

# Armazenamento e Indexação

## Banco de Dados: Teoria e Prática

André Santanchè  
Instituto de Computação – UNICAMP  
Setembro de 2019

# Recomendações de Leitura

- (Silberschatz, 2006, cap. 11)
- (Ramakrishnan, 2003, cap. 8)
- (Elmasri, 2011, cap. 11 e 12)

# Onde Armazenamos Dados?

# Onde Armazenamos Dados?

- Memória RAM

- Disco

  - HD

  - CD / DVD

- Fita magnética

- Solid State Drive (SSD)

  - usa circuitos integrados como a memória – sem partes mecânicas

  - retém os dados sem a necessidade de energia

  - Interface equivalente a de um disco

# Questão 1

- Para cada item abaixo, liste suas vantagens e desvantagens como opção de tecnologia para armazenamento de dados num SGBD. Dê exemplos de dados que se adequariam à tecnologia.
  - a) Memória RAM
  - b) Disco Magnético
  - c) Fita Magnética

# Questão 1

## Resposta

- a) Memória RAM: rápida/cara. Pequena quantidade de dados, índices, dados temporários etc.
- b) Disco Magnético: relativamente barato/relativamente lento. Grande quantidade de dados, dados institucionais, logs, etc.
- c) Fita Magnética: baixo custo/lenta. Dados de backup, dados históricos, logs, etc.

# Hierarquia de Armazenamento

## ■ Armazenamento Primário

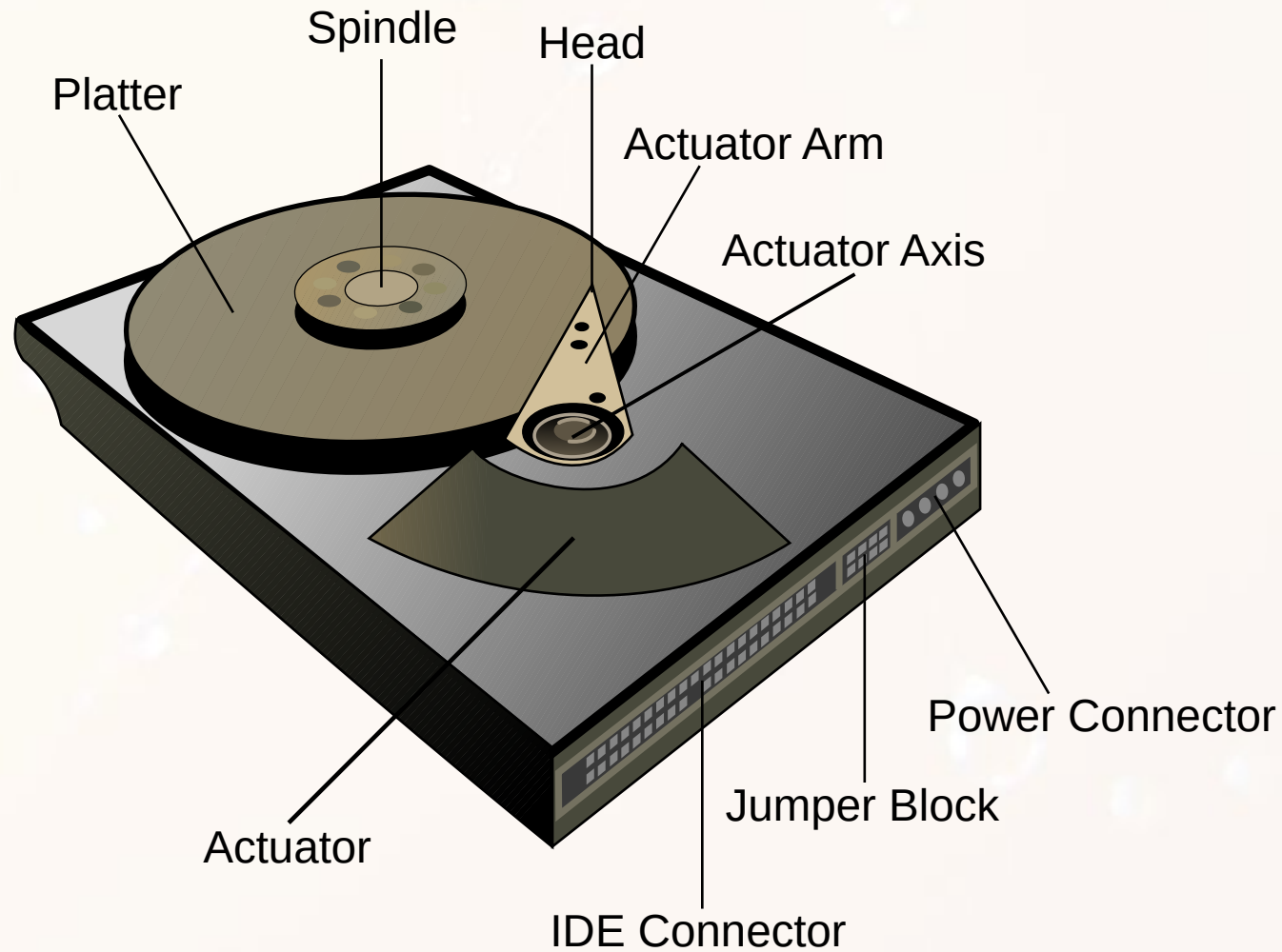
- Operado diretamente pela CPU
- Exemplos: memória RAM, cache

## ■ Armazenamento Secundário

- Usualmente mais barato e mais lento
- Não operado diretamente pela CPU
  - Exigem intermediação de armazenamento primário
- Exemplos: disco, fita magnética

(Elmasri, 2011)

# Estrutura do Disco

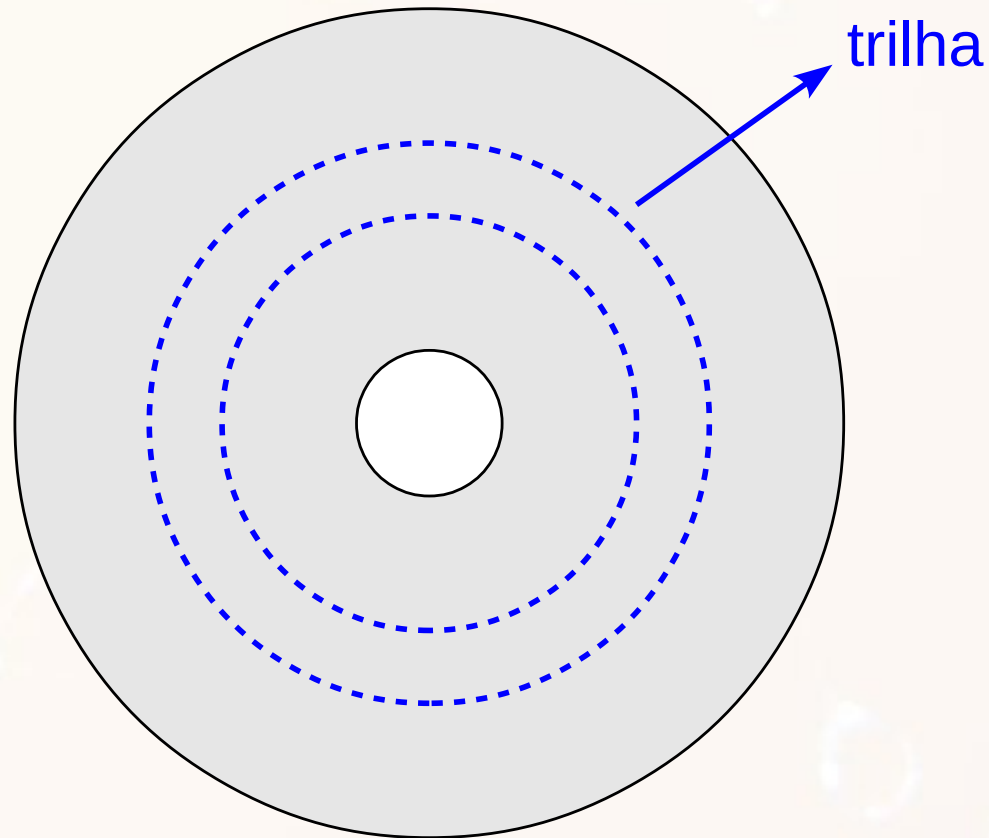


By Surachit [[http://en.wikipedia.org/w/index.php?title=File:Hard\\_drive-en.svg](http://en.wikipedia.org/w/index.php?title=File:Hard_drive-en.svg)]



# Estrutura do Disco

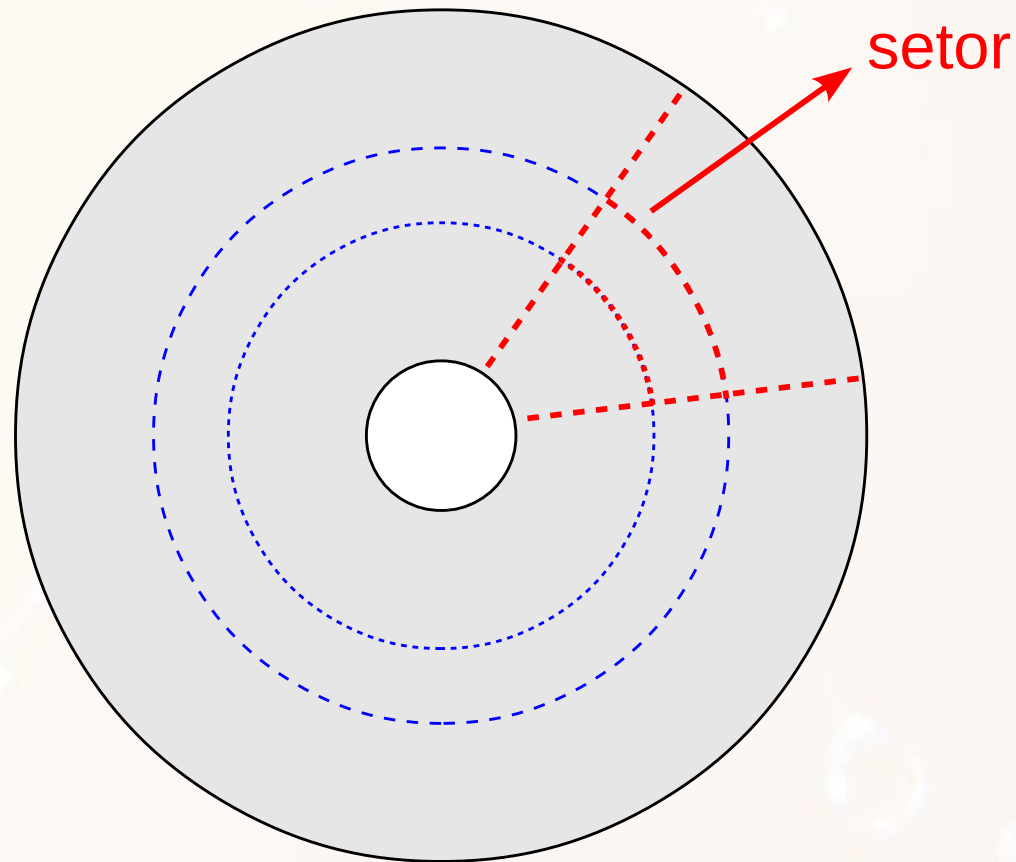
## Trilha



- Círculos magnéticos sobre a superfície
- Local onde são armazenados os dados

# Estrutura do Disco

## Setor



- Unidades de divisão da trilha
- Menor unidade de leitura/gravação

# Bloco de Disco ou Página

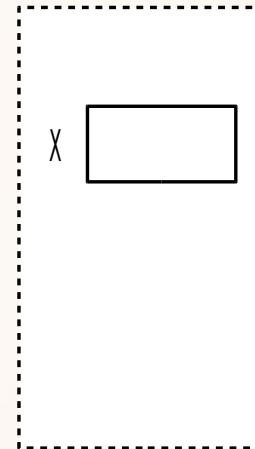
- Organização feita pelo SO sobre o disco
- Unidade de trabalho para o SO

**Armazenamento Secundário  
intermediado pelo Primário**

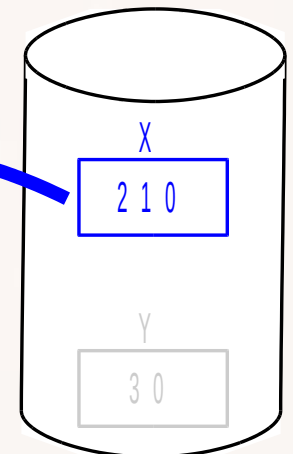
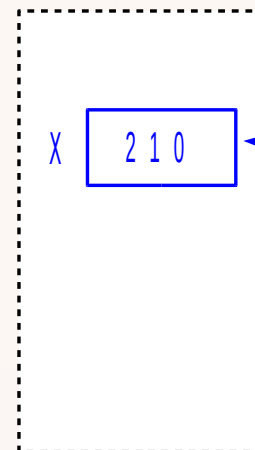
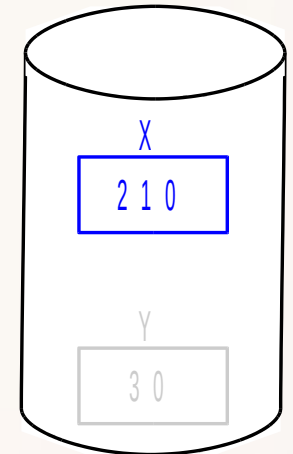
# Operação de Leitura Como Abstraímos

■ ler(X)

M em ória P rincipal



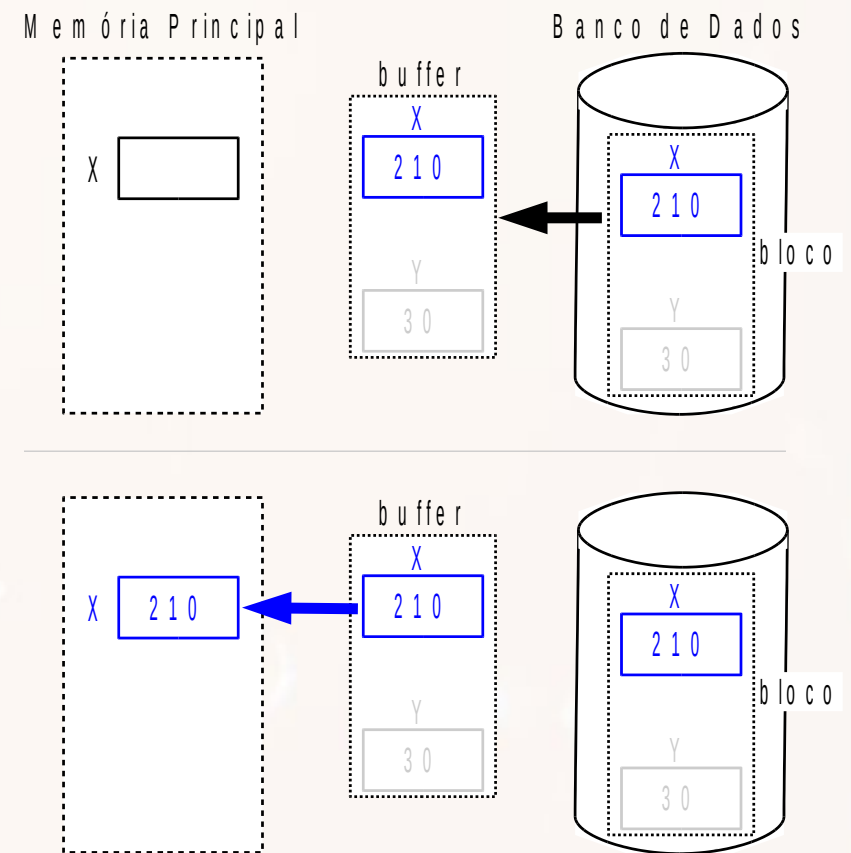
Banco de Dados



# Operação de Leitura Como Acontece

## ■ ler(X)

- encontra bloco X no disco
- copia bloco para buffer da memória principal (se ainda não estiver lá)
- copia o item X do buffer para a variável X da memória principal

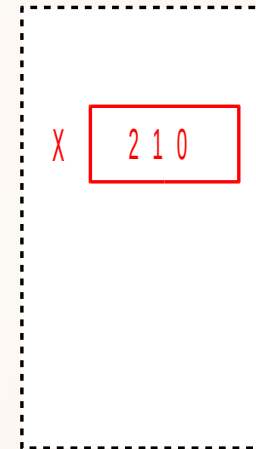


(Elmasri, 2010)

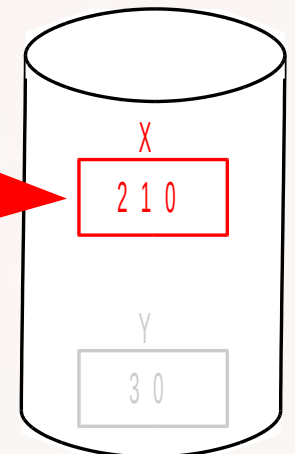
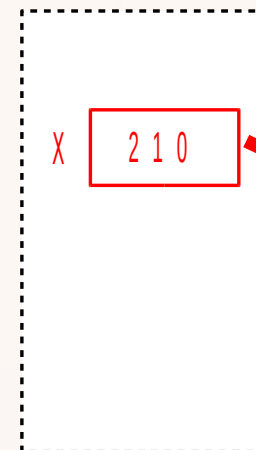
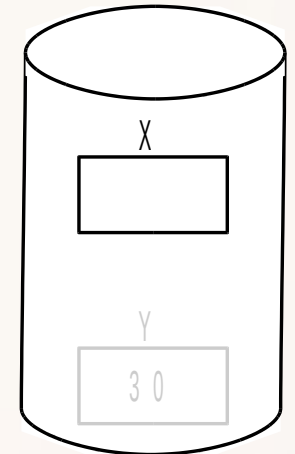
# Operação de Gravação Como Abstraímos

■ gravar(X)

