

GlusterFS

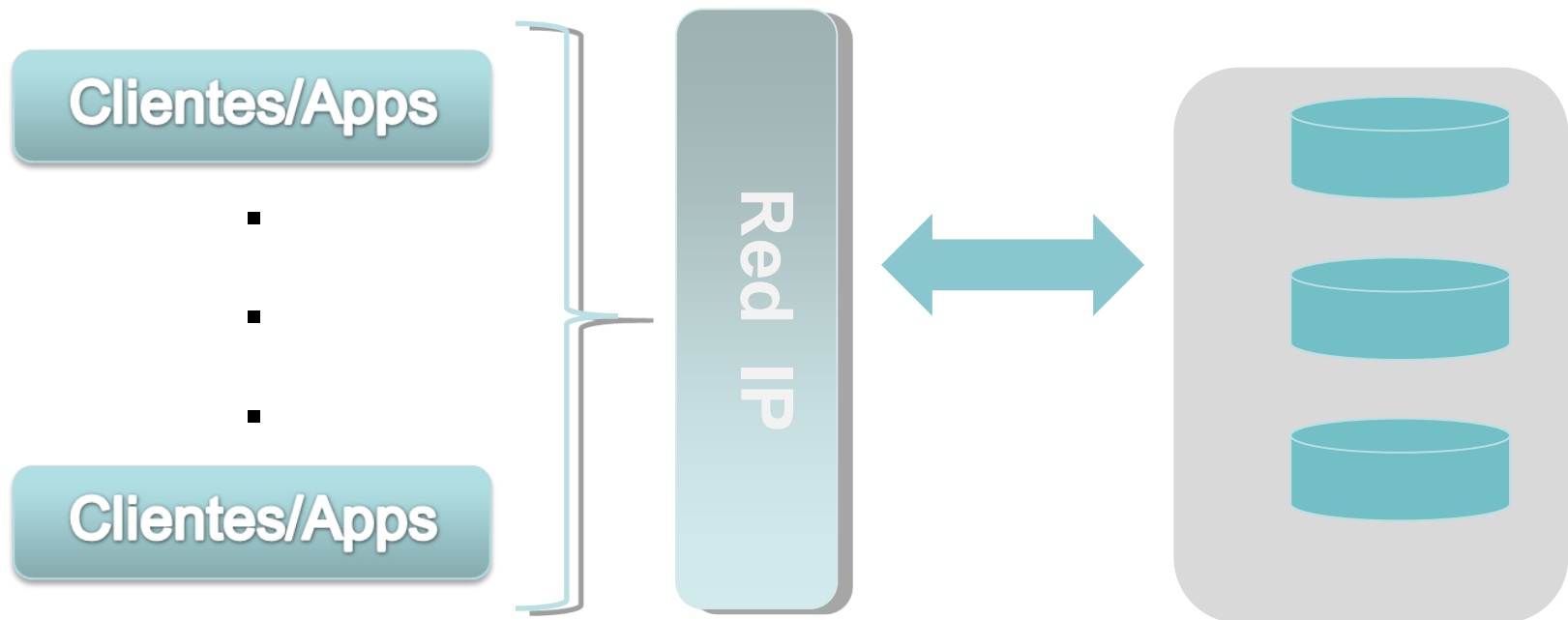
Una visión rápida a uno de los
más innovadores sistema de
archivos distribuido

¿Qué es GlusterFS?

- Es un sistema de archivos de alta disponibilidad y escalabilidad que puede brindar almacenamiento a gran escala (petabytes) a bajo costo (opensource) y manejo de hasta miles de clientes.

¿Qué es GlusterFS?

- GlusterFS agrupa dispositivos de almacenamiento a través de la red y maneja la data como si fuese un solo bloque



Ventajas de GlusterFS

- Flexibilidad para combinar diversos dispositivos, ya sean físicos virtuales o disco en la nube
- Muy fácil de instalar y configurar
- Diseño modular y apilable
- Escalabilidad lineal

Ventajas de GlusterFS

- Completamente implementado en user-space (FUSE)
- Fácil de migrar, depurar y mantener
- Diferentes modalidades de configuración
- Se crea sobre un sistema de archivos existente (ext3, ext4) lo que hace que se puedan recuperar los archivos aun si Gluster deja de funcionar

Ventajas de GlusterFS

- Los recursos de almacenamiento pueden ampliarse o disminuirse de acuerdo a las necesidades (on demand)
- Numerosas organizaciones de diversos tipos utilizando GlusterFS (servicios financieros, salud, web 2.0 etc.)

