私立東海大學

資訊工程學系研究所

碩士論文

指導教授：楊朝棟　博士

於雲端計算平台上 醫療影像檔案存取系統之實作

Implementation of a Medical Image File Accessing System on Cloud Computing

研　究　生：陳龍騰

中華民國一〇〇年一月

摘要

科技的進步已使醫院的 X 光片被數位影像儲存通訊系統（PACS）取代，然而在使用及管理 PACS 時我們遭遇到三大挑戰。第一、不同醫院之間的醫療影像交換及分享速度太慢，大量的醫療影像透過廣域網路（WAN）傳輸，在上傳和下載的速度都受限於網路頻寬。第二、為了安全穩定的儲存醫療影像，醫院在儲存設備的持續投資上過於昂貴，在管理上更是不方便。第三，醫療照護的過程中，醫療影像檔案系統提供的服務穩定性及可靠度是非常必然的。這篇論文的研究動機是為了試圖解決醫學影像在不同醫院之間的交換、共享、儲存、效能及可靠度等問題。

本研究中，我們基於雲端計算平台（Hadoop）設計了一個醫療影像檔案存取系統，稱為 MIFAS（Medical Image File Accessing System），透過這個系統可以查詢及取得醫療影像。系統容錯（System Fault Tolerance）架構設計，使前端應用的可靠度大為增加。HDFS（Hadoop Distributed File System）分散式檔案系統及複本配置機制的導入，使檔案儲存的穩定性得到保障，且投資及管理成本得以下降。並因為加入了下載策略的最佳化(Co-allocation Model)中介軟體，使得醫療影像在廣域網路(WAN)的傳輸速度得到提昇。此外我們還提供易於操作及管理的 WEB 化系統介面。相較於傳統的 PACS，MIFAS 更能提供良好的醫療影像檔案存取服務品質，並使不同醫院之間的醫學影像應用問題得到改善。

關鍵字：數位影像儲存通訊系統、雲端計算平台、系統容錯、醫療影像存取檔案系統、複本配置、分散式檔案系統

Abstract

Advances in technology, the X-ray films have been replace by the Picture Archiving and Communication System (PACS) in hospital. But we encountered three major challenges at the using and management of PACS. First, medical image exchange and sharing is too slow between the different hospitals, a large number of images through the wide area network (WAN) transmission, in both upload and download speed is limited by network bandwidth. Second, in order to store medical images safe and stable, the hospital's continued investment in the storage device is too expensive, the management is inconvenient. Third, the process of medical care, medical imaging services provided by the file system stability and reliability is very necessary. Therefore, the motivation of this paper is to attempt to resolve medical imaging in the exchange between the different hospitals, sharing, storage, performance and reliability issues.

In this research, we designed a Medical Image File Accessing System (MIFAS) based on Cloud-based computing platform (HADOOP). We can be queried by the system and access to medical images. The design in system fault tolerance architecture that greatly increased the reliability of front-end applications. Hadoop distributed file system (HDFS) and replication service, so that the stability of stored files are protected, and the investment and management costs can be reduced. Download the strategy and the addition of the Co-allocation model middleware that making the medical imaging in the wide area network (WAN) transmission speed has been improved. In addition, we also provide easy operation and management of WEB based interface. Compared with traditional PACS, our system provided the best quality medical image file access services to improve the medical imaging issue between the different hospitals.

Keywords: PACS,Cloud Computing,Hadoop,System Fault Tolerance,HDFS, MIFAS, Replication Service

Acknowledgements

回想當初考進入東海資工時的種種情事，就像是才剛發生的一樣，入學後煩腦著能跟隨那一位教授進行學習，在偶然的機會下，學長陳秋雄給了我一條明確的道路，並在他的推薦下得以加入高效能計算實驗室，參與一連串的研習。

學業完成在即，首先我要感謝指導教授－楊朝棟 博士，在學期間給予我最多耐心且最大支持與最真誠的鼓勵，無論是參與國際研討會的論文發表或者是在課堂之中及研究的過程或學習的過程中都充滿挑戰且富有成就感，何其有幸能成為楊老師實驗室的一員，在東海高效能計算實驗室培養之下，造就出的韌性及知識，以應付未來更多的挑戰。

論文口試期間，承蒙許慶賢教授、賴冠州教授、張玉山教授、時文中教授在百忙之中給予詳細的審閱與指導，並提供許多寶貴的意見，使論文的內容更加詳實與嚴謹，在此致上最誠摯的謝意。

接著我也要感謝我的家人，沒有家人的支持與體諒，我將無法順利完成論文，更遑論畢業取得碩士學位。

最後，我要感謝研究所合作的學長、同學們秋雄、裕翔、智霖、威利、冠傑等及提供支持與協助的公司資訊室主管孫培然主任及程式開發同仁政宏、少宏等，還要特別感謝系統組兄弟們這陣子的擔待，沒有你們無私的付出，我將不可能完成我的研究工作，謝謝你們大家，並且再一次深深感謝楊老師的指導。

Contents

List of Tables

List of Figures

Chapter 1

Introduction

EMR is an acronym from Electronic Medical Records. This refers to a paperless, digital and computerized system of maintaining patient data. It is designed to increase the efficiency and reduce documentation errors by streamlining the process. Implementing EMR is a complex, expensive investment that has created a demand for Healthcare IT professionals and accounts for a growing segment of the healthcare workforce [6].

An electronic medical record (EMR) is a computerized medical record created in an organization that delivers care, such as a hospital and doctors' surgery. Electronic medical records tend to be a part of a local stand-alone health information system that allows storage, retrieval and modification of records. According to the Medical Records Institute, the Electronic Medical Record can be described as five stages which are Automated Medical Record, Computerized Medical Record Provider-based, Electronic Medical Record, Electronic Patient Record and Electronic Health Record [6]. This work tries to solve the issues in Electronic Medical Records which are exchanging, storing and sharing in Medical Images.

Based on the "Medical Images Exchange" issue of EMR, we presented a system called MIFAS (Medical Image File Accessing System) in this work; it was built on the Hadoop [1] platform to solve the exchanging, storing, sharing issues in Medical Images. In this work, we also presented a new strategy for processing medical image inspecting, which was co-allocation mechanism for cloud environment. We utilized the Hadoop platform and Co-allocation mechanism to establish the Cloud environment for MIFAS. MIFAS could easily help users to retrieve, share and store Medical Images between different hospitals. The remainder of this work was organized as following. Background review and studies were presented in Chapter 2. The System Architecture was introduced in Chapter 3. Experimental results were presented in Chapter 4. Finally, Chapter 5 concluded this article.

Chapter 2

Background

2.1 Challenge in Medical Image Exchanging

For over a decade, the majority of all hospital and private radiology practices have transformed from film-based image management systems to a fully digital (filmless and paperless) environment but subtly dissimilar (in concept only) to convert from a paper medical chart to an EHR. Film and film libraries have given ways to modern picture archiving and communication systems (PACS). And they offer highly redundant archives that tightly integrate with historical patient metadata derived from the radiology information system. These systems may be not only more efficient than film and paper but also more secure as they incorporate with safeguards to limit access and sophisticate auditing systems to track the scanned data. However, although radiologists are in favor of efficient access to the comprehensive imaging records of our patients within our facilities, we ostensibly have no reliable methods to discover or obtain access to similar records which might be stored elsewhere [24, 25, 27].

According to our research, there were few Medical Image implementations on cloud environment. However, a familiar research presented the benefits of Medical Images on cloud were: Scalability, Cost effective and Replication [22]. In the same study, they also presented a HPACS system but lacked of management interface.

2.2 Hadoop and HDFS

Hadoop is one of the most salient pieces of the data mining renaissance which offers the ability to tackle large data sets in ways that weren't previously possible due to time and cost constraints. It is a part of the apache software foundation and its being built by the community of contributor in all over the world. The Hadoop project promotes the development of open source software and supplies a framework for the development of highly scalable distributed computing applications [23].

Hadoop is the top-leveled project in Apache Software Foundation and it supports the development of open source software [1]. Hadoop provides a framework for developing highly scalable distributed applications. The developer just focuses on applying logic instead of processing detail of data sets. The HDFS (Hadoop Distributed File System) file system stores large files across multiple machines, shown in Figure 2.1. It achieves reliability by replicating the data across multiple hosts, and hence does not require RAID storage on hosts. The HDFS file system is built from a cluster of data nodes, each of which serves up blocks of data over the network using a block protocol. They also serve the data over HTTP, allowing access to all content from a web browser or other client. Data nodes can connect to each other to rebalance data, to move copies around, and to keep the replication of data high. A file system requires one unique server, the name node. This is a single point of failure for an HDFS installation. If the name node goes down, the file system will be off-lined. When it comes back up, the name node must replay all outstanding operations. This replay process can take over half an hour for a big cluster [26].

