8-5-2011

Decision Modeling for Healthcare Enterprise IT Architecture Utilizing Cloud Computing

Charles Brust
Mayo Clinic, brust.charles@mayo.edu

Surendra Sarnikar
Dakota State University, ssarnikar@outlook.com

Decision Modeling for Healthcare Enterprise IT Architecture Utilizing Cloud Computing

Charles S. Brust
Mayo Clinic
brust.charles@mayo.edu

Surendra Sarnikar
Dakota State University
surendra.sarnikar@dsu.edu

ABSTRACT

In this paper, we present an overview of cloud computing, examine the potential uses for cloud computing in healthcare environments, and propose a framework to guide architectural selection decisions regarding information systems in both large and small healthcare organizations. The framework provides insight to both practitioners and academics by extending our understanding of the decisions regarding computing architectures within the healthcare system.

Keywords

Decision model, cloud computing, healthcare IT.

INTRODUCTION

Information and knowledge intensive organizations typically have heavy data and application infrastructure needs that vary significantly with market conditions and technology changes. In order to satisfy their data, and application requirements, many knowledge intensive organizations are increasingly exploring cloud computing models to efficiently meet such needs (Kim, 2009; Motahari-Nezhad et al., 2009). Since healthcare organizations are also knowledge intensive organizations and have similar problems of varying data and application needs that change with market demand and and technology changes, cloud computing based enterprise IT infrastructure models could potentially be of major benefit to healthcare organizations by lowering information technology costs while improving the availability and reliability of applications. However, given complexity of healthcare operations and the unique operating environment of healthcare organizations, a decision model is required to explore optimal enterprise IT model configurations and the suitability of cloud computing models for various healthcare organizations and operations.

Several issues exist in determining the suitability and effectiveness of cloud computing models for healthcare enterprise IT environments. First, there is inadequate clarity on the different types of cloud computing models and their suitability to various healthcare processes. Moreover, the specific factors that influence the design requirements of healthcare enterprise Information Technology architectures are not well understood. There is also limited literature that given IT architecture requirements, provides a decision model for analyzing and designing appropriate IT infrastructures for healthcare organizations. Healthcare organizations operate in a unique environment that includes strong regulatory requirements governing data sharing, storage and privacy safeguards, healthcare service provisioning requirements and standards, variety of applications that vary from data intensive applications such as genomic applications to patient critical systems such as operating room and emergency room information systems. Given the complexity and variety of requirements in the healthcare environment, a decision model is required to help the design of healthcare enterprise IT infrastructures that satisfy a wide variety of criteria while lowering costs.

This paper presents an extensive review of cloud computing models, specifically with a focus on the application of the technology in healthcare organizations. Based on the extensive review, we develop a model that will allow healthcare enterprises to make informed choices in the implementation of clouds where they are appropriate, and to consider all the criteria and possible options when making those selections. We present some sample examples through a series of simulations of cloud computing installations to illustrate a potential decision making process and conclude with an agenda for future research.

BACKGROUND AND LITERATURE REVIEW

Cloud Computing

Definition of Cloud Computing

There are many definitions of cloud computing, with each variation focusing on a different aspect of the technology. Some of these include "infrastructure from which businesses and users are able to access applications from anywhere in the world on demand" (Buyya, Yeo, Venugopal, Broberg and Brandic, 2009), "A cloud is a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service level agreements established through negotiation between the service provider and consumers" (Buyya, Yeo and Venugopal, 2008), "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 customized SLAs" (Vaquero, Rodero-Merino, Caceres and Lindner, 2008), and "being able to access files, data, programs and 3rd party services from a Web browser via the Internet that are hosted by a 3rd party provider" (Kim, 2009). In short, cloud computing can be equated to a utility – one taps into the service, just like a house taps into the electric grid or local water system, and the product (in this case computing capacity) is available to meet the level of the customer's needs. It is evident that there is a central theme through the definitions, and as such, for the purposes of this paper, we will use the following definition: "Cloud computing is the architecture by which customers may receive computing capacity in a utility fashion, allowing elasticity in demand to drive the cost and availability of the resource". By using such a definition, we allow for a relatively wide group of options to be included in the design, but at the same time we eliminate some of the more fringe definitions from consideration.

Public Clouds

The most common design for cloud computing is a public cloud. In this architecture, a customer is able to attach to remote computing resources which are completely controlled by a third party entity (Motahari-Nezhad, Stephenson and Singhal, 2009). As an example, Amazon provides their EC2 service, which allows users to provision infinite numbers of virtual servers at a fixed cost of $0.10 per server per hour (Armbrust, Fox, Griffith, Joseph, Katz, Konwinski, Lee, Patterson, Rabkin and Stoica, 2010). With this model, elasticity is built into the model, as systems can be added and removed as demand for computing resources dictates. For example, an electronics retailer may normally need 50 web servers to host their online presence, but due to a sale on big-screen televisions during Super Bowl week, their needs might spike to 500 servers. In the public cloud, these servers can be brought online in a matter of seconds, and can be deprovisioned just as quickly when demand goes down. In this scenario, the retailer might even reduce their server count overnight each evening as the volume of web traffic declines. It is conceivable that the retailer could host this traffic on in-house servers, but to do so would mean a large number of unneeded servers in the data center most of the time. Such a situation is wasteful of power and cooling resources, as well as requiring larger overhead from an administration standpoint.

