

Indexação Primária, Secundária e com Múltiplas Chaves

[Tipos de indexação em SGBD](#)

[Índice Clusterizado](#)

[Características e Vantagens](#)

[Índice Não-Clusterizado](#)

[Características e Vantagens](#)

[Índice primário](#)

[Índice Denso](#)

[Índice Esparsa](#)

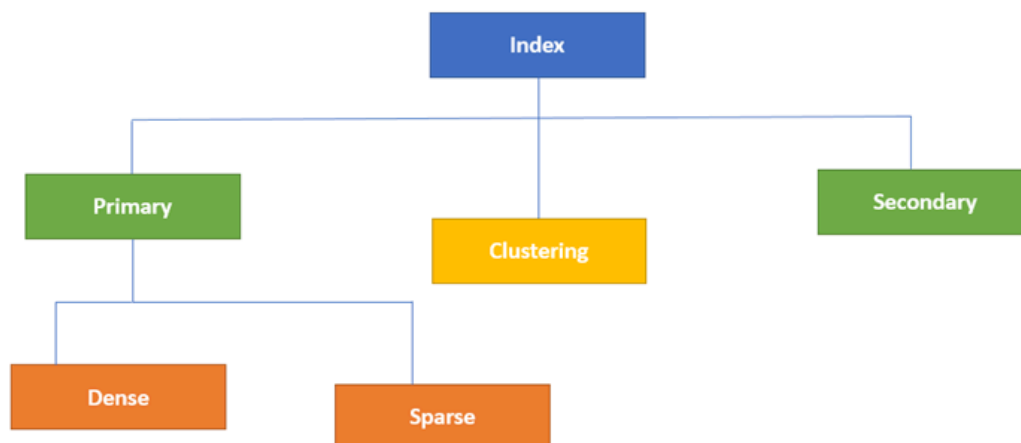
[Índice secundário](#)

[Índice secundário com múltiplas chaves](#)

[Índices Multiníveis](#)

[Características e Vantagens](#)

Tipos de indexação em SGBD



Tipo de índices no banco de dados

Os índices podem ser categorizados em dois tipos principais: **clusterizados** e **não clusterizados**.

Índice Clusterizado

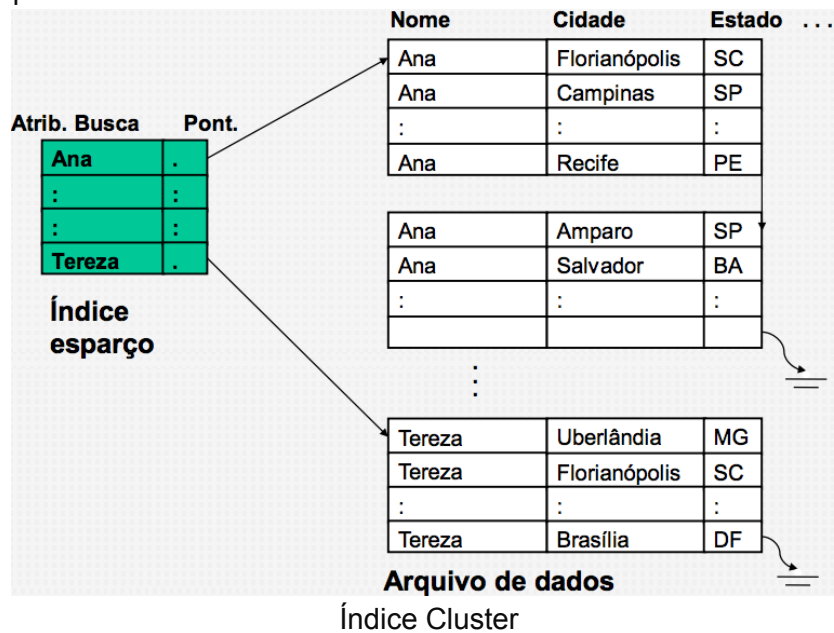
Um índice clusterizado determina a ordem física dos dados em uma tabela. Ou seja, a maneira como os registros são armazenados no disco corresponde à ordem do índice clusterizado.

