A Load Balancing Model Based On Cloud Partitioning For The Public Cloud

Swathi.m[1], Sowmya martha.m[2], Sowmya.k[3], Shanthamma.h.d[4], Basavaraj Jakkali[5]
[1]Student ,[2]student, [3]student,[4]student,[5]Associate Professor, CSE, BMSCE,Bangalore
Computer science and engineering
BMS College of Engineering,Bangalore,India

Abstract : **Cloud computing is one of the new style of computing.Network applications are growing very fast and the use of cloud computing has increased rapidly. Some of the advantages of cloud computing are low cost, on demand services pay per use basis,on-line and accessibility of data.Along with the advantages there are someissues to be resolved in order to improve reliability of cloud environment. Load Balancing is the main concern in cloud computing. In cloud computing load balancing has an important impact on the performance. User satisfaction and cloud computing efficiency can be improved by good load balancing.Some of the loadsaredelay or network load, CPU load, memory capacity. 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.Various load balancing techniques have been proposed by many researches.**

INTRODUCTION

Public cloud makes services such as computing, storage application which are available tothe public. Some of the servicesprovided by the public cloud are free and some are offered as payment as per usage. The major public cloudproviders are Amazon, Google, Microsoft, etc.Cloud system contains three major components they are clients, data center and distributed servers. Each element in cloud system has a definite purpose and plays a specific role. Load balancing is a process of reassigning the total load tothe individual nodes of the collective system.Load balancing makes resource utilizationeffective and improve the response time of the job.It also helps in removing a condition inwhich some of the nodes are overloaded while some othersare underloaded.
A load balancing algorithm used forbalancing purposes which is dynamic in nature does notconsider the previous state or behavior of the system, it depends on the current behavior of the system.It is also a relatively new technique thatfacilitates networking and resources by providing a maximumthroughput with minimum response time. Proper loadbalancing helps in utilizing the available resourcesoptimally. The load balancers in load balancing plays an important role, where each incoming request is redirected and it will betransparent to the client who makes the request.Loadbalancers are also available with variety of special features. Load balancing schemes depending onwhether the system dynamics are important can beeither static and dynamic. Static schemes are less complex and they do not use the system information where the dynamic schemes will bring additional costs for thesystem but can change as the system status changes.Dynamic schemes are more flexible than static schemes. The model consists main controller and balancers to gather andanalyze the information. The systemstatus provides a basis for choosing the right loadbalancing strategy.

RELATED WORK

In this section, we present literatures of several highly related research areas to load balancing model based on cloud partitioning for the public cloud. Load balancing in cloud computing was describe by Adler who introduced the tools and techniques commonly used for load balancing in the cloud. Load balancing in the cloud is still a new problem that needs new architectures to adapt to many changes.Chaczkoetal described the role that load balancing plays an important role in improving the performance and maintaining stability. Some of the load balancing algorithms are Round Robin, Equally Spread Current Execution Algorithm and Ant Colony algorithm.These algorithms applies the game theory to the load balancing strategy to improve the efficiency in the public cloud environment.Some of the classical loadbalancing methods are similar to the allocation methodin the operating system.Some of the examples are Round Robin algorithm and the First Come First Served (FCFS) rules. The Round Robin algorithm is used here because it is fairly simple. Many researchers have invented different techniques for load balancing and distributing the workload in a cloud environment. Many tools has been developed and some of them are c-meter, Amizone EC-2,GrenchMarketc.V. Mahadevan, S. Aslanzadeh andC. Mcdermid in their paper “Availability and load balancing in cloud computing 2011” described the role that

load partitioning can be accomplished . If the cloud partition load status is not normal then this job should be transferred to another partition.