Método de acceso a la Data

- Los volúmenes de Gluster se pueden acceder de diversas maneras, se puede utilizar NFS para exportar o también el protocolo nativo de Gluster que es el más recomendado. Como fue mencionado anteriormente este protocolo esta basado en FUSE (Filesystem user space) por lo cual hay que asegurarse que este modulo este cargado en el kernel. FUSE permite levantar el sistema de archivo de Gluster en el user space (espacio de memoria donde trabajan las aplicaciones del usuario)

Algunas aplicaciones comerciales

- Gluster Virtual Appliance for Amazon Web Services
 - Permite el desarrollo de un servidor de almacenamiento basado en Gluster en la nube de Amazon

Algunas aplicaciones comerciales

- Gluster Virtual Storage
 - Permite levantar un servidor de almacenamiento en Gluster corriendo sobre una servidor virtual (VMware, KVM, Xen) permitiendo aumentar o reducir la cantidad de nodos fácilmente, de acuerdo a las necesidades de la organización

Algunas aplicaciones comerciales

- Gluster Storage Software Appliance
 - Es una imagen (ISO) que se instala en un equipo no virtualizado que integra GlusterFS con Centos para formar una solución completa.

Gluster vs Otras Soluciones

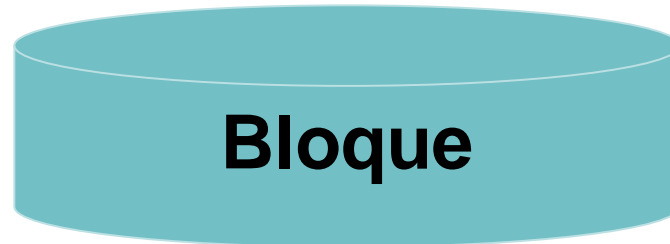
- Vs NFS:
 - Aunque NFS puede proveer de tiempos de lectura y escritura bajos y baja carga de CPU, NFS no provee alta disponibilidad ni redundancia, si el servidor NFS deja de estar disponible no hay acceso a la información

Gluster vs Otras Soluciones

- Vs Lustre:
 - Gluster es completamente distribuido, no utiliza servidor de meta-data lo que lo hace que sea sencillo escalar el servicio en cualquier momento.
 - Gluster esta diseñado manejar un número grande de archivos de pequeño tamaño (alrededor de 1Gb) Lustre esta optimizado para manejo de archivos grandes (>10Gb) Lo que hace que Gluster sea mucho mas adecuado como opción de almacenamiento para servicios web, correo electrónico etc.

