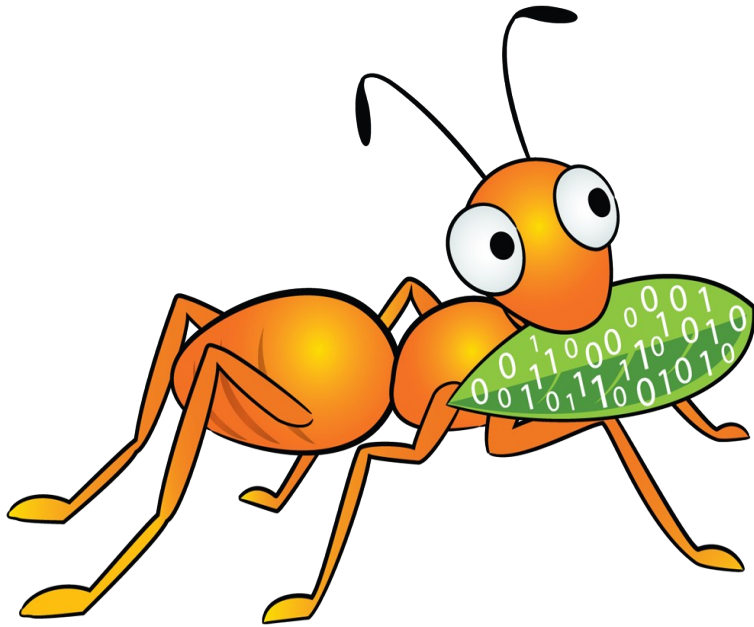# GlusterFS
Current Features & Roadmap

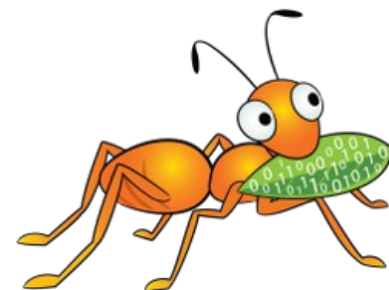**Niels de Vos**
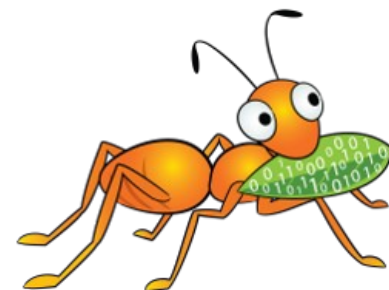**GlusterFS co-maintainer**

ndevos@redhat.com

Agenda

- Introduction into GlusterFS

- Quick Start

- Current stable releases

- History of feature additions

- Plans for the upcoming 3.7 release

- Expectations of the next major release

What is GlusterFS?

- Scalable, general-purpose storage platform
  - POSIX-y Distributed File System
  - Object storage (swift)
  - Distributed block storage (qemu)
  - Flexible storage (libgfapi)
- No Metadata Server
- Heterogeneous Commodity Hardware
- Flexible and Agile Scaling
  - Capacity – Petabytes and beyond
 - Performance – Thousands of Clients

Terminology

- Brick
  - Fundamentally, a filesystem mountpoint
  - A unit of storage used as a *capacity* building block
- Translator
  - Logic between the file bits and the Global Namespace
 - Layered to provide GlusterFS *functionality*
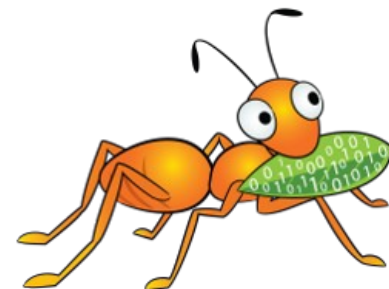
Terminology

