Evidence for a Cost Effective Cloud Computing Implementation Based Upon the NC State Virtual Computing Laboratory Model

Patrick DREHER[a,1], Mladen A. VOUK[b], Eric SILLS[b], Sam AVERITT[b]
[a]*Renaissance Computing Institute, Chapel Hill, NC 27517, USA*
[b]*North Carolina State University, Raleigh, NC 27695, USA*

Abstract. Interest in cloud computing has grown significantly over the past few years both in the commercial and non-profit sectors. In the commercial sector, various companies have advanced economic arguments for the installation of cloud computing systems to service their clients' needs. This paper focuses on non-profit educational institutions and analyzes some operational data from the Virtual Computing Laboratory (VCL) at NC State University from the past several years. The preliminary analysis from the VCL suggests a model for designing and configuring a cloud computing system to serve both the educational and research missions of the university in a very economical cost efficient manner.

Keywords. Cloud Computing, VCL, Cost Effective

Introduction

The concept of cloud computing has become a popular phrase in information technology (IT) over the past several years. Multiple commercial organizations have been aggressively building cloud computing capabilities with claims of economic advantages for their new architectures and operations methodology [e.g., 1, 2 and references therein]. These types of statements have sparked a vigorous debate with arguments both in favor and against the economic viability of cloud computing. [e.g., 3].

In this paper, we focus specifically on a cloud computing implementation within a research-oriented educational institution of higher learning, and we discuss some of the factors that demonstrate how such a system provides a scalable, sustainable, economically valuable and viable contribution to the campus layer IT cyberinfrastructure. We will leave aside a discussion regarding the economic pros and cons for a commercial based cloud computing operation and will only focus on a cloud computing implementation at an educational institution.

The first step in any discussion about the economics of cloud computing is to identify some of the basic components of such a system. Cloud computing can be

[1] Corresponding Author: Patrick Dreher, 100 Europa Drive Suite 540, Chapel Hill, North Carolina 27517, USA, dreher@renci.org

defined as a computational paradigm that assigns tasks to a combination of connections, computers, storage, software and services accessed over a network. This network of servers, storage and connections provides vast processing power for utility computing through a service architecture model. This system is sometimes also described or labeled generically as on-demand computing.

The concept of cloud computing has several key characteristics that provide users with a unique capability and niche among computational systems. Clouds can provide device independence from any particular hardware vendor and offer implementation of resource and cost sharing from among a large pool of users. Within this resource sharing concept, additional specific implementations help to enhance these general gains in technical performance, with potential follow-on economic savings. For example, technical efficiency and scalability is enhanced with centralization of infrastructure, location independence (as well as device independence), and efficiency in utilization through management of user demand load to the cloud system through implementation of software that controls simultaneous multi-user or project access.

Beyond these general technical enhancements, the idea of individual cloud architecture designs, specific implementations, and usage profiles have the potential for additional technical and economic impacts that can lead to better performance, throughput, and reduced costs. Areas at each specific site where such economies may be improved include:

- Network bandwidth and network load to the system
- Reliability and "up-time" of the system
- Site specific operational profile, including concurrent resource usage profile
- Services mix (IaaS, PaaS or SaaS - which ones and in what proportion) [2]
- Efficient on demand allocation and aggregation, and de-allocation and de-aggregation, of CPU, storage, and network resources
- Type of virtualization used (bare-metal to virtual load ratio)
- Scalability and rate of adaptability of the cloud to meet changing user demands
- Sustainability of the system under varying workloads and infrastructure pressures
- Serviceability and maintainability of the architecture along with the overall cloud computing system and user interfaces and application programming interfaces (API)
- Etc.

All of these factors can certainly apply to both commercial as well as educational/non-profit institutions. However, the weight or emphasis of each of these individual factors will be influenced by the design goals for each cloud computing installation and the intended use of the cloud computing system for a particular user base. This weighted emphasis on different aspects of cloud computing are driven by a combination of user requirements, and in the case of many commercial operations, a business plan that hopes to capture a niche of the business computational market by selling these services and capabilities at a profit.

[2] Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (Saas), e.g., http://www.webguild.org/2008/07/cloud-computing-basics.php, http://paastalk.com/cloud-saas-pass-market-overview/, http://www.webcloudworld.com/analysis/a-map-of-saas-paas-cloud-computing-players/

1. Building an Economically Viable Model for Cloud Computing

1.1. User Requirements

The first questions any IT architect must address when designing and configuring a system that provides IT services and capabilities is to understand who the users are and what they require from the system to support their work, i.e., the architect needs to construct and understand the operational profile of the system [4, 5]. There are several categories of users in a higher educational institution environment. There are **developers** of the base-line computing system infrastructure and of the base-line services (e.g., bare-metal images), there are software **authors**, and there are developers of enhanced images and services and **service integrators** (e.g., lecturers and teaching assistants). However, by far, the largest fraction of users in a higher educational institution are **students and faculty** (the end-users) who for most part are not interested in the underlying IT and prefer “one-button” access to services they need. It is this group that is also the driving engine and whose usage patterns can validate and justify the economic aspects of any cloud computing implementation.

