ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

COMPARING VARIOUS WORKFLOW ALGORITHMS WITH SIMULATED ANNEALING TECHNIQUE

Dr.V.Venkatesakumar [#1], R.Yasotha [#2]

[#]Department of Computer Science and Engineering, Anna University Regional Centre, Coimbatore, Tamil Nadu, India.
[1]email id: mail2venkatesa@gmail.com
[2]email id: yasotha.r60@gmail.com

***Abstract*— Cloud services are offered to the users through internet by pay-per-use basis in Cloud computing environments. It maintains data and resources using internet. By Cloud Computing technology resources, software and information are shared to several users through data centres. By Scheduling mapping of the tasks to specific resources are done using virtualization. To allocate resources to jobs workflow scheduling algorithms are employed. The problems that arise to the workflow are QoS parameters, cost, time, and information regarding jobs. Each problem has different behaviour in different environments and so specific scheduling algorithms have to be considered for that particular environment. In this paper survey of various workflows scheduling algorithms are done to focus on these problems. Various scheduling algorithms are compared with Simulated Annealing Technique and analyzed.**

***Keywords*— Cloud computing, Virtualization, Workflow Scheduling, Optimization Algorithms, Simulated Annealing**

I. INTRODUCTION

Cloud Computing employs pay-per-use model on large volume of computing resources with the help of virtualization. It stores data and access all resources from anywhere in the world by using internet. By virtualization flexibility and scalability is achieved. The services falls into three types: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS).

- Platform as a service (PaaS) allows to develop, run and manage web applications. PaaS can be provided as a public cloud service from a provider, where the consumer manages software deployment and configuration settings, or software installed in private data centres.

- In an Infrastructure as a Service (IaaS) model, a third-party provider grants hardware, software, virtualization servers, storage and other infrastructure components. IaaS providers handle tasks such as system maintenance, backup and planning. IaaS platforms offer highly scalable on-demand resources. This suite for workloads that are temporary or change unexpectedly.

- Software as a Service (SaaS) is a model in which applications are provided by a service provider and available to customers over Internet. SaaS supports Web services and service-oriented architecture (SOA). This model has indirect distribution and is deployed instantaneously. Traditional SaaS companies built their own pricing structure. It was difficult to work with this pricing structure because of varying price and decrease in the profit of companies. Later SaaS providers price an initial set up cost using a subscription fee and price application based on

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

number of users, processors, etc.

Cloud Computing helps to achieve efficiency in storage, processing, memory and bandwidth. Cloud Computing is a technique where large groups of servers are networked to provide centralized data storage and access to computer services or resources through online. Major types of cloud are public, private, hybrid and community.

- Private cloud focus on small scale organization that are internally managed.
- Public cloud provides applications and resources to the cloud user through brokers which can be fee or pay per use model.
- Hybrid cloud is the composition of public and private.
- Community cloud is those that shall resources between several organizations.

The cloud users, who are in need of a particular service, send a request to the cloud data centre. By the virtualization technology several virtual machines are created on physical machines residing in data centre. Cloud broker is the intermediate between the cloud users and cloud data centres. They tend to allocate the necessary cloud resources to application. By scheduling algorithms and Service Level Agreements (SLA) virtual machines are assigned to workflow application. Workflows are scheduled in order to improve QoS parameters and to perform load balancing among several virtual machines. This helps to improve the virtualization of resources at the service provider.

The objective of this paper is to be survey various workflow scheduling algorithms. The remainder of the paper is summarized as follows: Section 2 presents basic concepts workflow. Section 3 presents survey of workflow algorithms. Section 4 compares workflow algorithms with Simulated Annealing (SA) technique and Section 5 gives the conclusion the work.