- Volume
  - Bricks combined and passed through translators
  - Ultimately, what's presented to the end user
- Peer / Node
  - Server hosting the brick filesystems
  - Runs the Gluster daemons and participates in volumes
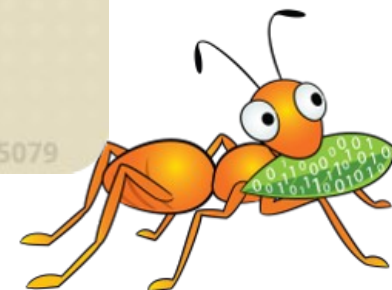- Trusted Storage Pool
  - A group of peers, like a "Gluster cluster"
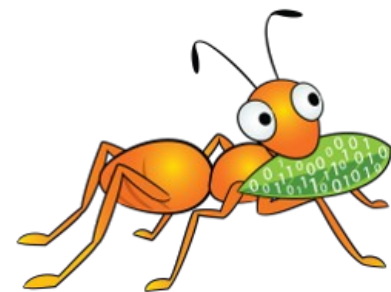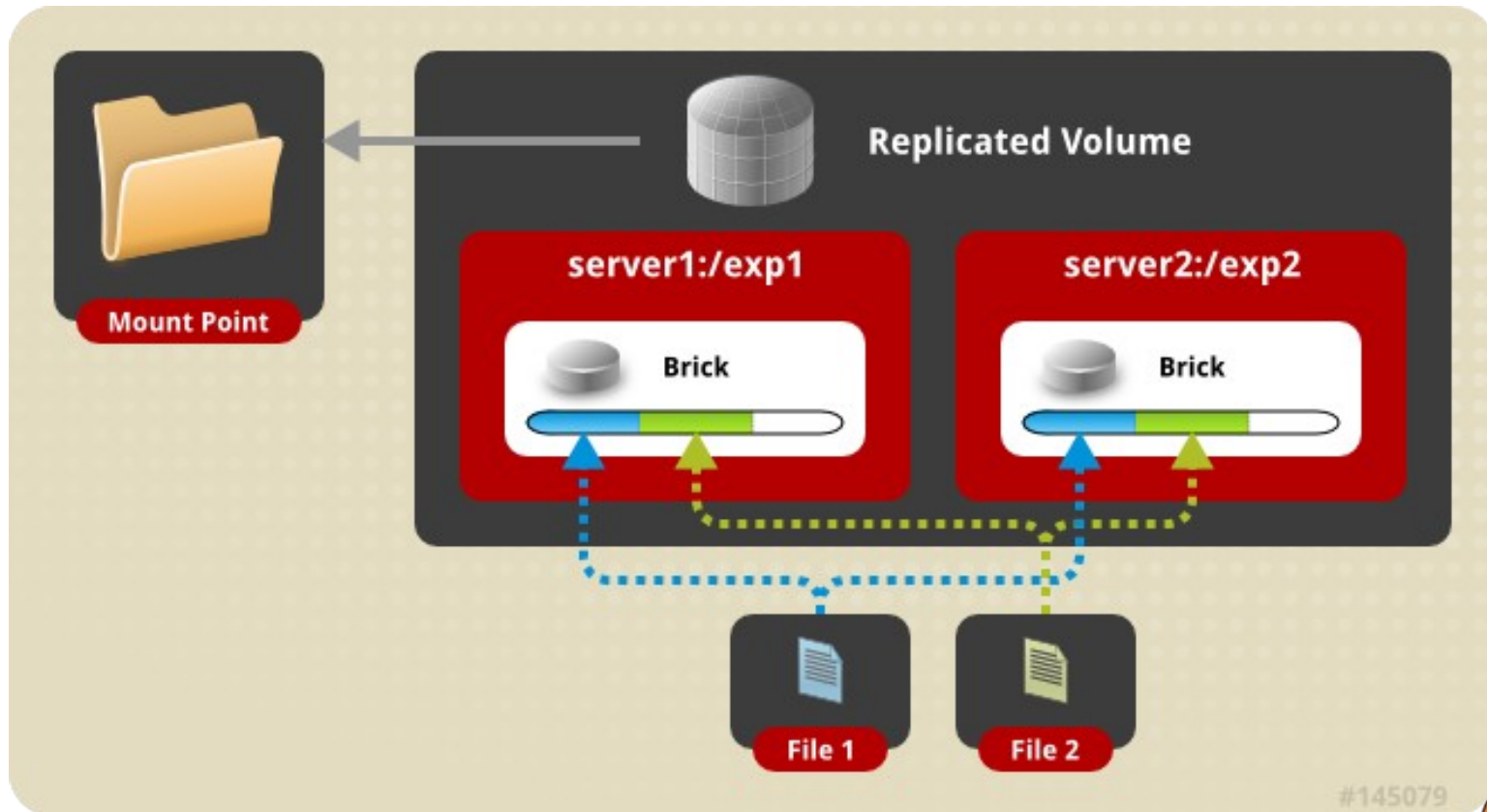
Distributed Volume

