



Task Scheduling Algorithms in the Cloud Computing Environment: Survey and Solutions

Naoufal Er-Raji*, Faouzia Benabbou, Ahmed Eddaoui

Laboratory of Modeling and Information Technology, The Faculty of Sciences Ben M'Sik, University Hassan II Casablanca, Morocco

Abstract— The cloud computing gives the possibility to have computing resources over the internet without owning the infrastructure. The relation between the cloud service consumers (CSC) and the cloud service providers (CSP) is formalized through the SLA (service level agreement). So the service provider must achieve best performance, shortest response time and optimal utilization of resources to respect this SLA. In cloud computing many tasks requires to be executed by the available resources, so the task scheduling become a big challenge in the cloud computing. To outperform this challenge there is a need to implement a good scheduling strategy with good scheduling algorithms looks at the requirements and priorities established. In this paper, a comparison of task scheduling algorithms in cloud computing is presented, in order to classify and identify those who may be most appropriate in the case of independent priorities tasks. We will show that the scheduling of tasks needs as well a good strategy which connects the customers' needs to scheduling of their tasks.

Keywords— Cloud computing, Task scheduling algorithm, scheduling strategy, task priority, SLA.

I. INTRODUCTION

In the literature there are a lot of definitions of Cloud Computing and this is one of them. The International Organization for Standardization defined the cloud as a paradigm for enabling network access to a scalable and elastic pool of shareable physical or virtual resources with self-service provisioning and administration on-demand. The cloud computing paradigm is composed of key characteristics, cloud computing roles and activities, cloud capabilities types, cloud service categories and cloud deployment models [1].

The main cloud characteristics are Broad network access, Measured service, Multi-tenancy, On-demand self-service, Resource pooling, rapid elasticity and scalability. The roles and activities of cloud are: Cloud service customer, Cloud service partner, Cloud service provider. The cloud service categories are: Communications as a Service (CaaS), Compute as a Service (CompaaS), Data Storage as a Service (DSaaS), Infrastructure as a Service (IaaS), Network as a Service (NaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). The Cloud deployment models are: Public cloud, Private cloud, Community cloud, and Hybrid cloud. The Cloud computing cross cutting aspects are Auditability, Availability, Governance, Interoperability, Maintenance and versioning, Performance, Portability, Protection of PII, Regulatory, Resiliency, Reversibility, Security, Service levels and service level agreement [1].

Cloud computing gives the possibility to use computing resources over the internet without owning the infrastructure. It provides the possibility to access applications and data from anywhere based on SLA (Service Level agreement). The management of resources is done by a set of existing and new technologies like virtualization and Service-Oriented Architectures (SOA) [2].

In the cloud computing the quality of service (QOS) is a measurement of user's satisfaction to the services provided. And, these services are defined in SLA, which is the contract between the cloud service consumers (CSC) and the cloud service providers (CSP) [3].

As the numbers of cloud users is increasing rapidly, the researchers have to find an optimal solution for tasks scheduling mainly when we should respect the SLA terms. The task scheduling in the cloud computing is the mechanism that maps user's tasks to the appropriate resources (VMs).

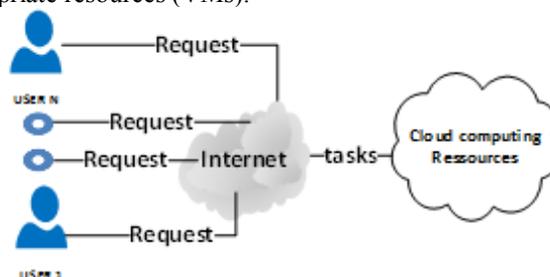


Figure 1 : The Cloud Computing Tasks

As depicted in the figure 1, in the cloud there are many users submitting their tasks which have different SLA constraints, to be executed by the available shared pool of resources simultaneously. For example, some consumers require more memory to store data, and some others may need more CPU time to compute complex task and so on. Thus, an effective strategy and effective task scheduling algorithms are needed.

