

Hardware y Software de Google

Blas Varela López

Índice

- Origen de Google
- El primer Hardware
- Hardware Actual
- Topología de Red
- Infraestructura de Servidores
- Centros de datos
- Google Web Server

- Sistemas de Almacenamiento
 - Google File System
 - BigTable
 - Spanner
 - Google F1
- Servicio de Bloqueo
- Sistemas de Indexación
 - Caffeine
 - HummingBird

Origen de Google

- Creadores Larry Page y Sergey Brin en 1996.
- Proyecto Universitario en Stanford.
- Originalmente conocido como Backrub.
- En 1997 cambian el nombre a Google.
- En 1998 se funda Google Inc.



El Primer Servidor

- Sun Ultra II con procesador 200MHz dual y 256 MB de RAM.
- Dos servidores Pentium II duales a 300 MHz con 512 MB de RAM y 9 discos de 9 GB.
- F50 IBM RS/6000 donado por IBM que incluía 4 procesadores, 512 MB de memoria y ocho discos duros de 9GB.
- Dos armarios con tres discos duros de 9 GB y seis de 4 GB.



El Primer Servidor

- Un armario de expansión de discos de IBM con otros ocho discos duros de 9 GB
- Armario con discos duros caseros que tenía 10 discos duros de 9 GB SCSI.



Hardware Actual

- Containers con hasta 1.160 servidores.
- Cada container consume hasta 250Kw.
- Ordenadores montados a medida.
 - Baterías de 12V.
 - Dos procesadores.
 - Dos discos duros.
 - 8 slots para memoria RAM.



Topología de Red

- Se traduce la dirección de Google a IP's de distintos clusters.
- Cada cluster envía la consulta al servidor menos ocupado en ese momento.
- Los racks de Google tienen entre 40 y 80 servidores.
- Conexión de 1 Gb por rack.



Infraestructura de Servidores

- Distribuidores de Carga.
- Servidores Proxy Squad.
- Servidores Web.
- Servidores de Recolección de datos.
- Servidores de Índices.
- Servidores de Documentos.
- Servidores de Anuncios.

Centros de Datos

- Seguridad
 - Recintos altamente controlados.
 - Control de accesos.
 - Datos en múltiples localizaciones.
 - Datos encriptados.
 - Monitorización discos duros.
- Eficiencia Energética
 - Elevar la temperatura de la zona fría.
 - Aprovechar el entorno.



Centros de datos

América

Condado de Berkeley, Carolina del Sur
Council Bluffs, Iowa
Condado de Douglas, Georgia
Condado de Jackson (Alabama)
Lenoir, Carolina del Norte
Condado de Mayes, Oklahoma
Condado de Montgomery (Tennessee)
Quilicura, Chile
The Dalles, Oregón

Asia

Condado de Changhua, Taiwán
Singapur

Europa

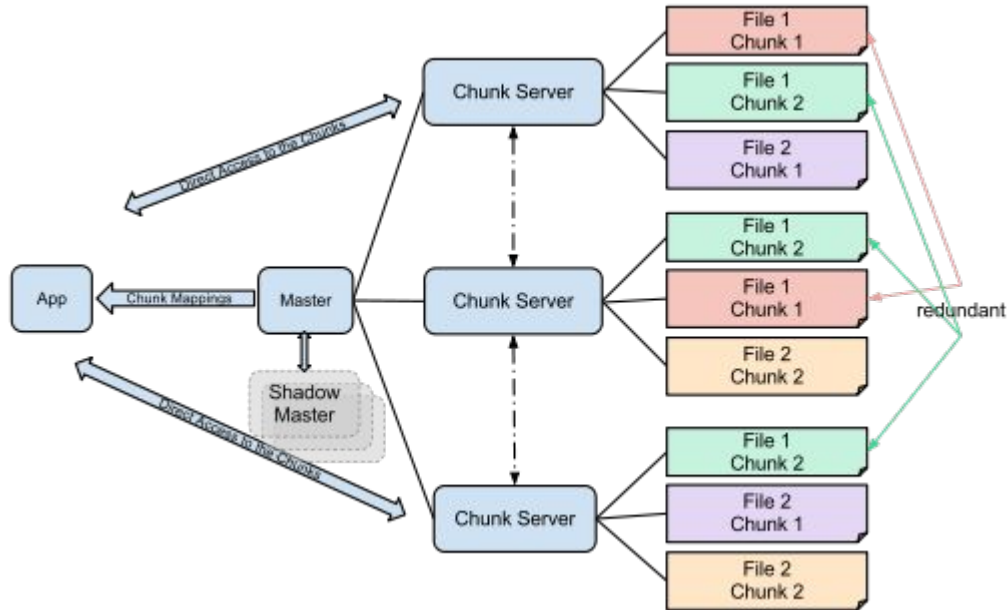
Dublín (Irlanda)
Eemshaven (Países Bajos)
Hamina, Finlandia
Saint-Ghislain, Bélgica



Google Web Server

- Servidor ejecutado por google.
- Utilizado en sistemas Unix.
- Versión modificada y adaptada de Apache.

