Advanced Materials Research *Online: 2013-09-18*
ISSN: 1662-8985, Vols. 756-759, pp 2037-2041
doi:10.4028/www.scientific.net/AMR.756-759.2037

CloudDICOM: A Large-scale Online Storage and Sharing System for DICOM Images

Lijun Liu[1, a], Qingsong Huang[2, b]

[1] Department of Biomedical Engineering, Kunming University of Science and Technology, Kunming, 650500, China

[2] Key Laboratory of Computer Application Technology of Yunnan, Department of Computer Science and Technology, Kunming University of Science and Technology, Kunming, 650500, China

[a]email:cloneiq@126.com, [b]email:kmustailab@hotmail.com (Corresponding Author)

Keywords: DICOM; WADO; Computing-cloud; Hadoop; HBase; Online Storage and Sharing

Abstract. Online storage and sharing for large-scale DICOM images becomes increasingly important for medical organizations or large hospitals. This paper presents a distributed architecture based on Hadoop and HBase to support online storage and sharing for DICOM images. An experimental system called CloudDICOM is designed and realized based on this architecture. The paper focuses on designing the architecture, workflow, data schema, and then on analyzing the components in CloudDICOM. Firstly, DICOM messages sent by clients will be received, converted and stored into Hadoop and HBase. Then, these messages will be indexed and generated query and WADO index database. The components of DICOM query and WADO based on this index are implemented to provide online DICOM query and WADO service for clients. The test results demonstrate that CloudDICOM can provide online storage and sharing service for large-scale medical images, and support standard DICOM Query and WADO service.

Introduction

With the development of medical imaging equipment (CT, MRI) technologies, the quality of the medical image has been greatly improved and the space occupied has also increased. The size of image data of an examination generated by CT or MRI can reach 50-100MBs, CR or DR equipment will produce more than 100MBs image data. In large hospital, the PACS system may generate 20GBs image data every day. Further more, the medical images are derived from different devices and have different data formats. Therefore, the medical image management is facing a lot of problems and challenges. For example, lack of uniform standards for data-sharing, high cost for independent construction, difficult to management and upgrades and maintenance [1][2][3], especially for online storage and sharing.

DICOM is a typical medical image standard and a network communications protocol for use with medical image diagnosis equipment, such as MRI and CT defined by the ACR (American College of Radiology) and NEMA (National Electrical Manufacturers Association) [4]. This standard has been developed with an emphasis on diagnostic medical imaging as practiced in radiology, cardiology and related disciplines, however, it is also applicable to a wide range of image and non-image related information exchanged in clinical and other medical environments [5]. All DICOM commands and most data attributes are always bonded with the four-level information model. Because the current radiology workflow is study-centric, the study root is the most widely supported in DICOM applications and is always provided by default [6]. WADO is neither a new medical communication protocol nor the transformation of a compatible DICOM service into its Web equivalent [7]. It is standardized as the Web extension to DICOM which provides a simple mechanism for accessing a DICOM persistent object from HTML pages or XML documents, through HTTP/HTTPs protocol, using DICOM UIDs (Unique Identifiers) [8]. The parameter of request type is required and its name shall be “requestType”, value shall be "WADO". The clients can also select the size, frame number, or other parameters of the resulting image for non-DICOM images. For Example, retrieving a region of a DICOM image, converted if possible in JPEG2000, with annotations burned into the image containing the patient name and technical information, and mapped into a defined image size [8].

Cloud Computing represents a new way of delivering computing resources in the high end computing environment which provides computation, software, data access, and storage services that do not require end-user knowledge of the physical location and configuration of the system that delivers the services [1]. Hadoop is a reliable and scalable architecture for large-scale distributed data storage and high-performance computation on a network of inexpensive pieces of commodity hardware [9][10][11]. At present, Hadoop is widely used in building cloud-based applications in many well-known companies, such as Yahaoo, Fackbook, Amazon, Taobao, Baidu, ect. Facebook, for example, is adding more than 15TBs of data into its Hadoop cluster every day and is subsequently processing it all [12]. Hbase is a distributed, fault-tolerant, highly scalable, column-oriented no-SQL database after Google's Bigtable which built on top of Hadoop distributed file system (HDFS) to create a massively scalable and high-performance platform for dealing with heterogeneous data including non-textual data types (blob, clob etc.) [13]. Hbase is a good choice when you need random, real-time read/write access to your Big Data, which provides Bigtable-like capabilities on the basis of Hadoop and HDFS [14] and has the ability to build a large-scale structured storage cluster based on cheap PC Server..

