

OTIMIZAÇÃO, MONITORAMENTO E OPERAÇÃO EM BANCOS DE DADOS





PUC Minas
Virtual

UNIDADE II – TÉCNICAS DE OTIMIZAÇÃO DE BANCOS DE DADOS



PUC Minas
Virtual

OTIMIZAÇÃO E TUNNING DE BANCO DE DADOS

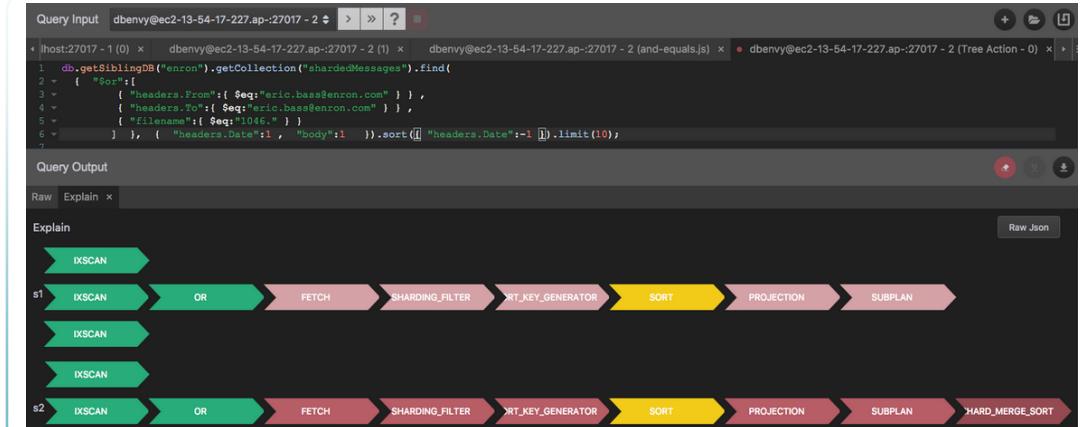


PUC Minas
Virtual

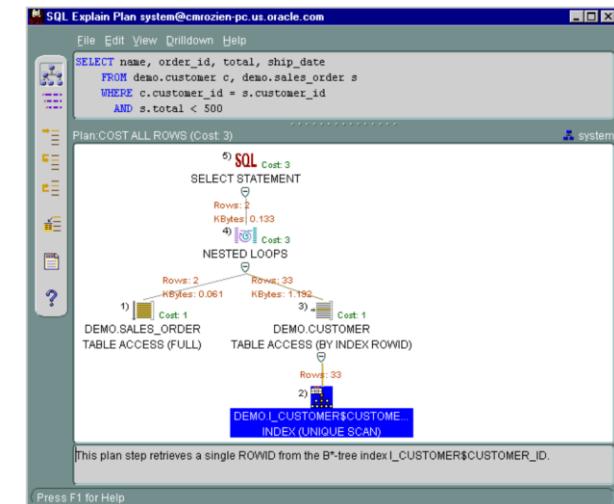
CÁLCULO DE CUSTOS E COMPLEXIDADE DE CONSULTAS

O DESEMPENHO DO BANCO DE DADOS

- Fato é que o desempenho de um banco de dados é obtido verificando-se a taxa entre os pedidos que o SGBD recebe e a entrega da informação.
- Dentro da construção da linha de raciocínio mais lógica precisamos observar alguns fatores que podem influenciar este desempenho: **Workload, Throughput, Recursos do Sistema, Otimização, concorrência.**



Fonte: HARRISON, 2017



Fonte: ORACLE, 2022

O DESEMPENHO DO BANCO DE DADOS

- **Workload:** caracteriza a demanda do banco de dados;
- **Throughput:** é a capacidade do computador de processar e entregar os dados;
- **Recursos do Sistema:** hardware e as ferramentas de software, geralmente o SGBD e seus componentes ou subsistemas, disponíveis no ambiente;
- **Otimização:** analisar e melhorar as pesquisas realizada internamente ao SGBD;
- **Concorrência:** condição em que dois ou mais componentes do *workload* estão tentando o mesmo recurso concomitantemente o que pode ocasionar eventos conflitantes.
- **PERCEBEMOS ASSIM PORQUE CONHECER A INFRA FÍSICA QUE SUPORTA O AMBIENTE, O MODELO DE NEGÓCIOS E O SGBD ESCOLHIDO!**

O OTIMIZADOR DE CONSULTAS

- Uma **MESMA** consulta complexa (tendo em vista que as simples seguem um padrão definido) de SQL pode ser processada de diversas formas com custos de avaliação **DIFERENTES**.
- O processo de otimização de consulta **CONSISTE** em **SELECIONAR** um dos **PLANOS DE EXECUÇÃO** que apresente o resultado mais eficiente, pois a ideia do SGBD é que não exigir que o usuário consiga escrever uma consulta que possa ser processada de forma eficiente.
- Em tese, consultas simples já tem seu estado otimizado e, desta forma, mecanismo de otimização só atua em consultas complexas.