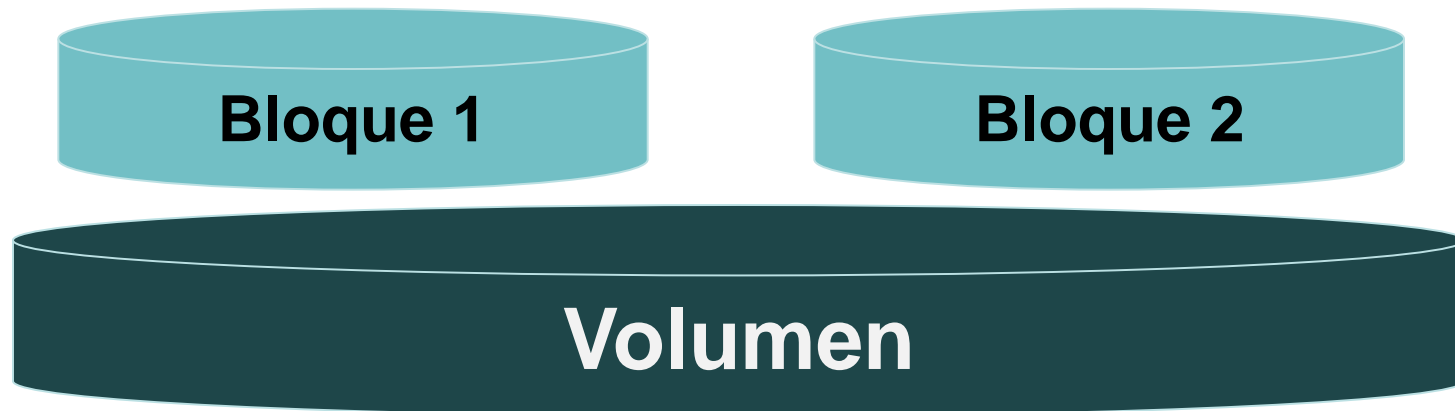
Terminología

- Bloque(Brick):
 - Es la capa física de almacenamiento para los volúmenes



Terminología

- Volumen:
 - Un volumen está compuesto por múltiples bloques

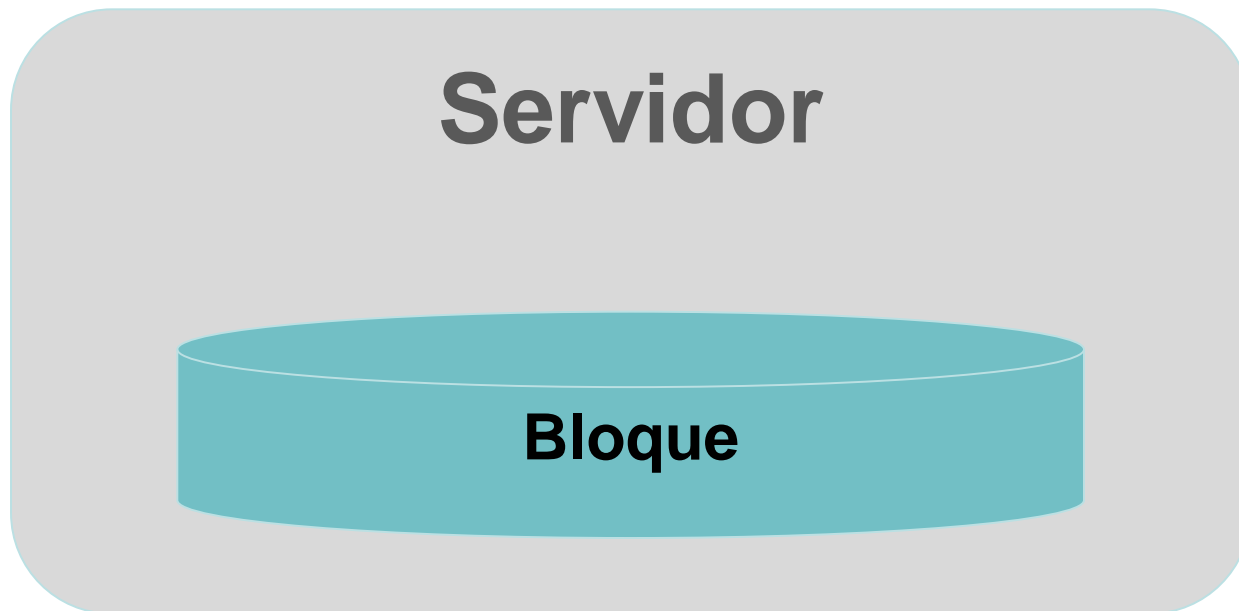


Terminología

- Servidor
 - Equipo (virtual o real) que exporta el volumen. Es el que alberga el sistema de archivos donde se guardará la data, y provee acceso a los volúmenes.

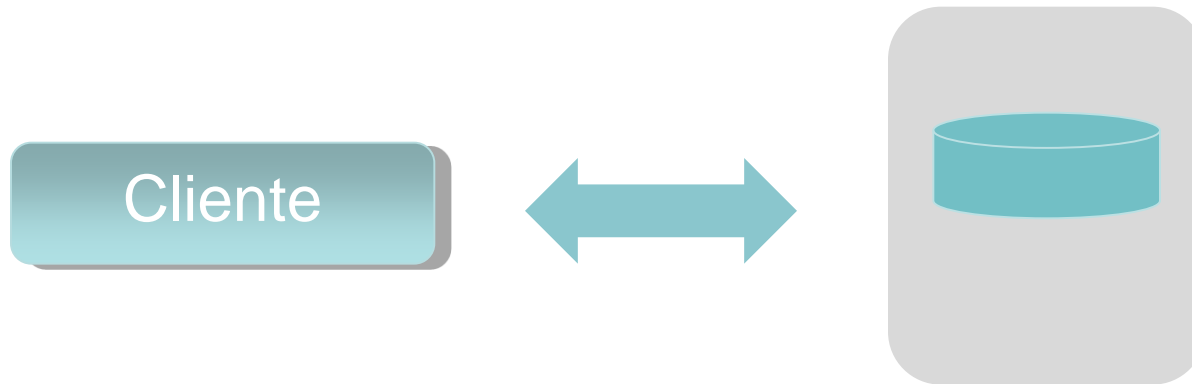
Terminología

- Servidor:



Terminología

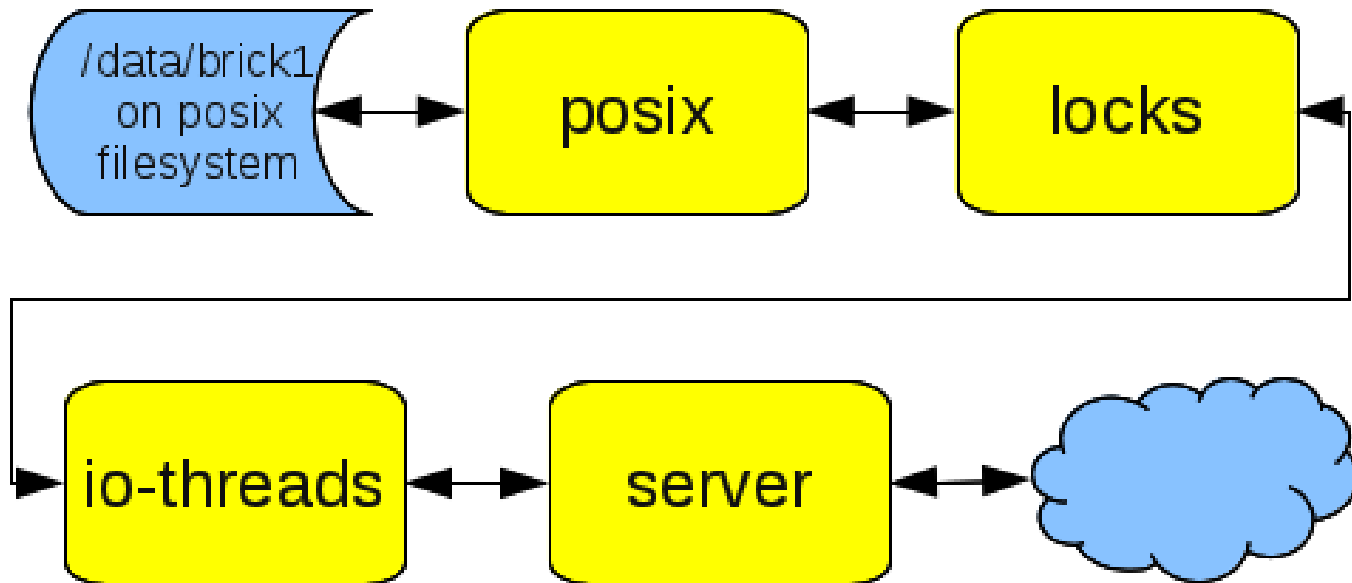
- Cliente
 - Es el equipo que monta el volumen, este equipo también puede ser un servidor



