

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO



Andréa Weberling Carvalho

Criação Automática de Visões Materializadas em SGBDs Relacionais

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Informática da PUC-Rio.

Orientador: Prof. Edward Hermann Haeusler

Rio de Janeiro

Abril de 2011



Andréa Weberling Carvalho

Criação Automática de Visões Materializadas em SGBDs Relacionais

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

Prof. Edward Hermann Haeusler

Orientador

Departamento de Informática - PUC-Rio

Prof. Sérgio Lifschitz

Departamento de Informática - PUC-Rio

Prof. Alessandro Fabricio Garcia

Departamento de Informática - PUC-Rio

Prof. José Eugênio Leal

Coordenador(a) Setorial do Centro

Técnico Científico - PUC-Rio

Rio de Janeiro, 29 de abril de 2011

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Andréa Weberling Carvalho

Possui graduação em Bacharelado em Ciência da Computação pela Universidade Estadual de Santa Cruz - Ilhéus/Bahia (2008). Tem experiência na área de Ciência da Computação, com ênfase em Banco de Dados e desenvolvimento Java WEB, atuando principalmente nos temas de Banco de Dados Auto-gerenciados e Autônomo.

Ficha Catalográfica

Carvalho, Andréa Weberling

Criação Automática de Visões Materializadas em SGBDs Relacionais / Andréa Weberling Carvalho; orientador: Haeusler, Edward Hermann – Rio de Janeiro: PUC. Departamento de Informática, 2011.

v., 83 f.: il. ; 29,7 cm

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

Inclui referências bibliográficas.

1. Informática - Dissertação. 2.Auto-sintonia. 3.Visões Materializadas. I. Haeusler, Edward Herman. II. Pontifícia Universidade Católica do Rio de Janeiro. III. Departamento de Informática IV. Título.

CDD: 004

Agradecimentos

À Deus em primeiro lugar, pois me possibilitou chegar até aqui. “Em tudo dai graças; porque esta é a vontade de Deus em Cristo Jesus para convosco. “ (1 Tessalonicenses 5:18).

Ao meu orientador, que sempre me auxiliou a fazer o melhor, pelo apoio, amizade e pela confiança. Muito obrigada.

Aos meus pais e demais familiares que me apoiaram e não mediram esforços, que deram o melhor que podiam para que eu chegasse até essa etapa de minha vida. Muito obrigada. À minha irmã, pois ela foi o incentivo para eu fazer esse mestrado.

À CAPES e à PUC-Rio, que concedeu-me bolsa e incentivou-me, sem os quais este trabalho não poderia ter sido realizado.

Aos professores que participaram da Comissão examinadora.

A todos os professores e funcionários do Departamento pelos ensinamentos e pela ajuda.

Aos meus amigos e colegas do laboratório de Bio-Informática da PUC-Rio que me apoiaram a apresentação do mestrado, me deram conselhos.

Aos meus amigos queridos Sergio Ruivace, Paulo Silveira, Allan Valeriano, Carlos Junior, Paulo França e Amadeu pelo companheirismo nas horas de estudo e apoio nas disciplinas mais exigentes do mestrado.

Resumo

Carvalho, Andréa Weberling; Haeusler, Edward Hermann. **Criação Automática de Visões Materializadas em SGBDs Relacionais**. Rio de Janeiro, 2011. 83p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Com o aumento da complexidade das aplicações de banco de dados a tarefa de sintonizar os sistemas de banco de dados para melhorar os tempos de resposta das consultas submetidas tem se tornado mais difícil. O uso de visões materializadas é uma forma de melhorar o desempenho do acesso aos dados. Trata-se de um objeto de um banco de dados relacional que armazena um conjunto de dados pré-calculados resultante de uma consulta. Esta dissertação propõe a criação automática de visões materializadas. Para tal utiliza uma arquitetura de auto-sintonia não-intrusiva, isto é, sem modificar o código do SGBD. São propostas métricas para estimar os custos envolvendo visões materializadas permitindo que se determine seu benefício efetivo. São apresentadas heurísticas para a tomada de decisão automática quanto à criação de visões materializadas em função de uma determinada carga de trabalho. São apresentados testes e simulações com o benchmark TPC-H na presença do SGBD MS SQL Server 2008.

Palavras-chave

Auto-sintonia; Visões Materializadas.

Abstract

Carvalho, Andréa Weberling; Haeusler, Edward Hermann (Advisor). **Automatic Creation of Materialized Views in Relational DBMS**. Rio de Janeiro, 2011. 83p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

As database applications become more complex, tuning database systems in order to improve query response times also become harder. One could consider materialized views, a relational database object that stores data resulting from specific queries, to obtain better performances. This dissertation proposes the automatic creation of materialized views. A non-intrusive architecture is used in order to keep DBMS source codes unchanged. There is a need to estimate the creation cost and the benefits obtained from the presence of the materialized views. Heuristics are proposed to help with the automatic decision upon creation of these materialized views for a given workload. Simulation tests for TPC-H benchmark and MS SQL Server DBMS are presented.

