

**Marcos Antonio Vaz Salles**

**Criação Autônoma de Índices  
em Bancos de Dados**

**DISSERTAÇÃO DE MESTRADO**

**DEPARTAMENTO DE INFORMÁTICA  
Programa de Pós-graduação em  
Mestrado em Informática**

Rio de Janeiro  
Julho de 2004

PONTIFÍCIA UNIVERSIDADE CATÓLICA  
DO RIO DE JANEIRO



**Marcos Antonio Vaz Salles**

**Criação Autônoma de Índices em Bancos  
de Dados**

**Dissertação de Mestrado**

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

Orientador: Prof. Sérgio Lifschitz

Rio de Janeiro  
Julho de 2004



**Marcos Antonio Vaz Salles**

## **Criação Autônoma de Índices em Bancos de Dados**

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

**Prof. Sérgio Lifschitz**

Orientador

Departamento de Informática — PUC-Rio

**Prof. Arndt von Staa**

Departamento de Informática — PUC-Rio

**Prof. Ana Carolina Brandão Salgado**

Centro de Informática — UFPE

**Prof. Ângelo Roncalli Alencar Brayner**

Departamento de Informática — UNIFOR

**Prof. José Eugenio Leal**

Coordenador Setorial do Centro Técnico Científico —

PUC-Rio

Rio de Janeiro, 15 de Julho de 2004

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

### **Marcos Antonio Vaz Salles**

Formou-se em Ciência da Computação pela Universidade Estadual de Campinas (UNICAMP). Trabalhou durante diversos anos com o desenvolvimento e a sintonia de aplicações de bancos de dados para o mercado financeiro. Desenvolveu durante o Mestrado, junto com o Prof. Sérgio Lifschitz e seu grupo de pesquisa, estudos sobre o uso de agentes de *software* para auto-sintonia de índices em SGBDs relacionais.

#### Ficha Catalográfica

Salles, Marcos Antonio Vaz

Criação Autônoma de Índices em Bancos de Dados/ Marcos Antonio Vaz Salles; orientador: Sérgio Lifschitz. — Rio de Janeiro : PUC-Rio, Departamento de Informática, 2004.

v., 113 f: il. ; 29,7 cm

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

Inclui referências bibliográficas.

1. Informática – Teses. 2. Auto-sintonia. 3. SGBD. 4. Agentes de software. 5. Índices. 6. PostgreSQL. I. Lifschitz, Sérgio. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

A meus pais.

## Agradecimentos

Ao Prof. Sérgio Lifschitz, orientador deste trabalho de pesquisa, pela motivação, apoio, opiniões e pela amizade que muito contribuíram para a condução e conclusão do trabalho.

Aos meus pais, pelo amor e apoio incondicionais que me permitiram trilhar um caminho de desafios e estudos.

À Capes, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, e à Faperj, Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro, pelo apoio financeiro.

Aos colegas do nosso grupo de pesquisa, pelo clima de intercâmbio e criação de novas idéias. Os colegas Rogério Costa, Maíra Ferreira de Noronha e Anolan Milanés contribuíram diretamente com este trabalho através de diversas discussões e de apoio na implementação.

Aos membros da banca examinadora, por seus relevantes comentários que ajudaram a tornar este texto mais palatável.

Aos professores do Departamento de Informática, pelo empenho na produção de trabalhos de pesquisa de relevância internacional. Os professores Arndt von Staa, Celso Ribeiro e Fábio Porto, em especial, me transmitiram questionamentos, idéias e ensinamentos importantes para a qualidade desta dissertação.

Aos demais amigos, colegas e funcionários do Departamento de Informática, pela ajuda, idéias trocadas e momentos de descontração.

A Deus e a São Miguel Arcanjo, pela luz e proteção concedidos em todos os passos da minha caminhada.

## Resumo

Salles, Marcos Antonio Vaz; Lifschitz, Sérgio. **Criação Autônoma de Índices em Bancos de Dados**. Rio de Janeiro, 2004. 113p. Dissertação de Mestrado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A escolha e materialização de índices são atividades comumente realizadas por administradores de bancos de dados (DBAs) para acelerar o processamento de aplicações de bancos de dados. Devido à complexidade da tarefa de seleção de índices e à pressão por maior produtividade sobre os profissionais que realizam sintonia, diversos trabalhos na literatura e em sistemas comerciais procuram obter ferramentas que possam apoiar o DBA na escolha dos melhores índices para uma dada carga de trabalho. Classificamos estes trabalhos como sendo de auto-sintonia local, uma vez que se focam em um problema de sintonia específico, em oposição a trabalhos de auto-sintonia global, que almejam obter um desempenho aceitável para o sistema como um todo. Esta dissertação propõe duas arquiteturas que permitem automatizar completamente a sintonia de índices. A independência de intervenção humana é obtida através do uso de agentes de *software*. A combinação de agentes com SGBDs torna os sistemas mais autônomos e capazes de auto-sintonia. Implementamos uma das arquiteturas propostas no SGBD de código fonte aberto PostgreSQL e obtivemos resultados experimentais com uma carga transacional que mostram a viabilidade de nossa abordagem.