Terminología

- Traductor
 - Existen varios traductores por los cuales el bloque debe pasar. Por nombrar alguno: existe un que traductor lo conecta con un sub-volumen y otros traductores (storage/postfix traslator) que maneja la comunicación entre la interfaz del sistema de archivo y el resto de los traductores.

Terminología



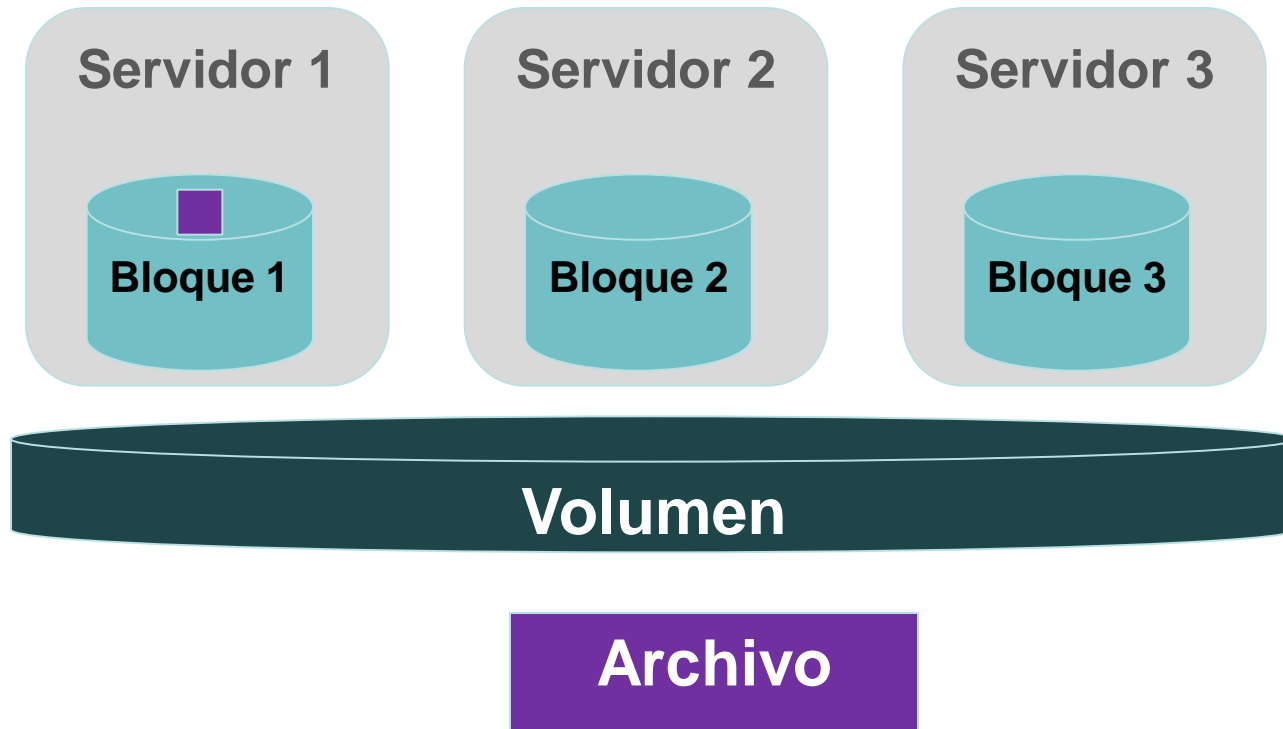
http://www.gluster.org/community/documentation/index.php/GlusterFS_Concepts

Distribución y Replicación de la Data

- Volumen Distribuido:
 - La configuración de Gluster en modo distribuido toma una lista de sub-volúmenes y distribuye los archivos entre ellos. Para calcular en cual de los servidores será almacenado el archivo se realiza un hash al nombre del archivo.

