A load aware matrix approach Load balancing in cloud computing

PRASANTHI G[1], G SRINIVASA RAO[2], N Sridhar Babu[3]

[1] *Research Scholar, Department of Information technology, GITAM University, Visakhapatnam*

[2] *Assistant Professor, Department of Information technology, GITAM University, Visakhapatnam.*

Abstract

Cloud computing is originate or grouping of Distributed computing, Grid computing and Parallel computing. Various computers and storages devices are connected to make heterogeneous resources of pool. Virtualization technology enhances the utilization of clouds resources by providing the different resources on the internet which users can buy pay on demand rule. Different service propose by the clouds computing are software as service (SaaS), Infrastructure as service (PaaS) and Platform as services (PaaS). For proper utilization of the resources on virtual machines and physical machines load balancer is need for distribution of resources. In this paper we discuss the virtualization technology and load balancing algorithm. A load aware matrix has been proposed and discussed.

Keywords: *Load Balancing, Virtualization, Resource Scheduling*

Introduction

Cloud computing is originated or combination of Distributed computing, Grid computing, and Parallel computing. Many computers and storages devices are connected to make heterogeneous resources of pool. Virtualization technology enhances the utilization of clouds resources by providing the different resources on the internet which users can buy pay on demand rule. Different service offer by the clouds computing are software as service (SaaS), Infrastructure as service (PaaS) and Platform as services (PaaS). For proper utilization of the resources on virtual machines and physical machines load balancer is need for distribution of resources. In this paper we discuss the virtualization *technology* and load balancing algorithm. A load aware matrix has been proposed and discussed.

Cloud computing is merges of distribution computing where resources and task are distributed to different geographical location [1]on different computing systems to perform the computational task in fast and efficient manners .Different service provided by the cloud computing are platform as service(PaaS) in which client use the different programming and networking tools to perform their jobs without any configuration or installation of these tools, Infrastructure as services(IaaS) no need to purchases the different computing resources like CPU, memory, storage and user can deploy or install their software and operating system which he want ,Software as services(SaaS) users can obtain the different application online they do not need to manage or install these application .The main aim of cloud computing is provides the different resources when the user demanded at low cast **[2]**.Load balancing is the technique which balanced the all these heterogeneous resources and different task which are performed in the distribution fashion to make sure that all resources and task assign to different nodes or virtual machines are properly balanced .

Cloud computing provide different services[3] pay on demand scale means user can rented the server or machine for completing is task after completing is task they can leave the services .The services provided by different cloud are *Platform as services PaaS* are the services where user can rented a machine and the services provides provide them runtime environment means operating system and different databases you are installed additionally they deploy their own software Google app is one the PaaS service provider, *Software as services SaaS* **[4]** are the services on the cloud which users can used online they do not need to install these services they just request to the cloud services provider or go to the internet and use the software they want like online word editor and playing online games, *Infrastructure as services IaaS* are the services which user or any organization needed computing resources or equipment i.e. memory ,CPU networking services or any other requirement they want can get .User can deploy these equipment and performs there computational task on the virtual machines .Amazon ,Azure are the IaaS services provider.

