IJCTA, 9(36), 2016, pp. 137-145

A Survey of Scheduling Algorithm in Cloud Computing Environment

A. Ramathilagam* and K. Vijayalakshmi**

Abstract : Due to the scientific development, cloud computing is providing several dynamic services by means of extremely large scalable and virtualized resources through the Internet. Even though the improvement of cloud computing and its utilization is growing rapidly, there is no typical effective task scheduling algorithm employed in the cloud environment. In certain scenarios, the communication charge is very high and that prevents well known task schedulers to be implemented in large scale distributed environment. At the present time, researchers try to build job scheduling algorithms that are compatible and applicable in the large scale cloud computing environment. Here, job scheduling is the most vital process in a cloud computing environment, since users are required to pay for resources utilized in accordance with time. Thus, well-organized operation of resources is extremely vital and in order to effectively carry out the process scheduling plays a very important part to accomplish maximum benefit from the resources. In this paper, several scheduling algorithms are investigated and issues associated with them in cloud computing are discussed concisely.

Keywords : Cloud computing, scheduling, make span, priority, quality of service, resource allocation, task scheduling, task completion time.

1. INTRODUCTION

Cloud computing is providing several dynamic services by means of extremely large scalable and virtualized resources through the Internet. Numerous descriptions and interpretations of “clouds” and “cloud computing” are available. In its broadest form ‘cloud’ can be defined as an elastic execution environment of resources concerning multiple stakeholders and offering a metered service at multiple granularities for a specific level of quality of service. In an alternate way, it can be defined as cloud is a platform or structural arrangement that allows implementation of code (services, applications etc.), in a managed and elastic manner, while “managed” indicates that reliability in accordance with pre-defined quality constraints is automatically ensured and “elastic” indicates that the resources are put to exploitation in keeping with actual current requirements observing overarching requirement definitions absolutely, elasticity includes both up- and downward scalability of resources and data, however it is too load-balancing of data throughput.

Job scheduling is one of the most important process done in all the computing situations. Cloud computing is one the approaching latest technology which is increasing considerably. To efficiently enhance the functioning of cloud computing atmosphere, job scheduling is one the process done with the aim of achieving maximum profit.

* Assistant Professor, Department of Computer Science and Engineering, AAA College of Engineering and Technology, Sivakasi,Tamilnadu.

** Professor, Department of Computer Science and Engineering, Ramco Institute of Technology,Rajapalayam,Tamilnadu. *E-mail: vijayalakshmik@ritrjpm.ac.in*

The attempt of the scheduling algorithms in case of distributed systems to distribute the load on processors and increasing their utilization at the same time reducing the total task execution time job scheduling, one of the most important optimization complications, plays a vital position to enhance elastic and consistent systems. The foremost task is to schedule jobs to the regulating resources in synchronization with flexible time, which engages finding out an appropriate progression in which jobs can be performed under transaction logic constraints. There are predominantly two groups of scheduling algorithm. They are, (1) Static scheduling algorithm and (2) Dynamic scheduling algorithm. Both have their own merits and demerits. Dynamic scheduling algorithms have better performance than static algorithm however have several overhead compare to static scheduling algorithm.

The rest of the article is organized as follows: Section II describes the General view of scheduling algorithm Need of scheduling is detailed in the Section III. Section IV describes the conventional scheduling algorithm to detect and prevent the black hole attack on routing protocols. Section V shows the comparison of various scheduling algorithm in cloud.Finally, the Section VI explains the inference from the existing work.Section VII describes the metrics for clouds.Finaly the Section VIII draws the conclusions and the future work of this article.

2. GENERAL VIEW OF SCHEDULING ALGORITHM

Scheduling parallel applications in a distributed environment, like a group of workstations, makes a vital and unsettled complication. One of the major research complications is competently exploiting unused resources and to share the time with system rather between the processes. The objective of scheduling algorithms is distributing the load on processors and increasing their utilization at the same time reducing the overall task execution time [1]. There has been a collection of types of scheduling algorithm exists in distributed computing system. Most of them can be functional in the scenario of cloud environment with suitable verifications. The foremost advantage of job scheduling algorithm is to accomplish a high recital computing and the best scheme throughput. Conventional job scheduling algorithms are not adequate enough to provide scheduling in the scenario of cloud environments. Job scheduling algorithms in cloud computing can be classified into two main groups; Batch Mode Heuristic scheduling Algorithms (BMHA) and Online Mode Heuristic Algorithms.

