International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

The Hybrid Algorithm for Load Balancing and Efficient Management of Resources in Cloud Computing

Shabir Ahmad Teli*1, Manoj Kumar Srivastava2

*1 M.Tech (Scholar), CSE Depatment Desh Bhagat University, Mandi Gobindgarh, Punjab, India

2CSE Department, Desh Bhagat University, Mandi Gobindgarh, Punjab, India

ABSTRACT

Article Info
Volume 7, Issue 4
Page Number: 341-349

Publication Issue :
July-August-2021

Article History
Accepted : 20 July 2021
Published : 27 July 2021

The field of parallel and distributed computing has considerably transformed in cloud computing and load balancer techniques play an important role in the next generation of cloud computing for the storage and access to the applications. Load balancing is a strategy for distributing workload among numerous computers or other resources through network links in order to optimise use of resources, maximum performance and minimise reaction time, and prevent overload. Load balance can be achieved by using resource efficiently to meet end user demands, and helps servers less computing, volunteer computing, software defining computing, etc. The recent development of technology is a key issue in cloud computing for the control of resources or load balancing. This report describes the conceptualization for the efficient management of resources, enhances the stability of web services and presents numerous approaches for load balance. Round Robin algorithms (SLBA), dynamic algorithms for load balancing (DLBA) and dynamic nature inspired algorithms for load balancing (NDLBA). DLBA and NDLBA are more efficient than SLBA according to experimental results. This article presents the future guidelines for cloud computing.

Keywords : Cloud computing, load balancing, simulation, CloudSim

The purpose of this work is to define qualitative components for cloud modelling and then carry out load balance analysis of algorithms based on these components

I. INTRODUCTION

The pay-per-use cloud computing paradigm offers on-demand access to a shared pool of customized computing resources (e.g., networks, servers, storage systems, applications, and services), which can be provided quickly and published with minimal administrative effort or inter-service provider interaction. As a result, new computer designs are emerging to fulfil market demands. Cloud computing is backed by a number of major companies, including Amazon, Yahoo, Sun, and Google[3]. The cloud-based deployment methodology makes use of assets from a variety of sources[3].

The Next Generation Cloud Computing

Distributed computing can be defined as the usage of several networked computers across a large geographical area or the world over the Internet. Cloud computing can act as a distributed computer that consists of a variety of virtual computer resources and allows them to be structured and extended quickly by supplying physical or virtual computers.

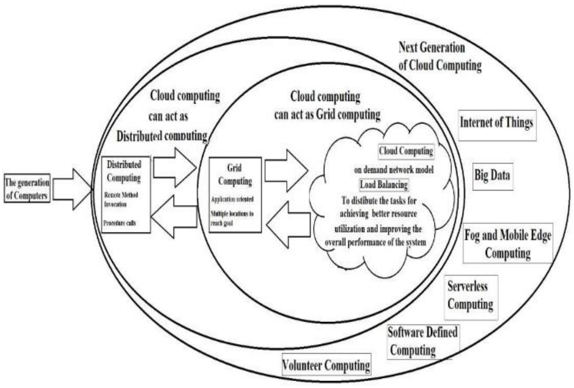

Figure 1: The Next Generation Cloud Computing

It provides highly scalable, redundant monitoring of resource utilisation, self-recovery of real-time programming models to re-balance assignments if required [5]. Cloud computing is also used to coordinate the resource sharing and resolution of problems in dynamic, multi-agency virtual organisations. A grid is a heterogeneous decentralised system in which multiple organisations' resources do not exercise ultimate control of these resources. In this circumstance, management of resources is subject to various and contradictory administrative organisational policies [6].

Load Balancing in Next generation Cloud computing

Load balancing can be achieved by efficient utilization of resources and enhancements in the implementation of the technology.. The following generation of cloud-dynamics includes multi-service concepts, which include a unique multimedia task for each servers cluster and which each client requires a special multi-media carrier at a certain time[15]. A charge balancing machine, which receives work from multiple sites and is then delivered to the data centre, is utilised to accomplish effective load distribution[16].

