IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-6, June 2019 E-ISSN: 2347-2693

Survey on Scheduling Algorithms for Multiple Workflows in Cloud Computing Environment

A. Aggarwal[1*], P. Dimri[2], A. Agarwal[3]

[1]Dept. of Virtualization, School of Computer Science, UPES, Dehradun, India
[2] Dept. of Computer Science and Engineering, GBPEC, Pauri, India
[3] Dept. of Virtualization, School of Computer Science, UPES, Dehradun, India

Corresponding Author: ambika2488@gmail.com, Tel.: +91-9456505869

DOI: https://doi.org/10.26438/ijcse/v7i6.565570 | Available online at: www.ijcseonline.org

Accepted: 11/Jun/2019, Published: 30/Jun/2019

Abstract**—** Cloud computing has been the buzzword of the ICT industry lately and has been widely accepted because of its pay-as-you-go pricing model and ease of use. Hence the bigger scientific applications which are generally represented as workflow or DAG are also moving on the cloud. Scheduling workflows on the cloud was becoming a problem as the existing traditional workflow scheduling algorithms (for Grid) were not perfectly suitable for cloud environment because of its dynamic nature. This paper tries to explore the existing algorithms for scheduling multiple workflows in cloud computing environment and presents a comparative analysis in tabular form of some existing algorithms along with their parameters, methods and tools used.

Keywords**—** Scheduling, Cloud Computing, Multiple Workflow

I. Introduction

In recent times, cloud computing has gained tremendous popularity in the ICT industry and has reached a maturity level that is capable of providing promising platforms and computing resources for developing and deploying large-scale applications. Cloud provides an on-demand service model which consists of a large shared pool of computational resources like storage, servers, network bandwidth etc. which can be easily made available to the user and they pay a minimal amount depending on the their usage[1].

Workflow is basically a set of dependent tasks which is typically represented using a DAG (Directed Acyclic Graph). Workflow is a commonly used model for describing scientific applications which usually comprises of a number of tasks and also the control or data dependency among those tasks. The problem of scheduling workflows is considered as NP-complete problem which is concerned with finding appropriate solution for assigning resources (from a pool of resources) to various tasks in an efficient manner.

A lot of algorithms exist for scheduling workflows in traditional grid computing environment. Two of the best known list-based heuristics are “Heterogeneous Earliest Finish Time (HEFT)” and “Critical Path On a Processor (CPOP)” which focus on the problem of effective performance in workflow scheduling [2]. LOSS and GAIN [3] are the two best known budget-constraint algorithms. They begin with taking existing schedules as the input and then keep reassigning every task to some other processor until it finds a match in the cost or the cost exceeds the budget. Some other examples of budget-constraint workflow scheduling algorithms are “Budget-constraint Heterogeneous Earliest Finish Time (BHEFT)” [4], which is an extension of HEFT, and “Heterogeneous Budget Constraint Scheduling (HBCS)” [5] algorithm.

Also, there exist some algorithms which try to optimize more than one objective simultaneously. One such algorithm is POSH [6] which combines two parameters together; makespan and cost. Some other algorithms like NSPSO [7] and ε-Fuzzy PSO [8] have used “particle swarm optimization (PSO)” algorithm to produce Pareto Optimal trade-offs between cost and makespan.

But now the computing environment is moving towards cloud with unanimous consensus and thus we need to schedule workflows on the cloud. While scheduling workflows on cloud, we might face some challenges because a) clouds have a complex pricing scheme and b) the resource pool in cloud is much larger [9]. Hence, the existing workflow scheduling algorithms may not be suitable for replicating in a cloud environment.

© 2019, IJCSE All Rights Reserved

The remainder of the paper is organized as follows; section I contains the introduction to workflows and the traditional algorithms used in grid computing for scheduling workflows. Section II explains the concept of workflow scheduling in cloud. Section III talks about multiple workflow scheduling in cloud. Section IV focuses on the related work and some existing algorithms are discussed and compared. Lastly, section V concludes the paper and future scope is discussed.