M e m ó r i a P r i n c i p a l



B a n c o d e D a d o s

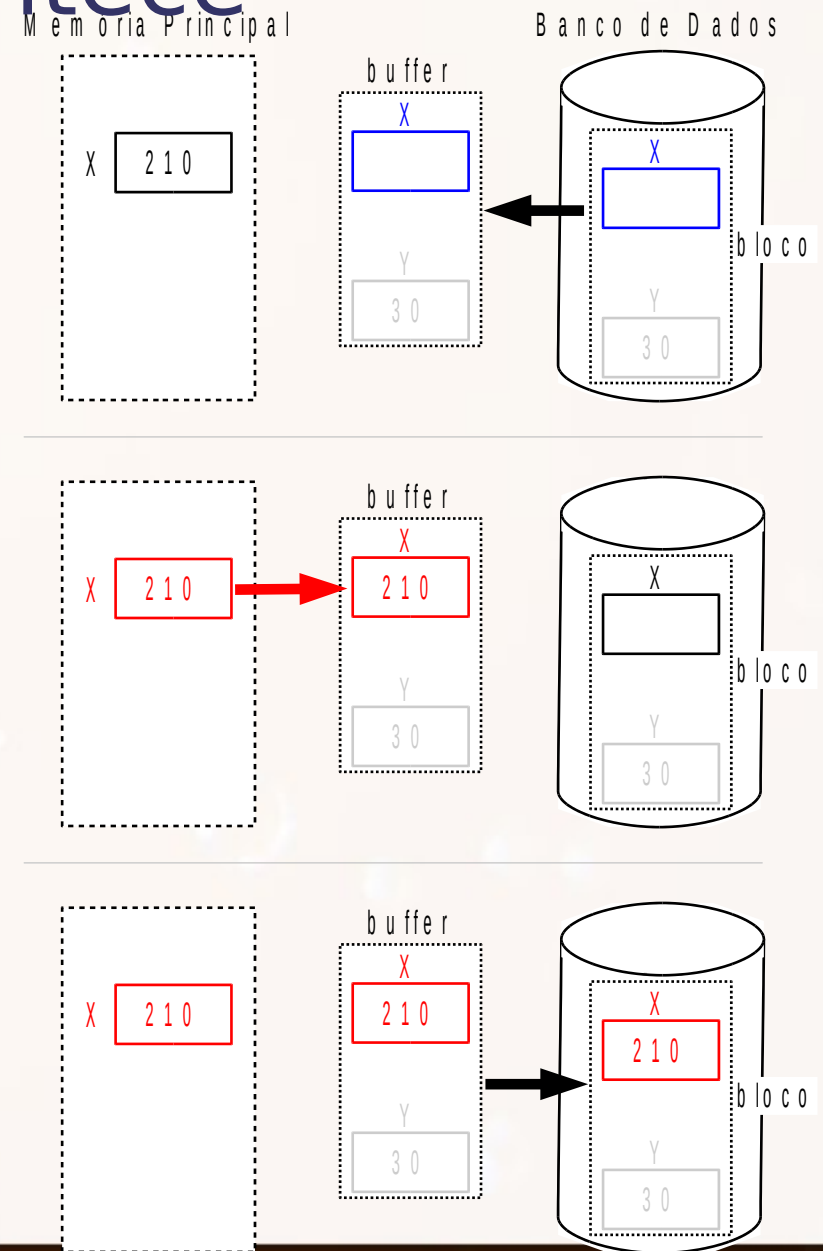


# Operação de Gravação

## Como Acontece

### ■ gravar(X)

- encontra bloco X no disco
- copia bloco para buffer da memória principal (se ainda não estiver lá)
- copia variável X da memória principal para o buffer
- atualiza o buffer no disco











# Arquivos e Registros

- Um arquivo de registros - abstração para SGBD

(Ramakrishnan, 2003)

- **Arquivo:** abstração criada pelo SO para os blocos de disco
- **Registro:** abstração de subdivisão do arquivo criada pela aplicação ou SGBD para o arquivo

# Ordem?

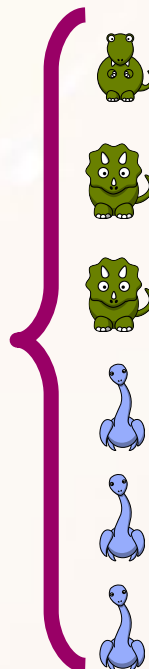
Id	Is a	Origin Place
 <b>FMNH PR2081</b>	Tyrannosaurus rex	Hell Creek
 <b>MNHN 1912.20</b>	Triceratops calicornis	Lance Creek
 <b>MNHN 1912.20b</b>	Triceratops horridus	Lance Creek
 <b>MNHN A. C. 8592</b>	Plesiosaurus dolichodeirus	Lyme Regis
 <b>SIPB R 90</b>	Plesiosaurus dolichodeirus	Lyme Regis
 <b>STC223</b>	Plesiosaurus gurgitis	St. Croix

?

}

# Ordem?

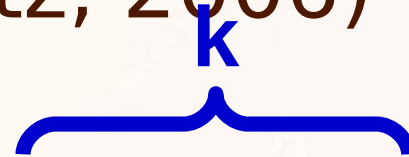
ordem?



Id	Is a	Origin Place
FMNH PR2081	Tyrannosaurus rex	Hell Creek
MNHN 1912.20	Triceratops calicornis	Lance Creek
MNHN 1912.20b	Triceratops horridus	Lance Creek
MNHN A. C. 8592	Plesiosaurus dolichodeirus	Lyme Regis
SIPB R 90	Plesiosaurus dolichodeirus	Lyme Regis
STC223	Plesiosaurus gurgitis	St. Croix

# Sequencial

- Gravação em ordem sequencial de acordo com a chave de busca (Silberschatz, 2006)



Id	Is a	Origin Place
FMNH PR2081	Tyrannosaurus rex	Hell Creek
MNHN 1912.20	Triceratops calicornis	Lance Creek
MNHN 1912.20b	Triceratops horridus	Lance Creek
MNHN A. C. 8592	Plesiosaurus dolichodeirus	Lyme Regis
SIPB R 90	Plesiosaurus dolichodeirus	Lyme Regis
STC223	Plesiosaurus gurgitis	St. Croix

sequencial



# Ordem?

ordem?



Id	Is a	Origin Place
MNHN A. C. 8592	Plesiosaurus dolichodeirus	Lyme Regis
SIPB R 90	Plesiosaurus dolichodeirus	Lyme Regis
STC223	Plesiosaurus gurgitis	St. Croix
MNHN 1912.20	Triceratops calicornis	Lance Creek
MNHN 1912.20b	Triceratops horridus	Lance Creek
FMNH PR2081	Tyrannosaurus rex	Hell Creek

# Sequencial

- Gravação em ordem sequencial de acordo com a chave de busca (Silberschatz, 2006)

**sequencial** {

**k**

Id	Is a	Origin Place
MNHN A. C. 8592	Plesiosaurus dolichodeirus	Lyme Regis
SIPB R 90	Plesiosaurus dolichodeirus	Lyme Regis
STC223	Plesiosaurus gurgitis	St. Croix
MNHN 1912.20	Triceratops calicornis	Lance Creek
MNHN 1912.20b	Triceratops horridus	Lance Creek
FMNH PR2081	Tyrannosaurus rex	Hell Creek

# Ordem?

ordem?

Id	Is a	Origin Place
STC223	Plesiosaurus gurgitis	St. Croix
MNHN 1912.20b	Triceratops horridus	Lance Creek
SIPB R 90	Plesiosaurus dolichodeirus	Lyme Regis
FMNH PR2081	Tyrannosaurus rex	Hell Creek
MNHN 1912.20	Triceratops calicornis	Lance Creek
MNHN A. C. 8592	Plesiosaurus dolichodeirus	Lyme Regis



# Heap

- Sem ordenação; gravação em qualquer posição (Silberschatz, 2006)

heap

Id	Is a	Origin Place
STC223	Plesiosaurus gurgitis	St. Croix
MNHN 1912.20b	Triceratops horridus	Lance Creek
SIPB R 90	Plesiosaurus dolichodeirus	Lyme Regis
FMNH PR2081	Tyrannosaurus rex	Hell Creek
MNHN 1912.20	Triceratops calicornis	Lance Creek
MNHN A. C. 8592	Plesiosaurus dolichodeirus	Lyme Regis



# Ordem?







ordem?



Id	Is a	Origin Place
FMNH PR2081	Tyrannosaurus rex	Hell Creek
MNHN 1912.20b	Triceratops horridus	Lance Creek
MNHN 1912.20	Triceratops calicornis	Lance Creek
MNHN A. C. 8592	Plesiosaurus dolichodeirus	Lyme Regis
STC223	Plesiosaurus gurgitis	St. Croix
SIPB R 90	Plesiosaurus dolichodeirus	Lyme Regis

# Hash


- Cálculo de função de hash sobre atributo para definir posição (Silberschatz, 2006)







Id	Is a	Origin Place
 <b>FMNH PR2081</b>	Tyrannosaurus rex	Hell Creek
 <b>MNHN 1912.20b</b>	Triceratops horridus	Lance Creek
 <b>MNHN 1912.20</b>	Triceratops calicornis	Lance Creek
 <b>MNHN A. C. 8592</b>	Plesiosaurus dolichodeirus	Lyme Regis
 <b>STC223</b>	Plesiosaurus gurgitis	St. Croix
 <b>SIPB R 90</b>	Plesiosaurus dolichodeirus	Lyme Regis

# Exercício 1







- Dado o arquivo a seguir, proponha uma função hash para posicionar os registros conforme  $k$ .

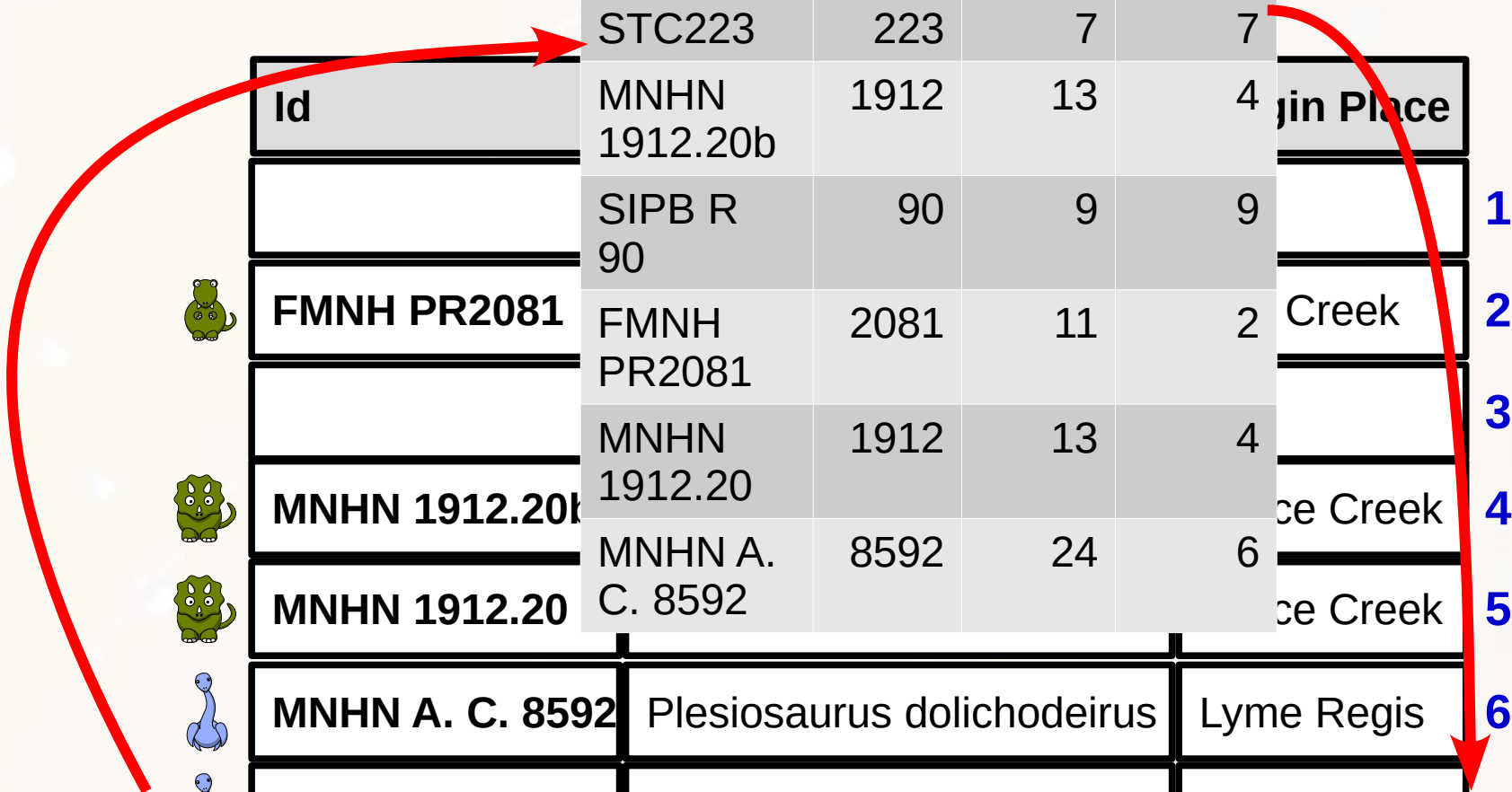
$k$



Id	Is a	Origin Place
 STC223	Plesiosaurus gurgitis	St. Croix
 MNHN 1912.20b	Triceratops horridus	Lance Creek
 SIPB R 90	Plesiosaurus dolichodeirus	Lyme Regis
 FMNH PR2081	Tyrannosaurus rex	Hell Creek
 MNHN 1912.20	Triceratops calicornis	Lance Creek
 MNHN A. C. 8592	Plesiosaurus dolichodeirus	Lyme Regis

# Exercício 1

	k	parte num.	soma díg.	soma díg. (pos.)		
	STC223	223	7	7		
<b>Id</b>	MNHN 1912.20b	1912	13	4	in Place	
	SIPB R 90	90	9	9		<b>1</b>
 <b>FMNH PR2081</b>	FMNH PR2081	2081	11	2	Creek	<b>2</b>
	MNHN 1912.20	1912	13	4		<b>3</b>
 <b>MNHN 1912.20</b>	1912.20				ce Creek	<b>4</b>
 <b>MNHN 1912.20</b>	MNHN A. C. 8592	8592	24	6	ce Creek	<b>5</b>
 <b>MNHN A. C. 8592</b>	Plesiosaurus dolichodeirus		Lyme Regis			<b>6</b>
 <b>STC223</b>	Plesiosaurus gurgitis		St. Croix			<b>7</b>
						<b>8</b>
 <b>SIPB R 90</b>	Plesiosaurus dolichodeirus		Lyme Regis			<b>9</b>



# Organização de Arquivos

## ■ Heap

- sem ordenação
- gravação em qualquer posição

## ■ Sequencial

- gravação em ordem sequencial de acordo com a chave de busca

## ■ Hash

- cálculo de função de hash sobre atributo para definir posição

(Silberschatz, 2006)

# Índice

# Processamento de Consulta

- Dada a consulta:  
SELECT nome FROM Pessoa  
WHERE id=146
- Se os dados estão espalhados em disco, precisaríamos acessar todos os blocos do arquivo da tabela Pessoa
- Índices de BDs ajudam neste processo
- Índices de de BDs funcionam como índices de livros, apontando para a localização do conteúdo

# Índice

- Estrutura de dados
- Organiza registros
- Otimiza certas operações de recuperação  
(Ramakrishnan, 2003)



# Entrada de Índice







- Entrada de índice (*data entry*) → registros armazenados em um índice
- Alternativas para a entrada de índice
  - (1)  $k^*$  - registro completo com chave  $k$
  - (2)  $(k, rid) \rightarrow rid = id$  do registro de chave  $k$
  - (3)  $(k, rid-list) \rightarrow rid-list =$  lista de registros de chave  $k$

(Ramakrishnan, 2003)

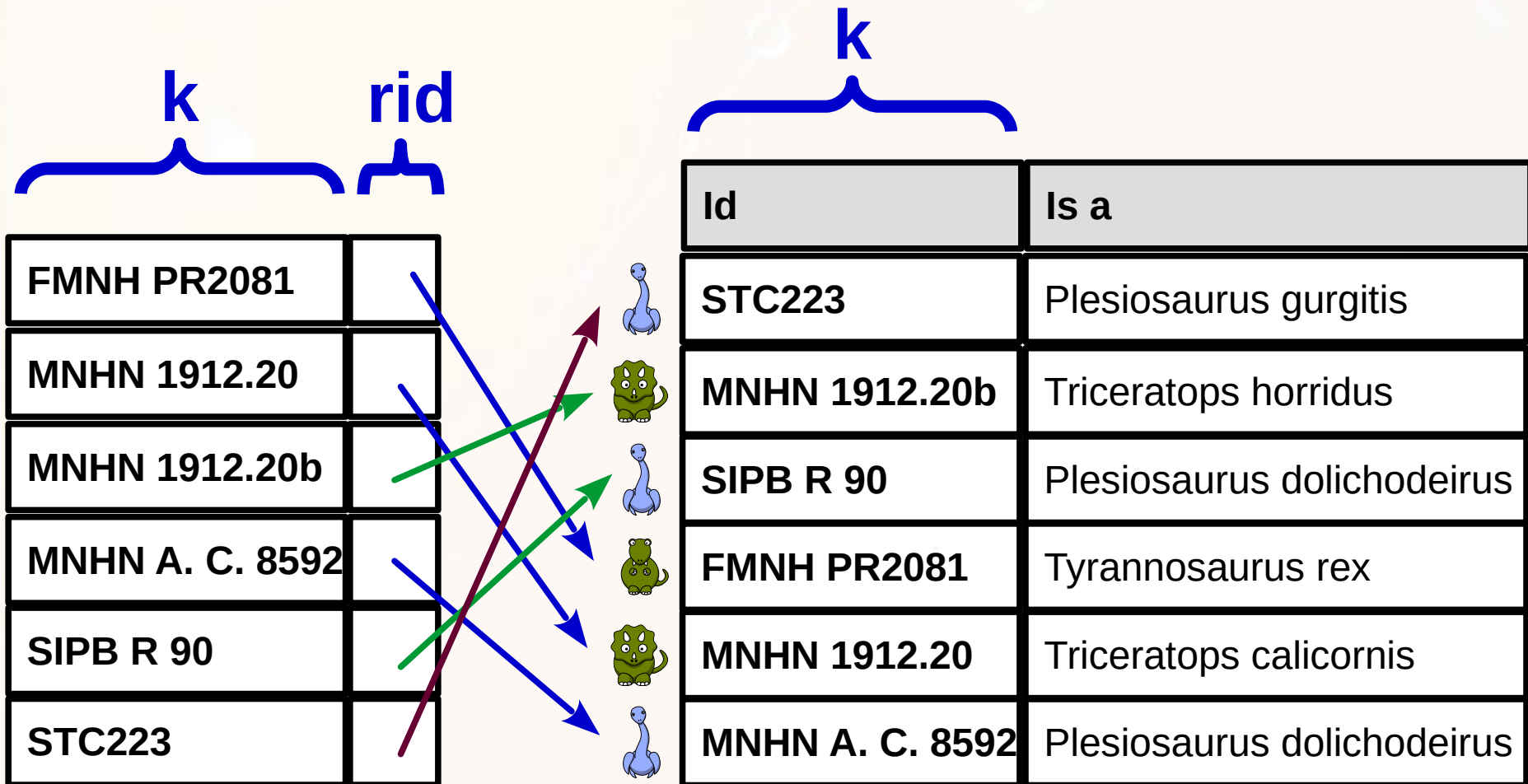
# Entrada de Índice (1)