Cloud innovation and the rise of the sensors and gadgets bring people, units and the corresponding computer closer together[33]. The Internet of Things (IoT) is the next generation technology, with trillions of intelligent items talking to others to make lives even more convenient. IoT is based on the community of wireless sensors (WSN). Zigbee is one of the most prominent WSN protocols (Figure 2).

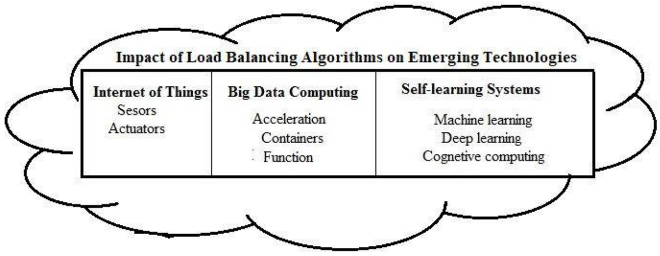

Figure 2: Impact on Emerging Technologies [33]

Load balancing data processing involves using in-infrastructure sensors which encompass transportation, communications, buildings, healthcare and various user devices, wearables, and home devices, which have led to the coming Internet of Things (IoT) [18]. Data processing with lock balancing includes Cloud load balancing with IoT enhances accuracy and efficiency and reduces human intervention as well as effective data access[32].

The fundamental objective of the cloud-based load balancing issue is to map the responsibilities for setting VMs efficiently[26]. Dynamic load balance inspired by the nature continues to provide a global problems of optimisation due to the following questions: heterogeneous cloud structures, computer resources and applications' quality of service (QoS) requirements[3,4]. As seen below, the main motivation behind cloud load balance solutions and new technologies is:

- Designing a system model to assign tasks to new technologies in cloud computing[26]..

- The problem of optimization that can be adapted to the cloud remains to produce nature-inspired dynamic load balance[27].

II. RELATED WORKS

Load balancing is a strategy for distributing workload among numerous computers or other resources through network links in order to optimise use of resources, maximum performance and minimise reaction time, and prevent overload.

This section discusses the efficient research carried out by the competent load balancing investigators..

A. Sharma, S. Verma et al. [7] Different load balancing methodologies were examined which play a major part in planning and load balancing based on the knowledge of many performance assessments that will boost the performance of parallel work calculations. They have made it clear that balancing weights are on various machines to carry out the operation quickly.

R. M. Singh, S. Paul, A. Kumar et al. [8] Proposed an effective load balancing architecture. The author shows that the task is about data processing, accessing software and storage tools. With a service level arrangement and required sources, the efficient software classifys process reliably.

The cloud schedule is illuminated by S. Nagadevi and K. Satyapriya, D. Malathy et al. [9], so that the system can reach high response times and lower error rates. In parallel processing, each process is assigned to one of the processing servers available.

W. Huai, Z. Qian, X. Li, G. Luo, and S. Lu et al. [10] Proposed heuristic load balancing scheduling solution in parallel computing. The meta-heuristic task planning provides an excellent amplification by practising the data for the results. Heuristic approaches can function either statically or dynamically. Cyclic methods such as round robin are also concerned with FIFO ways for carrying out such scheduling duties.

D. Kliazovich, S. T. Arzo, F. Granelli, P. Bouvry, S. U. Khan et al.[11] Works with defined periods of time on resource management for each task. As resources are one of the most important concerns in the actual load balance. If resources are not accessible, machines cannot handle heavy loads that make machines difficult to do jobs. The processes are then queued till the chance of the jobs is reached.

S. H. Jang, T. Y. Kim, J. K. Kim , J. S. L. School and S. Sudhir [12][13] The work on the balancing charge is also a heuristic approach to scheduling, depending on the completion time of the process, to schedule the following available apparatus.