In case of BMHA, Jobs are queued and gathered into a set when they enter in the system. The scheduling algorithm will begin following a permanent period of point in time. The major examples of BMHA based algorithms are; First Come First Serve (FCFS), Round Robin (RR), Min–Min and Max–Min algorithm. Through the On-line mode heuristic scheduling algorithm, jobs are taken when they enter in the system. In view of the fact that the cloud environment is a varied system and the pace of each processor differ quickly, the on-line mode heuristic scheduling algorithms are extremely appropriate for a cloud environment. Most fit task scheduling algorithm (MFTF) is a right algorithm of On-line mode heuristic scheduling algorithm.

A. First Come First Serve Algorithm

It is identical to FIFO. The job which comes first in the queue is served first. Average queuing time is long however the performance of FCFS algorithm is not so adequate.

B. Round Robin algorithm

It integrates the FCFS plus Preemptive Scheduling. In case of the round robin scheduling, processes are send out in a FIFO manner however are given a restricted amount of CPU time called a time-slice. In case when a process does not complete prior to its defined time-slice, the CPU is preempted and decided to the subsequent process, waiting in a queue. Works fine in case of small jobs; normally employed in timesharing systems. The preempted process is subsequently placed at the rear of the ready list.

C. Min–Min algorithm

It is a kind of stack task scheduling algorithm**.** This algorithm prefers small tasks to be finished earliest which consecutively large task delays for long time.

D. Max-Min algorithm

It is also a kind of stack task scheduling algorithm**.** This algorithm prefers huge tasks to be finished at the start, which consecutively small task delays for long time.

E. Most fit task scheduling algorithm

In this algorithm, job which fit best in queue are done first. This algorithm has high failure ratio.

F. Priority scheduling algorithm

The fundamental idea is uncomplicated: every process is allocated a priority, and priority is permitted to run. Equivalent-priority processes are scheduled in FCFS manner. The Shortest-Job-First (SJF) algorithm is an exceptional scheme of general priority scheduling algorithm. An SJF algorithm is simply a priority algorithm where the priority is the reverse of the subsequent CPU burst. Priority can be different either internally or externally. Internally determined priorities make use of certain quantifiable quantities or qualities to work out priority of a process.

Scheduling method in cloud can be partitioned into three stages explicitly,

- Resource discovering and filtering - Datacenter broker find out the resources adjacent in the network system and gathers standing information associated with them.
- Resource selection - Target resource is chosen in accordance with the assured constraints of task and resource.
- Task submission - Task is submitted to selected resource.

3. NEED OF SCHEDULING IN CLOUD

The most important advantage of moving to clouds is predominantly application scalability. Dissimilar to grids, scalability of cloud resources permits real-time provisioning of resources to meet application constraints. Cloud services, similar to, compute, storage and bandwidth resources are accessible at considerably lesser costs. Typically tasks are scheduled by means of user requirements. New scheduling policies require to be planned to overcome the complications posed by network properties among user and resources. Most recent scheduling policies possibly will use certain conventional scheduling ideas to merge them together with certain network aware policies to offer solutions for improved and more effective job scheduling .

In the beginning, scheduling algorithms were being executed in grids [2-3]. As a result of the decreased performance affected in grids, now there is a requirement to realize scheduling in cloud. This facilitates workflow management systems to readily satisfy the Quality of- Service (QoS) constraints of applications [4], as against to the conventional approach that needs advance reservation of resources in global multi-user grid settings. Cloud services, similar to, compute, storage and bandwidth resources are accessible at significantly lesser costs. Cloud applications typically need extremely difficult execution environments. These environments are complicated to produce on grid resources .

