IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

An Enhanced Load Balancing Mechanism in Cloud computing

Sahil Kumar[1], Pawan Luthra[2], Bohar Singh[3]

Research Scholar, Computer Science Engg, Shaheed Bhagat Singh State Technical Campus, Ferozepur [1]

Assistant Professor, Computer Science Engg, Shaheed Bhagat Singh State Technical Campus, Ferozepur [2,3]

Abstract: Cloud computing is the means of accessing a shared pool of configurable computing resources (including hardware, software, networks, servers, storage applications and services) that can be rapidly provided, used and released with minimal effort on the part of users or service providers. But it has some of the main concerns like load management and fault tolerance. In this paper we have put forward a load balancing approach in cloud computing. Load balancing is helped to distribute the workload across multiple nodes to ensure that no single node is overloaded. It helps in proper utilization of resources .It also improves the performance of the system.

Keywords: Cloudlets; Clusters; Data Center Broker; Data Center; Virtual Machine.

I. INTRODUCTION

Cloud is a computing framework which usually denotes storing and accessing data and programs over the internet, instead of your computer's hard drive. Cloud Computing is handled by means of the great prospective paradigm utilized for placement of applications taking place over Internet. This perception also elucidates about the applications which are widen on the way to be manageable over and done with the Internet. Cloud applications utilize huge information centers as well as operational servers which are utilized to host net applications plus services [1]. Cloud computing is a service that is distributed over the internet for data access, computing and cloud storage by creating scalability, elasticity and low cost. Second invention platform for division which suggests [1] a variety of services and applications to the user actually attain them. A cloud computing is one of the rising information technologies used in computation now days. It is green technologies which agree to accessing, computing and storing the assets by offering a variety of services. Cloud computing normally includes models like Infrastructure-as-a-service [1], Platform-as-a-service and Platform-as-a-service. To reduce the computation time and to conquer the storage space issues, most of the organization now a day's make regular use of cloud computing from the established process of calculation. It mainly focuses on allocating data and computations over a scalable information centres of network.

Cloud computing is actually a type of computing that depends on distribution of the computing resources rather than local servers and applications. Basically in cloud computing, the word cloud used as a symbol for 'internet', so cloud computing typically means 'a form of computing based on Internet' [2]. The web device storageand applications are delivering servicesto computers. Example: where the larger collections of the system are connected in private networks/public networks, the environment for the application, data and file storage are energetically scalable which leads to cost of estimation, application hosting, substance storage and delivery.

II.CLOUD MODELS AND ITS CHARACTERISTICS

Basically cloud models are of two types: Deployment model and Service model Cloud computing.

A. Deployment Model

Cloud deployment models refer to a particular type of cloudy environment that is primarily separated by size and access. Deployment models are of three types:

i. Public cloud computing: It mainly depends on third individual to suggest services by paying them on regular basis according to the procedure. Public Cloud environment is made available to all unrestricted consumers who can subscribe the needed services [3].

ii.Private Cloud Computing: The organization itself regulates the services. Usually administrations go for private cloud only in the case of involvement of sensible information. Scaling can be done very professionally by adding hardware and thus the environment can be expanded. The security will be more due to the control of internal structure contained in it and therefore data will be secured.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

iii.Hybrid Cloud Computing: It is the mixture of both public & private cloud computing. A less sensible data will be stored in public and all others in Private Cloud.

B. Service Models

i.Platform-as-a-service: The platform as services refers to the division on pay to alter the direct readying of application level assets or internet application, software deployment framework and runtime atmosphere. It is a platform wherever package will be developed, deployed and tested. They are resources of entire life cycle where software can be operated on a PAAS. For example: Microsoft Azure, Google app engine, Amazon EC2, IBM Smart cloud etc. In other words, the platform as a service helps the customer to create their own applications. The cloud application supports a set of applications program interface. It acts as an interface between application and hardware.

ii.Infrastructure as a service: The infrastructure as a service is also called hardware as service. They deliver computing capabilities as consistent services and basic storage over the network. Pooled and made available to holder workloads are services, storage schemes, networking tools, data centre space etc. Example: storage services provided by Amazon S3, Amazon EBS etc., where users can consume the property and virtual desktop like network, virtualized services, routers and storage etc. are main examples of this model.

iii.Software-as-a-Service:The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

CLOUD CHARACTERISITCS:

Cloud computing possess the following key characteristics [4]

1. On-demand self-service: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.

2. Broad network access: Cloud computing provide the users with various capabilities over the network which are accessed through standard mechanisms that promote their use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops etc.)

