Is Cloud Computing the Solution for Brazilian Researchers?[*]

Daniel de Oliveira
Federal University of Rio de Janeiro
UFRJ
Rio de Janeiro, Brazil

Eduardo Ogasawara
Federal University of Rio de Janeiro
Federal Center of Technological Education
Rio de Janeiro, Brazil

ABSTRACT

Over the last decade Brazilian government has considerably increased the investment in Research and Development (R&D) and especially in computer science research. Although investments have increased, they are not enough to fulfill the computational needs for all Brazilian researchers since there are many universities and research centers that are not able to acquire a cluster. Cloud computing can play an important role in providing an adequate infrastructure for researchers to do their research and produce valuable results, empowering Brazil in the scientific arena. This paper presents a brief evaluation on how cloud computing can improve Brazilian research by providing an adequate computational infrastructure for researchers.

General Terms

H.3 [**Information Storage and Retrieval**]: Systems and Software-Distributed Systems; C.2 [**Computer Communication Networks**]: Distributed Systems.

Keywords

Cloud Computing, Scientific Experiment, High Performance Computing

1. INTRODUCTION

Over the last years most of scientific research is becoming more and more dependent of high performance computing (HPC) environments such as clusters and computational grids [1,2]. Most of scientific research usually deals with large volume of data (structure and unstructured) that requires huge CPU power to produce a result in a reasonable time.

However, HPC environments such as clusters are usually very expensive to be acquired by a small research center. In Brazil, many research centers and universities do not have enough budgets to buy these computers. This problem can be observed in the TOP 500.org supercomputers list [3] in which there is just one supercomputer in Brazil (86th position). Meanwhile, there are 282 supercomputers in United States, 132 in Europe and 24 in China, illustrating the diverse scenario.

Additionally, according to one of the most important funding agencies in Brazil (Brazilian Scientific Research Council - CNPq) [4], most of the available budget is provided to urban areas while only a smaller part of the budget is provided to rural areas or small cities.

Meanwhile, Cloud Computing [5-7] has emerged as an alternative computing model in which Web-based services can be provided

---

[*] This paper presents the personal opinion of the authors and does not necessarily reflect the opinion of their universities and Brazilian Government.

for different types of researchers to have access to a huge variety of resources, such as software and virtualized hardware.

Cloud computing may be applied to solve problems in many domains of knowledge, including scientific researches. In fact, some researchers have adopted this paradigm for their research, thus moving their research apparatus (programs, systems and data) from local environments to the cloud [8-10].

The important advantage provided by clouds is that it is not necessary to acquire expensive computational infrastructure like clusters to perform their research or even configure many pieces of software. The cloud computing model defines that the resources can be used on demand, and researchers pay only while consuming a resource. Thus, researchers from small university or research center are now able to perform the research in huge HPC environments just by allocating the necessary resources in a cloud [8] for a given moment.

This paper briefly discusses problems in scientific research in Brazil and how cloud computing virtual resources can provide support for Brazilian researchers to execute their tasks since computational infrastructure, such as clusters and grids, are not always available.

This paper is organized as follows. Section 2 describes cloud computing concepts and presents referential background. Section 3 describes the current scenario in Brazil and how cloud computing can aid and we conclude in Section 4.

## 2. CLOUD COMPUTING: REFERENTIAL BACKGROUND

In the last few decades, clusters and grids [11,12] have emerged as a important solution for most HPC problems of industry and scientific research community. However, configuring and maintaining a cluster or a computational grid is usually a very complex activity that requires specialists to support it. In addition, the high cost to acquire this computational apparatus can be considered a serious problem to conduct researches.Cloud Computing [5-7] can be defined as a novel and promising computational model where Web-based services aimed to provide different virtualized capabilities for different kinds of users (in this case scientific researchers) on demand, such as software (Web Services, for example) and hardware (virtual machines, virtual clusters, and so on). The concept behind cloud computing relies on few concepts of several research fields like Service-Oriented Architectures (SOA) [13], distributed and grid computing.

