Efficient load balancing by optimized flexi max-min algorithm

P Praveen Kumar¹ and V Kanchana Devi*²

¹Student, Master of Computer Applications, Vellore Institute of Technology - Chennai Campus, Tamilnadu, India.
²Associate Professor, School of Computing Science and Engineering, Vellore Institute of Technology - Chennai Campus, Tamilnadu, India.

Email: *kanchanadevi@vit.ac.in

Abstract. Load Balancing is required to distribute tasks across Virtual Machines on one hand and on the other hand in a cloud Environment. Here, a new task scheduling algorithm: ‘Optimized Flexi Max-Min Scheduling Algorithm’ is proposed. The algorithm maintains a data structure which is modeled after a Binary Search Tree (BST) for estimations, enhanced searching, task allocation, and migration of tasks. Cloud-Sim is used to model and simulate the cloud computing environment in order to obtain simulated data. The result of this experiment shows that the proposed algorithm outperforms the traditional Max-Min task scheduling algorithms.

Keywords: Load Balancing, Virtual Machines, Cloud Environment, Max-Min Algorithm, Scheduling Algorithm, Encryption, Decryption

1. Introduction

The progression of innovation and how organizations and associations run in this cloud time are clear proof that distributed computing is the future stage for organizations and individual processing needs. Distributed computing has increased huge consideration and prominence in the IT segment since dominant part of associations, organizations and people can't purchase assets like equipment and programming. Because of the significant expense engaged with getting them. Distributed computing is progressively turning into an ideal stage for associations and people since; it has had the option to bring down the obstructions of the significant expenses related with giving the imperative programming and equipment. Distributed Computing has changed over capacity, for instance, into a 'pay more only as costs arise administration'. This has empowered people and associations not to require the acquisition of their own assets (equipment and programming) or paying for upkeep cost. The effect of distributed computing can be felt in our day by day exercises. It has affected our lives in a wide scope of developments which have totally changed how figuring administrations are created, estimated and conveyed.
2. Objective
Associations and people are driving on limiting cost, satisfying clients' need for asset sharing and information accessibility on request. All these are bolstered by distributed computing. This pivotal complex stage is progressively turning into the client/client decision in the Information Technology (IT) advertise as a result of its multi-perserverence, adaptability, accessibility, security, proficient capacity and asset sharing stage. Given the huge market section and focused on development of distributed computing, it is required to turn into a market head in the capacity business sooner rather than later. The paper centres around the key issues with cloud situations which incorporates the asset portion and cloudlets booking. When planning cloudlet in distributed computing condition, diverse cloudlets should be executed in equal by the accessible assets so as to live up to shoppers' desires and to accomplish better execution by limiting make length and expanding asset usage. Max-Min endeavors to handle these issues, yet creates higher length and poor asset usage. Also, the number of assignments with high culmination time is more than errands with low completion.

This paper tends to consider the issues of Max-Min calculation which incorporate the creation of higher make length; poor asset use and unbalance load by building up an appropriate methodology call Modified Max-Min (MMax-Min) which is equipped for choosing and relegate either task with most extreme fulfillment time or errand with least culmination time to an asset for improved effectiveness. What's more interesting is a nitty-gritty calculation which is an alteration of the customary Max-Min calculation that allot both errand with most extreme and least fulfillment time to an asset for improved effectiveness.

For the reason for benchmarking the outcomes, the investigation additionally executed the proposed component in a cloud sim, the outcomes were thought about, which shows that MMax-Min is proficient in creating less make range and adjusts load viably. The examination tries to analyze and assess the exhibition of the proposed Modified Max-Min calculation with other existing cloud booking calculations, for example, Max-Min and Round Robin. Cloud-Sim toolbox, a recreation domain in distributed computing was utilized to try and discover the presentation of these calculations according to task planning for the cloud. The main issue of this examination is to gauge the make span and adjustment in utilizing these calculations to plan an offered assignment to an asset.

