

SDDC: cluster de banco de dados definido por software

Islan Santos¹, Marta Valéria do Rosário¹, George Leite Junior¹

¹Instituto Federal de Sergipe (IFS)
Caixa Postal – 49.400-000 – Lagarto – SE – Brazil

{islanm, marttavaleria}@gmail.com, george.junior@ifs.edu.br

Abstract. *Many companies are looking for dynamic infrastructures that fit the needs of their applications and budget. We know that the information generated by a company is an essential asset. However, it has been a great challenge for companies to keep all the information generated available in an efficient, fast and secure way. With this in mind, this summary proposes a platform capable of replicating data between different DBMSs in a transparent way for the applications.*

1. Introdução

O elevado custo de soluções pagas de Sistemas Gerenciadores de Bancos de Dados (SGBD) tais como: Oracle, DB2 (IBM) e SqlServer (Microsoft), fazem com que pequenas e médias empresas acabem optando por opções gratuitas ou versões com limitações. Embora estas soluções possam atender grande parte dos sistemas, existem funcionalidades que apenas são disponibilizadas em versões pagas, e outras precisam utilizar soluções de terceiros para suprir certos recursos, como é o caso do Slony, que é um software de replicação para banco PostgreSQL [Marcotte 2005].

A replicação de dados já vem sendo utilizado há alguns anos na computação, e seu foco tem variado desde que começou a ser utilizado, inicialmente seu principal foco era a tolerância a falhas que ocorriam frequentemente nos sistemas de computação. Em seguida, a alta demanda por serviços na web fez com que a replicação fosse utilizada também a fim de propiciar maior disponibilidade aos arquivos mais requeridos. Atualmente, além de cumprir os papéis já mencionados, a replicação de dados é utilizada em larga escala para aumentar a velocidade de processamento [de Souza 2013].

Com foco em diminuir os custos com SGBDs, este trabalho propõe uma plataforma para Cluster Independente de Banco de Dados Definida por Software denominada SDDC (Cluster Database Defined by Software). Os recursos disponibilizados ajudarão os administradores de aplicações a analisar o nível de sobrecarga no acesso ao dado, à tolerância a falhas entre outros.

2. Solução Proposta

Para a solução proposta, será desenvolvido uma plataforma que será responsável por gerir toda a inteligência de banco de dados de uma aplicação e utilizará a replicação de dados em SGBDs de diferentes tipos. O *middleware* proposto será construído utilizando a plataforma Java, e inicialmente, trabalhará com 3 tipos de banco de dados: SqlServer, PostgreSQL e MySql, mas nada o impede que outros bancos sejam adicionados no futuro.

A técnica escolhida para a replicação dados será a *Primary Copy Replication*, também conhecida como replicação passiva [Wiesmann and Schiper 2005], nela existe

o conceito de réplicas primárias e secundárias, todas as transações são encaminhadas para um banco de dados principal, chamado de réplica primária, os demais bancos são chamados de réplicas secundárias e são utilizados como backups, em caso de uma falha na réplica primária uma das réplicas secundárias é promovida a primária e a camada de dados mantêm-se operando normalmente. O pedido é processado apenas pela réplica primária sendo o estado resultante transferido para as restantes [Costa 2015]. Entretanto, a plataforma permite que novas regras sejam construídas, assim pode ser definidas novas formas de replicação.

A plataforma também disponibilizará uma interface de gerenciamento web capaz de criar *dashboards* para ajudar o monitoramento do acesso à informação, permitindo detectar gargalos e prever sobrecargas.

3. Considerações Finais

A plataforma proposta fornecerá recursos para que pequenas e médias empresas, que não possuem grandes orçamentos para contratação de SGBDs proprietários, possam executar aplicações em um ambiente com tolerância a falhas, alta disponibilidade de dados e backups em SGBDs de diferentes desenvolvedores de forma transparente.

A conclusão deste trabalho será importante para futuras pesquisas na área, além de poder adicionar mais SGBDs ao middleware, também é necessário o estudo de novas técnicas para melhorar alguns pontos fracos que a replicação passiva possa apresentar, e por fim, propor o desenvolvimento de outro tipo de replicação para fazer um estudo comparando os tempos de operações entre as técnicas.

Referências

- Costa, A. D. (2015). Gestão de bases de dados relacionais em cloud computing.
- de Souza, M. F. (2013). *UMA ESTRATÉGIA PARA BALANCEAMENTO DE CARGA EM BANCO DE DADOS REPLICADOS*. PhD thesis, Universidade Federal do Rio de Janeiro.
- Marcotte, L. (2005). Database replication with slony-i. *Linux Journal*, 2005(134):1.
- Wiesmann, M. and Schiper, A. (2005). Comparison of database replication techniques based on total order broadcast. *IEEE Transactions on Knowledge and Data Engineering*, 17(4):551–566.