

Avaliando o Impacto da Modelagem de Dados em Aplicações OLAP Utilizando SGBD Relacional e Colunar

1

Abstract.

Resumo.

1. Introdução

Para tomar decisões de maneira rápida e inteligente Organizações têm como base os dados armazenados em seus repositórios. Conforme o volume destes dados aumenta a má estruturação do repositório pode degradar o desempenho do acesso às informações, e devido a isso a tecnologia de Data Warehouse é utilizada para persistência de dados e tomada de decisões de forma eficiente.

Além do armazenamento de dados, é necessário que se tenha uma aplicação que os acesse e traga informações de maneira tão eficiente quanto à estruturação do repositório. Aplicações OLAP analisam dados multidimensionalmente e executam consultas analíticas, fazendo com que estejam fortemente associadas à Data Warehouses. Ao conjunto de Data Warehouse e aplicações OLAP denomina-se ambiente OLAP.

Um ambiente OLAP deve ser projetado e implantado visando a rapidez na recuperação de dados [Wrembel and Koncilia 2007], [Codd et al. 1993], [Kimball and Ross 2002]. Sistemas Gerenciadores de Bancos de Dados (SGBD) surgem como uma solução para esta questão, e duas classes de SGBD podem atuar como gerenciadores de Data Warehouses, os relacionais e os NoSQL. Para que haja uma conclusão sobre qual abordagem é a melhor, a aplicação de um benchmark torna-se pertinente, e no escopo de banco de dados um benchmark que se destaca por tratar de sistemas de suporte à decisão é o TPC-H [TPC-H 2017].

Outro fator que interfere no desempenho de um ambiente OLAP é a forma como este é modelado. Ambientes normalizados são utilizados pela facilidade de manutenção, e o fácil entendimento acerca do relacionamento entre as entidades, que traz uma visão mais clara do sistema [Bax and Souza 2003]. Entretanto, modelos denormalizados surgem para trazer ganho no desempenho das consultas, por diminuir as junções entre tabelas devido ao menor número de entidades.

Dadas estas condições, o objetivo aqui é realizar um estudo comparativo entre um SGBD relacional e outro NoSQL como gerenciadores de Data Warehouse utilizando o TPC-H. Devido a seu amplo uso e poucas modificações na linguagem SQL, foi definido o PostgreSQL como SGBD relacional e o MonetDB como NoSQL, este por ainda utilizar SQL, possuir interface simples e ser o pioneiro entre os bancos NoSQL colunares.

2. Recuperação e Persistência de Dados

3. TPC-H

4. O Experimento

5. Conclusão

Referências

Bax, M. P. and Souza, R. (2003). Modelagem estratégica de dados: Normalização versus "dimensionalização". *KMBRASIL, Anais... São Paulo: SBGC*.

Codd, E., Codd, S., and Salley, C. (1993). Providing olap (on-line analytical processing) to user-analysis: An it mandate. *White paper*.

Kimball, R. and Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Wiley Computer Publishing, 2 edition.

TPC-H (2017). Tpc-h homepage. Acessado em: 19 de maio de 2017.

Wrembel, R. and Koncilia, C. (2007). *Data warehouses and OLAP: concepts, architectures, and solutions*. Igi Global.