Distribución y Replicación de la Data

- Volumen Distribuido:



Distribución y Replicación de la Data

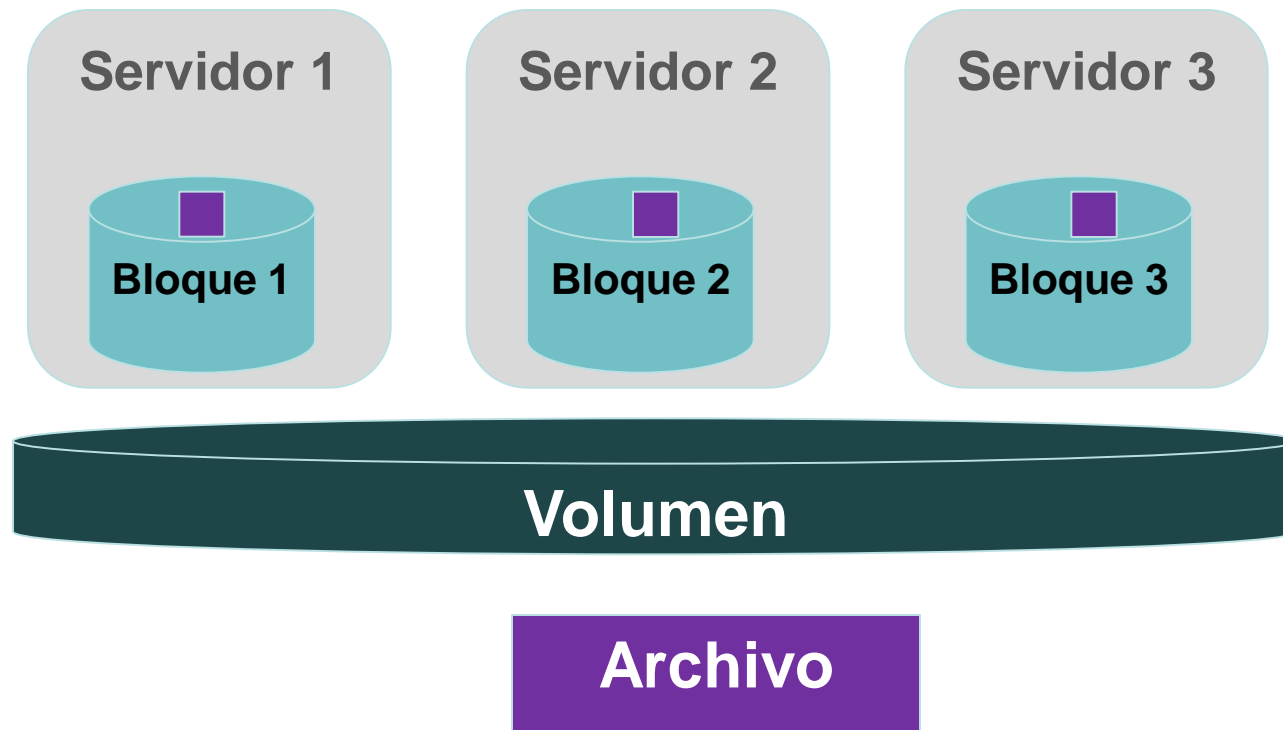
- Volumen Distribuido
 - Este modo tiene ciertas desventajas como que si uno de los servidores esta fuera de servicio no se pueden recuperar los archivos que está almacenados en él. También pueden presentarse problemas de lectura/escritura si un archivo es de un tamaño mayor al sub-volumen en el que debe ser almacenado.

Distribución y Replicación de la Data

- Volumen en Réplica:
 - Es el modo más utilizado, este modo provee redundancia en la información almacenada en los servidores, lo que se encuentra en un servidor esta exactamente igual en los otros servidores, lo que ofrece no solo redundancia en la data sino en la disponibilidad del servicio. En cada operación sobre los archivos se hace la copia en la réplica por lo que podemos decir que hay sincronización en la replica.
 - Funciona de manera análoga a RAID 1

Distribución y Replicación de la Data

- Volumen en Réplica:

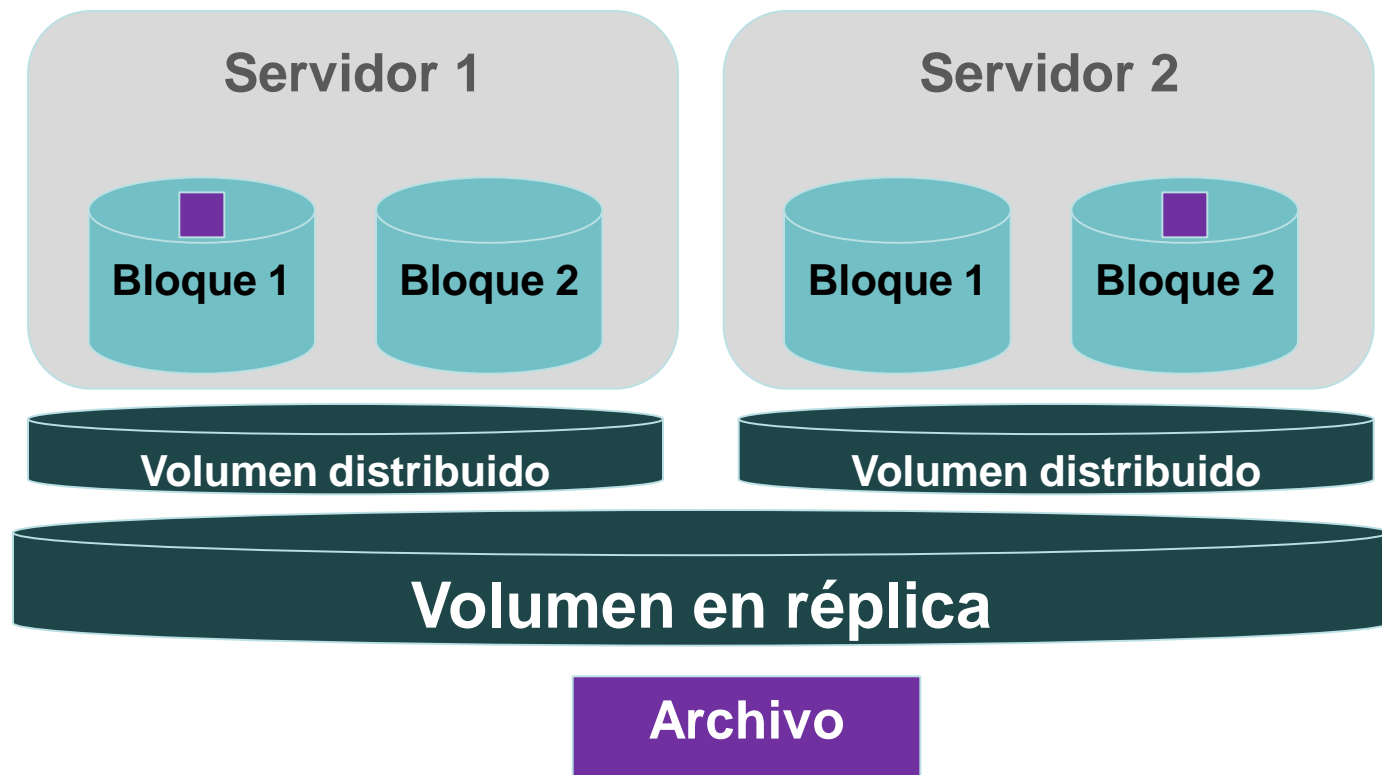


Distribución y Replicación de la Data

- Volumen Distribuido en Réplica
 - Este modalidad es una mezcla de las dos anteriores y distribuir los archivos a través de volúmenes en replica. Este tipo de configuración ofrece mejoras considerable en operaciones de lectura en la mayoría de los ambientes. Es el recomendado cuando la alta disponibilidad y la alta confiabilidad es critica

