

Analysis of Various Task Scheduling Algorithms in Cloud Computing

Patel Dhara R*, Dr. Chirag Thaker

Department of Information Technology, Shantilal Shah Engineering College, Bhavnagar, Gujarat, India

ABSTRACT

Cloud computing is a technology that provides data storage, pool of resources and online access to computer services. Scheduling defines the order in which the set of tasks to be completed. Though large number of resource is available in cloud computing the user tasks should be properly allocated to the resources so that maximum tasks can be executed by utilizing minimum resources. Cloud computing provides various scheduling algorithms considering various parameters that can increase the performance of the system. This review paper focuses on various Scheduling algorithms in detail and the issues and the challenges faced by those scheduling algorithms.

Keywords: Cloud Computing; Scheduling; Scheduling Algorithms

I. INTRODUCTION

Cloud computing is a distributed computing environment that includes a pool of resources, on-demand network access, various development platforms and useful software are delivered as a services to users on the basis of pay as per use over the internet. Rajkumar Buyya defined cloud computing as “Cloud is a parallel and distributed computing system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service level agreement established through negotiation between the service provider and consumers”. [2] Cost effectiveness, scalability, reliability, fault tolerance, service-orientation, and resource management and scheduling, utility based, portability, virtualization and service level agreement (SLA) are the features of Cloud Computing. Cloud computing components include the web and central server to take care about the resource like data, storage, applications, etc. Cloud computing provides various types of service and deployment models. The major service models are Infrastructure-as-a-service (IaaS), Platform-as-a-service (PaaS) and Software-as-a-service

(SaaS). The common deployment models are Public Cloud, Private Cloud, Hybrid Cloud and Community Cloud.

The main goal of cloud computing is to provide an efficient and easy access to remote and geographically distributed resources. Therefore the allocation of the resources and scheduling plays an important role on the performance and efficiency of the system. An efficient scheduling is a way to proper allocation of the tasks to different resources. User tasks need to be schedule properly with the resources for maximum utilization of the available resources. Scheduling can be categorized as job scheduling, task scheduling, resource scheduling and workflow scheduling.

The scheduling process in the cloud computing environment consists of several components as shown in the Figure 1. The user jobs/tasks which arrives are queued into the waiting queue. Waiting queue is the queue in which the jobs/tasks are waiting to get assigned on a particular machine for execution. The scheduler is the important component of the scheduling process, it determines the execution order of the jobs waiting in the

queue. The scheduler schedules the jobs/tasks to the computing entity. The computing entity provides all the computing facilities like operating system, software etc. to process the submitted jobs/tasks.

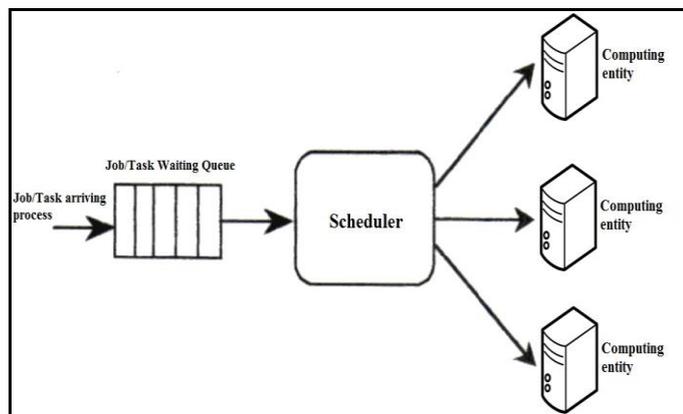


Figure 1. Scheduling Model

The paper is organized as follows: In section II, describes the various scheduling parameters used to increase the performance of the system. Section III presents the existing scheduling algorithms. Open issues have been discussed in section IV and finally the paper is concluded in section V.

II. METHODS AND MATERIAL

A. Comparison Parameters

Scheduling of the jobs/tasks, consider certain parameters through which the performance of the system can be improved by improving those parameters. Various parameters used to compare various scheduling algorithms are discussed in this section:

- **Makespan:** It is defined as the total completion time of all the tasks in a job queue. The makespan should be reduced to increase the performance of particular algorithm.
- **Resource utilization:** Resource utilization is the use of the resource in such a way that increases the throughput of the system.
- **Deadline:** It is the period of time from submitting a task to the time by which it must be completed.
- **Execution time:** The exact time taken to execute the given task is known as execution time. The ultimate goal of any scheduling algorithm is minimizing the execution time.

