Volume 8, No. 5, May-June 2017
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
REVIEW ARTICLE
Available Online at www.ijarcs.info

Scheduling Policies in Cloud Computing Environment : A Review

Akhil Chaurasia
Department of Computer Science and Engineering
Sant Longowal Institute of Engineering and Technology
Deemed University, Longowal, Sangrur, India

Jaspal Singh
Department of Computer Science and Engineering
Sant Longowal Institute of Engineering and Technology
Deemed University, Longowal, Sangrur, India

Abstract: In this modern era, cloud computing has become one of the fastest emerging technology. Cloud computing offers services as a request-response model, where a customer can request for computing resource in a pay-as-you-use manner. With its growing application and popularization, IT companies are rapidly deploying distributed data centers globally, posing numerous challenges in terms of scheduling of resources under the different administrative domain. Due to the increasing demands of the users for cloud services or resources, it becomes a complex problem of finding the best match of resource-workload combination and scheduling of suitable resources to cloud workloads depends on the QoS requirements of cloud applications. Researchers still face difficulties to select the efficient and suitable resource scheduling algorithm for a specific workload from the standing literature of resource scheduling algorithms. This paper gives an overview of the existing scheduling policies in cloud computing systems. The literature concerning to fifteen types of scheduling algorithms has been stated. Systematic analysis of this research work will help researchers to find the important characteristics of scheduling policies in cloud computing.

Keywords: Scheduling Policies, resource scheduling, task scheduling, virtual machine

I. INTRODUCTION

Cloud Computing, the long-held dream of computing as a utility, has the potential to transform a large part of the IT industry, making the software even more attractive as a service and shaping the way in which hardware is designed and purchased. The National Institute of Standards and Technology (NIST) defines, "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 [1]." Cloud computing is a utility-based system which gives users on-demand access to a shared pool of computing resources over the network. These resources can include network, server, storage, application, services and so on. They can be provisioned quickly as a service rather than a product as and when needed and released after usage without much effort from the user side. Cloud has three service models and four deployment models. Some of the features of cloud computing include flexibility, maintenance, multi-tenancy, cost etc. The service models of the cloud include Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). It is revolving upside down the way it comprehends computation by enabling the use of. It is also important to ensure that the cloud delivery models [2] provide common access control interfaces for appropriate interoperability, which demands a policy neutral access control requirement and implementation framework that can be used to address. The deployment models of the cloud include public cloud, private cloud, hybrid cloud and community cloud. Efficient scheduling policies should be used in order to make use of the cloud capabilities efficiently. Many researches have been carried out in the arena of resource allocation and scheduling in the cloud. The main enthusiasm of these scheduling policies is to minimize the execution cost and time and to increase resource utilization. The objective of this paper is to focus on various scheduling policies in cloud computing system. Since cloud computing is in initial stages, different kinds of researches are going on, the research areas which include cloud systems development and management, resource allocation and scheduling, security issues, cloud storage, elastic scalability, programming models etc. In cloud computing, user may face thousands of virtualized resources to utilize, it is impossible for anyone to allocate the jobs manually. Due to the virtualization properties, cloud computing leaves job scheduling complexity to the virtual machine layer through resource virtualization. Hence to allocate the resources to each job efficiently, scheduling plays more important role in cloud computing. In addition, difference in computing sources in different nodes adds to the complexity of task scheduling. Furthermore, common data exchange among nodes, hosts, and clusters in data-intensive cloud applications makes the task-scheduling procedure extremely complicated. Most of these methods focus on allocating CPU and memory resources to various cloud-computing tasks, assuming that all physical nodes and VMs have unlimited network bandwidth. Resource provisioning has effects in many areas like operational cost, quality of service, policy of the cloud provider, so it has a high impact while setting up a cloud system. Rest of the paper will be devoted towards analysis of cloud computing and its association with workload prediction. In next section of paper, features of cloud computing, method of resource provisioning, prediction algorithm, various tools and factors affecting implementation of cloud system have been discussed.

