

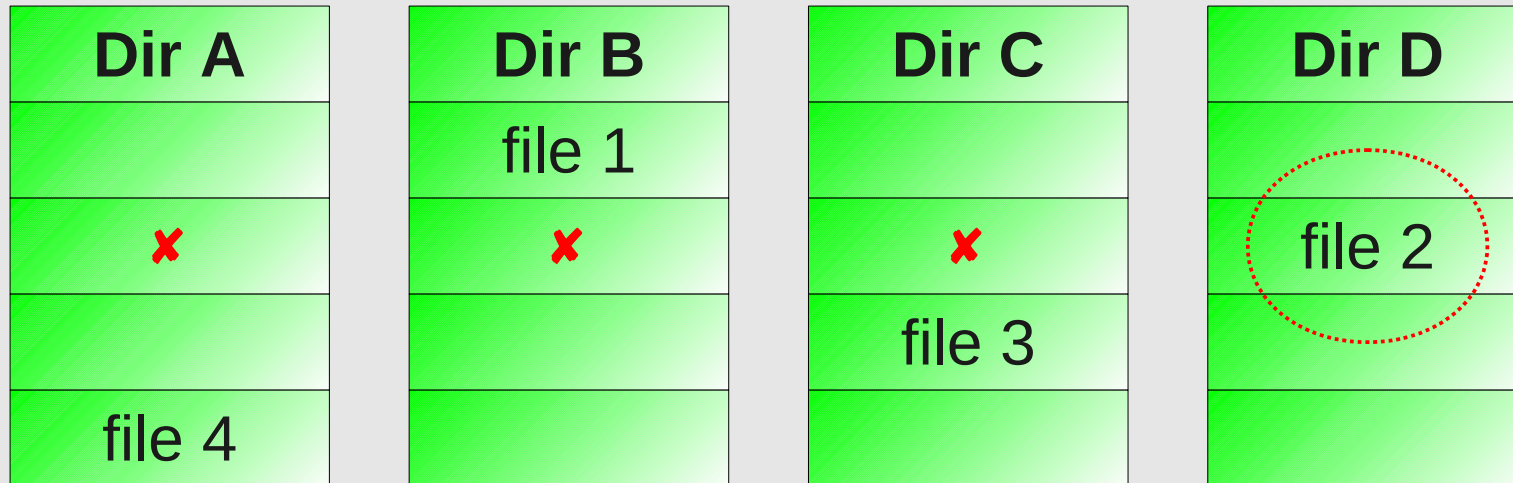
GlusterFS Translators Almost Live Tutorial

Jeff Darcy
August 28, 2012

Session Overview

- Build a “negative lookup” caching translator
 - add functionality
 - hook things together
 - build
 - configure
 - debug
- Almost live - see each step as code is written, built, and run

The Problem



- First process misses in A/B/C
- Second process has no “memory” of misses, searches from the beginning again
- Idea: cache the misses, return immediately

On the Shoulders of Giants

```
jdarcy@jdarcy-dt:~/projects/glusterfs_patches (master)
File Edit View Search Terminal Help
1226
1227 int32_t
1228 default_stat (call_frame_t *frame, xlator_t *this, loc_t *loc,
1229              dict_t *xdata)
1230 {
1231     STACK_WIND (frame, default_stat_cbk, FIRST_CHILD(this),
1232               FIRST_CHILD(this)->fops->stat, loc, xdata);
1233     return 0;
1234 }
1235
1236 int32_t
1237 default_lookup (call_frame_t *frame, xlator_t *this, loc_t *loc,
1238               dict_t *xdata)
1239 {
1240     STACK_WIND (frame, default_lookup_cbk, FIRST_CHILD(this),
1241               FIRST_CHILD(this)->fops->lookup, loc, xdata);
1242     return 0;
1243 }
1244 █
1245 int32_t
1246 default_fsetattr (call_frame_t *frame, xlator_t *this, fd_t *fd,
1247                  struct iatt *stbuf, int32_t valid,
1248                  dict_t *xdata)
1249 {
```

