International Journal of Engineering Research & Management Technology ISSN: 2348-4039

March- 2015 Volume 2, Issue-2

Email: editor@ijermt.org www.ijermt.org

Survey paper on cloud computing with load balancing policy

Anant Gaur, Kush Garg
Department of CSE
SRM University, NCR Campus

ABSTRACT
Cloud computing is an attracting technology in the field of computer science. The cloud is changing our life by providing users with new types of services. In cloud computing the most important process is load balancing which leads to faster response to requests of cloud users. Some of the algorithm will developed by using certain newer approaches that can be used to get a measurable improvement in the system response time as compared to the traditional approaches. Load balancing is the process of reassigning the total loads to the individual nodes of the collective system to make the best response time and also good utilization of the resources.
The optimization of load balancing policies along with the service broker strategies such as overall response time, total cost of data center is also reduced. In this experimental 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 with using the different broker policy relevant to the load balancing policy in cloud computing using cloud analyst simulation.

Keywords-Load balancing, Load balancing algorithms, cloud analyst, response time, cost, virtual machines, data center and broker policy.

1. INTRODUCTION

Cloud computing is one of the fastest implementing technology in the decade. Many companies are trying to implement and introduce clouds, due to its simple and flexible architecture. These result in the increasing number of users reaching cloud. Although clouds are bifurcated in public private and hybrid models but still problem of reliability may arise in these clouds. Cloud computing has been adopted by organization which includes, social networking websites, online application design by Google app managers and by Google doc which are some of the important implementation and a step ahead in cloud computing. Some clouds are also designed for online software testing. This all suggests that cloud computing will change the way we interact with the resources via Internet. Cloud models used virtualization technology. This technology helps in slicing a single data centre or high power server to act as multiple machines. It depends on the hardware configuration of the data centre or server in how may virtual machine they can be divided. To implement virtualization, additional software is also required. This software is the system software an operating system, can be from windows for example windows server 2008 or Hyper-V or for an open source environment like Linux [1].
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 [2].

Benefits and barriers of cloud computing:

Progress of Cloud Computing is enormous with respect to personal uses and business uses. Users of cloud computing can utilize or maintain the online resources. Among several advantages or benefits, few are discussed below:

a) Scalability: Scalability is the capability of a system to increase total throughput under an increased load when resources are added. Resources can be hardware, servers, storage, and network. The user can quickly scale up or scale down the resources in cloud computing according to their need without buying the resources.

b) Virtualization: In cloud computing, virtualization is a concept where users have a single view of available resources irrespective of their arrangement in physical devices. So it is advantageous for providing the service towards users with less number of physical resources.

c) Mobility: Cloud Computing means mobility because users can access applications through internet easily at any point of time.

d) Low Infrastructure Costs: The pay-per-usage model is supported in cloud computing. It actually helps an organization to pay for the resources they need, not to make any investment for the resources available in the cloud. Moreover, the provider does not require any infrastructural maintenance or upgrade costs.

e) Increased Storage: Users or clients in cloud computing can store more data in cloud than on private computer systems, which they use on regular basis. It not only relieves them from buying extra storage space, but also improves performance of their regular system, as it is less loaded. On the other hand, data or programs are accessed anytime through internet, since they are available in cloud.

Thus, the cloud computing provides several advantages with form of elasticity, availability and expandability on-demand. Still it has some constraints or limitations as discussed in the following:

a) Latency: Low latency has always been an important consideration in telecom networks for voice, video and data. As cloud-based architecture can easily be accessed through the internet, so high latency is an important issue in every communication between client and provider.

b) Platform or Language constraints: Adaptation of platform or language always plays an important role. Till today, cloud providers support specific language or platform that does not interoperable with other providers. So a universal set of standards needs to be defined in case of language or platform adaptation.

c) Resource Control: Controlling of resources in cloud is not always in scope of client. It may vary between different cloud providers. Sometimes resource isolation is very much needed but it is very hard to isolate for the client to identify the exact resource. At the same time, resources may exhaust for keeping data or providing services, so data or program may need to migrate over other resources. This is also major and challenging issues in cloud computing. So, controlling resources and distributing loads through migration (if possible) between different resources is very much essential [3].

