

Hardware y Software:

Google

Javier García Peña

Indice

Introducción	3
Software	4
Hardware	6

Introducción:

Google nació en 1996 como un proyecto de investigación de dos estudiantes de doctorado de la universidad de Standford, Larry Page y Sergey Brin.

Este proyecto consistía en estudiar la reciente World Wide Web y su contenido matemáticamente, como si de un grafo gigantesco de información se tratase.

Originalmente Larry trabajaba solo en esta definición hasta que se unió Sergey, y se dieron cuenta del potencial que este proyecto albergaba al tener tanta información mediante los vínculos externos dispusieron una araña (web crawler) para recorrer internet. Poco después desarrollaron el algoritmo de clasificación de página (PageRank), o lo que es lo mismo El Dorado dentro del campo SEO.

Sobre el año 1998 se forma de manera oficial la compañía Google, con sede en un garaje de una amiga.

Para finales de ese años, menos de 4 meses desde su fundación, el índice de páginas de Google ya contaba con 60 millones de páginas. A pesar de que sus fundadores no creían en el modelo de negocios de la publicidad, cedieron y empezaron con anuncios de textos simples.

A principios del siglo XXI, Google empezó a comercializar su gran ventaja, el aspecto SEO, al vender espacio publicitario enlazado a palabras claves, algo que al principio estaba siendo muy infravalorado.

Unos años entrados en el siglo, concretamente 2004, se celebró una OPV que le reportó un valor en el mercado final de aproximadamente 24 billones de dólares.

En la actualidad Google se ha convertido en una empresa gigantesca internacional que cuenta con muchos servicios que forman parte de nuestro día a día, como puede ser su calendario, correo electrónico, mapas, pero también proyectos tecnológicos que podremos ver en un futuro no muy lejano como vehículos autopilotados, dron de reparto o las famosas Google Glass.

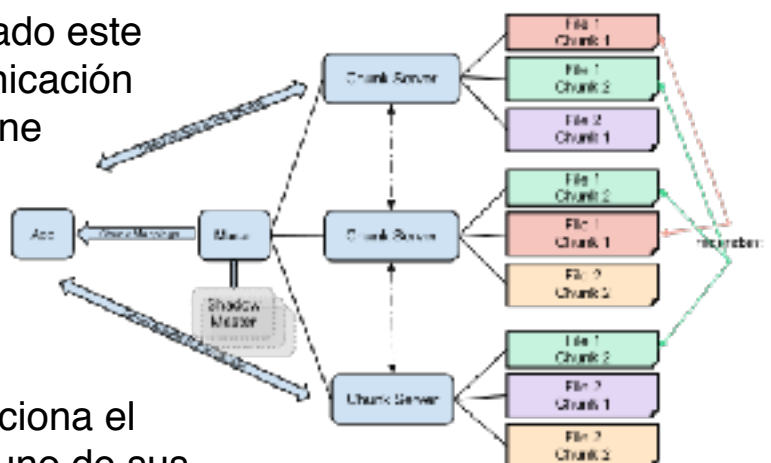
Software:

Para el software, Google hace a medida todo lo que necesita para sus sistemas, por ejemplo el sistema operativo que usa en sus máquinas, llamado Google Web Server(GWS) está basado en Linux, pero construido desde la base para las necesidades de la empresa.

Para el almacenaje de sus archivos, Google usa su propio sistema de archivos, Google File System (GFS). Este sistema de archivos está diseñado para trabajar a nivel de clusters ya que su antecesor BigFiles no era escalable.

Entre las peculiaridades de este sistema tenemos que los archivos son divididos en porciones de 64 megabytes, estos juegos son replicados en varios servidores, normalmente suelen ser 3.

La manera en la que está montado este sistema es a través de la comunicación de un nodo maestro, que contiene los metadatos de los archivos disponibles en los servidores esclavos (Chunkservers), de esta manera, y sumado a la clonación de los archivos mencionada antes, Google soluciona el caso de que una máquina o alguno de sus componentes muera.



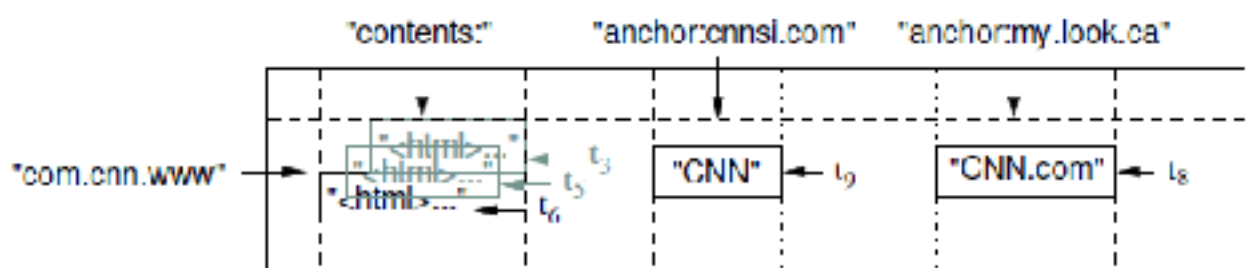
Aún así este sistema es antiguo y en la actualidad Google usa uno mas actualizado, Colossus, que subsana varios errores del Google File System como el fallo por único punto, ya que al tener solo un maestro si esa máquina dejaba de funcionar tenían que reemplazarla físicamente para que se pudiese continuar con el funcionamiento normal.