II. WORKFLOW SCHEDULING IN CLOUD

Workflow is generally represented with the help of Directed Acyclic Graph (DAG) G={V,E}, where V represents the set of vertices such that each vertex $v_i \in V$, and E represents the set of edges such that each edge $e_{i,j} \in E$. Every edge $e_{i,j} \in E$ shows the data dependency between two vertices v_i and v_j which means that data produced by v_i is consumed by v_j.
A sample workflow diagram is shown below:-

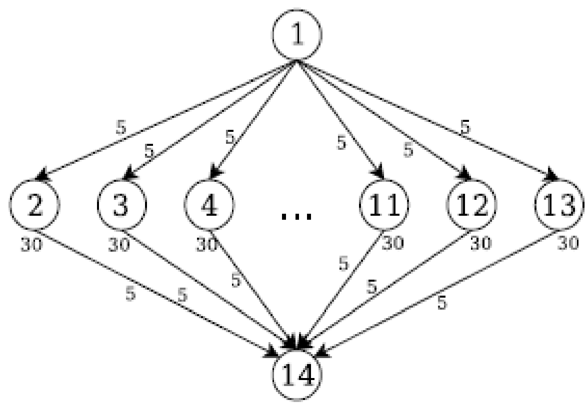

Figure 1. A sample workflow (Directed Acyclic Graph) [10]

Various workflow scheduling algorithms for cloud computing environment have been developed till date. Some focus on optimizing a single parameter while others are multi-objective scheduling algorithms. Different methods, different parameters and different tools have been adopted while developing these algorithms. All the proposed workflow scheduling algorithms for cloud environment are inspired by the existing scheduling algorithms for grid.

"Evolutionary Multi-Objective Optimization (EMO)" [9] based workflow scheduling algorithm, "Game-Multi-Objective (GMO)" [15] algorithm for scheduling workflows is based on Game Theory approach, another algorithm which is based on deadline for resource provisioning and scheduling [16] utilizes the meta-heuristic "particle swarm optimization (PSO)" technique, Workflow-tailored BOSS [17] multi-objective algorithm, MCPCPP [18] algorithm for Big-data applications, are some examples of the existing workflow scheduling algorithms in cloud environment.

III. MULTIPLE WORKFLOW SCHEDULING IN CLOUD

In cloud environment, where we have a large pool of resources, we can schedule multiple workflows in order to achieve greater efficiency and improved resource utilization. Workflow scheduling algorithms can be broadly categorised as follows: static and dynamic scheduling algorithms. The static scheduling algorithms are mostly concerned with scheduling a single workflow whereas the dynamic or the online scheduling approach can be utilized for scheduling multiple workflows [11]. It is mainly because the static approach is based on the assumption that the overall structure of the workflow or the DAG is known beforehand [12]. It also assumes that the various constraints are also known such that the approximate time of execution of all the tasks can be easily calculated. Thus, in static scheduling algorithms, all the required resources are allocated before beginning the execution of the task and no changes can be made once the execution has started. On the other hand, the dynamic or the online scheduling algorithms can allocate the resources dynamically ie during execution as it is aware of the dynamic nature of resource availability in a cloud environment and takes it into consideration. One way to look at the multiple workflow scheduling problem might be to merge all the workflows to form a single larger workflow. Some studies have been conducted in the recent past to focus on the above mentioned problem of scheduling multiple workflows in cloud computing environment. We are going to discuss these works in the further sections.

IV. RELATED WORK

A. *"CCRH"*

In this paper, Y. Wang et.al [12] have proposed a competitive and dynamic algorithm for scheduling multiple workflows (multiple DAGs) which takes into consideration the link communication processor. The communication competition model is used in the algorithm for describing the mode of communication among processors. It makes use of dynamic hierarchical approach for defining the scheduling processor unit and then calculates the unfair degree of each workflow. The proposed algorithm is more effective when there is large difference weights in multiple workflows. The simulations were compared with BMCT and HEFT for four different aspects such as: fairness in multiple workflow scheduling, time taken for scheduling, utilization of CPU and the execution time of tasks. The results show improved fairness in multiple workflow scheduling, reduced average scheduling time of multiple workflows and improved robustness of the algorithm.

