A Cloud Computing Based Platform for Geographically Distributed Health Data Mining

by

Yunyong Guo
BSc, Nankai University, 2000

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in the School of Health Information Science

Supervisory Committee

A Cloud Computing Based Platform for Geographically Distributed Health Data Mining

by

Yunyong Guo

BSc, Nankai University, 2000

Supervisory Committee

Dr. Alex M.H. Kuo, (School of Health Information Science)
Supervisor

Dr. Andre Kushniruk, (School of Health Information Science)
Committee member

Abstract

Supervisory Committee
Dr. Alex M.H. Kuo, (School of Health Information Science)
Supervisor
Dr. Andre Kushniruk, (School of Health Information Science)
Committee member

With cloud computing emerging in recent years, more and more interest has been sparked from a variety of institutions, organizations and individual users, as they intend to take advantage of web applications to share a huge amount of public and private data and information in a more affordable way and using a reliable IT architecture. In the area of healthcare, medical and health information systems based on cloud computing are desired, in order to realize the sharing of medical data and health information, coordination of clinical service, along with effective and cost-contained clinical information system infrastructure via the implementation of a distributed and highly-integrated platform. The objective of this study is to discuss the challenges of adopting cloud computing for collaborative health research information management and provide recommendations to deal with corresponding challenges. More specially, the study will propose a cloud computing based platform according to recommendations. The platform can be used to bring together health informatics researchers from the different geographical locations to share medical data for research purposes, for instance, data mining used for improving liver cancer early detection and treatment. Finding from a literature review will be discussed to highlight challenges of applying cloud computing in a wide range of areas, and recommendations will be paired with each challenge. A proof of concept prototype research methodology will be employed to illustrate the proposed

cross national cloud computing model for geographically distributed health data mining applied to a health informatics research.

Table of Contents

List of Tables

List of Figures

Acknowledgements

This thesis would not have been possible without the help, support and patience of my principal supervisor, Prof. Alex Kuo, not to mention his advice and unsurpassed knowledge of cloud computing and health informatics. The good advice, support and friendship of my committee member, Prof. Andre Kushniruk, have been invaluable on both an academic and a personal level, for which I am extremely grateful.

I would like to acknowledge the financial, academic and technical support of the University of Victoria and its staff.

Dedications

I would like to thank my wife, Man Wang, and my two little sons, Bryan Guo and Nathan Guo. They are always there cheering me up and stood by me through the good times and bad.

Chapter 1 **Introduction**

This chapter highlights the importance of cloud computing adopted to the field of health care by describing a few examples of application of cloud computing in healthcare. Additionally, this chapter provides the significance of the present thesis and research objectives for the study.

1.1. Cloud Computing Adopted in Healthcare

With rapid healthcare and economic development, more and more medical records are generated. The motivation to improving the level of modern records management by using innovative technology has dramatically increased. Information technology offers the potential to address healthcare's three primary challenges: rising costs, uneven quality and inadequate access. One of most popular and promising information technologies is cloud computing (Luis et al., 2008; Buyya et al, 2008). It is defined as an on-demand, self-service network architecture in which users are able to access computing resources and share information anytime from anywhere (Mell and Grance, 2010). Cloud computing systems provide many benefits to facilitate medical information resource sharing. Within cloud computing, users or organizations gain the right to access medical records online, to engage their providers via digital channels, and to share their records across their teams of providers. (Catteddu and Hogben, 2009; Chow et al., 2009; Jeffrey and Neidecker-Lutz, 2009). In addition, cloud computing reduces barriers to regulatory approval and licensing. Therefore, cloud computing accelerates the rapid sharing of clinical protocols, best practices and outcomes data without location restriction as best practice, standardized care procedures that can be supported.

It is well known that the costs of healthcare services constantly rise in every country and healthcare providers are more demanding. Therefore, adopting advanced health information technologies to reduce healthcare cost and improve quality is imminent (Saranummi, 2008, Saranummi, 2009, Saranummi, 2011, Vasilakos and Lisetti, 2010). Cloud computing is one of the most prominent technological trends as it offers an applicable platform for health information technology services over the Internet (Shimrat, 2013). Cloud computing represents a "fourth space" beyond those healthcare has traditionally delivered: hospitals, clinics and homes (Haughton, 2011, Teng et al, 2010). Since health informatics seek new ways of driving healthcare information sharing forward, for example, international health information research collaboration, growing demands are now placed on computer networks to provide hardware and software resources and pave a new avenue to share sensitive and private medical data from different geographic locations. Cloud computing demonstrates tremendous opportunities for the collaborative healthcare information sharing (Cloud Computing 2013). Users or the organization do not need to care about over-provisioning for a service whose popularity does not meet their predictions, thus wasting costly resources, or under-provisioning for one that becomes widely popular, thus missing potential customers and revenue. Nevertheless, cloud computing has also introduced a set of new and unfamiliar challenges (Andrei, 2009; Buyya *et al*., 2008; Catteddu and Hogben, 2009), such as lack of interoperability, standardization, privacy, network security and culture resistance.

In order to overcome obstacles of adopting the cloud computing service, there are many research efforts that contribute to build and examine cloud computing for the healthcare purposes. One example of cloud based system was designed for storage and file management system for healthcare. Guo et al. (2010) designed a cloud-based intelligent hospital file management system (HFMS) that removes some of the restrictions that existed in current hospital management systems. The restrictions consist of limited storage capacity due to inefficient hardware devices and low performance of information technology (IT) systems caused by the ocean of clinical data. The proposed cloud computing system for hospital file management systems includes a central server and many satellite servers. The central server controls the file system meta-data that consists of namespace, access control, file-block mapping and physical address of relevant information. The benefits of the system is to lower the cost of server clusters but increase the flexibility as the physical boundaries are minimized and the utilization of system resource is maximized. Chen et al. (2010) recommended a cloud computing based system to store clinical data in order to meet the growing need of storage space for EMR, meanwhile, the suggested cloud-based system satisfied robust data security and information privacy protection requirements. In the model, the EMR data can be stored in the local storage system and other two different commercial clouds using the algorithm of RAID 3 (redundant array of inexpensive/independent disks), so that the data stored in each cloud based server lose the meaning and use. A cryptographic method combined with RAID-3 is proposed to be applied to ensure the integrity of data upload and download. In addition, Teng et al. (2010) described a long term off-site medical image archive solution for medical digital imaging and communication (DICOM). Managing long-term onsite medical imaging archives is a big challenge for cost containing in healthcare area. The growing need for a high volume of medical images leads to the issue of scalability and maintenance in picture archiving and communication systems (PACS).

The Windows Azure Cloud platform was applied to implement the prototype of DICOM image archive service. The prototype system was examined with a wide range of public domain DICOM images. The examined image series were successfully sent from clients, received and indexed by the server in the cloud, retrieved as requested in queries and returned. With the help of Azure's functionality and features, the system has strong capability to decrease the cost of image archives storage and management budget, as well as improve the disaster recovery ability. Rolim et al. (2010) demonstrate a cloud-based model used to automatically collect, distribute and process patients' data. Obviously, the proposed system can significantly decrease the manual involvement, eliminate typing errors, and improve clinical data accessibility. The system has a network of sensors connected to legacy medical devices to collect patients' vital data and deliver it to the cloud server for storage, management and distribution. Recently, Lo et al. (2011) come up with a cloud-based Early Warning Service (EWS) that enable the simulation of patients' data. More interestingly, the system allows to automatically process and calculate the patients' risk index by capturing vital signs using the medical sensors, transmitting the received values to data storage room in the cloud, as well as monitoring the patient's status, and notifying doctors and nurses by calling or messaging their mobile phones as necessary.

This study will identify the challenges of applying the healthcare cloud in collaborative health information research and discuss potential approaches to conquer those barriers, such as organizational change, security, legal, regulatory and compliance.

1.2. Significance and Purpose of the Study

The primary goal of this research is to illustrate principle barriers in adopting cloud computing in healthcare organization, both technical and nontechnical aspects. In order to reconcile those challenges, some recommendations will be explored when applying or migrating cloud computing across different boundaries and countries. Furthermore, based on the review and discussion for the challenges and solutions of adopting cloud computing, the ultimate goal of the study is to use the proof-of-concept prototype methodology to illustrate how a proposed cloud computing platform is adopted to liver cancer data mining collaborative research for geographically distributed sites. It is well known that each country has different geographical variations, such as climate, race, living habits, diets, culture, social patterns etc., which lead to different cancer risks and medical treatments. Regarding the patient's demographic and geographic variance, the liver cancer patient data from the single country is no longer able to provide a comprehensive understanding and background to undertake the study of the cause of liver cancer. Some key information is most likely to be hidden by lack of patient data correlation between different continents. More and more researchers have realized the utilization and benefits of data mining techniques for studying the relationship between the cause of liver cancer and patient's demographic characters. Nevertheless, the current studies and research indicate the limited extent to which health data and information sharing across various countries is used to support collaborative liver cancer early detection research, in that the collaborative liver cancer research is still facing tremendous issues, such as data sharing, study communication and IT cost and maintenance. Therefore, as a promising concept and technique, our proposed cloud computing platform is designed to overcome barriers and may provide benefits to future liver cancer studies in data collection and sharing in geographically distributed locations.

1.3. Research Objectives

The study proposes a cloud computing based data mining platform to share different levels of clinical data and to extract knowledge from a huge number of raw data in different locations. Essentially, the platform can be applied to different hospitals, organizations as well as countries. More importantly, to our knowledge, there are not many existing models that have been specifically focused on health informatics studies. Given the research discussion, not only does the study propose a robus

t information system to contribute the health data sharing in each region, but also the study has potential to promote international health informatics research collaboration.

The objectives of the study are:

- Review the challenges associated with adopting cloud computing in healthcare management and health research
- Explore solutions to conquer the barriers of implementing cloud computing for health data sharing among different geographic jurisdictions.
- Design a proof-of-concept cloud computing architecture for geographically distributed health data mining research.

Chapter 2 **Research Background**

This chapter elaborates research questions, methods and limitations to the questions as well as the research process.

2.1. Statement of the Study Problems

According to the World Health Organization (WHO), cancer is the leading cause of death worldwide **(**7.6 million deaths in 2008**)**, and it is projected to continue rising, with an estimated 12.7 million deaths in 2030. One major contributor to this condition is liver cancer. Liver cancer has high prevalence in different countries. Liver cancer leaded to overall cancer mortality which was 696,000 deaths (9.2%) per year (Ferlay, et al. 2010). In Canada, there was an estimated 173,800 new cases of cancer according to the Canadian Cancer Statistics 2010 released by the Canadian Cancer Society. A recent report (Canadian Institute for Health Information, 2011) called "Learning From the Best: Benchmarking Canada's Health System", which looks at the latest statistics and indicators comparing health systems on quality of care and access to care, shows that cancer deaths remains relatively high in Canada, particularly for cancers that are hard to screen for and treat early, such as liver cancer. The report for Canadian males indicated liver cancer increased 2.2% compared with 2009 (Canadian Cancer Society, 2010). Economic burdens of liver cancer treatment are high and rising quickly. For example, treatment with the ribavirin/interferon alfa-2b combination can cost up to $30,000 per course of treatment for an infected person.

On the other hand, Canadian healthcare continues to face increased pressure to contain costs while maintaining or increasing quality healthcare service. Canadian federal and provincial governments, as well as a number of healthcare organizations had invested a great deal of manpower and financial resources to conduct the study of standardized clinical pathways, design clinical guidelines and interpret diagnosis data in order to improve liver cancer treatment. As health informatics plays a critical role in every aspect of the healthcare area, especially in the current information technology era, it is vital to collect data from different data sources, maintain the data, produce information, discover knowledge and disseminate data, information and knowledge to various stakeholders.

Therefore, to accomplish the above tasks, more and more health informatics researcher realized that the advent of cloud computing and its business model have been become some of the biggest changes impacting not only the computer industry but also collaborative health research. Several health research innovations have demonstrated that cloud computing has the potential to overcome collaborative health research data sharing management and IT cost containment. However, researchers agree that academic research on challenges and recommendations for applying cloud computing and in particular the adoption in collaborative health research, as well as healthcare industry, still needs to be significantly expanded in all aspects, even though some work has been performed on the security and adoption strategies for cloud computing in other business areas. Challenges and opportunities are constantly a hot topic that is receiving increasing focus as implementation of cloud computing in the healthcare area is demanded in order to replace legacy systems. So far, most industry publications emphasis is on the financial benefits of adopting cloud computing and the cost-effectiveness of migrating to cloud computing. There is little published literature on the solutions to potential barriers of applying cloud computing in healthcare, especially, in the collaborative health research that cloud computing will be applied in.

The study will address the following four broad questions:

1. What are the critical challenges associated with adopting cloud computing?
2. What are the possible solutions for users to successfully implement cloud computing infrastructure by addressing the challenges?
3. How do we form a cloud computing based data mining platform for sharing research resources/results in international collaborative health related research?
4. How can a cloud computing platform be used for liver cancer early detection study across different geographical-distributed locations?

2.2. Research Method

This study will utilize extensive secondary research on standard documents, industry periodicals, analysis reports, conference journals and published academic papers to investigate the research questions outline above. In order to summarize the current state of knowledge in the area to serve as a background for the study, a literature review of the development and implementation of various cloud computing and data mining technologies in the healthcare area are carried out. In addition, a proof of concept prototype study methodology is applied in the formulation of the framework and model proposed, which involve shared patients' clinical information from three different sites.

In this study, we adopt a proof of concept prototype research methodology. It is often employed in clinical research studies (Fardon, et al. 2007, Lawrence, 2005) and is defined as following (Wikipedia, 2013):

"A proof of concept prototype is used to test some aspect of the intended design without attempting to exactly simulate the visual appearance, choice of materials or intended manufacturing process. Such prototypes can be used to "prove" out a potential design approach such as range of motion, mechanics, sensors, architecture, etc. These types of models are often used to identify which design options will not work, or where further development and testing is necessary."

A proof of concept prototype is generally applied early in the system development cycle. It is used to validate technical feasibility, helps identify potential stumbling blocks, identifies what a platform can or can't provide, and helps determine the scope and level of customization necessary to complete the project. It can also help identify performance issues. Here, we assemble many of our applications / solutions in a "composite" fashion. We are re-using services, functions, etc. from other applications. This re-use requires integration points. It is these integration points in our overall "context" that we are vetting with the prototype effort. Using the proof of concept prototype methodology, we are able to examine some cloud computing based platform implementation success factors along with factors that impact overall scope and estimates of effort for the proposed platform.

2.3. Research Limitation

Although the research has achieved its aim to discuss and explore opportunities for the adoption of cloud computing in the health collaborative research, there were some unavoidable limitations. Firstly, the concept of cloud computing based data mining platform in this study is relatively new to many healthcare organizations. Most of them are focused on security, social-technical impact, business models. There is little literature related to our current clinical study for the comprehensive investigation. Secondly, due to time limits, this research was conducted only on a specific health care research topic (liver cancer early detection research), so, to generalize the results for other healthcare research interests, the study should have involved more studies at different levels. It is clear that distinctive disease detection studies have different characters and requirements for various aspects. Finally, other than the challenges and opportunities we have discussed in the thesis, regarding the necessities more fluid design specifications and challenges to our traditional thinking about jurisdiction related to data protection, there are still a great many open issues and potential opportunities in the adoption of cloud computing in the healthcare industry which need to be resolved. A more complete survey on this topic will be expected with the evolution of the cloud computing concept in the near future.

2.4. Research Process

Cloud computing is an emerging concept and technology for delivering computing resource and service. However, like any innovations, cloud computing has also faced challenges to the organization seeking to adopt it. Therefore, in the research, we firstly review the challenges which it raises, such as trust, security, legal, compliance and organizational challenges. Then, we will present the potential coupled solutions to tackle those challenges, thus facilitating the adoption of cloud computing in different organizations, in particular, for our international collaborative research.

For the last and most important part, the rest of thesis will be organized to propose a model of cloud computing based data mining platform to improve collaborative health research for liver cancer early detection with the further understanding of challenges and solutions of cloud computing. As we have indicated in the previous chapter, the failure to consider demographic differences among liver cancer patient's characters had led to overlooking certain root causes and correlations for liver cancer, thus limiting early detection options. Therefore, a collaborative research model will give us an opportunity to share clinical data and extract knowledge from different geographic distributed sites. For the collaborative model, we assume that researchers from many different countries work with partners who share medical data through a cloud computing architecture, then use data mining algorithms to analyze liver cancer data which are extracted from their own Electronic Health Record (EHR) systems. However, how to share the information across boundaries economically and efficiently in collaborative research is turning out to be a big challenge to accomplish the goal. Cloud computing as a latest technological trend provides a strong infrastructure and offers a true enabler for health information technology services over the internet. By doing so, the collaborative research can be attained on a pay-as-you-use cloud service model to help the health information researchers cope with current and future demands yet keeping their cost to a minimum (Cloud Computing ,2013, Shen et al., 2011). Herein, in order to design a cloud computing based platform for collaborative liver cancer early detection research, we expect to complete the following tasks:

(1) To review and discuss the challenges and solutions for adopting cloud computing in various organizations

We conduct a comprehensive review of recent publications related to cloud computing, thus summarizing barriers of adopting a cloud computing across boundaries. Then, we provide possible coupled solutions to deal with those potential challenges, thereby building up a solid base for our proposed cloud computing based data mining platform.

(2) To propose a cloud computing based data mining platform for sharing research resources/results

In this proposed model, we plan to design a cloud based data mining platform that allows researchers in three different locations easy sharing of research resources/results through the internet. In the platform, the cloud computing based data mining architecture will employ the Software as a Service (SaaS) model to run the cloud service and share the data through a data centre server.

(3) To illustrate how the proposed platform could be applied in collaborative liver cancer early detection research

Using the proposed cloud based architecture and sharing medical data from three different resources, we will be able to infer the relationship from a large number of medical records using the combination of distributed data mining algorithms and association algorithms (e.g. the Apriori algorithm). This analysis produces association rules that indicate what combinations of demographics, geographic locations and patient characteristics lead to liver cancer that can help health providers to provide early alerts to patients with high liver cancer risk.

Chapter 3 **Literature Review**

This Chapter presents research background information and the context of the present thesis, such as the definition of cloud computing, data mining, approaches and activities, along with cloud computing deployment models. In addition, this chapter also reviews the application of data mining in cancer study.

3.1. Definition of Cloud Computing

Cloud computing is a new model and concept in computing science. It has been defined as follows (Vaquero *et al*., 2008):

"Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically re-configured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customizedService-Level Agreements."

Mell and Grance (2010) give a definition of cloud computing that is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service-provider interaction. We have already seen similar more limited applications for years, such as Google Docs or Gmail. Nevertheless, cloud computing is different from traditional systems. Figure 1 shows the infrastructure of the NIST concept of cloud computing.

