

A Brief Review of Load Balancing Issue in Cloud Computing Environment

Kapil Dangl
Samrat Ashok Technology Institute
Vidisha(M.P)

Nirmal Gaud
Samrat Ashok Technology Institute
Vidisha(M.P)

ABSTRACT

The management of resources and request in cloud environment is big challenging task. The unbalanced scenario of resources and request raised situation of overloading and degraded the performance of cloud environment. Now a day's various authors used load balancing technique for the improvement of the cloud environment. For the balancing of load used load policy on the basis of static and dynamic nature. The static load balancing technique used CPU scheduling algorithm such as a round robin, first come first serve and shortest job first. All these technique is not schedule job in proper manner and the performance of cloud is degraded. Instead of these technique dynamic balancing used heuristic based technique such as ant colony optimization, particle swarm optimization and many more swarm based algorithm. now a day some authors used probability and time quantum based scheduling technique for cloud environment. In this paper present the review of load balancing technique for cloud environment.

Keywords

Cloud Computing, Load Balancing, policy of load balancing, swarm intelligence

1. INTRODUCTION

The applicability of cloud based services in every filed of enterprise such as big IT industry as well as small IT industry. The cloud based services reduces the cost of hardware and software for the establishment of IT industry. The cost casting and money-saving policy increase the popularity of cloud based services. The demand of cloud based services raised some issue related to the cloud computing environment [1]. The limited number of resource and unlimited number of request creates some overloading scenario in cloud environment. For the balancing of overloaded scenario used various job and task scheduling algorithm. the job and task scheduling algorithm basically based on the concept of CPU scheduling algorithm. the CPU scheduling algorithm faced a limited number of request, but here scenario of request is unlimited [2,3]. For the balancing of load cloud partition technique play efficient role. The cloud partition technique distributes the job according to the area of server and location. The distribution of area and server also faced a problem of overloading. Some authors give the probability based job scheduling technique over cloud computing. The probability based cloud job scheduling used theory of job entropy and

balance the job. For the improvement of load balancing technique swarm intelligence play major role. The family of swarm intelligence offering various algorithm for load balancing [4,5]. The swarm intelligence gives some memory based algorithm and some memory independent algorithm. the memory based algorithm performs very good scenario for load balancing technique. some authors used queuing theory and time quantum factor for the management of load over cloud computing. The queuing theory mange the task according to their time quantum factor. For the management of resource and allocation of resource the management of virtual machine is also important issue. The success of overall cloud environment depends on the management of virtual machines. Load balancing is done so that every virtual machine in the cloud system does the same amount of work throughout therefore increasing the throughput and minimizing the response time [8,9]. Load balancing is one of the important factors to heighten the working performance of the cloud service provider. Balancing the load of virtual machines uniformly means that anyone of the available machine is not idle or partially loaded while others are heavily loaded. One of the crucial issue of cloud computing is to divide the workload dynamically. The benefits of distributing the workload includes increased resource utilization ratio which further leads to enhancing the overall performance thereby achieving maximum client satisfaction [9]. The rest of paper organized in section 2. discuss load balancing policy. In section 3. Discuss related work in cloud computing and in section 4. discuss problem formulation of cloud computing and finally discuss conclusion and future work in section 5.

2. LOAD BALANCING

In cloud computing, the limited numbers of resource allocate the maximum consumer and handle maximum job profit. Due to this reason the process of cloud scheduling is overloaded and the performance of cloud computing is decrease. For the improvement of job scheduling and allocation of job required the process of load balancing technique [9,10]. Here discusses the load balancing selection algorithm. How to used different mode of load balancing algorithm for the process of job allocation and job selection. The process of categorization of algorithm depends on the process of work started and load balancing policy. The policy of load balancing divided into three section one is data sender, data receiver and combination of both.

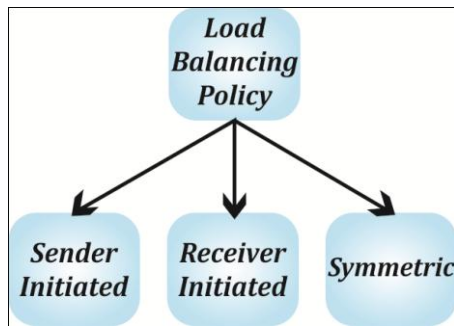


Figure 1 shows that distribution of load balancing policy [11].