1244,0-1 92%

Minor Changes 1

```
jdarcy@jdarcy-dt:~/projects/negative
File Edit View Search Terminal Help
19
20 #include <ctype.h>
21 #include <sys/uio.h>
22
23 #ifndef _CONFIG_H
24 #define _CONFIG_H
25 #include "config.h"
26 #endif
27
28 #include "glusterfs.h"
29 #include "xlator.h"
30 #include "logging.h"
31
32 #include "negative.h"
33
34 int32_t
35 negative_lookup (call_frame_t *frame, xlator_t *this, loc_t *loc,
36                 dict_t *xdata)
37 {
38     STACK_WIND (frame, negative_lookup_cbk, FIRST_CHILD(this),
39               FIRST_CHILD(this)->fops->lookup, loc, xdata);
40     return 0;
41 }
42
"example.c" 42L, 1190C written                               38,35                               Bot
```

Back for Seconds

```
jdarcy@jdarcy-dt:~/projects/negative
File Edit View Search Terminal Help
22
23 #ifndef _CONFIG_H
24 #define _CONFIG_H
25 #include "config.h"
26 #endif
27
28 #include "xlator.h"
29
30 /* _CBK function section */
31
32 int32_t
33 default_lookup_cbk (call_frame_t *frame, void *cookie, xlator_t *this,
34                   int32_t op_ret, int32_t op_errno, inode_t *inode,
35                   struct iatt *buf, dict_t *xdata, struct iatt *postparent)
36 {
37     STACK_UNWIND_STRICT (lookup, frame, op_ret, op_errno, inode, buf,
38                         xdata, postparent);
39     return 0;
40 }
41
42 int32_t
43 default_stat_cbk (call_frame_t *frame, void *cookie, xlator_t *this,
44                 int32_t op_ret, int32_t op_errno, struct iatt *buf,
45                 dict_t *xdata)
46
"../glusterfs_patches/libglusterfs/src/defaults.c" 1354L, 43459C 33,1 1%
```

Minor Changes 2

```
jdarcy@jdarcy-dt:~/projects/negative
Edit View Search Terminal Help
5 #include "config.h"
6 #endif
7
8 #include "glusterfs.h"
9 #include "xlator.h"
0 #include "logging.h"
1
2 #include "negative.h"
3
4 int32_t
5 negative_lookup_cbk (call_frame_t *frame, void *cookie, xlator_t *this,
6                     int32_t op_ret, int32_t op_errno, inode_t *inode,
7                     struct iatt *buf, dict_t *xdata, struct iatt *postparent)
8 {
9     STACK_UNWIND_STRICT (lookup, frame, op_ret, op_errno, inode, buf,
0                          xdata, postparent);
1     return 0;
2 }
3
4 int32_t
5 negative_lookup (call_frame_t *frame, xlator_t *this, loc_t *loc,
6                 dict_t *xdata)
7 {
8     STACK_WIND (frame, negative_lookup_cbk, FIRST_CHILD(this),
9               "sample.c" 52L, 1564C written, 39, 30, 85%
```

Boiler Plate

```
jdarncy@jdarncy-dt:~/projects/negative
File Edit View Search Terminal Help
41     return 0;
42 }
43
44 int32_t
45 negative_lookup (call_frame_t *frame, xlator_t *this, loc_t *loc,
46                 dict_t *xdata)
47 {
48     STACK_WIND (frame, negative_lookup_cbk, FIRST_CHILD(this),
49                FIRST_CHILD(this)->fops->lookup, loc, xdata);
50     return 0;
51 }
52
53 struct xlator_fops fops = {
54     .lookup = negative_lookup,
55 };
56
57 struct xlator_cbks cbks = {
58 };
59
60 struct volume_options options[] = {
61     { .key = {NULL} },
62 };
~
~
"example.c" 62L, 1723C written                               53,1           Bot
```


init

jdarcy@jdarcy-dt:~/projects/negative

File Edit View Search Terminal Help

```
53 int32_t
54 init (xlator_t *this)
55
56     negative_private_t *priv = NULL;
57
58     if (!this->children || this->children->next) {
59         gf_log ("negative", GF_LOG_ERROR,
60             "FATAL: negative should have exactly one child");
61         return -1;
62     }
63
64     if (!this->parents) {
65         gf_log (this->name, GF_LOG_WARNING,
66             "dangling volume. check volfile ");
67     }
68
69     priv = GF_CALLOC (1, sizeof (negative_private_t), gf_negative_mt_priv);
70     if (!priv)
71         return -1;
72
73     this->private = priv;
74     gf_log ("negative", GF_LOG_DEBUG, "negative xlator loaded");
75     return 0;
76
```

