Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

SYSTEM SIMULATION AS DECISION DATA IN HEATHCARE IT

Charles S. Brust, D.Sc.

Mayo Clinic
200 1st St SW
Rochester, MN 55905, USA

Robin Clark

QMT Group
1143 Oak Ridge Tpke
Suite 107A-134
Oak Ridge, TN 37830, USA

ABSTRACT

Information Technology in healthcare is an ever-growing enterprise, with medical providers becoming more and more reliant on data to make care decisions. With the increased reliance on these applications for care, questions arise around the availability and manageability of those systems. This paper examines a model which has been developed for the selection of computing infrastructure architectures in healthcare organizations. This model utilizes the Analytics Hierarchy Process (AHP) to weigh the various criteria that come into play for decisions of this nature. Further, to vet the recommendations of the AHP model, and to lend quantitative data to the decision making process, simulations of the various architectural options were built for various application scenarios. The results of these simulations thus serve as additional validation of the model's efficacy. This paper focuses on the use of discrete event simulation using ExtendSim® to assist in the architectural selection process for computing architectures.

1 INTRODUCTION

As the availability of technology in healthcare increases, the proliferation of information technology (IT) infrastructure projects and their potential solutions makes for a complex decision making process. Infrastructure (i.e. the hardware and operating system components of a computing platform) forms the backbone of any IT project – without the servers being in place, no software components can be installed or have a place to run. As the basis of the platform, it is vital that the infrastructure components be properly designed for the application being installed, as well as for interaction with other applications that are currently installed or may be in the future. Failure to do so may manifest itself in issues such as poor interoperability, suboptimal performance, higher cost, or a wide range of other issues. It is important then, that the proper infrastructure architecture be selected and implemented for each project.

1.1 Role of IT in the Healthcare Environment

The operating environment of healthcare organizations is characterized by several requirements that include strong regulatory requirements governing data sharing, storage and privacy safeguards, healthcare service provisioning requirements and standards. A variety of applications that range from data intensive functions such as genomic applications to patient critical systems such as electronic medical records and emergency room information systems are utilized throughout the institutions. As such, the most critical patients are reliant upon IT systems as a direct component of their care. Further, computing needs in both clinical and research arenas are increasing due to the proliferation of better technology, which in turn requires additional data storage and analysis. The complexity of the healthcare operating environment and the variety of requirements in the healthcare environment that need to be taken into account make the design of healthcare enterprise IT infrastructures a complex design and decision problem.

1.2 Role of IT in Providing Cost-Efficient Healthcare

As healthcare organizations work through the methods by which they will deal with declining governmental and insurance reimbursements, following IT infrastructure optimization practices becomes critically important. Such practices will not only reduce direct costs for the projects, but will also ensure that applications and processes in the practice are run in the most efficient manner possible.

The impact of the Affordable Healthcare Act and the healthcare reforms it puts in place is yet to be fully understood, but it is likely that these changes will add to the need for IT infrastructure. Data analytics will become a more important tool within the healthcare practice (i.e. "what percentage of our cardiac arrest patients had the full standardized protocol followed while they were in the Emergency Department?"), with analysis of processes, outcomes, etc. Therefore, the IT infrastructure in these areas must be designed in such a way as to fully support the needs of the practice and allow for the improvement of quality both clinically and administratively, but at the same time contain IT costs.

1.3 Healthcare IT Needs

As the medical field makes advances, there will continue to be growing needs for IT in healthcare organizations. Personalized medicine will require genomic fingerprinting of each patient – an extremely data intensive process. Virtual consultations, remote robotic surgeries, and other telemedicine practices increase the need for data and video bandwidth, and the importance of high availability in the systems supporting those functions. The availability of large datasets for research purposes means that researchers can now consider thousands or millions of patient histories when examining a disease, but to do so, the IT infrastructure must be in place which allows for the processing of queries and reports for big data scenarios (i.e. Hadoop, etc.). Each of these practice innovations will require the design of unique IT infrastructure improvements to best support the technological needs, and each of these improvements has the potential to further add to the complexity and costs of IT infrastructure in the healthcare environment.

Changes in medical practice, such as the advent of personalized medicine will require a significant increase in the IT capabilities of healthcare organizations also, as a wide array of high-powered computing resources are needed to process and store genomics data, and to apply that data in clinical practice (Ginsburg and McCarthy 2001, Sadee and Dai 2005, Burke and Psaty 2007, Ginsburg and Willard 2009, Ng et al. 2009). While these changes may be difficult and expensive to implement, there are many benefits to be realized, including increased patient safety, improved quality and efficiency from staff, and reduced patient cost due to interoperability and transferability of health information. The complexities of such implementations from an IT infrastructure perspective are immense, and as such, significant planning and care must be utilized from the onset of the project to allow for the optimal selection of architecture and configuration of IT resources, at the most cost-effective level possible.

2 BACKGROUND

In this section, the various technologies that may be used in enterprise computing architectures are examined. These technologies form the basis of the need for a decision model – there are many alternatives that IT groups need to consider when making the architectural decision around a computing platform for a particular application, and the wide variety of options, each with inherent strengths and weaknesses, makes this decision complex. An understanding of the architectural options is necessary to comprehend the need for a model to assist in such decisions.

2.1 Enterprise Computing Architectures

In today's medical science, there is a large need for high-throughput computing, and as the medical technology advances, that need for computing power grows exponentially. Today, much of the computing power needs fall into the research-oriented areas of medicine, but if current trends continue, physicians will soon be requesting genomics, gene sequencing and other highly specialized tests for day-

to-day clinical diagnoses, as well as utilizing these genomic markers to create customized treatments for cancer and other diseases (Donachy, Harmer and Perrott 2003). Areas in medicine that are using or planning to use grid or cloud technologies include genomics and bioinformatics, image rendering and storage for radiological functions such as CT scans, MRIs, and ultrasounds, the electronic medical record (EMR), and even drug development through human systems simulation.

To fully understand the scope of the options contained in the decision model, it is necessary to define the various computing architectures that will be considered.

2.2 Grid Computing

Grid computing can be defined as the use of multiple computing resources in parallel to arrive at the desired output results more quickly than would have been possible with a single system (Gagliardi et al. 2005). In some cases, grid computing has been performed by dedicated systems, tied together by specialized software for such tasks as large-scale computations or for 3-D animation rendering. In other situations, the processor time has been culled from "spare" CPU cycles on desktop systems in an enterprise, or throughout the world (i.e. SETI@ Home). In either situation though, the basic premise remains the same: use many low-cost computing devices together to do the work that once would have required large dedicated systems (Gentzsch 2002).

2.3 Cloud Computing

There are many definitions of cloud computing, with each variation focusing on a different aspect of the technology. 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" (Brust and Sarnikar 2011). 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.

2.4 Virtual Computing

Virtual machine technology allows for multiple operating systems to reside on a single piece of hardware, and to run simultaneously, through the help of a hypervisor layer. In this way, the physical resources of the hardware (CPU, memory, etc.) can be accessed by many systems. This allows for better resource utilization, since the virtual systems can be added incrementally to maximize usage, whereas an operating system running directly on that same physical hardware may not come close to using the available resources. This has become especially true in recent years as multi-core CPUs have come to the market – there is much more processing power available in most new computers than ever before, and the operating system and application market has not yet taken advantage of these added resources natively.

2.5 Application of Architectures in Healthcare IT

The possible uses for grid or cloud computing architectures 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 (Groen 2006). 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 et al. 2008). Among these are simulations, home health care, patient safety, electronic medical records, imaging, genomics, and bioinformatics (Breton, Dean and Solomonides 2005). While this certainly is not a comprehensive list, it does demonstrate the wide reaching functionality that the alternate architectures could provide.

Though there are many benefits to the use of public computing architectures in healthcare, there are also some barriers to wholesale integration. Chief among the concerns is security of data. Patient

criticality must also be examined, as patient care critical processes may be too risky to place in the cloud due to possible outages or connectivity issues (Leavitt 2009).

3 DECISION MODEL FOR COMPUTING ARCHITECTURE SELECTION

In order to make a selection of computing architecture, an Analytic Hierarchy Process (AHP)-based model was created (Brust and Sarnikar 2010) to enable multiple criteria (cost, patient criticality, etc.) to be analyzed through the use of weighted inputs from one or more users. These criteria are compared against each other in a pair-wise manner to determine the best fit of several possible outcomes (architectural selections) when compared to these criteria. By using such a model, it becomes possible to compare a wide variety of architectures by examining their individual strengths and weaknesses, and assessing those qualities against the desired qualities for the end system. The AHP model only gives qualitative data however, and as such, it is necessary to incorporate a simulation portion as well to gain quantitative data for total analysis of the solution.

4 SIMULATION ANALYSIS

Simulation of the computing architectures allows both for the validation of the recommendation given through the AHP model, as well as to gain quantitative data which will also be utilized in the final decision making process. As such, the simulation exercise is a vital component of the computing architecture decision method. The simulation was built as a discrete event model in ExtendSim.

4.1 Simulation Methodology

To evaluate the viability of modeling alternate computing architectures in a health care environment, scenarios are developed for use of these technologies in order to examine the possible architectures that might be employed in each instance. The simulations allow quantitative data to be gathered regarding several parameters of interest, and for this data to be utilized in the final decision process. By examining the implementation as it would be put in place through simulation first, we can also take uncertainties into account, by means of running simulations that would emulate the possible changes (i.e. higher traffic rates for a website, or a growth trend in the utilization of the application day-to-day) which could alter the computing architecture selected. The capability of doing such what-if scenarios in simulation also ties in well to the same ability of the AHP model. Such scenario manipulation is a core function of this modeling method, and will serve to improve the overall implementation of the projects.

To simulate a computing architecture, it is necessary to understand the particulars of that architecture, and to then generate a model which equates servers and processes to objects within the model. Taking a simple example of a physical architecture for a web-based application, one would need to understand the traffic pattern expected, the number of users that a server can handle concurrently, how long the web server takes to process each request, and whether those users are making single requests to the server, or if there is a back and forth conversation between user and server (i.e. looping requests). Once these parameters are understood, a model can be generated for simulations.

If we suppose that there are no looping requests in the example discussed previously, then our simulation model will contain a start point which generates the work units (requests) into the system, and which does this work unit generation at a specified rate and distribution. Downstream from the initiator will be the work processor – equivalent to the server in this example. The processor object (and associated queuing object if needed) take the requests from the initiator and hold them for a period of time (again, specified rates and distributions based on the projected real-time statistics for the application) to simulate processing of the work unit. Beyond the processor, the work unit exits the simulation as the request has been completed. Throughout the simulation model, it is possible to gather data regarding speeds of processing, number of servers required to process the workload, etc., and it is essential to gather the correct data in order to gain understanding from the simulation exercise.

In order to gather valid comparative data from each computing architecture, it is essential to ensure that the input metrics used are the same across the architectures, and that the output data is also

comparable. Specific inputs for the simulations include run rate and distribution, expected time to process work units, and projected outage parameters per server and for the entire application. From these parameters, the simulation outputs include the number of work units processed, the number of servers required to process that work, the average and maximum time to process each work unit, the cost to process each work unit, and the reliability and availability metrics for the application. With these quantitative outputs, essential data is made available to the decision makers regarding the most beneficial computing architecture to select for the application.

It is necessary to calculate some of the outputs of the simulation model in order to arrive at metrics that are easily compared between architectures. These include reliability, availability, and cost per work unit. Reliability (R) is defined as the percentage of time that all the servers/nodes in the architecture are up and available for users simultaneously (e.g. an outage of any single node reduces the reliability metric for the scenario). Availability (A) is defined as the percentage of time that the application is available for users. This means that there is no system-wide outage in place. Individual nodes may be down, however the overall architecture can still be utilized. Cost Per Work Unit (C) is defined as the cost of a server/node multiplied by the total number of servers/nodes required for the workload, divided by the number of work units processed.

Verification and validity testing of the simulation model are required to ensure that the simulation is designed in such a way as to emulate the real-world system and thus to allow accurate data collection to occur from the model. To accomplish these tests, several methods were employed, including phase validity testing consisting of output data comparison and expert verification of the model, and Turing testing (Turing 1950).

4.2 Sample Healthcare Scenarios for Evaluation

In order to evaluate the viability of simulation use, several scenarios are presented which represent applications that may be vital to the operation of a healthcare enterprise. Summaries of these scenarios are documented below.

Immunization / epidemic / pandemic – When viral pandemics, such as the H1N1 virus in 2009 hit, there is a need for large-scale immunization clinics to be held in order to ensure that all who want/need the vaccine are able to receive it (Stephenson et al. 2004). To support such an endeavor, all the IT infrastructure required for scheduling, tracking inventory, logistics, and immunization administration has to be in place prior to the opening of the clinic to patients. In order to prepare for such a potential situation, hospitals today would be required to maintain infrastructure that would not see any activity except in the event of such an emergency. This is not cost effective, and also requires additional administrator time to maintain these unused systems. The AHP model results for this scenario are as follows:

- Normal and Pandemic traffic cases:
    - Best option – private cloud web and interface tiers
    - Second option – public cloud web tier, private cloud interface tier

PHR / medical info website hosting (i.e. mayoclinic.com, WebMD, etc.) - As the public takes a more direct role in their healthcare, it is becoming more common for patients to maintain a personal health record (PHR) (Iakovidis 1998, Tang et al. 2006). Many healthcare providers are now hosting PHR sites - some with links to the EMR so that care is automatically updated in the PHR. Generally, these services are web-hosted so that patients have anywhere and anytime access to their data. Cloud computing could allow a low-cost method to accommodate both planned and unplanned scalability needs for this service. The AHP model results for this scenario are as follows:

- Normal and spiked traffic cases:
    - Best option – Private cloud
    - Second option – Hybrid cloud

