MOVING BIG DATA MEDICAL IMAGE EHR TO THE CLOUD: AN APPROACH TO DESIGN FAST PERFORMANCE TRANSPORT SOFTWARE USING FASP

Sreekanth R 1, Dr. Gondkar RR 2, Augustin Minalkar 3
1 Research Scholar, R&D Center, Bharathiyar University, Coimbatore, TamilNadu, India
2 Professor, Department of MCA, Brindavan College, Bangalore, India
3 HoD, Faculty of Computing, Botho University, Gabarone, Botswana

ABSTRACT

Cloud computing provides efficient and scalable resources to store, process and access data. Healthcare organizations are moving to store their Electronic Health Records (EHR) onto the cloud as it is an emerging computation model. With the increase in healthcare data collected from various geographical locations and to move such huge big data onto the cloud for efficient processing is a challenging issue. Most of the healthcare data is unstructured and moving this data to cloud for processing requires a fault-tolerant system that can be deployed on low cost hardware. Hadoop distributed file system (HDFS) is fault-tolerant and deployed on low cost hardware. Recently researchers have proposed online algorithms based on cost-minimizing approach. These algorithms at any point of time aggregate and efficiently process the data by selecting the data center. In this paper we first study the cost-minimizing algorithms online lazy migration (OLM) and randomized fixed horizon control (RHFC) proposed by researchers recently. We then study Fast and secure protocol (FASP) to move big data electronic health records to the cloud. As EHR data is unstructured and large image data like X-ray, scan images, graph images (ECG, Treadmill, Echo) need to be moved to cloud, an efficient fast performance transport software is required which uses FASP considering these parameters

Keywords— Algorithms, Cloud Computing, EHR, FASP, Hadoop

[1] INTRODUCTION

Cloud computing guarantee to provide on demand, scalable, pay-as-you-go computation and storage capacity and enabling server resources such as CPU, bandwidth to users with managed clouds. Cloud platforms from Amazon EC2 and S3, Microsoft Azure, Google App Engine, Rackspace, salesforce have demonstrated an advanced way of managing the cloud by the concept of do-it-together. These platforms provide the users a shared pool of servers from various data centers and provide services to all users by virtualization technologies. Data-intensive applications i.e. big data applications are depending on cloud platforms for execution
of their applications as cloud is elastic and resource provisioned. The move from paper based medical records to EHR systems has generated big data [1]. Big data provides an opportunity to healthcare experts to take better decisions and minimize costs [2] by providing personalized medical care to patients. As 80% of EHR data is unstructured and these data rely on clouds for analyzing and processing their data sets. Computing framework MapReduce and Hadoop provides framework to process large scale data sets. There are bottlenecks when data to be transferred to cloud especially when it is big data. The cost of transferring the data can add up quickly making the data transfer cost an issue. One of the solution given to avoid the huge data transfer cost is from Amazon web service [3] of its new cloud front service. This can reduce the bandwidth of about 1,500Mbit/sec when we intend to transfer 10TB of data [4]. Two online algorithms proposed for the data migration by the researchers is carefully studied. OLM and RFHC algorithm for optimal migration of large data sets of EHR on to the cloud can be discussed.

The remaining sections of the paper are organized as follows. Section II discusses the related work carried out, section III discuss about the EHR migration issues on to cloud. Section IV discuss about the solution for data migration to cloud. Section V focus on evaluation and Section VI concluded remarks are made to complete the paper.

