3ª Avaliação - CBD.2019.02

**UFRJ - Universidade Federal do Rio de Janeiro**

**COS480 – CBD – Construção de Banco de Dados**

**3ª Avaliação: Estudo comparativos entre resoluções de JOINS**

**Data de entrega: 12 de novembro de 2019**

***Prof. Milton R.Ramirez (***[*milton@matematica.ufrj.br*](mailto:milton@matematica.ufrj.br)***)***

Programar e comparar o desempenho entre algoritmos propostos na literatura para realizar a operação de JOIN entre duas tabelas relacionais. Os requisitos são os seguintes:

1. Coletar medidas de desempenho (custo) usuais, tais como, número de blocos acessados (lidos e escritos) etc.
2. As operações serão sempre EQUI JOINS entre duas tabelas envolvendo apenas um atributo em comum:

SELECT \*

FROM Tabela1 t1 INNER JOIN Tabela2 ON t1.atrb1 = t2.atrb2

1. Algoritmos para resolução de JOIN a serem comparados:
   1. JOIN baseado em Loop Aninhado. A ser testado em todas as organizações primárias, com e sem índices.
   2. JOIN baseado em Ordenação (Merge-Join). A ser testado nas organizações primárias sequenciais ordenadas, ou quando houver índices nos atributos de junção.
   3. JOIN baseado em Hash (Hash-join). A ser testado em todas as organizações primárias, com e sem índices.
2. Em relação ao tamanho das tabelas, comparar as seguintes situações:
   1. Nenhuma das tabelas pode ser carregada totalmente em memória principal
   2. Uma das tabelas pode ser totalmente carregada em memória e a outra não.
3. Em relação à organização primária dos arquivos de registros, comparar as seguintes organizações:
   1. Os dois Arquivos de Registros não estão ordenados pelos atributos utilizados na junção e não possuem nenhum índice secundário nesses atributos;
   2. Os dois Arquivos de Registros não estão ordenados pelos atributos utilizados na junção e apenas um deles possui um índice secundário (Árvore B) nesse atributo;
   3. Os dois Arquivos de Registros não estão ordenados pelos atributos utilizados na junção e ambos possuem índice secundário nos seus atributos de junção;
   4. Um dos Arquivos de Registros está ordenado pelo atributo utilizado na junção (arquivo sequencial) e o outro não está ordenado por esse atributo da junção, porém possui um índice secundário (Árvore B) nesse atributo.
   5. Os dois Arquivos de Registros estão ordenados pelos atributos utilizados na junção e não possuem nenhum índice primário ou secundário nesses atributos;
   6. Arquivo de Registros organizado por Hash estático no atributo utilizado na junção,
   7. Arquivo de Registros organizado pelo índice primário Árvore B+, cujo atributo chave é o atributo utilizado na junção.

O resultado a ser entregue será na forma de (i) um relatório descrevendo as opções adotadas no desenvolvimento dos testes de desempenho, tais como, o tamanho do bloco de memória adotado, o fator de blocagem, a descrição da base de dados usada no teste, a descrição dos atributos utilizados na junção, a descrição dos campos, das chaves, das estruturações dos arquivos adotadas, campos utilizados em ordenações, funções de Hash, etc; (ii) link para um repositório tipo GIT com os arquivos do projeto (iii) tabelas com os valores dos indicadores de desempenhos levantados durante os testes e (iv) uma conclusão a respeitos dos resultados obtidos.