

# Cost and Time Evaluation of Load Balancing and Service Broker Strategies in Multiple Data Centers

Shipra Gupta  
Ph.D Research Scholar  
MRIU  
Faridabad  
Haryana, India

Indu Kashyap, Ph.D  
Dept. of CSE  
MRIU  
Faridabad  
Haryana, India

## ABSTRACT

Cloud computing is a boon for almost every sector as it reduces the overall expenditure by making I.T. infrastructure, services and platform on “on-demand” basis. The “pay as you go” policy of cloud technology enables the users to use and pay for the services and computing infrastructure only when they need it without worrying about the installation and underlying hardware or software infrastructure. Data centers are located at different locations and provide services to the users in their closest proximity. As the user has to pay for the service utilization, so selection of data center plays an important role for maximum resource utilization at minimum expenditure and best response time. An important feature in cloud is that according to user location and demand, the closest and the cheapest data center is assigned in that region. In this paper, the authors have analyzed the existing service broker policies i.e., Closest Data center, optimize response time, and reconfigure dynamically load, in data center selection.

## Keywords

Cloud computing, data center, load balancing, service broker, User base, Region.

## 1. INTRODUCTION

Cloud computing is a technology that aims at large amount of computing in a virtualized way. Users can access the cloud services and reduce the capital costs as they have to pay for the services they need as for the time period they need the service. NIST defined cloud computing as “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources(e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

According to Gartner cloud computing can be defined as “a style of computing where massively scalable IT-enabled capabilities are delivered ‘as a service’ to external customers using Internet technologies”.

The cloud is operated and managed at a data center owned by a service vendor that hosts multiple clients and uses dynamic provisioning.

Some benefits of cloud computing are:

*Pay per use Service:* Organizations employing the cloud services don’t own the infrastructure on which they need to execute their applications. They rent the application usage from the cloud provider. This enables the organizations to pay for only the services they need and use, and avoids for

paying for resources they don’t need, thus helping them avoid a large capital expenditure for infrastructure.

*Resource Utilization:* The cloud service provider that delivers some or all of the services required to the organization can also share infrastructure between multiple clients. This helps improve utilization rates by eliminating a lot of wasted server idle time. The shared use of very high speed bandwidth distributes costs, enables easier peak load management, often improves response times, and increases the pace of application development.

*No need of Maintenance infrastructure:* The organizations employing the cloud services do not have to worry about hardware, software or environment maintenance as the user does not have to manage and control the underlying configuration setting of the applications it uses.

*Mobility of data:* Organizations/ users employing cloud services can access their data easily from anywhere through internet, which enables the professionals to work 24X7 on their own ease without the hassels of being available on their workplace.

*Rapid Elasticity:* In cloud computing, the users/ organizations can dynamically scale their space both up and down without any effect on the ongoing services as in cloud environment, the capabilities and services appear to be unlimited to the consumer.

Along with the benefits, Cloud model do have some limitations. Some of these limitations can be summarized as:

*Security and Compliance:* Security is a major issue in cloud environment. As the client data is stored virtually on the cloud, so proper security parameters are employed to protect data from vulnerable attacks. Data encryption is one technique for data security in cloud.

*In-house control:* Moving the data on cloud means that the data is no more “in-house” and so every time the user needs the data, he has to extract the data from the cloud. The difficulty in extracting the data from cloud is one of the reasons why organizations having a second thought in adopting cloud computing.

*Internet connection dependency:* In cloud computing, the data is stored on cloud and be accessed through internet browser. So, speed and reliability of internet connection is a major issue. Proper speed and continuous connectivity is mandatory for professionals working on cloud.

## 2. CLOUD REFERENCE ARCHITECTURE

Fig. 1 represents the overview of NIST cloud reference architecture [1]. The fig defines five major actors, i.e., cloud consumer, cloud provider, cloud carrier, cloud auditor and cloud broker.