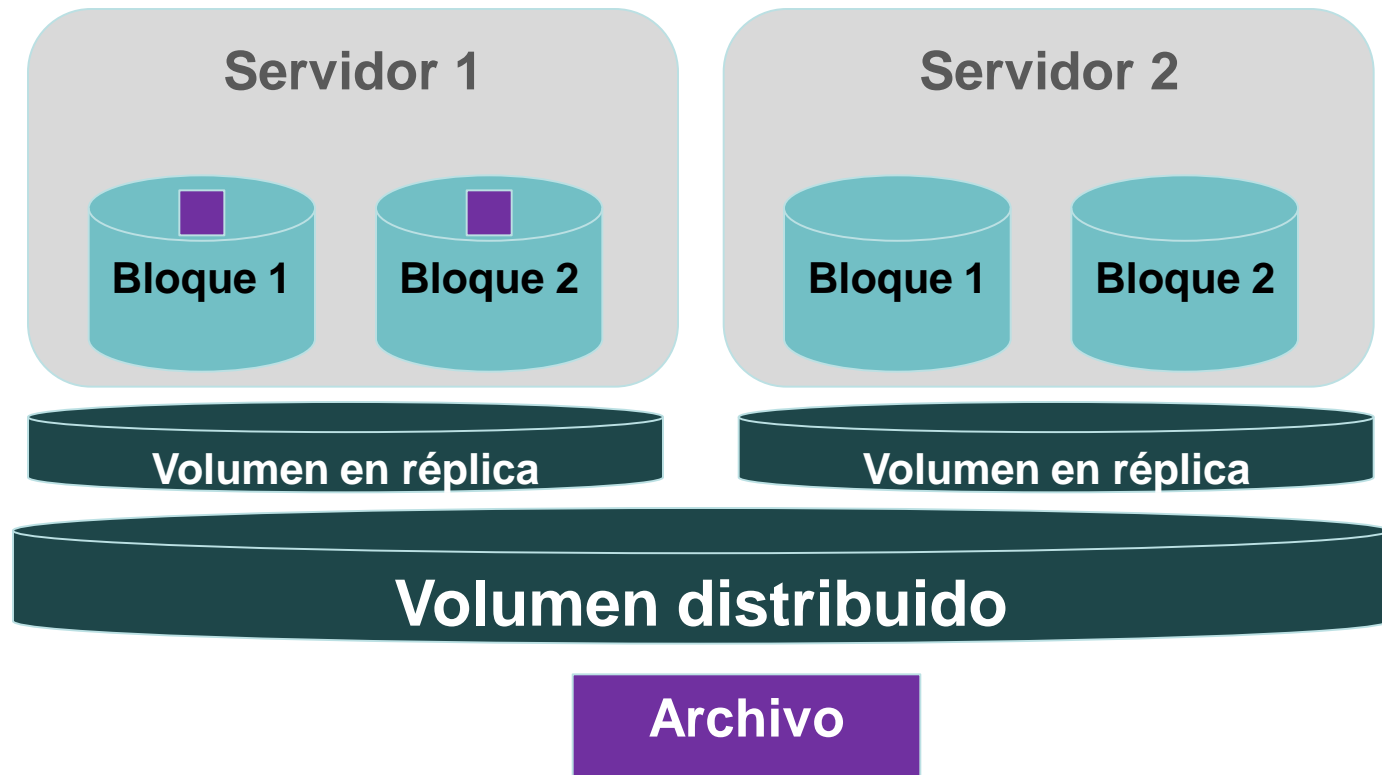
Distribución y Replicación de la Data

- Distribuido en Réplica



Distribución y Replicación de la Data

- Réplica Distribuido



Distribución y Replicación de la Data

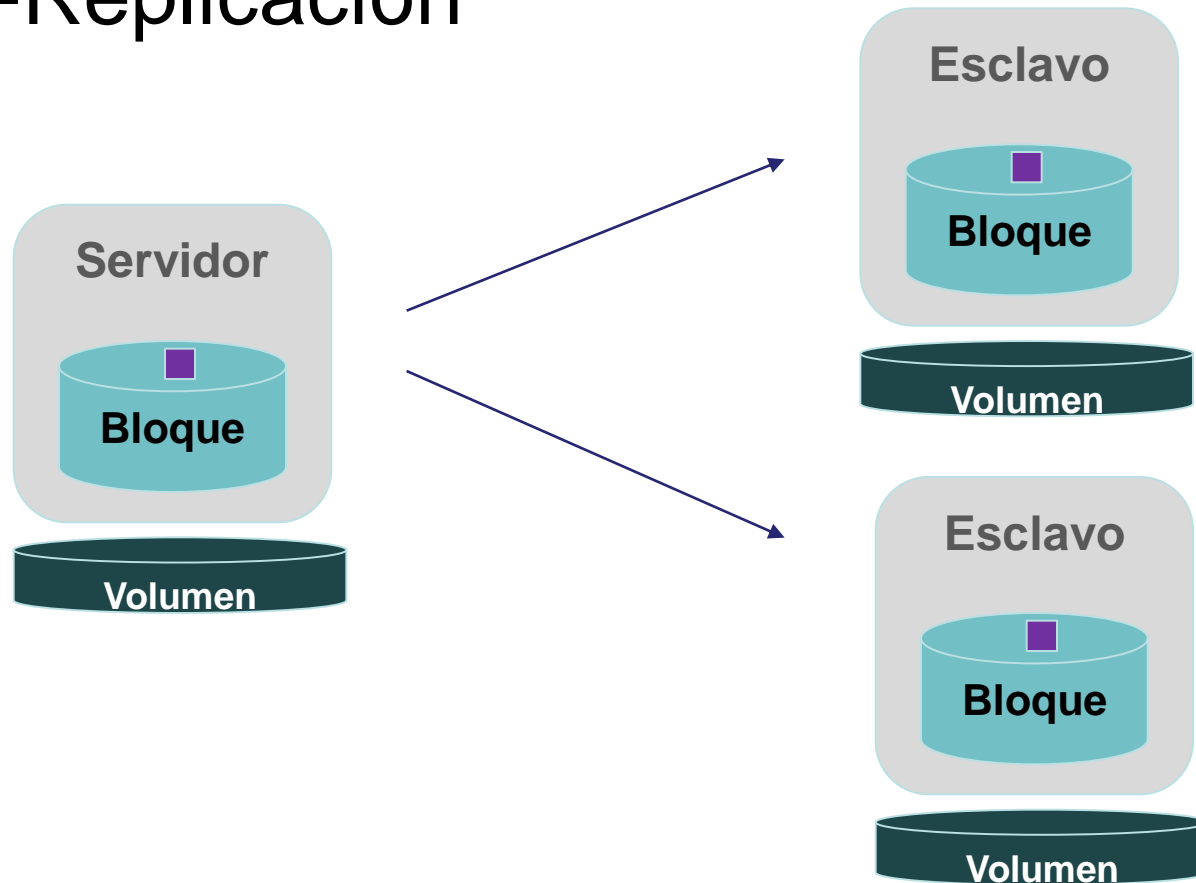
- Geo-Replicación:
 - Esta modalidad provee de un servicio de replicación asíncrona a través de una red LAN o WAN o Internet. Trabaja basándose en un modelo maestro-esclavo, siendo el maestro un volumen GlusterFS, y el esclavo puede ser
 - Un directorio Local
 - Otro volumen GlusterFS, que puede ser local o encontrarse en un host diferente.
 - Cuando la data en el maestro deja de estar disponible, puede restituirse a partir de cualquiera de los esclavos.

