

# Cloud Computing Scheduling Algorithms: A Review

<sup>1</sup> Varinder Kaur, <sup>2</sup> Gurjot Kaur

<sup>1</sup> Department of Computer Science and Engineering, Chandigarh University,  
Gharuan, Punjab/160055, India

<sup>2</sup> Asst. Professor, Department of Computer Science and Engineering, Chandigarh University,  
Gharuan, Punjab/160055, India

**Abstract** - A cloud consists of several elements such as clients, datacenter and distributed servers. Scheduling is a critical problem in Cloud computing, because a cloud provider has to serve many users in Cloud computing system. So scheduling is the major issue in establishing Cloud computing systems. The main goal of scheduling is to maximize the resource utilization i.e. energy consumption, power management and minimize processing time of the tasks using migration. In this paper, we will show how scheduling algorithms lead to optimization of the Qos.

**Keywords** - *Cloud Computing, Scheduling, FCFS, Priority, Qos.*

## 1. Cloud Computing

Cloud computing is an emerging technology that is becoming widespread as it enables accessing computing resources such as applications, storage, services, video games, movies and music on demand in such a way that the Cloud clients need not have any idea how or from where they are receiving these contents. The only thing they needed is a broadband connectivity to the Cloud. Cloud computing is an emerging area within the field of information technology (IT). Cloud computing is a model for enabling ubiquitous, convenient, on-demand network to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [6].

It is turning upside down the way it realize computation by enabling the use of It is also important to ensure that the cloud delivery models [7] provide generic access control interfaces for proper interoperability, which demands for a policy neutral access control specification and enforcement framework that can be used to address

cross-domain access control issues. Cloud computing is a concept involving different issues, concerns, technologies. Reaching to a global comprehensive definition seems to be defined arbitrarily for each IT organization or company. We rather to skip debates about our definition and use global definition announced by NIST [8] as follows:

“Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

## 2. Job Scheduling

Job scheduling is the key port of cloud computing job management. One excellent method may enhance the efficiency of the whole cloud computing system, and effectively share resource in wide area. The most important aim of cloud computing is to offer information services according to need. Scheduling theory for cloud computing is gaining consideration with day by day hike in cloud popularity. In general, scheduling is the process of mapping tasks to available resources on the basis of tasks' characteristics and requirements. It is an essential aspect in efficacious working of cloud as many task parameters need to be considered for proper scheduling. The available resources should be utilized efficiently without affecting the service parameters of cloud so that most of the jobs can be executed in sequential manner.

## 3. Scheduling Algorithms

There are many algorithms that are used to schedule the number of jobs at cloud.

### 3.1 FCFS Algorithm [10]

First Come First Served (FCFS)-Sometimes first-in first served and first-come, first choice is a service policy whereby the requests of customers or clients are attended to in the order that they arrived, without other biases or preferences. The policy can be employed when processing sales orders, in determining restaurant seating, on a taxi stand, etc. In western society, it is the standard policy for the processing of most queues in which people wait for a service that was not prearranged or pre-planned [10].

### 3.2 Priority Scheduling Algorithm [12]

In Priority Scheduling [12], each process is given a priority, and higher priority methods are executed first, while equal priorities are executed First Come First Served or Round Robin.

Priorities can be defined either internally or externally.

- Internally defined priorities use some measurable quantity or quantities to compute the priority of a process. For example, time limits, memory requirements, the number of open files, and the ratio of average I/O burst to average CPU burst have been used in computing priorities.
- External priorities are set by criteria outside the OS, such as the importance of the process, the type and amount of funds being paid for computer use, the department sponsoring the work, and other, often political, factors.

### 3.3 Round Robin Scheduling Algorithm [10]

The round robin scheduling works as following:

- Each process is provided a fix time to execute called quantum.
- Once a process is executed for given time period. Process is preempted and other process executes for given time period.
- Context switching is used to save states of preempted processes.

### 3.4 Multilevel Feedback queue Scheduling [5]

Multilevel feedback queue scheduling allows a process to move between queues. This movement is facilitated by the characteristic of the CPU burst of the process. If a process uses too much CPU time, it will be moved to a lower-

priority queue. This scheme leaves I/O-bound and interactive processes in the higher priority queues. In addition, a process that waits too long in a lower-priority queue may be moved to a higher priority queue. This form of aging also helps to prevent starvation of certain lower priority processes.

### 3.5 Least Slack Time Scheduling Algorithm [1]

Least Slack Time (LST) scheduling is a scheduling algorithm. It assigns priority based on the slack time of a process. Slack time is the amount of time left after a job if the job was started now. This algorithm is also known as Least Laxity First. It's most common use is in embedded systems, especially those with multiple processors.

Table 1: Comparative Table

<i>Algorithm</i>	<i>Comparison</i>
FCFS	The requests of customers or clients are attended to in the order that they arrived without other biases or preferences.
Priority	Each process is given a priority, and higher priority methods are executed first.
Round Robin	Each process is provided a fix time and it has to execute in that time.
Multilevel Feedback Queue Scheduling	If a high priority process uses too much CPU time, it will be moved to a lower-priority queue.
Least slack time scheduling	It assigns priority based on the slack time of a process and then executes.