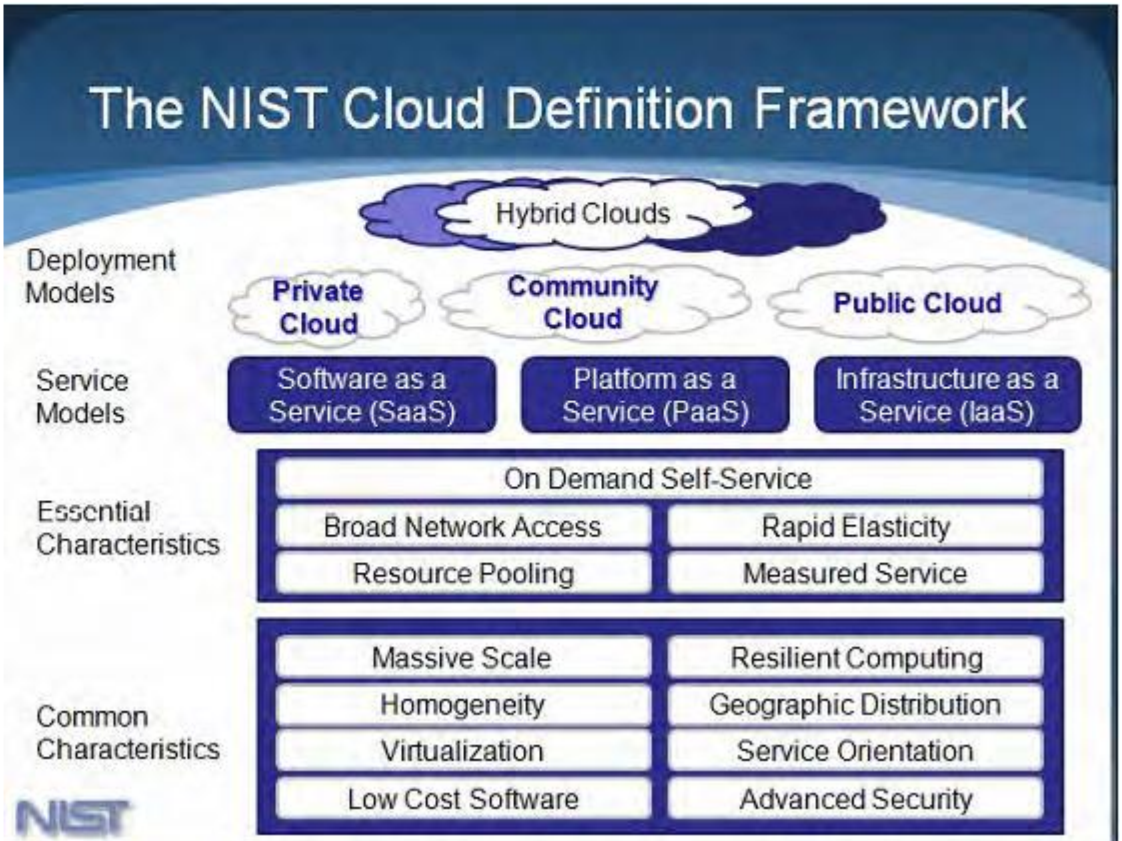

Figure 1 Cloud Computing Definition (Grace,2010.)

Armbrust et al. (2010) state that cloud computing offers a wide range of computing sources on demand anywhere and anytime; eliminates an up-front commitment by cloud users; allows users to pay for use of computing resources on a short-term basis as needed and has higher utilization by multiplexing of workloads from various organizations. Cloud computing includes three models: (1) Software as a Service (SaaS): the applications (e.g. EHRs) are hosted by a cloud service provider and made available to customers over a network, typically the Internet. (2) Platform as a Service (PaaS): the development tools (such as OS system) are hosted in the cloud and accessed through a browser (e.g. Microsoft Azure). (3) Infrastructure as a Service (IaaS): the cloud user outsources the equipment used to support operations, including storage, hardware, servers and networking components. The cloud service provider owns the equipment and is responsible for housing, running and maintaining it. In the clinical environment, healthcare providers are able to remotely access the corporate Intranet via a local Internet service provider, since they have the option to have an ISDN line installed to their home or hospital linking with Cloud Computing system, as figure 2 shown (Guo et al., 2010).

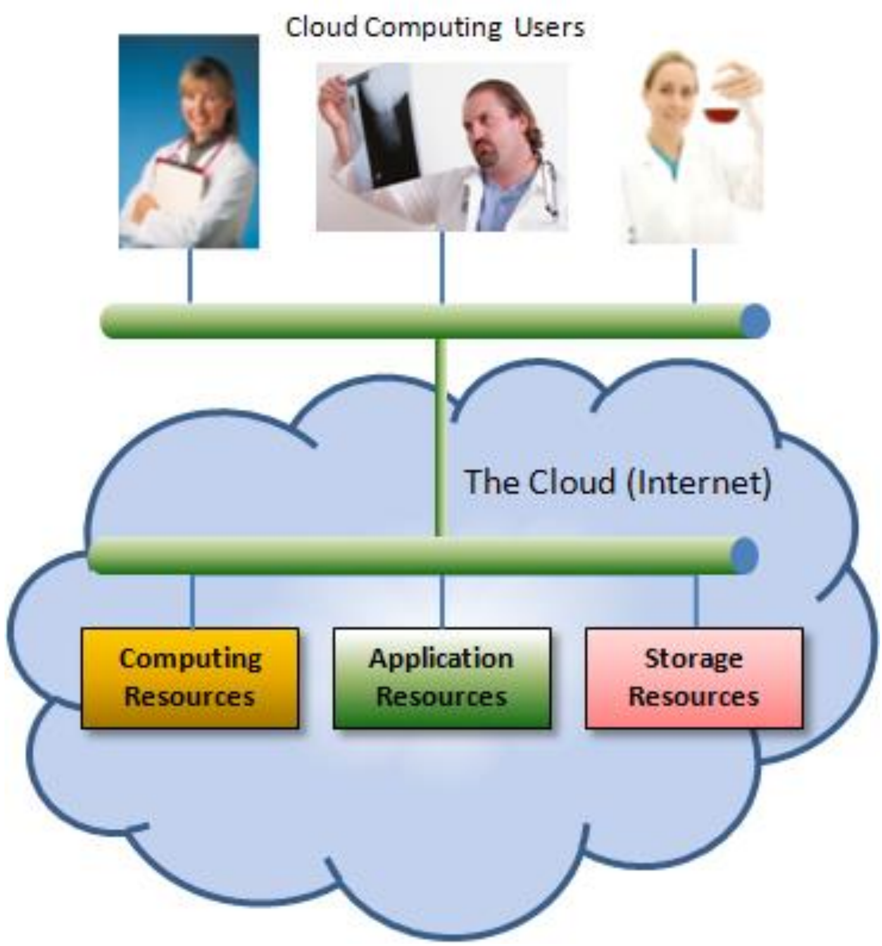

Figure 2: The Cloud Computing Platform

A Cloud Computing service stack model is introduced by Guo et al. The bottom two layers is a virtualization of resources in the form of storage and computing which is the foundation of cloud services. Virtual resource layer services lie on top of a cloud layer. It is an external application programming interface which provides the internal mechanism. Cloud service is not a separate service, but rather a collection of services. The model is shown in the figure 3 (Guo *et al*., 2010).

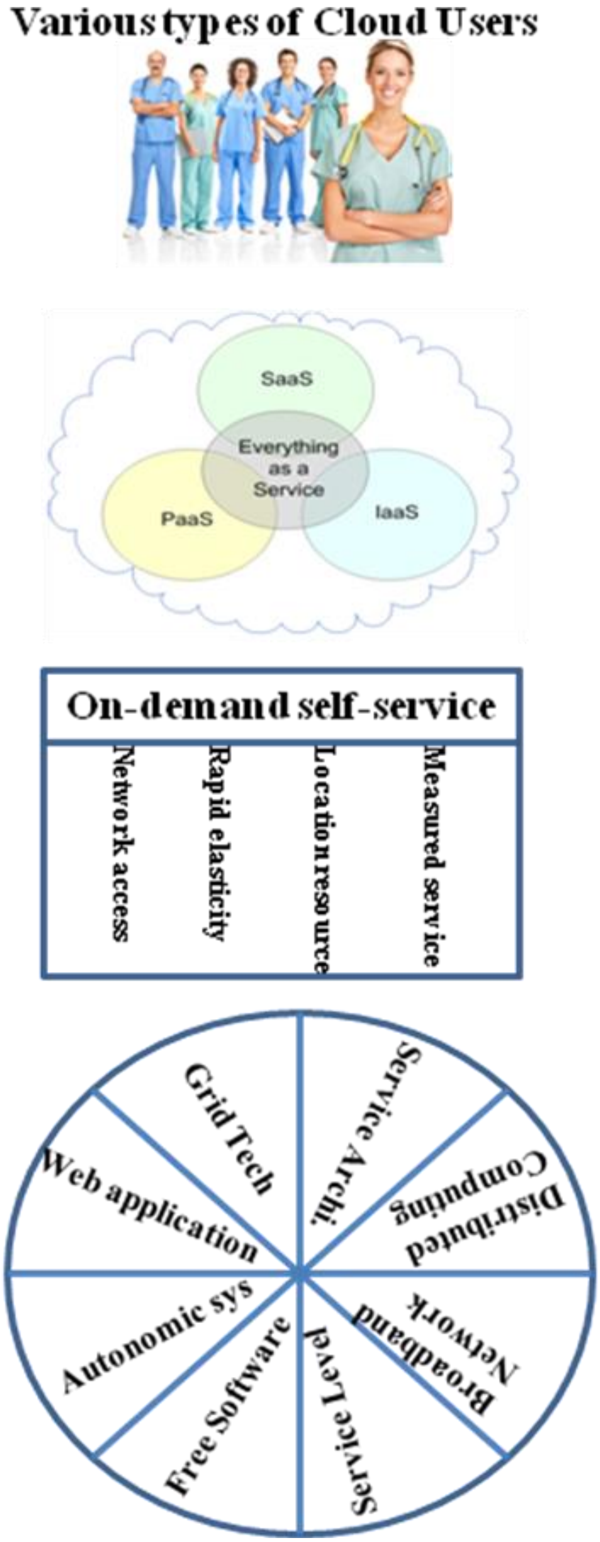

Figure 3: The Cloud Computing Service Stack Model

In table 1, a summarized definition for cloud computing from various experts is provided by Luis *et al*. (2009).

Table 1: Definition of Cloud Computing Adapted from Luis *et al*. (2009)

Author/Reference	Definition
M. Klems (Geelan, 2009)	*"you can scale your infrastructure on demand within minutes or even seconds, instead of days or weeks, thereby avoiding under-utilisation (idle servers) and over utilisation (blue screen)of in-house resources".*
P. Gaw (Geelan, 2009)	"*refers to the bigger picture...basically the broad concept of using the internet to allow people to access technology enabled services".*
R. Buyya (Buyya et al., 2008)	"*a type of parallel and distributed system consisting of collection of interconnected and virtualised computers that are dynamically provisioned and present as on or more unified computing resource based on service-level agreements established through negotiation between service provider and customer".*
R. Cohen (Geelan, 2009)	"*for me the simplest explanation for cloud computing is describing it as, 'internet centric software'. This new cloud computing software model is a shift from traditional single tenant approach to software development to that of scalable, multi-tenant, multiplatform, multi-network, and global".*
J. Kaplan (Geelan, 2009)	" *a broad array of web-based services aimed at allowing users to obtain a wide range of functional capabilities on a 'pay-as-you-go' basis that previously required tremendous hardware/software investment and professional skills to acquire".*
D. Gourlay (Geelan, 2009)	"*cloud will be the next transformation over the next several years, building off of the software models that virtualisation enabled"*
D. Edwards (Geelan, 2009)	"*...what is possible when you leverage web scale infrastructure (application and physical)in an on-demand way. ...anything as a service... all terms that couldn't get it done. Call it 'cloud' and everyone goes bonkers".*
B. De Haff (Geelan, 2009)	*"...there are really only three types of services that are cloud based: SaaS, PaaS, and Cloud Computing Platforms".*

B. Keppes (Geelan, 2009)	"*put cloud computing is the infrastructural paradigm shift that enables the ascension of SaaS".*
K. Sheynkman (Geelan, 2009)	"*the 'cloud' model initially focused on making hardware layer consumable as on demand compute and storage capacity. ... to harness the power of the cloud, complete application infrastructure needs to be easily configured, deployed, dynamically scaled and managed in these virtualised hardware environments".*
O.Sultan (Geelan, 2009)	"*... in a fully implemented Data center 3.0 environment, you can decide if an app is run locally (cook at home), in someone else's data center (take-out) and you can change your mind on the fly in case you are short on data center resources (pantry is empty) or you having environmental/facilities issues (too hot to cook)".*
K.Harting (Geelan, 2009)	"*cloud computing overlaps some of the concepts of distributed, grid and utility computing, however it does have its own meaning if contextually used correctly. Cloud computing really id accessing resources and services needed to perform functions with dynamically changing needs".*
J. Pritzker (Geelan, 2009)	"*cloud tend to be priced like utilities... i think is a trend not a requirement".*
T. Doerksen (Geelan, 2009)	"*cloud computing is... the user friendly version of grid computing".*
T. von Eicken (Geelan, 2009)	"*... outsourced, pay-as-you-go, on-demand, somewhere in the internet".*
M. Sheedan (Geelan, 2009)	"*... 'cloud pyramid' to help differentiate the various cloud offerings out there... top: SaaS; middle: PaaS; bottom: IaaS".*
A. Ricadela (Geelan, 2009)	*"... cloud computing projects are more powerful and crash proof than Grid systems developed even in recent years*"
I. Wladawsky Berger (Geelan, 2009)	*"... the key thing we want to virtualise or hide from the user is complexity. ...with cloud computing our expectation is that all that software will be virtualised or hidden from us and taken care of by systems and /or professionals that are*

	somewhere else – out there in the cloud".
B. Martin (Geelan, 2009)	"*cloud computing really comes into focus only when you think about what IT always needs: a way to increase capacity or add capabilities on the fly without investing in new infrastructure, training new personnel, or licensing new software"*
R. Bragg (Bragg, 2008)	" *the key concept behind the Cloud is Web application... a more developed and reliable Cloud".*
G. Gruman and E. Knorr (2008)	"*cloud is all about: SaaS... utility computing... Web services... PaaS... Internet integration... commerce platforms...".*
P. McFedries (McFedries, 2008)	"*cloud computing, in which not just our data but even our software resides within the cloud, and we access everything not only thorugh our PCs but also cloud-friendly devices, such as smartphones, PDAs... the megacomputer enabled by virtualisation and software as a service... this is utility computing powered by massive utility datacenter".*
Gartner(Plummer et al., 2009)	"*A style of computing where scalable and elastic IT-related capabilities are provided as-a-service using Internet technologies to multiple external customers"*

In addition, cloud computing has various characteristics that distinguish it from other computing paradigms, such as massive scale availability of computing and storage capabilities, homogeneity, use of virtualisation technology, resilient computing, and pay-as-you go model. More importantly, cloud computing benefits from low or no starting IT infrastructure costs, geographical distribution of clouds and limited administration personnel. Therefore, the above characteristics attract business organizations and government agencies to apply it in the different areas (GNI, 2009, Luis et al., 2008,Vouk, 2008) .

3.2. Four Cloud Computing Deployment Models

Four cloud computing deployment models have been developed recently, in order to address different requirements and environments (Dustin Amrhein et al, 2010, CSA, 2009). These four models are public cloud, private cloud, community cloud and hybrid cloud (Dustin Amrhein et al, 2010; CSA, 2009, Grance, 2010; Mell and Grance, 2009; Catteddu and Hogben, 2009).

As the name indicates, public cloud is mainly used by general public and shared in a pay as you go model of payment. Internet is used to transfer the information between different users, as the provider is responsible for ensuring the economies of scale and the management of the shared architecture. The model is illustrated in the figure 4 (Dustin Amrhein et al, 2010).

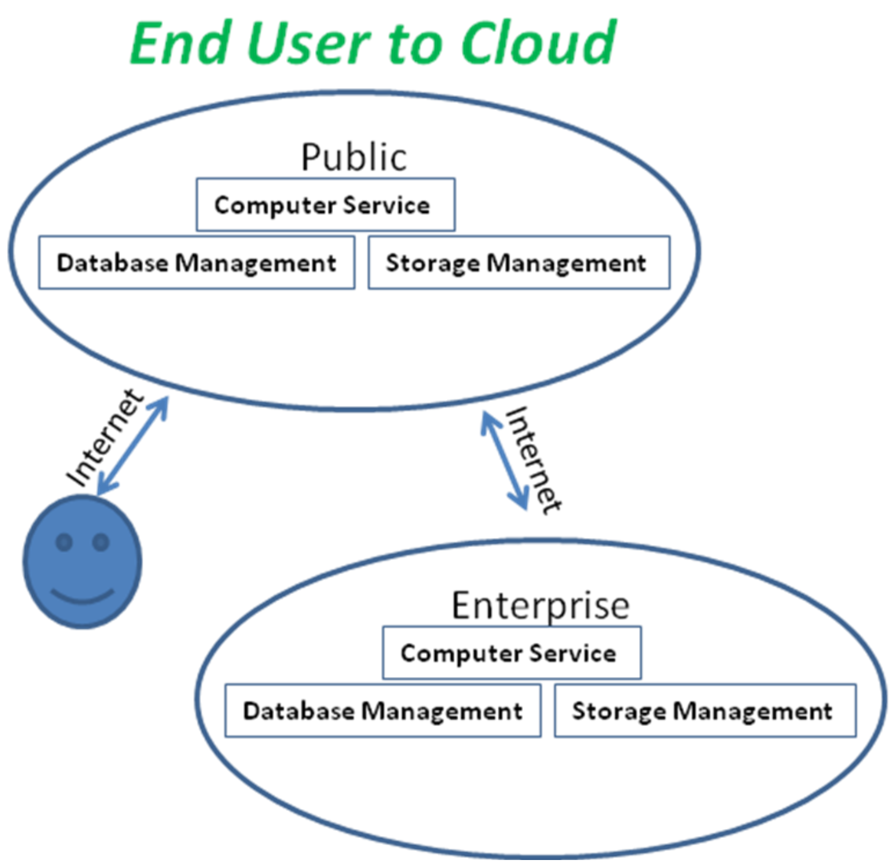

Figure 4: The Model of Public Cloud (Dustin Amrhein et al, 2010)

Private cloud is distinct from the public cloud, as it only opens up to limited users, not to any unknown third parties. The cloud resource in the model is managed by the user organization premises or offsite. This model will not significantly reduce the IT infrastructure investment as the public cloud does. Figure 5 shows the model (Dustin Amrhein et al, 2010).

