

El sistema de archivos de Google

Sánchez Hernández Marco Antonio

Teran García Rodolfo Mario

Importancia del sistema de archivos de Google

El sistema de archivos de Google no solo es importante para los servicios de cómputo en la nube ofrecidos por Google, sino que también los propios servicios de Google como YouTube, Drive y Gmail hacen uso de este.



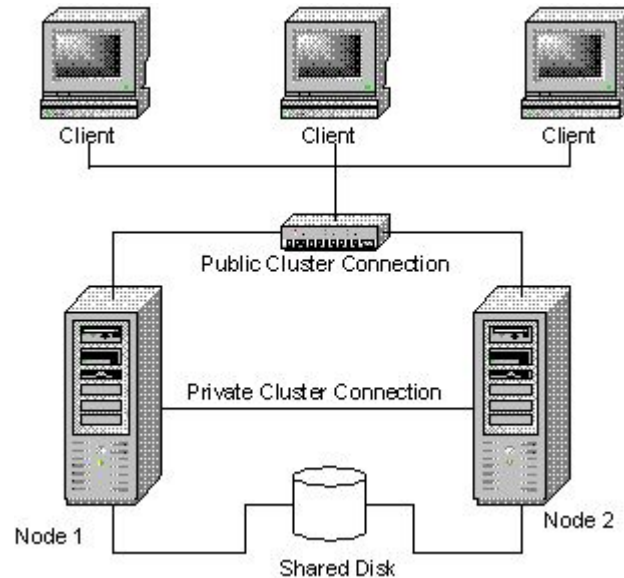
Componentes principales de los servicios de almacenamiento de Google

- **Sistema de archivos** (Google File System, Colossus)
- **Base de datos relacional** (Spanner)
- **Job Scheduler** (Borg)

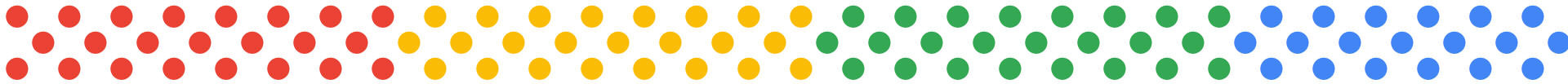
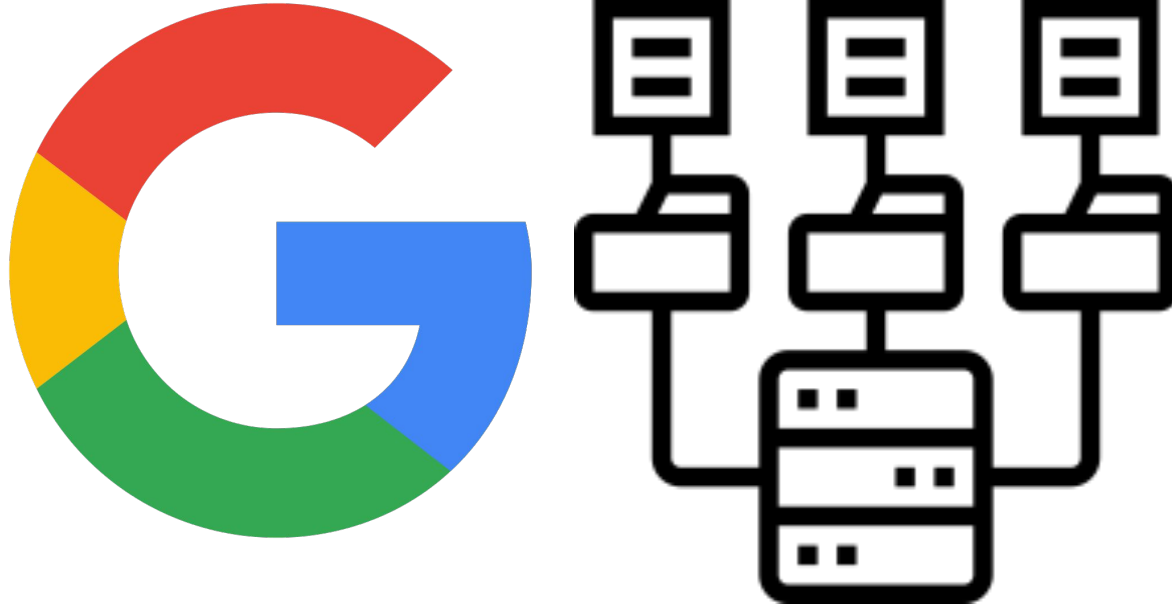


