

## Lista de Exercícios

MC536 - Bancos de Dados: Teoria e Prática  
Instituto de Computação  
Universidade Estadual de Campinas

Armazenamento e Indexação  
2016  
André Santanchè

A Asdrúbal&Asdrúbal Engenharia executa projetos em todo o Brasil. Em seu sistema de controle de projetos estão duas tabelas:

```
Projeto(IdProjeto, Orçamento, Cidade)
Despesa(TipoDespesa, NrDocumento, IdFornecedor, IdProjeto, Valor)
```

A tabela `Projeto` registra uma entrada para cada projeto:

Coluna	Descrição
<code>IdProjeto</code>	Identificador do projeto.
<code>Cidade</code>	Cidade em que está sendo executado o projeto.

A tabela `Despesa` registra cada despesa associada a cada projeto:

Coluna	Descrição
<code>NrDocumento</code>	Número do documento associado à despesa (e.g., nota fiscal).
<code>IdProjeto</code>	Identifica o projeto ao qual está associada a despesa. Chave estrangeira para a tabela de <code>Projeto</code> .
<code>TipoDespesa</code>	Tipo da despesa que pode ser: 'S' → serviço; 'M' → material de consumo; 'P' → peça; 'A' → aluguel de equipamento.
<code>PagamentoVista</code>	Campo booleano: verdadeiro se for pagamento a vista no ato da compra e falso se for a prazo.
<code>FontePagadora</code>	Dois possíveis estados: 'A' → pago pelo contratante; 'B' → pago pelo contratado. Este campo pode também ser nulo indicando que a fonte pagadora ainda não foi definida e o será posteriormente.
<code>Valor</code>	Valor da despesa.

### Questão 1

Considere três soluções de indexação ou ordenação da tabela de `Despesa`:

- índice na forma de árvore B+;
- índice hash;
- índice bitmap;
- não há índice, mas há uma rotina de ordenação externa para este campo.

Para os seguintes cenários alternativos apresentados a seguir recomende uma ou mais das soluções (a), (b), (c) ou (d) e comente a diferença da sua aplicação em relação às demais soluções. Indique em que campo ou campos serão aplicados os índices. Comente aspectos como consumo de espaço em disco (permanente ou temporário) e desempenho.

I) Uma rotina controle de projeto precisa do total de despesas de um dado projeto a partir do seu `IdProjeto`. Esta rotina é disparada uma grande quantidade de vezes diariamente.

II) Ao final de cada mês é gerado um relatório apresentando o total de despesas por projeto.

III) Os dados da tabela `Despesa` são usados diariamente por rotinas analíticas que fazem cruzamentos de dados e respondem perguntas como: Qual o percentual de despesas com serviço são pagas à vista? Dentre elas, qual o percentual pago pela contratada? Há despesas de material de consumo pagas pela contratante? Qual o tipo de despesa possui valores entre R\$ 5.000 e R\$ 10.000?

## Questão 2

Considerando-se que os valores das despesas são organizados nas seguintes faixas:

- baixa → despesa  $\leq 1.000$ ;
- regular →  $1.000 < \text{despesa} \leq 5.000$ ;
- alta →  $5.000 < \text{despesa} \leq 20.000$ ;
- muito alta → despesa  $> 20.000$ .

Projete um índice bitmap para a tabela *Despesa*, cujo objetivo é responder de forma eficiente às seguintes consultas:

- a) Despesas do tipo X apresentam valores na faixa Y (baixa, regular, alta ou muito alta)?
- b) Há alguma despesa do tipo X, cujo valor seja muito alto e que tenha sido paga à vista?
- c) Dentre as despesas de material de consumo, qual o percentual pago pela contratante?

Defina quantos bits serão aplicados em cada parte do índice e para que finalidade. Justifique porque seu índice atenderá às consultas (a), (b) e (c) de forma eficiente.

## Questão 3

Um sistema de monitoramento da bolsa de valores mantém um cadastro local que replica alguns dados das ações. Sabe-se que em geral aproximadamente 20% das ações têm valor acima da média. Como os dados destas ações é acessado com muita frequência pelo sistema, a empresa decidiu implantar um índice que otimize o acesso às ações acima da média. Indique que índices seriam adequados para este cenário e justifique, considerando os seguintes cenários:

- a) Há grande atualização dos dados das ações durante um mesmo dia.
- b) Os dados são atualizados uma vez por semana.

## Questão 4 (questão de prova)

Considere os comandos SQL a seguir para criar tabelas que controlam Produtos e Receitas, bem como o respectivo esquema relacional simplificado. A tabela de *Produto* mantém um cadastro de produtos, com seu *código*, *nome* e *custo\_unitario* que corresponde ao custo de aquisição de uma unidade do produto. Cada *Receita* tem um *código* e *nome*. Cada entrada nesta tabela *Ingrediente* indica que um *Produto* é componente de uma *Receita* em uma certa quantidade.

<pre>CREATE TABLE Produto (   codigo_produto VARCHAR(5),   nome_produto VARCHAR(80),   custo_unitario FLOAT,   PRIMARY KEY (codigo_produto) );</pre>	<pre>CREATE TABLE Ingrediente (   codigo_receita VARCHAR(5),   codigo_produto VARCHAR(5),   quantidade FLOAT,   PRIMARY KEY (codigo_receita, codigo_produto, quantidade),   FOREIGN KEY (codigo_receita)   REFERENCES Receita (codigo_receita),   FOREIGN KEY (codigo_produto)   REFERENCES Produto (codigo_produto) );</pre>
<pre>CREATE TABLE Receita (   codigo_receita VARCHAR(5),   nome_receita VARCHAR(80),   custo_total FLOAT,   PRIMARY KEY (codigo_receita) );</pre>	<p><b>Esquema Relacional:</b></p> <pre>Produto(codigo_produto, nome_produto, custo_unitario) Receita(codigo_receita, nome_receita, custo_total) Ingrediente(codigo_receita, codigo_produto, quantidade)</pre>

Considere a seguinte consulta SQL realizada na tabela *Ingredientes*:

```
SELECT codigo_receita
FROM Ingrediente
```

```
WHERE codigo_produto="LEITE" AND quantidade>2.0 AND quantidade<5.3
```

Nesta consulta, os valores de comparação para `codigo_produto` e `quantidade` podem variar e são definidos no momento da consulta.

Indique que estrutura de indexação seria a mais adequada para otimizar esta consulta e como a chave deste índice será constituída. Considere que será apenas um índice. Justifique as escolhas feitas.