An analysis of the performance and benefit of execution of the Open Porous Media (OPM) Reservoir Simulator in the Cloud

Agnaldo Silva Lima[1] and Edson Borin[2]

***Abstract*— The Interest in use the Cloud Computing has been growing lately, and as a result, the power demand used by Cloud providers with the Azure is growing. In view of this problem and the interest of the scientific community in scaling scientific tools in the Cloud, we have studied the behavior of a Reservoir Simulator running its processes in the Cloud. In this study, we aim to establish the best way of scaling Reservoir Simulator by using resource efficiently. We do a performance analysis to determine which approach to follow and lastly do a financial and energy consumption analysis. The experiments show that Cloud can be considered as a viable alternative economically and energetically to the traditional machines of High-Performance Computing for the execution of the Reservoir Simulators.**

I. INTRODUCTION

The use of Cloud Computing (CC) for High-Performance Computing (HPC) processes has grown lately. This is due to ease of use, improvements in computer network infrastructure, and increased competition among service providers such as Amazon Web Services (AWS) and Microsoft Azure. According to estimates, it is expected that by 2020 92% of workloads will be executed in Cloud environments [1][2]. By the definition of Vaquero et al. (2008) Cloud Computing is a set of easily usable and accessible virtual resources. These features can be dynamically configured to fit a variable load, allowing for optimized use of resources. This set of features is typically exploited through the pay-for-use model with guarantees offered by the provider through service level agreements [3][4].

High-Performance Computing can be approached as a practice that seeks to solve major problems using vast coordinated computational resources to obtain the best results in the shortest possible time [9]. Searching for a profile for High-Performance Computing evolution, Netto et al. [7] argue that the use of HPC environments became popular in the 1990s thanks to the use of clusters of computers since they were cheap alternatives to traditional Supercomputers. Moreover, by the different ways of using cluster around the world and the search for higher performance, the Grid of Computer was able to provide the necessary expansions and has become quite relevant in the scientific community.

In view of the above, CC has become an alternative to traditional HPC machines, and increasingly attractive from the economic point of view given its form of economies of scale [5][10]. The CC brings the elasticity, with which users can increase or decrease the use of resources. In addition, the user only pays for what is actually used, which provides a great reduction of costs when compared to the traditional form of HPC [5].

However, as a result of the great growth of CC, energy consumption has been increasing. Studies show that CC services around the world consume about 8% of the energy produced [5]. Given the current global climate change caused by Global Warming, reducing energy consumption and resource efficiency is of paramount importance. In order to make the great growth of the Cloud become sustainable from the energy point of view, efforts have been made by the scientific community. Therefore, there is a need to look for new ways to scale scientific applications in the Cloud of commercial providers. In order to contribute to these efforts, we study a Reservoir Simulator (SR) in the Cloud for cost-benefit and energy performance.

The SR is an important scientific tool of the Petroleum Industry that needs environments of HPC; through the use of SR, it is possible to predict the behaviour of fluids in real reservoirs using mathematical models [2]. There are increasing challenges in this area; for example, the depletion of oil reserves, the search for CO2 storage areas, the conscious use of aquifers, given the growing scarcity of drinking water, and, one of the most difficult, the field of pre-SAL oil [15].

For this work, we use the RS Open Porous Media (OPM) [11], which is a free software based on a collaborative effort between industry and academic sectors. Like other scientific software, OPM makes use of high-performance numerical libraries that are able to solve problems with Differential Equations (PDEs). Moreover, we make use of Cloud service of Microsoft Azure, which granted free way of use of credits. To scale the OPM by establishing a high-performance scientific environment, we ask the following questions:

- What is the performance of OPM in the high-performance Cloud over physical machines used in HPC environments without the use of a hypervisor (virtual machine manager)?
- Which feature does OPM use the most?
- Does OPM use any Acceleration (GPU) feature?
- What is the best way to implement OPM in the Cloud, taking into account cost and energy consumption?

The answer to these questions sums up our contributions with this article:

- We note that even with the overhead produced by the virtualization software, the runtime of the OPM in the Cloud is less than the simulation time in a

*This work was not supported by any organization

[1] a193381 at dac.unicamp.br

[2] Instituto de Computao Universidade Estadual de Campinas (UNICAMP) edson at ic.unicamp.br

similar machine used locally and without the use of a hypervisor;
- During performance analysis, we noticed that OPM only uses CPU intensive - CPU-bound;
- The Data show that the version used (binary package distribution) of OPM is not able to use the GPU feature to obtain acceleration in the simulations;
- We note that the Azure virtual machine model "F16s v2" with 16 VCPUs and 32 GB of memory is the best choice for implementing OPM in the Cloud; we note that it produces, using only one instance, the best simulation time considering energy consumption and price. Moreover, for even larger processes, the use of a Cluster of 4 machines of the F2s v2 version may be a better choice as the processes grow and larger times are required for the simulations.

II. Related work

Jobs that study the problem of consumption in excess of energy has grown in recent years. Initially, we cited the case of CloudLightning, an initiative that seeks to offer a new paradigm for delivering CC services to users [10]. Through a study such as de Guyon et al [5], it is proposed the service delivery where users consciously make the choice for energy consumption through the following parameters: "Little, Medium and Big". If it is limited, therefore, to academic sectors, there is no study that analyzes and proposes how to obtain the best energy performance in Provider Cloud with Microsoft Azure.

There are, however, works that are proposed to do a study of Reservoir Simulators in the Cloud. In the study of Alves et al is done the analysis of RS Mufits in the Clouds of providers such as Amazon EC2 and Microsoft Azure. As a result of the execution of such an application, it is obtained that the Cloud still does not present satisfactory performance when compared to traditional HPC machines and without a HyperVisor. However, greater numbers of instances are not evaluated in the Cloud, nor do they run to simulations using distributed computing, which would make it possible to assess the quality of communication in the Cloud.

III. Methodology

To perform our experiments that measure the viability of OPM in the Cloud, we compare two similar machines: a local physical machine available in our Search Cluster and a virtual machine (instance) in the Cloud of the Microsoft Azure provider. Each machine has 32 GB and 16 cores of processing; while the local machine has a 2.60 GHz Xeon E5-2630 processor, the machine selected in Azure has the Xeon Platinum 8166 2.7 GHz processor. In the experiments the following parameters are used: for a local machine we measure between the execution time, the purchase price of the machine and the energy consumption during a year; for the instance in the cloud, we have measured the time and the price to keep the machine active in the same amount of time.

TABLE I
INSTANCE CHOSEN IN AZURE TO PERFORM THE EXPERIMENTS

Machine	CPU	RAM	Storage	Payment (R$/h)
A8-v2	Xeon E5-2673 v4 2.30GHz	16 GB	80 GB	1,106
A8m-v2	Xeon E5-2673 v4 2.30GHz	64 GB	80 GB	1,451
B8ms	Xeon E5-2673 v4 2.30GHz	32 GB	64 GB	1,236
D32s-v3	Xeon E5-2673 v3 2.4GHz	128 GB	64 GB	1,236
Ds15-v2	Xeon E5-2673 v3 2.4GHz	140 GB	280 GB	4,964
E16-v3	Core i7-5820K 3.30GHz	16 GB	400 GB	3,533
F16s - v2	Xeon Platinum 81682,7 GHz	32 GB	128 GB	2,258
F64s - v2	Core i7-5820K 3.30GHz	128 GB	512 GB	9,031
H8	Core i7-5820K 3.30GHz	56 GB	1 GB	2,643
H8m	Core i7-5820K 3.30GHz	112 GB	1 GB	3,54
L32	Core i7-5820K 3.30GHz	256 GB	5630 GB	8,287
NC6	Core i7-5820K 3.30GHz	56 GB	340 GB	2,988
Cluster D2S-v3	8 x Xeon E5-2673 v4 2.30GHz	64 GB	16 GB	2,552
Cluster DS11 v2	4 x Xeon E5-2673 v3 2.4GHz	84 GB	28 GB	1,98
Cluster E2 - v3	4 x Xeon E5-2673 v4 2.30GHz	84 GB	50 GB	1,768
Cluster F2S v2	8 x Xeon E5-2673 v3 2.4GHz	256 GB	16 GB	2,264

Furthermore, in experiments carried out exclusively in the cloud, we chose multiple instances with different profiles to measure the best time and get a cost profile. The experiments were performed in 17 environments: 13 instances with a more robust profile and 4 clusters of the same instance with a simpler profile. Table 1 below shows the assembly of the experiment.

Finally, in order to calculate the energy consumption of each instance used in the Cloud experiments, we make an inference based on information about the consumption of each hardware present in each virtual machine.

To perform the reservoir simulations at all stages we perform the experiments 5 times to get accurate and we use the dataset Norne, a real case that simulates the behaviour of the fluids inside the reservoir for 3312 days.

IV. Results

In this section, we first show the results obtained between the comparison of the local machine and the Azure F16s - v2 instance. According to the experiments, we get the best simulation time in the Cloud: with 16 parallel processing threads, we get a time of approximately 3 minutes. Figure 1 shows the comparison between the performances

In a small financial analysis of Cloud viability compared to traditional HPC machines we get the following values. Our local research machine was acquired at a price of

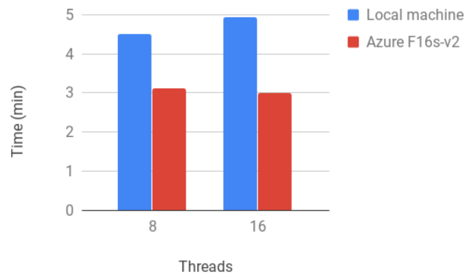

Fig. 1. Details of the execution time. We note that the f16s-v2 machine gets the best time using 16 threads when running its processes in parallel

R$ 10,042.60 and consumes an average of 460 KWh per month, 24 hours a day, equivalent to R$ 2541.16 per year, paid in energy consumption [16][17]. In comparison, if the machine selected - Azure F16s - v2 (R$ 2.26 per hour) in the experiment had a 24-hour availability, we paid the equivalent of R$ 19,509 at the end of a year [18]. That is, by comparison, the option to purchase a machine may be a more viable alternative from a financial point of view. However, Cloud may be a better choice since initially, the user would not have to bear the cost of purchasing the machine. In addition, it becomes easier for the user to keep track of the use of the service since if the instance is disconnected when there is no use, there will be no more charges. Table x measures our price comparison between machines

The following figure shows the results we get when running OPM in the Cloud using multiple instances. According to the results, it is possible to notice the form of economies of scale of the graph when comparing price per simulation time. In other words, we get a reduction in the time according to the price increase. We have found that the F16s - v2 and B8ms machines are the best choices from the point of view of performance and price. During the execution of the simulations, the average time was in 3 min and 30 s.

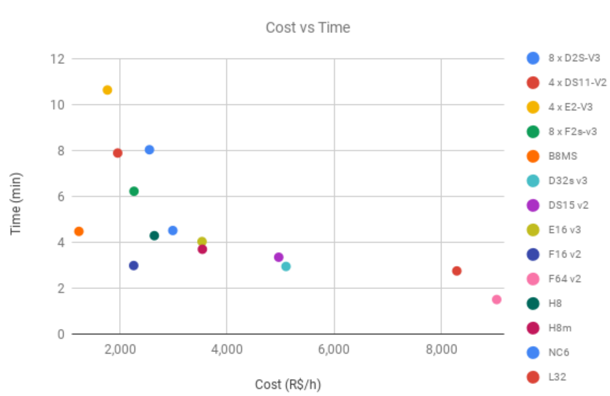

Fig. 2. Graph comparing cost by time. The best results were obtained by the instance F16s - v2 and B8ms.

Finally, figure 3 shows the energy consumption of each machine and the simulation time. According to the results, it is possible to subdivide the energy consumption into three categories: low, medium, high. In the high category, are the machines of the greater power of processing or were Cluster of a machine. We note that the F16s - v2 machine produces one of the best times and the lowest power consumption.

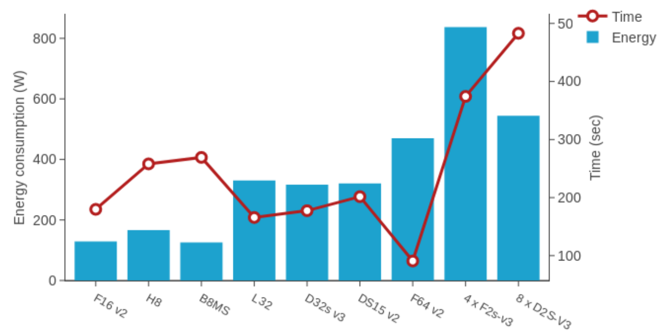

Fig. 3. Comparison between the simulation time and the energy consumption of each machine. It is possible to subdivide consumption into low, medium and high. For the calculation of energy, we are based on the average consumption of each hardware that constitutes a physical machine.

V. CONCLUSIONS

In this article, we have studied the performance and cost-effectiveness of scaling an HPC scientific tool in the Cloud of the Microsoft Azure provider. Data suggest that the Cloud can be considered a viable alternative to traditional HPC machines to run the OPM Reservoir Simulator. It is shown that the execution performance in the Cloud is similar to that obtained when compared to a physical machine without HyperVision. When comparing multiple instances in the cloud to measure cost-benefit and energy performance, the data show that the F16s-v2 instance may be considered the best choice for scaling OPM in the Cloud.

ACKNOWLEDGMENT

I would like to thank Professor Dr Edson Borin to be my advisor during the development of this paper, I introduce the tools of scientific simulations and available resources in Azure so that the simulations were performed. To the Microsoft Azure cloud service for establishing the company/university partnership and making it possible to use its services without charging for which values I also thank for Unicamp that through SAE (Student Support Service) offered student assistance without which this work could not have been developed. Last but not least, I want to thank the graduate students Vanderson do Rosrio, Alexandre Gomes and Jefferson Brunetta who helped me in important phases of the creation of this work; such help was essential to make this work complete in a much shorter time.

References are important to the reader; therefore, each citation must be complete and correct. If at all possible, references should be commonly available publications.

References

[1] Christos K. Filelis-Papadopoulos, George A. Gravvanis, Panagiotis E. Kyziropoulos, A framework for simulating large scale cloud infrastructures, Future Generation Computer Systems, Volume 79, Part 2, 2018, Pages 703-714, ISSN 0167-739X, https://doi.org/10.1016/j.future.2017.06.017.

[2] Cisco, Cisco Global Cloud Index: Forecast and Methodology, 20152020, 2016. http://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085pdf. (Accessed 14 December 2016).

[3] Rosso, Jonathan Patrick, Anlise de Desempenho de Aplicaes Cientficas em Ambiente de Nuvem Privada

[4] VAQUERO, L. M. et al. A break in the clouds: Towards a cloud definition. SIGCOMM Comput. Commun. Rev., ACM, New York, NY, USA, v. 39, n. 1, p. 5055, dez. 2008. ISSN 0146-4833. Disponvel em: ¡http://doi.acm.org/10.1145/1496091.1496100¿. Citado na pgina 19

[5] David Guyon, Anne-Ccile Orgerie, Chrtistine Morin, Deb Agarwal, How much energy can green HPC cloud users save?in: 2017 25th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, (PDP), IEEE, 2017, pp. 416420.

[6] A. Gupta, L. V. Kale, F. Gioachin, V. March, C. H. Suen, B.-S. Lee, P. Faraboschi, R. Kaufmann, and D. Milojicic, The Who, What, Why and How of High Performance Computing Applications in the Cloud, in IEEE International Conference on Cloud Computing Technology and Science (CloudCom), 2013, p. 12.

[7] Netto, Marco A. S. and Calheiros, Rodrigo N. and Rodrigues, Eduardo R. and Cunha, Renato L. F. and Buyya, Rajkumar,HPC Cloud for Scientific and Business Applications: Taxonomy, Vision, and Research Challenges,ACM Comput. Surv., April 2018, 51, 1, jan, 2018,8:1–8:29, http://doi.acm.org/10.1145/3150224

[8] Alves, Maicon Melo, Drummond, Lcia Maria de Assumpo,Anlise de Desempenho de um Simulador de Reservatrios de Petrleo em um Ambiente de Computao em Nuvem.

[9] Lynn, T., Liang, X., Gourinovitch, A., Morrison, J. P., Fox, G. and Rosati, P. (2018) 'Understanding the determinants of cloud computing adoption for high performance computing', Proceedings of 51st Hawaii International Conference on System Sciences (HICSS-51), Hawaii, 3-6 January.

[10] OPM: url: https://opm-project.org/ (Acessado em 05/10/2018 s 11:30)

[11] DUNE. http://www.dune-project.org/ (Acessado em 05/10/2018 s 11:30)

[12] LAPACK. http://www.netlib.org/lapack/ (Acessado em 05/10/2018 s 11:30)

[13] BLAS. http://www.netlib.org/blas/ (Acessado em 05/10/2018 s 11:30)

[14] PETSc. url: http://www.mcs.anl.gov/petsc/ (Acessado em 05/10/2018 s 11:30)

[15] Waisberg, Idel. Brazil's Pre-Salt Layer. (3 de dezembro de 2011). url: http:// large.stanford.edu/courses/2011/ph240/waisberg1/ (Acessado em 05/10/2018 s 11:30)

[16] CPFL (Acessado em 20/10/2018): https://www.cpflempresas.com.br/institucional/tarifas.aspx?emp=CPFL

[17] http://www.buildcomputers.net/power-consumption-of-pc-components.html (Acessado 05/10/2018 s 12:00)

[18] https://azure.microsoft.com/pt-br/pricing/details/virtual-machines/linux/