The main objective of this paper is to study and compare a variety of task scheduling algorithms that are used in the cloud computing. And, also give a classification of these algorithms according to the consumers and providers need. We will also show that studying the scheduling solutions with isolated manner is insufficient to meet the needs of customers and therefore compliance with the SLA.

The remaining parts of this paper are organized as follows: section II defined the task and scheduling, section III present State of Art, Section IV present comparison between the tasks scheduling algorithms and finally Section V concludes the paper and presents the future work.

II. TASK SCHEDULING

In the literature a task is a single process or multiple processes which will be executed on a compute node presented by VM [4]. The task scheduling in the cloud computing is a set of rules and parameters, which determine and choose the task to be executed on the resources between a variety of possible tasks at a particular time [3]. The Task scheduling algorithms are the responsible for distribute the tasks submitted by the users onto available resources [5].

In the cloud there are different types of tasks for example dependent or independent tasks, periodic or permanent tasks, preemptive or non-preemptive tasks and so on. The decision of a type of scheduling is related, implicitly or explicitly, to the nature of priority of the task.

The tasks priority can be defined with different parameters in different level on different time. For example, it can be based on simple criteria or the combination of them. Following are some ones who are used in some papers.

-The Task execution cost: the costs that need the task to be executed in the specific virtual machine.

-The task age: the time that waited the task to be executed and this age begin when the task is entered to the cloud and finished in the execution of the task in the VM.

-The task size: the length or the number of MIPS that needs this task to be executed in a specific VM.

-The task deadline: the time that the task considered failed.

After the definition of the priority of the tasks, we have to assign them to the appropriate resources. Thus, this chooses can be based on the load of VM (maximum load or minimum load), the execution time (minimum execution time, maximum execution time), the capability of VM(number of MIPS) and so on.

In the literature, there are different types scheduling techniques with different characteristic like Static or Dynamic one, centralized / Distributed one, preemptive or non-preemptive one, cooperative scheduling, immediate or online one, batch or offline mode.

III. STATE OF ART

In the cloud many scheduling algorithms have been proposed. In this part of paper we present some of them. Before talking about the task scheduling algorithms used in the cloud computing, we can't neglect the basic algorithms, so we began with the basic algorithms and then the others.

First Come First Serve (FCFS): in this algorithm the task that arrived first in the queue will be executed first. **Last in First out (LIFO):** in this algorithm the last task arrived in the queue will be executed first.

Shortest Job First (SJF): in this algorithm the task with shortest processing time will be executed first.

Priority Based: in this algorithm, a priority number (integer) is associated with each task; the one who has the highest priority is selected to be executed first. But this priority is changed depend the purpose that it used.

The Round Robin Algorithm (RR): in this algorithm the time is divided into multiple times quantum. Each task assign a time quantum, during this each task perform its operation.

The Minimum Execution Time (MET): in this algorithm the task is selected in random order, and assigned to the machine with the minimum expected execution time for that task, regardless of that machine availability.

The Minimum Completion Time (MCT): in this algorithm the task is selected in random order, and assigned to the machine that has the minimum expected completion time for that task.

The Min-Min Task Scheduling Algorithm: in this algorithm the task with minimum execution time is assigned to the machine with minimum execution time for that task.

The Max-Min Task Scheduling Algorithm: in this algorithm the task with maximum execution time is assigned to the resource that produces the minimum execution time for that task.

Other algorithms have been proposed which are either the combinations of basic algorithms or the new algorithms.

Genetic Algorithm Based Bi-Objective Task Scheduling in Hybrid Cloud Platform [6]: This algorithm is based on the genetic algorithm, it focuses on hybrid cloud, it sends jobs to private cloud to minimizing the cost and sends jobs to public cloud to minimizing the execution time. The intention of this algorithm is to improve both cost and execution time in the same time in Hybrid cloud.