**k\*: registro (incluindo chave)**

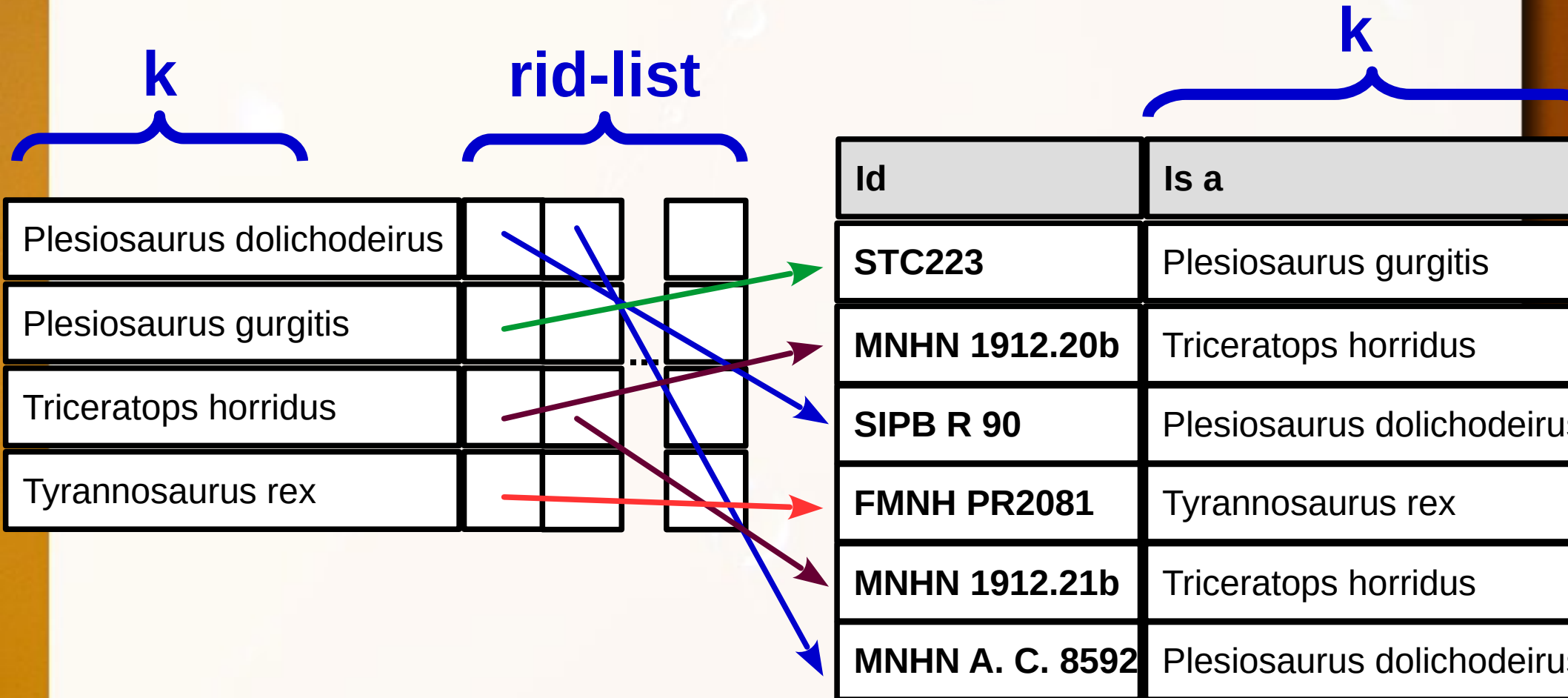
**k: chave**

Id	Is a	Origin Place
 <b>FMNH PR2081</b>	Tyrannosaurus rex	Hell Creek
 <b>MNHN 1912.20</b>	Triceratops calicornis	Lance Creek
 <b>MNHN 1912.20b</b>	Triceratops horridus	Lance Creek
 <b>MNHN A. C. 8592</b>	Plesiosaurus dolichodeirus	Lyme Regis
 <b>SIPB R 90</b>	Plesiosaurus dolichodeirus	Lyme Regis
 <b>STC223</b>	Plesiosaurus gurgitis	St. Croix

# Entrada de Índice (2)



# Entrada Índice (3)



# Entrada de Índice

## ■ Alternativas para a entrada de índice

(1)  $k^*$  - registro completo com chave  $k$

(2)  $(k, \text{rid}) \rightarrow \text{rid} = \text{id do registro de chave } k$

(3)  $(k, \text{rid-list}) \rightarrow \text{rid-list} = \text{lista de registros de chave } k$

(Ramakrishnan, 2003)

## ■ Vantagens das alternativas (2) e (3):

□ mais de um índice para o mesmo arquivo

□ menor: pode-se carregar mais ou inteiro na memória

□ suporta estruturas mais complexas

# Exercício 2

- Nas aulas anteriores, discutimos sobre redundância de informação e seus potenciais problemas. Índices são estruturas que introduzem redundância no banco de dados. Descreva o impacto da introdução deste tipo de redundância em termos de:
  - a) Consistência dos dados
  - b) Velocidade de leitura
  - c) Velocidade de gravação

# Índices Primários e Secundários

- Índice primário ou de agrupamento
  - arquivo ordenado sequencialmente
  - chave de busca define ordem do arquivo
- Índice secundário
  - índice de não agrupamento
  - índice não necessariamente único

# Índice Primário

- Entrada do índice (1):  $k^*$  - registro completo com chave  $k$

$k$

Id	Is a	Origin Place
FMNH PR2081	Tyrannosaurus rex	Hell Creek
MNHN 1912.20	Triceratops calicornis	Lance Creek
MNHN 1912.20b	Triceratops horridus	Lance Creek
MNHN A. C. 8592	Plesiosaurus dolichodeirus	Lyme Regis
SIPB R 90	Plesiosaurus dolichodeirus	Lyme Regis
STC223	Plesiosaurus gurgitis	St. Croix

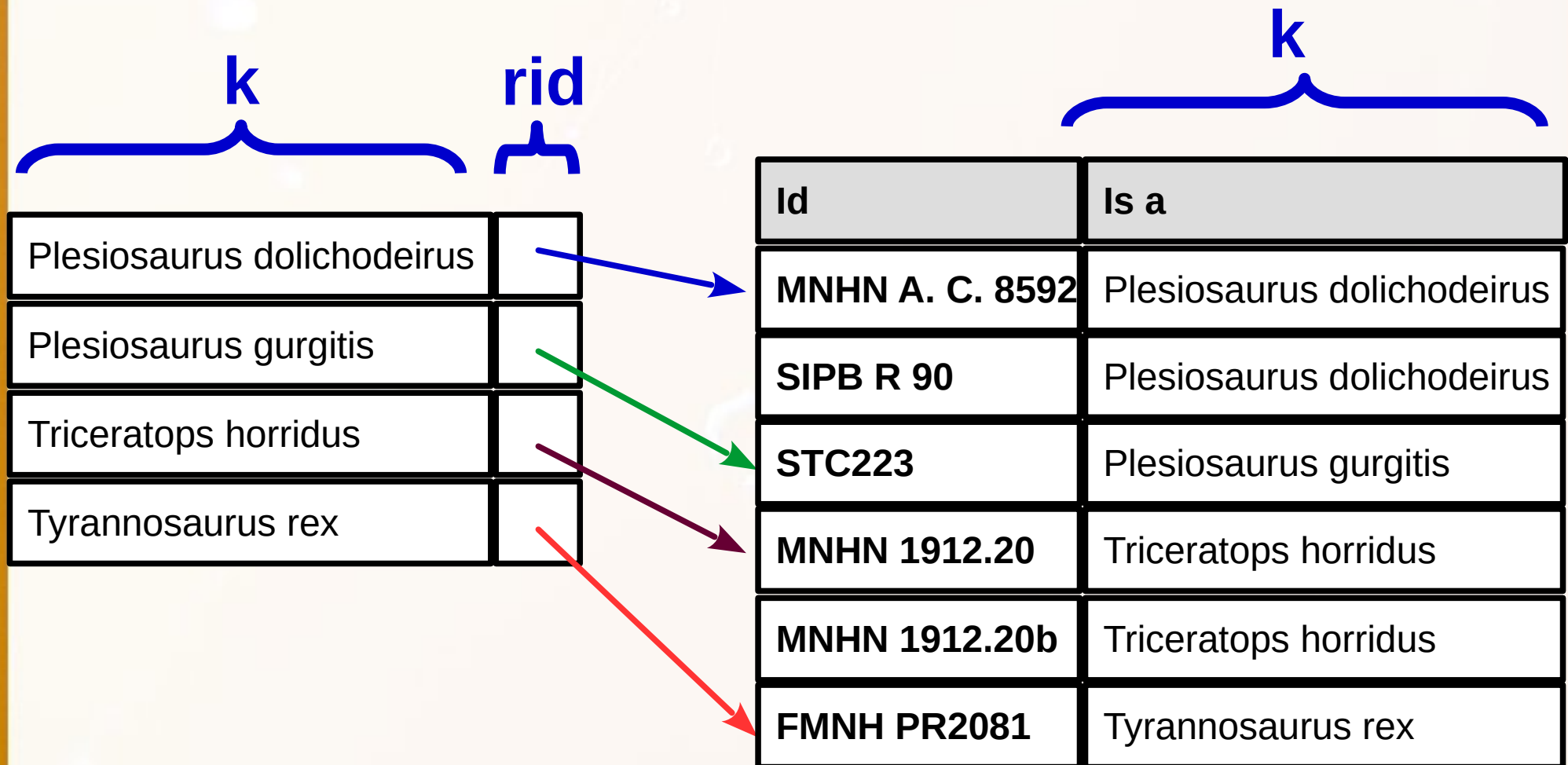
sequencial





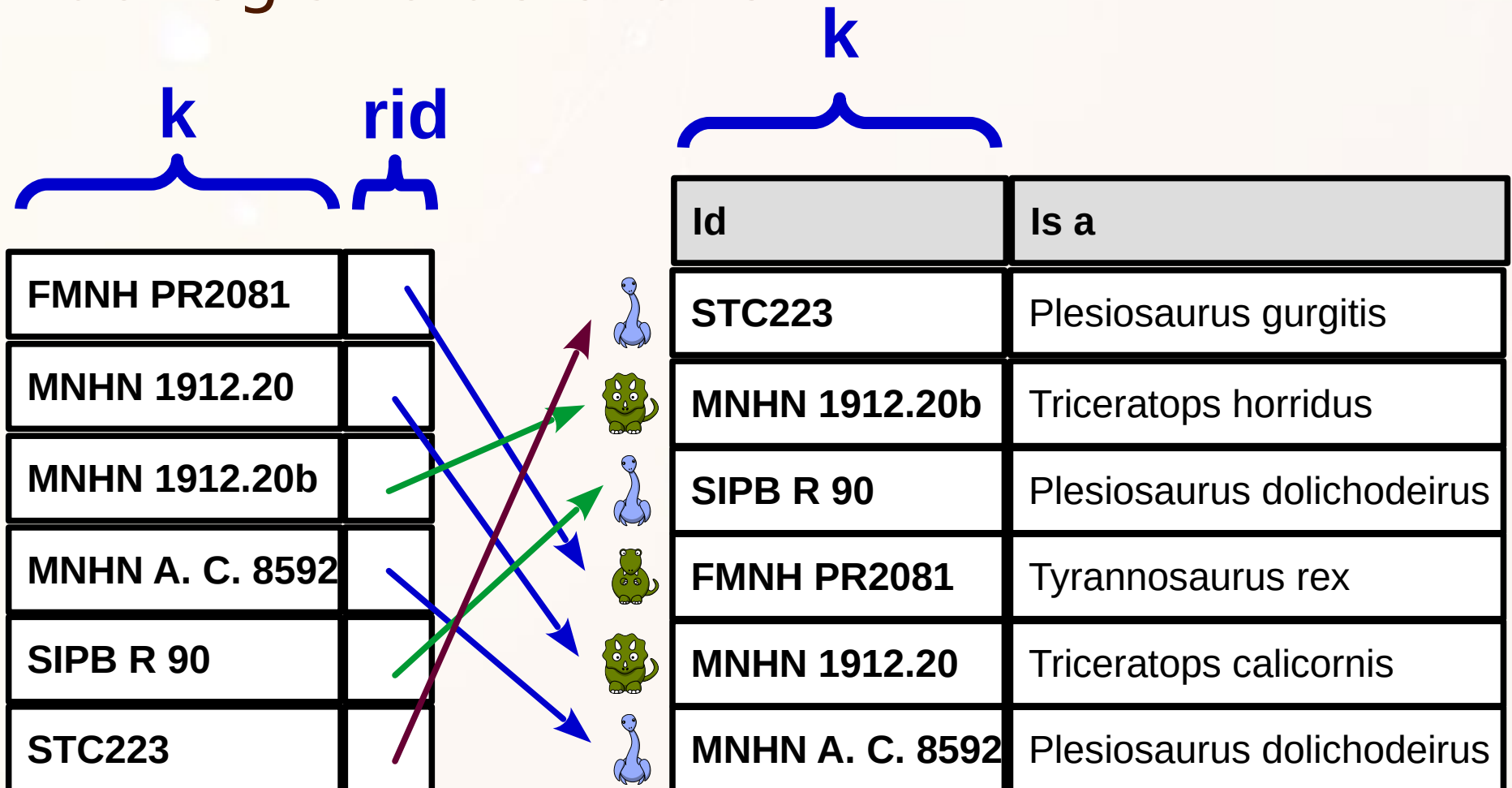
# Índice Primário

- Entrada do índice (2):  $(k, rid) \rightarrow rid = id$  do registro de chave  $k$



# Índice Secundário

- Entrada do índice (2):  $(k, \text{rid}) \rightarrow \text{rid} = \text{id}$  do registro de chave  $k$



# Índices Densos e Esparsos

## ■ Denso

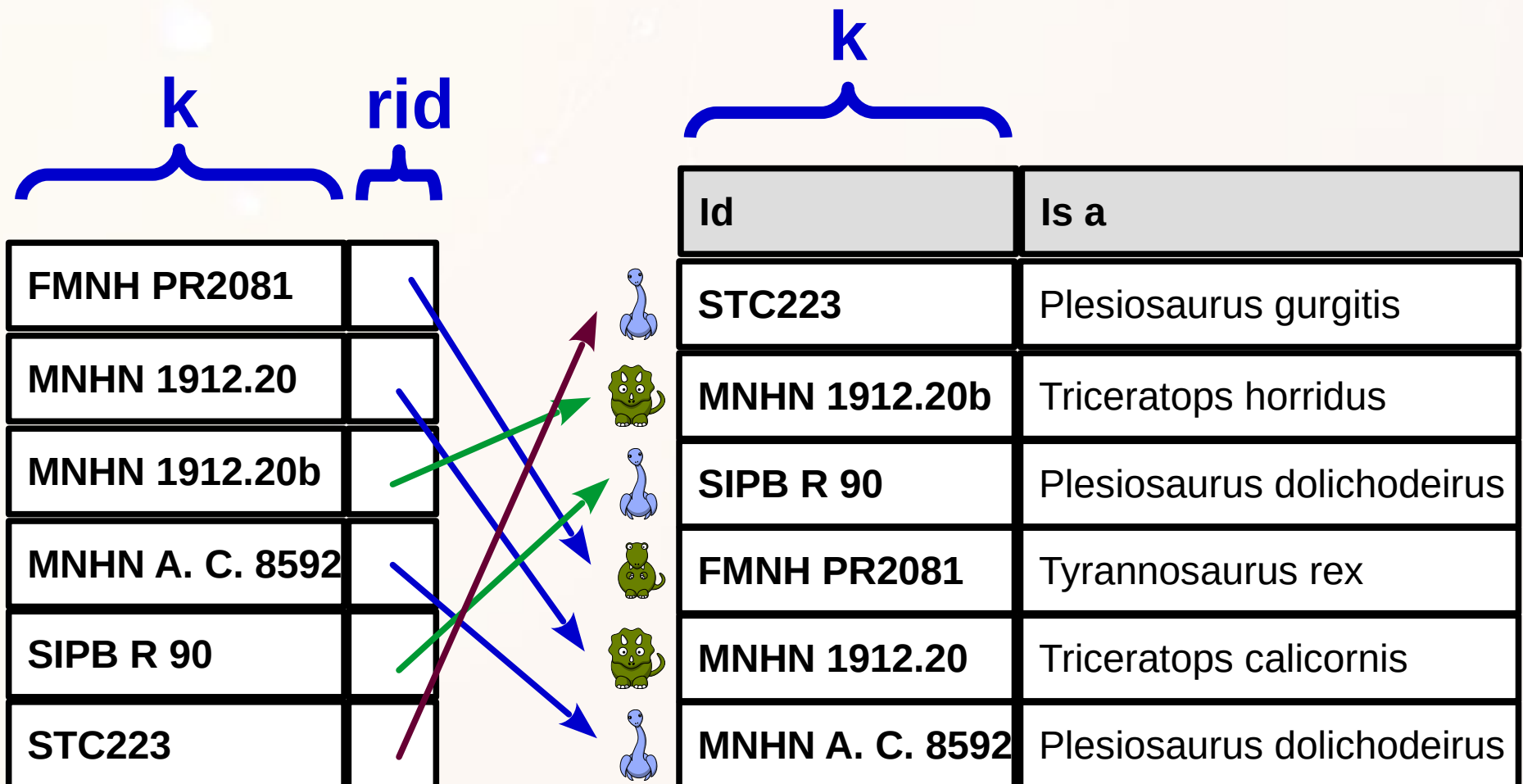
- uma entrada de índice para cada valor de chave

