

Organização e Recuperação da Informação

Organização de dados em memória secundária

DC – UFSCar
Jander Moreira

Conteúdo

- Agenda
 - Organização em memória secundária
 - Arquivos indexados
 - Índices primários
 - Índices secundários

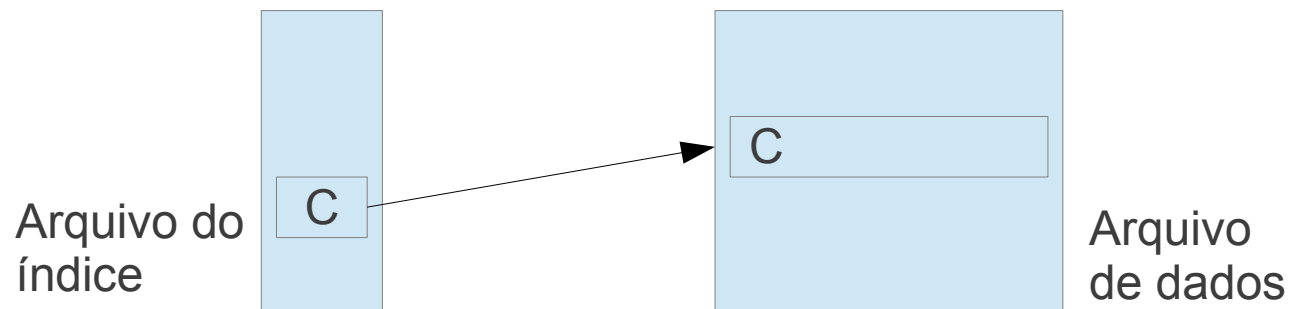
Índices

- Índice

- Estrutura auxiliar cujo uso permite o acesso a uma dada informação de forma "rápida"

- Organização básica

- Consiste em um arquivo auxiliar, associado ao arquivo de dados principal
 - Contém valores de chaves de interesse (chave de pesquisa) e ponteiros para os registros ou para os blocos que os contém
 - As chaves do índice são mantidas de forma ordenada para tornar a pesquisa delas eficiente



Índices

- Índices
 - Densos
 - Cada chave do arquivo de dados tem sua representação como uma entrada no índice
 - Relação 1:1
 - Esparso
 - Cada entrada do índice é associada a um conjunto de registros
 - Relação 1:n

Arquivos indexados

- Indexação
 - Índices primários
 - Índices secundários

Arquivos indexados

- Índices primários

- Arquivo de dados ordenado (arquivo sequencial) pela chave de pesquisa usada no índice

The diagram illustrates a primary index structure. It consists of two tables. The first table, on the left, is an index table with 10 rows. Each row contains a key value in the first column and a pointer (represented by a horizontal line) in the second column. The second table, on the right, is a data table with 10 rows and 8 columns. Each row contains the same key value in the first column, followed by 7 empty cells. Arrows point from the pointer column of the index table to the first column of the data table, indicating that the index points to the location of the data row.