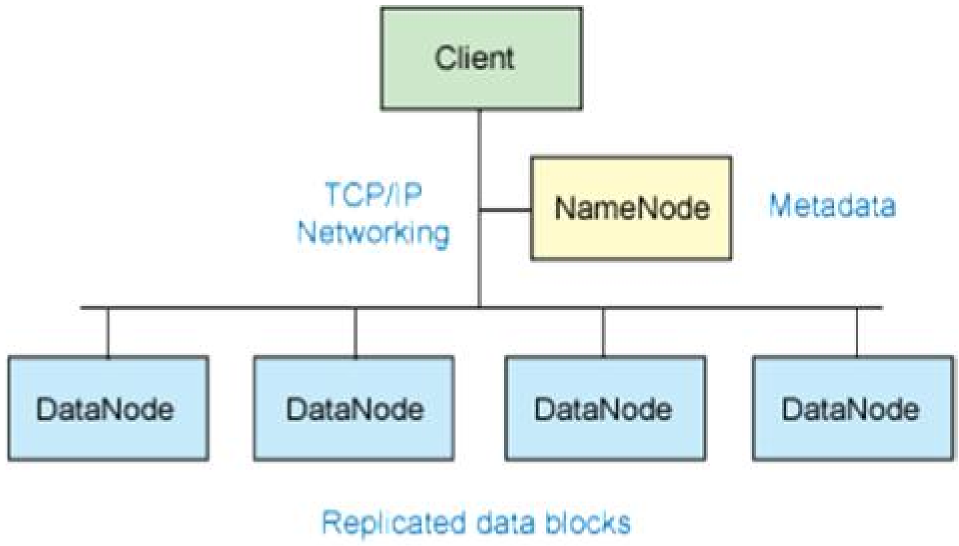

Figure 2.1 Hadoop HDFS Architecture

2.3 PACS

PACS is an acronym that stands for Picture Archiving and Communication System, shown in Figure 2.2. PACS revolutionized the field of radiology, which now consists of all digital, computer-generated images as opposed to the analog film of yesteryear. Analog film took up space and time for filing and retrieval and storage, and was prone to being lost or misfiled. PACS therefore saves precious time and money, and reduces the liability caused by filing errors and lost films. A PACS consists of four major components: the imaging modalities such as CT and MRI, a secured network for the transmission of patient information, workstations for interpreting and reviewing images, and archives for the storage and retrieval of images and reports. Combined with available and emerging Web technology, PACS has the ability to deliver timely and efficient access to images, interpretations, and related data. PACS breaks down the physical and time barriers associated with traditional film-based image retrieval, distribution, and display. PACS is primarily responsible for the inception of virtual radiology, as images can now be viewed from across town, or even from around the world. Additionally, PACS acts as a digital filing system to store patients' images in an organized way which enables records to be retrieved with ease as needed for future reference.

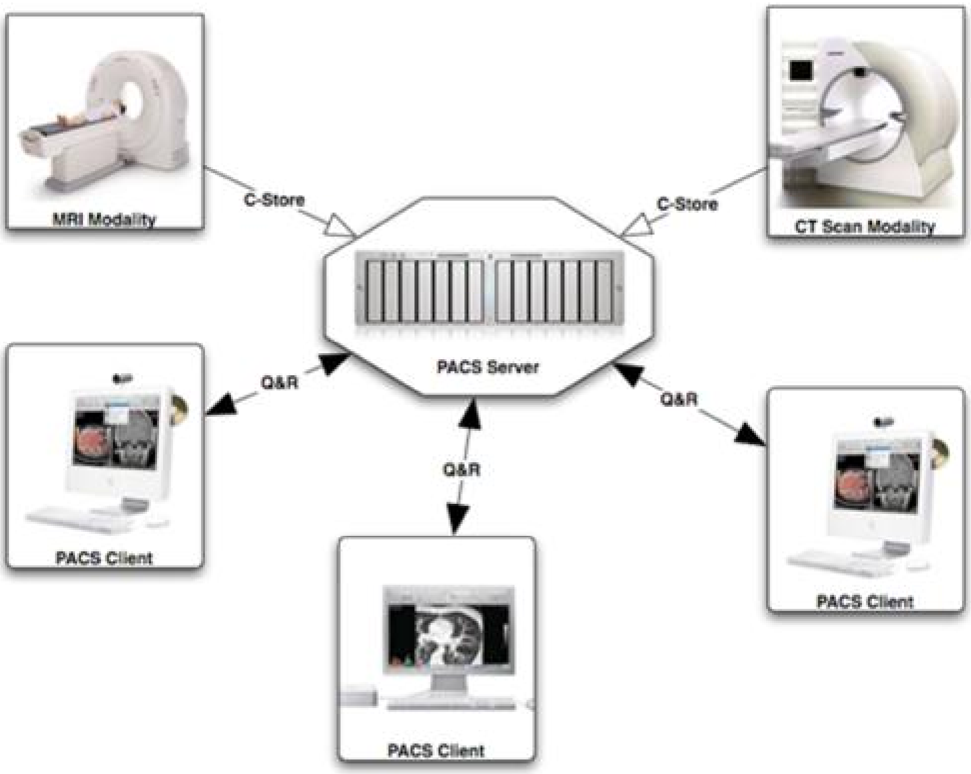

Figure 2.2 PACS Architecture

2.4 DICOM

What is DICOM? Short for Digital Imaging and Communications in Medicine, a standard in the field of medical informatics for exchanging digital information between medical imaging equipment (such as radiological imaging) and other systems, ensuring interoperability. The standard specifies: a set of protocols for devices communicating over a network the syntax and semantics of commands and associated information that can be exchanged using these protocols a set of media storage services and devices claiming conformance to the standard, as well as a file format and a medical directory structure to facilitate access to the images and related information stored on media that share information. The standard was developed jointly by ACR (the American College of Radiology) and NEMA (the National Electrical Manufacturers Association) as an extension to an earlier standard for exchanging medical imaging data that did not include provisions for networking or offline media formats [29].

DICOM enables the integration of scanners, servers, workstations, printers, and network hardware from multiple manufacturers into a picture archiving and communication system (PACS). The different devices come with DICOM conformance statements which clearly state the DICOM classes they support. DICOM has been widely adopted by hospitals and is making inroads in smaller applications like dentists' and doctors' offices.

DICOM differs from some, but not all, data formats in that it groups information into data sets. That means that a file of a chest X-Ray image, for example, actually contains the patient ID within the file, so that the image can never be separated from this information by mistake. This is similar to the way that image formats such as JPEG can also have embedded tags to identify and otherwise describe the image.

A DICOM data object consists of a number of attributes, including items such as name, ID, etc., and also one special attribute containing the image pixel data (i.e. logically, the main object has no "header" as such - merely a list of attributes, including the pixel data). A single DICOM object can only contain one attribute containing pixel data. For many modalities, this corresponds to a single image. But note that the attribute may contain multiple "frames", allowing storage of cine loops or other multi-frame data. Another example is NM data, where an NM image by definition is a multi-dimensional multi-frame image. In these cases three- or four-dimensional data can be encapsulated in a single DICOM object. Pixel data can be compressed using a variety of standards, including JPEG, JPEG Lossless, JPEG 2000, and Run-length encoding (RLE). LZW (zip) compression can be used for the whole data set (not just the pixel data) but this is rarely implemented.

The same basic format is used for all applications, including network and file usage, but when written to a file, usually a true "header" (containing copies of a few key attributes and details of the application which wrote it) is added.

2.5 Co-allocation Mechanism

Co-allocation architecture enables parallel downloading from data node. It can also speed up downloads and overcome network faults. The architecture proposed [14] consists of three main components: an information service, a broker/co-allocator, and local storage systems. Co-allocation of data transfers is an extension of the basic template for resource management [8]. Applications specify the characteristics of desired data and pass attribute descriptions to a broker. The broker searches for available resources, and gets replica locations from the Information Service [7] and Replica Management Service [11]; then, obtains the lists of physical file locations. We have implemented the following eight co-allocation schemes: Brute-Force (Brute), History-based (History), Conservative Load Balancing (Conservative), Aggressive Load Balancing (Aggressive), Dynamic Co-allocation with Duplicate Assignments (DCDA), Recursively-Adjusting Mechanism (RAM), Dynamic Adjustment Strategy (DAS), and Anticipative Recursively-Adjusting Mechanism (ARAM)[16,17,18].

2.6 Ganglia Monitor System

The Ganglia is a distributed monitoring system for high-performance computing systems. It is based on a hierarchical design targeted at federations of clusters. It leverages widely used technologies such as XML for data representation, XDR for compact, portable data transport, and RRDtool for data storage and visualization. It uses carefully engineered data structures and algorithms to achieve very low per-node overheads and high concurrency. In this research we used the ganglia monitor system to help us to monitor storage, memory and CPU usage to measure each node performance. As below picture (Seeing Figure 2.3) is our experimental ganglia monitor web page.

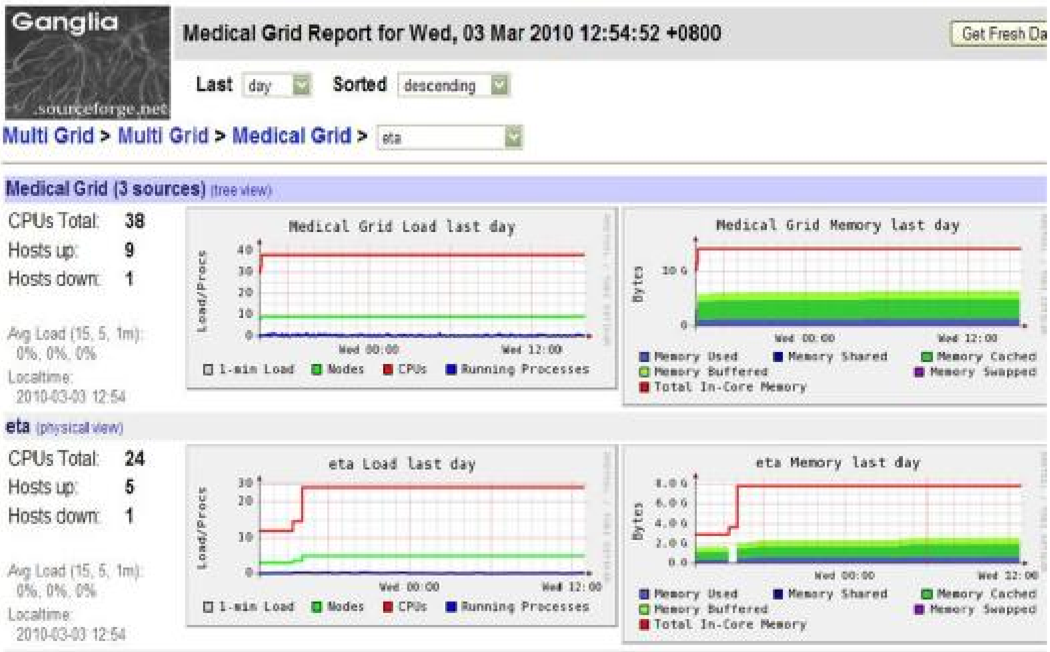

Figure 2.3 Ganglia Monitor System

2.7 Related Works

HDFS servers (i.e., DataNodes) and traditional streaming media servers are both used to support client applications that have access patterns characterized by long sequential reads and writes. As such, both systems are architected to favor high storage bandwidth over low access latency [32].