## ■ Esparso

- uma entrada índice para mais de um valor de chave

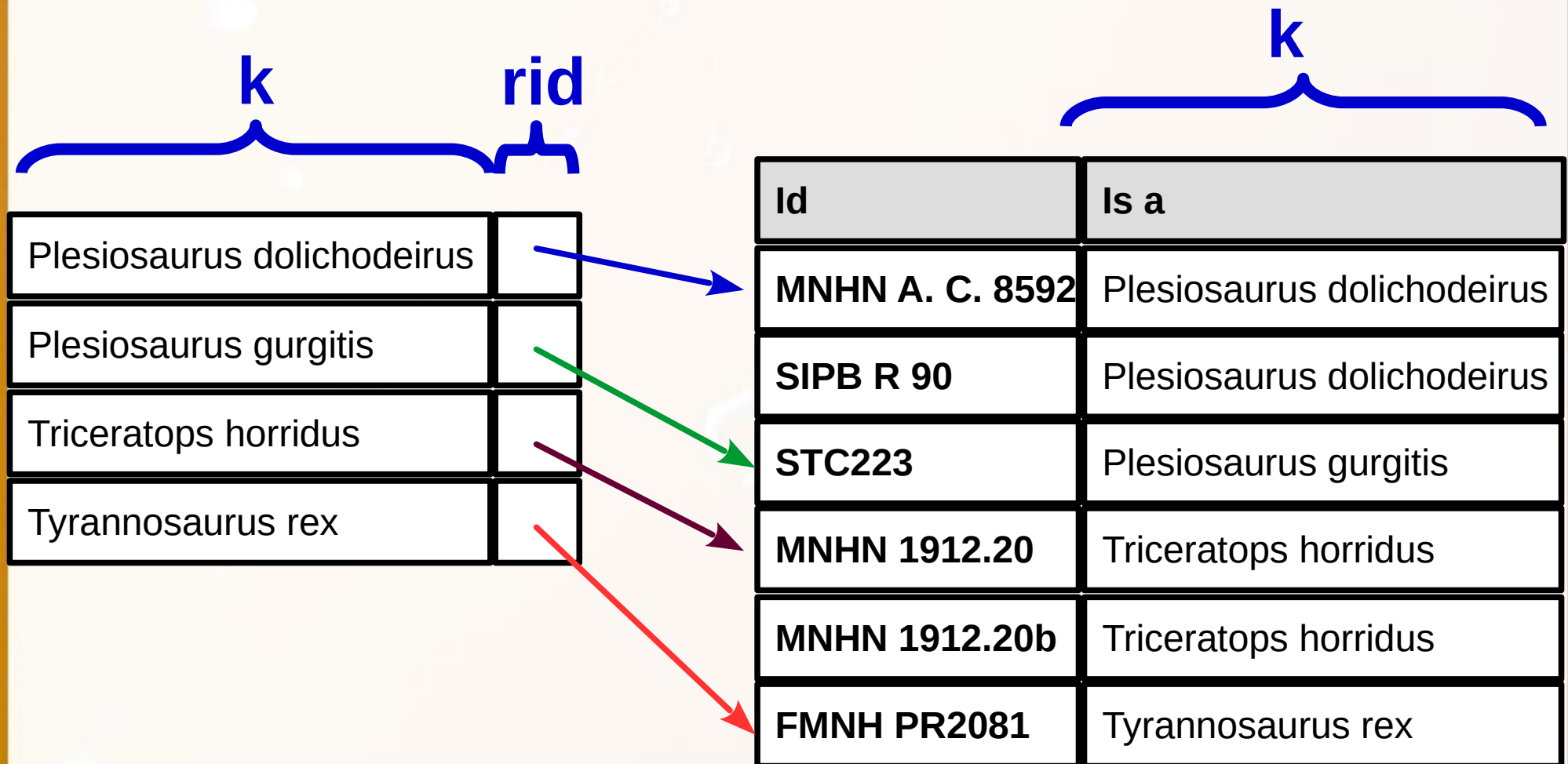
# Índice Denso

- Uma entrada de índice para cada valor de chave



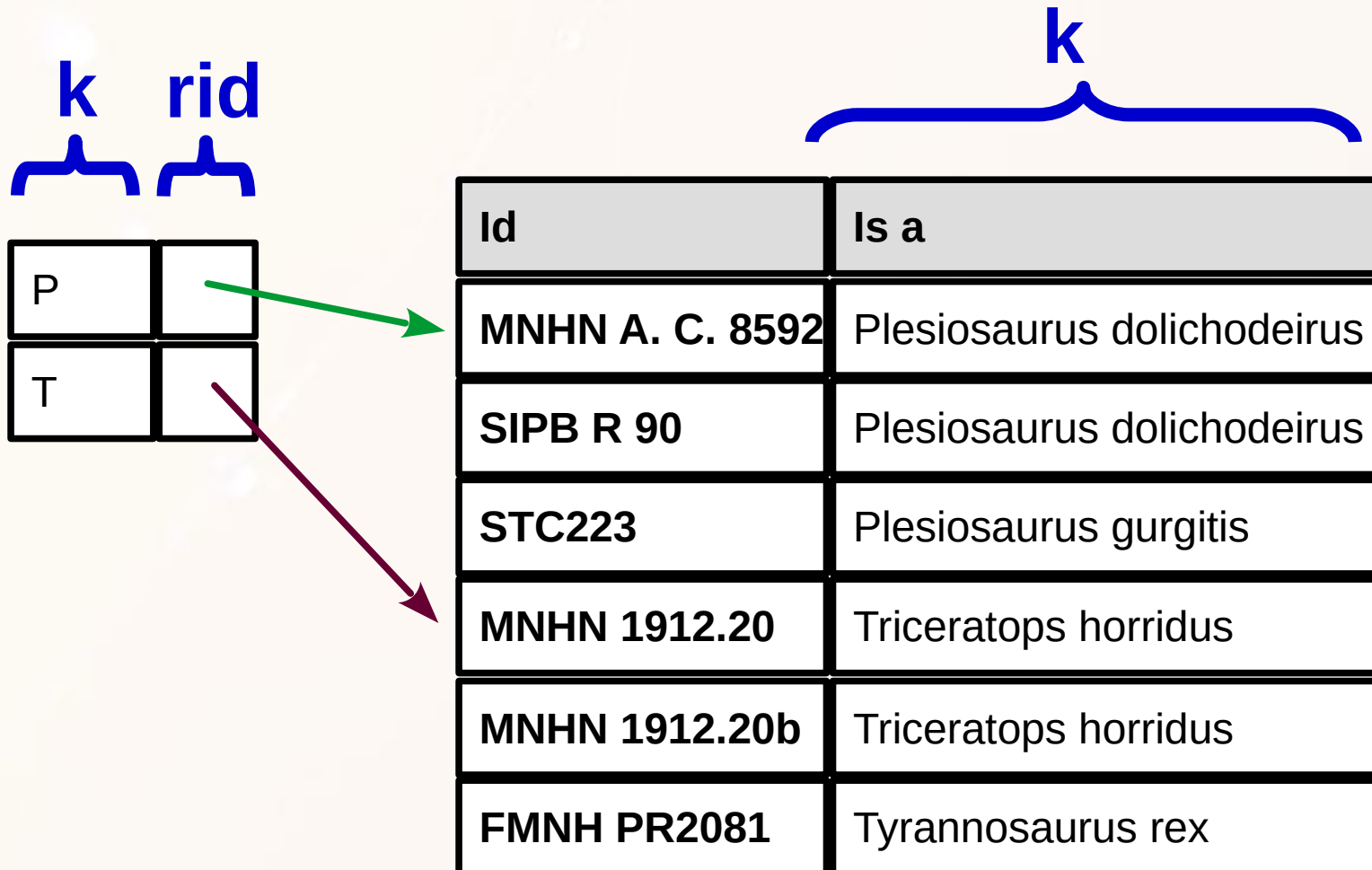
# Índice Denso

- Uma entrada de índice para cada valor de chave



# Índice Esparso

- Uma entrada índice para mais de um valor de chave



## Exercício 3

- Em uma relação com 5 atributos, qual o número máximo possível de índices primários e secundários ao mesmo tempo? Justifique.

# Estruturas de Índice



# Hashing

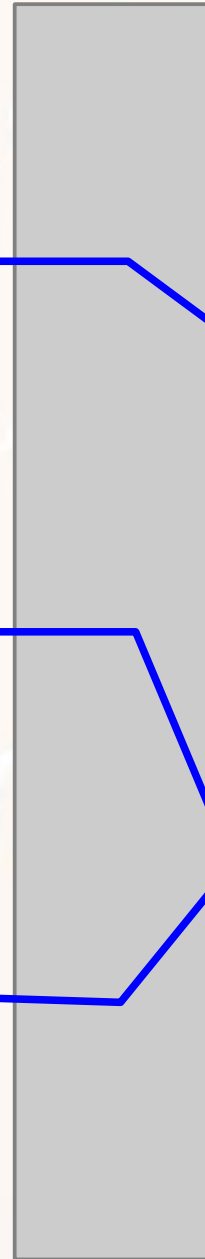
função hash

buckets

**1234** Quincas

**9876** Dorianana

**6543** Asdrúbabal



1

...

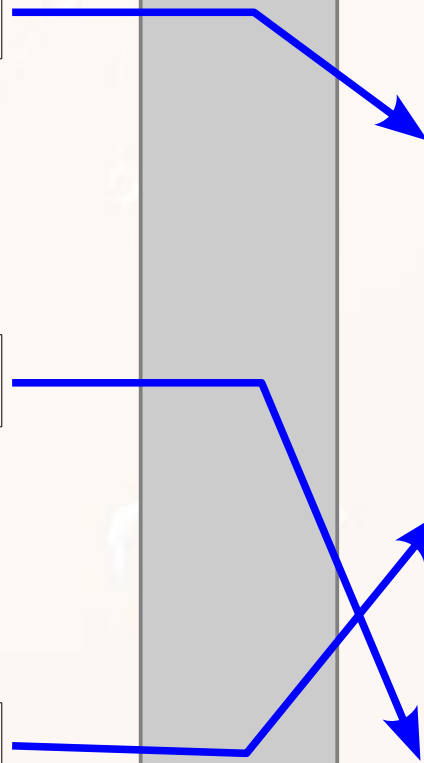
12 1234/Quincas

...

65 6543/Asdrúbabal

...

98 9876/Dorianana



# função hash

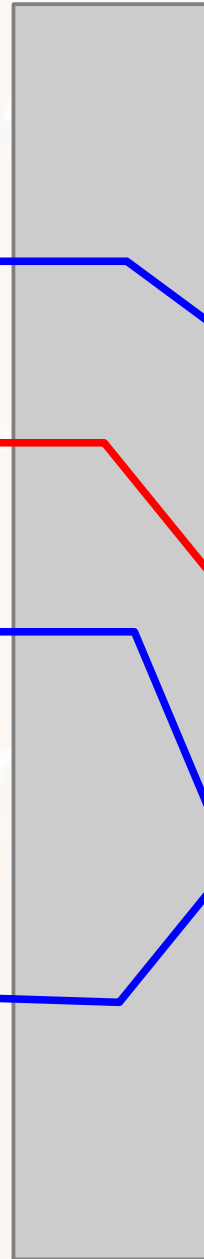
# buckets

**1234** Quincas

**6500** Lucinda

**9876** Doriana

**6543** Asdrúbal



1

...

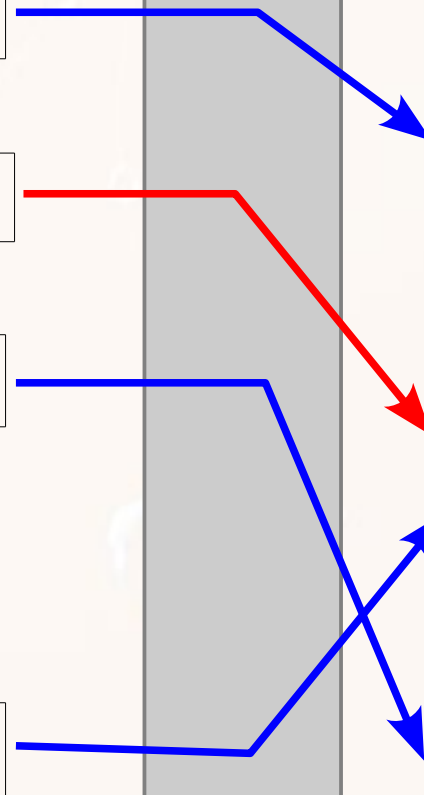
12 1234/Quincas

...

65 6543/Asdrúbal

...

98 9876/Doriana



# função hash

# buckets

**1234** Quincas

**6500** Lucinda

**9876** Dorianana

**6543** Asdrúbal

1

12 1234/Quincas

65 6543/Asdrúbal      6500/Lucinda

98 9876/Dorianana

...

...

...

# função hash

# buckets

**1234** Quincas

**6500** Lucinda

**9876** Doriana

**6543** Asdrúbal

000

001

010

011

100

101

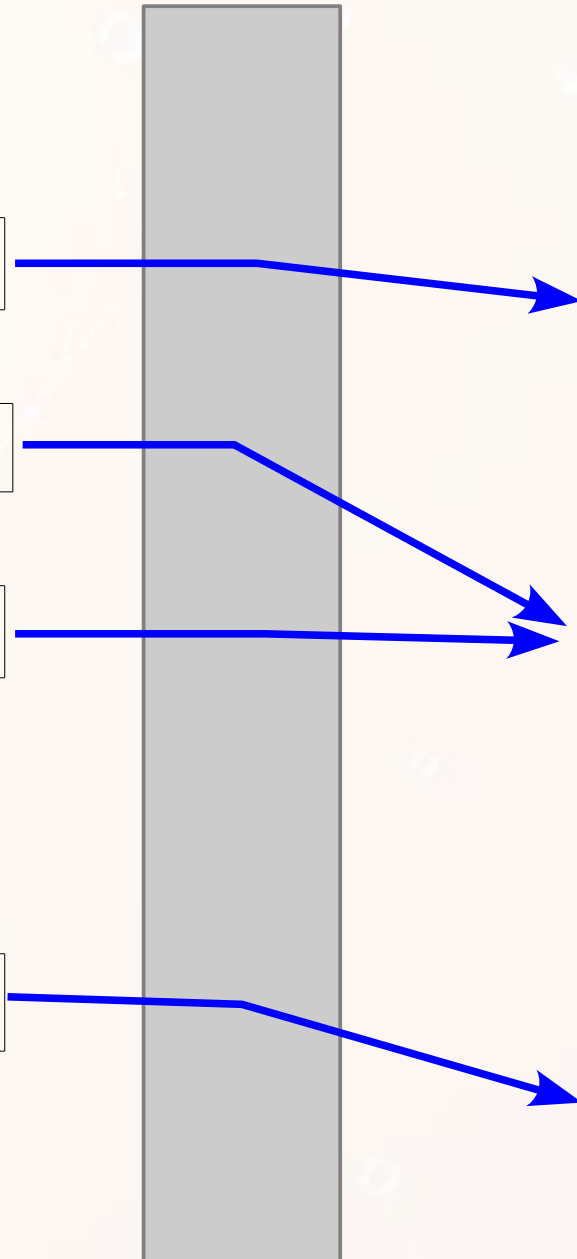
110

111

1234/Quincas

9876/Doriana      6500/Lucinda

6543/Asdrúbal

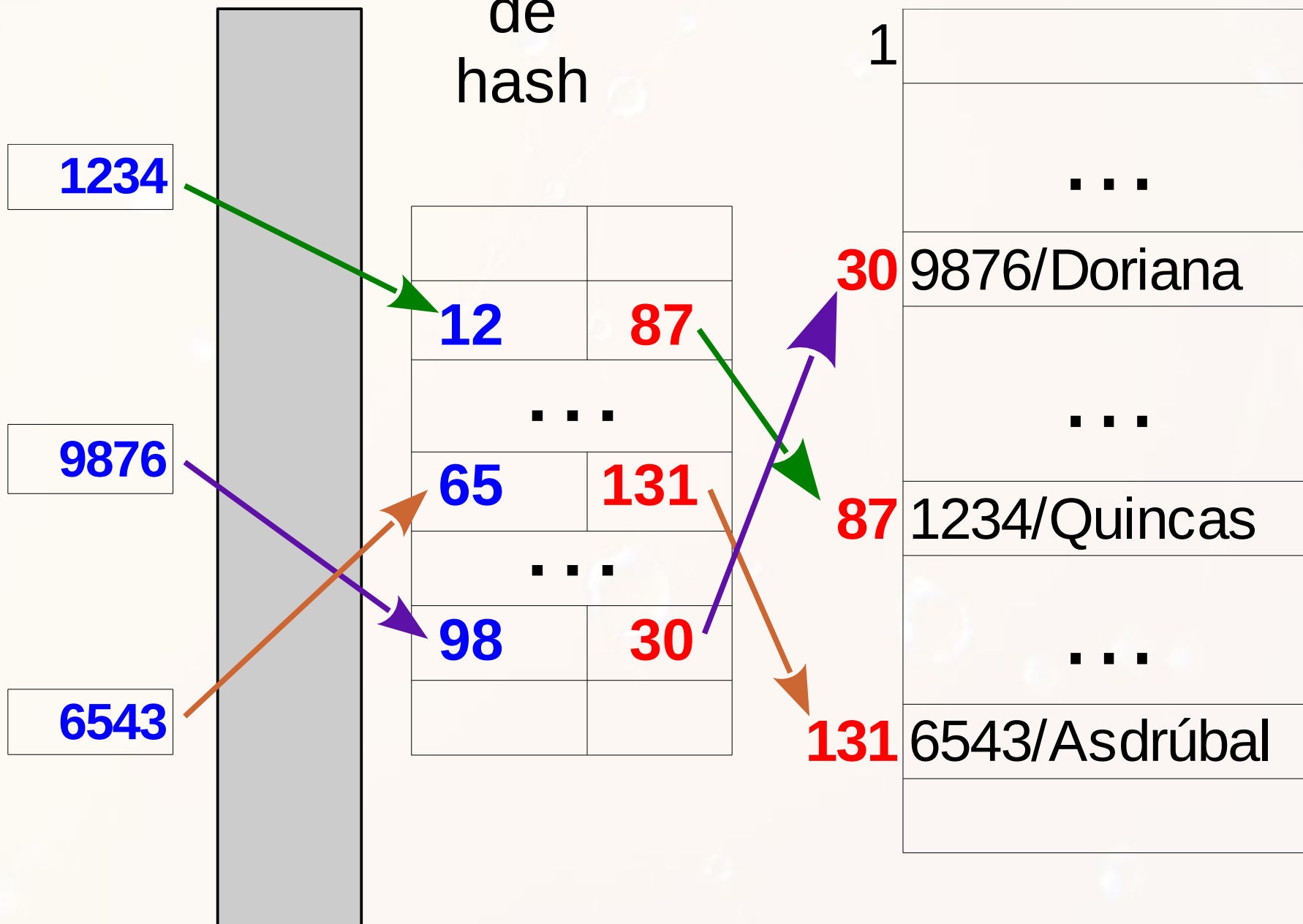


# Índice de Hash

função hash

índice de hash

registros



função hash

índice de hash

buckets

1234

9876

6543

<b>12</b>	<b>87</b>
...	
<b>65</b>	<b>131</b>
...	
<b>98</b>	<b>30</b>

1

...