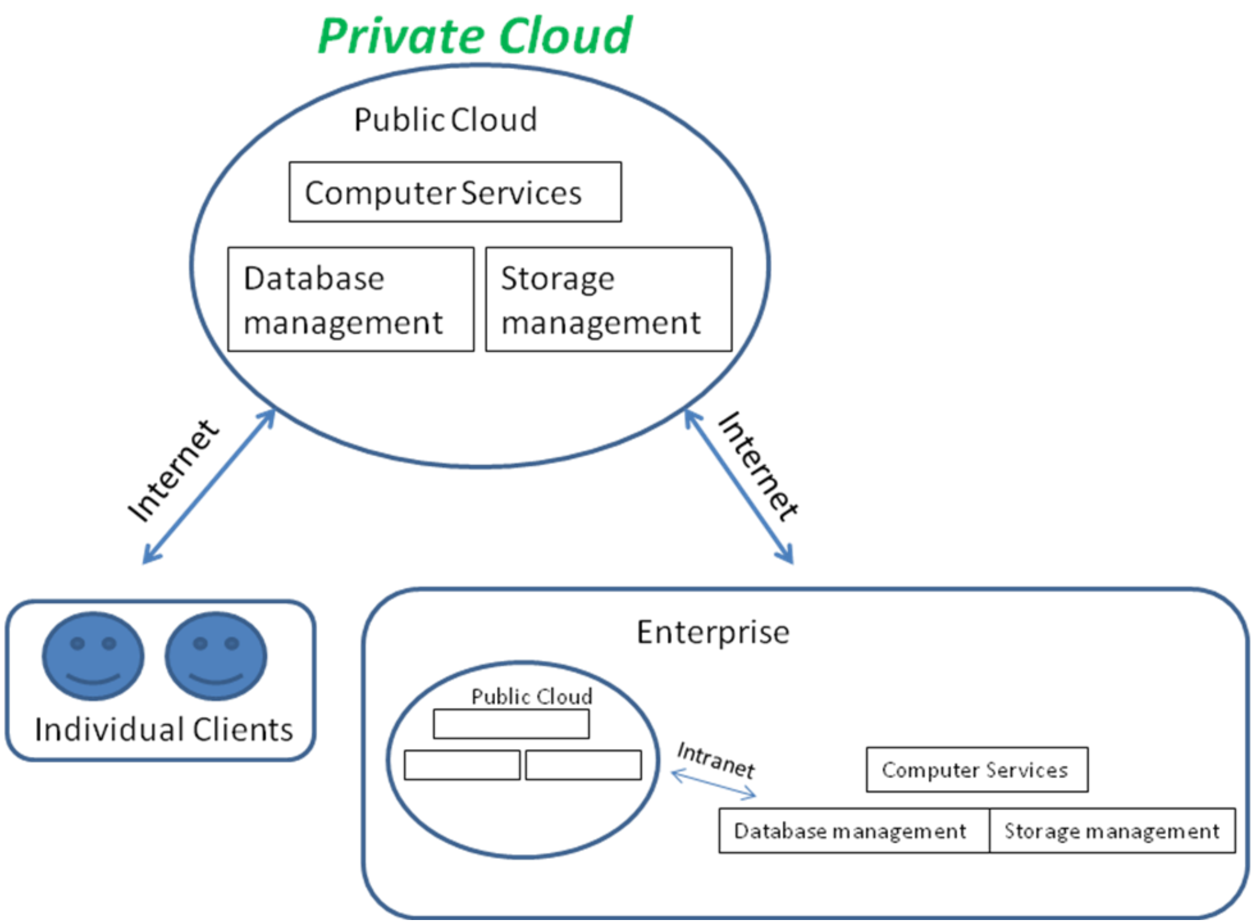

Figure 5: The Model of Private Cloud (Dustin Amrhein et al, 2010)

Hybrid cloud is model of adoption which combines different clouds, such as private and public clouds. The public and private cloud's functionalities are integrated together, shown in figure 6 (Dustin Amrhein et al, 2010):

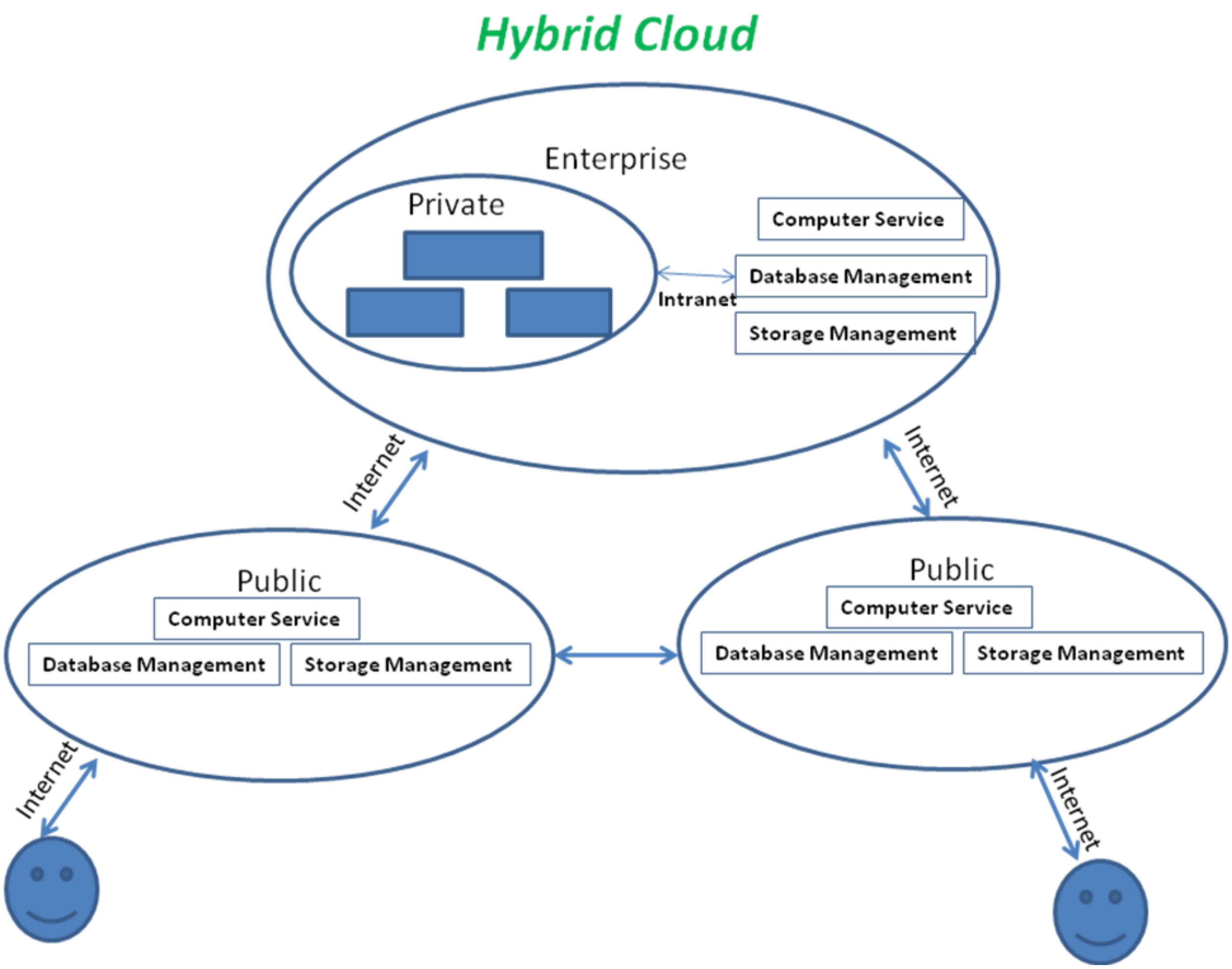

Figure 6: The Model of Hybrid Cloud (Dustin Amrhein et al, 2010)

Community cloud is the fourth model which is used by multiple organizations or institutions that have shared concerns or interest, for example, compliance considerations, privacy needs. The infrastructure may be operated by the third party. The model is depicted as the Figure 7 (Dustin Amrhein et al, 2010):

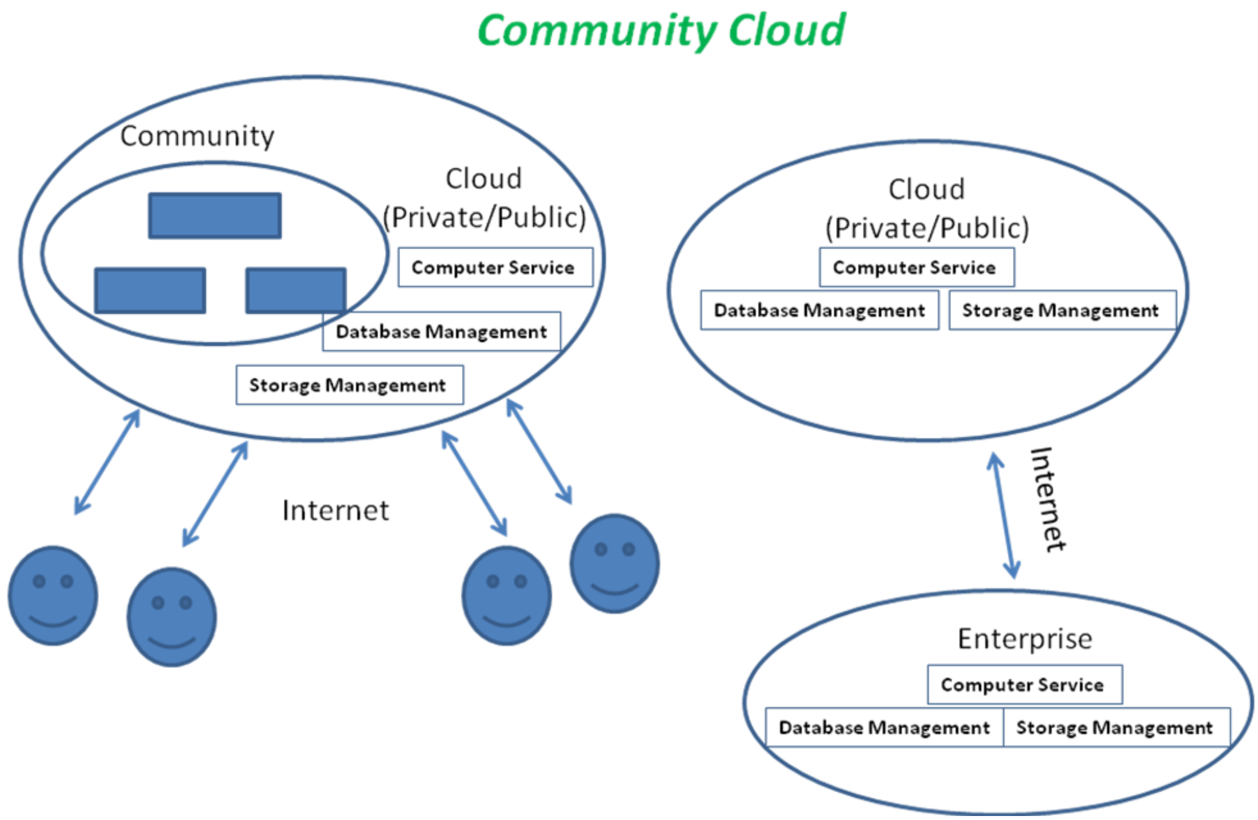

Figure 7: The Model of Community Cloud (Dustin Amrhein et al, 2010).

3.3. Application of Cloud Computing in Healthcare

Different types of organizations can benefit from cloud computing such as government agencies, financial enterprises, online entertainment companies, and healthcare providers. In this research, we focus on the healthcare industry. Currently, enhancing healthcare service quality and reducing the operational budget are the most important topics in the utilization of updated IT technologies in the healthcare area (Goldschmidt, 2005, Davidson and Heslinga, 2006, Klein, 2007). In order to achieve this goal, healthcare HIT is highly intended to move departmental solutions to encompass larger strategy at the enterprise level, and from standalone systems that offer limited and localized solutions to more integrated and interconnected ones that bring up comprehensive and effective solutions (Lenz and Reichert, 2007). Cloud computing has been deemed as a integrated solution that shifts the burden of managing and maintaining complex healthcare in-house high-cost hardware, software, and network infrastructure to the cloud, even the cloud service providers (Teng *et al*., 2010, Cloud Computing, 2013). More specifically, healthcare information systems confront the high cost of implementing and maintaining IT, fragmentations of HIT and insufficient exchange of patient data, lack of legal regulation mandating the use and protection of electronic health care data capture and communications, as well as lack of healthcare IT design and development standards (Kaletsch and Sunyaev, 2011, European Commission, 2013, U.S. Department of Health & Human Services, 2003).

Most health care systems are built on workflows that consist of paper medical records, duplicated test results, and fragmented IT systems. The majority of physicians in healthcare do not always have the information they require when they need to rapidly make patient-care decisions, and patients often have to carry a paper record of their health history information with them from visit to visit. To address the problems, IBM and Active Health Management (2010) collaborated to create a cloud computing technology-based Collaborative Care solution that gives physicians and patients access to the information they need to improve the overall quality of care, without the need to invest in new infrastructure. IBM facilitated the American Occupational Network and HyGen Pharmaceuticals to improve patient care by digitizing health records and streamlining their business operations using cloud-based software from IBM MedTrak systems, Inc. and The System House, Inc. Their technology handles various tasks as a cloud service through the internet instead of developing, purchasing and maintaining technology onsite. Rolim et al. (2010) designed a cloud computing platform used to collect patients' crucial information automatically from legacy medical systems through a network of sensors, and then transfer the data through cloud to central storage, processing, and distributing. Nkosi and Mekuria (2010) have reported that multiple medial sensor signals are processed and stored in a cloud computing protocol management system. The system significantly increases efficiency by utilizing various mobile devices for societal services and promotes health care services. Furthermore, Koufi et al. (2010) proposed a cloud computing based emergency medical system model for the Greek National Health Service embedding the emergency system with personal health record systems to offer doctors with easy and direct access to patient data from anywhere and at anytime with low cost and in any computer devices. Acumen solution's (2009) cloud computing CRM and project management system were selected by the U.S. Department of Health & Human Services' office of the National Coordinator for Health IT to manage the selection and implementation of EHR systems across the country. The software will enable regional extension centres to manage interactions with medical providers related to the selection and implementation of an EHR system. Sharp Community Medical Group in San Diego will be using the collaborative Care solution to change the way physicians and nurses access information throughout the hospital group's multiple electronic medical record system to apply advanced analytical and clinical decision support to help give doctors better insight and work more closely with patient care teams. Another similar example of applying cloud service in the healthcare area is the architecture of the hospital file management system (HFMS). A HFMS cluster contains a master server and multiple blocks of servers by multiple client access. HFMS application software can achieve optimal performance and availability, which is shown in figure 8 (Guo *et al*., 2010).

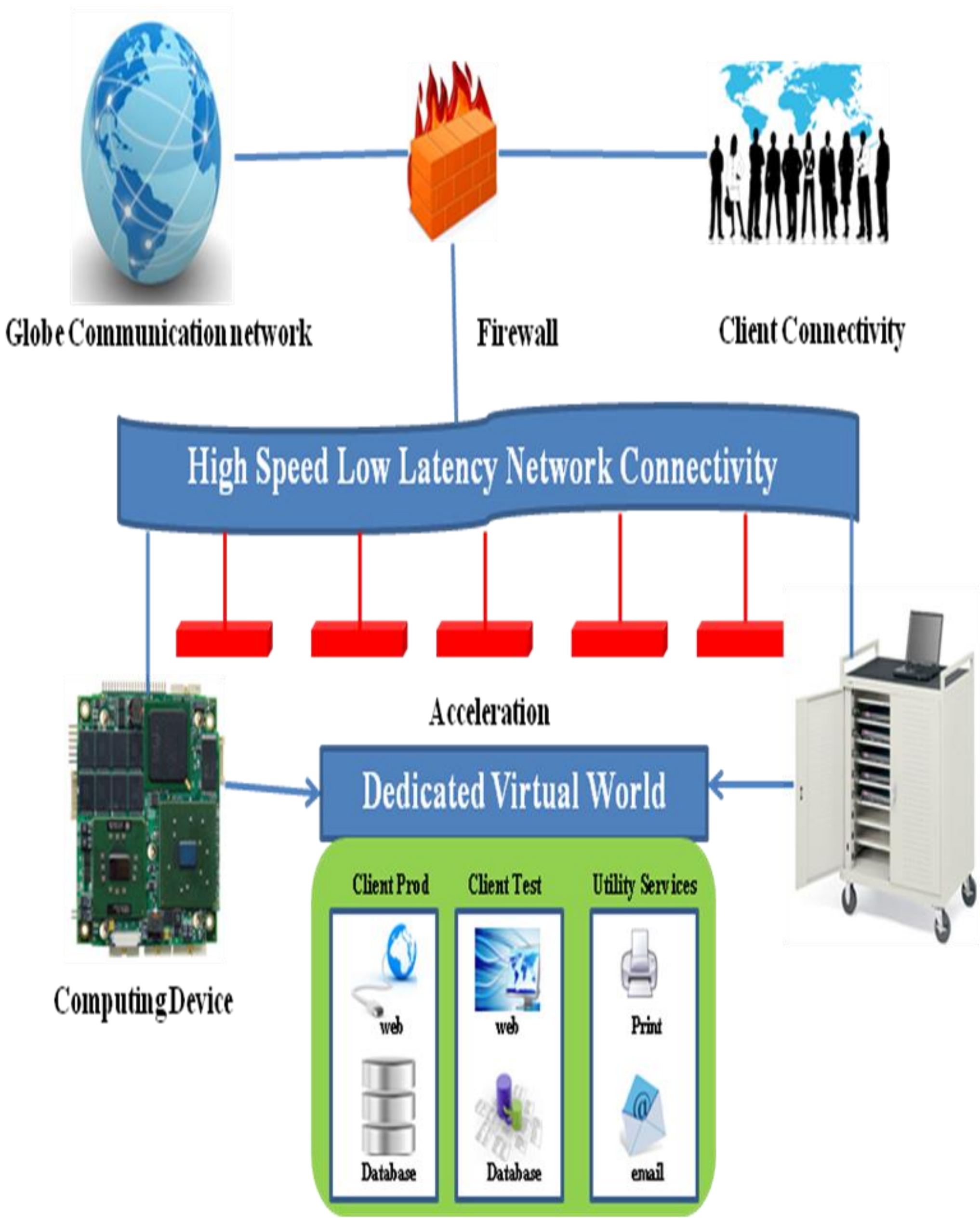

Figure 8: The Architecture of HFMS

One of eHealth cloud models is presented in the figure 9. It is a special-focused cloud computing targeted on healthcare area, which offers IT services to improve patient care while decreasing the operational cost and increasing system efficiency (Abukhousa et al., 2012).