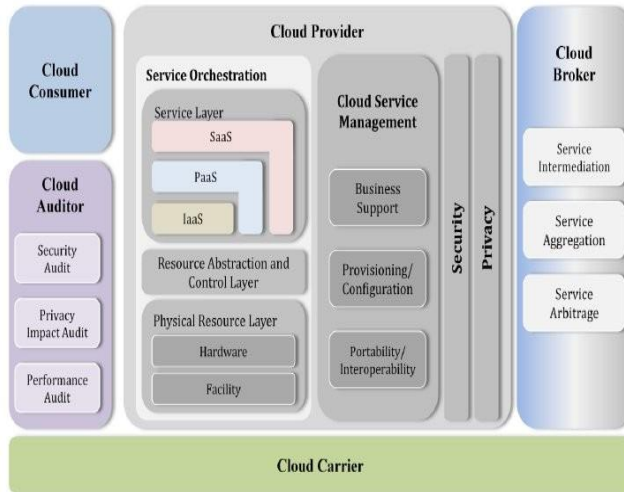


Figure 1. The Conceptual Reference Model

**Cloud Consumer:** This entity is defined as a person or organization which requests and uses services from cloud provider.

**Cloud Provider:** This entity is defined as a person or organization that makes the requested services available to the consumer.

**Cloud Auditor:** This entity is defined as a professional or organization which conducts independent audits of the availability, integrity and confidentiality of cloud services.

**Cloud Broker:** This entity can be defined as a professional or organization that intermediary between cloud consumer and provider. It maintains business relationships with various cloud providers and helps the consumer organization, the best services in terms of cost, performance and other QoS factors.

**Cloud Carrier:** It is intermediary entity responsible for providing connectivity and availability of cloud services between the cloud consumer and the provider.

In cloud computing, the user can request the level of services as per their requirement. So, the user requirements of services are broadly classified in terms of cloud delivery models.

## 3. CLOUD DELIVERY MODEL

Services are delivered in three different service models, i.e.

**SaaS (Software as a Service):** NIST defines SaaS as follows: “The capability provided to the consumer is to

use the provider’s applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser. The consumer does not manage or control the underlying cloud infrastructures including networks, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user specific application configuration settings.” The main advantages of SaaS can be the cost effectiveness in terms of licensing and maintenance. The practical example of SaaS is GoogleApps.

According to Google, Google Apps is a service from Google that provides independently customizable versions of several Google products using a domain name provided by the customer. It features several Web applications with similar functionality to traditional office suites, including Gmail, Google Calendar, Docs, Drive, Groups, News, Play, Sites, Talk, and Wallet.

**PaaS (Platform as a Service):** NIST describes PaaS as follows: “The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructures including networks, servers, operating systems, storage but has control over the deployed applications and possibly application hosting environment configurations.” The main advantage of PaaS is in SDLC (Software Development Life Cycle) as the developers can develop their applications without installing the development tools on their end. The practical example of PaaS is GoogleApp Engine.

According to Google, Google App Engine (GAE) is a platform as a service (PaaS) cloud computing platform for developing and hosting web applications in Google-managed data centers. Applications are sandboxed and run across multiple servers.

**IaaS (Infrastructure as a Service):** NIST describes IaaS as: “The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g. host firewalls).” The main advantage of IaaS can be the cost effectiveness and maintenance free infrastructure on demand basis. The practical example of IaaS is Amazon EC2.

According to Amazon, Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers. Amazon EC2’s simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon’s proven computing environment. Fig. 2 shows the configuration of these cloud delivery models in cloud environment.

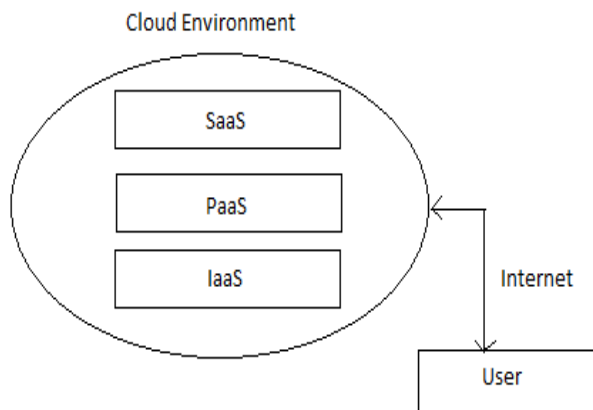


Fig. 2: Cloud delivery Models in Cloud environment

#### 4. CLOUD DEPLOYMENT MODEL

The cloud delivery models operate in different cloud deployment models which identify the cloud environment depending on size, access and ownership. There are four cloud deployment models, i.e.,

