil = 2 \times \lceil m/2 \rceil^3$
...	...
d	$2 \times \lceil m/2 \rceil^{d-1}$

nível 2: tem 2 páginas a partir da raiz, sendo cada uma com no mínimo $\lceil m/2 \rceil$ descendentes

Observação 2

- Propriedades da árvore-B de ordem m
 - cálculo do número mínimo de descendentes de um nível (**análise do pior caso**)

nível	número mínimo de descendentes
1	2
2	$2 \times \lceil m/2 \rceil$
3	$2 \times \lceil m/2 \rceil \times \lceil m/2 \rceil = 2 \times \lceil m/2 \rceil^2$
4	$2 \times \lceil m/2 \rceil \times \lceil m/2 \rceil \times \lceil m/2 \rceil = 2 \times \lceil m/2 \rceil^3$
...	...
d	$2 \times \lceil m/2 \rceil^{d-1}$

nível 3: cada uma das 2 páginas a partir da raiz tem $\lceil m/2 \rceil$ páginas, sendo cada uma com no mínimo $\lceil m/2 \rceil$ descendentes

Observação 2

- Propriedades da árvore-B de ordem m
 - cálculo do número mínimo de descendentes de um nível (**análise do pior caso**)

nível	número mínimo de descendentes
1	2
2	$2 \times \lceil m/2 \rceil$
3	$2 \times \lceil m/2 \rceil \times \lceil m/2 \rceil = 2 \times \lceil m/2 \rceil^2$
4	$2 \times \lceil m/2 \rceil \times \lceil m/2 \rceil \times \lceil m/2 \rceil = 2 \times \lceil m/2 \rceil^3$
...	...
d	$2 \times \lceil m/2 \rceil^{d-1}$

para qualquer nível d , com exceção da raiz (nível 1)



Complexidade (Pior Caso)

- Número de chaves (N)
 - $N + 1$ descendentes no nível das folhas
 - Profundidade da árvore-B no nível das folhas
 - d
 - Relacionamento
 - $N + 1$ descendentes e
 - número mínimo de descendentes da árvore-B com profundidade d
-

Complexidade (Pior Caso)

$$N + 1 \geq 2 \times \lceil m/2 \rceil^{d-1}$$

$$d \leq 1 + \log_{\lceil m/2 \rceil} \left((N + 1)/2 \right)$$

- Exemplo

- $m = 512$

- $N = 1.000.000$

- $d \leq 1 + \log_{256} (500.000,50) \Rightarrow d \leq 3,37$

- acesso a disco adicional: arquivo de dados

a árvore possui não mais do que 3 níveis de altura

