

## Armazenamento de Imagens Médicas DICOM em Sistemas de Arquivos Distribuídos

Douglas D. J. de Macedo<sup>1,2</sup>, Hilton G. W. Perantunes<sup>1,2</sup>, Aldo von Wangenheim<sup>1</sup>, M. A. R. Dantas<sup>2</sup>

Universidade Federal de Santa Catarina

Departamento de Informática e Estatística

<sup>1</sup>Laboratório de Telemedicina - LABTELEMED

<sup>2</sup>Laboratório de Pesquisa em Sistemas Distribuídos – LAPESD

88040-900 – Florianópolis – SC, Brasil

{macedo,william,awangenh,mario}@inf.ufsc.br

### Resumo

*Métodos para armazenamento e recuperação de informações em sistemas de telemedicina, em sua maioria estão relacionados a utilização de bancos de dados convencionais. Assim, desafios relacionados a escalabilidade, distribuição, desempenho e o alto custo computacional envolvido, induz em pesquisas de novas abordagens para solução destes problemas. Diante disto, neste trabalho de pesquisa utilizou-se um meio alternativo de alto desempenho para o armazenamento de imagens médicas DICOM. Foi proposto um modelo de dados, baseado na hierarquização das imagens, utilizando o formato de dados HDF5. Para distribuição dos dados, foi utilizado o sistemas de arquivos distribuídos PVFS, como plataforma básica para armazenamento. Desta forma, a presente pesquisa apresenta um método diferenciado de armazenamento e de acesso aos dados. Nos resultados experimentais, o armazenamento das informações obteve um desempenho da ordem de 17%, em relação a recuperação que ocorreu na casa dos 27%. Estes dados indicam um desempenho superior no armazenamento, quando comparado ao meio tradicional utilizando bancos de dados relacionais.*

### 1. Introdução

Na década de 60 [6] as primeiras aplicações de Telemedicina surgiram, e desde então é sugerida e aplicada como uma forma de prover acesso à saúde de pessoas que estejam isoladas ou desprovidas de acesso a pessoal médico qualificado. Neste tipo de abordagem, a sua utilização pode reduzir o custo de transporte de pacientes e maximizar a utilização do parque tecnológico instalado em hospitais e clínicas médicas [15].

No decorrer destes anos, a Telemedicina e conseqüentemente os Sistemas de Telemedicina evoluíram e vêm se disseminando por muitos países. Os principais fatores que culminaram neste desenvolvimento foram a sofisticação das redes de comunicação, em conjunto com os protocolos de rede, sistemas operacionais e os equipamentos de nova geração para imagens médicas.

Entretanto, ainda existem muitos desafios que cerceiam este crescimento, que vão desde melhores regulações e normas de boas práticas na medicina até melhores infraestruturas para suportar sua disseminação em escala. Por estes motivos, atualmente a grande maioria destes sistemas estão instalados dentro dos hospitais. Visando a ampliação destes serviços para fora destes limites, existem muitos desafios tecnológicos a serem resolvidos, tais como: indexação de informações de forma distribuída [12], mecanismos de integração mais eficazes [11], métodos de armazenamento alternativo aos bancos de dados (objetivo deste trabalho), entre outros.

É importante salientar que o desenvolvimento de novos métodos para tratamento destes problemas devem proporcionar níveis de escalabilidade sustentáveis e arquiteturas que suportem a inserção de mecanismos de alto desempenho. Esta problemática advém do fato de que bancos de dados médicos, que são compostos de dados de pacientes, imagens e vídeos, no geral podem ultrapassar facilmente vários *terabytes* de tamanho.

O projeto chamado Rede Catarinense de Telemedicina RCTM [14] do Estado de Santa Catarina, é um exemplo do uso de Telemedicina no Brasil. Este projeto é desenvolvido e mantido pelo Laboratório de Telemedicina [4] da Universidade Federal de Santa Catarina (UFSC) e é constituído por uma série de serviços para auxílio a saúde [19]. Entre estes serviços está o objeto de estudo deste trabalho, o CyclopsDCMServer, um servidor de imagens

médicas DICOM, projetado para trabalhar em redes de longas distâncias e que grava todas suas informações em bancos de dados relacionais.

No modelo adotado pela RCTM [13], os hospitais dos municípios realizam os exames nos pacientes, tais como: eletrocardiogramas, hemodinâmicas, cintilografias, tomografias e ressonâncias. Estes exames são enviados em tempo real no formato de imagens médicas DICOM para o *CyclopsDCMServer*, que segmenta as informações e trata as imagens, as enviando então para um banco de dados centralizado, no Hospital Universitário. Os médicos então podem acessar as informações dos exames, através do Portal de Telemedicina [19] para efetuar os procedimentos necessários com os dados.

