Performance Evaluation of MPI Parallel Transfer in Microsoft Azure Cloud

Eduardo Roloff, Emmanuell Diaz Carreo, Jimmy Valverde Sanchez, Philippe Navaux
Federal University of Rio Grande do Sul - UFRGS
Informatics Institut
Porto Alegre - RS - Brazil
{eroloff,edcarreno,jkmvsanchez,navaux}@inf.ufrgs.br

Abstract

Since the firsts appearances of the cloud computing paradigm, it is attracting attention of the High-Performance Computing (HPC) community. Due to its characteristic of elasticity and unlimited amount of resources, combined with the pay-per-use charge model. These key concepts makes T̈he Cloudän interesting alternative as an environment for HPC applications. This work presents a first step of a project were we will investigate the capabilities of cloud computing for HPC. In this work we executed the MPI Exchange Benchmark among four different Azure Data Centers to study the behavior of the network in the same provider. Our results shown that, compared with a traditional HPC cluster, the cloud presents a 10 times performance loss, but with a predictable behavior.

1. Introduction

Cloud Computing has two main characteristics: the pay-per-use cost model and the elasticity [1]. These characteristics offer an interesting alternative for High Performance Computing (HPC) applications. Due to them, it is possible to have access to, virtually, any amount of resources in little time without upfront costs.

This work is one of the firsts steps towards a project that aims to explore opportunities for symbiotic, coordinated use of HPC and cloud computing infrastructures. The idea is, starting from an HPC infrastructure such as a traditional HPC cluster. The purpose is to identify cloud computing aspects that could be explored to provide a far-reaching, flexible (e.g., in terms of adaptation to fluctuating demands) and efficient environment for running HPC.

It is a common sense that the main bottleneck of Cloud Computing is the network performance, a very important aspect for HPC. Since MPI is an important standard for HPC communication [2] its performance need to be measured.

In this paper, we provide a evaluation of MPI Parallel Communication in the Microsoft Azure public Cloud.

We compared the same type of virtual machine (VM) instance among four different Azure Data Centers and used the machines during working hours and during the night.

Our intention was to verify the allocation time presents performance impact in the machines and if they have different performance levels among different Data Centers using the same instance. To have a baseline for comparison purposes we used the traditional cluster *econome* from the GRID5000 project.

Our results shown that the execution during the day and night has minimum performance difference and could be ignored. We also conclude that there a slightly difference in the network performance when compared the execution times between the different Data Centers.

This work is organized as follow. Section 2 presents the methodology used to conduct our work. In the Section 3 we present the results of the execution of the network benchmark among the Azure Data Centers. Finally, Section 4 presents a discussion of the work and talks about the future work.

2. Methodology

We performed experiments on one traditional cluster system as well as four Data Center locations of Microsoft Azure using the G5 VM instance. The G5 instance is a VM with 32 cores, composed of a E5-2698v3 CPU running at 2.3 GHz with 448 GB of RAM, there is no precise information about the network interconnection. The traditional cluster is the *econome* machine from *GRID 5000* and is composed of two 8-core processors, the network interconnection is 10 Gbit Ethernet.

In all environments, we create systems with 128 cores in total to maintain a comparable baseline. The total number of nodes were four, for the G5 machines, and 8 for the econome cluster. The locations of Microsoft Azure used were: West Europe (WEU), West USA (WUS), East USA

Machine name	Processor model	Cores per instance	Network	Location
Econome	E5-2660	16	10 Gbit/s	France
G5	E5-2698 v3	32	—	West EU East US Singapore West US

Table 1. Configuration of the environments used in the experiments.

(EUS) and Southeast Asia (SAS). To the best of our knowledge, all systems are running without Hyper-Threading. All environments use Intel processors of recent generations, at least the Sandy-Bridge family.

Table 1 contains an overview of the machines used in the evaluation. Although main memory sizes vary between different instance sizes, all amounts were sufficient for our experiments and are therefore not mentioned in the table.

We use the Intel MPI Benchmark Exchange test. According with the Intel MPI Benchmark manual *"Exchange is a communication pattern that often occurs in grid splitting algorithms (boundary exchanges). The group of processes is similar to a periodic chain, and each process exchanges data with both left and right neighbor in the chain."* This test measures both the bandwidth and the latency of the network.

We executed the tests five time on each environment and allocation. Each one of the experiments was performed with different message sizes, 0, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, and 32768 bytes.