Cloud Computing Services
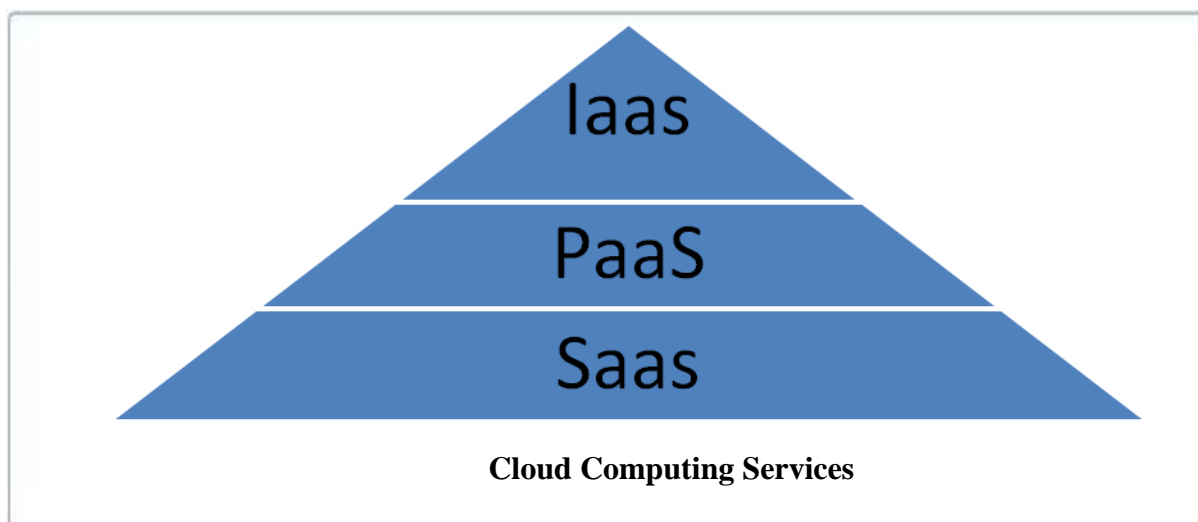
Virtualization in Cloud computing

Virtual environment to provide the service to client. Different and multiple operating system are running on the single physical machine by using the virtualization, so the resources can be efficiently used. To enhance the performance different hardware resources are combined. For the proper utilization of the resources different algorithms and software are used called hypervisor**[5]**.When a virtual machine is created on physical machine it requires some resources i.e. CPU ,memory and energy management are the important factors .In this paper author specially focus on the Memory management of the virtual machine .discus the different algorithm, framework and application that how memory is dynamically assign to a virtual machine .

Virtual Machines Architecture
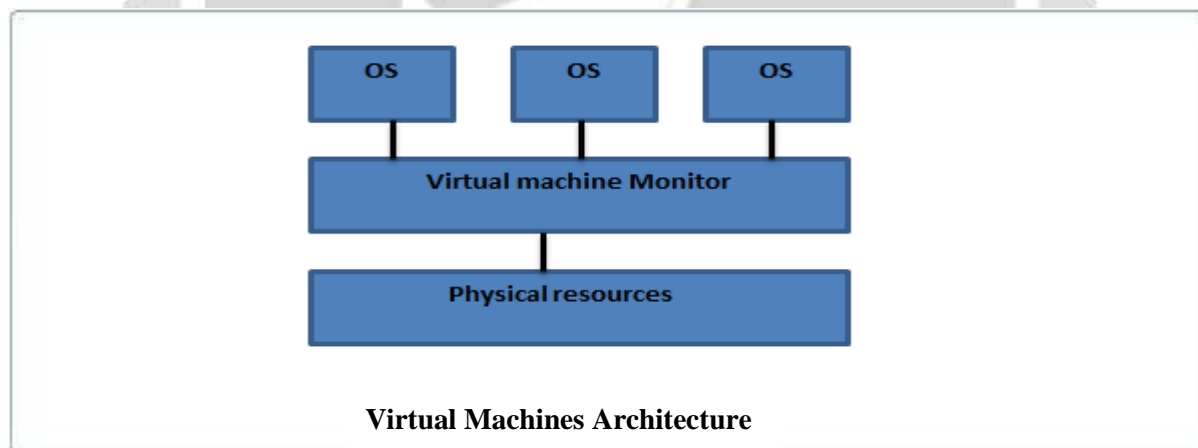
Load Balancing

Load balancing[6] is the method of optimizing the resources on clouds virtual machines .Main purpose is distribution of resources and task among all machine equally so that no node is overloaded or idle .load balancing is classified in two main classes' dynamic and static algorithms.

- Reason using of load balancing are following
- Reduce waiting time
- Reduce the response time
- Increase the utilization of resources
- Improve reliability
- Increase through output
- Load balancing enhance the performance of the system by managing every node

Load Balancing Working
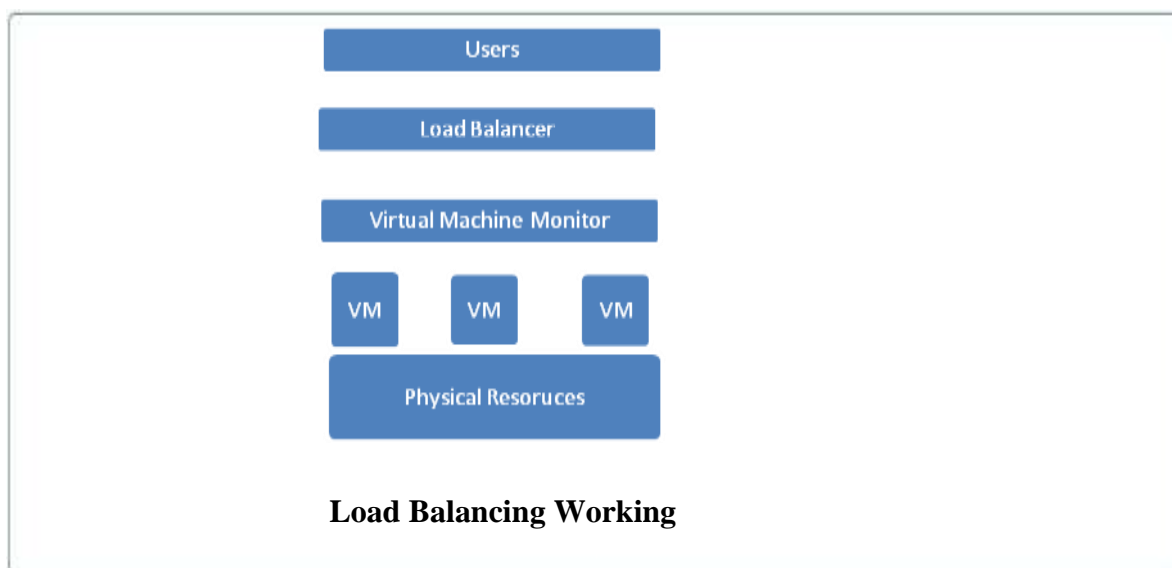
[7] the detail architecture of the load balancer that how it works is explain in and which factors are involved in the efficient utilization of resources and how a job or task is assign to the balancer and how it deals with assigned tasks. In *dynamic load balancing algorithms* task and resources to the nodes are managing dynamically as the requirement changes the resources are allocated according to them. In *static load balancing algorithm* the state of the nodes and the resources needed to complete are gathered before assigning a to any node .Dynamic algorithms are divided in two type distributed and non-distributed .In distributed mechanism use the dynamic load balancing .How the different nodes interact with each other have forms .One is non- cooperative in which each node work independently to achieve a common goal .Second is cooperative in which nodes work together to achieve a common object .

Motivation

virtualization in which all the cloud computing processing, storage, task assigning is done using this technology, so virtualization is core backbone of cloud computing. Furthermore to manage the resources of physical devices create different virtual machines now the problem arise that how much resources are assign to machines and which resources are needed to that machine after allocating the resources to the machines now the problem arise that which machine perform which task or which task is assign to which machine Here the   Load balancing in cloud computing comes to solve all these above problems which is very important task for the virtual machines and their task management **.**

Objectives

The main objectives of the study under consideration are as follows:

- Develop a dynamic load balancing algorithm
- Virtual machine monitoring
- Managing the virtually resources efficiently
- Increase the throughput of the system
- Reduce the overhead on machines
- Response Time, Performance, Waiting Time, Scalability, Power Consumption, Speed of the virtual machines

ssState-Of-The-Art: Literature Review and Existing Solutions

Xen [8] is a type-1 (bare metal) open-source virtual machine monitor (VMM)1 used in commercial clouds, such as Amazon EC2 and Rack Space. It sits between the hardware and operating systems. From the scheduling perspective, each virtual machine (called a domain in Xen) contains multiple VCPUs that are scheduled by a VMM scheduler on a host with multiple physical CPUs (PCPUs). At boot time, Xen creates a privileged domain called domain 0, which is responsible for managing the other guest domains. Round Ribbon based [9] fixed amount of time is assign to each task. It uses the ring to arrange the task. Because time is equally assign to each node. The data center controller assigns the task to the virtual machine in a ring form picked the VM randomly and assign the task. But the problem is that it does not check the state of the VM either its free or not .Throttled load Balancing Algorithm In [9] load balancer main the list of VM and their states .if the client request for VM then datacenter scan the list of virtual machines which are free and assign the task the suitable machine.

Min-Min Scheduling algorithm This algorithm [10]works on the set of the task which required less resources and less complication time . Min-Max scheduling algorithm In the algorithm[11] load balancer pick the task which needed more resources and takes some time to complete the task. The problem in this algorithm is the task completion time is too long so the other process has to wait for the complication of the task .Ant colony load balancing algorithm [11] when a task is received then the ant move to the source form food to the parent node or super node regional load balancing node is called the super node or head node .they the record of the all node which have been visited or has any assignment Honey bee load balancing algorithm [9] load balancing is achieved when there is heterogonous environment with the help of local server .current on the virtual machine and their states are recorded then the machines are grouped according to the load .when the new task is arrived then is allocated to the least loaded virtual machine.

Resource-aware scheduling algorithm It is combination of two algorithm min-min and min-mix it covers the advantages in these algorithms and remove the flaws in these algorithms [12].Algorithm Active Monitoring Load Balancing (AMLB) [13]I the author discussed the work same as mention above different load balancing algorithm static and dynamic and their categories .Also proposed a load balancing algorithm Active Monitoring Load Balancing (AMLB) Algorithm tested through simulation and also share the simulation result that how this is more useful in the cloud datacenters Least Connection Algorithm This algorithm[14] is basically depends on the connection on a single node. The load balancer of this algorithm firstly picks the node that have the smallest number of connection, so load can be transferred. The load balancer also maintains the set of connection on every node. Whenever a new connection is implemented the number of connection increases and whenever the old connection completes, or time out action occur the number of connection decreases. Load balancer maintains all the increasing and decreasing connection related information.

Stochastic Hill Climbing Algorithm [15]adapts the incomplete method for solving optimization problem. Stochastic Algorithm is like a loop which continuously keeps moving upward or uphill i.e. moving towards the direction of higher value. When there is no higher value neighbor, it stops. It has two elements one is candidate generator and the other an evaluation criterion. Candidate generator is the one which maps single solution to a set of possible successors. The Evaluation criteria evaluate the rank for each valid solution. The better ranking means closer to or valid solution .Integration Of A Real-Time Hypervisor And A Cloud Management System (Open stack) [16]through a real time resource interface second a real time VM scheduler to allow regular VMs to share hosts with RT VMs without interfering the real-time performance of RT VMs; and a VM-to-host mapping strategy that provisions real-time performance to RT VMs while allowing effective resource sharing with regular VMs .[17]Application used by the customer need large data retrieval then the efficiency of the system may be decreased if the task scheduled by the execution time. Moving applications to a cloud computing environment triggers the need of scheduling as it enables the utilization of various cloud services to facilitate execution.[18]Developing a algorithm relevant to the load balancing i.e. CPU management memory management and other factor that affect the performance of VM in cloud computing .Moreover also discussed the matrix through which we can test our algorithm performance relevant to the resources also mention the simulator which are used in testing the algorithms .He also suggest that algorithm can be tested more accurately in the real time .He also discussed the advantages and disadvantages of different matrix based of the classification of the algorithm

 RIAL: Resource Intensity Aware Load Balancing in Clouds[19]. Migration of live virtual machines new scenario that clouds currently perform load balancing by migrating virtual machines (VMs) from heavily loaded physical machines (PMs). The unique features of clouds pose formidable challenges to achieving effective and efficient load balancing. First, VMs in clouds use different resources (e.g., CPU, bandwidth, memory) to serve a variety of services (e.g., high performance computing, web services, file services), resulting in different over utilized resources in different PMs. Also, the over utilized resources in a PM may vary over time due to the

time-varying heterogeneous service requests. Second, there is intensive network communication between VMs Load Balance Min-Min (LBMM),[20] Two-Phase Load Balancing Algorithm. Load balancing algorithms are also classified on the bases of cost Load Balancing with Optimal Cost Scheduling Algorithm, Cost Effective Load Balanced Resource Allocation for Partitioned Cloud System. The author proposed a cluster-based load balancing algorithm proposed algorithm is divided in to two parts (MET) is produced by merging the best algorithms (Equally Spread Current Execution algorithm and throttled algorithm) which quite good.