76,1

82%

fini

```
jdarcy@jdarcy-dt:~/projects/negative
File Edit View Search Terminal Help
70     if (!priv)
71         return -1;
72
73     this->private = priv;
74     gf_log ("negative", GF_LOG_DEBUG, "negative xlator loaded");
75     return 0;
76 }
77
78 void
79 fini (xlator_t *this)
80 {
81     negative_private_t *priv = this->private;
82
83     if (!priv)
84         return;
85     this->private = NULL;
86     GF_FREE (priv);
87
88     return;
89 }
90
91 struct xlator_fops fops = {
92     .lookup = negative_lookup,
93 };
"example.c" 100L, 2444C written                                78,1                                90%
```

Header File

```
jdarcy@jdarcy-dt:~/projects/negative
File Edit View Search Terminal Help
23 #ifndef _CONFIG_H
24 #define _CONFIG_H
25 #include "config.h"
26 #endif
27 #include "mem-types.h"
28 #include "hashfn.h"
29
30 #define GHOST_BUCKETS 64
31 #define GHOST_HASH(x) (SuperFastHash(x,strlen(x)) % GHOST_BUCKETS)
32
33 typedef struct _ghost {
34     struct _ghost *next;
35     char *path;
36 } ghost_t;
37
38 typedef struct {
39     ghost_t *ghosts[GHOST_BUCKETS];
40 } negative_private_t;
41
42 enum gf_negative_mem_types_
43     gf_negative_mt_priv = gf_common_mt_end + 1,
44     gf_negative_mt_ghost,
45     gf_negative_mt_end
46 ;
"example.h" 46L, 1311C written 46,1 Bot
```

Updating the Cache (negatives)

jdarcy@jdarcy-dt:~/projects/negative

File Edit View Search Terminal Help

```
34 int32_t
35 negative_lookup_cbk (call_frame_t *frame, void *cookie, xlator_t *this,
36                     int32_t op_ret, int32_t op_errno, inode_t *inode,
37                     struct iatt *buf, dict_t *xdata, struct iatt *postparent)
38 {
39     negative_private_t *priv = this->private;
40     ghost_t *gp = NULL;
41     uint64_t ctx = 0;
42     uint32_t bucket = 0;
43
44     inode_ctx_get(inode, this, &ctx);
45     if (op_ret < 0) {
46         gp = GF_CALLOC(1, sizeof(ghost_t), gf_negative_mt_ghost);
47         if (gp) {
48             gp->path = (char *)ctx;
49             bucket = GHOST_HASH(gp->path);
50             /* TBD: locking */
51             gp->next = priv->ghosts[bucket];
52             priv->ghosts[bucket] = gp;
53             gf_log(this->name, GF_LOG_DEBUG, "added %s",
54                 (char *)ctx);
55             goto unwind;
56         }
57     }
```

"example.c" 127L, 3458C written

55,1

32%

Updating the Cache (positives)

```
jdarcy@jdarcy-dt:~/projects/negative
File Edit View Search Terminal Help
46     gp = GF_CALLOC(1, sizeof(ghost_t), gf_negative_mt_ghost);
47     if (gp) {
48         gp->path = (char *)ctx;
49         bucket = GHOST_HASH(gp->path);
50         /* TBD: locking */
51         gp->next = priv->ghosts[bucket];
52         priv->ghosts[bucket] = gp;
53         gf_log(this->name, GF_LOG_DEBUG, "added %s",
54              (char *)ctx);
55         goto unwind;
56     }
57 }
58 else {
59     gf_log(this->name, GF_LOG_DEBUG, "found %s", (char *)ctx);
60     exorcise(this, (char *)ctx);
61 }
62
63 /* Both positive result and allocation failure come here. */
64 GF_FREE((void *)ctx);
65
66 STACK_UNWIND_STRICT (lookup, frame, op_ret, op_errno, inode, buf,
67                     xdata, postparent);
68 return 0;
69 }
"example.c" 127L, 3458C written                                47,25                                43%
```

