

A Review on CPU Scheduling Algorithms in Cloud Environment

Kuldeep^{#1}, Ravinder Kumar^{#2}

¹kuldeepboora8[at]gmail.com

²rssdhanda[at]gmail.com

Abstract: *Cloud computing is a new paradigm in which computing is delivered as a service rather than a product, whereby shared resources, software, and information are provided to consumers as a utility over networks. Cloud computing is capable to provide massive computing or storage resources without the need to invest money or face the trouble to build or maintain such huge resources. Consumers only need to pay for using the services just like they do in case of another day to day utility services such as water, gas, electricity, etc. Scheduling algorithms are used for dispatching user tasks or jobs to a particular resource or data. Scheduling is a challenging job in the cloud because the capability and availability of resources vary dynamically. In this paper, we provide a review of various scheduling techniques used in cloud computing environment.*

Keywords: Cloud Computing, Scheduling, Round Robin, SRTF, Virtualization

1. Introduction

Cloud computing is a technique in which computing is delivered as a service rather than a product, whereby shared resources, software, and information are provided to consumers as a utility over networks. One of the main advantages and motivations behind Cloud Computing is reducing the CAPEX (capital expenditures) of systems from the perspective of cloud users and providers. By renting resources from cloud providers in a pay-per-use manner, cloud customers benefit from lowered initial investments and relief of IT maintenance [1].

Cloud computing is capable to provide massive computing or storage resources without the need to invest money or face the trouble to build or maintain such huge resources. Consumers only need to pay for using the services just like they do in case of another day to day utility services such as water, gas, electricity, etc. Cloud computing is now being used in many applications that are beyond distribution and sharing of resources [2]. The distributed resources are used only if the cloud resources are scheduled. The optimal scheduler results in high-performance cloud computing whereas poor schedulers produce substandard results. Scheduling algorithms are used for dispatching user tasks or jobs to a particular resource or data. Scheduling is a challenging job in the cloud because the capability and availability of resources vary dynamically. The goal of job scheduling is to properly dispatch parallel jobs to slave node machines according to scheduling policy meeting certain performance indexes and priority constraints to shorten total execution time and lower computing cost and improve system efficiency. In this paper we provide review of various scheduling techniques used in cloud computing environment.

2. Features of Cloud Computing

Cloud computing provides several features that make it attractive to IT industry, such as [1][2].

1)No Up-Front Investment: The pricing model in cloud computing is based on a pay-per-use principle. This model

gives the ability to rent services and resources from the cloud as he needs.

2)Lowering Operating Cost: Cloud environment resources are allocated and de-allocated on demand and this can provide a considerable saving in operating costs since resources can be released when service demand is low.

3)Scalability And Elasticity: the infrastructure providers have a large amount of resources and infrastructure. So they can easily expand its service to handle the growing service according to client demand. On the other hand, Elasticity is the ability to scale resources both up and down when required. Allowing the dynamic integration and extraction of physical resources to the infrastructure. That's mean elasticity enables scalability.

4)Easy access: the cloud services provided to users as web-based services. So, they can access the services through any devices supported with Internet connections.

5)Reducing business risks and maintenance expenses: Shifts the business risks such as hardware failures to infrastructure providers because providers have better expertise and resources to manage these risks.

6)Virtualization: Virtualization hides a computing platform's physical characteristics from users, It allows abstraction and isolation of lower level functionalities and underlying hardware.

7)Mobility: Cloud Computing implies mobility since clients can get to applications through web effectively anytime of time.

3. Basic algorithm of Scheduling

1)First Come First Serve: FCFS for parallel processing and is aiming at the resource with the smallest waiting queue time and is selected for the incoming task. The default policy implemented by the VM provisioned is a straightforward policy that allocates a VM to the Host in First-Come-First-Serve (FCFS) basis. The disadvantages of FCFS is that it is non preemptive. The shortest tasks which are at the back of the queue have to wait for the long task at the front to finish. Its turnaround and response are quite low[3].

