Cloud Resource Management – Virtual Machines Competing for Limited Resources

Aaron Blojay Grant
Computing/Apple Department
DSG International Limited
London, United Kingdom
engrgrant@yahoo.com

Opeoluwa Tosin Eluwole
Transaction Security Department
Underwriters Laboratories
Basingstoke, United Kingdom
eluwole.opeoluwa@gmail.com

***Abstract*—The present-day high-tech feats in data, mobile, wireless and broadband technologies cannot be overemphasized. Cloud computing technology makes it possible to backup, develop and host application design for the internet where information technology (IT) related facilities are provided "as a service"; allowing clients access technology-enabled services more economically and flexibly on a pay-as-you-use basis. This paper underlines effective resource management in Cloud environments. We utilize two Cloud simulations namely CloudSim and CloudAnalyst to analyze the way virtual machines (VMs) are mapped to hosts, the distribution of loads and management of host resources among VMs as well as their behavioural patterns under various deployment configurations. Real-world scenarios are considered and a resource management strategy performed on a personal computer (PC) using activity monitor and VMware fusion running three different operating systems (OSs). Amazon Elastic Compute Cloud (EC2) is used to evaluate and analyze resource consumption via CloudWatch. Results obtained indicate that effective resource management enhances the Quality of Service (QoS) offered by Cloud Service Providers (CSPs) as well as reduces the processing and response times. This effectually reduces the overwhelming cost of consumption; a great benefit to CSPs, and large-scale applications such as Facebook, Twitter and YouTube.**

Keywords—Cloud computing; EC2; CloudSim; CloudAnalyst

I. Introduction

The Cloud phenomenon is rapidly becoming the de facto standard for internet computing, storage and hosting in recent years. It is built on different research areas such as Service Oriented Architecture (SOA), grid computing and virtualization technology. Like every day public utilities such as gas and electricity, Cloud technology can provide computing resources with utility in a dynamically scalable, virtualized manner to end-users. CSPs utilize the concept to effectively provide infrastructural services to their clients via the Internet in a very flexible manner, enabling elastic scaling of resources. Since this is performed on usage or demand basis, cost is effectively reduced. As the prevalence of Cloud computing continues to grow, the need for resource management within the infrastructure layer also increases. The provisioning of the Cloud infrastructure in the data center (DC) is a fundamental prerequisite. This is followed by how well resources are allocated, migrated and managed. Also, dynamic usage patterns of the user, location and geographical distribution of DC, availability of Internet and Cloud service adaptation are all key factors in Cloud resource management. Overall, large-scaled software systems such as social networking sites can benefit tremendously from Cloud services to minimize costs, meet Service Level Agreements (SLAs) and improve QoS to customers [1], [2].

Further sections of the paper are organized as follows: In section II, concise but insightful background information on Cloud technology, its deployment and service delivery models are elaborated. Section III presents the concept of virtualization as well as resource monitoring and management techniques. Evaluation through simulation (CloudSim, CloudAnalyst, and EC2) in real-time on real-world scenarios are carefully presented in section IV. Section V gives results obtained from various scenarios/case studies examined while section VI concludes the paper.

II. Background

A. Cloud Computing – Types, Service & Deployment Models

A Cloud can be referred to as a huge pool of easily reachable, dynamically reconfigurable or scalable virtualized resources including hardware, development platforms and facilities in such a manner that optimum resource utilization can be achieved. A system designed based on the Cloud technology concept makes up a Cloud environment. Two main models in Cloud technology are deployment and service models. Clouds in their deployment models can be *public (external), private (internal) or hybrid.* A public Cloud is hosted, operated and managed by a third party vendor from one or several DCs with shared resources dynamically provisioned on self-service bases over the Internet via web application and bill on utility (pay-as-you-go) bases. A private Cloud is a simplified representation of complex internetworked devices and connections defined according to relationship with the enterprise network. Its pattern may be dedicated (hosted within the customer-owned DC), communal (located at the premises of a third party while owned, managed and operated by vendor-customer established SLAs) or managed (infrastructure owned by client but managed by a vendor). A hybrid Cloud combines both public and private Clouds, permitting the flexibility of running non-core applications in an external Cloud while maintaining core or sensitive data internally. On the other hand, the service models in Cloud computing form the Cloud architecture and are divided into three major layers viz *Software as a Service* (SaaS), *Platform as a Service* (PaaS) and *Infrastructure as a Service* (IaaS). Cloud computing simply implies the delivery of those virtualized resources and/or capabilities using any of the three service models [1] [3 – 4]. Figures 1, 2, and 3 show the Cloud concept, Cloud deployment models and Cloud service models respectively.

