International Journal of Research
Available at https://pen2print.org/index.php/ijr/
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 23
December 2018

A Review paper on CPU Scheduling in Cloud Environment

Nimmi Singh & Mamta Sohal

[1]Department of Computer Science
South Point Institute Of Technology And Management(SPITM),
Deenbandhu Chhotu Ram University of Science & Technology (DCRUST), Sonepat
[1]nimmi.singh46@@gmail.com

[2]Department of Computer Science
South Point Institute Of Technology And Management(SPITM),
Deenbandhu Chhotu Ram University of Science & Technology (DCRUST), Sonepat
[2]mamtakalra21@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 other 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 review of various scheduling techniques used in cloud computing environment.*

Keywords— Cloud Computing, Scheduling, Round Robin

I. 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 other 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 useful 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.

II. FEATURES OF CLOUD COMPUTING

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

- **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 cloud as he needs.

- **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.

- **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.

- **Easy access:** the cloud services provided to users as a web-based services. So, they can access the services

through any devices supported with Internet connections.

- **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.

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

- **Mobility**: Cloud Computing means mobility because users can access applications through internet easily at any point of time.

III. LITERATURE REVIEW

Job scheduling in cloud computing has attracted great attention. Most research in job scheduling adopt a paradigm in which a job in cloud computing system is characterized by its workload, dead-line and the corresponding utility obtained by its completion before deadline, which are factors considered in devising an effective scheduling algorithm. This paradigm is known as Utility Accrual (UA) paradigm.

Many researchers have proposed different scheduling algorithms that run under cloud computing environment. Most of the scheduling algorithms that have been proposed attempt to achieve two main objectives namely, to run the user task within the deadline and to maintain efficiency (load balancing) and fairness for all tasks Here, we reviewed the most relevant research works done in the literature for job scheduling in cloud computing.

Garg et al. (2009) [3] addressed "the issue of increases in energy consumption by data centers in cloud computing. A mathematical model for energy efficiency based on various factors such as energy cost, CO2 emission rate, HPC workload and CPU power efficiency was proposed. In the model a near-optimal scheduling algorithm that utilizes heterogeneity across multiple data centers for a cloud provider was introduced".

Li et al. (2009) [4] introduced "a novel approach named EnaCloud, which enables application live placement dynamically with consideration of energy efficiency in a cloud platform. They use a VM to encapsulate the application, which supports the applications scheduling and live migration to minimize the number of running machines to save energy".

Praveen Gupta et. al. (2010) [5] described "Cloud computing has come out to be an interesting and beneficial way of changing the whole computing world. In this paper, we deal with the various methodologies adopted to handle all the processes and jobs concurrently executing and waiting into the web application and web server housed into the same system or different systems. Also, these different methods will be compared taking into account the same number of jobs, but varied environmental conditions and hence, the result would be formulated. Various issues like virtual resources, queuing strategies, resource managers etc. has been discussed here apart from the main coverage points. All these aspects will be closely studied, observed and proved with proper explanations".

The study in (Yang et al., 2011) [6] highlighted "the issue of job scheduling in cloud computing. They argued that there is no well-defined job scheduling algorithm for the cloud that considers the system state in the future. The existing job scheduling algorithms under utility computing paradigm do not take hardware/software failure and recovery in the cloud into account. To tackle this issue they proposed a Reinforcement Learning (RL) based algorithm that helps the scheduler in making scheduling decision with fault tolerable while maximizing utilities attained in the long term".

Li et al. (2011) [7] introduced "a hybrid energy efficient scheduling algorithm using dynamic migration that handles job execution in private clouds. The algorithm concentrates on reducing the response time, con-serves more energy and performs higher level of load balancing. In addition, the work in (Lin et al., 2011) concentrated on the issue of power consumption in data centers. They proposed a scheduling policy named Dynamic Round-Robin (DRR) that effectively reduces power consumption for virtual machine scheduling and consolidation. The algorithm attempts to deploy the virtual machines to servers and migrate virtual machines among servers".

