

# Estruturas de Indexação de Dados

Cristina Dutra de Aguiar Ciferri

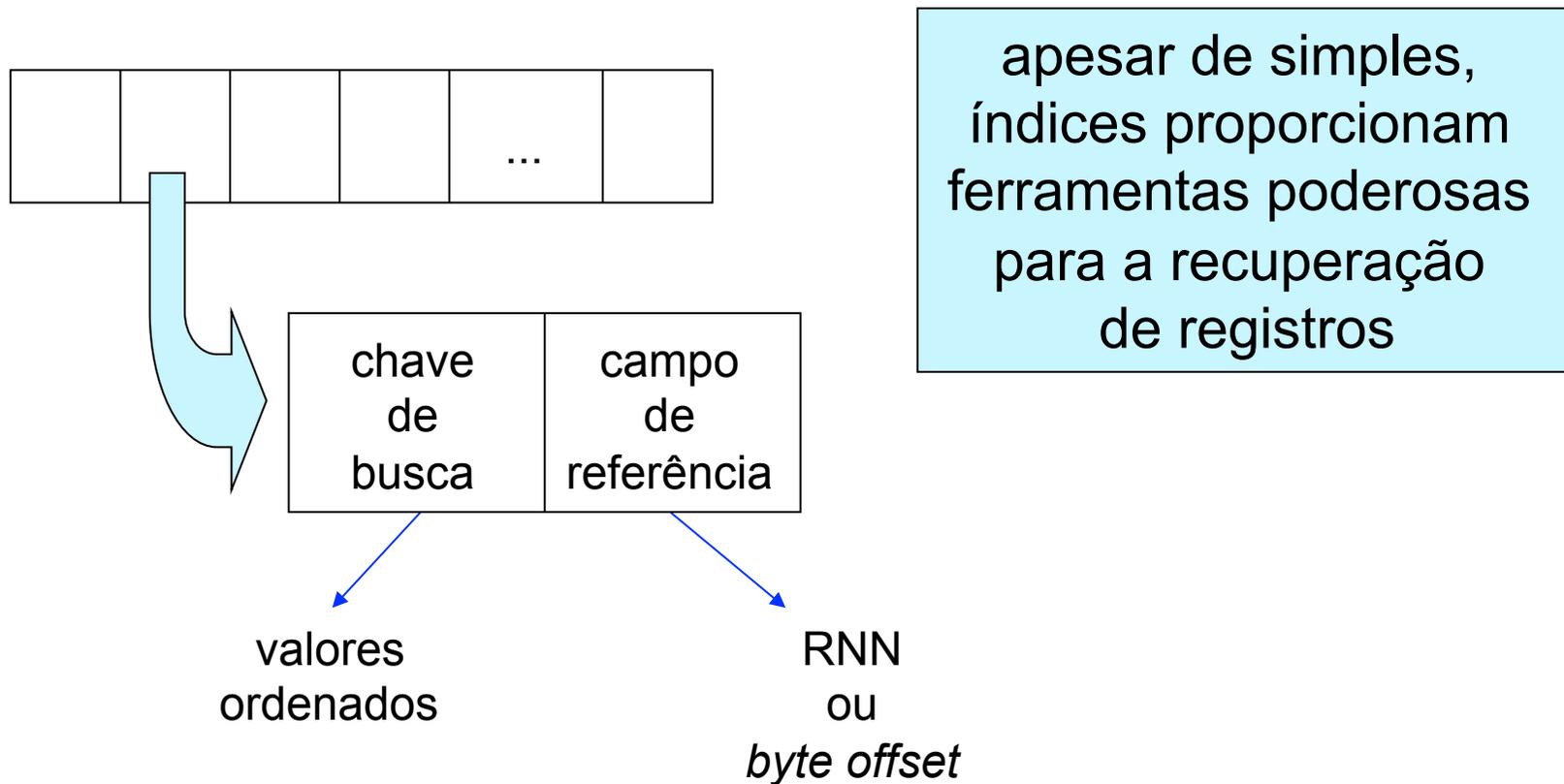
# Índice

- Estrutura de acesso auxiliar usada para melhorar o desempenho na recuperação de registros
- Pesquisa
  - restringida a um subconjunto dos registros, em contrapartida à análise do conjunto completo
  - realizada em resposta a certas condições

