

Curso de Ciência da Computação

Banco de Dados 2 Aula 04

Prof. Dr. rer. nat. Eros Comunello



1

Otimização de BD

Práticas de otimização em Banco de Dados SQL

Tuning

2

Práticas de otimização em BD

- ❑ Usar ÍNDICES?
- ❑ O povo diz: "índices são bons para consultas e ruins para atualizações"
- ❑ Pensar globalmente, fixar localmente
 - ❑ 80 x 20: qual consulta é responsável?
- ❑ Particionar para resolver gargalos
 - ❑ Espaço, transações
- ❑ Ao servidor o que é do servidor
 - ❑ triggers
- ❑ Esteja pronto para trade-offs
 - ❑ Cobertor curto!



Ciência da Computação

3

Práticas de otimização em BD

- ❑ Problemas de sintonia
- ❑ Seleção de índices
- ❑ Alocação de dados
- ❑ Controle de carga (ajuste de MPL - Controle adaptativo da Multiprogramação)
- ❑ Política de substituição de páginas em memória
- ❑ Ajuste de tamanhos/quantidades de buffers
- ❑ Refino automático de estatísticas



Ciência da Computação

4

Práticas de otimização em BD

Dificuldades:

- ❑ Perceber que um recurso está sendo mal utilizado => monitoramento é parte fundamental do processo;
- ❑ Localizar e entender a verdadeira fonte do problema => Mais de 90% do tempo para resolução de problemas de desempenho é gasto no diagnóstico.



Ciência da Computação

5

Práticas de otimização em BD

A atividade de tuning envolve:

- ❑ Hardware e sistemas operacionais;
- ❑ Gerência de memória e acesso a discos;
- ❑ Controle de concorrência e recuperação;
- ❑ Uso de índices adequados;
- ❑ Otimização e reescrita de consultas;
- ❑ Projeto de banco de dados adequado;

Ajuda conhecer SGBDs específicos!



Ciência da Computação

6

Práticas de otimização em BD

Self-Tuning:

- ❑ Capacidade de auto ajuste dos SGBDs ao ambiente existente para obtenção de melhor desempenho;
- ❑ Alguns SGBDs comerciais possuem implementação de algumas características;
- ❑ Trabalhos científicos: além dos índices
 - ❑ e.g. projeto físico, memória, ...



Ciência da Computação

7

Práticas de otimização em BD

Planos de execução:

- ❑ É o resultado da otimização;
- ❑ É especificado no plano de execução:
 - ❑ Ordem de acesso às tabelas;
 - ❑ Ordem de operações;
 - ❑ seleção, projeção e junção;
 - ❑ Índices utilizados;
 - ❑ Tipos de junção;
 - ❑ Ordenações;
 - ❑ Tabelas intermediárias.



Ciência da Computação

8

Práticas de otimização em BD

QEP: Query Execution Plan

Exemplo:

```
SELECT endereço, data-nascimento
FROM empregado
WHERE nome = 'Jhonatan'
```

Execution Plan

```
0      SELECT STATEMENT Optimizer=CHOOSE
1  0    TABLE ACCESS (BY INDEX ROWID) OF 'EMPREGADO'
2    1  INDEX (UNIQUE SCAN) OF 'PK_EMP' (UNIQUE)
```



Ciência da Computação

9

Práticas de otimização em BD

Métodos de Acesso

- ❑ Tipos básicos de operação:
 - ❑ Varreduras sequenciais (full scan);
 - ❑ Indexadas (index scan);
- ❑ Implementação de operadores
 - ❑ Junções;
 - ❑ Uniões;
 - ❑ Ordenações e eliminação de duplicatas.



Ciência da Computação

10

Práticas de otimização em BD

Junções:

- ❑ Principais algoritmos:
 - ❑ Laços Aninhados (Nested Loops Join);
 - ❑ Ordenação/Intercalação (Merge Join);
 - ❑ Hashing (Hash Join).



Ciência da Computação

11

Práticas de otimização em BD

Ordenações:

- ❑ Operação básica para diversas outras operações físicas;
- ❑ Resolve (usualmente) várias cláusulas:
Order By, Group By, Distinct, Union, Minus, ...
- ❑ Operação cara!
- ❑ Se o conjunto a ser ordenado não cabe em memória, pode ser ainda mais cara!



Ciência da Computação

12

Práticas de otimização em BD

Fator de Filtro (FF):

- ❑ Fator de filtro de um predicado:
 - ❑ percentual de linhas da tabela que satisfaz em determinado predicado.
- ❑ Utilizado pelo Otimizador para:
 - ❑ estimar tamanho da tabela resultado;
 - ❑ decidir sobre uso de índices e outras estratégias de acesso.
- ❑ Calculado com base nas estatísticas existentes.