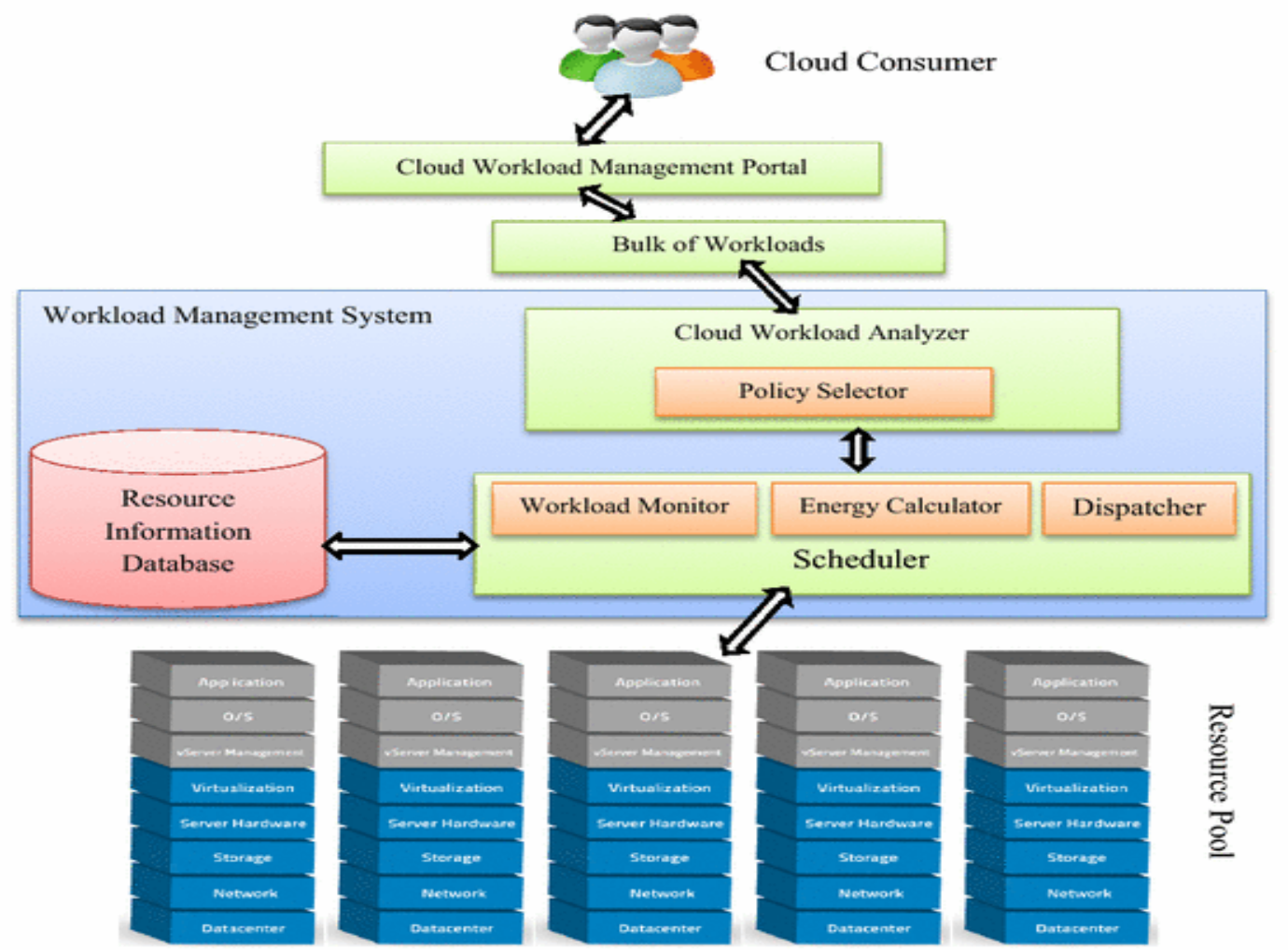

Figure1. Resource scheduling in Cloud computing

II. II. EXISTING SCHEDULING ALGORITHM

The following scheduling algorithms are presently established in the area of grids, clouds and various workflows and these algorithms have been summarized in table 1with the scheduling parameters.