Pensando neste problema, este trabalho abordará o uso de sistemas de arquivos distribuídos, baseados em agregados computacionais, para o armazenamento destas imagens DICOM. Entretanto, armazenar imagens em formato “puro” impossibilitaria ou dificultaria uma futura recuperação das informações. Para isto, as informações serão hierarquizadas e armazenadas no formato de dados HDF5. Desta forma, esta abordagem terá uma camada de alta performance para o armazenamento das imagens e metadados hierarquizados, o que possibilitará a recuperação dos dados no meio distribuído.

No restante deste artigo será apresentada uma breve revisão bibliográfica, alguns aspectos de trabalhos relacionados, a pesquisa realizada e os resultados experimentais. Assim, na seção 2 será apresentado um embasamento teórico sobre imagens médicas e ao formato de dados HDF5. Na seção 3 serão relacionados e comentados alguns trabalhos correlatos ao tema. Na seção 4, a arquitetura proposta do trabalho será apresentada. Na seção 5, serão discutidos o ambiente de trabalho, os experimentos realizados, bem como os resultados experimentais. Por fim, na seção 6 serão apresentadas as conclusões e os possíveis temas para trabalhos futuros.

## 2. Imagens Médicas e Armazenamento

### 2.1. Imagens Médicas

Com o avanço tecnológico, na década de 70, surgiu uma nova modalidade de equipamentos para diagnóstico médico, a Tomografia Computadorizada (*Computed Tomography*). Um exemplo deste tipo de imagem pode ser visualizado na Figura 1. Ela introduziu o uso de técnicas de arquivamento e transmissão de imagens digitalizadas nos ambientes hospitalares.

Após o surgimento deste tipo de exame, outras modalidades de diagnóstico elaborados por imagens digitais emergiram. Porém, a inexistência de um padrão único de troca de imagens tornou-se um problema, frente a necessidade de



Figura 1. Tomografia Computadorizada

interconectar sistemas de diferentes fabricantes. Por este fato, cada fornecedor criava sua própria solução para arquivamento, impressão, transmissão e visualização de imagens digitais e informações de cada modalidade de exame [8].

A partir deste problema o Colégio Americano de Radiologia (ACR – *American College of Radiology*), em conjunto com a Associação de Fabricantes de Equipamentos Elétricos dos Estados Unidos (NEMA – *National Electrical Manufacturers Association*), reconheceram a necessidade da existência de um padrão único para a intercomunicação entre equipamentos fabricados por diferentes fornecedores.

A primeira versão deste padrão foi publicada em 1985 e foi denominada ACR-NEMA Standards Publication No. 300-1985. Logo após a primeira versão, ocorreram duas revisões, em 1986 e 1988, sendo a última denominada ACR-NEMA Standards Publication No. 300-1988. Estas regulações proviam especificações de uma interface de hardware, um conjunto de formatos de dados e um conjunto mínimo de comandos de software.

A maior deficiência deste padrão era em relação ao suporte aos ambientes de redes de computadores. Por este motivo, o padrão resistiu até 1992, quando foi lançada a ACR-NEMA Standards Publication PS3, também chamada de DICOM 3 (*Digital Imaging and Communications in Medicine*), um padrão muito mais robusto que os primeiros.

Atualmente o DICOM 3 é o padrão de fato para os Sistemas de Arquivamento e Comunicação de Imagens (PACS – *Picture Archiving and Communications System*), sendo suportado pela grande maioria dos aparelhos que trabalham com informações médicas digitais. O fato de um aparelho o suportar é a garantia de que este poderá ser facilmente integrado em um PACS já existente devido à utilização de tecnologias de rede acessíveis e baratas para a sua implementação, e ao fato de este poder utilizar-se de serviços disponibilizados por outros aparelhos que também suportem o padrão DICOM [8].

A partir da versão 3 do DICOM, o Laboratório de

Telemedicina da Universidade Federal de Santa Catarina (UFSC), desenvolveu o seu próprio servidor para armazenamento e recuperação de imagens médicas, chamado CyclopsDCMServer. Ele foi desenvolvido para funcionar em ambientes PACS, dentro de hospitais ou clínicas radiológicas. Entretanto, devido às necessidades de expansão, ele foi remodelado para trabalhar em redes de longa distância, onde foram incluídos aspectos de tolerância a falhas, tratamento de erros, entre outros aspectos [16].

