

# HARDWARE Y SOFTWARE DE GOOGLE.

## **MOTIVACIÓN**

Hoy en día todos usamos Google cuando queremos buscar información acerca de algo de lo que no sabemos mucho o cuando queremos acceder a un sitio Web cuya URL desconocemos. Es una herramienta muy potente que usamos todos y que nos resulta muy útil a la hora de trabajar porque cuando tecleamos en la barra lo que queremos buscar nos suele aparecer enlaces donde acceder a la información de forma rápida y directa. Seguro que todos tenemos curiosidad por saber cómo es este servidor web. De eso tratará este trabajo, de ver cómo la plataforma Google en cuanto a software y en cuanto a Hardware.

## **HISTORIA BÁSICA DE** **GOOGLE.**

Larry Page y Serguéi Brin comenzaron Google como un proyecto universitario en enero de 1996. Ambos eran estudiantes de posgrado en ciencias de la computación en la Universidad de Stanford.

Originalmente el buscado se llamaba BackRub, pero en 1997 pasó a llamarse Google tras inspirarse en el término matemático “gúgol” ( un uno seguido de 100 ceros) en referencia a su objetivo de organizar una enorme cantidad de información en la web.

Fundan la compañía el 4 de Septiembre de 1998, y estrena su motor de búsqueda el 27 del mismo mes. Disponía de un armario lleno de servidores con unos 80 procesadores y dos routers.

- 2001: En marzo Eric Schmidt es nombrado presidente de la junta directiva. En julio lanza su servicio de búsqueda de imágenes.
- 2002: En febrero se lanza Google Search Appliance. En septiembre se lanza Google Noticias.

- 2003: En febrero adquiere Pyre Labs y con ellos el servicio de creación de blogs Blogger. En diciembre se lanza Google Print y Google Libros.
- 2004: En enero lanza la red social orkut. En marzo se lanza Google Local, que poco más tarde integra Google Maps. En abril se presenta Gmail con 1Gb de almacenamiento. En agosto salió a bolsa.
- 2005: Sale a la luz Google Maps y Google Earth. En agosto presenta Google Talk. En noviembre presenta Google Analytics y posteriormente Google Transit.
- 2006: lanza Picasa, Google Docs, Google Finance, Google Calendar y Google Apps, servicio orientado a empresas. En octubre adquiere por 1659 millones de dólares Youtube.
- 2007: lanza Android. Para abril de este año, Google ya se convirtió en la marca más valiosa del mundo, alcanzando los 66000 millones de dólares, superando a empresas como Microsoft, General Electric y Coca-cola. En noviembre presenta OpenSocial.
- 2008: En septiembre, presenta su propio navegador web Google Chrome. Lanza Friend Connect, Google Latitude y Google Voice, entre otros.
- 2010: Presenta su primer teléfono móvil, Nexus One. en octubre presenta su proyecto de vehículos autónomos. En diciembre presenta su segundo teléfono, Nexus S, fabricado por Samsung.
- 2011: Larry Page es nombre director ejecutivo. En mayor se presentan los Chromebooks. En junio presenta Google+. En agosto adquiere Motorola Mobility por 8.800 millones de euros. Presenta su tercer teléfono inteligente, el Galaxy Nexus, con la cuarta versión de Android.
- 2012: Presenta Project Glass, un proyecto para crear unas gafas de realidad aumentada. Se anuncia Android 4.1, la primera tableta Google, Nexus 7.
- 2014: Google compra SlickLogin, una compañía compuesta por grandes desarrolladores, expertos en seguridad, que han creado un innovador sistema de identificación a través del sonido para los teléfonos inteligentes.

# HARDWARE

Google tiene su propio departamento de diseño de hardware

## ¿CÓMO HA IDO EVOLUCIONANDO EL HARDWARE?

### 1. LA PLATAFORMA ORIGINAL DE GOOGLE

La plataforma original de Google se hizo en la Universidad de Stanford en un proyecto llamado backrub. Estaba escrita en Java y en Phyton y funcionaba de acuerdo con el siguiente hardware:

- Sun Ultra II con procesadores duales de 200MHz y 256 MB de memoria RAM. Esta era la máquina principal del sistema Backrub original.
- 2 Servidores de 300 MHz Dual Pentium II con 512 MB y 9 discos duros de 9GB entre los dos
- F50 IBM RS/6000 (donated by IBM) with 4 procesadores, 512MB of RAM y 8 x 9GB discos duros
- Dos contenedores con 3 discos duros de 9GB y 6 discos duros de 4GB respectivamente que se añadieron a Sun Ultra II.
- Contenedor de expansión de disco IBM con 8 discos duros de 9GB.
- Contenedor de disco casero que contenía 10 discos duros SCSI de 9GB.



## El aparato de búsqueda de Google. GSA (Google Search Appliance)

Este aparato fue presentado por Google hace una década como un cuadro amarillo. Google Search Appliance fue la primera aventura de empresa en el mundo de los negocios, abriendo el camino para una larga lista de otras herramientas de la empresa, incluyendo Gmail y Google Docs y Google Chromebooks. Aunque este “proyecto” se estancó debido a otras herramienta que se desarrollaron tanto de hardware como de software, Google siguió creyendo en el objetivo del GSA. Por ejemplo, resaltamos la versión GSA 7.0 que se sacó, en la cual se añadió una nueva interfaz, nuevas herramientas de búsqueda y capacidades de almacenar datos mucho más grandes de las que había en ese momento.