It is a fact that many applications have moved from local desktops to the cloud, such as Google Docs and Google Spreadsheet. These applications and data produced by them are hosted into the cloud and available at any time for any user. The users are able to access programs, documents and data from any computer that is connected to the Internet. Virtualization [14], scalability [15],
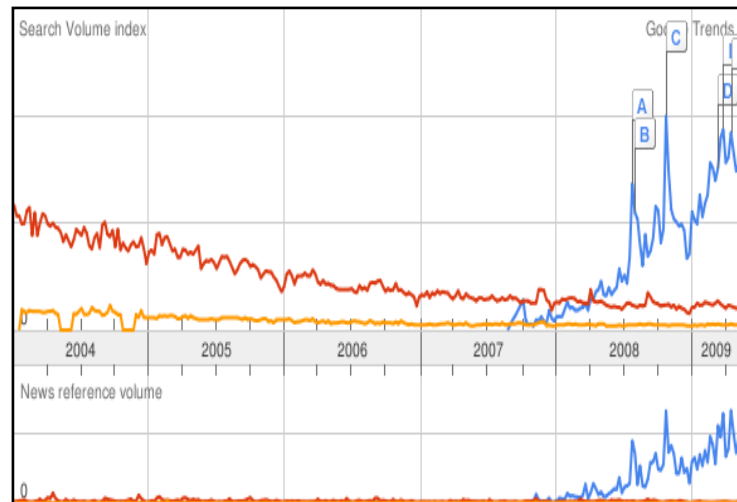
Figure 1 The volume of Google searches about Cloud Computing (blue) compared to Grid Computing (red) and Cluster Computing (yellow) [15]
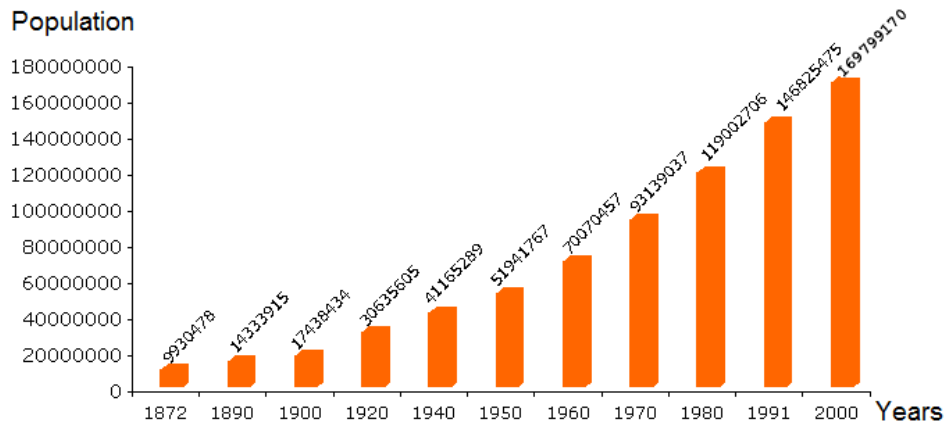
Figure 2 Demographic distribution of Brazilian population adapted from [18]

availability [16] and security [17] are important issues to be considered.

However, since Cloud Computing is a new computational model, it is not entirely clear what are its components and its definition. Actually there is no final definition for this term. Vaquero et al. [6] discusses several definitions and analyzing the different aspects of these definitions. For instance, some definitions focused on scalability and resource optimization while others focused on business model definition [18]. The definition proposed by Vaquero et al. [6] comprises all of the aspects of the existing definitions: "Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically re-configured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees

are offered by the Infrastructure Provider by means of customized Service Level Agreements (SLAs)".

In other words, a cloud may be considered as a distributed system (over the Web) that consists of a collection of virtualized machines (VM). Clouds may be physically implemented by Grids, P2P networks [13] or clusters. According to Buyya et al. [19] clouds are clearly next-generation data centers with virtualized nodes through hypervisor technologies such as VMs.