[2] RELATED WORK

Healthcare organizations are showing interest in migrating the various applications and huge EHR data to the cloud platform. There are models developed for migrating health care data to cloud. F. Desprez [5] proposed a cloud brokering algorithm improvement that optimizes the infrastructure deployment cost considering different data storage and transfer policies. Data transfer and storage is taken into account while designing the algorithms. The algorithms were validated on SimGrid Cloud broker (SGCB). They deployed n VMs, \(v_1, \ldots, v_n\) across the m available clouds \(c_1, \ldots, c_m\), using a given number instance types, \(i_1, \ldots, i_l\). Experiments have concluded that keeping images in every cloud results in higher storage cost but reduction is seen in the final bill. Linquan Zhang [6] proposed two online algorithms for moving the data to the cloud. The OLM algorithm achieves a worst-case competitive ratio of as low as 2.55 under real world settings and not considering future information. The RHFC algorithm provides a decreasing competitive ration with increasing size of the look ahead window. Further an algorithm for moving deferrable big data to cloud is proposed by Linquan Zhang [7]. The algorithm will minimize the traffic cost for uploading the big data to the cloud.

[3] MEDICAL IMAGES EHR DATA MIGRATION ISSUE ON TO CLOUD STORAGE

In this section we first discuss the issues of migrating medical images to the cloud for storage. As medical images are huge and are usually in size of terabytes the bandwidth will slow down when it is moved to the cloud over the internet. Medical images on cloud are regulated by HIPPA [10] and healthcare organizations have to comply with these requirements. Cloud computing should also adhere to EHR regulations. The traditional WAN based methodology cannot move the terabytes of medical image data to the cloud with the speed as required by the organizations. They just use the high speed bandwidth for minimal time and therefore it is unsuitable for huge volume of data and this causes delays. As cloud computing promises three key advantages such as reducing the risk of upfront investment, pay-as-you-go model help to improve the cash flow, consume the required resources at any point of time [11]. The way how the terabytes of data moved to the cloud is by traditional means of shipping hard disk to the
cloud providers, and transfer the data through web using the TCP based transfer methods such as FTP or HTTP.

These types of migrating the images to the cloud create issues of loss or damage of the data or engaging number of engineered IT systems. There are two bottlenecks while transferring such huge data to the cloud. One is while transferring the data over WAN; the other is at the cloud data center. For transfer of large data sets to the cloud these two bottlenecks to be addressed [8]. Fig1. Shows moving large medical images to the cloud using WAN. As WAN based transfers are TCP based a single HTTP uses less than 10 Mbps and multiple HTTP uses 10-100 Mbps [13]. This can slow down the transfer of terabytes of medical image data to the cloud. Transfer from local server to cloud storage can use multiple HTTP connections which again is a bottleneck while moving to a datacenter.

In HDFS the default size of the data chunk is 64MB. If we are trying to upload or download the data through applications which are larger sizes in Terabytes or Petabytes the file requires dividing it into thousands of chunks which is a bottleneck in transfer speed for the WAN. Cloud services need the technology to get the data transported at very high speed mechanism that can solve the issues of data migration to cloud. By optimizing the intra-cloud data transfer and providing consistent speed throughout the data path and medical image files of large data to be transferred at any distance over any network. High-speed software lines can be established which can transfer the image data directly to the cloud providers [8] and without any input and output hurdles.

**[4] SOLUTION FOR MEDICAL IMAGE DATA MIGRATION DIRECT TO CLOUD STORAGE**

The solution to transfer the medical image data to the cloud will be based on direct to cloud transport platform to transfer the data files or directories to, from and between cloud storage. The transport technology should be based on fast access and secure protocol so that the large image files can transfer with high speed transfer of files for uploading or downloading from the cloud. The technology used in transferring the image files to the cloud will provide the capabilities to perform the transfer at any distance up to the I/O limits of the platform. It will also support to transfer the image files in a single transfer session in a default size of 64MB chunks of various parts. For eg. Medical image data files up to 0.625TB of data can be transferred per single session on Amazon web services S3 [9]. The technology also provides adaptive bandwidth control to avoid network congestion. Security authentication and access control mechanisms and built in encryptions will help on client and server to control the secured access of files from the cloud storage. Fig2 shows the large medical image data can be transfer to cloud storage with fast and secure protocol which avoids the bottlenecks of input and output. This supports all the major cloud storage platforms.
[5] EVALUATION

The performance of the medical image file transfer can be tested on a single virtual machine with server software with FASP. Various test types conducted on the server for cloud storage with different dimension values and with platform limitations proved that the fast and secure protocol provides the best way to transfer large image files to the cloud without major bottlenecks. Table 1 below shows cloud storage with various test types and dimensions. Various platforms like Google, Microsoft azure, AWS S3 are taken into consideration for the tests.

