Volume 3, Issue 12, December 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Survey on Optimization Techniques for Task Scheduling in Cloud Environment

R. Angel Preethima*, Margret Johnson

Department of Computer Science and Engineering

Karunya University, India

Abstract— Cloud computing is developed on the base of distributed computing, grid computing and virtualization. Job Scheduling is much critical in cloud computing. By considering scheduling cost and job priority, job scheduling is done in cloud environment. This paper gives the detailed survey of metaheuristics algorithms to obtain an optimal solution in job scheduling. These algorithms are Ant Colony Optimization, Particle Swarm Optimization, Improved Genetic Algorithm, Modified Bee Life Algorithm, Artificial Bee Colony Optimization and Improved Intelligent Water Drops Algorithm. These job scheduling algorithms are evaluated by makespan, computation cost and data transmission cost.

Keywords— Cloud Computing, Job Scheduling, Makespan, Optimization Techniques, NP-hard problem.

I. INTRODUCTION

This Cloud Computing enables users to run an application across many computers which are distributed geographically and connected through a network. As said by Wang et al., a computing cloud is a set of network enabled services, providing scalable, QoS guaranteed, normally personalized, inexpensive computing platforms on demand, which could be accessed in a simple and pervasive way *[3]*. In cloud computing, the resources are provided to cloud users according to their needs. Cloud Computing is provided with vast facilities. Hence, it is most wanted in IT industries. The services which should be given are available through the internet. Hence, the internet facility should be must in order to access the services. The requirement needed in order to enhance those services of cloud by the system should have less memory, a very light operating system and browser. Few Benefits of cloud computing is listed: cost savings, no much resource to be installed; the services can be increased or decreased as per our requirements hence it is scalable or flexible. Maintenance is cheap and easier because all the resources are managed by the cloud providers itself. Scheduling is a NP-hard problem. The main goal is to schedule the task to the appropriate resources in accordance with appropriate time, in order to achieve the optimal solution by finding out a proper sequence of tasks that will be executed.

This paper is organized as follows; Section II includes a discussion about job scheduling techniques and section III gives the conclusion of the paper.

II. OPTIMIZATION TECHNIQUES

A. Ant Colony Optimization

Ant Colony Optimization which is proposed by Marco Dorigo in 1992. Ant Colony Optimization algorithm is a technique for solving optimization problems which can be reduced to finding good paths. Methodology used in *[2]*, employs Ant Colony inspired algorithm to minimize weighted flow time in parameter sweep experiment with job priorities. The most popular swarm intelligence techniques are Ant Colony Optimization (ACO). The new scheduler is based on the above technique for executing Parameter Sweep Experiments (PSEs) in cloud by considering the job priority.

The main aim is to achieve minimum weighted flowtime for a set of jobs and minimizing the makespan Scheduler operates at two levels: Cloud-wide or data center level and VM-level. In Cloud-level, user VMs is allocated to resources using ACO techniques. In VM-level, the jobs are allocated to a VM based on job's priority. Each physical machine is allotted with job by ACO algorithm in order to solve the distributed job scheduling problem. An ant works on each job. The ant search for the machines which has less load and transfer the jobs to those machines. This technique can achieve low makespan and flowtime.

B. Particle Swarm Optimization

Particle Swarm Optimization is a combined work of Kennedy and Eberhart in 1995. PSO is global optimization method which is used to find the global minimum of the objective function. In *[1]*, particle swarm optimization is proposed. Particle Swarm Optimization schedules applications to cloud resources that take into both computation cost and data transmission costs. The average computation cost is calculated for each and every task on all the resources. This can be calculated by executing the tasks in an application. The computation cost is inversely proportional to the computation time, the task that is completed quickly will have cost higher. As per the assumption, the sizes of input and output data of each task are known. If the cost of communication is increases, the time taken decreases or vice verse. The mapping of all tasks in the workflow is done initially.

The dependencies between the tasks are validated and the algorithm assigns the "ready" tasks as per the mapping given by PSO. The polling time is done after dispatching the task into the resources for execution. The waiting time is used for acquiring the status of tasks and it is called as middleware dependent. The ready list is updated when the task is completed. When communication cost changes, the PSO mapping has to be recomputed. Depending on the recomputed PSO mapping, the ready task are assigned to the compute resources.PSO optimizes the cost of task resource mapping based on the solution obtained.PSO find global minima very quickly. Particle Swarm optimization can attain good distribution of workload onto resources.

C. Genetic Algorithm

Genetic algorithm is proposed by Holand, who was inspired by biological evolution, in 1975. Genetic Algorithm is based on the biological concept of population generation. In *[4]*, improved genetic algorithm is proposed. Genetic algorithm is a method of scheduling the tasks to the resources based on their individual solutions. The main terms used in genetic algorithms are; Initial Population, Fitness Function, selection, crossover and mutation. Initial populations are generated by using the Min-Min and Max-Min techniques for Genetic Algorithms. Better initial population is obtained by Min-Min and Max-Min when they are used for individual generation. Improved Genetic Algorithm can achieve less makespan.

We have considered Virtual Machines as resource and Cloudlets as tasks/jobs. We have checked the performance of the algorithms in two cases: in first case, the number of virtual machines has made as fixed and the numbers of cloudlets are varied as per our need. But in the next case we have made the number of cloudlets as fixed and varied the number of virtual machines. In order to achieve minimum time of completion, jobs are scheduled in multiple machines efficiently such that the jobs are completed in minimum time.