Um índice de agrupamento possui dois campos, sendo o primeiro campo do mesmo tipo do campo de agrupamento e o segundo campo é um ponteiro para primeiro registro do bloco de disco (registro de âncora ou âncora de bloco). O primeiro campo possuirá um determinado valor e o segundo campo apontará para todos os dados que possuam o

mesmo valor para aquele campo. Este tipo de índice é esparsos, pois possui uma entrada para cada valor distinto do campo de indexação e não para cada registro no arquivo.

Este tipo de índice possui problemas de inclusão e exclusão de registros. O problema com este tipo de índice é a inclusão e exclusão de múltiplos registros, que podem modificar os registros de âncora. Os registros de dados estão fisicamente ordenados, repetindo-se os mesmos problemas observados para a indexação com índices principais.

Vamos supor que uma empresa queira encontrar o cadastro de seus clientes por nome. Nesse caso, a indexação do cluster em SGBD deve ser criada para todos os clientes que têm o mesmo primeiro nome.



Características e Vantagens

- **Ordenação física de registros:** Ao usar um índice clusterizado, os registros são fisicamente armazenados na ordem do índice.
- **Recuperação rápida:** As consultas que seguem a ordem do índice podem ser extremamente rápidas, já que os dados adjacentes são frequentemente lidos juntos.
- **Único por tabela:** Só pode haver um índice clusterizado por tabela, pois define a ordem física dos registros.

Índice Não-Clusterizado

Diferente do clusterizado, o índice não-clusterizado não reorganiza os registros físicos. Em vez disso, ele mantém um apontamento para os registros, como um índice de livro.

Características e Vantagens

- **Estrutura separada:** Mantém uma estrutura separada da tabela de dados original.
- **Múltiplos por tabela:** É possível ter vários índices não-clusterizados para uma única tabela.
- **Flexibilidade:** Pode-se otimizar diferentes consultas criando índices não-clusterizados em diferentes colunas.

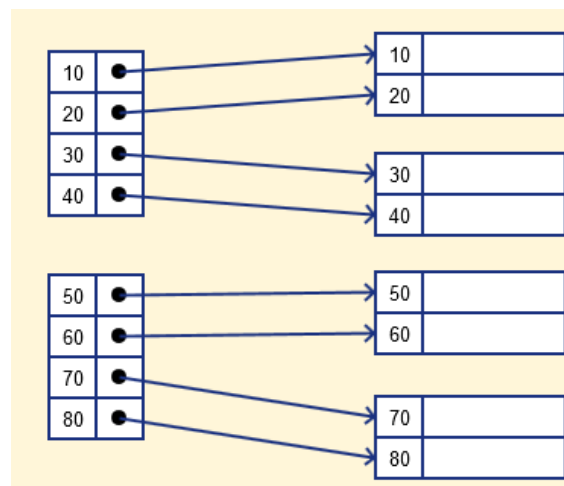
A indexação no banco de dados é definida com base em seus atributos de indexação. Dois tipos principais de métodos de indexação não clusterizados são: **Indexação Primária** e **Indexação Secundária**

Índice primário

O índice primário é um arquivo ordenado com tamanho fixo e dois campos. O primeiro campo é igual a uma chave primária e o segundo campo é apontado para aquele **bloco de dados** específico. Existe uma entrada de índice no arquivo de índice para cada bloco do arquivo de dados. Cada entrada no índice tem o valor da chave primária para o primeiro registro do bloco (registro de âncora ou âncora de bloco). A indexação primária em SGBDs também é dividida em dois tipos: **Índice Denso** e **Índice Esparso**

Índice Denso

Em um índice denso, um registro é criado para cada chave de pesquisa avaliada no banco de dados. Isso ajuda você a pesquisar mais rapidamente, mas precisa de mais espaço para armazenar registros de índice. Nesta indexação, os registros do método contém o valor da chave de pesquisa e apontam para o registro real no disco.