A. mantri, S. Nandi, G. Kumar, S. Kumar et al. [14] They have suggested their strategy in a minimum interval to complete the jobs. The MTT and the MTT are both based on heuristic procedures in which assigned processor processes function in an efficient way and do jobs in less execution times. Parallel task development is a way for assigning procedures to various processors. The parallel work arrangement uses several types of algorithms and recycling strategies to eliminate delayed tasks. Nowadays, several scheduling approaches and methods have been applied to shorten work time. The NP hard job arrangement is the best possible technique for some task arrangements and other approaches, and no-one can determine approximately.

Li et al. [21] Proposed a VM Energy Consolidation (ECVMC) Method based on controls on the potential for several types of resources to be overloaded. The methodology was proposed for VM migration. Multiple algorithms for several phases of VM dynamic consolidation were applied and the mapping relationship between PMs and VMs was used to establish a viable solution for this approach. These results have resulted in a better VM consolidation performance with efficient performance and maximum use and QoS guaranteed.

Li et al. [22] A Bayesian network-based VM consolidation for live VM migration based on energy and QoS parameters such as dynamic workload, CPU usage and the number of VM migrations has also been given. This strategy reduces energy consumption, prevents further inconsequential VM migrations, and enhances QoS and reduces unnecessary consumption of resources. These two strategies only evaluate the use of the CPU and avoid other resources.

Ranjbari and Torkestani [23] Introduced VM energy efficiency and SLA guarantee automatically-based learning mechanism. This method takes into account changes in the user requirement to predict the excessive PM and shut down idle servers to decrease energy consumption. But only the host overload is observed when the underused hosts are not efficiently recognised..

Mahdhi and Mezni [24] Proposed a predictive VM consolidated strategy using the Kernel Density Estimate technology to estimate the future traffic and resources of VM migration. Proposed This technique minimises energy use and guarantees QoS. This is a technique which takes into account CPU, RAM and VM consolidation storage. However, this model has tremendous traffic and safety issues. In the same way, The VM consolidation strategy used by Abdelsamea et al. [25] has been based on the CPU, RAM, and network bandwidth hybrid regression algorithms. This approach presented two models for effectively detecting host overload and underload with low energy usage and guaranteed QoS. However, the regression algorithm is used to standardise the expected use with a fixed point which can reduce efficiency. The research shows that most VM consolidation strategies solely use the CPU for overload and load situations detection. This strategy is easier to calculate but there are inconveniences when relying on one single aim in effective forecast to load situations. The use of inefficient thresholding techniques to determine overload thresholds is also an important issue. This challenge has been considered as the rationale for designing the suggested Host overload and underload monitoring MOSLO based multiple linear regression algorithms.

ALGORITHMS

As a cloud technology, solution vendor one of the ultimate aims is to design a suitable algorithm and to understand how to correctly deploy Cloud computing services for efficient and reliable computing. As a service, virtual machines (VMs) can be accessed on demand. In cloud computing, load-balancing strategies are used to assign virtual machines on request. This is because the load-balancing approach is so critical in choosing which VM should be allocated based on the user's demands. It is possible that a number of requests are received during service provision and that certain requestors need to stand in queue, but they can send requests to other service providers.

I. Simplest Algorithm of Round Robin

When assigning jobs, it doesn't take into account a node's historical load state. As part of the task distribution process, the round-robin scheduling method is employed. As a result, the first node is chosen at random, and all subsequent jobs are distributed to the remaining nodes in an orderly method [15]. In this process, the virtual machines are picked at random from a pool of candidates. The datacenter controller rotates requests among a list of virtual machines. The Data Center controller assigns the initial request to a random VM in the group, and assigns subsequent requests in a circular pattern. A request is moved to the bottom of the queue once the VM has been assigned it [13]. Load distribution with round-robin is by far the simplest approach. As a result, it is frequently the primary choice for creating a simple scheduler. The fact that all that is necessary is a list of nodes is one of the reasons for its simplicity.