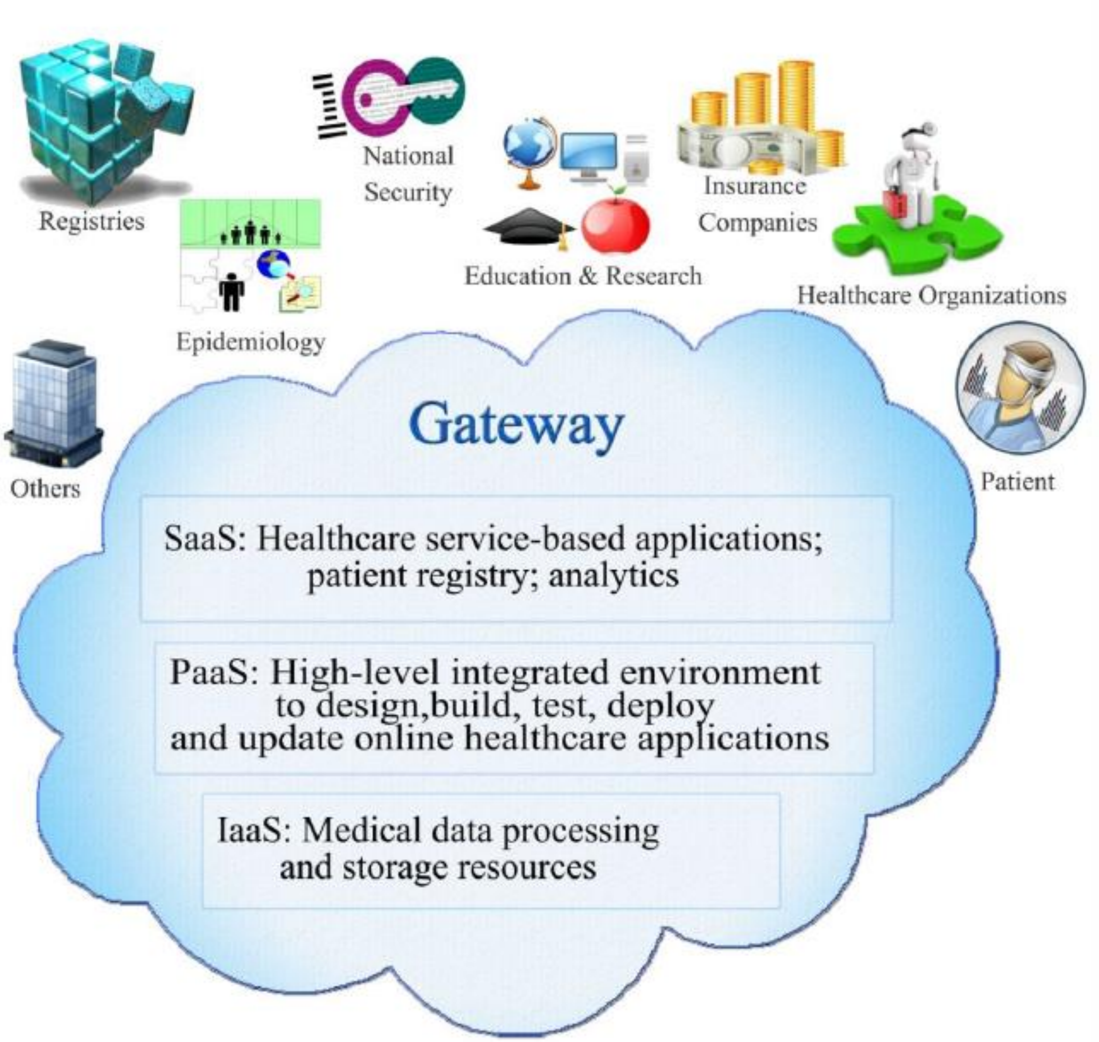

Figure 9: ehealth cloud model (Abukhousa et al., 2012)

3.4. Definition of Data Mining

Data mining is a critical step from knowledge discovery in database management processes, which refers to the “nontrivial process of identifying valid, novel, potentially useful and ultimately understandable pattern in data” (Fayyad *et al*., 1996). The term pattern here is defined as some abstract representation of a subset data of the data, that is, an expression in some language describing a data subset or a data subset or a model applicable to that subset. In order to perform descriptive and predictive analysis, data mining employs various analysis methods, including clustering, classification, regression and association analysis, to discover interesting patterns in the given data set that serve as the basis for estimating future trends. The data itself can be simple numerical figures and text documents and spatial data, multimedia data. Data mining is the extraction of hidden predictive information from large databases, thereby helping organizations focus on the most important information in their data repositories (Chapman, Clinton, Kerber, et al, 2005, Dunham, Sridhar, 2006, Larose, 2005). As well, data mining assists users and organizations to make proactive knowledge-driven decisions by forecasting future trends and characters (Larose, 2005). It provides automated prospective analyses which are far better than the analyses of past events offered by retrospective tools typical of decision support systems. More importantly, data mining is able to generate solutions for questions that traditionally were too time consuming to answer, since it can be used to find hidden patterns, and predict information that professionals may ignore because the data lies outside their expectations (DEshpande, Thakare, 2010).

3.5. Data Mining General Models

Predictive and Descriptive models are two fundamental models of the data mining system (Jensen, 2011, Larose, 2005, Tan, Steinbach, Vipin, 2009).

By using approaches of classification, regression, time series analysis, the predictive model permits one to anticipate unknown data values depending on the known values.

On the other hand, the descriptive model identifies the patterns or correlation in data and explores the properties of the data through the methods of clustering, summarization, association rule, sequence discovery, and so on.

Both data mining models with corresponding data mining algorithms are shown in Figure 10:

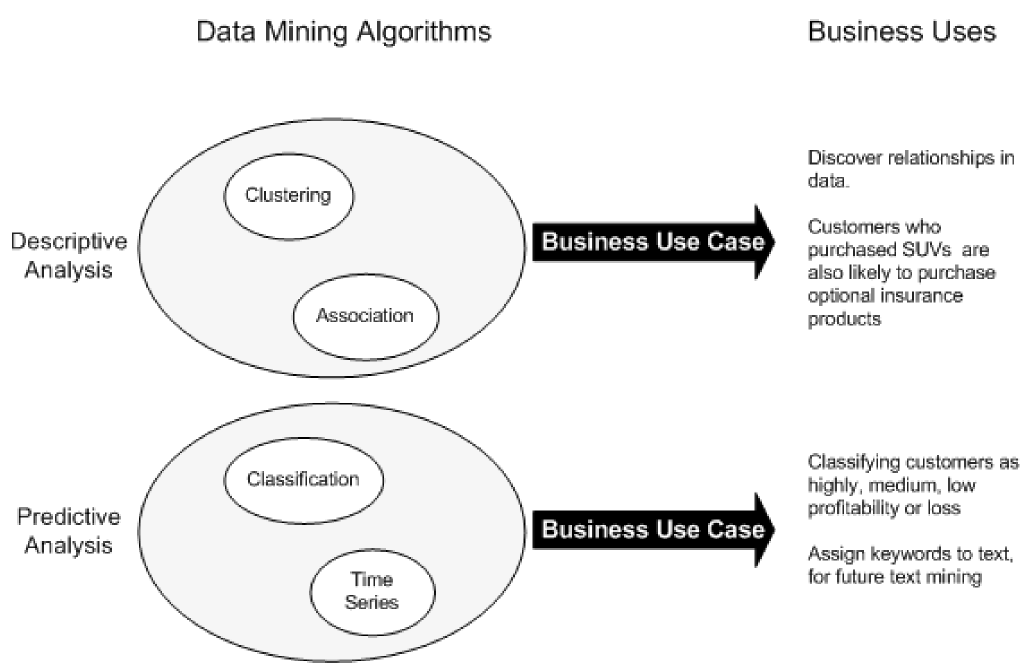

Figure 10: Data mining models with corresponding data mining algorithms (DWreview.com , 2007).

Each algorithms applied into these two models are overviewed below (DEshpande, Thakare, 2010):

In terms of the predictive model, prediction is the process of analyzing the current and past states of the attribute and prediction of its future state. Classification is a method of mapping the target data to the predefined groups or classes. Regression includes the learning of functions that maps data item to real valued prediction variables. As far as the time series analysis, the time increments are used to determine the similarity between different time periods, and the connective line is examined to decide its behavior and the historical time series plot is used to forecast future values of the variable.

Regarding the descriptive model, clustering is similar to classification except that the groups are not predefined, but are defined by the data alone. It is the partitioning or segmentation of the data into clusters. The clusters are defined by studying the attribution of the data by the domain experts. Summarization is the method of presenting the summarized information from the data. The association analysis was first introduced and formulated in 1993 (Agrawal, Imielinski and Swami, 1993). It is defined as the "market-basket problem". The problem is that we are given a set of items and a large collection of transactions which are sets of items. The task is to find relationships between the containments of various items within those baskets (Agrawal, Imielinski and Swami, 1993). Sequence discovery is a process of finding the sequence patterns in data and can be used to find trends.

3.6. General Application of Data Mining

Data mining applications can be recognized as a specific domain, as it focuses on the use of the domain specific data and data mining algorithms that aim for specific objectives. The goal of the applications studied in this context is to generate the specific knowledge within different fields. In the various domains the data generating sources generate different types of data. Data can be obtained from simple text, number figures to complicated audio-video data. Therefore, to detect the patterns and thus knowledge from this data, distinct types of data mining algorithms are applied.

Language research and language engineering take much time to extract linguistic information about a text. Data mining can be used to automatically output the huge number of linguistic features from text files in the linguistic profile (Halteren, 2004). Data mining is realized as quite effective for authorship verification and recognition. In Web-based Education (Romero, Ventura, De-Bra, 2004), data mining is applied to improve courseware. The relationships are discovered among the usage data picked up during student's sessions. The knowledge gained from data mining is very powerful for the teacher or the author of the course, who could decide what modifications will be the most suitable to improve the efficiency of the course. In crime analysis (Oatley, Ewart, 2011), data mining and decision support systems play a crucial role in assisting human inference in the forensic domain that produces one the most difficult decision-making circumstance, as an essential component of criminal investigation involves the interrogation of large database of information held by police and other criminal justice agencies. The development of data mining applied in criminal analysis is the ability to link crimes, past and present, as well as find the hidden pattern of criminal behaviours. In the manufacturing industry (Harding *et al*, 2006), data mining technologies have been widely applied in production processes, operations, fault detection, maintenance, decision support, and product quality improvement. It also involves the use of data mining in customer relationship management, information integration aspects, and standardization. Solieman (2006) has stated that data mining is a powerful tool in sports, since the huge amount of statistics are collected for each player, team, game, and season in the sport world. It is able to conduct the study of statistical analysis, pattern discovery, as well as outcome prediction. Meanwhile, patterns in the data are often viable in the prediction of future events. Data mining assists software maintenance engineers to understand the structure of software systems and assess their maintainability (Kanellopoulos *et al*., 2007). According to their similarity, the clustering algorithm effectively used to produce overviews of systems by creating mutually exclusive groups of classes, member data or methods. Bankruptcy is the major threat to the banking sector (Foster, Stine, 2004), as it can increase the cost of lending. The data mining algorithms can be used for forecasting of personal bankruptcy. Predicting bankruptcy has become the province of computer science rather than statistics. The data mining methods least squares regression; neural nets and decision trees are proved to be the appropriate for prediction of bankruptcy. E-commence is a most promising domain for data mining, as well (Ansari, 2001). It is prospective because many of the ingredients required for successful data mining are likely to be available, such as plentiful data records, reliable data provided by electronic collection, hence, insight can be turned to action and return on investment can be measured. The association of e-commence and data mining significantly enhances the results and guides the users in generating knowledge and making correct business decisions. Retailers have been gathering large amount of data like customer information and related transactions, product information, and so on. This significantly improves the application such as product demand forecasting, assortment optimization, product recommendation and assortment comparison across retailers and manufacturers (Ghani, *et al.*, 2008). Data mining is used in this context, as it can improve the work efficiency and accurately.

3.7. Application of Data Mining in the Cancer Studies

In medical science, there is broad spectrum for application of data mining, for instance, diagnosis of disease, patient profiling and history generation, and so on. Mammography is the method used in breast cancer detection. Radiologists face lot of difficulties in detection of tumors. Computer-based methods could aid medical staff and improve the accuracy of detection (Antonie, Zaiane, Coman, 2001). Neutral networks with back-propagation and association rule mining are used for tumor classification in mammograms. Data mining can be effectively used in the diagnosis of lung abnormality that may be cancerous or benign (Kuisak, *et al.*, 2000). Based on the experience and knowledge of application of data mining in the medical science, it shows that data mining algorithms largely reduce patients' risks and diagnosis costs. Using the prediction algorithms the observed prediction accuracy was 100% for 91.3% cases, although medical data is complex and difficult to analyze. In recent years data mining has received considerable attention as a tool that can be applied to cancer detection and treatment (Gong, et al., 2004, Pospisil, 2006, Barker, Clevers, 2006, Park, *et al*., 2008, Delen, 2009, Lisboa, *et al*.,2010. Hu, 2010).

Regarding data mining dedicated to liver cancer early detection and treatment, El-Serag (2002) introduced that the major risk factors of hepatocellular carcinoma (HCC) consist of chronic hepatitis virus infections, especially hepatitis B and hepatitis C; cirrhosis caused by either hepatitis or alcoholoism, and chronic exposures to various cytotoxic substances, for instance, arsenic, polyvinyl chloride (PVC), and so on. Serag developed data mining models for better understanding the fundamental mechanism leading to HCC development and early HCC detection. Applying data mining methods in medical database study, Wright and Sitting (2006) implemented the association rule technique to find out the similar medical pathways from an ambulatory computerized physician order entry system, such as child's vaccination, prostate specific antigen and the treatment of breast cancer. All of clinical orders can form an order set in which doctors are able to select proper orders, thus minimizing the mistakes and seeking relevant orders, in turn, these methods are more efficient to decrease medical errors and improve the quality of the healthcare service. Luk et al. (2007) used artificial neural network and classification and regression tree algorithms in an attempt to distinguish HCC from non-tumour liver tissues. They employed 2-dimensional gel electrophoresis to produce protein expression profiles of 66 tumour and 66 non-tumour paired samples. Eventually, they revealed that those classification algorithms were suited to be applied to the building of classification models based on the hidden pattern in the proteomic dataset. In addition, artificial neural network and classification and regression tree algorithms generated good predictive abilities in differentiation between tumour and non-tumour tissues for liver cancer. Lin (2009) proposed classification and regression tree (CART) and case-based reasoning (CBR) techniques to structure an intelligent diagnosis model aiming to provide a comprehensive analytic framework to raise the accuracy of liver disease (cancer) diagnosis. The major steps in applying the model include: (1) adopting CART to diagnose whether a patient suffers from liver disease; (2) for patients diagnosed with liver disease in the first step, employing CBR to diagnose the types of liver diseases. In the first step, the CART rate of accuracy is 92.94%. In the second step, the CBR diagnostic accuracy rate is 90.00%. The experimental results showed that the intelligent diagnosis model was capable of integrating CART and CBR techniques to support physician in making decisions regarding liver disease diagnosis and treatment.More recently, Rajeswari and Reena (2010) used the liver disease datasets obtained from UCI repository consists of 345 instances with seven different attributes to test three DM algorithms: Naive Bayes algorithm, FT Tree algorithm and KStar algorithm. The study results showed that FT Tree had better classification accuracy compared to other algorithms.

Chapter 4 Challenges and Recommendations for Applying Cloud Computing

To achieve the objective of the thesis, namely to provide an overview of the past and current research in the challenges of adopting cloud computing, a literature review was conducted. This chapter also presents details of recommendations for technical and non technical challenges.

4.1. Technical Challenges and Recommendations of Adopting and Growing Cloud Computing

Many challenges exist with maintaining the level of protection of data and cloud computing fundamental functionality required by current healthcare service in cloud computing infrastructure. Thus, there are restriction on cross-border patient data storage and transfer in the health cloud. Given recent investigation, the challenges we face are technical (operation availability, service reliability, interoperability, distributed system bug, complex maintain service, system security breach) and non-technical (organizational change, mutual trust, software licensing, standards, data ownership, information privacy) obstacles to the implementation of cloud computing. For our contribution, we make efforts to provide corresponding solutions to help users improve their understanding about how to design and operate cloud computing systems in a more effective and secure way.

Table 2: Summary of technical challenges and recommendations for adoption of cloud computing

	Challenge	Recommendation
1	Operation availability	Multiple back-up cloud providers
2	Service reliability	Increase the flash memory, network nodes and IT resource
3	Data interoperability	Form the standardized interface and open protocols and API
4	Distributed system bug	Distributed VMs
5	Complex maintaining service	Simplify the maintaining process and create the test domain
6	System security breach	Establish security standard, policies and encrypt and segregate the private data, as well as run regular audit and risk assessment

1. **Challenge**: *Operation availability*. According to technical restrictions, the first challenge for the adoption of cloud computing is about stability of operation. Organizations which use cloud computing will have concerns on sufficient availability. The service outage may bring up huge loss due to inadequate back up plans. Given customers' high expectation, cloud computing service providers use multiple network providers so that single point failure could not lead to take them off the air. More importantly, very high availability of cloud computing requires multiple cloud computing service providers, regardless of multiple internet providers, as they need to have the similar software infrastructure and accounting systems (Armbrust et al., 2009). According to the Los Angeles Times, the six-hour outage of Cerner's network late last month has raised fresh concerns about cloud hosting of patient records. Cerner declined to say how many facilities were affected by the July 23 outage, which it attributed to "human error." however, the outage affected Cerner's entire network nationally and possibly internationally.

The company serves about 9,300 facilities worldwide, including more than 2,600 hospitals. A Cerner spokeswoman told the *Times* that it was reviewing its training and procedures to improve its response. The University of Pittsburgh Medical Center, which runs Cerner systems across more than 20 hospitals, experienced a 14-hour outage in December, according to the *Times*. However, the hospital had a backup system that enabled doctors and staff to continue accessing patient records.

Recommendation: In order to ensure the business continuity, multiple cloud service providers are beneficial when we plan to set up a cloud computing based system. Whenever there is any problem with one provider, another cloud service could be offered from another provider as a back-up. For example, the current major cloud computing service providers across world are Amazon Web Services, Google, Microsoft Azure, Yahoo and Salesforce. Com. The organization is better to share their demand with some of those providers, not merely bind with a single one, thus ruling out the potential disruptive resource-dependent restriction. In addition, the service model in the most of cloud service providers is pay per use, which provides a doable management approach to connect to multiple service providers for an organization. If applicable, we need several servers to run test and development tasks to a cloud-based infrastructure, thus ensuring the cloud can be run safely, securely and speedily.