Google File System



BigTable

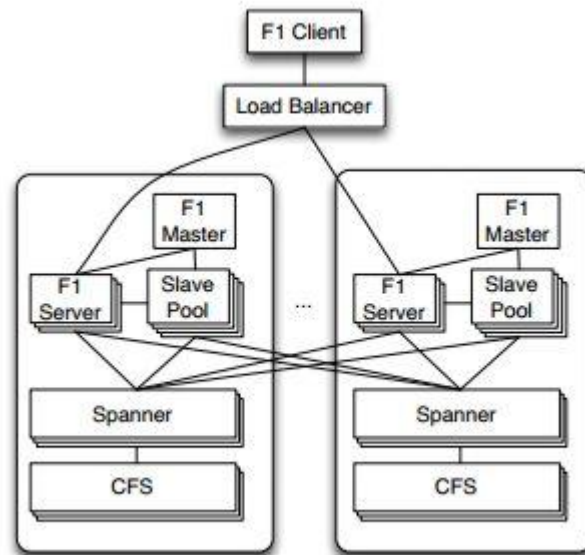
- Sistema de gestión de base de datos.
- Construido sobre Google File System.
- Tablas multidimensionales con la mayoría de celdas sin utilizar.
- Tablas divididas en columnas como tabletas de 200 MB.
- Cada máquina almacena 100 tabletas.

Spanner

- Base de datos NewSQL.
- Uso de Algoritmo Paxos para fragmentar los datos.
- Sincronización de tiempo asistida por Hardware.

Google F1

- Base de datos híbrida entre NoSQL y sistemas SQL tradicionales.
- Construida sobre Spanner.
- Motor de consultas.
- Sistema de seguimiento de cambios y publicaciones.



Servicio de Bloqueo

- Sincroniza los accesos a los recursos compartidos.
- Se soluciona el problema de la coherencia de cache.
- Google lo ha empezado a usar como servicio DNS.

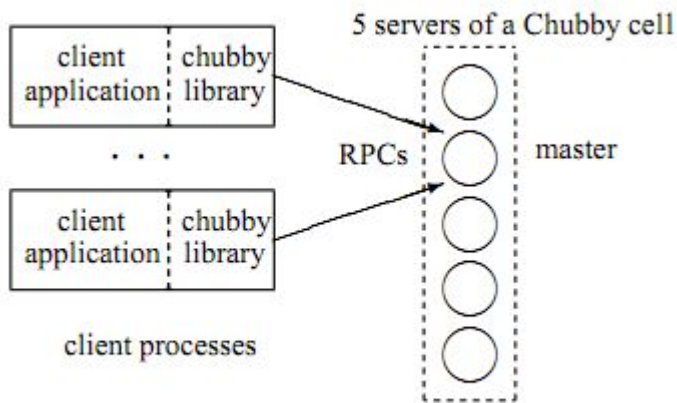
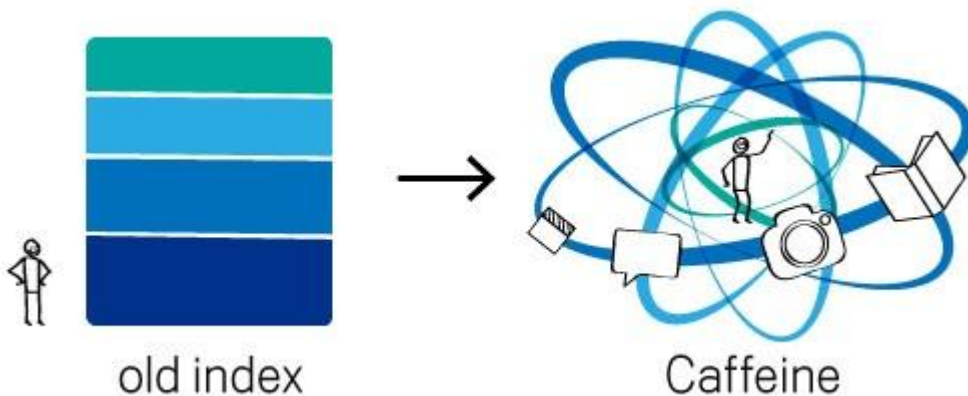


Figure 1: System structure

Caffeine

- Sistema de indexación lanzado en 2009.
- Analiza la web en pequeñas porciones de forma continua.
- Cientos de miles de páginas procesadas por segundo.
- Cerca de 100 millones de GB almacenados.



Hummingbird

- Sistema de indexación lanzado en 2013.
- Rápido y preciso.
- Capacidad de responder preguntas.
- Avance hacia la web semántica.

