International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

An Exploration of Multi-Objective Scientific Workflow Scheduling in Cloud

Arunkumar Panneerselvam[*1], Bhuvaneswari Subbaraman[2]
[*1]Department of Computer Science, Pondicherry University, Karaikal Campus, Karaikal, Puducherry, India
[2]Registrar, Central University of Tamilnadu, Thiruvarur, Tamilnadu, India

ABSTRACT

Today's scientific community highly relies upon scientific workflow applications for various operations related to their field. Distributed environment like cloud perfectly matches for executing large-scale scientific applications by providing enormous on demand resources. Hence scientific workflows scheduling in cloud environment gain importance. The cloud resources renting follows pay per use model and thus the cloud users are constrained to work under multiple objectives like time and cost. This paper explorers the work of research community in the area of Multi-Objective scheduling of scientific workflows in the cloud environment.
Keywords : Workflow scheduling, Multi-Objective scheduling, Cloud computing

I. INTRODUCTION

This is the era of big data where data drenches from various devices and applications. It is often needed a compatible platform to manage this enormous data. Cloud offers an excellent environment for handling big data and workflows that deals with the big data.

Scientific Workflows are data intensive and compute intensive workflows that are modelled to execute various components of the scientific applications to solve a particular problem.Various scientific workflow management tools that extensively use scientific workflows are now integrated with popular commercial cloud providers. The research community is actively involved in finding good scheduling techniques to schedule the scientific workflows in cloud environment. There are various research challenges and issues involved in scheduling these workflows in cloud Platforms. A workflow is often represented by DAG (Directed Acyclic Graph), G = (V, E) where the V represents vertices which are tasks (T1, T2, T3, ... ,Tn) that needs to be scheduled and executed in the cloud. E represents edges of the graph, which indicates the dependency between two tasks say for T_i and Tj. The dependency may be control dependency and/or data dependency. Data dependency between Tasks T1 and T2 means that task T2 cannot be executed until T1 completes its execution. In large-scale scientific applications, data dependencies and control dependencies between tasks are inevitable.

When scheduling the tasks of the scientific workflows to cloud virtual machines several objectives need to be satisfied. These objectives can be classified based on cloud user, cloud broker and cloud service provider. For example, executing the tasks with less cost and time may be the objective of the cloud user. Allocating the virtual machines so that less energy is conserved may be the objective of the cloud provider. In many situations to attain maximum benefits, we require multiple objectives to be satisfied.

The rest of the paper is organized in the following way. Section II introduces the concepts of multi-objective optimization problems. Section III discuss the various work contributed by the research community in the area of Multi objective scientific workflow scheduling. Section IV discuss about the techniques used for Multi objective optimization. Section V is the conclusion and future research direction.

II. MULTI-OBJECTIVE OPTIMIZATION

Multi-Objective optimization refers to simultaneously optimizing two or more objectives in a problem domain. Often these objectives are conflict in nature. For example, in cloud task allocation and resource scheduling problem the cloud user prefers to minimize the cost of hired virtual resources at the same time wishes to execute tasks in less time. These two objectives of the cloud user usually conflicts because high-end resources when used for task execution results in minimum execution time with increased cost. When low configuration virtual resources are hired, the cost is minimized but the execution time will increase. Hence, there is always a need for trade-off between these two objectives. In Multi objective Optimization, we try to obtain Pareto optimal solution, which is a set of non-dominated solutions. The Multi-Objective problem can be formulated as below,

Minimize $\{f_1(x), f_2(x), \ldots, f_k(x)\}$ (1)
subject to $x \in S$,

Where,

- ✓ f_i: $R^n \rightarrow R$, the objective function.
- ✓ k (≥ 2), the number of (conflicting) objectives functions.
- ✓ x, the decision vector (of n decision variables x_i).
- ✓ $S \subset R^n$, the feasible region formed by constraint functions.

The Multi-Objective optimization can be constrained or non-constrained. The goal of Multi-Objective optimization would be obtaining a set of non-dominated solutions. If P is said to be set of solutions, then the set of non-dominated solution P' are those that are non-dominated by any other member of the set P. The mathematical formulation of the Pareto optimality is given as, A point $\hat{x} \in X$ is said to be Pareto optimal solution (or non-inferior solution) to the problem P, if there is no $x \in X$ such that $f(x) \leq f(\hat{x})$ [1]. From the non-dominated solutions, finally one solution is selected to solve the problem in hand. The following Fig. 1 represents the concepts of Pareto optimality [2].

