Workload Management Using Load Balancing Algorithm in Cloud Computing

Khushbu shah[1], Prof. Ochchhav Patel[2]
[1, 2] Department of Computer Engineering
[1, 2] LDRP-ITRCE-IT Department,LDRP-ITR, Gandhinagar, India

Abstract- *Cloud computing helps to share data and provide many resources to users. Users pay only for those resources as much they used. Cloud computing stores the data and distributed resources in the open environment. The amount of data storage increases quickly in open environment. So, load balancing is a main challenge in cloud environment. Load balancing is helped to distribute the dynamic workload across multiple nodes to ensure that no single node is overloaded. It helps in proper utilization of resources .It also improve the performance of the system. Many existing algorithms provide load balancing and better resource utilization. There are various types load are possible in cloud computing like memory, CPU and network load. Load balancing is the process of finding overloaded nodes and then transferring the extra load to other nodes.*

Keywords:- Cloud Computing, Load Balancing, load balancing algorithms,cloudsim

I. INTRODUCTION

Cloud computing is a new technology .It providing online resources and online storage to the user's .It provide all the data at a lower cost. In cloud computing users can access resources all the time through internet. They need to pay only for those resources as much they use .In Cloud computing cloud provider outsourced all the resources to their client. There are many existing issues in cloud computing. The main problem is load balancing in cloud computing. Load balancing helps to distribute all loads between all the nodes. It also ensures that every computing resource is distributed efficiently and fairly. It helps in preventing bottlenecks of the system which may occur due to load imbalance. It provides high satisfaction to the users. Load balancing is a relatively new technique that provides high resource utilization and better response time.

A. Cloud computing consist of several characteristics:

- On demand service
- Broad Network Access
- Resource Pooling
- Measured Service
- Rapid Elasticity

Challenges in Cloud Computing

There are many challenges in cloud computing:-

1. Security
2. Efficient load balancing
3. Performance Monitoring
4. 4.Consistent and Robust Serviceabstractions
5. Resource Scheduling
6. Scale and QOS management
7. Requires a fast speed Internet connection.

II. CLOUD COMPUTING MODEL

Fig: 1 shows Cloud computing model which consist services of cloud and different deployment models as:

A. Services of Cloud Computing:

Service means different types of applications provided by different servers across the cloud. There are many services are provide to the users over cloud.

1) Software as a Service (SaaS): Saas provided all the application to the consumer which are provided by the providers. Applications are running on a cloud infrastructure. Interfaces (web browser) are used access the applications. The consumer does not control the internal function..

There are some of applications of software of services:-

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

Advantages:

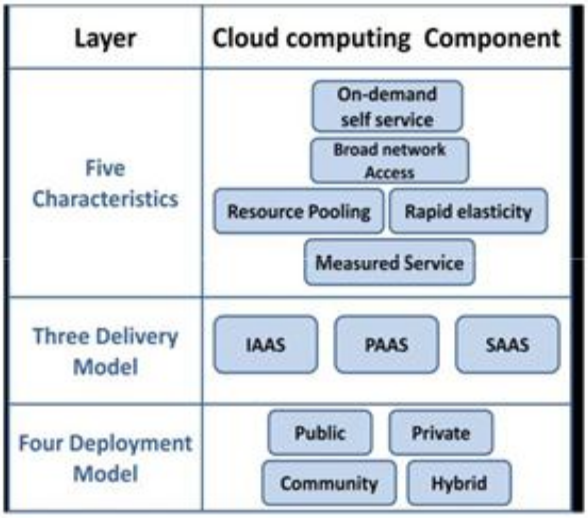
1) The main advantage of SaaS is costing less money than buying the whole application.
2) It provides reliable and cheaper applications.
3) More bandwidth.

2) Platform as a Service (PaaS): PaaS provides all the resources to the customers that are required for building applications. It provides all the services on the internet User not need to download and install the software. The consumer does not control network, servers, operating systems, or storage. Consumer controls all applications which they deploy.

