

An Analysis Report of Workflow Scheduling Algorithm for Cloud Environment

Akanksha Bhatia
Department of CSE
Rajasthan College of Engineering for Women,
RJ, India

Rakesh Sharma
Assistant Professor
Rajasthan College of Engineering for Women,
RJ, India

ABSTRACT

Cloud is almost an inseparable part of human's life, we are not aware but when we are sharing/storing our photographs online in our emails or social sites then we are using cloud services instead of storing them in our computer's hard drive or sharing them via hard devices. In the official works as well we many times comes into a situation to call a web service rather than themselves creating it in house and invoking, this comes under cloud computing. The workflows in cloud environment can be scheduled as needed based on the user's requirement as well as the scheduling algorithms concept. Different scheduling algorithm considers different parameters for the scheduling of task workflows and there outputs are also different. Some considers time or deadline as the base parameter and tries to complete the task within the provided deadline, they are also known as hard deadline based algorithms whereas some considers cost as there base parameter where they tries to reduce the cost as much as possible so that it will be economic to the end users. Whereas some are combination of both time and cost where they look after for a solution which can cover the deadlines as well as the solution will not cost much to the end users. There are some traditional scheduling algorithms also which schedules the task workflows like FCFS (First come first serve) and SJF (Shortest job first) which are also helpful in scheduling task as well as performance measurement of new algorithms. In this paper we have studies different and popular scheduling techniques of Workflows in different cloud models and we will analyze their experimental results.

Keywords

Cloud Computing, Hybrid Cloud, Heuristic Algorithm, Static Algorithm, Dynamic Algorithm, Virtual Machine

1. INTRODUCTION

New opportunities, simplified use and varied services are presented to the user by Cloud Computing. Scalability, cost effectiveness and resource pooling are some of the cloud attributes which made it more popular and easily approachable by a lot number of users. Scheduling of the available resources and there proper utilization is the biggest and core challenge for cloud computing so as to meet the user's demand efficiently. Scheduling techniques selects the best available resources from the pool and allocated it to the arrived task depending on some dynamic or static parameters.

A scheduler should be as smart as it can pick the appropriate resource to achieve high throughput of the overall system. When we consider the user's perspective then any of the

submitted task should not skip its deadline and when we consider from the cloud provider's point of view then the resource selection should be monetarily beneficial. The scheduling algorithm should be efficient enough to select the best available resource for a task as well as achieved the Quality of services of the submitted task in a pool.

Cloud computing is used to provide services of processing power as well as demand storage. In other words we can say that the cloud services provided the resources on lease for increasing the users processing power and computational capacity when necessary. The cloud users can easily lease or release the cloud resources. The main benefits to the end users are the avoidance of upfront investment, operating cost, maintenance cost and scalability on users demand. Elasticity is the feature which is provided by the cloud providers to the end users.

The process of a logical combination and presentation of the subset of resources from a pool of resources are also known as Virtualization, it is an abstract way by which the users can experience the benefits of the original configurations. The software for virtualization provides the interface to virtual machines as an abstract of hardware resources, network devices, software devices, memory etc.

Virtual Machines are the individual units which can execute isolate and individually without any dependency on any of the other virtual machine. Every Virtual machine will have its own processing unit, operating system, applications and network services. Virtualization is the concept of consolidated hardware, server and application within the same machine. Virtualization techniques examples are Xen which is an important player of the cloud computing, as by this users can dynamically select the software configurations as per there requirements in the requested resources. Fig 1.1 explains virtualization.

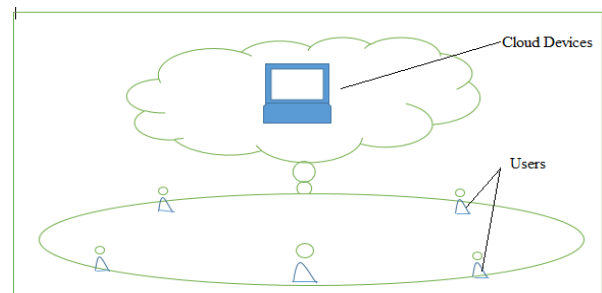


Figure 1.1 Concept of Virtualization

2. PROCESS OF SCHEDULING

Here we present the procedure of Scheduling in cloud environment by the steps as well as brief description of the same.