A*: *Optimal Cloud Resource Provisioning Algorithm

Sivadon et al. [3] proposed an optimal cloud resource provisioning (OCRP) algorithm is presented to solve the difficulties caused by the uncertainty of consumer's future requirement and provider's resource prices. This OCRP is obtained by formulating and solving a stochastic integer programming problem. The proposed algorithm can provision computing resources for being used in multiple provisioning stages as well as a long term plan. Here the demand and price uncertainty is considered. Various approaches like deterministic equivalent formulation, sample-average approximation and Bender's decomposition are considered to obtain the solution for OCRP algorithm.

B: Resource Scheduling Strategy based on Genetic Algorithm

Jianhua Gu et al. proposed [4] a scheduling strategy on load balancing of virtual machine resources using Genetic Algorithm (RSGA). It uses historical data and current states of VMs. In the proposed method starting from the initialization in cloud itself they look for the best scheduling solution by genetic algorithm in every scheduling and when there are no VM resources in the whole system use the algorithm to choose scheduling solution according to the computed probability. Even though this method can better realize load balancing and proper resource utilization, it does not deals with the dynamic behavior of resource allocation.

C*: *An energy-efficient scheduling approach based on private cloud

Jiandun li et al. proposed [5] a hybrid energy efficient approach, which is comprised of pre-power technique and least load first algorithm and the proposed approach can save more time for users, conserve more energy and achieve higher level of load balancing. Since private clouds have some unique characteristics and special requirements, it is still a challenging problem to effectively schedule virtual machine requests onto compute nodes, especially with multiple objectives to meet. Mainly two problem of resource scheduling are discussed. Pre-power technique is used to reduce the response time and it uses idle threshold value where least load first algorithm is used to balance the workloads when data centers are running on low power mode.

D: Energy-Efficient Scheduling Scheme (EESS) for Virtual Machines in Cloud Computing

Shailesh and Ashok [6] proposed Energy-Efficient Scheduling Scheme calls EESS for virtual machines that distribute maximum workload on minimum number of virtual machines. Thus, the consumption of energy will be in low amount, to test the proposed approach cloud environment created using Virtual Box. Sun-Oracle provide open source virtualBox3.1 support virtualization, para-virtualization, Xen support para-virtualization, DVF simulator run processor on low frequency, voltage so amount of energy conserve. EESS distribute workload on less number of virtual machines by using migration, pause, resume, clone of virtual machines. Energy-efficient scheduling scheme (EESS) is beneficial for power generation plant and their survive problems.

E*: *A scheduling algorithm for private cloud

Jiandun li et al. [7] proposed a hybrid energy efficient approach using dynamic migration. This paper is based on [1], but there is a issue related to the using of threshold value, powering down a busy nod is not feasible. Hence an expected spectrum set for the left capacity is used. It uses power up command to wake the sleep node as well as the idle node. Hence power efficiency is improved.

F: Power-aware provisioning of Cloud Resources for Real Time Services

Kyong Hoon Kim et al. [8] proposed power-aware provisioning approach of virtual machines for real-time services. This approach model a real-time service as a real-time virtual machine request and provision virtual machines of datacenters using DVFS (Dynamic Voltage Frequency Scaling) schemes. The proposed algorithm shows better reduced power consumption with high performance. A real-time Cloud service framework has proposed where each real-time service request is modeled as RT-VM in resource brokers. The proposed adaptive schemes, Adaptive-DVS and δ-Advanced-DVS, show more profit with less power consumption regardless of system load.

