

Centros de datos de Google

Los **Centros de datos de Google** (o *Google Data Center*) son instalaciones especialmente creadas por Google para el almacenamiento y gestión de sus servidores. Constan de distintos elementos electrónicos tales como sistemas de almacenamiento, dispositivos de comunicación, elementos de climatización y dispositivos de seguridad.

En este apartado se especificaran detalles referentes a Google. En el 2013 Google tenía 13 centros de datos en los que, según las estimaciones realizadas sobre los últimos datos energéticos emitidos se calcula que habría un total de 900,000 servidores.

Los servidores que se usan en los centros de datos son montados a medida por Google y funcionan bajo una distribución de Linux también personalizada por Google.

Ventajas

Distribuir los servidores de forma des-localizada en vez de en una única estructura tiene una serie de ventajas.

Rapidez de acceso, la distancia al punto de conexión con un servidor es uno de los factores que influyen en la velocidad con la que obtendremos los datos, cuanto más cerca estemos del centro de datos correspondiente, más rápida será nuestra búsqueda.

Fiabilidad del sistema, tener varios centros de datos en vez de uno solo de un tamaño mucho mayor, protege al sistema frente a caídas inesperadas, provocadas desde fallos en servidores, cortes de corriente, o incluso catástrofes naturales.

Coste reducido, construir un único centro de datos centralizado y de un tamaño desmesurado es mucho más caro que construir varios más pequeños.

Distribución en distintas franjas climáticas, los centros de datos están distribuidos a lo largo de diversas franjas climáticas con diferente estacionalidad, esto permite compensar altas temperaturas en un área, desviando el trabajo a las zonas más frías en esa época, lo cual repercute directamente en el gasto energético.

Distribución en distintas franjas horarias, existe también la posibilidad de desplazar carga hacia los centros de datos que puedan aprovechar tarifas más bajas de electricidad, como por ejemplo durante horas nocturnas de bajo consumo.

Flujo de datos

Google tiene esencialmente dos grandes redes separadas, la primera conecta los usuarios a los servicios ofrecidos por Google (Búsqueda, Gmail, YouTube, etc.) la segunda conecta los centros de datos entre sí. El sentido de esta separación de redes en vez de la integración de ambas en una sola red es debido a las diferentes características de demanda de tráfico que se aprecian entre lo que los consumidores necesitan y las necesidades de Google.

La red de usuarios tiene un tráfico constante aunque diferenciado en dos tramos, nocturno y diurno, el segundo con una mayor carga. Los niveles de exigencia así mismo son muy altos, todos los paquetes han de llegar intactos y en el menor tiempo posible. A pesar de los tramos mencionados, el tráfico en esta red es constante, el volumen dentro de uno de los tramos no varía excesivamente de forma imprevisible ya que depende del número de usuarios de la zona que es relativamente constante.

Sin embargo el tráfico interno es completamente cambiante frente a la regularidad del externo. Google tiene que mover petabytes de datos (índices de la WWW, backups, o incluso los datos derivados del lanzamiento de un nuevo producto) y necesita saber cuánto tiempo tardara el proceso y poder priorizar el tráfico más importante, lo cual es imposible con el sistema tradicional de transporte de datos.

Para ello Google usa OpenFlow, un sistema *Open Source* desarrollado por científicos de las Universidades de Stanford y Berkley. OpenFlow se basa en la llamada Red Definida por Software (SDN, *Software Defined Networking*) que consiste en separar las funciones de direccionamiento de la red, aislándolas en servidores que conocen el estado completo de la red, en vez de dejándolas a decisión de los *router* como en el sistema tradicional.

Tomando como ejemplo una compañía de taxis, el sistema tradicional sería en el que el taxista decide la ruta, sin saber si la calle por la que decide ir estará cortada por obras o congestionada por el tráfico. En el sistema Open Flow, el taxista preguntara su ruta a una central que conoce el estado exacto de tráfico en cada una de las calles, pudiendo tomar la decisión óptima. Esta central de tráfico podría incluso ordenar a los taxis (datos con poca importancia) permanecer en el garaje mientras pasan los camiones de bomberos (datos con mucha importancia).

Seguridad del sistema y protección de los datos

Google posee una importante infraestructura que no solo contiene el valor físico del hardware que la compone, sino también el valor de los datos que almacena. Algunos de esos datos son propios de la empresa (estadísticas, índices, etc.) mientras que otros son datos privados de los usuarios (correos de Gmail, videos de YouTube, etc.).

El funcionamiento de todos los servicios depende de este hardware y los datos que almacena, por tanto la gestión de la seguridad en estos centros de datos es una de las prioridades de la empresa.

Tal y como Google cuenta en uno de sus videos su política de seguridad se basa en tres áreas principales:

Física

Dentro de un centro de datos hay miles de máquinas que ejercen de servidores para los servicios ofrecidos por Google, contienen los datos privados de millones de usuarios con un valor económico incalculable, que además están protegidos bajo las leyes de protección de datos por ser material privado. El acceso está altamente controlado, no se permiten turistas o visitas, el perímetro está vallado y fuertemente vigilado, incluso los trabajadores de Google tienen que ser perfectamente identificados en los puntos de acceso, en los cuales hay personal de seguridad 24 horas al día que disponen de una red de vídeo cámaras que monitorizan el acceso de todos los vehículos.