SYSTEM MODEL

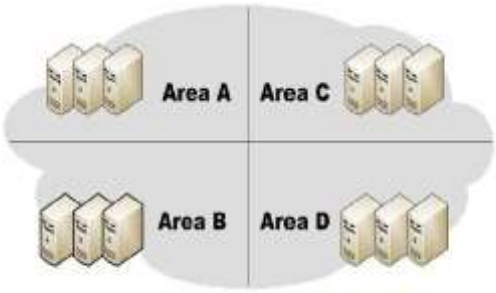
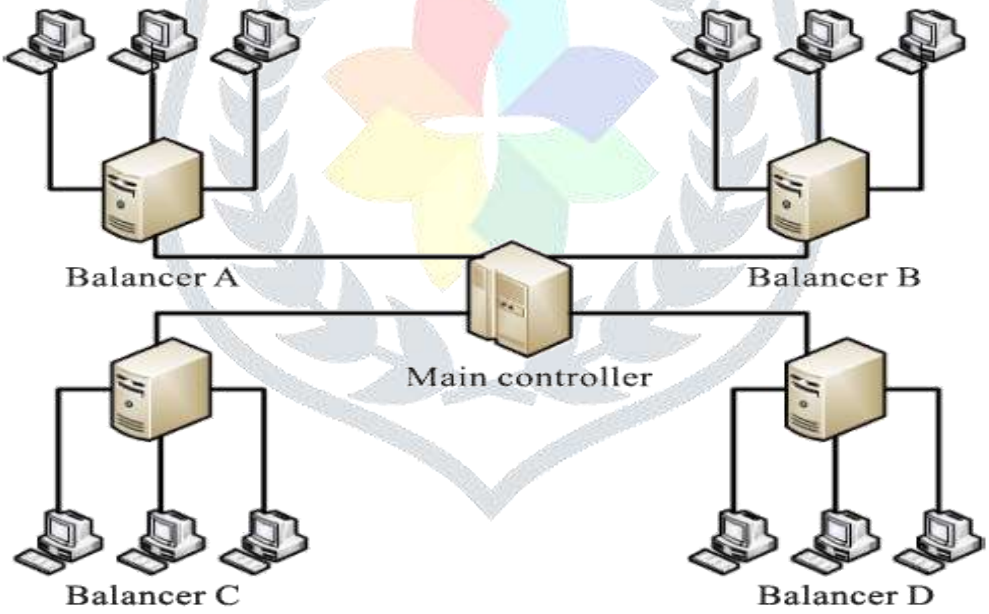


Fig. 1 Typical cloud partitioning.

In this section we are describing how the model works. There are several cloud computing categories with this work focused on a public cloud.It is based on the standard cloud computing model, and the service is provided by a service provider. A large public cloud consists of many nodes and the nodes will be in different geographical locations. This large cloud is managed by the cloud partitioning. Subarea of the public cloud with divisions based on the geographic locations is a cloud patition.The load balancing strategy is based on the cloud partitioning concept. Each partitions have a balancers that contain the status of particular partition. After creating the cloud partitions the load balancing then begins when a job arrives at the system with the main controller deciding which cloud partition should receive the job. The partitioned load balancer then helps in deciding how to assign the job to the each node. When the load status of a cloud partition is normal this partitioning can be accomplished locally. And if this cloud



parttionlaod status is not normal then this job should be transferred to one of the another partition. The main components of system model are main controller and balancers. The architecture is shown below.

MAIN CONTROLLER AND BALLANCER

The main controller contains the status of all the partition's balancer with in the cloud The main job of the main controller is when the job arrives, it will first check the status of the balancers. The status of the balancers are idle,normal and heavily loaded.The main controller checks the status of balancer whether it is idle or heavily loaded.If it is heavily loaded it checks the next balancer status to assign the job,or else it assigns the job to that particular balancer. The main controller has to communicate with the balancers frequently to refresh the status information. The cloud partition balancer gathers load information from every node to evaluate the

cloud partition status. The balancer holds the status of the each and every node present in that particular partition.So when the job arrives from the main controller it checks the status of nodes one by one. If the node is idle it assigns the job to that particular node or else it check the status of the next node in the same partition to assign the job. Hence the load is balanced between the nodes, thereby resource utilization is increased.The above figure depicts the relation between the main controller and the balancer.

PROPOSED WORK

The proposed work is divided into two modules.They are

1.User module

2.System module

[1]User module:

In User module each user should have an account in order to access or search the details. Otherwise they cannot access the details of the cloud. Each user is provided with the id and password for authentication process to access the cloud. If the user is accessing for the first time user has to get register first by filling the registration form.

[2]SYSTEM MODULE

The System module is focused on public cloud which is based on standard cloud computing model. The system model is based on the concept that a large public cloud will include many nodes and the nodes in different geographical locations.Cloud partitioning will be used here to manage this large cloud with the main controller deciding which cloudpartition should receive the job. Moreover as the number of users grows on in the cloud, thesystem model will provide ways to divide the public cloudinto more locations as clouds servers. System model will provide a way to add new locations, new servers and monitor all the servers on all locations for current connections and current status.