G: An ANT Colony Algorithm for balanced job scheduling in grids

Ku Ruhana and Husna Jamal [9] proposed an enhanced ant colony optimization algorithm for jobs resource scheduling in grid computing. A balanced Ant colony algorithm which uses pseudo random proportional rule to balance the entire system load while completing all the jobs at hand as soon as possible according to the environment status. Current scientific problems are very complex and need huge computing power and storage space. To utilize the grid resources efficiently, balanced ant colony algorithm is proposed by balancing the workload as well as minimizing the completion time of each job.

H*: *Job scheduling algorithm based on dynamic management of resources provided by grid computing system

Ungurean proposed [10] an algorithm of job scheduling and dynamic adjustment of nodes loading within a grid system. Within a distributed computing system, requests of processing are randomly received from the system's users. A good planning of these requests assumes their assigning towards available processors, so that all requests have to be solved as soon as possible. Considering the resources sharing in grid systems, a job scheduling algorithm are proposed with dynamic load balancing is proposed. The distribution of first come, first served (FCFS) with a round robin mechanism of the execution nodes is proposed.

I: Scheduling of scientific work flow using chaos genetic algorithm

Golnar Gharooni-fard et al. [11] proposed a novel genetic algorithm called chaos-genetic algorithm which is used to solve the scheduling problem considering both user's budget and deadline. Due to the nature of chaotic variables such as pseudo-randomness, ergodicity and irregularity, the evolutional process of chaos-genetic algorithm makes individuals of sub generations distribute ergodically in the defined space and circumvents the primary of the individuals of traditional genetic algorithms (TGA). In grid environment, by using the characteristic of chaotic variable in scattering the solutions among the whole search space and thus avoids the precipitate convergence of the solutions and produces better results within a shorter time. Investigation of scheduling workflows considering the QoS constraints (user budget, deadline) has done.

J: Evaluation of gang scheduling performance and cost in a cloud computing system

Ioannis and Helen [12] proposed a Gang Scheduling algorithm which is an efficient job scheduling algorithm for time sharing, already applied in parallel and distributed system. This paper is to study the performance of a distributed Cloud Computing model, based on the Amazon Elastic Compute Cloud (EC2) architecture and to revise, study, and estimate both the performance and the overall cost of two foremost gang scheduling algorithms. It utilizes the concept of virtual machines which acts as the computational units of the system. The proposed system implemented for adding and removing virtual machines from the system depending on the systems load at any specific time.

K: A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environment

Suraj et al. [13] proposed Meta heuristics method based on particle swarm optimization proposed. In grid environment, user applications may incur large amount of data retrieval and execution costs when they are scheduled taking into account only the execution time. Added to that optimizing the execution time, the cost arising from data transfer between resources as well as execution costs must also be taken into account, and focusing to minimize the total execution cost of applications on resources. PSO's ability to find near optimal solutions for mapping all the tasks in the workflow to the given set of computer resources. It takes both computation and communication cost into account, if the resource cost increases PSO minimizes the total maximum cost assigning all tasks to resources.

L: Cloud-DLS. Dynamic trusted scheduling for Cloud computing

Wei Wang et al. [14] proposed two algorithms, first a novel

Bayesian method based cognitive trust model, and second a trust dynamic level scheduling algorithm named Cloud-DLS by integrating the existing DLS algorithm. Because of the characteristics of cloud computing, obtaining trustworthiness in computing resources is difficult. Novel Bayesian method based cognitive trust model, trust relationship models of sociology used.This paper focuses on trustworthiness in cloud computing. Cognitive trust model is used. Two kinds of trust that is direct trust degree; recommendation trust degree is obtained to obtain the trusted scheduling, and extends the traditional DLS algorithm by considering trustworthiness of resource nodes. This algorithm meets the requirement of user tasks in trust, and makes tasks scheduling based on directed acyclic graph (DAG) more reasonable.

