Journal of Advances and Scholarly Researches in Allied Education
Vol. XIV, Issue No. 1, October-2017, ISSN 2230-7540

Cloud Computing With Green Virtual Systems

Jagdish Kaur*

Assistant Professor, Department of Computer Science, DAV College for Women, Ferozepur Cantt

Abstract – In cloud computing, systems require a hardware inventory that is often prohibitive in terms of cost and manpower involved in its installation and configuration. In addition, physical servers in general are highly underutilized and remain powered on 24 by 7 regardless of the extent of their utilization. This is highly inefficient given on average servers are underutilized and use only 10% of the overall resources of the server.

This goal of this research is to create cloud systems that are "leaner" or more efficient in terms of their use of resources in terms of time and the quantity used. A "Green" cloud system responds to peak utilization periods and adjusts availability of resources based on them expanding or shrinking the cloud as needed. Furthermore, this system will detect the nature of the demand on it and provide customized resources specific to the needs of the services required by the users of the cloud. For example, more if more processor-intensive applications are in demand, the amount of processor resources will dynamically increase to support the demand.

Keywords- Cloud computing, Virtualization, Green computing, Cloud optimization, Virtual Computing

----------X----------

I. INTRODUCTION

Cloud computing is an architecture which provides services to users under one entity, a "one-stop" shop for virtual services. On the backend, the equipment and staff required to support it can be prohibitive to maintain. So much so, that more and more "clouds" are being hosted by outside contractors in virtual computers. This trend is

A. The Potential "Green" benefits of the Cloud

The reason why cloud computing is a greener technology is that it looks to consolidate and to better *utilize* computer resources which ensures the reduction of physical resources and hardware components needed within companies. This therefore reduces the amount of power and resources needed to manage these resources from the physical power supply, to cooling equipment to the actual need to call on IT repair individuals and companies when things go wrong. All these activities require some type of emission which will be reduced or eliminated by choosing cloud computing technology.

Cloud computing allows applications and data to be accessed from anywhere which means that employees can access company information at home. This could see a rise in the number of people working at home remotely and therefore ensuring they no longer need to do their daily commute to the office. This in turn ensures fewer emissions to the environment from their cars etc.

Another "green" benefit of the cloud is that it can be delivered as an on-demand in a form of "utility computing" where data centers deliver computational resources as needed. For this reason, these hosting sites don't have to be up-and-running 24/7 as data can be accessed when required which will significantly lower energy consumption. For this reason, utility computing can directly support grid computing architectures and web sites which require very large computational resources or those that have sudden peaks in demand (see Figure 1 below).

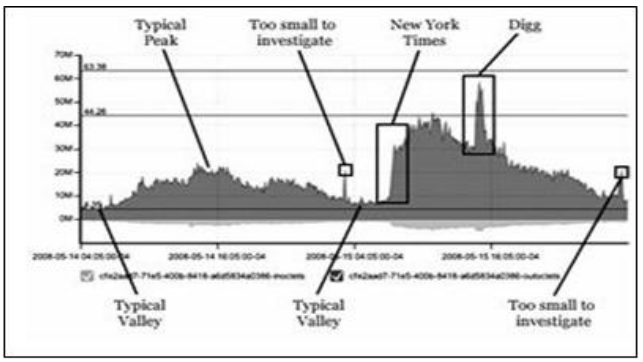


Figure 1. Internet traffic for major web sites over span of two days (Theo, 2011).

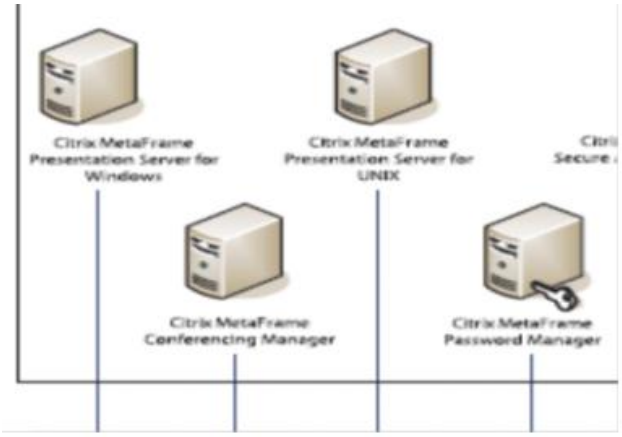
While the number of users accessing every cloud may not number in the millions, the challenge is still the same- as demand changes in the cloud, how to adjust the resources so as to maximize efficiency in times of peak usage and minimize the amount of resources consumed at others. In utility computing, this adjustment is typically performed at the level of the hosting facility using more efficient servers, power management, or server consolidation (Eric, 2011). While practical for larger hosting facility with access the latest technology and physical hardware, the same cannot be said for sites with more limited budgets or staff to support.