- Files "evenly" spread across bricks
- *Similar* to file-level RAID 0
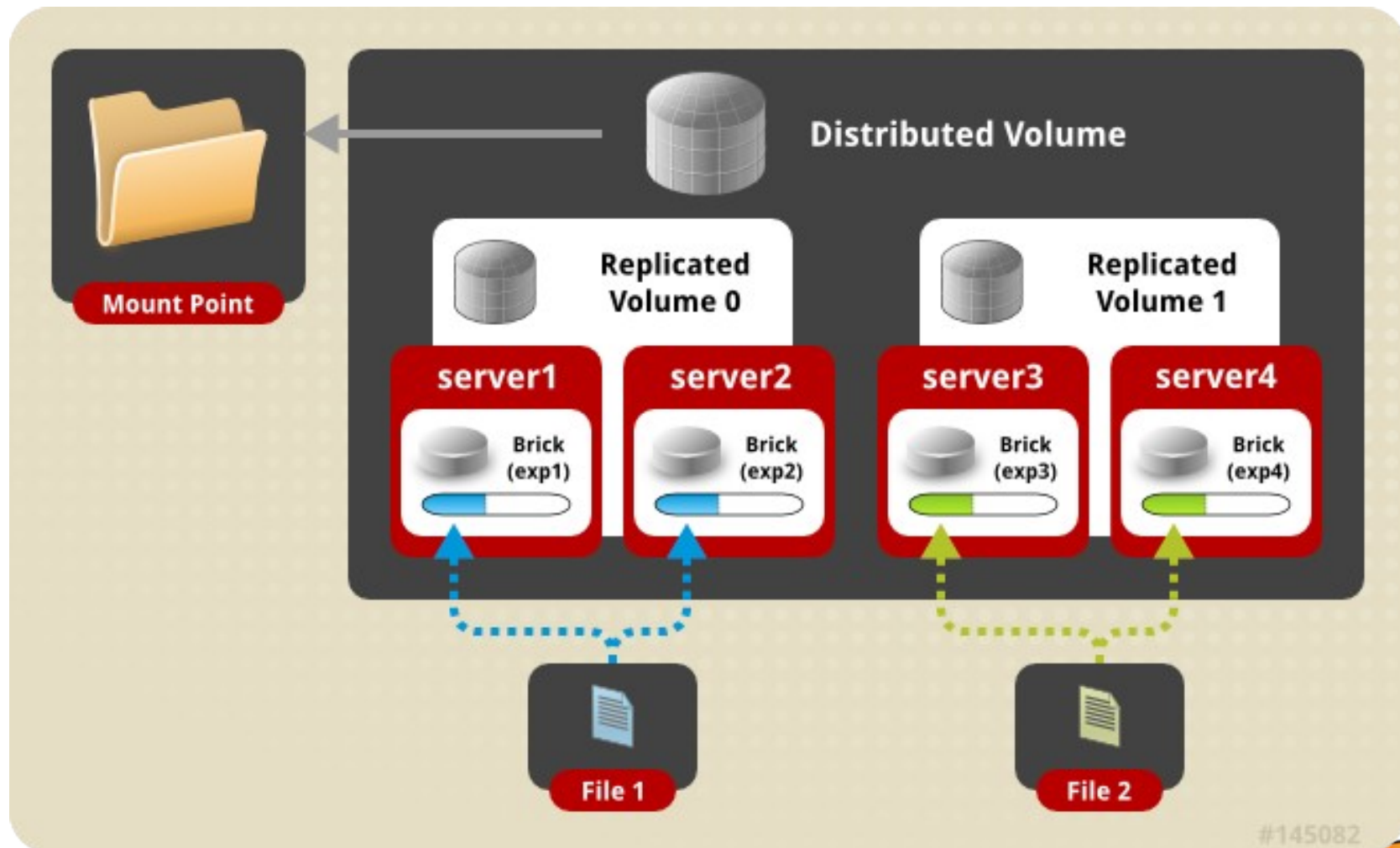