Clustered File System

Sistema de archivos distribuido en diferentes servidores

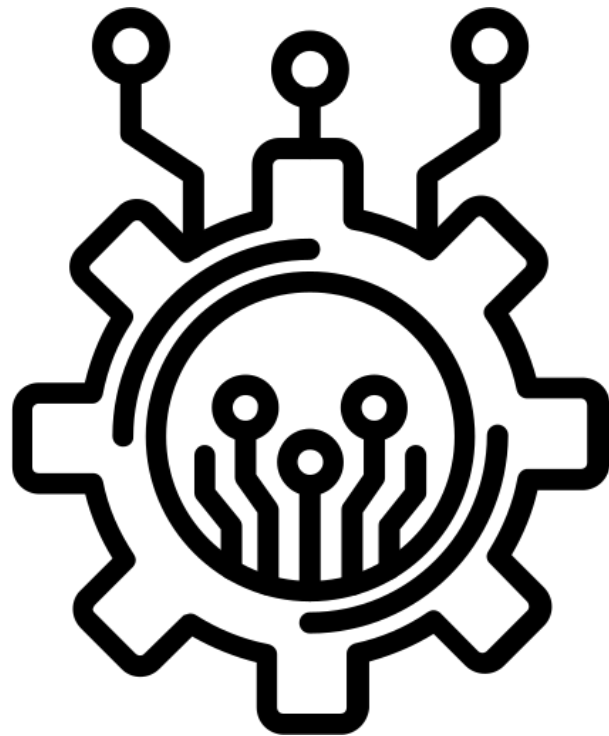


Google File System



Entorno

- Muchos componentes
- Gestión de archivos grandes (GB)
- Dos tipos de lectura:
 - Grandes lecturas de streaming
 - Pequeñas lecturas aleatorias
- Semántica bien definida
- Ancho de banda $>$ latencia



