

# MATA54 - Estruturas de Dados e Algoritmos II

## Busca por Chave Secundária

Flávio Assis

Versão gerada a partir de slides do Prof. George Lima

IC - Instituto de Computação

Salvador, outubro de 2021

# Busca por Chave Secundária

- ▶ Como recuperar eficientemente registros que possuem um determinado valor para uma determinada chave secundária ou uma combinação de valores para chaves secundárias?

# Busca por Chave Secundária

- ▶ Como recuperar eficientemente registros que possuem um determinado valor para uma determinada chave secundária ou uma combinação de valores para chaves secundárias?

Ex.:

- ▶ Em um arquivo com registros sobre funcionários de uma empresa, quais funcionários trabalham no departamento  $d$ ?

# Busca por Chave Secundária

- ▶ Como recuperar eficientemente registros que possuem um determinado valor para uma determinada chave secundária ou uma combinação de valores para chaves secundárias?

Ex.:

- ▶ Em um arquivo com registros sobre funcionários de uma empresa, quais funcionários trabalham no departamento  $d$ ?
- ▶ Quais funcionários trabalham no departamento  $d$  e são do sexo feminino?

# Busca por Chave Secundária

- ▶ Como recuperar eficientemente registros que possuem um determinado valor para uma determinada chave secundária ou uma combinação de valores para chaves secundárias?

Ex.:

- ▶ Em um arquivo com registros sobre funcionários de uma empresa, quais funcionários trabalham no departamento  $d$ ?
- ▶ Quais funcionários trabalham no departamento  $d$  e são do sexo feminino?
- ▶ Quais são os gerentes dos departamentos da empresa?

# Busca por Chave Secundária

- ▶ Como recuperar eficientemente registros que possuem um determinado valor para uma determinada chave secundária ou uma combinação de valores para chaves secundárias?

Ex.:

- ▶ Em um arquivo com registros sobre funcionários de uma empresa, quais funcionários trabalham no departamento  $d$ ?
  - ▶ Quais funcionários trabalham no departamento  $d$  e são do sexo feminino?
  - ▶ Quais são os gerentes dos departamentos da empresa?
- 
- ▶ Potencialmente vários registros podem satisfazer a consulta

# Busca por Chave Secundária

- ▶ Como recuperar eficientemente registros que possuem um determinado valor para uma determinada chave secundária ou uma combinação de valores para chaves secundárias?

Ex.:

- ▶ Em um arquivo com registros sobre funcionários de uma empresa, quais funcionários trabalham no departamento  $d$ ?
  - ▶ Quais funcionários trabalham no departamento  $d$  e são do sexo feminino?
  - ▶ Quais são os gerentes dos departamentos da empresa?
- 
- ▶ Potencialmente vários registros podem satisfazer a consulta

## Estrutura de dados a serem estudadas

- ▶ Arquivos multilista
- ▶ Arquivos invertidos
- ▶ Árvores  $k$ - $d$

# Arquivos Multilista



## Caracterização

- ▶ Para cada campo (chave secundária) de interesse, há um lista encadeada:

Arquivo principal: cada registro é um nó em uma das listas

chave primária	campo 1	ptr 1	campo 2	ptr 2	...	campo $k$	ptr $k$
----------------	---------	-------	---------	-------	-----	-----------	---------

- ▶ Cada ponteiro (ptr  $i$ ) liga registros pertencentes à mesma lista, de acordo com um critério pré-estabelecido.
- ▶ Uma matriz de cabeças de listas é criada para cada campo:

Cabeças da lista para campo  $i$

Valor 1	ptr $h_1$
Valor 2	ptr $h_2$
...	
Valor $m$	ptr $h_m$

# Arquivos Multilista: Exemplo

Suponha que os seguintes registros devam ser armazenados em um arquivo:

<b>Id</b>	<b>Nome</b>	<b>Depto</b>	<b>Função</b>	<b>Idade</b>
17	João	Financeiro	Gerente	40
30	Maria	Pesquisa	Gerente	30
28	Carlos	Pesquisa	Programador	35
19	Ana	Financeiro	Contador	25
42	Paulo	Financeiro	Contador	30
18	Joana	Pesquisa	Programador	32
20	José	Produção	Gerente	45

E que se deseje fazer consultas por **departamento** e **função** (chaves secundárias).

# Arquivos Multilistas

Consideremos (**por exemplo!**) que os registros serão armazenados em um arquivo com 11 posições, utilizando **hashing duplo** com as seguinte funções:

- ▶  $h_1(c) = c \bmod 11$
- ▶  $h_2(c) = (\lfloor c/11 \rfloor) \bmod 11$

Campos de apontadores são introduzidos para o atendimento a cada tipo de consulta por chave secundária:

	Id	Nome	Depto	Pt Depto	Função	Pt Função	Idade
0				λ		λ	
1				λ		λ	
2				λ		λ	
3				λ		λ	
4				λ		λ	
5				λ		λ	
6				λ		λ	
7				λ		λ	
8				λ		λ	
9				λ		λ	
10				λ		λ	

# Armazenamento do Início das Listas

## Uso de arquivos de hashing auxiliares

Consideremos (**por exemplo!**) arquivos usando **hashing duplo** de tamanho 5 e com as seguintes funções **hipotéticas**:

$$h_{1,d}(\text{Financeiro}) = 3, h_{2,d}(\text{Financeiro}) = 2$$

$$h_{1,d}(\text{Pesquisa}) = 2, h_{2,d}(\text{Pesquisa}) = 2$$

$$h_{1,d}(\text{Produção}) = 3, h_{2,d}(\text{Produção}) = 1$$

$$h_{1,f}(\text{Gerente}) = 1, h_{2,f}(\text{Gerente}) = 2$$

$$h_{1,f}(\text{Programador}) = 2, h_{2,f}(\text{Programador}) = 2$$

$$h_{1,f}(\text{Contador}) = 1, h_{2,f}(\text{Contador}) = 1$$

	Depto	Apont
0		$\lambda$
1		$\lambda$
2		$\lambda$
3		$\lambda$
4		$\lambda$

	Função	Apont
0		$\lambda$
1		$\lambda$
2		$\lambda$
3		$\lambda$
4		$\lambda$

# Arquivos Multilista

Inserção do registro:

Id	Nome	Depto	Função	Idade
17	João	Financeiro	Gerente	40

	Id	Nome	Depto	Pt Depto	Função	Pt Função	Idade
0				$\lambda$		$\lambda$	
1				$\lambda$		$\lambda$	
2				$\lambda$		$\lambda$	
3				$\lambda$		$\lambda$	
4				$\lambda$		$\lambda$	
5				$\lambda$		$\lambda$	
6	17	João	Financeiro	$\lambda$	Gerente	$\lambda$	40
7				$\lambda$		$\lambda$	
8				$\lambda$		$\lambda$	
9				$\lambda$		$\lambda$	
10				$\lambda$		$\lambda$	

	Depto	Apont
0		$\lambda$
1		$\lambda$
2		$\lambda$
3	Financeiro	6
4		$\lambda$

	Função	Apont
0		$\lambda$
1	Gerente	6
2		$\lambda$
3		$\lambda$
4		$\lambda$

# Arquivos Multilista

Inserção do registro:

Id	Nome	Depto	Função	Idade
30	Maria	Pesquisa	Gerente	30

	Id	Nome	Depto	Pt Depto	Função	Pt Função	Idade
0				λ		λ	
1				λ		λ	
2				λ		λ	
3				λ		λ	
4				λ		λ	
5				λ		λ	
6	17	João	Financeiro	λ	Gerente	8	40
7				λ		λ	
8	30	Maria	Pesquisa	λ	Gerente	λ	30
9				λ		λ	
10				λ		λ	

	Depto	Apont
0		λ
1		λ
2	Pesquisa	8
3	Financeiro	6
4		λ

	Função	Apont
0		λ
1	Gerente	6
2		λ
3		λ
4		λ

# Arquivos Multilista

Inserção do registro:

Id	Nome	Depto	Função	Idade
28	Carlos	Pesquisa	Programador	35

	Id	Nome	Depto	Pt Depto	Função	Pt Função	Idade
0				$\lambda$		$\lambda$	
1				$\lambda$		$\lambda$	
2				$\lambda$		$\lambda$	
3				$\lambda$		$\lambda$	
4				$\lambda$		$\lambda$	
5				$\lambda$		$\lambda$	
6	17	João	Financeiro	$\lambda$	Gerente	8	40
7				$\lambda$		$\lambda$	
8	30	Maria	Pesquisa	10	Gerente	$\lambda$	30
9				$\lambda$		$\lambda$	
10	28	Carlos	Pesquisa	$\lambda$	Programador	$\lambda$	35

	Depto	Apont
0		$\lambda$
1		$\lambda$
2	Pesquisa	8
3	Financeiro	6
4		$\lambda$

	Função	Apont
0		$\lambda$
1	Gerente	6
2	Programador	10
3		$\lambda$
4		$\lambda$

# Arquivos Multilista

Inserção do registro:

Id	Nome	Depto	Função	Idade
19	Ana	Financeiro	Contador	25

	Id	Nome	Depto	Pt Depto	Função	Pt Função	Idade
0				$\lambda$		$\lambda$	
1				$\lambda$		$\lambda$	
2				$\lambda$		$\lambda$	
3				$\lambda$		$\lambda$	
4				$\lambda$		$\lambda$	
5				$\lambda$		$\lambda$	
6	17	João	Financeiro	9	Gerente	8	40
7				$\lambda$		$\lambda$	
8	30	Maria	Pesquisa	10	Gerente	$\lambda$	30
9	19	Ana	Financeiro	$\lambda$	Contador	$\lambda$	25
10	28	Carlos	Pesquisa	$\lambda$	Programador	$\lambda$	35

	Depto	Apont
0		$\lambda$
1		$\lambda$
2	Pesquisa	8
3	Financeiro	6
4		$\lambda$

	Função	Apont
0		$\lambda$
1	Gerente	6
2	Programador	10
3	Contador	9
4		$\lambda$



# Arquivos Multilista

Inserção do registro:

Id	Nome	Depto	Função	Idade
42	Paulo	Financeiro	Contador	30

	Id	Nome	Depto	Pt Depto	Função	Pt Função	Idade
0				$\lambda$		$\lambda$	
1	42	Paulo	Financeiro	$\lambda$	Contador	$\lambda$	30
2				$\lambda$		$\lambda$	
3				$\lambda$		$\lambda$	
4				$\lambda$		$\lambda$	
5				$\lambda$		$\lambda$	
6	17	João	Financeiro	9	Gerente	8	40
7				$\lambda$		$\lambda$	
8	30	Maria	Pesquisa	10	Gerente	$\lambda$	30
9	19	Ana	Financeiro	1	Contador	1	25
10	28	Carlos	Pesquisa	$\lambda$	Programador	$\lambda$	35

	Depto	Apont
0		$\lambda$
1		$\lambda$
2	Pesquisa	8
3	Financeiro	6
4		$\lambda$

	Função	Apont
0		$\lambda$
1	Gerente	6
2	Programador	10
3	Contador	9
4		$\lambda$

# Arquivos Multilista

Inserção do registro:

Id	Nome	Depto	Função	Idade
18	Joana	Pesquisa	Programador	32

	Id	Nome	Depto	Pt Depto	Função	Pt Função	Idade
0				$\lambda$		$\lambda$	
1	42	Paulo	Financeiro	$\lambda$	Contador	$\lambda$	30
2				$\lambda$		$\lambda$	
3				$\lambda$		$\lambda$	
4				$\lambda$		$\lambda$	
5				$\lambda$		$\lambda$	
6	17	João	Financeiro	9	Gerente	8	40
7	18	Joana	Pesquisa	$\lambda$	Programador	$\lambda$	32
8	30	Maria	Pesquisa	10	Gerente	$\lambda$	30
9	19	Ana	Financeiro	1	Contador	1	25
10	28	Carlos	Pesquisa	7	Programador	7	35

	Depto	Apont
0		$\lambda$
1		$\lambda$
2	Pesquisa	8
3	Financeiro	6
4		$\lambda$

	Função	Apont
0		$\lambda$
1	Gerente	6
2	Programador	10
3	Contador	9
4		$\lambda$

# Arquivos Multilista

Inserção do registro:

Id	Nome	Depto	Função	Idade
20	José	Produção	Gerente	45

	Id	Nome	Depto	Pt Depto	Função	Pt Função	Idade
0	20	José	Produção	$\lambda$	Gerente	$\lambda$	45
1	42	Paulo	Financeiro	$\lambda$	Contador	$\lambda$	30
2				$\lambda$		$\lambda$	
3				$\lambda$		$\lambda$	
4				$\lambda$		$\lambda$	
5				$\lambda$		$\lambda$	
6	17	João	Financeiro	9	Gerente	8	40
7	18	Joana	Pesquisa	$\lambda$	Programador	$\lambda$	32
8	30	Maria	Pesquisa	10	Gerente	0	30
9	19	Ana	Financeiro	1	Contador	1	25
10	28	Carlos	Pesquisa	7	Programador	7	35

	Depto	Apont
0		$\lambda$
1		$\lambda$
2	Pesquisa	8
3	Financeiro	6
4	Produção	0

	Função	Apont
0		$\lambda$
1	Gerente	6
2	Programador	10
3	Contador	9
4		$\lambda$

## Características

- ▶ Facilitam operações para chaves secundárias

## Características

- ▶ Facilitam operações para chaves secundárias
- ▶ Exigem planejar o arquivo principal antecipadamente

## Características

- ▶ Facilitam operações para chaves secundárias
- ▶ Exigem planejar o arquivo principal antecipadamente
- ▶ Utilizam espaço extra para ponteiros

## Características

- ▶ Facilitam operações para chaves secundárias
- ▶ Exigem planejar o arquivo principal antecipadamente
- ▶ Utilizam espaço extra para ponteiros
- ▶ Requerem manutenção de consistência entre os arquivos auxiliares e o arquivo principal:

## Características

- ▶ Facilitam operações para chaves secundárias
- ▶ Exigem planejar o arquivo principal antecipadamente
- ▶ Utilizam espaço extra para ponteiros
- ▶ Requerem manutenção de consistência entre os arquivos auxiliares e o arquivo principal:
  - ▶ Movimentação de registros no arquivo principal?



## Características

- ▶ Facilitam operações para chaves secundárias
- ▶ Exigem planejar o arquivo principal antecipadamente
- ▶ Utilizam espaço extra para ponteiros
- ▶ Requerem manutenção de consistência entre os arquivos auxiliares e o arquivo principal:
  - ▶ Movimentação de registros no arquivo principal?
  - ▶ Uso de chaves primária vs. ponteiros?

## Características

- ▶ Facilitam operações para chaves secundárias
- ▶ Exigem planejar o arquivo principal antecipadamente
- ▶ Utilizam espaço extra para ponteiros
- ▶ Requerem manutenção de consistência entre os arquivos auxiliares e o arquivo principal:
  - ▶ Movimentação de registros no arquivo principal?
  - ▶ Uso de chaves primária vs. ponteiros?
  - ▶ Modificação do valor das chaves secundárias?

# Uso de Chaves ao invés de Índices

	<b>Id</b>	<b>Nome</b>	<b>Depto</b>	<b>Pt Depto</b>	<b>Função</b>	<b>Pt Função</b>	<b>Idade</b>
0	20	José	Produção	$\lambda$	Gerente	$\lambda$	45
1	42	Paulo	Financeiro	$\lambda$	Contador	$\lambda$	30
2				$\lambda$		$\lambda$	
3				$\lambda$		$\lambda$	
4				$\lambda$		$\lambda$	
5				$\lambda$		$\lambda$	
6	17	João	Financeiro	9	Gerente	8	40
7	18	Joana	Pesquisa	$\lambda$	Programador	$\lambda$	32
8	30	Maria	Pesquisa	10	Gerente	0	30
9	19	Ana	Financeiro	1	Contador	1	25
10	28	Carlos	Pesquisa	7	Programador	7	35

	<b>Depto</b>	<b>Chave</b>
0		$\lambda$
1		$\lambda$
2	Pesquisa	<b>30</b>
3	Financeiro	<b>17</b>
4	Produção	<b>20</b>

	<b>Função</b>	<b>Chave</b>
0		$\lambda$
1	Gerente	<b>17</b>
2	Programador	<b>28</b>
3	Contador	<b>19</b>
4		$\lambda$

# Arquivos Invertidos

## Caracterização

- ▶ Para cada campo (chave secundária) de interesse, há um arquivo, chamado de **invertido**. Cada valor do campo passa a ser índice (ou chave primária) no arquivo invertido.