30

1234/Quincas

...

87

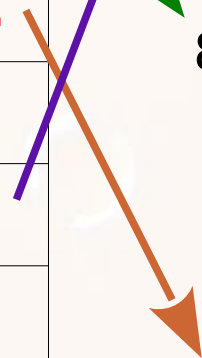
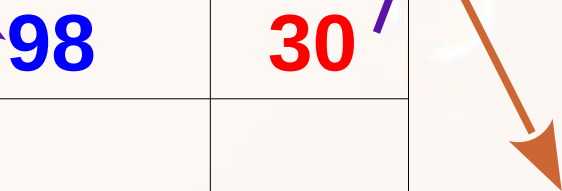
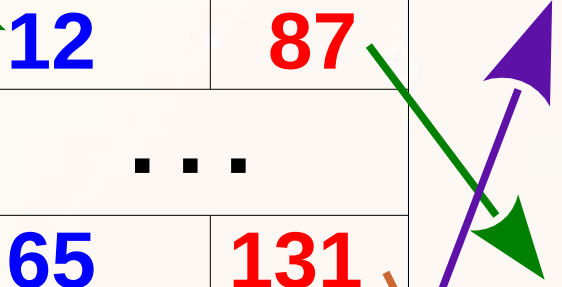
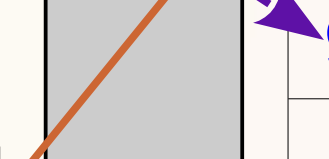
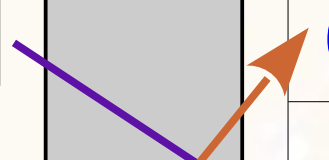
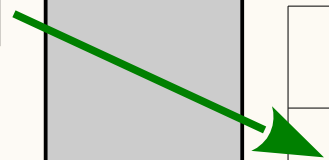
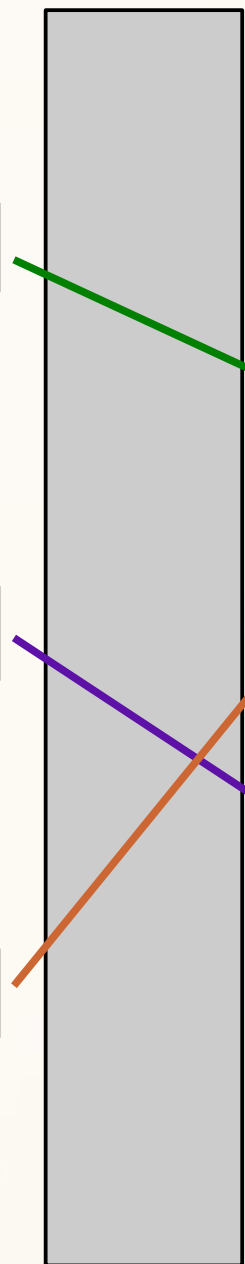
6543/Asdrúbal

6500/Lucinda

...

131

9876/Doriana





# Hashing Extensível

# função hash

# buckets

**1234** Quincas

**6500** Lucinda

**9876** Doriana

**6543** Asdrúbal

0 0 0	
0 0 1	
0 1 0	
0 1 1	
1 0 0	
1 0 1	
1 1 0	
1 1 1	

d'=2

d'=2

d'=2

d'=2

1234/Quincas

9876/Doriana      6500/Lucinda

6543/Asdrúbal

# função hash

# buckets

**1234** Quincas

**6500** Lucinda

**9876** Dorianana

**8500** D ó ris

**6543** Asdrúbal

0 0 0	
0 0 1	
0 1 0	
0 1 1	
1 0 0	
1 0 1	
1 1 0	
1 1 1	

d'=2

d'=2

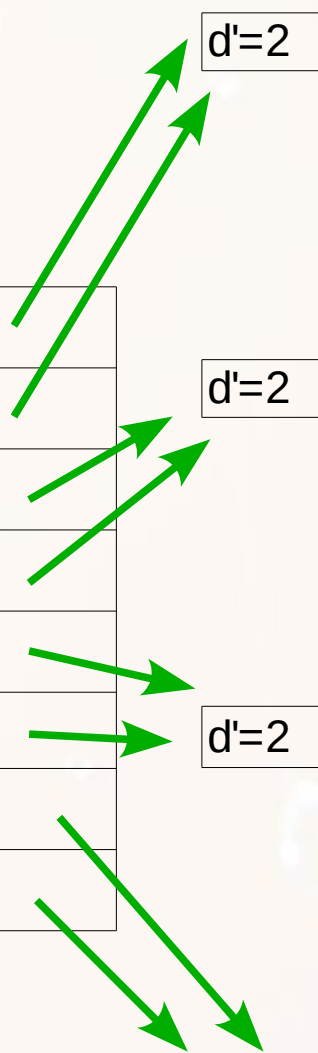
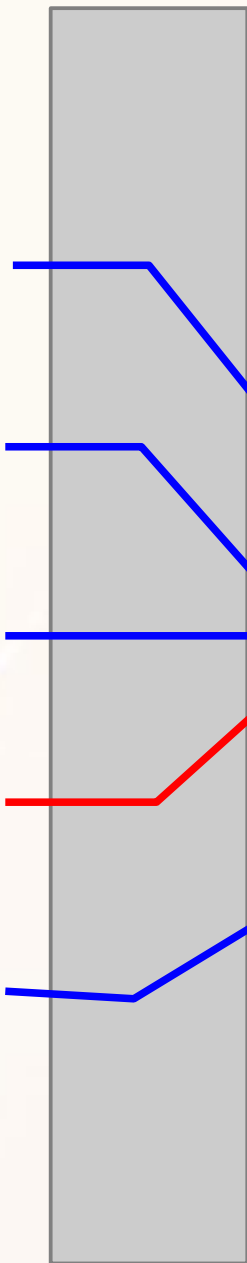
d'=2

d'=2

1234/Quincas

9876/Dorianana      6500/Lucinda  
----/Zandor            ----/Alcebíades  
----/Melissa            ----/Bonerges

6543/Asdrúbal



# função hash

# buckets

**1234** Quincas

**6500** Lucinda

**9876** Doriana

**8500** D ó ris

**6543** Asdrúbal

0 0 0	
0 0 1	
0 1 0	
0 1 1	
1 0 0	
1 0 1	
1 1 0	
1 1 1	

d'=2

d'=2

d'=3

d'=3

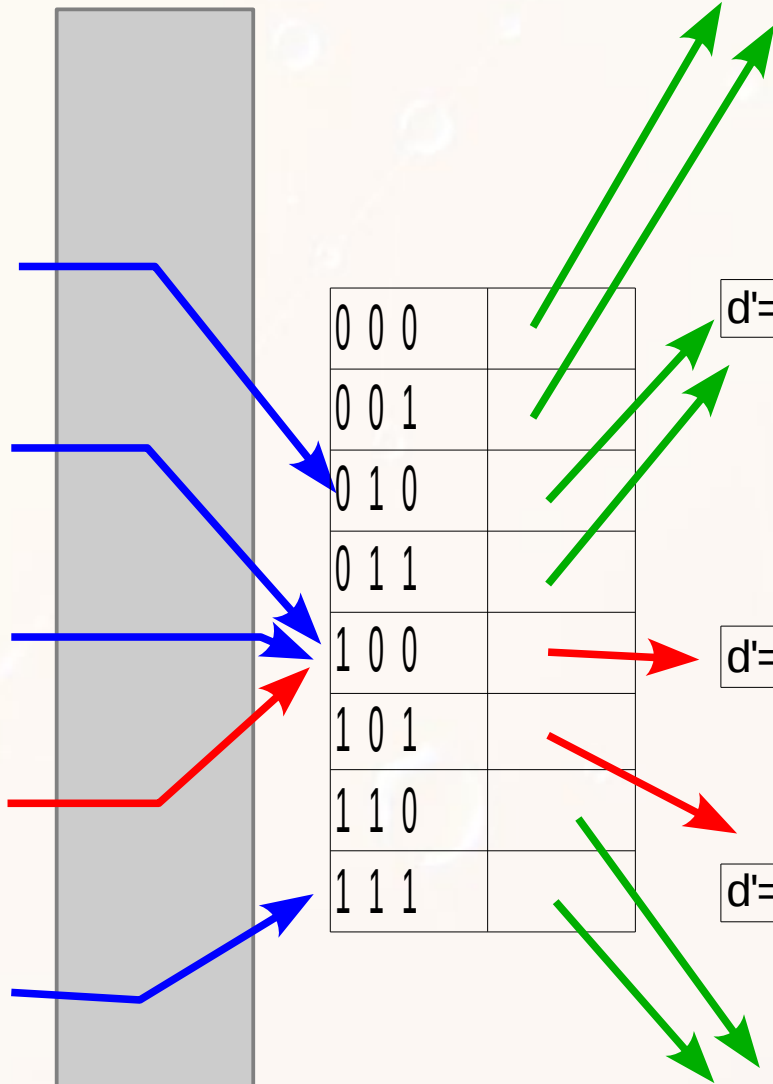
d'=2

1234/Quincas

9876/Doriana      6500/Lucinda  
----/Zandor      ----/Alcebíades  
----/Melissa      ----/Bonerges

8500/Dóris

6543/Asdrúbal



# Índices Multiníveis

# Hashing Dinâmico





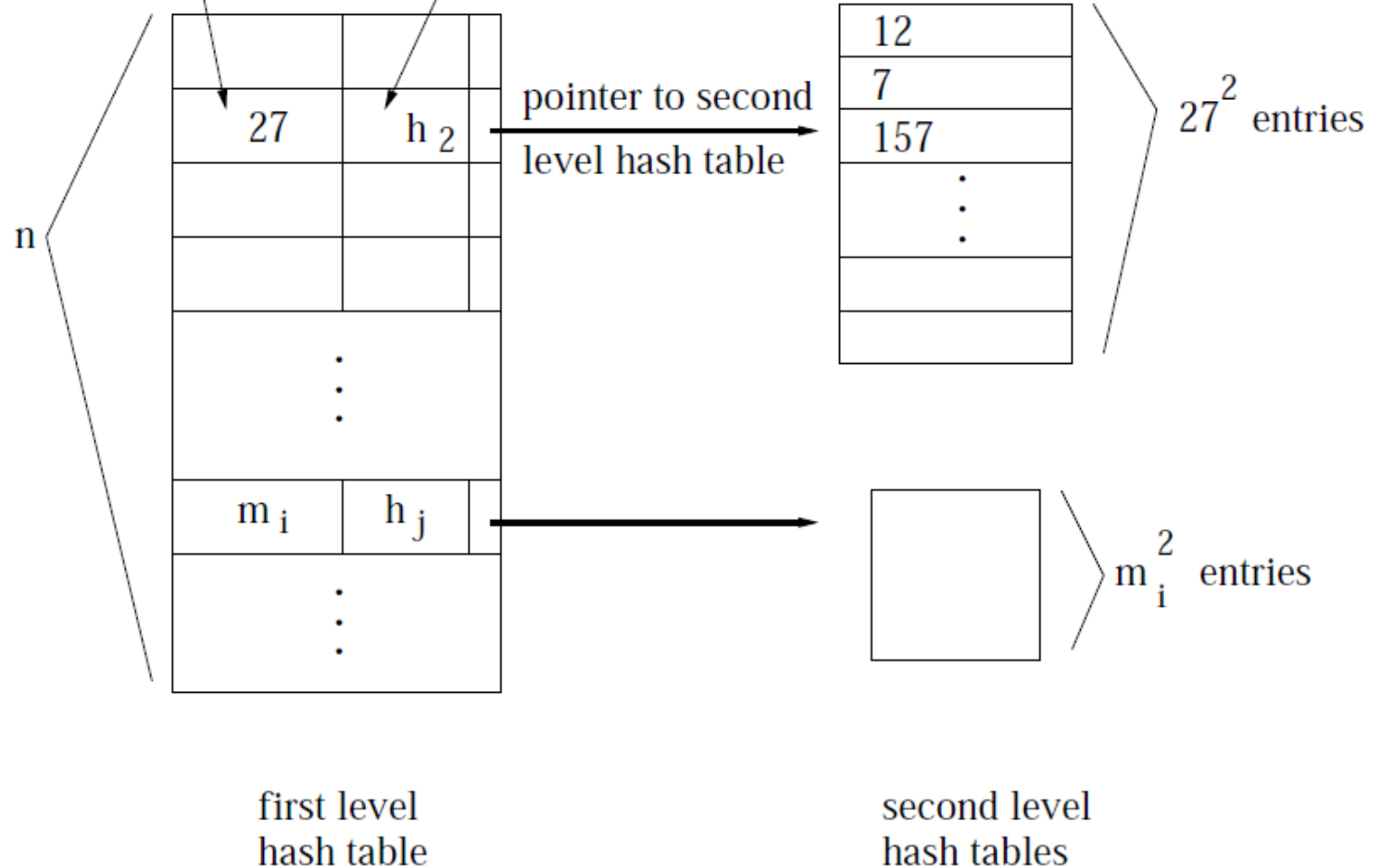




# Perfect Hashing

how many elements hash to this slot

a hash function from family H

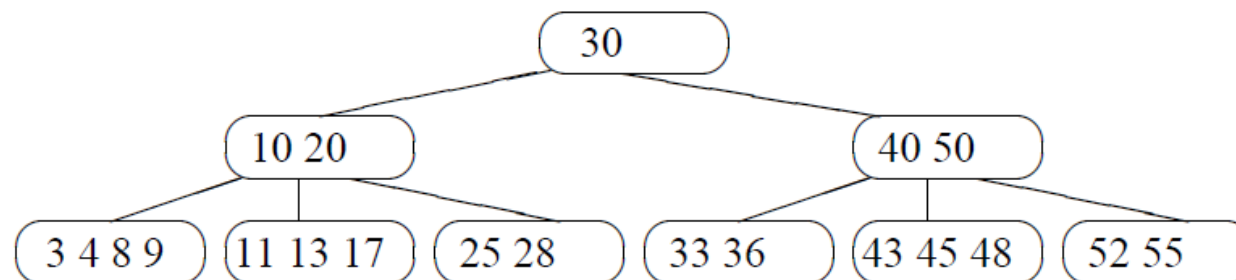


(Demaine, 2003)

# Árvores B

## Árvores B

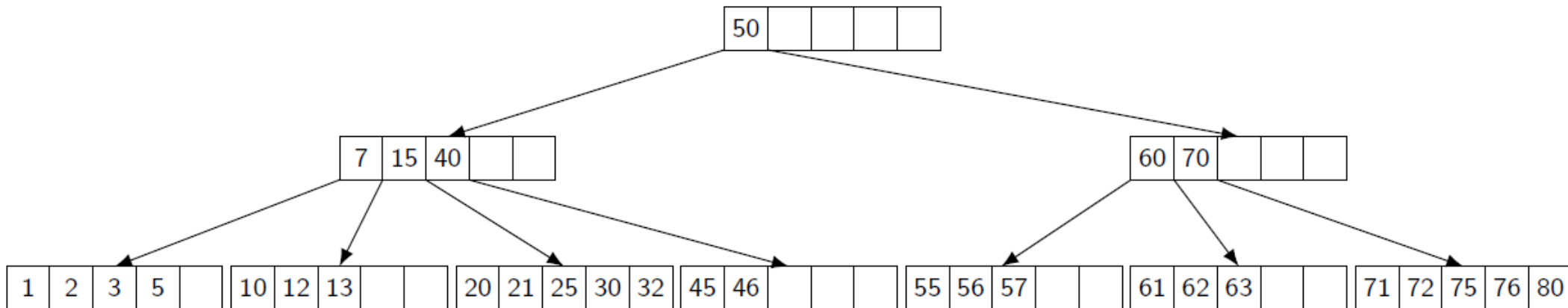
- Árvores  $n$ -árias: mais de um registro por nodo.
- Em uma árvore B de ordem  $m$ :
  - página raiz: 1 e  $2m$  registros.
  - demais páginas: no mínimo  $m$  registros e  $m + 1$  descendentes e no máximo  $2m$  registros e  $2m + 1$  descendentes.
  - páginas folhas: aparecem todas no mesmo nível.
- Registros em ordem crescente da esquerda para a direita.
- Extensão natural da árvore binária de pesquisa.
- Árvore B de ordem  $m = 2$  com três níveis:



(Almeida,  
2010)

# Exemplo de árvore B de ordem 5

Neste caso, cada nó tem no mínimo dois e no máximo cinco registros de informação.