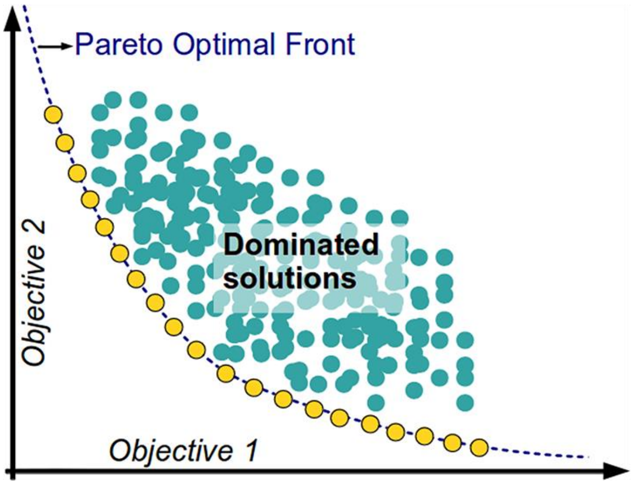

Figure 1. Representation of Pareto Optimal solution [2]

In Figure 1 the yellow circles represent the non-dominated solutions. These solutions are non-competing with each other. The blue circles represent the dominated solution, which means the solutions in the dominated set cannot be improved without the degradation of the objective function. The quality indicators like hypervolume can be used to measure the strength of the Pareto optimality.

III. RELATED WORK

Multi–Objective optimization has been applied for various real world problem domains such as Aerodynamics design, Industrial neural network design, Molecular structures for drugs, Medical decision making, Supply chain management, Interactive aircraft design, Land use planning, Lens and Bridge designs, Computers job scheduling etc. [3]. Many authors have studied the Multi objective task scheduling in cloud computing. The tasks scheduling in cloud environment involves various research challenges. The tasks may be dependent tasks or independent tasks. Many researchers have addressed the single objective task scheduling in their work. Scientific workflow contains dependent tasks. Scientific workflow scheduling with single objective has been addressed in the literature a lot. However, there is limited work in the area of Multi-Objective optimization scheduling with respect to scientific workflows. The following Table.1 list out the important contribution of the research community in the area of Multi-Objective scientific workflow task scheduling in distributed environments.

Table 1. Multi-Objective scientific workflow scheduling works

Ref. No	**Authors**	**Objectives Addressed**	**Technique used**	**Strength & Weakness**
[4]	PhyoThandar Thant, Courtney Powell et. al (2017)	Makespan, Machine instance deployment Cost and VM instance Failure	Level wise workflow Partitioning and Optimization	Budget and other QoS constraints not addressed
[5]	Miao Zhang, RajkumarBuyya et. al (2017)	Total Execution cost, Degree of imbalance	Adaptive individual assessment scheme and uses the number of Pareto individuals and Pareto Entropy	Budget constraint addressed
[6]	Zhaomeng Zhu, Gongxuan Zhang et.al (2016)	Makespan, Cost	Evolutionary Multi Objective Optimization - Problem specific encoding, Population initialization, fitness evaluation and Genetic Operators	Take account of real world Pricing strategies of cloud
[7]	Ji Liu, Esther Pacitti et.al (2016)	Execution time, Cost	Based on Multi Objective cost model	Addressing Workflow Scheduling at multi-site
[8]	Juan J. Durillo, Hamid MohammadiFard et. al (2012)	Makespan, Economic cost	Usage of List based Multi Objective Heterogeneous Earliest Finish time	Employs list based heuristics
[9]	Jia Yu, MiChaelKirley et.al (2007)	Execution time, Cost	Multi Objective Evolutionary Algorithm	Addresses Budget and Deadline constraints
[10]	HeyangXu, Bo Yang et.al (2016)	Makespan, Execution cost	Min-min based time and cost trade off	Addresses Fault Recovery

From the literature, it is evident that only there are limited works in the area of Multi-Objective scientific workflow scheduling in cloud. The survey shows that most of the authors concentrated only on two objectives, that is, Cost and Makespan. Cost refers to the cost of executing a task in the cloud environment and Makespan is the completion time of the last task in the workflow. In addition, it is evident that constraints are addressed less. For example, Budget of the cloud user and the time for executing a particular workflow are the important constraints since these constraints are associated with the Service Level Agreements of the cloud user and the cloud provider. Bi-Objective works are more and tri objective works are very less. The authors has used simple rule based heuristics and Meta-Heuristics techniques to solve the Multi-Objective scheduling optimization.

IV. TECHNIQUES USED FOR MULTI-OBJECTIVE WORKFLOW SCHEDULING

Generally, the task scheduling optimization problem is NP Hard problem and cannot be solved in a Polynomial time. There are many rule based algorithms, Heuristics and Meta-Heuristics algorithms used for solving the Multi-Objective Optimization problems. Hybrid techniques had been used by few researchers. Bio-inspired Population based search techniques are widely employed to solve the task scheduling problem in cloud environment. These algorithms have their own merits and demerits. These algorithms can fetch near optimal solutions that are acceptable. The following algorithms are latest and most often used algorithms for Multi-Objective Optimization.

