A Review on Various Workflow Scheduling Algorithms in Cloud Computing

S Sudheer Mangalampalli[1],
Department of CSE,
Shri Vishnu Engineering College For Women,
Bhimavaram, AP, India.

Dr. Sangram Keshari Swain[2],
Associate Professor,
Department of CSE,
Centurion University of Technology and Management,
Odisha, India.

Dr. P Kiransree[3],
Department of CSE,
Shri Vishnu Engineering College For Women,
Bhimavaram, AP, India.

K V Narayana Rao,
Department of CSE,
Shri Vishnu Engineering College For Women,
Bhimavaram, AP, India.

Abstract: **Cloud computing technology is one of the fastest growing technologies in IT industry and which is used in different fields for different purposes such as storage, compute and network virtually which are given to the users on demand in the form of services. Provisioning and deprovisioning of these virtual resources based on Service Level Agreement made between Cloud users and provider. Mapping of virtual resources to workflow in cloud computing is a tedious task as incoming workflow varies with respect to time. In Cloud computing, mapping of these workflows to virtual resources which were resided in physical hosts is to be fulfilled by a scheduler which assigns workflows automatically to virtual resources without any human intervention. In this paper, we have reviewed various workflow scheduling algorithms and studied different parameters which effects scheduling algorithms in cloud computing.**

Keywords: Cloud Computing, Workflow Scheduling, SLA

INTRODUCTION

According to NIST [1], "Cloud Computing can be defined as on demand network access to a shared pool of configurable, computational resources." It is a fast growing technology which is to be used in almost all applications i.e. in enterprise and academia. In the current era, due to this pandemic every organization is running online and if the organizations need to operate it from online they need more computing, storage and network. For compute, storage and network capabilities most of the organizations going on to cloud which reduces on premises infrastructure. An advantage to migrate on to cloud and manage the infrastructure on cloud virtually gives the flexibility for the organizations by scaling up and down their servers automatically on demand. Another advantage for the organizations and users of the cloud is management of the hardware and the cost optimization for the underlying hardware over the virtual infrastructure can be minimized. Generally Cloud computing can be depicted in two types of [2] architectures.1) Generic architecture which consists of three tiers i.e. front end, back end and a middle tier which is of network which connects both front and backends. Front end is an application which should run from a browser which is to be on any kind of device. Back end is a virtual infrastructure i.e. storage capacity runs on the physical hosts. These both front and back ends to be connected by a network. 2) Market-oriented architecture- typically which is to be used in real time scenarios. In this architecture, initially users submit requests on to interface of the cloud and then cloud broker which is an agent on behalf of users took the requests and gives these requests to the task manager. In this module [2], task manager receives requests from broker and then validates requests based on SLA made between each user and cloud provider. Simultaneous requests will be sent from each user to cloud console, so validating and assigning requests to corresponding virtual resources is a tedious task. Mapping of requests i.e. tasks or workflows to corresponding VMs is to be done by a scheduler which is employed after Task manager which also needs to communicate with resource manager which checks the availability of resources at the datacenters. There are two types of scheduling techniques available in cloud computing. They are 1) Task Scheduling – which considers scheduling tasks onto VMs which were to be executed independently. 2) Workflow Scheduling- which considers tasks which are dependent on one another and maps tasks to VMs resided in the physical hosts which in turn resided in datacenters. In this paper, we have studied about various workflow scheduling algorithms and their parameters which effects the scheduler and maps workflows to corresponding VMs.

The highlights of the paper are presented below.

- Studied about various workflow scheduling algorithms.
- Given few research directions about the parameters need to be addressed by using various workflow scheduling algorithms.

In the next section, we have discussed about various workflow scheduling algorithms and in the next section few research directions were given which effects various scheduling algorithms and finally we have given conclusion about analysis of workflow scheduling algorithms in cloud computing.

Literature Review

In [3], multi objective optimization scheduling is designed which aims at minimization of makespan, cost and load balance of workflows. It is modeled by hybridization of SOS and EFT algorithms. Simulation is carried out on Workflowsim and input to algorithm is given as random

workload and it surpasses existing algorithms in view of above specified parameters.

In [4], algorithm is formulated which is aimed at minimization of processing cost by using a workflow. It is modeled by using bat algorithm. It is simulated on MATLAB and input to algorithm is generated randomly. It is evaluated against PSO and CSO algorithms and it shows a great impact over existing algorithms in view of specified parameters.

In [5], multi objective algorithm is formulated which aims at makespan and load balancing. It is modeled by intelligent water drops algorithm. It is simulated on workflowsim and real time workflows were used for simulation. It is compared against existing PSO, Min-Min and Max-Min algorithms and it is outperformed over existing algorithms with above mentioned metrics.