D. Modified Bee Life Algorithm

In *[6]*, authors discuss Bee Life Algorithm with greedy method to minimize makespan (i.e. execution time). The main challenge of optimization methods is to find the global optimal. In BLA, set of tasks are chosen randomly for the first algorithm. Then the makespan is calculated for that set of tasks for particular CSC. Then finally the algorithm checks for stopping criteria. There are two stages .they are reproduction stage and foraging stage. In reproduction stage, by mutation and crossover the algorithm will find out which set of tasks will forwarded to which CSC. Fitness will compare the priority of the task chosen for particular CSC .If the chosen task has high priority, and then it will be scheduled first. Then the jobs are scheduled on their priority basis .The Shortest path algorithm is used for the foraging behavior in order to reach the first CSC.

In this Greedy methods try to enhance each single step. The features of tasks are Tasks are aperiodic, Tasks are non-preemptive and to the process unit each task has two ways of accessing: exclusive access and shared access. Tasks enter the center scheduler as a set and then through BLA algorithm and greedy method to generate an optimal schedule. Greedy approach is used as local search process in foraging part of BLA in order to reach the best individual in the neighborhood. In hybrid cloud, the greedy method will use shortest path algorithm to find the nearest cloud storage center and recourses.

E. Artificial Bee Colony Optimization

Artificial Bee Colony Optimization was proposed by Karaboga in 2005.Efficient Artificial Bee Colony Algorithm is proposed in *[5]*. There are four phases in ABC algorithm initialization phase, employed bees phase, onlooker bees' phase and scout bees phase. In initialization phase, from the search space the individuals are randomly selected. Mutation phase is added after the employed bee phase. The local search is done by Employed bee phase. The local best position can be changed through mutation, and the algorithm may not be trapped into local optima. By sharing the information individuals can make use of others' advantage. Mutation and crossover operator of Genetic algorithm (GA) in the classical ABC algorithm where used in an efficient artificial bee colony (ABC) algorithm. For solving the job scheduling problem with the criterion to decrease the maximum completion time, crossover operator after the employed bee phase and mutation operator after onlooker bee phase of ABC algorithm are added.

F. Improved Intelligent Water Drops Algorithm

The authors in *[7]* propose Intelligent Water Drops (IWD) algorithm for minimizing makespan and mean flowtime in job scheduling. After satisfying certain constrain and objectives for the given task, the resources are allocated to the tasks. This is the major purpose of the Job Shop Scheduling Problem (JSSP). The Intelligent Water Drops (IWD) algorithm is used for solving the Multi Objective Job Shop Scheduling Problem. The optimization objective of MOJSS is done by considering multiple criteria, as makespan, tardiness and mean flow time of the schedules to find the best compromising solutions. The MOJSS-IWD is proposed to solve the MOJSS problem. The algorithm of MOJSS-IWD is as follows: A global Pareto set is maintained. To select the best schedules an external iteration is executed for each objective. Each internal iteration is of two steps, namely, (i) identification of the initial schedules and (ii) a Pareto local search on the schedules identified, NIWD IWDs travel from the source node to the sink node. The IWD generates a feasible solution (schedule). Pareto local search is done in order to further improve the solutions.

The Pareto local search is based on a scoring function. During the local search, the schedule that yields the minimum scoring function value is selected. Then in order to add into the Pareto set dominance checking is performed in STB.

Then the result is updated and reported. For solving the multi-objective scheduling problem MOJSS-IWD is one of the best approaches.

III. CONCLUSIONS

In this paper, we have surveyed the various scheduling algorithms in cloud computing and described their performance. Existing scheduling algorithms does not consider some metrics such as energy consumption and availability. Therefore the scheduling algorithm should be involved in order to improve and achieve the optimal solution for job scheduling in cloud environment.

ACKNOWLEDGMENT

The authors would like to acknowledge the School of Computer Science and Technology and the Department of Computer Science for providing us with the opportunity to prepare this report.

.

REFERENCES

[1] Pandey S, Wu L, Guru S and Buyya R," A particle swarm optimization-based heuristic for scheduling workflow applications in Cloud Computing environments" In: International conference on advanced information networking and applications (AINA 2010).IEEE Computer Society; 2010. p. 400–7.

[2] Mateos, C., Pacini, E and Garcia Garino, C," An ACO-inspired Algorithm for Minimizing Weighted Flowtime in Cloud-based Parameter Sweep Experiments" Advances in Engineering Software 56, 2013, pp.38–50.

[3] Wang L, Tao J, Kunze M, Castellanos AC, Kramer D, Karl W. Scientific cloud computing: early definition and experience. In: 10th IEEE international conference on high performance computing and communications (HPCC 2008). Washington, DC, USA: IEEE Computer Society; 2008. p. 825–30.

[4] Pardeep Kumar and Amandeep Verma, "Independent Task Scheduling in Cloud Computing by Improved Genetic Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.2, Issue 5,May, 2012.

[5] Manish Gupta and Govind Sharma, "An Efficient Modified Artificial Bee Colony Algorithm for Job Scheduling Problem", International Journal of Soft Computing and Engineering, Vol.1, Issue 6, January, 2012.

[6] Tasquia Mizan, Shah Murtaza Rashid Al Masud and Rohaya Latip, "Modified Bees Life Algorithm for Job Scheduling in Hybrid Cloud". International Journal of Engineering and Technology (IJET 2012), Volume 2 No. 6, June, 2012.

[7] S.H.Niu, S.K.Ong and A.Y.C.Nee, "An improved intelligent water drops algorithm for solving multi-objective job shop scheduling", Engineering Applications of Artificial Intelligence 26, 2013, pp.2431–2442.