Recently “Cloud” became a hot word in this field. S. Sagayaraj [22] proposes that Apache Hadoop is a framework for running applications on large clusters built of commodity hardware. The Hadoop framework transparently provides both reliability and data motion. Hadoop implements a computational paradigm named Map/Reduce, where the application is divided into many small fragments of work. Each fragment of work may be executed or re-executed on any node in the cluster. So, by just replacing the PACS Server with Hadoop Framework can lead to good, scalable and cost effective tool for the Imaging solution for Health Care System. In the same study, they also presented a HPACS system but lacked of management interface.

It is complex for kind of performance issues, but J. Shafer, et al. [30] proposed "The Hadoop distributed filesystem: Balancing portability and performance" had a good view in this field. The poor performance of HDFS can be attributed to challenges in maintaining portability, including disk scheduling under concurrent workloads, file system allocation, and file system page cache overhead. HDFS performance under concurrent workloads can be significantly improved through the use of application-level I/O scheduling while preserving portability. Further improvements by reducing fragmentation and cache overhead are also possible, at the expense of reducing portability. However, maintaining Hadoop portability whenever possible will simplify development and benefit users by reducing installation complexity, thus encouraging the spread of this parallel computing paradigm.

In our previous reproaches [12, 14, 15, 17, 18, 20] use co-allocation to solve the data transfer problem in grid environment. It is the foundation of this thesis. But it is for grid not implement on cloud, so in this work we have a significant change from previous works, because we implement it on the cloud environment.

Chapter 3

System Design and Implementation

In Figure 3.1, we described the current overview of Distribution File System. The MIFAS has three HDFS groups. The first group, THU1 and the second group, THU2 are both in Tunghai University. And the third group is CSMU in Chung Shan Medical University Hospital. All of the groups are under 100Mbps network bandwidth in TANET (Taiwan Academic Network) network environment. The HDFS group number can be very flexible. The minimum is one but the maximum can be many. The more HDFS group we have the more duplication source we get. It means that the PCAS images source is from the HDFS. Thus, if we increase the source number (build more HDFS group), the effects will definitely be different based on source numbers.

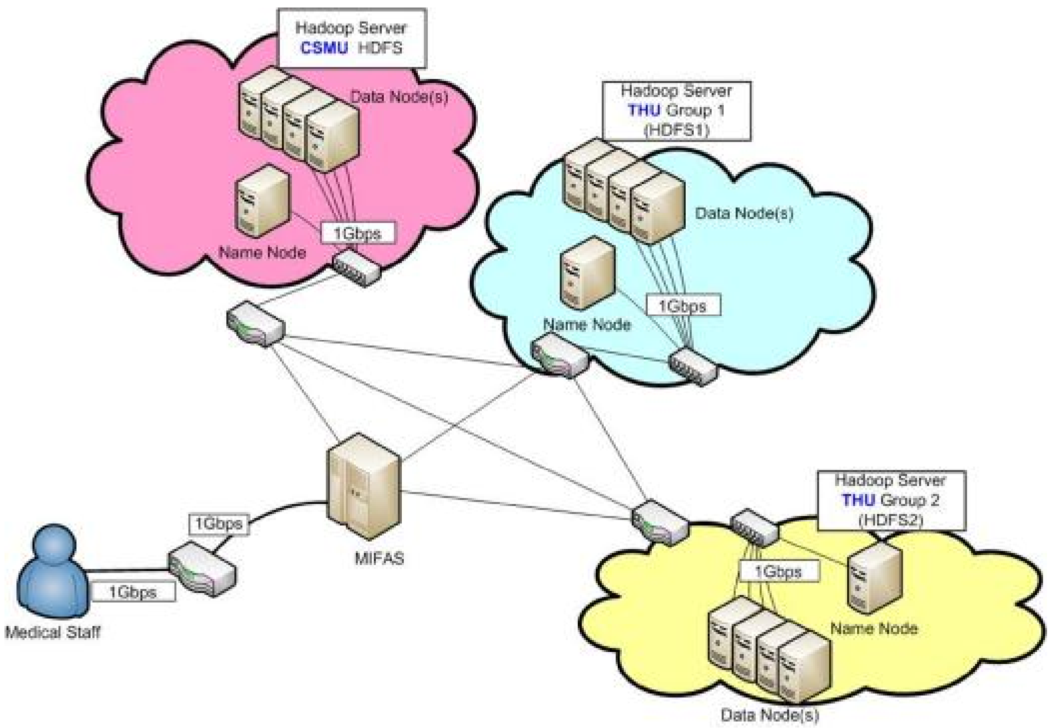

Figure 3.1 Overview of Distribution File System.

3.1 System Architecture

In Figure 3.2, MIFAS was developed on cloud environment. The distribution file system was built on HDFS of Hadoop environment (Chapter 2.2). This Hadoop platform could be described as PaaS (Platform as a Service). We extended a SaaS (Software as Service) based on PaaS. As the shown illustration, the top level of MIFAS was web-based interface. MIFAS provided a nice and friendly interface that users could easily queries the Medical Images.

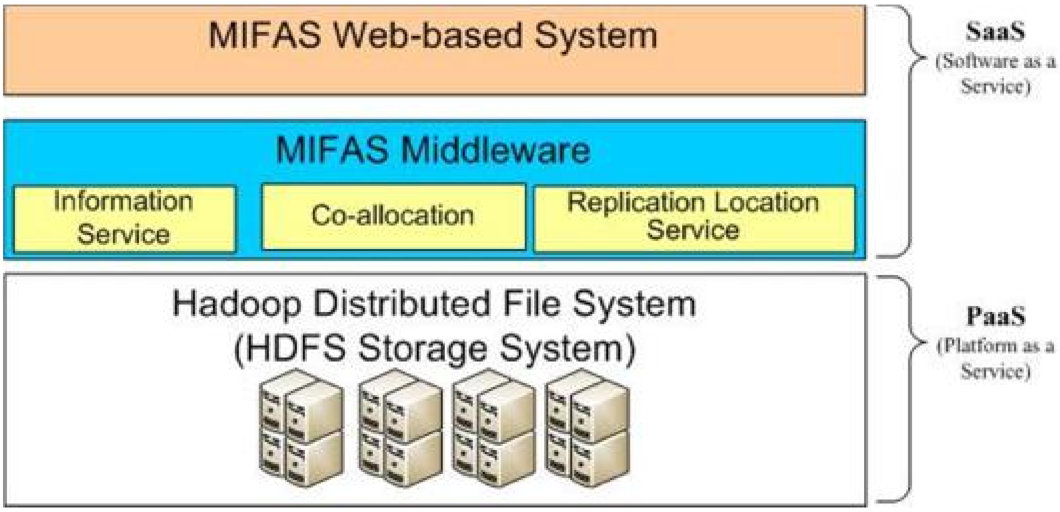

Figure 3.2 System Architecture of MIFAS

Middleware: There was a mechanism to handle the transmission issue in MIFAS. In this research, we called the mechanism as MIFAS Middleware. The Middleware's purpose is to assign/acquire the best transmission path to distribution file system. This Middleware also collected necessary information such as bandwidth between server and server, the server utilization rate, and network efficiency. The information provided entirety MIFAS Co-allocation Distribution Files System to determine the best solution of downloading allocation jobs (Figure 3.3).

Information Service: To obtained analysis in the status of host. The Middleware of MIFAS had a mechanism to fetch the information of hosts called Information Service. In this research, we installed the Ganglia [28] in each member of Hadoop node to get the real-time state from all members. Therefore, we could get the best strategy of transmission data from Information Service which is one of the components of MIFAS Middleware.

Co-allocation: As our researched into section 2.3, Co-allocation mechanism could conquest the parallel downloading from data nodes. Besides, it also sped up downloading and solved network faults problems.

Replication Location Service: In this research, we built three groups of HDFS in different locations, and each HDFS owned an amount of data nodes. The Replication Location Service means that the Service would automatically make duplication from private cloud to one another when medical images uploaded to MIFAS.

