Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A hyper-heuristic cost optimisation approach for Scientific Workflow Scheduling in cloud computing

Ehab Nabiel Alkhanak *, Sai Peck Lee

Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia

HIGHLIGHTS

- Proposes a completion time driven hyper-heuristic approach for cost optimisation.
- Provides a comprehensive background on scientific workflow scheduling in cloud.
- Proposed approach helps in saving the cost and time of the cloud service providers.
- Significantly optimises the execution cost and time compared to baseline approaches.

ARTICLE INFO

Article history:
Received 11 December 2017
Received in revised form 22 March 2018
Accepted 26 March 2018
Available online xxxx

Keywords:
Workflow Scheduling
Cost optimisation
Hyper-heuristic approach
Cloud computing

ABSTRACT

Effective management of Scientific Workflow Scheduling (SWFS) processes in a cloud environment remains a challenging task when dealing with large and complex Scientific Workflow Applications (SWFAs). Cost optimisation of SWFS benefits cloud service consumers and providers by reducing temporal and monetary costs in processing SWFAs. However, cost optimisation performance of SWFS approaches is affected by the inherent nature of the SWFA as well as various types of scenarios that depend on the number of available virtual machines and varied sizes of SWFA datasets. Cost optimisation performance of existing SWFS approaches is still not satisfactory for all considered scenarios. Thus, there is a need to propose a dynamic hyper-heuristic approach that can effectively optimise the cost of SWFS for all different scenarios. This can be done by employing different meta-heuristic algorithms in order to utilise their strengths for each scenario. Thus, the main objective of this paper is to propose a Completion Time Driven Hyper-Heuristic (CTDHH) approach for cost optimisation of SWFS in a cloud environment. The CTDHH approach employs four well-known population-based meta-heuristic algorithms, which act as Low Level Heuristic (LLH) algorithms. In addition, the CTDHH approach enhances the native random selection way of existing hyper-heuristic approaches by incorporating the best computed workflow completion time to act as a high-level selector to dynamically pick a suitable algorithm from the pool of LLH algorithms after each run. A real-world cloud based experimentation environment has been considered to evaluate the performance of the proposed CTDHH approach by comparing it with five baseline approaches, i.e. four population-based approaches and an existing hyper-heuristic approach named Hyper-Heuristic Scheduling Algorithm (HHSA). Several different scenarios have also been considered to evaluate data-intensiveness and computation-intensive performance. Based on the results of the experimental comparison, the proposed approach has proven to yield the most effective performance results for all considered experimental scenarios.

1. Introduction

Scheduling the submitted Scientific Workflow Application (SWFA) tasks to the available computational resources while optimising the cost of executing the SWFA is one of the most challenging processes of Workflow Management System (WfMS) in a cloud computing environment [1–3]. The cost optimisation challenge of SWFS in cloud requires the consideration of three main perspectives: (i) strong inter-dependencies between the SWFA tasks. This computational-intensiveness of tasks introduces complexity to the scheduling processes, since the data of the SWFA tasks needs to be transferred between the computational resources (i.e. VMs) of cloud computing, (ii) different sizes of SWFA datasets that need to be considered by the scheduler, which significantly

* Corresponding author.
E-mail addresses: ehabsoa@siswa.um.edu.my (E.N. Alkhanak), saipeck@um.edu.my (S.P. Lee).

https://doi.org/10.1016/j.future.2018.03.055
0167-739X/© 2018 Published by Elsevier B.V.

affect the execution time and execution cost [4,5]. Thus, data-intensiveness needs to be considered while calculating the completion time (makespan) and total cost of executing the tasks of SWFA on available resources (i.e. VMs), and (iii) different numbers of computational resource (VMs) based on the user requirements. Accordingly, considering the abovementioned perspectives makes the SWFS process more complicated and requires a great amount of computational resources in terms of completion time and total computational cost [6–8]. Moreover, a large size of SWFA datasets can cause a significant increase in the dependency among workflow tasks, and this resultant tasks dependency ultimately increases the completion time required to process the SWFA in the available resources. Ultimately, any delay in completion time can negatively impacts on cost optimisation of SWFS.