Image rendering farm - it is becoming more common for surgeons to simulate difficult surgeries before performing them on actual patients, and/or to create 3D image (or sometimes even physical via 3D printing) models based on an amalgamation of CT, MRI, ultrasound, and other imaging techniques(Leventon 2000). In a cloud or grid computing scenario, these resources would be more

readily available for such tasks as they arose, without making significant impact on other ongoing computing projects / applications. The AHP model results for this scenario are as follows:

- Normal and high utilization cases:
  - Best option – Server grid
  - Second option – Desktop grid

4.3 Usage Scenarios and Associated Decision Problems

For each of the proposed scenarios, there are variables that come into play when making the architectural decisions. Table 1 summarizes the criteria (depicted as outcome variables) being considered in each scenario, along with the architectural alternatives which are applicable to them. Some of these, such as cost and architectural limitations impact every scenario. Others, such as patient criticality, are not universally equal. To understand the decision problems that are inherent in each scenario, it is necessary to lay out the variables that apply to that application. For uniformity in comparisons, Linux operating systems will be assumed for all servers within a scenario. For all testing scenarios where public cloud is considered, Amazon EC2 pricing (East Coast) will be utilized. As of this writing, the fee for such service is $0.085 per hour for a small server.

Table 1: Scenario Summary

Simulation Case Title	Tier	Business Service Type	Usage Scenario	Architectures to Simulate							Allocation / Prioritization	Outcome Variable
				Standard Physical Servers	Standard Virtual Servers	Dedicated Grid	Desktop Grid	Private cloud	Public cloud	Hybrid cloud		
Immunization Normal Traffic Case	Web	Noncritical Clinical	Normal Distribution	Y	Y	N	N	Y	Y	Y	FIFO	Cost, Availability, Reusability of infrastructure
	Interface			Y	Y	N	N	Y	Y	Y		
Immunization Pandemic Scenario Traffic Case	Web	Critical Clinical	Normal Distribution with large volume	Y	Y	N	N	Y	Y	Y	FIFO	Cost, Availability, Time to spin up infrastructure to meet volume
	Interface			Y	Y	N	N	Y	Y	Y		
PHR Service Normal Traffic Case	Web	Noncritical Business	Non-homogeneous Poisson	Y	Y	N	N	Y	Y	Y	FIFO	Cost of providing service, Availability to end users
	Database			Y	Y	N	N	Y	Y	Y		
PHR Service Spiked Traffic Case	Web	Noncritical Business	Non-homogeneous Poisson + random traffic increases	Y	Y	N	N	Y	Y	Y	FIFO	Cost, Availability
	Database			Y	Y	N	N	Y	Y	Y		
Image Rendering Normal Utilization Case	Rendering	Noncritical Clinical	Normal Distribution	Y	Y	Y	Y	Y	Y	Y	FIFO	Availability, Scalability, Cost
Image Rendering High Utilization Case	Rendering	Critical Clinical	Normal Distribution	Y	Y	Y	Y	Y	Y	Y	FIFO	Availability, Scalability, Cost, Time to spin up infrastructure

5 RESULTS AND DISCUSSION

In order to determine the success of the model and the simulations, it is necessary to examine the results of each, and to compare and contrast them to determine whether the validity of the model has been supported. This section will delve into the results of the simulation model in the scenarios previously described, compare those simulation results with the AHP recommendations made in section 3, and examine the similarities and differences found.

5.1 Simulation Results

The initial simulations show that there are significant benefits to be realized through the use of alternate IT architectures, including cost savings, increased reliability, and faster return of results. While some results were predictable, others were less so, and it is possible that some of the conclusions reached might

have been overlooked without running through the simulation portion of the model. There must also be consideration of other factors, such as system availability, reliability, etc., and more granular results can be expressed for these factors through the simulation phase. Table 2 summarizes the results of the simulations for each of the scenarios.