- Server/Disk failure could be catastrophic
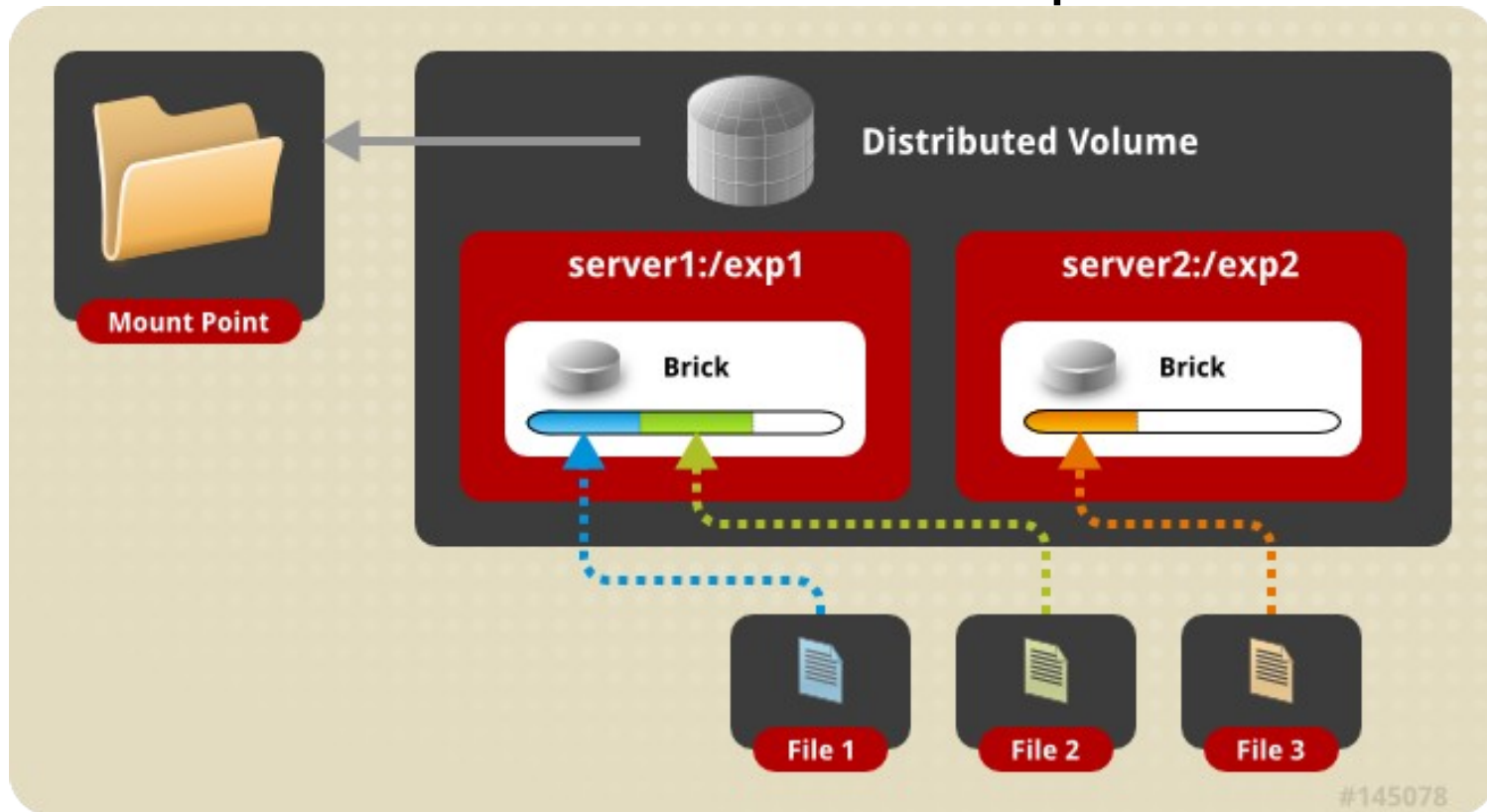
Replicated Volume