Cloud computing is a concept wide variety of differences. Scheduling process of cloud can be divided into below stages-

- 1) Resource Identification and status gathering – Brokers of datacenter identify the resources available in the cloud's network and all gather the information related to their status.
- 2) Resource selection depending on the parameters – This is a resource selection and decision making stage. The resource is selected for the task execution depending on certain parameters. These parameters can be time, cost etc.
- 3) Task Execution – At last the task is submitted to the selected resource and processing starts.

The users will only need to submit the task and specify the specifications with that of their requirements. Everything else is handled by the broker of the cloud provider. The task is assigned to the resource of the virtual machine and executed. A block diagram of the scheduling process is shown in Figure 1.2

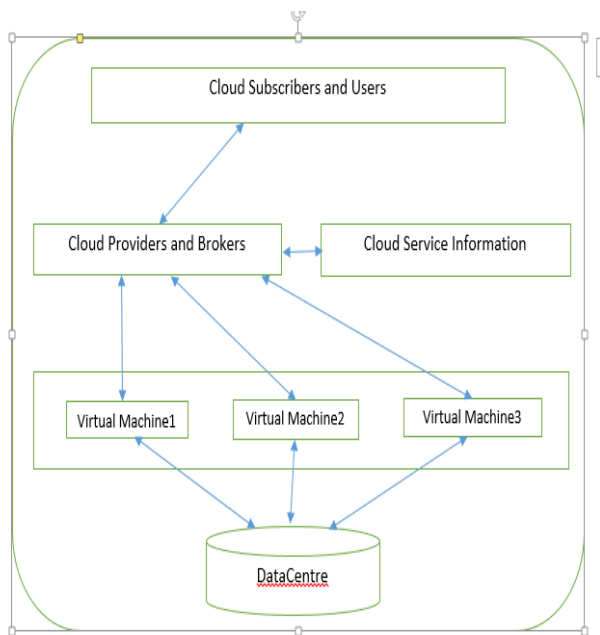


Figure 1.2 Scheduling in Cloud

3. RELATED WORK

There is many different models of cloud computing environment they are Public Cloud, Private Cloud and Hybrid Cloud. There are basically two sort of cost involved in the cloud environment, setup cost and Operational cost. Setup cost is one time cost which is needed to start a cloud environment at the user's premises and operational cost is the every time usage cost whenever a task is submitted for execution. The Public cloud works on pay per usage concept i.e. pay for the amount of resources you use or we can say pay as per usage. Higher the request of resources used then higher will be the charges of the same. Public clouds have no setup cost but they change as per usage i.e. they do have operational cost. Private cloud have high setup cost but no operational cost whereas Hybrid cloud is combination of both public cloud and private cloud.

We have studied the researches of some of the scheduling algorithms and below is the brief analysis of them. Table 1.1 shows a comparative study of the scheduling algorithms.

In paper [1] author has proposed a concept of the scheduling named generalized priority based scheduling, where both the task and the resources are selected depending on some criteria. Task which are submitted by the brokers are selected based on their size i.e. higher the size then higher the priority o that task. The resource is selected on the basis of their MIPS and the resource with highest MIPS will get the highest priority. The highest priority task is submitted to the highest priority resource. This generalized algorithm performs better than FCFS and Round Robin Algorithm and results are better.

In paper [2] author has proposed a new algorithm named cloud least laxity first where the laxity is the priority deciding parameter. Laxity is the time difference between the deadline and the time of execution of the task. The author deceives the priority of the job depending on the laxity and locality of the jobs. The lower the laxity of the job and higher its priority to be executed over a VM. This algorithm takes care of the soft deadline concept. Implementation of the algorithm is performed on CloudSim toolkit and the results are compared with the Time Shared and Space shared algorithms where the results are improved as compared to other algorithm.

