

5 Conclusões e Trabalhos Futuros

Este capítulo apresenta uma revisão do trabalho desenvolvido nesta dissertação, lista as contribuições obtidas, e mostra algumas direções em que o trabalho poderia ser estendido.

O objetivo da dissertação foi estudar soluções totalmente automáticas para a escolha e materialização de índices em SGBDs relacionais. Inicialmente mostramos o contexto em que se encaixa a nossa pesquisa comentando sobre trabalhos que vêm sendo desenvolvidos para tornar sistemas de bancos de dados capazes de auto-sintonia. Estes trabalhos, bem como esta dissertação, procuram reduzir ou eliminar a intervenção humana em atividades de sintonia que precisam ser realizadas para os atuais sistemas.

Apresentamos uma descrição do problema de escolha de índices em SGBDs relacionais. As propostas presentes na literatura se focam na escolha de quais tabelas e colunas devem ser indexadas e na obtenção de boas relações de custo-benefício levando em conta as freqüências e tipos de comandos da carga de trabalho. Outras características importantes do problema, como a existência de periodicidades e perfis para a carga de trabalho, a obtenção da carga de trabalho e a determinação do momento para criar ou remover índices, não recebem tratamento adequado.

Sugerimos que o uso de agentes pode endereçar os pontos pouco explorados do problema. A característica de autonomia de agentes é interessante para a construção de sistemas adaptáveis e que não necessitam de intervenção humana. Outros trabalhos de nosso grupo de pesquisa já exploraram o uso de agentes para auto-sintonia de sistemas de bancos de dados.

Propusemos duas arquiteturas baseadas em agentes para a auto-sintonia de índices em bancos de dados. Ambas possibilitam a total eliminação da intervenção de profissionais especializados na tarefa de sintonia de índices. A primeira arquitetura utiliza um único agente embutido no SGBD e apresenta, em relação à segunda arquitetura, menores custos de observação do sistema. As decisões de criação e remoção de índices são realizadas pela *heurística de benefícios*, revisada e estendida nesta dissertação a

partir do trabalho de [8]. Esta heurística dispara as ações de modificação da base imediatamente, sem apresentar nenhum mecanismo de agendamento. O uso desta arquitetura é recomendado para cargas de trabalho relativamente estáveis e para ambientes em que não haja restrições quanto à criação de índices durante a operação normal do sistema.

A segunda arquitetura, por outro lado, já possui mecanismos para agendamento de execução de ações e para a obtenção dos perfis e sazonalidades da carga de trabalho processada. O seu uso de múltiplos agentes importa, entretanto, em maiores custos para observação, armazenamento de crenças e atuação sobre o SGBD. Esta arquitetura deve ser empregada para cargas de trabalho com modificações periódicas e em ambientes em que as decisões de alteração sobre a base devam ser executadas em momentos “quietos” do sistema.

Implementamos a arquitetura com um único agente no SGBD de código fonte aberto PostgreSQL. Em primeiro lugar, alteramos o SGBD para permitir a simulação de índices hipotéticos. Conseguimos obter um mecanismo de simulação eficiente e com boa qualidade.

Conduzimos, então, estudos sobre como o agente para escolha automática de índices poderia ser integrado com o SGBD, respeitando seu modelo de processos. Apresentamos o projeto detalhado do agente utilizando um *framework* para implementação de agentes e mostramos as questões envolvidas na extensão de cada uma das camadas desse *framework*.

Avaliamos o agente implementado submetendo o sistema a diversos testes com uma carga transacional. Os resultados de vazão e de qualidade dos índices encontrados automaticamente indicam que o uso de nossa abordagem é tanto mais interessante quanto maior for o tempo de execução estável da carga de trabalho. Observamos, ainda, que com o tipo de aplicação utilizada o agente conseguiu manter boa qualidade de escolha de índices mesmo quando o sistema foi submetido a cargas de processamento mais intensas.