II. WORKFLOW

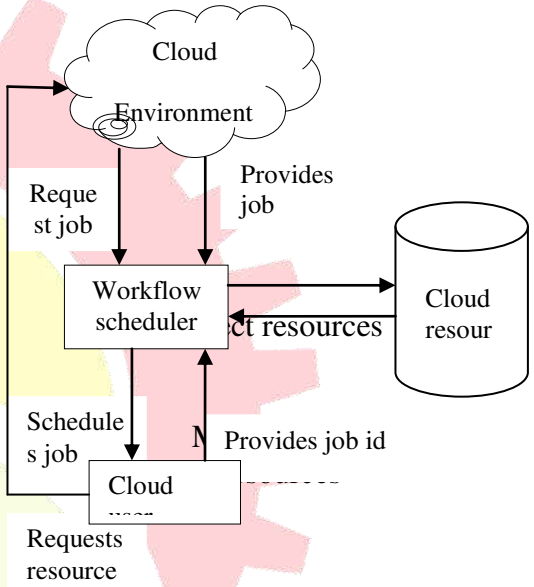

Fig. 1 Overview of various tasks at workflow scheduling environment

A workflow usually maps resources to takes that achieves performance single workflow consists of set of tasks which communicate with each other. Workflow is represented via DAG such as G (V, E) node (V) represents task and directed edge represents dependencies between tasks(E).There are many real time application of workflow where a particular sequence of subtasks has to be executed in order to carry out a main task. Hence in a Directed Acyclic Graph (DAG) series of nodes have to be linked according to set of rules to achieve parent child relationship.

The user objectives that are written in Service Level Agreement have to be satisfied by the workflows algorithm by finding the optimal schedule. The objectives of the user can be cost, optimal, scalabilities, reliability, budget constrain, deadline construction make span optimize.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

Workflow algorithm is classified as heuristic and Meta heuristic. In heuristic also user assigns priority to workflow application based on his own experience and is problem centric. Meta heuristic is independent of human interaction and provides general solution to workflow application.

III. WORKFLOW SCHEDULING ALGORITHMS

The various workflow scheduling algorithms that help in optimizing several QoS parameters are listed below:

- Genetic Algorithm
- List Scheduling Algorithm
- Ant Colony Optimization
- Particle Swarm Optimization
- Cat Swarm Optimization

A. *Genetic Algorithm*

In workflow scheduling mapping inter-dependent tasks to the distributed resources comes under "NP hard problem". To generate optimal results within polynomial time exhaustive search method is used. Genetic Algorithm provides an efficient solution using various search techniques. By the principle of evolution best solution can be obtained from large space in polynomial time.

Genetic Algorithm combines the best solutions obtained from the past searches with the new solution space. It maintains several search space which are represented as individuals.

Jia Yu et.al explains that Genetic algorithm generates an initial population with randomly generated solutions. It produces new solution by applying crossover and mutation and calculates fitness value. Quality of individual is found from fitness function. Hence Genetic Algorithm tends to minimize execution time and meeting user specified budget by evaluating the cost fitness and time fitness. This algorithm help to find the optimum schedule for workflow application that minimizes the time of execution and satisfies user budget.

B. *List Scheduling Algorithm*

Grid Application Development Software infrastructure is used for scheduling workflows on heuristic strategies. It makes use of heterogeneous grid resources which minimizes overall job completion time. The resources are given specific rank values which reflect the performance of a particular component. The performance matrix based on rank values and map components to corresponding resources is constructed.

Anoban Mandal et.al explains that multi-objective list scheduling in Grids tends to find a solution based on the constraint vector. Initially the workload of the workflow activities containing constraint vector. At first the workload of the workflow activities containing constraint vectors are partitioned. The portioning is done based on the user objectives. The ranks are assigned to each workflow order them in ascending manner and maps to appropriate resource.

C. *Ant Colony Optimization*

Timur keskinturk et.al presents that ant colony optimization helps users to achieve quality of service parameters in workflow scheduling. Ant colony optimization algorithm helps the ants to identify the optimal route between their colonies and food source using a kind of chemical to communicate. The chemical is called pheromone is deposited randomly on the path where ants move to find the food. After a certain time by sensing this chemical other ants follow this path and pheromone level increases on this path. When Ant colony optimization is applied on optimization problems these ants traces path or chooses next node towards optimal solution.

If several routes are present ants tend to move in all the routes equally. The preference of the

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

route choose by ants depend on the frequency of the route used. The pheromone level is greater for routes that are shorter distance to destination. This has 2 elements: evaporation of pheromone and amount of pheromone. The evaporation of pheromone reduces the importance of the earlier obtained solutions. If the pheromone level increases it helps to ensure the importance of the optimal solution. This algorithm tends to effectively balance the load when compared with heuristic and genetic algorithm.