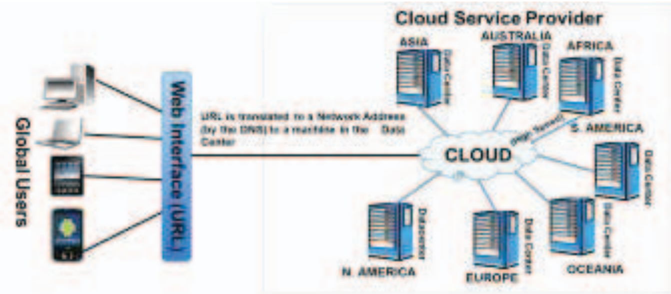

Figure 1. Cloud technology at a glance

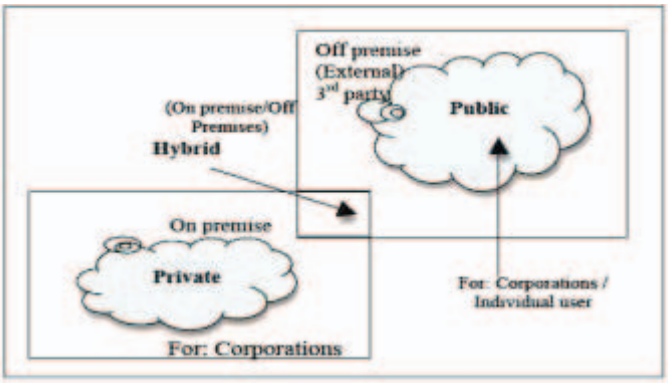

Figure 2. Cloud deployment models

Figure 3. Cloud service models

B. SaaS, PaaS and IaaS

In SaaS, access is purchased on a subscription or pay-as-you-use basis in order to use applications hosted in the Cloud. SaaS supports an architecture that can run multiple instances of itself regardless of location. The vendor does management and monitoring from the end-user prospective. Typical SaaS examples include Salesforce.com, Google Apps such as Google Docs, and Google Mail. In PaaS, consumers can procure access which enables them to deploy their own software and application in the Cloud. This offers a higher level of abstraction to make a Cloud easily programmable. There is an integrated management in PaaS. Typical PaaS examples are Salesforce's Force.com, and Google App Engine. In IaaS, highly-scaled redundant and shared computing infrastructure can be purchased and accessed by consumers using the Internet technology. Consumers control and manage the systems in terms of utilization but not the Cloud infrastructure itself. Typical IaaS examples are Amazon Web Services with Elastic Compute Cloud EC2 for processing and Simple Storage Service S3 for storage [1], [4 – 7].

III. VIRTUALIZATION AND RESOURCE MANAGEMENT

A. The Virtualization Phenomenon

The term virtualization cannot be dissociated from Cloud computing. In fact, the massive deployment of virtualization tools at the infrastructure layer typifies the uniqueness of Cloud computing technology. Virtualization implies the abstraction and democratization of computer resources; essentially the quartet of storage, processing power, memory and input/output (I/O) devices. Like the emulation concept, virtualization makes a system which pretends as two or more of the same system appearing as a shared resource at various layers of virtualized service. Figure 4 depicts the architecture concept of virtual of virtualization technology. The virtualization layer partitions the physical machine or resources of the underlying actual or physical hardware into VMs with workloads. This simplifies management of resources because utilization is by multiplexing lot more VMs on one physical system or host. Having a credible management strategy for managing Cloud resources or VMs' infrastructure is critical for any CSP that sell the IaaS. It can be passionately argued that the most significant driving force for Cloud resource management today is the advancement in virtualization technology. One of the key benefits of virtualization is the different techniques discovered to quickly scale workloads and resources up and down without any obstruction to the services being provided; thereby guaranteeing performance, adaptability, QoS and scalability. [1], [6], [8], [9 – 11].