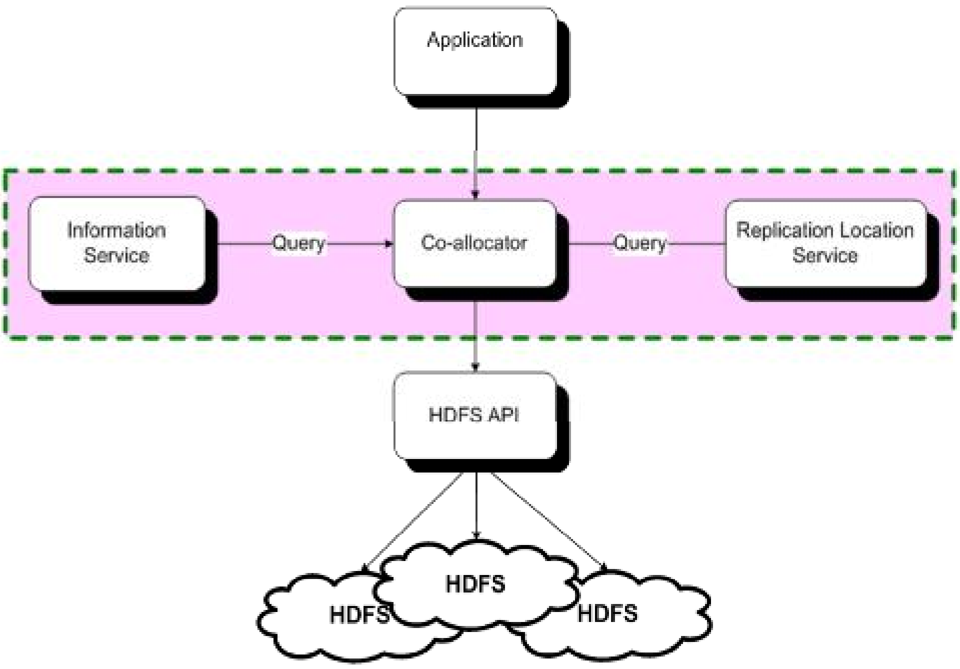

Figure 3.3 Middleware Mechanism

3.2 System Workflow

In Figure 3.4, it shows our efforts on MIFAS. In this research, we also made a real system to achieve our thesis. The system's workflow shows in the shown illustration. Firstly, users input username and password to authenticate. Secondly, users could input search condition to query patients' information. Thirdly, users could also view patients' Medical Images. Fourthly, users can configure in MIFAS. Fifthly, if users can present MIFAS downing mechanism, it means the Middleware is workable in MIFAS.

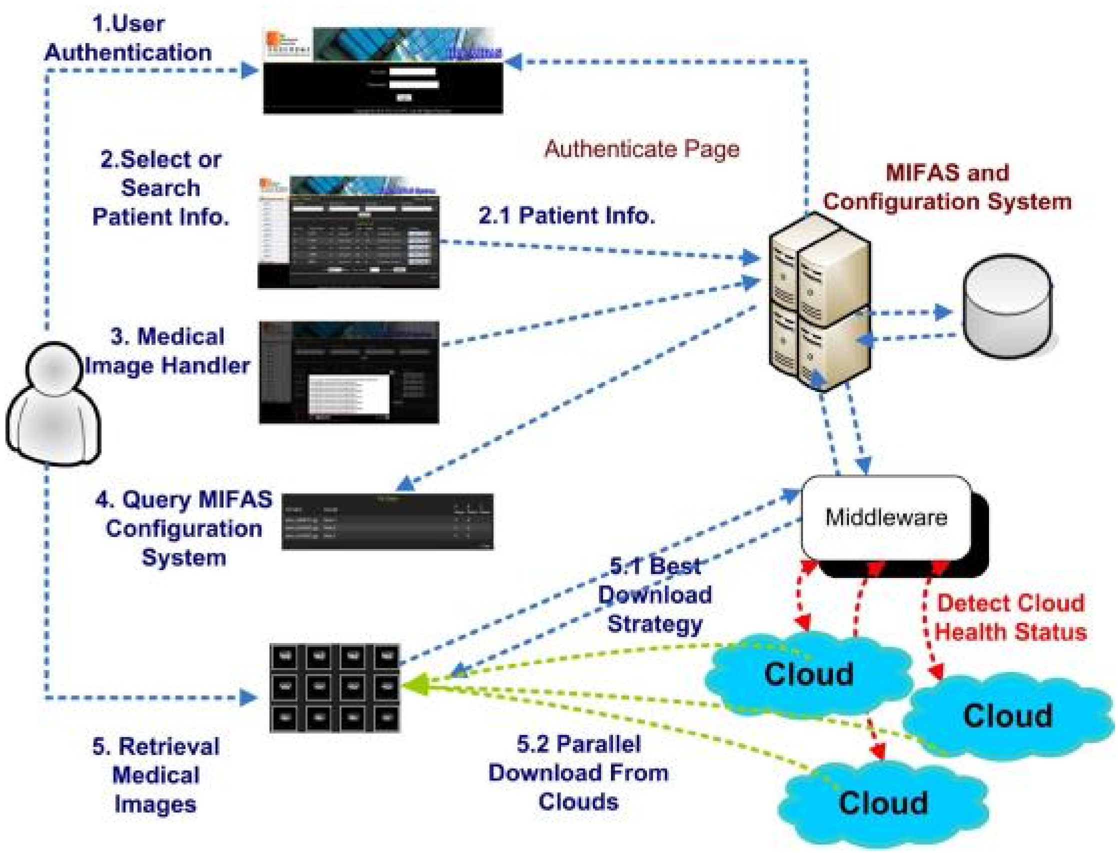

Figure 3.4 System Workflow of MIFAS

3.3 Sysem Inerface

MIFAS offer authentication interface as shown Figure 3.5, the users have to pass the validation before login to the MIFAS. After passing the validation, you will see the Portal of MIFAS (In Figure 3.6). In the Figure 3.6, there are 3 main functional block, each one has its functional purpose.

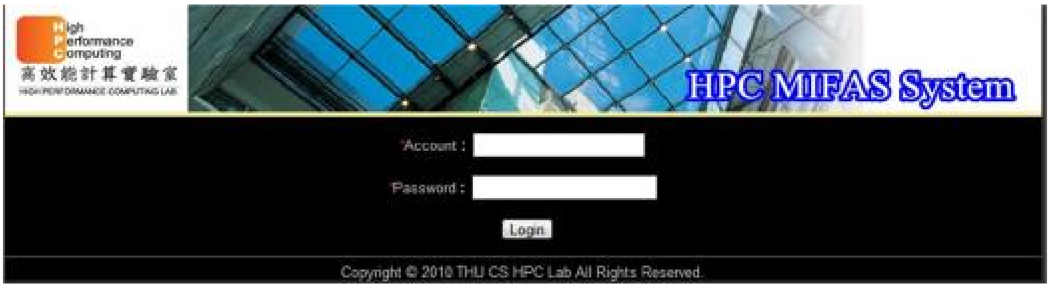

Figure 3.5 Authorization Interface

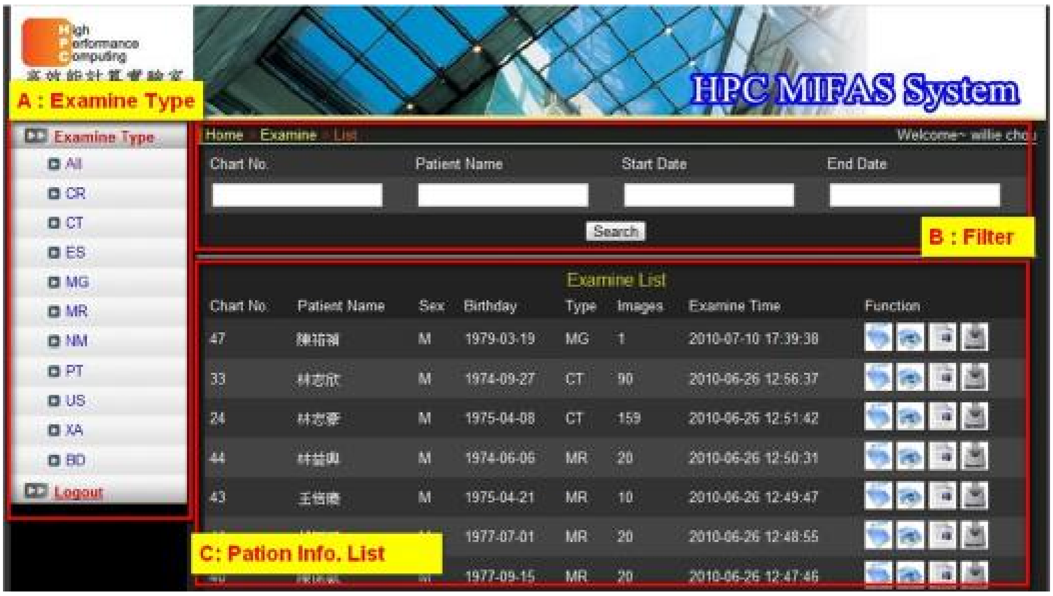

Figure 3.6 Portal of MIFAS System

Examine Type: it is the catalog of Medical Examination Type. Medical Images could from various medical imaging instruments, including ultrasound (US), magnetic resonance (MR), positron emission tomography (PET), computed tomography (CT), endoscopy (ENDO), mammograms (MG), direct radiography (DR), computed radiography (CR) etc. The examine type is cataloged depend on the definition of DICOM.

Filter: this block provided search function; users can get patient information through inputted keyword. In web-based interface system, it is easy to reach this goal. User can capture any information that he/she wanted by filter. Thus, like other system on the internet, MIFAS provide multidimensional information for the users. There are four main options in the filter function which are “Chart NO” (Examination No), “Patient Name”, “Start of Examination Date”, and “End of Examination Date”.

Patient Information List: in this block, you will see the detail information according to the search condition of block B and block A. And it is also including several important functions.

Function items as shown Figure 3.7, which are the 1st Image Status, 2nd Thumbnail viewer, 3rd PACS Reporting, 4th Download File. Display the information of file distribution status as Figure 3.8 , including photographic description and photographic catelog. Thumbnail viewer function is in Figure 3.9, this function show the thumbnail of Medical Image and examination report. PACS Reporting is as shown Figure 3.10, it is the detail report of patient medical record. For more detail of Medical Images, we can through Download File function then utilize other professional DICOM viewer. Regarding to how to upload Medical Images to MIFAS please see Figure 3.11. According to our thesis, the Replication Location Service will duplicate images to each cloud.

