

# What is GlusterFS?

## GlusterFS Basic Explanation

GlusterFS is an open-source distributed file system that allows you to pool storage resources from multiple servers (nodes) into a single file system. It's designed to handle large amounts of data by distributing it across many machines, making it scalable and fault-tolerant. Essentially, GlusterFS lets you combine the storage capacity of several machines into a shared system that all machines can access.

GlusterFS is managed by the `glusterd` service, which is the core of the system. It keeps track of which files are stored on which machines and ensures that data is replicated across different nodes. This replication is useful because it allows your data to survive even if one of your nodes fails, ensuring high availability.

## Why Use GlusterFS for Syncing Docker Swarm Nodes?

When running services in Docker Swarm's replicated mode, the same service runs on multiple nodes to ensure high availability and scalability. These nodes may need to access the same set of data files, and that's where GlusterFS becomes very useful. Instead of storing separate copies of files on each node (which could lead to inconsistencies), GlusterFS ensures that all nodes share the same file system and stay in sync.

Using GlusterFS on the host system (rather than inside Docker) ensures that the file syncing works across the entire system, regardless of Docker's configuration. This setup is independent of the containers, allowing for seamless file sharing even if containers are recreated or moved to different nodes. Additionally, since GlusterFS operates at the system level, it can sync files across nodes more efficiently, avoiding network bottlenecks inside Docker containers.

Setting up GlusterFS across the nodes ensures that any file written on one node is immediately available to the others, providing data consistency and reliability when scaling services across a Docker Swarm cluster.