

**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE INGENIERÍA

DIVISIÓN DE INGENIERÍA ELÉCTRICA

- **Sistemas Operativos**

- **Tarea 3**

**Sistemas de archivos remotos:
Google File System**

- **Profesor**

Wolf Iszaevich Gunnar Eyal

- **Grupo: 01**

- **Guadarrama Flores Edgar Alejandro.**

SEMESTRE: 2019-1



- **¿Es remoto o distribuido? ¿Por qué sienten que sea así?**

El sistema de archivos de google es un sistema distribuido porque permite compartir archivos, impresoras y otros recursos como un almacenamiento persistente en una red de computadoras. Lo cual posibilita a un usuario acceder, modificar, alterar, crear archivos guardados en un servidor (google) como si fuera almacenamiento local. Todo el procesamiento de información se hace desde la nube.

- **¿Cuáles son sus características principales? ¿Para qué tipo de uso está diseñado? ¿Para qué tipo de uso *no* es recomendado?**

En un principio las computadoras que forman los cluster son máquinas Linux que tenían un bajo costo (procesador xeon doble nucleo, 2Gb de ram y 800Gb de almacenamiento), esto se debía que en un principio los creadores de este sistema no tenían muchos recursos para comprar máquinas poderosas entonces utilizaban el software libre y equipos baratos.

Los archivos se organizan de manera jerárquica en directorios e identificados por rutas de acceso

La arquitectura sobre la que está basada en dos partes importantes: un nodo Master y múltiples nodos chunkservers.

Chunkservers. Su función es almacenar los archivos del sistema y enviar actualizaciones de su estado a los nodos maestro. El tamaño de los chunks tamaño es de 64MB¹. y enviar actualizaciones de su estado a los nodos maestro.

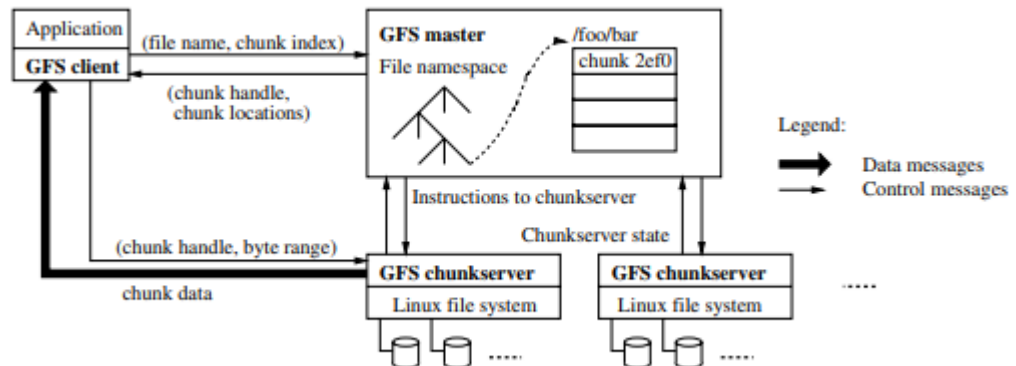
Master. Son los nodos que tienen mayor carga de trabajo. Se encargan de guardar los metadatos de cada chunk y administra los permisos de escritura y lectura de a cada uno de ellos, también guardan registro de los cambios críticos hechos a los archivos y realizan copias de los archivos que no han sido replicados.

Los metadatos contienen etiquetas de identificación de cada chunk los cuales alimentan tablas en el nodo maestro para dar seguimiento a su ubicación.

Los nodos maestros vigilan que aplicaciones quieren acceder a los chunks, cuando reciben una solicitud nueva responden con la ubicación del chunk pedido si este no está siendo utilizado por otra aplicación. Estos permisos se dan para cada aplicación por un tiempo determinado, cuando el proceso termina o se le agota el tiempo el chunkserver propaga los cambios a los servidores y los cambios no se hacen hasta que todos los servidores se enteran del cambio.

¹ El tamaño de los chunks ayuda a responder la última pregunta

Diagrama de la arquitectura (Obtenida de *The google File System*, Sanjav Ghemanwat.



¿Para qué sistema operativo está diseñado? ¿Saben si está implementado en algún otro sistema?

GoogleFS trabaja en el sistema operativo Unix por las razones de optimización de costos de los creadores y por uso de software libre.

- **¿Cómo es su *modelo de fallos* o *modelo de consistencia*? (para una mejor explicación de a qué me refiero con esto, revisen la definición de AFS — Más allá de los primeros dos párrafos que presentan al sistema, casi todo el resto es referente a su modelo de consistencia).**

El tamaño de los chunk no es al azar, este tamaño tiene como objetivo reducir los metadatos asociados a ellos y así evitar que los nodos se sobrecarguen individualmente, lo cual generaría problemas si varios clientes quieren acceder al mismo tiempo al mismo 'chunk' lo cual desembocaría en un cuello de botella y pérdida de información.

Los archivos se guardan tres veces lo cual permite tener respaldos de información mientras se facilita el acceso a un archivo que sea muy demandado, lo cual permite al sistema recuperarse de fallos.

Referencias

<https://static.googleusercontent.com/media/research.google.com/es/archive/gfs-sosp2003.pdf>
<https://computer.howstuffworks.com/internet/basics/google-file-system.htm>