Distribución y Replicación de la Data

- Geo-Replicación
 - La Geo-Replicación a diferencia del Volumen en Réplica mantiene la copia de la data de forma asíncrona, es decir chequea periódicamente cambios en la información sincroniza al detectar algún cambio.

Distribución y Replicación de la Data

- Geo-Replicación

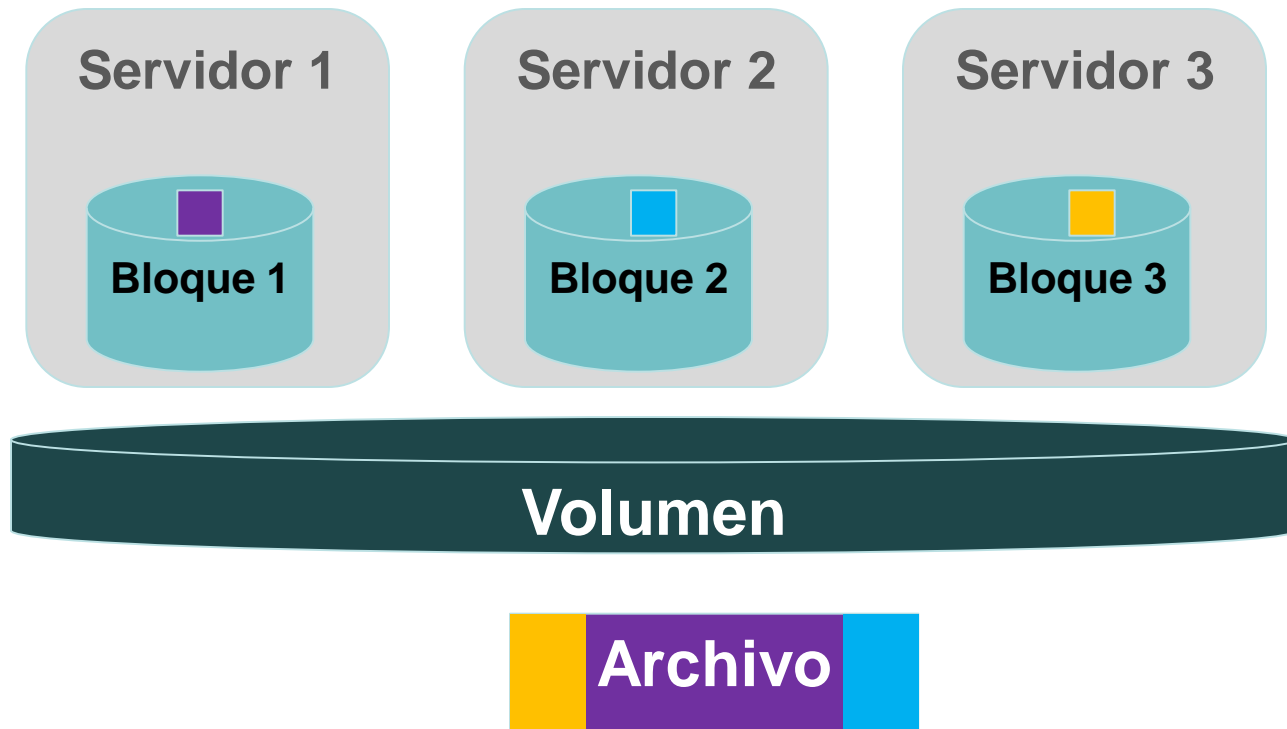


Distribución y Replicación de la Data

- Volumen Stripped
 - Esta modalidad es generalmente utilizada para almacenar datos de computación de alto rendimiento, funciona repartiendo cada archivo en diferentes bloques, con lo que podemos inferir que es mas óptimo para archivos de gran tamaño. Funciona de manera análoga a un RAID 0

Distribución y Replicación de la Data

- Stripped



Componentes de Gluster

- Glusterd
 - Demonio que gestiona el volumen. Debe correr en todos los servidores que exportan un sistema de archivo.

Componentes de Gluster

- Gluster
 - Es el comando para abrir la consola de Gluster. En la consola de Gluster también se pueden configurar manualmente los volúmenes (aumentar, reducir, balancear, migrar...) sin tener que parar el servicio