D. *Particle Swarm Optimization*

Suraj Pandal et.al explains that particle swarm optimization observes the behavior of the birds referred as particle locates the source of food. The movement of the particle has magnitude and direction as parameters referred as velocity of the particle. At instant of time the position of the particle is identified is identified influenced by its best location and location of best particle. The randomly initialized particles have fitness functions producing a fitness value for optimization. The best particle so far reached and best particles among entire population are identified. The previous fitness values are compared with new fitness value to find the optimized solution. The communication cost is updated at each schedule. This also can be used for any number of tasks and resources to minimize the total cost of execution. Particle swarm optimization algorithm optimizes the computation cost and data transmission cost.

E. *Cat Swarm Optimization*

The authors Shu-Chuan Chu et.al present Cat Swarm Optimization (CSO) based heuristic scheduling algorithm used to schedule the tasks to available resources in order to minimize the total cost of execution. The CSO algorithm considers cost of data transmission between dependent resources and execution of tasks on resources. The authors experiment with the proposed CSO algorithm which minimizes the total cost improves existing PSO and distributes load on the available resources.

CSO considers the behaviours of cats, and composed of two modes: tracing mode and seeking mode. Every cat has its own position composed of many dimensions, velocities, a fitness value for those dimensions. The seeking mode describes the cat in resting position, observing around and the next position to move. The various factors which influence seeking mode are size of seeking memory, number of dimensions, ratio for selected dimensions and the next position to move. Initially the cats created are randomly distributed in the solution space. The cats are set to seeking and tracing mode based on the velocities of the cats. By applying fitness function the fitness value of the cats the best solution are found.

IV. SIMULATED ANNEALING TECHNIQUE:

SIMULATED ANNEALING (SA) is a technique to find an improved solution for an optimization problem by demanding random variations of the existing solution. A worse variation may be established as a new solution with a probability. The capacity to avoid becoming trapped at local minima is its advantage. SA is can find the global optimal solution with probability value 1.

The annealing scheduling table has a set of parameters to run the process of SA, which is composed of the initial temperature T_0, the way of temperature reduction and increment counter L of certain temperature, the end rule S. If initial temperature is too high, it is easier to find the global solution but with slow computation and time-consuming. If it is too low down, it is easier to unite to a local optimal solution. If the reduction in temperature is too rapid, the minimum point might be vanished. If it reduces too slowly, the rate of convergence of the algorithm will be significantly

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

reduced. The initial solution and the neighbouring solution of SA are considered importantly.

V. COMPARING VARIOUS WORKFLOW ALGORITHMS WITH SIMULATED ANNEALING TECHNIQUE

- *Genetic Algorithm(GA) vs Simulated Annealing (SA)*

When comparing Simulated Annealing (SA) with Genetic Algorithm, SA provides a new solution by modifying only one solution with a local movement while GA provides solutions by combining two different solutions. If execution time is not limited, one could always perform a complete search, and get the best possible solution. SA obtains high-quality solutions in a shorter time, but cannot improve that when given more time while GA is able to progress the solution constantly when given more time.

- *List-based scheduling vs Simulated Annealing (SA)*

The list-based scheduling heuristic is a fast method to obtain a feasible solution in terms of execution time, but there is no assurance of achieving a good optimal one. In case of Simulated Annealing (SA) it is experimental that an optimal solution is approximated in less amount of time, avoiding local minimums. In short, exact methods might obtain good results but require very large execution times.

- *Particle Swarm Optimization (PSO) vs Simulated Annealing(SA)*

The performance of particle swarm optimization and simulated annealing algorithm is observed to achieve an optimal or near optimal solution. It is accomplished that the particle swarm optimization algorithm gives optimized solutions with local search. The particle swarm optimization algorithm provides a quality solution with local search which is better to that of the simulated annealing algorithm, but the simulated annealing algorithm takes shorter time to find a solution.

- *Ant Colony Optimization (ACO) vs Simulated Annealing (SA)*

The operating times of Ant Colony Optimization (ACO) are enhanced than those of Simulated Annealing (SA). SA takes more time to determine and estimate the neighbouring solutions at each iteration. ACO uses information from previous iterations which reduces the processing time needed to calculate the cost.

- *Cat Swarm Optimization (CSO) vs Simulated Annealing(SA)*

The CSO algorithm can reduce number of search locations and can achieve high accuracy and computation speed can be achieved. It has better performance in function minimization problem when compared to other optimization algorithms. When compared with Simulated Annealing CSO takes long time to find an acceptable solution. Moreover CSO can provide reasonable solution by use of high speed processor.