3. Resource pooling: The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. Examples of resources include storage, processing, memory, network bandwidth, and virtual machines

. 4. Rapid elasticity: Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out, and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.

5. Measured Service: Cloud systems automatically control and optimize resource use by leveraging a metering capability1 at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

III.LOAD BALANCING

Load balancing is a computer network method for distributing workloads across multiple computing resources, so that time efficiency can be increased and also proper utilization of resources takes place while also avoiding a situation where some of the nodes are heavily loaded while other nodes are idle or doing very little work. Load balancing techniques in clouds, consider various parameters such as performance, response time, scalability, throughput, resource utilization, fault tolerance, and associated overhead.

Goals of load balancing

- Goals of load balancing as discussed by authors of [6],[7] include:
- Substantial improvement in performance
- Stability maintenance of the system
- Increase flexibility of the system so as to adapt to the modifications.
- Build a fault tolerant system by creating backups.

Classification of Load Balancing Algorithm

Based on process orientation they are classified as:

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

a) Sender Initiated: In this sender initiates the process; the client sends request until a receiver is assigned to him to receive his workload

b) Receiver Initiated: The receiver initiates the process; the receiver sends a request to acknowledge a sender who is ready to share the workload

c) Symmetric: It is a combination of both sender and receiver initiated type of load balancing algorithm. Based on the current state of the system they are classified as:
1. Static Load Balancing: In the static load balancing algorithm the decision of shifting the load does not depend on the current state of the system. It requires knowledge about the applications and resources of the system. The performance of the virtual machines is determined at the time of job arrival. The master processor assigns the workload to other slave processors according to their performance. The assigned work is thus performed by the slave processors and the result is returned to the master processor. Static load balancing algorithms are not pre-emptive and therefore each machine has at least one task assigned for itself. Its aims in minimizing the execution time of the task and limit communication overhead and delays. This algorithm has a drawback that the task is assigned to the processors or machines only after it is created and that task cannot be shifted during its execution to any other machine for balancing the load.

2. Dynamic Load Balancing: In this type of load balancing algorithms the current state of the system is used to make any decision for load balancing, thus the shifting of the load is depend on the current state of the system. It allows for processes to move from an over utilized machine to an underutilized machine dynamically for faster execution. This means that it allows for process pre-emption which is not supported in Static load balancing approach. An important advantage of this approach is that its decision for balancing the load is based on the current state of the system which helps in improving the overall performance of the system by migrating the load dynamically.

IV.PROBLEM DESCRIPTION

After reading the existing research papers, we have found the below listed problems:

- Machines have been assigned priority only on the basis of MIPS.
- Low QoS value tasks have been assigned high priority and are assigned to high priority virtual machines.Similarly HighQoS value tasks have been assigned low priority and are assigned to low priority virtual machines.
- Non dominated list is created in which new cloudlets are compared with the previous cloudlets in case to give the priority.

V.CLOUDSIM

CloudSim is a famous tool that is actually a toolkit for simulation of cloud scenarios [4].CloudSim has been developed as a Cloud Bus project in Australia [4]. CloudSim actually enables the users to have a proper insight into cloud scenarios without worrying about the low level implementation details [5]. CloudSim is invented as Cloud Bus Project at the University of Melbourne, Australia and supports system and behaviour modeling of cloud system components such as data centres, virtual machines (VMs) and resource provisioning policies. It implements generic application provisioning techniques that can be extended with ease and limited efforts. CloudSim helps the researchers to focus on specific system design issues without getting concerned about the low level details related to cloud-based infrastructures and services [7]. CloudSim is an open source web application that launches preconfigured machines designed to run common open source robotic tools, robotics simulator Gazebo. SimJava is a toolkit for building working models of complex systems. It is based around a discrete event simulation kernel at the lowest level of CloudSim. It includes facilities for representing simulation objects as animated icons on screen [7, 8]

