



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

A Study of Load Balancing Issue Among Multifarious Issues of Cloud Computing Environment

Sonika Matele¹, Navneet Singh², Dr.K.James Mathai³

Department Of Computer Engineering & Application
National Institute of Technical Teacher's Training & Research
Bhopal, Madhya Pradesh
INDIA

Abstract: We are all moving, adamantly and incrementally, in the direction of cloud computing today. Due to its simplicity, its services are so popular that they take place in our daily life too. Cloud computing integrates a range of different technologies that have developed through various computing techniques. Among multifarious issues, the issue of load balancing means that every machine in the network does equal amount of work at any given time, no machine gets overloaded in companion to another one.

This paper has been prepared as a part of the initial research study carried out by Sonika Matele for Master of Computer Technology and Application PG Programme to increase the performance of the given algorithms by varying different parameters or developing an innovative algorithm for provisioning of cloud computing resources.

Keywords: Cloud Computing, Data as a service, Virtual Machine, Throttled Load Balancing Algorithm

I. Introduction

Cloud Computing (CC) is an emerging computing model that provides the mobile access to wide area distributed resources on demand. The cloud computing technology is implemented in different architectures using different services and also various deployment models. The basic difference is that it can also co-exist with various other technologies and soft design approach. It is also a model for enabling convenient on demand network access to a shared pool of configurable computing resources.

This paper, after briefly stating some of the concerns of cloud computing today and then focuses on the efficient and effective utilization or provisioning of shared computing resources available on the cloud. To fully utilize the benefits of using cloud computing; the resource allocation and scheduling features must also be used that is provided by a particular cloud. Instead of having such fascinating features, cloud still has some serious issues which arise various question marks in its performance. [1] The emerging technology of cloud computing addresses the serious issue of load balancing for handling massive amounts of data, so there is a need of the techniques that can process huge amount of data in a proper manner. Whenever the processes are generated by the users, firstly a decision is taken for the appropriate datacenter and after that the choice of appropriate virtual machine (VM) take place. As the handling of processes takes place firstly at datacenter level and then at VM, so is quite necessary to take care of this problem at both these levels. Several types of algorithms are introduced in these papers which are used at both these levels to handle the load balancing problem. The basic idea behind these algorithms is to distribute the load in such a way that a particular machine did not get too overloaded in comparison to other ones.

II. Characteristics

The five important characteristics of cloud computing defined by NIST includes: [2]

- *On-demand self-service:* when a consumer of cloud wants a service then consumer requests to cloud and it provide the service according to need.
- *Global network access:* Access the cloud is globally allowed but it requires internet connection. Both high bandwidth and low latency are expected.
- *Distributed resource pooling:* The provider's computing resources are pooled to serve all consumers in a distributed manner with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.
- *Scalable:* Cloud is scalable in nature; number of users can add or remove according to their need.
- *Measured service:* They primarily derive from business model properties and indicate that cloud service providers control and optimize the use of computing resources through automated resource allocation, load balancing, and metering tools.

III. Types Of Clouds

There are different types of clouds that you can subscribe to; depending on your needs. As a home user or small business owner, you will most likely use public cloud services. [3]

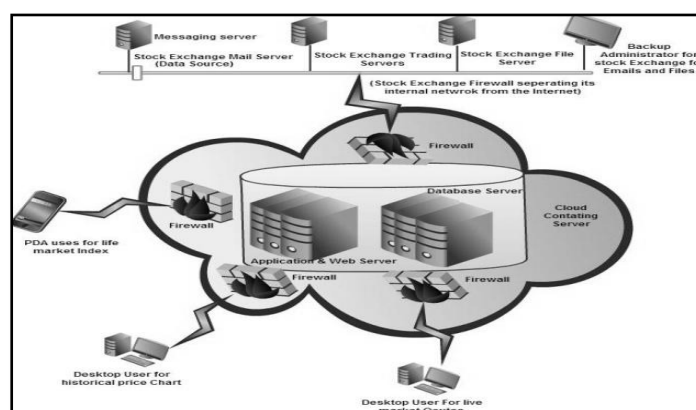
- **Public Cloud:** A cloud provider offers cloud services to the public at large.
- **Private Cloud:** An organization operates a cloud for its own use.
- **Community Cloud:** Here computing is done in a collaborative effort in which infrastructure is shared between several organizations.
- **Hybrid Cloud:** A hybrid cloud is essentially a combination of at least two clouds, where each one is either a mixture of public, private, or community.

