International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 10, October 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

A COMPARITIVE STUDY ON VARIOUS STATIC LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING

Abhijeet Malik[1], Prabhat Chandra[2]

Master of Computer Application, National Institute of Technology, Kurukshetra, (India)

ABSTRACT

Internet is growing day by day and so is their application due to which traffic on the network is increasing tremendously. To alleviate the effect of this increased traffic some effective load balancing algorithms is needed. It is a mechanism used for distributing workload on nodes present in cloud infrastructure. Load balancing algorithms are divided into two classes: Static and dynamic load balancing which are based on system topology. This paper first describes cloud computing and then gives brief overview of various static load balancing algorithms and then a comparative study of these algorithms based on various parameters is discussed. These algorithms try to reduce response time, reduces power consumption, increase system throughput and better resource utilization. After that an algorithm is proposed for improving round robin's response time.

Keywords— Cloud Computing; Load Balancing; Resource Utilization; Response time; Round Robin; Throughput;

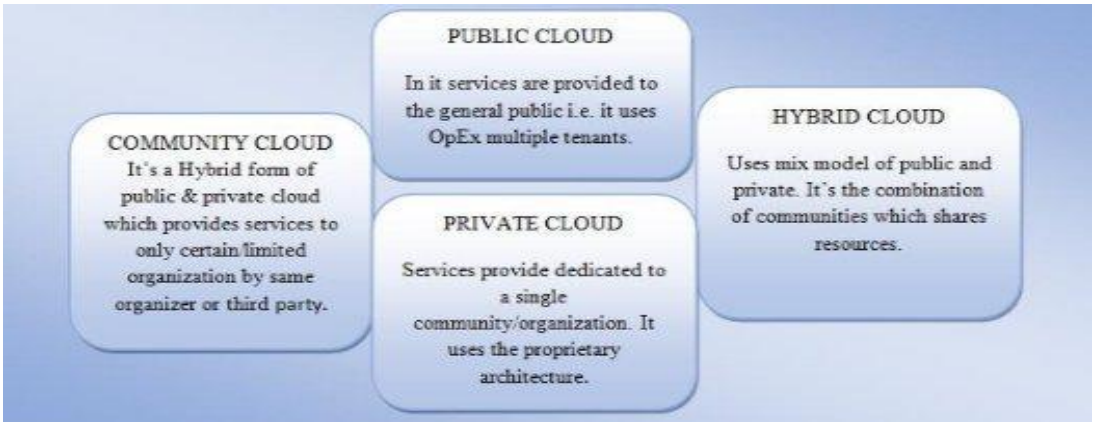
I. INTRODUCTION

The "cloud computing" is everywhere. In simple words cloud computing means store and access data and different programs on Internet in place of your computer's hard drive. Cloud computing move data from portable personal computers to large data centers. It can be adjusted as the demand changes which will eliminate the cost of software and hardware [1]. There is a need of an efficient working algorithm which helps in reducing the infrastructure costs, growing applications, overall environmental benefits (low carbon emissions) as many users shares large systems efficiently. Cloud is consisted of various types of records

II. TYPES OF CLOUD

There are various types of cloud categorized on the basis of their deployment and services provided to various types of users as shown in Fig.1

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 10, October 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354



III. CLOUD SERVICE MODEL

A. INFRASTRUCTURE as a SERVICES (IaaS)

It supplies virtualized computing infrastructure such as hardware, software, servers, storage and also hosts some user applications on-demand basis to their users [2], [3]. This model is suitable for temporary and experimental system or services. Cloud provides an environment that hosts and does not limit an application to particular resources.

B. PLATFORM as a SERVICES (PaaS)

It is related to that form of cloud computing that supplies a stage and environment which a user can exploit to build their applications and services online and users can access them using any web browser[2],[3]. It provides a platform in cloud such as execution runtime, database, web server, development tool. Examples related to PaaS are – Google App Engine, Windows Azure etc.

C. SOFTWARE as a SERVICES (SaaS)

End user can directly consume the services provided in the cloud [3]. Thus it removes the need for install and run application in their own data center. The client doesn't need to pay for the software or any other hardware supp ort rather the client can directly be benefited from SaaS service.

