Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 7, Issue. 12, December 2018, pg.154 – 157

Load Balancing Techniques in Cloud Computing: A Review

Manisha Thakur
Research Scholar, Bahra University, Waknaghat, Shimla Hills
tmanisha194@gmail.com

Parul Gazta
Assistant Professor, Bahra University, Waknaghat, Shimla Hills
parulgazta@gmail.com

Abstract: The cloud computing is the architecture in which cloudlets are executed by the virtual machines. The most applicable virtual machines are selected on the basis of execution time and failure rate. Due to virtual machine overloading execution time and energy consumption is increased at steady rate. In the previous year's many techniques are proposed for the virtual machine migration. In this paper, techniques of virtual machine migration are reviewed and analyzed in terms of certain parameters

KEYWORDS: VM migration, Cloudlet, virtual machines, cloudsim

Introduction

A large pool of systems which gather together in a public or private environment for providing resources in a shared manner is known as the cloud computing system. This provides a dynamic scalable infrastructure for the organization as well as the data stored. There is a reduction in the cost computation, application hosting, content storage and delivery with the help of cloud computing technique. There are direct cost advantages achieved due to the utilization of this technique. Also, the data center is to be transformed from a capital-intensive set up to a variable cost environment [1]. The availability of cloud computing requires Software, hardware, application, platform, infrastructure and storage with an internet connection. There are certain steps involved within the cloud computing systems. In grid computing large scale problems are solved with the help of parallel computing [2]. Utility computing provides the resources on the basis of metered services which means that the users will have to pay on the basis of amount they consumed the services. SAAS is a part of the service model of cloud

computing. It provides the network based subscription to the applications. Cloud Computing involves the services from all of the above given. In other terms it provides IT resources anywhere anytime to the user. Enterprises can choose to deploy applications on Public, Private or Hybrid clouds [3]. Cloud Integrators can have a crucial impact on determining what the right cloud path for each organization is. Public clouds are owned and operated by outsiders. They convey better economies of scale than customers, as the infrastructure costs are spread among a mix of users, giving every individual customer an attractive minimal cost, "Pay-as-you-go" model [4]. Private clouds are built exclusively for a single enterprise. They aim to address concerns on data security and offer greater control, which is typically lacking in a public cloud. Hybrid Clouds join both public and private cloud models. With a Hybrid Cloud, service providers can use outsider Cloud Providers in a full or partial manner which helps in expanding the flexibility of computing. The Hybrid cloud environment is capable of giving on-demand externally provisioned scale [5]. The ability to augment a private cloud with the resources of a public cloud can be utilized to manage any unexpected surges in workload. The cloud resources are to be allocated is a proper manner and this is a prior objective of this system. It is to be made sure that the financial profit is the highest in the selected architecture. One of the major function of cloud computing is the allocation of resources in which the resources are distributed in a proper manner [6]. For the purpose of satisfying the requirements of users, resource allocation is relates to provide commodities are services. It is done keeping is mind the perspective of the consumer. The economy of the industries is increased by the proficient resource allocation results provided. The clients enter the resources at very less chances for a small job when the skill is to be deployed as a service [7]. In the certain systems, the resource allocation as well as the scheduling for providing proper coordination for the job resources is done by including a key. There should be no violation of the set of constraints while this resource allocation process. There is no prediction of which resources are perfect for the allocation in a particular job provided by the cloud users [8]. The perfect resource allocation procedure involves the proficient allocation along with the least number of resources to be used. This will help in providing maximum profit for the users. The main aim of resource allocation strategy is to provide resources from team for the purpose of concurrent allocation in the cloud computing environment.

Literature Review

Huangke Chen, et.al (2015) introduced in this paper [9], an interval number theory to portray the uncertainty of the computing environment and a scheduling engineering to mitigate the impact of uncertainty on the task scheduling quality for a cloud data center. Based on this design, a novel scheduling algorithm (PRS1) is presented. The experimental results demonstrate that PRS performs superior to those algorithms, and can effectively improve the performance of a cloud data center.

Doulamis ND, et.al (2014) proposed in this paper [10], an algorithm for allocating tasks to resources that minimizes the infringement of the tasks' time requirements. The exact time scheduling of the tasks on the resources is then decided by considering the time constraints. Experimental results demonstrate that the proposed algorithm outperforms other scheduling algorithms for various values of the granularity and the load of the task requests.

Abdul Hameed, et.al (2014) discussed in this paper [11], open challenges connected with energy efficient resource allocation. Accessible techniques presented are summarized based on the energy-efficient research dimension taxonomy. The focal points and inconveniences of the existing techniques are comprehensively analyzed against the proposed research dimension taxonomy. On the basis of this study various techniques are enlisted which can be utilized as per their benefits.

Young Choon Lee, et.al (2012) proposed in this paper [12], two energy-cognizant task consolidation heuristics are presented, which aim to maximize resource utilization and explicitly consider both dynamic and sit without moving energy consumption. The heuristics assign every task to the resource on which the energy consumption for executing the task is explicitly or implicitly minimized without the performance degradation of that task. Based on the experimental results, the heuristics demonstrate their efficient energy-saving capability.

Zhanjie Wang, et.al (2015) proposed in this paper [13], a dynamically hierarchical, resource-allocation algorithm for multiple cloud nodes collaborating in big data environment. Both theoretical and experimental results represent that the proposed

algorithm outperforms the MinMin algorithm in terms of communication traffic and makespan. The results demonstrate that DHRA can reduce message number and communication traffic significantly, with the equal or even less tasks finish time compared with MinMin.