IV. Cloud Services

Following are the services offered by the cloud:

- **Infrastructure as a service (IaaS):** IaaS gives storage and computing services as standardized services over the network. Storage systems and servers and computing devices are pooled.
- **Platform as a Service (PaaS):** In PaaS a layer of software or development environment is encapsulated & offered as a service, upon which other higher levels of service can be built.
- **Software as a service (SaaS):** In this model, a complete application/Software is offered to the customer, as a service on demand examples is Google docs and Microsoft.
- **Data as a Service (DAAS):** It is almost same as Software as a Service. Like all other members of the cloud family it is also based on the concept of on demand service. DaaS is based on the concept that the product, data in this case, can be provided on demand to the user regardless of geographic or organizational separation of provider and consumer. Additionally, the emergence of service oriented architecture (SOA) has rendered the actual platform on which the data resides also irrelevant. This development has enabled the recent emergence of the relatively new concept of DaaS. [4]
- **Security-as-a-service:** It is an outsourcing model for security management. Typically, it involves applications such as anti-virus software delivered over the Internet but the term can also refer to security management provided in-house by an external organization. [5] Benefits offered by Security-as-a-Service are as follows:
 - Constant virus definition updates that are not reliant on user compliance.
 - Extreme level security
 - Faster user provisioning.
 - Outsourcing of administrative tasks, allow an organization to devote more time to its core competencies.

Figure 1 : Virtualized Infrastructure



As shown in figure1; when a firewall security is provided in a cloud computing environment, the cloud service provider manages the live-stock program via back end connectivity to one of its exchange trading platform. The end user can view the live-stock reports using the browser and the website that is hosted by the cloud service provider. [20] This secure environment building a firewall that is also working as a service in the cloud computing environment is known as security as a service. This concept can be extended to a fraud detection system in cloud environment where a third party trusted server will first verify the user's authentication then pass it on to the cloud data center.

V. Technical issues needed to overcome

In spite of being providing such useful services, there are still lots of issues that need to be concerned with a cloud environment. A list of some such of these challenging issues in the cloud is given as follows:

- **Security:** As a lot of personal and secure information is saved in the personal computers of the users which is now being transferred to the cloud, so it becomes crucial to maintain proper security in the cloud computing environment. Some of the issues which are related to the security of the cloud are as follows: [6] [7]

- Privileged access
- Data location
- Data segregation
- Long-term viability
- Data confidentiality assurance
- Data integrity protection
- Guarantee of data availability

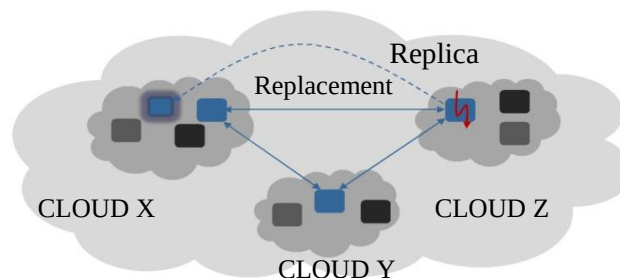
- **Technical Hardware & Software Expertise:** There are various types of users and their needs are also different. So as to fulfill the requirement we must customize our cloud services according to their need. For developing a quality suite application, a proper cloud expertise is required to fulfill business's needs. The availability of the computing components like hardware and software are needed to be ensured for realizing the benefits of cloud computing.

- **Load Balancing:** It is a mechanism to distribute the workload in such a fashion that it can provide appropriate resource utilization and also give better performance in terms of throughput etc. To achieve the entire aforesaid requirement it becomes necessary to use appropriate algorithms and methods, so the problem of overload can be tackled easily.