Arquivo invertido para o campo  $i$

Valor 1	ptr 11	ptr 12	...	ptr 1 $k_1$
Valor 2	ptr 21	ptr 22	...	ptr 2 $k_2$
Valor $m$	ptr $m1$	ptr $m2$	...	ptr $mk_m$

- ▶ O número de ponteiros para cada registro é variável, o que exige possíveis listas de índices/chaves no arquivo invertido

# Arquivos Invertidos - Exemplo

Considere novamente o problema apresentado anteriormente:

- ▶ mesmo conjunto de registros
- ▶ interesse em busca pelas mesmas chaves secundárias:  
**departamento** e **função**

Id	Nome	Depto	Função	Idade
17	João	Financeiro	Gerente	40
30	Maria	Pesquisa	Gerente	30
28	Carlos	Pesquisa	Programador	35
19	Ana	Financeiro	Contador	25
42	Paulo	Financeiro	Contador	30
18	Joana	Pesquisa	Programador	32
20	José	Produção	Gerente	45

# Arquivos Invertidos - Exemplo

Supondo (**por exemplo!**):

- ▶ Arquivos invertidos com 5 posições, com **hashing duplo** e com três apontadores por registro
- ▶ As seguintes funções de hashing **hipotéticas**:

$h_{1,d}(\text{Financeiro}) = 3$ ,  $h_{2,d}(\text{Financeiro}) = 2$   
 $h_{1,d}(\text{Pesquisa}) = 2$ ,  $h_{2,d}(\text{Pesquisa}) = 2$   
 $h_{1,d}(\text{Produção}) = 3$ ,  $h_{2,d}(\text{Produção}) = 1$

$h_{1,f}(\text{Gerente}) = 1$ ,  $h_{2,f}(\text{Gerente}) = 2$   
 $h_{1,f}(\text{Programador}) = 2$ ,  $h_{2,f}(\text{Programador}) = 2$   
 $h_{1,f}(\text{Contador}) = 1$ ,  $h_{2,f}(\text{Contador}) = 1$

	Depto	Aponts		
0		$\lambda$	$\lambda$	$\lambda$
1		$\lambda$	$\lambda$	$\lambda$
2		$\lambda$	$\lambda$	$\lambda$
3		$\lambda$	$\lambda$	$\lambda$
4		$\lambda$	$\lambda$	$\lambda$

	Função	Aponts		
0		$\lambda$	$\lambda$	$\lambda$
1		$\lambda$	$\lambda$	$\lambda$
2		$\lambda$	$\lambda$	$\lambda$
3		$\lambda$	$\lambda$	$\lambda$
4		$\lambda$	$\lambda$	$\lambda$

# Arquivos Invertidos

Inserção do registro:

Id	Nome	Depto	Função	Idade
17	João	Financeiro	Gerente	40

	Id	Nome	Depto	Função	Idade
0					
1					
2					
3					
4					
5					
6	17	João	Financeiro	Gerente	40
7					
8					
9					
10					

	Depto	Aponts		
0		λ	λ	λ
1		λ	λ	λ
2		λ	λ	λ
3	Financeiro	6	λ	λ
4		λ	λ	λ

	Função	Aponts		
0		λ	λ	λ
1	Gerente	6	λ	λ
2		λ	λ	λ
3		λ	λ	λ
4		λ	λ	λ



# Arquivos Invertidos

Inserção do registro:

Id	Nome	Depto	Função	Idade
30	Maria	Pesquisa	Gerente	30

	Id	Nome	Depto	Função	Idade
0					
1					
2					
3					
4					
5					
6	17	João	Financeiro	Gerente	40
7					
8	30	Maria	Pesquisa	Gerente	30
9					
10					

	Depto	Aponts		
0		λ	λ	λ
1		λ	λ	λ
2	Pesquisa	8	λ	λ
3	Financeiro	6	λ	λ
4		λ	λ	λ

	Função	Aponts		
0		λ	λ	λ
1	Gerente	6	8	λ
2		λ	λ	λ
3		λ	λ	λ
4		λ	λ	λ

# Arquivos Invertidos

Inserção do registro:

Id	Nome	Depto	Função	Idade
28	Carlos	Pesquisa	Programador	35

	Id	Nome	Depto	Função	Idade
0					
1					
2					
3					
4					
5					
6	17	João	Financeiro	Gerente	40
7					
8	30	Maria	Pesquisa	Gerente	30
9					
10	28	Carlos	Pesquisa	Programador	35

	Depto	Aponts		
0		λ	λ	λ
1		λ	λ	λ
2	Pesquisa	8	10	λ
3	Financeiro	6	λ	λ
4		λ	λ	λ

	Função	Aponts		
0		λ	λ	λ
1	Gerente	6	8	λ
2	Programador	10	λ	λ
3		λ	λ	λ
4		λ	λ	λ

# Arquivos Invertidos

Inserção do registro:

Id	Nome	Depto	Função	Idade
19	Ana	Financeiro	Contador	25

	Id	Nome	Depto	Função	Idade
0					
1					
2					
3					
4					
5					
6	17	João	Financeiro	Gerente	40
7					
8	30	Maria	Pesquisa	Gerente	30
9	19	Ana	Financeiro	Contador	25
10	28	Carlos	Pesquisa	Programador	35

	Depto	Aponts		
0		λ	λ	λ
1		λ	λ	λ
2	Pesquisa	8	10	λ
3	Financeiro	6	9	λ
4		λ	λ	λ

	Função	Aponts		
0		λ	λ	λ
1	Gerente	6	8	λ
2	Programador	10	λ	λ
3	Contador	9	λ	λ
4		λ	λ	λ

# Arquivos Invertidos

Inserção do registro:

Id	Nome	Depto	Função	Idade
42	Paulo	Financeiro	Contador	30

	Id	Nome	Depto	Função	Idade
0					
1	42	Paulo	Financeiro	Contador	30
2					
3					
4					
5					
6	17	João	Financeiro	Gerente	40
7					
8	30	Maria	Pesquisa	Gerente	30
9	19	Ana	Financeiro	Contador	25
10	28	Carlos	Pesquisa	Programador	35

	Depto	Aponts		
0		λ	λ	λ
1		λ	λ	λ
2	Pesquisa	8	10	λ
3	Financeiro	6	9	1
4		λ	λ	λ

	Função	Aponts		
0		λ	λ	λ
1	Gerente	6	8	λ
2	Programador	10	λ	λ
3	Contador	9	1	λ
4		λ	λ	λ

# Arquivos Invertidos

Inserção do registro:

Id	Nome	Depto	Função	Idade
18	Joana	Pesquisa	Programador	32

	Id	Nome	Depto	Função	Idade
0					
1	42	Paulo	Financeiro	Contador	30
2					
3					
4					
5					
6	17	João	Financeiro	Gerente	40
7	18	Joana	Pesquisa	Programador	32
8	30	Maria	Pesquisa	Gerente	30
9	19	Ana	Financeiro	Contador	25
10	28	Carlos	Pesquisa	Programador	35

	Depto	Aponts		
0		λ	λ	λ
1		λ	λ	λ
2	Pesquisa	8	10	7
3	Financeiro	6	9	1
4		λ	λ	λ

	Função	Aponts		
0		λ	λ	λ
1	Gerente	6	8	λ
2	Programador	10	7	λ
3	Contador	9	1	λ
4		λ	λ	λ

# Arquivos Invertidos

Inserção do registro:

Id	Nome	Depto	Função	Idade
20	José	Produção	Gerente	45

	Id	Nome	Depto	Função	Idade
0	20	José	Produção	Gerente	45
1	42	Paulo	Financeiro	Contador	30
2					
3					
4					
5					
6	17	João	Financeiro	Gerente	40
7	18	Joana	Pesquisa	Programador	32
8	30	Maria	Pesquisa	Gerente	30
9	19	Ana	Financeiro	Contador	25
10	28	Carlos	Pesquisa	Programador	35

	Depto	Aponts		
0		λ	λ	λ
1		λ	λ	λ
2	Pesquisa	8	10	7
3	Financeiro	6	9	1
4	Produção	0	λ	λ

	Função	Aponts		
0		λ	λ	λ
1	Gerente	6	8	0
2	Programador	10	7	λ
3	Contador	9	1	λ
4		λ	λ	λ

- ▶ Arquivos invertidos facilitam operações para chaves secundárias

# Arquivos Invertidos

- ▶ Arquivos invertidos facilitam operações para chaves secundárias
- ▶ O arquivo principal independente dos arquivos invertidos, que podem ser criados *a posteriori*



# Arquivos Invertidos

- ▶ Arquivos invertidos facilitam operações para chaves secundárias
- ▶ O arquivo principal independente dos arquivos invertidos, que podem ser criados *a posteriori*
- ▶ Manutenção da consistência com arquivo principal:
  - ▶ Movimentação de registros no arquivo principal?

# Arquivos Invertidos

- ▶ Arquivos invertidos facilitam operações para chaves secundárias
- ▶ O arquivo principal independente dos arquivos invertidos, que podem ser criados *a posteriori*
- ▶ Manutenção da consistência com arquivo principal:
  - ▶ Movimentação de registros no arquivo principal?
  - ▶ Uso de chave primária vs. ponteiros?

# Arquivos Invertidos

- ▶ Arquivos invertidos facilitam operações para chaves secundárias
- ▶ O arquivo principal independente dos arquivos invertidos, que podem ser criados *a posteriori*
- ▶ Manutenção da consistência com arquivo principal:
  - ▶ Movimentação de registros no arquivo principal?
  - ▶ Uso de chave primária vs. ponteiros?
  - ▶ Modificação do valor das chaves secundárias?

# Árvores k-d

- ▶ Consulta por faixa de valores de mais do que uma variável. Ex.:
  - ▶ Quais funcionários possuem salários maiores do que  $s$ ?
  - ▶ Quais são as pessoas que possuem idade entre  $a$  e  $b$ , peso entre  $x$  e  $y$  e altura entre  $r$  e  $s$ ?
- ▶ Árvores  $k$ - $d$  proporcionam uma forma de estruturar dados em  $k$  dimensões: **árvore 3-d (três dimensões), 4-d (quatro dimensões)**, etc.

- ▶ Árvore binária de busca adaptada para tratar faixa de valores (das chaves secundárias)
- ▶ Cada chave secundária representa uma dimensão
- ▶ Cada nível da árvore está associado a apenas uma dimensão
- ▶ Cada nó da árvore particiona a faixa de valores associada a sua dimensão em  $\leq$  e  $>$
- ▶ Árvore provê estrutura de índices. Os dados (ou ponteiros para os mesmos) são armazenados em páginas externas

# Ilustração de Árvores $k$ - $d$

Construção de uma árvore 2- $d$ , usando os seguintes registros como representativos para o índice.

Peso e altura são as chaves secundárias de interesse.