Cloud computing systems serving users within a university environment must at least provide the following capabilities for the faculty, researchers, staff and students

- Deliver services and support to a wide range of users from the novice to the most sophisticated expert researcher - from users who can barely find the terminal, to those who are expert supercomputer users.
- Deliver a wide-range of course materials and academic support tools to instructors, teachers, professors, and other educators and university staff as part of the academic mission of the institution
- Deliver research level computational systems and services in support of the research mission of the university

Fulfilling such a set of user requirements across components of a distributed system of hardware and software, that is often coupled with a given level of network support, is sometimes categorized under the term of a “service–oriented architecture”. These types of IT systems provide the user with a given functionality, capacity, and quality of delivery connected through a mix of some set of both tightly and loosely coupled components. The distributed components have qualities that are on-demand or batch, reusable, sustainable, scalable, customizable, secure, and reliable. In addition to all of these technical characteristics, the IT architect of such a cloud computing system must demonstrate that such a system is cost-effective to operate and maintain.

1.2. Planning for Cloud Computing User Demand in an Educational Environment

One of the most straightforward cost savings on an academic campus can be realized through the centralization of data centers and related computing support costs. Measurements from such implementation within the North Carolina Community College System (NCCCS) demonstrated it may be possible to realize savings of 50% in the infrastructure budget [6]. Assuming that equipment is consolidated and such savings are realized, are there additional economies that can also be realized through the use of the computational systems in a cloud configured environment?

For a university based cloud computing system to be economically viable it requires a scheduling process that carefully shepherds these resources in a way that efficiently and economically matches the demand load over time to the system resources. A typical university environment has its own unique ebbs and flows of campus activity over the course of a year. Computing resources that support the academic program typically may see large growth in user demand during assignment times, and perhaps near the end of each academic term. On the other hand, during fall and spring break, during winter holidays, and perhaps in the summer, academic demand can be considerably lower. Universities with a sizeable research presence on the campus however, have research projects and activities that are active year round and show less dependency on the academic calendar. Since research projects that use HPC are chronically short of computational resources, they are an excellent resource utilization backfill provided that the cloud dynamically transfer resources between single-"seat" and HPC use modes.

1.3. The Virtual Computing Laboratory

Today's educational environments are especially diverse, dynamic, and demanding. To address these challenges, in February 2004 Sam Averitt et al. [7] described a technology that was labeled as the Virtual Computing Laboratory (VCL) [e.g., 8, Vouk08a]. This cloud computing idea was developed and implemented at the North Carolina State University through a collaboration of its College of Engineering and Information Technology Division to address a growing set of computational needs and user requirements for the university.

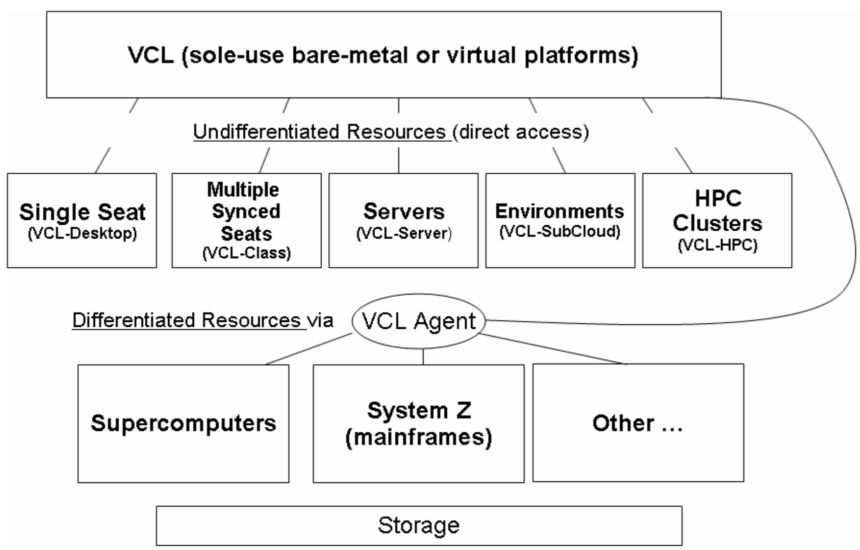

Figure 1. Illustration of the VCL service categories.

