

Uso de Índices e Visões para a Otimização de Banco de Dados

Digiampietri, L.A., Scarabelin, F.D.

Objetivos

O objetivo deste projeto foi ajudar a resolver o problema de otimização automática de Sistemas Gerenciadores de Bancos de Dados sugerindo a criação de índices e visões materializadas. Por se tratar de um problema extremamente complexo, o objetivo não era abordar todas as facetas do processo de otimização, mas sim tratar alguns dos pontos mais importantes possibilitando alcançar resultados interessantes do ponto de vista prático, científico e educacional. A solução almejada deveria ser genérica de forma a ser aplicada em qualquer sistema gerenciador de bancos de dados relacionais que utilizassem índices e visões materializadas em seus geradores de plano de execução de consultas. Além da otimização de cargas de trabalho para bancos de dados de leitura e escrita, este projeto também pretendia trazer contribuições para bancos de dados apenas de leitura (como é o caso de alguns "data warehouses") [1] e também para estimar o comportamento de um banco de dados na medida em que o banco de dados e/ou a carga de trabalho crescessem.

Métodos/Procedimentos

Usamos o PostgreSQL para a avaliação do desempenho dos SGBDs em relação ao tempo de execução de consultas. Através das formulas matemáticas que regem o comportamento dos SGBD (em relação à quantidade de registros dos bancos de dados, presença de índices, presença de visões materializadas, entre outros) especificamos um modelo matemático do comportamento geral dos SGBDs em relação ao tamanho das tabelas, índices e visões e com isso, estimamos automaticamente quais índices e visões materializadas poderiam ser criados para aumentar o desempenho do SGBD (considerando restrições de espaço em disco disponível).

Resultados

Com o problema de leitura e escrita, os melhores resultados foram obtidos nas cargas de trabalho que tiveram mais consultas do que atualizações. No caso de banco de dados que faziam uso somente de leitura, o custo total de cada consulta foi de 6 a 18 vezes mais rápida quando comparada com o custo que o banco de dados original debandava. Por último, temos também uma estimativa do comportamento para uma determinada carga de trabalho, que tinha como idéia avaliar se era possível conceber um banco de dados dado a sua carga de trabalho esperada. Os primeiros resultados foram satisfatórios, mas em futuros projetos a meta é obter resultados mais precisos. [2]

Conclusões

Os resultados dos testes com o benchmark foram considerados bem sucedidos, porém aspectos adicionais de otimização serão estudados para otimizar ainda o tempo de resposta das consultas feitas sobre o banco de dados. Em futuros projetos, pretende-se além de aprimorar o desempenho da otimização propriamente dita, tornar o projeto o mais automático possível, além de desenvolver uma ferramenta gráfica para a apresentação dos resultados.

Referências Bibliográficas

[1] Harizopoulos, S., Liang, V., Abadi, D. J., and Madden, S. (2006). Performance tradeoffs in read-optimized databases. In Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB 2006), pages 487–498. ACM.

[2] Alves Junior, A.C., Scarabelin, F.D., Digiampietri, L.A. (2010). A framework for database optimization and workload control. Revista de Sistemas de Informação da Faculdade Salesiana Maria Auxiliadora (ISSN 1983-5604) - aceito para publicação