Guiyi Wei, (2010) proposed a pragmatic approximated solution with the accompanying two steps. Every participant solves its optimal problem independently, without consideration of the multiplexing of resource assignments [14]. It is demonstrated that Nash equilibrium dependably exists if the resource allocation game has feasible solutions. The advancement problem considered in this paper relates to a large proportion of cloud-based computing services.

Author's Names	Year	Description	Outcomes
Huangke Chen, Xiaomin Zhu, Hui Guo, Jianghan Zhu, Xiao Qin, Jianhong Wu	2015	The authors introduced an interval number theory to portray the uncertainty of the computing environment and a scheduling engineering to mitigate the impact of uncertainty on the task scheduling quality for a cloud data center.	The experimental results demonstrate that PRS performs superior to those algorithms, and can effectively improve the performance of a cloud data center.
Doulamis ND, Kokkinos P, Varvarigos E	2014	The authors proposed an algorithm for allocating tasks to resources that minimizes the infringement of the tasks' time requirements.	Experimental results demonstrate that the proposed algorithm outperforms other scheduling algorithms for various values of the granularity and the load of the task requests.
Abdul Hameed, Alireza Khoshkbarforoushha, Rajiv Ranjan, Prem Prakash Jayaraman, Joanna Kolodziej, Pavan Balaji, Sherali Zeadally, Qutaibah Marwan Malluhi, Nikos Tziritas, Abhinav Vishnu, Samee U. Khan, Albert Zomaya	2014	The authors discussed open challenges connected with energy efficient resource allocation. Accessible techniques presented are summarized based on the energy-efficient research dimension taxonomy.	On the basis of this study various techniques are enlisted which can be utilized as per their benefits.
Young Choon Lee, Albert Y. Zomaya	2012	The authors proposed two energy-cognizant task consolidation heuristics are presented, which aim to maximize resource utilization and explicitly consider both dynamic and sit without moving energy consumption.	Based on the experimental results, the heuristics demonstrate their efficient energy-saving capability.
Zhanjie Wang, Xianxian Su	2015	The authors proposed a dynamically hierarchical, resource-allocation algorithm for multiple cloud nodes collaborating in big data environment.	The results demonstrate that DHRA can reduce message number and communication traffic significantly, with the equal or even less tasks finish time compared with MinMin.
Guiyi Wei, Athanasios V., Vasilakos, Yao Zheng, Naixue Xiong	2010	The authors proposed in this paper [14], a pragmatic approximated solution with the accompanying two steps.	It is demonstrated that Nash equilibrium dependably exists if the resource allocation game has feasible solutions.

Conclusion

In this paper, it is concluded that cloud computing has decentralized architecture due to which load balancing, security are the major issues. The load unbalancing affects network performance in terms of certain parameters. In this review paper, techniques of load balancing are reviewed and analyzed in terms of various parameters

References

[1] M. D. Dikaiakos, D. Katsaros, P. Mehra, G. Pallis and A. Vakali, "Cloud computing: Distributed Internet computing for IT and scientific research", 2009, Internet Computing, IEEE, 13(5), 10-13,

[2] T. Surcel and F. Alecu, "Applications of Cloud Computing", 2008, International Conference of Science and Technology in the Context of the Sustainable Development, pp. 177-180

[3] Sumit Goyal, "Public vs Private vs Hybrid vs Community - Cloud Computing: A Critical Review", 2014, I.J. Computer Network and Information Security, 3, 20-29

[4] G. Lewis, "Basics about cloud computing", 2010, Software Engineering Institute Carniege Mellon University, Pittsburgh

[5] M.A. Vouk, "Cloud computing–issues, research and implementations", 2004, Journal of Computing and Information Technology, 16(4), 235-246

[6] Christian V, Rodrigo NC, Dileban K, Rajkumar B," Deadline-driven provisioning of resources for scientific applications in hybrid clouds with Aneka", 2012, Future Gener Comput Syst 28:58–65

[7] Guiyi W, Athanasios V, Vasilakos YZ, Naixue X," A game-theoretic method of fair resource allocation for cloud computing services", 2012, J Supercomput 54:252–269

[8] Anton B, Jemal A, Rajkumar B," Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing", 2012, Future Gener Comput Syst 28:755–768

[9] Huangke Chen, Xiaomin Zhu, Hui Guo, Jianghan Zhu, Xiao Qin, Jianhong Wu," Towards Energy-Efficient Scheduling for Real-Time Tasks under Uncertain Cloud Computing Environment", 2015, J Syst Softw 99:20–35

[10] Doulamis ND, Kokkinos P, Varvarigos E," Resource selection for tasks with time requirements using spectral clustering", 2014, IEEE Trans Comput 63(2):461–474 Vol. 63, No. 2

[11] Abdul Hameed, Alireza Khoshkbarforoushha, Rajiv Ranjan, Prem Prakash Jayaraman, Joanna Kolodziej, Pavan Balaji, Sherali Zeadally, Qutaibah Marwan Malluhi, Nikos Tziritas, Abhinav Vishnu, Samee U. Khan, Albert Zomaya," A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems", 2014, Computing 1–24

[12] Young Choon Lee, Albert Y. Zomaya," Energy efficient utilization of resources in cloud computing systems", 2012, J Supercomput60:268–280

[13] Zhanjie Wang, Xianxian Su," Dynamically hierarchical resource-allocation algorithm in cloud computing environment", 2015, J Supercomput

[14] Guiyi Wei, Athanasios V., Vasilakos, Yao Zheng, Naixue Xiong," A game-theoretic method of fair resource allocation for cloud computing services", 2010, J Supercomput 54: 252–269