Nome	Peso (kg)	Altura (cm)
José	50	63
Maria	74	140
Joana	62	114
Alberto	55	145
Manoel	98	146
Daniel	48	120
Karen	35	91

# Ilustração: Construção do Índice

Inserção de:

Nome	Peso (kg)	Altura (cm)
José	50	63

Árvore 2-d:

Dimensão 1 (peso) → José, 50, 63



# Ilustração: Construção do Índice

Inserção de:

Nome	Peso (kg)	Altura (cm)
Maria	74	140

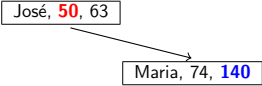
Árvore 2-d:

Dimensão 1 (peso)

José, 50, 63

Dimensão 2 (altura)

Maria, 74, 140



# Ilustração: Construção do Índice

Inserção de:

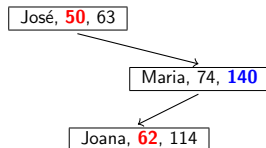
Nome	Peso (kg)	Altura (cm)
Joana	62	114

Árvore 2-d:

Dimensão 1 (peso)

Dimensão 2 (altura)

Dimensão 1 (peso)



# Ilustração: Construção do Índice

Inserção de:

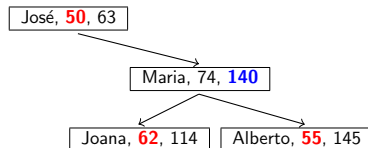
Nome	Peso (kg)	Altura (cm)
Alberto	55	145

Árvore 2-d:

Dimensão 1 (peso)

Dimensão 2 (altura)

Dimensão 1 (peso)



# Ilustração: Construção do Índice

Inserção de:

Nome	Peso (kg)	Altura (cm)
Manoel	98	146

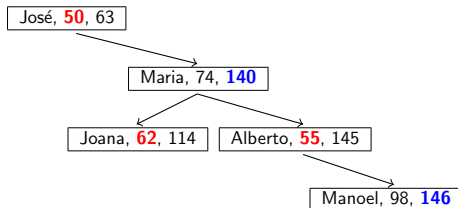
Árvore 2-d:

Dimensão 1 (peso)

Dimensão 2 (altura)

Dimensão 1 (peso)

Dimensão 2 (altura)



# Ilustração: Construção do índice

Inserção de:

Nome	Peso (kg)	Altura (cm)
Daniel	48	120

Árvore 2-d:

Dimensão 1 (peso)

José, 50, 63

Dimensão 2 (altura)

Daniel, 48, 120

Maria, 74, 140

Dimensão 1 (peso)

Joana, 62, 114

Alberto, 55, 145

Dimensão 2 (altura)

Manoel, 98, 146

# Ilustração: Construção do Índice

Inserção de:

Nome	Peso (kg)	Altura (cm)
Karen	35	91

Árvore 2-d:

Dimensão 1 (peso)

José, 50, 63

Dimensão 2 (altura)

Daniel, 48, 120

Maria, 74, 140

Dimensão 1 (peso)

Karen, 35, 91

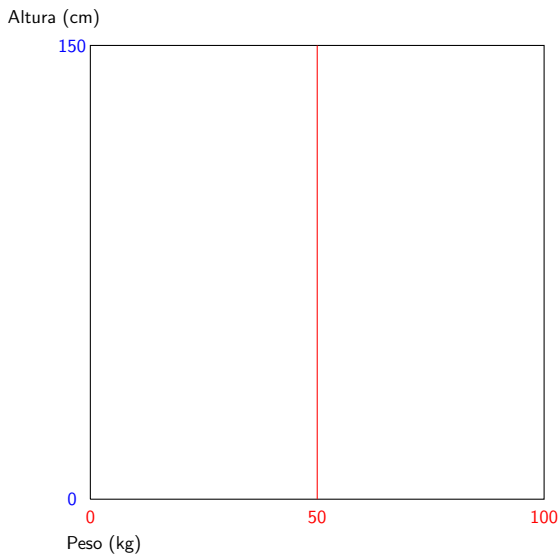
Joana, 62, 114

Alberto, 55, 145

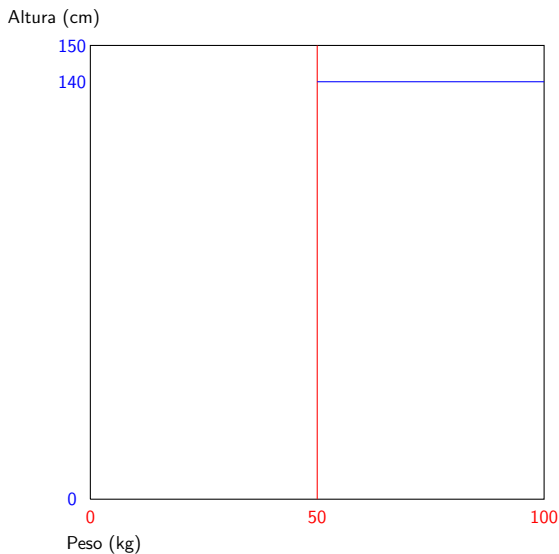
Dimensão 2 (altura)