Disadvantages

- There is very less portability among different providers.

3) Infrastructure as a Service (IaaS): In this service consumer does not manage or control the underlying cloud infrastructure. In infrastructure as a service consumer able to control operating systems, storage, and all applications which they deployed. Infrastructure Providers control storing and processing capacity. Virtualization is used assign and dynamically resizes these resources to build systems as demanded by customers.



B. Layers of Services

All the services have number of layers. Which manage by the users and providers.

Cloud Deployment Models:

1. Public Cloud: The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization .Anyone can use public cloud as they want without restriction.
2. Private Cloud: The cloud infrastructure is used by a single organization. Private cloud is only managed by the organization or a third party
3. Community Cloud: The cloud infrastructure is shared by many organizations .Community cloud supports a specific community that has shared concerns.Ex:-security requirements, policy, compliance considerations. It may be managed by the organizations or a third party.
4. Hybrid Cloud: Hybrid cloud is a combination of two or more clouds (private, community, or public). That remains unique entities but is bound together by standardized technology that enables data and application portability.
5. Ex:- cloud bursting for load-balancing between clouds.

III. OBJECTIVES

- Server balances the load by using external resources
- No proper utilization of resources
- Performed in private cloud
- Improves the response time and effective job processing
- Network reliability

IV. LOAD BALANCING

Load balancing is used to distributing a larger processing load to smaller processing nodes for enhancing the overall performance of system.

Load balancers can work in two ways: one is cooperative and non-cooperative. In cooperative, the nodes work simultaneously in order to achieve the common goal of optimizing the overall response time. In non-cooperative mode, the tasks run independently in order to improve the response time of local tasks.

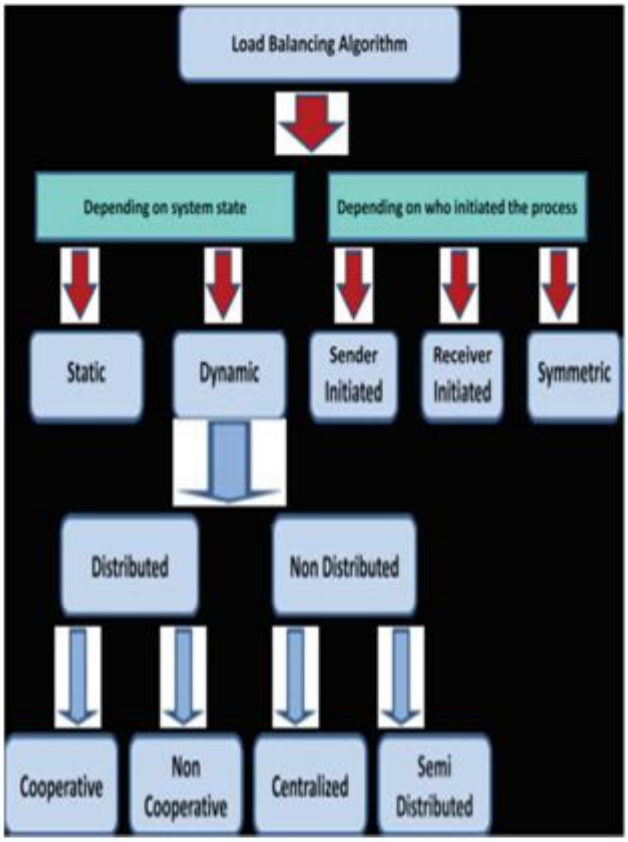
Load balancing algorithms, in general, can be divided into two categories: static and dynamic load balancing algorithm. A static load balancing algorithm does not take into account the previous state or behavior of a node while distributing the load. On the other hand, a dynamic load balancing algorithm checks the previous state of a node while distributing the load. The dynamic load balancing algorithm is applied either as a distributed or non-distributed. The advantage of using dynamic load balancing is that if any node fails, it will not halt the system; it will only affect the system Performance.

