

## **Google File System (GFS)**

Un sistema de archivos es una estructura de datos dentro de una unidad lógica que permite al sistema operativo almacenar información de forma organizada e independiente de los procesos que la utilizan. El sistema de archivos se encarga de abstraer propiedades físicas de diferentes dispositivos. La información, al ser independiente de los procesos, se mantiene tras su finalización y, además puede ser utilizada por varios procesos. Los sistemas de archivos son usados en dispositivos de almacenamiento secundario. Existen dos tipos de sistemas de archivos. Distribuidos y remotos.

Un Sistema de Archivos Distribuido (DFS) es un servicio que permite al usuario acceder y procesar archivos guardados en un servidor como si fuese local.

Un Sistema de Archivos Remotos (RFS) son caso particular de la abstracción de sistemas no-físicos, si bien el sistema operativo no tiene acceso directo a los archivos y directorios que le solicitará el usuario, por medio de los módulos de red, sabe cómo obtenerlos y presentarlos como si fueran locales.

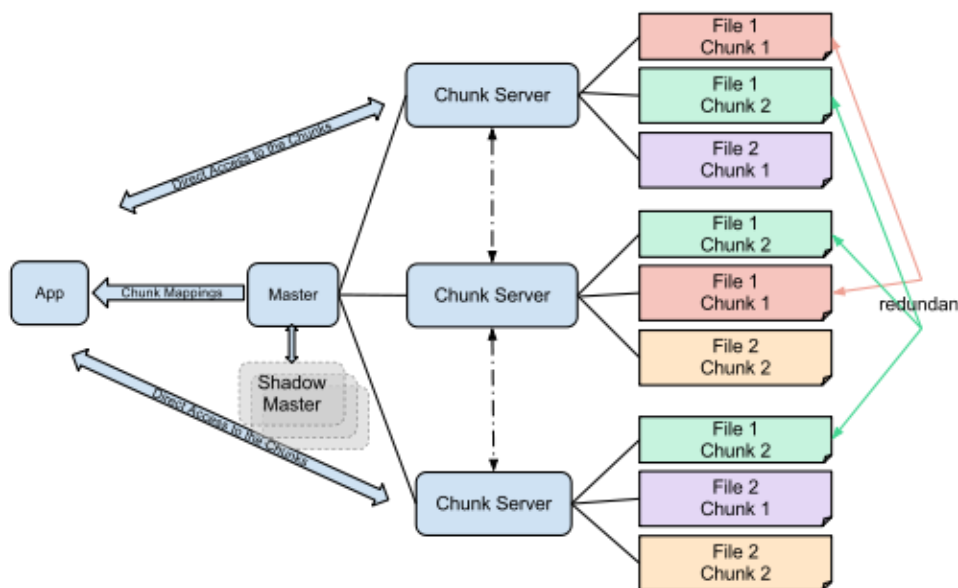
El Sistema de Archivos Google, en inglés Google File System (GFS, GooFS o GoogleFS), es un sistema de archivos distribuido propietario desarrollado por Google Inc., que soporta toda su infraestructura informática de procesamiento de información en nube. Está especialmente diseñado para proveer eficiencia, fiabilidad de acceso a datos usando sistemas masivos de clúster de procesamiento en paralelo. La actual versión de Google File System tiene el nombre clave Colossus.

Se puede mencionar que la estructura del sistema de archivos de Google está separada en dos partes importantes: nodos maestro y nodos Chunk.

El diseño de este sistema de archivos fue creado para manipular consultas de almacenamiento de datos, generando enormes cantidades de información en cada consulta, la cual deberá de servir en el futuro para contestar otras consultas, optimizando el manejo de información.

Los bloques que utiliza GFS son similares a los de un disco duro normal pero las diferencias consisten en que tienen un gran tamaño de 64 MB, rara vez se modifican, casi nunca se reducen de tamaño y normalmente, se adiciona información o simplemente se leen para su manipulación.

Estos bloques de datos están separados en nodos, estos nodos están en servidores los cuales a su vez están compuestos por computadoras de bajo costo con algunas de las siguientes especificaciones: 1.4 GHz de velocidad de procesador, 2 GB de memoria RAM, dos discos de 80 GB 5400 rpm y corren bajo el sistema operativo LINUX.



La decisión de los fundadores de Google de utilizar máquinas de bajo costo y software libre, fue porque en los inicios de su fundación no disponían del

dinero suficiente para comprar servidores y programas de pago, pero si disponían de creatividad e inteligencia para fundar la empresa aun cuando eran estudiantes de la universidad de Stanford, Estados Unidos.

El diseño de este sistema, esta creado para recuperarse de fallos sin ningún problema contando con distintas replicas en distintos lugares. Otra característica es la

utilización de archivos que almacenan una gran cantidad de información que satisfacen los requerimientos que las consultas en internet necesitan. Otro punto importante es la necesidad en el sistema de la velocidad para producir una gran fluidez en el procesamiento de información, esta característica es llamada baja latencia.

GFS funciona gracias a los Clusters basados en Linux, estos consisten en un nodo maestro y varios nodos Chunk. Los archivos de los nodos chunk se dividen en bloques de tamaño fijo que tienen asignados una etiqueta de 64 bits en el nodo maestro y están replicados en otros servidores de acuerdo a su demanda.

La función del nodo maestro es la manipulación de metadatos creados por cada bloque de los nodos chunk, es decir que su función es la de la localización de la información.

## **Bibliografía:**

- <https://ai.google/research/pubs/pub51>
- <https://www.fing.edu.uy/inco/cursos/sistoper/recursosEnlaces/presentacion-gfs.pdf>
- Fundamentos de sistemas operativos / GunnarWolf [y tres más]. – Primera edición. – México D.F. : Universidad Nacional Autónoma de México, Instituto de Investigaciones Económicas : Facultad de Ingeniería, 2015.  
367 p. : ilustraciones ; 28 cm.
- <https://www.elconspirador.com/2013/12/25/google-file-system/>