LOAD BALANCING WITH SWITCHING MECHANISM IN CLOUD COMPUTING ENVIRONMENT

M. Aruna[1], R. Punithagowri[2]

[1]*Assistant Professor (Sr.G), Erode Sengunthar Engineering College, Erode, India.*
[2]*Student, Computer Science and engineering, Erode Sengunthar Engineering College, India*

Abstract

Cloud Computing provides computation, software applications, data access management and storage resources without knowing its computing infrastructure. Cloud Computing is an emerging technology which has thousands of virtual machines consolidated to provide different services to the end users. End users can submit jobs or tasks to the cloud environment and these jobs are termed as Load which is executed by the cloud servers and the results are sent back to the end users. Load Balancing Algorithms are developed to distribute the workload evenly across all the nodes so that all the nodes are available and no node should be heavily loaded. Several Load Balancing algorithms are compared and the new approach called Switching Mechanism is applied to the partitioned Cloud Environment.

Keywords— *Cloud Computing, Load Balancing, Load Balancer, Main Controller, Partitions*

----------***----------

1. INTRODUCTION

All the large business and small business companies are moving to cloud environment because of its scalability. The jobs arriving to the Cloud Environment are executed by the large data centers which have thousands of blade servers. Cloud Computing is a Service Oriented Architecture (SOA). It provides different types of services to the users. Users can get the services with no need to know their infrastructure. That is, users do not know where the service is originated and its infrastructure. Users need to pay only for what they used from cloud in the form of services. This is the simplicity of Cloud.

There are 4 different types of cloud environment. They are,

- Public Cloud(Free of Cost, anyone can access)
- Private Cloud(Pay for what you used, only for single organization people)
- Hybrid Cloud(Combined both public & private Clouds)
- Community Cloud(For Communication purpose)

Users can access the cloud resources in the form of services. There are 3 basic services provided by the Cloud Environment. They are,

- Platform as a Service (PaaS)
- Software as a Service (SaaS)
- Infrastructure as a Service (IaaS)

1.1 Virtualization

Cloud Computing is based on the Concept of virtualization technology. Virtualization [1] means "something that is not real but gives all the facilities as a real one". It is the software implementation on the bare hardware so that the resources under the hardware can be utilized more effectively. Cloud Computing uses the virtualization technique to make use the cloud resources efficiently. Two types of virtualization can be used in Cloud Environment.

1. Full Virtualization
2. Para Virtualization

1.1.1 Full Virtualization

In Full Virtualization [1], the installation of one computer is done on the other computer. It will result in a virtual machine that have all the facilities and softwares that are present in the actual machine.

1.1.2 Para Virtualization

In Para Virtualization [1], the hardware allows multiple operating systems to be run on a single machine. For this, VirtualBox tool is used. Here all the services are not fully available, rather than the services are provided in a partial manner.

2. LOAD BALANCING

Load Balancing is a technique to balance the load across cloud environment. It is the process of transferring load from heavily loaded nodes to low loaded nodes. As a result, no node should be heavily loaded. Thereby it will increase the availability of nodes. If all the jobs are arrived to the single node, then its queue size is increased and it becomes overloaded. There is a need to balance the load across several nodes, so that every

node is in running state but not in overloaded state. The goals are as follows [2]:

- To increase the availability
- To increase the user satisfaction
- To improve the resource utilization ratio
- To minimize the waiting time of job in queue as well as to reduce job execution time
- To improve the overall performance of Cloud environment

Based on the current state of the system, load balancing algorithms can be divided into two types:

2.1 Static Schemes

The current status of the node is not taken into account [3]. All the nodes and their properties are predefined. The algorithm works based on the predefined information. Since it does not use current system status information, it is less complex and it is easy to implement.

2.2 Dynamic Schemes

This type of algorithm is based on the current system information [3]. The algorithm works according to the changes in the state of nodes. Dynamic schemes are expensive one and are very complex to implement but it balances the load in effective manner.

3. RELATED WORKS

Load Balancing is one of the main issues in Cloud Computing. Load Balancing is needed to distribute the workload to all the servers. There are several Load Balancing algorithms have been developed. Here, some of the Load Balancing algorithms are compared.