Es un aparato que permite indexar documentos. Se monta sobre un bastidor que se puede integrar en una red interna o en una página web usando solamente una interfaz de búsqueda de Google para la recuperación de resultados finales. El sistema operativo que usa es Centos. Como características principales de este aparato de búsqueda destacan:

- Contiene tecnologías de búsqueda de Google y proporciona medios para configurar y personalizar el aparato.
- Tiene capacidades de búsqueda que incluyen búsqueda de contenido web, tipos de archivos, bases de datos y contenidos de manejo de sistemas
- Indexación de contenido que se puede buscar
- Posee palabras clave que devuelven valores específicos cuando se emplean esas palabras.
- Incluye sinónimos que ofrecen términos alternativos de búsqueda.
- Proporciona estadísticas en los resultados que contienen número de resultados devueltos, duración de búsqueda, el título del documento, su web, la fecha... etc.
- Contiene términos de búsqueda resaltados para mostrar los éxitos de búsqueda y te permite ver las palabras en contexto sin la necesidad de abrir los documentos.

- Incluye la posibilidad de agrupar resultados parecidos para que la duplicidad se note lo menos posible.
- Emplea diferentes tipos de documentos en los resultados
- Permite ordenar resultados por fecha o importancia



- En cuanto a la escalabilidad, señalar que muchas aplicaciones se pueden enlazar juntar para escalar billones de documentos y que el hardware se puede distribuir entre alrededor de localidades

## **2. Año 2006:**

Cada clúster consta de 359 racks, 31.654 servidores, 63.184 CPUs, 126.368 GHz de potencia de procesamiento, 63.184 Gbytes de RAM y 2.527 Tbytes de espacio en Disco Duro.

Diariamente cada clúster atiende un máximo de 40 millones de búsquedas.

Google tiene repartido más de 20 centros de datos alrededor del mundo (Norteamérica, Asia y Europa) con al menos un clúster, se calcula que google está soportado por más de 600.000 servidores.

## **Project 02**

En el año 2006, Google construyó un centro de datos grande en Oregon, 130 millas al norte de Columbia River y The Dalles. Este centro fue llamado Project 02. Hubo un enfrentamiento entre Google, Yahoo y Microsoft provocado por el infértil desierto en los alrededores de Colombia, a lo largo de la frontera Oregon-Washington, junto con la barata electricidad y la conexión de datos fácilmente accesible. Por aquel entonces, Microsoft y Yahoo habían anunciado que estaban

construyendo grandes centros de información en Wenatchee y Quincy, Wash., 130 millas al norte, pero esto se trataba de una carrera en la cual ellos estaban compitiendo a contracorriente. Se pensó que Google había construido el ordenador más grande del mundo y que tenía más servidores que ninguna otra compañía. En ese centro de datos había dos edificios, y se pensó que en el futuro se alojarían allí decenas de miles de procesadores baratos y discos, mantenidos juntos con una cinta de Velcro en un ensayo que hace Google para fácil intercambio de sus componentes. Las plantas enfriantes jugaron un papel crucial, pues tanta energía de computación provocaba un calor sofocante.

Este centro de datos, tanto su naturaleza como su diseño fue mantenido en secreto durante años a pesar de la dificultad que conllevaba eso al tratarse de un centro de computación tan grande como dos campos de fútbol. El hecho de que Google estaba detrás de este centro de datos fue informado en el entorno local, pero muchos oficiales en The Dalles habían firmado acuerdos confidenciales con Google el año anterior, luego ellos no podían comentar en el proyecto.



### **3. AÑO 2009:** [1][2][3][4]

Nos preguntamos sobre cómo son por dentro físicamente los servidores web y los centros de datos que utiliza google. Ambos son diseñados dentro de la propia compañía. Ben Jai es uno de los diseñadores de estos servidores de Google.

- **Servidores web:**

Dispone de un conjunto de servidores web (sobre un millón y medio) a los cuales se les llama cluster. Para armar los clusters, Google emplea clones con su propio sistema operativo basado en Linux. Dichos clones son diseñados y contruidos por la propia empresa.

Dispone de sistemas propios para combatir las bajadas de los niveles de tensión, ya que estos pueden producir problemas en la memoria de los equipos.

Asimismo, disponen de una batería propia de 12 voltios que sirve para proporcionar energía suficiente en caso de producirse algún error en la red de alimentación.

Vamos a centrarnos un poco más en esta idea...¿por qué tenemos las baterías incorporadas?

- Por un ahorro económico:

Algunos centros de datos típicos se basan en sistemas de alimentación ininterrumpida (UPS) - dispositivos que gracias a sus baterías proporcionan energía eléctrica por un tiempo limitado y durante un apagón eléctrico a todos los dispositivos que tenga conectados. Sin embargo, la construcción de la fuente de alimentación directamente en el servidor, es más barato y además no se pierde capacidad. Estos beneficios permiten un mayor número de servidores.