- Load balancing helps in fair allocation of computing resource to achieve a high User satisfaction and proper Resource utilization .High resource utilization and Proper load balancing helps in minimizing resource consumption. It helps in implementing fail over, scalability, and avoiding bottlenecks.

- Load balancing is a techniques that helped networks and resources by providing a Maximum throughput with minimum response time. Load balancing is dividing the traffic between all servers, so data can be sent and received without any delay with load balancing.
- In our daily life example of load balancing is websites. Users could experience many Problems without Load balancing like delays, timeouts and long system responses.

A. Load balancing classification:

Fig. represents different load balancing algorithms. This is mainly divided into two categories: static load balancing algorithm and dynamic load balancing algorithm:



1) Static approach: - This approach is mainly defined in the design or implementation of the system. Static load balancing algorithms divide the traffic equivalently between all servers.
2) Dynamic approach:- This approach considered only the current state of the system during load balancing decisions. Dynamic approach is more suitable for widely distributed systems such as cloud computing. Dynamic load balancing approaches have two types
   a) Centralized approach: - In centralized approach, only a single node is responsible for managing and distribution within the whole system. Other all nodes are not responsible for this.
   b) Distributed approach: - In distributed approach, each node independently builds its own load vector. Vector collecting the load information of other nodes. All decisions are made locally using local load vectors. Distributed approach is more suitable for widely distributed systems such as cloud computing.

B. Metrics for Load Balancing:

1. Throughput: - It is used to calculate the all tasks whose execution has been completed. The performance of any system is improved if throughput is high.
2. Fault Tolerance: -It means recovery from failure. The load balancing should be a good fault tolerance technique.
3. Migration time: -It is the time to migrate the jobs or resources from one node to other nodes. It should be minimized in order to enhance the performance of the system.
4. Response Time: - It is the amount of time that is taken by a particular load balancing algorithm to response a task in a system. This parameter should be minimized for better performance of a system.
5. Scalability: - It is the ability of an algorithm to perform Load balancing for any finite number of nodes of a system. This metric should be improved for a good system.

C. EXISTING LOAD BALANCING ALGORITHMS

There are many load balancing algorithms which help to achieve better throughput and improve the response time in cloud environment. All the algorithms have their own benefits. [16] [17] [18]

1. Task Scheduling based on LB: This algorithm mainly consist two level task scheduling mechanism which are based on load balancing to meet dynamic requirements of users.It is improving the task response time.It also provide better resource utilization .
2. Opportunistic Load Balancing: OLB is to attempt each node keep busy, therefore does not consider the present workload of each computer. OLB assigns each task in free order to present node of useful .
3. Round Robin: - In this algorithm all the processes are divided between all processors. In this each process is assigned to the processor in a round robin order. The work load distributions between processors are equal. Different processes have not same job processing time.When time quantum is very large then RR Scheduling Algorithm is same as the FCFS Scheduling.

and when time quantum is too small then Round Robin Scheduling is known as Processor Sharing Algorithm.

4. Min-Min Algorithm: It starts with a set of all unassigned tasks .In this minimum completion time for all tasks is found. Then after that among these minimum times the minimum value is selected. The main problem of this algorithm is has a starvation.
5. Max-Min Algorithm: Max-Min algorithm is almost same as the min-min algorithm. The main difference is following: In this algorithm first finding out minimum execution times, then the maximum value is selected which is the maximum time among all the tasks on any resources. Then all assigned task is removed from the list that executed by the system
6. Honeybee Foraging Behavior: It is a nature inspired Algorithm for self-organization. Honeybee achieves global load balancing through local server actions. The performance of the system is enhanced with increased system diversity. The main problem is that throughput is not increased with an increase in system size. When the diverse population of service types is required then this algorithm is best suited.
7. Active Clustering:- In this algorithm same type nodes of the system are grouped together and they work together in groups. It works like as self-aggregation load balancing technique where a network is rewired to balance the load of the system. Systems optimize using similar job assignments by connecting similar services. System Performance improved with improved resources. The throughput is improved by using all these resources effectively.
8. Ant Colony Optimization :- Ant algorithms is a multiagent approach to difficult combinatorial optimization problems . Example of this approach is travelling salesman problem (TSP) and the quadratic assignment problem (QAP) . These algorithms were inspired by the observation of real ant colonies. Ant's behaviour is directed more to the survival of the colonies