Strategy of Load Balancing:-

When the cloud partition is idle few jobs are arriving and thus the cloud partition has the ability to process jobs as quickly as possible so a simple load balancing method such as "The Round Robin algorithm based on the load degree evaluation" will be used here for its simplicity. When the cloud partition is normal and job arrival pattern is much faster than in the idle state and the situation is far more complex. Hence we make use of the different strategy in load balancing as each user wants his jobs completed in the shortest time. The current model uses the game theory approach for non-cooperative games proposed by Grosu called "the best reply". Non-cooperative games and cooperative games are present in game theory. The decision makers come to an agreement which is called a binding agreement in cooperative games. Each decision maker decides by comparing nodes with each others. Each decision maker makes decisions only for his own benefit in non-cooperative games.

CONCLUSION AND FUTURE WORK:

In this paper we presented the Load balancing model based on partitioning the public cloud,this enhances the performance in the cloud computing environment.Cloud Computing provides user with following services iaas,paas.One of the major issues in cloud computing is load balancing. Load balancing distributes the load evenly among all servers in the cloud to maximize the resource utilization, increases throughput, to provide good response time, to reduce energy consumption. To increase the performance and user satisfaction this model is using a main controller and balancer. Main controller and balancerdistributes the load evenly. Good load balancing will leads to efficient cloud computing and higher user satisfaction. The paper proposed the new architecture for load balancing as well as performance enhancement in public cloud. The concept of load balancing is implemented on the basis of cloud partitioning method with different strategies for different cloud status. Ourmain aim is to achieve the results and compare those with the existing system. Significant level of development is running in order to satisfy the required objectives.

REFERENCES

[1] GaochaoXu, Junjie Pang, and Xiaodong Fu, *A Load Balancing Model Based on Cloud Partitioning for the Public Cloud*, IEEE TRANSACTIONS ON CLOUD COMPUTING YEAR 2013.

[2] R. Hunter, *The why of cloud*, http://www.gartner.com/DisplayDocument?doccd=2264 69&ref= g noreg, 2012.

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

[4] P.Mell and T.Grance,*The NIST definition ofcloudcomputing*, http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf, 2012

[5] Wayne Jansen, Timothy Grance, "Guidelines on Security and Privacy in Public Cloud Computing", National Institute of Standards and Technology Gaithersburg, January 2011.
[6] Jeep Ruiter, MartijnWarnier, "Privacy Regulations for Cloud Computing", Faculty of Sciences, VU University Amsterdam International Journal of Web & Semantic Technology (IJWesT) Vol.3, No.2, April 2012 50
[7] DanchoDanchev,"Building and Implementing a successful Information Security Policy" window security.com- Windows Security Resources for IT admins.
[8] David Escalante and Andrew J. Korty, Cloud Services: Policy and Assessment, *EDUCAUSE Review,* vol. 46, no. 4 (July/August 2011)
[9] Richard N. Katz, "Looking at Clouds from All Sides Now", *EDUCAUSE Review,* vol. 45, no. 3 (May/June 2010): 32-45
[10**]** Anthony T.Velte, Toby J.Velte, Robert Elsenpeter, Cloud Computing A Practical Approach, TATA McGRAW-HILL Edition 2010.
[11**]** 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.
[12] A. AlZain et al., "Cloud Computing Security: From Single to Multi-Clouds" 2012 45th Hawaii International Conference on System Sciences.
[13] OlfaNasraoui et al., "Ensuring Data Integrity and Security In Cloud Storage", IEEE, VOL. 20, No. 2,February 2013.
[14] Qin Liu et al., "Reliable Re-encryption in Unreliable Clouds" IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2011 proceedings
[15] Google Trends, Cloud computing, http://www.google.com/trends/explore#q=cloud%20computing, 2012.
[16] B. Adler, Load balancing in the cloud: Tools, tips and techniques, http://www.rightscale. com/infocenter/whitepapers/ Load-Balancing-in-the-Cloud.pdf, 2012
[17] Z. Chaczko, V. Mahadevan, S. Aslanzadeh, and C. Mcdermid, Availability and load balancing in cloud computing, presented at the 2011 International Conferenceon Computer and Software Modeling, Singapore, 2011.