

Megadados

Aula 12 – Índices, particionamento

2020 – Engenharia

Fábio Ayres <fabioja@insper.edu.br>

Indices

Índices

Índices são estruturas de dados que facilitam a localização de informação no banco de dados.

Cria-se um índice a partir de uma coluna de uma tabela.

CREATE INDEX

```
CREATE INDEX index_name [index_type]  
    ON tbl_name (index_col_name,...)
```

index_type:

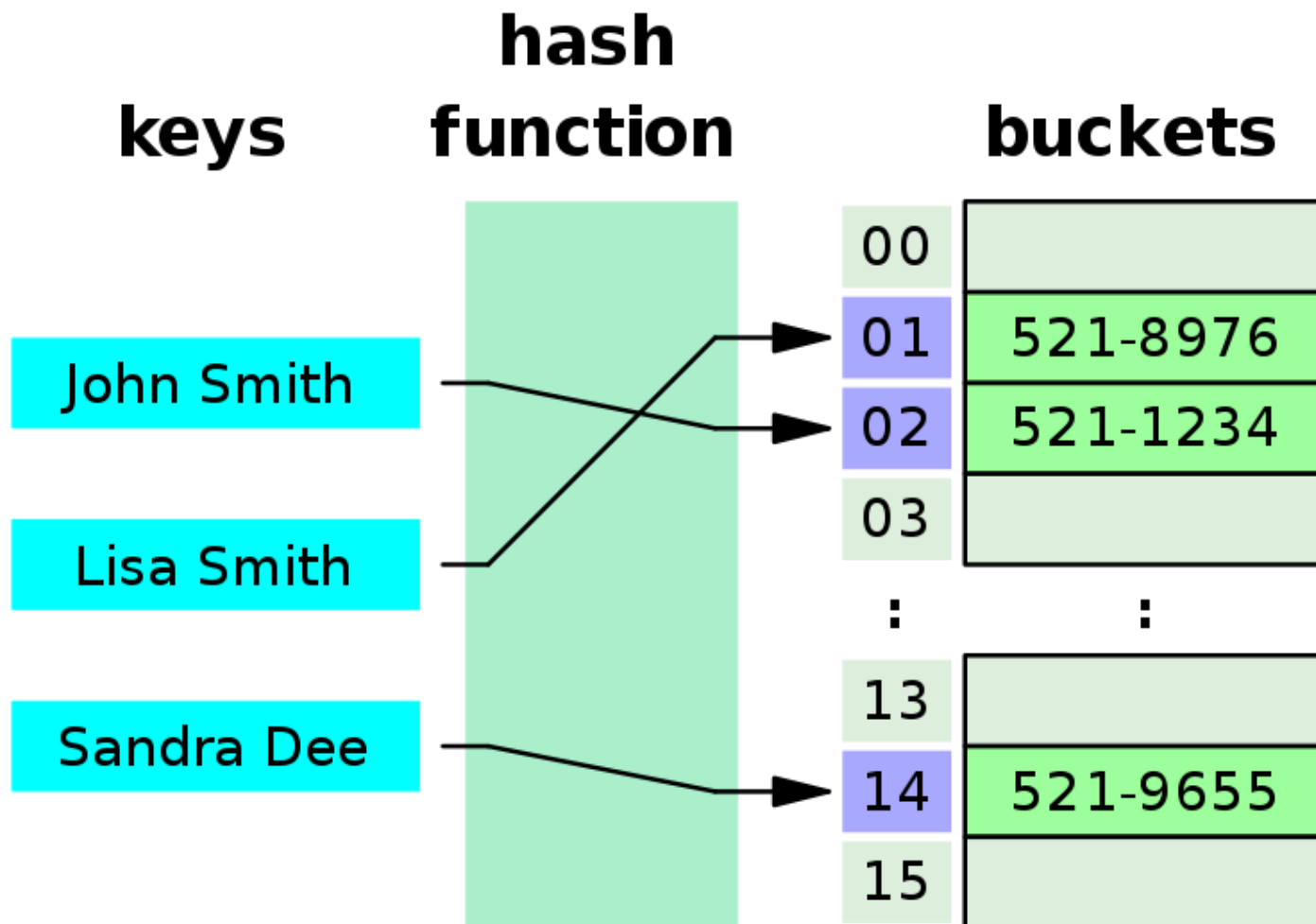
USING {BTREE | HASH}

- Como funcionam os índices? Implementações mais comuns:
 - Hash tables
 - B-trees

Hash tables

- Mapa chave-valor organizados pelo valor hash da chave
- Bom para buscas por valor exato
- Ruim para buscas por faixas de valor

Exemplo



Complexidade

Lembrando de “Desafios de Programação”:

- Busca? $O(1)$
- Inserção? $O(1)$
- Remoção? $O(1)$

Atividade

Explique porque uma hash table é

- Boa para buscas por valor exato
- Ruim para buscas por faixas de valor

B-trees

Um tipo de árvore balanceada de busca

- Razoável para buscas por valor exato
- Boa para buscas por faixa de valor

O professor demonstrará o processo de construção da árvore.

Complexidade das B-trees

- Busca? $O(\log(n))$
- Inserção? $O(\log(n))$
- Remoção? $O(\log(n))$

Atividade

Explique porque uma b-tree é

- Razoável para buscas por valor exato
- Boa para buscas por faixa de valor

Leituras interessantes

<http://www.jasimabasheer.com/posts/btrees.html>

Particionamento

Particionamento

Particionar é dividir as tabelas de um banco de dados em partes menores.

Permite distribuir o banco de dados em vários nós, aumentando o desempenho em situações de acesso concorrente intenso.

Tipos de particionamento

- **Vertical**: as tabelas são **separadas por colunas** em várias tabelas com relacionamento 1:1
 - A tabela original é o INNER JOIN das várias tabelas derivadas
- **Horizontal**: as tabelas são **separadas por linhas** em fragmentos (*shards*). Esse procedimento também é chamado de fragmentação (*sharding*)
 - A tabela original é o UNION das várias tabelas derivadas

Particionamento vertical

Id	Coluna_1	Coluna_2	Coluna_3
1	a	b	c
2	d	e	f
3	g	h	i
4	J	k	l



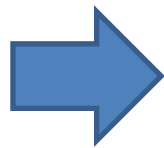
Id	Coluna_1	Id	Coluna_2	Coluna_3
1	a	1	b	c
2	d	2	e	f
3	g	3	h	i
4	J	4	k	l

Para refletir

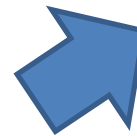
Pense em situações onde o particionamento vertical é benéfico, e onde é problemático.

Particionamento horizontal

id	colA	colB
1	a	b
2	c	d
3	e	f
4	g	h
5	i	j
6	k	l
7	m	n



Função de
partição



id	colA	colB
2	c	d
4	g	h
6	k	l

id	colA	colB
1	a	b
3	e	f
5	i	j
7	m	n

Particionamento horizontal

<https://dev.mysql.com/doc/refman/8.0/en/partitioning-pruning.html>

Para refletir

Pense em vantagens e desvantagens dos seguintes esquemas de particionamento:

- Particionamento por $h(x) \bmod n$, onde $h(x)$ é uma função de hash
- Particionamento por faixas de valores

Insper

www.insper.edu.br