V. SIMULATION IN CLOUD: CLOUDSIM

Resources and software are shared on the basis of client's demand in cloud environment. Essentially, dynamic utilization of resources is achieved under different conditions with various previous established policies. Sometime it is very much difficult and time consuming to measure performance of the applications in real cloud environment. In this consequence, simulation is very much helpful to allow users or developers with practical feedback in spite of having real environment. In this research work, simulation is carried out with a specific cloud simulator, CloudSim [7].

A. Importance of simulation technique:

Simulation means mimicking the actual environment towards benefit of the research. The user or researcher can actually analyze the proposed design or existing algorithms through simulation. They can check the efficiency and merit of the design before the actual system is constructed. Simulation is advantageous to the users, as they can explore the benefit of that design repeatedly.

This actually reduces the cost of reconstruct as changes have been made during design time. Even most of the time researchers could carry out benchmark experiments repeatedly in scalable environment for evaluating different aspects.

B. Cloud Simulator- CloudSim

Users are capable of accessing shared resources through utilizing available public cloud platform. However, accessing real cloud environment or public cloud is not always handy. Instead of the real environment, cloud simulator could facilitate the experiments. Simulation environment allows customers or users to tune the performance bottlenecks or evaluates different kinds of features under varying load distributions[7]. Different kinds of functionalities of CloudSim are presented in the following [4].

- support for modeling and simulation of large scale cloud computing data centers
- support for modeling and simulation of virtualized server hosts, with customizable policies for provisioning host resources to virtual machines
- support for modeling and simulation of energy-aware computational resources
- support for modeling and simulation of datacenter network topologies and message-passing applications.

VI. ARCHITECTURE OF CLOUDSIM

Equal load distribution may improve resource utilization by transferring load from heavily loaded server to the idle server. Existing scheduling algorithms [13], [19], [24],[25] estimate system parameters such as the job arrival rate, CPU processing rate, and load on the processor for migrating jobs into least loaded processors in order to balance load.

In [2], Bhathiya presents that Cloud Analyst can evaluate any algorithm or application deploying in the cloud. VM life cycle starts from provisioning of a host to a VM, VM creation, VM destruction, and VM migration[2]. Figure **3** represents a class diagram of the CloudSim components

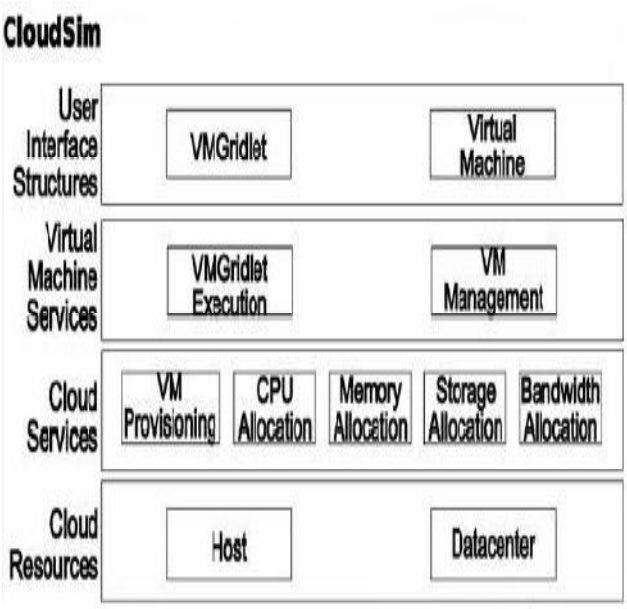
considered for execution analysis. In addition to that, a brief description of these components and the working relationship between them is presented in the following:

Datacenter: Datacenter encompasses a number of hosts in homogeneous or heterogeneous configurations (memory, cores, capacity, and storage). It also creates the bandwidth, memory, and storage devices allocation.