Checking the Cache

jdarcy@jdarcy-dt:~/projects/negative

File Edit View Search Terminal Help

```
72 negative_lookup (call_frame_t *frame, xlator_t *this, loc_t *loc,
73                 dict_t *xdata)
74 {
75     negative_private_t *priv    = this->private;
76     ghost_t            *gp      = NULL;
77     uint32_t          bucket = 0;
78
79     bucket = GHOST_HASH(loc->path);
80     for (gp = priv->ghosts[bucket]; gp; gp = gp->next) {
81         if (!strcmp(gp->path, loc->path)) {
82             gf_log(this->name, GF_LOG_DEBUG, "%s (%p) => HIT",
83                   loc->path, loc->inode);
84             STACK_UNWIND_STRICT (lookup, frame, -1, ENOENT,
85                                 NULL, NULL, NULL, NULL);
86             return 0;
87         }
88     }
89     gf_log(this->name, GF_LOG_DEBUG, "%s (%p) => MISS",
90           loc->path, loc->inode);
91     inode_ctx_put(loc->inode, this, (uint64_t)gf_strdup(loc->path));
92     STACK_WIND (frame, negative_lookup_cbk, FIRST_CHILD(this),
93               FIRST_CHILD(this)->fops->lookup, loc, xdata);
94     return 0;
95 }
```

"example.c" 144L, 4246C written

92,9

59%

Final Pieces (create and mkdir)

```
jdarcy@jdarcy-dt:~/projects/negative
File Edit View Search Terminal Help
97 int32_t
98 negative_create_cbk (call_frame_t *frame, void *cookie, xlator_t *this,
99                     int32_t op_ret, int32_t op_errno, fd_t *fd, inode_t *inode,
100                     struct iatt *buf, struct iatt *preparent,
101                     struct iatt *postparent)
102 {
103     uint64_t          ctx = 0;
104
105     inode_ctx_get(inode, this, &ctx);
106     exorcise(this, (char *)ctx);
107     GF_FREE((void *)ctx);
108
109     STACK_UNWIND_STRICT (create, frame, op_ret, op_errno, fd, inode, buf,
110                         preparent, postparent);
111     return 0;
112 }
113
114 int32_t
115 negative_create (call_frame_t *frame, xlator_t *this, loc_t *loc, int32_t flags,
116                mode_t mode, fd_t *fd, dict_t *params)
117 {
118     inode_ctx_put(loc->inode, this, (uint64_t)gf_strdup(loc->path));
119     STACK_WIND (frame, negative_create_cbk, FIRST_CHILD(this),
120               FIRST_CHILD(this)->fops->create, loc, flags, mode, fd,
"example.c" 199L, 6123C written                               120,1          54%
```

Time to Build

jdarcy@jdarcy-dt:~/projects/negative (master)

File Edit View Search Terminal Help

```
1 # Change these to match your source code.
2 TARGET = negative.so
3 OBJECTS = negative.o
4
5 # Change these to match your environment.
6 GLFS_SRC = /root/glusterfs_patches
7 GLFS_ROOT = /usr/lib64
8 GLFS_VERS = 3git
9 HOST_OS = GF_LINUX_HOST_OS
10
11 # You shouldn't need to change anything below here.
12
13 CFLAGS = -fPIC -Wall -O0 -g \
14         -DHAVE_CONFIG_H -D_FILE_OFFSET_BITS=64 -D_GNU_SOURCE -D$(HOST_OS) \
15         -I$(GLFS_SRC) -I$(GLFS_SRC)/libglusterfs/src \
16         -I$(GLFS_SRC)/contrib/uuid -I.
17 LDFLAGS = -shared -nostartfiles -L$(GLFS_ROOT) -lglusterfs -lpthread
18
19 $(TARGET): $(OBJECTS)
20     $(CC) $(CFLAGS) $(OBJECTS) $(LDFLAGS) -o $(TARGET)
21
22 install: $(TARGET)
23     cp $(TARGET) $(GLFS_ROOT)/glusterfs/$(GLFS_VERS)/xlator/features/
24
"Makefile" 26L, 737C 1,1 Top
```


