International Journal of Engineering & Technology, 7 (2.33) (2018) 56-58

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

State of the art - optimization techniques in cloud environment

B. Priya [1] *, T. Gnanasekaran [2]

[1] Research Scholar, Anna University, Asst Prof , Sri Sai Ram Engineering College, Chennai
[2] Professor , R.M.K Engineering College, Chennai
**Corresponding author E-mail: t.gnanasekaran@gmail.com*

Abstract

A Cloud is a network of a shared pool of configurable and computing resources providing efficient, on-demand pay-as-per-use access. Its main objectives being: Scalability, High availability of resources and to reduce the overhead incurred. Scheduling is the method of determining the order by which the jobs have to be executed. It determines the various tasks that are to be executed in parallel and the efficient resource to carry out the tasks. Load balancing is the method of balancing the load across various resources by fixing a threshold and migrating the tasks to the under loaded resources based on the threshold. Optimization techniques are used to find a finite solution for scheduling of tasks although not optimal. Various optimization techniques are employed based on Cost, CPU utilization and to balance the load. This paper deals with the importance of optimization, the various metrics and constraints associated. A literature survey on the various optimization techniques is also analyzed based on the attributes of the tasks.

Keywords: *HJSA; Hierarchical; Virtual Machine (VM); Throughput.*

1. Introduction

Cloud computing is a computing service paradigm that provides services to the user as per the resource usage constraint pay-as-per use. Cloud is a distributed system for accessing the applications across the network. Virtualization concept is used to provide services to the client.

The different categories of users in cloud are: Cloud users, Cloud Service Provider and Cloud resource providers. When a request is submitted to the Cloud Scheduler using a Portal, the Scheduler access the Cloud resources using Application Provisioning and Virtualization technologies to provide the response.

Cloud Computing is categorized based on

1) The services offered as: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS).
2) The Cloud location as: Public, Private, Hybrid Cloud.

Scheduling of tasks in cloud computing allocates the best suitable resources for the task to be executed with consideration of different parameters like time, cost, scalability, make span, reliability, availability, throughput, resource utilization and so on.

An efficient load balancing technique provides an ideal environment, which improves the user satisfaction. Load Balancing helps in meeting the QoS requirements as well as maximizing the profit of the Cloud Service Providers with optimal resource usage.

The load balancing algorithms are categorized based on the origin of the process, system's current state and on the spatial distribution of nodes. Centralized, distributed and hierarchical load balancing are employed to optimize the resources.

The various QoS parameters considered in Cloud are: CPU Utilization, RAM Size, bandwidth, execution time, idle time of VMs, utilization cost.

Optimization techniques are used to efficiently allocate the tasks to the resource and to balance the load based on the tasks attributes. Some of the optimization techniques employed are Ant Colony Optimization (ACO), Particle Swarm Optimization(PSO), Greedy and Dynamic Optimization.

Soft computing techniques such as Hill Climbing, variations of it, heuristic searching methods are also used in optimization in cloud. Using these techniques the tasks are efficiently allocated to the various Virtual Machines.

The rest of the paper is organized as follows: Section II deals with the various metrics associated with optimization. Section III deals with the various constraints for optimization. Section IV illustrates the literature review on the various optimization techniques proposed by different authors. Section V gives a comparative study on the literature review.

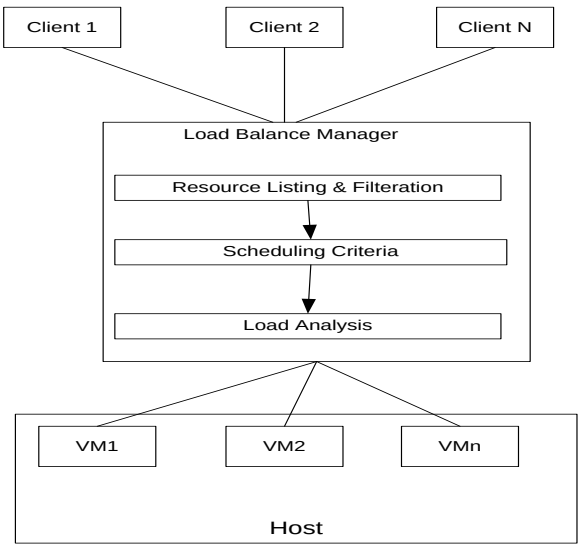

Fig. 1: Scheduling and Load Balancing in Cloud.