2. **Challenge**: *Service reliability*. A high quality of cloud service depends on how stable the cloud system can run all the time. The performance of cloud service from providers should have strong reliability to meet the need of users, especially

healthcare industry users. Healthcare services and data must be error-free, as all of decisions, whether about single persons or societal health, will take into account of data and service from eHealth cloud. Technically, cloud computing is based on multiple virtual machines sharing CPUs and main memory, as well as I/O channels. It may cause the problem in the I/O interface between virtual machines, when the bulk of data is transferred in the cloud computing system. Literally, a health related cloud system may need to handle hundreds of healthcare providers' data and millions of patient records synchronically.

Recommendation: From a technical point view, one way to minimize the I/O interface workload is to increase the flash memory, as it provides much faster to access and expenses less energy. As we know, flash memory made by semiconductor memory can maintain information even when electricity is cut off. In addition to flash memory, we can also increase the capacity and number of IT resources such as computer nodes, network connections, and storage unites.

3. **Challenge**: *Data interoperability* is another barrier to load, store and transfer data in distinctive organizations and sites, as we lack active standardization. For instance, service for the healthcare industry can be provided from different cloud service providers, such as medical images with different resolution, which bring some difficulties to store patient data and conduct data mining or analysis tasks in the same cloud server. Therefore, a proper degree of interoperability must allow customers to operate their data and programs from one site to run on another. As

well, a good degree of interoperability can facilitate smooth data migration among various available systems.

Recommendation: The obvious solution is to find a standardized interface. Then, a cloud computing software provider could deploy services and data across multiple cloud computing providers so that the failure of a single company would not take all copies of client data with it (Armbrust et al. 2009). In other words, the standardization of APIs provides the opportunity for cloud computing which the same software infrastructure has capability to apply in a private and a public cloud. Data migration between an old local application and a new cloud computing system can be easily achieved if open protocols and APIs are offered. As well, many researchers (Al-Jaroodi and Mohamed, 2012, Nguyen *et al*. 2012) manage to apply the concept of service-oriented Architecture (SOA) to the healthcare cloud system. SOA with standardized models and protocols is prone to make service available and easily accessible. We don't need to care about the underlying infrastructure, development models or implementation details. As a result, it assists to conquer the interoperability problem and loose coupling among cloud system components and users.

4. **Challenge**: *Distributed system bug*: One of the outstanding problems in cloud computing is getting rid of bugs in these quite large scale distribution systems. We barely see these bugs appearing in smaller configurations, so the debugging ought to take place at scale in the production datacenters. The plausible solution is the reliance on virtual machines in the cloud computing. A lot of traditional SaaS

providers developed their infrastructure without using virtual machines, or in that they preceded the recent popularity of virtual machines or they felt they could not afford the performance hit of virtual machines (Armbrust et al. 2009). The bugs occurring in the cloud based systems cannot be reproduced in smaller configurations, so the debugging must take place at scale in the production datacenters.

Recommendation: Distributed VMs need to be relied on for debugging process. The level of virtualization would be very helpful to capture useful information in ways that are impossible without VMs.

5. **Challenge**: *Complex maintaining service*: Cloud computing, whether private or public or hybrid type, is a complicated integrated system, which is unlike having a single operating system for individual client services. Herein, it will increase the complexity of system maintainability in the cloud computing system relative to an individual system. The increase is largely caused by the need to consider the requirement and characteristics of the multiple service providers and clients. These requirements can be completely distinctive while maintenance in the cloud infrastructures, software, or platform must be done without negative impacts on any services provided for any clients.

   Recommendation: In order to reconcile the issue, we need to simplify the maintenance process, as all cloud resource and provided services must be designed for easy and accountable maintenance. As well, testing domains can be deployed along with production domain to minimize the potential negative effect, thus reducing the time and cost used in the cloud service maintenance task.

6. **Challenge:** *System Security Breach*: Of the entire barriers list above, the system security issue emerges as the major concern for the adoption of cloud computing system, in particular for health related cloud services. Medical data and information is a security risk because they're vulnerable to breach, whether intentional or not. The security in the context of the cloud computing system is a complex situation that must be discussed. For a given cloud server centre that has literally thousands or tens of thousands of medical data from a wide spectrum of data sources, this becomes a very significant IT administrative task and operational task, and even makes the risk more complex. Developing balanced regulations to improve the system security are tough, since you want to create regulation to safeguard the data safety, while, at the same time, you want to ensure that regulation does not stifle innovation.

   Furthermore, according to Grobauer et al. is (2011) viewpoint, we should make a clear distinction between technical security risks facing any IT infrastructure and those risks imposed solely caused by the nature of the cloud itself. From the perspective of system security:

   (1) Unauthorized parties and individuals cannot access the system without any permission;

   (2) The data in the cloud server is accurate and consistent;

   (3) Users are the persons they claim to be;

(4) Users access only data they are allowed to access based on their authentication and access levels;

(5) The safety of data and the overall system are monitored and recording by the proper auditing system with user and data access activities.

(6) The wide range of security requirements among organizations and service providers may not be fully reflected in cloud service security policies.

The Cloud security alliance has proposed a cloud model to show how security and compliance map, shown in Figure 11 (CSA, 2009).

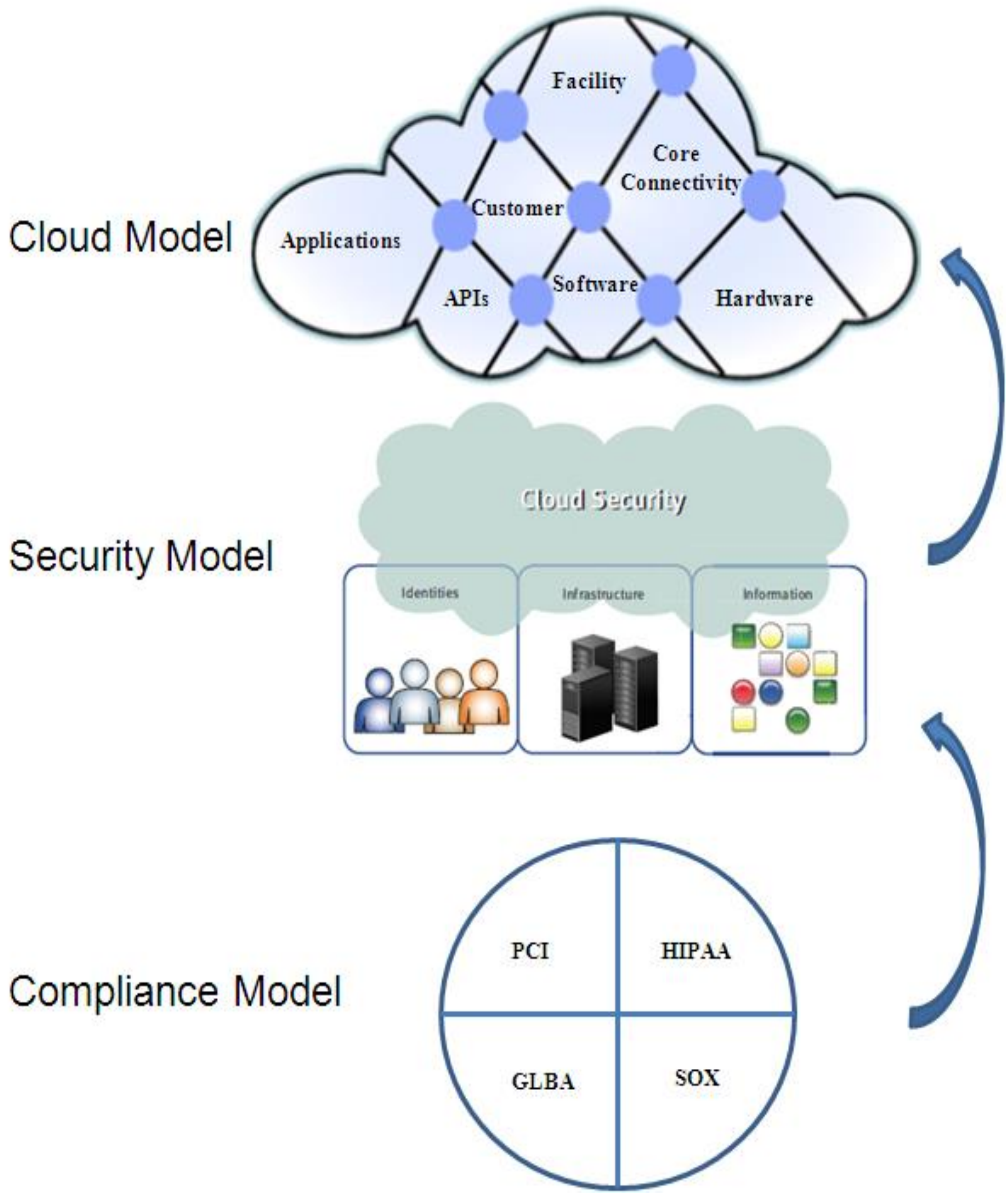

Figure 11: Mapping Cloud Model to Security and Compliance Model.

Considering the open environment of a cloud infrastructure, it is very critical to offer cloud services which support appropriate and sufficient access control and authentication mechanisms besides principles to secure the transfer of confidential data to and from clients and service providers. It is a baseline to keep secure in the multi-sites clouds where it is conserved along with other data.

Recommendation: Researchers have put effort to analyze the security risks to simplify the process of incorporating adequate security measures. The below table (table 3) is prepared to present some security solutions and approaches that have been proposed to fulfill the requirements for securing the storage and providing access control to health related data in the cloud computing based system.

Table 3: Summary of recommended security approaches for cloud services

Publication	Target	Solution	Functions	Restrictions
Li *et al*. (2010)	Secure personal health record (PHR).	Patient-Centric and fine-grained data access control through multiple-owner settings model.	Fine-grained access; user revocation; flexible data access policy; beak-glass access.	Computation overhead on data owner; risks of privacy exposure by cloud owner.
Yu *et al*. (2010)	Secure PHR, maintain data confidentiality against cloud owner; reduce overhead on the data owner for key distribution and data/user management.	Allow data owner to delegate computation tasks in fine-grained data access control to cloud owners without disclosing	Cloud owner is not able to learn any “plaintext medical data”; computation overhead is reduced on users which saves their efforts and	Limited data access policies management; no mechanisms to serve emergency scenarios.

		content by combined advanced cryptographic techniques: Key-Policy Attribute-Based Encryption (KP-ABE), Proxy Re-Encryption (PRE) and lazy re-encryption.	time online; scalability.	
Ahmed and Raja (2010)	Ensure privacy and confidentially of EHR; Track activities at the Cloud owner end of the e-Health Cloud infrastructure.	Install Computer forensic tools on the cloud machines to capture any volatile information. Different mechanisms of authentication , authorization and access control procedures to protect nonvolatile information.	Protect medical data confidentially against cloud owners; CFT provides digital evidence to be used to establish cyber crime in courts of law.	CFT promises to track volatile actions and to support cases in courts; but it does not physically stop volatile actions.
Lohr *et al*. (2010)	Secure end user platforms when accessing e-Health Cloud.	Trusted privacy domains (TVD): Systems in the end-user platform should be able to divide the execution environment	Overcome limitations of security features within end user platforms; reduce security risks; Automatic management	Hardware requirements; complexity and scalability

		for applications into separated domains isolated from each other.	(transparent TVD establishment, key management and policy enforcement).	
Kaletsch and Sunyaev (2011)	Secure EHR exchange when a referral is made.	EHR is only stored at the physicians' practices (the only trusted environment). EHRs are encrypted and signed before leaving the physician's practice and decrypted only by the receiving specialist.	Provides secure encryption and signatures for all documents transferred.	Patient and patient-centric functions are not integrated in the current model.
Fan (2011)	Develop and disseminate specialized secure services/platforms for e-Health Cloud.	Digital signature, hashing and encryption, integrity check-sum, Single Point of Contact (SPoC), access control, audit trial, identity mapping, *etc*	A complete e-Health Cloud services platform, provides mechanisms to reinforce medical data confidentiality, security, integrity & auditing.	Comprehensive evaluation in a real medical environment; Integration with other public e-Health Clouds.

The reality of security is based on the probability of different risks and how effective the various mitigation strategies are in place in dealing with the perceived risks. Simply speaking, cloud security is a sense based on the psychological reaction to both the risks and the countermeasures. Therefore, the common security challenges faced by healthcare research for cloud applications are less governance, data lock-in, lack of patient consent and authorization, lack of standard technologies and solutions integrity and key management procedure. With all the fears surrounding information security and cloud computing in particular, we stress that all medical information should be stored securely in a place which can be tracked by relevant staff. Therefore, segregated data plays an important role in maintaining the system security. Users and providers should work together to segregate the data according to its sensitivity, so the only users who can access certain things from a cloud server actually need to do so. As we categorize data based on the impact of a potential security breach, we should consider intruders’ adversaries' motivations. They are likely to be interested in data for monetary or intelligence reasons, for example, so not all data is equally interesting to them.

As well, we need to create an encryption scheme. The encryption scheme ought to be efficient and easy to be used by researchers. We also need to set up access control, as the database storing the composite EHR, where an organization’s database is stored at an external third-party service provider. A new distributed access control paradigm is defined as Cryptographic Access Control (CAC), which relies on cryptography to provide confidentiality and integrity of data managed by system (Harrington and Jensen, 2003). In addition to access control, signature and verification provide security for cloud computing. Digital signature is a very useful tool for offering authenticity, integrity and nonrepudiation while it has seldom been considered to provide researchers privacy on its own. It includes group signature, ring signature and threshold signature.

More importantly, many security standards need to be developed and used in current cloud computing applications, which ensure their security and build trust. For example, ISO 27001 is designed to “provide a model for establishing, implementing, operating, monitoring, reviewing, maintaining and improving an information security management system (ISMS)” (BSI, 2005). Given their needs, objectives and security requirements, the standard defines how an organization can manage them information security. The standard also can be used by both internal and external parties in assessing the security posture of the organization, which lead to certifications indicating that a party meets the standards requirement for information security. Therefore, health informatics researchers can develop a structure security and trust framework through collaboration, thus evaluating the security requirements of the user and vendor (Kuo, 2011). We can outline the organization guidelines and mandate for information security as the security polices, so that each party agrees upon security policies, standards and practices to complete the collaboration research project. In addition, we need to organization information security, asset management, human resource security, physical and environmental security, as well as communication and operations management, access control, and incidence management.

Lastly, cloud providers need to check for vulnerabilities, run risk assessments and conduct audits and correct weaknesses discovered. The data center can consider buying and running a data loss prevention software program, which runs on the perimeter server, and apply security patches to internet applications that are connected to the cloud based systems. Providers need to make sure that the firewalls are installed properly, and that the antivirus programs are operational. Cloud service stakeholders are better to comply with objective, specific measures, such as those recommended by the National Institute of Standards and Technology or HITRUST, so they can defend the adequacy of the safeguards their taking to protect private information. The provider and client need to work together to designate that is responsible for maintaining the integrity of the system.

4.2. Non-*Technical Challenges and Recommendations*

As of any innovation, cloud computing should be rigorously evaluated from different aspects, such as technical and non-technical. To fully assess challenges rooted from cloud computing implementation, those non-technical issues are also addressed in Table 4.

Table 4: Summary of non-technical challenges and recommendations for adoption of cloud computing

	Challenge	**Recommendation**
1	Organizational change	Review the adequacy of policies and offer training courses for stuff
2	Building mutual trust	Create the relevant policies and build up high reputation
3	Software licensing	Have more open sources
4	Standard	Establish the guideline and legislation, as well as adopting the current defined standard
5	Data ownership	Government issues the policies and form a framework
6	Information privacy	Minimizing collecting personal data, access control, client’s data encryption

1. **Challenge**: *Organizational change*: The move towards cloud computing system will need significant modifications to business processes and also to the organizational boundaries in the industry. Cloud service will involve a number of clients and service providers as participants. Given the emerging concept, not all participants in cloud service can understand the underlying mechanism and necessities. Therefore, there may be resistance and a gap to implement the cloud service in an organization, such as organization a culture change, work flow change, etc.

   Recommendation: Firstly, the organization establishes a new taskforce to review the adequacy of current policies and identify what steps need to be taken to advocate the advantages of cloud services in a supportive regulatory environment. Secondly, an efficient training and learning process will assist people to understand the change well, so that staff can merge into the new operational working process without reluctance. Cloud service providers need to offer short courses for industry, government, and university on cloud computing, including its data management process.

2. **Challenge**: *Building mutual trust* is another challenge for application of cloud computing in cloud computing service. “Trust is a judgment of unquestionable utility- as humans we use it every day of our lives. However, trust has suffered from an imperfect understanding, a plethora of definitions, and informal use in literature and in everyday life” (Marsh, 1994). Fundamentally, addressing the different situations or instance where trust is required is very crucial to adopt

the cloud computing service, especially when it is constructed across distinct organizations and countries for service providers and clients. The degree of trust requirement to which each group attaches to the different situations may differ but nevertheless, these different scenarios are an indispensible part in building a trust relationship in cloud computing adoption. The scenarios include (a) access to resource or access to trustor's resource, (b) service provision (researchers place their trust on the service providers), (c) certification of trustees/identity trust (requires the researcher to believe that the service provider is as it claims to be, which is based on certification by third party of the trustworthiness of the trustee), (d) infrastructure trust (describe the extent to which the trusting party believes that the necessary systems and institutions are in place in order to support transactions and provide safety net). Herein, trust as a characteristic and quality of relationship between service providers and clients need to balance between responsibility and diligence.

Recommendation: It is always a two way relationship in the cloud computing service. To build a trust, there are two approaches. The trust can be built by using policies to establish it among multiple users and provider groups. The focus is on managing the exchange of credentials and enforcement of access policies, as trust can be established by a sufficient amount of credentials supplied by an entity wishing to have access to another entity or resource. In another approach, trust is established through the use of reputations. The reputation of service providers is determined based on past interactions or performance and used to determine or assess its future behaviour. Past interaction is used to accumulate the reputation, thus building mutual trust.

3. **Challenge**: *Software licensing*. Software licensing can limit use of the computer system and cloud service. Users have to pay for the software and pay an annual maintenance and upgrade fee for the service and software. Software licensing could diminish the interest for cloud service users, since most of them intend to decrease the operational budget and save the enterprise cost by using the cloud service.

   Recommendation: Considering limited funds and enterprise cost containment, the solution can be used to open source to remain popular or simply ask that commercial software companies change their licensing structure to better fit it to applying cloud computing for the purpose of specific user groups, such as the academic research.