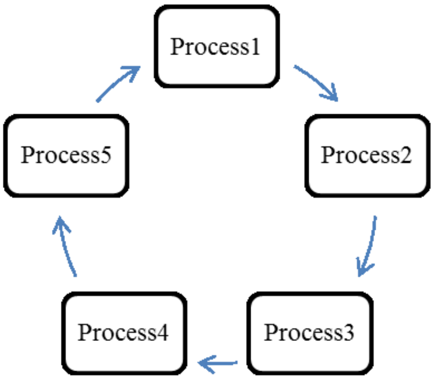

Fig. 3 : Processing of Round Robin

The round robin algorithm is as follows:

Step 1: A VM index is maintained by the Round Robin VM load Balancer. All VMs start with a zero allocation.

Step 2. a. The user requests/cloudlets are received by the data centre controller.

b. Requests are distributed to VMs in a cyclical fashion.

c. The round-robin Virtual machine balancer of load will distribute the time for user requests to be executed.

The VMs are unallocated by the Round Robin VM Load balancer after the execution of cloudlets.

Step 4: Any requests queries (new /pending/waiting) are detected by the controller of data centre.

Step 5: Carry on from step 2.

III. Dynamic Throttled Algorithm

In real time, a throttled balancer is a load balancer method. As part of this throttled algorithm, the client first requests the balancer to select the best VM performing the required activity. The procedure begins with a list of all the virtual machines. To speed up the lookup process, every row is separately indexed. According to the client's request, the load balancer assigns a VM if a match is found based on size and accessibility. Load balancer returns -1 if no VM satisfies the condition, and the request is queuing. The throttled algorithm:

Step 1: Throttled VM an index database of virtual machines (VMs) is maintained by the Load Balancer to keep track of their status (available/busy). Virtual machines are offered from the beginning..

Step 2: A new request is received by the DC Controller.

Step 3: The Datacentre Supervisor asks the Throttled Virtual machine Load Balancer for the next allocation.

Step 4: It scans the table from top to bottom until it discovers the first available VM or until the table is scanned fully.

i) Data Center Receiver takes an alert from Throttled Virtual Machine Load Balancer if the VM ID is detected.

ii) A request is routed to that VM by the Data Center Controller.

iii) The Data Center Controller notifies the Throttled VM Load Balancer of the new allocation

iv) The bottle necked VM Balancer updates the database of allocations.

If Throttled VM Balancer cannot find it, it returns -1.

ii) The request is queued by the Data Center Controller.

Step5. The Throttled VM Balancer is notified that the VM has been de-allocated when the VM finishes the request and the Datacenter Controller receives the response cloudlet.

Step 6: The Datacenter Controller examines the queue for any pending requests. If there are any, the process continues from step 3.

Step 7: Go back to step 2.

PROJECTED FUSION ALGORITHM(Hybrid Algo)

This study proposes a hybrid technique for vm load balancing that combines aspects from the Round Robin and Throttled algorithms. When assigning a new work to a virtual machine, round robin is a typical load balancing approach that ignores the current load state of the virtual machine. The throttled method served as motivation for implementing the current load status. In the hybrid approach, the current load situation of vm was factored while making judgments about job distribution to virtual machines. The proposed hybrid approach performed effectively when the prediction accuracy per request was varied.

With the hybrid technique proposed, an arbitrary VM is first chosen and checked for availability if a cloudlet is received to run on VMs. Cloudlet is assigned to it if available. Control will then move through the VM pool in a circular method until a cloud assignment VM is available. After cloud allocation, the next comparison occurs in this circular manner at the next VM.

This hybrid method derived the notion of cyclic ways to assign Virtual machines to cloudlets from the algorithm of Round Robin. And each step was motivated by a throttled algorithm-based availability check..

The algorithm proposed is the following:

Step1. Hybrid Load Balancer stores a VM index and VM status (complete/disponible). Every VM is 0 at the beginning.

Step2. a. User requests/cloudlets are being sent to the data centre controller.

b. Circular inspection of all VMs.

c. The VMs are assigned requests according to their VM queue status: if status is available: the cloudlet is assigned to the VM.

For that cloudlet, other next VM availability is checked.

d. The quantum time is allocated by the Hybrid Load Balancer for the effectiveness of the user request. Following the processing of cloudlets, the Hybrid Traffic Balancer de-allocates the VMs.