- **Completion time:** The time taken to complete the whole execution of a job. It also includes the execution time and the delay caused by the cloud system.
- **Load balancing:** It is the method of distributing the entire load in a cloud network across different nodes so that at a time no nodes remain under loaded. The load should be balanced to increase the efficiency of the system.
- **Energy consumption:** Many different scheduling algorithms has designed to reduce the power consumption and improving the performance.
- **Performance:** Performance indicates the overall efficiency given by the scheduling algorithm in order to provide good services to the user as per their requirements.
- **Quality of service:** This include various user input constraints like meeting execution, performance, deadline, cost, makespan, etc.
- **Response time:** The elapsed time between the end of an inquiry or demand on a computer system and the beginning of a response.
- **Scalability:** It is the ability of the system to function well when it is changed in size to satisfy the user need.

B. Existing Scheduling Algorithms

In this paper we are describing various task scheduling algorithms in detail.

FIRST COME FIRST SERVE ALGORITHM (FCFS): The First Come First Serve Algorithm (FCFS) collects the tasks in a queue until the resources are available and when the resources are available the tasks are assigned to them based on the arrival time of the task. It is the simplest algorithm and less complex in nature but it does not consider any other criteria for scheduling the tasks to the machine. Here always the first task will be executed first so the tasks arriving later will have to wait for a longer period of time. [12]

ROUND ROBIN ALGORITHM (RR) : RR scheduling is simple, easy to implement, and starvation-free. This algorithm uses the ring as its queue to store jobs. Each job has the same execution time and it will be executed in turn. If a job can't complete its work during its turn, it will be stored back to the queue waiting for

the next turn. Main feature of RR algorithm is execution of each job in turn and it doesn't have to wait for the previous one to get completed. But if the load is found to be heavy, RR will take a long time to complete all the jobs. [12]

EARLIEST DEADLINE FIRST ALGORITHM : In this scheduling algorithm, the scheduler points the task having the shortest deadline. Whenever a scheduling event occurs then the queue will be searched for the process that is closest to its deadline, the found process will be scheduled for execution. [10]

MIN-MIN ALGORITHM : The Min-Min Algorithm is based on the Minimum Completion Time (MCT) that is used to assign tasks to the resources having minimum expected completion time. It works in two phases, in the first phase, the expected completion time will be calculated for each task in a list and during the second phase, the task with the overall minimum expected completion time from the list is chosen and assigned to the corresponding machine. Then the task is removed from the list and the process is repeated until all the tasks in the list are mapped to the corresponding resources. This algorithm is unable to balance the load well as it usually schedules the smaller tasks first. [4]

MAX-MIN ALGORITHM : In this algorithm the scheduler schedules the tasks by expecting the Execution Time of the tasks and allocation of resources. Instead of selecting the minimum MCT, the maximum MCT is selected. It gives the priority to the larger tasks. The Max-Min algorithm is better than Min-Min algorithm in cases when the number of short tasks is more than the longer tasks. In this algorithm the larger tasks are given the highest priority then the smaller tasks so the smaller tasks have to wait for a longer period of time. [4]

RESOURCE AWARENESS SCHEDULING ALGORITHM (RASA) :

The Resource Awareness Scheduling Algorithm (RASA) is the hybrid of Min-Min and Max-Min algorithm. In this algorithm the Min-Min and Max-Min algorithms are applied alternatively to take advantage of both the algorithm and overcome their drawbacks. [6]

IMPROVED MAX-MIN ALGORITHM : This algorithm assigns task with maximum execution time to the resource which produce minimum completion time rather than Max-Min algorithm which assigns the task with maximum completion time to the resource which provides minimum execution time. [7]

ENHANCED MAX-MIN ALGORITHM : This algorithm modifies the Max-Min algorithm. It is based on the expected execution time in which it assigns a task with average execution time on the machine which gives minimum completion time.

IMPROVED COST BASED ALGORITHM : In this algorithm the scheduler schedules the tasks based on their cost to different resources. The cost of services varies for different tasks based on their complexity. The algorithm considers resource cost and processing capability of resources. They group tasks based on the processing capacity and selects best resource to schedule them so that the cost is reduced. [5]

An analysis of above mentioned scheduling algorithm and the different scheduling parameters considered by them, their advantages and disadvantages are listed in Table 1.

