2010 13th IEEE International Conference on Computational Science and Engineering

# Implementation of a Medical Image File Accessing System on Cloud Computing

Chao-Tung Yang Lung-Teng Chen Wei-Li Chou Kuan-Chieh Wang

High-Performance Computing Laboratory
Department of Computer Science, Tunghai University
Taichung, 40704, Taiwan R.O.C
ctyang@thu.edu.tw, g97357002@thu.edu.tw, g98357002@thu.edu.tw, g98357006@thu.edu.tw

***Abstract*—Large scale cluster based on cloud technologies has been widely used in many areas, including the data center and cloud computing environment. The purpose of presenting the research paper in this field was to solve the challenge in Medical Image exchanging, storing and sharing issues of EMR (Electronic Medical Record). In recent years, many countries invested significant resources on the projects of EMR topics. The benefit of the EMR included: Patient-centered Care, Collaborative Teams, Evidence-based Care, Redesigned Business Processes, Relevant Data Capture and Analysis and Timely Feedback and Education. For instance, the ARRA-HIT project in Untied States (2011 - 2015), Health Infoway project in Canada (2001 - 2015) and NHIP project in Taiwan, etc. Aim to the topic of EMR, we presented a system called MIFAS (Medical Image File Accessing System) to solve the exchanging, storing and sharing on Medical Images of crossing the different hospitals issues. Through this system we can enhance efficiency of sharing information between patients and their caregivers. Furthermore, the system can make the best-possible patient-care decisions.**

***Keywords- EMR, PACS, Hadoop, HDFS, Co-allocation, Cloud Computing***

## I. 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 paper 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 paper; it was built on the Hadoop [1] platform to solve the exchanging, storing, sharing issues in Medical Images. In this paper, 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 paper was organized as following. Background review and studies were presented in Section 2. The System Architecture was introduced in Section 3. Experimental results were presented in Section 4. Finally, Section 6 concluded this article.

## II. BACKGROUND

### A. 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 HER. 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

978-0-7695-4323-9/10 $26.00 © 2010 IEEE
DOI 10.1109/CSE.2010.48

same study, they also presented a HPACS system but lacked of management interface.

### B. 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. 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].

### C. 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].

## III. System Design and Implementation

In Figure 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.

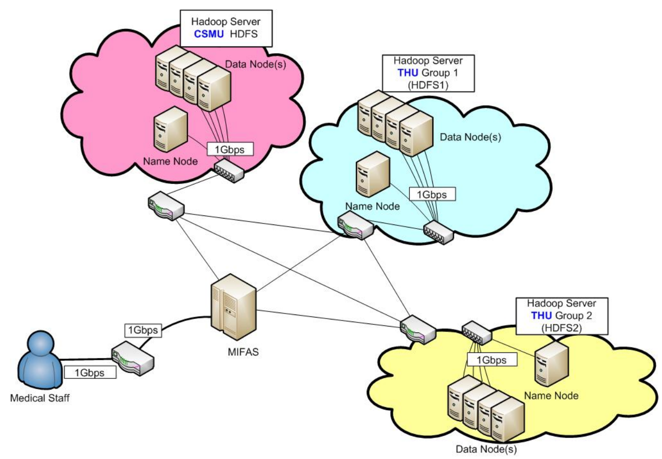


Figure 1. Overview of Distribution File System.

### A. System Architecture

In Figure 2, MIFAS was developed on cloud environment. The distribution file system was built on HDFS of Hadoop environment (Section 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.

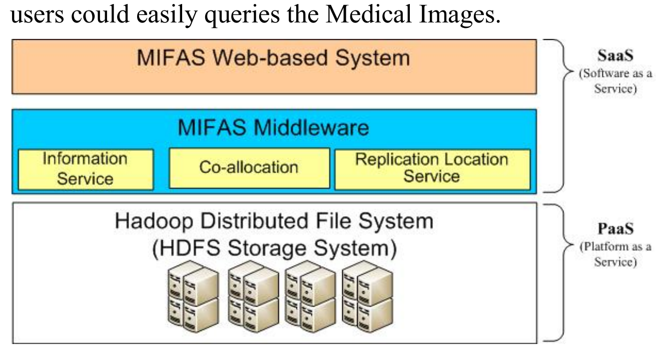


Figure 2. System architecture of MIFAS

**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).

**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.

### B. System Workflow

In Figure 3, 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.