Um dos principais objetivos deste servidor é armazenar e fornecer arquivos DICOM indexados em um repositório de dados gerenciado por SGBDs relacionais, tais como: PostgreSQL, MySQL, Oracle, etc. Toda a comunicação entre os equipamentos médicos e o servidor é realizada através de redes de comunicação baseadas em TCP/IP [1]. Atualmente o CyclopsDCMServer suporta as seguintes modalidades de exames médicos:

- Radiografia Computadorizada (CR)
- Tomografia Computadorizada (CT)
- Ressonância Magnética (MR)
- Medicina Nuclear (NM)
- Ultra-Som (US)
- Cintilografia (XA)
- Eletrocardiogramas (ECG)

## 2.2. Formato de Dados Hierárquico

O Formato de Dados Hierárquico [2] (HDF – *Hierarchical Data Format*) é um formato portátil de dados desenvolvido no parque de pesquisas da Universidade de Illinois nos Estados Unidos. Trata-se de uma biblioteca para manipulação de alto desempenho de dados científicos, que possibilita o armazenamento de objetos com grandes quantidades de dados, como *arrays* multi-dimensionais e tabelas com grandes conjuntos de dados, que podem ser utilizados em conjunto, de forma que atenda às aplicações.

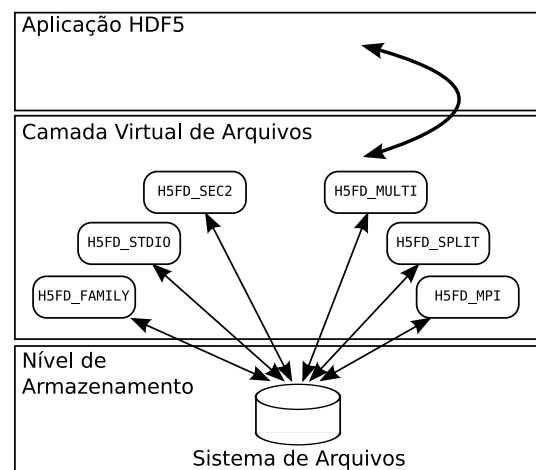
A sigla HDF também se refere ao conjunto de softwares, interfaces de aplicativos e utilitários que compõem a biblioteca e que possibilitam aos usuários a manipulação de arquivos no formato. O software está em desenvolvimento desde 1988 e atualmente encontra-se em sua quinta versão, a HDF5. Esta, por questões de projeto, não possui compatibilidade com as versões anteriores, mas possui diversos aprimoramentos em relação a elas, em especial no que se refere ao acesso paralelo a dados [18].

O HDF5 implementa um modelo para o gerenciamento e armazenamento de dados que é um modelo abstrato de representação e armazenamento. Também estão presentes bibliotecas para implementar o modelo e mapeá-lo para

diferentes mecanismos de armazenamento. A biblioteca também implementa um modelo de transferência de dados que provê uma forma eficiente de transferência de uma representação de armazenamento para outra representação. Entre suas características, destaca-se a capacidade de auto-descrição através de metadados sobre o conteúdo dos arquivos, flexibilidade no uso de diferentes tipos de dados em um único ou múltiplos arquivos, portabilidade entre plataformas, padronização de tipos e formatos de dados e o fato de a biblioteca ser um software de código aberto.

O seu modelo de programação suporta desde pequenos sistemas até grandes multi-processadores e clusters, manipulando por meio de instanciação, registro e recuperação de dados, objetos do modelo abstrato. A biblioteca é a implementação concreta do modelo de programação, e exporta as APIs HDF5 e suas interfaces.

As estruturas conceituais mais importantes do HDF5 são o *dataset* e o *group*. Um *dataset* é um *array* multi-dimensional de elementos de um tipo de dados (*datatype*) especificado. Os tipos de dados podem ser atômicos (números inteiros, de ponto flutuante, cadeias de caracteres, etc) ou tipos compostos. Os grupos são similares a estruturas de diretório, provendo uma forma de organizar explicitamente os conjuntos de dados em um arquivo HDF5.



**Figura 2. Relacionamento Entre os Modelos**

A biblioteca HDF5 possui uma Camada Virtual de Arquivos (VFL – *Virtual File Layer*), que consiste de uma API para o tratamento de operações de I/O em baixo nível, de forma que uma aplicação possa gravar dados utilizando diferentes recursos de armazenamento [17], como pode ser observado na Figura 2.

O Modelo Abstrato de Dados (ADM – *Abstract Data Model*) define conceitos para a descrição de dados complexos armazenados nos arquivos. É um modelo genérico projetado para conceitualmente abranger diversos modelos específicos de dados. Diferentes tipos de dados podem ser

mapeados para objetos do modelo de dados HDF5, e conseqüentemente armazená-los e recuperá-los utilizando a biblioteca HDF5 [2].

