Performance Analysis of a Numerical Weather Prediction Application in Microsoft Azure

Emmanuell D. Carreño, Eduardo Roloff, Jimmy V. Sanchez, and Philippe O. A. Navaux

Federal University of Rio Grande do Sul, Informatics Institutemendmend
Porto Alegre, Brazil
{edcarreno, eroloff, jkmvsanchez, navaux}@inf.ufrgs.br

Abstract

The Cloud Computing paradigm emerged as a practical solution to perform large-scale scientific computing. The elasticity of the Cloud and its pay-as-you-go model presents an attractive opportunity for applications commonly executed in clusters or supercomputers. In this context, the user does not need to make an upfront purchase of infrastructure. The resources can be rented from a provider to be used for a period of time. In this paper, we present the performance results of executing and scaling horizontally a numerical weather prediction application in a Cloud infrastructure. Results show that scalability and performance are acceptable and that the Cloud Computing paradigm represents an attractive environment for high performance applications.

1. Introduction

Computer generated weather forecast is a continuous developing area that has been executed mostly on grids or supercomputers. This is because applications in climate and weather research typically require huge amounts of data and computing power [1]. Numerical Weather Prediction (NWP) has been benefited by the computing area for years, the use of computers has allowed them to use more variables in their simulations and more data sources to perform precise predictions. This computing approach has achieved much, but can also be a significant limitation to progress [2].

The costs associated with hardware investment and the need for energy efficiency have motivated research in the usage of new technologies to optimize costs; Cloud Computing is one of them. Cloud Computing provides cost-effective, fast, and a vast amount of resources for large-scale applications. The Cloud is used as utility where computing, storage, and networking services are provided on-demand.

On-demand Cloud Computing offers an attractive new dimension to High Performance Computing (HPC), in which virtualized resources can be appropriated. These resources could be used in a customized manner, to target a particular scenario, at the time and in the form the user desires.

Some HPC initiatives have started utilizing Cloud Computing for research. However, it is clear that one of the major concerns for the HPC community is the performance [3]. For weather research, the need to reduce the waiting time found in the batch jobs schedulers common in HPC infrastructure, the possibility of quickly access and share the processed data, and the development of newer simulations or ensemble forecasts has motivated them to use the Cloud.

2. Numerical Weather Prediction

Numerical Weather Prediction (NWP), is the use of mathematical models to predict, accurately, the atmospheric behavior for an specific period. It represents the state of the atmosphere at a fixed time (temperature, wind components, etc.) over a discrete domain [1]. The results are denominated weather forecast when predicts only a few days and climate forecast when predicts months.

BRAMS was the NWP evaluated in this work. BRAMS is one of the most widely employed numerical models in regional weather centers in Brazil [4]. It is a model based on RAMS [5]. RAMS, developed by the atmospheric science department at the Colorado State University, is a numerical weather prediction model used to simulate atmospheric conditions. The new developments integrated into BRAMS include modified computational modules to better simulate the tropical atmosphere. The primary objective of BRAMS is to provide a single model for Brazilian regional weather centers.

3. Microsoft Azure

Microsoft started its initiative in Cloud Computing with the release of Windows Azure at its annual professional developer conference in 2008. The initial model was the platform as a service allowing developing and running applications written in the programming languages supported by the .NET framework. Nowadays, the Azure infrastructure covers all types of service models and has gained a significant market share. The service prices vary from location to location. Because of this factor, selecting the best price for each application can reduce or increase costs significantly.

We used the Infrastructure as a Service (IaaS) model, one of the Cloud service models. This model provide us complete management capabilities over the Virtual Machines (VMs) running on Microsoft's infrastructure. Azure offers multiple options regarding IaaS [6]. Each Virtual Machine configuration is called *instance size* using Azure's conventions. Some of the instances sizes are very interesting for HPC, especially the ones with Infiniband. Unfortunately, Infiniband was available only for Windows virtual machines using a proprietary MPI driver. We selected only virtual machines using Ubuntu Linux distribution, therefore we did not used any Infiniband for our experiments.

Microsoft Azure was chosen because previous work [7] has shown that performance and cost efficiency for HPC in Azure was better than on Rackspace and Amazon.

4. Methodology

The tests were performed by scaling the number of nodes available to the application. The experiments consisted of executing BRAMS to perform a forecast of 72 hours with a spatial resolutions of 20 km and 5 km covering a an area of 1500×1500 km. Each forecast simulation was performed five times using virtual machine instances with eight CPU cores. Table 1 contains the specifications of the VM instances used on the experiments. At every iteration, we added a new instance up to eight nodes (168 CPU cores).

Characteristic	Cloud Instance Size			
	A4	A7	A8	A9
Processor Model	Xeon E5-2660		Xeon E5-2670	
Processor Speed (GHz)	2.2		2.6	
Number of CPU Cores	8	8	8	16
Memory (GB)	14	56	56	112
Networking (Mbps)	800	2000	5000	10000

Table 1. Specifications of the Cloud instances used for the performance experiments.