The architecture of a public cloud system includes an internet connection and a large remote data center. In that data center there are some (usually large) number of servers running a virtualization platform such as VMware, by which each of the physical hosts can be divided into many virtual server instances for customer use. These computing resources are then assigned to customers based on demand – the provisioning may be done at the application, server, or platform level (see SaaS, PaaS, and IaaS section below). The customer will have a service level agreement (SLA) in place with the vendor to determine when servers should be added or removed from use based on load and/or user experience (Dikaiakos, Katsaros, Mehra, Pallis and Vakali, 2009). In the web services example given above, the SLA may call for auto-provisioning of additional servers when the load reaches 80% of capacity, and deprovisioning when load drops below 20%. In this way, there should always be sufficient computing power to service the users, while allowing the retailer to save costs by removing unneeded capacity.

Private Clouds

Server virtualization is not only the purview of cloud vendors like Amazon or Microsoft. In fact, virtualization is a fast growing technology across the industry due to its benefits and low cost to implement. As such, many IT shops already have the basic parts of cloud computing running in their environments, though without the automation tools that cloud computing includes. In some cases, large companies such as Boeing have taken the architectural ideas around cloud and brought the additional automation technology in-house rather than contracting with a hosting company (Motahari-Nezhadet al. , 2009). This self-hosting of cloud computing has been termed a private cloud (Rimal, Choi and Lumb, 2009). By bringing all the infrastructure components in-house, the corporation is able to mitigate some of the concerns that exist with public cloud

installations, including security and network bandwidth (Sotomayor, Montero, Llorente and Foster, 2009). While this may seem like the ideal solution, it must be noted that though there are benefits to a private cloud, there are also downfalls, including increased costs for infrastructure, the potential for reduced server utilization, and increased complexity which may lead to the need for additional administrative staff.

Other cloud computing architectures

Public and private clouds represent the two extremes in cloud computing architecture. In many cases, there are benefits to utilizing a mixture of these ideas. This brings the concept of hybrid clouds, and of federated clouds. Hybrid clouds are those installations which use a combination of public and private cloud architecture in their installations (Rimalet al. , 2009). To better understand the concept of hybrid clouds, consider the following example: An enterprise has chosen, for security reasons, to install a private cloud to host their entire website infrastructure (web servers, database, etc.). Occasionally, this website has traffic spikes that exceed the capabilities of the private cloud. In this instance, the organization could choose to bring up web servers through a public cloud provider, and at the same time, reduce the internal web server count in order to increase database computing power (Zhang, Jiang, Yoshihira, Chen and Saxena, 2009). In this way, the organization retains its security benefits while at the same time being able to respond to user demand. The only public cloud costs come during the time where those resources are being used, and the consumers remain happy with the levels of service and performance that the website exhibits.

Another possibility for mixed cloud installations involves the use of multiple public cloud providers – in that scenario, a company application may actually reside on servers from both Amazon and Microsoft, or any cloud host, due to restrictions on the cloud model (SaaS, PaaS, IaaS) that the host provides. Using our previous website example, one might find that the web servers are best hosted in Microsoft's Azure cloud due to the .NET nature of the application, while the databases must go in Amazon's EC2 because Oracle on Linux is the chosen direction. In that case, a federated cloud is created across Microsoft and Amazon to host the application (Rochwerger, Breitgand, Levy, Galis, Nagin, Llorente, Montero, Wolfsthal, Elmroth and Caceres, 2010). With cloud federation, a company can choose to have services hosted on the best-of-breed provider, while still retaining interoperability through the application stack.

Finally, the notion of a private cloud, hosted publicly has recently been proposed (Krautheim, 2009; Wood, Gerber, Ramakrishnan, Shenoy and Van der Merwe, 2009) as a more cost-effective option such as would be found in public clouds, but retaining the security functionality of private clouds. This virtual private cloud architecture allows a company to have a secure Virtual Private Network (VPN) connection to the cloud provider, and to have a dedicated, yet flexible set of hardware on which to run their virtual servers. In this way, there is separation on both the physical and virtual levels for corporate data, and the security function is handled exactly the same as if the data were hosted in the enterprise's local facilities. The scaling benefits are in place as well, since the customer can add and remove servers as needed to host the applications, and charges for virtual servers are metered by hourly rates just as with a standard public cloud, albeit the rate is likely higher for this specialized service.

SaaS, PaaS, and IaaS models

Within the various architectures of cloud computing, we also find varying service offerings. These take the form of choices by which the computing power is presented. They include Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS) (Wang, Tao, Kunze, Castellanos, Kramer and Karl, 2008). In a SaaS cloud, the connection is to a software application. All the servers in that cloud run the application, though there may be separate instances for separate customers. In this model, the end result is similar to server-based applications outside the cloud: customers connect to a given location, and the application is run on the server CPU. Benefits of this model include that the cloud provider is responsible for the licensing and maintenance of the application, however this model can be less flexible than others since there will not typically be variation within the application to allow end-user customization.

