ISSN: 2319-6505

Available Online at http://journalijcar.org

International Journal of Current Advanced Research
Vol 5, Issue 3, pp 639-642, March 2016

International Journal of Current Advanced Research

ISSN: 2319 - 6475

RESEARCH ARTICLE

LOAD SHARING TECHNIQUES AND CHALLENGES IN CLOUD COMPUTING

Vanithasree S[1] and Anjugam P[2]

[1,2] K.M.G College of arts and Science, Gudiyatham, 635803, Tamilnadu, India

A R T I C L E I N F O

Article History:

Received 16th December, 2015
Received in revised form 24th January, 2016
Accepted 23rd February, 2016
Published online 28th March, 2016

Key words:

Cloud Computing, Cloud Service Model, Load Balancing, Round Robin, Throttled, Equal Spread Current Execution.

A B S T R A C T

In present paper, discussed with the cloud computing requirements for access control, migration, security, data availability, trust issues and sensitive information. Cloud computing is emerging technology which is a new standard of large scale distributed computing and parallel computing. It provides shared resources, information, software packages and other resources as per client requirements at specific time. As cloud computing is growing rapidly and more users are attracted towards utility computing, better and fast service needs to be provided. One of the challenging scheduling problems in Cloud datacenters is to take the allocation and migration of reconfigurable virtual machines into consideration as well as the integrated features of hosting physical machines.

In order to select the virtual nodes for executing the task, Load balancing is a methodology to distribute workload across multiple computers, or other resources over the network links to achieve optimal resource utilization, minimum data processing time, minimum average response time, and avoid overload. The objective of this paper to propose efficient and enhanced scheduling algorithm that can maintain the load balancing and provides better improved strategies through efficient job scheduling and modified resource allocation techniques. Load balancing ensures that all the processors in the system as well as in the network do approximately the equal amount of work at any instant of time. The results discussed in this paper, based on existing Equally Spread Current Execution, Round Robin, Throttled and a new proposed enhanced and efficient scheduling algorithms.

INTRODUCTION

Cloud Computing refers to the use and access of multiple server based computational resources via a digital network(WAN).Cloud users may access the resources using computer note book, pad computer, smart phone, or other device. In cloud computing applications are provided and managed by the cloud server and data is also stored remotely in cloud configuration. As Cloud Computing is growing rapidly and clients are demanding more services and better results, load balancing for the Cloud has become a very interesting and important research area. Load balancing ensures that all the processor in the system or every node in the network does approximately the equal amount of work at any instant of time. our thesis discussed many different load balancing techniques used to solve the issue in cloud computing environment

A Cloud Computing Architecture

Cloud computing architecture refers to the components required for cloud computing. Cloud computing typically involves multiple cloud components communicating with each other over a loose coupling mechanism such as messaging queue. Cloud computing architecture can be divided into two sections: front end and back end. They both are connected with each other through a network, usually the internet.

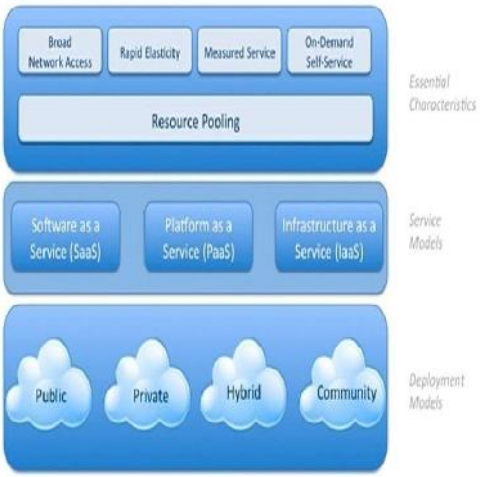


Figure 1 Cloud computing architecture

The front end is what the user (client) sees whereas the back end is the cloud of the system. Front end has client's computer and the application required to access the cloud and the back end has the cloud computing services like various computers, servers and data storage.