4. **Challenge**: *Standards*. The concept of cloud computing indicates that the data and information stored in the providers database server need to be shared and exchanged through different parties. Lack of standards for business informatics, policies, interoperability, and transmission methods will impede the development of cloud computing services. Therefore, the stakeholders in the cloud system do not have a solid base to start offering and using it in a convenient way. More and more issues and problems may occur due to this challenge, and social and ethical concerns will arise.

Recommendation: We need to write up clear and adequate legislation and guidelines for business, technical and clinical practices of organizations in the cloud service. For example, currently, there are some standards and classification for ehealth cloud service. WHO provides the International Classification of Disease tenth revision (ICD-10) (International Classification of Diseases, 2013). ICD-10 defines a medical classification list for the coding of disease, signs or abnormal findings, complaints, social conditions, and external causes of injury or disease. The cloud service stakeholders can agree on adopting some of existing standards and classifications to enable interoperability among various entities.

5. **Challenge**: d*ata ownership*. Cloud service clients need to send their data to a cloud centre, and service providers will store those data in a proper way in the data centre. Generally, ownership of data is an area with no clear guidelines, as clients think they made the data, but providers deem that data are stored in their space. For instance, a patient's record could be the sole property of the patient, yet can his physician also claims ownership. What about the patient's insurer or the hospital management? The ambiguous data ownership may bring up potential conflict in the cloud service and threaten to the relationship between providers and clients.

   Recommendation: The government could create clear policies and establish reliable frameworks to draw clear ownership boundaries. All organizations involved in the cloud service need to make agreements in accordance with the related policies.

6. **Challenge**: *Information privacy* is the most serious issue among the non-technical challenges that causes users and organizations to be wary of implementing cloud computing in the healthcare information sharing, since personally identifiable information or any sensitive information could be exposed to the public by any criminal attacks or accidental information breach. In some sensitive areas, such as healthcare, private data exposure, data leakage, and data loss in the cloud service can be perilous. As well, a high risk of liability in cases of data loss or leakage will threaten loss of reputation and client's trust. The privacy challenge for software engineers is to create cloud services in such a way as to decrease privacy risk, and to ensure legal compliance. There is no doubt that security and privacy protection of cross-institutional and cross-national electronic patient records is very crucial, especially in healthcare information sharing. For example, few data breaches are as malicious or as in-your-face as a recent attack on Surgeons of Lake County, a small practice in Libertyville, Ill. Hackers gained access to a server that stored emails and electronic medical records. They encrypted and password-protected the files and then posted a ransom note on the server demanding payment in exchange for the password to unlock the files. The practice instead shut off the server and called police. The Surgeons of Lake County breach, which affected the records of more than 7,000 patients,

according to the U.S. Department of Health & Human Services, happened in June, 2012. This story is so ironic--most people worry that their health records will be spread all over their local newspaper. In fact, not all hackers steal health records data for personal gain--some hack patient information just for the fun of it. Information security and privacy concerns have become so widespread that providers are increasingly at risk of not being able to defend their private information and other IT systems.

Recommendation: We need to apply legislation placing geographic and other restrictions on the collection, processing and transfer of personally identifiable information limiting usage of cloud services as currently designed. Eventually, any violation of privacy will be sued by the customers (Pearson,2009).The researchers who conduct health informatics topics, need to minimize personal information sent to and stored in the cloud, protect personal information in the cloud, maximize user control, specify and limit the purpose of data usage, and provide period providers' and users' feedback. In order to meet the privacy requirements, researchers need to analyze the system to assess how only the minimal amount of patient information necessary can be exposed and stored. Minimizing the collection of personal data decreases the need to protect data as strongly during storage and processing. Where possible, encryption technique are highly recommended. In addition, privacy-preserving data mining techniques may be used to correlate the relationship of two databases with different sources, in which the only information revealed to either of the database owners about the other's data is the information that can be understood from the output of the data mining algorithm (Lindell and Pinkas, 2002). Li *et al.* (2010) proposed that the data owner has full control over the selective sharing of their personal data. Therefore, instead of the cloud service provider encrypting the personal data, clients can generate their own decryption keys utilizing attribute-base encryption (ABE) and then distribute them to their authorized users. Clients are allowed to select in a fine-grained way which users can have access and to which parts of their personal data by encrypting the record in relation with a relevant set of attributes. In addition, the client can delegate most of the computation tasks involved in fine-grained data access control to the cloud owner without disclosing the original data content, which can be accomplished through utilizing combined advanced cryptographic techniques (Yu *et al*. 2010).

Therefore, an approach to investigate the challenges of applying cloud computing in the healthcare information sharing is necessary to offer us more chances to avoid potential risks. As we know, in spite of all the efforts, the cloud computing system is still in its infancy. In this chapter, a literature review of issues for adopting the cloud computing, especially in the health care area, was presented with focus on the importance of the concepts involved, implementation and challenges. Accordingly, we discussed many proposed solutions to address the challenges and we concentrated on the security and information privacy issues as they represent the biggest challenges in both technical and non-technical categories.

Chapter 5 **Cloud Computing Based Data Mining Platform**

This chapter describes the proposed cloud computing based data mining platform with regard to the proposed research questions. According to the challenges review and opportunities exploration about adopting the cloud computing system in the previous chapter, the chapter introduces a platform that could be potentially used for geographic-distributed health research.

5.1. The Cloud Computing Based Data Mining platform

Given the above discussion of challenges and recommendations of adopting cloud computing, cloud computing systems can bring tremendous benefits to build an efficient information sharing system with low cost. The cost effective cloud computing system will allow the large organization or collaboration teams to construct and operate large-scale, commodity-computer datacenters at low-cost locations, as they indicate the factors of 5 to 7 in decrease in cost for electricity, network bandwidth, operations, software, and hardware available at these very large economies of scale (Armbrust et al. 2009). Hamilton (2008) has summarized that huge datacenters with tens of thousands of computers can own hardware, network bandwidth and power for 1/5 to 1/7 the prices offered to a small size datacenter. Table 5 is listed the comparison:

Technology	Cost in small sized datacenter	Cost in large datacentre	ratio
Network	$95 per Mbit/sec/month	$13 per Mbit/sec/month	7.1
Storage	$2.20 per GByte/moth	$0.40 per GByte/ month	5.7

Administratin	≈140 Servers/ Administrator	>1000 servers/ Administrator	7.1

Table 5: Economies of scale in 2006 for small sized datacentre (≈1000 servers) vs. large datacentre (≈50,000 servers) (Hamilton, 2008).

Considering the economies of scale for cloud computing, we will propose a cloud computer based platform for geographic distributed health data mining. International collaborative health informatics research will serve as a case to be used for the proposed platform. It is hoped it will make sense for our potential health informatics research collaboration and speed our work as it is able to accelerate our operation time to revenue and reduce our operational costs.

However, it is extremely hard to assess the proposed cloud computing system due to current restriction to access medical datasets existing in different countries. In addition, we need longitudinal, randomized, double-blinded trials involving a homogeneous patient population and medical condition, in order to provide clinical evidence for medical benefits. It also requires the technology to fully developed and ready for large scale use. This is not simple feasible when doing technological research into new types of cloud computing based health data mining, as strong evidence based on medical outcomes is impossible to obtain while we are still in the research phase of new technology. Therefore, we use the research method of “Proof-of-Concept Prototype” for studying the proposed cloud computing based health data mining system, thus demonstrating the technical feasibility for the collaborative research before applying physical and organizational system in a real-world deployment setting, and involving real patient datasets and users.

As we design in the proposed model, research teams from various geographical sites will be allowed to focus on data management areas such as usage data collection, data pre-processing, data mining algorithm design and implementation, and mined pattern interpretation, since data coming from different sites will be stored in a datacenter which is provided by cloud service providers. In addition, each team of researchers will set up their own monitoring systems and communication channels in order to build up appropriate study environments. The platform is designed to adopt a cloud computing based Service-Oriented Architecture (SOA) in order to facilitate the development and integration of heterogeneous database and application systems (see Figure 12). If any team wants to access remote resources (data or applications), it merely requests web services through the SOA by applying the standardized programming grammar. This cloud based information sharing architecture can save much communication time, reduce cost, and preserve data security.

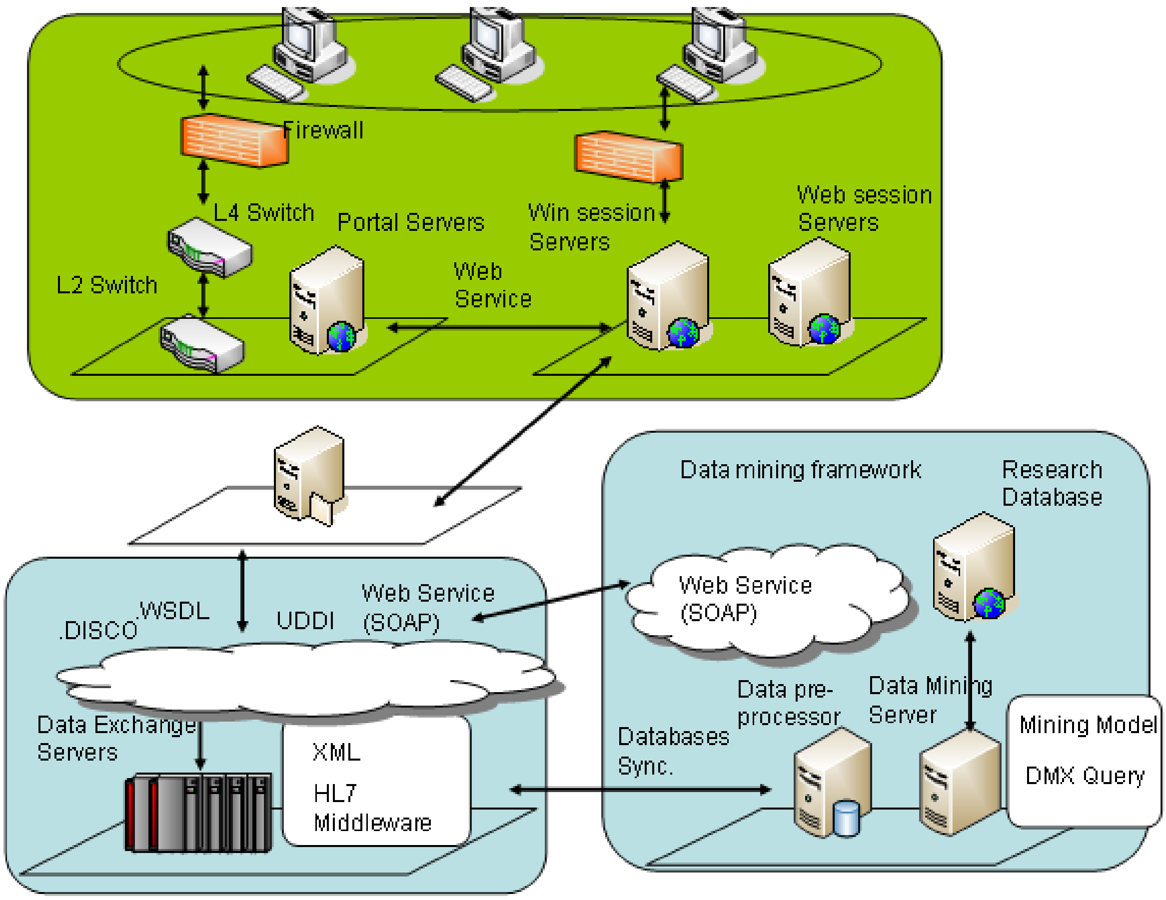

Figure 12: The Service-Oriented Architecture

The cloud algorithm was used to visualize servers with VMware in the platform (see Figure 13). It allows researchers to use VMware to open visualized server to conduct the same research study at different location. Different site researchers can share and transform outcomes via the visualized server while keeping their original patient data in local databases. The benefit of this structure is that it has preserved the original data security and privacy.

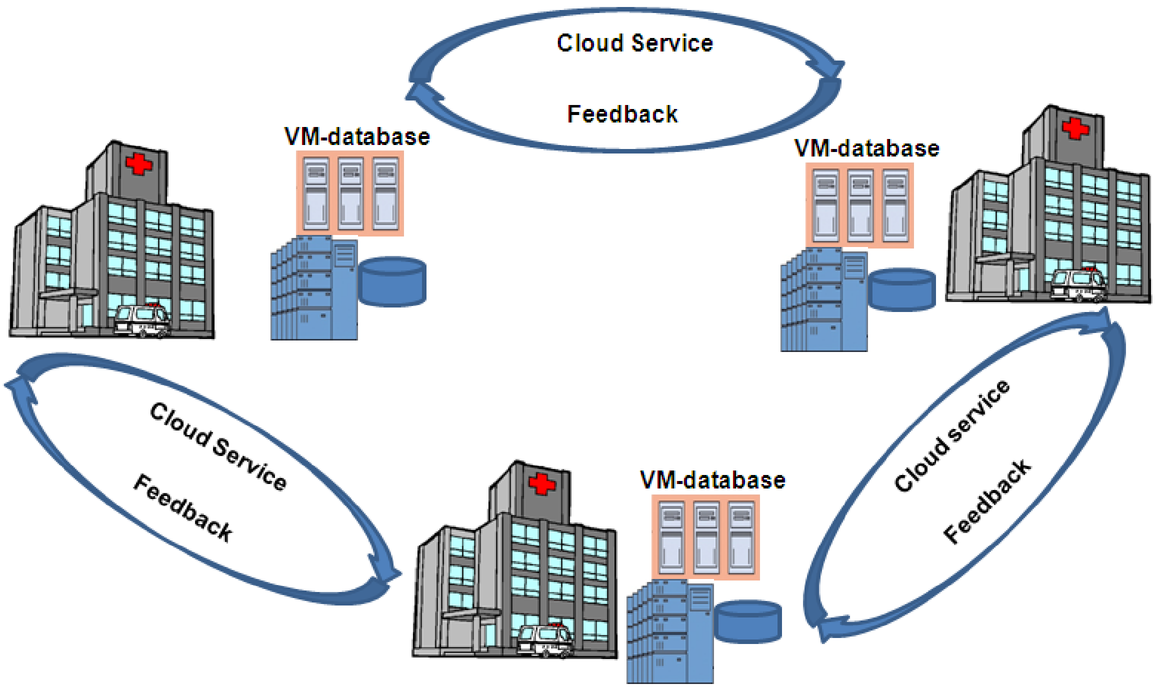

Figure 13: Visualizing server workflows

The SOA and visualizing servers form a so called "cloud architecture" which allows researchers to easily share study resources such as patient liver cancer related health data, data mining applications and storage spaces (see Figure 14) through internet. In the cloud architecture, the data mining applications can be viewed as a *Software as a Service (SaaS)* and the visualizing servers as a *Platform as a Service (PaaS)*. In other words, this is a very flexible model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service-provider interaction (Kuo, Kushniruk. Borycki, 2011).

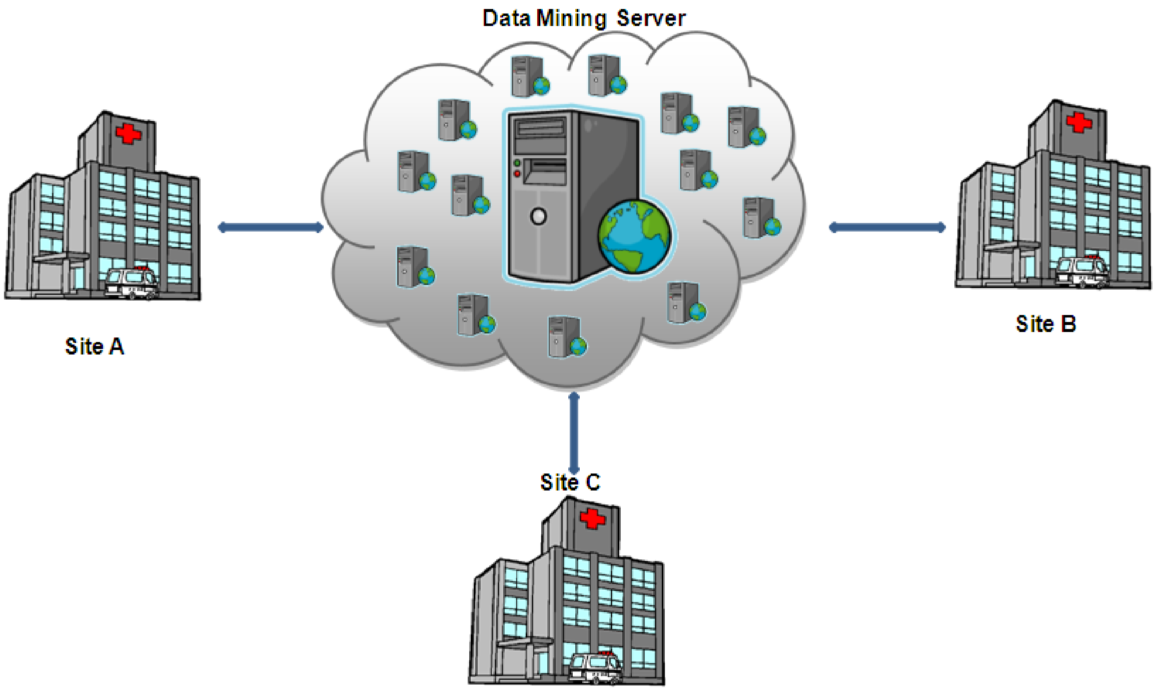

Figure 14: Cloud architecture for data mining platform across three countries

In summary, in order to conduct health information research, we need to deal with an ocean of patient data, thus finding out meaningful patterns and conclusions. In terms of our specific research target, we work on health information research that focuses on the effective and accurate investigation of the association between liver cancer and demographic and geographic characters of patients which is at the heart of liver cancer patient diagnosis, monitoring, and treatment planning. When patient's clinical information are accurately shared across distinct geographic locations, in a manner conducive to liver cancer early detection processes and systems, the business and clinical processes are improved, with possible value to the care of the liver cancer patient. The objective of the platform is to bring together researchers at the different sites that apply the state-of-the-art data mining algorithms to analyze liver cancer data extracted from their Electronic Health Record (EHR) systems. Specifically, we will develop an across national boundary cloud computing platform for sharing research data/resources across three geographic locations. Subsequently, using different algorithms, we need to further explore and interpret the relationship between the cause of liver cancer and patients' characters, such as demographic and geographic locations, thus discovering the early detection of liver cancer mechanism. Eventually, it would be ideal that, given our research outcome, we could provide the recommendations for establishing liver cancer standardized treatment pathway and guidelines. Cloud computing needs to appeal to the feeling of three parties and address the potential security risks in a manner that three parties will feel safe and secure. According to the discussion in the last chapter, by addressing the challenges to adopt cloud based service is this way various researchers from different locations will feel safer and more secure, hence, they trust cloud service providers and share patient data. The collaborative health informatics research model is demonstrated in the Figure 15:

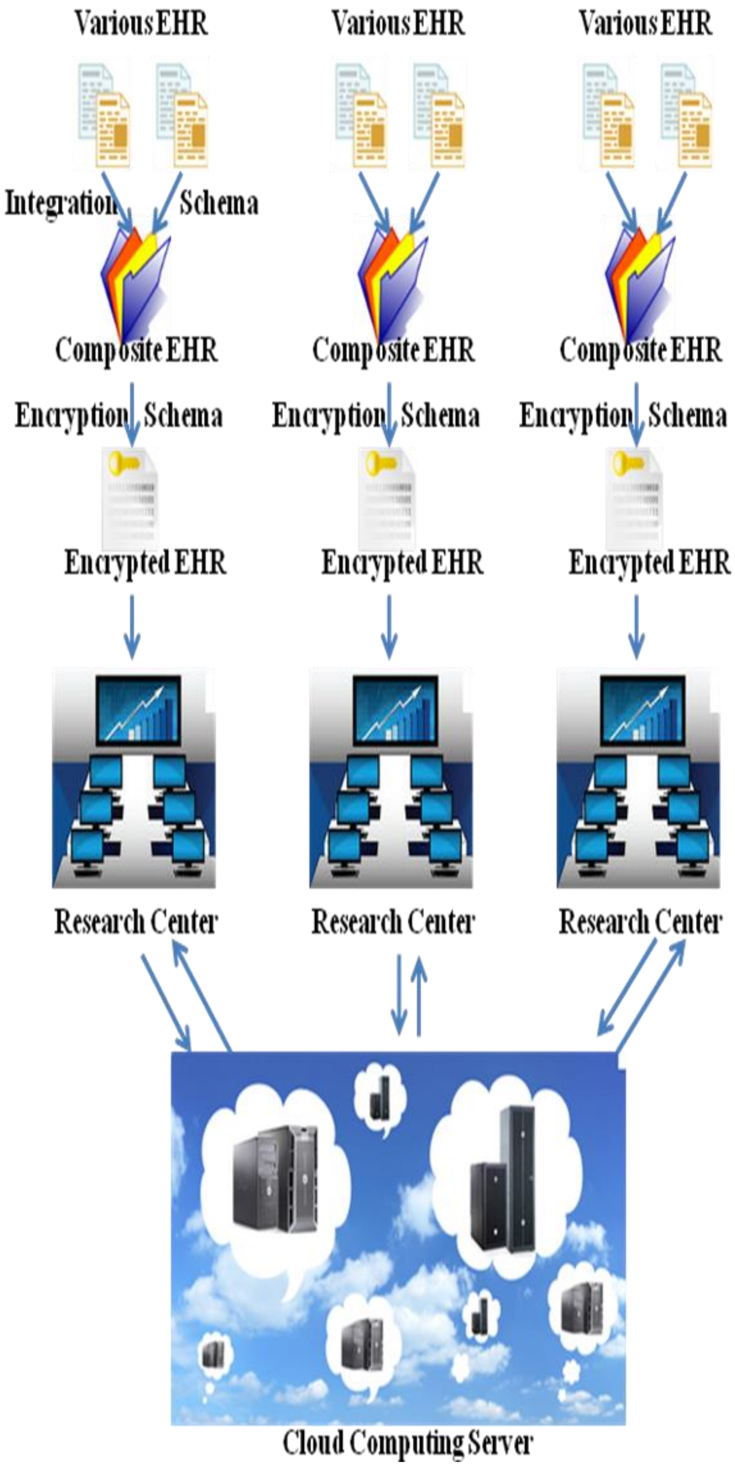

Figure 15: Collaborative Health Informatics Research Model Using Cloud Computing

Once the proposed cloud computing based system is set up, we manage to apply privacy-control distributed data mining to deliver, transfer and analysis the data from multiple data sources. In our proposed system, data will be produced from many physically distributed locations; therefore, the system requires a data centre which gathers data from distributed locations. Additionally, issues such as data privacy concerns (regarding clients information stored in a center servers) demonstrates that aggregating data for centralized mining is not an easy job. The privacy-control distributed data mining allows user to tackle issues via access-control rules, as it often applies the same or different mining algorithm to handle local data and communicate among many process units, and then combine the local pattern discovery by local data mining algorithm from local database into a globe knowledge solution. Hence, distributed data mining enables geographically distributed health informatics teams to collaborate in an innovative and privacy-controlled way. Cheung (2006) has illustrated a methodology termed learning-from-abstraction. The learning-from-abstraction methodology intends to identify data patterns in general merely, thereby discovering the local data's statistical abstractions to the outside world. Figure 16 demonstrates how the learning-from-abstraction methodology applied into our proposed cloud computing based geographical-distributed health data mining. As well, the local data owners are able to implement their own privacy policies on their ends, thus determining how much privacy protection is necessary. To abstract the local data using data mining technology, we could use various algorithms repeatedly to each abstraction level, for instance, Apriori algorithm combined with the association rule.

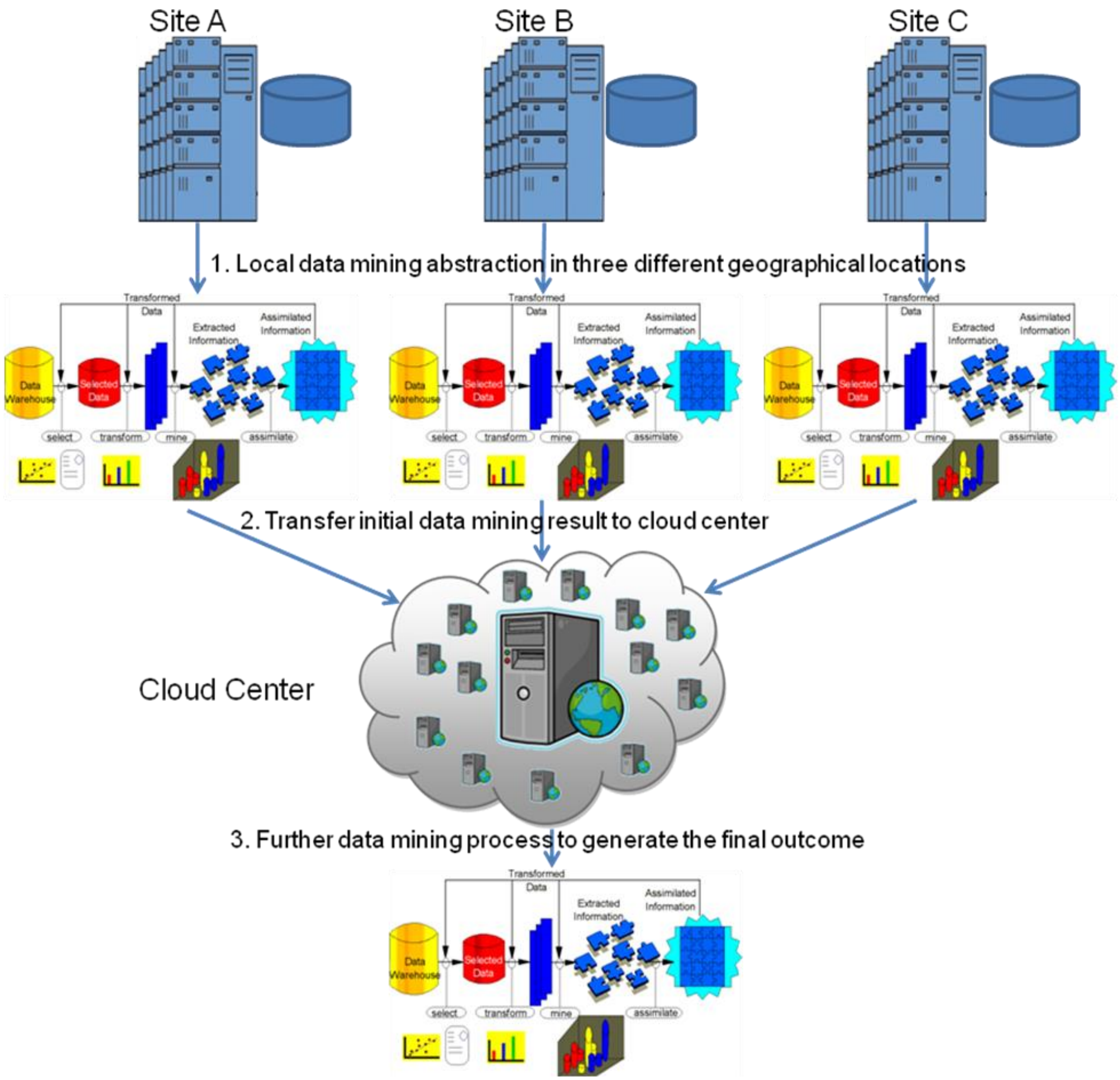

Figure 16: Suggested distributed data mining model based on learning-from-abstraction methodology

Chapter 6 **Conclusion**

Little literature in the healthcare information sharing addresses the critical challenges and solutions of applying cloud computing in collaborative health research. While the use of cloud computing continues to increase, legal and security concerns are also increasing. Although cloud computing providers may run many obstacles, we believe that over the long run providers will successfully navigate these challenges. By using a cloud computing security framework, the collaborative parties can answer questions related to governance and best practice and determine whether the organization is capable of IT governance in the cloud computing applications.

In this thesis, we introduced the concept of cloud computing, elucidated many technical and non-technical challenges confronting the success of adopting cloud computing, and explored different potential solutions to address the corresponding challenges for the application of cloud computing, such as system security and information privacy.

Furthermore, we proposed a cloud computing based platform for geographic distributed health data mining using a proof-of-concept methodology, which provides valuable information regarding the technical applicability of the system, its usability, and issues related to the physical an organizational deployment of the system. There is no doubt that cloud computing has the potential to overcome collaborative health research issues. However, in addition to ensuring that health informatics researchers and health providers are equipped with technology to help improve health research and provide adequate care for patients, they also need to be aware of how such systems impact business and administrative operations. Few studies have evaluated the impact of cloud computing applications to such collaborative research in terms of its opportunities and challenges. Proof-of-Concept methodology is a more cost-effective approach for demonstrating initial technical evidence as well as a invaluable source for improving the platform at a stage before resources are invested in final development and system.

The contributions we made in our study:

(1) Evaluated the technical and non-technical obstacles to adopt and grow the cloud computing system, coupling with exploring the corresponding opportunities to overcome those challenges.

(2) Designed and developed a flexible, secure and privacy-preserved cloud platform for researchers in different geographic locations for easily exploring diverse medical data for health informatics research through internet.

(3) The thesis results can be used as a reference by other similar research to determine whether (or how) to migrate from traditional isolated computer system to cloud-based services.

Besides the cloud computing platform, data mining could be designed to collect data from different data repositories consists of many records with different attributes. The records in the different datasets are pertaining to various characters of patients with different liver cancer diseases, which are thought to be sensitive to liver disorders that might arise from demographic and geographic locations. By using specific algorithms, it would be able to pinpoint that the relationship of geographic and demographic with incidences of liver cancer is strong. If possible, this platform could facilitate researchers to explore diverse medical data for other types of cancer early detection and developing better clinical pathways and clinical guidelines for the cancer treatment, provide a practical international collaboration between academic institutes and healthcare providers, and develop an across national country boundary architecture for sharing research resources/experiences.

In conclusion, each country has different geographical locations, climates, races, living habits, diets, culture, social patterns, etc which lead to different cancer risks and medical treatments. The effective and accurate investigation of the association between liver cancer and demographic and geographic characters of patients is at the heart of liver cancer patient diagnosis, monitoring, and treatment planning. The proposed cloud computing model is expected to be share clinical data and facilitate to extract knowledge from a huge number of raw data in different locations. Then, it could be used for liver cancer early detection study. When patient's clinical information are accurately shared across distinct geographic locations, in a manner conducive to liver cancer early detection processes and systems, the business and clinical processes would be improved, with possible value to the care of the liver cancer patient. If this platform can be further developed in the near future, we could explore a reliable information system that may be able to contribute the liver cancer research, and also create a useful international platform to share medical data across boundaries in the industry.

Chapter 7 the Future Study

In future study, we hope to further develop our cloud computing platform into a geographic-distributed international collaborative research for liver cancer early detection study. As of the international collaborative research, multiply groups of researchers from different countries could use their existing framework. The main benefit of the geographic distributed collaborative project is that researchers can apply data mining algorithms to analyze diverse clinical diagnosing records contained in distinct medical information systems to discover the hidden knowledge related to liver cancer. The future study includes three stages as follows:

Stage 1

- *Set up and test the cloud based platform*

  To meet all involved research teams joining the collaborative research need, the cloud based data mining platform must be adjusted and tested to meet all parties' requirement of sharing study resources and results. Each part of the platform will be examined by unit tests during the setup stage. Upon integrating all parts into a single unit, we will test the process of data sharing.

- *Data security and privacy protection*

  The cloud based data mining platform is usually accessible to many different researchers. Therefore, it poses increased data protection risks for users and providers when compared with traditional client-server systems. The non-identified client data must be enforced whenever the medical data is shared in the three research institutes.

- *Sample Selection*

Fundamentally, as the research design, patients from the three distinct geographic locations will be our target population. It is better to include at least two different types of ethnic or cultural groups. The research data will be obtained from EHR systems of different hospitals. We recommend that the population is to provide a test bed environment for liver cancer early detection and standardized treatment process studies. Statistically, in conducting any research study, proper sample and sampling techniques need to be selected prudently as they are related to time, money, and generalization (Jackson & Verberg, 2007). Random sampling methods are the most unbiased methods and easiest to generalize research findings to the greater populations; however, it is unrealistic for conducting most of studies because it is too expensive and time-consuming (Nieswiadomy, 2008). Despite nonrandom sampling being more likely to produce biased samples and restrict generalizations, the research literature confirmed that nonrandom convenient sampling method will continue to be the most popular sampling technique in sociological research because of savings in time and money (Nieswiadomy, 2008). Therefore, we are going to utilize convenience sampling technique to select patients as my samples. In terms of sample size, Nieswiadomy (2008) pinpointed that a sample size of 30 for a quantitative research is considered minimal size to represent the research subject.

Stage 2

- *Conduct medical data pre-process*

In the platform, patients' medical information would be stored in distributed heterogeneous databases which include the entire clinical orders generated from physicians. Those clinical records cover the medical process and knowledge based on various clinical diagnoses. However, there is a great deal of variation in the hardware, software, coding methods, terminologies/nomenclatures used and their definitions between systems. Furthermore, different users may interpret the same data (e.g. words or terms) in different ways. These problems are barriers to further data mining processes (Kuo, *et al*,2011). As data mining can only uncover patterns already present in the data, the target dataset must be large enough to contain these patterns while remaining concise enough to be mined in an acceptable timeframe. Thus, data pre-processing is essential to analyze the multivariate datasets before clustering or data mining.

- *Implement data mining algorithms*

Data mining algorithms for this study, such as distributed data mining algorithms, association rules, sequence clustering, etc, are required to be implemented, tested and evaluated. This allows research teams to determine if the algorithms fit the knowledge discovery requirements. For example, a sequence clustering algorithm was applied to find a pathway of chemotherapy which must be able to find out certain association rules between clinical order sets, and the rules can be used to actually improve treatment.

- *Data Mining Analysis*

The association rules can be used in the each unit. An association rule is an implication or if-then-rule which is supported by data. The motivation for developing association rules is market basket analysis which deals with the contents of point-of-sale transactions. A typical association rule resulting from such a study could be “90 percent of all customers who buy bread and butter also buy milk." Nevertheless, association rule is rarely explored to be used in health data. The data is arranged into "transactions" which contain a set of data items focused around a specific event, object, or time period. For our research a record is a collection of information about patient and the diagnostic outcome they experience over a short period of time. For example:

{Age: 30-39, female, white, liver cancer, Victoria Canada}

From these records association rules are generated, which is an implication of the form X→Y, where X and Y are disjoint subsets of all the possible data items.

The strength of the association rule can be measured by its support and confidence as follows:

Let I = {i1, i2….,im} be a set of items and X, Y belong to I, then the support of an association pattern is defined as equation (1).

$$\text{Support } (X \rightarrow Y) = P\,(X \cap Y) \quad (1)$$

, and the confidence of the association pattern is defined as equation (2).

$$\text{Confidence } (X \rightarrow Y) = \text{support } (X \cap Y)/\text{support } (X) \quad (2)$$

Support determines how often the data items in a rule are present together in a record in a given data set and is simply the count of records that contain X and Y. Support is applied to decide if a rule is of interest as high support indicates that the rule occurs often in the data. Confidence is used to determine the reliability of the inference made by the rule and is an estimate of the conditional probability of Y given X. It suggests "If X is present in a record, how likely is it that Y is also present". Therefore, these rules indicate a strong co-occurrence relationship between the given data item subset. It should be noted however that these rules do not automatically imply causality in a relationship. Fundamentally and ideally, we will show that multilevel characteristic association rules can reveal potential interaction between demographic/geographic and incidences of liver cancer. The platform could use the Apriori algorithm to perform association analysis on the characteristics and attributes of the liver cancer patients, the place where they lives, patient's race and the patient's age. The algorithm could generate association rules that have high levels of confidence and support. Our analysis will not include records with missing data, however, future work will include the missing data in the electronic health record systems. This might increase the performance as the size of data set will considerably increase.

- *Establish standardized clinical pathways and guidelines*

  If we apply data mining algorithms (e.g. sequence clustering) to mine the hidden liver cancer clinical pathway, a team including liver cancer specialists and domain experts will evaluate the pathway and follow the SAGE (Standards-Based Active Guideline Environment) guideline model (Tu, 2007) to establish standardized treatment guidelines.

Stage 3

- *Validate the results*

  The final stage of the knowledge discovery is to verify the patterns/guidelines produced by previous stage. Physicians and domain experts will again be involved in the validation process to interpret the mined patterns. Also, the validation uses a test data set which the data mining algorithm was not trained on. The learned patterns could be applied to the data set and the resulting output could be compared to the desired output. A number of statistical methods may be used to evaluate the data mining algorithms such as Receiver Operating Characteristic (ROC) curves.

Reference

Abukhousa, E.,Mohamed, N. And Al-Jaroodi, J. (2012) eHealth Cloud: Opportunities and Challenges. Future Internet. 4, 621-645.

Acmen Solution. (2009). Acumen nabs ONC cloud computing contract. http://www.cmio.net/index.php?option=com_articles&view=article&id=20648:acumen-nabs-onc-cloud-computing-contract&division=cmio. Accessed on September 1st, 2011.

Ahmed, S.; Raja, M.Y.A. (2010) Tackling cloud security issues and forensics model. In Proceedings of the High-Capacity Optical Networks and Enabling Technologies (HONET 2010), Cairo, Egypt, pp. 190–195.

Agrawal, R., Imielinski, T. and Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. In Proceedings of the International ACM SIGMOD Conference, Washington DC, USA, pages 207–216.