É importante ressaltar que trabalhos anteriores que endereçam a seleção de índices para bancos de dados relacionais não propõem soluções independentes de intervenção humana para a sintonia dos índices do sistema. Estes trabalhos são fundamentais para a formalização de heurísticas eficientes para a escolha de índices, porém recomendam o uso destas heurísticas em ferramentas de apoio ao DBA. Já trabalhos anteriores que utilizaram agentes em conjunto com SGBDs não abordaram o problema de sintonia de índices. Esta dissertação procura preencher o espaço deixado por esses trabalhos anteriores, utilizando agentes para alcançar uma solução inde-

pendente de intervenção humana para o problema de sintonia de índices em sistemas de bancos de dados.

Contribuições da Dissertação

As principais contribuições trazidas por esta dissertação foram:

- um estudo dos trabalhos desenvolvidos sobre o tema de auto-sintonia de bancos de dados, documentado em [36].
- a extensão de uma arquitetura com um único agente para auto-sintonia de índices em bancos de dados relacionais.
- a proposta de uma arquitetura com múltiplos agentes que permite a escolha e materialização de índices em SGBDs relacionais sem qualquer intervenção humana.
- a implementação de mecanismos para a simulação de índices hipotéticos em um SGBD de código fonte aberto, o PostgreSQL. Trabalhos anteriores haviam realizado este tipo de implementação em sistemas comerciais, o que impede a análise das alterações realizadas sobre o SGBD.
- a implementação de uma arquitetura para auto-sintonia de índices em um SGBD de código fonte aberto e a sua avaliação experimental, demonstrando a viabilidade do uso da abordagem de agentes para o problema de sintonia de índices.

Trabalhos Futuros

O trabalho de pesquisa realizado no contexto desta dissertação pode ser estendido de diversas formas. Seria interessante investigar a aplicação da abordagem de agentes aqui empregada em outros problemas de auto-sintonia local, como a gerência de memória do sistema ou a alocação de dados entre diferentes elementos de processamento. Um problema especificamente relacionado ao contexto de sintonia de índices é o da sintonia de fatores de preenchimento de páginas e a escolha do momento para a realização da reorganização física de um índice.

Arquiteturas de agentes como as utilizadas nesta dissertação permitem que reutilizemos o trabalho de pesquisa já presente na literatura sobre heurísticas de decisão para problemas de sintonia. A incorporação destas

heurísticas exige pequenas adaptações para que possam ser executadas em resposta a medições contínuas realizadas pelo agente sobre o SGBD.

Na implementação realizada, é interessante investigar se o uso de diferentes heurísticas para escolha de candidatos trariam escolhas de índices com qualidades muito diferentes para as cargas de trabalho processadas. Também é possível avaliar o impacto de desempenho de heurísticas que percorressem por mais tempo o espaço de busca de possíveis índices para cada comando. Estamos atualmente avaliando como implementar este tipo de heurística.

Nossa implementação pode ser estendida, ainda, para incorporar a destruição automática de índices anteriormente presentes na base de dados. Com isto poderíamos avaliar que tipos de ajustes o agente realizaria sobre bases de dados em que especialistas humanos já tivessem realizado a sintonia de índices. Outro ponto de extensão da implementação é a construção da arquitetura com múltiplos agentes proposta no Capítulo 3.

Por fim, seria interessante analisar como o agente implementado se comportaria em termos da qualidade dos índices selecionados com aplicações em que há uma maior variação nos comandos submetidos ao SGBD. Cenários típicos seriam sistemas utilizados para consultas *ad hoc* e para suporte a decisões. É possível especular que o agente enxergaria um maior percentual dos comandos submetidos ao sistema, mesmo sobre carga intensa de processamento, pelo fato de cada comando neste tipo de ambiente ser mais complexo e possuir maior duração para seu processamento. Isto causa a submissão efetiva de um número esperado menor de comandos ao sistema. De toda forma, resta saber se o maior percentual de comandos visto permaneceria significativo para a seleção automática dos índices para a base de dados. Repare que, em sistemas com este perfil de uso, até mesmo a criação de índices por parte de um especialista é uma tarefa de alta complexidade, uma vez que é difícil prever as porções da base de dados que são de maior interesse para os usuários.