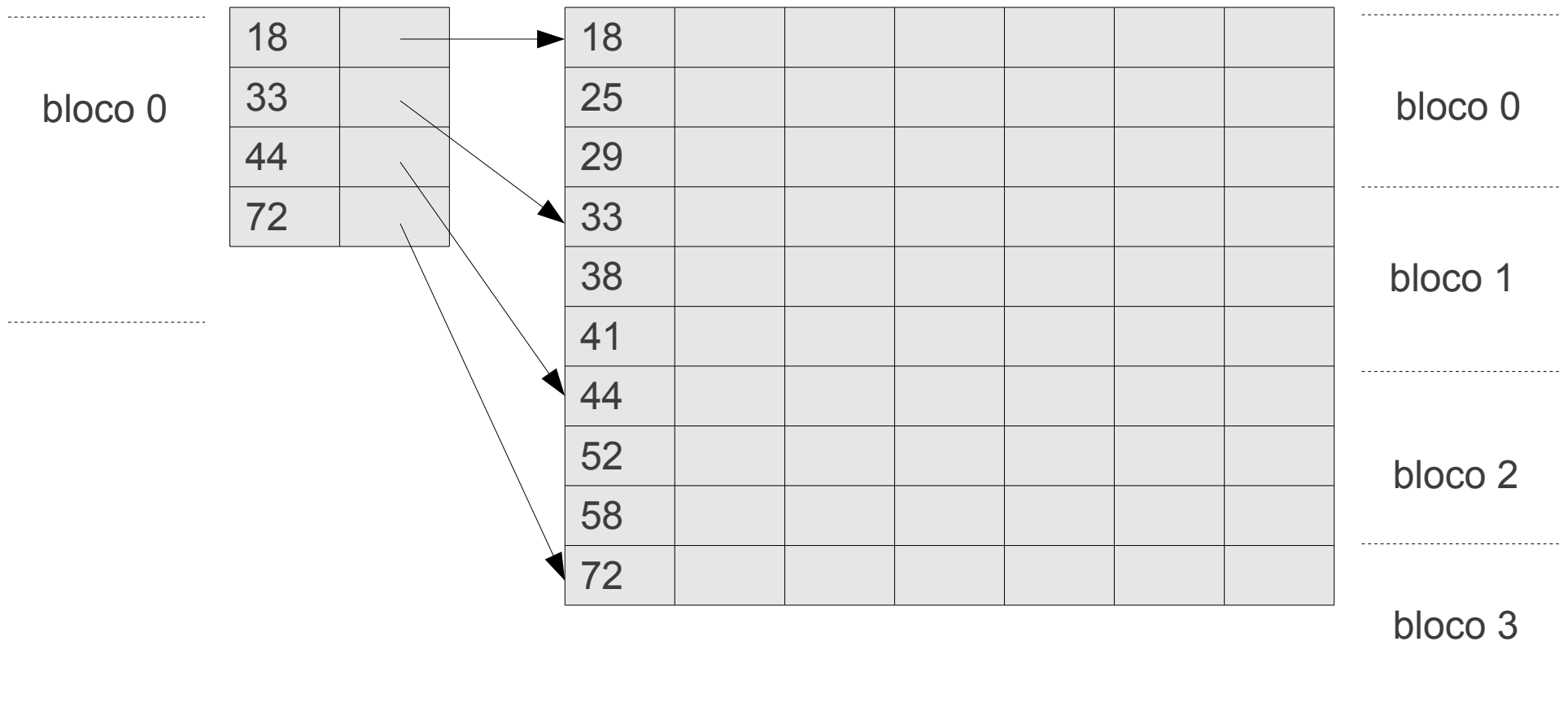
18	—	→	18						
25	—	→	25						
29	—	→	29						
33	—	→	33						
38	—	→	38						
41	—	→	41						
44	—	→	44						
52	—	→	52						
58	—	→	58						
72	—	→	72						

Arquivos indexados

- Índices primários
 - Registros nos arquivos
 - Cada registro de índice contém $\langle chave, ponteiro \rangle$
 - Cada registro de dados contém os dados completos
 - Espaço
 - Cada bloco de disco pode conter, geralmente, mais registros de índice que registros de blocos