2. ARCHITECTURE OF CLOUD

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

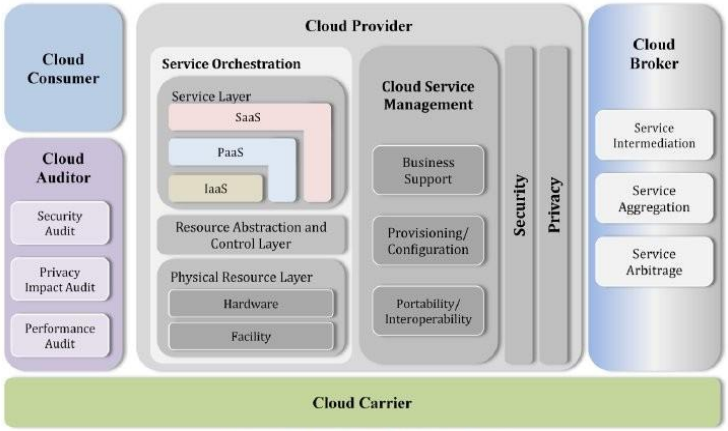

Figure 1. The Conceptual Reference Model

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 [2]

3. SERVICE MODELS CLOUD COMPUTING

Cloud computing provides their offers according to several models:

a. Infrastructure as a Service (IaaS),
b. Platform as a Services (PaaS),
c. Software as a Services (SaaS)

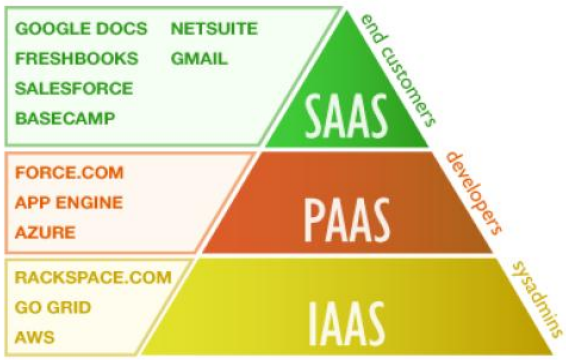

Fig 2: Cloud computing models

a. **Infrastructure as a services (IaaS)**
In IaaS grids clusters, virtualized server, its computational resources- CPU's, memory, network, storage and system software are delivered as a services [4]. Perhaps the best known example is Amazon's Elastic Computer Cloud (EC2) and Simple Storage Service's (S3) which provides (managed and scalable) resources as services to the user.

b. **Platform as a Services (PaaS)**
It typically makes use of dedicated API's to control the behavior of a server hosting engine which executes and replicates the execution according to user request eg. force.com, Google App Engine.

c. **Software as a Services (SaaS)**
Standard application software functionality is offered within a cloud. Eg. Google Docs, SAP Bussiness by design Load Balancing is one of prerequisites to utilize the full resource of parallel and distributed systems [4].

4. LOAD BALANCING

Distribute workload of multiple network links to achieve maximum throughput, minimize response time and to avoid overloading. We use three algorithms to distribute the load and check the performance, time and cost.

A. Round Robin Algorithm

Round robin algorithm is random sampling based. It means it selects the load randomly in case that some server is heavily loaded or some are lightly loaded.

B. Equally Spread Current Execution Algorithm

Equally spread current execution algorithm process handle with priorities. It distribute the load randomly by checking the size and transfer the load to that virtual machine which is lightly loaded or handle that task easy and take less time[5] , and give maximize throughput. It is spread spectrum technique in which the load balancer spread the load of the job in hand into multiple virtual machines.

C. Throttled Load Balancing Algorithm

Throttled algorithm is completely based on virtual machine. In this client first requesting the load balancer to check the right virtual machine which access that load easily and perform the operations which is give by the client or user. In this algorithm the client first requests the load balancer to find a suitable Virtual Machine to perform the required operation [5].