In paper [3] author has proposed an algorithm which is based on the score based deadline constraint algorithm for cloud environment. The score is the parameter which represents the capabilities of the hardware resource depending on the MIPS (Millions of instruction per second), MOPS (Millions of Operations per second) and Disk space. The higher the score means the higher the performance of the virtual machine. The deadline allocated by the user is divided into the jobs specific deadline depending on the size of the jobs. In this way the deadline is divided into the jobs as per the sizes of the jobs and while allocation of the Virtual machine the deadline of the individual job is taken care to not to be missed. In this way the jobs are allocated to the highest score machine for execution. CloudSim is used as the toolkit for the implementation of Algorithm and performance comparison.

In paper [4] author has proposed an algorithm which is designed to work on Hybrid clouds and also reduces the cost and makespan of accessing public cloud. Hybrid cloud is a combination of the environment where public cloud and private both are used. Setup up of private cloud is expensive but its operational cost is low were as the setup cost in the public cloud is almost zero but they charge you as per usage therefore operational cost higher than private cloud. As a hybrid cloud is combination of both therefore author proposed an algorithm which can first utilize the private cloud resources and then use the public cloud resources if needed and as less as possible. This will also decrease the makespan of the task. Cloudsim is used for performing experiments and comparison shows that this algorithm performs well in reducing cost and reaching deadline of the resources.

In paper [5] is also associated with the better usage of the hybrid cloud environment. Hybrid environment where private cloud resources and public cloud resources both are available for usage and the algorithm has to decide the best usage procedure so that all the deadlines are meet as well as we have to pay less i.e. use the private cloud resources as much as possible. The author in this algorithm copes with the challenge of known as well as unknown running time of the submitted tasks. A concept of randomization is used in this algorithm. Monte Carlo algorithm is used for the unknown

running time task execution. This algorithm provides optimal solution to the end users.

In paper [6] author has proposed an algorithm named HCOC hybrid cloud optimized cost scheduling algorithm which works for the hybrid cloud environment. The algorithm is based on the selection of the public cloud resources and then they are aggregated to the private cloud so that enough processing power can be collected and the tasks can be

completed within the users provided execution time. As of the procedure, the workflow is first prepared for scheduling in the private cloud as the operational cost of the private cloud is less than public cloud. If the execution estimation is too high then the deadline then the algorithm searches for resources on the public cloud and aggregate them for processing of the task. In this way the operational task is also minimum and deadline is also not missed and the task execution is completed before the deadline of the task.

Table 1.1 Comparative study of different Scheduling Algorithms

Algorithm	Description	Parameters	Tool
[1]Generalized Priority based Algorithm	Task which are submitted by the brokers are selected based on their size i.e. higher the size then higher the priority of that task. The resource is selected on the basis of their MIPS and the resource with highest MIPS will get the highest priority.	Size, MIPS	CloudSim
[2]The Cloud least laxity First Algorithm	Tasks are collected in order of the least laxity first and the task with least laxity is the one to be processed first. The algorithm then looks for the resource having a slot available for processing. If available then the task is assigned to the resource for processing else algorithm restarts to process another item of task in the second task of the list.	Deadline, Cost	CloudSim
[3]Score based deadline constraint workflow scheduling algorithm	The algorithm is based on the concept of score and deadline distribution. The resources or the virtual machines are assigned the score depending on the MIPS, MOPS and Storage space and the task, subtasks are assigned the sub deadline so the processing will never miss the deadline provided by the end users.	Deadline, Cost	CloudSim
[4]Deadline and cost based workflow scheduling in hybrid cloud	The algorithm is implemented for the Hybrid cloud which is combination of private and public cloud. The private clouds are cheaper to the users therefore first priority is given to the private cloud and the left task which cannot be completed in private cloud are sent to be executed over the public cloud which are comparatively costly to the user. Author also ensures that the deadlines are also not missed and for this he divided the task in level based scheduling. All the levels are independent and they are not dependent on each other for the results or input. Deadlines are also decided at each level for processing.	Deadline, Cost, Makespan	CloudSim
[5]Randomized Approximation Scheme for resource allocation in Hybrid cloud	This algorithm is based on the hybrid cloud environment where the user have both private and public cloud with limited bandwidth and we have to find out the algorithm which can effectively use the private cloud and minimize the usage as well as cost of public cloud. The algorithm is designed to cover both the challenges of scheduling where the task execution time is known and another is unknown to environment. Monte Carlo algorithm is used for sampling in unknown time of execution algorithm. Author has considered the scheduling of BOT task i.e. bag full of task or the lot number of task together.	Cost, Makespan	Amazon EC2
[6] HCOC : A cost optimization algorithm for workflow scheduling in hybrid cloud	The hybrid cloud optimized cost schedule decides which resource should be leased from the public cloud so as to minimize the cost of the allocated tasks. The algorithm first makes an initial schedule using the path clustering heuristic and check if private cloud resources can satisfy the deadlines, if not possible then algorithm identify which all resources needs to be scheduled in public cloud so that performance is not affected.	Cost, Performance	OGSA