2)Round Robin Scheduling: Round Robin (RR) algorithm focuses on the fairness. RR uses the ring as its queue to

store jobs. Each job in a queue has the same execution time and it will be executed in turn. If a job can't be completed during its turn, it will be stored back to the queue waiting for the next turn. The advantage of RR algorithm is that each job will be executed in turn and they don't have to be waited 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[3].

3) Priority scheduling algorithm: The basic idea is straightforward: each process is assigned a priority, and priority is allowed to run. Equal-Priority processes are scheduled in FCFS order. The shortest-Job-First (SJF) algorithm is a special case of general priority scheduling algorithm. An SJF algorithm is simply a priority algorithm where the priority is the inverse of the (predicted) next CPU burst. That is, the longer the CPU burst, the lower the priority and vice versa[3].

4) Min-Min algorithm: The Min-Min algorithm first finding the minimum expected time of all tasks in meta-task. The task having the minimum expected completion time is selected and assigned to the corresponding resource. This step is iterated until meta-task is not empty. Here, a big task has to wait for the completion of smaller ones[3].

5) Max-Min algorithm: The Max-Min algorithm expected completion time of each task as per the available resource is calculated. A task which has overall maximum completion time is scheduled over a resource with overall minimum execution time. This step is repeated until meta-task is not empty. Here, the waiting time of a larger task is reduced[3].

4. Literature Review

Job scheduling for cloud computing has pulled in awesome consideration. Most research in job scheduling receives a worldview in which a job in cloud computing system is characterized by its workload, deadline and the relating utility got by its finishing before the time limit, which are factors considered in contriving an effective scheduling algorithm. Utility Accrual (UA) paradigm is being known by this paradigm.

Numerous scientists have proposed distinctive scheduling algorithms that keep running under cloud computing circumstances. A Maximum number of scheduling algorithms have been proposed to accomplish two principal goals to be specific, to run the client assignment within time limit and to manage capability (load balancing) and reasonableness for all tasks. Here, we inspected the most applicable research works done in the literature for job scheduling in cloud computing.

A. Yang et al. (2011)

Highlighted “The crucial issue show in this paper of job scheduling in cloud computing that perceive the system state later on there is no very much characterized job scheduling algorithm. Already characterized job scheduling algorithms doesn't take hardware/software breakdown and restoration into account under utility computing paradigm in cloud. To catch this problem they proposed algorithm in perspective of

Reinforcement Learning (RL) that helps the scheduler in settling on scheduling decision with fault tolerable while boosting utilities accomplished in the long haul”[4].

B. Li et al. (2011)

Introduced “a hybrid energy efficient scheduling algorithm using dynamic migration that handles job execution in private clouds. The algorithm focus consideration on diminishing the response time, con-serves more energy and execute more raised measure of load balancing. Furthermore, the work in (Lin et al., 2011) concentrated on the problem of energy utilization in server. For virtual machine scheduling and consolidation a scheduling policy they show an algorithm named Dynamic Round-Robin (DRR) that drastically diminishes energy utilization. The algorithm endeavors to set up the virtual machines to servers and migrate virtual machines among data centers”[5].

C. Paul and Sanyal (2011)

Described “the issue of how to utilize cloud computing asset capably and increment benefits with the job scheduling system. Thus, they designed a credit based scheduling algorithm to figure the entire group of tasks in task queue and find the minimum completion time of all tasks. The proposed scheduling method consider the scheduling issue as a task issue in science where cost of a task to be appointed to a resource given by the cost matrix. Anyway, the algorithm does not recognize the handling time of a job, but rather distinct issues are viewed as the probability of a asset to be free not long after executing a task with the goal that it will be accessible for next waiting job”[6].

D. Vijindra and Sudhir Shenai (2012)

Proposed in their paper “an algorithm for a cloud computing environment that could consequently designate resources in view of energy optimization techniques. At that point, we demonstrate the effectiveness of our algorithm. In the experiments and results analysis, we find that in a practical Cloud Computing Environment, utilizing one entire Cloud node to calculate a single task or job will misuse a lot of energy, even when the structure of cloud framework generally support paralleled process. We have to convey an automatic process to locate the proper CPU recurrence, fundamental memory's mode or disk's mode or speed. We have also expand scalable distributed monitoring software for the cloud clusters”[7].