The major advantage behind the concept of Cloud Computing is that average users (in our case researchers) are able to access a great variety of resources without having to acquire the infrastructure. The pay-per-use model allows the user to make use of the infrastructure just paying for the used time, just like an ordinary rental.
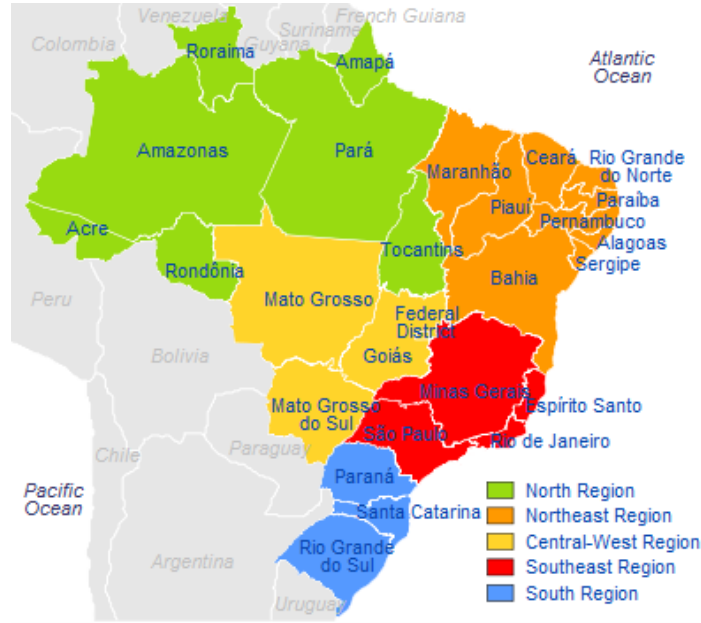
Figure 3 Brazil political map [18]

The popularity of the Cloud Computing is increasing in a fast pace. According to Google†, the search volume in Google labeled as "Cloud Computing" has overcome solid computational paradigms such as Grid Computing and Cluster Computing as shown on Figure 1. It evidences that Cloud Computing appears to be not another hype, but a new solid computational paradigm that is already being already adopted (mainly) by industry and scientific community.

3. CURRENT BRAZILIAN RESEARCH SCENARIO AND HOW CLOUD CAN IMPROVE IT

Brazil is one of the largest countries in extension in the world and has the eighth largest economy of the world. Along with India and China, Brazil has one of the fastest growing factors among all countries. Brazil is also one of the BRIC Countries (Brazil, Russia, India and China).

Scientific research in Brazil is mostly funded by government agencies in public universities and public research institutes. According to Brazilian Institute of Geography and Statistics (IBGE) [20] more than 73% of funding for research comes from government sources.

Although the population of Brazil is growing in a fast pace, as presented in Figure 2, the investment in research is not growing proportionally as we can see in Table 1 that brings the total investment by Brazilian government in US$ in R&D per capita. Although the Brazilian government rose the budget per capita from US$ 6.95 to US$ 37.36 in ten years, it is still far from the current US$ 738.41 per capita in the United States [21,22].

Table 1 R&D *Per capita* investment in research in Brazil over the last decade in US dollars

Year	Budget (in US$ per capita)
2000	6.95
2001	8.72
2002	8.20
2003	11.17
2004	12.12
2005	16.64
2006	19.75
2007	22.21
2008	25.50
2009	29.27
2010	37.36

Although the investments are increasing, it is still not enough to provide necessary computational infrastructure for all researchers in Brazil, especially in North, Central and Northeast regions of Brazil (Figure 3). In addition, the amount of money present in this paper represents all investments in P&D, not necessarily the computational investments only. The result of this lack of investment can be seen in Table 2. The percentage of indexed research papers of Brazilian researchers is 48 times smaller than the percentage of indexed papers of American researchers, for

---

† http://trends.google.com

example. The authors are aware that this percentage is impacted by many other factors; however we also believe that if researchers have better infrastructure this percentage could be increased.

