IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Augmenting QoS in Cloud by Utilizing load balancing algorithm

Prabhu.B [1], Rohit.V[2], Vengatesh.R[3], Vignesh.J[4], Harikrishna Pillutla[5]

Student, CSE Department, Rajiv Gandhi College of Engineering and Technology, Puducherry, India[1, 2, 3, 4]

Assistant Professor, CSE Department, Rajiv Gandhi College of Engineering and Technology, Puducherry, India[5]

Abstract: Cloud is combination of server and computer that are interconnected together to provide resources to clients. It expose as a new brand computing prototype that aims to supply reliable, custom-made and QoS (Quality of Service) and dynamic environments for end customers. In cloud environment the resource allocation plays a major role in the performance of entire system and also the customer satisfaction provided by the system. In the proposed system, the main goal is to provide equal and distributed load over all the processors. Various strategy algorithm and policy has proposed and implementing Load balancing in Cloud environment. The paper, we represent a WRR algorithm combined with queuing algorithm applied to expeditiously order computation jobs among the processors onto the cloud datacenters with less communication overhead.

Keywords: Cloud Computing, QoS, WRR algorithm, Load Balancing.

I. INTRODUCTION

Cloud computing provides a new computing paradigm which aim to provide a dependable, manual and QoS (Quality of Service) guaranteed a computing dynamic environments for end-users. The Combination of Distributed processing, parallel processing and grid computing together emerge as cloud computing [1]. The basic principle of cloud computing is that data is deposited in data center of an internet. The company provide cloud service and by handling and preserving the operation of the data center. The users can use the stored data at any time by using API given by cloud providers through terminal equipment connected to internet [1]. Cloud may reduce the cost by avoiding the capital expenditure of the company in lease and the physical infrastructure from a third party provider. Due to the adoptable nature of cloud, we can easily access more resources from the cloud providers when we need to enhance the business. To increase the higher value of above mentioned benefits, the services offered a term of huge resources by allocating desirable to the applications running in the cloud. The following segment discuss about the significance of resource allocation. Load balancing provides the technology that elevates the utilization of resources and thereby providing a throughput with less response time and by sharing the equal load between servers. To attain load WRR balancing algorithm and the resource consumption there are several algorithms that can be used. Best example for load balancing is online shopping cart. Without load balancing, users experiences a delay while ordering, transactions and buying. Load balancing solutions apply to the surplus servers which helps to a better distribution of the communication traffic and there by online purchasing will made easy [3]. Load balancing is the priority issues in cloud computing. Load balancing is to allot the workload dynamically across all partition nodes in the cloud data center and there by refusing the circumstances where some nodes are stored in huge manner while others are comfortable or doing some little work. By improving the performance of the system resource utilization, it helps to reduce the Load balancing issue in a cloud computing environment and also by making every computing resource in a distributed manner and there by efficiently performed. When several service fail, load balancing helps to providing and removing of instant applications without fail.

TABLE.1: Cloud Balancing Policies

Cloud	*Type*	R	RR	WRR	LC	WLC	S
EC2 (ELB&Route53)	IaaS		*	*	*		
GoGrid	IaaS		*	*	*	*	*
Rackspace	IaaS	*	*	*	*	*	
App Engine	PaaS						*
Azure	PaaS		*				
CloudFoundry	PaaS		*	*	*		
Heroku	PaaS		*				

II. SCOPE OF THE PAPER

Resource allocation includes touchable device such as hardware to the better utilization of softer assets such as human capital. It involves balancing competing needs and priorities and there by determining the most effective course of action and order to maximize the effective use of limited resources and to gain the best return on investment.

Resources are dynamic in nature and by varying the load resource with efficient change in Configuration of cloud. A poor scheduling policy may leave many processors in an inactive manner while a clever may consume a huge portion of the CPU cycles. In the active approach we face issue in distributed dispatching of task to resource [7].

