



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

A Survey on Workflow Scheduling in Cloud Computing Environment

Manish Gautam¹, Rakesh Kumar²

M.E Scholar, Dept. of Computer Science and Engineering, NITTTR, Panjab University, Chandigarh, India¹

Assistant Professor, Dept. of Computer Science and Engineering, NITTTR, Panjab University, Chandigarh, India²

ABSTRACT: Cloud computing is a practical approach which allows central resources to be served anywhere anytime on pay per basis. The fundamentals of cloud computing are based on principle of reusability of Information technology capabilities. Still there are many challenges in cloud computing. One of them is scheduling. To balance the load in cloud systems, the resources and workload must be scheduled in a good fashion so that maximum use of available resources can be made. There are various scheduling algorithms used by the load balancer to schedule the tasks that are running on any virtual machine. In this paper, we discuss the various scheduling algorithms that improve the utilization of resources allocated to the tasks and maximize the resource utilization.

KEYWORDS: Cloud Computing, Scheduling Algorithms, Load Balancing, Queen-Bee Algorithm.

I. INTRODUCTION

Cloud computing is new IT delivery model, which enables users to store and access data according to their need irrespective of time and place. The cloud is just a metaphor for the Internet. It will be the third revolution in the IT industry that models the development of software from hardware and services from software, and from centralized services to distributed services. Cloud computing is the latest technology that provides computational resources, storage and many more computing services on a pay per basis. Cloud computing provides all services on the basis of virtualization in which cloud provider provides virtual machine to the user on his demand [1]. In scheduling a big task is divided in sub tasks and resources are allocated to each of the subtask to be executed successfully in a well scheduled manner. Resources are allocated in best way to achieve some decided objective. Every workflow has a parent child relationship, in which each task is treated as a node and every edge represents the relationship between each node. Based on the task relationship with each other, the tasks can be categorized as dependent and independent tasks. The tasks which do not depend on each other are called independent tasks and can be executed in isolation on available resources and there is no need to check any precedence order. The dependent tasks are different from independent tasks since there is a parent- child relationship between tasks, we cannot provide resources to the child before completion of its parent task. The main objective in workflow scheduling is to minimize the make span and maximize the utilization of resources where makespan is the latest finish time of the last task in the workflow. Thus, the task dependency plays an essential role in deciding the appropriate scheduling strategy.

Rest of the paper is structured as follows. In section II, we discuss the related work. Section III gives research issues and challenges in workflow scheduling. In section IV, we discuss finding from literature survey. Finally, Section V gives the concluding remarks of the paper.

II. RELATED WORK

A lot of work has been done in the past in the area of cloud computing. Following are few of the important researches done.

Sheng-Jun Xue et al. [2] presented a QoS based hybrid particle swarm optimization (GHPSO) to map applications on to cloud resources. In GHPSO, crossover and mutation of genetic algorithm are embedded in the particle swarm optimization algorithm (PSO), so that it can play a role in the discrete problem, in addition, variability index, changing with the number of iterations, is proposed to ensure that the population can have the higher global search ability during



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

the early stage of evolution, without the premature phenomenon. The simulation results show that the GHPSO achieves better performance than the standard particle swarm algorithm used in minimizing costs within a given execution time.

Medhat A. Tawfeek et al. [3] proposed a modified ant colony optimization for cloud task scheduling. The goal of the modification is to enhance the performance of the basic ant colony optimization algorithm and optimize the task execution time in view of minimizing the makespan of a given task set. This approach proposes self-adapting criteria for the basic ant colony optimization control parameters. Ant colony optimization algorithm and modified algorithm have been simulated using the Cloudsim. Results obtained from the simulation experiment shows that the modified ant colony optimization performs way better than the prior.

Sourabh Bilgaiyan et al. [4] presented heuristic scheduling algorithm based on CSO. The main aim of this application was to map the tasks onto available resources to get desired results. The two parameters that were considered in proposed work include - execution cost of tasks on different resources and data transmission cost between two dependent resources. An imaginary workflow was used for the experimentation and the workflow scheduling results were compared with the existing PSO algorithm which shows that proposed work gives improved results over existing PSO in terms of number of iterations. The proposed work also ensures reasonable load distribution on the available resources.