In this paper, we design and implement a distributed storage and sharing system named CloudDICOM for DICOM Images based on Hadoop and HBase which has the ability to provide large-scale medical images online storage and sharing service for medical organization or large hospital, which support standard DICOM query and WADO service.

System Design and Implementation

1. Architecture Overview

In this section, we will illustrate how to integrate CloudCOM system with Hadoop and HBase. This system is designed as a fully distributed DICOM storage and sharing system based on Hadoop and HBase, which is consists of five components: DICOM parsing and conversion, DICOM storage, DICOM index, DICOM query and WADO Service. Figure 2 shows the architecture of this system.

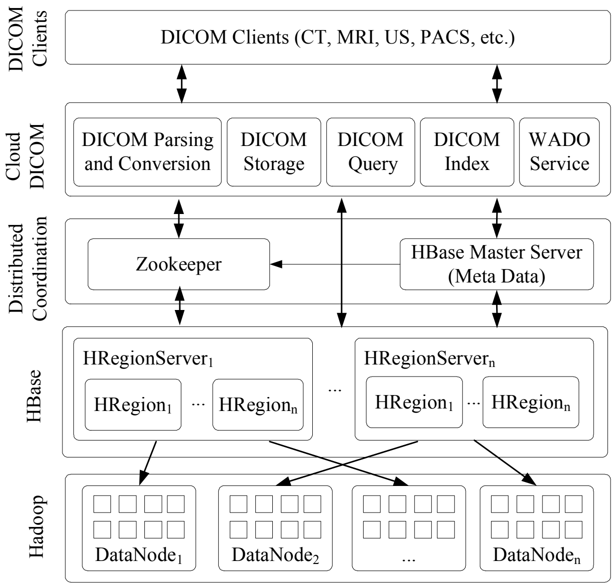

Fig.2. Architecture of CloudDICOM

2. System Data Schema Design

In HBase, the storage schema is very important, because it will affect the performance directly for reading and writing. According to the HBase official recommendations, the basic principles of schema design are to minimum column families and minimize the size of the row and column.

(1) DICOM Schema Design: According to the DICOM information model and the requirements of DICOM query and WADO, we designed one HTable named DicomTable to store the DICOM objects. As we known, data in HBase is stored sequentially in accordance with row-key, so making the frequently accessed data is stored together, and as much as possible try to make the data are distributed across various points of regions in the table are better for improving the reading and writing performance. So we use a composite row-key as the DicomTable's Rowkey which contains four parts: OrgUID, PatientID, Modility and Timestamps separated by "||". Thus, the DICOM images belong to the same organization or department and patient are stored together.

Despite one HTable only containing a column family is the best, we can introduce a second and third column family in the case where data access is usually column scoped. So the DicomTable consists of four column families: Patient, Study, Series and Image. Try to keep the column family names as small as possible, using P (Patient), D (Study), S (Series) and I (Image) as the name of column family. Each column family contains multiple columns which come from DICOM metadata. The figure 3 shows the DICOM data schema in HBase.

(2) Indexing Schema Design: The secondary index and join are basic characteristics of online business systems. HBase extends the Bigtable model, which only considers a single index, similar to a primary key in the RDBMS. To meet the DICOM online query and WADO services, we designed index schema to support the secondary index and join operation and the index is designed separately and stored in different HTable.

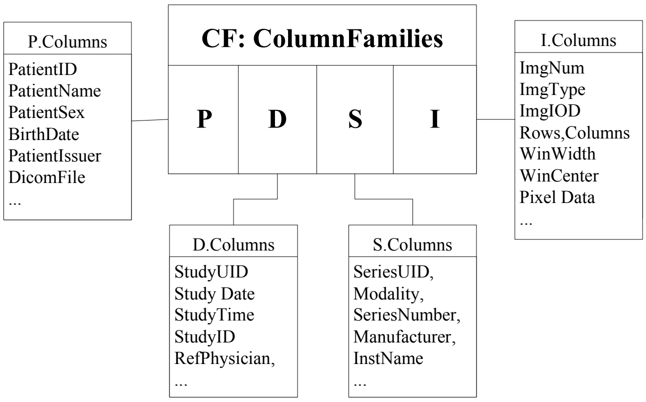

Fig.3. DICOM Data Schema

1) Query Index Schema: The row-key of query index consists of column family, column name, column value separated by "||". For example, P || PatientName || Rubo. Query index schema only has one family named IF (Index Family) which includes one column called ICV (Index Column Value) which stored a row-key list values come form DicomTable. The table1 shows the query index data in HBase.

