

Aplicando a Computação Distribuída no desenvolvimento de um *middleware* para a replicação de dados utilizando como laboratório de pesquisa a Universidade Estadual do Piauí.

Ana Patrícia de Sousa¹, João Batista Oliveira Silva², Laiton Garcia dos Santos³, Rubens dos Santos Lopes⁴

Instituto Federal do Piauí (IFPI) – Campus Floriano (PI) – Brasil

¹anapatricia3636@gmail.com; ²joaobatistatads@gmail.com;
³laitongarcia@hotmail.com; ⁴rubens_roc@outlook.com

Abstract. *Through Distributed Computing techniques, the present study develops a platform to support the management of educational institutions. The system performs the data management of Academic Control and Library of the State University of Piauí (UESPI) - Campus Dr. Josefina Demes. The system implements the data distribution by replication techniques using a middleware. The data of academic control system CAD are fragmented to the library management system Web Library in order to an automatic insertion of users' information of the library. The execution of this project is carried a process of technology transfer between UESPI and the Federal Institute of Piauí - Floriano Campus (IFPI).*

Resumo. *Por meio de técnicas de Computação Distribuída, este estudo desenvolve uma plataforma para gestão de instituições de ensino. O sistema faz a gestão de dados do Controle Acadêmico e da Biblioteca da Universidade Estadual do Piauí (UESPI) - Campus Dr^a Josefina Demes. O sistema implementa a distribuição de dados por meio de técnicas de replicação pelo uso de um middleware. Os dados do sistema de controle acadêmico CAD são fragmentados para o sistema de gestão de bibliotecas Web Library, inserindo automaticamente as informações dos usuários da biblioteca. A execução deste projeto realiza um processo de Transferência de Tecnologia entre a UESPI e o Instituto Federal do Piauí - Campus Floriano (IFPI).*

1. Introdução

Diante da crescente demanda pela disponibilidade, confiabilidade e integridade, requisitos imprescindíveis no âmbito do desenvolvimento de *software*, torna-se cada vez mais necessário que as informações estejam à disposição dos seus usuários no momento que são requisitadas e de forma segura. Os sistemas distribuídos contribuem com isso ao permitir a disponibilidade das informações através do uso de softwares de acesso, e gerenciamento de dados adequados interligados por redes de computadores.

As dificuldades de adaptar sistemas para comunicação e compartilhamento de informações com outros sistemas mais novos e avançados tecnologicamente são grandes e complexas. Assim, torna-se importante o desenvolvimento de sistemas que permitam o acesso e a integração de dados existentes em diferentes fontes.

Nesse contexto, o presente trabalho realizou a aplicação de procedimentos de computação distribuída por meio de técnicas de replicação de dados para o desenvolvimento de um *middleware* orientado a banco de dados, com a função de tornar acessíveis em diversas máquinas às informações, até então, centralizadas no banco de dados do software CAD. Este trabalho descreve ainda os processos seguidos para fragmentar e replicar dados de pessoas para o software gerenciador de bibliotecas *Web Library*. Para a aplicação das técnicas citadas foram utilizadas as melhores práticas de programação, manipulação de dados e configuração de redes de computadores.

2. Revisão da literatura

2.1 Distribuição de Sistemas

O ambiente computacional atual é composto por Sistemas de Informações autônomos, distribuídos e heterogêneos. Os sistemas são autônomos porque sua construção, manutenção e operação ocorrem de forma independente e sem a preocupação com integração com outros sistemas. “De forma mais simplificada, é possível dizer que um sistema autônomo é um grupo conectado, de um ou mais prefixos IP, executados por operadores de rede que têm uma única e bem definida política de roteamento.” [Hawkinson e Bates 1996].

Ainda conforme Tanenbaum e Steen (2007), de modo a suportar máquinas e redes heterogêneas e, ao mesmo tempo, oferecer uma visão de sistema único, os sistemas distribuídos são organizados costumeiramente por meio de uma camada de software, situada logicamente entre duas camadas denominada *middleware*, uma formada por usuários e aplicações e uma camada subjacente formada por sistemas operacionais e facilidades básicas de comunicação.

O *middleware* foi definido como uma camada de software cujo objetivo é mascarar a heterogeneidade e fornecer um modelo de programação conveniente para os programadores de aplicativos. Em particular, ele simplifica as atividades de comunicação de programas aplicativos por meio do suporte de abstração como a invocação a métodos remotos, a comunicação entre um grupo de processos, a notificação de eventos, o particionamento, posicionamento e recuperação de objetos de dados compartilhado entre computadores, a replicação de objetos de dados compartilhados e a transmissão de dados multimídia em tempo real [Coulouris e Dollimore e Kindberg 2007].

A figura 1 mostra a arquitetura de um sistema distribuído organizado como *middleware*. Veja que a aplicação B é distribuída para os computadores 2 e 3. O sistema distribuído oferece formas de comunicação entre os componentes de uma aplicação distribuída e entre diferentes aplicações ocultando, ao mesmo tempo, da melhor forma possível, as distinções entre hardware e sistemas operacionais para cada aplicação, tendo como principal objetivo facilitar o acesso a recursos remotos e seu compartilhamento de maneira controlada e eficiente [Tanenbaum e Steen 2007].

