

Exercício 7 - Otimização

Nome: Letícia Trein Medeiros

Especialização Ciência de Dados

1. Considere a consulta abaixo

```
SELECT E.NOME  
FROM EMPREGADO Emp, ALOCACAO Aloc, PROJETO Proj  
WHERE Proj.NOMEPROJ = 'BD1'  
AND Proj.NUM = Aloc.PNUM  
AND Emp.MAT = Aloc.MAT  
AND Emp.DATANASC > '10/10/1978';
```

a) Reescreva a consulta acima para otimizá-la do ponto de vista de heurística (ou seja, escreva a árvore de otimização).

Resposta:

```
SELECT Emp.NOME  
FROM EMPREGADO Emp  
JOIN ALOCACAO Aloc ON Emp.MAT = Aloc.MAT  
JOIN PROJETO Proj ON Aloc.PNUM = Proj.NUM  
WHERE Proj.NOMEPROJ = 'BD1'  
AND Emp.DATANASC > '1978-10-10';
```

- O filtro Proj.NOMEPROJ = 'BD1' reduz o conjunto de dados do projeto antes das junções.
- O filtro Emp.DATANASC > '1978-10-10' reduz o número de empregados que precisam ser considerados antes da junção.

b) Considerando a otimização baseada em custo, cite ao menos 6 parâmetros diferentes que podem otimizar esta consulta. Explique como estes parâmetros impactam o custo final.

1. **Existência de índices:** Utilização de índices em campos de junção ou colunas pode acelerar a execução da consulta, limitando o número de discos de leitura necessário, não necessitando de escaneamento completo da tabela durante a execução.
2. **Cardinalidade:** Número de tuplas em uma tabela influenciam no custo. Consultas que resultam em uma menor cardinalidade (menos linhas) são geralmente mais baratas.
3. **Tamanho das tabelas:** Tabelas grandes requerem mais leituras de disco e uso de memória, logo, realização de filtragens de tabelas podem ser eficazes.
4. **Ordem de execução das junções :** A ordem na qual o otimizador decide realizar as junções afeta diretamente o custo da consulta. Junções em tabelas menores ou filtradas primeiro podem reduzir a quantidade de dados processados nas etapas seguintes.
5. **Filtragem correta:** Aplicar corretamente filtros (WHERE) antes das junções pode reduzir o número de registros processados, diminuindo o custo.
6. **Tipo de dados:** Tipos de dados mais leves (como INT em vez de VARCHAR) consomem menos espaço e são processados mais rapidamente, impactando positivamente o custo.