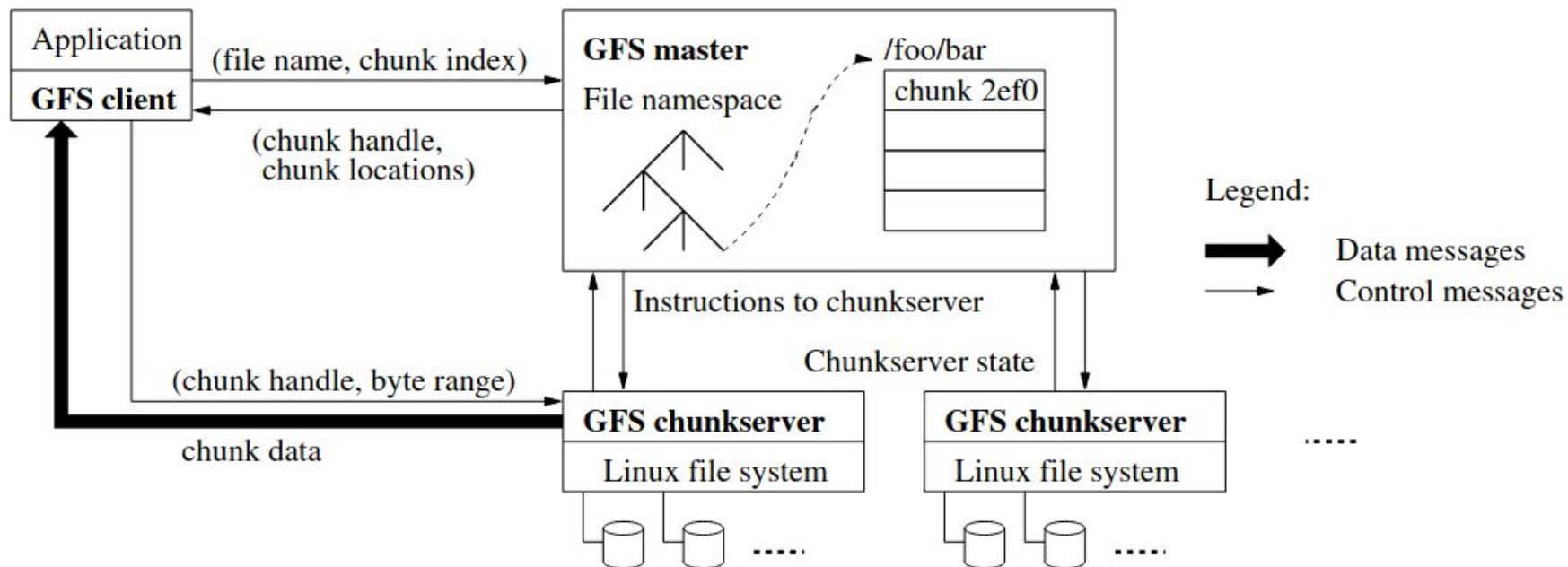
Interfaz



- Organización en directorios
- Operaciones habituales
 - delete, open, close, read, write
- Operaciones adicionales
 - snapshot
 - record append

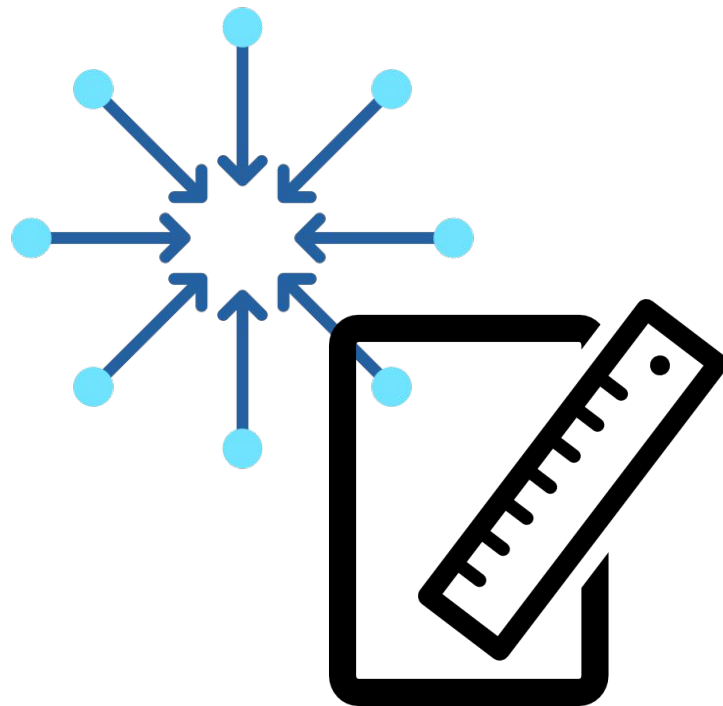


Arquitectura

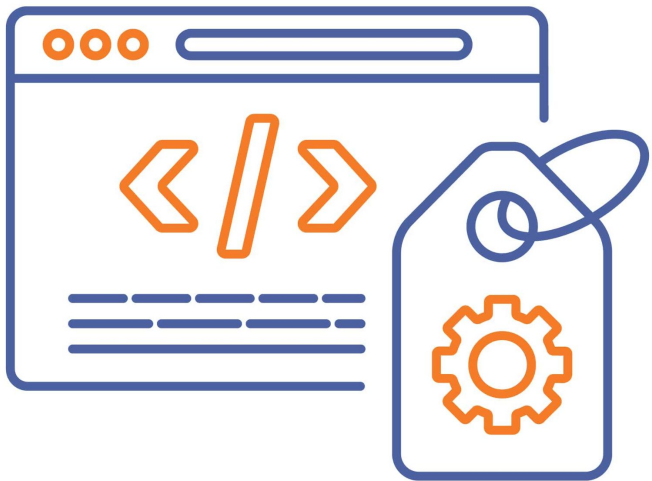


Master unico & tamaño del chunk

- Simplifica el diseño
- Se debe evitar convertirse en un cuello de botella
- Tamaño fijo de 64 MB
 - Reduce la necesidad de interacción con el master y la optimiza
 - Minimiza la saturación de la red
 - Menor cantidad de metadatos



