

Survey on Job Scheduling algorithms in Cloud Computing

Fazel Mohammadi¹, Dr. Shahram Jamali² and Masoud Bekravi³

¹Department of computer, Science and Research branch, Islamic Azad University, Ardabil, Iran

²Associate Professor, University of Mohaghegh Ardabili, Ardabil, Iran

³Department of Computer Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran

Abstract: *Cloud computing is a distributed computing model in which enables developers to automatically deploy applications during job allocation and storage distribution. Cloud computing intends to share a pool of virtualized computer resources and equipment's of computation, storage, information and knowledge for scientific researches over the Internet. Job scheduling (allocation hardware resources to the tasks) is one of the major and challenging issues in cloud computing. A good scheduling policy also helps in proper and efficient utilization of Virtual Machine's (VMs). Job scheduling algorithm is an NP-complete problem which places an important role in cloud computing. This paper surveys the existing different type of job scheduling algorithms in cloud computing environment and summarizes some method to improve the performance.*

Keywords: Survey, Cloud Computing, Job Scheduling.

1. Introduction

Cloud Computing is an essential ingredient of advanced computing systems. Computing concepts, technology and architectures have developed and consolidated in the last decades. Many aspects are subject to technological evolution and revolution. Cloud Computing is a computing technology that is rapidly consolidating itself as the next step in the development and deployment of increasing the number of distributed application [1].

To gain the maximum benefit from cloud computing, developers must design mechanisms that optimize the use of architectural and deployment paradigms. The role of Virtual Machine's (VMs) has emerged as an important issue because, through virtualization technology, it makes cloud computing infrastructures to be scalable. Therefore developing on optimal scheduling of virtual machines is an important issue.

The cloud computing architecture has three layers [2], for the software which require on demand services over Internet. Figure 1 shows different layers of cloud architecture.

1. Infrastructure as a Service (IaaS): This layer delivers hardware components (like server and storage) and software as services.

2. Platform as a Service (PaaS): Cloud application developers are the users of this layer. Automatic scaling, load balancing and integration with other services (e.g. email services) are the major benefits to cloud application developer.

3. Software as a Service (SaaS): This layer hosts the software and provide to the customer through Internet. It reduces the purchase and maintenance cost of the

customer.

The main Purpose is to schedule tasks to the Virtual Machines (VMs) in accordance with adaptable time, which involves finding out a proper sequence in which tasks can be executed under transaction logic constraints [3]. The job scheduling of cloud computing is a challenge. To take up this challenge we review the number of efficiently job scheduling algorithms. It aims at an optimal job scheduling by assigning end user task.

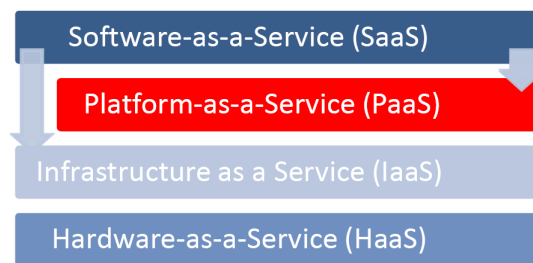


Figure 1 Cloud computing service

The rest of the paper is organized as follows. In next section Literature Survey about different scheduling algorithms of Virtual machine in cloud are discussed. Section 3 describes Existing Scheduling Algorithm in Cloud Computing. Section 4 discusses The Proposed Scheduling Algorithm. Section 5 discusses Experimental Setup and Results are analyzed. Conclusions are discussed in section 6.

2. Task Scheduling in Cloud Computing

Job Scheduling of cloud computing refers to dispatch the computing tasks to resource pooling between different resource users according to certain rules of resource use under a given cloud circumstances. At present there is not a uniform standard for job scheduling in cloud computing. Resource management and job scheduling are the key technologies of cloud computing that plays a vital role in an efficient cloud resource management.

3. Existing Scheduling Policy

The following task scheduling algorithms are presently established in the cloud environments and these algorithms have been summarized in table 1 with scheduling parameters.