The process of load balancing technique categorized into two sections one is static and other is dynamic.

Static: the method of static load reconciliation method work on the idea of permeable data, its don't depends on the current state data.

Dynamic: the construct of dynamic load equalization technique doesn't rely upon the receptive information; it acts on the idea of current conditions of method. currently a day's used four policies for load equalization technique these techniques used transfer policy, location policy, choice policy and data policy.

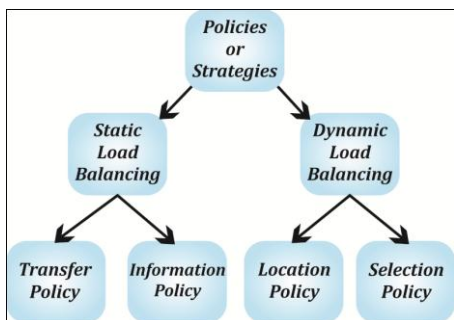


Figure 2: Shows that distribution of load balancing policy [12].

3. RELATED WORK

Here discusses some related work in the field of load balancing based on different function and balancing policy.

Yasser Alharbi and Kun rule Et al. [1] throughout this paper authors describe the strategy of implement is mentioned to the service of a Cloud Computing that works with computing resources and knowledge of research. in order that they get high performance efficiency overall of the implement result, completion time and increasing the production of cloud links. They mentioned the primarily useful observance throughout VMs placement to avoid uncalled-for migration to induce high efficiency for applications. The implementation results gift that their formula can update the jobs' completion time and maximize the absolutely applications performance.

Amir Nahir, Ariel Orda and Danny Raz Et al. [2] in line with the researchers throughout this implement, they have sure creating over one reproductions each{of each} job and effort each reproduction to a definite server. inside the premise of arrival of a replica to the very best of the sequence at its server, the latter signals the servers

holding replicas of that job, so on eliminate them from their sequences. To crucial the parameter configuration is very smitten by the system things, comprising the size of the system, load pattern, job interval, and inter-server delays. As fully completely different {completely different} things for several different systems, there is no single parameter configuration that is optimum to any or all systems. Consequently, to deploy the mentioned discovered, the system manager got to prepare a simulation-based analysis to look at the proper settings for the precise system. Since things may modification, system performance got to be constantly monitored thus on guage whether or not or not any of the parameter values got to be updated.

AlakaAnanth and K. Chandrasekaran Et al. [3] throughout this report, they analysis relating to cloud computing, it's future technology now-a-days, to emerged as a result providing computing resources as a service to the shoppers among the kind of platform, infrastructure and computer code. They together introduce new job coming up with technique practice the construct of theory, genetic rule and focuses on cooperative game supposed approach to provide Vilfredo Pareto optimum resolution. so the draw back was generated with practice Vilfredo Pareto optimality construct that uses NSGA II rule. Finally instrumental results gift that cooperative game supposed approach offers higher results once place next to non-cooperative coming up with.

Matthew Malensek, SangmiPallickara, and ShrideepPallickara Et al. [4] Reasearcher introduced to the analysis but proactive disk programming, backed by prognostic models and client-side coordination, can influence the output and responsiveness of a cluster in data-intensive computing environments. Proactively programming VM I/O bursts by foretelling and observation disk usage schemes gets long fairness across hosted VMs. Reason of I/O usage is host-specific and addicted to the band and behavior of VMs, offline approaches tend to be unsuitable. Minerva depends on on-line models trained with host-specific resource usage data to compared good capture usage patterns at a particular host. Time-series analysis of access schemes is important to identifying programming discontinuities. They live their framework with a representative MapReduce work on a 1,200-virtual-machine cluster, demonstrating a twenty initial improvement in completion time.

Nguyen KhacChien, Nguyen holmiumng Son and metal DacLoc Et al. [5] throughout this implantation, they made public that load exploit algorithm supported the technique of evaluating the highest of service time in heterogeneous cloud computing environments. designing cases of severally stages were taken into account once they describe formulas to calculate the standard virtual-core's method power. among the end of the analysis, results showed that the mentioned algorithm is easier. In four designing cases, the interval and quantity unit of measurement improved. Load exploit supports effectively utilize method resources, improve efficiency but it to boot causes the matter of greenhouse gas radiations and energy loss.