O OTIMIZADOR DE CONSULTAS

- As otimizações de consultas (**query**) com vistas à melhoria de performance do banco de dados, tem seu funcionamento baseado na estratégia formada pelo SGBD para execução de consultas com um SELECT e/ou JOIN (no caso dos bancos relacionais).
- O módulo de otimização de consultas, que é um dos **subsistemas** (*query optimizer*), só implementa as rotinas estratégicas que estão presentes no SGBD e, neste ponto, começamos a verificar as **diferenças de características e qualidades que os distinguem**.
- Existem duas técnicas principais utilizadas para otimização de consultas: **baseada em regras heurísticas e avaliação sistemática do custo dos diferentes planos de execução**.

QUAIS CUSTOS DEVEMOS OBSERVAR PARA CÁLCULO

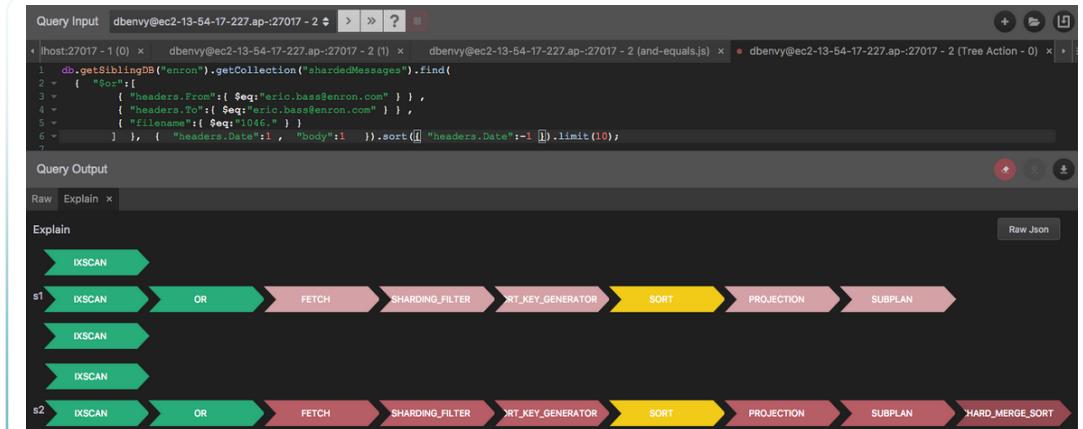
- **Custo de acesso a memória secundária (disco):** relacionado a ação de leitura/gravação de blocos de dados, lembrem-se que os blocos de arquivos são armazenados contiguamente, podemos fazer pesquisa em arquivos temporários, e os tipo de estrutura de acesso.
- **Custo de armazenagem:** ao executarmos um comando há geração de arquivos intermediários e temporários por um plano de execução da consulta definido pelo SGBD.
- **Custo de Computação:** custo de execução das operações **em memória** nos buffers de dados onde estão inclusas pesquisas em registros, classificação, merge para **JOIN**, e **cálculos** em campos com valores.

VOLTEMOS PARA A QUESTÃO DOS CUSTOS DE QUERIES

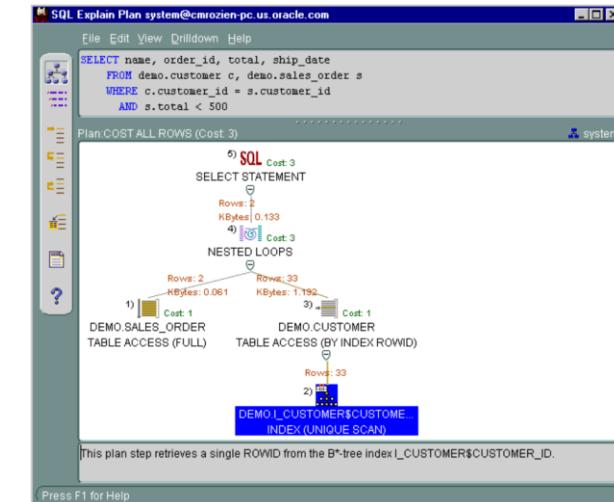
- Custos de Comunicação: Custo de transporte de dados da consulta e seus resultados. (Banco de dados Distribuído).
- E o **foco** passa a ser na atuação para resolver problemas em:
 - **Bancos de dados grandes:** Minimizar custo de acesso a memória secundária (armazenamento em disco, por exemplo).
 - **Bancos de dados menores:** Minimizar custo de computação.
 - **Bancos de dados distribuídos:** Diminuição de custo de comunicação.