E. Neetu Goel, Dr. R.B. Garg (2012)

Discussed “a state chart that illustrate the relative study of various scheduling algorithms for a solitary CPU and shows which algorithm is best for the specific circumstance. Utilizing this representation, it turns out to be significantly less demanding to understand what is happening inside the system and why an alternate arrangement of processes is a contender for the allotment of the CPU at various time. The target of the study is to dissect the high productive CPU scheduler on map of the high quality scheduling algorithms which suits the scheduling goals”[8].

F. Monica Gahlawat, Priyanka Sharma (2013)

Described "Scheduling as a arrangement of approaches and mechanisms to control the request of work to be executed by a computer system. CPU is by wide margin the most vital resource of the computer system. Recent advances in software and architecture of the system expanded the unpredictability of the processing as computing is now distributed and parallel. Job scheduling is complex in this environment. The VM (Virtual Machine) can use a distinctive VCPUs (Virtual CPU) while running queue for each physical CPU, which is alluded to Partition Queue Model (PQM). As a difference, a Sharing Queue Model (SQM) of CPU scheduling algorithm can be used. This paper is analyzing and accessing the execution of different CPU scheduling for cloud condition using CloudSim".[9]

G. Vignesh V, Sendhil Kumar (2013)

Proposed in their paper, "the objective of this paper was to introduce a model for job-oriented resource scheduling for a cloud computing environment. Resource allocation task is scheduled for the Process which gives the accessible resources and client preferences. The computing resources can be assigned by the rank of job .This paper builds the analysis of resource scheduling algorithms. The time parameters of three algorithms, viz. Round Robin, Pre-emptive Priority and Shortest Remaining Time First have been thought about. From this, it has been proposed that SRTF has the least time parameters in all regards and is the most proficient algorithm for resource scheduling"[10].

H. Dilshad H. Khan, Deepak Kapgate (2014)

Discussed in their paper,"they clarified distinct algorithms and techniques introduced for Virtual Machine Scheduling either at single server or multiple server. Additionally surmises their qualities to determine the issue of proficient Virtual Machine Management in Cloud Computing. We examine and analyze these algorithms and systems in respects of different execution networks to give a diagram of the most recent approaches in the field"[11].

I. Lipsa Tripathy, Rasmi Ranjan Patra (2014)

Described that "Cloud computing is an emerging technology. It process huge amount of data so scheduling mechanism works as a vital role in the cloud computing. Thus my protocol is designed to minimize the switching time, improve the resource utilization and also improve the server performance and throughput. This method or protocol is based on scheduling the jobs in the cloud and to solve the drawbacks in the existing protocols. Here we assign the priority to the job which gives better performance to the computer and try my best to minimize the waiting time and switching time. Best effort has been made to manage the scheduling of jobs for solving drawbacks of existing protocols and also improvise the efficiency and throughput of the server"[12].

J. Nima Jafari Navimipour and Farnaz Sharifi Milani (2015)

Proposed that "the task scheduling problem in Cloud computing is an NP-hard problem. In this manner, numerous heuristics have been proposed, from low level execution of tasks in different processors to abnormal state execution of tasks. In this paper, another developmental algorithm is proposed which named CSA to schedule the tasks in Cloud computing. CSA algorithm is depends on the commit brood parasitic conduct of some cuckoo species in mix with the Levy flight conduct of a few birds and fruit flies. The simulation results comes about showed that when the estimation of Pa is low, the speed and scope of the algorithm turn out to be high"[13].

K. S.Sujan, R.Kanniga Devi (2015)

Proposed in their paper, "a dynamic scheduling scheme for cloud computing was proposed. Considering the resource provisioning as the primary issue to be address, the scheduling is being considered as the unique circumstance. Here we utilize a dynamic scheduling scheme considering the makespan as the metric. Algorithms like min-min and round robin are contrasted with deference with the proposed scheme. By considering and assessing all the above given constraints and strategies, a makespan based dynamic scheduling scheme for cloud computing is being proposed"[14].