Jie Huang [5] proposed a genetic algorithm based model to solve problem of workflow scheduling in cloud environment. They mix characteristics of task scheduling with logical design genetic factors and apply it for scheduling of scientific workflow tasks in cloud environment. The proposed algorithm gives improved performance in terms of efficiency of task scheduling, and also satisfies the QoS requirements of the users.

S. Selvarani et al. [6] proposed an algorithm based on improved cost to perform scheduling in cloud environment to maximize utilization of resources. This algorithm gives a measure of both communication cost and computation depending on performance. This also improves the computation/communication ratio.

Yun Yang et al. [7] proposed an algorithm which considers both cost and time as QoS parameters. The result analysis shows that this algorithm can performs better in terms of cost that is lower as compared to other algorithms, deadline parameters are also considered.

Ke Liu et al. [8] gives a workflow scheduling algorithm which considers the properties of cloud computing, using fitness function based on combination of time and cost with user input enabled on the fly. Author performs workflow scheduling and tries to reduce makespan and cost along with maximum utilization of resources.

Suraj Pandey et al. [9] exhibited a heuristic approach named "Particle Swarm Optimization" in which applications are scheduled on cloud resources, as they considered both computation cost and data transmission cost. Using workflow application they performed various experiments by varying computation cost and communication cost. On comparing the results of PSO and 'Best Resource Selection' algorithm, they observed that PSO is three times more efficient than BRS in terms of cost and also provides fine allocation of workload onto resources. Experimental results demonstrate that PSO could accomplish as much as three times cost reduction as compared to BRS, and a fine allocation of workload onto resources. PSO also allocate the tasks to available resources in order to balance the load on computation resources.

Cui Lin et al. [10] presented a scheduling algorithm that is SHEFT (Static hybrid earliest finish time) to perform workflow scheduling in cloud computing environment, this algorithm is modified version of EFT (earliest finish time). The experimental results shows that it performs better mapping of tasks to resources and optimize makespan and also scales resources elastically at runtime.

Yu et al. [11] gave the idea that how constraint budget can be applied in the GA for task scheduling, because users want to use services of the grid based on QoS constraints. User specified budget is used in fitness function depending upon which the schedule of machines is evaluated that whether it should be taken in the population for



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

further operations of GA or not. Initial population is randomly generated and fitness operation is performed using cost fitness and time-fitness method. Two point crossover methods are used as crossover operation.

Zhang et al. [12] proposed an improved version of genetic algorithm to increase the efficiency. For this the chromosomes of poor fitness are hybridized and mutated with the overall best individual, it increases the population diversity and also increases the convergence rate of population. They used fitness function based on make span and “roulette wheel” as selection operator. The result analysis shows that IGA performs better than GA in grid environment.

Delavar et al. [13] proposed a hybridized genetic algorithm to obtain quick response. Author performed workflow scheduling algorithm using this algorithm (HSGA) and considers makespan, load balancing and speed up ratio as QoS parameter. Best fit and round robin algorithm to generate initial population are used. Crossover operation used in this algorithm “random gene selection crossover”.

Rodrigo A et al. [14] presented optimization and novel schema for mobile robot to improve its performance using a genetic algorithm Queen Bee. Control strategy used for the robot kinematics and a PID controller for the robot dynamics is based on fuzzy logic. The fuzzy control is then optimized using queen bee optimization algorithm. This optimized fuzzy method gives better performance. Finally, many optimal controllers are joined together in a way to create an adaptive controller that can handle common cases in an effective manner.

Salim Bitam [15] proposed BLA (Bee Life Algorithm) for efficient scheduling in cloud environments. Scheduling is NP-Complete problem and aim of scheduling is to spread the jobs in optimized way to improve utilization of resources. Results of simulation experiment shows that the BLA performs better than GA in term of execution time, with least complexity.