In [6], algorithm is formulated which aims at minimization of makespan and execution cost. It is modeled by using jaya algorithm. It is simulated on Cloudsim and input to algorithm is given with real time workflows. It is evaluated against existing PSO, GA, ACO, Honey bee and CSO algorithms and it is outperformed over existing algorithms in view of specified parameters.

In [7], algorithm is formulated which focuses at overall cost of the workflow. It is modeled by using BAT algorithm. Input to algorithm is generated randomly. Simulation is carried out on Cloudsim and it is evaluated against BRS algorithm and it shows significant improvement over the existing algorithm with the specified parameter.

In [8] an algorithm is designed which aims at optimization of workflow scheduling while minimizing makespan and cost of the workflow. It is modeled by using EIWD algorithm. Real time workflows are used for simulation. Simulation is carried out on Cloudsim. It is evaluated against PSO and C-PSO algorithms and surpasses existing algorithms in view of specified parameters.

In [9], algorithm is formulated which focuses on minimization of makespan and maximization of resource utilization. It is modeled by using PSO algorithm to get optimal solution. It is implemented on a customized environment by authors. Standard real time workflows were used as an input to algorithm and are evaluated against GA algorithm and it shows significant improvement over existing algorithms in view of specified parameters.

In [10], algorithm is designed which aims at minimization of makespan. It is modeled by using co-evolutional GA. It is implemented on Cloudsim and real time workflows are given as input to algorithm and it is evaluated against HEFT algorithm and it surpasses existing algorithm in view of specified parameters.

In [11], many objective algorithm is formulated which focuses at makespan, energy consumption, cost and Reliability. It is modeled by using IMAOPSO algorithm. It is simulated on workflowsim and input to algorithm is given from realtime workflows. It is evaluated against MAOPSO and leaf algorithms and it is outperformed over existing algorithms for the above specified metrics.

In [12], algorithm is formulated which focuses at parameters processing cost and energy consumption. It is modeled by using HPSO algorithm. Realtime workflows were given as an input to algorithm. It is simulated on cloudsim and evaluated against MOPSO and FDPSO algorithms and it surpasses the existing algorithms with specified parameters.

In [13] algorithm is designed which aims at execution cost while meeting deadline. It is modeled by using ASFLA algorithm. Realtime and random generated workflows were given as input to algorithm. It is implemented on customized java simulator and compared against existing algorithms PSO and SFLA and it shows huge impact over existing algorithms in view of specified parameters.

In [14], algorithm is formulated which focuses at minimization of execution time and maximizing reliability. It is modeled by using BAT algorithm. It is implemented on workflowsim and evaluated against BREA algorithm. It surpasses existing algorithm in terms of specified parameters.

In [15], algorithm is formulated which aims at minimization of total cost, time complexity and schedule length. It is modeled by hybridization of CRO and ACO algorithms. It is implemented on Cloudsim and realtime workloads were used for simulation. It is evaluated against CRO, ACO, PSO and CEGA algorithms and it surpasses existing algorithms in view of specified parameters.

In [16], algorithm is designed which focuses on minimization of execution cost while maintaining deadlines. It is modeled by using L-ACO algorithm. It is implemented on cloudsim and realtime workflows were given as input to algorithm. It is compared against existing PSO algorithm with the specified parameters.

In [17], multi-objective algorithm is formulated which focused at cost, energy consumption while maintaining deadlines. It is modeled by using ECMSMOO algorithm. Realtime workflows were given as input to algorithm. It is evaluated against existing algorithms such as MOHEFT, CMPSO and it outperforms existing algorithms in view of specified parameters.

In [18], algorithm is designed which focuses on execution cost and resource efficiency while meeting deadlines. It is modeled by using TOPSIS algorithm. It is simulated on Workflowsim and realtime workflows were given as input. It is compared against existing algorithms CWSA, BHEFT and it shows huge impact over existing algorithms in view of specified parameters.

In [19], algorithm is designed which aims at minimization of energy consumption, execution time while maximization of throughput. It is modeled by using EATTO based BAT algorithm. It is simulated by using a customized environment in python. It is evaluated against ACO, BSH and RN algorithms and it surpasses existing algorithms in view of specified parameters.

In [20], an algorithm is formulated which aims at minimization of makespan and energy consumption. It is modeled by using HCRO algorithm. It is implemented on customized java simulator and real time workloads are given as input to algorithm. It is compared against HCRO and MPQGA algorithms and is outperformed over existing algorithms in view of specified parameters.

