

CK0117 - Sistemas de Bancos de Dados - 2018-1

Javam Machado; Eder Clayton

TRABALHO II - Operadores Algébricos

1 Aspectos Gerais

O trabalho consiste em implementar operações básicas existentes em um SGBD convencional: **ordenação**, **seleção**, **projeção** e **junção** (descritos na seção 2). Na implementação, desconsidere o uso de SGBDs. Use um índice *Hash* nas operações em que a sua utilização for necessária, procure uma lib que já a implementa na linguagem escolhida.

Um banco de dados será mapeado na memória, onde cada tabela terá um conjunto de páginas sem quantidade definida, cada página terá um conjunto de 16 registros e cada registro terá como base uma estrutura que definirá o tipo da tabela, sendo do tipo Funcionário caso precise apenas de uma tabela para efetuar a operação ou duas tabelas, um do tipo Funcionário e outro do tipo Venda, caso a operação a ser implementada seja Junção como mostrado na imagem abaixo.

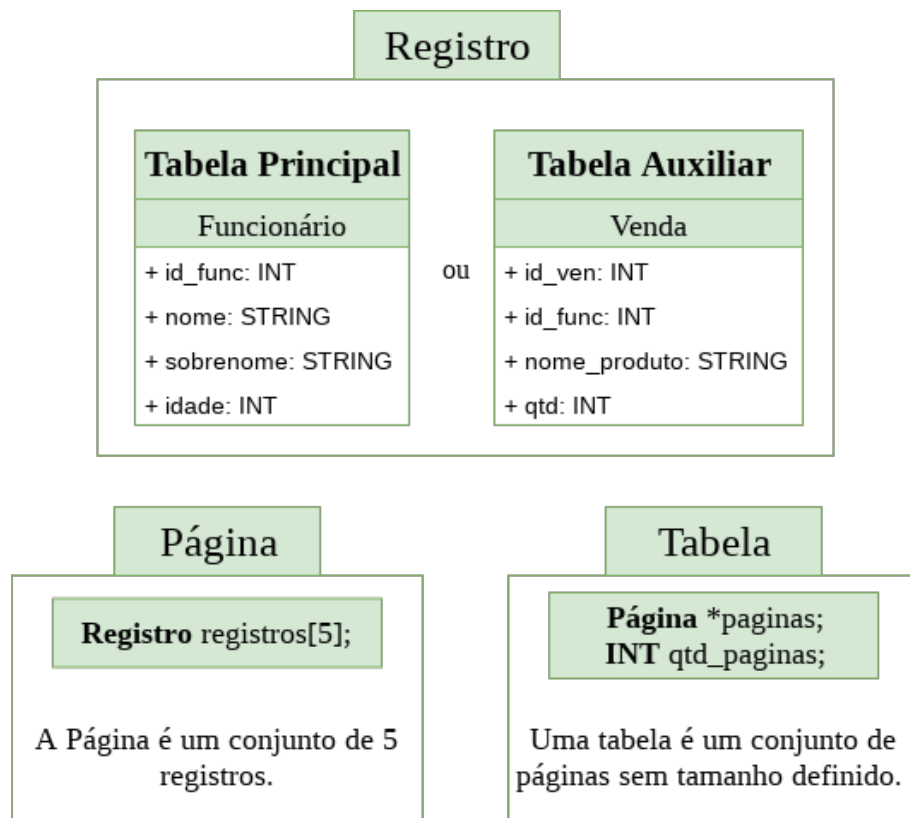


Figure 1: Abstração do Banco de Dados em Memória

Se necessário, utilize páginas temporárias para implementar o operador. Serão enviados arquivos com os registros para teste. Não serão os mesmos registros na avaliação. O resultado da operação deve conter as **tuplas retornadas da operação**, mas as tabelas originais não devem ser modificadas.

2 Lista de Trabalhos

Cada equipe de, *no máximo*, dois alunos fará um dos seguintes trabalhos.

External Merge Sort (EMS): Ordenação externa de uma tabela. O programa irá receber como entrada uma tabela do tipo *heap* e gerar uma saída ordenada da mesma tabela, além da quantidade de etapas necessárias para fazer a ordenação.

Index Selection (IS): Seleção sobre uma tabela utilizando um índice. Para tal considere um predicado de seleção com igualdade.

Distinct Sort Projection (DSP): Projeção com eliminação de duplicatas realizando a ordenação. Para tal você precisa realizar a ordenação externa da relação.

Distinct Hash Projection (DHP): Projeção com eliminação de duplicatas construindo as partições (*buckets*). Você deve utilizar uma implementação de índice *hash* no trabalho, como definido na seção 1.

Nested Loop Block Join (NLBJ): Junção com laço aninhado orientado a bloco - permitir a variação do número de frames de memória para a tabela *outer* a fim de testar o programa.

Nested Loop Index Join (NLIJ): Junção com laço aninhado utilizando índice. Você deve utilizar uma implementação de índice *hash* no trabalho, como definido na seção 1.

Sort Merge Join (SMJ): Junção utilizando a ordenação externa das tabelas. Grave as tabelas ordenadas em disco antes de fazer a etapa de comparação

Hash Join (HJ): Junção utilizando a criação de partições (*buckets*) das tabelas. Grave as tabelas particionadas em disco antes de fazer a partição.

Data da entrega: Segunda-feira - 07 de maio de 2018 com apresentação e arguição no LEC/DC. O código do trabalho deve ser enviado por email para eder.clayton@lsbd.ufc.br até o final do horário da entrega, i.e meio dia da mesma segunda-feira. Envios posteriores não serão aceitos.