© July 2015 | IJIRT | Volume 2 Issue 2 | ISSN: 2349-6002

Load Balancing Algorithms in Cloud Computing Environment : An Introspection

M.Vanitha[1], Dr.P.Marikkannu[2]
[1]*Department of Computer Science and Engineering,*
Karpagam College of Engineering, Coimbatore, India.
[2]*Department of Information Technology,*
Anna University Regional Centre, Coimbatore, India.

***Abstract*— Cloud computing provides services through online to users. The computing cloud is intended to allow the user to avail of various services without investing in the underlying architecture. Load balancing is one of the main issues in the field of cloud computing. The load is in the form of memory, CPU capacity, network or delay load. Cloud load balancing is the process of distributing workloads across multiple computing resources. Cloud load balancing reduces costs associated with document management systems and maximizes availability of resources. Instead of using physical machines, cloud computing uses virtual machines for hosting, storing and networking of different components. The load balancing needs to be done properly because failure in any one of the node can lead to unavailability of data.To provide a solution to load balancing several load balancing algorithms are available.**

Index Terms*— *Cloud Computing, Load balancing, virtual machines

I. INTRODUCTION

Cloud computing is an on demand service in which shared resources, information, software and other devices are provided according to the clients requirement at specific time. It's a term which is generally used in case of Internet. The whole Internet can be viewed as a cloud. Capital and operational costs can be cut using cloud computing.

II. CLOUD COMPONENTS

A Cloud system consists of the following three major components, each of which has a specific role.

Clients: **T**o manage information related to the cloud, end users interact with the clients. Clients generally fall into three categories.

Mobile: Windows Mobile Smartphone, smart phones, like a Blackberry, or an iPhone.

Thin: They don't do any computation work. They only display the information. Servers do all the works for them. Thin clients don't have any internal memory.

Thick: These use different browsers like IE or mozilla Firefox or Google Chrome to connect to the Internet cloud.
Now-a-days thin clients are more popular as compared to other clients because of their low price, security, low consumption of power, less noise, easily replaceable and repairable etc.

Datacenter: Datacenter is nothing but a collection of servers hosting different applications. A end user connects to the datacenter to subscribe different applications. A datacenter may exist at a large distance from the clients.
Now-a-days a concept called virtualization is used to install a software that allow multiple instances of virtual server applications.

Distributed Servers: Distributed servers are the parts of a cloud which are present throughout the Internet hosting different applications. But while using the application from the cloud, the user will feel that he is using this application from its own machine.

III. TYPE OF CLOUDS

Based on the domain or environment in which clouds are used, clouds can be divided into 3 catagories :

- Public Clouds
- Private Clouds
- Hybrid Clouds (combination of both the private and public clouds)

IV. VIRTUALIZATION

It is a very useful concept in context of cloud systems. Virtualization means "something which isn't real", but gives all the facilities of a real. It is the software implementation of a

computer which will execute different programs like a real machine.

Virtualization is related to cloud, because using virtualization an end user can use different services of a cloud. The remote datacenter will provide different services in a fully or partial virtualized manner.

Types of virtualization are:

- Full virtualization
- Paravirtualization

Full Virtualization

In case of full virtualization a complete installation of one machine is done on the another machine. It will result in a virtual machine which will have all the softwares that are present in the actual server.

Here the remote datacenter delivers the services in a fully virtualized manner. Full virtualization has been successful for several purposes, they are:

- Sharing a computer system among multiple users
- Isolating users from each other and from the control program
- Emulating hardware on another machine

Paravirtualization

In paravitualisation, the hardware allows multiple operating systems to run on single machine by efficient use of system resources such as memory and processor. e.g. VMware software. Here all the services are not fully available, rather the services are provided partially.

Paravirtualization has the following advantages:

Disaster recovery: In the event of a system failure, guest instances are moved to another hardware until the machine is repaired or replaced.

Migration: As the hardware can be replaced easily, hence migrating or moving the different parts of a new machine is faster and easier.

Capacity management: In a virtualized environment, it is easier and faster to add more hard drive capacity and processing power. As the system parts or hardwares can be moved or replaced or repaired easily, capacity management is simple and easier.