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

III. RELATED WORK

All Cloud computing is a representation of allowing expedient, on-demand system entry to a configurable calculating resources [5]. Aligning and the interior structures of resources have been studied, but the power management is not done. Amazon Elastic Compute Cloud (EC2) is an instance of HaaS, which is a structure of cloud computing [5]. Equity should be followed whereas combining several types of resource into deliberation. Cloud services are simple and there by decrease both trade costs and ecological loads. Quick flexibility and the measured services are highlighted for cloud computing scenario. There are several papers that transpose the algorithms to achieve fairness for cases where a combined resource allocation is not measured. To yield the cloud computing services reasonably, it is important for optimizing the resource distribution under the statement. Besides to be the current processing capability and storage ability, it is essential to declare the bandwidth to enter them at equivalent time. The unstable situation occurs when unnecessary produced data acquires and an important economic cost on transmit and inventory in cloud. A pecuniary approach presented the services "offer" for the commanding the distributed performance. They allows the consumer to charge only at the time when desirable, only a preferred quantity of calculating resources and without upsetting and there by concerning the position or interior structures of the resources [5]. The National Institute of Standards and Technology (NIST) established four necessary distinctiveness of cloud computing: resource pooling. It is predictable that the employing efforts will fasten up will their movement from construction and possessing their individual systems on renting cloud computing services. In this work, an optimal resource allocation method is used for optimization of resource usage based on user's task [8].

IV. SYSTEM ANALYSIS

The cloud users get a good quality of services from their service providers with desired cost. The quality and cost are based on their source allocation process in the particular service environment. The provider should assign the resource to the clients in an optimal way. There are so many resource allocation models that are used in loud computing area. The following surveys perform in the prior works of resource allocation models of the cloud computing environment [9]. We can quickly access more resources from the cloud providers when we need to enhance our business. The remote resources can access to the cloud services from anywhere at any time. To utilize the max effort, the services offered in term of resources should be orderly arranged to the applications in cloud [11].

A. EXISITING WORK

It activates the cloud service providers and the operating centers to meet an authorized customer and to increase the QoS levels and there by trusting the QoS metric collection and analysis implementation scheme that extends traditional monitoring, management, and the response for IaaS and SaaS to complete SOA stack that may comprise business logic in backend as a service and governance the gaming services. This paper contains the real-world scenarios that describe the applications to voice and data systems for the performance measures and to DDoS for security measures [14].

B. DRAWBACKS OF EXISTING SYSTEM:

1. Higher in overhead
2. Communicational cost increased
3. Low in QOS

C. PROPOSED WORK

We have collaborated WRR load balancing algorithm with queue algorithm. WRR load balancing is commonly routing policy available in cloud load balancers. However, there existing a deficiency of effective mechanisms to decide the weights assigned to a specified server to attain overall optimal revenue of the system. We concentrate in particular algorithms based on the closed queuing networks for multi-class workloads, which used to describe the application with the service level understanding differentiated across multi users. We also equate the efficiency of queuing theoretic methods with the simple heuristics that doesn't needs to specify a stochastic model of application. Cloud providers are largely increasing the load balancing. Requests are send to end servers by the load balancer and to the following certain load balancing policies. These policies normally intend to reduce the imbalance between different servers and to improve the system throughput or to shorten the response time. Among commonly existing load balancing polices, round robin plays a major role and there may increase utilization among cloud providers [3]. Considering the heterogeneous resources available in the cloud, the weighted round robin policy, also offered a several cloud providing and there by placing a huge load to the servers with high weight using the round robin policy and with some similarities. From the experimental results a similar behaviour is found between the two kinds of policies, which makes it possible to assign the weights of the weighted round robin policy, according to the probabilistic estimation the heuristic algorithm and the optimization introduced the probabilistic routing.

Our proposed system also includes a queue based job scheduling algorithm for efficient execution of user jobs. It paper also includes the relative performance analysis for a desired and the proposed job scheduling algorithm along with other well-known job scheduling algorithms and by considering the parameters like average waiting time, average response time. Job scheduling algorithms is one of the most challenging theoretical issues in the cloud computing area. Some intensive researches have been done in the area of job scheduling of cloud computing. Jobs are queued and collected into a set when they arrive in the batch mode. The scheduling algorithm will start after a fixed period time.

In our proposed paper we are combining the WRR algorithm and the queuing algorithm and there by getting the result by providing equal and distributed load over all the processors and expeditiously order the computation jobs among the processors onto the cloud datacenters with less communication overhead.

Algorithm 1: Weighted Round Robin

Require: W
i = -1
cw = 0
while TRUE do
i = (i + 1) mod n
if i == 0 then
cw = cw - gcd (W)
if cw < 0 then
cw = max (W)
if cw == 0 then
return null
end if
end while

D. ADVANTAGES OF PROPOSED SYSTEM

1) Improve the Load balancing in cloud environment by portioning the nodes into two idle and busy nodes
2) Communication overhead and queue waiting reduced
3) No longer wait for low priority task, this can be achieved by monitoring the resource execution and calculating the idle time
4) All service request will validate and completed within in mentioned SLA.
5) Overload resource avoidance and energy consumption.