Finally the 4th icon in Figure 3.7 could download DICOM format Medical Images from MIFAS. MIFAS will enable co-allocation mechanism to allocate file through a best strategy.

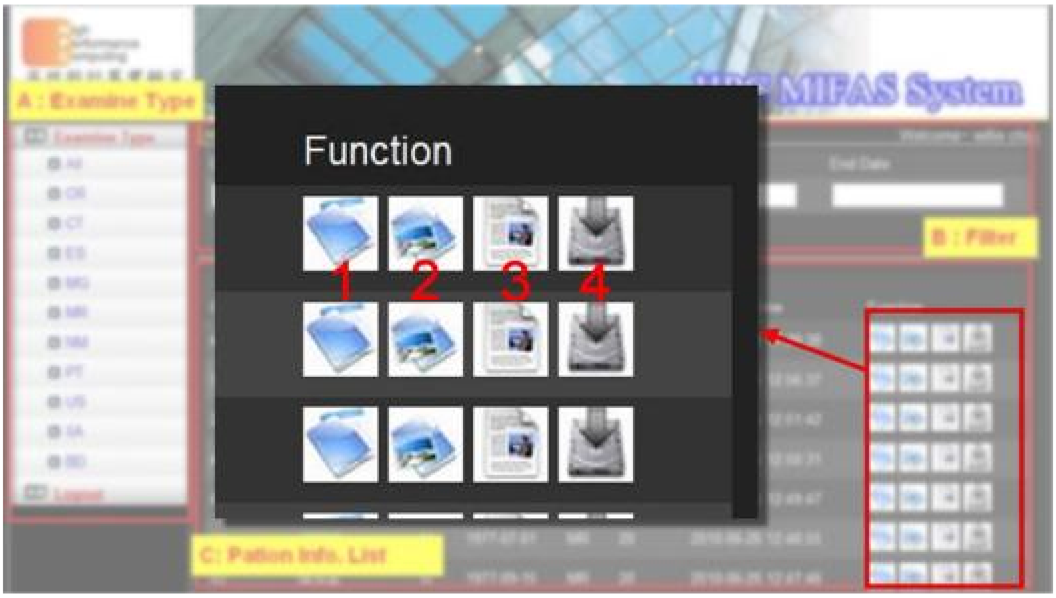

Figure 3.7 Functions

File Status

File Name	Descript	A Server	B Server	C Server
photo_9db96151.jpg	Brain-1	V	V	
photo_8c0c6a52.jpg	Brain-2	V	V	
photo_8c0c6a53.jpg	Brain-3	V	V	

» Top

Figure 3.8 File Status

Figure 3.9 Medical Image Preview

\

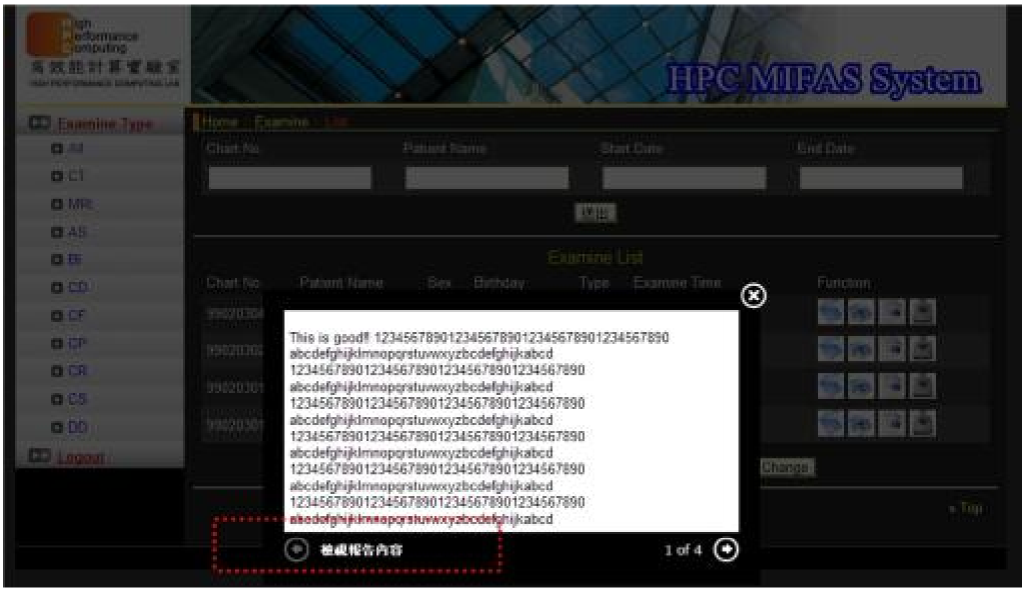

Figure 3.10 Patient Records

Figure 3.11 Upload Medical Images

Chapter 4

Experimental Environments and Results

4.1 Experimental Environments

In this section, we compared MIFAS with PACS. The PACS system in Chung Shan Medical University Hospital is shown in Figure 4.1 It is a production PACS in CSMU and there are total three PACS systems in CSMU, CSUM Chung Kang Branch Hospital and CSUM Tai-Yuan Branch Hospital. Each PACS system has a synchronization mechanism. The network bandwidth between each PACS is under 100Mbps.

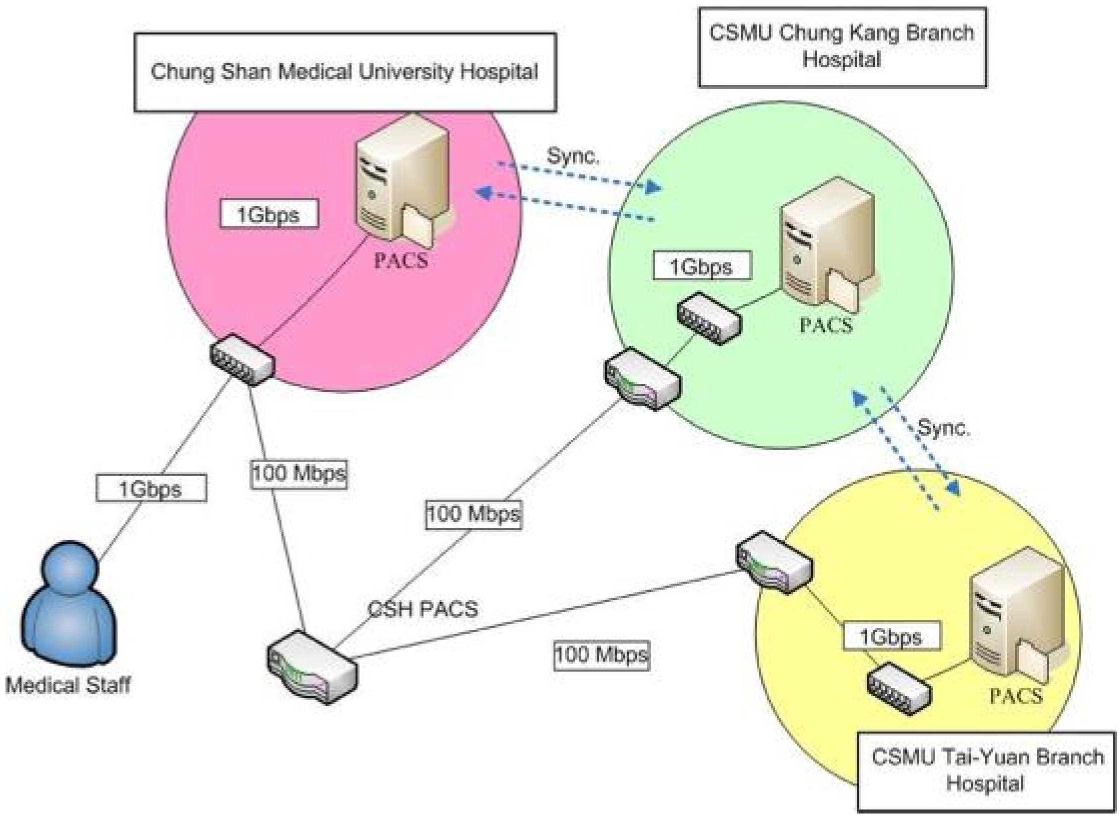

Figure 4.1 System workflow of MIFAS

Table 4.1 Bandwidth of PACS in CSMU

End-to-End Transmission Rates (Mbps) of PACS in CSMU		
Node From	***Node To***	***Bandwidth***
CSMU	Chung Kang Branch	100 Mbps
CSMU	Tai-Yuan Branch	100 Mbps
Chung Kang Branch	Tai-Yuan Branch	100 Mbps

4.2 Experiment 1: Compare Images Retrieval Times from PCSA and MIFAS

In this experiment, we gave same Medical Images as Table 4.2 in PCAS and MIFAS. In a general situation we can describe the PACS vs. MIFAS in Figure 4.2. The illustration shows the experiment 1 environment. All the nodes are under a good situation, then we download the same files from each PACS and MIFAS. The purpose of this experiment is to compare the images Retrieval Times from PCAS and MIFAS.