The study in (Sindhu and Mukherjee, (2011) [8] presented "two scheduling algorithms for scheduling task in cloud computing, taking into account their computational complexity and the computing capacity of the processing elements. The algorithms are designed for private cloud environment where the resources are limited. The first algorithm is named Longest Cloudlet Fastest Processing Element (LCFPE) which considers the computational complexity of the

cloudlets in the process of making scheduling decisions. The second algorithm is named Shortest Cloudlet Fastest Processing Element (SCFP). In this algorithm, the shorter cloudlets are mapped to Processing Elements (PEs) having high computational power so as to reduce flow time while at the same time taking into account that longer jobs are not starved".

In (Paul and Sanyal, 2011) [9] authors discussed "the issue of how to utilize cloud computing resources proficiently and gain maximum profits with the job scheduling system. For this purpose, they proposed a credit based scheduling algorithm to evaluate the entire group of tasks in the task queue and find the minimal completion time of all tasks. The proposed scheduling method considers the scheduling problem as an assignment problem in mathematics where the cost matrix gives the cost of a task to be assigned to a resource. However, the algorithm does not consider the processing time of a job, but other issues are considered such as the probability of a resource to be free soon after executing a task so that it will be available for the next waiting job".

Vijindra and Sudhir Shenai. A (2012) [10] in their paper, have presented "an algorithm for a cloud computing environment that could automatically allocate resources based on energy optimization methods. Then, we prove the effectiveness of our algorithm. In the experiments and results analysis, we find that in a practical Cloud Computing Environment, using one whole Cloud node to calculate a single task or job will waste a lot of energy, even when the structure of cloud framework naturally support paralleled process. We need to deploy an automatic process to find the appropriate CPU frequency, main memory's mode or disk's mode or speed. We have also deployed scalable distributed monitoring software for the cloud clusters".

Neetu Goel, Dr. R.B. Garg (2012) [11] wrote a paper. The paper presented "a state diagram that depicts the comparative study of various scheduling algorithms for a single CPU and shows which algorithm is best for the particular situation. Using this representation, it becomes much easier to understand what is going on inside the system and why a different set of processes is a candidate for the allocation of the CPU at different time. The objective of the study is to analyze the high efficient CPU scheduler on design of the high quality scheduling algorithms which suits the scheduling goals".

Monica Gahlawat, Priyanka Sharma (2013) [12] described "Scheduling as a set of policies and mechanisms to control the order of work to be performed by a computer system. CPU is by far the most important resource of the computer system. Recent advances in software and architecture of the system increased the complexity of the processing as the computing is now distributed and parallel. Job scheduling is complex in this environment. The VM (Virtual Machine) can use a distinctive VCPUs (Virtual CPU) running queue for each physical CPU, which is referred to Partition Queue Model (PQM). As a contrast, a Sharing Queue Model (SQM) of CPU scheduling algorithm can be used. This paper is analyzing and evaluating the performance of various CPU scheduling in cloud environment using CloudSim".

Swachil Patel, Upendra Bhoi (2013) [13] described that "in cloud computing, there are many jobs requires to be executed by the available resources to achieve best performance, minimal total time for completion, shortest response time, utilization of resource usage and etc. Because of these different objectives and high performance of computing environment, we need to design, develop, propose a scheduling algorithm to outperform appropriate allocation map of jobs due to different factors. In job scheduling priority is the biggest issue because some jobs need to scheduled first then the other jobs which can wait for a long time. In this paper, a systematic review of various priority based job scheduling algorithms is presented. These algorithms have different perspective, working principles etc. This study concludes that all the existing techniques mainly focus on priority of jobs and reduces service response time and improving performance etc. There are many parameters that can be mentioned as factor of scheduling problem to be considered such as load balancing, system throughput, service reliability, service cost, service utilization and so forth".

Vignesh V, Sendhil Kumar (2013) [14] wrote a paper. "The goal of this paper was to propose a model for job-oriented resource scheduling in a cloud computing environment. Resource allocation task is scheduled for the Process which gives the available resources and user preferences. The computing resources can be allocated according to the rank of job .This paper constructs 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 taken into consideration. From this, it has been computed that SRTF has the lowest time parameters in all respects and is the most efficient algorithm for resource scheduling".

Dilshad H. Khan, Deepak Kapgate (2014) [15] wrote a paper. In this paper, "they explained different algorithms and techniques proposed for Virtual Machine Scheduling either at single data centre or multiple data center. Also infers their characteristics to resolve the issue of efficient Virtual Machine Management in Cloud Computing. We discuss and compare these algorithms and techniques in regards of various performance matrices to provide an overview of the latest approaches in the field".