**Public Cloud:** A public cloud is a cloud computing environment available for use by general people. Fig. 3 shows a general view of public cloud. The examples of public cloud can be Amazon Web Services, Google App Engine.

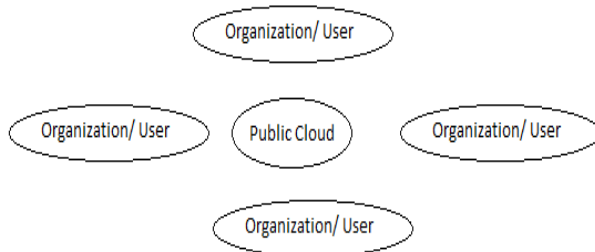


Fig. 3: Public Cloud

**Private Cloud:** NIST describes a private cloud as a cloud infrastructure operated solely for an organization, managed by the organization or a third party and existing either on premise or off premise. The private cloud is typically hosted within the boundaries of the owner organization. Fig. 4 shows a general view of private cloud.

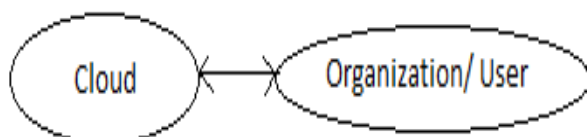


Fig. 4: Private Cloud

**Community Cloud:** The concept of community cloud lies in between public and private cloud. Community cloud constitutes of shared resources/ infrastructure between multiple number of organizations having shared requirements

such as pooling of high level security resources, firewalls etc.

**Hybrid Cloud:** NIST defines hybrid cloud as “a composition of two or more clouds (private, community or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability.”

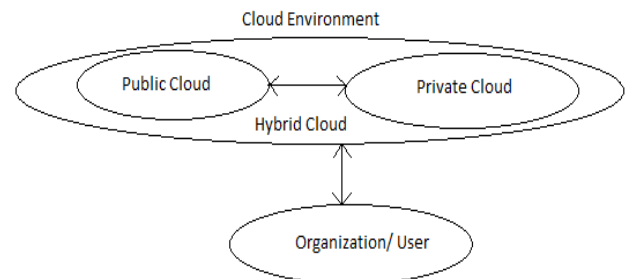


Fig. 4: Hybrid Cloud

#### 5. BUILDING BLOCKS OF CLOUD COMPUTING

Based on the above discussed parameters and models, a block diagram of cloud environment is shown in Fig. 5.

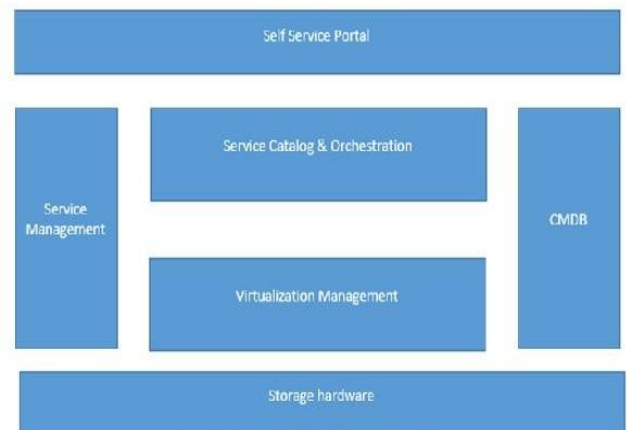


Fig. 5: Building Blocks of Cloud Computing

**Self Service Portal:** It is the web based interface between the user and the provider.

**Service Catalog and Orchestration:** The service catalog consists of the services offered to the user while the orchestration layer allows the user to control to rent the services.

**Virtualization Management:** Virtualization is a technique to partition the resources of a single physical machine into multiple virtual machine. To achieve virtualization, this layer consists of a hypervisor or a VMM (Virtual Machine Monitor). The basic functions of virtualization layer are [2]:

- Scheduling Virtual Machines
- Memory Management
- Emulating I/O devices and arbitrating access to them
- Network Packet Processing

*Service Management:* As the name suggest, this block/layer is responsible for virtual machine management service i.e. it manages all the VMs in the cloud environment.

*CMDB (Configuration Management Database):* CMDB is responsible for integrating all the service management activities.

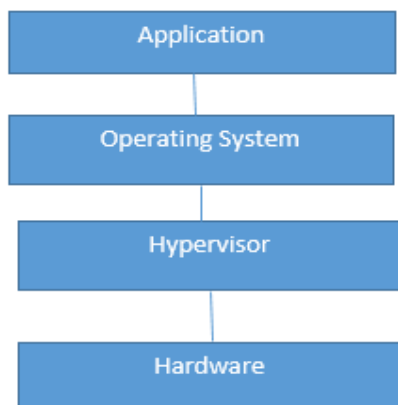
*Storage Hardware:* All VMs in a given environment are connected to the storage hardware. All activities of VMs are stored in the storage hardware which make up for the data in case of the virtual server failure as the other machines in the network copies the data and the user continues to work without any interruption.

## 6. HYPERVISOR

A hypervisor is basically a software that is used for creating and managing the virtual machines and is also known as virtual machine manager (VMM). It provides an abstraction layer that separates the VM from system hardware. In cloud environment, server is defined as host whereas the virtual machine is defined as guest machine. Hypervisor is a software that allows multiple operating systems to share a single hardware host. Each operating system appears to have the host's resources dedicatedly assigned to itself. The hypervisor controls the host processor and resources, allocating what is needed to each operating system such that the guest operating systems or the virtual machines do not disrupt each other so that if one VM is having some issues it does not affect the operation of the other.

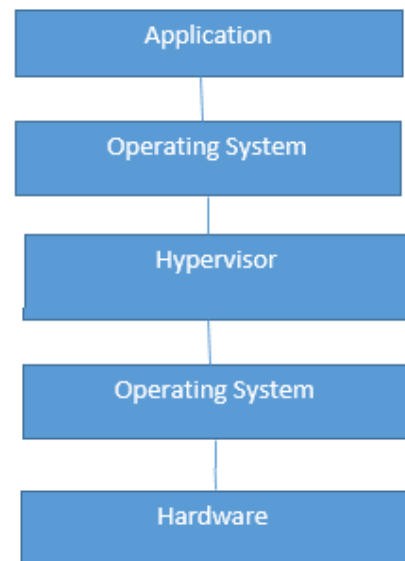
A hypervisor virtualizes the physical resources to the guest operating system. There are two types of hypervisors:

*Type1 Hypervisor:* It is also known as bare-metal hypervisor. It is installed directly on the hardware platform. After installation, it divides the hardware into a number of VMs with operating system on each VM. The guest OS runs on a layer above the type1 hypervisor. Fig. 6 illustrates the Type1 hypervisor. Example of Type1 (bare-metal) hypervisor is VmWare ESXi.



**Fig. 6: Type1 Hypervisor**

*Type2 Hypervisor:* It is a software level or hosted hypervisor. It is installed above the existing operating system. It installs host operating system as an application. Fig. 7 illustrates Type2 hypervisor. Example of Type2 hypervisor is VmWare Workstation.



**Fig. 7: Type2 Hypervisor**

A Type 1 Hypervisor is faster and more efficient as they eliminate the host OS layers processing time. Type1 hypervisors are also more secure as compared to Type 2.

Type 2 hypervisors are easier to set up as they have a hand on OS layer to work with.

There are a several number of parameters to be considered in cloud environment for the maximum utilization and cost control of the organization. In this paper, the authors have analyzed a number of scenarios in cloud environment and made a comparative study in terms of overall response time and cost of operations. In the study, the authors have considered the three load balancing policies i.e., Round Robin, Equally spread current execution load and throttled along with the three service broker policies i.e., closest data center, optimize response time and reconfigure dynamically with load. The authors have performed all the observations in CloudAnalyst simulator.

## 7. SIMULATION ENVIRONMENT

The authors have considered 5 user base in five different regions i.e., region 0, 1, 2, 3 and 4. The request per user per hour are different for each user base whereas the data size per request (bytes) has been considered as constant as 100 for each user base. The overall simulation environment is shown in Fig. 8.

The authors have considered three data centers in three different regions. The data center configuration is shown in Fig. 8 & 9. The advanced configuration of user bases as well as the data centers is illustrated in Fig.10. The internet characteristics representing the delay matrix and the bandwidth matrix is illustrated in Fig. 11.

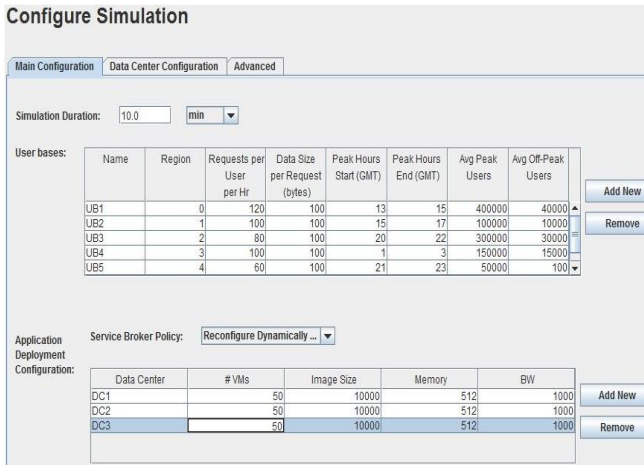


Fig.8: Simulation Environment

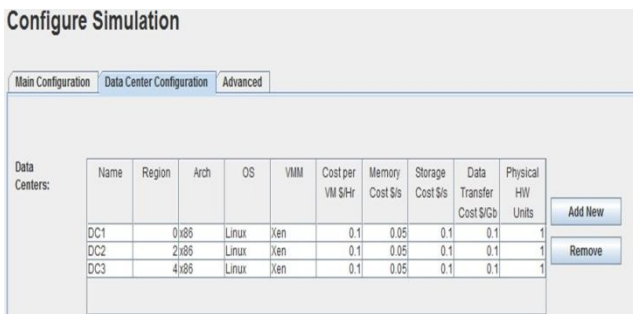


Fig. 9: Data Center Configuration

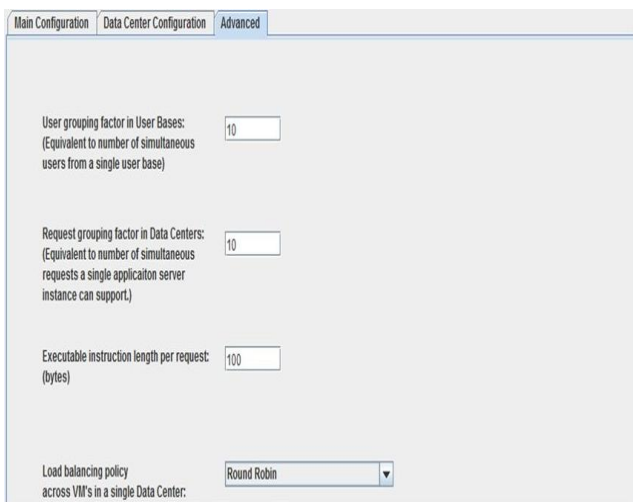


Fig. 10: Advanced Configuration

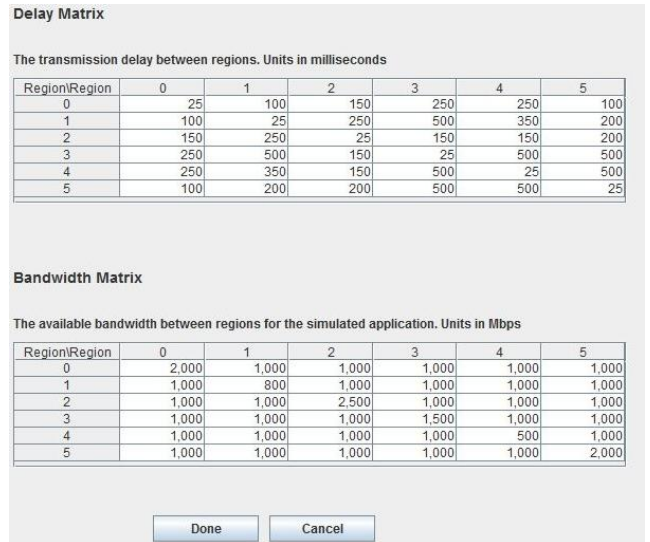


Fig.11: Internet Characteristics