- Copies files to multiple bricks
- *Similar* to file-level RAID 1

Distributed Replicated Volume

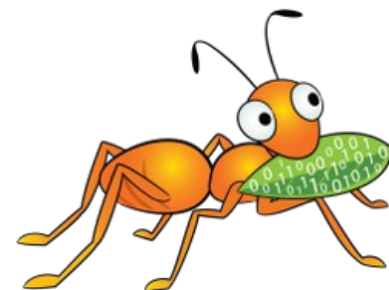- Distributes files across replicated bricks

Data Access Overview

- GlusterFS Native Client

  - Filesystem in Userspace (FUSE)

- NFS

  - Built-in Service, NFS-Ganesha with libgfapi

- SMB/CIFS

  - Samba server required (libgfapi based module)

- Gluster For OpenStack (Swift-on-file)

  - Simultaneous object-based access via Swift

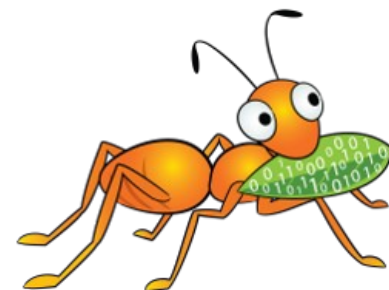- libgfapi flexible abstracted storage

  - Integrated with QEMU, Bareos and others

Quick Start

- Available in Fedora, Debian, NetBSD and others

- Community packages in multiple versions for different distributions on http://download.gluster.org/

- CentOS Storage SIG packages and add-ons

- Quick Start guides on http://gluster.org and CentOS wiki

Quick Start

1. Install the packages (on all storage servers)

2. Start the GlusterD service (on all storage servers)

3. Peer probe other storage servers

4. Create and mount a filesystem to host a brick

5. Create a volume

6. Start the new volume


7. Mount the volume

Current stable releases

- Maintenance of three minor releases

  - 3.6, 3.5 and 3.4

- Bugfixes only, non-intrusive features on high demand

- No fixed release schedule (yet?)

- Patches get backported to fix reported bugs
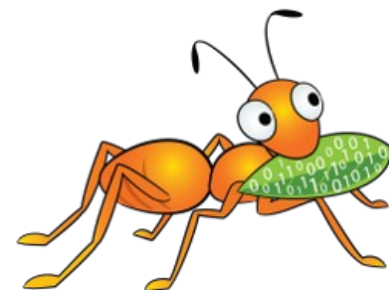
Features included in version 3.4

- WORM: Write Once Read Many

- Operating versions for GlusterD

- Block device translator

- Duplicate Request Cache (used with NFS)