- Mayor eficiencia:

Grandes UPSs pueden alcanzar un 92-95 por ciento de eficiencia, es decir, se desperdicia una gran cantidad de energía. Los servidores con baterías incorporadas de Google son capaces de alcanzar un 99.9 por ciento de eficiencia.

La eficiencia es importante porque mejora los costos de consumo de energía. Pero además es importante porque la ineficiencia produce pérdidas de calor que hace que sea necesario un gasto mayor para conseguir refrigeración.

Google tiene patentes en este diseño de la batería incorporada, sin embargo Hoelzle, vicepresidente senior de la infraestructura técnica en Google, ha afirmado que estarían dispuestos a conceder licencias a los vendedores.

Estos servidores no tienen un hardware demasiado raro o avanzado:

- Dos procesadores
- Dos discos duros

- 8 ranuras para memoria RAM
- Procesadores x86 AMD e Intel

En la siguiente imagen podemos ver el interior de uno de los miles de servidores web de los que dispone google.



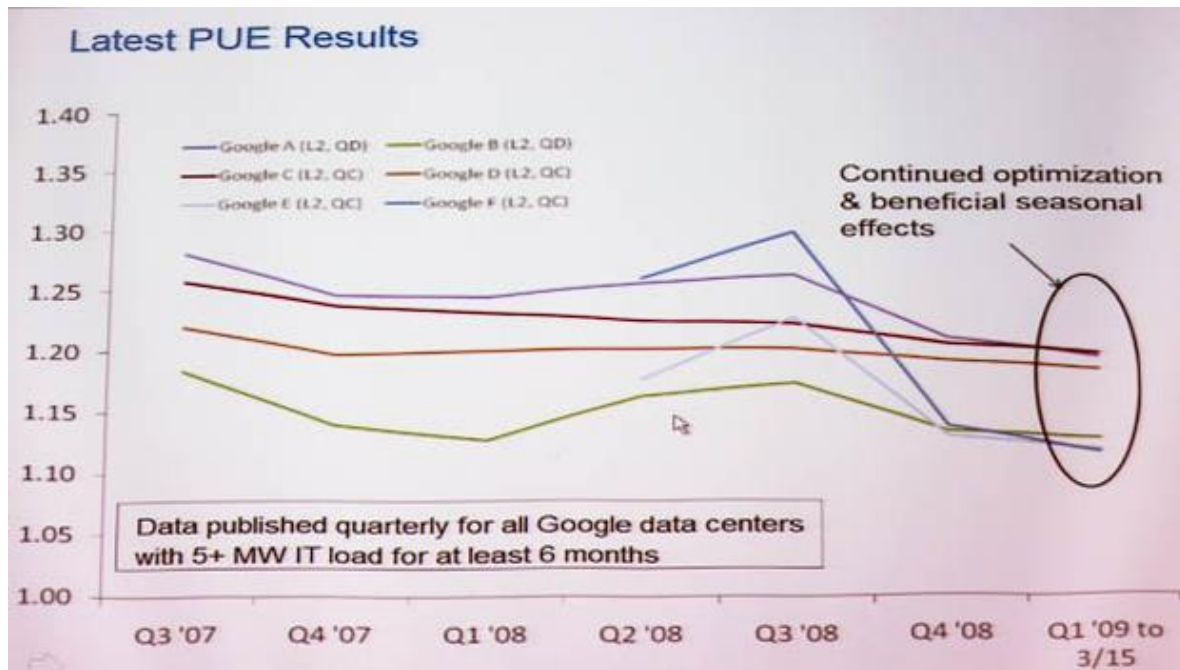
- **Centros de datos :**

Jimmy Clidaras, ingeniero distinguido de la empresa, reveló que el núcleo de los centros de datos de la compañía se construyen a partir de contenedores de carga 1AAA donde se introducen más de 1.600 servidores web. Cada uno de los centros de datos tienen suelo falso, cableado, energía eléctrica, aire acondicionado y una zona para la administración. Como dato curioso cabe destacar que cada contenedor necesita una energía de 250kW.

Google también desveló resultados sobre la eficiencia energética de su centro de datos, realizando las medidas usando el estándar PUE (Power Usage effectiveness). En ella se compara el total de energía consumida por el centro de datos con la cantidad de energía que realmente llega al equipamiento, lo que nos permite conocer la cantidad de energía perdida en los sistemas de refrigeración, iluminación.... Una puntuación



perfecta de 1 significa que no se pierde energía. Una puntuación de 1.5 significa que los servicios auxiliares consumen la mitad de la energía. Las puntuaciones que obtiene Google son envidiablemente bajas, sin embargo la compañía sigue tratando de reducirlas aún más. Veámos cómo fué disminuyendo desde 2007 al 2009.



Este tipo de centros de datos se han estado usando desde el año 2005. Chris Malone es una de las personas que están involucradas en el diseño de estos centros de datos. En la siguiente imagen vemos los contenedores de centros de datos. Como en los centros de datos convencionales, los contenedores de carga tienen los suelos elevados.



