Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 119 (2017) 112–120

Procedia Computer Science

www.elsevier.com/locate/procedia

6th International Young Scientists Conference in HPC and Simulation, YSC 2017,
1-3 November 2017, Kotka, Finland

Simplifying the Use of Clouds for Scientific Computing with Everest

Sergey Volkov, Oleg Sukhoroslov*

Institute for Information Transmission Problems of the Russian Academy of Sciences,
Bolshoy Karetny per. 19, build.1, Moscow 127051 Russia

Abstract

Cloud computing has emerged as a new paradigm for on-demand access to a wast pool of computing resources that provides an alternative to using on-premises resources. This paper discusses the challenges related to using the cloud computing infrastructures for scientific computing. An approach based on Everest platform addressing these challenges is presented along with the prototype integration of Everest with Google Compute Engine. The proposed integration enables Everest users to seamlessly provision and use cloud-based computing resources for running different types of workloads including HPC and HTC applications. In contrast to other efforts, the presented approach also supports building and sharing domain-specific web services that automate execution of applications on dynamically provisioned cloud resources or hybrid infrastructures.

Peer-review under responsibility of the scientific committee of the 6th International Young Scientist conference in HPC and Simulation

Keywords: cloud computing, high-performance computing, high-throughput computing, platform as a service, Google Compute Engine

1. Introduction

Computational methods are now widely used for solving complex scientific and engineering problems. These methods often require a large amount of computing resources. Traditionally such resources have been provided by dedicated computing clusters, supercomputing centers or distributed grid infrastructures. Recently cloud computing has emerged as a new paradigm for on-demand access to a wast pool of computing resources that provides an promising alternative to traditional approaches based on on-premises resources.

There are several advantages of using clouds for scientific computing. Clouds can significantly lower time-to-solution via quick resource provision, skipping the lengthy process of building a new cluster on-premises or avoiding long queue wait times on shared computing facilities. Clouds can also reduce the total cost of ownership by allowing dynamic scaling of computing resources depending on current workload, or by leveraging cheap spot instances that represent excess cloud capacity. In contrast to on-premises clusters that are often shared across multiple research

* Corresponding author.
E-mail address: sukhoroslov@iitp.ru

1877-0509
Peer-review under responsibility of the scientific committee of the 6th International Young Scientist conference in HPC and Simulation
10.1016/j.procs.2017.11.167

groups with diverse needs, clouds allow each group to have its own dedicated cluster that can be customized to specific applications of this group. This customization allows for better resource efficiency and reduced costs in comparison to the one-size-fits-all approach used with shared clusters.

Despite the described advantages, there are some organizational and technical challenges that hamper the use of clouds for scientific computing. While many large organizations have already invested into on-premises computing infrastructures, small research groups and individual researchers are often more agile and keen to explore the opportunities offered by clouds. However, since the use of clouds requires additional funding, a proper analysis should be performed in order to estimate the required costs and possibly compare them with the use of existing on-premises or community-provided resources [1, 2, 3].

While the mentioned organizational issues are important, in this paper we will focus on the technical challenges related to the use of clouds for scientific computing. The biggest challenge is related to the deployment, configuration and management of required software stack including the system software, job schedulers and end-user applications. While the on-premises infrastructures are managed by skilled IT staff that perform these activities for the users, the Infrastructure-as-a-Service (IaaS) cloud services provide only a bare-bones VM-based infrastructure. Another challenge is related to the selection of optimal hardware configuration and resource usage strategy for particular application in order to reduce cost and increase performance. While tightly-coupled high-performance computing (HPC) applications require the use of dedicated clusters with low inter-node latency, the embarrassingly parallel high-throughput computing (HTC) workloads can be run on a dynamically-sized pool of cheap spot instances. There are other challenges such as data movement and access, security and software license management.

Despite the abundance of cloud computing providers, the inherent complexity of these infrastructures and the mentioned technical challenges, along with the lack of required IT expertise among the researchers, limit the wide adoption of clouds for scientific computing. This problem can be solved by providing high-level services with domain-specific interfaces that hide the mentioned complexity from the user. These services should automate all common actions needed to allocate resources in the cloud, deploy and configure the required software and perform computations. The use of service-oriented approach can also improve the research productivity by enabling publication, sharing and composition of computing applications as services.

Everest [4, 5] is a web-based distributed computing platform that implements the described approach. The platform supports publication, execution and composition of computing applications in a distributed environment. Unlike other solutions, Everest is based on the Platform as a Service (PaaS) cloud computing model by providing its functionality via remote web and programming interfaces. A distinctive feature of Everest is the support for running applications on arbitrary combinations of computing resources attached by users. Currently the platform supports integration with standalone servers, computing clusters and European Grid Infrastructure [6].

This paper presents the prototype integration of Everest platform with cloud computing infrastructures on the example of Google Compute Engine (GCE). The proposed integration has the following benefits. First, it enables Everest users to seamlessly access computing resources of GCE and combine them with other types of resources already supported by the platform. Second, it makes it possible to use Everest for building generic or domain-specific web services for submission and automation of computations in the cloud. Hopefully, the proposed approach will make cloud computing services accessible to a wider group of researchers.

The paper is structured as follows. Section 2 discusses related work. Section 3 provides a brief overview and relevant technical details on Everest and GCE. Section 4 discusses the possible approaches for integration of Everest and GCE, presents the current prototype implementation and describes its experimental evaluation. Section 5 concludes and discusses future work.

2. Related Work

There exist several tools for deploying virtual clusters for scientific computing on resources provisioned from cloud infrastructures. StarCluster [7] is a toolkit for creating and managing distributed computing clusters hosted on Amazon's Elastic Compute Cloud (EC2). It supports dynamically adding and removing nodes to and from the cluster depending on the number of jobs queued up. The auto-scaling functionality requires a user to run load balancer that periodically polls the cluster. ElastiCluster [8] can be used to create, setup and manage computing clusters on several public cloud providers or a private OpenStack cloud. It relies on Ansible to describe and automate the steps required