- Server Quorum

- libgfapi for native GlusterFS support in applications

- Eager Locking

- NFSv3 ACL support
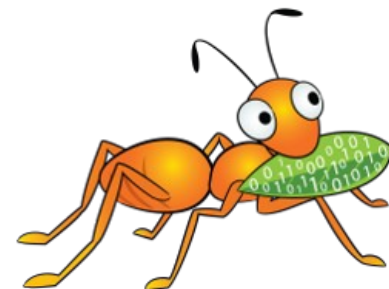
Features included in version 3.5

- File Snapshot for qcow2 files

- GFID access

- On-Wire (de)compression

- Quota Scalability

- Readdir ahead

- Zerofill

- Brick Failure Detection

- Parallel geo-replication
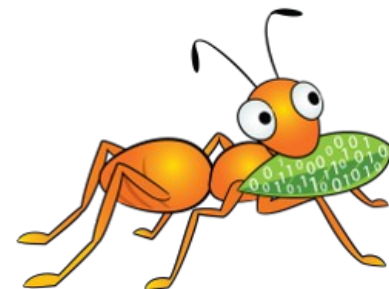
Features included in version 3.6

- Improved SSL support

- Heterogeneous bricks

- Volume wide locks for GlusterD

- Volume Snapshots

- User Serviceable Snapshots

- AFR refactor

- RDMA improvements

- Disperse translator for Erasure Coding
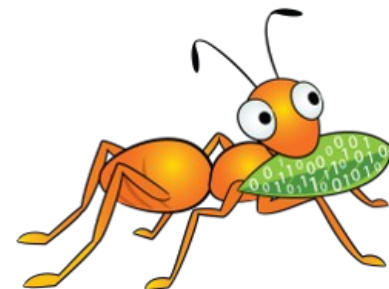
Plans for the upcoming 3.7 release

Feature freeze at the end of February

- Small-file performance enhancements
- Tiering, rack-aware placement and more
- Trash translator for undelete operations
- Netgroups and advanced exports configuration (NFS)
- BitRot detection
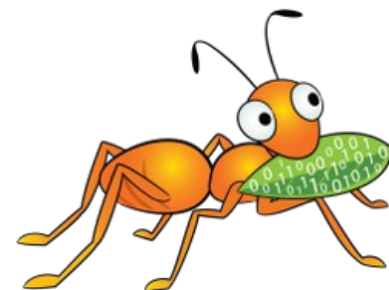- Support for NFS Ganesha clusters
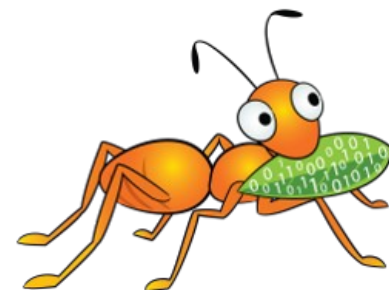
Data Classification in 3.7

- Mapping file characteristics to subvolume characteristics

- File characteristics:

  - Size, age, access rate, type (filename extension)

- Subvolume characteristics:

  - Physical location, storage type, encoding method

- User provided mapping via 'tags'
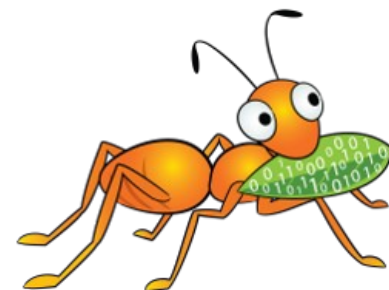
- Implemented using 'DHT over DHT' pattern
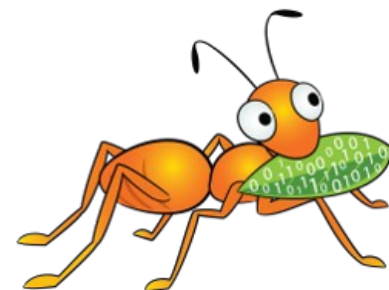
Netgroups and Exports for NFS in 3.7

- More advanced configuration for authentication based on /etc/exports like syntax