#### **4. AÑO 2012:**

##### **Datos y seguridad [8]:**

Mantienen los datos seguros mediante decenas de medidas de seguridad críticas.

Los servidores especializados exclusivamente para los centros de datos de google siguen sin ser distribuidos externamente. No incluyen hardware ni software innecesario para reducir el número de posibles vulnerabilidades.

También han incorporado fuertes medidas de recuperación ante desastres: si ocurre algún incendio u otro tipo de accidente, los datos se transfieren automáticamente a otro centro de datos para que los usuarios puedan seguir trabajando. Los generadores de emergencia seguirán alimentando los centros de datos incluso en caso de fallo de suministro eléctrico.

Protegen nuestros datos:

Distribuyen todos los datos en numerosos equipos situados en ubicaciones diferentes. A continuación, dividen y replican los datos en varios sistemas para evitar un único punto de fallo. Como medida de seguridad asignan nombres aleatorios a los distintos fragmentos de datos, de este modo, resultará prácticamente imposibles de leer para un usuario normal.

Mientras cualquier usuario trabaja, los servidores realizan copias de seguridad automáticas de los datos más importantes. Estos nos permite recuperar nuestro trabajo en cuestión de segundos si por cualquier motivo tenemos un accidente.

También realizan un seguimiento riguroso del estado de cada disco duro de los centros de datos. Destruyen dichos discos duros antes de desecharlos mediante un proceso meticuloso para evitar el acceso a sus datos.

Equipo de seguridad en constante trabajo:

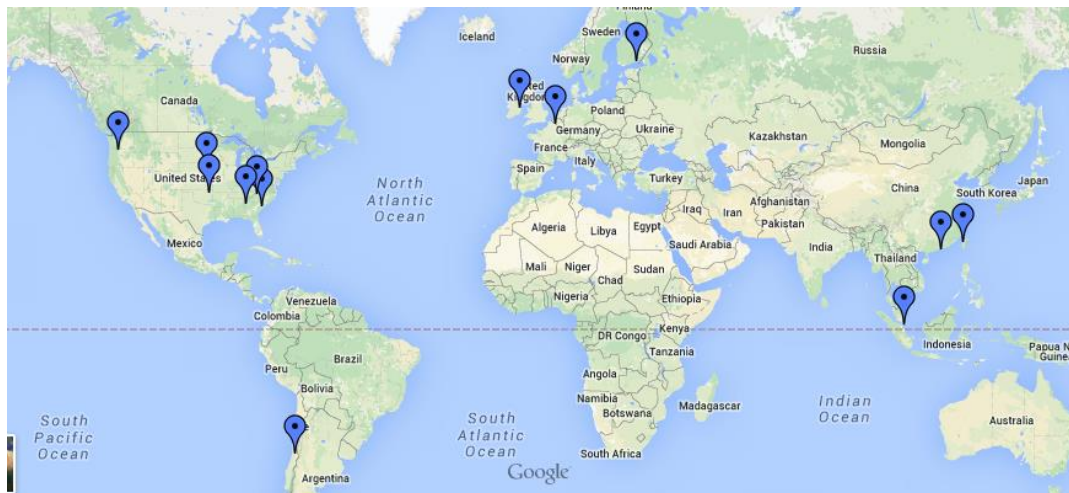
El sistema de seguridad de la información mantiene los sistemas de defensa del perímetro de la empresa, desarrolla procesos de revisión de seguridad y diseña la infraestructura de seguridad personalizada.

También juegan un papel fundamental a la hora de desarrollar e implementar los estándares y políticas de seguridad.

En los centros de datos se disponen de controles de acceso, guardias de seguridad, vigilancia por vídeo y una cerca de seguridad para proteger físicamente las instalaciones.

### **UBICACIONES GLOBALES: [9]**

Poseen centros de datos en todo el mundo para mantener el funcionamiento diario.



### **5- AÑO 2013: [5]**

Google gasta unos 7.3 billones de dólares en mejorar sus centros de datos. Este gasto está dedicado a una expansión masiva de la red de centros de datos global de la empresa, lo que representa, posiblemente, el mayor esfuerzo que se ha hecho nunca en la industria de los centros de datos. Veamos las ampliaciones en los centros de datos anunciadas en este año:

- Una expansión de 600 millones de dólares en su complejos de The Dalles, Oregón.
- Una expansion de 400 millones de dólares en su campus en Council Bluffs, Iowa.
- Nuevas inversiones con un coste de 600 millones de dólares en una fase adicional para su centro de datos en el campus situado en Lenoir, North Carolina.
- 600 millones de dólares de inversión para apoyar la nueva construcción en el centro de datos de South Carolina, condado de Berkeley.

- 390 millones de dólares para el centro de datos de Bélgica.
- Realiza una compra de terreno en Oklahoma para otra posible futura expansión.

La construcción de varias instalaciones en un solo sitio suele ser más económica que la construcción de un nuevo sitio, ya que la infraestructura básica para la energía y la conectividad se instala durante la construcción de la primera planta, dejando menos trabajo y gasto en fases posteriores. La mayoría de las inversiones de capital que realiza son destinadas a mejorar los centros de datos, servidores y equipos de red.

