A Review of Workflow Scheduling in Cloud Computing Environment

Anterpreet Kaur
Department of Information Technology,
CEC, Landran, Mohali, India
cecit.anter@gmail.com

Abstract

Over the years, distributed environments have evolved from shared community platforms to utility-based models; the latest of these being Cloud computing. This technology enables the delivery of IT resources over the Internet and follows a pay-as-you-go model where users are charged based on their usage. There are various types of Cloud providers each of which has different product offerings. They are classified into a hierarchy of as-a-service terms: Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). There are a mass of researches on the issue of scheduling in cloud computing, most of them, however, are about workflow and job scheduling. A cloud workflow system is a type of platform service which facilitates the automation of distributed applications based on the novel cloud infrastructure. Many scheduling policies have been proposed till now which aim to maximize the amount of work completed while meeting QoS constraints such as deadline and budget. However many of them are not optimal to incorporate some basic principles of Cloud Computing such as the elasticity and heterogeneity of the computing resources. Therefore our work focuses on studying various problems and issues related to workflow scheduling.

Keywords: Cloud computing, Scheduling, Workflow.

1. INTRODUCTION

With advancement in technology, processing and storage and also with the success of the Internet, computing resources have become cheaper, more powerful and more ubiquitously available than ever before, this technological trend has given birth to the realization of a new computing model called cloud computing, in which resources (e.g., CPU and storage) are provided as general utilities that can be leased and released by users through the Internet in an on-demand fashion. In a cloud computing environment, the traditional role of service provider is divided into two: the infrastructure providers who manage cloud platforms and lease resources according to a usage-based pricing model, and service providers, who rent resources from one or many infrastructure providers to serve the end users [1]. Cloud computing can be divided into three service models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) [2]. Scheduling in cloud computing is the process of mapping tasks onto resources and the systems (e.g. CPU time, bandwidth and memory) efficiently. Hence, to allocate the resources to each job effectively, scheduling plays an important role in cloud computing. On the basis of dependency of jobs, scheduling can be of two types: Independent Scheduling (Task Scheduling) and Dependent Scheduling (Workflow Scheduling). A workflow is a sequence of connected instructions. The motive of workflow scheduling is to automate the procedures especially which are involved in the process of passing the data and the files between the participants of the cloud, maintaining the constraints [3]. Workflow scheduling is NP hard problem therefore it is impossible to generate an optimal solution within polynomial time and algorithms focus on generating approximate or near-optimal solutions [4]. The multi-objective nature of the scheduling problem in Clouds makes it difficult to solve, especially in the case of complex jobs like workflows [5]. The objective of this paper is to focus on various workflow scheduling algorithms.

2. Algorithms and Metrics

Numbers of authors proposed work in the area of workflow scheduling algorithm. Table1.represents the description of work done and algorithm used considering the various parameters like time, cost, makespan, speed, scalability, throughput, resource utilization, scheduling success rate and so on.

Table1. Survey on workflow scheduling algorithms

Algorithm	Description	Parameter	Tool
An Algorithm in SwinDeW-C for Scheduling Transaction-Intensive Cost-Constrained Cloud Workflows [6]	Focuses on novel scheduling algorithm which considers the characteristics of cloud computing to accommodate transaction-intensive cost-constrained workflows by compromising execution time and cost with user input enabled on the fly.	Cost, Time	SwinDeW-C
A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows for Cloud Computing [7]	A multiple QoS constrained scheduling strategy of multi-workflows (MQMW) is introduced to schedule multiple workflows which are started at any time and the QoS requirements are taken into account. They have considered four factors which affect the total makespan and cost of workflow greatly.	Scheduling success rate, Cost, Time, Makespan	CloudSim
Ant Colony Optimization [8]	ACO finds the schedule that meets all user imposed QoS constraints. It calculates the pheromone values based on heuristics and experiments are done on ten workflow applications.	Deadline, Cost, Budget, Makespan and Reliability	Grid Environment
A Particle Swarm Optimization-Based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments [9]	A particle swarm optimization (PSO) based heuristic to schedule applications to cloud resources that takes into account both computation cost and data transmission cost. It is used for workflow application by varying its computation and communication costs.	Resource utilization, Time	Amazon EC2
A Revised Discrete Particle Swarm Optimization for Cloud Workflow Scheduling[10]	A Revised Discrete Particle Swarm Optimization (RDPSO) was proposed to schedule applications among cloud services that takes both data transmission cost and computation cost into account. Experiment was conducted with a set of workflow applications by varying their data communication costs and computation costs according to a cloud price model. Results show that the proposed RDPSO algorithm can achieve much more cost savings and better performance on make span and cost optimization.	Cost, Makespan	Amazon EC2
Ant Colony Optimization Based Service Flow Scheduling with Various QoS Requirements in Cloud Computing[11]	An Ant Colony Optimization (ACO) algorithm was used to optimize the service flow scheduling with various Quality of service (QoS) requirement. Service Level Agreement (SLA) monitoring module was also introduce in order to monitor the running state of cloud services	Reliability, Response time, Cost, Security	Java Environment
SHEFT workflow scheduling algorithm [12]	A SHEFT workflow scheduling algorithm to schedule a	Execution time, Scalability	CloudSim
Auto Scaling [13]	In this various computing element was virtual machines of various sizes and cost, jobs are specified as workflows.	Deadline, Cost	Amazon EC2
Deadline constrained workflow scheduling in software as a service Cloud [14]	A new QoS-based workflow scheduling algorithm based on a novel concept called Partial Critical Paths (PCP), was proposed which tries to minimize the cost of workflow	Cost, Deadline	Java environment

