ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

IARJSET

International Advanced Research Journal in Science, Engineering and Technology

AGNI-PANKH 16

Jawaharlal Darda Institute of Engineering and Technology, Yavatmal

Vol. 4, Special Issue 3, January 2017

Novel techniques of load balancing on Cloud Computing: A review paper

Prof.C.S.Dhamande[1], Miss. Vaishnavi J. Deshpande[2], Miss. Gautami S. Shinde[3]

Professor, Computer Science & Engg, Jawaharlal Darda Institute of Engineering & Technology, Yavatmal, India[1]

Student, Computer Science & Engg, Jawaharlal Darda Institute of Engineering & Technology, Yavatmal, India[2, 3]

Abstract: As Cloud Computing is play an vital role for data storage in very minimal cost and accessible for all time over the internet, load balancing for the cloud computing has turned into a very interesting and important study area. It is a kind of Internet-based computing that provides shared processing resources and data to computers and other devices on demand. Efficient load balancing schemes ensures efficient resource utilization by provisioning of resources to cloud users on demand basis in pay-as-you-say manner. Cloud computing has become a highly demanded service or utility due to the advantages of high computing power, cheap cost of services, high performance, scalability, accessibility as well as availability.

Keywords: cloud computing, shared resources, Load balancing.

I. INTRODUCTION

Cloud is the cluster of distributed computers that provides on-demand computational resources over a network. Cloud Computing has become one of the popular technology adopted by both industry academia providing a flexible and efficient way to store and retrieve files. The goal of cloud computing is to allow users to take benefit from all of these technologies, without the need for deep knowledge about or expertise with each one of them. Cloud computing provides all of its resources as services, and makes use of the well-established standards and best practices gained in the domain of SOA to allow global and easy access to cloud services in a standardized way.

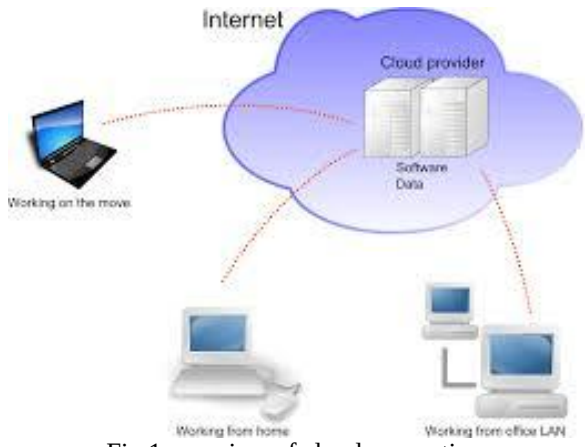

Fig 1 overview of cloud computing

Instead of paying high for IT hardware and software, it is now possible to rent (on demand) required infrastructure at cheap rates Clouds are made up of computing resources in the form of services to client and can be access remotely [1]. Cloud computing [2] led to an innovative approach in the way in which IT infrastructures, applications, and services. Cloud Structure is made up of resources like cloud servers which handle process and storage and client that ask for services. Information Technology is implementing in various organizations day by day are designed, developed, and delivered. [3] Cloud consists of several data centres, servers, clients which are interconnected in an efficient way. It includes fault tolerance, high availability, scalability, flexibility, reduced overhead for users, reduced cost of ownership, on demand services etc. With the rapid development of internet technology and network technology, more and more people use the internet to obtain information, shopping and entertainment. Cloud computing comes in three forms: public clouds, private clouds, and hybrids clouds. compare public, private, and hybrid clouds in terms of the different levels of security and management required.

A. Public cloud:

A public cloud is basically the internet. Service providers use the internet to make resources, such as applications (also known as Software-as-a-service) and storage, available to the general public, or on a 'public cloud.

B. Private Cloud:

Private clouds are data center architectures owned by a single company that provides flexibility, scalability, provisioning, and automation and monitoring. The goal of a private cloud is not sell "as-a-service" offerings to external customers but instead to gain the benefits of cloud architecture without giving up the control of maintaining your own data centre.