Manoel, 98, 146

# Ilustração: Construção do índice

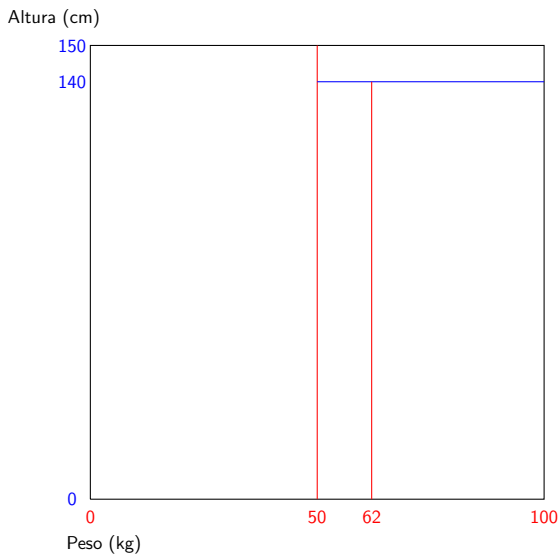


# Ilustração: Construção do índice

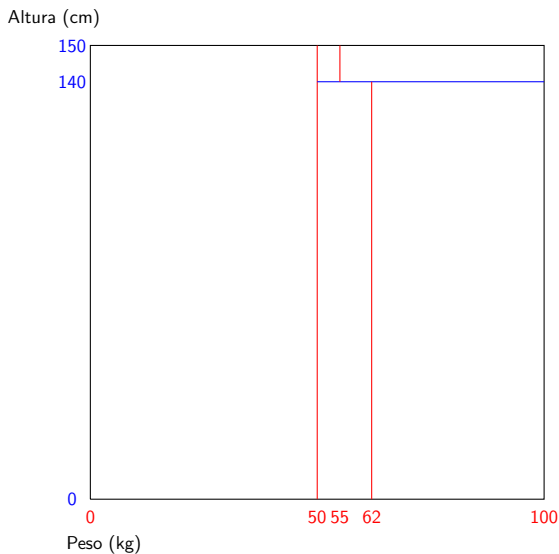




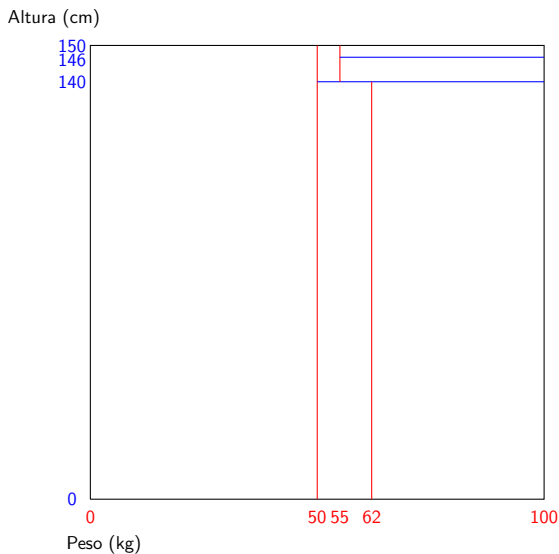
# Ilustração: Construção do índice



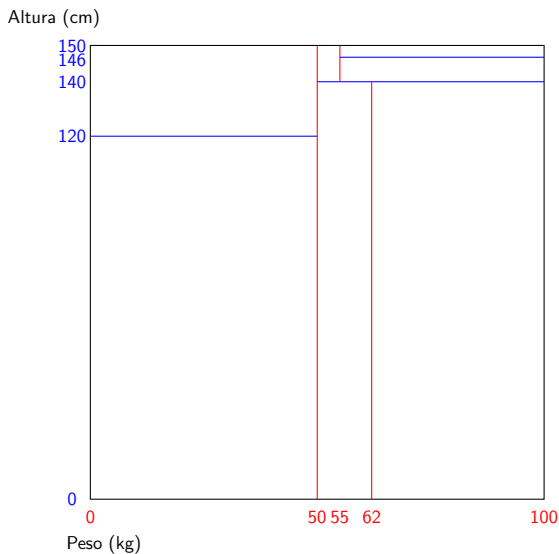
# Ilustração: Construção do índice



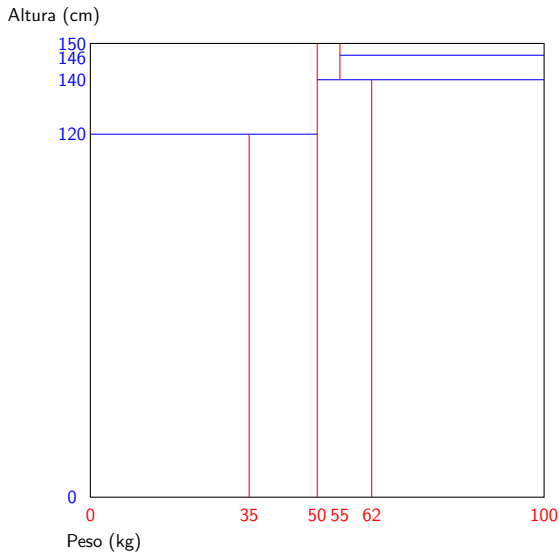
# Ilustração: Construção do índice



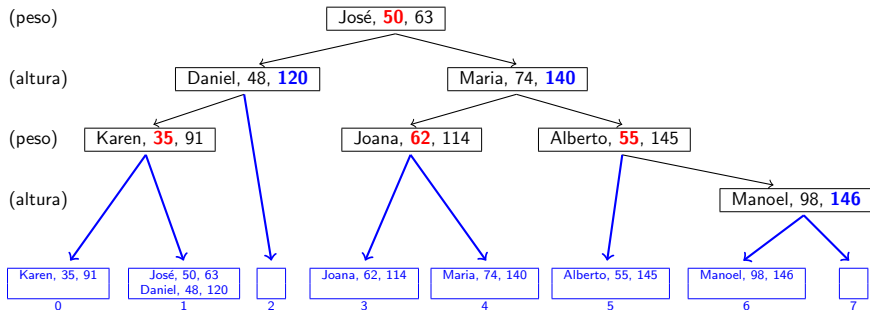
# Ilustração: Construção do índice



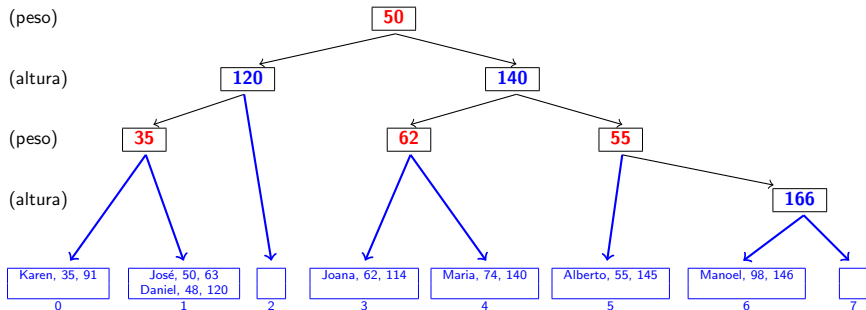
# Ilustração: Construção do índice



# Ilustração: Páginas



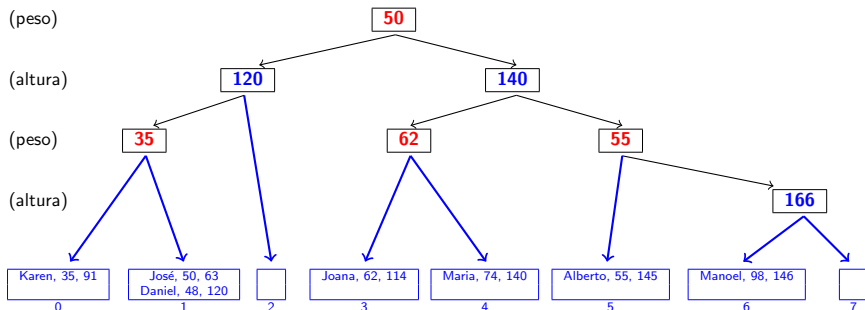
# Ilustração: Armazenamento apenas do Valor da Dimensão



# Exemplo de Consultas

Quais são as pessoas com:

- ▶ peso  $p$  maior ou igual a 30 kg e menor ou igual a 55 kg:  $30 \leq p \leq 55$
- ▶ altura  $h$  maior do que 90 cm:  $h > 90$

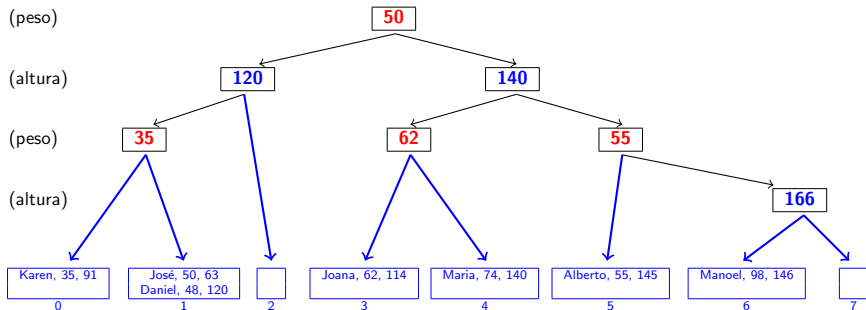




# Exemplo de Consultas

Quais são as pessoas com:

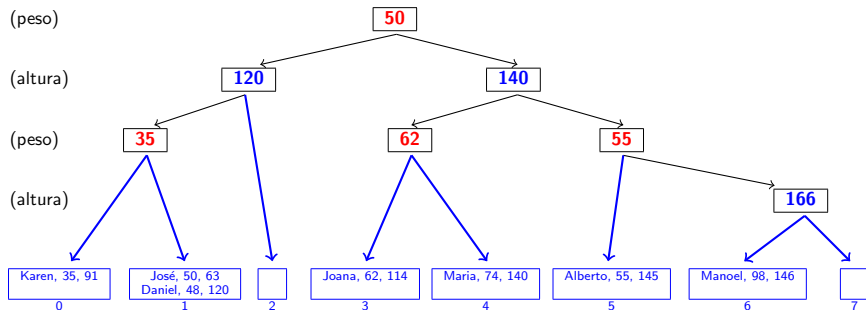
- ▶ peso  $p$  maior ou igual a 60 kg:  $p \geq 60$
- ▶ altura  $h$  maior do que 90 cm:  $h > 90$



# Exemplo de Consultas

Quais são os possíveis pesos e alturas dos registros que podem ser armazenados na página 1?

E na página 3?



- ▶ Árvores  $k$ - $d$  representam particionamentos por faixa de valores em função dos valores de interesse

