Extending Dynamic Scheduling Policies in WorkflowSim by Using Variance based Approach

Jyoti Thaman, Maharishi Markandeshwar University, Sadopur, India
Manpreet Singh, Maharishi Markandeshwar University, Sadopur, India

ABSTRACT

Workflow scheduling has been around for more than two decades. With growing interest in service oriented computing architecture among researchers and corporate users, different platform like clusters computing, grid computing and most recent cloud computing, appeared on computing horizon. Cloud computing has attracted a lot of interest from all types of users. It gave rise to variety of applications and tasks with varied requirements. Heterogeneity in application’s requirement catalyzes the provision of customized services for task types. Representation of tasks characteristics and inter-task relationship through workflows is in use since the ages of automation. Scheduling of workflows not only maintain the hierarchical relationship between the tasks but also dictates the requirement of dynamic scheduling. This paper presents a variance based extensions of few promising dynamic scheduling policies supported by WorkflowSim. An exhaustive performance analysis presents strength and weakness of the authors’ proposal.

KEYWORDS
Dynamic Scheduling, Makespan, Nearest Neighbor, Task Types, Variance, WorkflowSim, Utilization

1. INTRODUCTION

Cloud computing utilizes services which runs on internet that move computations from self-managed individual resources to on-demand pool of resources in centralized infrastructures. In cloud computing, the end users contract with cloud vendors for customized Virtual Machines (VM) and interact with the VMs using only a console/browser through the internet, with all the data and applications maintained on the remote servers accessible to end users from any device, anywhere and at any time (Kalagiakos & Karampelas, 2011). Considering that energy is converted to product or service through pool of hardware resources for communication and computation; initial work in parallel and distributed computing was to establish power efficient data centers using power aware technologies and components. Dynamic Voltage and Frequency Scaling (DVFS) and dynamic power management (DPM) (Horvath et al., 2007) like approaches are examples of this. That’s not the end. This in fact mandates further analysis and development in line with power efficiency of data centers through software or firmware or hybrid methodology.

User’s submissions are processed in clouds by subjecting tasks to resources. Resource usage in clouds depends upon the types and sequence of tasks and resources. Work flow technologies are used to deal with increasing complex data, data-intensive application, simulations and analysis. These technologies are also used to schedule computational tasks on distributed resources, to manage
dependencies among tasks and to stage data sets into and out of execution sites (Juve et al., 2013). These workflows are used to model computations in many scientific disciplines (Juve & Deelman, 2011). If resources can be used in optimal manner, a lot of financial liabilities of both, cloud users and cloud providers can be reduced. It is therefore important to understand workflows’ behavior so that improved algorithms can be designed for providing resources, scheduling computational jobs and managing data.

In this paper, workflows have been considered to represent tasks types. Through dynamic scheduling algorithms workflows can be scheduled and optimal resource usage becomes more probable. An extension of several dynamic scheduling algorithms is proposed in this paper. Section 2 presents a related work and identifies gaps. Section 3 present a brief of WorkflowSim and basic dynamic scheduling policies. Section 4 present proposed extensions of basic dynamic scheduling algorithms. Section 5 presents simulation and analysis on the basis of task types. Finally, concludes in Section 6.

2. RELATED WORK

This section presents a brief review of several research works done in the field of scheduling. Proposals in (Lee et al., 2009; Rahman et al., 2013; Abrishami et al., 2013; Abdelkader & Omara, 2012; Juve et al., 2013; Bittencourt & Madeira, 2011; Ji et al., 2015; Lee, Y. C., 2015; Ghafarian, T., & Javadi, B., 2015; Vasile et al., 2015) proposed scheduling solutions for workflows. Proposals in (Sakellariou & Henan, 2004; Li et al., 2012; Chen et al. 2013; Yu et al., 2008; Ding et al., 2015; Arabnejad et al., 2016; Xu, H., & Yang, B., 2015) refers to the scheduling solution for independent tasks.

Research work in (Juve et al., 2013) presented diversified workflows from sciences like earthquake, gravitational etc. Workflow profiling tools were used to carve out the characteristic of tasks available in workflow on the basis of computational, communication or other requirements. In (Lee et al., 2009) authors described an enhancement to PEGASUS by considering dynamics for allocation decision as compared to existing approach of allocation. The runtime performance of a resource was considered as a qualification for any resource allocation and the results are evaluated using grid middleware over clusters. In (Rahman et al., 2013) authors proposed highly adaptive workflow scheduling by calculating critical path and using this information for task-resource mapping. In (Abrishami et al., 2013) authors presented two phase scheduling which met two objectives simultaneously. Solution was based on Partial Critical Path and addressed makespan of a workflow and respected the timing-deadlines. Authors proposed two algorithms namely IaaS Cloud Partial Critical Paths (IC-PCP) and IaaS Cloud Partial Critical Paths with Deadline Distribution (IC-PCPD2). Both algorithms were based on PCP concept. Earlier approach optimized the makespan characteristics of a workflow, while second algorithm met deadlines simultaneously. Complexity analysis of both algorithms concludes that both algorithms were highly suitable for cloud environment. An extension of HEFT by using task duplication named (Abdelkader & Omara, 2012) Clustering Based HEFT with Duplication (CBHD) for scheduling in generic heterogeneous resource environments was proposed. The CBHD algorithm is considered an amalgamation between the most two important task scheduling in Heterogeneous machine, The Heterogeneous Earliest Finish Time (HEFT) and the Triplet Clustering algorithms. CBHD outperforms the HEFT and Triplet algorithm by decreasing the makespan by 2.5%. It also achieves better load balancing than the HEFT algorithm by 70%, and it increases processors utilization by 10% with respect to the HEFT and Triplet algorithms. In (Bittencourt & Madeira, 2011) the authors presented a Hybrid Cloud Optimized Cost (HCOC) scheduling algorithm which optimizes the resource migration/lease from any public cloud into a private cloud. Authors strived to achieve
A Study and Implementation of a Movie Recommendation System in a Cloud-based Environment
www.igi-global.com/article/a-study-and-implementation-of-a-movie-recommendation-system-in-a-cloud-based-environment/181034?camid=4v1a

Observations on Effect of IPC in GA Based Scheduling on Computational Grid
www.igi-global.com/article/observations-effect-ipc-based-scheduling/62998?camid=4v1a