3.1. Ant Colony Optimization (ACO)-inspired:

A new Cloud scheduler based on Ant Colony Optimization is the one presented by Cristian Mateos and et.al [4]. The goal of our scheduler is to minimize the

weighted flowtime of a set of PSE jobs, while also minimizing Makespan when using a Cloud.

In the ACO algorithm, the load is calculated on each host taking into account the CPU utilization made by all the VMs that are executing on each host. This metric is useful for an ant to choose the least loaded host to allocate its VM.

Parameter Sweep Experiments (PSE) is a type of numerical simulation that involves running a large number of independent jobs and typically requires a lot of computing power. These jobs must be efficiently processed in the different computing resources of a distributed environment such as the ones provided by Cloud. Consequently, job scheduling in this context indeed plays a fundamental role.

In this algorithm, Makespan and flowtime are evaluated. Evaluation results of this metrics show that ACO performance better than two other (Random and Best effort) algorithms.

3.2. Berger model:

Baomin Xu and et.al [5] proposed for the first time an algorithm of job scheduling based on Berger model in cloud environment. The Berger model of distributive justice is based on expectation states. It is a series of distribution theories of social wealth.

Expectation states formed by a series of theories are used to study actors and evaluate the impact of their behavior. Brief speaking, expectation states theories are to study the follow two issues. First, actor how to generate expectations of itself and other individual's according to the information (such as status, reward, and performance differences) around the world; Secondly, these expectations how to affect the behavior (such as participatory, and decision-making influence) of actors and others. Expectation states theories have been expanded and applied widely. State value theory of distribution justice is an important theoretical basis of the paper. It described that allocator how to use referential comparisons to establish the expectation for reward allocation. The expectations are used to evaluate the justice or injustice of distribution in a variety of circumstances.

Through the expansion of CloudSim platform, job scheduling algorithm based on Berger model is implemented. The validity of the algorithm is verified on the extended simulation platform. By comparing of simulation results with the optimal completion time algorithm, the proposed algorithm in this paper is effective implementation of user tasks, and with better fairness.

3.3. Dynamic level scheduling (DLS) algorithm:

Wei Wang and et.al extend the traditional DLS algorithm by considering trustworthiness of resource nodes [6]. This algorithm meets the requirement of user jobs in trust, and makes jobs scheduling based on directed acyclic graph (DAG) more reasonable. The main contribution of this study to scheduling systems is that it extends the

traditional formulation of the scheduling problem so that both execution time and reliability of applications are simultaneously accounted for.

Evaluation the trustworthiness of machines in Cloud environment by proposed algorithm shows decrease the failure probability of the task assignments, and assurance of the execution of tasks in a security environment.

3.4. Min-Min Algorithm:

The main idea of the Min-Min algorithm is as quickly as possible to dispatch each task to virtual machines as resources which can complete the task in the shortest possible time [7].

Min-Min algorithm will execute short jobs in parallel and the long jobs will follow the short jobs. The shortcoming of this algorithm is the short jobs scheduled first, until the machines are leisure to schedule and execute long jobs. Min-min can cause both the whole batch jobs executed time get longer and unbalanced load. Even long jobs cannot be executed. Compared with the traditional Min-min algorithm, improved algorithm adds the three constraints (quality of service, the dynamic priority model and the cost of service) strategy which can change this condition.

The experimental results of improved Min-Min algorithm show it can increase resource utilization rate, long tasks can execute at reasonable time and meet users' requirements.

3.5. Particle Swarm Optimization (PSO) Algorithm:

Particle Swarm Optimization (PSO) as a meta-heuristics method is a self-adaptive global search based optimization technique introduced by Kennedy and Eberhart [8]. The PSO algorithm is alike to other population-based algorithms like Genetic algorithms (GA) but, there is no direct recombination of individuals of the population [9].

The PSO algorithm focuses on minimizing the total cost of computation of an application workflow.

As a measure of performance, Authors used cost for complete execution of application as a metric. The objective is to minimize the total cost of execution of application workflows on Cloud computing environments. Results show that PSO based task-resource mapping can achieve at least three times cost savings as compared to Best Resource Selection (BRS) based mapping for our application workflow. In addition, PSO balances the load on compute resources by distributing tasks to available resources.