O HDF5 Lite possui dois conjuntos de funções, relacionados à manipulação de *datasets* e *atributos*. Os métodos da API consistem de funções de maior nível de abstração que realizam mais operações por chamada do que as funções básicas da biblioteca HDF5. O propósito é servir como interface por meio de funções intuitivas que utilizam conjuntos específicos de características da API existente. Mesmo com a API de alto nível, porém, algumas funções de criação e acesso aos arquivos do HDF5 básico podem precisar ser utilizadas por uma aplicação.

Por outro lado, a API HDF5 Image define um sistema padrão de armazenamento para *datasets* HDF5 que espera-se que sejam interpretados como imagens. Esta versão da API possui métodos para armazenar imagens indexadas de 8 bits, nas quais cada *pixel* é armazenado como um índice em uma paleta de cores, e imagens *true color* de 24 bits, em que o armazenamento de cada *pixel* contém os planos de cores vermelho, verde e azul.

Ambas as APIs serão utilizadas em uma ferramenta construída para servir como ponte de acesso entre o servidor de imagens DICOM e as funções da biblioteca HDF5 para a criação de arquivos neste formato a partir de um conjunto de imagens médicas. Na próxima seção serão apresentados alguns trabalhos relacionados à implementação de sistemas de armazenamento baseados em HDF5.

### 3. Trabalhos Relacionados

A aplicação do HDF5, para manipulação de dados médicos é um tema ainda muito pouco estudado. Um dos motivos é pelo fato de que a maioria dos servidores de imagens médicas, vem com *drivers* específicos para armazenamento de informações em bancos de dados relacionais. Entretanto, o HDF5 é utilizado em uma vasta área de aplicação, com diferentes abordagens, mas principalmente em aplicações científicas e tecnológicas, para tratamento de grandes quantidades de informações, como podemos constatar nos trabalhos de Cohen *et al.* [7], Gosink *et al.* [9], Lee e Hung [10] e Yu *et al.* [20].

No trabalho de Cohen *et al.* [7] é apresentado um estudo baseado em extensões de sistemas gerenciadores de bancos de dados relacionais, que permitem a representação de dados científicos e operações estatísticas comuns. Foram utilizados o NetCDF e o HDF, que são os dois dos formatos científicos mais populares. Neste trabalho, ainda foram realizadas operações estatísticas, usando as extensões dos SGBDs em comparação com as operações nativas do NetCDF e do HDF.

Gosink *et al.* [9] apresentam uma nova abordagem para acelerar os acessos a grandes arquivos HDF5, uti-

lizando indexação semântica multi-dimensional, denominada HDF5-FastQuery. Os resultados deste trabalho demonstraram que esta nova abordagem é duas vezes mais rápida que os métodos padrões do HDF5.

Lee e Hung [10] descrevem o desenvolvimento de uma ferramenta, elaborado pelo time CERES (*Clouds and Earth's Radiant Energy System*) de Gerenciamento de Dados, que é parte do Sistema de Observação da Terra da NASA. Esta ferramenta tem a função de visualização gráfica via interface com o usuário para análises de *datasets* HDF.

Yu *et al.* [20] realizam em seu trabalho uma série de avaliações de desempenho em gravações de dados paralelas. São realizados vários testes de leitura e escrita em interfaces paralelas de alto nível tais como HDF e o NetCDF, em ambientes com um grande número de processadores, no caso é utilizado o supercomputador Blue Gene/L.

### 4. Arquitetura Proposta

Com o uso da tecnologia descrita, foi iniciada a preparação de um ambiente em que imagens médicas no formato DICOM pudessem ser armazenadas em um meio distribuído no formato HDF5 em adição aos métodos já descritos, como gravação em banco de dados relacionais. Tal ambiente possibilita a realização de comparações, buscando analisar os custos de tempo de gravação e recuperação entre os dois métodos.

O primeiro passo para a preparação do ambiente, foi a criação do HDF5 Wrapper Library (H5WL). O H5WL é uma biblioteca composta de um objeto *wrapper* que contém os métodos de criação e armazenamento de informações específicas de imagens DICOM, obtidas através do servidor de imagens. O objeto é utilizado como um módulo a parte do CyclopsDCMServer e instanciado de forma similar aos outros métodos de armazenamento disponíveis no servidor. Esta arquitetura pode ser visualizada na Figura 3.

Em seguida, foi construído um sistema compartilhado de arquivos entre os diversos nodos de um agregado computacional. Esse sistema foi obtido por meio da utilização do *Parallel Virtual File System* (PVFS) [5] em cada um dos nodos do agregado. No momento em que o CyclopsDCMServer solicita a criação de um novo arquivo HDF5, este é criado diretamente no ambiente compartilhado.