Table.1. Query Index Data Example

No.	Rowkey	IF.ICV							
1	P||PatientName||Rubo	key_1	key_2	key_7	key_{12}	key_{15}	key_{18}	…	key_n
2	P||PatientName||Mike	key_3	key_5	key_6	key_8	key_{13}			
3	S||Modality||CT	key_3	key_5	key_8	key_{12}	key_{13}	key_{15}	key_{17}	
…	…	…	…	…	…	…	…	…	…
m	S||Modality||MR	key_4	key_6	key_7	Key_9	key_{10}	…	key_k	

Note: $key_{1\ldots n\ (k)}$ comes from the DicomTable's Rowkey.

For example, if a query is PatientName = 'Mike' and Modality = 'CT', we can directly navigate to no.2 and no.3 and get two row-key lists, and then the same row-key will be merged into one list as key_3, key_5 , key_8 and key_{13}.

2) WADO Index Schema: The WADO index row-key is consists of parameter values (studyUID, seriesUID, objectUID) separated by "||", which are required in one WADO request. It has the same family and columns with the Query Index, but only stored one row-key value that comes from DicomTable. Thus, a WADO request can be located to a DICOM object accurately in accordance with the parameters.

3. How the CloudDICOM System Works

In the following section, we will describe the system's workflow and functionality of the various components. At first, this system will parse DICOM objects sent from DICOM clients and convert them to HBase storage format, which will be stored into HBase database. Secondly, these DICOM objects will be scanned and indexed in order to generate Query and WADO index database. In the end, the Query and WADO components based on the index database will provide online queries and web access to DICOM object services for the DICOM clients. The workflow of CloudDICOM can be expressed in Fig 4. As figure 4 shown, each component plays an important and different role in CloudDICOM system as follows:

(1) DICOM Parsing and Conversion: Listening, receiving and parsing DICOM messages sent by clients, and converting them to HBase storage formats according to the HTable Schema designed in this paper.

(2) DICOM Storage: Storing these DICOM objects and their metadata into HBase database using MapReduce programs.

(3) DICOM Index: Scanning the metadata of the DICOM objects stored in HBase, generating query and WADO index, and then storing them into index database.

(4) DICOM Query: Receiving query requests from DICOM clients, parsing query parameters, searching DICOM objects from DICOM query index database in accordance with these parameters, obtaining DICOM object list from HBase according to index values and returning them to clients.

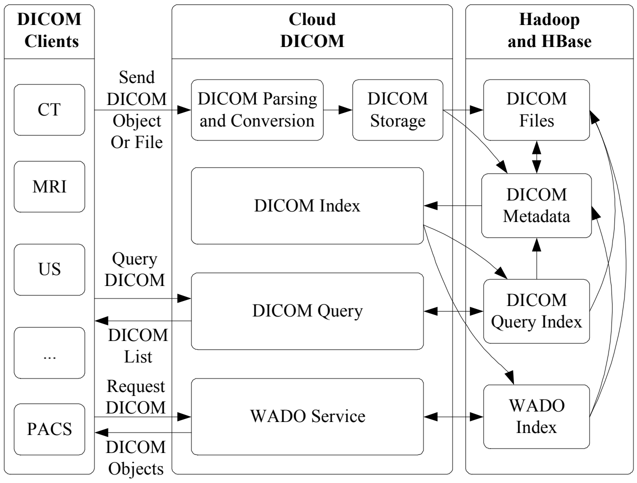

Fig.4. Workflow of CloudDICOM

(5) WADO Service: Receiving WADO requests from DICOM clients, parsing query parameters, retrieving DICOM objects from the WADO index database according to these parameters, getting DICOM objects from HBase according to index value (row-key), extracting requested information from the object and returning to clients.