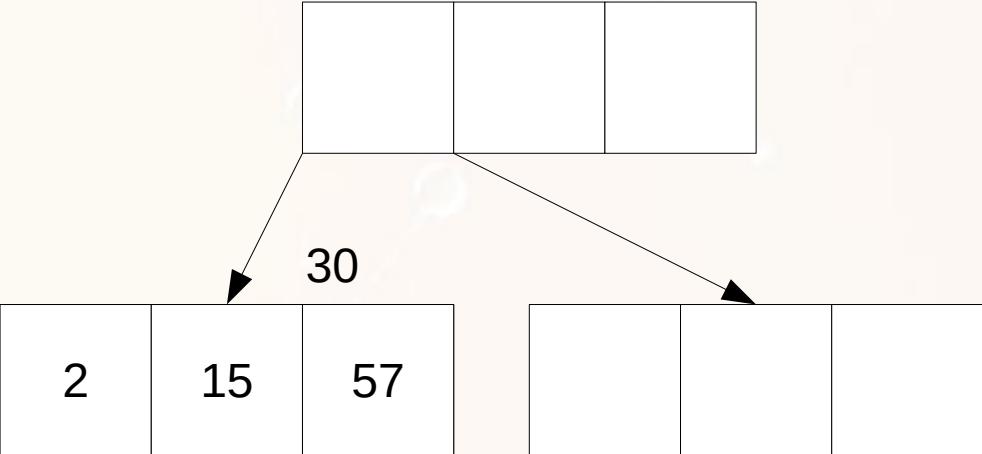


15, 2, 57, 30

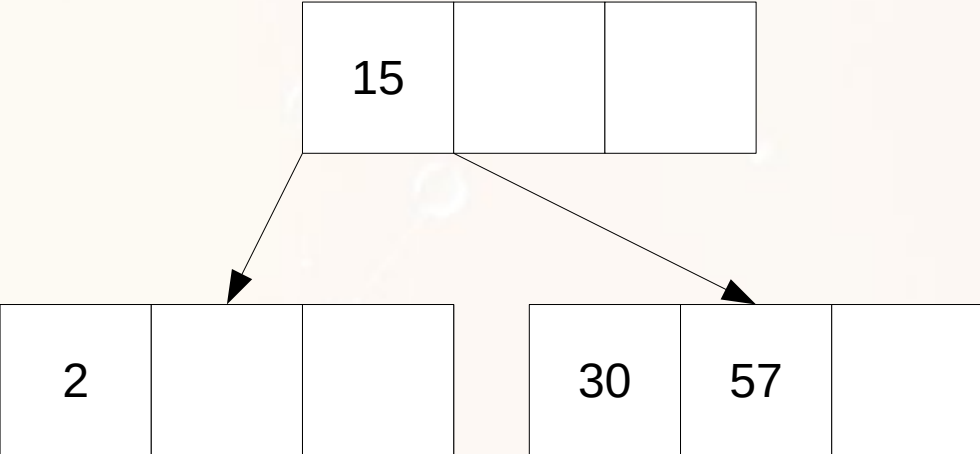
30

2	15	57
---	----	----

15, 2, 57, 30

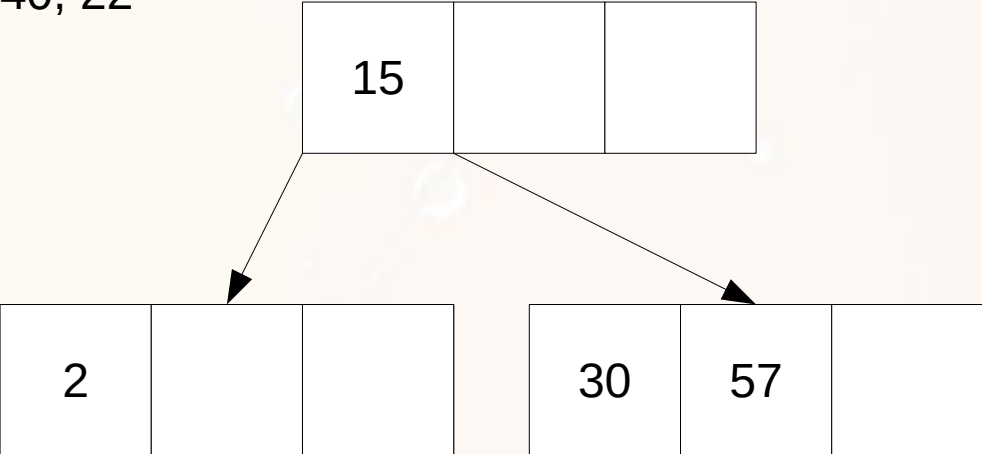


15, 2, 57, 30

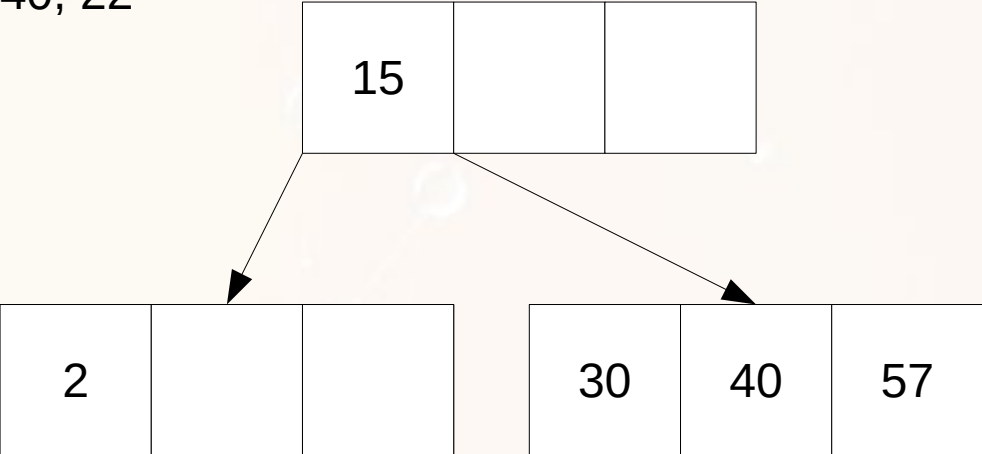




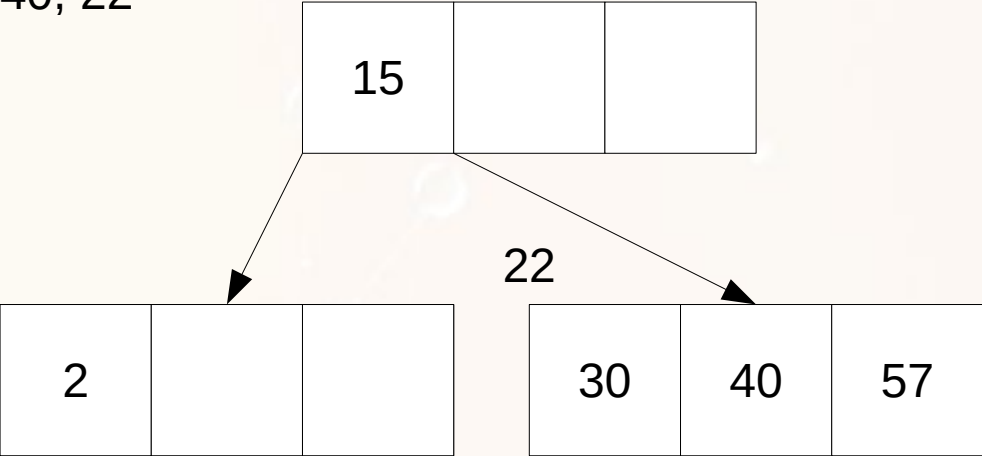
15, 2, 57, 30, 40, 22



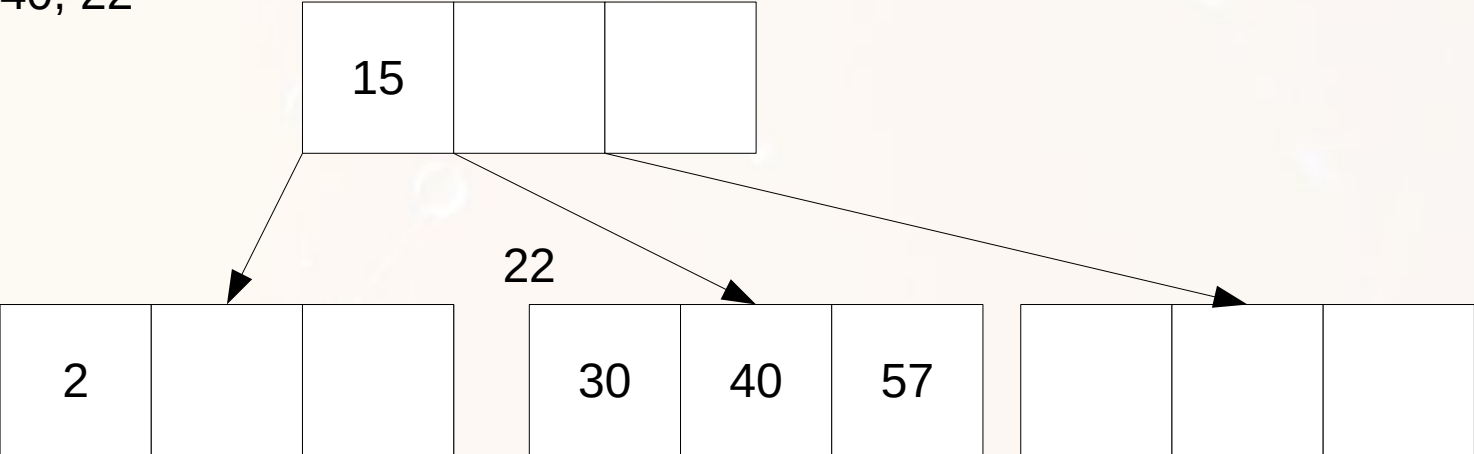
15, 2, 57, 30, 40, 22



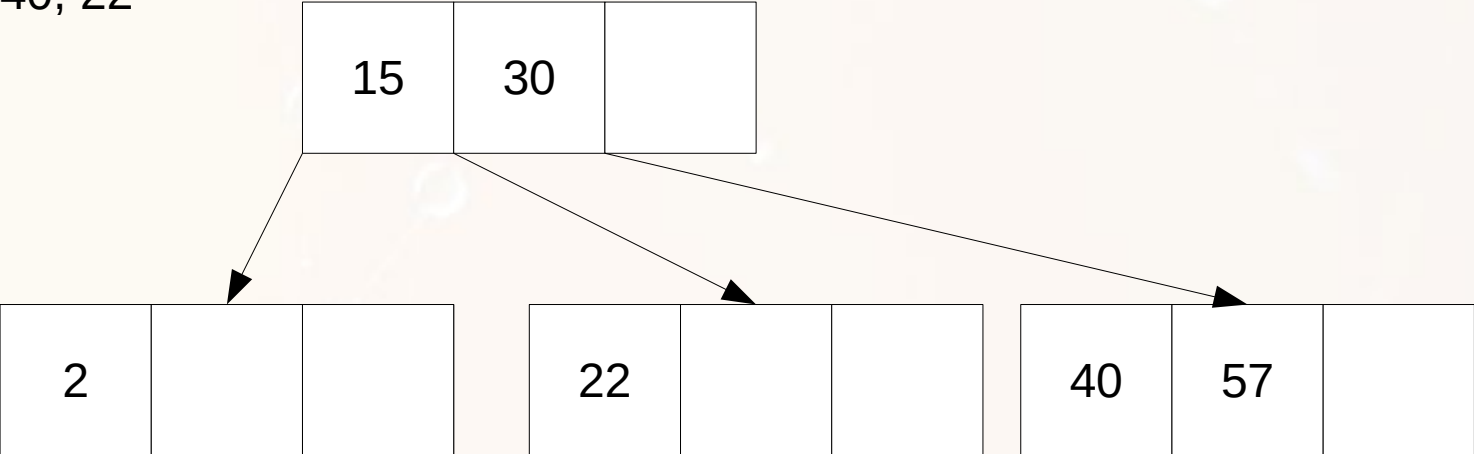
15, 2, 57, 30, 40, 22



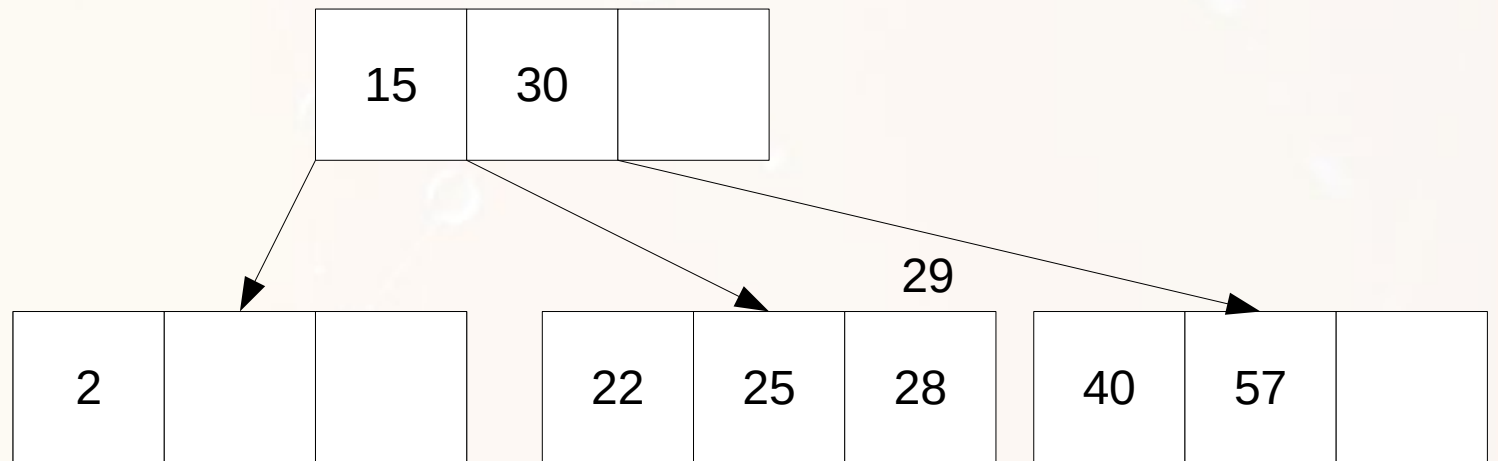
15, 2, 57, 30, 40, 22



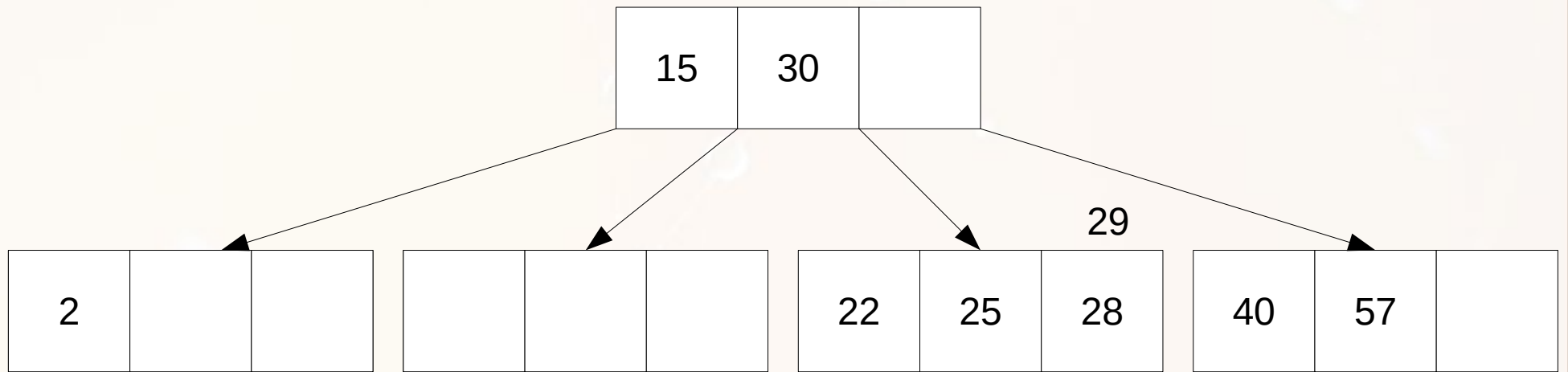
15, 2, 57, 30, 40, 22



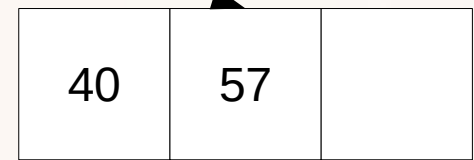
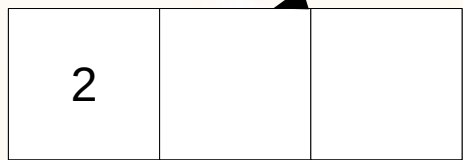
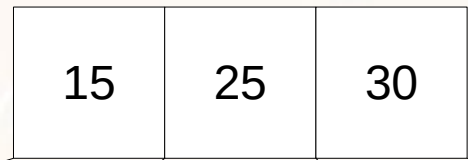
15, 2, 57, 30, 40, 22, 25, 28, 29



15, 2, 57, 30, 40, 22, 25, 28, 29

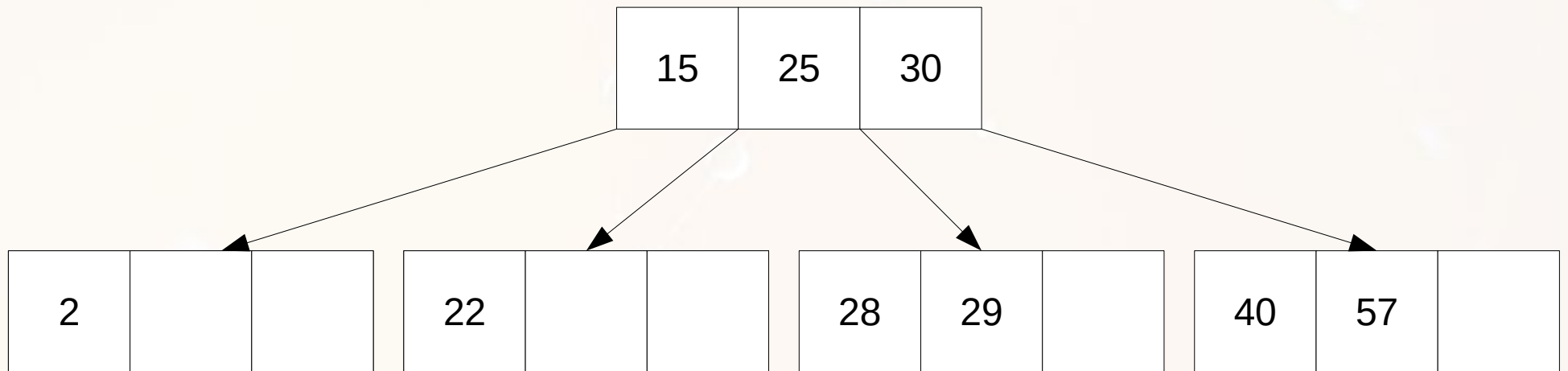


15, 2, 57, 30, 40, 22, 25, 28, 29





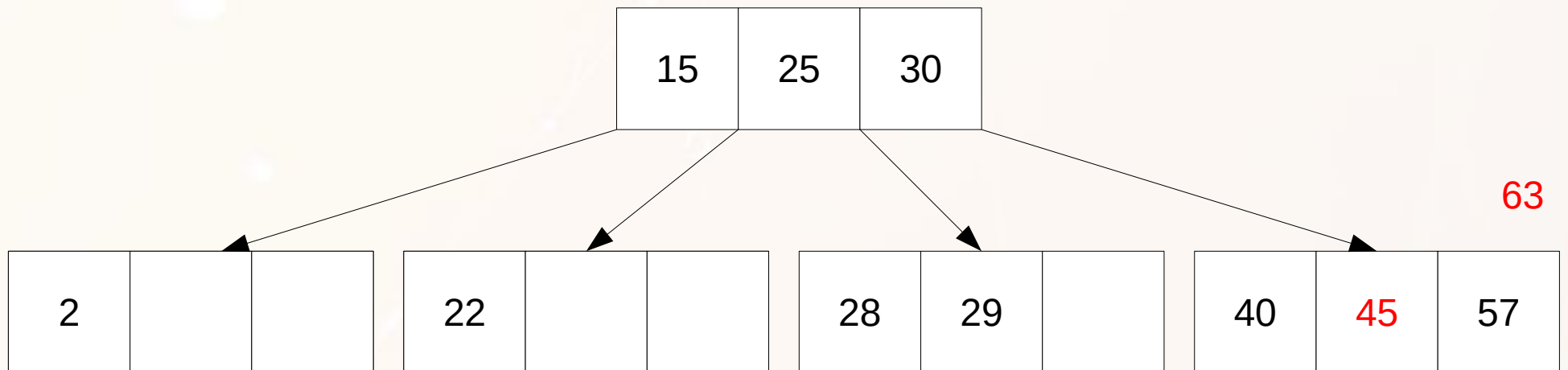
15, 2, 57, 30, 40, 22, 25, 28, 29, 45, 63, 70