Para a leitura inicial dos dados, foi necessário separá-los em duas categorias: elementos de dados comuns e imagem. Os elementos de dados comuns são informações sobre a imagem, como dados sobre o paciente (Nome, CPF, RG, etc.) para o qual a imagem foi gerada, dimensões e outras características da imagem. No caso da imagem, é a representação binária da figura gerada pelo equipamento médico, como uma tomografia ou cardiograma, por exemplo. Para o armazenamento no formato HDF5 optou-se

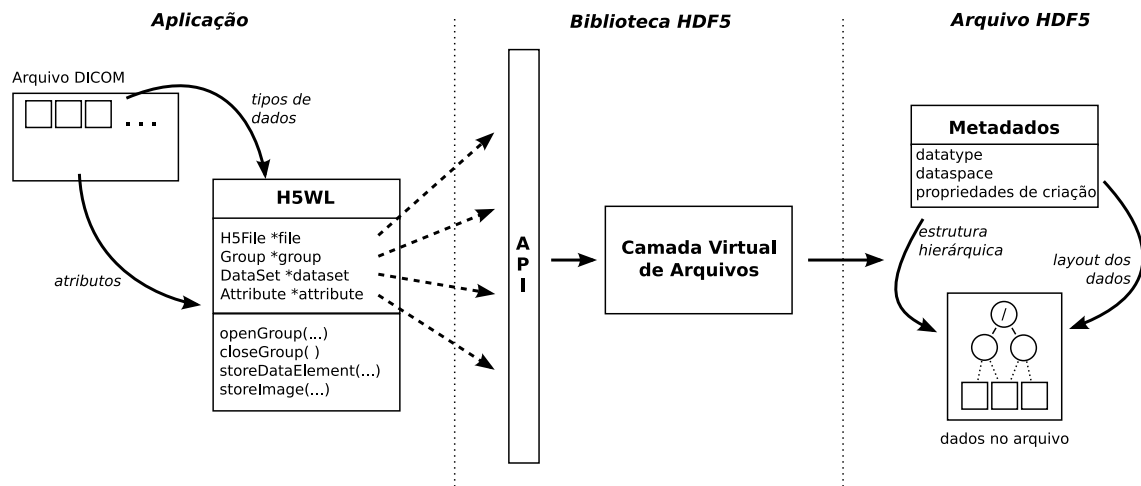


Figura 3. Modelo de Operação

por referenciar todos os elementos de dados comuns como *strings*, enquanto as imagens deveriam ser armazenadas em uma representação numérica gerada pela API HDF5 Image (H5IM) [2], própria para este fim.

Outro aspecto considerado durante o projeto foi a estrutura hierárquica adotada para representar cada imagem dentro de um arquivo HDF5. Considerando as informações disponíveis dentro das imagens em formato DICOM, optou-se por utilizar a estrutura exibida na Figura 4. As cinco camadas da hierarquia (abaixo do grupo raiz) possibilitam uma organização satisfatória para o experimento realizado, permitindo uma boa visualização, um excelente desempenho e praticidade de acesso aos dados.

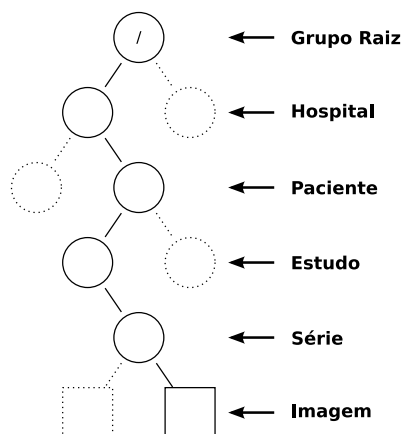


Figura 4. Hierarquização dos Dados em HDF5

No CyclopsDCMServer, no momento da obtenção dos dados da imagem, a hierarquia é criada por meio de chamadas aos métodos do H5WL que cria grupos, sendo ne-

cessário informar o “caminho” do grupo que está sendo criado no arquivo HDF5 e um nome, que será utilizado como identificador dentro do arquivo. Uma vez que o objeto esteja aberto no grupo desejado, dois métodos de acesso estão disponíveis para o armazenamento dos dois diferentes tipos de informações obtidos do arquivo original. Tais métodos também permitem que informações sobre esses dados sejam recebidos pelo objeto *wrapper* e associados às informações originais.

Uma vez que todas as informações de um determinado arquivo DICOM sejam obtidas e gravadas no arquivo HDF5, um ou mais grupos são fechados e uma nova iteração é realizada para o armazenamento de uma nova imagem no mesmo arquivo. Quando o ciclo acaba, observa-se a existência de possíveis erros em sua criação, e em caso negativo o processo de gravação iniciado pelo CyclopsDCM-Server é encerrado.