Z. Pooranian et al. [16] used Queen-bee algorithm in wireless sensor networks to create energy efficient clusters. Reproduction process in wireless sensor networks is optimized by Queen Bee algorithm. Author simulate proposed algorithm using J-Sim and results shows that the clustering by the QB algorithm decreases the energy consumption as compared to existing algorithms and increases the lifetime of the network.

Zahra Pooranian et al. [17] proposed Queen-Bee algorithm, was used for scheduling the independent tasks and was compared with four other offered algorithms. The calculation results showed that the proposed algorithm produces less makespan than GA, SA, GSA algorithms and it is also more suitable regarding the execution time.

Qin, L.D et al. [18] proposed a novel queen-bee evolution for solving the optimization problem of economic power dispatch. Queen-bee algorithm’s performance is tested using typical system of 6 generators, 13 generators and is found that the proposed algorithm outperforms GA (genetic algorithm), in terms of speed and robustness and is proved numerically.

Table 1: A Brief Comparison among Various Workflow Scheduling Algorithms

Algorithm Name	Basic Purpose	Nature of Algorithm	Factor Considered	Tool Used
Deadline and Cost based Workflow Scheduling in Hybrid Cloud [19]	Workflow Scheduling	Hybrid	Deadline-Constraint, Cost Minimization;	Cloud Sim
Genetic Algorithm Based QoS Aware Service Compositions[20]	Independent Scheduling	Genetic	Makespan, Different QoS;	Cloud Sim
Heuristic Algorithm for Multi-Objective [21]	Independent Scheduling	List Scheduling	Makespan, Cost Optimization, Energy Efficient, Reliability;	GroudSim

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

A Particle Swarm Optimization(PSO)[22]	Independent Scheduling	Particle Swarm Optimization	Computation and transmission cost	Amazon EC2
Improved Intelligent Water Drops Algorithm [23]	Independent Scheduling	Intelligent Water Drops	Makespan Optimization;	Cloud Sim
Improved Genetic Algorithm (IGA) [24]	Independent Scheduling	Genetic	Execution time, Resource Utilization, Speed, CPU utilization,	Eucalyptus
Multiple QoS Constrained Scheduling Algorithm[25]	Independent Scheduling	Hybrid	Time, Cost, Make span, Scheduling success rate	CloudSim
Heuristic based Genetic Algorithms (HGAs) [26]	Independent Scheduling	Genetic	Execution cost, Execution time, Data transfer cost	Java Environment
Integer linear program (ILP) formulation [27]	Independent Scheduling	List Scheduling	Makespan, Cost, Time	Java Environment
A Market Oriented Hierarchical Scheduling Strategy in Cloud Workflow Systems [28]	Workflow Scheduling	Genetic, Ant Colony Optimization and particle swarm Optimization	Makespan, Cost and Resource Utilization	Cloud Sim
Deadline-Constrained Workflow Scheduling Algorithm for Infrastructure as a Service [29]	Workflow Scheduling	PCP (Partial Critical Path)	Deadline-Constraint, Cost Minimization	Cloud Sim
The Hybrid Cloud Optimized Cost scheduling algorithm(HCOC)[30]	Independent Scheduling	Hybrid	Cost Optimization	Cloud sim
Revised Discrete Particle Swarm Optimization (RDPSO)[31]	Independent Scheduling	Particle Swarm Optimization	Makespan and Cost Optimization;	Amazon Elastic Compute Cloud
An Ant Colony Optimization Algorithm for Load Balancing [32]	Load Balancing	Ant Colony Optimization	Load Balance;	Cloud Sim
The look ahead genetic algorithm (LAGA)[33]	Workflow Scheduling	Genetic	Makespan, Reliability;	Cloud Sim
Queen-Bee optimization Algorithm[17]	Independent Scheduling	Genetic	Makespan	Matlab

III. RESEARCH ISSUES AND CHALLENGES IN WORKFLOW SCHEDULING

According to literature survey of workflow scheduling we found that there are various research issues and challenges.

