International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 7, ISSN No. 2455-2143, Pages 159-162
Published Online May - June 2016 in IJEAST (http://www.ijeast.com)

VARIOUS LOAD BALANCING TECHNIQUES AND CHALLENGES IN CLOUD COMPUTING

Brahm Prakash Dahiya
Ph. D. Scholar, CSE Department,
I.K. Gujral Punjab Technical University, Jalandhar
Punjab, India

Abstract— **Load balancing within the cloud computing atmosphere has a very important impact on the performance and outcomes. In the cloud storage, load balancing may be a key issue. It had consumed a lot of cost to keep up load data, since the system is just too large to timely disperse workload. Load balancing is main challenges in cloud computing that is work to distribute the dynamic work across multiple nodes to make sure that no single node is weak. It helps in optimum utilization of resources and thus in enhancing the performance of the system. Good load balancing makes cloud computing a more efficient and improves user satisfaction .A number of existing programming algorithms will maintain load balancing and supply higher ways through efficient job planning and resource allocation techniques furthermore. So as to achieve most profits with optimized load balancing algorithms, it's necessary to utilize resources with efficiency. This paper discusses a number of the existing load balancing algorithms in cloud computing and additionally their challenges.**

Keywords— Cloud computing; Load balancing; performance; programming algorithm.

I. INTRODUCTION

A Cloud computing is rising as a replacement paradigm of huge scale distributed computing. It has touched computing and knowledge away from desktop and portable PCs, into massive knowledge centers [1]. It provides the scalable IT resources like applications and services, in addition because the infrastructure on that they operate, over the web, on pay-per-use basis to regulate the capacity quickly and simply [2]. It helps to accommodate changes in demand and helps any organization in avoiding the Capital prices of software system and hardware. Thus, Cloud Computing could be a framework for facultative an appropriate on demand network access to a shared computing resources. These resources are often provisioned and de-provisioned quickly with lowest management effort or service supplier interaction [3]. This more helps in promoting accessibility. Attributable to the exponential growth of cloud computing, it has been widely adopted by the trade and there's a speedy enlargement in data-centers [4].

Recently, public cloud is formed accessible as a pay per usage model whereas non-public cloud is built with the infrastructure of the organization itself. Internet Services, Google Application Engine, and Microsoft Azure area unit samples of public cloud. The service provided by the general public cloud is understood as utility computing. As benefit, users will access this service “anytime, anywhere”, share information and collaborate a lot of simply, and keep information safely within the infrastructure. Though there is a unit risks involved with releasing information onto third party servers while not having the complete management of it. In cloud computing setting, the random arrival of tasks with random utilization of central processing unit service time necessities will load a selected resources heavily, whereas the opposite resources area unit idle or area unit less loaded [6]. Hence, resource management or load equalization is major difficult issue in cloud computing. Load equalization could be a methodology to distribute employment across multiple computers, or different resources over the network links to attain optimum resource utilization, maximize turnout, minimum reaction time, and avoid overload. This analysis work is predicated on simulation technique and it uses the cloud machine [5].

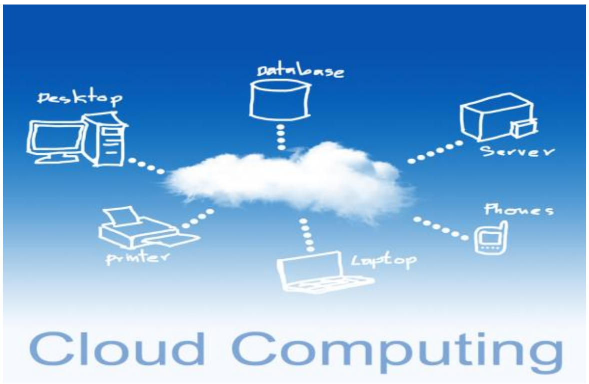

Fig. 1 cloud computing

The objective of this paper is study load balancing algorithms or techniques in cloud computing. The author explained load balancing in cloud computing in Section 2. Section 3 presents existing techniques in load balancing in cloud computing. Load balancing challenges in cloud computing explained in Section 4.Section 5 concludes the paper with directives of future work.

II. LOAD BALANCING IN CLOUD COMPUTING

Load balancing is one in all the most problems associated with cloud computing. The load will be a memory, C.P.U. capability, network or delay load. It's perpetually needed to share work load among the varied nodes of the distributed system to boost the resource utilization and for higher performance of the system. This will facilitate to avoid true wherever nodes area unit either heavily loaded or below loaded within the network. Load balancing is that the method of guaranteeing the equally distribution of work load on the pool of system node or processor in order that while not distressing, the running task is completed. The goals of load balancing [7] area unit to:

a. Improve the performance
b. Maintain system stability
c. Build fault tolerance system.
d. Accommodate future modification.