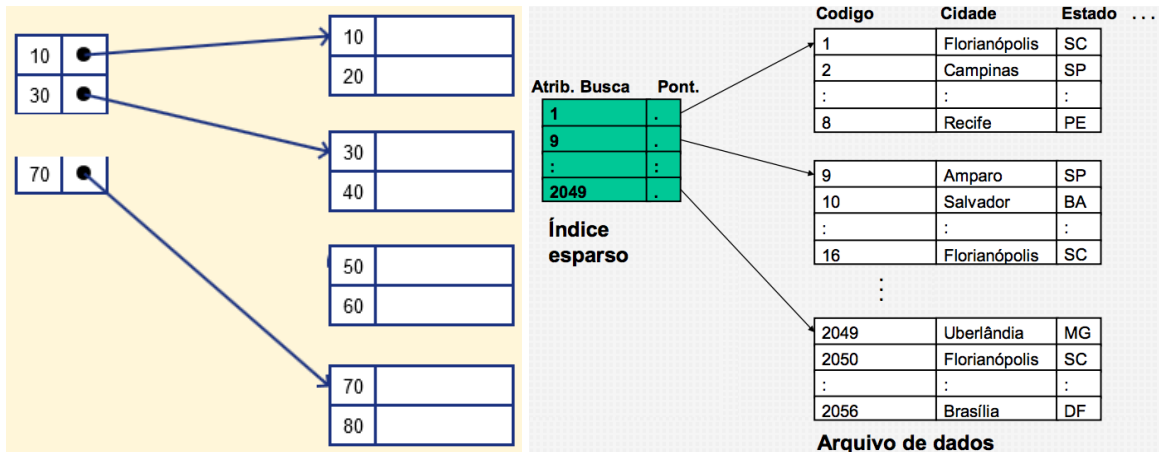


Índice Denso

Índice Esparso

É um registro de índice que aparece apenas para alguns valores do arquivo. O Sparse Index ajuda você a resolver os problemas de indexação densa em SGBDs. Neste método de indexação, um intervalo de colunas de índice armazena o mesmo endereço de bloco de dados e, quando os dados precisam ser recuperados, o endereço do bloco será buscado.

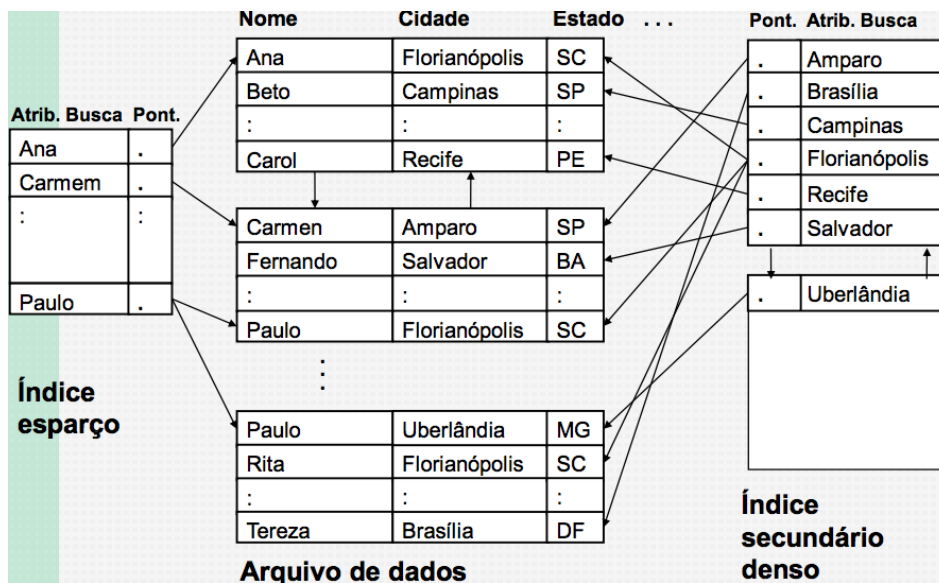
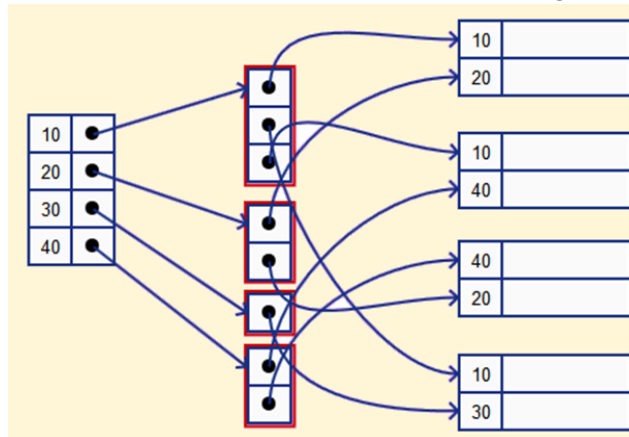
No entanto, o índice esparsos armazena registros de índice apenas para alguns valores de chave de pesquisa. Ele precisa de menos espaço, menos sobrecarga de manutenção para inserção e exclusões, mas é mais lento em comparação com o índice denso para localização de registros.