[7] Hyper heuristic scheduling algorithm	The hyper heuristic scheduling algorithm is based on the modification of the heuristic algorithm to provide better solutions for the makespan. The basic idea behind this algorithm is to use diversity detection and improvement detection operator in convergence process so as to balance the intensification and diversity in the search process.	Makespan	CloudSim and Hadoop
[8]Job Scheduling algorithm based on Berger model in cloud algorithm	The proposed algorithm uses the berger model for scheduling algorithm in cloud resources. It serves the two basic Quality of service parameters i.e. Completion time and bandwidth. The tasks are sorted on basis of the priority of the task in the list. Priority is as per the fairness parameter. The highest priority task get the resource VM for execution.	QoS(Completion time, Bandwidth)	CloudSim
[9]Priority based resource scheduling algorithm in cloudsims	The author presented a priority based algorithm which can provide an environment in the simulator which can be mapped to the real cloud environment. The algorithm is designed to handle the priority tasks by the VMs, the cloudlets are divided into jobs and the jobs are allocated to the appropriated VM. This algorithm provides a real time environment to researchers so that the researchers will have an exact scenario real cloud.	Priority	CloudSim
[10]A threshold based dynamic resource allocation scheme for	The algorithm is designed to allocate the Virtual machine resources to the jobs depending on the load changes rather than peak demands. This also avoids the situation of resource reallocation and related overhead. This algorithm performance results in better results of resource utilization as well as cost.	Resource Utilization, Cost	CloudSim

In paper [7] author has proposed an improved and efficient algorithm of Heuristic algorithm. As per the author there is an immense scope of improvement in heuristic algorithms therefore this subject is chosen for improvement. Author has used the Nobel heuristic algorithm and named it as hyper heuristic scheduling algorithm. Improvement and diversity detection operators are used by the designed algorithm to dynamically identify the lower level heuristic which can be used in finding better candidate solution. The performance of the proposed algorithm is measured by comparing it with different algorithms in Cloudsim simulator as well as Hadoop which is a real system.

In paper [8] author uses a concept of Berger model for scheduling of jobs in cloud environment. The algorithm takes care of the two QoS parameters one is deadline/completion time and another is bandwidth. The algorithm makes it sure that none of the deadlines are missed as well as there is best usage of the bandwidth. The tasks are assigned the priorities based on the fairness parameter and they are sorted ascending i.e. highest priority task is always on the top of the list and the lowest priority task will be in the end of the list. The VMs are assigned to the task as per the priority. Author has used cloudsim for experimental purpose.

In paper [9] author has proposed an algorithm where the priority based tasks are assigned to the resources. The author also proposed the model to reduce the differences of simulation and real environment so that researchers will get the better results with exact estimated results. The author has provided the algorithm which will fill the gap between the real and actual environment of scheduling. Author also concluded that the designed algorithm is easily map able to the real environment and this also helps the researchers to achieve their goals of scheduling in real cloud environment. Author has used cloudsim for the scheduling experiments.