There are primarily 2 kinds of load balancing algorithms:

2.1 STATIC formula

In static formula the traffic is split equally among the servers. This formula needs previous information of system resources, in order that the choice of shifting of the load doesn't depend upon this state of system. Static formula is correct within the system that has low variation in load.

2.2 DYNAMIC formula

In dynamic formula the lightest server within the whole network or system is searched and most well-liked for balancing a load. For this real time communication with network is required which might increase the traffic within the system. Here current state of the system is employed to create selections to manage the load.

III. EXISTING TECHNIQIES IN LOAD BALANCING

In this section we have a tendency to discuss existing load balancing techniques in cloud computing. Here we have a tendency to classify load balancing algorithmic rule in 2 main sorts that square measure Static load balancing and Dynamic load balancing. In 2009, B Soto mayor et al [11] introduced a static well-known load balancing technique known as round Robin, within which all processes square measure divided amid all accessible processors. The allocation order of processes is maintained regionally that is freelance of the allocation from the remote processor. During this technique, the request is shipped to the node having least variety of connections, and since of this at some purpose of your time, some node is also heavily loaded and alternative stay idle [11]. This drawback is resolved by CLBDM. In 2010, S C. Wang et al. [12] bestowed a dynamic load balancing algorithmic rule known as load balancing Min-Min (LBMM) technique that relies on 3 level frameworks. This method uses timeserving Load balancing algorithmic rule that keep every node busy within the cloud while not considering execution time of node. Attributable to this it causes bottle neck in system. This drawback is resolved by LBMM 3 layer design. Initial layer request manager that is answerable for receiving task and distribution it to at least one service manager to second level. On receiving the request service manager divide it into subtasks. Afterward service manager can assign subtask to service node to execute task. In 2011, B. Radojevic et al [13] introduced a static load balancing algorithmic rule known as CLBDM (Central Load balancing decision Model). CLBDM is sweetening of the spherical Robin technique. This can be supported session shift at application layer. In spherical robin, request is shipped to the node having least variety of connections. RR is increased and in CLBDM, the calculation of the affiliation time between the consumer and also the node is finished and if the affiliation time goes higher than the brink then drawback is raised. If a retardant is arises, then the affiliation between the consumer and also the node is terminated and also the Task is forwarded to the more node mistreatment spherical Robin law. In 2011, L. Colb et al [14] introduced the Map Reduced based mostly Entity Resolution load balancing technique that relies on massive datasets. During this technique, 2 main tasks square measure done: Map task and cut back task that the author has represented. For mapping task, the half methodology is dead wherever the request entity is partitioned off into elements. So COMP methodology is employed to check the elements and at last similar entities square measure classified by cluster methodology and by mistreatment cut back task. Map task reads the entities in parallel and method them, in order that overloading of the task is reduced. In 2011, J Hu et al. [15] introduced a static programming strategy of load balancing on virtual machine resource. This method considers the historical information and additionally this state of system. Here, central computer hardware and resource monitor is employed. The programming controller checks the provision of resources to perform a task and assigns identical. Resource availableness details square measure collected by resource monitor. In 2011, J Al-Jaroodi et al. [16] projected a dynamic load balancing technique named DDFTP (Duel Direction Downloading algorithmic rule from FTP server). This could even be enforced for load balancing in cloud computing. In DDFTP, file of size m is split into m/2 partition and every node starts process the task. for instance if one server begins from zero to progressive order than alternative can start from m to prejudicial order severally from one another. As on downloading 2 consecutive blocks the task is taken into account as finished and appointed next task to server. Attributable to reduction in network communication between consumer and node network overhead is reduced. In 2012, K. Nishant et al [17] introduced a static load balancing technique known as hymenopterans Colony optimization. During this technique, An hymenopterans starts the movement because the request is initiated. This method uses the Ants behavior to gather info of cloud node to assign task to the actual node. During this technique, once the request is initiated, the hymenopterans and also the secretion starts the forward movement within the pathway from the “head” node.The hymenopterans moves in forward direction from a full node craving for next node to ascertain whether or not it's a full node or not. Currently if hymenopterans notice underneath loaded node stills it move in forward direction within the path. And if it finds the full node then it starts the backward movement to the last underneath loaded node it found antecedently. Within the algorithmic rule [15] if hymenopterans found the target node, hymenopterans can kill in order that it'll stop needless backward movement. In 2012, T. Yu Shanghai dialect et al. [18] introduced a dynamic load balancing technique known as Index Name Server to attenuate the info duplication and redundancy in system. This method works on integration of First State duplication and access purpose optimization. To calculate optimum choice purpose some parameter square measure defined: hash code of knowledge block to be downloaded, position of server having target block of knowledge, transition quality and most information measure. Another calculation parameter to search out weather affiliation will handle extra node or is at busy level B(a), B(b) or B(c). B(a) denote affiliation is extremely busy to handle new affiliation , B(b) denotes affiliation isn't busy and B(c) denotes affiliation is restricted and extra study required to understand additional regarding affiliation. In 2012, B. Mondal et al [19] have projected a load balancing technique known as random Hill climb supported soft computing for finding the optimization drawback. This method solves the matter with high likelihood. It’s a straightforward loop getting direction of accelerating price that is uphill. And this build minor amendment in to original assignment in step with some criteria designed. It contains 2 main criteria one is candidate generator to line attainable successor and also the alternative is analysis criteria that ranks every valid answer. This results in improved answer. In 2013, D. adult male et al [20] projected a Honey Bee Behavior impressed Load balancing[HBB-LB] technique that helps to realize even load balancing across virtual machine to maximize output. It considers the priority of task waiting in queue for execution in virtual machines. Afterward work load on VM calculated decides whether the system is full, underneath loaded or balanced and supported this VMs square measure classified. New in step with load on VM the task is scheduled on VMs. Task that is removed earlier. To search out the proper low loaded VM for current task, tasks that square measure removed earlier from over loaded VM square measure useful. Forager bee is employed as a Scout bee within the next steps.