3. Related Work
The authors has identified the drawbacks of the formal research formulation to obtain solution in cloud computing. Later, many related topics, i.e., virtual machine migration techniques, forecasting techniques, stability and availability of the several nodes are highlighted. The existing algorithms are surveyed with top-do-down approach. Then it has been classified as 11 different models. [1]. Optimized use of resources with modified swarm optimization based task scheduling algorithms, in this paper, the authors adopted a modified algorithm to obtain maximum benefit from the available resources with optimized usage of resources is ensured and for this purpose scheduling plays a major role in cloud computing. The particle swarm optimization in the proposed algorithm, triggers annealing, particle swarm optimization and it generally utilizes less execution time [2].

In paper [3], authors proposed an effective and security enabled privacy-preserving approach for data of limited resource mobile devices in cloud computing. The adopted approach deploys probability based public-key encryption technique for the process of encrypting the available data and intimates top ranked keywords mainly to search over the encrypted document to retrieve the specified files from the cloud environment. The goal was to adopt an effective system for encryption of retrieved data without losing the seclusion of available data. Later, prompted index keyword searching highly increases the system usability by this kind of ranking approach on similar score for searched results, returns similar files instead of all files, and verifies accuracy of file retrieval. The result shows, processing time, data privacy of document, overheads due to communication are decreased. With
proper security and performance analysis of the proposed approach, the authors proved that the new technique is secure and effective[3].

The goals of the paper [4] minimizing turnaround time of the project and improvise priority scheduling by incorporating proper scheduling technique. The authors included a two-tier backfilling approach which develops suitable conservative backfilling approach with priority terms and slack-time. Firstly, the Two-Tier Strict Backfilling (2TSB) does not supports the pre-emption method in the task waiting queue of CPU. Secondly, pre-emption method is allowed in Two-Tier Flexible Backfilling (2TFB) approach with slack factor (SF) which has two different versions: 2TFB, 2TFB-SF in the proposed approach.

In former one, a new task is pre-empted before the waiting tasks but not waiting processes, where the 2TFB-SF allows pre-emption technique in task waiting queue and process waiting queue. Two-Tier Priority Backfilling (2TPB) approach takes the priority of the task so that the task which obtained high-priority are pre-empted than the one which obtain low-priority task. The results proves that the comparison with 2TSB. The 2TPB approach reduces quiet less turn-around time with high-priority processes by 25% more. In paper [6] Cloud Storage has been introduced, which describes the key technologies available behind the cloud computing technology. The Cloud Storage, management techniques regarding cloud computing, various types of services provided by cloud, driving forces behind the cloud computing and explanation of cloud storage, the attributes and disadvantages of cloud storage, which then concluded by specifying some of the issues to be identified by the cloud storage providers.

This paper incorporates the scheduling problem for the large-scale software inspired from real-world problem, categorized by a large number of homogeneous and heterogeneous bags-of-jobs which are considered as the important sources of bottlenecks in the scheduled task but opens-up a great potential for optimization of the same algorithm. The scheduling problem later is designed as a new sequential co-operative game based approach and proposes a new communication technique and along with storage-aware multi-objective approach that is optimized generally with two user functionalities like execution time of the task and cost while considering the available two constraint similar to the network bandwidth and storage sequence [7].

The authors [8] has addressed the workflow scheduling issues in cloud computing technology which later aimed to a complete execution of workflows by taking the QoS into account such that as the budget of the project and deadline of it in execution. Numerous workflow scheduling approaches are addressed in the workflows of cloud computing. Here, a complete survey and proper analysis of various techniques has been presented. It highlights the scheduling techniques in the cloud computing environment and provides a segregation of the proposed technique based on the various scheduling approach applied in each techniques.

4. Methodology

The proposed advances are characterized in the following steps

4.1 Creating Work Processes

Scientific work processes, for example, Inspiral, Montages, Cyber Shake and Sipht are first made. This structures the premise and connects to the dataset for recreation purposes.