	execution while meeting a user-defined deadline. This algorithm recursively schedules the partial critical paths ending at previously scheduled tasks.		
Workflow Scheduling for SaaS / PaaS Cloud Providers [15]	An integer linear program (ILP) formulation for the problem of scheduling SaaS customer's workflows into multiple IaaS providers where SLA exists at two levels.	Time, Cost, Makespan	Java environment
A Compromised-Time-Cost Scheduling Algorithm in SwinDeW-C for Instance-Intensive Cost Constrained Workflows on a Cloud Computing Platform[16]	In this compromised-time-cost scheduling algorithm was used which considers the characteristics of cloud computing to accommodate instance-intensive cost-constrained workflows by compromising execution time and cost with user input enabled on the fly. .	Cost, Time	SwinDeW-C
A Priority Constrained Scheduling Strategy[17]	Priority Impact Scheduling Algorithm was proposed that considered fairness and consumer's priority and task weight are taken into account.	Success rate	Cloud Sim
Deadline and Budget Distribution based Cost-Time Optimization Workflow Scheduling Algorithm for Cloud [18]	Deadline and Budget distribution-based Cost-Time Optimization (DBD-CTO) workflow scheduling algorithm was considered that minimizes execution cost while meeting timeframe for delivering results and analyze the behavior of the algorithm while considering the two constraints: deadline and budget.	Cost, Time	Java environment
A Multi-Objective Approach in Heterogeneous Environment [19]	The scheduling algorithm is implemented in ASKALON environment for Grid and Cloud Computing. It outperform when compared with bi-criteria heuristic and bi-criteria genetic algorithm.	Makespan, Economic Cost, Energy Consumption, Reliability	ASKALON environment
A market-oriented Hierarchical scheduling strategy in cloud workflow systems [20]	A market-oriented hierarchical scheduling strategy was proposed in which service-level scheduling and a task-level scheduling was considered. The service-level scheduling deals with the Task-to-Service assignment and the task-level scheduling deals with the optimization of the Task-to- VM assignment in local cloud data centers.	Makespan, Cost, CPU time	SwinDeW-C
A truthful dynamic workflow scheduling mechanism[21]	Introduced a pricing model and a truthful mechanism for scheduling single tasks considering two objectives: monetary cost and completion time. Theoretically analyze the truthfulness and the efficiency of the mechanism and significant impact of the selfish behavior of the cloud providers on the efficiency of the whole system.	Makespan, Cost	GridSim
Deadline based Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds [4]	An algorithm based on the meta-heuristic optimization technique, Particle Swarm Optimization (PSO), which aims to minimize the overall workflow execution cost while meeting deadline constraints on an IaaS cloud enviroment	Cost, Deadline	Cloudsim

3. CONCLUSION

In this paper, we have surveyed the various existing workflow scheduling algorithms in cloud computing and tabulated their various parameters along with tools and algorithm used. Through extensive literature survey it has been found that there are number of algorithm for scheduling, and these algorithms some how differ in scheduling factor and parameter. It has also been analyzed that workflow scheduling is NP hard problem therefore it is impossible to generate an optimal solution within polynomial time and algorithms focus on generating approximate or near-optimal solutions. Lots of research work has been done on workflow scheduling but still there are areas which require further attention e.g. elasticity and heterogeneous nature of cloud is not considered to meet the QoS requirement.

4. REFERENCES

[1] Zhang, Qi, Lu Cheng, and R. Boutaba "Cloud computing: state-of-the-art and research challenges" Journal of internet services and applications, vol. 1, pp. 7-18, 2010.

[2] Dillon, Tharam, Chen Wu, and C. Elizabeth "Cloud computing: issues and challenges." in 24th IEEE International Conference on Advanced Information Networking and Applications (AINA), pp. 27-33, 2010.