## **CATÁLOGO DE HARDWARE DE GOOGLE [16]:**

Google tiene interés en expandir sus líneas de negocio más allá de su núcleo principal. Para ello en los últimos años ha desarrollado un potente hardware en diversos campos tecnológicos.

- **“Cucharas inteligentes”[13]:**

Google trata de ayudar a los pacientes con Parkinson creando cucharas inteligentes. La idea original fue desarrollada por Lift Lab[14]. La tecnología que se ha usado detecta cuando la mano tiembla y hace unos ajustes instantáneos para mantener el equilibrio. Es capaz de reducir los temblores hasta en un 76%. Se puede adquirir por unos 295 dólares.

- **Trekker:**

Se trata de una pesada mochila que porta un dispositivo formado por 15 cámaras. Toma imágenes, que las va almacenando en un disco duro propio, cada dos segundos. Además dispone de un GPS para localizar dónde se van tomando. Ya ha permitido captar imágenes del Gran Cañón y el Monte Everest, entre otros.

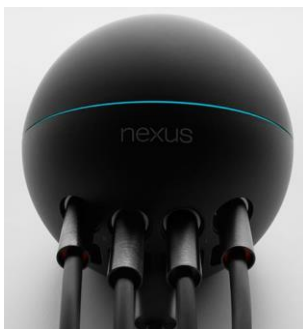
- **Robots de Boston Dynamics[15]:**

En 2013 Google compró esta firma de robótica incorporando una serie de robots como BigDog (permite desplazamientos sobre terreno helado), Petman (robot con forma y movimientos humanos) y Cheetah (permite desplazarse a 29 millas por hora), entre otros.



- **Nexus Q:**

Es un dispositivo que funciona para dar órdenes a los altavoces, tablets, televisiones o smartphones desde otro dispositivo android. Sin embargo no fue lo suficiente exitoso y a finales de 2012 fue descatalogado. Además en 2013 se rompió la compatibilidad entre los dispositivos Android y Nexus Q, lo que lo convirtió en un bonito pisapapeles.



- **Proyecto Tango[18]:**

Este Smartphone de Google cuenta con sensores que identifican el espacio que rodea al usuario, trata de mapear el mundo. Transforma la forma de entender los espacios: cualquier persona podría tener mapeado su propio hogar para poder realizar reformas, decoraciones....Otra ventaja importante es que las personas ciegas

podrían caminar tranquilamente guiados por la asistencia de su smartphone.

- **Proyecto Ara:**

Este Smartphone funciona como piezas de Lego que se conectan a una base. Es el usuario el que organiza las prioridades del espacio para poner un mejor procesador, más RAM, más memoria interna, más batería o mejor cámara. Si nos quedamos cortos de algún componente, podemos añadirlo sin tener que cambiar los demás. De este modo la producción es más barata y la contaminación es menor.

- **Google Glass.**
- **Chromecast.**
- **Termostato Nest.**
- **Coche autónomo.**
- **Google Chromebook.**

## TOPOLOGÍA DE RED DE GOOGLE

Los detalles de las redes mundiales de Google privadas no están públicamente disponibles, pero algunas publicaciones de Google hacen referencia al informe “Atlas Top 10”, que ubica a Google como el tercer proveedor de servicio de Internet (ISP) detrás de Level 10.

## The “ATLAS 10” Today

Rank	Provider	Percentage
1	Level(3)	5.77
2	Global Crossing	4.55
3	ATT	3.35
4	Sprint	3.2
5	NTT	2.6
6	Cogent	2.77
7	Verizon	2.24
8	TeliaSonera	1.82
9	Savvis	1.35
10	AboveNet	1.23

(a) Top Ten 2007

Rank	Provider	Percentage
1	Level(3)	9.41
2	Global Crossing	5.7
3	Google	5.2
4		
5		
6	Comcast	3.12
7		
8	<i>Intentionally omitted</i>	
9		
10		

(b) Top Ten 2009

- Based on analysis of anonymous ASN (origin/transit) data
  - Weighted average percentage
- Top ten has NO direct relationship to Observatory participation
- Tier1s still carry significant traffic volumes (and profitable services)
- But Comcast and Google join the top ten

Para funcionar una red tan grande con conexiones directas a tantos ISPs como se pueda al menor coste posible Google tiene una política (peering) muy abierta.

Se puede acceder a la red de Google desde 67 puntos de intercambio públicos y 69 localizaciones diferentes alrededor del mundo. En Mayo de 2012 Google tenía 882 Gbits por segundo de conectividad pública. Esta red pública se usa para distribuir el contenido a los usuarios de Google, además de navegar por Internet para construir los índices de búsqueda.

Se estima que Google tiene más de 2000000 servidores, que están ordenados en racks de clusters en varias ciudades del mundo. Un **rack** es un soporte metálico destinado a alojar equipamiento electrónico, informático y de comunicaciones. Un cluster es un conjunto de ordenadores que están conectados entre sí por una red de alta velocidad y que parecen ser un sólo ordenador. Cada cluster de Google tiene miles de servidores. Los centros más importantes están en Mountain View (California), Virginia, Memphis Atlanta y Dublín. Al estar



los servidores de Google geográficamente dispersos, Google puede ofrecer un servicio más rápido a sus usuarios. Esto es esencial, pues en el año 2005 Google había indexado 8.000 millones de webs.

