Available online at www.sciencedirect.com

ScienceDirect

IERI Procedia 10 (2014) 169 – 176

Procedia IERI

www.elsevier.com/locate/procedia

2014 International Conference on Future Information Engineering

Improving HPC Application Performance in Public Cloud

Rashid Hassani*, Md Aiatullah, Peter Luksch

University of Rostock, Rostock, 18059, Germany

Abstract

Improving as well as evaluating the performance of High Performance Computing (HPC) applications by migrating them to Cloud environments are widely considered as critical issues in the field of high performance and Cloud computing. However, poor network performance, heterogeneous and dynamic environments are some series of pitfalls for execution of HPC applications in Cloud. This paper proposes a new approach to improve the performance and scalability of HPC applications on Amazon's HPC Cloud. The evidence from our approach points a significant improvement in speed up and scale up with the response rate of more than 20 percent parallel efficiency on the Cloud in comparison to dedicated HPC cluster. We state that the EC2 Cloud system is a feasible platform for deploying on-demand, small sized HPC applications.

Selection and peer review under responsibility of Information Engineering Research Institute

Keywords: Cloud computing; HPC applications; parallel sorting algorithm; MPI; Amazon EC2

1. Introduction

Supercomputing, often called as High Performance Computing (HPC), is referred to as an attribute that employs parallel processing to perform huge computations in short time. HPC platforms are typically tightly-coupled and perform frequent inter-processor communication and synchronization. They require a cluster setup with massive number of computers which are expensive to install, maintain, and operate. They are mostly being used for scientific research in academia and industries. Therefore, supercomputers cannot be a

* Rashid Hassani. Tel.: +49-3814987565; fax: +49-3814987522.
E-mail address: rashid.hassani@uni-rostock.de.

Selection and peer review under responsibility of Information Engineering Research Institute
doi:10.1016/j.ieri.2014.09.072

feasible solution for the normal users and small businesses who intend to have on-demand access or to run their applications for a short period.

Cloud computing is emerging as a commodious resource of computational power in recent years which provides several possibilities to build HPC platform [1]. Traditional HPC platforms offer limited hardware access and fail to scale. The ability to scale up and down the computing platform according to the application requirements and users' budget, makes the Cloud a cost-effective, timely solution and emerging trend to the needs of HPC users. Some of the key features characterizing Cloud computing are virtualization, elasticity of resources and rapid growing with ability of delivering both infrastructure and software as a services. These features allow better flexibility and customization to specific application of HPC users. Therefore, HPC community has discovered Cloud computing facilities as a potential target system. This is one of the main reason motivating many users and organizations to port HPC applications to Cloud.

Elastic Compute Cloud (EC2) [2], introduced by Amazon Web Services (AWS), provides powerful compute and storage resources on demand through hardware level virtualization and also aims to be global. It provides the possibility of computing on virtual parallel clusters. Some studies investigated the benefits of performing HPC applications on the Amazon Cloud infrastructure. Despite the benefits offered by Cloud computing, it has not yet been established whether Cloud can offer a suitable alternative to supercomputers for HPC applications. Therefore, this motivated us to carry out a detail studies on HPC in the Cloud.

The results from past research of HPC applications on Cloud by focusing on performance as the metric have been pessimistic. They outlined major limitations which have been resulted from insufficient network performance, resource heterogeneity and multi-tenancy for HPC applications on Cloud [3,4,5,10,11]. However, Cloud advanced in recent years and now is going to solve some of these problems by leading to heterogeneous configurations in processors, memory, and network. They have already been adopted in different scenarios such as intensive and business applications in which, the user can scale up and down resources when required and finally drop them out when the task is done.

Sorting algorithms are considered as a core part used in HPC applications. There are many types of distributed sorting algorithms and many improvements have been done on them. Speed up is the performance attribute to analyze the parallel algorithms. Some of these sorting algorithms have been implemented on various computing infrastructures to analyze the performance of the system. It would be interesting to see on how the Cloud scales and what is the performance in Cloud vs. high-end machines when the parallel sorting algorithm is implemented. Despite the wide benefits the Cloud offers, the research question currently is "whether the Cloud is a feasible platform for parallel sorting HPC applications".

In this paper, we have implemented the MPI version of parallel Radix sort and analyzed its performance on Cloud infrastructure and finally compared it with dedicated high-end HPC platform.

The contributions of this work are the following:

- Investigating most recent works on HPC applications which have been ported to the Cloud environment in various fields and also identifying the challenges faced by these applications in Cloud environment.
- Implementing an efficient MPI version of parallel Radix sort algorithm to obtain scalability and good speed up in Cloud.
- Evaluating the performance and scalability of our proposed technique by deploying it both on a real dedicated HPC testbed (Sirius) [6] and on Amazon EC2 Cloud [2].

The rest of the paper is organized as follows:

Section II gives a brief overview of Cloud computing and also Amazon EC2 infrastructure by defining the reference model of this paradigm and discusses the potential opportunities and investigates the current state of the art of high performance computing in public Cloud specially Amazon EC2. Section III describes the implementation of our proposed parallel version of Radix sort and outlines the results of the scaling analysis