<table>
<thead>
<tr>
<th>Test Area</th>
<th>Test type</th>
<th>Dimensions</th>
<th>Values</th>
<th>Platform limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load Test</td>
<td>File size</td>
<td>400 GB</td>
<td>AWS S3 up to 5TB</td>
<td>Google max Session 62 GB</td>
</tr>
<tr>
<td></td>
<td>Bandwidth</td>
<td>10,100,500 Mbps</td>
<td>Azure max bandwidth 400 Mbps</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data Sets</td>
<td>Large files up to 100 GB</td>
<td>Google file size 62 GB</td>
<td></td>
</tr>
<tr>
<td>Stress Test</td>
<td>File size</td>
<td>0 byte to 400 GB</td>
<td>AWS S3 up to 5TB for single session</td>
<td>Google max Session 62 GB</td>
</tr>
<tr>
<td></td>
<td>Bandwidth</td>
<td>500 Mbps – 1 Gbps</td>
<td>Azure max bandwidth 400 Mbps</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data Sets</td>
<td>Large files upto 100 GB</td>
<td>Google file size 62 GB</td>
<td></td>
</tr>
<tr>
<td>WAN</td>
<td>Performance test</td>
<td>Bandwidth 512 kbps – 10Gbps</td>
<td>Google max Session 62 GB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Packet loss rate</td>
<td>0%,0.1%,1%,5%,10%,20%</td>
<td>Azure max bandwidth 400 Mbps</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Block sizes</td>
<td>16kb – 4MB</td>
<td>Google max Session 62 GB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Operating systems</td>
<td>All major OS</td>
<td>Google max Session 62 GB</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Cloud storage testing

[6] CONCLUSION
Cloud services can be considered for storing medical image files by the healthcare organizations. As the cloud provides virtual centralization of applications, storage, computing and can be accessed by web applications. Health care industries have enormous opportunities for storing, archiving, sharing and accessing medical images in the cloud. The cloud allows the health care industry to manage the data more efficiently and cost-effectively. In this paper we first discuss on various cost effective algorithms proposed by various researchers to transfer the data to the cloud. The cloud enables to efficiently manage the large set of images with high bandwidth transfer. It can scale the capacity as and when needed. It can manage the authentication encryption and security protocols. Large datasets of image files need to be transferred to cloud storage by avoiding major bottlenecks with regards to input and output. By using WAN and TCP based transfers it is not possible to transfer Terabytes of image data to the cloud. By using fast and secure protocol sessions and virtual on demand servers it is possible to transfer the image files to the cloud by utilizing the maximum bandwidth.

REFERENCES


Sreekanth R, Dr Gondkar RR and Augustin Minalkar

Author[s] brief Introduction

Sreekanth R received his Masters Degree in computer Applications in 2001, and MPhil (Computer Science) in 2006 from Bharathidasan University, India. He is currently pursuing his PhD from Bharathiya University, India. He has 15 years of work experience in teaching various universities in india and abroad. He has published more than 20 research papers in National and International Journals and conferences. His area of research includes Big Data analytics, Cloud Computing.

Dr RR Gondkar received his PhD from Kuvempu University, India. He is currently working as Professor & HoD, Brindavan college Bangalore. He has published more than 40 papers in national and international conferences nd journals. He is Research guide to various universities in india. He is guiding PhD, students in areas of Cloud Computing, Image processing, Datamining.

Augustin Minalkar received his MCA degree from From Mysore university in the year 2002. He has 11 years of teaching experience. He is currently HoD, FoC, Botho University, Botswana. He has Attended 3 conferences. His area of interest is Data Mining, Data warehouse, Big Data.