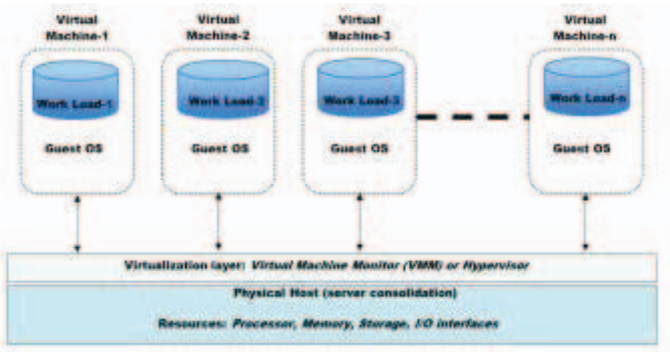

Figure 4. Architectural concept of virtualization technology

B. Virtual machines (VMs), Resources and Hypervisor

*VM*s are chain of imitation systems on physical machine (PM) pretended to be independent and share resources of the underlying PM. A VM's life cycle begins with a request delivered to the IT department (cycle 1). The IT administration processes the request by matching these resources with the requirements from the servers' resource pool and start provisioning of the needed VM (cycle 2). Once provisioned and started, it's ready to provide the required service according to the SLA (cycle 3), after a time period the machine is then released (cycle 4). A *hypervisor* sits at the lowest level of the hardware and acts as a "traffic cop" to schedule, split and manage the amount of access guest OSs have to the *resources* (processor, memory, disk I/O, storage and other I/O devices) [1].

C. Provisioning, Migration and Management of VMs

Provisioning a VM is performed by defining the server's configuration based on hardware requirements and software. The procedure can be summarized as follows: Load OS and appliances (step 1); customize and configure (step 2); install patches (step 3); start server (step 4); and run the provisioned VM (step 5).

In VM migration, key components such as CPU, storage disks and memory are completely virtualized. The migration can be *hot (live) migration*, *cold (regular) migration* or *live storage migration.* The former is performed when the machine is powered on and moves from one physical host to another with no noticeable effect. The intermediate occurs when the machine is powered off and moves from one physical host to another. The latter involves moving the virtual disks or configuration file of a running virtual machine to a new data store without any interruption in the availability of the VM's service.

VM management coordinates the activities of provisioning and migration based on SLAs. For this to be effective, there must be integration of virtualization's management and SLA's management tools so as to achieve balance in resources by transferring and monitoring the workloads, and accordingly meeting the agreed SLAs [1], [7], [9].

D. Scheduling-access and Dynamic Resource Techniques

Scheduling-access, operating via the hypervisor utilizes a *user-lever* job approach. A user-lever job contains a container (requested resource) and not necessarily the workload. It works with a possibility of obtaining a container of resources to which jobs can later be bound. In dynamic/automatic resource management, according to application requirements, up-down dynamic/automatic scaling can be applied to infrastructure using application-specific metrics such as transactions/second, number of simulation users, and request latency. This helps to achieve prompt response to traffic increase, greater fault tolerance, and minimized energy consumption [1], [10], [12 – 14].

IV. EVALUATION THROUGH SIMULATION AND REAL-TIME

A. CloudAnalyst, CloudSim and CloudWatch

CloudAnalyst and CloudSim (depicted in Figure 5) work in tandem with each other as the former is built on the latter's toolkit with a Graphic User Interface GUI for modeling and evaluating in Cloud environments [2]. Load requests between VMs are stabilized by the *VMLoadBalancers*. The CloudAnalyst simulations in this paper utilized the Round-Robin (RR) load balancer method. Its algorithm works on a rotating basis keeping a list of all VMs in the process. One VM is assigned to a workload, the next for the second request till the list is exhausted, and then it moves back to the beginning of the list to continue distribution. CloudWatch is a distinct application that measures resource consumption and level of usage.

