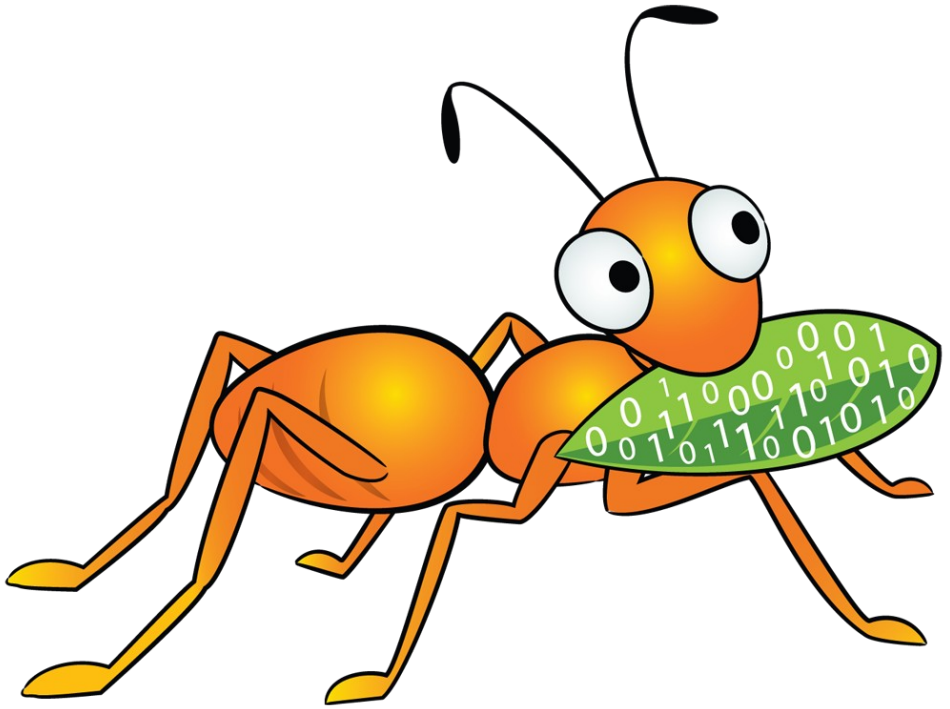
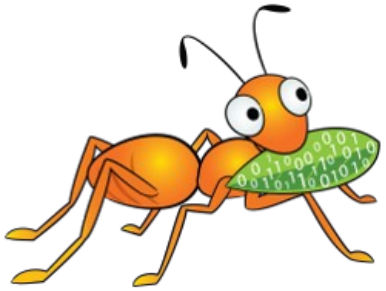
Gluster related development



Niels de Vos
Sr. Software Maintenance Engineer
Global Support Services
Red Hat

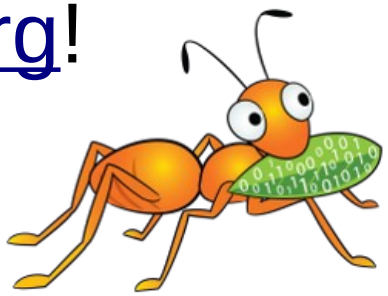
Agenda

- Gluster Community Forge
- libgfapi for reducing overhead
- MapReduce with pmux
- Hadoop
- SWIFT



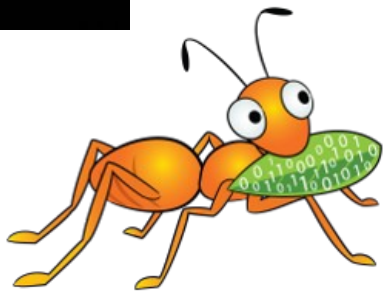
Gluster Community Forge

- Hosting for Gluster related projects
- Tools:
  - lsgvt, gtop, ...
- Pre-upstream translators:
  - Disperse, glupy, file versioning, ...
- Big Data, MapReduce:
  - Hadoop enablement, pmap
- And many more at http://forge.gluster.org!