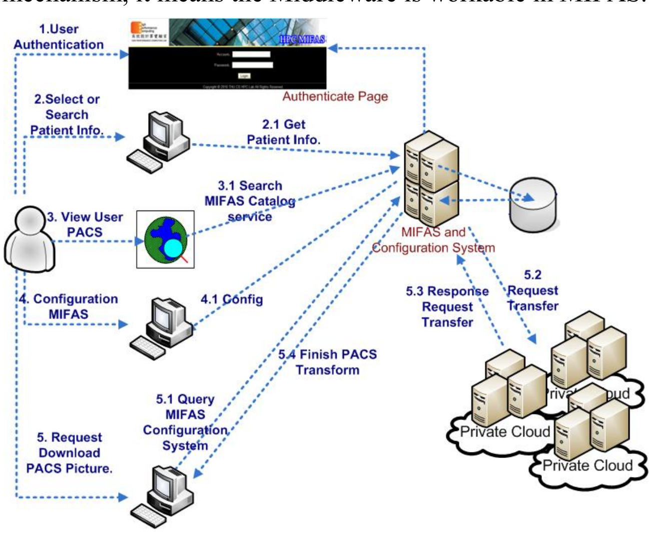


Figure 3. System workflow of MIFAS

## IV. EXPERIMENTAL ENVIRONMENTS AND RESULTS

### A. Experimental Environments

In this section, we compared MIFAS with PACS. The PACS system in Chung Shan Medical University Hospital is shown in Figure 4. 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.

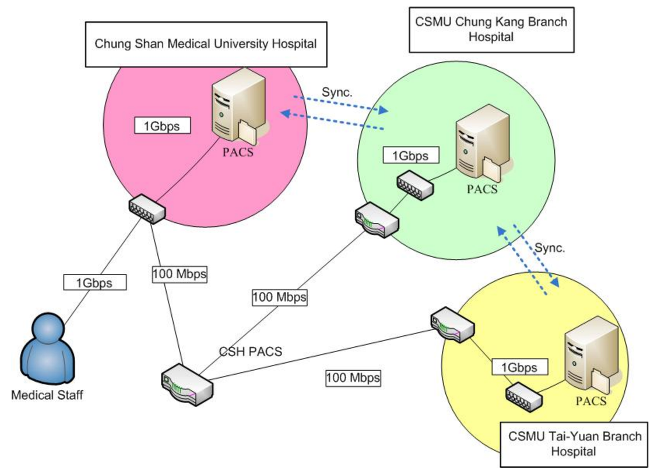


Figure 4. System workflow of MIFAS

TABLE I. 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 |

### B. Experiment 1: Compare Images Retrieval Times from PCSA and MIFAS

In this experiment, we gave same Medical Images as Table 1 in PCAS and MIFAS. In a general situation we can describe the PACS vs. MIFAS in Figure 5. 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 II. 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 |

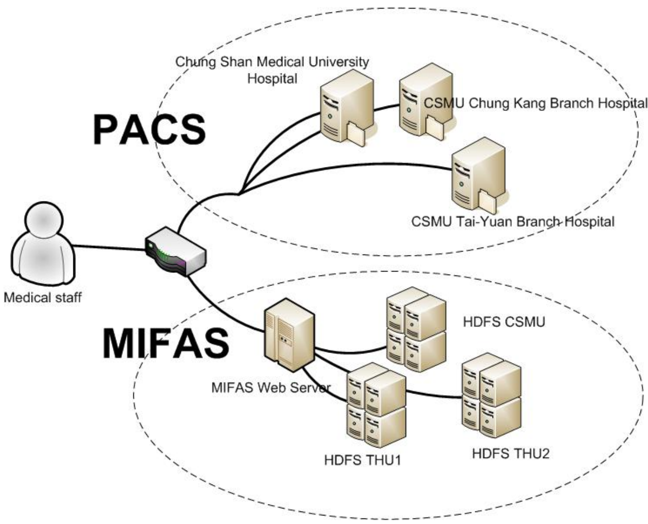


Figure 5. System workflow of MIFAS

We tested 500 times for each site, and get the average time result in Figure 6. Results in proximal site can be discovered to be more time-consuming in Figure 7. The result of Figure 6 and Figure 7 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.