IV. TECHNOLOGIES RELATED TO CLOUD COMPUTING

A. Grid Computing

This technology [4] is also considered as the future internet access technology. This technology is termed as - "Resource sharing and coordinated problem solving in dynamic, multi-institutional virtual organization" [2]. In it the idle system in the network is efficiently utilized by uniting pool of servers, storage system and network into a single large virtual system for resource sharing at run time.

B. Utility Computing

This computing enables service supplier to create computing resources and infrastructure administration accessible to customers as a metered services when require [2] and it is also helpful for cost-effective schemes. This technology is based on-demand to the user and involves – scalability, high availability, manageability, disaster recovery and value.

C. Virtualization

Virtualization [2] refers to creating a virtual version of anything such as server ,storage devices ,various devices, network elements and operating systems so that many users can work on them simultaneously on-demand by dividing the resources into one or more instances logically. One of the best popular examples of virtualization is VMware. Virtualization increases the efficiency, utilization and flexibility of the existing computer hardware system [5].

D. Autonomic computing

In this technology these are the four main aspects involved in it – Self heal, self configure, self protect and self optimization [6]. SELF HEAL- This is a technique which discovers and then diagnoses and react to disruptions caused of the service.

- SELF CONFIGURING- The installation, configuration and integration is very time and energy consuming. This self configuring technique removes this problem by automated configuration of components with high level policies.
- SELF PROTECTING- This technique identify, anticipate, detect and protect against unauthorized access or intrusion or any viruses [6].
- SELF OPTIMIZING- This technology involves monitoring and tuning of resources on its own. It helps in operating in unpredicted development working environment and efficiently utilization of the resources available.

V. LOAD BALANCING IN CLOUD COMPUTING

Load balancing is the way of distributing weight (jobs or tasks) transversely on set of processors which are linked to set of connections distributed throughout world. In load balancing approach every processor is tried to make them similarly busy to complete the works more or less at the same time. Load Balancing requires various scheduling algorithm to work efficiently and for utilizing the power of cloud computing. The profit of workload distribution includes superior resource consumption percentage which leads to increasing on the whole performance [1], [3]. A diagram of load balancing environment is shown in Fig. 2.

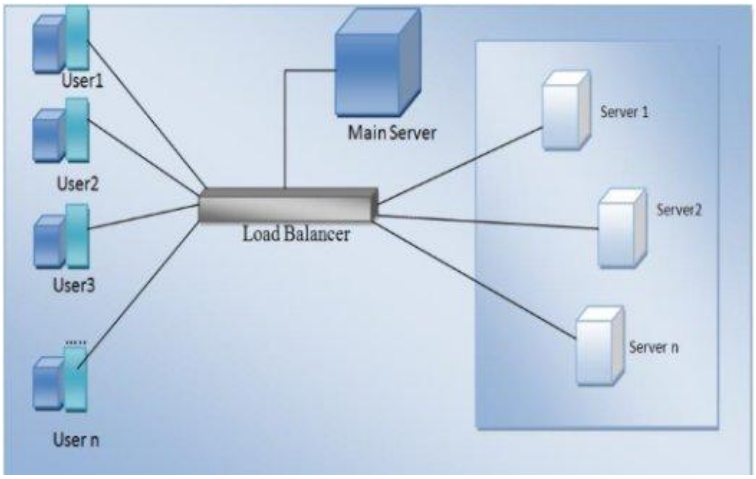


Fig. 2

VI.CHALLENGES IN LOAD BALANCING

Before study of a variety of load balancing algorithms we need brief introduction of issues and challenge that we can face in several load balancing techniques [7]. Some of these issue and challenges are summarized below:-

A. Spatial Distribution

Several load balancing techniques are designed for limited and closely located cloud nodes like in intranet but in real life applications this is not the case, therefore some other factor must be considered like distance among various service nodes, pace of set of connections associations amongst nodes and remoteness between client node and task processing nodes etc [7].

B. Full Replication Algorithm

Full replication algorithm are not always required because they do not provide efficient storage as same data is replicated to all of the nodes in the network so we need partially replicated algorithm which are complex but they save parts of memory in various cloud notes and provide efficient resource utilization.

C. Algorithm Complexity

Load balancing algorithm must be simple in terms of implementations and number of operations required to perform any task because more complexity may lead to various unwanted negative effects.

D. Point Of Failure

In load balancing the data of all nodes is collected and then this load is balanced. This data must be secured by using various techniques. Further more if the controller or any node collapses then also our whole system work properly and no data is lost. For this purpose we need distributed load balancing algorithm with proper control and coordination among various units to work properly.