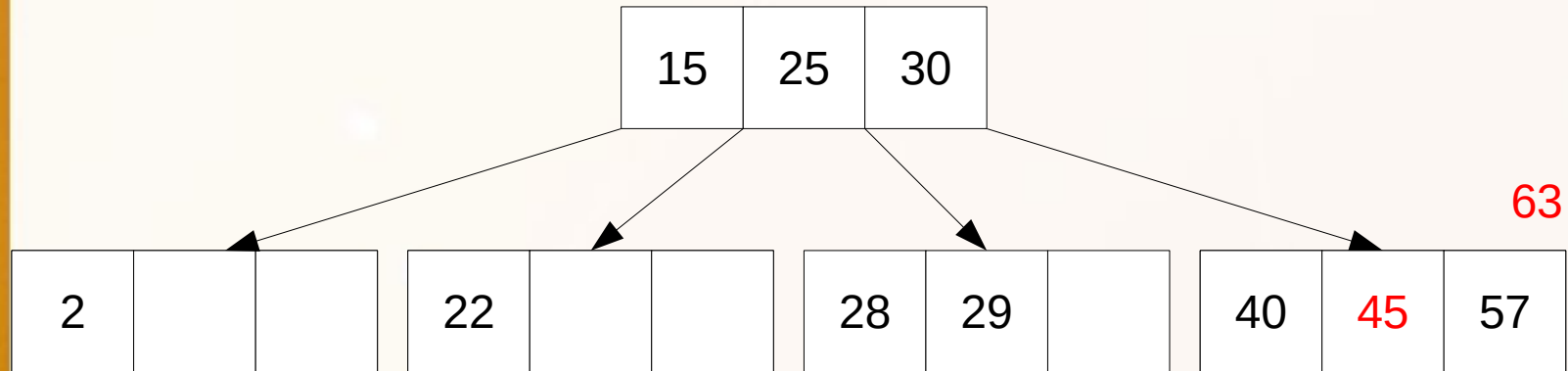
## Questão

- Como ficará a árvore com a introdução destes três números?

15, 2, 57, 30, 40, 22, 25, 28, 29, 45, 63, 70

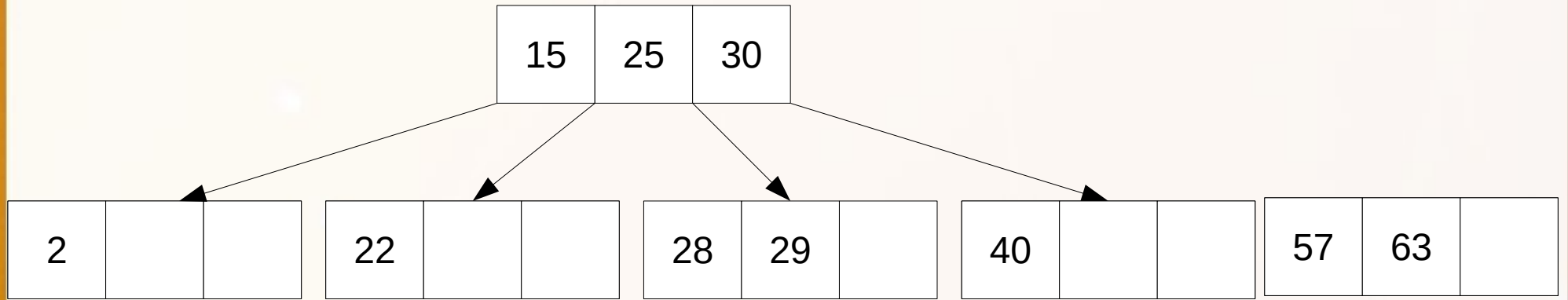


15, 2, 57, 30, 40, 22, 25, 28, 29, 45, 63, 70

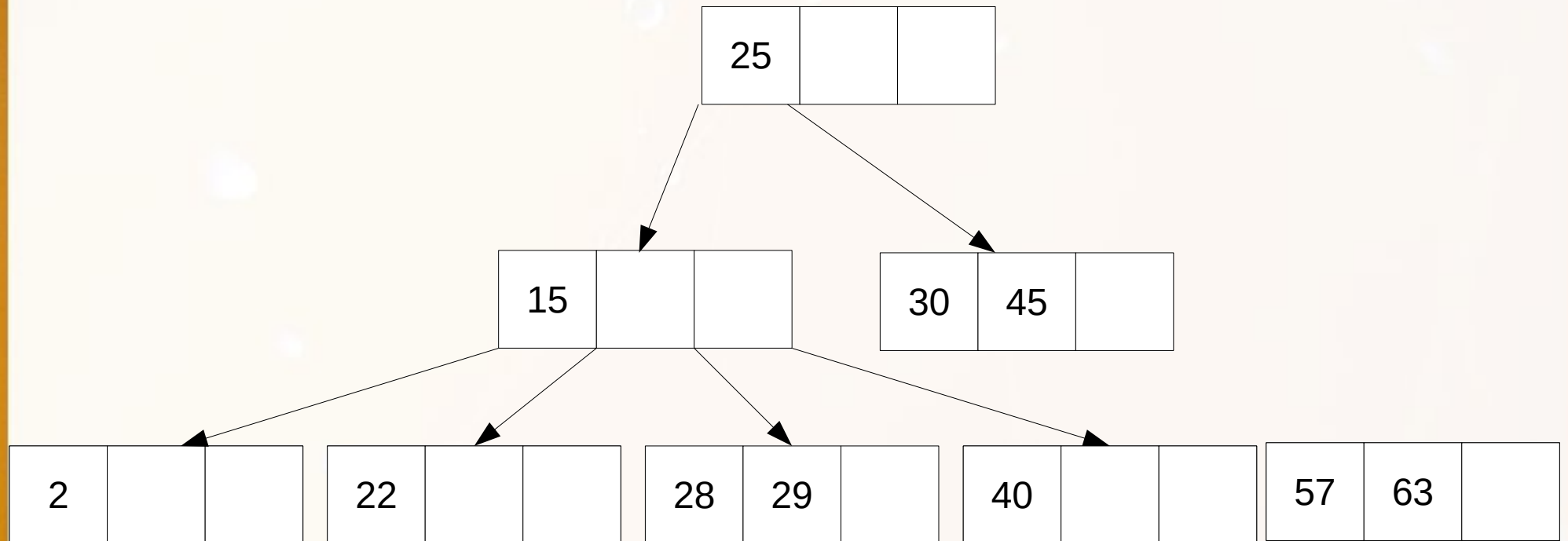


15, 2, 57, 30, 40, 22, 25, 28, 29, 45, 63, 70

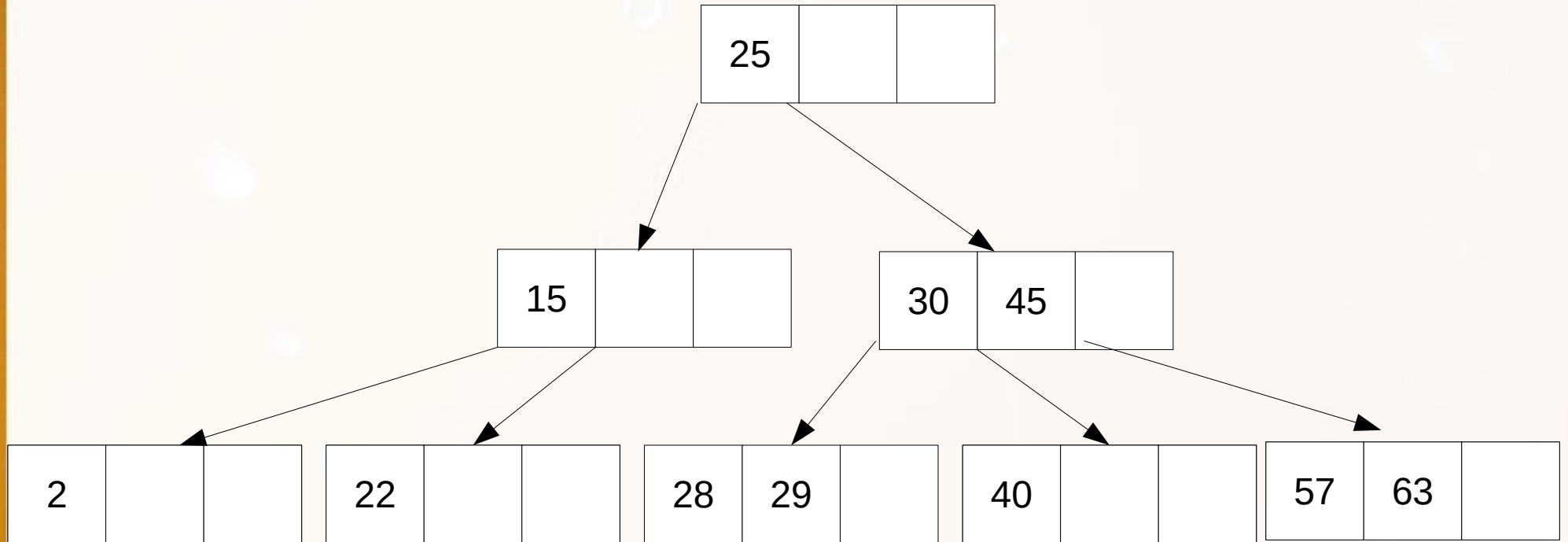
45



15, 2, 57, 30, 40, 22, 25, 28, 29, 45, 63, 70



15, 2, 57, 30, 40, 22, 25, 28, 29, 45, 63, 70



# Números mínimos e máximos de registros

Árvore B de ordem 255:

nível	mínimo		máximo	
	nós	registros	nós	registros
1	1	1	1	$1 \times 255$
2	2	$2 \times 127$	$256^1$	$256^1 \times 255$
3	$2 \times 128^1$	$2 \times 128^1 \times 127$	$256^2$	$256^2 \times 255$
4	$2 \times 128^2$	$2 \times 128^2 \times 127$	$256^3$	$256^3 \times 255$
5	$2 \times 128^3$	$2 \times 128^3 \times 127$	$256^4$	$256^4 \times 255$
Total	4.227.331	536.870.911	4.311.810.305	1.099.511.627.775

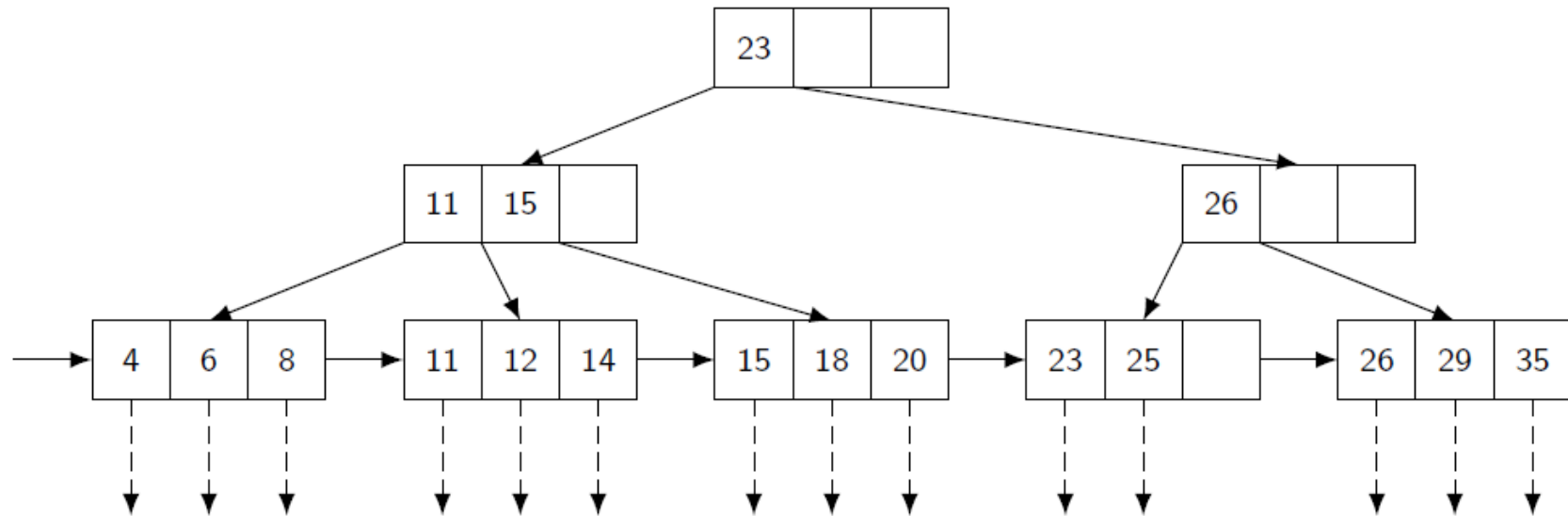
# Variantes de árvores B

- ▶ Árvores B\*: o número de registros ocupados de um nó é no mínimo  $\frac{2}{3}$  da sua capacidade.
- ▶ Árvores B<sup>+</sup>:
  - ▶ nós internos com chaves apenas para orientar o percurso
  - ▶ pares (*chave, valor*) apenas nas folhas
  - ▶ regra de descida:
    - ▶ subárvore esquerda: menor
    - ▶ subárvore direita: maior ou igual
  - ▶ apontadores em lugar de valores tornando mais eficiente a movimentação dos registros durante inserções e remoções
  - ▶ ligações facilitando percurso em ordem de chaves



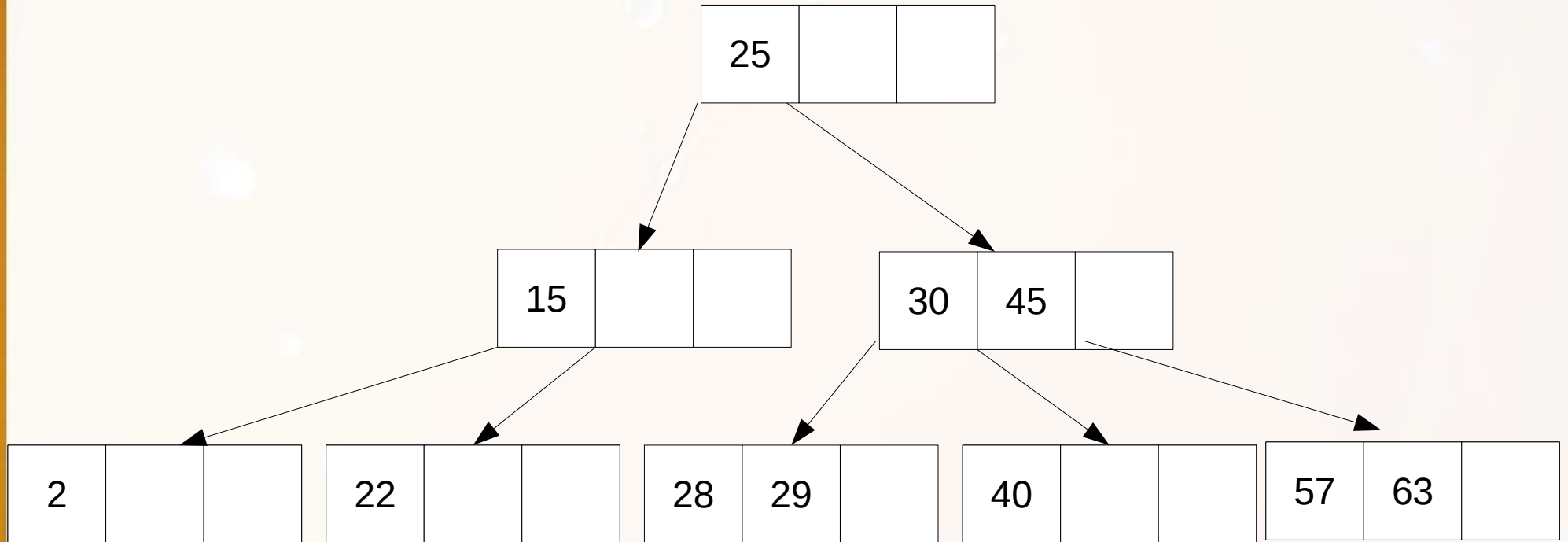
## Variante de árvores B (cont.)

Exemplo de árvore B<sup>+</sup> de ordem 3:



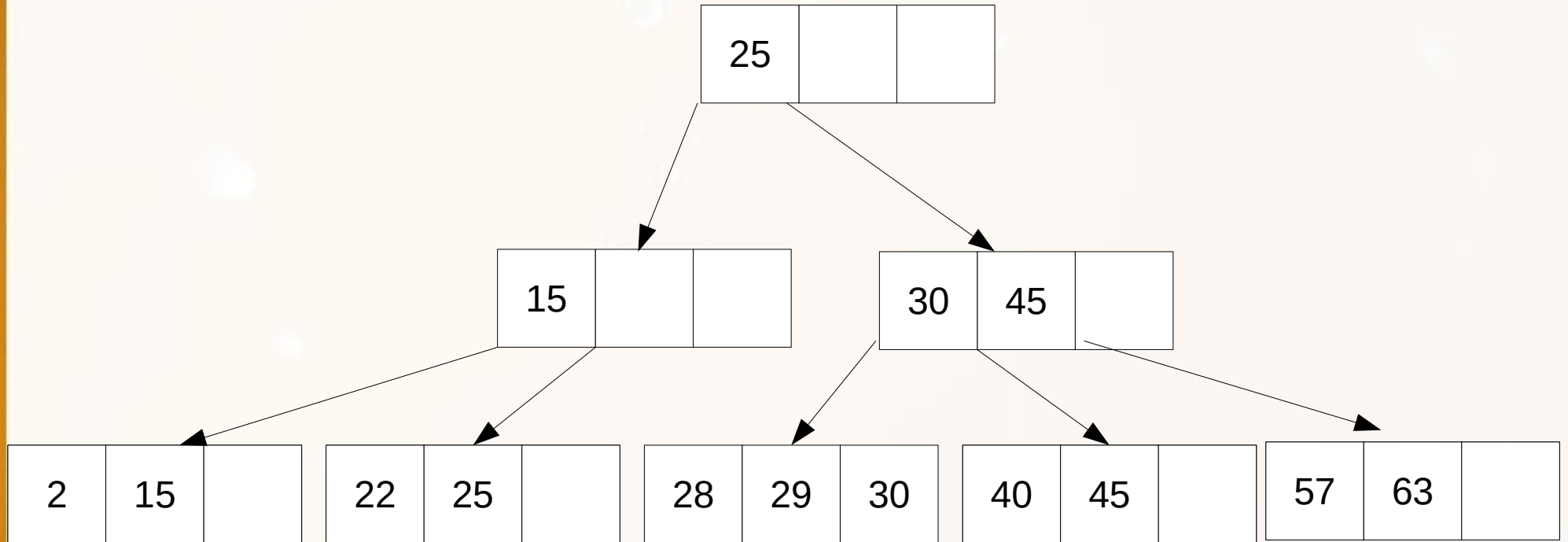
Setas tracejadas indicam apontadores para os valores da informação. A lista ligada das folhas permite percurso simples e eficiente em ordem de chaves.