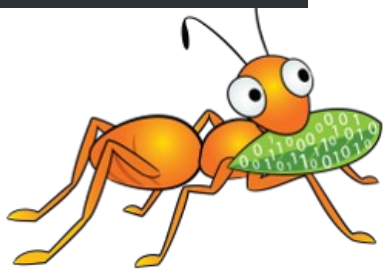
Tool: lsvgt

```
$ ./lsgvt -p tests -a distrep2x2

Topology for volume distrep2x2:

Distribute set
 |
 +---- Replica set 0
 |      |
 |      +---- Brick 0: vm-152.example.com:/bricks/distrep2x2
 |      |
 |      +---- Brick 1: vm-218.example.com:/bricks/distrep2x2
 |
 +---- Replica set 1
        |
        +---- Brick 0: vm-86.example.com:/bricks/distrep2x2
        |
        +---- Brick 1: vm-91.example.com:/bricks/distrep2x2
```



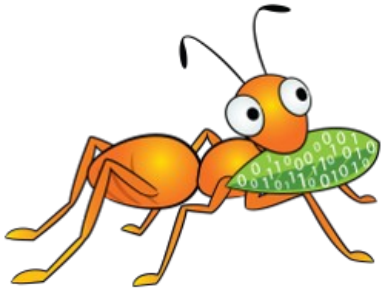
Tool: gtop

```
gtop - 3.3.0.7rhs   2 nodes,  2 active CPU%:  3 Avg,  3 peak Skew:  3s  07:54:11
Activity: Network:  44K in,  45K out    Disk:   0b reads,   8K writes
Storage:10 volumes, 20 bricks /   23G raw,  14G usable,   3G used,  11G free
Volume            Bricks   Type   Size   Used   Free   Volume Usage
ctdb                 2      R     491M    25M    466M   5%
ftp                  2      R     991M    33M    959M   3%
repl                 2      R    1015M   195M    820M   19%
smallfiles           2      R      6G      2G      4G   39%
temp1                2      D    1007M    51M    955M   5%
temp2                2      D    1007M    51M    955M   5%
temp3                2      D    1007M    51M    955M   5%
temp4                2      D    1007M    51M    955M   5%
temp5                2      D    1007M    51M    955M   5%
temp6                2      D    1007M    51M    955M   5%

                       CPU        Memory %   Daemons      Network      Disk I/O
S Gluster Node      C/T  %   RAM  Real|Swap C-S-N-H-G    In  | Out  Reads | Writes
▲ rhs5-1            2    3  491M   94  11  Y Y Y Y .    12K   11K     0b      3K
▲ rhs5-2            2    3  491M   93  14  Y . Y Y .    32K   34K     0b      5K
```