A. NSGA II (Non-Dominated Sorting Genetic Algorithm)

The Non-Dominated Sorting Genetic algorithm (NSGA) is said to be the first Evolutionary Algorithm for addressing Multi-Objective Optimization proposed by K. Deb et. al. in the year of 2000. This algorithm uses genetic operators like selection, mutation and crossover to generate Population of Pareto dominance. The computational complexities of NSGA are addressed and later in 2002 NSGA II was proposed which used crowding distance to choose the quality solutions. The NSGA II has been applied to many test problems and has been proved to perform better than author evolutionary algorithms.

B. SPEA2 (Strength Pareto Evolutionary Algorithm)

E. Zitzler et.al proposed the Strength Pareto Evolutionary Algorithm in 1999 and SPEA2 in 2001. It is also an Evolutionary Algorithm and is the extension of the original genetic algorithm. It chooses the set of Pareto optimal solutions by considering two important parameters; one is the strength Pareto and the fitness value. SPEA2 gives good results when applied to various test problems and real world multi-optimization problems.

C. MOPSO (Multi Objective Particle Swarm Optimization Algorithm)

The Original Particle Swarm Optimization was proposed by Kennedy et.al. in the year of 1995. In this Bio-Inspired Population based meta-heuristic approach inspired by birds flocking nature, local best and global best solutions are obtained and based on the fitness values final solution is decided. Many variations to Particle swarm optimization have been proposed by many authors for using it in Multi-Optimization Problems.

V. CONCLUSION

This paper explorers the various concepts and works related to Multi-Objective Optimization with respect to scientific workflow scheduling in cloud. Bio Inspired Population based Meta-Heuristics are better than traditional rule based algorithm. However, these algorithms poses computational complexities and to be tailored to particular problem domain. The future research direction can be in creating more general solutions like hyper heuristic algorithms for solving Multi-Objective scientific workflow scheduling in cloud environments.

VI. REFERENCES

[1]. Kaisa M. Miettinen, "Nonlinear Multiobjective Optimization," 1999 ed. Kluwer Academic Publishers, Boston.

[2]. Arthur Francisco, ThomasLavie (2013). Metamodel-Assisted Optimization of Connecting Rod Big-End Bearings", Journal of Tribology. DOI: 10.1115/1.4024555.

[3]. Theodor Stewart and Oliver Bandte, "Real world Applications of Multiobjective Optimization" in Multi objective Optimization: Interactive and Evolutionary Approaches, 2008. Springer Lecture Notes in Computer Science, Verlag Berlin Heidelberg.

[4]. Thant, P., Powell, P., &Schlueteer (2016). Multi-objective level wise Scientific workflow optimization in Iaas Public Cloud, Scientific Programming, vol. 2017, Article ID 5342727, 17 pages, 2017. doi:10.1155/2017/5342727.

[5]. Buyya, R., Li, H., Liu, L., & Zhang, M. (2017).An adaptive multi-objective evolutionary algorithm for constrained workflow scheduling in Clouds. Distributed and Parallel Databases, 1-30.

[6]. Zhu, Z., Zhang, G., Li, M., & Liu, X. (2016).Evolutionary Multi-Objective Workflow Scheduling in Cloud. IEEE Transactions on Parallel and Distributed Systems, 27(5),1344–1357. https://doi.org/10.1109/TPDS.2015.2446459.

[7]. Liu, J., Pacitti, E., Valduriez, P., Oliveira, D. D., Mattoso, M., Liu, J., …Mattoso, M. (2016). Scientific Workflow Execution with Multiple Objectives in Multisite Clouds ,HAL Id : lirmm-01409092 ScientificWorkflow Execution with Multiple Objectives in Multisite Clouds, 4–6.

[8]. Durillo, J. J., Fard, H. M., &Prodan, R. (2012). MOHEFT: A multi-objective list-based method for workflow scheduling. In CloudCom 2012 - Proceedings: 2012 4th IEEE International Conference on Cloud Computing Technology and Science (pp.185–192). https://doi.org/10.1109/CloudCom.2012.6427573.

[9]. Yu, J., Kirley, M., &Buyya, R. (2007).Multi-objective planning for workflow execution on Grids.In Proceedings - IEEE/ACM International Workshop on Grid Computing (pp.10–17).https://doi.org/10.1109/GRID.2007.4354110.

[10]. Xu, H., Yang, B., Qi, W., &Ahene, E. (2016).A multi-objective optimization approach to workflow scheduling in clouds considering fault recovery. KSII Transactions on Internet and Information Systems, 10(3), 976–995. https://doi.org/10.3837/tiis.2016.03.002