C. Hybrid Cloud:

Hybrid cloud, companies can maintain control of an internally managed private cloud while relying on the

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
AGNI-PANKH 16
Jawaharlal Darda Institute of Engineering and Technology, Yavatmal
Vol. 4, Special Issue 3, January 2017

public cloud as needed. For instance during peak periods individual applications, or portions of applications can be migrated to the Public Cloud.

Services in cloud

The overall objective of cloud computing is to provide this resources to client in the form of cloud services. Some of the services offered by cloud computing is Software as a Service (SaaS), Hardware as a Service (HaaS), Infrastructure as a Service (IaaS), Application as a Service (AaaS), Desktop as a Service (DaaS), Backup as a Service (BaaS), Network as Service (NaaS).

A. Software as a Service (SaaS):

SaaS user remove all the traditional parameter for data center such as licenses of software ,installation and support. SaaS uses the Web to deliver applications that are managed by a third-party vendor and whose interface is accessed on the clients' side. Most SaaS applications can be run directly from a Web browser, without any downloads or installations required. Examples for SaaS are Gmail, Google Apps, Microsoft Office 365, Google+, facebook, yahoo.

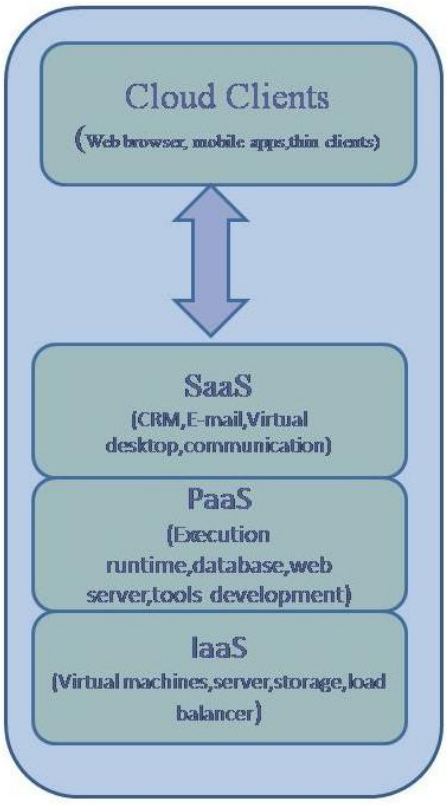

Fig 2. cloud computing services

B. PaaS: Platform as a Service (PaaS)[4] deliver computational resources through a platform. What developers gain with PaaS is a framework they can build upon to develop or customize applications. PaaS makes the development, testing, and deployment of applications quick, simple, and cost-effective, eliminating the need to buy the underlying layers of hardware and software. Paas user free from installation of hardware, software to run an a application , it handles from anywhere to any places.

Examples of Paas Beanstalk, Windows Azure, Heroku, Force.com and Google App Engine.

C. IaaS: Infrastructure as a Service (IaaS)[4] delivers computer infrastructure (such as a platform virtualization environment), storage, and networking. Instead of having to purchase software, servers, or network equipment, users can buy these as a fully outsourced service that is usually billed according to the amount of resources consumed. Basically, in exchange for a rental fee, a third party allows you to install a virtual server on their IT infrastructure. Compared to SaaS , PaaS and IaaS users are responsible for managing more: applications, data, runtime, middleware, and O/S.

II. LOAD BALANCING

Load balancing is a computer network method for distributing workloads across multiple computing resources, for example computers, a computer cluster, network links, central processing units or disk drives. Load balancing plans to optimize resource use, maximize throughput, minimize response time, and evade overload of any one of the resources. By the use of multiple components with load balancing instead of a single component may increase reliability through redundancy. Load balancing [5] is a process of reallocate the total load to the individual nodes of the collective system to make resource utilization effective and to develop the response time of the job, simultaneously removing a condition in which some of the nodes are over loaded while some others are under loaded. There are many different kinds of load balancing algorithms available, which can be categorized mainly into two groups. The following section will discuss these two main categories of load balancing algorithms.