## Palavras-chave

Auto-sintonia; SGBD; Agentes de software; Índices; PostgreSQL.

## Abstract

Salles, Marcos Antonio Vaz; Lifschitz, Sérgio. **Autonomic Index Creation in Databases**. Rio de Janeiro, 2004. 113p. MSc. Dissertation — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The choice and materialization of indexes are activities commonly done by database administrators to speed up database application processing. Due to the complexity of the index selection task and to the pressure for productivity increase put on tuning professionals, many works on the literature and on commercial systems seek for tools that can help the DBA choose the best indexes for a given workload. We classify these works as local self-tuning, once they are interested in a specific tuning problem, in opposition to global self-tuning work, which is targeted at obtaining acceptable performance for the system as a whole. This dissertation proposes two architectures that allow the complete automation of the index tuning task. Human intervention independence is achieved through the use of software agents. The combination of agents and DBMS makes systems more autonomous and self-tuning. We have implemented one of the proposed architectures in the open source DBMS PostgreSQL and obtained experimental results with a transactional workload that show the feasibility of our approach.

## Keywords

Self-tuning; DBMS; Software Agents; Indexes; PostgreSQL.

## Conteúdo

1	Introdução	11
2	Auto-sintonia de Bancos de Dados e Agentes de <i>Software</i>	19
2.1	Auto-sintonia de Bancos de Dados	19
2.2	Agentes	23
2.3	Comentários Finais	28
3	Auto-sintonia de Índices através de Agentes	30
3.1	Infra-estrutura no SGBD	30
3.2	Arquitetura com um Único Agente	38
3.3	Arquitetura com Múltiplos Agentes	50
3.4	Comentários Finais	58
4	Implementação e Resultados Experimentais	60
4.1	Arquitetura do PostgreSQL	60
4.2	Cargas de Trabalho e Bases de Dados de Teste	63
4.3	Ambiente de Implementação	67
4.4	Implementação de Índices Hipotéticos	68
4.5	Implementação da Arquitetura com um Único Agente	77
4.6	Comentários Finais	101
5	Conclusões e Trabalhos Futuros	104
	Referências Bibliográficas	107

## Lista de Figuras

1.1	Exemplo de árvore B+	13
2.1	Metodologia para seleção de índices	21
2.2	Integração de Agentes com SGBDs	26
2.3	<i>Framework</i> de construção de agentes de [35]	27
3.1	Modelo local para auto-sintonia usando agentes	40
3.2	Modelo global para auto-sintonia usando agentes	41
3.3	Estratégia de avaliação de consultas	44
3.4	Estratégia de avaliação de atualizações	46
3.5	Instanciação do <i>framework</i> de construção de agentes para auto-sintonia de índices	47
3.6	Arquitetura com múltiplos agentes	51
3.7	Instanciação do <i>framework</i> para construção do <i>Agente de Coleta de Comandos</i>	52
3.8	Instanciação do <i>framework</i> para construção do <i>Agente de Seleção de Índices</i>	54
3.9	Instanciação do <i>framework</i> para construção do <i>Agente de Execução de Sintonia</i>	55
4.1	Processos e módulos do PostgreSQL	62
4.2	Esquema do banco de dados simples de vendas	63
4.3	Arquitetura do <i>toolkit</i> DBT-2	65
4.4	Esquema do <i>benchmark</i> TPC-C	66
4.5	Tempo de Criação de Índices Reais e Hipotéticos	73
4.6	Integração entre o agente e o PostgreSQL	79
4.7	Diagrama de classes envolvidas na configuração das camadas	82
4.8	Diagrama de classes da camada <i>Sensor</i>	83
4.9	Uma interação típica na camada <i>Sensor</i>	84
4.10	Diagrama de classes da camada <i>Crença</i>	85
4.11	Uma interação típica na camada <i>Crença</i>	86
4.12	Diagrama de classes da camada <i>Raciocínio</i>	86
4.13	Uma interação típica na camada <i>Raciocínio</i>	87
4.14	Diagrama de classes da camada <i>Ação</i>	89
4.15	Uma interação típica na camada <i>Ação</i>	90
4.16	Vazão observada com testes de 30 minutos	94
4.17	Vazão observada com testes de 60 minutos	96
4.18	Vazão observada com testes de 90 minutos	97
4.19	Custos ponderados da carga de trabalho	99
4.20	Comandos vistos pelo agente nos testes	100

## Lista de Tabelas

4.1	Freqüências das transações executadas pelo <i>toolkit</i>	67
4.2	Algumas consultas utilizadas para avaliação de estatísticas para índices hipotéticos	74
4.3	Qualidade de índice hipotético em consulta por intervalo e agregação	75
4.4	Qualidade de índice hipotético em consulta de ordenação	76
4.5	Valores dos parâmetros de inicialização do PostgreSQL	92
4.6	Índices sempre criados pelo agente e pelo <i>toolkit</i>	98