Almost There

root@gfs-i8c-01:~/negative

File Edit View Search Terminal Help

```
[root@gfs-i8c-01 negative]# make install
cc -fPIC -Wall -O0 -g -DHAVE_CONFIG_H -D_FILE_OFFSET_BITS=64 -D_GNU_SOURCE -DGF_LINUX
_HOST_OS -I/root/glusterfs_patches -I/root/glusterfs_patches/libglusterfs/src -I/root
/glusterfs_patches/contrib/uuid -I. -c -o negative.o negative.c
cc -fPIC -Wall -O0 -g -DHAVE_CONFIG_H -D_FILE_OFFSET_BITS=64 -D_GNU_SOURCE -DGF_LINUX
_HOST_OS -I/root/glusterfs_patches -I/root/glusterfs_patches/libglusterfs/src -I/root
/glusterfs_patches/contrib/uuid -I. negative.o -shared -nostartfiles -L/usr/lib64 -lg
lusterfs -lpthread -o negative.so
cp negative.so /usr/lib64/glusterfs/3git/xlator/features/
[root@gfs-i8c-01 negative]# cp /var/lib/glusterd/vols/test/test-fuse.vol .
[root@gfs-i8c-01 negative]# █
```

Original Volfile

root@gfs-i8c-01:~/negative

File Edit View Search Terminal Help

```
27     subvolumes test-write-behind
28 end-volume
29
30 volume test-io-cache
31     type performance/io-cache
32     subvolumes test-read-ahead
33 end-volume
34
35 volume test-quick-read
36     type performance/quick-read
37     subvolumes test-io-cache
38 end-volume
39
40 volume test-md-cache
41     type performance/md-cache
42     subvolumes test-quick-read
43 end-volume
44
45 volume test
46     type debug/io-stats
47     option latency-measurement off
48     option count-fop-hits off
49     subvolumes test-md-cache
50 end-volume
"test-fuse.vol" 50L, 1031C written
```

Modified Volfile

root@gfs-i8c-01:~/negative

File Edit View Search Terminal Help

```
32     subvolumes test-read-ahead
33 end-volume
34
35 volume test-quick-read
36     type performance/quick-read
37     subvolumes test-io-cache
38 end-volume
39
40 volume test-md-cache
41     type performance/md-cache
42     subvolumes test-quick-read
43 end-volume
44
45 volume test-negative
46     type features/negative
47     subvolumes test-md-cache
48 end-volume
49
50 volume test
51     type debug/io-stats
52     option latency-measurement off
53     option count-fop-hits off
54     subvolumes test-negative
55 end-volume
"test-fuse.vol" 55L, 1120C written
```

Running in a Debugger

```
root@gfs-i8c-01:~/negative
File Edit View Search Terminal Help
[root@gfs-i8c-01 negative]# make install
cc -fPIC -Wall -O0 -g -DHAVE_CONFIG_H -D_FILE_OFFSET_BITS=64 -D_GNU_SOURCE -DGF_LINUX
_HOST_OS -I/root/glusterfs_patches -I/root/glusterfs_patches/libglusterfs/src -I/root
/glusterfs_patches/contrib/uuid -I. -c -o negative.o negative.c
cc -fPIC -Wall -O0 -g -DHAVE_CONFIG_H -D_FILE_OFFSET_BITS=64 -D_GNU_SOURCE -DGF_LINUX
_HOST_OS -I/root/glusterfs_patches -I/root/glusterfs_patches/libglusterfs/src -I/root
/glusterfs_patches/contrib/uuid -I. negative.o -shared -nostartfiles -L/usr/lib64 -lg
lusterfs -lpthread -o negative.so
cp negative.so /usr/lib64/glusterfs/3git/xlator/features/
[root@gfs-i8c-01 negative]# cp /var/lib/glusterd/vols/test/test-fuse.vol .
[root@gfs-i8c-01 negative]# vi test-fuse.vol
[root@gfs-i8c-01 negative]# gdb glusterfs
GNU gdb (GDB) Red Hat Enterprise Linux (7.2-50.el6)
Copyright (C) 2010 Free Software Foundation, Inc.
License GPLv3+: GNU GPL version 3 or later <http://gnu.org/licenses/gpl.html>
This is free software: you are free to change and redistribute it.
There is NO WARRANTY, to the extent permitted by law.  Type "show copying"
and "show warranty" for details.
This GDB was configured as "x86_64-redhat-linux-gnu".
For bug reporting instructions, please see:
<http://www.gnu.org/software/gdb/bugs/>...
Reading symbols from /usr/sbin/glusterfs...(no debugging symbols found)...done.
(gdb) r --debug -f test-fuse.vol /import/test
```