Table 2: Simulation Results

Scenario	Architecture	Mean # Servers required	Median # Servers required	Mean Reliability	Median Reliability	Mean Availability	Median Availability	Mean Time to process work unit (seconds)	Median Time to process work unit (seconds)	Mean Cost per work unit processed	Median Cost per work unit processed
Immunization Normal Traffic Case	Physical	15.6	20	99.60%	99.54%	99.96%	99.97%	4.59	3.934	$0.0145	$0.0186
	Virtual	14.95	20	99.96%	99.96%	99.94%	99.96%	4.799	4.092	$0.0098	$0.0131
	Private Cloud	5.65	6	99.92%	99.92%	99.95%	99.96%	3.5654	3.5652	$0.0013	$0.0014
	Public Cloud	6.2	6	99.92%	99.93%	99.91%	100%	4.9308	4.9309	$0.0014	$0.0014
	Hybrid Cloud	6.25	6	99.91%	99.91%	99.93%	99.95%	4.9304	4.9307	$0.0014	$0.0014
Immunization Pandemic Scenario Traffic Case	Physical	19.6	6	100%	100%	99.90%	100%	6.9	3.57	$0.0092	$0.0028
	Virtual	11.1	6	100%	100%	99.98%	100%	3.7255	3.5667	$0.0037	$0.0020
	Private Cloud	6	6	99.83%	99.83%	99.95%	100%	3.5659	3.5659	$0.0007	$0.0007
	Public Cloud	7	7	99.86%	99.88%	99.92%	100%	4.9309	4.9313	$0.0008	$0.0008
	Hybrid Cloud	7	7	99.84%	99.85%	99.96%	100%	4.9308	4.9306	$0.0008	$0.0008
PHR Service Normal Traffic Case	Physical	14.4	20	99.55%	99.55%	99.95%	99.96%	2.512	1.958	$0.0067	$0.0093
	Virtual	12.8	20	99.96%	99.96%	99.96%	99.98%	2.323	1.553	$0.0042	$0.0066
	Private Cloud	4	4	99.92%	99.92%	99.95%	99.96%	1.3332	1.3332	$0.00043	$0.0004
	Public Cloud	6	6	99.92%	99.92%	99.84%	100%	1.9995	1.9996	$0.0007	$0.0007
	Hybrid Cloud	6	6	99.92%	99.92%	99.95%	99.96%	1.9996	1.9996	$0.0007	$0.0007
PHR Service Spiked Traffic Case	Physical	11.2	4	100%	100%	99.94%	100%	2.523	1.334	$0.0026	$0.0009
	Virtual	9.4	4	100%	100%	99.97%	100%	1.831	1.334	$0.0016	$0.0007
	Private Cloud	4	4	99.83%	99.84%	99.90%	100%	1.3331	1.3331	$0.0002	$0.0002
	Public Cloud	6	6	99.84%	99.87%	99.88%	100%	1.9995	2.0001	$0.0003	$0.0003
	Hybrid Cloud	6.05	6	99.87%	99.89%	99.86%	100%	1.9997	1.9997	$0.0003	$0.0003
Image Rendering Normal Traffic Case	Physical	80.6	78.5	97.64%	97.59%	99.97%	99.97%	28925	28923	$11.3049	$11.0255
	Virtual	82.8	81	99.79%	99.79%	99.94%	99.95%	28924	28923	$8.1628	$7.9878
	Private Cloud	3950	80	99.56%	99.55%	99.97%	99.98%	28899	28909	$14.9780	$0.2165
	Public Cloud	2660	79	99.55%	99.56%	99.96%	100%	28911	28907	$8.6612	$0.2072
	Hybrid Cloud	4807	1290	99.53%	99.52%	99.95%	99.96%	28911	28917	$19.8994	$4.4095
	Server Grid	619.3	629.5	54.57%	54.80%	99.99%	100%	13100	13130	$14.8490	$15.0591
	Desktop Grid	1423.9	1372.5	99.99%	99.99%	99.08%	99.10%	18224	18215	$1.8758	$1.8089
Image Rendering Spiked Traffic Case	Physical	2776.15	2771	100%	100%	100%	100%	28894	28892	$7.4709	$7.4599
	Virtual	3057.15	3048	100%	100%	99.96%	100%	28893	28893	$5.2757	$5.2574
	Private Cloud	9143	3051	98.54%	98.55%	99.97%	100%	28990	28893	$6.1450	$1.8358
	Public Cloud	6390.95	3058	98.70%	98.71%	100%	100%	27618	28898	$4.0403	$1.8197
	Hybrid Cloud	5582.9	3050	98.70%	98.78%	99.95%	100%	28886	28894	$3.5037	$1.8301
	Server Grid	10913	11657	12.73%	15.29%	100%	100%	1884.5	1855	$62.5472	$66.8077

Scenario	Architecture	Mean # Servers required	Median # Servers required	Mean Reliability	Median Reliability	Mean Availability	Median Availability	Mean Time to process work unit (seconds)	Median Time to process work unit (seconds)	Mean Cost per work unit processed	Median Cost per work unit processed
	Desktop Grid	15309	15139	97.67%	97.67%	99.49%	100%	279903	279850	$36.9810	$36.6461

5.1.1 Immunization Scenario

In the immunization application scenario, two use cases are described – standard traffic, and pandemic situation, which equates to a burst of high traffic over a short period of time. In this case, the pandemic traffic is elevated for 3 weeks.