PaaS clouds provide a blank virtual server to the customer, and that customer is responsible for installing and licensing any applications that are hosted there (Grossman, 2009). Amazon's EC2 offering is the classic example of this model, whereas the service provides standard-configuration Linux based servers in whatever quantity is needed, and those servers are loaded with applications and deployed for production use. There are many tools that allow automation of the application installation and other processes, and as such, PaaS services can be just as responsive as SaaS clouds when it comes to elasticity of computing power. This model is much more flexible when viewed from an application perspective, as the customer has full control over the application install and feature customization, but with that flexibility also comes the burden of maintaining software licensure for the application.

IaaS takes another step back in the configuration: the service provider for this model simply presents blank virtual servers (or sometimes disk), and the customer loads both the operating system and application (Bhardwaj, Jain and Jain, 2010). Just as

in PaaS, this adds flexibility for the customer, but with the administrative and licensing burdens that go along with it. This can be beneficial to customers that require multiple operating systems, or who wish to have multiple virtual hardware configurations.

A new concept (Wanget al. , 2008) that has direct ties to IaaS is that of Data-as-a-Service (DaaS). This method is actually more of a data sharing service than a typical cloud – it has roots in research, where aggregate databases of information, such as medical records, could be stored in a DaaS cloud, and all the data would be available to subscribing customers for research purposes.

Cloud Computing in Healthcare

Today's healthcare environment is IT-intense, with healthcare providers relying on data from many different sources to make patient care decisions and recommendations. As the requirements for data processing increase, healthcare enterprises may find that it is not cost effective, or even possible to host all the computing power necessary for a modern practice internally. Further, cost drivers dictate that utilization of those resources must be at a high level to maintain efficiency. The need for reliable, scalable infrastructure solutions continues to increase in the healthcare arena, and the ability to meet these needs in a constrained budget will allow the necessary expansion of technology solutions to continue. Cloud computing offers a lower-cost option to computing resources, though not without possible concerns. If those concerns can be addressed through proper architecture, cloud computing has the potential to be a positive element in future healthcare IT projects.

Potential uses

The possible uses for cloud computing in healthcare are significant and widespread. The architecture has already been proven in other business arenas for standard functions such as finance, web services, etc., but healthcare organizations generally have not yet implemented cloud for patient care applications to date. There are several possibilities that stand out as early adoption options, due to their varied computing needs, and the potential for large blocks of compute time being consumed (Keahey, Figueiredo, Fortes, Freeman and Tsugawa, 2008). Among these are simulations, home health care, patient safety, electronic medical records, imaging, genomics, and bioinformatics. While this certainly is not a comprehensive list, it does demonstrate the wide reaching functionality that the cloud could provide.

Issues and Concerns with Cloud Computing in Healthcare Environments

Though there are many benefits to the use of cloud computing architectures in healthcare, there are also some barriers to wholesale integration, especially in public cloud scenarios. Chief among the concerns is security, and this comes on three fronts: first, the fact that data will not stay inside the corporate datacenter and is thus subject to outside parties attempting to intercept it; second, that a third-party is involved in the data processing, and that company must abide by HIPAA and other patient privacy regulations (Takabi, Joshi and Ahn, 2010). Finally, patient criticality must also be examined, and processes that are directly in the patient care critical path may not be the best options, as a provider outage at the cloud level infrastructure could cause an emergency in the clinical facility. Chief among these outages would be connectivity to the cloud – a simple backhoe cable cut might render the cloud based applications unreachable (Leavitt, 2009). Similarly, an outage at the hosting datacenter could have similar outcomes. These issues must be considered when planning a cloud architecture.

A DECISION MODEL FOR HEALTHCARE CLOUD COMPUTING

With the vast array of options that are possible in a cloud computing architecture, it is easy to be overwhelmed in the decision-making process. This complexity, coupled with the wide variety of healthcare applications that might be hosted by the cloud technology makes architectural decisions nearly impossible without guidance. To that end, we propose a decision model based on the Analytic Hierarchy Process (AHP) (Saaty, 1990) which will be used to assist enterprises in selecting the optimal cloud architecture for their environment. This model consists of an architecture selection procedure, coupled with an AHP model to assist in matching cloud characteristics to requirements for medical applications. This model has been extended from a previous AHP-based product selection framework paper (Wei, Chien and Wang, 2005), molding it for use in healthcare environments.

Requirements Framework for Healthcare Cloud Computing

In order to make sound decisions around the cloud architecture, all the varying contributing factors must be understood and taken into account. For healthcare organizations, these factors include costs, capacity on demand, security, regulatory concerns, and patient criticality. Each of these will play some part in the overall architectural decision, and in some cases, a single element can cause a potential solution to be thrown out completely.