Jing Tai Piao and Jun Yan Et al. [6] inside the implement, they depicted strategy places the VMs on physical machines with observation of the network

things between the data storage and thus the physical machines. throughout the implement, their strategy has in addition monitored the concept throughout that instable network situation changed the data access behaviors and deteriorated the appliance performance, and prohibited this idea by migrating the VM to totally different physical machines. Their simulation result on Cloudsim 2.0 shown that the mentioned strategy may update the task completion time.

Kien Le, Jingru Zhang, JiandongMeng, economic expert Bianchini, YogeshJaluria and Thu D. Nguyen Et al. [7] in step with the authors, the analysis analysis the flexibleness of the many cooling ways that to handle load spikes, compares the behaviors of their changeable cost-aware terms to cost-unaware and static terms, and explains the results of the numerous parameter settings. They investigate the particular truth of cooling on total worth, and careful scan whether or not or not cooling-aware load distribution can cause worth savings and in addition analyzed transient cooling effects succeeding from abrupt, huge changes in info center lots of. They complete that pre-cooling is needed to forestall warming in these ideas. Finally, they have shown that intelligent placement and migration of load can therefore cause important worth savings. Thus, all electricity-related costs ought to be motorized to maximise and en-sure consistent worth savings.

KyongHoon Kim, Anton Beloglazov and RajkumarBuyya Et al. [8] They investigate and analysis a amount of your time Cloud service framework where each amount of your time service request is sculptural as RT-VM in resource brokers and put together analyzed power-aware provisioning of VMs for soft and arduous amount of your time Cloud services. The experimental results have delineate that info centers can deduct power usage and increase their profits using DVFS ways that. The mentioned adaptative ways that, Adaptive-DVFS and δ -Advanced-DVFS, generate higher profit with lower power usage in spite of the system load. Once the analyzed is completed, they commit to implement the mentioned framework during a) very Cloud broker and run real implement to validate the wise connectedness of the strategy.

Saurabh Kumar Garg, Adel NadjaranToosi, Srinivasa K. Gopalaiyengar and RajkumarBuyya Et al. [9] They explained at intervals the analysis programming mechanism and admission management that not only maximizes the resource utilization and profit. By full performance thought, it's live that the mentioned mechanism MWAP reduces the number of servers utilized by hour over other ways same as consolidation and migration with the negligible SLA violation. Finally they found that it's a necessity to recollect of varied varieties of SLAs beside applicable penalties and additionally the mixture of workloads for higher resource provisioning and utilization of datacenters. The mentioned mechanism provides substantial improvement over static server consolidation and reduces SLA violations.

George Kousiouris, TommasoCucinotta and Theodora Varvarigou Et al. [10] researchers main goal of this implement is analyzed the results of kind of crucial parameters on the efficiency of VMs and these parameters likes' allocation percentages, fundamental quantity programing selections and co-placement of

VMs. The implement learned from this progress is effective however 5 p.c, generic and could be at any time extended therefore on involved different identified significant concepts or coefficients. Finally in experiment, Through the utilization of this mechanism, Associate in Nursing infrastructure/Cloud provider can have a priori information of the interference. therefore they are ready to optimize the management of the physical resources.

DeepalJayasinghe, CaltonPu, Tamar Eilam, MalgorzataSteinder, Ian Whalley and disfunction Snible Et al. [11] inside the implement authors analyzed the way to boost accessibility and efficiency of services hosted on IaaS clouds. They formulate SCAVP (structural constraint-aware virtual machine placement) as Associate in Nursing improvement draw back (design big hierarchical draw back sizes) and gift its hardness. in order that they discuss a hierarchical approach composed of four approximation algorithms to resolve the placement draw back for larger draw back sizes. Finally on a simulated atmosphere, they illustrated the importance of mentioned system.

SriramKailasam, Nathan Gnanasambandam, JanakiramDharanipragada and Naveen Sharma Et al. [12] according to authors of this implement, targeting inter-cloud distributed analytics on workloads that square measure predominantly computations on documents and image a section that provides apriori visibility into the properties of the works in sequence. they have mentioned three flavors of schedulers that are not entirely effective in cloud-bursting of monumental works. By a thin pipe, they have shown the viability of optimizing and guaranteeing a large number of quality of service metrics on a hybrid cloud connected. the pc hardware method's extension mentioned to multiple work classes would manufacture the cloud explosive strategy applicable to a large number of environments like instructional computing domain.

