Pontifícia Universidade Católica
DO RIO DE JANEIRO

Eduardo Maria Terra Morelli

Recriação Automática de Índices em um SGBD Relacional

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-graduação em Informática da PUC-Rio como requisito parcial para obtenção do grau de Mestre em Informática.

Orientador: Prof. Sérgio Lifschitz

Rio de Janeiro, setembro de 2006

Pontifícia Universidade Católica
DO RIO DE JANEIRO

Eduardo Maria Terra Morelli

Recriação Automática de Índices em um SGBD Relacional

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Mestrado em Informática do Departamento de Informática do Centro Técnico e Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Sérgio Lifschitz
Orientador
PUC–Rio

Prof. Rubens Nascimento Melo
Departamento de Informática - PUC–Rio

Prof. Renato Fontoura de Gusmão Cerqueira
Departamento de Informática - PUC–Rio

Prof. Geraldo Zimbrão da Silva
Departamento da Ciência da Computação - UFRJ

Prof. Asterio Kiyoshi Tanaka
Departamento de Informática Aplicada - UNIRIO

Prof. José Eugenio Leal
Coordenador(a) Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 19 de setembro de 2006

Eduardo Maria Terra Morelli

Formou-se em Tecnólogo em Processamento de Dados pela PUC-Rio, em 1990. Trabalha desde 1997 como administrador de bases de dados corporativas. Autor dos livros Oracle 8: SQL, PL/SQL e Administração (2000), SQL Server 2000 Fundamental (2001) e Oracle 9i Fundamental (2002), todos pela Editora Érica. Desde 1986 está envolvido com tarefas de treinamento, tendo ministrado cerca de 6.400 horas e desenvolvido mais de trinta apostilas. Também contribui para revistas do ramo tal como SQL Magazine.

Ficha Catalográfica

Morelli, Eduardo Maria Terra

Recriação automática de índices em um SGDB relacional / Eduardo Maria Terra Morelli ; orientador: Sérgio Lifschitz. – Rio de Janeiro : PUC, Departamento de Informática, 2006.

120 f. : il. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

CDD: 004

À Monick, Leonardo e Thiago,

cúmplices em cada pequena conquista que viabilizou este grande trabalho.

Agradecimentos

Ao longo do período no qual desenrolou-se esta dissertação, aconteceram inúmeros momentos de plena euforia e outros de completo desânimo. Felizmente, contei com precisas orientações do Professor Sergio Lifschitz, responsável direto pela manutenção dos trabalhos no rumo adequado. Os incontáveis encontros proporcionaram debates muito interessantes, fazendo as várias horas parecerem alguns minutos.

Devo agradecer a Marcos Salles pela sua excelente dissertação, que lançou as bases a partir das quais pude edificar este trabalho.

Obrigado a Anolan Milanés pelos inúmeros pedidos de socorro atendidos. Sua dissertação também contribuiu para determinar os rumos a seguir.

À amiga Janaína Oleinik, pelas horas a fio de estudos e pelo apoio constante.

Agradeço aos meus valorosos ajudantes, Felipe Rondon e Daniel Seabra, por valiosas contribuições de resultados experimentais.

Finalmente, não poderia esquecer de José Maria Monteiro, cujas dicas foram inestimáveis em diversos momentos.

Resumo

Morelli, Eduardo Maria Terra; Lifschitz, Sérgio. **Recriação Automática de Índices em um SGBD Relacional.** Rio de Janeiro, 2006. 120p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Uma dentre as muitas tarefas desempenhadas por DBAs consiste em tentar garantir que os tempos de respostas dos comandos submetidos por usuários a um grande SGBDR não excedam valores previamente acordados. Esta dissertação segue uma linha de estudos denominada auto-sintonia de índices, que preconiza a realização de ajustes automáticos na execução de consultas SQL, visando reduzir-lhes tempos de resposta, a partir de alterações no conjunto de índices: criação, eliminação e recriação. Este trabalho teve como ponto de partida a dissertação de Marcos Salles [32], que seguiu a mesma linha, propondo um mecanismo automático de criação de índices. Esta dissertação estende [32], primeiro submetendo sua implementação a uma carga de trabalho alternativa e depois realizando eliminações e reconstruções de índices automáticas, levando em consideração níveis de preenchimento de páginas alternativos. Também foram realizados testes utilizando ferramentas comerciais, Oracle 10g e SQL Server 2005, para avaliar quão eficaz comportou-se a implementação proposta em [32]. Vale ressaltar que os testes realizados limitaram-se à criação de índices, já que as ferramentas não oferecem facilidades de reconstrução automática. Diferentemente dos trabalhos publicados nessa linha de estudos e das ferramentas comerciais disponíveis, foi criado um protótipo que não se limita a sugerir novos índices; também são eliminados os que deixaram de ser interessantes, porém, antes ocorre uma avaliação para verificar se a reconstrução não seria mais adequada. Criou-se, inclusive, uma heurística rudimentar que avalia um índice a ser destruído e recomenda sua reconstrução, caso atenda a determinados requisitos.

Palavras-chave

Auto-sintonia; SGBD; Índices; PostgreSQL; Fragmentação

Abstract

Morelli, Eduardo Maria Terra; Lifschitz, Sérgio. **Automatic Reindexing in Relational Databases.** Rio de Janeiro, 2006, 120p. MSc. Dissertation – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

One of the most important tasks of Database Administrators certainly is to guarantee optimal response times to statements submitted by users of big RDBMS. Our dissertation deals with Index Self-tuning, which means creating, dropping or recreating indexes automatically, in order to decrease SQL queries durations. We start from Marcos Salles' dissertation [32], which proposed an automatic way of creating indexes. We extend [32] in many ways: first using a different workload, TPC-H like. Second, following created indexes inspecting its usage. Finally, we have gotten to drop and, mostly, recreate indexes using different fillfactor in leaf pages. Also, we have elaborated many tests using commercial tools, Microsoft SQL Server 2005 and Oracle 10g in order to ratify [32] ideas. Unfortunatelly, we could not test automatic dropping and recreating in these tools, as long as they do not offer this kind of functionalities. Unlike related work and commercial tools, we have created a code prototype that not only suggests new indexes creations, but also drops and recreates indexes using an own heuristics. To validate our ideas we have used a TPC-C like workload, but we had to make some changes to increase updates and force reindexing.

Keywords

Self-tuning; RDBMS; Indices; PostgreSQL; Fragmentation

Conteúdo

Lista de figuras

Lista de tabelas