Figure 5. CloudAnalyst/CloudSim Structure

B. Simulation Summary

Amazon Web Service (AWS) is used as case study for real-life example applying CloudWatch for monitoring resource consumption. Two real-world scenarios are demonstrated to illustrate how Cloud resource management is implemented in a Cloud environment. The first describes virtualization concept on PC where Facebook and YouTube resource consumptions are examined and analyzed. This was implemented using VMware fusion to install three different OSs namely Linux Ubuntu, Linus RedHat and Windows XP. Each OS runs independently or simultaneously. The specifications for virtualization on host machine are presented in Table I. The second scenario focuses on Amazon as a case study where EC2 is created and Amazon CloudWatch used to monitor metrics such as latency, and CPU utilization [2], [9], [15 – 16].

For the simulation-based investigation, Facebook was hosted on slow and fast processors with details summarized in Table II.

Table I. Specifications for Virtualization on Host Machine

Physical machine (Host)		
Operating system	Macintosh OSX	Version 10.6.8
Processor	Intel Core 2 Duo	(2.4GHz)
Memory	4GB	(667 MHz DDR2 SDRAM)
Storage	500 GB	
Number of VM	1 to 3	
VM software	VMware fusion	Version 2.0
Virtual machine (Guest)		
Number of OS	1 to 3	Windows Xp, RedHat, Ubuntu
Memory	1GB	
Storage	100GB	
Processor	1 core	

Table II. Simulation-based Scenario Details

SIMULATION-BASED SCENARIO	**CASE STUDY**		
	Case I	**Case II**	**Case III**
Scenario I Facebook hosted on slow processor	1 DC: 20 PMs, 50VMs/DC, 1R	2 DCs: 10 PMs/DC, 25 VMs/DC, 2R	3 DCs: 25 PMs/DC, 50 VMs/DC, 3Rs
Scenario II Facebook hosted on fast processor	1 DC: 20 PMs, 50VMs/DC, 1R	2 DCs: 10 PMs/DC, 25 VMs/DC, 2R	3 DCs: 25 PMs/DC, 50 VMs/DC, 3Rs

Where DC: Data Center; PM: Physical Machine; VM: Virtual Machine; R: Region

C. Simulation Setup, Assumptions and Evaluation Factors

- A Six User-Base (UB) pattern [2] (shown in Table III) is defined to represent the six continents of the world. In a real world scenario, each user is represented independently. The simulator is informed on the number of users that should be treated as a single bundle for traffic generation.

Table III. User Base Regional Time-zone Plan

User Base UB	**Region R**	**Time Zone**	**Peak Hours** (GMT)	**Simultaneous Facebook users** (peak hours)	**Simultaneous Facebook users** (off-peak hours)
1	0 (North America)	GMT-6	13:00-15:00	1,435,000	143,500
2	1 (South America)	GMT-4	15:00-17:00	345,000	34,500
3	2 (Europe)	GMT+1	20:00-22:00	1,210,000	121,000
4	3(Asia)	GMT+6	01:00-03:00	250,000	25,000
5	4 (Africa)	GMT-2	21:00-23:00	45,000	4,500
6	5 (Oceania)	GMT-10	09:00-11:00	120,000	12,000

- Single time zone is specified for each UB assuming that most users make use of Facebook application during evening hours after dinner for two hours and each user make a new request after every 180 seconds.

- For cases I, II and III (Table II), it is assumed that application is hosted in one (Region 0), two (Regions 0 & 2) and three (Regions 0, 2 & 3) DC locations respectively. RR technique is used for load balancing and UBs connect to DCs during either peak or off-peak hours.

- Key factors forming a bundle of QoS parameters used to analyze resources and load balancing among VMs in the simulation-based investigations are: *Response time, Processing time and Cost.* The simulation configuration parameter details are presented in Table IV.