Conclusion Future Work

Although different dynamic and static algorithms are proposed and developed in recent time each one has own advantages and disadvantages the problem I problems mainly I focused to solve the execution time of nodes, overloading on the different machines., Waiting of high time-consuming tasks is very long, Delay in operation, Delay in operation due to long search to find acceptable Virtual Machines. Our study tries to address all these issue in load balancing algorithms

References

1.   Priyanka Singh, P.B., *Assorted Load Balancing Algorithms in Cloud Computing: A Survey.* International Journal of Computer Applications (0975 – 8887), 2016.

2.   Jain, S., *A Survey of Load Balancing Challenges in Cloud Environment.* Proceedings of the SMART - 2016, IEEE Conference ID: 39669, 2016.

3.   P.S.I.T., P.S., *Assorted Load Balancing Algorithms in Cloud Computing: A Survey.* International Journal of Computer Applications (0975 – 8887), 2016.

4.   Ghomi, E.J., *Load-balancing algorithms in cloud computing: A survey.* Journal of Network and Computer Applications 2017.

5.   DING1, Y., *Review of Virtual Memory Optimization in Cloud Environment* 2016 International Conference on Intelligent Networking and Collaborative Systems, 2016.

6.   Ghomia, E.J., *Load-balancing algorithms in cloud computing: A survey.* Journal of Network and Computer Applications 88 50–71, 2017.

7.   Jain, S., *A Survey of Load Balancing Challenges in Cloud Environment* Proceedings of the SMART - 2016, IEEE Conference ID: 39669, 2016.

8.   Xi, S., *Ttowards real-time hypervisor scheduling in xen," in Embedded 185 Software (EMSOFT.* 2011 Proceedings of the International Conference on.

9.   Athokpam Bikramjit Singh1, *Survey on Various Load Balancing Techniques in Cloud Computing.* 2017.

10. Swathi I J Singh, *Review: Different Improvised Min-Min Load Balancing Algorithm in Cloud Computing Environment.* Journal of Computer and Mathematical Sciences, Vol.7(11), 540-550, , November 2016.

11. aslam, s., *Load Balancing Algorithms in Cloud Computing: A Survey of Modern Techniques.* National Software Engineering Conference (NSEC 2015), 2015.

12. Jiangtao Zhang, *Resource Provision Algorithms in Cloud Computing: A Survey.* Journal of Network and Computer Applications, 2015.

13. Shaw, S.B., *Balancing Load of Cloud Data Center using Efficient Task Scheduling Algorithm* International Journal of Computer Applications (0975 – 8887) Volume 159 – No 5, February 2017, 2017.

14.   Shaw, S.B., *Balancing Load of Cloud Data Center using Efficient Task Scheduling Algorithm*
        International Journal of Computer Applications (0975 – 8887) Volume 159 – No 5, February 2017,
        2017.

15.   Mondal B, D.K., Dutta P. , *Load   balancing in cloud computing using stochastic hill climbing-a soft
        computing approach.* . Procedia Technology. 31;4:783-9 Dec (2012).

16.   Xi, S., *RT-OpenStack: CPU Resource Management for Real-Time Cloud Computing.* 2015 IEEE 8th
        International Conference on Cloud Computing, 2015.

17.   Matloobi, R., *Managing performance degradation of collocated virtual machines in private cloud.*
        IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th
        International, 2016.

18.   Ghomi, E.J., *Load-balancing algorithms in cloud computing: A survey."* Journal of Network and
        Computer Applications

19.   (2017), H.S., *"RIAL: Resource Intensity Aware Load Balancing in Clouds.".* IEEE Transactions on
        Cloud Computing vol1.

20.   Minxian Xu1 Wenhong Tian2, *Load Balancing in Cloud Computing.* John Wiley & Sons, Ltd., 2017.