

Literature Review

Leo Unbekandt
leo@unbekandt.eu

CRANFIELD UNIVERSITY
School of engineering
Msc Computational and Software Techniques in Engineering

April 29, 2014

Contents

1	Background	2
2	Motivation	2
3	Algorithms	3
3.1	Linear Programming	3
3.2	Bin packing	3
3.3	Network flows	3
3.4	Others	3
4	Real data analysis	3

Introduction

These 10 last years, the Internet has grown at an exponential rate. More and more people are connected, and using always more numerous kinds of device. As the demand is raising strongly, service providers have to be able to support an increasingly amount of customers on their infrastructure. In 2006, Amazon launched Amazon Web Services (AWS) is born, popularising the concept of cloud computing. This platform allows developers and companies to allocate resources on demand, and these are provisionned instantly.

The market of cloud computing has grown really quickly and forecasts predict that this expansion won't stop in the near future. The following study gathers information from different suveys[1]. Currently, the public cloud providers represent a market of 47.4 billion of dollars and is expecting to grow up to 107.2 billion of dollars on only 4 years.

It is no coincidence if the industry and the academic environments are really interested by this domain. On the one hand, it is a field where money is present and on the second hand, we can consider that we are only at the beginning of an era: a lot of things have to be developed, to be optimized and to be found.

1 Background

An interesting definition of the Cloud Computing[2] has been written by the National Institute of Standards and Technology. Different kind of clouds are specified, if Amazon is a “public” cloud, this is not the only way to use a cloud infrastructure: private cloud or hybrid cloud mixing private and public cloud infrastructure are being developped more and more. Thanks to open-source projects like Openstack[3], clouds environment can be installed on private infrastructures. This is necessary, for security, performance or data control purposes.

The evolution of the paradigm of cloud computing has been made possible thanks to different technologies. The virtualisation, as explained by Paul Barham et al.[4] allows servers to be splitted in different sub-components, isolated from each other, sharing the resources of the physical machine.

Technologies have been developed to give people much more flexibility in the way to manage their virtual machines, also called instances. Actually, the concept of live migration, which is detailed in the work of Christopher Clark et al.[5], has been built to move instances from one physical host to another without interrupting the activity of anything running in the virtual machine. The memory is kept intact of course, but also the running connections. The instance may seem frozen for a few second when the migration is finalized, but nothing is disrupted.

2 Motivation

The legitimate question is “Why do people migrate their infrastructure to a cloud infrastructure?”. The answers are multiple, Valentina Salapura explains how a virtualized environment improves the resiliency of an infrastructure [6]. More precisely, when a service requires to be scalable, highly available and fault taulerant, using cloud technologies is essential. In the case of disaster recovery scenarios, they are highly simplified and cheaper thanks to those environments.

As a result the infrastructures are composed of a certain amount of physical machines (PMs) which could be dispatched among different data centers, and each of these PMs, contains a variable number of virtual machines (VMs). The problematic which is now interesting concerns the assignment of these VMs, what is the optimal distributions of the instances

among the different servers? It depends of what characteristic has to be optimized.

Thomas Setzer and Alexander Stage base their study on the statement that energy represents up to 50% of operating costs of an infrastructure. That's why there is a need to optimize it. Using the virtual machine reassignment through live migrations, they are looking at consolidating the VMs on the physical servers. Consolidating an infrastructure consists in reducing the number of PMs which are hosting instances without disturbing the performance of these. After this operation, useless PMs can be suspended and electricity is saved, then when more computational power is required they are resumed dynamically.

In the publication of Weihia Song et al.[7] also introduce their subject by explaining that it has been estimated that Amazon manages more than half a million of physical servers around the world and that it must be a priority for them to reduce their expenses by consolidating their infrastructure.

3 Algorithms

3.1 Linear Programming

3.2 Bin packing

3.3 Network flows

3.4 Others

4 Real data analysis

Conclusion

References

- [1] Louis Columbus. Roundup of cloud computing forecasts update, 2013. <http://www.forbes.com/sites/louiscolumbus/2013/11/16/roundup-of-cloud-computing-forecasts-update-2013/>, November 2013.
- [2] Peter Mell and Timothy Grance. The NIST definition of cloud computing. *Recommendation of National Institute of Standards and Technology*, September 2011.
- [3] Openstack Community. Official website for the openstack project. <http://www.openstack.org/>.
- [4] Barham Paul, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield. Xen and the art of virtu-

- alisation. In *19th ACM Symposium on Operating Systems Principles*, 2003.
- [5] Clark Christopher, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield. Live migration of virtual machines. In *2nd Symposium on Networked Systems Design and Implementation (NSDI'05)*, 2005.
 - [6] Valentina Salapura. Cloud computing: Virtualization and resiliency for data center computing. In *30th International Conference of Computer Design*, 2012.
 - [7] Weijia Song, Zhen Xiao, Qi Chen, and Haipeng Luo. An adaptive resource provisioning for the cloud using online bin packing. *Transactions on Computers*, 2013.
 - [8] Thomas Setzer and Alexander Stage. Decision support for virtual machine reassignments in enterprise data centers. In *Network Operations and Management Symposium*, 2010.