Experiment and Test result

The experimental system is constructed on a Hadoop-HBase cluster using one master, six slaves and a CloudDICOM server. The testing hardware for the master and CloudDICOM server was Intel(R) Xeon(R) CPU and 8GBs of RAM and All slaves have the same hardware environment: Intel (R) Core (TM) 2 Duo CPU and 4GBs of RAM.

All servers under Hadoop cluster are running under Ubuntu11.10 operating system connected by the same gigabit switch. Hadoop-0.20.203.0 and HBase-0.94.1 were chosen to complete the cluster. Dcm4che2 DICOM toolkit was used to implement the DICOM parsing and communication. We also developed a DICOM client which can produce a lot of storage, query and WADO requests for this system at the same time. Due to the limitation of experimental condition, only more than 20GBs DICOM image files were stored in HBase. The testing results show the performance of writing and reading can meet the requirements of online storage and query for large-scale DICOM images.

Conclusion

The proposed system focuses on how to store and share large-scale DCIOM images in a distributed way using Hadoop and HBase. An experimental system called CloudDICOM was designed and implemented which was able to receive DICOM messages and store them into Hadoop and HBase. We realized the components of DICOM index, DICOM query and WADO service. In our experimental condition, the system has the ability to meet the distributed online storage and sharing for large-scale DICOM images, but the ultra-large-scale images storage and sharing needs further testing and verification.

In future work, we will continue to optimize the system architecture and DICOM index, expand the cluster size and increase the amounts of data to support online storage and sharing for the ultra-large-scale DICOM images. And further more, the DICOM client should be closer to the actual application environment.

Acknowledgement

In this paper, the research was sponsored by Science and Technology Innovation Fund Project of China (Project No. 10C2615305099) and Social Development of Science and Technology Project of Yunnan Province (Project No. 2010CA016).

References

[1] G.Kanagaraj and A.C.Sumathi, Proposal of Sharing Medical Images in an OpenSource Cloud Computing System, International Journal of Computing and Corporate Research, 2012, vol. 2.

[2] C. C. Teng, J. Mitchell, C. Walker, et al, A Medical Image Archive Solution in the Cloud, 2010 IEEE International Conference on Software Engineering and Service Sciences (ICSESS), 2010.

[3] Frost and Sullivan, Prepare for Disaster & Tackle Terabytes When Evaluating Medical Image Archiving, 2008, http://www.frost.com.

[4] T. Hiroyasu, Y. Minamitani, M. Yoshimi and M. Miki, Distributed PACS using Network Shared File System, WORLDCOMP'11, 2011, USA, July 18-21.

[5] DICOM standard Part 1: Introduction and Overview, http://medical.nema.org/Dicom/2011/ 11_01pu.pdf .

[6] O. S. Pianykh, Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guid, 2008, Springer-Verlag.

[7] G. V. Koutelakis and D. K. Lymberopoulos, WADA Service: An Extension of DICOM WADO Service, IEEE trans on Information Technology in Biomedicine, 2009, vol. 13, no. 1, pp. 121-130.

[8] DICOM standard Part 18: Web Access to DICOM Persistent Objects (WADO) , http://medical.nema.org/Dicom/2011/11_18pu.pdf .

[9] D. Carstoiu, E. Lepadatu and M. Gaspar, "Hbase-nonSQL Database, erformances Evaluation", International Journal of Advancements in Computing Technology, vol. 2, no. 5, December 2010.

[10] R. C. Taylor, "An overview of the Hadoop / MapReduce/HBase Framework and Its Current Applications in Bioinformatics", BOSC2010, Boston, MA, USA. July 2010.

[11] A. Khetrapal and V. Ganesh, "HBase and Hypertable for Large Scale Distributed Storage Systems", (http://cloud.pubs.dbs.uni-leipzig.de/node/46) .

[12] M. N. Vora, Hadoop-HBase for Large-Scale Data, 2011 IEEE International Conference on Computer Science and Network Technology, December 24-26, pp. 601-605.

[13] W. Peng, Approached the cloud, 2009, Beijing: The People's Posts and Telecommunications Press.

[14] When Would I Use HBase? , http://hbase.apache.org.

Information Technology Applications in Industry, Computer Engineering and Materials Science
10.4028/www.scientific.net/AMR.756-759

CloudDICOM: A Large-Scale Online Storage and Sharing System for DICOM Images
10.4028/www.scientific.net/AMR.756-759.2037