A Performance Improvement of Load Balancing and its Status Evaluation in Public Cloud

Miss. Karuna Bakde   Prof.B.M.Patil

ABSTRACT- Load Balancing is one of the main challenges in cloud computing which is required the distributed the dynamic workload across multiple nodes to ensure that no single node is overloaded. Load Balancing in the cloud computing environment has an important impact on the performance. Good load balancing makes cloud computing more efficient and responsive. In this paper introduces a better load balance model for public cloud based on cloud partitioning concept with a switch mechanism for different situations to resolve the issue of load balancing in public cloud computing.

Index Terms- Cloud Computing, Cloud Partition. Load Balancing, Public Cloud.

————————————— ◆ —————————————

1. INTRODUCTION:-

Cloud is a technology discontinuity that, within next 10 years, is likely to dramatically change IT organizational missions, structures, roles, skills and operations [2]. Cloud computing is an internet computing in which the load balancing is the one of the challenging task [10]. Cloud computing is on demand service in which information, shared resources, software and other devices are provided according to clients request and its requirement at specific time. Users get service from a cloud without paying attention to the details [1].

Load balancing is the new technique that facilitates networks and resources by a maximum throughput with minimum response time [7]. Proper load balancing can help in utilizing the available resources optimally, thereby minimizing the resource consumption. Load balancing schemes depending on whether the system dynamics are important can be either static or dynamic [12]. Static load balancing scheme divide the traffic equivalently between the

services. Dynamic load balancing scheme chooses the lightest server preferred to balance the traffic and selecting an appropriate server needed real time communication with network. The advantage of using dynamic load balancing scheme is that if any node fails, it will not halt the system, it will only affect the system performance [3]. Dynamic load balancer uses policies for keeping track of updated information.

A Public cloud is one which the services and infrastructures are provided over the internet [4]. Public cloud environment is very large and it hard to maintain. Thus, this model divides the public cloud in several cloud divisions. When the environment is very large, divisions simplify the load balancing. The cloud has a main controller that chooses the suitable partitions for arriving jobs while the balancer for each cloud partition chooses the best load balancing strategy. There are many load balancing algorithm such as Round Robin load balancing algorithm, Throttled load balancing algorithm, Equally Spread Current Execution Algorithm, Ant Colony algorithm.

Randles et.al.[9] giving the comparative analysis by checking cost and performance.

2. RELATED WORK:-

There have been many studies of load balancing for the cloud environment. Cloud computing is a recent trend in IT that moves computing and data away from desktop and portable PCs into large data centers. It refers to applications delivered as services over the Internet as well as to the actual cloud infrastructure — namely, the hardware and systems software in data centers that provide these services [11]. However, load balancing in the cloud is still a new problem that needs new architectures to adapt to many changes. There are many load balancing techniques given by the researchers over time to time same have advantages over and vice-versa. Distribute workload of multiple network links to achieve maximum throughput, minimize response time and to avoid overloading. There are many load balancing algorithm such as Round Robin load balancing algorithm, Throttled load balancing algorithm, Equally Spread Current Execution Algorithm, Ant Colony algorithm. Randles et.al.[9] giving the comparative analysis by checking cost and performance. Modified Throttled Algorithm using here because it gives better performance.

3. SYSTEM MODEL:-

A Public Cloud is based on the standard cloud computing model, with service provided by a service provider. A public cloud area is very large environment. So, it hard to manage it and complexity arises. Partitioning the cloud in small divisions to simplify load balancing and improving the system performance.
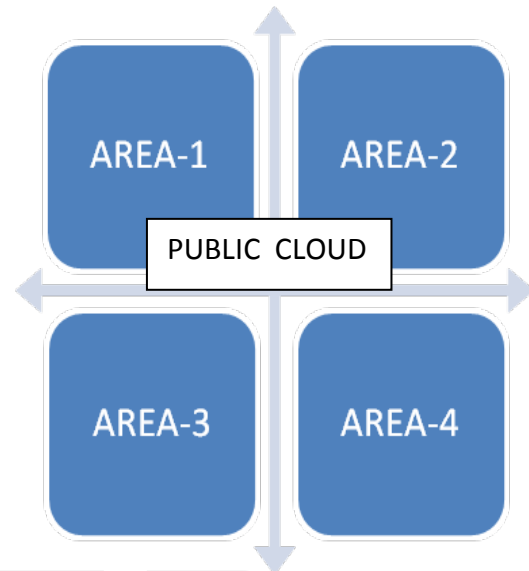

