



RENAN HENRIQUE GOMES DAMAZIO ASSUNÇÃO RA 21038114
ALEX ARANTES GONÇALVES RA 21011214

RELATÓRIO 6

Tópicos Emergentes em Bancos de Dados

QUESTÃO 1

Os scripts de criação final da base de dados se encontram no arquivo Q1.txt

Como nenhuma informação foi fornecida em relação as queries utilizadas, não foi feita análise da utilização de índices.

A restrição 2 trata a cerca da nulidade dos atributos de telefone da tabela Cliente, porém, como não foram apresentados dados sobre a porcentagem de nulos, essa informação não foi considerada para as análises.

A restrição 3 faz referência a criticidade das operações de leitura e escrita das tabelas NotaFiscal e NotaFiscal item para os registros dos últimos 2 meses. A princípio pensou-se em realizar o particionamento pela data, porém, o PostgreSQL não possui recurso de particionamento automático, de forma que todo particionamento deve ser feito no ato da criação da tabela.

Pesquisando mais a fundo as soluções disponíveis na internet encontramos a extensão “PG Partition Manager” que se propoe a fazer conjuntos de partições de tabela baseados em tempo e serial. A solução é de código aberto e está disponível no GitHub através do endereço: https://github.com/pgpartman/pg_partman.

Como as transações críticas de leitura sobre a tabela Cliente Observam somente clientes ativos, esta foi particionada pelo atributo SituacaoCliente, considerando 3 Partições:

- Clientes ativos: com valor de SituacaoCliente = 1
- Clientes inativos: com valor de SituacaoCliente = 0
- Outros valores: considerando qualquer outra situação de cliente possível

QUESTÃO 2

Uma análise envolvendo as relações Proj, Alocado e Emp dos últimos 5 meses revelou um crescimento dos padrões de consulta relacionado abaixo.

Transação	Descrição	Frequência Diária de usc	% Crescimento
T1	select ... from alocado natural join proj where nmlocal in (Londres, São Paulo, Auckland);	168	10
T2	select ... from proj natural alocado natural join emp where qtdehoras >= 160;	111	19
T3	select ... from emp where codemp not in (select codemp from alocado where qtdehoras < 160);	105	11
T4	select ... from emp natural join alocado where codproj not in (select codproj from proj where nmlocal not in (Buenos Aires, Montevideo, Sao Paulo, Campinas));	205	8

Tabela 1: Padrão de Transações sobre as relações Emp, Alocado e Proj

Uma análise de quantidade de tuplas que das relações acima revelou as seguintes informações:

- A quantidade de tuplas da relação T1 é de 33663, 41% do total de tuplas adquiridas pelo sistema para esta consulta..
- A quantidade de tuplas da relação T2 é de 20 tuplas, 12% do total de tuplas adquiridas pelo sistema para esta consulta.
- A quantidade de tuplas da relação T3 é de 366266, 91% do total de tuplas adquiridas pelo sistema para esta consulta..
- A quantidade de tuplas da relação T4 é de 40004, 50% do total de tuplas adquiridas pelo sistema para esta consulta.

Ao se particionar horizontalmente as relações listadas, uma quantidade menor de tuplas seria adquirida pelo sistema da memória física, reduzindo a velocidade das consultas e melhorando o desempenho dos sistema.

Embora a localização 'São Paulo' no atributo nmlocal apareça nas queries das relações T1 e T4, uma partição horizontal baseado neste critério não seria benéfico para T1, uma vez que as duas partições seriam requeridas para esta consulta. Beneficiaria T4, mas não substancialmente porque a sistema deixaria de carregar apenas as tuplas contendo 'São Paulo' ou seja 6667 de 40004 tuplas. Desta forma optamos por realizar uma partição levando em consideração o impacto em T4, pois, muito embora a frequência foi a que menos aumentou nos últimos 5 meses, ainda é a querie de maior frequência diária das relações apresentadas. Aplicamos, portanto, partição horizontal utilizando como critério o atributo 'Campinas' para nmlocal, 13544 tuplas de 40004 a menos seriam requeridas para esta relação.

Seria possível aplicar uma partição horizontal de qtdehoras < 160, mas a relação T2 pouco melhora, uma vez que a quantidade total de tuplas é consideravelmente pequena, 164 tuplas. De maneira similar não haveria vantagens para T3, uma vez que esta

relação requer a maior parte da partição de tuplas cujo atributo qtdehoras ≤ 160 ou seja, 9% a menos do total de tuplas que correntemente o sistema adquiri da memória.