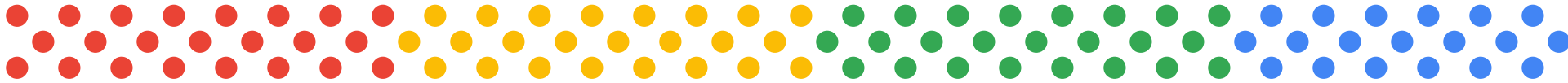
Metadatos



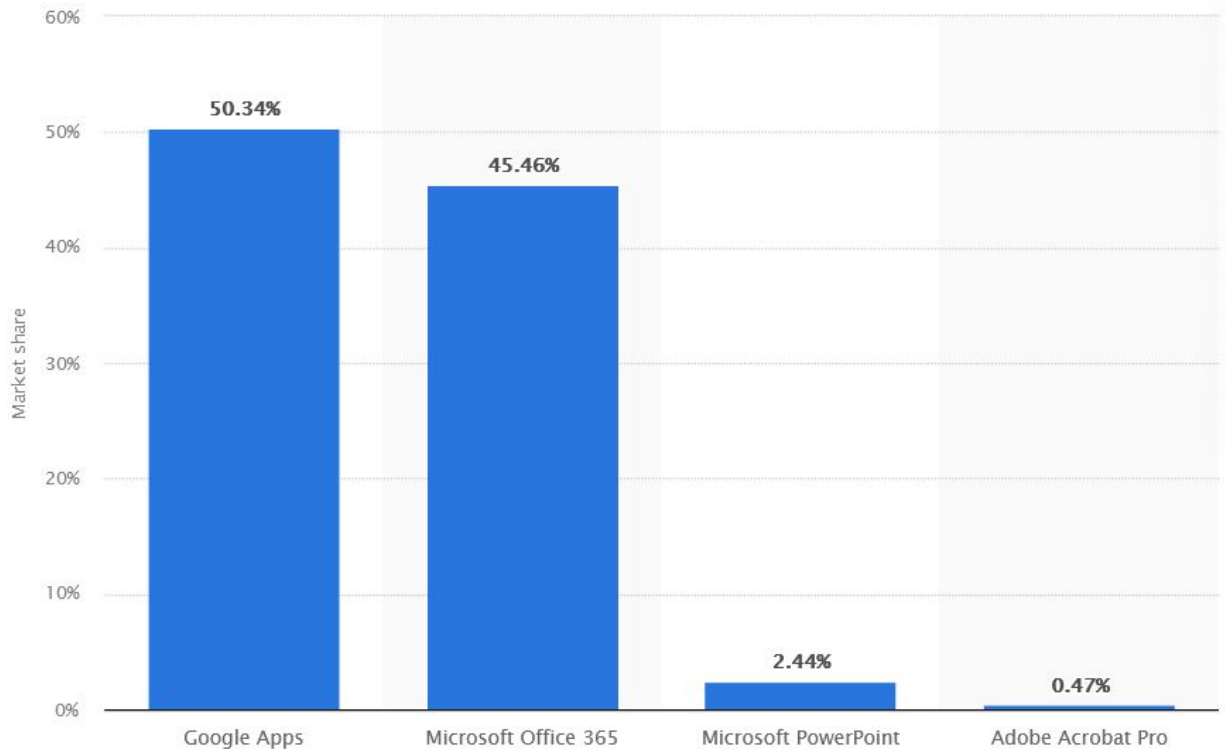
- Tres tipos
 - Espacios de nombres de archivos y chunks
 - Asignación de archivos y chunks
 - Ubicaciones de las réplicas de cada chunk
- No se almacena la ubicación de los chunks de manera persistente
- Estructura de datos en memoria
- Registro de operaciones



¿Por qué Google necesita un nuevo sistema de archivos?