In [21], algorithm is designed which focuses on makespan, resource utilization and computational cost. It is modeled by hybridization of SFLA-RSO algorithm. It is implemented on MATLAB and evaluated against SFLA and RSO algorithms

and it is outperformed over existing algorithms in view of specified parameters.

In [22], algorithm is designed which aims at failure probability, reliability and execution time. It is modeled by usig Tmin-min and Tmax-min algorithms. It is implemented on cloudsim and evaluated against max-min and min-min algorithms and shows huge impact over existing algorithms in the view of specified parameters.

In [23], adaptive algorithm is formulated which aimed at completion time while meeting deadline constraints. It is modeled by using ATSDS algorithm and implemented on cloudsim. It is compared against CTC and QDA algorithms and it outperforms over existing algorithms with the specified metrics.

The below table 1 represents summary of the above scheduling algorithms which are of the type workflow scheduling, tools used for algorithm and objective of the algorithm.

Table 1: Summary of Cloud Scheduling algorithms

Study	Algorithm Used	Type of Scheduling	Experiment Type	Tools Used
[3]	HSOSEFT	Workflow	Simulation	Workflowsim
[4]	BAT	Workflow	Simulation	MATLAB
[5]	IWD	Workflow	Simulation	Workflowsim
[6]	Jaya	Workflow	Simulation	Cloudsim
[7]	BAT	Workflow	Simulation	Cloudsim
[8]	EIWD	Workflow	Simulation	Cloudsim
[9]	PSO	Workflow	Simulation	Customized
[10]	Coevolutional GA	Workflow	Simulation	Cloudsim
[11]	IMAOPSO	Workflow	Simulation	Workflowsim
[12]	HPSO	Workflow	Simulation	Cloudsim
[13]	ASFLA	Workflow	Simulation	Customized
[14]	BAT	Workflow	Simulation	Workflowsim
[15]	HCROACO	Workflow	Simulation	Cloudsim
[16]	LACO	Workflow	Simulation	Cloudsim
[17]	ECMSMOO	Workflow	Simulation	Cloudsim
[18]	TOPSIS	Workflow	Simulation	Workflowsim
[19]	EATTO	Workflow	Simulation	Customized
[20]	HCRO	Workflow	Simulation	Customized
[21]	SLFA-RSO	Workflow	Simulation	MATLAB
[22]	Tmin-min, Tmax-min	Workflow	Simulation	Cloudsim
[23]	ATSDS	Workflow	Simulation	Cloudsim

From the above table1, summary of workflow scheduling algorithms were depicted and from the above table we can identify that many of authors are using nature inspired algorithms to solve workflow scheduling problem in cloud computing. Many of the authors have used workflowsim and cloudsim as tools to simulate scheduling algorithms.

From the above literature review, we have observed that many of the authors addressed parameters like makespan, Energy consumption, Processing cost, trust, reliability, execution cost, execution time and computational costs. Many of the authors haven't considered the metrics named as utilization of memory, network, and electricity unit cost.

In the next section, we have given few of research directions which needs to be addressed for workflow scheduling algorithms.

RESEARCH DIRECTIONS

Memory Utilization: Memory utilization is to be considered as one of the important parameter as cloud is to be depicted as in terms of compute, storage and network. Whenever utilization of memory is to be high automatically it impacts maximization of resource utilization. So We can consider this parameter as one of the important research direction.

Network Utilization: Network utilization is also to be considered as one of the important Parameters as network is also one of the resources in the cloud. In Cloud Computing, network is also to be given as a virtual service called as SDNs so that we can also consider utilization of network as one of the research direction.

Electricity unit cost: It is an important metric which needs to be considered. Price of electricity varies from place to place around the world and datacenters were spread around globe so based on priority of workflow scheduler needs to migrate workflow to a datacenter where price of electricity is low.

Conclusion and Future Work

Workflow scheduling is a huge challenge in cloud computing as the tasks were dependent on one another and it is very difficult to assign tasks onto corresponding VMs which were resided on physical hosts. In this paper, we have reviewed about the various workflow scheduling algorithms in which they have used various nature inspired algorithms and addressed various parameters makespan, processing cost, reliability and computational costs. Authors haven't addressed the metrics memory utilization, Network utilization and electricity unit cost are some of the important research directions needed to be considered as future research .