5.1.2 Immunization Scenario Simulation Results

An examination of the simulation data reveals that overall, there is only a slight difference in the number of servers required to support the high traffic vs. that of standard traffic, regardless of the architecture chosen. This then means that the determinate factor for cost per work unit is the base cost of each server, and as such, the cloud options are much less expensive than physical or virtual servers. Further, fewer servers are required overall in a cloud architecture, since spikes in utilization are more easily smoothed by allocating additional servers in 1-hour blocks. This further reduces the per-work-unit costs for cloud architectures. Private cloud architecture was shown to be slightly less expensive than other cloud options ($0.0002 per work unit difference, which equates to approximately $1,665 per year for the total solution).

The final consideration to be examined is that of time to process a work unit. In this scenario, a work unit is equated to a patient contact –the nurse requests a list of the immunizations the patient has had or needs, and the application returns the list. In a normal traffic situation the difference in timing for such a query shown for various architectures is inconsequential; however in the high-traffic situation, this becomes much more likely to affect patient care, and thus is a direct tie-in to the patient criticality criteria in the AHP model. In a pandemic, where a mass immunization clinic is being run, there may be tens or even hundreds of concurrent patients being seen, with a continual queue of additional patients following them. A 1 second difference in the timing of the response equates to nearly 180 fewer patients going through the system in an 8-hour shift. This additional delay can thus have a significant impact on the care of patients, and must be considered to be a determinate factor in the architectural decision.

It is necessary to note that while the AHP results were in line with the simulation results in this case, that may not be true in all situations. This is due to inherent constraints in any multiple criteria decision making problem, where optimization of the criteria can in some cases return different qualitative results when compared to the quantitative results from the scenario. In these cases, it is imperative that the institution examine all the data from both the AHP model and the simulation to make a determination as to the best path to follow.

5.1.3 PHR Scenario

The Patient Health Record scenario is similar to the Immunization case, in that there are two use cases described - standard traffic, and public health alert situation, which equates to a burst of high traffic over a short period of time. In the PHR case however, the estimated time traffic is elevated is 2 weeks.

5.1.4 PHR Scenario Simulation Results

Here again, the simulation data shows that overall, there is only a slight difference in the number of servers required to support the high traffic vs. that of standard traffic, regardless of the architecture chosen, though the overall required servers in any architecture for this scenario is smaller than its equivalent immunization scenario counterpart. Once again, with cloud architecture, we see a smaller number of servers required overall due to the allocation of additional servers in 1-hour blocks. In the

PHR example, we once again find private cloud architecture to be marginally less expensive than other cloud options (approximately $0.0001 per work unit difference, which equates to an annual difference of $1,636 for the total solution).

The availability and reliability scores for the PHR application were similar to those that were experienced in the immunization application, due to the nature of the similar infrastructure requirements for these scenarios. As such, the lower reliability of physical servers is again demonstrated through the simulations, and higher scores for those architectures utilizing alternate technologies that do not tie the operating system to the physical device are verified.

When considering the time to process a work unit, the workflow must be understood. For the PHR scenario, each work unit is a query or record update within the PHR application by the patient user. In a normal traffic situation the difference in timing for such a query shown for various architectures is inconsequential; in the high-traffic situation, it may have some effect, but not to the extent of that seen in the immunization application. Here, there is not a direct linkage between the application and patient care, since it is used by external customers who may or may not be patients of the institution. This does not mean though, that slow response times are inconsequential – the PHR application can be a method by which an institution cultivates new patients, and as such, a good user experience is essential to establishing the appearance of a high quality practice. The response times across the architectures tested are in the 2 second range, which has been shown (Nah 2004) to be acceptable for a web application.

5.1.5 Image Rendering Scenario

In the final scenario, there is a significantly different workload, and as such, a much changed landscape when looking at architectural options. The Image Rendering scenario examines an application which has work units that take a large block of system work time – 8 hours each on average. With this large delay, the time to process a work unit becomes much more significant than in the other scenarios that have been examined. Additionally, the availability and reliability criteria take a larger role in the decision regarding architecture, since the loss of a single node, or of the entire environment, will have much larger effects on the response times for work completion.

5.1.6 Image Rendering Scenario Simulation Results

Again, there are standard and high-traffic workloads to be considered for this application, with the high-traffic bursts taking a 1 week timeframe. The difference between the required numbers of servers is significant in this scenario when comparing the standard traffic load to that of high-traffic periods, and this difference will certainly contribute heavily to the architectural recommendation. Further, the difference between server requirements for the various considered architectures is also significant, as the requirement for server quantities could be less than 100 or over 4800 for the standard traffic case depending on the architecture selected. Finally, the cost component criterion shows a wide variety of options, with up to $18 per work unit separating the high and low architectures in the standard traffic case. This large difference in costs points quickly to a solution for institutions that are cost-conscious, though it cannot be the only concern considered.