- **Fault Tolerance Awareness:** Fault tolerance is a major concern in the cloud computing so as to achieve availability, reliability and transparency of critical files and services. Fault tolerance awareness is used to predict the failures and take appropriate action before the actual failure occurs. [8] There are two different types of fault tolerant techniques which are as follows:

- Reactive techniques which reduces the effect of failure during execution of the application when actual failure occurs.
- Proactive fault tolerance which avoids the recovery from faults, errors and failures rather than waiting for the failure to occur and then reducing its effect.

Figure 1: Fault Tolerance Cloud Computing Environment



Fault tolerance in a cloud computing environment is same as that of a distributed computing environment. Here above figure 2; tries to depict the replication of services, where if one server or service at location Z fails then the replacement can be done at cloud X or at cloud Y. The advantage of this replication is that the complete process is transparent to the user i.e. the user is unknown with the failure of the service and migration of the service from cloud Z to cloud X.

- **Loss of control:** Cloud users manage their IT infrastructure through their relationship with their Customer Service providers (CSPs) and through service level agreements (SLAs). This requires skills that IT organizations typically do not possess today, so they will need to reinvent themselves to make this shift.

VI. Problem undertaken

From all the above mentioned issues, this paper is centric to the issue of the load balancing. It is one of the core and the crucial issue in the cloud computing environment. Actually in the cloud, users are allowed to utilize the

potential of other nodes on the internet and then use the same node's services to perform the desired task. [9] Thus load balancing is the mechanism that decides which requesting nodes/client will use which virtual machine and which one will be put on hold. It can be done individually as well as on grouped basis. Furthermore, it is also required to minimize the cost of the machine and maximize the profit for the service being offered. [10]

The various goals of load balancing are: [11]

- Improvement of performance
- Backup Recovery
- To stable the system
- For Further Modifications

A lot of algorithms have been proposed to do this task. But there are some important things which might be considered while the developments of such algorithm like:

- Estimation of load
- Comparison of load
- Stability of different system
- Performance of system
- Interaction between the nodes
- Nature of work to be transferred
- Selecting of nodes and many other ones. [11]

Performance of various load balancing algorithms can be analyzed by different cloud simulators. CLOUD ANALYST is a CLOUD SIM based tool which is developed by Clouds Laboratory and provide a simulation framework for research work. [12] [13] It uses algorithms at two different levels. First one is at datacenter level where various broker algorithms decide which datacenter is taken into consideration for the processing of the incoming request. Three different broker algorithms are used at this level namely as Service Proximity Based Routing, Performance Optimized Routing, Dynamic Service Broker. Second is at Virtual Machine (VM) level where VM management and load balancing of VM's (within a single datacenter) is taking place. Various algorithms are used to choose a particular VM in a single datacenter to process the incoming request, namely as Round-Robin Algorithm, Equally Spread Current Execution Algorithm, Throttled Load Balancing Algorithm.[14] The above algorithms can be understood as explained below.

Data Centre Load Balancing Algorithms

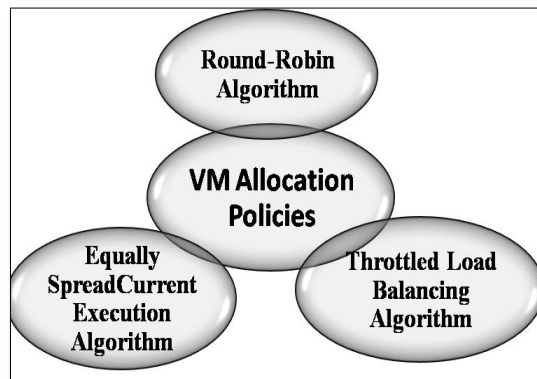
- *ServiceProximityServiceBroker*: It maintains an index table of all Data Centers indexed by their region. When the Internet receives a request it queries the *ServiceProximityServiceBroker* for the destination Datacenter. Algorithm firstly retrieves the region of the incoming request and queries for the region proximity list for that region. This list orders the remaining regions in the order of lowest network latency first when calculated from the given region. The algorithm picks the first data center located at the earliest/highest region in the proximity list. If more than one data center is located in a Region, one is selected randomly.
- *Performance Optimized Routing*: It is a *BestResponseTimeServiceBroker* which maintains an index of all Data Centers available. When the Internet receives a request, it queries the *Best Response-TimeServiceBroker* for the destination datacenter. Then it identifies the closest (in terms of latency) data center using the *ServiceProximity-ServiceBroker* algorithm. Then it iterates through the list of all data centers and estimates the current response time at each data center. If the least estimated response time is in the closest data center, the *Best Response-TimeServiceBroker* selects the closest data center. Else, it picks either the closest data center or the data center with the least response time.
- *DynamicServiceBroker*: It should extend either *ServiceProximityServiceBroker* or the *Best Response - TimeServiceBroker*. Firstly it maintains a list of all data centers and another list with the best response time recorded so far for each data center. When request arrives it queries the *DynamicServiceBroker* for the destination datacenter. The *DynamicServiceBroker* uses the *ServiceProximityServiceBroker/ BestResponse-TimeServiceBroker* algorithm to identify the destination. The *DynamicServiceBroker* updates the best response time records if the current response time is better than previous.