M: A Framework for Resource Allocation Strategies in Cloud Computing Environment

M. Asad et al. [15] presents a critical evaluation of current network resource allocation strategies their possible applicability in Cloud Computing. Framework for resource allocation problem based on online tailored active measurements was developed. Focusing towards network awareness and consistent optimization of resource allocation strategies and identifies the issues which need further investigation by the research community. There is a need for developing new methods of reliable active measurements aimed at capturing global Internet behavior online, and for enabling prediction of the most critical performance parameters. In this paper, the resource allocation problems are discussed that are based on Artificial Intelligence based, Theory of random graphs, Peer to peer based approaches and resource allocation framework is proposed in cloud computing.

N: A Priority based Job Scheduling Algorithm

Shansollah and Mohamed [16] proposed in cloud computing environment based on multiple criteria decision making model, using analytical hierarchy process. Provided a discussion about some issues related to the proposed algorithm such as complexity, consistency and finish time. The proposed algorithm has reasonable complexity. But the main disadvantage is that the finish time cannot be calculated and response time is more .Also for more number of jobs allocations it is not suitable since finding priority of each job is tedious one.

O: Heterogeneity-Aware Resource Allocation and Scheduling

Gunhoo Lee et al. [17] proposed Heterogeneity-Aware Resource Allocation and Scheduling (HARS) algorithm which allocate a metric of share in a heterogeneous cluster to realize a scheduling policy that gives high performance and fairness. The heterogeneity of the environment should be built along with the performance and cost effectiveness. The data analytics system must report for heterogeneity of the situation and workloads. It also needs to provide fairness among jobs when multiple jobs share the cluster. Hence architecture to allocate resources to a data analytics cluster in the cloud proposed.

Table No. 1 Comparison between Scheduling Policies

Sr. no.	Algorithm/technique used	Scheduling Parameter considered	Goals Achieved
1	Optimal Cloud Resource Provisioning Algorithm	Cost	Reduced cost
2	Resource Scheduling Strategy based on Genetic Algorithm	Number of virtual machines, Execution Time	This method can better realize load balancing and resource utilization
3	An energy-efficient scheduling approach based on private cloud	Response Time, Energy Consumption	Reduced Execution time
4	Energy-Efficient Scheduling Scheme(EESS) for Virtual Machines in Cloud Computing	Execution Time, Load Balancing	Distribute maximum workload on minimum number of virtual machines
5	A scheduling algorithm for private cloud	Execution Time, Energy Consumption	Improve power efficiency
6	Power-aware provisioning of Cloud Resources for Real Time Services	Response Time, Energy Consumption	Reduced power consumption with high performance
7	An ANT Colony Algorithm for balanced job scheduling in grids	Make span, Load Balancing	Balancing the workload as well as minimizing the completion time of each jobs
8	Job scheduling algorithm based on dynamic management of resources provided by grid computing system	Execution Time,	Improved resource provisioning
9	Scheduling of scientific workflows using chaos genetic algorithm	Execution Time, Cost	Produces better results within a shorter time
10	Evaluation of gang scheduling performance and cost in a cloud computing system	Response Time, Cost	Improved performance
11	A Particle Swarm	Cost	Reduce the cost

	Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments		
12	Cloud-DLS. Dynamic trusted scheduling for Cloud computing	Trust	Algorithm meets the requirement of user tasks in trust
13	A Framework for Resource Allocation Strategies in Cloud Computing Environment	Resource Utilization, Load Balancing	Improve resource utilization
14	A Priority based Job Scheduling Algorithm	Make span	Since priority is considered no important task will not be lagged
15	Heterogeneity-Aware Resource Allocation and Scheduling	Response Time, Cost, Fairness	High performance and fairness

