

Sistemas de archivos remotos: Ceph y Gluster

Estamos en una época donde las necesidades de almacenamiento crecen exponencialmente, simplemente un usuario común necesita cada vez más espacio en algo tan esencial como su smartphone. Un profesional audiovisual no se da abasto con los gigas y gigas de vídeos y material que puede generar diariamente y es cuando el modelo de almacenamiento en la nube es una fuerte necesidad. Ahora bien, lo anterior palidece frente a las necesidades de almacenamiento que requieren miles de compañías: desde bases de datos de empleados, clientes y transacciones (como un banco), hasta miles de horas de contenido para streaming (como Netflix y YouTube por ejemplo).

Ante ello, existen Ceph y Gluster de Red Hat, que son arquitecturas de almacenamiento denominadas SDS (Almacenamiento Definido por Software por sus siglas en inglés) de código abierto diseñadas para almacenar grandes cantidades de datos.

Gluster es muy útil para almacenamiento de objetos gran escala (terabytes y petabytes) y distribuye datos entre computadoras conectadas, pero lo hace en bloques (de un tamaño predeterminado de 12KB). Una característica importante que tiene es que utiliza protocolos tradicionales de transferencia y almacenamiento, sin mayor riesgo de seguridad o de pérdida de datos (caso contrario a Ceph).

Ceph está más enfocado a almacenamiento rápido y a corto plazo de bloques pequeños (64K es el predeterminado) reduciendo los cuellos de botella en los accesos de almacenamiento de datos en objetos binarios distribuidos en muchas computadoras, creando una nube privada y los datos se pueden combinar.

Su ventaja es que al estar distribuido normalmente entre más nodos que un sistema Gluster, en caso de alguna falla de disco se puede recuperar la información de forma más veloz porque permite que varias unidades ingresen los datos requeridos en paralelo.

Para ambos casos, el usuario interactúa con sus datos con un sistema de archivos común en una interfaz tipo NFS o POIX, con la posibilidad de hacer búsquedas y recuperar datos.

Ambos son sistemas de archivos distribuidos, ya que cada nodo guarda una parte del total de los datos y todos trabajan de forma conjunta, solicitando o enviando datos.

Finalmente cada uno está destinado a públicos distintos. Gluster es ideal para empresas donde sus usuarios requieren consultar y modificar constantemente la información de manera rápida y actualizada, mientras que Ceph le sirve más a empresas que solamente requieren mover grandes cantidades de información de un lugar a otro (por ejemplo big data). Ambas opciones por ser de código abierto suelen implicar bajos costos.

Fuentes:

<https://technologyadvice.com/blog/information-technology/ceph-vs-gluster/>

<https://www.networkcomputing.com/storage/gluster-vs-ceph-open-source-storage-goes-head-head/8824853>