The Virtual Computing Laboratory is characterized by the design parameters for a secure, scalable, maintainable, and sustainable service-oriented architecture (SOA). This system can deliver user required solutions for a variety of diverse service environments anytime and anyplace on demand or by reservation. VCL delivers configurations to serve single real or virtual computer laboratory "seats" or desktops, single applications on-demand, classroom size groups of seats, enterprise server solutions, implementation of research clusters for specific calculations, aggregates of equipment to deliver (sub-)cloud service, and high performance computing services. Figure 1. illustrates the spectrum of VCL service categories that can be delivered in a university environment in an on-demand user model.

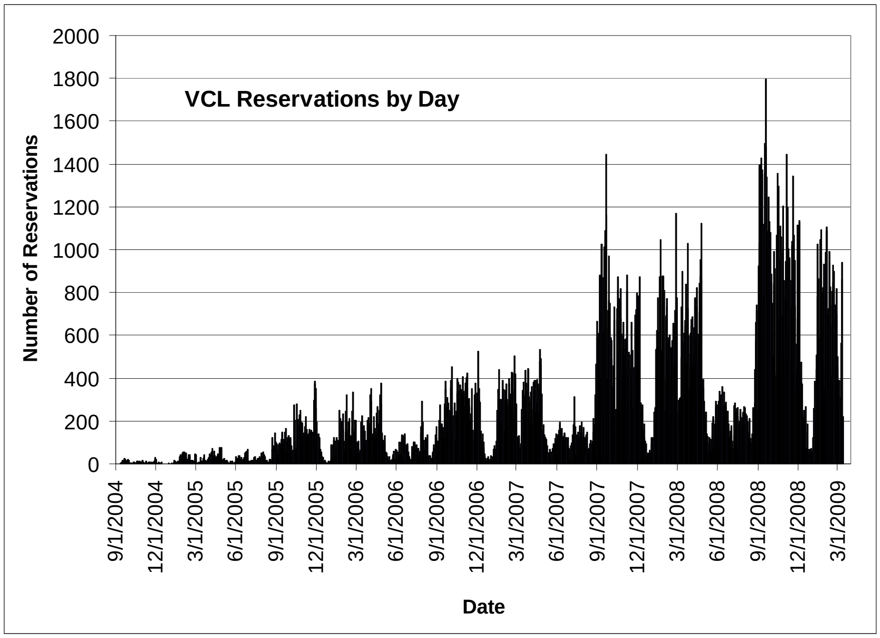

Figure 2. VCL reservations as a function of time from September 2004 through February 2009.

At NC State, the demand load on the VCL computational systems is monitored with care. Over the course of a calendar year, non-HPC usage shows recurring peaks and troughs in the demand load. For example, during semester there is a pronounced rise in the user demand level for VCL non-HPC resources. Figure 2 illustrates this periodic demand pattern for the period of September 2004 through February 2009.

Inspection of the usage pattern data shows that such demand levels are reached at only specific times throughout the year. Not surprisingly major troughs can be identified as corresponding to summer and winter holiday time periods.

Maintaining the necessary hardware capabilities to exclusively service these peak demand loads will leave large fractions of the VCL idle for large fractions of the year. On the other hand, having insufficient resources at peak periods will lead to dissatisfied users because they cannot access and schedule computing resources when they are needed. Having such a large excess (standby) computing capacity to assure availability of cloud resources at all times is not an economically viable path unless "idle" resources are put to some alternative use while they are not needed for VCL desktop student usage.

2. Operating VCL as an Economically Viable Model for Cloud Computing

Today both educational institutions and commercial vendors are deploying cloud computing resources and services, each with a somewhat different emphasis. On the commercial side, companies such as Amazon, Microsoft, Google and others have entered this area, each offering users a different mix of capabilities. For example, Amazon provides a platform with physical hardware and user controllable kernels and software stacks, whereas companies such as Force.com offer cloud resources that run against a very constrained set of applications. Depending on the commercial operation, users may have flexibility in assembling the cloud hardware but lack the additional systems to support such type of configurations without additional development, or they may be highly constrained to run only specific types of applications. [2]

Unlike commercially focused cloud computing operations, an environment in an educational institution has a complex set of operational requirements that may include openness, accessibility, mobility in transferring project information to and from the cloud system, control of the configuration of the hardware and software stacks, a capability for a richness and flexibility and a level of security of their computations, data and intellectual content [1, 10]. Such a list of requirements is difficult to fulfill in the commercial world within one commercial cloud environment.