One alternative to making cloud hosting sites greener is to make the components of the cloud more environment-friendly. The remainder of this paper will describe an implementation which uses virtual machines to model a green implementation of a cloud-computing system. Section 2 explains in detail in how virtual machines are traditionally implemented. Section 3 describes a nontraditional implementation using adaptive infrastructure design. Section 4 measure the performance of this design against the traditional implementation. Given the results, Section 5 makes recommendations for future work, and Section 6 provides a synopsis of conclusions drawn from this research.

II. TRADITIONAL CLOUD COMPUTING ARCHITECTURES

The cloud-computing environment in today's business world requires users have access to business applications that may be hosted in multiple remote locations. These applications must appear as if the application is running locally on the user's computer while in reality it is actually being served from a host server or servers.

A typical deployment is displayed below.



Remote users access application servers hosted with virtual infrastructure. It is important to note also, there is often a need for identical application servers for the purpose of load balancing to accommodate a great number of users or for the purpose of high availability in case one server should fail.

However, creating a cloud with hundred or even thousands of physical servers can be economically impractical. More often than not, the resources provided by a physical server are severely underutilized. In additional, the average lifespan of a typical server is from two to four years. For these reasons, an increasing number of cloud architectures involve implementations using virtual machines.

A. The Virtual Cloud

Virtual machines within a cloud environment provide the flexibility and performance necessary to provide critical infrastructure services in today's business environment. They are easily replicated, maintained, monitored, and altered with limited investment in physical resources.

Virtual machines, in any environment, cloud or otherwise, still consume resources which, if unregulated, can cause considerable resource and power consumption. According to Cohen (Cohen, 2010), by 2013, approximately 60 percent of server workloads will be virtualized. He further anticipates 10 percent of the total number of physical servers sold will be virtualized with an average of 10 VM's per physical server sold.

At 10 VM's per physical host that means about 80-100 million virtual machines are being created per year or 273,972 per day or 11,375 per hour. Data centers utilizing virtual servers require enormous amount of resources- not the least being energy. "Attacking this data center consumption issue is really about green technology," said Gary Shambat, a researcher at Stanford University's Electrical Engineering Department.

B. Need for a Greener Virtual Cloud

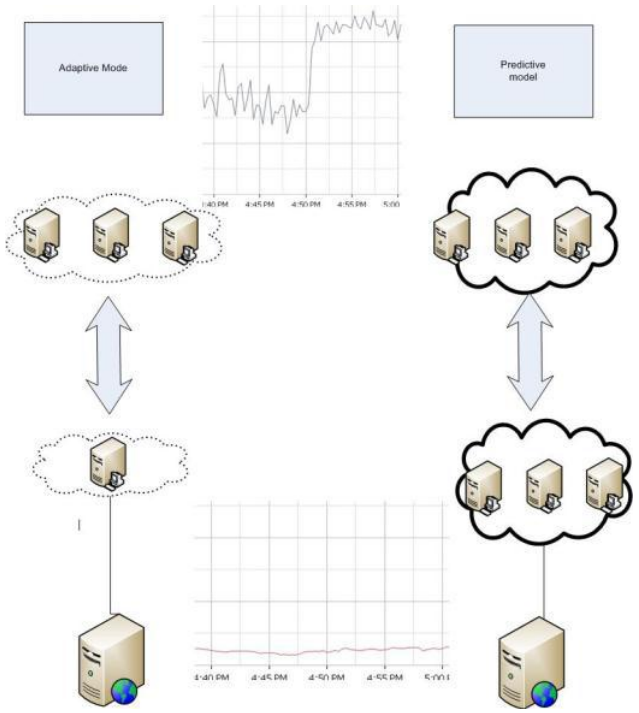
The reason why cloud computing is a greener technology is that it looks to consolidate and to better *utilize* computer resources which ensures the reduction of physical resources and hardware components needed within companies. With the cloud being an on-demand model it also means that data centers don't have to be up-and-running 24/7 as data can be accessed when required which will significantly lower energy consumption. In addition, there is also the fact that cloud computing allows applications and data to be accessed from anywhere which means that employees can access company information at home.