ZeratulIzzahMohdYusoh and Maolin Tang Et al. [13] authors printed that the matter of composite SaaS resource management in Cloud data centre, specifically on its initial placement and resource improvement problems aiming at rising the SaaS performance supported its execution time additionally as minimizing the resource usage. They pictured problems ar developed as combinatorial improvement problems goal at rising the SaaS efficiency suppose execution time. They use two combinatorial improvement problems with paradigms of process algorithms. Finally experimental tried that the mentioned algorithms regularly generate a potential and satisfactory answer to any or all the check problems.

Abhishek Gupta, Laxmikant V. Kale, DejanMilojicic, Paolo Faraboschi and Susanne M. Balle Et al. [14] throughout this analysis, they address application-aware allocation of n VM instances to physical hosts from one pool. They vogue ANd implement associate degree HPC-aware hardware on high of Open-Stack calculate (Nova) and jointly incorporate it during a) very machine. Through varied optimizations (specifically topology and hardware-awareness, cross-VM interference accounting and application-aware consolidation) incontestable edges of up to thirty second increase in job turnout and efficiency improvement up to forty fifth whereas limiting the impact of disturbance to eight.

YueGao, Yanzhi Wang, Sandeep K. Gupta and MassoudPedram Et al. [15] throughout this paper, they the matter of world operation optimization in cloud computing from the angle of the cloud service provider (CSP). Their goal is to permit the CSP with a versatile programing and optimization framework that aims to at constant time maximize energy efficiency and meet all user deadlines. throughout this implement a pair of employment models area unit taken in cloud computing systems. Finally they discuss "Guided Migrate and Pack" (GMaP) as a unified programing and optimization framework for the CSP that addresses these issues in a {very} very holistic fashion. GMaP is in addition versatile in search space size and formula run time management. Implementation results prove that after GMaP is deployed for the CSP, world energy consumption costs updates by over twenty third once pairing thirty - fifty users, and over 16 PF once pairing sixty - 100 users.

Anton Beloglazov and RajkumarBuyya Et al. [16] throughout this analysis paper, researchers printed to grasp the implications of the net nature of the matter. in order that they evaluated the mentioned algorithms through intensive simulations on a large-scale implement setup mistreatment work traces from over cardinal Planet-Lab VMs. The implement's result have drawn native regression based formula combined with the MMT VM selection policy significantly outperforms various dynamic VM consolidation algorithms in relevance the ESV metric due to a well reduced level of SLA violations and so the vary of VM migrations. thus on choose the mentioned system in associate passing real Cloud infrastructure, they decide to implement it by extending a real-world Cloud platform, like Open Stack.

4. PROBLEM FORMULATION

Load balancing technique provides great utility of cloud computing over the sharing of resource and allocation of jobs. The cloud based services including hardware infrastructure and resource pooling such as platform as service and software as service. In all service application cloud computing enforced the process of virtual machine and data center. For the maximum utilization of virtual machine and data center cloud computing required job scheduling and load balancing. Job scheduling and load balancing is very critical issue in cloud computing environment. The distribution of load on cloud computing is basically a principle of distribution of task and job on dedicated resource such as datacenter and virtual machine. The datacenter and virtual machine is very important component of cloud computing. The proper utilization of resource and network cloud computing process required the process of load balancing technique. in cloud computing the load are distributed in the manner of distributed computing. The process of distributed computing allows the concept of load sharing and job sharing during the processing of task. Now a day's used various dynamic load balancing technique in cloud computing. The dynamic load balancing technique depends on the process of job transfer policy and sharing of information. The job transfer policy used heuristic based operation such as genetic algorithm, ant colony optimization and other classical set theory. In some heuristic based function provide limited search space for the processing of job. Some major point observe during the survey of load balancing technique.

1. Maximum time span for completion of job[11].
2. Overload condition of job scheduler[12].
3. Network congestion[13].
4. Garbage of resource[14].
5. The process state in overflow[16].