Fig. 1   Cloud partitioning.

3.1 Main controller and balancers

Load balancing done by main controller and the balancers. When a job arrives at public cloud, first step is choosing the right partition to perform the required operation. Main controller first assigning job to suitable partition and communicate with balancers to refresh the status information of all the partition giving idea of status (busy/ideal). Balancer gathers the status information of all partition of every node.

3.2 Assigning jobs to the cloud partition

When a job arrives at the public cloud, the first step is deciding to choose right partition. The cloud partition status can be divided into three types:
(1)Idle: When the percentage of idle nodes exceeds, change to idle status.

(2) Normal: When the percentage of the normal nodes exceeds, change to normal load status.

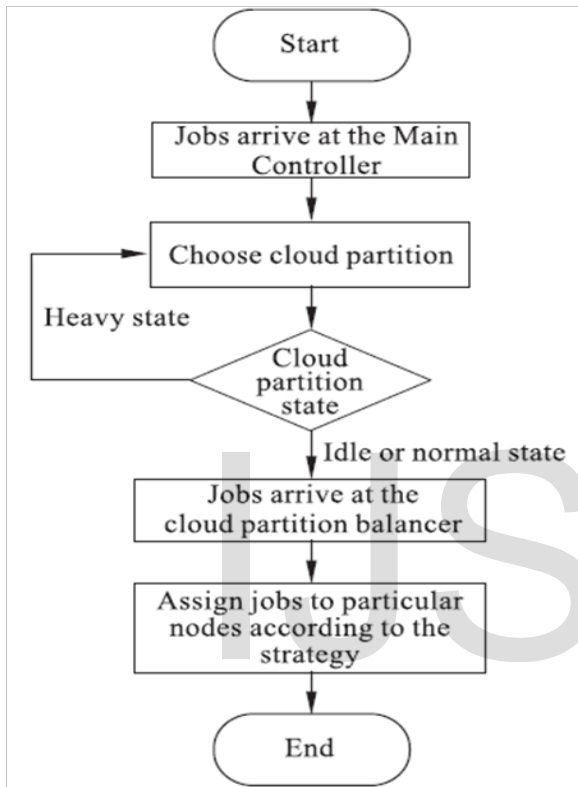(3) Overload: When the percentage of the overloaded nodes exceeds, change to overloaded status [5].

Fig.2 Job Assignment Strategy

3.3 Assigning job to the suitable cloud partition

Cloud partition state gather all status information from the node and compute load degree of the partition. Load degree is depend on various static and dynamic parameters. Static parameter is memory speed of CPU, number of CPU include in load balancing and memory size. Dynamic parameters is CPU utilization ratio, Network bandwidth, memory utilization ratio [1].

BEST PARTITION SEARCHING ALGORITHM

```
begin
 while job do
   searchBestPartition (job);
     if partitionState == idle || partitionState    ==
normal  then
       Send Job to Partition;
   else
     search for another Partition;
   end if
 end while
 end
```

Load degree is computed from these parameters are as follows:

Compute Load Degree

Inputs:

The static parameters include the number of CPU's, the CPU processing speeds, the memory size, etc. Dynamic parameters are the memory utilization ratio, the CPU utilization ratio, the network bandwidth.

Process:

1. Define a load parameter set:

   F= {F1,F2…Fm} with each Fire presents the total number of the parameters.