Availability and reliability criteria take on a higher significance in this scenario, and there is a large gap between the high and low scores with regard to reliability shown in the simulation results. While most of the architectural options have fair to good reliability, all are lower than the other two scenarios, and dedicated server grid reliability is significantly lower, especially in the high-traffic workload simulation. This is a direct result of the longer time to process work units – if a node experiences an outage in this scenario, the downstream effect of having to restart the work elsewhere is much more noticeable. Further, the large number of servers required to handle the workload of this application makes it much more likely that nodes will go offline during the simulation.

5.2 Discussion – AHP results v. Simulation Results

Overall, the comparison between the findings in AHP and those in the simulations shows that the AHP model works well. The simulations supported the validity of the AHP model, while at the same time adding a factor of qualitative data that would not have otherwise been gained. This was shown specifically in the strong results for the hybrid cloud architecture and its ability to quickly circumvent or recover from node outages. Similarly, in the image rendering scenarios, the reliability of grid computing architectures was shown to be much lower than expected, which could easily influence the final architectural decision.

One purpose of running the simulations of the various architectural options is to verify the results from the AHP model to ensure the selection of the best architecture based on the input given regarding the criteria. With this goal in mind, it is clear that the simulation results are undoubtedly positive, and show strong support for the selections made in the AHP model. Further, the simulations lend stronger credence to the recommendations of the model by allowing numeric comparisons on cost, reliability, speed of results, and other factors which were not directly available from the AHP model. To that end, the simulations become more valuable to real-world users, especially those just starting to utilize the AHP model and who require additional data for making a final decision.

As supporting data, the simulations allow for a quantitative inspection of the expected real-world results, potentially bringing facts to light that might not otherwise have been discovered until the actual implementation of the system. For example, the image rendering application shows a much lower reliability score (55% in the standard traffic model, 15% in the high-traffic scenario) for server grid architecture. While this may or may not be acceptable for any given institution, the importance of the simulation is evident – the quantitative data would not have been available without running the simulations. In this way, the simulation exercise serves as a feedback loop for the AHP results as well, and additional considerations that are brought forth by the quantitative data can be examined before the final architectural decisions are made.

6 CONCLUSIONS

This paper has explored an important problem in healthcare information technology – the selection of the most appropriate computing architecture for use in a given project. A solution has been developed and proposed here, and the viability of that solution has been validated through multiple methods. It was found that in the decision modeling realm, a gap exists for situations requiring both subjective and objective evaluation of user data and opinions. The selection framework discussed addresses both of these gaps.

6.1 The Model as the Solution to the Problem

With the state of healthcare IT and the need for improved processes around architectural selection having been established, the AHP model has been discussed as a method by which healthcare IT enterprises can better examine the architectural options for a given application project, and to make the best selection of those architectures based on the patient care and business needs of the institution at that point in time.

6.2 Validation of the Model

The simulation experiments have shown that the proposed model has the capability of analyzing and recommending the best-fit architecture for a given application. Each of three scenarios was examined through the AHP model for architectural recommendations, and these recommendations were then compared to quantitative data generated through the simulation experiments for each architecture. By simulating the real-world performance of each scenario for every applicable architecture, it becomes possible to compare the results side-by-side to allow for the selection of the best solution for a given set of business requirements, both from a qualitative as well as quantitative basis. Some results of the

simulation were surprising when compared to initial reactions before running the AHP model, such as showing that grid computing architectures had comparatively low reliability per node as compared to other architectures. These unexpected results serve to reinforce the need to utilize simulation to gain quantitative data to examine the recommended architecture and other options, allowing a more informed decision process.

6.3 Contributions to Theory and Practice

This paper has shown several contributions which advance the field, including the establishment of a simulation model for the purpose of validating the architectural decisions made within the AHP model. Through the illustrative examples, it has been demonstrated that a decision model can be evaluated to explore whether an alternative enterprise computing architecture model will allow healthcare organizations to meet their ever-increasing computing capacity requirements, while still optimizing the architecture of the infrastructure to best-fit the needs of the institution. In the example simulations, several cases have been illustrated which allow healthcare enterprises to project capacity for internal and external clouds, grids, physical or virtual servers, or any combination thereof, with an added benefit of giving expected cost comparisons for the various scenarios.

In addition, the theories utilized in creation of this model allow it to be abstracted for other uses – the criteria are flexible and thus can be altered for use in other industries. Similarly, the simulations are not healthcare specific, and could be used to emulate the IT installations in various businesses. This flexibility across all the parts of the model ensure that it is thus not limited to a niche use in healthcare, but rather can be used throughout a variety of industries with small modifications to tune it best for those enterprises.