Step3. The VMs are de-allocated with the Hybrid Load Balancer after the processing of cloudlets.

Step 4: The Data Center Supervisor scans the queue for any new /pending/waiting requests.

Step 5: Carry on from step 2.

IV. RESULT ANALYSIS

The hybrid strategy generated better performance in terms of responsiveness, datacentre demand serving time, and data centre activity cycle time when compared to the results of the Round robin algorithm, Throttled algorithm, and equitably dispersed current execution algorithm separately. The presented hybrid approach performed effectively when the data size per request was the same so when the size of data per request was different. In the future, a comparison of these four methods will be offered.

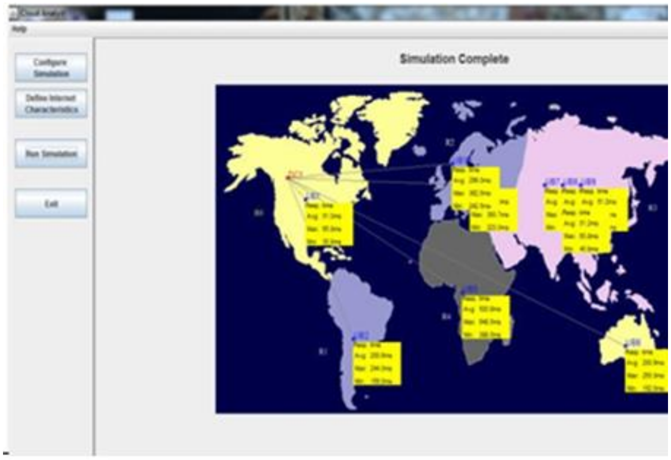

Fig.4. For the same data quantity per request, simulation statistics for a hybrid load balancing algorithm

	Hybrid	Round Robin	Throttled
Overall Response	202.55	202.57	
Data Center	1.32	1.33	
Data Center	3.87	8.38	4.26
Serving Time for DC1 Data	1.27	1.28	

Serving Time for DC2 Data	1.59	1.65	1.60
Serving Time for DC1 Data	2.31	2.88	
Serving Time for DC2 Data	3.87	8.38	4.26

Table1. Algorithms for load balancing requests with different data sizes

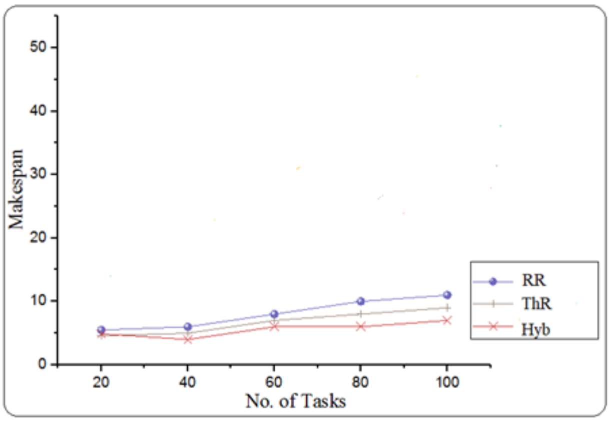

Figure 5 : Makespan of the algorithm

V. CONCLUSION

Load balancing is amongst the most critical features of cloud computing. This research focuses on the effects of load balancing on cloud computing's dependable new technologies. In addition, proper load balancing on the web with IoT, Data Science, and computational modeling can approach the current levels of secure cloud computing. The Hybrid algorithm has been shown to be more efficient and advantageous for cloud technology than the RR and Throttled algorithms. To maximize overall performance and resource usage, jobs are distributed across VMs. This paper discussed a variety of load balancing techniques, both simple and complicated, as well as their types. The need to develop fully independent code snippets for dynamic load balancing in the long term will allow for greater resource utilization, less machining, improved imbalance, efficient work migrations, and a longer cloud computing time frame in the next generation.

VI. REFERENCES

[1]. Buyya, R., Ranjan, R., Calheiros, R. N. (2009) „Modeling and simulation of scalable Cloud computing environments and the CloudSim toolkit: Challenges and opportunities." International Conference on High Performance Computing & Simulation', HPCS '09., pp.1-11.