4.2 Creating Server Farm

A datacenter is made to fill in as a capacity stage to run applications, process information, and store information. The datacenter is a pool of servers that contain an assortment of uses/programming that give space for different examples of virtual server to be worked and gotten to by clients by means of the Internet. The production of a datacenter is to guarantee that assets are made accessible on cloud for forward dissemination to satisfy the need of cloud clients successfully and gainfully.

4.3 Creating Cloud server farm representative
A cloud server farm intermediary is made. In actuality, a cloud server farm merchant might be an individual, gathering of people or a business substance that goes about as a delegate between the cloud specialist organization and the cloud administration customer/client. The cloud server farm merchant will be answerable for finding assets, choosing the assets and presenting the assets to the system framework and gathering status data identified with the asset for planning.

4.4 Resource Discovering and Selecting
In asset finding, a rundown of every single accessible asset in the cloud is acquired and recorded by cloud server farm dealer who presents it in the system framework and gathers status data identified with the asset for planning. During asset determination, data is gathered on the accessible assets and afterward the best asset is chosen to coordinate the application necessities for powerful planning.

4.5 Task accommodation
Task accommodation is the last stage in planning an undertaking to an asset. The cloud server farm specialist is liable for presenting the chosen errands to the accessible or inactive assets for planning.

4.6 Applying Max-Min, Round Robin and the MMax-Min calculations
After presenting an errand to an asset, the proposed Modified Max-Min calculation (M Max Min) is utilized to plan an undertaking, and from that point, the current Max-Min and Round Robin calculations are applied to benchmark the presentation of the proposed Modified Max-Min calculation.

4.7 Simulation improvement/examination of result
The reason for the recreation extract is to assess the presentation of the proposed Modified Max Min calculation to benchmark it against the current calculations. To accomplish this, the scientists recreated the proposed calculation in a re-enactment domain known as cloud sim toolbox. The outcomes were contrasted and benchmarked and the current calculations, for example, Max-Min and Round Robin calculations.

Pseudo-Code of Modified Max-Min calculation
1. While there are cloudlets in Cloudlet List
2. for all submitted cloudlets in the set C_i
3. for all VMs; VM_j is calculated finishing time (CT_ij) = et_ij + rt_j;
4. for each cloudlet in all VMs, discover the cloudlet with least execution time (MinClt)
5. else discover the cloudlet with greatest execution time (MaxClt)
6. Select a cloudlet (C_i) in the Cloudlet List and contrast and MaxClt.
7. If the ci has most extreme execution time, allocate the ci to the asset that produce it least fruition time for execution
8. Else appoint MinClt to the assets that produce it least execution time.
9. End if
10. Expel the cloudlet from the Cloudlet List
11. Update prepared time (rt_j) of the chose vm, R_j
12. Update CT_ij for all C_i
13. End While

MMax Min with other recently utilized cloud task planning calculations, for example, Max-Min and Round Robin, a recreation situation called Cloud Sim toolbox was utilized. The reenactment was run on Intel® center i3, 500GB HDD and 4GB Ram on 64 piece Windows 8 working framework. The two work processes are, Epigenomics and Sipht which contrast from one another as far as cloudlet size. The parade speed of each cloudlet is estimated in Million Instructions Per Second (MIPS).
5. Performance Matrix
In breaking down the exhibition of the proposed calculation (MMax-Min), we utilized the combined presentation network to quantify, look at and assess the presentation of the proposed calculation with the standard Max-Min and Round Robin.

5.1 Make span
It is all about the time taken to get a cloudlet planned. This is mostly used to quantify the effectiveness of computation with respect to time. It is estimated by the following conditions 1 and 2.
1. Make span = max (CT of j(t, m))
2. CT – the activity end time – the activity start time

5.2 Load balance
It is a way of distributing load on multiple cloud resources evenly for the purpose of increasing throughput and optimizing the use of cloud resource.
1. Set Server_Status as Empty // for handling the request from the queue
2. Set Hash_Map to the initial value 0 // Initialization of HashMap
3. Check Queue_Status to full or empty //Handled by Data Centre
4. Move the Request from Queue and update the Queue // Data Centre move incoming request from queue
5. Check Hash_Map is not 0. // To verify whether Hash Map handles the taken request
6. Update Server_Status as ready
7. Reassign_Server();
8. update the incoming request entries in the server and the queue; // update entries