VII.OBJECTIVE OF LOAD BALANCING

∑*FLEXIBLITY AND SCALABILITY*- Algorithm used must possess flexibility and scalability.

∑ *PRIORITIZATION*- Priority based queue of the various resources or jobs

∑ *COST EFFECTIVE* – Algorithm which works with lower cost and energy efficient.

∑ *IMPROVED REACTION TIME*- Time that is taken by particular algorithm to respond to a task. Algorithms which take smallest amount of time for responding to tasks or jobs that is given to processors.

∑ *OPERATING COST ASSOCIATE*- It depends on the inter process communication between the tasks.

VIII. LOAD BALANCING ARE CLASSIFIED IN TWO WAYS

A. Static Algorithm

In Static load balancing algorithms, task is assigned to node on the basis of its ability to process the user request. Tasks are settled equally to servers. Due to increase of traffic performance decreases.

B. Dynamic Algorithm

Dynamic algorithms take current position and state to balance the load which changes the status by altering the parameters. Tasks are assigned dynamically to the server.

IX. FOLLOWING ARE SOME ALGORITHMS WHICH ARE RELATED FOR LOAD BALANCING

A. Shortest Job First

This algorithm reduces the waiting time of processes by assigning the CPU to the process having least processing time i.e. the shortest job is the one which is processed before the heavy load jobs. This algorithm provides efficient utilization of resources and minimizes time consumption.

B. Max-Min Algorithm

This algorithm is related to the work or task having maximum finishing time is chosen and is allocated to the cpu with least finishing time[3],[8],[9]. A disadvantage of Max-Min algorithm is that cloudlets with minimum completion time have maximum waiting time.

C. Opportuninistic Load Balancing (Olb)

OLB load balancing algorithm [8], [9] is static in nature. It keeps every virtual-machine in the system busy and it never considers the currently running tasks assigned to each and every virtual machine which helps in achieving proper load balancing. Disadvantage regarding OLB algorithm is that it certainly not uses the expected Processing time of the task which leads to less optimized result [2].

D. Min-Min Algorithm

In this approach the job with minimum size is assigned to the resource which can produce the output in shortest span of time [3], [9]. Thus more jobs could be assigned to faster processors to increase system throughput. While when there are scheduling requests from the diverse applications, the scheduler allocates the application to the host with selection of most excellent match from the group of applications and group of different offered hosts. The selected approach based on forecast of computing supremacy of related hosts.

E. Task Scheduling On Load Balancing

This task Scheduling algorithm is consisted of two phase mechanism on task scheduling which is based on load balancing for the fulfillment of users needs. This helps in obtaining elevated resources consumption or utilization. This algorithm in starting mapped all the tasks to virtual machines and then all virtual machines to host resources for achieving load balancing. It is improving the task reaction time. It also gives superior and enhanced resource utilization [3]

F. Simple Round Robin Algorithm

Processes are sent through first in first out method with the help of queue but it has specified limited amount of time slice or time quantum (CPU time). If a running job or process does not get finished before its time slice expires, the CPU is preempted from running process and then it is allocated to next process in queue which is waiting for the CPU. Then preempted process is send to back of the queue consisting of ready processes to be accomplished. Queue is circular (assumption). It is combination of the first come first serve, preemption of cpu

from running processes according to time slice [9]. Minimum Response time is main advantage of this algorithm for the time sharing systems.

Comparison of above stated algorithms on the basis of various specification such as Throughput, Performance, Fault Tolerance of algorithm, Response Time(reaction time), Resource Utilization, Scalability, and Power Saving.

X. ALGORITHM SCHEDULING CRITERIA

CPU UTILIZATION- making the system as demanding as possible so as to make a proper use of processors.

∑ *THROUGH-PUT*- It is the number of processes accomplished in per unit time of cpu cycle.

∑ *TURNAROUND TIME*- time of starting of a process and time when it finished its job or task.

TAT = C.T - A.T, C.T=COMPLETION TIME, A.T=ARRIVAL TIME

∑ *WAITING TIME*- it is the time spends in the ready queue waiting for the cpu. WT = TAT - B.T B.T=BURST TIME.

∑ *RESPONSE TIME* –Time from the submission of a request until the first response is produced.

