ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

A Novel Approach of Load Balancing Strategy in Cloud Computing

Antony Thomas[1], Krishnalal G[2]

PG Scholar, Dept of Computer Science, Amal Jyothi College of Engineering, Kanjirappally, Kerala, India[1]

Assistant Professor, Dept of Computer Science, Amal Jyothi College of Engineering, Kanjirappally, Kerala, India[2]

ABSTRACT: The term 'cloud computing' has grown in popularity, especially with the its widespread use. The very concept of cloud computing is complex and consists of a number of related concepts. However, as the cloud computing environment is growing in number and kinds of services being provided, the need of efficient load balancing and scheduling is also significantly increasing. The task scheduling and resource management are closely related to the efficiency of the whole cloud computing facilities. One such very important and necessary concept is that of load balancing. This importance arises from the fact that the performance and efficiency of cloud computing is based on how effectively load balancing is done. There are many load balancing algorithms available, but the relevant point regarding this is how these algorithms are effectively used. An effective system with regard to cloud computing is one in which the tasks provided by the user are completed within the stipulated time period. This paper proposes to develop and implement an approach to effectively introduce controller and balancer into data centres of a cloud computing system. For this, using different mechanisms, by which each task is optimally assigned to a virtual machine. To do this effectively, task priority algorithm can also be used.

KEYWORDS: Cloud computing, load balancing, Round robin

I. INTRODUCTION

The cloud computing technology is an attractive field in the area of computer science. All kinds of requirements posed by the user can be accomplished through cloud computing. There are numerous ways to define the term 'cloud computing'. NIST defines cloud computing as " as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources(example : network, servers, storage, applications and services) that can be rapidly that can be rapidly provisioned and released with minimal management effort or service provider interaction. From the definition, it is obvious that, for a cloud computing system to be worthy of the definition, we should first focus on the primary concept that is necessary for the efficiency of the system. This concept is called load balancing.

At present, there is a lot of research underway in the field of load balancing. However, load balancing is not an easy task. There are a number of policies available for load balancing. Some of the policies are static, while others are dynamic. In spite of the many load balancing schemes available, if these load balancing schemes are not effectively used, it becomes impossible to accomplish the definition of cloud computing. This paper discusses how this can be done effectively and efficiently. The tasks provided by the user go through different levels. Only if load balancing is achieved at each and every level, will there be any benefit to the user. Data centres in association with a cloud computing system could be located anywhere in the world. The first priority involves selecting the correct data centre. If this is done effectively, we can say that almost 10% of the work is done and we are one step closer to reaching the definition of a cloud computing system. Similarly, load balancing schemes have to be correctly used in the remaining levels.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

The decision regarding which algorithm to use at each level has to be made correctly. As mentioned before, there are lots of load balancing schemes, but their effectiveness varies across levels. Using a particular scheme in one level may be optimal, but using the same algorithm in another level may not be as effective as using an alternative technique. On inspection of each task, it becomes obvious that each node may not be equally capable of handling each task. Each node has its own capacity. Understanding the capacity of each node is necessary in assigning tasks to each node. However, if these assignments are not done in a controlled fashion, it may lead to an imbalance situation. Brokers that act between the user and the data centre should take this into consideration. Many companies provide cloud computing services. Satisfying the requirements of the user is a prime factor for every cloud provider. So, the basic objective of each cloud provider is to correctly complete the users task within the stipulated time. Thus it becomes a primary task for developers to make headway in this field.

The rest of this paper is organised is as follows. Literature review is discussed in section 2. Section 3 present the proposed approach. Section 4 tells analysis of the system. Section 5 concludes this paper.

II. RELATED WORK

Load balancing can affect the overall performance of a system executing an application. Load balancing algorithms can be classified in two different ways [1] : Static and Dynamic.

The main focus is on the efficient utilization of the virtual machines and balancing the virtual machines with the incoming request. Load balancing is defined as a process of making effective resource utilization by reassigning the total load to the individual nodes of the collective system and thereby minimizing under or over utilization of the available resources or virtual machines [2]. Shridhar G. Domanal and G. Ram Mohana Reddy [3] have developed Modified Throttled algorithm which maintains an index table of virtual machines and also the state of VMs similar to the Throttled algorithm [4]. There has been an attempt made to improve the response time and achieve efficient usage of available virtual machines.