Table IV. Simulation Parameters (Adapted from: [2])

Server Parameters (Per machine)		
Number of servers	*x*	Depends on the case study
Architecture	x86	
Operating System	Linux	
Virtual Machine Monitor (VMM)	Xen	
Number of processor	Four (4)	Per host (PM)
Processor speed (Intel Pentium):	100MIPS at 100MHz	Per host (PM)
Memory (RAM)	2048 MB	2GB = Per host (PM)
Storage (Hard-disk)	100,000 MB	100GB = Per host (PM)
Bandwidth	10000 bytes	10GB = Per host (PM)
VM allocation policy	Time-Shared	Time-shared = Per host (PM)
DC Characteristics (Billing) Note: The billing plan is assumed but follow the actual pricing pattern of Amazon (EC2)		
Cost per VM (Hourly)	0.10 USD	Per Hour
Cost per Memory (Hourly) 1MB	0.05 USD	1MB/Hour
Cost per Storage (1GB)	0.10 USD	1GB/Storage
Cost per Data transfer (1GB)	0.10 USD	1GB/transfer (to and fro)
Virtual Machines (VMs)		
Name	*x*	Depends on case study
Number of VM	*x*	The number of VMs to be allocated to an application from the Selected DC.
Imagine size	10000bytes	A single VM imagine size
Memory	1024MB	Amount of memory made available
Bandwidth (BW)	1000byte	Amount of BW made available
User Base (UB)		
Name	*x*	Depend on case study
Region	R0 – R5	Divided into the six continents
Request/User (Hourly)	*x*	Depends on case study
Data size request	*x*	Depends on case study
User grouping factor	1000	
Request grouping factor	100	
Executable instruction length	250	

"Xen" is a VMM for x86 architectures (x86 is the instruction set architecture based on Intel 8086 CPU). Allow several guests OS's to execute on the computer hardware concurrently. *x* is a whole number depending on case study.

V. RESULTS AND DISCUSSION

A. Virtualization on PC

Using Macintosh in-built resource management software called activity monitor (refer to Table I), the following results depicted in Figures 6 – 8 below are obtained.

In Figure 6, no VM is used. Out of the host machine specifics (4GB of memory, Intel Core 2 Duo at 2.4GHz of processor and 500GB of storage), the result shows that when streaming a video on YouTube for five minutes, up to 46.21% of processing power, 31.75% of memory and 34.49% storage is utilized. Conversely, when sharing, uploading or downloading pictures and other information on Facebook for five minutes, up to 21.74% of processing power, 31.50% of memory and 34.48% of storage is utilized.

Figure 7 shows the different percentages of utilizing the host machine specifics shared with the VM. When surfing Facebook to share, upload or downlaod pictures and other infomation for five minutes running Windows Xp alone on the VM, up to 39% of processing power, 71.25% of memory and 34.74% of storage is utilized. Running Windows Xp and Linux Ubuntu simultaneously, up to 58% of processing power, 92% of memory and 34.83% storage is utilized. By running the trio of Windows Xp, Linux Ubuntu & Linux RedHat concurrently, up to 65% of processing power, 99.25% of memory and 34.94% storage is utilized

In Figure 8, surfing YouTube for five minutes running Windows Xp alone on the VM shows that up to 61.15% of processing power, 72.25% of memory and 34.72% of storage is utilized. Running Windows Xp and Linux Ubuntu simultaneously, up to 80% of processing power, 93.5% of memory and 34.83% storage is utilized. Running Windows Xp, Linux Ubuntu & Linux RedHat concurrently indicates up to 90% of processing power, 96.5% of memory and 34.94% storage is utilized.

Table V gives a juxtaposition of the numerical consumption figures for Facebook and YouTube on the various OSs utilized. Markedly, YouTube consumed on average more processing power and slightly more memory due to streaming over Facebook and the fact that majority of the applications on it quintessentially entails the triad of data, voice and video transmission and/or reception. Both applications however consumed practically similar storage capacity of magnitude lesser than half of the total as there was no need for this to be fully utilized.