2. Compute the load degree as

   Load _Degree(N)= Σ $\alpha_i F_i$

   where i= 1…m

3. Average cloud partition degree from the node load degree statistics as:

Load degreeavg=Σi=1..n LoadDegree(Ni)

4. Three level node status are defined

IDLE:

Load_degree(N)=0

NORMAL:

0<Load_Degree(N)<Load_Degree(N)high

OVERLOAD:

Load_Degree(N)high <= Load_Degree(N)

Output :-

Idle or Normal Or Overloaded

4. CLOUD PARTITION AND LOAD BALANCING STRATEGY:-

Good load balancing algorithm improving the performance of cloud. Cloud partition state gather all status information from the node and compute load degree of the every partition. When a new job arrives, according the status of load degree which partition is suitable is decided. Balancer for each cloud partition chooses the best load balancing strategy. In this paper Throttled load balancing algorithm are used.

Throttled load balancing algorithm gives the better performance then Round Robin algorithm. This algorithm implements a throttled load balancer (TLB) to monitor the loads on each VM. Here each VM is assigned to only one task at a time and can be assigned another task only when the current task has completed successfully. Throttled load balancing algorithm is completely based on virtual machine. In

this algorithm 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. The process first starts by maintaining a list of virtual machines each row is individually indexed to speed up the look up process. The job of TLB is to maintain an index table of all VMs as well as their current states (Available or Busy). If match is found on the basis of size and availability of the machine, then the load balancer accepts the request of the client and allocates that virtual machine to the client [4]. The TLB scans the index table from top to bottom until the first available VM is found. If it finds, then TLB returns the VM id to the Data Center Controller. The Data Centre communicates the request to the VM identified by the id. On the other hand, if the TLB doesn't find any VM in the available state it simply returns null. In this case Data Center Controller queues the request until the availability of any VM.

IMPROVED THROTTLED LOAD BALANCING ALGORITHM FOR PUBLIC CLOUD

Step1: Main Controller maintains the status information of all partitions

Step2: Vm Load Balancer maintains an index table of VMs and the state of the VM (BUSY/AVAILABLE). At the start all partition and its VM's are available.

Where, Vm=Virtual Machine.

Step3: While (Vm_clicks=0) do

Step4: Evaluate fitness of Load efficiency.

$$fitness = \frac{\sum_{i=1}^{n} response\ time}{Vm\_clicks}$$

Where, n = total no. of users

i = no. of Vm

Step5: Searching for new Vm

Step6: Select the fittest Vm and Calculate Load.

Step7: Calculate Load ,check if the load on selected Vm is idle, normal or overloaded.

Step8: D0 Vm allocation or task shifting in continuous loop.

5. CONCLUSION:-

This paper demonstrated the applicability of using partition techniques and Improved Throttled Load Balancing Algorithm to obtain measurable improvements in resource utilization and availability of cloud-computing environment and increase the business performance in cloud based sector. According the partition load, status of partition is calculated.
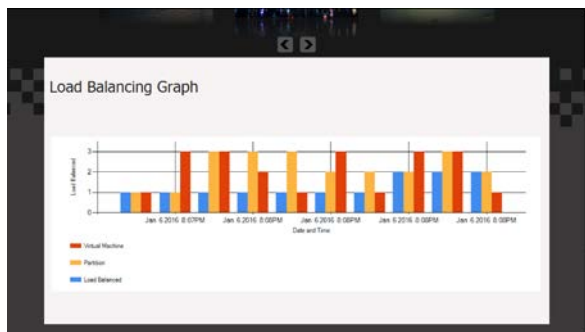

Fig3:Load balance graph

Above graph shows the virtual machine, partition and load balance. This graph shows the load balance status of the particular date. This also shows the partition of Vm and its status.