Priority determines the importance of the element with which it is associated. In terms of task scheduling, it determines the order of task scheduling based on the parameters undertaken for its computation [5]. In the present framework, the deadline based tasks are prioritized on the basis of task deadline. The tasks with shorter deadline need to be executed first. So they are given more priority in scheduling sequence. The task list is rearranged with tasks arranged in ascending order of deadline in order to execute the task with minimum time constraint first. The cost based tasks are prioritized on the basis of task profit in descending order. This is appreciable as tasks with higher profit can be executed on minimum cost based machine to give maximum profit [6].

A large public cloud will include many nodes and the nodes in different geographical locations. Cloud partitioning is used to manage this large cloud. A cloud partition is a subarea of the public cloud with divisions based on the geographic locations [7].In cloud analyst we can see three different algorithms. Round Robin is a random sampling based algorithm. It means it selects the load randomly in case that some server is heavily loaded or some are lightly loaded. 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. 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 [8]. Table 1 shows data transfer cost for all algorithm that are mentioned above [9].

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

Parameters	*Load Balancing Algorithm on cloud Analyst*		
	Round Robin	*ESCE*	*Throttled*
Data Centres	*2*	*2*	*2*
UB	*5*	*5*	*5*
H/W Unit	*2*	*2*	*2*
VM	*20*	*20*	*20*
Avg (ms)	*0.28*	*0.28*	*0.28*
Min (ms)	*0.02*	*0.02*	*0.02*
Max (ms)	*0.64*	*0.64*	*0.64*
Total $ (Cost)	*1.83*	*1.83*	*1.82*

Table 1 Cost Estimation and Average Requesting Time Table

III. PROPOSED APPROACH

In the proposed approach, different load balancing algorithms are used. Each and every task will pass through different stages. And ultimately, each task will reach its intended destination, referred to as the host. The overall system architecture is shown in fig 3.1. The three stages are given below:

1. Task to Data centre Controller
2. Data centre controller to each partition
3. Finally the task is assigned to each host

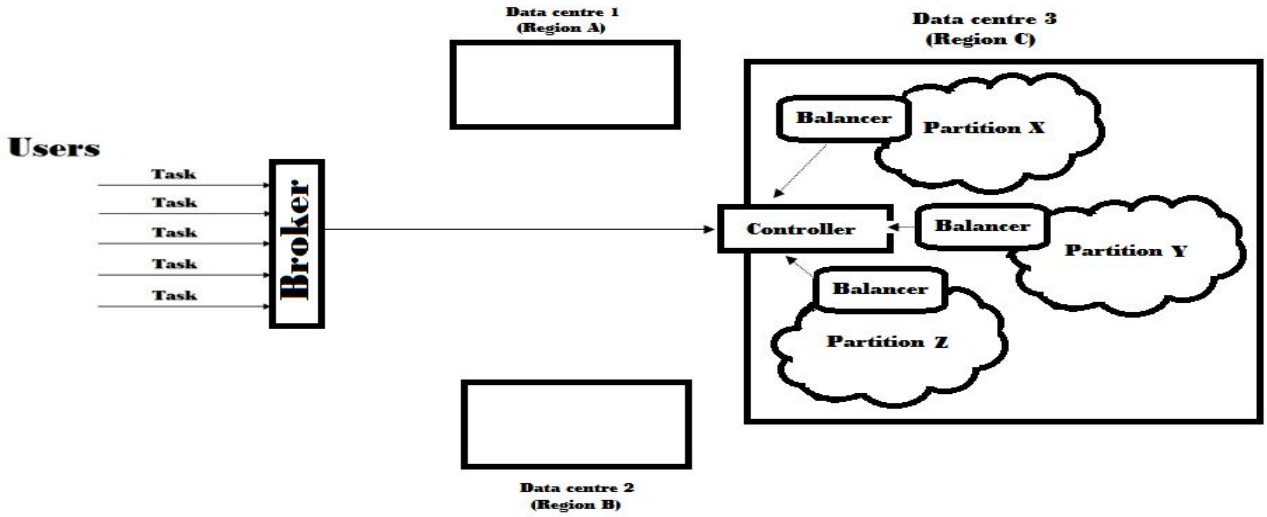

Fig 3.1 Overall architecture

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

First Stage: ***Task to Data centre Controller***

In the first stage, each task from the user has to reach a suitable data centre. The decision regarding this allocation is done by considering the locality of the data centre. The nearest data centre is chosen. This is done with the intention of reducing the execution time of the task. Fig 3.2 shows the model of this stage. The responsibility of this task is carried out by the cloud broker. Thus it is necessary for the broker to have a mechanism to locate the nearest data centre.