Lipsa Tripathy, Rasmi Ranjan Patra (2014) [16] 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".

Nima Jafari Navimipour and Farnaz Sharifi Milani (2015) [17] explained that "the task scheduling problem in Cloud computing is an NP-hard problem. Therefore, many heuristics have been proposed, from low level execution of tasks in multiple processors to high level execution of tasks. In this paper, a new evolutionary algorithm is proposed which named CSA to schedule the tasks in Cloud computing. CSA algorithm is based on the obligate brood parasitic behavior of some cuckoo species in combination with the Levy flight behavior of some birds and fruit flies. The simulation results demonstrated that when the value of Pa is low, the speed and coverage of the algorithm become very high".

S.Sujan, R.Kanniga Devi (2015) [18] wrote a paper. In this paper, "a dynamic scheduling scheme for cloud computing was discussed. Considering the resource provisioning as the main issue to be address, the scheduling is being considered as the context. Here we use a dynamic scheduling scheme considering the makespan as the metric. Algorithms like min-min and round robin are compared with respect to the proposed scheme. A real time scenario based model named Berger model is taken in account of comparison with the existing conventional algorithms. By considering and evaluating all the above given constraints and strategies, a makespan based dynamic scheduling scheme for cloud computing is being proposed".

Sagnika Saha, Souvik Pal (2016) [19] in their paper offered "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".

Sushil Kumar Saroj, Aravendra Kumar Sharma (2016) [20] 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".

Akilandeswari. P and H. Srimathi (2016) [21] 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".

Shridhar Domanal, Ram Mohana Reddy Guddeti, and Rajkumar Buyya (2016) [22] wrote a paper. In this 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".

IV. CONCLUSION

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 other 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 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] Garg, S.K., C.S. Yeo, A. Anandasivam and R. Buyya, 2009. Energy-efficient scheduling of HPC applications in cloud computing environments. Comput. Sci. Distributed, Parallel Cluster Computing.

[4] Li, B., J. Li, J. Huai, T. Wo and Q. Li et al., 2009, "EnaCloud: An energy-saving application live placement approach for cloud computing environments", Proceedings of the International Conference on Cloud Computing, Sept. 21-25, IEEE Xplore Press, Bangalore, pp: 17-24. DOI: 10.1109/CLOUD.2009.72

[5] Gupta, P.K. and N. Rakesh, 2010. "Different job scheduling methodologies for web application and web server in a cloud computing environment", Proceedings of the 3rd International Conference on Emerging Trends in Engineering and Technology, Nov. 19-21, IEEE Xplore Press, Goa, pp: 569-572. DOI: 10.1109/ICETET.2010.24

[6] 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

[7] Li, J., J. Peng and W. Zhang, 2011, "A scheduling algorithm for private clouds", J. Convergence Inform. Technol.", 6: 1-9. Li, J., M. Qiu, J. Niu, W. Gao and Z. Zong et al., 2011.

[8] Sindhu, S. and S. Mukherjee, 2011, "Efficient task scheduling algorithms for cloud computing environment". Commun. Comput. Inform. Sci., 169: 79-83. DOI: 10.1007/978-3-642-22577-2_11

[9] 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.

[10] 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.

[11] 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.

[12] Monica Gahlawat, Priyanka Sharma, "Analysis and Performance Assessment of CPU Scheduling Algorithms in Cloud using Cloud Sim", International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 5 – No. 9, July 2013.

[13] Swachil Patel, Upendra Bhoi, "Priority Based Job Scheduling Techniques In Cloud Computing: A Systematic Review", International Journal Of Scientific & Technology Research VOLUME 2, ISSUE 11, NOVEMBER 2013.

[14] 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.

[15] 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

[16] Lipsa Tripathy, Rasmi Ranjan Patra, "Scheduling In Cloud Computing", International Journal on Cloud Computing: Services and Architecture (IJCCSA) ,Vol. 4, No. 5, October 2014.

[17] 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.*

[18] 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.*

[19] 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.

[20] 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)

[21] 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.

[22] 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.