Índice Esparso

Índice secundário

Um índice secundário é outro meio para acessar um arquivo de dados, quando um índice primário já existe. Um índice secundário pode ser criado em um campo qualquer único ou em um campo não chave com valores duplicados. Um mesmo arquivo pode ter diversos arquivos e ser indexado por mais de um índice secundário. Em geral são densos.



Com base nessas duas categorias fundamentais, emergem diversas abordagens de indexação (ex. Hashing, árvore B, ordem invertida, multinível, etc..).

Índice secundário com múltiplas chaves

É o uso de uma ou mais chaves para localizar conjuntos de registros do arquivo de dados, fazendo uma busca em vários índices e uma combinação (**AND e/ou OR**) dos resultados. Com o uso de **índices secundários**, este tipo de busca é simples e rápida, pois as operações lógicas são realizadas nos arquivos de índices, reduzindo assim os seeks.

Busca usando múltiplas chaves

Exemplo: arquivo de CDs, e dois arquivos secundários para ele, com chaves **Artista** e **Rótulo**.

Algumas consultas possíveis:

- Encontre CD com número= 54 (acesso por chave primária);
- Encontre CD's com Artista= "Chico Buarque" (chave secundária: Artista);
- Encontre todos os CD's com Rótulo= "Philips" (chave secundária: Rótulo).
- Encontre Artista="Milton Nascimento" e Rótulo="Emi-Odeon"
 - Busca nos índices secundários: por Artista e por Rótulo
 - Fazer um AND dos resultados (referência à chaves primárias) retornadas.
 - Buscar no índice primário os endereços dos registros resultantes no arquivo de dados.
 - Fazer seeks no arquivo de dados para buscar os registros.

16

Busca usando múltiplas chaves

Aplicações das chaves secundárias

→ Exemplo: arquivo de CDs, e dois arquivos secundários para ele, com chaves **Artista** e **Rótulo**

| | |
|---|--|
| 0 | 400 Minas Milton Nascimento Emi-Odeon 1975 |
| 1 | 323 Falso brilhante Elis Regina Philips 1976 |
| 2 | 329 A Arte de Chico Buarque PolyGram 1982 |
| 3 | 210 Chico Canta Chico Buarque Philips 1985 |
| 4 | 143 A Arte de Milton nascimento Universal 1988 |
| 5 | 54 Geraes Milton Nascimento Emi-Odeon 1976 |

| | |
|-----|---|
| 54 | 5 |
| 143 | 4 |
| 210 | 3 |
| 323 | 1 |
| 329 | 2 |
| 400 | 0 |

| | |
|-----------------|-----|
| CHICO BUARQUE | 210 |
| CHICO BUARQUE | 329 |
| ELIS REGINA | 323 |
| MILTON NASCIMEN | 54 |
| MILTON NASCIMEN | 143 |
| MILTON NASCIMEN | 400 |

| | |
|-----------|-----|
| EMI-ODEON | 54 |
| EMI-ODEON | 400 |
| PHILIPS | 210 |
| PHILIPS | 323 |
| POLYGRAM | 329 |
| UNIVERSAL | 143 |