2.1 Static Algorithm

[6]Static algorithms divide the traffic equivalently between servers. By this approach the traffic on the servers will be disdained easily and consequently it will make the situation more imperfectly. This algorithm, which divides the traffic equally, is announced as round robin algorithm.

2.2 Dynamic Algorithm

[6]Dynamic algorithms designated proper weights on servers and by searching in whole network a lightest server preferred to balance the traffic. However, selecting an appropriate server needed real time communication with the networks, which will lead to extra traffic added on system. Additionally, Cloud vendors are based on automatic load balancing services, which allowed entities to increase the number of CPUs or

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
AGNI-PANKH 16
Jawaharlal Darda Institute of Engineering and Technology, Yavatmal
Vol. 4, Special Issue 3, January 2017

memories for their resources to scale with the increased demands . This service is optional and depends on the entity's business needs. Therefore load balancers served two important needs, primarily to promote availability of cloud resources and secondarily to promote performance[7]. Load balancing is one of the major issues related to cloud computing. The load may be memory, CPU capacity, network load or delay load. It is always required that work load must be shared among the various nodes of the distributed system so as to improve the resource utilization and also for better performance of the computing system. In order to avoid system failure, load balancing is often used by controlling the input traffic and stop sending the workload to resources which become overloaded and non-responsive.

Types of Load balancing:

2.1. Carton

R. Stanojevic[8] proposed a mechanism CARTON for cloud control the unifies the use of load balancing(LB) and distributed rate limiting(DRL). LB is used to equally distribute the jobs to different Servers so that the associated costs can be minimized and DRL is used to make sure that the resources are distributed in a way to keep a fair resource allocation. With very low computation and communication overhead, this algorithm is simple and easy to implement

2.2 Compare and Balance

Zhao[9] addressed the problem of intra-cloud load balancing amongst physical hosts by adaptive live migration of virtual machines. A load balancing model is designed and implemented to reduce virtual machines' migration time by shared storage to balance load amongst servers according to their processor or IO usage and to keep virtual machines zero-downtime in the process. A distributed load balancing algorithm COMPARE and BALANCE is also proposed that is based on sampling and reaches equilibrium very fast. This algorithm assures that the migration of VMs is always from high-cost physical hosts to low-cost host but assumes that each physical host has enough memory which is a weak assumption

2.3 Events-Driven

V. Nae[10] presented an event-driven load balancing algorithm for real time massively multiplayer online games (MMOG). This algorithm after receiving capacity events as input, analyzes its components in context of the resources and the global state of the game session, thereby generating the game session load balancing actions. It is capable of scaling up and down a game session on multiple resources according to the variable user load but has occasional QoS breaches.

2.4 Scheduling Strategy on LB of VM Resource

A scheduling strategy on load balancing of VM resources that uses historical data and current state of the system. This strategy achieves the best load balancing and reduced dynamic migration by using a genetic algorithm. It helps in resolving the issue of load imbalance and high cost of migration thus achieving better resource utilization [11].

2.5 Honeybee Foraging Behavior

M.Randles investigated a decentralized honey-bee based load balancing technique that is a nature- inspired algorithm for self -organization. It achieves global load balancing through local server actions. Performance of the system is enhanced with increased system diversity but throughput is not increased with an increase in system size. It is best suited for the conditions where the diverse population of service types is required [10].

2.6 Biased Random Sampling

M. Randles investigated a distributed and scalable load balancing approach that uses random sampling of the system domain to achieve self organization thus balancing the load across all nodes of the system. The performance of the system is improved with high and similar population of resources thus resulting in an increased throughput by effectively utilizing the increased system resources [10].

2.7 Message Oriented Model

Zenon Chaczko presented a model that uses XMPP for load balancing. This technology is open for real time communication between various parties. XMPP clients send presence information to XMPP presence servers and XML streams containing details of presence information of clients produced by these servers. Using a load balancer on the top of an XMPP server allowed incoming requests to be prioritized and handled by a generic service [13].