## 8. SIMULATION RESULTS

The authors have analyzed the average response time and data center processing time for the three load balancing techniques along with the three service broker policies.

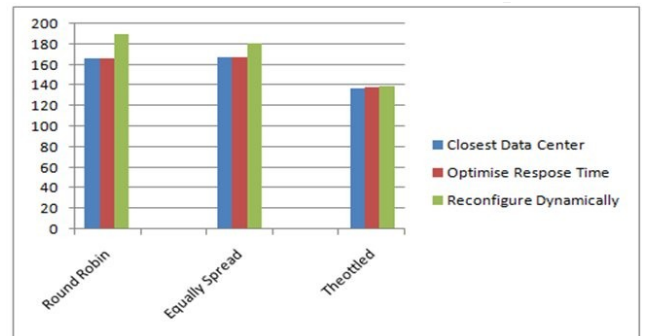


Fig.12: Average Response Time

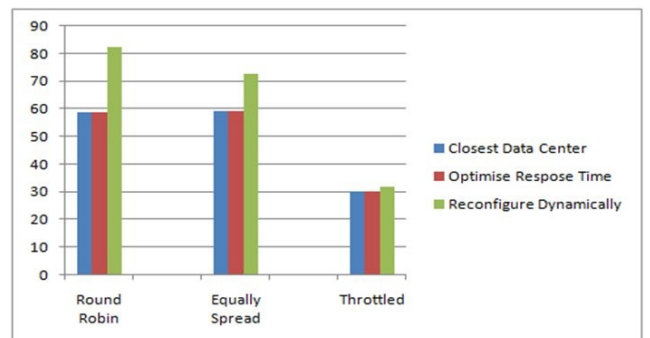


Fig. 13: Data Processing Time

Fig.12 shows that average response time for closest data center and optimize response time remains approximately equal for any of the load balancing techniques, whereas the fig shows that it increases remarkably for reconfigure dynamically in case of round robin and equally spread execution while the time difference is very low for throttled load balancing policy. It is quite evident that the average response time is minimum in case of throttled load balancing. The authors identify the same scenario for data processing time in Fig. 13.

So, it is evident that the overall response time minimizes in case of throttled load balancing policy in any of the service



broker policy. Further, we have analyzed the total data center cost for the three load balancing policies along with each service broker strategy.

Fig. 12 shows that the total data center cost is approximately the same for closest data center and optimize response time for all the three service broker policies, whereas it increases remarkably in reconfigure dynamically in all three service broker policies.

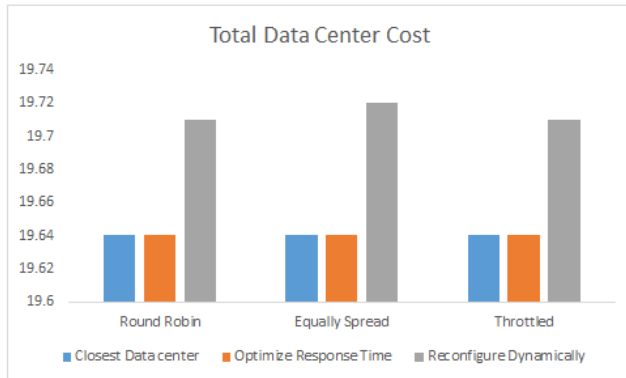


Fig. 14: Total Data Center Cost

In the graphs shown in Fig. 12, 13 and 14, the authors identified that the overall response time reduces for throttled load balancing but the overall cost of the data center is not optimized.