VM Level Load Balancing Algorithms:

- *Round-Robin Algorithm*: In this algorithm [15] the process are divided among processors in a round robin order. The distributions of all the processes are equal but their processing time is different. So at any point of time some nodes may be heavily loaded and others remain idle. This algorithm is mostly used in web servers where HTTP requests are of similar nature and distributed equally.

- **EquallySpreadCurrentExecution Algorithm:** This algorithm process handles with priorities. It distributes the load randomly by checking the size and transfer the load to that virtual machine which is lightly loaded or handle that task easier and take less time, and give maximize throughput. It is spread spectrum technique in which the load balancer spreads the load of the job at hand into multiple virtual machines. [16]
- **Throttled Load Balancing Algorithm (TLBA):** The TLBA maintains a record of the state of each virtual machine (busy/idle). If a request arrived concerning the allocation of virtual machine, then it sends the ID of the ideal VM to the data center controller and data center controller allocates the ideal virtual machine. [17]

Figure 3: Various VM Allocations polices used in the cloud



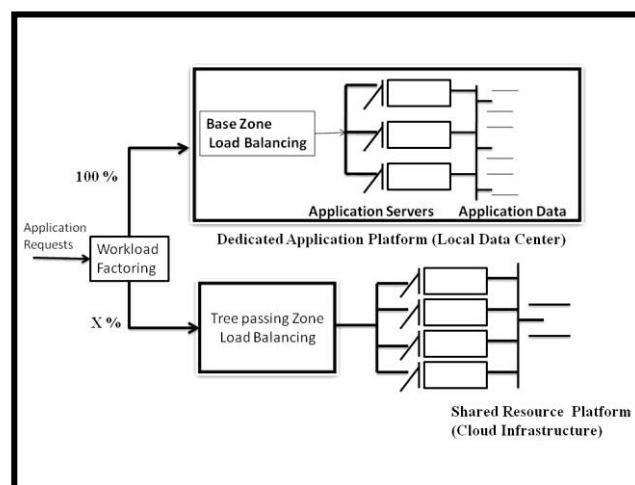
The figure 3; shows the main allocation policies which are applied at the virtual machine level. The performance of all the above mentioned algorithms can be raised by modifying their various parameters. As there is always a requirement of efficient and effective techniques, hence further modification can rectify the issue of load balancing in a cloud environment.

VI. Model for load balancing

Several models are applied for the purpose of load balancing. One of the most common models is discussed below: In figure 4; the model [18] describes two types of load balancing. One is trespassing load balancing and another one is based zone load balancing. The base zone is the dedicated application platform in the local data center whereas the trespassing is the application platform in a cloud data center.

The base zone runs all the time and processes the applications, the local data center works in more compact and highly utilized environment. Now here the trespassing differs from the base zone as it is **in demand** and works for transient period. It is expected to be utilized only in rare time.