TABLE 1. Comparison Of Existing Scheduling Algorithms

Scheduling Algorithms	Parameters Considered	Advantages	Disadvantages
First Come First Serve Algorithm	Arrival time	Simplest in implementation	The jobs/tasks that arrive later have to wait for a long period of time so their waiting time increases
Round Robin Algorithm	Arrival time, Time quantum	Less complex and starvation-free	Pre-emption of jobs/tasks is required

Earliest Deadline First Algorithm	Dynamic task priorities, Pre-emptive, Deadline	The algorithm is optimal, all deadlines are met	Needs priority queue for sorting deadlines, behaves badly under overload, need dynamic priorities
Min-Min Algorithm	Minimum Completion Time	It is simple	Does not consider the existing load on resources
Max-Min Algorithm	Makespan	Reduce the makespan	Smaller jobs/tasks have to wait for longer time
Resource Awareness Scheduling Algorithm	Makespan	Reduces the makespan, avoids delays of executing large tasks, also support concurrency in execution of large and small tasks	If the resources are odd and the Max-Min strategy have better situation than Min- Min, we have to select the Min-Min instead of Max-Min.
Improved Max-Min Algorithm	Makespan	concurrency execution of tasks using available resources and load balance in small distributed environment	Does not considered the other scheduling parameters
Enhanced Max-Min Algorithm	Makespan, Load balancing, Average execution time	Improves the makespan	Only theoretical analysis is performed
Improved Cost-Based Algorithm	Processing cost, Makespan	Reduce the cost	Dynamic cloud environment and other QoS attributes are not considered

III. RESULTS AND DISCUSSION

Open Issues

Based on the survey conducted on various task scheduling algorithms we came to know that there is still lots of improvements that can be carried out. The major issues in scheduling algorithms are response time, cost, resource allocation, deadline, energy consumption and many other. Some techniques can be adopted to improvise the various issues and ultimately increase the performance of the system.

IV. CONCLUSION AND FUTURE SCOPE

In this paper, we have conducted the survey of various task scheduling algorithms in cloud computing and on the basis of the comparison parameters we compared various task scheduling algorithms with each other. A brief description of each algorithm is described in this

paper. The issues of the algorithm are addressed so that more efficient scheduling technique can be developed in

future which can fulfil the various parameters and increase the performance of the system.

V. REFERENCES

- [1] Peter Mell, Timothy Grance, "The NIST definition of Cloud Computing (September, 2011)", Accessed on May, 2014.
- [2] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg and I. Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", Future generation computer systems, vol. 25, no. 6, pp. 599–616, June 2009.

- [3] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. "A view of cloud computing", *Communications of the ACM*, 53(4):50–58, 2010.
- [4] Etminani. K, and Naghibzadeh. M, "A Min-min Max-min Selective Algorithm for Grid Task Scheduling". *The Third IEEE/IFIP International Conference on Internet*, Uzbekistan, 2007.
- [5] Mrs.S.Selvarani1; Dr.G.Sudha Sadhasivam, improved cost-based algorithm for task scheduling in Cloud computing, *IEEE* 2010.
- [6] Saeed Parsa and Reza Entezari-Maleki, "RASA: A New Task Scheduling Algorithm in Grid Environment" in *World Applied Sciences Journal* 7 (Special Issue of Computer & IT): 152-160, 2009.
- [7] S. Devipriya, C. Ramesh, "Improved Max-Min Heuristic Model for Task Scheduling in Cloud". *International Conference on Green Computing, Communication and Conservation of Energy (ICGCE)*, 2013
- [8] El-Sayed T. El-kenawy, Ali Ibraheem El-Desoky, Mohamed F. Al-rahamawy "Extended Max-Min Scheduling Using Petri Net and Load Balancing" *International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, September 2012*
- [9] Wikipedia. Cloud computing. Retrieved from <http://en.wikipedia.org/wiki/Cloud-computing>, 2011
- [10] J. Singh, B. Patra, S. P. Singh, "An Algorithm to Reduce the Time Complexity of Earliest Deadline First Scheduling Algorithm in Real-Time System", (IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol. 2, No.2, February 2011.
- [11] Chunmei Chi, Feng Gao. "The Trend of Cloud Computing" in China[J]. *Journal of Software*, VOL.6, NO.7, 2011,7:1230~1234.
- [12] Pinal Salot, "A Survey of various Scheduling Algorithm in Cloud Computing Environment," *IJRET*, vol. 2, pp.131–135, 2013.