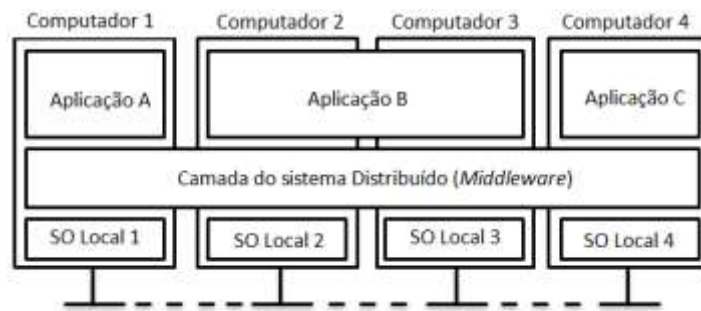


Figura 1- Sistema distribuído organizado como Middleware
Fonte: Tanenbaum e Steen (2007).

2.2 Sistema de Banco de dados distribuídos (SBDD)

Numata (2012) define um sistema de dados distribuídos como a junção de redes de computadores com os sistemas de banco de dados. O autor completa ainda que o sistema de banco de dados distribuídos é um conjunto de vários bancos de dados logicamente inter-relacionados, separados fisicamente em lugares diferentes, geograficamente dispersos, distribuídos por uma rede de computadores.

A figura 2 apresenta a arquitetura de um ambiente em que se configura a distribuição de dados.



Figura 2 – Ambiente de Sistemas de Banco de Dados Distribuídos

Existem duas alternativas básicas de posicionamento dos dados no SBDD: particionados, onde conforme Ozsu e Valduriez (2001), o banco de dados se divide em várias partições disjuntas posicionadas cada uma em um site distinto; e replicados, estes podem subdividir-se em parcialmente replicados (ou parcialmente duplicados) e totalmente replicados (ou totalmente duplicados).

A técnica de replicação de dados é aplicada para aprimorar a confiabilidade ou melhorar o desempenho de sistemas distribuídos. Em um sistema de arquivos replicado é possível continuar trabalhando após a queda de uma réplica, por meio da comutação para uma das outras réplicas, configurando assim a confiabilidade do sistema. A replicação para alcançar desempenho como finalidade é aplicada quando um sistema precisa ser ampliado quantitativa e geograficamente.

Na replicação assíncrona, se um banco é alterado, a alteração será propagada e aplicada para outro(s) banco(s) num segundo passo, sendo que esta poderá ocorrer em segundos, minutos, horas ou até dias depois [Elmasri e Navathe 2005].

2.3 Redes de computadores

Uma rede de computadores pode ser denominada como um conjunto de dispositivos conectados por *links* de comunicação (denominados nós). Um nó pode ser um computador, uma impressora ou qualquer outro dispositivo capaz de enviar e/ou receber dados gerados em outros nós. Atualmente, grande parte das redes usam processamento distribuído para executar uma tarefa entre muitos computadores, tal processo é bem mais eficiente que entregar todo poder de processamento a uma única máquina poderosa e deixá-la responsável por todo gerenciamento da rede [Forouzan 2006].

Ainda conforme Forouzan (2006), as redes de computadores podem ser comparadas conforme alguns critérios. Dentre esses se destacam pelo nível de importância: a performance, a confiabilidade e a segurança.

Há um grande mal entendido entre os conceitos de redes de computadores e sistemas distribuídos. A principal diferença entre esses dois temas é que em sistemas distribuídos, a existência das máquinas autônomas é transparente (ou seja, tem seus detalhes ocultados do usuário) [Tanenbaum 1997].

3. Materiais e Métodos

3.1 Levantamento Bibliográfico

- Foi realizada pesquisa bibliográfica sobre o que são Sistemas Distribuídos suas características e tipos de arquitetura;
- Estudou-se formas de implementar a camada de software que estabelece a abstração de programação, o *Middleware*;
- Estudou-se métodos e técnicas do desenvolvimento *web* utilizando ferramentas de Java EE e JPA.
- Foi levantada bibliografia sobre o funcionamento e alternativas de implementação de Banco de Dados Distribuídos seus conceitos, tipos de sistemas e suas arquiteturas;
- Realizou-se estudo sobre Redes de Computadores suas tecnologias, tipos e arquiteturas e a que se adequaria melhor ao projeto;
- Estudou-se os serviços oferecidos por servidores de modo a utilizar os que melhor se adequassem às necessidades do projeto.

3.2 Implementação

3.2.1 Análise

Foi feita a análise dos requisitos para o correto desenvolvimento do projeto baseado nas necessidades dos usuários para os quais estão direcionadas as plataformas.

3.2.2 Banco de dados

O banco de dados foi implementado utilizando o modelo relacional e desenvolvido na Linguagem de Consulta Estruturada (SQL) que representa o banco de dados como uma coleção de relações. Utilizou métodos necessários ao desenvolvimento do banco de dados da aplicação e técnicas de distribuição de dados, como replicação e fragmentação. O Sistema Gerenciador de Banco de Dados (SGBD) escolhido foi o PostgreSQL por sua

consistência, documentação e ser livre. Também para testes de heterogeneidade foi utilizado o SGBD Oracle.