5. BROKER POLICIES IN CLOUD COMPUTING

The entire process of serving a client is a part of any one of the services defined in the service model. It begins with a request for a particular resource or application, be it for development, or just accessing the storage of the service provider [6]. The request is serviced by the cloud service provider through a series of steps, the first one passing through a cloud service broker, which acts as the intermediary between a cloud consumer and the cloud service providers. The service broker makes use of any one of the available service broker policies in order to send the request to the most appropriate data centre [6].

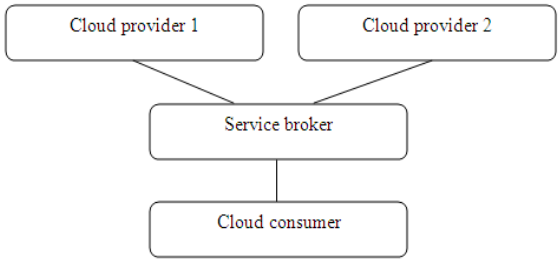

Figure 3: Role of service broker

After choosing the data centre that is going to perform computation, the load balancer at the data centre comes into action. It makes use of the implemented load balancing algorithms to select the appropriate virtual machine to which the request has to be sent for execution. The innermost abstraction layer comprises virtual machine management. The hypervisor or the virtual machine manager is responsible for the management and migration of virtual machines in the cloud data centers.

Out of the above tasks, the use of an efficient service broker policy is quite necessary to ensure that the later tasks are carried out with efficiency and least response time. Literature shows that quite a lot of research has been done in this regards. The three most frequently used service broker policies are explained below:

A. Closest data centre policy: This policy makes use of the concept of region proximity in selecting the data centre to which the user request has to be guided [6]. A region proximity list is maintained by making use of the "lowest network latency first" criteria to set the order of occurrence of data centers in the list. The data centre that occurs first in the list, i.e., the closest data centre is selected to fulfill the request using this policy. In case more than one data centers with the same latency are available, a random selection of the data centers is made. This policy is therefore beneficial in case the request can be satisfied by a data center that is quite close or within the same region.

B. Optimal response time policy: This service broker first identifies the closest data centre by making use of the network latency parameter, as in the previous policy. Then, the current response time is estimated for each of them. If the estimated response time is the one for the closest data centre, then the closest data centre is selected. Otherwise, the closest data centre or the data centre with the least response time is selected with a 50:50 chance of occurrence.

C. Dynamically reconfigurable routing with load balancing: This broker policy makes use of the current execution load in order to scale the application deployment. It also increases or decreases the number of virtual machines accordingly. The router needs to hold an additional responsibility of scaling the application deployment. This is done based on the load that it is currently facing. Scaling is done in this policy by considering the current processing times and the best processing time ever achieved [6].

6. LOAD BALANCING ALGORITHM

Equally spread current execution algorithm: Equally spread current execution algorithm process handle with priorities. It distribute the load randomly by checking the size and transfer the load to that virtual machine which is lightly loaded or handle that task easy and take less time , and give maximize throughput. It is spread spectrum technique in which the load balancer spread the load of the job in hand into multiple virtual machines [5].

- In this algorithm load balancer makes effort to preserve equal load to all the virtual machines connected with the data centre.
- Load balancer maintains an index table of Virtual machines as well as number of requests currently assigned to the Virtual Machine (VM).
- If the request comes from the data centre to allocate the new VM, it scans the index table for least loaded VM.
- In case there are more than one VM is found than first identified VM is selected for handling the request of the client/node, the load balancer also returns the VM id to the data centre controller.

ALGORITHM:

Step1:- find the next available VM.

Step2:-check for all current allocation count is less than max length of VM list allocate the VM.

Step3:- if available VM is not allocated create a new one.

Step 4:- count the active load on each VM. Step5:- return the id of those VM which is having least load [END][7].