In addition to the performance requirements, there are also capital equipment and operational considerations. When constructing a cloud computing system, there is a delicate balance between acquiring too many computing resources that are not efficiently utilized throughout the year and having an insufficient quantity to satisfy user demand during periods of maximum load. Having too many capital resources during large segments of the calendar year results in long periods of time where these resources are idle and being wasted. At the other end of the supply-demand spectrum, under-provisioning the cloud computing system can lead to serious dissatisfaction among users. Unacceptably poor performance levels for services can result because of the large number of users on the system and/or an inability to even access these advertised resources when needed.

Both scenarios are inefficient and each incurs a different economic cost. In one instance, excessive capital equipment expenditures are deployed with the resulting excess capacity, inefficiency, and underutilization. In the other scenario, there is a scarcity of resources because of the level of user requests and as a result, users are not serviced and are dissatisfied.

2.1. VCL Operations

Beginning in the Fall of 2004, VCL began its production operations, and today it serves a student and faculty population of more than 30,000 with both desktop and high performance computing services.- currently over 80,000 image reservations per semester, and about 7 million high-performance computing (HPC) CPU hours per year. The key differentiators that VCL offered to the university community to service the user requirements for a wide spectrum of computational environments are:

- simplicity of use and maintenance,
- required versatility, security and cost-effectiveness
- a broad resource-based approach to “virtualization,”

- flexible ways of delivering resource services through "images[3]."
- integrated delivery of both individual, group and HPC requests.

A more detailed explanation of the VCL architecture, the services and virtualization, and production operations were described by Averitt, et al. in 2007 and 2008 [8, 9].

The NC State cloud provides users with access to the reservations systems and management tracking through either a web portal or an API. Users are validated and authenticated into the VCL system using a variety of methods. Authorization to check availability and schedule VCL resources and image installation onto the cloud is controlled by one or more management nodes. These nodes can be within the same cloud, or among different clouds, and they allow extensive sharing of the resources. The management nodes are also involved in checking that licensing and other constraints are honored when resources are scheduled. In the case of NC State, all of its VCL images are equipped with middleware that allows the users to access NC State enterprise storage, storage on their own access computers, as well as any other network-accessible storage they have permissions to use.

NC State undifferentiated resources, i.e., resources that can be modified at will and on which users get full administrative privileges, are currently about 1,500 IBM Blade Center™ blades, but VCL can and does operate on other hardware such as Sun blades and Dell clusters. Typically about 40% to 60% of the blades serve high performance computing needs, the rest are in the individual "service" mode. VCL differentiated services, services that can be used with certain privileges but generally not modified at administrative level, are teaching lab computers that are adopted into VCL when they are not in use (e.g., at night), and other external resources and services that can be communicated with through the client-side VCL daemon or API. More detailed information about VCL user services, functions, security and concepts can be found in [1, 8].

Currently NC State's VCL is serving a student and faculty population of more than 30,000. Delivery is focused on augmentation of the student owned computing with applications and platforms that students may otherwise have difficulty installing on their own machines because of licensing, application footprint, or similar. We serve about 60,000 to 100,000 reservation requests (mostly of the on-demand or „now" type) per semester. A typical user reservation is 1-2 hours long. We currently have about 150 production images and another 450 or so other images. Most of the images serve single user requests and HPC cycles, with a smaller number focused on Environment- and Workflow-based services. Environments are aggregates of images that can be used to form virtual clouds.

In fact, the VCL implementation has most of the characteristics and functionalities discussed so far and considered desirable in a cloud. It can also morph into many things. Functionally it has a large intersection with the Amazon Elastic Cloud [11]. The difference between VCL and Elastic Cloud is that VCL is open source [12], allows users to request both bare-metal loads and virtualized loads, and construct their own cloud services. For example, by loading a number of resources (virtual or real) with

[3] The basic IT services delivery mechanism of VCL are its "images" and image aggregates. Conceptually, an image is a software stack that incorporates a) any base-line operating system, and if virtualization is needed for scalability, a hypervisor layer, b) any desired middleware or application that runs on that operating system or on a hypervisor, and c) any end-user access solution that is appropriate. Images can be loaded onto "bare-metal" (direct load onto hardware), or onto an operating system/application virtual environment (hypervisor) of choice.

Hadoop-enabled images [13] one can implement a Google-like map/reduce environment, or by loading and Environment or group composed of Globus-based images one can construct a sub-cloud for grid-based computing, and so on. It is also possible to integrate other clouds into VCL resource pool through its API and gateway images.

A typical NC State bare-metal blade serves as many as 25 student seats – a 25:1 ratio – a considerably better ratio than traditional physical labs at 5:1 to 10:1. Hypervisors and server-apps can increase utilization by another factor of 2 to 40 depending on the application and user profile. Our maintenance overhead is quite low – about 2 to 3 FTEs in maintenance and help-desk, for about 2,000 nodes, with another 3 FTEs in VCL development.