3.2.3 Aplicação

Foi implementada a camada de *software, middleware*, para se obter as melhores formas de estabelecer comunicação e compartilhamento de informações entre as aplicações envolvidas.

Aplicou-se métodos e técnicas do desenvolvimento *web* para implementação do software de gerenciamento de bibliotecas *Web Library*.

3.2.4 Arquitetura da rede

Para os testes do sistema houve a necessidade de construir uma rede física e lógica para simulação em ambiente real. Os computadores foram interligados por meio de uma rede local (*Local Area Network* - LAN) de topologia estrela, utilizando um modem como dispositivo responsável pela interligação. Foi configurado um servidor de Domínio *Active Directory*, DNS e de Arquivos, para gerenciamento de privilégios de usuários, endereços na rede e cotas de diretórios entre outras tarefas.

4. Resultados

O *middleware* desenvolvido neste trabalho é responsável pela realização da replicação e fragmentação de dados de forma automatizada. O desenvolvimento foi efetuado na linguagem Java utilizando recursos dos próprios SGBDs para obter as informações a serem replicadas. As tarefas realizadas e sua interação com o sistema estão organizadas no diagrama de atividade disposto na **figura 3**.

Os testes realizados, para o compartilhamento de dados entre os sistemas CAD e *Web Library*, não se preocuparam com o desempenho nem com o uso de memória foram focados na funcionalidade do aplicativo visto que aconteceu também em ambiente emulado. Por ser assíncrono superou falhas na comunicação dos sistemas conferindo consistência de informações entre as bases de dados. Existe ainda a necessidade de realizar testes de carga e stress ao *middleware*.

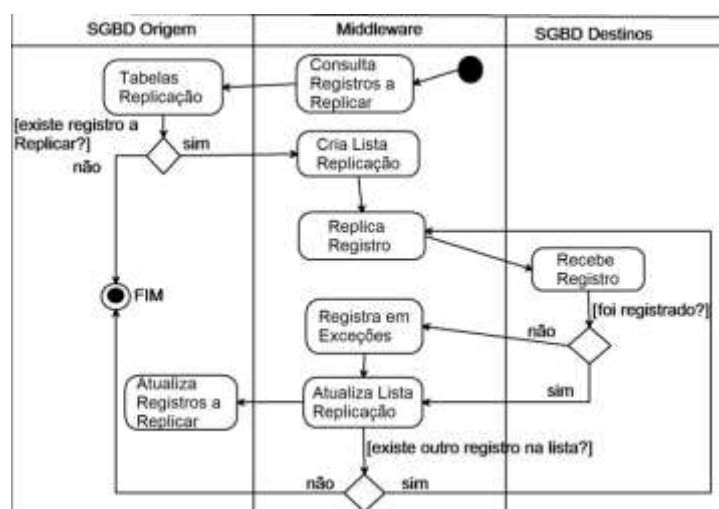


Figura 3: Diagrama de atividade do *middleware*

Conclusão

O trabalho desenvolvido contemplou os objetivos e atendeu os requisitos propostos no seu início. A replicação de dados assíncrona e a consistência das informações em todas as réplicas, juntamente com a possibilidade de atualização dos nodos, mesmo com falhas na comunicação, foi alcançada. Sendo um diferencial comparando-se a outros replicadores existentes que não possuem esta possibilidade como o Slony.

A utilização do Java como linguagem de programação possibilitou que a ferramenta desenvolvida seja executada em sistemas operacionais que deem suporte a ela, e assim permitindo que ela seja utilizada entre plataformas e SGBDs heterogêneos confiando a portabilidade do sistema. A manutenção constante desse sistema se faz necessário para manter sua integridade e segurança bem como melhorar seus diagnósticos de conflitos.

Referências

- Coulouris, G. e Dollimore, J. e Kindberg, T. (2007). Sistemas distribuídos: conceitos e projeto. 4. ed. Porto Alegre: Bookman.
- Elmasri, R. e Navathe, B. S. (2005). Sistemas de Bancos de Dados. São Paulo: Pearson Education.
- Forouzan, B. A (2006). Comunicação de Dados e Redes de Computadores. 3. ed. Porto Alegre: Bookman.
- Gouveia, J. e Magalhães, A. (2007). Redes de Computadores. 1. ed. Rio de Janeiro: FCA Editora de informática.
- Numata, C. A. (2012). Banco de dados distribuídos. São Paulo: Faculdade de Tecnologia de São Paulo.
- Ozsu, T. M. e Valduriez. P. (2001). Princípios de Sistemas de Banco de Dados Distribuídos. 2. ed. Vandenberg D. de Souza. Rio de Janeiro: Campus.
- Tanenbaum, A. S. (1997). Redes de Computadores. Insight Serviços de Informática. 3. ed. Rio de Janeiro: Campus.
- Tanenbaum, A. S. e Steen, M. V. (2007). Sistemas distribuídos: princípios e paradigmas. 2. ed. São Paulo: Pearson Prentice Hall.