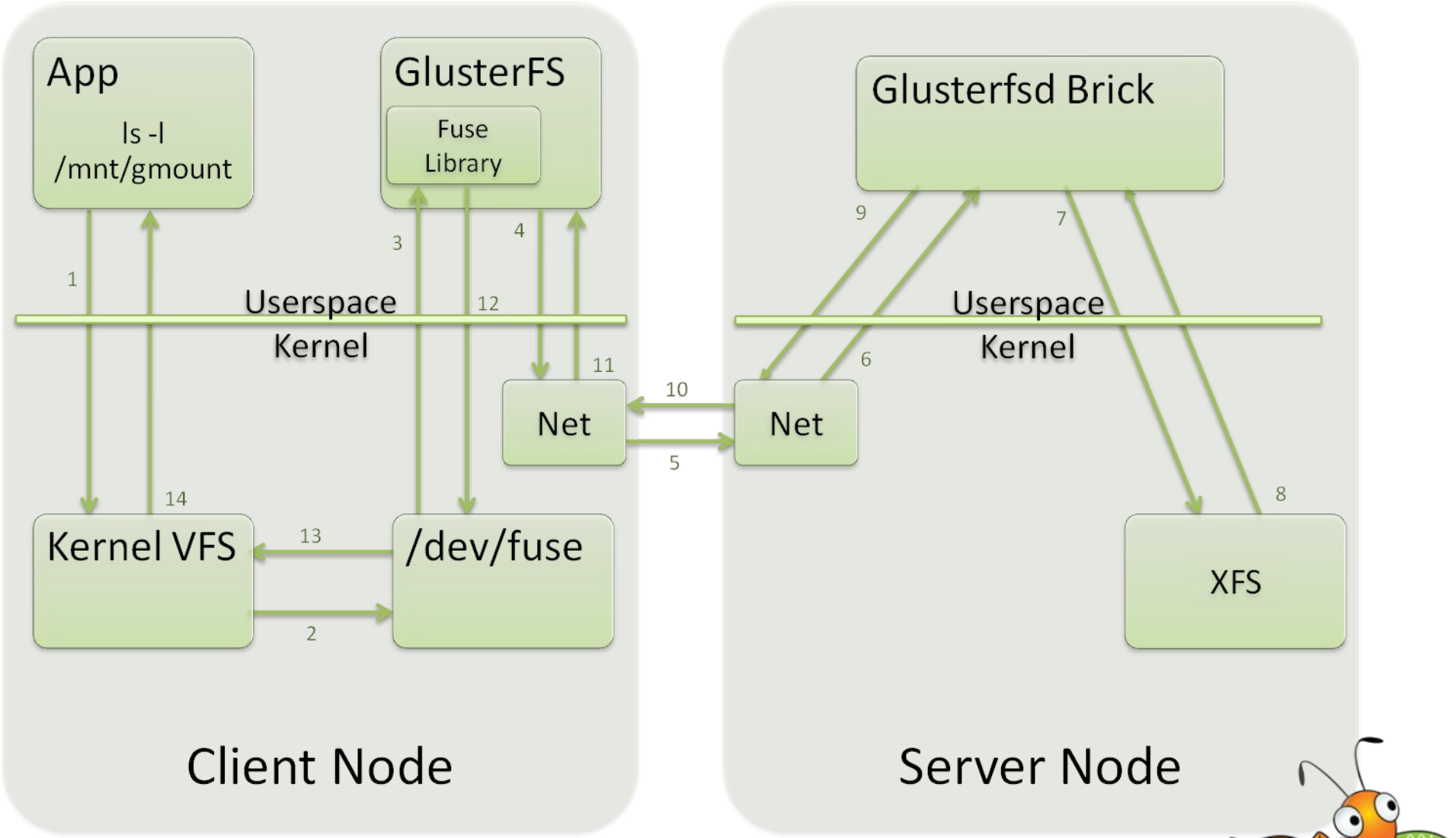


libgfapi for reducing overhead

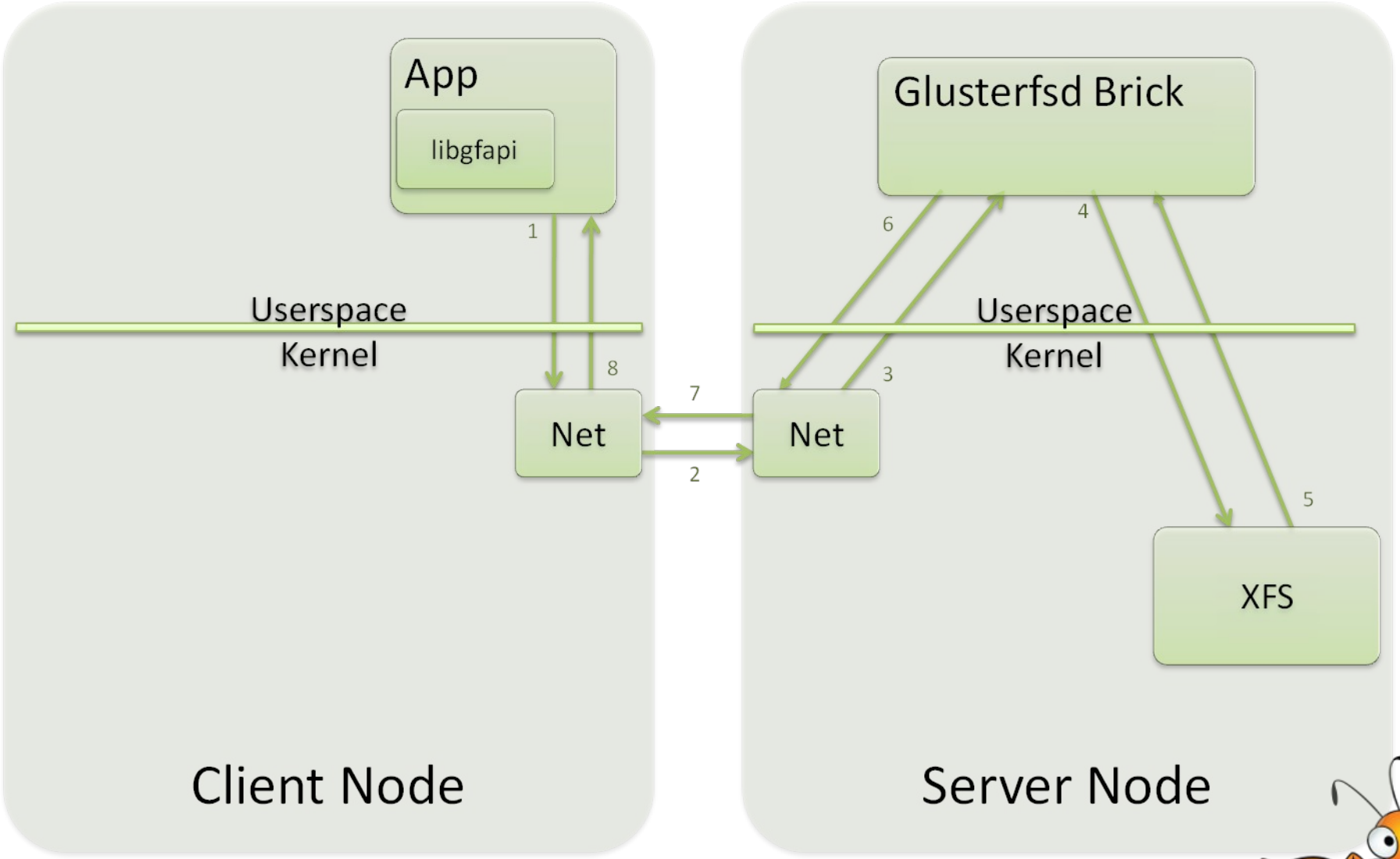
- Can be used to eliminate the FUSE and kernel VFS layer
- No requirement for mounting a volume
- Existing applications need to be extended
- Examples:
  - qemu storage layer
  - Samba VFS plugin



libgfapi: access through FUSE



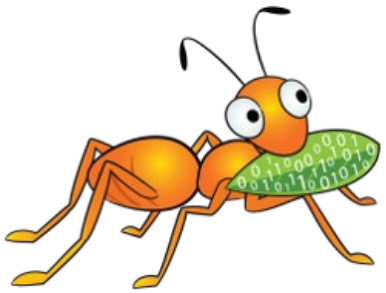
libgfapi: access through library



Samba libgfapi integration

- Prevents the need for volumes mounted through FUSE, improves performance
- Upstream Samba VFS plugin:
  `/usr/lib64/samba/vfs/glusterfs.so`
- Updated `gluster volume start/stop` hook scripts that modify `/etc/samba/smb.conf` on demand
- New `smb.conf` configuration snippet:

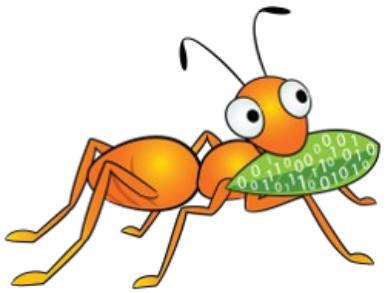
```
[gluster-$volname]
vfs objects = glusterfs
glusterfs:volume = $volname
path = /
```