Yet despite the inherent advantages, deployment and management of virtual machines within the cloud is a time- and resource-consuming process. For global organizations, paying third party hosting providers to monitor, maintain, the machines which comprise the

cloud is considerable as the machines typically run 24/7 to accommodate all regions.

C. Predictive versus Adaptive deployment

Systems which feature automated deployment of virtual machines within the cloud help to ease the load on the physical hardware which hosts them. The challenge becomes in how to optimize the rollout of the virtual workload so as to meet demand while minimizing consumption of resources during idle times. Two such approaches to deployment of virtual servers in the cloud are predictive and adaptive.



Adaptive deployment increases the number of servers as demand increases. The size of the cloud and the components of it are created in small packages or increments so that as demand increases new resources are added to support the demand. This model is preferred when cloud resources are underutilized prone to only periodic spikes in consumption.

Predictive deployment uses large number of servers support the periods of optimal demand. This model is based on historical data obtained to anticipate demand and dedicate resources accordingly. The advantage to this model is that resources are scheduled ahead of time thereby providing optimal performance metrics. In optimal application, this model operates when resources consumption is more or less consistent over time.

The models have each their inherent advantages, but also disadvantages as any models do. Adaptive deployments require adjustments in the cloud in terms of number or types of servers- the efficacy of which depends upon the algorithms used to make real-time adjustments to the cloud. Predictive models provide optimal performance at the cost of over allocating resources in the cloud.

The paper proposes a hybrid model which incorporates the best features of both.

III. A HYBRID APPROACH

The optimal model for a cloud deployment is one in which the services provided are well-suited for the demands of the environment as it changes over time. When demand is high, available resources should be increased. Conversely, during period of periodic or low utilization, more conservative options should be employed to accommodate the limited demand.

The purpose of this paper is to propose a general model that can be employed to optimize the resources of the cloud by combining the best features of the adaptive and predictive models. First, it will generate or reduce the number of virtual machines deployed to the cloud by logging and anticipating utilization based on prior history and powering on and off machines as required. As the load increases, new virtual machines are added to the cloud to accommodate. When load drops off and fewer sessions are anticipated, machines are powered off. This is important especially for the fact that most cloud systems have peaks and valleys of utilization as opposed to more uniform patterns.

Second, the machines that are deployed will have characteristics which vary dependent on the demands of the cloud. “Seed” machines with different characteristics, i.e. number of processors or amount of memory, will form the basis for expansion of the cloud. Other machines, which are termed “worker” machines are added and removed to accommodate changing demand.

For example, when the environment demands more processor-intensive operations, more virtual machines are created with more processers per machine.

A. Seeding the Cloud

The foundation of the cloud consist of seed virtual machines that are always powered on continually monitoring performance and demands on the cloud and initiating changes based on the readings obtained. This machine type is called a “seed” because it forms the foundation for producing other machines of similar type to the cloud when conditions are favorable for the type. The function of the seeds within a cloud is to monitor, analyze, and increase the resources to it as demand warrants.

B. Right-sizing the Cloud

Depending on the type of demand imposed on the cloud at a given time, a corresponding "worker" machine is deployed with characteristics best suited to accommodate the current load. These worker machines are active only as long as connections are active on them. Once demand lessens, these machines proactively power themselves off to conserve resources for the other machines. To effort this, each worker machine monitors its own usage and creates a log entry with a timestamp and the associated session count.

In addition to adjusting the number of worker machines to the demand, the cloud also tracks and maintains the type of machines deployed. By default, the two main factors included in this adjustment are processor and memory utilization. When utilization of memory or processors is high, the cloud initiates the deployment of worker machines with more processors or memory. On the other hand during normal or low levels of usage, the worker machines deployed are lightweight in terms of processor and memory footprints. The seed machine maintains similar log entries as far as processor idle times and memory usage in order to determine the type of machines to deploy.

C. Summary of Roles in the Green Cloud

The two roles of seed and worker complement each other in the hybrid design. The seed is always powered on, but uses few resources, while the workers can vary in the amount of resources used and maintain power only as needed.