QUAIS AS ENTRADAS PARA ANALISAR ESTES CUSTOS?

- **CATÁLOGO DO SGBD:** é o coração de qualquer SGBD, ele armazena todas as definições de esquemas e das bases de dados mantidas e utilizadas pelos componentes internos do SGBD.
- **ESTATÍSTICAS DE BANCO:** contêm informações estatísticas sobre a distribuição de valores em tabelas, colunas de uma tabela ou exibição indexada entre outros itens passíveis de análise para alimentar o otimizador de consultas.



Fonte: HARRISON, 2017



Fonte: ORACLE, 2022

A ESTATÍSTICA TE OFERECE MAIS POSSIBILIDADES!

- Tamanho de arquivo, o número de tuplas (**r**), o número de blocos (**b**), o fator de bloco (**bfr**). Dados do método de acesso primário e os atributos de acesso primário aos arquivos.
- Os índices e atributos dos índices, Número de níveis (**x**) de cada nível dos índices (primary, secondary, clustering).
- Quantidade de valores distintos e o numero de vezes que eles ocorre em cada atributo de cada variável de relação básica.



E ONDE QUEREMOS CHEGAR COM ISTO?

- O **otimizador de consultas** é um **programa**, assim é capaz de **simular** centenas de **estratégias de implementação** para uma única requisição, sendo que o ser humano se limita a realizar três ou quatro maneiras de implementação.
 - É uma fórmula, onde as variáveis são alimentadas pelos dados coletados no **CATÁLOGO DE DADOS e nas ESTATÍSTICAS DE BANCO DE DADOS**.
 - ✓ **Pesquisa linear (s1):** $Cs1 = (b/2)$
 - ✓ **Pesquisa binária (s2):** $Cs2 = \log_2 b + \lceil \frac{1}{bfr} \rceil - 1$
 - ✓ **Índice primário (s3a):** $Cs3a = x + 1$ ou $Cs3a = 1$ (hash)
 - ✓ **Índice ordenado (s4):** $Cs4 = x + (b/2)$
- ✓ O cálculo do custo de uma query completa ficaria algo parecido com isto:
- $$OP = (x - (bi1/2) + (r/2)) = 3 + (4/2) + (3.000/2) = 3 + 2 + 1.500 = 1.505$$

VOLTEMOS PARA A QUESTÃO DOS CUSTOS DE QUERIES

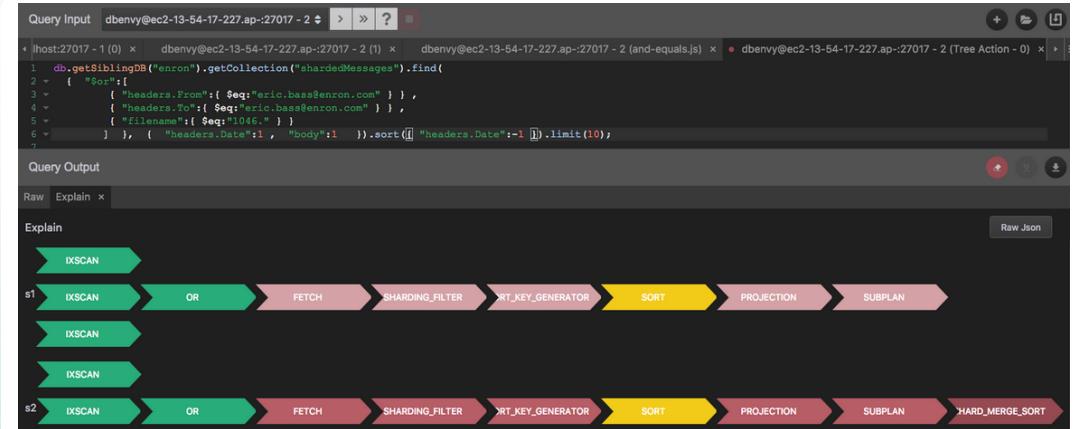
- A implementação do otimizador é o resultado da junção de conhecimentos de diversos programadores que ficam para uso comum de todos os usuários de um dado SGBD.
- Para ajustar o desempenho de um banco de dados devem ser feitos diversos ajustes e devemos observar diversos parâmetros além da projeção de escolhas para a melhoria para cada aplicação.
- Tamanhos do *buffer*, aspectos do projeto , constituição de esquemas e transações, atualizações de versão do SGBD e do sistema operacional principalmente pelas melhorias que podem trazer em cada atualização de versão.