Table 4.2 Experiment Medical Images

Image Type	**Medical Image Attribute**			
	Pixel	***QTY.***	***Total Size***	***Testing Times***
CR Chest	2804*2931	1	7.1MB	500 times
A series of CT	512*512	389	65MB	500 times

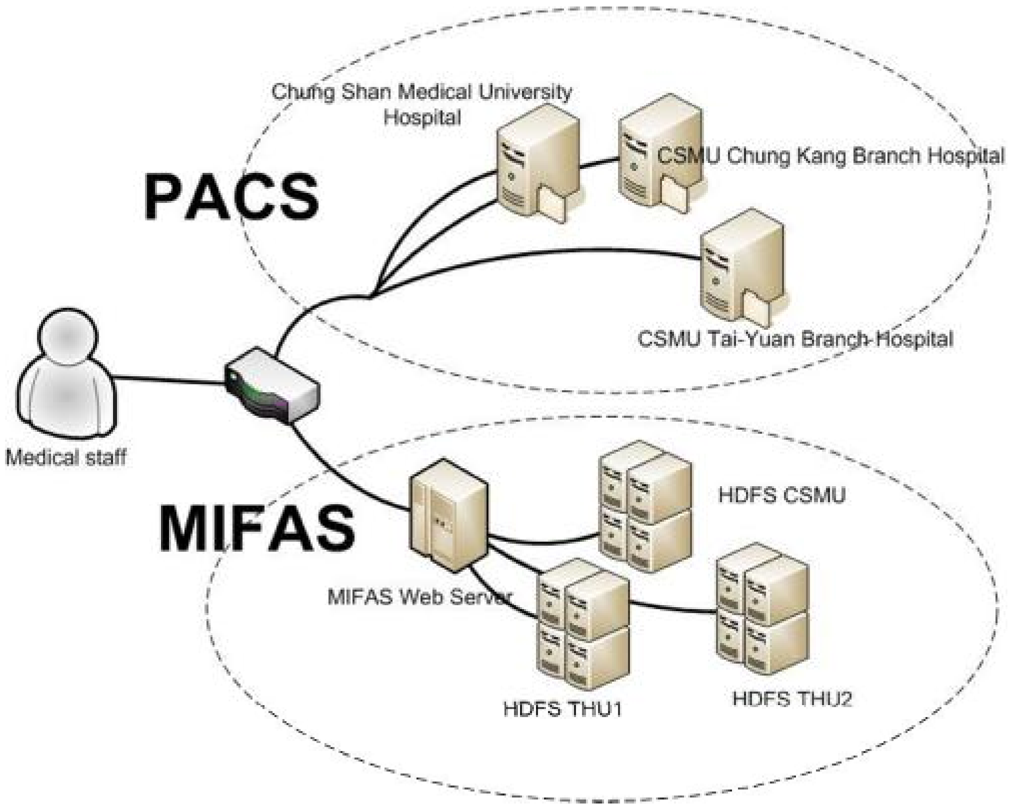

Figure 4.2 System workflow of MIFAS

We tested 500 times for each site, and get the average time result in Figure 4.3. Results in proximal site can be discovered to be more time-consuming in Figure 4.4. The result of Figure 4.3 and Figure 4.4 means the retrieve efficiency of PACS is better than MIFAS. The main reason of the result is that PACS was built on a High-Performance Server and is also a high cost Medical Image System. MIFAS is not easy to cross this threshold. But in experiment 2, the advantages of MIFAS can be displayed in the following experiment.

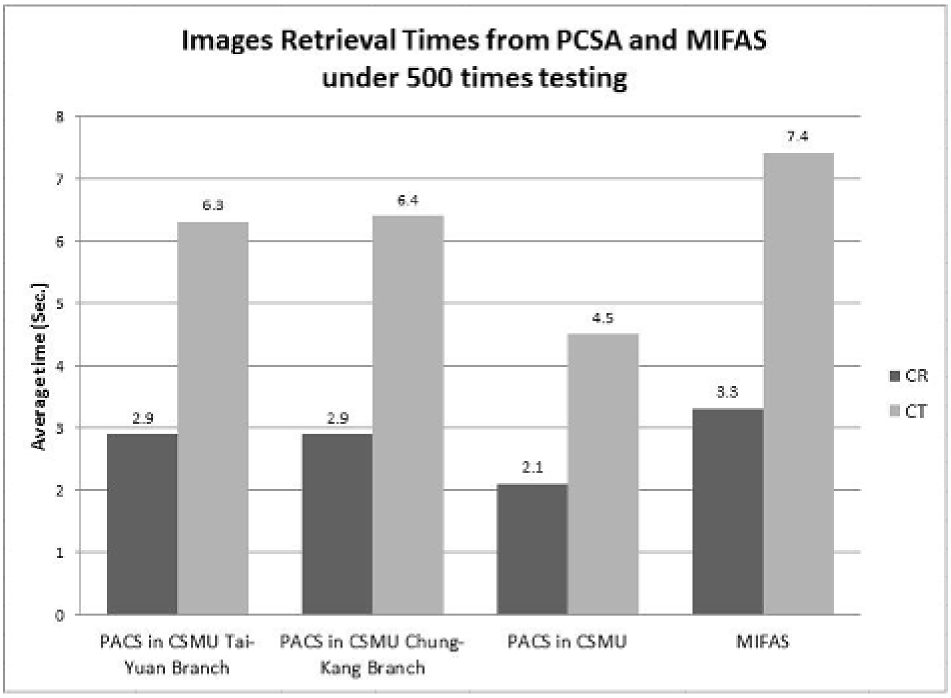

Figure 4.3 Image Retrieval Times Result

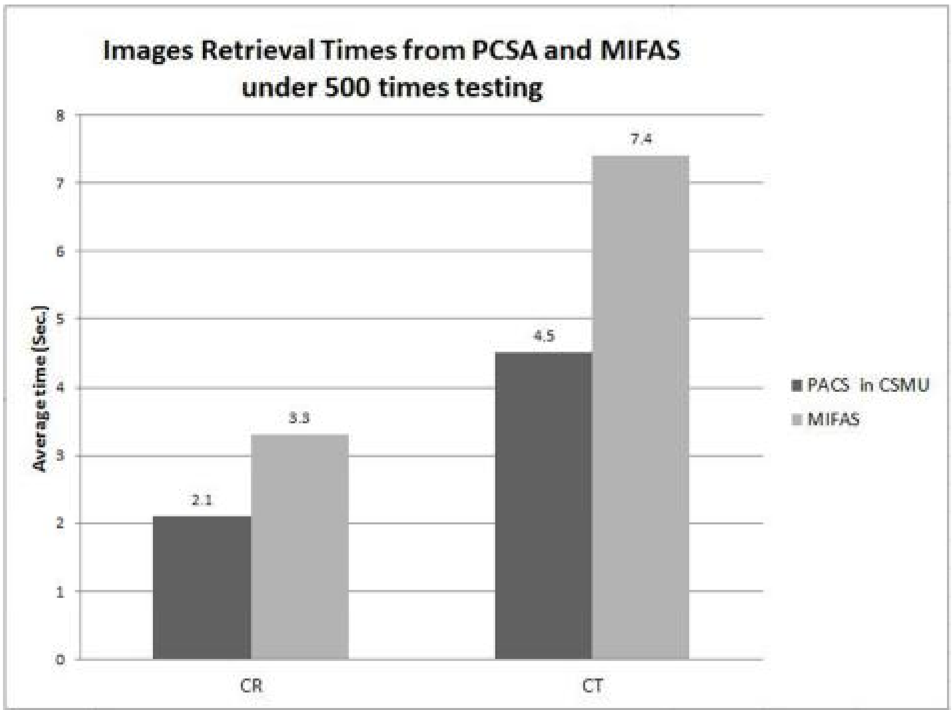

Figure 4.4 Image Retrieval Times Result in Proximal

4.3 Experiment 2: Compare PACS and MIFAS with Proximal Failure Problem

In this experiment, we want to compare PACS and MIFAS with proximal failure problem. As we addressed on MIFAS, the MIFAS provides a Co-allocation strategy, so basically the Single Site Failure issue will not affect the MIFAS operation. On the other hand, if the PACS system occurs to the same problem, the only one solution is to use another PACS site. In the experiment 2, we assume that PACS in CSMU occurs to system failure, and the Medical Staff must retrieval the Medical Images to another CSMU branch hospital. If the same situation, a HDFS group failure problem, occurs in MIFAS. Even the MIFAS has only two left HDFS groups, users still can retrieval the Medical Images from the others HDFS group. In this experiment, we assume both failures are in proximal site. This experiment result can be described in Figure 4.5, PACS in CSMU is not available, but the MIFAS can still work under a good situation.

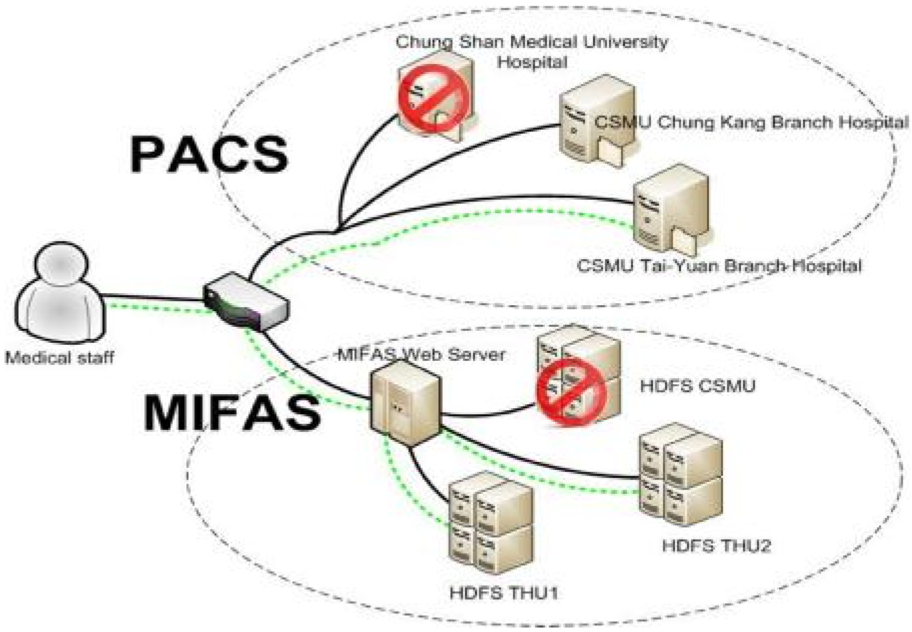

Figure 4.5 System configuration for Experiment 2

4.4 Hardware Broken in Both Environments

In Figure 4.6, we describe a hardware broken state in both systems. Obviously, only one PACS is left in CSUM Tai-Yuan branch but a strong contrast could be seen in MIFAS. And MIFAS still supports the medical images to transfer to users under a Co-allocation mechanism. In the other word, the benefit of MIFAS is that it conquests the network / hardware faults.