Setting First Breakpoint

```
root@gfs-i8c-01:~/negative
File Edit View Search Terminal Help
[2012-08-17 17:23:21.159021] D [fuse-bridge.c:4013:fuse_get_mount_status] 0-fuse: mount status is 0
[2012-08-17 17:23:21.159047] D [dht-diskusage.c:80:dht_du_info_cbk] 0-test-dht: on subvolume 'test-client-1': avail_percent is: 99.00 and avail_space is: 146706644992 and avail_inodes is: 99.00
[New Thread 0x7ffff15b2700 (LWP 16445)]
[2012-08-17 17:23:21.159871] I [fuse-bridge.c:3514:fuse_init] 0-glusterfs-fuse: FUSE inited with protocol versions: glusterfs 7.13 kernel 7.13
[2012-08-17 17:23:21.159920] D [negative.c:141:negative_lookup] 0-test-negative: 00000000-0000-0000-0000-0000-0000000000001/ => MISS
[2012-08-17 17:23:21.161108] D [negative.c:86:negative_lookup_cbk] 0-test-negative: found 00000000-0000-0000-0000-0000-000000000000/
^C
Program received signal SIGINT, Interrupt.
0x0000003a002e5d03 in epoll_wait () from /lib64/libc.so.6
Missing separate debuginfos, use: debuginfo-install glusterfs-3git-1.el6.x86_64
(gdb) b negative_lookup_cbk
Breakpoint 1 at 0x7ffff3ba4d7a: file negative.c, line 67.
(gdb) c
Continuing.
[2012-08-17 17:24:03.361634] D [client-handshake.c:184:client_start_ping] 0-test-client-0: returning as transport is already disconnected OR there are no frames (0 || 0)
[2012-08-17 17:24:03.361663] D [client-handshake.c:184:client_start_ping] 0-test-client-1: returning as transport is already disconnected OR there are no frames (0 || 0)
█
```

First Failed Lookup

```
root@gfs-i8c-01:~/negative
File Edit View Search Terminal Help
0000-0000-0000-0000-000000000000/fubar => MISS
Breakpoint 1, negative_lookup_cbk (frame=0x7ffff727c0d8, cookie=0x7ffffe4000920,
  this=0x84b1a0, op_ret=-1, op_errno=2, inode=0x0, buf=0x0, xdata=0x0,
  postparent=0x0) at negative.c:67
67         negative_private_t *priv  = this->private;
(gdb) n
68         ghost_t             *gp    = cookie;
(gdb)
69         uint32_t            bucket = 0;
(gdb)
71         if (!gp) {
(gdb)
76         if (op_ret < 0) {
(gdb)
77             bucket = GHOST_HASH(gp->gfid, gp->name);
(gdb) p gp->name
$1 = 0x7ffffe4000950 "fubar"
(gdb) n
79         gp->next = priv->ghosts[bucket];
(gdb)
80         priv->ghosts[bucket] = gp;
(gdb)
81         gf_log (this->name, GF_LOG_DEBUG,
(gdb)
```