In case of cloud computing user might handle numerous virtualized resources to exploit, it is not possible for anyone to assign the jobs manually. Because of commercialization and virtualization characteristics, cloud computing makes job scheduling complexity to the virtual machine layer through resource virtualization. For this reason, in order to assign the resources to each job competently, scheduling takes an important responsibility in the field of cloud computing [5].

4. CONVENTIONAL SCHEDULING ALGORITHM

In the current scenario, the following scheduling algorithms are employed in wide range in clouds.

4.1. Resource-Aware-Scheduling algorithm (RASA)

[2] formulated a new task scheduling algorithm RASA. It includes two traditional scheduling algorithms; Max-min and Min-min. RASA makes use of the meritsof Max-min and Min-min algorithms and wraps their demerits. Despite the fact that the limit of every task, incoming rate of the tasks, rate of the task execution on each of the resource, cost of the communication are not taken into account. Results demonstrate that RASA outperform the existing scheduling algorithms in the scenario of large scale distributed systems.

4.2. An Optimal Model for Priority based Service Scheduling Policy for Cloud Computing Environment

Dakshayini & Guruprasad [6] formulated a new scheduling algorithm in accordance with the priority and admission control system. Here, priority is allocated to each admitted queue. Admission of each queue is decided through manipulative acceptable delay and service outlay. The majot benefit of this algorithm is that its strategy with the future cloud architecture has accomplished extremely high (99%) service completion rate with distinct QoS. At the same time, this policy provides the maximum preference for well paid user service-requests, in general servicing cost for the cloud also increases.

4.3. A Priority based Job Scheduling Algorithm in Cloud Computing

Ghanbari and Othman [7] planned a new scheduling algorithm with the use of multi-criteria and multi-decision priority driven scheduling algorithm. This scheduling algorithm includes a three stage of scheduling: object, attribute and alternate level. In this algorithm priority is allocated through job resource fraction. Subsequently, priority vector can be compared against each queue. This scheme has greater throughput and less terminate time.

4.4. An Optimistic Differentiated Job Scheduling System for Cloud Computing

Ambike et al [8] has formulated a new scheduling algorithm with non-preemptive priority queuing model for the purpose of cloud user actions in the cloud computing environment. In this scheme, one web application is generated to do certain activity like file uploading and downloading, subsequently there is a requirement for efficient job scheduling algorithm. The Qos requirements of the cloud computing user and the maximum profits for the cloud computing service provider are accomplished with this algorithm.

4.5. Improved Cost-Based Algorithm for Task Scheduling

Selvarani and Sadhasivam [1] has formulated an improved cost-based scheduling algorithm for the purpose of making effective mapping of tasks to accessible resources in the cloud environment. The development of conventional activity based costing is formulated by innovative task scheduling strategy for cloud environment where there might be no association between the clouds application base and the method that several tasks source for overhead cost of resources. These scheduling schemes partition the entire user tasks in accordance with the precedence of each task into three different lists. This algorithm measures both resource cost and computation level, it furthermore enhances the computation/ communication association.

4.6. Extended Max-Min Scheduling Using Petri Net and Load Balancing

El-kenawy et al [9] has been implemented a new algorithm in accordance with the use of RASA algorithm. Petri nets are employed for the purpose of modeling the simultaneous performance of distributed systems. Max-min demonstrates achieving better schedules than RASA and original Max-min.

4.7. Performance and Cost evaluation of Gang Scheduling in a Cloud Computing System with Job Migrations and Starvation Handling:

Moschakis and Karatza [10] has formulated a gang scheduling algorithm with job migration and starvation management in which scheduling comparable jobs, previously functional in the fields of grid and cluster computing. The number of Virtual Machines (VMs) existing at several moment is dynamic and scales in accordance with the demands of the jobs being observed. The above mentioned model's performance is assessed with the help of simulation and overall cost of gang scheduling with migrations and starvation management. Results highlight that this scheduling scheme can be successfully deployed on clouds, and that cloud platforms can be feasible for HPC or even for high performance enterprise application scenarios.

4.8. A Cost-based Resource Scheduling Paradigm in Cloud Computing

Yang et al [11] has formulated cost-based resource scheduling model by means of leveraging market theory to schedule resources to satisfy user's necessity. The algorithm is employed for the purpose of reducing the overhead of running the algorithm in the cloud environment. Result demonstrates that, the algorithm is capable of achieving an ideal stability between complexity and performance.