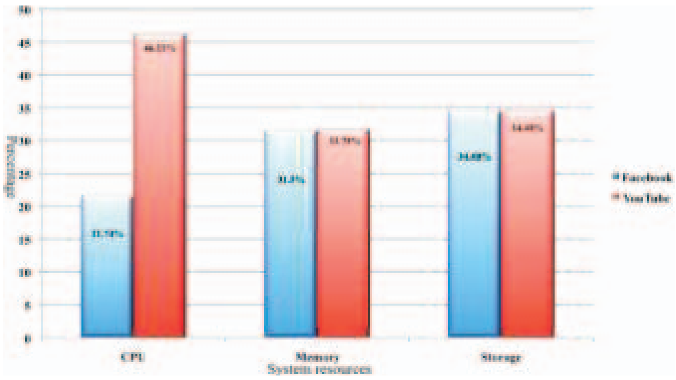

Figure 6. Percentage of resource consumption when running Facebook and YouTube independently on the host

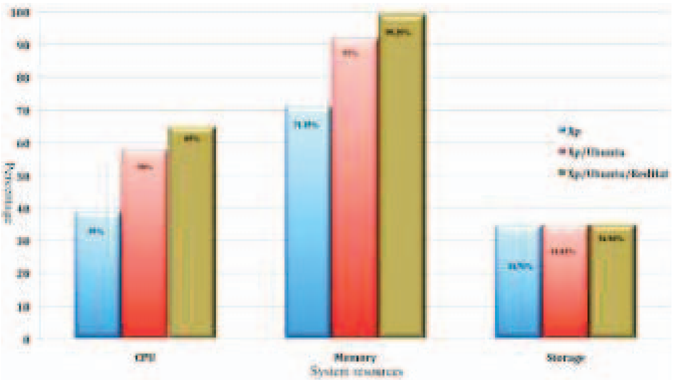

Figure 7. Percentage of resource consumption by Facebook

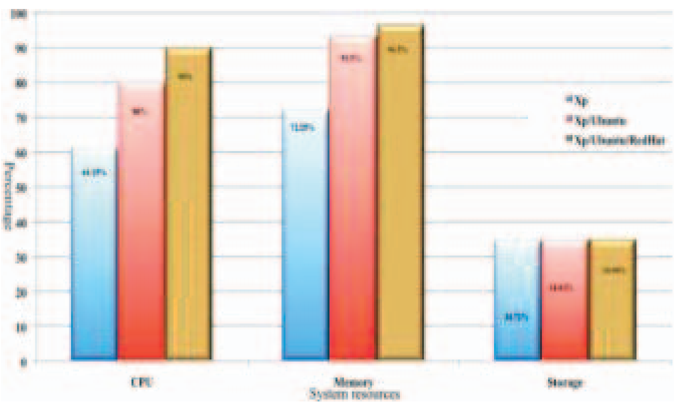

Figure 8. Percentage of resource consumption by YouTube

Table V. Numerical comparison of resource consumption by both applications

System Resources	FACEBOOK			YOUTUBE		
	Xp	Xp/ Ubuntu	Xp/ Ubuntu/ RedHat	Xp	Xp/ Ubuntu	Xp/ Ubuntu/ RedHat
Processor (2.4GHz)	39%	58%	65%	61.1%	80%	90%
Memory (4GB)	2.85	3.7	3.97	2.9	3.7	3.9
Storage (500GB)	173.6	174.1	174.7	173.6	174.1	174.7

B. *Amazon Web Service (AWS)*

Figure 9 shows two EC2 instances running and CloudWatch monitoring the level of resource consumed by each instance. The two EC2 instances (t1 micro) were launched with the following configuration: CPU=2 Core=1, Memory = 613 MB, basic 32-bit Linux Ubuntu AIM and CloudWatch was used as monitoring tools to monitor resource consumed by Facebook and YouTube. Results indicate resource utilizations were below 1%.

Figure 9. Amazon AWS running EC2 with CloudWatch

C. *Evaluation through Simulation*

Two scenarios are examined here using Facebook as case study (refer to Table II). The first entails simulation on a host machine having slow processor running at approximately 100 Million Instructions Per Second (MIPS). The other scenario has a host machine with fast processor running at 82300MIPS. For both scenarios, all the three cases listed in Table II were performed with three main parameters (processing time, response time and cost) analyzed on each successfully completed run.