The above mention problem found in various research papers during survey of load balancing

5. CONCLUSION AND FUTURE SCOPE

Now a day the performance of cloud computing is major issue. For the improvement of the performance of the cloud environment used load balancing and job scheduling technique. in this paper present the review and issue of cloud computing in concern of load balancing. In this review mainly focus on balancing technique. for the balancing of policy used various technique such as cloud partition, static and dynamic load policy. In this paper also focus on probability based theory using time quantum principle. The time quantum policy used the process of sharing virtual machine. the sharing of virtual machine improved the utilization of cloud resource. Paper also focus on some basic parameter of cloud computing such as completion of request and job during the management of load. The process of balancing reduces the uses of resource a boost the cost of economy. The current load balancing technique only considering the proper allocation of job, not consider the time complexity and scalability of cloud resource. This paper mainly focusses on local resource sharing based on probability based function. In future proposed a load balancing policy based on time quantum and probability function.

6. REFERENCES

- [1] Yasser Alharbi and Kun Yang "Optimizing jobs' completion time in cloud systems during Virtual Machine Placement", International Conference on Big Data and Smart City, 2016, Pp 1-6.
- [2] Amir Nahir, Ariel Orda and Danny Raz "Replication-Based Load Balancing", IEEE, 2016, Pp 494-507.
- [3] AlakaAnanth and K. Chandrasekaran "Cooperative Game Theoretic Approach for Job Scheduling in Cloud Computing", Computing and Network Communications, 2015, Pp 147-156.
- [4] Matthew Malensek, SangmiPallickara, and ShrideepPallickara "Minerva: Proactive Disk Scheduling for QoS in Multitier, Multitenant Cloud Environments", IEEE, 2016, Pp 19-27.
- [5] Nguyen KhacChien, Nguyen Hong Son and Ho DacLoc "Load Balancing Algorithm Based on Estimating Finish Time of Services in Cloud Computing", ICACT, 2016, Pp 228-233.
- [6] Jing Tai Piao and Jun Yan "A Network-aware Virtual Machine Placement and Migration Approach in Cloud Computing", IEEE, 2010, Pp 87-92.
- [7] Kien Le, Jingru Zhang, JiandongMeng, Ricardo Bianchini, YogeshJaluria and Thu D. Nguyen

- “Reducing Electricity Cost Through Virtual Machine Placement in High Performance Computing Clouds”, ACM, 2011, Pp 1-12.
- [8] KyongHoon Kim, Anton Beloglazov and RajkumarBuyya “Power-Aware Provisioning of Virtual Machines for Real-Time Cloud Services”, John Wiley & Sons, Ltd., 2011, Pp 1-19.
- [9] Saurabh Kumar Garg, Adel NadjaranToosi, Srinivasa K. Gopalaiyengar and RajkumarBuyya “SLA-based virtual machine management for heterogeneous workloads in a cloud datacenter”, *Journal of Network and Computer Applications*, 2014, Pp 108-119.
- [10] George Kousiouris, TommasoCucinotta and Theodora Varvarigou “The effects of scheduling, workload type and consolidation scenarios on virtual machine performance and their prediction through optimized artificial neural networks”, *The Journal of Systems and Software*, 2011, Pp 1270-1291.
- [11] DeepalJayasinghe, CaltonPu, Tamar Eilam, MalgorzataSteinder, Ian Whalley and Ed Snible “Improving Performance and Availability of Services Hosted on IaaS Clouds with Structural Constraint-aware Virtual Machine Placement”, IEEE, 2011, Pp 72-79.
- [12] SriramKailasam, Nathan Gnanasambandam, JanakiramDharanipragada and Naveen Sharma “Optimizing Service Level Agreements for Autonomic Cloud Bursting Schedulers”, *Parallel Processing Workshops*, 2010, Pp 285-294.
- [13] ZeratulIzzahMohdYusoh and Maolin Tang “Composite SaaS Placement and Resource Optimization in Cloud Computing using Evolutionary Algorithms”, IEEE, 2012, Pp 590-597.
- [14] Abhishek Gupta, Laxmikant V. Kale, DejanMilojicic, Paolo Faraboschi and Susanne M. Balle “HPC-Aware VM Placement in Infrastructure Clouds”, IEEE, 2013, Pp 11-20.
- [15] YueGao, Yanzhi Wang, Sandeep K. Gupta and MassoudPedram “An Energy and Deadline Aware Resource Provisioning, Scheduling and Optimization Framework for Cloud Systems”, IEEE, 2013, Pp 1-10.
- [16] Anton Beloglazov and RajkumarBuyya “Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers”, John Wiley & Sons, Ltd., 2011, Pp 1-24.