Coherent Genetic Algorithm for Task Scheduling [7]: This algorithm based on the genetic algorithm which considers resources, tasks cost and makespan as fitness criteria, the intention of this algorithm is to minimize the execution time and execution cost taking into consideration computational cost and computing capacity of the processing elements.

Genetic Algorithm Framework for Bi-objective Task Scheduling [8]: In this paper, the authors use genetic algorithm framework for their solution. They assume that the tasks are independent and can be executed in parallel on different

virtual machines (VMs). The intention of this work is to optimize makespan (running time) and the usage Virtual machine cost simultaneously. This solution can be used in public or hybrid cloud.

An improved Task Scheduling Algorithm based on Max-min [9]: This algorithm is the improvement over the improved Max-Min algorithm which use the initial step of Enhanced Max-Min algorithm, after that the improved Max-Min algorithm. Select the task just greater than average execution time and assigned it to the resource which gives minimum completion time. The intention of this algorithm is to optimize the scheduling polices and reduce the makespan.

Enhanced Max-min Task scheduling Algorithm [10]: In this algorithm the task with Average or nearest greater than the average execution time is selected to be executed by resource with minimum completion time (Slowest resource). It is based on comprehensive study of the impact of Improved Max-min task scheduling algorithm. It calculates the expected completion time of the submitted task on each resource. Then, the task with the overall maximum expected execution time (Largest Task) is assigned to a resource that has the minimum overall completion time (Slowest Resource). The intention of this algorithm is to reduce overall makespan and balance load cross resources.

Host Scheduling Algorithm Using Genetic Algorithm [11]: In this paper, the authors are proposed a new algorithm based on suffrage Heuristic coupled with Genetic algorithm. They assume that the tasks are independent. The goal of this algorithm is to minimize the makespan and enhance efficiency.

Optimization of FCFS Based Resource Provisioning Algorithm for Cloud Computing [12]: This algorithm is based on priority and uses the parameters to allocate resources; such as deadline or resources cost. The incoming tasks are grouped based on task requirements like minimum execution time or minimum resources cost and prioritized (FCFS manner). Resource selection based on task constraints using a greedy approach. The intention of this algorithm is to create a module depicting the normal FCFS algorithm in comparison to optimized version algorithm for resource provisioning in the cloud.

Job Scheduling Model for Cloud Computing Based on Multi-Objective Genetic Algorithm [13]: This algorithm is based on Multi-Objective Genetic Algorithm (MO-GA), it focuses on to minimize energy consumption and maximize the profit of service providers under the constraint of deadlines. It establishes a macroscopic scheduling model with cognition and decision components for the cloud computing, which considers both requirements of different jobs and the circumstances of computing infrastructure, then propose a job scheduling algorithm based on Multi-Objective Genetic Algorithm (MO-GA), taking into account the energy consumption and the profits of the service providers, and providing a dynamic selection mechanism of the most suitable scheduling scheme for users according to the real-time requirements.

A Priority Based Job Scheduling Algorithm in Cloud Computing [14]: this algorithm is based on the multi-criteria and multi-decision scheduling algorithm, also based on theory of Analytical Hierarchy Process (AHP) for calculating the priority. The Job requests a resource with a determined priority. This algorithm consists of three levels of priorities: object level (scheduling level), attribute level (the resources level) and alternative level (task level).

Improved Max-Min task algorithm [15]: in this algorithm the task with maximum execution time is assigned to the resources that give minimum completion time (slowest resource) for that task.

Bees Life Algorithm for job scheduling in cloud computing [16]: This algorithm is inspired from Bees behavior in nature. It basically runs in a cycle which has 13 steps of implementation. The Bees Life Algorithm cycle's iteration continues as long as the fitness changes and stops when fitness does not change. To optimize the Bees Life Algorithm two genetic operators are applied in reproduction phase, Crossover and Mutation.