Cost

Cost is a factor in any IT project, not just cloud computing. Here though, there are many varying options when it comes to cost that it bears close examination (Klems, Nimis and Tai, 2009). In a public cloud, one is only paying for the server time that is used, and the rate per server-hour seems low on initial examination. At a rate of $0.10/server-hour, a company is paying $876 per server, per year assuming that the server is always on. Depending on the cost for hosting the same server in-house, this may or may not be reasonable. Private clouds, on the other hand, do not have a per server-hour cost, however there is significant capital expense outlay at the time that servers are purchased. The addition of power, cooling, and administrative resources push the cost up even further, and we must also factor in that a company using private cloud will need to over-buy their infrastructure, so as to be able to handle on-demand elasticity requirements. In the hybrid cloud scenario, a company may be more cost-efficient if their normal load is known – in that case, the private cloud can be configured to host normal traffic, and the public cloud can be used for overflow.

Capacity on Demand

The concept of capacity on demand is the major highlight of cloud infrastructure. The idea that one can count on virtually unlimited server numbers and computing power drives many companies toward cloud initially. What may not get considered is that there is significant planning and resource time that must go into the planning for, and eventual maintenance of, such an infrastructure (Khajeh-Hosseini, Greenwood and Sommerville, 2010). If using public clouds, the applications must be written or modified to meet the standard deployment from the vendor. Further, negotiations for a service level agreement must take place, and once concluded, automation tools must be put into effect which will do the addition / subtraction of servers to meet that need. An enterprise that cannot or does not wish to dedicate internal IT staff to such tasks can purchase professional services from the hosting company, however this will add to the overall project cost. In private cloud installations, the same overhead costs for administration are present, along with infrastructure expenses. In that case however, there are a maximum number of servers that can be deployed – this is directly related to the capabilities of the infrastructure that has been put in place in the local datacenters. It will take much more careful planning to implement a private cloud that is capable of the elasticity needed within an enterprise without causing significant excess power, cooling, and hardware costs.

Security

As with any new technology, there are inherent technological problems in cloud computing that have not yet been fully addressed. Security is the chief concern with cloud computing, since the scenario of shared-everything is in use, coupled with a third-party that hosts all the services and does not necessarily provide complete transparency into their internal procedures (Brodkin, 2008). The popularity of cloud computing also creates a high profile target for hackers, who view the cloud as an unlimited source of computing power, if only they can slip through the back door undetected (Chen, Paxson and Katz, 2010). While several novel solutions have been proposed (Haeberlen, 2010; Jensen, Schwenk, Gruschka and Iacono, 2009; Wang, Wang, Ren and Lou, 2009, 2010), full implementation has not yet been achieved, and the consumer is at risk in the public cloud environment. Certainly mitigation strategies can be put in place, including hybrid clouds which keep critical data protected, or federation of clouds, spreading the load across multiple vendors. The highest safety, however, comes in the private cloud, since the infrastructure is fully controlled by the customer and can be secured and audited to whatever standard is needed.

Regulatory Concerns

The healthcare industry is governed by HIPAA, which states that all patient-identifiable data must be kept confidential and cannot be accessed, intentionally or accidentally, by parties other than the care team, unless explicitly approved by the patient. This could possibly be cause for concern, as in public clouds, the data is stored at a third-party site, where it is intermingled on servers and storage hosting services for other customers. A strict reading of HIPAA law could be interpreted as this being outside the regulations, even if there is a HIPAA agreement between the healthcare organization and the cloud hosting vendor (Osterhaus, 2010). If identifiable information is stripped from the records before being stored in the cloud, the concerns are significantly reduced; however this does then add complexity to the application, as the patient records must be reattached before the information is usable in a clinical setting.

Patient Criticality

Patient criticality measures the extent that the application being served by the cloud is involved in the direct care of the patient – this may range from an application that is directly in the clinical decision making process, to one that is strictly used for billing or other administrative purposes. The more critical the application is to care, the more caution that must be used in choosing to move to a cloud platform. This is especially true in the case of public clouds, where the infrastructure is not

under the control of the enterprise, thus involving third parties in the process of prioritization and correction of the issue. It is imperative that the service level agreements in these installations be explicit on times and priorities of problem correction.

Mix of Applications

Since healthcare IT enterprises are expected to support a wide range of applications, it must be noted that there could be several programs running in cloud services simultaneously. These may reside in a single cloud, in multiple isolated clouds, or in some sort of a federated cloud environment. At this time, the proposed model does not directly address optimization for multiple concurrent applications; however it will be extended in the future for this purpose.

Decision Model

Medical applications described above have varying technological needs. Table 1 summarizes the high-level technology requirements and capabilities for the various medical applications of cloud computing technology. For purposes of clarity, the classifications denote the relative amount of resources that an application would need to run properly in a mid-size, production medical environment. These thresholds may change depending on the specifics of the installation and the needs of the institution. As an example, genomics applications are ever-changing, as different report styles, etc. are created, whereas EMR databases will see little change over time with regard to the actual application. The cost per work unit measure in this context examines how fiscally valuable the data from an application may be to a medical institution, and to what level they may expect to see infrastructure investments take in order to implement that application. This may be equated to the amount of time that is required to see a return on investment (ROI) for the infrastructure.