Índice primário Índices secundários

17

Busca usando múltiplas chaves

| | |
|---|--|
| 0 | 400 Minas Milton Nascimento Emi-Odeon 1975 |
| 1 | 323 Falso brilhante Elis Regina Philips 1976 |
| 2 | 329 A Arte de Chico Buarque PolyGram 1982 |
| 3 | 210 Chico Canta Chico Buarque Philips 1985 |
| 4 | 143 A Arte de Milton nascimento Universal 1988 |
| 5 | 54 Geraes Milton Nascimento Emi-Odeon 1976 |

| | |
|-----|---|
| 54 | 5 |
| 143 | 4 |
| 210 | 3 |
| 323 | 1 |
| 329 | 2 |
| 400 | 0 |

Resultado da consulta:

1. encontre CD com número= 54 (acesso pela chave primária);

a) Buscar no índice primário o RRN correspondente ao 54

| |
|---|
| 5 |
|---|

b) Buscar no arquivo de dados, o registro com RRN = 5

| |
|--|
| 54 Geraes Milton Nascimento Emi-Odeon 1976 |
|--|

18

Busca usando múltiplas chaves

| | |
|---|--|
| 0 | 400 Minas Milton Nascimento Emi-Odeon 1975 |
| 1 | 323 Falso brilhante Elis Regina Philips 1976 |
| 2 | 329 A Arte de Chico Buarque PolyGram 1982 |
| 3 | 210 Chico Canta Chico Buarque Philips 1985 |
| 4 | 143 A Arte de Milton nascimento Universal 1988 |
| 5 | 54 Geraes Milton Nascimento Emi-Odeon 1976 |

| | |
|-----|---|
| 54 | 5 |
| 143 | 4 |
| 210 | 3 |
| 323 | 1 |
| 329 | 2 |
| 400 | 0 |

Resultado da consulta:

2. encontre CD's com Artista= "Chico Buarque"

a) Buscar no índice secundário Artista= "Chico Buarque"

| | |
|---------------|-----|
| CHICO BUARQUE | 210 |
| CHICO BUARQUE | 329 |

b) Buscar no índice primário o RRN correspondente 210 e o 329

| |
|---|
| 3 |
| 2 |

c) Buscar no arquivo de dados, os registros com RRN 3 e 2

| |
|--|
| 210 Chico Canta Chico Buarque Philips 1985 |
| 329 A Arte de Chico Buarque PolyGram 1982 |

| | |
|-----------------|-----|
| CHICO BUARQUE | 210 |
| CHICO BUARQUE | 329 |
| ELIS REGINA | 323 |
| MILTON NASCIMEN | 54 |
| MILTON NASCIMEN | 143 |
| MILTON NASCIMEN | 400 |

19

Busca usando múltiplas chaves

| | |
|---|--|
| 0 | 400 Minas Milton Nascimento Emi-Odeon 1975 |
| 1 | 323 Falso brilhante Elis Regina Philips 1976 |
| 2 | 329 A Arte de Chico Buarque PolyGram 1982 |
| 3 | 210 Chico Canta Chico Buarque Philips 1985 |
| 4 | 143 A Arte de Milton nascimento Universal 1988 |
| 5 | 54 Geraes Milton Nascimento Emi-Odeon 1976 |

| | |
|-----|---|
| 54 | 5 |
| 143 | 4 |
| 210 | 3 |
| 323 | 1 |
| 329 | 2 |
| 400 | 0 |

Resultado da consulta:

3. encontre todos os CD's com Rótulo= "Philips"

a) Buscar no índice secundário Rótulo=Philips

| | |
|---------|-----|
| PHILIPS | 210 |
| PHILIPS | 323 |

b) Buscar no índice primário o RRN correspondente 210 e o 329

| |
|---|
| 3 |
| 1 |

c) Buscar os registros com RRN 3 e 1 no arquivo

| |
|---|
| 210 Chico Canta Chico Buarque Philips 1985 |
| 323 Falso brilhante Elis Regina Philips 1976 |

| | |
|-----------|-----|
| EMI-ODEON | 54 |
| EMI-ODEON | 400 |
| PHILIPS | 210 |
| PHILIPS | 323 |
| POLYGRAM | 329 |
| UNIVERSAL | 143 |