Table 2 Percentage of indexed publications in ISI Web of Knowledge

Country	Percentage of Total number of indexed publications
United States	33.6
Japan	8.5
UK	7.7
Germany	6.8
France	5.1
Canada	4.1
Australia	2.3
India	1.7
China	1.5
South Korea	0.8
Brazil	**0.7**
Chile	0.5
Argentina	0.4

Since not even all researchers have enough budgets to fulfill their computational needs a new solution is indeed needed. Cloud computing presents as a promising solution for researchers that have not enough budgets to buy computational infrastructure. Using cloud environments researchers can only pay for the time they use resources instead of acquiring expensive computational infrastructure. The following subsections presents areas where cloud computing can aid researchers from various domains of knowledge.

3.1 Computational Infrastructure

By executing scientific research in clouds researchers are not required to assemble expensive computational infra-structure or even configure many pieces of software. An average researcher is able to only allocate the necessary resources in a cloud [8].

However, existing infrastructure for executing scientific research using the cloud as a HPC environments are still incipient. Nimbus science cloud [23] is one of these environments in which researchers can create an account and run their research using third party computational resources. This kind of environment manages some new important aspects such as initialization of virtualized instances and management of virtual instances.

This type of computational infrastructure should be able to provide ways to isolate researchers from the complexity of managing applications in the cloud (these applications may require HPC or not). Therefore, the infrastructure itself should be in charge of starting and monitoring executions of many types of applications. This way, researchers do not have to manage

infrastructure components or to worry about configurations, allowing them to focus only on their research process itself.

3.2 Data Sharing

Another advantage for researchers for using clouds is that the collaboration and sharing may be improved. The services available in the Web can share data produced in the researches thus allowing researchers to collaborate.

For example, there are many data repositories in the internet such as Google BigTable, Amazon SimpleDB and Microsoft Azure that allow researchers to store information and to share this information among the scientific community.

However, by using clouds privacy aspects have to be studied. Privacy is fundamental in the scientific research. Data produced in a research is private (and secret) until it is published in a scientific vehicle. In other words privacy in clouds is directly related to the problem of guaranteeing that important (and confidential) data and programs will be not available for researchers that are not authorized to access that. This problem becomes even more crucial since information about a research may be stored in the cloud. It means that important unpublished discoveries and results may be susceptible to external attacks.

There are many risks when researching in clouds. For example, the communication between researchers 'desktop and cloud environments may be monitored and important information can be captured, authentication mechanisms can be broken and so on.

Although this is a problem of almost every application that connects to the Internet and is in the context of a research, we must apply techniques for cloud environments. The first option (and obvious) is to use user-level encryption for research data, that is also effective in the cloud. This is already common for high-value data outside the cloud, and both tools and expertise are readily available.

4. CONCLUSION

Cloud computing is a brand new and fast evolving technology. By providing scalable resources in a pay-per-use model it can be easily used to provide computational infrastructure needed for scientific research. This paper discussed how cloud computing can improve research in Brazilian research scenario. Although it is an introductory and superficial work we can conclude that cloud computing is a serious candidate to aid research development of Brazil.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] M. Mattoso, C. Werner, G.H. Travassos, V. Braganholo, L. Murta, E. Ogasawara, D. Oliveira, S.M.S.D. Cruz, and W. Martinho, 2010, Towards Supporting the Life Cycle of Large Scale Scientific Experiments, *Int. J. Business Process Integration and Management*, v. 5, n. 1, p. 79–92.

[2] Y. Zhao, I. Raicu, and I. Foster, 2008, Scientific Workflow Systems for 21st Century, New Bottle or New Wine?, In: *2008 IEEE Congress on Services*, p. 467-471

[3] TOP 500, 2010. TOP500 List - June 2010 (1-100) | TOP500 Supercomputing Sites. Available at: http://www.top500.org/list/2010/06/100. Access: 13 Aug 2010.

[4] CNPq, 2010. CNPq. *Brazilian Scientific Research Council*

(CNPq). Available at: http://www.cnpq.br/. Access: 13 Aug 2010.