Una vez dentro del recinto, al acceder al edificio el personal de seguridad tiene que verificar de nuevo su identidad, no solo mediante las tarjetas electrónicas de identificación sino mediante análisis biométricos tales como el análisis de retina. Ya en las oficinas, un último control se impone antes de pasar a las salas de servidores (donde se guardan los datos).

Además del personal en los puntos de acceso, una sala de seguridad se encarga de la revisión de las cámaras de la instalación, un equipo de seguridad está presente 24 horas al día y actuaría en caso de alguna anomalía.

Protección de datos

Google se toma en serio la protección de los datos, que se guardan en múltiples localizaciones para asegurar su disponibilidad. Además Google tiene en cada uno de sus centros un equipo automático de vigilancia anti incendios, en caso de detectarse alguna anomalía en alguna de las secciones, automáticamente la actividad hace un puente a esa sección. Estos datos son críticos para el funcionamiento de la empresa, por tanto se guardan no solo con las medidas de seguridad físicas, sino además mediante algoritmos de encriptación, los datos no están guardados en texto plano sino cifrados

Los datos se guardan en discos duros que siguen un ciclo de vida, se monitoriza tanto su estado como su localización para evitar la pérdida o sustracción de alguno de ellos, cuando uno empieza a fallar, es reformateado y si supera una serie de pruebas, puesto de nuevo a funcionar.

En caso de no superar la prueba dada la cantidad de datos que contiene y la posible importancia de ellos pasa a un proceso para eliminar completamente la información que pudiera contener. Primero los datos serán sobrescritos y seguidamente los discos serán destruidos físicamente, una máquina aplastará y deformará el disco para que su contenido no pueda ser leído y seguidamente el equipo triturador que convertirá el disco duro en "viruta" metálica, imposible de recuperar.

Fiabilidad en las operaciones

Google mantiene una serie de generadores de emergencia, así como duplicidad en las conexiones tanto a la red eléctrica como a la red de datos, para evitar errores debidos a fallo en una conexión, todo ello destinado que sus servicios estén disponibles 24 horas al día, 365 días al año.

Eficiencia energética y sostenibilidad

La mayoría de los centros de datos utiliza la misma cantidad de energía general o de no computación (conversión de energía, refrigeración, etc) que la utilizada para suministrar energía a sus servidores.

En Google se ha reducido esta energía general a un 12 %. De esta forma, el 88 % de la energía que utiliza va dirigida a los equipos que suministran los productos y las búsquedas de Google.

El promedio ponderado de energía del PUE (siglas en inglés de *Power Usage Effectiveness*, Eficiencia en el Uso de Energía) de los últimos doce meses para todos los centros de datos de Google es de solo 1,12, lo que hace de estos centros de datos los más eficientes del mundo ya que según la encuesta sobre centros de datos de 2012 del Uptime Institute, el promedio global de los mayores centros de datos encuestados se encuentra entre el 1,8 y el 1,89.

Generalmente, todos los grandes servidores de otras empresas se suelen concentrar en una sala denominada sala fría, con una temperatura de unos 22 o 23 grados, una forma de mejorar la eficiencia energética de los servidores es tan simple como elevar la temperatura, en general las zonas de servidores de Google funcionan con una temperatura de 27 grados.

Sin embargo elevar la temperatura únicamente no es suficiente, la manera utilizada por Google para conseguir este rendimiento energético es el uso del entorno en el que emplazan sus centros de datos. En marzo de 2009, comenzaron la transformación de la fábrica de papel Summa Mill, de 60 años de antigüedad, para convertirla en un centro de datos. la actualidad, el centro de datos presta sus servicios a usuarios de Google de toda Europa y del resto del mundo.

La particularidad de esta planta es que utiliza el agua marina procedente de la bahía de Finlandia, aprovechando directamente la baja temperatura en la que se encuentra para enfriar sus equipos, una vez utilizada se mezcla con agua cogida nuevamente de la bahía para suavizar su temperatura y así devolverla con una temperatura parecida, minimizando el impacto ecológico.

Otro ejemplo de cómo Google usa nuevas formas para refrigerar sus instalaciones, es la ventilación por evaporación. Frente al uso de aire acondicionado, que puede aumentar entre un 30 % y un 70 % el consumo de energía, Google usa un sistema parecido al que usa nuestro propio organismo para refrigerarse.

Unos sistemas de refrigeración personalizados almacenan temporalmente el aire caliente que sale de los servidores y lo aíslan del resto del espacio del centro de datos. Los ventiladores situados sobre cada una de estas unidades extraen el aire caliente de la parte trasera de los servidores a través de conductos refrigerados con agua, que absorbe el calor del aire.

Esta agua caliente procedente del centro de datos baja a través de un material que acelera la evaporación y se convierte en parte en vapor, que es sacado al exterior mediante ventiladores. Finalmente se envía el agua refrigerada de vuelta al centro de datos para volver a empezar el ciclo.

Localización

La localización de los Google Data Center construidos en 2013 es la siguiente:

Estados Unidos:

- Berkeley County, Carolina del Norte
- Council Bluffs, Iowa (<https://osm.org/go/T87mBI9--?relation=128357>)
- Douglas County, Georgia (<https://osm.org/go/ZF~eZKw--?relation=1059843>)
- Mayes County, Oklahoma (<https://osm.org/go/T6VL09g-?relation=1822225>)
- Lenoir, Carolina del Norte (<https://osm.org/go/ZS0APRL--?relation=179912>)
- The Dalles, Oregón (<https://osm.org/go/WILzdaE?relation=186362>)

Sur América: