Accepted Manuscript

Scheduling dynamic workloads in multi-tenant scientific workflow as a service platforms

Maria A. Rodriguez, Rajkumar Buyya

PII: S0167-739X(17)30168-1
DOI: http://dx.doi.org/10.1016/j.future.2017.05.009
Reference: FUTURE 3457

To appear in: *Future Generation Computer Systems*

Received date: 31 January 2017
Revised date: 10 April 2017
Accepted date: 7 May 2017

Please cite this article as: M.A. Rodriguez, R. Buyya, Scheduling dynamic workloads in multi-tenant scientific workflow as a service platforms, *Future Generation Computer Systems* (2017), http://dx.doi.org/10.1016/j.future.2017.05.009

Scheduling Dynamic Workloads in Multi-tenant Scientific Workflow as a Service Platforms

Maria A. Rodriguez, Rajkumar Buyya

Cloud Computing and Distributed Systems (CLOUDS) Laboratory
School of Computing and Information Systems
The University of Melbourne, Australia

Abstract

With the advent of cloud computing and the availability of data collected from increasingly powerful scientific instruments, workflows have become a prevailing mean to achieve significant scientific advances at an increased pace. Emerging Workflow as a Service (WaaS) platforms offer scientists a simple, easily accessible, and cost-effective way of deploying their applications in the cloud at anytime and from anywhere. They are multi-tenant frameworks and are designed to manage the execution of a continuous workload of heterogeneous workflows. To achieve this, they leverage the compute, storage, and network resources offered by Infrastructure as a Service (IaaS) providers. Hence, at any given point in time, a WaaS platform should be capable of efficiently scheduling an arbitrarily large number of workflows with different characteristics and quality of service requirements. As a result, we propose a resource provisioning and scheduling strategy designed specifically for WaaS environments. The algorithm is scalable and dynamic to adapt to changes in the environment and workload. It leverages containers to address resource utilization inefficiencies and aims to minimize the overall cost of leasing the infrastructure resources while meeting the deadline constraint of each individual workflow. To the best of our knowledge, this is the first approach that explicitly addresses VM sharing in the context of WaaS by modeling the use of containers in the resource provisioning and scheduling heuristics. Our simulation results demonstrate its responsiveness to environmental uncertainties, its ability to meet deadlines, and its cost-efficiency when compared to a state-of-the-art algorithm.

Keywords: cloud computing, cost minimization, deadline, workflow as a service, resource provisioning, scheduling

1. Introduction

Workflows are defined by a set of computational tasks with dependencies between them and are a commonly used application model in computational science. They enable the analysis of data in a structured and distributed manner and have been successfully used to make significant scientific advances in various fields such as biology, physics, medicine, and astronomy [1]. Their importance is highlighted in todays big data era as they offer an efficient way of processing and extracting knowledge from the data produced by increasingly powerful tools such as telescopes, particle accelerators, and gravitational wave detectors. Hence, it is common for scientific workflows to be large-scale data and compute intensive applications that are deployed on distributed environments in order to produce results in a reasonable amount of time.

The emergence of cloud computing has brought with it several advantages for the deployment of scientific workflows. In particular, Infrastructure as a Service (IaaS) clouds allow Workflow Management Systems (WMSs) to access a virtually infinite pool of resources that can be acquired, configured, and used as needed and are charged on a pay-per-use basis. IaaS providers offer virtualized compute resources called Virtual Machines (VMs) for lease. They have a predefined CPU, memory, storage, and bandwidth capacity and different resource bundles (i.e., VM types) are available at varying prices. They can be elastically acquired and released and are generally charged per time frame, or billing period. While VMs deliver the compute power, IaaS clouds also offer storage and networking services, providing the necessary infrastructure for the execution of workflow applications.

Scheduling algorithms tailored for scientific workflows are crucial in taking advantage of the benefits offered by clouds and they have been widely studied in recent years. To achieve this, they need not only to focus on the task to resource mapping but also on deciding the number and type of resources to use throughout the execution of the workflow (i.e., resource provisioning). The majority of existing approaches focus on generating resource provisioning and scheduling plans for a single instance of a workflow. They assume application and resource models in which a single user submits a single workflow for execution to a

Email addresses: marodriguez@unimelb.edu.au.com (Maria A. Rodriguez), rbuyya@unimelb.edu.au.com (Rajkumar Buyya)