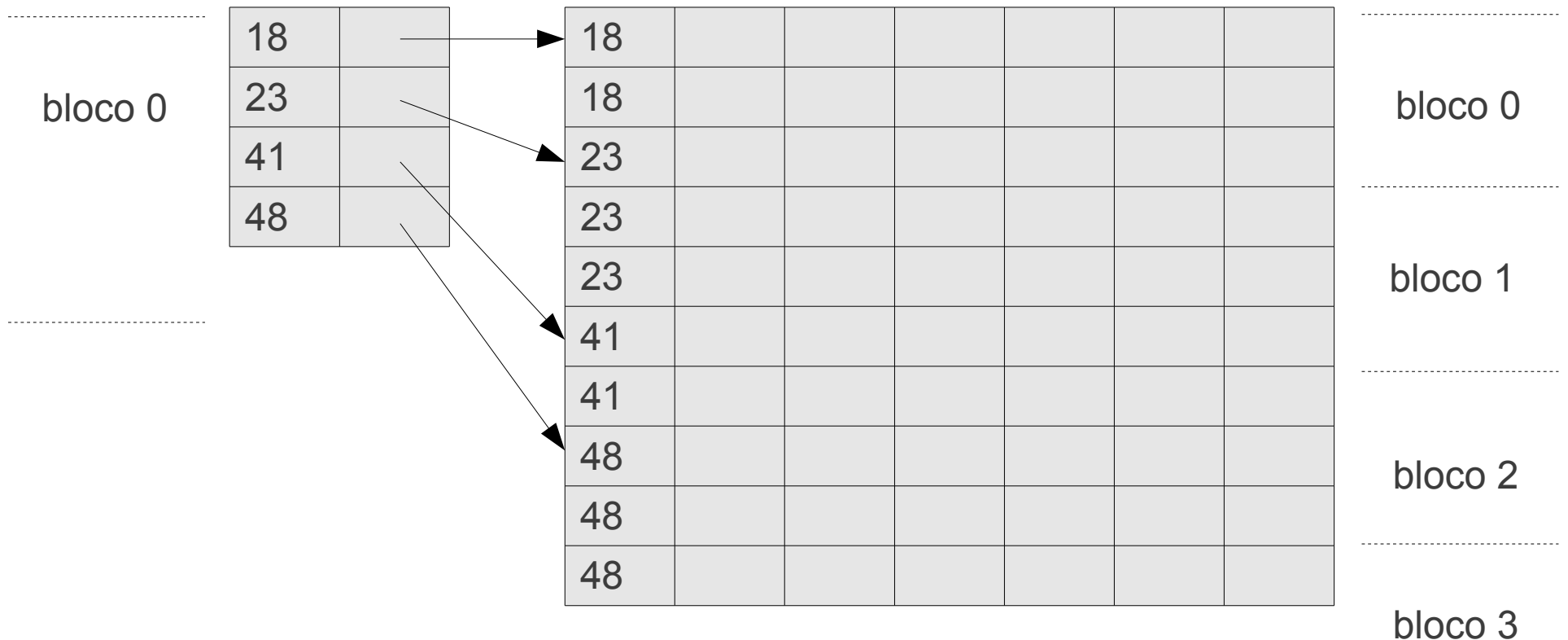
Arquivos indexados

- Índices primários
 - Esparsos



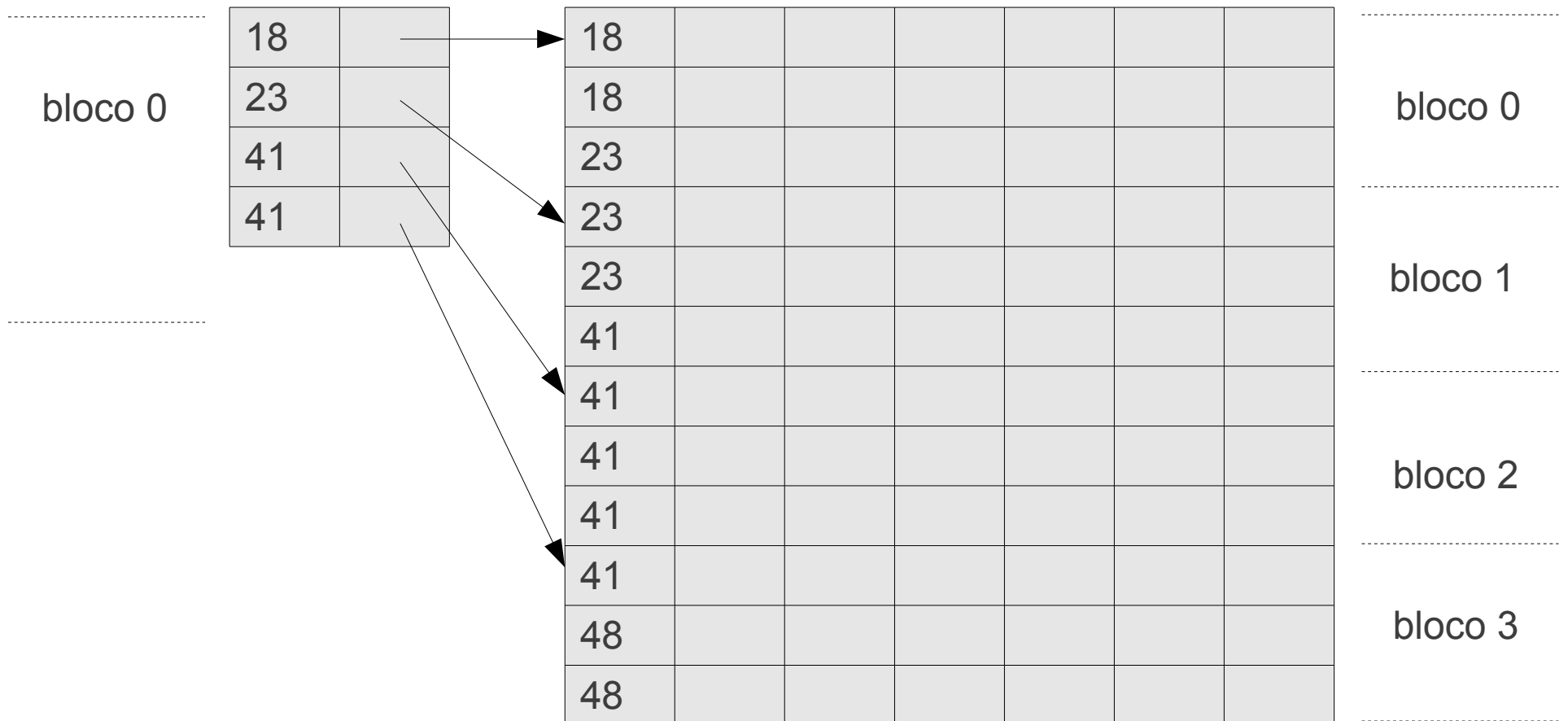
Arquivos indexados

- Índices primários
 - Com repetições de chaves: uma entrada por chave



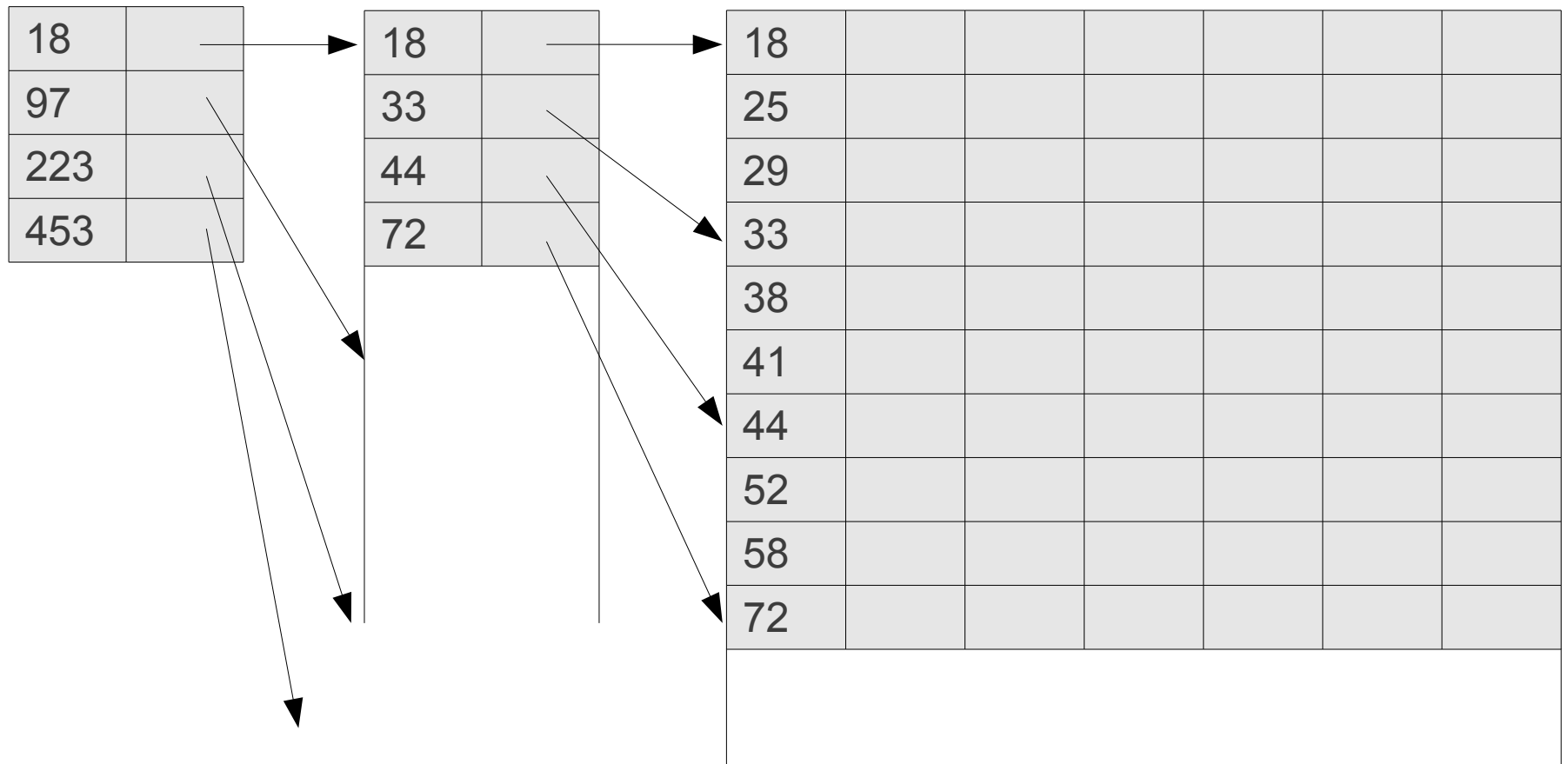
Arquivos indexados

- Índices primários
 - Com repetições de chaves: indicação de bloco