## 5. Resultados Experimentais

Para avaliar o modelo proposto, decidiu-se por realizar um estudo comparativo entre o armazenamento das imagens médicas no formato HDF5 e pelo método atualmente utilizado pelo CyclopsDCMServer, que é a gravação dos dados das imagens em um banco de dados relacional *PostgreSQL*. Neste banco de dados as informações sobre as imagens DICOM obedecem à seguinte estrutura: quatro tabelas para armazenar os elementos de dados comuns e meta-informações sobre as imagens DICOM e uma tabela com o objetivo de gravar os arquivos em formato binário, como *large objects*.

## 5.1. Ambiente

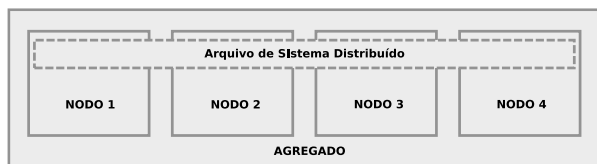
Todos os cenários avaliados foram desenvolvidos em um ambiente de teste. Para isto, foram utilizadas 4 máquinas, descritas na Tabela 1. Estas máquinas foram interconectadas por meio de uma LAN (*Local Area Network*) de 100Mbps. Todas as máquinas utilizam o sistema operacional Debian GNU/Linux, com *kernel* versão 2.6.22.

**Tabela 1.** Especificações das Máquinas do Ambiente

Nome	Descrição
Master	Intel Celeron 2.53GHz, 512Mb RAM, 40Gb HD
Nodo1	Intel Celeron 2.53GHz, 256Mb RAM, 40Gb HD
Nodo2	Intel Celeron 2.53GHz, 256Mb RAM, 40Gb HD
Nodo3	Intel Celeron 2.53GHz, 256Mb RAM, 40Gb HD

Para a realização dos experimentos sobre a arquitetura proposta anteriormente, um ambiente de testes foi criado de forma a atender aos requisitos necessários para a execução da ferramenta. Foi montado um agregado computacional composto por quatro nodos, cada um possuindo uma área compartilhada de disco rígido sob o PVFS.

Na Figura 5 pode-se visualizar um ambiente clássico de sistemas de arquivos distribuídos em agregados computacionais. Nela, os nodos do agregado computacional estão compartilhando espaço de armazenamento em disco, formando assim, um único grande volume de armazenamento.

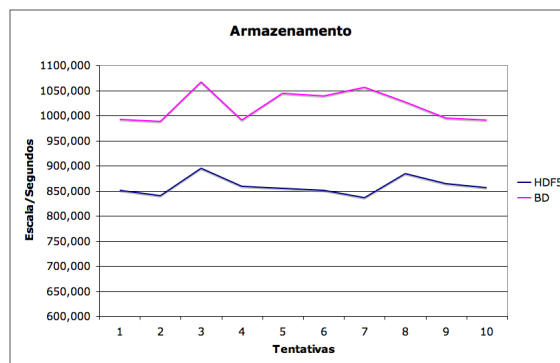


**Figura 5.** Sistemas de Arquivos Distribuídos

## 5.2. Experimentos

O primeiro experimento consistiu na gravação de um conjunto de dados, composto por **2.570** imagens DICOM, ocupando um espaço total em disco de **1.019** Mb. Os arquivos foram originalmente organizados pelo nome do paciente, sendo que cada paciente pode possuir um ou mais estudos com diversas séries de imagens. As imagens são geradas por equipamentos de tomografia computadorizada, gerando figuras monocromáticas de 512 *pixels* de altura e largura.

Na Figura 6 é possível visualizar um gráfico dos resultados obtidos no primeiro experimento, relativo ao armazenamento das imagens no ambiente distribuído. Pela análise



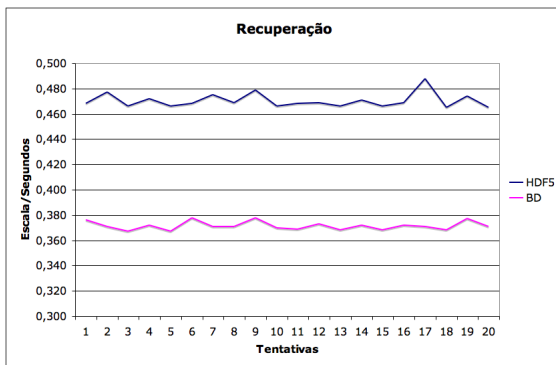
**Figura 6.** Gráfico de Desempenho de Armazenamento

dos resultados obtidos, pode-se perceber que existe um ganho de desempenho significativo no armazenamento dos arquivos DICOM no formato HDF5 em sistemas de arquivos distribuídos, em comparação ao armazenamento em um banco de dados relacional. Também é possível observar a menor variação nos tempos de armazenamento no sistema de arquivos distribuído. As médias de tempo foram de **858,73** segundos para a gravação no PVFS em formato HDF5 e de **1018,77** segundos para a gravação no banco de dados relacional.