	Simulation	**Home Health Care**	**Patient Safety**	**EMR**	**Imaging**	**Genomics**	**Bioinformatics**
Cost per Work Unit	Medium	Low	Low	Medium	High	Low	Low
Computing Capacity	High	Low	Low	Medium	High	High	High
Security	Low	Medium	Low	High	Medium	Medium	Low
Regulatory Concerns	Low	Low	Medium	High	High	High	Medium
Patient Criticality	Low	Low - Medium	Medium	High	High	Low	Low

Table 1: Need Thresholds for Various Medical Applications

Cloud Architecture Selection Framework

The purpose of this framework is to allow healthcare organizations to optimize their cloud computing architecture choices. The complexity of the cloud architecture can cause issues if poor choices are made at implementation, and these architectural decisions can mean limited performance or utility in the cloud in the future. The aim of this framework then is to guide these decisions so that each implementation can achieve its full potential. The specific steps within the cloud selection procedure are as follows:

Step 1: Identify the application characteristics and the available IT resources

Step 2: Determine the business needs for the application

Step 3: Match the application attributes to the known properties of the cloud model

Step 4: Remove nonviable cloud architectures from consideration

Step 5: Evaluate the cloud using the AHP model

Step 6: Determine final architecture based on the results obtained

Figure 1 shows the flowchart of the cloud architecture selection process. The comprehensive explanation of each criterion is presented in the context of the simulation detailed in the next section.

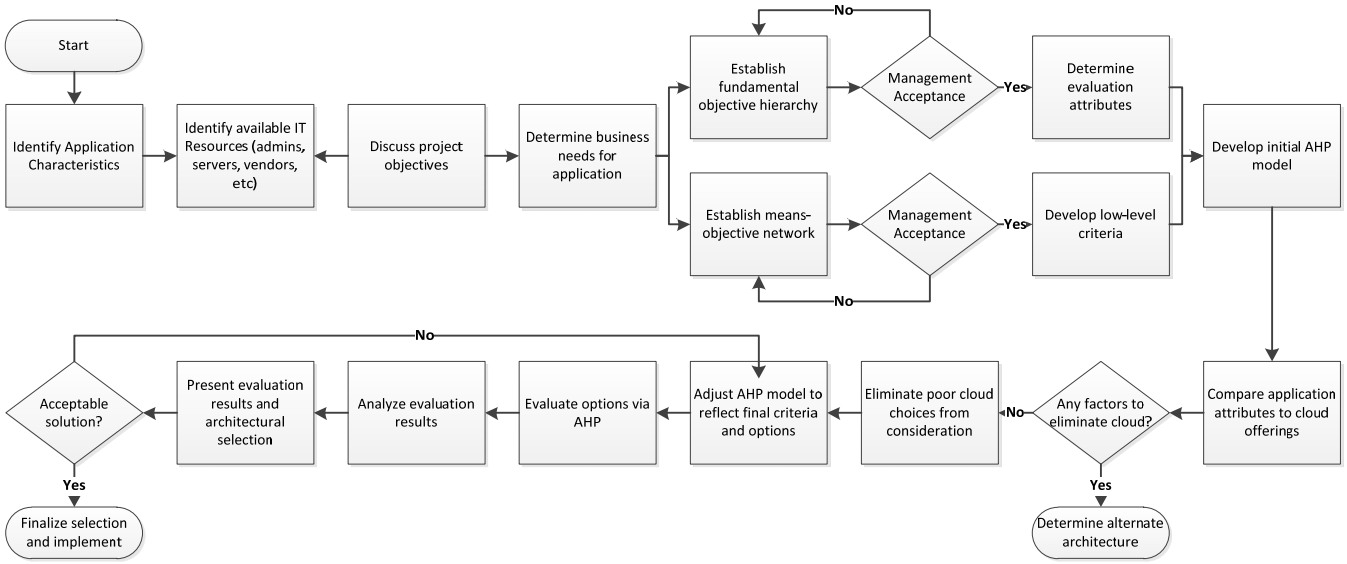

Figure 1: Comprehensive Cloud Architecture Selection Framework

Supporting AHP Model

In order to satisfy the requirements of such a diverse set of applications typically run in a large healthcare organization, it is necessary to identify a cloud solution that is optimized to handle the requirements for each application. This may mean that multiple clouds are needed within an organization, or that federation of clouds is used, depending on the application load. We propose the use of an AHP decision model, which factors all the various criteria that are involved with the decision making process for cloud architecture.

The AHP model allows for each enterprise to assign weights to the criteria at each level via a pair-wise comparison. In this way, the model can be specifically tuned to the factors in play at each institution, and as such, create the optimal solution for the individual situation.