- Support for netgroups

- Patches written by Facebook developers

- Forward ported from 3.4 to 3.7

- Cleanups and posted for review

NFS Ganesha support in 3.7

- Optionally replaces Gluster/NFS

- Supports NFSv4 with Kerberos

  - pNFS support for Gluster Volumes follows later

- Modifications to Gluster internals

  - Upcall infrastructure

  - Gluster CLI to manage NFS Genesha

  - libgfapi improvements

- High-Availability based on Pacemaker and Corosync

Plans for the next 4.0 release

- Intended for scalability and manageability improvements

- Support for multiple networks

- New Style Replication

- Improved Distributed hashing Translator

- Composite operations in the GlusterFS RPC protocol

- Coherent client-side caching

- Native ReST APIs for management and monitoring

- ... and much more
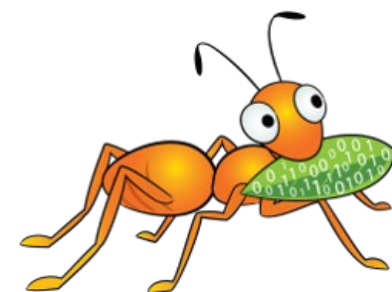
GlusterFS 4.0 – What's next?

- Code name for the release? Open to suggestions

- Submissions for feature proposals is still open!

- Implementing of key features has started

- Voting on feature proposals during design summit
  - Tentatively planned for March/April timeframe

Integration in other projects

- oVirt for easier installation, management and monitoring

- Nagios for improved monitoring and alerting

- OpenStack Manila (filesystem service)

- Hadoop plugin offers an alternative for HDFS

- Bareos Gluster File Daemon plugin
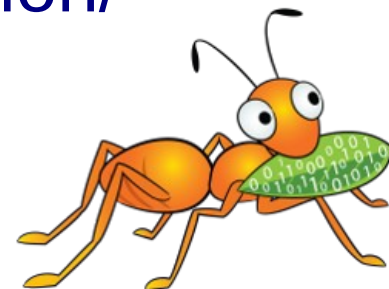
- … and many others

Resources

Mailing lists:
gluster-users@gluster.org
gluster-devel@gluster.org

IRC:
#gluster and #gluster-dev on Freenode

Links:
http://gluster.org/
http://forge.gluster.org/
http://www.gluster.org/community/documentation/
http://gluster.org/presos.php

Thank you!
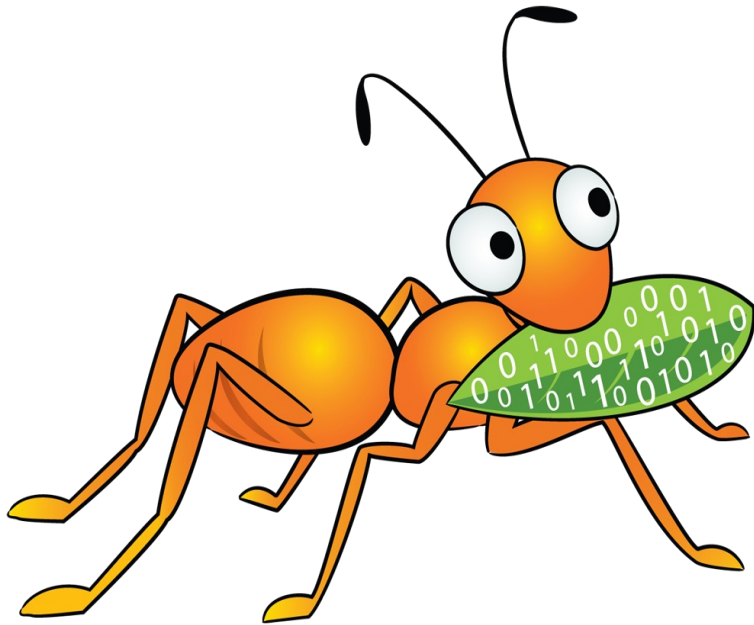
Niels de Vos
ndevos@redhat.com
ndevos on IRC