Uma segunda experiência foi realizada para a obtenção dos tempos de acesso a uma imagem específica, novamente em um banco de dados relacional e em um arquivo HDF5 gravado em um sistema de arquivos distribuídos. As tabelas utilizadas na consulta e o arquivo armazenado são as mesmas utilizadas na obtenção dos tempos de gravação dos dados. Iterações foram realizadas no acesso aos dados de uma imagem escolhida de forma aleatória, para ser recuperada das duas estruturas de armazenamento. No banco de dados relacional, a consulta aos dados da imagem foi realizada de forma que ao final da consulta a imagem estivesse carregada, na memória principal do servidor, em formato binário, pronta para ser armazenada em disco.

Já no acesso ao arquivo no formato HDF5, devido ao fato de os dados não estarem armazenados em formato binário, foi criada uma extensão no H5WL de forma que os dados de uma imagem DICOM pudessem ser armazenados em uma estrutura de dados projetada apenas para recuperar para a memória os elementos de dados de uma imagem específica, a fim de simular o mesmo método que o banco de dados trabalha.

Na Figura 7 é apresentado um gráfico relativo aos resultados do segundo experimento. É possível perceber que existe uma diferença constante entre os tempos obtidos a partir do banco de dados em comparação aos acessos ao arquivo HDF5, sendo que a vantagem agora passa a ser do banco de dados relacional.



**Figura 7. Gráfico de Desempenho da Recuperação**

Na Figura 8, é possível observar o registro dos tempos, em segundos, utilizado na gravação e leitura do conjunto de imagens para o servidor do banco de dados relacional e para o agregado computacional com sistema de arquivos distribuído, PVFS. É possível visualizar também que a leitura utilizando o método HDF5 foi cerca de **26,34%** mais lento, se comparando com a recuperação das imagens utilizando bancos de dados. Entretanto, a gravação das imagens DICOM hierarquizadas, utilizando sistemas de arquivos distribuídos foi **16,71%** superior.

Método	Leitura		Gravação	
	HDF	BD	HDF	BD
Min.	0,465	0,367	835,660	988,079
Max.	0,488	0,378	894,973	1066,320
Média	0,470	0,372	858,729	1018,765
Desvio (%)	26,34%		16,71%	

**Figura 8. Resultados**

## 6. Conclusões e Trabalhos Futuros

Neste trabalho foi abordado um método para armazenamento alternativo para imagens médicas DICOM, utilizando o formato de dados HDF5 em sistemas de arquivos distribuídos, em comparação com o método tradicional utilizando bancos de dados relacionais. Foi elaborado um modelo de dados a ser utilizado na hierarquização das imagens e ainda um modelo de acesso aos dados, utilizando o H5WL. Por fatores elucidativos, é possível visualizar no site oficial do H5WL [3] um exemplo simples de uma imagem DICOM formatada com o HDF5.

Como resultado deste estudo, podemos citar que no armazenamento das imagens médicas em sistemas de arquivos distribuídos, sob o formato HDF5, foi constatado que mesmo teve um desempenho cerca de 16,71% superior

comparando com o método usando bancos de dados. De outra forma, na recuperação das imagens o método utilizando bancos de dados foi superior cerca de 26,34%, em comparação a arquitetura proposta. É importante salientar, que nesta pesquisa não foram utilizados métodos de acessos e escritas paralelas no ambiente distribuído, sendo este tema de trabalhos futuros.

Em adição, um outro resultado alcançado pelo trabalho, é um modelo de operação e acesso aos dados, representado pelo H5WL, que pode interagir e integrar-se com outros tipos de estruturas distribuídas, como por exemplo, malhas computacionais. Desta forma, facilitando o desenvolvimento de novos métodos para acesso a imagens médicas DICOM no formato de dados HDF5.

Estudos mais profundos em relação a paralelização das consultas realizadas no arquivo HDF5 sobre o sistema de arquivo distribuído, sugerem uma melhora significativa no desempenho do sistema, visto que, da mesma forma, a aplicação de métodos de escrita paralela no ambiente distribuído também pode melhorar significativamente o desempenho do armazenamento.

Adicionalmente, um outro estudo possível para trabalhos futuros, seria a utilização de armazenamento das imagens em formato HDF5 em malhas computacionais, utilizando de software para armazenamento distribuídos, tal como o GridFTP. Para a recuperação das imagens, seria possível utilizar *webservices* para intermediar as operações entre as entidades do sistema.