4.9. GA-Based Task Scheduler for the Cloud Computing Systems

Ge and Wei [12] has formulated a new scheduler which constructs a scheduling judgment by means of evaluating the entire group of tasks in the job paddle. A Genetic Algorithm (GA) is taken as the optimization method for the new scheduler. In order to achieve an enhanced balanced load across the entire nodes in the cloud environment, here adjusted the scheduler by calculating the execution time of tasks allocated to certain processors and making an optimal decision over the entire group of tasks. Result shows that this scheduler can obtain a shorter make span for jobs than previous scheduling policies and attain a better balanced load across all the nodes in the cloud.

4.10. Minimum Completion Time for Power-Aware Scheduling in Cloud Computing

Mehdi et al [13] has formulated an algorithm for task scheduling that lowering the power consuming and dropping the total datacenter load. This algorithm gives enhanced results owing to its capability to examine all the available hosts. It develops a scheduling algorithm that maps the incoming tasks to a host and guarantees the tasks deadlines and the minimum power consumption. It concentrates of Datacenter Load and Power consumption (DLP). This scheme motivated by Minimum Completion Time (MCT), which maps the task to the host that provides smallest completion time.

4.11. A Resource Scheduling Algorithm of Cloud Computing based on Energy Efficient Optimization Methods

Luo et al [14] has formulated resource scheduling algorithm of cloud computing based on energy efficient optimization methods. Initially, investigated the energy consumption model, and organized the computing resource in four groups: CPU, memory, storage, and network. They designed a variety of regulation policies for different components. Subsequently, they formulated a dynamic resource scheduling algorithm in accordance with energy optimization of CPU, main memory and storage.

4.12. RSDC (Reliable Scheduling Distributed In Cloud Computing)

Delavar et al [15] formulated a consistent scheduling algorithm. Here, foremost work is divided to sub work. In order to balance the jobs the request and acknowledge time are computed independently. The scheduling of each job is performed one through the calculation of the request and acknowledges time in the form of a shared job. In order that efficiency of the system is enhanced and also the total time of processing in cloud computing is developed.

4.13. Multiple QoS Constrained Scheduling Strategy of Multi-Workflows (MQMW)

Xu et al [16] worked on multiple workflows and multiple QoS. They formulated an approach for multiple workflow management system with multiple QoS. The scheduling access rate is considerably enhanced by using this strategy. This scheme reduces the make span and cost of workflows for cloud computing platform.

4.14. A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications

Pandey et al [17] formulated a particle swarm optimization (PSO) based heuristic to schedule applications to cloud resources that considers both computation cost and communication cost. It is employed for workflow application by varying its computation and communication costs.

Mohanasundaram and Periasamy [18] formulated a hybrid swarm intelligence optimization approach for identifying the optimal data storage position in WSN which observed to attain the desired results.

4.15. A Compromised-Time-Cost Scheduling Algorithm

Liu et al [19] formulated a novel compromised-time-cost scheduling algorithm which considers the characteristics of cloud computing to accommodate instance-intensive cost-constrained workflows by compromising execution time and cost with user input.

5. COMPARISON OF VARIOUS CONVENTIONAL SCHEDULING ALGORITHM

Table 1

Algorithm	*Scheduling scheme*	*Parameter*	*Factors*	*Findings*	*Environment*	*Tools*
Improved Cost-Based Algorithm for Task Scheduling[1]	Batch mode	Cost, Performance	Unsche-duled task group	Measures both resource cost and computation performance. Improve the communication/ Computation relation.	Cloud Environment	Cloud Sim
An Optimal Model for Priority based Service Scheduling Policy for Cloud Computing Environment[6]	Batch mode	Processing time	Grouped task	Used to reduce processing time and also used for load balancing	Cloud Environment	Cloud Sim
A Priority based Job Scheduling Algorithm in Cloud Computing[7]	Batch mode	QoS,Service request time	An array of Workflow instances	High Qos and Superior Throughput	Cloud Environment	Cloud Sim