3.6. Priority based job scheduling (PJSC) Algorithm:

Priority of jobs is an important issue in scheduling because some jobs should be serviced earlier than other those jobs can't stay for a long time in a system.

In cloud computing, the main objective of this algorithm is to propose a new priority based job scheduling algorithm called PJSC. The proposed algorithm is based on the theory of Analytical Hierarchy Process (AHP). The Analytical Hierarchy Process is a multi-criteria decision-making (MCDM) and multi-attribute decision-making (MCDM) model [10].

The PJSC algorithm is based on multiple criteria decision making model.

The PJSC algorithm provided a discussion about some issues such as complexity, consistency and finish time. Evaluation result of this algorithm has reasonable complexity, also it decrease finish time (Makespan).

3.7. Round Robin Algorithm:

The Round Robin algorithm mainly focuses on distributing the load equally to all the resources [11]. Using this algorithm, the broker allocates one VM to a node in a cyclic manner. The round robin scheduling in the cloud computing is very similar to the round robin scheduling used in the process scheduling. The scheduler

starts with a node and moves on to the next node, after a VM is assigned to that node. This is repeated until all the nodes have been allocated at least one VM and then the scheduler returns to the first node again. Hence, in this case, the scheduler does not wait for the exhaustion of the resources of a node before moving on to the next.

Although round robin algorithms are based on simple rule, more load is conceived on servers and thus unbalancing the traffic.

Result of Round Robin algorithm shows better response time and load balancing as compared to the other algorithms.

Table 1. Scheduling Metrics considered by existing task Scheduling Algorithms

Paper	Execution Time	Response Time	Trust	Makespan	Utilization	Flowtime	Cost	Successful execution	Reliability	Fairness	Load Balancing	Reasonable Complexity
An ACO-inspired algorithm for minimizing weighted flowtime in cloud-based parameter sweep experiments [4]				✓		✓						
Job scheduling algorithm based on Berger model in cloud environment [5]	✓									✓		
Cloud-DLS: Dynamic trusted scheduling for Cloud computing [6]	✓		✓					✓	✓			
An Improved Min-Min Algorithm in Cloud Computing [7]	✓				✓							
A Particle Swarm Optimization (PSO) - based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments [8]							✓				✓	
A Priority based Job Scheduling Algorithm in Cloud Computing [10]				✓								✓
Analysis of variants in Round Robin Algorithms for load balancing in Cloud Computing [11]		✓									✓	
Load Balancing in Cloud Computing using Stochastic Hill Climbing-A Soft Computing Approach [12]		✓									✓	

3.8. Stochastic Hill Climbing-A Soft Computing Approach:

A local optimization approach Stochastic Hill climbing is used for allocation of incoming jobs to the servers or virtual machines (VMs) [12].

A variant of Hill Climbing algorithm Stochastic Hill Climbing (SHC) is one of the incomplete approaches for solving such optimization problems. The Incomplete methods may not guarantee correct answers for all inputs. Rather these methods find satisfying assignments for solvable problems with high probability.

With the scenario and configuration, overall average response time (RT) is calculated. The results show that in most of the case Stochastic Hill Climbing (SHC) decrease response time.

4. Our Contributions

A scheduling framework can be implemented by using different parameters. Good scheduling framework should include the following specifications. It must focus on:

- Load balancing and Energy efficiency of the data centers and virtual machines (VMs)
- Quality of Service parameters calculated by the user which contain execution time, cost and so on
- It should satisfy the security features.
- Fairness resource allocation places a vital role in scheduling

Considering all the metrics in a single scheduling framework is not a possible solution, hence it increases

the complexity of the design; the scheduling framework can be implemented.

5. Conclusion

Cloud computing is one of the user oriented technology in which user faces a pool of virtualized computer resources. In this paper we survey various existing scheduling algorithms in cloud computing. Table 1 is shown for additional reference. Since cloud computing is in infancy state, a scheduling framework should be implemented to improve the user acquiescence along with the service providers. The scheduling metrics can be coupled to prepare a framework for recourse allocation and scheduling in cloud computing. The scheduling framework should consider the user input limitations (deadlines, performance issues, execution cost, transmission cost, energy efficiency, Load Balancing, and Makespan) and so on.