B. *"OMPHC-PCPR and OPHC-TR"*

In this paper, S. Sharif et.al [11] have proposed two algorithms namely "Online Multiterminal Cut for Privacy in Hybrid Clouds using PCP Ranking (OMPHC-PCPR)" and

© 2019, IJCSE All Rights Reserved

"Online scheduling for Privacy in Hybrid Clouds using Task Ranking (OPHC-TR)". Both these algorithms have been developed for scheduling medical data (patient's) workflows. As per the study, these algorithms are capable of scheduling multiple workflows under two constraints, deadline and privacy. Both these algorithms have used different methods for ranking tasks in a workflow and they also take into consideration the dynamic aspects of hybrid cloud computing environment. The results also show that merging multiple workflows into a single workflow reduces the cost by 12%.

i. **"OMPHC-PCPR":** The input for the algorithm is a single DAG in which multiple workflows are merged.

This algorithm is invoked at the beginning of each schedule and then it calls the Monitoring procedure, which adds the finished tasks to the finished job list and removes them from instance list.

Another algorithm is used for finding the schedule path which can finish all the tasks before their latest finish time.

ii. **"OPHC-TR":** This algorithm also works similar to the previous algorithm in ranking the tasks except that it schedules the tasks individually unlike OMPHC-PCPR, which schedules the entire path.

C. *"EMuWS"*

In this paper, T. Thanavanich et.al [13] have proposed an algorithm "Energy-aware Multiple Workflows Scheduling (EMuWS)" for scheduling multiple workflows on a large-scale distributed system. The algorithm proposes to achieve scheduling quality and also focuses on reducing energy consumption while maintaining the completion time. The results shown in the paper are obtained by comparing EMuWS with HEFT and EHEFT while keeping HEFT as the base algorithm.

The authors have also proposed an algorithm IIP to identify the set of inefficient processors so that they can be shut down and their tasks can be rescheduled to other efficient processors.

D. *"Time dependence based Scheduling Strategy for Multiple Workflows"*

In this paper, S. Liu et.al [14] have proposed a time dependence based scheduling algorithm for multiple workflows which combines the advantages of three previously proposed static scheduling algorithms for multiple workflows with soft deadline. These algorithms are, Multiple "Deadline-constrained Workflows Task-based (MDW-T)", "Multiple Deadline-constrained Workflows Workflow-based (MDW-W)", and "Multiple Deadline-constrained Workflows Cluster-based (MDW-C)".

Scheduling Algorithm	**Year**	**Scheduling Approach**	**Scheduling Parameters**	**Scheduling Methods**	**Observations**	**Tool**
CCRH	2014	Static + dynamic hierarchical approach	Fairness, makespan, resource utilization, running time	Primary Backup Scheduling Algorithm	Improved fairness, reduced makespan, improved resource utilization and system uptime	Xen
OMPHC-PCPR and OPHC-TR	2014	Dynamic	Privacy, Budget and Deadline	Ranking Techniques	a) OMPHC-PCPR outperforms OPHC-TR by reducing the cost upto 50%. b) Both the algorithms take into account privacy and deadline.	Simulator
EMuWS	2016	Static	Energy consumption, makespan and quality	Traditional list scheduling with ratio of effectiveness applied on CyberShake, Montage, LIGO and Epigenomics	Energy consumption ratio is reduced upto 66.78%	Simulator
Time dependence based Scheduling Strategy for multiple Workflows	2016	Static	Time dependence, deadline and budget	Proposed strategy validated with real scientific workflows Montage, LIGO	a) It gives better completion rate when compared with HEFT and MDW-W b) Better performance when time interval of VM is smaller. c) With increase in the budget, the completion rate will also improve.	CloudSim

The proposed time dependence based scheduling strategy calculates priority with time dependence among workflows and uses budget and deadline constraints to remove those parts of workflows, which could not be completed. Results show that the completion rate of the proposed algorithm increased by 14-26% as compared to HEFT and it showed an increment of 8-25% as compared to MDW-W.