2.8 Open Flow Model

Hardeep Uppal presented a model in which open flowswitch is used. Open flow switches are like a standard switch with a flow table performing packet lookup and forwarding. The difference lies in how flow rules are inserted and updated inside the switch's flow table [14].

2.9 Self-organized Load Balancing

Algorithm Giuseppe has presented a new method for load balancing in which a node with highest capacity serves as super peers. At first level, algorithm find out the capacity of every peer, i.e., the amount of service requests that peer is able to fulfill in a client time unit. This, in turn, is reflected in the Myconet overlay as the target Bs of peers maintained by a super peer. This way, super peers are well positioned to effectively balance their neighbors' request queues [15].

III.CONCLUSION

The best load balancing approaches in cloud computing are those ones which have dynamic,

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

IARJSET

International Advanced Research Journal in Science, Engineering and Technology

AGNI-PANKH 16

Jawaharlal Darda Institute of Engineering and Technology, Yavatmal

Vol. 4, Special Issue 3, January 2017

distributed and non-cooperative features. Because it is not always practically feasible or cost efficient to maintain one or more idle services just as to fulfil the required demands. It is possible to make more efficient. As Cloud Computing is one of the best platform that gives storage of data in very minimal cost and accessible for all time over the internet, load balancing for the cloud computing has turned into a very interesting and important study area.

REFERENCES

[1]Anthony T. Velte, Toby J. Velte, Robert Elsenpeter, " Cloud Computing: A Practical Approach ", The McGraw-Hill Companies(2010), [Book]

[2] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, I. Brandic, Cloud computing and emerging IT platforms: vision, hype, and reality for delivering IT services as the 5th utility, Future Generation Computer Systems 25 (6) (2009) 599 –616.

[3] Reena Panwar "A Comparative Study of Various Load Balancing Techniques in Cloud Computing" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 3 Issue 9, September- 2014

[4] E.Gorelik, "Cloud Computing Models", Massachusetts Institute of Technology Cambridge, MA, 2013. Available: http://web.mit.edu/smadnick/www/wp/2013-01.pdf

[5] K. Ramana, A. Subramanyam and A. Ananda Rao, Comparative Analysis of Distributed Web Server System Load Balancing Algorithms Using Qualitative Parameters, VSRD-IJCSIT, Vol. 1 (8), 2011, 592-600

[6] L . Aruna,Dr. M . Aramudhan"Models of Load Balancing Algorithm in Cloud Computing "IPASJ International Journal of Computer Science (IIJCS), Volume 2, Issue 12, December 2014

[7] Gunpriya Makkar, Pankaj Deep Kaur "A Review of Load Balancing in Cloud Computing" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, 2015 ISSN: 2277 128X

[8] Deepak B S , Shashikala S V , Radhika K R "Balancing Techniques in Cloud Computing: A Study" International Conference on Information and Communication Technologies (ICICT- 2014)

[9] Zhao Y. and Huang W. (2009) 5th International Joint Conference on INC, IMS and IDC, 170-175.

[10] Nae V., Prodan R. and Fahringer T.(2010) 11th IEEE/ACM International Conference on Grid Computing (Grid), 9-17

[11] Randles M., Lamb D. and Taleb-Bendiab A. (2010) 24th International Conference on Advanced Information Networking and Applications workshops, 551-556.

[12] Meenakshi Sharma, Pankaj Shrama, Dr. Sandeep Sharma, Efficient Load [6] Hardeep Uppal, dane Brandon, OpenFlow based load balancing.

[13]Zenon Chaczko, Venkatesh Mahadevan, Shahrzad Aslanzadeh and Christopher Mcdermid, Availabilty and load balancing in cloud computing, 2011 International Conference on Computer and Software Modeling, IPCSIT vol.14 (2011) ACSIT Press, Singapore.

[14] Hardeep Uppal, dane Brandon, OpenFlow based load balancing.

[15] Ratan Mishra, Anant jaiswal, Ant colony optimization: A Solution of load balancing in cloud, International Journal of Web & Semantic Technology (IJWesT) Vol.3, No.2, April 2012