To arrive at a technology decision, users of this model need to step through the pair-wise comparisons for each of the criteria, paying careful attention to ensuring that the more important criteria for the enterprise - whether they are cost, patient criticality, security, etc. - are judged as such. With that information, the model can then return a recommendation of the optimal architecture for a given application.

Decision Criteria

Before making any architectural decisions, the enterprise must examine the type of application being proposed. This will assist the project team in making some initial determinations around the cloud model being used (i.e. public vs. private). This is potentially overridden by the available funding for the project (as described by the cost per work unit) – if this funding is not available at a sufficient level to meet the technically optimal architecture, then it may be possible to continue to deploy under alternate models if the enterprise is willing to accept any additional risks associated with such a change.

Cost per work unit assigns a cost to a measurable standardized unit. In the case of grids, this is the server-hour (one virtual server running for 1 hour). This is the billing unit used by public cloud vendors, and can be used to make determinations on the value of private, hybrid, or federated clouds as well, since in those scenarios, the total cost of private infrastructure, including hardware, software, power, cooling, building space, and administrators can be measured against the total capacity of the cloud and thus a theoretical server-hour cost can be deduced.

Computing capacity criteria examine the elasticity and flexibility of the application and cloud model. If using a fully public cloud, the capacity is nearly unlimited, while in hybrid or private installations, there may be less computing power available at a given moment to dedicate to the application. It must be noted that capacity and cost are two sides of the same equation – an increase on one side will mean an increase on the other, assuming all other factors remaining equal.

Security measures the ability of a cloud customer to keep their data safe from outside parties. In the healthcare field, the security of some (though not all) data is paramount to the institution, and an insecure cloud model could cause cloud technology to be eliminated as an option for the IT project. On the other hand, healthcare also has some data that does not

require the same security levels – for instance, web services normally don't have the same security level as patient data. These applications with lower security requirements may not take issue with the cloud offering.

Regulatory concerns and security, much like cost and capacity, are strongly linked criteria. In this case, the regulations such as HIPAA are one of the drivers toward a more secure environment for patient data. Since the cloud technology has not been fully vetted as HIPAA compliant in all areas, institutions need to take this status into account during decision making.

Patient Criticality criteria examine the proximity of the application to the line-of-care. This criterion exists to address the opinion that technology should never get in the way of the practice of medicine, but rather should complement and enhance that practice. This criterion and its sub-factors allow an enterprise to weight the relationship of technology and patient care, and further to allow model users to balance the use of applications across multiple usage scenarios.

Illustrative Examples

In order to demonstrate the viability of the proposed model, we have used CloudSim (Buyya, Pandey and Vecchiola, 2009; Calheiros, Ranjan, De Rose and Buyya, 2009) tools to simulate various possible small-scale cloud implementations. These have been compared against simulated selections from the AHP model to assess the recommendations given against the results of the simulated cloud. For all simulations, a standard of $0.10 per server-hour was used to calculate costs for public cloud infrastructure, and a standard of $0.15 per server-hour cost for private infrastructure. The higher cost for private infrastructure assumes a 50% overall utilization of the infrastructure over a 5 year period, and includes hardware, software, power & cooling, and administrative costs ($6000 for infrastructure, plus $12000 per year for other costs, and there are 10 virtual servers hosted on average).

Scenario 1 - Web services in a small institution

This scenario encompasses 2 virtual servers running on 1 host. The servers are running a website application for intranet use, so security is not a significant factor. Application load is relatively static, since the nursing staff uses the application around the clock to monitor the load of patients vs. emergency room beds. No patient-identifiable data is utilized, so HIPAA regulations do not come into play. The AHP model recommends public cloud implementation for this application, and we expect that costs will be reduced by doing so.

Public Cloud 71.2% system utilization – no additional virtual systems needed = $0.20 / hour total cost = $1752 / year.

Private Cloud 71.2% system utilization – no additional virtual systems needed = $0.30 / hour total cost = $2628 / year.

Scenario 2 – EMR Application with Database

In this application, we host an Electronic Medical Record application, which has an application component and a database. Like scenario 1, the application is only utilized inside the institution, but in this case, the EMR contains significant amounts of patient identifiable information, and thus the option for public cloud is not available due to security and regulatory concerns. In working through the model, the private cloud is suggested for this application (though a non-cloud implementation was close behind) – the difference being the flexibility provided in the cloud for instances where more users are using the application, and the elasticity that the cloud gives in adding more virtual servers to support those users.

Public cloud – not analyzed due to security requirements.

Private cloud 224.8% original system specification utilization (2 application servers, 2 database servers) – no additional virtual application servers needed, however database servers were doubled during simulation = $0.90 / hour total cost = $7884 / year.

Scenario 3 – Patient Health Record

With this final scenario, we look at a hybrid cloud solution, where web services are hosted publically and data is stored in a private cloud. The PHR model is a method by which patients can centralize their health records through uploads of data to a central storage application. Since this is patient-identifiable data, we are bound by HIPAA to secure it from any possible unauthorized release. As such, it cannot be stored in the public cloud, though the web servers can reside there. Our AHP model shows a hybrid solution as being the best balance between cost and security.

