A Survey of Scheduling Policies in Cloud Computing Environment

Aditya Makwe[#1], Priyesh Kanungo[#2]
[#1]M Tech Graduate, SCSIT, DAVV, Indore
[#2]Professor and Senior Systems engineer, SCSIT, DAVV, Indore

Abstract-*Cloud computing has created a new perception of computation with all its features and characteristics. With its profound computational capabilities, it provides a platform for solving complex computational problems. Cloud computing deals with different kinds of virtualized resources and scheduling plays a vital role in cloud computing. This paper focuses on that various scheduling policies with their features and utility. We give an elaborate idea of cloud computing and elaborate the advantages and disadvantages of various scheduling policies.*

Keywords— *Cloud Computing, Grid Computing, Data centre, Virtualization, Analytic Hierarchy process (AHP) model, Quality of Service (QoS), Gang Scheduling, Market Oriented Scheduling*

I. INTRODUCTION

Cloud computing is one step ahead of Grid Computing [2], reducing the cost of computing increasing reliability, flexibility by transforming computing as a utility. It is a form of specialized distributed computing [9] paradigm that is massively scalable and can be encapsulated as an abstract entity that delivers different forms of services to the customer These services are delivered by using the concept of virtualization or other approaches on user demands[10]. Efficient resource management is a very important aspect from cloud computing perspective and virtualization concept provides platform for completing the provisioning requirement of cloud based solution [5]. Using virtualization concept, a hardware entity is turned into software based component called as virtual machine (VM). Main advantage of this technique is that multiple VM's can run over a single physical host which improves resource utilization.

Since cloud computing is in its development state, a number of efforts and contribution are been made by researcher in its different areas like cloud system development and its management, resource allocation, scheduling [17], security issues, cloud storage etc. Different authors and organizations have given a different definition of cloud computing, but the community has considered the standard definition given by the National Institute of Standard and Technology (NIST) as it covers all the essential aspect of cloud computing. It states that "Cloud computing is a model for enabling ubiquitous, 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" [16]. The cloud computing has three primary service models based on the services provides viz. Software as Service (SaaS), Platform as Service (PaaS), Infrastructure as Service (IaaS). Four deployment models are also available based on the user's e.g. public, private, hybrid and community cloud. Also, cloud has five essential characteristics viz. on-demand self service, broad network access, resource pooling, rapid elasticity and metered services. In cloud environment, user may have a number of virtualized resources [12]. To allocate these virtual resources to users manually is a cumbersome process. With the development of virtualization technology, cloud computing leaves job scheduling complexity to virtual machine layer. In order to allocate resources to every job in an efficient manner and to achieve high throughput by fair utilization of resources, scheduling plays a key role.

II. CLOUD SCHEDULING POLICIES

Cloud scheduling algorithms can be categorized into batch mode heuristic scheduling algorithms (BMHA) and online heuristic algorithms. In BMHA, jobs are queued and collected into a set when they arrive in the system. Examples of BMHA based algorithms are; First Come First Served scheduling algorithm (FCFS), Round Robin scheduling algorithm (RR), Min–Min algorithm and Max–Min algorithm. In on-line model heuristic scheduling algorithm, jobs are scheduled when they arrive in the system. Since the cloud environment is a heterogeneous system and the speed of each processor varies quickly, on-line mode heuristic scheduling algorithms are appropriate for a cloud computing environment.

The cloud scheduling [4] can be static or dynamic. Static approach is used where the number of required virtual resources is constant (for example, a fixed-size service, a virtual classroom, etc.), and the cloud provider conditions (resource availability prices, , etc.) do not change throughout the service life-cycle. Resource selection can be online, once only, and in advance to service deployment. Contrary to this, the dynamic [13] approach is suitable for variable size services (e.g., a web server cluster with a fluctuating resource requirements), or

in the case of changing cloud provider conditions (dynamic resource availability, variable prices, etc.)

A. The Task Scheduling and resource allocation in Cloud computing environment using Analytic hierarchy process:

Management of Resource in Cloud Computing environment is a complex task. To solve this issue Ergu et al. presented a scheduling model based on the Analytic Hierarchy process model (AHP) where tasks are collected in a task pool. These tasks are then ranked by using pair-wise comparison matrix techniques and the AHP describing the available resources and user preferences [1]. The computing resources are then allocated to tasks in terms of their weights. An induced bias matrix is proposed to identify the inconsistent element and to improve the consistency ratio for solving the conflicting issues among the tasks. Tasks ranked using this method improve the resource utilization and also meet the user requirement. However, dynamic allocation of resources is not possible using this method.