6. Conclusion
To conquer the significant difficulties in task booking and asset portion utilizing Max-Min calculation in a distributed computing condition, a Modified Max-Min (MMax-Min) calculation was proposed. Re-enacted our methodology utilizing cloud sim toolbox which hugely observed a regular decrease in holding up time prompting less make span as increasingly more cloudlets showed up to be planned. This is on the grounds that our methodology can choose and allocate either cloudlet with greatest or least execution time to an asset. Execution examination, experimentation and benchmarking of the outcomes demonstrates that our drew closer (MMax-Min) can create great quality arrangements, delivering great estimations of make span and adjusting load successfully when contrasted with the standard Max-Min and Round Robin calculations.

References
[1] Zhang J, Huang H and Wang X 2016. Resource provision algorithms in cloud computing: A survey. *Journal of Network and Computer Applications* 64 p 23-42.
[2] Kaur G and Sharma E S 2014. Optimized utilization of resources using improved particle swarm Optimization based task scheduling algorithms in cloud computing. *International Journal of Emerging Technology in Advanced Engineering*. 4.6 p 110-115.
[3] Pasupuleti S K, Ramalingam S and Buyya R 2016. An efficient and secure privacy preserving approach for outsourced data of resource constrained mobile devices in cloud computing. *Journal of Network and Computer Applications*. 64 p 12-22.
[4] Lin Y, Thai M, Wang C and Lai Y 2015. Two-tier project and job scheduling for SaaS cloud service providers. *Journal of Network and Computer Applications*. 52 p 2636.
[5] Fei Teng 2011 Resource allocation and scheduling models for cloud computing. *Ecole Centrale Paris*.
[6] Rajan A P 2013. Evolution of cloud storage as cloud computing infrastructure service. *ArXiv Preprint arXiv* p 1308-1303.
[7] Duan R, Prodan R and Li X. 2014. Multi-objective game theoretic scheduling of bag-of-tasks workflows on hybrid clouds. *IEEE Transactions on Cloud Computing* **2.1** p 29-42.

[8] Masdari M, ValiKardan S, Shahi Z, and Azar S I 2016. Towards workflow scheduling in cloud computing: A comprehensive analysis. *Journal of Network and Computer Applications* **66** p 64-82.

[9] El-Sisi A B and Tawfeek M A 2014. Cloud task scheduling for load balancing based on intelligent strategy. *International Journal of Intelligent Systems and Applications* **6.5** p 25.

[10] Huang Q and Huang T 2010. An optimistic job scheduling strategy based on QoS for cloud computing. *2010 International Conference on Intelligent Computing and Integrated Systems (ICISS)*. p 673-675.

[11] Choudhary M, and Peddoju S K 2012. A dynamic optimization algorithm for task scheduling in cloud environment. *International Journal of Engineering Research and Applications (IJERA)*. **2.3** p 2564-2568.

[12] Fazel Mohammadi D, Jamali S and Bekravi M 2014 Survey on job scheduling algorithms in cloud computing.

[13] V K, R G 2019 Semi Markov process inspired selfish aware co-operative scheme for wireless sensor networks (SMPISCs). *Cybersecur* **2.4**

[14] V Kanchana Devi, R Ganesan. 2019 Trust-based Selfish Node Detection Mechanism using Beta Distribution in Wireless Sensor Network. *Journal of ICT Research and Applications* **13.1** p 79-92

[15] Devi V K, Gururaj A D M, Kavya A and Umamaheswari E 2018 Truck tracking and alerts monitoring system. *International Journal of Civil Engineering and Technology* **9** p 105-111.

[16] Devi Kanchana and R Ganesan 2017 A generic trust management framework for heterogeneous sensors in cyber-physical systems. *Asian Journal of Pharmaceutical and Clinical Research*. **10**. P 176.