∑ *FAULT TOLERANCE:* -fault tolerance means when the failure occurs, how it recovers. Load balancing techniques must be a good fault tolerant.

XI. PROPOSED ALGORITHM

Load balancing scheduling is the best way to decide how resources can be better utilized so that there is no failure of servers. In this algorithm first we put all the processes in an array and arrange them according to their priority as shown in Fig.3

For

All tasks T_i

If T_i is having high priority Put the task into Tl batch Else

Put the task into T2 batch End for

Do until all tasks in Tl mapped For

Sorting all tasks in Tl in decreasing order of Priority and adding them to request queue then

Dequeue a task from request queue and then allocate resources with Round Robin Algorithm

End for End do

Clear all task requests

Do until all tasks in T2 mapped For

Sorting all tasks in T2 in decreasing order of priority and adding them to request queue then

Dequeue a task from request queue and then allocate resources with Round Robin Algorithm

End for End do

Technique	Performance	Throughput	Fault Tolerance	Response Time	Resource Utilization	Scalability	Power Saving
Max-Min	Yes	Yes	No	No	Yes	Yes	No
OLB	Yes	No	No	No	Yes	No	No
Min-Min	Yes	Yes	No	Yes	Yes	No	No
TLB	Yes	No	No	Yes	Yes	No	No
Round Robin	Yes	Yes	No	Yes	Yes	Yes	No
Dynamic RR	No	Yes	Yes	No	Yes	No	No
SJFS	No	No	No	No	Yes	No	No

Table 1

Fig 3

XII. CONCLUSION

Load balancing is one of the major challenges in cloud computing. This will avoid the situation where some nodes are heavily loaded while others are idle or doing little work. It helps to achieve a high user satisfaction and resource utilization ratio. This paper describes various load balancing algorithms and a comparative study is also provided based on some parameters such as throughput, Resource utilization, Response time, Fault tolerance etc. Existing algorithms have some limitations and to remove these limitations to improve the efficacy of various algorithms a new load balancing algorithm can be designed by combining the features of Round robin and min-min algorithm. In this proposed algorithm tasks are first divided according to their priorities. Hence, it may improve the overall performance and resource utilization of the system and helps to take further steps towards green cloud computing. It also ensures that every computing resource is distributed efficiently and fairly.

REFERENCES

[1] N. K. Mishra, and N. Mishra,"Load Balancing Techniques: Need, Objectives and Major Challenges in Cloud Computing- A Systematic Review," in International Journal of Computer Applications, Volume 131 - Number 18, Year 2015.

[2] A. Vig, R. Singh Kushwah, and S. Singh Kushwah "An Efficient Distributed Approach for Load Balancing in Cloud Computing," in International Conference on Computational Intelligence and Communication Networks, pp: 751-755,year 2015.

[3] R. Kaur, and P. Luthra, "Load Balancing in Cloud System using Max Min and Min Min Algorithm," in

International Journal of Computer Applications (0975 – 8887), National Conference on Emerging Trends in Computer Technology (NCETCT-2014) ,pp:31-34 ,year 2014.

[4] U. Prajapati, and S. Shakya ,"Task Optimization in Grid Computing using Genetic Algorithm ," in The 10th International Conference on e-Business (iNCEB2015), Nov 2015.

[5] M. Sharma, P. Sharma, and S. Sharma, "Efficient Load Balancing Algorithm in VM Cloud Environment," in IJCST ,Vol.3, Issue 1, pp. 439-441, Jan. - Mar 2012.

[6] V. Sundramoorthy, P. Hartel, and H. Scholten "Taxonomy of Self-configuring Service Discovery Systems", pp: 1-29.

[7] K. Al Nuaimi, N. Mohamed, and M. Al Nuaimi, and Jameela Al-Jaroodi"A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms", IEEE Second Symposium on Network Cloud Computing and Applications ,pp: 137-142 ,year 2012

[8] K.Q. Yan, W.P. Liao, and Shun-Sheng Wang, "Towards a Load Balancing in a three-level cloud computing network ", Computer Science and Information Technology (ICCSIT),3rd IEEE International Conference on Computer,pp:108-113 ,IEEE 2010.

[9] S. Aslam, and M. Ali Shah "Load Balancing Algorithms in Cloud Computing: A Survey of Modern Techniques", 2015 National Software Engineering Conference (NSEC 2015) IEEE, pp: 30-35, Year 2015.