Market share of major office productivity software worldwide in 2022



Bases de datos no relacionales NoSQL



La información no se almacena de forma tabular.

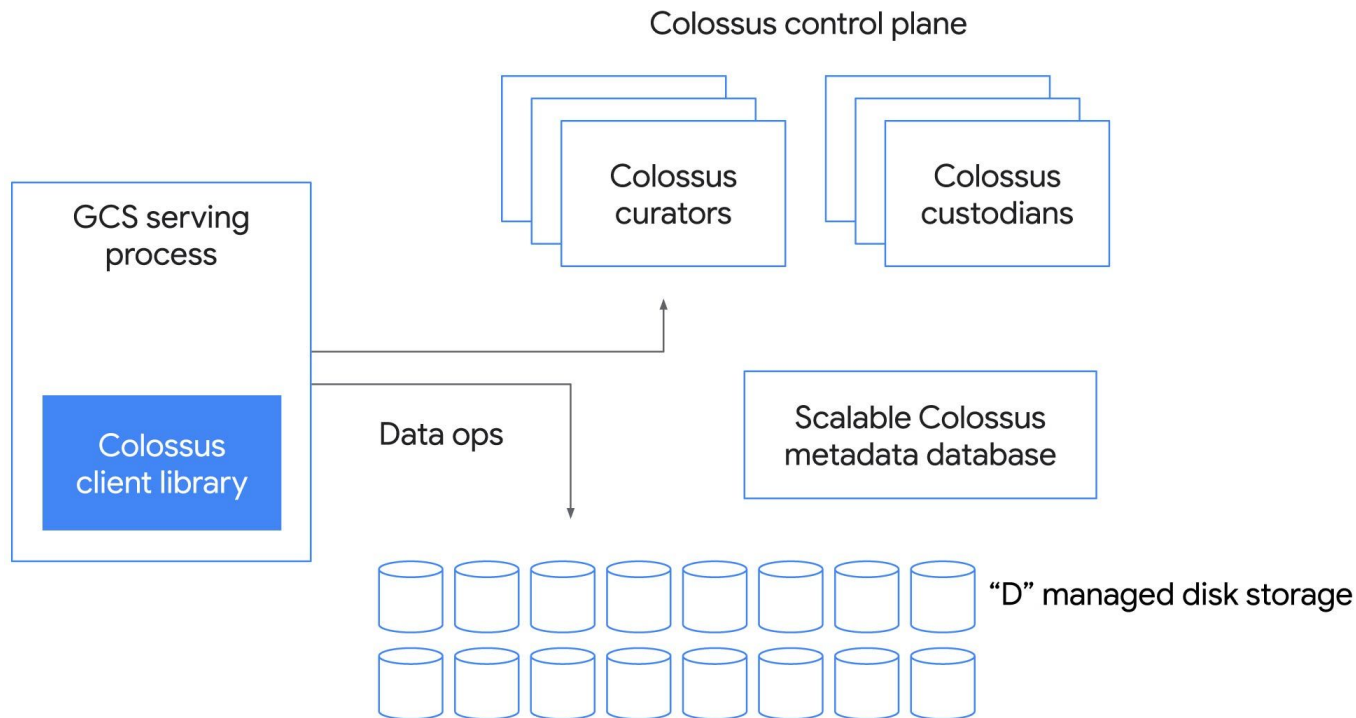
La información puede ser almacenada con una estructura similar a un documento.

Los documentos pueden contener una variedad de formatos y tipos de datos.

Permiten una mayor flexibilidad con respecto a las bases de datos relacionales debido a su capacidad de almacenar datos con estructuras diferentes.



Funcionamiento de Colossus (A alto nivel)



Typical cluster