IV. GREEN CLOUD IMPLEMENTATION

The beauty of implementing this green model is in its simplicity and ease of maintenance. It can be implemented in any type of virtual environment and with software that is readily available and has little or no cost. The implementation in this study used virtual machines running Windows 2003 operating system and Citrix to deploy an application as part of a simulated hosted cloud environment. Aside from the application installed, each machine contains scripts directory containing a prolog rule base, WMI (Windows Management Instrumentation) scripting files, and a java class for passing information between them.

A. Prolog Rule Base

The purpose of the rule base is to provide a logical framework for the cloud to manage itself when it comes to decisions regarding cloud performance. To facilitate this, each machine has a prolog program, status.pl, with information identifying the cloud components and how they should interact in a changing environment.

This rule base functions in three basic ways:

1) Represents components of the cloud as atoms/prolog facts: each machine is represented with one or more facts in the prolog program. For instance, the following facts associate two machines as either worker or seed as well as their tolerance in terms of session count, processor, and CPU utilization (1). seed('am-usa39-ctxi42',normal,normal,normal). (1) worker('am-usa39-ctxi41',normal,high,normal).

2) Models environmental factors using rules consisting of atoms and variables: The rules represent, in rudimentary form, how the cloud "interprets" changes in its environment. For instance, in the case of the number of user sessions, the following rules are formulated regarding what is considered low, medium, or high levels of utilization (2,3,4).

sess(high) :- the_hour_past(H),the_day(D), (2)

tss(D,H,_,S,_),S > 5.

sess(normal):the_hour_past(H),the_day(D), (3)

tss(D,H,_,S,_),S > 2, S<5.

sess(low) :- the_hour_past(H),the_day(D), (4)

tss(D,H,_,S,_),S < 3, S > 0.

3) Runs queries against collective rule base to determine actions to invoke: Prolog predicates query the rule base for data that is used to adjust the composition of the cloud. The poweroff predicate below queries for a machine candidate to power on when additional resources are needed(5).

poweroff(X) :- not(sess(present)),is_worker(X). (5)

This predicate checks to see if sessions(present) atom is true, that is, if there have recently been sessions on the machine, and is_worker verifies that the machine on which the script is running is classified as a worker machine. If so X, assumes the value of the machine name and is returned as part of the query.

B. WMI Scripting

Each machine executes Windows Management Instrumentation scripts (WMI) written in VBScript which query the operating system for performance-related data such as memory and processor utilization and session count. The machines store this information in separate prolog data files as facts. The prolog engine interprets these facts along with the prolog rule base as part of its scheduled operation. The fact below records the number of user sessions with a timestamp (day, hour, minute) with first three

values followed by the actual number of connections for that time (6).

tss(7, 15, 0, 23, a). (6)

In a similar fashion, the system logs processor idle time in the form of another prolog fact (7).

proc(7, 15, 40,98,i) (7)

The last argument in each case (6,7) simply designates the type of connection in the case of (6), "a" for “active” and in (7) “i” for “idle”.

C. Java Class for Cloud Management

A java class connects the logical foundation of the prolog rule base and logs to actions that manipulate machines within the cloud.

Using JPL, a java interface into prolog, the java program calls methods which execute prolog queries. These queries return values, which, in turn, are used in calls to a Perl scripting interface changing the power state of machines within the cloud accordingly.

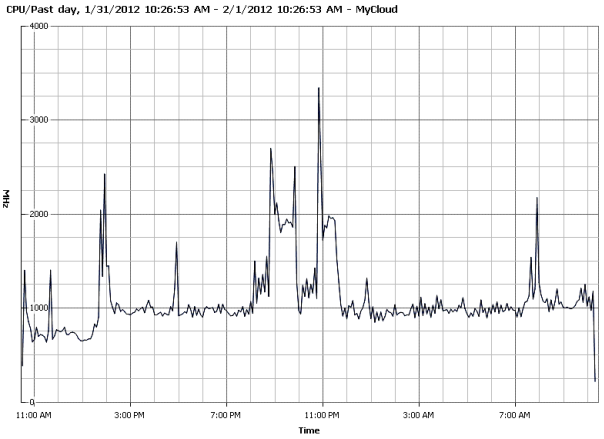
V. TESTING THE GREEN CLOUD

The purpose of this research is to demonstrate the efficacy of the model, but not in terms of measured energy savings. Other research has successfully developed tools to quantify the energy consumed by virtual machines (Kansal, et. al., 2010). Rather, the results here will quantitate efficiency of the model with the number of machine hours saved versus a more conventional architecture.