Virtual Machine (VM): VM characteristics comprise of memory, processor, storage, and VM scheduling policy.

Host: This experiment considers VM need to handle a number of cores to be processed and host should have resource allocation policy to distribute them in these VMs. So host can arrange sufficient memory and bandwidth to the process elements to exeute them inside VM. Host is also responsible for creation and destruction of VMs.

Cloudlet: Cloudlet is an application component which is responsible to deliver the data in the cloud service model. So the length, and output file sizes parameter of Cloudlet should be greater than or equal to Experimental results of executing task in CloudSim are represented in the next section.



VII. CONCLUSION

Load balancing is one of the main challenges in cloud computing. It is required to distribute the dynamic local workload evenly across all the nodes to achieve a high user satisfaction and resource utilization ratio by making sure that every computing resource is distributed efficiently and fairly. With proper load balancing, resource consumption can be kept to a minimum which will further reduce energy consumption and carbon emission rate which is a desire need of cloud computing. This study explains the concept of load balancing, general idea about static and dynamic load balancing algorithms and explains various comparative parameters of different load balancing algorithms. This study also gives an overall description of various distributed load balancing algorithms that can be used in case of clouds. So there is a need to develop an adaptive algorithm which is suitable for heterogeneous environment and should also reduce the cost.

REFERENCES

[1] Mell, Peter and Grance, Tim, “The NIST definition of cloud computing”, National Institute of Standards and Technology, 2009,vol53, pages50, Mell2009

[2] http://www.ancoris.com/solutions/cloudcomputing.html, "Cloud Computing |Google Cloud-Ancoris."

[3] Foster, I., Zhao, Y., Raicu, I. & Lu, S. (2008). Cloud computing and grid computing 360-degreecompared. IEEE Grid Computing Environment Workshops.

[4] S. S. MOHARANA, R. D. RAMESH, D. POWAR, “Analysis of load balancers in cloud computing” International Journal of Computer Science and Engineering,2013,vol 2,pages101-108

[5] NayandeepSran,NavdeepKaur , “Comparative Analysis of Existing Load Balancing Techniques in Cloud Computing ”,vol 2,jan 2013

[6] N. Ajith Singh, M. Hemalatha, “An approach on semi distributed load balancing algorithm for cloud computing systems” International Journal of Computer Applications Vol-56 No.12 2012.

[7] T. Kokilavani, Dr. D. I. George Amalarethinam “Load Balanced Min-Min Algorithm for Static Meta Task Scheduling in Grid computing” International Journal of Computer Applications Vol-20 No.2, 2011

[8] Chaczko, Zenon and Mahadevan, Venkatesh and Aslanzadeh, Shahrzad and Mcdermid, Christopher, “Availability and load balancing in cloud computing”, International Conference on Computer and Software Modeling, Singapore, chaczko2011availability

[9] Wang, L., Tao, J., Kunze, M., Castellanos, A., Kramer, D., & Karl, W(2008). “Scientific CloudComputing:

Early Definition and Experience," 825-830. Proceedings of the 10th IEEE InternationalConference on High Performance Computing and Communications (HPCC '08). Dalian, China, September 2008. IEEE

[10] ZenonChaczko, VenkateshMahadevan, Shahrzad Aslanzadeh, Christopher Mcdermid(2011)"Availabity and Load Balancing in Cloud Computing" International Conference on Computer and Software Modeling IPCSIT vol.14 IACSIT Press,Singapore 2011

[11] ZhongXu, Rong Huang,(2009)"Performance Study of Load Balanacing Algorithms in Distributed Web Server Systems", CS213 Parallel and Distributed Processing Project Report.

[12] http://www.writeaprogram.info/c/os-programs/priority-scheduling/"Priority Scheduling Algorithm Example – Write a Program.''