An Optimistic Differentiated Job Scheduling System for Cloud Computing[8]	Dependency mode	Quality of Service, Maximum Profit	Single job with multiple user	The Qos requirements of the cloud computing user and the maximum profits of the cloud computing service provider are achieved	Cloud Environment	Cloud Sim
Performance and Cost evaluation of Gang Scheduling in a Cloud Computing System [10]	Batch mode	Performance, Cost	Workflow with large number of jobs	The application of migrations and starvation handling had a significant effect on the model. It improves performance.	Cloud Environment	Cloud Sim
A Cost-based Resource Scheduling Paradigm in Cloud Computing[11]	Batch mode	Performance	Grouped Task	To achieve perfect balance between performance and complexity	Cloud Environment	Hypervisor
GA-Based Task Scheduler for the Cloud Computing Systems[12]	Batch mode	Makespan	Grouped Task	To achieve a better balanced load across all the nodes in the cloud	Cloud Environment	Hadoop map reduce
Minimum Completion Time for Power-Aware Scheduling in Cloud Computing[13]	Batch mode	Load, Task size	Grouped Task	1. Lowering the power consuming 2. Reducing the datacenter load	Cloud Environment	Green Cloud Simulator
A Resource Scheduling Algorithm of Cloud Computing based on Energy Efficient Optimization Methods[14]	Batch mode	Performance, Energy consumption	Grouped Task	Used to reduce energy consumption	Cloud Environment (Eucalyptus)	Hadoop
RSDC (Reliable Scheduling Distributed In Cloud Computing) [15]	Batch mode	Processing Time	Grouped Task	To Decrease Processing Time Capable for load balancing	Cloud Environment	Cloud Sim

A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows [16]	Batch/ Depen-dency mode	Cost,time, Makespan	Multiple workflows	Used to schedule the workflow dynamically Used to reduce the execution time and cost.	Cloud Environment	Cloud Sim
A Particle Swarm Optimization-based Heuristic for Scheduling [17]	Depen-dendency mode	Resource utilization, Time	Group of taks	Used for good distribution of workload onto resources	Cloud Environment	Amazon EC2
Compromised-Time-Cost Scheduling Algorithm [18]	Batch mode	Make span	Grouped task	Used to reduce the make span	Grid Environment	Grid Sim

6. INFERENCE FROM EXISTING WORK

Based on the extensive survey done in the previous section, a lot of good work has been reported in the literature. However, there is still a need of scheduling techniques that covers all requirements accurately. The main issues of scheduling algorithm in cloud computing are the power consumption, cost and time, resource allocation, response and access time which results in algorithms performance. Another issue is energy efficiency, which is a very vital factor in scheduling algorithms. Scheduling is one of the key issues in the management of application execution in a cloud environment.

7. METRICS FOR SCHEDULING IN CLOUDS

The existing workflow scheduling algorithms consider various parameters like time, cost, make span, speed, scalability, throughput, resource utilization, scheduling success rate and so on. But, for a multiple workflows, metrics like Reliability and Availability should also be considered.

8. CONCLUSION AND FUTURE WORK

Scheduling is one of the most important process in the field of cloud computing. In this paper, the various existing scheduling algorithms are analyzed. It is observed that the disk space management is one of the vital issues in a virtual environment in the cloud. Existing algorithm provides higher throughput and cost effective, however they do not take reliability and availability into account. As a result, it is necessary to develop an algorithm for the purpose of enhancing the availability and reliability in a cloud computing environment. Depending on surveying the various algorithm it can be concluded that, make span can be reduced by grouping the task. Since cloud computing systems have a high degree of unpredictability with respect to resource availability. In future as the cloud size increases, there is a need for better task scheduling algorithm.

9. REFERENCES

1. S. Selvarani, and G.S. Sadhasivam, “Improved cost-based algorithm for task scheduling in cloud computing,” IEEE international conference on Computational Intelligence and Computing Research (ICCIC), pp. 1-5), 2010.