References

- [1] Judeans and S. Ghemawat, "MapReduce: simplified data processing on large clusters," Sixth Symposium on Operating System Design and Implementation (OSDI'04), Dec. 2004, pp. 1-13.
- [2] Humayoun Israr, (2013), ARCHITECTURE LEVEL MAPPING OF CLOUD COMPUTING WITH GRID COMPUTING, International Journal of Engineering Sciences & Emerging Technologies, Volume 5, Issue 1, pp. 7-11.
- [3] XIAO Zhi-Jiao, CHANG Hui-You, YI Yang, (2007), "An Optimization Method of Workflow Dynamic Scheduling Based on Heuristic GA", Computer Science, Vol.1.34 No.2.
- [4] Cristian Mateos, Elina Pacini & Carlos Garc Garino, (2013), An ACO-inspired algorithm for minimizing weighted flowtime in cloud-based parameter sweep experiments.
- [5] Baomin Xu, Chunyan Zhao, Enzhao Hu, Bin Hu, (2011), Job scheduling algorithm based on Berger model in cloud environment, Advances in Engineering Software 42, PP. 419-425.
- [6] Wei Wang, Guosun Zeng, Daizhong Tang and Jing Yao, (2012), Cloud-DLS: Dynamic trusted scheduling for Cloud computing, Expert Systems with Applications 39, PP. 2321-2329.
- [7] Gang Liu, Jing Li and Jianchao Xu, (2013), An Improved Min-Min Algorithm in Cloud Computing, Department of Computer Science and Engineering, Changchun University of Technology, China.
- [8] Suraj Pandey¹, LinlinWu¹, Siddeswara Mayura Guru² & Rajkumar Buyya, A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments, Cloud Computing and Distributed Systems Laboratory, Department of Computer Science and Software Engineering, The University of Melbourne, Australia.
- [9] J. Kennedy and R. Eberhart, (1995), Particle swarms optimization In IEEE International Conference on Neural Networks, volume 4, pages 1942-1948.
- [10] Shamsollah Ghanbaria & Mohamed Othmana, (2012), A Priority based Job Scheduling Algorithm in Cloud Computing, Procedia Engineering 50, and PP. 778 – 785.
- [11] Pooja Samal and Pranati Mishra, (2013), "Analysis of variants in Round Robin Algorithms for load balancing in Cloud Computing", International Journal of Computer Science and Information Technologies, pp. 416-419, Vol. 4 (3).
- [12] Brototi Mondala, Kousik Dasgupta & Paramartha Duttba, (2012), Load Balancing in Cloud Computing using Stochastic Hill Climbing-A Soft Computing Approach, Procedia Technology 4, PP. 783 – 789.

AUTHORS



Fazel Mohammadi is currently a M.Sc. student in the Department of Computer in the Science and Research branch, Islamic Azad University, Ardabil, Iran. His research interests are in the areas of distributed systems with emphasis on scheduling algorithms for grids and clouds, cloud computing and machine learning. He is a teacher in Computer Science at technical high school and technical college.



Shahram Jamali is an associate professor leading the Computer Networking Group at the Department of Engineering, University of Mohaghegh Ardabili. He teaches on computer networks, network security, computer architecture and computer systems performance evaluation. Dr. Jamali received his M.Sc. and Ph.D. degree from the Dept. of Computer Engineering, Iran University of Science and Technology in 2001 and 2007, respectively. Since 2008, he is with Department of Computer Engineering, University of Mohaghegh Ardabil and has published more than 80 conference and journal papers. He also serves as editor member for some international journals



Masoud Bekravi received his B.Sc. and M.Sc. degrees in Electrical Engineering and Computer Engineering in 2004 (Islamic Azad university-Ardabil branch) and 2007 (Islamic Azad university-Arak branch) respectively. Now he is faculty member in the CE department, Islamic Azad University Ardabil branch. His research interests are power system, artificial intelligence, swarm robotics and Computer systems.