V. SOME OTHER WORKFLOW SCHEDULING ALGORITHMS

A. *"Cloud Workflow Scheduling Algorithm (CWSA)"*

According to B.P. Rimal et al. [19]This algorithm was proposed for multi-tenant cloud computing environment where the applications submitted are highly compute-intensive and require a lot of computing resources. The main objectives of the said algorithm was to minimize makespan,

© 2019, IJCSE All Rights Reserved

minimizing workflow execution cost, minimizing tardiness and efficient utilization of idle cloud resources. The said algorithm was compared with several standard algorithms for performance evaluation such as "First Come First Served (FCFS)", "Easy Backfilling", and "Minimum Completion Time (MCT)". The efficiency and effectiveness of the proposed algorithm was verified by conducting some experiments on CloudSim using some real-world scientific workflow applications which also demonstrated the scalability of thes algorithm.

B. *"Cloud Workflow Schedling with Deadlines and Time slot availability (WSDT)"*

In this paper[20], the authors have taken into account deadline, and time-slot availability for scheduling workflows in cloud. The available time slots are provided by the cloud vendors based on the availability of the capacities. The authors have presented an iterative heuristic model as a solution for their problem which consists of three major aspects: the initial solution, the improvement phase, and the perturbation phase. The proposed strategies have been verified by conducting some experimental analysis, and the results show that different strategies have different impact on the quality of solution.

C. *"Multi-level K-cut graph partitioning algorithm (MKCut)"*

The authors in this paper [21] have proposed an algorithm which will try to reduce the amount of data transfer from one datacentre to another, keeping in mind the load balancing issues and some constant data constraints. The fixed sized input datasets along with all the tasks are first contracted by the algorithm at the initial datacentre, and the resultant contracted graph is then scaled gradually following a step by step pattern. Later on the resultant graph is cut into K partitions using the K-cut algorithm in order to minimize the cut size. In order to maintain the load balancing constraint, the partitioned graph is eventually mapped to the initial workflow graph. Three real-world workflow applications have been used to evaluate the performance of this algorithm. The experimental results show that the proposed algorithm performs better than some of the existing algorithms.

D. *"Evolutionary Multi-Objective Optimization (EMO) based Workflow Scheduling Algorithm"*

In this paper, Z. Zhu et. al.[9] have focussed on the difficulties involved in directly applying the traditional workflow scheduling algorithms used in Grid computing to the Cloud Computing Environment. The proposed algorithm tries to optimize multiple objectives such as cost and makespan. The proposed algorithm is meant to work on infrastructure as a service platform, and it proposes several schemes for fitness evaluation, solution encoding, population initialization and genetic operators. Several experiments have been conducted on the Amazon EC2 instance types considering many real world as well as randomly generated workflows and the results show that the proposed algorithm performs better than the existing algorithms in most of the cases.

E. *"Cost and Energy Aware Scheduling (CEAS) Algorithm"*

Z. Li et. al [22] have proposed an algorithm for scientific workflows which aims at minimizing the energy consumption and reducing the cost of execution for cloud workflow schedulers. The said algorithm comprises of further five sub-algorithms. The first algorithm aims at minimizing the execution cost and is thus applied for VM selection. Next, in order to reduce the energy consumption and cost of execution of workflows, two algorithms have been utilized to merge tasks. Then there is another algorithm for utilizing the idle VMs by defining a reuse policy for already leased VMs. Finally, an algorithm is applied for minimizing the energy consumption of leased out VMs. The experimental results have been carried out using CloudSim to verify the performance of the proposed algorithms as compared to the existing algorithms.

F. *"Deadline based Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds"*

In this paper[23], the authors have proposed an algorithm for scheduling workflows in Infrastructure as a Service cloud architecture. The proposed algorithm is based on the traditional "Particle Swarm Optimization (PSO)" algorithm which adopts the meta-heuristic technique for optimization. The main objective of the proposed algorithm is to minimize the overall cost of execution of workflows without violating any deadline constraints. The verification experiments were carried out on CloudSim on some scientific workflows, and the results proved that the proposed algorithm out performed some of the existing scheduling algorithms.