These are pool of resources on the cloud that are provided on demand. It is not easy to find out that which resource will be available at the actual execution time of task scheduling.

It is difficult to handle the workflow of the application, which workflow graph is dynamic and change with time.

It is actually intricate to shrink the overhead involved while generating schedules for several task dependencies because there can be more than one users contending for common resources and decisions must be made in possible minimal time.

The virtual instances run on physical machines. When physical machine fails due to hardware or any other failure, the entire workflow application might need to be restart. It is complicated to transfer one workflow application running on a virtual machine to another one.

In the following table “Y” represent that this QoS parameter had been considered whereas “X” means there is a scope for further experiment.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Table 2: Table Showing Different Areas Which Require Further Attention and the Areas Which Have Already Been Explored

QoS constraints Scheduling Algorithms	Make Span	Cost	Budget Constrained	Deadline Constrained	Reliability	Energy Efficiency	Scheduling success rate	Resource Utilization	Load Balancing
Hybrid	Y	Y	X	Y	X	X	Y	X	Y
Modified Genetic	Y	Y	Y	Y	Y	X	X	Y	X
List Scheduling	Y	Y	Y	Y	Y	Y	X	X	Y
Particle Swarm Optimization	Y	Y	Y	Y	X	Y	X	Y	Y
Intelligent Water Drops	Y	X	X	X	X	X	X	X	X
Partial Critical Path	X	Y	X	Y	X	X	X	X	X
Ant Colony Optimization	Y	Y	Y	Y	X	X	X	Y	Y
Queen-Bee Optimization	Y	X	X	X	X	X	X	X	X

IV. FINDING FROM LITERATURE SURVEY

Numbers of authors have done work in the area of workflow scheduling. Literature review presents the description of work done in the field of scheduling using PSO, ACO, GA and modified versions of these algorithms [34][35][36]. In [14], author presented a control scheme for a mobile robot and an optimization method for improving its performance using fuzzy controller optimization. In [15] author proposed a new Bee Swarm optimization algorithm called Bees Life Algorithm (BLA) applied to efficiently schedule computation jobs among processing resources onto the cloud datacenters in a way to reduce execution time by spreading the jobs on to available resources. In [16] author proposed the Queen-bee algorithm to create energy efficient clusters in wireless sensor networks that shows that cluster creation using queen bee helps in decreasing the energy consumption as compared to other available algorithm. In [18], author proposed, evolution method termed queen-bee evolution is employed for solving the optimization problem of economic power dispatch to obtain a system that is fast and more robust. In [17] author proposed Queen-Bee algorithm, The Queen-bee (QB) is leader of mated female lives in bees hive or colony that is responsible for reproduction process, concept was used for scheduling the independent tasks and results was compared with four other offered algorithm.

V. CONCLUSION

Cloud Computing offers a way to increase data storage facility, computation power and add capabilities without investing in new infrastructure. In other words, everyone wants to access resources on the internet very fast for saving the time. So in order to do all that, it needs a good task scheduling algorithm. Scheduling is the main concern in the cloud computing system for satisfying the demands of cloud clients as well as of the cloud providers. That means if we could decrease the overall execution time of a set of tasks on a set of resources and increase the utilization of resources in a cloud computing system. So we can say that this is a good idea to apply Queen Bee Optimization Algorithm so that the intention to implement for better scheduling of tasks on resources with minimum completion time can be achieved.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