20

Busca usando múltiplas chaves

| | |
|---|--|
| 0 | 400 Minas Milton Nascimento Emi-Odeon 1975 |
| 1 | 323 Falso brilhante Elis Regina Philips 1976 |
| 2 | 329 A Arte de Chico Buarque PolyGram 1982 |
| 3 | 210 Chico Canta Chico Buarque Philips 1985 |
| 4 | 143 A Arte de Milton nascimento Universal 1988 |
| 5 | 54 Geraes Milton Nascimento Emi-Odeon 1976 |

| | |
|-----|---|
| 54 | 5 |
| 143 | 4 |
| 210 | 3 |
| 323 | 1 |
| 329 | 2 |
| 400 | 0 |

Resultado da consulta 4

Encontre CD's com Artista="Chico Buarque" e Rótulo="Philips"

a. encontre CD's com Artista= "Chico Buarque"

| | |
|---------------|-----|
| CHICO BUARQUE | 210 |
| CHICO BUARQUE | 329 |

b. encontre todos os CD's com Rótulo= "Philips"

| | |
|----------|-----|
| PHILIPS | 210 |
| POLYGRAM | 329 |

c. Faça AND dos resultados (referência à chaves primárias)

| |
|-----|
| 210 |
|-----|

d. Buscar registro com Artista="Chico Buarque" e Rótulo="Philips"

| |
|--|
| 210 Chico Canta Chico Buarque Philips 1985 |
|--|

| | |
|-----------|-----|
| EMI-ODEON | 54 |
| EMI-ODEON | 400 |
| PHILIPS | 210 |
| PHILIPS | 323 |
| POLYGRAM | 329 |
| UNIVERSAL | 143 |

21

Índices Multiníveis

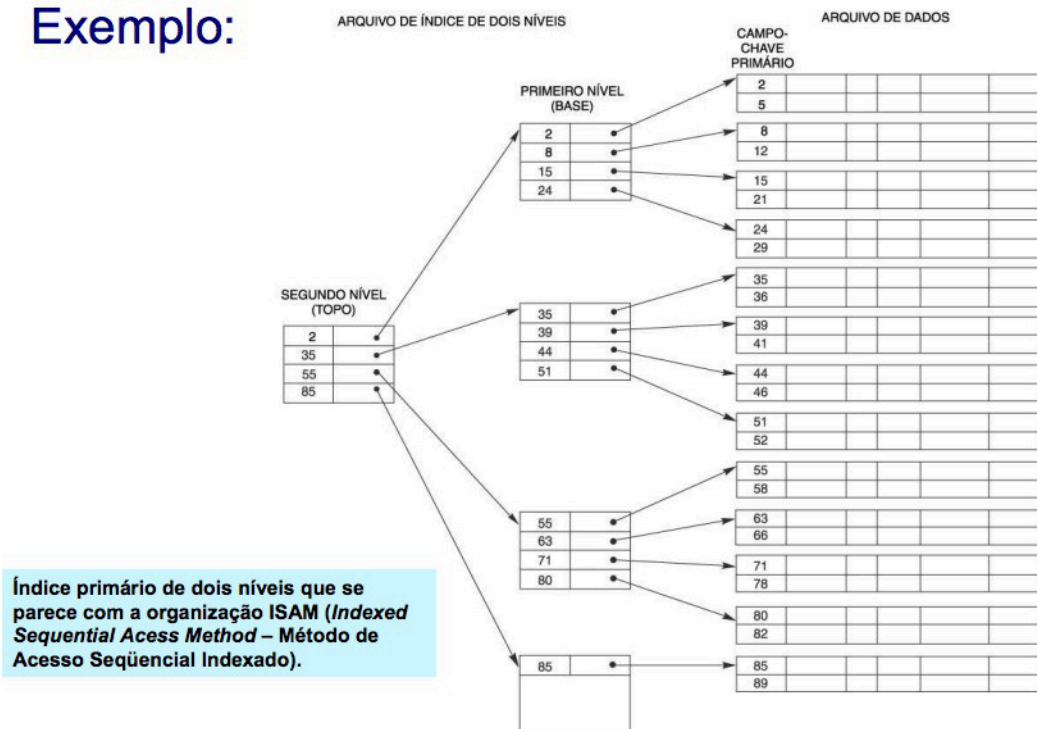
Os índices multiníveis são estruturas de índices hierárquicas que facilitam o acesso rápido a registros em grandes bancos de dados. O conceito central é semelhante ao de um índice de um livro. No primeiro nível, o índice aponta para um intervalo de chaves em um nível inferior. Esse nível inferior pode, por sua vez, apontar para outro nível de índices ou diretamente para os registros. Assim, os níveis superiores do índice oferecem uma visão mais generalizada, enquanto os níveis inferiores fornecem detalhamento, até que se chegue aos registros reais.