VI.CONCLUSION

Cloud computing provides everything to the user as a service which includes platform as a service, application as a service, infrastructure as a service. As cloud computing is a wide area of research but one of the major issues is dynamic load balancing because overloading of a system may lead to poor performance which can make the technology unsuccessful. So there is always a requirement of efficient load balancing algorithm for efficient utilization of resources. This paper presents a concept of Cloud Computing and focuses on load balancing algorithm This paper also presents a concept of Cloud Computing along with research challenges in load balancing followed by a comparative survey of the load balancing algorithm in cloud computing with respect, resource utilization, performance, response time and overhead time.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

REFERENCES

[1] Aiash, Mahdi, Glenford Mapp, and Orhan Gemikonakli."Secure live virtual machines migration: issues and solutions." Advanced Information Networking and Applications Workshops (WAINA), 2014 28th International Conference on. IEEE, 2014.
[2] Beloglazov, Anton, and RajkumarBuyya. "Energy efficient allocation of virtual machines in cloud data centers." Cluster, Cloud and Grid Computing (CCGrid), 2010 10th IEEE/ACM International Conference on. IEEE, 2010.
[3] Galloway, Michael, Gabriel Loewen, and Susan Vrbsky. "Performance Metrics of Virtual Machine Live Migration." Cloud Computing (CLOUD), 2015 IEEE 8th International Conference on. IEEE, 2015.
[4] Wu, H.-S., Wang, C.-J., and Xie, J.-Y. (2013a). Terascalerelb-an algorithm of prediction based elastic load balancing resource management in cloud computing. In Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on, pages 649-654. IEEE.
[5] M. Belkhouraf, A. Kartit, H. Ouahmane, H. K. Idrissi, Z. Kartit and M. . E. Marraki, "A secured load balancing architecture for cloud computing based on multiple clusters," IEEE, 2015.
[6] David Escalnte and Andrew J. Korty, "Cloud Services: Policy and Assessment", EDUCAUSE Review, Vol. 46, July/August 2011.
[7] N. K. Chien, N. H. Son and H. D. Loc, "Load Balancing Algorithm Based on Estimating Finish Time of Services in Cloud Computing," ICACT, pp. 228-233, 2016.
[8] H. H. Bhatt and H. A. Bheda, "Enhance Load Balancing using Flexible Load Sharing in Cloud Computing," IEEE, pp. 72-76, 2015.
[9] Nguyen, QuyetThang; Quang-Hung, Nguyen; Tuong, Nguyen Huynh; Tran,Van Hoai; Thoai, Nam, "Virtual machine allocation in cloudcomputing for minimizing total execution time o on each machine," Computing Management and Telecommunication (ComManTel), 2013International Conference on , vol.,no., pp.241,245, 21-24 Jan. 2013.
[10] Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, Cesar A. F. De Rose,and RajkumarBuyya, "CloudSim: A Toolkit for Modeling andSimulation of Cloud Computing Environments and Evaluation of Resource ProvisioningAlgorithms," Software: Practice and Experience(SPE), Volume 41, Number 1,Pages: 23-50, ISSN: 0038-0644, Wiley Press, New York, USA, January, 2011.
[11] R. K. S, S. V and V. M, "Enhanced Load Balancing Approach to Avoid Deadlocks in Cloud," Special Issue of International Journal of Computer Applications (0975 –8887) on Advanced Computing and Communication Technologies for HPC Applications -ACCTHPCA, June 2012, pp. 31-35, 2012.
[12] Kumar Nishant, P. Sharma, V. Krishna, N. and R. Rastogi, "Load Balancing of Nodes in Cloud Using Ant Colony Optimization," IEEE, pp. 3-9, 2012