IV. Load Balancing Challenges in cloud computing

1. Security

The main hurdle in the fast adoption of cloud is the security concerns of the customers [8]. Security issue has played the most important role in hindering Cloud computing acceptance. Various security issues, possible in cloud computing are: availability, integrity, confidentiality, data access, data segregation, privacy, recovery, accountability, multi-tenancy issues and so on. Solution to various cloud security issues vary through cryptography, particularly public key infrastructure (PKI), use of multiple cloud providers, standardization of APIs, improving virtual machines support and legal support [9].

2. Availability of Service

Since many systems have crashed on the cloud, like Amazon, so using only one Cloud Computing Service Provider (CCSP), services can result in a drawback as when a shutdown event happens on a cloud the service disappears and user cannot find that service. CCSP promises to provide infinite scalability for the customer but due to the fact that millions of users are now migrating to cloud computing so such promise is not fulfilled [8].

3. Third Party Dependence

Customers have no control over their own data as data is lost in the hands of the cloud computer service provider.

Although cloud computing has been wide adopted. analysis in cloud computing remains in its initial stages, and some scientific challenges stay unsolved by the scientific community, notably load balancing challenges [10].

•Automated service provisioning: A key feature of cloud computing is snap, resources may be allotted or free automatically. However then will we tend to use or unleash the resources of the cloud, by keeping constant performance as traditional systems and exploitation best resources?

•Virtual Machines Migration: With virtualization, a whole machine may be seen as a file or to unload a physical machine heavily loaded, it's attainable to maneuver a virtual machine between physical machines. The main objective is to distribute the load in a very datacenter or set of datacenters. However then will we tend to dynamically distribute the load?

•Energy Management: the advantages that advocate the adoption of the cloud is that the economy of scale. Energy saving could be a key purpose that permits a worldwide economy wherever a group of worldwide resources are going to be supported by reduced suppliers rather that each one has its own resources. However then will we tend to use a region of datacenter whereas keeping acceptable performance?

•Stored knowledge management: within the last decade knowledge keep across the network has associate exponential increase even for companies by outsourcing their knowledge storage or for people, the management of knowledge storage or for people, the management of knowledge storage becomes a serious challenge for cloud computing. However will we tend to distribute the info to the cloud for optimum storage of knowledge whereas maintaining quick access?

•Emergence of tiny knowledge centers for cloud computing: tiny datacenters may be a lot of helpful, cheaper and fewer energy client than massive datacenter. Tiny suppliers will deliver cloud computing services resulting in geo-diversity computing. Load reconciliation can become a haul on a worldwide scale to confirm associate adequate time interval with associate optimal distribution of resources.