Web server portion

Public cloud 35.6% system utilization – no additional virtual systems needed = $0.20 / hour total cost = $1752 / year, however this could have been reduced from 2 servers to 1, creating $0.10 / hour total cost = $876 / year.

Private cloud 35.6% system utilization – no additional virtual systems needed = $0.30 / hour total cost = $2628 / year, however this could have been reduced from 2 servers to 1, creating $0.15 / hour total cost = $1314 / year.

Data Storage / Database portion

Public cloud 57.2% system utilization – no additional virtual systems needed = $0.30 / hour total cost = $2628 / year.

Private cloud 57.2% system utilization – no additional virtual systems needed = $0.45 / hour total cost = $3942 / year.

Total cost / year

All private cloud: Web = $1314 + Database = $3942 = $5256 / year

Hybrid (public web, private database) = Web = $876 + Database = $3942 = $4818 / year

DISCUSSION

Cloud computing is a viable technology in healthcare, so long as the proper precautions are followed around data security and regulatory compliance. These current restrictions to usage mean that in many cases, healthcare enterprises will be required to utilize private or hybrid clouds to achieve both security and elasticity in infrastructure. By allowing the data that requires higher security to reside locally, or in a virtual private cloud (VPC), the concerns around regulation and data privacy are allayed, and the enterprise is still free to utilize public cloud infrastructure for services that do not require strict security. This hybrid or VPC approach may return a lower cost for the total computing requirements of the enterprise.

Through the illustrative examples, we demonstrate that a decision model can be evaluated to explore whether a cloud computing model will allow healthcare organizations to meet their ever-increasing computing capacity requirements, while still reducing cost for the infrastructure. In our sample simulations, we have illustrated simple cases which allows healthcare enterprises to project capacity for internal and external clouds, or any combination thereof, with an added benefit of giving expected cost comparisons for the various scenarios.

Our initial simulation experiments have shown that the proposed model has the capability of analyzing and recommending the best-fit architecture for a given application. While these simulations are limited in nature, they are sufficient to prove the concept of the model. Further case studies will certainly be necessary to extend the model and add to its rigor over time; however the base viability of the model's use has been established.

FUTURE RESEARCH

With the cloud computing concept being a relatively recent innovation in technology, there is significant opportunity for research in the arena of cloud computing in general, and more specifically, the healthcare focus for cloud computing is only beginning to be defined. As noted in the literature review, many of the papers that discuss cloud computing have been written in the last two years. In the near term, case studies of pilot or production implementations of cloud technologies would assist in gaining a better view of the initial uses of this concept in healthcare, thus allowing the discipline to grow organically. This is especially true in the security arena, since the existing articles give few solutions to the stated issues, and those that do suggest solutions have not yet been implemented in trials or case studies. Further, additional parallels between cloud and other enterprise architecture options such as grid computing need to be explored in more depth, as there may be concepts that can be lifted from one technology to another.

As cloud is being introduced into healthcare organizations, additional potential areas of research include:

- User acceptance – will physicians and other healthcare providers use a system in the cloud equally to one hosted locally? Does this acceptance have any correlation to the technology, or is it strictly a perception issue?
- Expansion of use – the initial areas listed above as possible use cases for cloud in healthcare are by no means a comprehensive list. As the technology is implemented, case studies looking at new usage scenarios would be beneficial.
- What can be done to mitigate the security risks that are currently perceived with cloud in regards to healthcare data? Novel solutions to the security issues would apply not only to healthcare, but to cloud implementations in general.
- Are solutions possible that will eliminate the regulatory concerns? This may be a legal issue rather than a technological problem, however allowing public cloud utilization for patient-identifiable information is paramount to large-scope acceptance of the technology in healthcare.

- Are there major benefits to utilizing private cloud computing vs. simply running applications on virtual machines within the data center? In other words, does the added complexity of the cloud software layer add benefit to the implementation? Many healthcare systems are already heavily invested in virtualization software today, and the addition of cloud software needs to be examined from this angle.