VI. CONCLUSION AND FUTURE SCOPE

Efficient scheduling of workflows in a cloud environment is the need of the hour and scheduling multiple workflows simultaneously and efficiently is of greater importance in order to improve resource utilization. Workflow is a set of dependent tasks and while scheduling a single workflow on cloud we might have some resources which remain unutilized for significant amount of time. In order to utilize these idle resources, we can simultaneously schedule another workflow interleaving the gaps left in the scheduler. We can identify some combinations of workflows which can be scheduled together on the existing resources to achieve better resource utilization, reduced finish time, and reduced power consumption.

The studies show that a lot of work has been done and is still being done in the field of single workflow scheduling in cloud environment whether it be optimizing a single objective or optimizing multiple objectives. But the field of multiple workflow scheduling in cloud environment is not much explored and a lot can be done this area. If all the workflows are arriving at the same time, then they can be merged into a single bigger workflow. If the arrival time of individual workflows is different, then they can be scheduled by clustering them depending on their time of arrival. Most of the existing algorithms focus on the static scheduling approach while a cloud environment is dynamic in nature. Thus, further research can be done while keeping the dynamic nature of cloud in mind.

REFERENCES

[1] P. Mell and T. Grance, *"The NIST definition of Cloud Computing"*, NIST, Gaithersburg, MD, USA, Tech. Rep. 6,2009.

[2] H. Topcuoglu, S. Hariri, and M.Y. Wu, *"Performance-effective and low-complexity task scheduling for heterogeneous computing"*, IEEE Transaction on Parallel and Distributed Systems, vol. 13, no. 3, pp. 260-274, March 2002.

[3] R. Sakellariou, H. Zhao, E. Tsiakkouri, and M.D. Dikaiakos, *"Scheduling workflows with budget constraints*", in Proc. Integr. Res. Grid Computing, 2007, pp.. 189-202

[4] W. Zheng and R. Sakellariou, *"Budget-deadline constrained workflow planning for admission control",* Journal of Grid Computing, vol. 11, no. 4, pp. 633-651, 2013.

[5] H. Arabnejad and J. G. Barbosa, *"A budget constrained scheduling algorithm for workflow applications",* Journal of Grid Computing, vol. 12, pp. 665-679, 2014.

[6] S. Su, J. Li, Q. Huang, X. Huang, K. Shuang, and J. Wang*, "Cost-efficient task scheduling for executing large problems in the cloud",* Parallel Computing, vol. 39, no. 4, pp. 177-188, 2013.

[7] R. Garg, and A.K. Singh, *"Multi-objective workflow grid scheduling based on discrete particle swarm optimization",* in Proc. Swarm, Evol., Memetic computing, 2011, pp. 183-190.

[8] R. Garg, and A.K. Singh, *"Multi-objective workflow grid scheduling using ε-Fuzzy dominance sort based discrete particle swarm optimization",* Journal of Supercomputing, vol. 68, no. 2, pp. 709-732, 2014.

[9] Z. Zhu, G. Zhang, M. Li, and X. Liu, *"Evolutionary Multi-objective Workflow Scheduling in Cloud",* IEEE Transactions on Parallal and Distributed Systems, vol. 27, no. 5, pp. 1344-1356, 2016.

[10] T.A. L. Genez, L.F.Bittencourt, E. R. M. Madeira, *"Workflow scheduling for SaaS/PaaS cloud providers considering two SLA levels"*, IEEE Network Operations and Management Symposium, 2012.

[11] S. Sharif, J. Taheri, A. Y. Zomaya, *"Online multiple workflow scheduling under Privacy and Deadline in Hybrid Cloud Environment",* 6th International Conference on Cloud Computing Technology and Science, IEEE Computer Society, 2014

[12] Y. Wang, C. Jia, Y. Xu, *"Multiple DAGs Dynamic Workflow Schedduling based on the Primary Backup Algorithm in Cloud Computing System",* 9th International Conference on Broadband and Wireless Computing, Communication and Applications, IEEE, 2014

[13] T. Thanavanich, A. Siri, K. Boonlom, and A. Chaikaew, *"Energy-aware Scheduling of Multiple Workflows Application on Distributed Systems",* 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), IEEE, 2016

[14] S. Liu, K.Ren, K. Deng, and J. Song, *"Time dependence based scheduling strategy for multiple workflows on IaaS cloud platform",* International Symposium on Computer, Consumer and Control, IEEE Computer Society, 2016.