Table.1 Comparison of Load Balancing Algorithms

S.No.	Title	Description	Limitation
1.	A Load Balancing Model based on Cloud Partitioning for Public Cloud [4]	Divide the Public Cloud Environment into several Partitions	Does not specify the Cloud Division rules and refreshing time period
2.	Enhanced Equally Distributed Load Balancing Algorithm [5]	Based on the counter variable allocated to each server, the requests are handled	Does not consider the availability of the resources in each server
3.	A New Approach for Load Balancing in Cloud Computing [6]	Combination of both Min-Min & Max-Min algorithms	Two algorithms need to be executed
4.	Power Aware Load Balancing for Cloud Computing [7]	Executed by Cluster Controller, based on threshold value 75% creates new VMs on PMs	Does not scale up for large data centers
5.	Efficient VM Load Balancing Algorithm for Cloud Computing Environment [8]	Enhances AMLB by assigning weight to each VM & maintains the index table	For every request, parses the index table, thereby increase in response time
6.	Improved Max-Min Algorithm in Cloud Computing [9]	Assigns task with maximum executing time to resource produces minimum completion time	Smaller tasks has to wait for a long time to execute

7.	Comparison of Load Balancing Algorithms in a Cloud [10]	Cloud Manager receives all requests & places in a queue then check for the availability and allocate to VM	Maintaining all requests by a Cloud Manager creates a bottleneck problem
8.	Availability and Load Balancing in Cloud Computing [2]	XMPP servers & XMPP clients communicate using MOM. XMPP clients sends its current status information to XMPP servers	Need to improve MOM for Load Balancing
9.	Energy Efficient Virtual Machine Provision Algorithm for Cloud System [11]	Proposes Dynamic Round Robin Algorithm which uses 2 rules to consolidate Virtual Machines	Does not scale up for large data centers

The Table.1 shows the comparison of different existing Load Balancing techniques and their limitations. Each technique has some limitation. Hence, the new approach need to be developed that enhances the Cloud Environment in all aspects.

4. PROPOSED WORK

All the Load Balancing Algorithms are applied to the entire Cloud Environment. Hence, it is difficult to manage the large Cloud Environment. Dynamic Load Balancing algorithms distribute the workload by checking the status of each server which is maintained by a table called Status table. Maintaining Status table for all the servers is a difficult one and it may not be updated periodically hence it may not be consistent one.

The proposed work is to apply Load Balancing algorithms only for the part of the cloud servers and not to the entire Cloud Environment. So that, the overhead of maintaining cloud servers is reduced. To distribute the workload in the partitioned Cloud Environment, two main components are needed [4]. They are,

1. Main Controller
2. Load Balancer

Before discussing about these components, the method for partitioning the cloud environment is to be known. The entire cloud environment is divided into partitions based on its geographical locations. Data Centers are reside in the country and each country has its own and unique co-ordinate value. Based on this co-ordinate value, the Cloud Environment is divided into partitions. The co-ordinate values are represented as $35^{O}41"22.22'$ N 139^{O} $41"30.12'$ E. Here, the 3 numeric values represents the degree, minutes and seconds values and it also represents the four directions East, West, North and South.

The Cloud is divided into four partitions as depicted in Fig.1. Partition-I includes the data centers that are reside in the North-West direction. Partition II includes the datacenters that are resided in the North-East direction. Partition III includes the data centers that are reside in the South-West area and Partition IV includes the data centers that reside in the South-East area. For Example, the co-ordinate value of Tokyo is $35^{O}41"22.22'$ N 139^{O} $41"30.12'$ E which reside in the North-East direction, hence Tokyo belongs to the Partition-II. In the representation of co-ordinate, it has three parts as degree, minutes and seconds.