# Árvore B



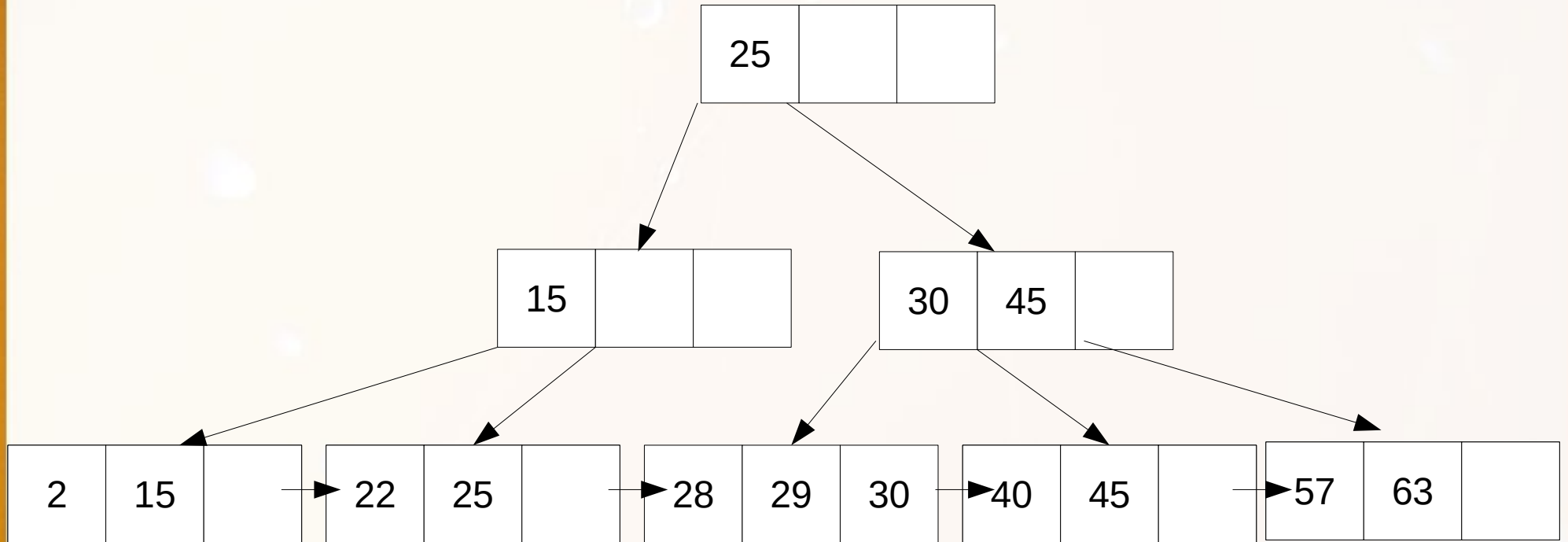
15, 2, 57, 30, 40, 22, 25, 28, 29, 45, 63, 70

# Árvore B+



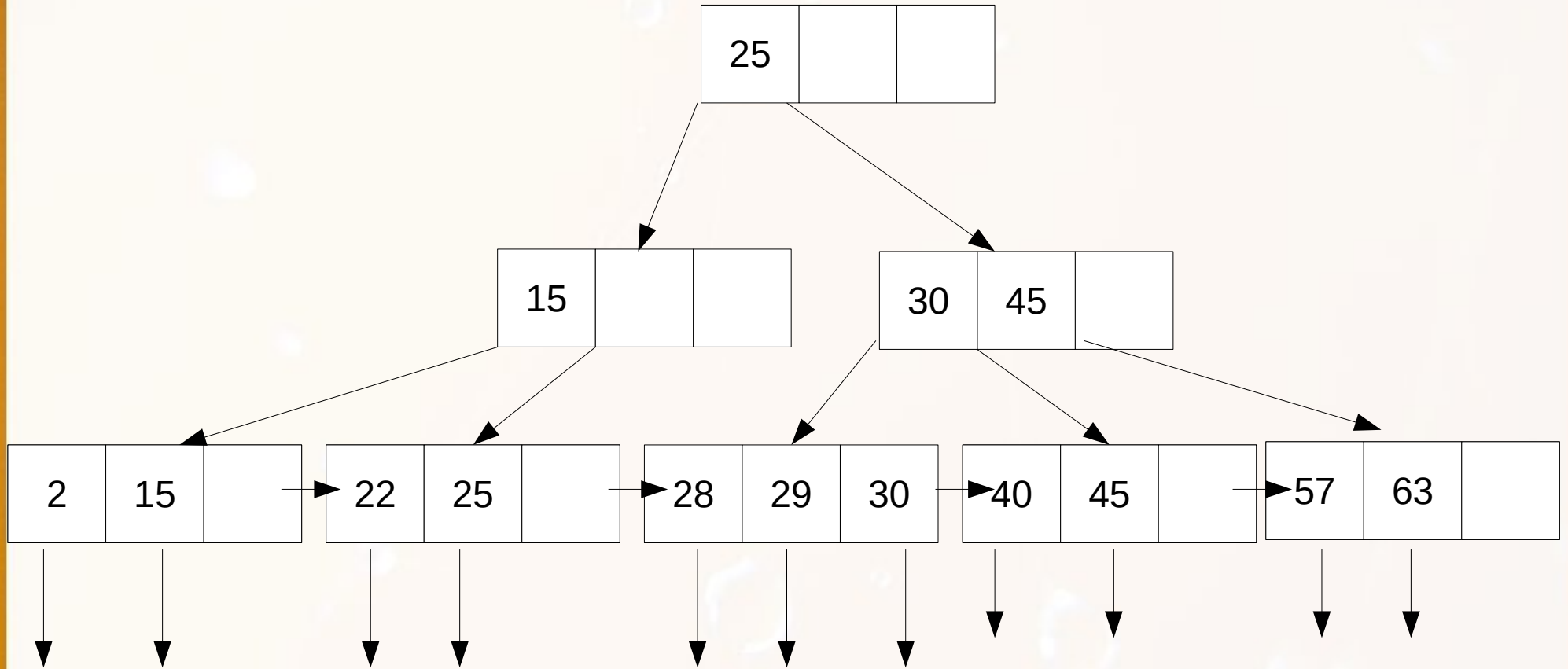
15, 2, 57, 30, 40, 22, 25, 28, 29, 45, 63, 70

# Árvore B+



15, 2, 57, 30, 40, 22, 25, 28, 29, 45, 63, 70

# Árvore B+



15, 2, 57, 30, 40, 22, 25, 28, 29, 45, 63, 70

**Índice Mapa de Bits**  
quando o foco é a **análise**

# OLTP x OLAP

## ■ OLTP

- Online Transaction Processing
- Bancos de dados “tradicionais”
- Operações de inserção, atualização e exclusão em pequenas partes do banco

## ■ OLAP

- Online Analytical Processing
- Operações de extração, recuperação e análise de dados

# Data Warehouse

## Comparada aos Bancos de Dados Tradicionais

### ■ Igual

- Coleção de dados relacionados
- Suportado por um sistema de gerenciamento

### ■ Diferente

- Orientada a aplicações de suporte a decisão
- Otimizada para recuperação de dados - não OLTP



# Índice Mapa de Bits

- Facilita consultas sobre chaves múltiplas
- Cada índice baseado em uma chave
- Mapa de bits sobre atributo **A** e relação **r**
  - mapa de bits = array de bits
  - tamanho do array = número registros de **r**
  - um mapa de bits (array) para cada valor de **A**

(Silberschatz et al., 2006)

# Mapa de Bits

S

S#	SNAME	STATUS	CITY
S1	SMITH	20	London
S2	JONES	10	Paris
S3	BLAKE	30	Paris
S4	CLARK	20	London
S5	ADAMS	30	Athens

P

P#	PNAME	COLOR	WEIGHT	CITY
P1	NUT	RED	12	London
P2	BOLT	GREEN	17	Paris
P3	SCREW	BLUE	17	Rome
P4	SCREW	RED	14	London
P5	CAM	BLUE	12	Paris
P6	COG	RED	19	London

SP

S#	P#	QTY
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

S<sub>b</sub>

	STATUS				CITY			
	10	20	30	40	London	Paris	Athens	Rome
S1	0	1	0	0	1	0	0	0
S2	1	0	0	0	0	1	0	0
S3	0	0	1	0	0	1	0	0
S4	0	1	0	0	1	0	0	0
S5	0	0	1	0	0	0	0	0

P<sub>b</sub>

	COLOR				WEIGHT					CITY				
	RED	GREEN	BLACK	BLUE	10	12	14	17	18	19	London	Paris	Athens	Rome
P1	1	0	0	0	0	1	0	0	0	0	1	0	0	0
P2	0	1	0	0	0	0	0	1	0	0	0	1	0	0
P3	0	0	0	1	0	0	0	1	0	0	0	0	0	1
P4	1	0	0	0	0	0	1	0	0	0	1	0	0	0
P5	0	0	0	1	0	1	0	0	0	0	0	1	0	0
P6	1	0	0	0	0	0	0	0	0	1	1	0	0	0

Fig. 1. (a,b) Binary representation of supplier—part data. (a) Supplier part data base (Date, 1982). (b) Binary representation.\*

\* Supplier and part names and SP relation are not shown to simplify the exposition.

(Spiegler & Maayan, 1985)

# Índice Mapa de Bits

## Otimizações

### ■ Codificação

- exemplo: codificação binária de possíveis valores

### ■ Compressão

- exigem descompressão para operações
- participam de operações sem descompressão

# Compressão de Mapa de Bits

A	1	2	3	4	5	6	7
2	0	1	0	0	0	0	0
4	0	0	0	1	0	0	0
1	1	0	0	0	0	0	0
3	0	0	1	0	0	0	0
2	0	1	0	0	0	0	0
1	1	0	0	0	0	0	0
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
4	0	0	0	1	0	0	0
5	0	0	0	0	1	0	0
6	0	0	0	0	0	1	0
7	0	0	0	0	0	0	1
1	1	0	0	0	0	0	0

(a)



A	{1}	{2,3}	{4,5,6,7}
2	0	1	0
4	0	0	1
1	1	0	0
3	0	1	0
2	0	1	0
1	1	0	0
1	1	0	0
2	0	1	0
4	0	0	1
5	0	0	1
6	0	0	1
7	0	0	1
1	1	0	0

(b)

(Koudas, 2000)

# Exercício para Casa 1

- Considere a relação Aluno(ra, curso, idade) que armazene estes dados para todos os alunos da Unicamp. Para cada uma das questões a seguir, defina qual o tipo de índice mais indicado.
  - a) `select * from Aluno where ra=5.`
  - b) `select * from Aluno where idade<70.`
  - c) `select * from Aluno where idade>27 and B<30.`
  - d) `select avg(idade) from Aluno.`
  - e) `select idade, count(*) from aluno  
where curso="Computação"  
group by idade`

# Referências

- Almeida, Charles Ornelas , Guerra, Israel; Ziviani, Nivio (2010) **Projeto de Algoritmos** (transparências aula).
- Demaine, Erik. 6.897: Advanced Data Structures – Lecture 2 (notas de aula). Fevereiro, 2003.
- Elmasri, Ramez; Navathe, Shamkant B. (2005) **Sistemas de Bancos de Dados**. Addison-Wesley, 4ª edição em português.
- Elmasri, Ramez; Navathe, Shamkant B. (2011) **Sistemas de Bancos de Dados**. Addison-Wesley, 6ª edição em português.

# Referências

- Ramakrishnan, Raghu; Gehrke, Johannes (2003) **Database Management Systems**. McGraw-Hill, 3<sup>rd</sup> edition.
- Sedgewick, Robert; Wayne, Kevin (2008) Princeton University: **Algorithms**. Maio, 2008.
- Silberschatz, Abraham; Korth, Henry F.; Sudarshan, S. (2006) **Sistema de Banco de Dados**. Elsevier, Tradução da 5a edição.



# Referências

- N. Koudas (2000). "Space efficient bitmap indexing". Proceedings of the ninth international conference on Information and knowledge management (CIKM '00). New York, NY, USA: ACM. pp. 194–201. doi:10.1145/354756.354819
- Spiegler I; Maayan R (1985). "Storage and retrieval considerations of binary data bases". Information Processing and Management: an International Journal 21 (3): 233–54. doi:10.1016/0306-4573(85)90108-6



# Agradecimentos

- Luiz Celso Gomes Jr (professor desta disciplina em 2014) pela contribuição na disciplina e nos slides.  
Página do Celso:  
<http://dainf.ct.utfpr.edu.br/~gomesjr/>
- Patrícia Cavoto (professora desta disciplina em 2015) pela contribuição na disciplina e nos slides.
- Luana Loubet Borges pelos exercícios.

André Santanchè

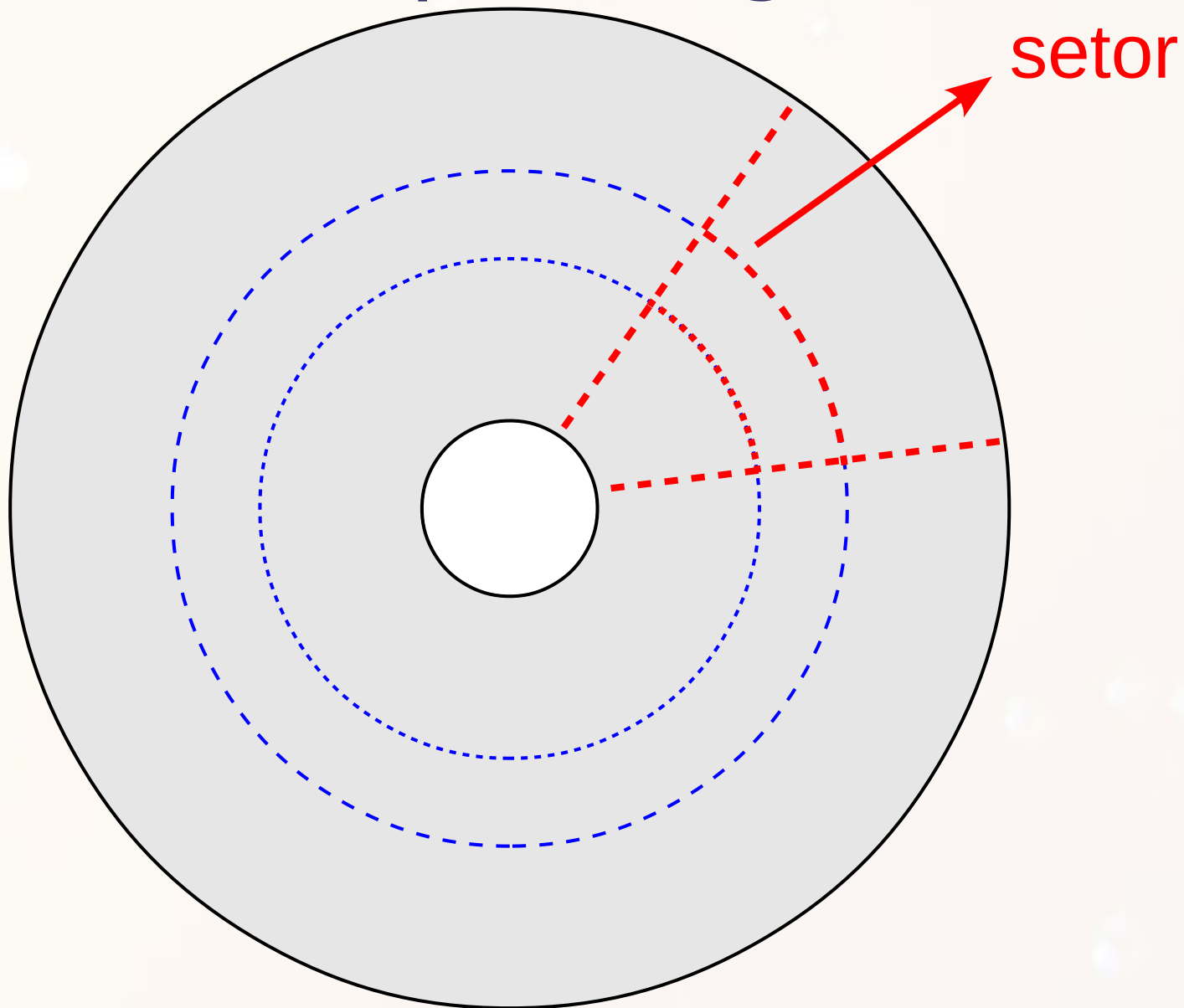
<http://www.ic.unicamp.br/~santanche>

# Licença

- Estes slides são concedidos sob uma Licença Creative Commons. Sob as seguintes condições: Atribuição, Uso Não-Comercial e Compartilhamento pela mesma Licença.
- Mais detalhes sobre a referida licença Creative Commons veja no link:  
<http://creativecommons.org/licenses/by-nc-sa/3.0/>
- Fotografia da capa e fundo por  
<http://www.flickr.com/photos/fdecomite/>  
Ver licença específica em  
<http://www.flickr.com/photos/fdecomite/1457493536/>

# Setor

## Divisão por Ângulo Fixo



# Setor

## Divisão por Densidade Constante