6.4 Limitations and Future Research

Limitations do exist within the simulation portion of the model. Primary among these is the validation of the simulation models. While steps were taken to ensure that the models created represented and acted as real-world systems would, there is opportunity for deeper validation of the models utilized in this paper. Further, the Turing test undertaken to establish validity of the simulation models was static in nature – result sets were utilized, however no real-time interaction or data extraction could be done as the simulations were run.

Finally, there is significant research to be done in the area of simulation modeling as it relates to computing architectures. While the principles of simulation are well documented in the literature, the more specific area of discrete event modeling for the simulation of a computing application has been virtually ignored academically. Much of the literature in this arena examines data flow, which while being a basis for the simulation model, falls short of the work undertaken in this paper. There is much to be learned through simulation, and there are significant potential cost savings to be realized through its use.

REFERENCES

Breton, V., K. Dean and T. Solomonides 2005. *The Healthgrid White Paper*, IOS Press.

Brust, C. and S. Sarnikar 2010. "Grid Computing in a Healthcare Environment: An AHP-based Framework for Enterprise Architecture and Design." In *Proceedings of the Decision Support Institute Conference* (DSI 2010), San Diego, CA.

Brust, C. and S. Sarnikar 2011. "Decision Modeling for Healthcare Enterprise IT Architecture Utilizing Cloud Computing." In *Proceedings of the AIS Americas Conference on Information Systems* (AMCIS 2011) Detroit, MI.

Burke, W. and B. M. Psaty 2007. "Personalized medicine in the era of genomics." *JAMA: the journal of the American Medical Association* 298(14): 1682-1684.

Donachy, P., T. J. Harmer and R. H. Perrott (2003). "Grid Based Virtual Bioinformatics Laboratory." In *Proceedings of the UK e-Science All Hands Meeting (2003)*: 111-116.

Gagliardi, F., B. Jones, F. Grey, M. Bégin and M. Heikkurinen 2005. "Building an infrastructure for scientific Grid computing: status and goals of the EGEE project." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 363(1833): 1729.

Gentzsch, W. 2002. "Grid computing, a vendor's vision." In *Cluster Computing and the Grid, 2002. 2nd IEEE/ACM International Symposium on,* (pp. 290-290).

Ginsburg, G. S. and J. J. McCarthy 2001. "Personalized medicine: revolutionizing drug discovery and patient care." *TRENDS in Biotechnology* 19(12): 491-496.

Ginsburg, G. S. and H. F. Willard 2009. "Genomic and personalized medicine: foundations and applications." *Translational research* 154(6): 277-287.

Groen, P. 2006. "Grid Computing, Health Grids, and EHR Systems." *Virtual Medical Worlds Monthly*.

Iakovidis, I. 1998. "Towards personal health record: current situation, obstacles and trends in implementation of electronic healthcare record in Europe1." *International journal of medical informatics* 52(1-3): 105-115.

Keahey, K., R. Figueiredo, J. Fortes, T. Freeman and M. Tsugawa 2008. "Science clouds: Early experiences in cloud computing for scientific applications." *Cloud Computing and Applications* 2008.

Leavitt, N. 2009. "Is cloud computing really ready for prime time?" *Growth***27**: 5

Leventon, M. E. (2000). "Statistical models in medical image analysis" Ph.D. Thesis Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Nah, F. F.-H. 2004. "A study on tolerable waiting time: how long are web users willing to wait?" *Behaviour & Information Technology* 23(3): 153-163.

Ng, P. C., S. S. Murray, S. Levy and J. C. Venter 2009. "An agenda for personalized medicine." *Nature* 461(7265): 724-726.

Sadee, W. and Z. Dai 2005. "Pharmacogenetics/genomics and personalized medicine." *Human molecular genetics* 14(suppl 2): R207-R214.

Stephenson, I., K. G. Nicholson, J. M. Wood, M. C. Zambon and J. M. Katz 2004. "Confronting the avian influenza threat: vaccine development for a potential pandemic." *The Lancet infectious diseases* 4(8): 499-509.

Tang, P. C., J. S. Ash, D. W. Bates, J. M. Overhage and D. Z. Sands 2006. "Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption." *Journal of the American Medical Informatics Association* 13(2): 121-126.

Turing, A. 1950. "Computing intelligence and machinery." *Mind* **59** (2236).

AUTHOR BIOGRAPHIES

CHARLES BRUST is Lead Systems Engineer at Mayo Clinic in Rochester, MN. He holds a B.S. in Information Technology from the University of Phoenix, and an M.S. and D.Sc. in Information Systems from Dakota State University in Madison, SD. His email address is brust.charles@mayo.edu.

ROBIN CLARK is the founder of the QMT Group where he is a simulation instructor, model builder, and block-level developer. He holds a B.S. in Physics from Tennessee Technological University and an M.S. in Management Science from the University of Tennessee. His e-mail addresses is: Clark@QMTGroup.com.