



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 Rodríguez

Date: A Coruña, junio de 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.*

Proverbio Chino

*Great spirits have always encountered violent opposition from mediocre
minds.*

Albert Einstein

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

Especificación

Título: Evaluación, diseño e implementación de un cloud privado para aplicaciones MapReduce

Clase: Proyecto de investigación y desarrollo

Autor: Marcos Salgueiro Balsa

Tutores: Patricia González Gómez
Tomás Fernández Pena
José Carlos Cabaleiro Rodríguez

Tribunal:

Fecha: A Coruña, junio de 2013

Calificación:

Resumen

La historia de la computación ha visto como la incesante evolución de la tecnología ha propiciado cambios en sus modos y formas de proceder. Hoy, los *tablets* y los *smartphones*, cuantitativamente inferiores manejando y memorizando números, campan a sus anchas en un mercado global saturado de opciones. La tendencia es clara: los usuarios pasarán a manejar más de un único dispositivo para acceder a la red y querrán disponer de toda su información sincronizada cuando la necesiten.

Pero esto es sólo una parte de la ecuación. Al otro lado de toda petición de consumo de un servicio se encuentra un servidor que ha de lidiar un volumen problemático de tráfico, al tiempo que mantiene la entrega de resultados en tiempos excelentes. Probablemente la baja latencia de acceso a los resultados de Google en su infancia ayudase en su ascenso. Añadamos que la idea de rodear toda implementación con eficacia energética es un requisito trascendental y no simplemente una buena elección, y tendremos un caldo de cultivo óptimo para la proliferación de nuevos paradigmas como el *cloud*. El cloud no es una idea intrínsecamente nueva sino que representa la generalización de un concepto más antiguo: la *virtualización*. La piedra angular de los cloud es la flexibilidad.

Otra tecnología que también ha florecido recientemente es *MapReduce*. Si bien los cloud se centran en facilitar la explotación de infraestructura, MapReduce mejora el manejo de enormes flujos de información no estruc-

turada; lo cual los convierte en un poderoso tándem computacional. Este proyecto propone una solución que permite aprovechar, utilizando ambas tecnologías, los recursos computacionales de los que se disponga. Se ha puesto especial énfasis en la flexibilidad de acceso, siendo único requisito un navegador web; en la simplicidad de configuración del clúster virtual, incluyéndose un despliegue mínimo automático; y en la transparencia y extensibilidad, al ser liberados tanto los fuentes como los diagramas, pudiendo ser utilizados como punto de acceso a instalaciones mucho más complejas.

Palabras clave

Computación Distribuida, Virtualización, Cloud, MapReduce, OpenStack, Hadoop.

Contents

List of Figures

