ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Cloud Partitioning is an Optimal Approach for Public Cloud

S.Vinod Kumar,
M.Tech,
Department of CSE ,
Global Institute of Engineering and Technology,
Chilkur,RRDistrict,Telangana

Mrs. Deeba khan,
Associate Professor,
Department of CSE ,
.Global Institute of Engineering and Technology,
Chilkur,RRDistrict,Telangana

Mrs. M. Jhansi Lakshmi,
Associate professor,
HoD of CSE,
Global Institute of Engineering and Technology,
Chilkur,RRDistrict,Telangana

ABSTRACT

Cloud computing is a provision of providing networked, on-line, on-demand services pay per use basis. Several issues as scalability, security, performance etc are discussed so far by many researchers for the cloud computing.. In public cloud environment various nodes are used with required computing resources situated in different geographic locations, so this strategy simplifies the load distribution across the multiple nodes, but fault tolerance and load balancing are most important problems obtaining high performance in the system. Load balancing is the process of distribution of workload among different nodes or processor. The purpose of load balancing is to improve the performance of a cloud environment through an appropriate distribution strategy. Game theory is the formal study of conflict and cooperation. Game theoretic concepts apply whenever the actions of several agents are interdependent. The game theoretic algorithms help to obtain a user optimal load balancing which ultimately improves overall performance of cloud computing. This paper introduces a better approach for public cloud load distribution using partitioning and game theory concept to increase the performance of the system.

Keywords— Cloud computing, Dynamic Load Balancing (DLB), Game Theory, Public Cloud, Cloud Partitioning.

INTRODUCTION

Cloud Computing is a concept that has many computers interconnected through a real time network like internet. cloud computing means distributed computing. Cloud computing enables convenient, on-demand, dynamic and reliable use of distributed computing resources. The cloud computing model has five main characteristics on demand service, broad network access, resource pooling, flexibility, measured service. Cloud computing is efficient and scalable but to maintain the stability of processing many jobs in the cloud computing is a very difficult problem. The job arrival pattern cannot be predicted and the capacities of each node in the cloud differ. Hence for balancing the usage of internet and related resources has increased widely. Due to this there is tremendous increase in workload. So there is uneven distribution of this workload which results in server overloading and may crash. In such the load, it is crucial to control workloads to improve system performance and maintain stability. The load on every cloud is variable and dependent on various factors. To handle this problem of imbalance of load on clouds and to increase its working efficiency, this paper tries to implement "A Model for load balancing by Partitioning the Public Cloud". Good load balancing makes cloud computing more efficient and also improves user satisfaction [1]. This article is aimed at the public cloud which has numerous nodes. A system having main controller, balancers, servers and a client is implemented here. It introduces a switch mechanism to choose different strategies for different situations. This paper divides the public cloud into cloud partitions and applies different strategies to balance the load on cloud. This paper gives an idea for balancing the load on clouds. It helps to avoid overloading of servers and improve response times. The basic designs of the system and algorithms to implement it are described in this paper [1]

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Goals of Load Balancing

To improve the performance substantially
To have a backup plan in case the system fails even partially.
To maintain the system stability.
To accommodate future modification in the system

RELATED WORK

There have been many studies of load balancing for the cloud environment. Load balancing in cloud computing was described in a white paper written by Adler[3] who introduced the tools and techniques commonly used for IEEE TRANSACTIONS ON CLOUD COMPUTING YEAR 2013 load balancing in the cloud. However, load balancing in cloud is still a new problem that needs new architectures to adapt too many changes. Chaczko et al.[4] described the role that load balancing plays to improve the performance and maintaining stability. There are many load balancing algorithms, such as Round Robin, Game theory Algorithm, and Ant Colony algorithm. Nishant et al.[5] used the ant colony optimization methods in nodes load balancing. Randles et al.[2] gave a compared analysis of some of algorithms in cloud computing by checking the performance time and cost. They concluded that the ESCE algorithm and throttled algorithm are better than Round Robin algorithm. Some of the classical load balancing methods is similar to the allocation method in the operating system, for example, the Round Robin algorithm and the First Come First Served (FCFS) rules. The Round Robin algorithm is used here because it is fairly simple.

