GlusterFS 1010

Author: Aaron Toponce
Email: aaron.toponce@gmail.com
Date: March 19, 2013

Contact and Details

You can find the source code, PDF, compressed tarball and HTML presentation at http://aarontoponce.org/presents/glusterfs.

My email address: aaron.toponce@gmail.com.

License

What is GlusterFS?

- Clustered filesystem
- Designed for commodity hardware

- Linear scalability
- High availability

Why GlusterFS?

"cloud server" vs "cloud storage"

Largest Data Usage

FUSE

- Uses the FUSE module
- Not in kernelspace
- Not slow - runs at wirespeed

Architecture

- Strongly consistent
- Peer consisted of "bricks"
- Volume consisted of connected peers
- Different from MooseFS or Ceph
- No metadata server - elastic hashing
- Files not split into chunks
- Supports TCP and RDMA

Volume Types

- Distributed
- Replicated
- Striped
- Distributed replicated
- Distributed striped

Distributed

- Distribute files throughout the bricks in the volume.
- No redundancy (JBOD).

Striped

- Stripes data across bricks in the volume.
- No redundancy, high performance (RAID-0).

Replicated

- Replicate files throughout the bricks in the volume.
- High redundancy (RAID-1).

Distributed replicated

- Distributes files across replicated bricks in the volume.
- Number of bricks should be a multiple of the replica count.
- High redundancy, high availability.

Distributed striped

- Stripe data across two or more nodes in the cluster.
- Number of bricks should be a multiple of the stripe count.
- High performance, high concurrency, no redundancy.

General Idea

- Bricks created on peer
- Peers connected together
- Volume mounted by FUSE clients

Paired Server Topology

- Servers added/removed in pairs
- Simple administration

Linked List Volume Topology

- Servers added/removed individually
- Complex administration

Creating A Volume

- `gluster peer probe server1`
- `gluster peer probe server2`
- `gluster volume create clstrfck replica 2 transport tcp server1:/vol1 server2:/vol1 server2:/vol1 server2:/vol2`
- `gluster volume start clstrfck`

Mounting A Volume

- `mount -t glusterfs server1:clusterfck /srv`

Additional Administration

- Adding peers: `gluster peer probe server3`
- Removing peers: `gluster peer detach server1`
- Adding bricks: `gluster volume add-brick clstrfck server3:/vol1 server3:/vol2`
- Adding geo-replication: `gluster volume geo-replication clstrfck offsite:/backup`

Help

- Offical Gluster Docs
- gluster-users mailing list
- Q&A Forums
- Bug Tracker
- IRC: irc://irc.gnu.org#gluster, irc://irc.freenode.net:#gluster