2. P. Saeed and R. Entezari-Maleki, "RASA: A New Task Scheduling Algorithm in Grid Environment," in World Applied Sciences Journal 7 (Special Issue of Computer & IT), pp. 152-160, 2009.
3. M.W. Berry, S.T. Dumais, and G.W. O'Brien, "Using linear algebra for intelligent information retrieval," SIAM Review, Vol. 37, pp. 573-595, 1995.
4. N. Muthuvelu, J. Liu, N.L. Soe, S. Venugopal, A. Sulistio, and R. Buyya, "A Dynamic Job Grouping-Based Scheduling for Deploying Applications with Fine-Grained Tasks on Global Grids," in Australasian Workshop on Grid Computing and e-Research (AusGrid2005), Newcastle, Australia. Conferences in Research and Practice in Information Technology, Vol. 44, pp. 41-48, 2005.
5. Green Technologies for the Energy-optimized Clouds" in Asian Journal of Research in Social Sciences and Humanities, Vol. 6,Issue 6, Special Issue June 2016
6. Dr.M. Dakshayini, and Dr. H.S. Guruprasad, "An Optimal Model for Priority based Service Scheduling Policy for Cloud Computing Environment," International Journal of Computer Applications, Vol. 32, No. 9, pp. 23-29, 2011.
7. S. Ghanbari, and M. Othman, "A priority based job scheduling algorithm in cloud computing," Procedia Engineering, Vol. 50, pp. 778-785, 2012.
8. S. Ambike, D. Bhansali, J. Kshirsagar, and J. Bansiwal, "An optimistic differentiated job scheduling system for cloud computing," International Journal of Engineering Research and Applications (IJERA) ISSN, pp. 2248-9622, 2012.
9. E.S.T. El-kenawy, A.I. El-Desoky, and M.F. Al-rahamawy, "Extended max–min scheduling using petri net and load balancing," International Journal of Soft Computing and Engineering (IJSCE), Vol. 2, No. 4, pp. 2231-2307, 2012.
10. I.A. Moschakis, and H.D. Karatza, "Performance and cost evaluation of Gang Scheduling in a Cloud Computing system with job migrations and starvation handling," 2011 IEEE Symposium In Computers and Communications (ISCC), pp. 418-423, 2011.
11. Z. Yang, C. Yin, and Y. Liu, "A cost-based resource scheduling paradigm in cloud computing," In 2011 12th International Conference on Parallel and Distributed Computing, Applications and Technologies, pp. 417-422, 2011.
12. Y. Ge, and G. Wei, "GA-based task scheduler for the cloud computing systems," In Web Information Systems and Mining (WISM), 2010 International Conference on, Vol. 2, pp. 181-186, 2010.
13. N.A. Mehdi, A. Mamat, A. Amer, and Z.T. Abdul-Mehdi, "Minimum completion time for power-aware scheduling in cloud computing," In Developments in E-systems Engineering (DeSE), pp. 484-489, 2011.
14. "Cluster based Key Management Authentication in Wireless Bio Sensor Network ", ,International Journal of pharma and bio sciences, Impact Factor = 5.121(Scopus Indexed).
15. A.G. Delavar, M. Javanmard, M.B. Shabestari, and M.K. Talebi, "RSDC (reliable scheduling distributed in cloud computing)." International Journal of Computer Science, Engineering and Applications, Vol. 2, No. 3, pp. 1, 2012.
16. M. Xu, L. Cui, H. Wang, and Y. Bi, "A multiple QoS constrained scheduling strategy of multiple workflows for cloud computing," In2009 IEEE International Symposium on Parallel and Distributed Processing with Applications, pp. 629-634, 2009.
17. S. Pandey, L. Wu, S.M. Guru, and R. Buyya, "A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments," In 2010 24th IEEE international conference on advanced information networking and applications, pp. 400-407, 2010.
18. R Mohanasundaram and PS Periasamy, "Clustering Based Optimal Data Storage Strategy Using Hybrid Swarm Intelligence in WSN", Wireless Personal Communications (Springer) Vol. 85, Issue 3, Pages 1381-1397, 2015.
19. Green Technologies for the Energy-optimized Clouds" in Asian Journal of Research in Social Sciences and Humanities, Vol. 6,Issue 6, Special Issue June 2016.