In the literature, several types of approaches have been proposed for SWFS [2,7,9–11]. The existing population-based meta-heuristics solutions have shown good performance for the optimisation of the large search space problem. In contrast, single-based meta-heuristic solutions do not exhaustively search within the scheduling problem space, yet they use different underlying strategies to find the desired solution based on defined fitness criteria. Therefore, the population-based meta-heuristic solution takes less computational effort as compared to single-based solution while it can often find good solutions [12–15]. However, each of these approaches has its strengths and limitations, which affect the SWFS processes. The hybrid meta-heuristics use the best features of two or more traditional meta-heuristics (e.g., Genetic Algorithm, Ant Colony Optimisation) in each iteration to provide a better optimal solution. Due to the complexity of hybrid meta-heuristics method, it might take a longer convergence time than the traditional meta-heuristics for each iteration [3,16,17]. On the other hand, the hyper-heuristic approach is an emerging class of meta-heuristic algorithms, which is integrated in such a manner that allows utilising the maximum strengths of the employed meta-heuristic algorithms to obtain an optimal solution. The hyper-heuristic mechanisms achieve better performance in terms of shorter execution time compared to other optimisation mechanisms. Additionally, hyper meta-heuristic is a class of new advanced techniques that are capable of accelerating the run-time of a single meta-heuristic algorithm. There are only few works [18–21] that have considered utilising hyper-heuristic for SWFS, while hyper-heuristic can always find the most cost optimal solutions for different scenarios. Thus, the aim of this research is to propose a completion time driven hyper-heuristic approach for cost optimisation of SWFS in a cloud environment for several scenarios. Proposing such approach can optimise the completion time as well as the total computational cost by dynamically selecting the most suitable meta-heuristic algorithm based on completion time performance of the employed meta-heuristic algorithms.

The proposed Completion Time Driven Hyper-Heuristic (CTDHH) approach employs four well-known population-based meta-heuristic algorithms, which act as Low Level Heuristic (LLH) algorithms (i.e., genetic algorithm, particle swarm optimisation, invasive weed optimisation, and hybrid invasive weed optimisation). In addition, the proposed algorithm enhances the native random selection way of existing hyper-heuristic solutions by incorporating the best computed workflow completion time to act as a high-level selector to pick a suitable algorithm from the pool of LLH algorithms after each run. The main aim of the proposed approach is to reduce the completion time and total computational cost to execute the SWFA. Based on the lowest achieved completion time, the proposed algorithm dynamically guides the searching processes to find an optimal solution by continuously sorting the computed time scores (i.e. completion times of previous runs) of all the employed LLH algorithms for each considered scenario and after every run. The computed time scores are listed in a scoreboard table. Next, for each single run, the high-level selector adopts the LLH algorithm that has the lowest computed time score for each scenario. The proposed dynamic hyper-heuristic algorithm continuously updates the scoreboard table by replacing the existing time score with the lowest computed time score, which ultimately affects the total computational cost value for that run. Finally, based on the scoreboard table, the proposed approach selects the most appropriate LLH algorithm for the next run. Consequently, the mechanism of the proposed completion time driven hyper-heuristic approach becomes more effective in allowing to reuse and utilise the maximum strengths of the employed LLH algorithms in searching for the optimal solution of the targeted cost optimisation problem.

To evaluate and analyse the performance of cost optimisation parameters (i.e. completion time and total computational cost) of the proposed approach, the authors have evaluated the proposed approach in a real-world cloud based experimental environment by comparing the proposed approach with the baseline approaches, i.e. four population-based approaches and an existing hyper-heuristic approach named Hyper-Heuristic Scheduling Algorithm (HHSA). Furthermore, there are several considered scenarios with different numbers of VMs, and different sizes of SWFA datasets. The data that are collected from the real-world based environments have been analysed based on the completion time (makespan) and total computational cost parameters.

There are several contributions that have been gained from conducting this research. The following are the key contributions of this research:

- A completion time driven hyper-heuristic approach for cost optimisation of SWFS has been proposed. The proposed CTDHH approach helps in optimising the completion time (makespan) and total execution cost of SWFS in the cloud computing environment.
- The proposed CTDHH approach is profitable for service consumers, by way of reducing the total computational cost utilising the computational resources of the cloud. At the same time, the proposed approach provides more satisfying user requirements (i.g. shorter completion time and cheaper computational cost).
- The proposed CTDHH approach helps in saving the energy and time of the service providers, by judiciously utilising the computational resources. This would ultimately help in reducing the computation cost as well as handling the computation-intensive and data-intensive SWFAs.
- This research would open new doors for a high impact research with innovative values through SWFA and cloud computing.

In order to simplify the readability of the paper, the abbreviation list containing terms used in this paper is shown in Table 1.

This paper is organised as follows: In Section two, a comprehensive definition about the SWFS, and cost optimisation of SWFS is provided. Section three describes the proposed Completion Time Driven Hyper-Heuristic (CTDHH) approach along with an example of the proposed algorithm. Section four discusses the evaluation part by using a real-world cloud based experimentation environment. Next, in Section five, the results and discussion of this research paper are described. Then, Section six provides the related works and finally, Section seven provides a conclusion and future work.

2. Background

In this section, the authors provide a comprehensive detail about the main stages of Scientific Workflow Scheduling (SWFS) and cost optimisation of SWFS.

Please cite this article in press as: E.N. Alkhanak, S.P. Lee, A hyper-heuristic cost optimisation approach for Scientific Workflow Scheduling in cloud computing, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.03.055.