[15] R. Duan, R, Prodan, and X. Li, *"Multi-Objective Game Theoretic Scheduling of Bag-of-Tasks Workflows on Hybrid Clouds"*, IEEE Transactions on Cloud Computing, vol. 2, no. 1, 2014.

[16] M. A. Rodriguez, and R. Buyya, *"Deadline based resource provisioning and scheduling algorithm for scientific workflows on clouds",* IEEE Transactions on Cloud Computing, 2014

[17] H. M. Fard, R. Prodan, and T.Fahringer, *"A Truthful Dynamic Workflow Scheduling Mechanism for Commercial Multicloud Environments",* IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 6, pp. 1203-1212, 2013.

[18] B. Lin, W. Guo, N. Xiong, G, Chen, A. V. Vasilakos, and H.Zhang, *"A Pretreatment Workflow Scheduling Approach for Big Data Applications in Multicloud Environments"*, IEEE Transactions on Network and Service Management, vol. 13, no. 3, pp. 581-594, 2016.

[19] B.P. Rimal , and M. Maier, *"Workflow Scheduling in Multi-Tenant Cloud Computing Environments"*, IEEE Transactions on Parallel and Distributed Systems, 2015.

[20] X. Li, L. Quian, and R. Ruiz, *"Cloud Workflow Scheduling with Deadlines and Time Slot Availability"*, IEEE Transactions on Parallel and Distributed Systems, 2015.

[21] K. Deng, K. Ren, M. Zhu, and J. Song, *"A Data and Task Co-scheduling Algorithm for Scientific Cloud Workflows"*, IEEE transactions on Cloud Computing, Dec. 2015.

[22] Z. Li, J. Ge, H. Hu, W. Song, and B. Luo, *"Cost and Energy Aware Scheduling Algorithm for Scientific Workflows with Deadline Constraint in clouds"*, IEEE Transactions on Services Computing, 2015.

[23] M. A. Rodriguez, and R. Buyya, "Deadline Based Resource Provisioning and Scheduling Algorithms for Scientific Workflows on Clouds",, IEEE transactions on Cloud Computing, Vol. 2. No. 2, April-June 2014.

Authors Profile

Ms. Ambika Aggarwal pursued Bachelor of Technology from Uttarakhand Technical University, Dehradun in 2010 and Master of Technology from Uttarakhand Technical University, Dehradun in the year 2012. She is currently pursuing Ph.D. and currently working as Assistant Professor (Senior Scale) in the Department of Virtualization, School of Computer Science, UPES, Dehradun since 2016. She has published various research papers/book chapters in reputed international journals. Her main research work focuses on Scheduling Algorithms, Cloud Computing, and Big Data Analytics. She has more than 7 years of teaching experience.

Dr. Priti Dimri pursued MCA from Gurukul Kangri University, in 1995 and MBA (Finance) from Sikkim Manipal University in 2010. She then pursued Ph.D. (Computer Graphics) from Uttarakhand Technical University, Dehradun in 2013 and iscurrently working as Associate Professor and Head, Department of Computer Science and Applications, GBPEC, Pauri since 2010. She is a member of Indian Society for Technical Education. She has published various research papers/book chapters in reputed international journals. Her research area includes Green Computing, Fractals, ITSM, Software Engineering. She has more than 10 years of teaching experience and approx. 10 years of Industry Experience.

Dr. Amit Agarwal pursued B.tech from GBPEC, Pauri and M.Tech from Punjabi University, Patiala. He then pursued Ph.D. from IIT, Roorkee and is currently working as Professor and Head, Department of Virtualization, School of Computer Science, UPES, Dehradun. He is a member of various Technical bodies and has published various research papers/books/chapters in reputed international journals. His research area includes Grid Computing, Cloud Computing, Scheduling Algorithms and IoT etc. He has approximately 15 years of teaching experience .

© 2019, IJCSE All Rights Reserved