Cache Kicks In

```
root@gfs-i8c-01:~/negative
File Edit View Search Terminal Help
Continuing.
[2012-08-17 17:35:21.603276] D [client-handshake.c:184:client_start_ping] 0-test-client-0: returning as transport is already disconnected OR there are no frames (0 || 0)
[2012-08-17 17:35:21.603331] D [negative.c:106:negative_lookup] 0-test-negative: filling in loc->pargfid
Breakpoint 2, negative_lookup (frame=0x7ffff727c0d8, this=0x84b1a0,
    loc=0x7ffffe4001a70, xdata=0x0) at negative.c:118
118         gf_log(this->name,GF_LOG_DEBUG,"%s/%s => HIT (%p)",
(gdb) list
113             continue;
114         }
115         if (strcmp(gp->name,loc->name)) {
116             continue;
117         }
118         gf_log(this->name,GF_LOG_DEBUG,"%s/%s => HIT (%p)",
119                 uid_utoa(loc->gfid), loc->name, loc->inode);
120         STACK_UNWIND_STRICT (lookup, frame, -1, ENOENT,
121                             NULL, NULL, NULL, NULL);
122         return 0;
(gdb) p *loc
$1 = {path = 0x7ffffe4000990 "/fubar", name = 0x7ffffe4000970 "fubar",
    inode = 0x7ffff238d0c0, parent = 0x7ffff238d02c,
    gfid = '\000' <repeats 15 times>, pargfid = '\000' <repeats 15 times>, "\001"}
(gdb) █
```

How Well Does It Work?

- Simulate PHP workload
 - 1000 files spread across 10 directories
 - 80% of requests to 10% of files
 - test time for 10000 searches
- Numbers for comparison only
 - debug build
 - wimpy test machines (GigE)

Getting a Baseline

```
root@gfs-i8c-04:/import/test
File Edit View Search Terminal Help
[root@gfs-i8c-04 negative]# mount -t glusterfs gfs1:test /import/test
[root@gfs-i8c-04 negative]# cd /import/test
[root@gfs-i8c-04 test]# ~/phpsim/mkdirs.py &> /dev/null
[root@gfs-i8c-04 test]# ls
dir0 dir2 dir4 dir6 dir8 file1 file11 file3
dir1 dir3 dir5 dir7 dir9 file10 file2 file4
[root@gfs-i8c-04 test]# time ~/phpsim/worker.py
average latency = 7.357ms

real    1m13.723s
user    0m2.464s
sys     0m1.566s
[root@gfs-i8c-04 test]#
```

Are You Ready?

```
root@gfs-i8c-04:/import/test
File Edit View Search Terminal Help
[root@gfs-i8c-04 ~]# cd ~/negative
[root@gfs-i8c-04 negative]# glusterfs -f test-fuse.vol /import/test
[root@gfs-i8c-04 negative]# cd /import/test
[root@gfs-i8c-04 test]# time ~/phpsim/worker.py
average latency = 1.095ms

real    0m11.089s
user    0m2.560s
sys     0m1.239s
[root@gfs-i8c-04 test]#
```

Interlude: glupy

- Write translators in Python
- Uses extending/embedding, ctypes module
- Brand new project
 - weekend hacking, mid-August 2012
 - still implementing 70+ operation types
- Article coming soon in \$major_publication

Python Lookup Function

```
root@gfs-i8c-04:~/glupy
File Edit View Search Terminal Help
# TBD: we need a better way of handling per-request data (frame->local in C).
requests = {}
dl.get_id.restype = c_long
dl.get_id.argtypes = [ POINTER(call_frame_t) ]

def uuid2str (orig):
    return "%08x%08x%08x%08x" % (orig[0], orig[1], orig[2], orig[3])

@lookup_fop_t
def lookup_fop (frame, this, loc, xdata):
    pargfid = uuid2str(loc.contents.pargfid)
    log ("lookup FOP: %s:%s" % (pargfid, loc.contents.name))
    # Check the cache.
    if cache.has_key(pargfid) and (loc.contents.name in cache[pargfid]):
        log ("short-circuiting for %s:%s" % (pargfid, loc.contents.name))
        dl.unwind_lookup(frame, 0, this, -1, 2, None, None, None, None)
        return 0
    key = dl.get_id(frame)
    requests[key] = (pargfid, loc.contents.name[:])
    # TBD: get real child xl from init, pass it here
    dl.wind_lookup(frame, POINTER(xlator_t)(), loc, xdata)
    return 0

"negative.py" line 43 of 98 --43%-- col 1
```