In case I of *scenario I*, increased traffic was observed during the two-hour busy period for each UB region. The peak period of one UB which increased traffic affected other UBs, resource consumption, response and processing times. For instance, UB 1 peak period occurring between 13:00-15:00GMT affected other UBs slightly. Due to the small number of requests generated from other UBs during this period the impact was somewhat less. Also, due to the huge traffic generated by UB 1 and UB 3 during the peak period, approximately 30 and 27 million requests respectively were processed per hour. This increased the DC processing time to 70seconds and over 60seconds per user. This means that users will experience response times of approximately 33seconds which might rise when more new users log on.

In case II of *scenario I*, there was no significant difference in terms of response times compared to case I. However, substantial difference was observed in processing times. Approximately 20million requests were processed per hour during the peak periods. This increased the 1 DC processing times to 120seconds. In the same vein, approximately 18millions requests were processed per hour. This dropped the 2 DC processing times just over 120seconds. This translates to a user response time of approximately 49seconds which might rise when more and more users log on the social network. Also, CloudAnalyst calculations show that VM cost was cheaper in case I while data transfer was cheaper in case study II.

In case III of *scenario* I, significant differences were observed in terms of response times across each region compared to cases I and II. Each DC processed less than 20million requests per hour during the peak period for less than 60seconds. This means that users will experience response times of approximately 22seconds which might rise during certain periods of time when more new users emerge on the network.

The results obtained examining all the three cases hosted on slow/fast processors are summarized graphically and numerically in Tables VI & VII and Figures 10 & 11 respectively.

Table VI. Data Centre overall average processing and times

Simulation Scenario	Average Processing Time (ms)	Average Response Time (ms)	Average Processing Time (ms)	Average Response Time (ms)
	Slow processors at 100MIPS		Fast processor at 82,300MIPS	
Case I	32316.78	32615.99	4.13	260.62
Case II	48079.31	48340.52	20.67	141.08
Case III	20882.45	21113.99	3.99	114.41

Table VII. Total cost per case study (in USD)

Data Centre ID	VM Cost Slow/Fast processor	Data transfer cost Slow/Fast processor	Total Slow/Fast processor
Case I	120.07/120.07	2702.35/2702.35	2822.42/2822.42
Case II	120.07/120.07	2702.35/2702.35	2822.42/2822.42
Case III	360.21/252.15	2702.35/2702.35	3062.56/2954.49

Inspection of results depicted in Figure 10 & 11 shows that the processing time reduces drastically comparing Figure 11 compared to Figure 10. Response time drops substantially as well with the fast processor. Increase in the number of DCs reduces both processing and response times due to the fact that the services are taken closer to clients. Also, effective load balancing between DCs and VMs and at the application layer enhances the response and processing times even more. To summarize, for QoS to be efficient, sufficient capacity is required to ensure effective service is offered during peak demand. This is evident when the fast processor was deployed. The response time was very fast compared to the slow processor. A linearly reduced response time was equally experienced with the fast processor. Also, better response time can be achieved by deploying a perfect load balancing technique or algorithm for VMs at the DC and application level. Also, bringing the service closer to the customers can improve response time.