Keywords

Self-tuning; Materialized View.

Sumário

| | | |
|--------|---|----|
| 1 | Introdução | 11 |
| 1.1. | Motivação | 11 |
| 1.2. | Objetivo | 12 |
| 1.3. | Organização deste trabalho | 12 |
| 2 | Conceitos Básicos e Trabalhos Relacionados | 14 |
| 2.1. | Índices e Visões Materializadas | 14 |
| 2.2. | Auto-Sintonia de Banco de Dados | 15 |
| 2.3. | Gerência de Visões Materializadas | 17 |
| 2.4. | Seleção de Visão Materializada | 19 |
| 2.5. | Arquitetura não-intrusiva para manutenção do projeto físico do banco de dados | 20 |
| 2.6. | Seleção automática de Visão Materializada | 21 |
| 2.7. | Consultas e Visões Materializadas | 24 |
| 2.8. | Resumo do capítulo | 26 |
| 3 | Heurísticas para sintonia de Visão Materializada na Arquitetura | 27 |
| 3.1. | Arquitetura não-intrusiva para criação de visões materializadas | 27 |
| 3.1.1. | Monitoramento da Carga de Trabalho e de seleção de Visões Materializadas Hipotéticas | 29 |
| 3.1.2. | Seleção Final de Visões Materializadas proposta | 36 |
| 3.1.3. | Materialização de Visões | 40 |
| 3.2. | Resumo do capítulo | 40 |
| 4 | Implementação e Resultados Experimentais | 41 |
| 4.1. | Instanciação da arquitetura não-intrusiva para criação de visões materializadas no SQL Server | 42 |
| 4.2. | Resultados Experimentais com SQL Server 2008 | 46 |
| 4.2.1. | Primeira carga de trabalho submetida | 47 |
| 4.2.2. | Segunda carga de trabalho submetida | 48 |
| 4.2.3. | Terceira carga de trabalho submetida | 49 |
| 4.3. | Resultados de desempenhos | 50 |

| | |
|-----------------------------------|----|
| 4.4. Resumo do capítulo | 52 |
| 5 Conclusões e Trabalhos Futuros | 53 |
| 5.1. Principais contribuições | 53 |
| 5.2. Trabalhos Futuros | 54 |
| Referências Bibliográficas | 55 |
| A Visões sobre parte de Consultas | 59 |
| B Benchmark TPC-H | 67 |
| B.1 Consultas do TPC-H | 68 |
| C Cargas de trabalho submetidas | 80 |
| D Visões Sugeridas | 82 |

Lista de figuras

| | |
|--|----|
| Figura 1 - Consulta A e Visão V para a consulta A | 15 |
| Figura 2 - Heurística para avaliação de consultas | 23 |
| Figura 3 - Consulta A do Benchmark TPC-H e Visão VM_1 | 24 |
| Figura 4 - Consulta B obtida do Benchmark TPC-H | 25 |
| Figura 5 - Consulta C obtida no Benchmark TPC-H | 25 |
| Figura 6 - Visão VM_2 | 25 |
| Figura 7 - Arquitetura para gerência de Visão Materializada | 29 |
| Figura 8 - Algoritmo para encontrar subconjunto de tabelas interessantes (<i>InterestingTableSubsets</i>) | 31 |
| Figura 9 - Passo 1 para definição de Visão Materializada | 32 |
| Figura 10 - Passo 2 para definição de Visão Materializada | 32 |
| Figura 11 - Passo 3 para definição de Visão Materializada | 33 |
| Figura 12 - Passo 4 para definição de Visão Materializada | 33 |
| Figura 13 - Passo 5 para definição de Visão Materializada | 33 |
| Figura 14 - Algoritmo DefineView | 34 |
| Figura 15 - Algoritmo <i>MountVM</i> para criar as definições de consultas de visões materializadas | 34 |
| Figura 16 - Heurística para selecionar visões materializadas hipotéticas | 35 |
| Figura 17 - Heurística para seleção final de Visão Materializada | 38 |
| Figura 18 - Classes dos <i>drivers</i> da arquitetura | 45 |
| Figura 19 - Classe de CWO | 45 |
| Figura 20 - Classes dos componentes CSO, CE e CMVM | 46 |
| Figura 21 - Gráfico de tempo de execução Carga 1 | 51 |
| Figura 22 - Gráfico de tempo de execução Carga 2 | 52 |
| Figura 23 - Gráfico de tempo de execução Carga 3 | 52 |
| Figura B.1 - Modelo do TPC-H | 67 |

Lista de tabelas

| | |
|---|----|
| Tabela 1 - Tamanhos de tabelas versus quantidade de páginas | 39 |
| Tabela 2 - Visões Materializadas utilizadas para primeira carga | 48 |
| Tabela 3 - Visões Materializadas utilizadas para segunda carga | 48 |
| Tabela 4 - Visões Materializadas utilizadas para quarta carga | 50 |
| Tabela A1 – Tabela de comparação de custo estimado utilizando visões sobre parte de consultas | 66 |