REFERENCES

- [1] A. Quarati, A. Clematis, A. Galizia, D. Agostino, M. Mangini, "Delivering cloud services with QoS requirements: an opportunity for ICT SMEs", 9th International Conference on Economics of Grids, Clouds, Systems, and Services (Springer, Berlin), Page No. 197–21, 2012
- [2] S.J. Xue, W. Wu, "Scheduling Workflow in Cloud Computing Based on Hybrid Particle Swarm Algorithm," *Telkommnika*, Volume.10, Number.7, pp. 1560-1566, November 2012.
- [3] M.A. Tawfeek, A. El-Sisi, A.E. Keshk, F.A. Torkey, "An Ant Algorithm for cloud task scheduling," *Proceedings of International Workshop on Cloud Computing and Information Security*, pp. 169–172, 2013.
- [4] S. Bilgaiyan, S. Sagnika, M. Das, " Workflow scheduling in cloud computing environment using Cat Swarm Optimization," *Advance Computing Conference, IEEE International*, pp- 680 – 685, February 2014.
- [5] J. Huang, "The Workflow Task Scheduling Algorithm Based on the GA Model in the Cloud Computing Environment," *Journal of Software*, volume. 9, pp. 873-880, April 2014.
- [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", *Proc. of 4th IEEE International Conference on e-Science*, 374-375, Indianapolis, USA, December 2008.
- [7] K. Liu, Y. Yang, J. Chen, X. Liu, D. Yuan and H. Jin, "A Compromised-Time- Cost Scheduling Algorithm in Swin De W-C for Instance-intensive Cost-Constrained", *International Journal of High Performance Computing Applications*, vol.24, pp. 445-456, May, 2010.
- [8] Ke Liu, "Workflows on Cloud Computing Platform", *International Journal of High Performance Computing Applications*, vol.24 no.4 445-456, May, 2010.
- [9] S. Pandey, L. Wu, S. M. Guru, R. Buyya, " A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments", *24th IEEE International Conference on Advanced Information Networking and Applications (AINA)*, pp. 400 - 407, 2010 .
- [10] C. Lin, S. Lu, "Scheduling Scientific Workflows Elastically for Cloud Computing", in *IEEE 4th International Conference on Cloud Computing*, 2011.
- [11] Z. Wu, X. Liu, Z. Ni, D. Yuan, Y. Yang, "A Market-Oriented Hierarchical Scheduling Strategy in Cloud Workflow Systems" , in *JSC 2010*.
- [12] J. Yu, and R. Buyya, "A Budget Constrained Scheduling of Workflow Applications on Utility Grids Using Genetic Algorithms", *IEEE Workshop on Workflows in Support of Large-Scale Science*, Paris, pp. 1-10, 2006.
- [13] X. Zhang, and W. Zeng, "Grid Workflow Scheduling Based on Improved Genetic Algorithm", *IEEE International Conference on Computer Design and Applications (ICCD)*, Qinhuangdao, vol. 5, pp. 270-273, 2010.
- [14] A. G. Delavar and Y. Aryan, "HSGA: A Hybrid Heuristic Algorithm for Workflow Scheduling in Cloud Systems", *Springer Conference on Cluster Computing*, pp. 1-9, 2013.
- [15] A. Rodrigo, C. Schmidt, "Queen Bee" genetic optimization of an heuristic based fuzzy control scheme for a mobile robot", *IEEE First Latin American Conference on Robotics and Automation* , November 2003.
- [16] S. Bitam "Bees Life Algorithm for Job Scheduling in Cloud Computing", *International Conference on Communications and Information Technology*, 2012.
- [17] Z. Pooranian, A. Barati and A. Movaghar, "Queen-bee Algorithm for Energy Efficient Clusters in Wireless Sensor Networks" *World Academy of Science, Engineering and Technology Vol:5*, 2011.
- [18] N. Chopra, S. Singh, "Deadline and Cost based Workflow Scheduling in Hybrid Cloud", *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI): IEEE*, 2013, pp. 840-846.
- [19] S. Sawant, "A Genetic Algorithm Scheduling Approach for Virtual Machine Resources in a Cloud Computing Environment", 2011, *Master's Projects*. Paper 198.