B types of cloud

On the basis of accessibility, clouds can be deployed using any of the following strategies:

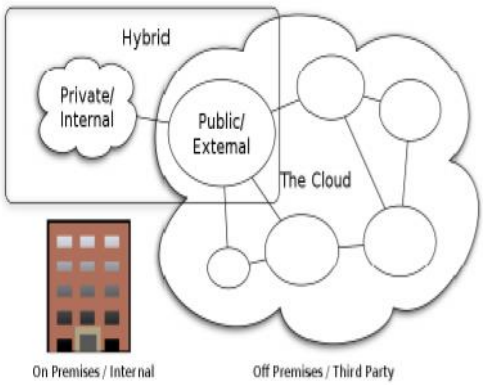


Figure 2 Load balancing in cloud computing

Public Cloud

Public clouds are made available to the general public by a service provider who hosts the cloud infrastructure. Generally, public cloud providers like Amazon AWS, Microsoft and Google own and operate the infrastructure and offer access over the Internet. Users need to pay only for the time duration they use the service i.e. pay-per-use. A public cloud does not mean that a user's data is publically visible; public cloud vendors typically provide an access control mechanism for their users. Clients do not need to purchase hardware to get service and can also scale their user on demand. Public clouds provide an elastic, cost effective means to deploy solutions.

Private Cloud

Private cloud means using a cloud infrastructure solely by one customer or organization. In a private cloud-based service, resources are deployed inside a firewall and managed by client's organization without the restriction of network bandwidth and security exposures. Organization owns the hardware and software infrastructure, manages cloud and controls access to its resources. The main advantage of private cloud is, it is easier to manage security, maintenance and upgrades and also provides more control over the deployment and use. As compared to public clouds, where all resources and applications were managed by the service provider, in private cloud these services are pooled together and made available for users at the organizational level.

Community Cloud

A community cloud is a cloud service model that provides a cloud computing solution to a limited number of individuals or organizations that is governed, managed and secured commonly by all the participating organizations or a third party managed service provider. It is multi-tenant infrastructure that is controlled and used by a group of organizations that have shared interests such as specific security requirements. Community clouds are designed for business and organizations working on joint projects, applications or research which requires a central cloud computing facility for building, managing and executing such projects.

Hybrid Cloud

A hybrid cloud is a cloud computing environment in which an organization provides and manages some in-house resources and has others provided externally. Hybrid cloud is a combination of private and public cloud. In this, a private cloud is linked to one or more external cloud services. The goal is to combine services and data from a variety of cloud models to create a unified automated and well-managed computing environment

LOAD BALANCING

Load balancing is one of the main issues related to cloud computing. The load can be a memory, CPU capacity, network or delay load. It is always required to share work load among the various nodes of the distributed system to improve the resource utilization and for better performance of the system. This can help to avoid the situation where nodes are either heavily loaded or under loaded in the network.

The goals of load balancing [16] are to

- Improve the performance
- Maintain system stability
- Build fault tolerance system
- Accommodate future modification.

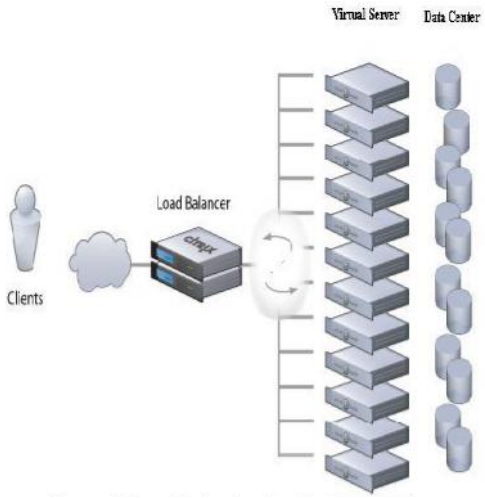


Figure 2 Load balancing in cloud computing

Types of Load Balancing Algorithms

The load balancing algorithms are classified into two categories: static and dynamic load balancing

- ***Sender Initiated****:* If the load balancing algorithm is initialized by the sender
- ***Receiver Initiated****:* If the load balancing algorithm is initiated by the receiver
- ***Symmetric:*** It is the combination of both sender initiated and receiver initiated Depending on the current state of the system, load balancing algorithms can be divided into 2 categories
- ***Static****:* It doesn't depend on the current state of the system. Prior knowledge of the system is needed
- ***Dynamic****:* Decisions on load balancing are based on current state of the system. No prior knowledge is needed. So it is better than static approach. Here we

will discuss on various dynamic load balancing algorithms for the clouds of different sizes.