Arquivos indexados

- Índices primários
 - Vários níveis

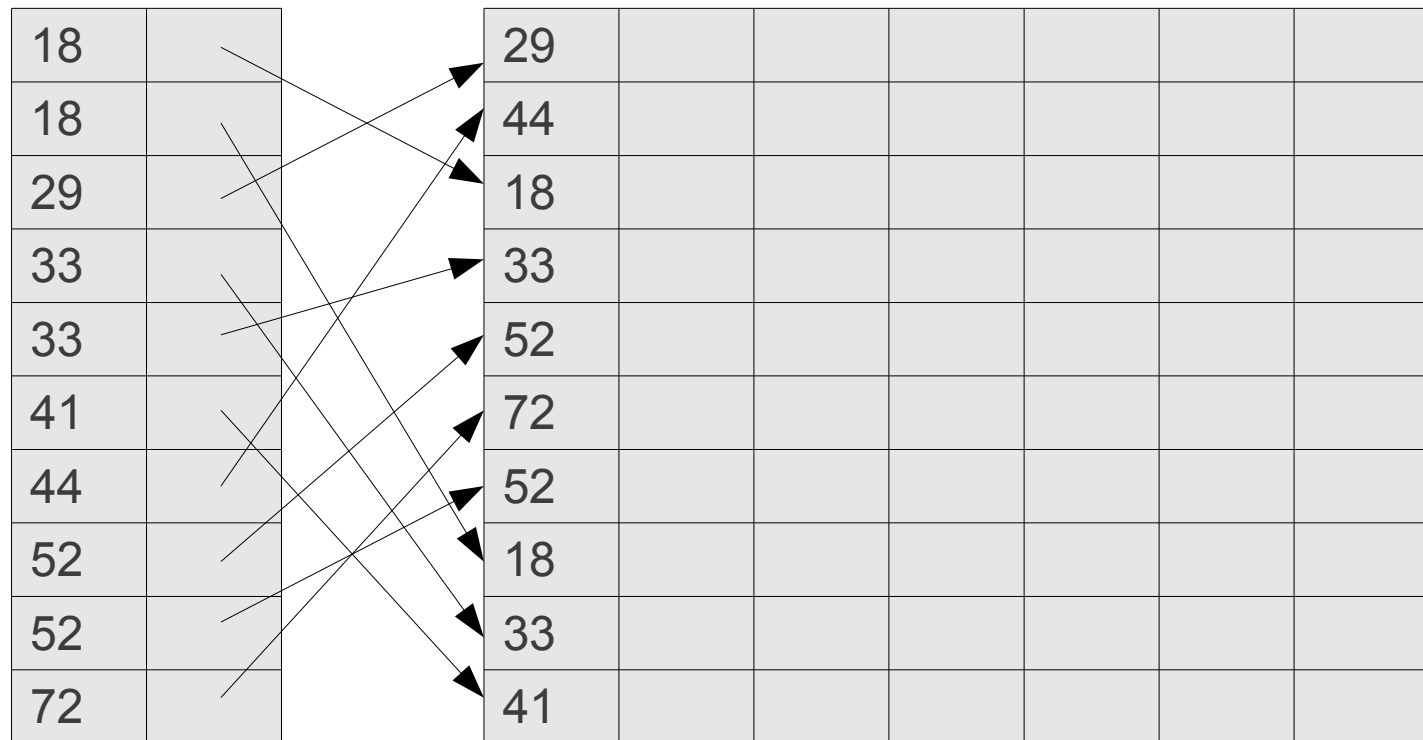


Arquivos indexados

- Índices primários
 - Considerações
 - Modificações no arquivo de dados implicam em modificações no arquivo de índice
 - Minimizar o impacto das modificações é interessante para reduzir os acessos a disco e aumentar o desempenho
 - Aproveitamento do espaço interno em cada bloco
 - Uso de blocos de estouro
 - Remoções lógicas

Arquivos indexados

- Índices secundários
 - Arquivo de dados não ordenado



Arquivos indexados

- Índices secundários
 - Árvores B+

Leitura

- Leituras
 - Moreira, apostila.
 - Garcia-Molina, capítulo 4, seções 4.1 a 4.3