Example of libgfapi

- Examples are available in the sources
  `<glusterfs.git>/api/examples/`
- Example execution of the test-suite:
  `$ ./gfapi.py playground gfapi-example`
- Example Python code:

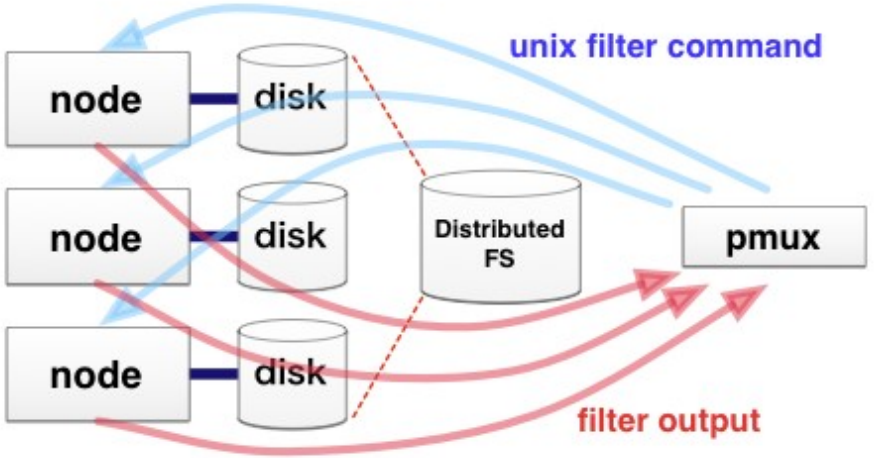
```
import gfapi
STORAGE_SERVER = "localhost"
VOLUME = "playground"
DATA_BASENAME = "gfapi-example"
vol = gfapi.Volume("localhost", "playground")
vol.mount()
vol.unlink(DATA_BASENAME + ".xa")
```

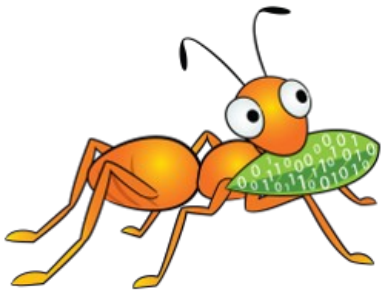


MapReduce with pmux

- Lightweight MapReduce
- Usage is similar to pipes in a shell
- Example:

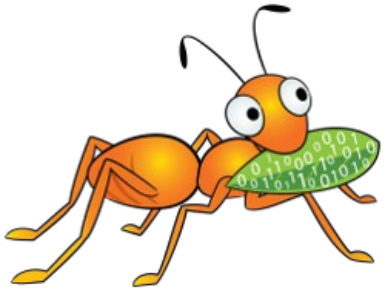
```
$ pmux --mapper="grep PATTERN" /mnt/vol/*.log
```





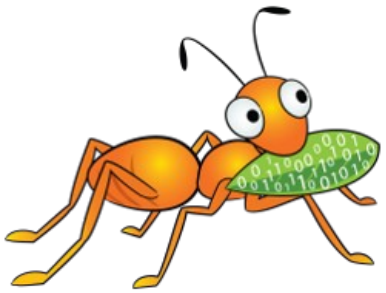
pmux internals and helpers

- Written in Ruby
- Uses `gflocator` to detect the locality of data
  - Runs as a daemon (requires root privileges)
  - Requests special extended attributes of files `trusted.glusterfs.pathinfo`
- pmux-gw provides a HTTP interface



