

ANTONIO CARLOS SALZVEDEL FURTADO JUNIOR

## AUTOMATIC ELASTICITY IN CLOUD COMPUTING

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Ciência da Computação do Departamento de Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Eduardo Almeida

CURITIBA

2011

ANTONIO CARLOS SALZVEDEL FURTADO JUNIOR

## AUTOMATIC ELASTICITY IN CLOUD COMPUTING

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Ciência da Computação do Departamento de Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Eduardo Almeida

CURITIBA

2011

## CONTENTS

ABSTRACT	ii
1 INTRODUCTION	1
BIBLIOGRAFIA	4

## ABSTRACT

The concept of cloud computing has been gaining a lot of attention lately, both in the business world and in research papers. Its popularity resides at the low costs it propitiates, opposed to traditional ways of providing computational resources. The main advantage is its elasticity, in which they can be offered on-demand. On the other hand, new challenges arise because of this dynamism. A system needs to guarantee a proper scheduling of these resources, in order to fulfil an organization's demands. At the same time, it needs to meet its scheduling policies ( minimize costs, heigh availability, and so on ) and deal with peaks of demand.

The paper is dedicated to the study of an efficient algorithm.

## CHAPTER 1

### INTRODUCTION

Cloud computing has a big potential to change the Information Technology (IT) world. It was popularized as a business model by Amazon's Elastic Compute Cloud, which started selling virtual machines (VMs) in 2006. Over time, more cloud providers have appeared in the market, offering their computational resources (CPU, memory, and I/O bandwidth). This type of cloud, in which the IT infrastructure is deployed through virtual machines, is referred as Infrastructure-as-a-service (IaaS). One of the most appealing benefits of this paradigm, when associated to cloud providers, is the ability to cut costs. Companies may base their IT strategies on cloud-based resources, spending very little or no money managing their own IT infrastructure. They pay for these resources on-demand, in contrast to the traditional resource provisioning model, in which they would need to deal with both under- and over- utilization of their own resources. Also, the cloud providers may offer lower prices because they are benefited with the economy of scale.

The benefits mentioned earlier are only noticeable when using public clouds – clouds made available in a pay-as-you-go manner to the public by a cloud provider. Although the market has evolved around this type of cloud, organizations might build IaaS clouds using their own infrastructure, known as private clouds. Their aim is not sell capacity over the internet, but give local users an agile and flexible infrastructure to run service workloads in their administrative domains. These users are offered VMs, which are scheduled among a group of physical machines from their organization, which we call a cluster. This leads to a better utilization of resources, since services with little demands can be packed into the same machine, process known as server consolidation. Other benefits, such as the migration of VMs between hosts and the ability to dynamically change the amount of resources provided to it, which are enabled by the technology present in virtual machine monitors (VMMs), make it possible to deal with fluctuations in the workload. This

provides an elastic environment, which is good for a private cloud, and vital for the pay-on-demand model used in public clouds. It's also possible for an organization to create a hybrid cloud, which are useful to supplement a private cloud's infrastructure with external resources from a public one.

The machine virtualization proposed by the cloud concept is essential to achieve its benefits. Database management systems (DBMS), like other software systems, are also increasingly being run on virtualized environments with the same goal. In [1], it is discussed the virtualization design problem for relational database workloads, which can be defined as follows: "*Given  $N$  database workloads that will run on  $N$  database systems inside virtual machines, how should we allocate the available resources to the  $N$  virtual machines to get the best overall performance?*". According to it, the problem of virtualization design may find a better solution when applied to relational database systems due to three factors. First, relational database workloads consist of SQL queries with constrained and highly specialized resource usage patterns. Second, queries are highly variable in the way they use resources – one query might heavily need CPU, while another needs I/O bandwidth instead. Thus, they could benefit from elasticity. Third, database systems already have a way of modeling their own performance, namely the query optimizer.

As mentioned, DBMSs have particularities involving their workloads. Therefore, the application running inside a VM shouldn't be treated as a black box. Instead, it should exploit the database system cost model. As VMMs have parameters to control the share of physical resources, database systems have tuning parameters to manage their own performance. These two sets need to be simultaneously analyzed and tuned. In [2], it is proposed a *virtualization design advisor*. It works by setting the configuration parameters of a VM containing a DBMS. These parameters determine how the shared resources will be allocated to each DBMS instance. It uses information about their anticipated workloads to specify the parameters offline. Furthermore, runtime information collected after the deployment of the recommended configuration can be used to refine this recommendation and to handle fluctuations in the workload. It was not proposed to run this advisor in a cloud, rather it was implemented in a single physical machine, in which two DBMS

instances were deployed.

Our study proposes to implement this virtualization design advisor in a cloud environment. The advisor should be able to configure all the VMs deployed in a cluster, considering the resources of all physical machines. The rest of this paper is structured as follows. In Section 2, the *virtualization design advisor* is described. Section 3 is used to show how a cloud infrastructure is managed. In section 4, we propose an integration of the advisor within the cloud management system.

## BIBLIOGRAFIA

- [1] A.A. Soror, A. Abounaga, and K. Salem. Database virtualization: A new frontier for database tuning and physical design. In *Data Engineering Workshop, 2007 IEEE 23rd International Conference on*, pages 388 –394, april 2007.
- [2] Ahmed A. Soror, Umar Farooq Minhas, Ashraf Abounaga, Kenneth Salem, Peter Kokosielis, and Sunil Kamath. Automatic virtual machine configuration for database workloads. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, SIGMOD '08, pages 953–966, New York, NY, USA, 2008. ACM.



ANTONIO CARLOS SALZVEDEL FURTADO JUNIOR

## AUTOMATIC ELASTICITY IN CLOUD COMPUTING

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Ciência da Computação do Departamento de Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Eduardo Almeida

CURITIBA

2011