7. EXPERIMENTAL WORK

The experimental work is performed using simulation software named as cloud analyst. The simulation is based on the cloud sim simulator; cloud sim is based on java and consist of a GUI interface which helps in easy configuration of attributes required for experiment. The diagram shows the environment for cloud analyst simulation tool. The simulation comes with three important menus, configure simulation, define Internet characteristics and run simulation. This menu is for configuring the experiment and setting up the load balancing algorithms. Simulation tool is having options to switch algorithm according to the requirement[1].

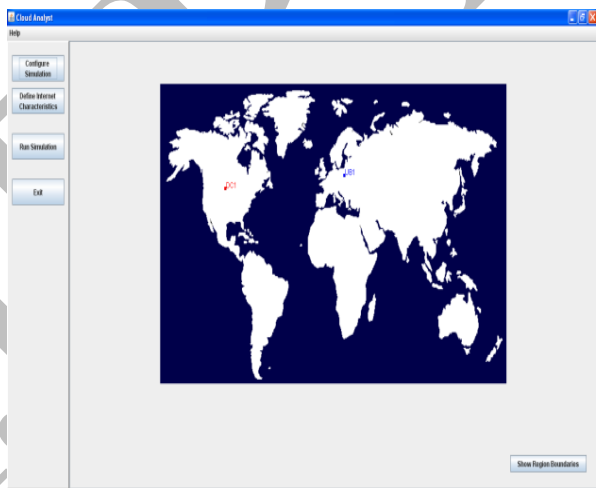

Fig 4: Cloud Analyst Interface

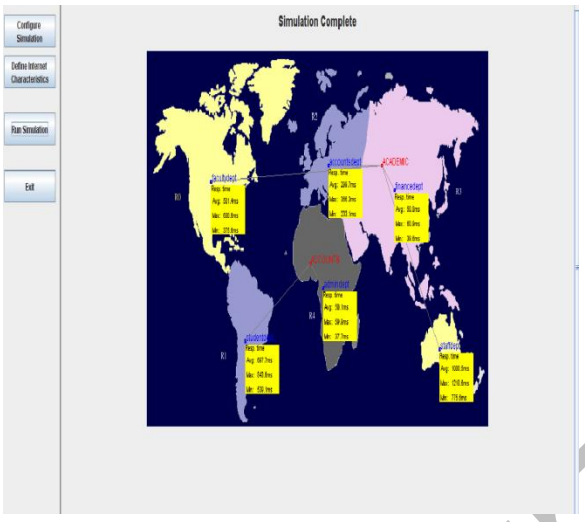

Fig 5: Partially simulation

8. CONCLUSION AND FUTURE WORK

The load balancing of the current system is one of the greatest issues in cloud computing. Various algorithms are used to solve this problem. In this paper we survey various existing load balancing methods.

9. REFERENCES

1. Dr. Hemant S. Mahalle, Prof. Parag R. Kaveri ,Dr.Vinay Chavan," Load Balancing On Cloud Data Centres" , International Journal of Advanced Research in Computer Science and Software Engineering,Vol.3,Issue 1,January 2013
2. Shipra gupta, ,Indu Kashyap,"Cost and time evaluation of load balancing and service broker strategies in multiple data centres"International journal of computer application,October 2014
3. Soumya Ray and Ajanta De Sarkar ,"Execution analysis of load balancing algorithm in cloud computing environment"International journal on cloud computing services and architecture"(IJCCSA),October 2012
4. Nusrat pasha and Dr. Ravi rastogi," Round Robin Approach for VM Load Balancing Algorithm in Cloud Computing Environment"International journal of advancd research in computer science and software engineering"(IJARCSSE),May 2014.
5. Ms.Nitika,"Comparative analysis of load balancing algorithms in cloud computing"International journal of engineering science.
6. Kunal kishore ,Vivek thapar"An efficient service broker policy for cloud computing environment" International Journal of Computer Science Trends and Technology (IJCST),August 2014.
7. Jaspreet kaur,"Comparison of load balancing algorithms in cloud".International journal of engineering research and applicationsMay-june 2012.