- [20] H. M. Fard, R. Prodan, J. J. D. Barrionuevo and T. Fahringer, "A Multi-Objective Approach for Workflow Scheduling in Heterogeneous Environment," *Cluster, Cloud and Grid Computing 12th IEEE International Conference*, 2012, pp.300-309.
- [21] S. Pandey, L. Wu, S. M. Guru and R. Buyya, "A Particle Swarm Optimization-Based Heuristic for Scheduling Workflow Application in Cloud Computing Environments", *Advance Information Networking and Applications: IEEE International Conference*, April 2010, pp. 400-407.
- [22] S. H. Niu, S. K. Ong and A. Y. C. Nee, "An Improved Intelligent Water Drops Algorithm for Achieving Optimal Job Shop Scheduling Solution", *International Journal of Production Research*, vol. 50, Jun. 2012.
- [23] H. Zhong, K. Tao, X. Zhang, "An approach to optimized resource scheduling algorithm for open-source cloud systems", *Fifth annual china grid conference (IEEE)*, pp. 124-129, 2010.
- [24] M. Xu, L. Cui, H. Wang, Y. Bi, "A multiple QoS constrained scheduling strategy of multiple workflows for cloud computing", *IEEE international symposium on parallel and distributed processing with applications*, pp. 629-634, 2009.
- [25] A. Verma, S. Kaushal, "Deadline constraint heuristic based genetic algorithm for workflow scheduling in cloud", *Forthcoming article in international journal of grid and utility computing*, vol 5, pp.96-106, 2014.
- [26] 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 (NOMS): mini-conference*, pp. 906-912, 2012.
- [27] Z. Wu, X. Liu, Z. Ni Dong Yuan and Y. Yang, "A Market Oriented Hierarchical Scheduling Strategy in Cloud Workflow Systems. *The Journal of Super Computing*", Vol. 63, pp. 256-293, Springer US, 2010.
- [28] S. Abrishami, M. Naghibzadeh and Dick H. J. E. Pema, "Deadline-Constrained Workflow Scheduling Algorithm for Infrastructure as a Service. *Journal Future Generation Computer Systems*", Vol. 29, Issue 1, pp. 158-169, 2013.
- [29] Jia Yu, Raj Kumar Buyya and Kotagiri Ramamohanarao, "Workflow Scheduling Algorithm for Grid Computing Meta-heuristics for Scheduling in Distributed Computing Environment", *Springer Berlin Heidelberg*, Vol. 146, pp. 173-214, 2008.
- [30] L. Fernando, B. E. Roberto, and M. Madeira, "HCOC: A Cost Optimization Algorithm for Workflow Scheduling in Hybrid Clouds, in *cloud computing*", vol. 2, pp. 207-227, 2011.
- [31] Z. Wu, Z. Ni, L. Gu and X. Liu, "A Revised Discrete Particle Swarm Optimization for Cloud Workflow Scheduling", *Computational Intelligence and Security IEEE International Conference*, pp. 184-188, 2010.
- [32] T. Keskinurk, M. B. Yildirim and M. Barut, "An Ant Colony Optimization Algorithm for Load Balancing in Parallel Machines with Sequence-Dependent Setup Times, *Computer and Operations Research*", vol. 39, pp. 1225-1235, June 2012.



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

- [33] X. Wang, C. S. Yeo, R. Buyya and J. Su, "Optimizing the Makespan and Reliability for Workflow Applications with Reputation and a Lookahead Genetic Algorithm", *Journal Future Generation Computer Systems*, vol. 27, pp. 1124-1134, 2011.
- [34] Poonam, M. Dutta, and N. Aggarwal, "Meta-Heuristics based approach for workflow scheduling in cloud computing : A survey", *Artificial Intelligence and Evolutionary Computations in Engineering Systems Volume 394 of the series Advances in Intelligent Systems and Computing*, Springer, pp 1331-1345, 2015.
- [35] L. Singh and S. Singh, "A Survey of Workflow Scheduling Algorithms and Research Issues", *International journal for scientific research and publications*, vol. 74, no. 15, pp. 21–28, 2013.
- [36] S. Xavier and S. P. J. Lovesum, "A Survey of Various Workflow Scheduling Algorithms in Cloud Environment", *International journal for scientific research and publications*, vol. 3, no. 2, pp. 2–4, 2013.