## Referências

- [1] The cyclops project. Disponível em: <http://www.cyclops.ufsc.br>. Acessado em: 17/06/2008.
- [2] Hdf5 – hierarchical data format 5. user's guide. Disponível em: <http://hdf.ncsa.uiuc.edu/products/hdf5/>. Acessado em: 16/06/2008.
- [3] Hdf5 wrapper library. Disponível em: <http://www.cyclops.ufsc.br/h5wl/>. Acessado em: 17/06/2008.
- [4] Laboratório de telemedicina. Disponível em: <http://www.telemedicina.ufsc.br>. Acessado em: 18/06/2008.
- [5] Pvfs – parallel virtual file system. Disponível em: <http://www.pvfs.org>. Acessado em: 18/06/2008.
- [6] R. L. Bashshur. Telemedicine and health care. *Telemedicine Journal and E-Health*, 8, 2002.
- [7] S. Cohen, P. Hurley, K. W. Schulz, W. L. Barth, and B. Benton. Scientific formats for object-relational database systems: a study of suitability and performance. *SIGMOD Record*, 35(2):10–15, 2006.
- [8] P. R. Dellani. Desenvolvimento de um servidor de imagens médicas digitais no padrão dicom. Master's thesis, Universidade Federal de Santa Catarina, 2005.
- [9] L. J. Gosink, J. Shalf, K. Stockinger, K. Wu, and W. Bethel. HDF5-fastquery: Accelerating complex queries on HDF datasets using fast bitmap indices. In *18th International*

- Conference on Scientific and Statistical Database Management (SSDBM'06)*, pages 149–158. IEEE Computer Society, 2006.
- [10] K.-P. Lee and P. Spence. view-hdf: visualization and analysis tool for hierarchical data format files. *Oceans '02 MTS/IEEE*, 2:744–750 vol.2, Oct. 2002.
- [11] D. D. J. Macedo, H. W. G. Perantunes, R. Andrade, A. von Wangenheim, and M. A. R. Dantas. Asynchronous data replication: A national integration strategy for databases on telemedicine network. In *21th IEEE International Symposium on Computer-Based Medical Systems, 2008. CBMS 2008.*, pages 495–500, 2008.
- [12] D. D. J. Macedo, H. W. G. Perantunes, E. Comunello, A. von Wangenheim, and M. A. R. Dantas. An interoperability approach based on asynchronous replication among distributed internet databases. In *IEEE Symposium on Computers and Communications – ISCC, 2008*, 2008.
- [13] D. D. J. Macedo, H. W. G. Perantunes, L. F. J. Maia, A. v. Wangenheim, and M. A. R. Dantas. Replicação assíncrona entre bancos de dados médicos distribuídos. In *Escola Regional de Bancos de Dados*, 2008.
- [14] R. Maia, A. von Wangenheim, and L. Nobre. A statewide telemedicine network for public health in brazil. In *19th IEEE International Symposium on Computer-Based Medical Systems, 2006. CBMS 2006.*, pages 495–500, 2006.
- [15] K. M. McNeill, R. S. Weinstein, and M. J. Holcomb. Arizona telemedicine program: Implementing a statewide health care network. *Journal of American Medical Informatics Association*, Jan. 18 2002.
- [16] L. A. Ribeiro, P. R. Dellani, A. von Wangenheim, M. M. Richter, K. Maximini, and E. Comunello. Cyclopsdistmeddb. - a transparent gateway for distributed medical data access in dicom format. In *CBMS '02: Proceedings of the 15th IEEE Symposium on Computer-Based Medical Systems (CBMS'02)*, page 315, Washington, DC, USA, 2002. IEEE Computer Society.
- [17] R. Ross, D. Nurmi, A. Cheng, and M. Zingale. A case study in application I/O on linux clusters. In *ACM/IEEE Supercomputing Conference (SC'2001)*, Denver, Nov. 2001. ACM SIGARCH/IEEE.
- [18] S. G. Shasharina, N. Wang, and J. R. Cary. Grid service for visualization and analysis of remote fusion data. In *Proceedings of the Second International Workshop on Challenges of Large Applications in Distributed Environments (CLADE'04)*, page 34. IEEE Computer Society, 2004.
- [19] J. Wallauer, D. D. J. Macedo, R. Andrade, and A. von Wangenheim. Building a national telemedicine network. *IT Professional*, 10(2):12–17, March-April 2008.
- [20] H. Yu, R. Sahoo, C. Howson, G. Almasi, J. Castanos, M. Gupta, J. Moreira, J. Parker, T. Engelsiepen, R. Ross, R. Thakur, R. Latham, and W. Gropp. High performance file i/o for the blue gene/l supercomputer. In *The Twelfth International Symposium on High-Performance Computer Architecture*, July 2006.