V. SERVICES PROVIDED BY CLOUD COMPUTING

Service means different types of applications provided by different servers across the cloud. It is generally given as ”as a service”. Services in a cloud are of 3 types, they are:

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

Software as a Service (SaaS)

In SaaS, the user uses different software applications from different servers through the Internet. The user uses the software as it is without any change and does not need to make lots of changes or doesn’t require integration to other systems. The provider does all the upgrades and patching while keeping the infrastructure running.

The client will have to pay for the time he uses the software. The software that does a simple task without any need to interact with other systems makes it an ideal candidate for Software as a Service. Customer who isn’t inclined to perform software development but needs high-powered applications can also be benefitted from SaaS.

Some of these applications include,

- Customer resource management (CRM)
- Video conferencing
- IT service management
- Accounting
- Web analytics
- Web content management

Benefits: The biggest benefit of SaaS is costing less money than buying the whole application. The service provider generally offers cheaper and more reliable applications as compared to the organisation. Some other benefits include,

- Familiarity with the Internet
- Better marketing,
- Smaller staff,
- reliability of the Internet,
- data Security,
- More bandwidth etc.

Obstacles: SaaS isn’t of any help when the organization has a very specific computational need that doesn’t match to the SaaS services while making the contract with a new vendor, there may be a problem. Because the old vendor may charge the moving fee. Thus it will increase the unnecessary costs.

SaaS faces challenges from the availability of cheaper hardwares and open source applications.

Platform as a Service (PaaS)

PaaS provides all the resources that are required for building applications and services completely from the Internet, without downloading or installing a software.

PaaS services are software design, development, testing, deployment, and hosting. Other services can be team collaboration, database integration, web service integration, data security, storage and versioning etc.

Hardware as a Service (HaaS)

It is also known as Infrastructure as a Service (IaaS). It offers the hardware as a service to a organization so that it can put anything into the hardware according to its will.

HaaS allows the user to "rent" resources as,

- Server space
- Network equipment
- Memory
- CPU cycles
- Storage space

VI. LOAD BALANCING

One of the important features of cloud computing is scalability. Cloud computing resources can be scaled up on demand to meet the performance requirements of applications. Load balancing distributes workloads across multiple servers to meet the application workloads. The goals of load balancing techniques are to achieve maximum utilization of resources, minimizing the response times, maximizing throughput. Load balancing distributes the incoming user requests across multiple resources. With load balancing, cloud based applications can achieve high availability and reliability. Since multiple resources under load balancer are used to serve the user requests, in the event of failure of one or more of the resources, the load balancer can automatically reroute the user traffic to the healthy resources. To the end user accessing a cloud-based application, a load balancer makes the pool of servers under the load balancer appear as a single server with high computing capacity. The routing of user requests is determined based on a load balancing algorithm.

VII. EXISTING LOAD BALANCING ALGORITHMS

TASK SCHEDULING BASED ON LOAD BALANCING

To meet the dynamic requirements of users, this algorithm performs two levels of task scheduling mechanism which are based on load balancing. It achieves high resource utilization. This algorithm performs load balancing by

i) Mapping tasks to virtual machines.

ii) Mapping all virtual machines to host resources.

The advantages are:

- Improved task response time.
- Better resource utilization.

1.6.1 MIN-MIN ALGORITHM

Load balance Min-Min (LBMM) assigns sub-tasks to the node which requires minimum execution time. The pseudo-code is following:

```
Minmin()
{
        generate a completionTime matrix
        for each task in taskList
        {
                find minimum completionTime from matrix;
                assign task to respective vm;
                update the completionTime;
        }
}
```

1.6.2 HONEYHIVE ALGORITHM

The Honeyhive algorithm is inspired by the "behavior of a colony of honeybees foraging and harvesting food."[7] Forager bees search for food, return to the hive and describe the food they found through a "waggle dance." The "waggle dance" can show the quantity, quality and distance of the food. For the Honeyhive algorithm, every server first plays a forger bee role and satisfies requests from virtual servers. With the service done, each server evaluates the profitability of its just-serviced virtual server. Then a server will adjust the advert board, which serves as a "waggle dance", and record the profitability of virtual servers. If the calculated profitability is high, a server will continue to serve the current virtual server. Otherwise, it will keep waiting.