REFERENCES

1. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., and Stoica, I., (2010) A view of cloud computing, *Communications of the ACM*. 53(4): p. 50-58.
2. Bhardwaj, S., Jain, L., and Jain, S., (2010) Cloud computing: A study of infrastructure as a service (IAAS), *International Journal of engineering and information Technology*. 2(1): p. 60-63.
3. Brodkin, J., (2008) Gartner: Seven cloud-computing security risks, *Infoworld*: p. 1–3.
4. Buyya, R., Pandey, S., and Vecchiola, C., (2009) Cloudbus toolkit for market-oriented cloud computing, *Cloud Computing*: p. 24-44.
5. Buyya, R., Yeo, C.S., and Venugopal, S. (2008) Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities. *Proceedings of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid* May 18 - 21 Shanghai, China.
6. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., and Brandic, I., (2009) Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility, *Future Generation Computer Systems*. 25(6): p. 599-616.
7. Calheiros, R.N., Ranjan, R., De Rose, C.A.F., and Buyya, R., (2009) Cloudsim: a novel framework for modeling and simulation of cloud computing infrastructures and services, *Computing Research Repository, vol. abs/0903.2525*.
8. Chen, Y., Paxson, V., and Katz, R., (2010) What's New About Cloud Computing Security, *University of California, Berkeley Report No. UCB/EECS-2010-5 January*. 20(2010): p. 2010-5.
9. Dikaiakos, M.D., Katsaros, D., Mehra, P., Pallis, G., and Vakali, A., (2009) Cloud computing: Distributed internet computing for it and scientific research, *Internet Computing, IEEE*. 13(5): p. 10-13.
10. Grossman, R.L., (2009) The case for cloud computing, *IT Professional*. 11(2): p. 23-27.
11. Haeberlen, A., (2010) A case for the accountable cloud, *ACM SIGOPS Operating Systems Review*. 44(2): p. 52-57.
12. Jensen, M., Schwenk, J., Gruschka, N., and Iacono, L.L. On technical security issues in cloud computing.*Proceedings of the 2009 IEEE International Conference on Cloud Computing*, Sep 21-25, Bangalore, India.
13. Keahey, K., Figueiredo, R., Fortes, J., Freeman, T., and Tsugawa, M., (2008) Science clouds: Early experiences in cloud computing for scientific applications, *Cloud Computing and Applications*. 2008.
14. Khajeh-Hosseini, A., Greenwood, D., and Sommerville, I. Cloud Migration: A Case Study of Migrating an Enterprise IT System to IaaS. *Proceedings of the IEEE 3rd International Conference on Cloud Computing*, July 5-10, Miami, FL, USA.
15. Kim, W., (2009) Cloud computing: Today and tomorrow, *Journal of object technology*. 8(1): p. 65-72.
16. Klems, M., Nimis, J., and Tai, S., (2009) Do clouds compute? a framework for estimating the value of cloud computing, *Designing E-Business Systems. Markets, Services, and Networks*: p. 110-123.
17. Krautheim, F.J. *Private virtual infrastructure for cloud computing*. 2009: USENIX Association.
18. Leavitt, N., (2009) Is cloud computing really ready for prime time?, *Growth*. 27: p. 5.
19. Motahari-Nezhad, H.R., Stephenson, B., and Singhal, S., (2009) Outsourcing Business to Cloud Computing Services: Opportunities and Challenges, *LABs of HP*.
20. Osterhaus, L.C., (2010) Cloud Computing and Health Information.
21. Rimal, B.P., Choi, E., and Lumb, I. A taxonomy and survey of cloud computing systems. *Proceedings of the Fifth International Joint Conference on INC, IMS and IDC*, Aug 25-27, Seol, South Korea.
22. Rochwerger, B., Breitgand, D., Levy, E., Galis, A., Nagin, K., Llorente, I.M., Montero, R., Wolfsthal, Y., Elmroth, E., and Caceres, J., (2010) The reservoir model and architecture for open federated cloud computing, *IBM Journal of Research and Development*. 53(4): p. 4.
23. Saaty, T., (1990) Multicriteria decision making: the analytic hierarchy process: planning, priority setting, resource allocation. RWS Publications Pittsburgh.
24. Sotomayor, B., Montero, R.S., Llorente, I.M., and Foster, I., (2009) Virtual infrastructure management in private and hybrid clouds, *IEEE Internet Computing*. 13(5): p. 14-22.
25. Takabi, H., Joshi, J., and Ahn, G., (2010) Security and Privacy Challenges in Cloud Computing Environments, *Security & Privacy, IEEE*. 8(6): p. 24-31.

26. Vaquero, L.M., Rodero-Merino, L., Caceres, J., and Lindner, M., (2008) A break in the clouds: towards a cloud definition, *ACM SIGCOMM Computer Communication Review*. 39(1): p. 50-55.
27. Wang, C., Wang, Q., Ren, K., and Lou, W. Ensuring data storage security in cloud computing.*Proceedings of the 17th International Workshop on Quality of Service*, July 13- 15, Charleston, SC.
28. Wang, C., Wang, Q., Ren, K., and Lou, W. Privacy-preserving public auditing for data storage security in cloud computing.*Proceedings of IEEE INFOCOM*, March 14 -19, San Diego, CA.
29. Wang, L., Tao, J., Kunze, M., Castellanos, A.C., Kramer, D., and Karl, W. Scientific cloud computing: Early definition and experience. *Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications*, Sep 25-27, Dalian, China.
30. Wei, C., Chien, C., and Wang, M., (2005) An AHP-based approach to ERP system selection, *International Journal of Production Economics*. 96(1): p. 47-62.
31. Wood, T., Gerber, A., Ramakrishnan, K., Shenoy, P., and Van der Merwe, J. *The case for enterprise-ready virtual private clouds*. 2009: USENIX Association.
32. Zhang, H., Jiang, G., Yoshihira, K., Chen, H., and Saxena, A. Intelligent workload factoring for a hybrid cloud computing model. Proceedings of the 2009 World Conference on Services, July 6-10, Los Angeles, CA.