2. Metrics for optimization

Optimization of tasks in cloud environment is based on the following: [1]

1) Makespan: It is the sum of the completion time to schedule all the tasks to the resource. Makespan has to be minimized for a good optimization.
2) Cost: It is the cost associated with the execution, transmission and transfer cost for the various resources.
3) Waiting time: The time taken to start executing the task. It is the difference between the submission time and the time taken to start responding.
4) Turnaround time: It is calculated as the sum of the time taken to wait in the queue and the completion time of the tasks.
5) Fairness: This metric ensures that no tasks should be starved of the resources and should get a fair share of the CPU utilization.
6) Resource utilization: For efficient resource utilization, all the resources should be kept busy. Efficient resource utilization results in profit.
7) Throughput: It is the measure of the tasks completed per unit time. Throughput should be high for efficient scheduling.

3. Constraints for optimization

Some of the constraints associated with optimization are:

1) Assigning priorities to the task based on the CPU utilization, Average Completion time, Execution time, Resource utilization etc.
2) Considering the dependency of tasks: A directed acyclic graph structure can be used with the nodes representing the incoming tasks and the edges indicating the dependency of the tasks. The predecessor and successor of the various tasks are determined.
3) Deadline of the tasks and the delay in execution of the job.
4) The cost associated with completion of the tasks. The various cost parameters are evaluated.

4. Related work

In [2] Artificial Immune System Optimization technique has been employed to prioritize the jobs based on their attributes. Jobs with same type of requirements and those with reduced execution time are assigned the higher priority. The Fitness function is also calculated to evaluate the jobs. This algorithm resulted in better resource utilization.

In [3] Cost optimization technique has been used where the attributes of jobs are evaluated based on the following costs: transmission, transfer and execution cost. The optimal Virtual machine is selected first for the data centers and then an optimum Data Centre is selected hierarchically. The proposed HJSA (Hierarchical Job Scheduling Algorithm) achieves efficient cost, execution time and throughput.

The method proposed in [4] HLBSGNN is based on hierarchical load balancing based on generalized Neural Network. The optimization is carried out based on Communication Overhead for hierarchical load balancing.

The load associated with each node is calculated and GreedyCommLB, a load balancing strategy is used for migrating the tasks from overloaded node to the underloaded node. If a threshold is set, then based on it, the tree height can be reduced in hierarchical strategy and also the overhead associated with memory.

A Greedy optimization technique is employed in [5] considering the parameters: Task Completion time, Profit associated with the task, space required for execution and also based on resource utilization requirement. An Activity Based Costing method is used for optimizing cost.

In [6], an Enhanced Load Balancing Mutation Algorithm based on PSO is used to reschedule the tasks that have failed. The proposed method resulted in the minimization of cost and round trip time.

Approximate solutions can be found out by using approximation algorithms to find the solutions for the problem for which the exact polynomial time is known.

Stochastic Hill Climbing, a variation of Hill climbing was proposed in [7] in a centralized system. It is one of the incomplete method for making assignments based on high probability. A candidate generator is used to map the candidate solution with the set of successors possible. An evaluation criteria is used to find a closer solution. This method resulted in improved response time.

Dynamic Optimization of resources was carried out in [8] based on the energy load. The average occupancy of resources of the virtual clusters are calculated. The performance index of resources is considered for balancing the load which resulted in reduced time and efficient load balancing method.

In [9], an artificial intelligent technique for efficient searching and optimization, Genetic Algorithms are employed to find the most suitable processors for job in cloud environment invoking the operations of search, operations and replacement. The under loaded sites are found out and the jobs are migrated from other sites. The performance showed a improved response time compared to FCFS, Round Robin and stochastic hill climbing algorithms.