1.6.3 ROUND ROBIN ALGORITHM

The round robin algorithm works in a round robin fashion. When the client provide a request, then the datacenter accept the request and it notifies the round robin load balancer to allocate a new virtual machine id to data center controller for processing. In this way the subsequent requests are processed in a circular order.

Weighted round robin assigns weight to each virtual machine . so that if one virtual machine capable of handling twice as much load as the other, then former gets the weight of 2 where as the later gets the weight of 1.

The major issue in this allocation is that it does not consider the advanced load balancing requirements such as processing times for each individual result.

1.6.4 ANT COLONY OPTIMIZATION

At first, the ants wander randomly. When an ant finds a source of food, it walks back to the colony leaving "markers" (pheromones) that show the path has food. When other ants come across the markers, they are likely to follow the path with a certain probability. If they do, they then populate the path with their own markers as they bring the ant colony algorithm is an algorithm for finding optimal paths that is based on the behaviour of ants searching for food. As more ants find the path, it gets stronger until there are a

couple streams of ants travelling to various food sources near the colony.

Because the ants drop pheromones every time they bring food, shorter paths are more likely to be stronger, hence optimizing the "solution." In the meantime, some ants are still randomly scouting for closer food sources. A similar approach can be used find near-optimal solution to the travelling salesman problem.

Once the food source is depleted, the route is no longer populated with pheromones and slowly decays.

Ant algorithms is a multiagent approach to difficult combinatorial optimization problems. Example include, Travelling salesman Problem and the quadratic assignment problem . Behavior of ant is directed more to the survival of the colonies, but not for individual.

1.6.5 RANDOMIZED ALGORITHM

This algorithm is static in nature. Here a process can be handled by a particular node n with probability p. This algorithm works well when all the processes are of equal loaded. It is not suitable when the loads are of different computational complexities. This algorithm is not maintaining a deterministic approach.

1.6.6 OPTIMISTIC LOAD BALANCING ALGORITHM (OLB)

This algorithm is to attempt each node keep busy. So, it doesn't consider the current workload of each computer. OLB assigns each task in free order to present node of useful. The advantage is that it is simple. The disadvantage is that it doesn't consider each expectation execution time of task. It leads to the whole completion time is very poor.

1.6.7 BIASED RANDOM SAMPLING

Biased Random Sampling bases its job allocation on the network represented by a directed graph. For each execution node in this graph, in-degree means available resources and out-degree means allocated jobs. In-degree will decrease during job execution while out-degree will increase after job allocation.The pseudo-code is following:

```
BiasedRandomSampling()
{
for each task in task queue
        init walklength=0;
        while (task is assigned to a vm) or
        (walklength > threshold)
        {
        Increment walklength
        assign task to vm if indegree > 0; decrement indegree
        }
        Remove task from task queue
}
Process completed tasks()
{
        Increment degree of vm assigned to the task
}
```

1.6.8 ACTIVE CLUSTERING

Active Clustering is a self-aggregation algorithm to rewire the network.Active Clustering works on the principle of grouping similar nodes together and working on these groups. The process involved is:

- A node initiates the process and selects another node called the matchmaker node from its neighbors satisfying the criteria that it should be of a different type than the former one.
- The so called matchmaker node then forms a connection between a neighbor of it which is of the same type as the initial node.
- The matchmaker node then detaches the connection between itself and the initial node.

The above set of processes is followed iteratively.

1.6.9 A LOCK-FREE SOLUTION FOR LOAD BALANCING

Liu et al. proposed a lock-free multiprocessing load balancing solution that avoids the use of shared memory in contrast to other multiprocessing load balancing solutions which use shared memory and lock to maintain a user session. It is achieved by modifying Linux kernel. This solution helps in improving the overall performance of load balancer in a multi-core environment by running multiple load-balancing processes in one load balancer. By modifying Linux kernel

1.6.10 SHORTEST RESPONSE TIME FIRST

Shortest Response Time First (SJF) algorithm associates with each process the length of the process's next CPU burst. When the CPU is available, it is assigned to the process that has the smallest next CPU burst. If the next CPU bursts of two processes are the same, FCFS scheduling is used. The SJF policy selects the job with the shortest processing time first.