IV. COMPARISON OF TASK SCHEDULING ALGORITHMS

In this section we propose a comparison of the above algorithms. The criteria of comparison are:

- Factor considered for task priority
- Factor considered for VM selected
- Deployment model
- Objectives
- Simulation tools

Algorithm ms	Factor considered for task priority	Factor considered for VM selected	Deployment model	Objectives	Simulation Tool
[6] 2016	The multi-objective optimization algorithm.	The capability of VM (MIPS)	Hybrid Cloud	-Developing a hybrid cloud scheduler that Optimize both cost and execution time simultaneously.	Cloudsim
[7] 2015	The task Cost Value	The result of mean function of each resource	Not Specified	Minimize the execution time (makespan) and execution cost on resources.	Cloudsim
[8] 2015	The task length	The types of VM (small, medium, large...)	-Public Cloud -Hybrid Cloud	Optimize simultaneously the makespan and cost of tasks.	Not Specified
[9] 2014	The average execution time.	The slowest resource	Not Specified	Optimize the scheduling polices and reduce the makespan	the Inspiral_100 , Montage_100 and Sipt_100

					workflows
[10] 2013	The nearest The Average execution time.	The slowest resource	Not Specified	Reduce overall makespan and balance load cross resources.	Cloudsim
[11] 2013	The expected completion time.	The genetic algorithm result	Not Specified	Minimize the makespan and enhance efficiency.	Cloudsim
[12] 2013	The task requirement	The task constraints using a greedy approach by the resource allocator	Not Specified	Optimized the normal FCFS algorithm for resource provisioning in the cloud.	Not Specified
[13] 2013	The task status (task manager)	-The task manager decision - The resource cognition component	Not Specified	Minimize energy consumption and maximize the profit of service providers under the constraint of deadlines.	Cloud Sim
[14] 2012	Analytical Hierarchy Process.	-The resource priority	Not Specified	Propose a new priority based job scheduling.	Not Specified
[15] 2012	genetic operation, crossover and mutation	-The greedy approach -The nearest datacenter	Not Specified	-Reduced the execution time (Makespan) with the least complexity. -spreading the workloads among the processing.	C++ programming language

The above algorithms try to optimize the task scheduling in the cloud computing, based each one on kind of priority parameter (implicitly). The most of papers use the task length as criteria to choose the tasks and use the capability (speed) to choose the resources.

Here there are some observations.

-There isn't enough information on the deployment model (public, hybrid, private) so we can't use some algorithms in some model, and we can't say if an algorithm has the same behavior indifferently of the deployment method.

-The main objective of all this work is to reduce the makespan.

-There isn't a work that addressed all **problems** accurately.

-A global strategy doesn't exist and the relationship with SLA contract is not clear.

-The most of paper use the CloudSim in their experiences.

So it is necessary to implement an effective strategy that gives more details on managing of tasks and their scheduling methods and show the details of how to schedule the task from the first step when the tasks are presented in the cloud until the last one.

We think that the choice of efficient algorithms is not an objective in itself, but rather, the choice of the approach should be based on customers' needs that are formulated in the SLA. We have more need of strategies that collects customer constraints, which vary depending on applications, and translate these constraints to priority parameters. Then, in the cloud, these parameters should be analyzed at every level of cloud (VM, node, cluster) to take the best decision in the selection of resources.

V. CONCLUSION AND FUTURE WORK

In this paper, we made a comparison of the different approaches for scheduling tasks. We have shown that these Approaches do not give an overview of scheduling from the client to the VM. These approaches do not show how the SLA is respected, for each customer, knowing that the constraints are not the same for all. Based on a single approach can't respond of the varied needs of customers. This is why our future work is to propose a comprehensive approach that can respect the constraints of clients and better manage resources in cloud computing.

