

**FACULDADES INTEGRADAS DO BRASIL - UNIBRASIL
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO**

LEONARDO ANTÔNIO DOS SANTOS

**COMPARTILHAMENTO DE ARQUIVOS EM STORAGE DE ALTA
DISPONIBILIDADE**

PROJETO DE TCC

CURITIBA

2012

LEONARDO ANTÔNIO DOS SANTOS

**COMPARTILHAMENTO DE ARQUIVOS EM STORAGE DE ALTA
DISPONIBILIDADE**

Projeto de TCC apresentado ao Curso de Bacharelado em Sistemas de Informação da Faculdades Integradas do Brasil - Unibrasil.

Orientador: Msc. Dionei Marcelo Moraes

Co-orientador: Msc. Pedro Eugênio Rocha

CURITIBA

2012

SUMÁRIO

1	Introdução	3
1.1	Problema	4
1.2	Objetivo Geral	4
1.3	Objetivos Específicos	4
1.4	Justificativa	4
2	Metodologia da Pesquisa	6
2.1	Parte I - Natureza da pesquisa	6
2.2	Parte II - Tipo de pesquisa	6
2.3	Parte III - Levantamento de requisitos	6
2.4	Descrição e Justificativa das Tecnologias Aplicadas no Trabalho	6
2.5	Arquitetura da Solução	7
2.6	Análise de Performance	7
3	Conclusão	8
4	Cronograma	9
	REFERÊNCIAS	10

1 INTRODUÇÃO

As tecnologias têm se tornado cada vez mais essenciais ao dia a dia das empresas e das pessoas. Grandes corporações têm utilizado sistemas informatizados para gerenciar as suas linhas de produção, sistemas de controle de estoque e compartilhamento de arquivos. Especialmente sobre o compartilhamento de arquivos, os volumes de dados têm aumentado em proporção cada vez maior. Basta observar o crescimento de serviços de armazenamento de dados famosos como YouTube(1), 4shared(2) e Source Forge(3).

Além da necessidade de compartilhamento de grandes quantidades de dados, grande parte das empresas possuem informações que precisam ser armazenados de forma segura e com alta disponibilidade, ou seja, a informação necessita estar disponível quase todo tempo, senão todo o tempo.

Assim, é fácil pensar em juntar vários HDs (*Hard Disk* - Disco Rígido) a fim de solucionar o problema do compartilhamento de arquivos. Mas esta solução, apesar de parecer fácil, na prática não funciona lançando mão para isto de métodos de compartilhamento de arquivos tradicionais, como CIFS, NFS ou SAMBA - protocolos para compartilhamento de arquivos para sistemas de arquivos (*filesystems*) como *Windows* ou *Linux*. A justificativa para isto é porque estes métodos compartilham apenas os arquivos e não os dispositivos de armazenamento, tornando inviável o gerenciamento deste dispositivo de forma remota.

Neste contexto, surgiram outros protocolos para compartilhamento de dispositivos de armazenamento de dados (ou *block devices*), sendo os principais deles o iSCSI(4) e o Fibre Channel(5). Ambos os protocolos cumprem os objetivos do armazenamento de dados de forma segura, com alta disponibilidade e para grandes quantidades de arquivos. Os principais problemas que envolvem estes protocolos são o alto custo de tecnologia e a dependência de hardware específico para implementação.

O ATA Over Ethernet é um protocolo *open source* (código aberto) desenvolvido pela empresa Coraid (6) e disponibilizado à comunidade de ciências da computação a fim de ser estudado e melhorado em sua implementação. Ele resolve os problemas mais comuns do Fibre-

Channel e iSCSI que são, respectivamente, alto custo e sobrecarga de rede.

1.1 PROBLEMA

Falta de uma área de armazenamento de dados compartilhada entre todo o corpo docente e discente da UniBrasil.

Alto custo nas tecnologias de armazenamento de dados para compartilhamento de alta disponibilidade.

1.2 OBJETIVO GERAL

Implementar o uso de um protocolo e uma aplicação web que, integrados, permitam compartilhar arquivos com alta disponibilidade na intranet da UniBrasil e na internet.

1.3 OBJETIVOS ESPECÍFICOS

- Estudar sobre protocolos de rede, alta disponibilidade de dados, análise de performance e aplicações web;
- Avaliar tecnologias paralelas que já encontram-se em uso no mercado;
- Implementar o uso do protocolo adaptando-o a uma infraestrutura capaz de compartilhar dados pela intranet e pela internet via web de forma distribuída;
- Desenvolver sistema web que facilite o uso desta tecnologia ao usuário final.

1.4 JUSTIFICATIVA

O compartilhamento de arquivos, apesar de ser simples utilizando as ferramentas conhecidas, torna-se muito custoso quando se busca implementar grandes áreas de armazenamento de dados, principalmente se forem distribuídos e com alta disponibilidade. Isto acontece devido ao fato desta implementação ser possível somente com a compra de equipamentos especializados para este fim.

O protocolo AoE aparece com a finalidade de resolver problemas de compartilhamento de arquivos, só que a um baixo custo de tecnologia, se comparado a protocolos para filesystems (sistemas de arquivos) distribuídos.