B. Priority based job scheduling algorithm:

Ghanbari et al. have proposed a new propriety based job scheduling algorithm in a cloud computing environment based on multi criteria decision making approach, using the Analytic Hierarchy Process model [3]. They have also discussed some issues related to algorithm like complexity, consistency and finish time. The main disadvantage of this algorithm is that it is reasonably complex and does not calculate the finish time of tasks while the response time is poor. For a large number of tasks, this algorithm does not perform well since dynamic calculation of priority of tasks is a complex.

C. Adaptive Market-oriented Scheduling policies for Cloud Computing:

Provisioning of additional resources is necessary when the local resources does not complete user requirement. Buyya et al. Have developed two market oriented policies aiming at application deadline by increasing the computational capacity of local resources via hiring the resources from an Infrastructure as a Service (IaaS) provider [14]. The policies consider user provided deadlines and budget as a scheduling parameter. The policies don't have any prior knowledge of the applications' execution times. Time optimization scheduling policy minimizes the application completion time, whereas cost optimization scheduling policy minimizes cost involved in running the application. Furthermore, the authors conclude that efficiency of cost optimization and time optimization policies can be potentially increased by increasing the budget. By using the proposed policy, different work load types can get completed before the deadline within the budget.

D. Analysis of Variants in the Round Robin Algorithm for Load Balancing in Cloud Computing:

In order to manage the resources in cloud environment, load balancing is required for the user tasks that are submitted for execution. Samal et al. have used the concept of Round-Robin for distributing the load equally among the compuitng resources. The scheduler starts with a node and moves on to the next node. After assigning a Virtual Machine (VM) to the node, this process continues until all nodes have at least one VM and returns back to starting node [15]. Although the Round-Robin algorithm is based on simple rule, more loads are conceived on servers thus resulting in unbalancing the traffic.

E. An energy efficient Scheduling approach based on Private Clouds:

Private clouds have some unique characteristics and requirements. To schedule a virtual machine requests on computing nodes in such environment is a challenging task. To address this problem, Zhang et al. have discussed about Hybrid energy efficient Scheduling application based on pre-power techniques and least load first algorithm [6]. They described two problems for VM scheduling. Pre-power technique uses idle threshold value and improves the response time. Least load scheduling algorithm is used to balance workloads when the data centres are running in low power mode.

F. Zhang's Scheduling algorithm for private clouds:

Zhang et al. have developed a hybrid energy-efficient algorithm using dynamic migration. In this algorithm the power efficiency is improved by using the power up command to wake up the sleep node as well as idle node [7].

G. Heterogeneity aware Resource Allocation and Scheduling in Cloud:

To improve performance and cost-effectiveness of a data analytics cluster in the cloud, the data analytics system should account for heterogeneity of the environment and work load [8]. In this Algorithm Lee et al. have described cost effective resource allocation and job scheduling scheme for a heterogeneous cluster. The data analytics system must report for heterogeneity of the workloads. To address fairness among multiple jobs sharing the cluster, architecture has been designed to allocate resources to a data analytics cluster in the cloud environment.

H. Evaluation of Gang Scheduling Performance and Cost in a Cloud Computing System:

Ioannis et al. developed an efficient job scheduling algorithm for time sharing, using gang scheduling.

They analysed the performance of a distributed cloud computing model based on Amazon Elastic Compute Cloud (EC2) architecture that implements a Gang Scheduling scheme [11]. They also estimated the performance and overall cost of two main gang scheduling algorithm, Adaptive first come first served (AFCFS) and largest job first (LJFS) both algorithm can be applied in an environment where virtual machines (VM) vary dynamically.

I. Cloud-DLS, Dynamic Trusted Scheduling for Cloud:

Wang et al. developed a trusted dynamic scheduling algorithm named Cloud-DLS. Getting firmness in computing resources is difficult due to the characteristics of cloud computing [18]. The used Bayesian method based cognitive trust model and true relationship model of sociology. To decrease the failure probability of the task assignment and assurance of the execution of tasks in security environment, a trusted dynamic level scheduling algorithm is proposed. Primary objective of was to extend the traditional formulation of the scheduling problem so that both execution time and reliability of application are simultaneously accounted for scheduling system.