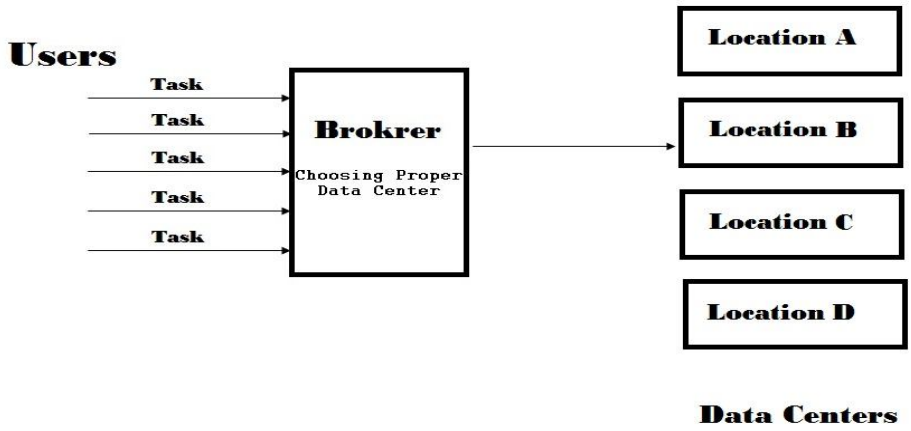

Fig 3.2 Task to Data centre controller

Second Stage: ***Data centre controller to each partition***

In the next level, note that the task has been allocated to a data centre. Each data centre has its own controller. Each data centre has a number of partitions and each partition has its own nodes (hosts). In the second stage, the controller chooses a suitable partition by considering the load status. For this, each partition has a balancer. This balancer maintains a status of loads on different hosts under that partition. This status has to be periodically updated and sent to the main controller. The partition can be of 3 types:

1. **Idle**: The partition is idle when the percentage of idle nodes exceeds α
2. **Normal**: The partition is idle when the percentage of normal nodes exceeds β
3. **Overloaded**: The partition is overloaded when the percentage of overloaded nodes exceeds γ

The parameters are set by the cloud partition balancers [7]. Status of the partitions may change from one state to another. For example, a partition in normal state will move to an idle state on completion of all the task assigned to it. Before assigning the tasks to different partitions, each task has to be prioritised based on its deadline. If the partition status is idle or normal, the tasks can be assigned to that partition. If there is no other partition with the status 'idle' or 'normal' , the task is assigned to the partition which is having a a lower load when compared to others. The diagram of third stage is shown in fig 3.3

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

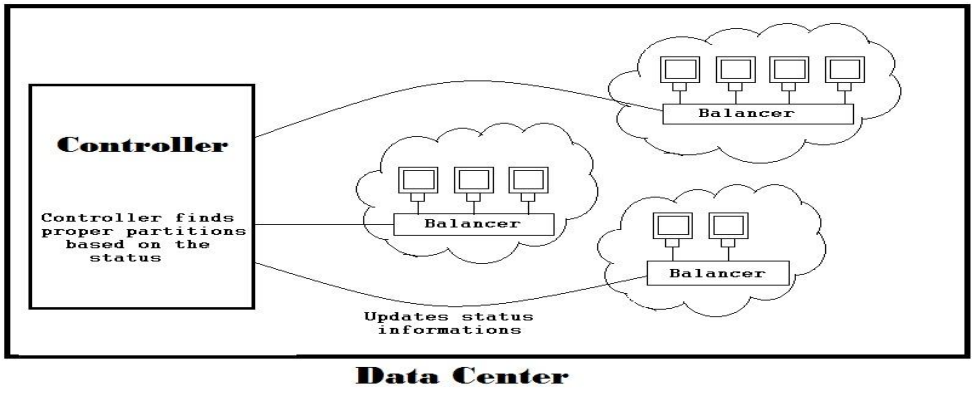

Fig 3.3 Second stage

Third Stage: **Finally task assigned to each host**

In the third stage, the tasks are assigned to corresponding hosts. This can be done in 2 different ways. The load degree has to be computed for each node in the partition. The load degree is computed from some static and dynamic parameters. The static parameters are number of CPUs, memory size etc. Dynamic parameters are memory utilisation ratio, CPU utilisation ratio, network bandwidth.

