Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 201 (2017) 534–542

Procedia Engineering

www.elsevier.com/locate/procedia

3rd International Conference "Information Technology and Nanotechnology", ITNT-2017, 25-27 April 2017, Samara, Russia

Job flow management for virtualized resources of heterogeneous distributed computing environment

I. Bychkov[a], A. Feoktistov[a]*, I. Sidorov[a], R. Kostromin[a]

[a]*Matrosov Institute for System Dynamics and Control Theory of SB RAS, 134, Lermontov st., Irkutsk, 664033, Russia*

Abstract

We address to the problem of an integration of heterogeneous computational clusters to the united environment applying the OpenStack platform for a cluster resource virtualization. However, this platform does not interact with the traditional resource management systems such as the PBS Torque or SLURM that are used in cluster resources. To this end, we developed an additional hypervisor shell to run virtual machines through queues of the aforementioned systems. Thus, we expand the OpenStack capabilities for an application job management and provide a computation scalability for virtualized resources of clusters. In order to show the benefits of our approach to the scalable application management in integrated cluster environments, we developed a parameter sweep application for simulation modeling of warehouse logistics and solved three optimization tasks for the real refrigerated warehouse. We compared the selected criteria of task solving processes with the hypervisor shell applying and without it by means of simulation modeling. Experiments show that the hypervisor shell can significantly improve computation speedup and task solving time through reallocating virtual machines to queues of the resource management systems.

Peer-review under responsibility of the scientific committee of the 3rd International Conference "Information Technology and Nanotechnology".

Keywords: heterogeneous distributed computing environment; HPC-clusters; OpenStack

1. Introduction

Nowadays, an effective solution of large-scale computational problems in various subject domains demands the use of various high-performance computing (HPC) systems. Computational clusters represent the essential part of

* Corresponding author. Tel.: +7-924-711-6704; fax: +7-395-251-1616.
E-mail address: agf65@yandex.ru

1877-7058
Peer-review under responsibility of the scientific committee of the 3rd International Conference "Information Technology and Nanotechnology".
10.1016/j.proeng.2017.09.651

such systems. The rapid development of HPC leads to the large diversity of their hardware and software platforms, architectures and interconnects. Thus, clusters differ their software and hardware characteristics, and hence computational capabilities.

An integration of clusters within a heterogeneous distributed computing environment has the goal of expanding their total capabilities for solving the wide spectrum of scientific and applied computational problems. It is an important approach to researches in the field of parallel and distributed computing. In such approach, clusters are the main components of the environment.

To solve a computational problem, a user should form a job for a resource management system (RMS) used in a cluster. The job is a specification of solving problem that include information on the required computing resources, executable applied programs, inputs and outputs, and other required data.

When cluster resources are virtualized, they are managed by hypervisors. RMSs can be included into virtual machine images. However, neither hypervisors nor hypervisor middleware interact with RMSs used in non-virtualized resources.

In this paper, we present an approach to creating an integrated cluster environment (ICE) that includes virtualized resources. It based on the OpenStack platform usage jointly with the specialized hypervisor shell allowing to start virtual machines through queues of the traditional RMSs such as the PBS Torque [1] or SLURM [2] in non-virtualized resources.

We developed parameter sweep application for solving three optimization tasks for real refrigerator warehouse. We also provided practical and simulation experiments that illustrate utility of the developed application and benefits of the proposed hypervisor shell for the application job management jointly with the OpenStack platform.

2. Related work

The research field related to the problem of integrating heterogeneous resources of centers of collective usage to a distributed computing environment for large-scale scientific studies is sufficiently novel. An analysis of a research state in this field shows that at present there is no unified approach for solving this problem. Various tools of the computing environment virtualization, resource management systems and web-based platforms are used for the resource integration [3-5]. These tools have a set of advantages and drawbacks. Usually, a common drawback is their narrow specialization of a job management in the integrated environment.

The important component of the integration technology is an implementation of the ability to take into account a subject domain specifics of the problems [6]. An increase of a computations management efficiency can be provided by a multi-agent technology use [7-9]. A functioning efficiency of such systems is enhanced by the use of elements of economic theory [10].

The developed multi-agent systems are often highly specialized and do not provide an interaction with other systems for the computation management based on the traditional models that used in the workflow management [11], parameter sweep applications [12], resource management systems [13] and meta-scheduling through a middleware [14].

However, the common problems for the multi-agent and listed above traditional models are the following [15]:

- Variety of software and hardware platforms, architectures and interconnects that are used in nodes of heterogeneous distributed computing environments,
- Ensuring the reliability and safety of computational processes in nodes of such environments,
- Accounting both the subject domain specifics of the problems and features of heterogeneous environment resources,
- Providing the multi-level parallelism in a problem solving process,
- Organization of the interdisciplinary research.

In this context, virtualization technologies of distributed computing environments are actively developed to partially solve the above-listed problems [16]. Nevertheless, an integration of traditional models of the computations management with models of the virtualized resources management is still an important and non-trivial problem.

Virtual machines and containers underlie two major approaches to cluster resources virtualization.