All the experiments were executed in the same Azure datacenter location using Affinity Groups. Using this feature we could achieve better networking performance as the VM would be as physically close to each other as possible. The datacenter we selected was Europe North.

Scalability is used to establish the possibility of an application to use more resources in case it needs to use them. This could be seen on two different axes, horizontal or vertical scaling. In the Cloud, horizontal scaling is referred to the possibility of adding more virtual machines for processing. The adding process could reach a limit in which adding more machines causes only the increasing in experiment costs without increasing the performance.

Vertical scaling in the Cloud is used commonly to upgrade the characteristics of the instance running a service. Usually, this process occurs on-demand and over a running instance. BRAMS uses homogeneity to balance resource us-

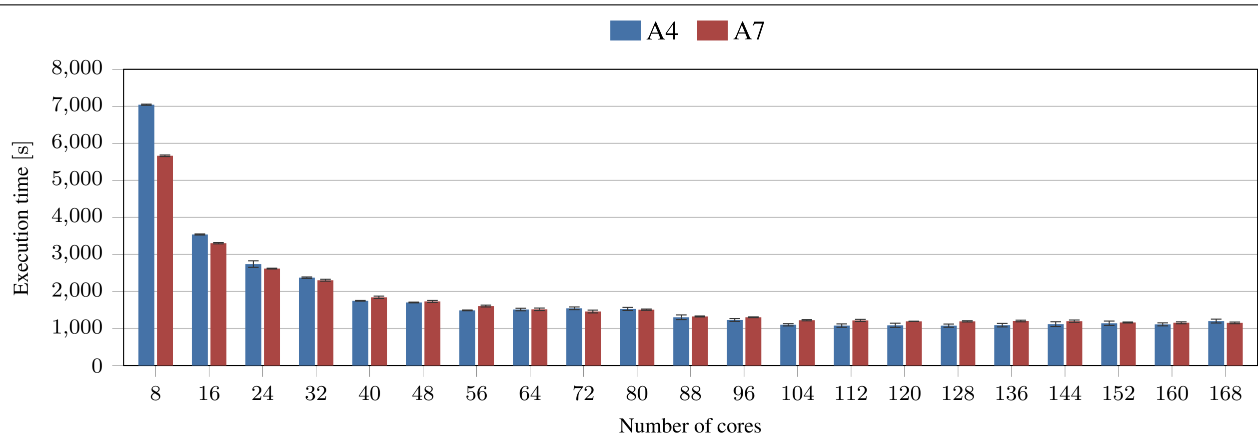

Figure 1. Scaling experiments with 168 cores using a 20 km model resolution with A4 and A7 VM instance sizes. Error bar indicates standard deviation.

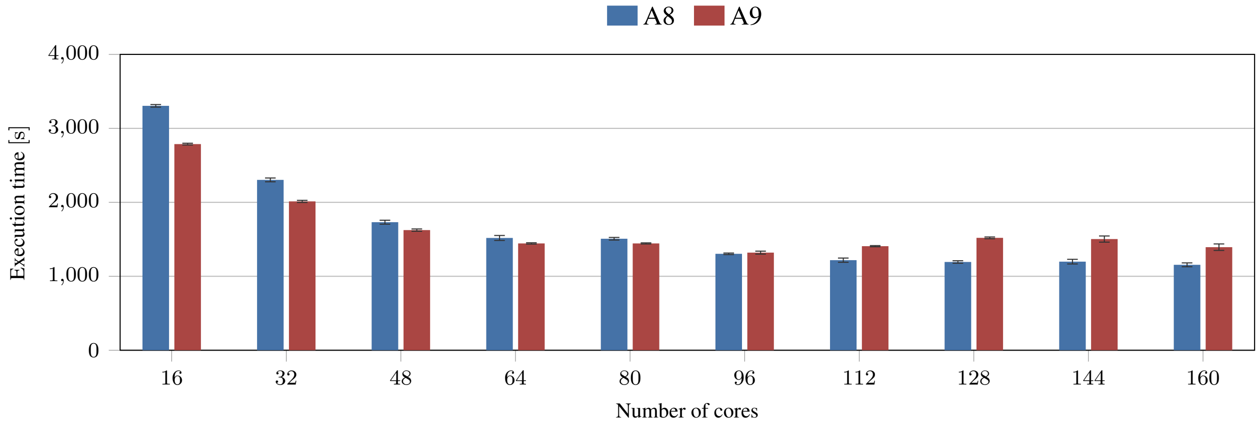

Figure 2. Experiments with 160 cores and a 5 km model resolution using A8 and A9 VM instance sizes. Error bar indicates standard deviation.

age. Due to this behavior is not wise to upgrade the instance characteristics in the middle of an experiment while running the application. Our vertical scaling experiments were based on using Cloud instances with different characteristics, but maintaining the same configuration for all the machines during each experiment. Using this approach we could measure in which point was not possible to get more performance gains scaling BRAMS in the Cloud.