Figure 4: Application Hosting Platform Architecture in Cloud computing Environment



VII. Conclusion

Cloud computing is a vast concept and load balancing plays a very important role in case of clouds. There is a huge scope of improvement in this area. In this paper many load scheduling algorithms have been discussed that can be applied to a cloud environment, but there are still so many other techniques that can be applied to balance the loads in clouds. The performance of the given algorithms can also be increased by varying different parameters or developing an innovative algorithm for provisioning of cloud computing resources.

References

- [1] Kevin Hamlen "Security Issues for Cloud Computing" The University of Texas at Dallas, USA.
- [2] Mell, P., Grance, T. "The NIST definition of Cloud Computing" version 15. National Institute of Standards and Technology (NIST), Information Technology Laboratory (October 7, 2009)
- [3] AlexaHuth and JamesCebula "The Basics of Cloud Computing", www.us-cert.gov/reading.../USCERTCloudComputingHuthCebula.
- [4] Data_as_a_service: Wikipedia
- [5] Security-as-a-Service www.searchsecurity.techtarget.com
- [6] Puneet Jai Kaur and Sakshi Kaushal, "Security Concerns in Cloud Computing", University Institute of Engineering and Technology, Punjab University, Chandigarh
- [7] Brodtkin, J., "Gartner: Seven cloud-computing security risks", Infoworld 2008, www.infoworld.com
- [8] Anju Bala, Inderveer Chana, "Fault Tolerance- Challenges, Techniques and Implementation in Cloud Computing". Computer Science and Engineering Department, Thapar University Patiala, Punjab, India
- [9] Anthony T. Velte, Toby J. Velte, Robert Elsenpeter, "Cloud Computing A Practical Approach by Tata McGraw-HILL" Edition 2010
- [10] Tanveer Ahmed, Yogendra Singh "Analytic Study of Load Balancing Techniques Using Tool Cloud Analyst", University School of Information Technology New Delhi, India
- [11] Ali M. Alakeel "A Guide to Dynamic Load Balancing in Distributed Computer Systems", IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.6, June 2010.
- [12] Rajkumar Buyya, Rajiv Ranjan and Rodrigo N. Calheiros, "Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges and Opportunities"
- [13] Bhathiya Wickremasinghe, Rodrigo N. Calheiros, and Rajkumar Buyya "CloudAnalyst: A CloudSim-based Visual Modeller for Analyzing Cloud Computing Environments and Applications"
- [14] Bhathiya Wickremasinghe. "Cloud Analyst: A CloudSim-based Tool for Modelling and Analysis of Large Scale cloud Computing Environments". A MEDC Project Report under the guidance of Rajkumar Buyya., June-2009
- [15] Zhong Xu, Rong Huang, (2009) "Performance Study of Load Balancing Algorithms in Distributed Web Server Systems", CS213 Parallel and Distributed Processing Project Report
- [16] Ms. NITIKA, Ms. SHAVETA, Mr. GAURAV RAJ "Comparative Analysis of Load Balancing Algorithms in Cloud Computing", International Journal of Advanced Research in Computer Engineering & Technology Volume 1, Issue 3, May 2012:
- [17] Jasmin James "Efficient Vm Load Balancing Algorithm For A Cloud Computing Environment" International Journal of Computer Science and Engineering (IJCSE), Vol. 4 No. 09 Sep 2012, P-1658- 1663
- [18] Hui Zhang; Guofei Jiang; Yoshihira, K.; Haifeng Chen; Saxena, A., "Intelligent Workload Factoring for a Hybrid Cloud Computing Model," Services - I, 2009 World Conference on, vol., no., pp. 701,708, 6-10 July 2009 IEEE transactions.
- [19] Ronald L. Krutz and Russell Dean Vines: Cloud Security "A Comprehensive Guide to Secure Cloud Computing" Published by Wiley Publishing, Inc. 10475 Crosspoint Boulevard
- [20] Information System Audit And Control Association, Issue 2009, Volume 6, An Auditors Perspective

Acknowledgments

I would like to give a special thanks to my guide Dr.K James Mathai, who always support and encourage me towards this work. He has guided me in all the aspects whether it is technical or non-technical. Also I would also like to thank my family and all my friends who directly or indirectly helped in completing this paper.