L. Sagnika Saha, Souvik Pal (2016)

Proposed "a genetic based scheduling algorithm that reduces the waiting time of the overall system. However the tasks enter the cloud environment and the users have to wait until the resources are available that leads to more queue length and increased waiting time. This paper introduces a Task Scheduling algorithm based on genetic algorithm using a queuing model to minimize the waiting time and queue length of the system"[15].

M. Sushil Kumar Saroj, Aravendra Kumar Sharma (2016)

Described that "CPU scheduling has significant contribution in efficient utilization of computer resources and increases the system performance by switching the CPU among the various processes. However, it also introduces some problems such as starvation, large average waiting time, turnaround time and its practical implementation. Many CPU scheduling algorithms are given to resolve these problems but they are lacked in some ways. Most of the given algorithms tried to resolve one problem but lead to others. To remove these problems, we introduce an approach that uses both average and variable time quantum. In this approach, some processes are served with average time quantum and others with variable time quantum. This approach not only provides the minimum average waiting time and turnaround time but also try to prevent the starvation problem"[16].

N. Akilandeswari. P and H. Srimathi (2016)

Described "Cloud computing was utility based environment as pay per use model achieved by Parallel, Distributed and

Cluster computing accessed through the Internet. A key advantage of cloud computing is on-demand self-service, scalability, and elasticity. In on-demand self-service, the cloud user can request, deploy their own software, customize and pay for their own services. Scalability is achieved through virtualization. Being elastic in nature, cloud service gives the infinite computing resources (CPU, Memory, Storage). In cloud environment to achieve the quality of service many scheduling algorithms are available, but the scalability of task execution increases, scheduling becomes more complex. So there is a need for better scheduling. This paper deals with the survey of dynamic scheduling, different classification and scheduling algorithms currently used in cloud providers"[17].

O. Shridhar Domanal, Ram Mohana Reddy Guddeti, and Rajkumar Buyya (2016)

Proposed in their paper, "they proposed a novel hybrid Bio-Inspired algorithm for task scheduling and resource management, since it plays an important role in the cloud computing environment. Conventional scheduling algorithms such as Round Robin, First Come First Serve, Ant Colony Optimization etc. have been widely used in many cloud computing systems. Cloud receives clients tasks in a rapid rate and allocation of resources to these tasks should be handled in an intelligent manner. In this proposed work, we allocate the tasks to the virtual machines in an efficient manner using Modified Particle Swarm Optimization algorithm and then allocation / management of resources (CPU and Memory), as demanded by the tasks, is handled by proposed HYBRID Bio-Inspired algorithm (Modified PSO + Modified CSO). Experimental results demonstrate that our proposed HYBRID algorithm outperforms peer research and benchmark algorithms (ACO, MPSO, CSO, RR and Exact algorithm based on branch-and-bound technique) in terms of efficient utilization of the cloud resources, improved reliability and reduced average response time"[18].

5. Conclusions

Cloud computing is capable to provide massive computing or storage resources without the need to invest money or face the trouble to build or maintain such huge resources. Consumers only need to pay for using the services just like they do in case of another day to day utility services such as water, gas, electricity, etc. Scheduling algorithms are used for dispatching user tasks or jobs to a particular resource or data. Scheduling is a challenging job in the cloud because the capability and availability of resources vary dynamically. In this paper, we provide a review of various scheduling techniques used in cloud computing environment.