É necessário caso o arquivo de índice se torne muito grande. Se este for o caso, é interessante indexá-lo em mais de um nível. Mantendo os índices em arquivos menores, é possível mantê-lo na memória primária, melhorando dessa forma o desempenho e a eficiência. O problema é que um número maior de índices aumenta logicamente a complexidade do sistema.

A organização então segue:

- Primeiro nível: arquivo ordenado pela chave de indexação, valores distintos, entradas de tamanho fixo (como se fosse um índice primário).
- Demais níveis: índice primário sobre o índice do nível anterior e assim sucessivamente até que no último nível o índice ocupe apenas um bloco.
- Número de acessos a bloco: um a cada nível de índice, mais um ao bloco do arquivo de dados.

Exemplo:



Características e Vantagens

- **Buscas Otimizadas:** A busca pode ser feita em vários níveis, tornando-se mais rápida, especialmente quando se lida com uma grande quantidade de dados.
- **Menor Espaço de Armazenamento:** Ao contrário de manter um índice único e massivo, os índices multiníveis podem ser mais compactos e ocupar menos espaço.
- **Minimização do Overhead:** Ao reduzir o espaço de busca em cada etapa, os índices multiníveis minimizam o overhead de I/O e melhoram a performance geral de buscas.

É essencial entender que não há uma técnica superior universalmente; cada uma se destaca em determinados cenários. Assim, ao projetar um sistema ou aplicativo, é vital analisar cuidadosamente as necessidades e características dos dados para escolher a técnica de indexação mais apropriada. Ao fazer a escolha correta, os benefícios em termos de desempenho e eficiência podem ser significativos.

Referencias:

- Sistemas de Gerenciamento de Banco de Dados - Ramakrishnan Gehrke (3ª edição)
- <https://www.ic.unicamp.br/~thelma/gradu/MC326/2010/Slides/Aula05-indexacao.pdf>
- https://sae.unb.br/cae/conteudo/unbfga/lbd/banco2_indices.html
- <https://professorluisleite.files.wordpress.com/2015/03/aula-6-armazena-mento-e-indexaccca7acc83o3.pdf>

Isenção de Responsabilidade:

Os autores deste documento não reivindicam a autoria do conteúdo original compilado das fontes mencionadas. Este documento foi elaborado para fins educativos e de referência, e todos os créditos foram devidamente atribuídos aos respectivos autores e fontes originais.

Qualquer utilização comercial ou distribuição do conteúdo aqui compilado deve ser feita com a devida autorização dos detentores dos direitos autorais originais. Os compiladores deste documento não assumem qualquer responsabilidade por eventuais violações de direitos autorais ou por quaisquer danos decorrentes do uso indevido das informações contidas neste documento.

Ao utilizar este documento, o usuário concorda em respeitar os direitos autorais dos autores originais e isenta os compiladores de qualquer responsabilidade relacionada ao conteúdo aqui apresentado.