Python Callback Function

```
root@gfs-i8c-04:~/glupy
File Edit View Search Terminal Help
@lookup_cbk_t
def lookup_cbk (frame, cookie, this, op_ret, op_errno,
                inode, buf, xdata, postparent):
    log ("lookup CBK: %d (%d)" % (op_ret, op_errno))
    key = dl.get_id(frame)
    pargfid, name = requests[key]
    # Update the cache.
    if op_ret == 0:
        log ("found %s, removing from cache" % name)
        if cache.has_key(pargfid):
            cache[pargfid].discard(name)
    elif op_errno == 2: # ENOENT
        log ("failed to find %s, adding to cache" % name)
        if cache.has_key(pargfid):
            cache[pargfid].add(name)
        else:
            cache[pargfid] = set([name])
    del requests[key]
    dl.unwind_lookup(frame, cookie, this, op_ret, op_errno,
                    inode, buf, xdata, postparent)
    return 0

@create_fop_t
def create_fop (frame, this, loc, flags, mode, umask, fd, xdata):
    "negative.py" line 62 of 98 --63%-- col 2-9
```

Python Setup

```
root@gfs-i8c-04:~/glupy
File Edit View Search Terminal Help
@create_cbk_t
def create_cbk (frame, cookie, this, op_ret, op_errno,
                fd, inode, buf, preparent, postparent, xdata):
    log ("create CBK: %d (%d)" % (op_ret, op_errno))
    key = dl.get_id(frame)
    pargfid, name = requests[key]
    # Update the cache.
    if op_ret == 0:
        log ("created %s, removing from cache" % name)
        if cache.has_key(pargfid):
            cache[pargfid].discard(name)
    del requests[key]
    dl.unwind_create(frame, cookie, this, op_ret, op_errno,
                    fd, inode, buf, preparent, postparent, xdata)
)
    return 0

class xlator ():
    def __init__ (self, xl):
        dl.set_lookup_fop(xl, lookup_fop)
        dl.set_lookup_cbk(xl, lookup_cbk)
        dl.set_create_fop(xl, create_fop)
        dl.set_create_cbk(xl, create_cbk)

"negative.py" line 98 of 98 --100%-- col 1
```

Yet Another Volfile

```
root@gfs-i8c-04:~/glupy
File Edit View Search Terminal Help
end-volume
volume test-quick-read
    type performance/quick-read
    subvolumes test-io-cache
end-volume
volume test-md-cache
    type performance/md-cache
    subvolumes test-quick-read
end-volume
volume test-negative
    type features/glupy
    option module-name negative
    subvolumes test-md-cache
end-volume
volume test
    type debug/io-stats
    option latency-measurement off
    option count-fop-hits off
    subvolumes test-negative
end-volume
"test-fuse.vol" line 56 of 56 --100%-- col 1
```

Scripting Languages Aren't So Slow

```
root@gfs-i8c-04:/import/test
File Edit View Search Terminal Help
[root@gfs-i8c-04 glupy]# export PYTHONPATH=$PWD
[root@gfs-i8c-04 glupy]# glusterfs -f test-fuse.vol /import/test
[root@gfs-i8c-04 glupy]# cd /import/test
[root@gfs-i8c-04 test]# time ~/phpsim/worker.py
average latency = 1.286ms

real    0m12.989s
user    0m2.076s
sys     0m1.092s
[root@gfs-i8c-04 test]#
```


Conclusion

- Even simple translators can produce big effects
- Writing translators isn't that hard
- Writing glupy translators will be even easier
- What would you like GlusterFS to do today?

Resources

<http://hekafs.org/index.php/2011/11/translator-101-class-1-setting-the-stage/>

<http://hekafs.org/index.php/2011/11/translator-101-lesson-2-init-fini-and-private-context/>

<http://hekafs.org/index.php/2011/11/translator-101-lesson-3-this-time-for-real/>

<http://hekafs.org/index.php/2011/11/translator-101-lesson-4-debugging-a-translator/>

<https://github.com/jdarcy/negative-lookup>

<https://github.com/jdarcy/glupy>

<http://www.gluster.org/>