In SJF, it is very important to know or estimate the processing time of each job which is major problem a

VIII. CONCLUSION

Computation in cloud is done with the aim to achieve maximum resource utilization with higher availability at minimized cost. Load balancing is the main task to achieve maximum utilization of resources. In this paper we addressed various load balancing techniques and their applicability in cloud computing environment. Here we categorized the

algorithms as static and dynamic. First we described the algorithms which can be applied to static cloud computing environment next we described various dynamic algorithms for cloud computing environment. Static load balancing is fail to model heterogeneous nature of cloud. But provide easiest simulation and monitoring environment. Dynamic load balancing algorithms are best suited for heterogeneous nature cloud computing environment, but it is difficult to simulate. However, the performance of the cloud computing environment can be improved by modeling the dependencies between the tasks using workflows.

REFERENCES

[1] Klaithem Al Nuaimi, Nader Mohamed, Mariam Al Nuaimi and Jameela Al-Jaroodi" A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms" 2012 IEEE Second Symposium on Network Cloud Computing and Applications.

[2] Venubabu Kunamneni" Dynamic Load Balancing for the Cloud" International Journal of Computer Science and Electrical Engineering (IJCSEE) ISSN No. 2315-4209, Vol-1 Iss-1, 2012

[3] Ratan Mishra, Anant Jaiswal" Ant colony Optimization: A Solution of Load balancing in Cloud" International Journal of Web & Semantic Technology (IJWesT) Vol.3, No.2, April 2012

[4] Tushar Desai, Jignesh Prajapati "A Survey Of Various Load Balancing Techniques And Challenges In Cloud Computing" INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 2, ISSUE 11, NOVEMBER 2013

[5] Harpreet Kaur,Maninder Singh"A Task Scheduling and Resource Allocation Algorithm for Cloud using Live Migration and Priorities" International Journal of Computer Applications (0975 – 8887) Volume 84 – No 13, December 2013

[6] Prabavathy.B, Priya.K, Chitra Babu "A Load Balancing Algorithm For Private Cloud Storage" IEEE - 316614th ICCCNT 2013 July 4-6, 2013, Tiruchengode, India

[7] Gaochao Xu, Junjie Pang, and Xiaodong Fu "A Load Balancing Model Based on Cloud Partitioning for the Public Cloud" IEEE TRANSACTIONS ON CLOUD COMPUTING YEAR 2013

[8] Martin Randles, David Lamb, A. Taleb-Bendiab "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing" 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops

[9] Albert Y. Zomaya, Senior Member, IEEE, and Yee-Hwei The "Observations on Using Genetic Algorithms for Dynamic Load-Balancing" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 12, NO. 9, SEPTEMBER 2001

[10] Shu-Ching Wang, Kuo-Qin Yan, Wen-Pin Liao and Shun-Sheng Wang "Towards a Load Balancing in a Three-level Cloud Computing Network" 978-1-4244-5540-9/10/©2010 IEEE

[11] Jenn-Wei Lin ,Chien-Hung Chen, Chi-Yi Lin "Integrating QoS awareness with virtualization in cloud computing systems for delay-sensitive applications" © 2014 Elsevier B.V.

[12] A.Meera , S.Swamynathan "Agent based Resource Monitoring system in IaaS Cloud Environment" International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013

[13] Radojevic, B., & Zagar, M. (2011). Analysis of issues with load balancing algorithms in hosted (cloud) environments. In MIPRO, 2011 Proceedings of the 34th International Convention, 416-420. IEEE.

[14] Jorge M. Cortés-Mendoza, Andrei Tchernykh, Johnatan Pecero, Pascal Bouvry, Laura Cruz-Reyes " Distributed VoIP Load Balancing in Cloud Computing"

[15] Ms.Shilpa D.More, Mrs.Smita Chaudhari "Reviews of Load Balancing Based on Partitioning in Cloud Computing" (IJCSIT) International Journal of Computer Science