GLUSTER 4.0
WHAT'S NEXT IN GLUSTER?

Kaushal (kshlm/kshlmster)
GlusterD Maintainer
AGENDA

- Quick Gluster intro
- Gluster-4.0
WHAT IS GLUSTER?

- Software defined network storage system
  - Distributed, Replicated/EC
  - No metadata server
- Commodity hardware and scale-out
- Posix compatible
  - Multiple-access methods
GLUSTERFS TERMS

- **Peer/Node/Server** - A computer with the GlusterFS server packages installed
- **Trusted Storage Pool** - The GlusterFS cluster
- **Client** - accesses GlusterFS using the native protocol
GLUSTERFS TERMS

- Brick - Smallest unit of storage in GlusterFS
- Volume - A logical collection of bricks, that appears as a single export to clients
GLUSTERFS TERMS

- Translators - Modular bits of GlusterFS that implement the actual features
- Volume graph - A graph of translators arranged together to create a volume
USING GLUSTER

- `gluster peer probe <hostname>`
- `gluster volume create <name> replica 2 <peername>:/path <peername>:/path ...`
- `gluster volume start <volumename>`
- `mount -t glusterfs <peername>:/<volname> /<path to mountpoint>`
PAST AND PRESENT
ORIGIN

- Gluster - GNU Cluster
  - Linux distro
  - Clustered supercomputer
- GlusterFS provided storage
  - GlusterFS v1 - Part of main Gluster project
  - GlusterFS v2 - Split into a separate project
  - GlusterFS v3 - Primary project
GLUSTERFS-3.X

- 3.0 - December 2009
  - Protocol defined
- 3.1 - October 2010
  - GlusterD, NFS
GLUSTERFS-3.X

• 3.2 and beyond
  ▪ EC, Snapshots
  ▪ GFAPI, NFS Ganesha, Samba
  ▪ Heketi

• 3.13 - December 2017
  ▪ Current release
GLUSTER-4.0

- Next major release
- Late February 2018
- Short term maintenance
- Drop support for older distros
  - EL6
GLUSTER-4.0

- Protocol changes*
- GlusterD2*
- Metrics
- FIPS
- Performance enhancements
GLUSTER-4.X

- RIO - Relation Inherited Object distribution
- JBR - Journal based replication
- GFProxy
- Halo replication
- Thin arbiter
- +1 scaling
- More automated cluster management
GLUSTER-4.0
PROTOCOL CHANGES

- New on-wire RPC
- Better XDR structs
  - More defined members
  - New dictionary
- Old RPC version still available
GLUSTERD2

- New management system for Gluster-4.0
  - Not backwards compatible with GD1
  - Backwards compatible with clients
- From scratch rewrite, written in Go
- Better scalability, integration and maintenance
GLUSTERD2

- ReST/HTTP API
  - JSON requests and responses
- New CLI
- More flexible and pluggable internal frameworks
  - Transaction, Volgen, ReST, Events
- Uses Etcd internally
  - automatic Etcd setup/management
GD2 IN GLUSTER-4.0

- Technology preview version
- Most of GD1 commands to be implemented
- Preliminary automatic volume creation
- No upgrade/migration support from 3.x
GD2 IN GLUSTER-4.1+

- Stabilize
- Documentation on commands, APIs and different workflows
- Support upgrade/migration from 3.x
- Centralized logging and tracing
- Fully automatic volume management aka dynamic volume provisioning, +1 scaling
- Automatic cluster formation
- More native APIs for integration and workflows
UPGRADING TO 4.X

- Avoid filesystem downtime
  - Existing clients can retain continuous access
- GD1 and GD2 shipped in 4.0 and 4.1
  - GD1 planned to be removed from 4.2
  - No GD1 feature updates
UPGRADING TO 4.X

- Rolling upgrade from 3.x to 4.0 using GD1
  - Install 4.0 on one node
  - Restart Gluster on the node
  - Heal volume
  - Continue on next node
- Kill GD1 and start GD2 everywhere
- Migrate/import data into GD2
- GD2 picks up running bricks and daemons
- Upgrade done without total downtime
THANK YOU!