Modified Bacterial Foraging Optimization(MBFO) algorithm in [10] ia aimed at the parameters of CPU utilization, Memory size and bandwidth measurement. Directed Acyclic Graph has been analyzed and an improvized Classification tree was used for choosing a particular Virtual Machine. Determining the resources to be alloted to the users is predicted by Regression tree analysis. The experiment results were comapared with PSO algorithm and showed an improvement in minimizing the execution time, cost and an increased throughput.

5. Comparison of the various optimization techniques

Table 1: Comparison of Optimization Techniques

S. No	Technique	Parameters Considered	Performance
1	Artificial Immune System	Execution time	Efficient resource utilization.
2	Hierarchical Job Scheduling Algorithm(HJSA)	Transfer Cost, Transmission and Execution cost	Applied in hierarchical environment – Efficient throughput and cost.
3	Hierarchical Load Balancing based on generalized neural network(HLBSGNN)	Communication Overhead	Better Migration of tasks. Reduced tree height.
4	Greedy Optimization	Task Completion time, memory space	Optimized Cost
5	Enhanced Load Balancing Mutation Algorithm.	Particle Swarm optimization, round trip time	Optimized Cost.
6	Stochastic Hill Climbing	Probability factors	Applied in Centralized environment - Improved response time.
7	Dynamic Optimization	Energy load	Reduced Execution time
8	Modified Bacterial Foraging Optimization(MBFO)	CPU utilization, RAM Size , Bandwidth	Minimized execution time, cost. Efficient throughput.

6. Conclusion

In this paper, the various optimization techniques have been analyzed based on various attributes of the tasks. It is intended to propose a optimization based on improving response time in cloud environment.

References

[1] Mala Karla, Sarbjeet Singh, "A review of metaheuristic scheduling techniques in cloud computing", .Egyptian Informatics Journal (2015) 16, 275–295.

[2] R. Valarmathi, T. Sheela, "A Novel Hierarchical Scheduling Method for Managing Parallel Workloads in Cloud", Global Journal of Pune and Applied Mathematics (2016).

[3] Pown Kamarajapandian, Chitra, HJSA: A Hierarchical Job Scheduling Algorithm For Cost Optimization In Cloud Computing Environment, Economic Computation and Economic Cybernetics Studies and Research, Issue 2/2016, Vol. 50.

[4] Jixiang Yang, Limg Ling, Haibin Liu, "A hierarchical load balancing strategy considering communication delay overhead for large distributed computing systems", Hindawi Publishing Corporation, 2016.

[5] Asif Mohammad, Prof. Ashish Kumar, Lal Shri Vratt Singh, "A Greedy Approach for Optimizing the Problems of Task Scheduling and Allocation of Cloud Resources in Cloud Environment", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056, Volume: 03 Issue: 09 | Sep-2016.

[6] A.I.Awada, N.A.El-Hefnawyb, H.M.Abdel_kaderc, "Enhanced Particle Swarm Optimization For Task Scheduling In Cloud Computing Environments ", Elsevier - Procedia Computer Science 65 (2015) 920 – 929.

[7] Brototi Mondal, Kouik Dasgupta, Paramatha Dutta, "Load Balancing in Cloud Computing using Stochastic Hill Climbing- A Soft Computing Approach", Elsevier- Procedia Technology 4(2012).

[8] Lao Zhihong Larisa Ivascu, "Cloud Computing Resource Dynamic Optimization Considering Load Energy Balancing Consumption", TELKOMNIKA, Vol.14, No.2A, June 2016.

[9] Kousik Dasguptaa, , Brototi Mandalb, Paramartha Duttac, , Jyotsna Kumar Mondald, Santanu Dame, "A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing", ScienceDirect, Elsevier -Procedia Technology 10 (2013) 340 – 347.

[10] Anusha Bamini and Sharmini Enoch "Optimization of Resource Allocation Parameters in Cloud Environment Using Design of Experiments", International Journal of Pure and Applied Mathematics, Volume 116 No. 22 2017, 217-232.