References

[1] Zhang, Q., Cheng, L., and Boutaba, R., "Cloud computing: state-of-the-art and research challenges", *Journal of Internet Services and Applications*, 1(1): pp. 7-18,(2010).
 [2] Sareen P., "Cloud Computing: Types, Architecture, Applications, Concerns, Virtualization and Role of IT Governance in Cloud", *International Journal of*

Advanced Research in Computer Science and Software Engineering, 3(3): pp. 533-538, (2013).
 [3] Teena Mathew, K. Chandra Sekaran, and John Jose, "Study and Analysis of Various Task Scheduling Algorithms in the Cloud Computing Environment", *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, 2014.
 [4] Yang, B., X. Xu, F. Tan and D.H. Park, 2011, "An utilitybased job scheduling algorithm for cloud computing considering reliability factor", *Proceedings of the 2011 International Conference on Cloud and Service Computing*, Dec. 12-14, IEEE Xplore Press, Hong Kong, pp: 95-102. DOI: 10.1109/CSC.2011.6138559.
 [5] Li, J., J. Peng and W. Zhang, 2011, "A scheduling algorithm for private clouds", *Journal Convergence Information Technology*, 6: 1-9. Li, J., M. Qiu, J. Niu, W. Gao and Z. Zong et al., 2011.
 [6] Paul, M. and G. Sanyal, 2011, "Task-scheduling in cloud computing using credit based assignment problem", *Int. J. Comput. Sci. Eng.*, 3: 3426-3430, 2011.
 [7] Vijindra and Sudhir Shenai. A, "Survey of Scheduling Issues in Cloud Computing", 2012, *ICMOC-2012*, 1877-7058, Elsevier Ltd, Doi: 10.1016/j.proeng.2012.06.337, page no: 2881 – 2888.
 [8] Neetu Goel, Dr. R.B. Garg, "A Comparative Study of CPU Scheduling Algorithms", *International Journal of Graphics & Image Processing*, Vol 2, issue 4, November 2012.
 [9] Monica Gahlawat, Priyanka Sharma, "Analysis and Performance Assessment of CPU Scheduling Algorithms in Cloud using Cloud Sim", *International Journal of Applied Information Systems (IJ AIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 5 – No. 9, July 2013*.
 [10] Vignesh V, Sendhil Kumar KS, Jaisankar N, "Resource Management and Scheduling in Cloud Environment", *International Journal of Scientific and Research Publications*, Volume 3, Issue 6, June 2013.
 [11] Dilshad H. Khan, Prof. Deepak Kapgate, "Efficient Virtual Machine Scheduling in Cloud Computing", *International Journal of Computer Science and Mobile Computing*, Vol.3 Issue.5, May- 2014, pg. 444-453
 [12] Lipsa Tripathy, Rasmi Ranjan Patra, "Scheduling In Cloud Computing", *International Journal on Cloud Computing: Services and Architecture (IJCCSA)*, Vol. 4, No. 5, October 2014.
 [13] Nima Jafari Navimipour and Farnaz Sharifi Milani, "Task Scheduling in the Cloud Computing Based on the Cuckoo Search Algorithm", *International Journal of Modeling and Optimization*, Vol. 5, No. 1, February 2015.
 [14] S.Sujan, R.Kanniga Devi, "A Dynamic Scheduling Scheme for Cloud Computing", *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* Volume 4 Issue 3, March 2015.
 [15] Sagnika Saha, Souvik Pal and Prasant Kumar Pattnaik, "A Novel Scheduling Algorithm for Cloud Computing Environment", *Advances in Intelligent Systems and Computing* 410, DOI 10.1007/978-81-322-2734-2_39, © Springer India 2016.
 [16] Sushil Kumar Saroj, Aravendra Kumar Sharma, Sanjeev Kumar Chauhan, "A Novel CPU Scheduling with

Variable Time Quantumbased on Mean Difference of Burst Time”, International Conference on Computing, Communication and Automation (ICCCA2016)

- [17] Akilandeswari. P and H. Srimathi, “Survey on Task Scheduling in Cloud Environment”, I J C T A, 9(37) 2016, pp. 693-698 © International Science Press.
- [18] Shridhar Domanal, Ram Mohana Reddy Guddeti, and Rajkumar Buyya, “A Hybrid Bio-Inspired Algorithm for Scheduling and Resource Management in Cloud Environment”, IEEE Transactions On Services Computing, VOL. X, NO. X, JULY 2016