

Google File System.

Por: Sánchez Espinosa Andrew Blaise.

Sistemas Operativos.

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 (No confundir con GMail File System).

El GFS es un sistema de archivos distribuido en este caso porque presenta una interacción entre dos sistemas (de archivos) o sistema a sistema, además de conectar en red varias computadoras juntas y aprovechar sus recursos individuales de forma colectiva sin necesidad de estar en un mismo punto geográfico.

El GooFS es un sistema de archivos que está optimizado por Google para el almacenamiento de datos básicos y sus necesidades de uso (sobre todo el motor de búsqueda de Google), y puede generar enormes cantidades de datos que deben ser mantenidas para optimizar la siguiente respuesta; El actual sistema de archivos surgió como una mejora a su BigFiles, desarrollado por Larry Page y Sergey Brin en los inicios de Google, cuando estudiaban en Stanford. Los archivos son divididos en porciones de tamaño fijo de 64 megabytes, similar a los cluster o sectores de las unidades de disco duro tradicional, donde muy rara vez son sobrescritos, o reducidos, por lo general los archivos se adicionan o se leen.

También está diseñado y optimizado para funcionar con los clusteres de servidores de Google, nodos de alta concurrencia formado por computadoras de bajo coste, donde deben tomarse precauciones contra un alto índice de fallos por sobrecarga en los nodos individuales y por ende la probable pérdida de algunos datos. Otros puntos en el diseño apuntan a manejar una gran caudal de datos, e incluso resolución de problemas de latencia.

Google File System está optimizado para el almacenamiento de datos básicos y sus necesidades de uso, esto puede generar enormes cantidades de datos que deben ser mantenidas para optimizar la siguiente respuesta. También está diseñado y optimizado para funcionar con los clústeres de los servidores de Google, nodos de alta concurrencia formada por computadoras de bajo coste, donde deben tomarse precauciones contra un alto índice de fallos por sobrecarga en los nodos individuales y por ende la probable pérdida de algunos datos. Otros puntos en el diseño apuntan a manejar un gran caudal de datos, e incluso resolución de problemas de latencia.

La principal diferencia entre los demás sistemas de archivos, es que el Google File System no está implementado en el kernel del sistema operativo, sino que funciona como una librería en el espacio de usuario (userspace). Google usa su Google File System para organizar y manipular archivos enormes y para permitir a los desarrolladores de aplicaciones los recursos de investigación y desarrollo que requieren. El Google File System es exclusivo de Google y no está a la venta. Pero podría servir como modelo para sistemas de archivos para organizaciones con necesidades similares.

En cuanto a las operaciones o comandos que se pueden realizar en el Google File System resultan ser las operaciones básicas de archivos: creación, apertura, lectura, escritura y cierre de archivos, también en el desarrollo de este sistema de archivos se incluyeron algunos comandos especiales que son “snapshots” que permiten copiar rápidamente el contenido de una computadora y append que permite agregar información a n archivo existente sin sobrescribir los datos previamente escritos. El Google File System divide los archivos en fragmentos de 64 M Bytes (como ya se había mencionado) para hacer más eficiente el ancho de banda del sistema.

Para saber más:

<https://queue.acm.org/detail.cfm?id=1594206>

<http://www.baselinemag.com/c/a/Infrastructure/How-Google-Works-1>