LOAD BALANCING

In cloud computing, load balancing is required to distribute the dynamic local workload evenly across all the nodes. It helps to achieve a high user satisfaction and resource utilization ratio by ensuring an efficient and fair allocation of every computing resource. Proper load balancing aids in minimizing resource consumption, implementing fail-over, enabling scalability, avoiding bottlenecks and overprovisioning. There are mainly two types of load balancing algorithms: In static algorithm the traffic is divided evenly among the servers. This algorithm requires a prior knowledge of system resources, so that the decision of shifting of the load does not depend on the current state of system. Static algorithm is proper in the system which has low variation in load. In dynamic algorithm the lightest server in the whole network or system is searched and preferred for balancing a load. For this real time communication with network is needed which can increase the traffic in the system. Here current state of the system is used to make decisions to manage the load. Load balancing based on Cloud Partitioning There are several cloud computing services with this work focused on a public cloud. A public cloud is based on the standard cloud computing model, with service provided by a service provider. A large public cloud will include many nodes and the nodes in different geographical locations. Cloud partitioning is used to manage this large cloud. A cloud partition is a subarea of the public cloud with divisions based on the geographic locations. The architecture is shown in Fig.1. The load balancing strategy is based on the cloud partitioning concept. After creating the cloud partitions, the load balancing then starts, when a job arrives at the system, with the main controller deciding which cloud partition should receive the job. The partition load balancer then decides how to assign the jobs based on load status of nodes. When the load status of a cloud partition is normal, this partitioning can be accomplished locally. If the cloud partition load status is not normal, this job should be transferred to another partition. Here we are going to discuss some load balancing technique for both the partition having either load status=idle or load status=normal based on load degree. The node load degree is based on different static and dynamic parameters of each node.

EXISTING SYSTEM

Since the job arrival pattern is not predictable and the capacities of each node in the cloud differ, for load balancing problem, workload control is crucial to improve system performance and maintain stability. Load balancing schemes depending on whether the

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

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.

DISADVANTAGES:

- Cloud computing environment is a very complex problem with load balancing receiving.
- The job arrival pattern is not predictable and the capacities of each node in the cloud differ, for load balancing problem, workload control is crucial to improve system performance and maintain stability.

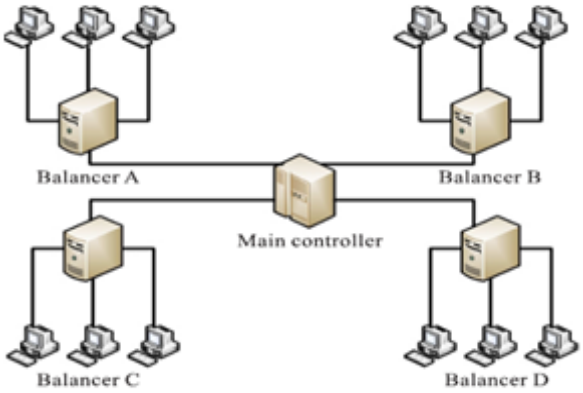
PROPOSED SYSTEM

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.

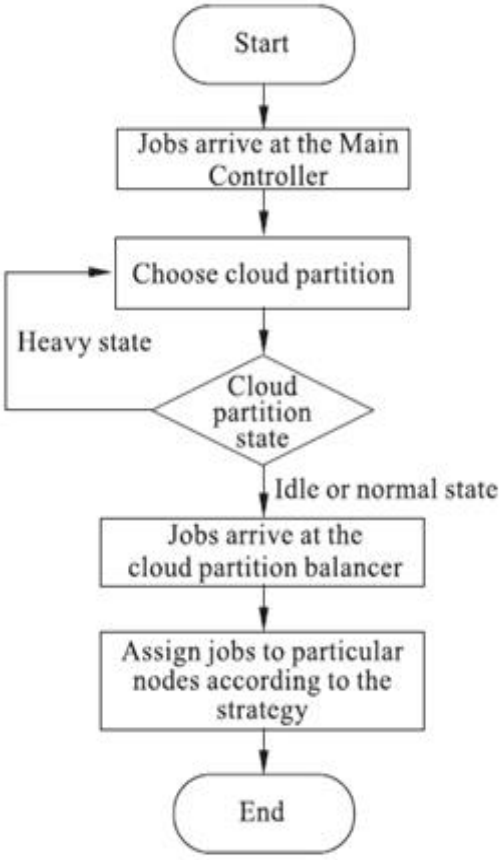
Load balancing schemes depending on whether the system dynamics are important can be either static or dynamic. Static schemes do not use the system information and are less complex while dynamic schemes will bring additional costs for the system but can change as the system status changes. A dynamic 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.

ADVANTAGES:

- 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 role that loads balancing plays in improving the performance and maintaining stability.



PROJECT FLOW:



ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Implementation

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

USER MODULE:

In this module, Users are having authentication and security to access the detail which is presented in the ontology system. Before accessing or searching the details user should have the account in that otherwise they should register first.

System Model:

There are several cloud computing categories with this work focused on a public cloud. A public cloud is based on the standard cloud computing model, with service provided by a service provider . A large public cloud will include many nodes and the nodes in different geographical locations. Cloud partitioning is used to manage this large cloud. A cloud partition is a subarea of the public cloud with divisions based on the geographic locations. with the main controller deciding which cloud partition should receive the job. The partition load balancer then decides how to assign the jobs to the nodes. When the load status of a cloud partition is normal, this partitioning can be accomplished locally. If the cloud partition load status is not normal, this job should be transferred to another partition.

Main controller and balancers:

The load balance solution is done by the main controller and the balancers.

The main controller first assigns jobs to the suitable cloud partition and then communicates with the balancers in each partition to refresh this status information. Since the main controller deals with information for each partition, smaller data sets will lead to the higher processing rates. The balancers in each partition gather the status information from every node and then choose the right strategy to distribute the jobs.

Cloud Partition Load Balancing Strategy:

When the cloud partition is idle, many computing resources are available and relatively few jobs are arriving. In this situation, this cloud partition has the ability to process jobs as quickly as possible so a simple load balancing method can be used. There are many simple load balance algorithm methods such as the Random algorithm, the Weight Round Robin, and the Dynamic Round Robin .The Round Robin algorithm is used here for its simplicity.

CONCLUSIONS AND FUTURE WORKS

A public cloud is one based on the standard cloud computing model, in which a service provider makes resources, such as applications and storage, available to the general public over the Internet [6]. Public cloud is made up of several nodes situated in deferent geographic location. Cloud partitioning is a method to make partitions of huge public cloud is some segment of cloud. A public cloud is one based on the standard cloud computing model, in which a service provider makes resources, such as applications and storage, available to the general public over the Internet. Public cloud is made up of several nodes situated in deferent geographic location. Cloud partitioning is a method to make partitions of huge public cloud is some segment of cloud. The object of study in game theory is the game, which is a formal model of an interactive situation. It typically involves several players; a game with only one player is usually called a decision problem. In future study we will try to find other load balance strategy because other load balance strategies may provide better results, so tests are needed to compare different strategies. Many tests are needed to guarantee system availability and efficiency. Also we

will address the development of game theoretic models for load balancing in the context of uncertainty as well as game theoretic models for dynamic load balancing in future. We also plan to develop dynamic load balancing schemes based on dynamic game theory that provide fairness by taking the current system load into account and also consider other aspects of heterogeneity.

References

[1] R. Hunter, The why of cloud, http://www.gartner.com/ DisplayDocument?doc cd=226469&ref= g noreg, 2012.

[2] M. D. Dikaiakos, D. Katsaros, P. Mehra, G. Pallis, and A. Vakali, Cloud computing: Distributed internet computing for IT and scientific research, Internet Computing, vol.13, no.5, pp.10-13, Sept.-Oct. 2009.

[3] P. Mell and T. Grance, The NIST definition of cloud computing, http://csrc.nist.gov/ publications/nistpubs/ 800- 145/SP800-145.pdf, 2012.

[4] Microsoft Academic Research, Cloud computing, http://libra.msra.cn/Keyword/6051/cloud.computing?q uery= cloud%20computing, 2012.

[5] Google Trends, Cloud computing, http://www.google.com/trends/explore#q=cloud%20co mputing, 2012.

[6] N. G. Shivaratri, P. Krueger, and M. Singhal, Load distributing for locally distributed systems, Computer, vol. 25, no. 12, pp. 33-44, Dec. 1992.

[7] B. Adler, Load balancing in the cloud: Tools, tips and techniques, http://www.rightscale. com/info center/whitepapers/ Load-Balancing-in-the-Cloud.pdf, 2012

[8] Z. Chaczko, V. Mahadevan, S. Aslanzadeh, and C. Mcdermid, Availability and load balancing in cloud computing, presented at the 2011 International Conference on Computer and Software Modeling, Singapore, 2011.

[9] K. Nishant, P. Sharma, V. Krishna, C. Gupta, K. P. Singh, N. Nitin, and R. Rastogi, Load balancing of nodes in cloud using ant colony optimization, in Proc. 14th International Conference on Computer Modelling and Simulation (UKSim), Cambridgeshire, United Kingdom, Mar. 2012, pp. 28-30.

[10] M. Randles, D. Lamb, and A. Taleb-Bendiab, A comparative study into distributed load balancing algorithms for cloud computing, in Proc. IEEE 24th International Conference on Advanced Information Networking and Applications, Perth, Australia, 2010, pp. 551-556.

[11] A. Rouse, Public cloud, http://searchcloudcomputing.techtarget.com/definition/ public-cloud, 2012.

[12] D. MacVittie, Intro to load balancing for developers — The algorithms, https://devcentral.f5.com/blogs/us/introto- load-balancing-for-developers-ndash-the-algorithms, 2012.

[13] S. Penmatsa and A. T. Chronopoulos, Game-theoretic static load balancing for distributed systems, Journalof Parallel and Distributed Computing, vol. 71, no. 4, pp. 537-555, Apr. 2011.

[14] D. Grosu, A. T. Chronopoulos, and M. Y. Leung, Load balancing in distributed systems: An approach usingcooperative games, in Proc. 16th IEEE Intl. Parallel and Distributed Processing Symp., Florida, USA, Apr. 2002,pp. 52-61.

[15] S. Aote and M. U. Kharat, A game-theoretic model for dynamic load balancing in distributed systems, in Proc. The International Conference on Advances in Computing, Communication and Control (ICAC3 '09), New York, USA, 2009, pp. 235-238.