EXISTING LOAD BALANCING ALGORITHMS

A round robin algorithm (rr)

It is the simplest algorithm that uses the concept of time quantum or slices Here the time is divided into multiple slices and each node is given a particular time quantum or time interval and in this quantum the node will perform its operations. The resources of the service provider are provided to the client on the basis of this time quantum. In Round Robin Scheduling the time quantum play a very important role for scheduling, because if time quantum is very large then Round Robin Scheduling Algorithm is same as the FCFS Scheduling. If the time quantum is extremely too small then Round Robin Scheduling is called as Processor Sharing Algorithm and number of context switches is very high. It selects the load on random basis and leads to the situation where some nodes are heavily loaded and some are lightly loaded. Though the algorithm is very simple but there is an additional load on the scheduler to decide the size of quantum and it has longer average waiting time, higher context switches higher turnaround time and low throughput.

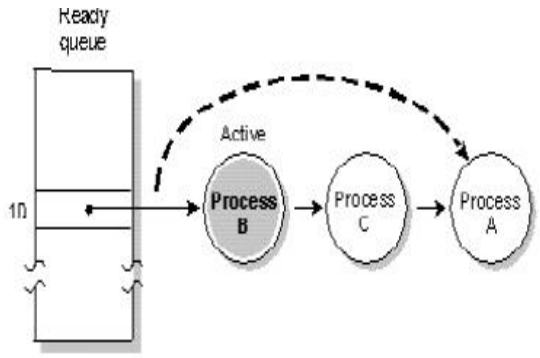


Figure 3 Round Robin Algorithm

PROPOSED WORK

A modified throttled load balancing algorithm

Giving some modification to this available algorithm we can give flexibility to the user for getting services from the cloud provider. As we know that while providing resources as services it is possible to have a number of request at a same time and due to that some requestor need to remain in the queue though they have possibility to send request to other service provider.

Thus to overcome such situations we can modify some existing algorithm and make it available to user to decide whether they need to be in queue or can have a service from the other cloud provider. We have decided to do the modification to the throttled load balancer as it is founded by some experts of cloud computing that throttled load balancer is much better as compare to the other available load balancing algorithms.

The Throttled Load balancing algorithm is divided into three parts. The first phase is the initialization phase. In the first phase, the expected response time of each VM is to be found. In second Phase find the efficient VM. Last Phase return the ID of efficient VM.

B Efficient and Enhanced Algorithm (eea)

Algorithm

- ***Step1***. Initially VM index table will be 0 as all the VMs are in available state.
- ***Step2***. DataCenter Controller receives a new request.
- ***Step3***. DataCenter Controller queries new Load Balancer for next allocation
- ***Step4***. Datacenter Controller parses the VM list to get next available VM: If found: Load Balancer returns the VM id to Datacenter Controller Step2 continues If not found Using round robin fashion VM index is reinitialized to 0 and in increment manner VMs are checked to find VM in available state
- ***Step5***. When the VM finishes the processing the request, and the Datacenter Controller receives the cloudlet response, it notices the load balancer of the VM de-allocation
- ***Step6***. The Load Balancer updates the status of VM in allocation table to available.
- ***Step7***. Continue from Stp2 The purpose of the algorithm is to find the expected Response Time of each Virtual Machine, which is calculated as:

Response Time = Fint - Ant + TDelay (1)

Where, Ant is the arrival time of user request and Find is the finish time of user request and the transmission delay can be determined using the following formulas

Delay = Latency + Ttransfer ….(2)

Where, TDelay is the transmission delay T latency is the network latency and T transfer is the time taken to transfer the size of data of a single request (D) from source location to destination.

Ttransfer = D / Bwperuser (3)

Bwperuser = Bwtotal / Nr (4)

Where, Bwtotal is the total available bandwidth and Nr is the number of user requests currently in transmission. The Internet Characteristics also keeps track of the number of user requests in between two regions for the value of Nr.

C Challenges for Load Balancing

There are some qualitative metrics that can be improved for better load balancing in cloud computing.