3. Results

Figure 1 shows the bandwidth results of the Exchange test in logarithmic scale. The cloud instances present the same behavior pattern of increasing the bandwidth when the package size increases. We could observe that we have a very predictable bandwidth capacity according with the increase of the package size.

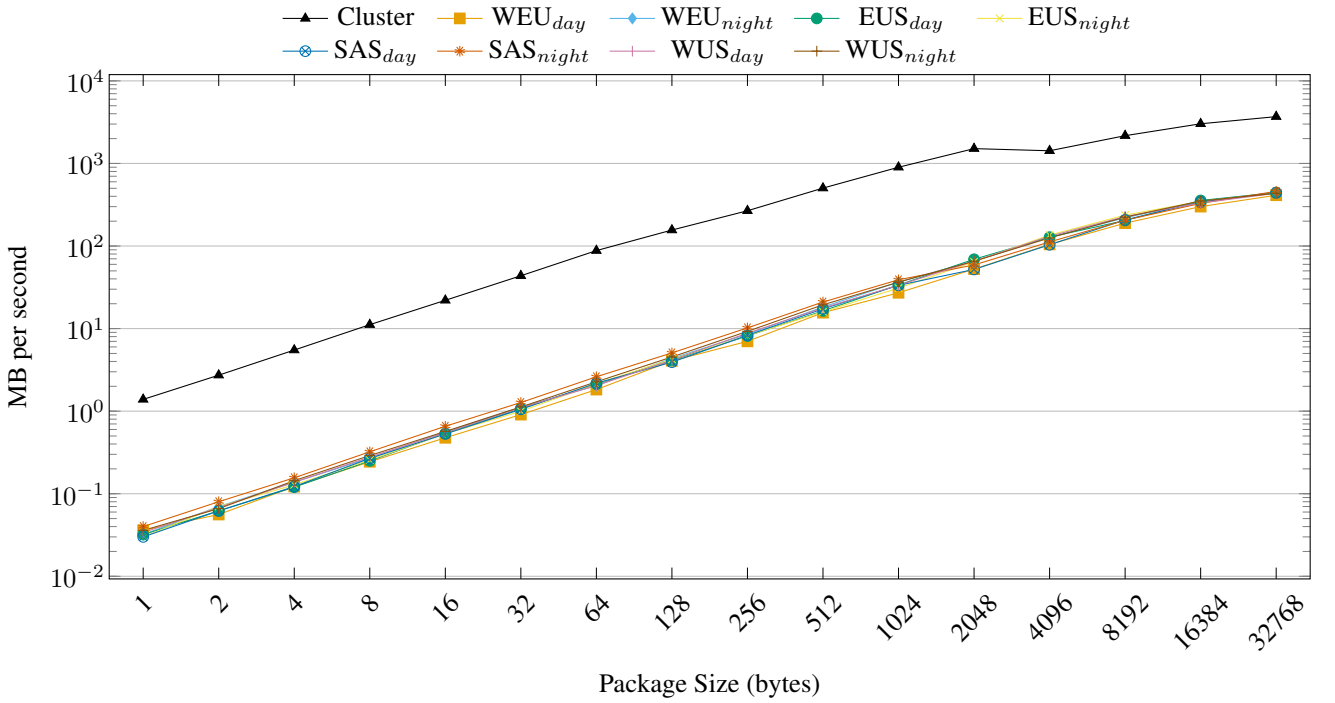

Figure 1. Bandwidth Results for Exchange benchmark.