Colossus, además de tener varios maestros como backup usa los códigos de Reed-Solomon para la detección de errores en su sistema, elegidos porque su coste es más flexible que el utilizaban antes y la mejora del MTTF(mean time to failure) en las simulaciones con su uso. También en este sistema se reduce todavía más el peso de los bloques de datos, de 64MB a 1MB, con lo cual pueden almacenar más bloques en un mayor número de máquinas, reduciendo aún más la posibilidad de pérdidas de datos.

En cuanto a las bases de datos, Google tiene BigTable que es un gestor de base de datos desarrollado, por ellos mismos, sobre Google File System y basado en NoSQL.

La elección de NoSQL es debido, principalmente, a que Google necesita disponer de una de las mayores, si no la mayor, base de datos del mundo, y NoSQL frente a SQL proporciona una mayor adaptación a la escalabilidad horizontal a la que se enfrenta esta empresa.

A la hora de la estructura de almacenaje utiliza un mapa tridimensional que consta de dos cadenas de caracteres y una fecha (fila, columna, timestamp)



Las filas están ordenadas lexicográficamente ya que por este campo es por donde se agrupan campos que tienen buenas propiedades locales, es decir campos que identifican archivos que un cliente puede usar frecuentemente.

En cada celda de ese mapa la información se divide en tabletas de aproximadamente unos 200 megabytes por lo que en cada máquina se pueden almacenar sobre una o dos centenas de tabletas.

Al estar divididas están optimizadas para el Google File System y además BigTable cuenta con una tabletas especiales, llamadas "META1", la cuales son localizadas mediante otra tableta especial(META0) y que sirven para localizar las tabletas que hay en el servidor, es decir, con estas tabletas especiales se respeta el modelo maestro-esclavo del Google File System.

Hardware:

Originalmente cuando Google no pasaba de proyecto universitario en Stanford contaban con unas máquinas que bien podrían encontrarse en algún hogar de la época.



Contaban con:

- Sun Microsystems Ultra II con un procesador a 200Mhz dual y 256MB RAM, siendo esta la máquina principal.
- Dos servidores Pentium II duales a 300Mhz a 512MB de RAM y 9 discos de 9GB, en estos servidores se ejecutaba la búsqueda.
- Un F50 IBM RS/6000 con 4 procesadores, 512MB de RAM y 8 discos duros 9GB.
- Dos armarios en el que había 3 discos duros de 9GB y 6 de 4GB, que iban conectados a la máquina principal.

- Un armario de expansión de discos IBM con 8 discos de 9GB.
- Armario fabricado caseramente con 10 discos de 9GB SCSI.

Sin embargo, en la actualidad y dado la envergadura de la compañía, Google cuenta con algo mas de un millón servidores que usan una arquitectura x86 sobre una versión modificada de Linux.

Estos servidores están repartidos en varios centros de datos distribuidos en distintos países del mundo, entre los que se encuentran, Estados Unidos, Chile, Finlandia o Taiwan.

De entre estos centros de datos, destaca el “Proyecto O2”, un centro situado en The Dalles(Oregón), construido en 2006 que alberga cerca de 12km cuadrados, equivalentes a dos campos de fútbol americano.



La localización es óptima para el centro de datos pues la electricidad es barata y hay una fuente cercana de fibra, lo que minimiza mas los costes del complejo y sus objetivos.

Lo que ocurre concretamente en este complejo no se sabe con certeza, pero si su objetivo, crear un “pulmón” que supla gran parte de la capacidad de red global de la compañía para ejecutar millones de búsquedas.

Bibliografía

XLDB 2013: Evolution of Systems Infraestructure at Google - <https://www.youtube.com/watch?v=gCGvneeHbPQ>

Colossus successor to google file system GFS - <https://www.systutorials.com/3202/colossus-successor-to-google-file-system-gfs/>

Google remakes online empire with 'Colossus' - <https://www.wired.com/2012/07/google-colossus/>

Project 02: Google's Secret Data Center - <https://googlesystem.blogspot.com.es/2006/06/project-02-googles-secret-data-center.html#gsc.tab=0>