CONTRIBUTION TO VIRTUALIZATION AND DISTRIBUTION TECHNIQUES FOR COMPUTING INFRASTRUCTURES

José Luis Otero Pena, PhD Student

Supervised by: Dr. Felipe Gil Castiñeira[1]
Dr. Pedro Rodríguez Hernández[1]

1: Telematics Engineering, University of Vigo

Universidade Vigo

Gradiant

Motivation of the work

Cloud computing paradigm adoption is driving evolution of IT departments.

New complex needs arise:

- **autoscaling**: to avoid overprovisioning and resources waste
- **advanced monitoring** to predict behavior and prevent problems
- **multi-cloud** or **hybrid cloud** management in order to avoid vendor locking
- **cost-aware** services orchestration to obtain optimal performance at a minimum infrastructure cost.

Figure 1: Hybrid cloud computing and multi-cloud services orchestration avoids vendor locking

In **Big Data** scenarios large amount of data must be read, written and transferred.

As **data increases** each day the challenges to optimize this computations are far from be resolved.

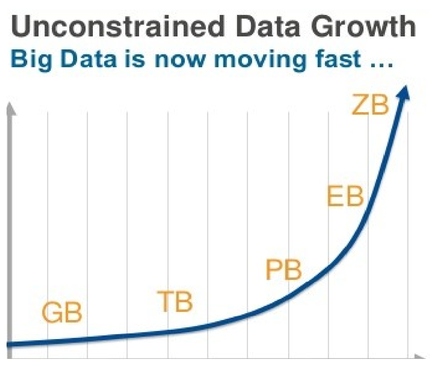

Figure 2: New approaches are needed to deal with the increasing growing of data

Thesis Objectives

1. Reduce the **data storage in unsafe environments** like hybrid clouds
2. Contribute to the **optimization of bandwidth** [1] in data transferring in environments with moving datasets as big data clusters.
3. Obtain **accurate predictions**, alarms and real-time status of deployed services by processing **heterogeneous metrics** captured from different services, operative systems or cloud providers.
4. Determine the **optimal parameters to perform autoscaling** [2] avoiding as much as possible the awareness of the services' characteristics.

Next Year Planning

2015-2016

- **2 conferences or workshops**
- Design and implementation of a new scheme to **optimize the communication and storage**
- **Comparison** of main optimization techniques

Results & Discussions

First year of PhD, part time

Performed literature review

Obtained knowledge about the environment:

Distributed storage Cloud Computing
Virtualization Containers
Big Data Deduplication
Monitoring Private virtual cloud
Autoscaling Hybrid Cloud

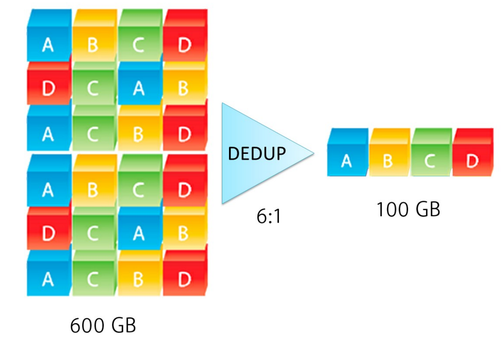

Figure 3: Deduplication removes duplicated data and change it for a link to the only copy.

Main conclusions :

- Most of data have a **considerable rate of duplicated data** [3].
- Hybrid cloud computing, and container virtualization (with **Docker** [4] main)
- Container vs VM: estimated **efficiency rate of 14:1**.
- Hybrid cloud needs **multi-cloud monitoring.**
- Automated autoscaling by using black box approaches like **extended Kalman filters** [5]

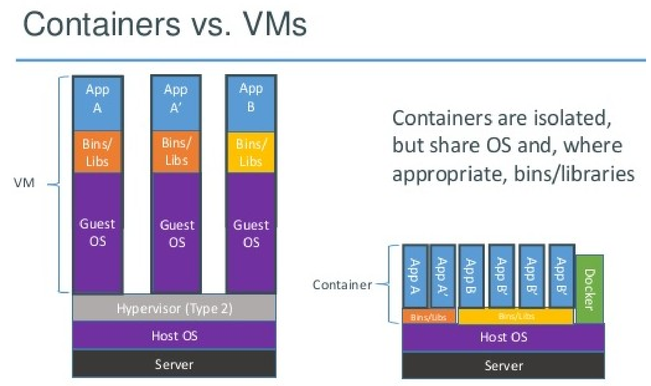

Figure 4: Layers of Container and VM virtualization shows why containers are computationally more efficient.

Research Plan

Milestones:

1. **State of the art analysis**
2. **Data storage deduplication** in untrusted providers
3. **Optimization of data transferring** in Big Data and cloud environments
4. Uniformize **metrics** extracted **from heterogeneous cloud providers**
5. Optimal parameters to perform **automatic scaling**.

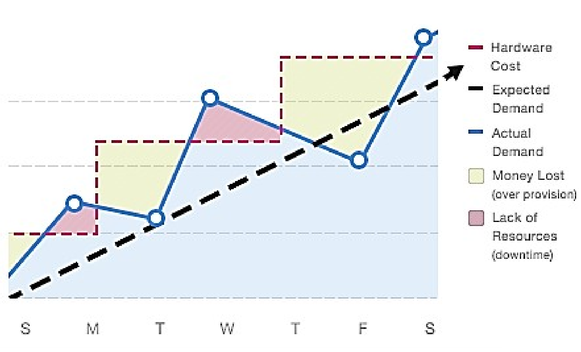

Figure 5: Infrastructure scaling problems

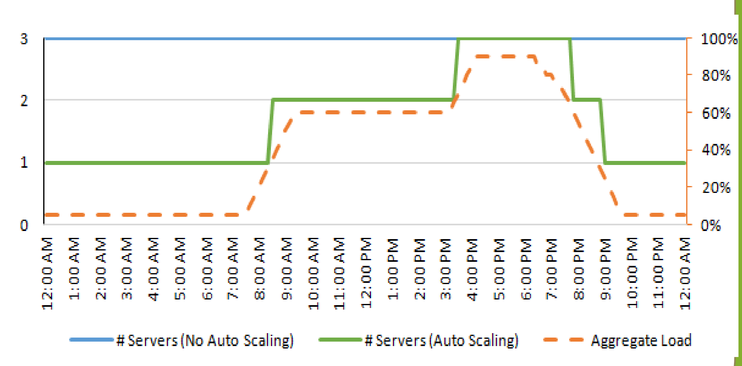

Figure 6: Autoscaling benefits vs over provisioning

References

[1] Zhang, Y., Ansari, N., Wu, M., & Yu, H. (2012). On wide area network optimization. *Communications Surveys & Tutorials, IEEE*, *14*(4), 1090-1113.

[2]Gandhi, A., Dube, P., Karve, A., Kochut, A., & Zhang, L. (2014). Adaptive, Model-driven Autoscaling for Cloud Applications. In *Proceedings of the 11th International Conference on Autonomic Computing, Philadelphia, PA, USA*.

[3] Jin, K., & Miller, E. L. (2009, May). The effectiveness of deduplication on virtual machine disk images. In *Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference* (p. 7). ACM.

[4]Merkel, D. (2014). Docker: lightweight linux containers for consistent development and deployment. *Linux Journal*, *2014*(239), 2.

[5] Alipour, H., Liu, Y., & Hamou-Lhadj, A. (2014). Analyzing Auto-scaling Issues in Cloud Environments. CASCON.