El funcionamiento es el siguiente: Cuando un ordenador cliente intenta conectarse a google, varios servidores DNS resuelven el nombre de dirección [www.google.com](http://www.google.com) en múltiples direcciones IP siguiendo la política Round-Robin, es decir, el orden en el que se traducen las direcciones es racional, con tiempos equitativos. Esto es el primer nivel de balanceo de carga y dirige el cliente a diferentes clusters de Google. Una vez que el cliente se conecta al servidor adicional de balanceo de carga se le envían las peticiones al servidor web menos cargado. Esto hace que Google sea uno de los CDN (content delivery network) más grandes y más complicados del mundo.

Un content delivery network es un gran sistema de servidores distribuidos desplegados en múltiples centros de datos a través de Internet.

## SOFTWARE

La mayoría de los subsistemas y componentes software que usa Google en sus servidores fueron desarrollados “en casa”.

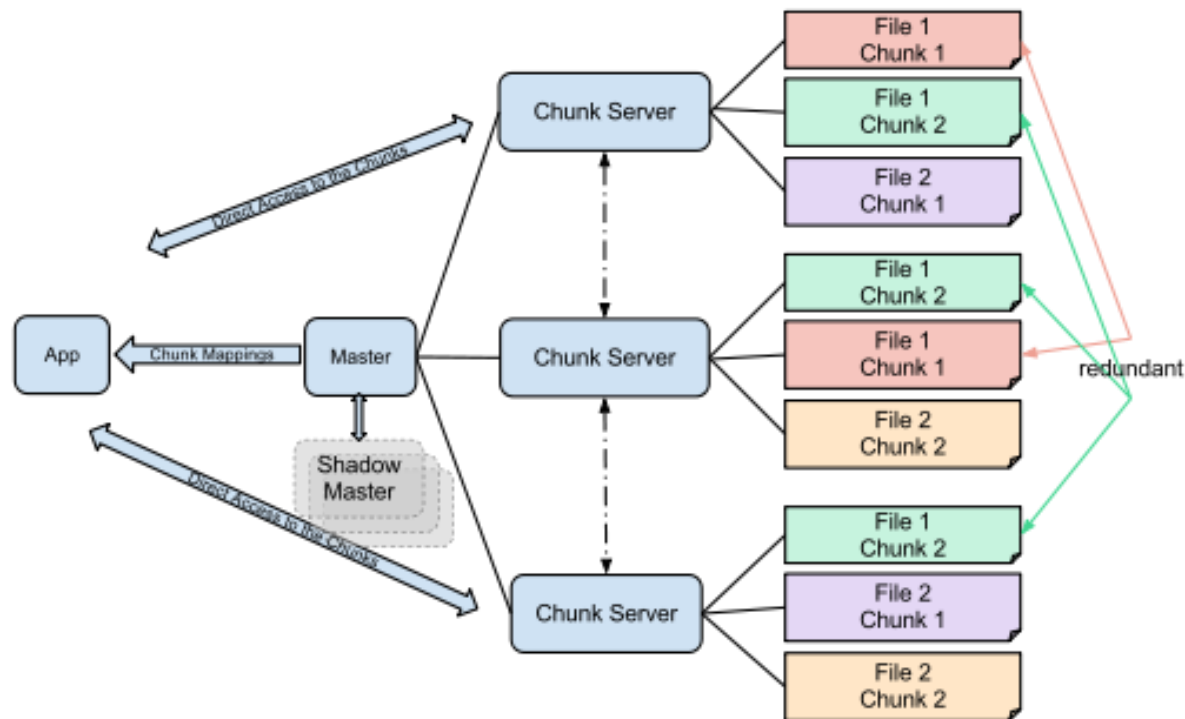
El software que se ejecuta sobre la infraestructura de Google incluye:

- Un servidor Web de Google (GWS): Este servidor está hecho a medida y es sólo usado por Google. Ahora funciona casi al 13% de todos los sitios web activos. Es el servidor web que google usa para los servicios on-line. Aunque Apache sigue siendo el servidor Web más frecuente, seguido de los servidores de Microsoft, Google encabeza a todos los demás.
- Sistemas de almacenamiento, entre los que destacan:
  - Sistema de archivos de Google (Google File System). Está diseñado del siguiente modo:
    - Los archivos están almacenados en bloques de tamaño fijo 64 MB



- Consigue la fiabilidad por medio de la replicación: cada bloque está replicado a través de 3 o más chunkservers.
- Hay un único maestro con el fin de coordinar los accesos y mantener la información mediante simple manejo centralizado.
- No hay caché de información, lo que supone un pequeño beneficio debido a los grandes conjuntos de datos

La siguiente imagen muestra el funcionamiento del sistema de archivos de Google:



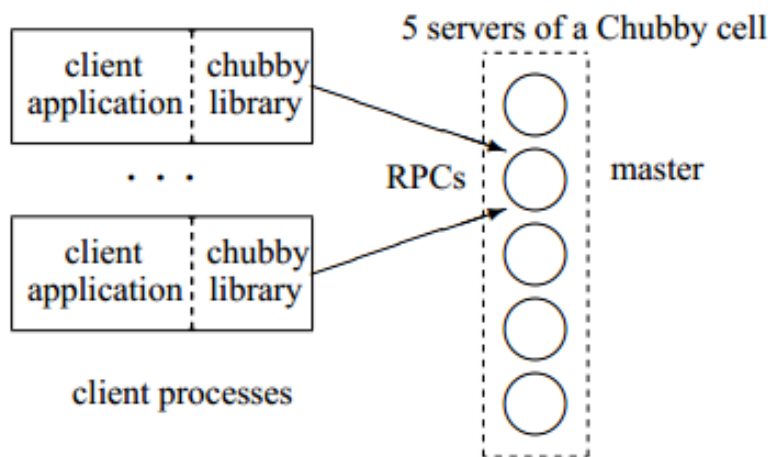
- Bigtable: Es un sistema de almacenamiento distribuido para manejo de estructuras de datos. No soporta un modelo completo de bases de datos. Está construido sobre GFS/Colossus. Algunas características destacables de la bigtable de Google son:
  - Está estructurado a nivel de Cluster
  - Exporta un mapa distribuido y ordenado
  - Divide y equilibra información basada en tamaño y carga
  - Para almacenamiento de archivos emplea GFS/Colossus.

- Spanner : Es una base de datos de Google globalmente distribuida NewSql. NewSql es un clase de modernos sistemas de manejo de bases de datos relacionales.
  - Google F1: Un sistema de manejo de bases de datos basado en Spanner, sustituyendo una versión de MySql.
- Chubby lock service: Es un sistema distribuido diseñado para actividades de grano grande. Inicialmente fue diseñado para actividades en los sistemas distribuidos de Google, pero actualmente se usa ampliamente como un repositorio y servicio de nombres para información de configuración. Proporciona bajo volumen de almacenamiento pero un almacenamiento fiable para sistemas de almacenamiento loosely-coupled (tienen poco conocimiento de los otros componentes) hechos para máquinas pequeñas que se comunican por redes de alta velocidad.
- Una célula de Chubby servirá unas 10000 máquinas de 4 cuatro procesadores conectadas por un ethernet de 1GB. La mayoría de las células están limitadas a un centro de datos, aunque es posible hacer funcionar a una célula de Chubby con réplicas localizadas alrededor de miles de kilómetros.

Este servicio está diseñado para permitir a los clientes sincronizar actividades y ponerse de acuerdo en detalles básicos de su entorno. Puede ser usado de varias maneras como designar a un servidor maestro, permitir al maestro descubrir a los servidores bajo su control, permitir a los clientes encontrar a su maestro, particionar pequeñas cantidades de metainformación o distribuir el trabajo entre varios servidores.

La interfaz de Cubby es similar a sistemas de archivos distribuidos simples que hacen escrituras y lecturas de archivo entero y tiene y locks consultivos. Pero también está diseñado para hacer más hincapié en la fiabilidad y disponibilidad para muchos clientes que en alto rendimiento. El servicio lock enfoca menos en capacidad de almacenamiento y rendimiento.

Aquí mostramos una imagen que refleja la interfaz de Chubby:



- Lenguaje de programación Sawzall
- MapReduce: Un modelo de programación y una implementación asociada para procesamiento y generación de grandes conjuntos de información con un algoritmo distribuido o paralelo en un cluster. Un programa de MapReduce está compuesto de una función Map que realiza filtramiento y ordenación y de una función reduce que realiza una operación de resumen.
- Google usa los tres siguientes sistemas de indexación y búsqueda:
  - TeraGoogle: Fue un gran índice de búsqueda de Google lanzado en 2006. Google intentó ser todo lo que posiblemente puede ser para todos, un one-stop-shop.
  - Caffeine: Conocido como el sistema de indexación de reparación. Su índice fue separado previamente en capas, con algunas capas actualizadas más rápidamente que otras. Caffeine adoptó una nueva forma en la que el índice completo sería actualizado continuamente. Google pasó de un sistema indexado en lotes a un sistema de indexación que era actualizado en el acto. Con Caffeine se podía actualizar el índice continuamente porque se podía examinar la web en porciones pequeñas, mientras que antes para actualizar el índice se necesitaba analizar la web entera. Si se descubrían nuevas páginas eran añadidas de forma automática al índice de búsqueda.

Caffeine tenía capacidad para procesar cientos de miles de páginas cada segundo. El índice tenía una gran capacidad de almacenamiento en una base de datos y se podía añadir nueva información a gran velocidad.