APLICAÇÃO DA HEURÍSTICA NA OTIMIZAÇÃO

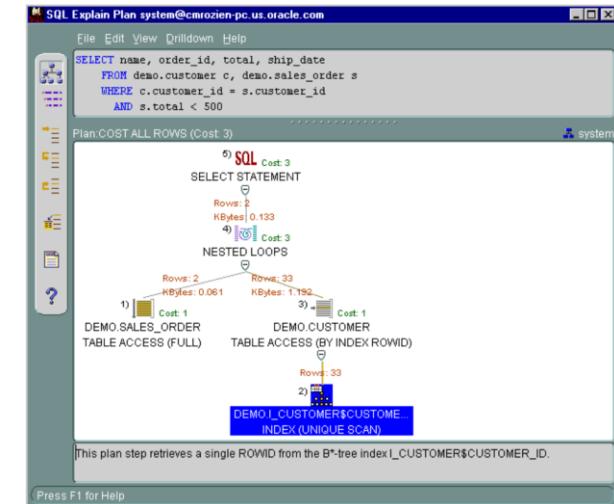
- **A HEURÍSTICA** trabalha diretamente na álgebra relacional, aplicando **regras definidas** gerando um **plano de execução mais eficiente** sem submeter a qualquer tipo de análise de estatística.
- **Não utiliza as estimativas** que estão armazenados nos catálogos do sistema de banco de dados pois o processo requer muito esforço.
- O algoritmo gera uma nova forma algébrica a partir das **regras de equivalência** da **ÁLGEBRA RELACIONAL**.
- **Regras heurísticas** são aplicadas para modificar consultas normalmente em forma de árvore ou **grafos de consulta** para melhorar seu desempenho.

PRINCIPAIS REGRAS DA HEURÍSTICA

- Executar operações de seleção e projeção primeiramente;
- A junção só deve ser realizada depois da seleção e projeção;
- Somente os atributos solicitados para o resultado da consulta e os que realmente são necessários em consultas subsequentes é que devem ser projetados;
- Evitar geração de múltiplas tabelas intermediárias;



Fonte: HARRISON, 2017



Fonte: ORACLE, 2022

APLICAÇÃO DA HEURÍSTICA NA OTIMIZAÇÃO

“As técnicas de otimização de consulta que integram a seleção de heurística e a geração de planos de acesso alternativo foram adotados em diversos sistemas[...]. A busca pelo plano ideal termina quando o orçamento do custo de otimização é ultrapassado e o melhor plano encontrado até esse ponto é retornado.”
(SILBERSCHATZ, 2012, p. 379).

■ REFERENCIAS BIBLIOGRÁFICAS

HARRISON, Guy & Michael. MongoDB Performance Tuning: Optimizing MongoDB Databases and their Applications. Apress. 2021.

SIQUEIRA, Daniel 2021. Histograma: O que é, Exemplos, Gráficos e Tipos. Disponível em: <<https://www.alura.com.br/artigos/o-que-e-um-histograma>>. Acesso em: 19 Fev. 2023.

PEDROZO, Wendel Góes & **VAZ** Maria Salete Marcom Gomes, 2007. Arquitetura para Seleção de Índice no SGBD PostgreSQL, utilizando abordagem baseada em custos do Otimizador . Disponível em: <<https://www.inf.unioeste.br/~clodis/BDI/A11.pdf>>. Acesso em: 19 Mar. 2023.

CANTARELLI, Elisa Maria Pivetta, 2022. Árvores Binárias Balanceadas. Disponível em: <<https://docplayer.com.br/10942308-Arvores-binarias-balanceadas.html>>. Acesso em: 19 Mar. 2023.

HARRISON, Guy, 2017. Getting started with MongoDB explain(). Disponível em: <<https://www.dbkoda.com/blog/2017/11/12/MongoDBExplain>>. Acesso em: 19 Mar. 2023.

SILBERSCHATZ, Abraham. Sistema de banco de dados / Avi Silberschatz, Henry F. Korth, S. Sudarshan; tradução Daniel Vieira. Rio de Janeiro: Elsevier, 2012.