- ▶ Árvores  $k$ - $d$  representam particionamentos por faixa de valores em função dos valores de interesse
- ▶ Árvore não necessariamente balanceada
  - ▶ Construção da árvore depende da escolha dos valores representativos
  - ▶ Valores adequados podem ser escolhidos para a construção do índice, independentemente dos registros armazenados

# Árvore $k-d$

- ▶ Árvores  $k-d$  representam particionamentos por faixa de valores em função dos valores de interesse
- ▶ Árvore não necessariamente balanceada
  - ▶ Construção da árvore depende da escolha dos valores representativos
  - ▶ Valores adequados podem ser escolhidos para a construção do índice, independentemente dos registros armazenados
- ▶ Manutenção da consistência com arquivo principal se página de dados não armazena os registros (e sim ponteiros para eles)

# Árvore $k-d$

- ▶ Árvores  $k-d$  representam particionamentos por faixa de valores em função dos valores de interesse
- ▶ Árvore não necessariamente balanceada
  - ▶ Construção da árvore depende da escolha dos valores representativos
  - ▶ Valores adequados podem ser escolhidos para a construção do índice, independentemente dos registros armazenados
- ▶ Manutenção da consistência com arquivo principal se página de dados não armazena os registros (e sim ponteiros para eles)
- ▶ Índice pode crescer sob demanda ou ser previamente projetado (caso se saiba dimensionar os dados alvo)