PACS has its own synchronization mechanism. However, if there is a hardware or network broken in PACS, the only way is to use the survival site and try to reduce the resume time. The PACS system also has its limitation on concurrent user numbers. Compare to the MIFAS, it can distribute the workload from users accessing under the Co-allocation mechanism. This section shows MIFAS can effectively reduce the single site fault, problems of broken network and hardware.

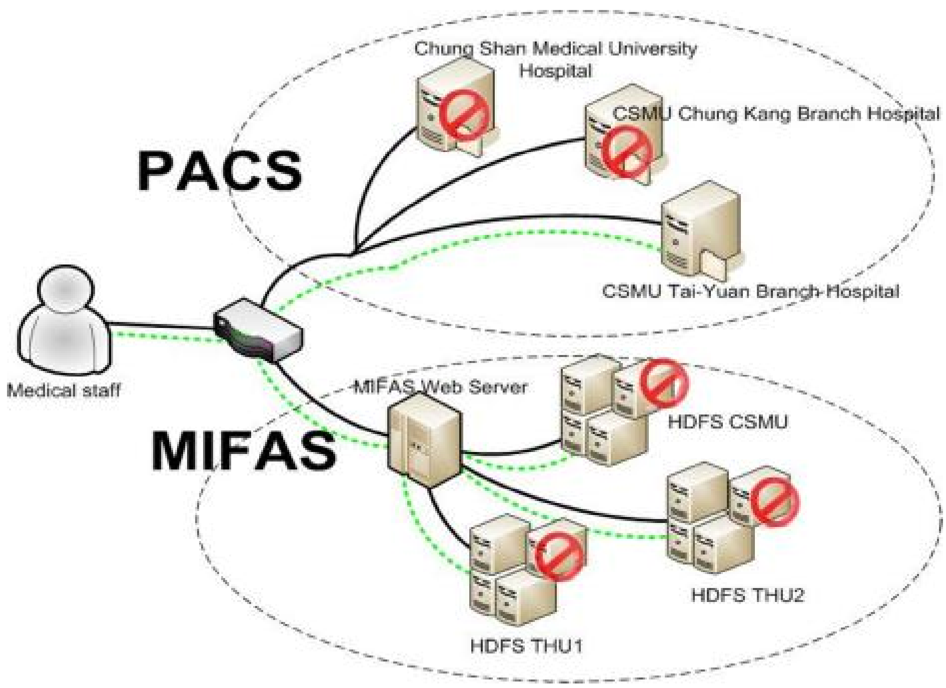

Figure 4.6 Hardware broken in both environments

4.5 Experiment 3: Compare Retrieve Times from MIFAS and PACS Cross Hospitals

Due to HDFS have good performance under WAN, we simulated the environment as shown Figure 4.7. Generally, PACS in hospital always provide a single download node which exists many potential problems, on the contrary, the MIFAS was more suit for multi-users accessing under WAN because the flexible architecture. In Figure 4.8, we simulated PACS and MIFAS accessing same CR medical record (7.1MB) under diverse user numbers. When users number increased, MIFAS could more effectively to handle more users. In same experiment, we tested it again, and this time we enabled co-allocation mechanism in MIFAS as shown Figure 4.9.

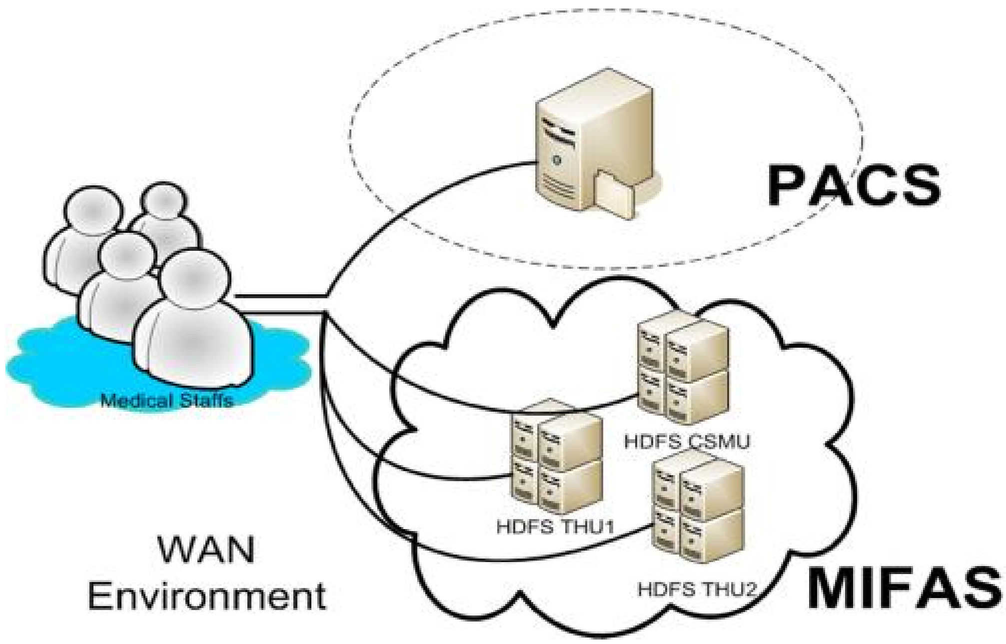

Figure 4.7 User Face to PACS and MIFAS on WAN

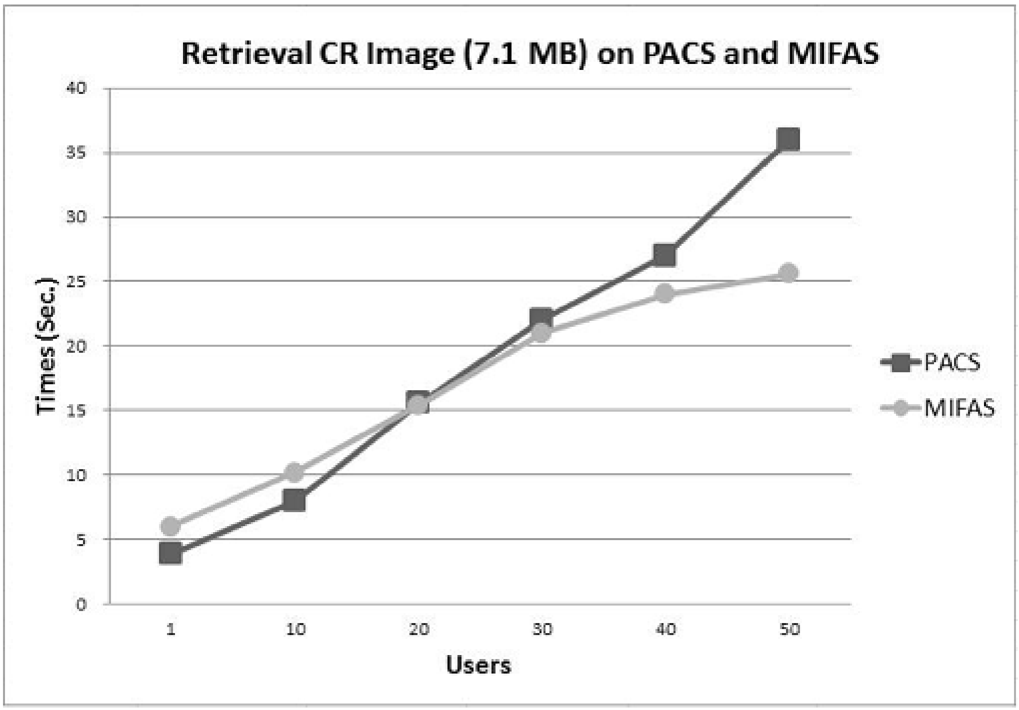

Figure 4.8 Retrieval CR Image on Both System

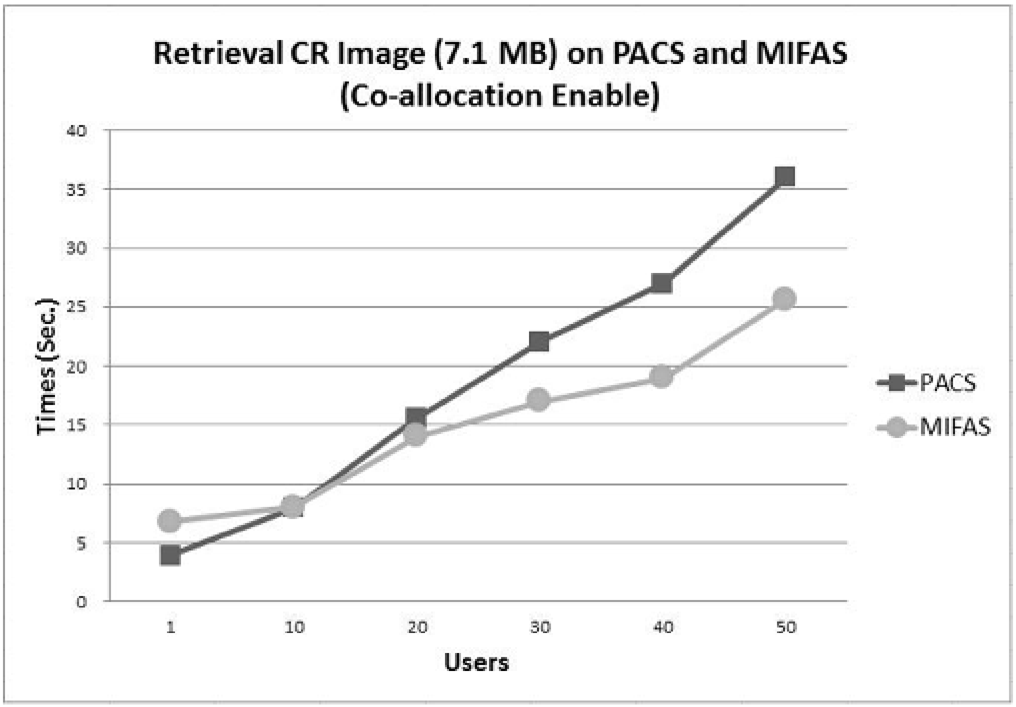

Figure 4.9 Retrieval CR Image on PACS and MIFAS (Co-allocation Enable)

4.6 Experiment Summary

According to experimental results, we obtained system features of MIFAS and PACS, as Table 4.3.