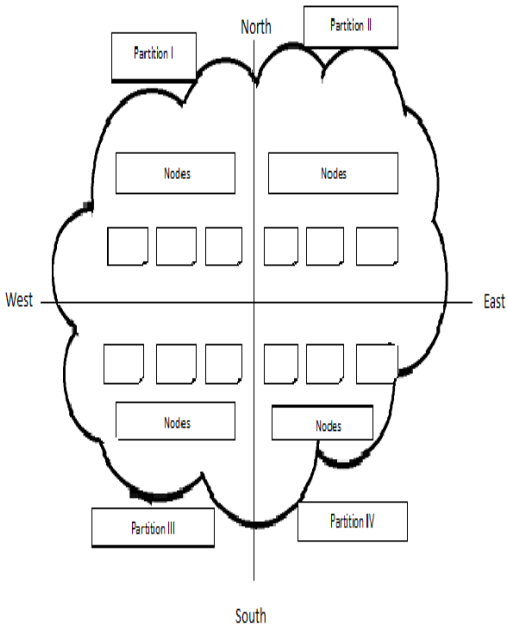

Fig 1 Cloud Partitions

End users from different locations submit their jobs to the Cloud Environment. All these jobs are received by a Main Controller which is a single node to manage all the partitions. Nodes under each partition are managed by a Load Balancer. Main Controller distributes the jobs to the Load Balancer by checking its partition status. The partition may be in 3 states as

Idle, Normal and Overload states. The partition status is set by the Load Balancer based on the parameters as Number of CPUs, the CPU processing speeds, the available memory size, the memory utilization ratio, the CPU utilization ratio and network bandwidth etc.,

The jobs are received by the Load Balancers and the Load Balancing algorithms are applied to the partitions. Here the Switching Mechanism is applied. Switching Mechanism is the process of switching over to the 2 different algorithms according to the 2 different situations. Switching Mechanism contains two different algorithms, one simple algorithm for Idle state partitions and another one effective algorithm for Normal state partitions. Round Robin is a simple and cheap algorithm that can be used for Idle state partitions. An effective algorithm for Normal state partition should prevent the partitions becoming overloaded state.

5. CONCLUSIONS AND FUTURE ENHANCEMENT:

By partitioning method, the overhead in the management of cloud environment is reduced. Based on the co-ordinate values, Cloud Environment is divided into partitions. Cloud division rule [4] issue is resolved by using co-ordinate values for partitioning. For Idle state partitions, simple existing algorithm can be used but for Normal state partitions, a new algorithm is needed to be developed. In future, the concept is to be implemented using CloudSim tool and also to propose an effective algorithm that will work for Normal state partitions.

REFERENCES

[1] Ratan mishra & anant jaiswal, "ant colony optimization: a solution of load balancing in cloud", international journal OF web & semantic technology, VOL. 3, NO. 2, PP. 33-50, 2012.

[2] Zenon chaczko, venkatesh mahadevan, shahrzad aslanzadeh AND christopher mcdermid, "availability and load balancing in cloud computing", international conference ON computer AND software modeling, VOL. 14, PP. 134-140, 2011.

[3] Venubabu kunamneni, "dynamic load balancing for the cloud", international journal OF computer science AND electrical engineering, VOL. 1, NO. 1, PP. 33-37, 2012.

[4] Gaochao xu, junjie pang & xiaodong fu, “a load balancing model based on cloud partitioning for public cloud ”, ieee transactions ON cloud computing, VOL. 18, NO. 1, PP. 34-39, 2013.

[5] Shreyas mulay & sanjay jain, “enhanced equally distributed load balancing algorithm for cloud computing”, international journal OF research IN engineering AND technology, VOL. 02, NO. 06, PP. 976-980, 2013.

[6] Mohana priya.s & subramani.b, "a new approach for load balancing in cloud computing", international journal OF engineering and computer science, VOL. 2, NO. 5, PP. 1636-1640, 2013.

[7] Jeffrey m. Galloway, karl l. Smith & susan s. Vrbsky(2012), "power aware load balancing for cloud computing", international journal OF science AND research, VOL. 1, 2012.

[8] Jasmin james & dr. Bhupendra verma, "efficient vm load balancing algorithm for cloud computing environment", international journal OF computer science AND engineering, VOL. 4, NO. 09, PP. 1658-1663, 2012.

[9] Elzeki.o.m, reshad.m.z & elsoud.m.a, "improved max-min algorithm in cloud computing", international journal OF computer applications, VOL. 50, NO. 12, PP. 22-27, 2012.

[10] Jaspreet KAUR, "comparison of load balancing algorithms in a cloud", SINTERNATIONAL journal OF engineering research AND applications, VOL. 2, NO. 3, PP. 1169-1173, 2012.

[11] Ching-chi lin, pangfeng liu & jan-jan wu, "energy-efficient virtual machine provision algorithm for cloud system", ieee 4TH international conference ON cloud computing, PP. 81-88, 2011.