Uma implementação completa deste protocolo juntamente com uma ferramenta Web que facilitará e orquestrará o seu uso.

2 METODOLOGIA DA PESQUISA

2.1 PARTE I - NATUREZA DA PESQUISA

Redes de Computadores e Sistemas Distribuídos (Protocolos de Comunicação, Arquitetura de Clusters e Servidores, Sistemas Distribuídos, Alta Disponibilidade).

2.2 PARTE II - TIPO DE PESQUISA

Aplicada e Bibliográfica.

2.3 PARTE III - LEVANTAMENTO DE REQUISITOS

Levantamento das necessidades de compartilhamento de arquivos utilizando-se de mecanismos como entrevistas e análise dos sistemas existentes.

2.4 DESCRIÇÃO E JUSTIFICATIVA DAS TECNOLOGIAS APLICADAS NO TRABALHO

As principais tecnologias utilizadas serão:

PROTOCOLOS:

AoE (Ata over Ethernet): protocolo utilizado na comunicação entre os dispositivos de armazenamento de dados, ou discos rígidos.

LDAP (Lightweight Directory Access Protocol): protocolo de autenticação dos usuários que terão acesso aos dados compartilhados. Possui sua base de dados centralizada.

LINGUAGENS:

PHP (Hypertext Preprocessor): linguagem de programação web que será utilizada no desenvolvimento do site para compartilhamento de dados entre os usuários.

MySQL: banco de dados onde ficarão armazenados os meta-dados dos arquivos armazenados, assim como os respectivos relacionamentos destes arquivos com os seus usuários.

FERRAMENTAS:

AoE Tools: conjunto de ferramentas utilizadas para implementação do protocolo AoE.

CakePHP Development Framework 1.3: framework que agiliza o desenvolvimento de sites através da linguagem PHP.

MySQL Server 5.5: implementação do servidor de banco de dados MySQL.

OpenLDAP: Implementação *open source* do servidor de diretórios LDAP.

Linux Debian 5: sistema operacional *open source* que será utilizado para a implementação.

Apache HTTP Server 2.2: servidor web *open source*.

2.5 ARQUITETURA DA SOLUÇÃO

A arquitetura para a solução apresentada é semelhante à Figura 1.

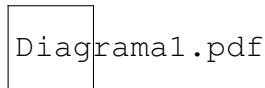


Figura 1: Arquitetura da Solução.

2.6 ANÁLISE DE PERFORMANCE

Serão executados testes de análise de performance sobre o protocolo, a rede, os discos e outras tecnologias envolvidas, como o servidor Web e o servidor de banco de dados.

3 CONCLUSÃO

Este trabalho visa facilitar o uso de um protocolo código-aberto para o compartilhamento de arquivos em alta disponibilidade, tornando esta tarefa acessível a qualquer pessoa que assim o desejar.

Com a crescente demanda sobre esta área, principalmente no que se refere à virtualização de servidores, o que necessita de um armazenamento distribuído, este trabalho tenta trazer simplicidade ao uso de protocolos de armazenamento de dados para estas novas tendências, em especial por este recurso poder ser utilizado por qualquer pessoa, por conta de sua característica de software livre.

Para trabalhos futuros, podem ser efetuadas mudanças no próprio protocolo, visando o aumento da sua performance em situações específicas, como no caso da sua utilização em conjunto com RAID (*Redundant Array of Independent Drives*).

4 CRONOGRAMA

O cronograma 2 representa apenas uma hipótese temporal projeto a ser desenvolvido, podendo ser adaptado ao longo do projeto.

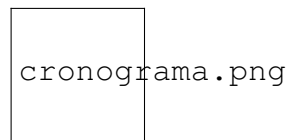


Figura 2: Cronograma do Projeto.

REFERÊNCIAS

- 1 YOUTUBE - Broadcast Yourself. [S.l.]: 2011 YouTube, LLC, 2011. <http://www.youtube.com/>. Acesso em: 20 set. 2011.
- 2 4SHARED.COM - free file sharing and storage. [S.l.]: 4shared.com, 2005. <http://www.4shared.com/>. Acesso em: 20 set. 2011.
- 3 SOURCEFORGE.NET: Find, Create, and Publish Open Source software for free. [S.l.]: 2011 Geeknet, Inc, 2011. <http://sourceforge.net/>. Acesso em: 20 set. 2011.
- 4 SATRAN, J. et al. **Internet Small Computer Systems Interface (iSCSI)**. IETF, abr. 2004. RFC 3720 (Proposed Standard). (Request for Comments, 3720). Updated by RFCs 3980, 4850, 5048. Disponível em: <<http://www.ietf.org/rfc/rfc3720.txt>>.
- 5 CLARK, T. **Designing storage area networks: a practical reference for implementing Fibre Channel SANs**. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999. ISBN 0-201-61584-3.
- 6 HOPKINS, S.; COILE, B. **ATA over Ethernet Specification**. [S.l.]: The Brantley Coile Company, Inc, 2001. <http://support.coraid.com/documents/AoEr11.txt>. Acesso em: 20 set. 2011.