[2]. Y. Mansouri, A. N. Toosi, R. Buyya, Data Storage Management in Cloud Environments: Taxonomy, Survey, and Future Directions, ACM Computing Surveys (2017) 1-51.

[3]. D. M. Shila, W. Shen, Y. Cheng, X. Tian, a. X. S. Shen, AMCloud: Toward a Secure Autonomic Mobile Ad Hoc Cloud Computing System, IEEE Wireless Communications 24 (2) (2017) 74-81.

[4]. D. Petcu, G. Macariu, S. Panica, C. Crciun, Portable Cloud Applications: From Theory to Practice, Future Generation Computer Systems 29 (6) (2013) 1417-1430.

[5]. Gill, Sukhpal Singh, and Rajkumar Buyya. "SECURE: Self-Protection Approach in Cloud Resource Management." IEEE Cloud Computing 5, no. 1 (2018): 60-72.

[6]. Haiying Shen, "RIAL: Resource Intensity Aware Load Balancing in Clouds", Volume: PP, Issue: 99, 2017, DOI 10.1109/TCC.2017.2737628, IEEE Transactions on Cloud Computing.

[7]. A. Sharma, S. Verma, "A Survey Report on Load Balancing Algorithm in Grid Computing Environment", Int. J. Adv. Engg. Res.

Studies/IV/II/Jan.-March 128, pp:132 -138, 2015.

[8]. R. M. Singh, S. Paul, A. Kumar, "Task scheduling in cloud computing: Review." International Journal of Computer Science and Information Technologies 5.6, pp: 7940-7944, 2014.

[9]. S. Nagadevi, K. Satyapriya, Dr.D.Malathy, “A Survey on Economic Cloud Schedulers for Optimized Task Scheduling”, International Journal of Advanced Engineering Technology, 2013.

[10]. W. Huai, Z. Qian, X. Li, G. Luo, and S. Lu, “Energy Aware Task Scheduling in Data Centers, Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications”, volume: 4, number: 2, pp. 18-38, 2013.

[11]. D. Kliazovich, S. T. Arzo, F. Granelli, P. Bouvry, S. U. Khan, “e-STAB: Energy-Efficient Scheduling for Cloud Computing Applications with Traffic Load Balancing”, IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing e-STAB, 2013.

[12]. S. H. Jang, T. Y. Kim, J. K. Kim , J. S. L. School, “The Study of Genetic Algorithm-based Task Scheduling for Cloud Computing”, International Journal of Control and Automation Vol. 5, No. 4, December, 2012

[13]. S. Sudhir, “Survey on Scheduling Issues in Cloud Computing”, Procedia Engineering, Elsevier, 2012.

[14]. A. mantri, S. Nandi, G. Kumar, S. Kumar, High Performance Architecture and Grid Computing Computing, International Conference HPAGC 2011, Chandigarh, India, July, Proceedings, 2011.

[15]. Ashkan Paya, Dan C Marinescu, "Energy-aware Load Balancing and Application Scaling for the Cloud Ecosystem", IEEE Transactions on Cloud Computing, Vol: 5, Issue: 1, Jan.-March 1, 2017, pp. 1-14.

[16]. Calheiros R. N., Ranjan, R., Rose, C. A. F. D., Buyya, R. (2009) „CloudSim: a novel framework for modeling and simulation of cloud computing infrastructures and services‟ Computing Research Repository, vol. Abs/0903.2525.

[17]. Sharma S, Singh S, Sharma M. Performance analysis of load balancing algorithms. World Academy of Science, Engineering and Technology. 2008 Feb 21; 38(3):269-72.

[18]. Hou L, Zhao S, Xiong X, Zheng K, Chatzimisios P, Hossain MS, Xiang W. Internet of things cloud: Architecture and implementation. IEEE Communications Magazine. 2016 Dec; 54(12):32-9.