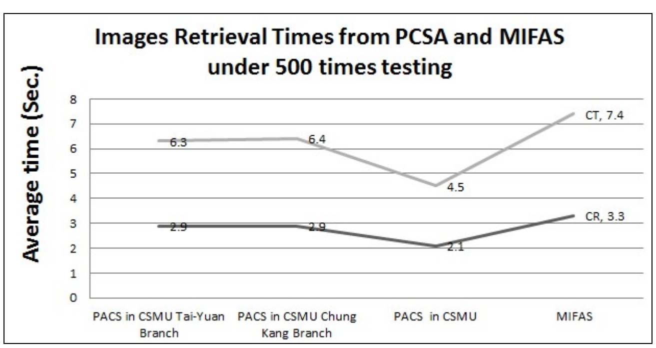


Figure 6. Image Retrieval Times Result

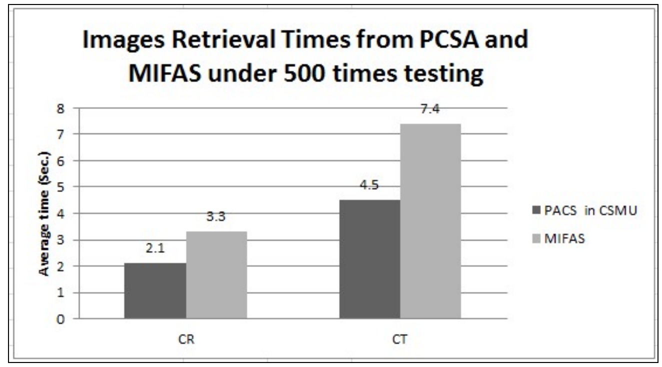


Figure 7. Image Retrieval Times Result in Proximal

### *C. 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 8, PACS in CSMU is not available, but the MIFAS can still work under a good situation.

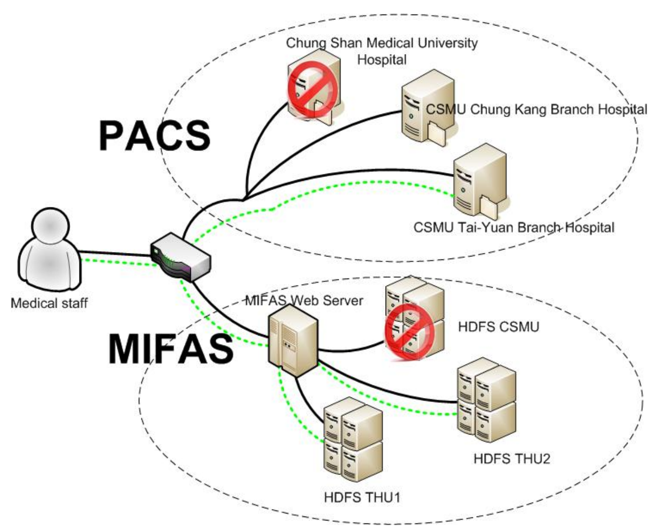


Figure 8. System configuration for Experiment 2

*D. Hardware Broken in Both Environments*

In Figure 9, 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.

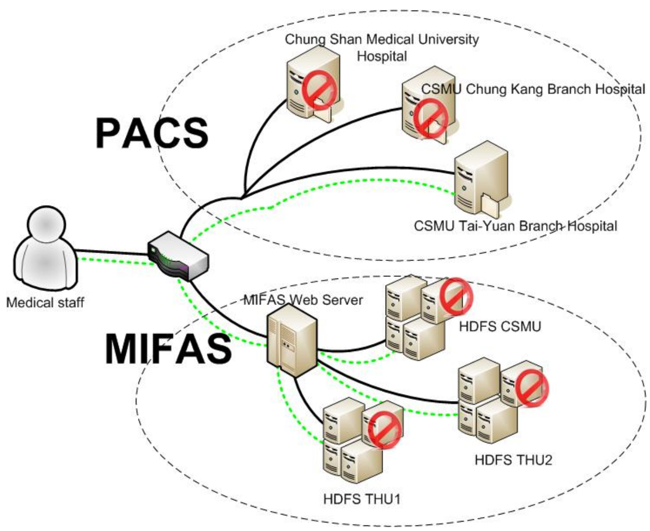


Figure 9. Hardware broken in both environments

## V. CONCLUSIONS

MIFAS is a flexible, stable and reliable system and proves Medical Images sharing, storing and exchanging issues is available under our thesis. Therefore, Medical Images can easily share cross different hospitals. MIFAS offers advantages as below:

**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.

## ACKNOWLEDGMENT

This work is supported in part by the National Science Council, Taiwan R.O.C., under grants no. NSC 99-2220-E-029-004 and NSC 99-2622-E-029-001-CC2.

## 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] Z. Zhou, S. S. Chao, J. Lee, B. Liu, J. Documet, H.K Huang "A Data Grid for Imaging-based Clinical Trials" Medical Imaging 2007: PACS and Imaging Informatics, edited by Steven C. Horii, Katherine P. Andriole, Proc. Of SPIE Vol. 6516, pp. 65160U, 2007.

[20] Chao-Tung Yang, Shih-Yu Wang, and William C. Chu, "A Dynamic Adjustment Strategy for Parallel File Transfer in Co-Allocation Data Grids," Journal of Supercomputing, Springer Netherlands, Article in Press, June 2009. (DOI 10.1007/s11227-009-0307-4)

[21] 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.

[22] 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

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

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

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

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

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

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