# Índice

- Observações
  - existe uma variedade de índices, cada qual com uma estrutura de dados particular
  - qualquer campo em um arquivo pode ser usado para criar um índice
  - vários índices podem ser definidos para um mesmo arquivo

# Índice Simples ou Linear



# Índice Simples ou Linear

ANG3795	167
COL31809	353
COL38358	211
DG139201	396
DG18807	256
FF245	442
LON2312	32
MER75016	300
RCA2626	77
WAR23699	132

*arquivo de índice*

32	LON   2312   Romeo and Juliet   Prokofiev ...
77	RCA   2626   Quartet in C Sharp Minor ...
132	WAR   23699   Touchstone   Corea ...
167	ANG   3795   Symphony No. 9   Beethoven ...
211	COL   38358   Nebraska   Springsteen ...
256	DG   18807   Symphony No. 9   Beethoven ...
300	MER   75016   Coq d'or Suite   Rimsky ...
353	COL   31809   Symphony No. 9   Dvorak ...
396	DG   139201   Violin Concerto   Beethoven ...
442	FF   245   Good News   Sweet Honey In The ...

chave 

*arquivo de dados*

**arquivo auxiliar em disco**

**arquivo armazenado em disco**

# Índice Simples ou Linear

ANG3795	167
COL31809	353
COL38358	211
DG139201	396
DG18807	256
FF245	442
LON2312	32
MER75016	300
RCA2626	77
WAR23699	132

*arquivo de índice*

**valores ordenados**

32	LON   2312   Romeo and Juliet   Prokofiev ...
77	RCA   2626   Quartet in C Sharp Minor ...
132	WAR   23699   Touchstone   Corea ...
167	ANG   3795   Symphony No. 9   Beethoven ...
211	COL   38358   Nebraska   Springsteen ...
256	DG   18807   Symphony No. 9   Beethoven ...
300	MER   75016   Coq d'or Suite   Rimsky ...
353	COL   31809   Symphony No. 9   Dvorak ...
396	DG   139201   Violin Concerto   Beethoven ...
442	FF   245   Good News   Sweet Honey In The ...

*arquivo de dados*

**geralmente registros desordenados**

# Índice Simples ou Linear

ANG3795	167
COL31809	353
COL38358	211
DG139201	396
DG18807	256
FF245	442
LON2312	32
MER75016	300
RCA2626	77
WAR23699	132

*arquivo de índice*

*campos e*

*registros de tamanho fixo*

32	LON   2312   Romeo and Juliet   Prokofiev ...
77	RCA   2626   Quartet in C Sharp Minor ...
132	WAR   23699   Touchstone   Corea ...
167	ANG   3795   Symphony No. 9   Beethoven ...
211	COL   38358   Nebraska   Springsteen ...
256	DG   18807   Symphony No. 9   Beethoven ...
300	MER   75016   Coq d'or Suite   Rimsky ...
353	COL   31809   Symphony No. 9   Dvorak ...
396	DG   139201   Violin Concerto   Beethoven ...
442	FF   245   Good News   Sweet Honey In The ...

*arquivo de dados*

*campos e*

*registros de tamanho fixo ou variável*

# Características

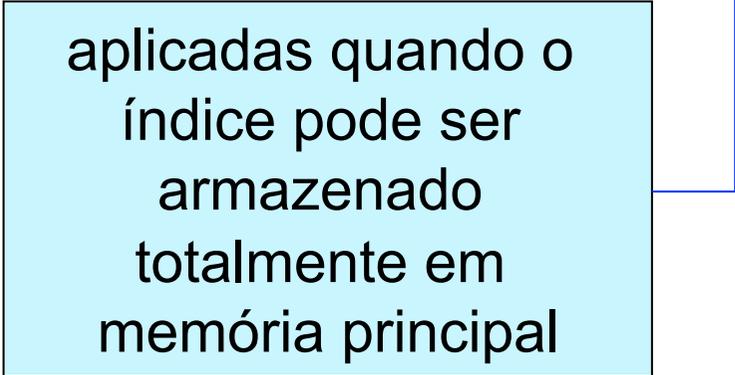
- Arquivo de índice
  - possui a forma de organização mais fácil de se trabalhar
    - campos e registros de tamanho fixo
  - possui tamanho muito menor do que o tamanho do arquivo de dados
  - possibilita a busca binária de seus registros
  - pode ser replicado em memória principal caso seja pequeno o suficiente

# Importante

A existência de índices não afeta a localização física dos registros dos arquivos de dados

# Operações em um Índice

- Pesquisa
- Criação
- Inserção
- Remoção
- Atualização
- Destruição
- Carregamento
- Reescrita



aplicadas quando o índice pode ser armazenado totalmente em memória principal

# Pesquisa

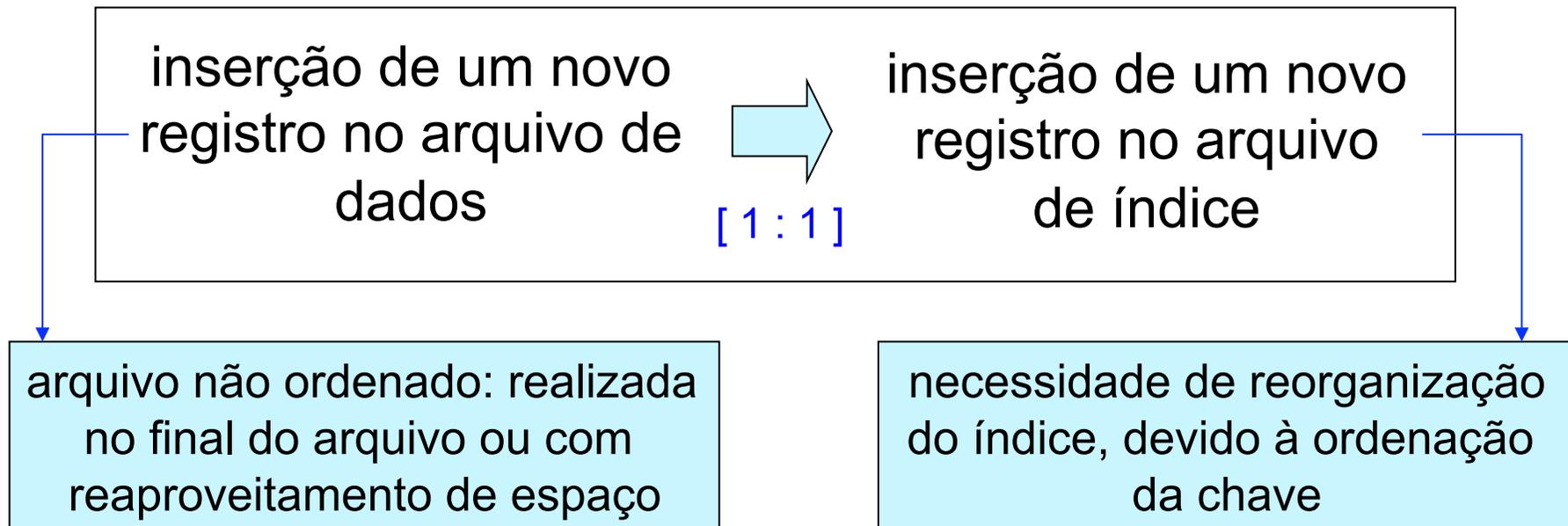
- Baseada na chave de busca
  - encontra a posição da chave no **arquivo de índice**
  - obtém o RRN ou o *byte offset* do registro correspondente à posição encontrada
  - encontra o registro no **arquivo de dados**
  - recupera o registro solicitado do arquivo de dados

# Criação

- Cria o índice juntamente com a criação do arquivo de dados
  - apenas registro de cabeçalho
- Cria o índice baseado em um arquivo de dados já existente
  - registro de cabeçalho
  - demais registros (chave de busca + campo de referência), obtidos a partir de uma varredura no arquivo de dados