5. Results

We performed the first scaling experiments using standard A4 and A7 instance sizes. As depicted in Figure 1 the application was not able to scale well beyond 88 cores. Even using that amount of cores is not worth, over 56 cores the time reduction is less than two percent at each scaling step.

Generally speaking the performance gets worst as more nodes are added to the experiment for both instance sizes. This can be attributed to the fact that BRAMS is an application with a high amount of communication between cores and nodes. Other factors influencing the scalability could be related to network latencies or access to the Cloud storage implemented while migrating [8] BRAMS to Azure.

We evaluated if the reason why BRAMS did not scale was because the instances used were not the ones with the best performance. We performed another pair of experiments using the instance sizes with the best networking capabilities.

The results are shown on Figure 2. This experiment was performed using A8 and A9 instances; the A8 instance supports eight virtual cores with 56GB of memory while the A9 instance supports 16 virtual cores with 112GB of memory. Due to the difference in the number of cores, we run the forecasts only using 16 cores at a time for the A9 instance or pairs of A8 instances.

In this case, BRAMS behaves a little different but also confirms that scalability was limited. The application scaled correctly to near 80 cores. After that point, performance worsened with each node added using A9 instances. In the case of A8 instances, after 88 cores performance gains were low.

It is interesting to notice that an A8 instance is similar to an A9 in most of the characteristics but the number of cores. However, A8's performance results and scaling were better than A9's. Besides, A9's price is twice the price of an A8. The results show that in the case of BRAMS a virtual machine twice as expensive is not better in the same magnitude.

Regarding the experimental results in this section, we conclude that there is a very specific point in which is better to use fewer instances to execute an application in the Cloud. That point would depend on the characteristics of the application. In some cases, running the application on instances with lower performance characteristics can be a better option than scaling. This is true when the performance benefits are so minimal that running the application would mean wasting the money.

6. Conclusion

Performance analysis show that Cloud Computing infrastructure offers possibilities to migrate applications from legacy systems and that it is possible to provide an environment for the application while reducing the setup complexity.

Performance variability between experiments was low in Azure, which is an important characteristic for HPC as it leads to a higher predictability of experiments, both in the time it takes and how much it will cost.

The scalability experiments show that it is possible to scale BRAMS up to 168 CPU cores with certain mistrust. Our results show that beyond 88 CPU cores the reduction in execution time is less than two percent at every scaling step.

In the future a comparison with a physical cluster could be performed to improve the soundness of the results. Also, we are going to test performance and scalability using Infiniband equipped instances.

References

[1] R. Souto, R. Avila, P. Navaux, M. Py, N. Maillard, T. Diverio, H. Velho, S. Stephany, A. Preto, J. Panetta, E. Rodrigues, and E. Almeida, "Processing mesoscale climatology in a grid environment," in *Cluster Computing and the Grid, 2007. CCGRID 2007. Seventh IEEE International Symposium on*, pp. 363–370, May 2007.

[2] C. Evangelinos and C. N. Hill, "Cloud Computing for parallel Scientific HPC Applications: Feasibility of Running Coupled Atmosphere-Ocean Climate Models on Amazon's EC2," in *In The 1st Workshop on Cloud Computing and its Applications (CCA*, 2008.

[3] S. Benedict, "Performance issues and performance analysis tools for hpc cloud applications: a survey," *Computing*, vol. 95, no. 2, pp. 89–108, 2013.

[4] CPTEC-INPE, "Brazilian Regional Atmospheric Modelling System (BRAMS)." `http://www.cptec.inpe.br/brams`, 2014. Accessed August 10, 2014.

[5] R. Pielke, W. Cotton, R. Walko, C. Tremback, W. Lyons, L. Grasso, M. Nicholls, M. Moran, D. Wesley, T. Lee, and J. Copeland, "A comprehensive meteorological modeling system - RAMS," *Meteorology and Atmospheric Physics*, vol. 49, no. 1-4, pp. 69–91, 1992.

[6] Microsoft, "Sizes for virtual machines." `https://azure.microsoft.com/en-us/documentation/articles/virtual-machines-size-specs/`. Accessed July 1, 2015.

[7] E. Roloff, F. Birck, M. Diener, A. Carissimi, and P. Navaux, "Evaluating high performance computing on the windows azure platform," in *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*, pp. 803–810, June 2012.

[8] E. D. Carreño, E. Roloff, and P. O. A. Navaux, "Challenges and Solutions in Executing Numerical Weather Prediction in a Cloud Infrastructure.," in *International Conference on Computational Science*, (Reykjavík, Iceland), p. 6, 2015.