[3] Zhu, Linan, Qingshui Li, and Lingna He "Study on cloud computing resource scheduling strategy based on the ant colony optimization algorithm." in International Journal of Computer Science Issues, vol.9, no. 5, 2012.

[4] Rodriguez Sossa, M., and Rajkumar Buyya "Deadline based Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds." in IEEE Transactions on Cloud Computing, vol. 2, no. 2, pp.222-235, 2014.

[5] A.KBardsiri and S.M Hashemi "A Review of Workflow Scheduling in Cloud Computing Environment" in International Journal of Computer Science and Management Research, vol. 1, no.3, pp. 348-351,2012

[6] Y. Yang, K. Liu, J. Chen, X. Liu, D. Yuan and H. Jin, "An Algorithm in SwinDeW-C for Scheduling Transaction-Intensive Cost-Constrained Cloud Workflows" in Proceeding of 4th IEEE International Conference on e- Science, pp. 374-375, 2008.

[7] M. Xu, L. Cui, H. Wang, Y. Bi, "A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows for Cloud Computing" in IEEE International Symposium on Parallel and Distributed Processing, pp. 629-634, 2009.

[8] Wei Neng Chena and Jun Zhang, "An Ant Colony Optimization Approach to a Grid Workflow Scheduling Problem with Various QoS Requirements" in IEEE Transactions on System, Man and Cybernetics, Applications and Reviews, , Vol. 39, no.1, pp. 29-43,2009.

[9] S. Pandey, L. Wu, S.M. Guru, R Buyya, "A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments" in 24th IEEE International Conference on Advanced Information Networking and Applications, pp. 400-407, 2010.

[10] Z. Wu, Z. Ni, L. Gu, "A Revised Discrete Particle Swarm Optimization for Cloud Workflow Scheduling" Computational Intelligence and Security (CIS), pp. 184-188, 2010.

[11] H. liu, D. Xu, H. Miao, "Ant Colony Optimization Based Service flow Scheduling with Various QoS Requirements in Cloud Computing", in First ACIS International Symposium on Software and Network Engineering, pp. 53-58, 2011.

[12] C. Lin, S. Lu, "Scheduling Scientific Workflows Elastically for Cloud Computing", in IEEE 4th International Conference on Cloud Computing, pp. 246-247, 2011.

[13] Ming.M and Marty Humphrey. "Auto-scaling to minimize cost and meet application deadlines in cloud workflows" in Proceedings of International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 49. ACM, 2011.

[14] S. Abrishami and M. Naghibzadeh, "Deadline-constrained workflow scheduling in software as a service Cloud" in Scientia Iranica, Transactions D: Computer Science & Engineering and Electrical Engineering, pp. 680-689, 2012.

[15] T.A.L. Genez, L.F. Bittencourt, E.R.M. Madeira, "Workflow Scheduling for SaaS / PaaS Cloud Providers Considering Two SLA Levels" in IEEE Network Operations and Management Symposium (NOMS): Mini- Conference, pp. 906-912, 2012.

[16] K. Liu, Y. Yang, J. Chen, X. Liu, D. Yuan and H. Jin, "Compromised-Time-Cost Scheduling Algorithm in SwinDeW-C for Instance-Intensive Cost-Constrained Workflows on Cloud Computing Platform" in International Journal of High Performance Computing Applications, Vol. 24, No. 4, pp. 445-456, 2010.

[17] Wu, Hu, Zhuo Tang, and Renfa Li "A priority constrained scheduling strategy of multiple workflows for cloud computing" in 14th International Conference on Advanced Communication Technology (ICACT), pp. 1086-1089, 2012.

[18] A. Verma, S. Kaushal, "Deadline and Budget Distribution based Cost- Time Optimization Workflow Scheduling Algorithm for Cloud" in International Journal of Computer Applications (IJCA), 2012.

[19] H.M Fard, R. Prodan, J.J.D Barrionuevo, and T. Fahringer "A multi-objective approach for workflow scheduling in heterogeneous environments" in Proceedings of 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012). IEEE Computer Society, 2012.

[20] Z. Wu, X. Liu, Z. Ni, D. Yuan, Y. Yang, "A Market-Oriented Hierarchical Scheduling Strategy in Cloud Workflow Systems" in Journal Supercomputing, 2013.

[21] H.M Fard, R. Prodan, and T. Fahringer. "A truthful dynamic workflow scheduling mechanism for commercial multicloud environments" in IEEE Transactions of Parallel and Distributed Systems, vol. 24, no. 6 pp. 1203-1212, 2013.