

1

Introdução

1.1. Motivação

Muitos sistemas de informações atingiram tamanha importância que a sua indisponibilidade, ainda que temporária, acarreta grandes prejuízos econômicos. Assim, é de grande motivação o estudo de estratégias para torná-los tolerante a falhas. Nesse sentido, a estratégia de replicar e distribuir os dados aumenta a probabilidade de que estes sistemas continuem ativos mesmo diante de problemas de energia elétrica, *software* e *hardware*, inclusive problemas de rede.

Garcia-Molina [GARCIA-MOLINA, 2000] mostra que a redundância de dados produzida com a replicação também permite distribuir a demanda de serviços entre dois ou mais pontos de acesso. Dessa forma, é possível que cada cliente acesse o ponto mais próximo, diminuindo o tráfego na rede. Isso também possibilita trabalho *offline*.

A replicação de dados ainda permite o estabelecimento de *backups contínuos*, isto é, com intervalos consideravelmente menores que os processos tradicionais de *backup*. Assim, ao invés de manter uma cópia diária de toda a base de dados, pode-se manter uma replicação com atrasos de minutos. Conseqüentemente, é obtida uma cópia mais recente e sem a necessidade de realizar um grande processamento na base de dados. Dessa forma, evita-se que as aplicações fiquem temporariamente indisponíveis.

No entanto, para sistemas cujas bases de dados são heterogêneas, a implementação de um serviço de replicação ainda é complexa. Por exemplo, para um banco de dados que guarda o endereço de arquivos externos, um serviço de replicação deve replicar tanto os dados do banco quanto os dados do arquivo. E esta replicação deve ocorrer de tal forma que o banco de dados não deve armazenar o caminho para um arquivo que não existe.

Uma das dificuldades para replicar uma base heterogênea como esta ocorre porque os tradicionais sistemas de replicação se concentram em apenas um tipo de dados. Ou seja, existem serviços de replicação para bancos de dados e serviços de

replicação para arquivos. Para replicar esta base seria necessário garantir que o serviço de replicação de banco de dados atuaria em sincronia com o serviço de replicação de arquivos.

1.2. Objetivos

Os principais objetivos deste trabalho são:

- Projetar uma arquitetura flexível para serviços de replicação de bases heterogêneas. Esta arquitetura utilizará a técnica de projeto orientado a domínio e será baseada em agentes, fornecendo autonomia e escalabilidade aos serviços.
- Implementar um sistema de replicação baseado nesta arquitetura, provendo serviços que garantam a qualidade dos dados replicados, através de versionamento, verificação de consistência e demais regras de negócios. Dessa maneira pretende-se demonstrar que a arquitetura proposta é viável.
- Prover utilitários que facilitem a configuração e manutenção do sistema de replicação, implementando mecanismos que verifiquem se as bases de dados mantêm a consistência transacional e referencial.

1.3. Organização da Dissertação

No capítulo 2 são abordados, de forma introdutória, os principais conceitos utilizados nesta dissertação, tais como replicação de dados, modelagem orientada a agentes e projeto orientado a domínio.

Já no capítulo 3 são descritas as estratégias e decisões que formam a arquitetura, numa visão conceitual. Os principais agentes são abordados em alto nível. Além disso, é passada uma visão rápida sobre os utilitários que colaboram com o sistema de replicação.

Logo após, no capítulo 4, são explicados os detalhes que envolvem as decisões técnicas, tecnologias e testes realizados girando em torno da implementação de referência desenvolvida.

Finalmente, no capítulo 5, são discutidas as contribuições acadêmicas deste trabalho. Também é feita uma visão comparativa deste com trabalhos relacionados. Finalmente, são descritas as propostas para trabalhos futuros.

No Anexo A se encontra um vocabulário para os termos definidos neste trabalho.