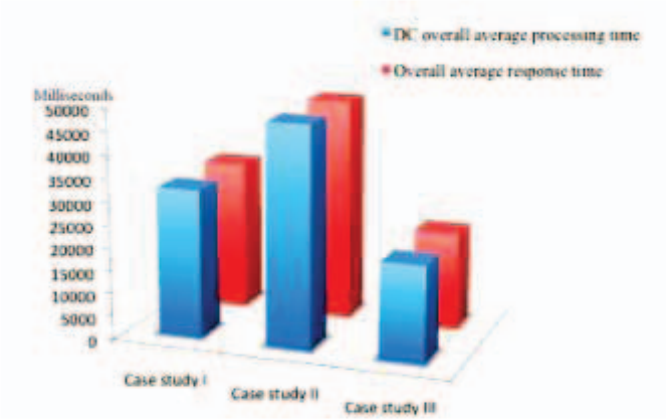

Figure 10. Overall processing and response times using slow processor

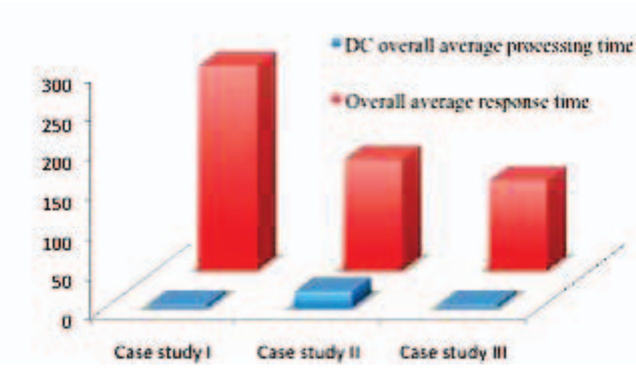

Figure 11. Overall processing and response times using fast processor

VI. CONCLUSION

In this paper, Cloud resource management entailing efficient load balancing distribution among VMs has been investigated via CloudAnalyst/CloudWatch/CloudSim tool and a RR VMLoadBalancer technique. The scenario with a slow processor showed how the overall processing and response times affect the usage patterns of the DCs hosting the application. In the second scenario with a fast processor, an effective resource load distribution improves the response times drastically. Overall, simulation results indicate that effective resource and load management strategies enhance performance, reduce cost of consumption and improve processing and response times when loads and resources are properly monitored and shared dynamically among VMs and/or DCs. These culminate in improved QoS to clients. Facebook and YouTube have been used as cases because of their wide global audience and the huge benefits they can acquire from Cloud technology. It can be said that Cloud resource management is not only about providing pay-as-you-use, multi-tenancy or other vital services; but better processing and response times to reduce customers' consuming cost and meet SLAs. Evidently, well-ordered resource allocation mechanisms, monitoring tools and load balancing algorithms are vital in order to thoroughly achieve the management objectives.

REFERENCES

[1] R. Buyya, J. Broberg and A. M. Goscinski (Editors), "Cloud Computing: Principles and Paradigms", John Wiley & Sons Inc., 2011.

[2] J B. Wickremasinghe, "MEDC Project Report 2009" Available from: http://www.Cloudbus.org/students/MEDC_Project_Report_Bhathiya_318282.pdf. [Accessed July 20, 2011].

[3] T. Mather, S. Kumaraswamy and S. Latif, "Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance", O'Reilly, 2009.

[4] S. Patidar, D. Rane, and P. Jain, "A Survey Paper on Cloud Computing", *Second International Conference on Advanced Computing & Communication Technologies* (ACCT), pp. 394 – 398, 2012.

[5] N. Antonopoulos and L. Gillam (Editors), "Cloud Computing: Principles, Systems and Applications: Computer Communications and Networks, Springer, 2010.

[6] T. Velte, A. Velte and R. Elsenpeter: "Cloud Computing: A practical Approach", McGraw Hill, 2010.

[7] C. W. Dawson, "Project in Computing and Information Systems: A Student Guide", Pearson Education Limited, 2005.

[8] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose and R. Buyya, "Research Article, CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms". First published online: 24 AUG 2010. John Wiley & Sons, Ltd, 2010

[9] B. Golden, "Virtualization for dummies", Wiley Publishing Inc. 2008.

[10] J. Hurwitz, R. Bloor:, M. Kaufman and F. Halper, "Cloud Computing for Dummies, Wiley Pushing Inc., 2010.

[11] F. Wuhib, R. Stadler and M. Spreitzer "Gossip-based Resource Management for Cloud Environments", *6th International Conference on Network and Service Management,* Ontario, Canada, October 25 – 29, 2010.

[12] VMware,Availablefrom:http://www.vmware.com/pdf/vmware_drs_wp.pdf. [Accessed August 05, 2011].

[13] The Cloud Computing and Distributed Systems (CLOUDS) Laboratory, University of Melbourne. Available from: http://cloudbus.org/cloudsim/, [Accessed May 26, 2011].

[14] T. Wood, "Sandpiper Technique", Available from: http://www.cs.umass.edu/~twood/pubs/NSDI07.pdf. [Accessed August 05, 2011].

[15] Youtube: Available from http://www.youtube.com. [Accessed August 22, 2011]

[16] Facebook:Availablefrom:http://www.facebook.com/press/info.php?statistics, (http://www.allfacebook.com/facebook-becomes-largest-display-advertiser-2010-05). [Accessed: August 22, 2011].