## 9. CONCLUSION AND FUTURE WORK

In the study, the authors have analyzed a scenario in cloud environment with the help of CloudAnalyst simulator. The results show that average response time as well as the data center processing time is remarkably reduced in case of throttled load balancing policy but the total data center cost does not show any noticeable difference from the other two load balancing policies. So, the authors analyze that the selection of load balancing policies along with the service broker strategies affects the overall response time but has the least impact on the total data center cost. Further In this paper, we have analyzed the scenario where the peak hours of the user bases have been assumed to be different and the overall response time has been evaluated, which is not practically always true. We have also considered identical configurations for all the data centers, which is again not a practically always true condition.

So, as a future work, the optimization of load balancing policies along with the service broker strategies such that along with the overall response time the total data center cost is also reduced can be a consideration. Further, the practical scenarios where the on-peak and off-peak hours of two or more data centers coincides and the data center configurations are also different can be considered.

## 10. REFERENCES

[1] Fang Liu, Jin Tong, Jian Mao, Robert Bohn, John Messina, Lee Badger and Dawn Leaf, NIST Cloud

Computing Reference Architecture, National Institute of Standards and Technology.

- [2] Eric Keller, Jakub Szefer, Jennifer Rexford and Ruby B. Lee, NoHype: Virtualized Cloud Infrastructure without the Virtualization, ISCA'10, June 19–23, 2010, Saint-Malo, France.
- [3] <http://www.cloudbus.org/cloudsim> : (Cloud Analyst can be downloaded from here)
- [4] Saif U. R. Malik, Samee U. Khan, and Sudarshan K. Srinivasan, Modeling and Analysis of State-of-the-Art VM-Based Cloud Management Platforms, IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 1, NO. 1, JANUARY- JUNE 2013
- [5] Joseph Doyle, Robert Shorten, and Donal O'Mahony, Stratus: Load Balancing the Cloud for Carbon Emissions Control, IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 1, NO. 1, JANUARY-JUNE 2013
- [6] Jenn-Wei Lin, Chien-Hung Chen, and J. Morris Chang, QoS-Aware Data Replication for Data-Intensive Applications in Cloud Computing Systems, IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 1, NO. 1, JANUARY- JUNE 2013
- [7] Xinyu Lei, Xiaofeng Liao, Tingwen Huang, Huaqing Li, and Chunqiang Hu, Outsourcing Large Matrix Inversion Computation to a Public Cloud, IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 1, NO. 1, JANUARY- JUNE 2013
- [8] Kousik Dasgupta, Brototi Mandal, Paramartha Dutta, Jyotsna Kumar Mondal, and Santanu Dame, "A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing", International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013, Elsevier Science Direct
- [9] Dhinesh Babu L.D., P. Venkata Krishna, "Honey bee behavior inspired load balancing of tasks in cloud computing environments", Applied Soft Computing, [elsevier.com/locate/asoc](http://elsevier.com/locate/asoc)
- [10] Johan Tordsson, Rubén S. Montero, Rafael Moreno-Vozmediano, Ignacio M. Llorente, Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers, Future Generation Computer Systems [www.elsevier.com/locate/fgcs](http://www.elsevier.com/locate/fgcs)
- [11] Bhathiya Wickremasinghe, Rodrigo N. Calheiros, and Rajkumar Buyya, "CloudAnalyst: A CloudSim-based Visual Modeller for Analyzing Cloud Computing Environments and Applications", 24th International Conference on Advance e Information Networking and Applications (AINA) IEEE Computer Society, 2010
- [12] Rajkumar Buyya, Christian Vecchiola, S. Thamarai Selvi, "Mastering Cloud Computing", Mc Graw Hill Publication, 2013