Table 4.3 System Features of MIFAS and PACS

Feature	MIFAS	PACS
Stable	Yes	Yes
Reliable	Yes	Yes
Easy Implement	Yes	No
Easy Exchange	Yes	No
HTTP Access	Yes	Yes
Speed Up on WAN	Yes	No
Flexible Storage	Yes	No
Total Cost Ownership	Low	High

Chapter 5

Conclusions

This paper proposes the MIFAS solution Medical Image transfer rate time, which integrates Medical Image into cloud environments. MIFAS is a flexible, stable and reliable system and proves Medical Images sharing, storing and exchanging issues is available under our thesis. Currently, MIFAS offers five advantages.

Scalability: Always extend the nodes as per requirement. Adding a node to the network is as simple as hooking a Linux box to the network and copying few configuration files. Also, Hadoop provides details about the available space in the Cluster. Therefore, according to this report, one can decide on adding a node or not.

Cost effective: Since the Linux nodes are always cheap; there is no need to invest much on the hardware as well as OS.

Best Strategy: Besides the Hadoop platform could offer us a distributed file systems, we also use the Co-allocation mechanism to get the best strategy to retrieve Medical Images.

Replication: Because the Replication Location Service in MIFAS Middleware data can be saved completed, the data can be easily shared through different private cloud.

Easy Management: We provide friendly management interface. And through the interface we can easy setting and management of the private cloud environment in MIFAS.

Furthermore, the MIFAS is not only a good practice of implementation medical image file accessing on cloud but also the goal of Medical Image Exchanging to enhance patients and their caregivers share information can be fully achieved. Through the MIFAS system with a lot less expense, redundancy in medical resources is believed to be acceptable.

References

[1] Apache Hadoop Project, http://hadoop.apache.org/hdfs/

[2] Canada Health Infoway, http://www.infoway-inforoute.ca

[3] American Recovery And Reinvestment Act(ARRA),http://www.ama-assn.org

[4] Department of Health, Executive Yuan, R.O.C. (TAIWAN) “Electronic medical record project”, http://emr.doh.gov.tw

[5] Mental Research Institute, http://www.mri.org

[6] Wiki-Electronic Medical Record, http://en.wikipedia.org/wiki/Electronic_medical_record

[7] Chervenak, E. Deelman, I. Foster, L. Guy, W. Hoschek, A. Iamnitchi, C. Kesselman, P. Kunszt, and M. Ripeanu, B. Schwarz, H. Stockinger, K. Stockinger, and B. Tierney. “Giggle: A Framework for Constructing Scalable Replica Location Services,” in Proc. SC, pp. 1-17, 2002.

[8] Chervenak, I. Foster, C. Kesselman, C. Salisbury, and S. Tuecke, “The data grid: Towards an architecture for the distributed management and analysis of large scientific datasets,” Journal of Network and Computer Applications, 23(3), pp. 187-200, 2001.

[9] G. V. Koutelakis, D. K. Lymperopoulos, Member, IEEE, “A Grid PACS Architecture: Providing Data-centric Applications through a Grid Infrastructure”, Proceedings of the 29th Annual International Conference of the IEEE EMBS Cite International, Lyon, France, ISBN: 978-1-4244-0787-3, August 23-26, 2007.

[10] N. E. King , B. Liu , Z. Zhou , J. Documet , H.K. Huang “The Data Storage Grid: The Next Generation of Fault-Tolerant Storage for Backup and Disaster Recovery of Clinical Images” Medical Imaging 2005: PACS and Imaging Informatics, edited by Osman M. Ratib, Steven C. Horii, Proceedings of SPIE Vol. 5748, pp. 208-217, 2005.

[11] S. Vazhkudai, “Enabling the Co-Allocation of Grid Data Transfers,” Proceedings of Fourth International Workshop on Grid Computing, pp. 44-51, 17 November 2003.

[12] C.T. Yang, S.Y. Wang, C.H. Lin, M.H Lee, and T.Y. Wu, “Cyber Transformer: A Toolkit for Files Transfer with Replica Management in Data Grid Environments,” Proceedings of the Second Workshop on Grid Technologies and Applications (WoGTA’05), pp. 73-80, December 2005.

[13] C.T. Yang, C.H. Chen, K.C. Li, and C.H. Hsu, “Performance Analysis of Applying Replica Selection Technology for Data Grid Environments,” PaCT 2005, Lecture Notes in Computer Science, vol. 3603, pp. 278-287, Springer-Verlag, September 2005.

[14] C.T. Yang, S.Y. Wang, and C.P. Fu, “A Dynamic Adjustment Mechanism for Data Transfer in Data Grids,” Network and Parallel Computing: IFIP

International Conference, NPC 2007, Lecture Notes in Computer Science, vol. 4672, pp. 61-70, Springer, ISSN 1611-3349, September 17-20, 2007.

[15] C.T. Yang, Y.C. Chi, T.F. Han and C.H. Hsu, “Redundant Parallel File Transfer with Anticipative Recursively-Adjusting Scheme in Data Grids”, ICA3PP 2007, pp. 242–253, 2007.

[16] C.T. Yang, M.F. Yang, Wen-Chung Chiang “Enhancement of anticipative recursively adjusting mechanism for redundant parallel file transfer in data grids,” J Network Comput Appl (2009), doi:10.1016/j.jnca.2009.02.002

[17] C.T. Yang, I.H. Yang, K.C. Li, and C.H. Hsu “A Recursively-Adjusting Co-Allocation Scheme in Data Grid Environments,” ICA3PP 2005 Algorithm and Architecture for Parallel Processing, Lecture Notes in Computer Science, vol. 3719, pp. 40-49, Springer-Verlag, October 2005.

[18] C.T. Yang, I.H. Yang, K.C. Li, and S.Y. Wang, “Improvements on Dynamic Adjustment Mechanism in Co-Allocation Data Grid Environments,” The Journal of Supercomputing, Springer, vol. 40, no. 3, pp. 269-280, 2007.

[19] Zheng Zhou, Sander S. Chao, Jasper Lee, Brent Liu, Jorge Documet and H. K. Huang, "A data grid for imaging-based clinical trials", Proc. SPIE 6516, 65160U (2007); doi:10.1117/12.710636

[20] Chao-Tung Yang, Yao-Chun Chi, Ming-Feng Yang and Ching-Hsien Hsu, “An Anticipative Recursively-Adjusting Mechanism for Redundant Parallel File Transfer in Data Grids,” Proceedings of the Thirteenth IEEE Asia-Pacific Computer Systems Architecture Conference (ACSAC 2008), Aug. 4-6, 2008 in Hsinchu, Taiwan.

[21] Gopinath Ganapathy and S. Sagayaraj, "Circumventing Picture Archiving and Communication Systems Server with Hadoop Framework in Health Care Services," Journal of Social Sciences, vol. 6, pp. 310-314, 2010. http://www.scipub.org/fulltext/jss/jss63310-314.pdf

[22] Venner, J., 2009. Pro Hadoop. 1st Edn., Apress, ISBN:13: 9781430219439, pp: 440.

[23] L. Faggioni, et al., "The future of PACS in healthcare enterprises," European Journal of Radiology, 2010.

[24] E. Bellon, et al., "Trends in PACS architecture," European Journal of Radiology, 2010.

[25] R. Grossman, et al., "Compute and storage clouds using wide area high performance networks," Future Generation Computer Systems, vol. 25, pp. 179-183, 2009.

[26] Sutton LN. PACS and diagnostic imaging service delivery—A UK perspective. Eur J Radiol (2010), doi:10.1016/j.ejrad.2010.05.012.

[27] Ganglia Monitoring System, http://ganglia.sourceforge.net/

[28] DICOM, http://medical.nema.org/

[29] J. Shafer, et al., "The Hadoop distributed filesystem: Balancing portability and performance," in Performance Analysis of Systems & Software (ISPASS), 2010 IEEE International Symposium on, White Plains, NY 2010, pp. 122-133.

[30] L. Xuhui, et al., "Implementing WebGIS on Hadoop: A case study of improving small file I/O performance on HDFS," in Cluster Computing and Workshops, 2009. CLUSTER '09. IEEE International Conference on, 2009, pp. 1-8.

[31] A. L. N. Reddy, J. Wyllie, and K. B. R. Wijayaratne. Disk scheduling in a multimedia I/O system. ACM Transactions on Multimedia Computing, Communications, and Applications, 1(1):37–59, 2005.