V. SYSTEM ARCHITECTURE

It is a conceptual model that defines the structure, behavior, and more views of a system. It is a formal description and there by representing of a system and planned in such a way that supports reasoning about the structures and behaviors of the system [16].

VI. EVALUATION

Cloud computing is an future technology and the several researches have been carried out in order to solve the challenges faced by cloud. There are multiple challenges where the cloud is facing the resource allocation techniques.

This paper provides entire description of several resource allocation techniques. Load balancing is one of the priority issues in cloud computing [17]. By enhancing the system resource performance and there by utilizing it and helps to reduce the Load balancing issue in cloud computing and also make sure that every computing resource is distributed efficiently and reasonably. When multiple services tend to fail the load balancing helps to provisioning and de-provisioning of instances of applications without fail.

VII. CONCLUSION

This paper concludes by an queue based scheduling algorithm and the weighted round robin in a load balancing policy leads to increasing a load balancing in an cloud environment. However, there will be increase in effectiveness of the mechanisms to decide the weight assigned to each server and to achieve an overall optimal revenue of the system.

REFERENCES

[1]. M. A. Adnan, R. Sugihara, and R. K. Gupta, "Energy efficient geographical load balancing via dynamic deferral of workload", in Proc. of IEEE Cloud. pp.188-195, 2012.
[2]. H. Goudarzi and M. Pedram, "Geographical load balancing for online service applications in distributed datacenters", in Proc. of IEEE Cloud. pp.351-358, 2013.
[3]. M.F. Arlitt, C.L. Williamson, "Web- server workload characterization: The search for invariants", IEEE/ACM Trans. on Networking, vol. 5, no. 5, pp. 631-645, Oct1997.
[4]. W. C. Cheng and R. R. Muntz, "Optimal routing for closed queuing networks", Performance Evaluation13 Vol.1, pp.3-15, 1991.
[5]. A. Hordijk and J.A. Loeve, "Optimal static customer routing in a closed queuing network", Statistica Neerlandica, vol.54, Issue.2, pp.148-159, July 2000.
[6]. P.Boyle, "Load balancers can provide the busiest Web sites with nonstop performance", PC magazine, Feb. 1997.
[7]. R. Ranjan, L. Zhao, X. Wu, A. Liu, A. Quiroz, and M. Parashar, "Peer to-peer cloud provisioning: Service discovery and load-balancing", in Cloud Computing. Springer, pp.195-217, 2010
[8]. V.Cardellini, M. Colajanni, P.S. Yu, "Redirection algorithms for load sharing in distributed Web- server systems", Proc. of 19th IEEE Int'l. Conf. on Distributed Computing Systems (ICDCS'99), Austin, TX,, June 1999.
[9]. Li.Yetal, "A new Class of Priority-based Weighted Fair Scheduling Algorithm ", Physics Procedia , Issue.33, pp.942– 948,2012.
[10]. Monir Abdullah Moetal, "Optimal Workload Allocation Model for Scheduling Divisible Data Grid Applications", Future Generation Computer Systems,Vol.26, pp.971-978,2010 .
[11]. Mousumi Petal, "Dynamic job Scheduling in Cloud Computing Based on Horizontal Load Balancing", Int.J.Computer. Tech.Appl, Vol.2, no.5, pp.1552-1556,2011
[12]. Tai-Lung Chen et al, "Scheduling of Job Combination and Dispatching Strategy for Grid and Cloud System", GPC, pp. 612–621,2010
[13]. T. Cui, T. Xue and K. Nahrstedt."Optimal Resource Allocation in Overlay Multicast", ICNP, pp.137-172, 2003
[14]. K. Czajkowsk, I. Foster, C. Kesselman, V. Sander and S. Tuecke, "SNAP: A protocol for negotiating service level agreements and coordinating resource management in distributed systems", 2002.
[15]. L. Badger, T. Grance, R. Patt-Corner, and J. Voas, "Cloud computing synopsis and recommendations" ,National. Institute. Standards Technol. (NIST),U.S Department of Commerce, Gaithersburg, MD, USA, NIST Special Publication 800, May 2012.
[16]. C. Tang, M. Steinder, M. Spreitzer, and G. Pacifici, "A Scalable Application Placement Controller for Enterprise Data Centers", Proc. Int'l World Wide Web Conf. (WWW '07), May 2007.
[17]. Weikun Wang, Giuliano Casale, "Evaluating Weighted Round Robin Load Balancing for Cloud Web Services", Department of Computing, Imperial College ,London., UK