2.2. Evidence for VCL as an Economically Viable Cloud Computing System

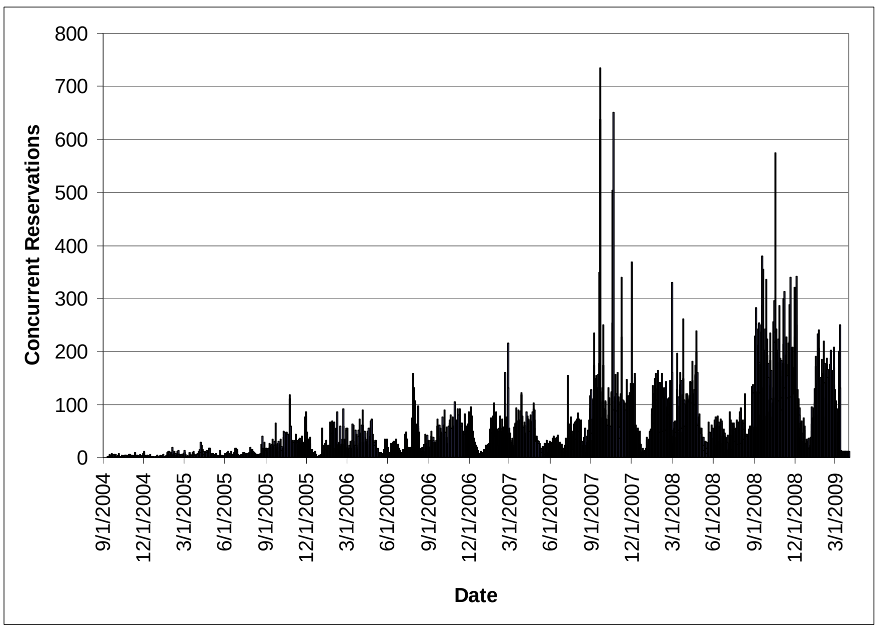

Figure 3. VCL concurrent reservations - from September 2004 through February 2009.

Figure 3 shows the number of non-HPC concurrent VCL reservations for the same period as Figure 2. Figure 4 shows the total and the maximum number of concurrent reservation for November 2008. At any given time during November, the maximum number of concurrent blade reservations by day remains around a value of 300 to 350. From Figures 3 and 4, we see that in our case concurrent usage is about 20% to 25% of the daily usage. It is important to note that this fraction does depend on the operational profile, so a large number of concurrent class (group) reservations may increase this fraction. Figure 5 shows that the average daily demand (for November 2008) itself varies by the hour of the day, thereby further refining the time window where a larger number of blades need to be operational and available for VCL users. This average shows the same consistency when measured over longer periods of time. This suggests that perhaps it is not necessary to keep 500 blades active and available for VCL use to provide demand surge protection, and some of them could be used in some other way.

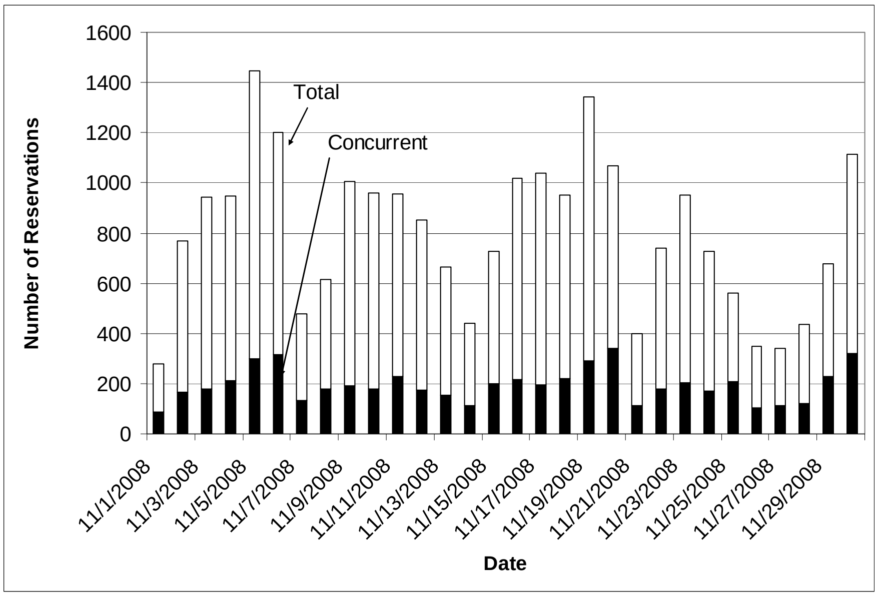

Figure 4. Total and maximum concurrent number of reservation per day – Nov. 2008.