J. Job scheduling algorithm based on Berger model in cloud environment:

Commercialization and virtualization are the main characteristics of Cloud Computing. Considering these characteristics, Xu et al. have described the scheduling algorithm based on Berger model. The algorithm contains two constraints. In the first constraint, user tasks are classified of on the basis of quality of service (QoS) preferences. In the Second constraint is to define resource fairness, justice function to judge the fairness of the resource allocation [19]. This model agrees with QoS parameters like completion time and bandwidth requirements..

III. CHARACTERISTICS OF SCHEDULING ALGORITHMS

Performance of the system has major impact on the revenue of service provider in cloud environment, therefore scheduling becomes of jobs becomes a crucial task in a cloud framework. Since scheduling is a NP hard problem, there is always scope of improvement in a algorithm. A good scheduling framework should focus on:

***A.* Energy consumption:**

It is the degree to which the resources of the system are being utilized. A good scheduling algorithm must focus on Energy efficiency and load balancing of the Datacentres.

B. QoS Parameters:

These parameters are determined by the user and includes execution time, cost etc.

C. Security features:

It should satisfy the security needs.

D. Fair resource allocation:

The fairness in allocation of resources effects the overall performance.

E. Resource utilization:

The overall revenue depends on efficient utilization of resources A good scheduling algorithm must focus on maximum utilization of resources. Table 1 depicts the advantage and disadvantages of the above policies.

IV. TABLE I
COMPARISON BETWEEN DIFFERENT SCHEDULING POLICIES

S.No	Algorithm	Methodology used	Scheduling parameters considered	Advantage	Dis-advantage
1	The Analytic hierarchy process: Task Scheduling and resource allocation in Cloud Computing environment	Theory of AHP model and reciprocal pair-wise comparison matrix is used for ranking of tasks	Response time, task expenses	Improve resource utilization	Dynamic allocation of tasks is not possible.
2	A priority based job scheduling algorithm in the Cloud Computing environment	Theory of AHP is used	Make span	Considers the important tasks first based on priority	Increase Make span
3	Adaptive Market-oriented Scheduling	Deadline and budget constraint based time and	Response time, execution time	Increases computation	Completion time is more

	policies for Cloud Computing	cost optimization policy		al capacity by hiring resources from IaaS Cloud	
4	Analysis of variants in the Round Robin algorithm for load balancing in Cloud Computing	Based on Round Robin scheduling technique	Response time, load balancing	Fairness in allocation of VM's	Unbalancing of traffic when load increases
5	An energy efficient Scheduling approach based on Private Clouds	Energy efficient algorithm using dynamic migration	Execution time, energy consumption	Improved power efficiency	Can't meet user requirements efficiently
6	A scheduling algorithm for private clouds	Energy efficient algorithm using dynamic migration	Response time, energy consumption, load balancing	Improved power efficiency	Can't meet user requirements efficiently
7	Heterogeneity aware Resource allocation and scheduling in cloud	Data analytic cloud architecture	Response time, cost	Fairness	Tasks with limited instances may not perform well
8	Evaluation of gang scheduling performance and cost in a cloud computing system	Based on gang scheduling	Response time, cost	Support dynamic environment	Complex model
9	Cloud-DLS : Dynamic trusted scheduling for cloud computing	Based on Bayesian cognitive model and trust relationship models of sociology	Execution time	Meets the requirement of cloud computing work load	Does not meets all user requirements
10	Job scheduling algorithm based on Berger model in cloud environment	Based on Berger model	Execution time	Fairness	Complex model

V. CONCLUSIONS

Cloud Computing is a multi featured platform that helps to get the maximum benefits of computational resources and aggregate them to achieve the high throughput of the system by solving the complex problems under the define criteria. In order to manage the resources a good scheduling policy is required. In this paper, a survey of various scheduling policies of cloud computing environment has been provided. The scheduling algorithm should consider various parameters (execution time, cost, reliability, energy efficiency, performance, makespan) for allocation of resources, whereas addressing all parameters in a single algorithm is a cumbersome process.