III. CONCLUSION

Cloud computing is one of the user oriented technology in which user faces hundreds of thousands of virtualized resources for each task. In this paper we survey various existing scheduling policies in cloud, grid and workflows. Since cloud computing is in infancy state, a scheduling framework should be implemented to improve the user satisfaction along with the service providers. The scheduling parameters can be coupled to prepare a framework for resource allocation and scheduling in cloud computing. The scheduling framework should consider the user input constraints (execution cost, deadlines, transmission cost, energy efficiency, performance issues, and make span) and so on. During scheduling they had considered various techniques and applied constraints but as the cloud computing is too vast that they had not been able to capture all aspects at the same time but they mentioned these facts that there is a chance of modification of algorithms and which part has to be modified. The main objective of the scheduling policies is to maximize utilization of resources and to reduce make span. A lot of policies are proposed to achieve effective scheduling, but since the scheduling in cloud computing is heuristic problem the more research can be done in this field and more optimized solutions can be achieved.

IV. REFERENCES

[1] P. Mell and T. Grance,"The NIST definition of cloud computing," In Communications of the ACM, vol. 53(6), December 2010.

[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] Sivadon Chaisiri, Bu-Sung Lee, Dusit Niyato, "Optimization of resource provisioning cost in cloud computing" IEEE transactions on services Computing,Vol.5, No.2, April- June 2012.

[4] Jianhua Gu, Jinhua Hu, Tianhai Zhao, Guofei Sun, A New Resource Scheduling Strategy Based on Genetic Algorithm in Cloud Computing Environment," Journal of Computers, vol. 7, January 2012.

[5] Jiandun Li, Junjie Peng, Zhou Lei, Wu Zhang, "An Energy-efficient Scheduling Approach Based on Private Clouds", Journal of lnformation & Computational Science, April 2011.

[6] Shailesh S. Deore, Ashok Narayan Patil, "Energy-Efficient Scheduling Scheme for Virtual Machines in Cloud Computing" International Journal of Computer Applications, Volume 56– No.10, October 2012.

[7] Jiandun Li, Junjie Peng, Zhou Lei, Wu Zhang, "A Scheduling Algorithm for Private Clouds", Journal of Convergence Information Technology, Volume6, Number 7, July 2011.

[8] Kyong Hoon Kim, Anton Beloglazov, Rajkumar Buyya, "Power-aware Provisioning of Cloud Resources for Real-time Services", ACM Publications, December 2009.

[9] Ruay-Shiung Chang, Jih-Sheng Chang, Po-Sheng Lin, "An ant algorithm for balanced job scheduling in grids", Future Generation Computer Systems, June 2008.

[10] Ungurean, "Job Scheduling Algorithm based on Dynamic Management of Resources Provided by Grid Computing Systems", Feb 2010.

[11] Golnar Gharoonifard, Fahime Moeindarbari, Hossein Deldari, Anahita Morvaridi, "Scheduling of scientific workflows using a chaos-genetic algorithm", Elsevier publications, 2010.

[12] loannis A. Moschakis, Helen D. Karatza, "Evaluation of gang scheduling performance and cost in a cloud computing system", Springer publications, 2010.

[13] Suraj Pandey, LinlinWu, Siddeswara Mayura Guru, Rajkumar Buyya, "A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments", IEEE International Conference on Advanced Information Networking and Applications, 2010.

[14] Wei Wang, Guosun Zeng, Daizhong Tang, Jing Yao, "Cloud-DLS: Dynamic trusted scheduling for Cloud computing", Expert Systems with Applications, 2011.

[15] Muhammad Asad Arfeen, Krzysztof Pawlikowski, Andreas Willig, "A Framework for Resource Allocation Strategies in Cloud Computing Environment", IEEE International Computer Software and Application Conference, July 2011.

[16] Shamsollah Ghanbari, Mohamed Othman, "A priority based job scheduling algorithm in cloud computing" International Conference on Advances Science and Contemporary Engineering, 2012.

[17] Gunho Lee, Byung-Gon Chun, Rangy H. Katz, "Heterogeneity-aware resource allocation and scheduling in the clouds".