## 4. Literature Survey

**M. Hwang et al. [1]** proposed LSTR (Least Slack Time Rate first), a new and simple scheduling algorithm, for a multi-processor environment, and demonstrate its efficient performance through various tests.

**Rohit et al. [2]** various scheduling algorithms, namely Short Job Scheduling, Job Scheduling Model based on

MultiObjective Genetic Algorithm, Priority based Job Scheduling Algorithm, SLA-Tree and Enhanced Max-min Task Scheduling Algorithm have been studied and analyzed.

**Altino et al. [3]** proposed to dynamically manage the execution of CPU-bounded tasks, deployed in a private cloud infrastructure. Additionally, co-located VMs suffer from performance interference due to contention in onchip resources. Our strategy aims at improving energy efficiency, while maximizing the rate of completed jobs by their deadlines.

**Bui Hwang et al. [4]** focused on finding platform scheduling policy for pilot-agent platform shared by many virtual screening users. They need a suitable scheduling algorithm at platform level to ensure certain fairness between users. So this paper includes multilevel scheduling algorithm.

**Kumaresh et al. [5]** evaluated the realization factor as the product of number of operations per cycle per processor and the speed of the processor. The subtasks are assigned high priority when both complexity factor and realization factor are high. Once the processes are classified into three queues they make use of a technique similar to round robin that reduces starvation of low and medium priority subtasks. The effectiveness of Starvation free (SF) Scheduling algorithm is evaluated through simulation result.

**Chuliang Weng et al. [6]** presented hybrid scheduling framework for the CPU scheduling in the virtual machine monitor. Two types of applications are high-throughput type and concurrent type. Virtual machine sets concurrent type when majority of workload is concurrent applications in order to reduce cost of synchronization. Otherwise it is set to high-throughput type by default. Experiments and results show that framework and scheduling strategy is flexible to improve performance of virtual machine.

## 5. Proposed Solution

Most data centers, by design, consume vast amounts of energy in an incongruously wasteful manner, interviews and documents show. The current review concern is the unwanted power utilized, energy consumed and more time consumed in data center which is exceptionally gaining attention of researchers with respect to scheduling of the computing resources. But one of major issue in datacenters found is to manage optimum energy, power usage in the systems.

In this review, we propose hybridization of least slack time scheduling and multilevel feedback scheduling algorithm to achieve above problem.

## 6. Conclusion

This paper presents common scheduling approaches in computational cloud. Cloud Resources are heterogeneous in nature, owned and managed by different organizations with different allocation policies. So scheduling problem must be handled in cloud computing. To perform better cloud needs a good scheduling algorithm. Using scheduling bandwidth of network can be utilized efficiently and response time can be deducted. Scheduling is the process to schedule data during transmission for uploading and downloading. Scheduling schedule application jobs and distribute load between machines to avoid circumstances of hanging. If proper scheduling is not achieved according to our requirement several errors can occur and it will produce errors like a few numbers of resources as their full capacity is not used and it is going vain.

## References

- [1] Dognjin Choi, M. Hwang and Pankoo Kim, "Least Slack Time Rate First: an Efficient scheduling Algorithm for Pervasive Computing Environment", *Journal of Universal Science*, vol.17, no. 6, 2011.
- [2] Rohit. O. Gupta, Tushar Champaneria, "A Survey of Proposed Job Scheduling Algorithms in Cloud Computing Environment", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3, Nov. 2013.
- [3] Altino M. Sampaio, Jorge G. Barbos, "Estimating Effective Slowdown of Tasks in Energy-Efficient Cloud Systems", *IEEE*, 2013.
- [4] Bui Kwang, "Multi level Queue Based Scheduling for Virtual Screening Application on Pilot-Agent Platforms on Grid/Cloud to Optimize the Strench", *The Sixth International Conference on future Computational Technologies and Applications*, 2014.
- [5] Kumaresh, "Multilevel Queue-Based Scheduling for Heterogeneous Grid Environment", *IJCSI*, Vol. 9, Issue 6, No 3, Nov. 2012.
- [6] Chuling Weng, Minglu Li., Xinda Lu and Zhingang Wang, "The Hybrid Scheduling Framework for Virtual Machine Systems", *Proc. Conf. VEE09*, 113-120.
- [7] M. Sheikhalishahi, M.Devare, L. "A General purpose And Multi-level Scheduling Approach in Energy Efficient Algorithm", *CLOSER Conference*, 2011.
- [8] Susane Albers, "Energy Efficient Algorithms", *Communication of ACM*, Vol.53, No.5, 2010.

- [9] D.Nurmi, R.Wolski, C.Grzegorzyc (2009), "The eculyptus open-source cloud-computing system", In Proc. Of CCGRID'09, 124-131.
- [10] Round Robin (RR)  
[http://en.wikipedia.org/wiki/round\\_robin\\_dchedilorg](http://en.wikipedia.org/wiki/round_robin_dchedilorg).
- [11] Jiandun Li, Junjie Peng and Wu Zhang, "A Scheduling Algorithm for Private Clouds", Journal of Convergence Information Technology, Vol. 6, No. 7, 2011.
- [12] Jing Lui, "Job Scheduling Model for Cloud Computing Based on Multi-Objective Genetic Algorithm", IJCSI, 2013.