[5] D. Oliveira, F. Baião, and M. Mattoso, 2010, "Towards a Taxonomy for Cloud Computing from an e-Science Perspective", *Cloud Computing: Principles, Systems and Applications (to be published)*, Heidelberg: Springer-Verlag

[6] L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, 2009, A break in the clouds: towards a cloud definition, *SIGCOMM Comput. Commun. Rev.*, v. 39, n. 1, p. 50-55.

[7] I. Foster, Y. Zhao, I. Raicu, and S. Lu, 2008, Cloud Computing and Grid Computing 360-Degree Compared, In: *Grid Computing Environments Workshop, 2008. GCE '08*, p. 10, 1

[8] L. Wang, J. Tao, M. Kunze, A.C. Castellanos, D. Kramer, and W. Karl, 2008, Scientific Cloud Computing: Early Definition and Experience, In: *10th IEEE HPCC*, p. 825-830, Los Alamitos, CA, USA.

[9] A. Matsunaga, M. Tsugawa, and J. Fortes, 2008, CloudBLAST: Combining MapReduce and Virtualization on Distributed Resources for Bioinformatics Applications, *IEEE eScience 2008*, p. 229, 222.

[10] C. Hoffa, G. Mehta, T. Freeman, E. Deelman, K. Keahey, B. Berriman, and J. Good, 2008, On the use of cloud computing for scientific workflows, In: *IEEE Fourth International Conference on eScience (eScience 2008), Indianapolis, USA*, p. 7–12

[11] I. Foster, C. Kesselman, and S. Tuecke, 2001, The Anatomy of the Grid: Enabling Scalable Virtual Organizations, *Lecture Notes in Computer Science*, v. 2150, p. 1―??

[12] I. Foster and C. Kesselman, 2004, *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann.

[13] T. Erl, 2009, *SOA Design Patterns*. 1 ed. Prentice Hall PTR.

[14] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, 2003, Xen and the art of virtualization, In: *Proceedings of the nineteenth ACM symposium on Operating systems principles*, p. 164-177, Bolton Landing, NY, USA.

[15] J. Napper and P. Bientinesi, 2009, Can cloud computing reach the top500?, In: *Proceedings of the combined workshops on UnConventional high performance computing workshop plus memory access workshop*, p. 17-20, Ischia, Italy.

[16] H. Yu and A. Vahdat, 2006, The costs and limits of availability for replicated services, *ACM Trans. Comput. Syst.*, v. 24, n. 1, p. 70-113.

[17] P.D. Leusse, P. Periorellis, P. Watson, and A. Maierhofer, 2008, Secure \& Rapid Composition of Infrastructure Services in the Cloud, In: *Proceedings of the 2008 Second International Conference on Sensor Technologies and Applications*, p. 770-775

[18] J. Geelan, 2009, Twenty-One Experts Define Cloud Computing, *Cloud Computing Journal*

[19] R. Buyya, C.S. Yeo, and S. Venugopal, 2008, Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities, In: *Proceedings of the 2008 10th IEEE International Conference on High Performance Computing and Communications*, p. 5-13

[20], 2010. IBGE :: Instituto Brasileiro de Geografia e Estatística. *IBGE :: Instituto Brasileiro de Geografia e Estatística*. Available at: http://www.ibge.com.br/home/. Access: 13 Aug 2010.

[21] U.S. Census Bureau, 2008, *Current population reports*. US Population.

[22] Ipeadata, 2010, *Ipeadata database*, http://www.ipeadata.gov.br.

[23] K. Keahey and T. Freeman, 2008, Science Clouds: Early Experiences in Cloud Computing for Scientific Applications. Cloud Computing and Applications

[24] Amazon EC2, 2010. Amazon Elastic Compute Cloud (Amazon EC2). *Amazon Elastic Compute Cloud (Amazon EC2)*. Available at: http://aws.amazon.com/ec2/. Access: 5 Mar 2010.