Throughput: It is the total number of tasks that have completed execution for a given scale of time. It is required to have high through put for better performance of the system.

Associated Overhead: It describes the amount of overhead during the implementation of the load balancing algorithm. It is a composition of movement of tasks, inter process communication and inter processor. For load balancing

technique to work properly, minimum overhead should be there.

Fault tolerant: We can define it as the ability to perform load balancing by the appropriate algorithm without arbitrary link or node failure. Every load balancing algorithm should have good fault tolerance approach.

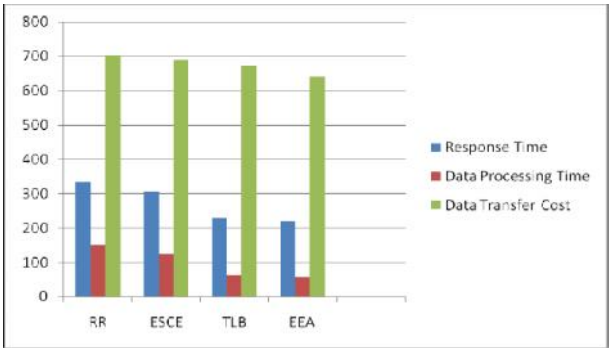
Migration time: It is the amount of time for a process to be transferred from one system node to another node for execution. For better performance of the system this time should be always less.

Response time: In Distributed system, it is the time taken by a particular load balancing technique to respond. This time should be minimized for better performance.

Resource Utilization: It is the parameter which gives the information within which extant the resource is utilized. For efficient load balancing in system, optimum resource should be utilized.

Scalability: It is the ability of load balancing algorithm for a system with any finite number of processor and machines. This parameter can be improved for better system performance.

Performance: It is the overall efficiency of the system. If all the parameters are improved then the overall system performance can be improved.



CONCLUSION AND FUTURE WORK

In present paper, discussed with the cloud computing requirements main purpose of load balancing is to satisfy the customer requirement by distributing load dynamically among the nodes and to make maximum resource utilization by reassigning the total load to individual node. This ensures that every resource is distributed efficiently and evenly.

So the performance of the system is increased. We have also discussed virtualization of cloud and required qualitative matrix for load balancing. The future work includes overcoming the problem of deadlocks and server overflow. We can also implement a new service broken policy in the simulator

References

1. GautamShroff, “Enterprise Cloud Computing,” Cambridge University Press, June, 2011.
2. Qi Zhang, Lu Cheng, RaoufBoutaba, “Cloud Computing: State of Art and Research Challenges,” Springer, April 2010.
3. http://thecloudtutorial.com/related.html.
4. GolamMoktaderNayeem, Mohammad Jahangir Alam,”Analysis of Different Software Fault Tolerance Techniques”, 2006.
5. L. M. Vaquero, L. Rodero-Merino, J. Caceres and M.Lindner, “A break in the clouds: towards a cloud definition,” SIGCOMM Computer Communication Review, vol. 39, pp. 50–55, December 2008.
6. GeoffroyVallee, Kulathep Charoenpornwattana, Christian Engelmann, AnandTikotekar, Step hen L.Scott,” A Framework for Proactive Fault Tolerance”, Feburary, 2008.
7. ElvinSindrilaru, Alexandru Costan,, Valentin Cristea,”Fault Tolerance and Recov ery in Grid Workflow M anagement Sy stems”, 2010 International Conference on Complex, Intelligent and Software Intensive Systems.
8. M.Armbrust, A.Fox, R. Griffit,et al., “A view of cloud computing”, Communications of the ACM , vol. 53, no.4, pp . 50– 58, 2010.
9. Hong Luo, Dejun Mu, Zhiqun Deng, et. Research of Task Scheduling in Grid Computing. Application Research of Computers. 2005; (5): 16-19.
10. Basic concept and terminology of cloud computing- http://whatiscloud.com
11. L. Wang, J. Tao, M. Kunze,”Scientific Cloud Computing: Early Definition and Experience”, the 10th IEEE International Conference Computing and Communications 2008.
12. Load Balancing in Cloud computing, http://community.citrix.com/display/cdn/Load+Balancing

****