Ciência da Computação

13

Práticas de otimização em BD

Fator de Filtro (FF):

- ❑ Fator de filtro de um predicado:
 - ❑ número entre 0 e 1:
$$0 \leq FF \leq 1$$
- ❑ $FF \rightarrow 0$ poucas linhas
- ❑ $FF \rightarrow 1$ todas as linhas da tabela



Ciência da Computação

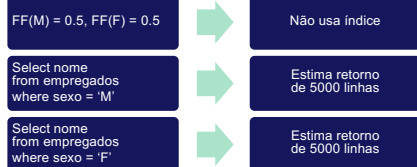
14

Práticas de otimização em BD

Fator de Filtro (FF):

- ❑ Estatísticas:
 - ❑ Tabela de Empregados com 10.000 linhas
 - ❑ Coluna Sexo: 2 valores distintos.

❑ Estimativas:



Ciência da Computação

15

Práticas de otimização em BD

Fator de Filtro (FF):

- ❑ Estatísticas com a presença de **Histograma**: otimizador pode perceber distribuição não-uniforme:
 - ❑ tabela de Empregados com 10.000 linhas:
 - ❑ coluna Sexo: distribuição não uniforme.
 - ❑ estatísticas de distribuição não uniforme:
 - M: 9.000 empregados
 - F: 1.000 empregados
 - ❑ $FF(F) = 0.1$
 - ❑ Para consulta "sexo = 'F'", otimizador pode decidir utilizar o índice da coluna sexo



Ciência da Computação

16

Práticas de otimização em BD

Fator de Filtro (FF):

- ❑ Predicado composto: AND

```
select * from empregados
where dept = '600'
and sexo = 'm' and cargo = 'programador'
```
- ❑ $FF(sexo) = 0.9$
- ❑ $FF(dept) = 0.2$
- ❑ $FF(cargo) = 0.1$
- ❑ $FF = 0.9 * 0.2 * 0.1 = 0.018$
- ❑ Estimativa para a tabela resultado: 180 linhas



Ciência da Computação

17

Práticas de otimização em BD

Fator de Filtro X Índices:

- ❑ O fator de filtro influencia diretamente na utilização de índices => quando consulta retorna poucas linhas e existem índices adequados, otimizador pode utilizar índices;
- ❑ "Poucas linhas" depende de SGBD para SGBD e de versão para versão.



Ciência da Computação

18

Práticas de otimização em BD

Tuning:

- ❑ Problema está em uma consulta apenas...
- ❑ Tentar primeiro métodos com impactos locais
 - ❑ reescrita da consulta;
- ❑ Criação de índices, desnormalização do esquema e reordenação de transações:
 - ❑ podem afetar todo o sistema!



Ciência da Computação

19

Práticas de otimização em BD

Reescrita:

- ❑ Evitar having se pudermos utilizar where.
- ❑ Where elimina linhas antes da operação de agrupamento e facilita a utilização de índices na comparação.

```
Select max(salario)
From empregado
Group by depnum
Having depnum = 100
```



```
Select max(salario)
From empregado
Where depnum = 100
```



Ciência da Computação

20

Práticas de otimização em BD

Reescrita:

- ❑ Eliminar "DISTINCT"
 - ❑ Verificar a lógica da aplicação;
 - ❑ Para consultas em uma única tabela, verificar se o resultado da consulta contém alguma coluna definida como chave única;
 - ❑ Para consultas com junções verificar o tipo de junção, as colunas da junção e as colunas resultado.
- ❑ Exemplo Oracle com 100k rows:
 - ❑ `select distinct ident, nome from empregado;` -> Custo 937
 - ❑ `select ident, nome from empregado;` -> Custo 102



Ciência da Computação

21

Práticas de otimização em BD

Outras atividades de Tuning:

- ❑ Uso de hints (sugestões)
 - ❑ Cuidadoso!
- ❑ Operações vs instruções SQL:
 - ❑ Ex. distinct --> internal sort
- ❑ Ausência de PKs e/ou FKs
 - ❑ Full scan only



Ciência da Computação

22

Práticas de otimização em BD

Outras atividades de Tuning:

- ❑ Métodos com maior abrangência:
 - ❑ Criação/destruição de índices;
 - ❑ Utilização de índices cluster;
 - ❑ Visões materializadas;
 - ❑ Escolha do tamanho adequado de bloco;
 - ❑ Posicionamento dos arquivos em disco;
 - ❑ Particionamento de objetos;
 - ❑ Ajuste das áreas de memória;
 - ❑ Ajuste dos parâmetros do SGBD;
 - ❑ Dentre outros...



Ciência da Computação

23

Link

- ❑ <https://www.postgresql.org/docs/9.3/sql-vacuum.html>
- ❑ <https://www.postgresql.org/docs/12/sql-analyze.html>
- ❑ <http://www.inf.puc-rio.br/~postgresql/conteudo/publicationsfiles/TutorialSBBD09SergioLifschitz.pdf>



Ciência da Computação

24