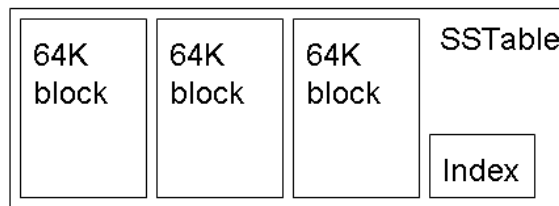
Con Caffeine no sólo logró Google un nuevo sistema de indexación, sino que se estrenaría una arquitectura software que probablemente apoyarían todas las aplicaciones on-line durante los siguientes años.

- Hummingbird :Fue diseñado en Septiembre de 2013. Este sistema de búsqueda ordena a través de las búsquedas (y correspondientes respuestas) que se han sido realizadas. Google dice que este sistema adoptó este nombre de ser preciso y rápido.

Hummingbird le presta más atención a cada palabra en una petición teniendo más en cuenta el significado petición entera que el significado de alguna/s palabra/s en concreto.

Hummingbird está diseñado para aplicar la tecnología a billones de páginas que están en Internet, además de Knowledge Graph facts. No se sabe hasta qué punto este nuevo sistema de búsqueda supone una mejora, pero lo cierto es que no ha habido muchas quejas por parte de los clientes de Google de que los resultados empeoren, cosa que los clientes suelen hacer cuando los resultados dejan de ser favorables.

- Google ha desarrollado varias abstracciones que usa para almacenar la mayoría de su información, entre las que destacan:
- Protocol Buffers: Es como XML pero más pequeño, simple y rápido. Con este mecanismo, con que se defina como se quiere que structure la información una sola vez, se puede usar un código fuente especial generado para leer y escribir fácilmente la información y usar una variedad de lenguajes de programación como Java, C++, o Python.
- SSTable(Sorted Strings Table): Es un archivo ordenado de parejas clave-valor. Los bloques de información contienen un índice añadido. Hay que señalar que en este caso el índice no es un rango de bloques, no un valor de bloques .



## Conclusiones:

En años anteriores no había mucho beneficio intentando predicar cómo trabaja Google internamente, pero ahora la gente está más interesada en la eficiencia energética y Google considera que es importante compartir algo de su experiencia. Con todas estas mejoras que hemos visto, Google trata de desarrollar tecnologías que ahorren energía a la hora de procesar y servir la información de forma mundial. Además se sigue tratando de buscar estándares de alimentación que sean más sostenibles.

Todas estas mejoras proporcionan un ahorro económico importante y un mejor servicio para los usuarios.

### Bibliografía:

- [1]<http://www.cnet.com/news/google-uncloaks-once-secret-server-10209580/>
- [2]<http://www.guatemewireless.org/hardware/el-hardware-de-los-servidores-de-google.html>
- [3]<http://google.dirson.com/post/4337-servidores-web-datacenters/>
- [4]<http://hipertextual.com/2009/04/revelan-datos-de-los-servidores-de-google>
- [5]<http://www.datacenterknowledge.com/archives/2014/02/03/google-spent-7-3-billion-data-centers-2013/>
- [6][http://www.theregister.co.uk/2010/01/29/google\\_web\\_server/](http://www.theregister.co.uk/2010/01/29/google_web_server/)
- [7]<https://www.cs.umd.edu/class/spring2011/cmsc818k/Lectures/gfs-hdfs.pdf>
- [8]<http://www.google.com/about/datacenters/inside/data-security/index.html>
- [9]<http://www.google.com/about/datacenters/inside/locations/index.html>
- [10]<http://dejanseo.com.au/chubby-lock-service/>
- [11]<https://developers.google.com/protocol-buffers/>

- [12]<http://www.slideshare.net/tomcythankachan1/google-bigtable-15370454>
- [13]<http://www.abc.es/tecnologia/informatica-hardware/20141128/abci-google-cuchara-inteligente-201411281113.html>
- [14]<http://www.liftware.com/>
- [15][http://www.bostondynamics.com/robot\\_bigdog.html](http://www.bostondynamics.com/robot_bigdog.html)
- [16]<http://www.ticbeat.com/tecnologias/amplio-catalogo-hardware-google/>
- [17]<http://www.wired.com/2012/10/google-search-appliance/>
- [18]<http://www.enter.co/eventos/googleio/2014/tango-ara-sdc-los-tres-ases-de-google-que-cambiaran-el-mundo/>

Páginas auxiliares por si queremos usar: Posibles enlaces software:  
[http://es.wikipedia.org/wiki/Categor%C3%ADa:Software de Google](http://es.wikipedia.org/wiki/Categor%C3%ADa:Software_de_Google)  
<http://google.software.informer.com/>

Posibles enlaces hardware:  
<http://www.google.com/about/datacenters/>

Desarrollos de google:  
<http://www.abc.es/tecnologia/informatica-hardware/20141128/abci-google-cuchara-inteligente-201411281113.html>  
<http://www.ticbeat.com/tecnologias/amplio-catalogo-hardware-google/>

Otros enlaces interesantes:  
<https://prezi.com/akvy29uydav0/servidores-de-google/> (no se hasta qué punto es fiable)  
<https://www.youtube.com/watch?v=TjD1HixyKL0>  
<http://www.datacenterknowledge.com/archives/2015/05/07/google-launches-bigtable-cloud-nosql-service/> (noticia 2014)