Of course, having some level of slack in the system is very advantageous for several reasons. A spare capacity allows for maintenance flexibility and upgrades without disruption to the production systems. It also provides options to support other types of projects that can take advantage of intermittent levels of spare CPU cycles.

High performance computing is an excellent candidate to absorb this level of spare capacity in the system. Figure 6 shows the demand load for high performance computing as a function of time by month over the course of a year. Most of the demand is of the batch type. The data shows that throughout the year, the demand for HPC computational resources remains relatively constant. In the past year, on-campus HPC demand has consumed over 7,000,000 CPU hours .

By shifting some of the relatively constant but high HPC computing demand load onto spare capacity cycles of the non-HPC VCL resources, it is possible to make more efficient use of the VCL system, minimize the fluctuation in unused non-HPC VCL capacity over time, and provide the HPC computing systems with incremental boosts of computational power.

To understand the economics of VCL, it is important to look at its two operational profile components – the HPC and the non-HPC aspect. Current yearly non-HPC VCL usage is approximately 160,000 reservations and over 300,000 hours. At any one time, up to 500 blades of the VCL cloud are in the non-HPC mode. Current yearly HPC VCL usage is approximately 7,000,000 CPU hours on approximately 500-1,000 blades (most of them are two-processor variants with one, two or four cores each). Support for the desktop and HPC services is interconnected and includes a hardware support person, a system administrator, three developers, help desk support, and one person for HPC code optimization.

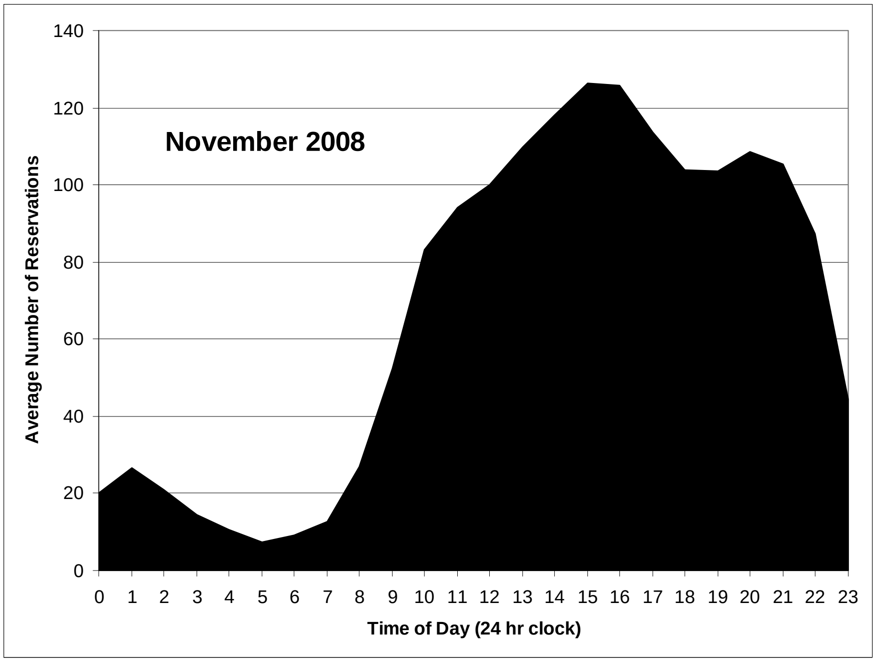

Figure 5. Average daily number of active reservations for November 2008.

Assuming that both desktop cloud computing and HPC cloud computing are both provided as VCL services to the community, the annualized capital and operational cost for this cloud is approximately $2,000,000[4]. This investment delivers 7.3 to 7.6 million combined desktop and HPC CPU hours, yielding a cost per CPU hour of approximately $0.26 to $0.27.

While this may seem high, it is important to note two things: a) NC State is still ramping up full production and we have extra capacity anticipating growth in usage, and b) currently both on-campus VCL modes utilize primarily bare-machine loads. The reason for (b) is that HPC users are not yet comfortable using virtual cores, and non-HPC engineering applications, which form the core of the requests, tend not to perform that well in a virtualized mode. We expect that this will change as virtualization becomes more efficient and more of the reservations are made for virtualized resources.

