

Survey Paper on Workflow Scheduling Algorithms Used in Cloud Computing

Murli Manohar Sharma¹ and Anju Bala²

^{1,2}*CSE Department, Thapar University
Patiala, Punjab*

Abstract

Cloud Computing is one of the emerging technologies which expand the boundary of internet by using the centralized servers to maintain data and resources. It enables the users and the consumers to use various applications provided by the cloud provider. But one of the major problems in front of this is workflow scheduling. The workflow scheduling is nothing but the scheduling algorithm, which is employed for the purpose of mapping the requests of users to the appropriate resources available. The scheduling of the workflow is typically performed manually by the IT staff. In other words, workflow scheduling is a kind of automation of the workflows, by using any algorithm. In this paper, importance of workflow, its basic reference model and the various existing scheduling algorithms along with their various facets are also tabulated.

Keywords- Cloud Computing, Workflow management, Reference Model, Existing workflow scheduling algorithm.

1. Introduction

1.1 Cloud Computing: Cloud Computing is receiving a great deal of attention both in publications and among users. Cloud Computing is the name which comes in front of everyone who uses the internet anyway. Cloud Computing is internet based computing, whereby computers, the other shared resources, software and information are provided to the other computers and the resources, on demand and generally pay-per-usage basis. In other words it is a subscription based service where the users can obtain networked storage space and various other computer resources. The cloud makes it possible to access the information from anywhere, anytime. This is especially helpful to the businesses that cannot afford the same amount of hardware and the other resources that a bigger can do. Cloud Computing allows consumers, the

businesses and the other participants to use applications, infrastructure, utilities etc. without any installation. The consumers can access their files by using the Cloud services but the only thing they need is to have a standard connection to the internet. This technology allows much more computing efficiency by centralizing storage, memory, processing and bandwidth.

1.2 Workflow Management: Basically Cloud Computing may have several definitions. According to R.Buyya and S.Venugopal[1] “Cloud Computing is a type of parallel and distributed system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreement established through negotiation between the service provider and the consumer”. Sun Microsystems [2] gives a view that there are many different types of clouds like public cloud, private cloud, community cloud and hybrid cloud. These are the deployment models of the cloud. Cloud can generally provide three levels of services: Iaas (Infrastructure as a service), Paas (Platforms as a service) and Saas (Software as a service). Iaas clouds like Amazon provide virtualized hardware and storage on which the user can deploy their own applications and services. Paas clouds provide an application development environment in which the user can run their applications on the cloud. Saas clouds deliver the software applications to the customers like Google Document or Google Calendar.

Workflow scheduling is a big issue in the era of computing. Basically it is the issue related to the mapping of each task to an appropriate resource and allowing the task to satisfy some performance constraints. 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.

The WfMC (Workflow Management Coalition) defined a workflow as “The automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules.”[3]

WfMC published its reference model, which identifies the interfaces which enabled products to interoperate at a variety of levels. This model elaborates the workflow management and their interfaces. This model is shown in Fig 1.

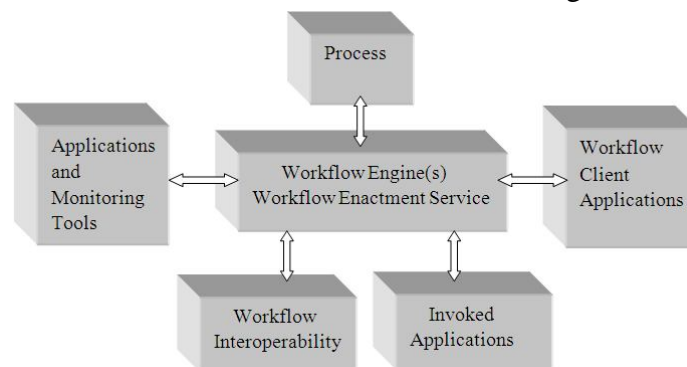


Fig. 1 WfMC Workflow reference Model

Various components are as follows.

1. **Workflow Engine-** A software service which provides the run-time environment to create, manage and execute workflows.
2. **Process-** It is the delegacy of a workflow process which supports automated manipulation.
3. **Workflow Interoperability-** Interfaces to support interoperability between different workflow systems.
4. **Invoked Applications-** Interfaces to support various other IT applications.
5. **Workflow Client Applications-** Interfaces for the interactions with the users.
6. **Administration and Monitoring-** Interfaces to provide system monitoring and metric functions to provide management of composite workflow application environment.

In other words the workflow scheduling is a functional module of the workflow engine(s) and is a significant portion of the workflow management system. Workflow scheduling focuses on mapping and managing the execution of interdependent tasks on the shared resources. A workflow is described by a Directed Acyclic Graph (DAG), in which each computational task is denoted by a node and the data dependencies as well as the control dependencies are represented by edges. With the increasing demand for process automation of the processes in the cloud, the improvement strategies of the workflow scheduling algorithms and the management strategies, is becoming a significant consequence. The workflow scheduling discovers the resources and allocates tasks on the suitable resources. Workflow scheduling plays a vital role in the workflow management as well as in the field of cloud computing.

2. Existing Scheduling Algorithm

Following workflow scheduling algorithms are being used and implemented in the workflow management systems and are summarized in Table-1.