As a test of the green cloud model, two different cloud simulations were carried out consisting of virtual application servers hosting an application published in Citrix. A load generator created sessions over the span of 12 hour period during which data was collected.

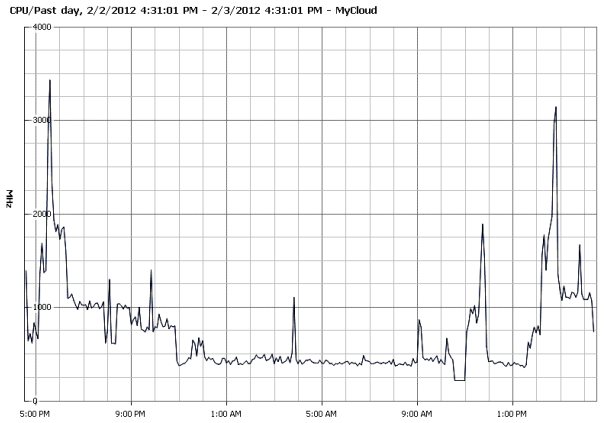
A. Traditional Simulation

The first simulation used three machines configured in the traditional design, load-balancing taking place across all machines with no changes in the power state or distribution of resources throughout. In the simulation, the load was steadily increased over the span of 2 hours with processor and memory monitored throughout (Fig. 4).



B. Green Simulation

The same load simulation was conducted on identical machines running as designed in the “predictive” architecture. In this simulation, all machines except for a seed machine were powered off. As the load increases, the machine initiated the powering on of the worker machines. When the simulation was finished, the worker machines powered themselves off according to the rules set up in the rule base (Fig. 5).



VI. CONCLUSION

In the implementation described here, these rules defined a cloud where a small representative set of machines initiated the expansion of the cloud as needed. This resulted in lower consumption of resources throughout the duration of the load testing. As load subsided, the same rules initiated the powering down of machines no longer required given the diminished load.

The predictive model as described is not a single type of implementation as much as a design for a more flexible cloud. This flexibility depends upon a rule base which can be as complex or as simple as the environment demands. As with any model, its application versus one using a more traditional

model depends on the scenario in which it is deployed.

VII. FUTURE WORK

While the research to date demonstrates some promising features, it is limited in terms of scope of deployment and depth of testing. Moving forward, the research will seek to expand upon the potential of the model by making it easier to deploy to a larger cloud and on creating a more complex rule base through which to increase the flexibility of the model as described.

REFERENCES

acts-computer-data-centers-with-huge-carbon-footprints/.[Accessed 04 January 12].

Aman Kansal, Feng Zhao, Jie Liu, Nupur Kothari, and Arka Bhattacharya (2010). Virtual Machine Power Metering and Provisioning , in *ACM Symposium on Cloud Computing (SOCC)*, Association for Computing Machinery, Inc., 10 June 2010

Erik Silk (2011). Santa Clara's cheap electricity attracts computer data centers with large carbon footprints. [ONLINE] Available at: http://peninsulapress.com/2011/02/16/santa-claras-cheap-electricity-attracts-computer-data-centers-with-huge-carbon-footprints/. [Accessed 04 January 12].

Reuven Cohen (2010). The Cloud Computing Opportunity by the Numbers -[update]-. [ONLINE] Available at: http://www.elasticvapor.com/search?q=60+percent&SEARCH=SEARCH. [Accessed 14 November 11].

Schlossnagle, Theo (2011). "Dissecting Today's Internet Traffic Spikes." OmniTI.com. http://omniti.com/seeds/dissecting-todays-internet-traffic-spikes. (3 Feb 2011).

Silk, Eric (2011). "Santa claras cheap electricity attracts computer data centers with huge carbon footprints." Peninsulapress.com. http://peninsulapress.com/2011/02/16/santa-clara-cheap-electricity-attracts-computer-data-centers-with-huge-carbon-footprints. (16 Feb 2011).

V.M. Ware (2010). "Reference and Capacity Planning with Citrix Presentation Server"[Online].Available: http://www.vmware.com/pdf/esx2_citrix_planning.pdf [Accessed: Mar. 31, 2010].

Corresponding Author

Jagdish Kaur*

Assistant Professor, Department of Computer Science, DAV College for Women, Ferozepur Cantt

E-Mail – armaanpreet29@gmail.com