Further, inspection of the two modes of service individually can provide some useful insights. If only VCL desktop services were provided today (with the current VCL capacity and small to moderate virtualization), the total capital and operational annualized cost, including software, would drop to approximately $1,000,000. At the current level of annual usage of about 300,000 hours of non-HPC VCL services this would result in a cost of approximately $0.83 to $1.68 per CPU hour (assuming an average of 4 to 2 cores per blade respectively) or about $1.5 to $3 per seat hour (assuming from two to one concurrent images per blade respectively). Since non-HPC VCL usage is still experiencing a super-linear growth, and our new blades (eight cores,

[4] This number assumes use of existing data center facilities, and it does not include cost of power and cooling. The latter can add another 10 to 20% to the cost.

16 GB of memory) can safely host as many as 10-16 virtual machines per blade, we expect as much as a 10-fold drop in the non-HPC CPU-hour cost as the capacity fills up. If only VCL HPC services (using bare-machine loads) were provided today, the total capital and operational annualized cost would be approximately $1,400,000 to service about 7,000,000 CPU hours at approximately $0.20 per CPU hour.

We see that sharing of the resources provides an up front reduction in the cost (from more than $2.4 million down to about $2 million) as well as an increased average utilization of the resources that also lowers the cost per CPU hour.

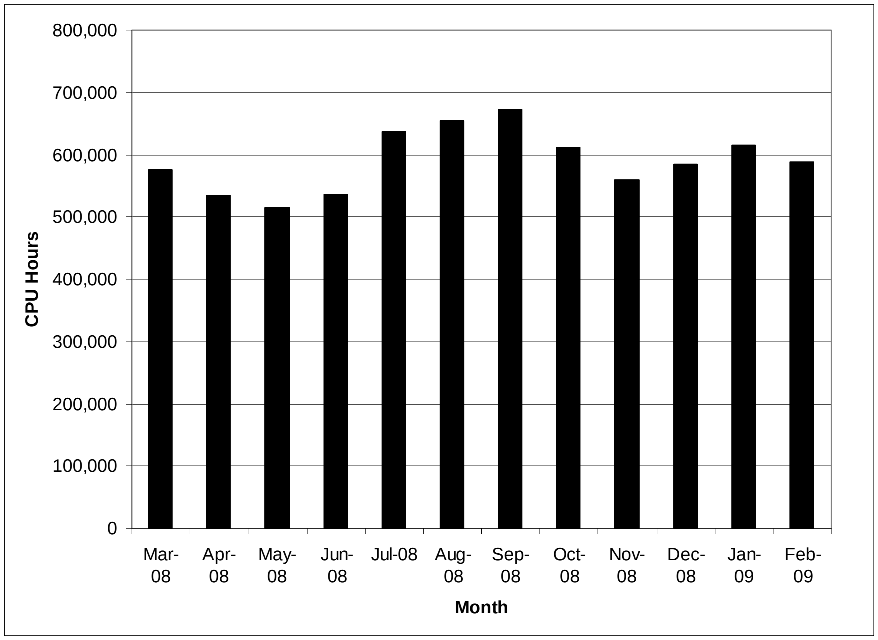

Figure 6. VCL-HPC usage (in CPU hours) March 2008 – February 2009.

3. Observations and Insights

Several observations and insights can be gleaned from an analysis of the faculty and student academic and research IT user requirements combined with the actual VCL operational data that has been gathered over the past several years. Probably one of the most important trends that can be inferred from the analysis of the combined VCL desktop and HPC utilization data is in the area of efficient utilization of the computational infrastructure.

The desktop Virtual Computing Laboratory serves an important function, delivering both educational IT support as well as providing small desktop analysis resources for research data. However, in order to provide these capabilities to users on across widely varying demand loads requires that the university make a large capital investment to assure this "on-demand" level of service availability.

The user demand over time for the educational aspects of the VCL is governed by the academic calendar of the university. Therefore, when users are to be able to access these academic cloud computing services with on demand reliability over 96% [8, 9] throughout the academic year, it means that a considerably amount of equipment needs to be in standby, or idle mode, for long periods of time, yielding a low average utilization rate over time and an expensive total cost of ownership.

A way to markedly decrease the total cost of ownership for the original system is to co-locate a potential complementary computational mode, with a higher and more consistent utilization rate over time, and seamlessly integrate the two systems. The VCL operational statistics over the past several years strongly support this idea and suggest that by designing and building a coherent integrated campus IT layer for faculty and student academic and research computational needs, it will allow the institution flexibility in servicing both of these university functions. It also allows the educational institution itself to maximize the return on their capital investment in the IT equipment and facilities and decrease the total cost of ownership.

The ability for university teaching and learning activities to have a strong IT capability that complements the classroom work has the best chance of being successfully implemented if there is also a large strong research component on the campus utilizing a common research computing infrastructure. Our paper indicates that the incremental cost to provide both efficient and economical academic and research computing services with minimal underutilization of equipment is enhanced by integrating the university IT teaching and learning aspects into the same capital equipment infrastructure that serves the HPC computing cloud needs. In particular, at NC State, using the VCL blades for both HPC and VCL desktop work provides economical services with minimal underutilization of equipment.