V. CONCLUSION

In this paper we surveyed multiple algorithms for load balancing in cloud computing. We discussed and challenged that must be considered the most suitable and effective load balancing. Which wastages the storage resources and therefore as our future work this algorithm is modified in terms of storage utilization, heterogeneity, energy efficient metric also considered.

VI. REFERENCE

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “Above the Clouds: A Berkeley View of Cloud Computing”, EECS Department, University of California, Berkeley,Technical Report No., UCB/EECS-2009-28, pages 1-23, February 2009.

[2] R. W. Lucky, “Cloud computing”, IEEE Journal of Spectrum, Vol. 46, No. 5, May 2009, pages 27-45.

[3] M. D. Dikaiakos, G. Pallis, D. Katsa, P. Mehra, and A. Vakali, “Cloud Computing: Distributed Internet Computing for IT and Scientific Research”, IEEE Journal of Internet Computing, Vol. 13, No. 5, September/October 2009, pages 10-13.

[4] G. Pallis, “Cloud Computing: The New Frontier of Internet Computing”, IEEE Journal of Internet Computing, Vol. 14,No. 5, September/October 2010, pages 70-73.

[5] CloudSim: A Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services, The Cloud Computing and Distributed Systems (CLOUDS) Laboratory, University of Melbourne, (2011) available from: http://www.cloudbus.org/cloudsim.

[6] Livny, M.; Melman, M. (2011): Load Balancing in Homogeneous Broadcast Distributed Systems. Proceedings of the ACM Computer Network: Performance Symposium, pp. 47-55.

[7] D. Escalnte, Andrew J. Korty, “Cloud Services: Policy and Assessment”, Educause review July/August 2011.

[8] Han Qi, Abdullah Gani (2012). Research on Mobile Cloud Computing: Review, Trend *and Perspectives*. Second International Conference on Digital Information and Communication Technology and it's Applications (DICTAP), pp. 195-202.

[9] Bhushan Lal Sahu,Rajesh Tiwari, Journal of Advanced Research in Computer Science and Software Engineering 2(9) (2012) 33-37.

[10] A. Khiyaita, M. Zbakh, H. El Bakkali and Dafir El Kettani, “Load Balancing Cloud Computing: State of Art” , 9778-1-4673-1053-6/12/$31.00, 2012 IEEE.

[11] Sotomayor, B., RS. Montero, IM. Llorente, and I. Foster, "Virtual infrastructure management in private and hybrid clouds," in IEEE Internet Computing, Vol. 13, No. 5, pp: 14-22, 2009.

[12] Wang, S-C., K-Q. Yan, W-P. Liao and S-S. Wang, "Towards a load balancing in a three-level cloud computing network," in proc. 3rd International Conference on. Computer Science and Information Technology (ICCSIT), IEEE, Vol. 1, pp: 108-113, July 2010.

[13] Radojevic, B. and M. Zagar, "Analysis of issues with load balancing algorithms in hosted (cloud) environments." In proc. 34th International Convention on MIPRO, IEEE, 2011.

[14] Kolb, L., A. Thor, and E. Rahm, E, "Load Balancing for MapReduce based Entity Resolution," in proc. 28th International Conference on Data Engineering (ICDE), IEEE, pp: 618-629, 2012.

[15] J. Hu, J. Gu, G. Sun, and T. Zhao, “A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud computing Environment”, Third International Symposium on Parallel Architectures, Algorithms and Programming (PAAP), 2010.

[16] Al-Jaroodi, J. and N. Mohamed. "DDFTP: Dual-Direction FTP," in proc. 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), IEEE, pp:504-503, May 2011.

[17] Nishant, K. P. Sharma, V. Krishna, C. Gupta, KP. Singh N. Nitin and R. Rastogi, "Load Balancing of Nodes in Cloud Using Ant Colony Optimization." In proc. 14th International Conference on Computer Modeling and Simulation (UKSim), IEEE, pp: 3-8, March 2012.

[18] T-Y., W-T. Lee, Y-S. Lin, Y-S. Lin, H-L. Chan and J-S. Huang, "Dynamic load balancing mechanism based on cloud storage" in proc. Computing, Communications and Applications Conference (ComComAp), IEEE, pp: 102-106, January 2012.

[19] Brototi M, K. Dasgupta, P. Dutta, “Load Balancing in Cloud Computing using Stochastic Hill Climbing-A Soft Computing Approach”, in proc. 2nd International Conference on Computer, Communication, Control and Information Technology(C3IT)-2012.