

A LOAD BALANCING MODEL BASED ON CLOUD COMPUTING FOR HETEROGENEOUS PARTITION

Ms. Pranita Narayandas Laddhad¹, Prof. Nitin Raut²

¹U.G Dept Of CSE, ²Asst. Prof. GNIET ,Dept. Of Comp. Science & Engg. Nagpur

¹prnita.laddhad@gmail.com, ²infotech.nitin@gmail.com

Abstract

Load balancing is the process of distributing the load among various nodes of a distributed system to improve both resource utilization and job response time while also avoiding a situation where some of the nodes are heavily loaded while other nodes are idle or doing very little work Good load balancing makes cloud computing more efficient and improves user satisfaction. This article introduces a better load balance model for the cloud based on the cloud partitioning concept with a switch mechanism to choose different strategies for different situations. The load balancing model given in this article is aimed at the cloud which has numerous nodes with distributed computing resources in many different geographic locations. Thus, this model divides the cloud into several cloud heterogeneous partitions. When the environment is very large and complex, these 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. Some of the classical load balancing methods are similar to the allocation method in the operating system, for example, the Round Robin algorithm and the First Come First Served rules..

Keywords— load balancing model; public cloud; heterogeneous cloud partition; cloud computing;

INTRODUCTION :

A Cloud computing is emerging as a new paradigm of large scale distributed computing. It has moved computing and data away from desktop and portable PCs, into large data centres Cloud Computing is widely used to deliver services over the Internet for both technical and economical reasons.[2]The cloud is changing our life by providing users with new types of services. Users get service from a cloud without paying attention to the details NIST gave a definition of cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction More and more people pay attention to cloud computing.[4]

In case of Cloud computing services can be used from diverse and widespread resources, rather than remote servers or local machines. There is no standard definition of Cloud computing. Generally it consists of a bunch of distributed servers known as masters, providing demanded services and resources to different clients known as clients in a network with scalability and reliability of data center. The distributed computers provide on-demand services. Services may be of software resources (e.g. Software as a Service,SaaS) or physical resources (e.g. Platform as a Service, PaaS) or hardware/infrastructure(e.g. Hardware as a Service, HaaS or Infrastructure as a Service, IaaS). Load balancing schemes depending on whether the system dynamics are important can be either static and dynamic. Static schemes do not use the system information and are less complex while dynamic schemes

will bring additional costs for the system but can change as the system status changes. A dynamic scheme is used here for its flexibility. The model has a main controller and balancers to gather and analyze the information. Thus, the dynamic control has little influence on the other working nodes. The system status then provides a basis for choosing the right load balancing strategy. The load balancing model given in this article is aimed at the public cloud which has numerous nodes with distributed computing resources in many different geographic locations. Thus, this model divides the public cloud into several cloud partitions. When the environment is very large and complex, these 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.

LOAD BALANCING ALGORITHMS :

ROUND-ROBIN LOAD BALANCER

It is a static load balancing algorithm, which does not take into account the previous load state of a node at the time of allocating jobs. It uses the round robin scheduling algorithm for allocating jobs. It selects the first node randomly and then, allocates jobs to all other nodes in a round robin manner. This algorithm will not be suitable for cloud computing because some nodes might be heavily loaded and some are not. Since the running time of any process is not known prior to execution, there is a possibility that nodes may get heavily loaded. Hence, weighted round-robin algorithm was proposed to solve this problem. In this algorithm, each node is assigned a specific weight. Depending on the weight assigned to the node, it would receive appropriate number of requests. If the weights assigned to all the nodes are equal, then each node will receive same traffic. In cloud computing system, precise prediction of execution time is not possible therefore, this algorithm is not preferred.

A FAST ADAPTIVE LOAD BALANCING METHOD

Zhang et al.[1] proposed a binary tree structure that is used to partition the simulation region into sub-domains . The characteristics of this fast adaptive balancing method are to be adjusted the workload between the processors from local areas to global areas. According to the difference of workload, the arrangements of the cells are obtained. But the main workload concentrates on certain cells so that the procedure of adjusting the vertices of the grid can be very long because of the local workload can be considered. This problem can be avoided by the fast load balancing adaptive method. Here the region should be partitioned by using the binary tree mode, so that it contains leaf nodes, child nodes, parent nodes etc. There were partition line between the binary tree and the indexes of the cells on the left are smaller that of right and the indexes on the top are smaller than the bottom. Calculate the workload based on the balancing algorithm. This algorithm has a faster balancing speed, less elapsed time and less communication time cost of the simulation procedure. Advantages are Relative smaller communication overhead relative smaller communication overhead, faster balancing speed, and high efficiency and the disadvantage is it cannot maintain the topology that is neighbouring cells cannot be maintained.

HONEY BEE BEHAVIOR INSPIRED LOAD BALANCING

Dhinesh et al. [5] proposed an algorithm named honeybee behavior inspired load balancing algorithm. Here in this session well load balance across the virtual machines for maximizing the throughput. The load balancing cloud computing can be achieved by modeling the foraging behavior of honey bees. This algorithm is derived from

the behavior of honey bees that uses the method to find and reap food. In bee hives, there is a class of bees called the scout bees and the another type was forager bees .The scout bee which forage for food sources, when they find the food, they come back to the beehive to advertise this news by using a dance called waggle/tremble/vibration dance. The purpose of this dance, gives the idea of the quality and/or quantity of food and also its distance from the beehive. Forager bees then follow the Scout Bees to the location that they found food and then begin to reap it. After that they return to the beehive and do a tremble or vibration dance to other bees in the hive giving an idea of how much food is left. The tasks removed from the overloaded VMs act as Honey Bees. Upon submission to the under load VM, it will update the number of various priority tasks and load of tasks assigned to that VM. This information will be helpful for other tasks , i.e., whenever a high priority has to be submitted to VMs, it should consider the VM that has a minimum number of high priority tasks so that the particular task will be executed earlier. Since all VMs are sorted in an ascending order, the task removed will be submitted to under loaded VMs. Current workload of all available VMs can be calculated based on the information received from the data center. Advantages are maximizing the throughput; waiting time on task is minimum and overhead become minimum. The disadvantage is if more priority based queues are there then the lower priority load can be stay continuously in the queue.

BIASED RANDOM SAMPLING

Biased Random Sampling [2] is a dynamic load balancing algorithm. It uses random sampling of system domain to achieve self-organization thus, balancing the load across all nodes of system. In this algorithm, a virtual graph is constructed with the connectivity of each node representing the load on server. Each node is represented as a vertex in a directed graph and each in-degree represents free resources of that node. Whenever a client sends a request to the load balancer, the load balancer allocates the job to the node which has at least one in-degree. Once a job is allocated to the node, the in-degree of that node is decremented by one. After the job is completed, the node creates an incoming edge and increments the in-degree by one. The addition and deletion of processes is done by the process of random sampling. Each process is characterized by a parameter know as threshold value, which indicates the maximum walk length. A walk is defined as the traversal from one node to another until the destination is found. At each step on the walk, the neighbour node of current node is selected as the next node. In this algorithm, upon receiving the request by the load balancer, it would select a node randomly and compares the current walk length with the threshold value. If the current walk length is equal to or greater than the threshold value, the job is executed at that node. Else, the walk length of the job is incremented and another neighbour node is selected randomly. The performance is degraded as the number of servers increase due to additional overhead for computing the walk length.

HEAT DIFFUSION BASED DYNAMIC LOAD BALANCING

Yunhua.et al.[10] proposed an efficient cell selection scheme and two heat diffusion based algorithm called global and local diffusion. Considered the distributed virtual environments there were various number of users and the load accessing by the concurrent users can cause problem. This can be avoided by this algorithm. According to the heat diffusion algorithm, the virtual environment is divided in to large number of square cells and each square cells having objects. The working of the heat diffusion algorithm is in such a way that every nodes in the cell sends load to its neighbouring nodes in every iteration and the transfer was the difference between the current node to that of neighbouring node. So it was related to heat diffusion process. That is the transfer of heat from high to low object, when they were placed adjacently In local diffusion algorithm, there were local decision making and efficient cell

selection schemes are used. Here they simply compared the neighbouring node loads to the adjacent node loads. If load is small then the transfer of load becomes possible. When global diffusion algorithm considered, it has two stages that is global scheduling stage and local load migration stage. From various experimental results the global diffusion algorithm becomes the better one. Advantages are communication overhead is less, high speed and require little amount of calculations. Disadvantages are network delay is high and several iterations are taken so there was a waste of time.

LOAD BALANCING IN DYNAMIC STRUCTURED P2P SYSTEMS

Brighten et al. [11] proposed an algorithm for load balancing in dynamic peer-to-peer system and other hybrid environments. In most peer-to-peer system the non uniform of objects in the space and also the load of the node can be changed continuously due to the insertion, deletion and other various operations. This will leads to decrease the performance of the system. So the concept of virtual server can be introduced. In this proposed load balancing algorithm, the load information of the peer nodes are stored in different directories. These directories help to schedule reassignment of the virtual servers to develop a better balance. Greedy heuristic algorithm used to find out a better solution for the proper utilization of the nodes. The huge number of virtual servers in the system helps to increase the utilization. The various load information in to the corresponding pool and then the virtual server assignments are to be done. This proposed algorithm should be applied to different types of resources like storage, bandwidth etc, It was designed to handle the various situations like varying load of the node, node capacity, entering and leaving of nodes and also insertion and deletion of the nodes. Advantages are high node utilization and increasing scalability. Disadvantage is the reassignment of the virtual server is difficult, template and simply type your text into it.

TABLE I
COMPARISON TABLE OF LOAD BALANCING TECHNIQUES

<i>Load Balancing Methods</i>	<i>Parameters</i>	<i>Merits</i>	<i>Demerits</i>
Round-Robin Load Balancer	Weight, cycle	Inexpensive and easy setup, simplicity	No support for high availability
Fast Adaptive Load Balancing Method	Efficiency Communication Cost	Faster Balancing Speed High Efficiency Low Communication Overhead	Cannot Maintain the Topology of Cells
Honey Bee Inspired Load Balancing Method	Makespan Task Migration Execution Time	Maximizing the Throughput Waiting Time Minimum Low Overhead	Low Priority Load Become Stay Continuously on the Queue
Biased Random Sampling	Threshold ,	Good Resource utilization, Low overhead	performance is degraded with an increase in population diversity
Heat Diffusion Based	Number of Migrated	Require Very Little Amount	Wastage of Time

Dynamic Load Balancing Method	Users Number of Overload Servers	of Calculation High Speed	Network Delay is High
Load Balancing in Dynamic Structured P2P Systems	Node Utilization Load Movement Factor	Increasing Scalability High Node Utilization	Assignment of Virtual Server is Difficult

CONCLUSIONS :

The load balancing of the current system is one of the greatest issues. 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. Various techniques and algorithms are used to solve the problem. In this paper we survey various existing load balancing methods in different environments. A large number of parameters and different types of soft computing techniques can be included in the future for the better utilization and needs of the user. So there is a need to develop an algorithm which is suitable for heterogeneous partition of public cloud and should also reduce the cost.

REFERENCES :

- [1] Dongliang Zhang, Changjun Jiang, Shu Li, "A fast adaptive load balancing method for parallel particle-based simulations", *Simulation Modelling Practice and Theory* 17 (2009) 1032–1042.
- [2] Shanti Swaroop Moharana¹, Rajadeepan D. Ramesh² & Digamber Powar³ "Analysis Of Load Balancers In Cloud Computing" *International Journal of Computer Science and Engineering (IJCSE)* ISSN 2278-9960 Vol. 2, Issue 2, May 2013, 101-108
- [3] Jaspreet kaur "Comparison of load balancing algorithms in a Cloud" Vol. 2, Issue 3, May-Jun 2012, Gaochao Xu, Junjie Pang, and Xiaodong Fu "A Load Balancing Model Based on Cloud Partitioning for the Public Cloud" ISSN 1007 - 0214 04 /12 pp 34-39 Volume 18, Number 1, February 2013.
- [4] Dhinesh Babu L.D, P. VenkataKrishna, "Honey bee behaviour inspired load balancing of tasks in cloud computing environments", *Applied Soft Computing* 13 (2013) 2292–2303.
- [5] Amandeep Kaur Sidhu¹, Supriya Kinger "Analysis of Load Balancing Techniques in Cloud Computing" Volume 4 No. 2, March-April, 2013, ISSN 2277-3061
- [6] Zhong Xu, Rong Huang, (2009) "Performance Study of Load Balancing Algorithms in Distributed Web Server Systems", *CS213 Parallel and Distributed Processing Project Report*.
- [7] P. Warstein, H. Situ and Z. Huang (2010), "Load balancing in a cluster computer" In proceeding of the seventh International Conference on Parallel and Distributed Computing, Applications and Technologies, IEEE.
- [8] Ms. Nitika, Ms. Shaveta, Mr. Gaurav Raj; "Comparative Analysis of Load Balancing Algorithms in Cloud Computing", *International Journal of Advanced Research in Computer Engineering & Technology* Volume 1,
- [9] Yunhua Deng, Rynson W.H. Lau, "Heat diffusion based dynamic load balancing for distributed virtual environments", *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, ACM, 2010, pp. 203–21.