If there is no large research computing user base, it may still be possible to achieve an efficient utilization of resources applied to desktop virtualization. However, this requires a much larger and diversified base of users to effectively be the large solid computational base in the business model that will allow a fraction of the users at any particular time to have the on-demand desktop VCL capabilities. An important aspect to consider in any integration of such capabilities is to make sure that the additional services not all have the same user demand cycle over time. It may be possible to construct a diversified user base from a combination of K-12 users, community colleges, and university teaching and learning students that each operate on a slightly different academic calendar. Another very attractive options is to share some of the cloud resources with small to medium business and start-ups which may have computational needs that are flexible in terms of computational schedules (e.g., running of payrolls, running of asynchronous simulations, etc.). Yet another option is to share the resources across different time-zones.

A much longer and more detailed analysis of the economic model and business case for VCL cloud computing in a higher institution education environment utilizing the experience and operational data from the NC State Virtual Computing Cluster is in preparation at this time

4. Acknowledgments

This work has been supported in part by IBM, Intel, NetApp, Cisco, NC State, UNC General Administration and State of North Carolina. We would like to thank all our colleagues on the VCL project, and especially Andy Kurth, Aaron Peeler, Henry Schaffer, Sarah Stein, and Josh Thompson, for their input and support

References

[1] Mladen Vouk, "Cloud Computing – Issues, Research and Implementations," Journal of Computing and Information Technology, **16** (4), 2008, pp 235-246

[2] Michael Armbrust, Armando Fox, Rean Griffith, Anthony Joseph, Randy Katz, Andrew Konwinski, Gunho Lee, david Patterson, ariel Rabkin, Ion Stoica, Matei Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing" Technical Report No UCB/EECS-2009-28, http://www.berkeley.edu/Pubs/TechRpts.EECS-2009-28.html

[3] Bernard Golden, The Case Against Cloud Computing, Parts 1-5, CIO Magazine, Jan-Feb 2009 (http://www.cio.com/article/print/477473, http://www.cio.com/article/print/478419 , http://www.cio.com/article/print/479103, http://www.cio.com/article/print/480595 , http://www.cio.com/article/print/481668)

[4] J.D. Musa, "Operational profiles in Software-Reliability Engineering," IEEE Software, **10** (2), pp. 14-32, March 1993.

[5] J.D. Musa, *Software Reliability Engineering*, McGraw-Hill, New York, 1998.

[6] Darryl McGraw, "The Cost of Virtualizing: Student Computing at Wake Tech," Information technology Services, Wake Tech Community College, a 27th January 2009 presentation at the RTP, NC virtualization workshop.

[7] UNCGA grant proposal entitled "Support for Distributed Scalable High Performance and Grid-based Computing and Problem Solving with Emphasis in Science and Engineering," by Sam Averitt, Mladen Vouk, Henry Schaffer, Eric Sills, and Gary Howell, NC State University, Raleigh, February 2004.

[8] Sam Averitt, Michael Bugaev, Aaron Peeler, Henry Schaffer, Eric Sills, Sarah Stein, Josh Thompson, Mladen Vouk, "The Virtual Computing Laboratory," Proceedings of the International Conference on Virtual Computing Initiative, May 7-8, 2007, IBM Corp., Research Triangle Park, NC, pp. 1-16 (http://vcl.ncsu.edu/news/papers-publications/virtual-computing-laboratory-vcl-whitepaper).

[9] Mladen Vouk, Sam Averitt, Michael Bugaev, Andy Kurth, Aaron Peeler, Andy Rindos*, Henry Shaffer, Eric Sills, Sarah Stein, Josh Thompson "'Powered by VCL' – Using Virtual Computing Laboratory (VCL) Technology to Power Cloud Computing ." Proceedings of the 2nd International Conference on Virtual Computing (ICVCI), 15-16 May, 2008, RTP, NC, pp 1-10 (http://vcl.ncsu.edu/news/papers-publications/powered-vcl-using-virtual-computing-laboratory-vcl)

[10] Sam Averitt, E-Research In the Clouds, Educause Net@Edu, 2009 http://connect.educause.edu/Library/Abstract/EResearchintheClouds/48135

[11] Amazon Elastic Compute Cloud (EC2):http://www.amazon.com/gp/browse.html?node=2 01590011, accessed Dec 2008

[12] VCL incubation project at Apache, http://incubator.apache.org/projects/vcl.html, 2009

[13] Hadoop: http://hadoop.apache.org/core/, accessed May 2008