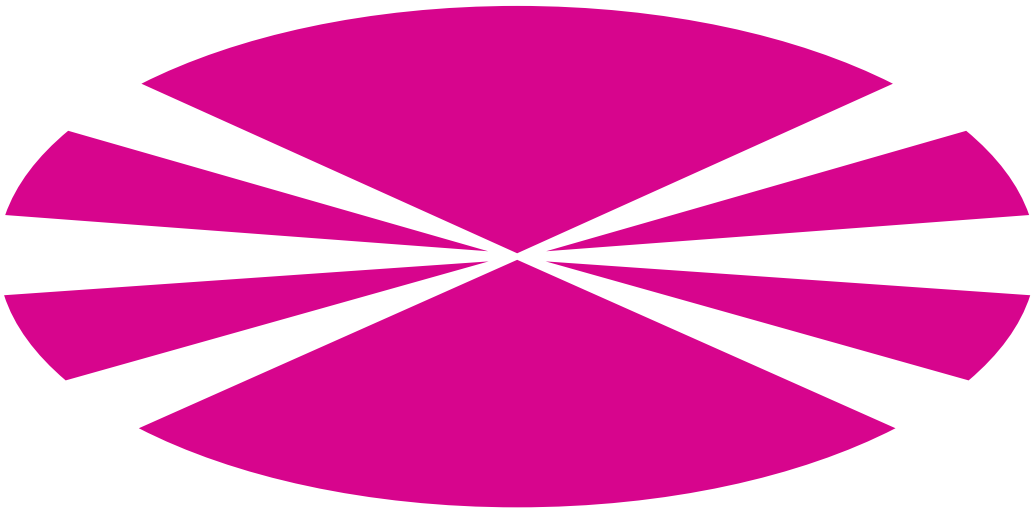

UNIVERSIDADE DA CORUÑA

SCHOOL OF INFORMATICS

Department of Electronics and Systems

Assessment, Design and Implementation of a Private Cloud for MapReduce Applications

Author: Marcos Salgueiro Balsa

Patricia González Gómez

Tutoring: Tomás Fernández Pena

José Carlos Cabaleiro Domínguez

Date: A Coruña, June 2013

Give a man a fish, and you'll feed him for a day.
Teach a man to fish, and you'll feed him for a lifetime.
Anne Isabella Thackeray Ritchie

Great spirits have always encountered violent opposition from mediocre minds.
Albert Einsten

The supreme art of war is to subdue the enemy without fighting.
Sun Tzu, *The Art of War*

[...] It takes these very simple-minded instructions – "Go fetch a number, add it to this number, put the result there, perceive if it's greater than this other number" – but executes them at a rate of, let's say, 1,000,000 per second. At 1,000,000 per second, the results appear to be magic.
Steven Paul Jobs

Summary

The history of computation has seen how the technology's unending evolution has promoted changes in its ways and means. Today, *tablets* and *smartphones*, quantitatively inferiors managing and memorizing numbers, camp freely in a global market saturated with options. The tendency is clear: users will get to use more than one device to access the Internet and will like to have all of their data synchronized and at hand, all the time.

But that is only a part in the equation. At the other side of every service request there lays a server that must deal with an ever increasingly troubling traffic volume, while it maintains response delivery at outstanding delay times — low latency "may" have helped the infant Google rise above the competition. If we also added that the idea of surrounding every implementation effort with energetic efficiency is a transcendental requisite and not simply a good practice, we would have a perfect environment for the proliferation of new distributed paradigms as the *Cloud*. The Cloud is not an intrinsically new idea but an old concept abstraction: *virtualization*. The clouds' cornerstone is flexibility.

Another technology that is constantly making it to the headlines is *MapReduce*. If the Cloud centers around easing infrastructure exploitation, MapReduce's core strength lies in its speeding up driving large masses of unstructured data; with makes them an extraordinary computational tandem. This project puts forth a solution that allows for drawing on computational resources available exploiting both technologies together. Special emphasis has been placed in flexibility of access, being a web browser the only application

required to use the service; in simplifying the virtual cluster configuration, by including a self-managed minimum deployment; and in transparency and extensibility, by freeing source code and documentation as *OSS*, favoring its usage as starting point for larger installations.

Palabras clave

Distributed Computing, Virtualization, Cloud Computing, MapReduce, OpenStack, Hadoop.

Contents

List of Figures