VI. CONCLUSION

One of the major issues in cloud computing environment is Workflow scheduling. In this paper, survey of various existing workflow scheduling algorithms are done on the basis of nature of

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

scheduling algorithm, objective criteria and the environment to which it was applied. The various performance parameters are estimated on various algorithms and are compared to Simulated Annealing technique. Further from the literature reviewed, there is a need to implement a energy competent algorithm for workflow application that minimize the execution time in cloud environment. Scheduling workflows in a cloud computing environment enables the use of various cloud services.

REFERENCES

[1] Rizos Sakellariou and Henan Zhao. 2004. *Hybrid Heuristic for DAG Scheduling on Heterogeneous Systems. Parallel and Distributed Processing Symposium*, 18th IEEE International Conference

[2] A. Mandal, K. Kennedy, C. Koelbel, G. Marin, J. Crummey and B. Liu. 2005. *Scheduling Strategies for Mapping Application Workflows onto the Grid. High Performance Distributed Computing,* 14th IEEE International Conference.

[3] Marek Wieczorek, Radu Prodan and Thomas Fahringer. 2005. *Scheduling of Scientific Workflows in the ASKALON Grid Environment.* ACM SIGMOD, Vol. 34, Issue 3, Pg. 56-62.

[4] Jia Yu and Raj Kumar Buyya. 2006. *A Budget Constrained Scheduling of Workflow Applications on Utility Grids using Genetic Algorithms. Workflows in Support of Large-Scale Science*, IEEE Conference, Pg. 1-10.

[5] Jai YU and Raj Kumar Buyya. 2006. *Scheduling Scientific Workflow Applications with Deadline and Budget Constraints using Genetic Algorithms. Scientific Programming Journal*, Pg. 217-230, Vol. 14, Issue 3-4.

[6] M. Rahman, S. Venugopal and R. Buyya. 2007. *A Dynamic Critical Path Algorithm for Scheduling Scientific Workflow Applications on Global Grids. E-Science and Grid Computing*, IEEE International Conference, Pg. 35-42.

[7] Wei Neng Chena and Jun Zhang. 2009. *An Ant Colony Optimization Approach to a Grid Workflow Scheduling Problem with Various QoS Requirements. System, Man and Cybernetics, Applications and Reviews, IEEE Transactions*, Vol. 39, Issue 1, Pg. 29-43.

[8] Qian Tao, Hui You Chang, Yang Yi, Chunqin Gu and Yang Yu. 2009. *QoS Constrained Grid Workflow Scheduling Optimization Based on a Novel PSO Algorithm*, Grid and Cooperative Computing. 8th IEEE International Conference, Pg. 153-159.

[9] Yanli Hu, Lining Xing, Weiming Zhang, Weidong Xiao and Daquan Tang. 2010. *A Knowledge Based Ant Colony Optimization for a Grid Workflow Scheduling Problem.* First International Conference Beijing, China, Pg. 241-248, Springer Berlin Heidelberg.

[10] Suraj Pandy, Linlin Wu, Siddeshwara Mayura Guru and Raj Kumar Buyya. 2010. *A Particle Swarm Optimization-Based Heuristic for Scheduling Workflow Application in Cloud Computing Environments. Advance Information Networking and Applications*, IEEE International Conference, Pg. 400-407.

[11] Yong Wang, R. M. Bhati and M. A. Bauer. 2011. *A Novel Deadline and Budget Constrained Scheduling Heuristic for Computation Grids.* Journal of Central South University of Technology Vol. 18, Issue 2, Pg. 465-472.

[12] Eugen Feller, Louis Rilling and Christine Morin. 2011. *Energy-Aware Ant Colony Based Workload Placement in Clouds,* Grid Computing. 12th IEEE International Conference, Pg. 26-33.

[13] S. H. Niu, S. K. Ong and A. Y. C Nee. 2012. *An Improved Intelligent Water Drops Algorithm for Achieving Optimal Job Shop Scheduling Solution.* International Journal of Production Research, Vol. 50, Issue 15.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

[14] Zhangjun Wu, Xiao Liu, Zhiwei Ni Dong Yuan and Yun Yang. 2013. *A Market Oriented Hierarchical Scheduling Strategy in Cloud Workflow Systems*. The Journal of Super Computing, Vol. 63, Issue 1, Pg. 256-293, Springer US.

[15] Simrat Kaur and Sarbjeet Singh. 2012. *Comparative Analysis of Job Grouping Based Scheduling Strategies in Grid Computing*. International Journal of Computer Applications, Vol. 43, Issue 15.