Al-Jaroodi, J.; Mohamed, N. (2012) Service-oriented middleware: A survey. J. Netw. Comput. Appl. 35, 211–220.

Antonie, M. L., Zaiane, O.R., Coman, A. (2001). Application of Data Mining Techniques for Medical Image Classification. Proceedings of the Second International Workshop on Multimedia Data Mining in Conjunction with ACM SIGKDD conference, San Francisco.

Andrei, T. (2009): Cloud Computing Challenges and Related Security Issues

Ansari,S., Kohavi, R., Mason, L., and Zheng,Z. (2001). Integrating E-Commence and Data Mining : Architecture and Challenge. Proceeding of IEEE International Conference on Data Mining.

Armbrust M, Fox A, Griffith R, Joseph AD, Katz RH, Konwinski A, et al. (2009) EECS Department, UC Berkeley. 2009. Above the Clouds: A Berkeley View of Cloud Computing. Technical Report. URL: http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf accessed on January 28, 2013.

Armbrust, M., et al. (2010). A view of cloud computing. Communication of the ACM, 53 (4), page 50-58.

Barker, N. and Clevers, H. (2006). Mining the Wnt pathway for cancer therapeutics. *Nature reviews-Drug discovery*. 5(12), 997 -1014.

Benbasat,I. (1987).The Case ResearchStrategy in Studies of Information Systems. MIs Quarterly, 11(3), 369-386.

Bingelow, S. J. (2010):Pro and Cons of Moving to the Cloud;Virtual Data Center E-ZineVol 21;TechTarget. www.techtarget.com.

BSI (2005) BS ISO/IEC 27001: 2005/BS 7799-2:: Information Technology-Security Techniques-Information Security Management Systems-Requirements. British Standards Institution. (2005 a).

Buyya, R., Yeo, C. S., et al. (2008) Market-oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities. 10th IEEE Conference on High Perfomance Computing and Communications. IEEE

Bragg, R. (2008) Cloud Computing: When computers really rule. Tech News World.12/12/2009

Canadian Cancer Society. (2010). Canadian Cancer Statistics 2010.

Canadian Institute for Health Information. (2011). Learning From the Best: Benchmarking Canada's Health System. https://secure.cihi.ca/estore/productFamily.htm?locale=en&pf=PFC1676 Accessed on November 24, 2012

Catteddu, D. & Hogben, G. (2009): Cloud Computing: Benefits, risks and recomendations for information security; European Network and Information Security Agency (ENISA);

Chapman, P., Clinton, J., Kerber, R, et al. (2005). CRISP-DM 1.0 : Step-by-step data mining guide. NCR Systems Engineering Copenhagen (USA and Denmark), DailerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringenen Bank Group B.V (The Netherlands).

Chen, P.;Freg,C.;Hou,T.; Teng, W.-G. (2010). Implementing RAID-3 on Cloud Storage for EMR System. *In Proceeding of the 2010 Internatinoal Computer Symposium* (*ICS*), Taiwan; pp.850-853.

Cheung, WK., Zhang, XF., et al. (2006). Service-Oriented Distributed Data Mining. The IEEE Computer Society.1089-7801.

Chow, R., Golle, P., et al. (2009) Cloud Computing: Outsourcing Computation Without Outsourcing Control. 1st ACM Cloud Comptuting Security Workshop. ACM

Cloud Computing (2013): Clear Benefits: The Emerging Role of Cloud Computing in Healthcare Information Systems. Available online: http://www.techrepublic.com/whitepapers/ cloud-computing-clear-benefits-the-emerging-role-of-cloud-computing-in-healthcare-information-systems/2384337 (accessed on February 5 2013).

CSA (2009) Security Guidance for Critical Areas of Focus in Cloud Computing V2.1. Cloud Security Alliance

CSA (2010): Top Threats to Cloud Computing V1.0; Cloud Security Alliance; 14

Davidson, E.; Heslinga, D. (2006) Bridging the IT adoption gap for small physician practices: An action research study on electronic health records. Inf. Syst. Manag. 24, 15–28.

DEshpande, S.P. and Thakare, V.M. (2010). Data Mining System and Applications : A Review. International Journal of Distributed and Parallel Systems (IJDPS). 1 (1), Pages 32-44.

Delen, D. (2009). Analysis of cancer data: a data mining approach. *Expert Systems*, 26(1), 100 -112.

Dunham,M.H., Sridhar S.(2006). Data Mining: Introductory and Advanced Topics. Pearson Education, New Delhi, ISBN : 81-7758-785-4, 1st Edition,.

Dustin Amrhein, P. A., Andrew De Andrade, Joe, Armstrong, e. A. B., James Bartlett, Richard
Bruklis, Ken Cameron, et al. (2010) Cloud Computing Use Cases White Paper. Version 3.0 ed., Cloud COmputing Use Case Discussion Group

DWreview.com.(2007). Web site accessed in 2011 http://www.dwreview.com/Data_mining/Future_data_mining.html.

El-Serag, H.B. (2002). Hepatocellular Carcinoma: An Epidemiologic View. *J. Clin. Gastroenterol*, 35, S72-S78.

European Commission. (2013) Protecting Your Personal Data. Available online: http://ec.europa.eu/ justice/data-protection/individuals/index_en.htm (accessed on Feb 21, 2013)

Fan, L.; Buchanan, W.; Thummler, C.; Lo, O.; Khedim, A.; Uthmani, O.; Lawson, A.; Bell, D. (2011) DACAR Platform for eHealth Services Cloud. In Proceedings of the 4th International Conference on Cloud Computing, Miami, FL, USA, pp. 219–226.

Fardon T, Haggart K, Lee DK, Lipworth, BJ(2007). A proof of concept study to evaluate stepping down the dose of fluticasone in combination with salmeterol and tiotropium in severe persistent asthma. Respir Med, 101:1218-1228.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, R. (1196). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, 39(11), 27-34.

Ferlay, J., Shin, H.R., Bray, Forman, F., D., Mathers, C. and Parkin, D.M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. International Journal of Cancer, 127(12), 2893-2917.

Foster, D.P., and Stine, R.A. (2004). Variable Selection in Data Mining: Building a Predictive Model for Bankruptcy. Journal of the American Statistical Association, Alexandria, VA, ETATS-UNIS, 99, ISSN 0162-1459, Pages 303-313.

Geelan, J. (2009) Twenty-One Experts Define Cloud Computing. Cloud Computing Journal. Cloudcomputing.sys-con.com,12/4/2010 12/4/2010

Ghani, R., Probst, K.,Liu,Y.,Krema,M.,Fano,A. (2008). Text Mining for Product Attribute Extraction. SIGKDD Explorations, 8(1).

GNI (2009) Demystifying the cloud: Important opportunities, crucial choices. GNi

Gong, J.L., et al. (2004). Data mining techniques for cancer detection using serum proteomic profiling. Artificial intelligence in medicine, 32(2), 71 -83.

Goldschmidt, P.G. (2005) HIT and MIS: Implications of health information technology and medical information systems. Commun. ACM, 48, 69–74.

Guo L, Chen F, Chen L, Tang X. (2010). The building of cloud computing environment for e-health. In: Proceedings of the 2010 International Conference on E-Health Networking, Digital Ecosystems and Technologies (EDT). New York, NY: IEEE; 2010 Presented at: The IEEE International Conference on E-Health Networking; Lyon, France.

Grance, T. (2010) The NIST Cloud Definition Framework. NIST

Grobauer, B.; Walloschek, T.; Stocker, E. (2011) Understanding cloud computing vulnerabilities. IEEE Secur. Privacy Mag. 9, 50–57.

Halteren, H. Van. (2004). Linguistic Profiling for Author Recognition and Verification. Proceeding of the 42nd Annual Meeting on Association for Computational Linguistics USA, Barcelona, Spain, Article No. 199.

Hamilton, J. (2008). Cost of Power in Large-Scale Data Centers [online]. November 2008. Available from: http://perspectives.mvdirona.com/2008/11/28/CostOfPowerInLargeScaleDataCenters.aspx. Accessed on May 10th, 2013

Hamilton, J. (2008). Internet-Scale Service Efficiency. In Large-Scale Distributed Systems and Middleware (LADIS) Workshop.

Harrington, A. and Jensen, C.. (2003) Crytographic Access Control in a Distributed File System. Proceeding of the eight ACM symposium on Access control models and tecnologies, 158-165

Harding, J.A., Shabbaz, M., Srinivas, Kusiak, A. (2006). Data Mining in Manufacturing: A Review. Journal of Manufacturing Science and Engineering. 128,Pages 969-976.

Haughton J. (2011). Year of the underdog: Cloud-based EHRs. Health Manag Technol;32(1):9.

Hu, R. (2010). Medical Data Mining Based on Association Rules. *Computer and information science*, 3(4), 104.

IBM. (2009). IBM and Partners Help Healthcare Clients Adopt Electronic Health Records and Improve Operations with Cloud Software. http://www-03.ibm.com/press/us/en/pressrelease/26963.wss. Accessed on September 1st, 2011.

IBM and ActiveHealth Management. (2010). ActiveHealth and IBM Pioneer Cloud Computing Approach to Help Doctors Deliver High Quality, Cost Effective Patient Care. http://www-03.ibm.com/press/us/en/pressrelease/32267.wss. Accessed on September 1st, 2011.

International Classification of Diseases (ICD). (2013) Available online: http://www.who.int/ classifications/icd/en/ (accessed on 11 May 2013).

Jackson, W. & Verberg, N. (2007). Methods: Doing social research (4th ed.). Toronto: Pearson.

Jeffrey, k. & Neidecker-lutz, b. (2009): The Future of Cloud Computing: Opportunities for European Cloud Computing beyond 2010; 66

Jensen, Christian, S. (2011). " Introduction to Temporal Database Research. Accessed by http://www. Cs.aau.dk/~csj/Thesis/pdf/chapter1.pdf.

Kaletsch, A.; Sunyaev, A. (2011) Privacy Engineering: Personal Health Records in Cloud Computing Environments. In Proceedings of the International Conference on Information Systems (ICIS 2011), Shanghai, China.

Kuo, M. H. (2011) Opportunities and Challenges of Cloud Computing to Improve Health Care Services, J Med Internet Res;13(3):e67-Hankel type involving products of

Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

Kuo,M.H., Kushniruk A.W. and Borycki,E.M. (2011). Can Cloud Computing Benefit Health Services? A SWOT Analysis. International Conference of the European Federation for Medical Informatics (MIE2011), Oslo, Norway, August 28-31.

Kuo, M. H., et al. (2011). National Strategies to Health Data Interoperability. Information Technology and Communications in Health (ITCH), Victoria, February 24 - 27.

Kuisak, A., Kernstine, K.H., Kern, J.A., McLaughlin, K.A., and Tseng, T.L. (2000). Data Mining: Medical and Engineering Case Studies. Proceedings of the Industrial Engineering Research 2000 Conference, Cleveland, Ohio, Pages 1-7.

Larose, D.T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining, ISBN 0-471-66657-2, John Wiley& Sons, Inc,

Lawrence Gould A. (2005). Timing of futility analyses for 'proof of concept' trials. Stat Med, 24:1815-1835.

Lenz, R.; Reichert, M. (2006) IT support for healthcare processes—Premises, challenges, perspectives. Data Knowl. Eng., 61, 39–58.

Li, M.; Yu, S.; Ren, K.; Lou, W. (2010) Securing Personal Health Records in Cloud Computing: Patient-centric and Fine-grained Data Access Control in Multi-owner Settings. In Proceedings of the 6th International ICST Conference on Security and Privacy in Communication Networks (SecureComm 2010), Singapore, pp. 89–106.

Lin, R.H. (2009). An intelligent model for liver disease diagnosis. Artificial intelligence in medicine, 47(1), 53 -62.

Lindell, Y. and Pinkas, B. (2002)“Privacy Preserving Data Mining”, Journal of Cryptology, vol 15, no 3,.

Lisboa,P.J.G. et al.(2010). *Data Mining in Cancer Research*. IEEE computational intelligence magazine, 5(1), 14 -18.

Lo, O.; Fan, L.; Buchanan, W.;Thummler,C.;Khedim, A.; et al. (2011). Twoards Testing and Validation of e-Health Infrastructure. *In Proceedings of the Pervasive Health*, Dublin, Ireland.

Lohr, H.; Sadeghi, A.; Winandy, M. (2010) Securing the E-Health Cloud. In Proceedings of the 1st ACM International Health Informatics Symposium (IHI 2010), Arlington, VA, USA,; pp. 220–229.

Luk, J.M., Lam,B.Y. et al. (2007). Artificial neural networks and decision tree model analysis of liver cancer proteomes. Biochemical and biophysical research communications, 361, 68 -73.

Luis, M. V., Luis, R.-M., et al. (2008) A break in the clouds: towards a cloud definition. SIGCOMM Comput. Commun. Rev., 39, 50-55.

Marsh, S.P. (1994) "Formalising Trust as a Computational Concept. Computing Science and Mathematics". University of Stirling.

Mcfedries, P. (2008) The Cloud is the Computer; IEEE Spectrum, www.spectrum.ieee.org/computing/hardware/the-cloud-is-the-computer, Accessed on Feb 15, 2013.

Mell, P., Grance, T. (2010). The NIST definition of cloud computing. Communication of the ACM, 53 (6),50.

Nguyen, D.K.; Lelli, F.; Papazoglou, M.P.; van den Heuvel, W.-J. (2012) Blueprinting Approach in Support of Cloud Computing. Future Internet, 4, 322–346.

Nieswiadomy, R.M. (2008). Foundations of Nursing Research (5th ed.). New Jersey: Pearson Prentice Hall.

Park, Y.J. et al,. (2008). Data mining technique for medical informatics: detecting gastric cancer using case-based reasoning and single nucleotide polymorphisms. *Expert Systems*, 25(2), 163 -172.

Pearson S. (2009) Taking account of privacy when designing cloud computing services. In: Proceedings of the ICSE Workshop on Software Engineering Challenges of Cloud Computing (CLOUD'09). New York, NY: IEEE; 2009 Presented at: the IEEE First international workshop on software engineering challenges for Cloud Computing (ICSE);; Vancouver, BC, Canada.

Pospisil, P., et al,. (2006). A combined approach to data mining of textual and structured data to identify cancer-related targets. *BMC bioinformatics*, 7, 1471-2105.

Plummer, D. C., Smith, D. M., et al. (2009): [Research] Five Refining Attributes of Public and Private Cloud;

Kanellopoulos, Y., Dimopulos, T., Tjorjis,C., Makris, C. (2007). Mining Source Code Element for Comprehending Object-Oriented Systems and Evaluating Their Maintainability. SIGKDD Explorations. 8 (1).

Klein, R. (2007). An empirical examination of patient-physician portal acceptance. Eur. J. Inf.Syst. 16, 751–761.

Oatley, G., Ewart, B.. (2011). Data Mining and Crime Analysis. John Wiley & Sons, Inc. 1,Pages 147-153.

Rajeswari, P. and Reena, G. S.(2010). Analysis of Liver Disorder Using Data Mining Algorithm. *Global Journal of Computer Science and Technology*, 10 (14), 57-61.

Rolim, C.O.; Koch, F. L.; Wetphall, C.B. (2010). A Cloud Computing Solution for Patient’s Data Collectionin Health Care Institutions. In proceeding of the Second International Conference on eHealth, Telemedicine, and Social Medicine (ETELEMED’10), St.Maarten, The Netherlands, pp.95-99.

Romero, C., Ventura, s. And De-Bra, P. (2004). Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Authors. Kluwer Academic Publisher, Printed in the Netherlands.

Saranummi, N. (2008). In the spotlight: Health information systems. IEEE Rev. Biomed. Eng., 1, 15–17.

Saranummi, N. (2009). In the spotlight: Health information systems—PHR and value based healthcare. IEEE Rev. Biomed. Eng. 2, 15–17.

Saranummi, N. (2011). In the spotlight: Health information systems—Mainstreaming mHealth. IEEE Rev. Biomed. Eng., 4, 17–19.

Sharp Community Medical Group. (2008). DiskAgent Launches New Remote Backup and Loss Protection Software as a Service Offering. https://www.diskagent.com/press_releases/remote_backup.php. Accessed on September 1st, 2011.

Shen, Chia-Ping, et al. (2011). A Data-Mining Framework for Transnational Healthcare System. J Med Syst. DOI 10. 1007/s 10916-11-9729-7.

Shimrat, O. Cloud Computing and Healthcare. Available online: http://www.himss.org/content/ files/Code%2093_Shimrat_CloudComputingandHealthcare_2009.pdf (accessed on 2 Febuary 2013).

Solieman, O.K.(2006). Data Mining in Sports: A Research Overview, A Technical Report, MIS Masters Project, August 2006. Web site accessed in 2011: http://ai.arizona.edu/hchen/chencourse/Osama-DM_in_Sports.pdf

Teng, C.C.; Mitchell, J.; Walker, C. (2010). A Medical Image Archive Solution in the Cloud. In Proceedings of the 2010 IEEE International Conference on Software Engineering and Service Sciences (ICSESS), Beijing, China,; pp. 431–434.

Tu, S.W. et al. (2007). The SAGE Guideline Model: Achievements and Overview. *Journal of the American Medical Informatics Association*, 14(5), 589–598.

U.S. Department of Health & Human Services. (2003) Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule. Available online: http://privacyruleandresearch.nih.gov/pdf/HIPAA_Booklet_4-14-2003.pdf (accessed on February 16, 2013).

Vasilakos, A.V.; Lisetti, C. (2010). Special section on affective and pervasive computing for healthcare. IEEE Trans. Inf. Technol. Biomed., 14, 183–359.

Vaquero LM, Rodero-Merino L, Caceres J, Lindner M. (2008). A break in the clouds: towards a cloud definition. ACM SIGCOMM Comput Commun Rev;39(1):50-55.

Vouk, M. (2008) Cloud Computing - Issues, Research and Implementations. Journal of Computing and Information Technology.

Williams, A. (2010) Top 5 Cloud Outages of the Past Two Years: Lessons Learned; ReadWriteWeb, http://www.readwriteweb.com/cloud/2010/02/top-5-cloud-outagesof-the-pas.php, 06/07/2010.

Wikipedia. (2013). http://en.wikipedia.org/wiki/Prototype.Accessed April 28,2013.

Wright, Adam and Sittig, D. F. (2006). Automated Development of Order Sets and Corollary Orders by Data Mining in an Ambulatory Computerized Physician Order Entry System. *AMIA Annual Symp. Proc*. 819–823.

Yu, S.; Wang, C.; Ren, K.; Lou, W. (2010) Achieving secure, scalable and fine-grained data access control in cloud computing. In Proceedings of INFOCOM 2010, San Diego, CA, USA, pp. 1–9.