This graph giving idea about server and virtual machine load. In this partition server is fully loaded and virtual machines are in normal or idle state. From this status evaluation of partition is calculated.
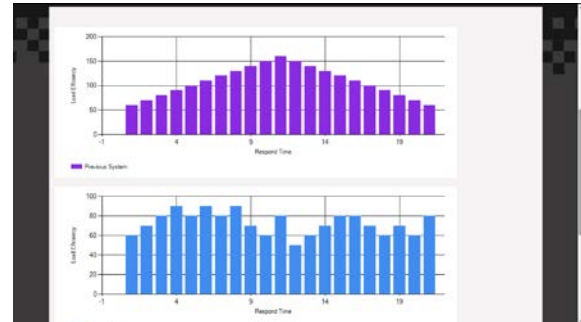

Fig4: Difference between throttled and Improved throttled Load balancing algorithm.

Above graph shows the difference between throttled and improved throttled load balancing algorithm with respect to load efficiency and response time.

6. FUTURE WORK:-

Cloud partitioning is a method to make partitions of public cloud. Public cloud is made up of several nodes situated in deferent geographic location. Cloud partitioning is a method to make partitions of huge public cloud in small partitions. Other partitioning concept can used in public cloud because public cloud is very area and hard to maintain it. In this paper Improved Throttled load balancing algorithm are used for balancing the load in public cloud. In this algorithm there improvement or development can be occur. Other load balancing strategy can give better

result and improving the performance. So, tests are needed to compare different load balancing strategies. **7. REFERENCES:-**

[1] Gaochao Xu, Junjie Pang, Xiaodong Fu "A Load Balancing Model Based on Cloud Partitioning for the Public Cloud" IEEE Transactions on Cloud Computing Year 2013.

[2] M. D. Dikaiakos, D. Katsaros, P. Mehra, G. Pallis, and A. Vakali, "Cloud Computing: Distributed Internet Computing for IT and Scientific Research", Internet Computing, Vol.13, no.5, Sept.-Oct. 2009, pp.10-15.

[3] S.S.Moharana, R.Ramesh, D.Powar, "Analysis of Load Balancers in Cloud Computing" International Journal of Computer Science and Engineering Vol.2, May 2013, pp.101-108.

[4] A. Rouse, Public cloud, http://searchcloudcomputing.techtarget.com/defination/public-cloud, 2012.

[5] Shrikant M. Lanjewar, Susmit S. Surwade, Sachin P. Patil ,Pratik S. Ghumatkar, Prof Y.B. Gurav "Load balancing in public cloud" IOSR Journal of Computer Engineering (IOSR-JCE) Volume 16, Issue 1, Ver. VI (Feb. 2014), PP 82-87.

[6] Ms.Nitika, "Comparative Analysis of Load Balancing Algorithms In Cloud Computing" International Journal of Engineering and Science , Vol 1, Issue-1.

[7] R. Shimonski. Windows 2000 & Windows Server 2003 Clustering and Load Balancing. Emeryville. McGraw-Hill Professional Publishing, CA, USA (2003), p 2, 2003.

[8] N. G. Shivaratri, P. Krueger, and M. Singhal "Load distributing for locally distributed systems" , Computer, Vol.25, no.12, pages.33-44, December.1992.

[9] M. Randles, D. Lamb, and A. Taleb-Bendiab, "A comparative study into distributed load balancing algorithms for load balancing algorithms for cloud computing", in Proc. IEEE 24th International Conference on Advanced Information Networking and Applications, Perth, Australia, 2010, pp.551-556.

[10] Rajesh George Rajan, V.Jeyakrishnan, "A Survey on Load Balancing in Cloud Computing Environments", in International Journal of Advanced Research in Computer and Communication Engineering" Vol.2, Issue 12, December 2013.

[11] Ankit Bahuguna, "CLOUD COMPUTING Vision, Hype and Reality about Distributed Internet Computing for IT and Scientific Research" Emerging Trends in Computer Applications.

[12] N. G. Shivaratri, P. Krueger, and M. Singhal "Load distributing for locally distribute systems" , Computer, Vol.25, no.12, pages.33-44, December.1992.