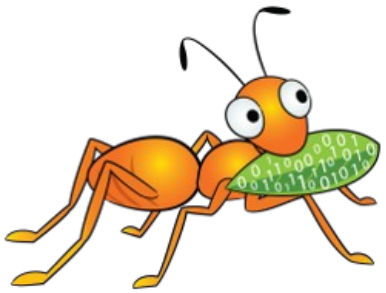
MapReduce with Hadoop

- Plugin for Hadoop
  - `glusterfs-hadoop-<version>.jar`
  - Support for Hadoop 1.x and 2.x
- Very clear installation guide:
  - https://forge.gluster.org/hadoop/pages/Home
  - https://forge.gluster.org/hadoop/pages/Configuration
- Works with Apache Hadoop
  - Tested with Hortonworks and Intel Hadoop Distributions



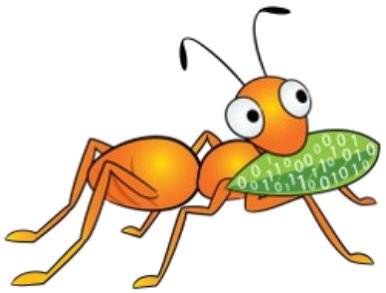
OpenStack SWIFT backend

- Unified File and Object (UFO) Storage based on the OpenStack SWIFT API
  - Access through HTTP(S)
  - Simple protocol, easy for client development
- Compatible with OpenStack 1.8
- Patches needed for multi-volume support
  - Need update/testing for OpenStack 1.9
  - Work is being done on upstreaming

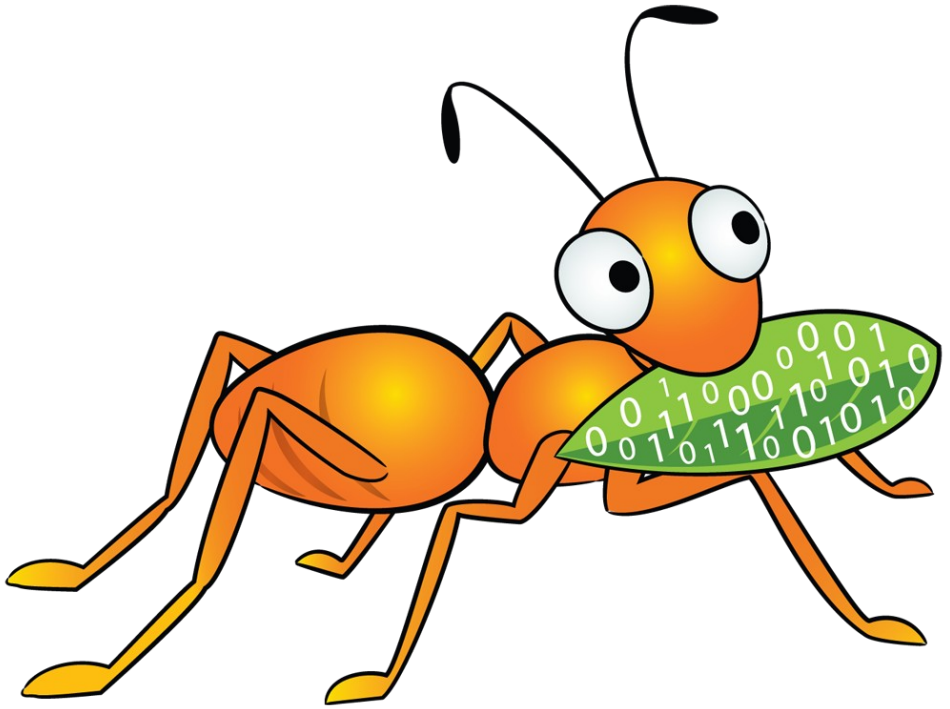


Questions?

- User questions and experiences
  - gluster-users@gluster.org
  - #gluster on Freenode IRC
- Developer questions and discussions
  - gluster-devel@nongnu.org
  - #gluster-dev on Freenode IRC
- Announcements mailinglist
  - announce@gluster.org



Thanks!



Contact details:

Niels de Vos
ndevos@redhat.com
ndevos on Freenode