If the cloud partition is idle then the round robin algorithm is used for assigning task to the hosts. If the cloud partition is normal, another load balancing strategy has to be used. This situation is a bit more complex than the first one. Shridhar G.Damanal and G. Ram Mahana Reddy [2] proposed a static load balancing strategy called VM-Assign Load Balance Algorithm. This can be applied here. The diagram of third stage is shown in fig 3.4

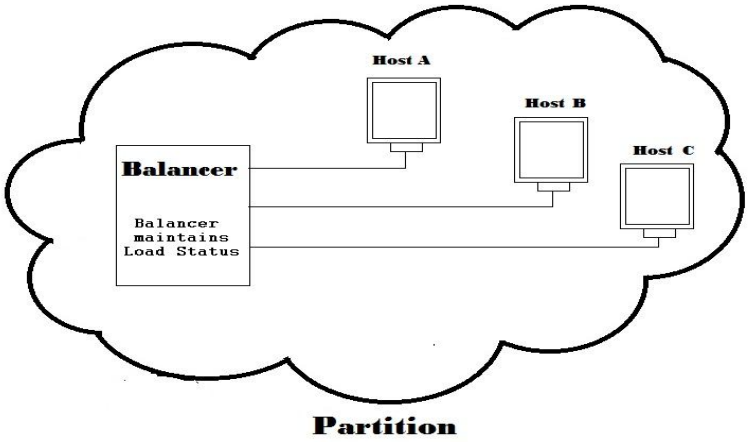

Fig 3.4 Third stage

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

IV. ANALYSIS

The performance of a cloud computing system depends on the task execution time. This paper proposes a cloud system model in which a mechanism is used to select a data centre such that the execution of task is quite efficient. This is because the data centre nearest to the user is selected. The task prioritization mechanism used in this paper improves the task scheduling policy. Due to the existence of partitions in the system, it is possible to use optimal algorithms based on the state of the partition. Thus, the task execution is done effectively and deadlines are met.

V. CONCLUSION

We have proposed a load balancing approach for cloud computing environment. It has been noted that traditional load balancing algorithms are usually not flexible and cannot match the dynamic changes to the attributes during the execution time. Dynamic algorithms are more flexible and take into consideration different types of attributes in the system both prior to and during run-time. Selecting nodes for executing a task in cloud computing is a major concern. They have to be properly selected according to the properties of the task.

REFERENCES

[1] A. Revar, M. Andhariya, D. Sutariya, M. Bhavsar, Load Balancing In Grid Environment Using Machine Learning-Innovative Approach, International Journal Of Computer Applications 8 (10 (Oct)) (2010) 975–8887.

[2] Shridhar G.Damanal And G. Ram Mahana Reddy ,Optimal Load Balancing In Cloud Computing By Efficient Utilization Of Virtual Machines - Ieee 2014.

[3] Shridhar G. Domanal And G. Ram Mohana Reddy," Load Balancing In Cloud Computing Using Modified Throttled Algorithm" Ieee, International Conference. Ccem 2013. In Press.

[4] Brototi Mondal, Kousik Dasgupta And Paramartha Dutta, "Load Balancing In Cloud Computing Using Stochastic Hill Climbing-A Soft Computing Approach" In Procedia Te Chnology 4 (2012) 783 - 789,Elsevier C3it-2012.

[5] Q. Cao, B. Wei And W. M. Gong, "An Optimized Algorithm For Task Scheduling Based On Activity Based Costing In Cloud Computing," In International Conference On Esciences 2009, Pp. 1-3.

[6] Monika Choudhary, Sateesh Kumar ,A Dynamic Optimization Algorithm For Task Scheduling In Cloud Environment , Peddoju International Journal Of Engineering Research And Applications (Ijera), May-Jun 2012, Pp.2564-2568

[7] Gaochao Xu, Junjie Pang, And Xiaodong Fu, A Load Balancing Model Based On Cloud Partitioning For The Public Cloud , Tsinghua Science And Technology Issn 1007-0214 04/12 pp34-39 Volume 18, Number 1, February 2013

[8] Ms.Nitika , Comparative Analysis Of Load Balancing Algorithms In Cloud Computing, International Journal Of Engineering And Science

[9] Dr. Hemant S. Mahalle , Prof. Parag R. Kaveri And Dr.Vinay Chavan , Load Balancing On Cloud Data Centres , International Journal Of Advanced Research In Computer Science And Software Engineering , Volume 3, Issue 1, January 2013