directions

REFERENCES

[1] Mell, Peter, and Tim Grance. "The NIST definition of cloud computing." (2011).
[2] Sudheer, M. S., and M. Dr Vamsi Krishna. "Dynamic PSO for task scheduling optimization in cloud computing." International Journal of Recent Technology and Engineering 8 (2019): 332-338.
[3] Anwar, Nazia, and Huifang Deng. "A hybrid metaheuristic for multi-objective scientific workflow scheduling in a cloud environment." Applied sciences 8.4 (2018): 538.
[4] Sagnika, Santwana, Saurabh Bilgaiyan, and Bhabani Shankar Prasad Mishra. "Workflow scheduling in cloud computing environment using bat algorithm." Proceedings of First International Conference on Smart System, Innovations and Computing. Springer, Singapore, 2018.
[5] Kalra, Mala, and Sarbjeet Singh. "An intelligent water drops-based approach for workflow scheduling with balanced resource utilisation in cloud computing." International Journal of Grid and Utility Computing 10.5 (2019): 528-544.
[6] Gupta, Swati, Isha Agarwal, and Ravi Shankar Singh. "Workflow scheduling using Jaya algorithm in cloud." Concurrency and Computation: Practice and Experience 31.17 (2019): e5251.
[7] Raghavan, S., et al. "Bat algorithm for scheduling workflow applications in cloud." 2015 International Conference on Electronic Design, Computer Networks & Automated Verification (EDCAV). IEEE, 2015.
[8] Elsherbiny, Shaymaa, et al. "An extended Intelligent Water Drops algorithm for workflow scheduling in cloud computing environment." Egyptian informatics journal 19.1 (2018): 33-55.
[9] Gupta, Rishabh, Vatsal Gajera, and Prasanta K. Jana. "An effective multi-objective workflow scheduling in cloud computing: a PSO based approach." 2016 Ninth International Conference on Contemporary Computing (IC3). IEEE, 2016.
[10] Butakov, Nikolay, and Denis Nasonov. "Co-evolutional genetic algorithm for workflow scheduling in heterogeneous distributed environment." 2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT). IEEE, 2014.
[11] Saeedi, Sahar, et al. "Improved many-objective particle swarm optimization algorithm for scientific workflow scheduling in cloud computing." Computers & Industrial Engineering 147 (2020): 106649.
[12] Verma, Amandeep, and Sakshi Kaushal. "A hybrid multi-objective particle swarm optimization for scientific workflow scheduling." Parallel Computing 62 (2017): 1-19.
[13] Kaur, Parmeet, and Shikha Mehta. "Resource provisioning and work flow scheduling in clouds using augmented Shuffled Frog Leaping Algorithm." Journal of Parallel and Distributed Computing 101 (2017): 41-50.
[14] Kaur, Navneet, and Sarbjeet Singh. "A budget-constrained time and reliability optimization bat algorithm for scheduling workflow applications in clouds." Procedia Computer Science 98 (2016): 199-204.
[15] Nasr, Aida A., et al. "Cost-Effective Algorithm for Workflow Scheduling in Cloud Computing Under Deadline Constraint." Arabian Journal for Science and Engineering 44.4 (2019): 3765-3780.
[16] Wu, Quanwang, et al. "Deadline-constrained cost optimization approaches for workflow scheduling in clouds." IEEE Transactions on Parallel and Distributed Systems 28.12 (2017): 3401-3412.
[17] Yao, Guangshun, et al. "Endocrine-based coevolutionary multi-swarm for multi-objective workflow scheduling in a cloud system." Soft Computing 21.15 (2017): 4309-4322.
[18] Chakravarthi, K. Kalyan, L. Shyamala, and V. Vaidehi. "TOPSIS inspired cost-efficient concurrent workflow scheduling algorithm in cloud." Journal of King Saud University-Computer and Information Sciences (2020).
[19] Gu, Yi, and Chandu Budati. "Energy-aware workflow scheduling and optimization in clouds using bat algorithm." Future Generation Computer Systems 113 (2020): 106-112.
[20] Singh, Vishakha, Indrajeet Gupta, and Prasanta K. Jana. "An energy efficient algorithm for workflow scheduling in IAAS cloud." Journal of Grid Computing (2019): 1-20.
[21] Karpagam, M., K. Geetha, and C. Rajan. "A reactive search optimization algorithm for scientific workflow scheduling using clustering techniques." Journal of Ambient Intelligence and Humanized Computing (2020): 1-9.
[22] Gupta, Punit, et al. "Trust aware workflow scheduling in Scalable Cloud Environment." (2019).
[23] Khorsand, Reihaneh, et al. "ATSDS: adaptive two-stage deadline-constrained workflow scheduling considering run-time circumstances in cloud computing environments." The Journal of Supercomputing 73.6 (2017): 2430-2455.