Table-1 Comparison of Existing Workflow Scheduling Algorithms

Sr. No.	Scheduling Algorithm	Scheduling Parameters	Scheduling factors	Findings	Tools
1	Market Oriented Hierarchical Scheduling Algorithm[4]	Makespan, CPU time	DAG scheduling	Minimizing the execution time.	Amazon EC2
2	SHEFT workflow scheduling Algorithm[5]	Scalability and Execution time	Group of tasks	Used to optimize execution time of workflow as well as helps to scale up.	Cloudsim

3	HEFT workflow scheduling Algorithm[7]	Makespan	Highest upward rank	Reduce make span in a DAG	GridSim
4	Ant colony Optimization Based Service flow Scheduling[16]	Resource Utilization, time	QOS	Optimizes the service flow scheduling	Amazon EC2
5	A Particle Swarm Optimization-based Heuristic for Scheduling.[6]	Resource utilization time.	Group of tasks	Good distribution of workload in a cost saving manner.	Amazon EC2
6	Multiple QOS constrained scheduling Algorithm[8]	Success rate, cost, time, Makespan	Multiple Workflows	Scheduling the workflow dynamically.	CloudSim
7	Cost based scheduling on utility grids.[9]	Cost	Task Scheduling	Reschedule the unexpected tasks	GridSim
8	Optimal Workflow scheduling Algorithm[11]	CPU utilization, Execution Time	Multiple Workflows	Improves CPU utilization	Open Nebula
9	RASA Workflow scheduling algorithm[10]	Makespan	Grouped tasks	Reduces makespan	GridSim
10	Workflow with budget constraints[12]	Makespan, budget	DAG scheduling	Minimize the execution time and the makespan	Amazon EC2
11	Improved cost-based scheduling algorithm for task scheduling in cloud computing.[14]	Cost and performance	Computation/Communication Ratio	Make efficient mappings of the tasks	CloudSim
12	An approach to Optimized Resource Scheduling Algorithm[15]	Speed, resource Utilization	IGA(Improved Genetic Algorithm) and automation	Achieves Optimization for cloud scheduling	Eucalyptus

3. Conclusion

With the emergence of cloud computing, the workflow scheduling also becomes the major issue for the purpose of mapping and assignment of resources. So workflow

scheduling is one of the key aspects of cloud computing. Moving workflow to the cloud computing environment enables the use of various cloud services to facilitate workflow execution. In this survey paper, various facets of cloud computing, workflow management system, its reference model (provided by WfMC) and its various components, various workflow scheduling algorithms which are available in the market, are surveyed and analysed.

References

- [1] R. Buyya, C. S. Yeo, and S. Venugopal, "Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities", Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC 2008, IEEE CS Press, Los Alamitos, CA, USA), Sept. 25-27, 2008, Dalian, China.
- [2] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg, and Ivona Brandic, "Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility", Future Generation Computer Systems, Elsevier Science, Amsterdam, June 2009, Volume 25, Number 6, pp. 599-616
- [3] Workflow Management Coalition, Workflow Management Coalition Terminology & Glossary, February 1999.
- [4] Salehi, M.A. and Buyya, R. "Adapting market-oriented scheduling policies for cloud computing", Proceedings of the 10th Int'l Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2010), Busan, Korea, pp. 351-362 (2010).
- [5] Zhangjun Wu¹, 2, Xiao Liu², Zhiwei Ni¹, Dong Yuan², Yun Yang," A Market-Oriented Hierarchical Scheduling Strategy in Cloud Workflow Systems" in JSC2010.
- [6] Pandey, S., Wu, L., Guru, S. and Buyya, R. "A particle swarm optimization based heuristic for scheduling workflow applications in cloud computing environments", *24th IEEE Int'l Conference on Advanced Information Networking and Applications (AINA), Perth, Australia*, pp. 400-407 (2010).
- [7] Wiczorek, M., Prodan, R. and Fahringer, T. "Scheduling of scientific workflows in the ASKALON grid environment", *SIGMOD Rec.*, 34(3), pp. 56-62 (2005).
- [8] Xu, M., Cui, L., Wang, H. and Bi, Y. "A multiple QoS constrained scheduling strategy of multiple workflows for cloud computing", *IEEE 11th Int'l Symposium on Parallel and Distributed Processing with Applications, Chengdu, China*, pp. 629-634 (2009).
- [9] Yu, J., Buyya, R. and Tham, C.K. "Cost-based scheduling of scientific workflow applications on utility grids", *First Int'l Conference on e-Science and Grid Computing, Melbourne, Australia*, pp. 140-147 (2005).
- [10] R. Sakellariou and H. Zhao, "A Hybrid Heuristic for DAG Scheduling on Heterogeneous Systems", The 13th Heterogeneous Computing Workshop (HCW 2004), Santa Fe, New Mexico, USA, April 26, 2004.

- [11] Saeed Parsa and Reza Entezari-Maleki, "RASA: A New Task Scheduling Algorithm in Grid Environment" in *World Applied Sciences Journal 7* (Special Issue of Computer & IT): 152-160, 2009.
- [12] Sakellariou, R., Zhao, H., Tsiakkouri, E. and Dikaiakos, M.D. "Scheduling workflows with budget constraints", *In Integrated Research in GRID Computing*, S. Gorlatch and M. Danelutto, Eds Springer- Verlag., pp. 189–202, (2007).
- [13] S. Selvarani, G.S. Sadhasivam, "Improved cost-based algorithm for task scheduling in Cloud computing", *Computational Intelligence and Computing Research (ICCIC)*, pp.1-5, 2010.
- [14] 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.
- [15] H. Zhong, K. Tao, X. Zhang, "An Approach to Optimized Resource Scheduling Algorithm for Open-source Cloud Systems", in *Fifth Annual China Grid Conference (IEEE)*, pp. 124-129, 2010.
- [16] H. liu, D. Xu, H. Miao, "Ant Colony Optimization Based Service flow Scheduling with Various QoSRequirements in Cloud Computing", *First ACIS International Symposium on Software and Network Engineering*, pp. 53-58, 2011.