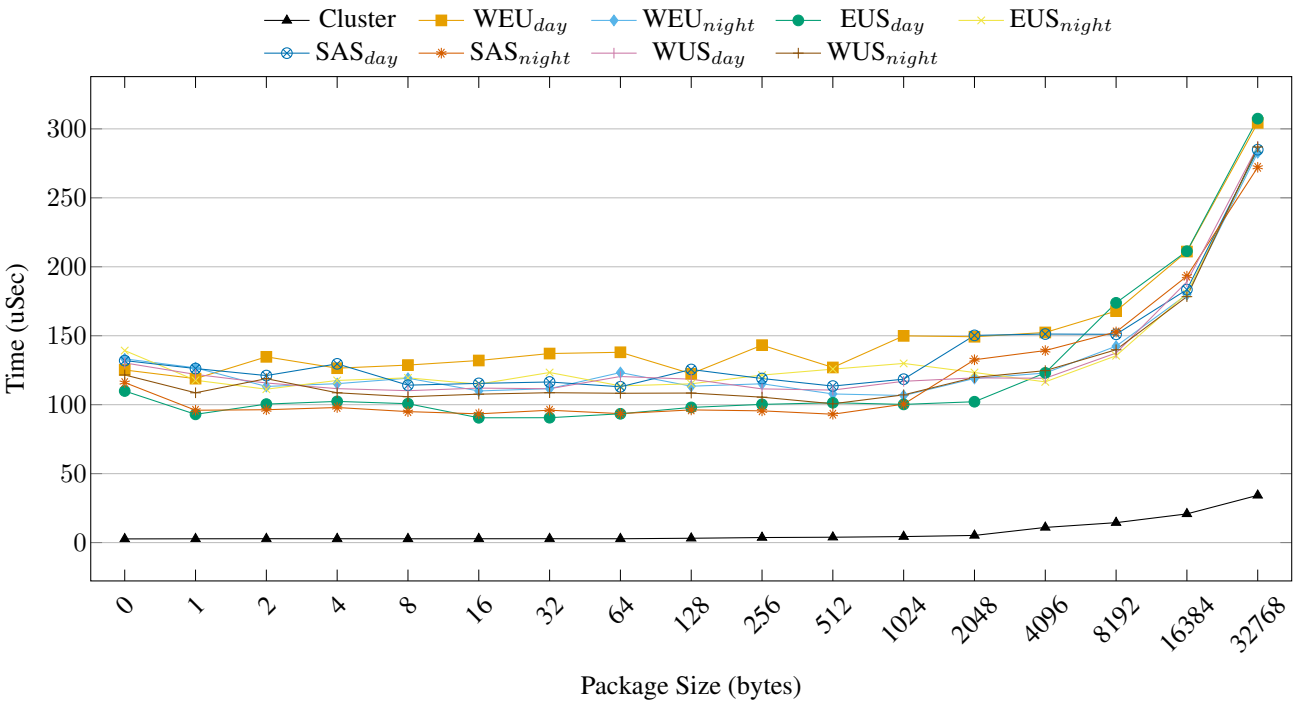

Figure 2. Latency Results for Exchange benchmark.

The bandwidth achieved by the cluster was 3 Gb/sec when the cloud allocations were around 350 Mb/sec for a 32 KB package. Comparing these numbers, we could observe that the cluster achieved a bandwidth around 10 times better than the cloud allocations. This indicates that the cloud network presents less capacitate at all as possible a certain level of contention, due to the virtualized and shared environment. However, the predictable behavior and the bandwidth increase of the cloud allocations could benefit the user when configuring his application to be executed in the cloud.

Figure 2 shows the latency results for the Exchange test. It is possible to verify that the cluster latency has a increase when the package size increases. This behavior could mean a level of network contention, the reason could be that this test performs a lot more concurrent communication in the network.

The cloud results showed a decent latency and a more predictable behavior, all the cloud instance allocations displayed the same pattern. It is interesting to note that all of the cloud allocations present a huge increase when the package size turns 8KB, and then all of them follow the standard pattern. This could be explained for a possible network optimization. The network is configured to handle a certain number of bytes at same time, for optimization, and when this number is reached the switches need to go to the controller to get a new configuration. This took some time, then the latency increases a little.

4. Conclusions and Future Work

In this works we performed an evaluation of the Exchange benchmark in the Microsoft Azure cloud. We shown that the cloud has a performance loss compared with a physical cluster and it is not ready for any HPC application. However for some kinds of applications, with a well defined communication pattern, it is suitable to be used.

As future work, we will expand this tests and provide a deep investigation of the cloud bottlenecks.

Acknowledgments.

This research received funding from the EU H2020 Programme and from MCTI/RNP-Brazil under the HPC4E project, grant agreement no. 689772. Experiments presented in this paper were carried out using the Grid'5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see https://www.grid5000.fr). Additional funding was provided by CAPES and Microsoft.

References

[1] R. d. R. Righi, V. F. Rodrigues, C. A. da Costa, G. Galante, L. C. E. de Bona, and T. Ferreto. Autoelastic: Automatic resource elasticity for high performance applications in the cloud. *IEEE Transactions on Cloud Computing*, 4(1):6–19, Jan 2016.

[2] J. A. Zounmevo, D. Kimpe, R. Ross, and A. Afsahi. Using mpi in high-performance computing services. In *Proceedings of the 20th European MPI Users' Group Meeting*, EuroMPI '13, pages 43–48, New York, NY, USA, 2013. ACM.