A Two-Phase Load Balancing Algorithm for Cloud Environment

Archana Singh, Madan Mohan Malaviya University of Technology, India

Rakesh Kumar, Madan Mohan Malaviya University of Technology, India

ABSTRACT

Load balancing is the phenomenon of distributing workload over various computing resources efficiently. It offers enterprises to efficiently manage different application or workload demands by allocating available resources among different servers, computers, and networks. These services can be accessed and utilized either for home use or for business purposes. Due to the excessive load on the cloud, sometimes it is not feasible to offer all these services to different users efficiently. To solve this excessive load issue, an efficient load balancing technique is used to offer satisfactory services to users as per their expectations also leading to efficient utilization of resources and applications on the cloud platform. This paper presents an enhanced load balancing algorithm named as a two-phase load balancing algorithm. It uses a two-phase checking load balancing approach where the first phase is to divide all virtual machines into two different tables based on their state, that is, available or busy while in the second phase, it equally distributes the loads. The various parameters used to measure the performance of the proposed algorithm are cost, data center processing time, and response time. Cloud analyst simulation tool is used to simulate the algorithm. Simulation results demonstrate superiority of the algorithm with existing ones.

KEYWORDS

Cloud Analyst, Equally Spread Current Execution, Load Balancing, Throttled, Virtual Machine (VM)

1. INTRODUCTION

Cloud computing started way back in the 1960's, when an intergalactic computer network was first introduced. After that many innovation reforms came and made cloud as the most favorable and preferable network for any kind of resource. Cloud creates condition or circumstance that puts one in a favorable or superior position for their uses. In the present era cloud computing is an emerging technology. With its lots of innovative ideas, it is becoming popular in many fields like say education, business, research, etc. Day by day researches and services providers doing their effort to make it more efficient as compared to the previous one. Today cloud users are not limited, due to its ease of access and reliability, the cloud is being used in numerous fields like industries, factories, healthcare, business organizations and institutions etc. Cloud has a vast number of users. Cloud gives three kinds of services which are namely platform as a service (PaaS), software as a service (SaaS) and infrastructure as a service (IaaS). These services are provided by different cloud service providers. Robustness, flexible, cost-effective, multi-valued, scalable, and many more; these are some features or quality belongs to the cloud. Load balancing is a mechanism for handling load on the cloud or a cloud computing part working together with other cloud services for providing efficient services to its user.

This paper proposes an algorithm for handling some existing problems on load balancing and give minimized datacenter processing time-based load balancing algorithm on the cloud. This algorithm is also cost-effective as compared to some existing load balancing approaches. The proposed algorithm is a combination of two phases one phase is about the virtual machine division and the second phase is about the equal load distribution among only available virtual machine i.e. in the second phase actual allocation and de-allocation of a virtual machine are performed.

Rest of the paper is organized as follows. Section 1 is about the introduction of cloud computing along with some history of cloud. Section 2 concludes previously purposed some research work by different authors on load balancing with their summarized discussion. Section 3 is about the load balancing requirement in the cloud. Section 4 is about the proposed load balancing approach and architecture of Two-phase load balancing algorithm. Section 5 shows provisional results between the proposed approach and the existing load balancing approaches. Finally, Section 6 gives the conclusion and future work.

2. LITERATURE REVIEW

When we say about load, that means it shows the number of users demand for the server. There are various reasons behind the high demand of the Internet where the first reason is the industrial use of cloud. Ttoday, many technical firms use cloud services for their work. Using a cloud for industrial purposes is not only the infrastructure-less service, apart from this, but cloud services are also cost-efficient, reliable for work from anywhere any time and there are many more advantages of using cloud services for industrial purposes. The reason for the high demand of the internet is social networking sites. There are many forms of social media such as blogs, forums, chat apps, photo sharing, Facebook, Instagram, Twitter and these are reasons for the increasing popularity of Online Social Networking (OSN). As per the report of Statista, currently 2.95 billion users of OSN. When the popularity of OSN increasing, it creates many kinds of serious cybercrime issues such as identity theft, malware, fake profile, spam ware, social bots, Sybil attack, sexual harassment, and many more (Sahoo & Gupta, 2018).

These issues are the most challenging factor for the Internet and this could be overcome by using the article (Gupta et al., 2018) where Gupta et el. (Gupta et al., 2018) describe a client-server JavaScript code rewriting-based framework that protects online users' privacy against XSS worms in OSN. The server-side generates a prediction graph that looks for extracting the JavaScript code and converting it to a separate file. This transfer takes place in complete isolation of unreliable JavaScript code and data. It conducts runtime monitoring of client-side JavaScript code to detect tainted flow of trusted JavaScript variables. The context of such dynamically tilted variables is determined by examining string analysis to see if this is a weak point. Finally, the decoding function takes place in the embedded JavaScript code and the parameter values of the HTTP request in the fuzzy malicious JavaScript code. If the match is found, XSS will be the attack vector. If not, it's not.

For handling cloud computing challenges, there are various algorithms already designed in which salient features are depicted in Table 1. Apart from the many types of research work on the enhancement of cloud load balancing algorithm and some of them thoroughly described here. In (Sekarani et al., 2019), authors show the relative analysis of existing load balancing algorithms based on their processing cost, response time, and data center processing time. Behalf of comparative study authors found that when concerning overall response time, FCFS is best among round-robin, equally spread current execution, throttled and shortest job first but when concerning about data center processing time and cost of a virtual machine then ESCE performed well among them. Xiao et al. (Xiao et al., 2019) present a load balancing approach which is based on request balancing among users in multiple autonomous cloud provider environment where they have described how other user's decision impact on request strategy and payment of each user. Variational Inequality (VI) theory along

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/a-two-phase-load-balancing-algorithm-for-cloud-environment/266227

Related Content

Saliency Priority of Individual Bottom-Up Attributes in Designing Visual Attention Models
Jila Hosseinkhaniand Chris Joslin (2018). *International Journal of Software Science and Computational Intelligence (pp. 1-18).*
www.irma-international.org/article/saliency-priority-of-individual-bottom-up-attributes-in-designing-visual-attention-models/223491

ILP Applications to Software Engineering
Daniele Gunetti (2007). *Advances in Machine Learning Applications in Software Engineering (pp. 74-102).*
www.irma-international.org/chapter/ilp-applications-software-engineering/4857

Entropy Quad-Trees for High Complexity Regions Detection
Rosanne Vetro, Dan A. Simoviciand Wei Ding (2013). *Advances in Abstract Intelligence and Soft Computing (pp. 299-316).*
www.irma-international.org/chapter/entropy-quad-trees-high-complexity/72788

Using Clustering for Forensics Analysis on Internet of Things
Dhai Eddine Salhi, Abelkamel Tariand Mohand Tahar Kechadi (2021). *International Journal of Software Science and Computational Intelligence (pp. 56-71).*
www.irma-international.org/article/using-clustering-for-forensics-analysis-on-internet-of-things/266228

Emergence Phenomenon and Fuzzy Logic in Meaningful Image Segmentation and Retrieval
Sagarmay Deband Siddhivinayak Kulkarni (2012). *Machine Learning Algorithms for Problem Solving in Computational Applications: Intelligent Techniques (pp. 167-178).*
www.irma-international.org/chapter/emergence-phenomenon-fuzzy-logic-meaningful/67702