REFERENCES

- [1] International Organization for Standardization "Information technology — Cloud computing — Overview and vocabulary" ISO/IEC 17788 First edition 2014-10-15, Reference number ISO/IEC 17788:2014(E) © ISO/IEC 2014.
- [2] P.D. Saronrex, R.K. Maheswari "Dynamic Consolidation of Virtual Machines In Cloud Data Centers For Managing Overloaded Hosts Under Quality of Service Constraints" International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com Volume 10, Issue 3 (March 2014), PP.40-47.
- [3] V R Bithiah Blessie,A.Stanislas,Dr L.Arockiam "A State of Art Scheduling Algorithms in Cloud Environment" Dept. of Computer Science, St.Joseph College Trichy, Tamilnadu, India, International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014) Vol. 2, Issue 2, Ver. 3 (April - June 2014).
- [4] Definition of jobs and tasks, [https://msdn.microsoft.com/en-us/library/bb525214\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/bb525214(v=vs.85).aspx)
- [5] Santhosh B , Dr. Manjaiah D.H "An Improved Task Scheduling Algorithm based on Max-min for Cloud Computing", International Journal of Innovative Research in Computer and Communication Engineering , Vol.2, Special Issue 2, May 2014, Department of CSE & ISE, Vemana Institute of Technology, Bengaluru, India.

- [6] Leena V. A., Ajeena Beegom A. S., and Rajasree M. S., Member, IACSIT "Genetic Algorithm Based Bi-Objective Task Scheduling in Hybrid Cloud Platform" International Journal of Computer Theory and Engineering, Vol. 8, No. 1, February 2016.
- [7] M. Krishna Sudha and, Dr. S. Sukumaran "Coherent Genetic Algorithm for Task Scheduling in Cloud Computing Environment" Australian Journal of Basic and Applied Sciences, 9(2) February 2015, Pages: 1-8 ISSN:1991-8178.
- [8] A.S. Ajeena Beegom and M.S. Rajasree "Genetic Algorithm Framework for Bi-objective Task Scheduling in Cloud Computing Systems" R. Natarajan et al. (Eds.): ICDCIT 2015, LNCS 8956, pp. 356–359, 2015. Springer International Publishing Switzerland 2015.
- [9] Santhosh B, Dr. Manjaiah D.H "An Improved Task Scheduling Algorithm based on Max-min for Cloud Computing" International Conference On Advances in Computer & Communication Engineering (ACCE - 2014) on 21st & 22nd April 2014, Organized by Department of CSE & ISE, Vemana Institute of Technology, Bengaluru, India
- [10] Upendra Bhoi, Purvi N. Ramanuj "Enhanced Max-min Task Scheduling Algorithm in Cloud Computing" international journal of application or innovation in engineering and management (IJAIEM) Volume 2, Issue 4, April 2013 ISSN 2319 – 4847.
- [11] Tarun Goyal & Aakanksha Agrawal "Host Scheduling Algorithm Using Genetic Algorithm In Cloud Computing Environment" International Journal of Research in Engineering & Technology (IJRET) Vol. 1, Issue 1, June 2013, 7-12 © Impact Journals.
- [12] Aditya Marphatia, Aditi Muhnot, Tanveer Sachdeva, Esha Shukla, Prof. Lakshmi Kurup "Optimization of FCFS Based Resource Provisioning Algorithm for Cloud Computing" IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 10, Issue 5 (Mar. - Apr. 2013), PP 01-05 www.iosrjournals.org.
- [13] Jing Liu, Xing-Guo Luo, Xing-Ming Zhang, Fan Zhang and Bai-Nan Li "Job Scheduling Model for Cloud Computing Based on Multi- Objective Genetic Algorithm" IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 3, January 2013 ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814 www.IJCSI.org.
- [14] Shamsollah Ghanbari, Mohamed Othman "A Priority based Job Scheduling Algorithm in Cloud Computing" International Conference on Advances Science and Contemporary Engineering 2012 (ICASCE 2012).
- [15] O. M. Elzeki, M. Z. Reshad, M. A. Elsoud "Improved Max-Min Algorithm in Cloud Computing" International Journal of Computer Applications (0975 – 8887) Volume 50 – No.12, July 2012.
- [16] Salim Bitam "Bees Life Algorithm for Job Scheduling in Cloud Computing" © ICCIT 2012, International Conference on Communications and Information Technologie.