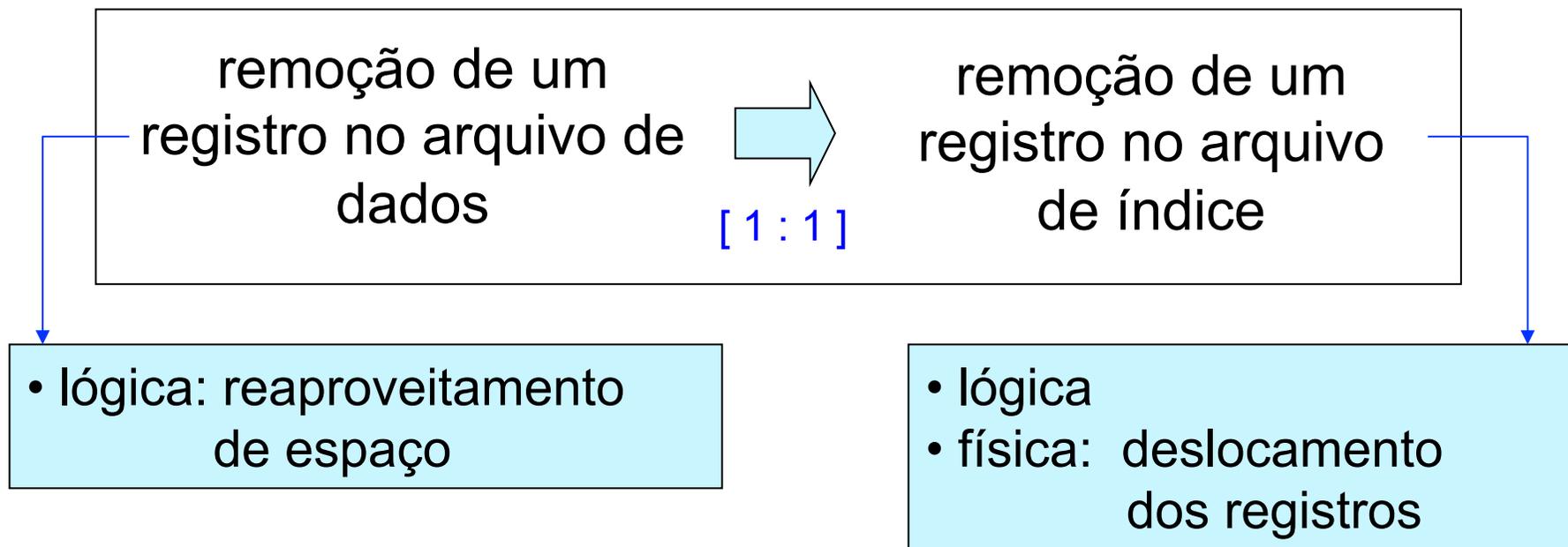
# Inserção

- Adiciona registros no índice devido às inserções no arquivo de dados



# Remoção

- Remove registros no índice devido às remoções no arquivo de dados



# Atualização

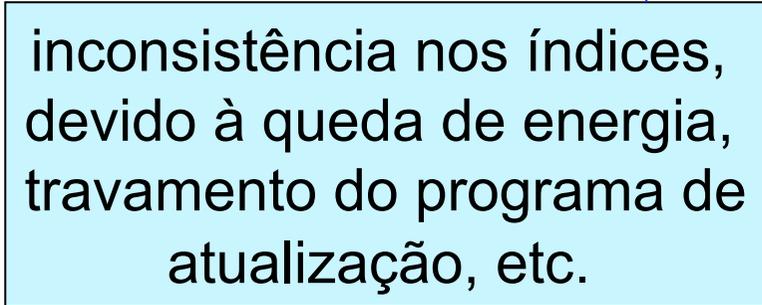
- Modifica registros no índice devido às modificações no arquivo de dados
- Tratamento
  - remoção seguida de inserção ← técnica mais utilizada
  - ou*
  - campo chave: reordenação do índice
  - campo não chave: ajuste do campo de referência se o registro mudar fisicamente no arquivo de dados

# Carregamento

- Carrega o arquivo de índice na memória principal antes de usá-lo
- Passos
  - aponta para o primeiro registro do arquivo de índice em disco
  - varre o arquivo de índices sequencialmente
  - cria o índice em memória principal, em geral implementado como um vetor

# Reescrita

- Atualiza o arquivo de índice em disco com base no arquivo de índice em memória principal, quando necessário
- Informação adicional
  - *status* no registro de cabeçalho
    - verdadeiro/falso



inconsistência nos índices,  
devido à queda de energia,  
travamento do programa de  
atualização, etc.

# Observações

- Índice simples ou linear
  - adequado quando cabe em **memória primária**
- Armazenamento em **memória secundária**
  - pode requerer vários acessos a disco, por causa da busca binária
  - pode ter manutenção cara, devido ao à adição e remoção de registros
  - requer o uso de outras organizações mais apropriadas, como **hash** e **árvore-B**.