Template was provided by courtesy of Causal Productions (www.causalproductions.com)".

REFERENCES

[1] Ergu, D., Kou, G., Peng, Y., Shi, Y. and Shi, Y. 2011. The analytic hierarchy process: task scheduling and resource allocation in cloud computing environment. *The Journal of Supercomputing*. (2011).

[2] Foster, I., Zhao, Y., Raicu, I. and Lu, S. 2008. Cloud Computing and Grid Computing 360-Degree Compared. *2008 Grid Computing Environments Workshop*. (2008), 1–10.

[3] Ghanbari, S. and Othman, M. A priority based job scheduling algorithm in cloud computing. *Procedia Engineering*. 50, Icasce (2012), 778–785, 2012.

[4] Kanungo P "Exploring Load Balancing To Solve Various Problems In Distributed Computing"*International Journal of Computer Trends and Technology (IJCTT)*,V4(10):3660-3664 October Issue 2013 .ISSN 2231-2803.www.ijcttjournal.org. Published by Seventh Sense Research Group.

[5] K C Gouda, Anurag Patro , Dines Dwivedi , Nagaraj Bhat."Virtualization Approaches in Cloud Computing". International Journal of Computer Trends and Technology (IJCTT) V12(4):161-166, June 2014. ISSN:2231-2803. www.ijcttjournal.org. Published by Seventh Sense Research Group.

[6] Li J, Peng J,Lei Z, Zhang W, An Energy-efficient Scheduling Approach Based on Private Clouds, Journal of Information & Computational Science. April 2011.

[7] Li J, Peng J, Lei Z, Zhang W, A Scheduling Algorithm for Private Clouds, Journal of Convergence Information Technology, Volume6, Number 7, 2011.

[8] Lee G, Chun B G, Katz R H, Heterogeneity-Aware Resource Allocation and Scheduling in the Cloud.

[9] Mehta, H., Kanungo**,** P**.** and Chandwani, M., "Towards Development of a Distributed e-Learning EcoSystem," *2nd International Conference on Technology for Education (TforE-2010)*, IIT Mumbai, July 1-3, 2010.

[10] Mahiba, T. 2013. Live Virtual Machine Migration in Dynamic Resource Management of Virtualized Cloud Systems. 2, 4 (2013), 226–233.

[11] Moschakis, I. A. and Karatza, Evaluation of gang scheduling performance and cost in a cloud computing system, *Journal of Supercomputing*, Springer Publications, 2010

[12] Makwe A. and Kanungo P., Scheduling in Cloud Computing Environment Using Analytic Hierarchy Process Model IEEE International Conference, Medicaps Institute of Technology and Science, Indore 10-12 Sept 2015

[13] Sharma, R.K., Kanungo**,** P. and Chandwani, M., "Performance enhancement of Network of Workstations using Dynamic Load Balancing," *Bhartiya Vigyan Sammelan, Indore*, 1-3 December, 2009.

[14] Salehi, M.A. and Buyya, R.. Adapting market-oriented scheduling policies for cloud computing. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 6081 LNCS, PART 1 (2010), 351–362, 2010.

[15] Samal P and Mishra P, Analysis of variants in Round Robin Algorithms for load balancing in Cloud Computing, International Journal of Computer Science and Information Technologies, pp. 416-419, Vol. 4 (3), 2013.

[16] The NIST Definition of Cloud computing http://www.nist.gov/itl/csd/cloud-102511.cfm

[17] Tiwari A. and Kanungo P**.**, "Dynamic Load Balancing Algorithm for Scalable Heterogeneous Web Server Cluster with Content Awareness," 2nd International Conference on Trendz in Information Sciences & Computing, (TISC) 2010, Satyabhama University, Chennai, India, pp. 143-148 (Print ISBN: 978-1-4244-9007-3, Paper available on IEEE Xplore.

[18] Wang, W., Zeng, G., Tang, D. and Yao, J. 2012. Cloud-DLS: Dynamic trusted scheduling for Cloud computing. *Expert Systems with Applications*. 39, 3 (2012), 2321–2329, 2012.

[19] Xu, B., Zhao, C., Hu, E. and Hu, B. 2011. Job scheduling algorithm based on Berger model in cloud environment. *Advances in Engineering Software*. 42, 7 (2011), 419–425.