In paper [10] author has proposed a scheduling algorithm at application level rather than assigning the resources to the task. Author has proposed a threshold based scheduling algorithm that allocates the virtual machines dynamically rather than static assignment. This assignment takes place at applications depending on their load changes rather than allocation based on the requirement. The threshold parameter can be used for resources reallocation as required. Author has used cloudsim toolkit for experimental purpose.

4. CONCLUSION AND FUTURE WORK

Cloud computing is one of the emerging technology and as of its increased usage the effective scheduling algorithm are needed. The demand of resources are increasing and the number of customer participants are increasing from day by day as everyone is now aware of benefits of cloud environment hiring. For this increasing demand we need to consider all the required parameters like cost, operational time, deadline etc. Cloud is pay per usage therefore we need advance techniques to minimize cost. Meeting deadlines are also a big challenge in cloud environment. Most of the algorithms designed in the cloud are designed for soft deadline and cannot fit for the model where hard deadlines are the major challenge. And the algorithms which are designed to meet deadline are costly to the end users. It is very difficult to have a one point solution where the solution and algorithm is both cost effective as well as consider the deadline to not to be missed at all. So in future we will work on algorithm designing which can meet the deadline based on priority as well as cost will be as less as possible to the users. This proposed work will serve the dual purpose of scheduling which will be in benefit of end users.

5. ACKNOWLEDGMENTS

Our thanks to the researchers who have contributed towards the researches and professors for their guidance.

6. REFERENCES

- [1] Dr. Amit Agarwal, Saloni Jain, X. 2014 “Efficient Optimal Algorithm of Task Scheduling in Cloud Computing Environment” *International Journal of Computer Trends and Technology (IJCTT) – volume 9 number 7*
- [2] Quentin Perret¹, Gabriel Charlemagne¹, Stelios Sotiriadis², Nik Bessis² “A deadline scheduler for jobs in Distributed systems” 2013 27th International Conference on Advanced Information Networking and Applications Workshops
- [3] Ranjeet Singh, Sarbjeet Singh “Score based deadline constrained workflow Scheduling algorithm for cloud systems” *International Journal on Cloud Computing: Services and Architecture (IJCCSA)*, Vol.3, No.6, December 2013
- [4] Nitish Chopra, Sarbjeet Singh “Deadline and Cost based Workflow Scheduling in Hybrid Cloud” 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)
- [5] Mohammad Reza Hoseiny Farahabady, Young Choon Lee, Albert Y. Zomaya “Randomized approximation scheme for resource allocation in hybrid-cloud environment” Springer Science Business Media New York 2014
- [6] Luiz Fernando Bittencourt, Edmundo Roberto Mauro Madeira “HCOC: A Cost Optimization Algorithm for Workflow Scheduling in Hybrid Clouds” *Journal of Internet Services and Applications* December 2011, Volume 2, Issue 3, pp 207-227
- [7] Chun-Wei Tsai et.al. “A Hyper-Heuristic Scheduling Algorithm for Cloud” *IEEE TRANSACTIONS ON CLOUD COMPUTING*, VOL. 2, NO. 2, APRIL-JUNE 2014
- [8] Baomin Xu ET. AL. “Job scheduling algorithm based on Berger model in cloud environment” *Advances in Engineering Software* 42 (2011) 419–425
- [9] Satish Kumar Srivastava, Kumar Rangasamy “Priority Based Resource Scheduling Algorithm in CloudSim” *International Journal of Science and Research (IJSR)* ISSN (Online): 2319-7064 Volume 3 Issue 4, April 2014
- [10] Weiwei Lin ET AL “A Threshold-based Dynamic Resource Allocation Scheme for Cloud Computing” *Pir/odciaed Eiang Einnggeinreinegri 2n3g (0200 (1210)1 619)5 0 –0 07–00300*