[19]. X. Xu, X. Zhao, A Framework for Privacy-Aware Computing on Hybrid Clouds with Mixed-Sensitivity Data, in: 17th IEEE International Conference on High Performance Computing and Communications, 7th IEEE International Symposium on Cyberspace Safety and Security, and 12th IEEE International Conference on Embedded Software and Systems, 2015, pp. 1344-1349.

[20]. A. N. Toosi, R. O. Sinnott, R. Buyya, Resource Provisioning for Data-intensive Applications with Deadline Constraints on Hybrid Clouds Using Aneka, Future Generation Computer Systems

[21]. Li, Z., Yan, C., Yu, L., & Yu, X. (2018). “Energy-aware and multi-resource overload probability constraint-based virtual machine dynamic consolidation method.” Future Generation Computer Systems, 80, 139-156.

[22]. Li, Z., Yan, C., Yu, X., & Yu, N. (2017). “Bayesian network-based virtual machines consolidation method.” Future Generation Computer Systems, 69, 75-87.

[23]. Ranjbari, M., & Torkestani, J. A. (2018). “A learning automata-based algorithm for energy and SLA efficient consolidation of virtual machines in cloud data centers.” Journal of Parallel and Distributed Computing, 113, 55-62.
[24]. Mahdhi, T., & Mezni, H. (2018). “A prediction-Based VM consolidation approach in IaaS Cloud Data Centers.” Journal of Systems and Software, 146, 263-285.
[25]. Abdelsamea, A., El-Moursy, A. A., Hemayed, E. E., & Eldeeb, H. (2017). “Virtual machine consolidation enhancement using hybrid regression algorithms.” Egyptian Informatics Journal, 18(3), 161-170.
[26]. Haiying Shen, “RIAL: Resource Intensity Aware Load Balancing in Clouds”, Volume: PP, Issue: 99, 2017, DOI 10.1109/TCC.2017.2737628, IEEE Transactions on Cloud Computing.
[27]. Wang SC, Yan KQ, Liao WP, Wang SS. Towards a load balancing in a three-level cloud computing network. In 2010 3rd International Conference on Computer Science and Information Technology 2010 Jul 9 (Vol. 1, pp. 108-113). IEEE
[28]. R. Trapero, J. Modic, M. Stopar, A. Taha, N. Suri, A Novel Approach to Manage Cloud Security SLA Incidents, Future Generation Computer Systems 72 (2017) 193 - 205.
[29]. T. A. Nguyen, D. S. Kim, J. S. Park, Availability Modeling and Analysis of a Data Center for Disaster Tolerance, Future Generation Computer Systems 56 (2016) 27 -50.
[30]. X. Yuan, G. Min, L. T. Yang, Y. Ding, Q. Fang, A Game Theory-based Dynamic Resource Allocation Strategy in Geo-distributed Datacenter Clouds, Future Generation Computer Systems 76 (2017) , 63 - 72.
[31]. A. Hameed, A. Khoshkbarforoushha, R. Ranjan, P. P. Jayaraman, J. Kolodziej, P. Balaji, S. Zeadally, Q. M. Malluhi, N. Tziritas, A. Vishnu, S. U. Khan, A. Zomaya, A Survey and Taxonomy on Energy E cient Resource Allocation Techniques for Cloud Computing Systems, Computing 98 (7) (2016) 751-774.
[32]. J. Lopez, R. Rios, F. Bao, G. Wang, Evolving Privacy: From Sensors to the Internet of Things, Future Generation Computer Systems 75 (2017) 46 -57.
[33]. Blesson Varghese and Rajkumar Buyya, Next Generation Cloud Computing: New Trends and Research Directions," Future Generation Computer Systems, ISSN: 0167-739X, Elsevier Press, Amsterdam, The Netherlands, 2017 (in press).

Cite this article as :

Shabir Ahmad Teli, Manoj Kumar Srivastava, "The Hybrid Algorithm for Load Balancing and Efficient Management of Resources in Cloud Computing", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 4, pp. 341-349, July-August 2021. Journal URL : https://ijsrcseit.com/CSEIT217477